The identification of novel *Mycobacterium tuberculosis* immunogenic adhesin proteins and the evaluation as a potential vaccine and diagnostic candidate

by

Rivesh Maharajh

Submitted in fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY (Medicine)

Discipline of Medical Microbiology

School of Medicine

Faculty of Health Science

UNIVERSITY OF KWAZULU-NATAL

2026

Declaration

I, Rivesh Maharajh, declare that

(i) The research reported in this dissertation, except where otherwise indicated, is my original work.

(ii) This thesis has not been submitted for any degree or examination at any other university.

(iii) This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

(iv) This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

a) their words have been re-written but the general information attributed to them has been referenced;

b) where their exact words have been used, their writing has been placed inside quotation marks, and referenced.

(v) Where I have reproduced a publication of which I am an author, co-author or editor, I have indicated in detail which part of the publication was actually written by myself alone and have fully referenced such publications.

(vi) This thesis does not contain text, graphics, or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References section.

Signed: [redacted] Date: 28/08/2026

PhD Candidate: Rivesh Maharajh

As the candidate's supervisor and co-supervisor, this Doctoral Thesis has been approved for submission:

Signed: [redacted] Date: 28/08/2026

S[redacted]usiso Senzani

Signed: [redacted] Date: 28/08/2026

Co-supervisor: Prof. Manormoney Pillay

Manuscripts for Publication

List of publications emanating from the PhD degree:

1. **Rivesh Maharajh**, Manormoney Pillay, and Sibusiso Senzani. **"A Computational Method for the Prediction and Functional Analysis of Potential *Mycobacterium tuberculosis* Adhesin-Related Proteins."** Published by the *Expert Review of Proteomics* Journal in October 2023; 20(12): 483-493. https://doi.org/10.1080/14789450.2023.2275678.

Contributions: The complete publication is presented in Chapter 3 of the thesis. The article was written by myself and approved by my supervisors, Dr. S Senzani and Prof. M Pillay. All three authors were involved in the conception and design of the study. I have provided analysis and interpretation of the data with support of Dr. S Senzani. I had drafted the manuscript, while my supervisors revised the manuscript for intellectual content. All authors provide the final approval of the version to be published and agree to be accountable for all aspects of the work.

2. **Rivesh Maharajh**, , Sibusiso Senzani, Balakrishna Pillay, and Manormoney Pillay. **" The use of GST/His affinity tag expression systems for recombinant production of solubilized *Mycobacterium tuberculosis* conserved hypothetical proteins with adhesin potential."**. Manuscript currently prepared for the *Molecular Microbiology* Journal

Contributions: The completed manuscript is presented in Chapter 4 of the thesis. The manuscript was written by myself and approved by my supervisors, Dr. S Senzani and Prof. M Pillay with collaborator Professor Balakrishna Pillay. All four authors were involved in the conception and design of the study. I performed all the laboratory experiments and data analysis with support from Dr. S Senzani and Prof. M Pillay. The final manuscript was drafted by myself and revised by all the authors.

3. **Rivesh Maharajh**, Manormoney Pillay, Ndumiso Mhlongo and Sibusiso Senzani. **" Validation of Novel *Mycobacterium tuberculosis* adhesin with associated extracellular matrix components: a Molecular docking and Phenotypic Approach"**. Manuscript under review by *Proteomics* (Wiley) Journal. Submitted May 2026.

Contributions: The completed manuscript is presented in Chapter 5 of the thesis. The manuscript was written by myself and approved by my supervisors, Dr. S Senzani and Prof. M Pillay, with collaborator Dr. N Mhlongo. All three authors were involved in the conception and design of the study. I performed the molecular docking analysis with guidance from Dr. N Mhlongo, completed all experimental assays and data analysis with support from Dr. S Senzani and Prof. M Pillay. The final manuscript was drafted by myself and revised by all the authors prior to submission to the journal.

4. **Rivesh Maharajh**, Sibusiso Senzani, Veron Ramsuran Balakrishna Pillay, and Manormoney Pillay. **"Evaluation of novel adhesin-related *Mycobacterium tuberculosis* conserved hypothetical protein biomarkers: an immunoinformatic and serological approach."** Manuscript currently prepared for the *Tuberculosis and Respiratory Disease* Journal

Contributions: The completed manuscript is presented in Chapter 6 of the thesis. The manuscript was written by myself and approved by my supervisors, Dr. S Senzani and Prof. M Pillay, with collaborators Professor Veron Ramsuran and Professor Balakrishna Pillay. My supervisors and I were involved in the conception and design of the study. Prof Veron Ramsuran provided biological samples and statistical advice. Prof. Balakrishna provided scientific advice and guidance. I performed all the laboratory experiments and data analysis with support from Dr. S Senzani and Prof. M Pillay. The final manuscript was drafted by myself and revised by all the authors.

Conference Proceedings

1. **Rivesh Maharajh.** "*in Silico* identification of potential *Mycobacterium tuberculosis* conserved hypothetical Adhesin Proteins using SPAAN Bioinformatics". Poster Presentation at the 7th South African TB Conference in September 2022.
2. **Rivesh Maharajh. "**Identification of *Mycobacterium tuberculosis* immunogenic adhesin proteins and their application as a potential vaccine and diagnostic candidate". Oral Presentation at the South African Medical Research Council 16th Early Scientist Convention in October 2022.
3. **Rivesh Maharajh.** "A Computational method for the prediction and functional analysis of potential *Mycobacterium tuberculosis* adhesin-related proteins with prospects for future research". Oral Presentation at the South African Medical Research Council RCD Beneficiary Conference in November 2023.
4. **Rivesh Maharajh.** "A Computational Method for the Prediction and Functional Analysis of Potential *Mycobacterium tuberculosis* Adhesin-Related Proteins". Oral Presentation at the University of KwaZulu-Natal College of Health Science Research Symposium in August 2024.
5. **Rivesh Maharajh. "** The structural interactions and validation *of Mycobacterium tuberculosis* adhesins with associated extracellular matrix components". Oral Presentation at the University of KwaZulu-Natal College of Health Science Research Symposium in August 2025.

Dedication

I dedicate my thesis to my family: My wife ***Nikisha Maharajh****, to my Daughter* ***Kitana Maharajh*** *and to my parents* ***Mr. Royheath Maharajh*** *and* ***Mrs. Leelawathie Maharajh***

And to all that has led me to this path in life:

“"If I have seen further, it is by standing on the shoulders of giants."

-Sir Isaac Newton

Acknowledgements

I would like to thank God Almighty for all the blessings I have received throughout this academic journey. Without prayer and belief, I would not have reached this destination.

The work undertaken would not have been possible without the support and contributions of the individuals listed below with utmost gratitude:

First and foremost, I wish to express my deepest heartfelt appreciation to my supervisors, Dr. Sibusiso Senzani and Professor Manormoney Pillay. Your unwavering support, patience, and encouragement have carried me through every stage of this PhD journey. I am profoundly grateful for the invaluable advice you have shared and the commitment you have shown to my growth as a researcher. Your scientific knowledge and guidance have not only shaped this project but have also laid the foundation for my future career. Thank you for giving me the incredible opportunity to pursue this PhD and for equipping me with skills and insights that I will cherish and take forward with me throughout my life.

I would like to extend my sincere gratitude to our collaborators. Dr. Ndumiso Mhlongo, your generosity in sharing your expertise in bioinformatics and critical thinking has been invaluable, and I am deeply appreciative of the knowledge I have gained through your guidance. Professor Balakrishna Pillay, your exceptional advice and the thoughtful direction you provided during our meetings have greatly enriched this work. I am profoundly thankful to you both for contributing to the development of this PhD and for sharing your intellectual insights. To Professor Veron Ramsuran, for his kind contribution of patient samples and advice provided for the statistical analysis, we are grateful for your contribution to this PhD project.

To the staff and students at the Discipline of Medical Microbiology. Zareena Solwa, Inga Elson, and Palesa Shemfe, each one of you have assisted me whenever I needed support within the laboratory. It has been a pleasure being in a laboratory where I can have a good relationship with those who are always willing to assist. To the students, each one has supported my work in some small way, especially my research colleagues Koobashnee Pillay and Nikita Deyal. It has been a pleasure working in the protein laboratory, sharing your expertise and knowledge with me among

the challenges shared with our projects. It made this journey comfortable knowing I could talk to someone who truly understands and inspires me to succeed.

To the College of Health Science scholarship programme, UKZN APACHE flagship grant, and the South African Bongani Mayosi Health Scholars Programme, I am grateful for the financial support provided throughout my PhD.

And finally, to my family, words cannot truly capture the depth of love and support you have given me throughout this journey. To my parents, thank you for standing beside me and encouraging me to pursue this chosen path in my career. Dad, your constant reminder that “an education is something no one can ever take away from you” has motivated me throughout my life as I strive to make you proud. Mum, your example that hard work builds character has inspired me to always perform at my best. To my wife, I am endlessly grateful for your patience and strength in enduring this journey with me. From the very first day, your unwavering support has been a blessing, and you are truly God-sent. To my siblings and in-laws, your encouragement has been a source of motivation, helping me push forward in my studies. To my daughter, you are my everything. Your love and presence have been a constant source of joy and strength. I hope that through this achievement, I continue to make you proud, just as you have made me proud every single day.

List of abbreviations

Acronym	Field Name
αβγ	Alpha beta gamma
AA	Amino acids
Ab	Antibody
AM	Alveolar macrophages
ATB	Active tuberculosis
AUC	Area under the curve
BCG	Bacillus Calmette-Guérin
BREC	Biomedical Research Ethics Committee
CH	Conserved hypothetical
COVID-19	Coronavirus disease 2019
COV	Coefficient of variation
CFP-10	Culture filtrate protein-10
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
CTAB	Cetyl trimethylammonium bromide
DR-TB	Drug resistant tuberculosis
DNA	Deoxyribonucleic acid
DST	Drug susceptibility testing
DTT	Dithiothreitol
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
ESAT-6	Early secretory antigen 6 kDa
ESX-1	ESAT-6 secretion system-1
ELISA	Enzyme-linked immunosorbent assay
FnBPs	Fibronectin binding proteins
EPS	Extracellular polymeric substance

GST	Glutathione-S-transferase
His	Histidine
HIV	Human immunodeficiency virus
Ig	Immunoglobulin
IFN-γ	Interferon gamma
kDa	Kilodalton
LB	Luria-Bértani
LF-LAM	Lateral flow urine lipoarabinomannan
LTBI	Latent tuberculosis infection
M	Molar
MDR-TB	Multidrug Resistant Tuberculosis
MOTT	Microorganisms other than Tuberculosis
MmpL	Mycobacterial membrane protein large
NAATs	Nucleic acid amplification tests
PAMPs	Pathogen-associated molecular patterns
PBS	Phosphate buffered saline
PBSt	Phosphate buffered saline with Tween 20
PDB	Protein database
POC	Point of care
P/N	Positive-to-negative
PstS1	Phosphate-binding periplasmic lipoprotein
Rec-	Recombinant
RIPE	Rifampicin, isoniazid, pyrazinamide, and ethambutol
RNA	Ribonucleic acid
ROC	Receiver operating characteristics
RR-TB	Rifampicin resistant tuberculosis
SA	South Africa
SCL	Subcellular localization
SDS	Sodium dodecyl sulfate

SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SPAAN	Software program to predict adhesin and adhesin-like protein using artificial neural networks
TB	Tuberculosis
TGF-β	Transforming growth factor-beta
TLRs	Toll-like receptors
TMB	3,3'5,5'Tetramethylbenzidine
TNF	Tumor necrosis factor
TPT	TB preventive treatment
VDW	Van der Waals
WHO	World Health Organization
WRD	WHO recommended rapid diagnostics

Abstract

Background: *Mycobacterium tuberculosis*, the etiological agent of tuberculosis, has been the leading cause of mortality among communicable diseases, with 10.7 million reported active cases worldwide. Currently, the Bacillus Calmette-Guérin vaccine remains the only commercially available tuberculosis vaccine with suboptimal protection in adult populations. In addition, early diagnosis and treatment monitoring remain challenging, particularly in high-burden countries. The *M. tuberculosis* proteome contains more than 4000 proteins, of which 25 percent remain functionally unexplored and are designated conserved hypothetical proteins. These conserved hypothetical proteins are genetic constituents of *M. tuberculosis* and are preserved across all lineages. Host-pathogen interaction promotes colonization of *M. tuberculosis* and subsequently provides a target for the host immune response. Therefore, adhesins can enhance the pathogenicity of *M. tuberculosis* as these surface proteins are the initial contact with the host cellular environment. The aim of the study was to identify novel adhesins from conserved hypothetical proteins for their immunogenic biomarker potential as rapid diagnostic tools and for novel therapeutic evaluation to combat *M. tuberculosis* infection.

Methods: This study integrated bioinformatics and experimental approaches to identify novel immunogenic adhesin proteins. Initially, SPAAN (a software program to predict adhesins and adhesin-like proteins using artificial neural networks) was used to screen the conserved hypothetical protein group to predict adhesins with the highest probability of adhesin potential. Coupled with systematic computational tools, putative adhesins were selected and assessed for their interaction with host cellular proteins. Seven *M. tuberculosis* conserved hypothetical proteins (Rv3811, Rv3717, Rv3491, Rv3209, Rv2730, Rv2598, and Rv2272) were selected to validate adhesin function and immunogenic properties. Recombinant production of these *M. tuberculosis* surface proteins was performed using Glutathione-S-transferase (GST)/Histidine (His) affinity tag expression systems specific to downstream applications. Phenotypic validation ensued via an indirect enzyme-linked immunosorbent assay (ELISA) to determine the binding affinity of recombinant GST-affinity tag proteins to human-derived extracellular matrix components (laminin, fibronectin, and collagen). In addition, molecular docking techniques were employed to expand on the molecular interactions between the extracellular matrix component (ECM) and

putative adhesins. The immunogenic potential of these *M. tuberculosis* conserved hypotheticals was screened using immunoinformatic tools that identify epitope regions and simulate humoral and cellular responses. A serological evaluation was performed among GeneXpert-confirmed 50 TB-positive and 50 TB-negative South African plasma samples (N=100). An ELISA was performed to detect antigen-specific immunoglobulin G (IgG) antibodies towards recombinant His-affinity tag proteins (Rv3811, Rv3717, Rv3491, Rv3209, Rv2598).

Results and Discussion: The adhesin-related bioinformatic analysis identified 167 conserved hypotheticals within the probability of adhesin (P_{ad}) threshold (≤ 0.51). Compared to known *M. tuberculosis* adhesins (n=34), approximately one-third fell below the threshold. This study established a selection criteria based on the P_{ad} index (and immunological properties), predicted function using homology modelling, the use of translocation tools, and the identification of potential secreted proteins. A selection of the dataset was used to verify adhesin potential as well as immunogenic potential through the production of recombinant *M. tuberculosis* conserved hypothetical proteins. The hydrophobic nature of these *M. tuberculosis* proteins proved challenging to solubilize while retaining biological conformation. The GST-affinity tag proteins were solubilized with 1% sarkosyl, and His-affinity tag proteins were solubilized with 8 Molar urea to dissociate aggregates of recombinant proteins from inclusion bodies. However, the GST-tag proved difficult to cleave, and cloning into a His-affinity vector poorly expressed Rv2272, but not Rv2730, as this inhibited *Escherichia coli* BL21 (DE3) growth. Evaluation of adhesin function was performed by molecular docking, which identified a map of key binding residue sites with a high number of Van der Waals atomic bonds between adhesin-ECM interactions. The dominant interactive ECM binding ranges were predicted for collagen (range: L1470-T1690), fibronectin (range: D330-R938, H1135-Q1594), and laminin (Range: E496-V1152). Phenotypic validation by the adhesin assay revealed that Rv3491 (binding intensity: 81%, 83%, and 92%) and Rv2730 (binding intensity: 61, 71%, and 89%) were associated with strong binding affinities to collagen, fibronectin, and laminin, respectively. In addition, Rv3811 (binding intensity: 74%) was associated with very strong fibronectin-binding potential. Immunological evaluation revealed that Rv3811(p = 0.036) and Rv2598 (p = 0.012) showed moderate detection in antigen-specific IgG antibodies compared to Rv3717 (p = 0.873) and Rv3209 (p = 0.535). However, Rv3491 (p < 0.001) reported higher IgG detection in TB-negative plasma samples, attributed to early correlates of clearance.

The receiver operating characteristics curve analysis showed a better performance for Rv3811 (Area under the curve, AUC = 0.622, $p = 0.030$, 50% sensitivity and 74% specificity) and Rv2598 (AUC = 0.646, $p = 0.008$, 32% sensitivity and 98% specificity), whereby Rv3717, Rv3209 and PstS1 showed poor outcome (AUC = 0.466-0.511, $p > 0.050$, < 46% sensitivity and < 88% specificity). The serological study showed diversity among individual patients in detecting single antigens within the South African cohort.

Conclusion

The implementation of bioinformatics and experimental approaches has shown a synergistic relationship within the study. Computational analysis revealed various functional and structural proteomic attributes of the conserved hypothetical proteins, leading to the identification of novel adhesins (Rv3811, Rv3491, and Rv2730) and potential immunogenic biomarkers (Rv3811 and Rv2598). Although these novel adhesins demonstrated poor diagnostic outcomes and are unable to meet the minimum World Health Organization requirement of 70% sensitivity for target product profiles by point-of-care testing, the selected proteins may hold therapeutic potential due to their ability to induce a host immune response. Furthermore, the adherence function and immunological properties of the dataset of adhesins/adhesin-related proteins require further evaluation to support the discovery of novel biomarkers.

Table of contents

Declaration ..... **Error! Bookmark not defined.**
Manuscripts for Publication .....
Conference Proceedings ii
Dedication iii
Acknowledgements iv
List of Abbreviations vi
Abstract ix
Table of Contents xii
List of Tables xv
List of Figures xvi
Chapter 1: Introduction 1
1.1 Background 1
1.2 Novelty and innovation 2
1.3 Rationale for the PhD thesis 2
1.4 Aim and objectives 3
Chapter 2: Literature Review 6
2.1 Introduction 6
2.2 The global concern of tuberculosis infection 6
2.2.1 The worldwide epidemic 7
2.2.2 The burden of tuberculosis infection in South Africa 8
2.3 Current diagnosis and treatment 8
2..3.1 Diagnosis of tuberculosis 8
2.3.2 Tuberculosis treatment strategies 11
2.4. Pathogenesis of tuberculosis 12
2.4.1 Etiology of tuberculosis 12
2.4.2 Structural characteristics of *Mycobacterium tuberculosis* 13
2.4.3 Modulation of the host response 15
2.5 Adhesin molecules in *Mycobacterium tuberculosis* 17
2.5.1 Role of adhesins 18

2.5.2 Human extracellular matrix components 19
2.5.3 Currently known *Mycobacterium tuberculosis* associated adhesins. 22
2.6. Therapeutic application 24
2.6.1 Vaccine developmental research 25
2.6.2 Serodiagnosis development 28
2.7 Mycobacterial proteomics 30
2.7.1 Bioinformatic approach 30
2.7.1.1 Adhesin and adhesin-like prediction 30
2.7.1.2 Structural prediction of adhesins 31
2.7.1.3 Molecular docking approach for adhesin identification 32
2.7.2 Mycobacterial protein synthesis 34
Chapter 3: Identification of Adhesin-related Conserved Hypothetical Proteins of *Mycobacterium tuberculosis* 53
3.1. Abstract 54
3.2. Introduction 55
3.3. Materials and Methods 57
3.4. Results and Discussion 60
3.5. Conclusion 71
Chapter 4: The Cloning, Expression, and Purification of Recombinant Proteins of *Mycobacterium tuberculosis* 81
4.1. Abstract 82
4.2 Introduction 83
4.3. Materials and Methods 84
4.4. Results 95
4.5. Discussion 103
4.6. Conclusion 105
Chapter 5: Validation of *Mycobacterium tuberculosis* Conserved Hypothetical Proteins with Adhesin Potential 114
5.1. Abstract 115
5.2. Introduction 116
5.3. Materials and Methods 117
5.4. Results 121
5.5. Discussion 136
5.6. Conclusion 139

Chapter 6: Immunological Evaluation of Conserved Hypothetical Proteins of *Mycobacterium tuberculosis* 146
6.1. Abstract 147
6.2. Introduction 148
6.3. Materials and Methods 150
6.4. Results 156
6.5. Discussion 171
6.6. Conclusion 178
Chapter 7: Synthesis of Research Findings 189
Appendices 203
Appendix A: Research approval 203
Appendix B: Supplementary data 205
B1. Chapter three supplementary data 205
B2. Chapter four supplementary data 217
B3. Chapter five supplementary data 227
B4. Chapter six supplementary data 241
Appendix C. Approved permission to use figures in chapter 2 literature review. 258
Appendix D. Turnitin report 261

List of tables

Table 2.1 List of known *Mycobacterium tuberculosis* adhesin proteins. 23
Table 2.2: Current tuberculosis vaccine candidates under clinical development and active trials27
Table 3.1. Conserved hypothetical genes with predicted function using Phyre2 homology prediction analysis (P_{ad}≥0.51)..... 66
Table 3.2. Subcellular localization and signal peptide analysis of the SPAAN predicted conserved hypothetical genes. 68
Table 4.2. Optimized expression conditions of novel glutathione-S-transferase (GST) and 6x Histidine (His) – affinity tagged proteins. 88
Table 4.3. Purification of GST/His-affinity tag proteins using various parameters to improve protein yields..... 93
Table 5.1. Computational characteristic analysis of potential adhesin proteins 123
Table 5.2. Representation of the most stable receptor-ligand binding affinity model according to docking scores (kcal/mol). 124
Table 5.3. Identification of *Mycobacterium tuberculosis* adhesin key binding residues interacting with extracellular matrix (ECM) proteins. 130
Table 5.4. Molecular interactive forces between *Mycobacterium tuberculosis* protein and ECM based on the PDBsum simulation. 131
Table 5.5. Comparison of the binding intensity of *Mycobacterium tuberculosis* conserved hypothetical proteins to extracellular matrix proteins. 135
Table 6.1. Demographic data of participants with GeneXpert confirmed TB status (N=100)... 153
Table 6.2. Optimized parameters for improved ELISA performance and detection of TB-positive and TB-negative blood plasma samples. 155
Table 6.3. Characteristics of *Mycobacterium tuberculosis* conserved hypothetical proteins. 157
Table 6.4. Fibronectin (Fn) binding potential of *Mycobacterium tuberculosis* Conserved Hypothetical proteins at $Absorbance_{450nm.}$ 162
Table 6.5. Multivariable logistic regression model of demographic factors associated with GeneXpert tuberculosis status (N=86). 164
Table 6.6. Performance characteristics of the ELISA to determine optimal sensitivity and specificity of *Mycobacterium tuberculosis* conserved hypothetical proteins in TB-positive and TB-negative plasma samples. 167
Table 6.7. The association of immunoreactivity of the *Mycobacterium tuberculosis* protein detection in ELISA in association with the TB outcome using a binary regression model..... 168
Table 6.8. Binary logistic regression analysis for predicting TB outcomes, showing the main and interaction effects for all combinations..... 169
Table 6.8. Binary logistic regression analysis for predicting TB outcomes, showing the main and interaction effects for all combinations. Continued… 170
Table 6.9. Comparison of *Mycobacterium tuberculosis* antigen-specific antibody detection among HIV/ TB groups (p<0.05 significance) 171

List of figures

Figure 2.1. The estimated tuberculosis incidence in countries that reported over 100,000 cases, as reported by the WHO in 2024. 7
Figure 2.2 Universal access to the World Health Organization has recommended rapid diagnostics (WRD) to increase detection and reduce diagnostic time in 2023. 9
Figure 2.3. Anti-TB drug development pipeline for the treatment of tuberculosis. 12
Figure 2.4. The pathway of *Mycobacterium tuberculosis* infection. 13
Figure 2.5. An illustrative composition of the mycobacterial cell envelope. 14
Figure 2.6. Eukaryotic basement membrane comprises an array of extracellular matrix components (ECM). 20
Figure 2.7. The spectrum of tuberculosis infection illustrates the stages of progression with prevention and control strategies. 25
Figure 3.1. Flow diagram of the selection of conserved hypothetical proteins predicted by computational analysis. 58
Figure 3.2 Functional categories of the *Mycobacterium tuberculosis* (MtbH37Rv) proteome identified by SPAAN analysis categorized by the Mycobrowser tool (N= 776). 63
Figure 3.3. Protein homology prediction of conserved hypothetical proteins (n=167) using Phyre2 analysis. 64
Figure 4.1. Solubilization of GST-affinity-tag proteins. 90
Figure 4.2. Solubilization of His-affinity-tag proteins. 91
Figure 4.3. The detection of the signal peptide cleavage site (CS) of the conserved hypothetical *Mycobacterium tuberculosis* protein sequence, as shown by SignalP 6.0. 96
Figure 4.4. SDS-PAGE gels (10%) of pilot expressed GST-affinity-tag induced *Mycobacterium tuberculosis* conserved hypothetical proteins at various time intervals. 98
Figure 4.5. SDS-PAGE gels (12%) of pilot expressed His-affinity-tag induced *Mycobacterium tuberculosis* conserved hypothetical proteins at 16°C overnight. 99
Figure 4.6. SDS-PAGE gel (12%) of purified GST-affinity tag proteins by Pierce glutathione agarose (Thermo Fisher Scientific) by batch method purification. 101
Figure 4.7. SDS-PAGE gel (12%) of purified His-affinity tag proteins using HisPur Cobalt Resin (Thermo Fisher Scientific) by batch method purification. 102
Figure 5.1. Flow diagram of the ELISA for adhesin binding detection. 121
Figure 5.2. Predicted molecular binding interaction between collagen (receptor) and *Mycobacterium tuberculosis* adhesin proteins (ligand) within an ECM range of L1470-T1690. 126
Figure 5.3. Predicted molecular interaction between fibronectin (receptor) and *Mycobacterium tuberculosis* adhesin proteins (ligand) within an ECM range of D330-R938 and H1135-Q1594. 127

Figure 5.4. Predicted molecular interaction between laminin (receptor) and *Mycobacterium tuberculosis* adhesin proteins (ligand) within an ECM range of E496- V1152..... 129
Figure 5.5. A. SDS-PAGE (15% gel) of batch method purified GST-fused proteins using sarkosyl purification method, and B. Identification of GST-fused proteins by western blot. 132
Figure 5.6. Adhesin-binding potential of conserved hypothetical *Mycobacterium tuberculosis* GST-recombinant proteins among host extracellular matrix components. 134
Figure 6.1. Flow diagram of the ELISA detection of the fibronectin binding potential of His-affinity tag recombinant *Mycobacterium tuberculosis* conserved hypothetical proteins. 152
Figure 6.2. Exposed epitope regions of *Mycobacterium tuberculosis* conserved hypothetical proteins are shown as discontinuous regions determined by 1) Discotope 2.0 (visualized using Jmol) and 2) Ellipro (visualized using ChimeraX). 159
Figure 6.3. Simulation of the (A) Humoral response following simulated immunization and (B) levels of cytokine response to *Mycobacterium tuberculosis* antigens 161
Figure 6.4. Fibronectin (Fn) binding activity at OD absorbance at 450nm of 6xHis-affinity tag *Mycobacterium tuberculosis* conserved hypothetical proteins. 163
Figure 6.5. ELISA detection of IgG antibodies against *Mycobacterium tuberculosis* Conserved hypothetical proteins among TB-positive (n=50) and TB-negative (n=50) patient plasma samples..... 165
Figure 6.6. Comparison of ROC curves showing detection efficiency of the ELISA for *Mycobacterium tuberculosis* antigens among TB-positive (n=50) and TB-negative (n=50) patient plasma samples. 166

Chapter 1: Introduction

1.1 Background

Mycobacterium tuberculosis, the causative agent of tuberculosis (TB), remains a global public health epidemic and has the highest rate of mortality of any infectious disease, with varying populations affected [1]. Currently, the only commercially available vaccine, Bacille Calmette-Guérin (BCG), is highly effective against meningeal and extrapulmonary disseminated TB in children; however, its effectiveness diminishes as individuals reach adulthood [2,3]. It is therefore imperative that, in order to combat these challenges, there is a need for effective drug treatments, novel, effective vaccine development, and early detection diagnostic assays. Although TB is a curable disease, several challenges impede TB control, including patient non-compliance with drug therapy, bacterial drug resistance, inadequate or delayed diagnosis, co-infection with human immunodeficiency virus (HIV), and limited availability of point-of-care testing [4-6]. Furthermore, in 2020, the Coronavirus disease 2019 (COVID-19) pandemic posed significant obstacles to the fight against TB worldwide, as it disrupted surveillance and treatment, thus placing additional strain on public healthcare [7,8]. Early and accurate diagnosis is essential for effective tuberculosis control, as it plays a critical role in preventing increased incidence, particularly within these high-burden countries.

In *M. tuberculosis*, virulence gene clusters encoding specialized cell-surface proteins that interact with the host cellular components, called adhesins, have been shown to influence TB pathogenesis [9,10]. These adhesins exhibit strong virulence properties that are genetically specific to the *M. tuberculosis* complex, and it is evident that the mycobacterial proteome has multiple groups of extracellular matrix-binding proteins mediating host-pathogen interactions [11-13]. The *M. tuberculosis* proteome consists of 25% functionally unexplored proteins, called conserved hypotheticals, which are shown to be preserved across all mycobacterial lineages. These proteins are uniquely expressed in bacterial cells and may pose a viable source for exploitation as potential diagnostic biomarkers and anti-adhesion therapy [14]

This thesis investigated the functional role of uncharacterized conserved hypothetical (CH) proteins, with particular emphasis on their adhesin properties. The recombinant production of these potential adhesins was explored for functional and immunological evaluation. A systematic

bioinformatic approach was employed, followed by phenotypic validation of these novel adhesins. In addition, the immunogenic properties of these adhesins were evaluated in a proof-of-concept serological study, revealing their potential for both serodiagnostic applications and integration into the TB vaccine development pipeline.

1.2 Novelty and innovation

This is a novel study that undertakes the identification and characterization of potential adhesin proteins from an unknown functional group of *M. tuberculosis* CH proteins using bioinformatic analysis. Furthermore, combined molecular methods were used to design and synthesize novel recombinant CH proteins from a selection of potential *M. tuberculosis* adhesins. Validating these novel *M. tuberculosis* adhesins may address current gaps in understanding the functions of CH proteins and provide a foundation for further research into these newly identified adhesin molecules. The detection of anti-adhesin antibodies (Immunoglobulin G) in patients' plasma samples may identify strong serodiagnostic and therapeutic markers.

The research questions that were addressed in this study:

1. Can conserved hypothetical proteins of *Mycobacterium tuberculosis* identified in the study be functionally characterized as adhesins?
2. What are the immunological implications of identifying novel adhesin proteins in *M. tuberculosis* , particularly regarding their potential adaptation for diagnostic testing and their integration into vaccine development pipelines?

1.3 Rationale for the PhD thesis

Existing serological diagnostic assays lack the sensitivity and specificity of antigen biomarkers required for reliable point-of-care use, thereby contributing to delayed diagnosis and continued transmission [7,8]. In addition, the only commercially available vaccine is BCG, which is only effective up to adolescence [2,3]. There is, therefore, an urgent need for more effective vaccine candidates and diagnostic biomarkers. Anti-adhesin-based strategies have shown promising outcomes in protection against bacterial pathogens, as demonstrated by human enterotoxigenic *Escherichia coli* by immunization with purified fimbriae [15], thus providing a platform for targeted adhesin proteins to confer protective immunity. Despite the biological importance of CH proteins, their functional association as adhesins has not been explored entirely [14]. Therefore,

this study addresses the gap by systematically identifying potential *M. tuberculosis* adhesins from a lesser-explored class of conserved hypothetical proteins, which may be a step closer to creating a diagnostic/vaccine tool for TB.

