CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

Identification of Potential Drug Targets in *Chlamydia pneumoniae* and Inhibition with Plant Derive Natural Compounds

by

Amna Noor

A thesis submitted in partial fulfillment for the degree of Master of Science

in the

Faculty of Health and Life Sciences
Department of Bioinformatics and Biosciences

2021

I dedicated this thesis to my beloved parents and my teachers.

CERTIFICATE OF APPROVAL

Identification of Potential Drug Targets in *Chlamydia pneumoniae* and Inhibition with Plant Derive Natural Compounds

by

Amna Noor

(MBS193018)

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Abdur Rahman	NUST, Islamabad
(b)	Internal Examiner	Dr. Erum Dilshad	CUST, Islamabad
(c)	Supervisor	Dr. Syeda Marriam Bakhtiar	CUST, Islamabad

Dr. Syeda Marriam Bakhtiar
Thesis Supervisor
November, 2021

Dr. Sahar Fazal
Head
Dept. of Bioinfo. and Biosciences
November, 2021

Dr. M. Abdul Qadir
Dean
Faculty of Health and Life Sciences
November, 2021

Author's Declaration

I, **Amna Noor** hereby state that my MS thesis titled **"Identification of Potential Drug Targets in *Chlamydia pneumoniae* and Inhibition with Plant Derive Natural Compounds"** is my own work and has not been submitted previously by me for taking my degree from Capital University of Science and Technology, Islamabad or anywhere in the country or Abroad. At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my MS Degree.

(Amna Noor)
Registration No: MBS193018

Plagiarism Undertaking

I hereby declare that this research work under this thesis title **"Identification of Potential Drug Targets in *Chlamydia pneumoniae* and Inhibition with Plant Derive Natural Compounds"**, neither as a whole nor as a part there has been copied out from any source, wherever references have been provided. It is further declared that we have developed this graduate project and accompanied report entirely on the basis of our knowledge. If any part of this graduate project is proved to be copied out from any source and found to be reproduced from other sources without proper acknowledgement ,we will stand by the consequences, and the university will take no responsibility.

(Amna Noor)
Registration No: MBS193018

List of Publications

It is certified that no publication(s) have been made out of the research work that has been carried out for this thesis:-

(Amna Noor)

Registration No: MBS193018

Acknowledgement

I would like to thanks ALLAH Almighty. The Most Magnificent and Compassionate , indeed, all praises are due to Him and His Holy Prophet Muhammad (PBUH). They gave us strength and aptitude to complete this target.

I want to acknowledge the efforts of my supervisor Dr. Syeda Marriam Bakhtiar, Assistant Professor, Capital University of Science and Technology , Islamabad, whose encouragements, guidance and support helped me to complete my thesis.

I want to acknowledge Head of Department of Health and Life sciences and Biosciences, Dr Sahar Fazal for giving me the opportunity to pursue my degree with research thesis. I owe a great deal of appreciation and gratitude to all faculty members Dr. Shoukat Iqbal Malik, Dr. Erum Dilshad, Dr. Arshia Butt for their guidance.

A special thanks to my all friends and especially to my best friend batch mate Iqra Riasat who always guide me regarding research work and other friends Maria Khadija , Izhar , Jibran for their support, coordination and help.

At the end I deeply gratefully thanks to my Parents my siblings for their countless contributions and moral support , encouragement which brightens my vision.

(Amna Noor)

Abstract

Chlamydia pneumoniae is amongst the commonly occurring Gram negative, intracellular bacterial pathogen found all across the globe. It is reported to be associated with asthma, arthritis, atherosclerosis and bronchitis along with various other respiratory tract infections. The treatment against this bacterial infection is usually comprising general purpose antibiotics, due to which the infection may reappear again. Therefore, this research was designed with the objective to identify drug targets which are common in all the strains. For further analysis, 12 strains were selected based on availability of complete and annotated genome, the core genome of all the selected strains was identified and 942 genes were found to be present in all the strains. Subtractive genomic analysis was performed on the core genes and non-homologous genes were retrieved. In total 16 essential, core and non-homologous genes were prioritized. These 16 genes were functionally annotated and were found to play important role in the growth and survival of the C. pneumoniae as well as in disease development and progression. As drug target 12 genes were short listed based on their sub cellular localization and function. These potential drug targets were docked against plant derived natural compounds which were nine in number. the results of the research would contribute towards devising more effective and efficient treatment against pneumonia caused by Chlamydia

Keywords: *Chlamydia pneumoniae*, drug target Identification, Pangenome, subtractive approach, target Priotization.

Contents

List of Figures

List of Tables

Abbreviations

Blast n	Basic local alignment research tool for nucleotides
C.P	Chlamydia pneumoniae
CELLO	Cellular Localization of the essential proteins
Chimera	Visualization and analysis of molecular structure
DEG	Database of Essential Genes
Edgar	Efficient Database framework for comparative Genome Analysis
KEGG	Kyoto Encyclopedia of Genes and Genomes
MOE	Molecular Operating Environment
NCBI	National Center for Biotechnology Information
PAIDB	Pathogenicity Island Database
Uniprot	Universal Protein Resource
Vfdb	Virulence factor database

Chapter 1

Introduction

Pneumonia is an infection which can damaged lungs and it can be caused by bacteria and viruses. The infection cause inflammation in the air sacs of the lungs which are known as alveoli. Inflammation caused difficulty in breathing. It is a contagious disease it can spread from person to person [1]. The bacterial and viral pneumonia is airborne that they can spread from a sneeze or cough and by the surface objects. The symptoms of the pneumonia include cough, fever, sweating, shortness of the breath and chest pain. The most common cause of the bacterial pneumonia is *Streptococcus pneumoniae*. Another type of the bacterial pneumonia is *C. pneumoniae*. The general characteristics of *C. pneumoniae* includes that it is rod shaped species and it is Gram negative bacteria. It has been associated with asthma, arthritis, and atherosclerosis. bronchitis, respiratory infection, coronary heart disease in humans [1].

The prevalence of the *C. pneumoniae* have been shown that about 50% of the population have affected worldwide. The prevalence estimates were similar to those published in other population-based studies, with prevalence rates ranging from 4.1 percent to 5.2 percent. It is parasitic organisms that are unable to reproduce outside of the host cell it depends on the host for its survival [2]. The genome structure of the *C. pneumoniae* is sequenced in 1999 in United States the most common and first discovered strain (CWL029) has been sequenced in its entirety with other strains. Circular DNA makes up 1,230,230 base pairs in the

genome [3]. A total of 1,052 protein genes and 43 RNA genes are found in the human genome. There have been no plasmids discovered so far.

The pathology of the *C. pneumoniae* is that the bacteria's basic form is transferred from small water droplets into the host's lungs, where it is phagocytized and transformed into cells. When they take the elementary body, it converts into the reticulate body, which can reproduce itself within a cell [3]. The optimum temperature where the replication is done is 37 C. *C. pneumoniae* can infect the reptiles such as snakes, frogs, turtle's iguanas and also infects the mammals like koalas. Dry cough, weariness, pain on one side of the chest, fever, loss of appetite, and aches are some of the symptoms. There is no drug or vaccine is available yet which can properly cure the diseases. Studies have been shown that clinicians can treat these diseases by first line therapy in which Tetracycline's and Flouroquionolones should be given [3].

C. pneumoniae is a pathogenic bacterium and causes the upper tract respiratory diseases. *C. pneumoniae* causes diseases by damaging the lining of the windpipes, throat and lungs [1]. It belongs to the Chlamydiaceae family and classified as Chlamydiae and its scientific name is Chlamydophila pneumoniae. The incubation period of the *C. pneumoniae* is between 3 to 4 weeks. The *C. pneumoniae* is very different from the other bacteria's which caused the pneumonia.

The *C. pneumoniae* is primary infection and the prevalence of the *C. pneumoniae* is 75% worldwide [2] and it can cause the reinfection which have been reported in previous studies [3]. *Genus Chlamydia* are obligate eu-bacterial parasites and they are further classified into four species. *C. pneumoniae* is one of them it is a intracellular Gram negative pathogenic bacteria. All classes Chlamydia share common characteristics. The organisms grow in the specialized vacuole in the Golgi apparatus vesicular compartment of the eukaryotic cell [1].

It is a parasitic bacterium and it have highly streamlined genome. *C. pneumoniae* is capable of persisting in the immune cells of the host. Infection causes host body to respond a number of pro-inflammatory cytokines, which are thought to have a role in disease pathogenesis. The *C. pneumoniae* genome contains 1230

nucleotide and 1052 estimated protein coding genes. *C. pneumoniae* has no extra chromosomal identified elements. The compact genome of *C. pneumoniae* makes it suitable for the rapid genomic sequence analysis [2].

Members of the Chlamydia phylum have a biphasic developmental cycle that is unique in the bacterial kingdom because it consists of an extracellular infectious elementary body and an intracellular metabolically active non-infectious reticulate body that can replicate inside inclusions, which are parasite-o-porous vacuoles derived from the host [1]. *C. pneumoniae* is found from USA, Taiwan, Germany, Australia, Belgium, Syria. Epidemiologic surveys have demonstrated that more than 70% of the adults have been exposed to this onset diseases during their lifetimes. The Prevalence rate of this diseases is start to rise from the childhood to the adults. The most rapid rise is occurred during the age limit from 5 to 20 years that is from early childhood to the adults [3]. *C. pneumoniae* was first identify as the causative agent of human pneumonia epidemics in 1985 later it was associated with a range of the other chronic dieses such as asthma, CVD, Alzheimer's diseases. The first report of the *C. pneumoniae* infection is in non-human host which is isolated from horses in 1993 [3]. it is also identified that the *Genus Chlamydia* species also infected the koalas in 1994 [3]. It also has been found in studies that *Genus Chlamydia* has been identified worldwide that they cause infection in wide range in the cold blooded and warm-blooded species.

Genus Chlamydia have distinguishing characteristics that they have their compact and highly conserved genomes in spite they have a wide range of the hosts and they caused diseases all members of that phylum share their characteristics i.e.; they have highly conserved core genome of over 500 genes and they present the half set of genes to the other members of the *Genus Chlamydia.*[2]. They have strain differences among the *Genus Chlamydia* also differs in the nucleotide sequences also in the various metabolic genes that includes purine and tryptophan and the potential virulence factors such as the inclusions membranes proteins and polymorphic membranes proteins [6]. Human strains of the *C. pneumoniae* are highly conserved in both their genes and nucleotide contents. They have around 300 single nucleotide polymorphisms (SNPs) have identified which separate the

four human *C. pneumoniae* respiratory strains with the whole genome sequences available yet to the date. *C. pneumoniae* were recently sequenced and it have been found that the strains of human and animals are much more virtually identical [4]. *C. pneumoniae* is an obligate intracellular pathogen that infects human and causes pneumonia. It was also known as the Taiwan acute respiratory agent (TWAR) because it is causes from the strain (TW-183) and acute respiratory isolate from another strain (AR-39) and it is commonly known as the *C. pneumoniae*. *C. pneumoniae* has very complex life cycle it can infect the other cell to reproduce. It can also infect the Koalas, emerald tree boas chameleons, frogs and turtles [3].

The first case of this infection is reported in Taiwan in 1950. There are no evidence and cases of this infection in human history before 1950. They also cause the other diseases atypical pneumoniae, bronchitis etc [1]. *C. pneumoniae* is found as a causative agent for (CAP) community acquired pneumonia and it accounts for the 6% to 10% cases. In studies it has been found that *C. pneumoniae* is co-pathogen of the (CAP) community acquired pneumonia. A total of 176 cases of C. pneumonia were discovered in a compilation of eight studies of CAP analyzed by [Kauppinen and Saikku] [2], with 68 percent representing pneumonia with more than one apparent origin. *Streptococcus pneumoniae*is the most prevalent organism associated with this [5].

In such lower and upper respiratory tract infection caused by *C. pneumoniae* may pave the way for invasion by the other bacteria such as *Streptococcus pneumoniae*. The clinical studies have been found that the characteristics of these infections reflects manifestation of the infections with the associated pathogen rather than the manifestation of the *C. pneumoniae* infection [4,5]. Although most of *C. pneumoniae* infections are mild, severe lung infection and extensive lymphocytic alveolitis have been reported [2]. In addition, adult-onset asthma [3], chronic obstructive pulmonary disease (COLD), and the development of vascular lesions may all be linked to chronic-persistent or recurring infections. Until now it had been found that about 50 strains have been isolated worldwide. DNA homology between these strains is 94 to 100 %. *C. pneumoniae* have been placed in their own order. Chlamydia are bacteria, according to rRNA sequence analysis, however

they have a very distant association with other bacterial divisions [5]. Although it has been reported that *C. pneumoniae* belongs to separate genera, their gene content and genome organization, as well as their structure, are very similar [6]. Though due to high percentage of *C. pneumoniae* infection in adults there is many chances of re-infections may occur. According to studies they suggest that *C. pneumoniae* infection seems to be both epidemic and endemic with ability of frequent re-infections during lifetime [7,8]. According to the currently known data, *C. pneumoniae* is a primary human-to-human disease that does not have an animal host. There is currently no vaccination available for the proper treatment of these disorders. There is a course of a antibiotics to cure this onset diseases through medication but the chances of re-infections is occurred and no proper treatment is available [8].

So as studies have shown that there is no proper drug/vaccine which can targets the infections and cure the onset diseases so the current study is designed for the identification of the therapeutics targets. In this study the Insilco approach is applied for Drug target Identification of *C. pneumoniae* in host (Human) to overcome the onset of the diseases. In this Insilco approach different computational tools are applied to retrieve the data for Drug target identification of the *C. pneumoniae* [9].

Chlamydia pneumonia is the pathogenic bacteria and the bio sample of the *C. pneumonia* were found worldwide i.e., from USA, Germany, Veinnie, Taiwan. The complete genome of the C. pneumonia was retrieved from the NCBI. Out of 56 strains they have twelve strains which having complete genome which are written below in (Table 1). The proteomes from the twelve genome of the *C. pneumoniae* were compared using the Pan-Genomic approach. The set of entire genes of all strains is known as the Pan-Genomic. Then its core genome is retrieved from the EDGAR web tool. The core genome represents the basic genes necessary for information transmittal and production of gene products. After the core genome prediction, it can be further filter out on the basis of the essentiality of the bacteria. Then the subtractive genomic approach is used to find all the essential proteins and they were checked against the non-homologous to the host (human). And

after this the non-homologous proteins were subjected for the virtual screening using the compound/ligand library which are retrieved from the literature review. This novel approach is applied in this study to targets the potent therapeutics to overcome the onset of the diseases.

1.1 Aims and Objectives

C. pneumoniae is one of the frequent causes responsible for infection in the upper respiratory tract. The treatment regimen available to cure this infection involves antibiotic treatment which are time borne and reinfection may occur. Resistance to antibiotic is another causes require more effective and long lasting treatment strategy. In this regards this project is designed with an aim to explore genomes of the *C. pneumoniae* to find new proteins/genes which could be together to develop drug

The main objectives of this study include:

1. Identification of core genes in selected strains of *C. pneumoniae*.
2. Prioritization of non-homologous essential genes as drug targets
3. Inhibition of drug targets with plant drive therapeutically active constituents.

Chapter 2

Literature Review

This chapter describes the background of the *Genus Chlamydia* , its etiology , prevalence and the genomic analysis its essential genes , subtractive analysis , pan genomic analysis, annotated genes, genes variants against available drugs/vaccine.

2.1 *Chlamydia pneumoniae*

Genus Chlamydia are obligate eu-bacterial parasites and they are further classified into four species. *C. pneumoniae* is one of them it is a intracellular gram negative pathogenic bacteria. All classes of *Genus Chlamydia* share common characteristics. The organisms grow in the specialized vacuole in the golgi apparatus vesicular compartment of the eukaryotic cell [1].

It is a parasitic bacterium and it have highly streamlined genome. *C. pneumoniae* is capable of persisting in the immune cells of the host. Infection causes host cells to release a number of pro-inflammatory cytokines, which are thought to have a role in disease pathogenesis. *C. pneumoniae* is a species of bacteria that causes pneumonia and some other respiratory tract disorders. CAP (community-acquired pneumonia) and lung infection are also caused by *C. pneumoniae* [8]. *Chlamydia* has been reported that it is the most commonly sexually transmitted infection (STI) in the United state and Europe. The infection is caused by the bacterium

C. pneumoniae and its infection is named after its pathogen. *C. pneumoniae* was first associated with two human pneumonia epidemics in 1985, and has since been linked to a variety of chronic conditions including asthma, cardiovascular disease, arthritis, and Alzheimer's disease. The first report of *C. pneumoniae* infection in a nonhuman host occurred in 1993 [9], when it was isolated from horses, and it was later identified as one of two Chlamydia species infecting koalas in 1994 [10]. *C. pneumoniae* has since been involved in a wide range of cold and warm-blooded animals all across the world, making it the most ubiquitous of all Chlamydia species [9].

The *Genus Chlamydia*les are characterized by their highly conserved and compact genome and they have wide range of the host cell in which they caused the diseases. All member of *Chlamydiales* phylum shares highly conserved core genome of over 500 genes [37]. The difference among the species and strain specific species is of the nucleotides differences in the metabolic genes which includes the tryptophan , purine and potential virulence factors such as their polymorphic membrane proteins and inclusion membrane proteins and functionally uncharacterized hypothetical proteins which are located in the plasticity zone [40]. Studies have been shown that the human strain of *C. pneumoniae* has highly conserved in their both regions which includes the genes and nucleotide contents. They have around 300 SNPs which can differentiate the four human *C. pneumoniae* respiratory strains from the available whole genome sequences [28].

The genomes of the koala and the bandicoot *C. pneumoniae* were recently sequenced which reveals that *C. pneumoniae* strains between the human and animals are variable but the interesting fact about it is that the human strains and the animals strains are virtually identical to each other [38]. The major genetic differences has been observed between the respiratory *C. pneumoniae* strains are limited to the encoding genes to their polymorphic membrane proteins and the outer membrane proteins. The sequence encoding genes in the *C. pneumoniae* makes up to 22% coding region as compared to the human *C. pneumoniae*. In *C. pneumoniae* strains differences have been noted in their nucleotide and the gene expression [41]. The *C. pneumoniae* were compared with the other members of

the Chlamydia family it have been noted that over one hundred characterized putative incs identified in the human *C. pneumoniae* strains. Although the genetic changes to potential phenotypes is limited due to lack of the genome sequences for the *C. pneumoniae* strains isolated from the other human tissues.

