COMPUTATIONAL ANALYSIS OF MEMBRANE ASSOCIATED GENES OF *MYCOBACTERIUM FORTUITUM* TO VALIDATE AS DRUG TARGETS

Dissertation submitted in partial fulfillment of the requirement for the degree of

BACHELOR OF TECHNOLOGY

IN

BIOTECHNOLOGY

By

Shubham Mittal (Enrollment number: 141810)

Under the guidance of

Dr. Rahul Shrivastava

&

Dr. Ragothaman M. Yennamalli

DEPARTMENT OF BIOTECHNOLOGY AND BIOINFORMATICS

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY,

WAKNAGHAT, SOLAN, HIMACHAL PRADESH -173234.

May – 2018

TABLE OF CONTENTS

TOPICS **PAGE NUMBER**

DECLARATION

I hereby declare that the work reported in the B.Tech thesis entitled **"Computational analysis of membrane associated genes of *Mycobacterium fortuitum* to validate as drug targets"** submitted for the degree of of B.Tech. in Biotechnology at **Jaypee University of Information Technology, Waknaghat India,** is an authentic record of my work carried out under the supervision of **Dr. Rahul Shrivastava & Dr. Ragothaman M. Yennamalli**. I have not submitted this work elsewhere for any other degree or diploma.

Shubham Mittal (141810)

Department of Biotechnology and Bioinformatics

JUIT, Solan.

Date: 15.05.2018

CERTIFICATE

This is to certify that the work titled **"Computational analysis of membrane associated genes of *Mycobacterium fortuitum* to validate as drug targets"**, submitted by **Shubham Mittal (141810)** in partial fulfillment for the award of degree of **B.Tech.** in **Biotechnology** at **Jaypee University of Information Technology, Solan** is a bonafide record of his original work carried out 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. Rahul Shrivastava

(Associate Professor)

Department of Biotechnology and Bioinformatics

JUIT, Solan

Date: 10/5/18

Dr. Ragothaman M. Yennamalli

(Assistant Professor (Grade II))

Department of Biotechnology and Bioinformatics

JUIT, Solan

Date: 16/5/18

ACKNOWLEDGEMENT

I would like to take this opportunity to thank all those individuals who have, directly or indirectly, contributed in the progress and completion of this project.

I thank **Dr. Sudhir Kumar,** Professor and Head, Department of Biotechnology and Bioinformatics, for providing me the opportunity and the facilities to carry out the project.

My profound gratitude goes to my project supervisors **Dr. Rahul Shrivastava** and **Dr. Ragothaman M. Yennamalli** for their unmatched patience, attention and care without which this project would not have been possible. Their knowledge and acute understanding of related subjects were what guided me all along in my project work titled **— Computational analysis of membrane associated genes of *Mycobacterium fortuitum* to validate as drug targets.** I am indebted to them.

My sincerest gratitude and heartfelt thanks goes to **Ms. Poonam Katoch** (PhD) a positive person in the very essence of her being. She has been a constant pillar of support during this endeavor. Her never-give-up attitude and constant guidance were essential factors in inspiring me to consistently research. I am also thankful to her for providing me with the sequences of *M. fortuitum* upon which I based my study.

It is my privilege and a matter of pride to thank my parents, family and friends for their constant encouragement and a shoulder to lean upon when things did not go as planned. Their kindness, support and care can never be repaid.

Shubham Mittal (141810)

ABSTRACT

Mycobacterium fortuitum is a rapidly growing, non-tuberculous mycobacterium which was discovered in 1938. It is known to be the cause of opportunistic infections such as osteomyelitis, joint infections, nosocomial infections, local cutaneous diseases and post-surgical infections. However recently it is emerging as a pathogen and garnering attention world over due to increase in its incidence in immuno-competent individuals as well. Membrane genes are responsible for a variety of functions in both Eukaryotic and Prokaryotic organisms. They are involved in transport, cell signaling, antibiotic resistance, and several other important functions. Their importance makes them essential to the virulent mechanisms of pathogens which is why 50% of all drugs target membrane genes for inhibition.

Tryptophan is a amino acid necessary to the survival of all mycobacterial species. Therefore, creating an absence of Tryptophan in target host cells of mycobacteria i.e. the macrophages in humans could have been a strategy against the pathogen, however most mycobacterial species including *M. fortuitum* contain within themselves a Tryptophan biosynthesis machinery. Inhibiting the genes of this machinery or associated membrane genes may serve as a potential drug action against *M. fortuitum.*

Random transposon mutagenesis experiments done *in vivo* provided us with the nucleotide sequence whose absence causes attenuation in the *M. fortuitum* infection. Computational *in silico* analysis of the sequence obtained was performed using a variety of tools and databases such as – BLAST (blastn, blastx, blastp), ExPASy, MSA – Jalview and Tcoffee, for homological studies and STRING, MINT, InAct, BioGRID, for protein interaction studies for functional interactors. Studies revealed the proteins ‘Anthranilate synthase component I’ and ‘Anthranilate synthase component II’ which may be involved in virulence of *M. fortuitum.* These genes identified may serve as potential drug targets in the development of novel drugs or other effective intervention strategies.

LIST OF ABBREVIATIONS AND SYMBOLS

°C	Degree Celsius
%	Percentage
MTBC	Mycobacterium tuberculosis complex
NTM	Non-tuberculosis Mycobacterium
CAPD	Continuous ambulatory peritoneal dialysis
CAMs	Cell Adhesion Molecules
IDO	indoleamine-2,3-dioxygenase
CD4	Cluster of differentiation 4
IFN	Interferon
TNF	Tumor necrosis factor
BLAST	Basic local alignment search tool
MSA	Multiple Sequence Alignment
STRING	Search tool for the retrieval of interacting genes/proteins

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1. INTRODUCTION

Membrane associated genes are those which either code for membrane proteins or proteins that interact with membrane proteins. Membrane proteins are either embedded inside the membrane or are attached to the membrane surface of cells or organelles. Many membrane proteins are transport proteins which carry ions and molecules across the membrane in either direction. Cell Adhesion Molecules (CAMs) are membrane proteins upon which a cell relies in order to recognize, make contact and interact with other cells. Membrane protein enzymes play a role in transferase, hydrolase, oxidoreductase, aminopeptidase and many similar activities. Membrane receptor proteins are responsible for cell signaling and for cell communication [1],[2]. Apart from these providing antibiotic resistance is a very important function of some membrane proteins. Such proteins are linked to the immune response of the organism against pathogenic attack. Interestingly, membrane and membrane associated proteins are potential targets for greater than sixty percent of all modern drugs [3].

Membrane proteins play a major role in the pathogenesis of microorganisms. Mycobacterium species belonging to the phylum Actinobacteria are notoriously known to make use of membrane proteins for their survival [5]. Mycobacterium tuberculosis complex (MTBC) and Non-tuberculosis Mycobacterium (NTM) are the two major groups of Mycobacterium species [6],[7]. Due to their response to a laboratory staining technique *Mycobacteria* are often called "acid-fast bacteria". Lately *Mycobacterium fortuitum,* an NTM species is emerging as a pathogenic species responsible not just for immunocompromised by also immunocompetent infections up till a varying degree. It is this incidence of infection that is garnering major attention from the Mycobacterial community all over [10],[11].

Mycobacterium fortuitum is a rapidly growing species that is very commonly found in the environment. It can grow within 3 to 4 days on laboratory media at an optimal temperature of 30-37°C while other Mycobacterial species can take weeks to grow [8],[9]. This is the reason why it is classified under Runyon Group IV (Edward Runyon was the scientist who categorized all NTMs). *M. fortuitum* can cause osteomyelitis (inflammation of the bone), joint infections, infections of

the eye after trauma and local skin disease, however pulmonary infections of *M. fortuitum* are uncommon [14].

M. tuberculosis was famously discovered by Robert Koch in 1882, and *M. fortuitum* was discovered in 1938. Most surgical site infections by *M. fortuitum* occur due to the ability to persist as biofilm within healthcare units. Biofilm formation involves membrane genes and the survival of these pathogenic membrane genes is dependent upon the continuation of particular molecular mechanisms of the microorganism [17].

Tryptophan Biosynthesis is one such molecular mechanism necessary for the survival of all Mycobacterial species. Amino acid such as Tryptophan is important for protein synthesis and eventually for the survival of many microorganisms [18]. In human beings Tryptophan is an essential amino acid and it needs to be taken in through diet. Macrophages are the target sites for Mycobacterial attack and survival. It has been reported that if Tryptophan is depleted inside a macrophage then in order to survive Mycobacterium kick starts Tryptophan biosynthesis pathway [21]. If this pathway can be inhibited then there is a chance of killing the pathogenic microbe and preventing infection. Post wet lab experiments use of computational tools to identify 'key' genes in Tryptophan biosynthesis pathway, upon inhibiting which the infection can be prevented, is the strategy of the current study.

CHAPTER 2

REVIEW OF LITERATURE

2. REVIEW OF LITERATURE

2.1 Introduction

The Greek prefix myco means “fungus” due to the mold-like pattern of growth of Mycobacterium on the surface of cultures, however Mycobacterium is a bacterial genus. It belongs to the family Mycobacteriaceae and the phylum Actinobacteria which naturally categories them among gram positive bacteria. Previously *Mycobacterium tuberculosis* and *Mycobacterium leprae* were the majorly known Mycobacterial species and other mycobacterial strains were thought to be unusual *M. tuberculosis* strains - they were hence called “atypical” [8],[9]. But with time a more extensive classification was developed according to which Mycobacterium species can be classified into *Mycobacterium tuberculosis* complex (MTBC) and Non-tuberculosis Mycobacterium (NTM). *Mycobacterium tuberculosis* complex (MTBC) comprises of *M. tuberculosis (MTB), M. bovis, M. microti, M. orygis, M. africanum, M. caprae, M. pinnipedii, M. canettii, M. mungi and M. suricattae,* all of which are characterized by their genetic similarity and the ability to cause tuberculosis in living organisms including humans [26].

Nontuberculous Mycobacterium (NTM) comprise those mycobacterial species that are not members of the *Mycobacterium tuberculosis* complex. *M. abscessus, M. kansasii, M. fortuitum, M. avium, M. chelonae, M. fortuitum, M. gordonae, M. xenopi, M. szulgai, M. haemophilum* are the major NTM species out of all 186 different species of NTM [17]. In the last few decades a growing number of fatal and treatment-resistant diseases caused by strains of NTM species have come to notice. In countries of Europe and in the West it has been found that bacteria in the *Mycobacterium avium* complex (MAC) preponderate, followed by *M. xenopi* and *M. gordonae.* Rapidly growing *M. fortuitum* is the most frequently occurring NTM in the European Union [13]. In India, a similar pattern is observed with the most commonly isolated NTM being *M. fortuitum* [4],[16].

Mycobacterium fortuitum is a rapidly growing, gram positive, rod shaped mycobacterium. It belongs to the phylum Actinobacteria and is ubiquitous with a generation time of about 4 hours. It is known to be prevalent in processed water and natural water, in sewage and in dirt. It is one of the major species behind

infections related to hospitals making it a prime source of nosocomial infections [11]. The mechanism of pathogenicity of *M. fortuitum* involves virulent membrane genes of the microorganism. These membrane genes are associated with effectively facilitating its survival strategy in host macrophage cells [22]. Not many drugs have been developed against the pathogen, two examples of such drugs are Ciproflxacin and Amikacin. Moreover, obstruction is imposed on the way of its treatment as it is continuously evolving resistance to the currently prescribed drugs, such as against Amikacin [23]. *M. tuberculosis* drugs are also known to be used against *M. fortuitum,* although ineffectively. The clinical manifestations of *M. fortuitum* infections are similar to tuberculosis, however a prolonged treatment to *M. tuberculosis* drugs results in enhanced genetic resistance in *M. fortuitum* along with more side effects [23]. For this reason, computational study of virulent factors that can be used as drug targets against *M. fortuitum* infections is an essential topic.

2.2 Epidemiology

M. fortuitum cause infections other than tuberculosis. It is not mandatory to report *M. fortuitum* infections. Hence, it is almost impossible to estimate the exact prevalence and incidence of *M. fortuitum* diseases. Nevertheless, it is one of the most clinically important NTM with highest prevalence in Asian countries including Saudi Arabia and India [13],[14]. NTMs are found in the environment particularly in water and the soil and apart from animals they may infect humans occasionally causing a variety of pathological infections. Prevalence of NTM infections is on the rise today and they can affect a wide number of tissues and body fluids causing both respiratory (chronic bronchopulmonary illness) diseases and dermal infections [17]. *M. fortuitum* is theoretically an opportunistic pathogen and causes infection in immunocompromised humans. However, an increasing trend of *M. fortuitum* infections in healthy individuals i.e. those who are immunocompetent, is observed lately.

2.3 Symptoms and Diseases of *M. fortuitum*

M. fortuitum can cause infection in different parts of the body such as the eyes, bones, skin and/or blood. It is one of the leading causes of opportunistic infections,

be it local cutaneous diseases, osteomyelitis, joint infections, ocular disease, post-surgical infections as well as disseminated infections [28],[29].

Disseminated disease, usually with soft tissue lesions and disseminated skin lesions, is almost restricted to immunocompromised individuals, especially AIDS patients. Endocarditis has also been found to be reported. Surgical-site infections due to *M. fortuitum* infection are common, and especially associated with cardiothoracic surgery. It is due to frequent contamination of the wound, directly or indirectly, by colonized tap water.

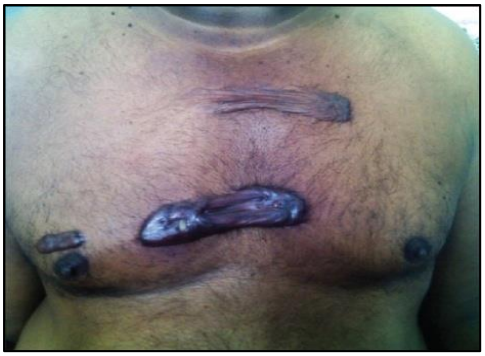

Figure 2.1: Multifocal keloids associated with *M. fortuitum* following intralesional steroid therapy [41].

