In silico Integrative Approach to Identify Druggable Candidates in *Klebsiella pneumoniae*: A Multi-drug Resistant Pathogen

By

Noor-Ul-Ain Sajid Mughal

National Center for Bioinformatics

Faculty of Biological Sciences

Quaid-i-Azam University Islamabad, Pakistan

2016

In silico Integrative Approach to Identify Druggable Candidates in *Klebsiella pneumoniae*: A Multi-drug Resistant Pathogen

A Thesis submitted in the partial fulfillment of the requirements for the degree of

Master of Philosophy in Bioinformatics

By

Noor-Ul-Ain Sajid Mughal

Supervisor

Dr. Syed Sikander Azam

National Center for Bioinformatics

Faculty of Biological Sciences

Quaid-i-Azam University Islamabad, Pakistan

2016

CERTIFICATE

This thesis submitted by **Miss Noor ul Ain Sajid Mughal** from National Centre for Bioinformatics, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad, Pakistan. is accepted in its present form as satisfying the thesis requirement for the Degree of Master of Philosophy in Bioinformatics.

Internal Examiner:_____
Dr. Syed Sikander Azam
Assistant Professor & Supervisor
Quaid-i-Azam University Islamabad.

External Examiner:_____
Dr. Muhammad Saeed
Assistant Professor
Department of Bio-Sciences,
COMSATS Institute of Information Technology
Islamabad.

Chairman:_____
Dr. Sajid Rashid
Associate Professor
National Centre for Bioinformatics
Quaid-i-Azam University Islamabad

Date:March 10,2016

DECLARATION

I hereby solemnly declare that the work ***"In silico* Integrative Approach to Identify Druggable Candidates in *Klebsiella pneumoniae*: A Multi-drug Resistant Pathogen"** presented in the following thesis is my own effort, except where otherwise acknowledged and that the thesis is my own composition. No part of the thesis has been previously presented for any other degree.

Dated: _____

_____

Noor-Ul-Ain Sajid Mughal

Dedication

To my Mom and Dad

Mr. and Mrs. Sajid Mughal

You always picked me up on time and encouraged me to go on every adventure, especially this one.

AND

To my Siblings

Arman (1981-2006),

Soban, Neel Kanwal and Usman

You have been my inspiration, and my soulmates.

Acknowledgments

In the name of ALLAH (SWT), The Most Bounteous and The Most Merciful, all praises to ALLAH (SWT) for strength and His blessing in the accomplishment of this thesis. No words can express enough gratitude for His knowledge and His wisdom.

I would like to express my gratitude to Dr. Syed Sikander Azam for being an outstanding advisor and excellent supervisor. His constant encouragement, support, and invaluable suggestions made this work successful. I am indebted and thankful to him.

I would like to acknowledge Chairman Dr. Sajid Rashid for the academic support.

I am thankful to Prof. Dr. Klaus R. Liedl, for lending us computational help for the extension of simulations, I am utterly obliged for his support. My sincere thanks to our Ph.D student Saad Raza for the Perl scripts, his help was really valuable for this work. I am indebted to Computational biology group members Asma Abro, Samra Wajid Abbasi, Amen Shamim, Seemab Khursheed, Hira Jabeen, Sanum Javed, Nimra Sabir, Iqra Ahmad, Sundus Iqbal, Fouzia Shaheen, Gul Sanober and Farhan Ul Haq for providing a stimulating and constructive environment in lab.

I would like to pay my highest regards to my parents, my family members Aminah, Haalah, Haadiyah, Muhammad, Bilal, Nazif and Nayel, along with my friends Amina, Maria, Iqra, Sundus, Abida and Shabana for their sincere encouragement and unconditional love throughout my research work. I owe everything to them.

Noor-Ul-Ain Sajid Mughal

Table of Contents

Table of Contents

List of Abbreviations

3-Dimensional	3D
Absorption, Distribution, Metabolism, Excretion and Toxicity	ADMET
Assisted Model Building with Energy Refinement	AMBER
Basic Local Alignment Search Tool	BLAST
Bovine Pancreatic Trypsin Inhibitor	BPTI
BRaunschweig ENzyme Database	BRENDA
Broyden-Fletcher-Goldfarb-Shanno	BFGS
Computer Aided Drug Design	CADD
Cluster Database at High Identity with Tolerance	CDHIT
Database of Essential Genes	DEG
Discrete Optimized Protein Energy	DOPE
Discover Studio Visualizer	DS Visualizer
Enzyme Commission	EC
Expert Protein Analysis System	ExPASy
Expectation Value	E-value
General Amber Force Field	GAFF
Genetic Optimization for Ligand Docking	GOLD
Iterative-TASSER	I-TASSER

List of Abbreviations

KEGG Automatic Annotation Server	KASS
Klebsiella pneumoniae	*K. pneumoniae*
Kyoto Encyclopedia of Genes and Genomes	KEGG
Molecular Dynamics	MD
Multi Drug Resistant	MDR
Nanosecond	ns
National Center for Biotechnology Information	NCBI
Process TRAJectory	PTRAJ
Protein Data Bank	PDB
Quantitative Structure Activity Relationships	QSAR
Radius of Gyration	R_g
Root Mean Square Deviation	RMSD
Root Mean Square Fluctuation	RMSF
Simulated Annealing with NMR Derived Energy Restraints	SANDER
Support Vector Machines	SVMs
Three-Point Transferable Intermolecular Potential	TIP3P
Universal Protein Resource Knowledge Base	UniProtKB
Visual Molecular Dynamics	VMD

List of Figures

List of Tables

Abstract

Untimely discontinuation and misuse of therapeutic agents paired with emergence of new strains has imparted resistivity to pathogens. This demands development of more directed antibacterial agents. *Klebsiella pneumoniae,* a Gram negative, rod shaped bacterium is a life threatening multi-drug resistant (MDR) pathogen primarily involved in pneumonia. Multi-drug resistant behaviors of this emerging pathogen, initiate the research towards the identification of novel drugs. Starting point of this study was to identify and characterize the potential druggable targets in multiple strains of *K. pneumoniae* via a hierarchical *in silico* genome subtraction. Druggable targets were characterized qualitatively through a number of filters leading to the identification of a promising drug target candidate, citrate lyase subunit beta (citE). Unavailability of experimentally determined structure of citE, directed utilization of homology modeling technique to predict the structure. Best modeled structure was selected and used for molecular docking studies. Molecular docking protocol classified compound 47, as the ligand best able to fit the binding pocket. Molecular dynamics (MD) simulations are playing a promising role in the development of therapeutic drugs. Thus, molecular dynamics was steered to simulate citE; a seventy nanosecond MD simulation illustrated a steady binding pattern of ligand in the protein's active site. Pursuing this, trajectory analysis was implemented to assess various characteristics of the docked system in terms of function of time. Dynamical flexibility of C-terminal domain was observed. It is established from the current work that more relaxed and improved structure of citE was produced which can be scrutinized for further analysis. Key findings of the study can be employed to guide the design of targeted citE inhibitory drugs.

INTRODUCTION

1. Introduction

Bacterial infections are the leading contributors to global mortality, affecting millions of people; with the issue worsened by the emerging antibiotic resistant bacterial strains (WHO, 2014). ESKAPE, a class of bacteria, comprising of *Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa, and Enterobacter* spp. are predominantly considered as the troublemakers in the current century due to their multidrug resistance abilities (Pendleton, et al., 2013; Xu, et al., 2014). Focus on these pathogenic species is due to their contribution to the health care associated infections leading to thousands of deaths (WHO, 2014).

Centers for Disease Control and Prevention (CDC) lately took the initiative to classify major microbial organisms that pose a threat to our health care system. The MDR *Klebsiella pneumoniae* is listed as one of the top microorganism with a serious threat level, requiring instantaneous attention in order to eradicate the associated health problems (CDC, 2013).

1.1. *Klebsiella pneumoniae*

Klebsiella pneumoniae (*K. pneumoniae*), a member of class gammaproteobacterium (Berman, 2012), is the most prevalent and clinically significant, human pathogen amidst rod shaped members of the genus *Klebsiella,* belonging to the family *Enterobacteriaceae. K. pneumoniae* are non-motile, rod-shaped, facultative anaerobic, Gram-negative bacteria (Janda and Abbott, 2008). A distinctive characteristic of *K. pneumoniae* is a thick polysaccharide coat, which aids its invasion of the host defenses. Inter and intra-species' transmission of plasmids and insertion elements, is facilitated in close members of the family, that leads to the horizontal exchange of antibiotic resistance genes.

Although found in the normal flora of the mouth, skin, and intestines, it can cause destructive changes to human lungs if inhaled specifically to the alveoli, resulting in bloody mucus. The most prevalent diseases associated with *K. pneumoniae* comprise of pneumonia (and other respiratory diseases), urinary tract infections (UTIs) and blood stream infections

in neonates, elderly and immune-compromised patients (Liu, 2011). In recent years, *K. pneumoniae* has been labelled as one of the most troublesome nosocomial infective agent (Nordmann, Cuzon and Naas, 2009).

1.1.1. *K. pneumoniae* Associated Mortality

Multidrug resistant (MDR) *K. pneumoniae*, associated infections exhibit a globally erratic pattern of prevalence and disease associated mortalities (Ko, et al., 2002). The mortality rates of up to 50% have been linked with *K. pneumoniae* inflicted hospital acquired infections (HAIs), particularly pneumonia, bacteraemia, Spontaneous Bacterial Peritonitis (SBP) and neonate septicaemia (WHO, 2014).

K. pneumoniae linked infections are not restricted to some specific strain but with the passage of time, new strains have been identified in different countries, with an increased mortality rate especially in Eastern Europe and Latin America (Paterson, et al., 2004). The World Health Organization approximates that one in three newborn infant deaths are due to pneumonia. Over two million children under five die each year worldwide. Elderly individuals, too, are at a particular risk for pneumonia and associated mortality (WHO, 2014). In the United States, community-acquired pneumonia affects 5.6 million people per year, and ranks sixth among leading causes of death. Furthermore, individuals with underlying chronic illnesses, such as Alzheimer's disease, cystic fibrosis, emphysema, and immune system problems as well as tobacco smokers, alcoholics, and individuals who are hospitalized for any reason, are at significantly increased risk of pneumonia (Almirall, et al., 1999). Spontaneous bacterial peritonitis (Hassan, El-Rehim and EL-Din, 2015), too has been reported with a high mortality rate of thirty percent (30%), standing second to pneumonia (Nobre, et al., 2008).

1.1.2. Multiple-Drug Resistance

Multiple drug resistance property present in pathogens is due to the presence of efflux pump genes and proteins, which are located on plasmids or chromosomes (Piddock, 2006). The issue of increasing prevalence of *K. pneumoniae* infections has been complemented by the rapid transmission of resistance against all the major classes of antibiotics among

different strains (Mantzarlis, et al., 2013). The antibiotics with high intrinsic activity, majorly the third-generation cephalosporins (e.g. cefotaxime, ceftriaxone), carbapenems (e.g. imipenem/cilastatin), aminoglycosides (e.g. gentamicin, amikacin), and quinolones are used as a treatment for infections caused by *K. pneumoniae* (Tsai, et al., 2013). More troublesome is the acquirement of carbapenemases, belonging to a series of Extended Spectrum β-lactamases (ESBLs), capable of hydrolysing essentially all β-lactums including the last resort antibiotics, carbapenems (Mosca, et al., 2013). ESBL phenotypes, particularly carbapenemase KPC2 producing strains have been associated with large scale endemics in health-care facilities world-wide (Shafiq, et al., 2013). Recent researches in Pakistan have revealed persistence of MDR *K. pneumoniae* infections in health care facilities (Saleem, et al., 2013). The combination of ESBLs with KPCs in *K. pneumoniae* strains, imparts alarming resilience against antibacterial drugs, leaving combinatorial therapy as the only choice of treatment in certain instances of severe infections. Researches have been instigated to evaluate the limitations of existing treatment options and to formulate novel therapeutic interventions against these resilient pathogens (Hirsch and Tam, 2010).

1.1.3. Genomic Features

Eight thirty nine (839) strains of the pathogen, *K. pneumoniae* have been reported till 2015, however only forty three (43) of these strains have been completely sequenced to date. The sequenced data is available on National Center for Biotechnology Information (NCBI) *(http://www.ncbi.nlm.nih.gov/genome/genomes/815?).* This study emphasizes on exploration of the genomes of the selected strains for identifying potential drug targets. The three strains selected are 30660NJST258_1, 30684NJST258_2, and JM45. The genome of *K. pneumoniae* 30660NJST258_1 comprises of a single chromosome, with six plasmid replicons, *K. pneumoniae* 30684NJST258_2 contains one chromosome accompanied by 4 plasmid replicons, while *K. pneumoniae* JM45 consists of one chromosome and two plasmids. Detailed genomic features of the selected strains are given in the Table 1.1.

Table 1.1. *Genomic features of 30660NJST258_1, 30684NJST258_2 and JM45.*

Strain Name	Size (Mb)	GC%	Genes	Protein
30660NJST258_1	5.54094	57.15	5587	5598
30684NJST258_2	5. 41722	57.43	5468	5478
JM45	5.60317	57.24	5659	5673

1.2. Applied *In silico* Approach

Current study incorporates sequential integration of various *in silico* approaches. The modular application of these methods follows the following route: genome subtraction, drug target selection, comparative homology modeling, molecular docking and molecular dynamic simulations.

1.2.1. Genome Subtraction

The conventional therapeutic discovery is a lengthy process, with its pace being incomparable for the issues raised by rapidly growing drug resistance issues. To complement the experimental procedures, computer aided drug design (CADD) plays an important role. Genome subtraction is a fine example of an integrative approach which provides exploration of the biological space, by identifying novel drug targets. This approach is speedy, powerful, and cost effective and has been successfully used to identify druggable targets in pathogens including: *Burkholderia pseudomallei* (Chong, et al., 2006), *Aeromonas hydrophila* (Sharma, Gupta and Dixit, 2008), *Clostridium botulinum* (Prajapati and Bhagat, 2012), *Neisseria meningitides* (Sarangi, et al., 2009), *Mycobacterium tuberculosis* (Hosen, et al., 2014), *Staphylococcus aureus* (Uddin and Saeed, 2014), *Streptococcus gordonii* (Azam and Shamim, 2014), *Edwardsiella tarda* (Neema, Karunasagar and Karunasagar, 2011), *Leptospira interrogans* (Pradhan, et al., 2013), *Streptococcus pneumoniae* (Munikumar, et al., 2013), *Clostridium botulinum* (Reddy and Rao, 2012) *Mycoplasma genitalium*

(Butt, et al., 2012), *Salmonella typhi* (Rathi, Sarangi and Trivedi, 2009) and *Mycobacterium leprae* (Shanmugam and Natarajan, 2010). The current investigation employs the concept of genome subtraction to identify novel biological drug targets within the genome of *K. pneumoniae*.

The procedure begins with the elimination of paralogous proteins of the microorganism, followed by a comparative analysis of the remaining proteome of the pathogen and host to discover non homologous macromolecular entities. Subsequently, proteins that are essential to pathogenic survival are identified and their druggability potential is determined. Cytoplasmic proteins are preferred on other cellular localization areas, because the mode of action of the drugs in the cytoplasm is aided in an enhanced manner, than the membrane. Thus, cytoplasmic proteins are more potent candidates for drug development (Parvege, Rahman, and Hossain, 2014). The resulting set of proteins is essential and unique to the pathogen, ensuring prevention of cross reactivity when a drug is targeted against the pathogenic protein.

1.2.2. Drug Target Selection

Target selection, usually implies finding a therapeutically significant agent (Knowles and Gromo, 2003). Besides, a proper target identification suggests the relationship between the drug and the disease which can further be investigated for possible side-effects (Hughes, et al., 2011). Druggability is an important feature that should be focused on, while probing for a potential target. The targets that pose both, structural and functional features of druggability are the most favorable in the course of study (Russ and Lampel, 2005). In the current study, the druggable target is citrate lyase subunit beta (EC: 4.1.3.34), which is a metalloprotein and the gene name is citE.

1.2.3. Comparative Homology Modeling

Proteins mainly control the important functions and behavior of living cells. However, function of a protein molecule is completely determined by its structure (Genheden, 2012). Three dimensional structural data of large molecules can easily obtained from Protein Data Bank (RCSB PDB), which is the central archive of experimentally obtained biomolecular

structures. Protein Databank (PDB) is expanding and improving each year (Berman, et al., 2000). However, the number of structurally characterized proteins is still less compared to the available known protein sequences (Schwede, et al., 2003). Therefore, the lack of experimental structures in databases can be accomplished by using various theoretical techniques, including homology or comparative protein modeling. Homology modelling is based on the simple, yet powerful concept that similar sequences have similar structures. Accordingly, the three dimensional space can be explored to generate a theoretical model for the target protein if a known template structure with a suitable level of similarity to the target is available (Krieger, Nabuurs and Vriend, 2003).

Owing to lack of crystallographic structure for citE from *K. pneumoniae*, the selected protein target was assessed for its suitability to be modeled through comparative modeling approach. To enhance objectivity of the approach, several webservers, namely, SWISS-MODEL (Schwede, et al., 2003), ModWeb (Pieper, et al., 2004) and I-TASSER (Zhang, 2008) were used in addition to the modeling program, MODELLER (Eswar, et al., 2008), version 9.10. Constructed homology models from the aforementioned sources were further subjected to validation check to evaluate the quality of the three dimensional protein structures, certifying that the structure is in a thermodynamically stable state (Garza-Fabre, Toscano-Pulido and Rodriguez-Tello, 2012). To this end, PROCHECK (Laskowski, Moss and Thornton, 1993), ProSA (Sippl, 1993; Wiederstein and Sippl, 2007), Errat value (Colovos and Yeates, 1993), G-Factor and Bad Contacts (Morris, et al., 1992) were measured. The structural quality assessment is a fundamental prerequisite for accurate application of drug design procedure.

1.2.4. Molecular Docking

Improved computational power and the growing ease of data availability in protein databases have led to the emergence of the field called molecular docking. It plays an essential role in structural molecular biology and structure-based drug designing by computationally simulating the molecular recognition process. It is the main tool frequently used to predict the binding modes in drug-receptor interactions by achieving an optimized conformation for receptor and ligand (Morris and Lim-Wilby, 2008). It is routinely used in modern drug

discovery research for hit identification, lead optimization and bioremediation. Drug discovery demands altering or preventing the biochemical reaction catalyzed via target molecule by identifying a potential inhibitor against a given protein which binds more strongly than its natural substrate (Thomsen and Christensen, 2006). It helps in the antimicrobial drug research, by predicting the preferred binding orientation of small molecule drug candidates against their macromolecular target.

Molecular docking studies explore the possible binding modes of a substrate to a given receptor but lack receptor flexibility. On the contrary molecular dynamics simulations treat protein-ligand as flexible and explore the best conformation of receptor ligand complexes. MD simulations are employed in the current study to further investigate the exact binding conformations and time dependent behavior of protein-ligand complex.

1.2.5. Molecular Dynamics Simulation

Molecular dynamics (MD) simulation has been applied successfully to study the behavior of protein and its structural dynamics that gives even minor and specific details on the time scale. Advancement in this brilliant technique enables to resolve biological problems. The work on MD simulation was first started by Alder and Wainwright in 1959. First studied protein via MD simulation was bovine pancreatic trypsin inhibitor (BPTI). Over the past years, MD simulation has become a rapidly established field and a standard tool to study biomolecules (McCammon, Gelin and Karplus, 1977).

