Characterization of Druggable Genome and Identification of Putative Drug Targets in Multi-drug Resistant *Cronobacter Sakazakii*

By

Fozia Shaheen

National Center for Bioinformatics

Faculty of Biological Sciences

Quaid-i-Azam University

Islamabad, Pakistan

2016

Characterization of Druggable Genome and Identification of Putative Drug Targets in Multi-drug Resistant *Cronobacter Sakazakii*

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Philosophy in Bioinformatics at Quaid-i-Azam University, Islamabad.

By

Fozia Shaheen

National Center for Bioinformatics

Faculty of Biological Sciences

Quaid-i-Azam University

Islamabad, Pakistan

2016

DEDICATED TO MY PARENTS

DECLARATION

The work reported in this thesis was carried out by Fozia Shaheen and I hereby declare that the title of thesis, **"*Characterization of druggable genome and identification of putative drug targets in multidrug resistant Cronobacter Sakazakii*"** and the contents of thesis are product of my own research and no part has been copied from any published source (except the references, standard mathematical or genetic models /equations /formulas /protocols etc). I further declare that this work has not been submitted for award of any other degree /diploma. The University may take action if the information provided is found inaccurate at any stage.

_____

Signature of Scholar

List of Abbreviations

Assisted Model Building with Energy Refinement	AMBER
Basic Local Alignment Search Tool	BLAST
Cluster Database at High Identity with Tolerance	CD-HIT
Database of Essential Genes	DEG
Discovery Studio Visualizer	DS Visualizer
Enzyme Commission	EC
Expectation-Value	E-Value
Expert Protein Analysis System	ExPASy
General Amber force field	GAFF
Genetic Optimization for Ligand Docking	GOLD
Iterative TASSER	I-TASSER
KEGG Automatic Annotation Server	KAAS
Kyoto Encyclopedia of Genes and Genome	KEGG

List of abbreviations

Molecular dynamics	MD
Nanoseconds	ns
Picoseconds	ps
Process TRAJectory	PTRAJ
Protein Data Bank	PDB
Protein Structure Analysis-web	ProSA-web
Root Mean Square Deviation	RMSD
Root Mean Square Fluctuation	RMSF
Simulated Annealing with NMR Derived Energy Restraints	SANDER
Three-Point Transferable Intermolecular Potential	TIP3P
Tripos Force Field	TFF
Universal Protein Resource Knowledgebase	UniProtKB
Visual Molecular Dynamics	VMD

List of Figures

List of Figures

List of Tables

ACKNOWLEDGEMENT

In the name of Allah, the most bounteous and the most merciful, all praises to Allah for strength and His blessing in the accomplishment of this thesis. No words can express enough gratitude for knowledge and wisdom He bestow upon us as minute reflection of His limitless 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. He has been everything that one could want in a supervisor. I could not have imagined having a better advisor and mentor for my MPhil study. I am indebted and thankful to him.

I would like to acknowledge Chairperson Professor Dr. Wasim Ahmad for the academic support

I am indebted to Computational biology group members Samra Wajid Abbasi, Saad Raza, Iqra Ahmed, Amen Shamim, Asma Abro, Sundus Iqbal, and Nur ul Ain Sajid.for providing a stimulating and constructive environment.

Last but not least, I would like to pay high regards to my parents, family and my husband Arif Hussain for their sincere encouragement and unconditional love throughout my research work and lifting me uphill this phase of life. I owe everything to them

ABSTRACT

Cronobacter Sakazakii is a life threatening multi drug resistant, Gram-negative, human opportunistic and food borne pathogen primarily involved in life threatening meningitis with a mortality rate of 40 to 80% and necrotizing enterocolitis with mortality rate of 10 to 55%, septicemia and bacteremia. Misuse of therapeutic drugs and self-medication along with selection pressure have granted resistivity to pathogen which fuels towards identification of newer and effective drugs. Aim of the study is the identification and characterization of potential drug targets in *Cronobacter sakazakii* via hierarchical *in silico* subtractive genomics approach. Qualitative characterization of druggable targets was done by applying various filters for identification of a potential drug target candidate in multiple strains of *Cronobacter sakazakii* that gives a common target: cell division protein (FtsZ). It is involved in cell cycle-*caulobacter* pathway which is one of the most promising candidates for drug development. Unavailability of experimentally determined structure homology modeling approach was used to model structure via MODELLER and other web servers. Best modeled structure was selected and used for molecular docking studies. A total of 102 potential inhibitors library with reported activities against FtsZ protein was prepared. The best affinity with the FtsZ was displayed by Compound 78 with GOLD Score of 86. Moreover, time dependent dynamic behavior of docked complex was analyzed using MD simulation studies. MD trajectory analysis revealed the flexibility of loop region to stabilize the binding of ligand and target protein and hydrogen bonding pattern was also rearranged. These conformational changes suggested the potential of 2‴-hydroxy-5″benzylisouvarinol-B to act as lead compound.

1. INTRODUCTION

Scientific world is waging a war against antibacterial agents and bacterial resistance since discovery of penicillin. Antibiotic resistance is a grave threat to global health which demands intense need to grapple this huge and complex problem. Antibiotic resistance has reached to an alarming point (WHO, 2014) Antibiotic resistance is due to deploying them wrongly which enables bacteria to boost resistance, ultimately depleting the effective stock of antibiotics. Selection pressure is also a leading contributor to emergence of resistant bacterial strains. Morbidity and mortality rates are increased due to the widespread of multiple drug resistant (MDR) strains, which poses a serious threat to public health sector. The conventional methods are also ineffective due to formidable challenges created by MDR bacterial infections. Therefore, antibacterial drug development needs to be continued so that novel and effective antibiotics can be developed to control serious health menace caused by multi-drug resistance organisms worldwide. In the current study subtractive genomics and *in silico* proteome analysis was carried out on multiple strains of emerging pathogen *Cronobacter sakazakii* to identify candidate drug targets.

1.1 The Genus Cronobacter.

Cronobacter is the officially known bacterial genus name for the organism which before 2007 was named as *Enterobacter sakazakii. Cronobacter* has been accepted as a genus in International Journal of Systematic and Evolutionary Microbiology the genus has six genomospecies and the classification of these organisms was revised based on a detailed polyphasic taxonomic study. *Cronobacter sakazakii* is named after the Japanese bacteriologist Richii Sakazakii (Iverson, et al., 2008).

1.2 Cronobacter Sakazakii

Cronobacter sakazakii belongs to *Enterobacteriaceae* family. It is formerly known as yellow pigmented *Enterobacter sakazakii* , is a Gram-negative bacillus, rod shaped, peritrichous, non-spore forming, human opportunistic and food borne pathogen. The primary reservoir of *Cronobacter sakazakii* has yet to be determined but it is postulated that plant material may be an important source. *Cronobacter sakazakii* can be isolated from a wide variety of foods including milk, cheese, dried foods, meats, water, vegetables, rice, bread, tea, herbs, spices, and powdered infant formula. Scrutiny studies have spotted

Cronobacter sakazakii in households, livestock facilities, food factories, and PIF production facilities (Yan and Fanning, 2015). Clinically, *Cronobacter sakazakii* has been isolated from cerebrospinal fluid, bone marrow, blood, intestinal and respiratory tracts, urine, ear and eye swabs, and skin wounds (Healy, et al., 2010).

1.2.1 Genomic Features

The complete genomes of three strains of *Cronobacter sakazakii* were published recently in 2013 which is available on National Center for Biotechnology Information (NCBI) (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Detail of genomes of all three strains (ES15, CMCC45402, and SP291) is presented in Table 1.1.

Table 1.1 Genomic features of ES15, SP291and CMCC45402.

Strains	Size(Mb)	Protein	GC%
ES15	4.26867	3807	57.10%
SP291	4.51878	4066	56.82%
CMCC45402	4.55994	4129	56.83%

1.2.2 Pathogenicity and Mortality rate

The high resistance of *Cronobacter sakazakii* to unusually dry conditions supports a high survival rate in powdered infant formula. *Cronobacter sakazakii* is having a greater ability of stress response due to its adaptive physiological traits. (Zhao, et al., 2014).*Cronobacter sakazakii* has viscous capsular and film barrier formation that assist in its adherence to all cell types and resist desiccation for a long interval of time in harsh conditions. (Iversen, et al., 2004). In many studies the significant contribution of OmpA to the virulence potential of Cronobacter Sakazakii is revealed. (Mohan, et al., 2009).*Cronobacter sakazakii* has been recognized as a critical pathogenic bacterium, especially for powdered formula-fed infants, due to high (up to 80%) mortality rate (Drudy, et al., 2006),primarily in newborns (Bar-Oz, et al, 2001).

Due to its pathogenicity and virulence, it mainly causes meningitis with a mortality rate of 40 to 80% and necrotizing enterocolitis with mortality rate of 10 to 55%, septicemia and bacteremia (Forsythe SJ, 2005). *Cronobacter Sakazakii* infections are also associated with significant morbidity. Those children who sustain *Cronobacter Sakazakii* related meningitis (94%) develop impaired sight and hearing, developmental impedance and irreparable, wide variety of sequelae such as quadriplegia. These sequelae are frequently attributed to secondary cerebral infarcts. (Drudy, et al., 2006)

FDA has spotted the presence of *Cronobacter Sakazakii* in baby formula in 2002. (CDC, 2003). *Cronobacter Sakazakii* infections are worrisome in infant population. The mortality rate in infected infants due to *Cronobacter Sakazakii* is 40%–80 %.(Bowen, et al., 2006). However cases are also reported in immunocompromised adults as well. The colonies of *Cronobacter Sakazakii* are also isolated from the mouths of stroke victims suffering with pneumonia (Gosney, et al., 2006)

1.2.3 Multi drug resistance

Different antibiotics are employed to control this pathogen but recent studies investigated emergence of antibiotic-resistant strains (Zhou, et al., 2011).*Cronobacter Sakazakii* is normally resistant to all macrolids including Fosfomycin, Fusidicacid, Rifampicin, Streptogramins, Clindamycin, Leucomycin, Tetracycline and Aminoglycosides. (Stock and Wiedemann, 2002).*Cronobacter Sakazakii* is resistant to Ampicillin due to acquisition of transposable elements and production of beta-lactamase and Cephalosporinase which inactivate third generation cephalosporin and wide spectrum of penicillin (Pitout, et al., 1997).

Computer aided drug designing provide alternative and cost effective approaches for addressing the problems of resistive infections. These approaches are proving to be very proficient in identification of pathogen drug targets and then development of targeted and robust drugs for those targets. The idea of interpretation and analysis of large amount of available proteome sequence data through various *in silico* approaches has revolutionized the drug discovery process.

1.3 *In silico* Approach in Current study

The current study includes different *in silico* approaches which can be integrated into various steps. These steps follow a specific sequential route: subtractive genomics approach, selection of therapeutic target, homology modeling, molecular docking and molecular dynamic simulations as shown in Figure 1.1.

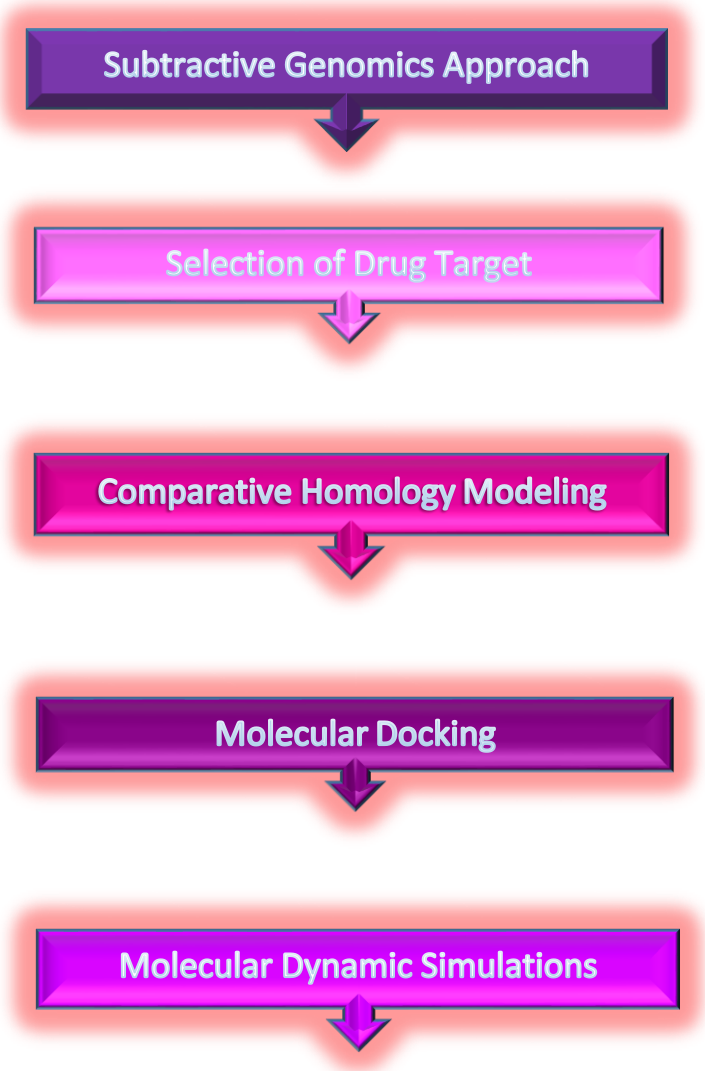


Figure 1.1 In silico steps adopted in the current study

1.3.1 Subtractive Genomics Approach

The conventional approaches of drug development result in dead ends or host toxicity which can be replaced by applying the in silico subtractive genomics approach to design potential drugs against bacterial diseases. Subtractive genomics basically deals with identification of novel targets, essential for pathogen survival and important component of various metabolic pathways .these targets should not be in homology with the host so that a drug or lead compound will be designed against these targets. This approach is fast, robust cost effective. The speed of drug discovery process is increased due to rapid increase in availability of genome sequences of various pathogens. This approach has successfully applied in identification of drug targets in pathogens like *Mycobacterium tuberculosis* (Hosen, et al., 2014), *Leptospira Interrogans* (Pradhan, et al., 2013), *Streptococcus pneumoniae* (Munikumar, et al., 2013*), Clostridium botulinum* (Prajapati and Bhagat, 2012*) Mycoplasma Genitalium* (Butt, et al., 2012), and *Salmonella Typhi* (Rathi, et al., 2009). *Mycobacterium Leprae* (Shanmugam and Natarajan, 2010), and *Mycoplasma Pneumonia* (Gupta, et al., 2010).

In the current study subtractive genomics approach has been utilized for multiple strains of *Cronobacter Sakazakii*: ES15, SP296, and CMCC45402 to identify putative drug targets. This will facilitate not only in understanding the biology of the infectious pathogen but will also help in developing a more effective medication. Different pipelines and filters are employed in the approach to find out important targets imperative for pathogen survival which could be easily targeted in human host due to lack of homology of these targets in human host. The designated drugs should effect the given target without adverse effect in host. Hence optimal targets are obtained in this way which provide selectivity. Thus, optimal target obtained from screening of a set of essential proteins are not only crucial for the viability of the organism, but also provide selectivity.

The drug target identified and characterized in the current study is bacterial cell division protein (FtsZ).FtsZ is selected as it is involved in cell division of bacterial cell which can be targeted to inhibit the growth of bacteria specifically in human host. FtsZ is an important protein of cell division cycle in almost all bacterial strains known yet.

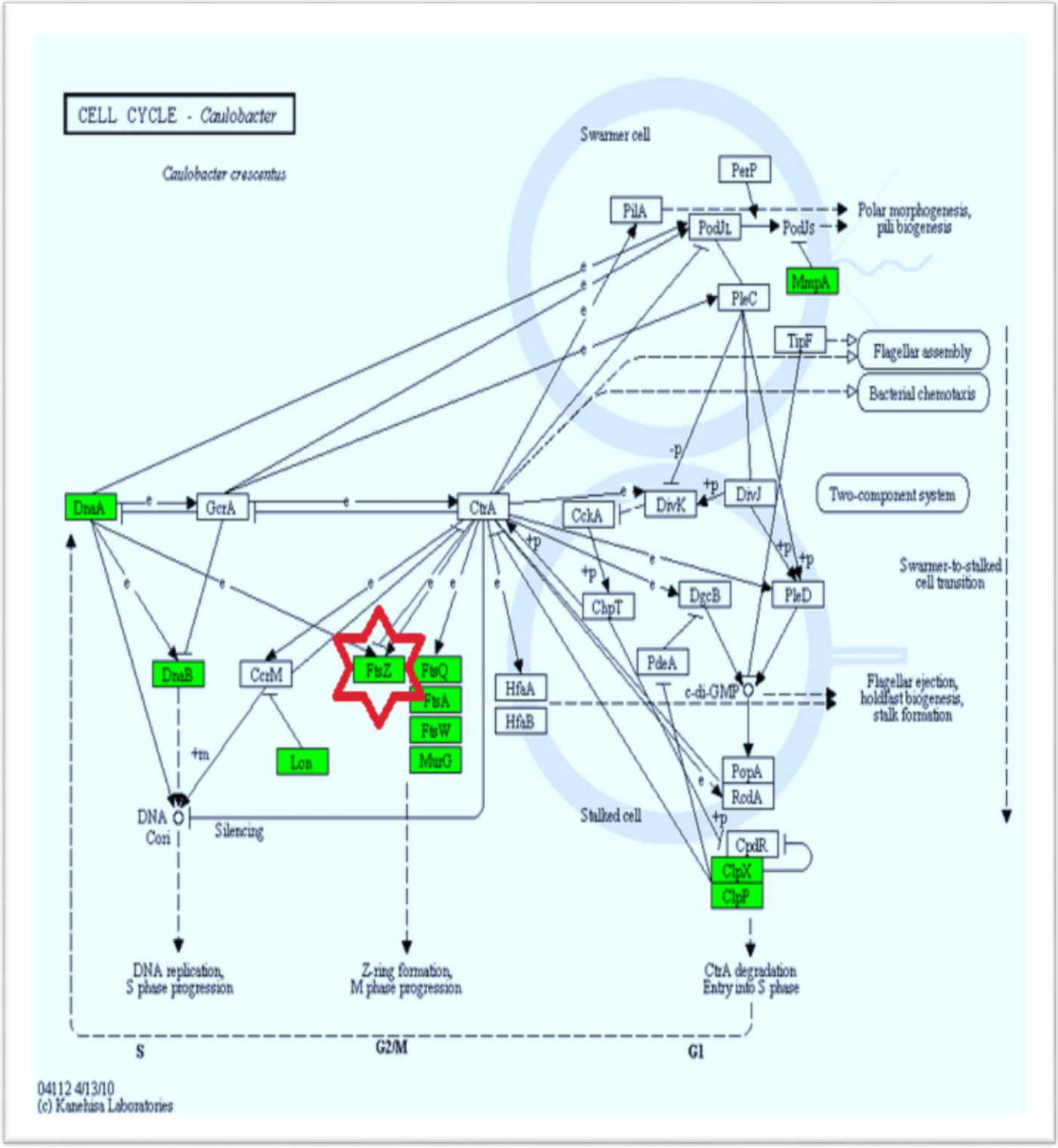


Figure 1. 1. Cell cycle-caulobacter pathway of human pathogen *Cronobacter sakazakii* showing durggable target FtsZ (EC: 2.2.1.6)

1.3.2 Bacterial cell division protein (FtsZ)

FtsZ is an important cell division protein which makes a contractile ring at the bacterial cell division site. The ring assembly regulation controls the timing and location of cell division. The function of FtsZ is to recruit the other proteins involved in cell division to produce a new cell wall between the dividing cells. Subtractive genomic approach predicted FtsZ as an ideal drug target for novel antibacterial drugs that is common in all three strains of *Cronobacter sakazakii.* It is found to be conserved among various bacterial species .FtsZ has a weak homology with human tubulins (Salimnia, et al., 2000). FtsZ taken from one bacterial specie can work better in other. The functional properties and its interaction with other proteins enables it as a potential therapeutic drug target, to find lead molecule with enhanced efficacy and reduced toxicity. The approximate molecular weight of FtsZ is 50kDa.

1.3.3 Homology modelling

Since last few decades there is an enormous increase in the sequence information at both, genome and proteome level, due to availability of high throughput sequencing techniques and prompt annotation of data (Khafizov, et al., 2014).Biological knowledge is more easily accessible due to large number of interlinked data repositories. However the problem arises in understanding biological data due to lack of structural information regarding the biomolecules. As of May 2014, Universal Protein Resource Knowledgebase (UniProtKB) (Boutet, et al., 2007) has 56555610 protein sequences reported in May 2014 while in Protein Data Bank (PDB) there are 79,356 experimental protein structures and this number is increasing exponentially (Westbrook, et al., 2003).Comparative homology modelling is a valuable tool to close the gap between sequence and structural space (Moult, et al., 2014).Hence a theoretical model can be generated if a known template structure is available with a suitable similarity is present . (Krieger, et al., 2003). The structure of target protein FtsZ from *Cronobacter sakazakii* is unavailable in the Protein Databank (PDB). Therefore, protein modeling was performed using MODELLER and various other web-based tools including SWISS-MODEL (Schwede, et al., 2003), ModWeb (Pieper, et al., 2004), I-TASSER (Zhang, 2008), and 3D-JIGSAW (Bates, et al., 2001). A whole comparison is also between MODELLER and other web servers is also carried out. PROCHECK

(Laskowski, et al., 1993), ProSA (Sippl, 1993), Verify3D (Eisenberg, et al., 1997), ERRAT value (Colovos and Yeates, 1993), G-Factor and Bad Contacts (Morris, et al., 1992) were measured and the best model was selected after evaluating all generated models from the above sources.

1.3.4 Molecular docking

Molecular docking is a powerful approach for structure based drug discovery. Molecular docking is an important tool in CADD (computer aided drug design) and structural molecular biology which is employed to predict the binding modes of three dimensional structure of target protein with ligand. (Morris, et al., 2008).The current study incorporates molecular docking to predict the inhibition mechanism of target protein. The information about the inhibitors and the active site of the target macromolecule has been obtained after an in depth study of literature.

Molecular docking comprises of two interrelated steps. First generating all possible binding configurations of ligand with receptor molecule through search algorithms and secondly prioritizing the binding modes to predict the most feasible and accurate binding conformation through scoring function (Taylor, et al., 2002). MD simulations are utilized in the current study to further explore the exact binding conformations and time dependent behavior of protein-ligand complex.

1.3.5 Molecular dynamic simulations

In late 1950's Alder and Wainwright introduced MD methods which were employed in study of interaction of hard spheres (Alder and Wainwright, 1959). Later in 1964 Rahman carried out molecular dynamics simulation study for liquid argon utilizing Lennard-john potential. The first protein which was simulated was bovine pancreatic trypsin inhibitor (BPTI) in 1977 (McCammon, et al, 1977).Molecular dynamics simulation is a principal tool in the theoretical study of biological molecules, using simple approximations based on Newtonian physics. It calculates time dependent behavior of a system.MD provide detailed information about conformational changes and fluctuations of nucleic acid and proteins.

MD method has two major families which can be categorized on the basis of the model selected to describe a physical system. The classical mechanic approach in MD simulations defines the dynamics of the system while quantum approach in MD simulations deals with

the nature of chemical bond. Quantum MD simulations are more powerful than classical mechanics approach as it deals with diverse biological problem. Due to computational complexity of quantum approach, classical mechanics is practical at present for simulations of biological systems containing many thousands of atoms over a time scale of nanoseconds. (Jarosaw, et al., 2015). Over the past MD has emerged as a major area in understanding the atomic basis of complex biological systems such as transport of small molecules and ions across membranes, protein folding and molecular recognition. Along with experimental approaches, MD simulation has a better understanding of protein structure-function relationship and manifest the drug discovery process. (Pramod, et al., 2014).

1.3.5.1 Statistical Mechanics – Foundation of Theoretical Framework.

Molecular dynamics simulations produce data at the microscopic level of a system, comprising atomic positions and velocities. The transformation of this microscopic data to macroscopic information such as heat capacities, pressure, energy etc., utilizes statistical mechanics. Hence the study of molecular dynamic of biological systems require statistical mechanics. (Wereszczynski and McCammon, 2012).Statistical mechanics makes a connection between microscopic simulation and macroscopic properties of system under study. The collection of all possible systems which have identical thermodynamic state but different microscopic state constitute an ensemble. (Wilde and Singh, 1998).

Four major ensembles are present in statistical mechanics each with varying properties. (Chandler, 1987).

- **Micro canonical ensemble (NVE)** – constant number of particles N, constant volume V and constant amount of energy E.
- **Grand canonical ensemble (μVT) -** constant chemical potential μ, constant volume V and constant value of temperature T.
- **Canonical ensemble (NVT)** - constant number of particles N, constant volume V and constant value of temperature T.
- **Isobaric-isothermal ensemble (NPT)** – constant number of particles N, constant pressure P and constant value of temperature T.

1.3.5.2 Classical Mechanics – Principle of Motion

Newton second law of motion proposed by Sir Isaac Newton is foundation of classical molecular dynamics. The acceleration of each atom in a given system can be predicted from the information of force applied on each atom. A trajectory is produced through integration of equations of motion which describe positions, velocities and acceleration of particles at different time intervals. The trajectory gives knowledge about average values of these properties .The essence of this method is deterministic as if the positions and velocities of each atom are known, the state of system in past or future can be easily detected .Solvated proteins can be simulated up to nanosecond scale however the regime of simulation up to millisecond is also reported. (Petrenko and Meller, 2010).

Newton's equation of motion at ith particle can be given as:

$$F_i = m_i a_i \qquad \textbf{(1.1)}$$

Where F*i* is force exerted on the particle, m*i* is the mass of the particle and a*i* denotes the acceleration produced. The acceleration is a second derivative of distance "d" and time "t". The above equation can be hence written as

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

Change in potential energy V of mass m of system can be calculated as the distance covered by the application of force F

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

Combining equation 1.2 and 1.3 yields

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

Therefore acceleration "a" of particle can be defined as time dependent behavior of velocity "v" at time "t":

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

1.3.5.3 Molecular Mechanics – Calculation of Molecular Structure

To compensate the need of description of molecular structure and properties practically, molecular mechanics was developed, which is an extension of classical mechanics. The limit of applicability of molecular mechanics depend on following factors.

- Molecules comprising thousands of atoms.
- and saccharides , peptides, Organics, oligonucleotides
- Vacuum, explicit, or implicit solvent environments.
- Ground state only.
- Thermodynamic and kinetic (via molecular dynamics) properties.

The powerful computational speed of molecular mechanics makes it useful in molecular dynamics, docking and conformational energy searching.

Molecular mechanics is based on following principles.

- Nuclei and electrons are united into atom-like particles.
- Atom-like particles have a net charge and spherical
- Interactions are based on classical potentials and springs.
- Interactions should be allotted to specific sets of atoms.
- Interactions calculate the spatial distribution of atom-like particles and their energies.

Force field of a molecular system can be determined by the sum of individual energy terms

$$E = E_{\text{covalent}} + E_{\text{noncovalent}}$$

Where the components of the covalent and noncovalent contributions are given by the following summations:

$$E_{\text{covalent}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}}$$

$$E_{\text{noncovalent}} = E_{\text{electrostatic}} + E_{\text{van der Waals}}$$

1.4 Aims and objective

The current study highlighted the different computational approaches to explore the druggable genome of human pathogen *Cronobacter sakazakii* strains, for possible identification of potential drug target. *In silico* subtractive genomics approach is applied to explore the complete genome of *Cronobacter's* three strains (ES15, CMCC45402, SP291) then common novel drug target was identified, which in turn will be effective against all three strains. *Cronobacter sakazakii* is the most prevalent strain involved in meningitis with high mortality rate. So this strain was selected for structure based drug designing process. Comparative modeling approach yielded the structural coordinates of target, which was common to three strains. Molecular docking and MD simulation illuminated the binding of different putative inhibitors, and dynamics of proposed target with candidate inhibitor, respectively. This study aims at finding the potential drug targets along with potent inhibitors against meningitis and simulation study explore the dynamics of docked target .This inhibitor binding process will help to develop more powerful inhibitor against pathogen which will also deal with resistance evolving ability of pathogen.

2. METHODOLOGY

2.1 System information

Computational structure comprises of both hardware and software components guarantee smooth execution of the research activities. For major portion of the current work, computational facilities provided at the Computational Biology Lab, National Center for Bioinformatics, Quaid-i-Azam University, Islamabad, Pakistan, were utilized. The entire procedure was divided into four stage; *in silico* subtractive genomics approach, homology modeling, molecular docking analysis and molecular dynamics simulation. Complete work was done using Intel (R) Core(TM) 2 Duo CPU E8600 @ 3.33 GHZ and the operating system used was Linux open SUSE 11.4. The production run of 100 nanoseconds for docked protein was carried out on high performance computer cluster in above mentioned lab given in Figure 2.1.