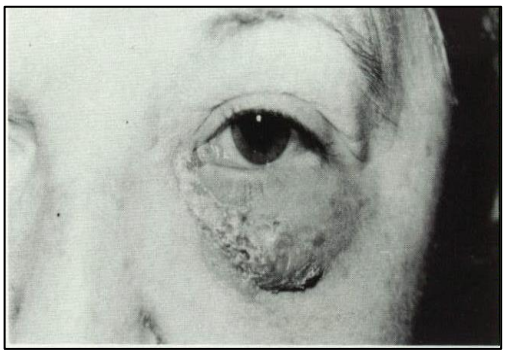

Figure 2.2: Patient with Dacryocystitis caused by *M. fortuitum* [42].

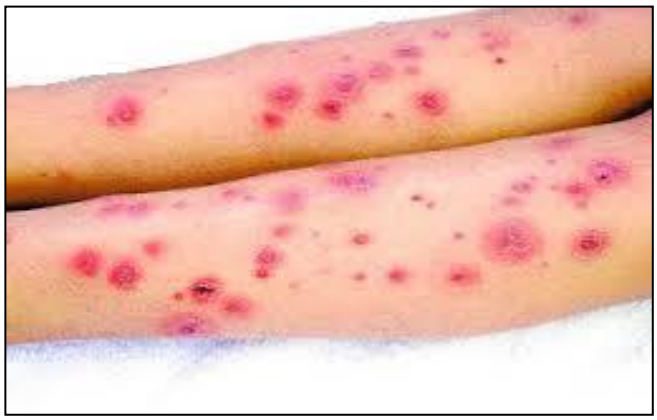

Figure 2.3: Skin lesions caused by *M. fortuitum* [43].

Other nosocomial infections with this organism include infections due to implanted devices and abscesses at injection-site. One case of meningitis due to *M. fortuitum* has been reported in association with a puncture wound and foreign body, while another has been reported with AIDS. *M. fortuitum* has been

previously documented as be the cause of peritonitis in patients who received continuous ambulatory peritoneal dialysis. Now, it is being more commonly isolated from soft tissue and skin infections. Emphasis was given on the fact that the patients with CAPD peritonitis in which conventional culturing repeatedly yields no organisms, the culturing for organisms such as mycobacteria should be done. Late episodes have additionally been depicted in immuno-competent has after utilization of debased whirlpool footbaths in nail salons [31].

2.4 Membrane associated proteins

Plasma membrane is the outermost layer of a cell and it serves to separate the cell from the external environment. While in Prokaryotes it is the only membrane, Eukaryotes have membrane-bound organelles as well. All types of membranes are made up of the same elemental phospholipid bilayer. Membrane proteins are seated in the phospholipid bilayer. Membrane associated proteins are proteins that interact with or are part of these membranes proteins [1]. Some proteins are situated inside the membrane and are called transmembrane proteins while others on the surface on one side of the membrane and are called integral monotypic proteins [2].

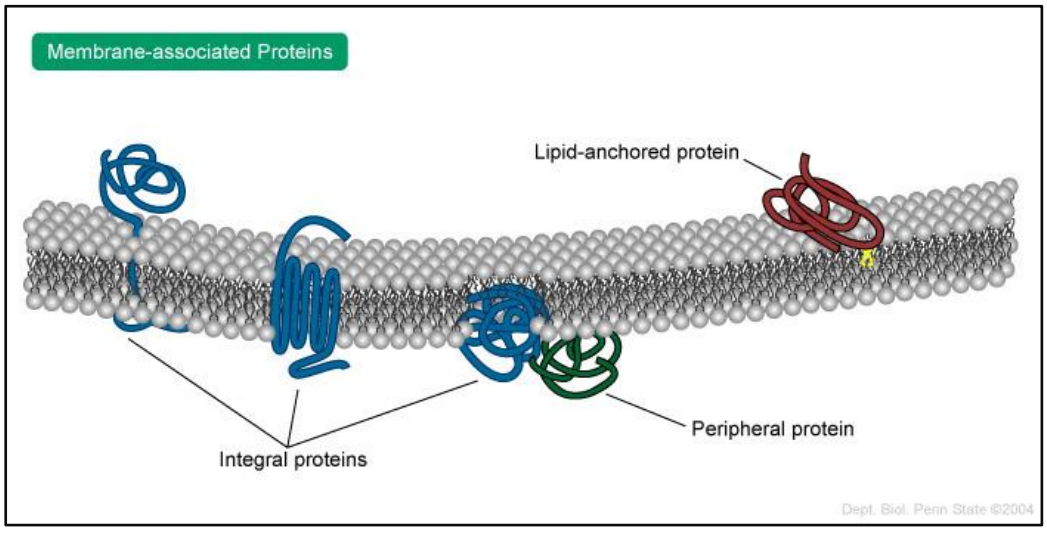

Figure 2.4: Membrane-associated proteins
(Source: Dept. Biology Penn State, 2004)

Membrane receptor proteins are responsible for cell signaling for cell communication. Membrane protein enzymes play a role in transferase, hydrolase, oxidoreductase, aminopeptidase and many similar activities. Many membrane

proteins are transport proteins which carry ions and molecules across the membrane in either direction. Cell Adhesion Molecules (CAMs) are membrane proteins upon which a cell relies in order to recognize, make contact and interact with other cells. Some of the very important membrane proteins are the ones that provide antibiotic resistance. For example, Colicins and Hemolysins are antibacterial peptides. Such proteins are linked to the immune response of the organism against pathogenic attack. Therefore, the strategic location and vital functions of membrane associated proteins establish their importance for the survival of any living organism. An estimate suggests that 23% of of human proteome is made up of membrane proteins. Moreover membrane and membrane associated proteins are potential targets for greater than 60% of all modern drugs that block receptors or inhibit enzymes related to diseases [3].

Macrophages are immune cells that engulf foreign materials that enter the body including microorganisms. However, the ability of *M. fortuitum* to persist inside the macrophages themselves, which is similar to *M. tuberculosis* is the major step and mechanism required for its pathogenesis [32]. There are several pathways via which *M. fortuitum* achieves this, many of which involve virulent membrane genes.

2.5 Tryptophan Biosynthesis – a survival strategy

Various amino acids are required by living organisms, including microorganisms, for their survival. Tryptophan is one such amino acid. It is evident that it is nutritionally essential for monogastric animals and preweaning ruminants because it cannot be synthesized in the body. Besides serving as a building block for proteins, Tryptophan is a critical nutrient for the functions of nervous and immune systems. Recent studies show that Tryptophan and its metabolites (e.g., serotonin (5-hydroxytryptamine, 5-HT) and melatonin) can regulate feed intake, reproduction, immunity, neurological function, and anti-stress responses [33]. Hence, over the past decades, much attention has been directed to study the role of Tryptophan as a limiting amino acid in mammalian and avian nutrition. Additionally, Tryptophan may modulate gene expression and nutrient metabolism to impact whole-body homeostasis in organisms. Thus, adequate intake of this amino acid from the diet is crucial for growth, development, and health of animals

and humans. Studies on *M. tuberculosis* showed that Tryptophan is a highly important amino acid when it comes to survival and pathogenicity. Most microorganisms require Tryptophan from external sources i.e. the host and cannot synthesize it but Mycobacterium species such as *tuberculosis* and *fortuitum* have a Tryptophan biosynthesis machinery [21]. In fact, Tryptophan biosynthesis protects mycobacteria from CD4 T cell-mediated killing thus emphasizing its role in virulence.

Previously it was found that Tryptophan biosynthesis pathway is an important drug target pathway against intracellular pathogen as IFN gamma activated macrophages induces indoleamine-2,3-dioxygenase (IDO) that catabolizes the available Tryptophan into kynurenines inside the macrophages to make it unavailable for pathogen. Therefore the pathogen must synthesize Tryptophan *de novo*, to survive inside macrophages. Loss of this pathway makes the pathogen hypersusceptible to IDO induced immunoclearance. Mammals do not synthesize Tryptophan, thus enzymes of this pathway are an important drug target [34].

Also, CD4 T cells induced the requirement for bacterial Tryptophan biosynthesis suggesting that interfering with Tryptophan metabolism might offer an appealing therapeutic strategy, however Zhang et al. found that trpE deletion mutants of M. tuberculosis attenuated in macrophages in the absence of CD4 T cells, IFN gamma and TNF alpha indicating, CD4 independent mechanism of Tryptophan starvation by host [21]. TrpE deletion mutants also showed limited growth in vitro inside macrophages as well as in mouse model indicating the role of Tryptophan synthesis in virulence.

Tryptophan operon was first discovered in 1953 by Jaques Monod in *Escherichia coli*. After much study its gene expression was concluded to be one of repressible negative regulation which means that Tryptophan synthesis can be inhibited by the chemical ‘Tryptophan’ i.e. the trp operon is repressed or ‘turned off’ when Tryptophan levels are high and expressed or ‘turned on’ when they are low.

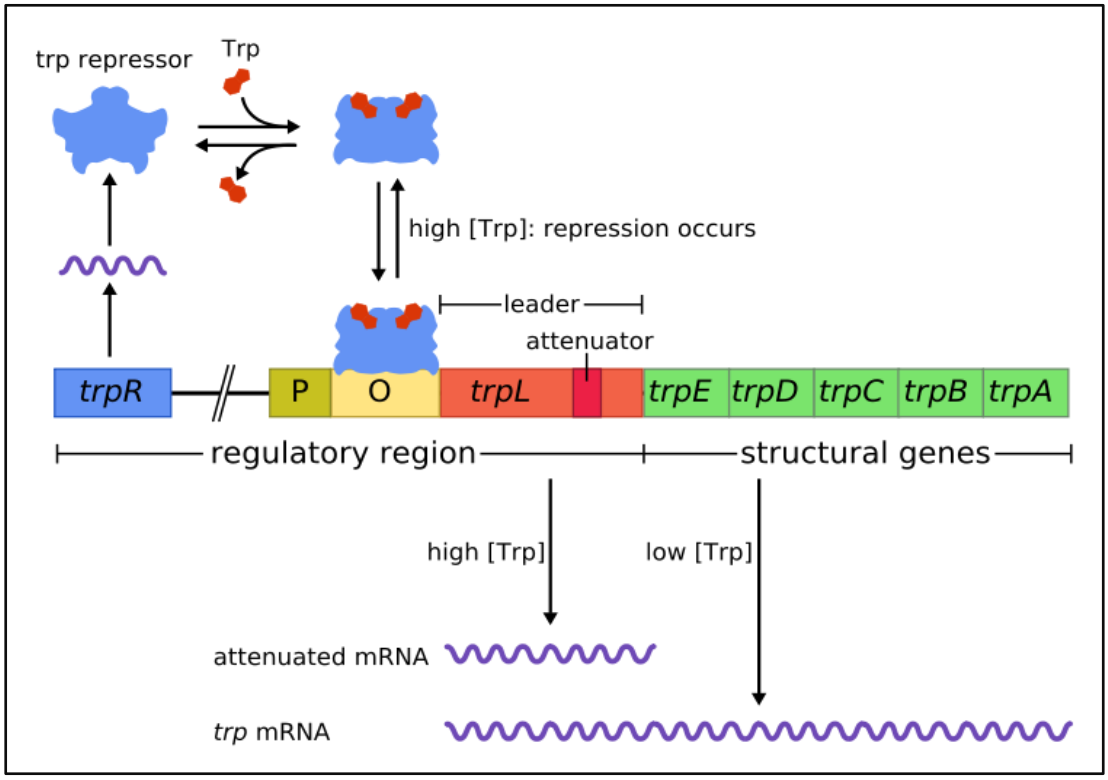

Figure 2.5: Tryptophan operon in *E.Coli*

(Source: Creative Commons)

The Tryptophan operon comprises of many different genes all of which are involved in the various steps in the synthesis of Tryptophan. TrpE, trpD, trpC, trpB and trpA are the structural genes that encode enzymes needed for Tryptophan biosynthesis. trpR is the Tryptophan repressor which regulates expression of the operon [18].

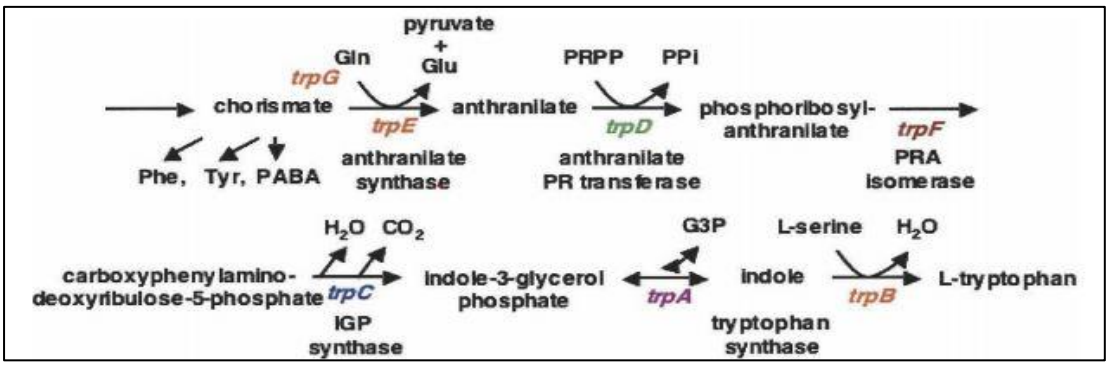

Figure 2.6: The genes, enzymes, and reactions of the Tryptophan biosynthetic pathway in *E.Coli*
(Source: *E.Coli* Trp repressor, iGEM 2009)

trpE codes for Anthranilate Synthase protein and is involved in the first step of synthesis of L-Tryptophan from chorismate. trpD codes for Anthranilate

phosphoribosyltransferase and is involved in the second step. trpC codes for Phosphoribosyl Anthranilate Isomerase/Indole-3-glycerol-phosphate Synthase, while trpA and trpB code for Tryptophan Synthase Alpha and Tryptophan Synthase Beta proteins, which are involved in the final steps of Tryptophan biosynthesis pathway [18].