2.1.1 Pathogenesis of *C. pneumoniae*

C. pneumoniae is a Gram-negative bacterium with a size range of (0.2 to 1 m) that goes through a number of modifications during its life cycle. Inside the host cell, it acts as an elementary body (EB). The elementary body is found in inactive form but it is resistant to the environmental stresses and they can also survive the outside of the host cell but for a limited time duration. The elementary body moves from an infected lungs of persons and affects the uninfected/healthy person [13]. A very small droplet of the infection is sufficient to affect the healthy person. The elementary body is taken up in the lungs by the cells in a pouch like structure which is known as endosome by the process which is known as phagocytosis.

The elementary body cannot be destroyed by the fusion with the lysosomes. Instead of this they can transform itself into the reticulate body (RB) and start to replicate inside the endosome. To complete its reproduction inside the endosome, the reticulate body has used the host cellular metabolic system [14]. When the replication is finally completed, the reticulate bodies convert back to the elementary body and release themselves back into the lung, continuing the process that eventually leads to the death of the host cell. The elementary body are thereafter are able to infect either in the same organisms or new host cell and repeat its cycle. The life cycle of the *C. pneumoniae* depends on the elementary body and the reticulate body. As the elementary body has the ability to infect the new host cell or same organisms cell but it cannot replicate by itself it depends on the reticulate body which can undergoes the process of the replication and causes the replication. They both elementary body and the reticulate body depends on each other for the cell cycle of the *C. pneumoniae*. As shown in the figure 2.1 the life cycle of the *C. pneumoniae* that how the elementary body can undergoes the

phagosome stage and then differentiate the elementary body into the reticulate body and the process of replication is done and the cell lysis occurred and cells are released into the host cell cause the ultimately the death of the cell and process is repeated in the same as well as new host cells [14].

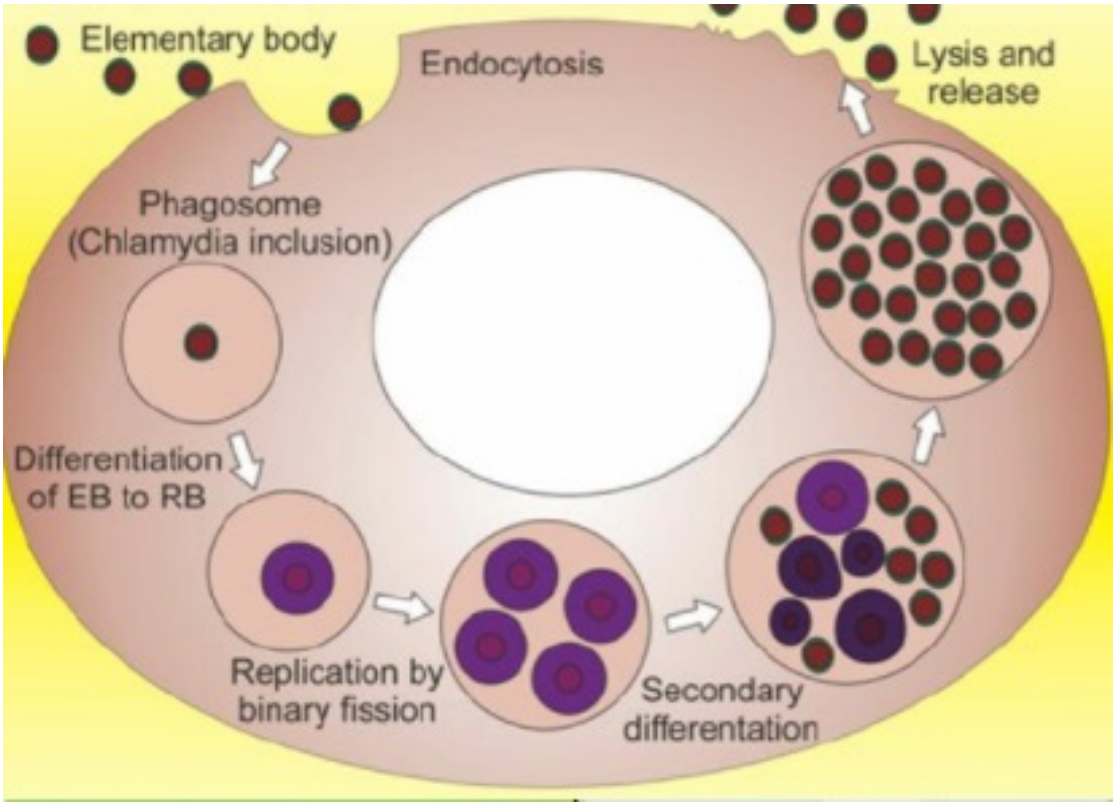

FIGURE 2.1: Life Cycle of *C. pneumoniae* [14].

2.1.2 Etiology & Mode of Transmission

C. pneumoniae infection is commonly spread in the peoples by the coughing and sneezing which creates small respiratory droplets that can contains bacteria and infects the lungs of the healthy persons. This can also be transmitted when the people can touch something where the droplets of the bacteria is present and then they touch their mouth or nose [14]. Most of the people do not infect with *C. pneumoniae* when they spend some of time with them but the bacteria spread to those peoples who will live together because it is a sexually transmitted diseases. *C. pneumoniae* have long incubation period that is the time in which the bacteria is breathing in the host cell and developed the symptoms of the *C. pneumoniae*.

The symptoms of the *C. pneumoniae* usually begin between 3 to 4 weeks after exposure [15]. In this figure 2.2 it is shown that the *C. pneumoniae* directly infects the lungs of the person. They affects the normal function of the bronchiole and alveoli and causes the accumulation of the fluid in the alveoli and caused the *C. pneumoniae* [13].

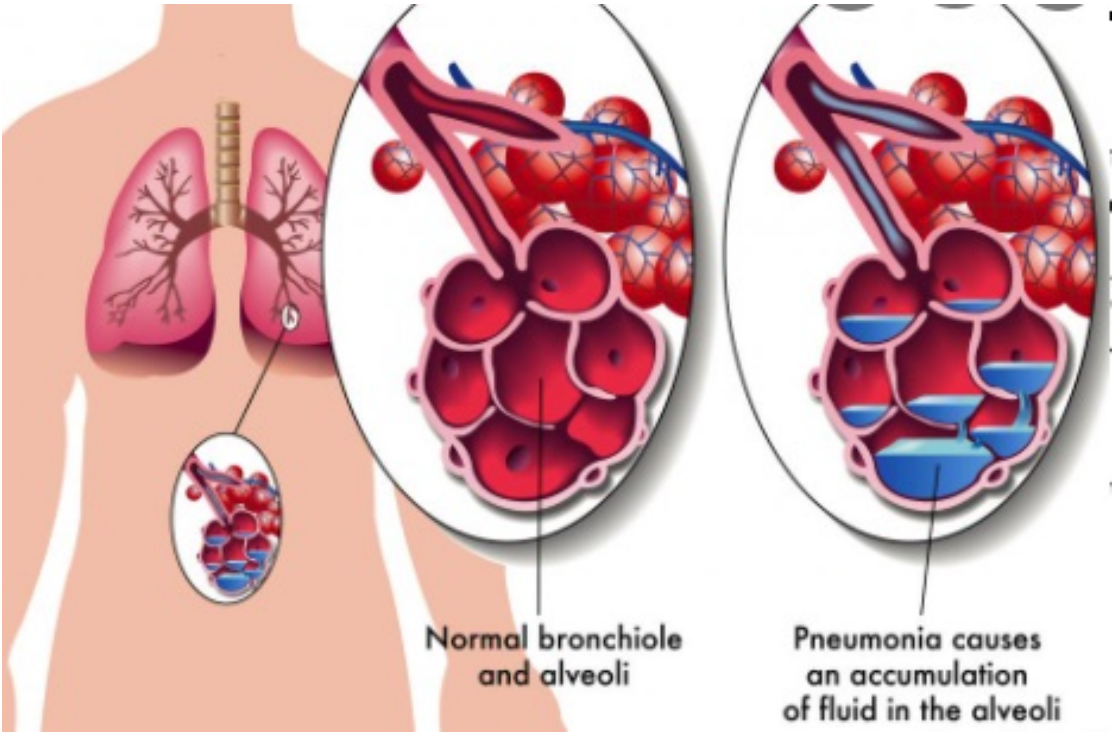

FIGURE 2.2: Lungs of Infected and Uninfected Person [13].

Many cases of the *C. pneumoniae* are diagnosed by Pediatrician when they physical examine the children. *C. pneumoniae* can be diagnosed by the blood test in which they detect the antibodies to the bacteria. It can take a week to shows results of the antibodies in the blood. There are special labs that can evaluate the specimens which are taken from the nose throat. But there is no proper cure of this diseases because *C. pneumoniae* shows the in vitro resistance to the antibiotics such as Penicillin , ampicillin, sulfa drugs. So, they can be temporary treated by the medications but infection can re-appear after conventional course of the antibiotics. So, there is need to novel therapeutics approach which can prevent the onset of the diseases [17]. In figure 2.3 it shows the blockage of the alveoli when is infected with the diseases. The alveoli are blocked and cannot be able to functions properly which results in infection.

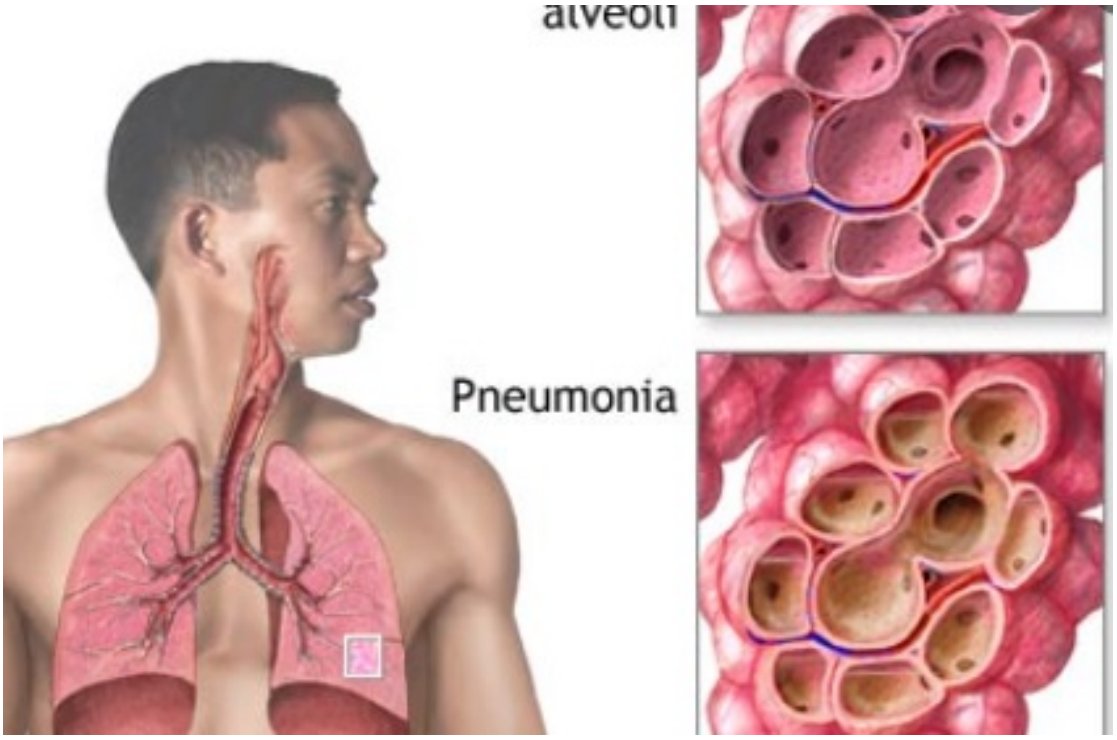

FIGURE 2.3: Pneumonia Affected the Alveoli [15].

C. pneumoniae is an obligate intracellular bacterial pathogen with potential developmental features in diseases, according to studies. Its distinctive lifestyle and capacity to propagate across the host cell while maintaining an immunological response. There are some common signs and symptoms of *C. pneumoniae* which causes a prolonged cough, bronchitis , pneumonia , sore throat , ear infection, breathing problems, runny or stuffy nose, fatigue, fever, sore throat and headache. The cough in this case prolongs for 2 to 6 weeks [15]. In figure 2.4 it is shown the signs and the symptoms of the *C. pneumoniae* which is commonly occurred and diagnosed at the early onset of the diseases.

2.1.3 Prevalence

The studies have been reported that majority of the individuals are exposed to the *C. pneumoniae* throughout their lifetime. The prevalence of the *C. pneumoniae* shows that the 50% people with the age group of 20 years (adults) are infected by this disease and 80% by the age group of 60 – 70 years old. The *C. pneumoniae* infects the lungs as it is a upper tract respiratory diseases so it can affects the lungs

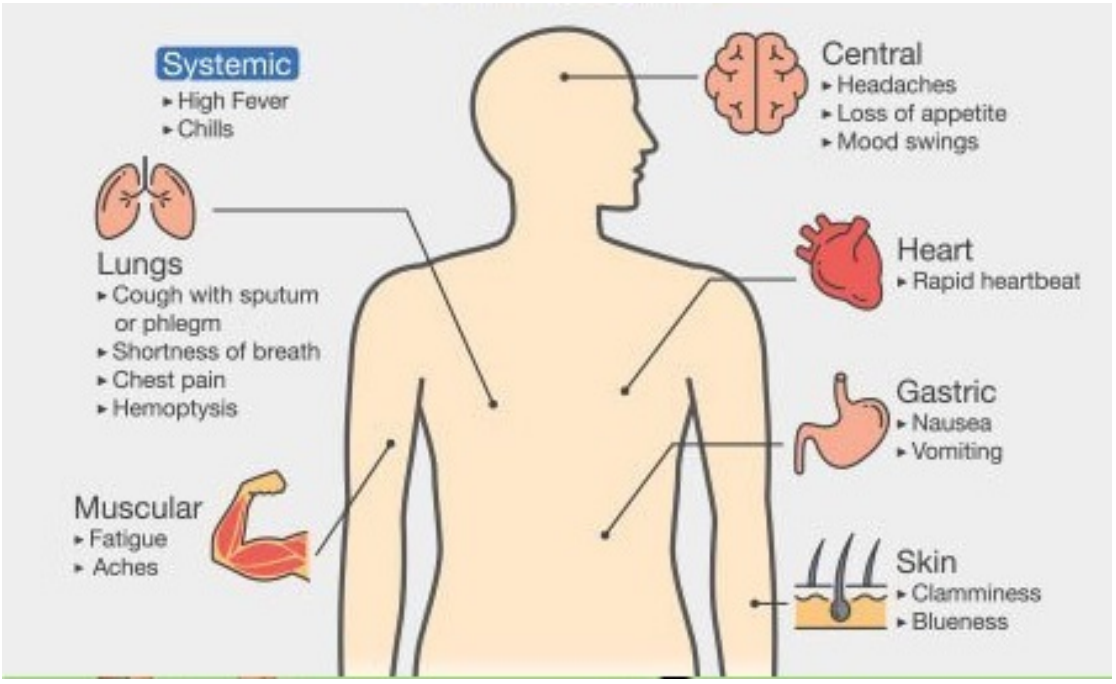

FIGURE 2.4: Sign and Symptoms of *C. pneumoniae* [16].

of the old peoples [18]. According to research around world it have been reported that the worldwide prevalence of this diseases is 50% it have been reported that the *C. pneumoniae* is most commonly occurred in male than the female [67]. According to the survey the males are exposed to this diseases. The cases of *C. pneumoniae* is reported worldwide that is from Taiwan , Veinnie , Germany , Australia, Hungry, South Africa ,USA, Netherlands etc. The first case of the *C. pneumoniae* is reported from the Taiwan in 1950 [4] after this it named as TWAR stands for Taiwan Acute Respiratory dieases. In this figure 2.5 is shows the prevalence of the *C. pneumoniae* in 2009. The blue colors is for the old and orange colors indicates the adults it shows that it can infect the adults.

