Artificial Genetically Encoded Peptides and Proteins as Next-Generation Therapeutics: Selection of ligands for *Mycobacterium tuberculosis* UDP-Galactopyranose Mutase as Potential Inhibitors.

Trisha Ghosh

A Thesis

in

Department of Biology

Presented in Partial Fulfilment of the Requirements

For the Degree of Master of Science (Biology) at

Concordia University, Montreal Quebec, Canada

August 2020

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Trisha Ghosh

Entitled: Artificial Genetically Encoded Peptides and Proteins as Next-Generation Therapeutics: Selection of ligands for *Mycobacterium tuberculosis* UDP-Galactopyranose Mutase as Potential Inhibitors.

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Biology)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

_____ Chair

Dr. Malcolm Whiteway

_____ Examiner

Dr. Peter Pawelek

_____ Examiner

Dr. Malcolm Whiteway

_____ Examiner

Dr. Steve Shih

_____ Thesis Supervisor

Dr. David Kwan

Approved by

_____ Chair of Department

Dr. Selvadurai Dayanandan

Date: August 14th 2020

ABSTRACT

Artificial Genetically Encoded Peptides and Proteins as Next-Generation Therapeutics: Selection of ligands for *Mycobacterium tuberculosis* UDP-Galactopyranose Mutase as Potential Inhibitors.

Trisha Ghosh

Peptides and small proteins provide a dynamic platform for drug discovery and therapeutics. They have a wide range of applications including inhibition of protein-protein interaction, inhibition of transporter and enzyme activity, imaging, and as co-crystallisation ligand for structural studies. In this study, we employed *in vitro* selection of ligands, using mRNA display, from an artificial genetically encoded library of peptides/proteins to identify candidates that bind the enzyme *Mycobacterium tuberculosis* UDP-Galactopyranose mutase (*Mt*UGM). The enzyme catalyzes the reversible conversion of UDP-galactopyranose (UDP-Gal*p*) to UDP-galactofuranose (UDP-Gal*f*), which is then assembled as the galactofuran layer in *Mycobacterium tuberculosis* (*Mtb*) cell wall. Cell wall biosynthesis is essential for *Mtb* survival and pathogenicity, deletion in genes involved in this process have proven lethal. We successfully identified macrocyclic peptides (MCPs) and affibodies specific for *Mt*UGM from a library with a diversity $>10^{12}$ clones through random non-standard peptide integrated discovery (RaPID) system and mRNA display, respectively. Enrichment of positive binders was observed for both selection processes suggesting binding specificity. Previous studies reveal that natural product-like MCPs can inhibit enzyme activity; their small size, conformational stability and high affinity for the target makes them attractive ligands. Thus, we hypothesize that identified MCPs may show inhibition of *Mt*UGM thereby halting cell wall biosynthesis. Discovered affibodies will be used for structural analysis of *Mt*UGM. The World Health Organisation (WHO) reported tuberculosis (TB) caused by *Mtb* to be one of the leading causes of death worldwide, with increasing incidences of drug resistant strains necessitating discovery of novel therapeutics. Through our study, we present a novel approach for discovering peptide/protein-based ligands for *Mt*UGM and hope to develop an assay for screening MCPs for potential inhibitors.

ACKNOWLEDGEMENTS

I take this opportunity to thank the Department of Biology, Faculty of Arts and Science, Concordia University, Montreal for giving me the opportunity of being a part of this program.

I am immensely thankful to Dr. David Kwan for supervising me during the program and providing me opportunities of international collaboration and networking. I would like to thank Dr. Christopher Hipolito at University of Tsukuba for his guidance and help during my work in Japan. I would also like to thank my committee members Dr. Malcolm Whiteway and Dr. Peter Pawelek for providing valuable insights on this project.

I thank Lauren Narcross, Kasper Kevvai and Daniel Tsyplenkov for helping me setup the HPLC.

I am immensely grateful to the members of the Kwan lab Dr. Lan Huong Thi Nguyen, F. Ifthiha Moheeden, Mohamed Nasr, Jamaac Ahmed, Haouyu Wu, Maxim Soroko, and Bo Yi Han for their constant support. I extend my gratitude to Dr. Lan Huong Thi Nguyen for initiating me in the lab and for her inputs on the project. A special thanks to Mohamed Nasr and F. Ifthiha Mohideen for always encouraging me. It was a pleasure to work with Bo Yi Han on parts of the project during her BIOL490.

Lastly, I would like to thank my father Mr. Binodranjan Ghosh and mother Mrs. Swapna Ghosh for their love, support and blessings. A big thanks to my sister Soumi Ghosh, without her, this journey would have been impossible.

CONTRIBUTIONS

This project was conceived by Dr. David Kwan[1]. The chemical synthesis of macrocyclic peptides was done by Dr. Christopher Hipolito[2], he also helped in method development for *in vitro* selection of peptides employed in this project. Dr. Lan Huong Thi Nyugen[1] helped with the scale up and purification of His_6-MBP-*Mt*UGM and His_6-MBP used in this project. The TEV protease purification and HPAEC assay development were performed with help of Bo Yi Han[1]. All other experiments were designed, performed, and analyzed by Trisha Ghosh[1].

1. **Department of Biology, Concordia University, Montreal, QC, Canada**
2. **University of Tsukuba, Tsukuba, Ibaraki, Japan**

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

BCA	Bicinchoninic acid
BSA	Bovine serum albumin
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
EMB	Ethambutol
FPLC	Fast protein liquid chromatography
FIT	Flexible *In vitro* translation
Gal*f*	Galactofuranose
Gal*p*	Galactopyranose
GFP	Green fluorescent protein
His-tag	Histidine-tag
HPAEC	High performance anion-exchange chromatography
IMAC	Immobilized affinity chromatography
INH	Isoniazid
IPTG	Isopropyl-ß-thiogalactoside
LB	Luria broth

M. tuberculosis	*Mycobacterium tuberculosis*
MBP	Maltose-binding protein
MCP	Macrocyclic peptide
MDR-TB	Multiple drug resistant TB
*Mt*UGM	*Mycobacterium tuberculosis* UDP-galactopyranose mutase
Ni-NTA	Nickle-nitrilotriacetic acid
PCR	Polymerase chain reaction
RaPID	Random nonstandard peptide integrated discovery
RIF	Rifampicin
RPM	Rotation per minute
RNA	Ribonucleic acid
SDS-PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis
TB	Tuberculosis
TDR-TB	Total drug resistant TB
TEV	Tobacco etch virus
UDP	Uridine diphosphate
UDP-Gal*f*	UDP-galactofuranose
UDP-Gal*p*	UDP-galactopyranose
XDR-TB	Extremely drug resistant TB

CHAPTER 1: INTRODUCTION

1.1 Tuberculosis and Current Treatment

Tuberculosis (TB) is an infectious disease caused by *Mycobacterium tuberculosis*. It remains the leading cause of death from a single infectious agent. In 2018, approximately 10 million people in total were diagnosed with TB of which 1.5 million cases were fatal [1]. TB is primarily a pulmonary disease, affecting the lungs, but it can be extrapulmonary affecting other parts of the body [2, 6]. *M. tuberculosis* spreads from an affected individual to another in the form of aerosol droplets. The virulent bacteria are phagocytosed by macrophage cells in the alveoli via macrophage mannose receptors [3]. Upon entry into the host macrophage, the bacteria reside in phagosomes and escapes degradation by inhibiting phagosome-lysosome fusion [4]. Studies have shown that infected macrophages in the lung, through their production of chemokines, cause an immune response. Inactivated monocytes, lymphocytes, and neutrophils form a granulomatous lesion to contain the spread of infection. Active pulmonary TB is the stage when the bacteria escapes this granuloma stage and spread in the lung tissue; it can also spread to other tissues via lymphatic vessels and blood resulting in extrapulmonary TB [5, 6].

A combination of chemotherapeutic agents is administered for six months as first line drugs for treating active TB. This includes mainly four drugs- isoniazid (INH), rifampicin (RIF), pyrazinamide (PZA), and ethambutol (EMB) prescribed for an initial 2-month phase followed by a continuation phase with INH and RIF for 4 months [7]. In the past decade, a rise in cases of multiple drug resistant-TB (MDR-TB) caused by strains of *M. tuberculosis* resistant to both rifampicin and isoniazid with or without resistance to other drugs has been reported. Total drug resistant-TB (TDR-TB) is caused by strains that are resistant to all the four first line drugs [8, 9]. At this stage second line of drugs are prescribed, this mainly includes aminoglycosides, cycloserine (analogue of D-alanine) and fluoroquinolones. These drugs are less effective than the first-line drugs (e.g., p-aminosalicylic acid); or may have toxic side-effects (e.g., cycloserine).

Extensively drug resistant TB (XDRTB) is a rare type of MDR-TB, which is resistant to both RIF and INH, in addition to any fluoroquinolone and at least one of three second line drugs [10]. Demographic data shows that treatment success rate varies with respect to strain, country, and the method of treatment. As of 2018, the success rate for non-resistant strain has been 85% using first line drugs, however, it is >70% for MDR-TB/TDR-TB in developed countries and as low as 56% in developing countries [11]. The rapid evolution of bacteria leading to resistance against anti-TB drugs poses a challenge to the community to develop an effective drug regimen [12]. This necessitates the discovery of novel therapeutics that will increase the efficacy of existing drugs, promise administration at a lower dosage, fewer side effects and overcome drug resistance.

1.2 Mycobacterium Cell Wall Biosynthesis and The Role of *Mt*UGM

M. tuberculosis is a facultative intracellular pathogen with the ability to infect and live in the host for decades, as in the case of latent TB, despite a functioning host immune system [13]. The success of this pathogen has been associated with the complex structure of the cell envelope, which is credited for the bacteria's virulence and antibiotic resistance [14]. The currently administered first line drugs target the process of cell wall biosynthesis, elucidating the importance of cell wall for the survival of the bacteria. INH inhibits mycolic acid elongation, PZA inhibits fatty acid synthase, and EMB affect the arabinan assembly **(Fig 1)** [15]. However, the bacteria have evolved to escape drug action either by mutating its native enzymes which are drug targets, increasing drug efflux or creating low permeability to the drug [16]. Nonetheless, the cell wall remains to be an attractive target for drug development.

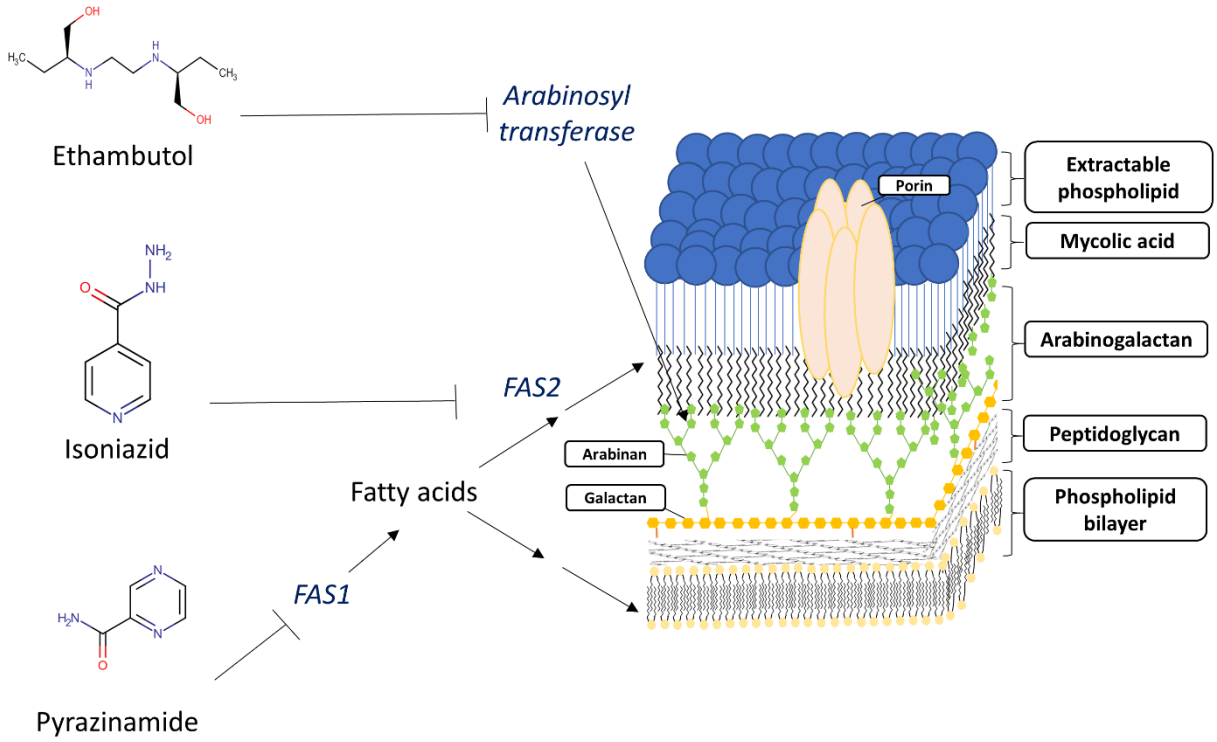


Figure 1. Anti-TB drugs targeting enzymes involved in cell wall biosynthesis. Arabinosyl transferase is inhibited by Ethambutol, further disrupting the arabinan assembly. Pyrazinamide and Isoniazid obstructs lipid chain formation by inhibiting fatty acid synthase I and II respectively. However, resistant *Mtb* strains reported mutations in these target enzymes to escape drug binding.

The cell wall is thick, waxy and hydrophobic. The structural complexity of cell wall—formed by peptidoglycan, lipid mycolic acid, and polysaccharide arabinogalactan—makes mycobacterium unique. Arabinogalactan (AG) is a heteropolysaccharide that is covalently linked to muramic acid residues of peptidoglycan by a phosphodiester bond [17, 18]. It is comprised of arabinose and galactose sugars that are present in the furanose ring form, D-galactofuranosyl (Gal*f*) and D-arabinofuranosyl (Ara*f*). The galactan layer is linear, with alternating β-(1→5) and β-(1→6) glycosidic linkages (5-linked β-D-Gal*f* and 6-linked β-D-Gal*f* residues). Previous studies have identified three key enzymes—UGM, GlfT1 and GlfT2—that help in galactan biosynthesis. UDP galactopyranose mutase (UGM), converts UDP-galactopyranose (UDP-Gal*p*; the nucleotide sugar donor of the six-membered ring form of galactose) to UDP-galactofuranose

(UDP-Gal*f*). The galactofuranosyltransferases GlfT1 and GlfT2, respectively prime and elongate the alternating β-(1→5) and β-(1→6) galactan using the UDP-Gal*f* donor substrate and a glycan acceptor substrate **(Fig 2)** [19-24]. UGM is found in several pathogens including bacteria, fungi, and nematodes. However, it is not found in mammals making it an attractive drug target [25, 26]. Efforts towards finding inhibitors for UGM have shown promising results. Inhibitors designed against UGM were essentially substrate analogs, some resembling UDP-Gal*f* which binds to the active site of the enzyme [27, 28].

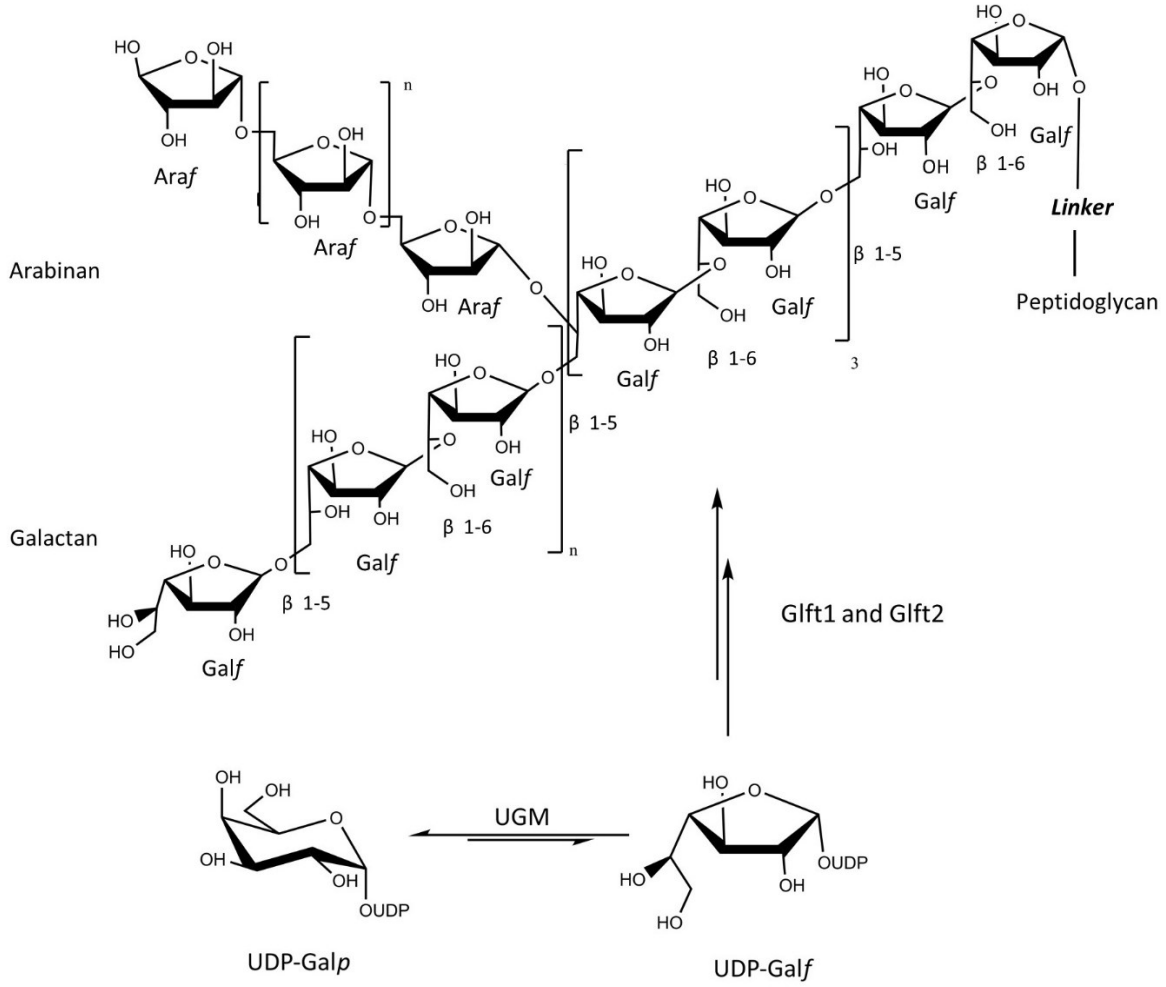


Figure 2. Role of UDP-Galactopyranose mutase in cell wall biosynthesis. UGM catalyzes the interconversion of UDP-Gal*p* and UDP-Gal*f*. Glft1 and Glft2 prime and elongate the alternating β-(1→5) and β-(1→6) galactan using the UDP-Gal*f* donor substrate and a glycan acceptor substrate to synthesize the galactan layer of mycobacteria.

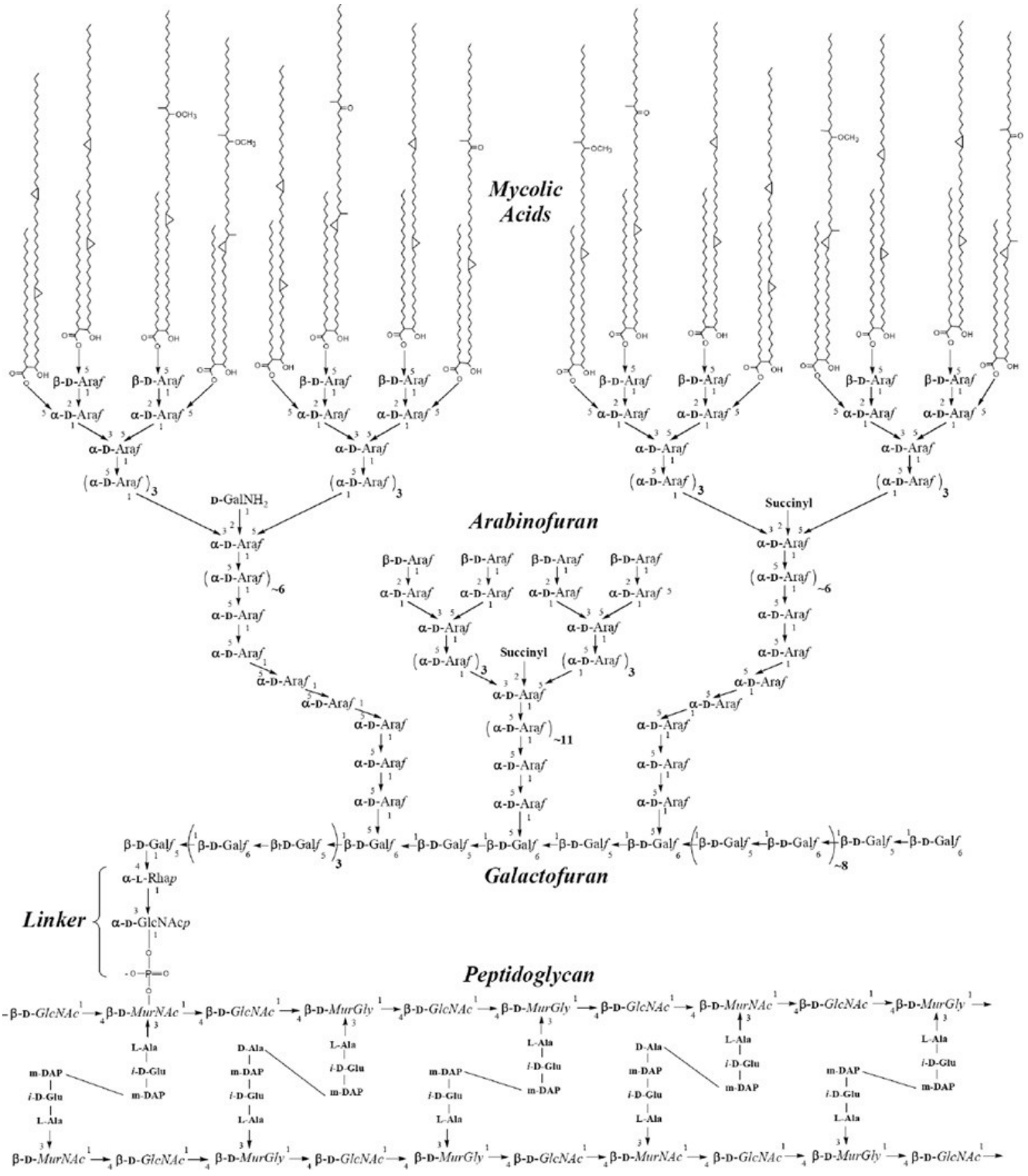


Figure 3. ***Mycobacterium tuberculosis*** **cell wall structure.** It has a unique structure, it is composed of peptidoglycan layer, arabinogalactan layer and mycolic acids. (Figure reprinted from Barry C. E., et al., 2007 [16]).

