COMPUTATIONAL STUDIES OF VIRULENT PROTEINS AND ANTIBIOTIC RESISTANCE IN DIARRHEAL PATHOGENS

A THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF DOCTOR OF PHILOSOPHY

IN

BIOINFORMATICS

BY

TAMANNA

Enrollment No. 116502

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY

WAKNAGHAT

October, 2017

Dedicated to

My Loving Parents

And

Lovely Husband

DECLARATION

I certify that:

a. The work contained in this thesis is original and has been done by me under the guidance of my supervisor.

b. The work has not been submitted to any other organisation for any degree or diploma.

c. Wherever, I have used materials (data, analysis, figures or text), I have given due credit by citing them in the text of the thesis.

Tamanna **Date:**
(Enrollment No. 116502)
Department of Biotechnology and Bioinformatics
Jaypee University of Information Technology, Waknaghat, India

CERTIFICATE

This is to certify that the thesis entitled, "**Computational Studies of Virulent Proteins and Antibiotic Resistance in Diarrheal Pathogens**" which is being submitted by **Tamanna (Enrollment No. 116502)** in fulfillment for the award of degree of **Doctor of Philosophy** in **Bioinformatics** at **Jaypee University of Information Technology**, **India** is the record of candidate's own work carried out by her under my supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of this or any other degree or diploma.

Dr. Jayashree Ramana
Assistant Professor (Senior Grade),
Email: Jayashree.ramana13@gmail.com

Date:

ACKNOWLEDGEMENT

Though only my name appears on the cover of this dissertation, a great many people have contributed to its production. I owe my gratitude to all those people who have made this dissertation possible and because of whom my graduate experience has been one that I will cherish forever.

At the very outset, I would like to thank ***Prof. (Dr.) Vinod Kumar****Vice Chancellor, JUIT;* ***Prof. (Dr.) Samir Dev Gupta****, Director and Academic Head andDean (Academic & Research),JUIT,* ***Maj Gen Rakesh Bassi (Retd.)****Registrar and Dean of StudentsJUIT,* ***Dr. Sudhir KumarSyal****Acting Head- Department of Biotechnology and Bioinformatics JUIT; for providing me an opportunity to pursue a Doctorate Degree, teaching assistantship and advanced lab infrastructure to accomplish my research work at Jaypee University of Information Technology, Waknaghat, India.*

I feel privileged to express my deep sense of reverence and gratitude to my mentor ***Dr. Jayashree Ramana, Assistant Professor (Bioinformatics)*** *for her constant encouragement & guidance in pursuing my research. Her positive attitude and zest for quality research always encouraged me and bought the best out of me. Her enthusiasm, encouragement and faith in me throughout have been extremely helpful. I am deeply indebted for her support, guidance and constructive criticism. She was always available for my questions and gave generously of her time and vast knowledge. I am also thankful to her for reading my reports; commenting on my views and helping me understand and enrich my ideas. I thank her from bottom of my heart.*

I gratefully acknowledge the help rendered by ***Prof. (Dr.) R. S. Chauhan*** *for his encouragement, timely help and cooperation throughout my research work.He sets high standards for his students and he encourages and guides them to meet those standards. It gives me immense pleasure to express my gratitude for his ever smiling disposition coming to my rescue in solving my problems and suggestions which helped me in maintaining my confidence.*

Also, I would like to thank ***Dr. Hemant Sood****for constantly encouraging and supporting me in my research endeavours. She played a very important role by being there so as to make me sail through all the twists and turns of this journey. She has been a constant source of love, concern, support and strength all these years and always there to listen and give advice. I am really very grateful to her for every time she helped me and stood there for me.*

*I am thankful to support staff*especially ***Mrs. Somlata Sharma*** *who maintained all the machines in my lab so efficiently that I never had to worry about viruses, losing files, creating backups or installing software**.*** *I would also like to thank* ***Mrs. Mamta Mishra*** *for her helping attitude and moral support.*

*Most importantly, none of this would have been possible without the love and patience of my family. It was the blessing and love of my grandparents (**Lt. Sh. Beli Ram** and **Smt. Pushpa Devi)** and my parents **(Mr. Ajay Kumar and Mrs. Kiran Lata)**, which have constantly been with me so as to make my life and career successful. Apart from my parents my brother **Akshat** and my sister **Kanika**have been my angels.My heart felt regard goes to my father in law**(Mr. Ravinder Kumar Bansal)**, mother in law**(Mrs. Rita Bansal)** and sister in law **Mrs. Jyoti**for their love and moral support. It was only because of their support, constant encouragement, prayers and blessings that I could overcome all frustrations and failures.*

I am deeply thankful to my husband ***Mr. Gaurav Bansal*** *for his continued and unfailing love, support and understanding during my pursuit of Ph.D degree He was always around at times I thought that it was impossible to continue. I greatly value his contribution and deeply appreciate his belief in me.*

Also, I would like to thank all my friends especially ***Bharti Ma'am***, ***Seneha Ma'am***, ***Manika***, ***Shweta***, ***Garima***, ***Tamanna***, ***Asha***, ***Kusum***, ***Tarun*** and ***Archit Sir***, *for their support and care that helped me overcome setbacks and stay focused on my graduate study. I greatly value their friendship and I deeply appreciate their belief in me. It is my pleasure to express my gratitude to all research scholars of the Biotechnology & Bioinformatics Department who helped me all possible aspects.*

I would like to express my heartfelt gratitude to all those who have contributed directly or indirectly towards obtaining my doctorate degree and apologize if have missed out anyone.

Last, but not the least, I thanks the one above all of us, omnipresent God, for answering my prayers, for giving me the strength to plod on during each and every phase of my life.

Tamanna

TABLE OF CONTENTS

CHAPTER 2

CHAPTER 3

***Developing machine learning tool for the prediction of Multidrug And Toxin Extrusion (MATE) proteins based on Artificial Neural Network (ANN) and Support Vector Machine (SVM)*39-59**

CHAPTER 4

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

WHO	World Health Organization
GEMS	Global Enteric Multicenter Study
ETEC	Enterotoxigenic E.coli
EPEC	Enteropathogenic E.coli
EHEC	Enterohemorrhagic E.coli
EAEC	Enteroaggregative E.coli
EIEC	Enteroinvasive E.coli
DAEC	Diffusely Adherent *E.coli*
Sd1	Shigella dysenteriae type 1
MATE	Multidrug And Toxin Extrusion
MFS	Major Facilitator Superfamily
SMR	Small Multidrug Resistance
RND	Resistance Nodulation and Cell Division
ABC	ATP Binding Cassette
T3SS	Type III Secretion System
BLAST	Basic Local Alignment Search Tool
PDB	Protein Data Bank
ANN	Artificial Neural Network
SVM	Support Vector Machine
PSSM	Position Specific Scoring Matrix
MDR	Multiple Drug Resistance

NR	Non Redundant
SNNS	Stuttgart Neural Network Simulator
SN	Sensitivity
SP	Specificity
MCC	Matthew correlation coefficient
AAC	Amino Acid Composition
DPC	Dipeptide Composition
CC	Charge Composition
HC	Hydrophobicity Composition
MC	Multiplets Composition
RBF	Radial Basis Function
ROC	Receiver Operating Characteristic
AUC	Area Under Curve
QRDR	Quinolone Resistance-Determining Region
Ser	Serine
Leu	Leucine
Asp	Aspartic acid
Gly	Glycine
Asn	Asparagine
Ile	Isoleucine
Gln	Glutamine
Arg	Arginine
Val	Valine

Ala	Alanine
Thr	Threonine
Glu	Glutamic acid

ABSTRACT

ABSTRACT

Diarrhea is a condition that involves the frequent passing of loose or watery stools. Diarrhea may be caused by Inflammatory Bowel Syndrome (IBS), an allergy, or an infection due to a virus, bacteria, or parasite. Diarrhea is also associated with other infections such as malaria and measles. Chemical irritation of the gut or non-infectious bowel disease can also result in diarrhea. According to the World Health Organization (WHO) each year approximately 1.7 billion deaths are attributable to diarrhea. In highest burdened regions like Southeast Asia and Africa, diarrhea is responsible for as much as 8.5% and 7.7% of all deaths respectively. In children under the age of 5 years, 80% of the deaths occur due to diarrhea only. Children are more susceptible to the complications of diarrhea because a smaller amount of fluid loss leads to dehydration, compared to adults.Although usually not harmful, diarrhea can become dangerous or signal a more serious problem. Major contributors to the diseases are bacteria, viruses and parasites. Enormous data about these pathogenic organisms is available from different information portals which need to be compiled for providing better treatment strategy. Although, currently available treatment methods which include Oral Rehydration Therapy (ORT), antibiotics and vaccines had reduced the diseases burden to some extent, but due to the increasing problem of drug resistance, control of infectious disease is becoming more difficult. Antibiotic resistance in case of Traveler's Diarrhea (TD) is an important public health concern. Large numbers of antibiotics are being employed to cure traveler's diarrhea, but widespread use of these antibiotics has developed resistant strains of pathogenic bacteria.Hence, it is crucial to understand the resistance mechanism and devising novel solution to combat this problem.

In this study, we first developed a database named dbDiarrhea, where Pathogen proteins, host proteins, Type Three Secretion System (T3SS) Effectors and T3SS Inhibitors information is available in a distinctive manner and is available for academic andresearch use at "http://www.juit.ac.in/attachments/dbdiarrhea/diarrhea_home.html". It also serves as a repository of the research articles of trials related to subunit and whole organism vaccines, high-throughput screening of Type III secretion system inhibitors and diagnostic assays, for various diarrheal pathogens. The user friendly interface allows querying proteins and research articles for different organism either by keywords or accession number. It alsoprovides sequence similarity search with the BLAST tool.

Multidrug And Toxin Extrusion (MATE) plays very important functions in the secretion of cationic drugs across the cell membrane and is utilized by several bacteria to evade the toxic effect of

antibiotics. Therefore, we proposedmachine learning method for prediction of MATE proteins. Here Artificial Neural Network and Support Vector Machine based approaches are applied to predict MATE proteins. The data set employed for training consists of 189 non-redundant protein sequences that comprises of 63 sequences from MATE family and 126 other protein sequences obtained NCBI protein databank. The fully-connected network was derived using amino acid composition, which yielded an overall accuracy of 84.45%, in three fold cross validation. But it failed to perform remarkably well on independent dataset. So, we generated SVM based Position Specific Scoring Matrix (PSSM) model and achieved an overall accuracy 92.06%infive-fold cross validation. The prediction algorithm presented here is implemented as a freely available web server MATEpred and is openly accessible at http://www.bioinformatics.org/matepred_hos, which will assist in rapid identification of MATE proteins.

Finally, we performed docking studies of DNA gyraseA (GyrA) and Topoisomerase IV (parC) of S*higella flexneri* and its mutants with two different fluoroquinolones, ciprofloxacin and norfloxacin to understand its resistance mechanism at structural level. *Shigella flexneri* DNA GyraseA with mutations at serine 83 to leucine and aspartic acid 87 to glutamate and serine 80 to isoleucine of parC have shown resistance to these fluoroquinolones.The amino acid residue Asp 87 in GyrA and Ser 80 in parC makes direct hydrogen bonds with both ciprofloxacin and norfloxacin (wild type), so the mutations at this point leads to drastic changes in molecular interactions. From this analysis, it was observed that there is a weaker interaction of ciprofloxacin/norfloxacin with all the mutants as compared to the wild type. The work presented here gives a good explanation for quinolone resistance in S*higella flexneri* and can be used further to design new drug targets against the resistant strains.

CHAPTER –1

INTRODUCTION

1.1. INTRODUCTION

Diarrhea is an increase in the frequency of bowel movements or a decrease in the form of stool. Diarrhea is a neglected tropical disease despite being a global scourge and international health challenge. Diarrhea exacts large tolls of morbidity and mortality among all age groups and is particularly endemic in developing countries. It causes about 1.7 billion deaths worldwide [1]. The global burden of incidence and severity for diarrhea is highest in Southeast Asia and Africa (Figure 1.1).

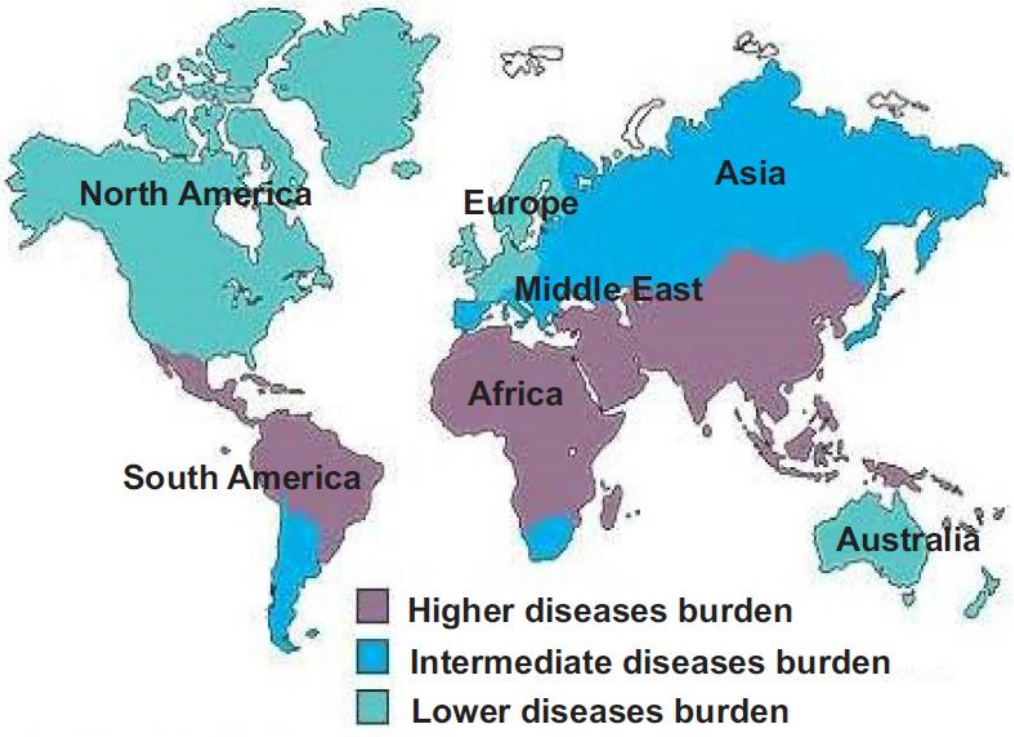


Figure 1.1 Worldwide Distribution of Diarrheal diseases. This compilation clearly notes Africa as highly burdened region with diarrheal diseases followed by Southeast Asia.

According to the report released in May 2017 by WHO, Diarrheal diseases account for 1 in 9 child deaths worldwide, making diarrhea the second leading cause of death among children under the age of 5. Diarrhea is also associated with other infections such as malaria and measles. For children with HIV, diarrhea is even more deadly; the death rate for these children is 11 times

higher than the rate for children without HIV. Diarrhea kills more children than malaria, measles, and AIDS combined [2](figure 1.2).

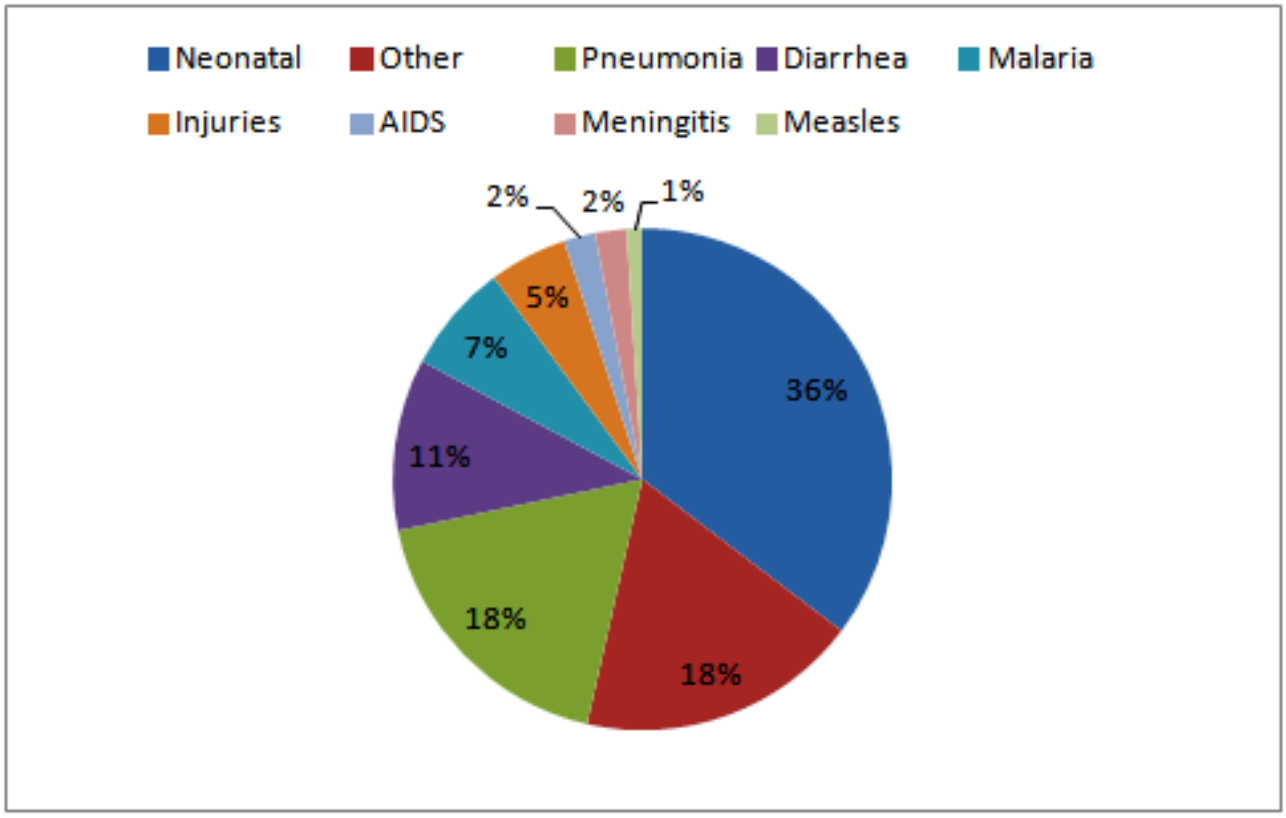


Figure 1.2 Proportional distribution of cause-specific deaths among children under five years of age [2].

In the context of India, according to Global Enteric Multicenter Study (GEMS), published in May 2013, about 18% of deaths are due to diarrheal diseases. The WHO estimates that diarrhea induced by the rotavirus kills between 90,000 and 153,000 children in India every year [1]. Around 1.5 million children below the age of five in India die annually and out of this 334,000 are due to diarrhea-related diseases. The rotavirus alone was responsible for moderate-to-severe diarrhea [3].

1.2. PREDISPOSING FACTORS FOR DIARRHEA

Most diarrheal germs are spread from the stool of one person to the mouth of another. These germs are usually spread through contaminated water, food, or objects.Water, food, and objects become contaminated with stool in many ways:

- People and animals defecate in or near water sources that people drink.
- Farmers use contaminated water to irrigate their crops.
- Crops irrigated with contaminated water are used to prepare meals
- People use contaminated water for drinking and food preparation
- Caregivers prepare foods with unwashed hands, contaminating the food.

1.3. Causative Agents OF Diarrhea

Diarrhea is caused by a league of heterogeneous pathogenic groups comprising of various bacteria, e.g. *E.coli, Vibrio, Shigella, Campylobacter*, *Yersinia*, etc. viruses e.g. *Rotavirus, Norovirus,* and protozoan parasites e.g. *Giardia, Entameoba.* The diarrheagenic mechanisms and the associated symptoms are as diverse as the etiological agents. 14 major bacterial, viral and parasitic species are known to be involved in the pathogenesis of diarrhea. Bacteria include *Escherichia coli, Salmonella enterica, Shigella species,Campylobacter jejuni, Vibrio cholerae, Vibrio parahaemolyticus, Yersinia enterocolitica, Clostridium difficile and Aeromonas hydrophila.* Viruses include *Rotavirus, Adenovirus* and *Norovirus.* The parasites include *Cryptosporidium parvum, Entamoeba histolytica and Giardia lamblia* (Figure 1.3).

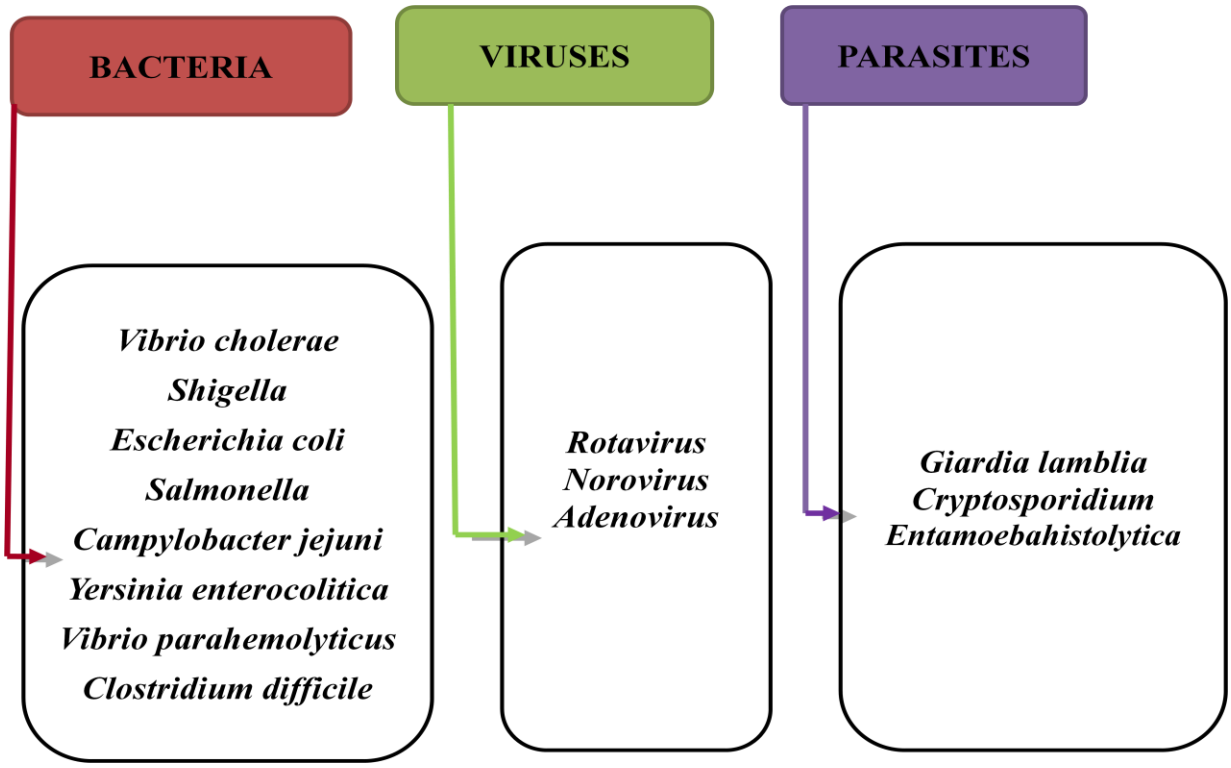


Figure1.3Major bacterial, viral and parasitic species involved in the pathogenesis of diarrhea.

1.3.1. Bacteria

1.3.1.1. *Escherichia coli*

Escherichia coli (E.coli) usually remains harmlessly confined to the intestinal lumen; however, in the debilitated or immunosuppressed host, or when gastrointestinal barriers are violated, even normal "nonpathogenic"strains of *E.coli* can cause infection [4]. The diarrheagenic *Escherichia coli* pathotypes (DEPs) include Enterotoxigenic (ETEC), Enteropathogenic (EPEC), Enterohemorrhagic (EHEC), Enteroinvasive (EIEC), Enteroaggregative (EAEC), and Diffusely Adherent *E.coli (DAEC),* all causing infections to the human intestinal tract. *Enterotoxigenic E.coli* **(**ETEC**)** elaborates at least one member of two defined groups of enterotoxins: ST (Heat-stable toxins) and LT(Heat-labile toxins) [4]. *Enteropathogenic E.coli* (EPEC) is known for its attaching-and-effacing (A/E) mechanism. *Enterohemorrhagic E.coli* **(**EHEC) also known as Shiga toxin producing *E.coli* causes bloody diarrhea. *Enteroaggregative E.coli (EAEC)* characteristically enhances mucus secretion from the mucosa, with trapping of the bacteria in a bacterium-mucus biofilm [4]. *Enteroinvasive E.coli (EIEC*)strains are biochemically, genetically, and pathogenetically related closely to *Shigella* spp. [4].

1.3.1.2. *Salmonella enterica*

Followed by *E.coli***,** *Salmonella enterica* is the second most widely studied organism. *Salmonella enterica* are enteropathogenic bacteria capable of causing a wide range of illnesses ranging from mild food poisoning to life-threatening systemic infections [5]. All serotypes are pathogenic for humans. Diarrhea with or without fever develops and lasts for about 3 weeks or more [6].

1.3.1.3. *Shigella species*

Shigella species are the causative agents of bacillary dysentery or shigellosis, which remains a threat to public health worldwide, particularly in developing countries [7]**.** It is caused by four major species: *S. sonnei, S. flexneri, S. dysenteriae type 1 (Sd1), S. boydii.* Among all these, S*higella flexneri* is more prevalent followed by *Shigella dysenteriae type 1*.