1.4 Aim and objectives

The study aimed to identify and validate potential novel immunogenic adhesin/adhesin-related proteins among the conserved hypothetical group of *Mycobacterium tuberculosis* proteins for the subsequent evaluation in serodiagnostic and vaccine development pipelines.

Primary Phase

1. To identify novel adhesin/adhesin-related proteins from the conserved hypothetical *Mycobacterium tuberculosis* using SPAAN and a robust pipeline of bioinformatic tools.
2. To clone, express, and purify recombinant conserved hypothetical *Mycobacterium tuberculosis* proteins using a pGEX6-P-1 expression system.
3. To investigate the interaction profiles of Adhesin–ECM models by in silico molecular docking simulations.
4. To determine the adhesin properties of the selected purified GST-affinity tag recombinant proteins with host extracellular matrix proteins such as collagen, laminin, and fibronectin.

Secondary Phase

1. To clone, express, and purify recombinant conserved hypothetical *Mycobacterium tuberculosis* proteins using a pET-28a expression system.
2. To validate the biological activity of His-affinity tag recombinant proteins by measuring adhesin capacity to bind fibronectin.
3. Establish optimized ELISA parameters to maximize sensitivity and specificity in serological testing.
4. Identification of anti-adhesin IgG antibodies specific to novel *Mycobacterium tuberculosis* adhesin-related antigens compared among TB-positive and TB-negative patient plasma.

References

1. World Health Organization. Global tuberculosis report 2025 Geneva: World Health Organization: WHO. 2025. Available from: https://iris.who.int/server/api/core/bitstreams/e97dd6f4-b567-4396-8680-717bac6869a9/content.

2. Starshinova A, Kudryavtsev I, Rubinstein A, *et al.* BCG vaccination: historical role, modern applications, and future perspectives in tuberculosis and beyond. Frontiers in Pediatrics. 2025;13(2025). https://doi.org/10.3389/fped.2025.1603732

3. Hatherill M, Cobelens F. Infant BCG vaccination is beneficial, but not sufficient. The Lancet Global Health. 2022;10(9):e1220-e1221. https://doi.org/10.1016/s2214-109x(22)00325-4

4. Bhatt A, Syed ZQ, Singh H. Converging epidemics: A narrative review of tuberculosis (TB) and human immunodeficiency virus (HIV) coinfection. Cureus. 2023;15(10). https://doi.org/10.7759/cureus.47624

5. Farhat M, Cox H, Ghanem M, *et al.* Drug-resistant tuberculosis: a persistent global health concern. Nature Reviews Microbiology. 2024;22(10):617-635. https://doi.org/10.1038/s41579-024-01025-1

6. Heidary M, Shirani M, Moradi M, *et al.* Tuberculosis challenges: Resistance, co-infection, diagnosis, and treatment. European Journal of Microbiology and Immunology. 2022;12(1):1-17. https://doi.org/10.1556/1886.2021.00021

7. Falzon D, Zignol M, Bastard M, Floyd K, Kasaeva T. The impact of the COVID-19 pandemic on the global tuberculosis epidemic. Frontiers in Immunology. 2023;14:1234785. https://doi.org/10.3389/fimmu.2023.1234785

8. McQuaid CF, Vassall A, Cohen T, Fiekert K, White R. The impact of COVID-19 on TB: a review of the data. The International Journal of Tuberculosis and Lung Disease. 2021;25(6):436-446. https://doi.org/10.5588/ijtld.21.0148

9. Vinod V, Vijayrajratnam S, Vasudevan AK, Biswas R. The cell surface adhesins of *Mycobacterium tuberculosis*. Microbiological Research. 2020;232:126392. https://doi.org/10.1016/j.micres.2019.126392

10. Squeglia F, Ruggiero A, De Simone A, Berisio R. A structural overview of mycobacterial adhesins: Key biomarkers for diagnostics and therapeutics. Protein Science. 2018;27(2):369-380. https://doi.org/10.1002/pro.3346

11. Halper J. Basic Components of Connective Tissues and Extracellular Matrix: Fibronectin, Fibrinogen, Laminin, Elastin, Fibrillins, Fibulins, Matrilins, Tenascins and Thrombospondins. In:

Halper J, editor. Progress in Heritable Soft Connective Tissue Diseases. Cham: Springer International Publishing; 2021. 105-126. https://doi.org/10.1007/978-3-030-80614-9_4

12. Ramsugit S, Pillay M. Identification of *Mycobacterium tuberculosis* adherence-mediating components: a review of key methods to confirm adhesin function. Iranian Journal of Basic Medical Science. 2016;19(6):579-584. https://dx.doi.org/10.22038/ijbms.2016.7124

13. Berdiaki A, Neagu M, Tzanakakis P, Spyridaki I, Pérez S, Nikitovic D. Extracellular matrix components and mechanosensing pathways in health and disease. Biomolecules. 2024;14(9):1186. https://doi.org/10.3390/biom14091186

14. Yang H, Ma Y, Lei X, *et al.* Stopping Tuberculosis at the Gate: The role of *M. tuberculosis* adhesins in infection and intervention. Vaccines. 2025;13(7):676. https://doi.org/10.3390/vaccines13070676

15. Klemm P, Schembri MA. Fimbrial surface display systems in bacteria: from vaccines to random libraries. Microbiology. 2000;146(12):3025-3032. https://doi.org/10.1099/00221287-146-12-3025

Chapter 2: Literature Review

2.1 Introduction

Tuberculosis (TB) is one of the leading causes of mortality among communicable diseases globally [1]. Although TB is curable by adhering to a strict 4–6-month short-term treatment regimen, non-compliance with completing treatment results in failure to eradicate the bacterial infection [2-4]. These factors increase the risk of drug-resistant (DR-) TB among individuals, therefore leading to a prolonged treatment regimen and resulting in increased transmission rates [5]. Tuberculosis is inherently transmitted by aerosolized nucleic droplets containing bacilli through the cough of an infected person, in proximity to healthy individuals, leading to pulmonary TB. This form of TB is one of the most frequently diagnosed modes of infection and can spread throughout the body, manifesting as meningeal and extrapulmonary TB, impacting the blood and vital organs [6]. In certain instances, latent TB infection may exist in an individual without symptoms and may reactivate when conditions permit, such as during a compromised immune response [7].

Since Robert Koch first isolated *Mycobacterium* bacilli in 1882, coordinated global efforts in treatment, diagnostics, and prevention have increasingly driven research toward innovative strategies to reduce and ultimately eliminate tuberculosis infections by 2030 [1,8]. Current molecular-based technologies have greatly enhanced our understanding of the genomic dissemination of diverse *Mycobacterium tuberculosis* strains, enabling the characterization of transmission patterns, the identification of novel therapeutic drug targets, and the discovery of immunogenic proteins with potential as biomarkers for diagnostic and vaccine development [9-11]. Nevertheless, significant challenges remain in curbing TB transmission and reducing its global incidence.

2.2 The global concern of tuberculosis infection

Tuberculosis continues to represent one of the most formidable infectious disease threats to global public health, with high transmission rates and an onset of latent infection. Its persistent global burden affects socioeconomic and healthcare system factors, making it a priority for worldwide control efforts and targeted research.

2.2.1 The worldwide epidemic

The World Health Organization (WHO) has classified TB as a worldwide crisis, with approximately 10.7 million active TB cases and 1.23 million deaths reported in 2024 [1]. It has shown a gradual decline from the 1.25 million deaths reported in 2023 [12]. The global mortality rate of communicable diseases increased to 28% in 2021 from 18.2% before the outbreak of the COVID-19 pandemic in 2019 [12]. This surge has increased the burden on global efforts to reduce TB incidence and death, further placing financial strain on health care systems and reducing monitoring and treatment efforts during the COVID-19 pandemic [13]. The thirty high-burden countries accounted for nearly 90 percent of all estimated incidence cases globally, with the highest burden seen in South-East Asia, followed by the Western Pacific and Africa (Figure 2.1). Furthermore, efforts to contain TB infection remain challenging, as *M. tuberculosis* latency affects approximately 25% of the world population [5]. In the absence of clinical symptoms, resurgence can be initiated by a range of immune-modulatory effects. Therefore, significant challenges for TB preventative treatment arise, particularly among people living with human immunodeficiency virus (HIV), who have a significantly higher risk of reactivation [7,14].

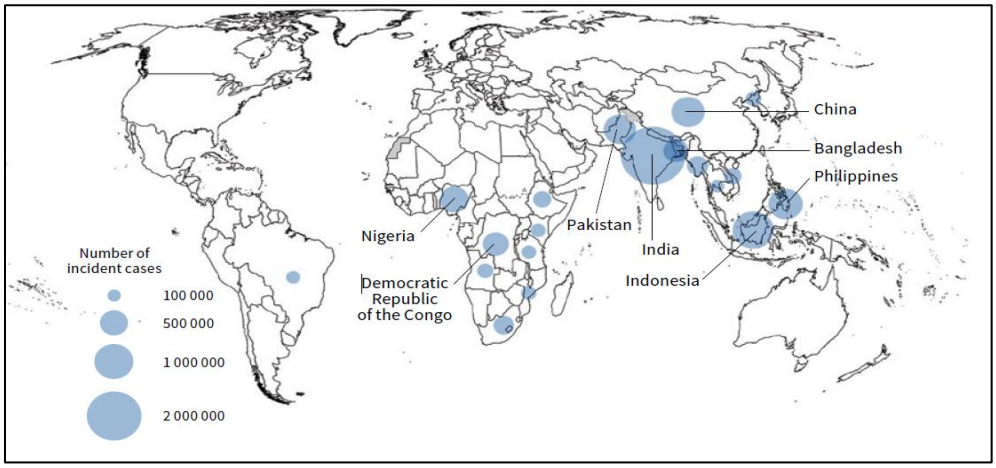


Figure 2.1. The estimated tuberculosis incidence in countries that reported over 100,000 cases, as reported by the WHO in 2024. Eight of the high-burden countries have accounted for two-thirds of the total incidence, with India remaining the highest-incidence country, accounting for a quarter of global cases [1].

2.2.2 The burden of tuberculosis infection in South Africa

According to the 2025 global TB report, South Africa (SA) remains among the 30 high TB burden countries (Figure 2.1). This is attributable to a high HIV associated TB and multi-drug resistant (MDR-) and rifampicin resistant (RR-) TB [1]. A national study by Statistics SA in 2016 reported that TB remains the leading cause of death among communicable diseases, with a steady decline from 8.3% to 6.5% in 2014 and 2016, respectively [15]. An updated report in 2022 showed TB to be second to HIV related deaths and in groups aged 18-24, TB remained the number one cause of death [16]. Although SA has a high burden of TB, it has now been classed as highly endemic since 2024, from severely endemic in 2015 [1]. The governmental efforts to curb the spread of TB by introducing the End TB Campaign, aimed at testing five million people in 2025-2026, have improved monitoring and treatment coverage by 79%, reduced the incidence of TB in 2023 to 427 per 100,000, which was a 57% decline since 2015 [17]. The successful treatment of DR-TB has improved to 62 percent in 2023, and the decrease in deaths among TB cases has reduced to 16 percent. Even though the trajectory is in the right direction, there remain challenges such as rapid diagnosis for early detection, high drug resistance cases, and HIV-related co-infection [5,18,19].

2.3 Current diagnosis and treatment

Early diagnosis and effective treatment remain critical pillars in TB control; however both pathways are burdened with persistent challenges. These include limitations in detection in remote settings, resistance profiles, and socio-economic factors that affect patient adherence leading to drug resistant (DR-) TB

2..3.1 Diagnosis of tuberculosis

The diagnosis of TB remains one of the most challenging aspects of TB control. In regions with high rates of HIV co-infection, treatment failure may arise from factors such as poor patient compliance, drug–drug interactions, and immune suppression. These challenges compromise treatment efficacy and monitoring and contribute to the emergence and spread of drug-resistant TB. In a 2025 WHO report, there was an attempt to reduce the time in which patients are correctly diagnosed by the implementation of approved rapid diagnostics benchmarks (Figure 2.2), which improves screening guided towards equitable TB treatment [20].

Point-of-care (POC) strategies represent the primary diagnostic approach for TB, typically initiated when patients present to healthcare facilities with symptoms [21]. Clinical diagnosis often relies on symptom screening, such as persistent cough, night sweats, fever, and weight loss, conventionally confirmed by chest radiography and sputum microscopy for pulmonary TB [21]. In contrast, extrapulmonary TB presents additional challenges: while symptoms may overlap with pulmonary disease, they are often organ-specific, and obtaining appropriate clinical samples is difficult. Therefore, it may complicate diagnosis, frequently requiring invasive procedures across multiple sites, which can cause significant patient discomfort and delay confirmation of the disease.

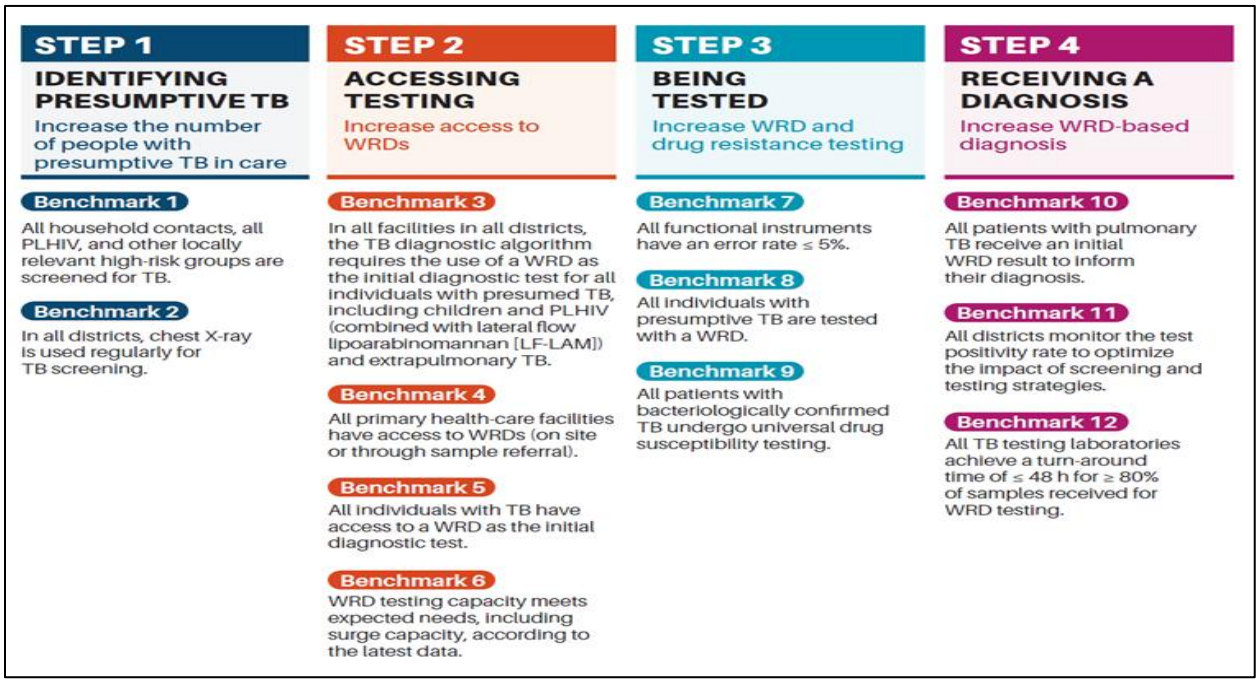


Figure 2.2 Universal access to the World Health Organization has recommended rapid diagnostics (WRD) to increase detection and reduce diagnostic time in 2023. A 12-benchmark-based programme that outlines the strategic framework to improve TB diagnosis outcomes by improving screening, monitoring, and optimal detection time [20].

Conventional laboratory diagnosis of TB must be rapid and efficient to ensure timely treatment. Microscopy remains the gold standard and most cost-effective method for detecting acid-fast bacilli in sputum, using Ziehl–Neelsen techniques [22,23]. While these methods provide quick results, they require skilled personnel and are limited by low sensitivity in paucibacillary TB samples [23]. To improve accessibility in resource-limited settings, light-emitting diode

fluorescent microscopy using auramine O staining has been introduced in middle- and low-income countries (LMICs) as a cost-effective alternative to conventional fluorescent microscopy [22,24]. In addition, liquid culture-based methods such as the BD BACTEC MGIT (Becton Dickinson Biosciences, Sparks, Md.) offer higher sensitivity and detection rates compared to smear microscopy but are costly and require up to four weeks to yield positive results [25]. Conventional phenotypic methods of drug susceptibility testing are time-consuming and require trained personnel to interpret outcomes [26,27]

Rapid molecular diagnostic techniques are an integral part of prompt, effective TB diagnosis [28]. The cost and management of novel diagnostic techniques are essential to their feasibility from a POC perspective. Indirect detection methods for TB diagnosis include *M. tuberculosis* antigen-based skin tests, tuberculin skin tests, and interferon-γ release assays (IGRAs). However, these tests provide an initial screening for TB detection, and sensitivity varies across populations and cannot discriminate between *M. tuberculosis* infection and TB as the clinical disease [28]. Currently, low-complexity manual nucleic acid amplification tests (NAATs), such as loop-mediated isothermal amplification and lateral flow urine lipoarabinomannan (LF-LAM) assay such as Determine™ TB LAM Ag assay (Abbott, MA, USA), are used to diagnose TB without resistance. These are cost-effective laboratory detection assays, and LF-LAM has shown improved sensitivity in detecting TB in patients co-infected with HIV [24,28]. Since the WHO's approval in 2010, the low-complexity automated Xpert MTB/RIF (Cepheid, CA, USA) has been recommended for high-burden countries to detect drug resistance and HIV-associated pulmonary TB, with results available after a 90-minute cycle [28,29]. The application of Xpert MTB/RIF in TB diagnostics allows smear-negative samples to be tested and to detect rifampicin-resistant strains of TB Xpert MTB/RIF (Cepheid, CA, USA) [30]. In the case of isoniazid- and rifampicin-resistant TB, the WHO has recommended moderate-complexity automated NAATs [28]. In addition to Xpert MTB/XDR (Cepheid, CA, USA), line probe assays such as Genotype MTBDRplusV1/V2/*s*l (Hain Lifescience, Nehren, Germany) or GenoScholar NTM+MDRTB II/PZA-TB II (Nipro, Osaka, Japan), and targeted next-generation sequencing such as deeplex Myc-TB (GenoScreen, Lille, France), AmPore-TB (Oxford Nanopore technologies, Oxford, UK) and TBseq (Sengting, Zhejiang, China), have been approved for follow-on tests for detecting TB drug resistance. These molecular assays outperform the phenotypic approach of DST, for high throughput and quicker results.

2.3.2 Tuberculosis treatment strategies

Tuberculosis is often described as a social disease due to overcrowding, vulnerable populations that are poverty-stricken, and limited access to healthcare [1]. The standard first-line therapy for active TB, consisting of rifampicin, isoniazid, pyrazinamide, and ethambutol (RIPE) requires adherence for two months, followed by rifampicin and isoniazid only for four months [23,31]. However, many patients default on treatment before completing the whole course, leading to severe consequences, ranging from relapse, prolonged infection, and the emergence of drug resistance [32]. The onset of drug resistance resulting from poor susceptibility to rifampicin or both rifampicin and isoniazid results in RR-TB or MDR-TB, respectively [1]. Therefore, second-line therapy is required, previously based on a cocktail of injectables, which required up to 24 months of treatment [33]. This had a higher adverse drug reaction rate compared to the current Bedaquilline, Pretomanid, and linezolid (BPaL) therapy has a reduced nine-month course [1,34].

To reduce the progression of LTBI to active TB, the WHO has approved the use of TB preventive treatment (TPT), particularly among people living with HIV (PLHIV) [1]. The treatment of LTBI is generally isoniazid for 6-months or 9-months by self-administering, which leads to poor adherence outcomes. However, TPT with rifapentine can reduce treatment to a 3-month short course [35]. The success of TPT among PLHIV infections reduce the risk of TB by 76% [7]. However, in addition to non-compliance, other challenges include stigma relating to fears of activation of TB and adverse drug reactions when coupled with antiretroviral treatment [7]. Therefore, ensuring treatment completion is as much a social and behavioral challenge as it is clinical. Education on the importance of adherence, socioeconomic barriers, and patient support systems is an important measure to reduce default rates [21]. In SA, the reduction in TB cases was achieved with high alert monitoring with the directly observed treatment-short course programme, now moving forward with the National TB recovery programme 4.0 to test over five million people for TB and end TB by 2035 [17,36]. Without these measures, the curability of TB risks being overshadowed by preventable treatment failures and the increasing threat of multidrug-resistant tuberculosis (MDR-TB) . As a result of drug resistance, novel drugs are being developed, currently in various clinical phases, with only three anti-TB drugs approved as per the 2025 global new TB drug pipeline (Figure 2.3).

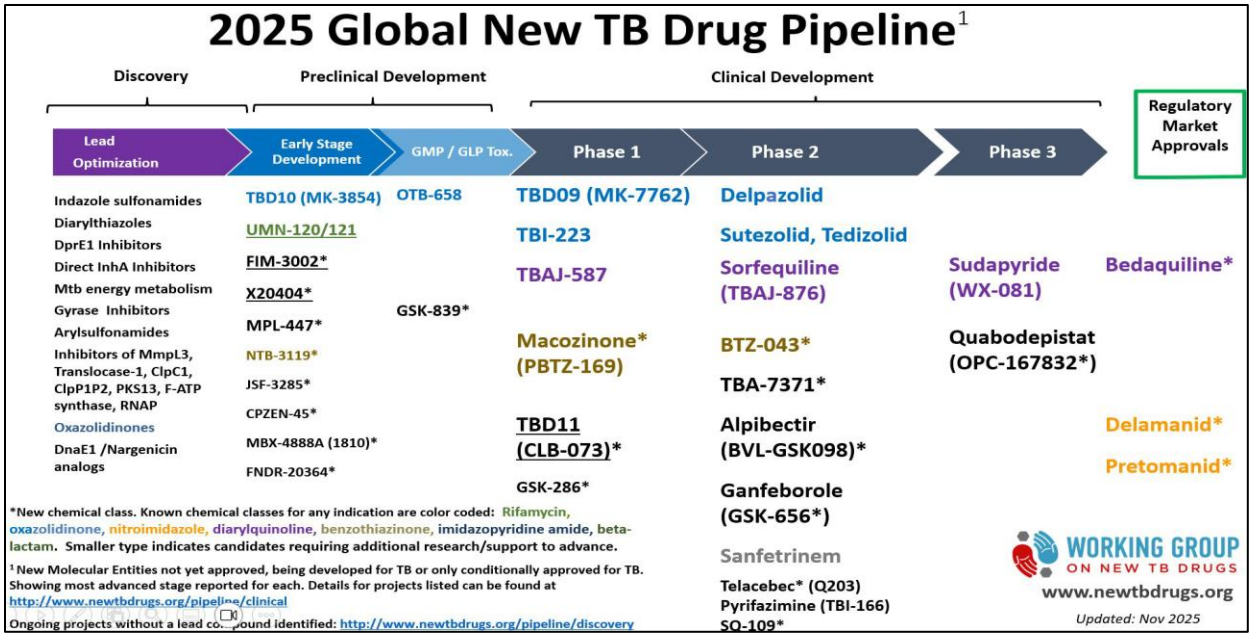


Figure 2.3. Anti-TB drug development pipeline for the treatment of tuberculosis. The drug pipeline shows anti-TB compounds in various stages of development, from discovery and preclinical testing to clinical trials and regulatory approval. While only three drugs (bedaquiline, delamanid, and pretomanid) have received regulatory approval to date, multiple candidates are advancing through Phase 2 and Phase 3 trials, showing promising progress in expanding treatment options for drug-resistant TB [37].

2.4. Pathogenesis of tuberculosis

The pathogenesis of TB is a systematic process that is initiated by transmission and culminates in the spectrum of clinical outcomes, ranging from latent infection to active disease. This section focuses on the etiological basis of infection, the structural composition of the *M. tuberculosis* cell envelope and the host immunomodulatory strategies that enhance the virulence of this mycobacterial pathogen.

2.4.1 Etiology of tuberculosis

The survival of *M. tuberculosis* begins when aerosolised bacilli enter the lungs, initiating the primary infection. *M. tuberculosis* evades phagolysosome fusion by binding directly to cellular receptors, invading alveolar macrophages, and replication within host cells [38]. This subsequently, cellular apoptosis is inhibited, and macrophage necrosis is induced [39,40]. The host immune response in the lungs is primarily mediated by tumor necrosis factor (TNF) cytokine activity, which promotes the formation of granulomas, compacted aggregates of immune cells that

may fuse into multinucleated giant cells and epithelioid histiocytes [41]. Poorly formed granulomas create a microenvironment that allows *M. tuberculosis* to replicate, persist, or remain dormant until conditions favor reactivation [38]. Although infection primarily occurs in the lungs, referred to as pulmonary TB, infection can also spread to other organs throughout the body, known as extrapulmonary infection [23]. As shown in Figure 2.4, upon active TB infection progression, clinical symptoms typically manifest in immunocompromised individuals, or the bacilli may remain dormant, leading to latent TB infection [42].

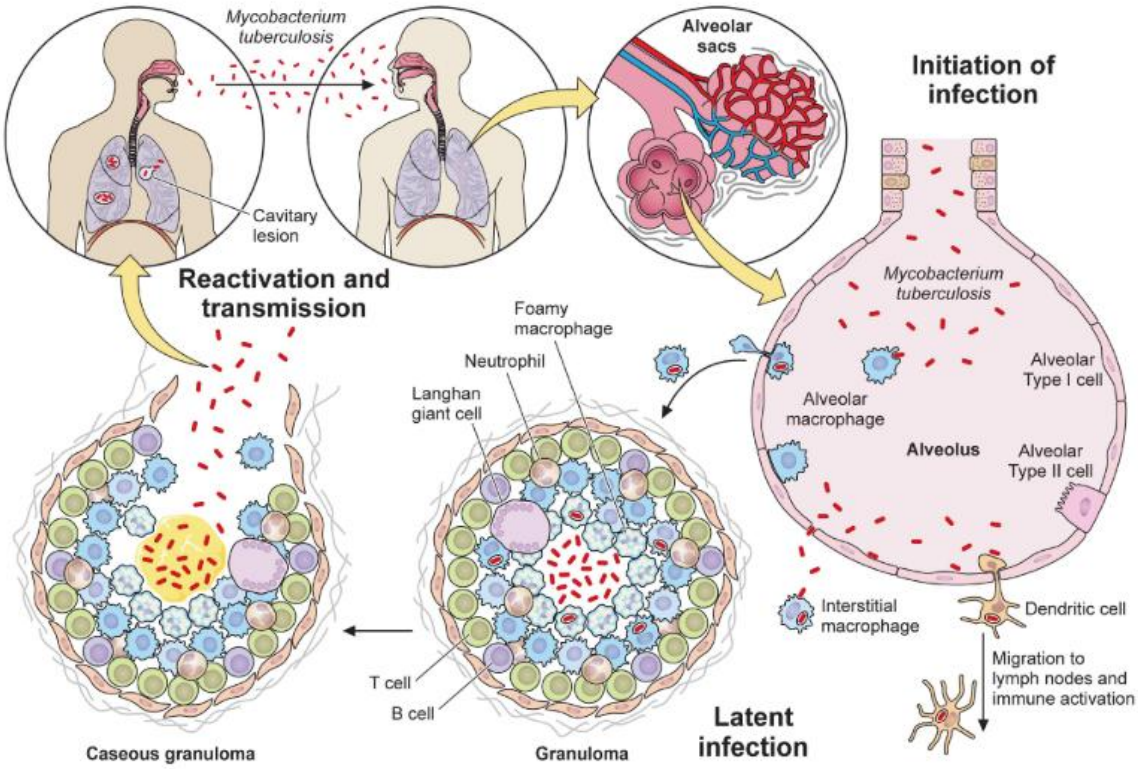


Figure 2.4. The pathway of *Mycobacterium tuberculosis* infection. The spread of tuberculosis occurs through direct contact with an infected individual. This leads to an initiated infection within the alveolus of the lung. The host immune response contains infection within granulomas. This either remains dormant as a latent infection or disseminates bacilli, leading to reactivation and transmission of disease [42]

2.4.2 Structural characteristics of *Mycobacterium tuberculosis*

Tuberculosis is associated with the group of closely related species within the *Mycobacterium tuberculosis* complex, namely, *M. bovis, M. canettii, M. caprae, M. mungi, M. orygis, M. pinnipedii, M. suricattae,* and *M. tuberculosis* [43,44]. The defining characteristic of the *M.*

tuberculosis structure is a dual outer membrane cell envelope, comprising free lipids and mycolic acids, which is covalently attached to the polysaccharide cell wall [45]. Daffé and Marrakchi [46] illustrated that the transporter proteins of mycobacterial membrane protein large (MmpL) are associated with high lipid content (Figure 2.5). Furthermore, the MmpL protein has an important physiological function that contributes to cell envelope integrity and membrane permeability [46]. The cell envelope promotes bacterial virulence properties, attributed to the low-permeability cell wall, which confers natural resistance to an array of antibiotics [47].

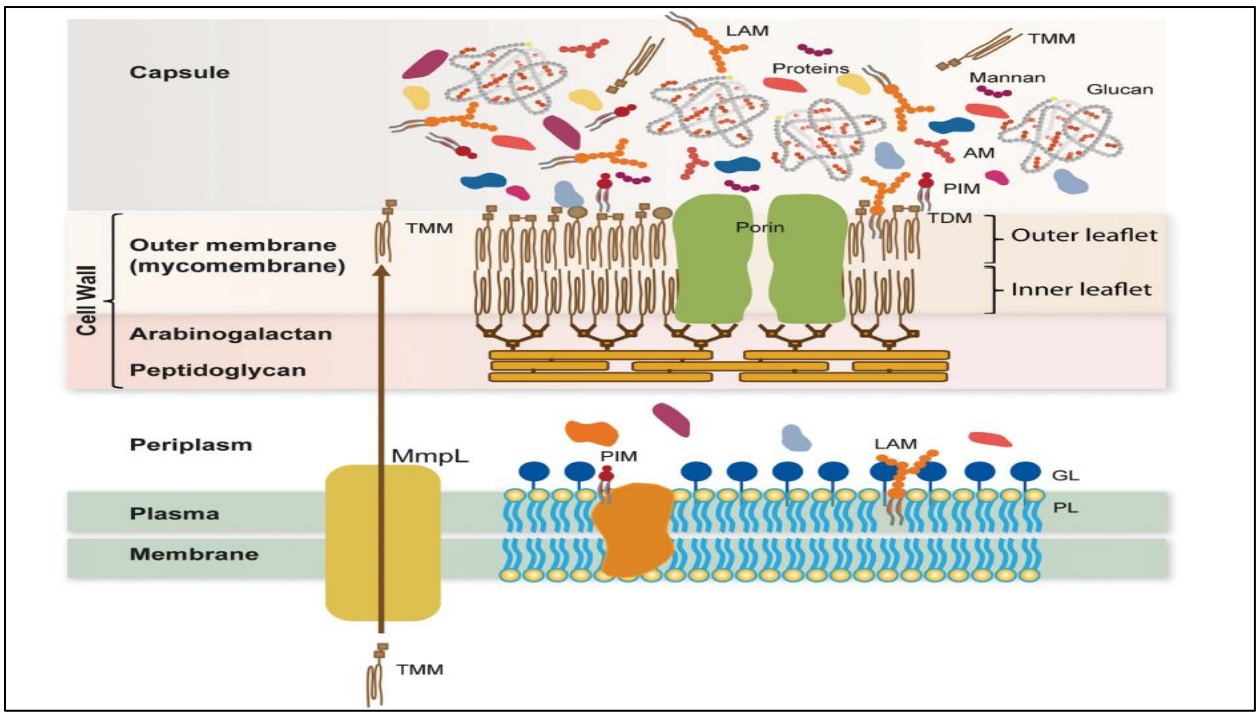


Figure 2.5. An illustrative composition of the mycobacterial cell envelope. Structurally, the envelope consists of multiple layers forming the plasma membrane, a cell wall made of peptidoglycan, arabinogalactan, and mycolic acids, surface lipids of the outer membrane, and a capsule with a matrix of α-glucan that is composed of proteins, α- glucan, mannan, arabinomannan, and lipids which are transported by the MmpL [46].