2.6 Hypothesis

Recollecting data from all of previous literature we observe that *M. fortuitum* is a non-tuberculous emerging pathogen whose infections are proving more fatal as seen from its growing incidence in immunocompetent individuals when compared to greater number of immunocompromised individuals getting infected before. With this increase in lethality of infection most drugs developed previously are proving to be ineffective against *M. fortuitum* [23]. Therefore there is an urgent need to identify new drug targets in *M. fortuitum* and develop new drugs for them. Most effective drugs are those whose drug targets are genes directly involved in the virulence of the microorganism (*M. fortuitum*). The inhibition of such a drug target greatly diminishes the chances of the microorganism (*M. fortuitum*) of causing an infection [12]. Tryptophan biosynthesis pathway, an effective survival strategy of mycobacterium in macrophages of host, has proven to be an important pathway whose genes can be explored as drug targets [18]. Furthermore, trpD, trpE and trpG, which are among the first few genes in the Tryptophan operon and the Tryptophan biosynthesis pathway have proven to be essential for the pathway to occur [21]. Study of these genes computationally after experimentally emphasizing their importance in pathogenicity of *M. fortuitum*, can identify new drug targets among these genes or among the membrane genes associated with them. Homology study and protein-protein interaction analysis of genes identified from experimental data could help in conducting an extensive search for potential drug targets.

CHAPTER 3

MATERIALS AND METHODS

3. MATERIALS AND METHODS

3.1 Obtaining Nucleotide Sequences

Nucleotide sequence used for study were obtained after random mutagenesis of *M. fortuitum* ATCC 6841 followed by cloning and sequencing. Random mutagenesis was done by using transposon, where pRT291 vector containing Transposon TnphoA, Reporter gene phoA, Kanamycin resistant gene and Transposase are used. A library was constructed and *in vivo* analysis of one mutant in mice model showed attenuated virulence. The *M. fortuitum* mutant was further subjected to genomic analysis by which a nucleotide sequence was obtained after cloning and sequencing of mutated membrane gene [Acknowledgement: Poonam].

3.2 Homological Analysis of Nucleotide Sequence obtained

The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity (homology) between biological sequences. The program helps in identifying or characterizing a nucleotide or protein sequence by comparing it to sequence databases and calculating the statistical significance of matches [36]. BLAST is greatly used to infer and identify members of gene families. A closer association of query sequence with a known sequence, protein or nucleotide, is identified by the alignment with maximum query coverage, least e-value (Expectation value, which represents the number of different alignments with score better than or equivalent to the score of current alignment), maximum score (of alignment, calculated by various statistical parameters) and maximum identity (percentage which describes the extent to which two sequences have same residues at the same positions). BLAST is fast and can be used to search extremely large databases. It is sufficiently sensitive and selective for most purposes. and uses a pre-indexed database. Moreover, it is robust i.e. the default parameters can usually be used.

There are five basic types of BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) which are a collection of five different programs that allow different combinations of query and database sequences to be used for alignment.

- blastn – DNA probe aligned against DNA database
- blastp – protein probe aligned against protein database

- blastx – translated DNA probe aligned against protein database
- tbastn – protein probe aligned against translated DNA database
- tblastx – translated DNA probe aligned against translated DNA database

3.2.1 Blastn

3.2.1.1 Not limited to '*Mycobacterium'*

Procedure:

1. The query sequence (nucleotide) obtained from sequencing after Random Mutagenesis experiments was taken.
2. blastn tool was used for alignment as it checks alignment of nucleotide sequence against nucleotide sequence database.
3. The query sequence was copied onto the space labelled 'Enter Query sequence' in blastn.
4. The database selected was 'Non-redundant protein sequences' (nr).
5. The alignment was performed against all known organisms.
6. All other parameters were set at default.
7. Local alignment command was given.

3.2.1.2 Limited to '*Mycobacterium (taxid:1763)'*

Procedure:

1. The query sequence (nucleotide) obtained from sequencing after Random Mutagenesis experiments was taken.
2. blastn tool was used for alignment as it checks alignment of nucleotide sequence against nucleotide sequence database.
3. The query sequence was copied onto the space labelled 'Enter Query sequence' in blastn.
4. The database selected was 'Non-redundant protein sequences' (nr).
5. The search was limited to '*Mycobacterium (taxid:1763)*'.
6. All other parameters were set at default.
7. Local alignment command was given.

3.2.2 Blastp

3.2.2.1 Not limited to '*Mycobacterium'*

Procedure:

1. The query sequence (protein) obtained after translating the nucleotide sequence obtained from Random Mutagenesis experiments was taken.
2. blastp tool was used for alignment as it checks alignment of protein sequence against protein sequence database.
3. The query sequence was copied onto the space labelled 'Enter Query sequence' in blastp.
4. The database selected was 'Non-redundant protein sequences' (nr).
5. The alignment was performed against all known organisms.
6. All other parameters were set at default.
7. Local alignment command was given.

3.2.2.2 Limited to '*Mycobacterium (taxid:1763)*'

Procedure:

1. The query sequence (protein) obtained after translating the nucleotide sequence obtained from Random Mutagenesis experiments was taken.
2. blastp tool was used for alignment as it checks alignment of protein sequence against protein sequence database.
3. The query sequence was copied onto the space labelled 'Enter Query sequence' in blastp.
4. The database selected was 'Non-redundant protein sequences' (nr).
5. The search was limited to '*Mycobacterium (taxid:1763)*'.
6. All other parameters were set at default.
7. Local alignment command was given.

3.2.3 Blastx

3.2.3.1 Not limited to '*Mycobacterium*'

Procedure:

1. The query sequence (nucleotide) obtained from sequencing after Random Mutagenesis experiments was taken.
2. blastx was the tool that was used for alignment as it translates the nucleotide sequence to protein sequence and then checks alignment which enables a more extensive search in terms of all possible alignments.

3. The query sequence was copied onto the space meant labelled ‘Enter Query sequence’ in blastx.
4. The database selected was ‘Non-redundant protein sequences’ (nr).
5. The alignment was performed against all known organisms.
   The search was limited to ‘*Mycobacterium (taxid:1763)*’.
6. All other parameters were set at default.
7. Local alignment command was given.

3.2.3.2 Limited to ‘*Mycobacterium (taxid:1763)’*

Procedure:

1. The query sequence (nucleotide) obtained from sequencing after Random Mutagenesis experiments was taken.
2. blastx was the tool that was used for alignment as it translates the nucleotide sequence to protein sequence and then checks alignment which enables a more extensive search in terms of all possible alignments.
3. The query sequence was copied onto the space meant labelled ‘Enter Query sequence’ in blastx.
4. The database selected was ‘Non-redundant protein sequences’ (nr).
5. The search was limited to ‘*Mycobacterium (taxid:1763)*’.
6. All other parameters were set at default.
7. Local alignment command was given.

3.3 ExPASy

ExPASY Translate tool is used to translate nucleotide sequence to amino acid sequence and identify the ORFs in the sequence. An ORF or Open Reading Frame is part of a reading frame that has the ability to be translated. A continuous set of codons, an ORF contains, a start codon (usually AUG) and an end codon (usually UAA, UAG or UGA). Essentially ORFs help in gene prediction. Long ORFs are often used to initially identify candidate protein-coding regions or functional RNA-coding regions in a DNA sequence [37].

Procedure:

1. ExPASy Translate tool was opened in the browser (https://web.expasy.org/translate/).
2. The nucleotide sequence used for study was entered into the section asking for query sequence.
3. Rest of the parameters were set at default.
4. The 'Translate Sequence' command was given.
5. ORFs were obtained from the results.

3.4 Multiple Sequence Alignment (MSA)

Software used: Jalview (stand-alone MSA)

Online Tool used: Tcoffee

A Multiple Sequence Alignment (MSA) is the alignment of three or more biological sequences of protein, DNA or RNA and their comparison with one another in order to factorize the percentage and extent of similarity. This is done by inserting gaps in the residues such residues with common structural positions or ancestral residues are aligned in the same column. Phylogenetic analysis can be subsequently performed from MSA which is the prediction of evolutionary relationships between the sequences based on the amount and location of conserved sequences. Examples of MSA programs are Clustal W, COBALT tool (NCBI), Jalview (Standalone software for MSA editing, visualization and analysis) [38].

Procedure:

1. ORFs obtained from ExPASy results were taken.
2. Proteins from previous literature which might show potential alignment with the obtained ORFs were chosen.
3. Sequences of these proteins were obtained from NCBI database.
4. About five mycobacterial and six non-mycobacterial species sequences were curated from previous literature for MSA with ORFs.
5. The sequences were together pasted in the text box in Jalview software along with an ORF each time.
6. Web service alignment was performed using 'Tcoffee with defaults'
7. 'Zappo' colour coding was selected.

8. Scores of alignment were noted.

3.5 Protein-Protein Interaction Study

The STRING database (Search Tool for the Retrieval of Interacting Genes/Proteins) is a tool to critically integrate and assess protein-protein interactions, be it direct (physical) associations or indirect (functional) associations. It is a web based resource with predicted interactions between about 9.6 million proteins from more than 2000 organisms. STRING curates data from numerous sources, which include public text collections, high throughput experiments, conserved coexpression, and computational prediction methods. Graphical interaction map of protein-protein interaction network, explanation in detail, and ease of use makes STRING a very powerful computational tool [39].

The network of proteins formed in STRING assimilates the network of predicted interactions for a certain group of proteins. The nodes of the network are the proteins while the edges between the nodes represent predicted functional interactions. In *evidence mode*, seven coloured lines are used to denote seven different types of predicted associations between the nodes. The Red line - indicates presence of fusion evidence, Blue line indicates cooccurrence evidence, Black line indicates coexpression evidence, Yellow line indicates textmining evidence, Green line indicates neighborhood evidence, Purple line indicates experimental evidence and the Light blue line indicates database evidence. In the *confidence mode* the thickness of an edge implies the degree of confidence in the prediction of an interaction.

A Red coloured node represents the query protein and the first shell of interactors. The white nodes represent the second shell of interactors. Empty nodes represent proteins of unknown 3D structure, while filled nodes represent proteins whose 3D structure is somewhat predicted or known.

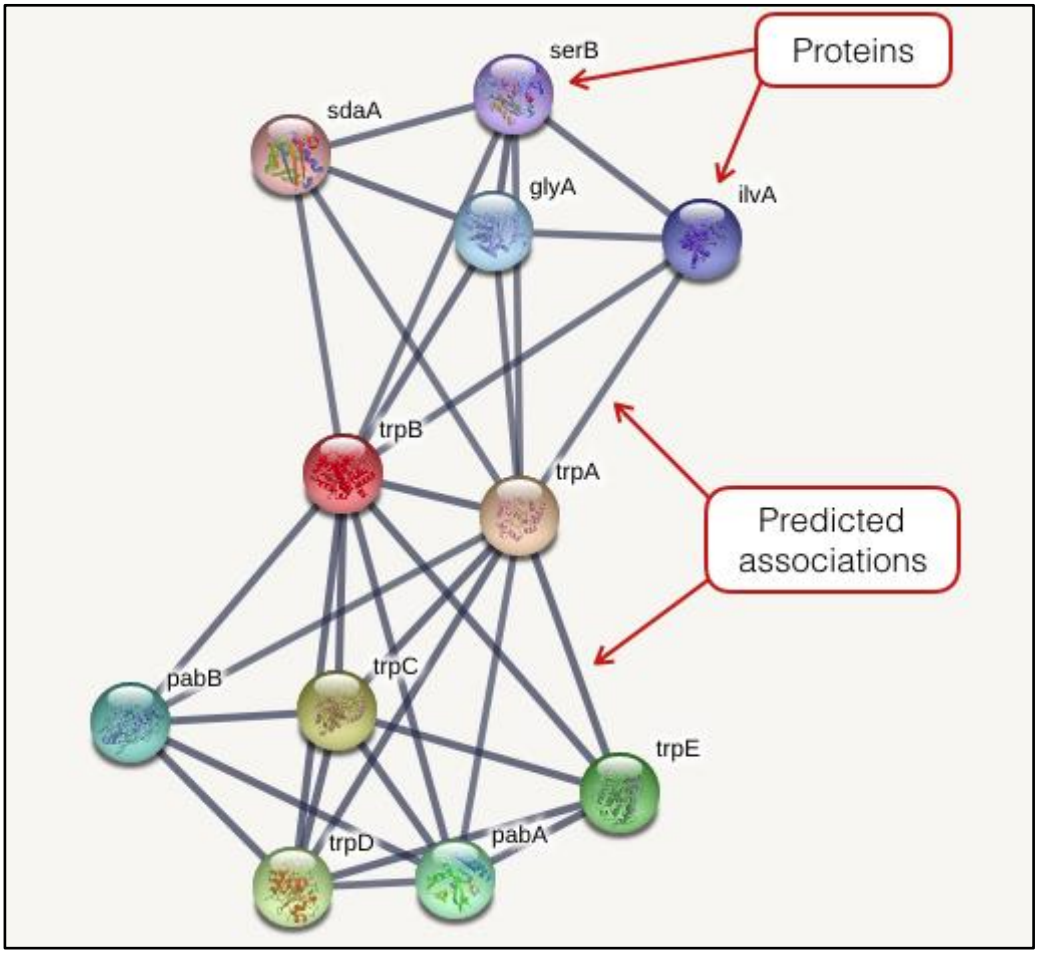

Figure 3.1: Protein predicted functional association network in Confidence mode.

Clicking on a node gives information about the protein from different databases. It also provides links to databases for further details. Clicking on an edge provides a detailed information of the protein-protein interaction.

Besides STRING other major databases are also available and can be used to find protein-protein interactions, such as, IntAct (Molecular Interaction Database), BioGRID (Biological Repository for Interaction Datasets), MINT (Molecular INTeraction Database), BIND (Biomolecular Interaction Network Database), DIP (Database of Interacting Proteins), HPRD (Human Protein Reference Database) and MyBASE (Mycobacterial database).

Procedure:

1. STRING database was opened in the browser (https://string-db.org/).

2. Amino acid sequence i.e. protein sequence obtained from best alignments from blastx of the nucleotide sequence under study was entered in the database.
3. *'Mycobacterium'* was chosen as the organism within which the search for the sequence was to be limited to.
4. The sequence showed a similarity in the proteome of multiple Mycobacterial species.
5. The results of species showing the closest similarity i.e. the ones giving a maximum score and least e-value were selected.
6. A network of proteins linked with our query protein within the selected species was obtained.
7. Proteins in network were studied for virulence by individually searching for their functions and role in Mycobacterial mechanisms.
8. Network of proteins was increased to 31 and all the interacting proteins were studied for a potential function in pathogenicity of the microorganism.