1.3 Peptide/protein-based Therapeutics

Peptides and proteins are made of amino acids connected with peptide bonds; they are 2-50 amino acids and >50 amino acid in length, respectively. They are a dynamic class of biomolecules with a wide range of applications with respect to drug discovery and therapeutics [29]. In biological systems proteins are notable for their ligand specific binding, for example, enzyme reactions, receptor binding and signaling. They can form specific chemical interactions with other proteins or ligands; therefore, they are of high interest with respect to drug development [30]. Protein based therapeutics, commonly called as 'biologics' have advanced leaps and bounds since the introduction of insulin therapy in 1920s [31]. Ensuing developments in protein expression, purification, and chemical synthesis; progress in peptide/protein-based therapies have been observed with successful drugs reaching the market. This includes growth factors, hormones, engineered antibodies and synthetic peptide drugs (oxytocin, vasopressin) [32] commercialized in the 1980s [33]. During this time, the primary method for discovering bioactive peptides relied on screening natural product libraries. However, naturally occurring peptides may not be suitable for direct administration because of its physical and chemical instability. Nonetheless, we see have witnessed research and the use of peptide-based drugs for addressing various diseases overt the past two decades [29].

Initial studies in peptides as therapeutics agents against pathogenic diseases started with the discovery of naturally occurring antimicrobial peptides (AMPs) [34]. Also known as host defense peptides (HDPs), these are endogenous peptides produced by different organisms including plants, fungi, animals, humans and microbes, against pathogenic infection. AMPs have been widely studied for their antimicrobial properties. AMPs also have anticancer, antifungal, and antiviral activity [35-37]. AMPs have exhibited many advantages over classical antibiotics, for example they are less likely to initiate drug resistance and they have a synergistic effect when used in combination with existing antibiotic drugs. Regardless of these advantages, not many naturally occurring AMPs have made it to market [38]. Of various structures of AMPs reported, most of the successful antibiotics have a cyclic structure, which indicates the importance of

macrocycles in peptide bioactivity [39]. Macrocyclic peptides (MCPs) also make up a major class of peptide-based antibiotic drugs.

Though, peptides are comparatively easy to design and synthesize, larger molecules like proteins have also served as therapeutic agents [40]. Antibodies are widely studied proteins with therapeutic effects, however, their large structure (>50 kDa) presents limitations, hindering cell penetration [41-43]. They are surface binders, thus engineering antibodies for small the catalytic grooves of enzyme active site are difficult. Moreover, most require post-translation modifications making the production process difficult and expensive. As a result, smaller antibody-like molecules were required to overcome said limitations [42, 44]. Rational design approaches have been adapted based on the ligand binding domains of antibodies to design libraries of novel proteins with similar scaffolds or structures. Some examples include, nanobodies (12-15 kDa) known as single domain antibody (variable domain of the heavy chain); monobodies, synthetic proteins much smaller (10 kDa) than antibodies were engineered based on fibronectin III domain; affibodies (<7 kDa), engineered from Z domain of SPA [45-48]. These engineered proteins have been studied in a variety of contexts, from probing and imaging to inhibiting enzymes, thus, providing an array of libraries.

1.3.1 Macrocyclic Peptides as Potential Inhibitors

The discovery of macrocyclic peptide-based natural products has led scientists to explore MCPs as drug candidates and scaffolds. The naturally occurring MCPs form a diverse family of peptides found within a range of organisms such as bacteria, fungi, plants, and mammals [49]. These peptides consist of non-standard amino acids and modified side chains. Cyclosporin A, an immunosuppressant drug, is one such example, extracted from an ascomycete fungus *Tolypocladium inflatum* Gams [50-53]. The macrocyclic backbone imparts structural rigidity and often possess modifications such as *N*-methylation. This increases cell permeability and provides resistance to protease [49, 54]. Among potent biologically active peptides, macrocyclization can occur in a variety of modes, including head to tail cyclisation **(Fig**

4a), formation of multiple macrocycles **(Fig 4b)**, as well as the formation of macrocyclic head with a linear tail **(Fig 4c)**.

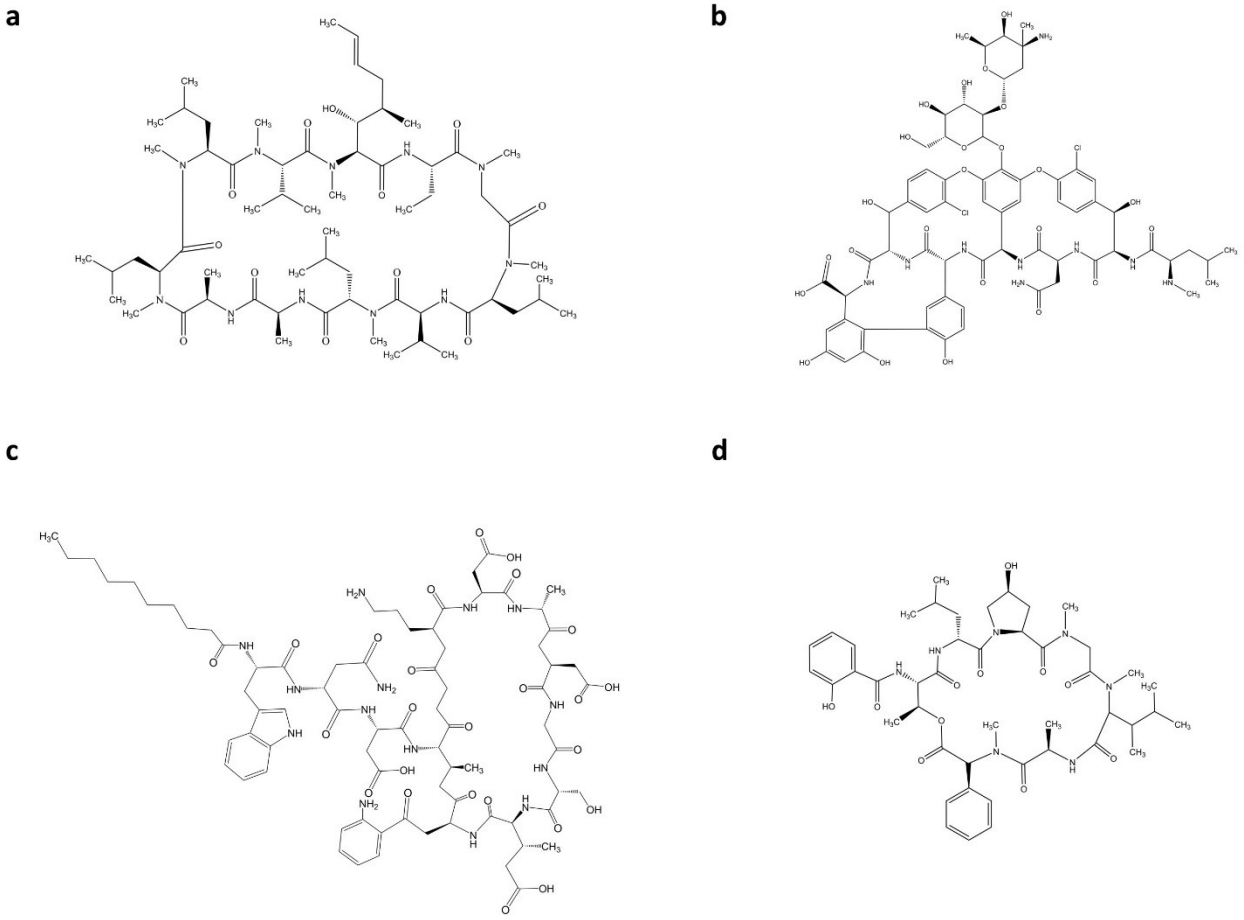


Figure 4. Naturally occurring functional MCP: a. Cyclosporin A isolated from the fungus *Tolypocladium inflatum*; used as immunosuppressant drug b. Vancomycin is a glycopeptide antibiotic produced by the soil bacterium *Amycolatopsis orientalis* c. Daptomycin is a lipopeptide antibiotic used against Gram-positive bacteria isolated from *Streptomyces roseosporus* d. Etamycin A, antibiotic first discovered in *Streptomyces virginiae*

MCPs have been shown to have favorable pharmacological properties compared to standard linear peptides. They are metabolically stable, target-specific, and have high affinity with the ability to modulate protein-protein interactions. These properties make MCPs excellent drug candidates [55, 88]. Traditional approaches in the discovery of antibiotics have mainly involved phenotypic screening of natural compounds. Naturally occurring peptides with macrocyclic structures represents a class of successful antibiotic drugs which includes compounds such as Vancomycin, Etamycin A and Daptomycin, **(Fig 4)** [56-58]. Thus, peptide-based macrocycles have been an inspiration for developing novel antibiotics. Scientists have explored macrocyclization of oligopeptide libraries to improve their bioactivity. In the early nineties, with the advent of high-throughput screening (HTS) of large compound libraries paved the way for screening MCP libraries [59, 60]. Subsequent developments in target-specific drug design has directed discoveries of MCPs binding to distinct targets. Advances in HTS enabled *de novo* screening of functional peptides. Phage display was the earliest techniques used to generate peptide libraries coupled with chemical cyclisation steps [61]. Ribosomal synthesis is carried out in *E. coli* transduced by phage that contains genes for coat proteins fused to randomized peptide. Cyclisation chemistries are performed for the peptides displayed on the phage surface. Such display techniques can then be used to study protein-protein interaction and find high affinity binders for a molecular target. *De novo* discovery of macrocyclic peptides was also explored through *in vivo* processes in *E. coli* called split-intein circular ligation of peptides and proteins (SICLOPPS) [62]. Biosynthesis of cyclic peptide libraries containing 10^8 sequences can be achieved with this method. In recent years, cell-free techniques such as mRNA display combined with flexible *in vitro* translation (FIT) have been widely applied to screen genetically encoded peptide libraries (section 1.4) [63-67]. Researchers have reported that MCPs successfully inhibit protein-protein interactions. A study detailed that an MCP discovered using the FIT-RaPID strategy selectively binds and inhibits hepatocyte growth factor, a protein involved in cancer progression [68]. The same technique has been applied to find inhibitory MCP against Ebola viral protein 24 (VP24) [69]. We hypothesize that, screening MCP libraries for potential inhibitors against UGM can lead to the development of novel therapeutics against TB.

1.3.2 Affibodies as Ligands

Affibody molecules are engineered small proteins comprised of three alpha helices bundled together with molecular weight approximately 6.5 kDa **(Fig 5)**. These small proteins have been largely studied as novel and selective binders since their introduction in 1995 by Nord and group [70]. During this time, phage display was a well-established method used to screen peptide libraries for target-specific binding [48, 70]. The construction of affibodies was carried out to create new ligands for biotechnology applications and to serve as "artificial antibodies". They can fold easily in reducing environments of the cell cytoplasm because of smaller structure while resembling antibody-like affinity and binding. The scaffold is based on the synthetic Z domain of Staphylococcus protein A (based on B domain of SPA) binding to Fc domain of immunoglobulin G. Affibodies have been used as high affinity ligands and have many applications like fusion proteins (for e.g., affibody fused with F_c domain of antibody), imaging, and detection [71, 72]. A recent study highlights selection of target-based affibody using the phage display platform, where affibody ligands were selected for extracellular domain of human epidermal growth factor receptor 2 (HER/neu) [73]. The selected affibodies were successfully used in imaging tumors in mice [74, 75].

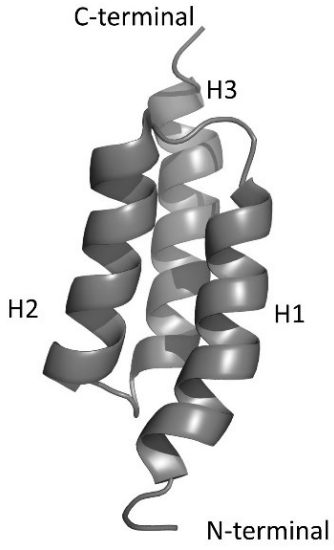


Figure 5. Structure of affibody: synthetic Z domain of SPA. It consists of 58 residues that form three α helices (H1, H2, and H3) which are stacked together.

1.4 *In vitro* Display Technique for Selecting Peptide/protein Ligands: mRNA Display

Display techniques are effective strategies for screening polypeptide libraries. The principle for all display techniques is, that the genotype coding sequence is physically linked with the encoded phenotype, the corresponding peptide sequence. This enables for selection of specific binders among trillions of clones. The first display technique to screen peptide library for binding was phage display [61]. Since then various display techniques have been developed to assist with finding peptide binders. The technology of display systems has evolved from cellular approaches such as phage display, yeast display, and bacterial display to *in vitro* display systems [76, 77]. *In vitro* display technologies have proven advantageous compared to *in vivo* display systems. They have emerged as innovative approaches that allow for screening of large diversity of clones in a cell free environment, thus, eliminating biological constraints of a cell [78]. *In vitro* display systems include mRNA display, ribosome display, and DNA based displays [79-82]. The mRNA display technique is used to select for peptide/protein binders against a target through evolution of the library. An mRNA-peptide fusion is created through the conjugation of puromycin to the mRNA enables the linkage of translated peptide/protein to its coding mRNA. Puromycin mimics the 3' end of an amino-acyl tRNA which forms a covalent bond with the nascent peptide. To stabilize the mRNA in the mRNA-peptide complex, a DNA strand is synthesized via reverse transcription forming RNA-DNA hybrid while still being attached to the peptide. The target for the display is immobilized on a bead, and the fusion library is screened for binders. Each selection round ends in washing of unbound peptides followed by PCR amplification of the corresponding cDNA (of the bound peptides), thus, enriching the library with clones specific for the target. The major advantage of genotype linkage with the presented peptide scaffold is that it is possible to isolate and amplify the positive clones [79, 80] **(Fig 6)**.

mRNA display has been adapted for screening of MCP libraries. The display process constitutes of *in vitro* translation system that incorporates non-canonical amino acids required for the cyclisation of the peptides. The genetic code is reprogrammed using a tRNA-priming catalyst called flexizyme. It is a specialized *de novo* ribozyme that can charge tRNA with non-canonical amino acids, thus increasing the

diversity of peptide library [64]. Flexizyme along with the recombinant translation system can help generate macrocyclic peptide libraries which can then be used for discovery of MCP ligands, this is referred to as the flexible *in vitro* translation system (FIT). Random non-standard peptide integrated discovery (RaPID) system uses FIT system to discover MCPs for the given target. MCPs discovered through this process have high specificity for the target and have been used in wide range of applications. It has been successfully used to discover MCPs with inhibitory activity against enzymes [65-68].

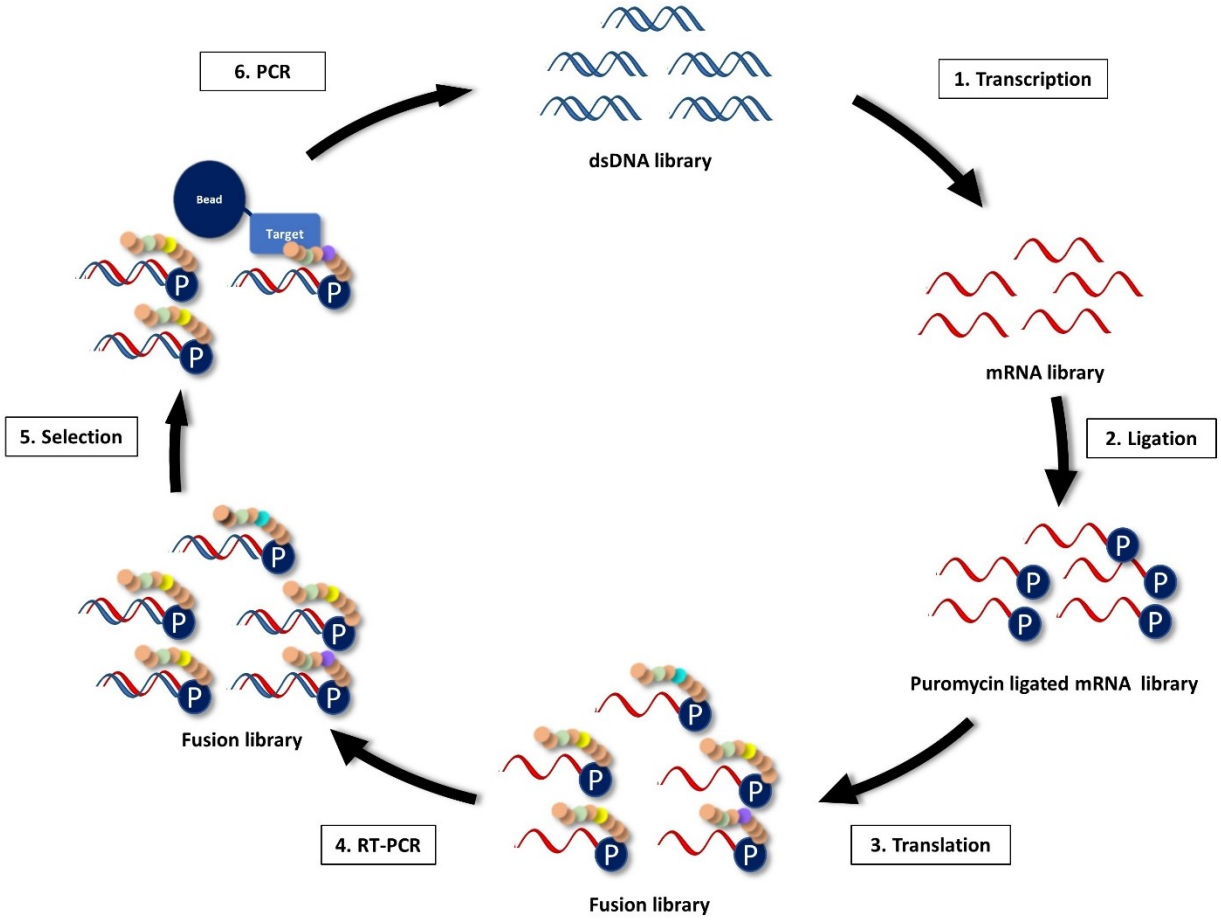


Figure 6. *In vitro* Selection of peptide binders using mRNA display technique. The process starts with a library of dsDNA, it is then transcribed using *in vitro* transcription machinery, the mRNA is then ligated with puromycin to create puromycin-ligated mRNA library. This is followed by *in vitro* translation to generate a fusion library of mRNA linked with its coded peptide/protein. This fusion library of mRNA-peptide is reverse transcribed to make RNA-DNA hybrid linked to its corresponding peptide. It is this library, that is used for the selection of binders (peptides/proteins that bind to the target). The target is immobilized on affinity beads and the library is introduced for binding, followed by several washing step to remove unbound proteins. The cDNA of the bound peptides is recovered, PCR amplified and used as the input library in the next round. This process is repeated, until a significant enrichment of clones is observed.