The collaborations of MD simulation and docking studies can give realistic information on structural level. Binding modes, stability, and ligand conformations; in all aspects, MD simulation is playing its vital role. MD simulation provides insight of dynamic properties like structural compactness and conformational changes, of the system (Azam, Uddin and Wadood, 2012). Therefore, in the context of drug designing, MD simulations can be considered as a very interesting application (Hansson, Oostenbrink and van Gunsteren, 2002). The progression in computational powers has now allowed, simulating a protein system for an even longer time-scale. MD simulation is based on a simple concept of mechanics which involves Newton's equations of motion (Binder, et al., 2004). Molecular dynamics simulations are remarkably used to study macroscopic thermodynamics. MD simulations also

assist to analyze and understand experimental data. Improvements in therapeutics can be attained by combining MD simulations, docking and experimental data (Alonso, Bliznyuk and Gready, 2006). Biological processes take place at different time scales on the basis of their complexity and dynamic nature. The dynamic behavior of biological molecules can be classified over the varying time scales into three parts. Local motions that describe the side chain movements, atomic fluctuations and loop motions over the time range of 10^{-15} to 10^{-1} second with the distance of 0.01 to 5 Å. Rigid body motions that cover the helix, domain and subunit movements over the distance of 1 to 10 Å and time range of 10^{-9} to 1 second. And large scale motions that require longer time scale usually 10^{-7} to 10^{4} second to complete. They deal with very small movements such as folding or unfolding of proteins, helix coil transitions, winding and unwinding of DNA or association and dissociation of a receptor and its ligand (Allen and Tildesley, 1989).

1.2.5.1. Statistical Mechanics

MD simulation deals with the information at microscopic level which includes atomic positions and relative velocities. To extend its applicability to biological systems, microscopic information is transformed into macroscopic entities such as energy, pressure and temperature using statistical mechanics. Macroscopic properties like fluctuations in binding energy or conformational changes involved in protein-ligand complex are explored using mathematical expressions of statistical mechanics. The objective is to investigate macroscopic phenomena from the properties of individual molecules in a system. Statistical mechanics prompts a sequence of points in multidimensional space as a function of time. All these points generated belong to the same states or microstates which are collectively referred as an ensemble. Ensemble of any system is the representative of different conformations of the system, having different microscopic states but same thermodynamic macroscopic states (Wilde and Singh, 1998). In statistical mechanics, four major ensembles exist with varying characteristics (Wereszczynski and McCammon, 2012).

1. Canonical Ensemble (NVT) is characterized by constant number of particles N, constant volume V and constant temperature T, in thermodynamic studies.

2. Micro canonical Ensemble (NVE) has a constant number of particles N, constant volume V and constant energy E.

3. Isobaric-Isothermal Ensemble (NPT) deals with constant particle number N, constant pressure P and constant temperature T.

4. Grand Canonical Ensemble (μVT) has a constant chemical potential μ, a constant volume V and a constant temperature T.

1.2.5.2. Classical Mechanics

The treatment of microscopic systems through laws of classical mechanics gives rise to molecular dynamics (Tuckerman, 2010). Classical mechanics is centered on Newton's second law and therefore describes the dynamic motion of biological bodies over a given time period. It states that when a force *"F"* is applied on a particle *"i"*, having a mass *"m"*, it will accelerate by a quantity "*a*" (McCall, 2010). Newtonian mechanics is applied at the molecular level by extrapolating this concept, to deduce the acceleration produced by "*N*" particles in a system, given that force applied on them is known. A specific particle "*i*" therefore experiences a cumulative force because of all the other particles that interact with it. This force is dependent on the positioning of interacting constituents and velocities responsible for their acceleration through the system over a certain time scale (Tuckerman, 2010). Mathematical equations, describing these concepts, allowing calculation of macro-canonical observables, are presented below.

The equation of motion for "i^{th}" particle:

$$F_i = m_i a_i \qquad (1.1)$$

Where F_i is force applied on the particle, "m_i" is the mass of the particle "*i*" and "a_i" is the acceleration produced. The acceleration is the second derivative of distance "*r*" and time "*t*", thus, substitution of these quantities gives the following equation:

$$F_i = m_i \frac{d^2 r_i}{dt^2} \qquad (1.2)$$

When a force "F_i" is applied on a particle of mass "m_i" it can be described in terms of change of potential energy "v" to give the following:

$$F_i = -\frac{dv}{dr_i} \qquad (1.3)$$

So, "F_i" in equation 1.2 can be replaced by equation 1.3, and then they can be equated to give:

$$-\frac{dv}{dr_i} = m_i \frac{d^2 r_i}{dt^2} \qquad (1.4)$$

During production run, the next coordinates for every particle in a system at any given time can be calculated using previous coordinates "r_i", velocity "v" and acceleration "a" at a given time "t".

$$\mathrm{a} = \frac{dv}{dt} \qquad (1.5)$$

Equation 1.5 denotes that for a particle in motion, its acceleration is dependent on the rate of change of velocity with respect to time.

1.2.5.3. Molecular Mechanics

Molecular mechanics utilizes classical mechanics to model molecular systems and determines the structure and physical properties of biological molecules. Molecular mechanics aims at calculating the energy associated with a given conformation of the molecule. Potential energy function is the sum of individual functions for angle bending, bond stretching, torsional energies and non-bonded interactions. Force field of a molecular system can be given by the sum of individual energy terms:

$$E = E_{covalent} + E_{non\text{-}covalent} \qquad (1.6)$$

$$E_{covalent} = E_{bond} + E_{angle} + E_{dihedral} \qquad (1.7)$$

$$E_{non\text{-}covalent} = E_{electrostatic} + E_{van\ der\ waals} \qquad (1.8)$$

Where the covalent interactions account the bond, angles and dihedrals while non-covalent interactions are the sum of electrostatic and van der Waals forces.

1.3. Aims and Objectives

Progressive expansions of computer aided drug design approaches have aided in removing multiple obstacles of the drug discovery process. Current research is driven by the necessity to address the incessantly increasing health-risk posed by life threatening nosocomial infections. Thus, this study proposes to scrutinize several strains of multiple drug resistant human pathogen *Klebsiella pnuemoniae,* and suggest a substitutive therapy to overcome the failure of preexisting antimicrobial treatments. With the shift from genetic to genomic approach, however, the conventional route to drug target mining, associated with strenuous procedure, is no longer desirable. *In silico* subtractive genomics approach was thus applied to explore the complete genome of three *K. pneumoniae* strains (30660NJST258_1, 30684NJST258_2, and JM45), which resulted in a common novel drug target. The resulting target represented a druggable protein, with the potential to be modulated by small drug-like inhibitors. Comparative analysis was carried out among diverse range of inhibitors by using molecular docking methods to screen out, the best target binding inhibitor. Computational aspects of the present study are enhanced additionally by dynamic view of the docked system, provided by simulation studies carried out. Simulated view of the target protein in ligand-bound form has substantial role in illustrating the structural conformations and stability.

METHODOLOGY

2. Methodology

In context to the current work, the procedural details including several computational expertise used and parametric observations made during the various phases are outlined in the following sections.

2.1. System Specification

The entire scheme comprises of four major steps; subtractive genomics, homology modeling, molecular docking and molecular dynamics simulations. The whole research was performed on Intel (R) Core(TM) 2 Duo CPU E8600 @ 3.33 GHZ and operating system Linux openSUSE 11.4 was used. This computational facility was provided at Computational Biology Lab of National Centre for Bioinformatics, Quaid-i-Azam University Islamabad, Pakistan (Figure 2.1). Furthermore, for extension of the simulations, assistance was acquired from Center for Chemistry and Biomedicine (CCB), University of Innsbruck, Austria.

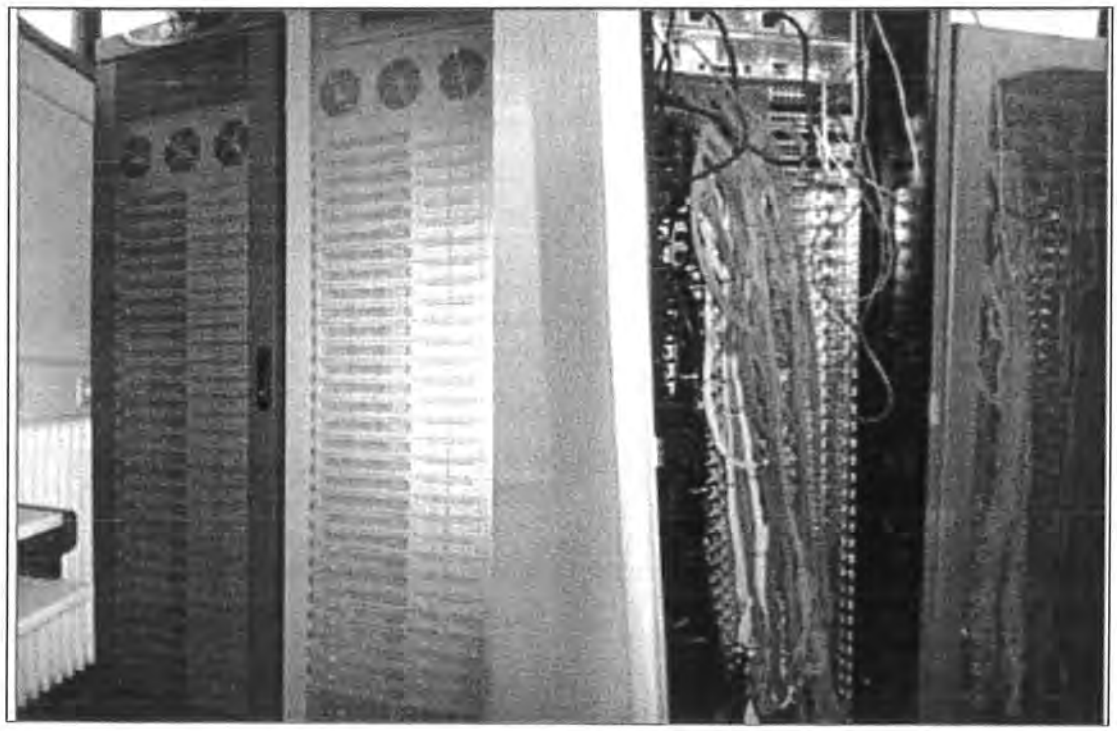

Figure 2. 1. *Cluster system used for computational studies carried out.*

2.2. *In silico* Approach Overview

The computational stream of work performed can be divided into different divisions. Each of these divisions implied a diverse set of tools and softwares, which were integrated together to insinuate into the pharmacologically essential, unique druggable target. This process started with the exploration of complete genome of *K. pneumoniae*. Three strains were selected and the rest was performed on all three selected strains. The therapeutic candidate that demonstrated strong functional significance was then subjected to comparative model building. The inhibition mechanisms were consequently comprehended through the use of molecular docking. MD simulations, provided a dynamic interpretation of time dependent variation in the behavior of the macromolecular systems. The schematic overview of the performed activities is provided in Figure 2.2.

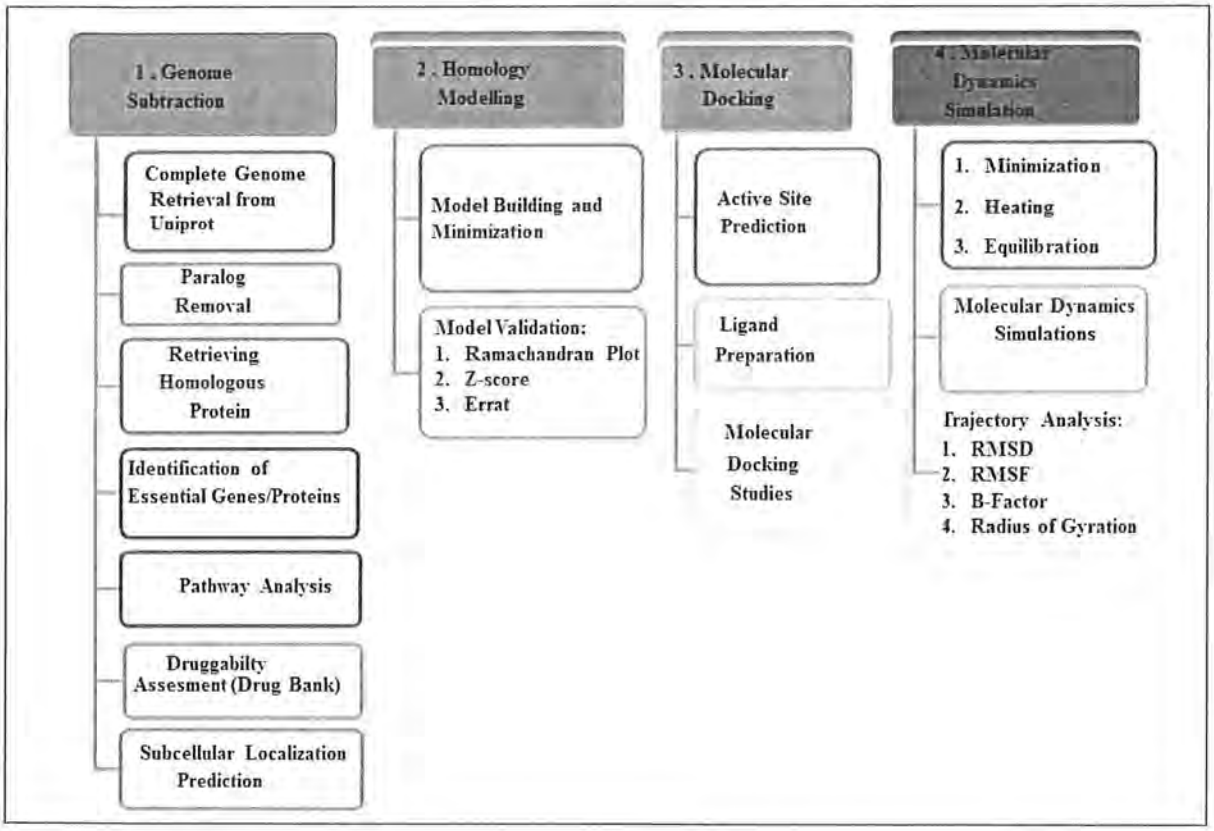

Figure 2. 2. *Schematic work flow, highlighting the major steps employed in study.*

2.2.1. Genome Subtraction

One of the most recently used strategies for drug target identification, in pathogens is subtractive genomics approach. The subtraction process began with the extraction of complete genome of the pathogen. UniProtKB (Boutet, et al., 2007) was used to retrieve the genome of three completely sequenced strains, of the pathogen under study. This database consists of both manually curated (UniProtKB/Swiss-Prot) and automatically annotated (UniProtKB/TrEMBL) protein sequences. After complete genome retrieval, a sequential series of analysis were performed that are discussed below.

2.2.1.1. Paralog Removal

Complete genomes of multiple strains of *K. pneumoniae* namely 30660NJST258_1, 30684NJST258_2, and JM45 were retrieved from Uniprot (*http://www.uniprot.org/*) in FASTA format. In order to eliminate the redundant paralogous proteins, the web server, CDHIT Suite (Cluster Database at High Identity with Tolerance) was used (Li, Jaroszewski and Godzik, 2001). It is a user friendly web interface that accepts genomes in the FASTA format and removes redundant data. The cut off value was set to 0.6 (i.e 60%).

2.2.1.2. Homolog Removal

Next two steps were enclosed using Perl script. The Perl script saves time in exploration and filtration of sequences in multiple strains. These scripts were provided at Computational Biology Lab of National Centre for Bioinformatics, Quaid-i-Azam University Islamabad, Pakistan. The comparative alignment of non-homologous *K. pneumoniae* proteins to the human proteome was carried out using the protein Basic Local Alignment Search Tool (BLASTp) service at National Center for Biotechnology Information (NCBI) (*http://blast.ncbi.nlm.nih.gov/Blast.cgi*). In the first step, whole proteome of the pathogen was searched against human proteome using Basic Local Alignment Search Tool (BLASTp) at threshold expectation value (E- value) of 10^{-4}. *K. pneumoniae* sequences for which no corresponding hits were found or which showed < 35% identity with human proteins, were selected as non-homologous, which were subsequently subjected to essentiality check.

2.2.1.3. Essential Proteins Identification

Non-human homolog sequences from the preceding step were used for essential proteins' identification in which Database of Essential Genes (DEG) (*http://tubic.tju.edu.cn/deg/*) (Zhang, Ou and Zhang, 2004) was employed. The number of essential genes in prokaryotes and eukaryotes have amplified about 10-fold and 5-fold, respectively in DEG (Zhang and Lin, 2009). BLASTp search was carried out against DEG with E-value cut off at 10^{-10}, bit score = 100 and sequence identity of >= 30%. All parameters were set to the Perl script and the resultant file consisted of the shortlisted essential proteins.

2.2.1.4. Metabolic Pathway Characterization

The allotment of metabolic pathways to essential proteins was the next step of the screening procedure. Essential non-homologous proteins were then functionally classified through the use of KEGG Automatic Annotation Server (KAAS) (*http://www.genome.jp/kegg/kaas/*), an analytical tool supported by the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (*http://www.genome.jp/kegg/*). KAAS server used *K. pneumoniae* genome as reference data set, to increase the accuracy of results. KAAS server assigns KEGG orthology (KO) IDs to the proteins using BLASTp, to search against the KEGG GENES repository (Kanehisa and Goto, 2000). It also allows access to analytical resources to transfer annotations, predict pathways and conduct similarity searches. On the basis of KO numbers, metabolic pathways are consigned to the set of sequences under study. (Moriya, et al., 2007). Since only a part of the complete pathogen proteome was being assessed, single-directional best hit method was employed. The predicted pathogenic pathways were manually matched to the human pathways to classify them into two categories: common and unique. Common pathways were those present in both the bacterium and the host. Whereas, unique, were exclusive to bacterium and hence, focus of current work.

2.2.1.5. Druggability Assessment using Drug Bank

Succeeding the characterization of pathogen specific proteins, druggability screen was applied as a filtering criterion. DrugBank is an online tool comprising details regarding the biochemical and pharmacological properties of a drug and alongside drug targets and their

mechanisms. Besides this, information about the quantitative structure activity relationships (QSAR) and absorption, distribution, metabolism, excretion and toxicity (ADMET) properties are also incorporated in DrugBank (Law, et al., 2014). To evaluate the druggability of potential target, DrugBank version 4.2, with default parameters, was used and only the proteins having a bit score > 100 or belonged to either experimental, investigational or approved category were selected.

2.2.1.6. Localization Prediction

The concluding step of the genome subtraction was prediction of subcellular location of the screened, unique and essential protein targets. PSORTb 3.0.2 (*http://www.psort.org/psortb/*) and subCELlular LOcalization predictor (CELLO) version 2.5 (*http://cello.life.nctu.edu.tw/*) were used to predict the location of the protein targets. The protein sequences were submitted in FASTA format in both tools, with organism type set to bacterial and gram stain set to negative. Both PSORTb (Nancy, et al., 2010) and CELLO (Yu, et al., 2006) implement Support Vector Machines (SVMs), a machine learning technique that mines the curated dataset to predict the localization through the use of suffix tree algorithm. Similar to PSORTb, CELLO adopts functionality of SVMs at two levels. The preliminary classification of proteins, conferring to subcellular location is performed on the basis of molecular descriptors derived from the protein sequence, followed by a final decision centered on the probability of the subcellular location.

2.2.2. Drug Target Selection

Subtractive genomics screening process resulted in shortlisting a few potential proteins from complete proteome of *K. pneumoniae*. Further on, these proteins were evaluated for their potency of becoming probable drug targets. Out of six essential druggable proteins shortlisted, one protein was selected for further computational analysis. UniprotKB was used to accumulate information regarding the functional role of the proteins, active isoforms and catalytic requirements for enzymatic activity like dependency on cofactors, subunit structure and associated post translational modifications. Additionally, structural details about the identified prospective targets were explored to establish availability of

experimental structures. In absence of such structures, the possibility of comparative homology based modeling was considered by assessing the template availability. The potential, druggable bacterial proteins, brought forth by the genomic screening were evaluated for feasibility to be carried forward towards the CADD process. The aforementioned aspects were considered cumulatively, which led to the selection of the most suitable drug target to undergo the CADD phase. Thus, the selected druggable target was citE.