1.3.1.4. *Campylobacter jejuni*

It is prevalent in adults and is one of the most frequently isolated bacteria from the feces of infants and children in developing countries. Infection is associated with watery diarrhea and on occasion dysentery (acute bloody diarrhea) [6].

1.3.1.5. *Vibrio cholera*

Many species of *Vibrio* cause diarrhea in developing countries. *V. cholerae* serogroups O1 and O139 cause rapid and severe depletion of volume. Stools are watery, colorless, and flecked with mucus. Vomiting is common; fever is rare [6].

1.3.1.6. *Vibrioparahaemolyticus*

Vibrio parahaemolyticus is a human pathogen that naturally inhabits marine and estuarine environments. Infection with *V. parahaemolyticus* is often associated with the consumption of raw or undercooked seafood, causing gastroenteritis with watery diarrhea. The presence of two type III secretion system (T3SS) proteins, thermostable direct hemolysin (TDH) and TDH-related hemolysin (TRH), has been closely associated with the severity of diarrheal illness [8]. Recent studies have also uncovered the indispensability of T3SS to the pathogenesis of diarrhea in various pathogens like *Vibrio cholerae AM-19226* [9], *Shigella* [10], *Salmonella enterica*[11]. T3SS is a common stratagem deployed by several enteric bacteria to translocate toxins and

effector proteins across the host cell via an injectisome and subvert normal cell functions to aid infections [12].

1.3.1.7. *Yersinia enterocolitica*

Yersinia entercolitica is a human foodborne pathogen that interacts extensively with tissues of the gut and the host's immune system to cause disease. As part of their pathogenic strategies, *Yersinia* have evolved numerous ways to invade host tissues, gain essential nutrients, and evade host immunity [13].

1.3.1.8. *Clostridium difficile*

Clostridium difficile is a Gram-positive, spore-forming rod that is responsible for 15 to 20 percent of antibiotic-related cases of diarrhea. *C. difficile* could be isolated from the gastrointestinal tracts of most neonates; thus, it was believed to be a commensal organism [10].

1.3.1.9.*Aeromonas hydrophila.*

Aeromonas hydrophila is the causative agent of a number of human infections such as septicemia and gastroenteritis. Isolation of *A. hydrophila* from water and food sources, and the increasing resistance of this organism to antibiotics and occurrence in chlorinated water supply[14] , presents a significant threat to public health [15].

1.3.2. Viruses

1.3.2.1. *Rotavirus*

Rotavirus is the leading cause of diarrhea hospitalization among children worldwide [16]. Nearly all children in both industrialized and developing countries get infected with rotavirus by the time they are 3–5 years of age [6].

1.3.2.2. *Norovirus*

Norovirus is the group of viruses responsible for causing gastroenteritis in humans. Norovirus infection, a major cause of acute epidemic diarrhea, has been described as a cause of chronic diarrhea in patients who are immunosuppressed, including transplant recipients and the very young [17].

1.3.2.3.*Adenovirus*

Adenoviruses are important etiologic agents of gastroenteritis. *Adenoviruses*, particularly enteric adenoviruses (EAds) type 40 (Ad40) and type 41(Ad41), can cause acute and severe diarrhea in young children worldwide[18].

1.3.3. Parasites

1.3.3.1.*Cryptosporidium parvum*

Cryptosporidium parvum is a protozoan parasite that causes cryptosporidiosis, a disease affecting the mammalian intestinal tract and mainly characterized by a diarrheal illness. Cryptosporidiosis can be found world-wide, and in developing countries 8–19% of diarrheal diseases are attributed to *Cryptosporidium*[19].

1.3.3.2.*Giardia lamblia*

Giardiasis is a parasitic disease caused by *Giardia* species, a flagellated protozoan parasite that occupies the small intestine of numerous hosts including humans [19]. Ithas a low prevalence (approximately 2–5%) among children in developed countries, but as high as 20–30% in developing regions [6].

1.3.3.3.*Entamoeba histolytica*

Amebiasis is caused by *Entamoeba histolytica*, a protozoan parasite that occurs worldwide. It occurs usually in the large intestine and causes internal inflammation. *Entamoeba histolytica* had high prevalence and unusual presentation by affecting high proportion of infants under 1 year [20]. *E. histolytica* can be a re-emerging serious infection when it finds favorable environmental conditions and host factors.

1.4. TYPES OF DIARRHEA

Episodes of diarrhea can be classified into following categories based on duration and mechanism:

Duration Based:

Table 1.1Classification of diarrhea based on the duration of occurrence**.**

Type	Duration
Acute	5-10 days in duration
Persistent	more than 14 days in duration
Chronic	more than 30 days in duration

Mechanism Based:

Table 1.2 Classification of diarrhea based on mechanism of occurrence**.**

Type	Duration
Osmotic	when too much water is drawn into the bowels
Secretory	increase in the active secretion or reduced absorption of ions and salts
Exudative	presence of blood and pus in the stool

1.5. SYMPTOMS OF DIARRHEA

- Nausea
- Abdominal pain
- Cramping
- Bloating
- Dehydration
- Fever
- Bloody stools
- Frequent Urge to evacuate the bowels

1.6. TREATMENT OPTIONS AND PREVENTION

The addition of zinc to oral rehydration solution has been proven effective in children with acute diarrhea in developing countries [21][22]. Several randomized controlled trials and meta-analyses suggested that probiotics are effective in primary and secondary prevention of gastroenteritis and its treatment. Their efficacy is less convincing in adults, but promising in

antibiotic-associated diarrhea [23]. Though the current treatment methods available to cure diarrhea have reduced the severity of the diseases but due to emergence of drug resistant bacteria, development of new drugs and vaccine antigens is required.

Table 1.3 List of currently available drugs and vaccines against diarrheal pathogens.

Organism	**Drugs**	**Vaccines**	
		Live attenuated	**Subunit**
Amebiasis	Metronidazol	-	-
Giardiasis	Metronidazole and Ornidazole	α1-Giardin	-
Campylobacter	Azithromycin	-	-
Cholera	Doxycycline and Tetracycline	Mutacol	Dukoral
Shigellosis	Ciprofloxacin and Norfloxacin	SC602, WRSs2 and WRSs3	-
Escherichia coli	Ciprofloxacin and Azithromycin	ACE527	-
Rotavirus	-	Rotarix, Rotateq and Rotavac	-

1.7. ANTIBIOTIC RESISTANCE – A SERIOUS THREAT

Infections have been the major cause of disease throughout the history of human population. With the antibiotics introduction, it was thought that this problem should disappear [24]. Two major ways that modern medicine saves lives are through antibiotic treatment of severe infections and under the antibiotic protection performance of medical and surgical procedures [25]. Discovery of antibiotics took place in the middle of the nineteenth century and brought down the threat of infectious diseases [26]. However, with the emergence of antibiotic resistant pathogens, currently available antibiotics are becoming ineffective [27]. Antibiotic resistance is the ability of a microorganism to survive and multiply in the presence of an antimicrobial agent that would normally inhibit or kill this particular kind of organism. Soon after the discovery of penicillin in 1940, a number of treatment failures and occurrence of some bacteria such as *Staphylococci* which were no longer sensitive to penicillin started being noticed [26]. In organisms that encountered the first commercially produced antibiotics, resistance to single antibiotics became prominent [28]. In the past decade, various

key organizations, such as the Infectious Diseases Society of America, the Centers for Disease Control and Prevention, the World Health Organization (WHO), and the World Economic Forum, have made antibiotic resistance the focus of highly visible reports, conferences, and actions [25]. Increasing prevalence of resistance has been reported in many pathogens over the years in different regions of the world including developing countries [26]. This has been attributed to changing microbial characteristics, selective pressures of antibiotic use, and societal and technological changes that enhance the development and transmission of drug-resistant organisms. Although antibiotic resistance is often enhanced as a consequence of infectious agents' adaptation to exposure to antibiotic used in humans or agriculture and the widespread use of disinfectants at the farm and the household levels[28, 29].

Bacteria have evolved several mechanisms of rendering antibiotics inactive such as the enzymatic hydrolysis of antibiotics, group transfer and the redox process [30]. Microorganisms have increasingly become resistant to ensure their survival against the antibiotics to which they are bombarded. They achieved this through different means but primarily based on the chemical structure of the antibiotic and the mechanisms through which they act. The resistance mechanisms therefore depend on which specific pathways are inhibited by the drugs and the alternative ways available for those pathways that the organisms can modify in order to survive [26]. Resistance can be described in two ways:

a) Intrinsic or natural whereby microorganisms naturally do not posses target sites for the drugs and therefore the drug does not affect them or they naturally have low permeability to those agents because of the differences in the chemical nature of the drug and the microbial membrane structures especially for those that require entry into the microbial cell in order to effect their action [26] .

b) Acquired resistance whereby a naturally susceptible microorganism acquires ways of not being affected by the drug [26, 31]. Acquired resistance mechanisms can occur through various ways as described below :

(1) By prevention of the antimicrobial from reaching its target by reducing its ability to penetrate into the cell (Figure 1.4). For example; *Pseudomonas aeruginosa* against imipenem (a beta-lactam antibiotic) [32].

Figure 1.4 Prevention of antibiotic from reaching its target site [33].

(2) By expulsion of the antimicrobial agents from the cell via general or specific efflux pumps (Figure 1.5). For example; *Escherichia coli* against tetracyclines[34]. These efflux pumps are variants of membrane pumps possessed by all bacteria, both pathogenic and non pathogenic, to move molecules in and out of the cell [29].

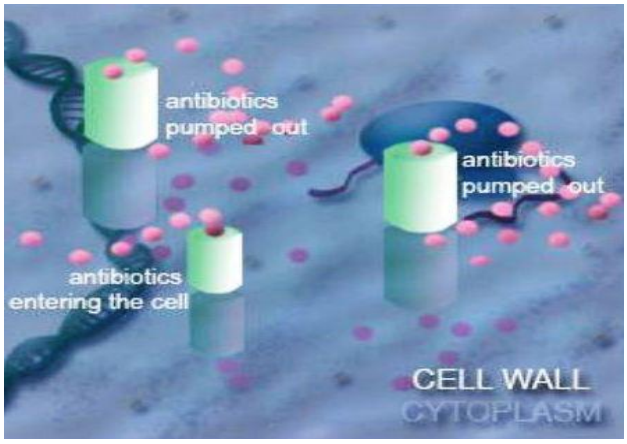


Figure 1.5 Expulsion of the antimicrobial agents from the cell via efflux pumps [33]

(3) By inactivation of antimicrobial agents via modification or degradation (Figure 1.6). For example; Enterobacteriaceae against chloramphenicol (acetylation) [35].

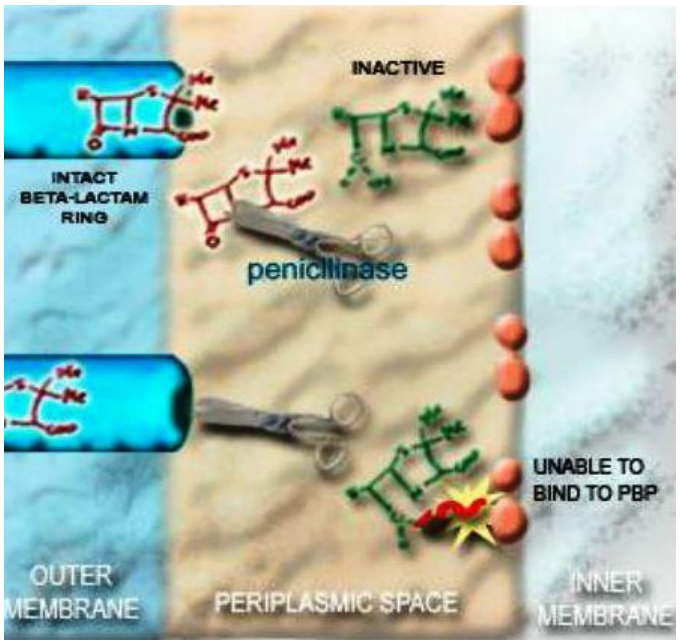


Figure 1.6 Inactivation of antimicrobial agents via modification or degradation [33]

(4) By modification of the antimicrobial target within the bacteria (Figure 1.7). For example; Mutations in DNA gyrase of *Shigella spp*. resulting in resistance to quinolones [36].

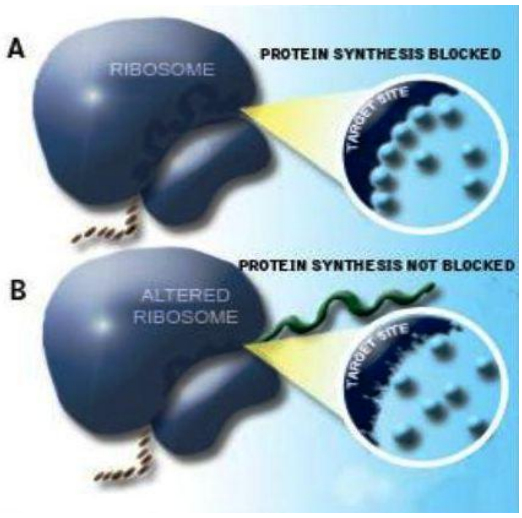


Figure 1.7 Modification of the antimicrobial target within the bacteria [33]

1.8. MULTIDRUG AND TOXIN EXTRUSION (MATE) PROTEINS

One of the mechanisms that bacteria utilize to evade the toxic effects of antibiotics is the active extrusion of structurally unrelated drugs from the cell [27]. Multidrug efflux is an important

mechanism of biocide and antimicrobial agent resistance in bacteria. Efflux is the pumping of a solute out of a cell. Efflux pump genes and proteins are present in both antibiotic-susceptible and antibiotic-resistant bacteria [37]. Poole in 2005, reported that efflux was first used as a mechanism of resistance to tetracycline in *Escherichia coli*[38]. These pumps have been divided into various groups (Figure 1.8), which include the Major Facilitator Superfamily(MFS), the Small Multidrug Resistance (SMR) family, the Resistance Nodulation and Cell Division (RND) family, the ATP Binding Cassette (ABC) family, and the Multidrug And Toxin Extrusion (MATE) family [39].

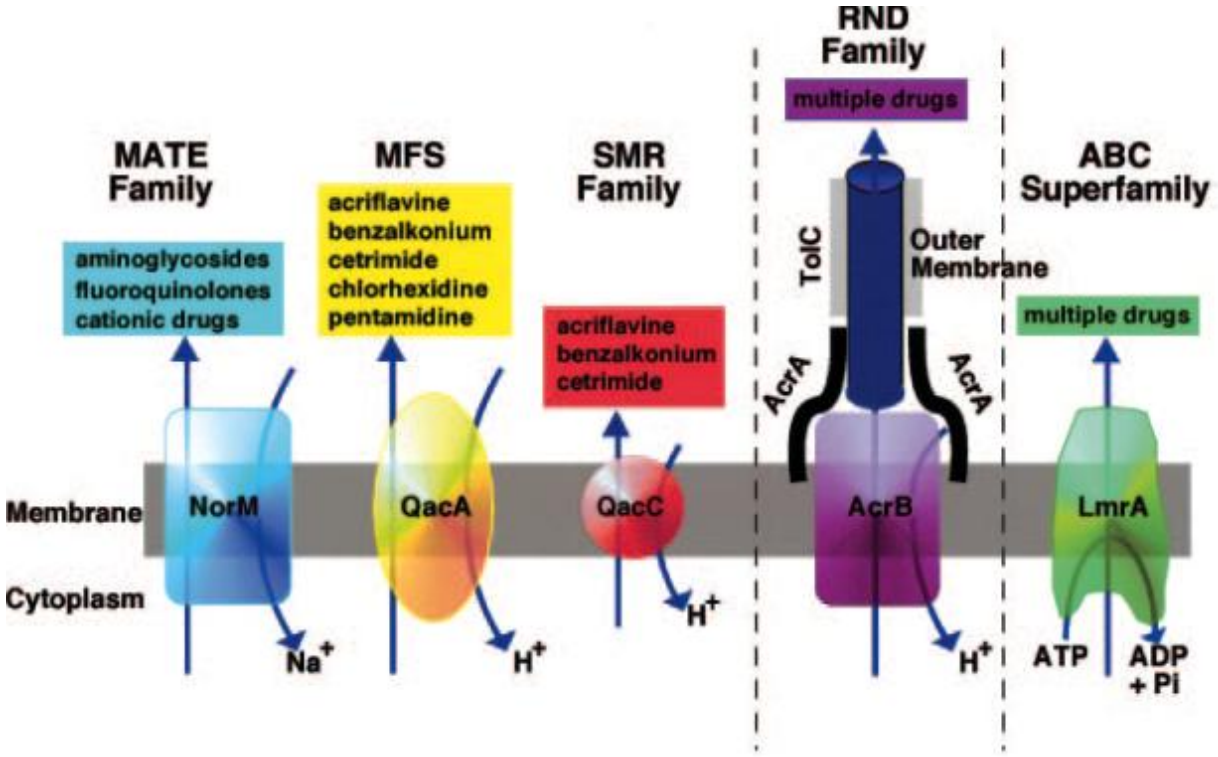


Figure 1.8 Diagrammatic comparison of the five families of efflux pumps [37].

Multidrug and Toxin Extrusion (MATE) proteins form a class of proteins that function as drug and proton antiporters. Initially, due to the presence of 12 transmembrane helices, they were designated as the member of MFS family. Shortly afterwards, it was reported that they showed no sequence identity to other known multidrug transporters, therefore, categorized as a new family of multidrug transporters, and are widely propagated in all realms of living beings [40]. MATE proteins have been characterized as important transporters that mediate the final excretion of cationic drugs into bile and urine [41]. In plants, transporter proteins from the

MATE family are essential in metabolite transport, which directly changes crop yields. In bacteria and mammals, these MATE transporters facilitate multiple-drug resistance (MDR), thus regulating the efficacy of many pharmaceutical drugs used in curing a variety of diseases [42]. MATE family transporters are conserved in the three pinion domains of life (Archaea, Bacteria and Eukarya), and export xenobiotics using an electrochemical exchange of H+ or Na+ across the tissue layer. MATE transporters confer resistance to bacterial pathogens and cancer cells, thus causing critical reductions in the curative efficacies of antibiotics and anti-cancer drugs, respectively [43]. An example of one such protein is NorM, of *Vibrio parahaemolyticus* which isa multidrug Na+-antiporter, and was found to confer resistance to dyes, fluoroquinolones and aminoglycosides[44, 45].

Multidrug and toxin extrusion protein (MATE1) is another type of efflux transporter identified quite recently, which is expressed in the kidney and liver, being localized at the apical membranes facing the lumen of the renal tubules and bile canaliculi, respectively. It mediates the excretion of organic cations, such as TEA and cimetidine, involving transmembrane proton gradient as a driving force [46]. Although MATE1 has been characterized as an organic cation/H^+ antiporter, it has recently been shown that human MATE1 (hMATE) can also transport some organic anions and amphoteric compounds [47].

As reported, MATE family efflux pumps depend upon Na+/H+ gradient for transport and have three major branches: the NorM branch, a branch containing several eukaryotic proteins and a branch containing *E. coli* DinF [44]. MATE protein length varies from 400 to 700 residues comprising of 12 transmembrane helices (Figure 1.9). In MATE proteins, there is no conserved consensus sequence; however they share ~40% sequence similarity[40]. It has been noted that extremely conserved regions of varied length ranging from 17 to 25 amino acid short stretch are located near close to transmembrane helix 1 (TM1) and TM7; in extracellular loops between TM1 and TM2, and TM7 and TM8; in cytoplasmic loops between TM2 and TM3, and TM8 and TM9; and in loops between TM4 and TM5, and TM10 and TM11 [40].

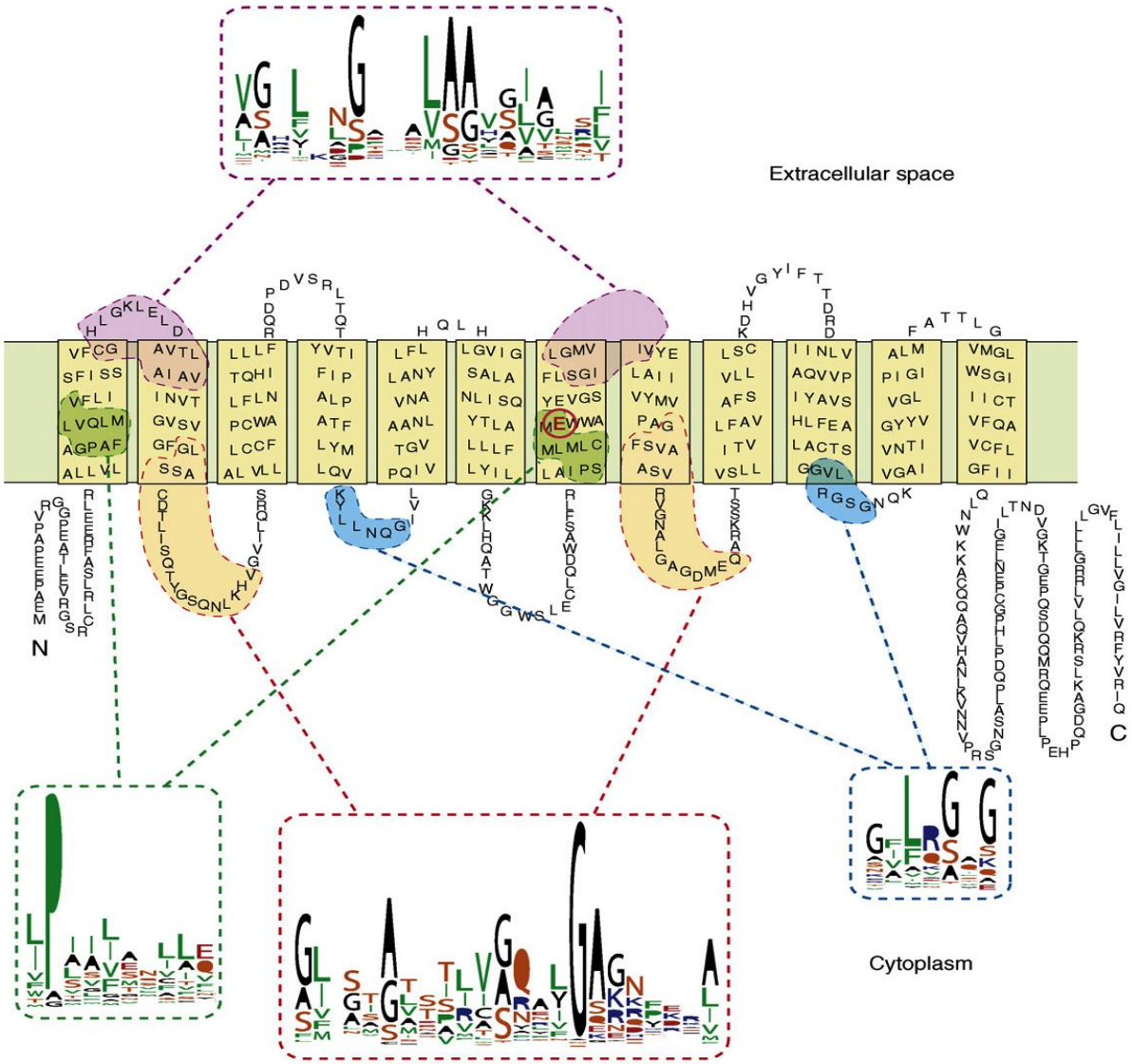


Figure 1.9 Typical secondary structure of a MATE-type transporter. Shown is a secondary structure model of human MATE1 [38]. Hydropathy analysis of typical MATE-type transporters predicts 12 transmembrane domains. A glutamate residue, E273, which is essential in human MATE1, is circled. The regions that are relatively well conserved among all known MATEs (Figure 1) are delineated by broken lines and colored as follows: TM1 and TM7 (green); extracellular loops between TM1 and TM2, and TM7 and TM8 (purple); cytoplasmic loops between TM2 and TM3, and TM8 and TM9 (orange); and loops between TM4 and TM5, and TM10 and TM11 (blue) [40].

1.8.1. MATE and Diarrhea

Vibrio cholerae, an important gram-negative enteric pathogen, is the causative agent of the severe diarrheal disease cholera [27]. Colmer*et al.* in 1988, identified VceAB, a multidrug resistance pump that provides *V. cholerae* with resistance to several toxic compounds, such as deoxycholate and the antibiotics nalidixic acid and chloramphenicol [48]. Coupling with Na^+ and substrate is an interesting characteristic of MATE-type efflux pumps, and this phenomenon has been reported in most MATE-type efflux pumps from several diarrheal pathogens which includes NorM from *V. parahaemolyticus*, VcmA from *V. cholerae* non-O1, VmrA from *V. parahaemolyticus* and VcrM from *V. cholerae* non-O1 [49]. NorM and its *Escherichia coli* homolog YdhE mediate resistance to dyes, hydrophilic fluoroquinolones, and aminoglycosides and thus facilitate the bacterial growth [27]. Mdtk, a MATE type efflux system have been reported in *Salmonella enterica* which is are responsible for causing acute gastroenteritis and typhoid [50]. A MATE type MDR pump was identified for *Clostridium difficile* also, but no member of this family has yet been described in other gram-positive bacteria [51].

Granting to the studies, it has been reported that sequence information for very few MATE proteins is available till date. Also, due to its primary structure heterogeneity, it is hard to recognize these proteins based on sequence similarity. To combat the problem of drug resistance, it is all important to extensively understand and identify multidrug resistance proteins at a faster pace.