CHAPTER 4

RESULTS

4. RESULTS

4.1 Homological Analysis of Nucleotide Sequence

4.1.1 BLAST

1. blastn (nucleotide alignment) did not give significant results, neither when alignment was checked with all organisms nor when the alignment was limited to *'Mycobacterium (taxid:1763)'.*
2. blastx, which serves to check alignment of a nucleotide sequence after translating it to a protein, with a protein database is used and it enables the most extensive alignment since it takes into account all the possible amino acids that can be formed upon translation.

   Nucleotide sequence used in blastx:

   TGGCCGGCGAATCCATGCTCCGGCCGCCATGGCGGCCGCGGGAAT
   TCGATTGTAAAACGACGGCCAGTGCCAAGCTTGCATGCCTGCAGG
   TCGACTCTAGAGGATCCCCATCGGTAGAATTGCTGAAACGGCAAG
   TGTCCAAGTCACCAAGTATATGGAAGTGGAGCTCTTCCGCTATGT
   CATGCATTTGACCAGCGTGGTCAAGGGACGTTTGCTTCCAGTACT
   CAATGCCATGGATGCCTTGAAAGCTACACTTCCAGCTGGAACAGT
   GTCAGGAGCTCCAAAGATTCGGGCCATGAGACGCATCTATGAACT
   GGAGACGGAAAAACGAGGCGTATACGCAGGAGCAATCGGCTACT
   TGTCTGCGACGGGTGATATGGATTTCGCCATTGCCATCCGAACTA
   TGATTCTCAAAAATCAAACAGCCTATGTGCAGGCTGGGGCAGGG
   ATTGTCTATGATTCTATCGCCCAAAACGAATACCAAGAAACCATT
   AACAAGGCTAAATCTATGACTAGAATTGGAGAACTAAGACCATG
   ATTTTATTGATTGATAACTATGATTCTTTTACTTATAACTTGGCTC
   AATACATTGGGAATTTTGCAGAGGTGCAGGTTTTGAGAAATGATG
   ATCCCAAGCTGTATGAAGAAGCTGAAAAAGCAGATGGTCTGGTTT
   TTTCTCCCGGTCCTGGTTGGCCAATTGATGCCGGAAAGATGGAAG
   ACATGATTCGTGACTTTTCAGGCAAGAAGCCAATTCTAGGGATTT
   GTTTGGGTCATCAAGCCATCGCAGAAGTCTTTGGTGGGAAGCTAG
   GCTTGGCTCCAAAAGTCATGCATGGGAAGCAGAGCCATATCAGCT
   TTGAAGCGCCATCTGTTCTCTATCAAGGCATTGAGGATGGTCGTC
   CAGTCATGCGTTACCACAGTATTTTGATTGAAGAAATGCCAGAAG
   ACTTTGAAGTGACAGCTCGTTCGACTGATGACCAAGCTATTATGG
   GAATTCAACACAAAAGCCTGCCGATTTATGGCTTCCAGTACCATC
   CAGAAAGTATCGGAACGCCAGATGGCTTGTCTTCTATTCGGAATT
   TTATCGAGAAGGTTGTAAAGTGAGGAAACTAGGATGAAAGAGAT
   TATTGAAAAACTAGCAAAATTTGAACATTTATCAGGTGTGGAAAT
   GACGGACGTCATTGAGCGTATCGTAACTGGCCGTCGTTTTACAAT
   CACTAGTGAATTCGCGGCCGCCTGCAGGTCGACTCATATGGGAGA
   GCTCCCAACGCTTGGAAAACACGCCCGG

blastx gave the best alignment with Anthranilate synthase component I (TrpE) and anthranilate synthase component II (TrpG) [*Mycobacterium abscessus subsp. abscessus*] with 94% and 93% identity respectively. Alignment with glutamine amidotransferase of anthranilate synthase or aminodeoxychorismate synthase [*Mycobacterium tuberculosis*] gave an identity of 49% while the alignment with para-aminobenzoate synthase component II [*Mycobacterium tuberculosis*] gave an identity of 45%.

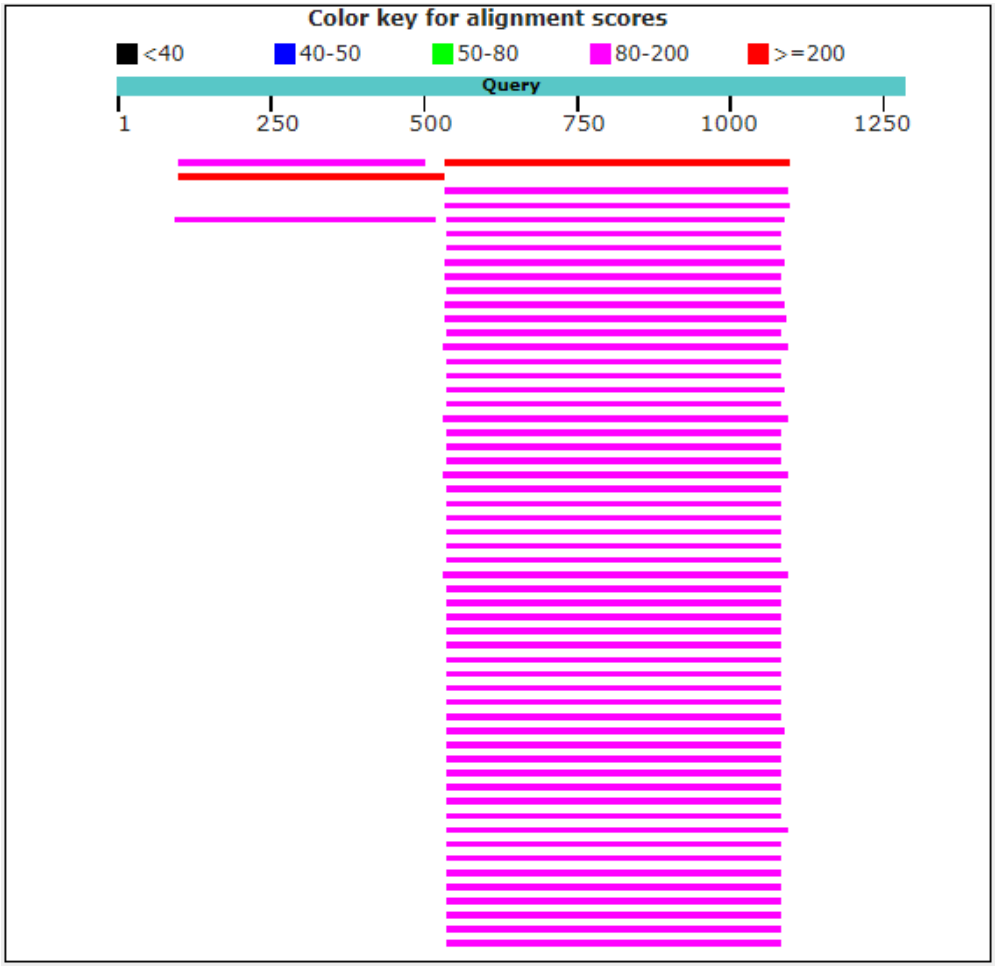

Figure 4.1: Blastx of obtained nucleotide sequence with protein database, with alignment limited to *Mycobacterium* (taxid:1763)

From these results we observed that our nucleotide sequence under study translates to two protein coding domains. The two different genetic regions in the sequence can be seen to probably overlap a little from the image of blastx. This possibility of overlap or gene fusion will be explored later.

4.1.2 ExPASy

1. The nucleotide sequence under study was entered in the ExPASy Translate tool available online.
2. Rest of the parameters were set at default.
3. Translate sequence command was given.
4. The ExPASy Translate tool shows translated protein sequence repeated several times with different ORFs (Open Reading Frames) of different length in each repeated protein sequence.
5. Two of the largest ORFs were taken.

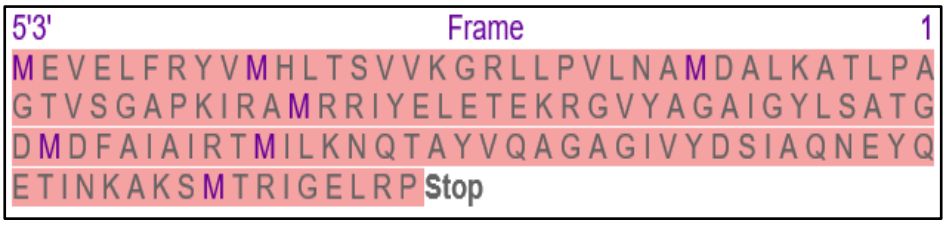

Figure 4.2: Open Reading Frame 1 (ORF #1)

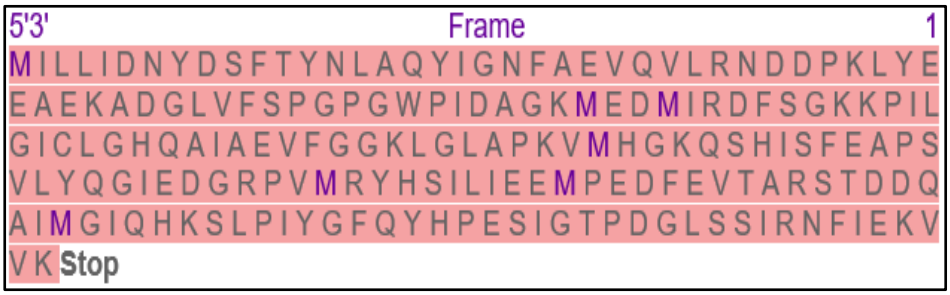

Figure 4.3: Open Reading Frame 2 (ORF #2)

4.1.3 Blastp

1. The two ORF regions obtained from ExPASy were taken.
2. They were entered in blastp software and their alignment studied.
3. blastp checked alignment of protein query with protein database. It did not give results significant to our study when alignment was checked with all organisms.
4. When the alignment was limited to *'Mycobacterium (taxid:1763)'* however significant alignments results were obtained.
5. The first Open Reading Frame showed a 95% identity with anthranilate synthase component I (TrpE) [*Mycobacterium abscessus subsp.*

abscessus], 55% identity with anthranilate synthase subunit I [*Mycobacterium tuberculosis*] and 54% identity with anthranilate synthase component I [*Mycobacterium fortuitum*].

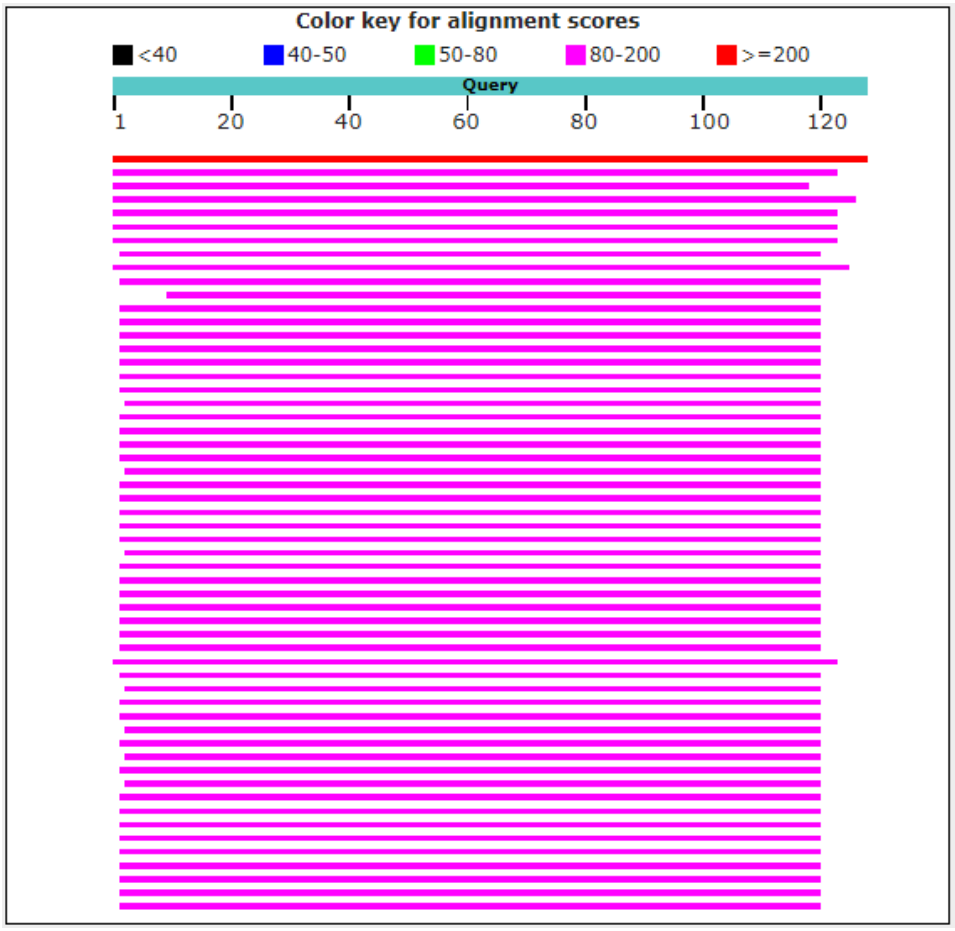

Figure 4.4: Blastp alignment of ORF #1

6. The second ORF showed 93% identity with anthranilate synthase component II TrpG [*Mycobacterium abscessus subsp. abscessus*], 49% identity with glutamine amidotransferase of anthranilate synthase or aminodeoxychorismate synthase [*Mycobacterium tuberculosis*] and 46% identity with para-aminobenzoate synthase component II [*Mycobacterium tuberculosis*].