2.2.3. Comparative Homology Modeling

A three dimensional crystallographic structure was unavailable for the selected target hence comparative modeling was carried out to elucidate a homology based structure. The peptide sequence for *K. pneumoniae* - citE (Uniprot ID: W8UQT9) was taken from UniprotKB and compared to PDB via alignment through BLASTp. The X-ray crystallographic structure for mcl from *Rhodobacter sphaeroides* (Zarzycki and Kerfeld, 2013) having a 33% sequence identity, 96% coverage, and a similarity of 71%, was chosen for model building procedure. An important catalytic requirement for citE, being metalloprotein, is its dependency on a magnesium ion as cofactor. Therefore it was imperative that a magnesium ion be modeled along with the citE protein. This functionality of incorporating heteroatoms (metal ions, ligands etc.) in the target structure, given they are present in template, is uniquely inherent in the program MODELLER (Eswar, et al., 2008). Therefore, it was exploited to generate citE structure with magnesium ion included. The initial alignment between target and template was carried out via the align2d method of MODELLER9.10. The model building script was then modified to include a variable that allows magnesium (a heteroatom) to be recognized by the program while parsing the template pdb file. MODELLER9.10 considers heteroatoms as rigid entities and uses their spatial information such as bond distances and angles from the template pdb to place them in to the target protein structure (Sánchez and Šali, 2000). In addition to MODELLER9.10, structure for the citE protein, was also attained through three web servers: SWISS-MODEL (Schwede, et al., 2003), ModWeb (Pieper, et al., 2004) and I-TASSER (Zhang, 2008) for comparative purposes.

2.2.3.1. MODELLER

MODELLER is used to predict three-dimensional structures of proteins via homology or comparative approach. Alignments of a sequence with known related structures are provided as input and MODELLER generates a model containing all non-hydrogen atoms.

2.2.3.2. SWISS-MODEL

SWISS-MODEL is an automatic server used to model tertiary and quaternary structures of proteins. It can be accessed through ExPASY web server and Swiss PDB-Viewer:

(http://swissmodel.expasy.org/).

2.2.3.3. I-TASSER

I-Tasser is an online server used for prediction of protein structures as well as their functions. It is based on multiple threading alignments by LOMETS. It can be accessed from: *http://zhanglab.ccmb.med.umich.edu/I-TASSER/.*

2.2.3.4. ModWeb

ModWeb, an online server, is used for comparative protein structure modeling. It can be used via ModPipe, which is a large-scale, protein structure modeling pipeline and used for its proper functionality. It can be accessed from the following URL: *https://modbase.compbio.ucsf.edu/scgi/modweb.cgi.*

2.2.3.5. Structure evaluation

An important step that concluded the model generation procedure was the evaluation of model quality. To this end, an array of tools was available which helped quantification and therefore validation of stereochemical properties and structural constraints. Highly precise tools can be readily accessed at National Institute of Health (NIH) server that provides all major structure validation tools through (SAVES) Structural Analysis and Verification Server. These tools are: PROCHECK (Laskowski, Moss and Thornton, 1993), Errat (Colovos and Yeates, 1993) and ProSA-web (Wiederstein and Manfred, 2007).

PROCHECK estimates stereochemical properties of a protein model; like Ramachandran plot, G-Factor and Bad Contacts. It calculates the energy of each residue of the overall model and the results are displayed in the form of a graph. Ramachandran plot, the graphical image from PROCHECK represents the distribution of individual protein residues within the predefined allowed and disallowed regions derived from evaluation of phi and psi angles of experimental structures (Ramakrishnan and Ramachandran, 1965). G-Factor and Bad Contacts are measures of the main chain reliability reflecting the relative positioning of non-bonded atoms relative to each other (Morris, et al., 1992). Errat is used for evaluating and refining the protein model. This verification algorithm works by statistically inspecting the non-bonded interactions among different atom types. ProSA-web is a tool used for validation of protein structures on the basis of z-score.

2.2.4. Energy Minimization

Among the evaluated models, best structure was selected and energy optimization of the protein model was carried out to improve its quality. The energy minimization procedure was performed by UCSF Chimera, resourceful visualization software offering various capabilities for structural analysis (Pettersen, et al., 2004). Gasteiger charges were assigned to the protein and structural relaxation was achieved by application of 1500 rounds of minimization runs (750 steepest descent followed by 750 conjugate gradient) with a step size of 0.02 Å, under ff03.r1 force field. The minimized protein structure was subsequently evaluated through the use of aforementioned quality assessment tools and utilized in the molecular docking.

2.2.5. Molecular Docking Protocol

Docking protocol is divided into three steps: active site identification, inhibitors/ligands preparation and molecular docking. Detailed methodology of these steps as follows:

2.2.5.1. Active Site Identification

Active pocket on a protein surface is assessed using 3D structure of a protein as structural information, as the role of important active site amino acid residues in the enzyme activity,

is critical for developing enzyme-specific antibacterial drugs. Multiple computational approaches with varying accuracies have been developed for prediction of active site in a 3D structure of protein. Same proteins share structural features to great extent even in different species specifically active site residues. Active site in citE was identified via literature search (Zarzycki and Kerfeld, 2013) and DoGSiteScorer, which predicts binding pocket on the basis of druggability (Volkamer, et al., 2012) and can be accessed at *http://dogsite.zbh.uni-hamburg.de/*. Active site mentioned in literature was then explored in our target sequence manually and through sequence alignment. UniprotKB was also explored for active site residue indicators, which turned out be to the same as reported in literature. The conserved residues were further prospected for their role in ligand binding.

2.2.5.2. Ligand Preparation

Potential inhibitors against the target protein were collected from BRENDA, which is a database of enzymes and contains relational information for inhibitors, cofactors, reaction kinetics and molecular functions etc. (Schomburg, Chang and Schomburg, 2002). It is available at *http://www.brenda-enzymes.org/*. A total of one hundred and six (106) inhibitors were collected. The 2D structures of the selected inhibitors were drawn and minimized through ChemOffice 2012 packages (Li, et al., 2004). ChemDraw Ultra 12.0 was used for structure drawing while the structures were minimized via Chem3D Pro 12.0 using MM2 force field.

2.2.5.3. Molecular Docking

The minimized protein and inhibitors were used to carry out molecular docking via the docking tools like Genetic Optimization for Ligand Docking (GOLD) (Jones, et al., 1997) and AutoDock Vina (Trott and Olson, 2010). Best docked complexes were characterized on the basis of Gold fitness score and binding affinities, respectively. For visualization of the docked protein complexes and to understand the interactions that contributed to binding of both the ligands and the metal cofactor in detail, LIGPLOT (Wallace, Laskowski and Thornton, 1995), Visual Molecular Dynamics (VMD) (Humphrey, Dalke and Schulten,

1996), UCSF Chimera (Pettersen, et al., 2004), Discovery Studio (DS) Visualizer 3.5 (Visualizer, 2012), and ligand interaction mode of Molecular Operating Environment (MOE) (Chemical computing Group, Inc., 2013) were used.

2.2.5.3.1. Docking Via GOLD

"Gold Genetic Optimization for Ligand Docking" (GOLD) Hermes package was used for docking. The standard default settings were adapted for docking poses such as population size was set to 100, niche size 2, selection pressure 1.1, operator weights for migrate 0, crossover 100, number of islands being 1, number of operations 10,000, number of dockings 10, and hydrogen atoms were added in the protein model.

Genetic Algorithm and GoldScore fitness function were implicated in this study using GOLD. The fitness function used in GOLD is as follows:

$$\text{GOLD Fitness} = S_{hb\_ext} + S_{vdw\_ext} + S_{hb\_int} + S_{vdw\_int} + S_{tor} \qquad (2.1)$$

Where S_{hb_ext} represents the hydrogen bonding score between protein-ligand, S_{vdw_ext} represents the Van der Waals score between protein-ligand, S_{hb_int} represents intramolecular hydrogen bonds within the ligand, S_{vdw_int} represents intramolecular strain within the ligand and S_{tor} represents the ligand's torsional energy.

Gold-Score fitness function (Eq. 2.1) was used as a benchmark to pick the best conformation and was taken in to account for further analysis.

2.2.5.3.2. Docking Via AutoDock Vina

Binding affinities of ligands along with GoldScore were calculated using AutoDock Vina. Vina implements Broyden-Fletcher-Goldfarb-Shanno (BFGS) method for local optimization, which is a fast method and combines both scoring functions and its derivatives (Trott and Olson, 2010). In the current study, the protocol adopted for AutoDock Vina included the ligand and protein pdbqt files; while the size of the grid set for docking was 30 Å × 30 Å × 30 Å, which was centered at x,y,z dimensions with the values of -5.056, 39.457 and 95.677, respectively with a grid box spacing of 1 Å. All the grid box parameters were defined for protein and ligand in the configuration file. Raccoon (Forli, 2010) was used to

make pdbqt files for ligands. One hundred and six ligands were docked via this setting and their binding affinities were calculated. Ligands with lowest binding affinity were extracted to make poses with receptor, which were used for further assessment.

2.2.6. Molecular Dynamics Simulation

Molecular dynamics (MD) simulations is a significant approach to get insight into the conformational aspects of biological systems and also to get considerable understanding of protein-ligand interactions and conformational changes involved in the process. MD simulations for the best docked complex were performed using SANDER module in Assisted Model Building with Energy Refinement (AMBER) program (Weiner and Kollman, 1981).

Simulation of biomolecules is performed in four steps, followed by the trajectory analysis, which are shown in Figure 2.3

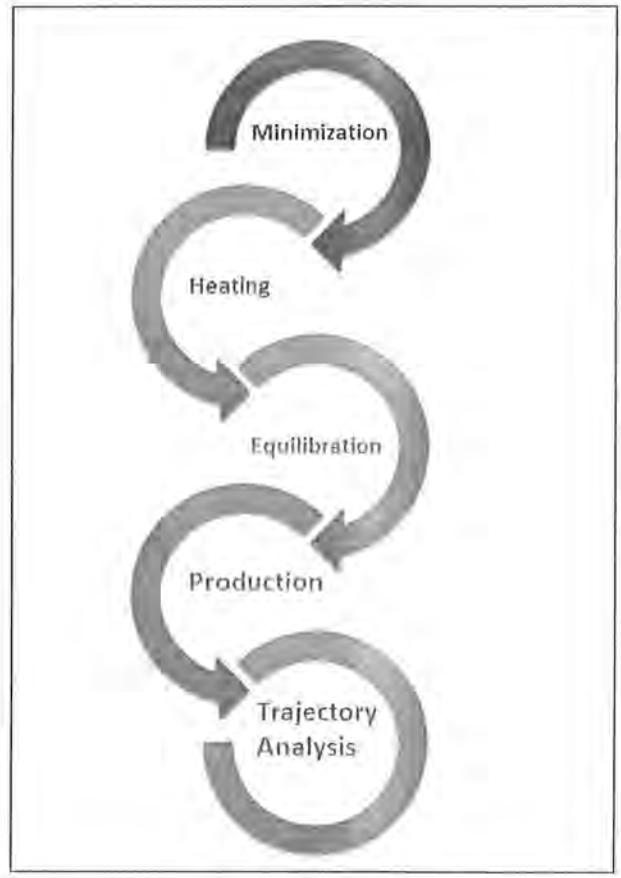

Figure 2. 3. *Steps involved in molecular dynamics simulations.*

2.2.6.1. System Preparation

MD simulations were performed using SANDER (Simulated Annealing with NMR Derived Energy Restraints) module in AMBER (Assisted Model Building with Energy Refinement) 12.0 suite of molecular dynamics program. The tLEaP module in Amber12 tools was employed to record the topology of the protein and the ligand. tLEaP, an AMBER 12 module, is an interface for primary coordinates and topology files preparation. Force fields used were ff03.r1, GAFF and ff99SB (Salomon-Ferrer, Case and Walker, 2013) and the system was solvated with three-point transferable intermolecular potential (TIP3P) water box with 8.0 Å (Figure 2.4). Ten (10) sodium (Na+) ions were added in order to neutralize the system. The parameters for magnesium ion (Mg^{+2}) were explicitly available in the antechamber program and therefore the cofactor library was loaded automatically. MD studies were carried out to investigate the conformations of protein receptor, compute the accurate energies and optimize the structures of docked complexes.

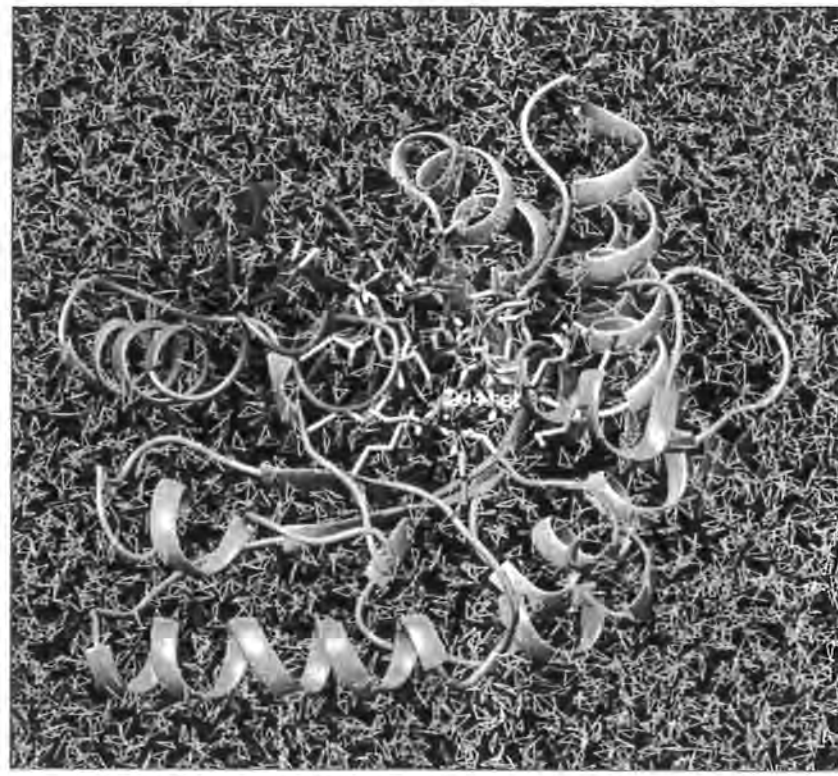

Figure 2. 4. *Solvation box surrounding docked protein.*

2.2.6.2. Minimization, Heating, Equilibration and Production

Functionality of tLEaP was extended to carry out minimization procedure of the docked system where a total of 5000 steps, split between 2500 steps of steepest descent and 2500 steps of conjugate gradient minimization were performed. For heating, equilibration and production run, periodic boundary conditions with a cut off value of 8.0 Å were used while maintaining a constant volume, temperature (300K) and pressure (1atm) (Berendsen, et al., 1984). During the production run, SHAKE algorithm (Ryckaert, Ciccotti and Berendsen, 1997) was applied to restrain hydrogen bond lengths of the system. Since canonical ensemble was used i.e. constant temperature, Langevin dynamics were employed with collision frequency set to 3.0.

Heating was performed for 10 picoseconds while equilibration was done for 100 picoseconds. The production run was performed in the similar manner using SANDER module to obtain trajectories. A total of 70 ns simulation run was obtained for the docked target. The analysis files were saved after 2 picosecond time step.

2.2.6.3. Simulation Trajectory Analysis

The trajectories generated as a result of simulation procedure were subjected to detailed analysis using PTRAJ module of AMBER to calculate four different quantities, namely, Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), Beta factor (B factor) and Radius of gyration (R_g). Graphical representations of these quantities for analytical purpose were viewed using Xmgrace (Vaught, 1996).

2.2.6.3.1. Root Mean Square Deviation

The coordinates of alpha carbon (Cα) are generally perceived as the representatives of the position of an amino acid in the three dimensional space. RMSD is a measure that allows comparison of relative positions of protein Cα atoms by computation of their averaged distances over a specific time interval (Maiorov and Crippen, 1994). RMSD is mathematically represented as:

$$RMSD = \sqrt{\frac{1}{N}\sum_i d_i^2} \qquad (2.2)$$

Where N is number of compared atoms, d_i is the distance between the i^{th} pair of atoms.

2.2.6.3.2. Root Mean Square Fluctuation

Root mean square fluctuation (RMSF) is defined as the root mean-square average distance of the given residue from its mean position (Kuzmanic and Zagrovic, 2010). The fluctuation of carbon alpha is calculated for each residue using the following equation:

$$RMSF = \sqrt{\frac{\sum_{t_k}^{T}(x_i(t_k)-x)}{T}} \qquad (2.3)$$

In this equation, x_i is position of C-α, x is averaged position and T is the time interval of that atom.

2.2.6.3.3. Beta factor

Beta factor is a term that is closely linked to the RMSF and measures the spatial displacement of atoms around their mean positions, generated as a consequence of the local vibrational and thermal movements (Kuzmanic and Zagrovic, 2010). Since they measure fluctuations they can be equated in terms of RMSF:

$$\beta Factor = RMSF^2\left(\frac{8\pi^2}{3}\right) \qquad (2.4)$$

2.2.6.3.4. Radius of Gyration (R_g)

Radius of gyration is a measure of the overall packing quality and density of a structure (Goodfellow, 1990). It is a physical property that can also be experimentally calculated, often through the application of small-angle X-ray scattering (SAXS) (Hong and Lei, 2009). Quantification of the compactness of macromolecular systems was achieved by the implementation of the following equation:

$$R_g = \frac{\sum_{i=1}^{N} m_i(r_i-r_{cm})^2}{\sum_{i=1}^{N} m_i} \qquad (2.5)$$

Where N is the total number of atoms, m_i is the mass of atom 'i', r_i is the position vector of atom 'i', and r_{cm} is the center of mass of the molecule under consideration.

RESULTS

3. Results

The procedural strategy of the drug discovery process aimed at the identification of *K. pneumoniae* potential targets led to conspicuous findings at various stages. Subtractive genomic strategy screened out the potential targets at the end of each step. The productive outcomes achieved during this course of study are elucidated below.

3.1. Subtractive Genomics Approach

The subtractive genomics approach is divided into various steps which screen the genome of the organisms under study. In the current study, genomes of three strains of *K. pneumoniae* were screened and the proteins were sifted at each step on the basis of some user-defined thresholds. This approach helped in identifying potential drug targets, common to all the three strains. An overview of the progressive subtractive genomic screening procedure is shown in Figure 3.1.

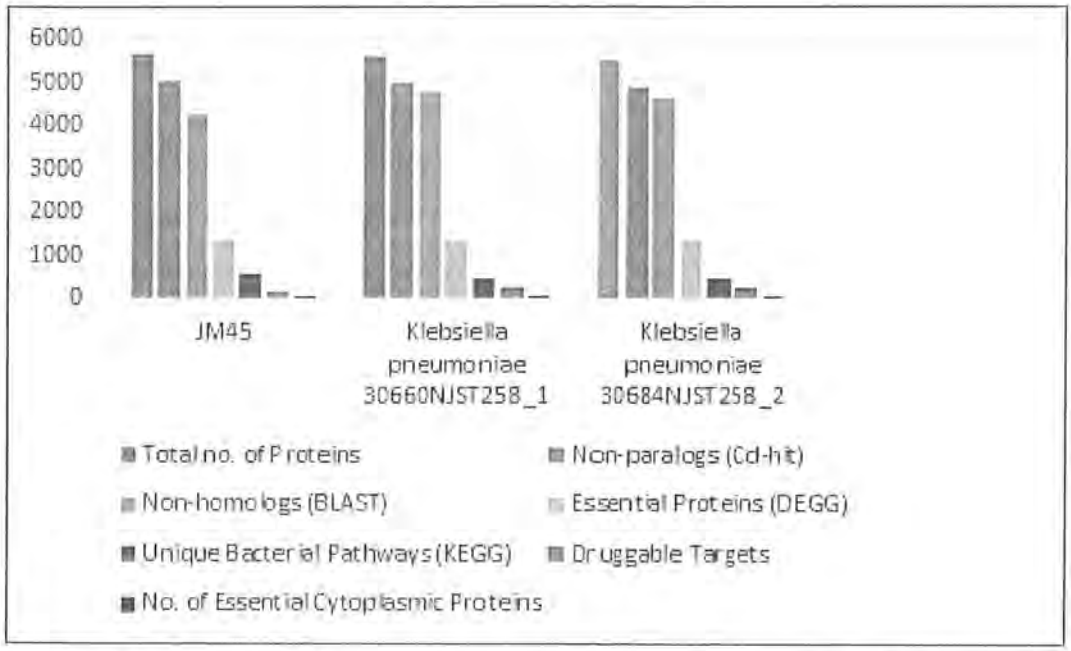

Figure 3. 1. *Overview of the screened proteins obtained at the end of each step of subtractive genomics.*

3.1.1. Genome Retrieval

K. pneumoniae strains namely, 30660NJST258_1, 30684NJST258_2, and JM45 were selected for current subtractive genomics study. These strains are completely sequenced and their genome was retrieved from UniProtKb. Genomics features of the selected strains are mentioned in Table 1.1.