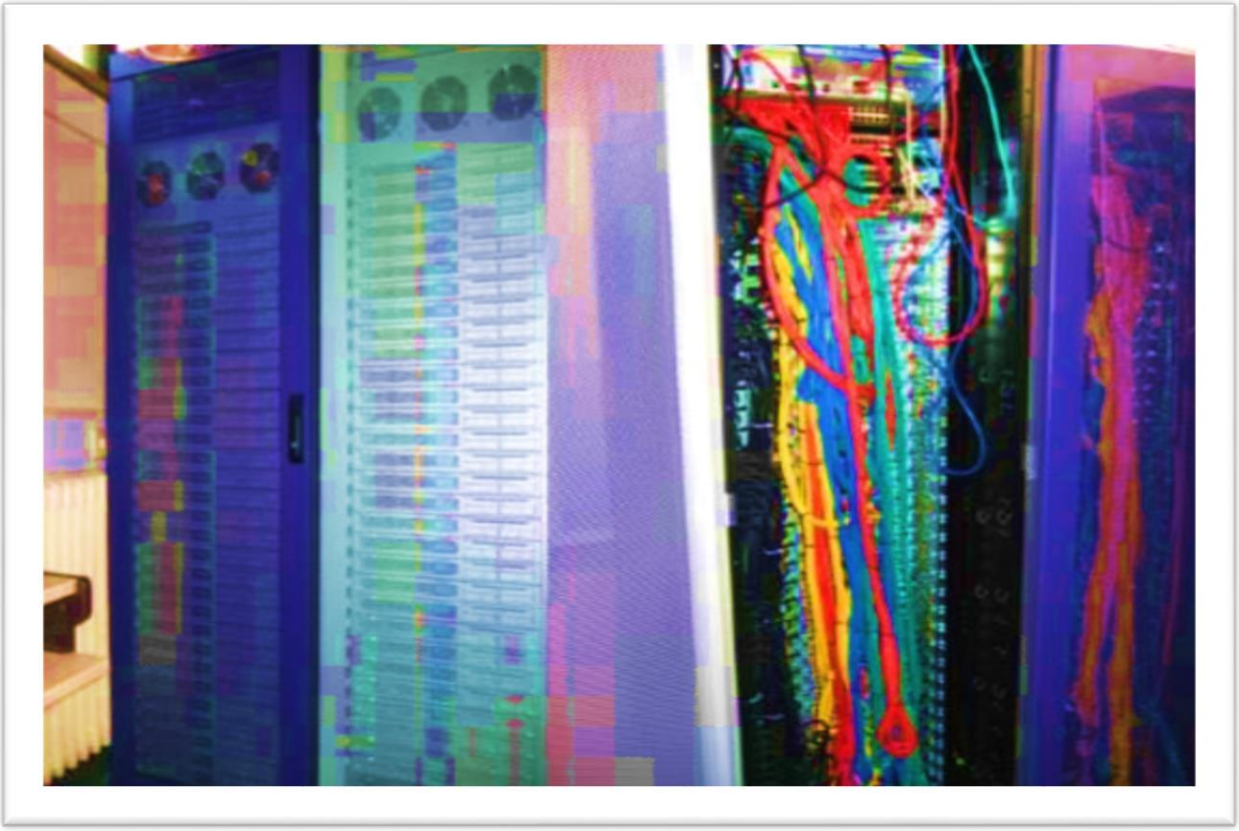


Figure 2. 1. Cluster system used for computational studies

2.2 Applied *in silico* Approach

The present strategy follow an *in silico* protocol which is partitioned into different modules, which used different soft wares and databases. At first genomes of multiple strains of *Cronobacter sakazakii* (ES15, CMCC45402, and SP291) were compared using a technique of subtractive genomics which directed to identify the pathogenic target. In second step homology modeling was done to anticipate the structure of potent target. Molecular docking was done to study system interactions with inhibitors which drove towards structure based drug designing. Finally MD simulations were performed to study time dependent behavior of target. The overall workflow of this study is shown in flowchart which given in Figure 2.2.

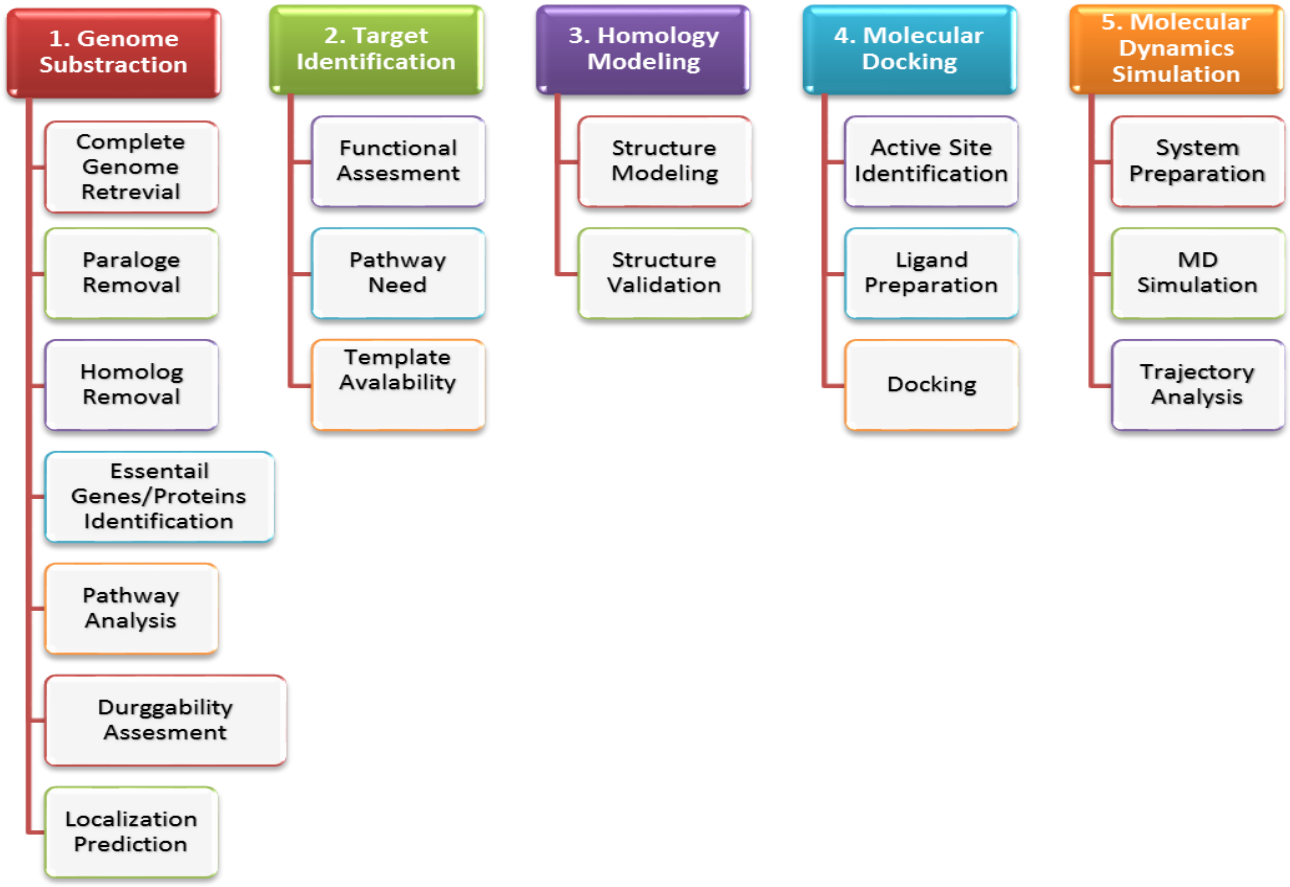


Figure 2. 2. In silico protocol followed in the current study

2.2.1 Subtractive Genomics Approach

Subtractive genomics approach is a standout amongst the most currently used methodologies, which is used in this study to predict the essential proteins in genome of *Cronobacter sakazakii*. This methodology involved the steps given in Figure 2.3.

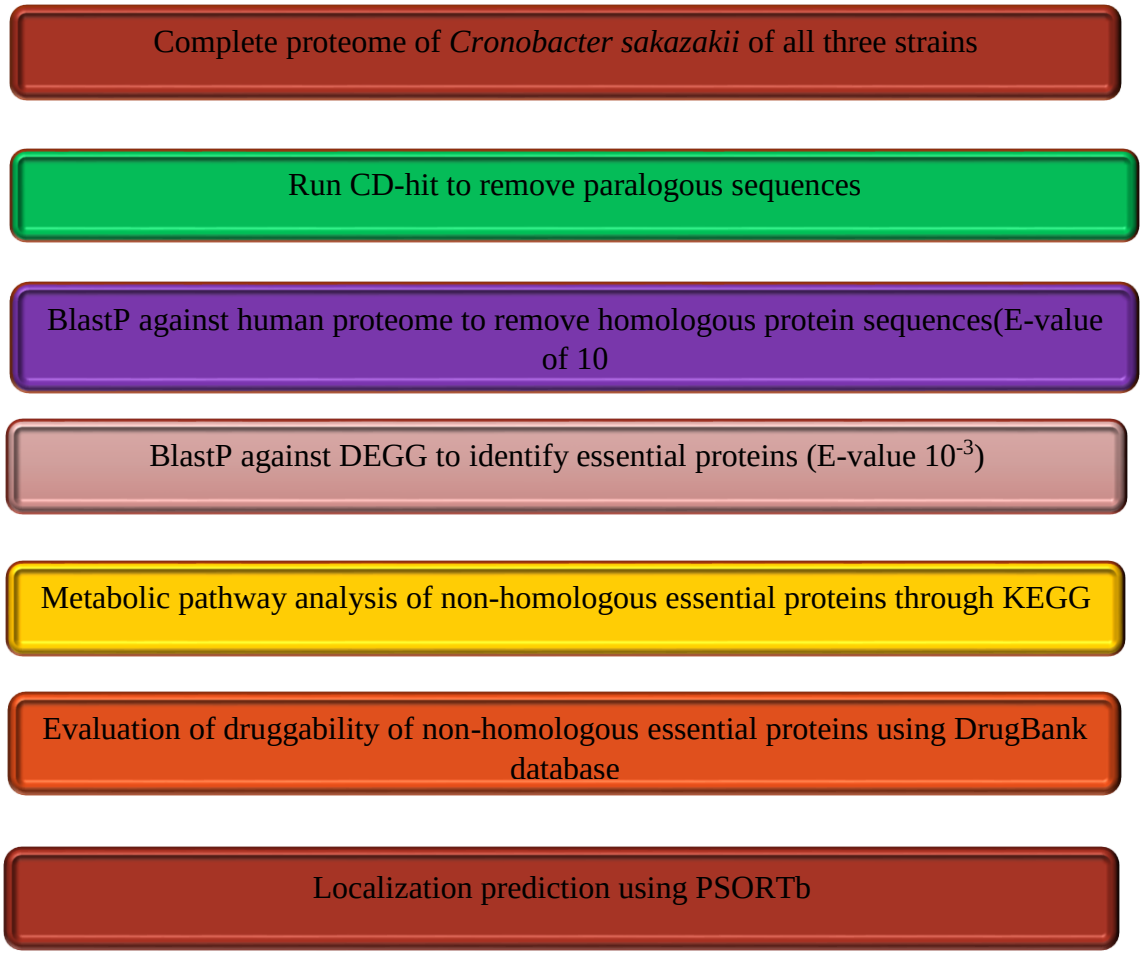


Figure 2.3. Flow chart depicting steps of subtractive genomics.

2.2.1.1 Complete proteome retrieval and paralog removal

Complete proteome of multiple strains: ES15, CMCC45402, and SP296 of *Cronobacter sakazakii* were retrieved from Uniprot (http://www.uniprot.org/) in fasta format. Uniprot consists of manually curated protein sequences under UniprotKB/Swiss-Prot and automated curated protein sequences under UniprotKB/TrEMBL (Boutet, et al., 2007). Web server CD HIT Suite (Cluster Database at High Identity with Tolerance) was used (Li, et al., 2001) to remove the redundancy of proteins from *Cronobacter sakazakii* strains. The cutoff value was set as 0.6 (60%) along with clustering and alignment parameters set as default to remove paralogous proteins. Finally non redundant sequences were obtained for further processing.

2.2.1.2 Retrieving of non-homologous protein

Removal of homologous sequences was done using Perl script provided at Computational Biology Lab of National Centre for Bioinformatics, Quaid-i-Azam University Islamabad, and Pakistan. In the initial step whole proteome of the pathogen was aligned against human proteome using Basic Local Alignment Search Tool (BLASTp) at threshold expectation value of (E- value) of 10^{-4} (Altschul, et al., 1990). Refseq NCBI database (Pruitt, et al., 2007) was locally remade on a server; these databases were then subjected to automated Blast search using Blast+ (Camacho, et al., 2013) and indigenously made Perl script, which resulted in the removal of homologs and compilation of non-human homologs for further analysis. The Perl script takes less time for investigating and filtering of the sequences in multiple strains. It makes the work fast. Manually it takes 2 weeks to a month for one genome. In correlation to this, scripts make the work finished in a couple of hours.

2.2.1.3 Identification of essential genes

Database of Essential Genes (DEG) (http://tubic.tju.edu.cn/deg/) (Zhang, et al., 2004) was used to outline set of essential *Cronobacter sakazakii* , non-human homologs. Essential gene filtering is an essential step since it is not much important that all homologous proteins obtained are also crucial to support life. Database of Essential Genes (DEG) (Zhang, et al., 2004) is repository that are integral to the life form survival, containing essentiality information for 43 representative eukaryotic and prokaryotic organisms (Zhang and Lin,

2009). Prokaryotic section of the database was taken as reference and BLASTp was done with E-value cut off at 10^{-4}, bit score = 100 and sequence identity of >= 30 %. *Cronobacter sakazakii* proteins. All parameters were set to the Perl script. Resultant list of fundamental proteins for all three strains were obtained by utilizing perl script.

2.2.1.4 Metabolic pathway analysis

The next step of screening procedure is to classify essential non-homologous proteins functionally and assigning them metabolic pathways using the KAAS (KEGG Automatic Annotation Server) (http://www.genome.jp/kegg/kaas/), which is supported by Kyoto Encyclopedia of Genes and Genomes (KEGG) database (http://www.genome.jp/kegg/). KEGG is conspicuously known for the assortment of information collection it holds among others, including genome information and functional annotation. KAAS specifically is a server that allots KEGG orthology (KO) numbers to the proteins by applying BLASTp to search against the KEGG GENES repository. The sequences under study are assigned metabolic pathways on basis of KO numbers (Moriya, et al., 2007).

The essential non homologous proteins of *Cronobacter sakazakii* genome obtained after search against DEG, were in this way submitted to KAAS utilizing *Cronobacter sakazakii* genome as reference data set to expand specificity of results. Since just a part of the complete pathogen proteome was being assessed, single-directional best hit system was used. The predicted pathogenic pathways were manually coordinated to the human pathways to arrange them into two classifications: common and unique. Common pathways were those present in both the host and bacterium. Though, unique, were selective to bacterium and thus, center of current work.

2.2.1.5 Druggability Assessment using DrugBank

The DrugBank database is a unique bioinformatics and cheminformatics asset that joins elaborated drug (i.e. compound, pharmacological and pharmaceutical) information with far reaching medication target (i.e. sequence, structure, and pathway) data available on http://www.drugbank.ca/ .Next step in genome subtraction was screening of chosen target proteins by means of Drug Bank. DrugBank version 3.0 was employed with default parameters to assess the druggability capability of non-homologous essential proteins by performing BLASTp against all the essential protein targets of drug compounds within

Drug Bank (Knox, et al., 2011). Drug bank contain 1634 FDA approved small molecule drugs, 89 nutraceuticals, 169 FDA approved biotech drugs and more than 6000 experimental drugs.

2.2.1.6 Subcellular Localization Prediction

In the last step the retrieved sequences were examined to recognize their subcellular location of the unique, and essential protein targets. For this purpose PSORTb 3.0.2 (http://www.psort.org/psortb/) and subcellular localization predictor (CELLO) version 2.5 (http://cello.life.nctu.edu.tw/) were utilized.

PSORTb a pioneer exertion, keeps up a database of localization data on an extensive variety of bacterial and archaeal species. It applies Support Vector Machines (SVMs), a machine learning method that searches the curated dataset to foresee the localization through the utilization of suffix tree algorithm. Bayesian classifier is employed in last step allocates a probable localization site to a protein (Nancy, et al., 2010). Additionally, it allocates the five subcellular areas i.e. cytoplasm, inner membrane, periplasm, outer membrane an extracellular, to gram-negative microorganisms (Yu, et al., 2010). Basically cytoplasmic proteins are selected as they have potential of becoming possible drug targets (Shoukat, et al., 2012). Similar to PSORTb, CELLO also utilizes the functionality of SVMs to predict the probability of the subcellular location (Yu, et al., 2006).

2.2.2 Selection of Druggable Target

Druggable essential proteins which were important in unique and essential metabolic pathways amongst different strains of *Cronobacter sakazakii* were identified and those which were common in all three were further used for analysis. At last, FtsZ playing a vital role in the cell cycle caulobacter pathway were selected for further in *silico* analysis. UniprotKB was used to retrieve information regarding the functional role, active isoforms, cofactors, subunit structure and associated post translational modifications. Furthermore, the presence of experimental structure was established by predicting the details about the target and in case of absence of such structure, the possibility of comparative homology based modeling was considered by assessing the template accessibility. Structural templates that showed at least 30% identity with > 90% query coverage were considered satisfactory. This assessment was done by comparing protein sequence against the

structural resource: PDB, using BLASTp functionality supported by NCBI. All these steps were carried out to select current target for further CADD analysis.

2.2.3 Model Construction, Refinement, and Validation

Comparative modelling was done to retrieve a homology based structure as the three dimensional crystallographic was not available for the selected target. The amino acid sequence for Cronobacter Sakazakii CMCC45402- FtsZ (Uniprot ID: V5U3P9) was taken from UniprotKB and compared to the PDB via alignment through BLASTp. The X-ray crystallographic structure for GDP bound FtsZ from Pseudomonas aeruginosa GDP (PDB ID:2VAW) (Olivia, et al., 2007), with sequence identity of 57% and query coverage of 94% was fulfilling the criteria of both the query coverage and sequence identity was selected for model building procedure. MODELLER (Eswar, et al., 2008) version 9.10 was used to build the structural model of target protein In addition to MODELLER9.10, structure for the FtsZ, was also obtained for comparative purposes through five web servers: SWISS-MODEL (Schwede, et al., 2003), Mod Web (Pieper, et al., 2004), I-TASSER (Zhang, et al., 2008), 3D-JIGSAW (Bates, et al., 2001) and ESyPred3D (Lambert, et al., 2002).

Validation is required to check the reliability of model generated. Models were evaluated using PDBSum (Laskowski, 2001), ERRAT value (Colovos and Yeates, 1993), Verify3D (Eisenberg, et al., 1997), ProSA (Wiederstein and Sippl, 2007), G-factor and bad contacts (Morris, et al., 1992). All these measures were used to make selection of the model. The results obtained after evaluation were used to select the best model. Functional basis of these structure assessment tools is elaborated further.

2.2.3.1 MODELLER

MODELLER is used to retrieve three-dimensional comparative structure of protein through homology or comparative modelling .The sequence is aligned with known related structure are provided as input and MODELLER generates a model containing all non-hydrogen atoms. Modeller also perform energy structure comparison, function calculations, and minimization.

2.2.3.2 SWISSMODEL

SWISSMODEL is an automated server. It is used for homology based protein modeling. 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 .It is used for the predictions of protein structures and functions. It can be accessed from http://zhanglab.ccmb.med.umich.edu/I-TASSER/. Three dimensional structures are generated on the basis of multiple-threading alignments by LOMETS. Function is predicted by matching 3D structure with Bio Lip protein function database. It can be accessed from: http://zhanglab.ccmb.med.umich.edu/I-TASSER/.

2.2.3.4 MODWEB

Mod Web is an online server .It is mainly used for comparative protein structure modeling. It is available on URL:https://modbase.compbio.ucsf.edu/scgi/modweb.cgi.

2.2.3.5 3D-JIGSAW

The 3D-JIGSAW is an automated server. It uses comparative modelling for prediction of structure and function against given protein sequence. It can be accessed from: http://bmm.cancerresearchuk.org/~3djigsaw/.

2.2.3.6 PROCHECK

A PROCHECK calculates stereo chemical property of a protein models like Ramachandran plot, Bad contacts and G-Factor. It estimates the energy of each residue of the overall model and results are shown in the form of a graph.

2.2.3.7 Verify3D

Verify3D is used in model evaluation to refine the structure. Each residue is allotted a structural preference based on its location and environment.

2.2.3.8 ERRAT

ERRAT is used in evaluation and refining the structure of model. It statistically inspect the non-bonded interactions among different atom types. Output is given in form of a plot of error function versus to position of residue.

2.2.3.9 ProSA-web

ProSA-web is an online tool used for validation of protein structure. The overall quality of protein structure is estimated on the basis of z-score. Thus, it is also capable to validate the low resolution structures.

2.2.4 Energy Minimization

The best model of target protein was then subjected to energy minimization process to improve its quality. UCSF Chimera (Pettersen, et al., 2004) ws used to minimize energy. Gasteiger charges were assigned to protein and structural constraints was removed by 1500 rounds of minimization runs (750 steepest descent followed by 750 conjugate gradient) with a step size of 0.02 Å, under Tripos Force Field (TFF) (Pettersen, et al., 2004).The minimized protein model thus obtained was evaluated through validation process. Finally the best minimized model was further taken for docking studies.

2.2.5 Molecular Docking Protocol

Docking protocol is divided into following steps: Active site identification, Inhibitors/ligands preparation and Molecular docking. Detailed methodology of these steps is given in following sections

2.2.5.1 Active Site Identification

Active site determination was carried out for the modeled protein to further work on its docking studies. Active site determination creates an idea to build a grid before docking. Active site of FtsZ was identified through extensive literature search and an online tool DoGSiteScorer. It predicts binding pocket on the basis of druggability (Volkamer, et al., 2012) and accessed at http://dogsite.zbh.uni-hamburg.de/, was used for further validation. Active site mentioned in literature was then predicted in our target sequence manually.

Active site residue were also confirmed by multiple sequence alignment (Edgar and Batzoglou, 2006) using ClustalO 2.0(Sievers, et al., 2011).

2.2.5.2 Ligand Preparation

A total 102 potential inhibitors with reported activities against FtsZ protein of several bacteria were identified from thorough literature search (Reynolds, et al., 2004; Paradis, et al., 2008; Jennings, et al; 2006; Catherine, et al., 2006; Haydon, et al., 2010; Jarosiaw, et al., 2010; Shanmugham, et al., 2011; Alberto, et al., 2010; Keffer, et al., 2013; Domadia, 2007; Haydan, et al., 2010; Awasthi, et al., 2011; Wang, et al., 2003; Kanemori, et al., 2014; Huang, et al., 2006; Waldemar, 2006; Beuria, et al., 2005; Lucile, et al., 2002; David, et al., 2009; Jaiswal, et al., 2007).The 2D chemical structures of these inhibitors were drawn using ChemDraw Ultra 8.0 in ChemOffice 2012 packages (Li, et al., 2004) which were further used in docking procedure. Standard bond angles, lengths and charges were used for structure of all inhibitors. The 2D structures drawn were further minimized using Chem3D Pro 12.0 using MM2 force field.

2.2.5.3 Molecular Docking

Molecular docking was carried out using the minimized protein and minimized ligand molecules. The docking software utilized are Genetic Optimization for Ligand Docking (GOLD) (Jones, et al., 1997) and AutoDock Vina (Trott and Olson, 2010). Gold Score and binding affinities were calculated from each ligand respectively .Hence on the basis of Gold Score best docked ligands were selected. LIGPLOT (Wallace, et al., 1995), Visual Molecular Dynamics (VMD) (Humphrey, et al., 1996), UCSF Chimera (Pettersen, et al., 2004), and Discovery Studio (DS) Visualizer 3.5 (Accelrys Software Inc., 2012), software were used for visualization and interaction studies of ligands in detail.

2.2.5.4 GOLD

GOLD Hermes package is a comprehensive docking setup wizard used in docking of flexible ligand with a partial flexible protein (Jones, et al., 1995).The standard default parameters were kept for docking poses of complexes such as population size is 100, niche size is 2, selection pressure is 1.1, operator weights for migrate is 0, crossover is 100,

number of islands is 1, number of operations is 10,000, number of dockings is 10. Hydrogen atoms were added in the protein model.

GOLD uses genetic algorithm and it implements four user defined scoring functions, namely, GoldScore, ChemScore, CHEMPLP and User Defined Score. GoldScore was used in the following study, which include energy terms describing the van der Waals interaction, hydrogen bonds and torsional energy (Verdonk, et al., 2003). Gold Score function is given mathematically as follow:

$$\textbf{Scoring function} = S_{hb\_ext} + S_{vdw\_ext} + S_{hb\_int} + S_{vdw\_int} + S_{tor} \qquad \textbf{(2.1)}$$

Where

S_{hb_ext} represents: protein ligand hydrogen bond score

S_{vdw_ext} represents: protein ligand van der Waals score

S_{hb_int} represents: fitness because of intramolecular hydrogen bond

S_{vdw_int} represents: intramolecular, ligand van der Waals energy

S_{tor} represents: ligands torsional energy

Gold- Score fitness function was used as criteria to select the best conformation in current study

2.2.5.5 Autodock Vina

AutoDock Vina is used for molecular docking which implements Broyden-Fletcher-Goldfarb-Shanno (BFGS) method for local optimization (Trott and Olson, 2010).. Docking poses are generated by AutoDock Vina. AutoDock (ADT tools 1.5.4 was used to generate the Pdbqt of the protein. The polar hydrogen atoms were added initially followed by addition of Kollman and Gasteiger charges. The file was saved as "pdbqt" file format. Graphical interface of Raccoon was used to generate Ligand PDBQT files (Forli, 2010). A grid box with a number of points of 20 Å × 20 Å × 20 Å box in x, y, z dimensions was created to enclose the ligand. Grid centered at x, y, z dimensions, was selected with values of 23.558, 34.790 and 10.381 respectively, with 1 Å box spacing. The grid box parameters were defined for protein and ligand in configuration file .During the docking procedure both ligand and protein were kept rigid. Ligands having lowest binding affinity were selected to make poses with receptor which were taken in further study.

2.2.6 Molecular Dynamics Simulation

Molecular dynamics simulation study was carried out to analyze the dynamic behavior of docked protein. Simulation was carried out by using Simulated Annealing with NMR-Derived Energy Restraints (SANDER) module in AMBER(Assisted Model Building with Energy Refinement) software (Case, et al., 2005). Docked protein simulation was carried out on Intel (R) Core(TM) 2 Duo CPU E8600 @ 3.33 GHZ with Linux open SUSE 11.4 as operating system. Simulation studies comprises of following five steps illustrated in Figure 2.4.

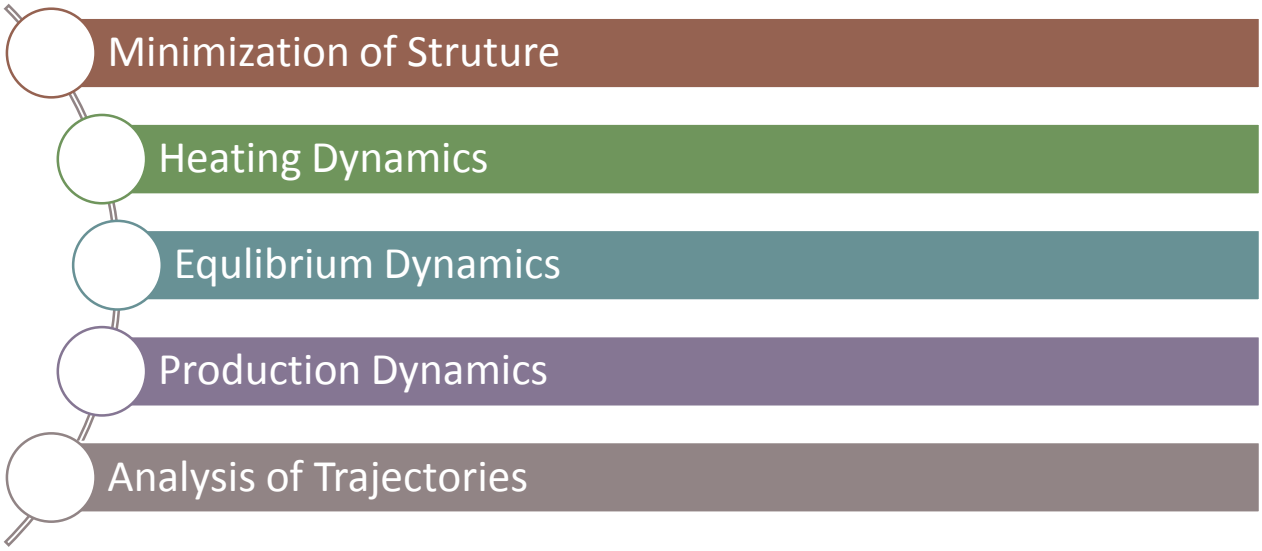


Figure 2.4. Steps of molecular dynamics simulations.