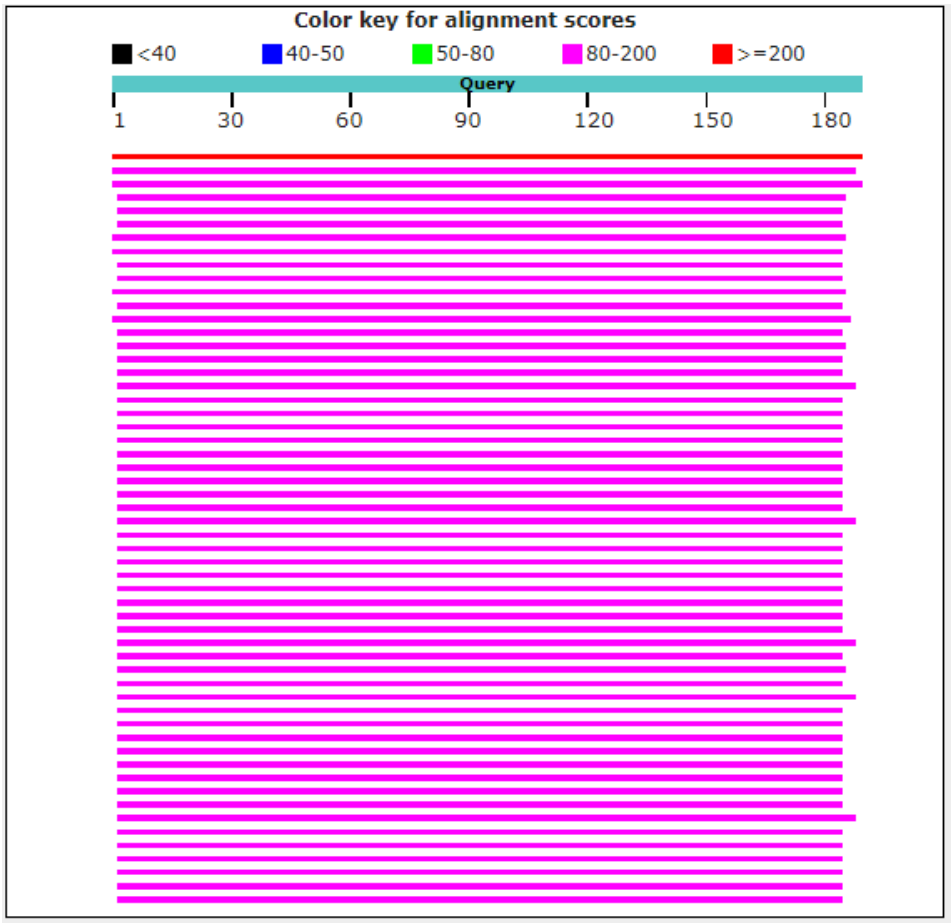

Figure 4.5: Blastp alignment of ORF #2

7. These results confirm that the ORFs obtained from the nucleotide sequence code for the proteins similar to those coded by trpE and trpG respective genes.

4.2 Multiple Sequence Alignment (MSA)

1. In order to explore the possibility of an overlap between proteins coded by trpE and trpG, MSA was performed.
2. Published literature was studied and five mycobacterial and six non-mycobacterial species were curated in which the protein coded by trpE is prominently found and may indicate, by resemblance, the possibility of overlapping genes.
3. These species and their proteins were: SIA63604.1 Anthranilate synthase component I (TrpE) [*Mycobacterium abscessus subsp. abscessus*], CKR88645.1 anthranilate synthase subunit I [*Mycobacterium*

tuberculosis], WP_003885741.1 anthranilate synthase component I, partial [*Mycobacterium fortuitum*], WP_087023174.1 anthranilate synthase component I [*Mycobacterium aurum*], WP_062828592.1 anthranilate synthase component I [*Mycobacterium brisbanense*], GAN27011.1 anthranilate synthase component 1 [*Legionella pneumophila*], OPF97106.1 anthranilate synthase component I [*Rhodopseudomonas palustris*], WP_011028832.1 anthranilate synthase component I [*Streptomyces coelicolor*], WP_061214194.1 anthranilate synthase component I [*Syntrophomonas wolfei*], WP_015032901.1 anthranilate synthase component I [*Streptomyces venezuelae*] and EAI13061.1 unknown [*environmental sequence*] respectively.

4. Following sequence identity (%) was obtained upon performing the MSA of above mentioned proteins coded by trpE and query (with ORFs trpE & trpG).

Table 4.1: Sequence identity (%) obtained upon performing the MSA of mycobacterial and non-mycobacterial proteins coded by trpE with query (containing ORFs of trpE and trpG).

Name of Microorganism	**Sequence identity (%)**
M. abscessus	95.28
M. tuberculosis	47.69
M. fortuitum (partial)	47.76
M. aurum	46.72
M. brisbanense	47.76
Legionella pneumophila	34.91
Rhodopsudomonas palustris	35.86
Streptophomonas coelicolor	52.31
Syntrophomonas wolfei	33.14
Streptomyces venezuele	34.12

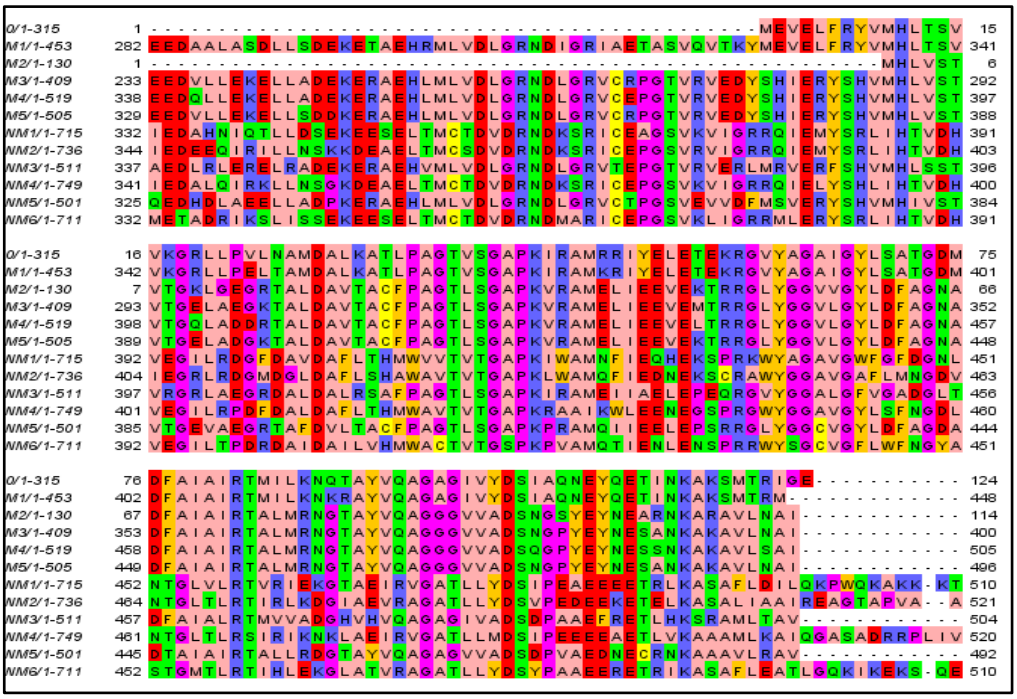

Figure 4.6: Multiple Sequence Alignment of query (ORFs of trpE and trpG) with trpE coded proteins of mycobacterial and non-mycobacterial species.

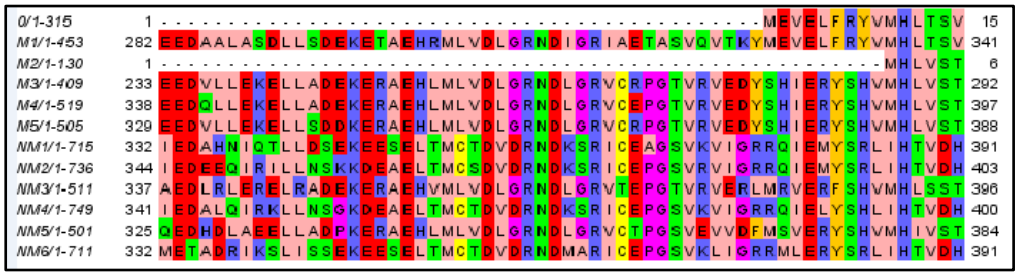

Figure 4.7: Multiple Sequence Alignment of query protein with *M. abscessus* (95.28% identity) starts at 326th amino acid of the trpE coded protein sequence of *M. abscessus*.

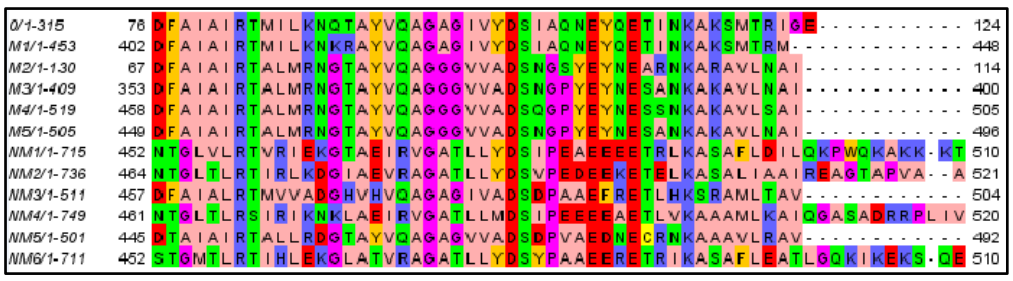

Figure 4.8: Multiple Sequence Alignment of query protein with *M. abscessus* (95.28% identity) ends at 448th amino acid of the trpE coded protein sequence of *M. abscessus*.

5. Similarily, five mycobacterial and six non-mycobacterial species were identified in which the proteins coded by trpG are prominently found and may indicate, by resemblance, the possibility of overlapping genes.

6. These species and their proteins were: SIA63626.1 anthranilate synthase component II TrpG [*Mycobacterium abscessus subsp. abscessus*], WP_085199178.1 aminodeoxychorismate/anthranilate synthase component II [*Mycobacterium fragae*], WP_085254181.1 aminodeoxychorismate/anthranilate synthase component II [*Mycobacterium saskatchewanense*], WP_085168619.1 aminodeoxychorismate/anthranilate synthase component II [*Mycobacterium celatum*], WP_075542580.1 aminodeoxychorismate/anthranilate synthase component II [*Mycobacterium kansasii*], ABQ56380.1 anthranilate synthase component II [*Legionella pneumophila str. Corby*], KIZ38196.1 anthranilate synthase [*Rhodopseudomonas palustris*], ABJ08370.1 anthranilate synthase, component II / anthranilate synthase, component I [*Rhodopseudomonas palustris BisA53*], NP_627427.1 anthranilate synthase component II [*Streptomyces coelicolor A3(2)*], ABI67704.1 anthranilate synthase, component I / anthranilate synthase, component II [*Syntrophomonas wolfei subsp. wolfei str. Goettingen G311*], WP_015034836.1 aminodeoxychorismate/anthranilate synthase component II [*Streptomyces venezuelae*], respectively.

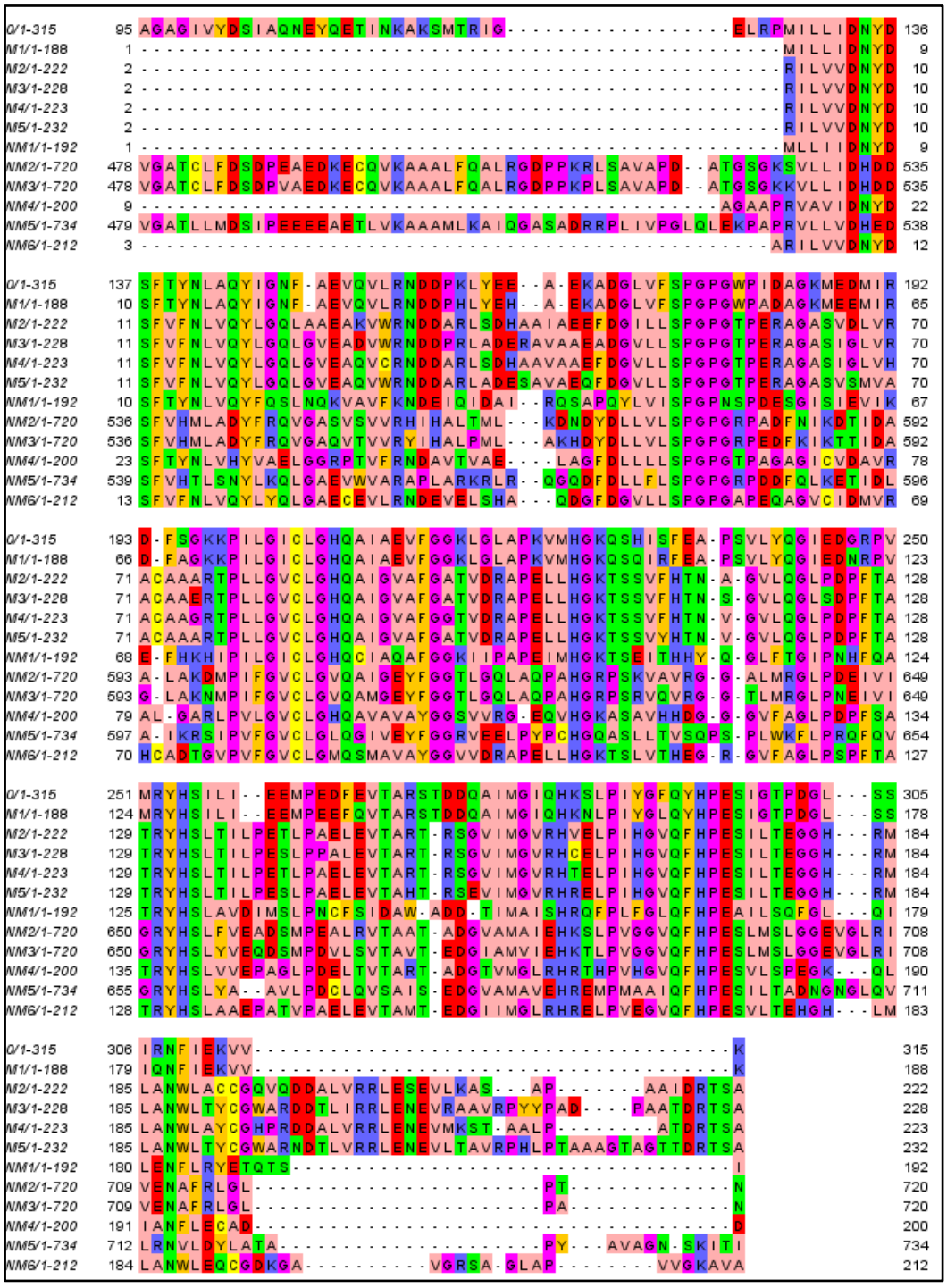

Figure 4.9: Multiple Sequence Alignment of query (ORFs of trpE and trpG) with trpG coded proteins of mycobacterial and non-mycobacterial species.