1.5 Rationale

1.5.1 Targeting Cell Wall Biosynthesis in *Mycobacterium tuberculosis*

The literature study and success of peptide/protein-based therapeutics has inspired us to screen artificial genetically encoded libraries of macrocyclic peptide and affibodies for discovery of potential inhibitors of *Mt*UGM. The cell wall biosynthetic process can be targeted to tackle *Mtb* [21]. To do so, we have chosen *Mt*UGM as our target. The enzymatic activity catalyzed by *Mt*UGM does not exist in mammals, therefore, potential drugs against the enzyme will be specific for the pathogen making it an excellent target [83]. With inhibition of *Mt*UGM, UDP-Gal*f* will not be available for priming the galactan layer of the *Mtb* cell wall, subsequently affecting the assembly of the layers above. A dysfunctional cell wall affects multiplication, reduces pathogenicity, and makes the bacteria susceptible to other drugs. We believe that potential peptide inhibitors can act synergistically with the existing drug regimen. Numerous antibiotics targeting the cell wall biosynthesis of bacterial pathogens have been discovered since the first use of penicillin in 1930. However, we have witnessed antibiotic resistance towards these drugs, making treatment of bacterial diseases difficult. Such is the case with tuberculosis. Though being mostly eradicated in the North Americas, the reported TB cases as of 2018 cover 3% of the of the total incidences, including cases of resistant TB. Novel therapeutics would thus be beneficial [1, 11].

1.5.2 Artificial Genetically Encoded Libraries of Macrocyclic Peptides and Affibodies

Artificial genetically encoded peptide libraries were designed in earlier research by Dr. C. J. Hipolito. Construction of an MCP library containing up to 10^{12} clones was done using the degenerate codon NNK to incorporate random amino acids. The peptides to be displayed on mRNA are 15 amino acid in length and, in place of *N*-formylmethionine, have an *N*-terminal *N*-(2-chloroacetyl)-L-phenylalanine (introduced via a semisynthetic tRNA) to facilitate macrocyclization with the first occurring cysteine in the chain. The peptides each have a $(Gly\text{-}Ser)_3$ linker at the C-terminal, at which they can be attached to mRNA for display. This library of MCPs was used in the FIT-RaPID technique to select binders for *Mt*UGM. The

affibody library was generated using the original 58 residue SPA-derived Z domain with 13 degenerate positions.

1.5.3 mRNA Display

The target *Mt*UGM was expressed and purified as fusion protein with MBP bearing an *N*-terminal His_6 tag for protein immobilization, the selection was performed against His_6-MBP-*Mt*UGM (with His_6-MBP used for negative selection to remove unspecific binders). The mRNA display technique involves *in vitro* transcription that allows the synthesis of a short mRNA sequence to be synthesized from a DNA template in a cell-free environment [79]. This is followed by the unique *in vitro* translation system called FIT; a genetic code reprogramming technique based on the combined use of flexizyme-catalyzed aminoacylation with a purified recombinant translation system allowing the incorporation of non-canonical amino acids. FIT has been applied to the synthesis of proteins containing non-canonical residues and, because it allows modification of proteins at an atomic rather than a residue level, is a powerful tool for the analysis of protein structure–function relationships [64]. The FIT reaction leads to the synthesis of a library of macrocyclic peptides, each of which is covalently linked through the puromycin moiety to its cognate mRNA. Each selection round is followed by PCR amplification thus, enriching clones of positive binders. The binding ability of individual clones can be determined by using single-clone display (FIT and RaPID system in mRNA display of a single sequence rather than a library). Real time polymerase chain reaction is a method to efficiently quantify the DNA in real time. This method was employed to quantify the recovered cDNA from each selection round from the display. This enables us to quantify the peptides that are positive or negative binders. Next-generation sequencing was used to sequence the mRNA library after each selection round. MCPs discovered by these means can be chemically synthesized using standard Fmoc solid phase synthesis. Selection of affibodies through mRNA display will differ from the FIT-RaPID system. *In vitro* translation of affibody mRNA library does not involve use of unnatural amino acids nor a macrocyclization step, but otherwise follows the same steps as MCP selection. The selected affibody can then be recombinantly expressed and purified for assays with *Mt*UGM.

1.5.4 Development of HPAEC Activity Assay

For testing the potential inhibitory activity of discovered ligands, in our lab we reproduced a previously described assay to detect the enzyme activity of *Mt*UGM. We know that *Mt*UGM catalyzes the interconversion of isomers UDP-Gal*p* and UDP-Gal*f*, and high-performance anion-exchange chromatography (HPAEC) can be used to detect the enzyme activity by distinguishing and quantifying these isomers at low levels [83, 84].

1.6 Objectives

The broader goal of the project is to develop and find novel therapeutics for treating TB, by targeting the process of cell wall biosynthesis in *Mtb*. This will be achieved by developing peptide/protein based potential inhibitors of *Mt*UGM. Towards this goal, we have employed an *in vitro* selection technology that can be adapted to select ligands from two libraries, MCPs and affibodies using mRNA display. Ligands discovered through this process will be screened for its inhibitory activity and may be used for studying substrate binding interactions which can provide a basis for developing potential drug candidates.

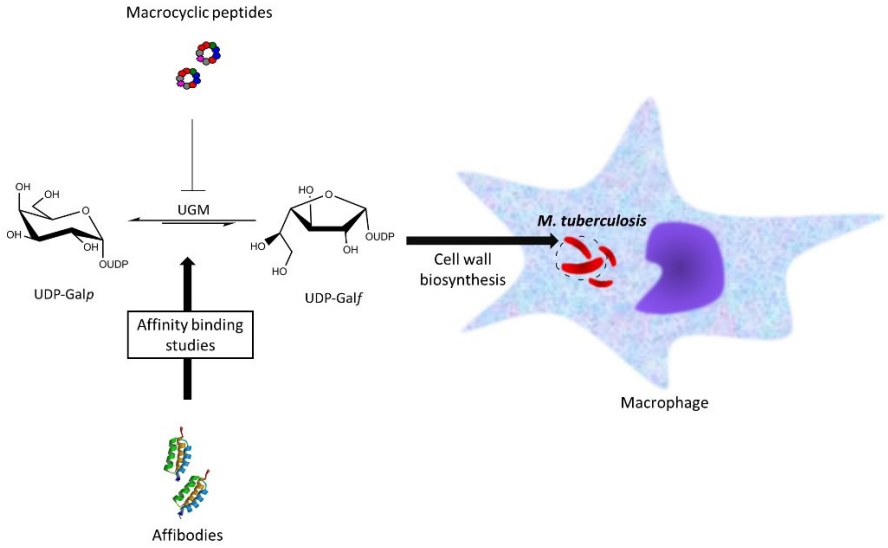


Figure 7. Overview of the project. The aim of this project is to identify the peptide/protein ligands that can inhibit *Mt*UGM thereby disrupting the cell wall biosynthesis process in *Mtb*. Discovered affibodies will be used as a tool to study affinity binding interactions.

CHAPTER 2: MATERIALS AND METHODS

2.1 Recombinant Expression and Purification of Proteins

2.1.1 Recombinant Expression of Proteins

For all protein expressions, the recombinant plasmid was expressed using BL21 (DE3) *E. coli* cells unless specified. The plasmid was isolated from overnight grown culture of DH10b using ThermoFisher GeneJet column Miniprep Kit and expressed in chemically competent BL21 (DE3) *E. coli* cells. The transformants were selected on LB media supplemented with 100 µg/mL ampicillin agar plates. Seed culture was prepared, transformants were inoculated into 20 mL LB media supplemented with 100 µg/mL ampicillin and grown overnight at 37 ˚C in shaking incubator. Seed culture was prepared, transformants were inoculated into 20 mL LB media supplemented with 100 µg/mL ampicillin and grown overnight at 37 ˚C in shaking incubator. 10 mL of seed culture wasp inoculated into an 800 mL LB media supplemented with 100 µg/mL ampicillin and grown at 37 ˚C in shaking incubator. This main culture was induced with IPTG to a final concentration of 1 mM at log phase (OD_{600nm} = 0.4-0.6). Induced culture was grown overnight, at 15˚C in shaking incubator.

NOTE: The choice of antibiotic will depend on the expression plasmid used.

2.1.2 Immobilized Metal Affinity Chromatography Purification Using Ni-NTA Column

For protein purification of his-tagged protein, the overnight grown cultures were harvested by centrifuging at 4200g, the pellet was resuspended in 15 ml of binding buffer/resuspension buffer, DNAse I, RNAse A and lysozyme were added to a final concentration of 5 µg/mL with the addition of one Roche protease inhibitor tablet per 30 ml of resuspended cells. The cells were lysed by sonication and centrifuged to remove cell debris. The crude lysate was filtered sterilised using 22 µM filter and applied onto amylose resin column preequilibrated with binding buffer. Protein fractions were eluted by AKTA-FPLC with elution buffer. The fractions were analysed on SDS-PAGE. The His_6-MBP-*Mt*UGM fractions were pooled

and concentrated using Vivaspin 6 (30,000 M.W.) concentrator spin columns and stored in storage buffer. The final concentration of purified protein was determined by Bradford's assay or BCA assay. Small aliquots of purified protein were flash frozen using liquid nitrogen and stored at -80˚C.

NOTE: Each protein uses specified buffers suitable for its purification and storage **(Table A1)**. The IPTG induction conditions vary for each protein.

*Mt*UGM

The pDEST-His_6-MBP-*Mt*UGM was obtained from Dr. David Sanders, University of Saskatchewan. It encodes wild-type *Mt*UGM fused with maltose-binding protein (MBP) for improved solubility and an N-terminal His-Tag used for affinity purification [84]. Protein expression was carried out as per section 2.1.1. The protein purification was carried out as per section 2.1.2 with modifications to increase yield. After IPTG induction, the cells were incubated in 18°C, harvested after 26 hours, resuspended in *Mt*UGM binding buffer. Amylose column was used for affinity purification, fractions were eluted with MtUGM elution buffer and stored in *Mt*UGM storage buffer. Final concentration of purified TEV protease was determined by Bradford′s assay. This fusion protein was used as a target for selection of ligands in mRNA display, where the histidine tag was used to immobilize the protein on magnetic beads.

TEV Protease

The pRK793 encodes S219V mutant TEV protease fused with maltose-binding protein (MBP) for improved solubility and an N-terminal His-Tag used for affinity purification [85, 86]. The recombinant plasmid was designed and deposited on Addgene by Dr. S. Waugh [86]. The plasmid was transformed into BL21 (DE3), transformants were selected on LB media supplemented with 100 µg/mL ampicillin and chloramphenicol agar plates. Expression was carried out as per section 2.1.1 and cells were cultured in LB supplemented with 100 µg/mL ampicillin and chloramphenicol. The harvested cells were resuspended in 15 ml of TGI buffer. Proteins purification was carried out as per section 2.1.2. with TEV protease elution

and storage buffer. Final concentration of purified TEV protease was determined by BCA assay. This TEV protease was then used to cleave MBP tag from His_6-*Mt*UGM-MBP for enzyme activity assay.

2.2 Selection of Macrocyclic Peptides Against UGM Using the FIT and RaPID system

2.2.1 Library Construction

The NNK15 DNA library of MCPs was designed according to the protocol [67]. The mRNA library was transcribed, and it encode 15 NNK (where N = A, U, G, C; K = U/G) positions sequentially to incorporate 15 random amino acids. The peptides have an N-terminal N-(2-chloroacetyl)-L-phenylalanine (introduced via a semisynthetic tRNA) instead of N-formylmethionine to facilitate macrocyclization with the first occurring cysteine in the chain. The NN13 affibody DNA library is based on the template of Z domain of SPA with 13 degenerate positions. The peptides each have a $(Gly\text{-}Ser)_3$ linker at the end, at which they can be attached to mRNA for display. The libraries were amplified using the primers mentioned in **Table A2**.

2.2.2 Preparation of N-(2-chloroacetyl)-L-phenylalanine-$tRNA^{fMet}_{CAU}$ and Flexizyme eFx

Initiator $tRNA^{fMet}_{CAU}$ and flexizyme dsDNA was amplified using primer extension by PCR. To amplify $tRNA^{fMet}_{CAU}$ 0.5µM fMetE-F, 0.5µM fMetE-R_2, and 5 ηM fMetE-R_1 was used in the final reaction mixture for PCR. Amplification of flexizyme dsDNA was done using 0.5µM eFxF66 and 0.5 µM eFx-AmC-R18 in the final reaction mixture. Phenol-Chloroform-isoamyl alcohol extraction and precipitation was performed to extract dsDNA and visualized on 3% agarose gel. *In vitro* transcription was carried out using NEB's HiScribe T7 Quick Yield RNA synthesis kit, the reaction was incubated overnight in the air incubator at 37 ˚C. $tRNA^{fMet}_{CAU}$ and flexizyme eFx was purified by denaturing PAGE with 8% polyacrylamide gel. Purified RNA was standardized to 250 µM in the final concentration. Initiator $tRNA^{fMet}_{CAU}$ (5.25 nmol) was charged with either *N*-(2-chloroacetyl)-L-phenylalanine or charged onto according to literature protocols. Charged tRNAs for the first rounds of the sub-selections were stored at −80 °C as dry pellets.

2.2.3 Protein Immobilization on Magnetic Beads

Invitrogen (Grand Island, NY, USA) Dynabeads® His-Tag Isolation & Pulldown magnetic beads (40 µg/µl) was saturated with His_6-MBP-*Mt*UGM (100 pmol/µl of beads) in *Mt*UGM selection buffer (500 mM NaCl, 25 mM Tris-Cl pH 7.5, 0.005% TWEEN 20) for positive selection. His_6-MBP (50 pmol/µl of beads) was also partially saturated on the beads to perform negative selection. The protein-bead mixtures were incubated with gentle rotation at room temperature for 20 min. After incubation, the beads were washed three times with *Mt*UGM selection buffer to remove unbound proteins. For all selection rounds, His-tagged *Mt*UGM (100 pmol) and His-tagged MBP was used to saturate his-tag binding sites on Invitrogen Dynabeads® His-Tag Isolation & Pulldown magnetic beads (40 µg) in selection *Mt*UGM buffer for positive and negative selection respectively.

2.2.4 Preparation of Puromycin-mRNA Conjugated Library

A ligation reaction mixture (1x T4 RNA ligase buffer, 7.5 µM puromycin linker, 10 µM mRNA library, 60 Units of T4 ligase, 20 µL) was incubated at room temperature for 3 hours. The reaction was quenched with a stop solution (0.6 M NaCl, 10 mM EDTA, pH = 7.5, 20 µL). The solution was extracted with 25:24:1 phenol–chloroform–isoamyl alcohol (40 µL) followed by a second extraction with chloroform–isoamyl alcohol (40 µL). The mRNA-puromycin conjugates were precipitated with the addition of ethanol (80 µL) to the aqueous phase. The resultant precipitate was pelleted by centrifugation (15,000g, 15 min), washed with 70% ethanol (20 µL) followed by centrifugation (15,000g, 3 min) and air-dried. The precipitate was dissolved in water (3 µL). Ligation efficiency was analysed on 8% polyacrylamide gel, stained with ethidium bromide, and visualized under UV.

NOTE: All concentrations and volumes are 1x; they are subject to change depending on the selection round parameters

2.2.5 Flexible *In vitro* Translation of Puromycin-mRNA Conjugated Library

NEB's *in vitro* translation system with customised reagents (NEB Solution A without amino acid and tRNA, NEB solution B without RF) called PURExpress® *in vitro* protein synthesis kit was used. The flexible *in vitro* translation mixture was set up with 19 proteinogenic amino acids, charged *N*-(2-chloroacetyl)-L-phenylalanine-$tRNA^{fMet}_{CAU}$, flexizyme eFx and 10 µM puromycin-mRNA library. It was incubated at 37˚C for 30 minutes followed by incubation at 25˚C for 12 minutes to facilitate cyclisation of the peptides. To stabilize the mRNA-peptide conjugate, mRNA was reverse transcribed to form cDNA-mRNA hybrid using Promega reverse transcriptase (-H). The reaction was carried out at 42 ˚C for 1 hour. Selection buffer was added to this reverse transcription mix in 1 : 6 ratio and de-salting was done by passing the translation reactions through Sephadex G-25 (beads swollen in *Mt*UGM selection buffer for at least 3 hours) columns that were pre-washed and equilibrated with *Mt*UGM selection buffer (500 mM NaCl, 25 mM Tris-Cl pH 7.5, 0.005% TWEEN 20). The library is then used for selection.

2.2.6 Selection of *Mt*UGM Binding Macrocyclic Peptides

2x blocking solution (0.2% acetylated BSA in selection buffer, 300 µL) was added to desalted peptide-mRNA solution. Preclear with magnetic beads partially saturated with His-tagged MBP was used to remove peptides that bind to His_6-MBP or the beads. The mixture was incubated at 4 ˚C for 10 minutes with gentle rotation. This step was repeated three times, where each time the library was added to a new microfuge tube. The supernatant was separated from the beads and 0.5 µL of supernatant was used for reverse transcription and real-time PCR to measure the concentration of the input peptide-mRNA. From this, 300 µL of peptide-mRNA library was added to magnetic beads immobilized with His_6-MBP-*Mt*UGM and incubated at 4 ˚C for 1 hour with gentle rotation. For negative selection, the remaining peptide-mRNA library was added to magnetic beads immobilized with His_6-MBP and incubated at 4 ˚C for 1 hour with gentle rotation. The supernatant was removed after the incubation period. The beads were washed three times with ice cold selection buffer to remove unbound mRNA-peptides. The beads were stripped by adding

polymerase free PCR reaction mix (total volume 50 ul) using primers CGS3an13.R39 and T7-CH-F46, incubated at 95 ˚C for 5 minutes. For quantification of the peptide-mRNAs recovered by bead binding, 1 µL of this mixture was removed for use in real-time PCR. The remaining of the solution was thermo cycled 16 × (95 °C/40 s, 50 °C/40 s, 72 °C/40 s). The PCR product was extracted with one equivalent volume of 25:24:1 phenol–chloroform–isoamyl alcohol followed by extraction using one equivalent volume of 24:1 chloroform–isoamyl alcohol. A solution of 3M NaCl was added to the aqueous phase and the DNA product was precipitated with the addition of ethanol. Subsequent rounds of selection were performed with modification. Total five rounds of selection were done. The DNA molecules obtained after the 5th round of both libraries were ligated into the plasmid pGEM-T Easy using TA-cloning. Individual clones were selected at random and sequenced by standard sequencing. DNA molecules obtained after each round were transcribed and deep sequence analysis was performed using Next-Gen Mi-seq.

2.2.7 Single Clone Assay

The binding ability of individual clones of peptides was determined by doing the selection process using individual peptide-mRNA and the recovered cDNA was quantified using qPCR.

NOTE: All primers are listed in **Table A2.**

2.3 Selection of Affibodies That bind to *Mt*UGM Using mRNA Display

The selection process for affibodies differs from that of MCPs. The FIT system is not applied to this process because, the affibodies include all the 20 proteinogenic amino acids. The above-mentioned steps in sections 2.3.3 to 2.3.6 were followed for selection of affibodies against *Mt*UGM.

2.4 Chemical Synthesis of Selected Macrocyclic Peptides

MCPs were synthesised using standard FMOC solid phase synthesis. The N-terminus was protected with FMOC (9-fluorenylmethoxycarbonyl) group prior to deprotection with piperidine. This was followed by a coupling reaction that added the incoming amino acid and reprotection. This was repeated until the

linear peptide chain synthesis was complete. The peptide was removed from the resin by treatment with trifluoroacetic acid (TFA). One-step purification by reverse-phase HPLC was performed to obtain purified peptide. The fractions were confirmed for the desired peptide mass using MALDI-TOF mass spectrometry and lyophilised.

2.5 Amplification and Bacterial Cloning of Affibody Genes

The top four affibodies as per mRNA display final round were selected for recombinant expression. For cloning into pET21a-GFP the genes were designed to have BglII at 5' end followed by the T7 promoter, affibody gene, $(Gly\text{-}Ser)_3$ linker, TEV protease recognition site and 3' NdeI restriction site. The gene synthesized by Twist Bioscience. Affibody sequences were PCR amplified using primers AFR6-F and AFR6-R, purified by using Bio Basic PCR purification kit. Amplified affibodies were cloned into pET21-GFP by restriction digestion of said restriction sites. The recombinant plasmid encoded His-tagged affibody tagged with GFP, with a TEV protease recognition site in between. DH10b cells were transformed with the recombinant plasmid and plated on LB-agar supplemented with 100 µg/mL of ampicillin. Six colonies were selected arbitrarily from the four plates and inoculated in 5 ml culture, grown overnight at 37 ˚C in shaking incubator. To ensure proper cloning plasmid isolated from these overnight cultures were screened by colony PCR (primers: AFR6-insert-F and AFR6-insert-R). Additionally, standard sequencing was performed to confirm successful insertion of affibody genes, sequences were aligned in SnapGene.

NOTE: All primers are listed in **Table A2**.

2.6 Recombinant Expression and Purification of Affibodies (AFR6-1,4)

The recombined plasmids (section 2.6) isolated from overnight cultures of DH10b cells. Affibodies were expressed and purified as per section 2.1.1 and 2.1.2. Affibody elution buffer and storage buffer were used for the process. **(Table A1)**. Concentration of purified affibody-GFP fusion protein was determined by Bradford's assay.