2.2.6.1 System preparation

Preprocessing which include minimization is done for system preparation before production run to remove all the steric clashes, equilibrating and heating the system. It involve four different steps: I) minimization II) heating III) equilibration IV) production run. The macromolecular system is prepared for simulation using xLEaP which is a graphical user interface to the LEaP module of AMBER 12 (Case, et al., 2012; Wang, et al., 2004). The Leap module in Amber12 tools was used to record the topology of the ligand and protein (Case, et al., 2005). Force fields used for protein and ligands was ff03.r1, GAFF and ff99SB (Salomon-Ferrer, et al., 2013) The system was neutralized by addition of 21 Na+ ions and solvated with three-point transferable intermolecular potential (TIP3P) water

box with 8.0 Å as shown in Figure 2.5. Antechamber program was employed for docked protein complex to make sure the correctness of bonds, angles and atom types.

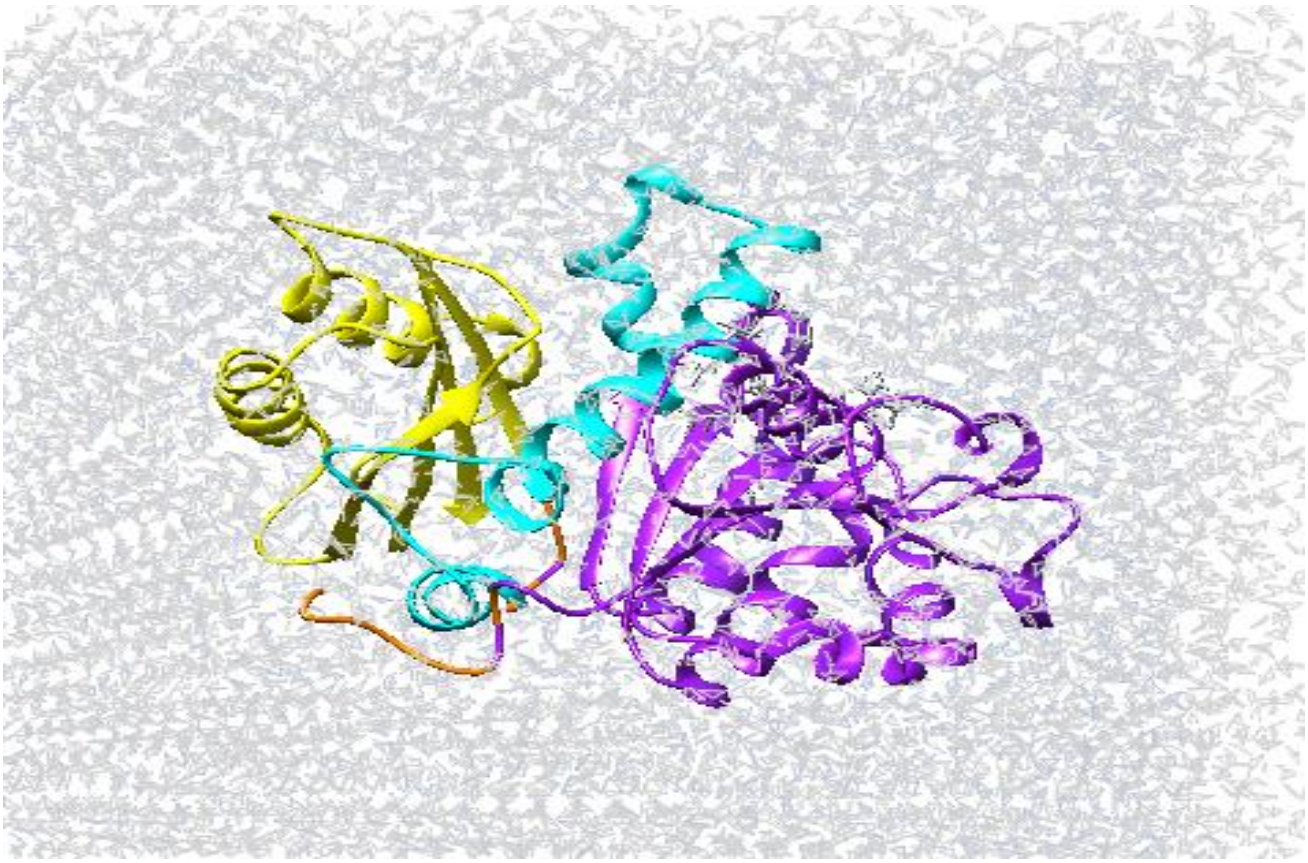


Figure 2.5 Solvation box surrounding docked protein.

2.2.6.2 Minimization, Heating, Equilibration and Production

After the system preparation minimization of complex was done thoroughly to remove unfavorable clashes before production run of MD simulation. 1500 steps for steepest descent method followed by 1000 steps for conjugate gradient were applied at non bonded cutoff value of 8 Å. Langevin dynamics algorithm (Feller, et al., 1995) was applied and heating of the system was performed for 10 picoseconds starting slowly from 0 k to a temperature of 300 k and pressure of 1 atm. Equilibration was done at a constant temperature of 300 for 10 picoseconds. Following equilibration production run of 100 ns

for the docked complex was done. Total energy of the system remains constant throughout the equilibration except kinetic and potential energy.

Production run was done finally by employing SHAKE algorithm (Ryckaert, et al., 1977) for bond constraints. SANDER (Simulated Annealing with NMR Derived Energy Restraints) module was used for production run .following production run, simulation run of 100ns was done. In the simulation box, periodic boundary conditions were applied with canonical ensemble. Berendsen coupling integration algorithm was used to keep temperature constant and consequently, MD simulations were achieved by means of Ewald summation method (Darden, et al., 1993).

2.2.6.3 Trajectory Analysis

For analysis of trajectory created after simulation procedure PTRAJ (Process TRAJectory) module of AMBER10 was used to calculate four different properties. In current study following four properties were calculated using PTRAJ and graphical representation were visualized in Xmgrace for analytical purpose (Vaught, 1996).

I) Root mean square deviation (RMSD)

II) Root mean square fluctuation (RMSF)

III) Radius of gyration (Rg)

IV) Beta factor (B-Factor)

2.2.6.4 Root Mean Square Deviation

The analogy in three-dimensional structure is measured for globular protein conformations by the RMSD of the Cα atomic coordinates after superposition of optimal rigid body. Root Mean Square Deviation (RMSD) is a measure of distance between two atoms of Cα, which are superimposed over each other for a specific frame.

$$RMSD = \sqrt{\frac{1}{N}\Sigma_i \ d_i^2}$$

In the above given equation, N signify the number of atom while di signify the distance between ith pairs of atom.

2.2.6.5 Root Mean Square Fluctuation

Root Mean Square Fluctuation (RMSF) is a measure of averaged root mean square Distance between the positions of atom from its mean position. RMSF calculates the local structural changes along the protein chain. In case of molecular dynamics, it can be defined as the set of positions for an atom achieved over a given time period. Following equation is used to calculate RMSF:

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

Where

T: represent the time interval.

x_i :represent the position of an atom at a particular time .

x :represent the averaged position of the atom.

2.2.6.6 Beta Factor

Beta factor is a term closely related to RMSF employed for dynamic studies. Local thermal and vibrational movements in the system create spatial displacement of atoms around their mean position .RMSF is a measure of these spatial movements of atoms around the reference position. (Kuzmanic and Zagrovic, 2010).following equation shows the Beta Factor measured in terms of RMSF.

$$\beta\ Factor = RMSF^2 (\frac{8\pi^2}{3}$$

2.2.6.7 Radius of Gyration

Radius of gyration is a measure the compactness of the macromolecular system including overall packing quality and density of a structure (Britton, et al., 1990). It can be calculated by the following equation:

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

Where "*N*" signify the total number of atoms, "m_i" signify the mass of "i^{th}"atom, "r_i" signify the position vector of "i^{th}" atom, and "r_{cm}" signify the center of mass of the molecule under consideration.

3. RESULTS

Methodological design of drug discovery process, generated different screened out results at the end of each step. A descriptive view of results accomplished over the span of current work is given in the following sections.

3.1 Genome Subtraction

The streamlining of the functional *Cronobacter sakazakii* genome is split into a few stages, expected to frame a list of the most conceivable drug targets. A diagram of this progressive screening strategy is given in Figure 3.1.

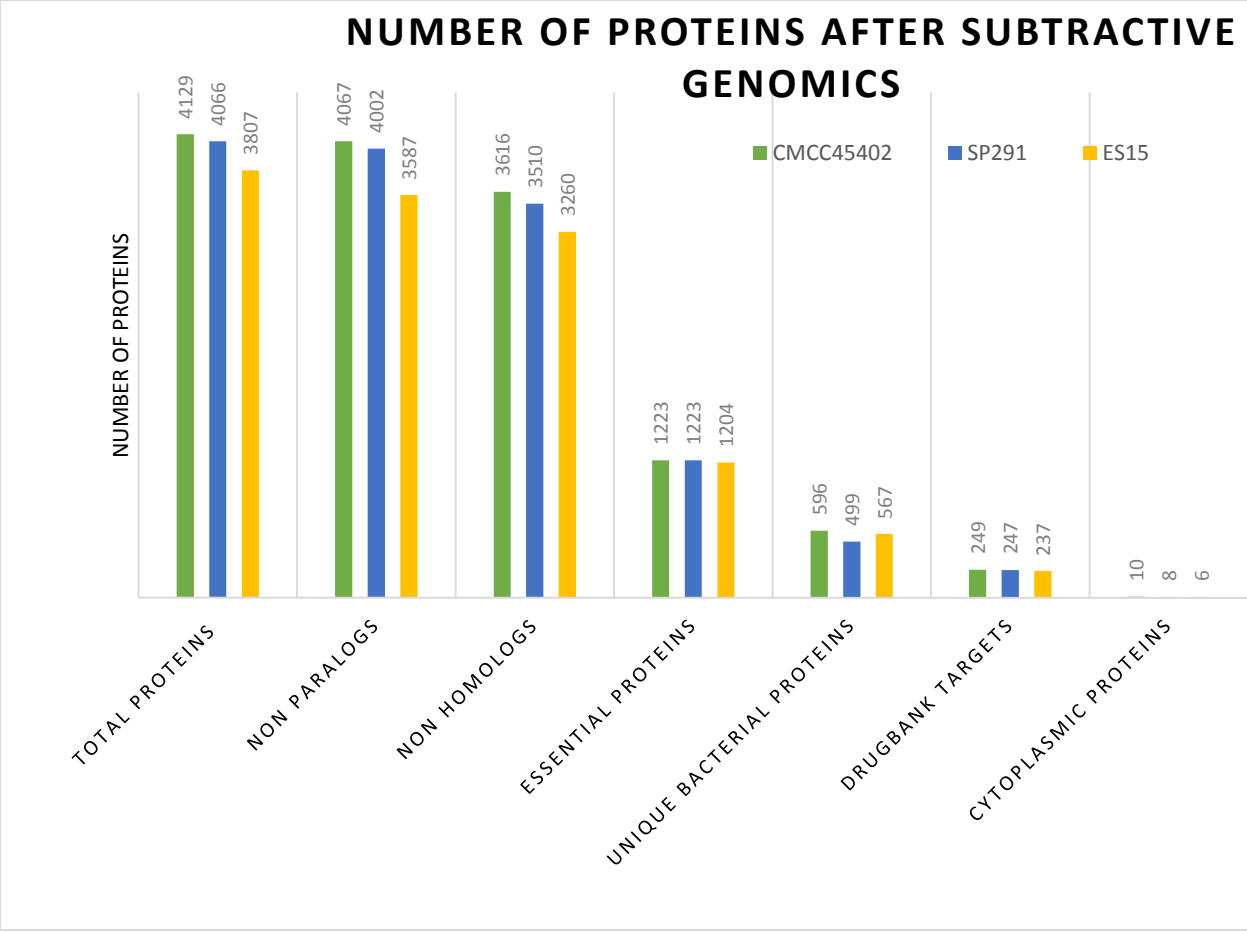


Figure 3. 1. Overview of screened proteins obtain at the end of each subtractive genomic steps.

3.1.1 Genome Retrieval

Three strains of *Cronobacter sakazakii* namely, CMCC45402, SP291 and ES15 were selected for current subtractive genomic study. These are completely sequenced strains and their genome was retrieved from UniprotKB comprising of 4349, 4805, and 4706 number of protein sequences respectively.

3.1.2 Non-Paralogous Proteins

Redundant protein sequences were removed via CD-HIT program at 60% identity leaving 4254, 4688, and 4623 non-paralogous protein sequences in three strains respectively. This led to the prediction of non-paralogous proteins i.e. 4067 in CMCC45402, 4002 in SP291 and 3587 in ES15 were subjected for further subtraction analysis.

3.1.3 Non-Homologous Proteins

These non-paralogous protein sequences were then subjected to Perl script for BLASTp to eliminate homologous protein sequences to Human homologous proteins were removed from proteome of CMCC45402, SP291 and ES15 strains. Finally 3616, 3510 and 3260 proteins were left in three strains, respectively.

3.1.4 Pathogen Essential Proteins

Pathogen essential genes are imperative for survival of pathogen which are predicted by search via DEG database. After this screening, essential genes identified were 1223 in CMCC45402, 1223 in SP291, 1204 in ES15 respectively.

3.1.5 Metabolic Pathway Analysis

Metabolic pathways of the essential proteins were investigated through KEGG Automatic Annotation Sever (KAAS) leading to the identification of potential drug targets involved in various crucial metabolic pathways of the pathogen. After accomplishing the KEGG, number of sequences for CMCC45402, SP291, and ES15 were reduced to 596, 499, and 567 respectively.

3.1.6 Druggability Assessment

To explore the relation between screened target and drug binding ability, DrugBank database was used to check the druggability potential of non-homologous essential proteins. This gives the essential drug target proteins for CMCC45402: 249, for SP291: 247, and for ES15: 237.During the course of this screening procedure many novel target proteins were also identified for which no hits were available in DrugBank.

3.1.7 Subcellular Location Prediction

The retrieved sequences from all three strains were further analyzed using PSORTb version 3.0.2 to identify their subcellular locations. In CMCC45402 strain, 10 were cytoplasmic, 15 were membranous, similarly in SP291 strain there were 8, cytoplasmic proteins, 14 were membranous and in ES15, 6 were cytoplasmic, 18 were membranous. Cytoplasmic proteins were selected for identification of putative drug target as cytoplasmic proteins are mostly consist of enzymes which are important for bacterial growth.

Table 3.1. Subtractive genomic analyses scheme in tabular form.

Features	CMCC45402 strain	SP291 strain	ES15 strain
Total no of sequences	4129	4066	3807
After CD hit at 60%	4067	4002	3587
After Blastp (non-homolog)	3616	3510	3260
DEGG Essential protein	1223	1223	1204
Essential proteins involved in metabolic pathways (KEEG)	596	499	567
Essential drug target proteins	249	247	237
Number of essential cell membrane proteins (PSORT)	15	14	18
Number of essential cytoplasmic proteins (PSORT)	10	8	6

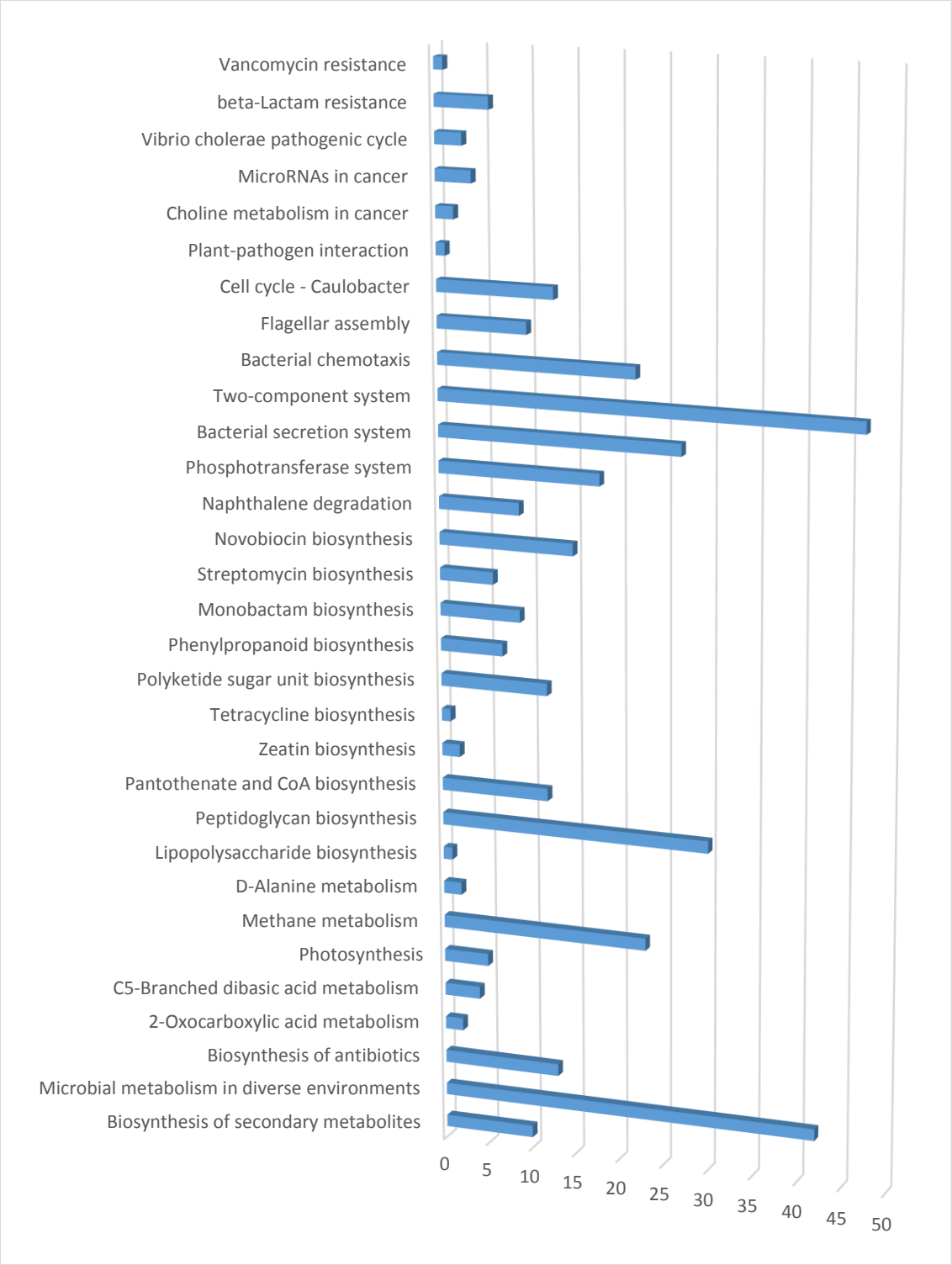


Figure 3.2. Number of proteins involved in the unique metabolic pathways of Cronobacter *sakazakii* strains

3.2 Drug Target Selection

The final step of the genome subtraction is a starting point for the drug discovery process. The selection of the best drug target from the shortlisted candidates was challenging. Since the aim of the current study was on identification of drug targets rather than vaccine targets. Proteins having cytoplasmic subcellular location were selected for further analysis. Six cytoplasmic proteins were then assessed on the basis of various features as outlined in Table 3.3. In the present study, FtsZ has been identified as potential drug target .It is an important cell division protein. An increase in FtsZ increases cell division and vice versa (Gilson, et al., 2001.) among all gram positive bacteria. FtsZ is essential protein in all three strains of *Cronobacter sakazakii* and further study was carried out on FtsZ as it is an effectual and attractive drug target. Important features of the selected target protein is listed in Table 3.2.

Table 3.2. Important features Of FtsZ

Sr .#	Property	Protein
1	Gene name	FtsZ
2	Protein name	Cell division protein
3	Uniprot ID	V5U3P9
4	PDB ID/template	2VAW
5	Molecular weight	41488 Dalton
6	Protein length	379
7	Protein family	Belongs to FtsZ Family
8	Subcellular location	Cytoplasm
9	Biological process	Cytokinesis by binary fission

Table 3.3. Common cytoplasmic potential drug targets in strains of *Cronobacter sakazakii*.

Sr #	Gene name	Protein name	Protein EC#	KO#	Unique pathway name	Pathway KO number
1	PTS-EI.PTSI	Phosphoenolpyruvate -protein Phosphotransferase	EC:2.7.3.9	K08483	Phosphotransferase system	ko02060
2	GlrR	two-component system, NtrC family, response regulator GlrR	……..	K07715	Two-component system	ko02020
3	BarA	two-component system, NarL family, sensor histidine kinase BarA	EC:2.7.13.3	K07678	Two-component system	ko02020
5	FliI	flagellum-specific ATP synthase	EC:3.6.3.14	K02412	Flagellar assembly	ko02040
6	FtsZ	cell division protein FtsZ	…….	ko04112	Cell cycle - Caulobacter	ko04112

3.3 Comparative Homology Modeling

Template of FtsZ from *pseudomonas aueroginosa* GDP (PDB ID: 2VAW) was selected as a reference structure that showed maximum coverage of 94% and identity of 57%. Multiple sequence alignment was done using ClustalO (1.2.1).Using ClustalO, alignment was built from sequences of template and protein as input which confirmed the conservation pattern at the sequence level as shown in Figure 3.4. Five models were generated through MODELLER9.10 along with Discrete Optimized Protein Energy (DOPE) score calculation listed in Table 3.4.Validation of the generated models is carried out by comparative analysis between the models generated by MODELLER and those generated by different automated web-based modeling tools such as SWISS-MODEL, 3D-JIGSAW, I-TASSER and ModWeb. For further validation the generated models were then subjected to (SAVES) Structural Analysis and Verification Server. The comparative analysis was done between web-based servers and model generated by MODELLER is shown in Table 3.4. PROCHECK generated Ramachandran plot, to inspect bad contacts and steric clashes to select the most accurate and reliable model as shown in Figure 3.3. On basis of quality assessment measures Model 2 was selected which was used for further processing. Model 2 exhibited strong stereochemistry with 94.7% residues in most favored region, 4.5% residue in additional allowed region, 0.4% residues in disallowed regions, and lowest Z score of 9.93 as shown in Table 3.4. Then this model is superimposed with template (2VAW) and its RMSD value was 0.546 Å, which ensure the similarity of the protein structure with the template as shown in Figure 3.6. The selected model is subjected to energy minimization with the tripos force field (TFF) using UCSF Chimera to relax the overall structure and remove steric clashes of side chains.

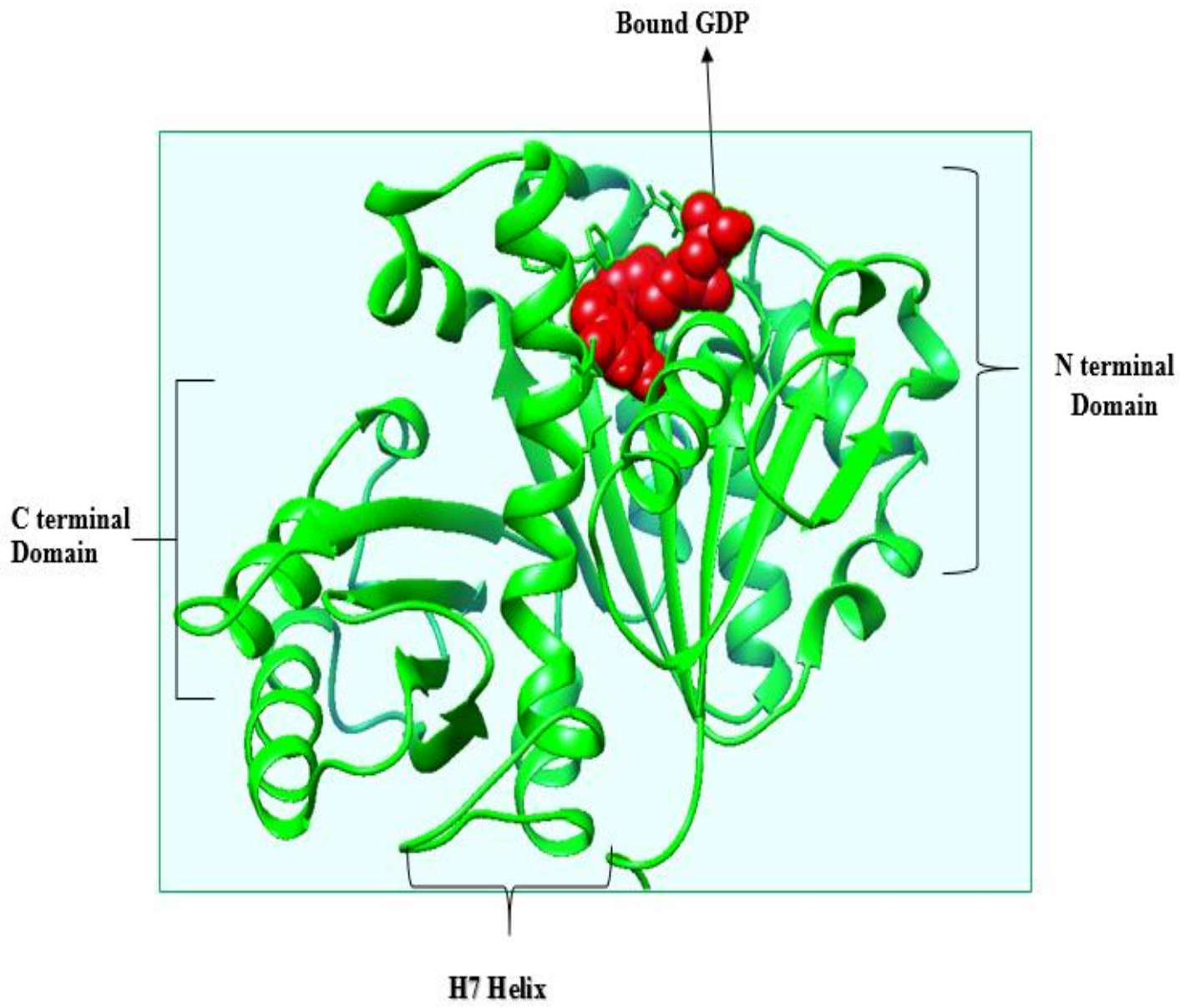


Figure 3.3 Homology model of FtsZ generated by MODELLER

Table 3. 4. Stereo-chemical properties of comparative homology modeled structures.