7. Following sequence identity (%) was obtained upon performing MSA of the above mentioned proteins coded by trpG with query (ORFs of trpE and trpG).

Table 4.2: Sequence identity (%) obtained upon performing the MSA of mycobacterial and non-mycobacterial proteins coded by trpG with query (containing ORFs of trpE & trpG).

Name of Microorganism	Sequence identity (%)
M. abscessus	93.09
M. fragae	44.90
M. saskatchewanense	45.41
M. celatum	45.41
M. kansasii	44.90
Legionella pneumophila	41.67
Rhodopsudomonas palustris	34.43
Rhodopsudomonas palustris BisA54	32.63
Streptophomonas coelicolor	38.73
Syntrophomonas wolfei	34.91
Streptomyces venezuele	40.51

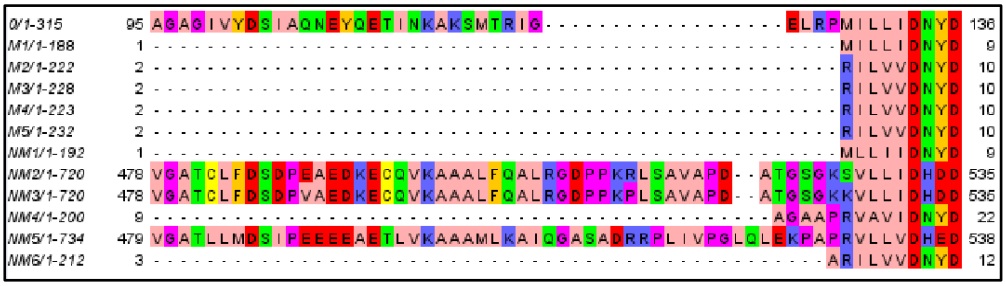

Figure 4.10: Multiple Sequence Alignment of query protein with *M. abscessus* (93.09% identity) starts at 1st amino acid of the trpG coded protein sequence of *M. abscessus* and 123rd amino acid of query (containing ORFs of trpE and trpG).

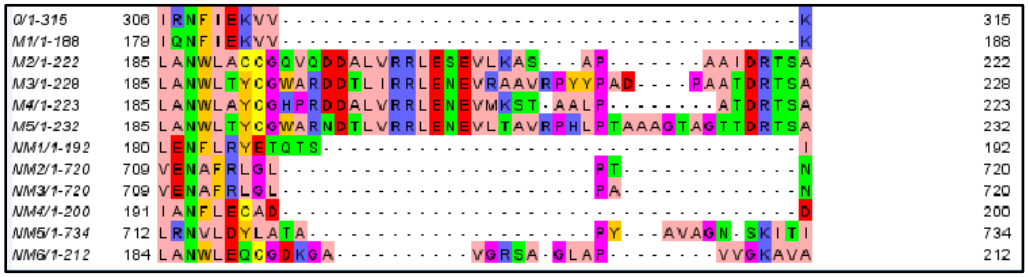

Figure 4.11: Multiple Sequence Alignment of query protein with *M. abscessus* (93.09% identity) ends at last amino acid (188th) of the trpE coded protein sequence of *M. abscessus* and last amino acid of query (containing ORFs of trpE and trpG).

It is clear from the results that different proteins coded by different genes trpE and trpG align at different locations of the query having ORFs of both trpE and trpG. Therefore, significant conclusion could not be ascertained with respect to the fusion or overlap of the two genes.

4.3 Overlapping Genes - evidence

1. In order to further explore the possibility of an overlap between the obtained domains trpE and trpG previous literature was read.
2. Favourable results were obtained in the paper 'Comparative study of overlapping genes in bacteria, with special reference to *Rickettsia prowazekii and Rickettsia conorii'* published in 2005 in the International Journal of Systematic and Evolutionary Biology [40].
3. The research talks about an overlapping sequence of 4 nucleotides i.e. 'ATGA' similar to that of common stretch of sequence between trpE and trpG i.e. 'ATGAT'.
4. This highly close resemblance of four out of five nucleotides in the same order significantly indicates the overlap between trpE and trpG genes.

Therefore, trpE and trpG are genes overlapping over 5 nucleotides which code for two different polypeptides, forming two different proteins.

Following are the nucleotide sequences that codes for the two ORFs obtained containing a common overlapping region of five nucleotides:

ATGGAAGTGGAGCTCTTCCGCTATGTCATGCATTTGACCAGCGTGGT
CAAGGGACGTTTGCTTCCAGTACTCAATGCCATGGATGCCTTGAAAG

CTACACTTCCAGCTGGAACAGTGTCAGGAGCTCCAAAGATTCGGGCC
ATGAGACGCATCTATGAACTGGAGACGGAAAAACGAGGCGTATACG
CAGGAGCAATCGGCTACTTGTCTGCGACGGGTGATATGGATTTCGCC
ATTGCCATCCGAACTATGATTCTCAAAAATCAAACAGCCTATGTGCA
GGCTGGGGCAGGGATTGTCTATGATTCTATCGCCCAAAACGAATACC
AAGAAACCATTAACAAGGCTAAATCTATGACTAGAATTGGAGAACTA
AGACCATGATTTTATTGATTGATAACTATGATTCTTTTACTTATAACT
TGGCTCAATACATTGGGAATTTTGCAGAGGTGCAGGTTTTGAGAAAT
GATGATCCCAAGCTGTATGAAGAAGCTGAAAAAGCAGATGGTCTGGT
TTTTTCTCCCGGTCCTGGTTGGCCAATTGATGCCGGAAAGATGGAAG
ACATGATTCGTGACTTTTCAGGCAAGAAGCCAATTCTAGGGATTTGTT
TGGGTCATCAAGCCATCGCAGAAGTCTTTGGTGGGAAGCTAGGCTTG
GCTCCAAAAGTCATGCATGGGAAGCAGAGCCATATCAGCTTTGAAGC
GCCATCTGTTCTCTATCAAGGCATTGAGGATGGTCGTCCAGTCATGCG
TTACCACAGTATTTTGATTGAAGAAATGCCAGAAGACTTTGAAGTGA
CAGCTCGTTCGACTGATGACCAAGCTATTATGGGAATTCAACACAAA
AGCCTGCCGATTTATGGCTTCCAGTACCATCCAGAAAGTATCGGAAC
GCCAGATGGCTTGTCTTCTATTCGGAATTTTATCGAGAAGGTTGTAAA
GTGAG

It is seen that the two sequences overlap over five nucleotides namely ‘ATGAT’ (yellow highlighted region).

4.4 Protein Interaction Study Results using STRING

4.4.1 STRING

1. The alignments with maximum query coverage were taken from blastx [*M. abscessus*] and were together entered in the STRING database.

   i.e. the following sequence.

MERIIHGDVLSPILAYMRLNGQHKVILESIPREKENARFSILAYNPVFEIKF
ENGVLYQNGQVIDRDPLDFLYEVTHKSQHHSELPFGGGAIGFVGYDMIS
LYEEIGQLPEDTIGTPDMHFFVYESYIVFDHKKEKIHIIEDALYSERSNEEL
EVALDQVLEELKIPAPNEFEDLDVSPLQFRPHIAPQKFEEMVETARDLIR
NGDMFQCVLSQRFSAEVTGNPFDFYRNLRVTNPSNYLYFYDFGDYQIIG
ASPESLVSVKNGIVTTNPIAGTRPRGANDEEDAALASDLLSDEKETAEHR
MLVDLGRNDIGRIAETASVQVTKYMEVELFRYVMHLTSVVKGRLLPEL
TAMDALKATLPAGTVSGAPKIRAMKRIYELETEKRGVYAGAIGYLSATG
DMDFAIAIRTMILKNKRAYVQAGAGIVYDSIAQNEYQETINKAKSMTRM
GELRPMILLIDNYDSFTYNLAQYIGNFAEVQVLRNDDPHLYEHAEKADG
LVFSPGPGWPADAGKMEEMIRDFAGKKPILGICLGHQAIAEVFGGKLGL
APKVMHGKQSQIRFEAPSVLYQGIEDNRPVMRYHSILIEEMPEEFQVTAR
STDDQAIMGIQHKNLPIYGLQYHPESIGTPDGLSSIQNFIEKVVK

2. Following Network of functional interactors was obtained.

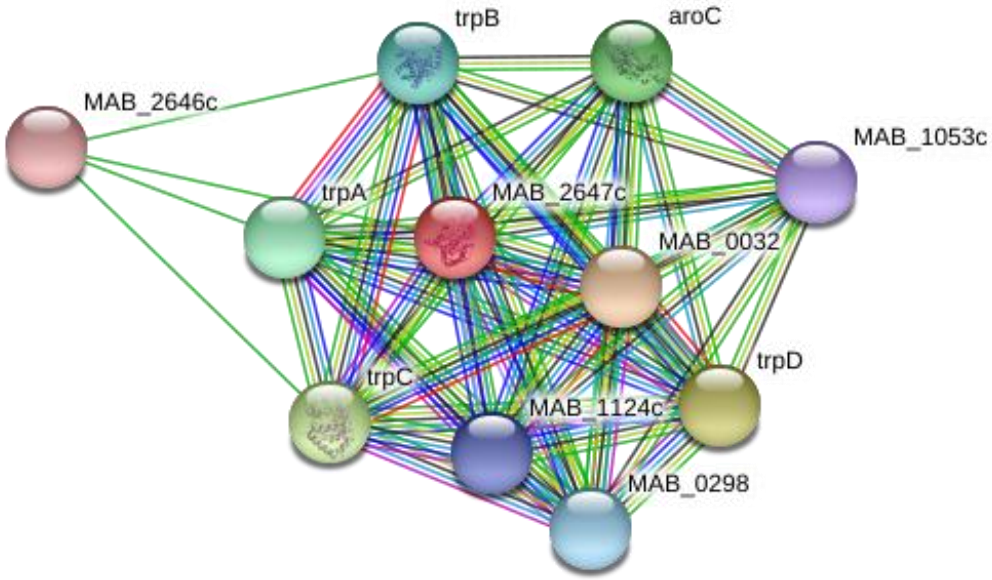

Figure 4.12: STRING Network with first shell of functional protein-protein interactors obtained using 'Evidence' mode of Network edges. Seven coloured edges denote seven different types of predicted interactions. The Red line indicates presence of fusion evidence, Blue line indicates cooccurrence evidence, Black line indicates coexpression evidence, Yellow line indicates textmining evidence, Green line indicates neighborhood evidence, Purple line indicates experimental evidence and the Light blue line indicates database evidence. Red coloured node represents the query protein.

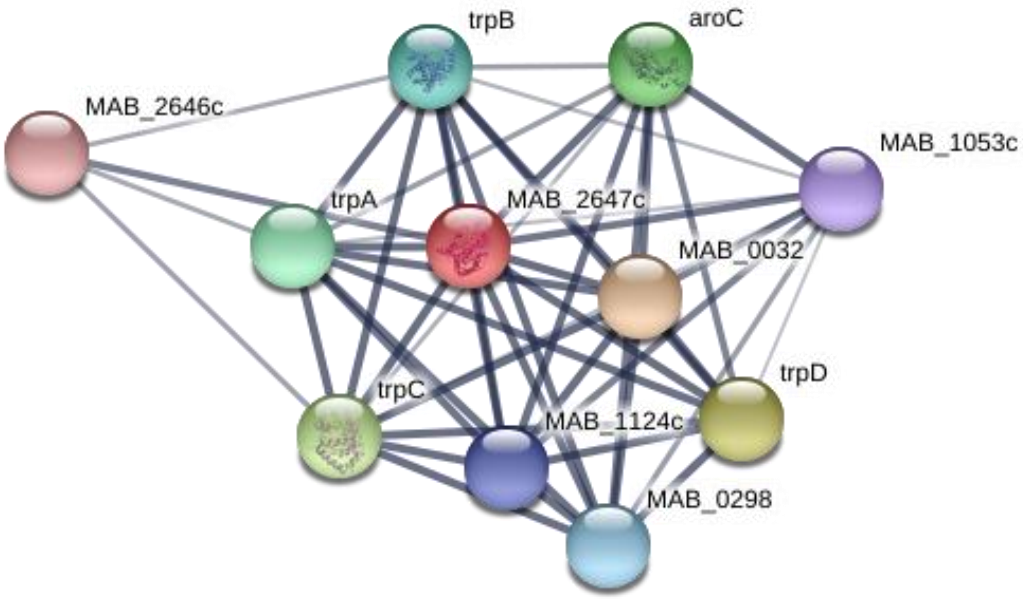

Figure 4.13: STRING Network with first shell of functional protein-protein interactors obtained using 'Confidence' mode of Network edges. Greater the thickness of an edge , greater is the degree of confidence in the prediction of an interaction.

3. The following were the first shell of predicted interactors.

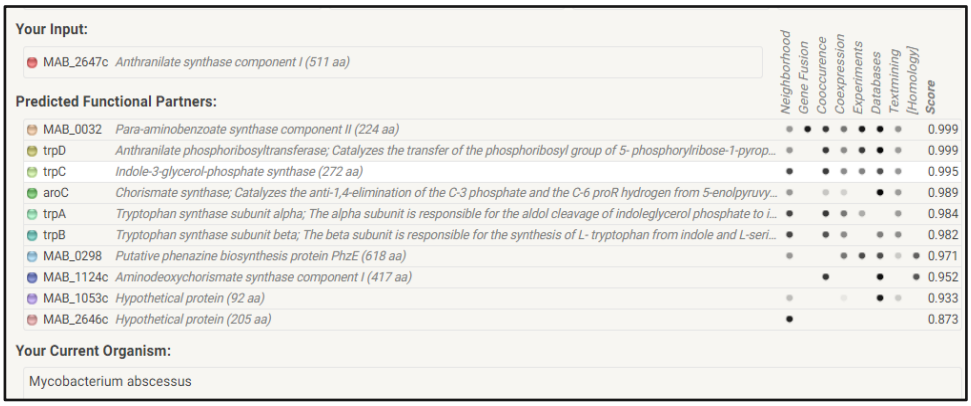

Figure 4.14: Names, interaction type and scores of first shell of interactors.

