STRUCTURAL INVESTIGATION OF *MYCOBACTERIUM TUBERCULOSIS*

DRUG TARGETS AND THE EVALUATION OF NATURAL PRODUCTS DERIVED

INHIBITOR CANDIDATES

A Dissertation

by

NISHANT DEVIDAS SHETTY

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	James C. Sacchettini
Committee Members,	Kevin Burgess
	Daniel Romo
	Thomas R. Ioerger
Head of Department,	Simon W. North

December 2012

Major Subject: Chemistry

ABSTRACT

This dissertation uses two different approaches to the identification of inhibitors of *Mycobacterium tuberculosis* (*Mtb*) - a structure-based drug discovery approach and a high-throughput screening of natural product libraries based approach. In the structure-based approach the structural characterization of *Mtb* s-adenosylhomocysteine hydrolase (SAHH) enzyme as a drug target using x-ray crystallography is described. Crystal structure of *Mtb* SAHH protein was solved in complex with the substrate adenosine and the product s-adenosylhomocysteine at 1.6 Å and 2.0 Å resolutions respectively. Additionally, crystal structures of *Mtb* SAHH in complex with inhibitors, aristeromycine (ARI), deazaadenosine and 2-fluoroadenosine were also solved at 2.1 Å, 2.2 Å and 2.4 Å resolutions respectively. The complex structure with ARI is the first structure reported and confirms the proposed type-I mechanism based inhibition of *Mtb* SAHH. Differences in the active site of *Mtb* SAHH and human SAHH are identified and the design of lead molecules selective towards the *Mtb* SAHH is described using the fragment-based lead identification method. The structural characterization of a nitrogen regulatory *Mtb* PII protein is also described. The crystal structure of *Mtb* PII protein in the apo form and adenosine triphosphate bound form was solved to 1.4 Å and 2.4 Å resolutions respectively. The crystal structures suggest an alternate annotation of the protein as GlnK and also provide insights into the mechanism of action of the *Mtb* PII protein. The *Mtb* PII protein plays a versatile role in the nitrogen regulatory pathway of the microorganism and represents a potential drug target in *Mtb*.

Through the alternate approach to drug discovery involving the screening of natural products for whole-cell bactericidal activity a novel natural product inhibitor of *Mycobacterium tuberculosis* and *Mycobacterium smegmatis* was isolated, purified and characterized. Challenges encountered in the large scale data processing involving high-throughput screening and high performance liquid chromatography (HPLC) / mass spectrometric analysis in terms of prioritizing the crude extracts, the HPLC fractions and the masses corresponding to the compounds of interest are listed and methods for data reduction and efficient data analysis are presented. The successful identification of a novel natural product with inhibitory activity towards the human and yeast proteasome in an *in-vitro* enzyme assay is also described. The novel polyphenolic natural product was discovered through the screening of crude extracts in a proteasome targeted *in vitro* enzyme assay followed by activity based fractionation, isolation, purification and structure elucidation using analytical techniques. A technique for the chemical derivatization of a mixture of unknown secondary metabolites in crude extracts is also described, which can potentially increase the existing diversity of natural product libraries used in high-throughput screening.

DEDICATION

To Ajja and Amma

NOMENCLATURE

TB	Tuberculosis
BC	Before Christ
BCE	Before the Common Era
WHO	World Health Organisation
Mtb	*Mycobacterium tuberculosis*
BCG	Bacilli Calmette-Guérin
NIAID	National Institute of Allergy and Infectious Diseases
CDC	Center for Disease Control
TBSGC	Tuberculosis Structural Genomics Consortium
XDR	Extensively Drug Resistant
M. smeg	*Mycobacterium smegmatis*
E. coli	*Escherichia coli*
Pf	*Plasmodium falciparum*
M. jann	*Methanococcus jannaschii*
SAHH	S-adenosyl-L-homocysteine Hydrolase
SAH	S-adenosylhomocysteine
NAD^+	Nicotinamide Adenine Dinucleotide (oxidized form)
NADH	Nicotinamide Adenine Dinucleotide (reduced form)
SAM	S-adenosylmethionine
HCY	Homocysteine

ADO	Adenosine
2OG	2-Oxoglutarate
Gln	Glutamine
GS	Glutamine Synthetase
GOGAT	Glutamine: 2-Oxoglutarate Aminotransferase
Glu	Glutamate / Glutamic acid
AmtB	Ammonium Transporter protein
Lys	Lysin
Asp	Aspartate / Aspartic acid
Leu	Leucin
Thr	Threonine
GlnE / Atase	Adenylyl Transferase
GlnD	Uridylyl Transferase
BME	β – mercaptoethanol
HTS	High-Throughput Screening
ARI	Aristeromycin
DZA	3-Deazaadenosine
2FA	2-Fluoroadenosine
PCR	Polymerase Chain Reaction
RMSD	Root Mean Square Deviation
PDB	Protein Data Bank
His	Histidine

Gly	Glycine
Ala	Alanine
Arg	Arginine
Asn	Asparagine
Cys	Cystine
Ile	Isoleucine
Met	Methionine
Phe	Phenylalanine
Pro	Proline
Ser	Serine
Trp	Tryptophan
Tyr	Tyrosine
Val	Valine
ETA	Ethylthioadenosine
OH	Hydroxyl
SDS PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
EDTA	Ethylenediaminetetraacetic Acid
PEG	Polyethylene Glycol
DSF	Differential Scanning Fluorimetry
STD NMR	Saturation Transfer Difference Nuclear Magnetic Resonance
ATP	Adenosine Triphosphate
ADP	Adenosine Diphosphate

S. griseus	*Streptomyces griseus*
DMSO	Dimethyl Sulfoxide
HPLC	High Performance Liquid Chromatography
LC	Liquid Chromatography
UV	Ultraviolet
PDA	Photo Diode Array
FA	Formic Acid
MS	Mass Spectrometry
m/z	Mass to Charge Ratio
HRMS	High Resolution Mass Spectrometry
TOF	Time-of-Flight
Q-TOF	Quadrupole Time-of-Flight
DNP	Dictionary of Natural Products
NMR	Nuclear Magnetic Resonance
IC_{50}	50 % Inhibitory Concentration
MIC	Miminal Inhibitory Concentration
DTNB	5,5'-dithiobis-(2-nitrobenzoic acid) / Ellman's reagent
ESI	Electrospray Ionization
DEPT	Distortionless Enhancement by Polarization Transfer
SFC	Supercritical Fluid Chromatography

TABLE OF CONTENTS

LIST OF FIGURES

xv

LIST OF TABLES

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

The History of Tuberculosis

Tuberculosis (TB) is the most widely recognized and mortal diseases known to mankind. Throughout history TB has been responsible for more deaths than the plague, leprosy or HIV with an estimated 1 billion fatalities due to TB in the past two centuries [1]. TB has aptly been labeled as the "Captain of all these men of death" [2]. TB is also one of the most ancient diseases known to affect humans. Lesions characteristic of TB in the skeletal remains of humans in the eastern Mediterranean from 7000 BC showed that prehistoric humans suffered from the disease. Many early Egyptian mummies from 3000 – 2400 BCE have been found with spinal lesions that are characteristic of spinal tuberculosis (also called Pott's disease) [3]. Hence it is certain that tuberculosis was widely prevalent in predynastic and early dynastic Egypt. Ancient Indians, in 2000 BCE described a disease called "Yakshma", the causes of which included over-fatigue, sorrow, fasting and pregnancy. Today we know that fatigue, depression and pregnancy can all lead to a weakening of the immune system thus making an individual more susceptible to active tuberculosis infection. Symptoms of the disease were similar to those found in consumptive patients in modern times.

Throughout the middle ages the disease continued to wreak havoc in the European countries. According to John Locke, a well-known physician, twenty percent of all deaths in London were due to tuberculosis in 1667 [4]. About 1000 to 1250 new

cases of TB per 100,000 people were reported every year in London in the seventeenth century [5]. Early European explorers to Africa, Americas and the Arctic are believed to have carried the disease with them and infected the indigenous population who had very little natural resistance to the disease. South East Asia was also not left untouched as TB followed the European colonizers. By 1900 approximately 10 % of all the deaths reported in Indian cities were due to TB.

Over the centuries, before the real cause of TB was yet to be discovered, the disease was known by many different names, like scrofula (tuberculosis of the lymph glands of the neck), *tabes*, consumption or *phthisis* (Greek for consumption), because the patients displayed characteristic traits of TB like losing excessive body weight and getting wasted. It has been called the white plague due of the pale appearance of the patient and the scale of the epidemic. Symptoms of the disease have been well documented in medical literature since the time of Hippocrates in 460 BCE. Fever, night sweats, shortness of breath, pain in the side or shoulder, excessive coughing with blood stained sputum are common symptoms associated with pulmonary TB.

Franciscus Delaboe Sylvius coined the word "tubercles" for the first time in 1679 in his book Opera medica, due to the presence of characteristic nodules that were found in the bodily organs, especially the lungs, of the victims of consumption. These tubercles were described to resemble millet seeds and would sometimes be hard as bones and at other times soft with a cheese like appearance. This similarity with millet seeds would later lead to the term 'miliary' tuberculosis, to describe a generalized infection throughout the body. In other instances the disease would be localized to certain parts of

the body, like the membranes surrounding the brain (meningeal tuberculosis), kidney tissues (renal tuberculosis), the spine (Pott's disease or spinal tuberculosis) and lupus (tuberculosis of the skin). Since the disease could infect virtually any part of the body, physicians of the seventeenth and eighteenth century were inclined to regard these symptoms as signs of unrelated diseases and even went to the extent of separating one type of infection from the other. For example, it was hard to imagine that the scrofulous glands in the neck were caused due to the same pathological condition that also caused the ulcerations in the lungs. The famous French physician Laënnec, through his studies on hundreds of consumptive cadavers, was able to postulate that the infiltrations, tubercles and cavities caused by the ulcers were all expressions of a single disease. Shortly, in 1839, J.L. Schönlein suggested that the disease officially be called as tuberculosis and used to cover all forms of phthisis.

The cause of the disease, however, was still unknown and debated until the eighteenth century. Indeed many prominent physicians of the time discounted the very contagious nature of phthisis. This was attributed to the fact that a person exposed to the sputum or infected tissue of another tuberculous patient did not always immediately develop the symptoms characteristic of the disease. We now know that the disease causing organisms may lay dormant inside a person's macrophages for many years before flaring up into its active form. Not surprisingly the more popular belief of the time was that the disease was passed on genetically from parents to their offspring, making some population predisposed to it. This was derived by the observation that if one or both parents in a household were tuberculous, their children were most likely to

develop the disease in their lifetimes. Indeed successive generations of a tuberculous person have been found to be particularly vulnerable to the disease given credence to the theory that TB must be hereditary. It took many years and much effort by physicians like William Budd, Thomas Young and Jean-Antoine Villemin to establish the contagion theory of tuberculosis among the medical professionals. Jean-Antoine Villemin, a French army surgeon, is credited for being able to demonstrate for the first time that phthisis could be transmitted from human to cattle and other animals. Villemin went on to show that the material taken from the scrofula could be used to induce ulceration and cavity formation in the lungs of another animal, thus proving that scrofula and tuberculosis were etiologically the same. With advances in microbiological techniques the germ theory slowly gained prominence in the world of medicine. It was back in 1722 that the English physician, Benjamin Marten had for the first time proposed the germ theory when he said that diseases were caused by … "some certain species of *Animalculae* or wonderfully minute living creatures that, by their peculiar shape or disagreeable parts are inimicable to our nature;" But limited by the insufficient technological advancement of his time, Marten was unable to experimentally demonstrate the relevance of his germ theory and wisely left it to the future generation of scientists to "… carry the theory much farther than I have done and … bring it to absolute demonstration in an extensive degree" . It was not until one hundred and sixty years later in 1862 that Robert Koch used the germ theory to actually demonstrate the constant presence of bacteria in the tuberculous tissues of human as well as livestock. He also showed that it was possible to cultivate the bacilli in a pure culture and produce

the disease in other normal animals by inoculating the culture into them. Koch's work was of great significance because man was finally able to confront the real enemy he was up against.

The next big step was to detect the presence of the infection in patients. Since the symptoms of TB appear only when the disease has progressed to an advanced stage, detecting the infection early was crucial to treatment. Incidentally all the currently available detection methods were discovered more or less by accident. In attempting to find a cure against TB, Koch derived a glycerine extract of the tubercle bacillus and named it tuberculin. While this discovery was widely publicized, it was immediately evident that tuberculin was not a cure against TB. However it found use as a test to determine the presence of tuberculosis infection in patients. Today the Mantoux skin test used worldwide for the detection of tuberculosis infection consists of a purified protein derivative (ppd) of the tuberculin. The accidental discovery of X rays by Roentgen in 1895 also provided a way to look directly into the lungs of a patient to detect the tuberculous lesions long before the symptoms appear in the patient.

Cause of TB and Early Treatment Methods

After it was identified as a single disease in the 1820s, the disease got its name as 'tuberculosis' in 1839 by Dr. J. L. Schönlein. The bacillus causing TB was first identified by Robert Koch in 1882 as tubercle bacillus and later termed *Mycobacterium tuberculosis*. Infection of the lungs (pulmonary tuberculosis) is the most common

infection site in the human body; however *Mtb* can infect virtually any part of the body including the spine, brain and skin.

Consumptives were recommended a healthy diet, lots of fluid, physical exercise and fresh air. During the Middle Ages many European consumptives found relief when they removed themselves from the cold and damp English weather to warm, bright and sunny climate of the Mediterranean countries. One superstitious yet popular treatment method in the Middle Ages was to be touched by a king. The French and English Kings were believed to possess a divine healing power to cure people with scrofulous glands in the neck. This ritualistic healing process was performed in large ceremonies, where the kings touched the patients mostly suffering from scrofula. While the ritual was started in the fifth century, it continued till as late as the eighteenth century by Queen Anne. Scrofula, being an indolent form of the disease, usually resolved by itself, leading credence to the popularity of the phrase - "royal touch". Cod-liver oil was first prescribed by Dr. C. J. Blasius Williams and since then has been widely recommended by physicians as a dietary supplement of tuberculosis patients [6]. The patients frequently sought opium to calm the continuous coughing and to ease the diarrhea and pain caused by intestinal tuberculosis. Studies conducted in the early twentieth century showed that animal protein consumed in the form of meat offered enhanced resistance against active tuberculosis [7]. While this claim has not been substantially proved or disproved, it is generally agreed upon that poor nutrition contributes to greater susceptibility to tuberculosis. In 1822, a physician named James Carson suggested the use of artificial pneumothorax as a surgical procedure to treat pulmonary consumption. The procedure

involved introducing air into the chest cavity by making incisions into the thorax on the infected side of the lungs. This allowed the collapse of the affected lobes thereby healing the patient. Other similar methods like thoracoplasty [8] and pneumoperitoneum [9] were soon developed based on the same principle of collapsing the infected lungs.

Many physicians in the eighteenth and early nineteenth century believed in the doctrine that pure air, pure water and pure food that are essential for a healthy living were somehow also curative to consumptive patients. Doctors were recommending long walks and horse rides in the open air to patients suffering from consumption. However, it was not until 1854 that a German physician Hermann Brehmer started and popularized the concept of sanatorium. Brehmer, a student of J. L. Schönlein opened the first sanatorium in Görbersdorf for treating tuberculous patients. Others were soon established all over Europe. Edward Livingston Trudeau, a physician and himself a tuberculosis patient established the first sanatoriums in America in 1882 after he read about the success of Brehmer's sanatorium treatment in Germany. Treatment methods in these sanatoriums consisted of nothing more than good food, fresh air and plenty of rest. Other prevention and treatment methods of tuberculosis included quarantine, good sanitary practices and exposure to fresh air.

Modern Treatment Methods

Based on the success of vaccines against the small pox and rabies virus, it was believed that a similar vaccine could be developed to immunize people against tuberculosis. In 1890, Koch prepared a glycerin extract of the tubercle bacilli and used it

to inject into guinea pigs and people, including himself. The extract, which he called "tuberculin", was tolerated well by individuals who were not previously infected by the tubercle bacilli but elicited a strong allergic reaction in people who were previously exposed to infection. Although tuberculin was heralded as a breakthrough in the fight against tuberculosis by the scientific community, it was soon evident that tuberculin was not the 'magic bullet' that everyone was looking for. It killed more people than it cured. Moreover it did not immunize against a future infection. Although Koch's work fell into discredit as a touted cure for tuberculosis, tuberculin is still in use in a modified form as a method for the early diagnosis of tuberculosis infection. Undaunted by the failure of tuberculin to cure or immunize against TB, other microbiologists continued their work on developing a vaccine against TB. Leon Charles Albert Calmette and his assistant Jean Marie Camille Guérin started working on a tuberculosis vaccine in 1900. They isolated a strain of bovine tubercle bacilli and cultured it in a medium containing ox bile. After many subsequent cultures in the media a change in the morphology of the bacillus was noted. The colonies containing the microorganism were smoother and smaller than those containing the virulent strains of the same microorganisms. When they inoculated guinea pigs and rabbits with this culture, the animals did not get sick. Therefore Calmette and Guérin concluded that they had achieved a strain of the tubercle bacilli of low virulence. After reculturing the vaccine strain over two hundred times they observed that the bacillus was unable to produce the disease in cows, horses and monkeys. They called it the *Bacilli Bilié Calmette-Guérin* strain which later became the BCG vaccine. On July 18, 1921, the vaccine was administered for the first time to a newborn baby at Hôspital

8

charité in Paris. The child did not develop any signs of TB and was healthy for the rest of its life. This event marked the beginning of man's first successful attempt to bring the spread of TB under control. Although the effectiveness of TB varied from country to country, BCG gained increasing acceptance in the world and today it is generally accepted as a good immunization technique against TB. Back in the twentieth century, in the absence of any other known cure against TB, BCG stood out as mankind's only weapon against the onslaught of a growing TB epidemic.

Since the beginning of the twentieth century investigations had started into the discovery and development of antibacterial chemotherapeutic agents. Arsphenamine, an arsenic containing small molecule synthesized in 1909 was the first antibacterial drug, and was used in the treatment of syphilis. Promin, a sulfur containing drug, synthesized in 1940 by W. H. Feldman was reported to have a mild activity against tuberculosis in guinea pigs [10,11]. Another compound called promizole was also synthesized and had similar activity but less toxicity than promin [12]. While synthetic molecules have been used in antibacterial chemotherapy, nature has provided us with some of the best antimicrobials known. So it was also in the case of tuberculosis that nature came to the rescue. Streptomycin, a natural product isolated from soil bacteria that belongs to the genus *Actinomycetes*, ultimately proved to be highly effective in killing the pathogen. Albert Schatz, a graduate student and his mentor, Selman Abraham Waksman spent many years isolating secondary metabolites from actinomycetes and studying their antitubercular activities. Streptomycin was isolated from a strain of *actinomyces griseus* (later called *Streptomyces griseus*) [13]. While it was found to have a significantly higher

activity against tubercle bacilli, it was also much less toxic compared to other known antitubercular drugs. Soon other antibiotics like oxytetracycline [14], viomycin [15] and cycloserine [16], derived from different species of *Streptomyces* were found to be active against tuberculosis. On the other hand, research on the development of synthetic antibiotics continued and a number of drugs like, para aminosalicylic acid [17], isonicotinic acid hydrazide (isoniazid) [18,19] and pyrazinamide, with good antitubercular activities were reported. Other drugs like rifampin, kanamycin and ethambutol were subsequently discovered. It was also observed that different TB drugs taken together were more effective than single drugs by themselves due to synergy [20,21].

Today drug treatment against TB consists of a cocktail of drugs taken over a period of time ranging from six months to one year. Ethambutol, isoniazid, pyrazinamide and rifampin are considered as first line of drugs as these are the most effective against TB (*Source – National Institute of Allergy and Infectious Diseases (NIAID)*). In many countries a second and third line of drugs are also administered to TB patients. The second-line drugs include aminoglycosides like amikacin and kanamycin, polypeptides like capreomycin, viomycin and enviomycin, fluoroquinolones like ciprofloxacin, levofloxacin and moxifloxacin, and thioamides like ethionamide and prothionamide. A third-line of drugs may consist of rifabutin, macrolides like clarithromycin, linezolid, thioacetazone, thioridazine, arginine, vitamin D and R207910. The second and third-line drugs are used either because these drugs are not as effective as the first line of drugs against *Mtb*, they are more toxic compared to the first line of drugs or because the first line of drugs are not easily available in these countries. Resistance of *Mycobacterium*

tuberculosis to the first line drugs may also prompt the usage of second-line or third-line of drugs (*NIAID*).

Even though drug treatment exists, eradication of this 'curable disease' has been like a 'mirage'. The disease has proved to be more than a match for all the drugs approved for treatment for various reasons. There are several reasons for the failure of drug treatment regimens. One of the reasons is the resurgence of drug resistant strains of the bacteria. Natural mutations exist in any random population of microorganisms that render a small population of the same resistant to the drugs. This is especially true with *Mtb*, which is known for its high mutation rate. Some drugs were more prone to develop drug resistance in the patients than others. For example, streptomycin, when it was discovered, was the best form of treatment against TB, however resistance to streptomycin soon began to appear and it had to be taken out of the first-line of preferred drugs. Poor patient compliance is another factor that has contributed to the rise in drug resistant strains of the bacteria. TB chemotherapy involves lengthy treatment regimens of six to twelve months. A lack of education about the mode of action of the antibiotics leads the patients to stop their treatment as soon as they start feeling better. High costs associated with the treatment also lead to a failure to comply with the treatment regimen in poor and developing countries. Since the antibiotic course is stopped abruptly, the pathogenic microorganisms see a suboptimal dose of the drugs and the treatment is unsuccessful in killing off the entire population of the microorganism. The surviving drug resistant pathogens begin to multiply in number and cause a drug resistant form of infection. Moreover, most of the known drugs against TB attack only the actively

replicating *Mtb* cells, but are unable to reach and kill the persistent *Mtb* that survives in the macrophages. From a drug development perspective it is difficult to design drugs that can enter the macrophages as well as penetrate the thick lipid layers of *Mtb*.

Today, the Center for Disease Control (CDC) lists TB as one of the world's most deadly diseases with about 1.4 million TB related deaths per year worldwide. About 9 million people get infected with the tubercle bacillus every year (*Source – CDC*). There are two forms of the infection – a latent infection and an active disease state. When an individual gets infected with *Mtb*, there is about one in ten chance that the person will develop the active form of the disease over the course of his lifetime. Nine times out of ten, the person carries a latent form of the infection showing no symptoms of the disease. What causes some people to remain symptom free while others to develop the disease is still not clearly understood. The change from the latent form to the active form has to do with the lifecycle of the bacillus inside a human body. Once inside the body, the tubercle bacilli are immediately attacked by the natural immune response of the individual and get engulfed by macrophages. Inside the macrophages the bacillus encounters a completely different and hostile environment, which leads to a drastic altering of the bacteria's metabolic processes. The mycobacterium goes into a dormant state where it doesn't grow actively but remains nested in macrophages. When external conditions lead to a weakening of the immune system, the bacteria seizes the opportunity and bursts out of the macrophage to begin its active life cycle. This is when the bacteria begin to divide rapidly, spreading the infection to various body parts. This is also why TB is one of the

leading causes of death by infectious diseases in immune-compromised people such as those with certain types of cancer or people infected with HIV.

The Tuberculosis Structural Genomics Consortium (TBSGC) was initiated to streamline the process of drug discovery against *Mtb* [22]. The goal of the consortium was to solve the three dimensional structures of proteins from *Mtb* that are considered important drug targets. The important drug targets could include those proteins found to be essential to the survival and growth of *Mtb* through gene knock-out studies, or the genes that are identified through comparisons of the whole genome sequencing of wild type and the drug resistant strains of *Mtb* genome. With the rapid improvements in technologies pertaining to gene cloning, protein expression, purification, crystallization and x- ray crystal structure determination, the consortium also aimed to solve the protein crystal structures in a high-throughput manner. Members of the consortium could apply their areas of technological expertise to collaborate with other members specializing in other areas of the protein structure determination process. Solving the crystal structure of proteins or enzymes that are known drug targets with known inhibitors, could help in understanding the mechanisms of inhibition and design new inhibitors or modify existing inhibitors that exhibit higher binding affinities by efficiently exploiting the hydrophobic, polar and electrostatic interactions in the active site. Nearly half of the approximately 4000 open-reading frames (ORFs) remain unannotated or wrongly annotated in the *Mtb* genome [23]. Thus, apart from providing a structural basis for the drug discovery efforts, a secondary goal of the consortium was to identify the function of unknown genes in the *Mtb* genome. Newly annotated genes may highlight the presence

of important biochemical pathways in *Mtb* that can further fuel the drug discovery process by identifying new protein structures that may be of biological relevance in *Mtb*. Thus the defined goals of the consortium were;

1. To construct a knowledge base to contribute to the understanding of *Mtb* pathogenesis and aid in drug design,

2. To develop scalable technologies to make structural genomics research accessible to a larger number of researchers,

3. To solve the structures of biologically important proteins as a basis for drug discovery efforts, and

4. To demonstrate the benefits of a distributed consortium approach for large-scale research [24].

Strategies for Drug Discovery

This dissertation uses two traditionally different but complementary drug development strategies to find inhibitors of *Mycobacterium tuberculosis* (*Mtb*), the cause of TB. TB, called the scourge of mankind, has gone from being completely untreatable to being on the verge of absolute control by the use of chemotherapy. For various reasons discussed above and in spite of numerous attempts made by the scientific and medical establishment to find a lasting cure, the eradication of TB has been out of reach. TB has reemerged as a major health concern in the twenty first century, especially in the developing countries but also in the countries of America and Europe. It is now again threatening to become a pandemic due to a period of negligence in keeping up with the

treatment regime against this opportunistic pathogen and also due to the unusually high mutation rate observed in *Mtb* that allows *Mtb* to develop resistance rapidly against drug therapy. New drugs are urgently needed to combat TB.

In order to design and develop an effective and lasting therapy for the elimination of TB the drug discovery approach needs to be designed rationally rather than through random discoveries that have dominated the past century. A planned research results in a reduced amount of risk and serendipity in the discovery process. As Henderson points out that drug discovery process is no longer a '… *random procedure in which inspired scientists, working around the clock, come upon breakthroughs in the middle of the night*' [25]. Given the imminent threat posed by rapidly spreading XDR strains of TB, it is imperative that a rational drug discovery approach be employed to hasten the process of identifying good inhibitors and lead molecules. This dissertation will explore two fundamentally different but complimentary approaches to drug discovery. These are;

1. Structure – based drug design approach and
2. Screening of natural products extract libraries for inhibition

The structure–based drug design approach, also called a target based approach, begins with the identification of a druggable *target* enzyme in the pathogen. Accurate knowledge of the three dimensional structure of the target enzyme and its catalytic active site is used to design small molecule ligands that can bind the target with high affinity [26]. X-ray crystallography or NMR spectroscopic methods are typically used to elucidate the three dimensional structure of the target macromolecule. Improvements over the past

decade in the instrumentation and methods of macromolecular structure determination have resulted in an exponential growth in the availability of the structures of target proteins in the available databases [27,28]. A structure guided medicinal chemistry approach is used to design the small molecule inhibitors of the target enzyme based on the knowledge of the active site architecture. Such inhibitors are predicted to be highly selective to the active site of the target of interest.

The target enzymes are selected based on their essentiality to the survival of the microorganism in different growth conditions. Enzymes considered as virulence factors make an interesting drug target in this perspective as these are not essential to the microorganisms under normal growth conditions but are required to establish an infection in the host organism. Many enzymes involved in different metabolic and biosynthetic pathways have been successfully targeted for inhibition studies [29]. The complete genome of *Mtb* has been sequenced and is available [30]. The TB structural genomics consortium (TBSGC) pipeline was established to solve the crystal structures of as many proteins from the TB genome as possible in a high-throughput manner [22]. The goal of the consortium is to have the structural information on all the potential protein targets available in a database [31]. The three dimensional structure of the protein in its apo form and substrate / product bound form helps to identify the presence and orientation of the different domains, the active sites and the preferred oligomeric state of the protein. With a detailed knowledge of the enzyme activity, its active site architecture, the chemical environment afforded by the active site residues and using the structural scaffold of the enzyme's natural ligands, chemical synthetic strategies are used to design

inhibitors that can potentially bind in the enzyme's active site and inhibit or alter its function. Based on the importance of an enzyme to the survival of the microorganism a loss of enzyme function can have a potentially detrimental cascading effect on one or more metabolic pathways of the organism ultimately resulting in cell death or cell growth arrest. As enzyme inhibitors are typically analogs of the substrate or product of the enzyme an effective inhibitor is one that binds with greater affinity than the enzyme's natural substrates and is both specific and selective towards the targeted enzyme.

Selectivity of an inhibitor to an enzyme arises when the designed inhibitor is unique to the active site of the targeted enzyme. For example, when developing inhibitors against protein kinases, it is important to consider the fact that there are many different types of kinases in an organism and that the active sites of most kinases are identical. Hence substrate analogs designed to inhibit one type of kinase may very well inhibit another kinase, making it a promiscuous inhibitor.

The concern over inhibitor specificity arises when the host organism encodes a protein or an enzyme that is homologous to the enzyme being targeted in the pathogen. If the host's as well as the pathogen's enzyme are conserved through evolution, the architecture of the active site would also have evolved to encode similar amino acids to perform the enzymatic reaction. Thus an inhibitor designed to inhibit the reactive site of the pathogenic enzyme would follow the same mechanism to bind the host enzyme resulting in cytotoxicity issues. Essential enzymes in the pathogens that do not have a homologous host enzyme are considered good targets because there are no other

17

enzymes competing for specificity of the potential enzyme inhibitor. In the absence of such a scenario, differences are sought between the host and the pathogen enzymes. Such differences in the architecture of the active site or the amino acids involved in the activity are exploited to develop the inhibitors so that they have high affinity for the target enzyme in the pathogen while having the least possible affinity for the host enzyme. A series of inhibitors designed in this manner are typically tested in an *in vitro* assay with the host and pathogen enzymes and the minimum fifty percent inhibitory concentrations (IC_{50} values) are calculated for both enzymes. Minimum *in vivo* inhibitory concentrations (MIC) are also measured wherever appropriate. However a problem with the target based approach to drug design is that, while the inhibitors are designed based on the thermodynamic parameters affecting their interaction with the enzyme's active site, not much consideration is given to the permeability of the inhibitors through the complex cell wall of the pathogenic organism. This is an inherent problem with the target based approach to drug design, where many potentially good inhibitor candidates are rendered useless due to their poor cell permeability. Chemical modifications carried out to increase cell permeability may also increase the *in vitro* IC_{50} values, thus affecting the selectivity and specificity of the inhibitors.

In the current research two key proteins functioning in different essential metabolic pathways of *Mtb* are described. The two pathways are;

1. Methylation pathway and
2. Nitrogen regulatory pathway

18

A brief introduction is provided for each pathway and the role of the selected target protein is described.

The Methylation Pathway

Methylation is central to the survival of a cell [32]. Simply explained, methylation is the addition of a methyl group to a biomolecule. The biomolecule, called the acceptor, could be DNA, proteins or other biomolecules like phospholipids and secondary metabolites. S-adenosyl methionine (SAM) is the most common methyl donor used in methylation reactions. The reaction is also referred to as transmethylation. SAM is thus ubiquitous in the cell and is the second most frequently used cellular substrate of an enzyme after ATP [33]. The methylation reactions are carried out enzymatically by methyl transferase enzymes. Numerous methyltransferases such as lysine methyltransferases, histone methyltransferases and DNA methyltransferases perform the methylation reactions in the cell to perform various cellular functions including quorum sensing [34], phospholipid biosynthesis [35] and DNA replication and repair [32].

Methyl transferases play the central role in methylation pathway. As a methyl group from SAM is transferred on to the acceptor biomolecule, SAM is converted to s-adenosyl homocysteine (SAH). SAH has a higher binding affinity for methyl transferases and is shown to be a potent feedback inhibitor of the transmethylation reaction. Thus, SAH needs to be eliminated from the cellular environment in order to facilitate continuation of the transmethylation process. An intracellular ratio of SAM to SAH is maintained at 10:1 and is called the methylation ratio [36]. The elimination of SAH

19

is processed by different methods. SAH can be hydrolysed by nucleoside phosphorylase into adenine and ribosyl homocysteine, or it is converted to s-inosyl homocysteine by methyl thioadenosine (MTA) / SAH deaminase (also called SAH aminohydrolase) [37], or it is hydrolysed to adenosine and homocysteine by SAH hydrolase (SAHH). Hydrolysis by SAHH leads to the formation of adenosine and homocysteine. The methylation pathway generates essential byproducts as each of the products mentioned above are fed back into other pathways like the purine salvage pathway, methionine biosynthesis, cysteine biosynthesis, taurine biosynthesis and α-ketobutyrate biosynthesis.

Since methylation reactions are central to the survival of bacterial cells, enzymes involved in this pathway have been targeted for inhibition studies [38]. For example, DNA methylation has been recognized as the driving force behind epigenetics with the methylation pattern often profoundly altered in cancer cells. Hence DNA methylation has been targeted for cancer chemotherapy [39]. Enzymes of the purine salvage pathway have been targeted in plasmodium falciparum for the treatment of malaria [40]. Methylenetetrahydrofolate reductase, a key enzyme in the methionine biosynthesis has also been targeted for cancer therapy [41]. Similarly, inhibitors of SAHH have also been investigated for antiviral chemotherapy [42] and human coronary artery disease [43].

SAHH represent a particularly good target in the methylation pathway because the enzyme is crucial to the maintenance of the intracellular SAM to SAH concentration. Inhibition of SAHH results in the accumulation of intracellular SAH, which in turn inhibits the methyltransferases, affecting the entire transmethylation pathway. Essentiality of SAHH in mammalian model was shown by the fact that chromosomal

20

deletion of the gene encoding SAHH caused embryonic lethality in mice [44]. SAHH is

also the only source of homocysteine (Hcy) in the cell and as Hcy is the precursor to the

formation of cystathione, cycteine and taurine in the transsulfuration pathway, SAHH

becomes an important checkpoint in the biosynthesis of these important metabolites. *Mtb*

has no homologs of SAH deaminase or SAH halogenase. *Mtb* encodes for a putative

SAH nucleosidase, which has 30 % protein sequence identity with the *E. coli* SAH

nucleosidase. *Mtb* SAHH was also shown to be essential for growth *in vitro* and thus it

represents a potential drug target in *Mtb* [45].

The Pathway of Nitrogen Control

Nitrogen is essential to life regardless of the type of organism or the environment

they grow in. Nitrogen is a component in some of the most basic molecules in the living

cell such as nucleic acids, amino acids, cell wall constituents and other secondary

metabolites. Microorganisms are constantly challenged for survival by a changing

nitrogen environment. Thus microorganisms have developed complex and failsafe

pathways for the regulation of nitrogen [46]. *E. coli* is considered a model organism among

prokaryotes whose nitrogen regulation system has been studied extensively [47]. It is

widely established that *E. coli* uses glutamine (Gln) levels as the nitrogen status and 2-

Oxoglutarate (2OG) levels as the carbon status of the cell [48]. A drop in the Gln level or

an increase in the 2OG level is perceived as a nitrogen deficient condition and

mechanisms are immediately kicked into action to allow nitrogen to be taken up from

the environment (usually in the form of ammonia /ammonium, nitrates, dinitrogen).

Conversely, drop in 2OG levels or increase in Gln levels signal nitrogen excess.

Consequently, the 2OG and Gln concentrations are closely monitored in the cell using a

highly conserved and ancient set of nitrogen sensing proteins called PII [49,50]. The PII

constitutes a family of small proteins that work in tandem with other nitrogen regulatory

pathway enzymes to convey the nitrogen status of the cell, ultimately controlling the

action of glutamine synthetase / glutamine: 2-Oxoglutarate aminotransferase enzymes

(GS/GOGAT) [47,51].

The GS/GOGAT system plays key role in the nitrogen regulation, converting

2OG to glutamate (Glu) and further on to Gln, which is then utilized by downstream

pathways that incorporate the nitrogen [52]. The action of GS is controlled by post

translational modification as well as at the transcriptional level [53]. In *E. coli* both modes

of regulation have been shown to be dependent on a PII family protein called GlnB [54].

A second homolog of the PII protein, GlnK, controls the cellular nitrogen levels

by directly modulating the activity of an ammonium transporter, AmtB [55,56]. AmtB is a

membrane bound channel that allows the direct influx of ammonium into the cell [57,58].

GlnK forms a complex with AmtB through its T-loop and blocks the channel, while

uridylylated GlnK is unable to form complex with AmtB [59].

There is high sequence and structural homology between GlnB and GlnK

proteins. However, there are differentiating factors between the PII proteins. For

instance, GlnK proteins carry a 3_{10} helix at their C-terminal ends, which is absent in the

GlnB proteins. Further, GlnB proteins bear a lysine (Lys) residue at the third position

and an aspartate (Asp) or glutamate (Glu) at the fifth position. Additionally, the *glnK*

gene is typically found next to the *amtB* gene forming an *amtB-glnK* or *glnK-amtB* operon, while the *glnB* gene typically lies in an operon with *glnA*.

Curiously, *Mtb* is found to carry a single homolog of the PII protein [60]. The *Mtb PII* gene is found in the *glnD-PII-amtB* operon. The protein sequence consists of a Leu at the third position and Thr at the fifth position, both consistent with the *E. coli* GlnK protein. The *Mtb glnD-PII-amtB* operon was upregulated under conditions of nitrogen starvation and found to work independent of the GlnE and Gs nitrogen regulatory enzymes suggesting that PII plays an important role in nitrogen control in the mycobacteria through interactions with the AmtB and GlnD [61].

Screening of Natural Products Extract Libraries for Inhibition

An alternate strategy for drug discovery uses a high-throughput screening (HTS) approach. In this method large libraries of synthetic / combinatorial compounds or natural products are screened to search for lead compounds or classes of compounds against a select protein target in an *in vitro* assay or tested for whole-cell growth inhibition of microorganisms in a cell culture. The key to such HTS approach are the size and quality of the library of compounds used. Big pharmaceutical companies possess large chemical collection of molecules that can range in their size from one hundred thousand to a million molecules. The molecules are sourced from diverse areas like randomly generated individual chemicals in the laboratory, or arrays of biased chemical libraries of compounds developed using combinatorial chemistry, or natural products. While randomly generated compounds may not appear to be suitable in an *in*

vitro assay where the target is well characterized and biophysical information about the target active site is known, such libraries do provide a source of diversity in the screening process, that can lead to a chance discovery of a novel structural class of lead compounds that can be further explored. Libraries of random chemical entities, if sufficiently large, are capable of sampling the 'chemical space' and are especially useful in HTS efforts where not much structural information is available going into the experiment. Combinatorial chemistry approach is used to generate analogs of a particular structural scaffold by exploiting the functionalities existing in the core molecule [62]. For example, if the core molecule has three functionalities and if each functional position can be modified with ten different monomers then potentially a library of one thousand such analogs can be generated using the ten monomers. Combinatorial chemistry technique thus emerged in the nineties as an efficient way of feeding the high-throughput screening libraries by generating large volumes of compounds based on a structural scaffold [62]. Such libraries are useful in a biased structure guided assay where the scaffold of a lead molecule is already known. Natural product libraries, on the other hand, represent the widest collection of molecules that can be generated as these are sourced from nature which has evolved and perfected its synthetic methodologies over millions of years. In terms of drug discovery natural product sources remain unrivaled as the number of approved drugs originating from natural products out compete those originating via synthetic methodologies [63,64]. The different sources of natural products include microorganisms like bacteria and fungi dwelling in the soil [65] or sea [66], marine sponges [67,68] and plants [69,70].

This dissertation will provide examples of high-throughput screening approach applied to *in vivo* (whole cell bactericidal) activity specifically against *Mtb* and *M. smegmatis* cells as well as *in vitro* assays directed against selected targets from *Mtb*. One of the main advantages of HTS studies on *in vivo* systems is that the cell permeability of the inhibitor is taken care of at the outset of the screening experiment as compounds exhibiting poor cell permeability are automatically screened out of the library. All the first-line therapy drugs currently used for TB like rifampicin, isoniazid, pyrazinamide and ethambutol were identified by this method of screening [71]. Other factors affecting drug development like cytotoxicity can also be taken care of by designing appropriate screens to eliminate the cytotoxic compounds. For example the collection of compounds can be screened simultaneously against a human cell line or other eukaryotic models to test for promiscuity and hence cytotoxic properties. Compound exhibiting *in vivo* activity against the mycobacterial cells as well as the eukaryotic cells are considered as likely cytotoxic agents and eliminated from the screen. Once a library is narrowed down to a list of prioritized hits, the target for the hits can be identified by a number of methods like growing resistant mutants for the hits and employing whole genome sequencing to pin point the target gene or genes in the microorganism [72]. Other methods include supplementing the cell culture medium with various secondary metabolites to determine which metabolites can rescue the culture from the bactericidal effects of the drug. Identifying such secondary metabolites provide useful insights into the metabolic pathways that the drug candidate might be targeting. Success of the high-throughput whole-cell screening approach is evident by the number of drugs that were identified by

this method and are currently undergoing clinical trials for TB therapy, namely, SQ-109, OPC-67583, LL3858, TMC 207 and PA -824 [73,74].

The structure-based drug discovery and the high-throughput screening methods have their advantages as valid drug development approaches to identify an inhibitor or a lead molecule in the shortest possible time. At the same time the challenges encountered in one type of approach seem to have a solution in the other approach. Thus it appears that a combined approach incorporating the structure–based and the high–throughput screening based methods has the potential to complement each other and work synergistically to significantly reduce both the cost and time of developing useful pharmacological agents [75]. In the forthcoming chapters, examples will be provided for each of the drug discovery approach, including the various experimental challenges encountered in the process followed by a conclusion.

CHAPTER II

MTB S-ADENOSYL-L-HOMOCYSTEINE HYDROLASE

Background and Introduction

S-adenosylhomocysteine hydrolase (SAHH) is a ubiquitous enzyme that

catalyzes the reversible hydrolysis of s-adenosylhomocysteine (SAH) to adenosine and

homocysteine (Hcy) in the presence of a bound NAD^+ as cofactor. The mechanism of

SAH hydrolysis was first reported in 1979 by Palmer and Abeles and is represented in

Figure 1 [76]. The first step involves the oxidation of hydroxyl group at C3' position of

SAH to form a 3' keto intermediate. In the first step NAD^+ is converted to NADH. This

is followed by the abstraction of a labile proton at the C4' position to form a carbanion

intermediate. Then the thioether linkage breaks to eliminate the homocysteine moiety

and a 4', 5' didehydro-3' keto adenosine intermediate is left at the active site. Michael

addition of a water molecule across the 4', 5' double bond results in the formation of 3'

keto adenosine. This is then followed by reduction at the C3' keto position to yield

adenosine and the simultaneous conversion back of NADH to NAD^+. Thus NAD^+ is

regenerated during every turnover to catalyze the next batch of SAH hydrolysis. SAHH,

through its function regulates the intracellular concentrations of SAH, adenosine and

_____

homocysteine. Inhibition of SAHH results in the accumulation SAH inside the cell

which in turn increases the concentration of SAM due to feedback inhibition of the

methyl transferases by SAH. Concentrations of adenosine and Hcy are also reduced

upon the inhibition of SAH hydrolysis. Thus SAHH inhibition perturbs not only the

methylation cycle but also the purine salvage pathway and the transsulfuration pathway.

While the methyl transferases have many copies, there is a single copy of the

SAHH enzyme in most organisms. Being unique and centrally located in important

metabolic pathways SAHH has traditionally been a good target for inhibition. For

example; SAHH has been targeted for inhibition in hypercholesterolemia [77], malaria [78],

Chagas disease [79], Cryptosporidiosis [80] and cancer [81].

Mtb SAHH is encoded by the gene *Rv3248c*. It is a 54 Kd protein that is essential

for *in vitro* growth of *Mtb* [45]. Although it is down-regulated two- to three fold during

starvation conditions *in vitro* [82], and is not significantly up- or down-regulated in

activated macrophages [83], it appears to be up-regulated in infected mouse lung tissue

(determined using promoter-trap experiments) [84], a condition which also shows up-

regulation of other infection – related genes, like isocitrate lyase. SAHH is also
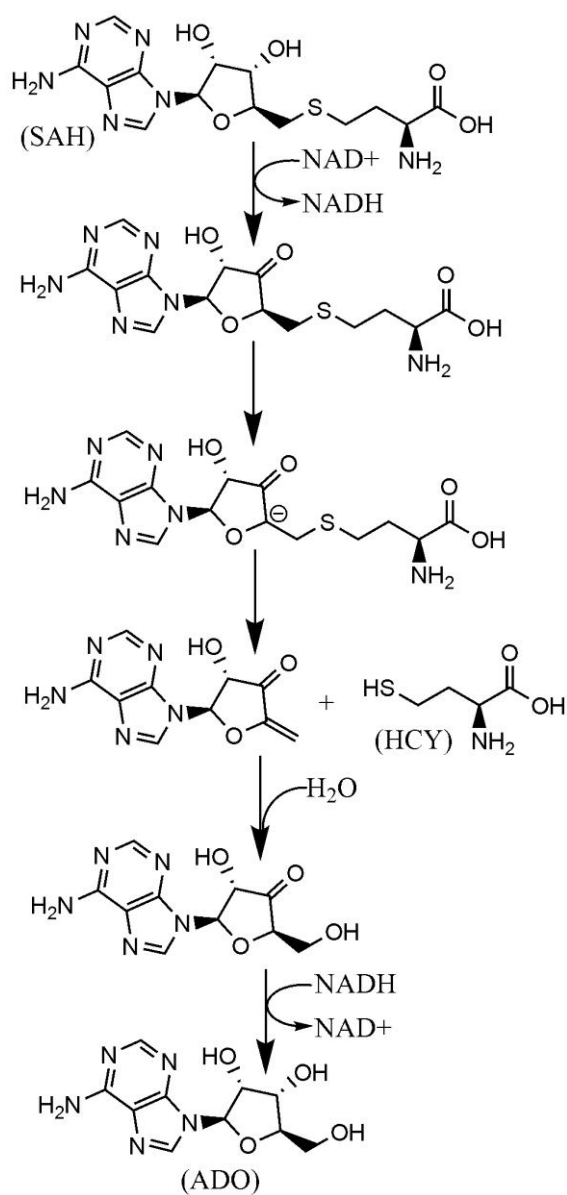
Figure 1: The mechanism of hydrolysis of SAH to ADO and HCY by *Mtb* SAHH.

considered ''druggable,'' given the large number of nucleoside analogs with

activity that have been discovered, such as aristeromycin (ARI) [85], neplanocin A[81,86], and

other ADO analogs [87]. However, these are not considered clinically relevant due to cytotoxicity issues [88] (**Figure 2**).

The activity of ADO analogs as competitive inhibitors can be rationalized through similar recognition and affinity compared to substrate, but inability to complete the reaction [85]. For example, 2',3'-dihydroxy-cyclopent-4'-enyl adenine (DHCeA) inhibits SAHH via a type-I mechanism [85] by oxidation of the 3' hydroxyl to a 3' keto group, which reduces the bound NAD^+ to NADH, thus inactivating the enzyme [89]. This compound represents a family of carbocyclic pyrimidine nucleosides, like neplanocin A [90], that form similar chemically-abortive complexes [91]. ARI is also expected to inhibit SAHH via a type-I mechanism [92]. Attempts have also been made to synthesize compounds exhibiting type-II inhibition (activation followed by covalent attachment to active site) [85]. For example, 5'-carboxyldehyde and vinyl-fluoride derivatives of ADO appear to react with a protein nucleophile at the 5' position [92]. There are also more conventional reversible inhibitors, including acyclic analogs such as D-eritadenine, whose affinity appears to be due to reorganized hydrogen bonding, which locks the enzyme into a closed conformation [93].

While SAHH is an essential enzyme, it is not unique to *Mtb*. Copies of SAHH gene exists in the human genome as well. The protein sequence identity between the human and *Mtb* SAHH is 61 %. With such a high sequence identity between the two

Figure 2: Chemical structures of analogs of ADO and inhibitors of SAHH.

enzymes, there is a risk of cross reactivity of designed *Mtb* SAHH inhibitors, inhibiting

the human SAHH, and hence causing cytotoxicity. Hence one challenge was to design

drugs that would be specific to the *Mtb* SAHH. A second challenge was that of

selectivity of the inhibitor to the target protein of interest. As mentioned earlier, a good starting point for designing inhibitors is by making analogs of the natural substrates or products of the enzyme. SAHH is a nucleotide binding protein and a typical genome encodes for many nucleotide binding proteins in the cell such as kinases and phosphorylases. Therefore nucleotide and nucleoside analogs suffer from the disadvantage that they may have affinity towards other ADO binding proteins and hence lose selectivity.

The first step was to crystallize and solve the crystal structure of the *Mtb* SAHH. The crystal structure for SAHH has been solved for three different organisms to date, all eukaryotic: human, rat (*Rattus rattus*), and malaria (*Plasmodium falciparum*) [77,89,91,93-98]. The *Mtb* SAHH structure would then be compared with the human and any other available SAHH structure to identify differences in or near the active site between the different structures. Such differences can then be exploited to design specific inhibitors of *Mtb* SAHH.

Methods

Cloning and Expression of **Mtb** *SAHH*

A 1.49 kb DNA fragment containing the *SAHH* gene (*Rv3248c*) was amplified by PCR from *Mtb* H37Rv genomic DNA as a template, using the following oligonucleotides as the forward and reverse primers, respectively:

5'- AGA TGA AGC CAT ATG ACC GGA AAT TTG GTG ACC 3'

5'- AGA GTA AGC TTA GTA GCG GTA GAG GTC CGG CTT-3'.

The amplified DNA fragment was digested with *Nde*I and *Hind*III and subcloned in the corresponding restriction sites in a pET28b vector to yield an N-terminal 6 X (His) tag. The kanamycin resistant SAHH-pET28b vector was transformed into *E. coli* BL21 (DE3) cells, and grown to the mid-log exponential phase at 37°C in LB media (Difco). Cells were induced at 18°C at an $O.D_{600}$ of 0.8 with 1 mM IPTG and grown for 16 hrs.

Purification of **Mtb** *SAHH and Preparation of the Apo Enzyme*

The harvested cell pellet was resuspended in buffer containing 20 mM Tris-HCl pH 8.0, 10 mM imidazole, 0.5 M NaCl and 10 % glycerol with 1 mM phenylmethanesulfonyl fluoride and DNAse. The cells were lysed using French press. SAHH was purified on a Hi-Trap Ni^{2+} chelating metal affinity chromatography column (GE Healthcare). The peak fractions were pooled and dialysed against 20 mM Tris pH 8.0, 50 mM NaCl, 10 % glycerol, 1 mM dithiothreitol (DTT) and concentrated to 18 mg/mL in an Amicon Ultra Centrifugal filter device (MILLIPORE). The enzyme was found to be about 95 % pure as observed on SDS PAGE.

The purified enzyme was stripped of both forms of the dinucleotide to prepare an apo-enzyme. The apo-enzyme was made as described for the preparation of rat liver apo-SAHH [99]. It involved an acidic (pH 3.3) ammonium sulfate precipitation of the enzyme followed by re-suspension in a potassium phosphate buffer (pH 7.2) containing 5 mM DTT and 1 mM EDTA. The step was repeated a second time followed by precipitation with a neutral saturated ammonium sulfate solution. The precipitate was re-dissolved in 20 mM potassium phosphate containing 0.1 mM EDTA. Any undissolved protein was

collected by centrifugation and discarded. The apo-enzyme was inactive and could not catalyze the hydrolysis reaction by itself. Activity was reconstituted by incubating it with 2 μM NAD$^+$ for 10 min before carrying out the activity assay. The reconstituted holo-enzyme was used for the enzyme activity assays.

Crystallization

Crystallization screening of *Mtb* SAHH was performed using the sitting drop and hanging drop methods with Crystal Screen I and II, Index (Hampton Research) and Wizard I and II (Emerald Biosystems), as well as random screening kits (Lawrence Livermore National Laboratory) using a crystallization robot (Robbins Scientific). Initial crystallization screening was carried out with SAHH alone, and pre-incubated with ADO or an analog inhibitor for 30 min at a molar ratio of ~1:15. Crystals for ADO-, DZA-, and ARI- bound SAHH were obtained at 18°C when 1:1 protein:reservoir solution drops were equilibrated within a hanging drop against 500 uL of mother liquor solution containing 20% w/v polyethylene glycol (PEG) 1000, 200 mM imidazole (pH 8.0), and 100 mM Ca(OAc)$_2$. Crystals for 2FA-bound SAHH were obtained in the presence of mother liquor solution containing 0.2 M MgCl$_2$, 0.1 M Tris pH 8.5 and 25% w/v PEG 3350. Diffraction quality crystals were obtained after 3-4 days. To determine the structure of SAHH in complex with SAH, 10 mM SAH was co-crystallized with SAHH in the same condition as for the complex with 2FA.

Data Collection, Structure Determination and Refinement

Crystals from a droplet were transferred directly to a cryoprotectant, N-paratone (Hampton Research). The crystals were mounted on nylon loops and flash-frozen in a liquid nitrogen stream at 100 K before data collection. The high resolution data of the ternary complexes of SAHH:ADO:NAD, SAHH:DZA:NAD, SAHH:2FA:NAD and SAHH:ARI:NAD were collected at beamline 19-ID on an Area Detector Systems Corporation Q315 area detector at the Advanced Photon Source, Argonne National Laboratory, Chicago, IL. The data were reduced using DENZO, and intensities were scaled with SCALEPACK [100]. Integrated and scaled data indicated that *Mtb* SAHH belongs to the $P2_1$ space group. The ternary SAHH:ARI:NAD, SAHH:DZA:NAD, and SAHH:2FA:NAD complexes were crystallized in the same space group with similar cell dimensions. Solvent content calculations indicated the presence of a tetramer in the asymmetric unit [101]. Data for SAHH:SAH co-crystals were also collected at APS on beamline 19-ID. The crystallographic data collection statistics are summarized in Table 1.

The crystal structures of *Mtb* SAHH bound with ADO or its analogs, ARI, 2FA, and DZA, have been determined to 1.60, 2.01, 2.42, and 2.20 Å, respectively. Attempts to crystallize apo SAHH did not yield diffraction quality crystals. The structures of the SAHH:ADO:NAD ternary complex of *Mtb* was solved by molecular replacement method, EPMR, using the human SAHH model (PDB ID: 1LI4) as a search model for the extending data from 25 to 3.5 Å [91]. Complexes with DZA, 2FA, and ARI were solved by molecular replacement from the ADO complex. Crystallographic refinement was

performed using REFMAC (5.02) [102]. NCS restraints were applied for each domain of the model. After model building and fitting, bias-minimized electron density maps were obtained using Shake&wARP protocol [103]. Model building was done using the programs XTALVIEW [104] and Coot [105]. After repeated cycles of refinement and manual model building, water molecules were added to the structure using Xfit [104]. The refinement continued with structure factors measuring 1.0 σ or better within the resolution range 30 to 2.48 Å to reduce the effects of poorly measured reflections. At the final stage of the refinement, the NCS restraints were removed for the entire model. The refinement statistics and the model stereochemistry are summarized in Table 1.

The SAHH:SAH:NAD complex structure was determined to a resolution of 2.0 Å. Our best data showed overall completeness of 82.5% in the resolution range of 30 – 2.0 Å. The low completeness may be attributed to the high crystal mosaisity (3.42). The data was processed using d*TREK [106]. After running Shake&wARP, clear density in the unbiased map was observed contiguous with the 5' carbon of the ADO, showing covalent attachment of additional atoms. Although positive density was apparent for most of the expected atoms of the HCY appendage, the density was not defined well enough to model the terminal atoms (amino and carboxylate groups). Therefore coordinates were only built and refined for Cβ, Cγ, and Sδ. The model was refined with Refmac5, and the statistics are also presented in Table 1.

In vitro *Assay*

HCY production was monitored using a spectrophotometric assay. Ellman's reagent (5,5'-dithiobis(2-nitrobenzoic acid), or DTNB), was used to convert HCY into a HCY-TNB complex [109,110] with an absorption maximum at 412 nm. The reaction mixture contained 0.5 μM apo-SAHH, 250 μM NAD$^+$, 100 μM DTNB, and 50 mM potassium phosphate (pH 7.2) for a total assay volume of 200 μL. Readings were observed on a POLARstar OPTIMA plate reader from BMG LABTECH, using an absorbance filter of 410 nm.

The apo-enzyme was incubated with cofactor NAD$^+$ and the inhibitors (where applicable) for 10 min. The reaction was initiated by addition of 100 μM SAH. The reaction progress was measured for approximately 2 min. The linear region of the curve was used for measuring the initial velocity parameters. Inhibition constants were determined by assaying 10 nM to 1 mM concentrations of the inhibitors in the presence of a constant SAH concentration.

Table 1: Crystallographic and refinement statistics for *Mtb* SAHH.

	SAHH:ADO:NAD	SAHH:ARJ:NAD	SAHH:2FA:NAD	SAHH:AD3:NAD	SAHH:ETA
Ligand Name	Adenosine	Aristeromycin	2-fluoroadenosine	3-deazaadenosine	adenosyl-L-homocysteine
Space Group	P21	P21	P21	P21	P21
Wavelength (Å)	0.964	0.964	0.964	0.964	0.979
Temperature (K)	120	120	120	120	120
a (Å)	97.04	94.85	96.25	95.74	94.2
b (Å)	112.13	111.85	112.08	112.25	111.69
c (Å)	100.23	100.4	100.45	100.85	100.13
α	90	90	90	90	90
β	92.41	96.49	95.15	95.84	96.01
γ	90	90	90	90	90
Resolution (Å)	1.6	2.01	2.42	2.2	2
Highest resolution bin (Å)	1.50 - 1.60	2.01 - 2.11	2.42 - 2.49	2.2 - 2.26	2.0 - 2.05
Observed Reflections	325473	136379	74187	93598	114623
Unique Reflections	308888	129513	70373	89824	108886
% completeness	94.9	98.9	91.91	97.9	82.5
R (merge)	0.049	0.067	0.0809	0.057	0.127
<I/σI>	13.1	10.12	10.52	13.06	8.36
Vm (Matthews Coefficient	2.51	2.44	2.48	2.48	2.48
% Solvent	50.92	49.49	50.41	50.42	50.41
Refinement					
Free R value, random, 5%	23.77	26.35	23.25	28.45	33.11
R value	20.04	21.2	21.21	21.19	26.02
protein residues	495	495	495	495	495
water molecules	1066	450	721	676	370
Rmsd bond length (Å)[a]	0.01	0.02	0.01	0.03	0.014
Rmsd bond angle (Å)[a]	1.38	1.88	0.83	2.42	1.63
Rmsd between subunits (Å	0.1	0.23	0.31	0.27	0.22
Overall coordinate error (Å	0.099	0.19	0.35	0.27	0.32
RSCC (*Refmac5*)[c]	0.95	0.951	0.948	0.947	0.893
Residues phi-psi angles					
Most favored (%)	91.3	90.7	89	89.5	89.4
Allowed (%)	8.5	8.9	10	10.1	10
Generously allowed (%)	0.2	0.2	0.3	0.4	0.4
Disallowed (%)	0	0	0	0	0.1

[a] Deviations from restraint targets.

[b] Estimated standard uncertainty, diffraction precision index (DPI) based on free R [107].

[c] Real space correlation coefficient, maximum likelihood mFo–DFc map, reported by REFMAC (5.02) [108].

Whole-Cell Assay

SAHH shares 90 % sequence homology between *Mtb* and *Mycobacterium smegmatis*. *M. smegmatis* has a shorter doubling time of approximately two to three hours compared to *Mtb* and has been used as a surrogate for *Mtb* [111]. Therefore *M. smegmatis* cells were used to study the MIC values for the inhibitors using whole cell inhibition assay. All inhibitors were dissolved in water to a concentration of 256 mg/L. First, 100 μL of Middlebrook 7H9 media (Difco) was dispensed into all 96 wells of a microtiter plate. Then 100 μL of the 256 mg/L stock inhibitor solution were dispensed into column # 2. Two fold serial dilutions were carried out for each column up to column # 10, giving inhibitor concentrations from 256 mg/L to 0.5 mg/L. Column # 11, containing only media served as a negative control. *M. smegmatis* mc^{2}155 [112] cells, diluted in 7H9 media, were added to all the wells to a final O.D$_{600}$ of 0.001. Cells were grown at 37˚C for 72 hrs. After this period, the wells were measured visually for cell growth. MIC values were determined where the wells had no visible bacterial growth.

Fragment-Based Screening for Lead Discovery

High-throughput screening (HTS) initiatives in the late twentieth century were designed to boost the rate of drug discovery rapidly and in a cost effective manner. The first HTS techniques were pioneered by industries to cope with their fast paced drug discovery process. The original screening libraries contained natural products in fermentation broths, which soon translated to dimethyl sulfoxide stocks of small molecules synthesized by combinatorial chemistry methods [113]. The synthesized

compounds had an average mass of five hundred daltons with hydrogen bond acceptors ≤ 10, hydrogen bond donors ≤ 5 and a cLogP ≤ 5. To satisfy the properties aspired in a hit molecule and at the same time efficiently cover the compound chemical space required the synthesis of large number of compounds. The hit rates were typically low at about $0.5 - 1$ % of the screening libraries which further required the screening of tens of thousands of compounds to identify statistically significant number of hits to proceed to the next step in the lead development. It has also been observed that the quality of the hits are crucial to the success of a HTS program as the lead compounds go through relatively minor chemical changes before they are launched as drug candidates [114]. The observation that despite a large increase in the number of screening compounds in the HTS libraries over the years, there has been no corresponding increase in the launch of new drugs suggests that the HTS methods are limited in their effectiveness. Better screening methods were pursued to increase the effectiveness of the screening while at the same time reducing the time and cost involved in the process.

In the past decade a fragment based approach has emerged as a viable alternative to HTS methods utilizing large number of compounds. The fragment-based lead discovery approach uses compounds with molecular weight less than three hundred daltons and contains a reduced number of functional groups with a lipophilicity value (ClogP) of < 3 [115]. The smaller size of the molecules makes it easier to sample the chemical space more efficiently while keeping the size of the compound library relatively small at hundreds or at the most thousands of compounds [116,117]. The leads identified by this approach tend to be weak binding (mM to µM range), but structural

studies of the target protein in complex with these fragments give valuable information about the interactions of the small molecules with the active site residues. Such interactions can be used as starting points for structure guided chemical synthesis of larger lead compounds that have higher binding affinities, thus increasing ligand efficiency and selectivity [118]. The building up of a lead molecule from fragments, also called tethering, has been shown to be an effective method for designing structure based inhibitors of target enzymes [119].

A 760 fragments library was used to screen against the *Mtb* SAHH enzyme. The fragment library was built using molecules with core structures that were consistently observed in lead compounds that exhibited drug like properties. This library was constructed in collaboration with the Prof. Chris Abell's laboratory at the University of Cambridge.