2.7 Enzymatic Assay using High-Performance Anion-Exchange Chromatography

Digested His_6-MBP-*Mt*UGM was used for enzymatic assay. The dual tag was removed via TEV protease cleavage. The digestion mixture was setup with 1:10 ratio His-tagged TEV protease to fusion protein; incubated overnight at 4˚C. The His-tagged TEV protease and His_6-MBP were removed using IMAC by loading the digested fusion protein onto Ni-NTA column. Flow-through fractions containing untagged *Mt*UGM were collected and analysed on SDS-PAGE. Final concentration was determined by Bradford's assay. Enzymatic activity of purified *Mt*UGM was determined using high proficiency anion-exchange chromatography by monitoring the conversion of UDP-Gal*p* to UDP-Gal*f* at 262nm. The final reaction mixtures included a fixed amount of enzyme *Mt*UGM (200 nM) and varying amounts of substrate UDP-Gal*p* (25-200µM), in 50 mM sodium phosphate buffer (pH 7) with 20 mM freshly prepared sodium dithionite to a final volume of 100 µl. The control reaction contained only UDP-Gal*p*. The reaction was carried out for 2 min at 37˚C and quenched with equal volume of n-butanol. The aqueous layer, containing UDP-Gal*p* and UDP-Gal*f*, was obtained by centrifugation and 10 µl was injected into CarboPac PA1 HPLC column. The sugars were eluted isocratically with 200 mM ammonium acetate (pH 7) buffer with a flow rate of 0.5mL/min; run time was set at 20 min. The area under the curve and peaks were integrated using Agilent Chemstation software. The % conversion of UDP-Galp to UDP-Galf was calculated using the following formula:

$$\%\ conversion = \frac{Area\ of\ UDP\text{-}Galf\ peak}{Area\ of\ UDP\text{-}Galp\ peak + Area\ of\ UDP\text{-}Galf\ peak} \times 100$$

CHAPTER 3: RESULTS

3.1 Selection of *Mt*UGM binding MCPs and Affibodies

*Mt*UGM plays a key role in the assembly of the cell wall in *Mtb*, it is responsible for the interconversion of sugar isomers UDP-Gal*p* and UDP-Gal*f*. To find binders specific for *Mt*UGM we recombinantly expressed and purified his-tagged, MBP-fused *Mt*UGM. The his-tag is important for attaching the enzyme on magnetic beads for carrying on the selection process. We selected two ligand libraries for mRNA display the affibodies and MCPs. The selection process is same except the display of affibodies does not require FIT.

3.1.1 Construction of plasmids encoding mutants of *Mt*UGM

The enzyme has been co crystallised with its substrate and inhibitors giving insights about the binding pockets. Dr. David Sanders' lab has successfully expressed *Mt*UGM and elucidated the crystal structure [84]. Originally, pDEST-His_6-MBP-*Mt*UGM was constructed via gateway cloning **(Fig 8a)**. Based on the crystal structure we decided to create point mutants of the wild type enzyme. These mutations are made in the active site of *Mt*UGM. These residues are important for making interactions with the substrate at the active site and change in which will affect the binding efficiency [87]. The active site mutants can be used in the negative selection (section 3.1.8) to remove enzyme surface binders. Two mutants were constructed, 3F*Mt*UGM **(Fig 8b)** with three mutations in the active site and 7x*Mt*UGM **(Fig 8c)** with seven mutations in the active site.

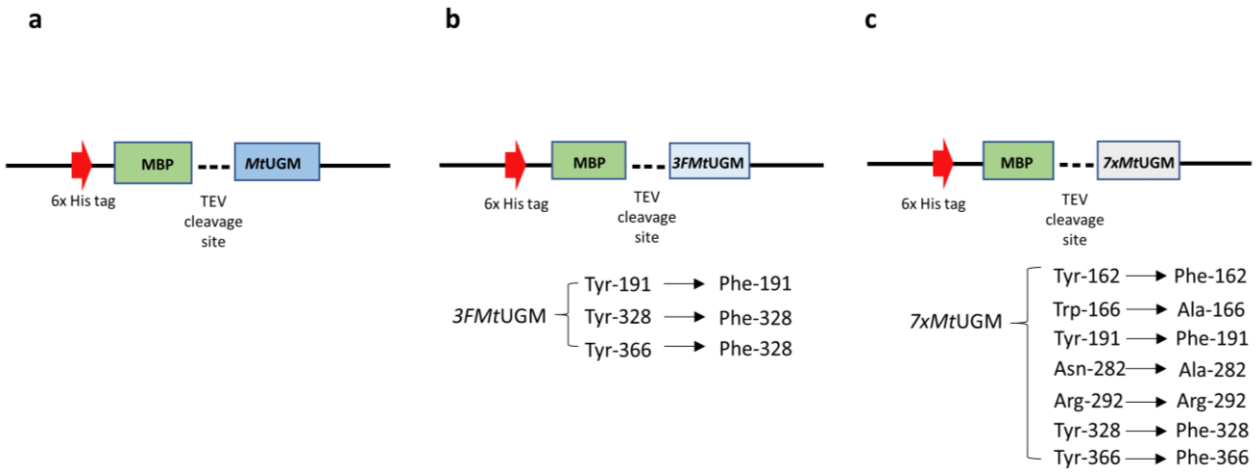
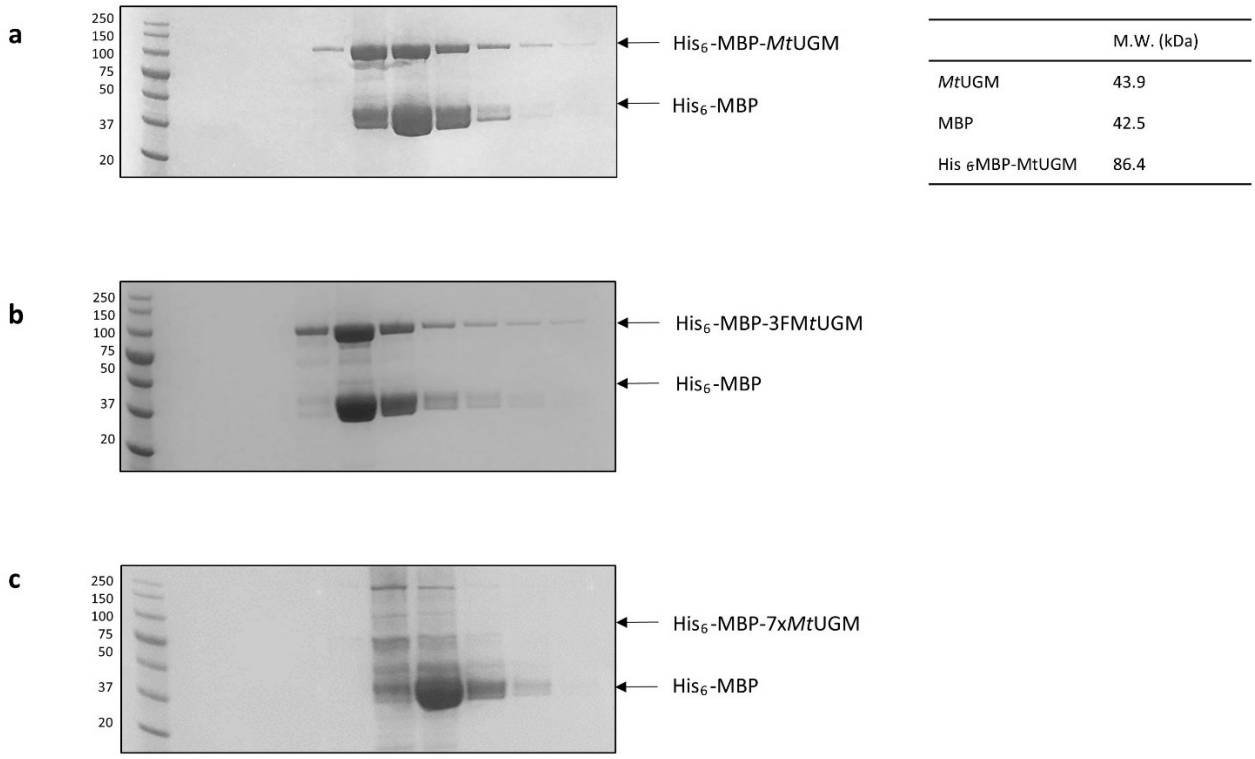


Figure 8. Plasmid constructs for expressing the target protein *Mt*UGM as a his-tagged, MBP fusion(a) pDEST-His_6-MBP-*Mt*UGM and the construction mutants (b) 3F*Mt*UGM and (c) 7x*Mt*UGM.

3.1.2 His-tagged, MBP-fused MtUGM and *Mt*UGM

The recombinant plasmid encoding *Mt*UGM as a His-tagged MBP fusion protein was obtained from Dr. David Sanders. pDEST-His_6-MBP-MtUGM **(Fig 9a)** is encoded with a TEV cleavage site between the dual tag and *Mt*UGM sequence. The recombinant His-MBP-*Mt*UGM fusion protein was purified by affinity chromatography. The MBP tag of the fusion protein binds to the amylose resin and was eluted with elution buffer containing maltose. SDS-PAGE analysis of purified protein fractions is shown in **Fig 9**. The His_6-MBP-*Mt*UGM fusion protein was purified and was observed at ~87 kDa **(Fig 9a)**. The second band at ~43 kDa presumably corresponds to His-MBP. It is possible that the dual tag had cleaved prior to or during the purification process or was expressed as a prematurely truncated protein. We successfully obtained, a final concentration of 0.5 mg/mL, determined by BCA assay, yielding a total of 1.5mg of fusion protein from 800ml of culture. However, we optimised the protocol to obtain the purest fraction of His_6-MBP-*Mt*UGM by pooling the single-band fraction, the concentration of which was 0.3 mg/mL in a total volume of 2ml. Expression and purification of the mutants followed the same procedure as the wild type. The SDS-PAGE gel for mutants 3F*Mt*UGM **(Fig 9b)** show that expression and purification was successful and ~1.5 mg of

protein was obtained from 800ml culture. The expression of 7x*Mt*UGM **(Fig 9c)** was not successful, it is possible that the protein folding efficiency was affected due to the mutations.



	M.W. (kDa)
*Mt*UGM	43.9
MBP	42.5
His 6MBP-MtUGM	86.4

Figure 9. SDS-PAGE analysis for recombinant expression of *Mt*UGM (a) His_6-MBP-*Mt*UGM (b) His_6-MBP-3F*Mt*UGM and (c) His_6-MBP-*Mt*UGM.

The recombinant plasmid pRK793 encodes a MBP-His_6-TEV for improved expression of soluble protein, however the fusion of the TEV protease is self-cleaving at a sequence following the MBP portion, leaving an active His_6-TEV that was purified via IMAC affinity chromatography **(Fig A4 a)**. Elution fractions were analysed on SDS-PAGE for size and purity. MBP-His_6-TEV fusion protein has theoretical molecular weight of ~70 kDa whereas cleaved His_6-TEV protease has a molecular weight ~30 kDa. **Fig A4 b** shows the successful purification of His_6-TEV protease. Despite the presence of uncleaved fusion protein,

catalytically active TEV protease was obtained. A final concentration of 1.77 mg/mL was determined by BCA assay and a total of 5.30 mg TEV protease was obtained from a 2 L culture.

The protocol followed to obtain purified *Mt*UGM from His_6-MBP-*Mt*UGM by TEV protease, has been previously established. To obtain pure *Mt*UGM, the His_6-MBP-*Mt*UGM protein was digested with TEV protease, which cleaves a recognition sequence following the MBP portion. After digestion, the cleaved fusion protein, resulted in a His_6-MBP (~42.5 kDa) fraction and a purified *Mt*UGM (~45 kDa). This mixture was filter sterilised, applied to N-NTA column and IMAC was performed. The untagged *Mt*UGM was eluted in the flow-though fractions, and His_6-MBP along with TEV protease were eluted with buffer (20 mM Tris, 250 mM NaCl, 220 mM Imidazole, pH 7.5). However, we observed another band at ~55 kDa, the reason for which is unknown **(Fig 10)**. Nonetheless, A final concentration of purified *Mt*UGM after TEV digest was determined to be 0.278 mg/ml by BCA assay.

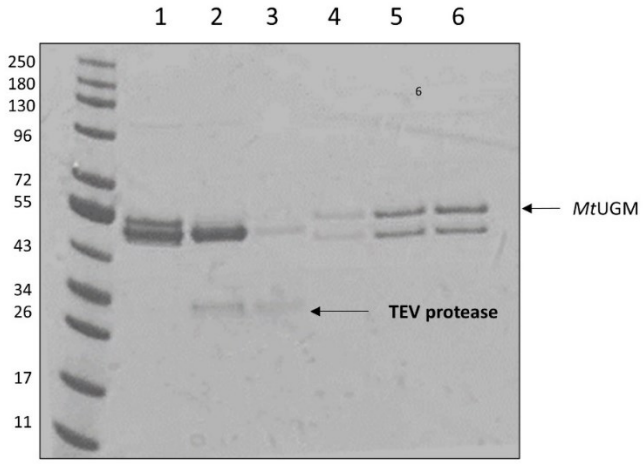


Figure 10. SDS-PAGE analysis of *Mt*UGM digested with TEV protease. Lanes 1, 2 and 3: fractions containing His_6-MBP and TEV protease eluted with imidazole. Lanes 4, 5 and 6: flow-through fractions with untagged *Mt*UGM.

3.1.3 Immobilisation of His_6-MBP-*Mt*UGM and His_6-MBP on Dynabeads.

To carry out the selection process, we immobilized the His_6-MBP-*Mt*UGM selection target and His-MBP counter-selection target. The latter was used for negative selection (preclearing) to remove non-specific peptides since our goal was to identify binders to *Mt*UGM. His-tag isolation magnetic beads were used for the immobilization. We determined the concentration of proteins that was ideal to saturate the bead surface by observing the beads coated with the protein and the supernatant on SDS-PAGE gels. We tested different concentrations and the results suggest that for our His_6-MBP counter-selection target (~43kDa), 50 pmol/µl of protein partially coat the beads **(Fig 11a)**. We found that 100 pmol/µl of beads is the concentration at which the beads are completely saturated with our His_6-*Mt*UGM-MBP selection target (~87kDa) **(Fig 11b)**. The beads saturated with the proteins were used for the selection process.

3.1.4 Preparation of eFX and $tRNA^{fMet}_{CAU}$ for Flexible *in vitro* Translation of Macrocyclic Peptides

Catalytic RNA ribozymes known as flexizymes were previously developed via *in vitro* evolution as a tool to expand the genetic code for cell-free protein translation, they can charge any *E. Coli* of tRNA with a variety of non canonical amino acids. Flexizyme eFx is capable of charging tRNA with a chloroacetyl modified aromatic amino acid. We synthesised dsDNA of flexizyme and $tRNA^{fMet}_{CAU}$ by primer extension polymerase chain reaction. The dsDNA of tRNA and eFx was visualised in 3% agarose gel at 99bp and 66bp, respectively **(Fig 12a)**. The dsDNA was reverse transcribed, and RNA was purified denaturing PAGE visualised under UV **(Fig 12b)**. *N*-(2-chloroacetyl)-L-phenylalanine-$tRNA^{fMet}_{CAU}$ was prepared by incubating eFx and initiator $tRNA^{fMet}_{CAU}$ which served as methionine during flexible *in vitro* translation. The chloroacetyl group is involved in cyclization of the peptide.

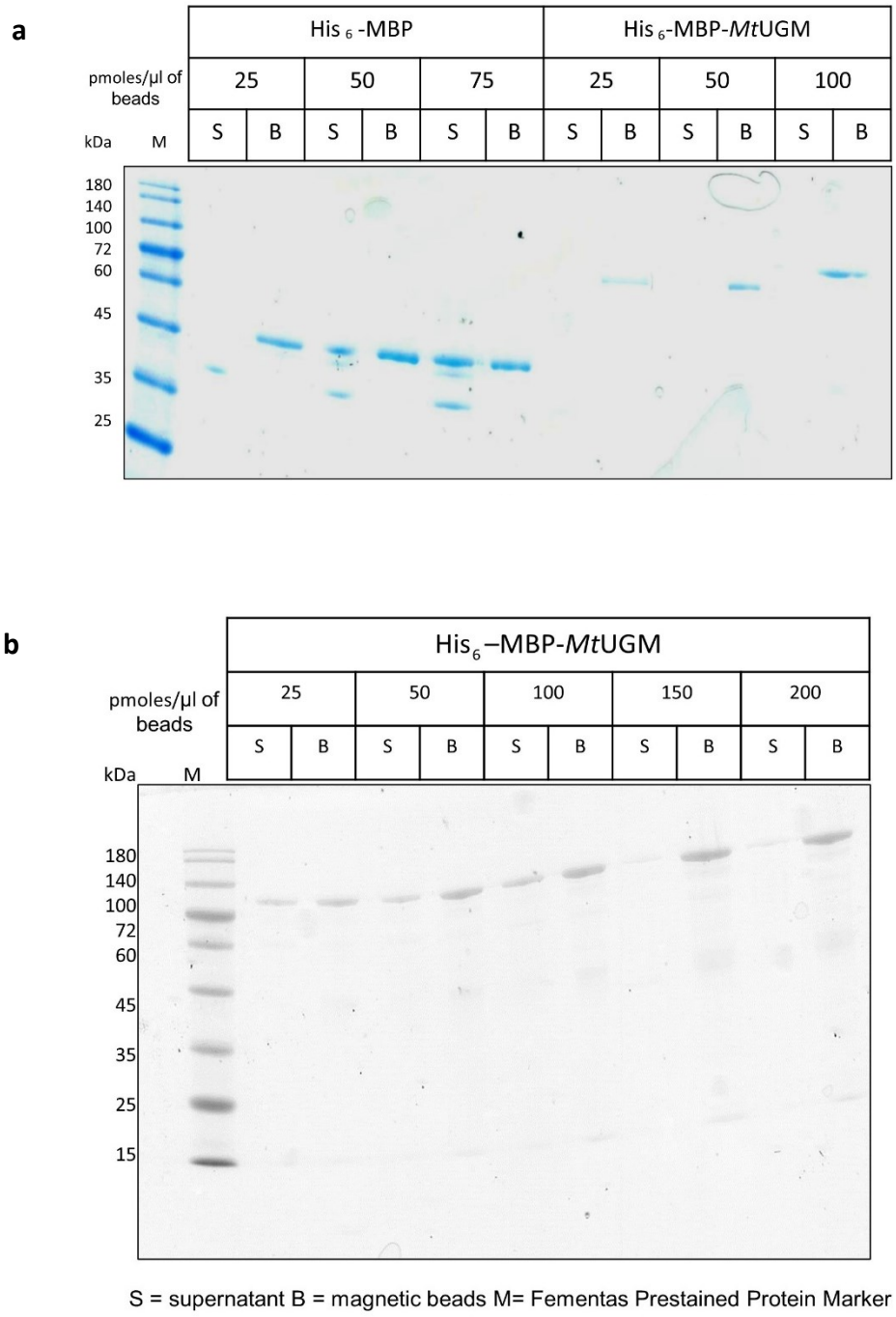


Figure 11. SDS-PAGE analysis of protein immobilisation on his-tag isolation beads (a) 25, 50 and 75 pmol of His_6-MBP was saturated on 1 ul of beads, and 50 pmol of His_6-MBP showed complete saturation, 25, 50 and 100 pmol was saturated on 1 ul of beads, the gel was inconclusive thereby, was repeated again (b) 25, 50, 100, 150 and 200 pmol of His_6-MBP-*Mt*UGM was saturated on 1ul of beads and 100 pmol of His_6-MBP-*Mt*UGM showed complete saturation.

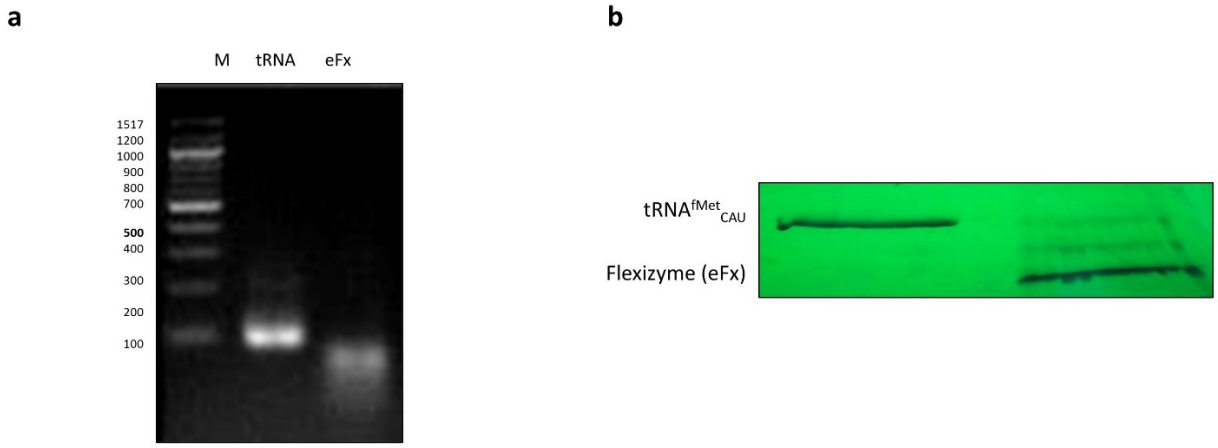


Figure 12. Preparation of $tRNA^{fMet}_{CAU}$ and Flexizyme (eFx) (a) PCR amplification of tRNA and flexizyme dsDNA (b) RNA purification using denaturing PAGE.