1.9. ANTIBIOTIC RESISTANCE IN TRAVELER'S DIARRHEA

Among all pervading diarrheal types, Travelers' Diarrhea (TD) is one of the most frequent illness among individuals travelling to developing countries [52]. It usually begins within the first week of travel and usually resolves within 3 to 5 days [53].However, symptoms can be severe enough to force a change in travel plans and to result in confinement to bed or, rarely, hospitalization [54]. It is induced by an infection acquired by consuming contaminated food or drinks or due to climatic conditions [55] . It brings heavy economic costs, both to the people who travel and to developing countries through loss of tourism income and loss of business investment opportunities caused by the threat of disease [53]. Various pathogens including *Enterotoxigenic E.coli (ETEC*), E*nteroaggregative E.coli (EAEC)* and *Campylobacter* have been

identified as the pathological agents of traveler's diarrhea (TD), with S*higella* spp. being one of the most common etiological agents. Other bacteria that cause diarrhea, such as *Salmonella, Yersinia, Aeromonas,* and *Plesiomonas* spp., are isolated less often [56]. Several antibiotics such as quinolones (ciprofloxacin, norfloxacin), rifaximin and azithromycin were reported to be effective and safe to use against travelers' diarrhea [53]. But it has been found that S*higella* spp. acquired resistance to these clinically important antibiotics [57]. In the treatment of enteric infections, antibiotic resistance is becoming increasingly important, particularly those due to *Shigella, Vibrio cholerae,* enterotoxigenic *Escherichia coli* (associated with traveler's diarrhea), and *Salmonella typhi*. The rate of resistance is highest in the regions, where the use of antibiotics is relatively unrestricted [58]. In the last few years, a dramatic escalation has been seen in the antibiotic resistance profile of S*higella* spp. [59]. Increased antibiotic resistance is a great impediment in control of the traveler's diarrhea and thus results in greater disease burden globally [60]. Of greatest immediate concern is the need for an effective, inexpensive antibiotic that can be used safely as treatment to Travelers diarrhea due to *Shigella,* primarily *Shigella dysenteriae* type 1 and *Shigella flexneri* [58]. Emerging resistance to fluoroquinolones such as ciprofloxacin has been studied in several bacteria, such as in *Staphylococcus aureus, Streptococcus pneumoniae, Pseudomonas aeruginosa,* and *Mycobacterium tuberculosis* [61]. Fluoroquinolones are one of the most commonly prescribed classes of antibacterials in the world and are used to treat a variety of bacterial infections in humans [62]. These agents generally consist of a 1-substituted-1, 4-dihydro-4-oxopyridine-3-carboxylic acid moiety combined with an aromatic or hetero-aromatic ring fused at the 5- and 6-positions [63]. They interact with 2 bacterial targets, the related enzymes DNA Gyrase A (GyrA) and topoisomerase IV (ParC), both of which are involved in DNA replication [64]. Fluoroquinolones form complexes of these enzymes with DNA, complexes that block movement of the DNA-replication fork and thereby inhibit DNA replication. DNA gyrase is the only bacterial enzyme that introduces negative superhelical twists into DNA, which is responsible for initiation of DNA replication [65]. Removal of positive superhelical twists that accumulate ahead of the replication fork or as a result of the transcription of certain genes is also facilitated by DNA Gyrase [65, 66]. Fluoroquinolones inhibit enzymes by stabilizing the DNA-DNA gyrase complex [64, 67], causing formerly reversible DNA-enzyme complexes to become irreversible due to inhibition of replication fork movement [68]. Damage to DNA and the generation of DNA-strand breaks then

trigger a set of events, as yet poorly defined, that follow the rapid inhibition of DNA synthesis and result in eventual cell death [66, 67]. Topological stress that arises from the translocation of transcription and replication complexes along DNA is relieved by DNA gyrase; wheras topoisomerase IV (ParC) being a decatenating enzyme resolves interlinked daughter chromosomes following DNA replication [69]. In last few years, large numbers of studies related to resistance mechanism have been reported, but structural level analysis revealing the mode of interaction of GyrA and ParC with fluoroquinolones yet needs to be explored. A study reported structural insights into the fluoroquinolone resistance mechanism of *Mycobacterium tuberculosis* DNA gyrase at atomic level [70]. This analyzed the functional, biophysical and structural studies of the two individual domains constituting the catalytic DNA gyrase and thus identified original mechanistic properties of quinolone binding that represent relationships between amino acid mutations and resistance phenotype [70]. Due to its ability to control the topological state of DNA molecule during replication process [71], DNA gyrase plays a significant role in survival of *Shigella flexneri* . In some of the bacteria like *Shigella flexneri*, *Escherichia coli* etc., DNA gyrase acts as the primary target for quinolones [72]. Therefore, it is a suitable candidate to study the effect of mutations on quinolone resistance.

REFERENCES

[1] World Health Organisation, “Diarrhoeal disease Fact sheet,” 2017.

[2] Centers for Disease Control and Prevention, 2013.

[3] S. Lakshminarayanan and R. Jayalakshmy, “Diarrheal diseases among children in India: Current scenario and future perspectives,” *Journal of Natural Science, Biology and Medicine*, vol. 6, (no. 1), pp. 24-28, 2015.

[4] J.P. Nataro and J.B. Kaper, “Diarrheagenic Escherichia coli,” *Clinical Microbiology Reviews*, vol. 11, (no. 1), pp. 142-201, 1998.

[5] J.E. Galan, “Salmonella Interactions with Host Cells: Type III Secretion at Work,” *Annual Review of Cell and Developmental Biology*, vol. 17, (no. 1), pp. 53-86, 2001.

[6] World Gastroenterology Organisation, “Acute Diarrhea,” 2008.

[7] J. Yang, L. Chen, J. Yu, L. Sun, and Q. Jin, “ShiBASE: an integrated database for comparative genomics of Shigella,” *Nucleic Acids Research*, vol. 34, (no. suppl 1), pp. D398-D401, 2006.

[8] T. Shimohata and A. Takahashi, "Diarrhea induced by infection of Vibrio parahaemolyticus," *The Journal of Medical Investigation*, vol. 57, pp. 179-182, 2010.

[9] O.S. Shin, V.C. Tam, M. Suzuki, J.M. Ritchie, R.T. Bronson, M.K. Waldor, and J.J. Mekalanos, "Type III Secretion Is Essential for the Rapidly Fatal Diarrheal Disease Caused by Non-O1, Non-O139 Vibrio cholerae," *mBio*, vol. 2, (no. 3), 2011.

[10] G.N. Schroeder and H. Hilbi, "Molecular Pathogenesis of Shigella spp.: Controlling Host Cell Signaling, Invasion, and Death by Type III Secretion," *Clinical Microbiology Reviews*, vol. 21, (no. 1), pp. 134-156, 2008.

[11] A.J. Müller, C. Hoffmann, M. Galle, A. Van Den Broeke, M. Heikenwalder, L. Falter, B. Misselwitz, M. Kremer, R. Beyaert, and W.-D. Hardt, "The S. Typhimurium Effector SopE Induces Caspase-1 Activation in Stromal Cells to Initiate Gut Inflammation," *Cell host & microbe*, vol. 6, (no. 2), pp. 125-136, 2009.

[12] G.R. Cornelis, "The type III secretion injectisome," *Nat Rev Micro*, vol. 4, (no. 11), pp. 811-825, 2006.

[13] P. Dube, "Interaction of Yersinia with the Gut: Mechanisms of pathogenesis and immune evasion," *Current Topics in Microbiology and Immunology*, vol. 337, (no. 1), pp. 61-91, 2009.

[14] S. Massa, R. Armuzzi, M. Tosques, F. Canganella, and L.D. Trovatelli, "Susceptibility to chlorine of Aeromonas hydrophila strains," *Journal of Applied Microbiology*, vol. 86, (no. 1), pp. 169-173, 1999.

[15] M. Aslani and M. Alikhani, "The Role of Aeromonas hydrophila in Diarrhea.," *Iranian Journal of Public Health* vol. 33, pp. 60-67, 2004.

[16] U.D. Parashar, C.J. Gibson, J. Bresee, and R.I. Glass, "Rotavirus and Severe Childhood Diarrhea," *Emerging Infectious Diseases*, vol. 12, (no. 2), pp. 304-306, 2006.

[17] T. Capizzi, G. Makari-Judson, R. Steingart, and W. Mertens, "Chronic diarrhea associated with persistent norovirus excretion in patients with chronic lymphocytic leukemia: report of two cases," *BMC Infectious Diseases*, vol. 11, (no. 1), pp. 131, 2011.

[18] M. Aminu, A. Ahmad, J. Umoh, M. de Beer, M. Esona, and A. Steele, "Adenovirus infection in children with diarrhea disease in Northwestern Nigeria", *Annals of African Medicine*, vol 6, (no. 4), pp. 168-173, 2007.

[19] A. Ghouil, N.T., O. Gascuel, F. Z.Guerfali, D. Laouini, EricMarechal, and LaurentBrehelin., "EuPathDomains: The divergent domain database for eukaryotic pathogens.," *Infection, Genetics and Evolution*, vol. 11, pp. 698–707, 2010.

[20] M.A. Hegazi, T.A. Patel, and B.S. El-Deek, "Prevalence and characters of Entamoeba histolytica infection in Saudi infants and children admitted with diarrhea at 2 main hospitals at south Jeddah: a re-emerging serious infection with unusual presentation," *The Brazilian Journal of Infectious Diseases*, vol. 17, (no. 1), pp. 32-40, 2013.

[21] C.L.F. Walker and R.E. Black, "Zinc for the treatment of diarrhoea: effect on diarrhoea morbidity, mortality and incidence of future episodes," *International Journal of Epidemiology*, vol. 39, (no. suppl 1), pp. i63-i69, 2010.

[22] M.K. Munos, C.L.F. Walker, and R.E. Black, "The effect of oral rehydration solution and recommended home fluids on diarrhoea mortality," *International Journal of Epidemiology*, vol. 39, (no. suppl 1), pp. i75-i87, 2010.

[23] A. Guarino, A.L. Vecchio, and R.B. Canani, "Probiotics as prevention and treatment for diarrhea," *Current Opinion in Gastroenterology*, vol. 25, (no. 1), pp. 18-23, 2009.

[24] Senka Dzidic, J. Suskovic, and B.e. Kos, "Antibiotic Resistance Mechanisms in Bacteria:Biochemical and Genetic Aspects," *Food Technology & Biotechnology*, vol. 46, (no. 1), pp. p11, 2008.

[25] C. Nathan and O. Cars, "Antibiotic Resistance â€" Problems, Progress, and Prospects," *New England Journal of Medicine*, vol. 371, (no. 19), pp. 1761-1763, 2014.

[26] A. Sefton, "Mechanisms of Antimicrobial Resistance," *Drugs*, vol. 62, (no. 4), pp. 557-566, 2002.

[27] M. Putman, H.W. van Veen, and W.N. Konings, "Molecular Properties of Bacterial Multidrug Transporters," *Microbiology and Molecular Biology Reviews*, vol. 64, (no. 4), pp. 672-693, 2000.

[28] M.N. Alekshun and S.B. Levy, "Molecular Mechanisms of Antibacterial Multidrug Resistance," *Cell*, vol. 128, (no. 6), pp. 1037-1050, 2007.

[29] C. Walsh, "Molecular mechanisms that confer antibacterial drug resistance," *Nature*, vol. 406, (no. 6797), pp. 775-781, 2000.

[30] J. Davies, "Inactivation of antibiotics and the dissemination of resistance genes," *Science*, vol. 264, (no. 5157), pp. 375-382, 1994.

[31] A.C. Fluit, M.R. Visser, and F.-J. Schmitz, "Molecular Detection of Antimicrobial Resistance," *Clinical Microbiology Reviews*, vol. 14, (no. 4), pp. 836-871, 2001.

[32] P.G. Higgins, A.C. Fluit, D. Milatovic, J. Verhoef, and F.J. Schmitz, "Antimicrobial susceptibility of imipenem-resistant Pseudomonas aeruginosa," *Journal of Antimicrobial Chemotherapy*, vol. 50, (no. 2), pp. 299-301, 2002.

[33] http://amrls.cvm.msu.edu/microbiology.

[34] N. Karami, F. Nowrouzian, I. Adlerberth, and A.E. Wold, "Tetracycline Resistance in Escherichia coli and Persistence in the Infantile Colonic Microbiota," *Antimicrobial Agents and Chemotherapy*, vol. 50, (no. 1), pp. 156-161, 2006.

[35] D.M. Livermore, M. Warner, S. Mushtaq, M. Doumith, J. Zhang, and N. Woodford, "What remains against carbapenem-resistant Enterobacteriaceae? Evaluation of chloramphenicol, ciprofloxacin, colistin, fosfomycin, minocycline, nitrofurantoin, temocillin and tigecycline," *International Journal of Antimicrobial Agents*, vol. 37, (no. 5), pp. 415-419, 2011.

[36] Y.L. Jeon, Y.-s. Nam, G. Lim, S.Y. Cho, Y.-T. Kim, J.-H. Jang, J. Kim, M. Park, and H.J. Lee, "Quinolone-resistant Shigella flexneri isolated in a patient who travelled to India," *Annals of laboratory medicine*, vol. 32, (no. 5), pp. 366-369, 2012.

[37] L.J.V. Piddock, "Clinically Relevant Chromosomally Encoded Multidrug Resistance Efflux Pumps in Bacteria," *Clinical Microbiology Reviews*, vol. 19, (no. 2), pp. 382-402, 2006.

[38] K. Poole, "Efflux-mediated antimicrobial resistance," *Journal of Antimicrobial Chemotherapy*, vol. 56, (no. 1), pp. 20-51, 2005.

[39] M. Otsuka, T. Matsumoto, R. Morimoto, S. Arioka, H. Omote, and Y. Moriyama, "A human transporter protein that mediates the final excretion step for toxic organic cations," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, (no. 50), pp. 17923-17928, 2005.

[40] H. Omote, M. Hiasa, T. Matsumoto, M. Otsuka, and Y. Moriyama, "The MATE proteins as fundamental transporters of metabolic and xenobiotic organic cations," *Trends in pharmacological sciences*, vol. 27, (no. 11), pp. 587-593, 2006.

[41] A.T. Nies, K. Damme, E. Schaeffeler, and M. Schwab, "Multidrug and toxin extrusion proteins as transporters of antimicrobial drugs," *Expert Opinion on Drug Metabolism & Toxicology*, vol. 8, (no. 12), pp. 1565-1577, 2012.

[42] X. He, P. Szewczyk, A. Karyakin, M. Evin, W.-X. Hong, Q. Zhang, and G. Chang, "Structure of a cation-bound multidrug and toxic compound extrusion transporter," *Nature*, vol. 467, (no. 7318), pp. 991-994, 2010.

[43] Y. Tanaka, C.J. Hipolito, A.D. Maturana, K. Ito, T. Kuroda, T. Higuchi, T. Katoh, H.E. Kato, M. Hattori, K. Kumazaki, T. Tsukazaki, R. Ishitani, H. Suga, and O. Nureki, "Structural basis for the drug extrusion mechanism by a MATE multidrug transporter," *Nature*, vol. 496, (no. 7444), pp. 247-251, 2013.

[44] P. Mohanty, A. Patel, and A. Kushwaha Bhardwaj, "Role of H- and D- MATE-type transporters from multidrug resistant clinical isolates of Vibrio fluvialis in conferring fluoroquinolone resistance," *PloS one*, vol. 7, (no. 4), pp. e35752, / 2012.

[45] X.-Z. Li and H. Nikaido, "Efflux-Mediated Drug Resistance in Bacteria," *Drugs*, vol. 64, (no. 2), pp. 159-204, 2004.

[46] K.-y. Ohta, Y. Imamura, N. Okudaira, R. Atsumi, K. Inoue, and H. Yuasa, "Functional Characterization of Multidrug and Toxin Extrusion Protein 1 as a Facilitative Transporter for Fluoroquinolones," *Journal of Pharmacology and Experimental Therapeutics*, vol. 328, (no. 2), pp. 628-634, 2009.

[47] Y. Tanihara, S. Masuda, T. Sato, T. Katsura, O. Ogawa, and K.-i. Inui, "Substrate specificity of MATE1 and MATE2-K, human multidrug and toxin extrusions/H+-organic cation antiporters," *Biochemical Pharmacology*, vol. 74, (no. 2), pp. 359-371, 2007.

[48] J.A. Colmer, J.A. Fralick, and A.N. Hamood, "Isolation and characterization of a putative multidrug resistance pump from Vibrio cholerae," *Molecular Microbiology*, vol. 27, (no. 1), pp. 63-72, 1998.

[49] K. Hashimoto, W. Ogawa, T. Nishioka, T. Tsuchiya, and T. Kuroda, "Functionally Cloned pdrM from *Streptococcus pneumoniae* Encodes a Na+Coupled Multidrug Efflux Pump," *PLoS ONE*, vol. 8, (no. 3), pp. e59525, 2013.

[50] K. Nishino, E. Nikaido, and A. Yamaguchi, "Regulation and physiological function of multidrug efflux pumps in Escherichia coli and Salmonella," *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, vol. 1794, (no. 5), pp. 834-843, 2009.

[51] G.W. Kaatz, F. McAleese, and S.M. Seo, "Multidrug Resistance in Staphylococcus aureus Due to Overexpression of a Novel Multidrug and Toxin Extrusion (MATE) Transport Protein," *Antimicrobial Agents and Chemotherapy*, vol. 49, (no. 5), pp. 1857-1864, 2005.

[52] S. Alajbegovic, J. Sanders, D. Atherly, and M. Riddle, "Effectiveness of rifaximin and fluoroquinolones in preventing travelers' diarrhea (TD): a systematic review and meta-analysis," *Systematic Reviews*, vol. 1, (no. 1), pp. 39, 2012.

[53] D.J. Diemert, "Prevention and Self-Treatment of Traveler's Diarrhea," *Clinical Microbiology Reviews*, vol. 19, (no. 3), pp. 583-594, 2006.

[54] R.A. Kuschner, A.F. Trofa, R.J. Thomas, C.W. Hoge, C. Pitarangsi, S. Amato, R.P. Olafson, P. Echeverria, J.C. Sadoff, and D.N. Taylor, "Use of Azithromycin for the Treatment of Campylobacter Enteritis in Travelers to Thailand, an Area Where Ciprofloxacin Resistance Is Prevalent," *Clinical Infectious Diseases*, vol. 21, (no. 3), pp. 536-541, 1995.

[55] Javier De la Cabada Bauche and H.L. DuPont, "New Developments in Traveler's Diarrhea," *Gastroenterology and Hepatology*, vol. 7, (no. 2), pp. 88-95, 2011.

[56] V. Jordi, R. Joaquin, G. Francisco, V. Martha, S. Lara, F. Maria Jose, and G. Joaquin, "Aeromonas spp. and Travelers Diarrhea: Clinical Features and Antimicrobial Resistance," *Emerging Infectious Disease journal*, vol. 9, (no. 5), pp. 552, 2003.

[57] L. Mensa, F. Marco, J. Vila, J. Gascón, and J. Ruiz, "Quinolone resistance among Shigella spp. isolated from travellers returning from India," *Clinical Microbiology and Infection*, vol. 14, (no. 3), pp. 279-281, 2008.

[58] R.B. Sack, M. Rahman, M. Yunus, and E.H. Khan, "Antimicrobial Resistance in Organisms Causing Diarrheal Disease," *Clinical Infectious Diseases*, vol. 24, (no. Supplement 1), pp. S102-S105, 1997.

[59] K.D. von Seidlein L, Ali M, Lee H, Wang X, Thiem VD, et al., "A multicenter study of Shigella diarrhea in six Asian countries: disease burden, clinical manifestation, and microbiology," *PLoS Med*, vol. 3, (no. e353), 2006.

[60] M.B. Zaidi, T. Estrada-Garcia, F.D. Campos, R. Chim, F. Arjona, M. Leon, A. Michell, and D. Chaussabel, "Incidence, clinical presentation, and antimicrobial resistance trends

in Salmonella and Shigella infections from children in Yucatan, Mexico," *Frontiers in Microbiology*, vol. 4, 2013.

[61] F.D. Lowy, "Antimicrobial resistance: the example of Staphylococcus aureus," *The Journal of Clinical Investigation*, vol. 111, (no. 9), pp. 1265-1273, 2003.

[62] K.J. Aldred, R.J. Kerns, and N. Osheroff, "Mechanism of Quinolone Action and Resistance," *Biochemistry*, vol. 53, (no. 10), pp. 1565-1574, 2014.

[63] Vashist J, Vishvanath, Kapoor R, Kapil A, Yennamalli R, Subbarao N, and R. MR., "Interaction of nalidixic acid and ciprofloxacin with wild type and mutated quinolone-resistance-determining region of DNA gyrase A.," *Indian Journal of Biochemistry and Biophysics*, vol. 46, (no. 2), pp. 147-153, 2009.

[64] D.C. Hooper, "Mechanisms of Action and Resistance of Older and Newer Fluoroquinolones," *Clinical Infectious Diseases*, vol. 31, (no. 2), pp. S24-S28, 2000.

[65] J.C. Wang, "DNA Topoisomerases," *Annual Review of Biochemistry*, vol. 65, (no. 1), pp. 635-692, 1996.

[66] D.C. Hooper, "Bacterial Topoisomerases, Anti-Topoisomerases, and Anti-Topoisomerase Resistance," *Clinical Infectious Diseases*, vol. 27, (no. 1), pp. S54-S63, 1998.

[67] K. Drlica and X. Zhao, "DNA gyrase, topoisomerase IV, and the 4-quinolones," *Microbiology and Molecular Biology Reviews*, vol. 61, (no. 3), pp. 377-92, 1997.

[68] H. Hiasa, D.O. Yousef, and K.J. Marians, "DNA Strand Cleavage Is Required for Replication Fork Arrest by a Frozen Topoisomerase-Quinolone-DNA Ternary Complex," *Journal of Biological Chemistry*, vol. 271, (no. 42), pp. 26424-26429, 1996.

[69] a.Z.X. Drlica K., "DNA gyrase, topoisomerase IV, and the 4-quinolones.," *Microbiology and Molecular Biology Reviews* vol. 61, pp. 377-92., 1997.

[70] J. Piton, S. Petrella, M. Delarue, G. Andre-Leroux, V. Jarlier, A. Aubry, and C. Mayer, "Structural Insights into the Quinolone Resistance Mechanism of *Mycobacterium tuberculosis* DNA Gyrase," *PLoS ONE*, vol. 5, (no. 8), pp. e12245, 2010.

[71] R.J. Reece, A. Maxwell, and J.C. Wang, "DNA Gyrase: Structure and Function," *Critical Reviews in Biochemistry and Molecular Biology*, vol. 26, (no. 3-4), pp. 335-375, 1991.

[72] J. Ruiz, "Mechanisms of resistance to quinolones: target alterations, decreased accumulation and DNA gyrase protection," *Journal of Antimicrobial Chemotherapy*, vol. 51, (no. 5), pp. 1109-1117, 2003.

CHAPTER –2

To develop the database dbDiarrhea: The database of pathogen proteins and vaccine antigens from diarrheal pathogens

ABSTRACT

Diarrhea occurs world-wide and is most commonly caused by gastrointestinal infections which kill around 1.7 billion people globally each year, mostly children in developing countries. We describe here dbDiarrhea, which is currently the most comprehensive catalog of proteins implicated in the pathogenesis of diarrhea caused by major bacterial, viral and parasitic species. The current release of the database houses 820 proteins gleaned through an extensive and critical survey of research articles from PubMed. The major contributors to this compendium of proteins are *Escherichia coli* and *Salmonella enterica*. These proteins are classified into different categories such as Type III secretion system effectors, Type III secretion system components, and Pathogen proteins. There is another complementary module called 'Host proteins'. dbDiarrhea also serves as a repository of the research articles describing 1) trials of subunit and whole organism vaccines 2) high-throughput screening of Type III secretion system inhibitors and 3) diagnostic assays, for various diarrheal pathogens. The database is web accessible through an intuitive user interface that allows querying proteins and research articles for different organism, keywords and accession number. Besides providing the search facility through browsing, the database supports sequence similarity search with the BLAST tool. With the rapidly burgeoning global burden of the diarrhea, we anticipate that this database would serve as a source of useful information for furthering research on diarrhea. The database can be freely accessed at http://www.juit.ac.in/attachments/dbdiarrhea/diarrhea_home.html.

2.1 INTRODUCTION

Diarrhea is an increase in the frequency of bowel movements or a decrease in the form of stool. Diarrhea is a neglected tropical disease despite being a global scourge and international health challenge. Diarrhea exacts large tolls of morbidity and mortality among all age groups and is particularly endemic in developing countries. It causes about 1.7 billion deaths worldwide [1]. Diarrheal disease is the leading infectious cause of childhood morbidity and mortality [2] and is responsible for killing around 7,60,000 children every year [1].

The most diarrhea episodes are self limiting and dehydration can usually be controlled with oral rehydration therapy [3, 4], which is the key to management of acute watery diarrhea, whereas

antimicrobial agents are indicated for acute invasive diarrhea, particularly shigellosis and amebiasis to reduce the duration of the disease. Though current treatments for diarrhea have made significant strides in reducing deaths, the associated contraindications coupled with escalating resistance of pathogens to existing anti-microbial agents [5, 6] have underscored the need for the search of more effective drugs and novel vaccine candidates. Understanding the burden of pathogen specific diarrheal disease and the variation by region is important for planning effective control programs for the overall reduction of diarrhea disease among persons of all ages [3]. This warrants increased attention and directed research efforts toward understanding the diarrheagenic mechanisms of different pathogens. In furtherance of this goal, we have developed dbDiarrhea to serve as a central compendium of the critical protein players central to the etiology of diarrheal diseases. It also houses research articles related to vaccine and diagnostic trials for various diarrheal pathogens.