In addition, the cell envelope components are relatively hydrophobic by masking pathogen-associated molecular patterns (PAMPs) and suppressing macrophage antimicrobial activity [48]. The multilayered mycobacterial cell envelope comprises a plasma membrane, a complex cell wall separated by a periplasmic space, with an additional capsule layer that enhances pathogenicity [46]. Mycobacterial components may provide potential therapeutic target sites for antimicrobial drugs and surface associated proteins for vaccine development [49-51].

2.4.3 Modulation of the host response

The pathogenesis of *M. tuberculosis* represents a remarkable evolutionary adaptation, enabling the bacterium to manipulate host immunity while simultaneously evading host defenses to sustain the initial infection. This section examines the complex host immune interaction during TB infection.

2.4.3.1 Macrophage evasion

During the early stages of *M. tuberculosis* infection in the lung, alveolar macrophages (AM) constitute the first line of defense [52]. Subsequently, macrophage phagocytosis is then initiated by opsonization, whereby bacilli are recognized following coating with immune components, including surfactant protein A, complement factors, soluble mannose, and immunoglobulin isotypes [53]. In addition, bacterial PAMPs, identified by pattern recognition receptors on AMs, further promote phagocytic uptake [40]. Internalization of bacilli by AM leads to the formation of the phagosome-lysosome, which triggers the production of reactive oxygen and nitrogen species, lowering pH, and nutrient depletion [40,52]. However, *M. tuberculosis* has developed multiple virulence mechanisms, including the type VII secretory system ESX-1, lipid-associated phthiocerol dimycocerosates, and secreted proteins such as nucleoside diphosphate kinase and the phosphatase SapM. These aid in evading phagosome maturation, thereby enhancing pathogen survival within the host [40,42,54]. In addition, macrophage autophagy is inhibited by a 42-kDa secretory enhanced intracellular survival protein, which induces interleukin (IL) -2 and histone acetylation, promotes protein kinase G binding to the pleckstrin homology domain, and prevents phagosome-lysosome fusion [40].

2.4.3.2 Cytokine signaling

Through cytokine signaling, macrophages differentiate into diverse/heterogenetic phenotypes within granuloma formations that engulf the *M. tuberculosis* infection [52]. Heterogenetic M1 and M2 macrophages are shown to initiate cytokine and chemokine signaling pathways, initiating T-Helper (TH)-1 proinflammatory responses (IL-1, IL-6, IL-12, IL-23, and C-X-Chemokine Ligand 9) and subsequently TH2 anti-inflammatory responses (IL-4, IL-10, and IL-13) [40,52]. The signaling of IL-4 induces M2 macrophages to increase acidic pH levels during phagosome maturation, whereas IL-1 suppresses interferon-β [53]. In contrast, mycobacterial proteins, such

as early secreted antigenic target 6kDa (Esat-6), expressed as secreted proteins within the macrophages, are known to trigger Interferon-β expression, which facilitates *M. tuberculosis* pathogenesis by promoting anti-inflammatory macrophage polarization and IL-10 production, and phagosome rupture for the translocation to the cytosol [42,52]. The sustained production of tumor necrosis factor (TNF) -α and IL-1 remain crucial for mycobacterial control as it promotes protective immune responses [55]. Host immune mechanisms can restrict *M. tuberculosis* within granulomas, establishing latent infection; however, dysregulated cytokine signaling and immune exhaustion may compromise containment and initiate reactivation.

2.4.3.3 T-lymphocyte interference

Activation of T-lymphocytes is integral to innate immune defenses against *M. tuberculosis* and constitutes a central mechanism in controlling TB infection. An effective T-lymphocyte response, including mucosal-associated invariant T-cells, must be tightly regulated to ensure pathogen clearance while preventing excessive inflammatory damage within pulmonary lesions, as uncontrolled inflammation can accelerate disease progression [56,57]. T-lymphocyte exhaustion was reported as a path used by *M. tuberculosis* to evade the innate response by persistent exposure to antigens, resulting in reactivation of the T-cell response [58]. In addition, Toll-like receptors (TLRs) and nucleotide and oligomerization domain-like receptors detect *M. tuberculosis* components, initiating immune activation. However, it has been shown that ESAT-6 can inhibit TLR signaling, thereby curbing innate immune activity [59]. *M. tuberculosis* induces regulatory T-cells that suppress effector T-cell activity, reducing immune pressure and promoting persistence [60]. By modulating these pathways, *M. tuberculosis* interferes with antigen presentation and weakens T-cell recognition, ultimately compromising host defense.

2.4.3.4 Granuloma formation

A cell-mediated immune response eventually leads to granuloma formation, which can contain, neutralize, or progress to active TB disease [56,61]. The transition from innate to adaptive immunity involves T-lymphocytes and production of IFN-γ, TNF-α, and IL-12, although other immune pathways are also critical [55,62]. Granulomas are characterized by macrophage aggregates, neutrophils, dendritic cells, T-cells, B-cells, and natural killer cells that form in

response to persistent *M. tuberculosis* infection [63]. Granulomas exhibit varying phenotypes, including fibrotic, caseous, and non-necrotizing lesions [61,64,65]. However, *M. tuberculosis* has evolved strategies within granulomas that influence bacterial replication, metabolic activity, and antibiotic susceptibility, thereby contributing to the distinction between latent and active infection [62,63,65]. Due to the presence of the many layers of the granuloma containing the *M. tuberculosis* infection, there is limited exposure of antigen presentation and recognition for the T-cell response at the site of infection [66]. Inflammation-induced immune response caused by neutrophils and T-cells may in turn exacerbate granuloma liquification in which *M. tuberculosis* may disseminate, causing active TB [52]. In contrast, successful containment of infection during which inflammation subsides leads to non- replicating latent infection activation of the *M. tuberculosis* dormancy survival regulon within the hypoxic dormant granuloma [42,62].

2.4.3.5 Cell death-specific mechanisms

Mycobacterium tuberculosis employs various strategies to inhibit host pathways that lead to cell death. This may be beneficial as the initial infection creates a replicative niche by preventing nutrient-rich cell death. Ubiquitination inhibition by mycobacterial enzymes, including protein kinase G, ubiquitin ligase, PPE38, and Rv0222, have been identified in suppressing host immunological pathways for such benefits [40]. Effector protein, *M. tuberculosis* lipoarabinomannan is known to inhibit infection-induced apoptosis by promoting the release of TNF-α and downregulating Bcl2, in addition to upregulation of host NuoG, Mcl1, Scavenger receptor-A1, Bcl-w, SecA2, and PknE anti-apoptosis-related proteins [67,68]. Nisa *et al.* [69] reported various *M. tuberculosis* associated proteins remarkably identified as cell wall or secreted proteins, which are strong inhibitors/inducers of programmed cell death.

2.5 Adhesin molecules in *Mycobacterium tuberculosis*

Mycobacterial proteins that interact with the host cellular components enhancing the virulence of the pathogen by promoting colonization and evasion of the host immune response. Adherence is facilitated by adhesin molecules that bind to an array of host extracellular matrix components and are discussed in this section.

2.5.1 Role of adhesins

It is widely understood that commensal and pathogenic bacteria have adapted to the host environment by expressing macromolecules that identify and bind to host cellular targets. These macromolecules include an array of lipids, proteins, pili, fimbriae, and glycopolymers that are associated with the bacterial cell surface [70]. While adherence facilitates colonization to the host surfaces, their presence presents a challenge for pathogenic bacteria. The host may recognize attachment of these macromolecules and elicit an immune response to the bacterial interactions to clear the infection [71]. Subsequently, virulence factors are expressed to evade the host immune response to progress bacterial colonization [40].

Adhesins can help mycobacteria adhere to cells and thus gain entry to cells such as host macrophages [38]. The strategic entry of MTB into macrophages entails adherence to host cellular receptors via adhesion-specific molecules [72,73]. These receptors, known as cell adhesion molecules (CAMs) or adhesins, are produced on the bacterial cell surface and further influence pathogenicity [70,74]. For example, within the *M. tuberculosis* complex, adhesins have been identified to promote virulence by inhibiting phagosome-lysosome fusion in macrophages, such as mycobacterial PE-PGRS30/62, protein kinase G, *M. tuberculosis* low molecular weight tyrosine phosphate A, lipoamide dehydrogenase C, and the antigen-85 complex (composed of three proteins Ag85A, B and C) , which are associated with suppressing the host immune response [42]. Mycobacterial extracellular polymeric substances (EPS), which include adhesins, have been associated with biofilm formation, thereby increasing drug tolerance and promoting *M. tuberculosis* survival in the host by resisting environmental stress [75,76].

In addition to mediating direct host cell entry, MTB adhesins also facilitate binding to ECM components to colonize and disseminate within the host microenvironment and evade host immune cellular receptors which serve as strategic advantages for the pathogen. The ECM comprises an array of host cellular proteins including fibronectin, collagen, elastin, laminin, fibrillin, and vitronectin that are located throughout the eukaryotic microenvironment [72,78]. Adherence to ECMs allows MTB to establish infection within lung parenchyma by anchoring to fibronectin and laminin facilitating persistence and dissemination within the host [82, 86]. Subsequently, ECM binding strategically evades host immune detection, allowing the pathogen to establish a niche before triggering the inflammatory response [71]. Indirect binding pathways such as fibronectin

binding may serve as a bridging mechanism to bind to macrophage integrins modulating an active state [81, 84]. Furthermore, collagen I has been shown to promote granuloma formation during initial infection [93], which may facilitate bacterial retention within granulomas thereby contributing to establishment of latent infection.

Adhesins play a pivotal role in enhancing the virulence of pathogenic bacteria by mediating host–pathogen interactions [70, 71]. In addition to contributing to pathogenesis, adhesins represent promising therapeutic and diagnostic targets, as they constitute the primary point of contact with the host and are capable of eliciting immune activation [72].

2.5.2 Human extracellular matrix components

The complexity of eukaryotic host cellular composition increases their susceptibility to pathogenic infection by providing an array of binding sites [70]. Rich in nutrients and providing a stable environment for replication and transmission, the human body constitutes an ideal niche for mycobacterial persistence and homeostasis. Extracellular matrix components on the surface of host cells function as a support structure of eukaryotic cells but also provide a site of attachment or adherence for mycobacterial adhesins [74]. The following ECMs constitute binding sites for *M. tuberculosis* related adhesins as well as for other pathogenic bacteria. The basement membrane (Figure 2.6) forms a selective barrier of epithelial or endothelial cells from lower connective tissue and comprises a variety of extracellular material that provides integrity to eukaryotic cells [77].

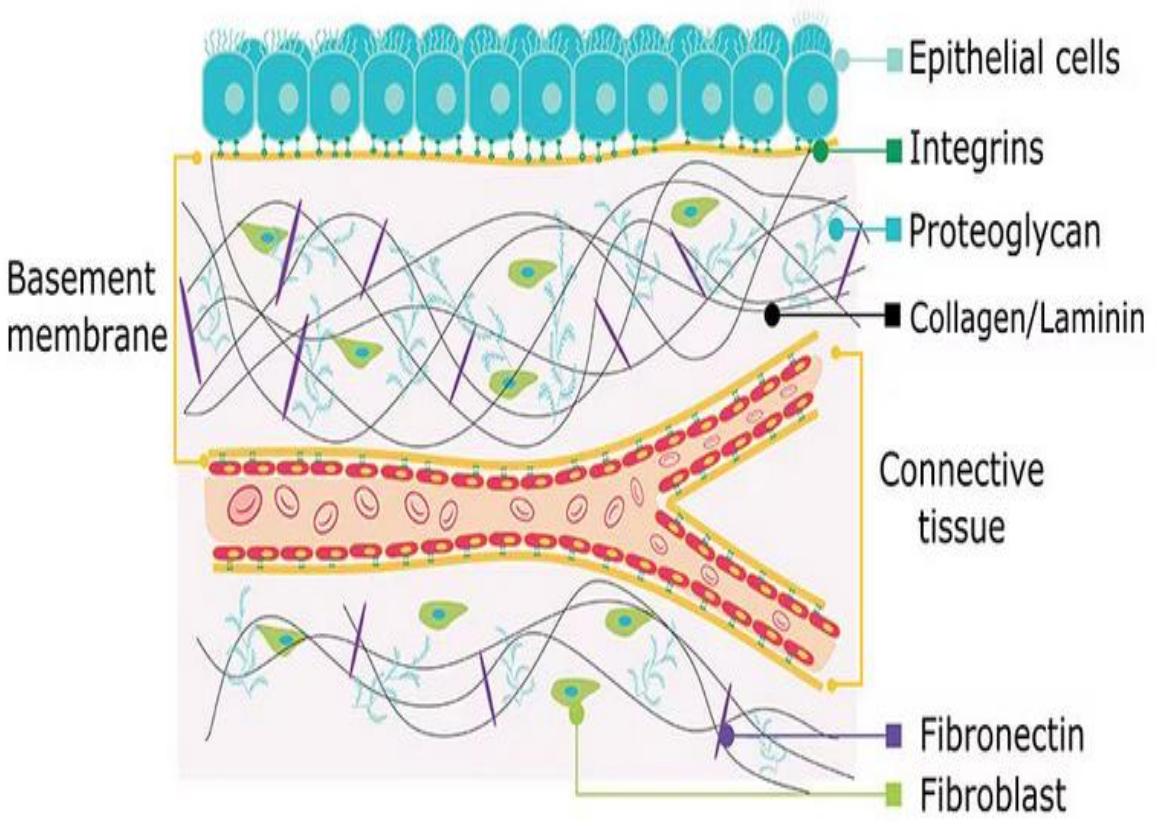


Figure 2.6. Eukaryotic basement membrane comprises an array of extracellular matrix components (ECM). The range of ECM components provides structural support and communication with a variety of biological processes including cell adhesion, migration and proliferation [77].

2.5.2.1 Fibronectin

Fibronectin, type I-III, exists as a multidomain glycoprotein (240-270 kDa) consisting of a dimer polypeptide linked with a disulfide bond, forming within ECMs as well as plasma [78]. Fibronectin was discovered to interact with many cellular components and other macromolecules [79]. As a key adhesion binding protein within the ECM, fibronectin participates in a wide range of biological activity including wound healing, cellular proliferation and differentiation, regulating phagocytosis and cellular migration [79,80]. Fibronectin binds to a host of eukaryotic cells, including integrin, collagen, proteoglycans and other adhesion-related complexes [78,81]. Even

though fibronectin has a diverse physiological role in the host, pathogens exploit the structural function for attachment, dissemination, and persistence [78]. Described as fibronectin attachment proteins or fibronectin binding proteins (FnBPs), mycobacterial adhesins have been identified to utilize fibronectin for invasion [79]. These adhesins include Glutamine Synthetase A1 (Rv2220), Wag22 (Rv1759c), PE-PGRS33 (Rv1818c), PE-PGRS (Rv0109), Alanine and proline-rich antigenic (Rv1860), Malate synthase G (Rv1837c), and Antigen 85 complex (comprising Rv3804c, Rv1886c, Rv0129c) [82-84]. Moreover, the role of mycobacterial FnBPs modulates antigen presentation among major histocompatibility complex II and T-lymphocytes and promotes inhibitory effects on IFN-γ in infected host cells [74].

2.5.2.2 Laminin

Laminin has been identified to form part of the heterotrimeric glycoprotein complex, ranging from 500-900 kDa in size [78]. The family contains more than sixteen members comprising αβγ chains with laminin-521/511 found in the adult connective tissue and laminin-411/421 in basement membrane of endothelial cells [85]. Similar to fibronectin, laminin modulates cell differentiation, migration, structural stability, and, in addition, inhibition of apoptosis [78]. *M. tuberculosis* associated malate synthase (Rv1837c) was demonstrated to bind to laminin of A549 lung epithelial cells [86]. Similarly, ESAT-6 (Rv3875, Rv3874), which induces cytolysis in pneumocytes have been identified to bind to laminin of A549 lung epithelial cells [86]. In addition, *M. tuberculosis* curli pili have been shown to bind to laminin and contribute to biofilm formation [87].

2.5.2.3 Collagen

Collagens represent the most abundant structural proteins among ECM components, comprising 25% of proteins, forming the principal structural framework of eukaryotic connective tissues [88,89]. Fibril-forming collagens, particularly types I and III, provide tensile strength and mechanical stability, while their dynamic regulation allows tissues to adapt to mechanical stress and injury [78]. While non-fibrillar collagen forms less than 10% of collagen, it is found in vital organs within the human body [88]. Beyond their structural role, collagens actively participate in signaling processes, interacting with proteoglycans, glycoproteins, and growth factors to maintain tissue homeostasis and initiate tissue repair responses [78,80]. Dysregulation of collagen synthesis or remodeling contributes to pathological states such as fibrosis and connective tissue anomalies,

highlighting its dual function as both a structural scaffold and a mediator of disease-associated tissue remodeling [90,91]. Squeglia, Ruggiero and Berisio [92] had emphasized that collagen binding and degradation play an initial role in *M. tuberculosis* pathogenesis, as collagen I is a major constituent of lung parenchyma and promotes the formation of granulomas during initial infection [93], therefore, these proteins may very well contribute intrinsically to pathogenesis [94]. Strong binding affinity of Rv2599 (membrane-bound protein) to collagen observed, indicating that *M. tuberculosis* may acquire adhesin function to bind to collagen during various stages of infection (95).

2.5.2.4 Additional extracellular matrix components

In addition to the most abundant composition of fibronectin, laminin, and collagen, the ECM comprises a multitude of macromolecules. Glycosaminoglycans and proteoglycans, forming the cellular glycocalyx, in addition to structural support, are responsible for mechano-sensing within the microenvironment [80]. Subsequently, mycobacterial glyceraldehyde-3-phosphate dehydrogenase binds to fibrinogen, which promotes iron uptake from host cells [95]. Mycobacterial antigen 85 complex was reported to bind to elastin, an insoluble polymer of tropoelastin that accounts for the elasticity of the ECM tissue (97). In identifying ECM components, Halper [78] outlined the association of fibrillin, latent Transforming growth factor (TGF)-β binding proteins, fibulins, matrilins, tenascins, and thrombospondins as products found within the ECM.

2.5.3 Currently known *Mycobacterium tuberculosis* associated adhesins.

Mycobacterium tuberculosis is a highly persistent pathogen that has specialized mechanisms to evade the host immune response. Cell adhesion molecules, adhesins located on the surface of the *M. tuberculosis* complex have enhanced the pathogenesis for survival by assisting with biofilm formation and facilitating attachment to ECM components and host cellular receptors [72,73]. These cellular receptors include toll-like receptors, heparan sulfate, macrophage receptors, scavenger receptor A, cluster of differentiation (1b, 14,36, and 43), surfactant protein-A, mannose receptors, Fc complement receptor-3, dendritic cell-specific intracellular adhesin molecule-3-grabbing non-integrin, and c-type lectins, among other host receptors [70,72]. The cell envelope of *M. tuberculosis* is highly hydrophobic, consisting of glycolipids, peptidoglycan, arabinoglycan,

and glycoproteins contributing to its virulence within the host environment by recognizing and attaching to host cellular components [73].

Mycobacterial adhesins have been of interest due to their interaction with the host components, rendering them ideal biomarkers for therapeutic uses. Current literature has identified a collation of *M. tuberculosis* adhesins, corroborated by Squeglia, Ruggiero [39] and Ramsugit and Pillay [96] (Table 2.1).

In addition, surface protein Rv3194c, lipoprotein aminopeptidase (Rv0418), and GroEL1 (Rv3417c) have been identified as *M. tuberculosis* -associated adhesins [97-99]. Subsequently, novel *M. tuberculosis* associated adhesins: *M. tuberculosis* protein (Mpt) 64 (Rv1980c), fructose-1,6-bisphosphate (Rv0363c), LpdC (Rv0462), isocitrate lyase (Rv0467), proteasome B subunit (Rv2110c), chaperone protein DnaK (Rv0350), S-adenosyl-L-homocysteine (Rv3248c), S-

Table 2.1 List of known *Mycobacterium tuberculosis* adhesin proteins. [39,96]

Adhesin name	Gene loci	Virulence	Host receptor	Reference
LpqH (19-kDa antigen)	Rv3763	TLR 2 agonist, T-cell proliferation	Mannose Receptor of THP-1 monocytes and macrophages	[103]
PstS-1	Rv0934	TLR2 and TLR4 agonist	Macrophages	[104]
Cpn60.2/GroEL2	Rv3418c, Rv0440	Modulation of dendritic cells response	Macrophages	[105]
Alanine- and proline-rich antigen (Apa)	Rv1860	Not specified	Macrophages	[106]
PE-PGRS proteins	Rv1759c, Rv1818c	TLR2 agonist, induces TNF-α release, induces apoptosis and mitochondrial fusion	Fibronectin and macrophages	[107]
Antigen 85 complex	Rv3804c, Rv1886c, Rv0129c, Rv3803c	Inhibit phagosome-lysosome fusion	Fibronectin and elastin	[108]
Glutamine synthetase A1	Rv2220	Not specified	Fibronectin	[109]
Early secreted antigen ESAT-6	Rv3875, Rv3874	Phagosome rupture and translocation to the cytosol	Laminin and lung epithelial cells	[86]
N-acetylmuramoyl-L-alanine amidase	Rv3717	Not specified	Fibronectin and laminin	[110]
Membrane-bound protein	Rv2599	Not specified	Collagen, fibronectin, and laminin	[110]

L, D- transpeptidase	Rv0309	Not specified	Fibronectin and laminin	[110]
HU laminin-binding protein	Rv2986c	Not specified	Laminin	[111]
Protein Kinase D	Rv0931c	Not specified	Brain endothelia and laminin	[112]
Malate synthase	Rv1837c	Not specified	Fibronectin, laminin, and epithelial cells	[113] [114]
Curli pili	Rv3312A	Not specified	Laminin, macrophages, and epithelial cells	[115], [87]
Mycobacterium cell entry-1 protein	Rv0169	Not specified	Epithelial cells	[116]
Type IV pili	Rv3654c-Rv3660c	Inhibits apoptosis	Macrophages and epithelial cells	[87]
Heparin binding haemagglutinin	Rv0475	Induces apoptosis, progression of extrapulmonary infection	Epithelial cells	[117] [118]

adenosylmethionine synthetase (MetK, Rv1392), and iron-regulated elongation factor- Tu (Rv0685), were shown to bind to plasminogen [100].

The surface-associated properties of adhesins may provide biomarker potential, as these proteins may stimulate humoral immune responses, thereby serving as potential serological biomarkers through the induction of antibody production. Importantly, *M. tuberculosis* specific adhesin molecules represent promising diagnostic biomarkers due to their unique composition, which distinguishes them from eukaryotic components. Ongoing validation studies to assess the immunogenic and serological potential of several well-characterized adhesins, including ESAT-6, the Ag85 complex, and MPT64, emphasize their relevance as candidates for *M. tuberculosis* therapeutics [101,102].

2.6. Therapeutic application

Despite the widespread use of the BCG vaccine and conventional screening methods, significant gaps in adult pulmonary protection and reliable antigen-based detection persist, driving ongoing research into next-generation vaccine candidates and serological biomarkers. This section examines the current developmental landscape of TB vaccines and explores the advancement of antigen-based serodiagnostic platforms, outlining the potential to address the critical shortcomings of existing prevention and diagnostic approaches.

2.6.1 Vaccine developmental research

The spectrum of TB disease moves beyond the historical distinction of latent TB infection and active TB infection, rather identifying the gap in between these stages, a progression of intervals that arises from infection to disease [119]. This progression identifies incipient TB, subclinical TB with no symptoms, subclinical with unrecognized symptoms, and then active TB disease that is symptomatic [119]. At the subclinical stage, it surpasses the incipient stage of bacilli replication, moving towards metabolically active infectious status [119]. Therefore, the ideal coverage for all stages would be to rely on effective vaccination strategies that prevent infection within the host across the spectrum of disease (Figure 2.7) [120].

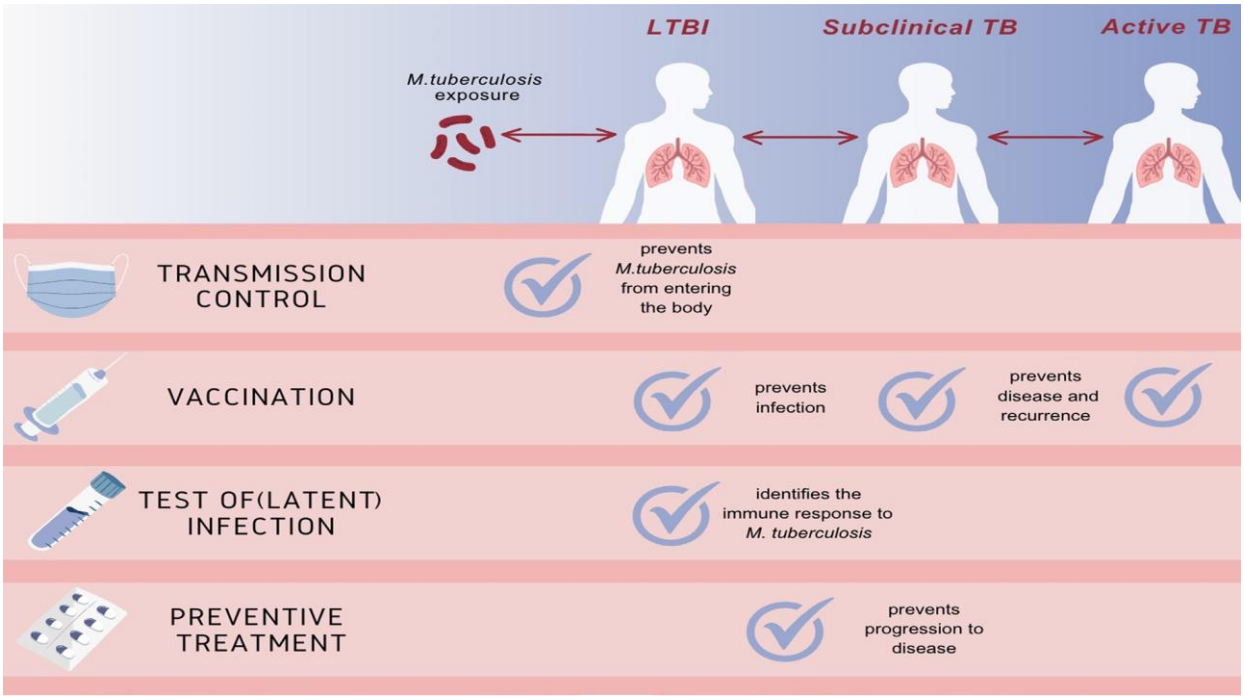


Figure 2.7. The spectrum of tuberculosis infection illustrates the stages of progression with prevention and control strategies. While various strategies of transmission control, testing, and preventive treatment each contribute to TB control, effective vaccination represents the most comprehensive intervention, preventing infection, disease, and transmission across the spectrum [120].

An optimal long-term immunity to TB would either eliminate the pathogen or maintain control by suppressing pathogenesis at the time of host infection that leads to active TB [121]. The Bacille Calmette-Guérin (BCG) vaccine was discovered over 100 years ago, in 1921, and isolated from *Mycobacterium bovis*, the causative agent among cattle [122,123]. Currently, the WHO-

approved BCG vaccine is the only effective and safe vaccine commercially available for TB immunization [1]. It is administered worldwide, with approximately 85% coverage, and provides higher protection against miliary and meningitis TB in children; however, its efficacy against pulmonary TB diminishes in adults [1,124].

In addition, the BCG vaccine was remarkably effective, correlating with lower mortality and morbidity rates during the corona virus disease-19 period and providing a protective immune response among immunized individuals [125]. The development of novel vaccines capable of conferring effective, long-term protection comparable to that of Bacille Calmette–Guérin (BCG) across all age groups remains a critical priority.

The ideal TB vaccine strategy encompasses pre-exposure and post-exposure establishment. Primarily, pre-exposure effectiveness should induce an immune-regulated response that is able to clear infection before dissemination and is shown to be more effective than the current BCG immunization [122]. Secondly, post-exposure, an effective vaccine should eradicate the latent state of TB, thus prevent reactivation of the infection and elicit a long-term immune response [122]. Current research explores the efficacy of novel vaccine candidates and the development of recombinant BCG variants to offer stronger, universal protection (Table 2.2), across various clinical phases [126].

Current vaccine approaches include whole cell, adjuvanted protein subunit, and DNA vaccine derivatives [123]. Whole-cell vaccines refers to recombinant mycobacterial strains that are either activated or live-attenuated [123]. These include phase 2b and 3, DAR-901, RUTI, BCG traveler vaccine, GamTBvac, Immuvac, MTBVAC, BBV169, and VPM1002 [126]. In addition, vaccine development utilizing fusion-derived antigen constructs to increase immune-modulatory efficacy offers an enhanced direction [122]. These included at various phases of the vaccine pipeline and are incorporated into the vaccine platform, namely 1) adjuvant-based: H107/CAF10b, AEC/BC02, ID93 + GLA-SE (Rv3620, Rv3619, Rv2608 and Rv1813), and M72/AS01E (Rv0125 and Rv1196), 2) RNA-based: BNT164a1/BNT164b1, and 3) viral delivery-based: AdAg85A(Ag85A), Ad5-105K, and TB/FLU-05E(ESAT-6 and Ag85a).

However, previously included subunit vaccines, namely MV85A, Ad-Ag85A (Ag85a), IC31 (ESAT-6, Rv2660, and Ag85B), ARERAS-422 (Ag85 complex), and DEAE (ESAT-6, Culture Filtrate Protein: CFP-10, and Ag85A), were removed as these did not pass the various phases of clinical trials due to safety, effectiveness, and immunogenicity requirements [127].

Although DNA vaccines are potential commercial vaccine candidates, they are still being developed and yet to progress to clinical phase trials [122].

Table 2.2: Current tuberculosis vaccine candidates under clinical development and active trials [126].