3.1.5 Preparation of Puromycin-Ligated mRNA Library

Puromycin plays a critical role in mRNA display, it attaches the translated peptide to its encoding mRNA. Ligation of puromycin was achieved by incubating the mRNA libraries coding the affibody and macrocyclic peptides with puromycin. The ligated library was purified and analysed on denaturing PAGE. The puromycin ligated library can be observed as a higher band on the gel compared to the unligated library **(Fig 13)**. The ligation round was repeated after each selection round.

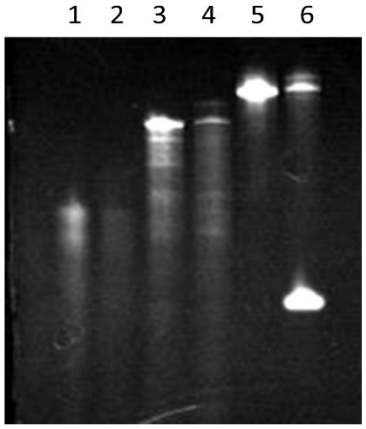


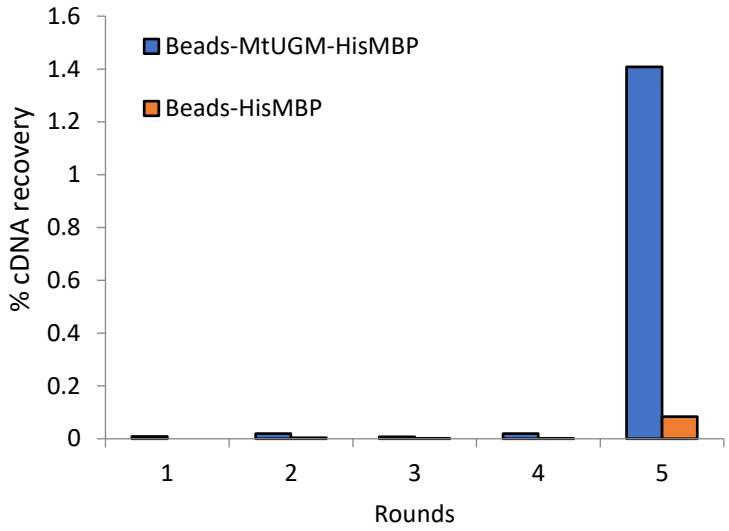
Figure 13. Denaturing PAGE analysis of puromycin ligated mRNA libraries (1) Macrocyclic peptide mRNA library (2) Macrocyclic peptide mRNA-Puromycin (3) Affibody mRNA library (4) Affibody mRNA-Puromycin (5) Monobody mRNA library and (6) Monobody-Puromycin.

3.1.6 Selection of *Mt*UGM Binding Macrocyclic Peptides

To identify MCPs for *Mt*UGM we employed the RaPID system. The mRNA template library was designed to have 5′-AUG-(NNK)n-UGC-3′, where AUG and UGC assign chloroacetyl amino acid (ClAc^{L}Phe) and Cys, respectively, and where, NNK15 (N represents any of four bases and K represents U or G) assign all possible 20 amino acids at 15 degenerate positions in the coded peptide. The library is ligated with puromycin prior to peptide translation, which allows the peptide synthesised to remain attached to its coding mRNA. The *in vitro* translation using *N*-(2-chloroacetyl)-L-phenylalanine enables spontaneous cyclisation with the first occurring cysteine in the peptide chain by forming thioether bond creating a library of 10^{12} macrocyclic peptides. This is followed by reverse transcription to stabilize the mRNA by creating mRNA-DNA hybrid which is linked with the corresponding peptide. Prior to the selection process, the library was subjected to negative selection (preclearing) using His_6-MBP (our counter-selection target) immobilised to beads. The library was then subjected to RaPID selection using His_6-MBP-*Mt*UGM that was immobilized on Ni^{2+}-NTA magnetic beads via the interaction with the hexa-histidine tag. We also incubated the precleared library with His_6-MBP as a negative control. The binding steps were carried out at 4°C. After the first round of selection, the cDNA was recovered and qPCR was performed to analyse the recovered amount of DNA, which was then PCR amplified for the next round. This was followed by *in vitro* transcription of the recovered library and puromycin ligation. Subsequently, *in vitro* translation and reverse transcription were performed to generate an enriched mRNA-DNA-peptide fusion library. This was repeated and the selection process continued until we observed enrichment of positive binders.

Figure 14a depicts the progress of selection process. At the fifth round, we observed an appreciable enrichment of MCPs binding to the target in the pools monitored by the recovery amount of selected cDNA by RTPCR, while the binding to His_6-MBP (immobilized to magnetic beads) was significantly low. The recovered library was 1.5% of the input library. We plotted the positive to negative ratio, where positive refers to the MCPs binding to His_6-MBP-*Mt*UGM and negative refers to His_6-MBP binders **(Fig 14b)**. This ratio however does not indicate affinity of binding or the frequency of selected clones.

a



b

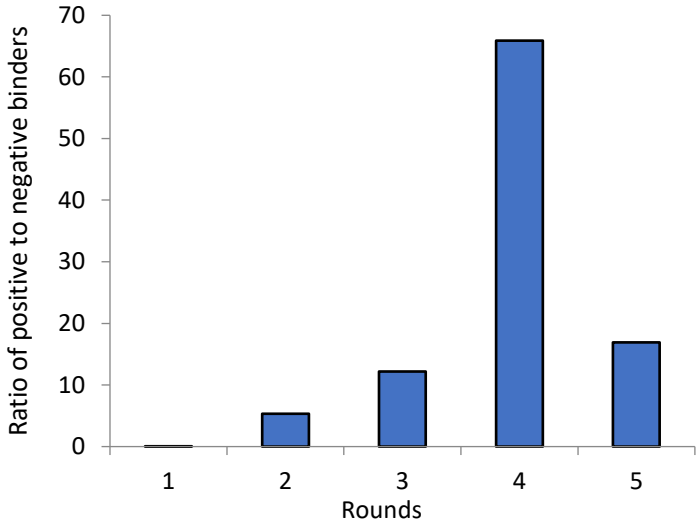


Figure 14. Selection of *Mt*UGM binding macrocyclic peptides (a) Progression of selection through individual rounds, percentages of peptides bound were determined by dividing the amount of recovered cDNA by the amount of input macrocyclic peptide-mRNA conjugate. (b) The positive to negative ratio determined by dividing the percentage recovery of macrocyclic peptide-mRNA conjugate binding to His_6-MBP-*Mt*UGM by recovery of macrocyclic peptide-mRNA conjugate binding to His_6-MBP.

3.1.7 Single Clone Assay

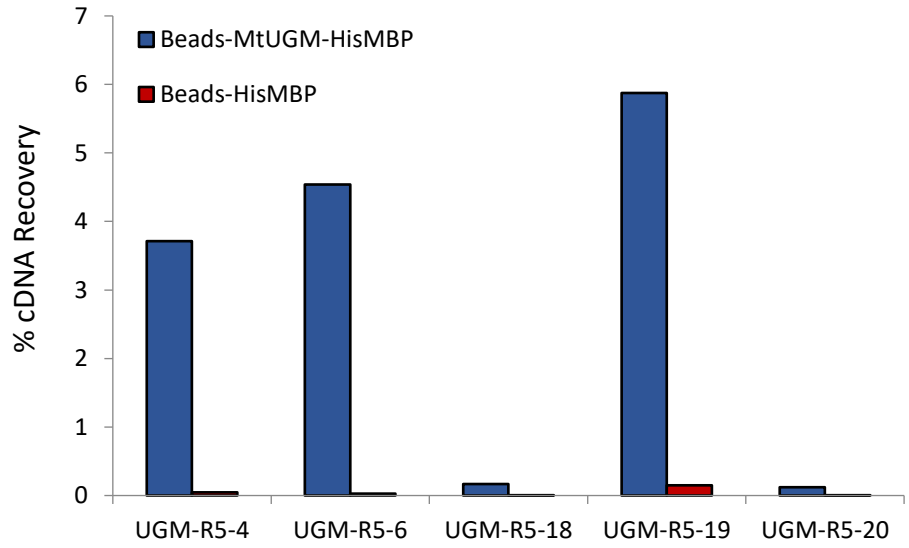
To assess the selection process at the end of round 5 we randomly isolated 20 clones by TA cloning and performed standard sequencing. The sequences reveal the appearance of HHR motif and macrocycles with one representing a mini cycle at the C terminal (Table 1).

Table 1. Macrocyclic peptides isolated for single clone assay

Clone	Sequence
UGM-R5-4	**ClAc^{L}F** RHSCHTQHHRLTHHQCGSGSGS
UGM-R5-6	**ClAc^{L}F** HKHTAHHRTHTHRRCGSGSGS
UGM-R5-18	**ClAc^{L}F** HSTHHWHQHHRQCGSGSGS
UGM-R5-19	**ClAc^{L}F** KTPWRHHHNTLTHRNCGSGSGS
UGM-R5-20	**ClAc^{L}F** KQKRHHYHRHKHQLNCGSGSGS

The clones were then tested for their binding and specificity against *Mt*UGM. UGM-R5-19 was observed with highest recovery at 5.8% of the input amount, and UGM-R5-20 reports the lowest at 0.1% of the input amount **(Fig 15a)** , however, the positive to negative ratio is significantly higher indicating that the binders are specific to *Mt*UGM **(Fig 15b)**.

a



b

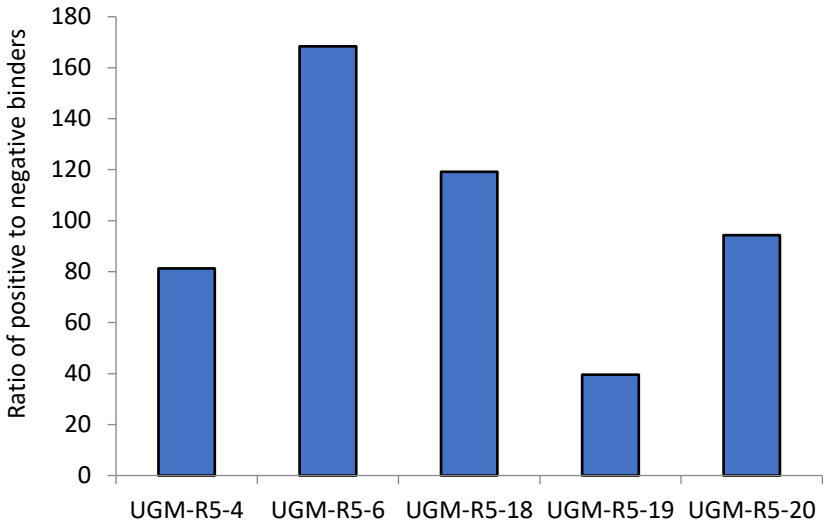


Figure 15. Single clone assay for selected MCPs (a) %cDNA recovery of Macrocyclic peptides binding to His_6-MBP-MtUGM and His_6-MBP (b) The positive to negative ratio for individual clones determined by dividing the percentage recovery of macrocyclic peptide-mRNA conjugate binding to His_6-MBP-*Mt*UGM by recovery of macrocyclic peptide-mRNA conjugate binding to His_6-MBP.

3.1.8 Selection of *Mt*UGM Binding Affibodies

Affibodies are relatively larger (6.5 kDa) than MCPs (2 kDa) in size. mRNA display of affibodies is a novel approach to find ligands for the target enzyme. The affibody library template consists of NNK 13 codons that code for 13 random amino acids shown below **(Fig 16).** The codons that encode a random amino acid are placed on the helix front of the two out of three helices, which make interactions with a target protein. Affibodies are high affinity ligands, thus, studying the evolution of these degenerate positions will provide information on ligand binding and highlight the essential amino acid motifs.

Helix front Helix front Helix back

MVDNKFNKEXXXAXXEIXXLPNLNXXQXXAFIXSLXDDPSQSANLLAEAKKLNDAQAPKGSGSGS*

Linker

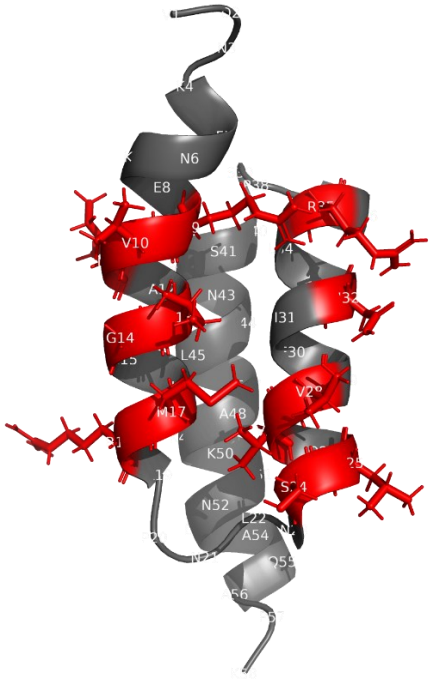


Figure 16. Affibody library template. The X denotes the amino acids encoded by NNK degenerate positions. The residues in red represents the random amino acids, essentially placed on the helix front.

In vitro translation of the Affibody mRNA library conjugated with puromycin incorporates all 20 natural amino acids. The translated affibody is thus attached to its coding mRNA creating an affibody-mRNA conjugate. This library was precleared by applying it to beads immobilised with His_6-MBP to remove binders to beads and His_6-MBP prior to the selection process. The precleared library was then applied for selection using His_6-MBP-*Mt*UGM immobilised on beads. We also incubated the precleared library with His_6-MBP as a negative control. The binding steps were carried out as described in section 3.1.6. **Fig 17a** show that at round 5 we observe the steep increase in % cDNA recovered as positive binders (i.e. affibodies that bind to His_6-MBP-*Mt*UGM) at 0.1 % and the negative binders at 0.005%. We plotted the positive to negative ratio, where positive refers to the MCPs binding to His_6-MBP-*Mt*UGM and negative refers to His_6-MBP binders **(Fig 17b)**. This ratio however does not indicate affinity of binding or the frequency of selected clones.

3.1.9 Next-Gen MiSeq Analysis of Sequence Enrichment During Selection Rounds.

mRNA libraries from each round were sequenced using Next-gen MiSeq to analyze the sequence enrichment during the selection progress. For the MCP library we selected the top 20 sequences **(Table 2)** from round 5 and plotted the gene frequencies from each round to generate an evolution graph. We can see the increase in the gene frequency of the selected clones in round 3, with highest frequency observed at round 5 **(Fig 18a)**. Similarly, we plotted the evolution graph for affibodies, which shows that the sequences enriched in the last round of selection begin to appear at round 4 and we observe a significant enrichment in round 6 depicted by the highest frequency **(Fig 18b)**. After which we stopped the selection process.

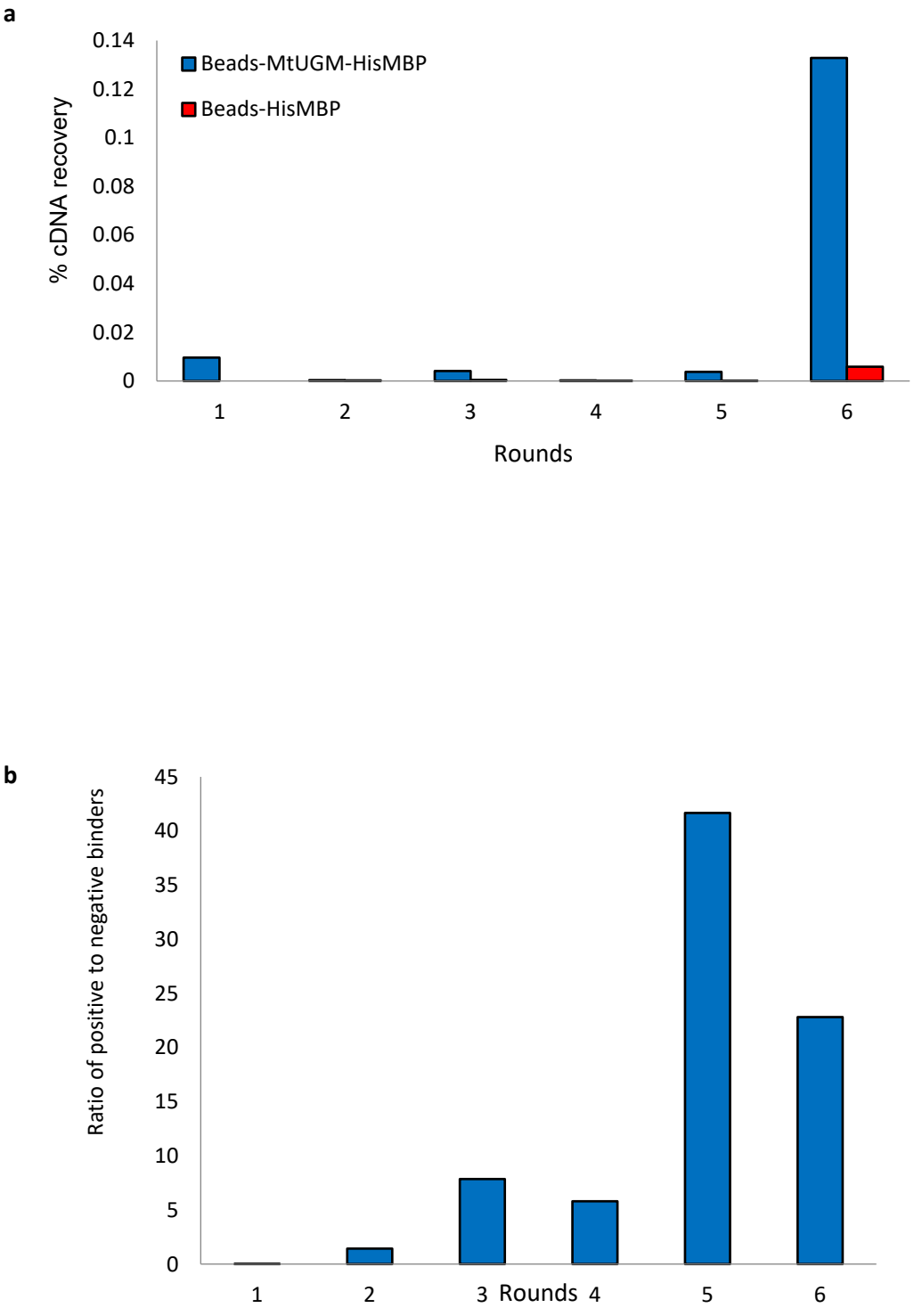


Figure 17. Selection of *Mt*UGM binding affibodies (a) Progression of selection through individual rounds, percentages of peptides bound were determined by dividing the amount of recovered cDNA by the amount of input affibody-mRNA conjugate. (b) The positive to negative ratio determined by dividing the percentage recovery of affibody-mRNA conjugate binding to His_6-MBP-*Mt*UGM by recovery of affibody-mRNA conjugate binding to His_6-MBP.