Differential Scanning Fluorimetry

A quick, easy and inexpensive method to perform fragment-based screening against a purified protein is by using the differential scanning fluorimetry (DSF). DSF typically uses a real time polymerase chain reaction (RT-PCR) to monitor the temperature-dependent unfolding of the protein and determine its melting temperature [120]. The protein is incubated with a SYPRO® orange protein stain and subjected to a gradual increase in temperature. As the protein begins to unfold, SYPRO® orange reacts with exposed hydrophobic patches in the protein increasing its fluorescent intensity, which can be used to determine the melting temperature (Tm) of the protein. Stabilizing

interactions of the protein with small molecules cause a shift in the Tm (ΔTm), which indicates protein: ligand complex formation (**Figure 3**). While the shift in the melting temperature of the protein is indicative of a binding event of a ligand to the protein, the amount of shift in Tm is not an indicator of the binding affinity of the ligand. Other methods like saturation transfer difference NMR or isothermal calorimetry are employed to confirm binding of the ligands, which will be discussed in the next section.
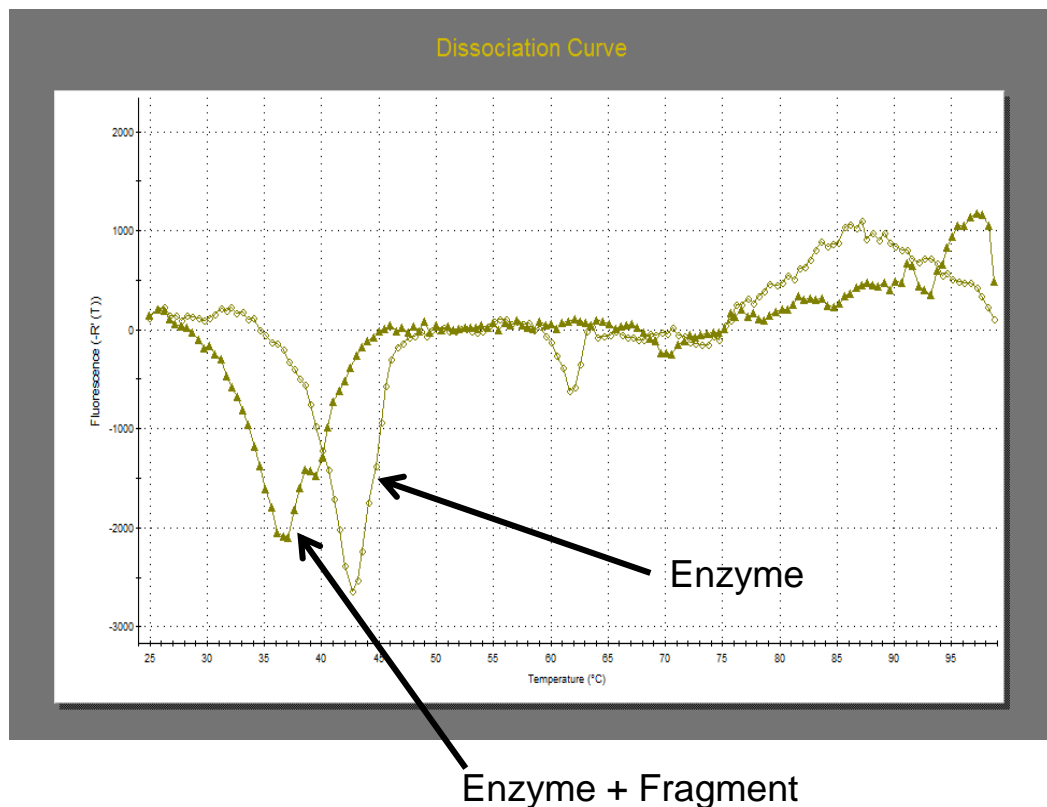
Figure 3: Representative thermal shift assay showing a shift in melting temperature ΔT of the enzyme upon binding the fragment.

As mentioned above, the DSF experiment was performed using an RT-PCR instrument to gradually increase the temperature of the reaction mixture consisting of *Mtb* SAHH and SYPRO® Orange dye in a suitable buffer while the protein unfolding event was monitored at the excitation and emission wavelength of the dye. First step in the experiment was to optimize the buffer. A wide selection of buffers covering their optimal pH range was used as controls to optimize the melting temperature of the enzyme. The different buffer solutions used in the experiment were; citric acid (pH 2.2 – 3.8), sodium acetate (pH 4.0 – 4.8), sodium citrate tribasic (pH 5.0 – 5.8), sodium cacodylate (pH 6.0 – 6.8), sodium salt of 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (Na HEPES) (pH 7.0 – 7.8), Tris (pH 8.0 – 8.8), 3-(Cyclohexylamino)-2-hydroxy-1-propanesulfonic acid (CAPSO) (pH 9.0 – 9.8) and 3-(Cyclohexylamino)-1-propanesulfonic acid (CAPS) (pH 10.0 – 11.0)

The experimental set up was as described in the protocol suggested by Niesen et al., [121]. To optimize the buffer conditions, *Mtb* SAHH at a final concentration of 4 µM was aliquoted into each assay well of a 96 well PCR plate. The different buffer concentrations were maintained at 0.2 mM final concentration. 1 µL DMSO was added to each well, which was replaced by the same amount of a fragment solution in DMSO during the actual fragment screening. SYPRO Orange dye which is obtained as a 5000 X concentrated solution was diluted in distilled water 500 times to a 10 X stock concentration. The 10 X dye solution was aliquoted to yield a 2.5 X final concentration in a final assay volume of 20 µL per well. The PCR plate was firmly sealed with a transparent plastic film. The plate was then spun down at 200 x g for 1 minute to settle

down the assay components and to eliminate any bubble formed during the pipetting steps. The RT PCR temperature settings were set to increase from 25 °C to 99 °C at the rate of 0.5 °C per minute. The excitation and emission of the SYPRO Orange dye was monitored at 492 nm and 610 nm respectively.

Once the suitable buffer was established the same experiment was run in the presence of the fragment library, while keeping the buffer constant. 1 μl of the DMSO was replaced by 1 μL of the fragment solution which was maintained as a 100 mM DMSO stock in a storage plate. Thus the final concentration of the fragment in the 20 μL assay volume was 5 mM.

Saturation Transfer Difference NMR

Saturation transfer difference nuclear magnetic resonance (STD-NMR) has been effectively used to detect the enzyme – ligand interactions [122]. In this technique the protein is selectively irradiated by Gaussian pulses causing saturation of the protein resonances, which is transferred to a bound ligand through intermolecular $^1H – ^1H$ cross relaxation. This is the on-resonance spectrum. Through exchange between bound and free ligand states the saturation is transported to solution and detected by subtracting a spectrum with saturation from an off resonance spectrum (**Figure 4**).
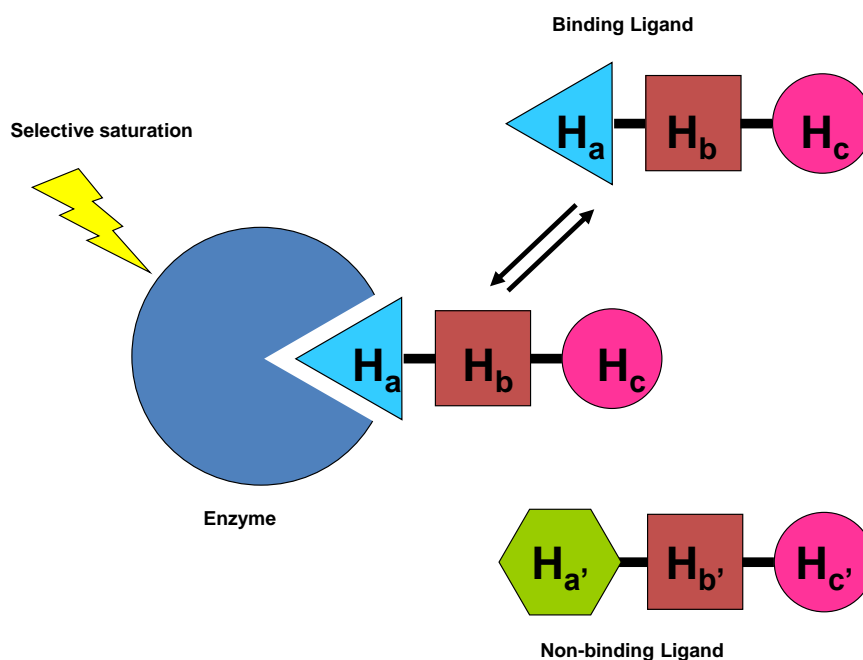
Figure 4: Principle of STD NMR. Cartoon depicting the transfer of saturation from enzyme to the bound ligand - $H_aH_bH_c$. The unbound ligand, $H_{a'}H_{b'}H_{c'}$ receives no saturation.

The STD NMR was measured in a buffer solution containing 20 mM Tris pH 7.5 and 50 mM sodium chloride. *Mtb* SAHH was pre incubated with ADO and NAD^+. A master mix of the enzyme at ~ 220 µM stock concentration with the buffer solution was prepared containing 10 % deuterated water. The fragments were added at a concentration such that the enzyme: fragment concentration was maintained at a ratio of 1: 100. The fragment stock solutions were prepared in 100 % DMSO. The final assay mixture consisted of 20 mM Tris pH 7.5, 50 mM sodium chloride, 30 µM *Mtb* SAHH, 3 mM fragment molecule and 10 % D_2O. In order to confirm the binding of a fragment to the enzyme, additional tests were performed by varying the concentration of ADO while

45

keeping the fragment concentration constant. Accordingly molar excess of ADO was used at 10 fold, 100 fold and a 1000 fold concentration of the enzyme, while maintaining a 100 fold excess of the fragment. Thus, to an assay containing 30 µM *Mtb* SAHH, ADO was added at 300 µM, 3 mM and 30 mM concentrations, while the fragment was maintained at a final concentration of 3 mM.

Results and Discussion

Crystal structure of **Mtb** *SAHH*

     Mtb SAHH crystallized as a homotetramer in the space group $P2_1$. Each subunit of *Mtb* SAHH (495 residues) consists of two α/β domains (**Figure 5**), as observed in previous structures, with domain I being a substrate-binding catalytic domain and domain II being a dinucleotide-binding domain (Rossmann fold). Each subunit is bound to one NAD^+ molecule. Domain I consists of residues 11-247 plus 423-466 (281 total), and domain II consists of residues 248-422 (175 total). In addition, there is a C-terminal extension of 29 residues (467- 495) observed in SAHH from other organisms (except Archaea, where it is truncated by 8 residues) [123] that covers the NAD- binding site in an adjacent subunit. This interaction is complemented by the other subunit in a local two-fold symmetry, making the tetramer a "dimer of dimers" (**Figure 6**). Upon a dimer formation (between chains A and B, for example (**Figure 6**)), ~5,700 Å of surface area is buried within each subunit, whereas a total of only ~3,440 Å of surface area is buried between chain A and the other two subunits (C and D) in the tetramer combined. The N-terminal 10 residues are disordered. This includes half of the 20-residue N-terminal

extension found in *Mtb* and other prokaryotic sequences but not seen in eukaryotes; residues 11-19 form an additional β-strand that packs against and extends the edge of the core β-sheet in domain I.
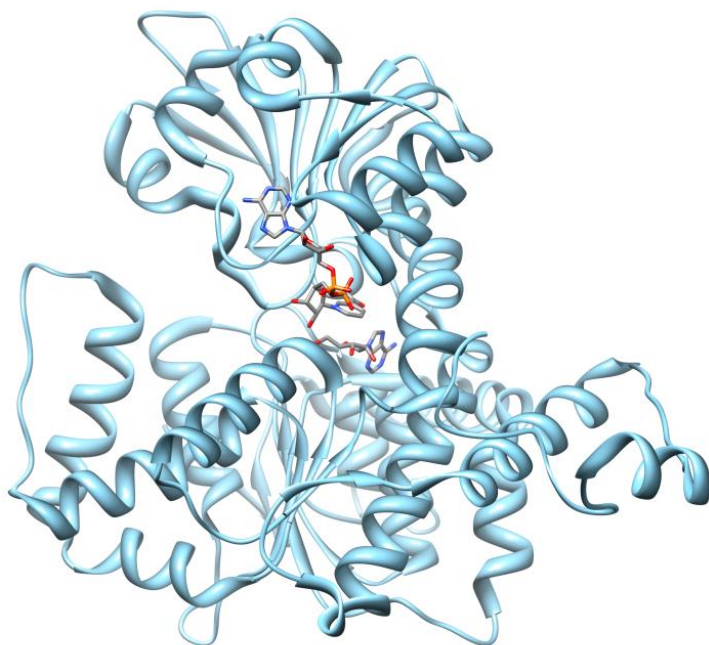
Figure 5: A single subunit of *Mtb* SAHH. The bound ADO and NAD$^+$ are shown in stick representation with the atoms colored by elements. Molecular graphics images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081) [124].
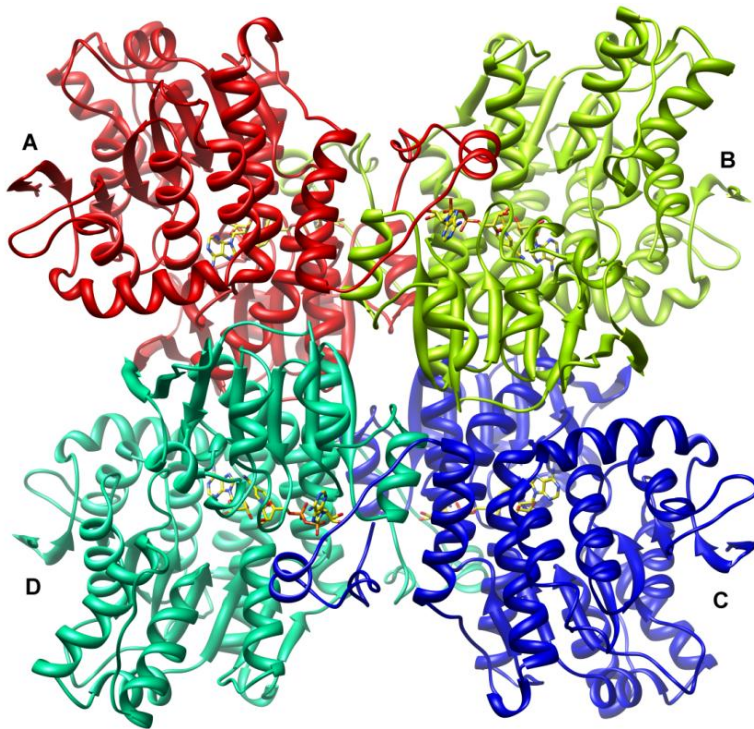
Figure 6: A tetramer of *Mtb* SAHH. The bound ADO and NAD$^+$ are shown in stick representation. Subunits A and B form a dimer and subunits C and D form a dimer, making this tetramer a 'dimer of dimers'.

Mtb *SAHH bound to NAD$^+$ and Ado*

The *Mtb* SAHH was crystallized in the presence of the product Adenosine (Ado) and the cofactor NAD$^+$ to look at the specific interactions of Ado with residues in the substrate binding domain. Comparison with the human SAHH, the crystal structure of which was also solved in complex with Ado, revealed no significant differences in the active site with respect to the Ado binding **(Figure 7)**.
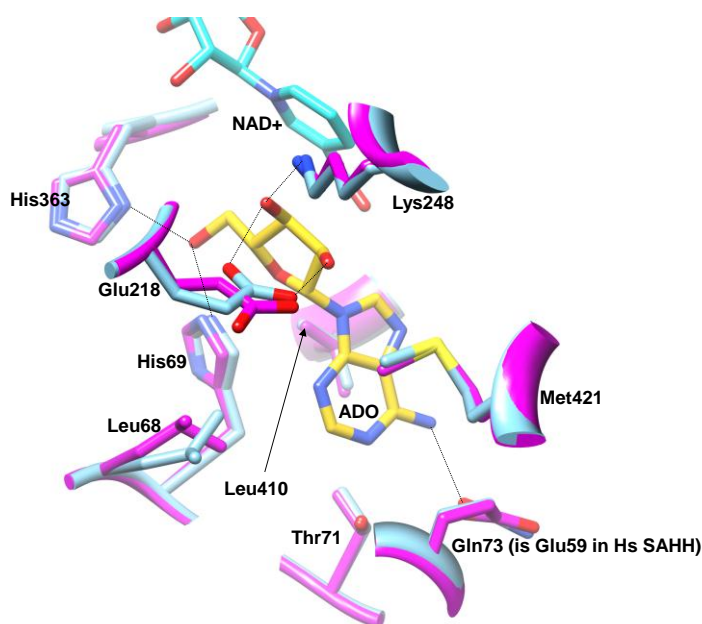
Figure 7: Superposition of *Mtb* and human SAHH active sites. The *Mtb* SAHH residues are shown in cyan and the human SAHH residues are shown in magenta. ADO is shown in yellow stick representation with nitrogen and oxygen colored by elements.

Mtb SAHH structure superimposes well on previously determined ligand-bound SAHH structures. The backbone Cα RMSD between the *Mtb* SAHH:ADO structure and the ADO-complexed structure from *P. falciparum* (PDB ID: 1V8B) (55% amino acid identity) is 0.84 Å (over 461 residues, or 94% of the sequence), as determined by SSM [125]. Similarly, the RMSD to the human SAHH (PDB ID: 1A7A) is 0.77 Å over 422 residues, and the RMSD to the rat SAHH (PDB ID: 1KY5) is 0.90 Å over 415 residues. A superposition of the backbones of these four enzymes is shown in Figure 8.The domains are in a "closed" conformation, with the hinge between domains flexed by ~17 degrees (based on comparison with apo-structure from the rat liver enzyme [126], since all

four structures reported herein were complexed with a product or an inhibitor).There is a 9-residue insertion (residues 109-117) that is unique to Actinobacteria [127], but this occurs in a surface loop and has no role on the active site or tetramer formation.
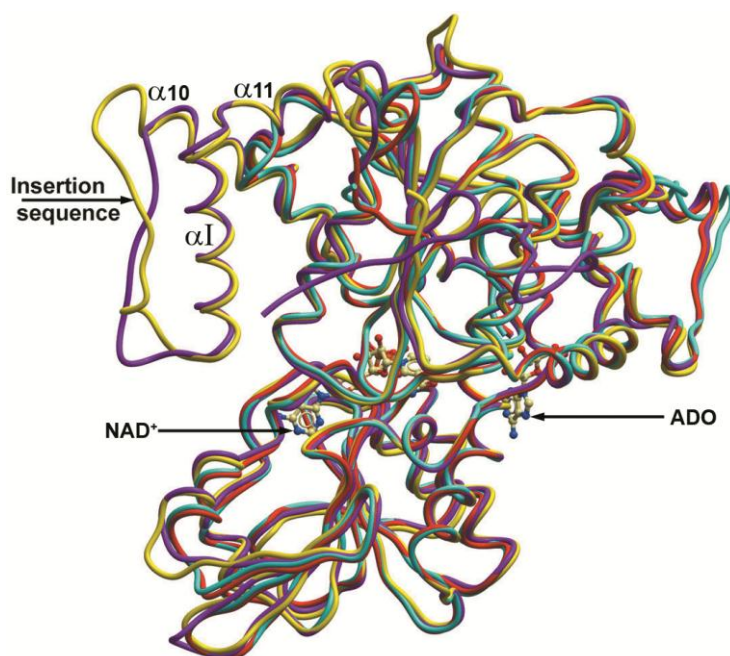
Figure 8: Superposition of *Mtb* SAHH (violet) with human (red, 1A7A), rat (cyan, 1KY5), and *P. falciparum* SAHH (yellow, 1V8B). The bound ADO and NAD^+ from the *Mtb* structure are shown for reference. The 37- residue insertion in the *Mtb* and Plasmodium structures can be seen forming a strand-turn-helix at the surface (upper left, in violet and yellow). α10, αI (insertion), and α11 have been labeled for reference.

Domain II binds NAD^+ in a similar fashion to that observed in all previous SAHH crystal structures. The dinucleotide-binding domain is considered an atypical

Rossmann fold because of variation in the topological connections of the $\beta\alpha\beta\alpha\beta$ core [89]. Specific interactions with NAD^+ include the side-chains of Thr20, Thr21, Asn253, Val286, Thr219, Thr220, Thr221, Thr338, Asn340, Thr304, Glu305, Ile306, Asn310 and Ile343, as well as numerous backbone atoms. All are conserved identically between human and *Mtb* SAHH, except Asn340 contacting the adenine ring, which replaces Cys278 in human SAHH. In addition, several residues from the C-terminal tail of an adjacent subunit make contact with the NAD^+, covering over the binding site. These include Tyr493 and Lys489 (from subunit B), which hydrogen bond with the diphosphate and ribose moieties, respectively.

The *Mtb* SAHH structure has a 37-residue insertion (residues 167-203) that is observed in other prokaryotic but not mammalian sequences [127]. The insertion extends helix α10 (human SAHH numbering [89]) by two turns, and then forms an anti-parallel strand-turn-helix (see **Figure 8**) that packs on the surface of domain I, before re-joining helix α11. The insertion is similar to that observed in the *P. falciparum* structure (PDB ID: 1V8B [97]). However, in 1V8B, helix α10 is extended by one additional turn before the strand begins, so they do not perfectly superimpose (**Figure 8**). The insertion structure lies on the surface of the substrate binding domain of a subunit and does not interact with adjacent subunits (i.e. does not influence packing of the tetramer). The insertion residues are far away from the active site (at least 12.7 Å away from the ADO in *Mtb* SAHH).
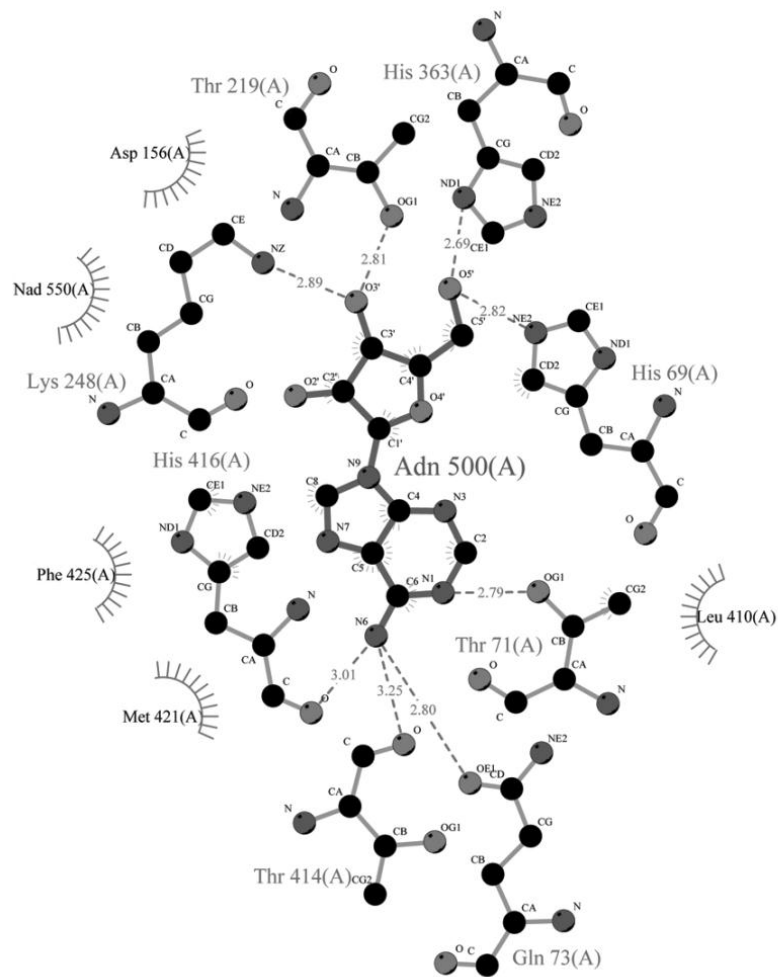
Figure 9: Interactions of ADO in the active site of *Mtb* SAHH with residues within 4 Å of ADO.

The ADO-binding portion of the active site (**Figure 9**) looks similar to that of previous structures from other organisms complexed with ADO analogs. The active site is enclosed in a buried cavity approximately 14 x 9 Å in size, with the nicotinamide ring of NAD^+ forming part of the boundary (packing against the substrate ribosyl ring, where it oxidizes and reduces C3'). All of the residues thought to be catalytically important are conserved, including Lys248 and His69 (equivalent to Lys185 and His54 in the rat numbering). The adenine ring is surrounded by hydrophobic residues: Leu68, Thr71, Gln73, Leu410, Met421, and Phe425. Ionizable residues Asp156, Asp252, Lys248 and Glu218 interact with the hydroxyls of the ribose. His69 and His363 form hydrogen-bonds with the 5' hydroxyl (2.82 and 2.69 Å, respectively). The only residue that is not perfectly conserved is a substitution of Glu58 (rat) by Gln73 in *Mtb* SAHH, both of which form a hydrogen bond with the N6 amino group on the adenine ring. In either case the carbonyl functional group on the side chain of Gln73 or Glu59 made hydrogen bonding interactions with the N6 amino group on Ado and did not represent an avenue for the design of a novel inhibitor specific towards the *Mtb* SAHH.

Mtb *SAHH bound to NAD^+ and Inhibitors*

Mtb SAHH was crystallized with three known inhibitors of SAHH in other organisms to study their modes of inhibitory actions. These were Aristeromycin (ARI), 2-Fluoroadenosine (2FA) and 3-Deazaadenosine (DZA) at resolutions of 2.01 Å, 2.42 Å, and 2.20 Å, respectively and all exhibited very similar binding-site geometry, with no significant changes in side-chain conformation. Superpositions with the ADO-bound
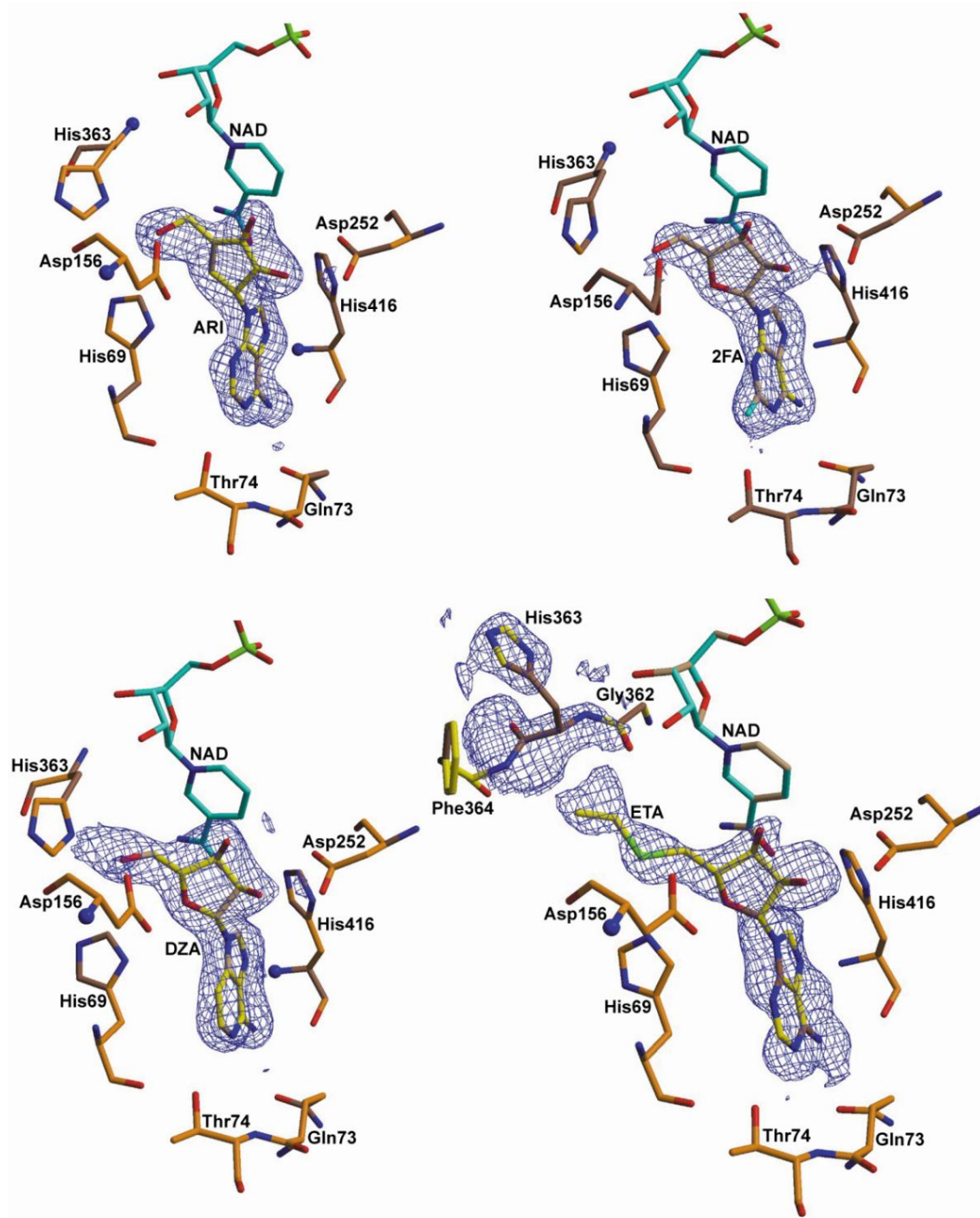
Figure 10: Active sites of *Mtb* SAHH bound to the inhibitors ARI, 2FA, DZA and the product SAH (represented as ETA).

structure had RMSD values of 0.18, 0.22, and 0.25 Å for complexes with ARI, 2FA, and DZA, respectively.

The complex with ARI (**Figure 10**) is the first crystal structure of SAHH to be solved with 3'keto aristeromycin. The ribosyl ring oxygen is replaced with a carbon atom in ARI, which is in van der Waals contact with Leu410 (3.65 Å), as in the ADO complex. Thus no significant hydrogen bonds observed in the other crystal structures were disrupted by this substitution. His69 and His363 still form hydrogen bonds with the 5' hydroxyl group of ADO. Importantly, C3' is observed to be planar, consistent with O3' being oxidized to the keto form. This confirms the prediction based on the proposed mechanism of action for ARI (type-I inhibitor, activated by NAD^+) [85]. This abortive complex bears resemblance to the complex of rat SAHH with 3'-keto adenosine (PDB ID: 1KY5) [96], in which C3' is also planar. O3' nonetheless continues to hydrogen bond with the nearby Lys248 side-chain, which is close enough (3.23 Å) to abstract a proton, as well as Thr219 (2.82 Å). ARI strongly inhibits the *Mtb* SAHH enzyme, with an IC_{50} of 0.2 μM (see **Table 2**).

Table 2: Enzyme activities of the inhibitors against *Mtb* SAHH.

Compound	IC$_{50}$	MIC	logP
Adenosine	~500 μM	not determined	-1.6
Aristeromycin	0.2 μM	>100 μM	-1.0
2-fluoro adenosine	66 μM	32 μM	-1.4
3-deaza adenosine	20 μM	>100 μM	-1.3

IC$_{50}$'s were determined by enzyme assay; MIC represents minimum inhibitory concentration against *M. smegmatis* MC^{2}155 cells in a whole-cell assay. LogP values (estimated partition coefficients) were determined by the XLogP v2.0 method of Wang et al. [128].

This is commensurate with inhibition activity reported for ARI against SAHH from other organisms (1.1 μM vs. *H. sapiens*, 3.1 μM vs. *P. falciparum* [97]). The strong inhibitory activity of ARI can be explained based on the mechanism. The proton abstraction from C4' of ADO is assisted by the presence of an electronegative keto group on one side (at the C3') and the ribosyl oxygen on the other side, which makes the C4' proton more acidic. In ARI, the ribosyl oxygen is replaced by a less electronegative carbon and hence the intermediate gets stuck in the 3' keto form. ARI was the most potent inhibitor of the three with an *in vitro* IC$_{50}$ value of 0.2 uM.

Fluoroadenosine (2FA) was chosen for the crystallization studies based on reports of good enzyme inhibition by a similar molecule in the *P. falciparum* SAHH enzyme. The reported IC$_{50}$ value for 2-Fluoro noraristeromycin in *P. falciparum* SAHH is 13 uM. The complex with 2FA (**Figure 10**) is similar to that of ADO, including the pucker of the ribose ring, and the position of the adenine ring (RMSD = 0.22 Å). The

fluoro group fits snuggly into a hydrophobic pocket, surrounded by Phe425, Leu68, and the backbone carbonyl of His69 (all at 3.13 - 3.30 Å distance). However, the closest contact is with the hydroxyl group of Thr74, to which the fluorine atom forms a hydrogen bond (2.48 Å). While not as potent as ARI, 2FA exhibits better inhibition (66 μM IC_{50}) compared to ADO, which was only found to inhibit the enzyme at concentrations above 500 μM (i.e. via product inhibition) (**Table 2**). Thus the fluorine atom, which is the only chemical difference, has a significant effect on binding affinity. A similar IC_{50} for 2FA has been observed for the human SAHH homolog (63 μM) [97]. However, in *P. falciparum*, polar substitutions at the 2-position have been found to have increased relative inhibition, which was attributed to the replacement of Thr60 (human) by Cys59, creating additional space in the ADO-binding pocket [97]. Importantly, however, in contrast to ARI, 2FA was found to have bacteriocidal activity against *M. smegmatis* in a whole-cell assay (MIC of 32 μM, whereas ARI and DZA had MICs greater than 100 μM, **Table 2**). Fluorine substituted nucleosides are known to be better antibacterials [129]. This may reflect differences in cell-wall permeability or uptake mechanism (e.g. by an active nucleoside transporter, which might not recognize carbocyclic compounds like ARI).

The complex with DZA (**Figure 10**) superposes almost perfectly onto the ADO complex, with an r.m.s.d. of 0.25 A° between the ligands. The N6 amino group still hydrogen bonds with the carbonyl oxygen atom on Gln73 (3.02 A°). The C3 position of the adenine ring (substituted for N3) is 3.63–4.52 A° away from Phe425, Leu68, and Thr74. The IC_{50} for DZA is 20 mM (**Table 2**). DZA did not show whole-cell bactericidal

activity, which again could be due to inability of DZA to exploit the same uptake mechanism as 2FA.

Comparison with the Human and **P. falciparum** *SAHH*

Although the three inhibitor molecules that the *Mtb* SAHH was crystallized with had their advantages, none of them represented a truly specific inhibitor of *Mtb* SAHH. The primary reason for this was that the product binding site was identical to the Human SAHH and offered almost no avenues for the design of novel specific inhibitors of *Mtb* SAHH. Hence the focus had to be shifted to other areas of the enzyme.

When the *Mtb* SAHH active site was overlaid on top of the *P. falciparum* SAHH active site, a curious observation was made (**Figure 11**). The active site residue His345 of *P. falciparum* SAHH was flipped out away from the 5' hydroxyl group of Ado. The analogous residue in *Mtb* SAHH, His363, on the other hand was pointing towards the 5' hydroxyl group of Ado and was within hydrogen bonding distance. This curious observation lead to the hypothesis that His 363 of *Mtb* SAHH may act as a gate, opening up to allow access to the substrate, SAH to enter and the product, homocysteine (Hcy) to leave the active site. If this were true, the definition of the active site would have to be expanded to beyond just the Ado binding region and to include the region occupying the Hcy moiety of the substrate SAH as well. The Hcy binding region may unveil differences in the architecture of the active site between the Human and *Mtb* SAHH, thereby opening up avenues to exploit the differences and design more specific inhibitors of *Mtb* SAHH.
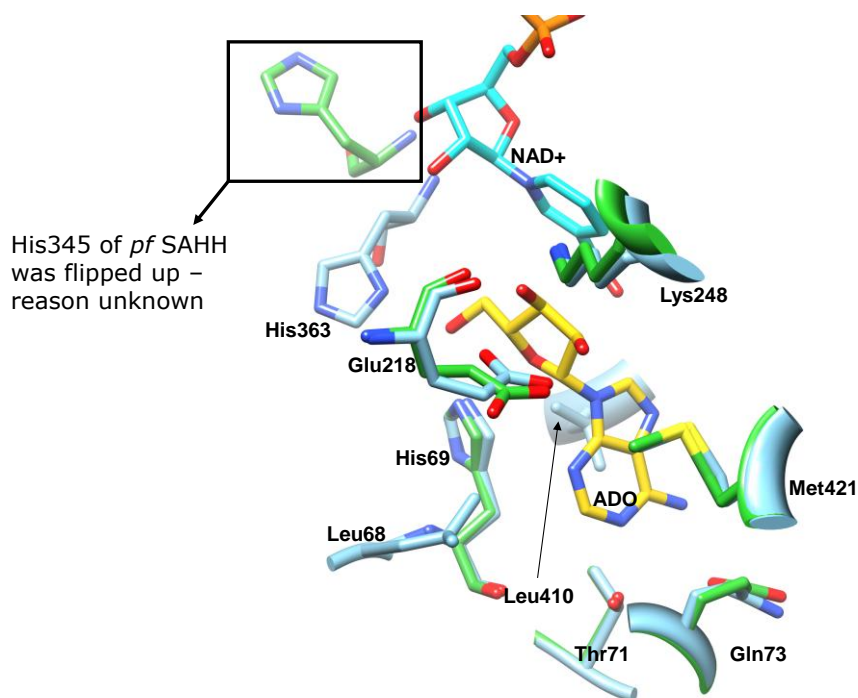
Figure 11: Superposition of the active sites of *Mtb* SAHH and *Pf* SAHH. The *Mtb* SAHH residues are shown in cyan and the *Pf* SAHH residues are shown in green color. ADO is shown in yellow stick representation.

Mtb *SAHH bound to NAD$^+$ and SAH*

To prove the above hypothesis, we crystallized *Mtb* SAHH in the presence of the substrate SAH. Since SAHH catalyzes the hydrolysis of SAH to Ado and Hcy in the presence of NAD$^+$, whenever *Mtb* SAHH was crystallized with SAH the crystal structure showed the presence of Ado bound in the active site. Hence to get SAH to bind and remain in the active site required the soaking of SAHH crystals with high concentrations of SAH for short periods of time. Even then the occupancy of SAH in the active site was low and part of the homocysteine moiety was visible. Thus, SAH was modeled into the

active site as ethylthioadenosine. With the truncated SAH in the active site the binding

mode of the substrate was apparent and clearly exposed the residues involved in the

binding of the Hcy moiety of the substrate. Density corresponding to the covalently-

attached HCY was observed in an access channel (**Figure 10 and 12**), adjacent to the 5'

position on the Ado.
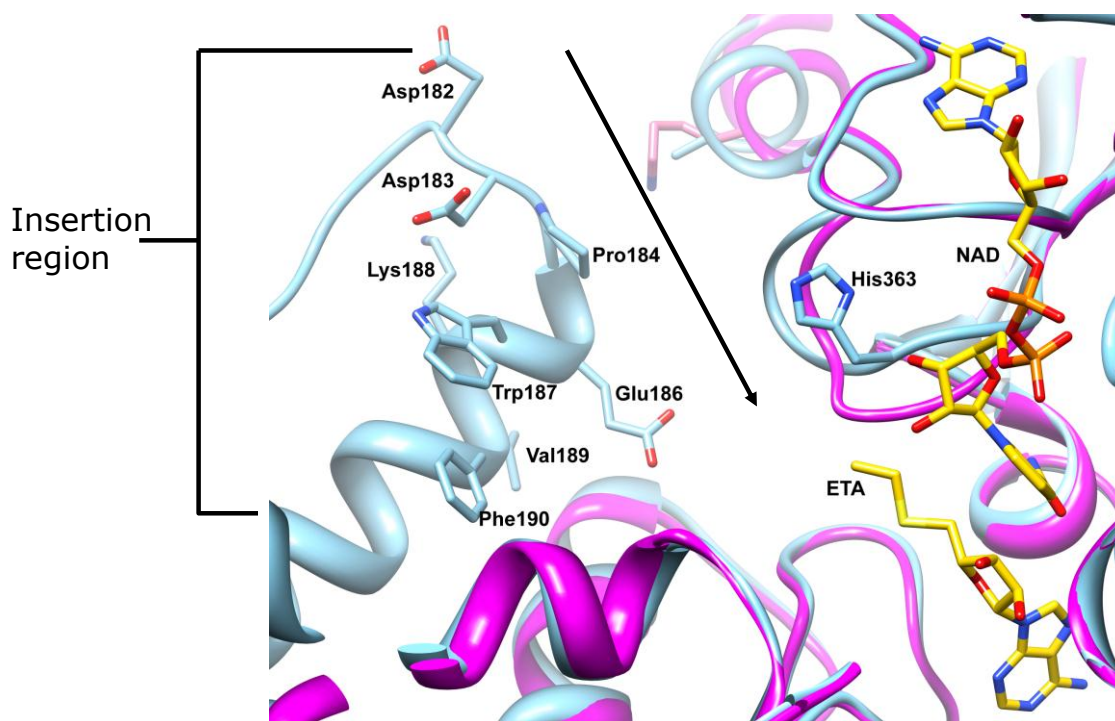
Figure 12: SAH bound *Mtb* SAHH showing the access channel from the SAH to the surface of the protein. Human SAHH is overlaid and represented in magenta ribbons. *Mtb* SAHH is represented in cyan ribbon. The arrow indicates the access channel from the surface of the enzyme to the active site. The region of the access channel contributed by the 37 residue insertion region is highlighted and the prominent residues displayed. ETA and NAD$^+$ bound in the active site are shown in yellow stick representation.

The channel is formed by rotation of the peptide plane between Cα363 and Cα364 by nearly 180°, swinging the imidazole ring of His363, which was at 2.69 Å from the 5'OH of ADO (SAHH:ADO:NAD structure), to 8.3 Å away from the Sδ of SAH (SAHH:SAH:NAD structure). This rotation opens up an access channel to the buried ADO-binding pocket that can accommodate the HCY moiety of the SAH substrate (compare **Figures 13A and 13B**). The channel leads to the surface (approximately 20 Å away), and could permit release of the liberated HCY while the rest of the catalytic cycle continues (i.e. hydration at C5', followed by reduction of keto at C3'). Sδ of SAH makes a hydrogen bond with His69 (3.03Å). The Cβ and Cγ atoms make van der Waals contacts (3.66 - 4.09Å) with Gly362, Thr219, Leu407, and the 2'OH of the NAD. Phe364 is also close-by (4.65Å), contributing to the hydrophobicity of the pocket. Density was not well-defined enough to accurately determine coordinates for the terminal amino and carboxyl groups (and Cα) of HCY. Therefore the ligand was modeled, refined, and deposited in truncated form as ethylthioadenosine (ETA), which was taken to partially represent SAH. The Cβ atom of HCY occupies the same position as the imidazole ring of His363 in the other complexes (**Figure 13A and 13B**). Therefore, to accommodate the binding of SAH, His363 is forced to flip out of the channel, rotating approximately 180 degrees around the backbone. Importantly, binding of SAH clearly proved the hypothesis that His363 acts as a gate to allow the substrate to bind. The side chain of His363 in the *Mtb* SAHH:SAH complex was rotated about 180 degrees relative to its position in the Ado bound *Mtb* SAHH crystal structure.

Solvent Access Channel of **Mtb** *SAHH*

The "opening up" of the His 363 gate revealed an additional active site for the binding of the Hcy moiety. Additionally, a solvent access channel that leads from the surface of the enzyme to the active site was also discovered. Comparisons could now be made among the residues lining up this additional active site and the adjoining access channel between the *Mtb* and Human SAHH. The crystal structure of *Mtb* SAHH was superimposed with the human homolog (PDB ID: 1A7A; complexed with DHCeA [89]), and the active sites were examined to look for differences that could be exploited for design of selective inhibitors. All of the residues lining the ADO-binding pocket are identical between *Mtb* and human, limiting the possibilities for inhibitor design. This even includes Thr74, which is equivalent to Thr60 in human, though is replaced by Cys59 in *P. falciparum* and has been investigated for design of selective anti-malarial drugs [97]. The only exception is Gln73, which replaces Glu59 in the human enzyme, though both form hydrogen bonds with the N6 amino group of the adenine moiety. While the residues binding the Hcy binding site were conserved between the *Mtb* and Human enzymes, there were significant differences in the access channel architecture between the *Mtb* and Human enzymes (**Figures 12, 13A and 13C**). When this structure is compared with the human SAHH, many residues lining the putative access channel (HCY-binding region) are also found to be conserved, including Cys93, Asn94, Tyr133, Gly157, Asp159, Phe364, and Arg406. The residues corresponding to Phe364 (Phe 302) and Arg 406 (Arg 343) in the human SAHH were previously proposed to be involved in hydrophobic and hydrogen bonding interactions respectively with the side chain of

homocysteine [126]. Upon analysis of the access channel, the importance of the 37 residue insertion region in the *Mtb* SAHH, which until now was unclear, became apparent. This insertion segment completely changed the architecture of the active site of *Mtb* SAHH relative to the Human SAHH. Most notably, the mouth of the channel is significantly more constrained in *Mtb* than in human (compare **Figures 12 and 13A and 13C**).

First, there are several substitutions by large, bulky hydrophobic residues that build up the wall of the channel, including Trp134 (Leu111 in human), Met162 (Asn137), Arg225 (Asn163), Phe190, and Trp187. The latter two are contributed by the 37-residue insertion (surface helix) that is unique to *Mtb* (relative to the human sequence). While in the Human SAHH the access channel is more exposed to the environment being shallow and funnel like, the insertion residues transformed the access channel in the *Mtb* SAHH into a deep and narrow entry way (**Figure 12 and 13**). In addition, the insertion region projects the side-chain of Glu186 into the access channel. The distance from the carboxylate group on Glu186 to the 5'OH on ADO is 12.66 Å (**Figure 13B**). While this residue was not initially considered relevant [97], its proximity to the HCY portion of SAH bound in the channel (approximately 10Å from Cβ, see **Figure 13A**) makes it a potentially important residue in the design of competitive inhibitors. Such differences in the access channel architecture between the *Mtb* and Human SAHH could be exploited to design small molecules that could bind in the narrow hydrophobic channel of the *Mtb* SAHH and block access to the active site. Such molecules could effectively inhibit the *Mtb* SAHH.
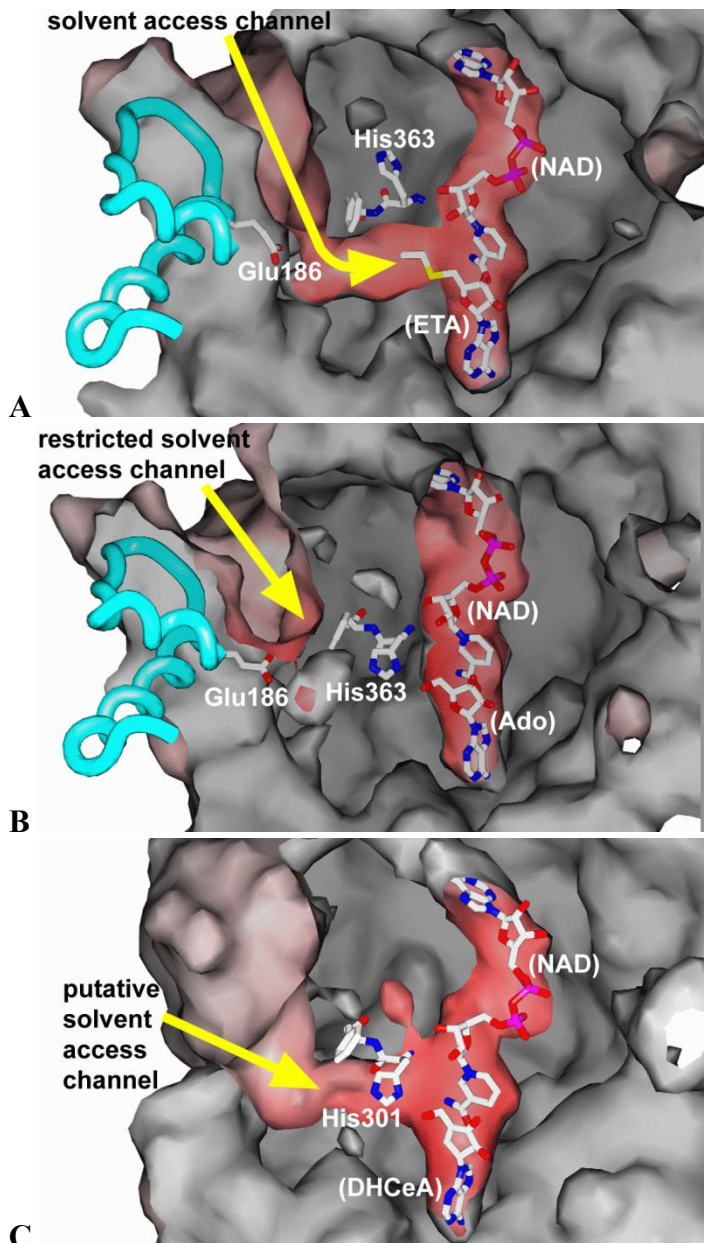
Figure 13: Differences in human and *Mtb* forms of SAHH enzyme near the entrance to the access channel / HCY-binding site.

A) *Mtb* SAHH bound to ETA and NAD showing a constricted solvent access channel lined by residues from the insertion region. Residues 167-203 from strand-turn-helix insertion are shown in cyan, with Glu186 highlighted. The figures were made using SPOCK [130].

B) Surface of *Mtb* SAHH binding site with access channel closed off by His363, and ADO enclosed in a buried pocket.

C) Human SAHH complexed with DHCeA and NAD showing a shallower solvent access channel compared to the *Mtb* SAHH. The surface of the active site has been drawn without His301 to simulate the open form of the channel. His301 would presumably flip out of the channel, allowing the SAH substrate to bind and HCY cleaved from the substrate to be released to solvent.

64

Fragment Based Screening

In order to search for small molecules that could bind in the narrow hydrophobic *Mtb* SAHH access channel a fragment based approach was used. The advantage of the fragment based screening over regular high throughput screening (HTS) approach using large chemical compound libraries is that a relatively smaller library of compounds can be used to cover the chemical space. A fragment library consists of compounds with molecular weights of ≤ 300 g/mol and the molecules have at the most three functional groups. As a result the fragments are capable of making fewer interactions than a larger molecule with higher molecular weight and much functionality, which in turn results in lower binding affinities for the fragments. The advantage however is that the smaller size allows the fragments to sample many different orientations in the active site and bind in its most optimal orientation whereas a large molecule would have no such flexibility in sampling the active site.

The fragment library consisted of 750 fragments ranging in molecular weights from 96 Da to 364 Da. Screening was performed using the differential scanning fluorimetry (DSF) technique. The first step of the process was to determine the optimal

buffer for subsequent screening. The different buffer conditions that were tested included citric acid (pH 2.2 – 3.8), sodium acetate (pH 4.0 – 4.8), sodium citrate tribasic (pH 5.0 – 5.8), sodium cacodylate (pH 6.0 – 6.8), sodium salt of 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (Na HEPES) (pH 7.0 – 7.8), Tris (pH 8.0 – 8.8), 3-(Cyclohexylamino)-2-hydroxy-1-propanesulfonic acid (CAPSO) (pH 9.0 – 9.8) and 3-(Cyclohexylamino)-1-propanesulfonic acid (CAPS) (pH 10.0 – 11.0). Through comparison of the melting temperature curves with all the buffers, the Na-HEPES buffer at pH 7.0 was found to be the most consistent. Na HEPES buffer facilitated a sharp unfolding event at 43.6 °C, which corresponded to the melting temperature of the protein. Keeping this buffer constant, the fragment library was then screened at 5 mM fragment concentration, while the enzyme concentration was empirically arrived at 10 μM. Since the aim of this experiment was to identify a fragment (or fragments) that bound to the narrow access channel leading up to the active site, the assay was set up in the presence of the natural substrates of the enzyme, namely NAD^+ and adenosine. Incubating the enzyme with the natural substrates was an attempt to ensure that the active site of the enzyme would be inaccessible to the fragment molecules being screened. As the fragments typically bind with low millimolar affinity, they would also be unable to replace the natural substrates in the active site. The only other binding site accessible for the fragment molecules was the narrow access channel. The hydrophobic nature of the residues lining the access channel and the small surface area available to bind the fragments were in sync with the lipophilic nature of the fragments and their small molecular weights.

While a number of fragment compounds were able to produce between two to five degrees shift in the melting temperature, the curious observation was that none of the 750 fragments produced a positive shift in the enzyme melting temperature. A positive shift in the melting temperature suggests that the protein unfolding event takes place at a higher temperature, which is indicative of a stabilizing interaction between the ligand and the protein. This is usually observed when a fragment binds in the active site of the protein, thereby stabilizing the protein's native conformation. On the other hand, a negative shift in the melting temperature suggests that the protein – ligand interaction has a destabilizing effect on the protein's native conformation. This is usually observed when the protein has to adopt an unnatural conformation in order to accommodate the ligand binding. The following seven fragments summarized in Table 3 were identified by DSF experiment to have the most shifts in Tm with *Mtb* SAHH enzyme. The interactions with hit fragments were reasoned as destabilizing to the protein because they lowered the melting temperature of the protein in the DSF studies. The change in melting temperature was $\geq 5\ ^{\circ}C$. A closer look at the fragment identified through DSF studies revealed that all of the molecules contain two closed ring systems. All fragments contained at least one six membered aromatic ring with either one or no substitution. The closed ring systems were either fused together or separated by a single bond as in a biaryl system. The polar substituents included amide, chlorine, hydroxyl, tifluoro methyl or oxime. The nonpolar substituent included methyl group.

Table 3: Fragment hits identified through DSF studies along with their molecular structures, molecular weights and the ΔTm values.
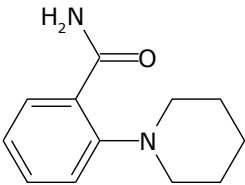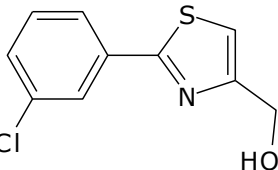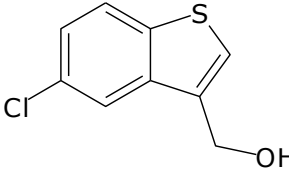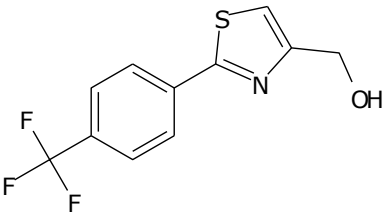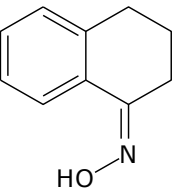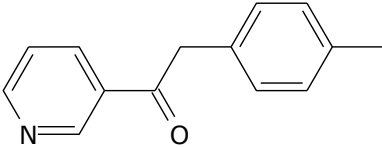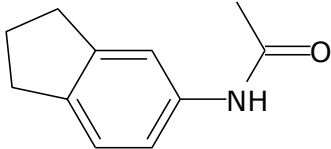
#	Name	Molecular structure	M.W.	Tm	ΔTm (°C)
1	2-(piperidin-1-yl)benzamide		204.3	36.6	7
2	(2-(3-chlorophenyl)thiazol-4-yl)methanol		225.7	36.6	7
3	(5-chlorobenzo[b]thiophen-3-yl)methanol		198.7	38.0	5.6
4	(2-(4-(trifluoromethyl)phenyl)thiazol-4-yl)methanol		259.3	37.1	6.5
5	(E)-ydronaphthalen-1(2H)-one oxime		161.2	36.2	7.4

Table 3: Continued.

#	Name	Molecular structure	M.W.	Tm	ΔTm (°C)
6	1-(pyridin-3-yl)-2-p-tolylethanone		211.3	38.0	5.6
7	N-(2,3-dihydro-1H-inden-5-yl)acetamide		175.2	34.6	9

Saturation transfer difference nuclear magnetic resonance spectroscopy (STD NMR) technique was used to validate the hits obtained through DSF studies. Each hit fragment was tested using the protocol mentioned in the methods section. 500 mM stock solutions of the hit fragments were prepared in 100 % DMSO and added to the separate NMR samples containing the enzyme and buffer to a final concentration of 3 mM. Out of the seven fragments, five fragments displayed ^{1}H chemical shift in the STD NMR experiment. The five fragments were (#1) 2-(piperidin-1-yl)benzamide, (#2) (2-(3-chlorophenyl)thiazol-4-yl)methanol, (#4) (2-(4-(trifluoromethyl)phenyl)thiazol-4-yl)methanol, (#6) 1-(pyridin-3-yl)-2-p-tolylethanone and (#7) N-(2,3-dihydro-1H-inden-5-yl)acetamide (see **Table 3**). An example of the normal ^{1}H NMR spectrum and the STD NMR spectrum for the fragment (#1) 2-(piperidin-1-yl)benzamide is shown in Figure 14 and 15.
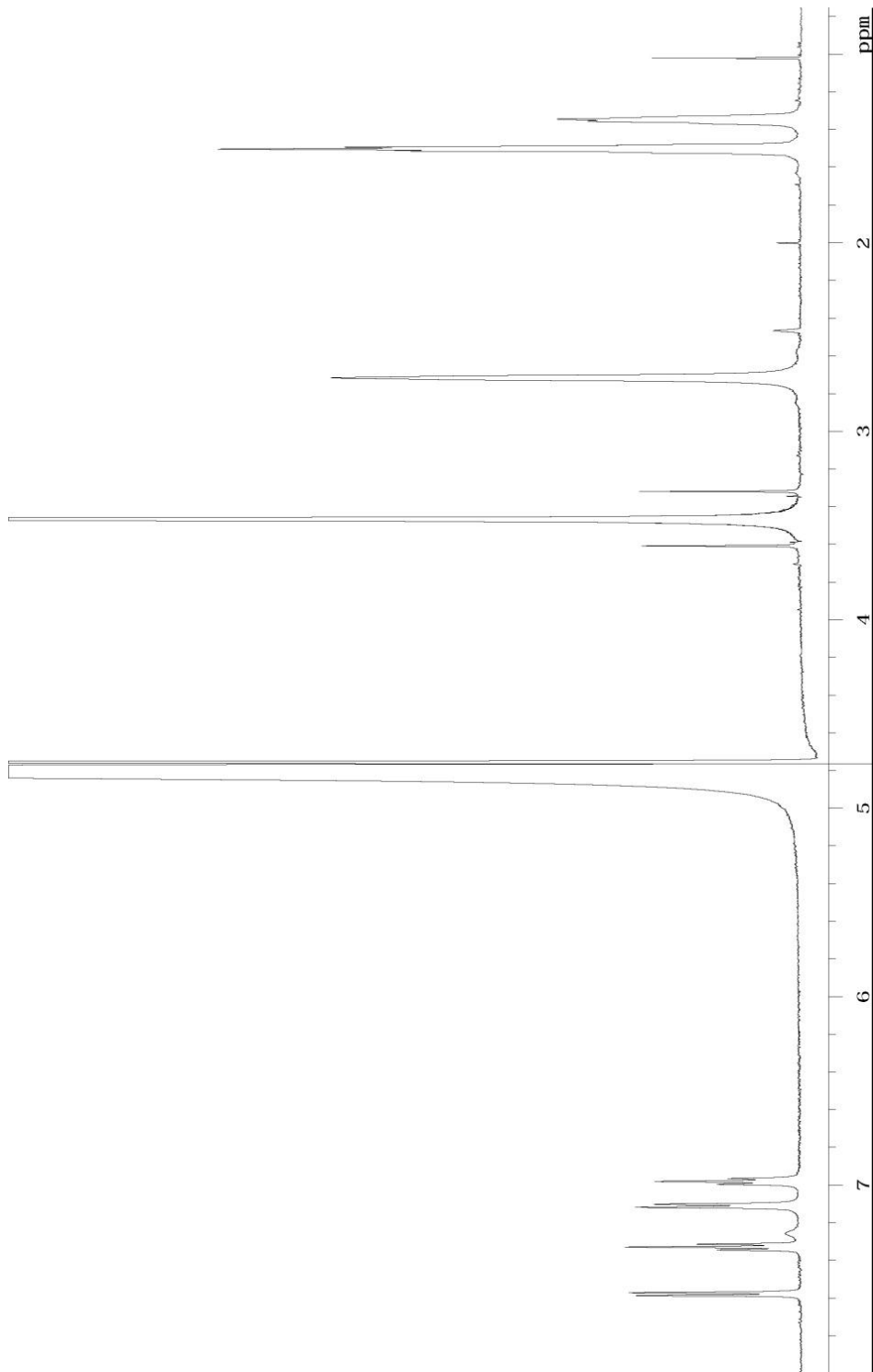
Figure 14: ^{1}H NMR spectrum for fragment #1, 2-(piperidin-1-yl)benzamide. The chemical shift observed at δ 3.5 ppm belongs to impurities in the mixture and the shift observed at δ 4.8 ppm belongs to the solvent.
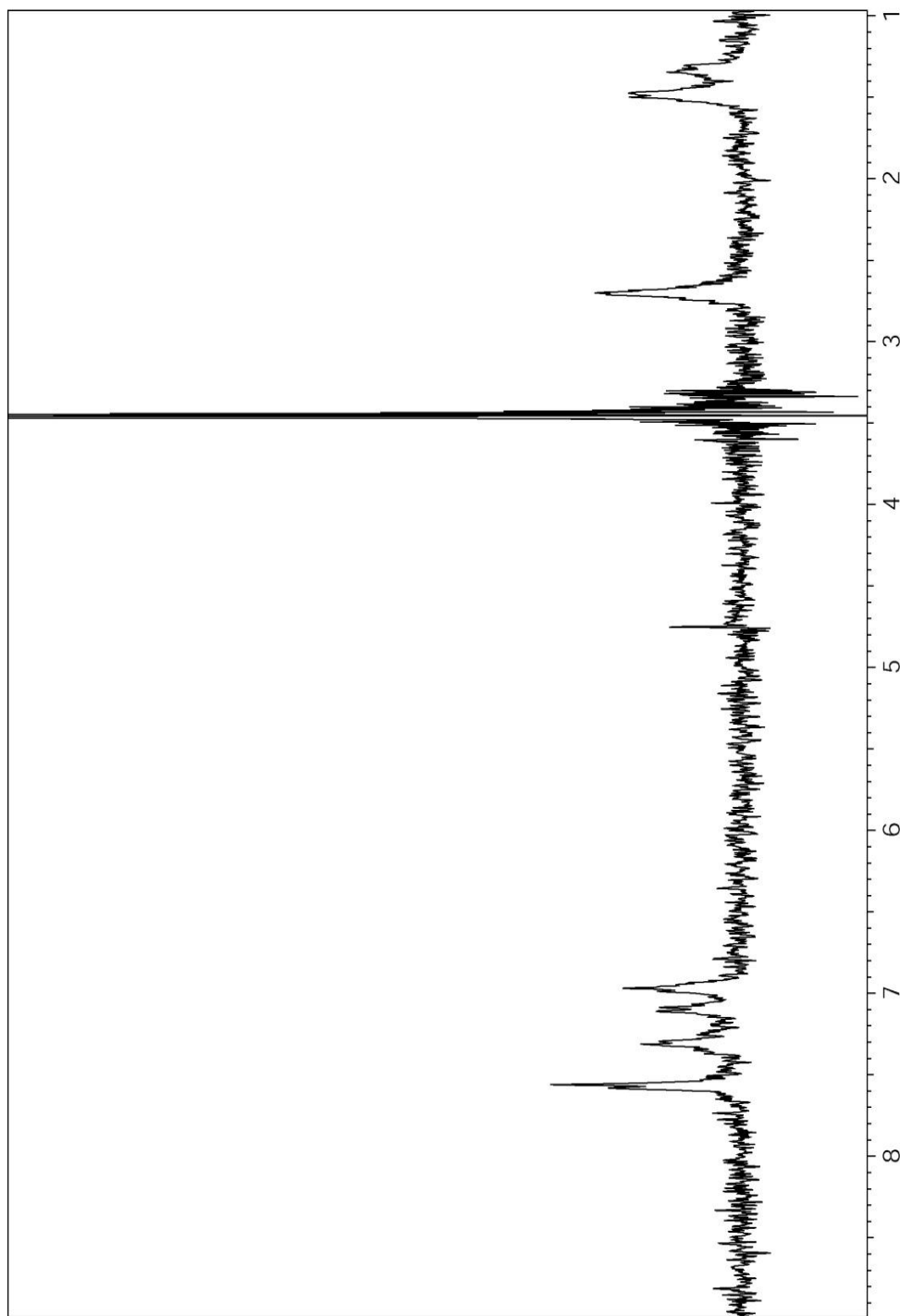
Figure 15: STD NMR signal for the fragment 2-(piperidin-1-yl)benzamide (3 mM concentration). The four peaks observed in the δ 6.8 ppm to δ 7.8 ppm correspond to the aromatic group of 1, 2-(piperidin-1-yl)benzamide. The chemical shift observed at δ 2.7 ppm corresponds to the protons that are adjacent to the nitrogen in the piperidine ring. The chemical shift observed at δ 1.3 and δ 1.4 ppm corresponds to the remaining protons on the piperidine ring.