4. The first 10 interactor genes of our query sequence are MAB_0032, trpD, trpC, aroC, trpA, trpB, MAB_0298, MAB_1124c, MAB_1053c and MAB_2646c.
5. The query sequence i.e. MAB_2647c is identified as the protein 'Anthranilate synthase component I' (i.e. trpE gene).

Table 4.3: Functions and scores of first shell of interacting proteins with query.

Name of Gene	Name of protein	Score of protein-protein interaction with query	Function
MAB_0032	Para-aminobenzoate synthase component II	0.999	Transferase activity
trpD	Anthranilate phosphoribosyltransferase	0.999	Catalytic activity in step 2 of L-tryptophan biosynthesis
trpC	Indole-3-glycerol-phosphate synthase	0.995	Catalytic activity in step 4 of L-tryptophan biosynthesis
aroC	Chorismate synthase	0.989	Catalytic activity in step 7 of chorismate biosynthesis
trpA	Tryptophan synthase subunit alpha	0.984	Catalytic activity in step 5 of L-tryptophan biosynthesis
trpB	Tryptophan synthase subunit beta	0.982	Catalytic activity in step 6 of L-tryptophan biosynthesis
MAB_0298	Putative phenazine biosynthesis protein PhzE	0.971	Involved in biosynthetic process
MAB_1124c	Aminodeoxychorismate synthase component I	0.952	Involved in folic acid-containing compound biosynthetic process
MAB_1053	Hypothetical protein	0.933	Involved in chorismate metabolic process
MAB_2646c	Hypothetical protein	0.873	Unknown (integral component of membrane)

6. trpC, trpD, trpA, trpB are among the first six interactors and are involved in Tryptophan biosynthesis pathway, which is the survival strategy in *Mycobacterium* genus.

4.4.2 MINT database interaction study

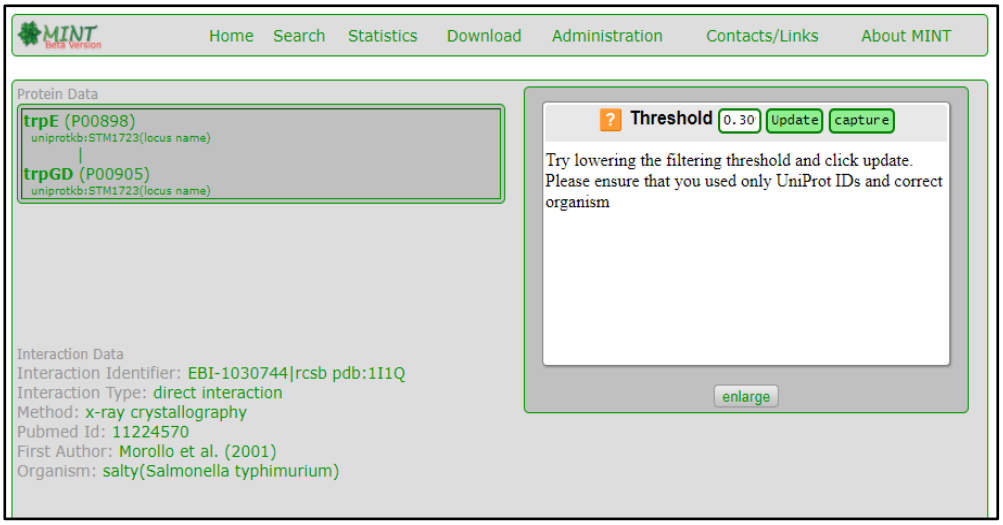

Figure 4.15: Interaction found between trpE and trpGD of *Salmonella typhimurium*

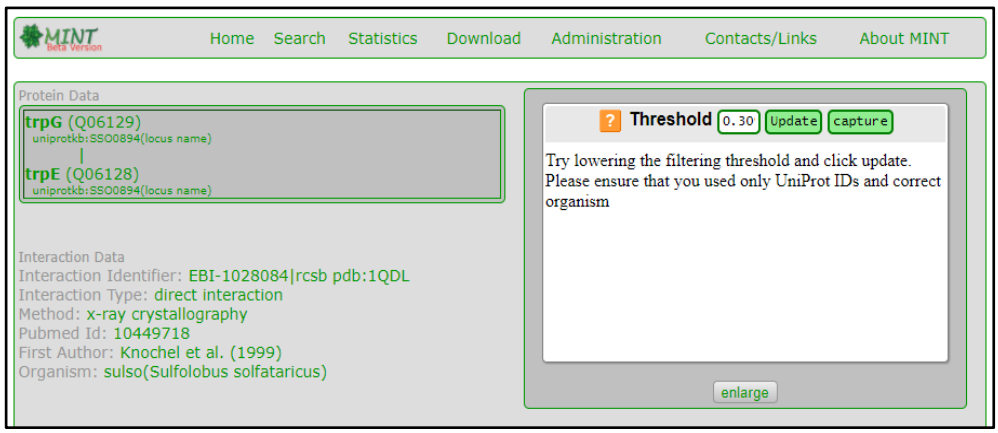

Figure 4.16: Interaction found between trpG and trpE of *Sulfolobus solfataricus*

CHAPTER 5

DISCUSSION & CONCLUSION

5. DISCUSSION & CONCLUSION

Mycobacterium fortuitum is one of the most studied microorganisms due to its role in nosocomial infections and its rapid emergence as a causative agent of disease even in immunocompetent individuals [18]. Virulent genes have been reported to play a significant role in the effective pathogenicity of *M. fortuitum.* Moreover, evidence suggests that virulence genes influence factors or mechanisms involved in protecting mycobacteria from immune system of human host as well as various environmental factors [5]. Therefore there is a need to sufficiently identify major virulent genes and study their mechanisms and functions in order to reduce the incidence of infection and improve its treatment.

In the present study hypothesis use of computational tools was proposed to analyze experimental data obtained from the study for genes involved in virulence of *M. fortuitum.* In this respect nucleotide sequence was obtained by random transposon mutagenesis technique since whole genomic sequence of *M. fortuitum* is not known till now. Computational tools were used to validate role of genes in virulence and also identify interacting partners all of which may function as potential drug targets.

Initially homology studies were performed using nucleotide sequence. In BLAST (Basic Local Alignment Search Tool) blastn did not provide any favorable results i.e. related to *M. fortuitum*, neither when alignment was performed with all organisms nor when it was limited to *Mycobacterium (taxid:1763).* Interestingly, blastx gave significant results when organisms for alignment were limited to *Mycobacterium (taxid:1763).* The sequence was made up of two domains of proteins. Protein sequences were obtained from the best two alignments for further studies.

ExPASy Translate tool is used to translate the original sequence whose attenuation may be indicative of its role in virulence. Upon translation, two major ORF (Open Reading Frames) regions were found. Furthermore, Blastp was performed to clearly state the identity of ORFs. Results of blastp showed that ORF1 belongs to protein coded by trpE gene and ORF2 belongs to protein coded by trpG gene; both

the genes belong to Tryptophan operon and function in the initial steps of Tryptophan biosynthesis.

All of the above were performed on sequences of *M. abscessus* that are evolutionarily similar to *M. fortuitum. M. abscessus* and *M. fortuitum* more often than not are found together or in close proximity in nature. This is because *M. abscessus* is also a NTM with genomic composition similar to *M. fortuitum.* Also, whole genomic sequence of *M. fortuitum* is not known till date and it makes sense to utilize similar genomic sequences of *M. abscessus.*

It was observed from blastx that the two domains recognized in the sequence had a minor common region, maybe over a few nucleotides. This region may either be fused or just overlapped. Fused genes synthesize only one polypeptide giving only one protein sequence, whereas overlapping genes separately synthesize two different polypeptides that function as two different proteins.

To identify if our sequence is overlapped or fused we performed MSA (Multiple Sequence Alignment). About 11 proteins (five Mycobacterial and six Non-Mycobacterial), coded by trpE and trpG each of various microorganisms were taken for alignment with two ORFs obtained from ExPASy. While we succeeded in clearly establishing the identity of ORF1 and ORF2 as anthranilate synthase component I (coded by trpE) and anthranilate synthase compenent II (coded by trpG), respectively, identifying sequence overlap could not be achieved by MSA. However, the fact that two different proteins are coded by the two domains in the nucleotide sequence indicates that the sequences are simply overlapped.

Reverse sequencing the two ORFs provided us with their common nucleotide sequence ‘ATGAT’ and by comparing this with an established ‘overlapping’ sequence ‘ATGA’ [40] we could observe that our sequences are ‘overlapping’.

Further analyzing the *M. abscessus* sequence obtained from top two alignment results of blastx, in STRING provided us with a network of functional interacting partners. Study of these functional interacting partners in the first shell of interactors brought our attention towards those interacting genes and proteins that can be utilized as drug targets.

REFERENCES

[1] A. Aperia, "Membrane transport proteins in health and disease", *Journal of Internal Medicine*, vol. 261, no. 1, pp. 2-4, 2007.

[2] H. Watson, "Biological membranes", *Essays In Biochemistry*, vol. 59, no. 0, pp. 43-69, 2015.

[3] H. Yin and A. Flynn, "Drugging Membrane Protein Interactions", *Annual Review of Biomedical Engineering*, vol. 18, no. 1, pp. 51-76, 2016.

[4] A. Maurya, V. Nag, S. Kant, R. Kushwaha, M. Kumar, A. Singh and T. Dhole, "Prevalence of Nontuberculous Mycobacteria among Extrapulmonary Tuberculosis Cases in Tertiary Care Centers in Northern India", *BioMed Research International*, vol. 2015, pp. 1-5, 2015.

[5] M. Forrellad, L. Klepp, A. Gioffré, J. Sabio y García, H. Morbidoni, M. Santangelo, A. Cataldi and F. Bigi, "Virulence factors of the Mycobacterium tuberculosis complex", *Virulence*, vol. 4, no. 1, pp. 3-66, 2013.

[6] J. Eisenstadt, "Microbiology and classification of mycobacteria", *Clinics in Dermatology*, vol. 13, no. 3, pp. 197-206, 1995.

[7] E. Shitikov, S. Kolchenko, I. Mokrousov, J. Bespyatykh, D. Ischenko, E. Ilina and V. Govorun, "Evolutionary pathway analysis and unified classification of East Asian lineage of Mycobacterium tuberculosis", *Scientific Reports*, vol. 7, no. 1, 2017.

[8] C. Daley and D. Griffith, "Nontuberculous Mycobacterial Infections", *Murray and Nadel's Textbook of Respiratory Medicine*, pp. 629-645.e6, 2016.

[9] I. Porvaznik, I. Solovič and J. Mokrý, "Non-Tuberculous Mycobacteria: Classification, Diagnostics, and Therapy", *Advances in Experimental Medicine and Biology*, pp. 19-25, 2017.

[10] E. Henkle and K. Winthrop, "Nontuberculous Mycobacteria Infections in Immunosuppressed Hosts", *Clinics in Chest Medicine*, vol. 36, no. 1, pp. 91-99, 2015.

[11] S. Tang, D. Lye, R. Jureen, L. Sng and L. Hsu, "Rapidly growing mycobacteria in Singapore, 2006–2011", *Clinical Microbiology and Infection*, vol. 21, no. 3, pp. 236-241, 2015.

[12] S. Bull and A. Doig, "Properties of Protein Drug Target Classes", *PLOS ONE*, vol. 10, no. 3, p. e0117955, 2015.

[13] B. Varghese, Z. Memish, N. Abuljadayel, R. Al-Hakeem, F. Alrabiah and S. Al-Hajoj, "Emergence of Clinically Relevant Non-Tuberculous Mycobacterial Infections in Saudi Arabia", *PLoS Neglected Tropical Diseases*, vol. 7, no. 5, p. e2234, 2013.

[14] J. Umrao, D. Singh, A. Zia, S. Saxena, S. Sarsaiya, S. Singh, J. Khatoon and T. Dhole, "Prevalence and species spectrum of both pulmonary and extrapulmonary nontuberculous mycobacteria isolates at a tertiary care center", *International Journal of Mycobacteriology*, vol. 5, no. 3, pp. 288-293, 2016.

[15] R. Parti, R. Shrivastava, S. Srivastava, A. Subramanian, R. Roy, B. Srivastava and R. Srivastava, "A transposon insertion mutant of Mycobacterium fortuitum attenuated in virulence and persistence in a murine infection model that is complemented by Rv3291c of Mycobacterium tuberculosis", *Microbial Pathogenesis*, vol. 45, no. 5-6, pp. 370-376, 2008.

[16] K. Kannaiyan, "Surgical Site Infections Due to Rapidly Growing Mycobacteria in Puducherry, India", *JOURNAL OF CLINICAL AND DIAGNOSTIC RESEARCH*, 2015.

[17] M. Donohue, "Increasing nontuberculous mycobacteria reporting rates and species diversity identified in clinical laboratory reports", *BMC Infectious Diseases*, vol. 18, no. 1, 2018.

[18] C. Yanofsky, "RNA-based regulation of genes of tryptophan synthesis and degradation, in bacteria", *RNA*, vol. 13, no. 8, pp. 1141-1154, 2007.

[19] T. Parish and N. Stoker, "The common aromatic amino acid biosynthesis pathway is essential in Mycobacterium tuberculosis", *Microbiology*, vol. 148, no. 10, pp. 3069-3077, 2002.

[20] P. Carroll, D. Muttucumaru and T. Parish, "Use of a Tetracycline-Inducible System for Conditional Expression in Mycobacterium tuberculosis and Mycobacterium smegmatis", *Applied and Environmental Microbiology*, vol. 71, no. 6, pp. 3077-3084, 2005.

[21] Y. Zhang, M. Reddy, T. Ioerger, A. Rothchild, V. Dartois, B. Schuster, A. Trauner, D. Wallis, S. Galaviz, C. Huttenhower, J. Sacchettini, S. Behar and E. Rubin, "Tryptophan Biosynthesis Protects Mycobacteria from CD4 T-Cell-Mediated Killing", *Cell*, vol. 155, no. 6, pp. 1296-1308, 2013.

[22] A. Helguera-Repetto, R. Chacon-Salinas, J. Cerna-Cortes, S. Rivera-Gutierrez, V. Ortiz-Navarrete, I. Estrada-Garcia and J. Gonzalez-y-Merchand, "Differential Macrophage Response to Slow- and Fast-Growing Pathogenic Mycobacteria", *BioMed Research International*, vol. 2014, pp. 1-10, 2014.