2.2 METHODOLOGY

2.2.1 Construction and Architecture of the Database

The database entries were manually curated by thoroughly searching research articles from PubMed as well as proteins from Uniprot, T3SEdb and GenBank using various keywords like ‘diarrhea’, ‘cholera’ and the names of various diarrheal pathogens (Figure 2.1).

This search was carried out by exploring the articles available in PubMed published during the period from 1990 to 2012. The result page showed a number of articles which were then filtered out to retrieve the relevant data with respect to each diarrheal pathogen and proteins involved in the pathogenesis. This was followed by the inclusion of additional information about the corresponding protein, Pfam domain using Pfam [7] database, functions and PDB identifiers wherever available. The complete protein information was extracted from GenBank (http://www.ncbi.nlm.nih.gov/genbank/) database and the related protein structure information from the PDB [8] database. Each protein was assigned to one of the following functional categories: adhesin, invasin, toxin, signal transduction, transporter, proteases and iron acquisition system protein. The information regarding T3SS effectors was retrieved from T3SEdb [9] database using in-house PERL script. Few reports have also indicated that T3SS inhibitors have the potential to be developed into novel antibacterial therapeutics [10, 11]. In this context, it was tenable to include proteins for Type III secretion system (T3SS) effectors and T3SS components.

These were retrieved from PubMed as well as the T3SEdb database [9] The database has been compiled through an extensive and thorough survey of the literature to include all the information available till date. The database can be easily updated by limiting the search using publication date using the 'Limits' option in NCBI.

The assembled proteins were categorized into different modules including module I called 'Pathogen Proteins'. In complementation to this is module II called 'Host Proteins' which refer to the human proteins involved in diarrhea infection. The modules III and IV include T3SS components and effectors respectively.

The modules V, VI and VII entail the studies describing (1) vaccine trials for whole organism as well as subunit vaccines. These vaccine candidates are also compared with the other vaccine databases such as VIOLIN [12] and it was found that many candidate strains such as SC599, WRSS1, WRSs2, and WRSs3 from *Shigella sp.* and 116E, MMU18006 from *Rotavirus* and many more are present in dbDiarrhea but untouched in VIOLIN and also VIOLIN doesn't cover the pathogens like *Clostridium difficile* and *Norovirus* which are the part of dbDiarrhea. Moreover, dbDiarrhea is focused only on the diarrheal pathogens but this is not the case with other databases that constitutes other pathogens also. (2) high-throughput screening of Type III secretion system inhibitors and (3) diagnostic assays respectively, for various diarrheal pathogens. Tables 2.1 to 2.3 enumerate the distribution of the proteins and research articles for different organisms in every module. Each entry in the modules I, II, V, VI and VII is linked to PubMed records corroborating the significance of the protein/vaccine strain in the context of diarrhea.

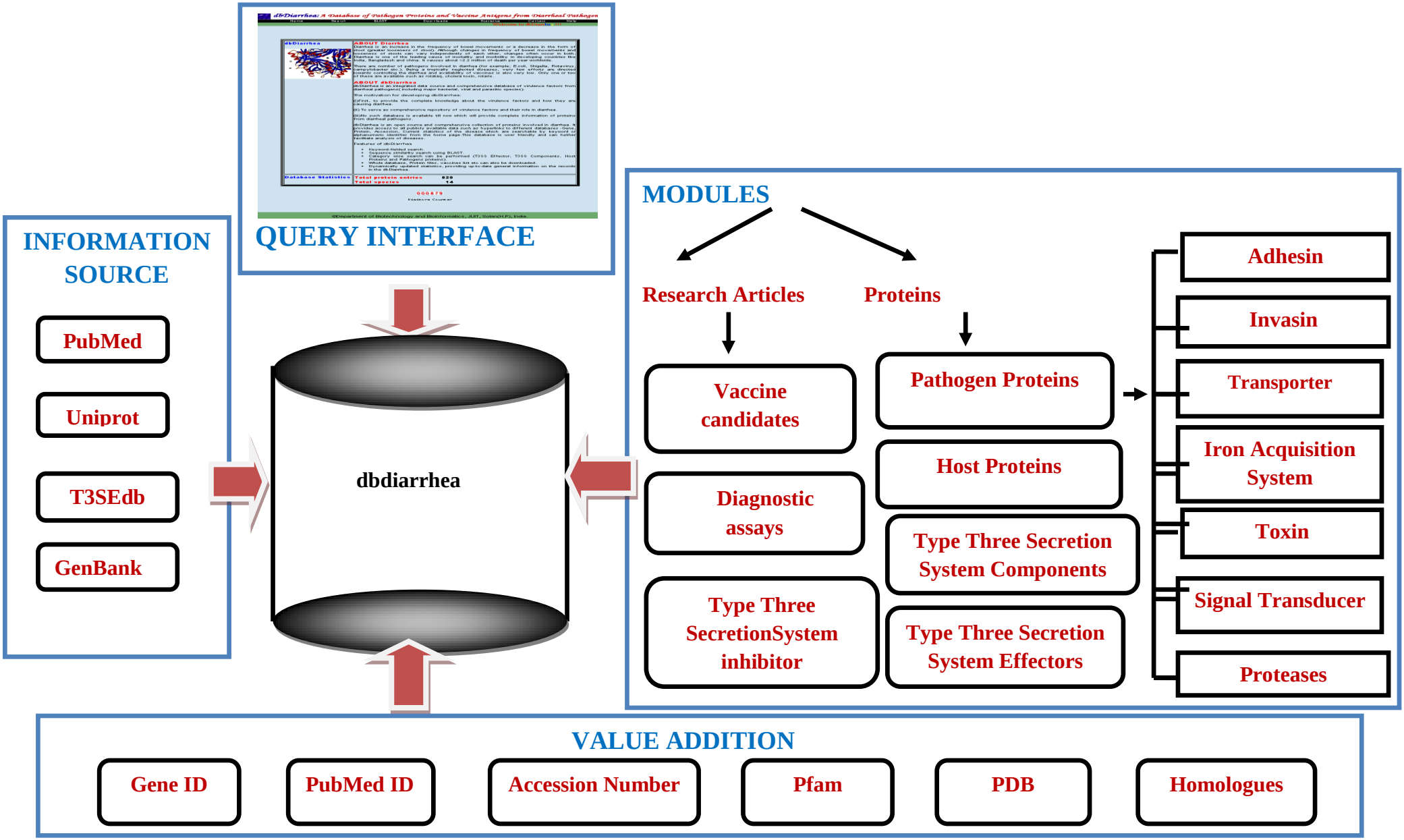

QUERY INTERFACE
INFORMATION SOURCE
PubMed
Uniprot
T3SEdb
GenBank
dbdiarrhea
MODULES
Research Articles
Proteins
Vaccine candidates
Diagnostic assays
Type Three SecretionSystem inhibitor
Pathogen Proteins
Host Proteins
Type Three Secretion System Components
Type Three Secretion System Effectors
Adhesin
Invasin
Transporter
Iron Acquisition System
Toxin
Signal Transducer
Proteases
VALUE ADDITION
Gene ID
PubMed ID
Accession Number
Pfam
PDB
Homologues

Figure 2.1 dbDiarrhea database schema. The protein sequences were collected by keyword search from different databases including PubMed, Uniprot, T3SEdb and GenBank. Value addition included the incorporation of Gene ID, PubMed ID, PDB, protein accession numbers, Pfam domain and homologs. The database includes modules containing proteins and research articles respectively. The proteins are grouped into four categories: Pathogen Proteins, Host Proteins, Type III secretion system (T3SS) components and T3SS effectors. The research articles include those covering vaccine trials, diagnostic assays and T3SS inhibitor studies

.Table 2.1 Organism-wise distribution of proteins in the database

Organism	Number of Proteins in the database
Escherichia coli	421
Shigella flexneri	88
Shigella dysenteriae	5
Shigella boydii	3
Shigella sonnei	3
Salmonella enteric	162
Yersinia enterocolitica	34
Cryptosporidium	2
Vibrio cholerae	27
Vibrio parahaemolyticus	16
Aeromonas hydrophila	7
Rotavirus	20
Campylobacter jejuni	25
Clostridium difficile	7
Total number of proteins	820

Table 2.2 Category-wise distribution of proteins in the database

	Categories			
Organisms	**Pathogen Proteins**	**Host proteins**	**T3SS Components**	**T3SS Effectors**
Escherichia coli	72	142	5	202
Shigella flexneri	19	32	21	16
Shigella dysenteriae	2	0	0	3
Shigella boydii	1	0	0	2
Shigella sonnei	0	0	0	3
Yersinia enterocolitica	5	0	13	16
Salmonella enterica	46	0	2	114
Cryptosporidium parvum	2	0	0	0
Vibrio cholerae	27	0	0	0
Vibrio parahaemolyticus	12	0	0	4
Aeromonas hydrophila	3	0	0	4
Rotavirus	6	14	0	0
Campylobacter jejuni	25	0	0	0
Clostridium difficile	7	0	0	0
Total	227	188	41	364

Table 2.3 List of total number of articles in the database describing vaccines candidates, Type Three Secretion System Inhibitors and Diagnostic assays, for various diarrheal pathogens present in the database

Organisms	Live attenuated		Subunit vaccines		T3SS inhibitors	Diagnostic assays
	No. of Strains	No. of References	No. of Strains	No. of References	No. of articles	No. of articles
Escherichia coli	25	46	20	27	18	43
Shigella dysenteriae	4	4	0	0	0	7
Shigella flexneri	10	10	4	4	3	0
Shigella sonnei	5	5	0	0	0	0
Salmonella enterica	4	4	1	1	9	14
Salmonella typhimurium	15	15	8	8	0	0
Yersinia enterocolitica	1	1	0	0	9	6
Campylobacter jejuni	1	1	2	5	0	5
Clostridium difficile	0	0	8	11	0	5
Vibrio cholerae	24	39	13	20	0	6
Vibrio parahaemolyticus	0	0	0	0	0	2
Rotavirus	11	19	2	2	0	14
Norovirus	0	0	2	3	0	10
Aeromonas hydrophila	0	0	0	0	0	0
Cryptosporidium	0	0	0	0	0	3
Entamoeba histolytica	0	0	0	0	0	3
Giardia lamblia	0	0	0	0	0	1
Total	100	146	6 0	81	39	119

dbDiarrhea is implemented as a MySQL database, which is connected to the HTML front-end through PHP using Microsoft IIS web server (Figure 2.2).

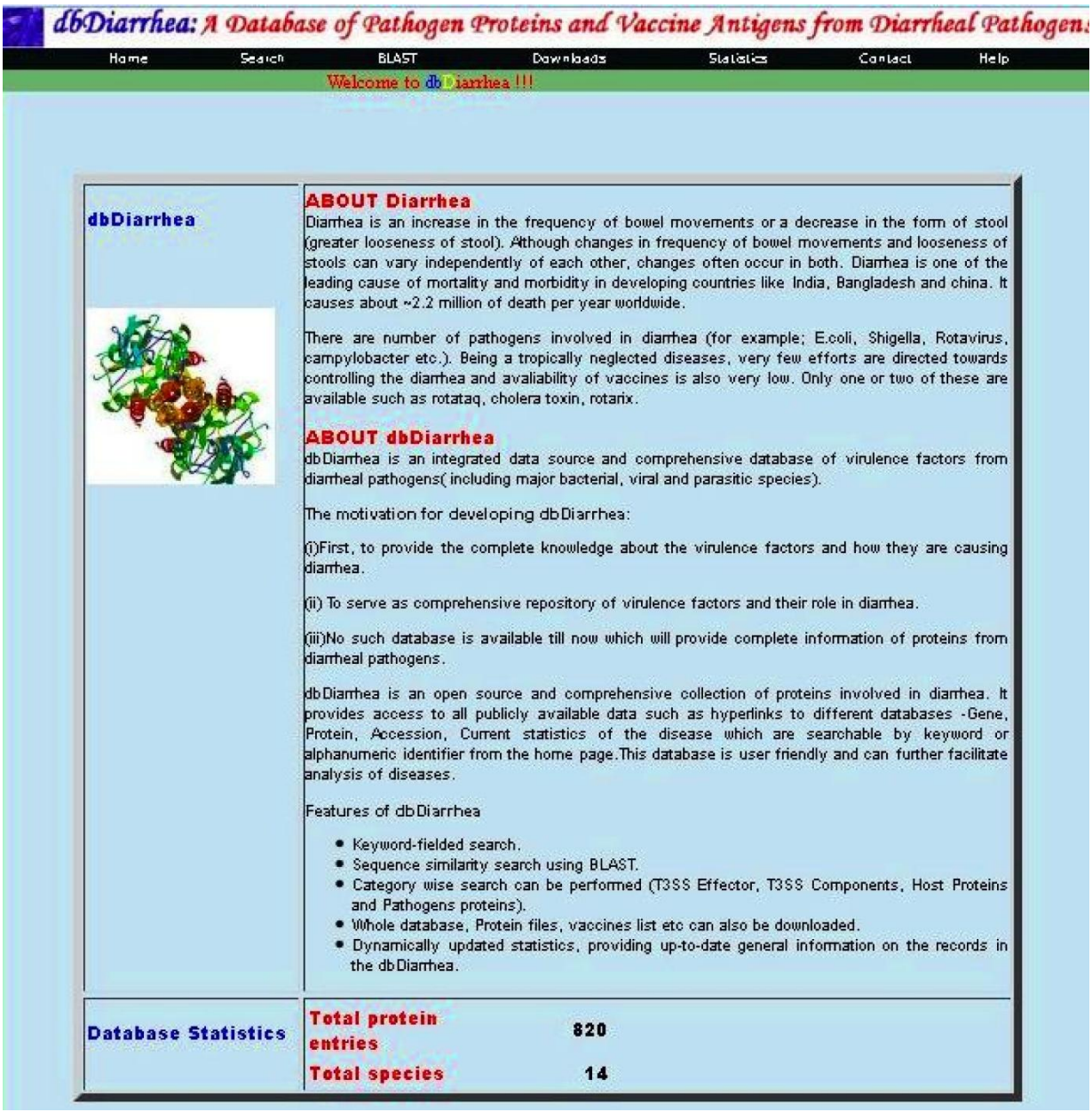


Figure 2.2 Snapshot of the database: dbDiarrhea.

2.3 RESULTS AND DISCUSSIONS

dbDiarrhea is the first web-based database that collects and compares diverse types of information about the virulence factors of multiple diarrheal pathogens. The database is openly accessible at http://www.juit.ac.in/attachments/dbdiarrhea/diarrhea_home.html. A user-friendly interface allows easy browsing and querying of information in various ways. The web query form allows users to selectively retrieve records from any module or functional category, for a

single or multiple species tabulating brief information for every protein or vaccine trial as shown in Figure 2.3.

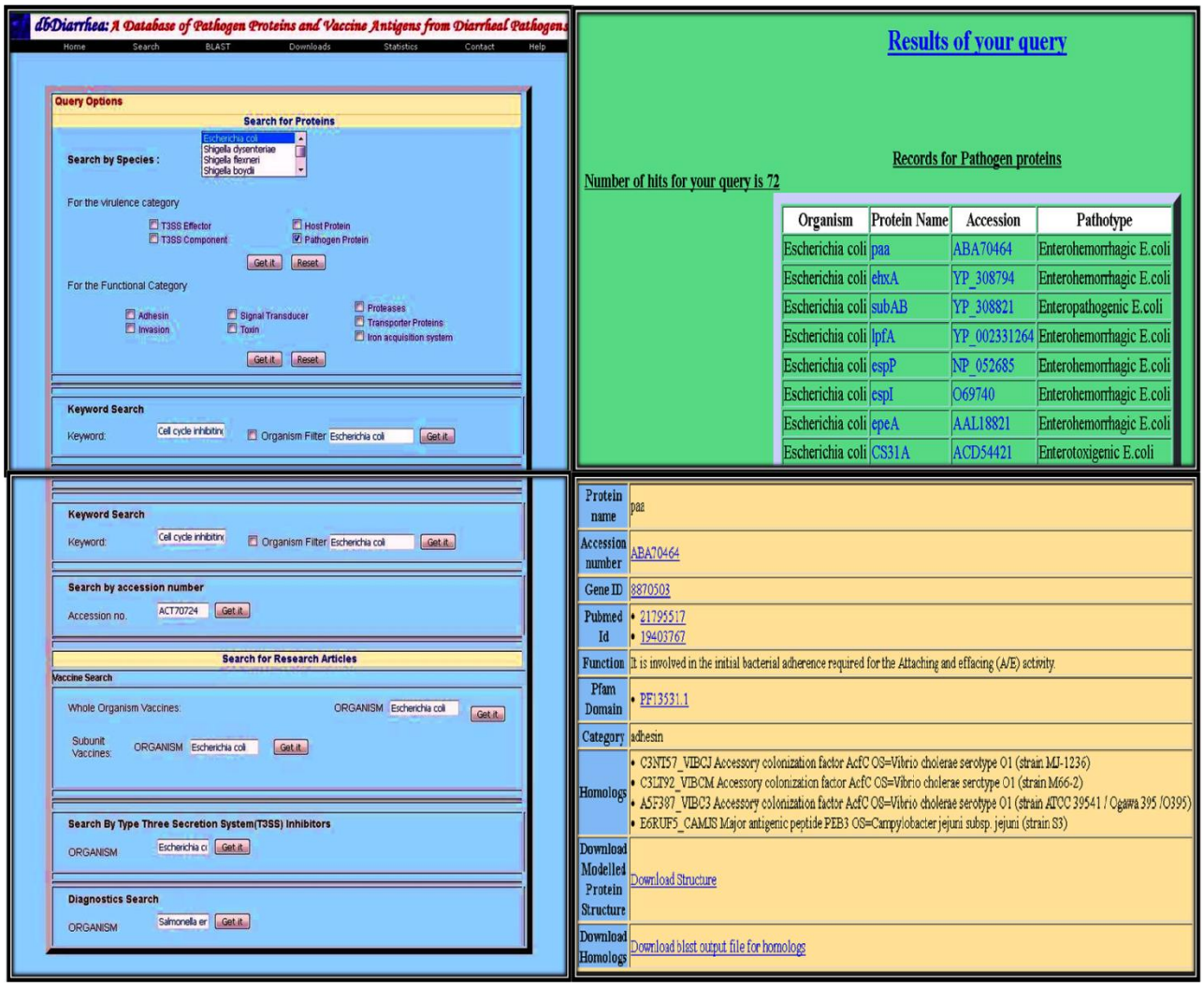


Figure 2.3 Snapshot of the search page of dbDiarrhea

The hyperlinks to each record further consolidate information on gene, protein, sequence and structure wherever available. The database may be queried with user-defined keywords, accession numbers and/or organism name. dbDiarrhea is integrated with BLAST (Basic Local Alignment Search Tool) [13] to provide sequence similarity search. Thus dbDiarrhea allows a researcher to identify the key players in the pathogenesis of diarrhea and search for the effective drugs and novel vaccine candidates. It would allow comparative analysis between different species, e.g. a protein responsible for diarrheal pathogenesis from one species like *Shigella*

dysenteriae might be closely related to some other pathogen protein whose function is not yet deciphered.

Many proteins in the database represent potential drug targets. The database enlists the PDB codes for the ones with solved three dimensional structures; for those with hitherto unsolved structures, we have modeled the eight proteins (aer, afa1, CDT-I, csgC, hlyA, paa, ssph1, Trh and Trk) using the Modeller module of Discovery Studio version 3.5 [14]. These models have also been uploaded on the web server. The accuracy of these models was validated using Ramachandran plot where the models with >98% residues in the allowed regions were selected as the final models. These models may be utilized by the user for the virtual screening of these models against compound libraries.

In future, dbDiarrhea will continue to be updated and refined with increased data so as to make it more worthwhile for the general users and also for the researchers working in this field.

2.4 CONCLUSION

The proteins in the database represent potential drug targets and/or vaccine candidates against the principal causal agents of diarrhea; some of these have already shown promise while others are still undergoing experimental investigation and many remain to be explored. We believe that this database will provide useful information portal for researchers working on diarrhea and various diarrheal diseases to translate the currently available knowhow into novel management and intervention strategies.

REFERENCES

[1] World Health Organisation, “Diarrhoeal disease Fact sheet,” 2013.

[2] C.L.F. Walker, I. Rudan, L. Liu, H. Nair, E. Theodoratou, Z.A. Bhutta, K.L. O'Brien, H. Campbell, and R.E. Black, “Global burden of childhood pneumonia and diarrhoea,” *The Lancet*, vol. 381, (no. 9875), pp. 1405-1416, 2013.

[3] C.L. Fischer Walker, D. Sack, and R.E. Black, “Etiology of Diarrhea in Older Children, Adolescents and Adults: A Systematic Review,” *PLoS Negl Trop Dis*, vol. 4, (no. 8), pp. e768, 2010.

[4] M.K. Munos, C.L.F. Walker, and R.E. Black, "The effect of oral rehydration solution and recommended home fluids on diarrhoea mortality," *International Journal of Epidemiology*, vol. 39, (no. suppl 1), pp. i75-i87, 2010.

[5] L.F. Dawson, E. Valiente, E.H. Donahue, G. Birchenough, and B.W. Wren, "Hypervirulent *Clostridium difficile* PCR-Ribotypes Exhibit Resistance to Widely Used Disinfectants," *PLoS ONE*, vol. 6, (no. 10), pp. e25754, 2011.

[6] S. Ghosh, G.P. Pazhani, G. Chowdhury, S. Guin, S. Dutta, K. Rajendran, M.K. Bhattacharya, Y. Takeda, S.K. Niyogi, G.B. Nair, and T. Ramamurthy, "Genetic characteristics and changing antimicrobial resistance among Shigella spp. isolated from hospitalized diarrhoeal patients in Kolkata, India," *Journal of Medical Microbiology*, vol. 60, (no. 10), pp. 1460-1466, 2011.

[7] R.D. Finn, A. Bateman, J. Clements, P. Coggill, R.Y. Eberhardt, S.R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E.L.L. Sonnhammer, J. Tate, and M. Punta, "Pfam: the protein families database," *Nucleic Acids Research*, vol. 42, (no. D1), pp. D222-D230, January 1, 2014.

[8] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne, "The Protein Data Bank," *Nucleic Acids Research*, vol. 28, (no. 1), pp. 235-242, 2000.

[9] D. Tay, K. Govindarajan, A. Khan, T. Ong, H. Samad, W. Soh, M. Tong, F. Zhang, and T. Tan, "T3SEdb: data warehousing of virulence effectors secreted by the bacterial Type III Secretion System," *BMC Bioinformatics*, vol. 11, (no. 7), pp. S4, 2010.

[10] T. Kline, H. B. Felise, S. Sanowar, and S. I. Miller, "The Type III Secretion System as a Source of Novel Antibacterial Drug Targets," *Current Drug Targets*, vol. 13, (no. 3), pp. 338-351, 2012.

[11] T. Kline, K.C. Barry, S.R. Jackson, H.B. Felise, H.V. Nguyen, and S.I. Miller, "Tethered thiazolidinone dimers as inhibitors of the bacterial type III secretion system," *Bioorganic & Medicinal Chemistry Letters*, vol. 19, (no. 5), pp. 1340-1343, 2009.

[12] Z. Xiang, T. Todd, K.P. Ku, B.L. Kovacic, C.B. Larson, F. Chen, A.P. Hodges, Y. Tian, E.A. Olenzek, B. Zhao, L.A. Colby, H.G. Rush, J.R. Gilsdorf, G.W. Jourdian, and Y. He, "VIOLIN: vaccine investigation and online information network," *Nucleic Acids Research*, vol. 36, (no. suppl 1), pp. D923-D928, 2008.

[13] S.F. Altschul, T.L. Madden, A.A. SchÃ¤ffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Research*, vol. 25, (no. 17), pp. 3389-3402, 1997.

[14] D.S.M.E. Accelrys Software Inc., Release 4.0, San Diego: Accelrys Software Inc., 2013.

CHAPTER –3

Developing machine learning tool for the prediction of Multidrug And Toxin Extrusion (MATE) proteins based on Artificial Neural Network (ANN) and Support Vector Machine (SVM)

ABSTRACT

The growth and spread of drug resistance in bacteria have been well established in both mankind and beasts and thus is a serious public health concern. Due to the increasing problem of drug resistance, control of infectious diseases like diarrhea, pneumonia etc. is becoming more difficult. Hence, it is crucial to understand the underlying mechanism of drug resistance mechanism and devising novel solution to address this problem. Multidrug And Toxin Extrusion (MATE) proteins, first characterized as bacterial drug transporters, are present in almost all species. It plays a very important function in the secretion of cationic drugs across the cell membrane. In this work, we propose SVM based method for prediction of MATE proteins. The data set employed for training consists of 189 non-redundant protein sequences, that are further classified as positive (63 sequences) set comprising of sequences from MATE family, and negative (126 sequences) set having protein sequences from other transporters families proteins and random protein sequences taken from NCBI while in the test set, there are 120 protein sequences in all (8 in positive and 112 in negative set). The model was derived using Position Specific Scoring Matrix (PSSM) composition and achieved an overall accuracy 92.06%. The five-fold cross validation was used to optimize SVM parameter and select the best model. The prediction algorithm presented here is implemented as a freely available web server MATEpred, which will assist in rapid identification of MATE proteins.