2.1.4 Epidemiology

C. pneumoniae is the Gram negative intracellular pathogenic bacteria. It plays role in the respiratory tract infection and extra pulmonary diseases. Seroepidemiological survey have been found that *C. pneumoniae* infections seems to be both epidemic and endemic. This studies have been shown that *C. pneumoniae* is widespread and it can caused reinfection during lifetime again and again. The

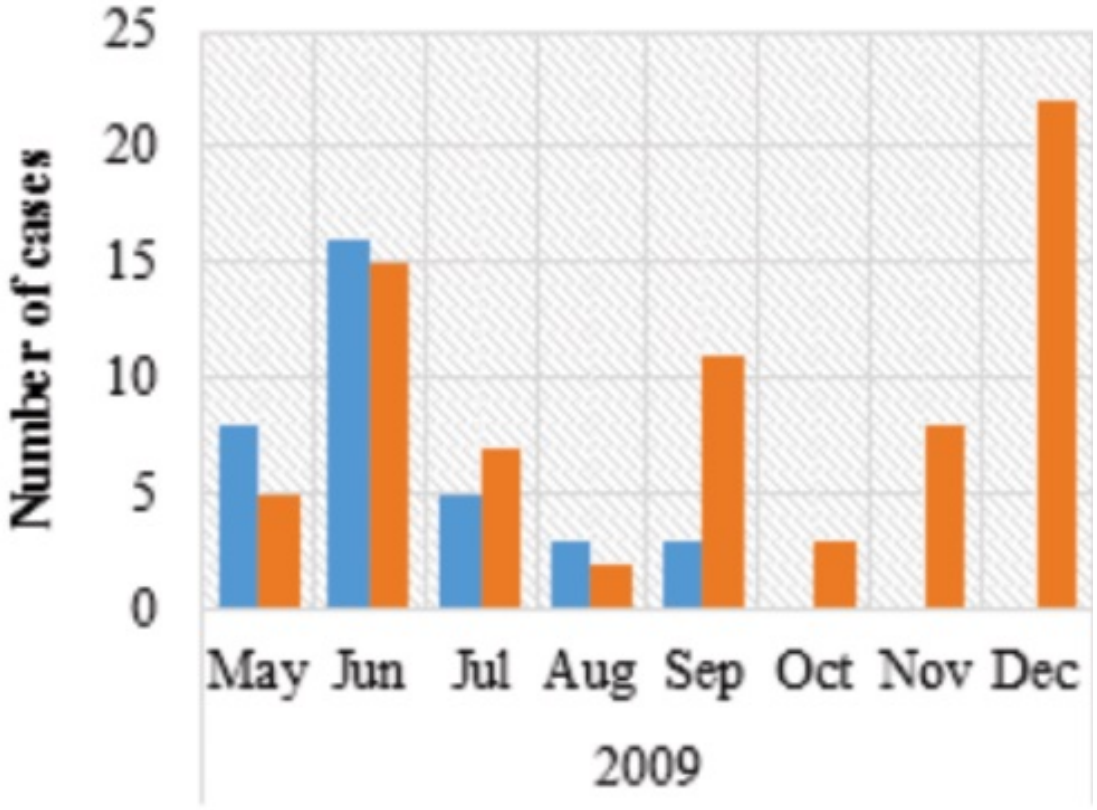

FIGURE 2.5: Prevalence of the *C. pneumoniae* [19].

maximum rate of new infections, according to surveys undertaken in Western countries, occurs between the ages of 5 and 15 years old [18]. All across globe, adult males have a higher prevalence of the antibody than females. *C. pneumoniae* is primarily transferred from human to human without any animal reservoir, according to currently available data. It have been found in literature that the household outbreaks is highly reported which helps in transmission of diseases. *C. pneumoniae* is the most widely reported it is associated with CAP (community acquired pneumonia), which affects between 6% and 25% of people and causes a moderate infection with a self-limiting clinical manifestations in certain cases [19].

Recently it have been reported that *C. pneumoniae* plays role in the community acquired pneumoniae and also in respiratory infections in immune compromised patients *C. pneumoniae* infections have been involved in the development of the pathogenesis of asthma in both adults and children. Furthermore the *C. pneumoniae* have also detected in the coronary plaques by different mode of transmission by the ways. It have been found that available data suggested that is transmitted

human to human without any animal reservoir. Outbreak in closed communities such as families, schools, colleges and military garrisons have been observed repeatedly [20]. *C. pneumoniae* can intact on the environmental surfaces for the long time period and it can transfers through hands and on skin. They have good survival rate in aerosolized particles of <5 μm and it can favor high humidity and low temperature for the survival [21].

C. pneumoniae have spread slowly in time durations of the 30 days. Serologic data have been show that durations of the epidemics vary from the 5 to 8 months it have relatively long incubation period. In the developed countries the infections is not common before the age of the 5 years. Seroepidemiological data shows higher prevalence in developing countries. It is more prevalent in adults, with a prevalence of 70%. Child-to-child transmission has been observed in Japan, as well as a family outbreak [22].

2.2 Gap Analysis

Gaps analysis are used in bioinformatics to account for genetic alterations caused by insertions or deletions in the sequence, also known as indels. As a result, when aligning two sequences of DNA, the gaps must be assessed as a whole [21].

The gaps are the chunks that arise while trying to align DNA or protein sequences and having to use padding or null characters to match homologous residues. The gap analysis of the diseases is that it have no available drug/vaccine which can cure it. It can causes the reinfection and diseases occurred.

2.2.1 Available Drug/Vaccine

C. pneumoniae diseases is caused by Gram negative pathogenic bacteria. There is no drug or vaccine is available yet which can properly cure the diseases. Studies have been shown that clinicians can treat this diseases by first line therapy in which Tetracycline's and Flouroquionolones should be given. But it could not

be prescribed to the young children's because they can affect the immune system directly. If a person is suffering from this diseases they should be diagnosed by the laboratory test which involves the taking a sample of sputum or swab from the nose/throat and test it in the laboratory for the further diagnosis. The proper treatment of *C. pneumoniae* is not discovered yet but the doctors can treat with different antibiotics.

There are many complications which clinicians facing while treating the *C. pneumoniae* the complications includes Encephalitis (swelling in brain) Myocarditis (swelling of the heart) they also can cause the chronic infections they can directly affects the respitorty tracts along with the brain and heart in severe conditions and causes asthma, arthritis [39]. So in this study the novel approach is applied to identify the potent therapeutics and drug / vaccine targets to overcome the onset of the diseases.

2.2.2 Reoccurrence of the Diseases

This diseases is treated by antibiotics but it has no proper drug/vaccine which can cure it properly. so the re-infection occurred which caused the diseases again.

2.2.3 Drug Resistance

The proper treatment of *C. pneumoniae* is not discovered yet but the doctors can treat with different antibiotics. This diseases is treated by first line therapy in which Tetracycline's and Flouroquionolones should be given. But re-infection may occurred which results in the drug resistance to antibiotics.

2.3 Drug Targets

A drug target is a molecule in the body, typically a protein, that is intrinsically linked to a specific disease process and could be targeted by a medicine to have

a desired therapeutic effect. The four main targets for pharmacological activity are receptors, ion channels, enzymes, and carrier molecules. Because they bind to specific target proteins, most drugs are effective in each of these four conditions [31]. G-protein-coupled receptors (GPCRs) account for 44% of human pharmacological targets, enzymes for 29%, and transporter proteins for 15%. DNA is used as many drugs used in cancer therapy target particular molecular target, which also serves as a non-specific target for cytotoxic agents. A covalent interaction takes place when the drug and its target share a pair of electrons, resulting in the formation of a new molecule. The interaction is extremely strong, leading in irreversible drug-target binding. This usually has a long-term biological consequence that can't be changed [32].

2.3.1 Pan Genome Analysis

A Pan genome is a set of orthologous and unique genes found in a group of organisms. The core genome, accessory genome, and species- or strain-specific genes make up the pan-genome. The homogeneity of genome annotation is the first step in a pan-genome analysis. Over the decades the advances in bioinformatics have facilitate the implementation of the Pan-genome analysis that gives the information about the associations of specific group of organisms.

Pan genome analysis have been shown effective approach to better understand the clade of the pathogenic bacteria and also helps to therapeutics targets on the basis of their biological similarities and differences [23]. The pan genome analysis of the pathogenic bacteria focus on the computational techniques that allow efficient pan genome analysis. A pan genome is the set of entire genes from all the strains within the clade .The pan genome is then broken down into three parts the one is the core pan genome that contains the genes which are present in all individual and the shell pan genome that conations genes present in two or more strains known as cloud pan genome the third one is the accessory genome which contains the dispensable genes which is present in the subset and they are strain specific genes. The overall studies of the pan genome is known as pan genomics [24].

Genetic repositories of the bacteria is large than the gene content of the human strains. There are further two types of pan genome in the closed genome few genes are added after sequencing many strains. In the open pan genomes many genes are added and predicted the size of the full pan genome is not possible. Pan genomes were constructed for the bacterial species identification but recently eukaryotic pan genome have been developed for the plant species. The pan genome has three parts core , shell and dispensable. The core pan genome is the part of the genome which is shared by all genomes. The shell pan genome is shared by the majority of the genomes it is shared by the 50% of the genome. The third part is the dispensable which is shared by the minimal genome families.

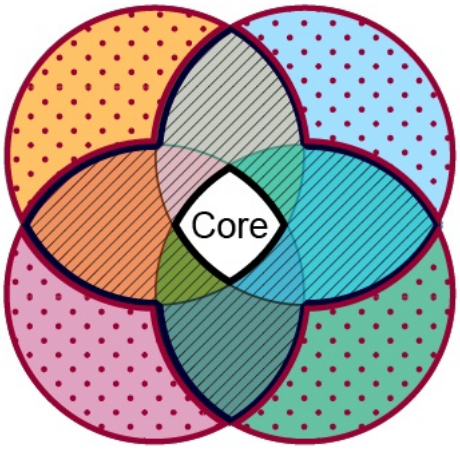

FIGURE 2.6: Pan Genome Analysis [23].

TABLE 2.1: List of Publications using Pan Genomic Based Drug Target Identification

Pathogenic Bacteria	**Application**	**Ref.**
Clostridium botulinum	To understand the symptoms of pathogen pan genome analysis was done. They found high genomic similarity from the genomes	[71]
Streptococcus pneumoniae	A new method for identifying possible treatment targets in human bacterial infections has been developed.	[72]

Haemophilus influenzae	H. influenzae strains with multidrug resistance in clinical isolates for the discovery of better/new medications to overcome this disease.	[61]
Corynebacterium pseudotuberculosis	In Corynebacterium pseudotuberculosis, comparative modeling at the proteome scale was used to identify conserved therapeutic and vaccination targets.	[73]

2.3.2 Subtractive Genome Approach

Subtractive genomics is the process in which the subtraction of sequences between host and the pathogen. Proteome which helps in providing information for a set of proteins which are essential to pathogen but are not present in the host.

Subtractive genomic analysis is an effective tool for determining protein identities in the pathogenic bacteria that are toxic but not present in the host. The identified proteins serve as the potential therapeutics targets against the pathogens which are responsible for infecting the host tissues. They insure the essentiality of the pathogen specific protein. Proteins that are necessary for a cell survival are known as essential proteins. The essentiality of gene is depends on the specific environment and cellular circumstances. The essentiality of the protein targets is a positive key factor for the druggability of the potential target therapeutics [23]. Subtractive genomic approach is used to analyze the whole proteome of the host and its pathogen and also identify the protein with the potent therapeutics which are present in the genome of the pathogen. They can subtract pathogenic genes which are required for the survival of the pathogen but it is not present in the

host. They play important role in the therapeutics development and identify the druggable proteins for the potent targets [42].

2.3.3 Genes Variants

Genetic variation is the difference among the individual DNA. There are two types of genetic variation which includes mutation and genetic recombination. Mutation is the main source of the genetic variation other factors also contribute such as sexual reproduction and genetic drift. Genetic variation is identified as phenotypic variation either in the quantitative traits or discrete traits for example white , pink , red petal color in flowers [43]. The technique of protein electrophoresis can be used to identify genetic variation at the enzymatic level. At each locus of polymorphic genes, there are many alleles. They are most common in the plants and insects and less in the vertebrates [32].

Genetic variation is caused by the variation in their nucleotides genes. When changes in the order of nucleotides in a DNA sequence result in differences in the order of amino acids in proteins coded by that DNA sequence, phenotypic variation occurs, and these variants in amino acid sequence alter the shape, and consequently the function, of the enzyme. [41]. According to the magnitude and kind of genomic variation that underpins genetic change, genetic variation can be classified into several categories. Base-pair substitution and indels are examples of small-scale sequence change (1 kilo base, kb). Large-scale structural variation ($>$ 1 kb) can be caused by copy number variation (loss or gain) or chromosomal rearrangement (translocation, inversion, or Segmental acquired). Genetic variation and recombination are produced by transposable elements and endogenous retroviruses, which are sometimes boosted by a variety of persistent viruses and their defectives, which provide genetic innovation in host genomes. The numerical variability in whole chromosomes or genomes is termed to as polyploidy or aneuploidy [41]. A gene variation is a change in the DNA sequence that makes up a gene that is irreversible. Initially, this form of genetic change was known to as a gene mutation, but because changes in DNA do not necessarily result in disease,

the term gene variation is regarded more accurate. A gene's variants can alter one or more DNA building units (nucleotides) [42].

Mosaicism is a condition in which a group of cells in the body has a distinct genetic composition than others due to variants acquired during development. The genetic alteration in Mosaicism does not occur in a parent's egg or sperm cells, nor in the fertilized egg, but rather later in embryonic development or maturity. Cells that emerge from the cell writhe and divide as the cell grows and divides [42]. Somatic Mosaicism occurs when a subset of somatic cells has a gene variation while others do not. Somatic Mosaicism may or may not produce health problems, depending on the variant and the number of cells affected. Germ line Mosaicism occurs when a section of egg or sperm cells has a variation while the rest do not. An unaffected parent can convey a genetic problem to their child in this situation [42]. The majority of variations do not cause disease, and those that do are rare in the general population. Some genetic variations are prevalent enough in the population to be classified as such. Differential traits such as eye color, hair color, and blood type are caused by a variety of such variants. Although many of these frequent DNA variants have no deleterious consequences [42].

2.3.4 Genome Annotation

Annotating a genome entails explaining the function of an anticipated gene's product through an Insilco approach. This can be accomplished with bioinformatics software that includes signal sensors (e.g., for TATA box, start and stop codons, or poly-A signal detection), content sensors (e.g., for G+C content, codons usage, or decoding frequency detection), and simian bioinformatics software. similarity detection (for example, for G+C content, codon use, or di codon frequency detection e.g., between proteins from closely related organisms, mRNA from the same organism, or reference genomes (Stein, 2001) [44]. However, the technique for predicting gene and genome structures (such as tRNAs, rRNA, and promoter regions) is dependent on the sequencing platforms and assembly approaches used [45]. Annotation of the genome can be divided into three categories. The first

is a nucleotide-level annotation, which attempts to identify the physical position of DNA sequences in order to locate components like genes, RNAs, and repetitive patterns. The genomic annotation can be divided into three categories. The first is a nucleotide-level annotation, which tries to specify the physical location of DNA sequences so that components such as genes, RNAs, and repeating patterns may be identified [45]. The third type of annotation is a process-level annotation, which aims to identify the networks and processes via which different genes interact, resulting in an effective functional annotation. Sequencing and/or assembly errors may risk the inference of true gene function due to the decreasing similarity at the last two levels (Miller et al., 2010; Reeves et al., 2009; Stein, 2001) [48].

2.3.5 Genomic Analysis

Genus Chlamydia bacteria are intracellular pathogens with significant medical implications. *Chlamydia trachoma* is a common sexually transmitted infection and the leading source of preventable blindness in underdeveloped nations, whereas one of the most common causes of pneumonia worldwide is *C. pneumoniae* [51]. As described in the dynamic interaction of six major evolutionary mechanisms targeted at genome reduction or complexification shapes a typical bacterial genome [42]. Genome contraction is caused by both genome streamlining and degradation, while the underlying evolutionary mechanisms are different. Strong positive selection pressure removes nodes from the genome, resulting in genome streamlining [51].

Genome degradation refers to the loss of genes due to weak or neutral selection, as seen by a large number of pseudo genes and incorporated selfish elements. Genome streamlining is found in extremely numerous and evolutionary successful organisms, but genome degradation has been shown in parasitic and symbiotic bacteria with reduced effective population densities due to environmental factors [52]. Genome decrease is countered by genome complexity and innovation, such as gene duplications operon shuffling, horizontal gene transfer, and mobile element propagation. And, while all of these pressures may be at work at the same time,

their impact on individual prokaryotic genomes varies dramatically, reflecting the biological niche and life cycle of the organism [53].

Genus Chlamydia has genomes of roughly 1 Mb and 850–1100 genes as a result of their mandatory intracellular lifestyle. There are only 14 transcription factors (TFs) predicted to control gene expression [38]. Unlike many other viruses with smaller genomes [31], however, this apparent simplification was most likely due to genome streamlining rather than deterioration. Specifically, genomes of all Chlamydia species have a low number of pseudo genes. Outside the plasticity zone, a genomic area of about 81 kB around the replication terminus [29], the gene order is highly stable. The gene content is similarly conserved, with the majority of genes shared with other members of the phylum [21]. Multiple studies have demonstrated that among *Genus Chlamydia* species, there is genome-wide homologous recombination, which may limit the accumulation of deleterious mutations. Finally, with the exception of the IS-associated tetracycline-resistance genomic island in C. suis and residues of IS-like elements and prophages in other genomes, the genomes of *Genus Chlamydia* spp. are largely free of disruptive mobile elements. Furthermore, the tiny genome of the *Chlamydia genus* allows for phenotypic variety. *C. pneumoniae* is one of the most common causes of respiratory infections in humans, and it can also infect horses, marsupials, and frogs. Other species infect mice, guinea pigs, birds, cattle, sheep, swine, horses, cats, koalas, frogs, and snakes, resulting in a wide range of diseases that can be transmitted to humans under certain conditions [54].

2.3.6 Essential Genes

Essential genes are those type of genes which are necessary for the cellular life. These genes constitute the minimum gene set which is required for the living cell. The genes which are encoded by the gene set are the essential genes and they are foundation of the life. The minimal genes set of the bacteria is the advent of the completion of the whole genome sequencing. Essential genes plays important role in different organisms [27]. Database of Essential Genes (DEG) have been

constructed that contain all the essential genes which are currently available. The essential genes in *C. pneumoniae* were retrieved from a website where essential genes were collected in vast numbers and references were provided. Each important gene entry contains a unique identification number, gene reference number, gene function, and sequencing information. The open database stores and manages all of the information. In DEG database user can search the gene by its name or its function. After that all the essential genes are blast against non homologous. In BLAST user can give the query sequences against the essential genes in DEG [30].

The essential genes are inferred using homologous sequences discovered using the DEG method. Essential genes encode tasks that are regarded to be universally important for all cells. Some basic functions and principles are thought to be shared by all cellular life on this planet. [44]. As a result, if the query sequences compared using BLAST include comparable genes in DEG, the questioned genes are likely to be essential as well. Furthermore, by doing a BLAST search against DEG for all of a genome's protein-coding genes, it is possible to identify the potential essential genes for freshly sequenced genomes' proteomes [43].