Structure Resource	Number of Residues				G Factor	Bad Contacts	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)	93.9%	5.3%	0.4	0.4%		-0.2	-9.92
MODELLER (2)	**94.7%**	**4.5%**	**0.4%**	**0.4%**	**0.7**	**-0.2**	**-9.93**
MODELLER (3)	94.3%	4.9%	0.3%	0.4%	0.6	-0.1	-9.83
MODELLER (4)	93.6%	5.7%	0.4%	0.2%	0.5	-0.1	-9.84
MODELLER (5)	93.2%	5.3%	0.8%	0.0%	0.6	-0.1	-9.90
I-TASSER	69.5%	25.8%	2.7%	2.0%	0.1	-0.4	-10.6
ModWeb	75.0%	17.9%	1.8%	5.4%	0.8	-0.2	-9.04
Swiss-Model	84.1%	12.5%	1.7%	1.7%	0.7	-0.2	-9.55
Phyre2	85.9%	12.8%	1.6%	1.4%	1.6	0.8	-9.86
3D-JIGSAW	69.2%	24.4%	3.3%	2.1%	-3.8	0.2	-8.76

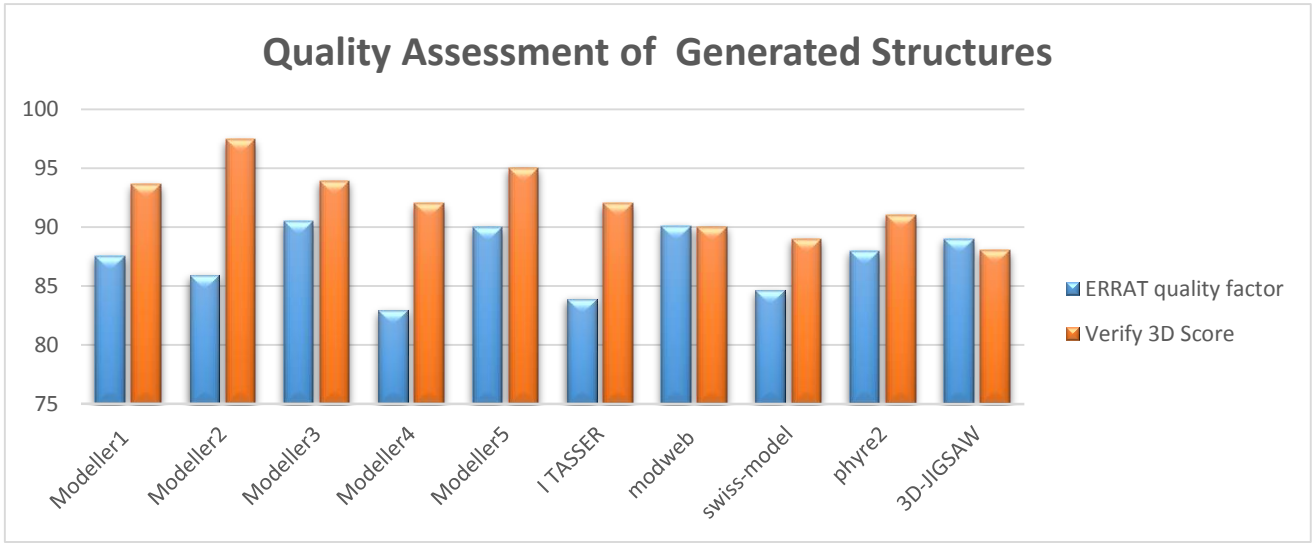


Figure 3.4 Comparison of Verify 3D score and ERRAT quality factors of modeled protein

Table 3.5. Physicochemical properties of FtsZ through ExPASy ProtParam tool

Physicochemical properties	Values
Number of amino acid	379
Molecular weight	39806.2
Theoretical Pi	4.67
Total number of negatively charged residues (Asp + Glu)	53
Total number of positively charged residues (Arg + Lys)	35
Aliphatic index	91.93
Instability index	25.41(stable)
Grand average of hydropathicity (GRAVY)	-0.033

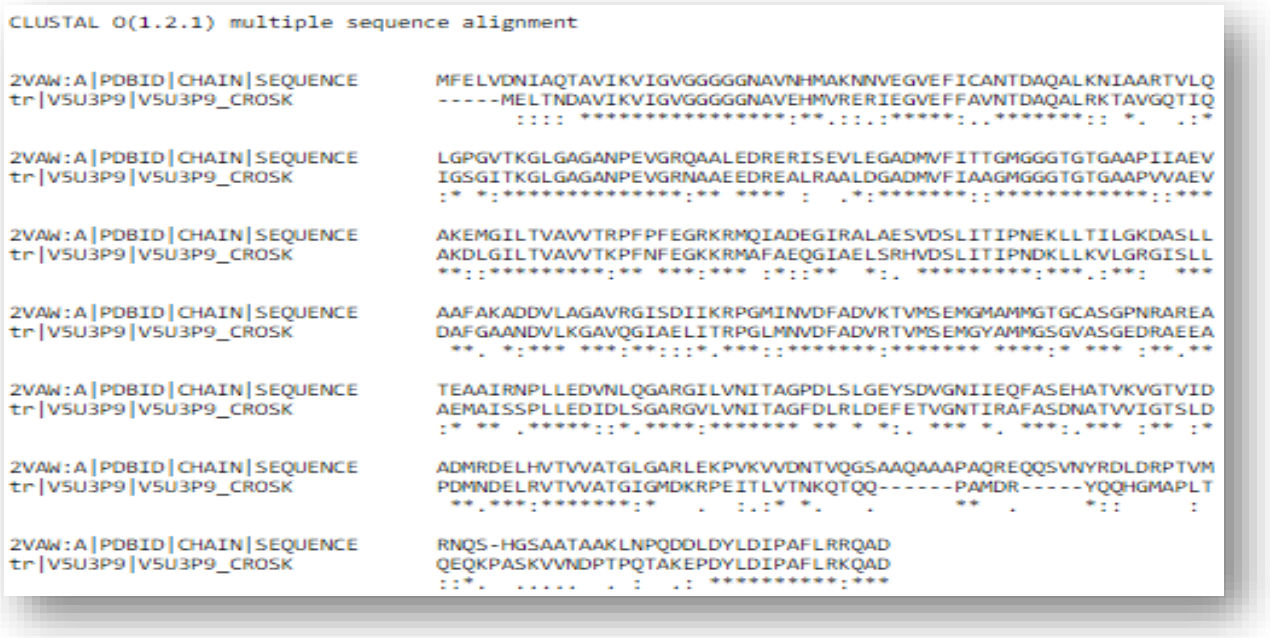


Figure 3.5. Alignment of template and target using ClustalO.

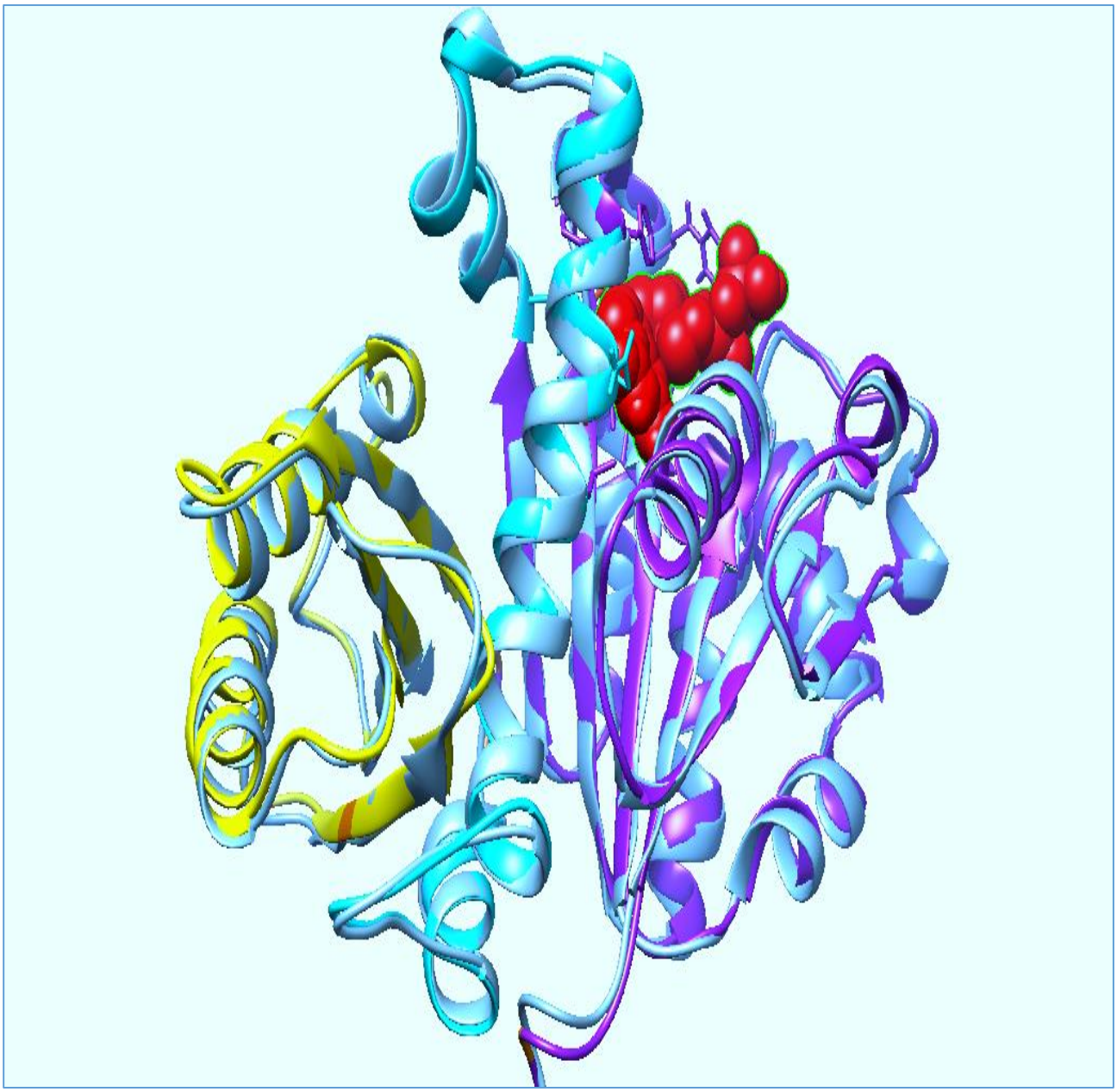


Figure 3.6. Superimposed structure of target and template FtsZ

3.4 Molecular Docking

Molecular docking aims at predicting the binding modes of ligands with protein which is the most important step as it plays the crucial role in drug designing. FtsZ is composed of three evolutionarily conserved domains which are shown in Figure 3.3. In the present study inhibitors were docked into the active site of protein.

3.4.1 Active Site Identification

Active site information is essential to start with the docking procedure. Active site of FtsZ was predicted through extensive literature search (Kaur, et al., 2010; Heger, et al., 2014; Hsin, et al., 2012) and DoGSiteScorer. Conserved active site is observed in different orthologous of FtsZ. DoGSiteScorer predicted three pockets and from second pocket active residue was picked which was same as given in literature.

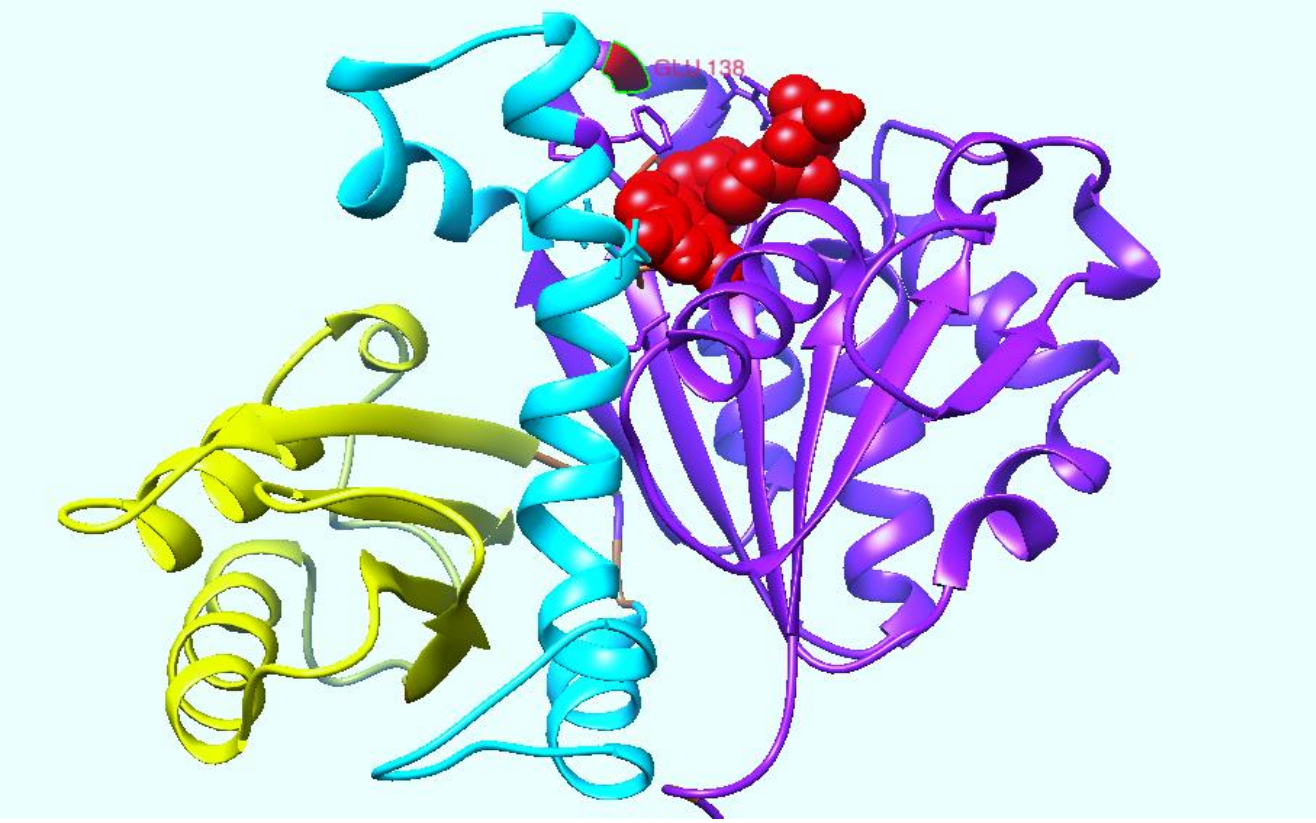


Figure 3.7. Topological view of FtsZ domains along with active residue (red)

3.4.2 Secondary structure prediction.

Psipred provides the secondary structure prediction results. Beta sheets alpha helices and random coils are secondary structure elements. Specific secondary structure is defined by position of every amino acid as shown in Figure 3.8 highlights the amino acids which make secondary structures formation. Results shows that secondary structure prediction contains more helixes than beta sheets which means that the protein is highly flexible.

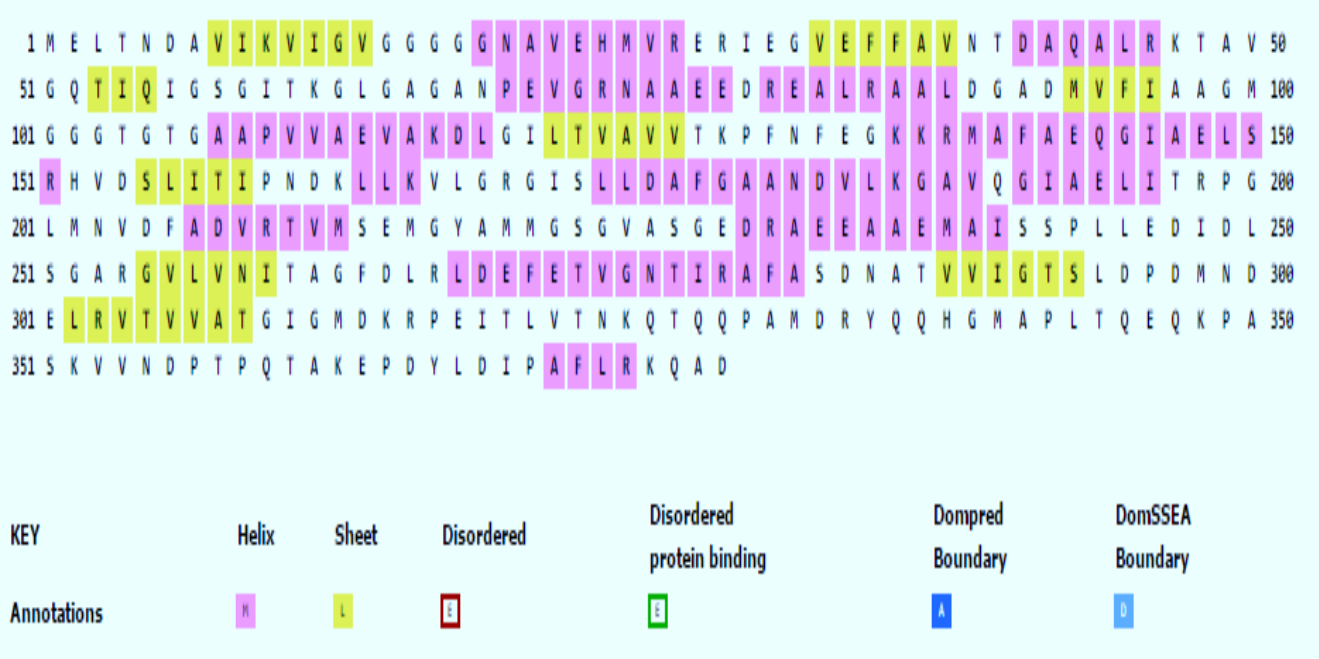


Figure 3.8 Secondary structure prediction of FtsZ.

3.4.3 Inhibitors Selection

Extensive literature search predicted important inhibitors against FtsZ. FtsZ inhibitors were mostly consist of derivatives of 3-methoxybenzamide,zantrins(polyphenols), Berberine an alkaloid product from several plant species like blackberry and golden scale, Cinnam-aldehyde from spices etc. which block substrate channel and inhibit the activity of enzyme (Boberek, et al.,2010). The current study is carried out to predict the inhibition mechanism of these inhibitors on protein functionality by hindering the activity of active site. These inhibitors were accessed from the BRaunschweig Enzyme Database (BRENDA) (Schomburg, 2002) . A total one hundred and two inhibitors were docked into FtsZ active site to analyze enzyme inhibition mechanism.

3.4.4 Docking Interaction Analysis

One hundred and two ligands molecules were prepared, docked into the active site of FtsZ using GOLD. GOLDScore ranged from 35.2 to 86. While corresponding binding affinities were also calculated using AutoDock Vina which were ranging from -10.3 to -4.9 kcal/mol. The highest score of 86.0 was achieved for compound 78, 2'"-hydroxy-5"benzylisouvarinol-B a polyphenol extracted from X. Afticana (Urgaonkar, et al. 2005) with binding affinity of -7.6 kcal/mol which was chosen for further analysis to find its interactions(hydrophobic and hydrogen bond interactions). Docking scores and respective binding affinities for top 10 compounds, arranged in descending order of Gold Score values are provided in Table 3.6. 2D and 3D depiction of protein and ligand interactions analysis was carried out employing UCSF Chimera, Discovery Studio, LIGPLOT and MOE which revealed the conformational details and preferred orientation of the ligand binding.

Table 3.6 Top ten docking results ranked on basis of GoldScore in descending order with respective binding affinities.

Compound Number	GOLDScore	Binding Affinities (kcal/mol)
Comp78	86.0	-9.6
Comp95	74.6	-9.4
Comp58	73.4	-7.9
Comp57	66.0	-8.6
Comp91	65.2	-7.4
Comp66	64.0	-8.0
Comp37	63.9	-9.3
Comp94	63.1	-8.3
Comp90	62.4	-7.0
Comp97	55.2	-8.1

The binding of compound 78 was observed at N terminal domain of protein .The visual assessment revealed that the following residues were involved in framing the inhibitor in active site of N terminal domain of FtsZ: Ile176, lys141, Gly139, Phe137, Ser177, Leu178, Phe182, Arg122, Glu138, Lys140, Asn136 and Arg142 as shown in chimera visualization of ligand dock pose with FtsZ in Figure 3.9. Non-bonded hydrophobic and electrostatic interactions were exhibited by, lys141, Gly139, Phe137, Ser177, Leu178, and Phe182; while Glu138, Arg122, and Lys140 showed prominent hydrogen bonds with the ligand. This can be seen in 2D interaction images of LIGPLOT and MOE as shown in Figure 3.10 and 3.11. Moreover, LIGPLOT image i.e. Figure 3.10 illustrated the presence of hydrogen bonds between ligand and target. Ligand oxygen moiety formed two hydrogen bonds with active residue Glu138, and Arg142 with a distance of 2.94 Å and 2.83 Å. Lys140 nitrogen atom developed hydrogen bonding with ligand at the distance of 2.70 Å. Hydrogen bonding details are provided in Table 3.7.

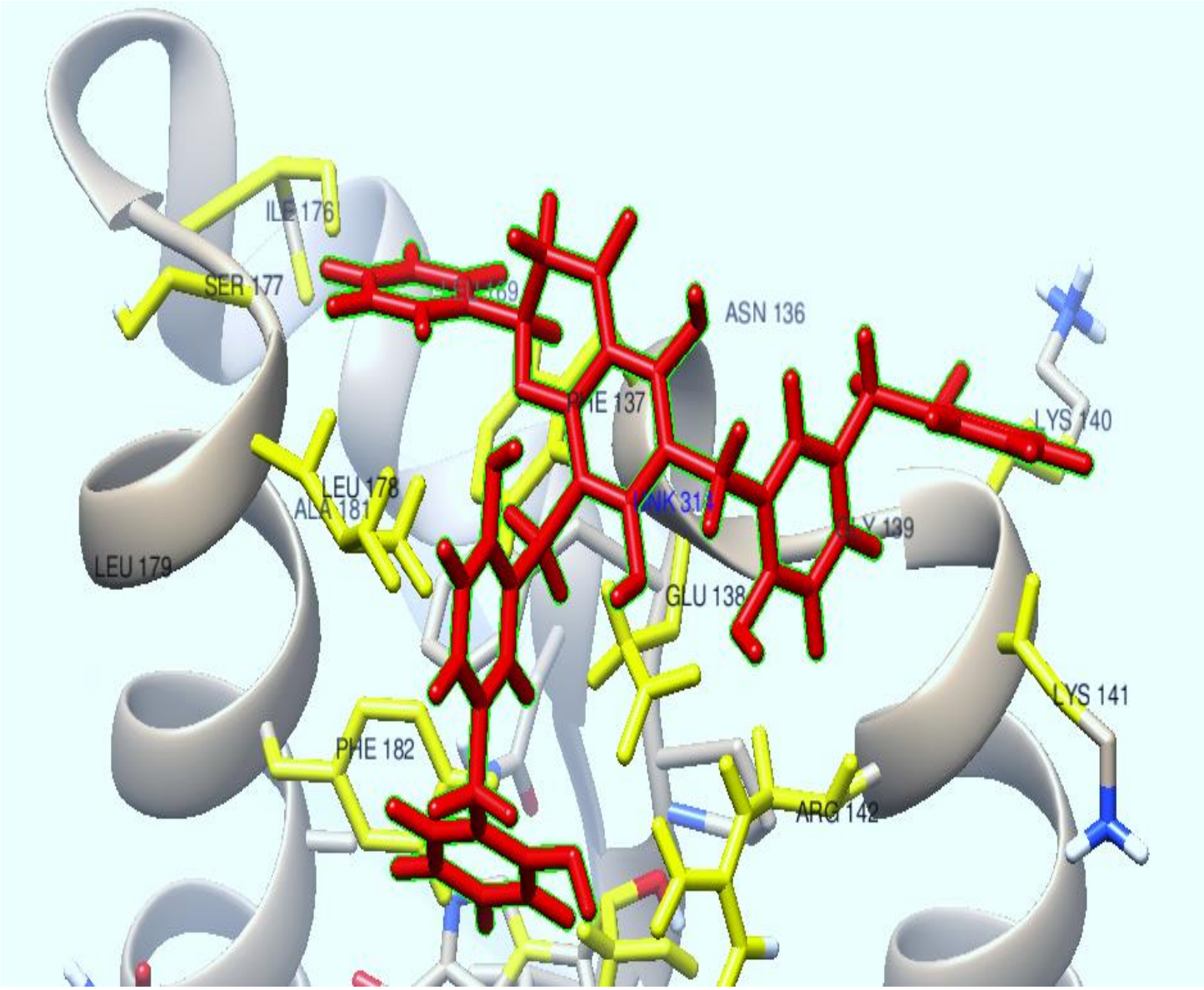


Figure 3.9 Visualization of ligand dock pose with FtsZ protein using UCSF chimera

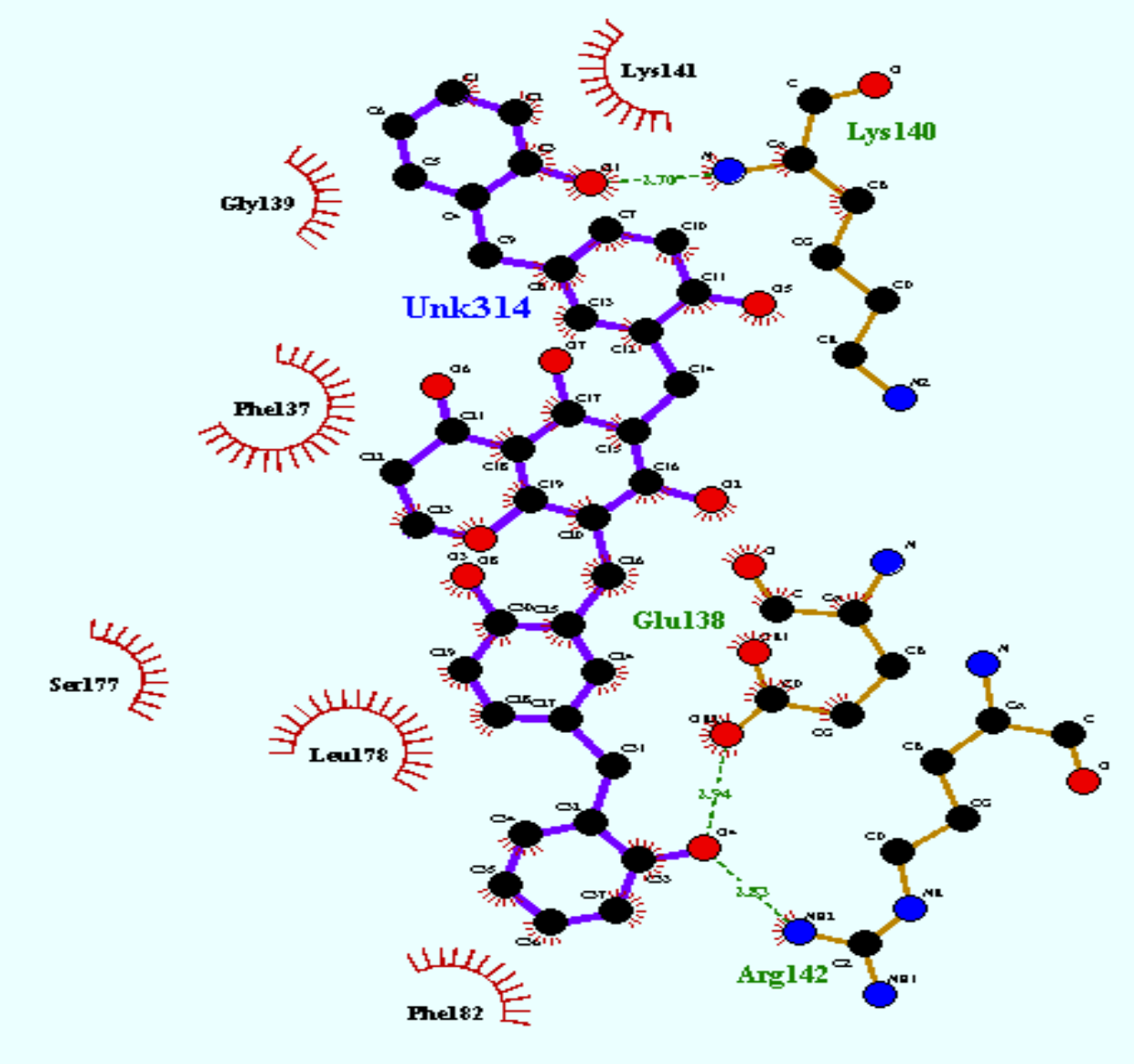


Figure 3.10. Interaction of ligand with FtsZ, highlighting interacting residues through LIGPLOT.

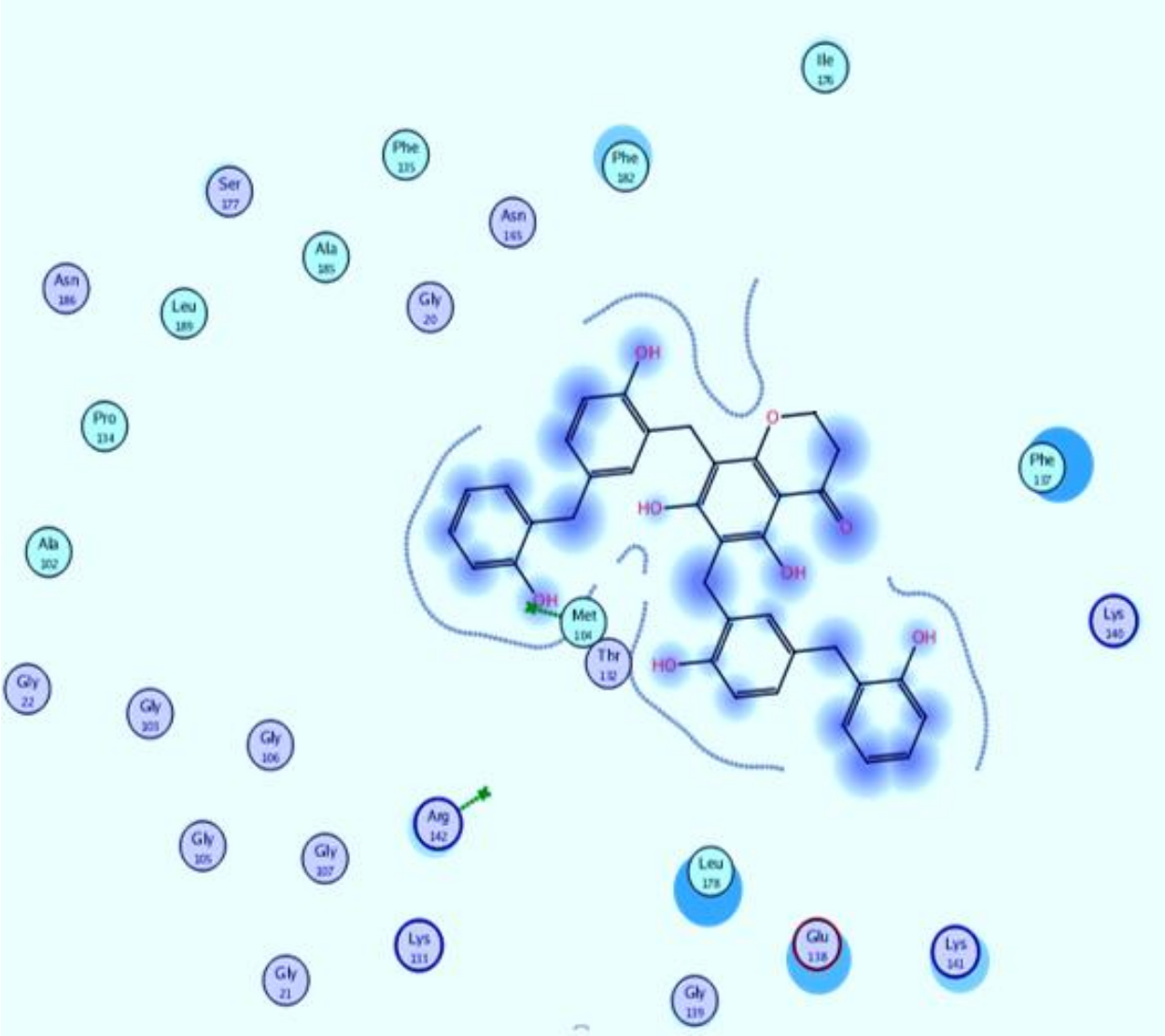


Figure 3.11. MOE ligand interaction image showing bonded and non-bonded interactions of inhibitor bound FtsZ.