Vaccine Candidate	Vaccine platform	Primary endpoint	Target Population	Trial Status
Phase 1				
H107/CAF10b	Protein/Adjuvant	Prevention of disease	People with/without MTB	Yes
AdAg85A (McMaster Univ.)	Viral vector	Prevention of disease	People with/without MTB	No
Ad5-105K (CanSino, Etana)	Viral vector	Prevention of disease	People with/without MTB	Yes
Phase 2a				
AEC/BC02 (Anhui Zhifei Longcom)	Protein/Adjuvant	Prevention of disease	People with MTB	Yes
ChAdOx1.85A + MVA85A (Univ. Oxford)	Viral vector	Prevention of disease	People with MTB and people cured of active TB	Yes
ID93 + GLA-SE (NIAID/NIH, Quratis)	Protein/Adjuvant	Prevention of disease	People with MTB	Yes
BNT164a1/BNT164b1 (BioNTech)	RNA	Prevention of disease	Not specified	Yes
TB/FLU-05E (Smor Res Inst, Russia MoH)	Viral vector	Prevention of infection	People with/without MTB	No
Phase 2b				
DAR-901 (Dartmouth)	Mycobacterial (Inactivated)	Prevention of disease	PLHIV, adults, adolescents	Yes
RUTI® (Archivel Farma)	Mycobacterial (Inactivated)	Therapeutic	People with MTB, MDR, active TB and people cured of TB	Yes
Phase 3				
BCG, traveler vaccine (HJF)	Mycobacterial (Live attenuated)	Prevention of infection	People without MTB	Yes
GamTBvac (Gamaleya Res. Centre)	Protein/Adjuvant	Prevention of disease	People without MTB	Yes
Immuvac, MIP (ICMR, Cadila Pharma)	Mycobacterial (Inactivated)	Prevention of disease	Not specified	Yes
MTBVAC (Biofabri) **BBV169** (Bharat)	Mycobacterial (Live attenuated)	Prevention of disease	Not specified	Yes
VPM1002 (SIIPL, VPM)	Mycobacterial (Live attenuated)	Prevention of disease/ infection	People cured of MTB	Yes
M72/AS01E (Gates MRI, GSK)	Protein/Adjuvant	Prevention of disease	People with/without MTB and PLHIV	Yes

MTB- *Mycobacterium tuberculosis* infection, PLHIV- People living with HIV, RNA- Ribonucleic acid

Among the highly anticipated candidate vaccines, M72/AS01E has been extensively evaluated among three African populations, this showed a 54% vaccine efficacy within the first 24-month and 49% vaccine efficacy within the 36-month follow-up [128]. These encouraging outcomes increase efforts to identify novel mycobacterial proteins capable of eliciting protective immune-modulatory response, characterized as adhesins. Furthermore, the region of deletion (RD1) within BCG that removes the virulent ESX-1 secretory system, expresses immunodominant ESAT-6 and CFP-10, could pose for prospective vaccine development and investigation [123].

2.6.2 Serodiagnosis development

The importance of TB diagnosis remains a core aspect of TB control and elimination. Immunological assays that incorporate mycobacterial biomarkers have the potential to detect various stages of infection within the spectrum, as well as monitoring treatment response to anti-TB therapy [129]. Currently, the only commercially produced biomarker-based detection of TB is the LF-LAM assay, which detects the 17.5 kDa cell wall-associated LAM antigen in PLHIV [28]. Other LTBI detection methods, tuberculin skin tests that identify tuberculin purified protein derivative products have been shown to produce false positives among BCG-immunized populations [22]. Subsequently, IGRAs that rely on T-cell activation of CFP-10 and ESAT-6 are not affected by BCG; however, both tuberculin skin and IGRA assays cannot differentiate between latency and active TB [22]. Antigen-based biomarkers may represent a key strategy for TB treatment management, as they can reflect therapeutic efficacy from initiation through completion of therapy, while also enabling clinical prediction of relapse and guiding the use of alternative anti-TB regimens [130]. Ideally, antigen selection should distinguish between latent and active TB infection, exhibit high human leukocyte antigen (HLA)-binding affinity, demonstrate immunodominance, retain memory responses upon re-exposure to active infection, and identify the levels of cytokines induced [131].

Serological detection assays to monitor the host antibody response to viable mycobacterial antigens have been explored. A study published in Shanghai, China demonstrated that the area under the curve for Rv0310c (AUC: 0.800) and Rv1255c (AUC: 0.808), within the RD region, had a higher performance in the receiver operating characteristic analysis when compared to that of

ESAT-6 (AUC: 0.655) and CFP-10 (AUC: 0.623) among BCG vaccinated population samples [132]. Similarly, a study conducted in Agra, India, determined that mycolyl transferase (Ag85 complex) had a higher sensitivity of 84.1% when compared to ESAT-6 (64.9%) and CFP-10 (66%) with high specificity (≥85%) across all antigens [133]. A study conducted in Beijing, China, showed the 38kDa and *M. tuberculosis* 48 protein having a higher sensitivity and specificity (73.6% and 85.4%, respectively) in comparison to ESAT-6 (73.2% and 77.7%, respectively) and CFP-10 (60.4% and 73.8%, respectively) [134]. These studies showed variable outcomes especially in the detection of CPF-10 and ESAT-6, which supports the heterogeneity of antigen recognition and reflects the difference in immune response, stage of infection and host genetics among individually TB infected patients [131,135].

An evaluation of the *M. tuberculosis* adhesins malate synthase and MPT51 in combination showed an 80% sensitivity among HIV negative smear positive TB patients in India, but when conducted in the USA, the sensitivity (58%) was lower in detecting *M. tuberculosis* [136] . In addition to single antigen identification, the construction of fusion proteins with a combination of antigenic epitope regions has enabled improvement of sensitivity and specificity of detection. A comparison study evaluated the sensitivity of fusion constructions of EspC/CFP7 (43%), HspX/EspC/CFP7 (58%) and HspX/EspC/CFP7/PPE57 (69%) and compared it to singular antigens [137]. This study demonstrated that individual antigens yielded low sensitivity (≤35%), whereas exposure to a multi-epitope fusion construct resulted in a more favorable outcome with enhanced sensitivity [137]. Lipolytic enzymes such as lipases and hydrolases produced by *M. tuberculosis* utilize host fatty acids as a carbon source and which are stored as triacylglycerol [138]. Specific lipolytic antigens (Rv0183, Rv1984c, Rv3097c and Rv3452) were evaluated to distinguish between latent and active TB, with sensitivity ranging from 73-90% and specificity ranging from 93-97%, with Rv3452 outperforming all antigens [138].

Enzyme-linked immunosorbent assay-based, multiplex microbead, lateral flow, and protein chip array assays hold considerable promise as serological diagnosis tools for *M. tuberculosis* detection. The ease of specimen collection, operational simplicity, low cost, and compatibility with point-of-care formats make them especially valuable in resource-limited settings [139]. Notably, antigen-based serodiagnosis platforms have the potential to complement conventional sputum microscopy, thereby enhancing diagnostic accessibility in TB-endemic regions [136].

2.7 Mycobacterial proteomics

The integration of multi-omics technologies comprising genomics, proteomics, transcriptomics, and metabolomics has substantially advanced our understanding of infectious disease pathogenesis [11]. This approach provides insights not only from a clinical perspective but also into the underlying molecular systems and host-pathogen interactions, facilitating the discovery of novel biomarkers and therapeutic targets for diagnostic and therapeutic interventions [140]. The current expansion of structural proteogenomic databases over the past two decades has enabled a new template for computational research, which serves to complement and support mycobacterial experimental validation.

2.7.1 Bioinformatic approach

Bioinformatic approaches have become indispensable for the systematic identification and functional characterization of adhesins, surface-exposed molecules, and immunomodulatory biomarkers that are critical for establishing infection before phenotypic laboratory driven assays [76]. Computational strategies employed in mycobacterial proteomics enable the prediction and analysis of functionally uncharacterized mycobacterial proteins, with a focus on systematic pipelines that combine multiple analytical tools for the benefit of potential vaccine construction [141]. Conserved hypothetical (CH) proteins are open reading frame sequences within a bacterial genome that encode proteins of unknown functions. Conserved hypothetical proteins have remained preserved across all phylogenic lineages and therefore of biological importance [142]. Such tools can unravel the predicted roles these CH proteins may play and further our understanding of these proteins in pathogenesis, leading to therapeutic development.

2.7.1.1 Adhesin and adhesin-like prediction

Computational analysis employs the prediction mechanisms of bioinformatic tools to identify specific traits that these genomic and proteomic structures align with based on algorithms [143]. Given the limitations of these algorithms, multiple tools may be strategically employed to enhance the prediction of these proteins. Adhesins are critical determinants of virulence in pathogenic bacteria such as *M. tuberculosis*, and their characterization holds substantial therapeutic potential [73]. Consequently, the computational prediction of novel adhesins represents a valuable strategy for identifying new targets for drug and vaccine development. A range of bioinformatic tools designed to analyze adhesin and adhesin-like proteins, notably, the Malarial Adhesin and Adhesin-

like Protein Predictor [143], the FungalRV [144] and FaaPred [145] predictors for fungal adhesins, the AdhesiomeR [146] database for classifying *E. coli* adhesins, a Software Program to predict Adhesin and Adhesin-like protein using Artificial Neural Networks (SPAAN) tool [147]. The latter tool, SPAAN, demonstrated high sensitivity (89%) and specificity (100%) among well-characterized adhesins from a range of virulent bacteria [147]. SPAAN operates by calculating the probability of adhesin (P_{ad}) based on five algorithmic attributes of this advanced artificial neural network that focuses on 105 compositional properties [147]. These include amino acid frequencies, multiplet frequencies, dipeptide frequencies, charge, and hydrophobic composition of the protein sequences [147].

The SPAAN tool has been the primary tool for the identification of adhesins across a range of bacterial species. A study conducted in 2018 identified 98 adhesins and adhesin-like proteins with a high P_{ad} (greater than 0.7) among *Lactobacillus fermentum*, a probiotic with potential anti-adhesive properties against methicillin-resistant *Staphylococcus aureus* [148]. Subsequently, a 2024 study predicted that eight outer membrane proteins of *Leptospira borgpetersenii* possess adhesin or adhesin-like functions using SPAAN, which were experimentally validated for binding to fibronectin and laminin [149]. Similarly, a recent study conducted in 2026 investigated the adhesin potential of outer membrane proteins in Gram-negative *Brucella melitensis*, whereby the majority of the proteins had a P_{ad} of greater than 0.7 [150]. There has been one study, conducted by Kumar *et al.* in 2013, identified novel *M. tuberculosis* adhesins using SPAAN, in which three novel adhesins, Rv3717, Rv2599, and Rv0309, were shown to bind to ECM using an image-based ELISA developed on activated polypropylene micro test plates [110]. The SPAAN tool has provided a basis for the prediction of functional adhesin potential. However, continued integrated pipelines of such predictive tools with validation is essential for fully characterizing the adhesive mechanisms of bacteria, potentially uncovering new targets for therapeutic interventions and vaccines.

2.7.1.2 Structural prediction of adhesins

In addition to SPAAN adhesin prediction, a robust pipeline of bioinformatic tools can be used to further assess protein characteristics which include translocation analysis, homology modelling and subcellular localization. This approach potentially improves the predictive quality for identifying mycobacterial adhesins and adhesin-like proteins.

Homology modelling and domain prediction methods improve the prediction analysis of the unknown protein putative function. Bioinformatic tools such as Phyre2 and Pfam provide protein predictions based on structural homology to known proteins from a protein database [151,152]. An online server for Pfam is based on sequence similarity using a database that predominantly relies on domain and family classification has recently been updated to include artificial intelligence and machine learning protocols [153]. While both tools predict homology and function of unknown proteins, Phyre2 relies on structural prediction, showing alignment of the protein sequence to known proteins with an analysis of the coverage and significance (Phyre2 rating) to the template by comparison across a database of genomes [151]. A study conducted used Phyre2 homology modelling to identify therapeutic targets of conserved hypothetical proteins of *Pseudomonas aeruginosa* by screening a large dataset of known proteins [154]. In a proteogenomic study in 2026, short open reading frames (ORFs) within the *M. tuberculosis* genome were investigated for protein function. Phyre2 was used to determine the predicted function of these unknown sequences with more than 50% confidence [155].

In addition to homology modelling, subcellular localization prediction tools such as PsortB, Subloc, and Loctree3 have been useful tools to identify surface-accessible proteins that potentially mediate host interactions [76]. In addition, Mycosub is a specialized subcellular localization tool designed to identify mycobacterial proteins translocated to the cell envelope or outer membrane regions [156]. In addition, secretory pathway mechanisms enhance the pathogenic properties of adhesins [157]. Various prediction tools may be used to identify secreted proteins, similar to ESAT-6 and CFP10 [123]. SignalP identifies a short sequence located at the amino (N)-terminus that codes for the secretory pathway [158]. However, other proteins may adapt to be secreted using the non-classical pathway in which leaderless proteins (not having a translocation sequence at the N-terminus) can be detected using the SecretomeP tool [157,159]. The identification of adhesins through the bioinformatic approach requires a multi-parametric pipeline that incorporates various systematic frameworks, particularly for *M. tuberculosis* and mycobacterial species alike.

2.7.1.3 Molecular docking approach for adhesin identification

Molecular docking is a structure-based computational approach that investigates the molecular binding interaction between two molecules, namely a ligand and a receptor, forming a stable

complex [160]. As two molecules are introduced in the simulation, atomic bond interactions, including hydrogen bonds, Van der Waals forces, and electrostatic forces, form a measurable binding affinity among the two structures [161]. The interactive state in which the two interactions occur is measured by a docking score that ranks the binding affinity and the strength of the interaction, generally related to the free energy of binding [161]. Since the availability of crystalline structures of proteins or the high-throughput server prediction of the tertiary structure using Alphafold or I-Tasser, these tools can identify specific amino acid sequences and potential binding residues [162,163]. The advantages of performing molecular docking have drastically accelerated the way in which the development of drug targets, as well as protein-protein interactions, is understood. Therefore, numerous molecular docking tools focus on various search algorithms that rely on shape matching, genetic algorithms, evolutionary optimization/algorithms, or a hybrid approach [160].

Subsequent visualization tools, such as UCFS Chimera (University of California, San Francisco) or Maestro (Schrodinger, Cambridge, USA), can be used to prepare the protein structures, which include structure clean-up, charge assignment, and energy minimization for downstream applications such as molecular dynamics investigation [160,164]. Numerous studies have explored the relationship between protein-protein interactions by adopting the molecular docking approach [165-168]. A study in 2013 produced findings that *M. tuberculosis* adhesin, secreted protein Ag85 complex, has a strong phenotypic association with ECM-associated elastin and its precursor tropoelastin [165]. Molecular docking was shown to identify a strong binding site with electrostatic interactions among key binding residues [165].

Other significant studies utilized molecular docking techniques to ascertain drug target sites within mycobacterial proteins, namely membrane-bound efflux pump Rv1250 susceptible to isoniazid [166], peptidoglycan biosynthesis protein MurI susceptible to benzodiazepine amine derivatives [167], and LAM biosynthesis protein arabinosyltransferase susceptible to isoniazid, ethambutol, and derivatives [168]. In addition, molecular docking applications extend to identifying epitope regions for vaccine targets, as shown in a study in 2025. This study identified multiple epitope regions of cell-wall-associated PE-PGRS33 protein to have a high binding affinity to toll-like receptor-2 vaccine, which was validated in an in vivo model to stimulate the immune system [169]. Paired with immunoinformatic tools such as C-ImmSim for immune simulation, and epitope prediction tools including Bepipred, ABCpred, Ellipro, and Discotope for linear and

discontinuous epitopes, these tools can help create an improved pipeline for an adhesin-related multiepitope vaccine [141,157,170].

Computational analysis using molecular docking has emerged as an ideal tool for structural proteomics and therapeutic drug discovery, providing insight into atomic interactions and key binding sequences among diverse mycobacterial targets. As computational methods continue to advance alongside experimental validation techniques, molecular docking is expanding its applications in therapeutic interventions.

2.7.2 Mycobacterial protein synthesis

Experimental validation of adhesin and adhesin-like mycobacterial proteins entails the process of synthesizing gene sequences, molecular cloning, protein expression, and purification to produce recombinant proteins. The selection of *M. tuberculosis* ORFs can be retrieved from mycobacterial genomic databases, such as the Mycobrowser online tool, which annotates over 4000 genes related to the *M. tuberculosis* genome [171]. The Mycobrowser online tool (https://mycobrowser.epfl.ch) provides the complete DNA sequence for each gene, allowing it to be synthesized.

Primer design is an important component of molecular cloning , as it is essential to include the full gene transcript to code for the amino acid sequence of the expressed target protein [172-175]. However, certain secretory sequences identified by signal peptide prediction may be excluded, thereby improving protein solubility during downstream expression and purification [110,172]. Primers typically include overhangs containing restriction enzyme recognition sites at the 5' and 3' ends, chosen to match the requirements of the selected cloning vector [173]. *In silico* SnapGene software facilitates the simulation of cloning strategies and expression of the specified protein prior to gene amplification using the polymerase chain reaction (PCR), in addition to sequence analysis [174]. Various cloning vectors, such as plasmids, have been developed as essential tools for molecular cloning [175]. Key elements of a plasmid vector include: (i) the origin of replication, which enables autonomous replication and is often regulated by an inducer at the promoter region activated by a substrate added to the growth media, (ii) a selectable marker, such as antibiotic resistance, to ensure successful transformation of the recombinant plasmid, (iii) multiple cloning sites for inserting the amplified DNA with restriction ends, and an affinity tag like a 6x histidine (His-) affinity tag or glutathione-S-transferase (GST-) affinity tag, fused to the expressed protein to aid in downstream detection after purification [175,176]. Cloning in a

bacterial system such as *E. coli*, particularly the DH5α strain, benefits from specific genetic mutations that improve plasmid stability. Important mutations, including recombination protein (Rec)A1 and endonuclease (End)A1, enhance plasmid maintenance and prevent non-specific degradation of the recombinant plasmid DNA, respectively [177].

Subsequently, the *E. coli* BL21 (DE3) strain has been genetically engineered to improve the expression of recombinant proteins by deleting the cytoplasmic (Lon) and outer membrane (OmpT) proteases, both of which are known to degrade extracellular proteins [177]. When performing expressions at various scales, optimization should be carried out to improve protein yields for downstream applications. This includes optimizing the concentration of inducers such as isopropyl β-D-1-thiogalactopyranoside (IPTG), the growth temperature, and the incubation time required to achieve optimum yields [178,179].

The purification of mycobacterial recombinant proteins, particularly adhesin-associated, can be challenging due to inherent physicochemical properties [110]. Overexpression of the recombinant protein during induction within the *E. coli* expression system can lead to improper folding and formation of aggregates called inclusion bodies [180,181]. This is frequent among cell-wall or surface-exposed bacterial proteins due to their solubility issues [110]. Therefore, recombinant proteins are subjected to a solubilization process using denaturant reagents to improve purification, followed by resuspension of all the proteins to refold into their natural conformation for downstream biological assessment [182].

When conventional bacterial cell lysis buffers pose a challenge for the solubilization of total cell lysates, denaturing detergents and chaotropic agents such as sodium dodecyl sulfate (SDS), urea (up to 8 Molar), or guanidine hydrochloride are used to disaggregate inclusion bodies [183,184]. Additionally, detergent analogs of SDS, like sarkosyl and sodium lauroyl glutamate, have been shown to gently dissociate protein conformations and promote protein refolding[185]. Purification techniques can be rather specific to the type of affinity tag adopted within the recombinant protein [186] .

A study by Massiah *et al.* [186] improved solubilization for batch method purification of both GST and His-tag fused proteins using sarkosyl, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS), and Triton X-100 [186]. This study showed that more than 95% of the recombinant protein was able to be solubilized with 10% sarkosyl from the expressed inclusion bodies. The addition of CHAPS and Triton X-100 was incorporated to improve the binding of the

recombinant protein to the affinity tag-related matrix [186]. In addition to chaotropic reagents, it has been shown that adjusting expression conditions, particularly by lowering the temperature during incubation, reduces the aggregation of overexpressed protein and improves the folding of the native protein conformation [187]. There are, therefore, a great number of factors that should be taken into consideration when designing the framework for the synthesis of mycobacterial recombinant protein.

References

1. World Health Organization. Global tuberculosis report 2025 Geneva: World Health Organization: WHO. 2025. Available from: https://iris.who.int/server/api/core/bitstreams/e97dd6f4-b567-4396-8680-717bac6869a9/content.

2. Koo H-K, Min J, Kim HW, *et al.* Prediction of treatment failure and compliance in patients with tuberculosis. BMC Infectious Diseases. 2020;20(1):622. 10.1186/s12879-020-05350-7

3. Nezenega ZS, Perimal-Lewis L, Maeder AJ. Factors influencing patient adherence to tuberculosis treatment in Ethiopia: a literature review. International Journal of Environmental Research and Public Health. 2020;17(15):5626.

4. Farhat M, Cox H, Ghanem M, *et al.* Drug-resistant tuberculosis: a persistent global health concern. Nature Reviews Microbiology. 2024;22(10):617-635. https://doi.org/10.1038/s41579-024-01025-1

5. Heidary M, Shirani M, Moradi M, *et al.* Tuberculosis challenges: Resistance, co-infection, diagnosis, and treatment. European Journal of Microbiology and Immunology. 2022;12(1):1-17. https://doi.org/10.1556/1886.2021.00021

6. Moliva JI, Turner J, Torrelles JB. Immune responses to Bacillus Calmette–Guérin vaccination: why do they fail to protect against *Mycobacterium tuberculosis*? Frontiers in Immunology. 2017;8:407. https://doi.org/10.3389/fimmu.2017.00407

7. Bares SH, Swindells S. Latent tuberculosis and HIV infection. Current Infectious Disease Reports. 2020;22(7):17. https://doi.org/10.1007/s11908-020-00726-x

8. Barberis I, Bragazzi N, Galluzzo L, Martini M. The history of tuberculosis: from the first historical records to the isolation of Koch's bacillus. Journal of Preventive Medicine and Hygiene 2017;58(1):E9. https://doi.org/10.15167/2421-4248/jpmh2017.58.1.728

9. Huang Y, Ai L, Wang X, Sun Z, Wang F. Review and updates on the diagnosis of tuberculosis. Journal of Clinical Medicine [Internet]. 2022; 11(19):5826. https://doi.org/10.3390/jcm11195826

10. Nogueira BMF, Krishnan S, Barreto-Duarte B, *et al.* Diagnostic biomarkers for active tuberculosis: progress and challenges. EMBO Molecular Medicine. 2022;14(12):e14088. https://doi.org/10.15252/emmm.202114088

11. Zhuang L, Yang L, Li L, Ye Z, Gong W. *Mycobacterium tuberculosis*: immune response, biomarkers, and therapeutic intervention. MedComm. 2024;5(1):e419. https://doi.org/10.1002/mco2.419

12. World Health Organization. Global tuberculosis report 2024. Geneva: World Health Organization, 2024. https://www.who.int/publications/i/item/9789240089488

13. Falzon D, Zignol M, Bastard M, Floyd K, Kasaeva T. The impact of the COVID-19 pandemic on the global tuberculosis epidemic. Frontiers in Immunology. 2023;14:1234785. https://doi.org/10.3389/fimmu.2023.1234785

14. Gong W, Wu X. Differential Diagnosis of latent tuberculosis infection and active tuberculosis: A key to a successful tuberculosis control strategy. Frontiers in Microbiology 2021;12 (2021): 5592. 10.3389/fmicb.2021.745592

15. Statistics South Africa. Mortality and causes of death in South Africa, 2016: Findings from death notification. Pretoria: StatsSA. 2018 [accessed 2026 February 20]. Available from: https://www.statssa.gov.za/publications/P03093/P030932016.pdf.

16. Statistics South Africa. Mortality and causes of death in South Africa: Findings from death notification. Pretoria: StatsSA. 2022 [accessed 2026 January 6]. Available from: https://www.statssa.gov.za/publications/P03093/P030932022.pdf&ved=2ahUKEwjW9Nih5LOSAxVRUEEAHa1FMYwQFnoECBkQAQ&usg=AOvVaw26V1pll-Mxsl3wqhmmPmUT.

17. South African Department of Health. NATIONAL TB RECOVERY PLAN 4.0: April 2025 – MARCH 2026. In: SADOH, editor. Pretoria: South African Department of Health; 2025. https://www.health.gov.za/wp-content/uploads/2025/05/TB-Recovery-Plan-4_final_250526.pdf

18. Churchyard G, Mametja L, Mvusi L, *et al.* Tuberculosis control in South Africa: successes, challenges and recommendations. South African Medical Journal. 2014;104(3):244-248. https://doi.org/10.7196/samj.7689

19. Heyckendorf J, Georghiou SB, Frahm N, *et al.* Tuberculosis treatment monitoring and outcome measures: new interest and new strategies. Clinical Microbiology Reviews. 2022;35(3):e00227-00221. https://doi.org/10.1128/cmr.00227-21

20. World Health Organisation. WHO releases first-ever Standard on universal access to rapid TB diagnostics Geneva: World Health Organisation. 2023 [Accessed 2025 December 26]. Available from: https://www.who.int/news/item/19-04-2023-who-releases-first-ever-standard-on-universal-access-to-rapid-tb-diagnostics.

21. Pandey BD, Ngwe Tun MM, Shah Y, Suzuki Y, Morita K. Ending tuberculosis by 2030: understanding the transmission. The Lancet Regional Health – Western Pacific. 2023;38. https://doi.org/10.1016/j.lanwpc.2023.100851

22. Yadav R, Arora J, Bhalla M. Laboratory diagnosis of tuberculosis: conventional and newer methods. International Journal of Scientific Research. 2020;9(10):319-324. https://doi.org/10.21275/SR20930230327

23. Alsayed SSR, Gunosewoyo H. Tuberculosis: pathogenesis, current treatment regimens and new drug targets. International Journal of Molecular Sciences. 2023;24(6):5202. https://doi.org/10.3390/ijms24065202

24. Zaporojan N, Negrean RA, Hodișan R, Zaporojan C, Csep A, Zaha DC. Evolution of Laboratory diagnosis of tuberculosis. clinics and practice. 2024; 14(2):388-416. https://doi.org/10.3390/clinpract14020030

25. Stanikzai AU. Performance evaluation, comparison and degree of effectiveness of five different techniques for detection and isolation of *Mycobacterium tuberculosis*. Optimum Science Journal. 2025(3):1-15. https://doi.org/10.5281/zenodo.15119106

26. Salam MA, Al-Amin MY, Pawar JS, Akhter N, Lucy IB. Conventional methods and future trends in antimicrobial susceptibility testing. Saudi Journal of Biological Sciences. 2023;30(3):103582. https://doi.org/10.1016/j.sjbs.2023.103582

27. Yusoof KA, García JI, Schami A, *et al.* Tuberculosis phenotypic and genotypic drug susceptibility testing and immunodiagnostics: a review. Frontiers in Immunology. 2022;13(2022):6828. https://doi.org/10.3389/fimmu.2022.87076828.

28. World Health Organization. WHO consolidated guidelines on tuberculosis. Module 3: diagnosis–rapid diagnostics for tuberculosis detection: World Health Organization; 2024. Available from: https://www.who.int/publications/i/item/9789240089488

29. Qin ZZ, Pai M, Van Gemert W, Sahu S, Ghiasi M, Creswell J. How is Xpert MTB/RIF being implemented in 22 high tuberculosis burden countries? European Respiratory Journal. 2015;45(2):549-554. https://doi.org/10.1183/09031936.00147714

30. Rimal R, Shrestha D, Pyakurel S, *et al.* Diagnostic performance of GeneXpert MTB/RIF in detecting MTB in smear-negative presumptive TB patients. BMC Infectious Diseases. 2022;22(1):321. https://doi.org/10.1186/s12879-022-07287-5

31. Navasardyan I, Miwalian R, Petrosyan A, Yeganyan S, Venketaraman V. HIV–TB coinfection: current therapeutic approaches and drug interactions. Viruses. 2024;16(3):321. https://doi.org/10.3390/v16030321

32. Tiberi S, Utjesanovic N, Galvin J, *et al.* Drug-resistant TB – latest developments in epidemiology, diagnostics and management. International Journal of Infectious Diseases. 2022;124:S20-S25. https://doi.org/10.1016/j.ijid.2022.03.026

33. Quenard F, Fournier PE, Drancourt M, Brouqui P. Role of second-line injectable antituberculosis drugs in the treatment of MDR/XDR tuberculosis. International Journal of Antimicrobial Agents. 2017;50(2):252-254. https://doi.org/10.1016/j.ijantimicag.2017.01.042

34. Peloquin CA, Davies GR. The treatment of tuberculosis. Clinical Pharmacology and Therapeutics. 2021;110(6):1455-1466. https://doi.org/10.1002/cpt.2261

35. Doan TN, Fox GJ, Meehan MT, *et al.* Cost-effectiveness of 3 months of weekly rifapentine and isoniazid compared with other standard treatment regimens for latent tuberculosis infection: a decision analysis study. Journal of Antimicrobial Chemotherapy. 2018;74(1):218-227. https://doi.org/10.1093/jac/dky403

36. Teo AKJ, Singh SR, Prem K, Hsu LY, Yi S. Duration and determinants of delayed tuberculosis diagnosis and treatment in high-burden countries: a mixed-methods systematic review and meta-analysis. Respiratory Research. 2021;22(1):251. https://doi.org/10.1186/s12931-021-01841-6

37. Stop TB Partnership. 2025 Global new TB drug pipeline 2025 [accessed 2025 November 11]. Available from: https://www.newtbdrugs.org./pipeline/clinical.