3.1 INTRODUCTION

Multidrug efflux is an important mechanism of biocide and antimicrobial agent resistance in bacteria. They have been divided into various groups, which include the Major Facilitator Super (MFS) family, the Small Multidrug Resistance (SMR) family, the Resistance Nodulation and Cell Division (RND) family, the ATP Binding Cassette (ABC) family, and the Multidrug And Toxin Extrusion (MATE) family [1]. Multidrug and Toxin Extrusion (MATE) proteins form a class of proteins that acts as drug and proton antiporters. MATE family members are organic cation exporters that excrete metabolic or xenobiotic organic cations from the body [2]. Multidrug And Toxin Extrusion proteins are mediating the excretion of several antimicrobial drugs as well as other organic compounds into bile and urine, thereby contributing to drug disposition [3]. MATE family transporters are conserved in the three pinion domains of life (Archaea, Bacteria and Eukarya), and export xenobiotics using an electrochemical exchange of

H+ or Na+ across the tissue layer. Transporter proteins from the MATE family are vital in metabolite transport in plants, directly affecting crop yields worldwide. MATE transporters also mediate Multi Drug Resistance (MDR) in bacteria and mammals, modulating the efficacy of many pharmaceutical drugs used in the treatment of a variety of diseases [4]. The first MATE transporter NorM from *V. parahaemolyticus* and its homologue YdhE from *Escherichia coli* were identified in 1998 [5]. The X-ray structure of the MATE transporter NorM revealed a unique topology of the predicted 12 transmembrane helices which is a distinctive feature from any other known Multi Drug Resistance (MDR) transporter [4]. As reported MATE proteins play major role in conferring resistance to multidrug in several pathogenic bacteria, it is therefore important to enhance our understanding of the role of MATEs in drug extrusion and to identify these proteins at a faster pace. Owing to the time limit and cost of experiments, there is a demand to have computational methods to rapidly examine and interpret relevant data

3.2 METHODOLOGY

3.2.1 Datasets Generated for Training

MATE proteins (assigned as positive set) and all other types of proteins (assigned as negative set) were collected through a broad and critical study of research articles from PubMed. Using CD-HIT (http://weizhong-lab.ucsd.edu/cd-hit/) [6] program the redundancy in both the sets was scaled down to 40%. So we had two datasets positive and negative, each comprising of 63 and 126 sequences, respectively.

3.2.2 Benchmark Datasets for Testing

For checking the efficiency of the SVM model generated, its performance was tested on independent datasets consisting of 8 positive sequences and 112 negative sequences, obtained after scaling down its redundancy to 40% against NR database.

3.2.3 ANN and SNNS

The Artificial Neural Network (ANN) consists of nodes or neurons that receive signals through interconnecting arcs [7]. Signals are passed between neurons through connection links which carry an associated weight [8] as shown in (Figure 3.1). These neurons are organized in input, hidden and output layers. Each neuron applies an activation function to its net input to determine it output signal.

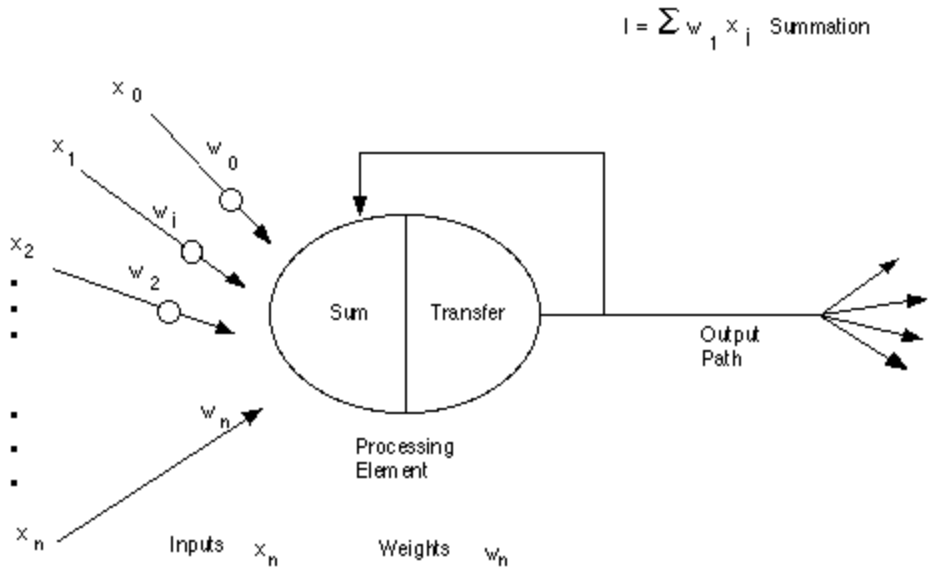


Figure 3.1 Basic Artificial Neural Network[9].

ANN was implemented using Stuttgart Neural Network Simulator (http://www.ra.cs.uni-tuebingen.de/SNNS/), SNNS version 4. The feed-forward back propagation type of neural networks was trained on different protein features. The number of hidden nodes, weights, number of cycles and other learning parameters were optimized for each network. The output unit consisted of target value 1 or -1, referring to positives and negatives respectively. The final number of cycles was determined where the Sum of Squared Error function (SSE) was the least.

3.2.4 SVM Algorithm

Support Vector Machine (SVM) is a supervised machine learning method first introduced by Vapnik in 1995 [10]. Support Vector Machines are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships.

Figure 3.2 shows the basic idea behind Support Vector Machines. Here the original objects (left side of the schematic) mapped, i.e., rearranged, using a set of mathematical functions, known as kernels. The process of rearranging the objects is known as mapping. The mapped objects (right side of the schematic) is linearly separable and, thus, instead of constructing the complex curve

(left schematic), an optimal line is to be found that can separate the GREEN and the RED objects.

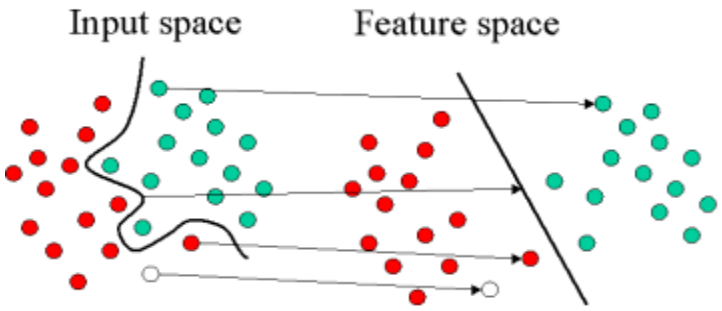


Figure 3.2 Schematic respresentation of Support vector machine

SVM in combination with kernel functions is used to map input data to some vector space. In order to avoid over fitting, SVM then finds a hyperplane separating the positive data from the negative ones in high dimensional *space*[11].

SVM in this approach was implemented using LibSVM package (http://www.csie.ntu.edu.tw/~cjlin/libsvm/) [12] which allows us to optimize a number of parameters [13] and to use kernels (e.g. linear, polynomial, radial basis function, sigmoid) for obtaining the best hyperplane [14]. LIBSVM supports various SVM formulations for classification, regression, and distribution estimation [15]. In this study Radial Basis Function (RBF) kernel was used. This kernel nonlinearly maps sample into high dimensional space so it, unlike the linear kernel can handle the case when the relation between the class labels and attributes is non linear [12].

3.2.5 Five-Fold Cross Validation

For evaluating the performance of modules generated in this study, we used five-fold cross validation in which the data is first partitioned into 5 equal sized datasets. Later, five iterations of training and validation are done such that within each iteration, a different fold of the data is held-out for validation while the remaining four folds are used for learning [16]. Several performance measures were then applied to evaluate the best parameters (γ and C) and then averaged to bring forth an overall assessment of the model [8].

3.2.6 Performance Measures

Applying the following equations accuracy, sensitivity, specificity and Matthew Correlation Coefficient (MCC) were calculated for evaluating the performance of SVM classifiers:

1) **Sensitivity**: It is determined as the percentage of MATE that is correctly predicted as MATE.

$$Sensitivity = \frac{TP}{TP+FN} \times 100$$

2) **Specificity**: It is the percentage of non-MATE that is correctly predicted as non-MATE.

$$Specificity = \frac{TN}{TN+FP} \times 100$$

3) **Accuracy**: It is the percentage of correct predictions out of the total number of predictions.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \times 100$$

4) **Matthews correlation coefficient** (MCC): It is a measure of both sensitivity and specificity. MCC = 0 is the indication of completely random prediction, while MCC = 1 indicates perfect prediction.

$$MCC = \frac{(TP \times TN)-(FN \times FP)}{\sqrt{(TP+FN)x(TN+FP)x(TP+FP)x(TN+FN)}}$$

5) **F-score:** It is the harmonic mean of precision and recall. The best value for F-score is 1 and worst score is 0.

$$F_1 = \frac{2xTP}{2xTP+FP+FN}$$

3.2.7 Feature Selection

3.2.7.1 Composition based SVM classifiers

a) **Amino Acid Composition (AAC**): It is the fraction of each of the 20 amino acids present in a protein sequence and generates an input vector of 20 dimensions.

b) **Dipeptide Composition (DPC)***:* It is the fraction of a dipeptide divided by the total number of possible dipeptides and gives information in the form of 400 dimensions (20*20).

c) **Charge Composition (CC)***:* It is the fraction of charged amino acids divided by the total length of the protein. The fractions of positively and negatively charged amino acids yields a fixed length input vector of 20 dimensions.

d) **Hydrophobicity Composition (HC)***:* Based on their hydrophobicity properties, the amino acids may be classified into five groups [17]. Moments of the positions of the five groups were calculated using the formula as below with r varying from 2 to 5. This yields a fixed length input vector of 25 dimensions.

$$Mr = \sum \frac{(Xi - Xm)^r}{N}$$

Where Xm = mean of all positions of hydrophobic amino acids, Xm = $\sum_{i=1}^{N}$Xi/N;

Xi = position of ith hydrophobic amino acid and N = total number of hydrophobic amino acids in the sequence [14].

e) **Multiplet Composition (MPC)***:* Multiplets are homopolymers (Y) n and yield an input vector of 20-dimensions.

Where, Y is any amino acid repeated n times with n≥2.

f) **Position Specific Scoring Matrix (PSSM) profile**

A Position Specific Scoring Matrix (PSSM) is a table that contains probability information of amino acids or nucleotides at each spot of an ungapped multiple sequence alignment. In such a table, the rows represent residue positions of a particular multiple alignment and the columns represent the names of residues or vice versa. The values in the table represent log odds scores of the residues calculated from the multiple alignments. PSSM consists of a set of 20 substitution scores at each position along the motif—one for each of the amino acids, thus generating an input vector of 400 dimensions. In this case, PSI-BLAST iterative search was performed against the non-redundant NCBI database, with a cut-off E-value of 0.001. In each of the 3 iterations, a profile or PSSM (Position Specific Scoring Matrix) is generated from a multiple alignment of the high scoring hits by calculating position specific scores for each position in alignments. After three iterations, PSI-BLAST generates a PSSM having the highest score. Sigmoid function ($f(x) =1/1+e^{-x}$) was used to normalize each element of the PSSM matrix whereby each element f(x) was scaled to a range of 0-1.

To make a SVM input of fixed length, we summed up all the rows in the PSSM corresponding to the same amino acid in the sequence, followed by division of each element by the length of the sequence. The steps followed to generate PSSM matrix with 400 dimensions is shown in Figure 3.3 [18]. Position-Specific Iterated BLAST (PSI-BLAST) provides an automated facility for constructing, refining, and searching PSSMs.

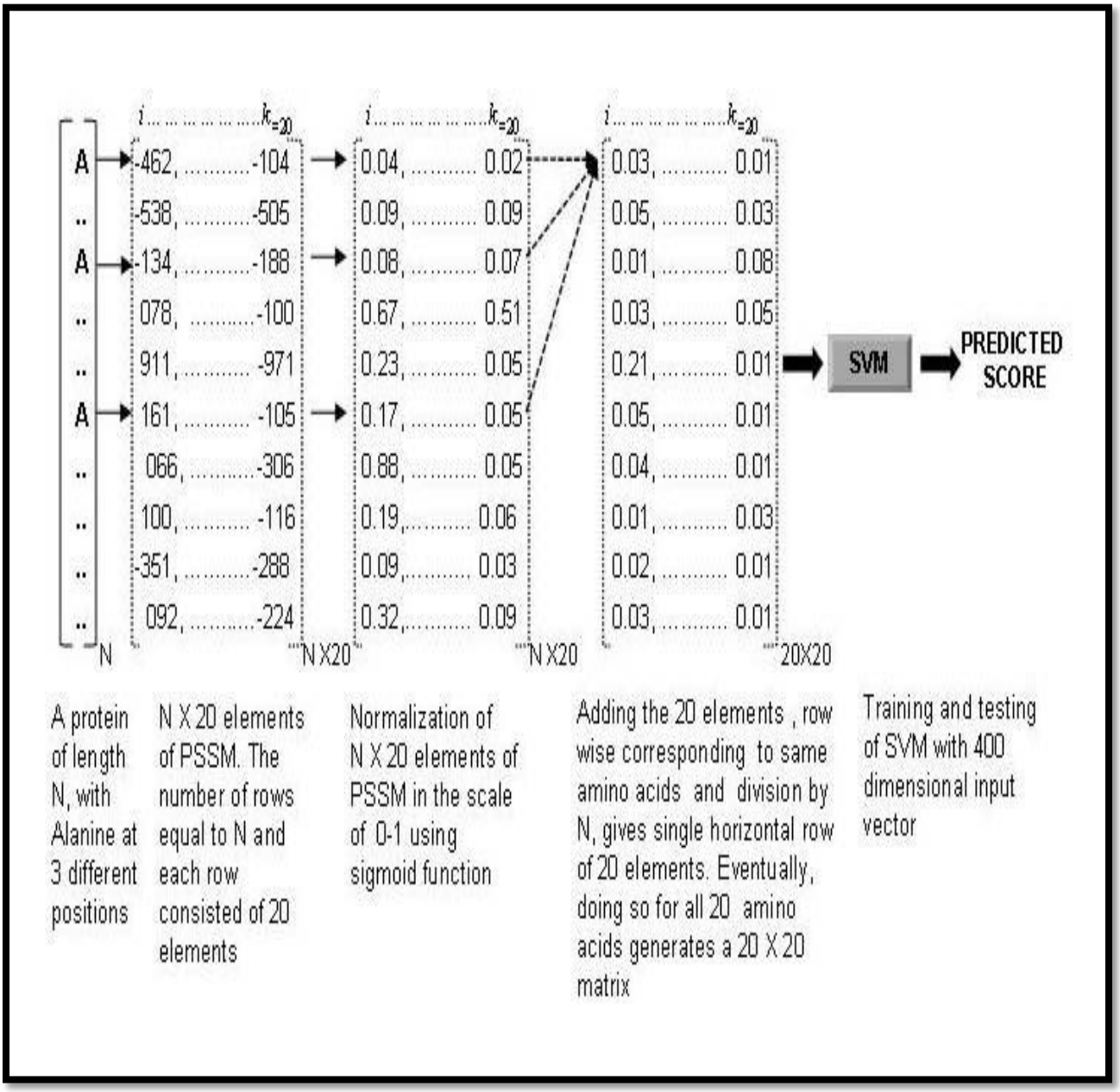
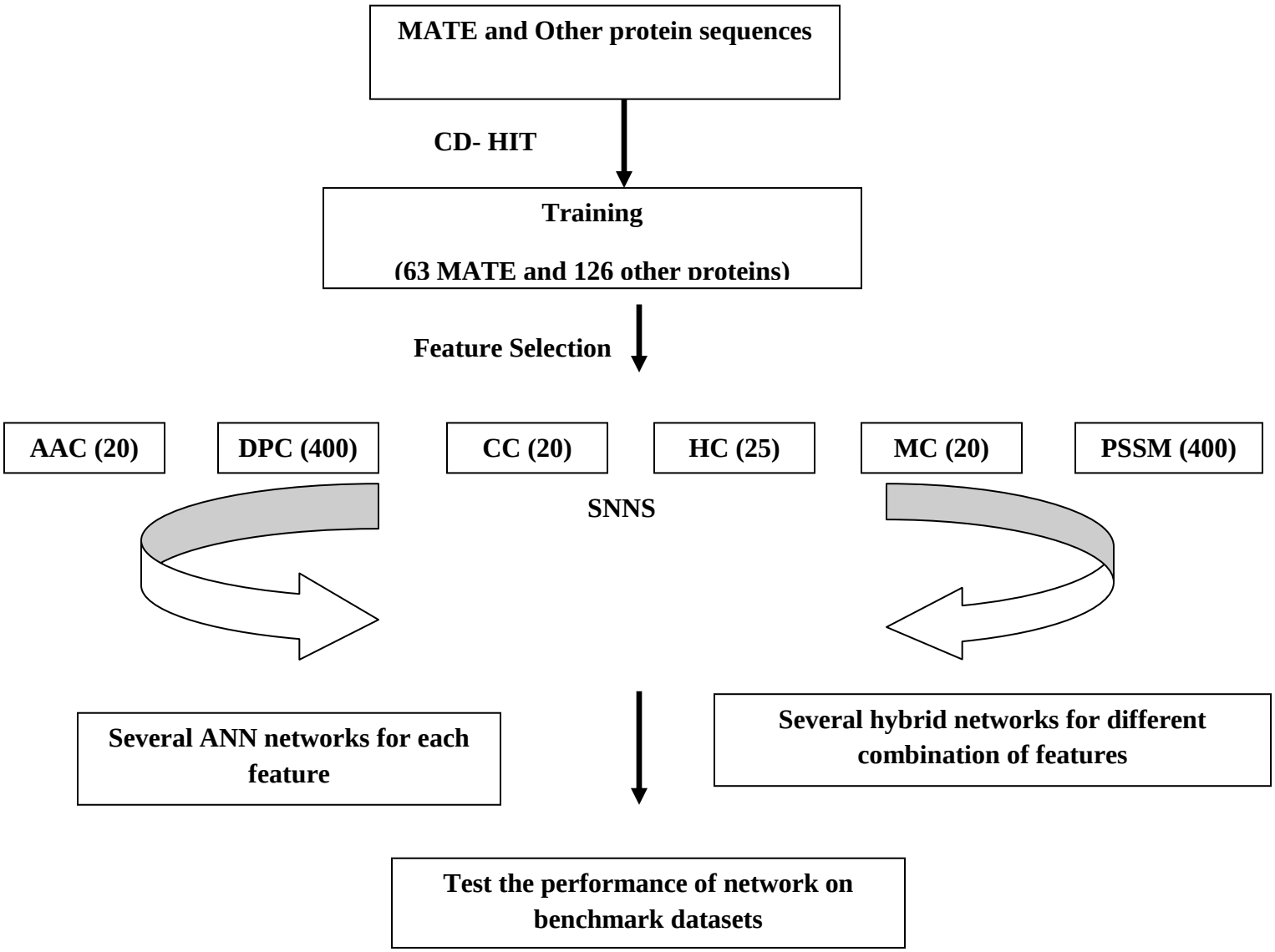


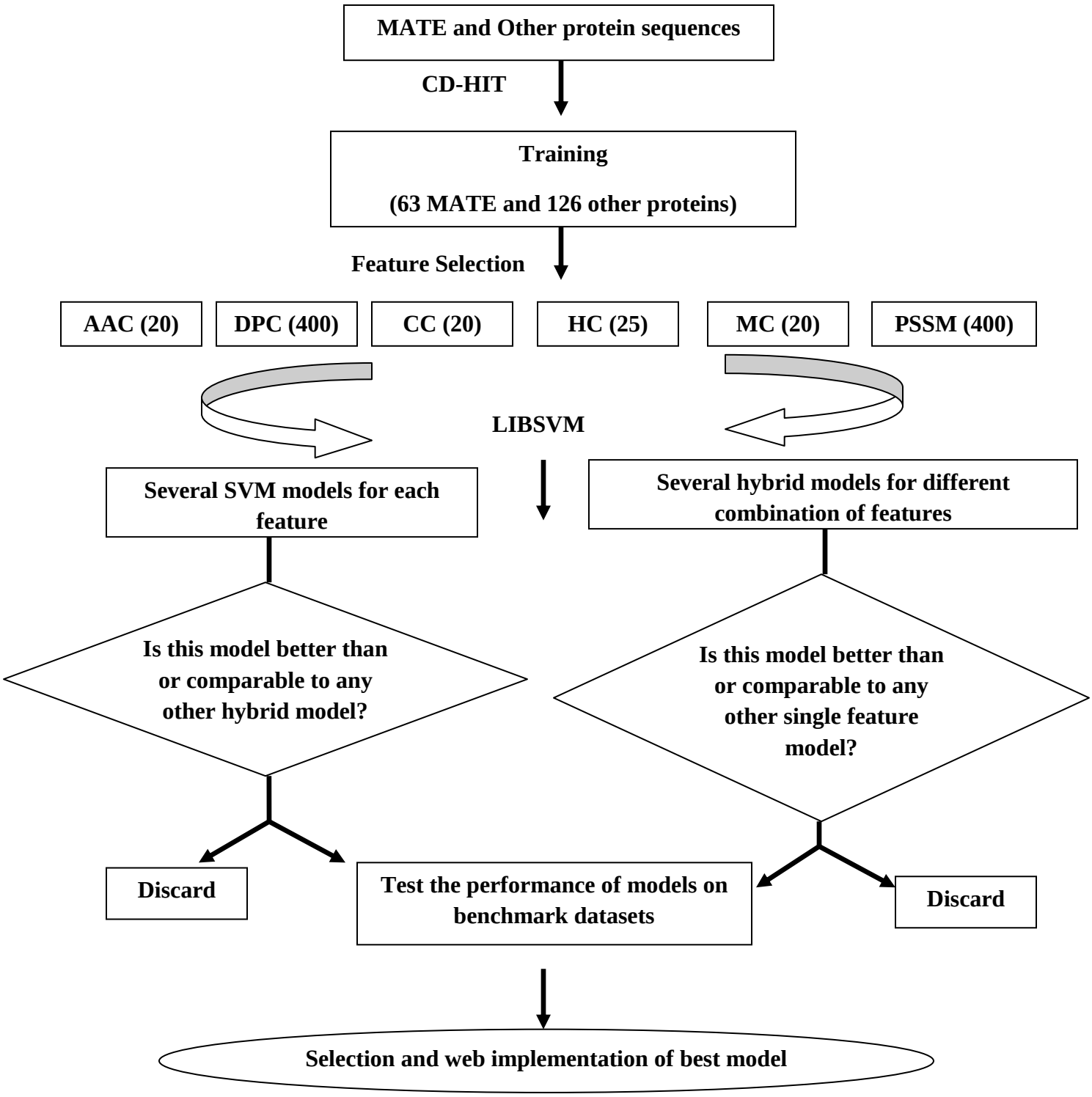
Figure 3.3 The steps used to convert PSSM profiles generated by PSI-BLAST into a training vector of 400 dimensions [18].

3.2.8 Flowcharts of the Experimental Procedure

a) Artificial Neural Network (ANN) based Approach



b) Support Vector Machine (SVM) based Approach



3.2.9 ROC Plot

LibSVM package was used to obtain the Receiver Operating Characteristic (ROC) plot for the SVM classifier developed in the study.

3.3 RESULTS AND DISCUSSION

3.3.1 Performance of ANN Based Networks

Different compositional features i.e .Amino Acid Composition (AAC), Dipeptide Composition (DPC), Charge Composition (CC), Hydrophobicity Composition (HC), Multiplet Composition (MPC) and Position-Specific Scoring Matrix (PSSM) were extracted from the positive and negative dataset sequences. Different ANNs were generated for the different types of features, while optimizing the learning parameters including activation function, number of hidden neurons, learning rate etc. The approach was to keep the number of hidden neurons and the number of training cycles as low as possible while simultaneously achieving good accuracies in threefold cross-validation. The training was carried out for different cycles and learning terminated when SSE (Sum of Squared Errors) was minimum. Random weights were used for initializing the network and Standard Backpropagation algorithm was used to minimize the differences between the computed output and the target value. The best ANN for each feature are described below and tabulated in Table 3.1.

Table 3.1 Performance of ANN classifiers in threefold CV

Network	Threshold	Accuracy (%)	Specificity (%)	Sensitivity (%)	MCC
AAC	0.4	84.45	99.20	64.90476	0.705
DPC	0.4	39.80	76.56	12.98	-0.051
CC	0.9	54.08	56.34	38.09	-0.053
HC	0.1	59.5	61.9	42.83	0.045
MC	0.9	76.16	32.5	84.12	0.177
PSSM	0.3	78.63	67.46	66.67	0.324

3.3.2 Performance of Alignment Based Techniques

In total 100 MATE sequences were collected through extensive survey of research articles from PubMed. Pfam and BLAST analysis were then performed on these protein sequences. The Pfam database contains one Pfam domain ‘MatE’ having Pfam ID PF01554.13 and it was found that

only 27 proteins out of the 100 MATE protein sequences showed presence of this Pfam domain. Position Specific Iterated (PSI) BLAST was also performed on a positive dataset comprising of 63 MATE protein sequence in a Leave-One-Out-Cross Validation (LOO CV) manner where once each sequence was used as the query sequence while the rest were used as the reference database at a threshold of 0.001. This process was repeated over each sequence present in a positive dataset. It was found that 21 sequences did not find any significant hit. As none of these similarity based search methods were sufficient to identify all the MATE proteins, therefore we explored SVM approach on various protein features for identification of MATE proteins.