However, keep in mind that many key genes are only required in specific growth circumstances, such as rich or minimum media, when interpreting the BLAST results [31]. Another benefit of researching all of the important genes in DEG is that some rules or regulations to solve the question of what basic tasks are required to support cellular life may be identified. These concepts could pave the way for the creation of new algorithms for predicting important genes. Some key gene processes, including as DNA replication, gene transcription, protein synthesis, energy production, and cell division, are expected. DEG, which contains all essential genes from many organisms, could aid in classifying those "unexpected" essential genes. Some important gene initiatives are still underway, and additional essential genes are expected to be discovered. DEG will be updated on a regular basis to include new entries as new important genes become available [49]. For each item, they intend to include more information regarding experimental approaches. They intend to incorporate the important genes of vertebrates, such as mice, in the next version of DEG. Database welcome user feedback, corrections,

and additional information, which will be used to keep the site up to date. DEG is freely available on the web [59].

2.3.7 Non-Homologous Genes

Non homologous genes are that type of the genes which do not belong to the same pair they are from the different pair of the chromosomes. In non homologous the shape and lengths of the arms and position of the centromere is different of the chromosomes. This type of genes do not pair during the meiosis [28].

2.4 Plant Derive Natural Compounds

Plant-derived compounds are frequently used as leading structures through chemical modifications. Salicylic acid, morphine derivatives, artemisinin, dicoumarol, warfarin, and camptothecin ,topotecan and irinotecan are the natural derive components. Natural drugs are manufactured from substances present in the natural world. Plants are the most common natural drug sources. Therapeutic ethno botany is a branch of study that explores the interaction between people and medicinal plants [21]. Natural antibiotics/antibacterial contain cephalosporins, cefamycins, benzylpenicillin, and gentamicin, to name a few. Natural antibiotics and antibacterial are considerably more harmful than synthetic antibacterial.

Antibacterial specifically refers to an inhibition of growth of bacteria, whereas antibiotics refer to a general inhibition of growth of microorganisms, which could include bacteria and viruses.

2.4.1 Flavonoids

Flavonoids are quite well antibacterial agents that can kill a variety of harmful microorganisms. It is found that hydroxyls at particular locations on Flavonoids aromatic rings boost action. Methylation of the active hydroxyl groups, on the

other hand, reduces activity. By significant inhibition of signalling pathways, Flavonoids reduce the generation of reactive oxygen species (ROS) and down-regulate several inflammatory mediators, resulting in anti-inflammatory effects [22]. In Flavonoids the functional hydroxyl groups reduce oxidative stress and/or chelate metal ions to mediate their antioxidant actions [20]. Flavonoids also serve as a secondary antioxidant defense in plant tissues that have been subjected to various abiotic and biotic stressors.

2.4.2 Alkaloids

Alkaloids are a vast and structurally diverse collection of chemicals that have been used as scaffolding for antibacterial medications like metronidazole and the quinolones. Antibacterial and antiviral activity of a molecule is directly associated with substances that kill bacteria and viruses locally or decrease their rate of growth without being hazardous to surrounding tissues [31].

2.4.3 Steroids

The latest discoveries not only indicate that combining antibiotics and steroids can help people with pneumonia recover faster, but they also point to a potentially more effective treatment for someone suffering from an asthma attack caused by M pneumoniae infection [32]. Antibiotics are anti-bacterial while steroids are anti-inflammatory and anti-allergic.

The plant derive natural compounds are used as ligand which binds to the selected targets and inhibits the drug targets of the *C. pneumoniae* these are list of the plant derive natural compounds which were retrieved from the literature review used in this research. The table 5.1 (ref to An Appendix) shows the compounds names , their type their molecular weight , and function which is retrieved from the Pubchem.

Chapter 3

Materials And Methods

The 12 strains of *C. pneumoniae* were selected for the core genomic identification using the pan genomic approach. The different steps were involved in this Insilco approach for the genomic based drug target identification. The C. Pneumonia were found worldwide i.e. from USA, Germany, Veinnie, Taiwan. The genome of the *C. pneumoniae* was retrieved from the NCBI. Then the core genome identification of the *C. pneumoniae* was done through the EDGAR web tool.

The genomic based drug target identification is done through the DEG database. In Drug target Priotization there are several factors which can help in determine potential therapeutic targets such as molecular weight, molecular function, cellular localization, pathway analysis and virulence factors. The molecular docking is done through the MOE-DOCKING tool. These parameters are used in this methodology to overcome this onset decease.

3.1 Genome Selection

Genome selection is a type of marker-assisted selection in which all quantitative trait loci are selected using genetic markers that includes the entire genome [29]. The genomes selection of the *C. pneumoniae* which is Gram negative intracellular pathogenic bacteria is done through NCBI. The genes and proteins sequences of

these twelve strains were retrieved from the NCBI (https://www.ncbi.nlm.nih.gov /genome/browse/prokaryotes/171/). The inclusion and exclusion criteria of the genome selection is the whole genome and exclusion criteria is annotated and variant genome.

3.2 Identification of Core Genomes

The core genome identification is done because the core genome represents the genes which are present in all strains of the species. It contains the housekeeping genes which is responsible for the regulatory function [38]. The pan-genome is the total number of genes that are present in given dataset. It includes core genome, accessory genome and strain specific genome.

Core genome includes all of the genes that are shared by all genomes, accessory genome consist of those gene which are absent in some of the strain sand strain-specific genes consist of those genes which are only present in single genome The core genome of the *C. pneumoniae* is retrieved from Edgar web tool involved some steps. To identify the core genome of the *C. pneumoniae* (https://edgar. computational. bio. unigiessen. de/ cgibin/ edgar. cgi?action = view type = core Genomeproject =EDGARChlamydia)

3.3 Identification of Non-Homologous Proteins

Non-homologues proteins are those types of proteins which are not present in the human. Subtractive genomics is the mechanisms which helps to provide the data for the proteins collection which is important for the pathogen but it is not present in the host [39]. The identification of the non-homologous and essential protein was carried out by using the NCBI BLASTn , by default parameters was used i.e (e-value = 0.0001, bit score $\geq$ 100). To find the essential genes from DEG database tool apply two parameters while using the tool that the cut off values are (Pct identity should be $>$ 35 and E value is = 0.001)

3.4 Genomic Based Target Identification

The first method of identifying and analyzing a potential therapeutic target gene or protein is to determine its function and relevance in the disease. Requires the identification of the target, the molecular mechanisms addressed by the target are characterized [39].

3.5 Drug Target Priortization

The drug target Priotization is to identify the suitable targets such as genes/proteins for the characterization for the annotating the gene function and the drug discovery. There are the several factors which can help out in determining the potential therapeutic targets such as the molecular weight, the pathway analysis of the genes , Virulence factors and the molecular and biological functions ,Cellular Localization [43].

3.5.1 Pathway Analysis (KEGG)

Pathway analysis is a collection of frequently used techniques for life science research that intends to make interpretation of high-throughput biological data [42]. The Pathway analysis of the essential proteins was find out from the Kyoto Encyclopedia of Genes and Genomes (KEGG) (https://www.genome.jp/kegg/).

3.5.2 Molecular Weight (ProtParam)

The Molecular weight of the protein is found out from the ProtParam. Molecular weight is the weight of the protein which is calculated in Kilo Dalton (kDa). In terms of drug target, the molecular weight of the target protein should be less than 100 kDa because it could easily pass through the cell membrane and reached its target site [44]. The molecular weight was calculated from ProtParam tool.

ProtParam is an online tool that is available on Expasy server. This tool computes physio chemical properties of the molecule.

3.5.3 Virulence (PAIDB)

Virulence is the relative ability of a microorganism to overcome the host defense. The Virulence of the essentials proteins were analyzed from the Pathogenicity Island Database (PAIDB) (http://www.paidb.re.kr/ aboutpaidb.php).

3.5.4 Molecular Function and Biological Process (Uniprot)

A biological process implies a specific goal for which the organism has been genetically programmed. Cell division, for example, produces two daughter cells (a split cell) from a single parent cell.[48] Molecular process that can be carried out by a single macromolecular machine, usually through direct physical contacts with other molecules. In this application, function refers to actions or activity that a gene product (or complex) performs. The important genes' molecular function and biological process were identified using the (Uniprot).[49]

3.5.5 Cellular Localization (CELLO)

The Cellular Localization of the essentials proteins were found out from the Sub cellular localization predictive system (CELLO). To analyze the distribution of potential therapeutic targets into different compartments of the cell, the PSORTb server was applied to predict their sub cellular localization. either they are cytoplasmic , intracellular or extracellular membrane proteins. [45] This cellular localization helps in the prediction that the selected potential targets are used as a drug or vaccine. If the selected potential targets are cytoplasmic it mean they are used to predict the drug targets but if they are extracellular then they were used as vaccine potential targets. The results were also cross-checked using the CELLO web server.

3.6 Catalytic Pocket Detection

The DogsiteScorer provides catalytic Pocket Detection of essential proteins with a specific drug score. DogsiteScorer is a highly autonomous pocket and druggability prediction method. To predict druggability scores, a support vector machine (SVM) is trained and evaluated on the druggability data set (DD) and its non redundant counterpart (NRDD). [32] The automated pocket detection and analysis tools enables the identification of the potential pocket and sub-pockets in the protein structure. The catalytic pocket detection have the global properties that describe its size , shape and chemical properties of the predicted pockets of the proteins.[54] A druggability score provides the linear combinations of the sub-pockets which describe its hydrophobicity, volume and enclosure.

3.7 Molecular Docking

Molecular Docking is done through MOE tool. Molecular docking is a technique of bioinformatics modeling that involves the combination of two or more molecules to form a stable adduct. It can predicts the three-dimensional structure of any complex based on the binding characteristics of the ligand and target. The 9 selected targets were subjected to the docking against the compound/ligand that are retrieved from the literature review. The plant derive natural compounds are used as ligand which binds to the selected targets and inhibits the drug targets of the *C. pneumoniae*. Out of 100 selected molecules these are re-docked again and top 10 molecules were selected. For each protein 1 of the best interaction were selected from top of these 10 molecules. Then after this for the results visualization 1 of the best interaction of protein-ligand is drawn in the chimera tool.

3.7.1 Docking Validation

Results visualization is done through Chimera tool. UCSF chimera tools prepare the structure of the receptor for the docking. It involves the deleting ligand and

solvent molecules and eliminate the alternate locations of the residues they change the selenomethiones to methionines and add the hydrogen atom and assign the charges to the protein atoms. In this design study the UCSF chimera tool is used in which the selected drug targets were visualized and delete the ligand and the solvent molecules and add the hydrogen atom and assign the charges to the proteins as they were shown in chapter 4 in the results figures that how they shows interaction with the active ligand.

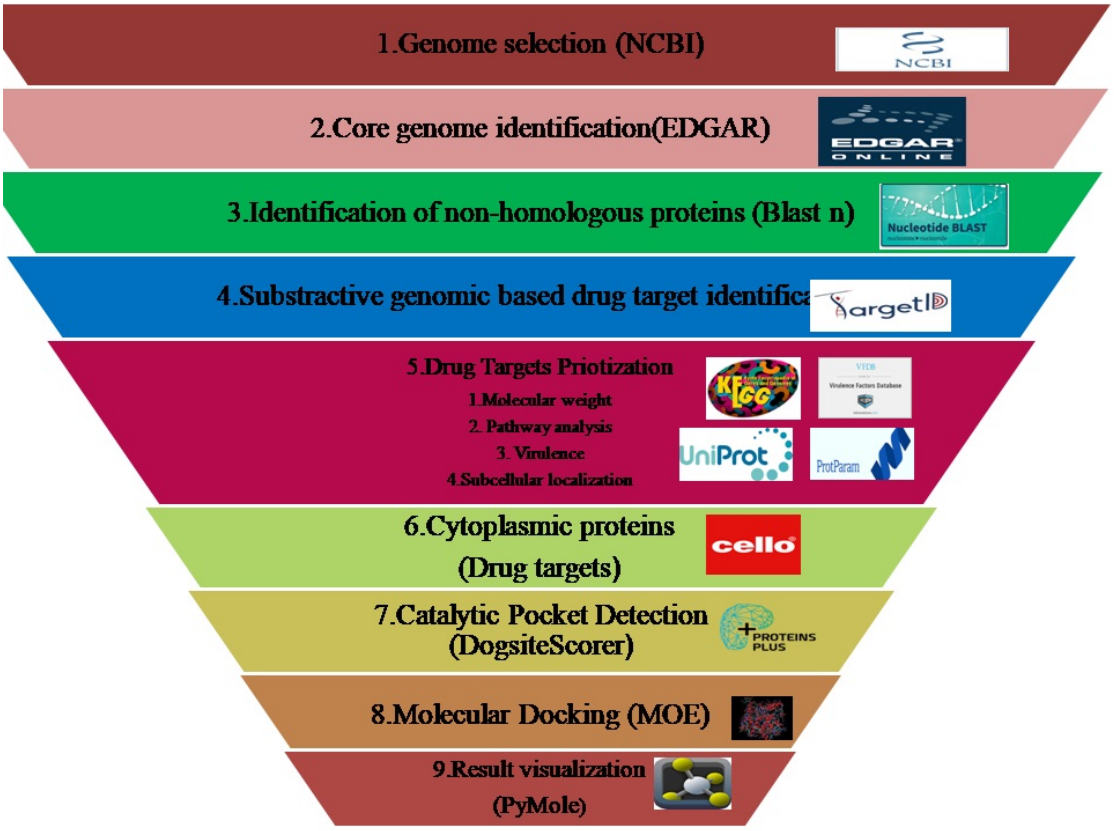

FIGURE 3.1: Flowchart of the Research Methodology

Chapter 4

Results and Discussion

New therapeutic targets were required to prevent the onset of the *C. pneumoniae*. A pan-genomic analysis were performed to identify the core genomes of the selected complete strains. Then the core genomes were subjected for the subtractive genomic analysis to identify the essential and non essential genes.

Then these essential and non essential genes were subjected for drug priotization which identify their cytoplasmic and membrane proteins with their druggability properties molecular weight, functionality, pathway, sub cellular localization , virulence. The membrane proteins were used as vaccine targets and cytoplasmic proteins were used ad drug targets. These targets were then used for the protein ligand interactions.

4.1 Identification of Core Genome of *C. pneumoniae* Strain

A Pan genome is a set of orthologous and unique genes found in a group of organisms. The core genome, accessory genome, and species- or strain-specific genes make up the pan-genome. The core genome identification includes selection of the genomes and the identifications of its genes by using the pan-genomic approach.

4.1.1 Selection of Genome

The twelve strains of the *C. pneumoniae* were selected for this study. The selection was based on the complete genomes strains data for the accuracy in the results. The gene and proteins data of these twelve strains were retrieved from the NCBI (https://www.ncbi.nlm.nih.gov/genome/browse /prokaryotes/171/). The table that is retrieved from the NCBI contains the whole genome information of the *C. pneumoniae*. The table which is shown in the NCBI home page contain 56 strains but after applying the filter on the table it gives the 12 strain which have complete genomes.

In this table 4.1 which is given below shows the organisms name their genomic statics which includes its bio sample , bio projects. This table also shown the strains names its GC contents also its size, level and assembly.

TABLE 4.1: Strains of *C. pneumoniae* Employed in the Drug Target Study with the Information on Genomics Statistics, Diseases Isolation, Location and Prevalence

Sr. No.	***C. pneumoniae* Strain**	**Bio Sample**	**Bio Project**	**Assembly**	**Level**	**Size (Mb)**	**GC %**
1	TW-183	SAMN0-2602979	PRJN-A420	GCA_000-007205.1	Complete	1.22-593	40.6
2	LPCo-LN	SAMN0-2604061	PRJN-A17947	GCA_000-024145.1	Complete	1.24-855	40.-4548
3	CWL-029	SAMN0-2603115	PRJN-A248	GCA_000-008745.1	Complete	1.23-023	40.6
4	CM1	SAMEA-3212818	PRJEB-8246	GCA_001-007125.1	Complete	1.22-989	40.6
5	AR39	SAMN0-2641564	PRJN-A247	GCA_000-091085.2	Complete	1.23-439	40.6
6	CWL-011	SAMEA-3212821	PRJE-B8246	GCA_001-007065.1	Complete	1.22-858	40.6

7	Wien3	SAMEA 3212835	PRJE B8246	GCA_001-007085.1	Complete	1.22-858	40.6
8	K7	SAMEA-3212826	PRJE-B8246	GCA_001-007045.1	Complete	1.22-852	40.6
9	PB2	SAMEA-3212830	PRJE-B8246	GCA_001-007145.1	Complete	1.22-813	40.6
10	Wien1	SAMEA-3212833	PRJE-B8246	GCA_001-007025.1	Complete	1.22-812	40.6
11	CV14	SAMEA-3212819	PRJE-B8246	GCA_001-007105.1	Complete	1.22-812	40.6
12	J138	SAMD0-0061097	PRJN-A257	GCA_000-011165.1	Complete	1.22-656	40.6

4.1.2 Identification of Core Genome Using Pan-Genomic Approach

The Pan-genomic analysis were performed using EDGAR tool. CWL029 strains were selected as reference genome and the rest of the strains were compared with (CWL029) reference strain. The results file includes the tags, locus and description of the genes along with FASTA format file (DNA and Proteins).The total genes which are identified in the in pan-genomes was 32,00 out of 972 are the core genes.

4.2 Subtractive Genome Analysis

Subtractive genome analysis includes two steps as follow

4.2.1 Identification of Non Host Homologous Proteins

The non homologous protein were identified by BLASTn using the default parameters against human genome to filter out the non homologous proteins. All the

35000 proteins sequences of the core genomes were subjected to BLASTn out of which 972 were non homologous proteins.

4.2.2 Identification of Essential Genes

The core 35000 non homologous proteins were subjected to the Database of Essential Genes (DEG) for the identification of the essential proteins, through which 16 proteins were obtained After finding the Non-Homologous sequences of the C.Pneumoniae. then blast these sequences against DEG databases.