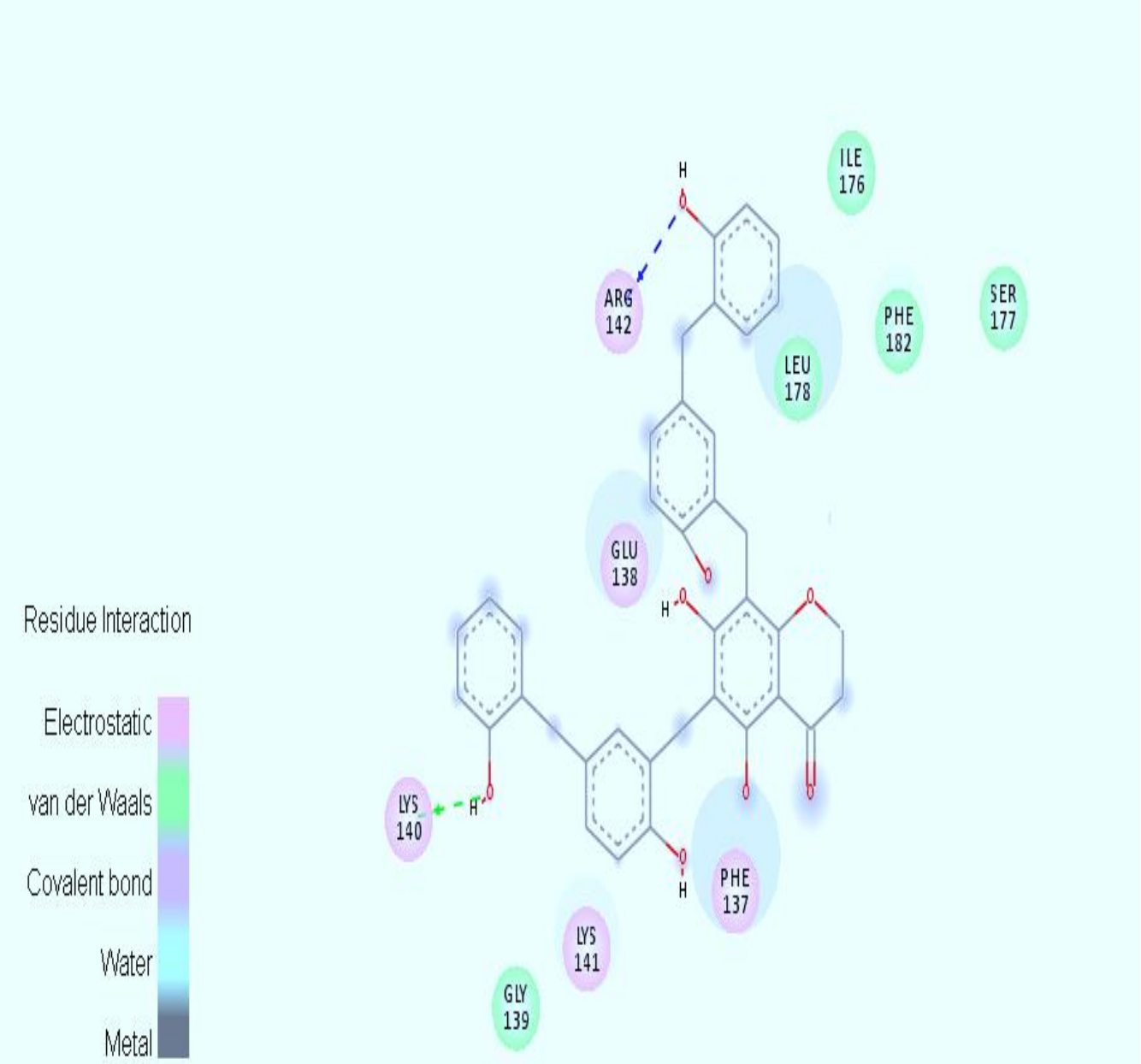


Figure 3.12. Interaction diagram of compound 78 with FtsZ protein using Discovery studio

Table 3.7. Hydrogen bond details of best docked compound with important interacting residues.

Interacting Atom of Protein	Interacting Atom of Ligand	Distance (A˚)
LYS141: N	O	3.61
LYS141: H	O	2.63
GLY139: HA3	O	2.55
GLY139: HB3	O	2.31
GLU138: OE1	O	2.95
PHE137: HD1	O	2.33
PHE137: HB2	O	3.10
ARG142: NE	H	2.32
GLU138: OE1	H	2.85
PHE137: HB2	H	3.68
LEU178: HD11	H	3.70
SER177: HA	H	3.01
ILE176: HG23	H	3.08
GLU138: OE2	O	2.84
GLY139: HA3	O	3.39
ARG142: HG3	O	2.31
GLU138: OE1	O	3.71

3.5 Molecular Dynamics Simulation

The docking study provided a meaningful knowledge about the druggability potential of *Cronobacter sakazakii* FtsZ within a static environment. Molecular dynamic study was carried out in order to understand the conformational changes (dynamic behavior) of FtsZ which is a potential drug target molecular dynamic simulation carried out followed by trajectory analysis was done utilizing PTRAJ module of AMBER to assess various properties of docked system. Properties including the RMSD, RMSF, B-factor and radius of gyration were plotted as a function of time to unravel the conformational changes of ligand bound FtsZ in a hydrated system.

3.5.1 Root Mean Square Deviation (RMSD)

The deviation of the backbone Cα atoms was noticed in the entire production run of docked protein for a time period of 100 ns. Average RMSD value calculated for docked complex over 100 ns time period was 2.8476 Å, plot showed the steady increase in RMSD in the start and then gradual increase was noticed as shown in Figure 3.13. At the end of graph some stability was seen as shown in Figure 3.13. During the simulation runs no major domain shifts were observed although there was some secondary structure rearrangements. The slight change in the structure is due to the elusive changes in backbone residues as shown in Figure 3.15.

3.5.2 Root Mean Square Fluctuations (RMSF)

RMSF is investigated to have insights into structural dynamics of residues. It provided a means to find and analyze the structurally flexible rigid regions of the drug target. Average root mean square fluctuations for the docked system for 100 ns was 1.39 Å which is well within the range and confirms the stability of system as shown in Figure 3.14. The maximum value of RMSF in case of docked protein was 10.4853 Å. RMSF trajectory analysis of normal binding site complex showed higher fluctuations for the amino acid residues present at N terminal domain whereas active site residues revealed lesser fluctuations. From above discussed observation it can be interpreted that active site is more

stable in a complex. In comparison with the entire protein, higher fluctuations were observed at regions: 1-20, 130-150, and 160-180, for docked protein signifying disorderness of C-alpha atoms in these regions.

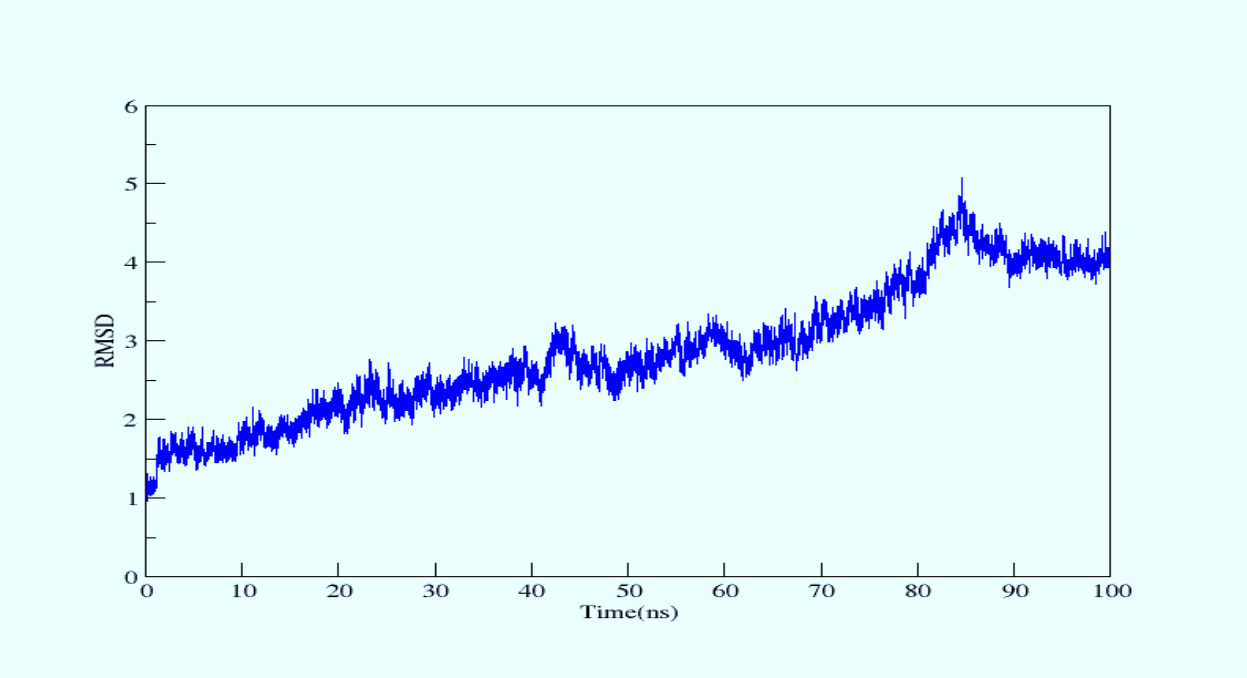


Figure 3.13. Root mean square deviation plot for 100 ns simulation run of docked FtsZ of *Cronobacter sakazakii*.

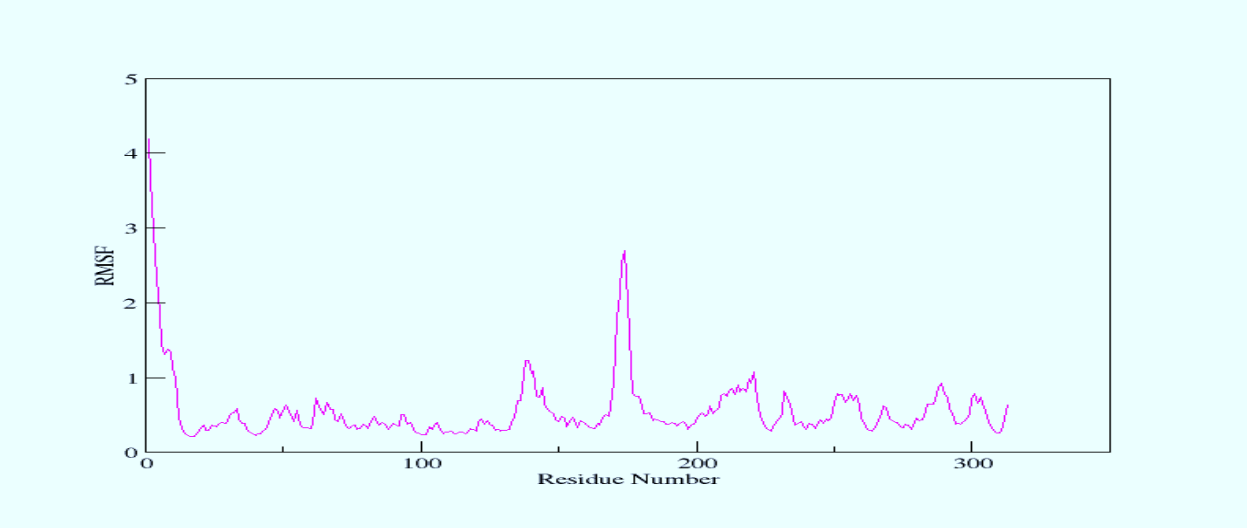
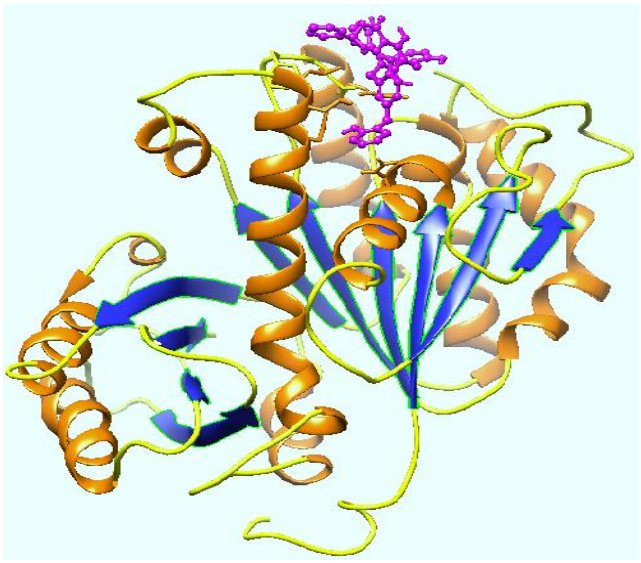
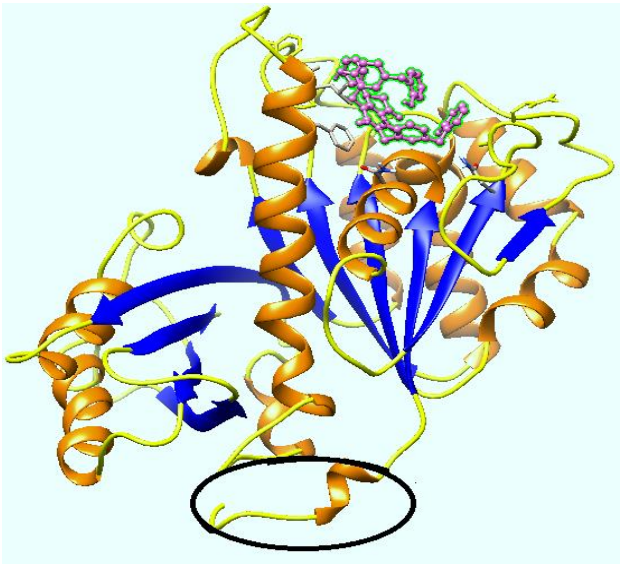


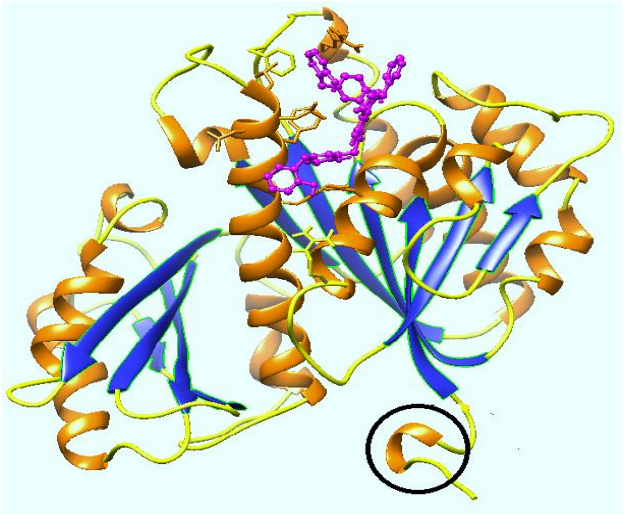
Figure 3.14. Root mean square fluctuation plot for docked FtsZ of *Cronobacter sakazakii*.



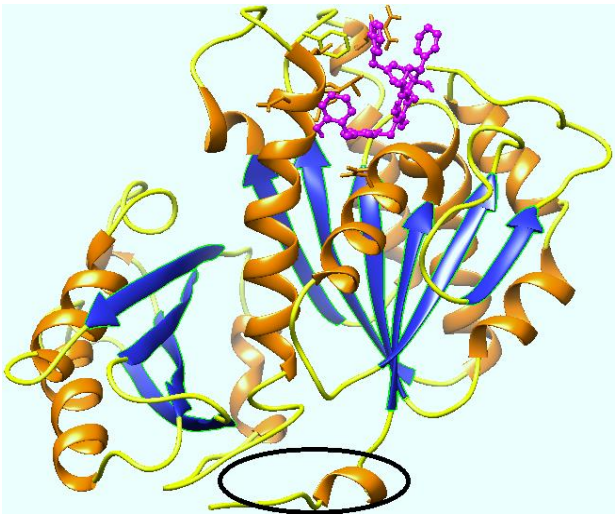
0ns



40ns



80 ns



100 ns

Figure 3.15. Snap shots of docked FtsZ of *Cronobacter sakazakii* at 0 ns, 40ns, 80 ns and 100 ns. Ligand is shown in stick representation in purple.

3.5.3 Beta Factor

Beta factor is measured in terms of RMSF. The value of beta factor is dependent on the level of localized atomic fluctuations which collectively contribute to the global vibrational movements of the protein and its thermal stability. B-factor value less than 30 $Å^2$ for an atom is the indication of confidence in its atomic position, but values greater than 60 $Å^2$ indicates possible atom's disorderness. The average value for the system of interest comes out to be 84.82.The slight increase in value of beta factor is probably due to residues 1 to 20 that exhibit high instability with a highest peak as shown in Figure 3.17 due to fluctuations in loop region. The loop is also replaced by helix at 40 ns, 80ns and 100ns as shown in Figure 3.15.These residues are not involved in ligand binding they do not affect the interaction of ligand and protein. The value for residues involved in ligand binding are less fluctuating signifying the uniformity of active site.

3.5.4 Radius of Gyration

Compactness and stability of a protein structure is determined by its radius of gyration (Rg). Reduction of radius of gyration values specified the stability of the system. It is measured as a root mean square distance between the protein‘s center of gravity and its ends. Docked system has an average value of 19.56 Å shown in Figure 3.16. This implies that docked complex exhibit stable and compact system according to the value of radius of gyration.

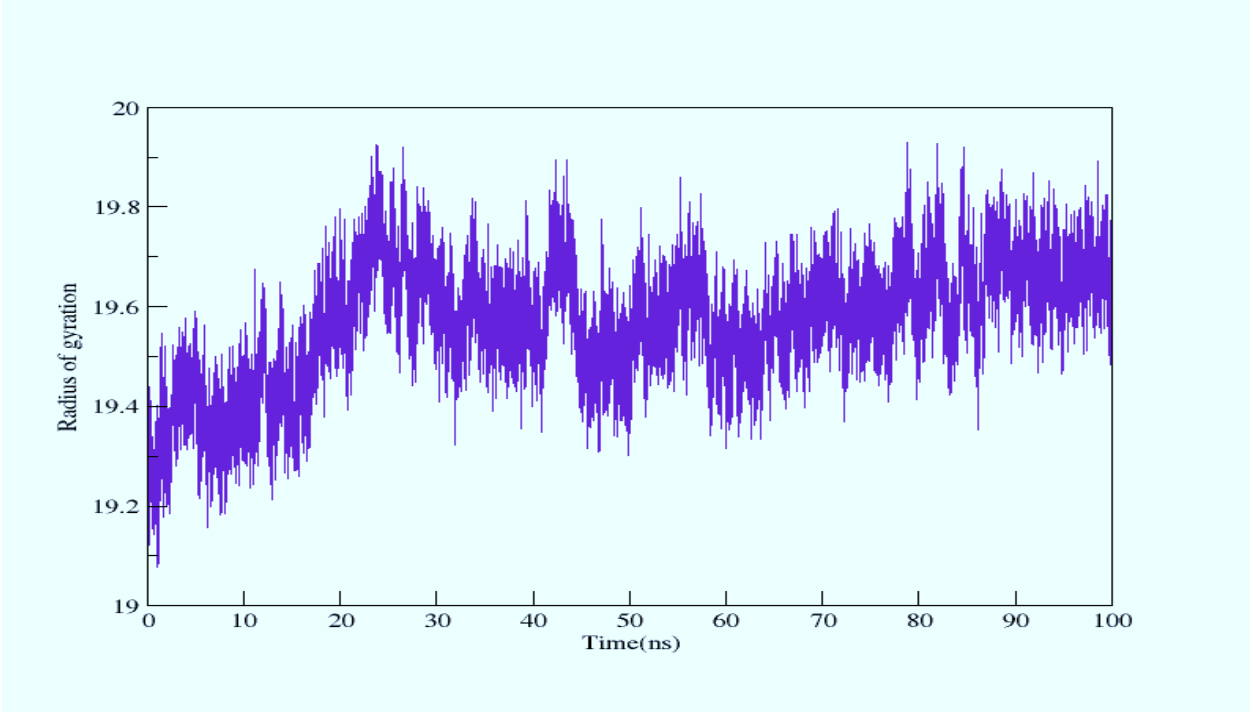


Figure 3.16. Radius of gyration of FtsZ over 100 ns simulation time period.

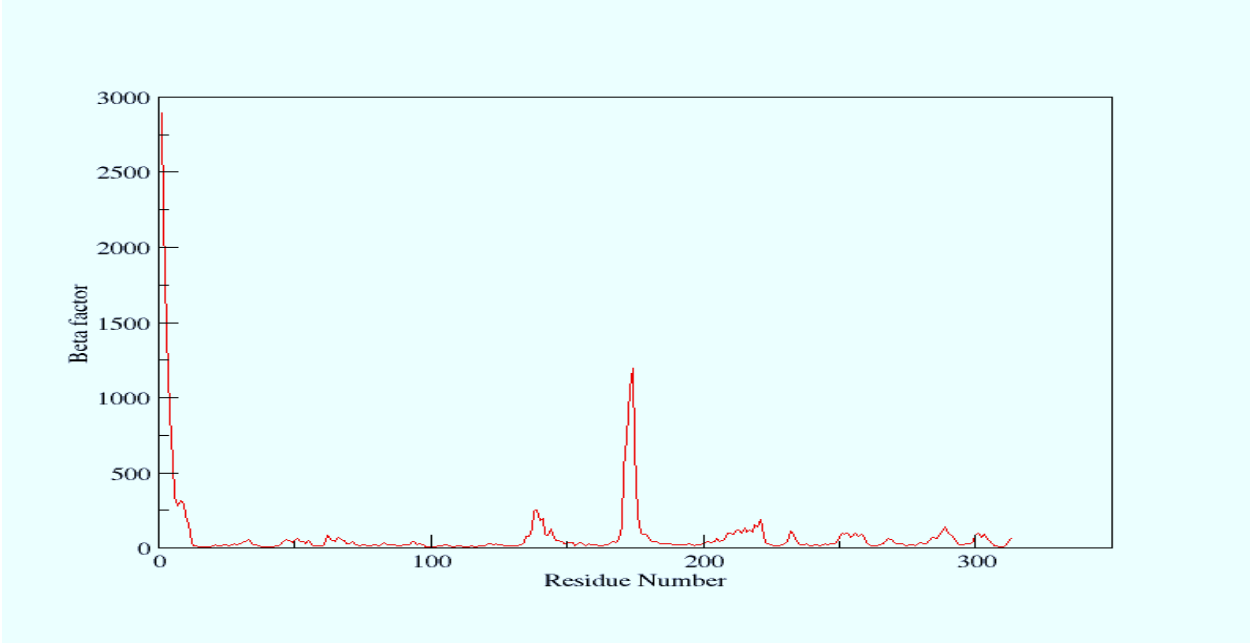


Figure 3.17. β-Factor graph of FtsZ protein over 100 ns simulation run.

4 DISCUSSION

Infectious diseases are alarming for the public health around the world. Capacity of adjustment and reshaping of pathogenic genome in bacterial species restricts the activity of antimicrobial agents against these pathogens. The rise of anti-toxin resistance strains are not as a result of high rate of bacterial genome reshaping by mutation but also activated additionally by antibiotics imposed selective pressure. Consequently, every one of these bottlenecks demand more up to date, specific and more intense therapeutic agents. The *in silico* drug design is an extensive field in which the distinctive sides of fundamental research and practice are amalgamated and inspire each other. Current procedures, for example, QSAR/QSPR, structure-based design, combinatorial library design, cheminformatics, bioinformatics and the expanding number of chemical and biological databases are utilized as a part of the field. (Wadood, et al., 2013). Computer aided drug designing over the past decades has been making a big splash in the area of drug design (Azam, et al., 2014). Subtractive genomics is a thriving in silico approach that filters out pathogenic novel drug targets from complete genome.

The present work aims on the identification of potential druggable candidates of *Cronobacter sakazakii*, also known as yellow pigmented *Enterobacter sakazakii* is a Gram-negative bacillus, rod shaped, non-spore forming, human opportunistic and food borne pathogen. The primary reservoir of *Cronobacter sakazakii* has yet to be determined but it is postulated that plant material may be an important source. Different antibiotics are employed to control this pathogen but recent studies investigated emergence of antibiotic-resistant strains due to which the situation demands a speedy course of drug discovery process (Zhou, et al., 2011). This went about as a boost for the present study where in silico drug design approach is utilized to facilitate the process and has given significant insight into the functional basis of the selected drug target.

This study is an aim to use *in silico* subtractive genomics approach to shortlist the possible drug targets in CMCC45402, SP2, and ES15 strains of *Cronobacter sakazakii*. Different steps of therapeutic target screening technique i.e. *in silico* subtractive genomic was used to filter pathogen genome.CMCC45402, ES15.SP291 strains of *Cronobacter sakazakii* were targeted to combat multiple strains with the same drug.

The number of proteins in these three strain are 3807, 4066 and 4129 respectively. Subtractive genomic steps were applied on each strain independently. Furthermore, toward the end accompanied 10, 8 and 6 potential drug target proteins which were cytoplasmic localized proteins. These common cytoplasmic targets were identified as a result of subtractive genomics approach which are involved in the following unique pathways: *Vibrio cholera* pathogenic cycle, Cell cycle - *Caulobacter*, Flagellar assembly, and Two-component system. After all evaluation of subtractive genomics study the target selected in current study was cell division protein (FtsZ), which is involved in "cell cycle-*caulobacter*" pathway. Multiple sequence alignment (MSA) of FtsZ sequences obtained from all three strains shows that they are highly identical, therefore targeting one protein sequence from one strain can work for all these three strains. FtsZ gene was used for comparative homology modeling to elucidate the binding interactions of ligands.

Comparative homology modeling was performed on FtsZ through MODELLER along with other automated web-based servers: I-TASSER, ModWeb, Swiss-Model, Phyre2, 3D-JIGSAW and the model was evaluated with different tools like ERRAT, PDBSum and Verify3D. Template FtsZ from *pseudomonas aeruginosa* GDP (PDB ID: 2VAW) was selected as a reference structure with the query coverage of 94% and 57% sequence identity. Comparison of model is shown in Table 3.4 which was helpful in selection of best model on basis of satisfactory values obtained from results. The best model came out through MODELLER. Furthermore molecular docking and molecular dynamic simulation was carried out on the best model selected. Selected model generated via MODELLER9.10 was monomer. RMSD is calculated to elucidate the similarity between two proteins, which analyze the ensembles of backbone atoms for proper conformation (Maiorov and Crippen, 1994). RMSD value calculated for target and template was 0.488 Å, which means that the generated model was of good quality and contain similar main chain fold as that of template chain. Physicochemical properties of the modeled structure are given in Table 3.5.

Molecular docking is utilized for identification of binding interactions of two molecules along with their binding affinities (Huang and Zou, 2010). To dock a ligand an active site is required in target protein. In current study Glu138, charged amino acid, was used as an active residue.