3.3.3 Performance of Composition based SVM classifiers

Fivefold cross validation of Amino Acid Composition (AAC), Dipeptide Composition (DPC), Charge Composition (CC), Hydrophobicity Composition (HC), Multiplet Composition (MPC) and Position-Specific Scoring Matrix (PSSM) was performed and all were trained using the Radial Basis Function (RBF) kernel. The kernel function was then optimized to obtain the best C and γ corresponding to the highest values of sensitivity, specificity and accuracy.

It was found that the Charge Composition (CC) model has an accuracy of 78.84%, 85.71% with Dipeptide Composition (DPC) and 65.08% with Multiplet Composition (MPC) based model. The Amino Acid Composition (AAC) model was found to exhibit over-fitting as it performed remarkably well (accuracy = 90.47%) in cross validation, but failed to perform well in testing set (accuracy = 50%). This suggests that amino acid composition used as an independent property is not enough to discriminate between MATE and non-MATE proteins.

3.3.4 Performance of Hybrid SVM Models

To enhance the prediction accuracy, we further developed several hybrid models with the combination of features. We obtained an accuracy of 72.75% with CH (hybrid of Charge and Hydrophobicity) based, 82.53% with DCP (hybrid of Dipeptide, Charge and PSSM) based and 74.60% with ACP (hybrid of Amino acid, Charge and PSSM) based model. But all of these hybrid models failed to perform well on independent test sets.

3.3.5 Performance of PSSM Profile Based SVM Classifier

PSSM profiles generated using PSI-BLAST provides valuable information about conserved residues present within the protein sequence. PSSM profiles for the training set sequences were

generated by performing the PSI-BLAST search against NR database. We employed PSSM profiles as a feature for training SVM. It was scaled between 0-1 and normalized using logistic function. This model was best among all the models and yielded an accuracy of 92.06%, with the sensitivity and specificity of 100% and 89.42% respectively along with an MCC of 0.82 and F-score of 0.83632 (Table 3.2) in 5-fold cross validation.

Table 3.2 Performance of different SVM classifiers in Five-Fold CV (Where SN- Sensitivity, SP- Specificity and MCC- Matthews correlation coefficient).

Model	**C**	**γ**	**SN (%)**	**SP (%)**	**Accuracy (%)**	**MCC**	**F-score**
AAC	5	0.06	73.02	99.92	90.47	0.78765	0.83632
DPC	4	0.01	68.25	94.44	85.71	0.67006	0.76111
CC	30	0.1	48.43	94.4	78.84	0.50581	0.60784
MPC	20	0.25	47.62	73.81	65.08	0.21428	0.25806
CH	25	0.9	27.34	96	72.75	0.34112	0.40462
ACP	10	5	76.8	89.09	74.60	0.65548	0.76042
DCP	2	6	73.68	86.36	82.53	0.59204	0.717794
PSSM	13	0.01	100	89.42	92.06	0.82436	0.86301

3.3.6 Performance on benchmarking datasets

Table 3.3 represents prediction results of the SVM model on a benchmark dataset yielding an overall accuracy of 72.22%.

Table 3.3 Performance on Benchmark DatasetsCV (Where SN- Sensitivity, SP- Specificity and MCC- Matthews correlation coefficient).

Model	**SN (%)**	**SP (%)**	**Accuracy (%)**	**MCC**	**F-score**
PSSM	100	89.4	92.5	0.6772	0.6677

3.3.7 Receiver Operating Characteristic (ROC) Plot

To evaluate the performance of the best model the ROC curve was used which shows the trade-off between true positive rate (sensitivity) and false positive rate (specificity) over their entire

range of possible values. The PSSM classifier had Area Under Curve (AUC) of 0.865 (Figure 3.3). This analysis confirmed the efficacy of the model.

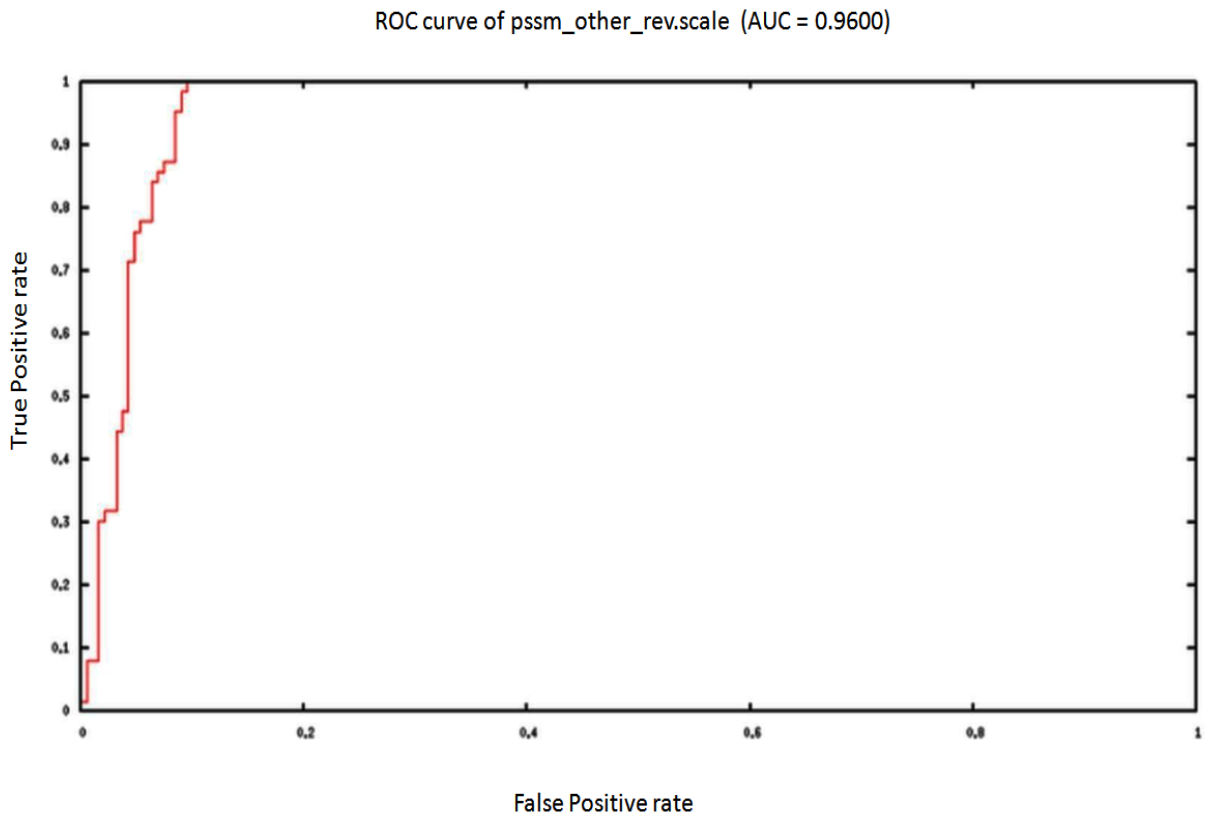


Figure 3.3 ROC curve of PSSM classifiers: ROC plot depicts relative trade-offs between true positive and false positives.

3.3.8 Web Implementation

The SVM classifier presented in this study is implemented as a freely available web tool 'MATEpred' to predict MATE proteins. The tool is openly accessible at http://www.bioinformatics.org/matepred_hos (Figure 3.4) and connected to the HTML front-end through PHP using the Apache web server.

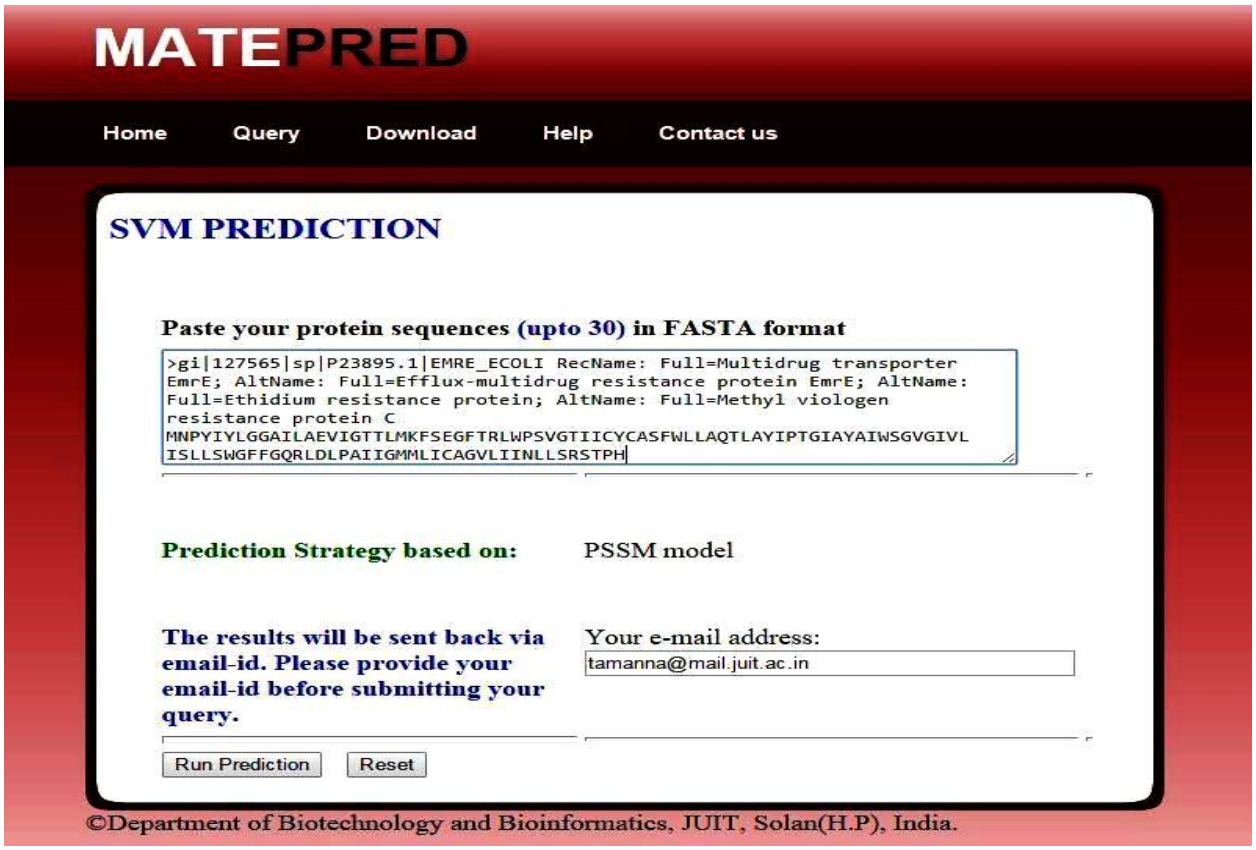


Figure 3.4 Snapshot of the prediction tool Matepred.

The prediction is made using PSSM classifier. The server accepts protein sequence in FASTA format as an input. The output is sent back to the user through e-mail which gives sequence number, predicted score and decision of the model as shown in Figure 3.5.

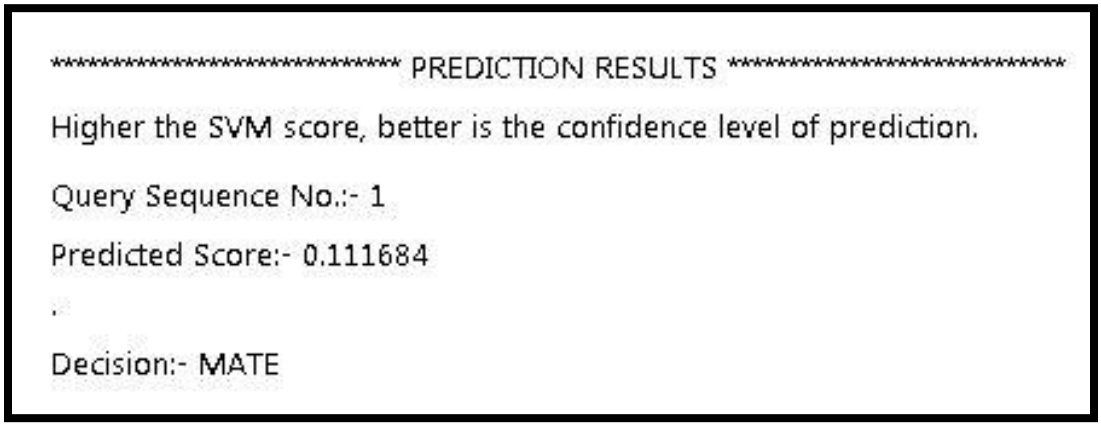


Figure 3.5 Results from MATEPred.

3.3.9 Application of MATEpred

Vibrio parahaemolyticus and S*higella* are two of major contributors in the epidemiology of diarrhea. We used the PSSM model to scan the proteomes of these organisms for the presence of MATE proteins. Initially, the model reported eight and sixteen positives for each of these species, respectively. Out of these predicted MATEs, seven proteins from *Vibrio parahaemolyticus* (Accession no. O82855, Q87FN2, Q87IY5, Q87HE9, Q87MO9, Q87FV4, and Q87QD3) (Table 3.4) and one protein from *Shigella* (Accession no Q323U7) (Table 3.5) are having 12 transmembrane helices predicted using HMMTOP [19] server. These were assigned as 'MATE-like' proteins that could be investigated for their role in drug resistance [14]. We also analyzed for the presence of Pfam and PROSITE which suggested that out of seven proteins, four proteins from *Vibrio* (Q87FN2, Q87HE9, Q87FV4, and Q87QD3) (Table 3.6 and 3.8) and one from *Shigella* (Q323U7) [Table 3.7 and 3.9] are the members of MFS family, and two more proteins from *Vibrio* (O82855 and Q87MO9) (Table 3.6) belongs to matE family, hence these represent potential MATE proteins. Out of the total predicted positives, there were some false positives also. We also performed BLAST [20] analysis for the predicted proteins against positive dataset employed for training and it was found that five proteins from *Vibrio* (O82855, Q87IY5, Q87HE9, Q87MO9 and Q87FV4) and one from *Shigella* (Q323U7) showed significant sequence similarity with the known MATE proteins. From all these analyses, it was observed that out of the total predicted proteins, two proteins from *Vibrio* with Accession number (O82855, Q87MO9) are MATE family proteins while two others with Accession numbers (Q87HE9 and Q87FV4) from *Vibrio* and one from *Shigella* (Q323U7) are the potential MATE candidates that can be further taken for experimental verification to study their role in drug resistance.

Table 3.4 Transmembrane regions of predicted proteins from *Vibrio parahaemolyticus.*

Sr. No	**Accession No**	**Protein Name**	**Transmembrane Helices**	**Transmembrane Region**
1	O82855	Multidrug resistance protein NorM	12	17-36, 49-70, 91-107, 120-143, 160-179, 194-213, 244-262, 275-296, 317-334, 349-369, 390-406, 419-437
2	Q87FN2	Multidrug resistance protein D	12	9-28, 45-62, 75-92, 99-118, 131-155, 164-182, 213-237, 250-267, 280-297, 302-321, 334-357, 364-383
3	Q87IY5	Multidrug efflux membrane fusion protein	12	12-31, 342-361, 368-387, 396-415, 446-465, 474-498, 535-554, 866-887, 894-917, 926-943, 974-993, 1002-1026
4	Q87HE9	Multidrug resistance protein E	12	7-24, 45-62, 71-88, 101-118, 131-148, 161-178, 209-227, 240-257, 270-287, 300-318, 331-348, 353-370
5	Q87MO9	Multidrug resistance protein	12	12-34, 43-65, 86-109, 130-149, 158-179, 188-210, 231-250, 263-286, 311-330, 349-373, 382-400, 409-428

6	Q87FV4	Multidrug resistance protein MdtL	12	6-25, 40-59, 70-88, 97-116, 129-152, 159-178, 211-235, 244-263, 270-289, 298-317, 330-354, 361-380
7	Q87QD3	Multidrug resistance protein	12	25-44, 59-82, 95-113, 124-141, 154-171, 184-203, 234-258, 269-288, 303-320, 327-350, 363-382, 391-409

Table 3.5 Transmembrane regions of predicted proteins from *Shigella boydii*

Sr. No	Accession No	Protein Name	Transmembrane Helices	Transmembrane Region
1	Q323U7	Multidrug transporter MdfA	12	17-36, 53-72, 85-102, 113-130, 143-162, 171-188, 221-240, 257-275, 288-307, 316-335, 348-365, 382-399

Table 3.6 Pfam results for *Vibrio parahaemolyticus*

Sr. No	Accession No.	Protein Name	Family	Description
1	O82855	Multidrug resistance protein NorM	MatE	MatE
2	Q87FN2	Multidrug resistance protein D	MFS_1	Major Facilitator Superfamily
3	Q87IY5	Multidrug efflux membrane fusion	ACR_tran	AcrB/AcrD/AcrF family

		protein		
4	Q87HE9	Multidrug resistance protein E	MFS_1	Major Facilitator Superfamily
5	Q87MO9	Multidrug resistance protein	MatE	MatE
6	Q87FV4	Multidrug resistance protein MdtL	MFS_1	Major Facilitator Superfamily
7	Q87QD3	Multidrug resistance protein	MFS_1	Major Facilitator Superfamily

Table 3.7Pfam results for *Shigella boydii*

Sr. No	**Accession No.**	**Protein Name**	**Family**	**Description**
1	Q323U7	Multidrug transporter MdfA	MFS_1	Major Facilitator Superfamily

Table 3.8 PROSITE Results for *Vibrio parahaemolyticus*

Sr. No	**Accession No.**	**Protein Name**	**PROSITE ID**	**FAMILY**
1	O82855	Multidrug resistance protein NorM	PS50156	MATE
2	Q87FN2	Multidrug efflux membrane fusion protein	PS50850	MFS
3	Q87HE9	Multidrug resistance protein	PS50850	MFS
4	Q87FV4	Multidrug resistance protein MdtL	PS50850	MFS
5	Q87QD3	Multidrug resistance protein	PS50850	MFS

Table 3.9 PROSITE Results for *Shigella boydii*

Sr. No	**Accession No.**	**Protein Name**	**PROSITE ID**	**Family**
1	Q323U7	Multidrug transporter MdfA	PS50850	MFS

3.4 CONCLUSION

MATEpred efficiently distinguishes MATE sequences from non-MATE sequences on the basis of PSSM profile. In future, MATEpred will continue to be updated with the inclusion of additional MATE sequences, which will further enhance the efficiency of MATEpred. This will make it more worthwhile for the general users and also for the researchers working in this field.

REFERENCES

[1] M. Otsuka, T. Matsumoto, R. Morimoto, S. Arioka, H. Omote, and Y. Moriyama, "A human transporter protein that mediates the final excretion step for toxic organic cations," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, (no. 50), pp. 17923-17928, 2005.

[2] H. Omote, M. Hiasa, T. Matsumoto, M. Otsuka, and Y. Moriyama, "The MATE proteins as fundamental transporters of metabolic and xenobiotic organic cations," *Trends in Pharmacological Sciences*, vol. 27, (no. 11), pp. 587-593, 2006.

[3] A.T. Nies, K. Damme, E. Schaeffeler, and M. Schwab, "Multidrug and toxin extrusion proteins as transporters of antimicrobial drugs," *Expert Opinion on Drug Metabolism & Toxicology*, vol. 8, (no. 12), pp. 1565-1577, 2012.

[4] X. He, P. Szewczyk, A. Karyakin, M. Evin, W.-X. Hong, Q. Zhang, and G. Chang, "Structure of a cation-bound multidrug and toxic compound extrusion transporter," *Nature*, vol. 467, (no. 7318), pp. 991-994, 2010.

[5] Y. Morita, K. Kodama, S. Shiota, T. Mine, A. Kataoka, T. Mizushima, and T. Tsuchiya, "NorM, a Putative Multidrug Efflux Protein, of Vibrio parahaemolyticus and Its Homolog in Escherichia coli," *Antimicrobial Agents and Chemotherapy*, vol. 42, (no. 7), pp. 1778-1782, 1998.

[6] W. Li and A. Godzik, "Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences," *Bioinformatics*, vol. 22, (no. 13), pp. 1658-1659, 2006.

[7] John Hertz, Anders Krogh, and R.G. Palmer, "Introduction to the theory of neural computation," *Addison-Wesley Longman Publishing Co., Inc.*, pp. 327, 1991.

[8] J. Ramana and D. Gupta, "Machine Learning Methods for Prediction of CDK-Inhibitors," *PLoS ONE*, vol. 5, (no. 10), pp. e13357, 2010.

[9] http://www.psych.utoronto.ca/users/reingold/courses/ai/cache/neural2.html.

[10] V.N. Vapnik, *Statistical learning theory* New York: Wiley-Interscience, 1998.

[11] A. Ben-Hur, C.S. Ong, S. Sonnenburg, B. Schölkopf, and G. Rätsch, "Support Vector Machines and Kernels for Computational Biology," *PLoS Comput Biol*, vol. 4, (no. 10), pp. e1000173, 2008.

[12] C. Chang and C. Lin, *{LIBSVM}: a library for support vector machines*, 2001.

[13] J. Ramana and D. Gupta, "LipocalinPred: a SVM-based method for prediction of lipocalins," *BMC Bioinformatics*, vol. 10, (no. 1), pp. 445, 2009.

[14] J. Ramana and D. Gupta, "FaaPred: A SVM-Based Prediction Method for Fungal Adhesins and Adhesin-Like Proteins," *PLoS ONE*, vol. 5, (no. 3), pp. e9695, 2010.

[15] C. Chih-Chung and L. Chih-Jen, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, (no. 3), pp. 1-27, 2011.

[16] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross Validation," in Encyclopedia of Database Systems, T. Özsu and L. Liu eds.: Springer, 2009.

[17] V. Brendel, P. Bucher, I.R. Nourbakhsh, B.E. Blaisdell, and S. Karlin, "Methods and algorithms for statistical analysis of protein sequences," *Proceedings of the National Academy of Sciences*, vol. 89, (no. 6), pp. 2002-2006, 1992.

[18] A. Garg and D. Gupta, "VirulentPred: a SVM based prediction method for virulent proteins in bacterial pathogens," *BMC Bioinformatics*, vol. 9, (no. 1), pp. 62, 2008.

[19] G.b.E. TusnÃ¡dy and I.n. Simon, "The HMMTOP transmembrane topology prediction server," *Bioinformatics*, vol. 17, (no. 9), pp. 849-850, 2001.

[20] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, and D.J. Lipman, "Basic local alignment search tool," *Journal of Molecular Biology*, vol. 215, (no. 3), pp. 403-410, 1990.

CHAPTER–4

Structural Insights into the Fluoroquinolone Resistance Mechanism of *Shigella flexneri* DNA Gyrase and Topoisomerase IV

ABSTRACT

Traveler's Diarrhea (TD) is an important public health concern that can result from a variety of intestinal pathogens including bacteria, parasites and virus. Large numbers of antibiotics are being employed to cure traveler's diarrhea, but due to widespread use of these antibiotics the pathogens are becoming resistant to it. In this work, we performed docking studies of DNA gyraseA (GyrA) and topoisomerase IV (ParC) of *Shigella flexneri* and its mutants with two different fluoroquinolones, ciprofloxacin and norfloxacin to understand its resistance mechanism at structural level. *Shigella flexneri* strains with mutations at serine 83 to leucine and aspartic acid 87 to glutamate or asparagine of GyrA and that of serine 80 to isoleucine in ParC have decreased susceptibility to fluoroquinolones. This analysis has revealed weaker interaction of ciprofloxacin/norfloxacin with all the mutants as compared to the wild type. The study highlights the importance of aspartic acid and serine in GyrA and that of serine in ParC forming bonds with ciprofloxacin/ norfloxacin, which may play a crucial role in antibiotic resistance. The work presented here co-relates very well with the experimental outcomes and gives a good explanation for fluoroquinolone resistance in *Shigella flexneri*.

4.1 INTRODUCTION

Traveler's Diarrhea (TD) is an important public health concern. Various pathogens including *Escherichia coli* (*ETEC*), S*almonellaspp*. and *Campylobacter* have been identified as the pathological agents of traveler's diarrhea (TD), with S*higella* spp. being one of the most common etiological agents. Several antibiotics such as quinolones (ciprofloxacin, norfloxacin), rifaximin and azithromycin were reported to be effective and safe to use against travelers' diarrhea [1] . But it has been found that S*higella* spp. acquired resistance to these clinically important antibiotics [2]. In the last few years, a dramatic escalation has been seen in the antibiotic resistance profile of S*higella* spp. [3]. Increased antibiotic resistance is a great impediment in control of the traveler's diarrhea and thus results in greater disease burden globally. Fluoroquinolone, one of the most effective second-line drugs are antibacterial compounds used to treat various kinds of bacterial infections [4, 5]. Fluoroquinolones, ciprofloxacin and norfloxacin have a broad spectrum of antibacterial activity and are used for treatment of large number of infectious diseases. But due to widespread use of these antibiotics the pathogens are becoming resistant to it. These mainly target DNA gyrase and Topoisomerase

IV [6]. DNA gyrase plays an important role in the regulation of DNA topology especially DNA super coiling activity DNA gyrase helps in the survival of bacteria inside the host cells. Topological stress that arises from the translocation of transcription and replication complexes along DNA is relieved by DNA gyrase; wheras topoisomerase IV being a decatenating enzyme resolves interlinked daughter chromosomes following DNA replication [7, 8]. Resistance to quinolones in DNA gyrase occurs through mutations in the Quinolone Resistance-Determining Region (QRDR) [9]. DNA gyrase and topoisomerase IV acts as the target for fluoroquinolones [10]. Therefore, these are suitable candidate to study the effect of mutations on fluoroquinolone resistance. Emerging resistance to quinolones such as ciprofloxacin has been studied in several bacteria, such as in *Staphylococcus aureus*, *Streptococcus pneumoniae*, *Pseudomonas aeruginosa*, and *Mycobacterium tuberculosis* [11]. In last few years, large numbers of studies related to resistance mechanism have been reported, but structural level analysis revealing the mode of interaction of GyrA and ParC with fluoroquinolones yet needs to be explored.