3.1.2. Non-Paralogous Proteins

After the genome retrieval, the first step was to remove the non-paralogous proteins from the genomes of selected strains. It was performed by CD-HIT application. This led to the identification of 77, 112 and 719 paralogs (duplicated proteins/genes) in 30660NJST258_1, 30684NJST258_2, and JM45 *K. pneumoniae* strains, respectively. The remaining non-paralogous proteins i.e. 5521 in 30660NJST258_1, 5366 in 30684NJST258_2 and 4954 in JM45 were then subjected to further subtraction analysis.

3.1.3. Non-Homologous Proteins

Using Perl scripts, BLASTp was performed against human proteome to identify non-homologous proteins and a total of 729, 723 and 696 homologous proteins were removed from the proteome of 30660NJST258_1, 30684NJST258_2, and JM45, leaving only 4793, 4643 and 4258 proteins in three strains, respectively.

3.1.4. Pathogen Essential Proteins

This search was performed against the DEG database to identify essential proteins in the pathogen, which were imperative for their survival. After the DEG screening, essential genes identified were 1367 in 30660NJST258_1, 1364 in 30684NJST258_2 and 1323 in JM45, with 3224, 3076 and 2731 non-essential proteins left, respectively. The non-essential proteins were not included in the further analysis, due to their insignificance in bacterial survival.

3.1.5. Metabolic Pathway Analysis

K. pneumoniae essential proteins obtained from the previous step, were subjected to KAAS that was used to annotate the essential pathogen proteins, in order to identify the metabolic

pathways in which these proteins were involved. KASS identified 628, 623 and 597 proteins involved in 30660NJST258_1, 30684NJST258_2, and JM45 respective pathways.

3.1.6. Druggability Assessment

To ascertain the relation between screened targets and their drug binding ability, DrugBank 4.2 was used and 257, 255 and 253 druggable targets while 378, 379 and 365 non-essential druggable targets were identified in 30660NJST258_1, 30684NJST258_2, and JM45, respectively. During this screening procedure, many novel target proteins were also identified, for which no hit was scored in DrugBank.

3.1.7. Subcellular Localization

Durggable targets were then further processed for subcellular localization predictions and 6, 9 and 10 cytoplasmic proteins were identified in 30660NJST258_1, 30684NJST258_2, and JM45, respectively. Remaining proteins in the respective strains were either periplasmic or membranous. Cytoplasmic proteins were considered to be a potent candidate for being a putative drug target as the cytoplasmic proteins are usually enzymatic in nature, and thus aid bacterial growth.

3.2. Drug Target Selection

Unique pathways along with the number of proteins that were identified in *Klebisella pneumoniae* are listed in Figure 3.2. These pathways were common in all three strains. A total of 10 proteins involved in these pathways were common to all three strains. Out of these drug targets, six proteins were short listed with their evident parameters outlined in Table 3.1. Out of the six shortlisted drug targets, for the current study "Citrate lyase subunit beta"; citE, a metalloprotein (EC: 4.1.3.34), was selected for CADD analysis. CitE is critical protein for bacterial pathogenesis, as it is essential for anaerobic energy metabolism (Goulding, et al., 2007). CitE is involved in two-component pathway, which allows the organisms to adapt to a wide range of environments and growth conditions including antibiotic stress (Srinivasan, et al., 2012).

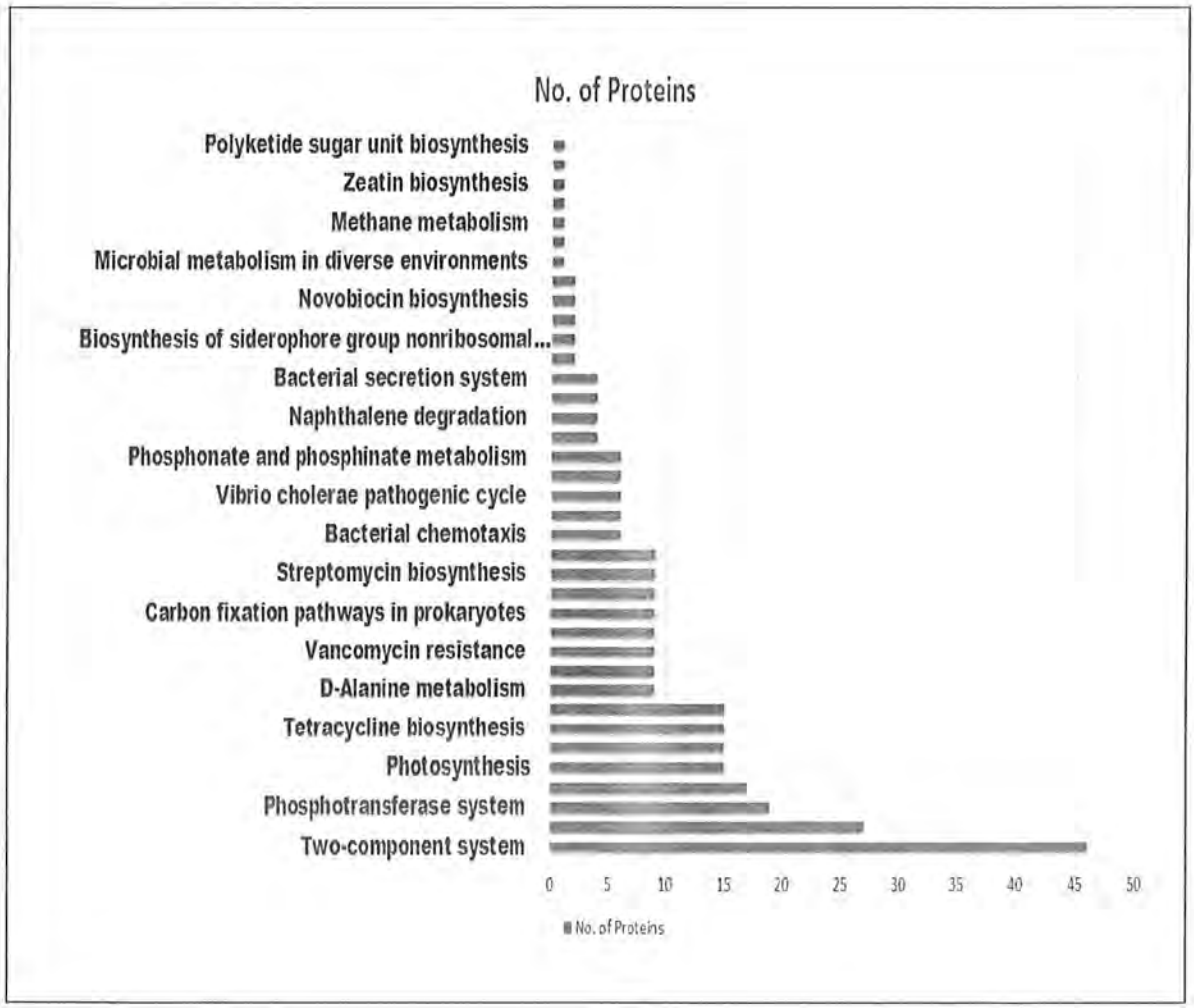

Figure 3. 2 *Number of proteins involved in the unique metabolic pathways of K. pneumoniae.*

Table 3.1. *Features used to identify feasibility of targets for CADD analysis.*

Gene Name	Unique Metabolic Pathway	Protein Length (AA)	Structure Available	Query Coverage with Target	Sequence Identity with Target	Cofactor	Subunit Structure
entE	Biosynthesis of siderophore group non-ribosomal peptides	535	No	98%	81%	No	Monomer
PTS-EI.PTSI	Phosphotransferase system	575	No	99%	96%	Mg^{+2}	Monomer
gspE	Bacterial secretion system	497	No	96%	61%	No	Dimer
glnG	Two-component system	471	No	98%	38%	No	Monomer
glrR	Two-component system	445	No	82%	41%	No	Monomer
cite	Two-component system	293	No	96%	33%	Mg^{+2}	Monomer

3.4. Comparative Homology Modeling

The CADD process begins with the structural availability of the chosen protein. Since experimental structure was unavailable for citE, comparative model building was carried out. At the sequence level, *K. pneumoniae* citE protein showed 96% coverage and 33% identity to the template structure PDB ID: 4L9Y, chain A. Using the template as a guide, structural models were generated through MODELLER9.10 and a variety of web servers. MODELLER 9.10 was used to generate five models along with Discrete Optimized Protein Energy (DOPE) score calculation. A thorough comparison of the stereochemical properties was subsequently carried out to select the best modeled structure (Table 3.2).

Based on the quality assessment measures obtained for the various homology models, Model number 3 generated via MODELLER9.10 was selected for further processing. In addition to providing significant coverage, Model 3 exhibited strong stereochemistry (Table 3.3). TIM-barrel structure was observed in citE which is considered as one of the most common conserved protein fold. It controls the enzymatic catalysis and maintains the structure of the protein (Nagano, Orengo and Thornton, 2002). The *K. pneumoniae* citE structure has the $(\beta/\alpha)_8$ TIM-barrel fold with an additional α-helix (Figure 3.7). TIM-barrel domain, showed the same depiction as in template structure. Moreover, when superimposed, the backbone atoms showed RMSD of 0.281 Å which is representative of high accuracy of the model. Another consideration made while selecting the model was that MODELLER9.10 fulfills the requirement of Mg^{+2} in the target protein structure. Since the web servers did not extend the capability of cofactor addition, Model 3 from MODELLER9.10 was selected for further processing. The model building procedure was followed by energy minimization in order to relax the overall structure and allow adjustment of side chains to remove steric clashes. An added benefit of optimization procedure was the improvement in the Errat quality factor which increased from 76.63 to 87.63 (Figure 3.3). Also, the optimization reduced the z-score from -5.1 to -5.9 (Figure 3.4). An illustrative view of the superimposed target-template structures depicting the accurate placement of the metal cofactor is shown in Figure 3.5. The homology model served as a starting point for the docking and subsequent simulation procedures.

Table 3. 2. *Stereo-chemical properties of comparative homology modeled structures.*

Structure Resource	Number of Residues				G Factor	Errat	Z Score
	[A,B,L] **Allowed region**	**[a,b,l,p]** **Additionally allowed region**	**[~a,~b,~l,~p]** **Generously allowed region**	**Disallowed region**			
MODELLER (1)	91.6%	7.3%	1.1%	0.0%	-0.61	80%	-6.18
MODELLER (2)	89.7%	8.4%	1.9%	0.0%	-0.57	78.5%	-5.44
MODELLER (3)	**92.0%**	**6.1%**	**1.5%**	**0.4%**	**-0.80**	**87.63%**	**-5.90**
MODELLER (4)	90.8%	6.9%	1.9%	0.4%	-0.44	82%	-6.02
MODELLER (5)	90.8%	8.0%	1.1%	0.0%	-0.52	72.2%	-4.95
I-TASSER	67.6%	25.6%	5.2%	1.6%	0.11	64%	-11.1
ModWeb	87.5%	5.3%	1.5%	0.4%	-0.27	71%	-9.03
SWISS-MODEL	89.2%	7.8%	0.4%	0.8%	-0.28	70%	-10.65

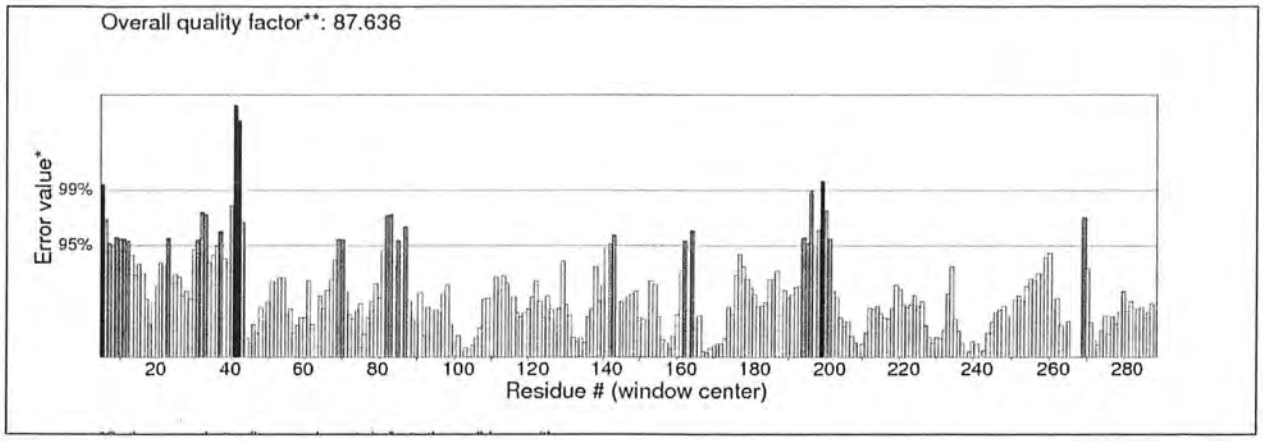

Figure 3. 3 *Errat plot of the selected MODELLER protein after minimization.*

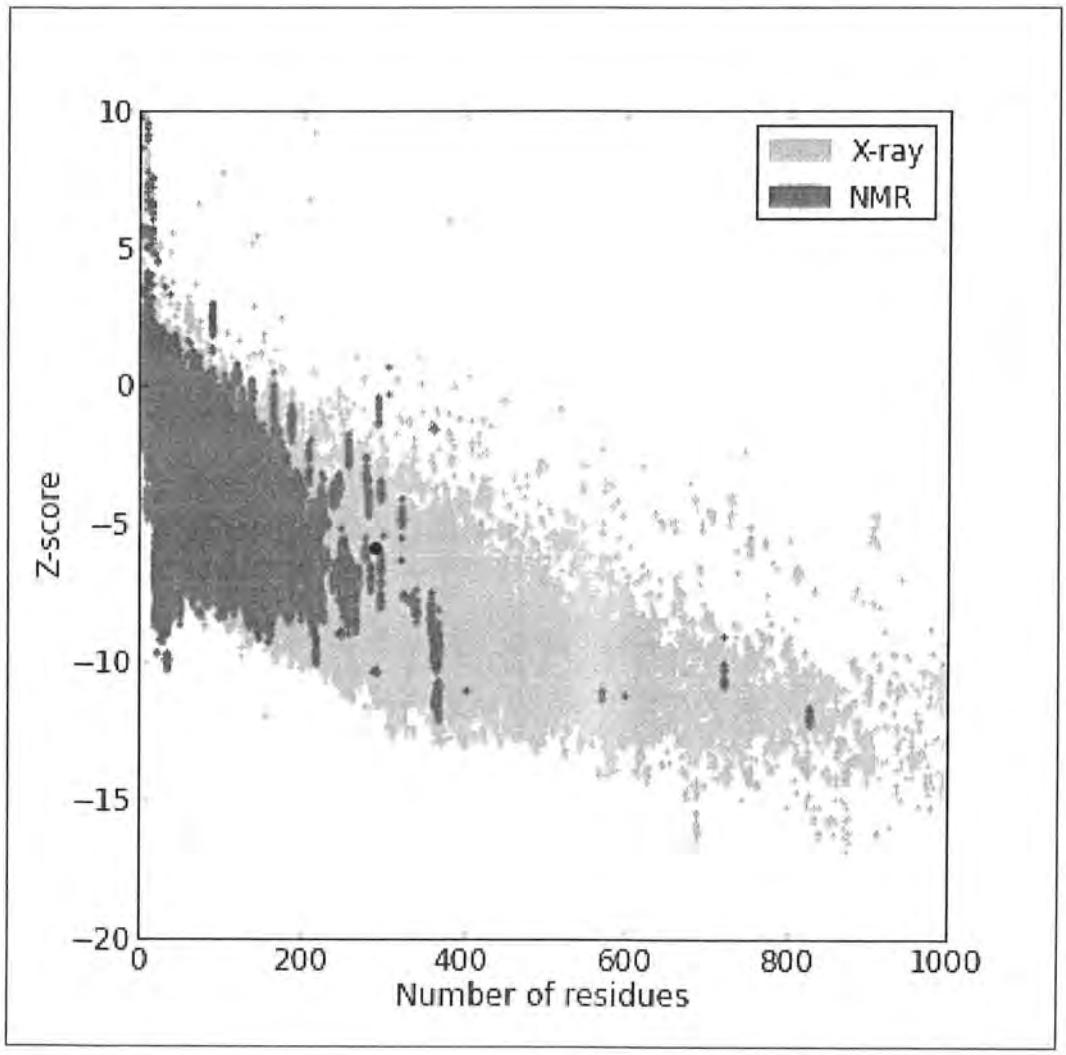

Figure 3. 4 *Z-score plot of the selected MODELLER protein after minimization.*

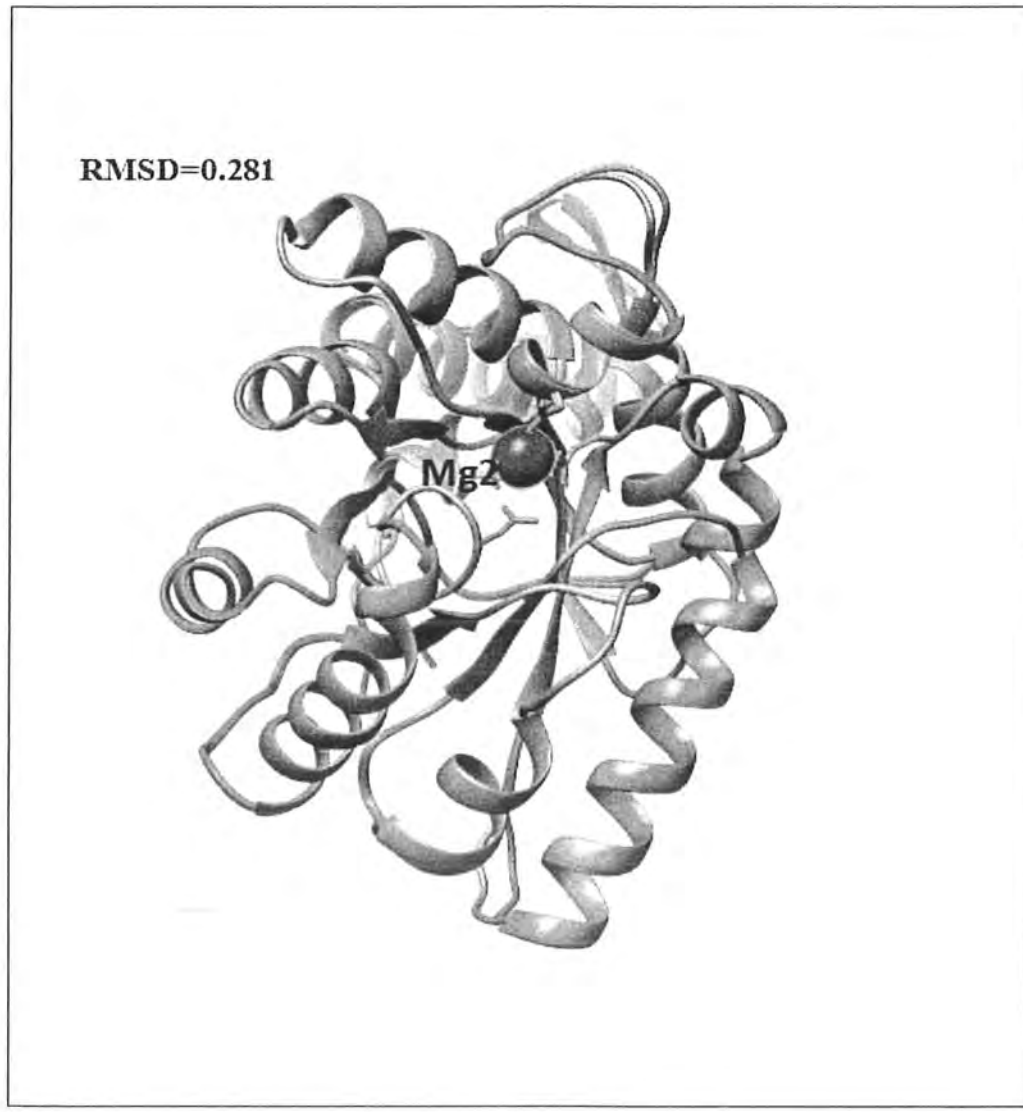

Figure 3. 5 *Superimposed structures of template 4L9Y (light green) and target citE (light brown) with magnesium ion.*

Table 3. 3. *Physicochemical properties of citE using ExPASy ProtParam tool.*

Physicochemical Properties	Values
Number of amino acids	293
Molecular weight	31780.4
Theoretical Pi	5.10
Total number of negatively charged residues (Asp + Glu)	42
Total number of positively charged residues (Arg + Lys)	30
Aliphatic index	102.29
Instability index	31.72 (Stable)
Grand average of hydropathicity (GRAVY)	0.054

3.5. Molecular Docking

The information about the domain organization within the citE protein combined with identification of the active site guided the molecular docking procedure.