38. Ryndak MB, Chandra D, Laal S. Understanding dissemination of *Mycobacterium tuberculosis* from the lungs during primary infection. Journal of Medical Microbiology. 2016;65(5):362-369. https://doi.org/10.1099/jmm.0.000238

39. Squeglia F, Ruggiero A, De Simone A, Berisio R. A structural overview of mycobacterial adhesins: Key biomarkers for diagnostics and therapeutics. Protein Science. 2018;27(2):369-380. https://doi.org/10.1002/pro.3346

40. Bo H, Moure UAE, Yang Y, *et al. Mycobacterium tuberculosis*-macrophage interaction: Molecular updates. Frontiers in Cellular and Infection Microbiology. 2023;13(2023). https://doi.org/10.3389/fcimb.2023.1062963

41. Ruiz A, Guzmán-Beltrán S, Carreto-Binaghi LE, Gonzalez Y, Juárez E. DNA from virulent *M. tuberculosis* induces TNF-α production and autophagy in M1 polarized macrophages. Microbial Pathogenesis. 2019;132:166-177. https://doi.org/10.1016/j.micpath.2019.04.041

42. Rahlwes KC, Dias BRS, Campos PC, Alvarez-Arguedas S, Shiloh MU. Pathogenicity and virulence of *Mycobacterium tuberculosis*. Virulence. 2023;14(1):2150449. https://doi.org/10.1080/21505594.2022.2150449

43. Riojas MA, McGough KJ, Rider-Riojas CJ, Rastogi N, Hazbón MH. Phylogenomic analysis of the species of the *Mycobacterium tuberculosis* complex demonstrates that *Mycobacterium africanum*, *Mycobacterium bovis, Mycobacterium caprae, Mycobacterium microti* and *Mycobacterium pinnipedii* are later heterotypic synonyms of *Mycobacterium tuberculosis*. International Journal of Systematic and Evolutionary Microbiology. 2018;68(1):324-332. https://doi.org/10.1099/ijsem.0.002507

44. Bespiatykh D, Bespyatykh J, Mokrousov I, Shitikov E. A comprehensive map of *mycobacterium tuberculosis* complex regions of difference. mSphere. 2021;6(4):10.1128/msphere.00535-00521. https://doi.org/10.1128/msphere.00535-21

45. Batt SM, Minnikin DE, Besra GS. The thick waxy coat of mycobacteria, a protective layer against antibiotics and the host's immune system. Biochemical Journal. 2020;477(10):1983-2006. https://doi.org/10.1042/bcj20200194

46. Daffé M, Marrakchi H. Unraveling the structure of the mycobacterial envelope. Microbiology Spectrum. 2019;7(4):10.1128. https://doi.org/10.1128/microbiolspec.gpp3-0027-2018

47. Dulberger CL, Rubin EJ, Boutte CC. The mycobacterial cell envelope — a moving target. Nature Reviews Microbiology. 2020;18(1):47-59. https://doi.org/10.1038/s41579-019-0273-7

48. Jacobo-Delgado YM, Rodríguez-Carlos A, Serrano CJ, Rivas-Santiago B. *Mycobacterium tuberculosis* cell-wall and antimicrobial peptides: a mission impossible? Frontiers in Immunology. 2023;14(2023):4923. https://doi.org/10.3389/fimmu.2023.1194923

49. Vilchèze C. Mycobacterial cell wall: A source of successful targets for old and new drugs. Applied Sciences. 2020;10(7):2278. https://doi.org/10.3390/app10072278

50. Romano M, Squeglia F, Kramarska E, *et al.* A structural view at vaccine development against *M. tuberculosis.* Cells. 2023;12(2):317. https://doi.org/10.3390/cells12020317

51. Bruch EM, Petrella S, Bellinzoni M. Structure-based drug design for tuberculosis: challenges still ahead. Applied Sciences. 2020;10(12):4248. https://doi.org/10.3390/app10124248

52. Ahmad F, Rani A, Alam A, *et al.* Macrophage: a cell with many faces and functions in tuberculosis. Frontiers in Immunology. 2022;13. https://doi.org/10.3389/fimmu.2022.747799

53. Liu D, Wang J, Xu Z, Chen X, Jiao Xa. Phagocytosis: strategies for macrophages to hunt *Mycobacterium tuberculosis*. One Health Advances. 2024;2(1):32. https://doi.org/10.1186/s44280-024-00065-9

54. Xander C, Rajagopalan S, Jacobs William R, Braunstein M. The SapM phosphatase can arrest phagosome maturation in an ESX-1 independent manner in *Mycobacterium tuberculosis* and BCG. Infection and Immunity. 2024;92(7):e00217-00224. https://doi.org/10.1128/iai.00217-24

55. Ravesloot-Chávez MM, Van Dis E, Stanley SA. The innate immune response to *Mycobacterium tuberculosis* infection. Annual Review of Immunology. 2021;39(2021):611-637. https://doi.org/10.1146/annurev-immunol-093019-010426

56. Flynn JL, Chan J. Immune cell interactions in tuberculosis. Cell. 2022;185(25):4682-4702. https://doi.org/10.1016/j.cell.2022.10.025

57. Carabalí-Isajar ML, Rodríguez-Bejarano OH, Amado T, *et al.* Clinical manifestations and immune response to tuberculosis. World Journal of Microbiology and Biotechnology. 2023;39(8):206. https://doi.org/10.1007/s11274-023-03636-x

58. Khan N, Vidyarthi A, Amir M, Mushtaq K, Agrewala JN. T-cell exhaustion in tuberculosis: pitfalls and prospects. Critical Reviews in Microbiology. 2017;43(2):133-141. https://doi.org/10.1080/1040841X.2016.1185603

59. Pathak SK, Basu S, Basu KK, *et al.* Direct extracellular interaction between the early secreted antigen ESAT-6 of *Mycobacterium tuberculosis* and TLR2 inhibits TLR signaling in macrophages. Nature Immunology. 2007;8(6):610-618. https://doi.org/10.1038/ni1468

60. Cardona P, Cardona P-J. Regulatory T cells in *Mycobacterium tuberculosis* infection. Frontiers in Immunology. 2019;10(2019):2139. https://doi.org/10.3389/fimmu.2019.02139

61. Cohen SB, Gern BH, Urdahl KB. The tuberculous granuloma and preexisting immunity. Annual Review of Immunology. 2022;40(2022):589-614. https://doi.org/10.1146/annurev-immunol-093019-125148

62. Chandra P, Grigsby SJ, Philips JA. Immune evasion and provocation by *Mycobacterium tuberculosis*. Nature Reviews Microbiology. 2022;20(12):750-766. https://doi.org/10.1038/s41579-022-00763-4

63. Chai Q, Lu Z, Liu CH. Host defense mechanisms against *Mycobacterium tuberculosis*. Cellular and Molecular Life Sciences. 2020;77(10):1859-1878. https://doi.org/10.1007/s00018-019-03353-5

64. Ahmad S. Pathogenesis, immunology, and diagnosis of latent *Mycobacterium tuberculosis* infection. Clinical Developmental Immunology. 2011;2011:814943. https://doi.org/10.1155/2011/814943

65. Kinsella RL, Zhu DX, Harrison GA, *et al.* Perspectives and advances in the understanding of tuberculosis. Annual Review of Pathology: Mechanisms of Disease. 2021;16(2021):377-408. https://doi.org/10.1146/annurev-pathol-042120-032916

66. Egen Jackson G, Rothfuchs Antonio G, Feng Carl G, Horwitz Marcus A, Sher A, Germain Ronald N. Intravital imaging reveals limited antigen presentation and t cell effector function in mycobacterial granulomas. Immunity. 2011;34(5):807-819. https://doi.org/10.1016/j.immuni.2011.03.022

67. Vergne I, Gilleron M, Nigou J. Manipulation of the endocytic pathway and phagocyte functions by *Mycobacterium tuberculosis* lipoarabinomannan. Frontiers in Cellular and Infection Microbiology. 2015;4(2014):187. https://doi.org/10.3389/fcimb.2014.00187

68. Krishnan V, Nath S, Nair P, Das B. *Mycobacterium tuberculosis* and its clever approaches to escape the deadly macrophage. World Journal of Microbiology and Biotechnology. 2023;39(11):300. https://doi.org/10.1007/s11274-023-03735-9

69. Nisa A, Kipper FC, Panigrahy D, Tiwari S, Kupz A, Subbian S. Different modalities of host cell death and their impact on *Mycobacterium tuberculosis* infection. American Journal of Physiology-Cell Physiology. 2022;323(5):C1444-C1474. https://doi.org/10.1152/ajpcell.00246.2022

70. Vinod V, Vijayrajratnam S, Vasudevan AK, Biswas R. The cell surface adhesins of *Mycobacterium tuberculosis*. Microbiological Research. 2020;232:126392. https://doi.org/10.1016/j.micres.2019.126392

71. Kline KA, Fälker S, Dahlberg S, Normark S, Henriques-Normark B. Bacterial adhesins in host-microbe interactions. Cell Host & Microbe. 2009;5(6):580-592. https://doi.org/10.1016/j.chom.2009.05.011

72. Bisht D, Meena LS. Adhesion molecules facilitate host-pathogen interaction & mediate *Mycobacterium tuberculosis* pathogenesis. Indian Journal of Medical Research. 2019;150(1):23-32. https://doi.org/10.4103/ijmr.IJMR_2055_16

73. Viljoen A, Dufrêne YF, Nigou J. Mycobacterial adhesion: from hydrophobic to receptor-ligand interactions. Microorganisms. 2022;10(2):454. https://doi.org/10.3390/microorganisms10020454

74. Kevlani N, Meena L. Prominent Role of FnBPs of *Mycobacterium tuberculosis* in cell adhesion, immune invasion and pathogenesis. Open Journal of Bacteriology. 2017;1(1):007-012. https://doi.org/10.17352/ojb.000002

75. Chakraborty P, Kumar A. The extracellular matrix of mycobacterial biofilms: could we shorten the treatment of mycobacterial infections? Microbial Cell. 2019;6(2):105-122. https://doi.org/10.15698/mic2019.02.667

76. Yang H, Ma Y, Lei X, *et al.* Stopping Tuberculosis at the Gate: The role of *M. tuberculosis* adhesins in infection and intervention. Vaccines. 2025; 13(7):676. https://doi.org/10.3390/vaccines13070676

77. Sigma-Aldrich. Extracellular Matrix Proteins and Tools for Cell Culture Optimization 2025. Available from: https://www.sigmaaldrich.com/ZA/en/technical-documents/technical-article/cell-culture-and-cell-culture-analysis/3d-cell-culture/attachment-factors-for-cell-culture?srsltid=AfmBOoqKDnr5njAvsTNh2N8b7lBRB1bgwIhs_yeMRu-UMRg3uuz3qtsd.

78. Halper J. Basic components of connective tissues and extracellular matrix: fibronectin, fibrinogen, laminin, elastin, fibrillins, fibulins, matrilins, tenascins and thrombospondins. In: Halper J, editor. Progress in heritable soft connective tissue diseases. Cham, Switzerland: Springer International Publishing; 2021. 105-126. https://doi.org/10.1007/978-3-030-80614-9_4

79. Mosher DF. Physiology of fibronectin. Annual Review of Medicine. 1984;35:561-575. https://doi.org/10.1146/annurev.me.35.020184.003021

80. Berdiaki A, Neagu M, Tzanakakis P, Spyridaki I, Pérez S, Nikitovic D. Extracellular matrix components and mechanosensing pathways in health and disease. Biomolecules. 2024;14(9):1186. https://doi.org/10.3390/biom14091186

81. Chagnot C, Listrat A, Astruc T, Desvaux M. Bacterial adhesion to animal tissues: protein determinants for recognition of extracellular matrix components. Cellular Microbiology. 2012;14(11):1687-1696. https://doi.org/10.1111/cmi.12002

82. Espitia C, Rodríguez E, Ramón-Luing L, Echeverría-Valencia G, Vallecillo AJ. Host-pathogen interactions in tuberculosis. understanding tuberculosis analyzing the origin of *Mycobacterium tuberculosis* pathogenicity 1st Ed InTech, Rijeka, Croatia. 2012:43-76. https://doi.org/10.5772/30310

83. Kuo C-J, Gao J, Huang J-W, *et al.* Functional and structural investigations of fibronectin-binding protein Apa from *Mycobacterium tuberculosis*. Biochimica et Biophysica Acta (BBA) - General Subjects. 2019;1863(9):1351-1359. https://doi.org/10.1016/j.bbagen.2019.06.003

84. Viljoen A, Alsteens D, Dufrêne Y. Mechanical forces between mycobacterial antigen 85 complex and fibronectin. Cells. 2020;9(3):716. https://doi.org/10.3390/cells9030716

85. Karamanos NK, Theocharis AD, Piperigkou Z, *et al.* A guide to the composition and functions of the extracellular matrix. The FEBS Journal. 2021;288(24):6850-6912. https://doi.org/10.1111/febs.15776

86. Kinhikar AG, Verma I, Chandra D, *et al.* Potential role for ESAT6 in dissemination of *M. tuberculosis* via human lung epithelial cells. Molecular Microbiology. 2010;75(1):92-106. https://doi.org/10.1111/j.1365-2958.2009.06959.x

87. Alteri CJ, Xicohténcatl-Cortes J, Hess S, Caballero-Olín G, Girón JA, Friedman RL. *Mycobacterium tuberculosis* produces pili during human infection. Proceedings of the National Academy of Sciences. 2007;104(12):5145-5150. https:doi.org/10.1073/pnas.0602304104

88. Owczarzy A, Kurasiński R, Kulig K, Rogóż W, Szkudlarek A, Maciążek-Jurczyk M. Collagen - structure, properties and application. Engineering of Biomaterials. 2020(156):17-23. https://doi.org/10.34821/eng.biomat.156.2020.17-23

89. Zhao C, Xiao Y, Ling S, Pei Y, Ren J. Structure of Collagen. In: Ling S, editor. Fibrous proteins: design, synthesis, and assembly. New York, USA: Springer US; 2021. 17-25.

90. Makareeva E, Leikin S. Chapter 7 - Collagen structure, folding and function. In: Shapiro JR, Byers PH, Glorieux FH, Sponseller PD, editors. Osteogenesis Imperfecta. San Diego: Academic Press; 2014. 71-84. https://doi.org/10.1016/B978-0-12-397165-4.00007-1

91. Kong W, Lyu C, Liao H, Du Y. Collagen crosslinking: effect on structure, mechanics and fibrosis progression. Biomedical Materials. 2021;16(6):062005. https://doi.org/10.1088/1748-605X/ac2b79

92. Squeglia F, Ruggiero A, Berisio R. Collagen degradation in tuberculosis pathogenesis: the biochemical consequences of hosting an undesired guest. Biochemical Journal. 2018;475(19):3123-3140. https://doi.org/10.1042/bcj20180482

93. González-Avila G, Sandoval C, Herrera MT, *et al. Mycobacterium tuberculosis* effects on fibroblast collagen metabolism. Respiration. 2008;77(2):195-202. https://doi.org/10.1159/000163064

94. Al Shammari B, Shiomi T, Tezera L, *et al.* The extracellular matrix regulates granuloma necrosis in tuberculosis. The Journal of Infectious Diseases. 2015;212(3):463-473. https://doi.org/10.1093/infdis/jiv076

95. Monfeli RR, Beeson C. Targeting iron acquisition by *Mycobacterium tuberculosis*. Infectious Disorders - Drug Targets. 2007;7(3):213-220. https://doi.org/10.2174/187152607782110031

96. Ramsugit S, Pillay M. Identification of *Mycobacterium tuberculosis* adherence-mediating components: a review of key methods to confirm adhesin function. Iran Journal of Basic Medical Science. 2016;19(6):579-584. https://dx.doi.org/10.22038/ijbms.2016.7124

97. Noens EE, Williams C, Anandhakrishnan M, Poulsen C, Ehebauer MT, Wilmanns M. Improved mycobacterial protein production using a *Mycobacterium smegmatis* groEL1ΔC expression strain. BMC Biotechnology. 2011;11(1):27. https://doi.org/10.1186/1472-6750-11-27

98. Brülle JK, Tschumi A, Sander P. Lipoproteins of slow-growing mycobacteria carry three fatty acids and are N-acylated by apolipoprotein N-acyltransferase BCG_2070c. BMC Microbiology. 2013;13(1):223. https://doi.org/10.1186/1471-2180-13-223

99. Zhao D, Lin D, Xu C. A protein fragment of Rv3194c located on mycobacterial cell surface efficiently prevents adhesion of recombinant *Mycobacterium smegmatis*, and promises a new anti-adhesive drug. Microbial Pathogenesis. 2020;149:104498. 10.1016/j.micpath.2020.104498

100. Xolalpa W, Vallecillo AJ, Lara M, *et al.* Identification of novel bacterial plasminogen-binding proteins in the human pathogen. Proteomics. 2007;7(18):3332-3341. https://doi.org/10.1002/pmic.200600876

101. Kaufmann SHE, Weiner J, von Reyn CF. Novel approaches to tuberculosis vaccine development. International Journal of Infectious Diseases. 2017;56:263-267. https://doi.org/10.1016/j.ijid.2016.10.018

102. Kim H, Choi H-G, Shin SJ. Bridging the gaps to overcome major hurdles in the development of next-generation tuberculosis vaccines. Frontiers in Immunology. 2023;14:2023. https://doi.org/10.3389/fimmu.2023.1193058

103. Diaz-Silvestre H, Espinosa-Cueto P, Sanchez-Gonzalez A, *et al.* The 19-kDa antigen of *Mycobacterium tuberculosis* is a major adhesin that binds the mannose receptor of THP-1 monocytic cells and promotes phagocytosis of mycobacteria. Microbial Pathogenesis. 2005;39(3):97-107. https://doi.org/10.1016/j.micpath.2005.06.002

104. Esparza M, Palomares B, García T, Espinosa P, Zenteno E, Mancilla R. PstS-1, the 38-kDa *Mycobacterium tuberculosis* glycoprotein, is an adhesin, which binds the macrophage mannose receptor and promotes phagocytosis. Scandinavian Journal of Immunology. 2015;81(1):46-55. https://doi.org/10.1111/sji.12249

105. Hickey TBM, Thorson LM, Speert DP, Daffé M, Stokes RW. *Mycobacterium tuberculosis* Cpn60.2 and DnaK are located on the bacterial surface, where Cpn60.2 facilitates efficient

bacterial association with macrophages. Infection and Immunity. 2009;77(8):3389-3401. https://doi.org/10.1128/iai.00143-09

106. Ragas A, Roussel L, Puzo G, Rivière M. The *Mycobacterium tuberculosis* cell-surface glycoprotein apa as a potential adhesin to colonize target cells via the innate immune system pulmonary c-type lectin surfactant protein a. Journal of Biological Chemistry. 2007;282(8):5133-5142. https://doi.org/10.1074/jbc.M610183200

107. Espitia C, Laclette JP, Mondragón- Palomino M, *et al.* The PE-PGRS glycine-rich proteins of *Mycobacterium tuberculosis* : a new family of fibronectin-binding proteins?The GenBank accession number for the sequence reported in this paper is AF071081. Microbiology. 1999;145(12):3487-3495. https://doi.org/10.1099/00221287-145-12-3487

108. Ronning DR, Vissa V, Besra GS, Belisle JT, Sacchettini JC. *Mycobacterium tuberculosis* antigen 85A and 85C structures confirm binding orientation and conserved substrate specificity. Journal of Biological Chemistry. 2004;279(35):36771-36777. https://doi.org/10.1074/jbc.M400811200

109. Puckett S, Trujillo C, Wang Z, *et al.* Glyoxylate detoxification is an essential function of malate synthase required for carbon assimilation in *Mycobacterium tuberculosis*. Proceedings of the National Academy of Sciences. 2017;114(11):E2225-E2232. https://doi.org/10.1073/pnas.1617655114

110. Kumar S, Puniya BL, Parween S, Nahar P, Ramachandran S. Identification of novel adhesins of *m. tuberculosis* h37rv using integrated approach of multiple computational algorithms and experimental analysis. PLOS ONE. 2013;8(7):e69790. https://doi.org/10.1371/journal.pone.0069790

111. Bhowmick T, Ghosh S, Dixit K, *et al.* Targeting *Mycobacterium tuberculosis* nucleoid-associated protein HU with structure-based inhibitors. Nature Communications. 2014;5(1):4124. https://doi.org/10.1038/ncomms5124

112. Be NA, Bishai WR, Jain SK. Role of *Mycobacterium tuberculosis* pknD in the pathogenesis of central nervous system tuberculosis. BMC Microbiology. 2012;12(1):1-7. https://doi.org/10.1186/1471-2180-12-7

113. Smith CV, Huang C-c, Miczak A, Russell DG, Sacchettini JC, Höner zu Bentrup K. Biochemical and structural studies of malate synthase from *Mycobacterium tuberculosis*. Journal of Biological Chemistry. 2003;278(3):1735-1743. https://doi.org/10.1074/jbc.M209248200

114. Kinhikar AG, Vargas D, Li H, *et al. Mycobacterium tuberculosis* malate synthase is a laminin-binding adhesin. Molecular Microbiology. 2006;60(4):999-1013. https://doi.org/10.1111/j.1365-2958.2006.05151.x

115. Naidoo N, Ramsugit S, Pillay M. *Mycobacterium tuberculosis* pili (MTP), a putative biomarker for a tuberculosis diagnostic test. Tuberculosis. 2014;94(3):338-345. https://doi.org/10.1016/j.tube.2014.03.004

116. Chitale S, Ehrt S, Kawamura I, *et al.* Recombinant *Mycobacterium tuberculosis* protein associated with mammalian cell entry. Cellular Microbiology. 2001;3(4):247-254. https://doi.org/10.1046/j.1462-5822.2001.00110.x

117. Esposito C, Cantisani M, D'Auria G, *et al.* Mapping key interactions in the dimerization process of HBHA from *Mycobacterium tuberculosis*, insights into bacterial agglutination. FEBS Letters. 2012;586(6):659-667. https://doi.org/10.1016/j.febslet.2012.01.047

118. Pethe K, Alonso S, Biet F, *et al.* The heparin-binding haemagglutinin of *M. tuberculosis* is required for extrapulmonary dissemination. Nature. 2001;412(6843):190-194. https://doi.org/10.1038/35084083

119. Migliori GB, Ong CWM, Petrone L, D'Ambrosio L, Centis R, Goletti D. The definition of tuberculosis infection based on the spectrum of tuberculosis disease. Breathe. 2021;17(3):210079. https://doi.org/10.1183/20734735.0079-2021

120. Vasiliu A, Martinez L, Gupta RK, *et al.* Tuberculosis prevention: current strategies and future directions. Clinical Microbiology and Infection. 2024;30(9):1123-1130. https://doi.org/10.1016/j.cmi.2023.10.023

121. Kaufmann SH, Lange C, Rao M, *et al.* Progress in tuberculosis vaccine development and host-directed therapies—a state of the art review. The Lancet. 2014;2(4):301-320. https://doi.org/10.1016/s2213-2600(14)70033-5

122. Fatima S, Kumari A, Das G, Dwivedi VP. Tuberculosis vaccine: A journey from BCG to present. Life Sciences. 2020;252:117594. https://doi.org/10.1016/j.lfs.2020.117594

123. Starshinova A, Kudryavtsev I, Rubinstein A, *et al.* BCG vaccination: historical role, modern applications, and future perspectives in tuberculosis and beyond. Frontiers in Pediatrics. 2025;13(1). https://doi.org/10.3389/fped.2025.1603732

124. Jones CE. Routine vaccination coverage—worldwide, 2023. Morbidity and Mortality Weekly Report. 2024;73:978–984. http://dx.doi.org/10.15585/mmwr.mm7343a4

125. Messina NL, Germano S, Chung AW, *et al.* Effect of Bacille Calmette–Guérin vaccination on immune responses to SARS-CoV-2 and COVID-19 vaccination. Clinical & Translational Immunology. 2025;14(1):e70023. https://doi.org/10.1002/cti2.70023

126. Stop TB Partnership. TB vaccine clinical pipeline 2026 [accessed 2026 February 14]. Available from: https://newtbvaccines.org/tb-vaccine-pipeline/clinical-phase/.

127. Martin C, Aguilo N, Marinova D, Gonzalo-Asensio J. Update on TB vaccine pipeline. Applied Sciences. 2020;10(7):2632. https://doi.org/10.3390/app10072632128.

128. Tait DR, Hatherill M, Meeren OVD, *et al.* Final Analysis of a Trial of M72/AS01E Vaccine to Prevent Tuberculosis. New England Journal of Medicine. 2019;381(25):2429-2439. https://doi.org/10.1056/NEJMoa1909953

129. Mahmoudi S, Pourakbari B, Mamishi S. Immunodiagnostics of tuberculosis: recent discoveries. In: Rezaei N, editor. Tuberculosis: integrated studies for a complex disease. Cham, Switzerland: Springer International Publishing; 2023. p. 139-150. https://doi.org/10.1007/978-3-031-15955-8_8

130. Yong YK, Tan HY, Saeidi A, *et al.* Immune biomarkers for diagnosis and treatment monitoring of tuberculosis: current developments and future prospects. Frontiers in Microbiology. 2019;10(2019): 2789. https://doi.org/10.3389/fmicb.2019.02789

131. Panda S, Kearns K, Cheng C, Lindestam Arlehamn CS. From antigens to immune responses: shaping the future of TB detection and prevention. International Journal of Infectious Diseases. 2024;141:106983. https://doi.org/10.1016/j.ijid.2024.106983

132. Luo L, Zhu L, Yue J, *et al.* Antigens Rv0310c and Rv1255c are promising novel biomarkers for the diagnosis of *Mycobacterium tuberculosis* infection. Emerging Microbes & Infections. 2017;6(1):1-8. https://doi.org/10.1038/emi.2017.54

133. Kumar G, Dagur PK, Singh PK, *et al.* Serodiagnostic efficacy of *Mycobacterium tuberculosis* 30/32-kDa mycolyl transferase complex, esat-6, and cfp-10 in patients with active tuberculosis. Archives of Immunology and Experimental Therapy. 2010;58(1):57-65. https://doi.org/10.1007/s00005-009-0055-4

134. Wu X, Yang Y, Zhang J, *et al.* Humoral Immune responses against the *Mycobacterium tuberculosis* 38-kilodalton, mtb48, and cfp-10/esat-6 antigens in tuberculosis. Clinical and Vaccine Immunology. 2010;17(3):372-375. https://doi.org/10.1128/CVI.00287-09

135. Lyashchenko K, Colangeli R, Houde M, Jahdali HA, Menzies D, Gennaro ML. Heterogeneous antibody responses in tuberculosis. Infection and Immunity. 1998;66(8):3936-3940. https://doi.org/10.1128/iai.66.8.3936-3940.1998

136. Achkar J, M., Dong Y, Holzman Robert S, *et al. Mycobacterium tuberculosis* malate synthase- and mpt51-based serodiagnostic assay as an adjunct to rapid identification of pulmonary tuberculosis. Clinical and Vaccine Immunology. 2006;13(11):1291-1293. https://doi.org/10.1128/CVI.00158-06

137. Arif S, Akhter M, Khaliq A, Nisa Zu, Khan IH, Akhtar MW. Serodiagnostic evaluation of fusion proteins from multiple antigens of *Mycobacterium tuberculosis* for active TB. Tuberculosis. 2021;127:102053. https://doi.org/10.1016/j.tube.2021.102053

138. Brust B, Lecoufle M, Tuaillon E, *et al. Mycobacterium tuberculosis* lipolytic enzymes as potential biomarkers for the diagnosis of active tuberculosis. PLOS ONE. 2011;6(9):e25078. https://doi.org/10.1371/journal.pone.0025078

139. McIntyre S, Warner J, Rush C, Vanderven HA. Antibodies as clinical tools for tuberculosis. Frontiers in Immunology. 2023;14(2023): 8947. https://doi.org/10.3389/fimmu.2023.1278947

140. Khan MM, Ernst O, Manes NP, *et al.* Multi-Omics Strategies Uncover Host–Pathogen Interactions. ACS Infectious Diseases. 2019;5(4):493-505. https://doi.org/10.1021/acsinfecdis.9b00080

141. Barazani O, Erdos T, Chowdhury R, Kaur G, Venketaraman V. New advances in the development and design of *mycobacterium tuberculosis* vaccines: construction and validation of multi-epitope vaccines for tuberculosis prevention. Biology. 2025;14(4):417. https://doi.org/10.3390/biology14040417

142. Vincent AT. Bacterial hypothetical proteins may be of functional interest. Frontiers in Bacteriology. 2024;3(2024). https://doi.org/10.3389/fbrio.2024.1334712

143. Ansari FA, Kumar N, Bala Subramanyam M, Gnanamani M, Ramachandran S. MAAP: malarial adhesins and adhesin-like proteins predictor. proteins: structure, function, and bioinformatics. 2008;70(3):659-666. https://doi.org/10.1002/prot.21568

144. Chaudhuri R, Ansari FA, Raghunandanan MV, Ramachandran S. FungalRV: adhesin prediction and immunoinformatics portal for human fungal pathogens. BMC Genomics. 2011;12(1):192. https://doi.org/10.1186/1471-2164-12-192

145. Ramana J, Gupta D. FaaPred: a SVM-based prediction method for fungal adhesins and adhesin-like proteins. PLOS ONE. 2010;5(3):e9695. https://doi.org/10.1371/journal.pone.0009695

146. Sidorczuk K, Burdukiewicz M, Cerk K, *et al.* adhesiomeR: a tool for *Escherichia coli* adhesin classification and analysis. BMC Genomics. 2024;25(1):609. https://doi.org/10.1186/s12864-024-10525-6

147. Sachdeva G, Kumar K, Jain P, Ramachandran S. SPAAN: a software program for prediction of adhesins and adhesin-like proteins using neural networks. Bioinformatics. 2005;21(4):483-491. https://doi.org/10.1093/bioinformatics/bti028

148. Jayashree S, Karthikeyan R, Nithyalakshmi S, Ranjani J, Gunasekaran P, Rajendhran J. Anti-adhesion property of the potential probiotic strain lactobacillus fermentum 8711 against methicillin-resistant staphylococcus aureus (MRSA). Frontiers in Microbiology. 2018;9(2018): 0411. https://doi.org/10.3389/fmicb.2018.00411

149. Kamaruzaman INA, Staton GJ, Ainsworth S, Carter SD, Evans NJ. Characterisation of putative outer membrane proteins from *Leptospira borgpetersenii* serovar hardjo-bovis identifies novel adhesins and diversity in adhesion across genomospecies orthologs. Microorganisms. 2024;12(2):245. https://doi.org/10.3390/microorganisms12020245

150. Kapoor J, Panda A, Naqvi I, Ganta S, Kumar S, Bandyopadhyay A. Identification and characterization of outer membrane proteins and membrane spanning protein complexes in *Brucella melitensis*. Proteins: Structure, Function, and Bioinformatics. 2026;n/a(n/a). https://doi.org/10.1002/prot.70118

151. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols. 2015;10(6):845-858. https://doi.org/10.1038/nprot.2015.053

152. Mistry J, Chuguransky S, Williams L, *et al.* Pfam: The protein families database in 2021. Nucleic Acids Research. 2021;49(D1):D412-D419. https://doi.org/10.1093/nar/gkaa913.