A group of researchers working at the National Institute of Health Korea reported fluoroquinolone resistant S*higella flexneri* isolates from a patient who had travelled to India [9]. In this study, it was reported that in the susceptibility test of fluoroquinolone family antibiotics, *Shigella flexneri* isolates with some mutations showed resistance to ciprofloxacin, norfloxacin, ofloxacin and nalidixic acid whose minimal inhibitory concentrations (MICs) are 8 μg/mL, 32 μg/mL, 8 μg/mL and 256 μg/mL [5] respectively. These mutations were Ser83→Leu and Asp87→Asn in gyrase A and Ser80→Ileu in parC. A third mutation corresponding to Asp87→Gly in gyraseA was also reported in another *Shigella flexneri* isolate from a Korean patient [5] . Here, we studied the aforesaid mutations to investigate the resistance mechanism at structural level.

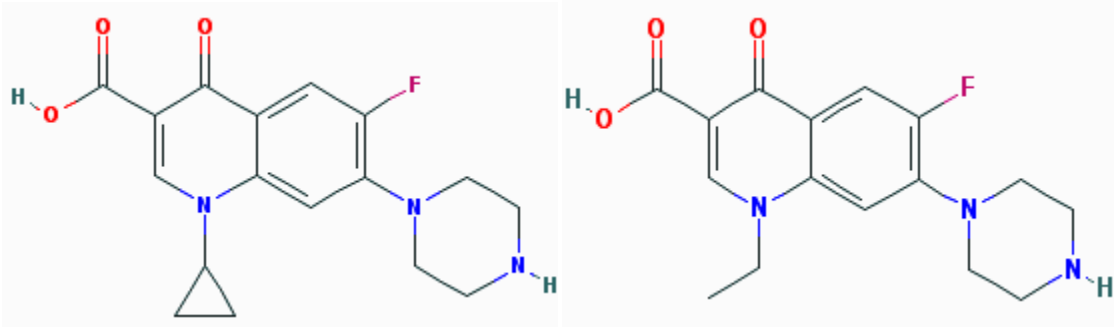
4.2 METHODOLOGY

The interaction study was carried out using LEADIT v2.1.6 package from BiosolveIT and Accelyrs Discovery Studio client 3.5 [12] was used for molecule preparation.

4.2.1 Ligand Preparation

The structures of the two ligand molecules ciprofloxacin (Figure 4.1 (A)) and norfloxacin (Figure 4.1 (B)) were taken from Pubchem (https://pubchem.ncbi.nlm.nih.gov/) having

identification numbers 2476 and 4359, respectively. The ligands were then prepared in Discovery Studio [12] and minimized by applying CHARMm force field and saved in MOL2 format for the further use in docking studies.



A. Ciprofloxacin **B.Norfloxacin**

Figure 4.1Chemical structures of (A) Ciprofloxacin (CID 2476) and (B) Norfloxacin (CID 4539)

4.2.2 Protein Preparation

4.2.2.1 Homology Modeling

The protein sequence of GyrA and ParC from the *Shigella flexneri* reference strains 2a (Accession number CEP59053.1) and 5a (Accession number EID62675.1), *Shigella sonnei* (Accession number CSP72916.1), *Shigella boydii* (Accession number WP_039060309.1) and *Shigella dysenteriae* (Accession number WP_001281279.1) showed 99% identity with the *Shigella flexneri* isolate used in this study.

The X-ray crystal structure of the target protein GyrA and ParC of *Shigella flexneri* was not available in the Protein Data Bank (PDB). The sequence of the protein GyrA (Accession number WP_001281258.1) and ParC (Accession number KFZ98372.1) from *Shigella flexneri* was retrieved from NCBI protein database (www.ncbi.nlm.nih.gov). The sequence homology of *Shigellaflexneri* DNA gyarse A and *Escherichia coli* (strain K12) determined using NCBI BLAST [13] against the PDB database was about 98%, and that of *Shigella flexneri* ParC and *Escherichia coli* was 99%, signifying that both the sequences are almost identical. *Escherichia coliK12* originated form a stool sample of a diarrhea patient had shown susceptibility to various drugs such as ampicillin, norfloxacin, ciprofloxacin, nalidix acid and erythromycin [14]. Also,

similar mutations at same positions in *Escherichia coli* GyrA and ParC had conferred resistance to two fluoroquinolones, ciprofloxacin and norfloxacin [4, 15, 16]. Therefore, crystal structure of *Escherichia coli* gyrase A (PDB_IDs: 1AB4) and *Escherichia coli* ParC (PDB_ID: 1ZVU) was retrieved from PDB (www.rcsb.org) and used as a template for homology modelling.

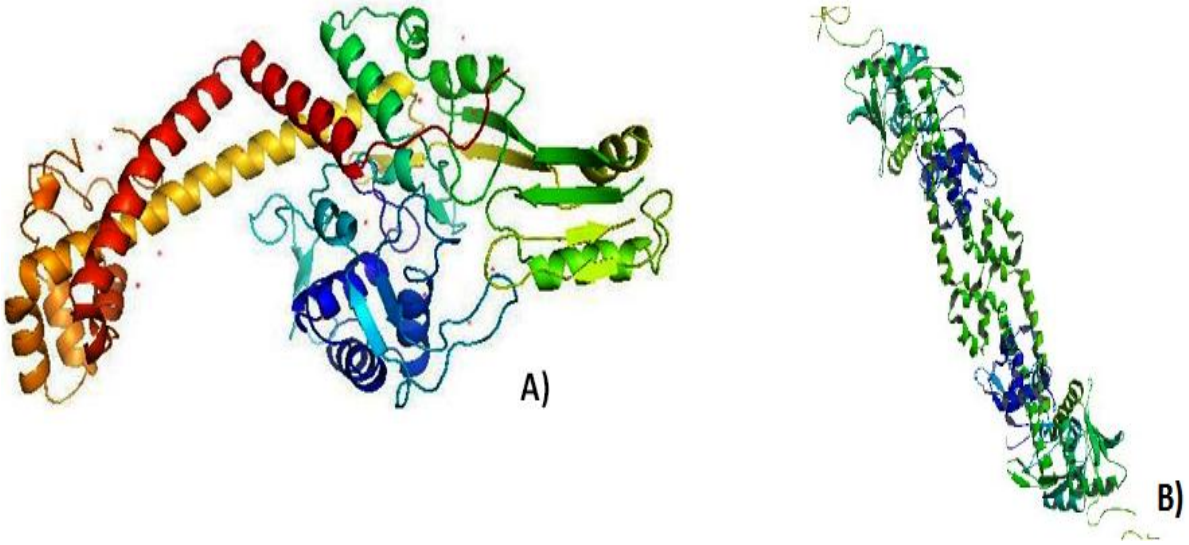


Figure 4.2Crystal structure of *Escherichia coli* used as template**A)** Gyrase A **B)** parC.

The protein structure was then modeled using Accelrys Discovery Studio v3.5. The length of the protein sequence retrieved after modelling was 875 amino acid residues and 752 amino acid residues for GyrA and ParC respectively. The best models chosen according to the lowest values of DOPE score were further evaluated using 3D verify (Accelrys Discovery Studio v3.5) [12] and ERRAT [17] program. This protein structure was referred in the study as wild type.

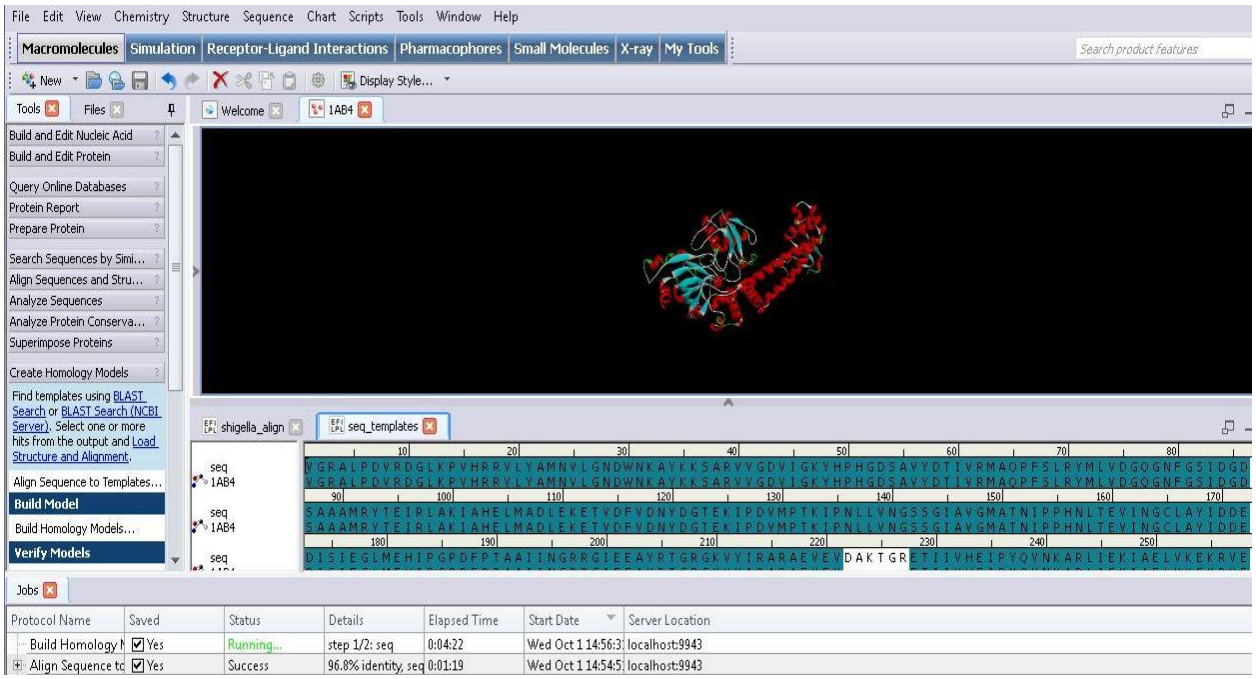


Figure 4.3Screenshot of the homology modelling performed using Discovery Studio

4.2.2.2 Mutated Protein Structures

Amino acid substitution corresponding to the mutation of Ser 83 to Leu and Asp 87 to Gly or Asn was introduced in the wild type protein structure of GyrA using Accelrys Discovery Studio v3.5 [12]. Since in our study, we had taken only N terminal sequence of *Shigella flexneri* DNA gyrase A so the positions corresponding to mutations Ser 83 and Asp 87 in the modeled wild type structure were residues 54 and 58 respectively. Hence two mutated structures were generated, one having mutations at Ser 54 to Leu and Asp 58 to Asn and was designated as mutant1. The other has mutations Ser 54 to Leu and Asp 58 to Gly and it was designated as mutant2.

Amino acid substitution corresponding to mutation Ser 80 to Ile was introduced in the wild type protein structure of ParC using Accelrys Discovery Studio v3.5 [12]. This mutated structure was designated as ParC mutant.

4.2.2.3 Structure Preparation and Minimization

The wild and mutated protein structures were then prepared and energy minimized using Conjugate Gradient and CHARMm force field with a gradient of 0.1 in Accelrys Discovery Studio v3.5.

4.2.3 Molecular Docking Studies

After ensuring the correct conformations of protein and ligands, molecular docking of the ligands to the wild and mutated structures was performed using BiosolveIT (version 2.1.6) [18] FlexX algorithm [19]. The binding site consists of 40 amino acids starting from 31st residue to 70th residues for both wild type and mutated molecules of *Shigella flexneri* gyrA, whereas in case of ParC binding site consists of 30 amino acids starting from 61st residue to 90th residues for both wild and mutated molecules. The reason behind choosing this site is that it encompasses all the reported residues involved in the interaction of these antibiotics with GyrA and ParC of *Shigella flexneri*. A total of 100 structures with best poses based on score and hydrogen bonds were screened out. Single best pose for each ligand was then chosen for further analysis.

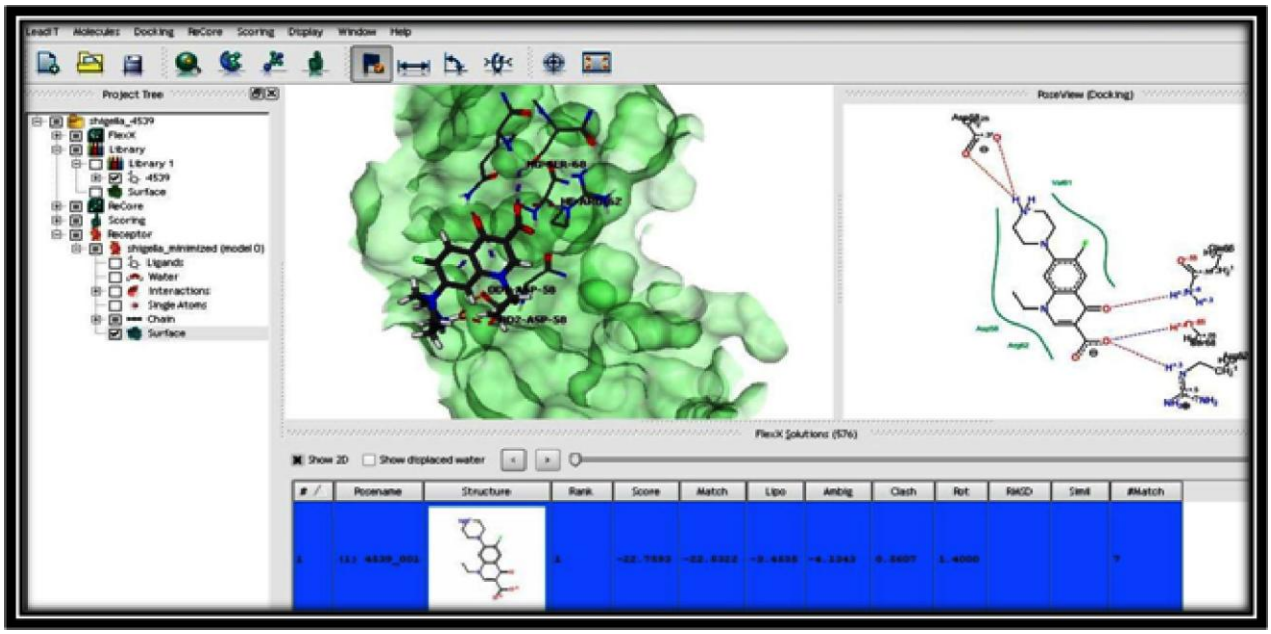
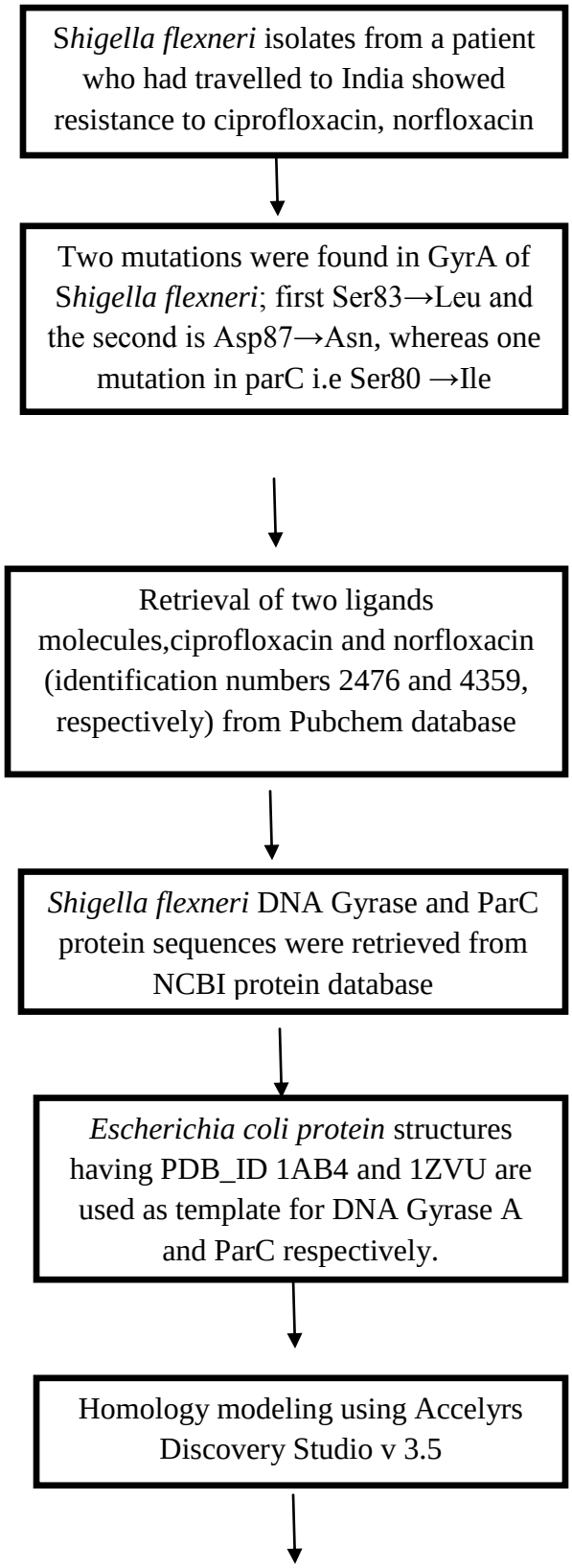
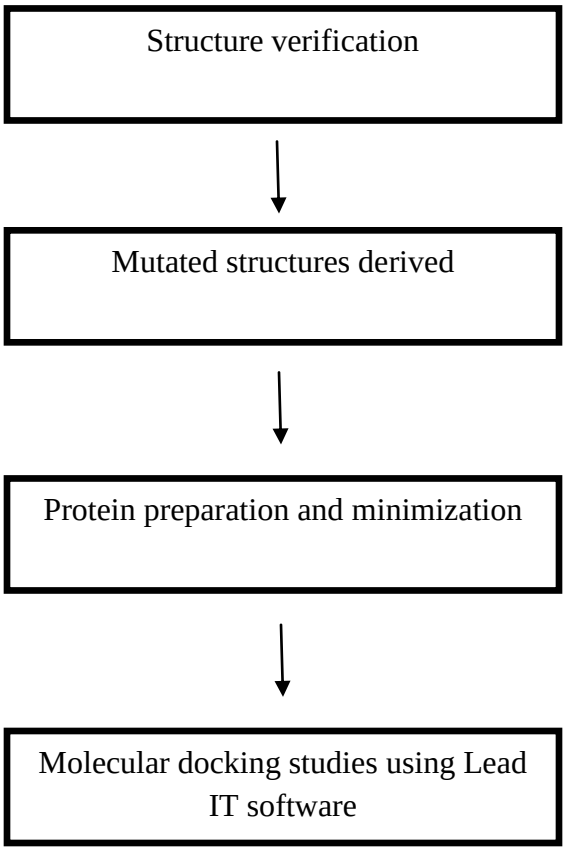


Figure 4.4Screenshot of the LeadIT interface used for docking of the proteins to ligand molecules**.**

4.2.4 Flowchart of the Experimental Procedure





4.3 RESULTS AND DISCUSSION

The results of ciprofloxacin and norfloxacin with respect to wild and two mutated GyrA are discussed below in terms of score and hydrogen bonds.

4.3.1 Ciprofloxacin Binding with Wild Type GyrA

The docking of ciprofloxacin with wild type protein showed energy of "-21.8005 kcal/mol" (Table 4.1), involving five hydrogen bonds via residues Asp 58, Ser 68, Arg 62 and Gln 65 as shown in Figure 4.5A. Also, fluorine atom of ciprofloxacin made close contact with the residue Val61 and also hydrophobically interacts with Asp 58. In general a high docking score (more negative value) reflects a strong interaction between the ligand and protein molecule. The docking score here suggested a strong binding between target protein and ciprofloxacin

Table 4.1 Residues and bonds involved in interactions of wild type and mutated protein molecule of *Shigella flexneri* DNA Gyrase A with ciprofloxacin and norfloxacin respectively

S.No	Compound Name*	Lead-IT Score (kcal/mol)	No of H-bond	Amino Acid	H-bond length (Å)
1	Ciprofloxacin-wild type	-21.8005	5	Asp58OD1 Asp58OD2 Gln65 Arg62 Ser68	2.22 2.05 1.88 2.16 2.36
2	Ciprofloxacin-Ser 54 to Leu and Asp 58 to Asn	-18.3914	3	Asp58OD1 Gln65 Arg62	2.14 1.89 1.75
3	Ciprofloxacin-Ser 54 to Leu and Asp 58 to Gly	-17.2574	2	Ala 55 Thr 59	1.59 1.56
4	Norfloxacin-wild type	-22.7593	5	Asp58OD1 Asp58OD2 Gln65 Arg62 Ser68	2.09 2.14 1.99 2.14 2.53
5	Norfloxacin-Ser 54 to Leu and Asp 58 to Asn	-18.1598	3	Asp58OD1 Gln65 Arg62	2.17 1.93 2.08
6	Norfloxacin-Ser 54 to Leu and Asp 58 to Gly	-18.3914	3	Leu54 Gln65 Arg62HE21 Arg62HH21	2.02 2.12 1.91 2.57

4.3.2 Ciprofloxacin Binding with GyrA Mutants

Decrease in the docking score was observed in both the structures, mutant1 (Score = -18.3914 kcal/mol) and mutant2 (Score = -17.2574 kcal/mol) involving only three bonds via residue Asp 58, Arg 62 and Gln 65 in case of mutant1 and two hydrogen bonds in mutant2 structure involving residues Ala 55 and Thr 59 (Figure 4.5B and 4.5C respectively). Also, interacting amino acids in mutant 2 were different from those in wild type. The major change noted here is the loss of two hydrogen bonds via residue Asp 58 and Ser 68 which showed the vital importance of these bonds in the binding of ciprofloxacin to GyrA of *Shigella flexneri*. Also, in this case there is a considerable shortening of hydrogen bond length (Table 4.1) which resulted in the distortion of some residues and their geometry.

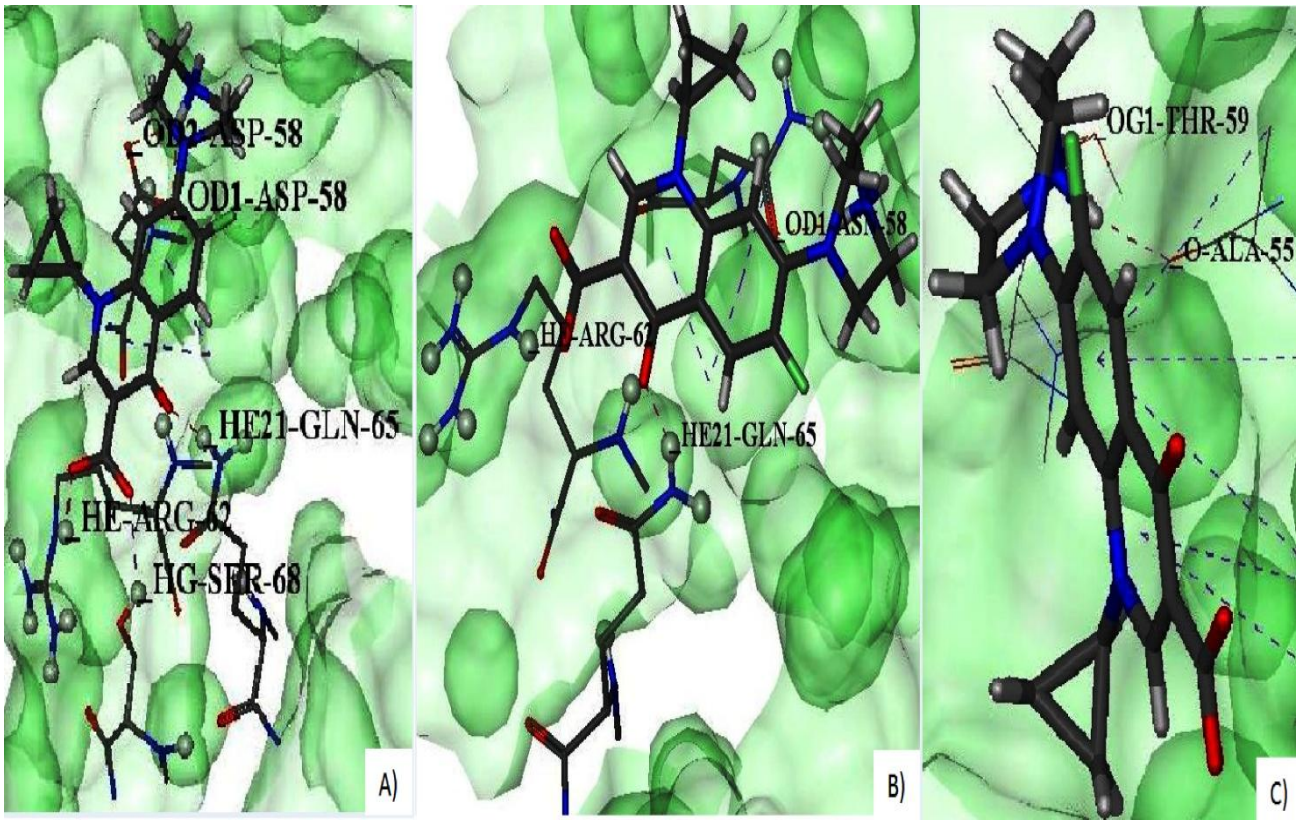


Figure 4.5Interaction of ciprofloxacin with *Shigella flexneri* DNA Gyrase A. A) with wild type. B) with mutant 1 C) with mutant 2**.**

4.3.3 Norfloxacin Binding with Wild Type GyrA

The docking of norfloxacin with wild type protein showed score of "-22.7593", involving five hydrogen bonds via residue Asp 58, Ser 68, Arg 62 and Gln 65 (Figure 4.6A). This suggested a strong binding affinity between target protein and the ligand molecule.