Molecular Docking was carried out using AutoDock Vina and GOLD which provided a detailed information about the protein interaction, conformation, orientation, and type of inhibition. Both these docking programs consist of different scoring algorithms which resulted in the slight change in results from both programs. Total 102 ligands were docked in active site which is present in N terminal domain of FtsZ protein. The highest GOLD score obtained is 86 of compound 78 with binding affinity of -9.6 kcal/mol. which was the best docking ligand that was taken for further detailed analysis. LIGPLOT, Discovery Studio, and UCSF Chimera was employed for detailed visualization of interactions of compound 78 with protein which include interactions like electrostatic interactions, hydrophobic interactions, and hydrogen bonding. Hydrogen bond interactions and binding energy calculations are important factors for investigation of representative docking studies. Hydrogen bonding plays a key role for inhibition mechanism in complex. If hydrogen bond interactions of ligand with active site residues is utilized to block a structure it becomes unstable and dispersal of ligand is seen. (Azam, et al., 2012). Structure stability and compactness of compound 78 was analyzed through hydrogen bonds networks involving the residues like Glu138, Arg122, and Lys140 showed prominent hydrogen bonds with the ligand. Non-bonded hydrophobic and electrostatic interactions were exhibited by lys141, Gly139, Phe137, Ser177, Leu178 and Phe182. Ligand oxygen moiety formed two hydrogen bonds with active residue Glu138, and Arg142 with a distance of 2.94 Å and 2.83 Å and Lys140 nitrogen atom developed hydrogen bonding with ligand at a distance of 2.70 Å.

Molecular dynamics simulations are employed to identify the dynamic and time dependent behavior of FtsZ. It provides knowledge about the region of molecule involved in dynamic behavior of system of interest (Azam, et al., 2012) . Inhibition process, with respect to time was elucidated in solvated medium, for docked complex of FtsZ in the current study. RMSD, RMSF, B-factor, and radius of gyration for docked system are generated. RMSD value of docked system calculated was >2 Å and docked complex was moving towards stability at the end of 100 ns after binding of FtsZ with its receptor. This implies that FtsZ is stable in its complex. RMSF averaged value calculated docked complex was >1 Å and few peaks were observed in the plot. Highest peak of RMSF was observed for first amino acid Met1 with RMSF value of 4.23 Å. Most fluctuating residues were Asn9, Gly20, Val12,

Val130, Phe135, Asp158, and Gly173.Most of the fluctuating residues are present in loop region of protein. Loops regions are considered as the most fluctuating residues containing region of protein (Ruvinsky and Vakser, 2010). Additionally, helices and sheets change their orientation. During the course of simulation helices and beta sheets also change their orientation slightly, during 100 ns. These movements are important for inhibition process. *In silico* subtractive genomics approach applied in current study depicted the significant results at different steps of analysis. Binding modes of inhibitor confirms the inhibition process. Similarly, structure dynamics of docked protein provided the eloquent insight, Docking studies filtered the best binding inhibitors among the given library of compounds whereas simulation studies identified the proper orientation and positioning of the ligand molecule in the active site. It also explored the timely dependent dynamic behavior of protein which can be practiced to increase the potency of drug against the *Cronobacter sakazakii* strains along with drug target specificity and selectivity which ultimately provide remedy for infections caused by studied pathogen.

Conclusion

The *in silico* based methodology includes a series of screening of proteins that can be utilized as potential drug targets and antibody candidates. These drug targets are important for the growth of the organisms and these proteins have unique functionality associated with a unique pathway. These proteins neither have a substitute protein nor an alternative pathway to perform the process. The current study carried out to design a drug-like molecule that can block cell division protein FtsZ. The microorganisms are rapidly getting resistant to the current drugs, so designing better and viable drugs ought to be made quicker. The current study was to explore the druggable genome of *Cronobacter Sakazakii* and identification of the putative therapeutic targets against emerging drug resistant pathogen. Subtractive genomic approach led to the identification of many pathogen specific targets, which were common to three strains under study. Parameters like durggability assessment and localization specifications led to the selection of FtsZ as drug target. Homology modeling was employed to model FtsZ which was further docked with library of 102 compounds. Furthermore, molecular docking was done which showed compound 78 is the highest GOLD scoring compound with the GOLD score of 86 which was the true consideration of the interaction strength. Polar amino acids was prominent in inhibitor moieties and clarifying the physicochemical properties of binding site. Moreover, MD simulation studies well explained the dynamic behavior of docked complex. Beside the side chain fluctuations and minor loop movement, stability of inhibitor and target protein complex was observed. All these function valuation fueled for rational structure based drug designing of novel inhibitors for *Cronobacter Sakazakii* with increase potency.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J., 1990. Basic local alignment search tool. *Journal of molecular biology*, *215*(3), pp.403-410.

Awasthi, D., Kumar, K. and Ojima, I., 2011. Therapeutic potential of FtsZ inhibition: a patent perspective. *Expert opinion on therapeutic patents*, *21*(5), pp.657-679.

Azam, S.S. and Raza, S., 2014. Structure modeling and hybrid virtual screening study of Alzheimer's associated protease kallikrein 8 for the identification of novel inhibitors. *Medicinal Chemistry Research*, *23*(7), pp.3516-3527.

Bar-Oz, B., Preminger, A., Peleg, O., Block, C. and Arad, I., 2001. Enterobacter sakazakii infection in the newborn. *Acta Paediatrica*, *90*(3), pp.356-358.

Bates, P.A., Kelley, L.A., MacCallum, R.M. and Sternberg, M.J., 2001. Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. *Proteins: Structure, Function, and Bioinformatics*, *45*(S5), pp.39-46.

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.

Beuria, T.K., Santra, M.K. and Panda, D., 2005. Sanguinarine blocks cytokinesis in bacteria by inhibiting FtsZ assembly and bundling.*Biochemistry*, *44*(50), pp.16584-16593.

Boberek, J.M., Stach, J. and Good, L., 2010. Genetic evidence for inhibition of bacterial division protein FtsZ by berberine. *PloS one*, *5*(10), p.e13745.

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

Bowen, A.B. and Braden, C.R., 2006. Invasive Enterobacter sakazakii disease in infants. *Emerging infectious diseases*, *12*(8), p.1185.

Britton, D.T., Odo, E.A., Gonfa, G.G., Jonah, E.O. and Härting, M., 2009. Size distribution and surface characteristics of silicon nanoparticles. *Journal of Applied Crystallography*, *42*(3), pp.448-456.

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.

Camacho, C., Madden, T., Ma, N., Tao, T., Agarwala, R. and Morgulis, A., 2013. BLAST command line applications user manual.

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.

Darden, T., York, D. and Pedersen, L., 1993. Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *The Journal of chemical physics*, *98*(12), pp.10089-10092.

Domadia, P., Swarup, S., Bhunia, A., Sivaraman, J. and Dasgupta, D., 2007. Inhibition of bacterial cell division protein FtsZ by cinnamaldehyde.*Biochemical pharmacology*, *74*(6), pp.831-840.

Drudy, D., Mullane, N.R., Quinn, T., Wall, P.G. and Fanning, S., 2006. Enterobacter sakazakii: an emerging pathogen in powdered infant formula.*Clinical Infectious Diseases*, *42*(7), pp.996-1002.

Eisenberg, D., Lüthy, R. and Bowie, J.U., 1997. VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods in enzymology*, (277), pp.396-404.

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.

Eweas, A.F., Maghrabi, I.A. and Namarneh, A.I., Advances in molecular modeling and docking as a tool for modern drug discovery.

Fiore, A., Casale, M. and Aureli, P., 2008. Enterobacter sakazakii: epidemiology, clinical presentation, prevention and control. *Ann Ist Super Sanita*, *44*(3), pp.275-280.

Forsythe, S.J., 2005. Enterobacter sakazakii and other bacteria in powdered infant milk formula. *Maternal & child nutrition*, *1*(1), pp.44-50.

Gilson, P.R., Yu, X.C., Hereld, D., Barth, C., Savage, A., Kiefel, B.R., Lay, S., Fisher, P.R., Margolin, W. and Beech, P.L., 2003. Two Dictyostelium orthologs of the prokaryotic cell division protein FtsZ localize to mitochondria and are required for the maintenance of normal mitochondrial morphology.*Eukaryotic cell*, *2*(6), pp.1315-1326.

Gosney, M.A., Martin, M.V., Wright, A.E. and Gallagher, M., 2006. Enterobacter sakazakii in the mouths of stroke patients and its association with aspiration pneumonia. *European journal of internal medicine*, *17*(3), pp.185-188.

Gupta, S.K., Imperiale, T.F. and Sarosi, G.A., 2001. Evaluation of the Winthrop-University Hospital criteria to identify Legionella pneumonia.*CHEST Journal*, *120*(4), pp.1064-1071.

Haydon, D.J., Bennett, J.M., Brown, D., Collins, I., Galbraith, G., Lancett, P., Macdonald, R., Stokes, N.R., Chauhan, P.K., Sutariya, J.K. and Nayal, N., 2010. Creating an antibacterial with in vivo efficacy: synthesis and characterization of potent inhibitors of the bacterial cell division protein FtsZ with improved pharmaceutical properties. *Journal of medicinal chemistry*,*53*(10), pp.3927-3936.

Healy, B., Cooney, S., O'Brien, S., Iversen, C., Whyte, P., Nally, J., Callanan, J.J. and Fanning, S., 2010. Cronobacter (Enterobacter sakazakii): an opportunistic foodborne pathogen. *Foodborne pathogens and disease*, *7*(4), pp.339-350.

Heger, M., van Golen, R.F., Broekgaarden, M. and Michel, M.C., 2014. The molecular basis for the pharmacokinetics and pharmacodynamics of curcumin and its metabolites in relation to cancer. *Pharmacological reviews*, *66*(1), pp.222-307.

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.

Hsin, J., Gopinathan, A. and Huang, K.C., 2012. Nucleotide-dependent conformations of FtsZ dimers and force generation observed through molecular dynamics simulations. *Proceedings of the National Academy of Sciences*, *109*(24), pp.9432-9437.

Huang, Q., Kirikae, F., Kirikae, T., Pepe, A., Amin, A., Respicio, L., Slayden, R.A., Tonge, P.J. and Ojima, I., 2006. Targeting FtsZ for antituberculosis drug discovery: noncytotoxic taxanes as novel antituberculosis agents. *Journal of medicinal chemistry*, *49*(2), pp.463-466.

Huang, S.Y. and Zou, X., 2010. Advances and challenges in protein-ligand docking. *International journal of molecular sciences*, *11*(8), pp.3016-3034.

Huang, Y., Niu, B., Gao, Y., Fu, L. and Li, W., 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics*,*26*(5), pp.680-682.

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

Iversen, C., Mullane, N., McCardell, B., Tall, B.D., Lehner, A., Fanning, S., Stephan, R. and Joosten, H., 2008. Cronobacter gen. nov., a new genus to accommodate the biogroups of Enterobacter sakazakii, and proposal of Cronobacter sakazakii gen. nov., comb. nov., Cronobacter malonaticus sp. nov., Cronobacter turicensis sp. nov., Cronobacter muytjensii sp. nov., Cronobacter dublinensis sp. nov., Cronobacter genomospecies 1, and of three subspecies, Cronobacter dublinensis subsp. dublinensis subsp. nov., Cronobacter dublinensis subsp. lausannensis subsp. nov. and Cronobacter dublinensis subsp. lactaridi subsp. nov. *International journal of systematic and evolutionary microbiology*, *58*(6), pp.1442-1447.

Iversen, C., Waddington, M., On, S.L. and Forsythe, S., 2004. Identification and phylogeny of Enterobacter sakazakii relative to Enterobacter and Citrobacter species. *Journal of clinical microbiology*, *42*(11), pp.5368-5370

Jaiswal, R., Beuria, T.K., Mohan, R., Mahajan, S.K. and Panda, D., 2007. Totarol inhibits bacterial cytokinesis by perturbing the assembly dynamics of FtsZ. *Biochemistry*, *46*(14), pp.4211-4220.

Jennings, Lee D., Ken W. Foreman, Thomas S. Rush, Desiree HH Tsao, Lidia Mosyak, Yuanhong Li, Mohani N. Sukhdeo et al. "Design and synthesis of indolo [2, 3-a] quinolizin-7-one inhibitors of the ZipA–FtsZ interaction."*Bioorganic & medicinal chemistry letters* 14, no. 6 (2004): 1427-1431.

Joseph, S., Cetinkaya, E., Drahovska, H., Levican, A., Figueras, M.J. and Forsythe, S.J., 2012. Cronobacter condimenti sp. nov., isolated from spiced meat, and Cronobacter universalis sp. nov., a species designation for Cronobacter sp. genomospecies 1, recovered from a leg infection, water and food ingredients. *International journal of systematic and evolutionary microbiology*, *62*(Pt 6), pp.1277-1283.

Kanemori, M., Yanagi, H. and Yura, T., 1999. The ATP-dependent HslVU/ClpQY protease participates in turnover of cell division inhibitor SulA in Escherichia coli. *Journal of bacteriology*, *181*(12), pp.3674-3680.

Kaur, S., Modi, N.H., Panda, D. and Roy, N., 2010. Probing the binding site of curcumin in Escherichia coli and Bacillus subtilis FtsZ–a structural insight to unveil antibacterial activity of curcumin. *European journal of medicinal chemistry*, *45*(9), pp.4209-4214.

Keffer, J.L., Huecas, S., Hammill, J.T., Wipf, P., Andreu, J.M. and Bewley, C.A., 2013. Chrysophaentins are competitive inhibitors of FtsZ and inhibit Z-ring formation in live bacteria. *Bioorganic & medicinal chemistry*, *21*(18), pp.5673-5678.

Khafizov, K., Madrid-Aliste, C., Almo, S.C. and Fiser, A., 2014. Trends in structural coverage of the protein universe and the impact of the Protein Structure Initiative. *Proceedings of the National Academy of Sciences*,*111*(10), pp.3733-3738.

Knox, C., Law, V., Jewison, T., Liu, P., Ly, S., Frolkis, A., Pon, A., Banco, K., Mak, C., Neveu, V. and Djoumbou, Y., 2011. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic acids research*, *39*(suppl 1), pp.D1035-D1041.

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.

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.

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

Miller, E.N., Jarboe, L.R., Turner, P.C., Pharkya, P., Yomano, L.P., York, S.W., Nunn, D., Shanmugam, K.T. and Ingram, L.O., 2009. Furfural inhibits growth by limiting sulfur assimilation in ethanologenic Escherichia coli strain LY180. *Applied and environmental microbiology*, *75*(19), pp.6132-6141.

Mittal, R., Wang, Y., Hunter, C.J., Gonzalez-Gomez, I. and Prasadarao, N.V., 2009. Brain damage in newborn rat model of meningitis by Enterobacter sakazakii: a role for outer membrane protein A. *Laboratory Investigation*, *89*(3), pp.263-277.

Mohan Nair, M.K., Venkitanarayanan, K., Silbart, L.K. and Kim, K.S., 2009. Outer membrane protein A (OmpA) of Cronobacter sakazakii binds fibronectin and contributes to invasion of human brain microvascular endothelial cells. *Foodborne pathogens and disease*, *6*(4), pp.495-501.

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.

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

Moult, J., Fidelis, K., Kryshtafovych, A., Schwede, T. and Tramontano, A., 2014. Critical assessment of methods of protein structure prediction (CASP)—round x. *Proteins: Structure, Function, and Bioinformatics*, *82*(S2), pp.1-6.

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.

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.

Paradis-Bleau, C., Lloyd, A., Sanschagrin, F., Clarke, T., Blewett, A., Bugg, T.D. and Levesque, R.C., 2008. Phage display-derived inhibitor of the essential cell wall biosynthesis enzyme MurF. *BMC biochemistry*, *9*(1), p.33.

Petrenko, R. and Meller, J., 2010. Molecular dynamics. *eLS*.

Pitout, J.D., Moland, E.S., Sanders, C.C., Thomson, K.S. and Fitzsimmons, S.R., 1997. Beta-lactamases and detection of beta-lactam resistance in Enterobacter spp. *Antimicrobial agents and chemotherapy*, *41*(1), pp.35-39.

Prajapati, C. and Bhagat, C., 2012. In-Silico Analysis and Homology Modeling Of Target Proteins for Clostridium Botulinum. *Int J Pharm Sci Drug Res*, *3*, pp.2050-2056.

Pruitt, K.D., Tatusova, T. and Maglott, D.R., 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic acids research*, *35*(suppl 1), pp.D61-D65.

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

Reynolds, R.C., Srivastava, S., Ross, L.J., Suling, W.J. and White, E.L., 2004. A new 2-carbamoyl pteridine that inhibits mycobacterial FtsZ.*Bioorganic & medicinal chemistry letters*, *14*(12), pp.3161-3164.

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.

Salimnia, H., Radia, A., Bernatchez, S., Beveridge, T.J. and Dillon, J.R., 2000. Characterization of the ftsZ cell division gene of Neisseria gonorrhoeae: expression in Escherichia coli and N. gonorrhoeae. *Archives of microbiology*, *173*(1), pp.10-20.

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.

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.

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

Shoukat, K., Rasheed, N. and Sajid, M., 2012. SUBTRACTIVE GENOME ANALYSIS FOR IN SILICO IDENTIFICATION AND CHARACTERIZATION OF NOVEL DRUG TARGETS IN C. trachomatis STRAIN D/UW-3/Cx.

Singamsetty, V.K., Wang, Y., Shimada, H. and Prasadarao, N.V., 2008. Outer membrane protein A expression in Enterobacter sakazakii is required to induce microtubule condensation in human brain microvascular endothelial cells for invasion. *Microbial pathogenesis*, *45*(3), pp.181-191.

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

Stock, I. and Wiedemann, B., 2002. Natural antibiotic susceptibility of Enterobacter amnigenus, Enterobacter cancerogenus, Enterobacter gergoviae and Enterobacter sakazakii strains. *Clinical microbiology and infection*, *8*(9), pp.564-57

Taylor, R.D., Jewsbury, P.J. and Essex, J.W., 2002. A review of protein-small molecule docking methods. *Journal of computer-aided molecular design*, *16*(3), pp.151-166..

Thanki, N., Umrania, Y., Thornton, J.M. and Goodfellow, J.M., 1991. Analysis of protein main-chain solvation as a function of secondary structure. *Journal of molecular biology*, *221*(2), pp.669-691.

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.

Umamaheswari, A., Pradhan, D. and Hemanthkumar, M., 2010. Virtual screening for potential inhibitors of homology modeled Leptospira interrogans MurD ligase. *Journal of chemical biology*, *3*(4), pp.175-187.

Urgaonkar, S., La Pierre, H.S., Meir, I., Lund, H., RayChaudhuri, D. and Shaw, J.T., 2005. Synthesis of antimicrobial natural products targeting FtsZ:(±)-dichamanetin and (±)-2'''-hydroxy-5''-benzylisouvarinol-B. *Organic letters*,*7*(25), pp.5609-5612.

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

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.

Wadood, A., Ahmed, N., Shah, L., Ahmad, A., Hassan, H. and Shams, S., In-silico drug design: An approach which revolutionarised the drug discovery process.

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

Wang, J., Galgoci, A., Kodali, S., Herath, K.B., Jayasuriya, H., Dorso, K., Vicente, F., González, A., Cully, D., Bramhill, D. and Singh, S., 2003. Discovery of a small molecule that inhibits cell division by blocking FtsZ, a novel therapeutic target of antibiotics. *Journal of Biological Chemistry*,*278*(45), pp.44424-44428.

Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A. and Case, D.A., 2004. Development and testing of a general amber force field. *Journal of computational chemistry*, *25*(9), pp.1157-1174.

Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A. and Case, D.A., 2004. Development and testing of a general amber force field. *Journal of computational chemistry*, *25*(9), pp.1157-1174.

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.

World Health Organization (2014). Antimicrobial resistance: global report on surveillance.

Yan, Q. and Fanning, S., 2015. Strategies for the identification and tracking of Cronobacter species: an opportunistic pathogen of concern to neonatal health. *Frontiers in pediatrics*, *3*.

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.

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.

Zhao, Z., Wang, L., Wang, B., Liang, H., Ye, Q. and Zeng, M., 2014. Complete genome sequence of Cronobacter sakazakii strain CMCC 45402.*Genome announcements*, *2*(1), pp.e01139-13.

Zhou, X., Fu, S., Gao, J. and Chen, H., 2012. Enterobacter sakazakii: an emerging foodborne pathogenic bacterium. *Annals of microbiology*, *62*(1), pp.1-5.

Thesis

by Fouzia Hussain

FILE	INTRODUCTION_OF_FTSZ_FINALIZEEEEED.DOCX (78.08K)		
TIME SUBMITTED	21-JAN-2016 09:58AM	WORD COUNT	9879
SUBMISSION ID	622464216	CHARACTER COUNT	57291

1. INTRODUCTION

Scientific world is waging a war against antibacterial agents and bacterial resistance since discovery of penicillin. Antibiotic resistance is a grave threat to global health which demands intense need to grapple this huge and complex problem. Antibiotic resistance has reached to an alarming point (WHO,2014) Antibiotic resistance is due to deploying them wrongly enable bacteria to boost resistance, ultimately depleting the effective stock of antibiotics. Selection pressure is also a leading contributor to emergence of resistant bacterial strains. Morbidity and mortality rates are increased due to the widespread of multiple drug resistant (MDR) strains, which poses a serious threat to public health sector. The conventional methods are also ineffective due to formidable challenges created by MDR bacterial infections. Therefore, antibacterial drug development needs to be continued so that novel and effective antibiotics can be developed to control serious health menace caused by multi-drug resistance organisms worldwide. In the current study subtractive genomics and in silico proteome analysis was carried out on multiple strains of emerging pathogen Cronobacter Sakazakii to identify candidate drug targets.

1.1 The Genus Cronobacter.

Cronobacter is the officially known bacterial genus name for the organism which before 2007 was named as Enterobacter sakazakii. Cronobacter has been accepted as a genus in International Journal of Systematic and Evolutionary Microbiology the genus has six genomospecies and the classification of these organisms was revised based on a detailed polyphasic taxonomic study. Cronobacter.Sakazakii is named after the Japanese bacteriologist Richii Sakazakii (Iverson C et al., 2008;Joseph et al., 2011).

1.2 Cronobacter Sakazakii

Cronobacter sakazakii belongs to Enterobacteriaceae family. It is formerly known as yellow pigmented Enterobacter sakazakii is a Gram-negative bacillus, rod shaped, peritrichous, non-spore forming, human opportunistic and food borne pathogen. The primary reservoir of Cronobacter Sakazakii has yet to be determined but it is postulated that plant material may be an important source.Cronobacter Sakazakii can be isolated from a wide variety of foods including milk, cheese, dried foods, meats, water, vegetables, rice, bread, tea, herbs, spices, and powdered infant formula. Scrutiny studies

have spotted Cronobacter Sakazakii in households, livestock facilities, food factories, and PIF production facilities.(Yan et al., 2015) Clinically, Cronobacter Sakzakii has been isolated from cerebrospinal fluid, bone marrow, blood, intestinal and respiratory tracts, urine, ear and eye swabs, and skin wounds (Healy B et al., 2010; Iversen C et al.,2007).

1.2.1 Genomic features

The complete genomes of three strains of Cronobacter Sakazakii were published recently in 2013 which is available on National Center for Biotechnology Information (NCBI) (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Detail of genomes of all three strains (ES15, CMCC45402, and SP291) is presented in Table 1.1.

1.2.2 Pathogenicity and Mortality rate

The high resistance of C. Sakazakii to unusually dry conditions supports a high survival rate in powdered infant formula.Cronobacter Sakazakii is having a greater ability of stress response due to its adaptive physiological traits. (Zhao et al., 2014;).Cronobacter Sakazakii has viscous capsular and film barrier formation that assist in its adherence to all cell types and resist desiccation for a long interval of time in harsh conditions.(Iversen C et al.,2004). In many studies the significant contribution of OmpA to the virulence potential of Cronobacter Sakazakii is revealed. (Singamsetty et al., 2008; Mittal et al., 2009; Mohan Nair et al., 2009).Cronobacter Sakazakii has been recognized as a critical pathogenic bacterium, especially for powdered formula-fed infants, due to high (up to 80%) mortality rate (Drudy D et al., 2006; Healy B et al., 2010),primarily in newborns (Iversen and Forsythe, 2003 Bar-Oz et al, 2001).

Due of its pathogenicity and virulence ,it mainly cause meningitis with a mortality rate of 40 to 80% and necrotizing enterocolitis with mortality rate of 10 to 55%, septicemia and bacteremia (Forsythe SJ., 2005)Cronobacter Sakazakii infections are also associated with significant morbidity. Those children who sustain Cronobacter Sakazakii related meningitis (94%) develop impaired sight and hearing, developmental impedance and irreparable, widevariety of sequelae such as quadriplegia. These sequelae are frequently attributed to secondary cerebral infarcts. (Drudy D et al., 2006)

FDA has spotted the presence of Cronobacter Sakazakii in baby formula in 2002. (CDC, 2003). Cronobacter Sakazakii infections are worrisome in infant population. The

mortality rate in infected infants due to Cronobacter Sakazakii is 40%–80 %.(Bowen et al., 2006).However case are also reported in immunocompromised adults as well. The colonies of Cronobacter Sakazakii are also isolated from the mouths of strokes victims suffering with pneumonia (Gosney MA et al., 2006)

1.2.3 Multi drug resistance

Different antibiotics are employed to control this pathogen but recent studies investigated emergence of antibiotic-resistant strains (Zhou X et al., 2011).Cronobacter Sakazakii is normally resistant to all macrolids including Fosfomycin, Fusidicacid, Rifampicin, Streptogramins, Clindamycins, Leucomycins, Tetrcylins and Aminoglycosides. (Stock I.Wiedemann B., 2002).Cronobacter Sakazakii is resistant to Ampicillin due to acquisition of transposable elements and production of beta-lactamase and Cephalosporinase which inactivate third generation cephalosporin and wide spectrum of penicillin (Pitout JD et al.,1997;Girlich D et al.,2001).

Computer aided drug designing provide alternative and cost effective approaches for addressing the problems of resistive infections. These approaches are proving to be very proficient in identification of pathogen drug targets and then development of targeted and robust drugs for those targets. The idea of interpretation and analysis of large amount of available proteome sequences data through various in silico approaches has revolutionized the drug discovery process.

1.3 In silico Approach in Current study

The current study includes different in silico approaches which can be integrated into various in silico approaches. These approaches follow a specific sequential route: subtractive genomics approach, selection of therapeutic target, homology modeling, molecular docking and molecular dynamic simulations as shown in Fig 1.1.

1.3.1 Subtractive Genomics Approach

The conventional approaches of drug development result in dead ends or host toxicity which can be replaced by applying the in silico subtractive genomics approach to design potential drug targets against bacterial desease.Subtractive genomics basically deals with identification of novel targets ,essential for pathogen survival and important component of various metabolic pathways .these targets should not be in homology with the host so

that a drug or lead compound will be designed against these targets. This approach is fast, robust cost effective. The speed of drug discovery process is increased due to rapid increase in availability of genome sequences of various pathogens.This approach has successfully applied in identification of drug targets in pathogens like Mycobacterium tuberculosis (Hosen et al., 2014), Leptospira Interrogans (Pradhan et al., 2013), Streptococcus pneumoniae (Munikumar et al., 2013), Clostridium botulinum (Prajapati and Bhagat, 2012; Reddy and Rao, 2012) Mycoplasma Genitalium (Butt et al., 2012), and Salmonella Typhi (Rathi et al., 2009). Mycobacterium Leprae (Shanmugam and Natarajan, 2010), and Mycoplasma Pneumonia (Gupta et al., 2010).

In the current study subtractive genomics approach has been utilized on multiple strains of Cronobacter Sakazakii: ES15, SP296, and CMCC45402 to identify putative drug targets. That will facilitate not only in understanding the biology of the infectious pathogen but will also help in developing a more effective medication. Different pipelines and filters are employed in the approach to find out important targets imperative for pathogen survival which could be easily targeted in human host due to lack of homology of these targets in human host. The designated drugs should effect the given target without adverse effects in host. Hence optimal targets are obtained in this way which provide selectivity. Thus, optimal target obtained from screening of a set of essential proteins are not only crucial for the viability of the organism, but also provide selectivity.

The drug target identified and characterized in the current study is bacterial cell division protein (FtsZ).FtsZ is selected as it is involved in cell division of bacterial cell which can be targeted to inhabit the growth of bacteria specifically in human host. FtsZ is an important protein of cell division cycle of almost all bacterial strain known yet.

