Introduction
1. Introduction

The successful completion of the Human Genome Project (HGP) and availability of complete genome sequences of pathogens, in combination with Genomics, Proteomics, Cheminformatics, Metabolomics, Interactomics, Systems biology and Drug designing, has revolutionized the field of Drug-discovery to identify the biomarkers and to develop vaccines against human pathogens. Development of various effective tools to analyze the molecules from the genome to proteome level has made a major impact on the advancement of in silico based studies on which we can rely upon.

The application of newer molecular and genomics research techniques/tools with the modern approach like Synteny Based Comparative Genomics (SBCG) (Lemoine et al., 2008) have helped to find precisely the important genes which are conserved among the Chlamydiaceae family members and also other respiratory pathogens which are also involved in causing the respiratory ailments like upper and lower respiratory infections, pneumonia and COPD with other systemic diseases which can be identified as therapeutic candidates.

The computational approach of studying newer molecular and genomics research techniques with the modern approach like comparative metabolic pathway analysis (Butt et al., 2012) and protein-protein interaction network studies are contributing to find precisely the important proteins/enzymes, which interact and play a role of pathogenicity in many infectious and systemic diseases which can be concluded as potential drug targets (Kushwaha et al., 2010).
Chapter 1

Introduction

The genus Chlamydia was established in 1966. *Chlamydophila pneumoniae* (*Cpn*) was separated as a distinct species in 1992. It is perhaps one among the successful Chlamydial species that have established a forte in a range homoeothermic and poikilothermic hosts, including humans, animals, amphibians and reptiles (Bodetti *et al.*, 2002). It is an obligate intracellular gram-negative bacterium, which possess a distinctive and a composite biphasic developmental cycle. It has been reported as TWAR (Taiwan acute respiratory) organism as it was isolated from Taiwan in acute respiratory isolate, also known as AR39 and along with other human pathogens like *Chlamydia trachomatis* (*Ct*) and *Chlamydia psitassi* (*Cps*).

*Cpn* is a common cause of upper and lower respiratory tract infections and pneumonia and has been associated with several chronic inflammatory conditions such as atherosclerosis and chronic obstructive pulmonary disease (COPD). It is also being actively investigated as a cause of several systemic diseases like COPD (Kuo *et al.*, 1995), chronic asthma (Grayston *et al.*, 1993), atherosclerosis, stroke (Sriram *et al.*, 2005), coronary artery disease, multisclerosis, Alzheimer’s disease (Balin *et al.*, 2008) and lung cancer etc. It is a very common bacterium worldwide, and almost everyone is infected at some point of their life.

The prevalence of COPD has emerged as the major cause of morbidity and mortality rate globally and it is anticipated as 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) (Murray & Lopez, 1997). The region wise projection of developing countries like India were even worse as published in GOLD report, 2013 (www.who.int/evidence, 2013). India contributes very
Chapter 1

Introduction

significantly to mortality from COPD 102.3 per 100,000 and 6,740,000 DALYs out of a world’s total of 27,756,000 DALYs; thus significantly affecting health related quality of life in India (Bhome, 2012).

It has been extremely difficult to diagnose and impossible to treat with current antibiotics. Thus, development of safe and effective vaccines, which represents a cost-effective approach that will have a greater impact on the inhibition of Chlamydia infections, is necessary by identifying new antichlamydial therapeutic targets. The present study mainly focuses on identifying novel therapeutic targets of pathogen using various strategies by efficiently utilizing the available genomic information. The genomic information is stored in large public repositories like NCBI (National Centre for Biotechnology Information), EMBL (European Molecular Biology Laboratory) and DDBJ (DNA DataBank of Japan). By using the latest developed advanced tools which are effective in annotating the genome information have paved a way to successful entry into the drug discovery pipeline.

Identification of the unique pathways has been an important strategy which is specific to the pathogen. By using comparative metabolic pathway analysis, the essentiality of enzymes is identified as they contribute to the survival of the pathogenic organisms. Later the shortlisted proteins can be considered for prioritization to conclude them as potential therapeutic targets. The genomic era allows better understanding of genomic data and is extremely vital for any study. The annotation of this information about genes/proteins that regulate important functions in organisms plays a pivotal role in combating the pathogens.
Recently, due to the development of novel tools, algorithms and methods, it has been able to predict protein functions and protein-protein interactions, which are experimentally proved to study the protein-protein interaction peaks. It is possible to model the protein structure of the drug targets by using available templates in PDB database (www.rcsb.org) (Sussman et al., 1998). Many effective tools like Modeler and Rosetta are helpful to model the structure. Screening of the approved drugs is important steps which are already available in the market using the database like DrugBank (Knox et al., 2011).

Docking studies helps to know the interaction of drug and the therapeutic target protein by docking. It is useful in calculating the binding energy and inhibition factors, and also to analyze the bonds formed between the drug molecule and the therapeutic target using tools like AutoDock (Seeliger et al., 2010) and Pymol (Delano, 2002).

The insilico approach adopted to identify the putative therapeutic targets will provide a deeper insight to the genes/proteins which are involved in pathogenicity of *C. pneumoniae* by acting as key virulence factors in inducing the disease by infecting the host. These identified therapeutic targets using various above discussed strategies will certainly facilitate in reducing the cost and time required to identify the effective therapeutics against the deadly pathogen.

1.2 Objectives of the study:

A comprehensive in silico analysis of *Cpn* has the potential to identify proteins that may be successfully targeted. This makes the primal goal of the thesis. *In silico* methods have the great advantage of speed, low cost, and provide a genomic view of the whole microbe at once, which
Chapter 1

is capable of answering the modern need of biomedical research, that are often difficult to address experimentally.

1. To perform genome analysis of *Chlamydyphila pneumoniae* (finding the homologous genes and non-homologous genes) by comparing among the other species of *Chlamydial* organisms.

2. To annotate all the uncommon/non-orthologous genes of *Chlamydyphila pneumoniae*, possibly because of which it gets distinguished as a causative organism for infective COPD, and to identify all the genes which code for secreted proteins among those identified set of uncommon genes.

3. To understand the interaction peak of the proteins between antigenic virulent factors of *Chlamydyphila pneumoniae* and the causative protein factors of chronic bronchitis. To conduct a comparative study of interaction peaks of the antigenic virulent factors of *Chlamydyphila pneumoniae* (*Chlamydia pneumonia*) and causative proteins for COPD.