TABLE 4.2: List of Essential Non-Homologous Proteins Pct Identity and E value

Sr. No	**Subject Id**	**E-value**	**Pct Identity**
1	DEG10380097	0.001	39.13
2	DEG10370096	0.001	38.239
3	DEG10580190	0.001	54.839
4	DEG10350210	0.001	48.485
5	DEG10330086	0.001	44.898
6	DEG10320168	0.001	35.922
7	DEG10110106	0.001	37.778
8	DEG10520181	0.001	35.897
9	DEG10470444	0.001	57.692
10	DEG10020286	0.001	37.313
11	DEG10520107	0.001	37.319
12	DEG10290105	0.001	35.052
13	DEG10270522	0.001	37.51
14	DEG10100462	0.001	36.667
15	DEG10450185	0.001	47.368

The 35000 genes of C.Pneumoniae were found from this database. To find the essential genes from this apply two parameters on this file the first parameter is

that the cut off values are (Pct identity should be > 35 and E value is = 0.001) Pct identity means (The percentage of cells where the gene is detected in the first group).

4.3 Drug Target Priotization Parameter

After finding the essential genes from DEG Database the multiple protocol was applied as shown in the table 4.2. The Query id is given to the Uniprot the molecular function, biological process, Protein and genes names is retrvied. After this the molecular weight of the genes which is find out by (ProtParam) tool.

The pathway analysis of the genes is done by the (KEGG). The Virulence is finding out by the (PAIDB) and the molecular and biological functions are checked from (Uniprot). The Cellular Localization is done through (CELLO) tool.

4.3.1 Molecular Weight

The Molecular weight of the protein is calculated from the ProtParam. (https://www.expasy.org/resources/protparam).

It calculates the MW in g/mol when it was converted into the kiloDa, all the Molecular Weight of the proteins was less than 100kDa , as shown in the table 4.3.

4.3.2 Sub cellular Localization

Out of 16 proteins 11 were cytoplasmic proteins and 1 were inner membrane protein 4 were extracellular protein as shown in the table 4.3.

4.3.3 Identification of Molecular and Biological Function

The molecular and biological functions of the proteins were retrieved from the Uniprot as well as their gene and protein name as shown in the table 4.3.

4.3.4 Identification of Virulence

All proteins shows the virulence property when these targets are BLAST against the Pathogenicity Island Database (PAIDB) (http://www .paidb .re.kr / about _ paidb .php). As shown in table 4.3.

4.3.5 Pathway Analysis

The Pathway analysis of the essential proteins were find out from the Kyoto Encyclopedia of Genes and Genomes (KEGG) (https://www.genome.jp/kegg/). The pathways of the targeted proteins were shown in table 4.3.

4.4 Protein Ligand Interactions

Protein-ligand interactions are facilitated through conformational shifts between low and high affinity states, as a result of molecular mechanics. Protein state and function are altered by ligand binding interactions [51].

To find the protein ligand interaction there is a two step identification process in first step is to detect the catalytic pocket of the targeted drug which shows the bindings sites for these drugs targets for the binding to the ligand and the second step is to perform the molecular docking. Protein state and function are altered by ligand binding interactions.

4.4.1 Catalytic Pocket Detection

The catalytic pocket detection of the essential proteins with specific drug score is find out through the DogsiteScorer. To identify the drug targets those catalytic pockets were selected whose druggability score was greater than 0.6.

Druggability score was greater than 0.6 is considered as good score but above than 0.8 score is not good.

Table 4.3: Drug Target Priotization Parameters of Twenty-three Non-Homologous Essential Genes

Sr No	Uniprot id	Molecular function	Biological Process	Protein name	Gene Name	Cellular localization	Vir.	Mol. Weight (Da)	Pathways
1	GATC_CHLPN	ATP binding glutaminyl-tRNA synthase (glutamine-hydrolyzing) activity	regulation of translational fidelity translation	Glutamyl-tRNA (Gln) amidotransferase subunit C	gatC	Cytoplasmic	Yes	1135-2.86	Mismatch repair Aminoacyl-tRNA biosynthesis Metabolic pathways Phosphotransferase system (PTS)

2	GATA_CHLPN	ATP binding Hydrolase activity	translation	Glutamyl-tRNA (Gln) amidotransferase subunit A	gatA	Cytoplasmic	Yes	5359-6.27	Aminoacyl-tRNA biosynthesis Metabolic pathways Phosphotransferase system (PTS)
3	GATB_CHLPN	ATP binding glutaminyl-tRNA synthase (glutamine-hydrolyzing) activity	translation	Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B	gatB	Cytoplasmic	Yes	5459-8.64	Aminoacyl-tRNA biosynthesis Metabolic pathways

									Phosphotransferase system (PTS)
4	A0A0F7WQS7_CHLPN	ATP binding glutaminyl-tRNA synthase (glutamine-hydrolyzing) activity	translation	DUF1978 domain containing protein	CPn_1054	Cytoplasmic	Yes	9345-6.03	No
5	NTPP_CHLPN	dTTP diphosphatase activity NADH pyrophosphatase activity	nucleotide metabolic process	Nucleoside triphosphate pyrophosphatase	2.2. CPn_0022	Cytoplasmic	Yes	2175-0.03	No

6	DNAJ_CHLPN	heat shock protein binding unfolded protein binding zinc ion binding	DNA replication protein folding response to heat	Chaperone protein DnaJ	dnaJ	Cytoplasmic	Yes	4210-8.55	No
7	XERC_CHLPN	DNA binding tyrosine-based sitespecific recombinase activity	cell cycle cell division chromosome segregation transposition, DNA-mediated	Tyrosine recombinase XerC	xerC	Cytoplasmic	No	3585-1.73	No

8	RNZ_ CHLPN	zinc ion binding 3'-tRNA processing endoribonuclease activity	tRNA processing	Ribonuclease Z	Rnz	Cytoplasmic	No	3489-9.83	No
9	LON_ CHLPN	sequence-specific DNA binding serine-type endopeptidase activity ATP binding	cellular response to heat protein quality control for misfolded or incompletely synthesized proteins	Lon protease	Lon	Cytoplasmic	No	9227-5.95	Terpenoid backbone biosynthesis Biotin metabolism Quorum sensing

									Base excision repair Peptidoglycan biosynthesis
10	TSAD_ CHLPN	iron ion binding N(6)-L-threonylcarbamoyladenine synthase activity	tRNA threonylcarbamoyladenosine modification	tRNA N6-adenosinethreonylcarbamoyltransferase	tsaD	Cytoplasmic Inner Membrane	Yes	3671-9.16	No

11	DNAJ_CHLPN	heat shock protein binding unfolded protein binding zinc ion binding	DNA replication protein folding response to heat	Chaperone protein DnaJ	dnaJ	Cytoplasmic	Yes	4210-8.55	No
12	PT1_CHLPN	kinase activity metal ion binding phosphoenol pyruvate protein phosphotransferase activity	phosphoenol-pyruvate-dependent sugar phosphotransferase system	Phosphoenol-pyruvate-protein phosphotransansferase	ptsI	Cytoplasmic	Yes	6347-9.05	Phosphotransferase system (PTS)

4.4.2 Molecular Docking

Molecular docking of the proteins is done through MOE tool. Molecular docking simulations for the selected set of targets proteins were performed by using MOE tool to predict the binding affinity between the molecules and ligand. For the molecular docking of the drug targets it includes the selection of the ligand , 3D structure prediction of the proteins and protein-ligand docking.

4.4.2.1 Selection of Ligand/Compounds

The structure of the selected ligand were constructed by using the MOE Builder tool. The selected compounds were modeled. The partial charges of these compounds was calculated and the energy of these compounds were minimized by using energy minimization algorithm with the default parameters. The minimized structure of all the compounds were saved in MDB format file. In the MOE-DOCK tool these prepared ligand were used as input source file for molecular docking.

4.4.2.2 3D Structure Prediction

The 3D structure prediction of the targeted proteins are predicted by the tool. The structure of all of the targets proteins was predicted as the structure of these proteins were not available in protein databank (PDB). Swiss Model web tool was used to predict the 3D of these targeted proteins.. The workflow of this tool includes the main steps first is data input. In data input all targeted proteins sequences (Fasta Format) were provided to this tool. Then second step is Data Search, in this for the provided data it searches its evolutionary related protein structure against Swiss-Model Template Library.

4.4.2.3 Validation of 3D Structure

Results visualization is done through Chimera tool. UCSF chimera tools prepare the structure of the receptor for the docking. It involves the deleting ligand and

solvent molecules and eliminate the alternate locations of the residues they change the selenomethiones to methionines and add the hydrogen atom and assign the charges to the protein atoms. In this design study the UCSF chimera tool is used in which the selected drug targets were visualized and delete the ligand and the solvent molecules and add the hydrogen atom and assign the charges to the proteins as they were shown in chapter 4 in the results figures that how they shows interaction with the active ligand.

4.5 Protein Model Validation

Protein model validation is done through the ERAAT and Ramachandra Plot. In Ramachandra plot when we plot a protein for the result validation if the value of the protein is it greater than 85% then the model is good/favorable if its value is lower than 85% than the model is not favorable. In DogSite scorer if the values lies 0.6 – 0.8 or above that this model is highly expectable if not then the model is not expectable.

Table 4.4: Protein Model Validation by ERAAT & Ramachandra Plot, DogSite Score

S. No	**Protein Name**	**ERAAT**	**Ramachandra Plot**	**DogSite Score**
1	Glutamyl-tRNA(Gln) amidotransferase subunit C	91.3793	89.9%	0.61
2	Glutamyl-tRNA(Gln) amidotransferase subunit A	88.0531	86.5%	0.85
3	Aspartyl/glutamyl-tRNA (Asn/Gln) amidotransferase subunit B	95.122	92.6%	0.81
4	DUF1978 domain-containing protein	86.9159	86.3%	0.81
5	Chaperone protein DnaJ	86.3563	83.7%	0.8

6	Tyrosine recombinase XerC	93.4256	90.6%	0.84
7	Ribonuclease Z	90.6087	86.1%	0.83
8	Lon protease	87.069	88.9%	0.81
9	tRNA N6-adenosine threonylcarbamoyltransferase	86.0248	88.0%	0.83
10	Chaperone protein	98.614	83.2%	0.89
11	Phosphoenolpyruvate-protein phosphotransferase	92.761	89.6%	0.87
12	Lon protease	87.069	88.9%	0.81

4.6 Molecular Docking

Molecular docking of the proteins is done through MOE tool. Molecular docking simulations for the selected set of proteins were performed by using MOE tool to predict the binding affinity between the molecules and ligand. Molecular docking were performed against 12 drug targets with drug like compounds in the MOE–DOCK tool. Out of 100 selected molecules these are re-docked again and top 10 molecules were selected. For each protein 1 of the best interaction were selected from top of these 10 molecules. Then after this for the results visualization 1 of the best interaction of protein-ligand is drawn in the chimera tool. The biological significance and the results analysis of the twelve predicted proteins are describe below.

4.6.1 Glutamyl-tRNA(Gln) Amidotranferase Subunit A

Glutamyl-tRNA(Gln) amidotranferase subunit A is a protein which plays important role in the protein biosynthesis. The glutamyl-tRNA protein has glutamine hydrolyzing activity its molecular function is that they play important role in the ATP binding. This protein belongs to amidase family. The primary function of this protein is to facilitate the synthesis of appropriately charged Gln-tRNA(Gln)

in species lacking glutaminyl-tRNA synthetase through transamidation of misacylated Glu-tRNA(Gln).The top 10 best confirmation of the protein is shown in table 4.5 along with their scientific names of compounds , S energy, number of interactions and its interactive residues. The residues Glu563,Lys556 were found to interact with the active ligand prostaglandin h2_SID_3717 which is found in the figure 4.1.

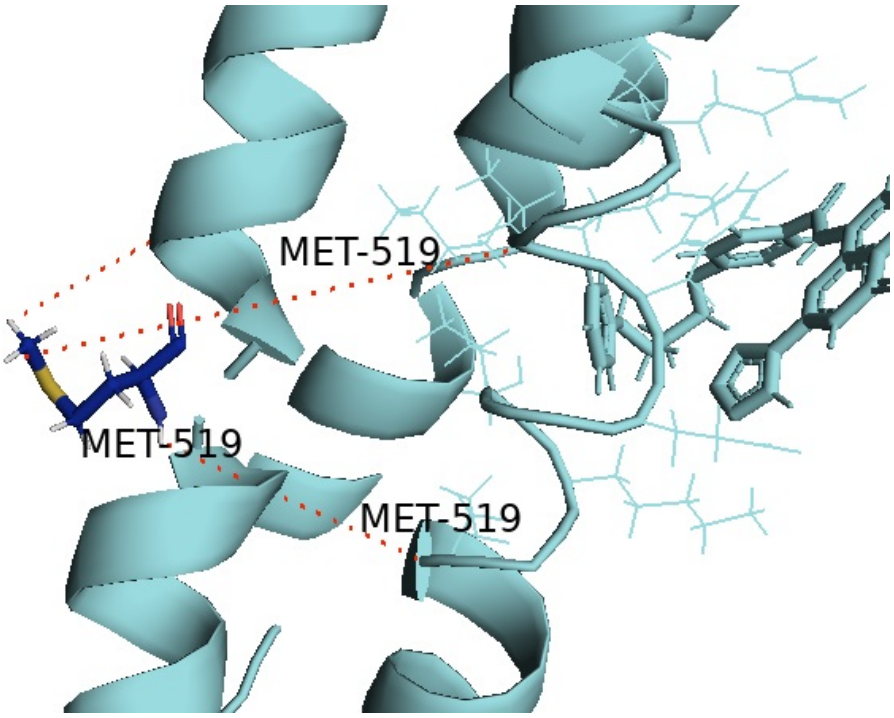

FIGURE 4.1: Interaction of Gln aminotransferase subunit A with SID_3717

TABLE 4.5: Scientific Names of Compounds, Minimized Energy, Number of Interactions and Interactive Residues of Glutamyl-tRNA(Gln) Amidotranferase Subunit A

Compound Names	**Interactions**	**S value**
Pranlukast_SID_96025094	Lys 556, Glu 563	-24.2984
Beraprost_SID_17396887	Phe 552,Arg 490	-22.9648
Arachindonate_SID_3519	Glu553, Lys 556	-22.0476
Prostaglandin h2_SID_3717	Glu563,Lys556,Met519	-21.2720
Montelukast_SID_96024917	Lys556	-21.1207
Zafirlukast_SID_7847477	Glu563	-20.8192
Arachindonate_SID_3519	Glu 505,Cys 607	-20.6823

5 (S)-HPETE_SID_7733	Glu 505,Lys 601	-20.5818
lloprost_SID_17396888	Glu 505,Glu 605	-20.5447
Seratrodast_SID_7848186	Lys 601	-20.5000

4.6.2 Aspartyl/glutamyl-tRNA (Asn/Gln) Amido- transferase Subunit

Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit is a protein which plays role in the regulation of the translational fidelity. It has asparaginyl-tRNA synthase (glutamine-hydrolyzing) activity. The biological function of this protein is proteins biosynthesis. It belongs to the Gatc family. This protein's primary function is to facilitate the production of properly charged Asn-tRNA(Asn) or Gln-tRNA(Gln) by transamidation of misacylated Asp-tRNA(Asn) or Glu-tRNA(Gln) in species lacking one or both asparaginyl-tRNA or glutaminyl-tRNA synthetases. In the presence of glutamine and ATP, the process is catalyzed by an activated phospho-Asp-tRNA(Asn) or phospho-Glu-tRNA (Gln).

The top 10 best confirmation of the protein is shown in table 4.6 along with their scientific names of compounds , S energy, number of interactions and its interactive residues. The residues Lys556 shows interaction with the active ligand Zafirlukast=SID=7847477 as shown in the figure 4.2.

TABLE 4.6: Scientific names of Compounds, Minimized Energy ,Number of Interaction and Interactive Residues of Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B

Compound Names	**Interactions**	**S value**
Pranlukast_SID_96025094	Glu 553, Arg 490	-24.2964
Beraprost_SID_17396887	Lys 556, Glu 563	-22.9642
Pranlukast_SID_960250094	Phe 552,Arg 490	-22.1334
Arachidonate_SID_3519	Glu553, Lys 556	-22.0476
Beraprost_SID_17396887	Glu563,Lys556	-21.7803
Beraprost_SID_17396887	Lys556	-21.4273

prostaglandin h2_SID_3717	Glu563	-21.2720
Montelukast_SID_96024917	Lys 556,Glu563	-21.1207
prostaglandin h2_SID_3717	Glu 563	-20.9270
Zafirlukast_SID_7847477	Glu553,ILE494,Arg490	-20.8192

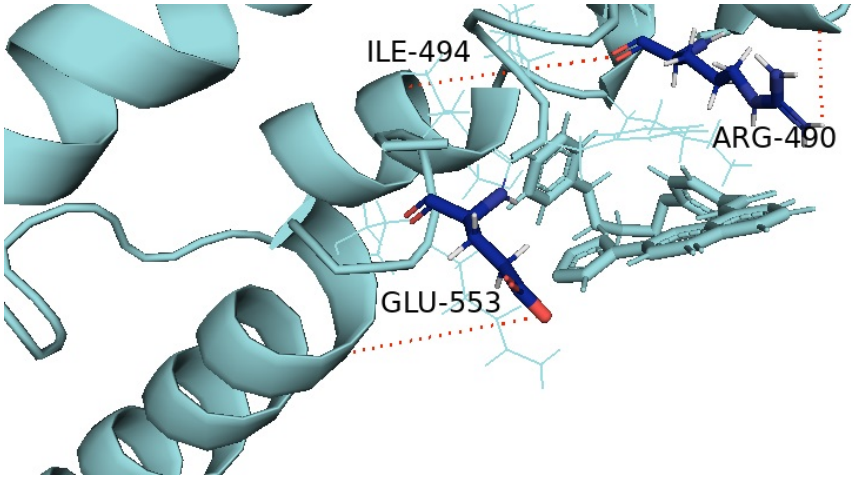

FIGURE 4.2: Interaction of Glu aminotransferase subunit B with SID_7847477

4.6.3 DUF1978 Domain-Containing Proteins

'DUF' families are annotated with the domain of unknown function. Members of this family have been discovered in a variety of putative proteins generated by the bacterium *C. pneumoniae*, but their exact function has yet to be determined. The top 10 best confirmation of the protein is shown in table 4.7 along with their scientific names of compounds , S energy, number of interactions and its interactive residues. The residues Lys 601 shows the interaction with the active ligand 5 (S)-HPETE_SID_7733 as shown in the figure 4.3.