153. Paysan-Lafosse T, Andreeva A, Blum M, *et al.* (2025) The Pfam protein families database: embracing AI/ML. Nucleic Acids Research. 2025; 53, D523-D534. https://doi.org/10.1093/nar/gkae997

154. Babic N, Kovacic F. Predicting drug targets by homology modelling of *Pseudomonas aeruginosa* proteins of unknown function. PLOS ONE. 2021;16(10):e0258385. https://doi.org/10.1371/journal.pone.0258385

155. Heiniger B, Schori C, Arefian M, *et al.* Proteogenomic discovery of novel small proteins in clinical *Mycobacterium tuberculosis* strains. BioRxiv. 2026; 2026.01.27.701740. https://doi.org/10.64898/2026.01.27.701740

156. Zhu P-P, Li W-C, Zhong Z-J, *et al.* Predicting the subcellular localization of mycobacterial proteins by incorporating the optimal tripeptides into the general form of pseudo amino acid composition. Molecular BioSystems. 2014;11(2):558-563. https://doi.org/10.1039/c4mb00645c

157. Chiliza TE, Pillay M, Pillay B. Identification of unique essential proteins from a *Mycobacterium tuberculosis* F15/LAM4/KZN phage secretome library. Pathogens and disease. 2017;75(1):ftx001. https://doi.org/10.1093/femspd/ftx001

158. Chaudhuri R, Kulshreshtha D, Raghunandanan MV, Ramachandran S. Integrative immunoinformatics for Mycobacterial diseases in R platform. Systems and Synthetic Biology. 2014;8(1):27-39. https://doi.org/10.1007/s11693-014-9135-9

159. Cornejo-Granados F, Zatarain-Barrón ZL, Cantu-Robles VA, *et al.* Secretome prediction of two *m. tuberculosis* clinical isolates reveals their high antigenic density and potential drug targets. Frontiers in Microbiology. 2017;8:128. https://doi.org/10.3389/fmicb.2017.00128

160. Prieto-Martínez FD, Arciniega M, Medina-Franco JL. Molecular Docking: Current advances and challenges. TIP Revista Especializada en Ciencias Químico-Biológicas. 2018;21:65-87. https://doi.org/10.22201/fesz.23958723e.2018.0.143

161. Pagadala NS, Syed K, Tuszynski J. Software for molecular docking: a review. Biophysical Reviews. 2017;9(2):91-102. https://doi.org/10.1007/s12551-016-0247-1

162. Gordon CH, Hendrix E, He Y, Walker MC. AlphaFold accurately predicts the structure of ribosomally synthesized and post-translationally modified peptide biosynthetic enzymes. Biomolecules. 2023;13(8):1243. https://doi.org/10.3390/biom13081243

163. Zhou J, Chen J, Peng Y, Xie Y, Xiao Y. A promising tool in serological diagnosis: current research progress of antigenic epitopes in infectious diseases. Pathogens. 2022;11(10):1095. https://doi.org/10.3390/pathogens11101095

164. Meng EC, Goddard TD, Pettersen EF, *et al.* UCSF ChimeraX: Tools for structure building and analysis. Protein Science. 2023;32(11):e4792. https://doi.org/10.1002/pro.4792

165. Kuo C-J, Ptak CP, Hsieh C-L, Akey BL, Chang Y-F. Elastin, a novel extracellular matrix protein adhering to mycobacterial antigen 85 complex. Journal of Biological Chemistry. 2013;288(6):3886-3896. https://doi.org/10.1074/jbc.M112.415679

166. Choudhary S, Kesavan AK, Juneja V, Thakur S. Molecular modeling, simulation and docking of Rv1250 protein from *Mycobacterium tuberculosis*. Frontiers in Bioinformatics. 2023;3(2023):5479. https://doi.org/10.3389/fbinf.2023.1125479

167. Fakhar Z, Naiker S, Alves CN, *et al.* A comparative modeling and molecular docking study on *Mycobacterium tuberculosis* targets involved in peptidoglycan biosynthesis. Journal of Biomolecular Structure and Dynamics. 2016;34(11):2399-2417. https://doi.org/10.1080/07391102.2015.1117397

168. Das N, Jena PK, Pradhan SK. Arabinosyltransferase C enzyme of *Mycobacterium tuberculosis*, a potential drug target: an insight from a molecular docking study. Heliyon. 2020;6(2). https://doi.org/10.1016/j.heliyon.2019.e02693

169. Yang H, Lei X, Chai S, Zhang S, Su G, Du L. A multi-epitope vaccine incorporating adhesin-derived antigens protects against *Mycobacterium tuberculosis* infection and dissemination. Frontiers in Immunology. 2025;16:1707471. https://doi.org/10.3389/fimmu.2025.1707471

170. Pillay K, Chiliza TE, Senzani S, Pillay B, Pillay M. In silico design of *Mycobacterium tuberculosis* multi-epitope adhesin protein vaccines. Heliyon. 2024;10(18). https://doi.org/10.2139/ssrn.4771732

171. Kapopoulou A, Lew JM, Cole ST. The MycoBrowser portal: A comprehensive and manually annotated resource for mycobacterial genomes. Tuberculosis. 2011;91(1):8-13. https://doi.org/10.1016/j.tube.2010.09.006

172. Skeiky Yasir AW, Lodes Michael J, Guderian Jeffrey A, *et al.* Cloning, expression, and immunological evaluation of two putative secreted serine protease antigens of *Mycobacterium tuberculosis.* Infection and Immunity. 1999;67(8):3998-4007. https://doi.org/10.1128/iai.67.8.3998-4007.1999

173. Basu C. PCR primer design: Springer; 2015. P276. https://doi.org/10.1007/978-1-0716-1799-1

174. Wong TS, Tee KL. Sequence analysis. A Practical Guide to Protein Engineering: Cham, Switzerland: Springer International Publishing; 2020. 11-27. https://doi.org/10.1007/978-3-030-56898-6

175. Kulkarni R, Bose R, Bose K. Selection of cloning and expression plasmid vectors. in: bose k, editor. textbook on cloning, expression and purification of recombinant proteins. Singapore: Springer Nature Singapore; 2022. 57-81. https://doi.org/10.1007/978-981-16-4987-5_3

176. Greenfield EA, DeCaprio J, Brahmandam M. Preparing gst-, his-, or mbp-fusion proteins from bacteria. Cold Spring Harbor Protocols. 2020;2020(9). https://doi.org/10.1101/pdb.prot100024

177. Banerjee R, Rudra P, Prajapati RK, Sengupta S, Mukhopadhyay J. Optimization of recombinant *Mycobacterium tuberculosis* RNA polymerase expression and purification. Tuberculosis. 2014;94(4):397-404. https://doi.org/10.1016/j.tube.2014.03.008

178. Najafi A, Tafaghodi M, Sankian M, Amini Y, Ghazvini K. Cloning, expression, and refolding of ppe17 protein of *mycobacterium tuberculosis* as a promising vaccine candidate. Iranian Journal of Medical Sciences. 2019;44(1):53-59. https://doi.org/10.30476/ijms.2019.40625

179. Goel N, Singh D, Bandhu A. Molecular cloning, expression, purification and functional characterization of Rv1588c of *Mycobacterium tuberculosis*. Biocatalysis and Agricultural Biotechnology. 2024;58:103239. https://doi.org/10.1016/j.bcab.2024.103239

180. Mustafa AD, Kalyanasundram J, Sabidi S, *et al.* Recovery of recombinant *Mycobacterium tuberculosis* antigens fused with cell wall-anchoring motif (LysM) from inclusion bodies using non-denaturing reagent (N-laurylsarcosine). BMC Biotechnology. 2019;19(1):27. https://doi.org/10.1186/s12896-019-0522-x

181. Gil-Garcia M, Navarro S, Ventura S. Coiled-coil inspired functional inclusion bodies. Microbial Cell Factories. 2020;19(1):117. https://doi.org/10.1186/s12934-020-01375-4

182. Piubelli L, Campa M, Temporini C, *et al.* Optimizing *Escherichia coli* as a protein expression platform to produce *Mycobacterium tuberculosis* immunogenic proteins. Microbial Cell Factories. 2013;12(1):115. https://doi.org/10.1186/1475-2859-12-115

183. Arakawa T, Akuta T, Ejima D, Tsumoto K. Solubilization and refolding of inclusion bodies by detergents. Protein Expression and Purification. 2025;236:106791. https://doi.org/10.1016/j.pep.2025.106791

184. Schlager B, Straessle A, Hafen E. Use of anionic denaturing detergents to purify insoluble proteins after overexpression. BMC Biotechnology. 2012;12(1):95. https://doi.org/10.1186/1472-6750-12-95

185. Arakawa T, Niikura T, Kita Y, Akuta T. Sodium dodecyl sulfate analogs as a potential molecular biology reagent. Current Issues in Molecular Biology. 2024;46(1):621-633. https://doi.org/10.3390/cimb46010040

186. Massiah MA, Wright KM, Du H. Obtaining soluble folded proteins from inclusion bodies using sarkosyl, triton x-100, and chaps: application to lb and m9 minimal media. Current Protocols in Protein Science. 2016;84(1):6.13.11-16.13.24. https://doi.org/10.1002/0471140864.ps0613s84

187. Peternel Š, Grdadolnik J, Gaberc-Porekar V, Komel R. Engineering inclusion bodies for non-denaturing extraction of functional proteins. Microbial Cell Factories. 2008;7(1):34. https://doi.org/10.1186/1475-2859-7-34

Chapter 3: Identification of Adhesin-related Conserved Hypothetical Proteins of *Mycobacterium tuberculosis*

This Chapter focused on the computational pipeline that sought to identify functional characteristics associated with enzymatic and adhesin-related properties of the conserved hypothetical classification of *Mycobacterium tuberculosis*. This study was published in the Expert Review of Proteomics (Taylor and Francis online, Impact Factor: 3.5) accepted on the 20th of October 2023.

Title: A computational method for the prediction and functional analysis of potential *mycobacterium tuberculosis* adhesin-related proteins

*Rivesh Maharajh, Manormoney Pillay and Sibusiso Senzani**

Discipline of Medical Microbiology, School of Laboratory Medicine and Medical Sciences, College of Health Sciences, University of KwaZulu-Natal, Durban 4013, South Africa

***Corresponding Author: Dr. S Senzani (PhD), Discipline of Medical Microbiology, University of KwaZulu-Natal, Durban, 4000, South Africa, Email address: SenzaniS@ukzn.ac.za**

3.1. Abstract

Objectives

Mycobacterial adherence plays a major role in the establishment of infection within the host. Adhesin-related proteins attach to host receptors and cell-surface components. The current study aimed to utilize *in-silico* strategies to determine the adhesin potential of conserved hypothetical (CH) proteins.

Methods

Computational analysis was performed on the whole *Mycobacterium tuberculosis* H37Rv proteome using a software program for the prediction of adhesin and adhesin-like proteins using neural networks (SPAAN) to determine the adhesin potential of CH proteins. A robust pipeline of computational analysis tools: Phyre2 and Pfam for homology prediction; Mycosub, PsortB, and Loctree3 for subcellular localization; SignalP-5.0 and SecretomeP-2.0 for secretory prediction, were utilized to identify adhesin candidates.

Results

SPAAN revealed 776 potential adhesins within the whole *M. tuberculosis* H37Rv proteome. Comprehensive analysis of the literature was cross tabulated with SPAAN to verify the adhesin prediction potential of known adhesin (n=34). However, approximately a third of known adhesins were below the probability of adhesin (P_{ad}) threshold ($P_{ad} \geq 0.51$). Subsequently, 167 CH proteins of interest were categorized using essential *in-silico* tools.

Conclusion

The use of SPAAN with supporting *in-silico* tools should be fundamental when identifying novel adhesins. This study provides a pipeline to identify CH proteins as functional adhesin molecules.

KEYWORDS: Bioinformatic analysis, Adhesion molecules, Mycobacterial, tuberculosis, Protein analysis, SPAAN tool

- Bacterial cell adhesion molecules known as adhesins are cell surface receptor proteins that facilitate host-pathogen interactions, which enhance pathogenesis in the host.
- Identification of novel *Mycobacterium tuberculosis* adhesins may provide pipeline for vaccine and serological biomarkers.
- This study focused on characterizing conserved hypothetical (CH) proteins in the *M. tuberculosis* proteome to ascertain adhesin role using a bioinformatic analysis approach.
- SPAAN analysis of the whole *M. tuberculosis* H37Rv genome identified 167 CH genes as potential adhesin/adhesin-like proteins.
- Further stringent *in-silico* analysis determined homology protein classification, subcellular localization, and translocation of individual CH proteins.
- Herein this study, a dataset was generated with comparison with known *M.*

3.2. Introduction

Understanding the link between bacterial pathogenic strategies and host immune responses is a critical aspect of identifying biomarkers for diagnostic applications and therapeutic development [1,2]. Commonly, pulmonary *Mycobacterium tuberculosis* infection commences when aerosolized bacilli enter the lungs [3]. The survival of *M. tuberculosis* entails interaction with a variety of host surface molecules to invade alveolar macrophages and inhibit phagolysosomal fusion while replicating within the host cells [3,4].

The strategic entry of *M. tuberculosis* into host alveolar macrophage involves the adherence to eukaryotic cellular receptors (T-lymphocyte receptors, C-type lectins, cluster of differentiation 43, heparin sulphate and mannose receptors) and extracellular matrix (ECM) components (fibronectin, collagen, elastin, laminin, thrombospondin, bone sialoprotein fibrillin and vitronectin) via mycobacterial adhesion specific receptors [4,5]. Cell adhesion molecules, known as adhesins, are receptor-bound proteins produced on the bacterial cell surface, facilitating host-pathogen interactions that are known to further enhance pathogenesis [6,7]. *Mycobacterium tuberculosis* specific adhesin molecules have the potential to be used as biomarkers in an array of applications because of their unique proteomic composition differing from eukaryotic

components. Squeglia and colleagues validated a range of *M. tuberculosis* adhesins and highlighted the versatility of these adhesion molecules to bind to specific ECM receptors [1]. Ideally, vaccine candidates and serological biomarkers are capable of eliciting host cell-mediated immune response due to cell surface exposure and inducing antibody production, which enhances the potential of bacterial adhesin molecules for such applications [8,9]. The potential use of anti-adhesin antibodies in serological diagnostics to identify antigenic proteins in tuberculosis infected patients can improve clinical identification of initial infections [1,10]. This suggests a growing interest in identifying novel adhesin proteins from a large selection of the mycobacterial proteome.

With the advancement of bioinformatic analytics, algorithms are used to identify compositional computational properties in the form of the arrangement of amino acid residue sequences to designate functional association and the characterization of proteins [11]. One such non-homologous bioinformatic tool is SPAAN (a software program for the prediction of adhesin and adhesin-like proteins using neural networks), which predicts adhesin and adhesin-like proteins from the bacterial proteome sequence [12]. The SPAAN tool, which has the potential to reduce exorbitant costs related to time-consuming experimental assays, was designed to identify adhesin and adhesin-like genes based on 105 compositional attributes using an artificial neural network system [12]. The computational arrangement of these attributes was reported to be based on five elements: amino acid frequencies, multiplet frequencies, dipeptide frequencies, charge, and hydrophobic compositions respectively [12]. Since the inception of the Linux-based adhesin prediction tool, SPAAN has been used to elucidate various potential adhesin proteins within the *M. tuberculosis* genome [2,13]. The *M. tuberculosis* H37Rv mycobacterial genome contains more than 4000 genes, of which approximately 1051 were coded as conserved hypothetical (CH) proteins that are functionally uncategorized [14]. A previous study by Kumar *et al.* (2013) investigated a small collection of genes that were evaluated for adhesin potential. Therefore there is a need to determine the functional characteristics of novel *M. tuberculosis* proteins [13]. The current study utilized the SPAAN adhesin prediction tool to identify potential adhesin and adhesin-like proteins within the *M. tuberculosis* proteome. Further classification of the CH proteins using recent in silico strategies was used to characterize these potential adhesins. This was used to identify an array of novel adhesin proteins that have experimental potential for vaccine and diagnostic candidacy.

3.3. Materials and Methods

3.3.1. Sequence retrieval (mapping out *Mycobacterium tuberculosis* H37Rv genome)

A comprehensive annotation of the complete *M. tuberculosis* H37Rv (strain ATCC 25618) proteomic reference data was retrieved from the Universal Protein Resource database (www.uniprot.org). Mycobrowser (https://mycobrowser.epfl.ch/) was used to analyze the genomic data by identifying sequence-specific domains that functionally characterized potential adhesin gene loci. Furthermore, the proteomic data was extracted for comparison analysis and predicted functionality of the selected gene profiles.

3.3.2 Adhesin-specific computational analysis

The probability of adhesin (P_{ad}) was predicted using the currently available SPAAN software (http://sourceforge.net/projects/adhesin/files/) [2,12]. A total of 3993 genes retrieved from the UniProt reference proteome (ATCC25618/*H37Rv*) were analyzed on SPAAN output (figure 3.1). The Probability of adhesin threshold used for adhesin identification included proteins of P_{ad} ≥0.510, which was determined to have an optimal 89% sensitivity and 100% specificity for adhesin prediction delineated by Sachdeva and colleagues [12]. The recommended criterion (P_{ad} ≥0.7) for whole genome analysis [12] was not suitable for the *M. tuberculosis* H37Rv proteome, as a number of experimentally characterized adhesins would be omitted such as HBHA, MTP, Hup, LpdC, Apa, PknD, type IV pili and Icl (see B1: supplementary table 1). This provided an inclusive threshold rather than that of HBHA only (P_{ad} ≥0.65) as indicated by Kumar *et al.* [13]. Furthermore, Kumar and colleagues [13] analyzed globular proteins that could be solubilized for protein purification. In contrast, our study included every protein structural conformation within

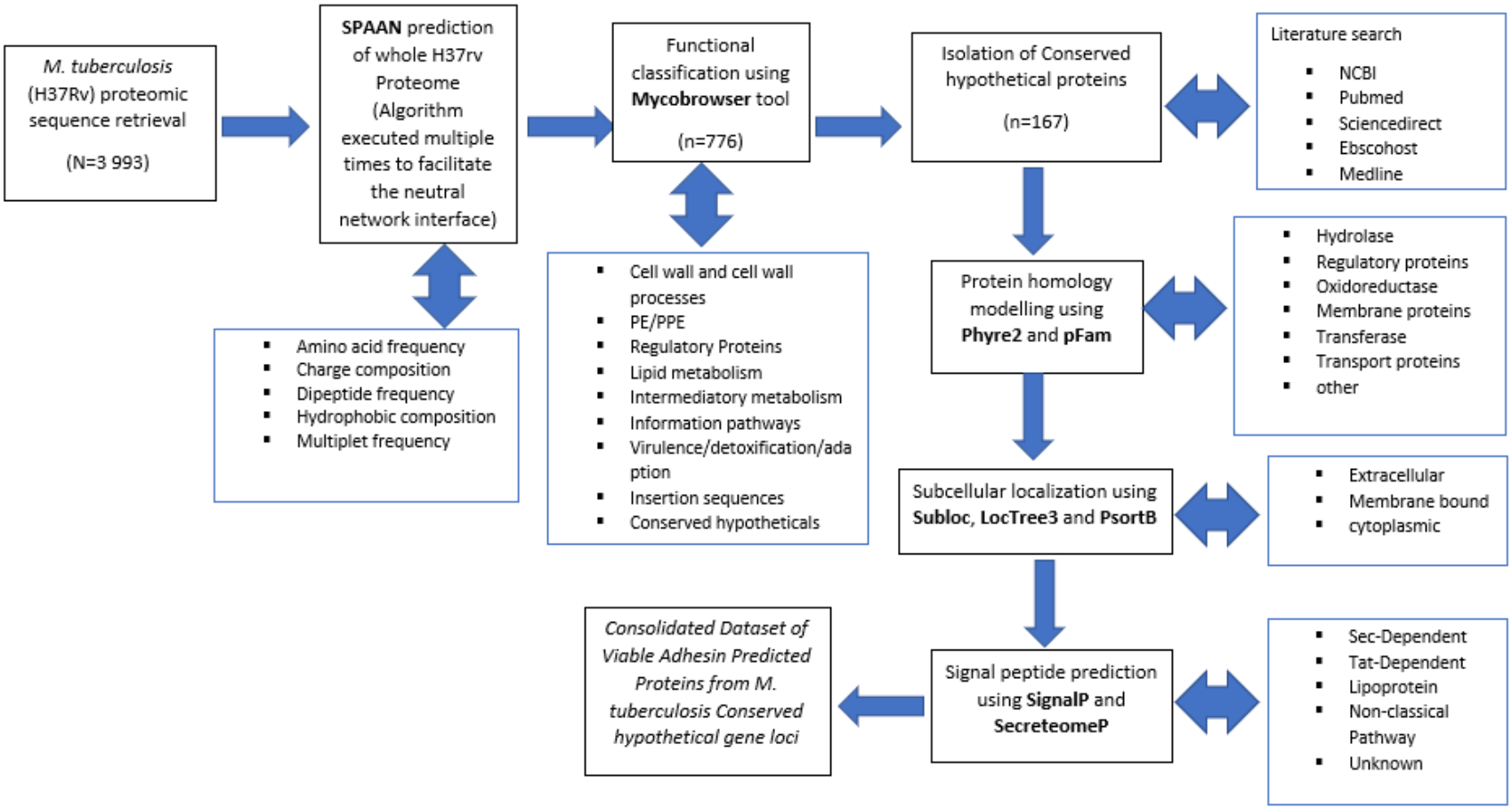


Figure 3.1. Flow diagram of the selection of conserved hypothetical proteins predicted by computational analysis.

the *M. tuberculosis* genome to identify other adhesin proteins. Based on the P_{ad} inclusion criterion, a more stringent in silico pipeline was developed, which included functional and structural prediction to support the identification of potential adhesin characteristics of individual proteins**.**

3.3.3 Homology modeling for conserved hypothetical genes

The hypotheticalconserved gene regions of the *M. tuberculosis* H37Rv genome were further investigated by cross aligning the protein domains against a known protein database (PDB). A protein model prediction system, Phyre2 (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index), was used to model the homology among the specific protein sequence and determine the function and structure by cross-referencing the available PDB server [15]. A degree of confidence ≥90% and alignment coverage ≥80% were interpreted to have strong homology with the protein sequence and a high-scoring template. The template information was further analyzed to include bacterial-derived sequence alignments only, which strengthened the probability of homologous identification to *M. tuberculosis* H37Rv. The unknown functional gene outputs from the Phyre2 analysis were further analyzed using the Pfam 35.0 tool to align functional association [16]. In addition, the literature searches using the following databases: NCBI, PubMed, EBSCOhost (CINAHL), Medline, and ScienceDirect were used to identify current studies that investigated known adhesins and hypothetical gene function (Figure 3.1).

3.3.4 Subcellular localization predictionPSORTb version 3.0.2 (https://www.psort.org/psortb/) was used to identify subcellular localization properties of the genes of interest [17]. Gene loci with had high localization scores (p≥8.00) were considered strong predictors of subcellular localization. In addition, MycoSub (http://lin-group.cn/server/Mycosub) and LocTree3 (https://rostlab.org/services/loctree3/) were used to verify subcellular localization (p≥0.5) The three subcellular localization tools were compared to best select the most probable localization prediction based on individual scores.

3.3.5 Translocated amino acid sequences.

SignalP- 5.0 server (https://services.healthtech.dtu.dk/services/SignalP-5.0/) was used to predict signal peptides that are expressed through the classical secretory pathway. Protein sequences that may utilize the non-classical secretory pathway were analyzed using the SecretomeP 2.0 prediction tool [18]. Individual protein sequences that were above the probability threshold ($p \geq 0.5$) were predicted to utilize the non-classical protein secretion pathway.

3.3.6. Statistical analysis

Data was analyzed using SPSS (IBM Statistical Package for Social Sciences for Windows, Version 24.0, Armonk, NY, IBM Corporation). Descriptive statistical analysis were used to generate frequencies and the mean ± standard deviation was calculated for data that was normally distributed. Probability of adhesin (P_{ad}) values was grouped according to ranges: higher P_{ad} index (≥ 0.80); moderate P_{ad} index (0.6-0.79); and lower P_{ad} index (≤ 0.59) for comparison analysis.

3.4. Results and Discussion

3.4.1. SPAAN prediction analysis

The SPAAN adhesin prediction tool generated 776 potential adhesin-related gene sequences from the *M. tuberculosis* H37Rv whole proteome (P_{ad} value range: 0.510-0.954). A comprehensive analysis of the literature was cross tabulated with the SPAAN outcomes to verify the adhesin prediction potential. A total of 24 known *M. tuberculosis* adhesin proteins were identified within the P_{ad} threshold (Figure 3.2) [7,19]. However, 10 known adhesin proteins were below the P_{ad} threshold. According to Sachdeva and colleagues (2005), SPAAN prediction was 89% specific among known adhesins. Therefore predicted adhesins that were below the P_{ad} threshold did not meet the compositional criteria for identification [12]. Interestingly, half of those known mycobacterial adhesins (GroES, PrcB, SahH, Ef-Tu, and MetK) were identified to mediate adhesion to host plasminogen [20]. Plasminogen, a precursor to plasmin, is known to be involved in the degradation of fibrin and other ECM products related to thrombus activity post-infection [21]. Furthermore, these mycobacterial plasminogen-binding proteins enhance invasion by stimulating pathogenic migration by attaching to the host plasminogen, thereby, activating fibrinolysis [21,22]. The analysis demonstrates a low probability of the adhesin potential of SPAAN to predict adhesins that are specific to the host plasminogen. The SPAAN prediction was further classified into the corresponding functional categories identified by the Mycobrowser tool

(Figure 3.1). From the analysis, gene loci that were associated with cell wall and cell wall processes were identified as the largest functional group (n=243, 31.30%) and were predicted to have strong adhesin potential in the SPAAN analysis. Cell wall proteins presented a higher proportion as predicted adhesins/ adhesin-like proteins possibly due to the SPAAN compositional attributes, particularly computing more favorably towards transmembrane associated hydrophobic proteins [12]. Relative transmembrane helix domains of cell wall proteins are highly associated with hydrophobic bilayer regions of the cell wall [23]. Such attributes, including increased hydrophobicity especially in biofilm formations, have been well established to improve bacterial adhesion to biotic and abiotic surfaces [5,24]. The SPAAN tool therefore supports the prediction of adhesin proteins that have the potential to be localized on the cell surface and possess the intrinsic hydrophobic composition of *M. tuberculosis* [1].

The majority of the known adhesin proteins were reportedly associated with the cell wall and cell wall processes namely, early secreted antigenic target 6kDa (ESAT-6), phosphate binding lipoprotein (PstS1), 19 kDa *M.tuberculosis* lipoprotein (LpqH), lipoprotein aminopeptidase (LpqL), heparin-binding hemagglutinin adhesin (HBHA), *M. tuberculosis* Curli Pili (MTP), Rv0309, Rv1806 (Apa protein) and recently Rv3194c [7,25]. In principle, cell wall-associated adhesins promote direct host-pathogen interactions, which increases the propensity for pathogenesis. Proline-glutamate/ proline-proline-glutamate (PE/PPE) proteins were the second largest identified functional group which included the known adhesins, Rv1759c (Wag22 protein), and PE domain polymorphic guanine-cytosine-rich repetitive sequence 33 (PE-PGRS33) [25,26]. PE/PPE functional proteins are commonly composed of cell surface components and contribute to the pathogenicity of mycobacteria owing to their unique intrinsically disordered protein structure that amplifies sequence variation so that the host has difficulty identifying target sites on the cell surface [27]. This further suggests that the SPAAN output would anticipate a large proportion of PE/PPE proteins establishing adhesin or adhesin-like functions based on the hydrophobic compositional attributes. Notably, a third of the PE/PPE proteins significantly displayed an average P_{ad} value (mean= 0.811, SD±0.125) and higher P_{ad} index (64,7%, n= 99) that was greater than any other categorized group, considering that these proteins are known to interact with host cell surface receptors to activate downstream signaling pathways [28]. Among the mycobacterial genome of *M. tuberculosis*, 168 genes are encoded for PE/PPE proteins [29]. Studies have shown that in addition to PPE10, PPE13 and PE-PGRS45 [30], the PE25-PPE41 complex [31], PE8-

PPE15[32] and PE35-PPE68_1 complex [33] were translocated to the cell surface via the type VII secretion system, particularly the ESX-5 secretion pathway for *M. tuberculosis*. The well-established PE-PGRS adhesin proteins namely Rv1759c and Rv1818c, that bind to host fibronectin and macrophage respectively, were interestingly not predicted within the SPAAN threshold in the current study [34,35].

Intermediatory metabolic proteins are mainly involved in intracellular chemical processes. However, these proteins have been discovered to be exported extracellularly [36]. The SPAAN analysis generated a total of 80 (10.31%) intermediatory metabolism/respiratory genes that may have adhesin potential, including fructose-1,6-bisphosphate aldolase (Fba), lipoamide dehydrogenase C (lpdC) and isocitrate lyase (icl), has been shown to possess receptors that bind to human plasminogen [20].

A small percentage of the functional categorized groups including lipid metabolism have well-established hydrophobic receptor-ligand interactions that reside within the antigen 85 complex of mycolyl transferases (Ag85a, Ag85b, and, Ag85c) known to bind to fibronectin and elastin, and secreted proteins (Mpt51 and, Mpt64) were found to bind to plasminogen [20,37,38]. Transmembrane serine/threonine- protein kinase D (PknD) is the only regulatory protein known as an adhesin that binds to laminin on brain endothelial cells [39]. Adhesin proteins, histone-like binding protein (HupB), and iron-regulating elongation factor tu (EF-tu) are exclusively involved in the biosynthetic information pathway and have been shown to bind to laminin and plasminogen respectively [20,40]. Virulence, detoxification, and adaptation grouped proteins included molecular chaperones growth E long (GroEL)1 and 2, which have been known to bind to toll like receptors on host epithelial macrophages and *Mycobacterium* cell entry 1 (Mce-1) has demonstrated to bind to host epithelial cells [41,42]. In addition, two other molecular chaperones, 10 kDa chaperone (GroES) and DnaK, are identified as adhesins due to their binding to plasminogen [42].

The Mycobrowser online tool identified almost a third (31.31%, n=167) of the SPAAN output, which has been categorized as CH genes. There are currently only three CH proteins that have been identified with adhesin characteristics, two ECM-binding adhesins (Rv2599 and Rv3717) discovered previously by Kumar *et al.*[13] using the SPAAN tool, and type IV pili

initially identified by Alteri [43], which potentially binds to host macrophages and epithelial cells [43].

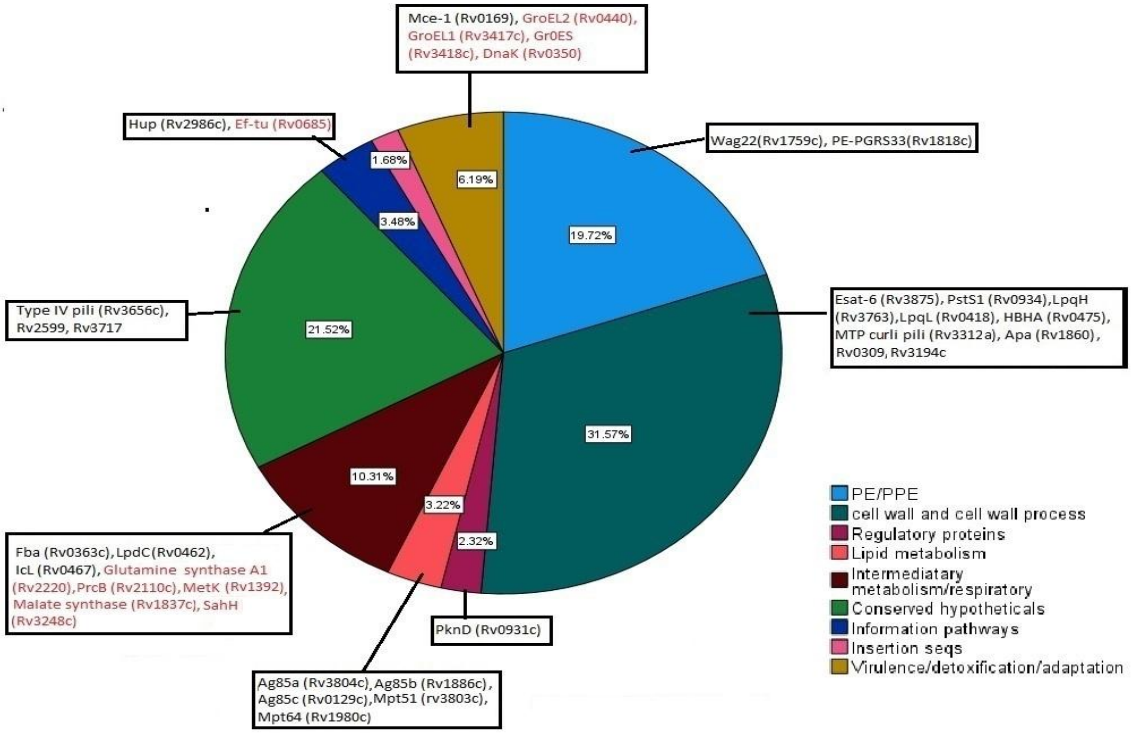


Figure 3.2 Functional categories of the *Mycobacterium tuberculosis* (MtbH37Rv) proteome identified by SPAAN analysis categorized by the Mycobrowser tool (N= 776). Currently, there are 24 known adhesins that are within the SPAAN threshold, although 10 known adhesins listed (in red) are below the selection criteria ($P_{ad} \geq 0.51$).

3.4.2. Homology classification of conserved hypothetical proteins identified from SPAAN.

The individual CH protein sequences (n=167) were aligned to well-characterized protein domains from the protein database (PDB) using the Phyre2 homology modeling tool used to predict protein function (Figure 3.3). Furthermore, unaligned phyre2 predictions were analyzed using the Pfam tool to collect additional protein information to determine functionality. The protein prediction analysis was performed on all CH proteins by aligning the domains with protein sequences from the Protein database (PDB) to determine the coverage and the probability of the gene of interest and the PDB template being homologous. Prediction alignment showed that hydrolase activity was the most frequently aligned group (n=42) with half of CH proteins rated at a moderate P_{ad} index, followed by regulatory proteins (n=27), and oxidoreductase activity (n=22), both having the majority at the lower P_{ad} index.