3.5.1. Active Site Identification

Active site of citrate lyase subunit beta was identified via literature search and also reconfirmed using DoGSiteScorer (Volkamer, et al., 2012) . DoGSiteScorer predicted thirteen pockets and from the first pocket active residue was selected which was the same as cited in literature. The active site of citE constitutes of negatively charged amino acid residues, glutamic acid and aspartic acid, coupled with a positively charged magnesium ion (Zarzycki and Kerfeld, 2013), illustrated in Figure 3.6. The selected amino acid residue in

the current course of study is Asp160. A conserved active site is observed in different orthologues of citE.

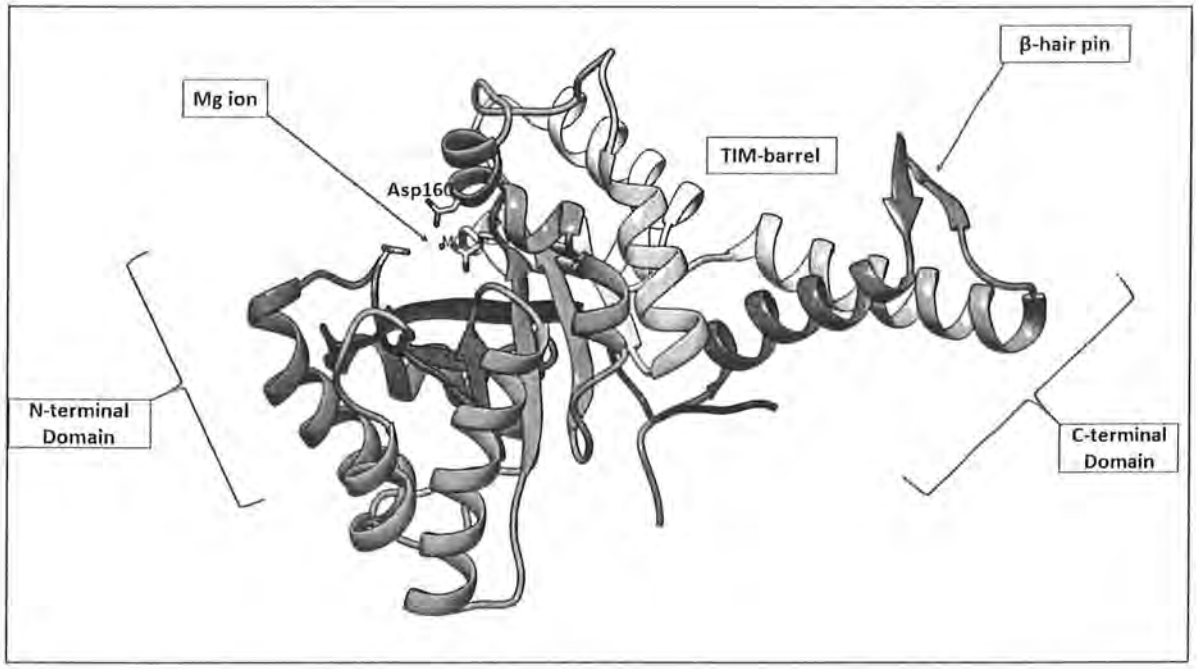

Figure 3. 6 *Topological view of citE, C and N-terminal domains with Mg^{+2} represented in the vicinity of the active site residue i.e Asp160.*

***Figure 3.** 7 Topological view of citE highlighting the TIM-barrel through the Bendix representation.*

3.5.2. Inhibitors Selection

The inhibitors selected to be docked into the active site of citrate lyase subunit beta, were mostly selected from the literature. Compounds from literature, and their analogues were used as potential inhibitors against citrate lyase subunit beta (Aoshima, Ishii and Igarashi, 2004). In the current study, the inhibitors were accessed from the BRaunschweig ENzyme Database (BRENDA) (Schomburg, Chang and Schomburg, 2002) and a total of one hundred and six (106) inhibitors were docked into the active site of citE.

3.5.3. Interaction Analysis

A total of 106 ligands were docked into the active site of the target, using GOLD and AutoDock Vina for calculation of GOLDScore and binding affinities, respectively. Along with this, preferred binding pocket orientation of active compounds was also identified. Conezyme-A (CoA) was the top scoring compound in both GOLD and Vina docking results. Current docking results depict the π-interactions and hydrogen bonding between the potential inhibitor and active site residues.

3.5.4. Active Site Binding Analysis

The prepared ligand molecules were docked into the active site of the target using GOLD. Corresponding binding affinities were also calculated using AutoDock Vina. The highest GoldScore of 102.8 was achieved for compound 47, with a binding affinity of -8.0 kcal/mol. Docking scores and respective binding affinities for top 10 compounds, arranged in descending order of GoldScore values are provided in Table 3.4. Detailed visualization analysis carried out through UCSF Chimera, LIGPLOT and MOE revealed the conformational details and preferred orientation of the ligand binding. Additionally it was seen that the metal ion, Mg^{+2} conferred no visible changes upon ligand binding. A graphical representation of the docked ligand via GOLD and Vina, positioned within the active site is outlined in Figure 3.8 and Figure 3.9, respectively.

Table 3. 4. *Docking results of top ten docked inhibitors in descending order of GOLD Score with the corresponding binding affinities.*

Sr. No.	Compound No.	GoldScore	Binding Affinity (kcal/mol)
1.	47	102.8	-8.0
2.	55	96.5	-7.9
3.	10	94.1	-7.9
4.	33	91.0	-6.5
5.	105	90.6	-7.1
6.	61	90.5	-6.9
7.	57	89.3	-7.5
8.	52	87.9	-7.3
9.	92	87.1	-6.8
10.	78	85.4	-7.5

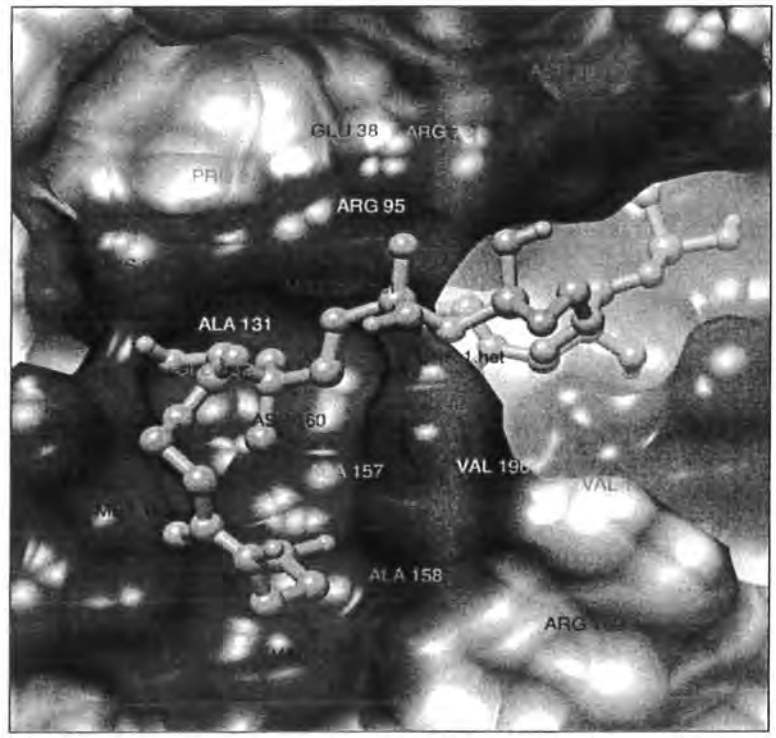

Figure 3. 8 *Best docked inhibitor (spring green) in the active site of citE with Mg^{+2} also shown to be a part of the active site.*

Figure 3. 9 *An overview of compound 47 docked into the active site domain of the selecetd target, citE (VINA).*

The binding of compound 47 was observed at the active site of the protein and residues involved in electrostatic interactions were Val197, Met163, Asp160, Phe159, Val158, Glu133, Lys98, Arg95, Arg70, Ala40, Asp39, Glu38, Asp36, Pro15, Ile14 and Phe13 (Figure 3.10). Besides this Val225, Val196, Val162, Ala158, Ala131 and Pro97 showed π-interaction with ligand, depicted in Figure 3.11. Moreover, LIGPLOT image i.e. Figure 3.12 illustrated the presence of hydrogen bonds between the ligand and the target. The Mg^{+2} interactions at the active site of citE were also illustrated using LIGPLOT and can be viewed in Figure 3.13. Ligand oxygen moiety formed two hydrogen bonds with the active residue Lys98 having 2.29 Å and 3.86 Å distance and nitrogen atoms of the ligand developed hydrogen bonding with Arg95, Glu133 and Asp160 residues of the protein, at a distance of 3.23 Å, 2.91 Å and 3.46 Å, respectively. Besides that, Mg^{+2} binding site exhibited hydrogen bonding with Glu38, Asp39, Glu133 and Asp160 residues. Hydrogen bonding depicted by Mg^{+2} did not alter after the docking of ligand in the active site. The hydrogen bonds present between the ion and the above mentioned residues had 1.87 Å, 1.85 Å, 1.87 Å and 1.88 Å distances, respectively (Figure 3.14). Along with these, hydrogen bond details of ligand with the target residues are given in Table 3.6.

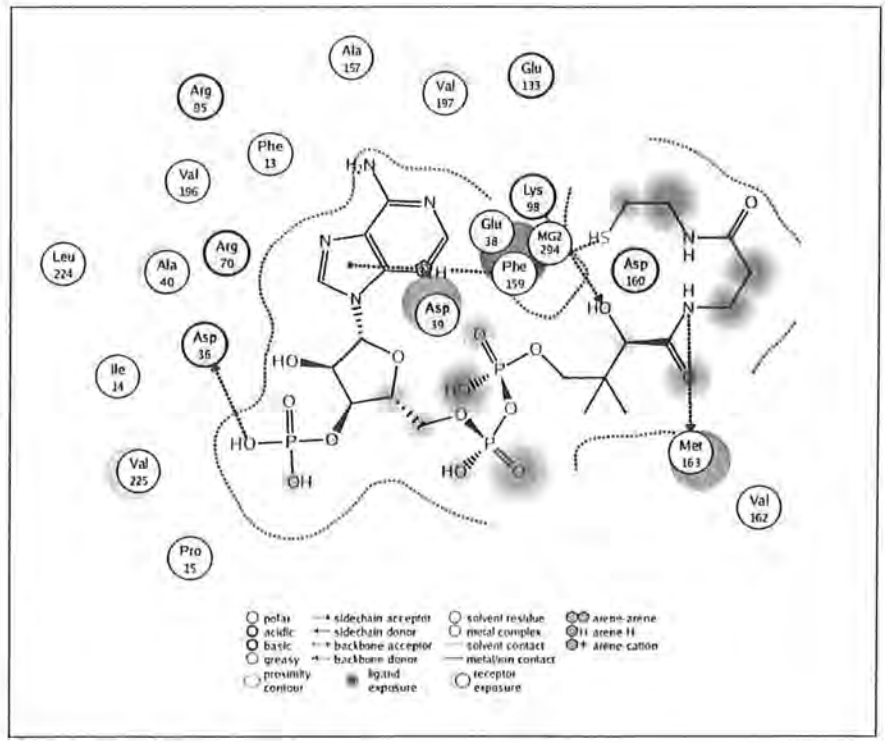

Figure 3. 10 *MOE ligand interaction image showing bonded and non-bonded interactions of inhibitor bound citE.*

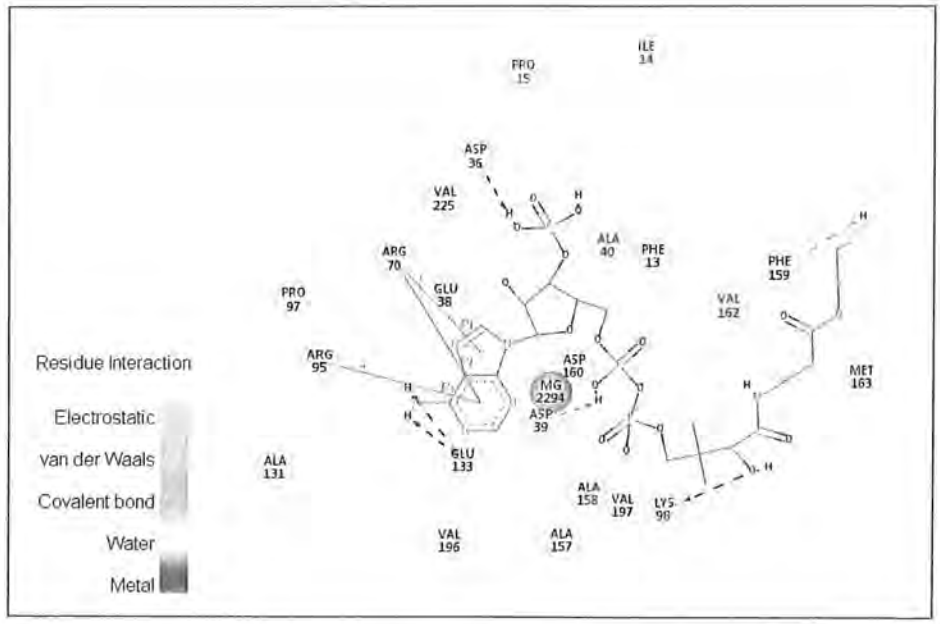

Figure 3. 11 *DS Visualizer 2D depiction of compound 47 representing hydrogen bonds and π interactions.*

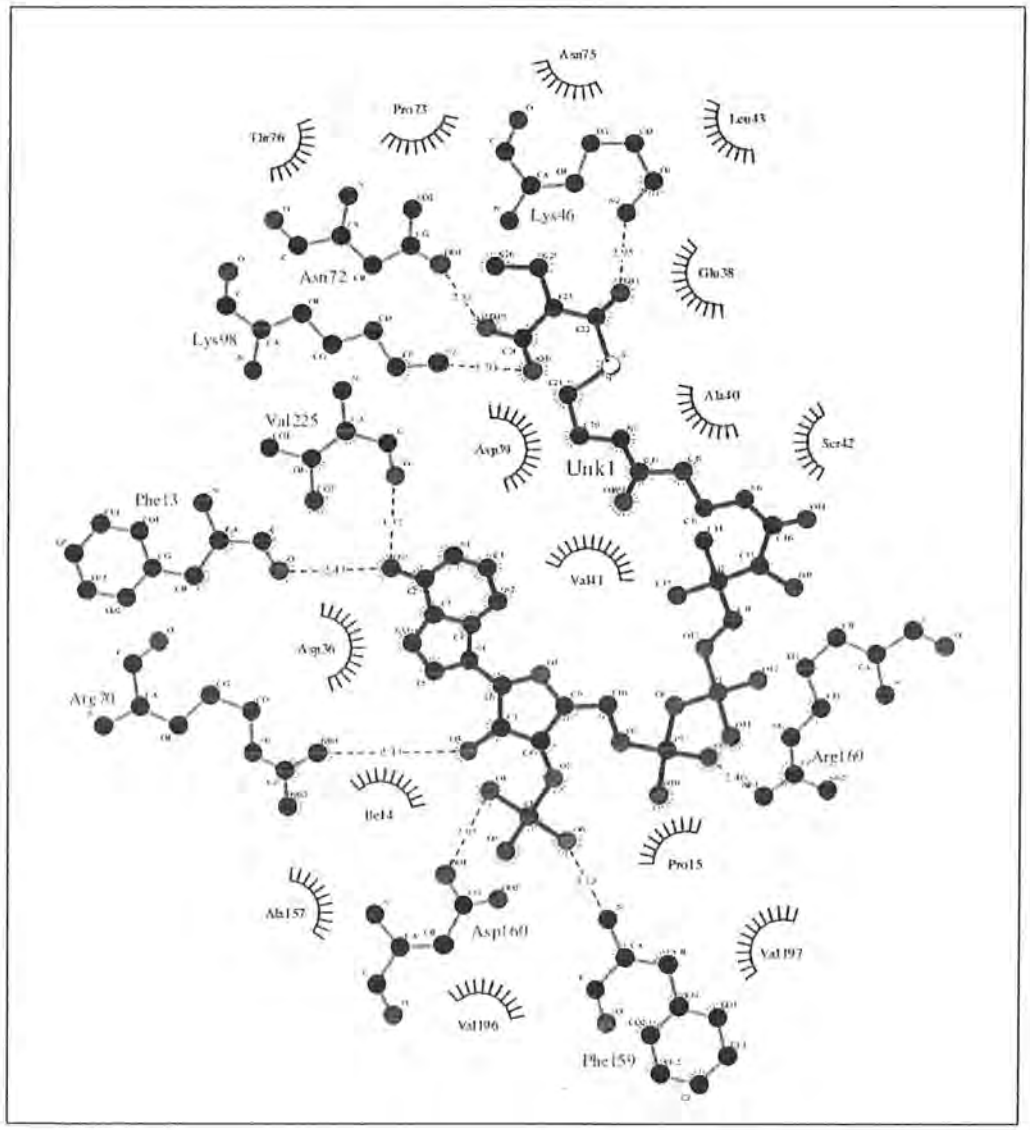

Figure 3. 12 *Interaction of ligand with citE, highlighting interacting residues through LIGPLOT.*