The aim of fragment based screening was to find small hydrophobic molecules that could bind in the narrow access channel of the *Mtb* SAHH enzyme and block the channel thereby inhibiting the enzyme function. Therefore the fragment based screening through differential scanning fluorimetry was conducted with the enzyme preincubated with NAD^+ and ADO. While the natural substrates occupied the active sites, it was reasoned that the weak binding fragment molecules possibly will interact with the residues in the access channel. In order to confirm the independent binding event of the fragments and the natural substrate ADO, a control experiment was performed where the concentration of ADO was varied while the concentration of the fragment 2-(piperidin-1-yl)benzamide was kept constant at 3 mM (**Figures 16, 17 and 18**). Three different concentrations of ADO were used, 0.3 mM, 3 mM and 30 mM. It was observed that while the NMR peak intensity of ADO increased at increasing concentrations, the peaks corresponding to the aromatic protons on 2-(piperidin-1-yl)benzamide were observed at a constant intensity in all three instances.
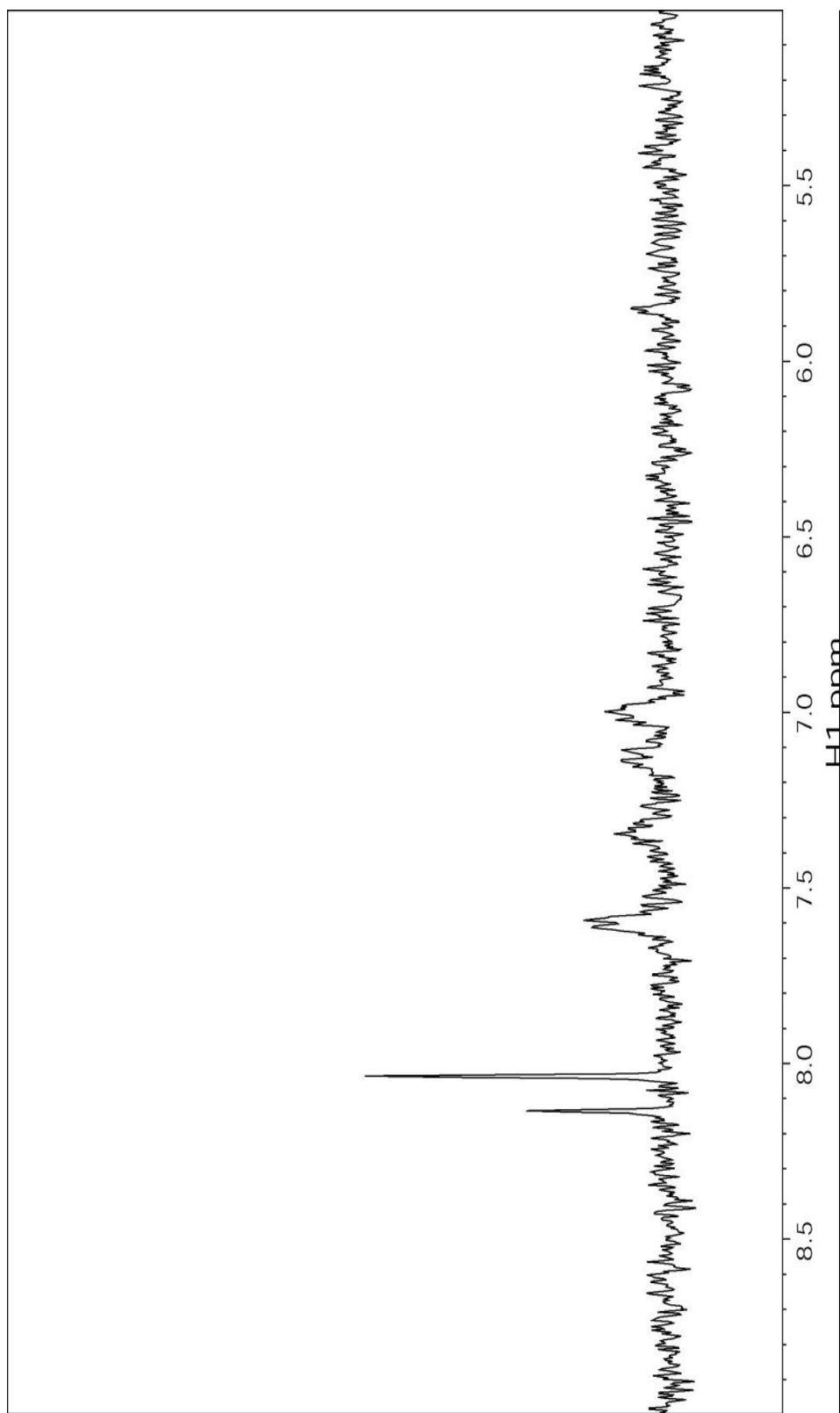
Figure 16: STD NMR signal for the fragment 2-(piperidin-1-yl) benzamide in the presence of 0.3 mM ADO. The peaks at δ 6.8 ppm to δ 7.8 ppm correspond to the fragment 2-(piperidin-1-yl) benzamide and the peaks at δ 8.0 ppm to δ 8.5 ppm correspond to the fragment 2-(piperidin-1-yl) benzamide and the peaks at δ 8.0 ppm to δ 8.5 ppm correspond to ADO.
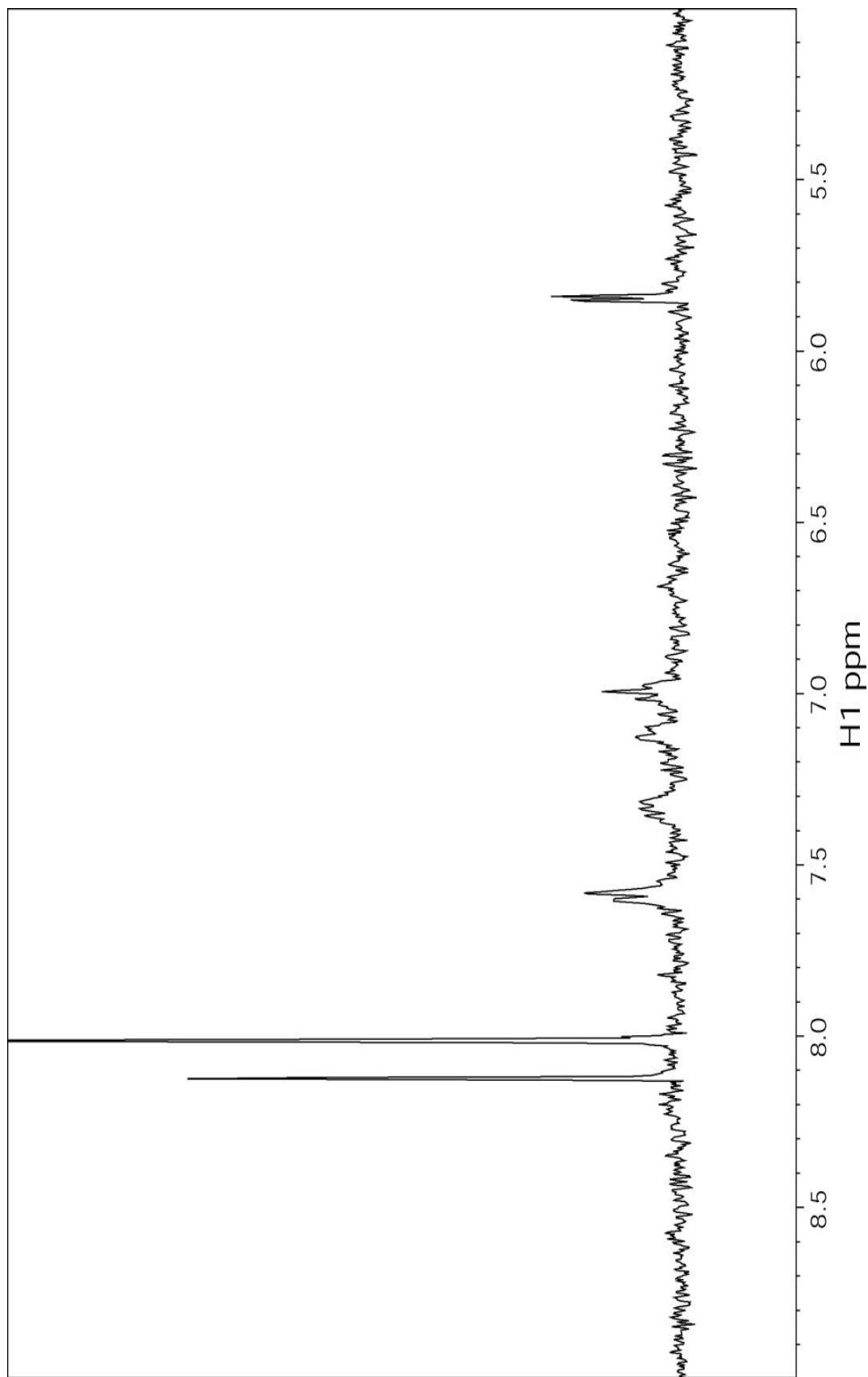
Figure 17: STD NMR signal for the fragment 2-(piperidin-1-yl)benzamide in the presence of 3 mM ADO. The peaks at δ 6.8 ppm to δ 7.8 ppm correspond to the fragment 2-(piperidin-1-yl)benzamide and the peaks at δ 8.0 ppm to δ 8.5 ppm correspond to ADO.
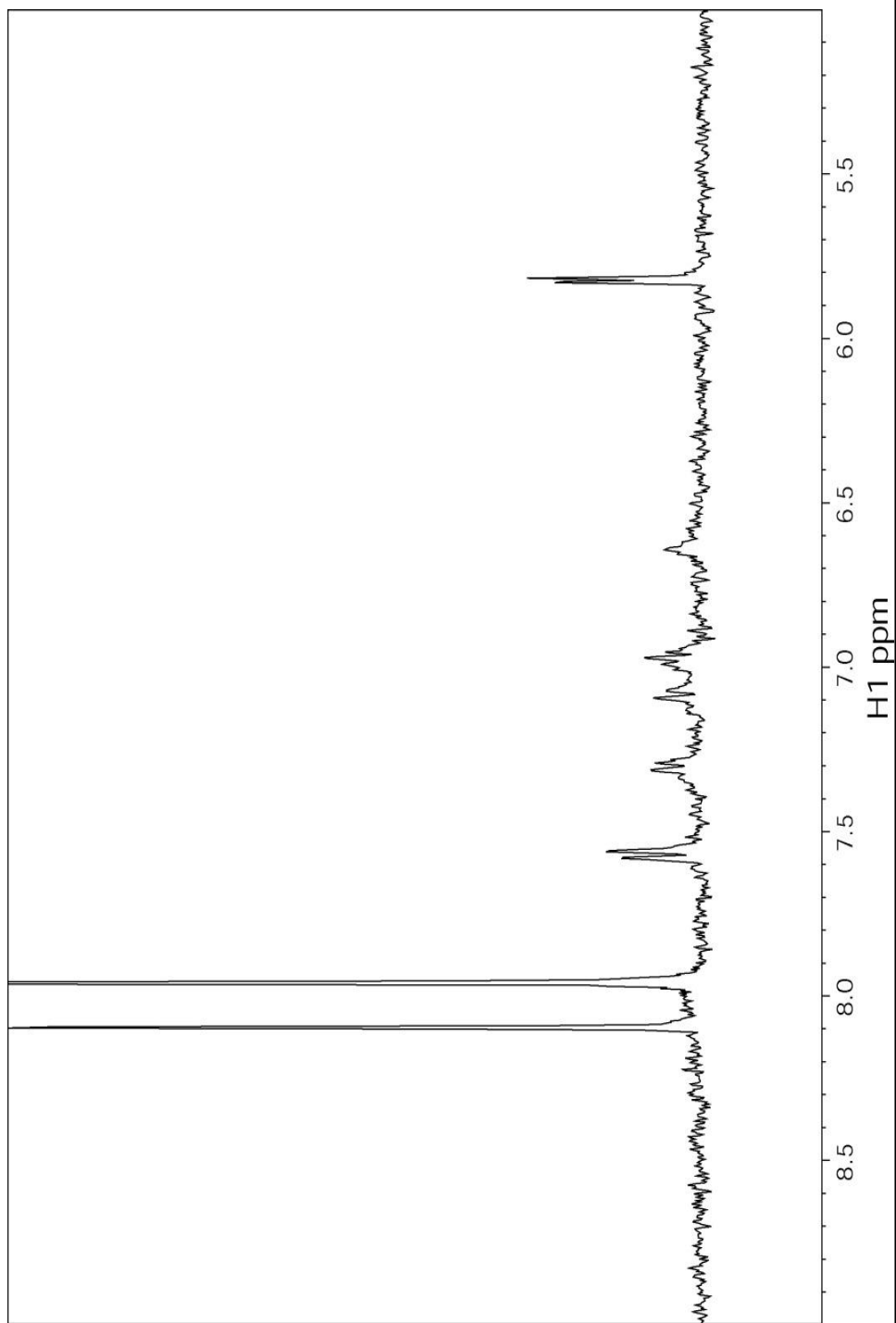
Figure 18: STD NMR signal for the fragment 2-(piperidin-1-yl)benzamide in the presence of 30 mM ADO. The peaks at δ 6.8 ppm to δ 7.8 ppm correspond to the fragment 2-(piperidin-1-yl)benzamide and the peaks at δ 8.0 ppm to δ 8.5 ppm correspond to ADO.

This observation provided evidence that the binding of adenosine and the fragment were two independent events. Analysis of the active site and the access channel leading up to the active site shows that the binding site for the fragment was most likely in the access channel. The side chains of Pro184 and Phe364 could provide the hydrophobic environment for nonbonding interactions with the aromatic ring of 2-(piperidin-1-yl)benzamide while the polar side chains of Glu186 or Asp183 could make hydrogen bonding interactions with the amide functional group on 2-(piperidin-1-yl)benzamide. Attempts to soak the fragments into crystals of *Mtb* SAHH or to co-crystallize the fragments with *Mtb* SAHH were unsuccessful. The fragments identified in this research represent a good starting point to build upon the chemical scaffold to develop a lead compound with tight binding affinities.

Conclusion

The crystal structure of the SAHH enzyme from *M. tuberculosis*, in complex with ADO, has been solved at 1.6 Å. The structure is largely similar to previously reported SAHH structures from human, rat, and malaria, exhibiting an α/β fold with two domains: an N-terminal substrate-binding domain bound to NAD^+, and a C-terminal dinucleotide-binding domain. The protein is in a closed conformation, with the two domains rotated to enclose the ADO ligand in a buried pocket. The *Mtb* SAHH has a 37-residue loop that forms a helix-turn-strand on the surface, similar to that found in the *Pf* SAHH, but not found in the mammalian structures. Complexes with three inhibitors were also reported: ARI, 2FA, and DZA. The 2FA and DZA complexes demonstrate

expected interactions between the ligands and side-chains lining the pocket, as seen in previous structures. However, the ARI complex is the first reported structure of SAHH with ARI, and shows a 3' keto group with planar carbon in the cyclopentanyl ring, confirming the prediction that it is oxidized by NAD^+ as a mechanism-based inhibitor. While most of the residues lining the ADO-binding pocket are identical between *Mtb* and human SAHH, allowing few opportunities for design of selective inhibitors, less is known about the binding mode of the HCY appendage of the full substrate. We determined the co-crystal structure of SAHH with SAH. Contiguous density was found adjacent to the 5' position of ADO and modeled as atoms in the body of covalently attached HCY. Binding of SAH forces a rotation of His363 around the backbone to flip out of contact with the 5' position of ADO and opens up access to a nearby channel that leads to the surface. The structure, the first complex with the SAH substrate reported, suggests that His363 acts as a switch that opens up to permit binding of SAH, and closes down after release of the HCY appendage. While most of the residues in direct contact with the HCY are conserved between human and *Mtb* SAHH, the mouth of the access channel nearby is significantly different, with the *Mtb* SAHH having a much more enclosed entrance surrounded by hydrophobic residues, along with Glu186 (not found in human SAHH). If the active site is expanded to include this HCY-binding access channel, these differences could potentially be exploited to design inhibitors that compete with substrate binding and moreover are selective against *Mtb*.

An attempt was made to exploit this difference between the *Mtb* and Human SAHH by directing a fragment based screening approach to find small molecules that

could potentially bind in the *Mt*b SAHH access channel thereby blocking the channel for entry of the substrate or exit of the products of the hydrolysis reaction thereby inhibiting the enzyme function. Fragment-based screening presented an enticing opportunity to find lead compounds that could bind in the narrow access channel leading up to the active site. With its collection of small molecules with an average molecular weight ≤ 300 daltons, and lipophilicity (ClogP) value ≤ 3, the fragment library was well suited for this investigation. 750 fragments were screened by the method of differential scanning fluorimetry. Through measuring the differences in the melting temperature of the apo enzyme and the fragment bound enzyme it was possible to identify eight candidates which had a melting temperature difference of 5° C or more. STD NMR technique was successfully employed to validate five out of the eight fragments as true binders of *Mtb* SAHH enzyme. The five fragments were; 2-(piperidin-1-yl)benzamide, (2-(3-chlorophenyl)thiazol-4-yl)methanol, (2-(4-(trifluoromethyl)phenyl)thiazol-4-yl)methanol, 1-(pyridin-3-yl)-2-p-tolylethanone and N-(2,3-dihydro-1H-inden-5-yl)acetamide.

The purpose behind the fragment based screening was to identify small hydrophobic molecules that could potentially bind in the deep and narrow access channel provided by the *Mtb* SAHH. Fragment screening was therefore conducted in the presence of the natural substrates, NAD^+ and ADO bound into the active site of *Mtb* SAHH. STD NMR technique was used to prove the independent binding of the fragment and the natural substrates. Experiments were conducted using a fixed concentration of the fragment, 2-(piperidin-1-yl)benzamide while the concentration of ADO was varied

from 0.3 mM to 30 mM. Three different STD NMR measurements were taken at 0.3 mM, 3 mM and 30 mM concentrations of ADO. The STD NMR peaks corresponding to the fragment were consistently observed with increasing concentrations of ADO, suggesting that the binding of the fragment was independent of substrate binding. In the absence of a crystal structure of *Mtb* SAHH in complex with the fragments it may be speculated that the fragments bind into the access channel of *Mtb* SAHH away from the active site. The access channel offers hydrophobic residues like Pro184, Ala185 and Trp187 that are unique to *Mtb* SAHH.

Future Work

Crystal structures of the *Mtb* SAHH in complex with the hit fragments would explain the mode of binding of the fragments and the specific interactions they make in the active site or the channel leading up to the active site.

The goal of any fragment based screening is to identify fragments that can bind in the active site of an enzyme and then build up upon the fragments to exploit the surrounding active site residues available to make hydrogen bonding or hydrophobic interactions. Once the fragment binding position in the active site is confirmed through solving the crystal structure of the complex, medicinal chemistry efforts can be made to modify the fragment into a lead compound. Chemical modifications can be made to the hit fragments to either connect two or more fragments together to make a lead compound or add more functionality to the fragment molecules such that they can have additional

interactions with the active site residues in the binding environment resulting in tight

binding affinities.

CHAPTER III

MTB NITROGEN REGULATORY PII PROTEIN

Background and Introduction

The PII family of small signal transduction proteins play a pivotal role as nitrogen sensors in microorganisms and plants [49,50]. Nitrogen sensing is achieved by regulating the intracellular nitrogen to carbon ratio represented largely by glutamine and 2-oxoglutarate (2OG) respectively. Intracellular nitrogen regulation is carried out by direct and indirect control of glutamine synthetase (GS) activity through PII proteins. *E. coli* has two homologues of the PII protein, GlnB (also referred to as PII), encoded by the *glnB* gene, and GlnK, encoded by the *glnK* gene. Under nitrogen limiting conditions, *E. coli* GlnB is uridylylated by a uridyl transferase (GlnD), and then interacts with adenyl transferase to directly activate GS through deadenylylation [47,131]. Conversely, under nitrogen rich conditions, *E. coli* GlnB is deuridylylated by GlnD, and then interacts through the adenyl transferase to deactivate GS by adenylylation. GlnB has also been shown to form a direct complex with GS in its activation/deactivation cycle in Archaeal *methanosarcina mazei* [132].

In the indirect control system, *E. coli* GlnB interacts with the two component

signal transduction system, NtrB and NtrC, to regulate the transcription of the gene *glnA*

which encodes for GS [133,134]. When GlnB is deuridylylated it can interact with a histidine

kinase, NtrB, which regulates the phosphorylation state of NtrC, controlling the

expression of *glnA*. Thus GlnB is indispensible to the pathway of nitrogen control in *E.

coli*.

GlnK is another PII homolog that also plays an essential role in the nitrogen

control systems of several organisms including *E. coli* [56,135]. In all known instances, the

glnK gene lies in an operon immediately next to the *amtB* gene [136], which codes for an

integral membrane protein, AmtB, that acts as ammonia/ammonium channel [58]. Under

nitrogen rich conditions in *E. coli* the GlnK trimer complexes with the AmtB trimer

blocking the influx of ammonium into the cellular cytoplasm [59]. GlnK is also implicated

in the membrane sequestration of enzymes involved in nitrogen fixation. For example, in

Azospirillium brasiliense GlnK binds to one of two different proteins depending on the

nitrogen status. During nitrogen fixing conditions GlnK binds DraG, a glycohydrolase

that switches on nitrogenases in *Azospirillium* [137,138], in conditions of nitrogen excess

GlnK binds DraT, an ADP-ribosyl transferase, and switches off these nitrogenases.

Binding of GlnK to DraG or DraT was found not to be competitive with the AmtB:GlnK

complex suggesting that a ternary AmtB:GlnK:DraG/T complex exists [138,139]. The

formation of this complex defines a new regulatory role for GlnK: assisting in the

membrane sequestration of the nitrogenases, DraG and DraT. Another ternary complex

of the AmtB:GlnK with the nitrogen stress transcription factor, TnrA, has also been

reported in *Bacillus subtilis* [140]. Similarly, in the Gram-positive soil bacterium *Corynebacterium glutamicum,* GlnK interacts with the transcriptional repressor, AmtR [141], and in *Streptomyces coelicolor* the GlnK:AmtB complex may interact with an OmpR type transcriptional activator, GlnR, that is involved in nitrogen metabolism [142,143]. Thus the AmtB:GlnK complex is capable of functioning with many other protein partners like DraG, DraT, TnrA, AmtR, and GlnR to achieve nitrogen regulation.

Crystal structures of GlnB and GlnK have been reported for organisms such as *E. coli* [144-146], *Methanococuus jannaschii* [147], *Thermus thermophilus* [148], *Herbaspirillum seropedicae* [149], *Synechococcus sp.* [150], and *Arabidopsis thaliana* [151]; GlnB and GlnK have also been shown to form heterotrimers in solution [152,153]. *Mycobacterium tuberculosis* (*Mtb*) has a single *PII* homolog, Rv2919c [30]. The *PII* gene was found to be essential for survival in primary murine macrophages by transposon site hybridization (TraSH) in H37Rv [154]. Its essentiality, the various roles that the protein is proposed to play in nitrogen regulation, and its absence in humans makes it an attractive drug target in *Mtb,* the causative agent of tuberculosis (TB). Despite its attractiveness as a drug target, the nitrogen regulatory system in *Mtb* is not well understood and no crystal structures of the apo or the substrate bound *Mtb* PII protein exists.

The *Mtb* PII gene was annotated *glnB* based on its higher sequence homology to *E. coli* GlnB (*www.webTB.org*). However, it exists in an operon with *glnD-PII-amtB* (*Rv2918c-Rv2919c-Rv2920c*) leading to speculation that it might encode for a GlnK protein based on the operon organization observed in other microorganisms [50]. The apparent ambiguity requires clear structural and biochemical data to confirm its identity.

As part of the *Mtb* structural genomics consortium, the crystal structure of this protein was solved, both in the apo and ATP bound forms, with the aim of understanding its true identity (and hence its role) in *Mtb*, potentially opening avenues for inhibition studies [22]. The crystal structures presented in this report provide important structural and functional insights into the role of *Mtb* PII as GlnK.

Methods

Cloning, Expression and Purification of* Mtb *PII Protein

A 339 bp DNA fragment containing the *PII* gene (*Rv2919c)* was amplified by PCR with *Mtb* H37Rv genomic DNA as a template, using the following oligonucleotide primers:

5' - GGG AAT TCC ATA TGA AGC TGA TCA CTG CGA TCG TGA AGC - 3'

3' - CCC AAG CTT TCA TAA CGC GTC GTG TCC GCG TTC AC – 5'. The amplified DNA fragment was digested with *Nde*I and *Hind*III restriction enzymes and subcloned into the corresponding restriction sites in a pET28b vector containing an N-terminal His tag (Novagen). Following sequence confirmation, the plasmid was transformed into *E. coli* BL21(DE3) cells. Cells were induced with 1 mM isopropyl-1-thio-β-D-galactopyranoside at A_{600} of 0.8, and grown for 15 hrs at 25 °C. The cells were harvested by centrifugation and lysed by French-press in a buffer containing 20 mM tris-HCl, 500 mM NaCl, 5 mM β-mercaptoethanol (BME), pH 7.5 (buffer A), and a protease inhibitor (Roche Applied Science). The lysate was spun down and the supernatant loaded onto a nickel HiTrap column (GE healthcare). The 12.32 kDa *Mtb* PII protein was eluted

with a 100 mL linear gradient of 10 to 500 mM imidazole in buffer A. The peak fractions were collected and dialyzed in the presence of 20 mM tris-HCl (pH 7.5), 50mM NaCl, 5% glycerol, and 5 mM BME and concentrated to 15 mg/mL using an amicon filter (Amicon) as measured by the Bradford protein assay (Bio-Rad).

Crystallization

Initial crystallization conditions were obtained at 289 K in 96-well plates by sitting-drop vapor diffusion. Crystals of the *Mtb* apo PII were obtained by mixing equal volumes of 15 mg/mL protein with a crystallization solution from Crystal Screen^TM (Hampton Research) containing 0.1 M sodium cacodylate pH 6.5 and 1.4 M sodium acetate as precipitant. The *Mtb* PII:ATP binary complex was prepared by mixing the purified 15 mg/mL protein solution with ATP (Sigma) at a final concentration of 5 mM and incubating on ice for 30 min. Crystals of *Mtb* PII:ATP were obtained in condition containing 0.1 M imidazole pH 6.0 and 1.0 M sodium acetate, also from Crystal Screen^TM.

Data Collection, Structure Determination and Refinement

X-ray diffraction data for *Mtb* PII and *Mtb* PII:ATP crystals were collected on APS beamline 19-ID and 23-ID respectively. For data collection, a single crystal was cryoprotected by brief soaking in N-paratone and then flash-frozen in a liquid N_2 stream (100 K). The HKL-2000 suite [155] of programs was used for integration and scaling of PII

and PII-ATP complex structures. Crystals of the *Mtb* apo PII protein belong to the space group R3 with unit cell parameters $a = b = 77.15$ Å, $c = 51.84$ Å with one molecule per asymmetric unit and an estimated solvent content of 48.46%. Crystals of the *Mtb* PII:ATP belong to the space group P4$_3$2$_1$2 with unit cell parameters $a = b = 69.74$ Å, $c = 146.68$ Å with three molecules per asymmetric unit and an estimated solvent content of 48.08%. Data collection details are summarized in Table 4.

The structure of *Mtb* apo PII was solved by molecular replacement using Phaser [156]. The truncated *E. coli* PII structure (PDB ID: 1GNK) lacking the T-loop residues 38–54, was used as search model to solve the *Mtb* apo PII crystal structure [146]. The *Mtb* apo PII structure was used to solve the three positions of *Mtb* PII:ATP structure. After rigid body and restrained refinement using CCP4 – REFMAC5 [102,157] the model building was carried out in COOT [105] and XTALVIEW [104] for both the structures; solvent molecules were subsequently added. Bias-minimized electron density maps were obtained using the Shake & Warp (SNW) protocol [158]. Each subunit of the *Mtb* PII:ATP structure contained one molecule of ATP with clear electron density and were included in the Refmac refinement. The final R-factor for the *Mtb* apo PII crystal structure was 20.80% (R_{free} = 22.60%) at 1.4 Å resolution and for the *Mtb* PII:ATP structure it was 21.57% (R_{free} = 29.21%) at 2.4 Å resolution (**Table 4**).

Table 4: Crystallographic and refinement statistics for *Mtb* PII protein.

	Mtb **apo PII**	*Mtb* **PII:ATP**
PDB ID	3BZQ	3LF0
Data Collection		
Space Group	R3	$P4_3\,2_1\,2$
Wavelength (Å)	0.96	0.98
Temperature (K)	100	100
a (Å)	77.15	69.74
b (Å)	77.15	69.74
c (Å)	51.84	146.68
α (°)	90	90
β (°)	90	90
γ (°)	120	90
Resolution (Å)	1.40 (1.40 – 1.44)	2.40 (2.40 – 2.46)
Unique Reflections	21338 (1595)	14020 (982)
% completeness	99.10 (100.0)	99.78 (97.92)
R (merge)	0.06 (0.35)	0.07 (0.63)
<I/σI>	23.94 (7.77)	12.20 (2.02)
Refinement		
R value (%)	20.80	21.57
Free R value, random, 5%	22.60	29.21
Molecules per asymmetric unit	1	3
Protein residues	99	312
ATP molecules	0	3
Rmsd bond length (Å)[a]	0.006	0.017
Rmsd bond angle (Å)[a]	1.007	1.862
Rmsd between subunits (Å)	-	0.31, 0.39, 0.45
Residues ϕ - ψ angles		
Most favored (%)	98.80	96.60
Allowed (%)	1.20	3.40
Generously allowed / Disallowed (%)	0.0	0.0

[a] Deviations from restraint targets.

Results and Discussion

Crystal Structure of **Mtb** *apo PII Protein*

The crystal structure of the *Mtb* apo PII protein was solved at 1.4 Å resolution by molecular replacement method using the *E. coli* apo GlnK structure (PDB ID: 1GNK) as a search model. The data collection and refinement statistics are in Table 4. The crystallographic asymmetric unit contains one subunit, however a homotrimer with identical subunits is formed in the crystal lattice (**Figure 19 A**). Size exclusion chromatography confirmed that *Mtb* PII protein exists as a trimer in solution (data not shown).

The structural features of the PII family of proteins from different organisms are well conserved in the *Mtb* PII structure (**Figure 19 A and B**). The RMSD of a subunit of *Mtb* apo PII with *E. coli* GlnB (PDB ID: 2PII) is 0.78 Å (for 86 Cα atom pairs – excluding residues 36 to 55 and 109 to 112) [159] and the RMSD with *E. coli* GlnK (PDB ID: 1GNK - subunitA) is 0.70 Å (for 91 Cα atom pairs – excluding residues 26 and 36 to 55) [146]. Figure 19 B shows a subunit of the *Mtb* apo PII protein with the respective secondary features labeled. Each subunit of the *Mtb* apo PII structure consists of a double β-α–β (ferredoxin) fold that forms a β-sheet from four anti-parallel β-strands (β1 to β4); helix α1 connects strand β1 to β2 and helix α2 connects strand β3 to β4. Helices α1 and α2 lie on one side of the β-sheet. Connecting strands β2 and β3 is a large loop region that extends out away from the main body of the molecule, designated as the T-loop, made up of 20 residues (Tyr36 to Phe55). The flexible region of this T-loop,
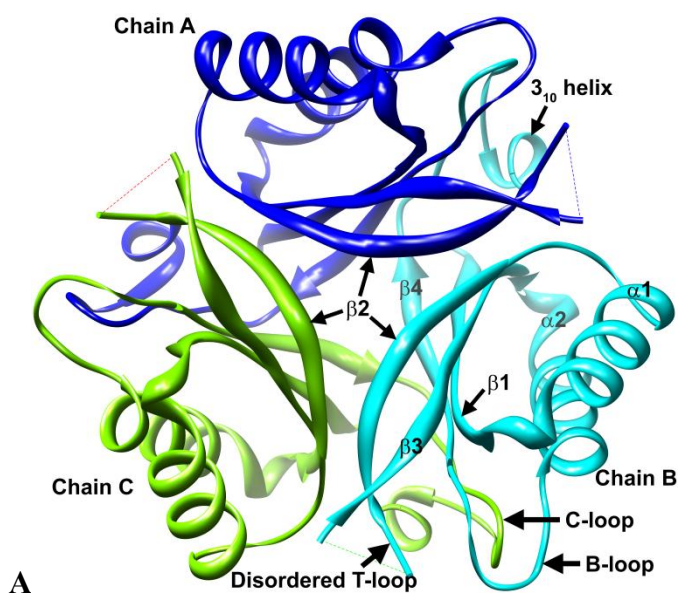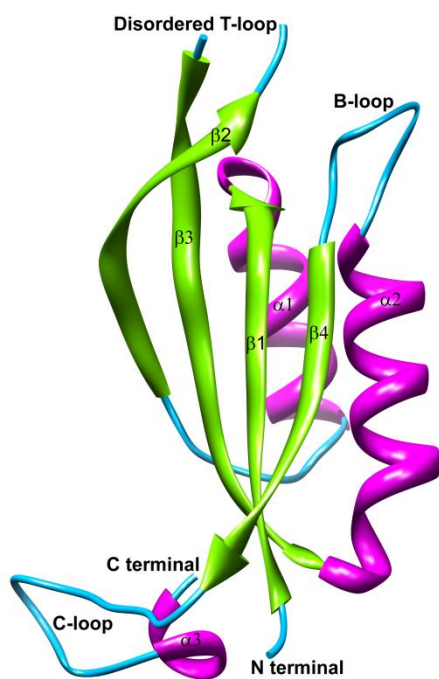
Figure 19: A) A ribbon diagram of the homotrimer of the apo *Mtb* PII protein looking down at the crystallographic triad. Molecular graphics images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081) [124]. **B)** A single subunit of the *Mtb* apo PII protein colored by the secondary subunits. The secondary features are labeled based on the *E. coli* GlnK structure [146].

encompassing residues Gln39 through Val53, are disordered in the electron density map and therefore omitted from the final model of the *Mtb* apo PII structure. The rest of the molecule, including the smaller loop connecting helix α2 and strand β4 (consisting of 8 residues, Arg82 to Gly89, referred to as the B-loop) which is oriented in the same direction as the T-loop, is fully ordered. The C-terminal residues extend out from β4 strand to form a fairly large C-loop (Asp97 to Leu112). The C-loop has a 3_{10} α-helix embedded within it between Gly108 and Ala111, consistent with the *E. coli* GlnK structure and in contrast to the *E. coli* GlnB structure [144].

Crystal Structure of the ATP bound Mtb *PII Protein*

The *Mtb* PII:ATP binary complex crystals were obtained by co-crystallization. X-ray data was collected to 2.4 Å resolution and the structure was solved by molecular replacement using the *Mtb* apo PII structure (PDB ID: 3BZQ) as the search model (Data collection and refinement statistics are presented in **Table 4**). Unlike the *Mtb* apo PII protein and the ATP bound *E. coli* GlnK, which have a single subunit in the asymmetric unit, the *Mtb* PII:ATP structure contained a trimer in the asymmetric unit with one ATP molecule bound per subunit (**Figure 20**). The molecular packing or binding of ATP to *Mtb* PII does not significantly change the overall structure of the protein. The trimeric *Mtb* PII:ATP structure is superposed onto the trimeric *Mtb* apo PII with an r.m.s.d. (between 94 x 3 Cα atom pairs) of 0.69 Å. The r.m.s.d. of *Mtb* PII:ATP with the *E. coli* GlnK:ATP trimer (PDB ID: 2GNK) is 0.80 Å (between 92 x 3 Cα atom pairs) [146], while the r.m.s.d. with *E. coli* apo GlnB (PDB ID: 2PII) was 0.74 Å (between the Cα atoms of

subunit B of *Mtb* PII:ATP and the monomer of *E. coli* GlnB). The r.m.s.d. between the three subunits of *Mtb* PII protein is 0.31 Å between subunits A and B, 0.39 Å between subunits B and C and 0.45 Å between subunits A and C. The differences lie in the positions of the residues in the ATP binding region, while the rest of the molecular backbone remains relatively unchanged compared to the *Mtb* apo PII. The disordered T-
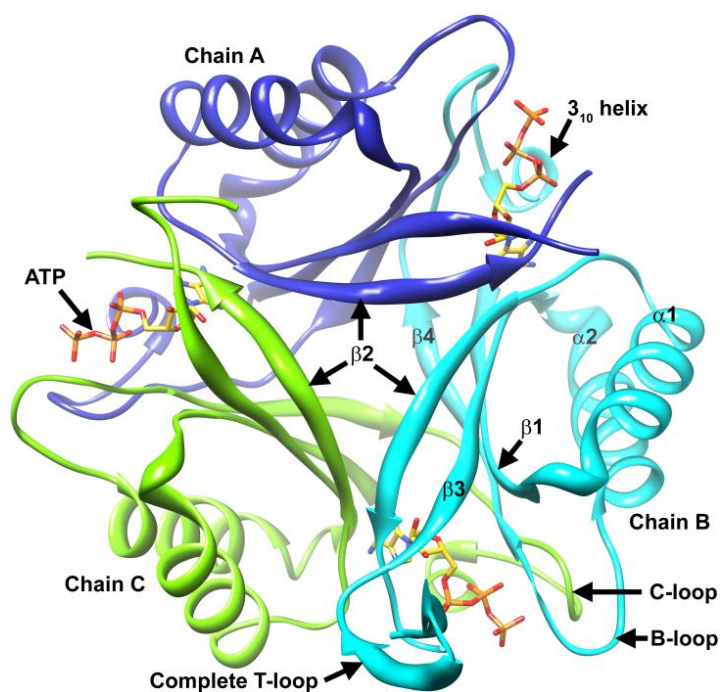
Figure 20: A ribbon diagram of the homotrimer of the ATP bound *Mtb* PII protein looking down the crystallographic triad. The three subunits of PII protein are represented in magenta, cyan and green ribbons. Bound ATP is shown in yellow stick representation.

loop of the apo PII structure is completely ordered in subunit B of the ATP bound *Mtb*

PII trimer with the exception of Lys40 and the partially ordered Gln39. However,

residues Gln39 to Val53 of subunit A and residues Gln39 to Glu50 of subunit C were

disordered and were omitted from the final model of *Mtb* PII:ATP structure.

Trimerization is Critical to the Function of **Mtb** *PII protein*

The trimer has some distinct structural features that indicate the possible function

of *Mtb* PII protein. In the trimer, the β-sheet from one subunit tilts toward the

neighboring subunit at a shear angle that is typical of β-barrel structures [160,161]. Three β-

sheets, one from each subunit, line the interior of the trimer and enclose a concave β-

barrel at the center of the trimer, while the α-helices of each subunit lie on the outside of

the trimer (**Figure 20**). The β2 strands from the individual subunits form the mouth at

the top of the β-barrel with Ser31 of each β2 strand making hydrogen bonding

interactions with Gln34 of the neighboring β2 strand. Additionally, Glu32 from each β2

strand comes closest to the center at the top of the β-barrel, lying symmetrically away

from each other at a distance of ~ 9.1 Å and creating a negatively charged region at the

mouth of the β-barrel (see **Figure 21**).
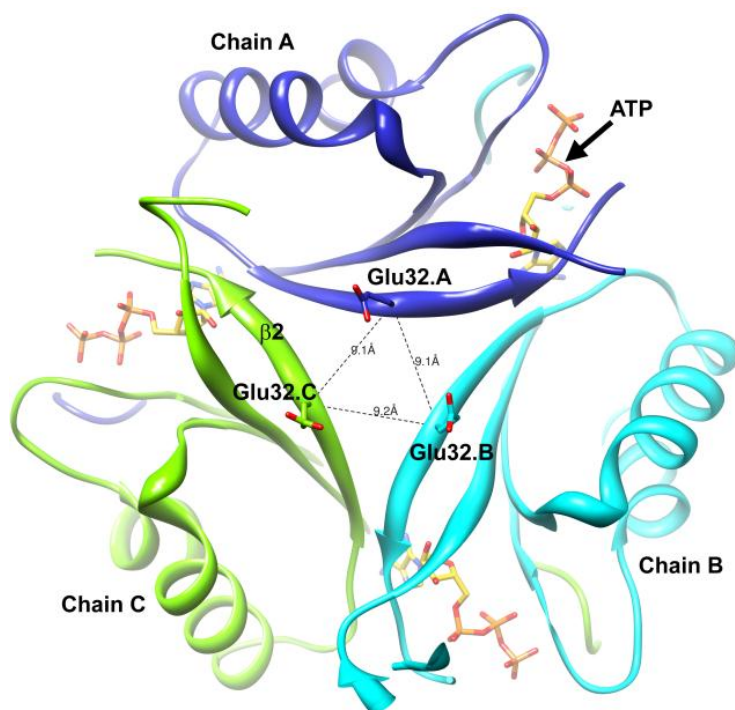
Figure 21: Glu32 from β2 strands of each subunit of the homotrimer form the mouth of the β-barrel lying at 9.1 Å distance away from each other.

The inner lining of this β-barrel encloses a concave cavity with hydrophobic residues Leu3 from the β1 strand and Trp92 from the β4 strand. In addition to Glu32 at the mouth of the cavity, Arg60 and Glu62 from the β3 strands of each subunit insert the polar ends of their side-chains into the center of the cavity forming hydrogen bonds to create a charged center, stabilizing the trimer assembly (see **Figure 22**).
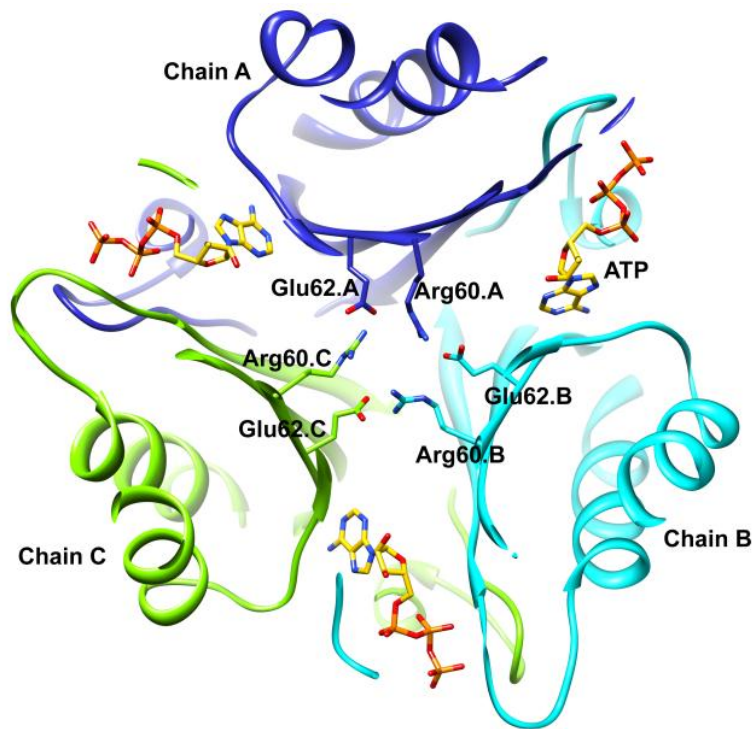
Figure 22: Arg60 and Glu62 inside the β-barrel of *Mtb* PII protein form a charged center.

Finally, the β4 strand of each subunit bends inward at the bottom to close the base of the concave β-barrel. Pro95 residues from each β4 strand are brought into close proximity at the bottom of this barrel at a distance of ~4.0 Å (see **Figure 23**).
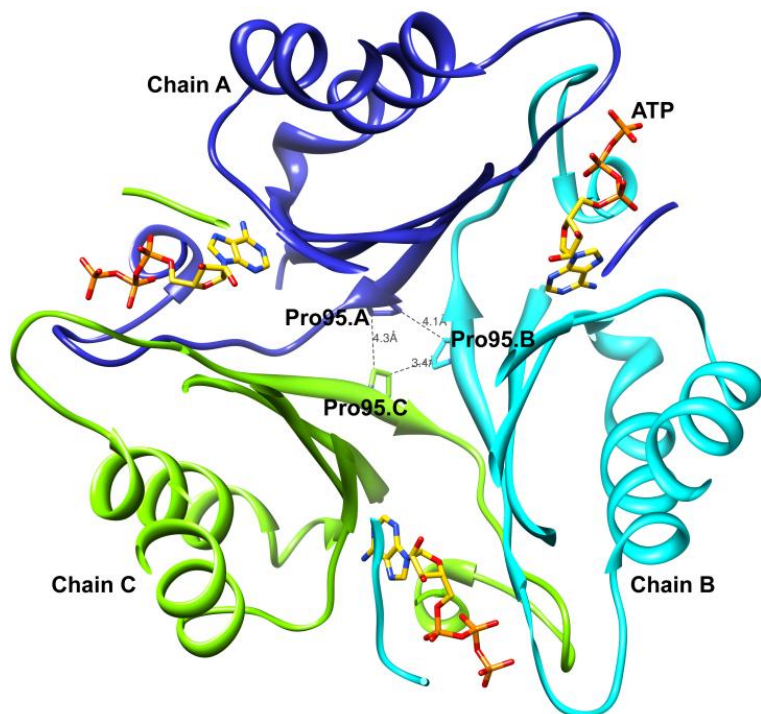
Figure 23: Pro95 from the β4 strands of each subunit form the base of the β-barrel.

The hydroxyl group of Ser94 in the β4 strand at the base of the β-barrel forms a hydrogen bond with the backbone carbonyl oxygen atom of Pro95 (of the same β4 strand) through a water molecule. Furthermore, the presence of a network of ordered water molecules observed at the bottom of the β-barrel might also serve to stabilize the trimer. While the rest of the trimer interface is conserved between most other organisms, including *E. coli* GlnB and GlnK, the Pro95 residue of *Mtb* PII that occurs at the bottom of the β-barrel is not conserved across species. It is an Asp95 in *E. coli* GlnB, *Klebsiell pneumoniae* GlnB, *Azospirillum brasilense* GlnK, *Herbaspirillium seropedicae* GlnK, and *rhodospirillum rubrum* GlnK, and a Glu95 in *E. coli* GlnK, *Klebsiell pneumoniae*

GlnK, and *Herbaspirillium seropedicae* GlnB. The amino group on Lys2 of *Mtb*

PII:ATP is positioned closely to Pro95 at a distance of 4.7 Å within the subunit. The

Lys2 forms a hydrogen bonding interaction with Asp97 of a neighboring subunit in the

trimer of *Mtb* PII:ATP at 2.7 Å distance. In the case of *E. coli* GlnK, the Lys2 is

involved in hydrogen bonding interactions with Glu95 (2.6 Å distance) of the same

chain and hence is not available for interaction with the residue at position 97 of the

neighboring subunit, which is Gln97 (6.7 Å). The functional significance of the

hydrophobic and hydrogen bonding interactions involving Pro95, Asp97 and Lys2 at the

bottom of the β-barrel in *Mtb* PII compared to *E. coli* GlnK/GlnB warrants further

investigation in relation to trimerization.

An important feature that becomes apparent in the trimeric state is that part of the

long C-loop of one subunit lies parallel to the β4 strand of the neighboring subunit

making several interactions. For instance, there are hydrogen bonding backbone

interactions between Thr98 (C-loop of subunit C) and Val93 (β4 strand of subunit B),

Val100 (C-loop of subunit C) with Val91 (β4 strand of subunit B) and Trp92 (β4 strand),

Val102 (C-loop) with Lys90 (β4 strand). The backbone amide of Arg103 (C-loop) also

hydrogen bonds with the backbone carbonyl of Gly89 (β4 strand) via a water molecule.

The extensive intersubunit hydrogen bonds between the C-loop and the β4 strand could

be the defining factor in the trimerization of *Mtb* PII. The C-loop of one subunit

approaches the B-loop and the ATP binding site of the neighboring subunit which may

also preserve the structural integrity of the functional trimer. For example, the backbone

carbonyl group of Thr104 (C-loop of subunit C) hydrogen bonds with the guanido group

of Arg82 (B-loop of subunit B) at 2.8 Å distance. The C-loop also encloses a 3_{10} helix turn at its C-terminal end, which contributes to the ATP binding site of the neighboring subunit and is partly stabilized by the hydrogen bonding of the carbonyl end of Ala111 (C-loop) with the amino side-chain of Lys90 ($\beta4$ strand) (2.8 Å distance). The importance of the trimeric architecture is that the secondary structural elements involved in the trimerization also form part of the ATP binding pocket. In particular, the $\beta2$ and $\beta4$ strands of *Mtb* PII contribute key residues to the ATP binding pocket (*vide infra*) indicating that the trimerization is critical for its function.

Comparison between the Apo and ATP Bound **Mtb** *PII Protein*

The crystal structure of *Mtb* PII:ATP shows that ATP binds in three symmetrically equivalent positions in the trimer in a pocket formed by the base of the T-loop, B-loop, the $\beta4$ strand of one subunit, and the $\beta2$ strand, $\beta3$ strand, and part of the C-loop of a neighboring subunit. In this structure the ATP exists in a C-3'-endo conformation with the adenine ring in the anti-orientation consistent with its position in the *E. coli* GlnK:ATP structure [146] (**Figure 24**).
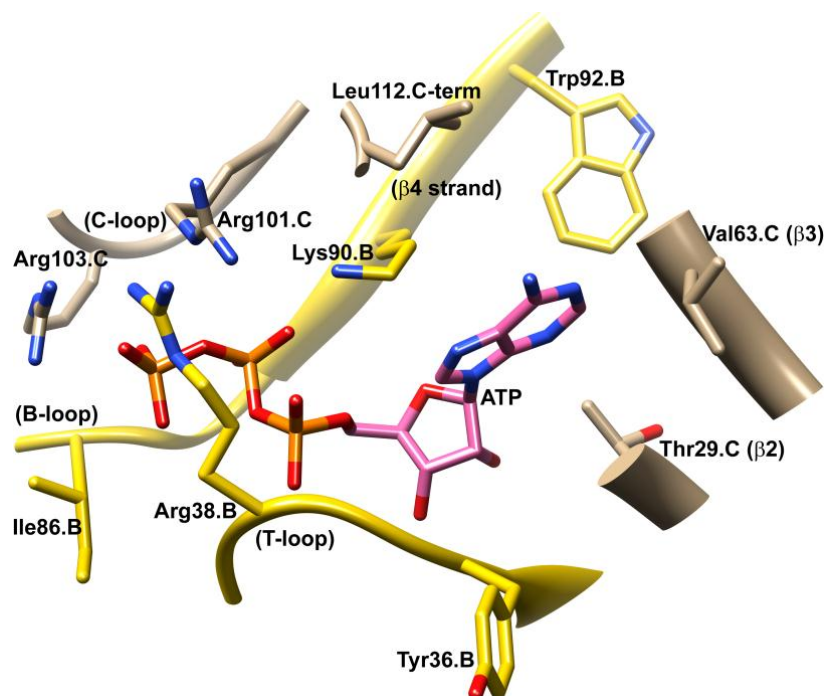
Figure 24: ATP bound in the active site of *Mtb* PII protein. Residues contributed by the neighboring subunit are shown in tan color while the residues contributed by the ATP binding subunit are shown in yellow. ATP is shown in pink stick representation.

Conserved residues Gly87 (B-loop), Gly89 (B-loop), Lys90 (β4 strand), and Arg103 (C-loop) are also found in the active site of *Mtb* PII, all of which hydrogen bond with the polar phosphate groups of ATP. The adenine ring of ATP sits in a hydrophobic pocket surrounded by residues contributed by two neighboring subunits. Residues contributing to adenine binding include Ile7 of the β1 strand and Trp92 of the β4 strand. The residues of the neighboring subunit contributing to adenine binding include Leu26, Gly27, Met28 and Thr29 of the loop connecting α1 helix and β2 strand, Val64 of the β3 strand, and Leu112 of the C-loop (colored tan in **Figure 24**). Since the ATP molecule

98

itself has non-bonded interactions with residues from the neighboring subunit it helps to further stabilize the trimer as compared to the apo trimer. While all the hydrophobic residues move closer into the binding pocket upon ATP binding, two residues move away; Gly27 moves by 2.5 Å and lies at a distance of 3.4 Å from the adenine ring, and the amide linkage between Gly27 and Leu26 rotates by ~180° such that the carbonyl carbon of Leu26 is 4.4 Å away from its original position in the *Mtb* apo PII (Subunit B of *Mtb* PII with a complete T-loop was used to calculate all inter-atomic distances in the ATP binding pocket). The movement of Leu26 and Gly27 away from their positions in the *Mtb* apo PII structure makes space for the adenine ring of ATP.

The protomer ATP interactions shown above can be viewed in reference to three important structural elements, namely the C loop (3_{10} helix), the B loop, and the T loop (**Figures 24**). The C-terminal end of the neighboring subunit of *Mtb* PII (at the end of the C-loop), aided by a 3_{10} helix, curls back to form part of the ATP binding pocket. The C-terminal residue Leu112 lines a hydrophobic pocket for the adenine ring of ATP, while the terminal carboxylic oxygen atom of Leu112 hydrogen bonds with Arg101 at a distance of 2.8 Å, which in turn hydrogen bonds with the phosphate of ATP (2.9 Å distance). The 3_{10} helix must play a role in the binding of the ATP because it brings Leu112 into proximity with ATP. Furthermore, in the *Mtb* apo PII structure, the Cα atom of Leu112 moves away from the ATP binding site by 1.6 Å and loses its hydrogen bonding interaction with Arg101 although the 3_{10} helix is conserved in the apo structure. Leu112 of the *Mtb* apo PII protein also makes hydrophobic contact with the N-terminal

Met1 (3.6 Å distance) of the same subunit. Upon binding ATP the Leu112 moves closer to ATP but still maintains the hydrophobic contact with the N-terminal Met1.

The B-loop carries the conserved mononucleotide binding motif 'Gly(84)-X-X-Gly(87)-X-Gly(89)-Lys(90)' and thus acts as the primary binding site for ATP [162]. Compared to the *Mtb* apo PII structure, the B-loop in *Mtb* PII:ATP moves away from the ATP binding region, making space for the ATP to bind. The Cα atom of Lys85 at the tip of the B-loop moves away by 1.5 Å. The backbone nitrogen atom of Gly89 makes hydrogen bonding interaction with the oxygen atom of the β-phosphate at 2.9 Å distance. While the Cα atom of ATP-bound and unbound Lys90 are in identical positions, the side-chain moves in the ATP bound structure so that the N-Z atom of Lys90 is 5.3 Å away from its position in the apo structure and hydrogen bonds with the β-phosphate of ATP at 2.5 Å distance.

Arg38 extends out from the base of the T-loop and hydrogen bonds with the oxygen atom of the γ-phosphate at 3.0 Å distance (**Figure 24**). The binding of ATP is involved in stabilizing the side-chain conformation of Arg38, and therefore should partly stabilize the T-loop conformation. In the *Mtb* apo PII structure, the side-chain of Arg38 is disordered (along with the T-loop) and is represented as alanine, which could result from a lack of stabilizing interactions in the absence of ATP. Indeed, the Cα atom of Arg38 in the *Mtb* apo PII moves away from the active site by 8.3 Å compared to the ATP bound structure, to open this cavity to solvent. This effectively changes the flexibility of the T-loop conformation in *Mtb* apo PII. Comparing the apo and ATP-bound structures of *Mtb* PII protein, it seems apparent that Arg38's interaction with the

γ-phosphate is the reason the base of the T-loop moves back into the active site (**Figure 25**). While Arg38 forms one part of the enclosure for the phosphates of ATP, Arg103 from the C-loop of the neighboring molecule forms another part of this enclosure forming a hydrogen bond with oxygen on the γ-phosphate (2.8 Å distance). Similarly, the C-loop also contributes Arg101, which further stabilizes the ATP binding site. Arg101 of subunits A, B and C hydrogen bond with the bridge oxygen between the β- and γ-phosphates of ATP (2.8, 2.9 and 3.2 Å distances respectively).
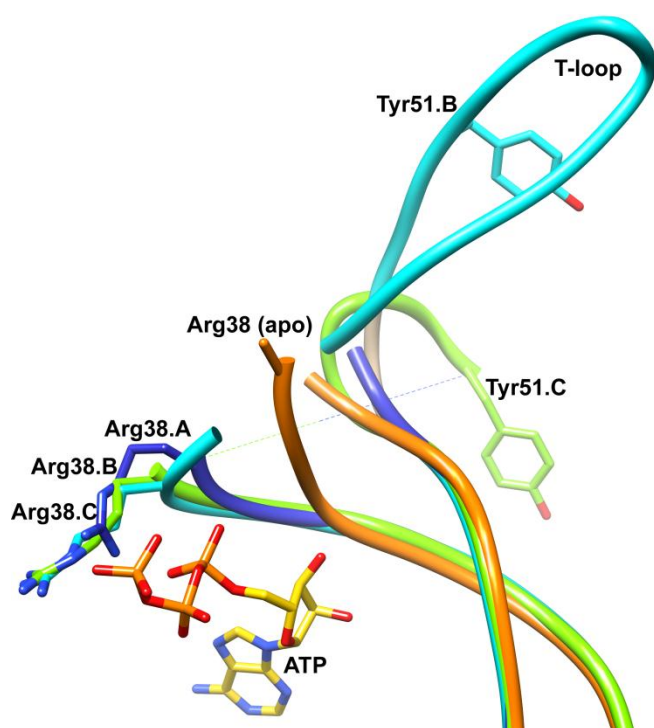
Figure 25: Superposition of the base of the T-loop of the three subunits of *Mtb* PII protein in the homotrimer and the apo *Mtb* PII protein.

While Arg38 moves in closer to the active site upon ATP binding, Tyr36 moves away (the C-α atom of Tyr36 moves by 2.9 Å distance) compared to its original position in the *Mtb* apo PII structure. In the ATP bound structure, the back bone nitrogen atom of the Tyr36 forms a hydrogen bond with the 3' hydroxyl group of ATP (3.2 Å distance). Therefore, in the ATP bound PII structure, the Tyr36 orients its side chain more closely to the side chain of Phe55 (at 3.5 Å distance). Since Tyr36 is the first residue and Phe55 is the last residue of the functionally important T-loop, ATP binding can be said to partly influence the orientation of the T-loop itself. Thus Tyr36 along with Gly37, Arg38 and Phe55 plays a role in the stability of the base of the T-loop. It is clear from the positions of residues leading up to the missing section of the T-loop in subunit C that the loop has a very different conformation compared to subunits A and B (**Figure 25**). While Tyr36 and Phe55 at the neck of the T-loop are in equivalent positions in subunit B and C, the Cα atoms of the functionally crucial Tyr51 are positioned 12.4 Å away compared to each other. Crystal packing might play a role in the solvent exposed part of the alternate T-loop conformation observed in subunit C of *Mtb* PII:ATP structure (as discussed below). Taken together, the binding of ATP along with the positions of Tyr36, Arg38, and Phe55 might contribute to the stability of at least the base of the T-loop of *Mtb* PII and point it in the right orientation to form a complex with the target proteins of PII, like AmtB or GlnD.

Although the residues at the base of the T loop including Arg38 is fully ordered in all subunits of *Mtb* PII: ATP (Arg38 is disordered in the apo structure), only subunit B contained the fully ordered T-loop. Upon analyzing the crystal packing effect from the

residues of symmetry related molecules on the stability and conformation of the visible

T-loop of subunit B in *Mtb* PII:ATP structure, we observed that part of the T-loop of

subunit B packs inside a groove created by two symmetry related molecules (*Mtb* PII'

and PII'') (**Figure 24**). Tyr51 of the T-loop makes hydrophobic interactions with Val53'

(3.8 Å distance) of the T-loop from the symmetry related molecule. Tyr46 makes

hydrophobic interactions with Ser52' (3.7 Å distance). Glu50 makes a hydrogen bond

interaction with Thr98'' (2.5 Å distance). The backbone amide group of Arg47 hydrogen

bonds with the carboxylic acid group of Asp71'' (3.0 Å distance), and the Cα atom of

Arg47 makes a van der Waal's interaction with the side-chain of Val70'' (3.8 Å).

Overlaying subunit B on subunit A reveals that the T-loop of subunit A is exposed to

solvent without any stabilizing interactions; and overlaying subunit B on subunit C of

Mtb PII revealed clashes of the T-loop with the T-loop from a neighboring symmetry

molecule, requiring drastic changes in the T-loop conformation in subunit C to avoid

these clashes. Because of this the T-loops in subunits A and C of ATP-bound *Mtb* PII

were only partially ordered. Collectively, the interactions of the T-loop of subunit B with

symmetry molecules help stabilize the flexible part of the T-loop, which would

otherwise need to be stabilized by forming a complex with a target protein or an effector

molecule such as 2OG [59]. The crystal packing effect upon stabilization of subunit B's T-

loop of *Mtb* PII protein mimics the formation of a complex with a target protein which

indicates that the T-loop requires additional stabilizing interactions mediated through the

binding of a target protein to *Mtb* PII (**Figure 27 A**). However, the present conformation

of the T-loop is subjective and may adopt different conformations in the presence of its

physiological binding partners like AmtB (the GlnK:AmtB homology model is presented in a later part).