1.3.2 Bacterial cell division protein (FtsZ)

FtsZ is an important cell division protein which makes a contractile ring at the bacterial cell division site. The ring assembly regulation controls the timing and location of cell division. The function of FtsZ is to recruit the other proteins involved in cell division to produce a new cell wall between the dividing cells. Subtractive genomic approach predicted FtsZ as an ideal drug target for novel antibacterial drugs that is common in all three strains of Cronobacter Sakazakii.It is found to be conserved among various

bacterial species .FtsZ has a weak homology with human tubulins (salimnia et al . 2000). FtsZ taken from one bacterial specie can work better in other. The functional properties and its interaction with other proteins enables it as a potential therapeutic drug target, to find lead molecule with enhanced efficacy and reduced toxicity(Kapoor et al.,2009). The approximate molecular weight of FtsZ is 50kDa.

1.3.3 Homology modelling

Since last few decades there is an enormous increase in the sequence information at both, genome and proteome level, due to availability of high throughput sequencing techniques and prompt annotation of data (Khafizov et al., 2014).biological knowledge is more easily accessible due to large number of interlinked data repositories. However the problem arises in understanding biological data due to lack of structural information regarding the biomolecules. As of May 2014, Universal Protein Resource KnowledgeBase (UniProtKB) (Boutet et al., 2007) has 56555610 protein sequences reported in May 2014 while in Protein Data Bank (PDB) there are 79,356 experimental protein structures and this number is increasing exponentially (Westbrook et al., 2003)Comparative homology modelling is a valuable tool to close the gap between sequence and structural space (Moult et al., 2014).Hence a theoretical model can be generated if a known template structure is available with a suitable similarity is present . (Krieger et al., 2003). The structure of target protein FtsZ from C.sakazakii is unavailable in the Protein Databank (PDB). Therefore, protein modeling was performed using MODELLER and various other web-based tools including SWISS-MODEL (Schwede et al., 2003), ModWeb (Pieper et al., 2004), I-TASSER (Zhang, 2008), and 3D-JIGSAW (Bates et al., 2001). A whole comparison is also between MODELLER and other web servers is also carried out. PROCHECK (Laskowski et al., 1993), ProSA (Sippl, 1993; Wiederstein & Sippl, 2007), Verify3D (Eisenberg et al., 1997), ERRAT value (Colovos & Yeates, 1993), G-Factor and Bad Contacts (Morris et al., 1992) were measured and the best model has been selected after evaluating all generated models from the above sources.

1.3.4 Molecular docking

Molecular docking is a powerful approach for structure based drug discovery. (Xuan-Yu Meng, et al., 2011)Molecular docking is an important tool in CADD (computer aided drug design) and structural molecular biology which is employed to predict the binding mode of three dimensional known structure of target protein and ligand. (Morris GMet al., 2008; Lengauer T et al., 1996).The current study incorporates molecular docking to predict the inhibition mechanism of target protein. The information about the inhibitors and the active site of target macromolecule has been obtained after an in depth study of literature.

Molecular docking comprises of two interrelated steps. first generating all possible binding configurations of ligand with receptor molecule through search algorithms and then prioritizing the binding modes to predict the most feasible and accurate binding conformation through scoring function (Morris and Lim-Wilby, 2008; Taylor et al., 2002). MD simulations are utilized in the current study to further explore the exact binding conformations and time dependent behavior of protein-ligand complex.

1.3.5 Molecular dynamic simulations

In late 1950's Alder and Wainwright introduced MD methods which were employed in study of interaction of hard spheres (Alder & Wainwright, 1959). Later in 1964 Rahman carried out molecular dynamic simulation study for liquid argon utilizing Lennard-john potential. The first protein which was simulated was bovine pancreatic trypsin inhibitor (BPTI) in 1977 (McCammon, et al, 1977).Molecular dynamic simulation is a principal tool in the theoretical study of biological molecules, using simple approximations based on Newtonian physics.it calculates time dependent behavior of a system.MD provide detailed information about conformational changes and fluctuations of nucleic acid and proteins.

MD method has two major families which can be categorized on basis of model selected to describe a physical system. The classical mechanic approach in MD simulations defines the dynamics of the system while quantum approach in MD simulations deals with the nature of chemical bond. Quantum MD simulations are more powerful than classical mechanic approach as it deals with diverse biological problem. Due to computational complexity of quantum approach, classical mechanics is practical at

present for simulations of biological systems containing many thousands of atoms over a time scale of nanoseconds. (Jarosaw et al., 2015). Over the past MD has emerged as a major area in understanding the atomic basis of complex biological systems such as transport of small molecules and ions across membranes, protein folding and molecular recognition. Along with experimental approaches, MD simulation has a better understanding of protein structure-function relationship and manifest the drug discovery process. (Pramod C et al., 2014).

1.3.5.1 Statistical Mechanics – Foundation of Theoretical Framework.

Molecular dynamics simulations produce data at the microscopic level of a system, comprising atomic positions and velocities. The transformation of this microscopic data to macroscopic information such as heat capacities, pressure, energy etc., utilizes statistical mechanics. Hence the study of molecular dynamic of biological systems require statistical mechanics.(Wereszczynski & McCammon, 2012).Statistical mechanics makes a connection between microscopic simulation and macroscopic properties of system under study. The collection of all possible systems which have identical thermodynamic state but different microscopic state constitute an ensemble. (Wilde and Singh, 1998).

Four major ensembles are present in statistical mechanics each with varying properties. (Chandler, 1987.

- **Micro canonical ensemble (NVE)** – constant atomic number of particles N, constant volume V and constant amount of energy E.
- **Grand canonical ensemble (μVT)** - constant chemical potential μ, constant volume V and constant value of temperature T.
- **Canonical ensemble (NVT)** - constant atomic number of particles N, constant volume V and constant value of temperature T.
- **Isobaric-isothermal ensemble (NPT)** – constant atomic number of particles N, constant pressure P and constant value of temperature T.

- **1.3.5.3 Molecular Mechanics – Calculation of Molecular Structure**

To compensate the need of description of molecular structure and properties practically, molecular mechanics was developed, which is an extension of classical mechanics. The limit of applicability of molecular mechanics depend on following factors.

- Molecules comprising thousands of atoms.
- and saccharides , peptides, Organics, oligonucleotides
- Vacuum, explicit, or implicit solvent environments.
- Ground state only.
- Thermodynamic and kinetic (via molecular dynamics) properties.

The powerful computational speed of molecular mechanics makes it useful in molecular dynamics, docking and conformational energy searching.

Molecular mechanics is based on following principles.

- Nuclei and electron are united into atom-like particles.
- Atom-like particles have a net charge and spherical
- Interactions are based on classical potentials and springs.
- Interactions should be allotted to specific sets of atoms.
- Interactions calculate the spatial distribution of atom-like particles and their energies.

1.4 Aims and objective

Dynamic extension of computer aided drug design methodologies has encouraged to relinquish numerous barriers of drug discovery process. Current study aims at finding the potential drug targets in human pathogen. The journey for new antibacterial therapeutics is because of the sensational heightening of Cronobacter Sakazakii resistance against numerous anti-toxins. Subtractive genomics approach on multiple strains of Cronobacter Sakazakii is applied for identification of potential drug targets which was selected to light the potent inhibitors against it.Hence modelling was performed to anticipate the structure of target protein. Molecular docking was done that gives knowledge into the binding association of protein with its inhibitors. The work was further supported by applying molecular dynamics on docked protein targets. Collectively, molecular docking studies together with MD simulations carried out will help in understanding the binding process of FtsZ inhibitors and will give further insights into structural alterations generated to

synthesize more powerful and selective inhibitors. Collectively, molecular docking concentrates together with MD reenactment will help in understanding the binding interactions of FtsZ inhibitors and dynamics of proposed target with candidate inhibitor, respectively.It will give further bits of knowledge into basic structural changes created to synthesize more effective and particular inhibitors.

2. METHODOLOGY

2.1 System information

Computational structure comprises of both hardware and software components guarantee smooth execution of of the research activities. For major portion of the current work, computational facilities provided at the Computational Biology Lab, National Center for Bioinformatics, Quaid-i-Azam University, Islamabad, Pakistan, were utilized. The entire procedure was divided into four stage; in silico subtractive genomics approach, homology modeling, molecular docking analysis and molecular dynamics simulation. Complete work was done using Intel (R) Core(TM) 2 Duo CPU E8600 @ 3.33 GHZ and the operating system used was Linux open SUSE 11.4.The production run of 100 nanoseconds for docked protein was carried out on high performance computer cluster in above mentioned lab (Figure 2.1).

2.2 Applied *In silico* Approach

The present strategy follow an in silico protocol which is partitioned into different modules, which used different soft wares and databases.At first genomes multiple strains of Cronobacter Sakazakii (ES15.CMCC45402.SP291) were compared using a technique of subtractive genomics which directed to identify the pathogenic target. In second step homology modeling was done to anticipate the structure of potent target. Molecular docking was done to study system interactions with inhibitors which drove towards structure based drug designing. Finally MD simulations were performed to study time dependent behavior of target. The overall workflow of this study is shown in flowchart which given in Figure 2.2.

2.2.1 Subtractive Genomics Approach

Subtractive genomics approach is a standout amongst the most currently used methodologies, which is used in this study to predict the essential proteins in genome of Cronobacter Sakazakii. This methodology involved the steps given in Figure 2.3.

2.2.1.1 Complete proteome retrieval and paralog removal

Complete proteome of multiple strains: ES15, CMCC45402, and SP296 of Cronobacter Sakazakii were retrieved from Uniprot (http://www.uniprot.org/) in fasta format. Uniprot consists of manually curated protein sequences under UniProtKB/Swiss-Prot and automated curated protein sequences under UniProtKB/TrEMBL (Boutet et al., 2007). Web server CD HIT Suite (Cluster Database at High Identity with Tolerance) was used (Li et al., 2001) to remove the redundancy of proteins from C.Sakazakii strains. The cutoff value was set as 0.6 (60%) along with clustering and alignment parameters set as default to remove paralogues proteins. Finally non redundant sequences were obtained for further processing.

2.2.1.2 Retrieving of non-homologous protein

Removal of homologous sequences was done using perl script provided at Computational Biology Lab of National Centre for Bioinformatics, Quaid-i-Azam University Islamabad, Pakistan.in the initial step whole proteome of the pathogen was aligned against human proteome using Basic Local Alignment Search Tool (BLASTp) at threshold expectation value of (E- value) of 10-4 (Altschul et al., 1990). Refseq NCBI database (Pruitt et al., 2007) was locally remade on a server; these databases were then subjected to automated Blast search using Blast+ (Camacho et al., 2013) and indigenously made Perl script,which resulted in the removal of homologs and compilation of non-human homologs for further analysis.The Perl script takes less time for investigating and filtering of the sequences in multiple strains. It makes the work fast like if manually the second step was done, it takes 2 weeks to a month for one genomeIn correlation to this, scripts make the work finished in a couple of hours .

2.2.1.3 Identification of essential genes

Database of Essential Genes (DEG) (http://tubic.tju.edu.cn/deg/) (Zhang et al., 2004) was used to outline set of essential C.sakazakii, non-human homologs. Essential gene filtering is an essential step since it is not much important that all homologs proteins obtain are also crucial to support life. . Database of Essential Genes (DEG) (Zhang *etal*., 2004) is repository that are integral to the life form survival, containing essentiality information

for 43 representative eukaryotic and prokaryotic organisms (Zhang & Lin, 2009). Prokaryotic section of the database was taken as reference and BLASTp was done with E-value cut off at 10-4, bit score = 100 and sequence identity of >= 30 %. Cronobacter Sakazakii proteins.All parameters were set to the Perl script. Resultant list of fundamental proteins for all three strains were obtained by utilizing per script.

2.2.1.4 Metabolic pathway analysis

The next step of screening procedure is to classify Essential non-homologous proteins functionally and assigning them metabolic pathways using the KAAS (KEGG Automatic Annotation Server) tool (http://www.genome.jp/kegg/kaas/), which is supported by Kyoto Encyclopedia of Genes and Genomes (KEGG) database (http://www.genome.jp/kegg/).KEGG is conspicuously known for the assortment of information collection it holds among others, including genome information and functional annotation.KAAS specifically is a tool that allots KEGG orthology (KO) numbers to the proteins by applying BLASTp to search against the KEGG GENES repository. The sequences under study are assigned metabolic pathways on basis of KO numbers (Moriya et al., 2007).

The essential non homologous proteins of Cronobacter Sakazakii genome obtained after search against DEG, were in this way submitted to KAAS utilizing Cronobacter Sakazakii genome as reference data set to expand specificity of results.Since just a part of the complete pathogen proteome was being assessed, single-directional best hit system was used.The predicted pathogenic pathways were manually coordinated to the human pathways to arrange them into two classifications:common and unique.Common pathways were those present in both the host and bacterium.Though, unique, were selective to bacterium and thus, center of current work.

2.2.1.5 Druggability Assessment using DrugBank

The DrugBank database is a unique bioinformatics and cheminformatics asset that joins elaborated drug (i.e. compound, pharmacological and pharmaceutical) information with far reaching medication target (i.e. sequence, structure, and pathway) dataavailable on http://www.drugbank.ca/ . Next step in genome subtraction was screening of chosen target proteins by means of Drug Bank. Drug Bank version 3.0 was employed with

default parameters to assess the druggability capability of non-homologous essential proteins by performing BLASTp against all the essential protein targets of drug compounds within Drug Bank (Knox et al., 2011).Drug bank contain 1634 FDA approved small molecule drugs, 89 nutraceuticals, 169 FDA approved biotech drugs and more than 6000 experimental drugs.

2.2.1.6 Subcellular Localization Prediction

In the last step the retrieved sequences were examined to recognize their subcellular location of the unique, and essential protein targets.for this purpose PSORTb 3.0.2 (http://www.psort.org/psortb/) and subCELlular LOcalization predictor (CELLO) version 2.5 (http://cello.life.nctu.edu.tw/) were utilized.

PSORTb a pioneer exertion, keeps up a database of localization data on an extensive variety of bacterial and archeal species.It appliesSupport Vector Machines (SVMs), a machine learning method that searches the curated dataset to foresee the localization through the utilization of suffix tree algorithm. Bayesian classifier is employed in last step allocates a probable localization site to a protein (Nancy et al., 2010). Additionally, it allocates the five subcellular areas i.e. cytoplasm, inner membrane, periplasm, outer membrane an extracellular, to gram-negative microorganisms (Yu et al., 2010). Basically cytoplasmic proteins are selected as they have potential of becoming possible drug targets (Shoukat et al., 2012).Similar to PSORTb, CELLO also utilizes the functionality of SVMs to predict the probability of the subcellular location (Yu et al., 2006).

2.2.2 Selection of Druggable Target

Druggable essential proteins which were important in unique and essential metabolic pathways amongst different strains of Cronobacter Sakazakii were identified and those which were common in all three were further used for analysis. At last, FtsZ playing a vital role in the cell cycle caulobacter pathway were selected for further in silico analysis. UniprotKB was used to retrieve information regarding the functional role, active isoforms, cofactors, subunit structure and associated post translational modifications.Furthurmore, the presence of experimental structure was established by predicting the details about the target and in case of absence of such structure, the possibility of comparative homology based modeling was considered by assessing the

template accessibility. Structural templates that showed at least 30% identity with > 90% query coverage were considered satisfactory. This assessment was done by comparing protein sequence against the structural resource: PDB,using BLASTp functionality supported by NCBI . All these steps were carried out to select current target for further CADD analysis.

2.2.3 Model Construction, Refinement, and Validation

Comparative modelling was done to retrieve a homology based structure as the three dimensional crystallographic was not available for the selected target. The amino acid sequence for Cronobacter Sakazakii CMCC45402- FtsZ (Uniprot ID: V5U3P9) was taken from UniprotKB and compared to the PDB via alignment through BLASTp. The X-ray crystallographic structure for GDP bound FtsZ from Pseudomonas aeruginosa GDP (PDB ID:2VAW) (Olivia *et al.*, 2007), with sequence identity of 57% and query coverage of 94% was fulfilling the criteria of both the query coverage and sequence identity was selected for model building procedure. MODELLER (Eswar et al., 2008) version 9.10 was used to build the structural model of target protein In addition to MODELLER9.10, structure for the FtsZ, was also obtained for comparative purposes through five web servers: SWISS-MODEL (Schwede et al., 2003), ModWeb (Pieper et al., 2004), I-TASSER (Wu et al., 2007; Zhang, 2008), 3D-JIGSAW (Bates et al., 2001) and ESyPred3D (Lambert et al., 2002).

Validation is required to check the reliability of model generated. Models were evaluated using PDBSum (Laskowski, 2001), ERRAT value (Colovos and Yeates, 1993), Verify3D (Eisenberg *et al.*, 1997), ProSA (Wiederstein and Sippl, 2007), G-factor and bad contacts (Morris *et al.*, 1992). All these measures were used to make selection of the model. The results obtained after evaluation were used to select the best model. Functional basis of these structure assessment tools is elaborated further.

2.2.3.1 MODELLER

MODELLER is used to retrieve three-dimensional comparative structure of protein through homology or comparative modelling .The sequence is aligned with known related structure are provided as input and MODELLER generates a model containing all non-hydrogen atoms.modeller also perform energy structure comparison ,function calculations,and minimization.

2.2.3.2 SWISSMODEL

SWISSMODEL is an automated server. It is used for homology based protein modeling. It can be accessed through ExPASY web server and Swiss PDB-Viewer (http://swissmodel.expasy.org/).

2.2.3.4 I-TASSER

I-Tasser is an online server .It is used for the predictions of protein structures and functions. It can be accessed from http://zhanglab.ccmb.med.umich.edu/I-TASSER/. Three dimensional structures are generated on the basis of multiple-threading alignments by LOMETS. Function is predicted by matching 3D structure with BioLiP protein function database. It can be accessed from: http://zhanglab.ccmb.med.umich.edu/I-TASSER/.

2.2.3.5 MODWEB

ModWeb is an online server .It is mainly used for comparative protein structure modeling. For proper functionality MODWEB depends on ModPipe which is a large scale protein structure modeling pipeline. It is available on URL: https://modbase.compbio.ucsf.edu/scgi/modweb.cgi.

2.2.3.6 3D-JIGSAW

The 3D-JIGSAW is an automatic server. It uses comparative modelling for prediction of structure and function against given protein sequence. It can be accessed from: http://bmm.cancerresearchuk.org/~3djigsaw/.

2.2.3.7 PROCHECK

A PROCHECK calculates stereo chemical property of a protein models like Ramachandranplot, Bad contacts and G-Factor . It estimates the energy of each residue of the overall model and results are shown in the form of a graph.

2.2.3.8 Verify3D

Verify3D is used in model evaluation to refine the structure. Each residue is alloted a structural preference based on its location and environment.

2.2.3.9 Errat

Errat is used in evaluation and refining the structure of model. It statistically inspect the non-bonded interactions among different atom types.Output is given in form of a plot of error function versus to position of residue.

2.2.3.10 ProSA-web

ProSA-web is an online tool used for validation of protein structure. The overall quality of protein structure is estimated on the basis of z-score. Thus, it is also capable to validate the low resolution structures.

2.2.4 Energy Minimization

The best model of target protein was then to energy minimization process to improve its quality. UCSF Chimera (Pettersen *et al.*, 2004) was used to minimize energy. Gasteiger charges was assigned to protein and structural constraints was removed by 1500 rounds of minimization runs (750 steepest descent followed by 750 conjugate gradient) with a step size of 0.02 Å, under TriposForce Field (TFF) (Pettersen et al., 2004; Sanner et al., 1996). The minimized protein model thus obtained was evaluated through validation process. Finally the minimized best model was further taken for docking studies.

2.2.5 Molecular Docking Protocol

Docking protocol is divided into following steps: Active site identification, Inhibitors/ligands preparation and Molecular docking. . Detailed methodology of these steps is given in following sections

2.2.5.1 Active Site Identification

Active site determination was carried out for the modeled protein to further work on its docking studies. Active site determination creates an idea to build a grid before docking. Active site of FtsZ was identified through extensive literature search and an online tool DoGSiteScorer.It predicts binding pocket on the basis of druggability (Volkamer *et al.*, 2012) and accessed at http://dogsite.zbh.uni-hamburg.de/, was used for further validation. Active site mentioned in literature was then predicted in our target sequence manually. Active site residue were also confirmed by multiple sequence alignment (Edgar and Batzoglou. 2006) using ClustalO 2.0(Sievers et al., 2011).

2.2.5.2 Ligand Preparation

A total 102 potential inhibitors with reported activities against FtsZ protein of several bacteria were identified from thorough literature search (Reynolds et al., 2004; Lioyd et al., 2008; Desiree et al; 2006; Catherine et al., 2006; Danielle et al., 2004; Neil et al., 2008; Jarosiaw et al., 2010; Shanmugham et al., 2011; Alberto et al., 2010; Jessica et al., 2013; Prerna, 2007; Haydan et al., 2010; Kumar et al., 2010; Jun wang et al., 2003; Takashi et al., 2014; Qing et al., 2008; waldemar vollmer ., 2006; Tushar et al., 2006; Lucile et al., 2002; David et al., 2009; Richa et al., 2006).. the 2D chemical structures of these inhibitors were drawn using ChemDraw Ultra 8.0 in ChemOffice 2012 packages (Li et al., 2004)which were further used in docking procedure. Standard bond angles, lengths and charges were used for structure of all inhibitors. The 2D structures drawn were further minimized using Chem3D Pro 12.0 using MM2 force field.

2.2.5.3 Molecular Docking

Molecular docking was carried out using the minimized protein and minimizaed ligand molecules. The docking software utilized are Genetic Optimization for Ligand Docking (GOLD) (Jones et al., 1997) and AutoDock Vina (Trott & Olson, 2010). GoldScore and binding affinities were calculated from each ligand respectively .Hence on the basis of GoldScore best docked ligands were selected. LIGPLOT (Wallace et al., 1995), Visual Molecular Dynamics (VMD) (Humphrey et al., 1996), UCSF Chimera (Pettersen et al., 2004), and Discovery Studio (DS) Visualizer 3.5 (Accelrys Software Inc., 2012), software were used for visualization and interaction studies of ligands in detail.

2.2.5.4 GOLD

GOLD Hermes package is a comprehensive docking setup wizard used in docking of flexible ligand with a partial flexible protein (Jones et al., 1995).The standard default parameters were kept for docking poses of complexes such as population size is 100, niche size is 2, selection pressure is 1.1, operator weights for migrate is 0, crossover is 100, number of islands is 1, number of operations is 10,000, number of dockings is 10. Hydrogen atoms were added in the protein model.

GOLD uses genetic algorithm and it implements four user defined scoring functions, namely , GoldScore, ChemScore, CHEMPLP and User Defined Score. GoldScore was used in the following study, which include energy terms describing the van der Waals interaction, hydrogen bonds and torsional energy (Verdonk et al., 2003). Gold Score function is given mathematically as follow:

Scoring function = $S_{hb\_ext} + S_{vdw\_ext} + S_{hb\_int} + S_{vdw\_int} + S_{tor}$

Where

S_{hb_ext} represents: protein ligand hydrogen bond score

S_{vdw_ext} represents: protein ligand van der Waals score

S_{hb_int} represents: fitness because of intramolecular hydrogen bond

S_{vdw_int} represents: intramolecular, ligand van der Waals energy

S_{tor} represents: ligands torsional energy

Gold- Score fitness function was used as criteria to select the best conformation in current study

2.2.5.5 Autodock Vina

AutoDock Vina is used for molecular docking which implements Broyden-Fletcher-Goldfarb-Shanno (BFGS) method for local optimization (Trott and Olson, 2010).. Docking poses are generated by AutoDock Vina.AutoDock (ADT tools 1.5.4 was used to generate th PdbQts of the protein.The polar hydrogen atoms were added initially followed by addition of Kollman and Gasteiger charges.The file was saved as "pdbqt" file format. Graphical interface of Raccoon was used to generate Ligand PDBQT files (Forli, 2010).A grid box with a number of points of 20 Å × 20 Å × 20 Å box in x,y,z dimensions was created to enclose the ligand. Grid centered at x,y,z dimensions, was selected with values of 23.558, 34.790 and 10.381 respectively, with 1 Å box spacing. The grid box parameters were defined for protein and ligand in configuration file .during the docking procedure both ligand and protein were kept rigid. Ligands having lowest binding affinity were selected to make poses with receptor which were taken in further study.

2.2.6 Molecular Dynamics Simulation

Molecular dynamics simulation study was carried out to analyze the dynamic behavior of docked protein. Simulation was carried out by using Simulated Annealing with NMR-Derived Energy Restraints (SANDER) module in AMBER(Assisted Model Building with Energy Refinement) software (Case et al., 2005). Docked protein simulation was carried out on Intel (R) Core(TM) 2 Duo CPU E8600 @ 3.33 GHZ with Linux open SUSE 11.4 as operating system. Simulation studies comprises of following five steps illustrated in Figure 2.4.

2.2.6.1 System preparation

Preprocessing which include minimization is done for system preparation before production run to remove all the steric clashes, equilibrating and heating the system. It involve four different steps: i) minimization ii) heating iii) equilibration iv) production run.The macromolecular system is prepared for simulation using xLEaP which is a graphical user interface to the LEaP module of AMBER 12 (Case et al., 2012; Wang et al., 2004). The Leap module in Amber12 tools was used to record the topology of the

ligand and protein (Case et al., 2005). Force fields used for protein and ligands was ff03.r1, GAFF and ff99SB (Salomon-Ferrer *et al.*, 2013) The system was neutralized by addition of 21 Na+ ions and solvated with three-point transferable intermolecular potential (TIP3P) water box with 8.0 Å (Figure 2.5). Antechamber program was employed for docked protein complex to make sure the correctness of bonds, angles and atom types.

2.2.6.2Minimization, Heating, Equilibration and Production

After the system preparation minimization of complex was done thoroughly to remove unfavorable clashes before production run of MD simulation. 1500 steps for steepest descent method followed by 1000 steps for conjugate gradient were applied at non bonded cutoff value of 8 Å. Langevin dynamics algorithm (Feller *et al.*, 1995) was applied and heating of the system was performed for 10 picoseconds starting slowly from 0 k to a temperature of 300 k and pressure of 1 atm. Equilibration was done at a constant temperature of 300 for 10 picosecounds. Following equilibration production run of 100 ns for the docked complex was done. Total energy of the system remains constant throughout the equilibration except kinetic and potential energy.

Production run was done finally by employing SHAKE algorithm (Ryckaert et al., 1977) for bond constraints. SANDER (Simulated Annealing with NMR Derived Energy Restraints) module was used for production run .following production run, simulation run of 100ns was done. In the simulation box, periodic boundary conditions were applied with canonical ensemble. Berendsen coupling integration algorithm was used to keep temperature constant and consequently, MD simulations were achieved by means of Ewald summation method (Darden et al., 1993).

2.2.6.3 Trajectory Analysis

For analysis of trajectory created after simulation procedure PTRAJ (Process TRAJectory) module of AMBER10 was used to calculate four different properties. In current study following four properties were calculated using PTRAJ and graphical representation were visualized in xmgrace for analytical purpose (Vaught, 1996).

i) Root mean square deviation (RMSD)

ii) Root mean square fluctuation (RMSF)

iii) Radius of gyration (Rg)

iv) Beta factor (B-Factor)

2.2.6.4 Root Mean Square Deviation

The analogy in three-dimensional structure is measured for globular protein conformations by the RMSD of the Cα atomic coordinates after superposition of optimal rigid body. Root Mean Square Deviation (RMSD) is a measure of distance between two atoms of Cα, which are superimposed over each other for a specific frame.

$$RMSD = \sqrt{\frac{1}{N}\Sigma_i \; d_i^2}$$

In the above given equation, N signify the number of atom while di signify the distance between ith pairs of atom.

2.2.6.5 Root Mean Square Fluctuation

Root Mean Square Fluctuation (RMSF) is a measure of averaged root mean square Distance between the positions of atom from its mean position. RMSF calculates the local structural changes along the protein chain. In case of molecular dynamics, it can be defined as the set of positions for an atom achieved over a given time period. Following equation is used to calculate RMSF:

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

Where

T: signify the time interval. xi :signify the position of an atom at a particular time .

x :signify the averaged position of the atom.

2.2.6.6 Beta Factor

Beta factor is a term closely related to RMSF employed for dynamic studies. Local thermal and vibrational movements movements in the system create spatial displacement of atoms around their mean position .RMSF is a measure of these spatial movements of atoms around the reference position. (Kuzmanic & Zagrovic, 2010).following equation shows the Beta Factor measured in terms of RMSF.

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

2.2.6.7 Radius of Gyration

Radius of gyration is a measure the compactness of the macromolecular system including overall packing quality and density of a structure (Goodfellow, 1990). It can be calculated by the following equation:

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

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

3. RESULTS

Methodological design of drug discovery process, generated different screened out results at the end of each steps. A descriptive view of results accomplished over the span of current work is given in the following sections.