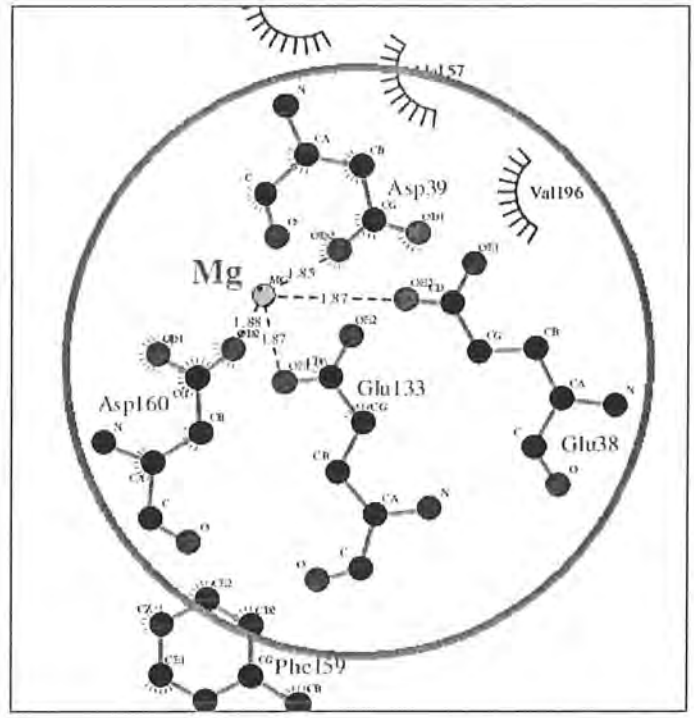

Figure 3. 13 *LIGPLOT illustration of Mg^{+2} interactions with their surrounding protein residues.*

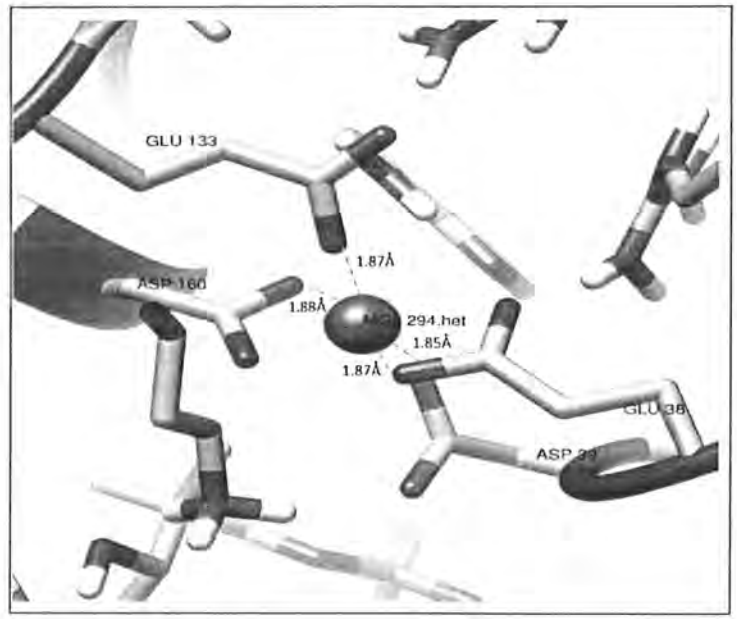

Figure 3. 14 *The placement of Mg^{+2} in the active binding pocket of citE.*

Table 3. 5. *Hydrogen bond details of best docked compound with important interacting residues.*

Protein Interacting Atom	Ligand Interacting Atom	Distance (Å)
36 ASP OD2	UNK H	3.20
38 GLU OE1	UNK N	3.72
39 ASP OD2	UNK H	2.75
70 ARG NH2	UNK H	3.85
70 ARG NH1	UNK H	3.20
95 ARG NH2	UNK H	2.57
95 ARG NH2	UNK H	3.39
95 ARG HH22	UNK N	3.23
98 LYS HZ1	UNK O	2.29
98 LYS HE2	UNK O	3.86
98 LYS NZ	UNK O	3.29
98 LYS NZ	UNK H	3.71
133 GLU HG2	UNK N	2.91
133 GLU OE1	UNK H	2.37
133 GLU OE2	UNK H	2.78
160 ASP OD1	UNK H	3.00
160 ASP OD1	UNK H	2.57
160 ASP H	UNK N	3.46
160 ASP OD1	UNK N	3.40

3.6. Molecular Dynamics Simulation

The docking study provided meaningful insights into the structural basis of druggability potential of *K. pneumoniae* citE. However, it provided this information within the context of a static environment. In order to infiltrate the dynamic behavior of citE, simulation protocol was carried out followed by trajectory analysis to assess various properties of the ligand bound protein. Simulation studies not only comprehend the dynamic behavior of proteins but also highlight important residues which play critical role in dynamic behavior (Azam, et al., 2013) Properties including the RMSD, RMSF, B-factor and radius of gyration were plotted as a function of time to understand the biomolecular movements within a solvated environment. Analysis of protein in ligand bound form led to evaluation of structural transformation and underlying atomic level transitions. It helped unravel ligand induced variability and the dynamic role of co-factor in the presence or absence of inhibitor.

3.6.1. Root Mean Square Deviations (RMSD)

The deviation of the backbone Cα atoms was noticed for the entire production run of the docked protein for a time period of 70 ns. The RMSD behavior of the inhibitor bound citE, over the studied time scale is mostly stable with an average value of 3.13 Å reaching the maximum value of 4.58 Å at the 15^{th} ns (Figure 3.15). Overall, the pattern of RMSD represents a single extensive domain shift within the structural framework of the protein-ligand complex, where a β-hairpin transforms into a loop (Figure 3.16), which occurs at the beginning (around 15 ns) of the simulation and continues till the end. The ligand placement was well complemented within the active site during the simulation and does not destabilize the protein (Figure 3.17).

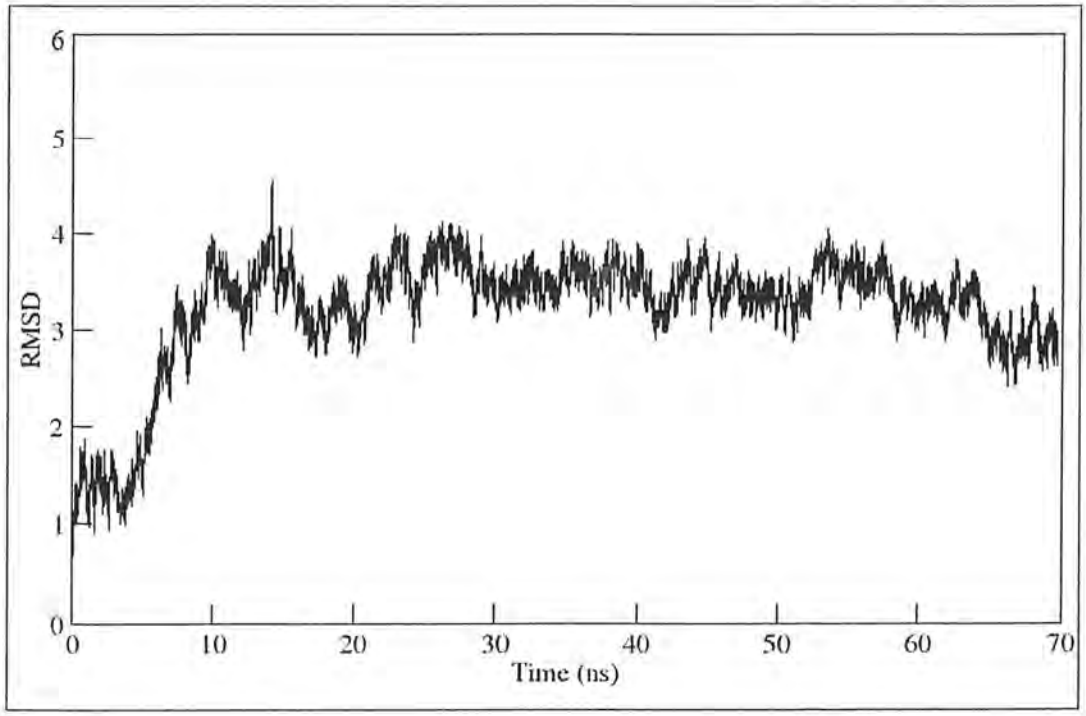

Figure 3. 15 *RMSD plot of docked citE protein complex for 70 ns simulation run*

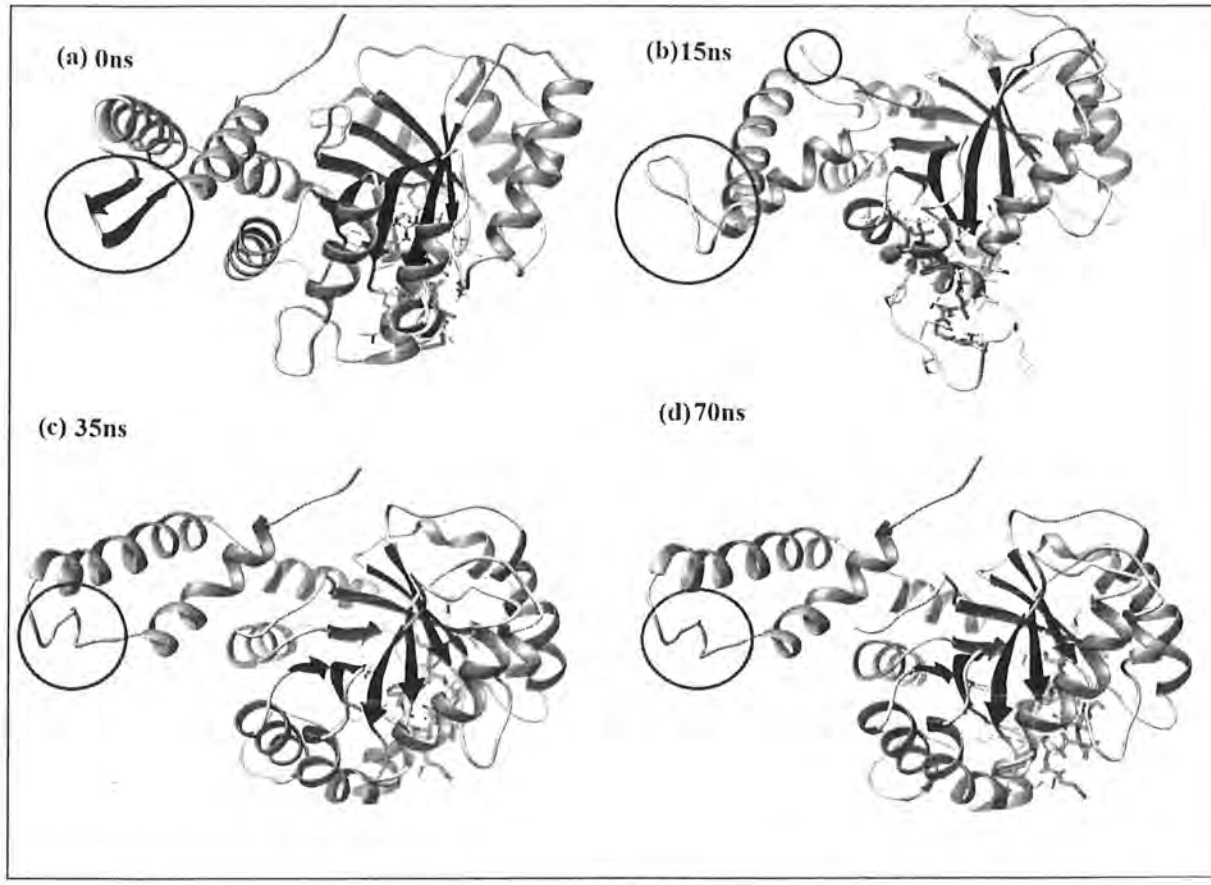

Figure 3. 16 *Snapshots of docked citE over a time lapse of 0 ns, 15 ns, 35 ns and 70 ns. The helices are depicted in spring green color, sheets in blue and the loops in pink color.*

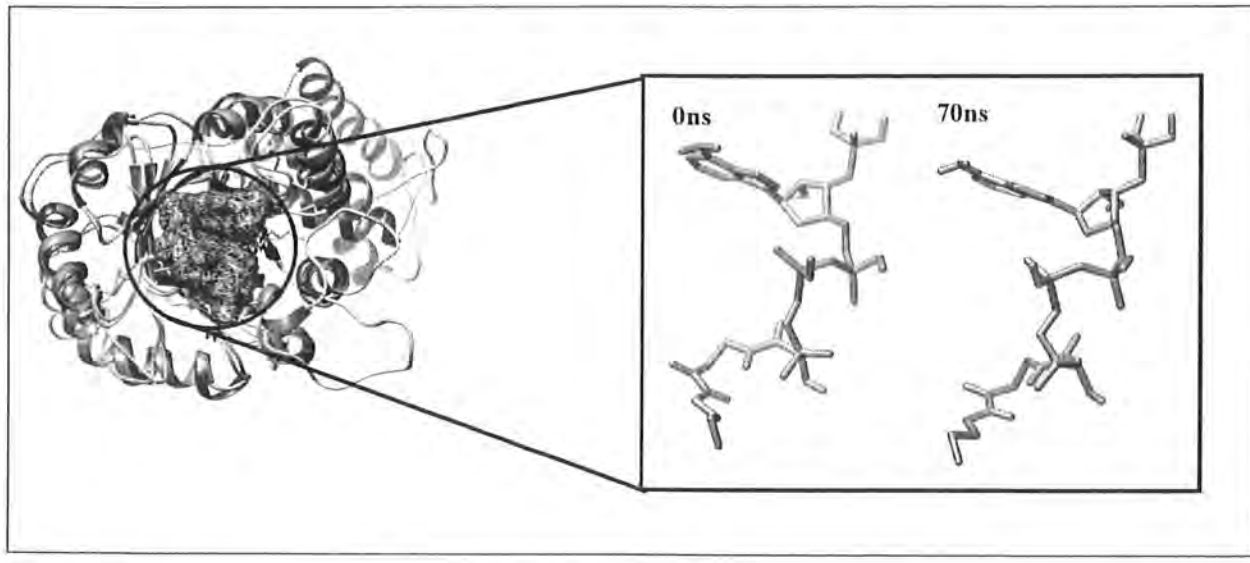

Figure 3. 17 *Ligand placement within the active site of citE between 0 ns and 70 ns.*

3.6.2. Root Mean Square Fluctuations (RMSF)

As a measure of atomic fluctuations, RMSF provides a means to recognize and comprehend the structurally flexible and rigid regions of the drug target. The average Cα fluctuation for the ligand-bound protein was observed to be 2.48 Å (Figure 3.18). The maximum value of RMSF for the ligand bound protein was 9.93 Å. Detailed analysis of the trajectories led to identification of those protein substructures that are responsible for the obtained RMSF trend. The most important observation in this context was that higher fluctuations were observed for regions that form loops and turns and are solvent exposed. In particular, the residues forming the C-terminal region, ranging between residues 257-271 exhibit a transient behavior where the β-hairpin transforms into a loop region, connecting two helices. The active site pocket inclusive of residues 36, 39, 95, 98, 129, 133, 157 and 160, had an RMSF value of less than 2.0 Å signifying the stability during the production run.

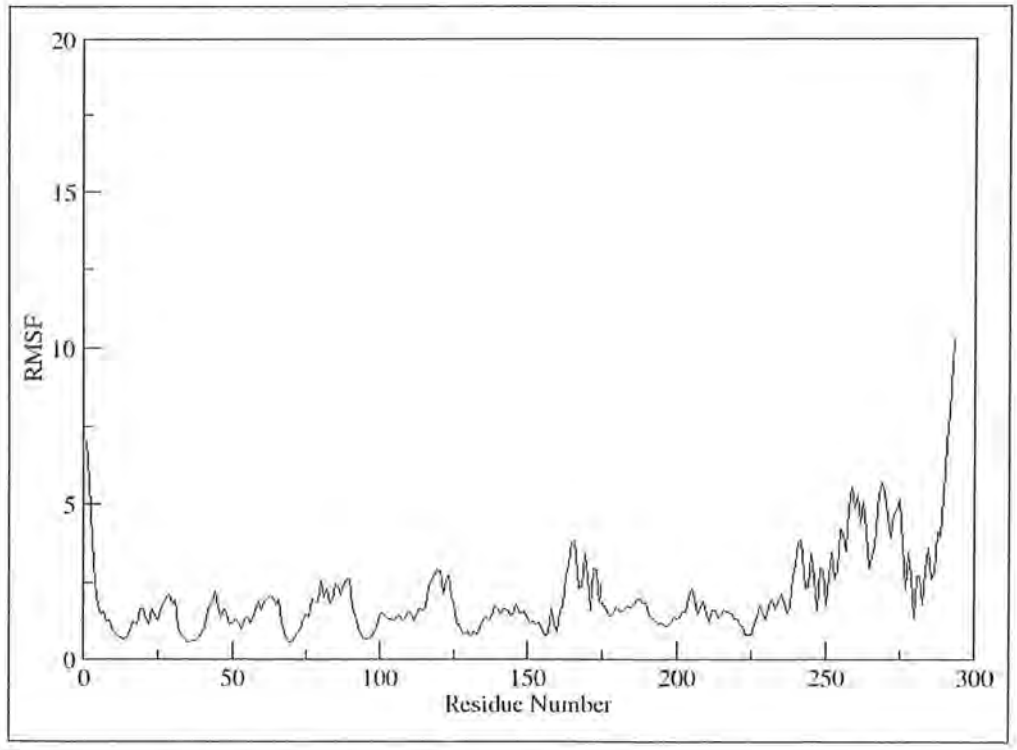

Figure 3. 18 *RMSF plot of docked citE protein over 70 ns simulation run.*

3.6.3. β-Factor Analysis

Beta factor is a quantity that is measured in terms of RMSF. Its value is therefore dependent on the level of localized atomic fluctuations which collectively contribute to the global vibrational movements of the protein and its thermal stability. The pattern of beta factor for citE is consistent with the RMSF trend. Regions having greater fluctuations as identified in the RMSF analysis exhibit beta factor of greater than 178 $Å^2$ with the maximum value for terminal residues: 2739.81 $Å^2$ (Figure 3.19).

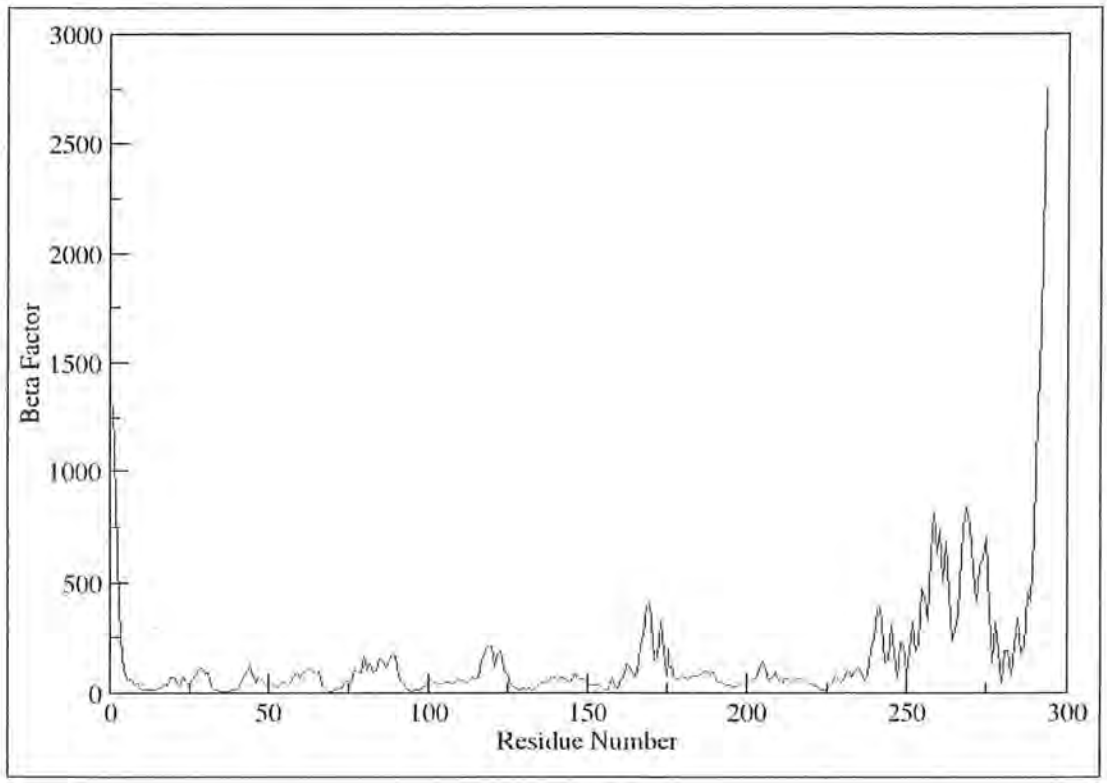

***Figure 3. 19** β-Factor graph of docked citE protein over 70 ns simulation run.*

3.6.4. Radius of Gyration (R_g)

As an evaluation parameter of the structural compactness, radius of gyration was calculated as a time function for the 70 ns simulation of the ligand-protein complex. The average value of 20.23 Å was observed that denotes the stability of the protein structure (Figure 3.20).