TABLE 4.7: Scientific Names of Compounds, Minimized Energy , Number of Interactions and Interactive Residues of DUF1978 Domain-containing Proteins

Compound Names	**Interactions**	**S value**
Pranlukast_SID_96025094	Glu598	-22.1996
Beraprost_SID_17396887	Glu 505	-22.0718
Pranlukast_SID_96025094	Glu 598	-21.8898

Arachindonate_SID_3519	Glu564,Thr568,His 509	-21.4028
prostaglandin G2_SID_8240	Lys601	-20.9432
Montelukast_SID_96024917	Leu 576	-20.2465
Pranlukast_SID_96025094	Glu 505,Cys 607	-20.0119
5 (S)-HPETE_SID_7733	Glu 505,Lys 601	-19.6989
Seratrodast_SID_7848186	Glu 505,Glu 605	-19.5409
5 (S)-HPETE_SID_7733	Cys 604,Glu505	-19.4721

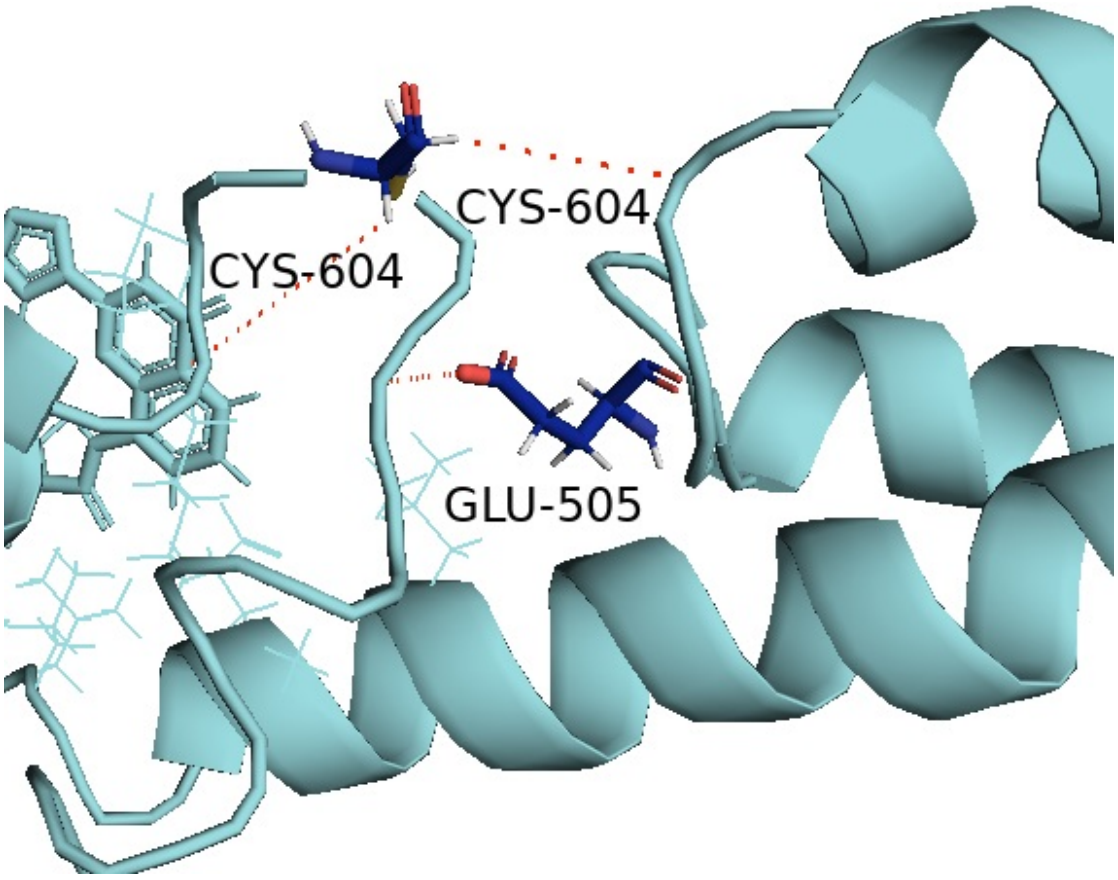

FIGURE 4.3: Interaction of DUF1978 domain-containing proteins with SID_7733

4.6.4 Chaperone Protein DnaJ

Chaperone protein DnaJ is a protein that plays role in the DNA replication , protein folding , protein refolding , chaperons bindings , zinc ion bindings , heat shock protein bindings. The biological process of this protein is that it to the heat

response. It belongs to the DnaJ family. The top 10 best confirmation of the protein is shown in table 4.8 along with their scientific names of compounds , S energy, number of interactions and its interactive residues. The residues Gln Ar97 interacts with active ligand Montelukast_SID_96024917 as shown in the figure 4.4.

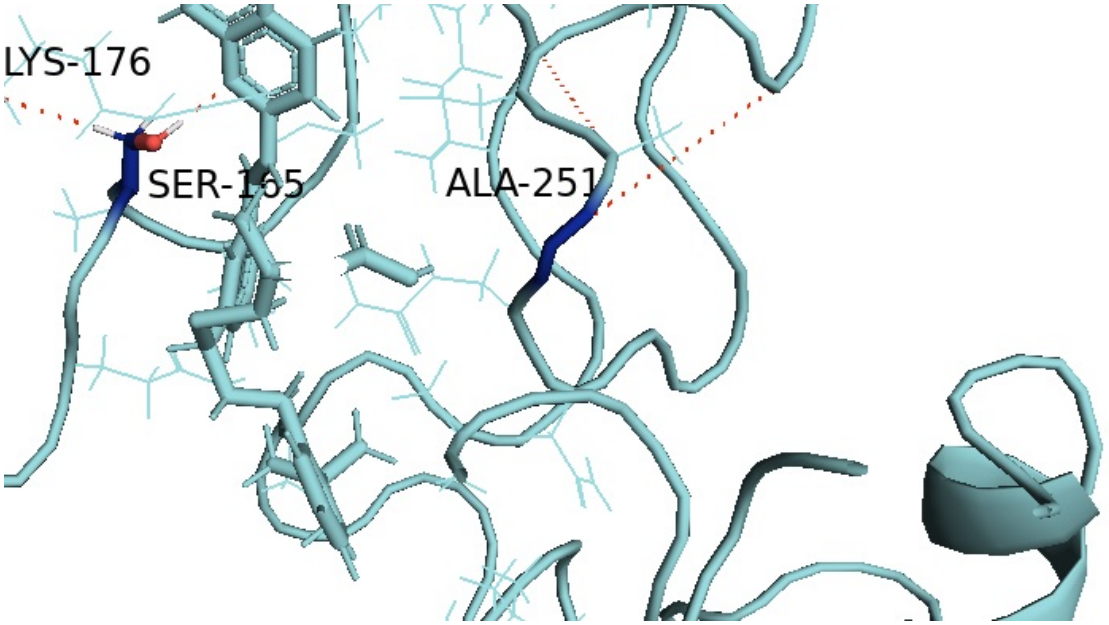

FIGURE 4.4: Interaction of Chaperone DNJ protein with SID_96025094

TABLE 4.8: Scientific Names of Compounds, Minimized Energy, Number of Interaction and Interactive Residues of Chaperone Protein DnaJ

Compound Names	**Interactions**	**S value**
Montelukast_SID_96024917	Gln Ar97	-28.8420
Montelukast_SID_96024917	Gln Ar97	-28.7557
Montelukast_SID_96024917	Gly A252,Lys Af76	-27.1135
Pranlukast_SID_96025094	Lys A176,Ser A165,ALA251	-26.9017
Pranlukast_SID_96025094	Lys Af78	-26.5864
Leukotriene A4_SID_4164	Asp A223	-25.8941
Ozagrel_SID_96025014	Gln A253,Asp A255,Lys A226	-25.2917
Montelukast_SID_96024917	Arg A222,Lys A225	-24.5875
Leukotriene B4_SID_5240	Gly A166	-24.5733
Montelukast_SID_96024917	Lys A176	-24.4978

4.6.5 Tyrosine Recombinase XerC

Tyrosine recombinase XerC is a protein which plays role in the transposition ,DNA mediated, chromosomal segregation , cell cycle and cell division. Site-specific tyrosine recombinase is a catalytic enzyme that catalyses the cutting and rejoining of recombining DNA molecules. Forms the heterotetrameric XerC-XerD complex, which recombines DNA substrates by attaching cooperatively to particular DNA consensus sequences separated from XerD binding sites by a short central region. The top 10 best confirmation of the protein is shown in table 4.9 along with their scientific names of compounds , S energy, number of interactions and its interactive residues. The residues Ser 286,Glu288,Glu54 interacts with active ligand of the prostaglandin h2_SID_3717 as shown in the figure 4.5.

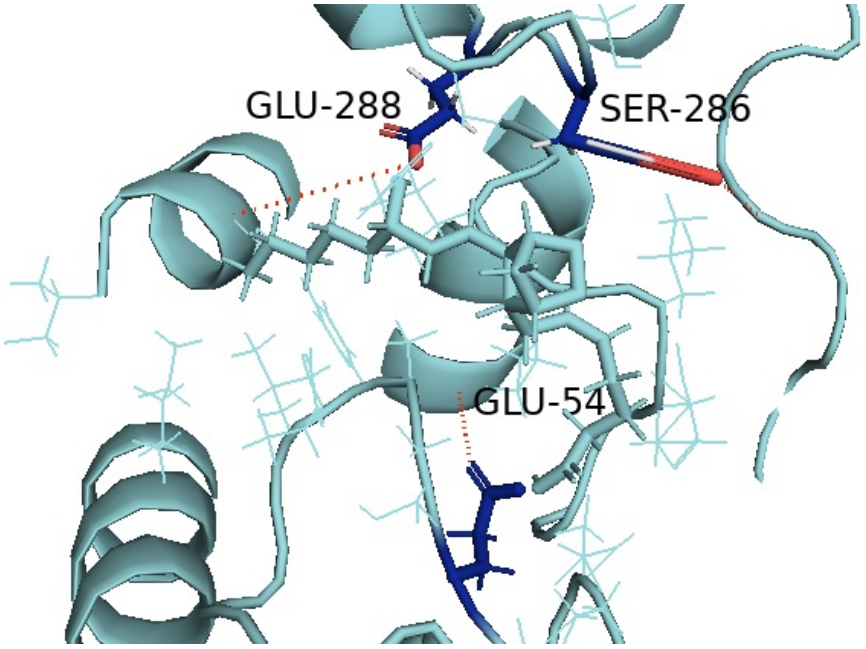

FIGURE 4.5: Interaction of Tyrosine XerC with SID_3717

TABLE 4.9: Scientific Names of Compounds, Minimized Energy, Number of Interactions and Interactive Residues of Tyrosine Recombinase XerC

Compound Names	**Interactions**	**S value**
Pranlukast_SID_96025094	Lys 34	-23.3437
Montelukast_SID_96024917	Arg 56	-22.1496
prostaglandin h2_SID_3717	Ser 286,Glu288,Glu54	-21.8706

Pranlukast_SID_96025094	Ala 52	-21.3714
Montelukast_SID_96024917	Ser 4	-21.2954
Zafirlukast_SID_7847477	Glu258	-21.1332
Montelukast_SID_96024917	Lys 55	-21.1099
Pranlukast_SID_96025094	Glu54, Thr67	-21.1010
Pranlukast_SID_96025094	Glu54,Glu110	-20.5464
Pranlukast_SID_96025094	Glu110	-20.1532

4.6.6 Ribonuclease Z

Ribonuclease Z plays role in the cell proliferation, regulation of cell growth. Its molecular function is metal ion binding and endoribonuclease activity. It also plays a function in the activity of nuclear and mitochondrial pre-tRNA tRNA 3'-processing endonuclease. By eliminating a 3'-trailer from precursor tRNA, it is most likely involved in tRNA maturation. In the developing embryo, it may play a role in tRNA processing. The top 10 best confirmation of the protein is shown in table 4.10 along with their scientific names of compounds , S energy, number of interactions and its interactive residues. The residues His B65 shows interaction with the active ligand 5 (S)-HPETE_SID_7733 as shown in figure 4.6.

TABLE 4.10: Scientific Names of Compounds, Minimized Energy, Number of Interactions and Interactive Residues of Ribonuclease Z

Compound Names	**Interactions**	**S value**
Montelukast_SID_96024917	His B242, His265	-27.3181
Montelukast_SID_96024917	Tyr A109.His B65, His B242	-26.0614
Montelukast_SID_96024917	Leu A31	-26.0465
Zafirlukast_SID_7847477	Asp A52	-24.7485
Beraprost_SID_17396887	His A110	-23.7591
5 (S)-HPETE_SID_7733	His B65,Tyr109,His242	-23.4856
Leukotriene A4_SID_4164	His A110	-23.1132
Montelukast_SID_96024917	His A110,His B242	-22.8109

Beraprost_SID_17396887	His B242	-22.6477
prostaglandin h2_SID_3717	His B242, His A110,Thr A107	-22.5890

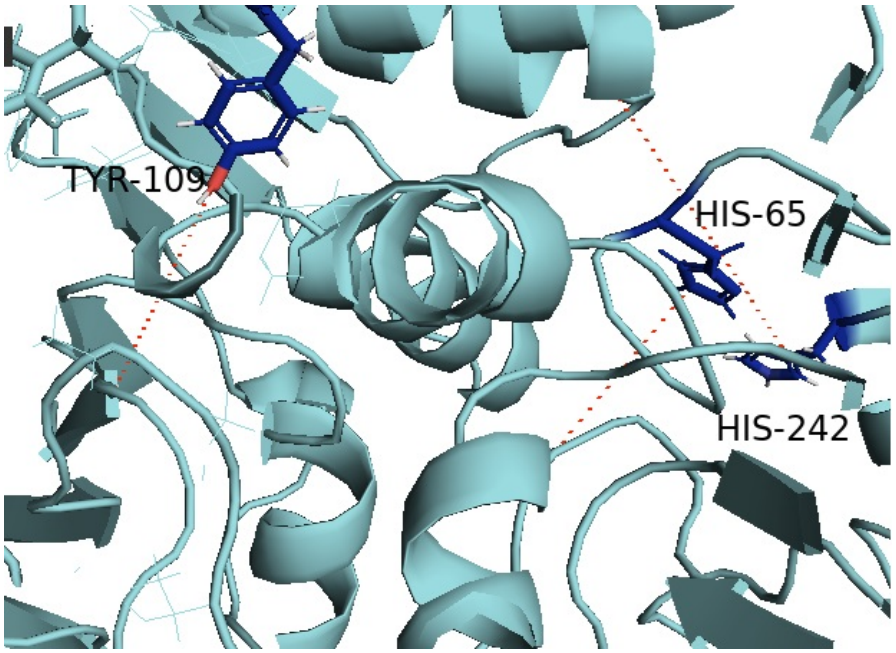

FIGURE 4.6: Interaction of Ribonulease with SID_7733

4.6.7 Lon Protease

Lon protease is a kind of protease. Lon, also known as Lonp1, is a nuclear-encoded mitochondrial ATP-dependent serine peptidase that aids in the maintenance of mitochondrial homeostasis by mediating the selective destruction of mutant and aberrant proteins in the organelle.. The top 10 best confirmation of the protein is shown in table 4.11 along with their scientific names of compounds , S energy, number of interactions and its interactive residues. The residues Lys90,His104,Arg213 shows interaction with the active ligand Pranlukast_SID_96025094 as shown in the figure 4.7.

TABLE 4.11: Scientific Names of Compounds, Minimized Energy, Number of Interactions and Interactive Residues of Lon protease

Compound Names	**Interactions**	**S value**
Pranlukast_SID_96025094	Lys90,His104,Arg213	-22.9733

thromboxane A2_SID_5266	Lys50,Arg213	-21.8209
5 (S)-HPETE_SID_7733	Glu84,Glu42,Phe44	-21.7034
Zafirlukast_SID_7847477	Arg213	-21.5945
Zafirlukast_SID_7847477	Lys59	-21.4095
Pranlukast_SID_96025094	Arg213,His104.Lys90	-21.2624
prostacyclin_SID_7847174	Arg212	-21.2179
Montelukast_SID_96024917	Thr210	-20.8665
5 (S)-HPETE_SID_7733	Glu218,Glu221	-20.6385
Zafirlukast_SID_7847477	Lys90,Glu91	-20.5336

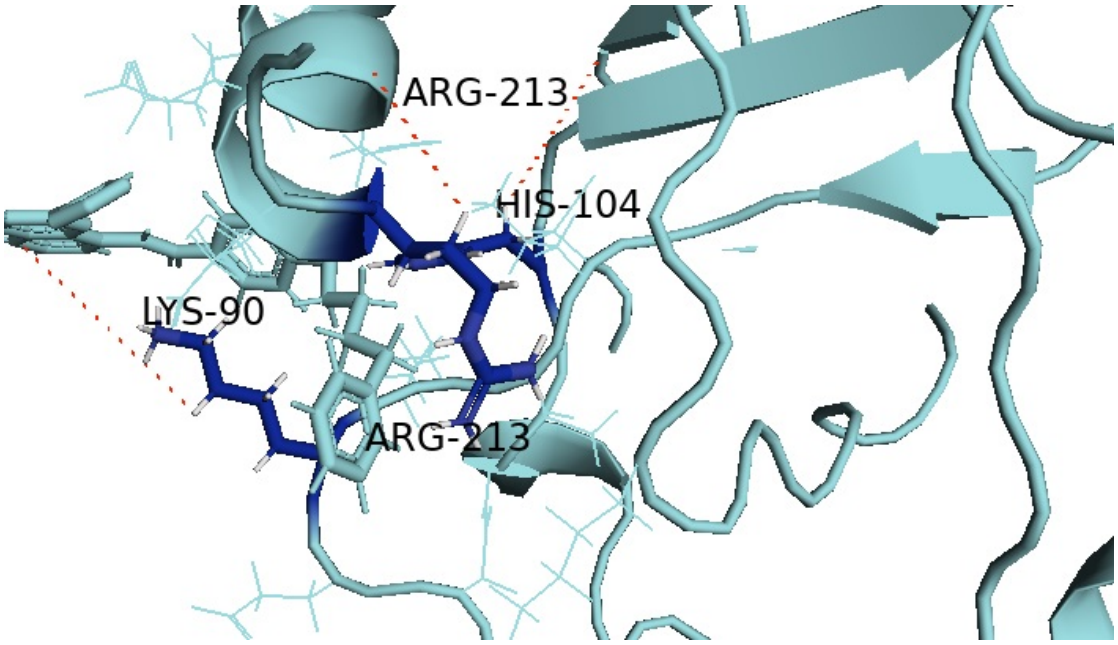

FIGURE 4.7: Interaction of Lon protease with SID_96025094

4.6.8 tRNA N6-Adenosine Theronylcarbamoyltransferase

In tRNAs that read codons that start with adenine, this component of the EKC/ KEOPS complex is essential for the production of a threonylcarbamoyl group on adenosine at position 37. (t6A37). The complex is thought to play a role in the transfer of the threonylcarbamoyl moiety of threonylcarbamoyl-AMP (TC-AMP) to the N6 group of A37. This reaction probably certainly includes OSGEP as a catalytic component. The top 10 best confirmation of the protein is shown in

table 4.12 along with their scientific names of compounds , S energy, number of interactions and its interactive residues. The residues Asp 303 shows interaction with the active ligand Pranlukast$_ SID $ _ 96025094 as shown in the figure 4.8.