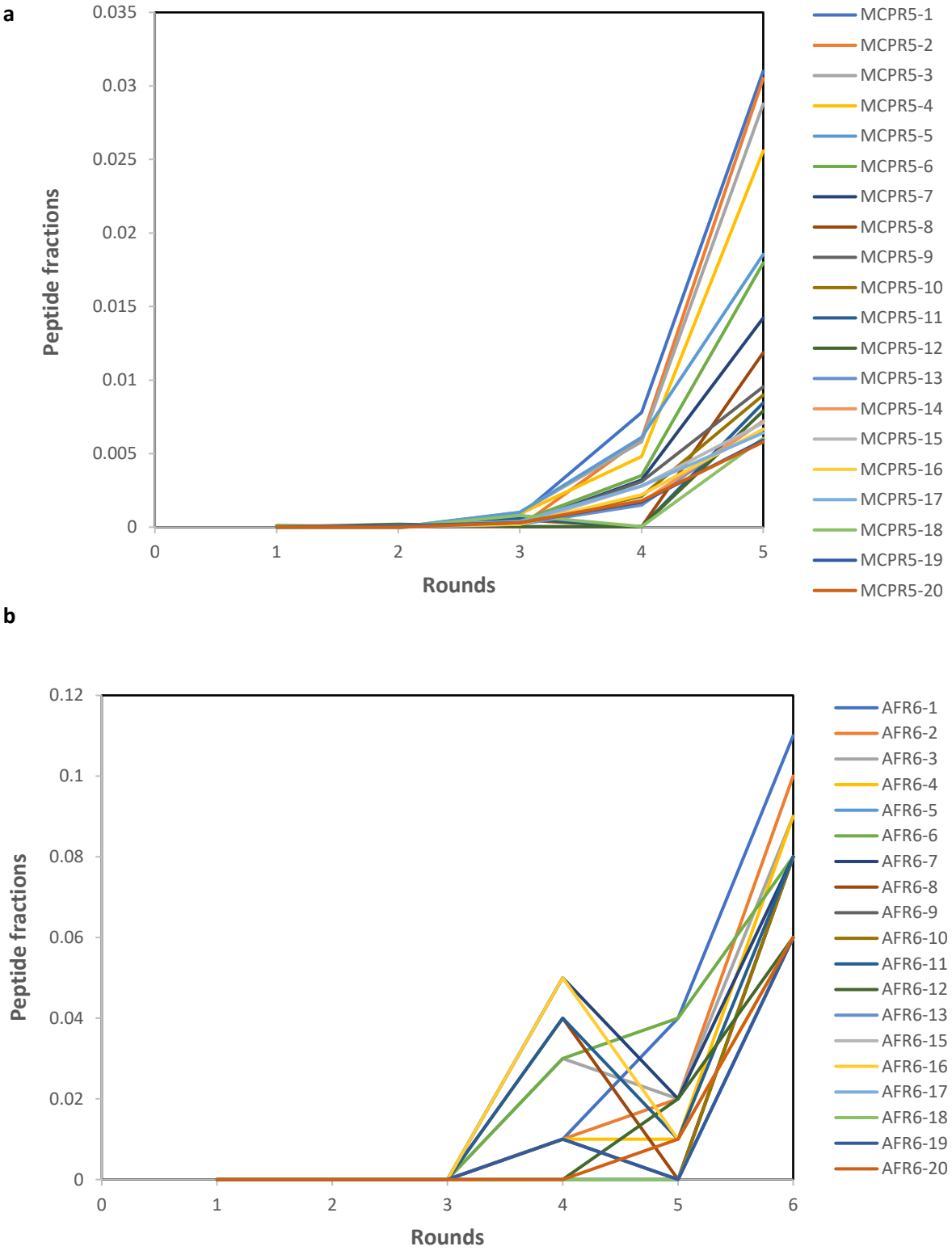


Figure 18. Next-gen analysis: evolution graph of peptide/protein sequences (a) Enrichment plot of the top 20 MCP sequences of round 5. (b) Enrichment plot of the top 20 affibody sequences of round 6.

Table 2. Macrocyclic peptide sequences from Round 5 and their frequency of occurrence.

Clone	Sequence	R1	R2	R3	R4	R5
MCPR5-1	ClAc^{L}F RHSCHTQHHRLTHHQCGSGSGS*	0	0	0.08	0.78	3.10
MCPR5-2	ClAc^{L}F KKHHRPIHRRHHTHCGSGSGS*	0.01	0.01	0.004	0.60	3.05
MCPR5-3	ClAc^{L}F HTQCSSHSRSHHHHCGSGSGS*	0.01	0	0.10	0.58	2.88
MCPR5-4	ClAc^{L}F TWNPHRHKCHTHTHHCGSGSGS*	0	0	0.09	0.48	2.56
MCPR5-5	ClAc^{L}F HHLHRRHHAHTHCGSGSGS*	0.01	0	0.10	0.61	1.86
MCPR5-6	ClAc^{L}F HHRHKIPAHTRHHRSCGSGSGS*	0.01	0	0.04	0.35	1.80
MCPR5-7	ClAc^{L}F HLKHHRAPHHRTHNCGSGSGS*	0	0	0.02	0.32	1.42
MCPR5-8	ClAc^{L}F VTHHHPKHRAHLHMHCGSGSGS*	0	0	0.05	0.002	1.19
MCPR5-9	ClAc^{L}F RHKHHRPRHHYPSHCGSGSGS*	0	0	0.02	0.31	0.95
MCPR5-10	ClAc^{L}F RHHRQDHHHTQHGSGSGS*	0	0	0.02	0.21	0.90
MCPR5-11	ClAc^{L}F HAHHDTHKTRHRHHCGSGSGS*	0	0	0.06	0.002	0.85
MCPR5-12	ClAc^{L}F HQKYHHSHRHDPAHKCGSGSGS*	0	0.02	0.004	0.002	0.79
MCPR5-13	ClAc^{L}F KKHHSHHPYHQTHKHCGSGSGS*	0	0	0.02	0.15	0.72
MCPR5-14	ClAc^{L}F YRSHHHHKHHDTRIHCGSGSGS*	0	0	0.03	0.17	0.72
MCPR5-15	ClAc^{L}F HKHTAHHRTHTHRRCGSGSGS*	0	0	0.02	0.28	0.71
MCPR5-16	ClAc^{L}F KRNHRIHHRYPHLHHCGSGSGS*	0	0.01	0.02	0.22	0.67
MCPR5-17	ClAc^{L}F MHRHHRHKHLNHTTCGSGSGS*	0	0.01	0.04	0.28	0.64
MCPR5-18	ClAc^{L}F KNLHHHRHTTHHHCGSGSGS*	0	0	0.08	0.002	0.60
MCPR5-19	ClAc^{L}F SNHHPRHHHVPHRLHCGSGSGS*	0	0.01	0.03	0.17	0.59
MCPR5-20	ClAc^{L}F QRTKHHHTHHHWHDCGSGSGS*	0	0.00	0.03	0.18	0.58

Table 3. Affibody sequences from Round 6 and their frequency of occurrence in round 4,5 and 6.

Clone	Sequence	R4	R5	R6
AFR6-1	MVDNKFNKEVLLAVIEIRVLPNLNLVQFSAFIWSLDDDPSQSANLLAEAKKLNDAQAPKGSGSGS*	0.01	0.04	0.11
AFR6-2	MVDNKFNKERCMALFEIVFLPNLNLIQFYAFIVSLTDDPSQSANLLAEAKKLNDAQAPKGSGSGS*	0.01	0.02	0.1
AFR6-3	MVDNKFNKECFLAILEIAWLPNLNCLQFFAFILSLGDDPSQSANLLAEAKKLNDAQAPKGSGSGS*	0.03	0.02	0.09
AFR6-4	MVDNKFNKELLLALVEIVFLPNLNLVQGFAFILSLVDDPSQSANLLAEAKKLNDAQAPKGSGSGS*	0.01	0.01	0.09
AFR6-5	MVDNKFNKEHWLAFMEIVVLPNLNLVQVFAFIWSLSDDPSQSANLLAEAKKLNDAQPKGSGSGS*	0.01	0	0.08
AFR6-6	MVDNKFNKEVTLAFWEIRWLPNLNVVQLYAFILSLGDDPSQSANLLAEAKKLNDAQAPKGSGSGS*	0.03	0.04	0.08
AFR6-7	MVDNKFNKEMSLAFVEITWLPNLNGWQLCAFIVSLRDDPSQSANLLAEAKKLNDAQAPKGSGSGS*	0.05	0.02	0.08
AFR6-8	MVDNKFNKEWLVAYIEIVWLPNLNFMQSYAFILSLVDDPSQSANLLAEAKKLNDAQAPKGSGSGS*	0.04	0	0.08
AFR6-9	MVDNKFNKEWVVAIVEIAYLPNLNKMQIFAFIISLSDDPSQSANLLAEAKKLNDAQAPKGSGSGS	0	0	0.08
AFR6-10	MVDNKFNKEACFALWEIWFLPNLNIVQGFAFIWSLRDDPSQSANLLAEAKKLNDAQAPKGSGSGS*	0.01	0	0.08
AFR6-11	MVDNKFNKELLVATLEIVFLPNLNWVQLVAFILSLSDDPSQSANLLAEAKKLNDAQAPKGSGSGS*	0.04	0.01	0.08
AFR6-12	MVDNKFNKEWLLAYYEILLLPNLNGVQFYAFIVSLSDDPSQSANLLAEAKKLNDAQAPKGSGSGS*	0	0.02	0.06
AFR6-13	MVDNKFNKECWLALLEIWVLPNLNVAQFSAFIWSLGDDPSQSANLLAEAKKLNDAQAPKGSGSGS*	0	0	0.06
AFR6-14	MVDNKFNKEWTLALIEITWLPNLNWFQWTAFILSLFDGPSQSANLLAEAKKLNDAQAPKGSGSGS*	0.13	0.02	0.06
AFR6-15	MVDNKFNKEELMAVIEIWFLPNLNSMQLFAFILSLMDDPSQSANLLAEAKKLNDAQAPKGSGSGS*	0	0	0.06
AFR6-16	MVDNKFNKELVIAYTEIVVLPNLNMIQVFAFIMSLSDDPSQSANLLAEAKKLNDAQAPKGSGSGS*	0.05	0.01	0.06
AFR6-17	MVDNKFNKERFLALIEIVKLPNLNLFQGFAFIVSLRDDPSQSANLLAEAKKLNDAQAPKGSGSGS*	0	0	0.06
AFR6-18	MVDNKFNKEFWVAHWEIWCLPNLNCVQGFAFIYSLVDDPSQSANLLAEAKKLNDAQAPKGSGSGS*	0	0	0.06
AFR6-19	MVDNKFNKEVFFAVFEIKWLPNLNYGQLQAFIVSLSDDPSQSANLLAEAKKLNDAQAPKGSGSGS*	0.01	0	0.06
AFR6-20	MVDNKFNKEMFFAFFEIVSLPNLNLMQLFAFILSLGDDPSQSANLLAEAKKLNDAQAPKGSGSGS*	0	0.01	0.06

3.2 Recombinant Expression and Purification of Affibody

The top 4 affibody sequences revealed after round 6 of mRNA display were used for recombinant expression of affibody genes (section 3.1.11, Table 3). The plasmid pET21a-GFP was used to express affibody conjugated with green fluorescent protein. The construct consists of N-terminal hexa-His tag for affinity purification and TEV protease recognition site between the affibody sequence and GFP (Fig 19). The affibody genes were cloned into pET21a-GFP. Cloning was performed via restriction digestion and ligation and transformation of DH10b cells. Expression was carried out by transforming BL21(DE3) cells induced with IPTG. Purification was performed by IMAC using an Ni-NTA column. The fractions were collected by Akta-FPLC and analysed on SDS-PAGE (Fig 20). His-tagged AFR6-1-GFP has a theoretical molecular size of ~35.7 kDa. The final concentration of 0.48 mg/ml per 1 L of culture was determined by the BCA assay, this fraction also contained non-specific proteins eluted along with AFR6-GFP which could not be separated. His-AFR6-2-GFP protein has a theoretical molecular weight of 36 kDa. Similarly, the eluted fractions seem to contain AFR6-2-GFP. The concentration was determined to be 0.17 mg/mL. AFR6-3-GFP has a theoretical molecular size of ~36 kDa. As seen on figure 9, AFR6-3-GFP corresponding bands were present.

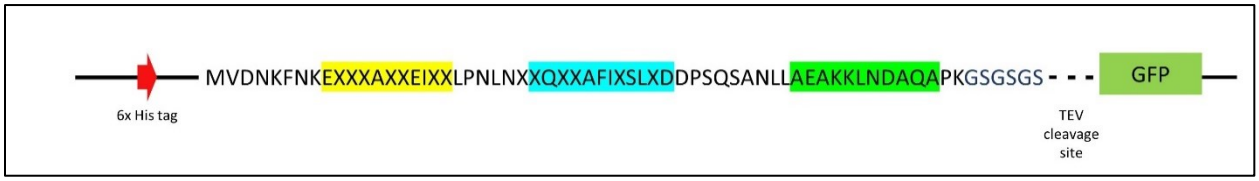


Figure 19. Affibody expression construct.

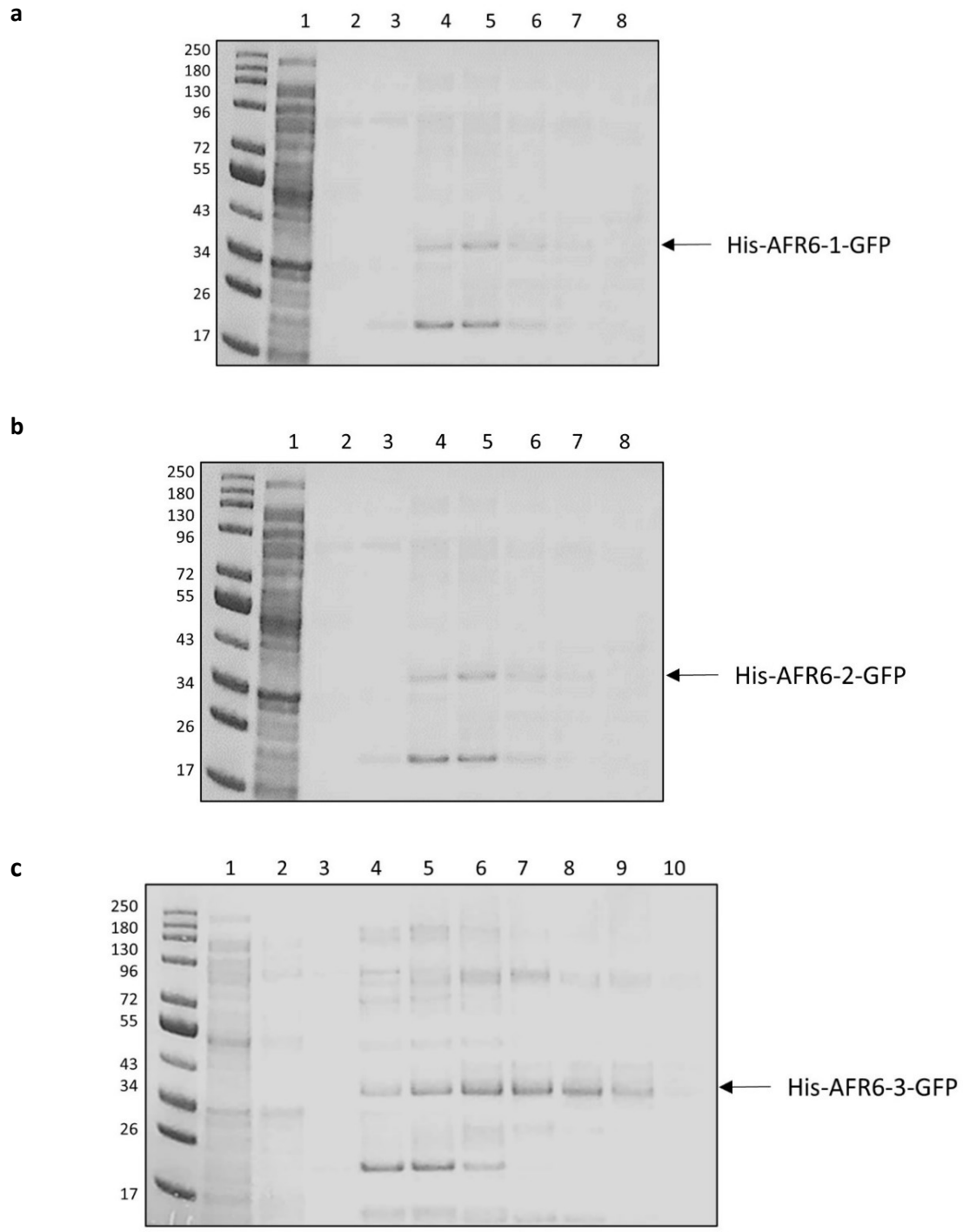
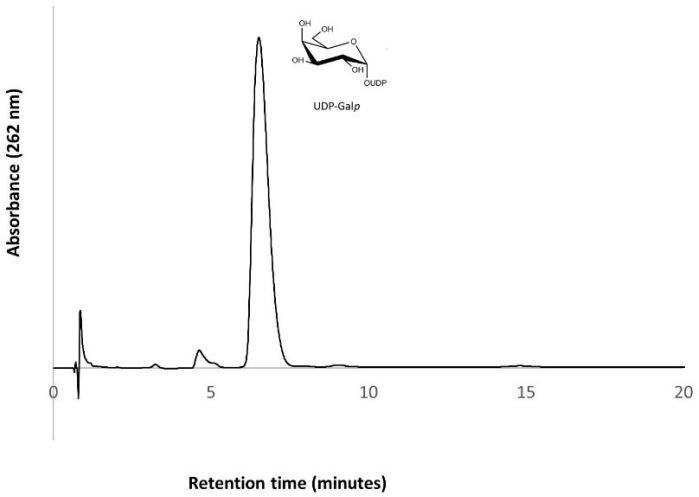


Figure 20. SDS-PAGE analysis of purified AFR6-1-GFP, AFR6-2-GFP and AFR6-3. (a) Lane 1: lysate, lane 2: the flow-through fractions, lane 3: wash fraction, and lanes 4 to 11: elution fraction containing AFR6-1-GFP. (b) Lane 1- flow-through fractions, lane 2-wash fraction, and lanes 3 to 8: elution fractions containing AFR6-2- GFP. (c) Lane 1: flow-through fractions, lane 2-wash fraction, and lanes 3- elution fraction containing AFR6-3- GFP.

3.3 HPAEC Activity Assay of *Mt*UGM

To detect *Mt*UGM enzyme activity we used previously described high-performance anion exchange chromatography with some modifications [83]. CarboPac PA1 column (Dionex Inc) preequilibrated with 200mM ammonium acetate was used to detect UDP-Gal*f* and UDP-Gal*p* peaks at 262 nm. The enzyme activity was confirmed by the detection of the product UDP-Gal*f* at 262 nm. The reaction was set up with the varying concentrations of the substrate UDP-Gal*p* concentrations (25-125μM), a fixed amount of *Mt*UGM (200nM) as per section 2.8. The negative control was set up with no enzyme. **Figure 21** shows the peaks for UDP-Gal*p* and UDP-Gal*f* separated by baseline resolution were detected at around 6.2 min and 8.5 min, respectively. At equilibrium, the pyranose form is favoured UDP-Gal*p* : UDP-Ga*lf* being 97:7 as previously reported [24]. Based on our assay, the % conversion of UDP-Gal*p* to UDP-Gal*f* was 6.32%, which is close to the reported data. Reported studies have successfully used this assay technique for testing the inhibitors of *Mt*UGM (and other bacterial UGMs) by calculating the % conversion of the substrate to the product. However, these research groups have used UDP-Gal*f* as the substrate, since, the equilibrium favors the conversion of UDP-Gal*f* to UDP-Gal*p*. Studies show that when UDP-Gal*p* is used as a substrate, most of the product (UDP-*Galf)* is converted to UDP-Gal*p*, because the equilibrium favours the formation of UDP-Gal*p* reaction. Thus, characterizing the MCPs and affibodies for inhibition using UDP-Gal*f* as substrate may be considered. Nonetheless, our results show the detection of UDP-Gal*f* peak, thus, this assay can be used to screen for potential inhibitors discovered through the mRNA display process.

a



	Reaction	UDP-Gal*p* (µM)	*Mt*UGM (200 nM)
	a	125	-
b	1	25	+
	2	50	+
	3	75	+
	4	100	+
	5	125	+

b

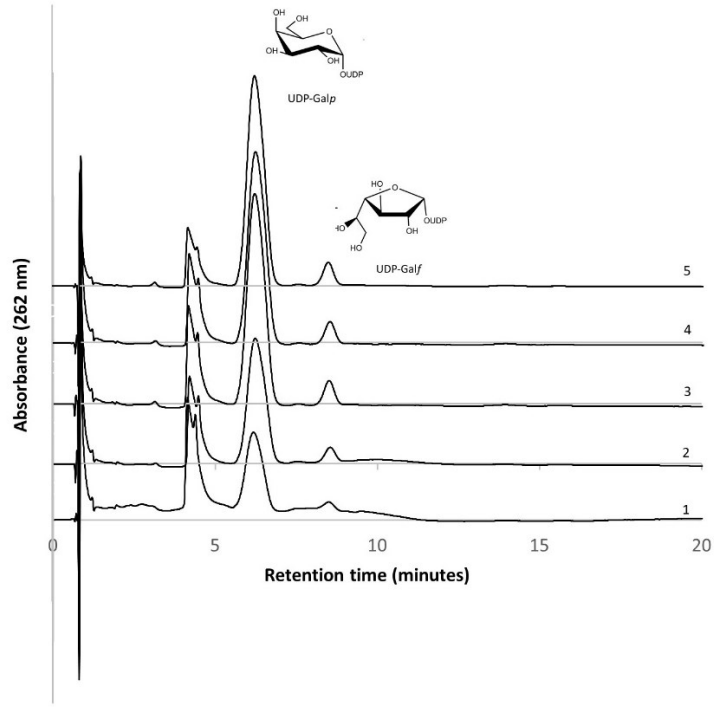


Figure 21. HPAEC activity assay of *Mt*UGM. The peaks for UDP-Gal*p* and UDP-Gal*f* is detected at around 6.2 minutes and 8.5 minutes, respectively. (a) Negative control- reaction was carried out with no *Mt*UGM. (b) Enzyme activity tested for varying concentrations of the substrate UDP-Gal*p* (1-5).

CHAPTER 4: DISCUSSIONS

The application of mRNA display platform as a tool to identify target-specific peptide/protein ligands from highly diverse libraries has been gaining popularity. Over the past two decades, a growing interest in rational design and screening of synthetic libraries of engineered proteins for therapeutic applications has been reported. The focus has been towards the generation of vast peptide libraries and developing high throughput screens to facilitate discovery of novel hits for specific targets. Apart from the broader goal of therapeutic application, peptide/protein-based ligands have been used to elucidate the binding interactions with the target protein; information that can play an essential role in developing novel drugs. With increasing incidences of antibiotic resistance, we notice a paradigm shift in the techniques involved in discovering novel antibiotic drugs. In our project, we have used an established display platform that relies on *in vitro* transcription-translation mechanisms to discover macrocyclic peptides and affibodies for our selected molecular target *Mt*UGM.