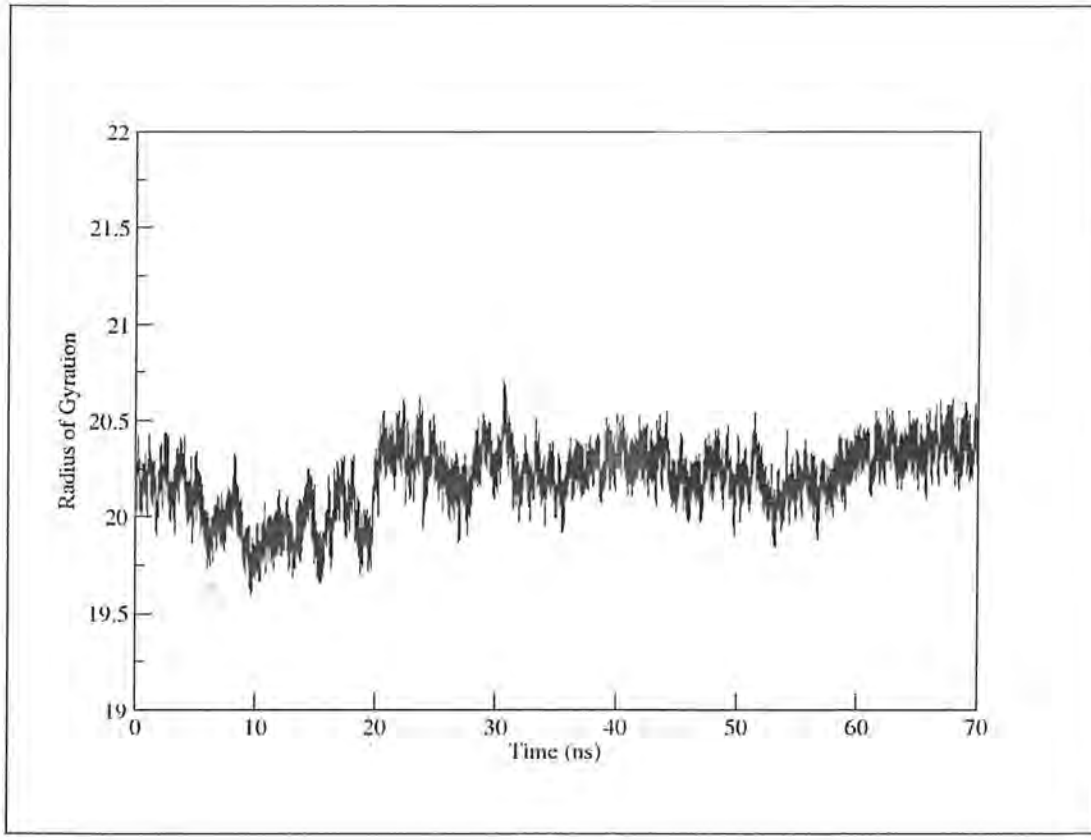

Figure 3. 20 *Radius of gyration of docked protein citE over 70 ns simulation time period.*

3.6.5. Cofactor Coordination Dynamics

A vital aspect, central to the functionality of citE is the presence of cofactor Mg^{+2} in its active site. The molecular dynamics simulation study helped to gain an understanding of the structural adjustments made upon ligand introduction into the system. However, the ligand addition, visibly did not affect the metal ion placement within the active site. No changes occurred in the binding pattern of the Mg^{+2} and it occupied the same location during the 70 ns simulation run (Figure 3.21).

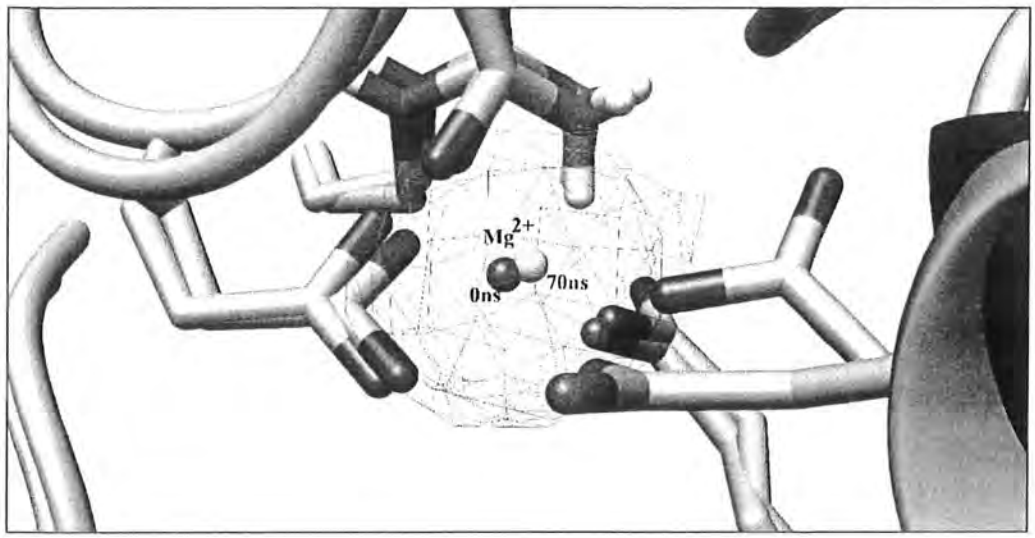

Figure 3. 21 *The co-factor Mg^{+2} in the superimposed poses of the docked protein at a leap of 0 ns and 70 ns.*

DISCUSSION

4. Discussion

The emergence of ongoing antibiotic resistance infections are a punitive epidemic to the global world. This phenomenon is not just because of high rate of bacterial genome reshaping by mutation but also triggered by antibiotics' imposed selective pressures. Mortality rates owing to infectious diseases caused by these resistant bacteria are increasingly affecting millions of people. Failures of antimicrobial drugs have been catastrophic. Thus, all these bottlenecks demand newer, specific and more potent therapeutic agents. Progressive expansion of bioinformatics based *in silico* techniques successfully are, therefore, assisting in overcoming different obstacles in drug discovery process.

The current work emphasizes on the identification of potential druggable candidates of *K. pneumoniae*, a Gram negative multidrug resistant, nosocomial pathogen (Berman, 2012). The procedure commenced with the differential proteome analysis whereby the obtained potential drug candidates reflected two important features: firstly, they were exclusive to the pathogen thereby avoiding cross reactivity issues; secondly, they were essential to the pathogen metabolism. Therefore, if targeted, their inhibition may have a bactericidal effect (Bakheet and Doig, 2010). Subtractive genomics is a thriving *in silico* approach that filters out pathogenic novel drug targets from complete genome. 30660NJST258_1, 30684NJST258_2, and JM45 are *K. pneumoniae* strains which were targeted to combat multiple strains with the same drug. Initially a total of 5598, 5478 and 5673 proteins were present in the three strains, respectively. CD-HIT tool removed all the duplicated sequences from pathogen genome so that the drug molecule can bind to the specific target. Furthermore, it was ensured that the host and pathogen have no homologous sequences, so the drug may not affect the host. BLASTp performed against RefSeq database filtered non-homologous pathogen sequences. Essentiality of non-homologous sequences for bacteria was investigated through homology search. Pathogen essential proteins are of great impact for interfering the pathogen survival within the host system thus endangering the bacterial growth. Essential proteins were thus subjected to metabolic pathways association predictions. During pathway analysis, a total of thirty eight unique bacterial pathways were identified which could be targeted. Conservation of these pathways among different pathogenic

strains suggests the importance of their constituent proteins in bacterial virulence and survival (Hosen, et al., 2014). Among the unique bacterial pathways higher number of unique, essential, non-homologous genes were present in "Two-component system", "Biosynthesis of secondary metabolites", "Phosphotransferase system", "Photosynthesis" and "Peptidoglycan biosynthesis". The unique proteins identified from the selected pathways were subjected to another round of analysis which incorporated a two-step evaluation where druggability was assessed followed by subcellular localization assessment. Protein functionality is highly dependent on its proper localization in living cell. Determination of target localization is important as, the drug has to agglomerate with the intended target. Cytoplasmic proteins can act as possible therapeutic targets, while membrane proteins are used to design vaccines (Barh, et al., 2011). Finally, towards the end of the subtractive genomics steps, 6, 9 and 10 potential drug target proteins were shortlisted in the three strains, respectively. Of the obtained potential drug targets, 5 cytoplasmic, bacterial enzymatic proteins were shortlisted. CitE, a critical protein for bacterial pathogenesis was selected as a therapeutic candidate, as it is essential for fatty acid biosynthesis or anaerobic energy metabolism (Goulding, et al., 2007). CitE is involved in two-component pathway, which serves as a basic stimulus-response coupling mechanism to allow organisms to sense and adapt to a wide range of environments, stressors, and growth conditions including antibiotic stress (Srinivasan, et al., 2012). This gene was common to all the three strains and was further extended to homology modeling and molecular docking to determine the binding modes of ligands.

At the sequence level, *K. pneumoniae* citE protein showed a 96% coverage to the template structure PDB ID: 4L9Y, chain A. Using the template as a guide, structural models were generated through MODELLER9.10 and a variety of web servers. The models were then evaluated with different tools like PROCHECK, Errat and ProSA-web, which tested the quality of structures generated. Based on the quality assessment measures obtained for the various homology models, Model number 3 generated via MODELLER9.10 was selected for further processing. In addition to providing significant coverage, Model 3 exhibited strong stereochemistry. ModWeb generated a model with stereochemistry and quality comparable to that of MODELLER model. However, citE is a magnesium dependent metallo-

enzyme, incorporation of Mg^{+2} in the structural model was a feature supported by MODELLER alone, thereby guiding the selection of the most structurally relevant as well a good quality model. RMSD has been established as an important structural measure that is directly representative of model quality, particularly in cases where the ultimate goal is molecular docking (Rodrigues, et al., 2013). When the template and target were superimposed, the backbone atoms showed RMSD of 0.281 Å, which is representative of high accuracy of the model. Model building procedure was followed by energy minimization in order to relax the overall structure and allow adjustment of side chains to remove steric clashes. An added benefit of optimization procedure was the improvement in the Errat quality factor which increased from 76.63 to 87.63. Also the optimization reduced the z-score from -5.1 to -5.9. TIM-barrel structure was observed in citE which is considered as one of the most common conserved protein fold. It controls the enzymatic catalysis and maintains the structure of the protein. The TIM-barrel domain, showed the same depiction as in template structure.

The homology model served as a starting point for the docking and subsequent simulation procedures. The ligand molecules were docked into the active site of the target using GOLD.and AutoDock Vina. The highest GoldScore of 102.8 was achieved for compound 47, with a binding affinity of -8.0 kcal/mol. In addition to highlighting the preferred ligand binding modes, it provided insight into the underlying chemical basis of interaction and type of inhibition. The binding of compound 47 was observed at the active site of the protein and the polar end of the ligand was establish interactions with residues, namely, Val197, Met163, Asp160, Phe159, Val158, Glu133, Lys98, Arg95, Arg70, Ala40, Asp39, Glu38, Asp36, Pro15, Ile14 and Phe13. The phenomenon of hydrogen bond formation has immense importance in relation to structural folding and stability and has been well documented with respect to protein interaction with ligand (Saenger and Jeffrey, 1991). Within the active site, ligand oxygen moiety formed two hydrogen bonds with the active residue Lys98 having 2.29 Å and 3.86 Å distance, respectively. While nitrogen atoms of the ligand developed hydrogen bonding with Arg95, Glu133 and Asp160 residues of the protein, at a distance of 3.23 Å, 2.91 Å and 3.46 Å, respectively. Besides that, Mg^{+2} binding site exhibited hydrogen bonding with Glu38, Asp39, Glu133 and Asp160 residues. Hydrogen

bonding depicted by Mg^{+2} did not alter after the docking of the ligand in the active site. Hydrogen bonds present between the Mg^{+2} and the above mentioned residues had 1.87 Å, 1.85 Å, 1.87 Å and 1.88 Å distances, respectively.

The docking study further needed an understanding of the structural adjustments made upon ligand introduction into the system. However, it provided this information within the context of a static environment. In order to allude the dynamic conduct, simulation protocol was carried out that provided eloquent insights into the structural basis of druggability potential of citE. Under physiological conditions, essentially all cellular processes are mediated within a hydrated environment. In this context, the next phase of molecular dynamics simulation facilitated the drug design approach by not only computationally placing the docked protein within solvated surroundings but also reflecting the time dependent behavior of the system. The biological importance of water is well known in relation to a range of biochemical process. Specifically, within the context of protein dynamics, several studies emphasize the importance of water in dictating the activity and the dynamic behavior of the protein particularly with reference to catalytic activity and functional inhibition (Suresh, et al., 2008). Analysis of protein, whilst in ligand bound form helped unravel the ligand induced variability and led to evaluation of organizational alterations of the protein-ligand system. The application of simulation to the *K. pneumoniae* citE, however, stipulated structural stability of system over the studied time scale.

Stability of the docked protein-ligand complex was explained by calculation of physical properties over a time period of 70 ns. The RMSD behavior of the inhibitor bound citE, over the studied time scale was mostly stable with an average value of 3.13 Å. The general pattern of RMSD represented a single extensive domain shift within the structural framework of the protein-ligand complex, where a β-hairpin transformed into a loop, which occurred at the 15^{th} ns, and persisted till the end of simulation. Overall, the ligand-bound citE showed a stable symmetry. This was accompanied by a well complemented ligand placement within the active site during the simulation. Also, the structural compactness of the ligand-bound citE, measured as radius of gyration (R_g) averaged at a value of 20.23 Å.

The RMSD plot was additionally analyzed for the fluctuations at the atomic levels, revealing the individual residues, being the cause of the alterations in the simulated protein-ligand complex. The average RMSF for the ligand-bound protein was observed to be 2.48 Å that helped in distinguishing the structurally flexible and rigid loci of the drug target. Higher fluctuations were observed for regions that formed loops and turns and were solvent exposed. In particular, residues forming the C-terminal region, ranging between residues 257-271 exhibited a transient behavior where the β-hairpin transformed into a loop region, connecting two helices. Amino acid residues that incur the fluctuations are away from the active site pocket of the protein. The active site pocket inclusive of residues 36-39, 95-98, 129-133 and 157-160, had an RMSF value of less than 2.0 Å signifying the stability during the production run. The pattern of beta-factor for citE was seen to be consistent with the RMSF trend. Changes in the RMSD and RMSF plots, thus illustrate the fact that the conformational deviations and alterations are in a way worthy for the protein-ligand complex. The unwavering active site residues ascertain the active site locus to remain stable throughout the simulation. All these changes however, are a requirement of the catalytic process and may depend on the nature of the protein whether it is enzymatic or not (Patel, Kumar and Durani, 2007).

In silico strategy adopted during the current project provided eloquent information at various stages of analysis. Collectively, the inferred knowledge about essential catalytic mechanisms, cofactor dynamics and effects of structural variability on inhibitor binding can be extended to increase efficacy of novel *K. pneumoniae* citE inhibitors in order to halt the lethal infections caused by the studied pathogen.

Conclusion

The current study highlights the route adopted for the identification of therapeutic candidates in multiple strains of the emerging Gram negative, MDR bacterial pathogen, *K. pneumoniae*. The subtractive approach has delivered findings of great pharmacological importance to aid in paving the way for development of putative drug candidates. The subtraction lead to a smaller subset of the functional genome, dispensing survival essential proteins for the pathogen. It was concurred that the potent drug target has no counterpart in human. The study characterized 6 putative druggable compounds common to genomes of three different strains of *K. pneumoniae*. CitE, the selected Mg^{+2} dependent metalloprotein affirms its pharmacological importance in antibacterial therapies targeting Gram negative bacteria. Thus, drugs can be developed against this pathogenic bacteria in order to block the infections it causes. Comparative homology modeling was applied to attain a high quality model for structurally uncharacterized citE. Molecular docking protocol classified compound 47 as the best potential inhibitory agent against citE. Insights from molecular docking and MD simulations led to the deduction that citE undergoes specific conformational changes explaining the dynamic behavior of the ligand-bound protein. Besides the side chain fluctuations and a sheet to loop transformation, stability of inhibitor and target protein complex was observed. The metal ion cofactor Mg^{+2} lying at the active site of citE, maintains its position and 4-fold coordination. Consequently, implying the molecular interactions in docked citE complex, it can be concluded that compounds mimicking the chemical structure of compound 47 can be employed to destabilize the target. It is concurred that molecular docking studies in combination with MD simulations can be useful in the discovery and development of more potent inhibitors and thus, specific and efficient drugs against MDR *K. pneumoniae*.

REFERENCES

References

Allen, M.P. and Tildesley, D.J., 1989. *Computer Simulation of Liquids*. Oxford university press.

Almirall, J., Bolibar, I., Balanzó, X. and Gonzalez, C.A., 1999. Risk factors for community-acquired pneumonia in adults: a population-based case–control study. *European Respiratory Journal*, *13*(2), pp.349-355.

Alonso, H., Bliznyuk, A.A. and Gready, J.E., 2006. Combining docking and molecular dynamic simulations in drug design. *Medicinal Research Reviews*, *26*(5), pp.531-568.

Aoshima, M., Ishii, M. and Igarashi, Y., 2004. A novel enzyme, citryl-CoA lyase, catalysing the second step of the citrate cleavage reaction in Hydrogenobacter thermophilus TK-6. *Molecular Microbiology*, *52*(3), pp.763-770.

Azam, S.S. and Shamim, A., 2014. An insight into the exploration of druggable genome of Streptococcus gordonii for the identification of novel therapeutic candidates. *Genomics*, *104*(3), pp.203-214.

Azam, S.S., Uddin, R. and Wadood, A., 2012. Structure and dynamics of alpha-glucosidase through molecular dynamics simulation studies. *Journal of Molecular Liquids*, *174*, pp.58-62.

Azam, S.S., Saroosh, A., Zaman, N. and Raza, S., 2013. Role of N-acetylserotonin O-methyltransferase in bipolar disorders and its dynamics. *Journal of Molecular Liquids*, *182*, pp.25-31.

Bakheet, T.M. and Doig, A.J., 2010. Properties and identification of antibiotic drug targets. *BMC Bioinformatics*, *11*(1), p.195.

Barh, D., Tiwari, S., Jain, N., Ali, A., Santos, A.R., Misra, A.N., Azevedo, V. and Kumar, A., 2011. In silico subtractive genomics for target identification in human bacterial pathogens. *Drug Development Research*, *72*(2), pp.162-177.

References

Berendsen, H.J., Postma, J.P.M., van Gunsteren, W.F., DiNola, A.R.H.J. and Haak, J.R., 1984. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, *81*(8), pp.3684-3690.

Berman, J.J., 2012. *Taxonomic guide to infectious diseases: understanding the biologic classes of pathogenic organisms*. Academic Press.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E., 2000. The protein data bank. *Nucleic Acids Research*, *28*(1), pp.235-242.