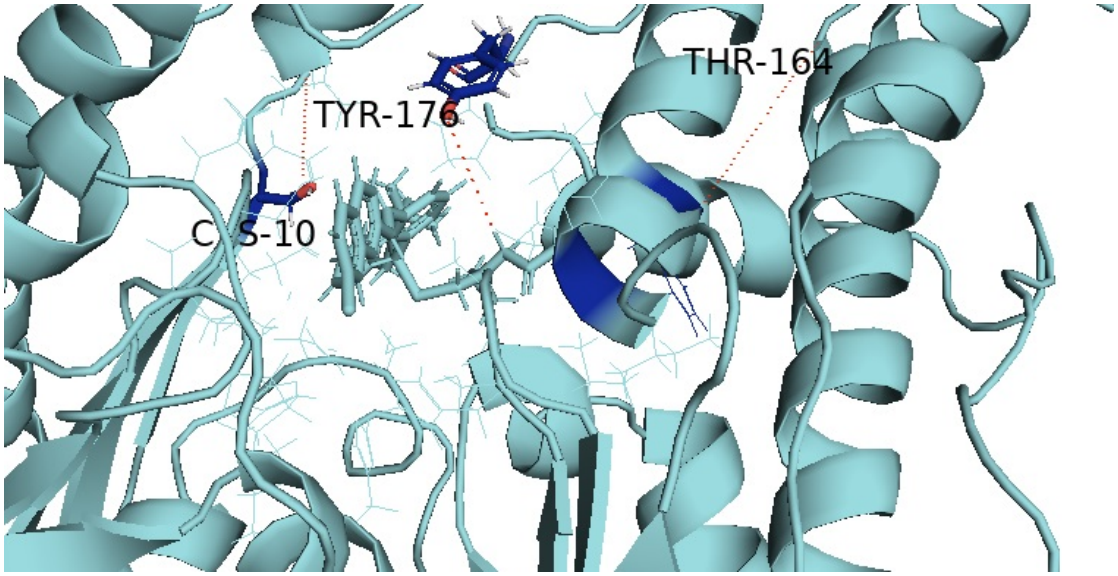

FIGURE 4.8: Interaction of tRNA N6 with SID_96025094

TABLE 4.12: Scientific Names of Compounds, Minimized Energy , Number of Interactions and Interactive Residues of tRNA N6-adenosine theronylcarbamoyltransferase

Compound Names	**Interactions**	**S value**
Pranlukast_SID_96025094	Asp 203	-26.0179
Leukotriene B4_SID_5240	Asp 303	-25.7693
Montelukast_SID_96024917	Ser 202	-25.3425
Pranlukast_SID_96025094	Asp 303,Glu163	-24.0489
Montelukast_SID_96024917	Ser21,Gly162	-23.9090
Pranlukast_SID_96025094	Tyr 176,Asp 303	-23.8904
Zafirlukast_SID_7847477	Asp 11, Glu12	-23.5767
Montelukast_SID_96024917	Gly 274,Tyr176	-23.2768
Pranlukast_SID_96025094	Tyr176,Thr164,Cys10	-21.7647
Pranlukast_SID_96025094	Asp 303	-21.1555

4.6.9 Phosphoenolpyruvate-protein Phosphotransferase

Phosphoenolpyruvate-protein phosphotransferase is a protein that plays important role in the phosophotransferase system and sugar transport system. It performs key role in the metal bindings. It belongs to the PEP-utilizing enzyme family. The top 10 best confirmation of the protein is shown in table 4.13 along with their scientific names of compounds , S energy, number of interactions and its interactive residues. The residues Arg A209, Arg A530 shows interaction with the active ligand Montelukast_SID_ 96024917 as shown in the figure 4.9.

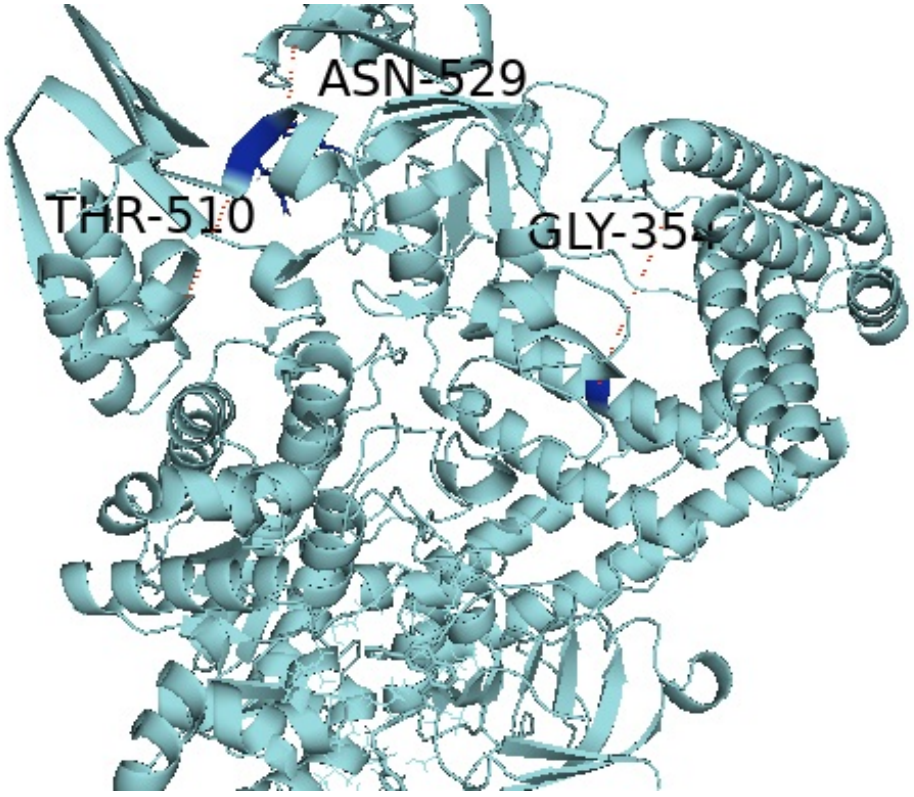

FIGURE 4.9: Interaction of Phosphotransferase with SID_96024917

TABLE 4.13: Scientific Names of Compounds, Minimized Energy, Number of Interactions and Interactive Residues of Phosphoenolpyruvate-protein phosphotransferase

Compound Names	**Interactions**	**S value**
Montelukast_SID_96024917	Met A23,Leu A25	-24.2922
Montelukast_SID_96024917	Glu A250	-24.1692
Pranlukast_SID_96025094	Arg A530	-23.5223

Montelukast_SID_96024917	His A470	-21.5141
Pranlukast_SID_96025094	Arg A530	-21.4228
5 (S)-HPETE_SID_7733	Tyr A240	-20.4494
5 (S)-HPETE_SID_7733	Ser A21	-20.1499
Montelukast_SID_96024917	Asn 529, Gly35,Thr 510	-19.9437
Ramatroban_SID_7848191	Thr A474,Arg A530	-19.6889
Montelukast_SID_96024917	Arg A209, Arg A530	-19.6264

Chapter 5

Conclusion and Future Prospects

Chlamydia Pneumoniae is a intracellular Gram negative bacteria which causes the respiratory tract diseases. As there is no specific drug orvaccine available against this bacteria therefore general antibiotics such as doxycycline , azithromycin are used, which are not only less effective but also do not treat the infection completely resulting in reoccurrence of infection. As the pathogen is present worldwide, therefore the variations in the genome also make a single treatment against the diseases less effective. There is a dire need to have an effective and efficient antibiotic against the pathogen. This research used pan genome and subtractive genome approach to identify core genome drug targets in the pathogen as potential drug targets. In order to identify the potential drug targets., the 12 strains were selected out of total 56 strains, selection of strains was done based on the availability of annotated whole genome. Pan genome analysis was performed and core as well as variable genome was identified. Out of core genome, essential genes were prioritized and subtractive genomic analysis was performed to identify non homologous genes. These 16 genes were found to play significant role not only in the growth and survival of *C. pneumoniae* but also in the development of the diseases/infections in the host (human).

Among these non homologous genes, the drug target Prioritization were performed. based on different factors including the molecular weight, pathway analysis ,virulence, molecular and biological processes, sub cellular localization. The first

parameter was molecular weight and the thresh hold was the proteins with molecular weight less than 100kDa, in addition to this pathways of the proteins were also studies to understand their biological role and significance. the proteins were found to be part of pathways including pathways for biosynthesis, DNA mediation, cell cycle, cell growth and cell proliferation and plays vital role in the survival of the bacteria to targets the host. Virulence factor was determined by using blast against the virulent factor database (Vfdb). Only virulent proteins were selected.

The biological and molecular roles provide the information about the process of replication, transcription and translation of the genetic material and also gives the functional role of the *C. pneumoniae*. Cellular localization was analyzed and it was found that out of 16 genes 11 were cytoplasmic 1 is inner membrane and the 4 were extracellular membrane protein. In this study the 12 cytoplasmic proteins were used for the further analysis.

Inhibition of the drug targets with plant drive natural compound was performed. The docking of these selected targets were done which shows the favorable interactions with the compound/ligand which were retrieved from the literature review. All of these interactions shows the binding affinity with the ligand at very low binding affinity which shows that they are stable molecules. The docking is done through the MOE-DOCKING tool , the docking results indicates that the selected 9 drug/vaccine targets can be used in future for designing the new drug/vaccine which have less possibility of side effect to prevent the onset of the diseases.

Bibliography

[1] Grayston, J. Thomas, Cho-Chou Kuo, San-pin Wang, and Jeff Altman. "A new Chlamydia psittaci strain, TWAR, isolated in acute respiratory tract infections." N0ew England Journal of Medicine 315, no. 3 (1986): 161-168.

[2] Kuo, C. C., & Grayston, J. T. (1990). A sensitive cell line, HL cells, for isolation and propagation of *Chlamydia pneumoniae* strain TWAR. Journal of Infectious Diseases, 162(3), 755-758.

[3] Sillis, M., & White, P. M. B. (1990). Rapid identification of Chlamydia psittaci and TWAR (C pneumoniae) in sputum samples using an amplified enzyme immunoassay. Journal of clinical pathology, 43(3), 260.

[4] MESSMER, T. O., BLACK, C. M., & THACKER, W. L. (1994). Mycoplasma contamination of chlamydiae isolated from clinical specimens. APMIS, 102(7-12), 793-796.

[5] Shor, A. C. C. K., Kuo, C. C., & Patton, D. L. (1992). Detection of *Chlamydia pneumoniae* in coronary arterial fatty streaks and atheromatous plaques. South African medical journal= Suid-Afrikaanse tydskrif vir geneeskunde, 82(3), 158-161.

[6] Davidson, M., Kuo, C. C., Middaugh, J. P., Campbell, L. A., Wang, S. P., & Finley, J. C. (1995). *Chlamydia pneumoniae* (TWAR) in Alaska Natives with coronary atheroma. Clin Infect Dis, 21, 726.

[7] Campbell, L. A., Kuo, C. C., Wang, S. P., & Grayston, J. T. (1990). Serological response to *Chlamydia pneumoniae* infection. Journal of Clinical Microbiology, 28(6), 1261-1264.

[8] Shor, A. C. C. K., Kuo, C. C., & Patton, D. L. (1992). Detection of *Chlamydia pneumoniae* in coronary arterial fatty streaks and atheromatous plaques. South African medical journal= Suid-Afrikaanse tydskrif vir geneeskunde, 82(3), 158-161.

[9] Muhlestein, J. B., Hammond, E. H., Carlquist, J. F., Radicke, E., Thomson, M. J., Karagounis, L. A., ... & Anderson, J. L. (1996). Increased incidence of Chlamydia species within the coronary arteries of patients with symptomatic atherosclerotic versus other forms of cardiovascular disease. Journal of the American College of Cardiology, 27(7), 1555-1561.

[10] Ong, G., Thomas, B. J., Mansfield, A. O., Davidson, B. R., & Taylor-Robinson, D. (1996). Detection and widespread distribution of *Chlamydia pneumoniae* in the vascular system and its possible implications. Journal of clinical pathology, 49(2), 102-106.

[11] Fischer, S. A., Gupta, S., Leatham, E. W., Carrington, D., Mendall, M. A., Kaski, J. C., & Camm, A. J. (1998). Elevated *Chlamydia pneumoniae* Antibodies, Cardiovascular Events, and Azithromycin in Male Survivors of Myocardial Infarction. Infectious Diseases in Clinical Practice, 7(5), 220.

[12] Hanski, L., & Vuorela, P. (2016). Lead discovery strategies for identification of chlamydia pneumoniae inhibitors. Microorganisms, 4(4), 43.

[13] Cho YS, Kim TB, Lee TH, Moon KA, Lee J, Kim YK, Lee KY, Moon HB (December 2005). "*Chlamydia pneumoniae* infection enhances cellular proliferation and reduces steroid responsiveness of human peripheral blood mononuclear cells via a tumor necrosis factor-alpha-dependent pathway". Clinical and Experimental Allergy. 35 (12): 1625–31.

[14] Takaoka N, Campbell LA, Lee A, Rosenfeld ME, Kuo CC (February 2008). *Chlamydia pneumoniae* infection increases adherence of mouse macrophages to mouse endothelial cells in vitro and to aortas ex vivo". Infection and Immunity. 76 (2): 510–4.

[15] Mussa FF, Chai H, Wang X, Yao Q, Lumsden AB, Chen C (June 2006). "*Chlamydia pneumoniae* and vascular disease: an update". Journal of Vascular Surgery. 43 (6)

[16] Hahn DL, McDonald R (October 1998). "Can acute *Chlamydia pneumoniae* respiratory tract infection initiate chronic asthma?". Annals of Allergy, Asthma & Immunology. 81 (4): 339–44.

[17] "*Chlamydia pneumoniae*". Taxonomy Browser. National Center for Biotechnology Information (NCBI), U.S. National Library of Medicine. Retrieved 2009-01-27.

[18] A. Collingro, P. Tischler, T. Weinmaier, T. Penz, E. Heinz, R.C. Brunham, T.D. Read, P.M. Bavoil, K. Sachse, S. Kahane, M.G. Friedman, T. Rattei, G.S.A. Myers, M. Horn Unity in variety—the pan-genome of the Chlamydiae Mol. Biol.

[19] World Health Organisation Department of HIV/AIDS Chlamydia Global Prevalence and Incidence of Selected Curable Sexually Transmitted Infections, World Health Organisation (2001) D.S.A. Beeckman, D.C.G. Vanrompay Zoonotic Chlamydophila psittaci infections from a clinical perspective Clin. Microbiol. Infect., 15 (2009), pp. 11-17

[20] Elwell C, Mirrashidi K, Engel J.Chlamydia cell biology and pathogenesis. Nat Rev Microbiol. 2016; 14(6):385–400.

[21] Nunes A, Gomes J. Evolution, phylogeny, and molecular epidemiology of Chlamydia. Infect Genet Evol. 2014; 23:49–64.

[22] Taylor-Brown A, Bachmann NL, Borel N, Polkinghorne A. Culture-independent genomic characterisation of Candidatus Chlamydia sanzinia, a novel uncultivated bacterium infecting snakes. BMC Genomics. 2016; 17(1):710.

[23] Staub E, Marti H, Biondi R, Levi A, Donati M, Leonard C, Ley S, Pillonel T, Greub G, Seth-Smith H, Borel N. Novel Chlamydia species isolated from snakes are temperature-sensitive and exhibit decreased susceptibility to azithromycin. Sci Rep. 2018; 8:5660.

[24] Seth-Smith H, Wanninger S, Bachmann N, Marti H, Qi W, Donati M, Francesco A, Polkinghorne A, Borel N. The Chlamydia suis genome exhibits high levels of diversity, plasticity, and mobile antibiotic resistance: Comparative genomics of a recent livestock cohort shows influence of treatment regimes. Genome Biol Evol. 2017; 9:750–60

[25] Taylor-Brown A, Spang L, Borel N, Polkinghorne A. Culture-independent metagenomics supports discovery of uncultivable bacteria within the genus Chlamydia. Sci Rep. 2017; 7:10661.

[26] European Centre for Disease Prevention and Control. Guidance on chlamydia control in Europe. Stockholm: ECDC; 2016.

[27] Hocking JS, Temple-Smith M, Guy R, Donovan B, Braat S, Law M, Gunn J, Regan D, Vaisey A, Bulfone L, Kaldor J, Fairley CK, Low N. Population effectiveness of opportunistic chlamydia testing in primary care in australia: a cluster-randomised controlled trial. Lancet. 2018; 392:1413–22.