4.1 mRNA Display: Discovery of *Mt*UGM Binding MCPs and Affibodies

Early on we hypothesized that selection process could be targeted towards the selection of macrocyclic peptide specific for the active site of *Mt*UGM by using an active-site triple-mutant negative selection target to exclude binders that interact outside of the active-site. The published crystal structure of *Mt*UGM highlights the role of active site residues in ligand binding. Thus, by having an active site mutant as a negative screen to fish out all surface binders seemed like a promising strategy. However, further consideration led us to believe that it will be difficult to achieve this goal as it would be a bottle neck for finding binders and thus we would lose the variety in the initial rounds if we include such a stringent negative selection, which is undesirable. Negative selection using active site mutants might also result in loss of potential allosteric inhibitors. Thus, for our study we concentrated on finding ligand binders and

later we would like to proceed with crystallography to identify the nature of binding, based on which we could try a directed selection process through which we can select for binders specific to the active site. The purification of 7x-*Mt*UGM mutant was not successful, we suspect this is due to misfolding of protein due to seven mutations. The 3F-*Mt*UGM active-site mutant created in this project may be useful in future studies as per the above stated reasons.

For the selection process, we obtained the target *Mt*UGM through recombinant expression and purification **(Fig 9)**. The purification of the target and the dual tag was carried out using affinity chromatography. It was a challenge to acquire a pure fraction of fusion protein His_6-MBP-*Mt*UGM, as it was suspected that the fraction would contain traces of His_6-MBP resulting from the premature termination of the His_6-MBP-*Mt*UGM fusion. This could pose a problem during the selection process. This was overcome by collecting the purest fractions (i.e. the first fraction collected, **Fig 9**) however, the yield was low. To overcome this, His_6-MBP-*Mt*UGM was expressed in Rosetta (DE3) following previously mentioned protocol in section 2.1, the purest fraction were collected from several batches and pooled together to obtain a final concentration of 0.3ml/mL in a total volume of 2 mL. Additionally, this was addressed by having His_6-MBP-bound beads for the pre-clearing of the peptide-mRNA library and as a negative control.

The selection of MCPs was done via the RaPID system. During the process, we faced difficulty in achieving a good efficiency of puromycin ligation between the selection rounds. Thus, we suspect that this may have resulted in a loss of diversity at this step. mRNA Display of affibodies was continued until round 6, appreciable enrichment was observed between the round 5 and 6. However, recovery of cDNA was at 0.1% which is low compared to MCPs. The PCR amplification of the affibody library also posed a challenge due to the amplification of non-target DNA sequence, thereby affecting the library diversity and template. We applied gel purification of mRNA to retrieve the correct size of the library however, we lost some library during the purification process.

4.2 Next-gen MiSeq Analysis

Through next-generation sequencing analysis, the evolution graph plotted for all the occurring MCP sequences shows a trend in the enrichment of clones as the selection progressed **(Fig A1)**. However, the evolution graph plotted for affibodies did not show a definite trend. Thus, we plotted the change in gene pools from individual rounds. The evolution graph for the top 20 sequences of affibodies in round 6 reveals that these sequences coding the corresponding affibodies were seen after the third round of selection. A general pattern of a dip in gene frequency before its increase was observed for most sequences except AR6-1, 2, 3, 6, 12, and 20 **(Fig 18b)**. It was also interesting to note that the evolution graph plotted for the top sequences appearing in round 5 of affibody selection shows that AFR5-10, (which is the same sequence as AFR6-6) dominates the selection **(Fig A2)**. Further bioinformatics analysis is required for elucidating the factors that influence a selection process; the knowledge of which can also be utilized as a basis for designing new ligand libraries. For example, investigating the emergence of motifs in the peptides/proteins can shed light on binding interactions. Ligands discovered through this process were synthesized. We aim to characterize them for specific functions.

4.3 Synthesis of Desired Ligands Identified Via *In vitro* Selection

Due to its smaller size and length (15 amino acids) in MCPs can be chemically synthesized. UGM-R5-19, UGM R5-20 (and fluorescently labeled MCPs) were chemically synthesised by Dr. Christopher Hipolito. Chemical synthesis of affibodies is not efficient, thus we recombinantly expressed and purified these affibodies. However, additional optimization of the protocol is required. The procedures were adapted from literature for affibody expression which stated use of denaturing conditions for purification followed by refolding; however, these were not tagged with GFP. We speculated that using denaturing conditions for GFP would result in an unfolded protein that would not be easy to refold, so we excluded denaturing conditions from our procedure. We were able to express low concentrations of affibodies fused to GFP, this could be due to insoluble protein. It was observed that the affibody tagged with GFP was present largely in

the insoluble fraction as we found GFP in the cell debris (visualised by fluorescence excitation under UV light, wavelength 350-400nm) after lysis and protein extraction, suggesting we lost recombinant protein with the cell debris. We can try optimising the buffer conditions to increase the solubility, use an MBP tag (for solubility), or increase sonication time. Though we were able to successfully express the affibodies, we found it challenging to obtain pure fractions during the elution process, and thus contaminating proteins eluted along with affibodies contributed to the final protein concentration. Improvements like the use of different affinity columns, such as a TALON resin, or use of size-exclusion chromatography to isolate desired proteins by its molecular size may be attempted to obtain better results.

4.4 Establishing HPAEC Assay for Testing Potential Inhibitors

Our HPAEC assay is based on previous literature, where it has been used to test enzyme activity and screen potential inhibitors. However, the method for running the HPLC varies, thus, the assay we used has been developed after considering some modifications. HPLC data is dependent on run time, flow rate, injection volume and the eluent concentrations. Previous papers have reported varying flowrate and run time. After the observation that the last peak is detected before 12 minutes, run time was set at 20 minutes. Even though the peaks for UDP-Gal*f* and UDP-Gal*p* appear much earlier than reported data, we confirmed that the identified peaks are correct. The peaks were compared against standards of UDP and UDP-Gal*p*, and against two blanks (enzyme in reaction buffer and reaction buffer without substrate or enzyme). However, we were not able to confirm the peak for UDP-Gal*f* against an authentic standard. To check if reaction incubation time (enzyme along with the substrate) affects the amounts of the product UDP-Gal*f*, we let the reaction proceed for 2 minutes and 10 minutes, no significant change was observed. The end goal for developing this assay was to test the synthesized MCP and affibodies for inhibition, thus, it will be our future goals.

4.5 Conclusions

The search for new ligand libraries and leads continues in the drug discovery industry. Peptides and small molecule proteins have served as excellent ligands with applications in PPI inhibition, imaging, probing, and co-crystallization ligands. Along with these applications, peptides and small proteins have shown therapeutic effects, prompting research studies that focus on a class of peptide/protein-based drugs. We present a selection method that uses the mRNA display technique to discover MCPs and affibodies from a library consisting of sequence diversity up to 10^{12}, specific to *Mt*UGM. The RaPID system incorporated the FIT system to generate MCPs by spontaneous cyclization facilitated by the thioether bond formation between non-natural amino acid ClAc^{L}F and first occurring Cys. The macrocyclic structures present many advantages as a drug-like molecule [88], thus, MCPs reported in this study can help provide insights on novel drug development. This technique of *in vitro* selection of MCP is not limited to discovering ligands for *Mt*UGM, it can be adapted to select ligands for other target proteins with elucidated crystal structures. Affibodies are robust engineered small proteins that have been displayed using phage display [79-74], mRNA display of affibodies reported in this study is a novel approach. We successfully selected affibodies specific to *Mt*UGM and hope to characterize them in future. Our study gives us an opportunity to explore applications of *in vitro* systems in discovering ligands from different type of compound libraries, in this case MCPs and affibodies, highlighting the dynamic nature and adaptability of the system. Together, the deep sequencing data generated and the mentioned display platform in this study will provide essential information for designing novel therapeutics, eventually assisting the development of new drug candidates for treating TB.

4.6 Future Directions

This project focussed on mRNA display as an *in vitro* selection platform for discovering peptide/protein-based ligands for the target *Mt*UGM. Moving forward, the first thing to do next will be to test the MCPs for inhibition using the established HPAEC assay. By, coupling the inhibition assay with

GlfT2 (an enzyme that uses UDP-Gal*f* as a substrate for galactosyltransferase activity for galactan synthesis) we can further validate our results. Our collaborator Dr. Christopher Hipolito and his student Ms. Nohara Goto together have discovered novel GlfT2-binding MCPs using the RaPID system. The discovered MCPs were then screened for inhibition by employing a novel biochemical assay. The assay is designed to measure fluorescence polarisation, by using a fluorescently labeled substrate; α-L-rhamnosyl-(1®3)-N-acetyl-α-D-glucosamine, linked by diphosphate to Alexafluor 488 dye. This disaccharide will act as an acceptor for GlfT1 in the transfer of two successive Galf residues from UDP-Gal*f*, generated in situ by *Mt*UGM (from the conversion of UDP-Gal*p*). Preliminary fluorescence polarization assays were performed by Kwan lab, $K_d \sim 9\mu M$ was observed, however, the assay was not performed in replicates, thus, repeating it would be worthwhile. Performing an *in vivo* assay will be essential to see if the MCPs can pass through the cell membrane and maintain its activity. Our collaborator, Yossef Av-Gay has previously carried out cell death assays with *Mtb* with the discovered GlfT2-targeting MCPs to check the cytotoxicity, however, they were not active inside the cells. Thus, modifications on GlfT2-MCPs were made to increase cell penetration. Other assays like SPR (surface plasmon resonance), fluorescence polarisation binding assay (using fluorophore tagged peptide), and cell-based assays can be performed additionally to calculate IC_{50} values.

Crystallization techniques have previously revealed a structural basis for binding interactions as a valuable tool for target-specific ligands. Thus, further work will be to co-crystallize the *Mt*UGM with the discovered ligands, both MCP and affibodies (in association with our collaborator, Kenneth Ng). This would provide information on the binding interactions made by these peptides. This can be used as a basis for designing novel drug-like molecules that inhibit this enzyme. Preliminary docking studies of discovered affibody by Thanasis Poullikas with *Mt*UGM have revealed that it binds close to the active site, thus, it will be worthwhile to proceed with co-crystallization studies **(Fig A3.)**

REFERENCES

1. Global Report on Tuberculosis, (2018-2020). World Health Organization.

2. Golden, M. P., & Vikram, H. R. (2005). Extrapulmonary tuberculosis: an overview. *American Family Physician*, ***72***(9), 1761–1768.

3. Schlesinger L. S. (1993). Macrophage phagocytosis of virulent but not attenuated strains of *Mycobacterium tuberculosis* is mediated by mannose receptors in addition to complement receptors. *Journal of Immunology (Baltimore, Md.: 1950)*, ***150***(7), 2920–2930.

4. Frehel, C., de Chastellier, C., Lang, T., & Rastogi, N. (1986). Evidence for inhibition of fusion of lysosomal and prelysosomal compartments with phagosomes in macrophages infected with pathogenic *Mycobacterium avium*. *Infection and Immunity*, ***52***(1), 252–262.

5. Ramakrishnan L. (2012). Revisiting the role of the granuloma in tuberculosis. *Nature Reviews. Immunology*, ***12***(5), 352–366.

6. Smith I. (2003). *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. *Clinical Microbiology Reviews*, ***16***(3), 463–496.

7. Bansal, R., Sharma, D., & Singh, R. (2018). Tuberculosis and its Treatment: An Overview. *Mini Reviews in Medicinal Chemistry*, ***18***(1), 58–71.

8. Heifets L. B. (1994). Antimycobacterial drugs. *Seminars in respiratory infections*, ***9***(2), 84–103.

9. Bernstein, J., Lott, W. A., Steinberg, B. A., & Yale, H. L. (1952). Chemotherapy of experimental tuberculosis. V. Isonicotinic acid hydrazide (nydrazid) and related compounds. *American review of tuberculosis*, ***65***(4), 357–364.

10. Zabinski R. F. & Blanchard J. S. (1997), The Requirement for Manganese and Oxygen in the Isoniazid-Dependent Inactivation of *Mycobacterium tuberculosis* Enoyl Reductase. *Journal of the American Chemical Society*, ***119*** (9), 2331-2332.

11. WHO demographic for TB success rate (2019).

12. Meacci, F., Orrù, G., Iona, E., Giannoni, F., Piersimoni, C., Pozzi, G., Fattorini, L., & Oggioni, M. R. (2005). Drug resistance evolution of a *Mycobacterium tuberculosis* strain from a noncompliant patient. *Journal of Clinical Microbiology*, ***43***(7), 3114–3120.

13. Horsburgh, C. R., Jr, & Rubin, E. J. (2011). Clinical practice. Latent tuberculosis infection in the United States. *The New England Journal of Medicine*, ***364***(15), 1441–1448.

14. McNeil, M. R., & Brennan, P. J. (1991). Structure, function and biogenesis of the cell envelope of mycobacteria in relation to bacterial physiology, pathogenesis and drug resistance; some thoughts and possibilities arising from recent structural information. *Research in Microbiology*, ***142***(4), 451–463.

15. Dover, L. G., & Coxon, G. D**.** (2011). Current status and research strategies in tuberculosis drug development. *Journal of Medicinal Chemistry*, ***54***(18), 6157–6165.

16. Rattan, A., Kalia, A., & Ahmad, N. (1998). Multidrug-resistant *Mycobacterium tuberculosis*: molecular perspectives. *Emerging Infectious Diseases*, ***4***(2), 195–209.

17. Besra, G. S., Khoo, K. H., McNeil, M. R., Dell, A., Morris, H. R., & Brennan, P. J. (1995). A new interpretation of the structure of the mycolyl-arabinogalactan complex of Mycobacterium tuberculosis as revealed through characterization of oligoglycosylalditol fragments by fast-atom bombardment mass spectrometry and 1H nuclear magnetic resonance spectroscopy. *Biochemistry*, ***34***(13), 4257–4266.

18. Jankute, M., Cox, J. A., Harrison, J., & Besra, G. S. (2015). Assembly of the Mycobacterial Cell Wall. *Annual Review of Microbiology*, ***69***, 405–423.

19. Beis, K., Srikannathasan, V., Liu, H., Fullerton, S. W., Bamford, V. A., Sanders, D. A., Whitfield, C., McNeil, M. R., & Naismith, J. H. (2005). Crystal structures of *Mycobacteria tuberculosis* and *Klebsiella pneumoniae* UDP-galactopyranose mutase in the oxidised state and *Klebsiella pneumoniae* UDP-galactopyranose mutase in the (active) reduced state. *Journal of Molecular Biology*, ***348***(4), 971–982.

20. Rose, N. L., Completo, G. C., Lin, S. J., McNeil, M., Palcic, M. M., & Lowary, T. L. (2006). Expression, purification, and characterization of a galactofuranosyltransferase involved in *Mycobacterium tuberculosis* arabinogalactan biosynthesis. *Journal of the American Chemical Society*, ***128***(20), 6721–6729.

21. Barry, C. E., Crick, D. C., & McNeil, M. R. (2007). Targeting the formation of the cell wall core of M. tuberculosis. *Infectious Disorders Drug Targets*, ***7***(2), 182–202.

22. Trejo, A. G., Chittenden, G. J., Buchanan, J. G., & Baddiley, J. (1970). Uridine diphosphate alpha-D-galactofuranose, an intermediate in the biosynthesis of galactofuranosyl residues. *The Biochemical Journal*, ***117***(3), 637–639.

23. Tanner, J. J., Boechi, L., Andrew McCammon, J., & Sobrado, P. (2014). Structure, mechanism, and dynamics of UDP-galactopyranose mutase. *Archives of Biochemistry and Biophysics*, ***544***, 128–141.

24. Soltero-Higgin, M., Carlson, E. E., Gruber, T. D., & Kiessling, L. L. (2004). A unique catalytic mechanism for UDP-galactopyranose mutase. *Nature Structural & Molecular Biology*, ***11***(6).

25. Misra S, Valicherla GR, Mohd Shahab, Gupta J, Gayen JR, Misra-Bhattacharya S. (2016). UDP-galactopyranose mutase, a potential drug target against human pathogenic nematode Brugia malayi. *Pathogens and Disease, FEMS*. ***74***(6): ftw072.

26. Beverley, S. M., Owens, K. L., Showalter, M., Griffith, C. L., Doering, T. L., Jones, V. C., & McNeil, M. R. (2005). Eukaryotic UDP-galactopyranose mutase (GLF gene) in microbial and metazoal pathogens. *Eukaryotic Cell*, ***4***(6), 1147–1154.

27. Dykhuizen, E. C., May, J. F., Tongpenyai, A., & Kiessling, L. L. (2008). Inhibitors of UDP-galactopyranose mutase thwart mycobacterial growth. *Journal of the American Chemical Society*, ***130***(21), 6706–6707.

28. Wangkanont, K., Winton, V. J., Forest, K. T., & Kiessling, L. L. (2017). Conformational Control of UDP-Galactopyranose Mutase Inhibition. *Biochemistry*, ***56***(30), 3983–3992.

29. Fosgerau K, Hoffmann T. Peptide therapeutics: current status and future directions. *Drug Discovery Today*. 2015; **20**(1):122-128.

30. Cunningham, A. D., Qvit, N., & Mochly-Rosen, D. (2017). Peptides and peptidomimetics as regulators of protein-protein interactions. *Current Opinion in Structural Biology*, ***44***, 59–66.

31. Banting, F. G., Best, C. H., Collip, J. B., Campbell, W. R., & Fletcher, A. A. (1922). Pancreatic Extracts in the Treatment of Diabetes Mellitus. *Canadian Medical Association Journal*, ***12***(3), 141–146.

32. du Vigneaud V, Ressler C, Swan JM, Roberts CW, Katsoyannis PG, Gordon S (1953). "The synthesis of an octapeptide amide with the hormonal activity of oxytocin". *Journal of American Chemical Society*. **75** (19): 4879–80.

33. Lau, J. L., & Dunn, M. K. (2018). Therapeutic peptides: Historical perspectives, current development trends, and future directions. *Bioorganic & Medicinal Chemistry*, ***26***(10), 2700–2707.

34. Zasloff M. (1987). Magainins, a class of antimicrobial peptides from Xenopus skin: isolation, characterization of two active forms, and partial cDNA sequence of a precursor. *Proceedings of the National Academy of Sciences of the United States of America*, ***84***(15), 5449–5453.

35. Hancock, R. E., & Chapple, D. S. (1999). Peptide antibiotics. *Antimicrobial agents and chemotherapy*, ***43***(6), 1317–1323.

36. Zasloff M. (2002). Antimicrobial peptides of multicellular organisms. *Nature*, ***415***(6870), 389–395.

37. Hancock, R. E., & Sahl, H. G. (2006). Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. *Nature Biotechnology*, ***24***(12), 1551–1557.

38. Kang, H. K., Kim, C., Seo, C. H., & Park, Y. (2017). The therapeutic applications of antimicrobial peptides (AMPs): a patent review. *Journal of Microbiology (Seoul, Korea)*, ***55***(1), 1–12.

39. Craik, D. J., Fairlie, D. P., Liras, S., & Price, D. (2013). The future of peptide-based drugs. *Chemical Biology & Drug Design*, ***81***(1), 136–147.

40. Köhler, G., & Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature*, ***256***(5517), 495–497.

41. Shah, D. K., & Betts, A. M. (2013). Antibody biodistribution coefficients: inferring tissue concentrations of monoclonal antibodies based on the plasma concentrations in several preclinical species and human. *mAbs*, ***5***(2), 297–305.

42. Skerra A. (2000) Engineered protein scaffolds for molecular recognition [published correction appears in J Mol Recognit 2001 Mar-Apr;14(2):141]. *J Mol Recognit*. ***13***(4):167-187.

43. Chames P, Van Regenmortel M, Weiss E, Baty D. (2009) Therapeutic antibodies: successes, limitations and hopes for the future. *British Journal of Pharmacology*. ***157***(2):220-233.

44. Binz, H. K., Amstutz, P., & Plückthun, A. (2005). Engineering novel binding proteins from nonimmunoglobulin domains. *Nature biotechnology*, ***23***(10), 1257–1268.

45. Hamers-Casterman, C., Atarhouch, T., Muyldermans, S., Robinson, G., Hamers, C., Songa, E. B., Bendahman, N., & Hamers, R. (1993). Naturally occurring antibodies devoid of light chains. *Nature*, ***363***(6428), 446–448.

46. Koide, A., Wojcik, J., Gilbreth, R. N., Hoey, R. J., & Koide, S. (2012). Teaching an old scaffold new tricks: monobodies constructed using alternative surfaces of the FN3 scaffold. *Journal of molecular biology*, ***415***(2), 393–405.

47. Stockbridge, R. B., Koide, A., Miller, C., & Koide, S. (2014). Proof of dual-topology architecture of Fluc F- channels with monobody blockers. *Nature communications*, ***5***, 5120.