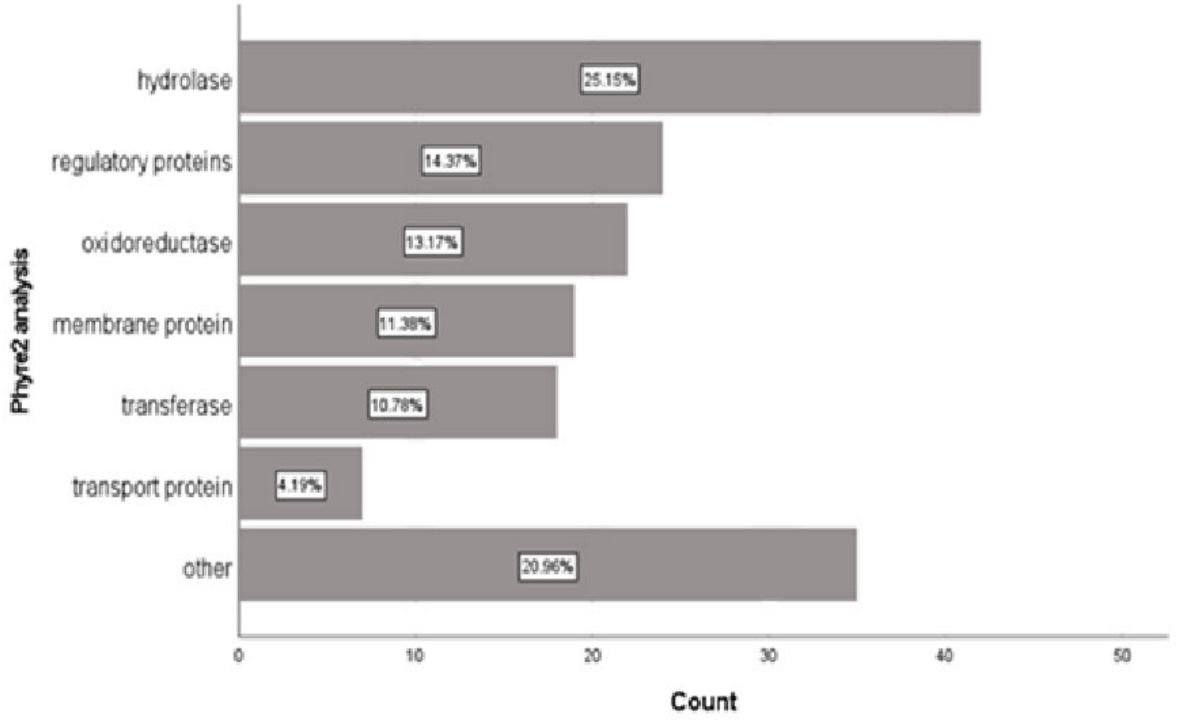


Figure 3.3. Protein homology prediction of conserved hypothetical proteins (n=167) using Phyre2 analysis. Six classes of proteins were categorized, including hydrolases were predicted as the largest proportion of conserved hypothetical proteins. The remainder of the classifications ranged from 4-15%, where the smallest proportion was identified as transport proteins. Other proteins (n=35) were predicted proteins that comprised an array of predicted protein functions that were not within the six classes of predicted proteins. All six protein classes of homology predictions are strongly associated with adhesin potential.

Amongst the predicted protein homology modeling, hydrolase activity was aligned with a high confidence rating (83%) when analyzing the protein structure with subsequent coverage (≥53%) using the known PDB template. When comparing Phyre2 analysis with known adhesins, the majority of adhesins were found to align with hydrolase activity namely, MTP, Rv0309, IpqL, Ag85 complex, MPT51, Fba, PrcB, SahH and Rv3717 (see B1: Supplementary table 1). Cell wall hydrolases have been suggested to serve as multifunctional proteins categorized as amidases, glycosidase, and peptidases with cell adhesion and virulence properties [44]. Kumar *et al.* identified two mycobacterial adhesins (Rv3717 and Rv0309) to have hydrolase activity, and that bind to fibronectin, collagen, and laminin [13]. In this study, Rv2575 was aligned with a high level of significance as a hydrolase and determined by Pfam analysis to be a zinc metallopeptidase enzyme. Commonly binding to a variety of host intestinal ECM components, fragilysin, a zinc-

dependent metallopeptidase, similarly binds and invades host epithelial cells attributing adhesin properties to enhanced pathogenicity [45].

Regulatory proteins are complex molecules that assist in the transcription and regulation of various proteins, which were predicted to have DNA/RNA/protein binding capability [46]. Mullen and colleagues (2008) identified the dual properties of competent DNA and fibronectin binding of *Pasturella* species, which supports the dual role of regulatory proteins as adhesins [47]. Identified as a mycobacterial regulatory protein, PknD is known to regulate signal transduction related to the phosphorylation of serine and threonine hydroxyl groups [48]. Subsequently, PknD-protein kinase plays an important role in adhesion to laminin on brain endothelial cells during central nervous system infection [39]. Oxidoreductases were the third largest group to be predicted from the CH SPAAN analysis, which included homology with ferredoxin and ferredoxin-like proteins. For instance, *Pseudomonas* species 120kDa oxidoreductase protein was discovered as a novel adhesin which provides evidence that these CH grouped proteins may also possess adhesin functions [49].

The category relating to "other" proteins comprised 20.96% uncategorized individual protein predictions (n=36) (Figure 3.2) and included homologous proteins with a high level of significance (Phyre2 rating) such as sugar-binding proteins (Rv1891, Rv3678c, and Rv3717) and cell invasion proteins (Rv0965c, Rv1078, Rv1950c, Rv2998, Rv3440c, and Rv3898c). Interestingly, the Phyre2 analysis aligned three hypothetical conserved gene loci (Rv2660c, Rv0979c, and Rv1078) as adhesin proteins, albeit with very low significance (≤23%) and low coverage (≤ 24%). The CH proteins with putative adhesin prediction identified in the present study were categorized in Table 3.1. These individual proteins listed were all predicted by Phyre2 and classified into protein groups with adhesin potential.

Table 3.1. Conserved hypothetical genes with predicted function using Phyre2 homology prediction analysis (P_{ad}≥0.51)

Predicted gene function (Phyre2 analysis)	Potential adhesin gene		
	higher P_{ad} index (≥0.80)	**moderate P_{ad} index (0.60-0.79)**	**lower P_{ad} index (≤0.59)**
Hydrolase	*Rv2575, Rv3337*	*Rv0340, Rv0381c, Rv0397a, Rv0577, Rv0963c, Rv1109, Rv1157c, Rv1754c, Rv1815, Rv1951c, Rv2114, Rv2240c, Rv2257c, Rv2292c, Rv2307c, Rv2422, Rv2525c, Rv2641, Rv3033, Rv3707c, Rv3811, Rv3822*	*Rv0274, Rv1716, Rv2517c, Rv2797c, Rv1154c, Rv1419, Rv1499, Rv1632c, Rv1728c, Rv2216, Rv2229c, Rv2256c, Rv2577, Rv3114, Rv3282, Rv3655c, Rv3683, Rv3796*
Membrane proteins	*Rv3209*	*Rv0679c, Rv0712, Rv2264c, Rv2730, Rv2744c, Rv3090, Rv3103c, Rv3267, Rv3706c, Rv3747, Rv3897c*	*Rv1083, Rv1489, Rv1762c, Rv3686c, Rv3484c, Rv3748*
Transferase	*Rv2077a, Rv3572*	*Rv0378, Rv0471c, Rv1137c, Rv1203c, Rv2302, Rv2960c*	*Rv0057, Rv0616c, Rv0725c, Rv0877, Rv0968, Rv1519, Rv1734, Rv1889, Rv2414c, Rv2598*
Oxidoreductase	*Rv1810*	*Rv0455c, Rv1134, Rv1804, Rv1875c, Rv2283, Rv3035, Rv3136a, Rv3353c, Rv3354, Rv3369*	*Rv0021c, Rv0799c, Rv0911, Rv1706a, Rv1117, Rv2083, Rv2728c, Rv2771c, Rv2774c, Rv3212, Rv3705a*
Regulatory Proteins	*Rv0387c, Rv1116a*	*Rv0614, Rv1057, Rv2078, Rv2082, Rv2820c, Rv3491*	*Rv0039c, Rv0185, Rv0500a, Rv0699, Rv0762c, Rv1669, Rv2312, Rv2619c, Rv2738c, Rv2821c, Rv3098c, Rv3333c, Rv3716c, Rv3745c, Rv3831, Rv3847*
Transport proteins	*None*	*Rv1268c, Rv2929*	*Rv0787, Rv0807, Rv1914c, Rv2387, Rv3005c, Rv3705c*
Other	*Rv0061c, Rv0991c, Rv1906c, Rv2599, Rv3207c, Rv3717*	*Rv0454, Rv1100, Rv1158c, Rv1891, Rv1975, Rv1998c, Rv2660c, Rv2706c, Rv2998, Rv3067, Rv3718c, Rv3898c*	*Rv0299, Rv0333, Rv0464c, Rv0965c, Rv0979c, Rv1078, Rv1950c, Rv1961, Rv2798c, Rv3182, Rv3440c, Rv3541c, Rv3656c (type IV pili), Rv3627c, Rv3678c, Rv3678a, Rv3900c,*

Approximately 30% of naturally occurring proteins are predicted to be membrane-associated, involved in cellular processes and virulence, and are thus, potential therapeutic drug targets [50]. These proteins, especially lipoproteins, have increased virulence properties because of their translocation to the cell surface [51]. This includes cell-wall metabolism, signal transducers, intracellular transport, and adhesion properties. [51-53]. In the present study, of the 18 proteins that were aligned with Phyre2 homology modeling to have membrane-associated functions, Rv0679c was identified to have a high activity binding peptide involved in host-pathogen recognition interactions [51]. The Phyre2 analysis of known proteins identified three mycobacterial adhesins (Rv3197c, Rv1759c, and Rv1818c) that were predicted as structural proteins, often a component of the cell wall. Structural proteins, Rv1759c and Rv1818c, belong to the PE/PPE family and Rv3197c was predicted to be located on the cell surface, as corroborated by the homology prediction outcome [25,28]. Membrane-associated proteins may prove as potential therapeutic drug targets due to their strong activation of human T-lymphocytes [53,54].

The present study identified 18 CH proteins that may have a dual function as adhesin and transferase activity. Transferase activity was notably observed in adhesins such as the Ag85 complex with a cell wall mycolyl transferase enzymatic role in *M. tuberculosis* [4,55]. Transferases have been identified among other bacterial species of *Streptococcus pneumoniae* to have adhesion properties and protein glycosylation [56].

3.4.3 Subcellular localization analysis

Subcellular localization (SCL) prediction is of interest to determine function, therapeutic potential, and host-pathogenic interactions due to the localized motif expression of these proteins [57,58]. In this study, the *in-silico* structural analysis included the use of three tools namely, PSORTb, LocTree3, and MycoSub to predict the SCL of the protein-based sequence in order to determine the probability of the expressed protein being translocated to a specific cell site. The three SCL tools were used to compare and determine the highest probability of predicted localization within the mycobacterial complex to improve prediction specificity and sensitivity. Computational analysis of SCL predicted potential adhesin proteins to be found in the cytoplasm, on the membrane or extracellular space (Table 3.2). The current study demonstrated that approximately 40.7% of the CH proteins were found to be largely secreted extracellularly (n=68), while approximately 38.3% were predicted to be cytoplasmic of origin (n=64) and only 21% of

Table 3.2. Subcellular localization and signal peptide analysis of the SPAAN predicted conserved hypothetical genes.

Subcellular localization	Gene loci				
	Sec-dependent signal peptide	TAT-dependent signal peptide	Lipoprotein signal peptide	Non-classical secretory pathway	unknown
Extracellular Secreted	Rv0061c, Rv0397a, Rv0455c, Rv1158c, Rv1419, Rv1815, Rv1891, Rv1906c, Rv1975, Rv3067, Rv3333c, Rv3572, Rv3655, Rv3656c, Rv3897c,	Rv1157c, Rv1669,	Rv0679c, Rv0699, Rv2240, Rv3353c,	Rv0340, Rv0500a, Rv0991c, Rv1078, Rv1100, Rv1109c, Rv1728c, Rv1951c, Rv2077a, Rv2083, Rv2706c, Rv2730, Rv2960c, Rv2998, Rv3098c, Rv3337, Rv3705a, Rv3707c, Rv3898c	Rv0057, Rv0185, Rv0299, Rv0378, Rv0387c, Rv0454, Rv0464c, Rv0577, Rv0614, Rv0799c, Rv0968, Rv0979c, Rv1137c, Rv1499, Rv1519, Rv1706a, Rv1914c, Rv1950c, Rv2078, Rv2283, Rv2312, Rv2517c, Rv2798c, Rv2929, Rv3103c, Rv3678a, Rv3745c, Rv3748,
Cytoplasmic	Rv1268c, Rv1804, Rv2216, Rv3207c, Rv3491, Rv3678c, Rv3705c, Rv3717,	Rv2577, Rv2599	Rv0381, Rv3033, Rv3212,	Rv0471c, Rv0877, Rv1134, Rv1716, Rv1961, Rv1998c, Rv2256c, Rv2292c, Rv2641, Rv2660c, Rv2774c, Rv3369, Rv3484, Rv3716c,	Rv0021c, Rv0274, Rv0333, Rv0712, Rv0725c, Rv0787, Rv0807, Rv0911, Rv0965c, Rv1116, Rv1117, Rv1154c, Rv1203c, Rv1632c, Rv1734c, Rv1762c, Rv1875, Rv2229c, Rv2302, Rv2422, Rv2575, Rv2598, Rv2619c, Rv2728c, Rv2738c, Rv2744c, Rv2771c, Rv2821c, Rv2797c, Rv3114, Rv3182, Rv3282, Rv3440c, Rv3541c, Rv3627, Rv3718, Rv3900c,
Membrane Bound	Rv0039c Rv1810, Rv2257c, Rv2264c Rv3267, Rv3354		Rv1889c, Rv3796	Rv0963c, Rv1057, Rv1083, Rv2114, Rv3035, Rv3090, Rv3686c, Rv3706c, Rv3811	Rv0616c, Rv0762c, Rv1489, Rv1754c, Rv2082, Rv2307c, Rv2387, Rv2414c, Rv2525, Rv2820c, Rv3005c, Rv3136a, Rv3209, Rv3683, Rv3747, Rv3822, Rv3831, Rv3847

subcellular localization prediction was membrane-bound (n=35). Proteins that were predicted to be secreted extracellularly comprised a larger proportion of predicted adhesins (Table 3.2), which align well with well-established known adhesins such as ESAT-6, LpqL, Mpt64, and Apa. However, the majority of known adhesins were within the membrane/ cell wall, including Rv2599, Wag22, PE-PGRS33, Rv0309, PstS1, Mce-1, HBHA, Mpt51, MTP, and type IV pili. The known adhesin proteins Rv3717, LpdC, and Icl, were predicted to be localized in the cytoplasm. The potential selection of predicted adhesin proteins for experimental progression would highly favor proteins that are better exposed to the cell surface, particularly those that have strong biomarker potential for therapeutic and diagnostic application.

3.4.5 Signal Peptide predictions

Excreted or secreted proteins have a potential role in extracellular biological processes conferring highly immunogenic properties, and therefore, are essential in developing diagnostic and therapeutic candidates [59,60]. In the present study, the protein prediction tool, SignalP (5.0) was used to verify the location of proteins as secretory signal peptides (Sec/Tat/lipoprotein translocon) as seen in Table 3.2. Proteins that identify as extracellular SCL consist of specialized amino acid sequences that express signal peptide amino acid sequences close to the N-terminus to translocate to the transmembrane surface via the Sec translocon and function as surface-expressed proteins. This is the most common secretory mechanism and requires a signal peptide for the Sec-dependent secretion [61]. Protein sequences that contain signal peptides have a higher probability of functioning as an adhesin located outside the cell [61]. In the current study, 43 CH proteins were predicted to contain a signal peptide sequence. Half of the predicted signal peptide genes (n=21) had a moderate P_{ad} value (0.6-0.79). However, this does not suggest that the predicted adhesin proteins are not secreted. The non-classical secretory pathway identified in mycobacterial species undergoes unconventional protein secretion and are termed leaderless proteins since there is a lack of the presence of a signal peptide sequence [62].

In the present study, the SecretomeP 2.0 prediction tool identified leaderless gene loci that were predicted to follow the non-classical secretory pathways (n=42). These gene loci were predicted to encode a secretory embedded peptide sequence that allows the proteins to be exported extracellularly. This tool may be used to validate the SPAAN output analysis to improve the prediction of adhesin and adhesin-like proteins. In support of this, *M. tuberculosis* has been

experimentally proven to contain an array of proteins that utilize the non-classical secretory pathway [59]. This includes a known adhesin protein, GlnA (Rv2220) which is localized in the cytoplasm, and possesses a signal protein embedded within the protein sequence. [61] In addition, the known adhesin proteins; EF-tu (Rv0685), GroES (RV3418c), GroEL (Rv0440), and DnaK (Rv0350) have been experimentally proven to utilize the non-classical secretory pathway [63].

In silico comparative proteomics may aid in predicting the function of these proteins [64] by determining the localization of the protein, particularly if the protein is secreted extracellularly or expressed on the cell surface. However, not all predictions of adhesin proteins may align with the former as there also exists anchorless adhesin proteins [2]. Signal peptides have various applications and roles especially in efficiently producing recombinant proteins by preventing inclusion body formation and protease degradation [65]. However, Kumar *et al.* (2013) reported great difficulty in purifying soluble proteins due to the formation of inclusion bodies and possible toxicity of membrane proteins [13]. To alleviate this, truncated genes were synthesized with the absence of signal peptide and hydrophobic sequences.

Although, there has been success in identifying *in-silico* predicted proteins with adhesin-like properties via experimental validation, there are still challenges in using such prediction tools. Protein homology for instance is limited, due to the prediction being dependent on the accessibility of known compiled protein databases to detect multidomain sites [2]. Adhesins have been known to contain unique amino acid sequences that are a diverse group, making it difficult to compare across other known proteins [2]. Adhesins may also mutate their common pattern recognition receptors within their protein sequences to avoid host immune recognition that precludes homology modeling [66]. Mutations including short nucleotide polymorphisms that result in strain variation may hinder homology detection of adhesin-specific sequences and require a more robust database, particularly among strain variants of *M. tuberculosis* [67]. Bioinformatics research should act sequentially as the initial step in aligning computational analysis with experimental corroboration, as databases across multiple genomic platforms continue to be consolidated. Implementation of these collective methods can improve the identification of novel genes with adhesin properties in an experimental setting.

Currently, there is only one published study that incorporates the *in-silico* analysis using SPAAN followed by experimental validation of predicted mycobacterial adhesin.

Kumar *et al.*[13] identified three novel hypothetical proteins (Rv0309, Rv3717, and Rv2599) experimentally by investigating the binding potential to ECM components [13]. The Rv3717 adhesin protein was shown to bind to ECM components with a high affinity for collagen, fibronectin, and laminin-binding [13]. The selection of these potential adhesin proteins was subjected to a stringent computational process to identify adhesin function. ELISA-based assays confirmed the binding potential of the expressed proteins to collagen, fibronectin, and laminin [13]. This systematic approach using both computational algorithms and experimental validation provides evidence for the successful use of the SPAAN analysis tool.

3.5. Conclusion

This bioinformatics-based study that used a systematic approach to *in-silico* comparative analysis that characterizes 167 potential adhesin proteins from the *M. tuberculosis* genome. SPAAN is currently the only adhesin prediction tool based on artificial neural networks to detect algorithms specific to adhesin attributes in *M. tuberculosis* (http://sourceforge.net/projects/adhesin/files). However, coupled with additional functional and structural based prediction tools, SPAAN was shown to be highly selective in identifying adhesin and adhesin like proteins, as evidenced by the number of known adhesins (n=34) detected. However, approximately one third of these known adhesins were demonstrated to be below the probability of adhesin (P_{ad}) threshold ($P_{ad} \geq 0.51$). The predicted novel adhesin dataset identified in this study can be referenced to explore experimental validation in future research. In addition, CH proteins have the potential of dual functional roles, enzymatic and adhesin properties that contain translocation sequences to determine their subcellular localization. Experimentally validated bacterial adhesins have been identified to have dual functions, which is supported by the findings of the present study. The *M. tuberculosis* proteome contains an array of unclassified proteins, particularly CH proteins, which once validated experimentally, potentially have novel therapeutic and serodiagnostic applications. This study therefore provides a pipeline to identify and functionally characterize SPAAN-related CH proteins to be critically validated experimentally for their adhesion potential.

Expert Opinion

Mycobacterium tuberculosis, the causative agent of tuberculosis, remains a global public health epidemic and has the highest cause of mortality of any infectious disease [68]. The development of effective drug therapies, ideal novel vaccines, and early detection diagnostic assays using stimulant antigenic biomarkers are essential to address this issue.

There is currently an increasing interest in the bioinformatics approaches, which offer a pre-selective foundation for novel and innovative research. More particularly in the scope of proteomics, whereby in silico tools can be used to predict the functional, genomic, and structural attributes of potential novel proteins. This has been developed into the primary focus when developing downstream high-cost experimentation. There is an abundance of valuable information gathered in large protein databases that provide an improvement in bioinformatic predictions. The new age of research should pivot around this available information. We are therefore able to channel *in silico* pipelines to challenge these novel proteins in an improved experimental setting to verify their potential.

One such pipeline that has drawn much attention is the dissemination and diversification of potential novel vaccine candidates and serological markers that may strengthen communicable disease control [69]. There is a need to explore novel antigenic proteins that may contribute to the battle against the spread of diseases, including TB, against which BCG is the only commercially available protective vaccine, with the limitation of reduced efficacy in adults [70]. Currently, various *M. tuberculosis* adhesin proteins are being validated for antigenic stimulus and serological potential, these include known adhesins such as ESAT-6, Ag85 complex, and MPT64 [8,71]. Furthermore, the use of these antigenic adhesin-related proteins can be enhanced in a fusion peptide design to improve immunogenic stimulus due to their cell-surface exposure to host cell-mediated mechanisms [72,73].

Future perspectives in the field of bioinformatics may enhance and accelerate vaccine and diagnostic development. SPAAN prediction can be a powerful tool, as shown in this study. However, when analyzing the *M. tuberculosis* genome, more adapted criteria were required, coupled with a pipeline of systematic *in-silico* tools. In a post-COVID-19 era, identification of mycobacterial adhesin-related proteins may provide a progressive approach for vaccine candidacy to effectively replace BCG or offer a prime boost strategy to reduce infections globally.

Author Contributions

RM, SS, and MP were involved in the conception and design of the study. RM and SS provided analysis and interpretation of the data. RM drafted the manuscript, SS and MP revised the manuscript for intellectual content. All authors provide the final approval of the version to be published and agree to be accountable for all aspects of the work. The authors declare no conflict of interest.

Funding Information

This study was funded by the University of KwaZulu-Natal (UKZN) College of Health Sciences Scholarship, UKZN APACHE Flagship funds, and the SAMRC Division of Research Capacity Development under the Bongani Mayosi Health Scholars Programme. The content hereof is the sole responsibility of the authors and does not necessarily represent the official views of the funding bodies.

Declaration of Interest

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Acknowledgments

This work was presented at the 7th SA TB Conference 2022 and the SAMRC 16 Early Career Scientific Convention 2022.

Ethical Consideration

The study was approved by the Biomedical Research Ethics Council, University of KwaZulu-Natal (approval ref: BREC/00001970/2020).

Supplementary data

Available from Appendices B1: Chapter three supplementary data of this thesis

References

1. Squeglia F, Ruggiero A, De Simone A, Berisio R. A structural overview of mycobacterial adhesins: Key biomarkers for diagnostics and therapeutics. Protein Sci. 2018;27(2):369-380. https://doi.org/10.1002/pro.3346

2. Chaudhuri R, Ramachandran S. Prediction of virulence factors using bioinformatics approaches. Methods Mol Biol. 2014;1184:389-400. https://doi.org/10.1007/978-1-4939-1115-8_22

3. Ryndak MB, Laal S. *Mycobacterium tuberculosis* primary infection and dissemination: a critical role for alveolar epithelial cells. Front Cell Infect Microbiol. 2019;9:299. https://doi.org/10.3389/fcimb.2019.00299

4. Bisht D, Meena LS. Adhesion molecules facilitate host-pathogen interaction & mediate *Mycobacterium tuberculosis* pathogenesis. Indian J Med Res. 2019;150(1):23-32. https://doi.org/10.4103/ijmr.IJMR_2055_16

5. Viljoen A, Dufrêne YF, Nigou J. Mycobacterial adhesion: from hydrophobic to receptor-ligand interactions. Microorganisms. 2022;10(2):454. https://doi.org/10.3390/microorganisms10020454

6. Kevlani N, Meena L. Prominent role of FnBPs of *mycobacterium tuberculosis* in cell adhesion, immune invasion and pathogenesis. Open J Bac. 2017;1(1):007-012. https://doi.org/10.17352/ojb.000002

7. Vinod V, Vijayrajratnam S, Vasudevan AK, Biswas R. The cell surface adhesins of *Mycobacterium tuberculosis*. Microbiol Res. 2020;232:126392. https://doi.org/10.1016/j.micres.2019.126392

8. Kim H, Choi H-G, Shin SJ. Bridging the gaps to overcome major hurdles in the development of next-generation tuberculosis vaccines. Front Immunol. 2023;14:1193058. https://doi.org/10.3389/fimmu.2023.1193058

9. Govender VS, Ramsugit S, Pillay MJM. *Mycobacterium tuberculosis* adhesins: potential biomarkers as anti-tuberculosis therapeutic and diagnostic targets. Microbiology. 2014;160(9):1821-1831. https://doi.org/10.1099/mic.0.082206-0

10. Attallah AM, Abdel Malak CA, Ismail H, El-Saggan AH, Omran MM, Tabll AA. Rapid and simple detection of a *Mycobacterium tuberculosis* circulating antigen in serum using dot-elisa for field diagnosis of pulmonary tuberculosis. J Immunoassay Immunochem. 2003;24(1):73-87. https://doi.org/10.1081/IAS-120018470

11. O'Rourke KF, Gorman SD, Boehr DD. Biophysical and computational methods to analyze amino acid interaction networks in proteins. Comput Struct Biotechnol J. 2016;14:245-251. https://doi.org/10.1016/j.csbj.2016.06.002

12. Sachdeva G, Kumar K, Jain P, Ramachandran S. SPAAN: a software program for prediction of adhesins and adhesin-like proteins using neural networks. Bioinformatics. 2005;21(4):483-491. https://doi.org/10.1093/bioinformatics/bti028

13. Kumar S, Puniya BL, Parween S, Nahar P, Ramachandran S. identification of novel adhesins of *m. tuberculosis* h37rv using integrated approach of multiple computational algorithms and experimental analysis. PLOS ONE. 2013;8(7):e69790. https://doi.org/10.1371/journal.pone.0069790

14. Kumari B, Saini V, Kaur J, Kaur J. Rv2037c, a stress induced conserved hypothetical protein of *Mycobacterium tuberculosis*, is a phospholipase: Role in cell wall modulation and intracellular survival. Int J Biol Macromol. 2020;153:817-835. https://doi.org/10.1016/j.ijbiomac.2020.03.037.

15. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015;10(6):845-858. https://doi.org/10.1038/nprot.2015.053

16. Mistry J, Chuguransky S, Williams L, *et al.* Pfam: The protein families database in 2021. Nucleic Acids Res. 2021;49(D1):D412-D419. https://doi.org/10.1093/nar/gkaa913.