3.1 Genome Subtraction

The streamlining of the functional Cronobacter Sakazakii genome is split into a few stages, expected to frame a list of the most conceivable drug targets. A diagram of this progressive screening strategy is given in Fig 3.1.

3.1.1 Genome Retrieval

Three strains of Cronobacter Sakazakii namely, CMCC45402, SP291 and ES15 were selected for current subtractive genomic study. These are completely sequenced strains and their genome was retrieved from UniProtKbcomprised of 4349, 4805, and 4706 number of protein sequences respectively.

3.1.2 Non-Paralogous Proteins

Redundant protein sequences were removed via CD-HIT program at 60% identity leaving 4254, 4688, and 4623 non-paralogous protein sequences in three strains respectively. This led to the prediction of non-paralogous proteins i.e. 4067 in CMCC45402, 4002 in SP291 and 3857 in ES15 were subjected for further subtraction analysis.

3.1.3 Non-Homologous Proteins

These non-paralogous protein sequences were then subjected to Perl script for BLASTp to eliminate homologous protein sequences to human homologous proteins were removed from proteome of CMCC45402, SP291 and ES15 strains. Finally 3616, 3510 and 3260 proteins were left in three strains, respectively.

3.1.4 Pathogen Essential Proteins

Pathogen essential genes are imperative for their survival of pathogen which are predicted by search via DEG database. After this screening, essential genes identified were 1223 in CMCC45402, 1223 in SP291, 1204 in ES15 respectively.

3.1.5 Metabolic Pathway Analysis

Metabolic pathways of the essential proteins were investigated through KEGG Automatic Annotation Sever (KASS) leading to the identification of potential drug targets involved in various crucial metabolic pathways of the pathogen. After accomplishing the KEGG, number of sequences for CMCC45402, SP291, and ES15 were reduced to 596, 499, and 567 respectively.

3.1.6 Druggability Assessment

To explore the relation between screened target and drug binding ability, DrugBank database was used to check the druggability potential of non-homologous essential proteins. This gives the essential drug target proteins for CMCC45402: 249, for SP291: 247, and for ES15: 237.During the course of this screening procedure many novel target proteins were also identified for which no hits were available in DrugBank.

3.1.7 Subcellular Location Prediction

The retrieved sequences from all three strains were further analyzed using PSORTb version 3.0.2 to identify their subcellular locations. In CMCC45402 strain, 10 were cytoplasmic , 15 were membranous , similarly in SP291 strain there were 8, cytoplasmic proteins , 14 were membranous and in ES15 ,6 were cytoplasmic ,18 were membranous. Cytoplasmic proteins were selected for identification of putative drug target as cytoplasmic proteins are mostly consist of enzymes which are important for bacterial growth.

3.2 Drug Target Selection

The final step of the genome subtraction is a starting point for the drug discovery process. The selection of the best drug target from the shortlisted candidates was challenging. Since the aim of the current study was on identification of drug targets rather than vaccine targets. Proteins having cytoplasmic subcellular location were selected for further analysis. Six cytoplasmic proteins were then assessed on the basis of various features as outlined in Table 3.3. In the present study, FtsZ has been identified as potential drug target .it is an important cell division protein. An increase in FtsZ increases cell division and vice versa (Gilson et al, 2001.) among all gram positive bacteria. .FtsZ is essential protein in all three strains of Cronobacter Sakazakii and further study was carried out on FtsZ as it is an effectual and attractive drug target.important features of the selected target protein is listed in Table 3.2.

3.3 Comparative Homology Modeling

Template FtsZ from pseudomonas aueroginosa GDP (PDB ID: 2VAW) was selected as a reference structure that showed maximum coverage of 94% and identity of 57%. Multiple sequence alignment was done using ClustalO (1.2.1).Using ClustalO, alignment was built from sequences of template and protein as input which confirmed the conservation pattern at the sequence level (Figure 3.4). Five models were generated through MODELLER v9.10 along with Discrete Optimized Protein Energy (DOPE) score calculation listed in Table 3.4.Validation of the generated models is carried out by comparative analysis between the models generated by MODELLER and those generated by different automated web-based modeling tools such as SWISS-MODEL, 3D-JIGSAW, I-TASSER and MODWEB.For further validation the generated models were then subjected to (SAVES) Structural Analysis and Verification Server. The comparative analysis was done between web-based servers and model generated by MODELLER is shown in Table 3.4. .PROCHECK generated Ramachandran plot, to inspect bad contacts and steric clashes to select the most accurate and reliable model (Figure 3.3). On basis of quality assessment measures Model 2 was selected which was used for further processing. Model 2 exhibited strong stereochemistry with 94.7% residues in most favored region, 4.5% residue in additional allowed region, 0.4% residues in disallowed regions , and lowest Z score of 9.93 (Table 3.4). Then this model is superimposed with

template (2VAW) and its RMSD value is 0.546 Å, which ensure the similarity of the protein structure with the template (Figure 3.6). The selected model is subjected to energy minimization with the tripos force field (TFF) using UCSF Chimera to relax the overall structure and remove sterric clashes of side chains

3.4 Molecular Docking

Molecular docking aims at predicting the binding modes of ligands with protein which is the most important step as it plays the crucial role in drug designing. FtsZ is composed of three evolutionarily conserved domains which are shown in Figure 3.3.In the present study inhibitors were docked into the active site of protein.

3.4.1 Active Site Identification

Active site information is essential to start with the docking procedure. Active site of FtsZ was predicted through extensive literature search (Kaur H et al., 2010, Heger et al., 2014,Hsin et al.,2012) and DoGSiteScorer. Conserved active site is observed in different orthologous of FtsZ. DoGSiteScorer predicted three pockets and from secound pocket active residue was picked which was same as given in literature.

3.4.2 Secondary structure prediction.

Psipred provides the secondary structure prediction results. Beta sheets alpha helices and random coils are secondary structure elements. Specific secondary structure is defined by position of every amino acid .Figure 3.8 highlights the amino acids which make secondary structures formation. Results shows that secondary structure prediction contains more helixes than beta sheets which means that the protein is highly flexible. Figure 3.9 shows the graphical view of secondary structure of target protein.

3.4.3 Inhibitors Selection

Extensive literature search predicted important Inhibitors against FtsZ. FtsZ inhibitors were mostly consist of Derivatives of 3-methoxybenzamide,zantrins(polyphenols), berberine an alkaloid product from several plant species like blackberry and goldenscale ,cinammaldehyde from spices etc. which block substrate channel and inhibit the activity of enzyme(boberek et al.,2010). The current study is carried out to predict the inhibition mechanism of these inhibitors on protein functionality by hindering the activity of active site. These inhibitors were accessed from the BRaunschweig ENzyme Database (BRENDA) (Schomburg, 2002) .A total one hundered and two inhibitors were docked into FtsZ active site to analyse enzyme inhibition mechanism.

3.4.4 Docking Interaction Analysis

One hundred and two ligands molecules were prepared, docked into the active site of FtsZ using GOLD.,GOLDScore ranged from 35.2 to 86.0. While corresponding binding affinities were also calculated using AutoDock Vina which were ranging from -10.3 to -4.9 kcal/mol. The highest score of 86.0 was achieved for compound 78, 2"'-hydroxy-5"benzylisouvarinol-B a polyphenol extracted from X. afticana (Urgaonkar et al. 2005) with binding affinity of -7.6 kcal/mol which was chosen for further analysis to find its interactions(hydrophobic nad hydrogen bond interactions). Docking scores and respective binding affinities for top 10 compounds, arranged in descending order of Gold Score values are provided in Table 3.6. 2D and 3D depiction of protein and ligand interactions analysis was carried out employing UCSF Chimera, Discovery Studio, LIGPLOT and MOE which also revealed the conformational details and preferred orientation of the ligand binding. The binding of compound 78 was observed at N terminal domain of protein .The visual assessment revealed the following residues were involved in framing the inhibitor in active site of N terminal domain of FtsZ: Ile176, lys141, Gly139, Phe137, Ser177, Leu178, Phe182, Arg122, Glu138, Lys140, Asn136, Arg142 as shown in chimera visualization of ligand dock pose with FtsZ Fig 3.10. Non-bonded hydrophobic and electrostatic interactions were exhibited by, lys141, Gly139, Phe137, Ser177, Leu178, Phe182; while Glu138, Arg122, and Lys140 showed prominent hydrogen bonds with the ligand. This can be seen in 2D interaction images of LIGPLOT and MOE as shown in Fig. Moreover, LIGPLOT image i.e. Figure 3.11 illustrated the presence of hydrogen bonds between ligand and target. Ligand oxygen moiety formed two hydrogen bonds with active residue Glu138, and Arg142 with a distance of 2.94 Å and 2.83Å and Lys140 nitrogen atom developed hydrogen bonding with ligand at the distance of 2.70Å. Hydrogen bonding details are provided in Table 3.7.

3.5 Molecular Dynamics Simulation

The docking study provided a meaningful knowledge about the druggability potential of C.Sakazakii FtsZ within a static environment.. Molecular dynamic study was carried out in oreder to understand the conformational changes(dynamic behaviour) of FtsZ which is a potential drug target molecular dynamic simulation was was carried out followed by

trajectory analysis was done utilizing PTRAJ module of AMBER to assess various properties of docked system. Properties including the RMSD, RMSF, B-factor and radius of gyration were plotted as a function of time to to unravel the conformational changes of ligand bound FtsZ in hydrated system.

3.5.1 Root Mean Square Deviation

The deviation of the backbone Cα atoms was noticed for the entire production run of docked protein for a time period of 100 ns. Average RMSD value calculated for docked complex over 100 ns time period was 2.8476 Å, plot showed the steady increase in RMSD in the start and then gradual increase was noticed as shown in (Fig 3.14) . At the end of graph some stability was seen (Figure 3.14). During the simulation runs no major domain shifts were observed although there was some secondary structure rearrangements as shown in (Fig 3.16). The slight change in the structure is due to the elusive changes in backbone residues. (Figure 3.16)

3.5.2 Root Mean Square Fluctuations (RMSF)

RMSF is investigated to have insights into structural dynamics of residues. It provided a means to find and analyze the structurally flexible rigid regions of the drug target. Average root mean square fluctuations for the docked system for 100 ns was 1.39 Å which is well within the range and confirms the stability of system as shown in (Figure 3.15). The maximum value of RMSF in case of docked protein was 10.4853 Å .RMSF trajectory analysis of normal binding site complex showed higher fluctuations for the amino acid residues present at N terminal domain whereas active site residues revealed lesser fluctuations. From above discussed observation it can be interpreted that active site is more stable in a complex. In comparison with the entire protein, higher fluctuations were observed at regions: 1-20, 130-150, and 160-180, for docked protein signifying disorderness of C-alpha atoms in these regions.

3.5. 3 Beta Factor

Beta factor is measured in terms of RMSF. The value of beta factor is dependent on the level of localized atomic fluctuations which collectively contribute to the global vibrational movements of the protein and its thermal stability. B-factor value less than 30

Å2 for an atom is the indication of confidence in its atomic position, but values greater than 60 Å2 indicates possible atom's disorderness. The average value for the system of interest comes out to be 84.82.The slight increase in value of beta factor is probably due to residues 1 to 20 that exhibit high instability with a highest peak as shown in Figure 3.18 due to fluctuations in loop region.The loop is also replaced by helix at 40 ns, 80ns and 100ns as shown in Figure 3.16.These residues are not involved in ligand binding they do not affect the interaction of ligand and protein. The value for atom residues involved in ligand binding are less fluctuating signifying the uniformity of active site.

3.5.4 Radius of Gyration

Compactness and stability of a protein structure is determined by its radius of gyration (Rg). Reduction of radius of gyration values specified the stability of the system. It is measured as a root mean square distance between the protein's centre of gravity and its ends. Docked system has an average value of 19.56 Å shown in Figure 3.17. This implies that docked complex exhibit stable and compact system according to the value of radius of gyration.

4 DISCUSSION

Infectious diseases are the main and prominent danger for the public health around the world.Capacity of adjustment and reshaping of pathogenic genome in bacterial species restricts the activity of antimicrobial agents against these pathogens.The rise of anti-toxin resistance strains are not as a result of high rate of bacterial genome reshaping by mutation but also activated additionally by anti-biotics imposed selective pressure.Consequently, every one of these bottlenecks demand more up to date, specific and more intense therapeutic agents.The in-silico drug design is an extensive field in which the distinctive sides of fundamental research and practice are amalgamated and inspire each other. Current procedures, for example, QSAR/QSPR, structure-based design , combinatorial library design, cheminformatics, bioinformatics and the expanding number of chemical and biological databases are utilized as a part of the field. (Wadood et al., 2013).Computer aided drug designing over the past decades has been making a big splash in the area of drug design (Azam et al., 2014). Subtractive genomics is a thriving in silico approach that filters out pathogenic novel drug targets from complete genome.

The present work aims on the identification of potential druggable candidates of Cronobacter Sakzakii, also known as yellow pigmented Enterobacter sakazakii is a Gram-negative bacillus, rod shaped, peritrichous,non-spore forming, human opportunistic and food borne pathogen. The primary reservoir of Cronobacter Sakazakii has yet to be determined but it is postulated that plant material may be an important source (Adeel et al., 2013). Different antibiotics are employed to control this pathogen but recent studies investigated emergence of antibiotic-resistant strains due to which the situation demands a speedy course of drug discovery process (Zhou X et al., 2011). This went about as a boost for the present study where in silico drug design approach is utilized to facilitate the process and has given significant insight into the functional basis of the selected drug target.

This study is an aim to use in silico subtractive genomics approach to shortlist the possible drug targets in CMCC45402, SP2, and ES15 strains of Cronobacter Sakazakii. Different steps of therapeutic target screening technique i.e. *in silico* subtractive genomic

was used to filter pathogen genome.CMCC45402, ES15.SP291 strains of Cronobacter Sakazakii were targeted to combat multiple strains with the same drug.

The number of proteins in these three strain are 3807, 4066.4129 respectively. Subtractive genomic steps were applied on each strain independently,furthermore, toward the end accompanied 10, 8 and 6 potential drug target proteins which were cytoplasmic localized proteins. These common cytoplasmic targets were identified as a result of subtractive genomics approach which are involved in the following unique pathways:Vibrio cholera pathogenic cycle, Cell cycle - Caulobacter, Flagellar assembly, and Two-component system. After all evaluation of subtractive genomics study the target selected in current study was cell division protein (FtsZ), which is involved in "cell cycle-caulobacter" pathway. Multiple sequence alignment (MSA) of FtsZ sequences obtained from all three strains shows that they are highly identical, therefore targeting one protein sequence from one strain can work for all these three strains. FtsZ gene was used for comparative homology modeling to elucidate the binding interactions of ligands.

Comparative homology modeling was performed on FtsZ through MODELLER along with other automated web-based servers: I-TASSER, ModWeb, Swiss-Model, Phyre2, 3D-JIGSAW and the model was evaluated with different tools like ERRAT, PDBSum and Verify3D. Template FtsZ from pseudomonas aueroginosa GDP (PDB ID: 2VAW) was selected as a reference structure with the query coverage of 94% and 57% sequence identity. Comparision of model is shown in Table 3.4 which was helpful in selection of best model on basis of satisfactory values obtained from results. The best model came out through MODELLER .Furthermore molecular docking and molecular dynamic simulation was carried out on the best model selected. . Selected model generated via MODELLER9.10 was monomer .RMSD is calculated to elucidate the similarity between two proteins, which analyze the ensembles of backbone atoms for proper conformation (Maiorov and Crippen, 1994). RMSD value calculated for target and template was 0.488 Å, which means that the generated model was of good quality and contain similar main chain fold as that of template chain. Physicochemical properties of the modeled structure are given Table 3.5.

Molecular docking is utilized for identification of binding interactions of two molecules along with their binding affinities (Huang and Zou, 2010). To dock a ligand an active site

is required in target protein. In current study Glu138, charged amino acid, was used as an active residue.

Molecular Docking was carried out using AutoDock Vina and GOLD which provided a detailed information about the protein interaction, conformation, orientation, and type of inhibition. Both these docking programmes cosist of different scoring algorithms which resulted in the slight change in results from both programs. Total 102 ligands were docked in active site which is present in Nterminal domain of FtsZ protein. The highest GOLD score obtained is 86.0 of compound 78 with binding affinity of -9.6 kcal/mol. which was the best docking ligand that was taken for further detailed analysis. LIGPLOT, Discovery Studio, and UCSF Chimera was employed for detailed visualization of interactions of compound 78 with protein which include interactions like electrostatic interactions, hydrophobic interactions, and hydrogen bonding. Hydrogen bond interactions and binding energy calculations are important factors for investigation of representative docking studies. Hydrogen bonding plays a key role for inhibition mechanism in complex. If hydrogen bond interactions of ligand with active site residues is utilized to block a structure it becomes unstable and dispersal of ligand is seen. (Azam et al., 2012). Structure stability and compactness of compound 78 was analyzed through hydrogen bonds networks involving the residues like Glu138, Arg122, and Lys140 showed prominent hydrogen bonds with the ligand. Non-bonded hydrophobic and electrostatic interactions were exhibited by, lys141, Gly139, Phe137, Ser177, Leu178, Phe182. Ligand oxygen moiety formed two hydrogen bonds with active residue Glu138, and Arg142 with a distance of 2.94 Å and 2.83Å and Lys140 nitrogen atom developed hydrogen bonding with ligand at the distance of 2.70Å.

Molecular dynamics simulations are employed to identify the dynamic and time dependent behavior of FtsZ. It provides knowledge about the region of molecule involved in dynamic behavior of system of interest (Azam *et al.*, 2012) . Inhibition process, with respect to time was elucidated in solvated medium, for docked complex of FtsZ in the current study.RMSD, RMSF, B-factor, and radius of gyration for docked system are generated. RMSD value of docked system calculated was >2 Å and docked complex was moving towards stability at the end of 100 ns after binding of FtsZ with its

receptor.This implies that FtsZ is stable in its complex. RMSF averaged value calculated docked complex was >1 Å and few peaks were observed in the plot.highest peak of RMSF was observed for frist amino acid Met1 with RMSF value of 4.23 Å. Most fluctuating residues were Asn9, Gly20, Val12, Val130, Phe135, Asp158, Gly173.Most of the fluctuating residues are present in loop region of protein. Loops regions are considered as the most fluctuating residues containing region of protein (Ruvinsky and Vakser, 2010). Additionally, helices and sheets change their orientation. During the course of simulation helices and beta sheets also change their orientation slightly, during 100 ns. These movements are important for inhibition process.

In silico subtractive genomics approach applied in current study depicted the significant results at different steps of analysis. Binding modes of inhibitor confirms the inhibition process. Similarly, structure dynamics of docked and undocked protein provided the eloquent insight, Docking studies filtered the best binding inhibitors among the given library of compounds whereas simulation studies identified the proper orientation and positioning of the ligand molecule in the active site. It also explored the timely dependent dynamic behavior of protein which can be practiced to increase the potency of drug against the Cronobacter Sakazakii strains along with drug target specificity and selectivity which ultimately provide remedy for infections caused by studied pathogen.

.

Conclusion

The in silico based methodology includes a series of screening of proteins that can be utilized as potential drug targets and antibody candidates. These drug targets are important for the growth of the organisms and these proteins have unique functionality associated with a unique pathway. These proteins neither have a substitute protein nor an alternative pathway to perform the process.The current study carried out to design a drug-like molecule that can block cell division protein FtsZ. The microorganisms are rapidly getting resistant to the current drugs, so designing better and viable drugs ought to be made quicker. The current study was to explore the durgable genome of *Cronobacter Sakazakii* and identification of the putative therapeutic targets against emerging drug resistant pathogen. Subtractive genomic approach led to the identification of many pathogen specific targets, which were common to three strains under study. Parameters like durggability assessment and localization specifications led to the selection of FtsZ as drug target. Homology modeling was employed to model FtsZ which was further docked with library of 102 compounds. Furthermore, molecular docking was done which showed compound 78 is the highest GOLD scoring compound with the GOLD score of 86.0 which was the true consideration of the interaction strength. Polar amino acids was prominent in inhibitor moieties and clarifying the physicochemical properties of binding site. Moreover, MD simulation studies well explained the dynamic behavior of docked complex. Beside the side chain fluctuations and minor loop movement, stability of inhibitor and target protein complex was observed. All these function valuation fueled for rational structure based drug designing of novel inhibitors for *Cronobacter Sakazakii* with increase potency.

Thesis

ORIGINALITY REPORT

7%	2%	7%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	John J. Georrge. "Subtractive Genomics Approach to Identify Putative Drug Targets and Identification of Drug-like Molecules for Beta Subunit of DNA Polymerase III in Streptococcus Species", Applied Biochemistry and Biotechnology, 03/14/2012 Publication	1%
2	Brendan Healy. "*Cronobacter* (*Enterobacter sakazakii*): An Opportunistic Foodborne Pathogen", Foodborne Pathogens and Disease, 04/2010 Publication	1%
3	www.geometry.net Internet Source	<1%
4	Lee, J.-H., Y. Choi, H. Shin, J. Lee, and S. Ryu. "Complete Genome Sequence of Cronobacter sakazakii Temperate Bacteriophage phiES15", Journal of Virology, 2012. Publication	<1%
5	Azam, Syed Sikander, Sumra Wajid Abbasi,	

Amina Saleem Akhtar, and Mah-laka Mirza. "Comparative modeling and molecular docking studies of d-Alanine:d-alanine ligase: a target of antibacterial drugs", Medicinal Chemistry Research, 2014.
Publication

<1%

---

6 Levan Tsereteli. "Study of Metals in Proteins-Deriving parameters for Molecular Dynamic Simulations from Quantum Mechanical calculations of active centers in metalloproteins", Repositório Aberto da Universidade do Porto, 2013.
Publication

<1%

---

7 Azam, Syed Sikander, Asma Abro, Saad Raza, and Ayman Saroosh. "Structure and dynamics studies of sterol 24-C-methyltransferase with mechanism based inactivators for the disruption of ergosterol biosynthesis", Molecular Biology Reports, 2014.
Publication

<1%

---

8 Peters, G√ºnther. "Computer Simulations : A Tool for Investigating the Function of Complex Biological Macromolecules", Enzyme Functionality Design Engineering and Screening, 2003.
Publication

<1%

---

9 Hosen, Md. Ismail, Arif Mohammad Tanmoy,

<1%

Deena-Al Mahbuba, Umme Salma, Mohammad Nazim, Md. Tariqul Islam, and Sharif Akhteruzzaman. "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, 2014.
Publication

---

10 Shin, H., J.-H. Lee, Y. Kim, and S. Ryu. "Complete Genome Sequence of Cronobacter sakazakii Bacteriophage CR3", Journal of Virology, 2012. <1%
Publication

---

11 biobytes.bicmku.in <1%
Internet Source

---

12 David Gurwitz. "In silico subtractive genomics for target identification in human bacterial pathogens", Drug Development Research, 03/2011 <1%
Publication

---

13 Doss, C. George Priya Chakraborty, Chira. "Integrating in silico prediction methods, molecular docking, and molecular dynamics simulation to pr", BioMed Research International, Annual 2014 Issue <1%
Publication

14	Guharoy, Mainak, Beata Szabo, Sara Contreras Martos, Simone Kosol, and Peter Tompa. "Intrinsic structural disorder in cytoskeletal proteins : Intrinsic Structural Disorder", Cytoskeleton, 2013. Publication	<1%
15	www.plosone.org Internet Source	<1%
16	www.baqiqau.org Internet Source	<1%
17	www.wag.caltech.edu Internet Source	<1%
18	Danilo Roccatano. "Theoretical Study of Nanostructured Biopolymers Using Molecular Dynamics Simulations: A Practical Introduction", NanoScience and Technology, 2007 Publication	<1%
19	www.incttox.com.br Internet Source	<1%
20	www.eurekaselect.com Internet Source	<1%
21	Azam, S. Sikander, and A. Hammad Mirza. "Role of thumb index fold in Wnt-4 protein and its dynamics through a molecular dynamics	<1%

simulation study", Journal of Molecular Liquids, 2014.
Publication

---

22 Kaur, Navkiran, Mansimran Khokhar, Vaibhav Jain, P. V. Bharatam, Rajat Sandhir, and Rupinder Tewari. "Identification of Druggable Targets for Acinetobacter baumannii Via Subtractive Genomics and Plausible Inhibitors for MurA and MurB", Applied Biochemistry and Biotechnology, 2013. <1%
Publication

---

23 metagenomaiberico.org <1%
Internet Source

---

24 Zhang, Xueyao, Tao Li, Jianqin Zhang, Daqi Li, Yaping Guo, Guohua Qin, Kun Yan Zhu, Enbo Ma, and Jianzhen Zhang. "STRUCTURAL AND CATALYTIC ROLE OF TWO CONSERVED TYROSINES IN DELTA-CLASS GLUTATHIONE S-TRANSFERASE FROM Locusta migratoria : Two Tyrosines of LmGSTD1 Are Important for Protein Structure and Enzyme Function", Archives of Insect Biochemistry and Physiology, 2012. <1%
Publication

---

25 Huan-Chang Tseng. "Nanocontraction flows of short-chain polyethylene via molecular dynamics simulations", Molecular Simulation, <1%

2009
Publication

26	Jadhav, Ankush, Vijayalakshmi Ezhilarasan, Om Prakash Sharma, and Archana Pan. "Clostridium-DTDB: A comprehensive database for potential drug targets of Clostridium difficile", Computers in Biology and Medicine, 2013. Publication	<1%
27	stat.duke.edu Internet Source	<1%
28	archiv.ub.uni-marburg.de Internet Source	<1%
29	Lopez-Camacho, E., M. J. Garcia Godoy, A. J. Nebro, and J. F. Aldana-Montes. "jMetalCpp: optimizing molecular docking problems with a C++ metaheuristic framework", Bioinformatics, 2014. Publication	<1%
30	Y. Sun. "The development of an O serotyping scheme for Cronobacter sakazakii", Applied and Environmental Microbiology, 02/04/2011 Publication	<1%
31	Sakkiah, S.. "Pharmacophore modeling and molecular dynamics simulation to identify the critical chemical features against human sirtuin 2 inhibitors", Journal of Molecular Structure,	<1%

20120314
Publication

32	Alvarez-Ordonez, A., M. Begley, and C. Hill. "Polymorphisms in rpoS and Stress Tolerance Heterogeneity in Natural Isolates of Cronobacter sakazakii", Applied and Environmental Microbiology, 2012. Publication	<1%
33	Azam, S. Sikander, Reaz Uddin, and Abdul Wadood. "Structure and dynamics of alpha-glucosidase through molecular dynamics simulation studies", Journal of Molecular Liquids, 2012. Publication	<1%
34	duepublico.uni-duisburg-essen.de Internet Source	<1%
35	Kotapati, Kasi Viswanath, Bhagath Kumar Palaka, Anithamma Kandukuri, Ramachandra Reddy Pamuru, Veeranjaneya Reddy Lebaka, and Dinakara Rao Ampasala. "Molecular cloning, characterization and three-dimensional structure prediction of Lipoxygenase from Finger millet [Eleusine coracana (L.) Gaertn.] germinating seedlings", Journal of Plant Biochemistry and Biotechnology, 2015. Publication	<1%
36	Azam, Syed Sikander, Asma Abro, and Saad	

	Raza. "Binding pattern analysis and structural insight into the inhibition mechanism of Sterol 24-C methyltransferase by docking and molecular dynamics approach", Journal of Biomolecular Structure and Dynamics, 2015. Publication	<1%
37	Monhemi, Hassan, Mohammad Reza Housaindokht, and Ali Nakhaei Pour. "Effects of Natural Osmolytes on the Protein Structure in Supercritical CO2: Molecular Level Evidences", The Journal of Physical Chemistry B, 2015. Publication	<1%
38	K. W. Osteryoung. "The Division of Endosymbiotic Organelles", Science, 12/05/2003 Publication	<1%
39	U. Amineni. "In silico identification of common putative drug targets in Leptospira interrogans", Journal of Chemical Biology, 05/14/2010 Publication	<1%
40	Kim, K.. "Prevalence and genetic diversity of Enterobacter sakazakii in ingredients of infant foods", International Journal of Food Microbiology, 20080229 Publication	<1%

41 Logue, Catherine M., and Lisa K. Nolan. "Emerging Bacterial Food-Borne Pathogens and Methods of Detection", Food Biochemistry and Food Processing Simpson/Food Biochemistry and Food Processing, 2012. <1%

Publication

---

42 Karunasagar. "In silico identification and characterization of novel drug targets and outer membrane proteins in the fish pathogen Edwardsiella tarda", Open Access Bioinformatics, 01/2011 <1%

Publication

EXCLUDE QUOTES	ON	EXCLUDE MATCHES	OFF
EXCLUDE BIBLIOGRAPHY	ON		