48. Nord K, Gunneriusson E, Ringdahl J, Ståhl S, Uhlén M, Nygren PA. (1997) Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain. *Nat Biotechnol*. ***15***(8):772-777.

49. Driggers, E. M., Hale, S. P., Lee, J., & Terrett, N. K. (2008). The exploration of macrocycles for drug discovery--an underexploited structural class. *Nature reviews. Drug discovery*, ***7***(7), 608–624.

50. Dreyfuss, M; Härri, E; Hofmann, H; Kobel, H; Pache, W; Tscherter, H (1976). "Cyclosporin A and C: new metabolites from *Trichoderma polysporum* (Link ex Pers.) Rifai". *European Journal of Applied Microbiology*. ***3***: 125–133.

51. Bissett, J. (1983). "Notes on Tolypocladium and related genera". *Canadian Journal of Botany*. ***61***: 1311–1329.

52. Petcher, T. J., Weber, H., & Rüegger, A. (1976). Crystal and molecular structure of an iodo-derivative of the cyclic undecapeptide cyclosporin A. *Helvetica chimica acta*, ***59***(5), 1480–1489.

53. Schreiber, S. L., & Crabtree, G. R. (1992). The mechanism of action of cyclosporin A and FK506. *Immunology today*, ***13***(4), 136–142.

54. Leenheer, D., Ten Dijke, P., & Hipolito, C. J. (2016). A current perspective on applications of macrocyclic-peptide-based high-affinity ligands. *Biopolymers*, ***106***(6), 889–900.

55. Vinogradov, A. A., Yin, Y., & Suga, H. (2019). Macrocyclic Peptides as Drug Candidates: Recent Progress and Remaining Challenges. *Journal of the American Chemical Society*, ***141***(10), 4167–4181.

56. McGuire, J. M., Wolfe, R. N., & Ziegler, D. W. (1955). Vancomycin, a new antibiotic. II. *In vitro* antibacterial studies. *Antibiotics annual*, ***3***, 612–618.

57. Tally, F. P., & DeBruin, M. F. (2000). Development of daptomycin for gram-positive infections. *The Journal of antimicrobial chemotherapy*, ***46***(4), 523–526.

58. Haste, N. M., Perera, V. R., Maloney, K. N., Tran, D. N., Jensen, P., Fenical, W., Nizet, V., & Hensler, M. E. (2010). Activity of the streptogramin antibiotic etamycin against methicillin-resistant Staphylococcus aureus. *The Journal of antibiotics*, ***63***(5), 219–224.

59. Martis E.A., Radhakrishnan R., Badve R.R. (2011) High-throughput screening: The hits and leads of drug discovery—An overview. J. Appl. Pharm. Sci. 1:2–10.

60. Brandish, P. E., Chiu, C. S., Schneeweis, J., Brandon, N. J., Leech, C. L., Kornienko, O., Scolnick, E. M., Strulovici, B., & Zheng, W. (2006). A cell-based ultra-high-throughput screening assay for identifying inhibitors of D-amino acid oxidase. *Journal of biomolecular screening*, ***11***(5), 481–487.

61. Smith G. P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science (New York, N.Y.)*, ***228***(4705), 1315–1317.

62. Tavassoli, A., & Benkovic, S. J. (2007). Split-intein mediated circular ligation used in the synthesis of cyclic peptide libraries in *E. coli*. *Nature protocols*, ***2***(5), 1126–1133.

63. Wilson, D. S., Keefe, A. D. & Szostak, J. W. (2001) The use of mRNA display to select high-affinity protein-binding peptides, *Proc Natl Acad Sci USA.* ***98***, 3750-3755.

64. Murakami, H., Ohta, A., Ashigai, H. & Suga, H. (2006) A highly flexible tRNA acylation method for non-natural polypeptide synthesis, *Nat Meth*. ***3***, 357-359.

65. Goto, Y., Katoh, T. & Suga, H. (2011) Flexizymes for genetic code reprogramming, *Nat Protocols*. **6**, 779-790.

66. Hipolito, C. J. & Suga, H. (2012) Ribosomal production and *in vitro* selection of natural product-like peptidomimetics: The FIT and RaPID systems, *Curr Opin Chem Biol*. ***16***, 196-203.

67. Hipolito, C., Tanaka, Y., Katoh, T., Nureki, O. & Suga, H. (2013) A Macrocyclic Peptide that Serves as a Cocrystallization Ligand and Inhibits the Function of a MATE Family Transporter, Molecules. ***18***, 10514.

68. Sakai, K., Passioura, T., Sato, H., Ito, K., Furuhashi, H., Umitsu, M., Takagi, J., Kato, Y., Mukai, H., Warashina, S., Zouda, M., Watanabe, Y., Yano, S., Shibata, M., Suga, H., & Matsumoto, K. (2019). Macrocyclic peptide-based inhibition and imaging of hepatocyte growth factor. *Nature chemical biology*, ***15***(6), 598–606.

69. Song, X., Lu, L. Y., Passioura, T., & Suga, H. (2017). Macrocyclic peptide inhibitors for the protein-protein interaction of Zaire Ebola virus protein 24 and karyopherin alpha 5. *Organic & biomolecular chemistry*, ***15***(24), 5155–5160.

70. Nord, K., Nilsson, J., Nilsson, B., Uhlén, M., & Nygren, P. A. (1995). A combinatorial library of an alpha-helical bacterial receptor domain. *Protein engineering*, ***8***(6), 601–608.

71. Ståhl, S., Gräslund, T., Eriksson Karlström, A., Frejd, F. Y., Nygren, P. Å., & Löfblom, J. (2017). Affibody Molecules in Biotechnological and Medical Applications. *Trends in biotechnology*, ***35***(8), 691–712.

72. Löfblom, J., Feldwisch, J., Tolmachev, V., Carlsson, J., Ståhl, S., & Frejd, F. Y. (2010). Affibody molecules: engineered proteins for therapeutic, diagnostic and biotechnological applications. *FEBS letters*, ***584***(12), 2670–2680.

73. Wikman, M., Steffen, A. C., Gunneriusson, E., Tolmachev, V., Adams, G. P., Carlsson, J., & Ståhl, S. (2004). Selection and characterization of HER2/neu-binding affibody ligands. *Protein engineering, design & selection: PEDS*, ***17***(5), 455–462.

74. Orlova, A., Magnusson, M., Eriksson, T. L., Nilsson, M., Larsson, B., Höidén-Guthenberg, I., Widström, C., Carlsson, J., Tolmachev, V., Ståhl, S., & Nilsson, F. Y. (2006). Tumor imaging using a picomolar affinity HER2 binding affibody molecule. *Cancer research*, ***66***(8), 4339–4348.

75. Rinne, S. S., Leitao, C. D., Mitran, B., Bass, T. Z., Andersson, K. G., Tolmachev, V., Ståhl, S., Löfblom, J., & Orlova, A. (2019). Optimization of HER3 expression imaging using affibody molecules: Influence of chelator for labeling with indium-111. *Scientific reports*, *9*(1), 655.

76. Chen, G., Hayhurst, A., Thomas, J. G., Harvey, B. R., Iverson, B. L., & Georgiou, G. (2001). Isolation of high-affinity ligand-binding proteins by periplasmic expression with cytometric screening (PECS). *Nature biotechnology*, ***19***(6), 537–542.

77. Boder, E. T., & Wittrup, K. D. (1997). Yeast surface display for screening combinatorial polypeptide libraries. *Nature biotechnology*, ***15***(6), 553–557.

78. FitzGerald K. (2000). *In vitro* display technologies - new tools for drug discovery. *Drug discovery today*, *5*(6), 253–258.

79. Roberts, R. W., & Szostak, J. W. (1997). RNA-peptide fusions for the *in vitro* selection of peptides and proteins. *Proceedings of the National Academy of Sciences of the United States of America*, ***94***(23), 12297–12302.

80. Baggio, R., Burgstaller, P., Hale, S. P., Putney, A. R., Lane, M., Lipovsek, D., Wright, M. C., Roberts, R. W., Liu, R., Szostak, J. W., & Wagner, R. W. (2002). Identification of epitope-like consensus motifs using mRNA display. *Journal of molecular recognition: JMR*, ***15***(3), 126–134.

81. Mattheakis, L. C., Bhatt, R. R., & Dower, W. J. (1994). An *in vitro* polysome display system for identifying ligands from very large peptide libraries. *Proceedings of the National Academy of Sciences of the United States of America*, ***91***(19), 9022–9026.

82. Reiersen, H., Løbersli, I., Løset, G. Å., Hvattum, E., Simonsen, B., Stacy, J. E., McGregor, D., Fitzgerald, K., Welschof, M., Brekke, O. H., & Marvik, O. J. (2005). Covalent antibody display--an *in vitro* antibody-DNA library selection system. *Nucleic acids research*, ***33***(1), e10.

83. Carlson, E. E., May, J. F., & Kiessling, L. L. (2006). Chemical probes of UDP-galactopyranose mutase. *Chemistry & biology*, ***13***(8), 825–837.

84. van Straaten, K. E., Kuttiyatveetil, J. R., Sevrain, C. M., Villaume, S. A., Jiménez-Barbero, J., Linclau, B., Vincent, S. P., & Sanders, D. A. (2015). Structural basis of ligand binding to UDP-galactopyranose mutase from Mycobacterium tuberculosis using substrate and tetrafluorinated substrate analogues. *Journal of the American Chemical Society*, ***137***(3), 1230–1244.

85. Tropea, J. E., Cherry, S., & Waugh, D. S. (2009). Expression and purification of soluble His(6)-tagged TEV protease. *Methods in molecular biology (Clifton, N.J.)*, ***498***, 297–307.

86. Kapust RB, Tözsér J, Fox JD, Anderson DE, Cherry S, Copeland TD, Waugh DS. (2001) Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng.* **14**(12):993-1000.

87. Chad, J. M., Sarathy, K. P., Gruber, T. D., Addala, E., Kiessling, L. L., & Sanders, D. A. (2007). Site-directed mutagenesis of UDP-galactopyranose mutase reveals a critical role for the active-site, conserved arginine residues. *Biochemistry*, ***46***(23), 6723–6732.

88. Dougherty, P. G., Qian, Z., & Pei, D. (2017). Macrocycles as protein-protein interaction inhibitors. *The Biochemical journal*, *474*(7), 1109–1125.

APPENDIX

LIST OF FIGURES

LIST OF TABLES

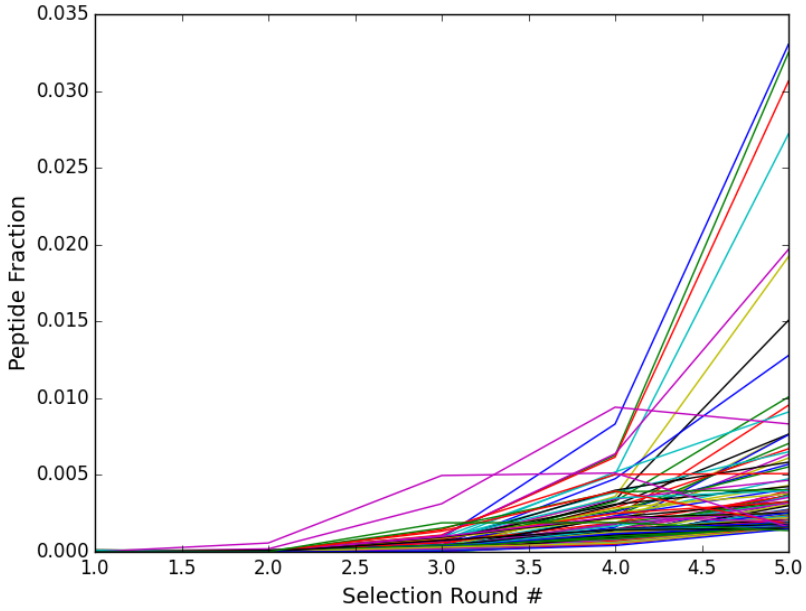


Figure A1. Evolution graph of all MCP sequences detected via Next-gen sequencing through the selection progress. (plotted by Hipolito C. J.)

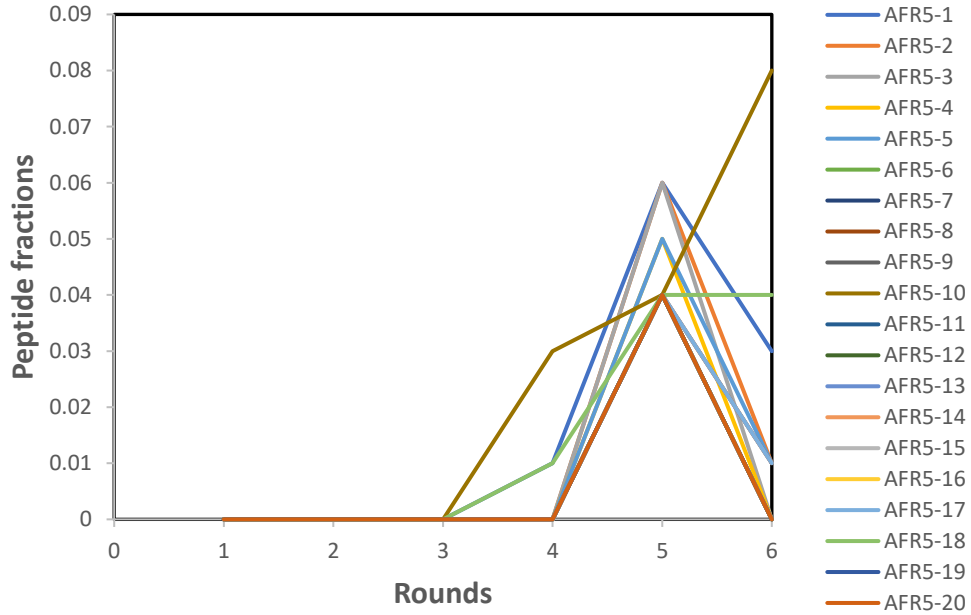
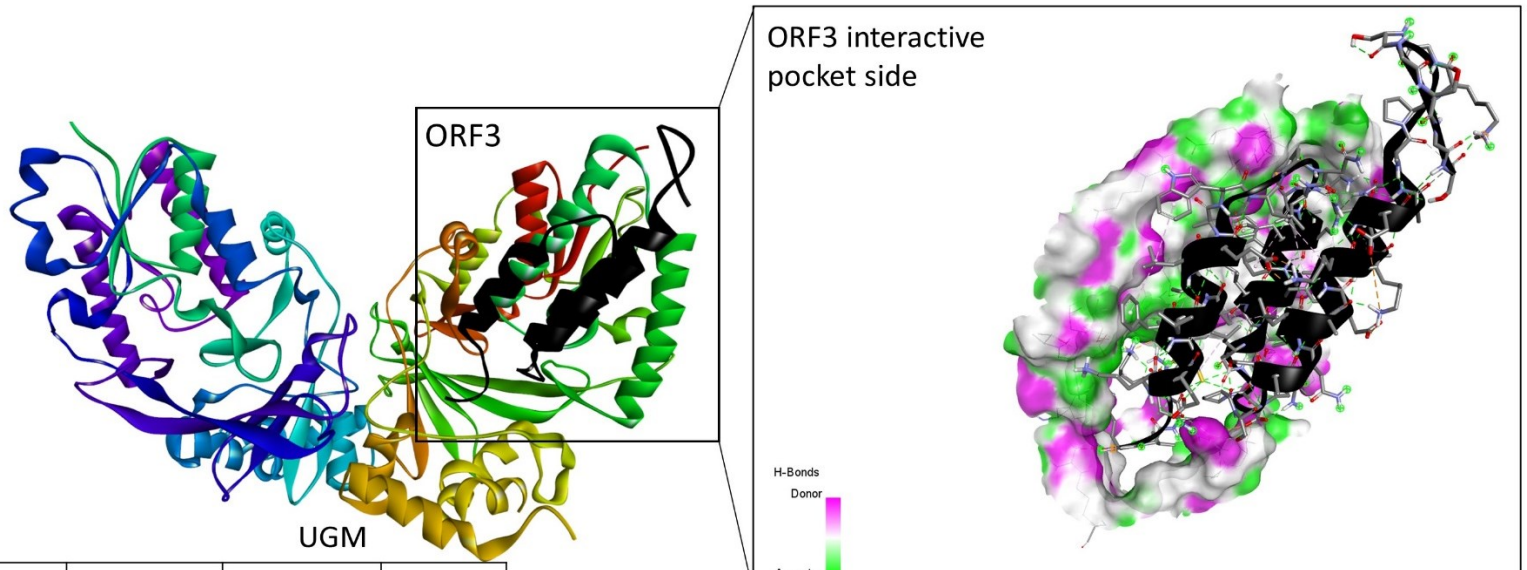


Figure A2. Evolution graph plotted for the top 20 sequences of round 5 affibody selection.



Model	Cluster Members	Representative	Weighted Score
1	159	Center	-1005.3
		Lowest Energy	-1211.6

Thanasis Poullikkas, University of Tsukuba JAPAN

Figure A3. Preliminary docking analysis against MtUGM with Affibody

a

	M.W. (kDa)
*Mt*UGM	43.9
MBP	42.5
His_6-TEV	30 (approx.)
MBP-His_6-TEV	70 (approx.)
His_6-MBP-MtUGM	86.4

b

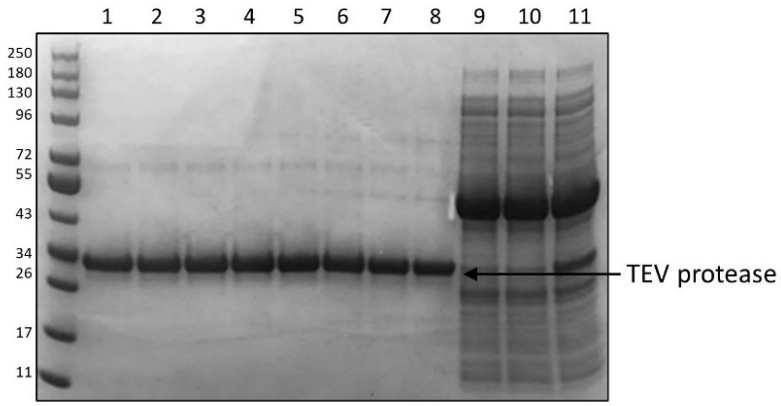


Figure A4. Recombinant protein expression of TEV protease (a) pRK793-MBP-His-TEV construct (b) SDS-PAGE analysis of purified TEV. Lanes 1 to 8 contains fractions with desired proteins, bands at ~70 kDa corresponds to uncleaved MBP-His-TEV protein and bands at ~30 kDa corresponds to His-TEV. Lane 9- wash fraction where buffer was applied to elute non-desired proteins; lane 10- flow-through fractions when lysate was applied to Ni-NTA column, and Lane 11- crude lysate. (performed by Bo Yi Han)

Table A1. List of buffers.

Name	Components	pH
***Mt*UGM Binding Buffer**	50 mM, NaCl 200 mM, EDTA 1 mM	8
***Mt*UGM Elution Buffer**	50 mM Tris, 200 mM NaCl, 1mM EDTA, 50 mM Maltose	8
***Mt*UGM Storage Buffer**	25 mM Tris, 500 mM NaCl	7.5
TGI Buffer	20 mM Tris, 250 mM NaCl, 20 mM Imidazole	7.5
TEV Protease Elution Buffer	20 mM Tris, 250 mM NaCl, 220 mM Imidazole	7.5
TEV Protease Storage Buffer	25 mM Tris, 500 mM NaCl	7.5
***Mt*UGM Selection Buffer**	500 mM NaCl, 25 mM Tris-Cl, 0.005% TWEEN 20	7.5
Affibody Binding Buffer	25 mM Tris, 10 mM imidazole, 500 mM NaCl	7.5
Affibody Elution Buffer	25 mM Tris, 200 mM NaCl, 500 mM Imidazole	7.5
Affibody Storage Buffer	50 mM Tris, 150 mM NaCl	7.5

Table A2. List of primers.

Name	Sequence
fMetE-F	G*TAATACGACTCACTATAG*GCGGGGTGGAGCAGCCTGGTAGCTCGTCG
fMetE-R_2	T**G**GTTGCGGGGGCCGGATTTGAACCGACGATCTTCGGGT
fMetE-R_1	GAACCGACGATCTTCGGGTTATGAGCCCGACGAGCTACCAGGCTGCTC
eFxF66	GGCGTAATACGACTCACTATAGGATCGAAAGATTTCCGCGGCCCCGAAAGGGGATTAGCGTTAGGT
eFx-AmC-R18	A[2OMeC] CTAACGCTAATCCCCT
AFR6-F	GAAGTGCCATTCCGCCTGAC
AFR6-R	CACTGAGCCTCCACCTAGCC
AFR6-insert-F	TAATACGACTCACTATAGGG
AFR6-insert-R	GCTAGTTATTGCTCAGCGG
T7-CH-F46	TAATACGACTCACTATAGGGTTGAACTTTAAGTAGGAGATATATCC
CGS3an13-R39	TTTCCGCCCCCCGTCCTAGCTGCCGCTGCCGCTGCCGCA
CGS3-CH-R24	TTTCCGCCCCCCGTCCTAAGACCC