Role of Mg2$^+$ in the ATP Binding Site

Analysis of available crystal structures of GlnB and GlnK in the RCSB Protein Data Bank revealed that all GlnK structures solved to date enclose a 3_{10} helix at their C-termini in contrast to GlnB structures (**Figure 19 B**). This includes the *E. coli* GlnK (PDB ID: 1GNK), *Thermus thermophilus* GlnK (PDB ID: 1UFL), *M. jannaschii (M. jann)* GlnK1 (PDB ID: 2J9C). Since both apo and the ATP bound *Mtb* PII proteins enclose a 3_{10} helix at their C-termini it might be implied that *Mtb* PII functions as a GlnK protein. It is unclear what causes the formation of the 3_{10} helix at the C-terminal of the GlnK protein since there is no conservation in the four residues – Arg107, Gly108, His109, Asp110 – that are involved in the formation of the 3_{10} helix among different organisms like *Mtb* PII, *E. coli* GlnK and *M.jannaschii* GlnK1. It is possible that the micro heterogeneity observed in the sequences of GlnB and GlnK might have resulted in the difference in propensity for 3_{10} helix formation. The 3_{10} helix brings the terminal residue, Leu112 in proximity with the adenine moiety of ATP. While the side-chain of Leu112 helps to maintain a hydrophobic pocket for the adenine ring of ATP, the carbonyl end of Leu112 hydrogen bonds with Arg101, which in turn hydrogen bonds with the γ-phosphate of ATP, similar to the interactions observed in *E. coli* GlnK:ATP structure. Thus the significance of the 3_{10} helix is that it contributes Leu112 towards forming the ATP binding site of *Mtb* PII.
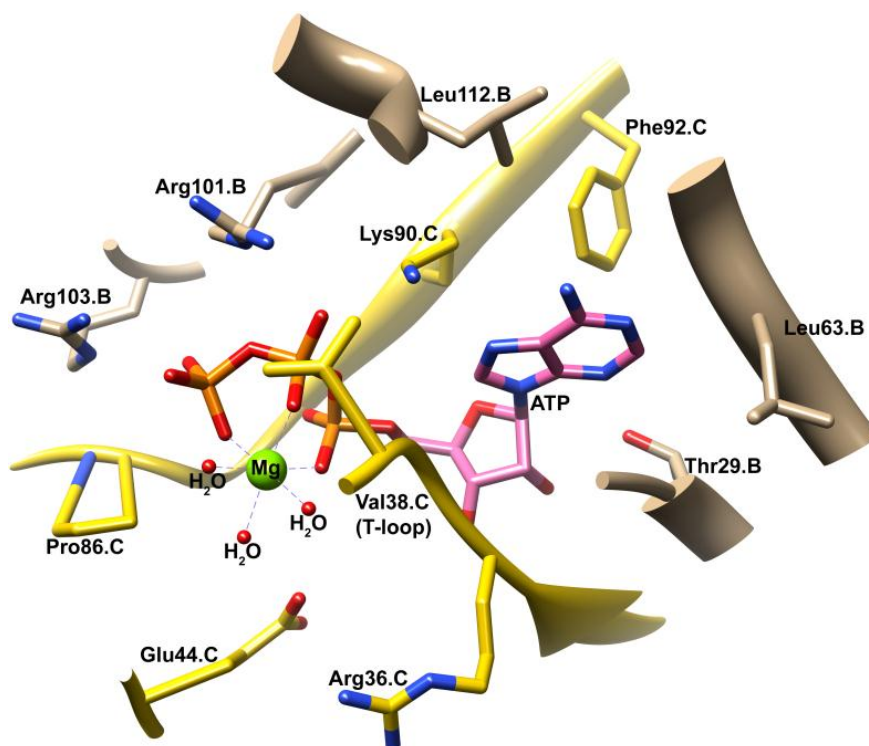
Figure 26: Active site of the *M. jannaschii* PII protein with the bound ATP shown in pink stick representation and Mg^{2+} ion shown as a green sphere. The coordinated water molecules are shown as red dots.

The *M. jannaschii* (*M. jann*) GlnK1 crystal structure has also been solved in complex with ATP, and 2OG molecules [147]. Out of three possible binding sites, only one 2OG molecule was bound per trimer and it bound next to Tyr51 of the T-loop which adopts a significantly different conformation with respect to the unbound T-loops. Binding of 2OG did not, however, influence the conformation of the rest of the molecule, including the C-terminal, which still accommodates a 3_{10} helix. Like the *Mtb* PII and the *E. coli* GlnK structures, *M. jann* GlnK1 also forms a homotrimer and has an

r.m.s.d. of 0.74 Å with the *Mtb* apo PII (between 95 Cα atom pairs from chain A of *Mtb* PII and *M. jann* GlnK1) and an r.m.s.d. of 0.87 Å with *E. coli* GlnK (between 95 Cα atom pairs from chain A of each). In addition, the crystal structure of the *M. jann* GlnK1 shows a Mg^{2+} ion bound next to ATP which coordinates with all three phosphate oxygen atoms [147] (See **Figure 24 and 26** for a comparison of the *Mtb* PII and *M. jann* GlnK1 ATP binding sites). Yildiz et al. have argued that Mg^{2+} and ATP are necessary to change the T-loop conformation which then creates the 2OG binding site. In the case of *Mtb* PII:ATP protein, Arg38 of the T-loop occupies the Mg^{2+} binding site (**Figure 26**) to form a direct hydrogen bond with the γ-phosphate oxygen atom of the ATP (3.0 Å distance). Arg38 of *Mtb* PII is a Val38 in *M. jann* GlnK1 suggesting that a Mg^{2+} is critical for *M. jann* similar to many other ATP binding proteins. Binding of Mg^{2+} ion pushes the base of T-loop away from the ATP phosphates in *M. jann* GlnK1, while in *Mtb* PII:ATP structure it is observed that the base of the T-loop comes close to ATP (the main-chain amide group of Arg38 hydrogen bonds with the ATP α-phosphate at 3.1 Å distance). The same amide group of Val38 in *M. jann* GlnK1 is 4.7 Å away from the ATP α-phosphate [147]. Therefore, it may be the case that a Mg^{2+} ion is not required by *Mtb* PII to change the T-loop conformation and create a 2OG binding site. Indeed, our attempts to crystallize *Mtb* GlnK in the presence of Mg^{2+} ions were unsuccessful.

Comparison with the **E.coli** *GlnK: AmtB Complex*

Mtb encodes a probable *amtB* gene (Rv2920c), which is next to the *glnB* gene [30]. Comparison between the 428 amino acid *E. coli* AmtB protein and the 477 amino acid

Mtb AmtB protein shows a 41% sequence identity. All essential residues involved in the interaction of AmtB with the T-loop of PII, specifically Phe107, His168, Phe215, Leu259, Ser263, Val299, Asp313, and His318 of the *E. coli* AmtB, were found to be conserved in *Mtb* AmtB [163]. We constructed a homology model of *Mtb* AmtB and the T-loop section of the *Mtb* PII protein based on the crystal structure of the *E. coli* GlnK:AmtB complex (PDB ID: 2NS1) (**Figure 27 B**) using the SWISS-MODEL program [164]. Comparing the T-loops of subunit B of the *Mtb* PII protein with the homology modeled *Mtb* PII, we noticed that the region of the T-loop that sticks out from the main body of the PII protein swings away from the T-loop binding pocket of AmtB by ~ 45°. The position of the T-loop observed in the crystal structure of *Mtb* PII:ATP could be due to crystal packing effects (*vide ante*). Thus, in order to bind *Mtb* AmtB, the T-loop would have to swing by ~ 45° with a maximum movement of ~ 11.0 Å (observed for the Cα atom of Arg47) to bring the T-loop inside the groove of AmtB (**Figure 27 B**). However, the base of the T-loop that forms and interacts with the ATP binding site does not move. This suggests that ATP binding stabilizes the base of the T-loop (residues 36-39), but that the rest of the T-loop must be stabilized by a target protein. Given that PII proteins can bind a multitude of target proteins (particularly, AmtB, and GlnD of the same operon [165]), it is logical to think that the T-loop exists in different conformations depending on the target protein it complexes with. For example, in the *Arabidopsis thaliana* PII:NAGK complex structure, the T-loop adopts a very different conformation compared to the *E. coli* PII:AmtB complex structure [166]. The crystal structure of PII:GlnD has not been solved for any organism to date despite strong biochemical

evidence for this protein-protein complex formation [167]. Obviously other parameters (like 2OG binding) would influence the conformational changes of the T-loop.
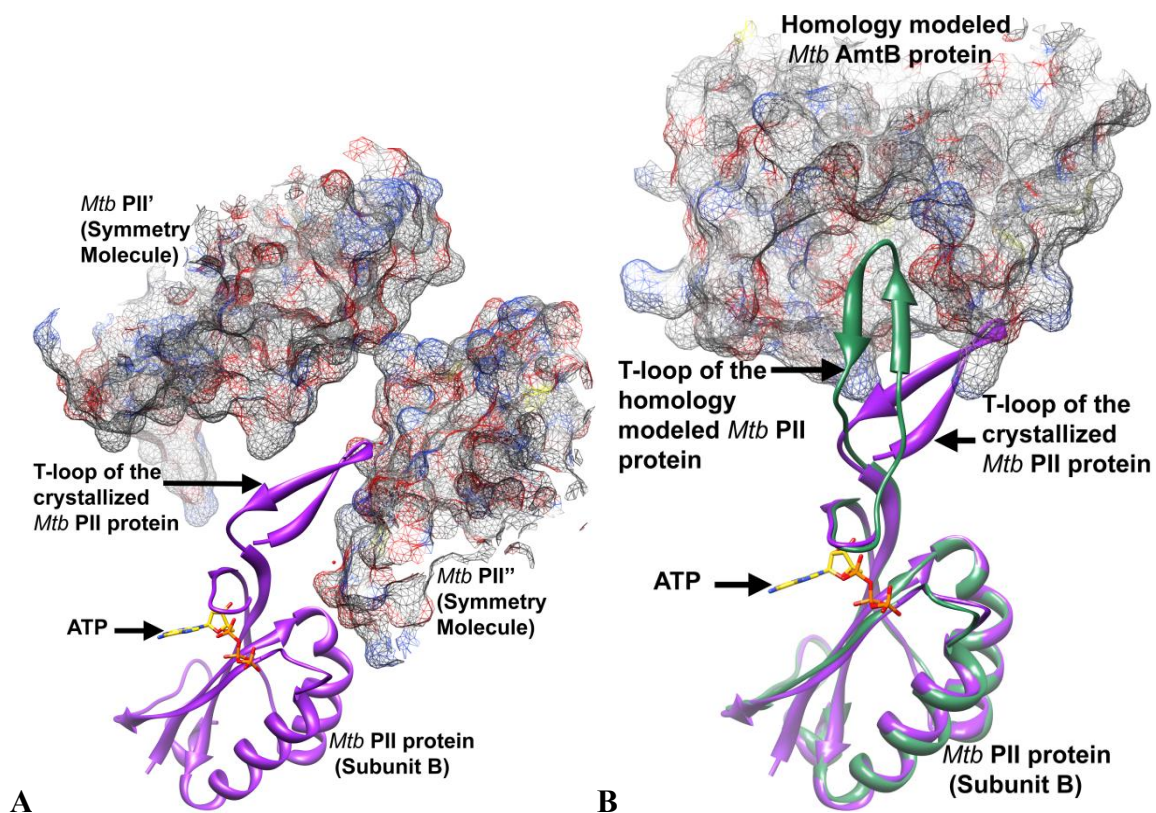
Figure 27: Homology model of the *Mtb* PII protein in complex with *Mtb* AmtB protein. **A)** The figure shows the observed position of the T loop in the crystal structure of *Mtb* PII protein with the neighboring symmetry molecules shown in surface representation. **B)** Figure shows the movement of the T loop of the *Mtb* PII protein in a homology model of the *Mtb* PII protein in complex with *Mtb* AmtB protein. Part of the homology modeled *Mtb* AmtB protein is represented in surface view.

Looking into the T-loop binding pocket of the *Mtb* AmtB homology model, there are few differences when compared with the *E. coli* AmtB structure. The terminal hydroxyl group of Tyr51 of *Mtb* PII protein makes favorable hydrogen bonding interactions with the backbone amide of Phe224 of *Mtb* AmtB (2.6 Å distance). The phenyl ring of Tyr51 makes additional hydrophobic contacts with the side-chain of Leu223 (3.3 Å distance), which does not exist in *E. coli* AmtB, instead it is an Ala192. However some hydrogen bonding interactions observed in *E. coli* are missing in *Mtb*, such as the interaction of the backbone carbonyl of Arg47 with Arg253 of *E. coli* AmtB; in *Mtb* it is an Asp284 and is 5.9 Å away. There is also a 36 residue insertion at the C-terminal of the *Mtb* AmtB that hangs on the cytoplasmic side of the membrane that was not included in the homology model, but will be involved in some form of interaction with the PII trimer. Thus *Mtb* PII protein would be expected to interact and form a complex with the *Mtb* AmtB.

Annotation of the **Mtb** *PII Protein*

The gene encoding *Mtb* PII has been annotated *glnB* in the Tuberculosis Structural Genomics Consortium (TBSGC). Comparing the sequences of other PII proteins from across different species, the *Mtb* PII protein shows a 2 to 7% greater sequence identity with known GlnBs than with their respective GlnKs. (See **Table 5** for comparison of percentage sequence identities with PII proteins from other organisms.) Thus, on the basis of sequence identity alone it could be termed *Mtb* GlnB. However, while the *glnB* gene in most organisms is found in an operon with the *glnA* gene

encoding glutamine synthetase, the *glnK* gene in most organisms is linked to an *amtB* gene forming an *amtB-glnK* or *glnK-amtB* operon (and may also contain a third gene, *glnD*). *C. glutamicum* and *S. coelicolor* are two other actinobacteria where the nitrogen regulation pathway has been studied in detail [168,169]. Like *Mtb* they too encode a single homolog of the *PII* gene that lies between the *amtB* and *glnD* genes and the gene product has been annotated as GlnK in both microorganisms based on operon organization. As *Mtb* follows the same operon organization as *C. glutamicum* and *S. coelicolor* it is logical that the gene would also translate into a GlnK protein. Furthermore, it has been suggested that to be annonated *glnB*, the protein product of these genes should consist of a conserved Lys3 and an Asp5 or Glu5 [50]. While the *E. coli* GlnB observes this rule, the *Mtb* PII protein has a Leu3 and Thr5, both consistent with *E. coli* GlnK. The Lys3 and Asp5/Glu5 residues among GlnB proteins and the Leu3 and Thr5 residues among GlnK proteins are fairly conserved across species, suggesting that the *Mtb PII* gene is in fact *glnK* [50].

Table 5: Percentage identity of *Mtb* PII protein with GlnB and GlnK of other organisms.

	GlnB (% identity with *Mtb* PII protein)	GlnK (% identity with *Mtb* PII protein)
α Proteobacteria		
Azospirillum brasilense	60	56
Rhodospirillum rubrum	57	60
Azorhizobium caulinodans	59	57
β Proteobacteria		
Herbaspirillum seropedicae	64	61
γ Proteobacteria		
Escherichia coli	61	54
Klebsiella pneumoniae	61	54
Firmibacteria		
Bacillus subtilis	--	7
Actinobacteria		
Corynebacterium glutamicum	--	68
Streptomyces coelicolor	--	68
Archaebacteria		
Methanococcus jannaschii	--	55
Cyanobacteria		
Synechococcus	61	--
Plants		
Arabidopsis thaliana	46	--

-- a GlnK or GlnB gene does not exist or has not yet been identified.

The *Mtb* PII protein projects the Leu3 and Thr5 residues from each subunit into the central cavity of the β-barrel formed by the trimer, while *E. coli* GlnB has Lys3 and Asp5 residues in these positions. The significance of these hydrophilic versus hydrophobic residues at the center of the β-barrel in relation to the function of GlnB and GlnK is unknown and calls for further investigation. Apart from operon organization and sequence specific characterization, the crystal structure of the *Mtb* PII protein gives a clear indication of the nomenclature of the protein. The presence of a 3_{10} helix at the C-terminus of the *Mtb* PII protein is consistent with other known GlnK structures, while in GlnB this region is a loop that is positioned away from the ATP binding pocket. It has been proposed earlier that GlnK structures may differ from GlnB due to the presence of a 3_{10} helix in the C-loop [145,146,150]. For these reasons it was suggested that the gene for the PII protein in *Mtb* is *glnK*.

Conclusion

The operon organization of *Mtb PII* gene, the existence of GlnK specific structural features in both apo and ATP bound *Mtb* PII crystal structures, and the homology model of *Mtb* PII:AmtB complex, suggest that the PII homolog of *Mtb* can play a key role as GlnK in the nitrogen regulatory pathway. In fact the up-regulation of the *PII*, *glnD*, and *amtB* genes in response to low nitrogen conditions in *Mtb* [61] implies a possibility that the PII and GlnD proteins may be active in nitrogen regulation through AmtB and possibly in the transcriptional regulation of the genes involved in nitrogen regulation. GlnD is another signal transducing protein that is active as a uridyl

transferase in the enterobacterial nitrogen regulation system [167]. Alternatively, the same enzyme has been shown to act as an adenyl transferase on GlnK proteins in the actinobacteria – *C. glutamicum* and *S. coelicolor* [165,170]. However, there is no evidence for the interaction between an adenylylated GlnK with GlnE in *C. glutamicum* and *S. coelicolor* [170,171]; this is in contrast to *E. coli*, where the GlnE and GlnB proteins work together to regulate GS (ammonium dependent glutamine synthatase). While GlnE was shown to be essential for *Mtb*, affecting the adenylylation and deadenylylation activity of *Mtb* glutamine synthetase, this pathway was found to be independent of the PII/GlnD/AmtB proteins [61], suggesting that the GlnE dependant regulation of GS does not require a PII homolog.

PII proteins have been implicated in the transcriptional regulation of genes involved in the nitrogen metabolism of actinobacteria [172]. In *S. coelicolor*, an OmpR type transcriptional regulator, called GlnR, has been suggested to carry out the transcriptional regulation of genes involved in nitrogen control, including *amtB*, *glnK*, and *glnD* among others. A *glnR* gene was also identified and characterized in *M. smegmatis* (*M. smeg*) and shown to regulate the transcriptional activation of genes involved in nitrogen control in the organism [173], although a reduced number of the target gene sequences were identified in *M. smeg* compared to *S. coelicolor*. *Mtb* encodes a putative GlnR, which shares a 73% amino acid sequence identity (83% homology) with the *M. smeg* GlnR protein and 64% identity (78% homology) with the *S. coelicolor* protein. *In vitro* experiments with the *S. coelicolor* GlnR showed that it was able to bind the *Mtb glnA* promoter region in electrophoretic mobility shift assays [142] suggesting that the *Mtb* GlnR

might be a good candidate to regulate the transcription of genes involved in nitrogen control. It has also been suggested in previous studies that the *C. glutamicum* GlnK can complex with a TetR type transcriptional repressor, AmtR [141], which is considered a master regulator of the nitrogen metabolism genes, involved in the repression of transcription of at least 33 genes during nitrogen surplus. *M. smegmatis* was the only mycobacterium to have a homolog of the AmtR (42% identity), while the rest of the mycobacteria had no homologs of AmtR [174]. Though *M. smeg* AmtR was considered to be an interesting candidate for a second nitrogen regulator [173], the exact role of the protein in the organism is unknown.

The model for nitrogen regulation in *Mtb* can be proposed from previous studies in *C. glutamicum, S. coelicolor* and *M. smegmatis* [60,173,174] based on the identification of the homologous proteins in the bacterial genome sequence database (listed above), and from our structural data for *Mtb* PII protein (**Figure 28**). Under nitrogen limiting conditions PII protein is adenylylated by GlnD, preventing complex formation with the ammonium permease AmtB, permitting an unhindered flow of ammonium into the cytoplasm (**Figure 28 A**). During this event, adenylylated PII may bind GlnR, preventing it from functioning as a transcriptional activator of the *amt* operon and various other nitrogen metabolism genes (**Figure 28 A**) [172]. At high intracellular nitrogen concentrations, GlnD is triggered to deadenylylate PII, freeing it to complex with the membrane bound AmtB and sequestering the PII protein to the membrane (**Figure 28 B**). The uncomplexed PIIs in the cytoplasm are degraded by protein kinases. The unavailability of cytoplasmic PII activates GlnR, which activates the transcription of the

nitrogen regulatory genes, however, the precise mechanism of PII/GlnR regulation needs be to be evaluated. Based on available data, the *Mtb* PII protein can interact with either AmtB (prevention of ammonium intake) or GlnR (transcriptional activation) to play a role as GlnK , guided by the adenylylation/deadenylylation by GlnD, in presence of ATP and possibly 2OG (**Figure 28**) [165,175]. Taken together, the diverse association of *Mtb* PII protein makes it an essential component of the bacterial nitrogen regulatory pathway.
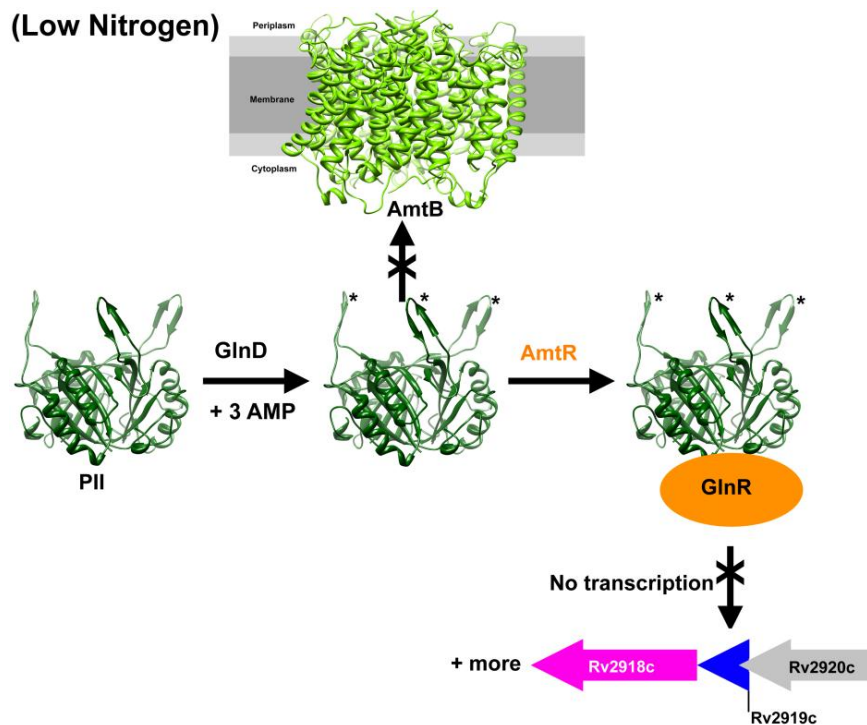
Figure 28: The putative nitrogen control pathway in *Mtb* in the presence of **A)** low nitrogen and **B)** excess nitrogen conditions.
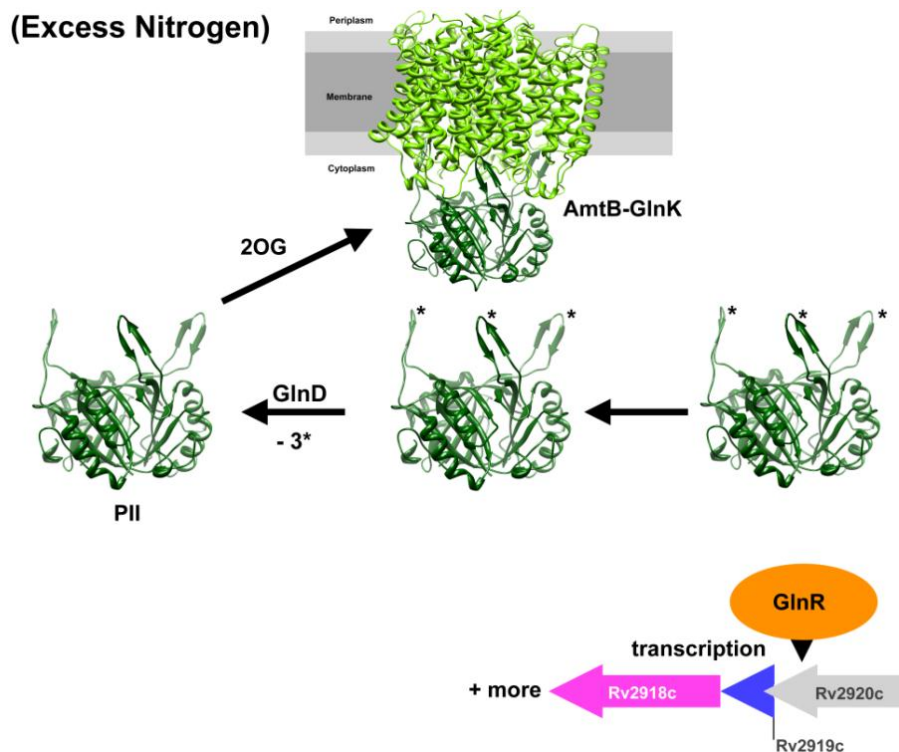
Figure 28: Continued.

Future Work

Our attempts to crystallize *Mtb* PII: ATP complex in the presence of Mg^{2+} ions were unsuccessful. However one cannot infer based on this observation that Mg^{2+} ion is not required for binding ATP. That a Mg^{2+} ion is not essential for binding of ATP or 2OG can be confirmed by mutation studies, where the Arg38 of *Mtb* PII protein is changed to Val38 as in the case of *M. jannaschii* PII protein. If such a Arg-Val mutant *Mtb* PII protein is crystallized in the presence of Mg^{2+} and ATP and Mg^{2+} ion is observed in the active site coordinating with the ATP molecule, then it can be confirmed that Arg38 replaces the function of Mg^{2+} ion in the *Mtb* PII protein and that Mg^{2+} ion is not required for the binding of ATP and orientation of the T-loop.

116

The actual binding site of 2OG on PII protein is not clearly established. A crystal structure of *Mtb* PII in complex with ATP and 2OG would shed light on the binding mode of the effector molecule and also help establish the mechanism by which *Mtb* PII functions.

Finally, while the homology model of *Mtb* PII: AmtB with the *E. coli* PII: AmtB suggests the possible role of *Mtb* PII as a GlnK protein, a crustal structure of such a complex would confirm the hypothesis.

CHAPTER IV

NATURAL PRODUCTS DERIVED INHIBITOR CANDIDATES

Background and Introduction

The Role of Microorganisms in Drug Discovery

Infectious diseases are caused by pathogenic microorganisms and are the leading cause of death worldwide. While such microbes were responsible for much of the suffering through the ages it was only recently that man was able to tap into the reserves provided by other nonpathogenic microbes to alleviate the pain caused by the former. It turns out that microbes are the best defense against microbes as microorganisms have contributed greatly in finding cure for the infectious diseases. Since the discovery of Penicillin in 1928, natural products derived from microorganisms like bacteria and fungi have played a major role in the treatment of infectious diseases. The discovery and the consequent application of streptomycin in the cure of tuberculosis exemplified the first instance where a major worldwide infectious disease could be controlled by the use of a naturally derived antibiotic and did much to raise the confidence of the scientific community towards natural products as an effective drug therapy. Streptomycin was isolated from *Streptomyces griseus*, an actinomycete, which subsequently proved to be the most prolific producer of antibiotics. New antimicrobials continue to be discovered from various strains of actinomycetes to this day, examples of which include the discovery of Abyssomicin from a rare actinomycete strain, *Verrucosispora* [176] and salinosporamide A from another strain of marine actinomycete, called *salinospora* [177].

Microorganisms have traditionally been a good source of drugs and lead compounds to the pharmaceutical industry. There are many popular examples of antibiotics, which include vancomycin, cephalosporins and rifamycins [178]. However, there is a continuous need to discover novel antibiotics. Antibiotics resistance among pathogens, evolution of new diseases like AIDS, Hanta virus and Ebola virus and the toxicity issues in some existing compounds make the process of drug discovery an ongoing effort. Fortunately, the metabolome of microorganisms contain a vast untapped reservoir of natural products that carry a wide range of structural diversity, of which many are yet to be discovered. It is fascinating to understand why microorganisms have evolved to produce such potent antimicrobial molecules.

Nature consists of microscopic ecosystems where the microorganisms are in constant struggle for survival amidst limited resources. Under such conditions, microorganisms produce biomolecules with antibiotic properties to defend themselves from the competing microbes in their immediate environment. In most instances these natural product antibiotics are secreted into the environment. The small molecules act on the competing pathogens either by inhibiting enzymes involved in important metabolic pathways or by targeting bacterial cell wall or cell membrane. They may also have bacteriostatic effects by targeting the protein synthesis in pathogens. The same logic applies to the treatment of infectious diseases in humans, where the prospective drug is chosen to target proteins in the invading pathogenic microorganisms like *Mycobacterium tuberculosis* (*Mtb*) in TB and *plasmodium falciparum* (*pf*) in malaria. It has been estimated that there are greater than ten thousand such protein targets for which small

molecule ligands are sought, and the small molecule space itself has been estimated to be in the range of 10^{14} to 10^{30} molecules [179]. To cover such a large chemical space requires a library of small molecules that constitute a wide variety of structural and chemical diversity. This is the primary advantage of using natural products for random screenings. The novel structural scaffolds provided by natural products owing to the presence of many chiral carbon centers is greater than can be provided by any available collection of compounds synthesized using combinatorial chemistry. Natural products also have the desired 'drug-like' properties, i.e. they can penetrate the cell walls and membrane of microorganisms and they can get metabolized to leave little or no toxic by-products. Indeed, between 60 – 80 % of all approved antibacterial and anticancer drugs were derived from natural product screenings [180]. It is believed that a large percentage of the world's biodiversity remains unexplored in terms of testing for their bioactivity. Technical difficulties ranging from the extraction, isolation and purification of novel natural products from nature has made them unattractive targets to the drug discovery companies which look for rapid returns in their investments. However the rapid growth in sophisticated analytical techniques and instrumentation like high performance liquid chromatography, mass spectrometry, nuclear magnetic resonance spectroscopy and high-throughput bioactivity assay methods in the recent years promises to considerably shorten the tedious and time consuming natural product extraction and identification procedures of the past.

Natural products from microorganisms are preferred as antibiotics because they have evolved over millions of years to perfect their function as an antimicrobial agent. It

would be rather unwise to turn away from this vast treasure trove of biodiversity available from nature when seeking novel compounds to fuel the drug discovery programs.

High-Throughput Screening Libraries

Improvements in screening technologies like high-throughput screening (HTS) and the development of efficient assay conditions for screening natural products for their bioactivity meant that a constant supply of newly discovered compounds was required. For example, HTS methods allow 100,000 or more chemicals to be screened in a single day (ref). In order to meet the challenge of the large screening capabilities of HTS systems combinatorial chemistry technology were developed through the 90s and have contributed considerably to the compounds database [179]. This approach involves the chemical syntheses of small molecules, largely based on the structural scaffold of natural products with drug like properties that are already known. While the synthetic procedures do afford chemical diversity to a given class of compounds, they are not a great source of structural diversity that is sought. 'Metabolic engineering' or 'combinatorial biosynthesis' methods have been used with success to produce diversity in natural products by manipulating the microorganisms at the genetic level [181,182]. The method was more target-specific and not capable of contributing to the natural compounds database at the desired rapid pace.

Microorganisms remain unrivalled in their ability to produce bioactive compounds that can be used to fuel drug discovery efforts. The sterically more complex

structures of natural products possess advanced binding characteristics compared to purely synthetic compounds [183]. Indeed about 40 % of the chemical scaffold found in nature are absent in today's medicinal chemistry [184,185]. New methods are therefore needed for discovering unique natural products with both structural and chemical diversity.

Whole-Cell Bioactivity Screening of Natural Product Extract Libraries

A vast majority of the antibacterials known today originate from the soil dwelling bacteria – Actenomycetes [186,187]. Yet actenomycetes continue to be the source of novel natural products to this day [188]. Continued discovery of novel natural products as lead candidates have sparked a renewed interest among the pharmaceutical industry to return to the field of natural product drug discovery [64]. Success in finding a novel natural product antibiotic depends upon the size and bio-diversity of the microbial extracts library. Since larger libraries present a higher success rate, collaborations were initiated with institutions that could offer diverse source of extracts for bioactivity studies. Thus, in addition to the soil bacteria that were isolated (described in the previous section), bio-activity guided fractionation were also performed on extracts obtained from different sources like the Vicuron library (FIIRV - Italy) and the Sarawak Biodiversity Center (SBC) (Malaysia).

As discussed previously, a collection of compounds can be screened against pathogenic microorganisms in a blind whole cell bioactivity assay or they can be screened against a selected target in the microorganism by conducting a directed *in vitro*

assay. Both methods have their advantages. The whole cell bioassay ensures that the compounds in the collection exhibiting poor cell permeability are removed from the priority list early in the screening process. However, the mode of cellular inhibition is unknown and compounds could be cytotoxic resulting in false positives. Also, in the case of a genuine inhibitor, the in-cell targets are unknown and the hit compounds present the possibility of reduced specificity towards the pathogen when tested in a host model. Cytotoxic compounds are usually easily identified and eliminated by performing simultaneous bioactivity assays with eukaryotic cells like yeast.

Target directed *in vitro* assays are performed with enzymes / proteins that have been validated as a good drug target. These are enzymes playing important roles in key metabolic pathways, are essential for the cell survival or act as important virulence factors. While the target directed *in vitro* assay screening with a compound library does not demand the knowledge of the biophysical and biochemical properties of the selected target, the routine does require that the desired target be over expressed and purified for the purpose of the assay. An inhibitor of the target protein identified in this manner has a greater probability of moving along as a lead candidate if it performs well in the whole-cell bioactivity assays as well. Natural products have an advantage in this regard since they are purported to have 'drug-like' properties, one of the criteria for which is good cell penetration power.

The following sections will discuss the screening of natural products in a targeted *in vitro* enzyme assay. In this section, examples will be provided of two commercially obtained natural products extract libraries that were screened for whole-cell bioactivity

123

assay. While the screening of the libraries is an ongoing process the section concludes with a discussion of the challenges encountered during the study, how they were overcome and the successful identification of a natural product inhibitor through the process of bioactivity guided fractionation of crude microbial extracts.

Methods

High Performance Liquid Chromatography Fractionation

The Vicuron library consisted of ninety six extracts. 200 µg / ml worth of extracts were provided for each of the ninety six extracts. The dry extracts were reconstituted with 150 µL water. The extracts were fractionated by injecting on to a reverse phase C18 column on a Waters® high performance liquid chromatography (HPLC) with a photodiode array detector, a manual injector and a WFC III® fraction collector, also from Waters®. 100 µL of the solution was injected on to the HPLC while the remaining 50 µL were stored for later LC / MS analysis. Forty fractions were collected per extract by running a gradient between water, methanol and acetonitrile, each containing 0.1 % formic acid. (LCMS grade solvents were used). The LC gradient conditions were as described below:

Column specification: 4.6 X 250 mm Waters Atlantis® T3, 5 µm (This column specification remains the same for all the subsequent LC fractionations on a semi preparative scale unless and otherwise mentioned)

Flow rate: 1 ml / min

Injection volume: 100 µL

Mobile phase: A = acetonitrile + 0.1 % formic acid, B = water + 0.1 % formic acid, C = methanol + 0.1 % formic acid.

Gradient conditions: 10 % C, 90 % B to 100 % C in 25 minutes; 100 % C to 100 % A in 10 minutes; 100 % A continued for 5 minutes.

The HPLC fractions for each extract were collected directly into the wells of a ninety six well deep well polypropylene plate. While the photodiode array ultraviolet detector maintains a record of all absorbance in the Ultraviolet (UV) range between 200 to 400 nm, the absorbance at 260 nm wavelength was extracted through the Empower® software for each of the ninety six LC chromatograms for the purpose of comparison. Out of the 1 ml fractions per well, 200 µL were aliquoted into a separate 96 well plate for later LC / MS analysis, while the remaining 800 µL fractions per well were evaporated and stock solutions made in DMSO to be used for whole-cell bioactivity assays. The ninety six extracts were grouped into similar UV chromatogram profiles, while taking into account the respective sources of the extracts. The UV data along with the bioactivity assay data was used to prioritize the active fractions of interest for further LC / MS analysis.

Twenty nine crude butanol extracts from the Sarawak library were fractionated by Sarawak Biodiversity Center (SBC) and sent for testing the whole-cell bioactivity against *M. smegmatis*. 50 µL, equivalent to 1 mg of crude extract, was injected into a Zorbax Eclipse XDB-C18, 5 µm, 4.6 X 150 mm reverse phase column. The LC gradient for the HPLC fractionation was as follows; 10 % methanol to 100 % methanol in 15 minutes; 100 % methanol maintained for another 15 minutes. The fractions were

collected directly into 96-well deep well plates at the rate of 1 ml per fraction per minute. 800 µL from each fraction was dried and sent for testing, while the remaining 200 µL were retained as reference.

LC gradient conditions for the semi-preparative scale fractionations of the six prioritized extracts – 25061, 26071, 26072, 32010, 32022 and 67037 were as follows;

Flow rate: 1 ml / min

Injection volume: 20 µL

Mobile phase: A = acetonitrile + 0.1 % formic acid, B = water + 0.1 % formic acid, C = methanol + 0.1 % formic acid.

Gradient conditions: 10 % C, 90 % B to 100 % C in 25 minutes; 100 % C maintained for 25 minutes; 100 % C to 100 % A in 10 minutes.

LC gradient conditions for the preparative scale fractionation of the six prioritized extracts – 25061, 26071, 26072, 32010, 32022 and 67037 were the same as that for the semi preparative scale fractionation, except the flow rate was maintained at 5.0 ml / min. A 10 X 250 mm Waters Atlantis® T3, 5 µm preparative scale HPLC column was used for the fractionation. The fractions were collected directly into 16 x 100 glass fraction tubes at the rate of 5 ml per fraction per minute.

The bioactive fractions number thirty / thirty one of extract 67037 were further fractionated on the semi preparative HPLC column to isolate the active compound. A randomly selected inactive fraction number 26 was also fractionated to serve as a negative control. The gradient conditions for both bioactive and inactive fractions were identical and consisted of an isocratic flow of 100 % C (methanol + 0.1 % formic acid)

for a period of 40 minutes at a flow rate of 1 ml / min. The forty fractions were collected in a 96-deep well polypropylene plate.

Upon testing the above fractions for bioactivity against whole cell *Mtb* and *M. smegmatis*, fraction number five was identified as bioactive. The remaining bioactive fraction number five was concentrated and subjected to further fractionation on the HPLC using the semi preparative C18 column. The gradient conditions for the fractionation were as follows;

10 % A - 90 % B to 100 % A in twenty five minutes; 100 % A maintained for fifteen minutes; 100 % A to 10 % A – 90 % B in five minutes. Forty five fractions were collected in a 96 – deep well polypropylene plate. The generated fractions were again tested for their whole cell bioactivity against *Mtb* and *M. smegmatis* cells. The bioactive fractions from this set of fractions were analyzed by LC / MS.

Whole-Cell Bioactivity Assay

An aliquot of each fraction was used for the *M. semegmatis* and *Mtb* whole cell bioassay. Yeast whole cell assay (against *C. albicans*) was tested as a surrogate for cytotoxicity assay. The bioactivity tests were performed in 384 well formats. Rifampicin, at concentrations ranging from 100 µM to 0.01 µM and an inhibitor, Vidya2, at concentrations ranging from 25 µM to 0.003 µM were used as positive controls against *M. smegmatis* and *Mtb* whole cells. Amphotericin, at concentrations ranging from – µM to – µM concentration was used as the positive control against *C. albicans*. Cell growth in the presence of media and absence of any inhibitor or fractions

were considered as positive controls. The 384 well plates containing the cells were
incubated at 37 ˚C for 48 hours for *M. smegmatis* cells and 120 hours for *Mtb* cells. After
this period, 1.25 μL of alamar blue (resazurin dye) was added to the wells and left for
overnight in the incubator at 37 °C. The wells were then measured
spectrophotometrically at 544 nm and 590 nm for cell growth. Wells with cell growth
are able to reduce the alamar blue which then turns pink. The resazurin dye in the wells
with no cell growth remained blue. Since the concentrations of the compounds in each
fraction were unknown, an arbitrary value of 50 X was assigned to the active compound
in each original extract dissolved in 150 μL solvent. Based on this arbitrary value and
taking into account the dilution factors during the fractionation, evaporation of the
fractions, resolubilization of the individual fractions and preparation of stock solutions
from the fractions, the fractions were tested at a final concentration of 0.5 X. A
representative layout of the fractions in a 384 well plate format along with the
appropriate controls is shown in the figure 29 below;

				67037.30.31-5/b								Prot-4 gradient											
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	DMSO	100uM A1	H2	A3	H4	A5	H6	A1	H2	A3	H4	A5	H6										25uM	
B	DMSO	50																					12.5	
C	DMSO	25 B1	G2	B3	G4	B5	G6	B1	G2	B3	G4	B5	G6										6.26	
D	DMSO	12.5																					3.125	
E	DMSO	6.26 C1	F2	C3	F4	C5	F6	C1	F2	C3	F4	C5	F6										1.5625	
F	DMSO	3.125																					0.78125	
G	DMSO	1.5625 D1	E2	D3	E4	D5	E6	D1	E2	D3	E4	D5	E6										0.390625	
H	DMSO	0.78125																					0.195313	
I	DMSO	0.390625 E1	D2	E3	D4	E5	D6	E1	D2	E3	D4	E5	D6										0.0976	
J	DMSO	0.195313																					0.0488	
K	DMSO	0.0976 F1	C2	F3	C4	F5	C6	F1	C2	F3	C4	F5	C6										0.0244	
L	DMSO	0.0488																					0.0122	
M	DMSO	0.0244 G1	B2	G3	B4	G5	B6	G1	B2	G3	B4	G5	B6										0.0061	
N	DMSO	0.0122																					0.00305	
O	DMSO	0 H1	A2	H3	A4	H5	A6	H1	AA	H3	A4	H5	A6										0	
P	DMSO	0																					0	
		Rifampicin																					Vidia2	

Figure 29: Layout of fractions, positive controls and negative controls in a typical 384 well format. DMSO was used instead of the fractions as a positive control. Rifampicin and the compound Vidya2 were used as negative controls. The wells labeled with A1, B1, C1… represent the wells with the fractions obtained from the HPLC.

129

Liquid Chromatography / High Resolution Mass Spectrometry

The bioactive fractions in the original fractions received from SBC were analyzed by direct infusion into the high resolution micrOTOF-Q II® mass spectrometer from Bruker Daltonics Inc to help prioritize the extracts of further interest. Liquid Chromatography / Mass spectrometry (LC / MS) data was obtained for the active fractions from the prioritized extracts of the Vicuron extract library as well as the SBC collection of extracts on an Agilent 1200 Infinity series HPLC with temperature controlled autosampler and a photodiode array detector and the in-line micrOTOF-Q II® mass spectrometer from Bruker Daltonics Inc. The LC conditions were as follows:

Column specification: 2.1 X 100 mm, Waters Atlantis T3®, 5 µm

Flow Rate: 0.5 ml / min

Mobile phase: A = water + 0.1 % formic acid, B = acetonitrile + 0.1 % formic acid.

Gradient conditions: 90 % A, 10 % B to 100 % B in 8 min; 100 % B continued for 4 min; 100 % B to 90 % A, 10 % B in 2 min; 90 % A, 10 % B continued for 3 min.

The high resolution mass spectrum (HRMS) was collected in electro spray ionization (ESI) positive mode in the mass range of 150 m/z to 2000 m/z. The ESI source settings were as follows;

End plate offset = - 500 V

Capillary = - 4500 V

Nebulizer = 4.0 Bar

Dry Gas = 11.0 L / min

Dry temp = 200° C

MS peaks of interest were identified by comparing the HRMS data for the active

fractions with those of the neighboring inactive fractions and also from solvent (blank)

runs to eliminate the insignificant MS peaks from consideration. Probable molecular

formula was generated using the high resolution MS data for the MS peaks of interest.

SmartFormula® prediction calculation available with the Data Analysis® software from

Bruker Daltonics Inc. was used for the generation of predicted molecular formulas for

the MS peaks of interest.

Dereplication of Actives

Dereplication is a process of screening the identified bioactive chemical entities

against a database of previously discovered molecules to recognize molecules that

already exist in the database [189]. In the natural products drug discovery process repeat

discoveries of bioactive compounds are a common occurrence and the dereplication

process is used to eliminate such entities, unless the bioactive compound represents an

activity that is different from the one published in which case it may be retained as a new

activity. This is a method of data reduction to eliminate redundant data and concentrate

on the unique MS data.

The molecular formula generated for the MS peaks of interest from the bioactive

fractions were subjected to a process of dereplication by matching the generated

molecular formula and the HRMS data against the Dictionary of natural products (DNP)

database. If the molecular weight, molecular formula, the biological source (where

available) and the bioactivity data of the DNP entry matched those of the fractions, then

the results were retained as a plausible hit. The fractions for which no match could be found were considered unique, possibly novel compounds and hence worthy of further investigation to identify the active compounds.

Results and Discussion

Vicuron Extract Library

Previously, a collection of 140,000 natural product extracts were obtained from FIIRV. The sources of these extracts were mainly rare actinomycetes, myxomycetes and fungi. 96 extracts were chosen from the total 140,000 extracts based on their previous bioactivity against gram positive bacteria. These were prioritized to enrich the collection with natural product extracts that were more likely to give anti mycobacterial hits. About 200 uL worth of fermentation broth from the selected 96 extracts were acquired in one 96 well microtitre plate to test their bio-activity against mc^2155 strain of *M. smegmatis* and the mc^27000 strain of *Mtb*. *M. smegmatis* is a fast growing generally nonpathogenic mycobacterium that is considered a surrogate for *Mtb*. The 96 extracts in the polypropylene plate were labeled according to their well positions.

All 96 extracts were fractionated into forty fractions per extract. The LC condition was primarily a gradient between 10 % methanol and 100 % acetonitrile. The fractions were collected directly into the wells of a 96 well polypropylene plate. 100 uL of each fraction was aliquoted into a separate plate and saved for later LC/MS analysis of the desired active fractions. The fractions were dried under vacuum and reconstituted

in dimethyl sulfoxide (DMSO). Stock solutions were made for testing the whole cell anti-mycobacterial activity against strains of *M. smegmatis* and *Mtb*.

Out of the 96 extracts, 33 extracts had fractions that were active against *M. smegmatis* and only three extracts had fractions that were active against *Mtb*. Given that the 96 extracts were selected based on their previous activity against gram positive bacteria, the results of the whole cell bioactivity assay came as a surprise as all 96 or at least a majority of the 96 extracts were expected to have positive bioactivity against *M. smegmatis* and *Mtb*. However it is interesting to note that the previous activity in the parent collection came from testing the whole extracts and not the various fractions. This suggests that the previous activity could have originated from a mixture of compounds that acted in a synergistic or additive fashion. This is typically observed in most herbal and ayurvedic medicines where the activity arises from multiple components of the extract acting in combination. When the 96 prioritized extracts were fractionated it is likely that compounds that previously acted in synergy got separated on the column and thus the activity was lost when each fraction was tested individually. Reproducibility is always an issue in any kind of high-throughput or low-throughput screening. For reasons including human error or experimentally introduced error, it is not always possible to exactly reproduce a screening assay. Such unforeseen errors in reproducibility may have also caused the extracts to stop showing whole-cell bioactivity. In either case, the number of extracts showing activity was considered significant and it was prudent to focus on the active extracts to look for novel bioactive natural products. The statistics of the active fractions were tabulated and are presented in the table 6 below. A total of 39

133

extracts had bioactivity on *Mtb*, *M. smegmatis* or *Candida albicans*. Cells with values that were considered significant for the purpose of prioritizing the extracts are colored (**Table 6**).

Two of the *Mtb* active extracts overlapped with the *M. smegmatis* extracts. These were extracts at well positions B3 and E8. Apart from the *M. smegmatis* and *Mtb* activity the fractions were also tested against *C. albicans* for their antifungal activity. The eukaryotic whole cell activity was considered as a surrogate for cytotoxicity assay. While a significant number (16) of the 96 extracts displayed antifungal activity, activities for six of these extracts overlapped with *M. smegmatis* whole cell activity and two of them overlapped with the *Mtb* activity (Highlighted in **Table 6**). Although the primary aim of the experiment was to find natural compounds that had anti-mycobacterial activity, out of curiosity the extracts showing antifungal activity were also added to the priorities list as these were also interesting and worthy of further investigation.

Table 6: A summary of the bioactivity observed in the 96 Vicuron extracts.

Type	summary FIIRV data	DB interpret	Compound ID	Strain ID	NP_original plate	Location	Actives in the 96 Vicuron extract library	M. smeg active fractions	M. smeg activity repeated	M.tb active fractions	M.tb activity repeated	In vitro MLS active fractions	In vitro retest @ dilutions	Antifungal active fractions	Associated UV Peak	MS peaks of interest	Extracts of further interest	Comments
Ractinos	bacterial PS	PS/MT	NP-874	96888	423	A2	A1	10, 30		No		No	No	No	30	618.1881, 376.3584, 458.3487	A1	
Ractinos	septoria tritici, bacterial ps	PS	NP-1286	92725	429	B4	A2	11 fractions		No		No	No	No				
Strepto	gram positive, septoria tritici	CW	NP-3507	97733	466	F7	A3	25 fractions		No		No	No	No				
Ractinos	NA	0	NP-6215	97476	501	E5	A4	9 fractions		No		No	No	No				
Ractinos	Hypersensitive S. cerevisiae	HYPER	NP-9471	57096	542	B10	A5	35 fractions		No		No	No	No				
Ractinos	NA	0	NP-13219	99511	589	A9	A6	39 fractions		No		No	No	No				
Fungi	NA	0	NP-16505	100381	630	B5	A8	40		No		No	No	No	No			
Fungi	NA	0	NP-23262	101801	714	F2	A9	12 fractions		No		No	No	No				
Fungi	NA	0	NP-26767	102799	758	D7	A10	16 fractions		No		No	No	No				
Fungi	NA	0	NP-31123	104451	812	H3	A11	38-40		No		No	No	No	No			
Myxo	NA	0	NP-203	116770	1185	E5	A12	36, 38		No		No	No	No	No			
Ractinos	NA	0	NP-1348	92747	429	H6	B2	2-5, 35, 36, 39		No		No	No	No	2, 4	368.203, 421.205, 518.2926, 735.3939	B2	Big UV peak at 4 min
Ractinos	NA	0	NP-4402	94093	477	H2	B3	22-25	22, 24, 30-35, 37	22		No	No	No	22	304.2645, 449.1721, 561.537, 563.5534	B3	Mtb activity
Ractinos	NA	0	NP-6640	98078	506	G10	B4	22, 35	40	No	15, 22	No	No	No	22		B4	
Ractinos	NA	0	NP-10279	98946	552	C9	B5	29, 30, 32-40		No		No	No	No	29, 30	307.1269, 365.1363, 591.2767, 348.2	B5	
Ractinos	NA	0	NP-13407	99738	591	D7	B6	29, 30, 34-37, 39		No		No	No	No	29	301.1404, 323.1237	B6	
Strepto	NA	0	NP-15259	100322	614	E9	B7	32-36		No		No	No	No	32	348.3267, 254.2497, 276.2314	B7	
Fungi	gram negative	MALT	NP-16817	100476	634	A7	B8	29, 31-37, 39		No		No	No	No	29	430.3, 348.3	B8	
Fungi	NA	0	NP-23636	101910	719	C6	B9	33, 35		No		No	No	No	No			
Fungi	NA	0	NP-26992	102838	761	C2	B10	32-37, 39		No		No	No	No	No			
Ractinos	Tolc - , S. maltophilia, septo	MALT	NP-888	96921	423	B6	C1	30		No		No	No	No	30	432.2734, 436.3193, 445.3181, 446.2882, 460.2712, 465.3368, 474.3472, 616.5767	C1	
Ractinos	E. coli Tolc - MR	MT	NP-4520	94232	480	C10	C3	36		No		No	No	No	No			
Ractinos	NA	GM-	NP-8101	62998	525	A11	C4	23, 24, 35		No		No	No	17	23, 24	361.2457, 432.2222, 634.2507, 1072.4385		
Fungi	NA	0	NP-18670	93268	657	B10	C8	35		No		No	No	No	No			
Fungi	NA	0	NP-27702	103077	770	B2	C10	36		No		No	No	No	No			
Ractinos	nial ps, bacterial RNA polym	DIAC	NP-11046	99167	561	H6	D5	21-25	29-37	No	6, 27-29	No	No	No	21-25	421.2092, 466.2317, 505.253	D5	
Fungi	gram positive, fungi, toxic	CW	NP-15508	91161	617	F8	E7	No		27, 28	6, 27-29	No	No	27, 28	27, 28	375.2003, 423.1644, 462.3191, 646.4674, 678.4541, 749.3921, 760.499, 808.63, 1060.4467, 355.2821, 381.2994, 602.9178, 609.9259, 768.5711, 850.7982, 906.8474, 928.8171, 944.7280, 960.6404, 966.7092, 990.6542, 1004.8076, 1045.6961, 1104.7803, 1118.7961, 1204.8382, 1218.8562	E7	2 intense MS peaks
Ractinos	NA	MT	NP-5426	94845	491	F6	E3	28, 29		No		No	No	No	28, 29		E3	good UV peaks
Ractinos	NA	0	NP-13865	99960	597	B5	E6	No		No		No	No	36, 37	No			
Fungi	m positive, gram negative, fu	ALL	NP-19602	100603	668	H2	E8	22-24	25-26	22-24	24-27	No	No	22-24	22-24	442.2848, 653.4432, 768.5694, 869.5457, 883.5608	E8	interesting
Myxo	NA	MT	NP-222	116764	1185	G4	E12	2		No		No	No	8 (doubtful)	No			
Fungi	NA	0	NP-16463	103169	629	F3	H7	35, 36, 38		No		No	No	31	No			
Myxo	NA	MT	NP-200	116746	1185	E2	H11	32, 36		No		No	No	No	No			
?	?	?	NP-24772	?	733	E2						4	No		4			
?	?	?	NP-57831	?	1146	F11						8, 38	38	8		1146-F11	
?	?	?	NP-58022	?	1149	B2						No	19	19		1149-B2	
?	?	?	NP-24852	?	734	E2						6, 30, 34, 35	34, 35	30		734-G10	
?	?	?	NP-24880	?	734	G10						No	32-35	32-35			
?	?	?	NP-98951	?	1598	F8						No	No	No			
							Total = 39										Total = 15	

A general observation among the active fractions was that activity was observed beyond the 20th fraction in most of the cases with some exceptions. The gradient system for HPLC fractionation from polar solvent to the non-polar solvent meant that the polar components of the extract mixtures eluted first followed by the nonpolar components of the extract in that order. Thus the compounds eluting beyond the 20th fraction can be expected to have nonpolar characteristics with few polar functionalities. It has been observed that compounds from a compound collection with good inhibitory activities that make it to the drug discovery stage are typically lipophilic (non-polar) molecules [190]. Thus the later fractions from the HPLC separations would be more desirable as possessing anti-mycobacterial activity.

After a high-throughput or low-throughput screening, the next step is to prioritize the compounds or extracts to be taken to the next level for further investigation. Thus the 33 *M. smegmatis* active extracts were narrowed down to the few most promising extracts that could potentially allow the identification of a novel anti-mycobacterial or antifungal compound. Since the collection of 96 extracts consisted of rare actinomycetes, myxomycetes and fungi, it was possible that one or more of the extracts belonged to the same species and strain of microbe. In such cases it would be unintuitive to select two or more extracts that belong to the same strain for further investigation. Thus an additional step in the prioritization was to analyze the LC chromatograms of all 96 extracts and group them into extracts with similar biological origins. When the LC profiles of all 96 extracts were compared to each other patterns began to emerge. Indeed it was seen that

the LC profiles of many extracts could be matched to other extracts in different well

positions. The groupings of the different chromatograms are shown in Figure 30 and 31.
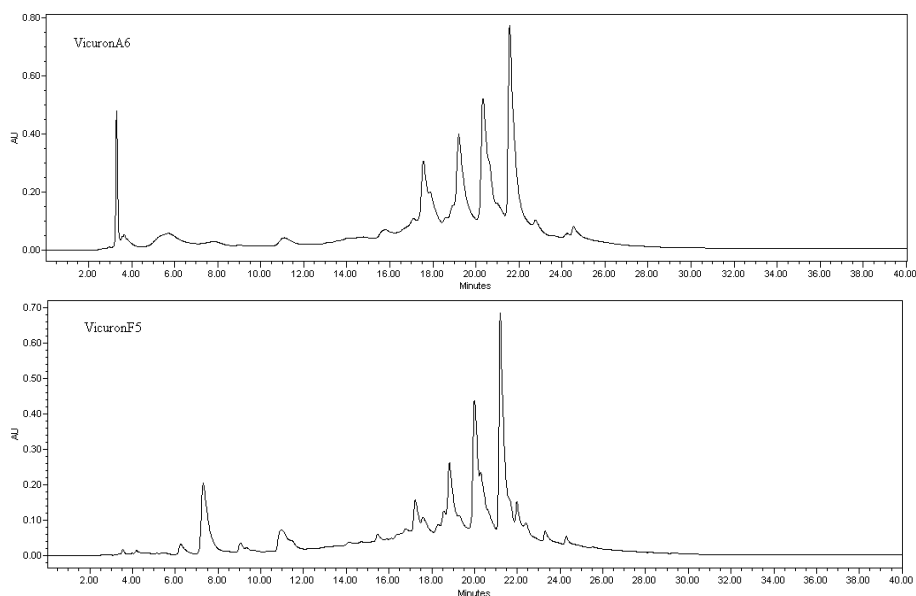
Figure 30: The LC UV profile chromatograms of two extract at well positions A6 and F5. The similarities in the LC profiles of the two are apparent. Both extracts were sourced from rare actinomycetes.
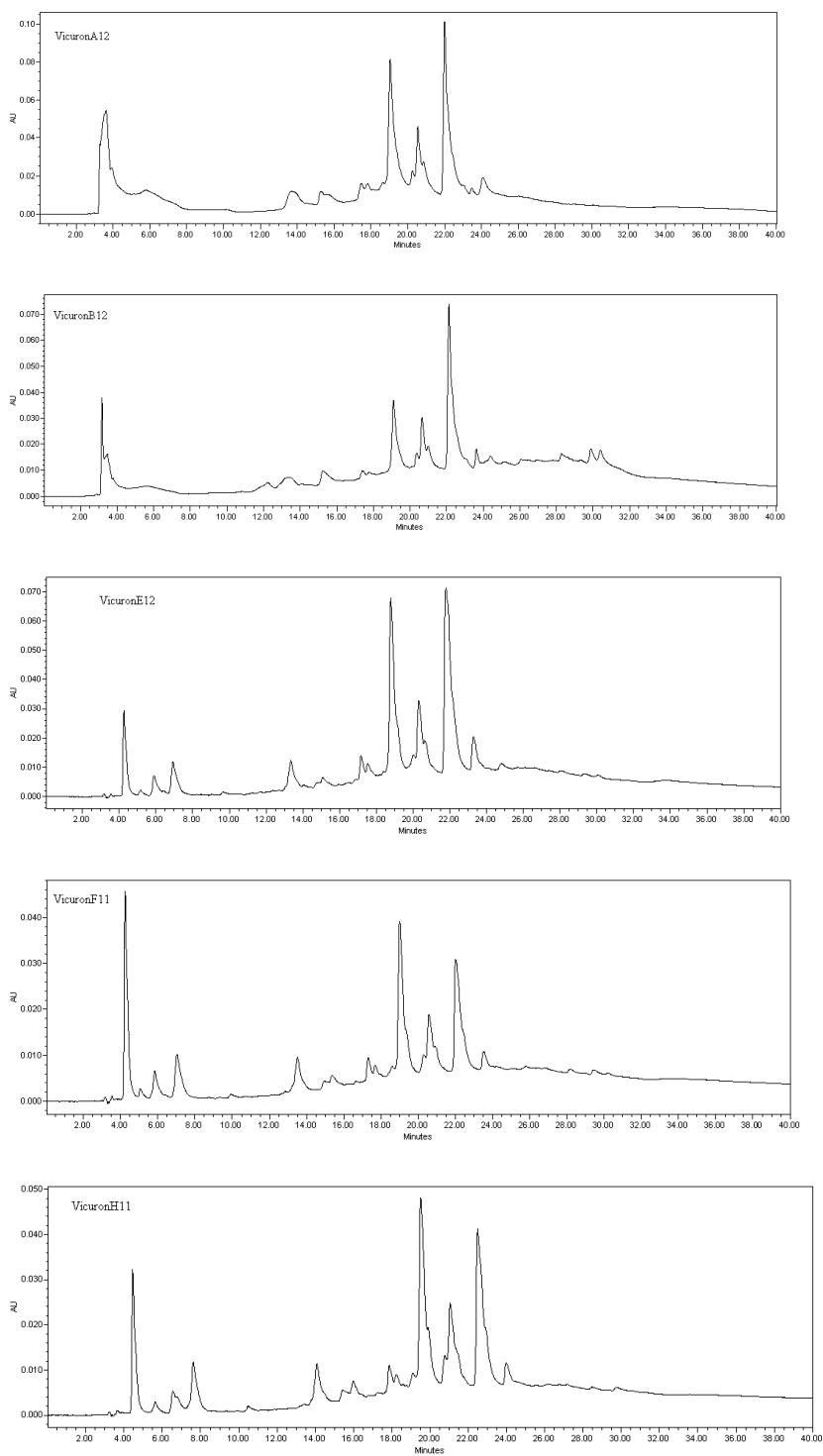
Figure 31: LC chromatograms of five different extracts sourced from myxomycetes. The position of the extracts in the 96 well plate were – A12, B12, E12, F11 and H11.

High resolution (HR) LC / MS data in the electrospray ionization (ESI) positive mode was collected for all bioactive fractions. The HRMS data was analyzed and compared with other active and inactive fractions and also with the blank solvent runs to identify MS peaks of interest. MS peaks that seemed to be unique to the active fractions and that could potentially represent the active component of the fraction were identified. The Data analysis® software on the Bruker micrOTOF Q-II MS instrument has a built-in SmartFormula® prediction program. The SmartFormula program was used to generate a list of predicted molecular formula for each MS peak of interest. High resolution MS data was used in the molecular formula prediction and in cases where the prediction list was greater than five, the results with the top five scores were retained.

The next step in the identification of the active compounds was dereplication. Databases exist of all the previously identified, characterized and published natural products, like ChemBank, PubChem antiBase and Chapman & Hall's Dictionary of Natural Products (DNP) [191]. In natural products drug discovery it is common to encounter a previously discovered compound in a given fermentation extract. Different strains of the same species are known to produce the same antibacterial compounds under different environmental conditions. In order to save time and to prevent the rediscovery of known natural products, the dereplication process is a necessity. In this process preliminary information about the bioactive compound like their molecular weight, MS fragmentation pattern or LC-NMR [192] data are compared with those existing in the various databases to identify already known compounds and their analogs [193]. The predicted molecular formula was very useful information about the bioactive compounds

and was used as the basis along with the high resolution mass to perform dereplication against the database of over 200,000 natural products contained in DNP. No natural products with the same high resolution mass and the same elemental composition as those predicted were found. Hence it was concluded that the observed activity most likely came from a novel natural product or at least an analog of a known natural product.

The UV absorbance data collected during the initial fractionation was also cross referenced with the activity data and the HRMS data on the active fractions to eliminate fractions that did not have any associated UV peak. Only the active fractions with an associated UV peak in the HPLC profile were considered because it would be easier to follow the bioactivity of compounds that exhibited a UV absorbance in the subsequent LC fractionation steps. As will be evident in the following paragraph, this feature was given much importance while prioritizing the extracts of interest for further investigation.

Nine extracts were selected for further investigation and the reason for their selection is given below:

B2 – This extract represented a unique biological source whose LC UV profile spectra did not match with that of any other extract in the collection. The extract had whole-cell bioactivity against *M. smegmatis*. The fractions that were bioactive also had strong UV peaks associated with them. The MS peaks of interest were in the range of 300 to 700 daltons, which are in the desired range for a drug-like molecule.

B3 - This extract had fractions that were active against *M. smegmatis* as well as *Mtb* in the whole-cell bioassay. This is one of the three extracts that showed bioactivity against *Mtb*. The same fraction that was active against *M. smegmatis* was also active against *Mtb*. Further the activity was reproducible in both bioassays. Besides, the active fraction also had an associated UV peak.

B5, B6 and B7 - These extract had fractions that were active against *M. smegmatis* with associated UV peaks for the fractions of interest. The active fractions were observed towards the nonpolar end of the LC gradient making the active compound or compounds possibly lipophilic in nature which is a desirable characteristic to have in a lead molecule.

C4 - This was among the few extracts that had both *M. smegmatis* activity and anti-fungal activity against *C. albicans*. However it did not show any activity against *Mtb* whole cells. The active fractions had associated UV peaks.