[28] Read, T. D., Brunham, R. C., Shen, C., Gill, S. R., Heidelberg, J. F., White, O., ... & Fraser, C. M. (2000). Genome sequences of Chlamydia trachomatis MoPn and *Chlamydia pneumoniae* AR39. Nucleic acids research, 28(6), 1397-1406.

[29] Shirai, M., Hirakawa, H., Kimoto, M., Tabuchi, M., Kishi, F., Ouchi, K., ... & Nakazawa, T. (2000). Comparison of whole genome sequences of *Chlamydia pneumoniae* J138 from Japan and CWL029 from USA. Nucleic Acids Research, 28(12), 2311-2314.

[30] Kuo, C. C., Jackson, L. A., Campbell, L. A., & Grayston, J. T. (1995). *Chlamydia pneumoniae* (TWAR). Clinical microbiology reviews, 8(4), 451-461.

[31] Wang SP ,Grayston JT. Immunologic relationship between genital TRIC, lymphogranuloma venereum, and related organisms in a new microtiter indirect immunofluorescence test, Am J Ophthalmol, 1970, vol. 70(pg. 367-74)

[32] Grayston JT. *Chlamydia pneumoniae*, strain TWAR pneumonia, Annu Rev Med, 1992, vol. 43 (pg. 317-23).

[33] Molestina RE, Dean D, Miller RD, Ramirez JA, Summersgill JT. Characterization of a strain of C. pneumoniae isolated from a coronary atheroma by analysis of the omp1 gene and biological activity in human endothelial cells. Infect Immun 1998;66:1370–6. 17.

[34] Jackson LA, Campbell LA, Kuo CC, Rodriguez DI, Lee A, Grayston JT. Isolation of *Chlamydia pneumoniae* from a carotid endarterectomy specimen. J Infect Dis 1997;176:292–5.

[35] Saikku P, Wang SP, Kleemola M, Brander E, Rusanen E, Grayston JT. An epidemic of mild pneumonia due to an unusual strain of Chlamydia psittaci. J Infect Dis 1985;151:832–9.

[36] Aldous MB, Wang SP, Foy HM, Grayston JT. *Chlamydia pneumoniae*, strain TWAR, infection in Seattle children and their families, 1965–1979. In: Bowie WR, Caldwell HD, Jones RP, et al., eds. Chlamydial infections. Cambridge, UK: Cambridge University Press, 1990:437–40.

[37] Patnode D, Wang SP, Grayston JT. Persistence of *Chlamydia pneumoniae*, strain TWAR micro-immunofluorescent antibody. In: Bowie WR, Caldwell

HD, Jones RP, et al., eds. Chlamydial infections. Cambridge, UK: Cambridge University Press, 1990:406–9.

[38] Wang SP, Grayston JT. Population prevalence antibody to *Chlamydia pneumoniae*, strain TWAR. In: Bowie WR, Caldwell HD, Jones RP, et al., eds. Chlamydial infections. Cambridge, UK: Cambridge University Press, 1990:402–5.

[39] Molestina RE, Dean D, Miller RD, Ramirez JA, Summersgill JT. Characterization of a strain of C. pneumoniae isolated from a coronary atheroma by analysis of the omp1 gene and biological activity in human endothelial cells. Infect Immun 1998;66:1370–6.

[40] Jackson LA, Campbell LA, Kuo CC, Rodriguez DI, Lee A, Grayston JT. Isolation of *Chlamydia pneumoniae* from a carotid endarterectomy specimen. J Infect Dis 1997;176:292–5.

[41] Saikku P, Wang SP, Kleemola M, Brander E, Rusanen E, Grayston JT. An epidemic of mild pneumonia due to an unusual strain of Chlamydia psittaci. J Infect Dis 1985;151:832–9.

[42] Schachter J, Dawson CR. Human chlamydial Infections. Littleton, MA: PSG Publishing, 1978:39–41.

[43] Wright SW, Edwards KM, Decker MD, Grayston JT, Wang SP. Prevalence of positive serology for acute *Chlamydia pneumoniae* infection in emergency department patients with persistent cough. Acad Emerg Med 1997;4:179–83.

[44] Schachter J, Dawson CR. Human chlamydial Infections. Littleton, MA: PSG Publishing, 1978:175–7.

[45] Beatty W, Morrison R, Byrne G. Persistent chlamydiae: from cell culture to a paradigm for chlamydial pathogenesis. Microbiol Rev 1994;58:686–99.

[46] Grayston JT, Mordhorst CH, Bruu AL, Vene S, Wang SP. Countrywide epidemics of *Chlamydia pneumoniae*, strain TWAR, in Scandinavia, 1981–1983. J Infect Dis 1989;159:1111–4.

[47] Bruu AL, Haukenes G, Aasen S, et al. *Chlamydia pneumoniae* infections in Norway 1981–87 earlier diagnosed as ornithosis. Scand J Infect Dis 1991;23:299–304.

[48] Fong IW, Chiu B, Viira E, Fong MW, Jang D, Mahony J. Rabbit model for *Chlamydia pneumoniae* infection. J Clin Microbiol 1997;35:48–52.

[49] Laitinen K, Laurila A, Pyha¨la¨ L, Leinonen M, Saikku P. *Chlamydia pneumoniae* infection induces inflammatory changes in the aorta of rabbits. Infect Immun 1997;65:4832–5.

[50] Mu¨hlestein JB, Anderson JL, Hammond EH, et al. Infection with *Chlamydia pneumoniae* accelerates the development of atherosclerosis and treatment with azithromycin prevents it in a rabbit model. Circulation 1998;97: 633–6.

[51] Campbell LA, O'Brien ER, Cappuccio AL, et al. Detection of *Chlamydia pneumoniae* in atherectomy tissue from patients with symptomatic coronary artery disease. In: Orfila J, Byrne GL, Cherneskey MA, et al., eds. Chlamydial infections. Bologna, Italy: Societa Editrice Esculapio, 1994: 212–5.

[52] Maass M, Gieffers J, Krause E, Engel PM, Bartels C, Solbach W. Poor correlation between microimmunofluorescence serology and polymerase chain reaction for detection of vascular *Chlamydia pneumoniae* infection in coronary artery disease patients. Med Microbiol Immunol 1998;187: 103–6.

[53] Moazed TC, Campbell LA, Rosenfeld ME, Grayston JT, Kuo CC. *Chlamydia pneumoniae* infection accelerates the progression of atherosclerosis in apolipoprotein (Apo E)-deficient mice. J Infect Dis 1999;180:238–41.

[54] Hu H, Pierce GN, Zhong G. The atherogenic effects of chlamydia are dependent on serum cholesterol and specific to C. pneumoniae. J Clin Invest 1999;103:747–53. 55. Fong IW. The value of animal models for C. pneumoniae related atherosclerosis. Am Heart J 1999;138:S512–3.

[55] Anderson JL, Mu¨hlestein JB, Carlquist J, et al. Randomized secondary prevention trial of azithromycin in patients with coronary artery disease and serologic evidence for *Chlamydia pneumoniae* infection. Circulation 1999;99:1540–7.

[56] Jackson LA, Campbell LA, Schmidt RA, et al. Specificity of detection of C. pneumoniae in cardiovascular atheroma. Evaluation of the innocent bystander hypothesis. Am J Pathol 1997;150:1785–90.

[57] Bachmaier K, Neu N, de la Maza LM, Pal S, Hessel A, Penninger JM. Chlamydia infections and heart disease linked through antigenic mimicry. Science 1999;283:1335–9.

[58] Kuo CC, Grayston JT, Campbell LA, Goo YA, Wissler RW, Benditt EP. *Chlamydia pneumoniae* (TWAR) in coronary arteries of young adults (15 to 35 years old). Proc Natl Acad Sci USA 1995;92:6911–4

[59] Jackson LA, Campbell LA, Schmidt RA, et al. Specificity of detection of C. pneumoniae in cardiovascular atheroma. Evaluation of the innocent bystander hypothesis. Am J Pathol 1997;150:1785–90.

[60] Laurila A, Bloigu A, Na¨yha¨ S, Hassi J, Leinonen M, Saikku P. Chronic *Chlamydia pneumoniae* infection is associated with a serum lipid profile known to be a risk factor for atherosclerosis. Arterioscler Thromb Vasc Biol 1997;17:2910–3.

[61] Kalayoglu MV, Byrne GI. Induction of macrophage foam cell formation by *Chlamydia pneumoniae*. J Infect Dis 1998;177:725–9

[62] Bachmaier K, Neu N, de la Maza LM, Pal S, Hessel A, Penninger JM. Chlamydia infections and heart disease linked through antigenic mimicry. Science 1999;283:1335–9

[63] Campbell LA, O'Brien ER, Cappuccio AL, et al. Detection of *Chlamydia pneumoniae* in atherectomy tissue from patients with symptomatic coronary artery disease. In: Orfila J, Byrne GL, Cherneskey MA, et al., eds. Chlamydial infections. Bologna, Italy: Societa Editrice Esculapio, 1994: 212–5.

[64] Jones RB, Priest JB, Kuo CC. Subacute chlamydial endocarditis. JAMA 1982;247:655–8.

[65] Marrie TJ, Harczy M, Mann OE, et al. Culture-negative endocarditis probably due to *Chlamydia pneumoniae*. J Infect Dis 1990;161:127–9.

[66] Odeh M, Oliven A. Chlamydial infections of the heart. Eur J Clin Microbiol Infect Dis 1992;11:885–93.

[67] Paróczai, D., Kókai, D., Ulbert, Á., Terhes, G., & Burián, K. (2020). Prevalence of *Chlamydia pneumoniae* antibodies in Southern Hungary.

[68] Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., ... & Mering, C. V. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic acids research, 47(D1), D607-D613.

[69] Grayston, J T., Kuo, C. C., Campbell, L. A. & Wang, S. P. *Chlamydia pneumoniae* sp-nov for Chlamydia sp strain TWAR. International Journal of Systematic Bacteriology (1989) 39:88 - 90.

[70] Stratton CW, Wheldon DB. Multiple sclerosis: an infectious syndrome involving Chlamydophila pneumoniae.. Trends Microbiol. (2006) 14(11):474-9.

[71] Stratton CW, Sriram S. Association of *Chlamydia pneumoniae* with central nervous system disease.. Microbes Infect. (2003) 5(13):1249-1253.

[72] Sriram S, Yao S, Mitchell WM, Calibresi P, Stratton CW, Ikejima H, Yamamoto Y. Comparative study of the presence of *Chlamydia pneumoniae* in cerebrospinal fluid from clinically definite and monosympomatic muliple sclerosis patinets. Clinical and Diagnostic Laboratory Immunology)2003) 9:1332-1337.

[73] Shirai, M., Hirakawa, H., Kimoto, M., Tabuchi, M., Kishi, F., Ouchi, K., Shiba, T., Ishii, K., Hattori, M., Kuhara, S., and Nakazawa, T. "Comparison of whole genome sequences of *Chlamydia pneumoniae* J138 from Japan and CWL029 from USA." Nucleic Acids Res.(2000) 28:2311-2314.

[74] Everett, K.D., Bush, R.M., Andersen, A.A. "Emended description of the order Chlamydiales, proposal of Parachlamydiaceae fam. nov. and Simkaniaceae fam. nov., each containing one monotypic genus, revised taxonomy of the

family Chlamydiaceae, including a new genus and five new species, and standards for the identification of organisms." Int. J. Syst. Bacteriol. (1999) 49:415-440..

[75] Cunningham, A. F., Johnston, S. L., Julious, S. A., Lampe, F. C. Ward, M. E. Chronic *Chlamydia pneumoniae* infection and asthma exacerbations in children. European Respiratory Journal. (1998) 11:345 - 349.

[76] Fukushi, H., and Hirai, K. "Restriction fragment length polymorphisms of rRNA as genetic markers to differentiate Chlamydia spp." Int. J. Syst. Bacteriol. (1993) 43:613-617.

[77] Gaydos, C.A., Palmer, L., Quinn, T.C., Falkow, S., and Eiden, J.J. "Phylogenetic relationship of *Chlamydia pneumoniae* to Chlamydia psittaci and Chlamydia trachomatis as determined by analysis of 16S ribosomal DNA sequences." Int. J. Syst. Bacteriol. (1993) 43:610-612.

An Appendix

Table 5.1: List of Plant Derive Natural Compounds used against targets in Molecular Docking

Sr No	Compounds Names	Definition	Mol. Weight	Mol. Formula	Type
1	5 (S)-Hpete	5(S)-HPETE(1-) is a conjugate base of 5(S) -HPETE that results from the carboxylic acid function being deprotonated. It functions as a metabolite in humans. It's a 5(S)-HPETE's conjugate base.	335.5	C20H31 O4-	Inhibitor
2	Beraprost Sodium	Beraprost is a synthetic derivative of prostacyclin that is now being tested in clinical trials to treat pulmonary hypertension. It's also being researched into as a way to avoid reperfusion harm.	420.5	C24H29 NaO5	Inhibitor

3	Epoprostenol Sodium	Epoprostenol Sodium is the sodium salt variant of epoprostenol, a synthetic prostacyclin and member of the prostaglandin family with anticoagulant and vasodilator properties. Epoprostenol sodium causes vasodilatation by directly simulating prostaglandin receptors in arterial vascular smooth muscle. This drug also prevents thrombus formation by inhibiting platelet aggregation by antagonizing platelet glycoprotein receptors.	374.4	C20H31 NaO5	Inhibitor

4	Lloprost Trometamol	Lloprost is a drug that is used to treat pulmonary arterial hypertension (PAH; high blood pressure in the vessels carrying blood to the lungs, causing shortness of breath, dizziness, and tiredness). In patients with PAH, lloprost may help them exercise more and reduce the progression of their symptoms.	121.14	C4H11 NO3	Inhibitor
5	Leukotriene A4	Leukotriene A4 is a Leukotriene that is the (5S,6S)-epoxy derivative of tetraenoic acid. It functions as a metabolite in mice. It's a Leukotriene, oxylipin, epoxy fatty acid, polyunsaturated fatty acid, and long chain fatty acid all combined into one.	318.4	C20H30 O3	Inhibitor

6	Leukotriene B4	In monocytes, lymphocytes, neutrophils, reticulocytes, platelets, and fibroblasts, Leukotriene B4 is produced from Leukotriene A4 by Leukotriene A-4 hydrolase. By boosting neutrophils activation and stimulating leukocyte adhesion, activation, and transendothelial migration, Leukotriene B4 reduces inflammation.	336.5	C20H32O4	I Inhibitor
7	Montelukast	Montelukast is a monocarboxylic acid and an aliphatic sulphide that belongs to the quinoline family. It works as a Leukotriene antagonist, antiasthmatic, and antiarrhythmia medication. It's a Montelukast conjugate acid (1-).	586.2	C35H36ClNO3S	Inhibitor

8	Montelukast Sodium	Montelukast sodium is a natural salt. Montelukast is an orally accessible Leukotriene receptor antagonist is often used for asthma prophylaxis and chronic therapy. It has been connected to a small number of cases of clinically significant liver impairment.	608.2	C35H35ClNNaO3S	Inhibitor
9	Ozagrel Hydrochloride Hydrate	Ozagrel (INN) is an antiplatelet drug that works by inhibiting thromboxane A2 production.	282.72	C13H15ClN2O3	Inhibitor
10	Ozagrel Sodium	Ozagrel sodium (OKY-046 sodium) is an inhibitor of the thromboxane A2 (TXA2) synthesis. Ozagrel sodium is an antiplatelet drug with an IC50 of 53.12 M that selectively inhibits human platelet aggregation.	250.23	C13H11N2NaO2	Inhibitor
11	Ozagrel	Ozagrel is a member of cinnamic acids.	228.25	C13H12N2O2	Inhibitor

12	Pranlukast Hydrate	Pranlukast is an antagonist for the cysteinyl Leukotriene receptor-1. It inhibits or lowers bronchospasm, which is produced mostly in asthmatics by an allergic reaction to allergens that are accidently or inadvertently met.	481.5	C27H23 N5O4	Inhibitor

13	Prostacyclin	Epoprostenol is an antihypertensive and platelet inhibitory oral prostacyclin and metabolite of arachidonic acid. Epoprostenol binds to prostacyclin receptors on platelet surfaces, activating platelet membrane adenyl cyclase and increasing cAMP levels in the process. The increased cAMP causes signal transduction, which causes vasodilatations. This substance also acts as a thromboxane A2 antagonist, causing immediate vasodilatations of the pulmonary and systemic arterial vascular beds as well as suppression of platelet aggregation.	352.5	C20H32O5	Inhibitor

14	Prostaglandin G2	The prostaglandin G2 is a kind of prostaglandin. It functions as both a mouse and a human metabolite. It's a prostaglandin G2 conjugate acid.	368.5	C20H32O6	Inhibitor
15	Prostaglandin H2	Prostaglandin H2 is a prostaglandin H that is an oxylipin, an olefin molecule, a secondary alcohol, and an oxylipin. It functions as a metabolite in mice. It's a prostaglandin H2 conjugate acid.	352.5	C20H32O5	Inhibitor
16	Ramatroban	Ramatroban is a molecular organic compound. Ramatroban has been used in clinical trials to treat asthma.	416.5	C21H21FN2O4S	Inhibitor
17	Seratrodast	Seratrodast is a molecular organic compound. Seratrodast (INN) is a thromboxane receptor antagonist that is commonly used to treat asthma.	354.4	C22H26O4	Inhibitor

18	Thromboxane	Thromboxane is significant mediators of the activities of cyclooxygenase-transformed polyunsaturated fatty acids.	296.5	C20H40O	Inhibitor
19	Zafirlukast	Zafirlukast is an orally active Leukotriene receptor antagonist that is commonly used for asthma prevention and therapy. Zafirlukast has been connected to a small number of severe episodes of acute liver damage in the past. Zafirlukast is a carbonate ester and an N-sulfonylcarboxamide belongs to the indole family. It works as a Leukotriene antagonist and an antiasthmatic agent.	575.7	C31H33N3O6S	Inhibitor