Binder, K., Horbach, J., Kob, W., Paul, W. and Varnik, F., 2004. Molecular dynamics simulations. *Journal of Physics: Condensed Matter*, *16*(5), p.S429.

Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M. and Bairoch, A., 2007. Uniprotkb/swiss-prot. In *Plant Bioinformatics* (pp. 89-112). Humana Press.

Butt, A.M., Tahir, S., Nasrullah, I., Idrees, M., Lu, J. and Tong, Y., 2012. Mycoplasma genitalium: a comparative genomics study of metabolic pathways for the identification of drug and vaccine targets. *Infection, Genetics and Evolution*, *12*(1), pp.53-62.

Centers for Disease control and Prevention, 2013. Antibiotic resistance threats in the United States, 2013. 2013.

Chemical Computing Group. 2013. Molecular Operating Environment (MOE), 2013.08. Chemical Computing Group Inc.

Chong, C.E., Lim, B.S., Nathan, S. and Mohamed, R., 2006. In silico analysis of Burkholderia pseudomallei genome sequence for potential drug targets. *In Silico Biology*, *6*(4), pp.341-346.

Colovos, C. and Yeates, T.O., 1993. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Science: a Publication of the Protein Society*, *2*(9), p.1511.

Eswar, N., Eramian, D., Webb, B., Shen, M.Y. and Sali, A., 2008. Protein structure modeling with MODELLER. In *Structural Proteomics* (pp. 145-159). Humana Press.

References

Forli, S., 2010. Raccoon| AutoDock VS: an automated tool for preparing AutoDock virtual screenings.

Garza-Fabre, M., Toscano-Pulido, G. and Rodriguez-Tello, E., 2012, July. Locality-based multi objectivization for the HP model of protein structure prediction. In *Proceedings of the 14th annual conference on Genetic and evolutionary computation* (pp. 473-480). ACM.

Genheden, S., 2012. Are homology models sufficiently good for free-energy simulations? *Journal of Chemical Information and Modeling*, *52*(11), pp.3013-3021.

Goodfellow, J.M., 1990. *Molecular Dynamics: Applications in Molecular Biology* (Vol. 16). CRC Press.

Goulding, C.W., Bowers, P.M., Segelke, B., Lekin, T., Kim, C.Y., Terwilliger, T.C. and Eisenberg, D., 2007. The structure and computational analysis of Mycobacterium tuberculosis protein CitE suggest a novel enzymatic function. *Journal of Molecular Biology*, *365*(2), pp.275-283.

Hansson, T., Oostenbrink, C. and van Gunsteren, W., 2002. Molecular dynamics simulations. *Current Opinion in Structural Biology*, *12*(2), pp.190-196.

Hassan, E.A., El-Rehim, A. and EL-Din, A.S., 2015. Creatinine modified Child-Turcotte-Pugh and integrated model of end stage liver disease scores as predictors of spontaneous bacterial peritonitis related in-hospital mortality: applicable or not. *Journal of Gastroenterology and Hepatology*.

Highsmith, A.K. and Jarvis, W.R., 1985. Klebsiella pneumoniae: selected virulence factors that contribute to pathogenicity. *Infection Control*, *6*(02), pp.75-77.

Hirsch, E.B. and Tam, V.H., 2010. Detection and treatment options for Klebsiella pneumoniae carbapenemases (KPCs): an emerging cause of multidrug-resistant infection. *Journal of Antimicrobial Chemotherapy*, p.dkq108.

Hong, L. and Lei, J., 2009. Scaling law for the radius of gyration of proteins and its dependence on hydrophobicity. *Journal of Polymer Science Part B: Polymer Physics*, *47*(2), pp.207-214.

References

Hosen, M.I., Tanmoy, A.M., Mahbuba, D.A., Salma, U., Nazim, M., Islam, M.T. and Akhteruzzaman, S., 2014. Application of a subtractive genomics approach for in silico identification and characterization of novel drug targets in Mycobacterium tuberculosis F11. *Interdisciplinary Sciences: Computational Life Sciences*, *6*(1), pp.48-56.

Hughes, J.P., Rees, S., Kalindjian, S.B. and Philpott, K.L., 2011. Principles of early drug discovery. *British Journal of Pharmacology*, *162*(6), pp.1239-1249.

Humphrey, W., Dalke, A. and Schulten, K., 1996. VMD: visual molecular dynamics. *Journal of Molecular Graphics*, *14*(1), pp.33-38.

Janda, J.M. and Abbott, S.L., 2008. The family enterobacteriaceae. *Practical Handbook of Microbiology, 2nd ed. Boca Raton: CRC Press*, *200*, pp.217-229.

Jones, G., Willett, P., Glen, R.C., Leach, A.R. and Taylor, R., 1997. Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology*, *267*(3), pp.727-748.

Kanehisa, M. and Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28*(1), pp.27-30.

Knowles, J. and Gromo, G., 2003. Target selection in drug discovery. *Nature Reviews Drug Discovery*, *2*(1), pp.63-69.

Ko, W.C., Paterson, D.L., Sagnimeni, A.J., Hansen, D.S., Von Gottberg, A., Mohapatra, S., Casellas, J.M., Goossens, H., Mulazimoglu, L., Trenholme, G. and Klugman, K.P., 2002. Community-acquired Klebsiella pneumoniae bacteremia: global differences in clinical patterns. *Emerging Infectious Diseases*, *8*(2), pp.160-166.

Krieger, E., Nabuurs, S.B. and Vriend, G., 2003. Homology modeling. *Methods of Biochemical Analysis*, *44*, pp.509-524.

Kuzmanic, A. and Zagrovic, B., 2010. Determination of ensemble-average pairwise root mean-square deviation from experimental B-factors. *Biophysical Journal*, *98*(5), pp.861-871.

References

Laskowski, R.A., Moss, D.S. and Thornton, J.M., 1993. Main-chain bond lengths and bond angles in protein structures. *Journal of Molecular Biology*, *231*(4), pp.1049-1067.

Law, V., Knox, C., Djoumbou, Y., Jewison, T., Guo, A.C., Liu, Y., Maciejewski, A., Arndt, D., Wilson, M., Neveu, V. and Tang, A., 2014. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Research*, *42*(D1), pp.D1091-D1097.

Li, W., Jaroszewski, L. and Godzik, A., 2001. Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics*, *17*(3), pp.282-283.

Li, Z., Wan, H., Shi, Y. and Ouyang, P., 2004. Personal experience with four kinds of chemical structure drawing software: review on ChemDraw, ChemWindow, ISIS/Draw, and ChemSketch. *Journal of Chemical Information and Computer Sciences*, *44*(5), pp.1886-1890.

Liu, D. ed., 2011. *Molecular Detection of Human Bacterial Pathogens*. CRC press.

Maiorov, V.N. and Crippen, G.M., 1994. Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. *Journal of Molecular Biology*, *235*(2), pp.625-634.

Mantzarlis, K., Makris, D., Manoulakas, E., Karvouniaris, M. and Zakynthinos, E., 2013. Risk factors for the first episode of Klebsiella pneumoniae resistant to carbapenems infection in critically ill patients: a prospective study. *BioMed Research International*, *2013*.

McCall, M., 2010. *Classical Mechanics: From Newton to Einstein: A Modern Introduction*. John Wiley & Sons.

McCammon, J.A., Gelin, B.R. and Karplus, M., 1977. Dynamics of folded proteins. *Nature*, *267*(5612), pp.585-590.

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C. and Kanehisa, M., 2007. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research*, *35*(suppl 2), pp.W182-W185.

References

Morris, G.M. and Lim-Wilby, M., 2008. Molecular docking. In *Molecular Modeling of Proteins* (pp. 365-382). Humana Press.

Morris, A.L., MacArthur, M.W., Hutchinson, E.G. and Thornton, J.M., 1992. Stereochemical quality of protein structure coordinates. *Proteins: Structure, Function, and Bioinformatics*, *12*(4), pp.345-364.

Mosca, A., Miragliotta, L., Del Prete, R., Tzakis, G., Dalfino, L., Bruno, F., Pagani, L., Migliavacca, R., Piazza, A. and Miragliotta, G., 2013. Rapid and sensitive detection of blaKPC gene in clinical isolates of Klebsiella pneumoniae by a molecular real-time assay. *SpringerPlus*, *2*(1), p.31.

Munikumar, M., Priyadarshini, V., Pradhan, D., Swargam, S. and Umamaheswari, A., 2013. 177 T-cell vaccine design for Streptococcus pneumoniae: an in silico approach. *Journal of Biomolecular Structure and Dynamics*, *31*(sup1), pp.114-115.

Nagano, N., Orengo, C.A. and Thornton, J.M., 2002. One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *Journal of Molecular Biology*, *321*(5), pp.741-765.

Nancy, Y.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S.C., Ester, M., Foster, L.J. and Brinkman, F.S., 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics*, *26*(13), pp.1608-1615.

Neema, M., Karunasagar, I. and Karunasagar, I., 2011. In silico identification and characterization of novel drug targets and outer membrane proteins in the fish pathogen Edwardsiella tarda. *Open Access Bioinformatics*, *3*, pp.37-42.

Nobre, S.R., Cabral, J.E.P., Gomes, J.J.F. and Leitao, M.C., 2008. In-hospital mortality in spontaneous bacterial peritonitis: a new predictive model. *Journal of Gastroenterology and Hepatology*, *20*(12), pp.1176-1181.

Nordmann, P., Cuzon, G. and Naas, T., 2009. The real threat of Klebsiella pneumoniae carbapenemase-producing bacteria. *The Lancet Infectious Diseases*, *9*(4), pp.228-236.

References

Parvege, M.M., Rahman, M. and Hossain, M.S., 2014. Genome-wide Analysis of Mycoplasma hominis for the Identification of Putative Therapeutic Targets. *Drug Target Insights*, *8*, p.51.

Patel, K., Kumar, A. and Durani, S., 2007. Analysis of the structural consensus of the zinc coordination centers of metalloprotein structures. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, *1774*(10), pp.1247-1253.

Paterson, D.L., Ko, W.C., Von Gottberg, A., Mohapatra, S., Casellas, J.M., Goossens, H., Mulazimoglu, L., Trenholme, G., Klugman, K.P., Bonomo, R.A. and Rice, L.B., 2004. Antibiotic therapy for Klebsiella pneumoniae bacteremia: implications of production of extended-spectrum β-lactamases. *Clinical Infectious Diseases*, *39*(1), pp.31-37.

Pendleton, J.N., Gorman, S.P. and Gilmore, B.F., 2013. Clinical relevance of the ESKAPE pathogens.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E., 2004. UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), pp.1605-1612.

Piddock, L.J., 2006. Clinically relevant chromosomally encoded multidrug resistance efflux pumps in bacteria. *Clinical Microbiology Reviews*, *19*(2), pp.382-402.

Pieper, U., Eswar, N., Braberg, H., Madhusudhan, M.S., Davis, F.P., Stuart, A.C., Mirkovic, N., Rossi, A., Marti-Renom, M.A., Fiser, A. and Webb, B., 2004. MODBASE, a database of annotated comparative protein structure models, and associated resources. *Nucleic Acids Research*, *32*(suppl 1), pp.D217-D222.

Pradhan, D., Priyadarshini, V., Munikumar, M., Swargam, S. and Umamaheswari, A., 2013. 161 Discovery of potent KdsA inhibitors of Leptospira interrogans through homology modeling, docking, and molecular dynamics simulations. *Journal of Biomolecular Structure and Dynamics*, *31*(sup1), pp.105-105.

References

Prajapati, C. and Bhagat, C., 2012. In-Silico Analysis and Homology Modeling Of Target Proteins for Clostridium Botulinum. *International Journal of Pharmaceutical Sciences and Drug Research*, *3*, pp.2050-2056.

Ramakrishnan, C. and Ramachandran, G.N., 1965. Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units. *Biophysical Journal*, *5*(6), pp.909-933.

Rathi, B., Sarangi, A.N. and Trivedi, N., 2009. Genome subtraction for novel target definition in Salmonella typhi. *Bioinformation*, *4*(4), p.143.

Reddy G, K. and Rao K, N., 2012. Identification of Drug and Vaccine Targets in Highly Mutant Clostridium botulinum B1 okra by Subtractive Genomics. *Research & Reviews: Journal of Genomics and Proteomics*, *1*(2).

Rodrigues, J.P.G.L.M., Melquiond, A.S.J., Karaca, E., Trellet, M., Dijk, M.V., Zundert, G.C.P., Schmitz, C., Vries, S.J., Bordogna, A., Bonati, L. and Kastritis, P.L., 2013. Defining the limits of homology modeling in information-driven protein docking. *Proteins: Structure, Function, and Bioinformatics*, *81*(12), pp.2119-2128.

Russ, A.P. and Lampel, S., 2005. The druggable genome: an update. *Drug Discovery Today*, *10*(23), pp.1607-1610.

Ryckaert, J.P., Ciccotti, G. and Berendsen, H.J., 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, *23*(3), pp.327-341.

Saenger, W. and Jeffrey, G.A., 1991. *Hydrogen Bonding in Biological Structures*. Springer-Verlag, Berlin.

Saleem, A.F., Qamar, F.N., Shahzad, H., Qadir, M. and Zaidi, A.K., 2013. Trends in antibiotic susceptibility and incidence of late-onset Klebsiella pneumoniae neonatal sepsis over a six-year period in a neonatal intensive care unit in Karachi, Pakistan. *International Journal of Infectious Diseases*, *17*(11), pp.e961-e965.

References

Salomon-Ferrer, R., Case, D.A. and Walker, R.C., 2013. An overview of the Amber biomolecular simulation package. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, *3*(2), pp.198-210.

Sánchez, R. and Šali, A., 2000. Comparative protein structure modeling: introduction and practical examples with MODELLER. In *Protein Structure Prediction* (pp. 97-129). Humana Press.

Sarangi, A.N., Aggarwal, R., Rahman, Q. and Trivedi, N., 2009. Subtractive genomics approach for in silico identification and characterization of novel drug targets in Neisseria Meningitides Serogroup B. *Journal of Computer Science & Systems Biology*, *2*(5), pp.255-258.

Schomburg, I., Chang, A. and Schomburg, D., 2002. BRENDA, enzyme data and metabolic information. *Nucleic Acids Research*, *30*(1), pp.47-49.

Schwede, T., Kopp, J., Guex, N. and Peitsch, M.C., 2003. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research*, *31*(13), pp.3381-3385.

Shafiq, M., Rahman, H., Qasim, M., Ayub, N., Hussain, S., Khan, J. and Naeem, M., 2013. Prevalence of plasmid-mediated AmpC β-lactamases in Escherichia coli and Klebsiella pneumoniae at tertiary care hospital of Islamabad, Pakistan. *European Journal of Microbiology and Immunology*, *3*(4), pp.267-271.

Shanmugam, A. and Natarajan, J., 2010. Computational genome analyses of metabolic enzymes in Mycobacterium leprae for drug target identification. *Bioinformation*, *4*(9), p.392.

Sharma, V., Gupta, P. and Dixit, A., 2008. In silico identification of putative drug targets from different metabolic pathways of Aeromonas hydrophila. *In Silico Biology*, *8*(3-4), pp.331-338.

Sippl, M.J., 1993. Recognition of errors in three-dimensional structures of proteins. *Proteins: Structure, Function, and Genetics*, *17*(4), pp.355-362.

Srinivasan, V.B., Vaidyanathan, V., Mondal, A. and Rajamohan, G., 2012. Role of the two component signal transduction system CpxAR in conferring cefepime and chloramphenicol resistance in Klebsiella pneumoniae NTUH-K2044. *PLoS ONE*, *7*(4), pp.e33777-e33777.

References

Suresh, C.H., Vargheese, A.M., Vijayalakshmi, K.P., Mohan, N. and Koga, N., 2008. Role of structural water molecule in HIV protease-inhibitor complexes: A QM/MM study. *Journal of Computational Chemistry*, *29*(11), pp.1840-1849.

Thomsen, R. and Christensen, M.H., 2006. MolDock: a new technique for high-accuracy molecular docking. *Journal of Medicinal Chemistry*, *49*(11), pp.3315-3321.

Trott, O. and Olson, A.J., 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, *31*(2), pp.455-461.

Tsai, Y.K., Liou, C.H., Fung, C.P., Lin, J.C. and Siu, L.K., 2013. Single or in combination antimicrobial resistance mechanisms of Klebsiella pneumoniae contribute to varied susceptibility to different carbapenems. *PLoS ONE*, *8(11)*, p.e79640.

Tuckerman, M., 2010. *Statistical Mechanics: Theory and Molecular Simulation*. Oxford University Press.

Uddin, R. and Saeed, K., 2014. Identification and characterization of potential drug targets by subtractive genome analyses of methicillin resistant Staphylococcus aureus. *Computational Biology and Chemistry*, *48*, pp.55-63.

Vaught, A., 1996. Graphing with Gnuplot and Xmgr: two graphing packages available under linux. *Linux Journal*, 28, p.7.

Visualizer, D.S., 2012. Release 3.5. *Accelrys Inc, San Diego, CA, USA*.

Volkamer, A., Kuhn, D., Rippmann, F. and Rarey, M., 2012. DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics*, *28*(15), pp.2074-2075.

Wallace, A.C., Laskowski, R.A. and Thornton, J.M., 1995. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Engineering Design and Selection*, *8*(2), pp.127-134.

References

Weiner, P.K. and Kollman, P.A., 1981. AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions. *Journal of Computational Chemistry*, *2*(3), pp.287-303.

Wereszczynski, J. and McCammon, J.A., 2012. Statistical mechanics and molecular dynamics in evaluating thermodynamic properties of biomolecular recognition. *Quarterly Reviews of Biophysics*, *45*(01), pp.1-25.

Wiederstein, M. and Sippl, M.J., 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, *35*(suppl 2), pp.W407-W410.

Wilde, R.E. and Singh, S., 1998. *Statistical Mechanics: Fundamentals and Modern Applications*. Wiley-Interscience.

World Health Organization, 2014. *Antimicrobial Resistance: Global Report on Surveillance*. World Health Organization.

Xu, Z.Q., Flavin, M.T. and Flavin, J., 2014. Combating multidrug-resistant Gram-negative bacterial infections. *Expert Opinion on Investigational Drugs*, *23*(2), pp.163-182.

Yu, C.S., Chen, Y.C., Lu, C.H. and Hwang, J.K., 2006. Prediction of protein subcellular localization. *Proteins: Structure, Function, and Bioinformatics*, *64*(3), pp.643-651.

Zarzycki, J. and Kerfeld, C.A., 2013. The crystal structures of the tri-functional Chloroflexus aurantiacus and bi-functional Rhodobacter sphaeroides malyl-CoA lyases and comparison with CitE-like superfamily enzymes and malate synthases. *BMC Structural Biology*, *13*(1), p.28.

Zhang, R. and Lin, Y., 2009. DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes. *Nucleic Acids Research*, *37*(suppl 1), pp.D455-D458.

Zhang, Y., 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, *9*(1), p.40.

Zhang, R., Ou, H.Y. and Zhang, C.T., 2004. DEG: a database of essential genes. *Nucleic Acids Research*, *32*(suppl 1), pp.D271-D272.

Thesis

ORIGINALITY REPORT

6%	2%	5%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Azam, Syed Sikander, and Amen Shamim. "An insight into the exploration of druggable genome of Streptococcus gordonii for the identification of novel therapeutic candidates", Genomics, 2014. Publication	1%
2	Azam, Syed Sikander, Sara Sarfaraz, and Asma Abro. "Comparative modeling and virtual screening for the identification of novel inhibitors for myo-inositol-1-phosphate synthase", Molecular Biology Reports, 2014. Publication	<1%
3	www.mswg.org.my Internet Source	<1%
4	Michael E. Budiman. "Using molecular dynamics to map interaction networks in an aminoacyl-tRNA synthetase", Proteins Structure Function and Bioinformatics, 08/15/2007 Publication	<1%