4.3.4 Norfloxacin Binding with GyrA mutants

In this case also, it was observed that there is a considerable decrease in the docking score in both the mutated structures, mutant1 (Score = -18.1598 kcal/mol) and mutant2 (Score = -18.3914 kcal/mol) (Table 4.1) involving only three bonds via residue Asp 58, Arg 62 and Gln 65 in mutant1 and four hydrogen bonds in mutant2 structure involving residues leu 54, Arg 62 and Gln 65 (Figure 4.6B and 4.6C, respectively). Here also, in mutant1 hydrogen bond loss was observed involving two residues Asp 58 and Ser 68. In case of mutant2 amino acids involved in interaction were different from those in wild type and there are only four hydrogen bonds participating in this particular interaction. These lost hydrogen bonds might be of great importance and play a significant role in binding of these residues with norfloxacin. Here also, in mutant type significant bond displacement was observed in some of the residues such as Asp 58 OD1 (where OD1 is the inner oxygen of the residue Asp forming hydrogen bond with corresponding ligand molecule), Gln 65 and Arg 62 (Table 4.1).

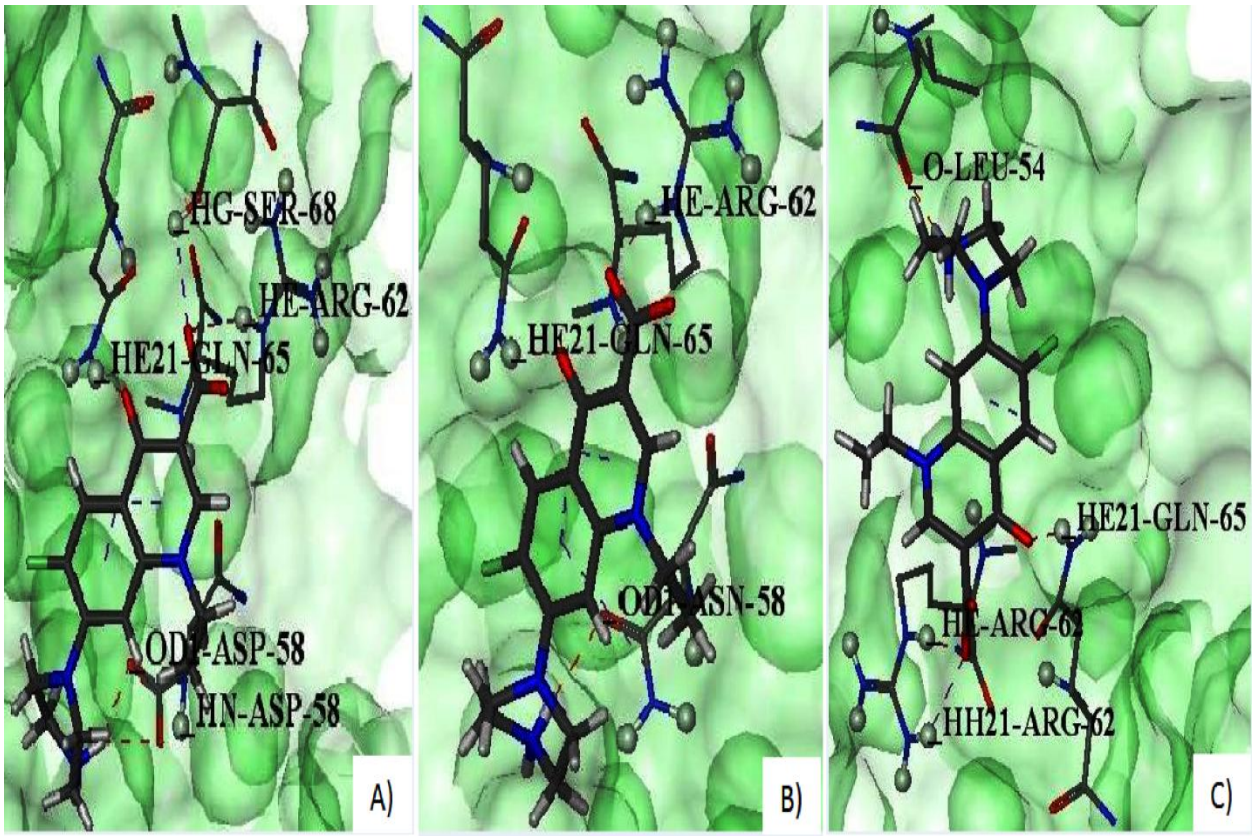


Figure 4.6 Interaction of norfloxacin with *Shigella flexneri* DNA Gyrase A. A) with wild type. B) with mutant 1 C) with mutant 2

4.3.5 Ciprofloxacin Binding with ParC

The docking of ciprofloxacin with wild type protein showed energy of "-13.8863 kcal/mol" (Table 2), involving two hydrogen bonds via residues Ser 80 and Glu 84 as shown in Figure 4.7A. The docking score suggested a strong binding between target protein and ciprofloxacin.

After introducing mutations in wild type structure of ParC, remarkable decrease in the docking score was observed (Score = -2.6835 kcal/mol) involving single bond via residue Gly 78 (Figure 4.7B). It was observed that unlike Ser 80, substituted Ile residues do not make direct hydrogen bond with ciprofloxacin and also there is increase in the hydrogen bond length in ParC mutant.

From these findings, it is clear that Ser 80 being directly hydrogen bonded to the ciprofloxacin in wild type protein plays a crucial role in the fluoroquinolone binding with ParC. This decreased docking score in mutant type accounts for increased resistance against fluoroquinolones.

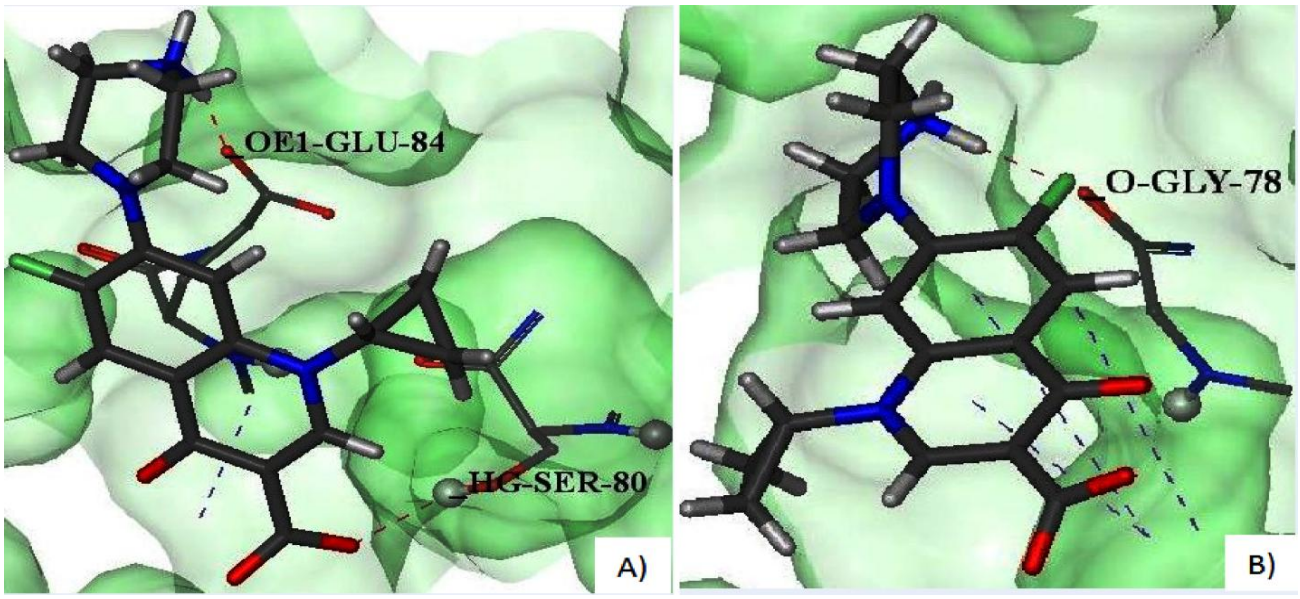


Figure 4.7 Interaction of ciprofloxacin with *Shigella flexneri* parC. A) with wild type B) with mutant type.

Table 4.2 Showing residues and bonds involved in interactions of wild type and mutated ParC protein molecule with ciprofloxacin and norfloxacin respectively

Sr. No	Compound Name	Lead-IT Score (kcal/mol)	No. of H-bond	Amino Acid	H-bond length (Å)
1	Ciprofloxacin-wild type	-13.8863	2	Ser80 Glu84	2.92 1.75
2	Ciprofloxacin-parC mutant	-2.6835	1	Gly78	3.01
3	Norfloxacin-wild type	-11.4378	2	Ser80 Glu84	2.74 2.12
4	Norfloxacin- parC mutant	-7.6866	1	Ala81	2.75

4.3.6 Norfloxacin binding with ParC

The docking of norfloxacin with wild type protein showed score of "-11.4378 kcal/mol", involving two hydrogen bonds via residue Ser 80 and Glu 84 (Figure 4.8A). This suggested a strong binding affinity between target protein and the ligand molecule.

Mutation of Ser 80 with Ile resulted in the considerable decreased docking score of -7.6866 kcal/mol involving only one hydrogen bond via residue Ala 81 as shown in figure 4.8B. Loss of Ser 80 in mutant type suggests a vital importance of this residue in norfloxacin binding with ParC.

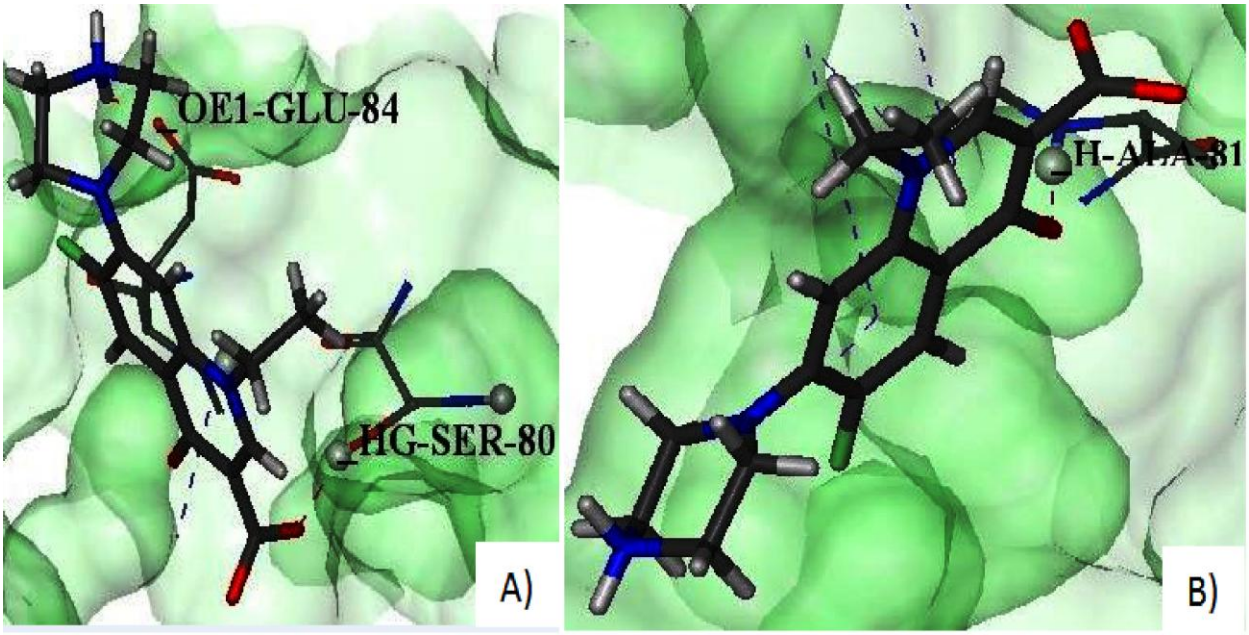


Figure 4.8 Interaction of norfloxacin with *Shigella flexneri* parC. A) with wild type B) with mutant type

4.4 CONCLUSION

The current computational studies provide insight into the interactions between target ligands (ciprofloxacin and norfloxacin) with *Shigellaflexneri* GyrA and ParC and its mutants using molecular docking to calculate binding energies and identifying key residues participating in interactions. From the results obtained it was evident that the decrease in docking score is more considerable in ParC than GyrA. This decreasing order of binding suggested that mutations in

ParC are more remarkable than mutations in GyrA, leading to induced resistance against the fluoroquinolones, ciprofloxacin and norfloxacin. This study reveals the relationship between the amino acid residues of the *Shigella flexneri* GyrA and ParC and the resistance mechanism to fluoroquinolones. Fluoroquinolone resistance takes place due to different mechanisms such as target site modification, by expulsion of the antimicrobial agents from the cell via general or specific efflux pumps or by plasmid-mediated fluoroquinolone resistance. Here we have studies the resistance mechanism due to modification of target binding site in bacteria. It was observed that both the mutations Ser 54 and Asp 58 in GyrA and Ser 80 in ParC are responsible for decreased interactions between fluoroquinolones, ciprofloxacin/norfloxacin and of *Shigella flexneri* DNA gyrase A. The amino acid residue Asp 58 in GyrA and Ser 80 in ParC makes direct hydrogen bonds with both ciprofloxacin and norfloxacin (wild type), so the mutations at this point leads to drastic changes in molecular interactions. The mutants have lower docking scores relative to the wild type proteins. These are not only due to hydrogen bond but also due to hydrophobic interactions that take place between these fluoroquinolones and active site residues of *Shigella flexneri*. In case of mutation Ser 54 to Leu in GyrA, Leucine being a bulkier molecule poses greater steric hindrance due to its side chain. From the above findings it is apparent that all the substitutions account for a decrease in the docking score and less efficient binding, ultimately leading to fluoroquinolone resistance in *Shigella flexneri* strains which were earlier sensitive to drugs. The molecular docking studies showed good correlation with experimental studies, hence provides a possible explanation for antibiotic resistance in *Shigella flexneri*. Further, these observations can be exploited to develop new drugs against the resistant strains of this pathogen.

Here, computational analysis on the experimentally proven mutations in *Shigella flexneri* GyrA and ParC against the two fluoroquinolones, ciprofloxacin and norfloxacin were performed and the results correlates very well with that of experimental results. The limitations to the studies are the inherent limitations of the docking algorithm. Some of these are: receptor flexibility, modelling cofactors, effectors and solvation effects[20]. Hence, the docking algorithms need to be further improved in these directions for an increased reliability of the results.

REFERENCES

[1] D.J. Diemert, "Prevention and Self-Treatment of Traveler's Diarrhea," *Clinical Microbiology Reviews*, vol. 19, (no. 3), pp. 583-594, 2006.

[2] L. Mensa, F. Marco, J. Vila, J. Gascón, and J. Ruiz, "Quinolone resistance among Shigella spp. isolated from travellers returning from India," *Clinical Microbiology and Infection*, vol. 14, (no. 3), pp. 279-281, 2008.

[3] K.D. von Seidlein L, Ali M, Lee H, Wang X, Thiem VD, et al., "A multicenter study of Shigella diarrhea in six Asian countries: disease burden, clinical manifestation, and microbiology," *PLoS Med*, vol. 3, (no. e353), 2006.

[4] Vashist J, Vishvanath, Kapoor R, Kapil A, Yennamalli R, Subbarao N, and R. MR., "Interaction of nalidixic acid and ciprofloxacin with wild type and mutated quinolone-resistance-determining region of DNA gyrase A.," *Indian Journal of Biochemistry and Biophysics*, vol. 46, (no. 2), pp. 147-153, April, 2009 2009.

[5] Y.L. Jeon, Y.-s. Nam, G. Lim, S.Y. Cho, Y.-T. Kim, J.-H. Jang, J. Kim, M. Park, and H.J. Lee, "Quinolone-resistant Shigella flexneri isolated in a patient who travelled to India," *Annals of laboratory medicine*, vol. 32, (no. 5), pp. 366-369, 2012.

[6] S.C. Kampranis and A. Maxwell, "The DNA Gyrase-Quinolone Complex: ATP HYDROLYSIS AND THE MECHANISM OF DNA CLEAVAGE," *Journal of Biological Chemistry*, vol. 273, (no. 35), pp. 22615-22626, 1998.

[7] J. Piton, S. Petrella, M. Delarue, G. Andre-Leroux, V. Jarlier, A. Aubry, and C. Mayer, "Structural Insights into the Quinolone Resistance Mechanism of *Mycobacterium tuberculosis* DNA Gyrase," *PLoS ONE*, vol. 5, (no. 8), pp. e12245, 2010.

[8] R.J. Reece, A. Maxwell, and J.C. Wang, "DNA Gyrase: Structure and Function," *Critical Reviews in Biochemistry and Molecular Biology*, vol. 26, (no. 3-4), pp. 335-375, 1991.

[9] J. Heddle and A. Maxwell, "Quinolone-Binding Pocket of DNA Gyrase: Role of GyrB," *Antimicrobial Agents and Chemotherapy*, vol. 46, (no. 6), pp. 1805-1815, June 1, 2002 2002.

[10] J. Ruiz, "Mechanisms of resistance to quinolones: target alterations, decreased accumulation and DNA gyrase protection," *Journal of Antimicrobial Chemotherapy*, vol. 51, (no. 5), pp. 1109-1117, 2003.

[11] F.D. Lowy, "Antimicrobial resistance: the example of Staphylococcus aureus," *The Journal of Clinical Investigation*, vol. 111, (no. 9), pp. 1265-1273, 2003.

[12] D.S.M.E. Accelrys Software Inc., Release 4.0, San Diego: Accelrys Software Inc., 2013.

[13] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, and D.J. Lipman, "Basic local alignment search tool," *Journal of Molecular Biology*, vol. 215, (no. 3), pp. 403-410, 1990.

[14] M.C. Sulavik, C. Houseweart, C. Cramer, N. Jiwani, N. Murgolo, J. Greene, B. DiDomenico, K.J. Shaw, G.H. Miller, R. Hare, and G. Shimer, "Antibiotic Susceptibility Profiles ofEscherichia coli Strains Lacking Multidrug Efflux Pump Genes," *Antimicrobial Agents and Chemotherapy*, vol. 45, (no. 4), pp. 1126-1136, 2001.

[15] P. Komp Lindgren, Ã.s. Karlsson, and D. Hughes, "Mutation Rate and Evolution of Fluoroquinolone Resistance in Escherichia coli Isolates from Patients with Urinary Tract Infections," *Antimicrobial Agents and Chemotherapy*, vol. 47, (no. 10), pp. 3222-3232, 2003.

[16] K. Poole, "Efflux-mediated multiresistance in Gram-negative bacteria," *Clinical Microbiology and Infection* vol. 10, pp. 12-26., 2004.

[17] C. Colovos and T.O. Yeates, "Verification of protein structures: Patterns of nonbonded atomic interactions," *Protein Science*, vol. 2, (no. 9), pp. 1511-1519, 1993.

[18] "Biosolveit LeadIT. version 2.1.6 http://www.biosolveit.de/LeadIT."

[19] R.M. Kramer B., and Lengauer T, " Evaluation of the FLEXX incremental construction algorithm for protein–ligand docking " *Proteins: Structure, Function, and Bioinformatics* vol. 37, pp. 228-241., 1999.

[20] M. Mihasan, "What in silico molecular docking can do for the ˜bench-working biologists," *Journal of Biosciences*, vol. 37, (no. 1), pp. 1089-1095, 2012.

CONCLUSION

AND

FUTURE PROSPECTS

CONCLUSION

Diarrhea is a very common term and is caused by different pathogens. It has long been considered as a major public health concern because of the morbidity and mortality it causes among all group of ages. In developing countries, where illnesses that cause diarrhea are more common and where health care is less readily available, diarrhea is a major health concern because of its potential to cause severe, life-threatening dehydration. Infants and the elderly are more prone to dehydration from diarrhea. Though the antibiotics and vaccines available to tackle diarrhea are effective to reduce the severity of the diseases but resistance of bacteria to these antibiotics has reached alarming levels in many parts of the world. Therefore, there is a continual need to develop new techniques and vaccine candidates that will be useful to reduce the burden of the disease.

In this work, known biological information, available sequence data and other associated information of diarrheal pathogens is used to develop novel methods in order to combat the increasing diarrheal disease. Future prospects and the practical applications of our approaches are also discussed briefly to provide new directions for the diarrhea related research.

Important findings of this thesis are summarized below:

- The database "dbDiarrhea" has been developed with an objective to provide all the relevant information about the diarrheal pathogens on a single platform. It is the first user-friendly interface that allows easy browsing and querying in various ways, thus selectively retrieving records from any module or functional category. dbDiarrhea provides important proteins from various diarrheal pathogens that could further be taken for experimental evaluation to identify new drugs or vaccine antigens against the major causative agents of diarrhea.
- In our second objective, for the rapid identification of Multidrug and Toxin Extrusion (MATE) proteins two approaches were applied. First is the Artificial neural Network (ANN) based approach and second is the Support Vector Machine (SVM) based approach. Different ANNs and SVM models were generated for the different types of features. But the results obtained using ANN approach were not as good as that of SVM based approach. So the web server "MATEPred" has been developed based on PSSM profiles using Support Vector Machine (SVM) approach yielding an overall accuracy of 92.06%, with the sensitivity and

specificity of 100% and 89.42% respectively along with an MCC of 0.82 and F-score of 0.83632.

- MATEPred efficiently distinguishes between MATE and Non-MATE sequences. Current MATE identification methods include experimental determination which require enormous efforts. The study presented here represents an initiative towards easy identification of MATEs from other proteins based on its PSSM profile.
- We further used the MATEPred server to scan the proteomes of two diarrheal species *Vibrio parahaemolyticus* and S*higella boydii*. Initially it reported eight and sixteen positives for each of these species, respectively. But in order to confirm whether these are actually MATE proteins or not, different types of analysis such as transmembrane helices prediction, Pfam domain and PROSITE analysis and BLAST analysis were performed. From all these analysis, five new potential MATE candidates (four from *Vibrio* and one from *Shigella*) are observed that can further be taken for experimental verification to study their role in drug resistance.
- The tool is expected to accelerate the identification of MATE proteins, thus providing new insights to find out the important therapeutic targets against resistant bacteria.
- As mentioned, the final study focused on performing molecular docking analysis of *Shigella flexneri* DNA Gyrase A and Topoisomerase IV with a fluoroquinolones, ciprofloxacin and norfloxacin. Both of these ligands showed stable interaction and the best binding affinity was calculated. The binding modes of ciprofloxacin and norfloxacin are quite similar with both the drugs showing strong interactions with wild type structures of *Shigella flexneri* GyrA and ParC as compared to that of mutants.
- The molecular docking studies presented are in good agreement with the experimental studies and provides a possible explanation for observed fluoroquinolone resistance. Further, the analysis of interaction can be exploited for better and more efficient design of new drugs against the resistant strains of this pathogen.

FUTURE PROSPECTS

- It is anticipated that this web based comprehensive resource "dbDiarrhea" would serve as a valuable accompaniment for analyzing proteins from major diarrheal pathogens and will also contribute scientific knowledge and help those working in this field. In future, dbDiarrhea will continue to be updated and refined with increased data so as to make it more useful.
- Five new potential MATE candidates from diarrheal pathogens identified using prediction server can be further taken for experimental verification.
- Interaction studies presented here will help in designing of new drug with some modifications in order to combat the problem of antibiotic resistance.

PUBLICATIONS

AND

PRESENTATIONS

PAPERS IN INTERNATIONAL REFEREED JOURNALS:

1. Jayashree Ramana, **Tamanna**. DbDiarrhea: The database of pathogen proteins and vaccine antigens from diarrheal pathogens. *Infection Genetics and Evolution*, vol. 12(8), pp. 1647-1651, 2012.[ISSN: 1567-1348, **IF: 2.885**]

2. **Tamanna,** Jayashree Ramana. MATEPRED- A SVM Based Prediction Method for Multidrug And Toxin Extrusion (MATE) Proteins. *Computational Biology and Chemistry*. vol. 58 , pp. 199-204, 2015 [ISSN: 1476-9271, **IF: 1.331**]

3. **Tamanna**, Jayashree Ramana. Structural Insights into the Fluoroquinolone Resistance Mechanism of *Shigella flexneri* DNA Gyrase and Topoisomerase IV. *Microbial drug Resistance.***[IF:2.306**]

PRESENTATIONS IN CONFERENCES:

Tamanna, Ramana J., "An ANN-based method for prediction of Multidrug And Toxin Extrusion (MATE) Proteins" **Poster Presentation** at 'World Congress on Stem Cell Research, Cancer Biology and Applied Biotechnology (Biotech-2014)' in *Jawaharlal Nehru University, New Delhi*, INDIAfrom 3-4, May 2014.

WORKSHOPS ATTENDED:

A **"Bioinformatics workshop"** organized jointly by **University of Nebraska**, Omaha, USA and **Jaypee University of Information Technology**, Solan, H.P., India from 8-10, May 2013.