[23] J. Rodríguez, A. Santos, R. Cremades, J. Rodríguez, E. Garcia-Pachon, M. Ruiz and G. Royo, "Activity of various drugs alone or in combination against Mycobacterium fortuitum", *Journal of Infection and Chemotherapy*, vol. 16, no. 1, pp. 64-67, 2010.

[24] S. Park, G. Suh, M. Chung, H. Kim, O. Kwon, K. Lee, N. Lee and W. Koh, "Clinical significance of Mycobacterium fortuitum isolated from respiratory specimens", *Respiratory Medicine*, vol. 102, no. 3, pp. 437-442, 2008.

[25] M. Adelman and D. Addrizzo-Harris, "Management of nontuberculous mycobacterial pulmonary disease", *Current Opinion in Pulmonary Medicine*, p. 1, 2018.

[26] S. Faria, I. Joao and L. Jordao, "General Overview on Nontuberculous Mycobacteria, Biofilms, and Human Infection", 2018. .

[27] R. Kothavade, R. Dhurat, S. Mishra and U. Kothavade, "Clinical and laboratory aspects of the diagnosis and management of cutaneous and subcutaneous infections caused by rapidly growing mycobacteria", *European Journal of Clinical Microbiology & Infectious Diseases*, vol. 32, no. 2, pp. 161-188, 2012.

[28] S. Okamori, T. Asakura, T. Nishimura, E. Tamizu, M. Ishii, M. Yoshida, H. Fukano, Y. Hayashi, M. Fujita, Y. Hoshino, T. Betsuyaku and N. Hasegawa, "Natural history of Mycobacterium fortuitum pulmonary infection presenting with migratory infiltrates: a case report with microbiological analysis", *BMC Infectious Diseases*, vol. 18, no. 1, 2018.

[29] S. Sousa, M. Bandeira, P. Carvalho, A. Duarte and L. Jordao, "Nontuberculous mycobacteria pathogenesis and biofilm assembly", *International Journal of Mycobacteriology*, vol. 4, no. 1, pp. 36-43, 2015.

[30] H. Jun, K. Jeon, S. Um, O. Kwon, N. Lee and W. Koh, "Nontuberculous mycobacteria isolated during the treatment of pulmonary tuberculosis", *Respiratory Medicine*, vol. 103, no. 12, pp. 1936-1940, 2009.

[31] J. Sangwan, "Mycobacterium fortuitum Peritonitis in a Patient on Continuous Ambulatory Peritoneal Dialysis (CAPD): A Case Report", *JOURNAL OF CLINICAL AND DIAGNOSTIC RESEARCH*, 2013.

[32] G. Lee, H. Lee, T. Kim, J. Kim, K. Sohn and E. Jo, "Mycobacterium fortuitum induces A20 expression that impairs macrophage inflammatory responses", *Pathogens and Disease*, vol. 74, no. 3, p. ftw015, 2016.

[33] K. Plain, K. de Silva, J. Earl, D. Begg, A. Purdie and R. Whittington, "Indoleamine 2,3-Dioxygenase, Tryptophan Catabolism, and Mycobacterium avium subsp. paratuberculosis: a Model for Chronic Mycobacterial Infections", *Infection and Immunity*, vol. 79, no. 9, pp. 3821-3832, 2011.

[34] A. Olive and C. Sassetti, "Metabolic crosstalk between host and pathogen: sensing, adapting and competing", *Nature Reviews Microbiology*, vol. 14, no. 4, pp. 221-234, 2016.

[35] A. de Jesus and T. Allen, "The role of tryptophan side chains in membrane protein anchoring and hydrophobic mismatch", *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1828, no. 2, pp. 864-876, 2013.

[36] S. Altschul, "Basic Local Alignment Search Tool", *Journal of Molecular Biology*, vol. 215, no. 3, pp. 403-410, 1990.

[37] E. Gasteiger, "ExPASy: the proteomics server for in-depth protein knowledge and analysis", *Nucleic Acids Research*, vol. 31, no. 13, pp. 3784-3788, 2003.

[38] C. Notredame, "Recent Evolutions of Multiple Sequence Alignment Algorithms", *PLoS Computational Biology*, vol. 3, no. 8, p. e123, 2007.

[39] C. Mering, "STRING: a database of predicted functional associations between proteins", *Nucleic Acids Research*, vol. 31, no. 1, pp. 258-261, 2003.

[40] K. Sakharkar, "Comparative study of overlapping genes in bacteria, with special reference to Rickettsia prowazekii and Rickettsia conorii", *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY*, vol. 55, no. 3, pp. 1205-1209, 2005.

[41] S. Kumar, N. Joseph, J. Easow and S. Umadevi, "Multifocal keloids associated with Mycobacterium fortuitum following intralesional steroid therapy", *Journal of Laboratory Physicians*, vol. 3, no. 2, p. 127, 2011.

[42] K. JA, "Mycobacterium fortuitum as a cause for nasolacrimal obstruction and granulomatous eyelid disease. - PubMed - NCBI", *Ncbi.nlm.nih.gov*, 1987[Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/3574874?dopt=Abstract.

[43] Z. ZOU, Y. LIU, B. ZHU and P. ZENG, "Direct identification of Mycobacterium abscessus through 16S rDNA sequence analysis and a citrate utilization test: A case report", *Experimental and Therapeutic Medicine*, vol. 8, no. 1, pp. 115-117, 2014.

PUBLICATIONS

1. **Identification of *Mycobaterium fortuitum* virulent membrane genes as potential drug targets**

   Shubham Mittal, Poonam, Rahul Shrivastava*

   Poster presentation at – Inbix'17, organized by BISR, Jaipur from November 7 – 9, 2017.

Identification of *Mycobaterium fortuitum* virulent membrane genes as potential drug targets

Shubham Mittal, Poonam, Rahul Shrivastava*
Email id: mshubham9606@gmail.com
*Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology
Waknaghat, Solan – 173234 (H.P), India

Abstract

Mycobacterium fortuitum is a rapidly growing non tuberculous pathogen attaining relevance in human health as it causes post-surgical infections. Membrane proteins are important in virulence as they play an important roles such as interaction with host cell, transport of biomolecules, and in maintenance of cellular stability. Therefore, the aim of present study was to identify and validate the membrane genes responsible for virulence to have an insight into the pathogenesis of *M. fortuitum*. The strategy applied for the study was random mutagenesis by using plasmid *pRT291* that contains transposon *TnphoA* characterized by inactive alkaline phosphatase gene & kanamycin resistance gene. Fifty mutants were screened *in vivo* for virulence attenuation. One mutant was found to have attenuated virulence which was further subjected to genomic analysis by cloning techniques followed by Bioinformatics analysis. The sequence obtained after genomic analysis was *trpD* gene which was further analysed by using STRING, MINT and BioGRID. It was inferred that two virulent genes *trpE* and *mbtI* have the best interaction with *trpD*. As there is strong interaction of *trpD* with these two virulent membrane genes it can be concluded that the virulence attenuation in the mutant MT721 can be due to the affect of *trpD* inactivation on the function of either of the two virulent genes. Thus, *trpD* can act as a potential drug target against *M. fortuitum* infections. Also, being a part of tryptophan biosynthesis pathway which is absent in humans, further adds up to the fact that it can act as a potential drug target.

Keywords

Mycobacterium fortuitum, Drug targets, STRING database, Anthranilate phosphoribosyl transferase

Background and Rationale

- *Mycobacterium fortuitum* is a rapidly growing nontuberculous mycobacterium.
- Major cause of nosocomial infections.
- Membrane genes play an important role in virulence.
- Hypothetical protein 'Anthranilate phosphoribosyl transferase' (coded by *trpD*), identified by random mutagenesis, is majorly involved in tryptophan biosynthesis.
- Validating virulence of *trpD* gene in *M. fortuitum* by studying interacting genes having a role in virulence.

Aims and Objectives

- Identification of transposon mutant library for virulence attenuation.
- Identification of genes involved in virulence by genomic analysis.
- Validation of identified gene as potential drug target using *in silico* techniques.

Methodology

Random mutagenesis of wild type *M. fortuitum* (ATCC 6841) by electroporation of vector *pRT291* containing *TnphoA*

↓

Selection of mutants done on the basis of alkaline phosphatase activity

↓

In vivo analysis of mutants for virulence characteristics.

↓

Sequencing of inactivated gene by cloning in *pUC19* vector.

↓

Homology study of the gene sequence done using BLAST tool.

↓

Interaction study of gene sequence using different Bioinformatics databases– STRING (Search Tool for Retrieval of Interacting Genes/Proteins), BioGRID (The Biological General Repository for Interaction Datasets) and MINT (the Molecular Interaction database)

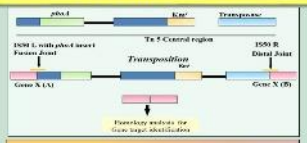

Fig 2. Strategy for Random Mutagenesis

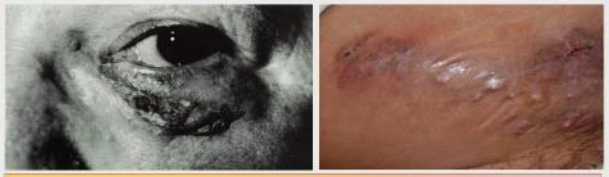

Fig 1. Infections caused by *M. fortuitum*[1]

Results

- The mutant showed 3 log decrease in virulence as compared to wild type over a span of forty days.
- Homologous genes for inactivated *M. fortuitum* gene identified by BLAST were Anthranilate phosphoribosyltransferase (*trpD*) of *Streptococcus oralis* (98% identity) and *M. tuberculosis* (39% identity).
- *trpE* & *mbtI* were found to show best interaction with *trpD* gene with score of 0.999 each and have been 'experimentally determined' previously.
- Analysis through MINT database also showed the interaction of *trpD* gene with *trpE* gene in *Salmonella typhimurium*.
- The interaction study with BIOGRID database did not provide any result regarding interaction of *trpD* with *trpE*.

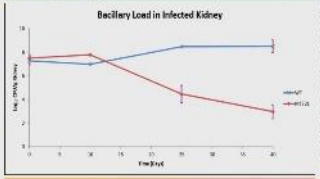

Fig 3. Graph of *In Vivo* analysis

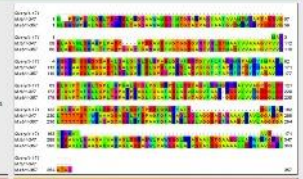

Fig 4. Multiple Sequence Alignment of *trpD* gene with *M. tuberculosis* (39%) and *M. abscessus* (92%)

Fig 5. Results of *trpD* in MINT database

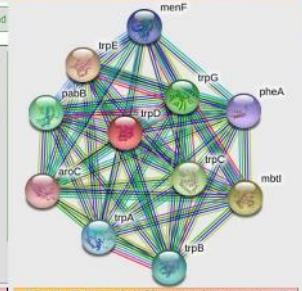

Fig 6. Network of interacting genes of *trpD* in STRING

Conclusions

Random mutagenesis followed by *in vivo* virulence study and genomic analysis identified the gene as *trpD* (codes Anthranilate phosphoribosyltransferase protein), involved in tryptophan biosynthesis. Virulence attenuation due to inactivation of *trpD* gene in mutant indicated the role of this gene in virulence, which was verified by *in silico* interaction study using STRING database resulting in the interaction of *trpD* gene with *trpE* gene (coding for anthranilate synthase component I) and *MbtI* (coding for salicylate synthase). Also the interaction of *trpD* gene was found with *trpE* gene using MINT database.

The results reveal the interdependency of these two genes. It may be hypothesized that the virulence attenuation in the *trpD* mutant of *M. fortuitum* is due to the effect of *trpD* inactivation on *trpE* gene function. The function of *trpE* gene in virulence has been very well known in *M. tuberculosis* which further validates our hypothesis.

Therefore, *trpD* gene have indirect role of in altering the function of *trpE* virulent genes. The results conclude the role of *trpD* gene in virulence which can be used as a potential drug target against *M. fortuitum* infections.

References

1. Tsai, Shih-Chen, et al. "Complicated skin and soft tissue infection with Mycobacterium fortuitum following excision of a sebaceous cyst in Taiwan." *The Journal of Infection in Developing Countries*10.12 (2016): 1357-1361.
2. Zhang, Yanjia J., et al. "Tryptophan biosynthesis protects mycobacteria from CD4 T-cell-mediated killing." *Cell* 155.6 (2013): 1296-1308.

Acknowledgements

The authors are thankful to DST–SERB, IAEC and Jaypee University of Information Technology for providing the resources essential for this research.

Fig 7. Description of interacting genes of *trpD* in STRING

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT
LEARNING RESOURCE CENTER

PLAGIARISM VERIFICATION REPORT

Date: 11 May 2018

Type of Document (Tick): Thesis | M.Tech Dissertation/ Report | B.Tech Project Report ✓ | Paper

Name: SHUBHAM MITTAL **Department:** Biotechnology & Bioinformatics

Enrolment No. 141810 **Registration No.** _____

Phone No. 9805619349 **Email ID.** mshubham9606@gmail.com.

Name of the Supervisor: Dr. Rahul Srivastava ; Dr. Ragothaman M. Yennamalli

Title of the Thesis/Dissertation/Project Report/Paper (In Capital letters):

COMPUTATIONAL ANALYSIS OF MEMBRANE ASSOCIATED GENES OF MYCOBACTERIUM FORTUITUM TO VALIDATE AS DRUG TARGETS

Kindly allow me to avail Turnitin software report for the document mentioned above.

Shubham Mittal

(Signature)

FOR ACCOUNTS DEPARTMENT:

Amount deposited: Rs. 310/- Dated: 11.5.2018 BRN1805/307 Receipt No. .....
(Enclosed payment slip)

(Account Officer)

FOR LRC USE:

The above document was scanned for plagiarism check. The outcome of the same is reported below:

Copy Received on	Report delivered on	Similarity Index in %	Submission Details	
12/05/2018	14/05/2018	7%	Word Counts	9,827
			Character Counts	61,209
			Page counts	59
			File Size	2.06M

Checked by
Name & Signature

Librarian

LIBRARIAN
LEARNING RESOURCE CENTER
Jaypee University of Information Technology
Waknaghat, Distt. Solan (Himachal Pradesh)
Pin Code: 173234