E3 - The extract had whole-cell bioactivity against *M. smegmatis* and strong UV peaks associated with the fractions containing activity. The active fractions of this extract also had the most number of unique MS peaks of interest with no matches against the natural products database during dereplication.

E7, E8 – These were the other two fractions that were active against both *M. smegmatis* as well as *Mtb* in the whole-cell assay. Further the activities were reproducible. Extract E7 had intense MS peaks of interest, which might be easy to follow upon refermentation and fractionation. During the dereplication of the masses obtained from the active fractions of extract E8, the antimicrobial compound fusaricidin

came up as one of the possible hits. This was an interesting find as antimycobacterial activity for this compound would be a novel activity. The source of the extract for E8 is known. It is a fungi and falls under the genus fusarium.

Sarawak Biodiversity Library of Extracts

29 extracts were short listed from a library containing a total of 145 microbial extracts, to test for their bioactivity against *M. smegmatis*. 17 of the extracts were sourced from different strains of the *Streptomyces species*. The sources for the remaining 12 extracts were unknown. The extracts were fractionated at Sarawak into 30 fractions per extract. The first 20 fractions were collected and the eluting solvents were removed by evaporation. The dry fractions were transferred into 96 well microtitre plates. The dry fractions were reconstituted into 20 µL per fraction and were assigned an arbitrary concentration of 50 X. The fractions were tested for bio-activity against *M. smegmatis* at a final concentration of the fractions of 0.5 X. The bioactive fractions were also tested for cytotoxicity against human cancer cell lines to eliminate the generally cytotoxic and nuisance compounds. Eight out of the 29 extracts had fractions that were active against *M. smegmatis*. The active extracts constituted 24 % of the total number of extracts tested. When the fractions were retested at double the concentration (1 X), 16 out of the 29 fractions showed activity, which included the previous eight extracts. The active extracts from the rescreening constituted 55 % of the total number of extracts. When compared with the cytotoxicity data, it was observed that six of the *M. smegmatis* whole-cell bioactive extracts were cytotoxic. The six cytotoxic extracts included four of the

original *M. smegmatis* bioactive extracts (that tested positive for whole-cell bioactivity at low concentrations of 0.5 X). HRMS data was obtained for all bioactive fractions that were non cytotoxic, by direct infusion into the MS using an ESI probe in both positive and negative modes.

The MS spectrums were compared with those of other bioactive fractions and with inactive fractions that were randomly picked. This was done to identify MS peaks of interest in the bioactive fractions that did not exist in the inactive fractions. If a certain MS peak appeared consistent in an adjacent bioactive fraction for the same extract but did not appear in an inactive fraction of the extract it was considered to be an MS peak of interest as it most likely corresponded to the bioactive component in the fraction. Likewise if an MS peak in a bioactive fraction was consistently observed in a neighboring inactive fraction it was most likely that the peak did not correspond to the active molecule. One possibility is that a bioactive compound may elute out of the HPLC column into two or more adjacent fractions, however, it may only be present in sufficiently high concentrations in one of the fractions to be able to respond to the bioactivity assay. A mass spectrometer with sufficiently high sensitivity can detect the mass of the active compound in all the fractions where it eluted even though it is at low concentrations. During analysis this phenomenon may cause a false negative data causing the MS peak corresponding to the active compound to be rejected as an uninteresting mass peak. In order to overcome this issue, random fractions that were far away from the bioactive fractions were picked as negative controls. It is unlikely for a

MS peak corresponding to the bioactive compound to occur in random fractions away from the bioactive fractions.

SmartFormula® prediction calculation was used to generate molecular formula for the MS peaks of interest. Among other criteria, the SmartFormula® calculation uses the error (in ppm) in the high resolution masses of the predicted molecular formula and the experimentally observed mass to score the generated molecular formulas. Depending on how well the mass spectrometer is calibrated, the predicted molecular formula with the lowest ppm error has the highest probability of being the real molecular formula. Since calibration of a Time-of-Flight (TOF) MS can deviate marginally over time the experimentally observed masses cannot be relied upon as the real masses. An error of up to 10 ppm was allowed in the experimentally observed masses for the SmartFormula® program to accurately predict the molecular formulas. For this reason, molecular formula predictions with up to 10 ppm or the top five predicted molecular formulas calculated by SmartFormula® were considered. Another factor to be considered in scoring the predicted molecular formulas is the mean standard deviation or the 'mSigma' value, which accounts for the variation in the isotopic peak patterns between the experimentally observed set of peaks and the predicted isotopic peak pattern of a molecular formula. For the Bruker micrOTOF Q-II® MS instrument a variation of 20 ppm or less was considered to be within the acceptable range for the SmartFormula® calculation to accurately predict a molecular formula.

The list of predicted molecular formulas thus generated and sorted for each MS peak of interest were collected and run through the process of dereplication. As

144

discussed in the previous sections dereplication process is helpful to identify previously published compounds and eliminate them from the list of prioritized extracts. While most of the MS peaks did not match with the available natural products in the dictionary of natural products (DNP), there were a few hits such as Manzamine for extract # 26072 that were interesting finds. They were of interest because Manzamine and its analogs are reported as inhibitors of *Mtb* [194,195], however this natural product is reported to be found typically in marine sponges. The source of extract # 26072 was *Streptomyces misionensis* and the genus *Streptomyces* are typically soil dwelling microorganisms. It would be an interesting find if Manzamine were indeed established to be the real source of the bioactivity as it would suggest a new source for the natural product. Natural products reported to be originating from certain marine sources like sponges have occasionally been corrected to be originating from other symbiotic and epiphytic microorganism that reside inside the sponges [196]. Examples of these exist in the dictionary of marine natural products like Swinholide A, which was originally believed to be produced by the sponge, *Theonella Sinhoei* [197]. This alkaloid was later isolated from the cyanobacterium *Aphanocapsa feldmanni* that exists inside the sponge [198].

The available LC / UV profiles were also compared to each other and with their respective activity profiles. Extracts with bioactive fractions that had associated UV peaks were desirable as these would be easy to follow through subsequent purification steps. Six extracts were selected for further analysis either because they presented the possibility of finding a novel antibiotic or to confirm the presence of a previously

published compound having a new bioactivity. These were extracts # labeled as 25061, 26071, 26072, 32010, 32022, and 67037 and will hence be referred to by their labels.

Large scale crude extracts were obtained for the six prioritized extracts. An aliquot consistent with the amount previously analyzed was processed in the same manner as before, then fractionated and the bio-activity tested to confirm that the activity was reproducible in the refermentation. Upon confirmation of activity, a preparative scale fractionation was performed using a preparative reverse phase C18 HPLC column at a flow rate of 5 ml/min. The same gradient of mobile phase solvents was used as for the small scale extracts. However the length of the gradient was expanded to cover 60 fractions. Stock solutions of the fractions were prepared so that they could be tested in the whole cell bioactivity assays at 0.5 X like the previous batch of small scale extracts. Two out of the six extracts – 26071 and 67037 had whole cell activity against *M. smegmatis* and 26071, 32022 and 67037 had whole cell activity against *Mtb*. The different bioactivity for each extract and the active fractions are represented below in Table 7.

Table 7: A summary of the 29 Sarawak extracts and their bioactivity

Source of the extract	Extract name	*M. smeg* active fractions	Cytotoxic fractions	Large scale extracts	New *M.smeg* activity (sm scale fractn.)	New *M. tb* activity (sm scale fractn.)	Prep *M. smeg* activity	Prep *M. tb* activity
Streptomyces sp. SXY66	8050	No	No	_				
Streptomyces parvulus	8062	16, 17	16-19	_				
unknown	8065	No	No	_				
unknown	9020	No	No	_				
Streptomyces seoulensis	9050	No	No	_				
unknown	13067	19, 20	16, 19, 20	_				
unknown	14056	17-19	15-19	_				
Streptomyces sp. M4032	25049	17	13-18	_				
Streptomyces sp. M4032	25054	17, 19	13-19	_				
unknown	25061	18, 19	No	Yes	36, 38, 40-43, 47-59	No	32, 33	33 (doubtful)
Streptomyces sp. 210801	26008	No	19					
Streptomyces sp. L-2-14	26071	4, 14-18	No	Yes	12-21, 30, 33-41, 45-57, 59	4, 12-23	4, 5, 8, 9, 14	6, 9, 10, 12, 13, 14, 15, 16
Streptomyces misionensis	26072	4	No	Yes	24, 34, 38, 40, 45, 47, 48, 51, 53, 55-56		No	No
unknown	31059	19	19					
unknown	32010	10	No	Yes	38, 40, 42-44, 47, 49-57, 59-60		No	No
unknown	32011	No	No					
Streptomyces sp. SXY66	32022	12	No	Yes	28-30, 34, 39-40, 44-45, 49-52, 54-57	40	No	15-16, 32, 47
Streptomyces sp. SXY66	32031	16, 19	No					
Streptomyces Chromofuscus	32033	No	No					
Streptomyces Chromofuscus	33059	No	No					
Streptomyces Chromofuscus	33068	No	No					
Streptomyces sp. IM - 7464	49045	No	No					
Streptomyces sp. ACT - 0095	50025	No	No					
unknown	51054	13	No					
unknown	54066	No	No					
Streptomyces sp. TA 4 - 8	65055	No	No					
unknown	65067	19	No					
unknown	67026	16	No					
Streptomyces Chromofuscus	67037	11	No	Yes	33-35	33-34	21, 30, 31, 46, 49, 50, 52	30, 14, 56-60
	67037-30-31						5,6,7	5, 6, 31

The extracts prioritized for further study are marked red. The fractions of the extracts that were bioactive against *M.smegmatis* in the original screen are shown in the blue cells. The cytotoxic extracts are shown in the grey cells. Fractions of the prioritized extracts that were refermented, refractionated and tested positive for bioactivity in the whole-cell assay against *M. smegmatis* cells are shown in the purple cells. Fractions that tested positive in the bioactivity assay against *Mtb* cells are shown in maroon. The last row colored dark red represents the active fractions of extract # 67037 that were further fractionated and tested for *M. smegmatis* and *Mtb* bioactivity.

The Story of Extract 67037

The activity profile of extract 67037 will be followed in this section. The 16s rDNA sequence of extract 67037 revealed the strain to belong to *Streptomyces* sp. SXY66. A butanol extract was shown to have positive antimicrobial activity against *Micrococcus luteus* and *Pseudomonas aeruginosa*. Further the crude extract was shown to have no inhibitory activity towards *Aspergillus niger*, *Bacillus subtilis*, *Escherichia coli* and *Saccharomyces cerevisiae*. Fractionation of the small scale extracts were conducted by S.B.C. and the dried fractions were sent for further testing against *M. smegmatis* and *Mtb*.When the fractionated extract was tested for whole-cell mycobacterial activity, fraction number 11 was bioactive against *M. smegmatis*. An aliquot of fraction number 11 was injected by direct infusion into the Bruker micrOTOF® Q-II mass spectrometer. MS peaks of interest were manually picked by comparing the MS data with those of the neighboring inactive fractions. As there was no match for the MS peak of interest in the dictionary of natural products database, the extract was considered worthy of further investigation. Large scale extracts for 67037 were obtained by refermentation on a large scale. An aliquot of the large scale fermentation was used to perform a small scale fractionation in order to check for reproducibility of the original bioactivity assay against *M. smegmatis* (shown in **Figure 32**).
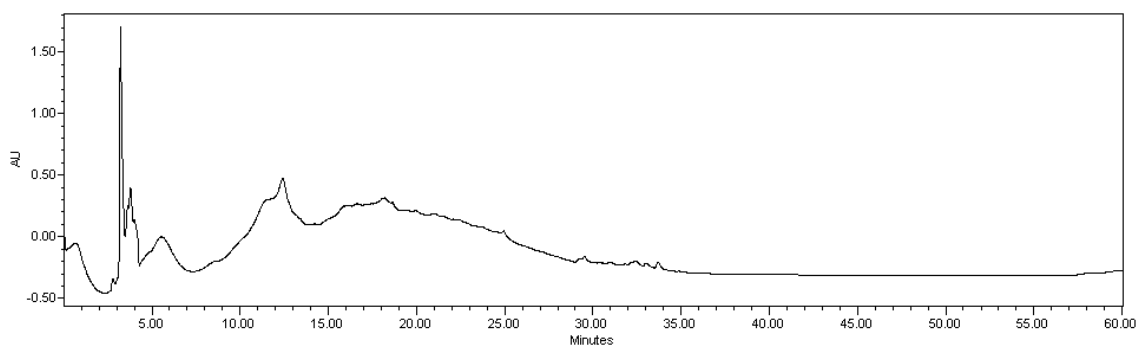
Figure 32: LC UV profile of extract 67037 on semi-prep C18 HPLC column.

When these fractions were tested in whole-cell assay against *M. smegmatis* bioactive fractions could be found, which confirmed the success of the large scale refermentation. The HPLC UV profile of the original and refermentation extract matched further confirming the reproducibility of the extract. The remaining extracts were then fractionated on a preparative scale HPLC column (Shown in **Figure 33**).
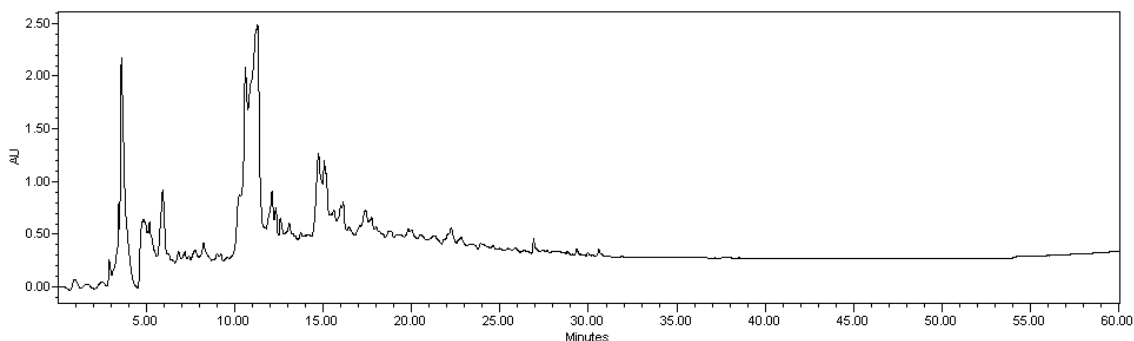
Figure 33: LC UV profile of extract 67037 on preparative scale C18 HPLC column.

Fractions were collected directly into glass test tubes at the rate of 5 ml/min/fraction. The large scale fractions were dried, reconstituted in dimethyl sulfoxide (DMSO) and retested in whole-cell bioactivity assay against *M. smegmatis* and *Mtb*. The same fractions were found to be active against both microorganisms (**Figure 34 and 35**).
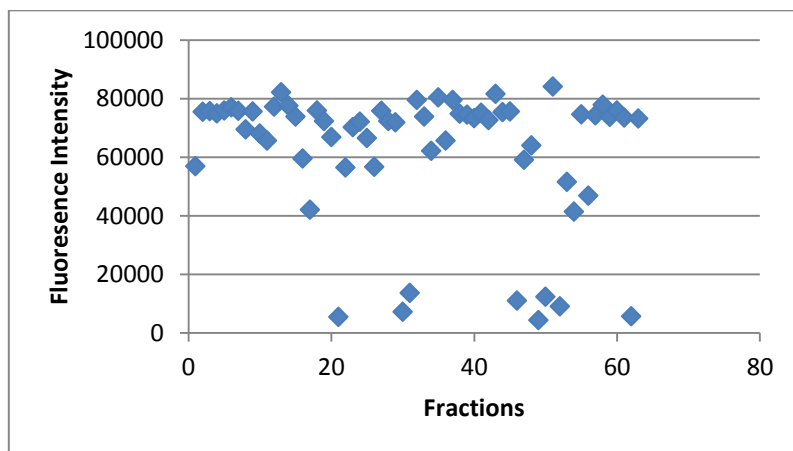
Figure 34: Whole-cell bioactivity profile of the 60 fractions of extract 67037 on *M. smegmatis* (mc^2 155) cells. Fraction numbers 30 and 31 from the preparative scale fractionation of extract 67037 was bioactive against *M. smegmatis* whole-cells. (The other fractions exhibiting bioactivity were not consistent with the bioactivity assay against *Mtb* and hence discarded.)

150
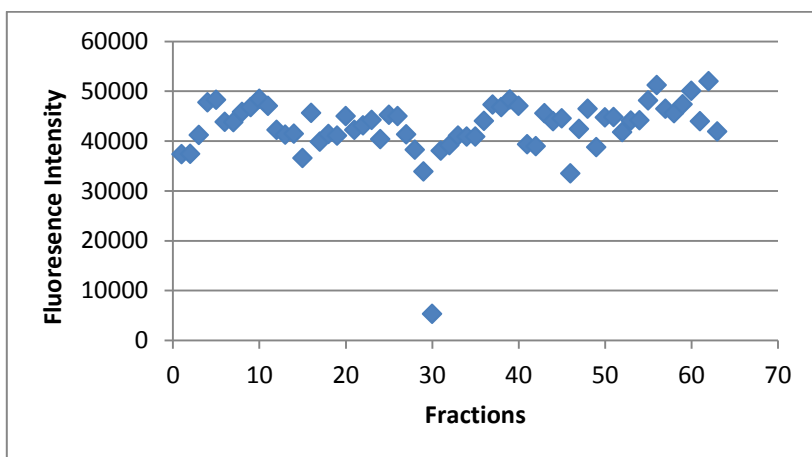
Figure 35: Whole-cell bioactivity profile of the 60 fractions of extract 67037 on *Mtb* (mc^2 7000) cells. Fraction number 30 from the preparative scale fractionation of extract 67037 was bioactive against *Mtb* whole-cells.

As this bioactivity originated from a preparative scale fractionation, there was sufficient amount of material to go through further rounds of HPLC separation to further isolate the active component of the fraction. Fraction numbers 30 and 31 of extract 67037 were bioactive against both *M. smegmatis* as well as *Mtb* cells. These two fractions were hence pooled together and refractionated on the semi-preparative HPLC column to fractionate them further (shown in **Figure 36**).
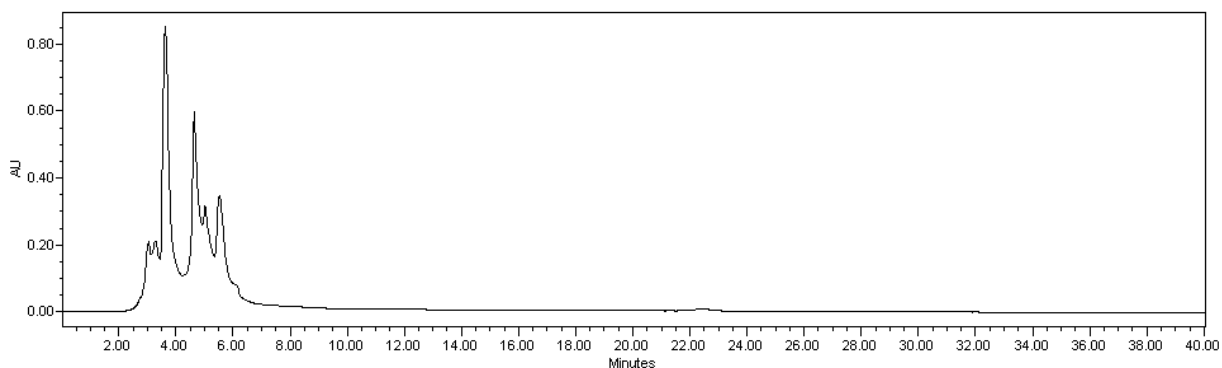
Figure 36: LC UV profile of fractions number 30 and 31 of extract 67037 on semi-prep C18 HPLC column.

The generated fractions were dried, reconstituted and retested for whole-cell bioactivity against *M. smegmatis* and *Mtb* (**Figure 37 and 38**).
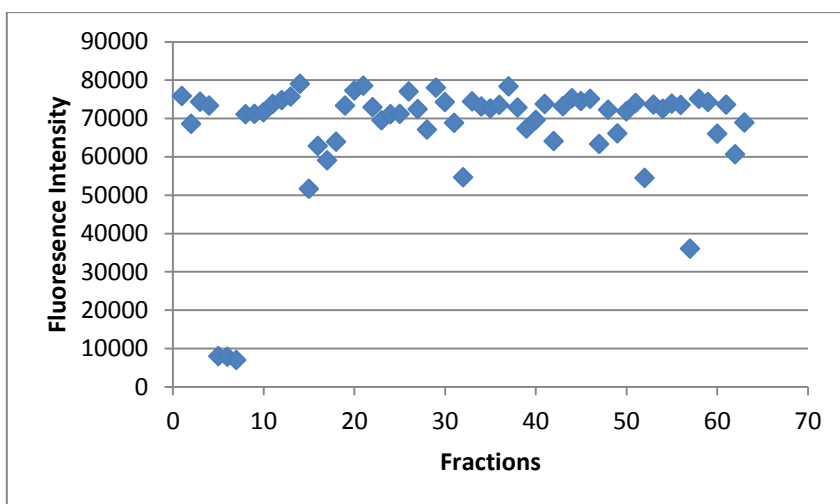
Figure 37: Whole-cell bioactivity profile of the fractionation of fractions number 30 and 31 of extract 67037 on *M. smegmatis* (mc^2 155) cells. Fractions number five, six and seven were active against *M. smegmatis* whole-cells.
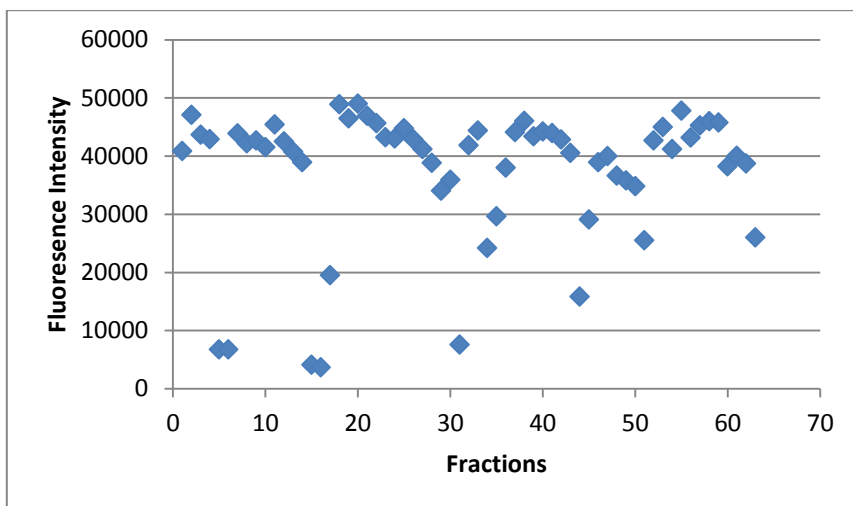
Figure 38: Whole-cell bioactivity profile of the fractionation of fractions number 30 and 31 of extract 67037 on *Mtb* (mc^2 7000) cells. Fractions number five and six were active against *Mtb* whole-cells.

In the retesting, fractions number five, six and seven were active against *M. smegmatis* whole-cells while fraction numbers five and six were active against *Mtb* whole-cells. Thus it was confirmed that the activity could be reproduced through the further rounds of purification. An aliquot of fractions number five, six and seven were analyzed on the LC / HRMS to get mass data for the possible bioactive compounds in the fractions.

As a control, a randomly picked inactive fraction number 26 from the preparative

scale fractionation of 67037 was refractionated using the same LC method as that used

for the fractions number 30 / 31 (**Figure 39**). Fraction number 26 represented a negative

control as none of the fractions generated from the fractionation would be expected to

show bioactivity against *Mtb* or *M. smegmatis* in the whole-cell assay. Indeed, when the

fractions were tested for bioactivity none of the fractions showed bioactivity against *Mtb*

or *M. smegmatis*.
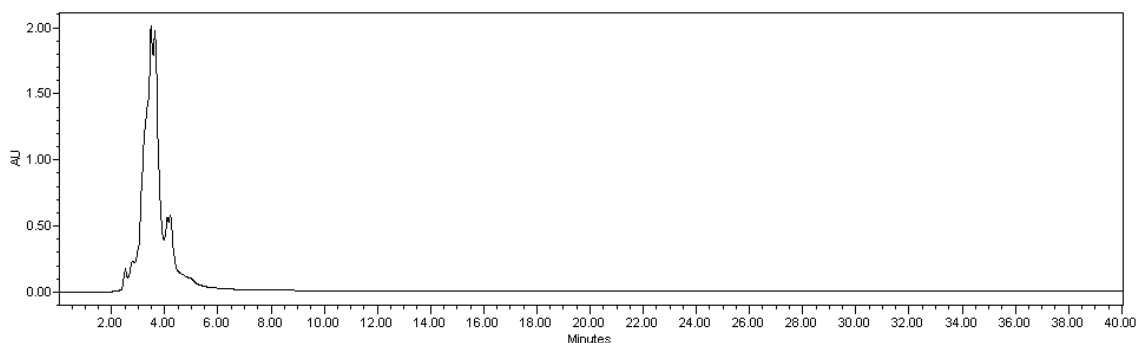
Figure 39: LC UV profile of fraction number 26 of extract 67037 on semi-prep C18
HPLC column. The UV profile of fraction number 26 is different from the UV profile of
fraction number 30 / 31 suggesting differences in the composition of the fractions.

Since fraction number five from the refractionation of fractions 30 / 31 was

active, fraction number five from the refractionation of 26 was also selected for LC /

HRMS analysis. This fraction was considered to represent a negative control. Apart from

this negative control other randomly selected inactive fractions from the fractionation of fractions 30 / 31, fractions two and 11 were also used as negative controls. The LC / HRMS data for all the selected fractions were collected using electrospray as a source of ionization in the mass spectrometer and in the positive and negative modes.

Manually sorting through the mass spectrum at every time point of the LC gradient is time consuming and introduces human error into the analysis. Metabolite Tools® suite of software package developed by Bruker Daltonics Inc. was used to simplify the data processing. The Metabolite Tools® software suite consists of two programs – Metabolite Detect® and Metabolite Predict®. The two programs were originally designed to analyze plasma, blood or serum samples and identify the administered drug molecules and metabolites thereof. The operating parameters of the Metabolite Tools® software package was modified to apply to the bioactive and inactive HPLC fractions of the extract 67037 and to identify MS peaks of interest in the bioactive fractions. The Metabolite Predict® part of the software is designed to predict metabolites of a known drug compound and was not considered useful in the context of the present research. Hence, the use of Metabolite Predict will not be discussed in this dissertation. Metabolite Detect® was a useful component of the software package as it can statistically compare the mass spectrum of a test sample with that of a reference (control) sample and subtract all the MS peaks occurring in the reference sample from the test sample. When applied to the HPLC fractions of 67037 with the appropriate operating parameters, this software could be used to quickly narrow down to the MS peaks of interest that existed in the bioactive fractions and did not exist in the inactive fractions. The LC / HRMS data

155

obtained from the bioactive fractions five and six from the refractionation of fractions 30 / 31 was used as the test sample. The LC / HRMS data obtained from the inactive fractions five and six from the refractionation of fraction 26 as well as the HRMS data obtained from the inactive fractions two and 11 from the refractionation of fractions 30 / 31 were used as reference samples. Upon subtracting all the MS peaks in the reference samples from the test samples, the MS data could be narrowed down to MS peaks of interest that could be followed through further rounds of purification.

The remaining amount of the bioactive fraction number five was concentrated and refractionated on HPLC to further isolate the bioactive component in the original fraction **(Figure 40)**.
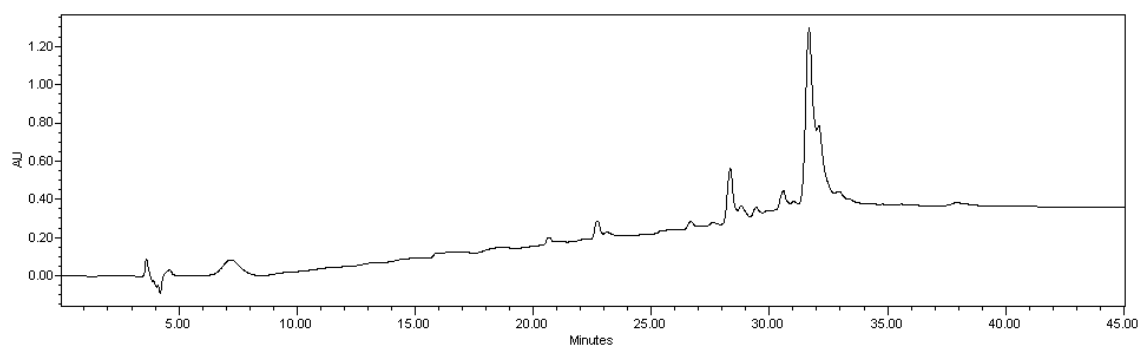
Figure 40: LC UV profile of fraction number five from the refractionation of fractions 30 and 31 of extract 67037 on semi-prep C18 HPLC column.

The newly generated fractions were again tested for their whole-cell bioactivity against *M. smegmatis* and *Mtb* **(Figure 41 and 42)**.

Figure 41: Whole-cell bioactivity profile of the fractionation of fraction five of extract 67037 on *M. smegmatis* (mc^2 155) cells. Fractions 32 and 33 exhibited whole-cell bioactivity against *M. smegmatis* cells. The fractions were tested at 0.05 X concentration.
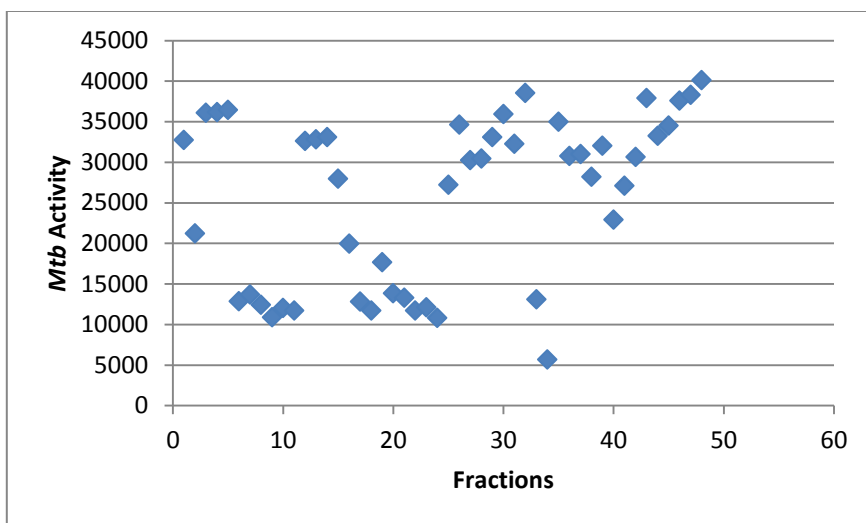
Figure 42: Whole-cell bioactivity profile of the fractionation of fraction five of extract 67037 on *Mtb* (mc^2 7000) cells. Fractions 32 and 33 exhibited whole-cell bioactivity against *Mtb* cells. The fractions were tested at 0.075 X concentration.

Fractions number 32 and 33 from the new set of fractions displayed consistent whole-cell bioactivity against both *M. smegmatis* and *Mtb* cells **(Figure 41 and 42)**. LC / HRMS analysis was performed for all the bioactive fractions through the successive rounds of fractionations starting from the crude extract to the semi purified fraction observed in **Figure 40**. The LC / HRMS data was analyzed to follow the MS peaks of interest that consistently appeared in the bioactive fractions (**Figures 43, 44 and 45**).
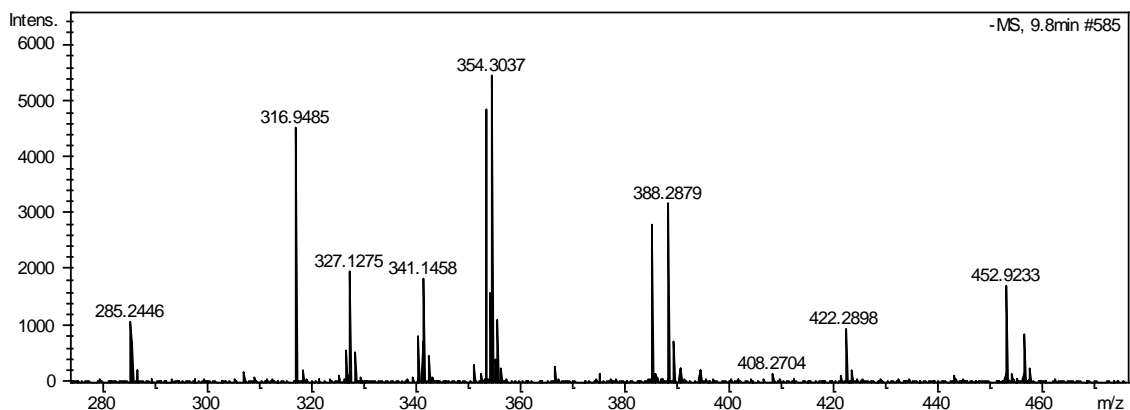
Figure 43: LC / HRMS analysis of fraction number 30 from the first round of fractionation of the crude extract. The MS profile at 9.8 minute time point in the ESI negative mode is shown.
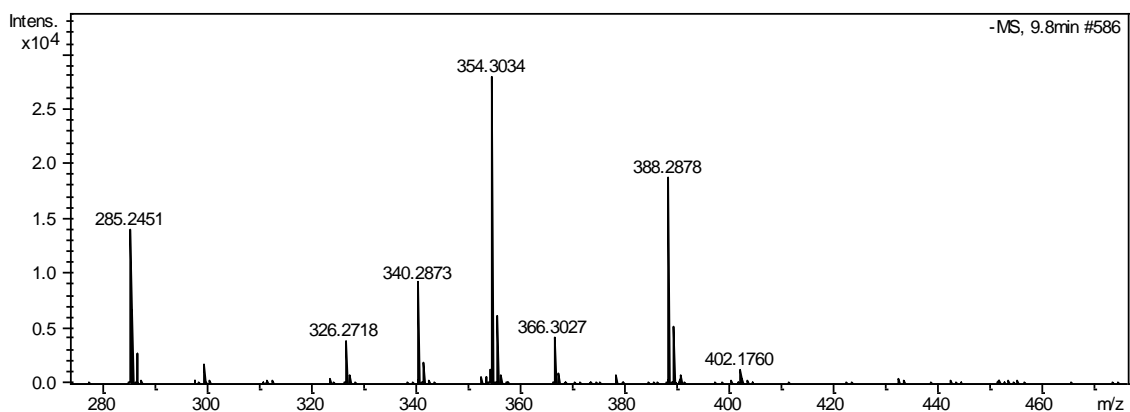
Figure 44: LC / HRMS analysis of fraction number five from the refractionation of fractions number 30 / 31. The MS profile at 9.8 minute time point in the ESI negative mode is shown.
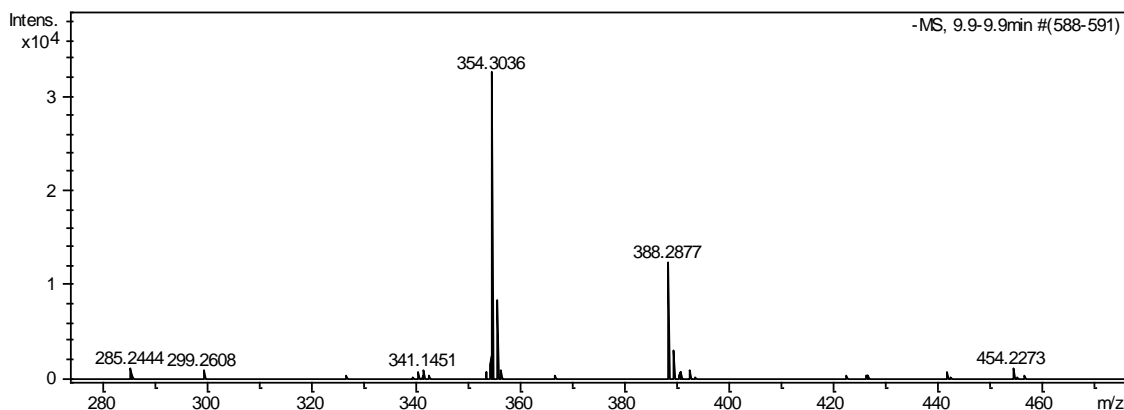
Figure 45: LC / HRMS analysis of fraction number 32 from the refractionation of fractions number five / six. The MS profile at 9.9 minute time point in the ESI negative mode is shown.

The LC / HRMS analysis of bioactive fractions through the rounds of HPLC fractionation showed that the MS peak at (m/z) 354.30 and 388.28 in the ESI negative mode appeared consistently through the rounds of purification suggesting that one of these two peaks might be the source of bioactivity against *Mtb* mc^{2}7000 and *M. smegmatis* mc^{2}155 whole cells (**Figures 43, 44 and 45**). Smartformula calculation for the MS peak at (m/z) 354.30 gave a molecular formula prediction of $C_{21}H_{40}NO_3$ [M-H]$^-$ with err [ppm] = -6.0 and mSigma [ppm] = 12.2 and for the MS peak at (m/z) 388.28 gave a molecular formula prediction of $C_{24}H_{38}NO_3$ [M-H]$^-$ with err [ppm] = -4.7 and mSigma [ppm] = 10.7.

At this stage of purification there was less than 1 mg of the fraction remaining, which did not allow for the NMR spectroscopic analysis of the fraction to elucidate the structure of the molecules. Hence, a large scale fermentation corresponding to 1 litre

160

fermentation broth of the bacteria producing the bioactivity was conducted. The

extraction of the fermentation broth yielded a total of 0.6747 gm dry extract. The extract

was dissolved in a minimum amount (~1 mL) of methanol and the clear solution was

injected onto the preparative scale C18 HPLC column (**Figure 46**).
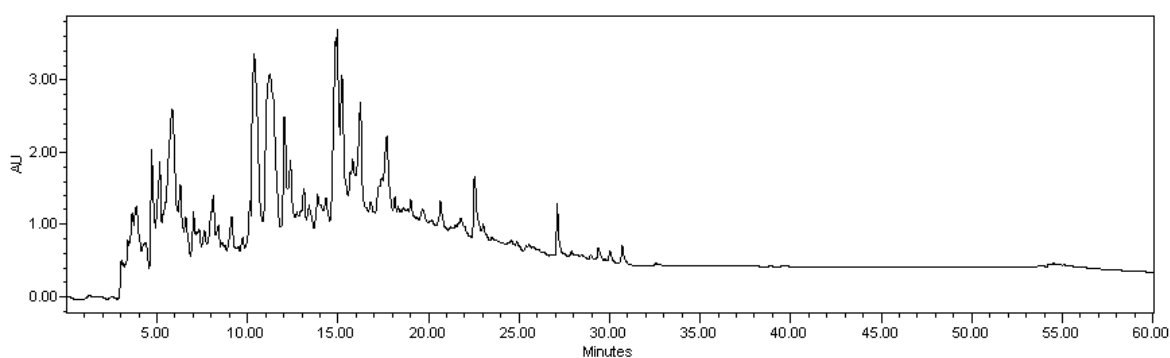
Figure 46: LC UV profile of the preparative scale fractionation of extract 67037. The same gradient conditions were used as for the previous preparative scale fractionation of extract 67037.

Fractions were collected at the rate of 5 ml/minute/fraction. 100 uL of each

fraction were aliquoted into the wells of a 96 well polypropylene plate and evaporated to

dryness. The wells were then reconstituted in 20 uL DMSO. 1 uL from each well was

tested against *M. smegmatis* whole cells in a 384 well polypropylene plate containing a
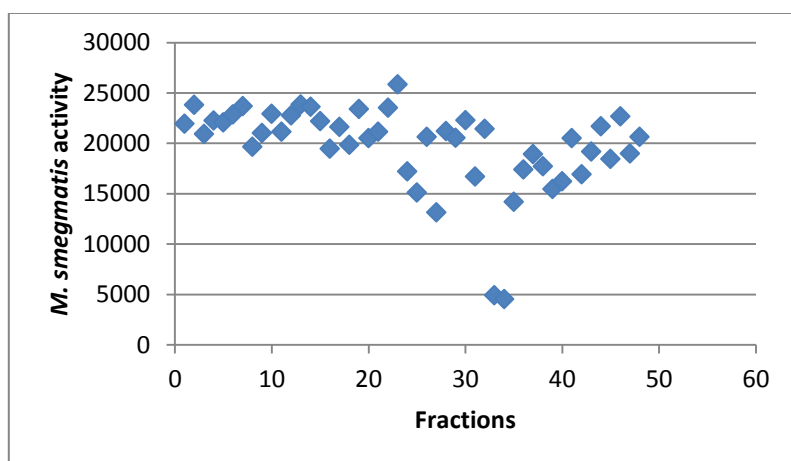
total of 49 uL of the cell culture in each well (**Figure 47**).
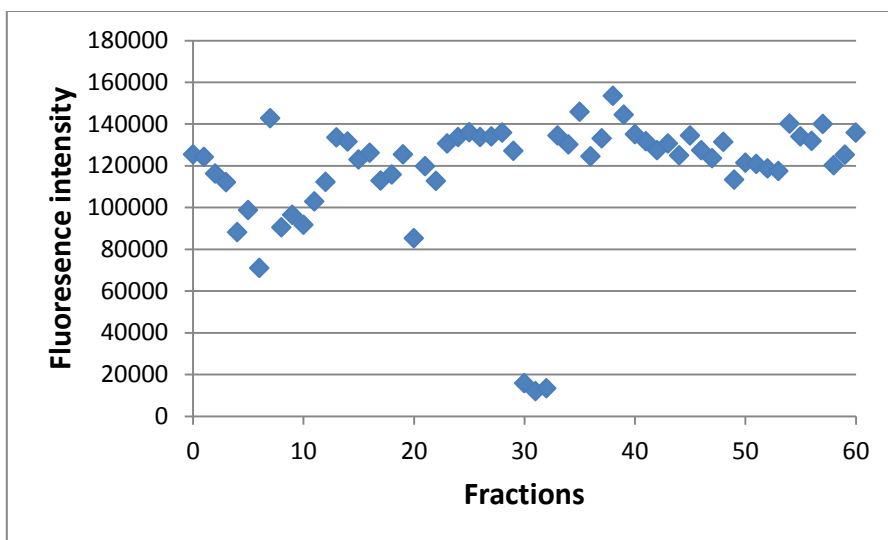
Figure 47: Whole-cell bioactivity profile of the fractionation of crude extract 67037 on *M. smegmatis* (mc^2 155) cells. Fractions 30, 31 and 32 exhibited whole-cell bioactivity against the *M. smegmatis* cells.

The bioactive fractions identified from the whole cell bioactivity assay against *M. smegmatis* matched the bioactivity profile of the fractions generated for the previous batch of the extract 67037. Thus the bioactivity was reproducible. The bioactive fractions were pooled together, dried and reconstituted in 200 uL methanol and injected into a semi-preparative C18 HPLC column **(Figure 48)**.
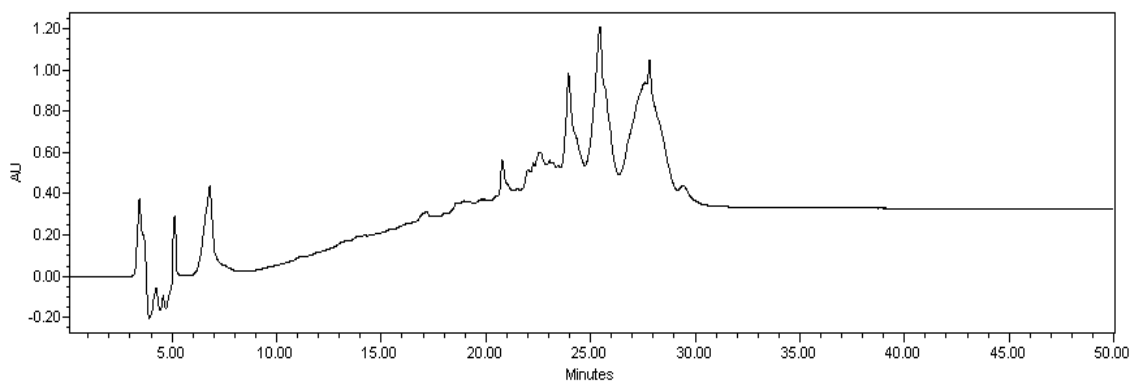
Figure 48: LC UV profile for the fractionation of fraction number 30 to 32 of extract 67037 on semi-prep C18 HPLC column. The UV profile at 220 nm is shown.

Fractions were collected at the rate of 1 ml/minute/fraction. 10 uL of each fraction were aliquoted into the wells of a 96 well polypropylene plate and evaporated to dryness. The wells were then reconstituted in 5 uL DMSO. 2 uL from each well was tested against *M. smegmatis* whole cells in a 384 well polypropylene plate containing a total of 49 uL of the cell culture in each well **(Figure 49)**.
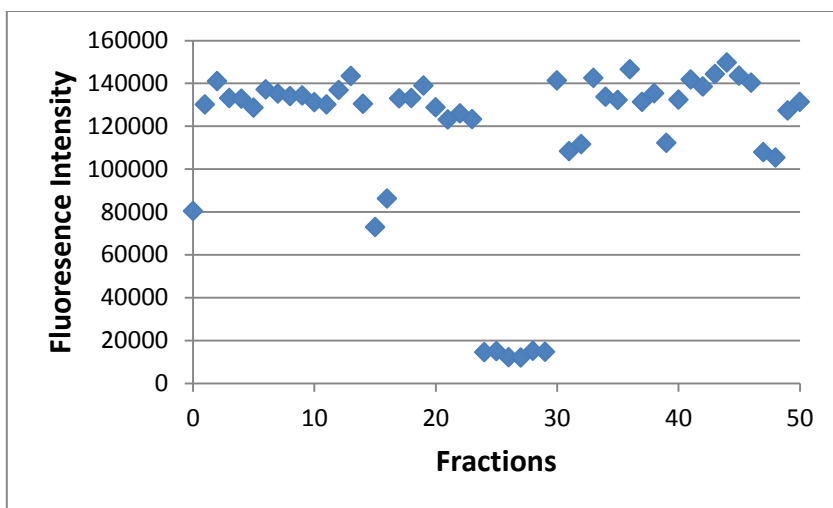
Figure 49: Whole-cell bioactivity profile of the fractionation of fractions 30, 31 and 32 from the crude extract 67037 on *M. smegmatis* (mc^2 155) cells. Fractions 24 through 28 exhibited whole-cell bioactivity against the *M. smegmatis* cells.

Bioactivity was obtained against *M. smegmatis* mc^{2}155 cells for fractions number 24 through 28 **(Figure 49)**. LC / HRMS analysis was performed for the bioactive fractions in the ESI negative mode **(Figure 50)**.
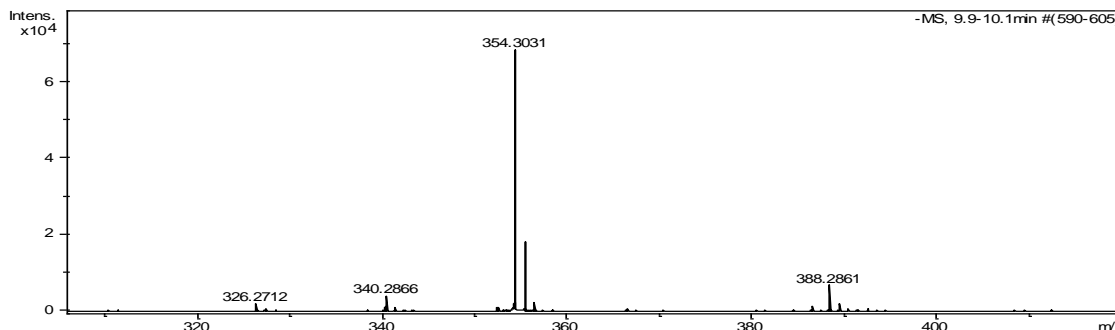
Figure 50: LC / HRMS analysis of fraction number 27 from the refractionation of fractions number 30 to 32 from extract 67037. The MS profile at 10.0 minute time point in the ESI negative mode is shown.

LC / HRMS analysis revealed that the same MS peak at (m/z) 354.30 and 388.28 appeared in the bioactive fractions **(Figure 50)**. A search in the dictionary of natural products (DNP) using the HRMS at (m/z) 354.30 [M-H]⁻ and 388.28 [M-H]⁻ and the generated molecular formula prediction revealed a single match for the 354.30 [M-H]⁻ peak and showed no matches for the 388.28 peak. The natural product that matched the MS peak at 354.30 [M-H]⁻ was *N*-Methyl-*N*-(1-oxooctadecyl)glycine with an accurate mass of 355.308644 and molecular formula of $C_{21}H_{41}NO_3$. In order to investigate the molecular structure of the natural product MS / MS fragmentation data was obtained from the MS molecular ion peak at (m/z) 354.30 [M-H]⁻.
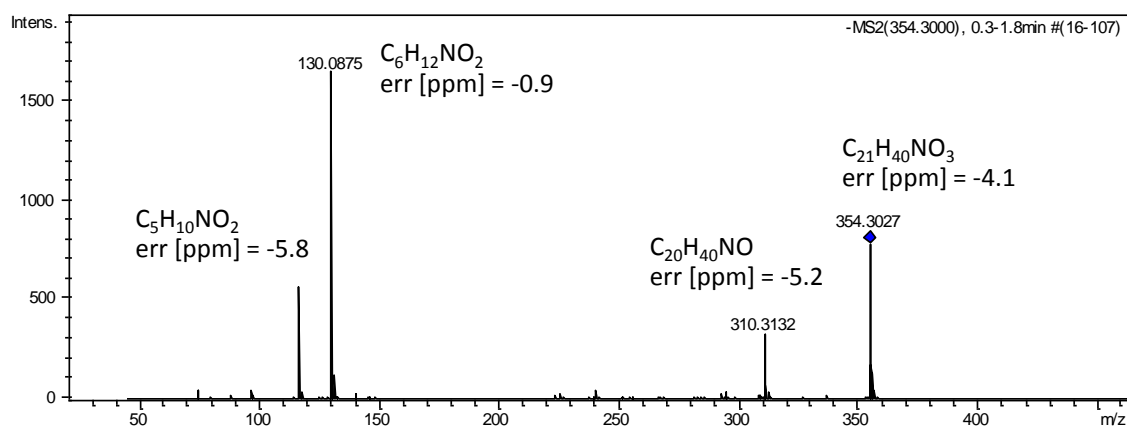
Figure 51: Fragmentation pattern observed for MS / MS of (m/z) 354.30 [M-H]⁻ in the ESI negative mode. The generated molecular formula for each fragment peak is shown.

The fragmentation pattern showed distinct peaks at (m/z) 116.07, (m/z) 130.08 and (m/z) 310.31. Molecular formulas were predicted for the fragment peaks using the SmartFormula® calculator and are shown in the figure 51. While a similar fragmentation pattern could be obtained for *N*-Methyl-*N*-(1-oxooctadecyl)glycine, the predicted molecular formulas of the fragments did not match with the molecular formulas for the predicted fragments of *N*-Methyl-*N*-(1-oxooctadecyl)glycine. Thus based on the HRMS and the predicted molecular formula of the observed fragments, a chemical structure was drawn for the molecular ion peak at (m/z) 354.30 [M-H]⁻ as follows (**Figure 52**);
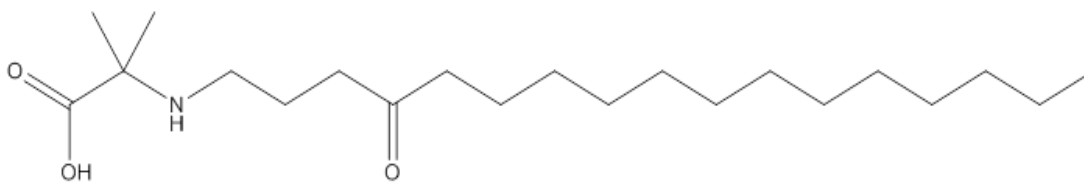
Figure 52: Predicted chemical structure for the MS peak at (m/z) 355.30 [M]. 2-methyl-2-[(4-oxoheptadecanoyl)amino]propanoic acid.

Chemical structures could be assigned to the observed fragment peaks in the tandem MS as follows (**Figure 53**);
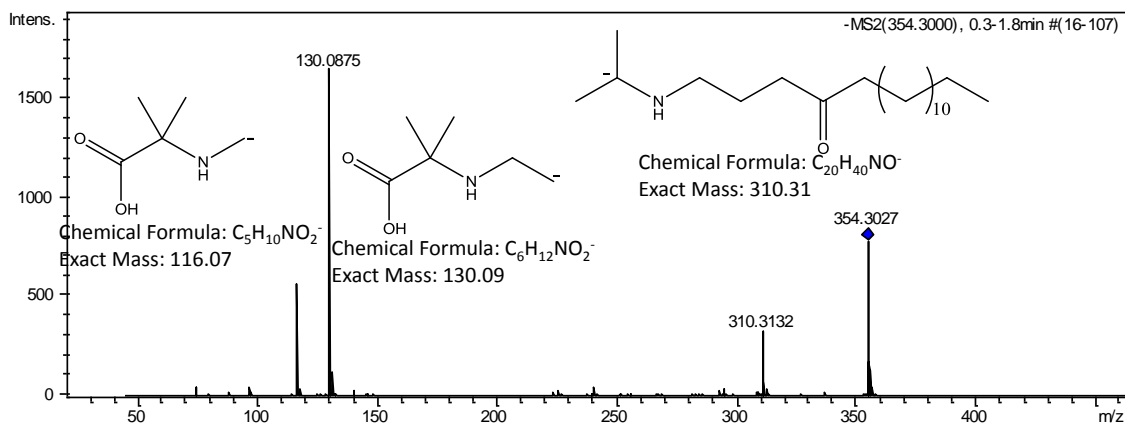
Figure 53: Fragmentation pattern observed for MS / MS of (m/z) 354.30 [M-H]⁻ in the ESI negative mode. The structure assignment for each fragment peak is shown.

The bioactive fraction number 27 was dried and reconstituted with ~ 400 µL (CD$_3$)$_2$CO (deuterated acetone). NMR spectroscopic data was obtained for the sample to gain insights into the structural characterization of the natural product (**Figure 54 and 55**).

The NMR data correlated with part of the predicted chemical structure. The broad peak observed at δ 1.3 ppm in the ^{1}H NMR and the cluster of peaks in the δ 10 – 40 ppm range in the ^{13}C NMR spectra correlate with the presence of a long aliphatic carbon chain.

Conclusion

Vicuron Extract Library

A collection of microbial extracts enriched for activity against gram positive bacteria were selected and taken through a process of fractionation, testing the fractions for whole-cell bioactivity followed by subsequent preliminary characterization of the active fractions. Nine extracts out of the 96 were chosen for further investigation based on the bioactivity data. The selected extracts had the highest potential for producing a natural product with novel properties. Preliminary characterization of the fractions producing activity showed great promise to this end.
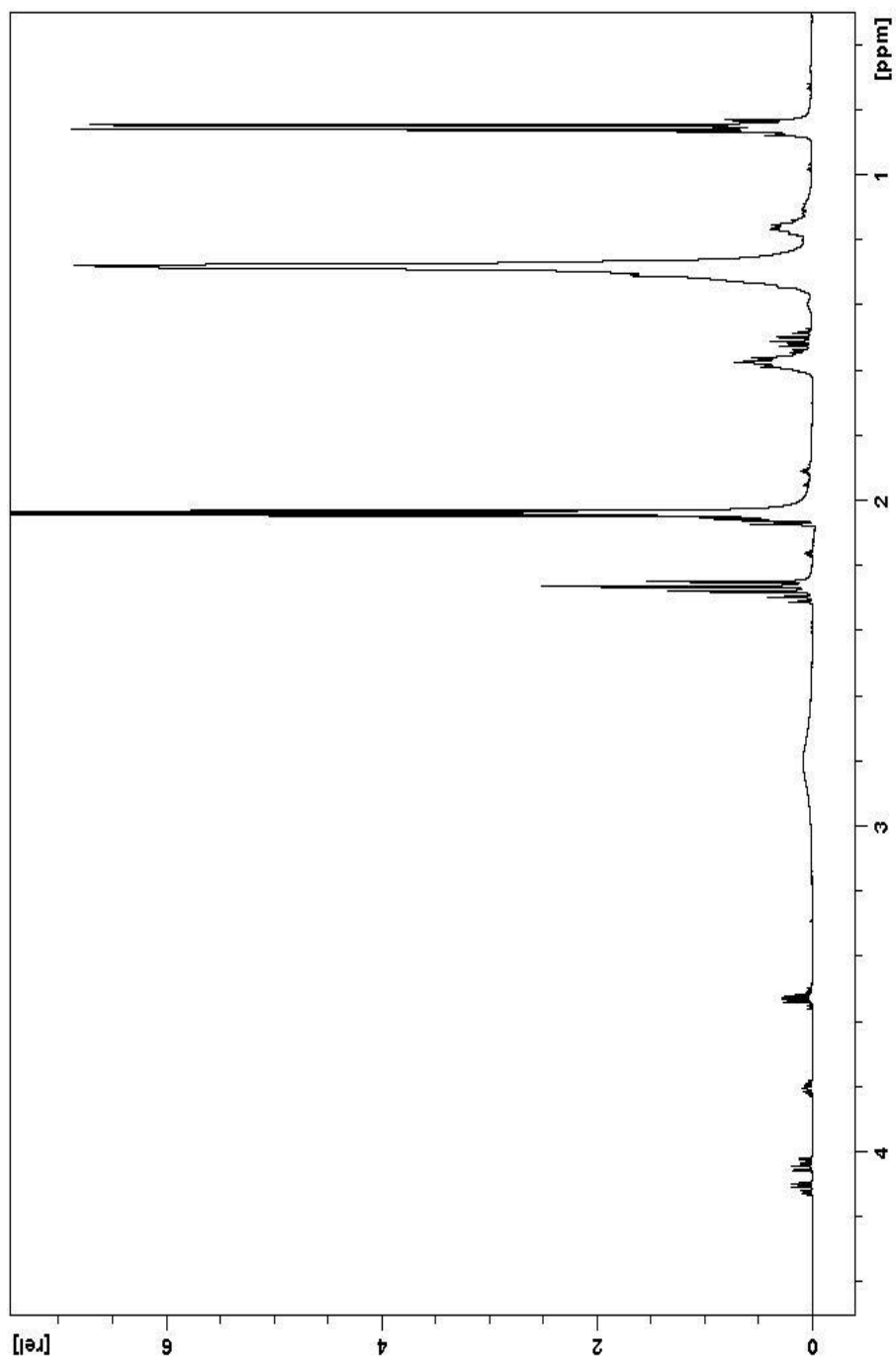
Figure 54: ^{1}H NMR data of the bioactive fraction number 27.
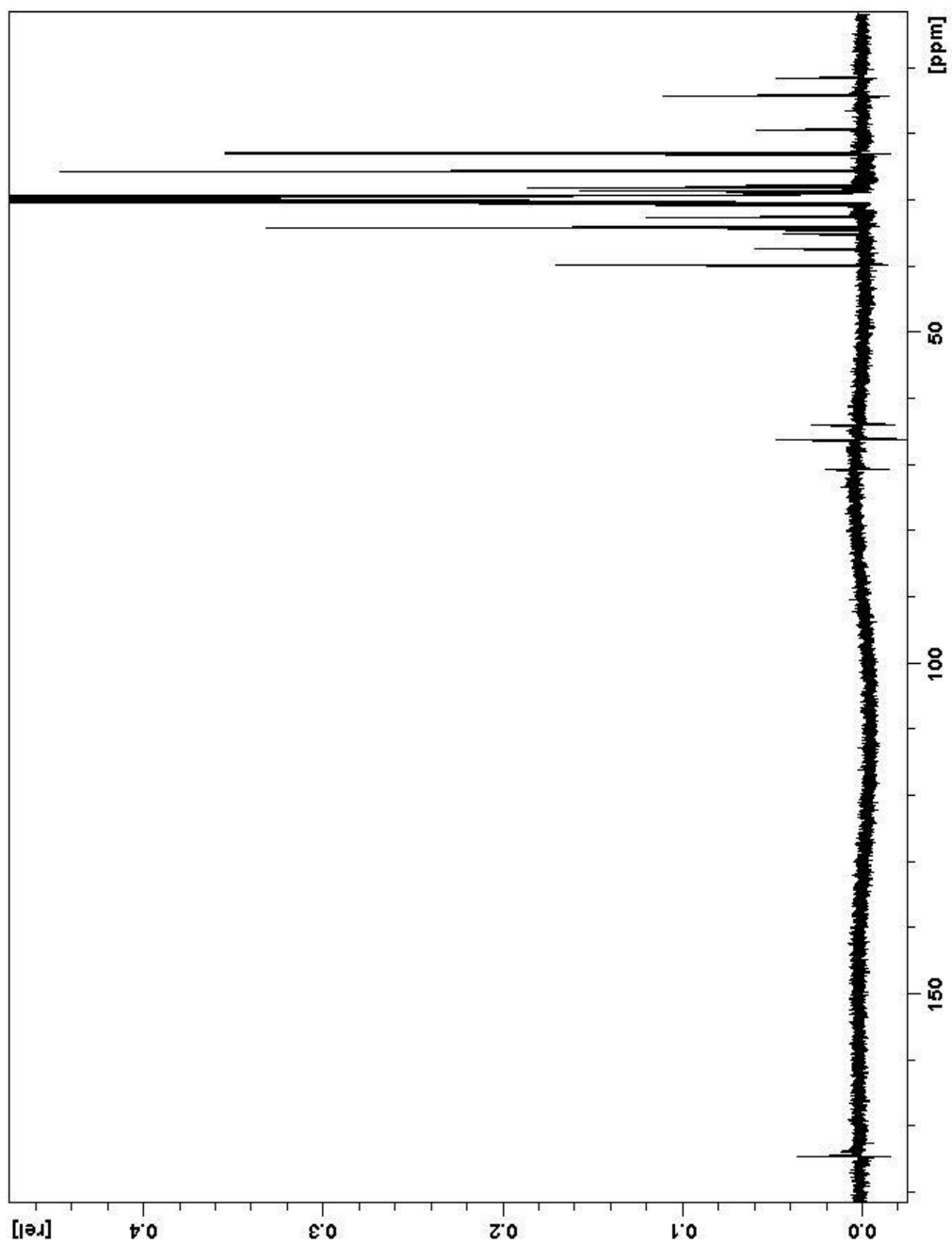
Figure 55: ^{13}C NMR data of the bioactive fraction number 27.

At the time of writing this dissertation the refermentation broths for the nine prioritized extracts were yet to be received. One of the biggest challenges in natural product drug discovery is reproducing the fermentation broth in an identical fashion so as to reproduce the biological activity present in the original fermentation. Many different parameter like the media composition, nutrient sources, temperature, time of incubation, etc., need to be strictly controlled in order to allow the reproduction of the bioactive component by the microorganism in sufficient amounts to be detected in an assay. Minor variations in the culture conditions can cause the microbe to stop producing the active secondary metabolite or to produce an inactive analog of the same. A way to determine the reproducibility of the fermentation broth is to perform a preliminary HPLC fractionation on an aliquote of the extract and compare the HPLC profile to the same fractionation carried out on the original extract. If the LC UV profiles of the two fermentation products can be matched then it can be concluded with confidence that the fermentation was reproducible. Other experimental challenges include appropriate sample preparation of the extracts for fractionation, optimization of the LC gradient to achieve the best possible separation profile and prompt handling of the LC fractions to prepare stock solutions for the whole-cell bioactivity assay.

The process of screening for whole-cell bioactivity was similar to what would be typically performed in a high-throughput fashion in a natural product drug discovery company. While a large natural products company would process in the order of 100,000 extracts simultaneously in a truly high-throughput fashion, the present study involved the fractionation of 96 prioritized extracts, which can be classified as a low-throughput

screening. However the methods and the number of steps involved in the process remain the same. The results obtained from this low-throughput screening effort are significant and suggest that even a low-throughput screening effort when performed with insight can yield positive results in a short amount of time.

Sarawak Biodiversity Library of Extracts

Fractions of extract 67037 had anti-mycobacterial activity against both *M. smegmatis* as well as *Mtb*. The further fractionation of the active fractions still contained the original activity. This indicated that the bioactive component of the extract was stable through the multiple steps involved in the HPLC purification. When tested against human cancer cell lines the crude extract showed no cytotoxicity indicating that the bioactivity was specific towards mycobacteria. Through the process of bioactivity guided fractionation the crude extract was purified and the potential source of bioactivity was isolated. The natural product likely to be active against *Mtb* and *M. smegmatis* was identified through HRMS and tandem MS as 2-methyl-2-[(4-oxoheptadecanoyl)amino]propanoic acid. NMR spectroscopic data obtained for the bioactive fraction was used to further validate the chemical structure of the natural product. The bioactive compound consists of a 17 carbon long aliphatic chain and a carboxylic acid head group. Based on the fatty acid like structure of the compound it may be speculated that the natural product can be easily taken up by the thick cell wall of the mycobacteria. In order to investigate the bioactivity of known naturally occurring fatty acids on *Mtb* and *M. smegmatis*, eight different fatty acids were selected and tested

for bioactivity against both mycobacteria in identically performed whole-cell assays at concentrations ranging from 20 µM to 0.15 µM. Surprisingly, the simple fatty acids showed whole cell inhibitory activity against the mycobacteria. Palmitic acid and stearic acid inhibited *Mtb* whole-cells with an MIC of ~ 5 µM and oleic acid and linoleic acid inhibited *M. smegmatis* whole-cells with an MIC of < 5 µM (**Figure 56 and 57**).
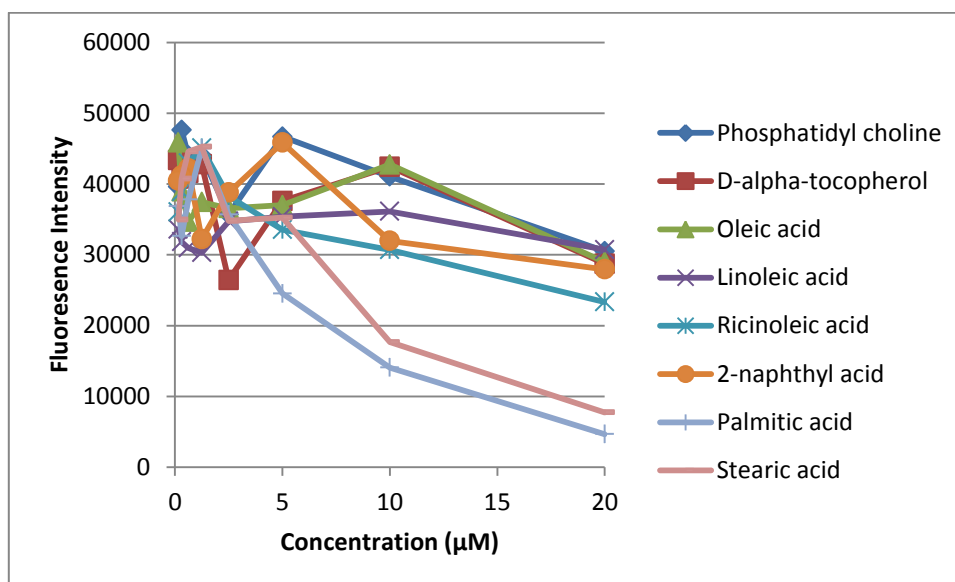
Figure 56: Whole-cell bioactivity assay of naturally occurring fatty acids against *Mtb* at 20 µM to 0.15 µM concentration. Palmitic acid and stearic acid inhibited whole-cell growth of *Mtb* with an MIC of ~ 5 µM.
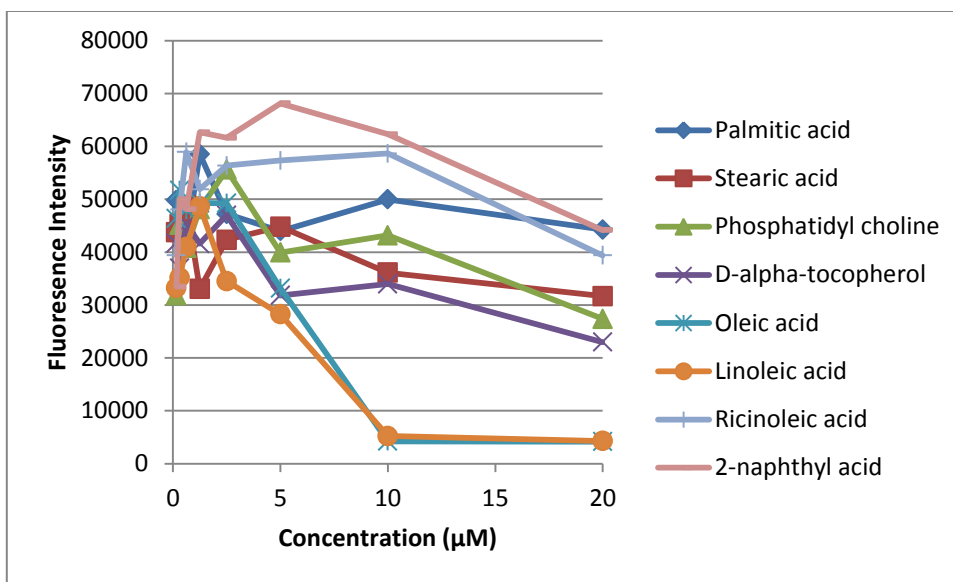
Figure 57: Whole-cell bioactivity assay of naturally occurring fatty acids against *M. smegmatis* at 20 µM to 0.15 µM concentration. Oleic acid and linoleic acid inhibited whole-cell growth of *M. smegmatis* with an MIC of < 5 µM.

Palmitic acid and stearic acid are long chain fatty acids with no unsaturations in the aliphatic carbon chain while oleic acid and linoleic acid are long chain fatty acids with one and two unsaturations in the aliphatic carbon chain respectively. The bioactivity of these simple fatty acids against the mycobacteria validates the conclusion that an atypical fatty acid like the one identified through the high-throughput screening process can also follow the same mechanism of cell uptake and inhibition. While fatty acids have been used as inhibitors of cholesterol and fatty acid biosynthesis [199] and in cancer therapy [200], the current findings may open new avenues for using synthetic or natural products containing a fatty acid like core structure as inhibitors of *Mtb*.

The process of screening natural product extracts for antibacterial activity is one that began with the discovery of Streptomycin back in 1944. Streptomycin was isolated from an actinomycete, which to this day continues to be the most prolific producer of antibacterial agents. Confirmation of 2-methyl-2-[(4-oxoheptadecanoyl)amino]propanoic acid as the bioactive agent in extract 67037, which is sourced from *Streptomyces chromofuscus*, will also reaffirm faith that 70 years after the discovery of the first naturally occurring antibiotic, natural products research still holds potential for discovering novel antibiotics against the dreaded disease of TB.

Future Work

An important aspect of natural product research is the investigation of the biological target of the molecule. Hence experiments will need to be performed to narrow down to the most probable enzyme or the biological pathway that the molecule targets. This can be achieved by several methods. One way is to isolate resistant mutants of the pathogen in the presence of the bioactive molecule. DNA of the mutant microorganism can then be extracted. A comparison of the whole genome sequence of the mutant and the wild type DNA will give clues to the location of the presence of the mutation. The genes carrying the mutations may be the ones conferring resistance to the pathogenic organism. Recombinant enzyme of the unmutated gene can be added to the microbial fermentation growing in the presence of the bioactive molecule. An increase in the dosing of the antibiotic needed to kill the bacterial culture would validate the gene or the translated enzyme as the true drug target.

Natural Product Inhibitors of the *Mtb* Malate Synthase Enzyme

Once a whole cell active compound is identified from a natural product library, the next objective is to find the pathway or the enzyme in the pathogen that the natural product targets. This requires growing cells that are resistant to the small molecule and sequencing the genome of the resistant strain. A comparison of the whole genome sequence of the resistant strain with the wild-type strain is then performed to find the mutation sites. This process enables identification of the target for the active compound. In order to speed up the process of lead identification, an alternate approach was used where *Mtb* malate synthase was selected as the target and the natural products library was screened in an *in vitro* enzyme assay. Malate synthase was selected as the target because it functions in the glyoxylate shunt pathway which has been shown to be essential for survival of *Mtb* inside the macrophages[201] and also due to the absence of a malate synthase homolog in the human.

An important factor contributing to the world wide spread of TB is the persistent nature of the *Mtb* bacillus. *Mtb* can survive inside the macrophages under harsh conditions by metabolizing inadequate nutrient sources such as fatty acids [202]. For example, when glucose is not available to *Mtb* as a carbon source, it can divert the acetyl CoA from β-oxidation of fatty acids into a glyoxylate shunt pathway [203,204]. The shunt pathway uses isocitrate lyase as a branch point in the tricarboxylic acid (TCA) cycle to generate glyoxylate and succinate. The glyoxylate is converted by malate synthase (*GlcB*) into malate using acetyl CoA as a substrate. As malate synthase is implicated as a

virulence or persistence factor in different organisms [205] [206] [207,208] and as it is absent in mammals it is an attractive target for drug discovery.

The vicuron library of natural compound extracts discussed in the previous sections was used as a screening library in an *Mtb* malate synthase enzyme assay targeted screening.

Methods

High-Throughput **in Vitro** *Assay*

The enzyme master mix constituted the following: 1800 µL of 10 X enzyme buffer (200 mM Tris pH 7.5, 50 mM $MgCl_2$, 0.8 mM EDTA), 6896 µL Milli-Q water and 4 µL of 2 mg / ml malate synthase enzyme.

74 µL of the master mix was added to each well of a 384-well microtitre plate, 1 µL of the natural products extract was added. (All solutions were added to the microtitre plate using the liquid handling robotic instrument, CyBio Vario®.) The mixture was incubated for 20 minutes. At the end of the 20 minutes, 5 µL of acetyl-CoA solution (12.5 mM) was added right before reading the plate on a plate reader.

The enzymatic reaction was started by adding a solution of 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB) / glyoxylate mixture. The DTNB / glyoxylate solution was prepared by mixing 5 mM DTNB / 6 mM glyoxylate in Milli-Q water.

The liquid handling robot was used to dispense 20 µL of DTNB / glyoxylate solution to complete the reaction. Plate reader was set at 412 nm wavelength. The reaction was monitored for 25 minutes. A zero time point reading was measured before

the addition of the DTNB / glyoxylate mixture. Another reading was taken, 20 minutes after addition of the DTNB / glyoxylate mixture. This was the end point reading.

Results and Discussion

A 640 extracts *Mtb* whole cell active hit plate generated from the 140,000 Vicuron microbial extracts library was used in the *in vitro* screening against *Mtb* malate synthase enzyme. While the complete set of natural products extracts could have been used in a high-throughput assay, it was reasoned that a subset of the 140,000 extracts which had already been shown to have anti-mycobacterial activity against whole-cell *Mtb* had a greater chance of containing a natural product inhibitor of the *Mtb* malate synthase. The inhibition assay was performed in a high throughput fashion using the Cy-Bio Vario® liquid handling robot. A dry run was first conducted using a compound, JSF 1038 as a positive inhibitory control and no inhibitors as negative controls to ensure that the enzyme assay was functioning as expected. To calculate a hit, the readings at the zero minute time point were subtracted from the 20 minute time point reading. By comparing with the positive controls, a value of below 0.5 Au were considered as hits. Seven of the extracts showed inhibitory activity in the malate synthase enzyme assay. One of these seven extracts was also part of the 96 Vicuron extracts tested for their whole cell anti mycobacterial activity. It was in position E6 of the Vicuron plate. The remaining 633 extracts that did not have any inhibitory activity against *Mtb* malate synthase but originally showed whole-cell inhibition against *Mtb* were reasoned to be targeting other enzymes and pathways in *Mtb*. The seven extracts were fractionated on a

HPLC using a C18 reverse phase column. 40 fractions were generated for each of the seven extracts. The 40 fractions were tested again in the same *in vitro* assay against *Mtb* malate synthase enzyme using the same robotic setup as the first round of screenings. Fractions from four out of the seven extracts retested positive. The active fractions from the above four extracts were then retested in a dilution series ranging from 2 X to 0.02 X relative to their concentrations in the previous activity assay, again using the same robotic setup as the first round of screenings. Three extracts out of the four retested in the dilution series. The previously mentioned extract number E6 from the Vicuron plate was one of these three active extracts that retested in the dilution series. Thus the subset of 640 extracts was narrowed down to the three extracts that were likely targeting *Mtb* malate synthase enzyme. These three extracts and their active fractions were labeled 597-B5 (fraction number 32 and 38), 1146-F11 (fraction number 38) and 734-E2 (fraction number 34 and 35) (**Figure 58 (1-5)**).
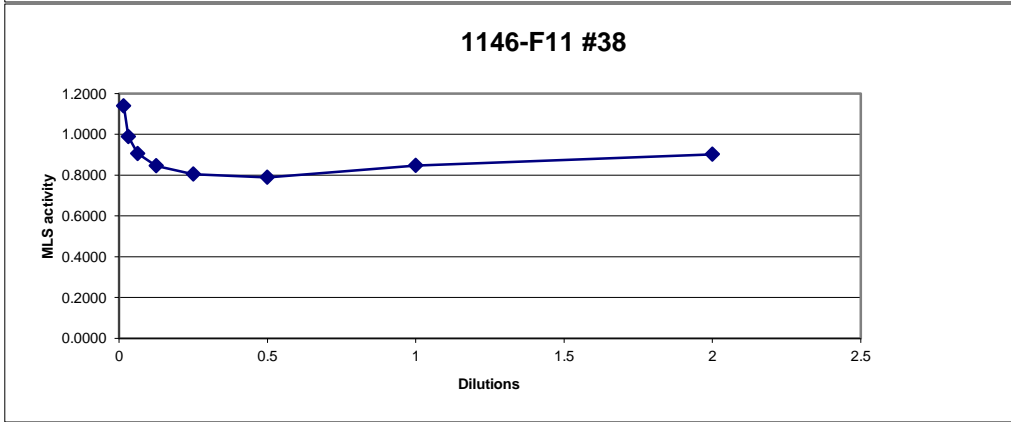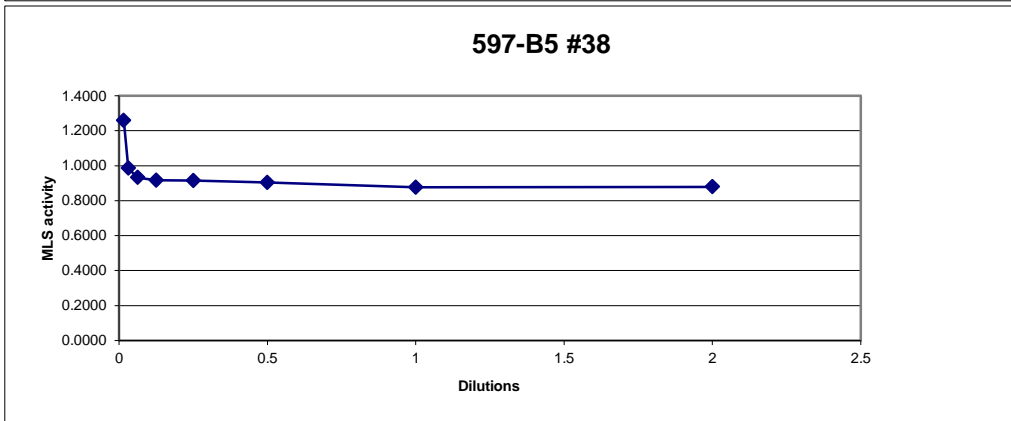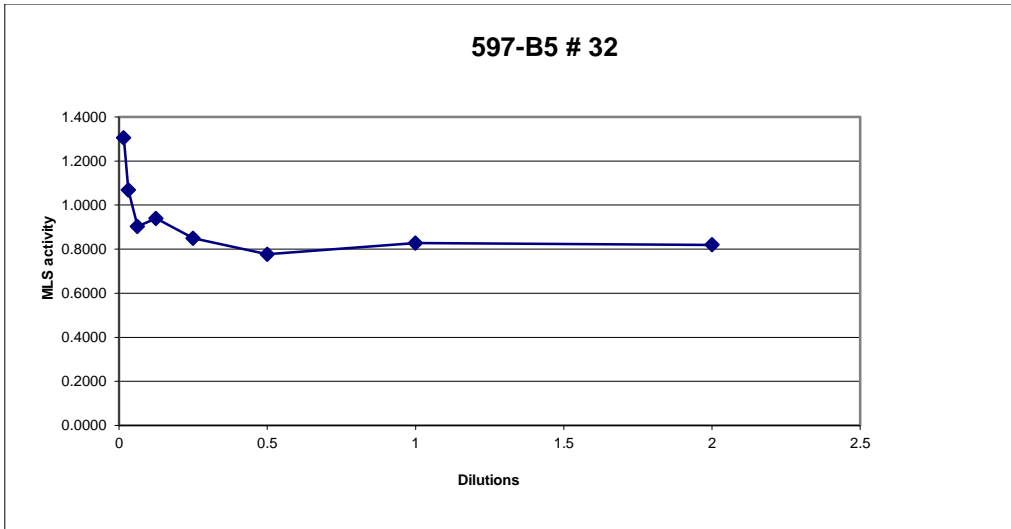
Figure 58 (1-5): *In vitro* enzyme activity assay of the extract 597-B5 (fractions 32 and 38), 1146-F11 (fraction 38) and 734-E2 (fractions 34 and 35) at dilution series of 0.2 X to 0.02 X against *Mtb* malate synthase enzyme.
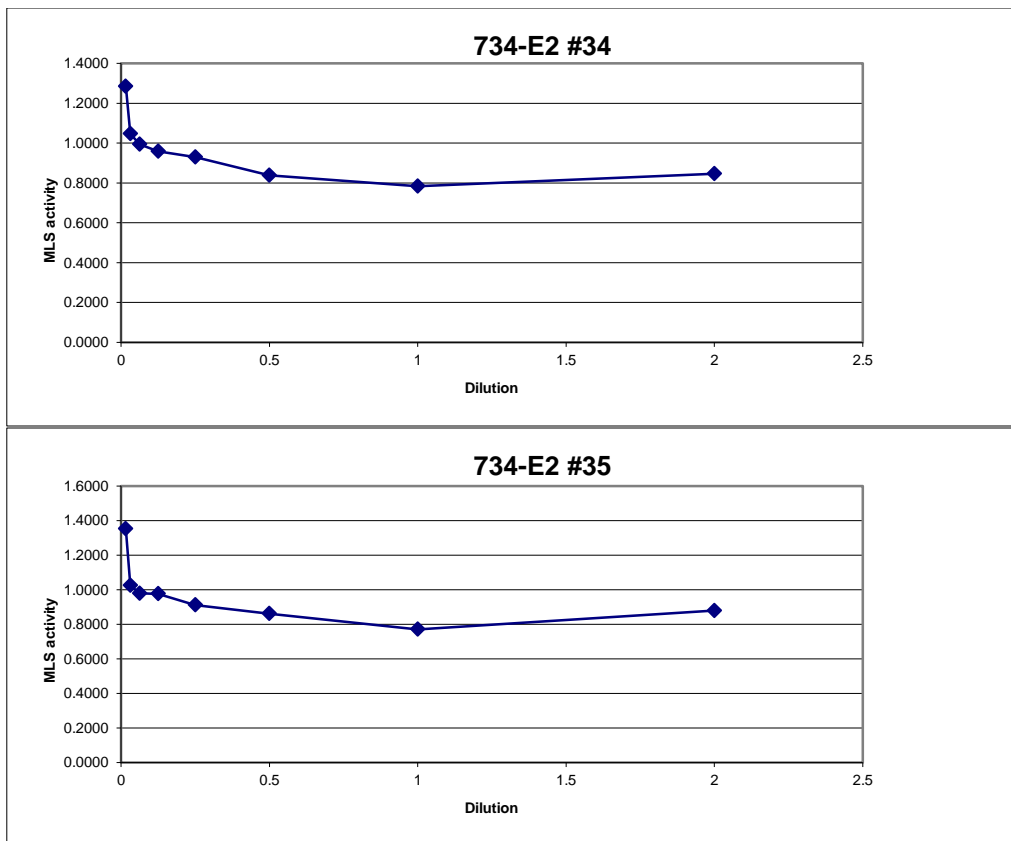
Figure 58 (1-5): Continued.

The active fractions were tested in a dilution series ranging from 2 X (most concentrated) to 0.02 X (least concentrated). Fraction number 32 of 597-B5 inhibited *Mtb* malate synthase at concentrations of up to 0.06 X. Fraction number 38 of 597-B5 inhibited *Mtb* malate synthase at concentrations of up to 0.03 X. Fraction number 38 of 1146-F11 inhibited *Mtb* malate synthase at concentrations of up to 0.03 X. Fraction numbers 34 and 35 of 734-E2 inhibited *Mtb* malate synthase at concentrations of up to 0.06 X.

Conclusion

The advantage of a target directed *in vitro* screening is that the target is already known. Malate synthase is an attractive drug target in *Mtb* due to its association with persistence factors in different organisms like *Candida albicans*, *Brucella abortus* and *Rhodococcus fascians* and the absence of the same enzyme in human beings. The only test that remains to be done for the enzyme inhibitor is to check for its cell permeability. As the subset of 640 extracts were selected based on their whole-cell bioactivity against *Mtb*, it is already known that the bioactive component of the extract is cell permeable. HPLC fractions of three extracts from the 640 extract library were shown to have *in vitro* inhibitory activity against *Mtb* malate synthase in a high-throughput enzyme assay. Active fractions from the three extracts were tested in serial dilutions and were found to exhibit *in vitro* activity at concentrations as low as 0.03 X. The absence of structural information on the *in vitro* active compounds in the extracts precludes the accurate determination of an IC_{50} value. The amount of material present in the extract mixture was not sufficient to perform NMR analysis. Efforts are currently underway to obtain more crude extracts for the active extracts. The isolation and purification of the natural products in the active HPLC fractions on a preparative scale will enable the accurate structural characterization of the natural products exhibiting the *in vitro* inhibitory activity against *Mtb* malate synthase.

CHAPTER V

NATURAL PRODUCT INHIBITORS OF THE PROTEASOME

Introduction

The high-throughput or low-throughput screening of a collection of synthetic

compounds generated by combinatorial chemistry in a whole-cell bioactivity assay can

be used to identify lead molecules with desirable cell wall penetrating capabilities.

Molecules synthesized to populate a high-throughput library of compounds are designed

to represent chemical diversity in relation to the functionalities or structural features.

However the structural and functional diversity may or may not assist in increasing the

cell penetrating capability of the combinatorial compound libraries. Lipophilicity is part

of the reason for the cell penetration power of a small molecule [209]. However, the

complete physico chemical properties of a desirable cell penetrating compound are not

completely understood. Natural products have a built-in advantage in this regard since

they are naturally built to traverse through cell membranes. Secondary metabolites with

bioactive properties that are biosynthesized in the cytoplasm of the producing

microorganism are through evolution perfected by nature to be secreted by one microbe

and taken up by another. Since the ability to penetrate cell walls constitute one of the

'drug-like' properties in a molecule, natural products have this natural 'drug-like'

property built into them. Hence, whole-cell bioactivity tests, while necessary for

synthetic molecules, may not be critical for natural products. A faster approach to

identifying natural product inhibitors is to conduct a targeted *in vitro* assay. Hence, in

this research instead of a whole-cell assay the high-throughput screening of the crude microbial extract libraries were directed in a target-based *in vitro* assay. The human proteasome was used as the target for the *in vitro* assay.

The proteasome serves as an important drug target due to its central role in degradation of misfolded proteins and in maintaining the cellular concentration of regulatory proteins involved in cell cycle regulation, apoptosis and angiogenesis [210]. The degradation process is modulated by the ubiquitin-proteasome pathway and generates peptides and amino acids, essential raw materials for synthesizing new proteins [211]. An increased activity of the human proteasome is shown in cancer cell lines owing to the increased rate of metabolism [212]. Cancer cells also exhibit higher levels of proteasome expression than normal cells which may contribute to the higher susceptibility of cancer cells to proteasome inhibitors [213]. Proteasome is an attractive drug target because the inhibition of proteasome results in the intracellular accumulation of proteins, which results in cell growth arrest and programmed cell death [214]. A number of different classes of synthetic compounds have been tested for anti-proteasome activity, including peptide aldehydes like MG-132, peptide vinyl sulphones like NIP-Leu$_3$-vinyl sulfone, peptide boronates like PS-341, peptide epoxyketones like epoxomicin and β-lactones like lactacyctones [215-218]. Natural product inhibitors of proteasome have also been identified, like TMC-95A derived from *Apiospora montagnei* [219], gliotoxin deried from fungus [220] and (−)-epigallocatechin-3-gallate (EGCG) found in green tea extracts [221]. Velcade (Bortezomib or PS-341), a dipeptide boronate compound was the first proteasome

184

inhibitor to be approved by the food and drug administration for the treatment of multiple myeloma [222,223].

Proteasome inhibition studies were initiated by testing aliquots of the crude microbial extracts in *in vitro* proteasome inhibition assays. The activity was measured against the chymotrypsin-like and caspase-like activity of the proteasome. The activities of the extracts were narrowed down to the bioactive components in the extracts by successive rounds of HPLC purifications followed by *in vitro* assays and mass spectrometric analysis to follow the MS peaks of interest that appeared consistently in the fractions showing *in vitro* proteasome activity. Finally NMR spectroscopy and MS fragmentation data was used to elucidate the structure of the active molecule.

Methods

High Performance Liquid Chromatography Fractionation

Crude material was obtained from Ohr pharmaceuticals in vials and round bottom flasks. The vials had insufficient material inside them to conduct HPLC fractionations and test the individual fractions. The crude material inside the round bottom flasks were sufficient to conduct multiple fractionations, followed by LC / MS and possibly NMR spectroscopic analysis. Crude material from five round bottom flasks were selected for HPLC fractionation and *in vitro* enzyme activity assays against the human and yeast proteasome. An aliquot of the crude was taken into vials to prepare an approximate stock solution of 20 mg / ml for each extract. 20 µL of the stock solution of the extracts were fractionated by injecting on to a reverse phase C18 column on a

Waters® high performance liquid chromatography (HPLC) with a photodiode array detector, a manual injector and a WFC III® fraction collector, also from Waters®. The gradient conditions were as follows;

Column specification: 4.6 X 250 mm Waters Atlantis® T3, 5 μm (This column specification remains the same for all the subsequent LC fractionations on a semi preparative scale unless and otherwise mentioned)

Flow rate: 1 ml / min

Injection volume: 20 μL

Mobile phase: A = acetonitrile + 0.1 % formic acid, B = water + 0.1 % formic acid, C = methanol + 0.1 % formic acid.

Gradient conditions: 50 % B, 50 % C to 100 % C in 15 minutes; 100 % C to 100 % A in 20 minutes; then 100 % A maintained for 10 minutes.

45 fractions were collected directly into the wells of a 96-well deep well polypropylene plate through the WFC III® fraction collector. 100 μL of fractions from each well was saved for LC / MS analysis while the rest was concentrated to test for proteasome activity in the *in vitro* enzyme assay.

The active fractions of Prot-2 were pooled together, the solvent was evaporated and the dry material was reconstituted in methanol to 20 μL. The 20 μL solution was injected into the HPLC column for further fractionation using a shallow gradient. The mobile phase gradient conditions were as follows; 50 % B, 50 % C to 100 % A in 5 minutes; 100 % A to 90 % C, 10 % A in 1 minute; 90 % C, 10 % A to 60 % C, 40 % A in 54 minutes; 60 % C, 40 % A to 100 % A in 5 minutes; 100 % A maintained for 5

minutes. Similarly, the active fractions of Prot-4 were pooled together, the solvent was evaporated and the dry material was reconstituted in methanol to 20 µL. The 20 µL solution was injected into the HPLC column for further fractionation. The mobile phase gradient conditions were as follows; 50 % B, 50 % C to 15 % B, 85 % C in 5 minutes; 15 % B, 85 % C to 100 % C in 30 minutes; 100 % C to 30 % A, 70 % C in 30 minutes; 30 % A, 70 % C to 100 % A in 5 minutes; 100 % A maintained for 5 minutes.70 fractions for Prot-2 and 75 fractions for Prot-4 were collected directly into the wells of 96-well deep well polypropylene plates.

Preparative scale fractionation for Prot-2 and Prot-4 were performed by dissolving 12.5 mg of the crude material for Prot-2 and 38 mg of crude material for Prot-4 in 200 µL of methanol. The stock solution was injected into a 10 X 250 mm Waters Atlantis® T3, 5 µm preparative scale HPLC column. A flow rate of 5 ml / min was used. The gradient condition for the preparative scale fractionation of Prot-2 was as follows; 50 % B, 50 % C to 100 % C in 15 minutes; 100 % C to 100 % A in 50 minutes; 100 % A maintained for 5 minutes.

The gradient condition for the preparative scale fractionation of Prot-4 was as follows;
95 % C, 5 % B to 100 % C in 30 minutes; 100 % C to 100 % A in 30 minutes; 100 % A maintained for 10 minutes.

The fractions were collected directly into glass tubes at the rate of 5 ml / min /fraction.

In Vitro *Assay*

The *in vitro* proteasome assay for Chymotrypsin-like activity was performed using a fluorogenic substrate Suc-Leu-Leu-Val-Tyr-AMC and the Caspase-like activity was performed using a fluorogenic substrate Z-Leu-Leu-Glu-AMC. 20S proteasome was isolated and purified from human red blood cells and used in the assay at a final concentration of 100 nM. The reaction buffer consisted of 20 mM Tris pH 7.5, 1 mM EDTA, 1 mM DTT and 0.03 % SDS. The proteasome was first incubated with the inhibitor at 37 °C for 20 minutes in the reaction mixture containing the buffer. 5 μM fluorogenic substrate was added to initiate the reaction. The increase in fluorescence was monitored for 30 minutes and the reaction rates were measured as slopes of fluorescence versus time. 2 % DMSO instead of the inhibitor was used as a control to determine the percent inhibition.

Liquid Chromatography / High Resolution Mass Spectrometry

The bioactive fractions were analyzed by Liquid Chromatography / Mass spectrometry (LC / MS) using the high resolution micrOTOF-Q II® mass spectrometer from Bruker Daltonics Inc to help prioritize the extracts of further interest. LC / MS data was obtained for the *in vitro* active fractions from the extracts on an Agilent 1200 Infinity series HPLC with temperature controlled autosampler and a photodiode array detector and the in-line micrOTOF-Q II® mass spectrometer from Bruker Daltonics Inc. The LC conditions were as follows:

Column specification: 2.1 X 100 mm, Waters Atlantis T3®, 5 μm

Flow Rate: 0.5 ml / min

Mobile phase: A = water + 0.1 % formic acid, B = acetonitrile + 0.1 % formic acid.

Gradient conditions: 90 % A, 10 % B to 100 % B in 8 min; 100 % B continued for 4

min; 90 % A, 10 % B in 2 min; 90 % A, 10 % B continued for 3 min.

The high resolution mass spectrum (HRMS) was collected in electro spray

ionization (ESI) positive mode in the mass range of 150 m/z to 2000 m/z. The ESI

source settings were as follows;

End plate offset = - 500 V

Capillary = - 4500 V

Nebulizer = 4.0 Bar

Dry Gas = 11.0 L / min

Dry temp = 200° C

MS peaks of interest were identified by comparing the HRMS data for the active

fractions with those of the neighboring inactive fractions and also from solvent (blank)

runs to eliminate the insignificant MS peaks from consideration. Probable molecular

formula was generated using the high resolution MS data for the MS peaks of interest.

SmartFormula® prediction calculation available with the Data Analysis® software from

Bruker Daltonics Inc. was used for the generation of predicted molecular formulas for

the MS peaks of interest.

Results and Discussion

Aliquots of five different crude extracts were weighed and concentrated stock solutions were prepared in DMSO. The five crude extracts were labeled Prot-1 through Prot-5 and will be referred by this label henceforth. Serial dilutions were prepared from the stock solutions and tested for proteasome activity in *in vitro* enzyme assays against the human and the yeast proteasome. The starting concentrations for the serial dilutions were at 20 mg / ml. All five extracts were found to display *in vitro* activity against the human and yeast proteasome. The process of activity based separation was started by injecting 20 mg / ml solutions of the crude extracts in methanol into a semipreperative C18 column on a HPLC (**Figure 59**). Solvent gradient systems starting at 50 % methanol and going up to 100 % acetonitrile in 45 minutes were used. 1 ml/min LC flow rate was maintained.
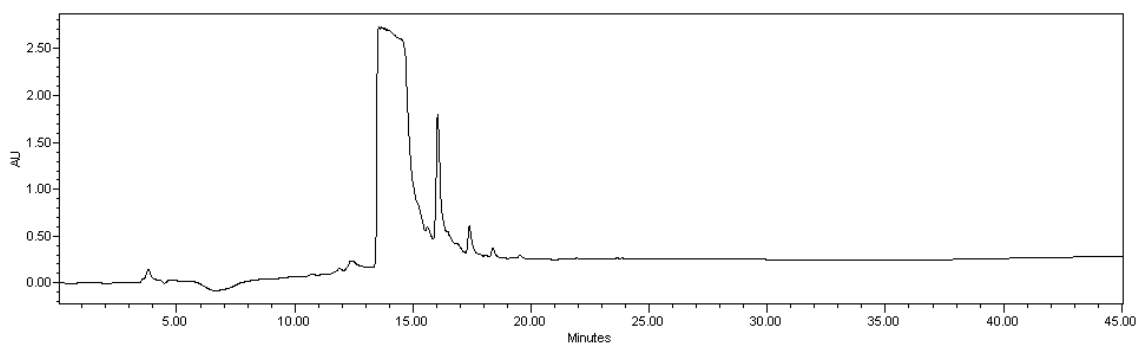
Figure 59: The LC UV profile for Prot-4 as a representative of the HPLC fractionation of the five different crude extracts.

190

All fractions were collected directly into the wells of 96-well deep well polypropylene plates. An aliquot of each of the LC fraction for all five extracts were again tested in *in vitro* assays. The activity profiles of the fractions are shown in **figure 60**.
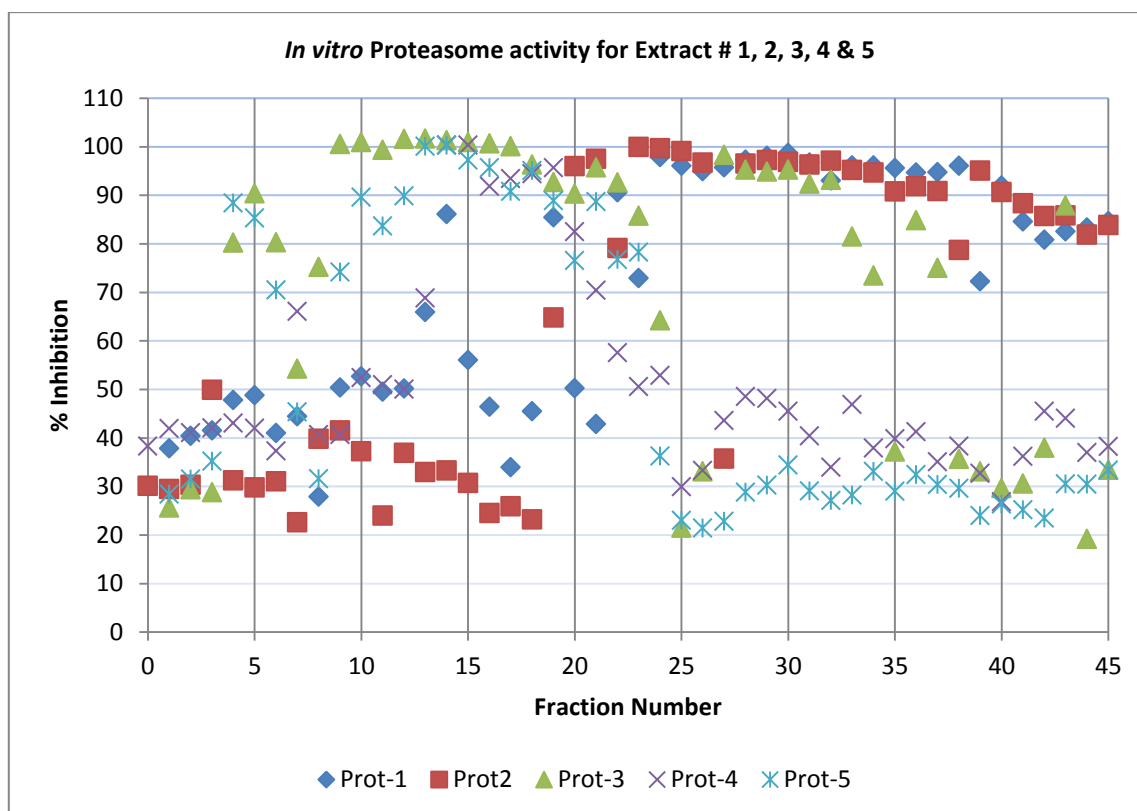
Figure 60: The percentage inhibition of human proteasome in an *in vitro* enzyme assay. The activity profiles of the fractions for all five crude extracts are overlaid in a single plot.

As the physical parameters like the molecular weight of the active component in the extracts were unknown it was not possible to estimate the concentration of extract solution. Therefore, the concentrations of solution were prepared empirically based on the *in vitro* activity of the fractions. The abundance of activity found in the fractionations of the five extracts in the figure above showed that a concentrated solution of extracts was injected into the HPLC column resulting in the spreading of the active component across a broad retention time range. The number of active fractions in each extract is shown below in **figure 61**.

The activity profile for all five extracts suggested that Prot-1 and Prot-2 might belong to one type of extract while Prot-4 and Prot-5 might belong to another type of extract with Prot-3 showing overlapping activity profiles with both types of the extracts. The LC UV profiles of Prot-1 and Prot-2 were similar while the LC UV profiles of Prot-

Prot-1

	1	2	3	4	5	6	7	8	9	10	11	12
a	x	15	16	31	32	x						
b	1	14	17	30	33	x						
c	2	13	18	29	34	45						
d	3	12	19	28	35	44						
e	4	11	20	27	36	43						
f	5	10	21	26	37	42						
g	6	9	22	25	38	41						
h	7	8	23	24	39	40						

Prot-2

	1	2	3	4	5	6	7	8	9	10	11	12
a	x	15	16	31	32	x						
b	1	14	17	30	33	x						
c	2	13	18	29	34	45						
d	3	12	19	28	35	44						
e	4	11	20	27	36	43						
f	5	10	21	26	37	42						
g	6	9	22	25	38	41						
h	7	8	23	24	39	40						

Prot-3

	1	2	3	4	5	6	7	8	9	10	11	12
a	x	15	16	31	32	x						
b	1	14	17	30	33	x						
c	2	13	18	29	34	45						
d	3	12	19	28	35	44						
e	4	11	20	27	36	43						
f	5	10	21	26	37	42						
g	6	9	22	25	38	41						
h	7	8	23	24	39	40						

Prot-4

	1	2	3	4	5	6	7	8	9	10	11	12
a	x	15	16	31	32	x						
b	1	14	17	30	33	x						
c	2	13	18	29	34	45						
d	3	12	19	28	35	44						
e	4	11	20	27	36	43						
f	5	10	21	26	37	42						
g	6	9	22	25	38	41						
h	7	8	23	24	39	40						

Prot-5

	1	2	3	4	5	6	7	8	9	10	11	12
a	x	15	16	31	32	x						
b	1	14	17	30	33	x						
c	2	13	18	29	34	45						
d	3	12	19	28	35	44						
e	4	11	20	27	36	43						
f	5	10	21	26	37	42						
g	6	9	22	25	38	41						
h	7	8	23	24	39	40						

Figure 61: The activity profile of 45 fractions for each of the five crude extracts (Prot-1 through Prot-5). The active fractions are colored in the figure.

193

4 and Prot-5 were similar. The UV absorbance values of the peak fractions of Prot-2 were about two and half times higher than the absorbance values of Prot-1. Similarly the absorbance values of the peak fractions of Prot-4 were about 15 times higher than Prot-5. Hence, Prot-2 and Prot-4 were taken to be representatives of the two types of extracts and were investigated further. The active fractions of Prot-2 were pooled together and evaporated to dryness. The dry extract was then reconstituted in methanol and another round of HPLC separation was performed using a shallower solvent gradient system (**Figure 62**). The gradient system is mentioned in the methods section above. Similarly, the active fractions of Prot-4 were pooled together and evaporated to dryness. The dry material was reconstituted in methanol and injected into the C18 column using a shallow solvent gradient as described in the methods section (**Figure 63**).
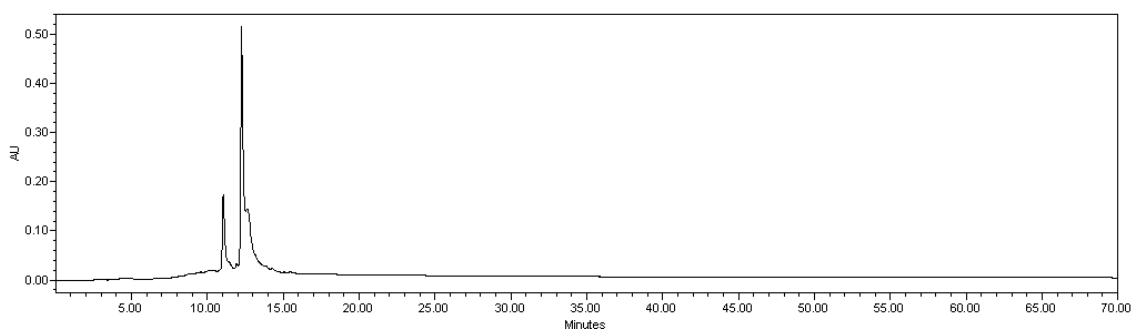
Figure 62: Fractionation of fractions 20 to 22 of Prot-2 on HPLC. UV profile at 230 nm is displayed.
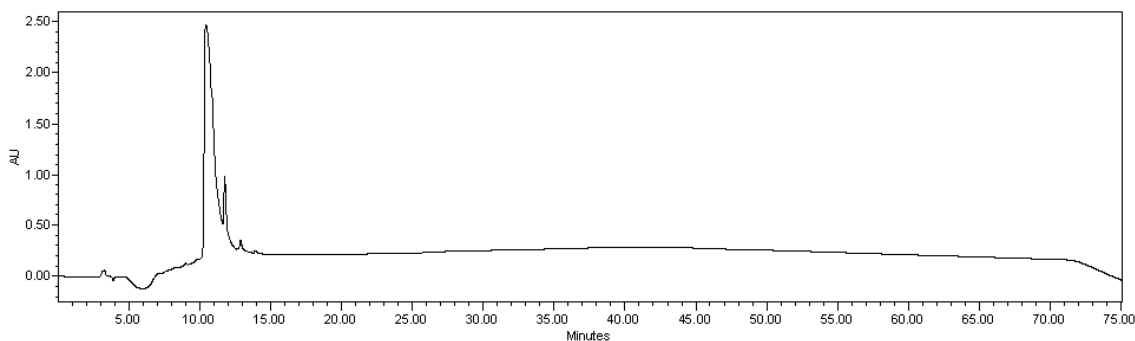
Figure 63: Fractionation of fractions 14 to 20 of Prot-4 on HPLC. UV profile at 230 nm is displayed.

After refractionation, the new fractions were retested for their *in vitro* proteasome activity. In the repeat fractionation, fractions number 12 and 13 were active for Prot-2 and fraction number 11 was active for Prot-4. The amount of material left in each fraction was not sufficient to conduct structural characterization studies. Hence, preparative scale HPLC fractionation was performed for Prot-2 and Prot-4 by injecting a concentrated solution into the preparative C18 column. Accordingly, 12.5 mg extract from Prot-2 dissolved in 200 µL methanol was injected and approximately 38 mg extract from Prot-4 dissolved in 200 µl of methanol was injected into the preparative scale C18 HPLC column (**Figure 64 and 65**).
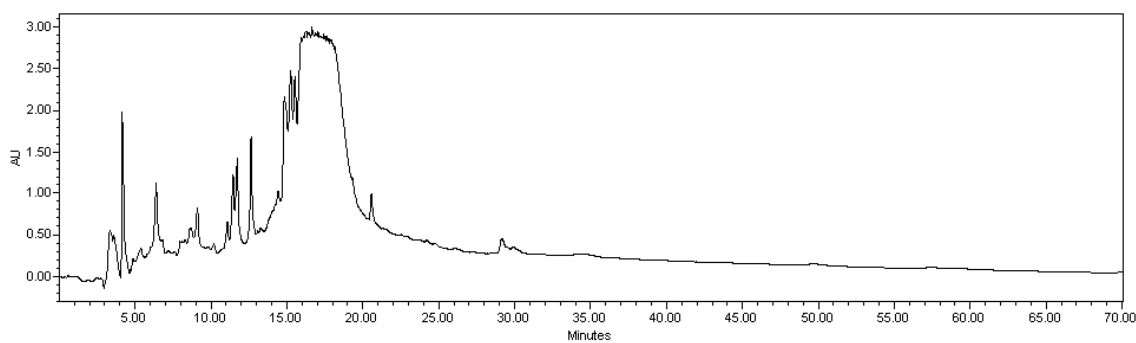
Figure 64: Preparative scale HPLC fractionation of Prot-2. UV profile at 230 nm is displayed.
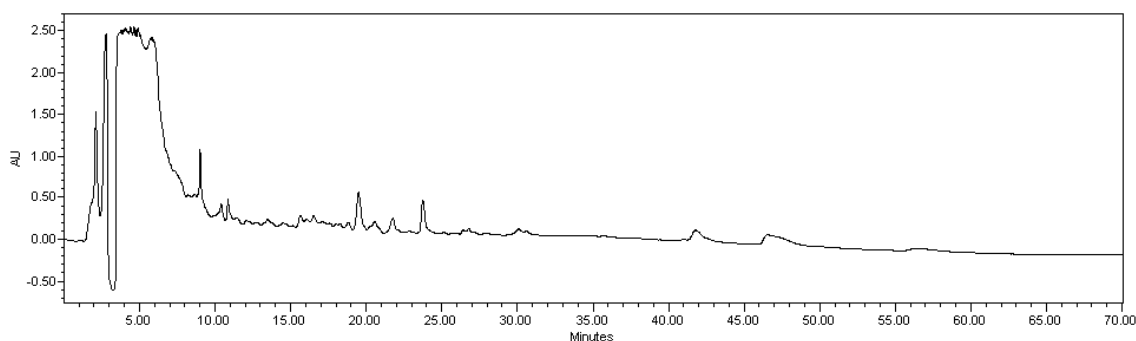
Figure 65: Preparative scale HPLC fractionation of Prot-4. UV profile at 230 nm is displayed.

The fractions were collected into glass fraction tubes at the rate of 5 ml / fraction / min. The active fractions were identified by performing the *in vitro* proteasome assay against all the fractions. Fraction number 18 for Prot-2 and fraction number five for Prot-4 were active.

196

The active fractions were analyzed by LC / HRMS to identify MS peaks of interest (**Figure 66 – 68 and 69 – 71**).
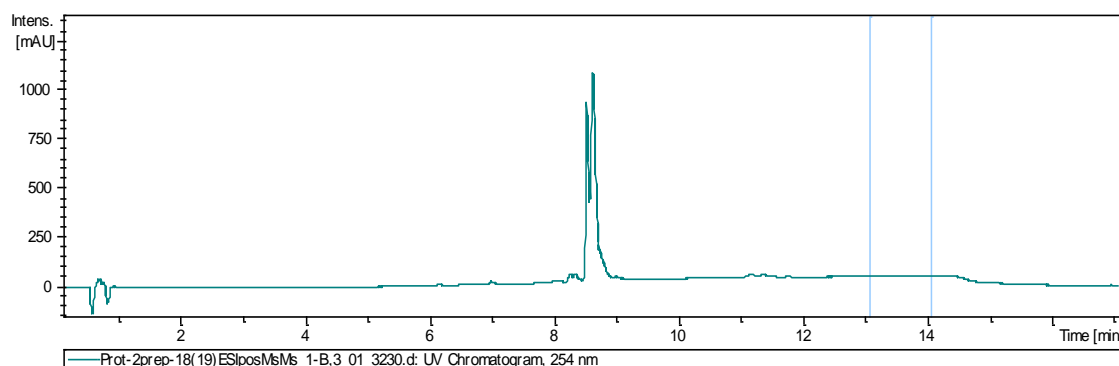
Figure 66: LC / MS analysis of the active fraction number 18 of Prot-2.
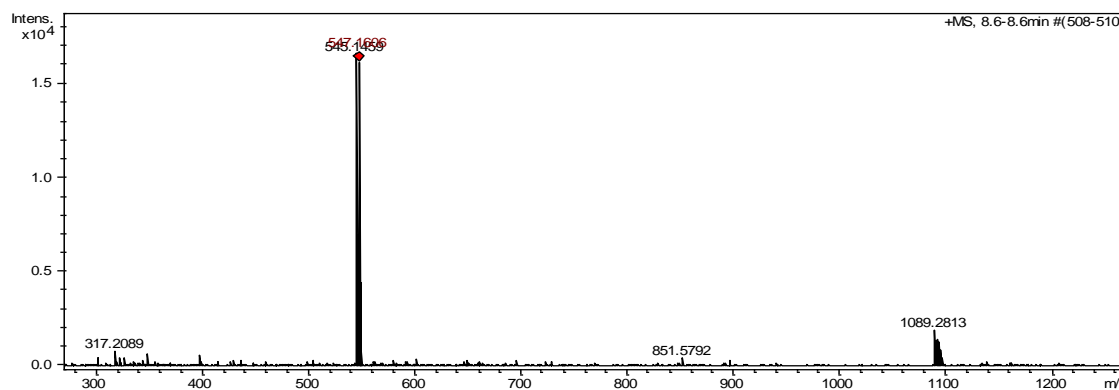
Figure 67: MS for the LC peak at 8.6 minutes in ESI positive mode.
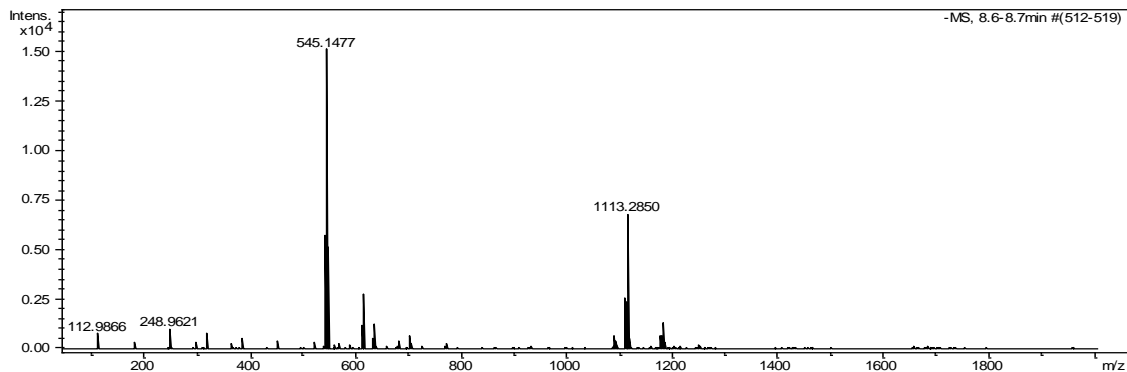
Figure 68: MS for the LC peak at 8.6 minutes in ESI negative mode.
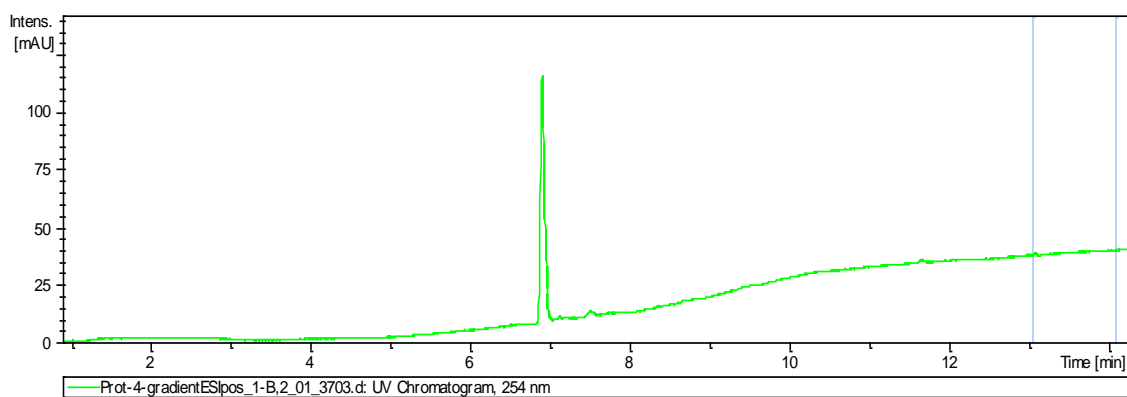
Figure 69: LC / MS analysis of the active fraction number five of Prot-4.
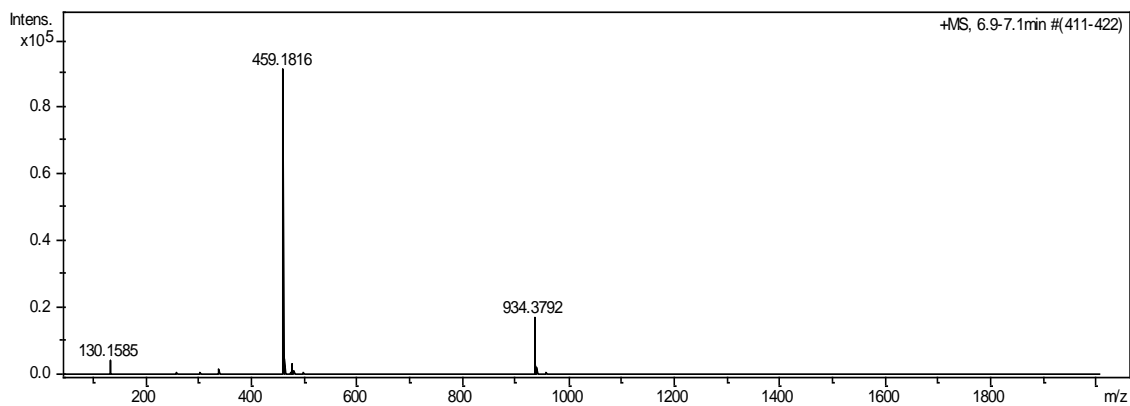
Figure 70: MS for the LC peak at 6.9 minutes in ESI positive mode.
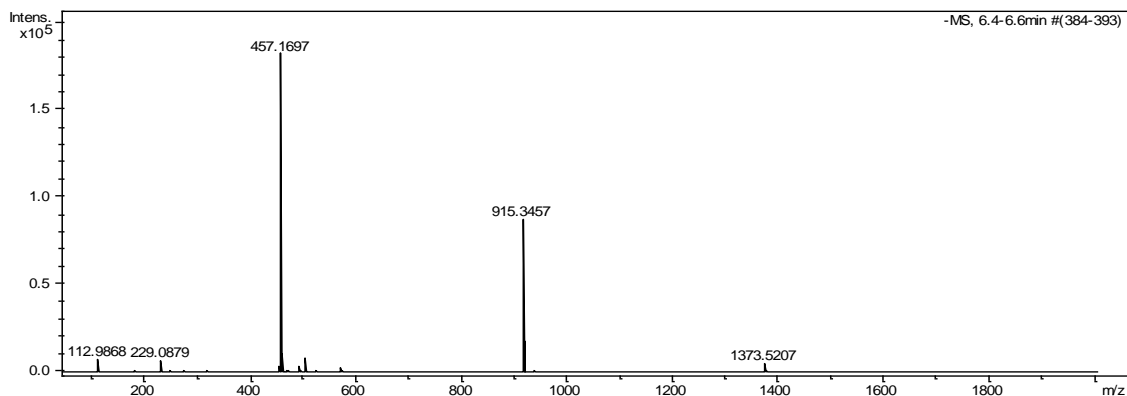
Figure 71: MS for the LC peak at 6.5 minutes in ESI negative mode.

Two MS peaks of interest were observed for Prot-2 at (m/z) 545.14 and (m/z) 1089.28 in the ESI positive while peaks at (m/z) 543.13 and (m/z) 1109.25 were observed in the ESI negative mode. Two MS peaks of interest were observed for Prot-4

at (m/z) 459.18 and (m/z) 934.38 in the ESI positive while a major peak at (m/z) 457.17 was observed in the ESI negative mode.

NMR studies were initiated on fraction number 18 of Prot-2 and fraction number five of Prot-4. The ^{1}H NMR data for fraction number 18 of Prot-2 was not conclusive. The ^{1}H NMR peaks suggested the presence of an impurity. The ^{13}C NMR data suggested that the concentration of the active compound was not sufficiently high to carry out further NMR studies.

Fraction number five of Prot-4 extract had a single sharp UV peak at the retention time of five minutes. The UV absorbance was sufficiently high at 2.5 AU. [1]H, ^{13}C, Distortionless Enhancement by Polarization Transfer (DEPT) - 135°, DEPT - 90°, COSY, HSQC, HMBC and NOESY experiments were performed on the active fraction to elucidate the chemical structure of the compound (**Figure 72 – 79**). ^{1}H NMR spectra and ^{13}C NMR spectra were recorded on a Bruker Avance III HD NMR system. $(CD_3)_2CO$ was used as the solvent with tetramethylsilane as the internal standard. Proton signals were observed in the ^{1}H NMR spectrum as follows; (500 MHz) ((CD3)2CO): δ = 2.01 (s, 3H), 2.11 (s, 3H), 6.14 (s, 1H), 6.15 (d, J=2 Hz, 1H), 6.20 (m, 1H), 6.35 (s, 1H), and 6.52 (d, J=2 Hz, 1H). ^{13}C NMR chemical shifts were observed at (125 MHz), ((CD3)2CO), 20.2, 21.3, 103.6, 104.2, 111.4, 111.7, 112.2, 119.8, 140.6, 140.7, 156.8, 157.7, 158.9 and 159.0. The DEPT – 90° and 135° revealed that there were 10 CH groups, 4 CH3 groups, no CH2 groups and 14 carbons had no attached protons. COSY, HSQC, HMBC and NOESY NMR experiments were performed to elucidate the correlation between the ^{1}H and ^{13}C NMR spectra.
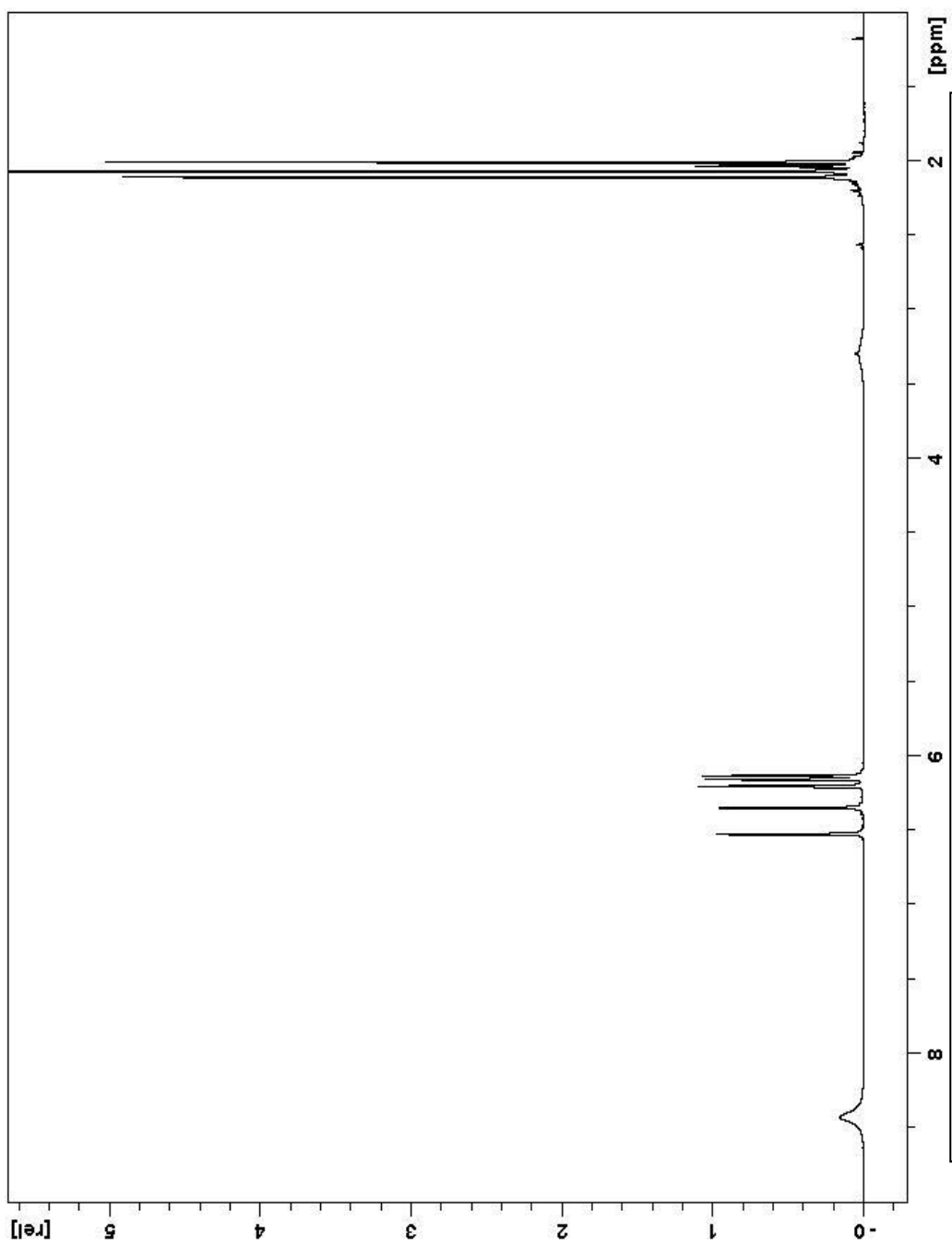
Figure 72: ^{1}H NMR spectrum for Prot-4 (full spectrum).
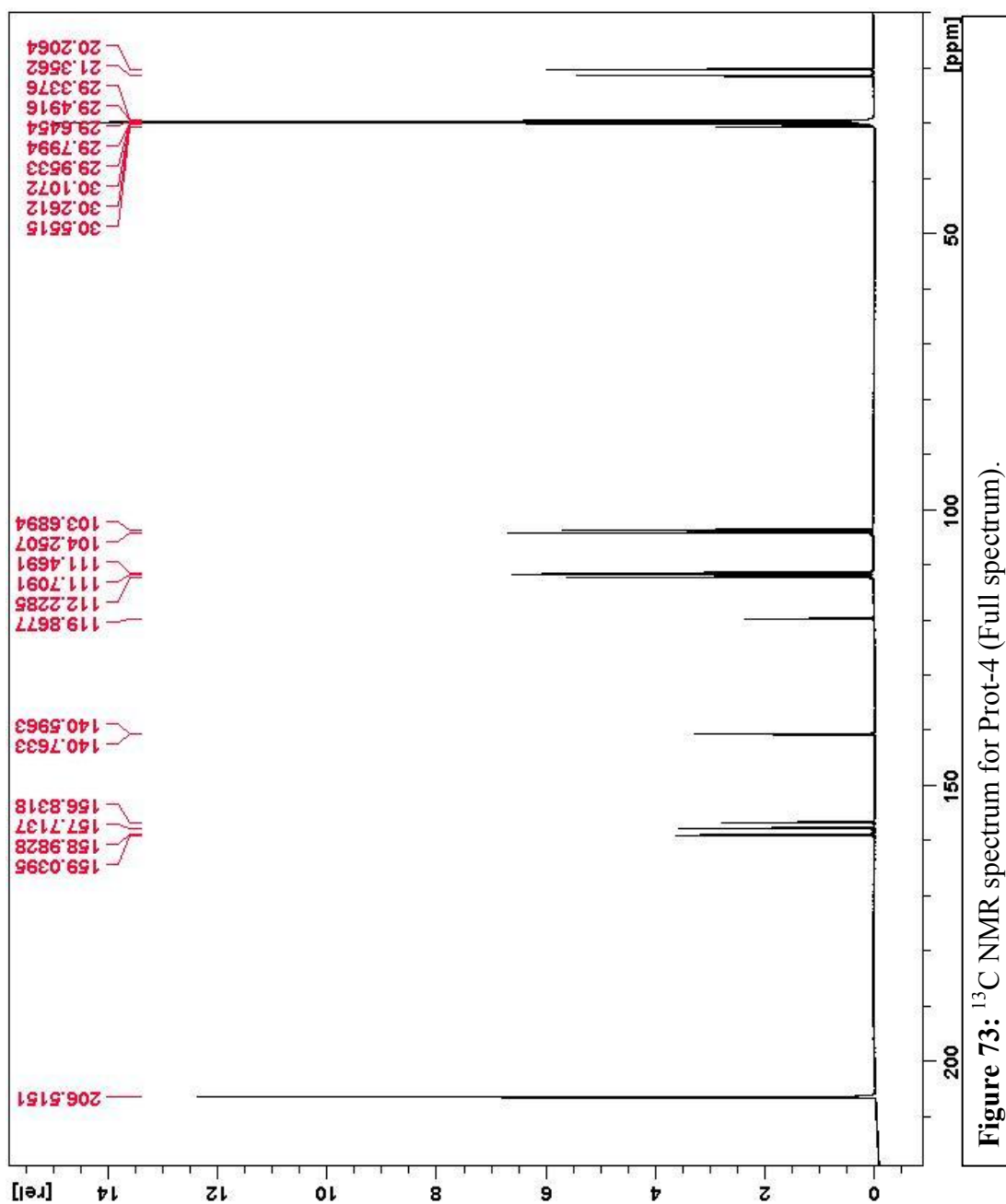
Figure 73: ^{13}C NMR spectrum for Prot-4 (Full spectrum).

202
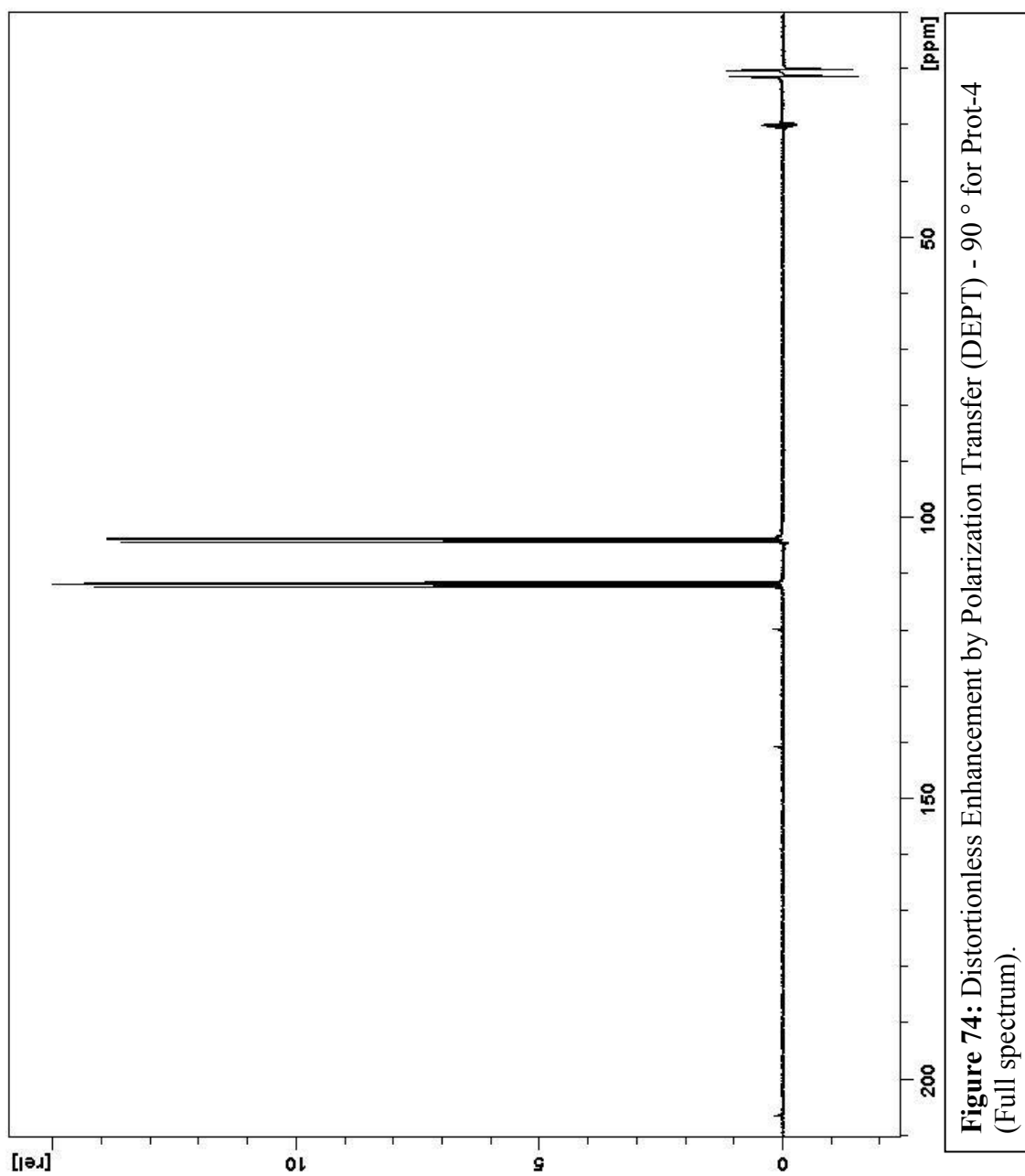
Figure 74: Distortionless Enhancement by Polarization Transfer (DEPT) - 90 ° for Prot-4 (Full spectrum).
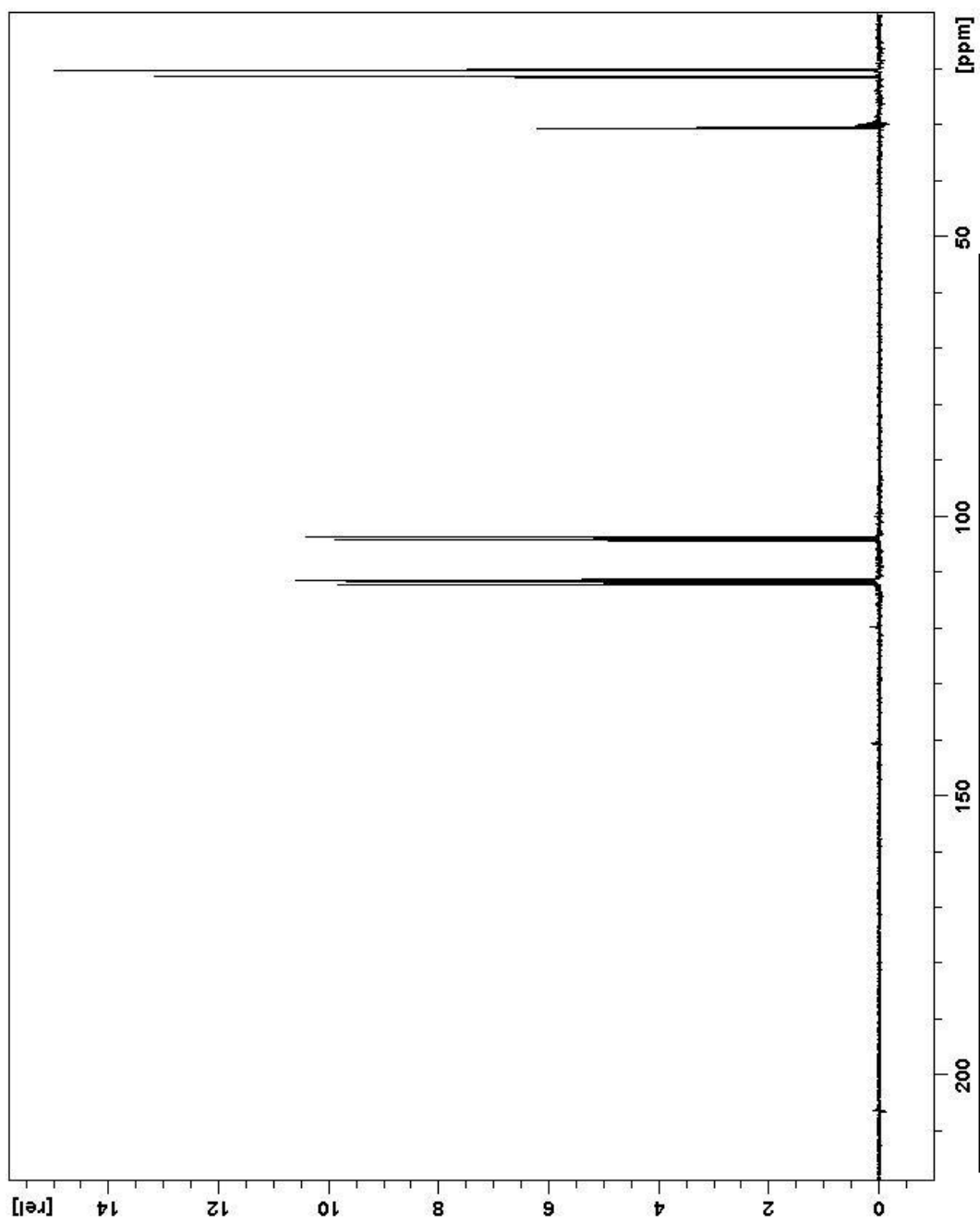
Figure 75: DEPT - 135 ° for Prot-4 (Full spectrum).
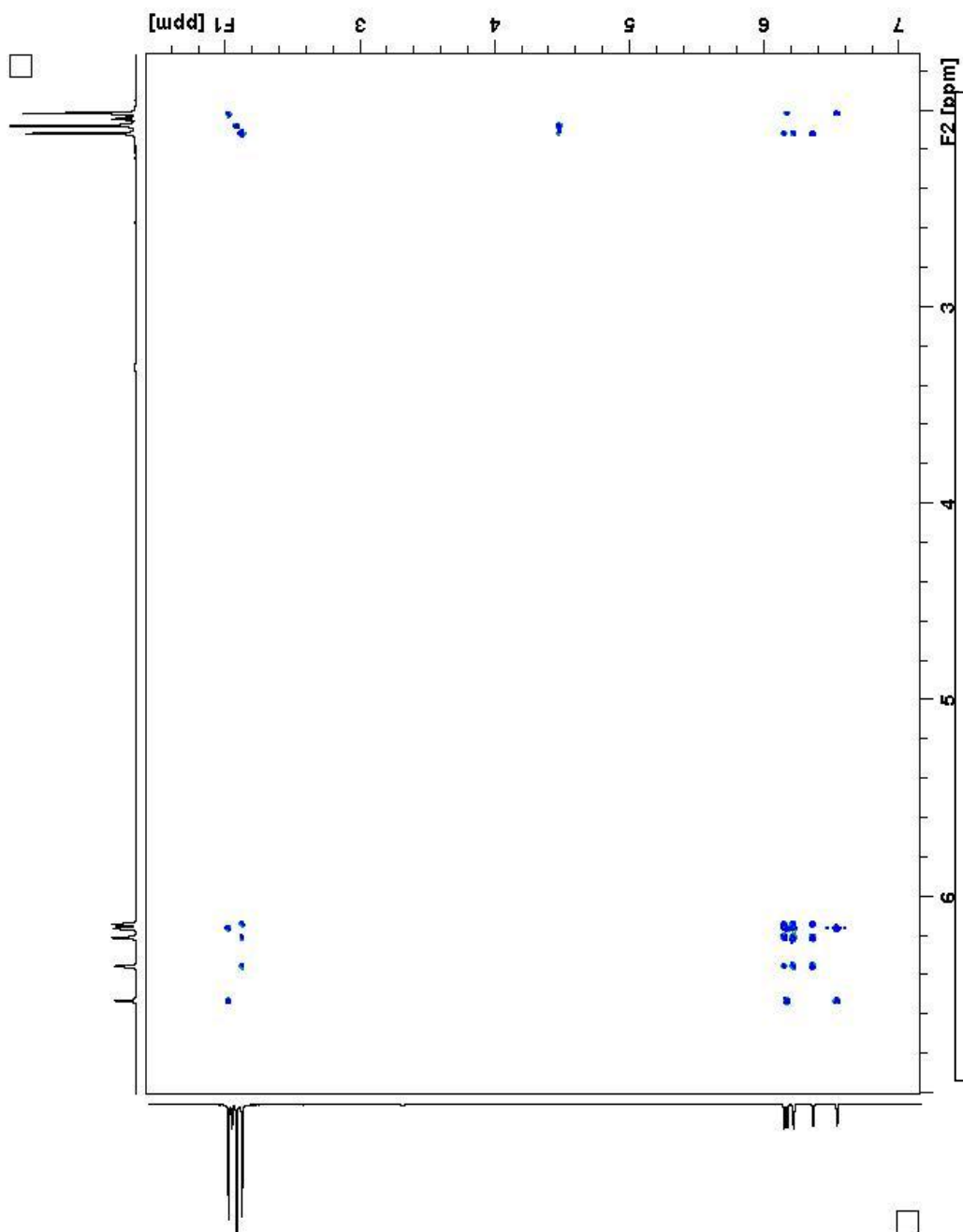
Figure 76: COSY spectrum for Prot-4.
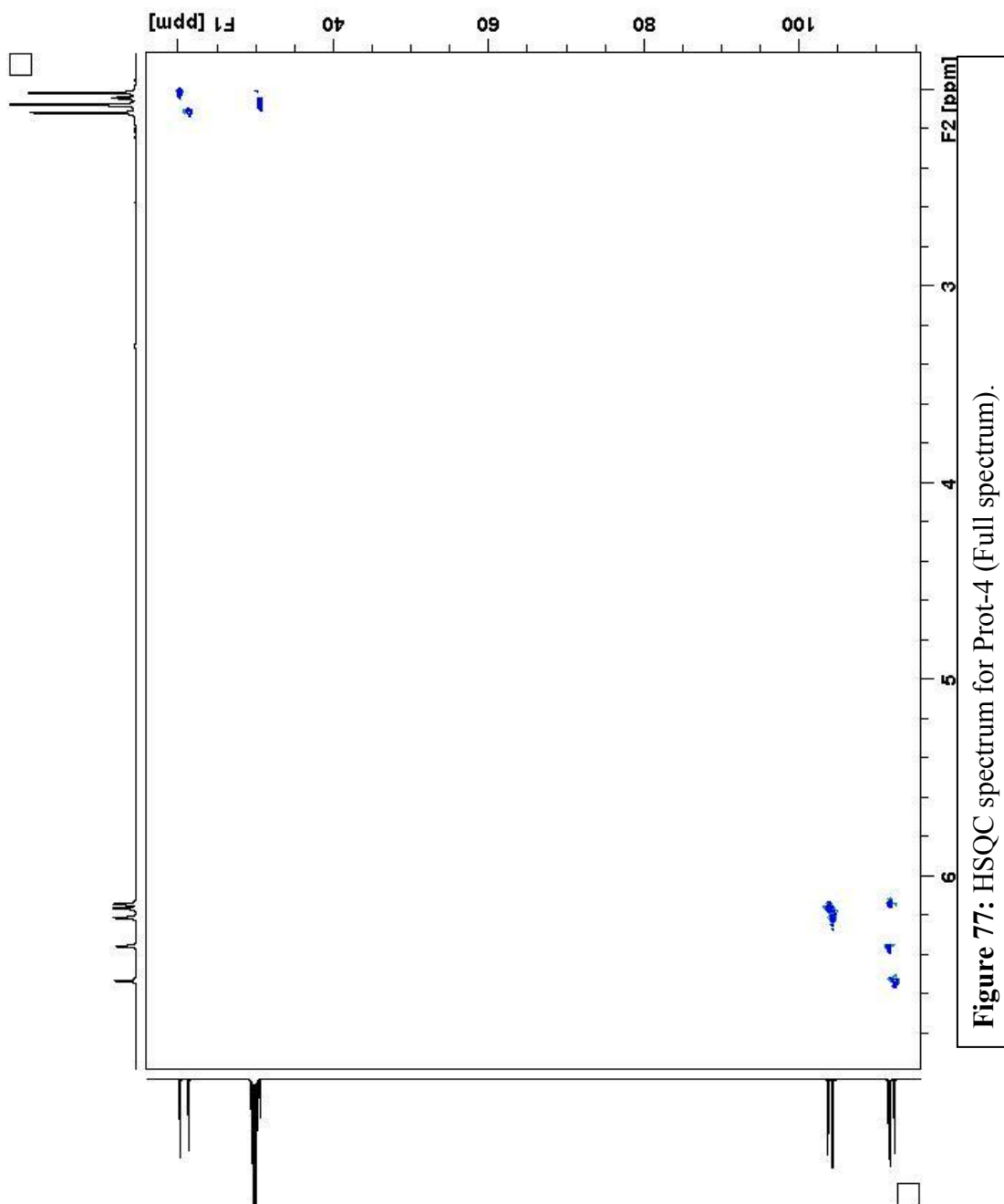
Figure 77: HSQC spectrum for Prot-4 (Full spectrum).
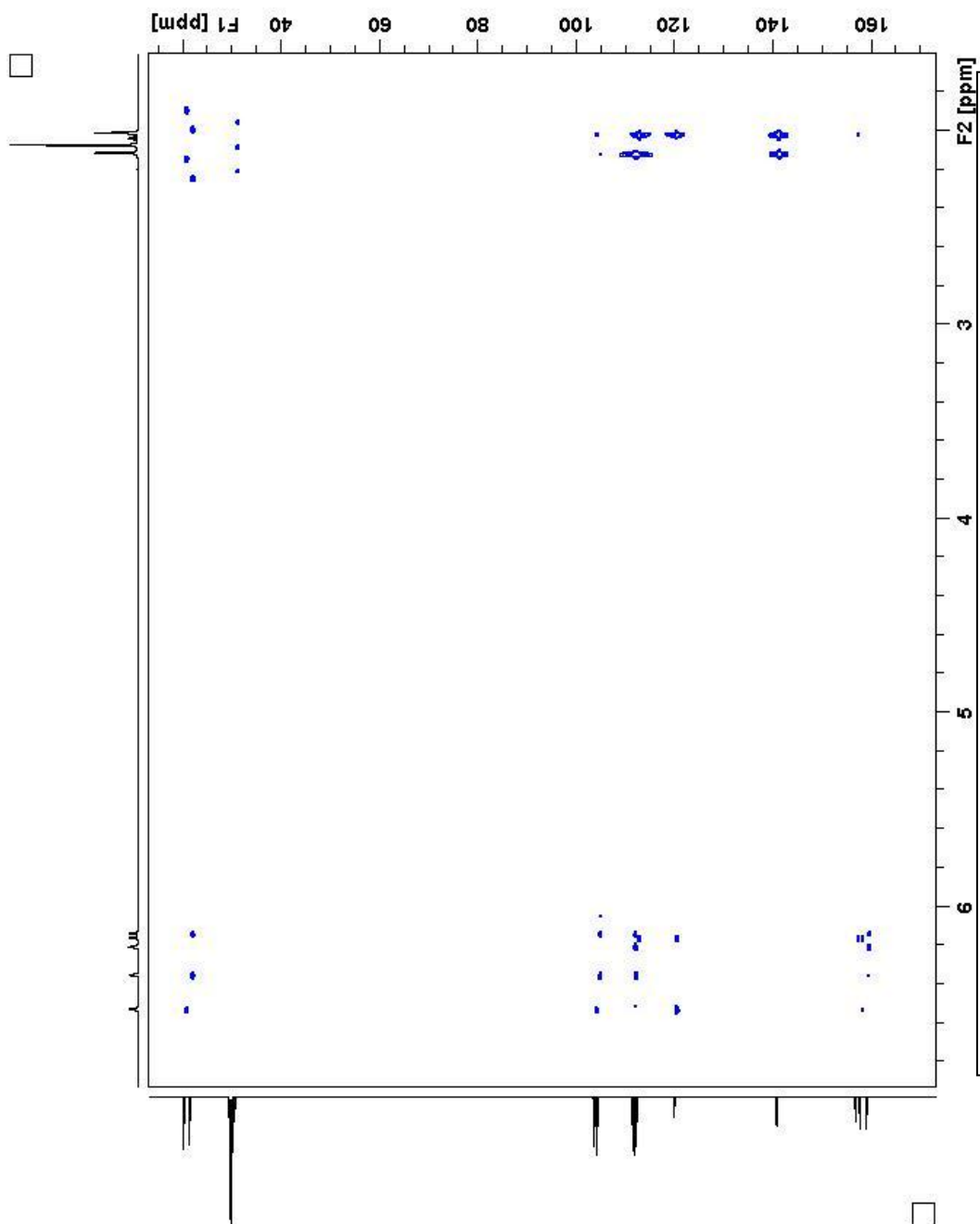
Figure 78: HMBC spectrum for Prot-4 (Full spectrum).
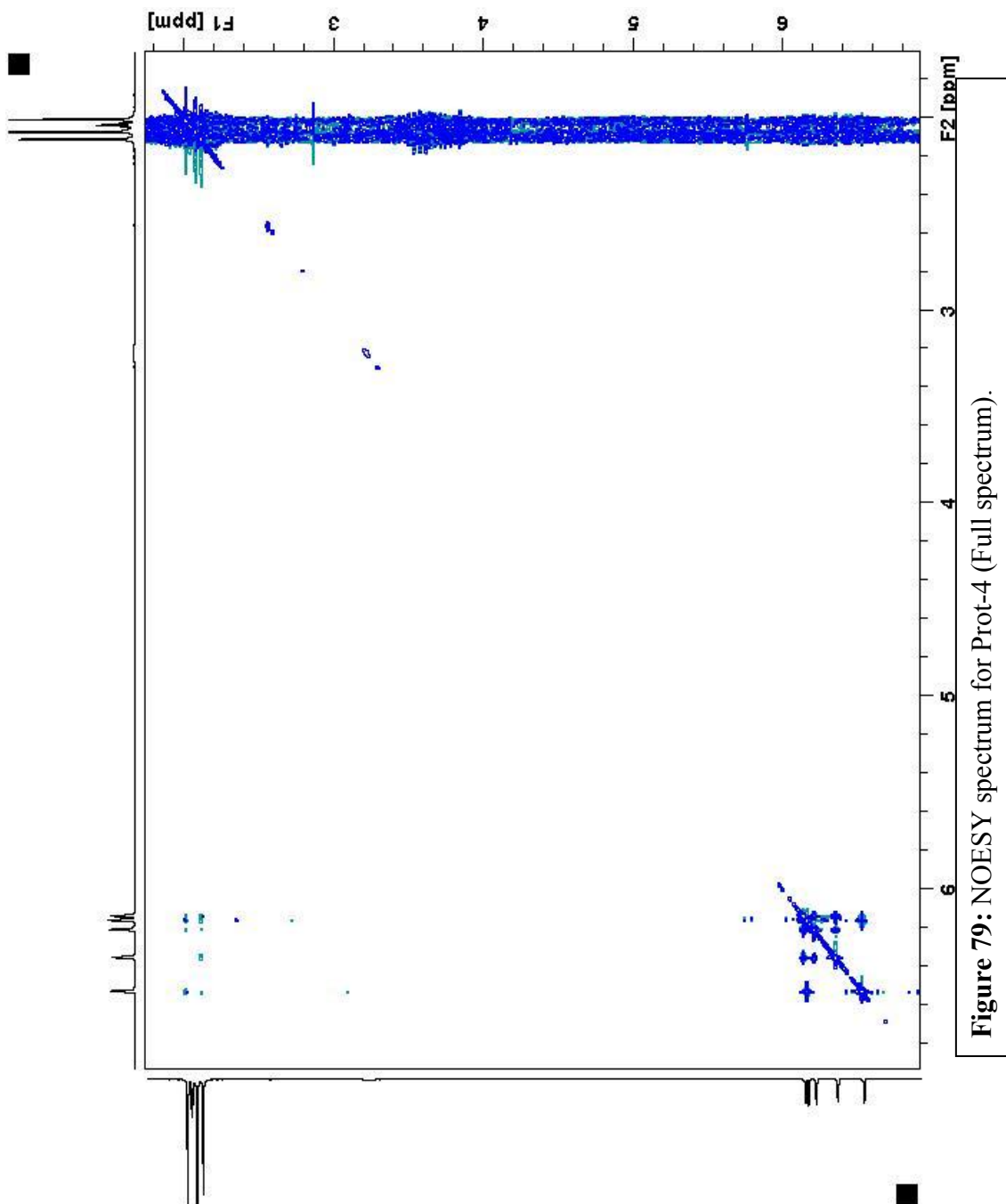
Figure 79: NOESY spectrum for Prot-4 (Full spectrum).

Structure was assigned based on the NMR data as shown in the Figure 80;
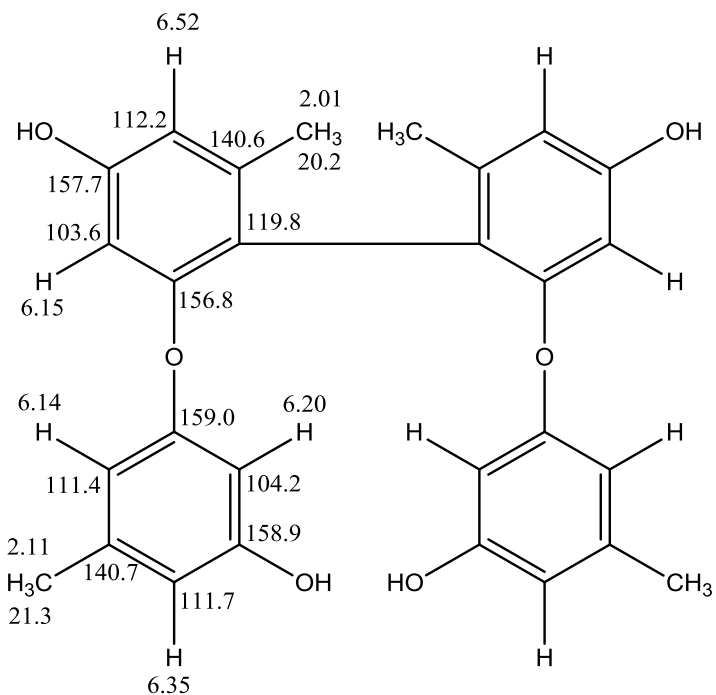
Figure 80: Assignment of ^{1}H and ^{13}C NMR chemical shifts to the chemical structure of Prot-4.

The HRMS in the ESI positive mode was observed at m/z = 459.1816 and in the ESI negative mode was observed at m/z = 457.1697. SmartFormula$^®$ prediction also calculated the molecular formula of the active molecule as $C_{28}H_{26}O_6$, which has a theoretical accurate mass calculated at 458.18. Collision induced dissociation (CID) of the molecular ion peak at (m/z) 459.18 by direct infusion into the Q TOF instrument

revealed mass peaks that correlated with the fragmentation pattern of the derived

structure of Prot-4 (**Figure 81 – 83**).
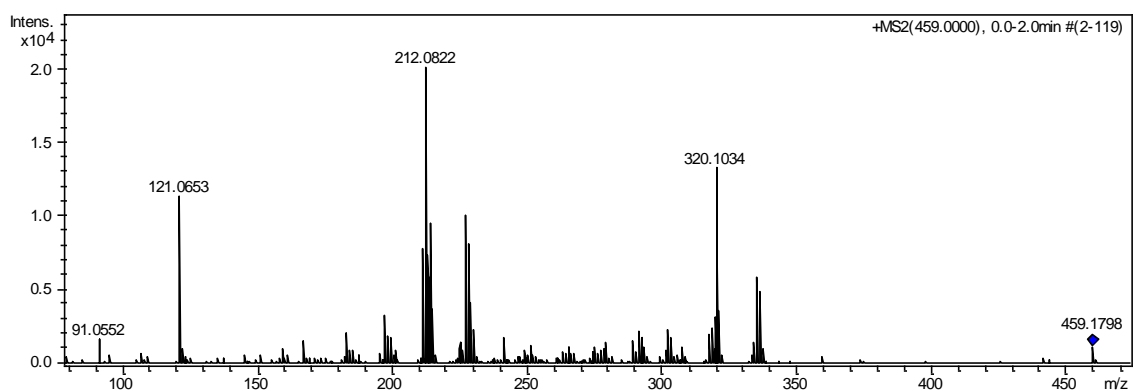
Figure 81: Fragment peaks from the MS / MS of (m/z) 459.18 of Prot-4 (Full spectrum).
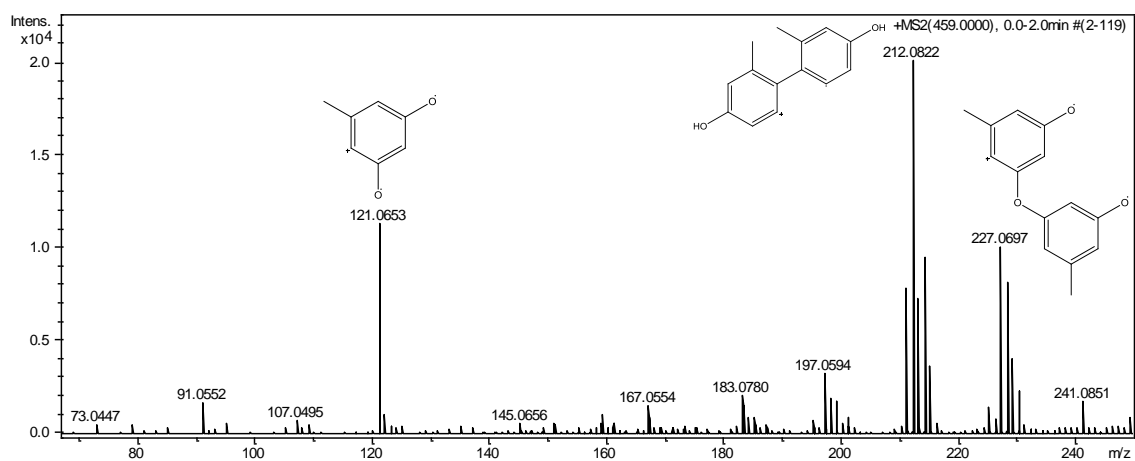
Figure 82: Fragment peaks from the MS / MS of (m/z) 459.18 of Prot-4 (zoomed into (m/z) 70 to (m/z) 250).
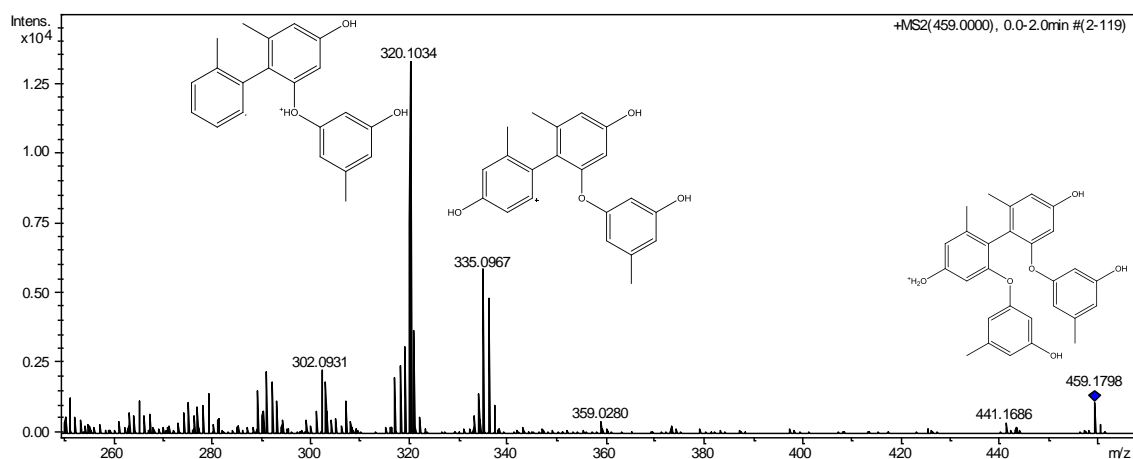
Figure 83: Fragment peaks from the MS / MS of (m/z) 459.18 of Prot-4 (zoomed into (m/z) 250 to (m/z) 460).

Chemical structures could be assigned to all the major fragments observed in the MS /MS of Prot-4 based on the polyphenolic structure of the parent molecule. Thus the structure of the bioactive natural product was confirmed by CID experiment (**Figure 84**).

211
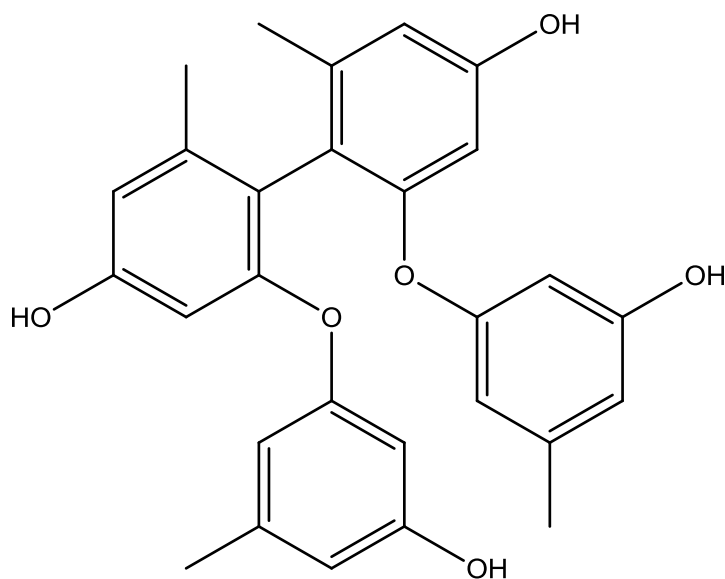
Figure 84: Molecular structure of the proteasome inhibitor - Prot-4.

The IUPAC name for the chemical structure of Prot-4 compound could be derived as 2,2'-di(3-hydroxy-5-methyl phenoxy)-6,6'-dimethyl biphenyl-4,4'-diol. With the knowledge of the molecular formula and the molecular weight information available for the discovered active compound, it was possible to calculate accurate IC_{50} values for the compounds inhibitory activity against the human proteasome. The IC_{50} value was determined to be ~15 µM for the human proteasome against the chymotrypsin-like as well as the caspase-like activity (**Figure 85 and 86**).
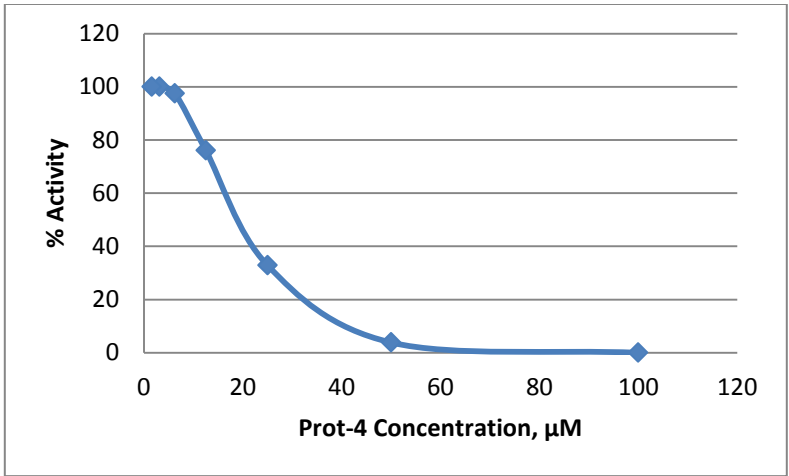
Figure 85: *In vitro* assay of Prot-4 against the chymotrypsin-like activity of the human proteasome.
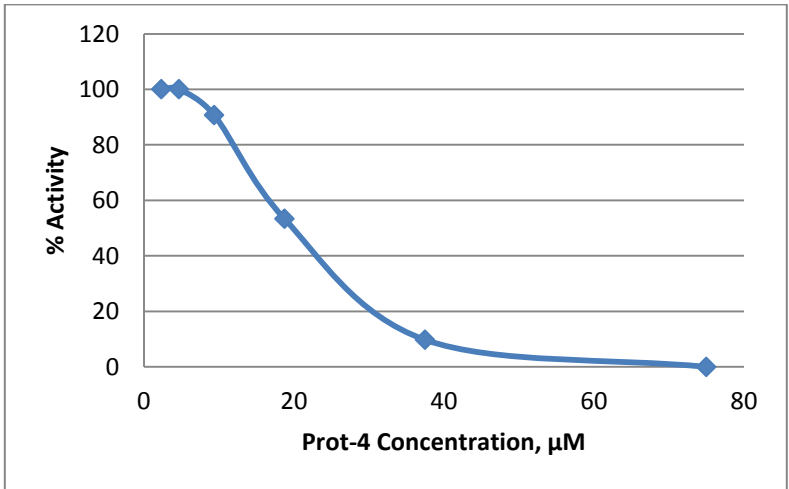
Figure 86: *In vitro* assay of Prot-4 against the caspase-like activity of the human proteasome.

Likewise, whole-cell bioactivity assay was also performed for Prot-4 against different cancer cell lines to establish an MIC value. The MIC against drug resistant ovarian adenocarcinoma cell line, drug resistant breast adenocarcinoma cell line, lymphoma cell line and drug resistant lymphoma cell line were found to be ~ 20 μM which was consistent with the *in vitro* activity.

Conclusion

The process of fractionating a crude extract and screening the fractions for proteasome activity in an *in vitro* assay was followed through successive fractionations of the active fractions until a sufficiently pure fraction was obtained. The pure fraction was analyzed and characterized by LC / HRMS and NMR spectroscopic methods. A previously unknown compound was identified as the source of activity against the human proteasomes.

The characterized natural product is a polyphenolic compound with a molecular weight of 458.18 and molecular formula of $C_{28}H_{26}O_6$. While polyphenols like EGCG have been investigated as inhibitors of proteasome, these polyphenols are commonly found in plants products like vegetables. The source of the extract from which the polyphenol was discovered in this research is unknown, except that the extract is of microbial origin. Polyphenols are generally found in plant products in nature. Microbial sources of polyphenols are rare. If indeed the source of the Prot-4 extract is microbial, then it represents the discovery of a proteasome active polyphenol of microbial origin.

The IC$_{50}$ values of the natural product against human proteasomes are modest at 15 μM. However, this can be used as a starting point to synthesize analogs of the natural product and improve the activity against proteasome. Structure activity relationship (SAR) of the natural product with the human will provide valuable insights into the binding mode of the inhibitor and clues to functional modifications of the inhibitor to enhance the binding affinities and increase potency of the inhibitor.

CHAPTER VI

CONCLUSION

TB is re-emerging and threatens to become a pandemic unless new therapeutic solutions are found. The urgency of the situation demands that all available avenues for drug discovery approach be explored to discover new inhibitors of the pathogenic microorganism, *Mtb*. A two pronged approach was applied to finding inhibitors of *Mtb* - A structure-based approach and a whole-cell inhibition based approach.

The structure-based approach involves the biophysical and biochemical characterization of important biological targets in the pathogen that are selected for inhibition studies. Enzymes involved in important biosynthetic and metabolic pathways or proteins required for virulence factors are considered good targets for the structure-based approach. The proteins or enzymes for which it is possible to design substrate or product analogs that can act as inhibitors of the protein by blocking the catalytic site are called druggable targets. For example, protein kinases are shown to be druggble targets in pathogenic microorganisms [224]. The knowledge of the three dimensional structure of the molecular target in the pathogen including the active site architecture, the domain organizations and protomer interaction surfaces are used to synthesize small molecules that are designed to fit in the active site or an allosteric site within the macromolecular structure of the protein target or at the interface of important protein complex formation sites with high affinity. Tight binding of the designed inhibitors is expected to interfere with the normal cellular functions of the proteins and cause a loss of function.

The whole-cell inhibition approach, on the other hand, assumes no prior knowledge of any intracellular protein or other molecular targets. This is analogous to a blinded study and involves the identification of any inhibitor that can arrest cell growth or cause rapid killing of the cells in a culture medium. At the outset, the intracellular targets of the inhibitor are not known, which precludes the knowledge of any biophysical and biochemical information about the possible intracellular target of the inhibitor. Hence the small molecule inhibitors cannot be designed a priori. Therefore, in order to increase the chances of finding an inhibitor the whole-cell inhibition approach works best when a large number of inhibitors comprising a wide range of structural and chemical scaffold are used in the inhibition assay. Whole-cell inhibition approach therefore requires collections of synthetic or naturally occurring small molecules that can cover the chemical space of a typical 'drug-like' molecule. A 'drug-like' molecule is expected to follow 'Lipinski's rule of five' [190]. As defined by the Lipinski's rule of five, a drug-like molecule contains ≤ 5 hydrogen bond donors, ≤ 10 hydrogen bond acceptors, a molecular weight of ≤ 500, a lipophilicity factor (cLogP) ≤ 5. Combinatorial chemistry or parallel synthesis approach are used to synthesize a large number of analogs of molecules of a given structural scaffold. Being constrained by the structural scaffold, the synthetically derived molecules can cover a limited chemical space. Natural products, on the other hand, are secondary metabolites biosynthesized in organisms like bacteria, fungi, higher invertebrates like marine sponges and plants through pathways that have evolved over millions of years. Due to this evolutionary advantage, natural products comprise structural scaffolds and chemical properties that are unparalleled in the field of

217

synthetic chemistry. Further, since natural products are biosynthesized by organisms to communicate with their immediate environments, they are naturally adept at penetrating through cell membranes and cell walls. The feature of being able to pass through membrane barriers makes the natural products desirable because such features are part of a 'drug-like' molecule. In addition, many secondary metabolites are biosynthesized by microorganisms and plants as interspecies defense mechanisms where the molecules are secreted out to protect the species against competing microbes in the immediate environment. The presence of natural antimicrobial properties makes natural products desirable candidates to be included in a screening library of compounds that is used for testing the whole-cell bioactivity against pathogenic microorganisms. Thus, natural products have advantages over synthetic molecules in the application of whole-cell inhibition studies.

Protein targets from two important pathways in *Mtb* were selected for the structure-based inhibition studies. These were the *Mtb* s-adenosylhomocysteine hydrolase (SAHH) in the methylation pathway and the *Mtb* PII protein in the nitrogen regulatory pathway of *Mtb*. The s-adenosylmethionine (SAM) based methylation pathway serves as an important metabolic pathway in all organisms including *Mtb* because methylation plays a central role in crucial intracellular processes like DNA methylation, amino acid biosynthesis, protein post translational modifications, phospholipid biosynthesis and intercellular processes like quorum sensing. The methyltransferase enzymes commonly use SAM as a donor of methyl group to methylate acceptor molecules like DNA, histones, enzymes, lipids and secondary metabolites. The

reactions are referred to as transmethylation reactions. After the transfer of a methyl group SAM is converted to s-adenosyl homocysteine (SAH). SAH is a potent feedback inhibitor of the methyltransferases and needs to be cleared from the cytoplasm in order to facilitate unhindered transmethylation reactions.

SAHH is a ubiquitous enzyme that assists transmethylation by breaking down the product of transmethylation reactions, SAH into adenosine and homocysteine (Hcy). SAHH also plays important part in other pathway like purine salvage and amino acid biosynthesis by generating the substrates for these pathways. Adenosine is processed in the purine salvage pathway by other enzymes like adenosine deaminase and adenosine kinases to generate adenosine triphosphate (ATP). Hcy is used in the biosynthesis of cysteine and the regeneration of methionine. *Mtb* SAHH was targeted for inhibition because of the essentiality of the enzyme to *Mtb* as observed by *in vitro* studies [45]. Human genome also encodes for SAHH and the amino acid sequence of human SAHH protein is 55 % similar to the *Mtb* SAHH amino acid sequence. Due to the high homology of the two proteins, it was essential to structurally characterize the *Mtb* SAHH and identify differences in the active site of the enzyme from the two species that could be used to design specific inhibitors of the *Mtb* SAHH. *Mtb* SAHH was crystallized in the presence of the product adenosine and the crystal structure solved by x-ray diffraction method to 1.6 Å resolution. Crystal structures of *Mtb* SAHH were also solved in the presence of known SAHH inhibitors like aristeromycin (ARI), deazaadenosine (DZA) and 2-fluoroadenosine (2FA) to 2.1 Å, 2.2 Å and 2.4 Å resolutions respectively. The crystal structure of *Mtb* SAHH in complex with ARI was the first one to be solved

and confirmed the proposed mode of inhibition for ARI. ARI was observed in the active site of the crystallized *Mtb* SAHH in the form of 3' keto aristeromycin. ARI is a mechanism based inhibitor of *Mtb* SAHH (0.2 µM) where it initiates the mechanism of hydrolysis similar to adenosine but gets trapped at the formation of the 3' keto intermediate. A less acidic proton at the 4' position of ARI compared to adenosine prevents the proton abstraction step and halts the further mechanism.

When compared to the active site of the human SAHH no significant differences were observed in the active site of *Mtb* SAHH. However, upon comparing with the active site of *Pf* SAHH a curious observation was made. The homologous histidine (His) residue in *Pf* SAHH corresponding to the His 363 of *Mtb* SAHH was flipped up by 180 °. The imidazole side chain of His 363 is proximal to the 5' hydroxyl group of adenosine at 2.69 Å. It was inferred that movement of the imidazole side chain of His 363 away from the 5' hydroxyl group of adenosine may be required for the binding of the Hcy portion of the full substrate, SAH, which is attached at the 5' position of adenosine. In order to understand the binding mode of the full substrate, SAH, *Mtb* SAHH was cocrystallized in the presence of SAH. While density for the complete SAH molecule was not observed in the binding site, SAH was modeled as ethylthioadenosine in the active site of *Mtb* SAHH. The binding mode of SAH in *Mtb* SAHH confirmed the hypothesis and for the first time showed that His 363 residue flips up and acts as a gate to allow the binding of the substrate, SAH. More importantly, crystal structure of *Mtb* SAHH revealed the architecture of the extended active site for binding SAH and also revealed an access channel that lead from the surface of the protein to the active site.

When comparisons were made with the human SAHH protein, differences could be identified between the two enzymes. Whereas in the human SAHH protein, the access channel was wide and funnel-like, the presence of a thirty six residue insertion in *Mtb* SAHH converted the access channel into a narrow path lined by hydrophobic residues. This difference between the human and *Mtb* SAHH could be used to design inhibitors that could bind in the narrow access way and block the function of *Mtb* SAHH. A fragment-based approach was adopted to identify small molecule inhibitors of the *Mtb* SAHH. A 750 compound fragment library was used to screen against *Mtb* SAHH in an assay using differential scanning fluorimetry technique. The technique uses a real time polymerase chain reaction (RT-PCR) to gradually increase the temperature of the assay mixture containing *Mtb* SAHH and a SYPRO orange dye in buffer. As the enzyme begins to unfold the dye interacts with the hydrophobic core of the enzyme causing it to fluoresce. The temperature at which the change in fluorescence emission is the highest is recorded as the melting temperature of the enzyme. Binding of a ligand to the enzyme causes changes in the stability of the enzyme leading to a change in the melting point temperature. Since the aim of the experiment was to find fragments that could bind in the narrow hydrophobic access channel of the enzyme, the assay was performed in the presence of bound adenosine and NAD^+, thus occupying the enzyme active sites. Through screening of the fragment libraries eight potential small molecule fragments were identified. Attempts to crystallize *Mtb* SAHH in the presence of the binding fragments were unsuccessful. Saturation transfer difference nuclear magnetic resonance (STD-NMR) spectrometry was used to validate the inhibitors. All the fragments

identified and validated in this study consist of an aromatic ring functionalized with a polar substituent like an amide, chloride, oxime or trifluoromethyl group with the exception of 1-(pyridin-3-yl)-2-p-tolylethanone. In the absence of crystal structures it may be speculated that the fragments take advantage of the hydrophobic nature of the access channel to form hydrophobic interactions involving the aromatic rings while the polar substituents on the aromatic ring may form hydrogen bonding interactions with residues like Glu186 which protrudes into the access channel and was identified as unique to the *Mtb* SAHH. Further studies on the design of inhibitors specific to the *Mtb* SAHH can be conducted by tethering the small molecules identified through fragment based screening and STD NMR to ADO and performing *in vitro* assays with the human and *Mtb* SAHH. Ligands that exploit the unique features offered by the narrow access channel would be expected to be selective towards the *Mtb* SAHH.

PII is a family of signal transduction proteins that act as nitrogen sensors in microorganisms and plants. The role of signal transduction is achieved by PII through complex formation with other proteins like ammonium transporter protein complex (AmtB), glycohydrolase (DraG), DraT and glutamine synthetase (GS) involved in the nitrogen regulation pathway. *Mtb* nitrogen regulatory PII protein was selected as a potential drug target because of the many role played by the protein in the direct and indirect regulation of intracellular nitrogen levels. The PII protein is functional through post translational modification of GS and through regulation of transcription of the gene *glnA* which encodes for GS. A functionally important T-loop on the PII protein consisting of 20 residues encompassing Tyr36 to Phe55 in *Mtb* PII protein is implicated

in the protein-protein complex formations. Factors influencing the orientation of the T-loop on the PII protein to facilitate the complex formation remain unknown.

Crystal structures of *Mtb* PII protein were solved in the apo form and in complex with ATP to 1.4 Å and 2.4 Å resolutions respectively. Density for the functionally important T-loop was missing in the apo structure while at least one of the subunits in the ATP bound PII contained a completely ordered T-loop. Comparisons with the *E. coli* and *M. jannaschii* PII proteins revealed that the structural features of *Mtb* PII are well conserved across species. The binding mode of ATP molecule in pocket formed by the interaction of two neighboring subunits in the functional trimer was also consisten with that observed in *E. coli* PII protein. Upon analysis it was apparent that the interaction of the residues at the base of the T-loop, particularly Arg38 with the ATP molecule was partially responsible for orienting the loop region towards target proteins for protein-protein complex formation. The presence of Arg38 in the position of Mg^{2+} in *M. jannaschii* PII protein and the interaction of side chain of Arg38 with the γ phosphate group of ATP also suggested the possibility that *Mtb* PII protein may not require Mg^{2+} ion for ATP binding. This observation was consistent with *in vitro* assay studies that showed that *Mtb* PII protein binds to ATP (K_d = 1.93 μM) and 2-oxoglutarate (2OG) in the absence of Mg^{2+} ions. Stabilization of the T-loop in subunit B of *Mtb* PII protein through crystal packing effect mimicked the formation of a protein-protein complex formation and suggested that while the base of the T-loop was partially stabilized by the ATP binding, additional stabilizing interactions mediated through the binding of a protein target were required to stabilize the T-loop. A homology model of the *Mtb* AmtB

and *Mtb* PII proteins based on the *E. coli* GlnK: AmtB complex showed that *Mtb* PII protein would have favorable interactions and would form a complex with *Mtb* AmtB protein. Structural studies of the *Mtb* PII protein revealed important insights into the function of the protein and its mechanism of action. Based on structural features like the presence of a 3_{10} helix at the C-terminal, the presence of Leu3 and Thr5 residues in the central cavity of the β-barrel formed by the trimeric assembly of *Mtb* PII protein and the operon organization of *Mtb PII* gene in between *amtB* and *glnD* genes suggests that the *Mtb PII* gene should translate into a GlnK protein. The biophysical and biochemical knowledge gained about *Mtb* PII protein will be useful in designing specific and selective inhibitors of the protein. Inhibition of *Mtb* PII protein can potentially disrupt the nitrogen regulation pathway in *Mtb* leading to cell death.

Chapter IV dealt with the whole-cell inhibition based approach to find inhibitors of *Mtb*. Natural products based drug discovery has been steadily declining in the recent years for reasons ranging from the time consuming processes of extraction, isolation and identification of new natural products to high costs. Big pharmaceutical companies have dropped their natural products discovery pipelines and shifted focus to other areas like combinatorial approaches to synthetic chemistry and genomics. A declining trend in the number of approved drugs introduced into the market is a clear testament that the alternative approaches to drug discovery are not working as expected in spite of the large amounts of investments being made. This trend holds true for tuberculosis therapy as no new drugs have been introduced against TB in the last forty years and a majority of the approved antitubercular drugs are natural products. In this research a whole-cell

inhibition based as well as a target-based *in vitro* enzyme assay approach was used to screen a collection of natural product extracts. Different available libraries of natural product extracts were screened against *M. smegmatis* and *Mtb* cells in an attempt to identify the most potent hits. The bioactivity based HPLC fractionation and isolation was performed with the collection of ninety six natural products extracts obtained from Vicuron and twenty nine natural products extracts obtained from Sarawak Biodiversity center (SBC). Whole-cell bioactivity data, HPLC UV profiles and LC / HRMS data were used to search for fractions with whole-cell or *in vitro* bioactivity. As is the case in any high-throughput screening experiment, a vast amount of data is generated and methods of data reduction to screen out the irrelevant data are necessary. LC / MS analysis resulted in a large amount of mass spectral data and methods were needed to narrow down to the promising MS peaks of interest that could lead to the identification of the natural product that was the cause of the bioactivity. A commercially available suite of software package called Metabolite Tools® developed by Bruker Daltonics Inc. was used to subtract LC / MS data of the inactive fractions from those of the bioactive fractions. Once the peak listings were narrowed down to the most interesting MS peaks, SmartFormula® prediction calculator, also developed by Bruker Daltonics Inc. was employed to generate molecular formula predictions using the high resolution MS data. A dereplication process was performed on the generated molecular formula and the associated HRMS data to identify the masses that corresponded to natural products that already existed in the available dictionary of natural products. The extract number 67037 from SBC showed consistent anti-mycobacterial activity against *Mtb* and *M. smegmatis*

through multiple rounds of fractionation and isolation. The bioactive fraction was purified by HPLC to isolate the source of the bioactivity. Through the application of LC / HRMS, tandem MS and NMR spectroscopy a compound with a fatty acid like core structure was identified. The molecule was named 2-methyl-2-[(4-oxoheptadecanoyl)amino]propanoic acid. No matches were observed for the chemical structure of the fatty acid against the dictionary of natural products. Therefore the identified compound is most likely a novel source of bioactivity against *Mtb* and *M. smegmatis* cells. Fatty acids have previously been used in the inhibition of fatty acid synthesis, cholesterol biosynthesis and in the inhibition of cancer cells [199,200]. The whole-cell inhibition of *Mtb* by saturated fatty acids like palmitic acid and stearic acid at MIC ~ 5 μM was an interesting find in this regard as it indicated that other long chain fully saturated fatty acids may also be able to inhibit the *Mtb* whole-cells. Palmitic acid and stearic acid have 16 and 18 carbon atoms long saturated aliphatic chains respectively. The identified natural product contains a saturated aliphatic chain that is 20 carbon atoms long. Based on the fatty acid like core structure it may be speculated that the natural product could be an inhibitor of enzymes involved in the fatty acid biosynthesis in *Mtb*.

The collections of natural product extracts were used in an *in vitro* enzyme assay against known drug targets in *Mtb* and human. A six hundred and forty microbial crude extract library that was previously determined to exhibit whole-cell bioactivity against *Mtb* cells was also directed in a target-based high-throughput *in vitro* assay to identify natural product inhibitors of *Mtb* malate synthase. *Mtb* malate synthase functions in the

226

glyoxylate shunt pathway. It is implicated as a virulence factor and represents an attractive drug target in *Mtb*. Screening of the extract library yielded three extracts that displayed reproducible *in vitro* inhibitory activity. The extracts were fractionated on HPLC to separate the components of the extract into forty fractions per extract. The individual fractions were then tested for *in vitro* inhibitory activity against *Mtb* malate synthase. The active fractions were then tested in serial dilutions of 2 X to 0.02 X concentrations in the assay. Five fractions in total originating from three different extracts were determined to exhibit *in vitro* inhibitory activity at as low as 0.03 X concentrations. The absence of structural information on the *in vitro* active compounds in the extracts precluded the accurate determination of an IC_{50} value. Efforts are underway to structurally characterize the natural products in the bioactive microbial extracts.

The bioactivity guided fractionation coupled with *in vitro* enzyme assay was used to identify a novel natural product inhibitor of the human proteasome. HPLC purification followed by high resolution mass spec data and NMR spectroscopy guided structure elucidation resulted in the identification a polyphenolic compound – 2,2'-di(3-hydroxy-5-methyl phenoxy)-6,6'-dimethyl biphenyl-4,4'-diol – that showed inhibitory activity against the human and yeast proteasome in an *in vitro* assay. The natural product had a modest IC_{50} value of 15 μM against the human proteasome. The natural product was active against the chymotrypsin-like as well as the caspase-like activities associated with the proteasome. The molecular structure of the discovered polyphenol is unlike any other previously investigated synthetic or natural compound inhibitors of the proteasome. The

binding mode and structure-activity relationship of the polyphenol can be better understood by crystallizing and solving the crystal structure of the polyphenol in complex with the human or yeast proteasome. The symmetric polyphenolic structure of the natural product affords an opportunity to create functional modifications to improve the binding affinity and therefore increase the inhibitory effect on the proteasome and hence represents a lead in the search for anticancer drugs.

In conclusion it may be stated that natural products research holds great potential in finding novel therapeutics against not only TB but also many diseases like cancer and AIDS. While the 'low hanging' fruits like streptomycin, erythromycin and vancomycin were discovered during the golden years of natural products drug discovery, many as yet undetected biosynthetic pathways exist in microorganisms, which produce novel natural products that are waiting to be discovered.

REFERENCES

1   Koehler, C. W. Consumption, the great killer. *Modern Drug Discovery* **5**, 47-49 (2002).

2   Bunyan, J. *The Life and Death of Mr. Badman*. (R. H. Russell, 1900).

3   Zink, A. R. *et al.* Characterization of *Mycobacterium tuberculosis* complex DNAs from Egyptian mummies by spoligotyping. *J Clin Microbiol* **41**, 359-367 (2003).

4   Myers, J. A. *Captain of All These Men of Death; Tuberculosis Historical Highlights*. (Warren H Green, 1977).

5   Krause, A. K. *Tuberculosis and Public Health; YA Pamphlet Collection (Library of Congress).* (Hermann M. Biggs memorial tuberculosis lecture, 1927).

6   Williams, C. J. B. On the use and administration of cod-liver oil in pulmonary consumption. *London Journal of Medicine* **1**, 1-18 (1849).

7   Baron John Boyd Orr Boyd-Orr, J. L. G. *Studies of Nutrition: the Physique and Health of Two African Tribes*. (H. M. Stationery Off., 1931).

8   Bull, P. Extrapleural thoracoplasty in the treatment of pulmonary tuberculosis, With an account of 37 cases. *Lancet* **2**, 778-783 (1920).

9   Barnes, J. Artificial pneumoperitoneum - In pulmonary tuberculosis and pregnancy. *Lancet* **2**, 976-977 (1939).

10   Feldman, W. H., Karlson, A. G. & Hinshaw, H. C. Promin in experimental tuberculosis; antituberculosis effects of sodium P,P'-diaminodiphenylsulfone-N,N'-didextroxe sulfonate (Promin) administered subcutaneously (a preliminary report). *Proceedings of the Staff Meetings. Mayo Clinic* **23**, 118-125 (1948).

11   William Steenken, J., Fred H. Heise. Action of Promin and Diamino-Diphenyl Sulfone upon Tubercle Bacilli; Antipromin Action of p-Aminobenzoic Acid. *Proceedings of the Society for Experimental Biology and Medicine*, 180-183, doi:10.3181/00379727-52-14077 (1943).

12   Feldman, W. H., Hinshaw, H.C., Mann, F.C. Promizole in tuberculosis. *Am Rev Tuberc*, 418-440 (1944).

13   Schatz, A., Bugie, E. & Waksman, S. A. Streptomycin, a substance exhibiting antibiotic activity against gram positive and gram-negative bacteria. *P Soc Exp Biol Med* **55**, 66-69 (1944).

14   Hobby, G. L., Lenert, T. F., Donikian, M. & Pikula, D. The tuberculostatic activity of terramycin. *American Review of Tuberculosis* **63**, 434-440 (1951).

15   Finlay, A. C. *et al.* Viomycin a new antibiotic active against Mycobacteria. *American Review of Tuberculosis* **63**, 1-3 (1951).

16   Epstein, I. G., Nair, K. G. & Boyd, L. J. Cycloserine, a new antibiotic, in the treatment of human pulmonary tuberculosis: a preliminary report. *Antibiotic Medicine & Clinical Therapy* **1**, 80-93 (1955).

17   Lehmann, J. Para-aminosalicylic acid in the treatment of tuberculosis. *Lancet* **1**, 15 (1946).

18   Benson, W. M., Stefko, P. L. & Roe, M. D. Pharmacologic and toxicologic observations on hydrazine derivatives of isonicotinic acid (rimifon, marsilid). *American Review of Tuberculosis* **65**, 376-391 (1952).

19   Bernstein, J., Lott, W. A., Steinberg, B. A. & Yale, H. L. Chemotherapy of experimental tuberculosis. V. Isonicotinic acid hydrazide (nydrazid) and related compounds. *American Review of Tuberculosis* **65**, 357-364 (1952).

20   D'Esopo, N. D. in *Tr VA-Armed Forces Conf on Chemotherap of Tuberc* 57-64.

21   Tempel, C. W., Hughes, F. J., Jr., Mardis, R. E., Towbin, M. N. & Dye, W. E. Combined intermittent regimens employing streptomycin and para-aminosalicylic acid in the treatment of pulmonary tuberculosis; a comparison with daily and intermittent dosage schedules. *American Review of Tuberculosis* **63**, 295-311 (1951).

22   Goulding, C. W. *et al.* The TB structural genomics consortium: providing a structural foundation for drug discovery. *Curr Drug Targets Infect Disord* **2**, 121-141 (2002).

23   Camus, J. C., Pryor, M. J., Medigue, C. & Cole, S. T. Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology* **148**, 2967-2973 (2002).

24   Musa, T. L., Ioerger, T. R. & Sacchettini, J. C. The tuberculosis structural genomics consortium: a structural genomics approach to drug discovery.

Advances in Protein Chemistry and Structural Biology **77**, 41-76, doi:10.1016/S1876-1623(09)77003-8 (2009).

25   Henderson, R. in *Harward Business Review* 100-105 (Harvard Business Publishing, 1994).

26   Amzel, L. M. Structure-based drug design. *Current Opinion in Biotechnology* **9**, 366-369 (1998).

27   Meyer, E. F. The first years of the Protein Data Bank. *Protein Sci* **6**, 1591-1597, doi:10.1002/pro.5560060724 (1997).

28   Kusunoki, M. [The protein data bank: recent trends and developments]. *Tanpakushitsu kakusan koso. Protein, Nucleic Acid, Enzyme* **42**, 1315-1319 (1997).

29   Andricopulo, A. D., Salum, L. B. & Abraham, D. J. Structure-based drug design strategies in medicinal chemistry. *Current Topics in Medicinal Chemistry* **9**, 771-790 (2009).

30   Cole, S. T. *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537-544 (1998).

31   Deshpande, N. *et al.* The RCSB Protein Data Bank: a redesigned query system and relational database based on the mmCIF schema. *Nucleic Acids Res* **33**, D233-237, doi:10.1093/nar/gki057 (2005).

32   Chiang, P. K. *et al.* S-Adenosylmethionine and methylation. *Faseb J* **10**, 471-480 (1996).

33   Fontecave, M., Atta, M. & Mulliez, E. S-adenosylmethionine: nothing goes to waste. *Trends in Biochemical Sciences* **29**, 243-249, doi:10.1016/j.tibs.2004.03.007 (2004).

34   Schauder, S., Shokat, K., Surette, M. G. & Bassler, B. L. The LuxS family of bacterial autoinducers: biosynthesis of a novel quorum-sensing signal molecule. *Mol Microbiol* **41**, 463-476 (2001).

35   Tehlivets, O., Hasslacher, M. & Kohlwein, S. D. S-adenosyl-L-homocysteine hydrolase in yeast: key enzyme of methylation metabolism and coordinated regulation with phospholipid synthesis. *FEBS Lett* **577**, 501-506 (2004).

36   Kramer, D. L., Porter, C. W., Borchardt, R. T. & Sufrin, J. R. Combined modulation of S-adenosylmethionine biosynthesis and S-adenosylhomocysteine

metabolism enhances inhibition of nucleic acid methylation and L1210 cell growth. *Cancer Res* **50**, 3838-3842 (1990).

37   Zulty, J. J. & Speedie, M. K. Purification and characterization of S-adenosylhomocysteine deaminase from streptonigrin-producing *Streptomyces flocculus*. *J Bacteriol* **171**, 6840-6844 (1989).

38   Worm, J. & Guldberg, P. DNA methylation: an epigenetic pathway to cancer and a promising target for anticancer therapy. *Journal of Oral Pathology & Medicine : Official Publication of the International Association of Oral Pathologists and the American Academy of Oral Pathology* **31**, 443-449 (2002).

39   Issa, J. P. & Kantarjian, H. M. Targeting DNA methylation. *Clinical Cancer Research : an Official Journal of the American Association for Cancer Research* **15**, 3938-3946, doi:10.1158/1078-0432.CCR-08-2783 (2009).

40   Donaldson, T. & Kim, K. Targeting *Plasmodium Falciparum* Purine Salvage Enzymes: A Look At Structure-Based Drug Development. *Infectious Disorders Drug Targets* **9** (2010).

41   Stankova, J., Lawrance, A. K. & Rozen, R. Methylenetetrahydrofolate reductase (MTHFR): a novel target for cancer therapy. *Curr Pharm Des* **14**, 1143-1150 (2008).

42   De Clercq, E., Cools, M. & Balzarini, J. Homocysteine potentiates the antiviral and cytostatic activity of those nucleoside analogues that are targeted at S-adenosylhomocysteine hydrolase. *Biochem Pharmacol* **38**, 1771-1778 (1989).

43   Malinow, M. R. Plasma homocyst(e)ine and arterial occlusive diseases: a mini-review. *Clinical Chemistry* **41**, 173-176 (1995).

44   Miller, M. W. *et al.* The mouse lethal nonagouti (a(x)) mutation deletes the S-adenosylhomocysteine hydrolase (Ahcy) gene. *Embo J* **13**, 1806-1816 (1994).

45   Sassetti, C. M., Boyd, D. H. & Rubin, E. J. Comprehensive identification of conditionally essential genes in mycobacteria. *Proc Natl Acad Sci U S A* **98**, 12712-12717 (2001).

46   Merrick, M. J. & Edwards, R. A. Nitrogen control in bacteria. *Microbiol Rev* **59**, 604-622 (1995).

47   Engleman, E. G. & Francis, S. H. Cascade control of *E. coli* glutamine synthetase. II. Metabolite regulation of the enzymes in the cascade. *Arch Biochem Biophys* **191**, 602-612 (1978).

48   Kamberov, E. S., Atkinson, M. R. & Ninfa, A. J. The *Escherichia coli* PII signal transduction protein is activated upon binding 2-ketoglutarate and ATP. *J Biol Chem* **270**, 17797-17807 (1995).

49   Ninfa, A. J. & Atkinson, M. R. PII signal transduction proteins. *Trends Microbiol.* **8**, 172-179 (2000).

50   Arcondeguy, T., Jack, R. & Merrick, M. P(II) signal transduction proteins, pivotal players in microbial nitrogen control. *Microbiol Mol Biol Rev* **65**, 80-105 (2001).

51   Jiang, P., Peliska, J. A. & Ninfa, A. J. Enzymological characterization of the signal-transducing uridylyltransferase/uridylyl-removing enzyme (EC 2.7.7.59) of *Escherichia coli* and its interaction with the PII protein. *Biochemistry* **37**, 12782-12794 (1998).

52   Leigh, J. A. & Dodsworth, J. A. Nitrogen regulation in bacteria and archaea. *Annu Rev Microbiol* **61**, 349-377 (2007).

53   Stadtman, E. R. The story of glutamine synthetase regulation. *J Biol Chem* **276**, 44357-44364 (2001).

54   Stock, A. M., Robinson, V. L. & Goudreau, P. N. Two-component signal transduction. *Annu Rev Biochem* **69**, 183-215 (2000).

55   Atkinson, M. R. & Ninfa, A. J. Role of the GlnK signal transduction protein in the regulation of nitrogen assimilation in *Escherichia coli*. *Mol Microbiol* **29**, 431-447 (1998).

56   Javelle, A. & Merrick, M. Complex formation between AmtB and GlnK: an ancestral role in prokaryotic nitrogen control. *Biochem Soc Trans* **33**, 170-172 (2005).

57   Liu, Y. & Hu, X. Molecular determinants for binding of ammonium ion in the ammonia transporter AmtB-A quantum chemical analysis. *J Phys Chem A* **110**, 1375-1381 (2006).

58   Khademi, S. *et al.* Mechanism of ammonia transport by Amt/MEP/Rh: structure of AmtB at 1.35 A. *Science* **305**, 1587-1594 (2004).

59   Durand, A. & Merrick, M. *In vitro* analysis of the *Escherichia coli* AmtB-GlnK complex reveals a stoichiometric interaction and sensitivity to ATP and 2-oxoglutarate. *J Biol Chem* **281**, 29558-29567 (2006).

60   Harper, C., Hayward, D., Wiid, I. & van Helden, P. Regulation of nitrogen metabolism in *Mycobacterium tuberculosis*: A comparison with mechanisms in *Corynebacterium glutamicum* and *Streptomyces coelicolor*. *IUBMB Life* (2008).

61   Read, R., Pashley, C. A., Smith, D. & Parish, T. The role of GlnD in ammonia assimilation in *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)* **87**, 384-390 (2007).

62   Dolle, R. E. Comprehensive survey of combinatorial library synthesis: 1999. *J Comb Chem* **2**, 383-433 (2000).

63   Newman, D. J., Cragg, G. M. & Snader, K. M. Natural products as sources of new drugs over the period 1981-2002. *Journal of Natural Products* **66**, 1022-1037, doi:10.1021/np030096l (2003).

64   Newman, D. J. & Cragg, G. M. Natural products as sources of new drugs over the last 25 years. *Journal of Natural Products* **70**, 461-477, doi:10.1021/np068054v (2007).

65   Reddy, B. V. *et al.* Natural product biosynthetic gene diversity in geographically distinct soil microbiomes. *Appl Environ Microbiol* **78**, 3744-3752, doi:10.1128/AEM.00102-12 (2012).

66   Haefner, B. Drugs from the deep: marine natural products as drug candidates. *Drug Discov Today* **8**, 536-544 (2003).

67   Newman, D. J. & Cragg, G. M. Advanced preclinical and clinical trials of natural products and related compounds from marine sources. *Curr Med Chem* **11**, 1693-1713 (2004).

68   Kong, D. X., Jiang, Y. Y. & Zhang, H. Y. Marine natural products as sources of novel scaffolds: achievement and concern. *Drug Discov Today* **15**, 884-886, doi:10.1016/j.drudis.2010.09.002 (2010).

69   Lee, K. H. Anticancer drug design based on plant-derived natural products. *Journal of Biomedical Science* **6**, 236-250 (1999).

70   Batista, R., Silva Ade, J., Jr. & de Oliveira, A. B. Plant-derived antimalarial agents: new leads and efficient phytomedicines. Part II. Non-alkaloidal natural products. *Molecules* **14**, 3037-3072, doi:10.3390/molecules14083037 (2009).

71   Handbook of anti-tuberculosis agents. Introduction. *Tuberculosis (Edinb)* **88**, 85-86, doi:10.1016/S1472-9792(08)70002-7 (2008).

72   Comas, I. *et al.* Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nature Genetics* **44**, 106-110, doi:10.1038/ng.1038 (2012).

73   Lalloo, U. G. & Ambaram, A. New antituberculous drugs in development. *Current HIV/AIDS Reports* **7**, 143-151, doi:10.1007/s11904-010-0054-4 (2010).

74   Koul, A., Arnoult, E., Lounis, N., Guillemont, J. & Andries, K. The challenge of new drug discovery for tuberculosis. *Nature* **469**, 483-490, doi:10.1038/nature09657 (2011).

75   Salemme, F. R., Spurlino, J. & Bone, R. Serendipity meets precision: the integration of structure-based drug design and combinatorial chemistry for efficient drug discovery. *Structure* **5**, 319-324 (1997).

76   Palmer, J. L. & Abeles, R. H. The mechanism of action of S-adenosylhomocysteinase. *J Biol Chem* **254**, 1217-1226 (1979).

77   Yamada, T. *et al.* Structure and function of eritadenine and its 3-deaza analogues: potent inhibitors of S-adenosylhomocysteine hydrolase and hypocholesterolemic agents. *Biochem Pharmacol* **73**, 981-989 (2007).

78   Bujnicki, J. M., Prigge, S. T., Caridha, D. & Chiang, P. K. Structure, evolution, and inhibitor interaction of S-adenosyl-L-homocysteine hydrolase from *Plasmodium falciparum*. *Proteins* **52**, 624-632 (2003).

79   Nakajima-Shimada, J., Hirota, Y. & Aoki, T. Inhibition of Trypanosoma cruzi growth in mammalian cells by purine and pyrimidine analogs. *Antimicrob Agents Chemother* **40**, 2455-2458 (1996).

80   Ctrnacta, V. *et al.* Efficacy of S-adenosylhomocysteine hydrolase inhibitors, D-eritadenine and (S)-DHPA, against the growth of *Cryptosporidium parvum in vitro*. *Exp Parasitol* **126**, 113-116, doi:10.1016/j.exppara.2010.04.007 (2010).

81   Yaginuma, S. *et al.* Studies on neplanocin A, new antitumor antibiotic. I. Producing organism, isolation and characterization. *J Antibiot (Tokyo)* **34**, 359-366 (1981).

82   Betts, J. C., Lukey, P. T., Robb, L. C., McAdam, R. A. & Duncan, K. Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. *Mol Microbiol* **43**, 717-731 (2002).

83   Schnappinger, D. *et al.* Transcriptional Adaptation of *Mycobacterium tuberculosis* within Macrophages: Insights into the Phagosomal Environment. *J Exp Med* **198**, 693-704 (2003).

84   Dubnau, E., Chan, J., Mohan, V. P. & Smith, I. responses of *Mycobacterium Tuberculosis* to growth in the mouse lung. *Infect Immun* **73**, 3754-3757 (2005).

85   Wolfe, M. S. & Borchardt, R. T. S-adenosyl-L-homocysteine hydrolase as a target for antiviral chemotherapy. *J Med Chem* **34**, 1521-1530 (1991).

86   Borchardt, R. T., Keller, B. T. & Patel-Thombre, U. Neplanocin A. A potent inhibitor of S-adenosylhomocysteine hydrolase and of vaccinia virus multiplication in mouse L929 cells. *J Biol Chem* **259**, 4353-4358 (1984).

87   Guranowski, A., Montgomery, J. A., Cantoni, G. L. & Chiang, P. K. Adenosine analogues as substrates and inhibitors of S-adenosylhomocysteine hydrolase. *Biochemistry* **20**, 110-115 (1981).

88   De Clercq, E. *et al.* Broad-spectrum antiviral activities of neplanocin A, 3-deazaneplanocin A, and their 5'-nor derivatives. *Antimicrob Agents Chemother* **33**, 1291-1297 (1989).

89   Turner, M. A. *et al.* Structure determination of selenomethionyl S-adenosylhomocysteine hydrolase using data at a single wavelength. *Nat Struct Biol* **5**, 369-376 (1998).

90   Mosley, S. L. *et al.* Carbocyclic pyrimidine nucleosides as inhibitors of S-adenosylhomocysteine hydrolase. *Bioorg Med Chem* **14**, 7967-7971 (2006).

91   Yang, X. *et al.* Catalytic strategy of S-adenosyl-L-homocysteine hydrolase: transition-state stabilization and the avoidance of abortive reactions. *Biochemistry* **42**, 1900-1909 (2003).

92   Liu, S., Yuan, C. S. & Borchardt, R. T. Aristeromycin-5'-carboxaldehyde: a potent inhibitor of S-adenosyl-L-homocysteine hydrolase. *J Med Chem* **39**, 2347-2353 (1996).

93   Huang, Y. *et al.* Inhibition of S-adenosylhomocysteine hydrolase by acyclic sugar adenosine analogue D-eritadenine. Crystal structure of S-adenosylhomocysteine hydrolase complexed with D-eritadenine. *J Biol Chem* **277**, 7477-7482 (2002).

94   Hu, Y. *et al.* Crystal structure of S-adenosylhomocysteine hydrolase from rat liver. *Biochemistry* **38**, 8323-8333 (1999).

95   Komoto, J. *et al.* Effects of site-directed mutagenesis on structure and function of recombinant rat liver S-adenosylhomocysteine hydrolase. Crystal structure of D244E mutant enzyme. *J Biol Chem* **275**, 32147-32156 (2000).

96   Takata, Y. *et al.* Catalytic mechanism of S-adenosylhomocysteine hydrolase. Site-directed mutagenesis of Asp-130, Lys-185, Asp-189, and Asn-190. *J Biol Chem* **277**, 22670-22676 (2002).

97   Tanaka, N. *et al.* Crystal structure of S-adenosyl-L-homocysteine hydrolase from the human malaria parasite Plasmodium falciparum. *J Mol Biol* **343**, 1007-1017 (2004).

98   Yamada T, T. Y., Komoto J, Gomi T, Ogawa H, Fujioka M, Takusagawa F. Catalytic mechanism of S-adenosylhomocysteine hydrolase: roles of His 54, Asp130, Glu155, Lys185, and Aspl89. *International Journal of Biochemistry and Cell Biology* **37**, 2417 - 2435 (2005).

99   Gomi, T., Takata, Y. & Fujioka, M. Rat liver S-adenosylhomocysteinase. Spectrophotometric study of coenzyme binding. *Biochim Biophys Acta* **994**, 172-179 (1989).

100   Borek, D., Minor, W. & Otwinowski, Z. Measurement errors and their consequences in protein crystallography. *Acta Crystallogr D Biol Crystallogr* **59**, 2031-2038 (2003).

101   Matthews, B. W. Solvent content of protein crystals. *J Mol Biol* **33**, 491-497 (1968).

102   Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-763 (1994).

103   Kantardjieff, K. A., Hochtl, P., Segelke, B. W., Tao, F. M. & Rupp, B. Concanavalin A in a dimeric crystal form: revisiting structural accuracy and molecular flexibility. *Acta Crystallogr D Biol Crystallogr* **58**, 735-743 (2002).

104   McRee, D. E. XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density. *J Struct Biol* **125**, 156-165 (1999).

105   Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004).

106   Pflugrath, J. W. The finer things in X-ray diffraction data collection. *Acta Crystallogr D Biol Crystallogr* **55**, 1718-1725 (1999).

107   Cruickshank, D. W. Remarks about protein structure precision. *Acta Crystallogr D Biol Crystallogr* **55**, 583-601 (1999).

108   Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-255 (1997).

109   Ellman, G. L., Courtney, K. D., Andres, V., Jr. & Feather-Stone, R. M. A new and rapid colorimetric determination of acetylcholinesterase activity. *Biochem Pharmacol* **7**, 88-95 (1961).

110   Lozada-Ramirez, J. D., Martinez-Martinez, I., Sanchez-Ferrer, A. & Garcia-Carmona, F. A colorimetric assay for S-adenosylhomocysteine hydrolase. *J Biochem Biophys Methods* **67**, 131-140 (2006).

111   Chaturvedi, V., Dwivedi, N., Tripathi, R. P. & Sinha, S. Evaluation of Mycobacterium smegmatis as a possible surrogate screen for selecting molecules active against multi-drug resistant *Mycobacterium tuberculosis*. *The Journal of General and Applied Microbiology* **53**, 333-337 (2007).

112   Snapper, S. B., Melton, R. E., Mustafa, S., Kieser, T. & Jacobs, W. R., Jr. Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Mol Microbiol* **4**, 1911-1919 (1990).

113   Pereira, D. A. & Williams, J. A. Origin and evolution of high throughput screening. *British Journal of Pharmacology* **152**, 53-61, doi:10.1038/sj.bjp.0707373 (2007).

114   Proudfoot, J. R. Drugs, leads, and drug-likeness: an analysis of some recently launched drugs. *Bioorg Med Chem Lett* **12**, 1647-1650 (2002).

115   Congreve, M., Carr, R., Murray, C. & Jhoti, H. A 'rule of three' for fragment-based lead discovery? *Drug Discov Today* **8**, 876-877 (2003).

116   Hajduk, P. J. Fragment-based drug design: how big is too big? *J Med Chem* **49**, 6972-6976 (2006).

117   Makara, G. M. On sampling of fragment space. *J Med Chem* **50**, 3214-3221 (2007).

118   Lesuisse, D. *et al.* SAR and X-ray. A new approach combining fragment-based screening and rational drug design: application to the discovery of nanomolar inhibitors of Src SH2. *J Med Chem* **45**, 2379-2387 (2002).

119 Erlanson, D. A., Wells, J. A. & Braisted, A. C. Tethering: fragment-based drug discovery. *Annual Review of Biophysics and Biomolecular Structure* **33**, 199-223, doi:10.1146/annurev.biophys.33.110502.140409 (2004).

120 Vedadi, M. *et al.* Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proc Natl Acad Sci U S A* **103**, 15835-15840 (2006).

121 Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* **2**, 2212-2221 (2007).

122 Mayer, M. & Meyer, B. Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *J Am Chem Soc* **123**, 6108-6117 (2001).

123 Porcelli, M. *et al.* S-adenosylhomocysteine hydrolase from the archaeon Pyrococcus furiosus: biochemical characterization and analysis of protein structure by comparative molecular modeling. *Proteins* **58**, 815-825 (2005).

124 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004).

125 Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **60**, 2256-2268 (2004).

126 Turner, M. A. *et al.* Structure and function of S-adenosylhomocysteine hydrolase. *Cell Biochem Biophys* **33**, 101-125 (2000).

127 Stepkowski, T., Brzezinski, K., Legocki, A. B., Jaskolski, M. & Bena, G. Bayesian phylogenetic analysis reveals two-domain topology of S-adenosylhomocysteine hydrolase protein sequences. *Mol Phylogenet Evol* **34**, 15-28 (2005).

128 Wang, R., Gao, Y. & Lai, L. Calculating partition coefficient by atom-additive method. *Perspectives in Drug Discovery and Design* **19**, 47-66 (2000).

129 Pittillo, R. F., Moncrief, C., Brockman, R. W. & Chambers, P. Antimicrobial Activity of 2-Fluoroadenosine and 2-Fluoroadenine. *Antimicrobial Agents Chemother (Bethesda)* **10**, 474-484 (1964).

130	Christopher, J. A. SPOCK: the Structural Properties Observation and Calculation Kit Program manual. *The Center for Macromolecular Design, Texas A&M University, College Station, TX.* (1998).

131	Jiang, P., Peliska, J. A. & Ninfa, A. J. The regulation of *Escherichia coli* glutamine synthetase revisited: role of 2-ketoglutarate in the regulation of glutamine synthetase adenylylation state. *Biochemistry* **37**, 12802-12810 (1998).

132	Ehlers, C., Weidenbach, K., Veit, K., Forchhammer, K. & Schmitz, R. A. Unique mechanistic features of post-translational regulation of glutamine synthetase activity in *Methanosarcina mazei* strain Go1 in response to nitrogen availability. *Mol Microbiol* **55**, 1841-1854 (2005).

133	Ninfa, A. J. & Magasanik, B. Covalent modification of the glnG product, NRI, by the glnL product, NRII, regulates the transcription of the glnALG operon in *Escherichia coli*. *Proc Natl Acad Sci U S A* **83**, 5909-5913 (1986).

134	Zhang, Y., Pohlmann, E. L. & Roberts, G. P. GlnD is essential for NifA activation, NtrB/NtrC-regulated gene expression, and posttranslational regulation of nitrogenase activity in the photosynthetic, nitrogen-fixing bacterium *Rhodospirillum rubrum*. *J Bacteriol* **187**, 1254-1265 (2005).

135	van Heeswijk, W. C. *et al.* An alternative PII protein in the regulation of glutamine synthetase in *Escherichia coli*. *Mol Microbiol* **21**, 133-146 (1996).

136	Thomas, G., Coutts, G. & Merrick, M. The glnKamtB operon. A conserved gene pair in prokaryotes. *Trends Genet* **16**, 11-14 (2000).

137	Dixon, R. & Kahn, D. Genetic regulation of biological nitrogen fixation. *Nat Rev Microbiol* **2**, 621-631 (2004).

138	Huergo, L. F. *et al.* Interactions between PII proteins and the nitrogenase regulatory enzymes DraT and DraG in *Azospirillum brasilense*. *FEBS Lett* **580**, 5232-5236 (2006).

139	Huergo, L. F. *et al.* Ternary complex formation between AmtB, GlnZ and the nitrogenase regulatory enzyme DraG reveals a novel facet of nitrogen regulation in bacteria. *Mol Microbiol* **66**, 1523-1535 (2007).

140	Heinrich, A. *et al.* Interaction of the membrane-bound GlnK-AmtB complex with the master regulator of nitrogen metabolism TnrA in Bacillus subtilis. *J Biol Chem* **281**, 34909-34917 (2006).

141   Jakoby, M., Nolden, L., Meier-Wagner, J., Kramer, R. & Burkovski, A. AmtR, a global repressor in the nitrogen regulation system of *Corynebacterium glutamicum*. *Mol Microbiol* **37**, 964-977 (2000).

142   Tiffert, Y. *et al.* The *Streptomyces coelicolor* GlnR regulon: identification of new GlnR targets and evidence for a central role of GlnR in nitrogen metabolism in actinomycetes. *Mol Microbiol* **67**, 861-880 (2008).

143   Wray, L. V., Jr., Atkinson, M. R. & Fisher, S. H. Identification and cloning of the glnR locus, which is required for transcription of the glnA gene in *Streptomyces coelicolor* A3(2). *J Bacteriol* **173**, 7351-7360 (1991).

144   Cheah, E. *et al.* Structure of the *Escherichia coli* signal transducing protein PII. *Structure* **2**, 981-990 (1994).

145   Xu, Y., Carr, P. D., Huber, T., Vasudevan, S. G. & Ollis, D. L. The structure of the PII-ATP complex. *Eur J Biochem* **268**, 2028-2037 (2001).

146   Xu, Y. *et al.* GlnK, a PII-homologue: structure reveals ATP binding site and indicates how the T-loops may be involved in molecular recognition. *J Mol Biol* **282**, 149-165 (1998).

147   Yildiz, O., Kalthoff, C., Raunser, S. & Kuhlbrandt, W. Structure of GlnK1 with bound effectors indicates regulatory mechanism for ammonia uptake. *Embo J* **26**, 589-599 (2007).

148   Sakai, H. *et al.* Crystal structures of the signal transducing protein GlnK from *Thermus thermophilus* HB8. *J Struct Biol* **149**, 99-110 (2005).

149   Machado Benelli, E. *et al. Herbaspirillum seropedicae* signal transduction protein PII is structurally similar to the enteric GlnK. *Eur J Biochem* **269**, 3296-3303 (2002).

150   Xu, Y. *et al.* The structures of the PII proteins from the cyanobacteria *Synechococcus sp. PCC 7942* and *Synechocystis sp. PCC 6803*. *Acta Crystallogr D Biol Crystallogr* **59**, 2183-2190 (2003).

151   Mizuno, Y., Moorhead, G. B. & Ng, K. K. Structural basis for the regulation of N-acetylglutamate kinase by PII in *Arabidopsis thaliana*. *J Biol Chem* **282**, 35733-35740, doi:M707127200 [pii]10.1074/jbc.M707127200 (2007).

152   Forchhammer, K., Hedler, A., Strobel, H. & Weiss, V. Heterotrimerization of PII-like signalling proteins: implications for PII-mediated signal transduction systems. *Mol Microbiol* **33**, 338-349 (1999).

153　van Heeswijk, W. C. *et al.* The *Escherichia coli* signal transducers PII (GlnB) and GlnK form heterotrimers *in vivo*: fine tuning the nitrogen signal cascade. *Proc Natl Acad Sci U S A* **97**, 3942-3947 (2000).

154　Rengarajan, J., Bloom, B. R. & Rubin, E. J. Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. *Proc Natl Acad Sci U S A* **102**, 8327-8332 (2005).

155　Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307 - 326 (1997).

156　McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D Biol Crystallogr* **63**, 32-41 (2007).

157　Vagin, A. A. *et al.* REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* **60**, 2184-2195 (2004).

158　Reddy, V. *et al.* Effective electron-density map improvement and structure validation on a Linux multi-CPU web cluster: The TB Structural Genomics Consortium Bias Removal Web Service. *Acta Crystallogr D Biol Crystallogr* **59**, 2200-2210 (2003).

159　Carr, P. D. *et al.* X-ray structure of the signal transduction protein from *Escherichia coli* at 1.9 A. *Acta Crystallogr D Biol Crystallogr* **52**, 93-104 (1996).

160　Murzin, A. G., Lesk, A. M. & Chothia, C. Principles determining the structure of beta-sheet barrels in proteins. II. The observed structures. *J Mol Biol* **236**, 1382-1400 (1994).

161　Murzin, A. G., Lesk, A. M. & Chothia, C. Principles determining the structure of beta-sheet barrels in proteins. I. A theoretical analysis. *J Mol Biol* **236**, 1369-1381 (1994).

162　Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *Embo J* **1**, 945-951 (1982).

163　Gruswitz, F., O'Connell, J., 3rd & Stroud, R. M. Inhibitory complex of the transmembrane ammonia channel, AmtB, and the cytosolic regulatory protein, GlnK, at 1.96 A. *Proc Natl Acad Sci U S A* **104**, 42-47 (2007).

164   Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res* **31**, 3381-3385 (2003).

165   Strosser, J., Ludke, A., Schaffer, S., Kramer, R. & Burkovski, A. Regulation of GlnK activity: modification, membrane sequestration and proteolysis as regulatory principles in the network of nitrogen control in *Corynebacterium glutamicum. Mol Microbiol* **54**, 132-147 (2004).

166   Mizuno, Y., Berenger, B., Moorhead, G. B. & Ng, K. K. Crystal structure of Arabidopsis PII reveals novel structural elements unique to plants. *Biochemistry* **46**, 1477-1483, doi:10.1021/bi062149e (2007).

167   Bueno, R., Pahel, G. & Magasanik, B. Role of glnB and glnD gene products in regulation of the glnALG operon of *Escherichia coli. J Bacteriol* **164**, 816-822 (1985).

168   Burkovski, A. Nitrogen control in *Corynebacterium glutamicum*: proteins, mechanisms, signals. *J Microbiol Biotechnol* **17**, 187-194 (2007).

169   Reuther, J. & Wohlleben, W. Nitrogen metabolism in *Streptomyces coelicolor*: transcriptional and post-translational regulation. *J Mol Microbiol Biotechnol* **12**, 139-146 (2007).

170   Hesketh, A. *et al.* The GlnD and GlnK homologues of *Streptomyces coelicolor* A3(2) are functionally dissimilar to their nitrogen regulatory system counterparts from enteric bacteria. *Mol Microbiol* **46**, 319-330 (2002).

171   Burkovski, A. Ammonium assimilation and nitrogen control in *Corynebacterium glutamicum* and its relatives: an example for new regulatory mechanisms in actinomycetes. *FEMS Microbiol Rev* **27**, 617-628 (2003).

172   Beckers, G. *et al.* Regulation of AmtR-controlled gene expression in *Corynebacterium glutamicum*: mechanism and characterization of the AmtR regulon. *Mol Microbiol* **58**, 580-595 (2005).

173   Amon, J. *et al.* Nitrogen control in *Mycobacterium smegmatis*: nitrogen-dependent expression of ammonium transport and assimilation proteins depends on the OmpR-type regulator GlnR. *J Bacteriol* **190**, 7108-7116 (2008).

174   Amon, J., Titgemeyer, F. & Burkovski, A. A genomic view on nitrogen metabolism and nitrogen control in mycobacteria. *J Mol Microbiol Biotechnol* **17**, 20-29 (2009).

175   Tremblay, P. L. & Hallenbeck, P. C. Of blood, brains and bacteria, the Amt/Rh transporter family: emerging role of Amt as a unique microbial sensor. *Mol Microbiol* **71**, 12-22 (2009).

176   Turnbull, A. P. & Boyd, S. M. Targeting cancer using fragment based drug discovery. *Anti-cancer Agents in Medicinal Chemistry* **12**, 40-48 (2012).

177   Erlanson, D. A. Introduction to fragment-based drug discovery. *Topics in Current Chemistry* **317**, 1-32, doi:10.1007/128_2011_180 (2012).

178   Newman, D. J., Cragg, G. M. & Snader, K. M. The influence of natural products upon drug discovery. *Natural Product Reports* **17**, 215-234 (2000).

179   Geysen, H. M., Schoenen, F., Wagner, D. & Wagner, R. Combinatorial compound libraries for drug discovery: an ongoing challenge. *Nat Rev Drug Discov* **2**, 222-230 (2003).

180   Cragg, G. M., Newman, D. J. & Snader, K. M. Natural products in drug discovery and development. *Journal of Natural Products* **60**, 52-60, doi:10.1021/np9604893 (1997).

181   Cropp, T. A. *et al.* Recent developments in the production of novel polyketides by combinatorial biosynthesis. *Biotechnol Genet Eng Rev* **19**, 159-172 (2002).

182   McDaniel, R. *et al.* Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proc Natl Acad Sci U S A* **96**, 1846-1851 (1999).

183   Henkel, T., Brunne, M. R., Muller, H. & Reichel, F. Statistical Investigation into the Structural Complementarity of Natural Products and Synthetic Compounds. *Angew. Chem. Int. Ed.* **38**, 643-647 (1999).

184   Muller-Kuhrt, L. Putting nature back into drug discovery. *Nat Biotechnol* **21**, 602 (2003).

185   Lee, M. L. & Schneider, G. Scaffold architecture and pharmacophoric properties of natural products and trade drugs: application in the design of natural product-based combinatorial libraries. *J Comb Chem* **3**, 284-289 (2001).

186   Berdy, J. Bioactive microbial metabolites. *J Antibiot (Tokyo)* **58**, 1-26, doi:10.1038/ja.2005.1 (2005).

187   Waksman, S. A. Antibiotics of Actinomycetes, an introduction and an outlook. *Antibiotics annual* **6**, 22-30 (1958).

188   Genilloud, O. *et al.* Current approaches to exploit actinomycetes as a source of novel natural products. *Journal of Industrial Microbiology & Biotechnology* **38**, 375-389, doi:10.1007/s10295-010-0882-7 (2011).

189   Corley, D. G. & Durley, R. C. Strategies for Database Dereplication of Natural-Products. *J Nat Products* **57**, 1484-1490 (1994).

190   Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* **46**, 3-26 (2001).

191   Fullbeck, M., Michalsky, E., Dunkel, M. & Preissner, R. Natural products: sources and databases. *Nat Prod Rep* **23**, 347-356, doi:10.1039/b513504b (2006).

192   Bobzin, S. C., Yang, S. & Kasten, T. P. LC-NMR: a new tool to expedite the dereplication and identification of natural products. *Journal of Industrial Microbiology & Biotechnology* **25**, 342-345, doi:10.1038/sj/jim/7000057 (2000).

193   Nielsen, K. F., Mansson, M., Rank, C., Frisvad, J. C. & Larsen, T. O. Dereplication of microbial natural products by LC-DAD-TOFMS. *Journal of Natural Products* **74**, 2338-2348, doi:10.1021/np200254t (2011).

194   El Sayed, K. A. *et al.* New manzamine alkaloids with potent activity against infectious diseases. *Journal of the American Chemical Society* **123**, 1804-1808 (2001).

195   Yousaf, M. *et al.* 12,34-Oxamanzamines, novel biocatalytic and natural products from manzamine producing Indo-Pacific sponges. *Tetrahedron* **58**, 7397-7402 (2002).

196   Haygood, M. G., Schmidt, E. W., Davidson, S. K. & Faulkner, D. J. Microbial symbionts of marine invertebrates: opportunities for microbial biotechnology. *J Mol Microbiol Biotechnol* **1**, 33-43 (1999).

197   De Marino, S. *et al.* Swinholide J, a potent cytotoxin from the marine sponge *Theonella swinhoei*. *Mar Drugs* **9**, 1133-1141, doi:10.3390/md9061133 (2011).

198   Andrianasolo, E. H. *et al.* Isolation of swinholide A and related glycosylated derivatives from two field collections of marine cyanobacteria. *Organic Letters* **7**, 1375-1378, doi:10.1021/ol050188x (2005).

199   Natali, F., Siculella, L., Salvati, S. & Gnoni, G. V. Oleic acid is a potent inhibitor of fatty acid and cholesterol synthesis in C6 glioma cells. *Journal of Lipid Research* **48**, 1966-1975, doi:DOI 10.1194/jlr.M700051-JLR200 (2007).

200   Rose, D. P. & Connolly, J. M. Effects of Fatty-Acids and Inhibitors of Eicosanoid Synthesis on the Growth of a Human Breast-Cancer Cell-Line in Culture. *Cancer Research* **50**, 7139-7144 (1990).

201   McKinney, J. D. *et al.* Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature* **406**, 735-738, doi:10.1038/35021074 (2000).

202   Segal, W. *The Mycobacteria: A Sourcebook* Vol. 2 (ed Lawrence G. Wayne George P. Kubica) 547–573 (Dekker, 1984).

203   Clark, D. P. C., J. E. Jr. in *Escherichia coli and Salmonella: Cellular and Molecular Biology*   (ed F. C. Neidhardt) 343–357 (ASM Press, 1996).

204   Cronan, J. E. J. L., D. in *Escherichia coli and Salmonella: Cellular and Molecular Biology*   (ed F. C. Neidhardt) 206–216 (ASM Press, 1996).

205   Lorenz, M. C. & Fink, G. R. The glyoxylate cycle is required for fungal virulence. *Nature* **412**, 83-86, doi:10.1038/35083594 (2001).

206   Hong, P. C., Tsolis, R. M. & Ficht, T. A. Identification of genes required for chronic persistence of *Brucella abortus* in mice. *Infect Immun* **68**, 4102-4107 (2000).

207   Vereecke, D., Cornelis, K., Temmerman, W., Holsters, M. & Goethals, K. Versatile persistence pathways for pathogens of animals and plants. *Trends Microbiol* **10**, 485-488 (2002).

208   Solomon, P. S., Lee, R. C., Wilson, T. J. & Oliver, R. P. Pathogenicity of *Stagonospora nodorum* requires malate synthase. *Mol Microbiol* **53**, 1065-1073, doi:10.1111/j.1365-2958.2004.04178.x (2004).

209   Kornhuber, J. *et al.* Lipophilic Cationic Drugs Increase the Permeability of Lysosomal Membranes in a Cell Culture System. *J Cell Physiol* **224**, 152-164, doi:Doi 10.1002/Jcp.22112 (2010).

210   Goldberg, A. L., Akopian, T. N., Kisselev, A. F., Lee, D. H. & Rohrwild, M. New insights into the mechanisms and importance of the proteasome in intracellular protein degradation. *Biological Chemistry* **378**, 131-140 (1997).

211   Glickman, M. H. & Ciechanover, A. The ubiquitin-proteasome proteolytic pathway: Destruction for the sake of construction. *Physiol Rev* **82**, 373-428, doi:DOI 10.1152/physrev.00027.2001 (2002).

212   Mani, A. & Gelmann, E. P. The ubiquitin-proteasome pathway and its role in cancer. *J Clin Oncol* **23**, 4776-4789, doi:Doi 10.1200/Jco.2005.05.081 (2005).

213   Kanayama, H. *et al.* Changes in expressions of proteasome and ubiquitin genes in human renal cancer cells. *Cancer Res* **51**, 6677-6685 (1991).

214   An, W. G., Hwang, S. G., Trepel, J. B. & Blagosklonny, M. V. Protease inhibitor-induced apoptosis: accumulation of wt p53, p21(WAF1/CIP1) and induction of apoptosis are independent markers of proteasome inhibition. *Leukemia* **14**, 1276-1283 (2000).

215   Groll, M., Koguchi, Y., Huber, R. & Kohno, J. Crystal structure of the 20 S proteasome: TMC-95A complex: A non-covalent proteasome inhibitor. *Journal of Molecular Biology* **311**, 543-548, doi:DOI 10.1006/jmbi.2001.4869 (2001).

216   Meng, L. H., Kwok, B. H. B., Sin, N. & Crews, C. M. Eponemycin exerts its autitumor effect through the inhibition of proteasome function. *Cancer Research* **59**, 2798-2801 (1999).

217   Fenteany, G. *et al.* Inhibition of Proteasome Activities and Subunit-Specific Amino-Terminal Threonine Modification by Lactacystin. *Science* **268**, 726-731 (1995).

218   Kisselev, A. F. & Goldberg, A. L. Proteasome inhibitors: from research tools to drug candidates. *Chemistry & Biology* **8**, 739-758 (2001).

219   Koguchi, Y. *et al.* TMC-95A, B, C, and D, novel proteasome inhibitors produced by *Apiospora montagnei Sacc. TC 1093*. Taxonomy, production, isolation, and biological activities. *J Antibiot (Tokyo)* **53**, 105-109 (2000).

220   Kroll, M. *et al.* The secondary fungal metabolite gliotoxin targets proteolytic activities of the proteasome. *Chem Biol* **6**, 689-698 (1999).

221   Nam, S., Smith, D. M. & Dou, Q. P. Ester bond-containing tea polyphenols potently inhibit proteasome activity in vitro and in vivo. *J Biol Chem* **276**, 13322-13330, doi:10.1074/jbc.M004209200 (2001).

222   Oakervee, H. E. *et al.* PAD combination therapy (PS-341/bortezomib, doxorubicin and dexamethasone) for previously untreated patients with multiple

myeloma. *Brit J Haematol* **129**, 755-762, doi:DOI 10.1111/j.1365-2141.2005.05519.x (2005).

223   Jagannath, S. *et al.* Bortezomib therapy alone and in combination with dexamethasone for patients with previously untreated multiple myeloma. *Blood* **106**, 231a-231a (2005).

224   Kurosu, M. & Begari, E. Bacterial Protein Kinase Inhibitors. *Drug Develop Res* **71**, 168-187, doi:Doi 10.1002/Ddr.20362 (2010).

225   Lopez, S. N., Ramallo, I. A., Sierra, M. G., Zacchino, S. A. & Furlan, R. L. Chemically engineered extracts as an alternative source of bioactive natural product-like compounds. *Proc Natl Acad Sci U S A* **104**, 441-444, doi:10.1073/pnas.0608438104 (2007).

226   Maharjan, R. P. & Ferenci, T. Global metabolite analysis: the influence of extraction methodology on metabolome profiles of *Escherichia coli*. *Anal Biochem* **313**, 145-154 (2003).

227   Lopez, S. N., Ramallo, I. A., Sierra, M. G., Zacchino, S. A. & Furlan, R. L. E. Chemically engineered extracts as an alternative source of bioactive natural product-like compounds. *P Natl Acad Sci USA* **104**, 441-444, doi:DOI 10.1073/pnas.0608438104 (2007).

228   Salazar, M. O., Ramallo, I. A., Micheloni, O., Sierra, M. G. & Furlan, R. L. E. Chemically engineered extracts: Bioactivity alteration through sulfonylation. *Bioorganic & Medicinal Chemistry Letters* **19**, 5067-5070, doi:DOI 10.1016/j.bmcl.2009.07.038 (2009).

229   Ramallo, I. A., Salazar, M. O., Mendez, L. & Furlan, R. L. E. Chemically Engineered Extracts: Source of Bioactive Compounds. *Accounts of Chemical Research* **44**, 241-250, doi:Doi 10.1021/Ar100106n (2011).

APPENDIX A

Chemical Derivatization of Microbial Extracts

Generating a sufficiently large collection of diverse natural products to feed a high-throughput screening effort in drug discovery requires the systematic extraction and isolation of natural products from a wide variety of sources, like soil bacteria, fungi, marine organisms and higher plant species. This can be a daunting process and consume a large amount of resources and time to accomplish. Since a single natural product extract mixture could contain potentially diverse set of structural scaffolds which is typically observed in secondary metabolites, such an extract mixture could be used as a starting point to introduce more chemical functionalities in a semisynthetic approach. Such an approach was successfully applied to a study conducted on the *n*-butanol extract of the plant *Polygonum ferrugineum* Wedd. (Polygonaceae) [225]. This approach does not assume any previous knowledge of the constituents of the mixture to be functionalized, nor does it require the isolation of the individual constituents. A chemical modification reaction is directed to a select functional group. Different secondary metabolites present in the extraction mixture containing the select functional group can get modified and give rise to analogs. Depending on the number of molecules with the functional group existing in the mixture, this technique can simultaneously generate a large number of new analogs, which can contribute to a greater diversity of natural products.

In the present study a similar approach has been applied to a mixture of secondary metabolites extracted from microbial species. The hypothesis was that

advantage can be taken of the structural diversity already existing in nature by using the secondary metabolites obtained from microorganisms as basis and starting points for carrying out chemical derivatizations. Single step chemical derivatization / functionalization reactions carried out on entire crude extracts from microorganisms have the potential to simultaneously generate a large number of chemically modified natural product analogs, thus introducing even more diversity into the extracts. Such an approach can provide a rapid system of generating libraries of structurally and chemically diverse set of natural products, thus saving the time and cutting costs.

Methods

Bacterial Fermentation

The microorganisms used to generate pools of natural product extracts included *E. coli*, *Mycobacterium smegmatis* (mc^{2}155) [112], *Streptomyces griseus*, *Streptomyces sahachiroi* and bacteria isolated from the soil. Each organism was grown independently in cultures containing a wide range of fermentation media recipes like Lysogeny broth (LB), terrific broth (TB), Middlebrook 7H9 media (Difco), pharmamedia, soybean meal, corn steep liquor. The common carbon sources were glucose, sucrose, mannitol, and glycerol. Salts like manganese chloride, calcium carbonate, copper sulfate, iron sulfate were used to help increase the secondary metabolite production. The cell culture volumes ranged from 50 mL – 2 L. The incubation temperatures ranged from 25 – 37 ˚C. The time period of incubation ranged from 5 – 7 days. At the end of the incubation period the cells were spun down and the pellets separated from the spent media. The cell

250

pellets and the spent media were independently treated to extract pools of secondary

metabolites. Alternately, the culture was treated to an extraction protocol without

separating the cells from the spent media.

Cell Lysis and Extraction

The cell pellets were extracted using a previously described cold methanol

extraction protocol that is shown to extract the maximum pool of secondary metabolites

from the pellets [226]. The spent media was extracted by equilibrating it with organic

solvents in a separating funnel. The organic solvents used for extraction included

dichloromethane, 1:1 dichloromethane / isopropanol or ethyl acetate. Extraction of the

spent media was carried out at neutral pH, acidic (3 – 4) pH, and basic (10 – 11) pH to

enrich the extracts with natural compounds containing either acidic or basic groups.

High Performance Liquid Chromatography Fractionation

The unmodified extracts were dissolved in methanol, fractionated on a reverse

phase C-18 Atlantis T3® (Waters) HPLC column to generate an ultraviolet (UV)

absorbance peak profile of the extract. An analytical HPLC system with a W600 pump

and a Photo Diode Array (PDA) detector from Waters was used. A solvent gradient was

run from 100 % water containing 0.1 % formic acid (FA) to 100 % acetonitrile. The

gradient was linear with a 1% change per minute and a flow rate of 1 ml /min. (The

chemically modified extracts were dissolved in methanol or methanol water mixtures

and crude fractionation was carried out identically). The fractions are collected at

1ml/min. The individual fractions were then dried out by lyophilization and reconstituted in water or methanol depending on solubility of the dried fraction. Thus stock solutions were made for each fractions and aliquoted into 96 well polypropylene plates. The plates were stored at 4 ˚C until further analysis.

Chemical Derivatization

p-bromobenzoyl chloride was used to carry out acylation of basic amine groups present in the natural products. Other chemical modifications carried out on the extracted metabolites included – amide bond formation using phenyl silane, benzoylation using p-trifluoromethyl benzoylchloride, chlorination using trichloroisocyanuric acid and fluorination using DAST and using Selectfluor reagents. Click chemistry was also used to generate derivatives of secondary metabolites in the natural product extracts. The reaction involved two steps; the first step was to react the nucleophilic residues like amines, hydroxyl groups and thiol groups on the secondary metabolites in the crude extracts with a bromomethyl alkyne to generate a propargylated species. The second step was to react the propargylated species with bromobenzyl azide to undergo a cyclization reaction and generate benzyl triazoles.

Whole-Cell Activity Assay

Whole-cell bactericidal activity was carried out on *M. smegmatis* (mc^2155) grown on 7H9 and M9 media and independently on *Mtb* (mc^27000) grown on 7H9 media. 196 μL of *M. smegmatis* mc^2155 cells diluted in 7H9 media to final $O.D_{600}$ of

0.001 were added to the wells of a sterile 96 well microtitter plate. 4 μL of the fractionated extracts were then added to each well. Rifampicin at concentrations of 2 μM to 250 μM was used as a negative control. Wells containing only 7H9 media served as another negative control, while wells with the *M. smegmatis* mc^{2}155 at the above mentioned final O.D$_{600}$ of 0.001 served as the positive control. The cells were grown in an incubator at 37 ˚C for 72 hours. After this period, 5 μL of alamar blue (resazurin dye) was added to the wells and left for a couple of hours in the incubator. The wells were then measured visually for cell growth. Wells with cell growth were able to reduce the alamar blue which then turned pink. The resazurin dye in the wells with no cell growth remained blue.

Liquid Chromatography /Mass Spectrometry

Fractions showing whole cell activity against *M. smegmatis* and *Mtb* mc^2 7000 were subjected to further fractionation on HPLC using a shallow gradient of the same mobile phase solvents used previously to allow for a better separation of the individual components of the active fraction. Simultaneously an aliquot of the same fractions were analyzed using liquid chromatography /mass spectrometry (LC /MS) to determine the mass composition of the fraction mixture. For the purpose of LC / MS, two different types of instrumentation were employed – a supercritical fluid chromatography (SFC) / MS (Waters) and an analytical LC /quadrupole Time of Flight (Q-TOF) MS. The analytical LC was obtained from Agilent whereas the Q-TOF was bought from Bruker

Daltonics. The Bruker microTOF-QII® is able to detect masses in high resolution

(HRMS) and at an accuracy rate of 5 ppm.

The SFC / MS use a normal phase column and hence was useful for analyzing

the nonpolar components of the mixture. However, due to the nature of the extraction

procedure, which selected for mainly polar compounds to be extracted, the SFC/MS

instrument had limited success in identifying nonpolar compounds. The regular LC /

HRMS on the other hand was able to detect several masses in the mass range of 100 to

1000 m/z (mass to charge ratio) which were representative of the active fractions.

Nuclear Magnetic Resonance Spectroscopy

500 MHz NMR spectrometers were used to generate ^{1}H NMR data for the active

natural compounds, when the activity came from a sufficiently pure fraction. ^{13}C NMR

data were also obtained whenever possible to assist in the assignment of the structure of

the bioactive natural compound.

Results and Discussion

The aim of this project was to generate diversity in any given crude extract

derived from a microorganism by employing relatively simple chemical derivatizations

on the crude microbial extracts. An example of the evidence for chemical modification

of the extracts was the p-bromobenzoylation reaction. The reagent, p-bromobenzoyl

chloride was used to react with primary or secondary amines present in the crude

microbial extract to form amide derivatives in a non-targeted fashion. Compounds

containing hydroxyl groups would also react with p-bromobenzoyl chloride to form the corresponding esters. Methanol extracts of *Streptomyces griseus* cells were used for the reaction. The cells grown in liquid media were harvested from the media and divided into two parts. While one part of the cells was subjected to the functionalization reaction, the other part was left unreacted and used as a reference. Both functionalized and the unfunctionalized extracts were fractionated on HPLC using the same solvent gradient system mentioned in the methods section above (**Figure 87 and 88**).
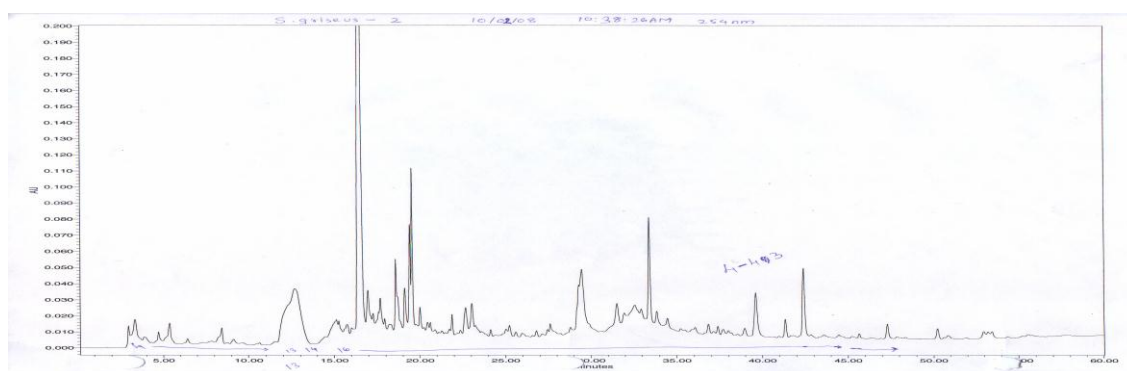

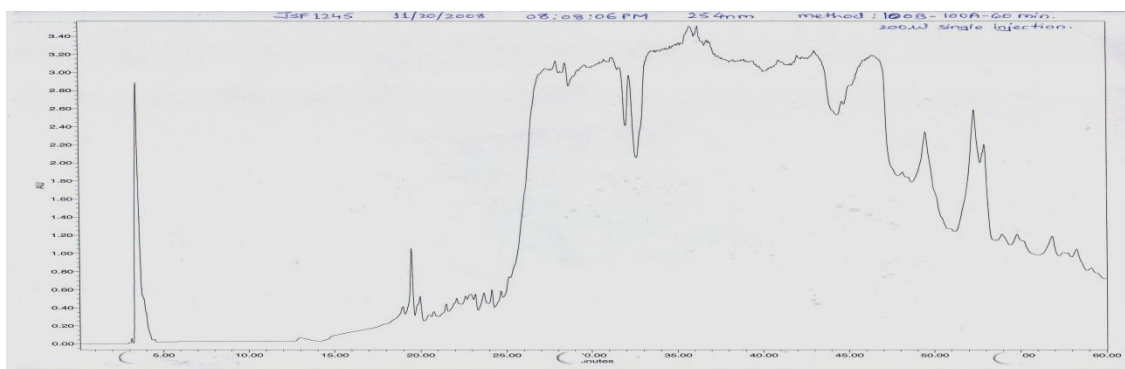**Figure 87:** HPLC UV profile of the unfunctionalized *S. griseus* extract.

Figure 88: HPLC UV profile of the p-bromobenzoylated *S. griseus* extract.

Success of the derivatization reaction was confirmed by three different methods:

SFC-MS runs of the before and after derivatization extracts were compared to look for characteristic bromine isotope peaks in the derivatized extracts. Indeed such isotope peaks were observed in the MS, which did not occur in the unmodified extracts, (**Figure 89**) thus proving success of the reaction.
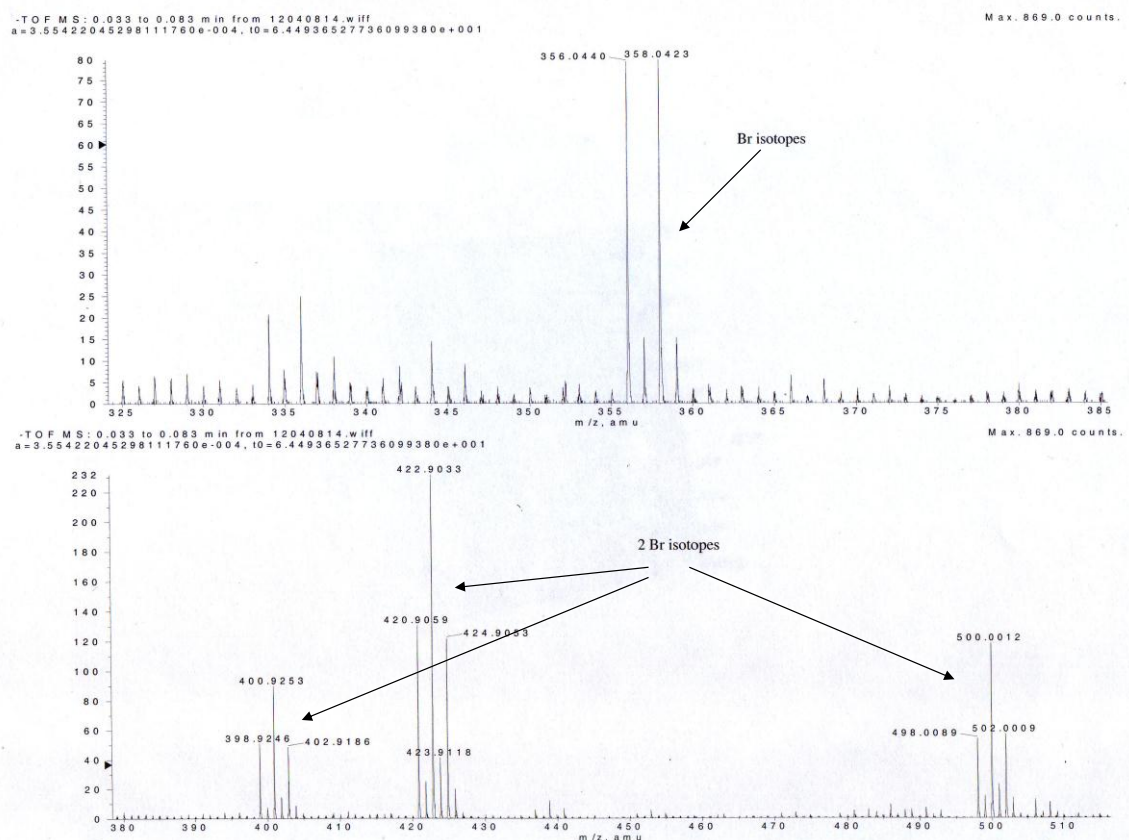
Figure 89: Characteristic bromine isotope peaks observed in the p-bromobenzoylated *S. griseus* extract.

To check for the reaction of amines to amides, chemical methods were employed: 70 mg /ml stock solutions of the unmodified and p-bromobenzoylated extract were prepared in methanol. Equal amounts of both the solutions were spotted on a TLC plate next to each other with a spot of plain water acting as control. A drop of the freshly prepared ninhydrin reagent was added to the spots and the plate was heated. The spot corresponding to the unmodified extract developed a strong red-blue coloration, while the spot corresponding to the modified extracts developed only faint coloration (**Figure**

257

90). This was evidence that the p-bromobenzoylation reaction of the extract was successful.

Unmodified *S. griseus*
extract (at 70 mg/ml)

P-bromo benzoylated *S. griseus*
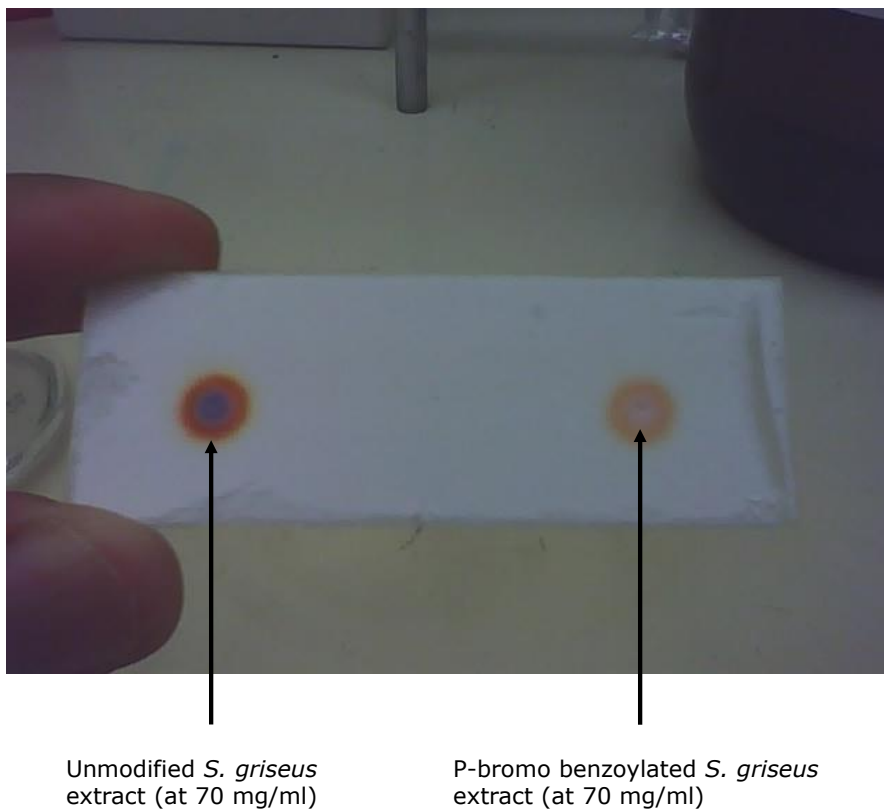extract (at 70 mg/ml)

Figure 90: Ninhydrin test to examine the functionalization of amines to amides.

p-bromobenzoylation reaction introduces a phenyl group in the product, which in turn introduces conjugation into the system. Such compounds would be expected to have higher absorbance in the UV region. 0.6 mg/ml solutions of the unmodified and p-

bromobenzoylated extract were prepared and scanned for their absorbance spectra in a
UV spectrophotometer. As expected, the p-bromobenzoylated extract showed a much
stronger absorbance in the UV region of 200nm to 400nm than the unmodified extract
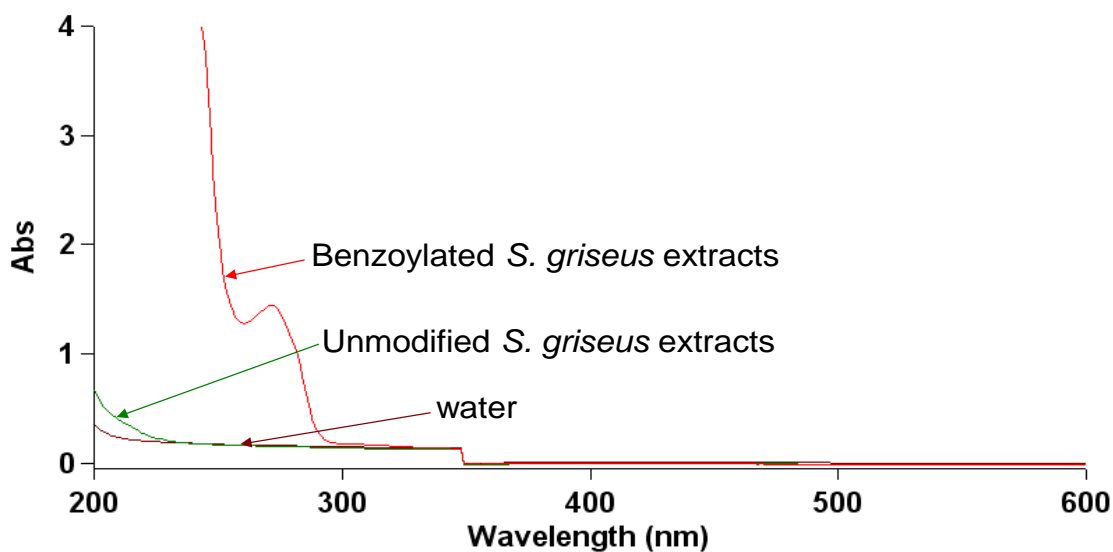(**Figure 91**). This was further evidence that the benzoylation reaction was successful.

Figure 91: The UV absorbance of unmodified and p-bromobenzoylated *S. griseus*
extracts at a concentration of 0.6 mg/ml. The p-bromobenzoylated extract showed a
stronger UV absorbance in the range of 200 – 300 nm than the unmodified extract.
Water was used as the blank negative control.

Conclusion

A method is described to increase the structural and functional diversity of
secondary metabolites in a crude mixture of unknown microbial extracts. Model

organisms like *E. coli*, *M. smegmatis* and *Streptomyces griseus* as well as randomly picked soil microbes such as *Bacillus thuringiensis* and *Bacillus megaterium* were cultured in the laboratory to a sufficiently high optical density. After harvesting the cell pellets secondary metabolites were extracted out of the cells by rupturing the cell wall using organic solvents like, dichloromethane and ethyl acetate. Chemical modifications were introduced in the pool of the secondary metabolites by employing single step derivatization reactions like chlorination, fluorination or p-bromobenzoylation carried out on the crude extracts without any purification. Click chemistry was one of the tools used for the derivatization reactions as well. The aim of the experiment was to introduce single modifications randomly in the pool of metabolites, which could potentially generate bioactive molecules where none existed before. For this purpose, the derivatized extracts were fractionated on HPLC and the fractions were tested for their whole-cell activities on *M. smegmatis* and *Mtb*. The underivatized extracts were also fractionated and tested for activity for reference and to keep track of the activity profile. This was followed by an activity-based fractionation and isolation of the active chemical component from the extract.

Characteristic bromine and chlorine isotope peaks were identified through LC-MS analysis for the respectively halogenated extracts confirming the success of the functionalization reactions. Chemical functionalization of the extract was also confirmed through chemical reactions such as the ninhydrin test to seek the functionalization of amine functional groups. High resolution mass spectrum (HRMS) data was obtained on

the Bruker microTOF-QII® mass spectrometer, which enabled preliminary identification of derivatized metabolites.

Secondary metabolites in a *S. griseus* fermentation extract were successfully p-bromobenzoylated by chemical derivatization on the crude extracts. Although similar derivatization experiments have been successfully carried out by other groups on a selected pool of known secondary metabolites [227-229], this is the first instance where a completely random set of unknown metabolites was used in the form of microbial extract. This experiment was a proof-of-concept that such derivatizations can be applied to any randomly selected metabolite pool or microbial extracts to generate a library of functionalized natural products. This process can significantly increase the size of the library of natural products that are used in the high throughput whole-cell screening or for *in vitro* enzyme inhibition assays. A larger library consisting of wider structural and chemical diversity of natural products will increase the probability of finding a drug lead compound.