17. Yu NY, Wagner JR, Laird MR, *et al.* PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes Bioinformatics. 2010;26(13):1608-1615. https://doi.org/10.1093/bioinformatics/btq249

18. Bendtsen JD, Kiemer L, Fausbøll A, Brunak S. Non-classical protein secretion in bacteria. BMC Microbiology. 2005;5(1):58. https://doi.org/10.1186/1471-2180-5-58

19. Ramsugit S, Pillay M. Identification of *Mycobacterium tuberculosis* adherence-mediating components: a review of key methods to confirm adhesin function. Iran J Basic Med Sci. 2016;19(6):579. https://dx.doi.org/10.22038/ijbms.2016.7124

20. Xolalpa W, Vallecillo AJ, Lara M, *et al.* Identification of novel bacterial plasminogen-binding proteins in the human pathogen *Mycobacterium tuberculosis*. Proteomics. 2007;7(18):3332-3341. https://doi.org/10.1002/pmic.200600876

21. Ayón-Núñez DA, Fragoso G, Bobes RJ, Laclette JP. Plasminogen-binding proteins as an evasion mechanism of the host's innate immunity in infectious diseases. Biosci Rep. 2018;38(5). https://doi.org/10.1042/bsr20180705

22. Jeffery C. Intracellular proteins moonlighting as bacterial adhesion factors. AIMS Microbiol. 2018;4(2):362-376. https://doi.org/10.3934/microbiol.2018.2.362

23. De Marothy MT, Elofsson A. Marginally hydrophobic transmembrane α-helices shaping membrane protein folding. Protein Sci. 2015;24(7):1057-1074. https://doi.org/10.1002/pro.2698

24. Krasowska A, Sigler K. How microorganisms use hydrophobicity and what does this mean for human needs? Front Cell Infect Microbiol. 2014;4:112. https://doi.org/10.3389/fcimb.2014.00112

25. Zhao D, Lin D, Xu C. A protein fragment of Rv3194c located on mycobacterial cell surface efficiently prevents adhesion of recombinant *Mycobacterium smegmatis*, and promises a new anti-adhesive drug. Microb Pathog. 2020;149:104498. https://doi.org/10.1016/j.micpath.2020.104498

26. Meena PR, Meena LS. Fibronectin binding protein and Ca2+ play an access key role to mediate pathogenesis in *Mycobacterium tuberculosis*: An overview. Biotechnol Appl Biochem. 2016;63(6):820-826. https://doi.org/10.1002/bab.1434

27. Mukhopadhyay S, Balaji KN. The PE and PPE proteins of *Mycobacterium tuberculosis*. Tuberculosis. 2011;91(5):441-447. https://doi.org/10.1016/j.tube.2011.04.004

28. Qian J, Chen R, Wang H, Zhang X. Role of the PE/PPE family in host–pathogen interactions and prospects for anti-tuberculosis vaccine and diagnostic tool design. Front Cell Infect Microbiol. 2020;10:594288. https://doi.org/10.3389/fcimb.2020.594288

29. Cole ST, Brosch R, Parkhill J, *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature. 1998;393(6685):537-544. https://doi.org/10.1038/31159

30. Abdallah AM, Verboom T, Weerdenburg EM, *et al.* PPE and PE_PGRS proteins of *Mycobacterium marinum* are transported via the type VII secretion system ESX-5. Mol Microbiol. 2009;73(3):329-340. https://doi.org/10.1111/j.1365-2958.2009.06783.x

31. Daleke MH, Ummels R, Bawono P, *et al.* General secretion signal for the mycobacterial type VII secretion pathway. Proc Natl Acad Sci. 2012;109(28):11342-11347. https://doi.org/10.1073/pnas.1119453109

32. Chen X, Cheng H-f, Zhou J, *et al.* Structural basis of the PE–PPE protein interaction in *Mycobacterium tuberculosis*. J Biol Chem. 2017;292(41):16880-16890. https://doi.org/10.1074/jbc.M117.802645

33. Damen MPM, Phan TH, Ummels R, Rubio-Canalejas A, Bitter W, Houben ENG. Modification of a PE/PPE substrate pair reroutes an Esx substrate pair from the mycobacterial ESX-1 type VII secretion system to the ESX-5 system. J Biol Chem. 2020;295(18):5960-5969. https://doi.org/10.1074/jbc.RA119.011682

34. Espitia C, Laclette JP, Mondragón- Palomino M, *et al.* The PE-PGRS glycine-rich proteins of *Mycobacterium tuberculosis* : a new family of fibronectin-binding proteins? Microbiol. 1999;145(12):3487-3495. https://doi.org/10.1099/00221287-145-12-3487

35. Brennan MJ, Delogu G, Chen Y, *et al.* Evidence that mycobacterial PE_PGRS proteins are cell surface constituents that influence interactions with other cells. Infect Immun. 2001;69(12):7326-7333. https://doi.org/10.1128/IAI.69.12.7326-7333.2001

36. Tsirigotaki A, De Geyter J, Šoštaric N, Economou A, Karamanou S. Protein export through the bacterial Sec pathway. Nat Rev Microbiol. 2017;15(1):21-36. https://doi.org/10.1038/nrmicro.2016.161

37. Kuo C-J, Bell H, Hsieh C-L, Ptak CP, Chang Y-F. Novel mycobacteria antigen 85 complex binding motif on fibronectin. J Biol Chem. 2012;287(3):1892-1902. https://doi.org/10.1074/jbc.M111.298687

38. Kuo C-J, Ptak CP, Hsieh C-L, Akey BL, Chang Y-F. Elastin, a novel extracellular matrix protein adhering to mycobacterial antigen 85 complex. J Biol Chem. 2013;288(6):3886-3896. https://doi.org/10.1074/jbc.M112.415679

39. Be NA, Bishai WR, Jain SK. Role of *Mycobacterium tuberculosis* pknD in the Pathogenesis of central nervous system tuberculosis. BMC Microbiol. 2012;12(7):2012. https://doi.org/10.1186/1471-2180-12-7

40. Kalra P, Mishra SK, Kaur S, *et al.* G-quadruplex-forming dna aptamers inhibit the dna-binding function of hupB and *Mycobacterium tuberculosis* entry into host cells. Mol Ther Nucleic Acids. 2018;13:99-109. https://doi.org/10.1016/j.omtn.2018.08.011

41. Chitale S, Ehrt S, Kawamura I, *et al.* Recombinant *Mycobacterium tuberculosis* protein associated with mammalian cell entry. Cell Microbiol. 2001;3(4):247-254. https://doi.org/10.1046/j.1462-5822.2001.00110.x

42. Hickey TB, Ziltener HJ, Speert DP, Stokes RW. *Mycobacterium tuberculosis* employs Cpn60. 2 as an adhesin that binds CD43 on the macrophage surface. Cell Microbiol. 2010;12(11):1634-1647. https://doi.org/10.1111/j.1462-5822.2010.01496.x

43. Alteri CJ. Novel pili of *Mycobacterium tuberculosis*: The University of Arizona; 2005.

44. Vermassen A, Leroy S, Talon R, Provot C, Popowska M, Desvaux M. Cell wall hydrolases in bacteria: insight on the diversity of cell wall amidases, glycosidases and peptidases toward peptidoglycan. Front Microbiol. 2019;10:331-331. https://doi.org/10.3389/fmicb.2019.00331

45. Ramírez-Larrota JS, Eckhard U. An Introduction to bacterial biofilms and their proteases, and their roles in host infection and immune evasion. Biomolecules. 2022;12(2):306. https://doi.org/10.3390/biom12020306

46. Bervoets I, Charlier D. Diversity, versatility and complexity of bacterial gene regulation mechanisms: opportunities and drawbacks for applications in synthetic biology. FEMS Microbiol Rev. 2019;43(3):304-339. https://doi.org/10.1093/femsre/fuz001

47. Mullen LM, Bossé JT, Nair SP, *et al. Pasteurellaceae* ComE1 proteins combine the properties of fibronectin adhesins and dna binding competence proteins. PLOS ONE. 2008;3(12):e3991. https://doi.org/10.1371/journal.pone.0003991

48. Metri R, Hariharaputran S, Ramakrishnan G, *et al.* SInCRe—structural interactome computational resource for *Mycobacterium tuberculosis*. Database. 2015;2015. https://doi.org/10.1093/database/bav060

49. Moreno-Brito V, Yáñez-Gómez C, Meza-Cervantez P, *et al.* A *Trichomonas vaginalis* 120 kDa protein with identity to hydrogenosome pyruvate: ferredoxin oxidoreductase is a surface adhesin induced by iron. Cell Microbiol. 2005;7(2):245-258. https://doi.org/10.1111/j.1462-5822.2004.00455.x

50. Tan S, Tan HT, Chung MC. Membrane proteins and membrane proteomics. Proteomics. 2008;8(19):3924-3932. https://doi.org/10.1002/pmic.200800597

51. Cifuentes DP, Ocampo M, Curtidor H, *et al. Mycobacterium tuberculosis* Rv0679c protein sequences involved in host-cell infection: Potential TB vaccine candidate antigen. BMC Microbiol. 2010;10(1):109. https://doi.org/10.1186/1471-2180-10-109

52. Vo JL, Martínez Ortiz GC, Subedi P, *et al.* Autotransporter adhesins in *Escherichia coli* pathogenesis. Proteomics. 2017;17(23-24):1600431. https://doi.org/10.1002/pmic.201600431

53. Sinha S, Kosalai K, Arora S, *et al.* Immunogenic membrane-associated proteins of *Mycobacterium tuberculosis* revealed by proteomics. Microbiol. 2005;151(7):2411-2419. https://doi.org/10.1099/mic.0.27799-0

54. Xiong Y, Chalmers MJ, Gao FP, Cross TA, Marshall AG. Identification of *Mycobacterium tuberculosis* H37Rv integral membrane proteins by one-dimensional gel electrophoresis and liquid chromatography electrospray ionization tandem mass spectrometry. J Proteome Res. 2005;4(3):855-861. https://doi.org/10.1021/pr0500049

55. Ptak CP, Kuo CJ, Chang YF. *Mycobacterium tuberculosis* antigen 85 family proteins: mycolyl transferases and matrix-binding adhesins. moonlighting proteins. 2017:357-369. https://doi.org/10.1002/9781118951149.ch20

56. Shi W-W, Jiang Y-L, Zhu F, *et al.* Structure of a novel O-Linked N-acetyl-d-glucosamine (O-GlcNAc) transferase, GtfA, reveals insights into the glycosylation of pneumococcal serine-rich repeat adhesins. J Biol Chem. 2014;289(30):20898-20907. https://doi.org/10.1074/jbc.M114.581934

57. Peabody MA, Laird MR, Vlasschaert C, Lo R, Brinkman FSL. PSORTdb: Expanding the bacteria and archaea protein subcellular localization database to better reflect diversity in cell envelope structures. Nucleic Acids Res. 2015;44(D1):D663-D668. https://doi.org/10.1093/nar/gkv1271

58. Tang S-N, Sun J-M, Xiong W-W, Cong P-S, Li T-H. Identification of the subcellular localization of mycobacterial proteins using localization motifs. Biochimie. 2012;94(3):847-853. https://doi.org/10.1016/j.biochi.2011.12.003

59. Cornejo-Granados F, Zatarain-Barrón ZL, Cantu-Robles VA, *et al.* secretome prediction of two *M. tuberculosis* clinical isolates reveals their high antigenic density and potential drug targets. Front Microbiol. 2017;8:128-128. https://doi.org/10.3389/fmicb.2017.00128

60. Gomez M, Johnson S, Gennaro ML. Identification of secreted proteins of *Mycobacterium tuberculosis* by a bioinformatic approach. Infect Immun. 2000;68(4):2323-2327. https://doi.org/10.1128/IAI.68.4.2323-2327.2000

61. Kang Q, Zhang D. Principle and potential applications of the non-classical protein secretory pathway in bacteria. Appl Microbiol Biotechnol. 2020;104(3):953-965. https://doi.org/10.1007/s00253-019-10285-4

62. Harth Gn, Horwitz MA. Export of recombinant *Mycobacterium tuberculosis* superoxide dismutase is dependent upon both information in the protein and mycobacterial export machinery. A model for studying export of leaderless proteins by pathogenic mycobacteria. J Biol Chem. 1999;274(7):4281-4292. https://doi.org/10.1074/jbc.274.7.4281

63. Wang G, Xia Y, Song X, Ai L. Common non-classically secreted bacterial proteins with experimental evidence. Curr Microbiol. 2016;72(1):102-111. https://doi.org/10.1007/s00284-015-0915-6

64. Gupta D, Banerjee S, Pailan S, Saha P. In silico identification and characterization of a hypothetical protein of *Mycobacterium tuberculosis* EAI5 as a potential virulent factor. Bioinformation. 2016;12(3):182-191. https://doi.org/10.6026/97320630012182

65. Owji H, Nezafat N, Negahdaripour M, Hajiebrahimi A, Ghasemi Y. A comprehensive review of signal peptides: structure, roles, and applications. Eur J Cell Biol. 2018;97(6):422-441. https://doi.org/10.1016/j.ejcb.2018.06.003

66. Hajishengallis G, Martin M, Sojar HT, *et al.* Dependence of bacterial protein adhesins on toll-like receptors for proinflammatory cytokine induction. Clin Diagn Lab Immunol. 2002;9(2):403-411. https://doi.org/10.1128/cdli.9.2.403-411.2002

67. O'Toole RF, Gautam SS. Limitations of the *Mycobacterium tuberculosis* reference genome H37Rv in the detection of virulence-related loci. Genomics. 2017;109(5):471-474. https://doi.org/10.1016/j.ygeno.2017.07.004

68. World Health Organization. Global tuberculosis report 2015: World Health Organization; 2015.

69. Cobelens F, Suri RK, Helinski M, *et al.* Accelerating research and development of new vaccines against tuberculosis: a global roadmap. Lancet Infect Dis. 2022;22(4):e108-e120. https://doi.org/10.1016/S1473-3099(21)00810-0

70. Moliva JI, Turner J, Torrelles JB. Immune responses to bacillus Calmette–Guérin vaccination: why do they fail to protect against *Mycobacterium tuberculosis*? Fron Immunol. 2017;8:407. https://doi.org/10.3389/fimmu.2017.00407

71. Kaufmann SHE, Weiner J, von Reyn CF. Novel approaches to tuberculosis vaccine development. Int J Infect Dis. 2017;56:263-267. https://doi.org/10.1016/j.ijid.2016.10.018

72. Khademi F, Yousefi-Avarvand A, Derakhshan M, *et al. Mycobacterium tuberculosis* HspX/EsxS Fusion Protein: Gene Cloning, Protein Expression, and Purification in *Escherichia coli.* Rep Bioc Mol Biol. 2017;6(1):15-21. https://iranjournals.nlai.ir/handle/123456789/505343

73. Arif S, Akhter M, Khaliq A, Akhtar MW. Fusion peptide constructs from antigens of *M. tuberculosis* producing high T-cell mediated immune response. PLoS One. 2022;17(9):e0271126. https://doi.org/10.1371/journal.pone.0271126

Chapter 4: The Cloning, Expression, and Purification of Recombinant Proteins of *Mycobacterium tuberculosis*

This chapter describes the molecular cloning, expression, and purification of conserved hypothetical recombinant proteins from *Mycobacterium tuberculosis* using GST/His-affinity pGEX-6P-1 and pET28a (+) expression vector systems, respectively. The purity and yield of the recombinant proteins obtained from both vectors were compared, and conditions for optimal protein purification were established. The findings highlight that successful production of conserved hypothetical proteins for the use in various downstream applications depends on the choice of expression vector, and that a systematic approach can be optimized to meet application-specific requirements. This manuscript was prepared for submission to the Molecular Microbiology Journal (Wiley publishing, Impact factor: 2.6)

Title: The use of GST/His affinity tag expression systems for recombinant production of solubilized *Mycobacterium tuberculosis* conserved hypothetical proteins with adhesin potential

*Rivesh Maharajh[1], Sibusiso Senzani[1], Balakrishna Pillay[2] and Manormoney Pillay[1]**

[1]Discipline of Medical Microbiology, School of Medicine, College of Health Sciences, University of KwaZulu-Natal, Durban 4013, South Africa

[2]Discipline of Microbiology, School of Life Sciences, College of Sciences, University of KwaZulu-Natal, Durban 4013, South Africa

***Correspondence: Prof. Manormoney Pillay (PhD), Discipline of Medical Microbiology, University of KwaZulu-Natal, Durban, 4000, South Africa. Email address:** Pillayc@ukzn.ac.za

4.1. Abstract

The *Mycobacterium tuberculosis* proteome contains numerous uncharacterized conserved hypothetical (CH) with potential adhesin properties for novel biomarker discovery. Therefore, the biological evaluation relies on high yields of purified protein in its native conformation state during cloning, expression, and purification in heterologous systems. Seven *M. tuberculosis* CH proteins were cloned into glutathione-S-transferase (GST)-pGEX-6P-1 and 6x Histidine (His)-pET-28a vectors, respectively. Recombinant proteins were expressed under optimal temperature, incubation time, and induction conditions. Solubilization methods of the GST-affinity-tag proteins (phosphate buffered saline-lysis, sodium dodecyl sulfate, and sarkosyl) and His-affinity-tag proteins (native and denaturing conditions) were compared. Purification was performed using gravity-flow column chromatography and the batch method for GST-affinity tag proteins. Only the batch method was performed for His-affinity tag proteins. In silico strategies were employed to improve primer design (restriction site and Gaunine-Cytosine overhangs) and assess signal peptide sequences to improve solubility. Expression was more suitable at low-temperature induction (16-30°C) and varied isopropyl β-D-1-thiogalactopyranoside (IPTG) concentrations (0.25-1.0 mM). The 1% sarkosyl solubilization method with addition of triton X-100 and 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS) to promote refolding significantly improved proteins yields for GST-affinity tag proteins, with batch-method purification yielding higher protein concentrations than gravity-flow chromatography. The His-affinity tag proteins required denaturing conditions (8 M urea) but required removal of steric interference by ethanol precipitation for serological assays. Rv2730 and Rv2272 expressed well in the GST-system but showed poor expression in the His-system, suggesting toxicity of proteins within the expression host stabilized with the larger GST-affinity tag. Both systems were shown to be viable. However GST is preferred for problematic CH proteins such as Rv2730 and Rv2272, whereas His-affinity system is suitable for serological applications when denaturing agents are removed providing practical application for purified recombinant *M. tuberculosis* proteins.

KEYWORDS: tuberculosis, conserved hypothetical, GST-affinity tag, His-affinity tag, protein solubility, recombinant production

4.2 Introduction

Proteins have diverse structural and biological functions within organisms. Classification of these proteins based on these functions warrants exploration for their potential serological and therapeutic benefits, particularly among pathogenic microbes [1]. Innovative tuberculosis (TB) research is highly dependent on the discovery of novel biomarkers to facilitate diagnosis at various stages within the spectrum of disease progression, and the ability to monitor treatment response [2-5]. Concurrently, effective TB vaccine candidates capable of immune protection at all stages of disease [4-6] are urgently needed to reduce the burgeoning global TB disease burden [7].

The proteomic map of *Mycobacterium tuberculosis* contains more than 4000 encoded protein sequences, many of which are functionally uncharacterized [2-4,8-10]. A classification of the functionally uncharacterized *M. tuberculosis* proteome, retained throughout its evolutionary genetic variation, designated as conserved hypothetical (CH) proteins, presents an opportunity for biomarker discovery. These include *M. tuberculosis* proteins linked to cell- surface adherence, known as adhesins, which have been reported to be promising biomarker candidates [11-17]. These adhesins participate in TB pathogenesis by interacting with host cellular targets, eliciting an immune response in the host, or upregulating virulence factors that evade the host response to enhance mycobacterial colonization [16-20]. Investigating the adhesin function of CH proteins provides a strategic approach to support novel diagnostic or therapeutic biomarker discovery [21,22]. This approach requires in silico analysis of the gene sequences, synthesis of high yields of biologically active *M. tuberculosis* proteins, and appropriate assays to evaluate the adhesin properties, as well as assessment of diagnostic and vaccine potential [23-27].

Recombinant protein synthesis involves a pipeline of complex laboratory-based molecular techniques to produce native, biologically functional active proteins [27-29]. Despite significant advances in molecular research, *M. tuberculosis* proteins remain difficult to purify [30], particularly since many of these are located on the cell surface [15,31] of the pathogen and are hydrophobic in nature. Factors that promote inclusion body formation and hinder the purification process should be addressed by in silico evaluation of the protein sequence to improve the stability and solubility of the recombinant protein [28,32,33]. Therefore, this study aimed to establish an optimized method for the efficient molecular cloning, expression, and affinity purification of CH *M. tuberculosis* associated proteins for downstream biological, functional, and serological evaluation by comparing GST- and His-affinity tag expression systems.

4.3. Materials and Methods

4.3.1 Ethical Approval

The study was approved by the Biomedical Research Ethics Council (BREC) of the University of KwaZulu-Natal (approval ref: BREC/00001970/2020).

4.3.2. Selection of novel conserved hypothetical *Mycobacterium tuberculosis* proteins

The SPAAN tool was used to identify functionally unknown conserved hypothetical *M. tuberculosis* proteins with predicted adhesin association, as reported in a previous study [10], and selected based on their probability of adhesin index for the adhesin assay. The MycobacRv database of potential adhesin vaccine candidates was considered with the abovementioned criteria [34]. Seven *M. tuberculosis* genes, Rv3811, Rv3717, Rv3491, Rv3209, Rv2730, Rv2598, and Rv2272, were selected for systematic cloning, expression, and purification to obtain glutathione-S-transferase (GST)-tagged recombinant proteins for functional biological assessment (described in Chapter 5). Subsequently, five *M. tuberculosis* genes, Rv3811, Rv3717, Rv3491, Rv3209, and Rv2272, were cloned and expressed as 6x histidine (His)-affinity tag fusion proteins for serological evaluation (described in Chapter 6). The hydrophobicity of the selected proteins (58-92%) was measured using the Thermo Fisher Scientific peptide synthesis and proteotypic peptide analyzing tool [35] (Chapter 5, Table 5.1).

4.3.3. Bacterial strains and plasmid vector design

Whole-genome DNA extracted from *Mycobacterium tuberculosis* (strain H37Rv) using the cetyltrimethylammonium bromide lysozyme method (Sigma-Aldrich, MO, USA) generated the genomic template DNA for amplifying of the full-length *M. tuberculosis* gene sequences [36]. Two modified strains of *Escherichia coli* Dh5α and BL21 (Invitrogen, CA, USA) were cultivated on Luria-Bertani Miller (LB, Difco, Detroit, USA) and used for cloning and expression, respectively. The plasmid vectors used in this study were pGEX-6P-1 (GE Healthcare, New Jersey, USA) containing the GST-affinity tag donated by National Bioproducts Institute (Durban, South Africa) and pET-28a (+) (Sigma-Aldrich, MO, USA) containing the N-terminal His-affinity tag (Discipline of Microbiology, UKZN, Westville Campus, South Africa). Plasmid vectors were propagated in *E. coli* DH5α cells and extracted using the GeneJET plasmid mini prep kit (Thermo Fisher Scientific, MA, USA).

4.3.4. Primer design and polymerase chain reaction

Forward and reverse primers (Inqaba Biotechnical Industries, Pretoria, South Africa) were designed with compatible restriction enzyme sites using NEBcutter 2.0 (New England Biolabs, MA, USA, https://tools.neb.com/NEBcutter) and were optimally synthesized for cloning into selected vector expression systems (Table 4.1). An in silico approach was used to analyze the cloning process of the *M. tuberculosis* proteins with the pGEX-6-P-1 and pET28a plasmids (B2: Supplementary Figure 1 and 2, respectively) using SnapGene software (GSL Biotech LLC, Boston, MA, USA, https://snapgene.com). The signal peptide sequence was predicted using the SignalP 6.0 server (https://services.healthtech.dtu.dk/services/SignalP-6.0/) and removed from the respective gene sequences through primer design (Table 4.1). The gene sequences for Rv3811 and Rv2730 were predicted to be extracellular and translocated via the non-classical Sec pathway, as described in a previous study [10]. Briefly, the primers, *M. tuberculosis* (H37Rv) template DNA and Q5 high-fidelity DNA polymerase master mix (New England Biolabs, MA, USA) were prepared according to manufacturer's instructions. The gene sequences were amplified by the polymerase chain reaction (PCR) in a T100 thermal cycler (Bio-Rad Laboratories Ltd., CA, USA) at optimal annealing temperatures (Table 4.1) according to Q5 high fidelity master mix cycling conditions. The PCR product size was verified by gel electrophoresis (1.5%) using a Generuler 1Kb plus DNA ladder (Thermo Fisher Scientific, MA, USA) to distinguish sequence size and compared with a DNA template-free negative control. PCR products were purified and concentrated using the PureLink DNA Purification Mini Kit (Invitrogen, CA, USA).

4.3.5 Cloning into Plasmid Vector

The native plasmid vector and amplified PCR products were restricted using FastDigest *BamHI* and *EcoRI*/*XhoI* (Thermo Fisher Scientific, MA, USA) according to the manufacturer's protocol. Briefly, the restriction reactions were setup in separate microcentrifuge tube and incubated at 37°C for 10 minutes in a thermostatic water bath (Labotec, Durban, South Africa). The restricted products were electrophoresed on a 0.5% agarose gel for DNA gel extraction and excised under ultraviolet light. The restricted DNA was extracted from the agarose gel using a QIAquick gel extraction kit (Qiagen, Hilden, Germany) as per manufacturer's instructions. The DNA was quantified, and the purity analysed using a NanoDrop spectrophotometer (Thermo Fisher Scientific, MA, USA). The optimal vector-to-insert molar ratio was determined by an online

ligation calculator (https://www.qiagen.com/us/applications/enzymes/tools-and-calculators/ligation-calculator). The ligation reaction that was performed using the T4 ligase rapid ligation kit (Thermo Fisher Scientific, MA, USA) as per manufacturer instructions and incubated at 22°C for 30 minutes. The recombinant plasmid was transformed by heat shock at 42°C for 45 seconds into calcium chloride-competent *E. coli* DH5α cells. The transformed *E. coli* DH5α cells were grown on LB agar media supplemented with ampicillin (100 µg/mLSigma-Aldrich, MO, USA) for GST transformants and LB agar media supplemented with Kanamycin (50 µg/mL, Sigma-Aldrich, MO, USA) for His-transformants at 37°C overnight. Single colonies were selected from the agar media, and a master plate was used to subculture potential transformants. Transformants were then screened by colony PCR [37] and plasmid restriction profiling [38].

Table 4.1. Design of primer sequence of selected *Mycobacterium tuberculosis* conserved hypothetical genes.

Genes of interest	**Forward primer (5′-3′)** ***BamHI***	**Reverse primer (5′-3′)** ***EcoRI/XhoI***	**Annealing temperature (°C)**	**Fragment size**
Probable adhesins				
Rv3209	GCCGGCG-*GGATCC*-GTGGCGCTTGGCGCGGTCGCG	GCGGCGCC-*GAATTC*-CTAGCGAGTGCATGTCGCGAT	68	569 bp
Rv2598	GCCGGCG-*GGATCC*-GTGCCACTTCATCAGCTCGCC	GCGGCGCC-*GAATTC*-TCATGGACGCCATCGGCTCGT	69	495 bp
Rv2272	GCCGGCG-*GGATCC*-GTGGCTGATGACTCAAACGAC	GCGGCGCC-*GAATTC*-TCACCCGGTGATCGCCTTGGC	67	369 bp
MycobacRv selection				
Rv3811	GCCGGCG-*GGATCC*-ATGGCAGCGACCGTCGTCATC	GCGGCGCC-*GAATTC*-TCAGGTGATCGGATGCGTTGGCAGC	68	1620 bp
*Rv3491**	GCCGGCG-*GGATCC*-CTCGGACCGCCGCCGGATGGCAGT	GCGGCGCC-*GAATTC*-TTAATAGCACATCGCAATG	63	486 bp
Rv2730	GCCGGCG-*GGATCC*-ATGATGATGAACTGGCGCCAGAC	GCGGCGCC-*CTCGAG*-TCACCTACCTTGCAACCCTTC	68	477 bp
Known Adhesin				
*Rv3717**	GCCGGCG-*GGATCC*-ACCCCCGCCAACATCGCCGGCATG	GCGGCGCC-*GAATTC*-CTAACGCGCCTGGCCCTGGGTG	70	669 bp

* Signal peptide sequence removed

4.3.6 Expression of recombinant proteins

The recombinant plasmid was extracted using the GeneJET plasmid mini prep kit (Thermo Fisher Scientific, MA, USA) and transformed into *E. coli* BL21(DE3) (Invitrogen, CA, USA) competent high-expression cells. The expression cells were cultivated in LB (Difco, Detroit, USA) media supplemented with kanamycin (50 µg/mL, Sigma-Aldrich, MO, USA) for the pET-28a (+) vector and with ampicillin (100 µg/mL, Sigma-Aldrich, MO, USA) for the pGEX-6P-1 vector as an antibiotic selective marker at 37°C until an optical density (OD_{600nm}) of 0.4-0.6 was reached measured on a UV spectrophotometer (Jenway 7205, Staffordshire, UK). A pilot study was conducted for optimization by inducing expression at various optimized concentrations with isopropyl β-D-1-thiogalactopyranoside (IPTG; 0.25, 0.50, 0.75, and 1.0 mM). Optimal timeframe (1-5 hours) and temperatures (25°C, 30°C and 37°C) were monitored for the pGEX-6P-1 and low bearing temperature of 16°C for the pET-28a (+) vector (Table 4.2). The expression process was upscaled, and the over-expressed recombinant protein-containing cultures were centrifuged in a Heraeus multifuge (Thermo Fisher Scientific, Kalkberg, Germany) at 3,020 xg. The supernatant from total cell-culture was decanted and the pellet was collected, weighed on a precision scale, and stored at -20°C. The GST-affinity tag (26.4 kDa with 0.9 kDa PreScission site) and His-affinity tag (0.8 kDa with 0.6 kDa thrombin site) were combined with the protein of interest to determine the recombinant size.

4.3.7 Solubilization of the conserved hypothetical *Mycobacterium tuberculosis* proteins

A solubilization assay was used to determine the presence of the recombinant protein in cell lysate, separated into the cleared lysate (soluble fraction) or within the pellet (inclusion body).

4.3.7.1. Glutathione-S-transferase affinity-tag proteins

Initially, the solubility of the expression culture pellet was analysed using the phosphate-buffered saline (PBS) lysis buffer, achieved by chemical and enzymatic lysis (Figure 4.1). A second solubility method involved the use of an anionic detergent, 1% sodium dodecyl sulfate (SDS), in a lysis buffer [39]. In addition, mechanical lysis using a Precellys 24 homogenizer (Bertin Technologies, Paris, France) was performed at 6,800 rpm, three cycles, and a 30-second pause time to improve solubilization. The 1% sarkosyl (sodium lauroyl sarcosine) lysis solubilization

method proposed by Massiah *et al.* [33] was used to solubilize the *M. tuberculosis* conserved hypothetical protein by unfolding and mildly denaturing inclusion bodies in LB media [33]. Furthermore, the freeze-thaw method was employed to improve lysis: total lysates were placed at –80°C for three minutes, followed by three minutes of agitation on an orbital shaker and 30 seconds of vortexing, repeated three times.

Table 4.2. Optimized expression conditions of novel glutathione-S-transferase (GST) and 6x Histidine (His) – affinity tagged proteins.

Protein	Optimal IPTG (mM)	GST- affinity tag protein		His-affinity tag protein	
		Recombinant size (kDa)	Optimal conditions	Recombinant size (kDa)	Optimal conditions
Rv3811	0.25	83.6	30°C for 4 hours	58.7	16°C
Rv3717	1.0	49.5	25°C for 4 hours	24.6	16°C
Rv3491	1.0	43.8	30°C for 4 hours	18.8	16°C
Rv3209	1.0	45.5	30°C for 2 hours	20.6	16°C
Rv2598	1.0	44.0	30°C for 4 hours	19.1	16°C
Rv2730	0.25	42.9	30°C for 4 hours	17.9	No expression
Rv2272	1.0	39.4	25°C for 4 hours	14.4	No over-expression

A flow diagram is presented to outline the solubilization of the GST-affinity tag proteins using PBS lysis, SDS lysis and sarkosyl lysis methods (Figure 4.1). Briefly, the expressed pellet was thawed on ice and resuspended in 20 mL PBS lysis buffer (50 mM phosphate buffered saline pH 5.2, 1 mM phenylmethylsulfonyl fluoride, 1% Triton X-100, 1 mM dithiothretol, 1mM ethylenediaminetetraacetic acid, and 1mg/mL lysozyme). The resuspended pellet was incubated at 25°C on an end-over-end rotator (Dynal, BainBio, MD, USA) for 30 minutes to ensure homogenization and improve the lysis reaction. The suspension was then centrifuged at 4,600 xg for 30 minutes (Heraeus multifuge, Thermo Fisher Scientific, Kalkberg, Germany). The SDS method required resuspension of the pellet in 1% SDS lysis buffer (50 mM TrisCl pH 7.2, 100

mM NaCl, and 1% SDS) and incubated on ice for 30 minutes. This method employed mechanical lysis using the Precelleys 24 homogenizer (Bertina Technologies, Paris, France) at 5,500 rpm for two cycles. The sarkosyl method outlined by Massiah *et al.* [33] was performed by resuspending 1 g of pellet in 10 mL of sarkosyl lysis buffer (50 mM TrisCl pH 7.2, 200 mM NaCl, 20 mM Mercaptoethanol). The enzymatic lysis using 1 mg/mL was conducted on an end-over-end rotator (Dynal, BainBio, MD, USA) for 30 minutes at 4°C. Thereafter, 1% sarkosyl final concentration was added and placed on the Dynal end-over-end rotator (BainBio, MD, USA) for 30 minutes at 25°C. In addition, a freeze-thaw process was employed to assist lysis of the bacterial cells by placing the suspension at -80°C for 3 minutes and vortex mix for 30 seconds, repeating the cycle three times. The suspension was then centrifuged at 20,000 xg (Heraeus multifuge, Thermo Fisher Scientific, Kalkberg, Germany) for 20 minutes at 25°C. The supernatant containing the clear lysate for all methods was collected and stored at – 4°C.

4.3.7.2.His-affinity-tag proteins

Cleared lysate preparation was performed according to the Institute for BioAnalytics (IBA) Life Sciences protocol [40]. Briefly, the expression culture pellet was thawed on ice for 15 minutes and lysed under native conditions (50 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole at pH 8.0), using mechanical lysis (Precellys 24 homogenizer at 5,500 rpm for 2 cycles, with a 20-second pause between cycles) and enzymatic lysis (1 mg/mL lysozyme and 5 mg/mL DNase I) to recover the recombinant protein in a clear lysate. Centrifugation was carried out at 10,000 xg (Heraeus multifuge, Thermo Fisher Scientific, Kalkberg, Germany). for 30 minutes at 4°C. Recombinant proteins associated with inclusion bodies were solubilized via a chemically assisted denaturing lysis buffer (100 mM sodium phosphate, 10 mM tris HCl, 8 M urea at pH 8.0) (Figure 4.2). The suspension was place on an end-over-end rotator (Dynal, BainBio, MD, USA) for 60 minutes at 25°C. The clear lysate was collected after centrifugation at 10,000 xg for 20 min at 25°C. Additionally, the pellet obtained from the native solubilization method was subjected to the denaturing protocol to recover recombinant protein from the inclusion body.

4.3.8. Purification of the recombinant proteins

The GST- and His- affinity tag recombinant proteins were purified as described in Table 4.3. Post-solubilization using the sarkosyl method, the clear lysate containing sarkosyl was treated with 4%

Triton X-100 and 40 mM 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS) prior to purification to remove intercalated sarkosyl from the GST-protein. This promoted refolding of the affinity tag to enable binding to the glutathione resin.

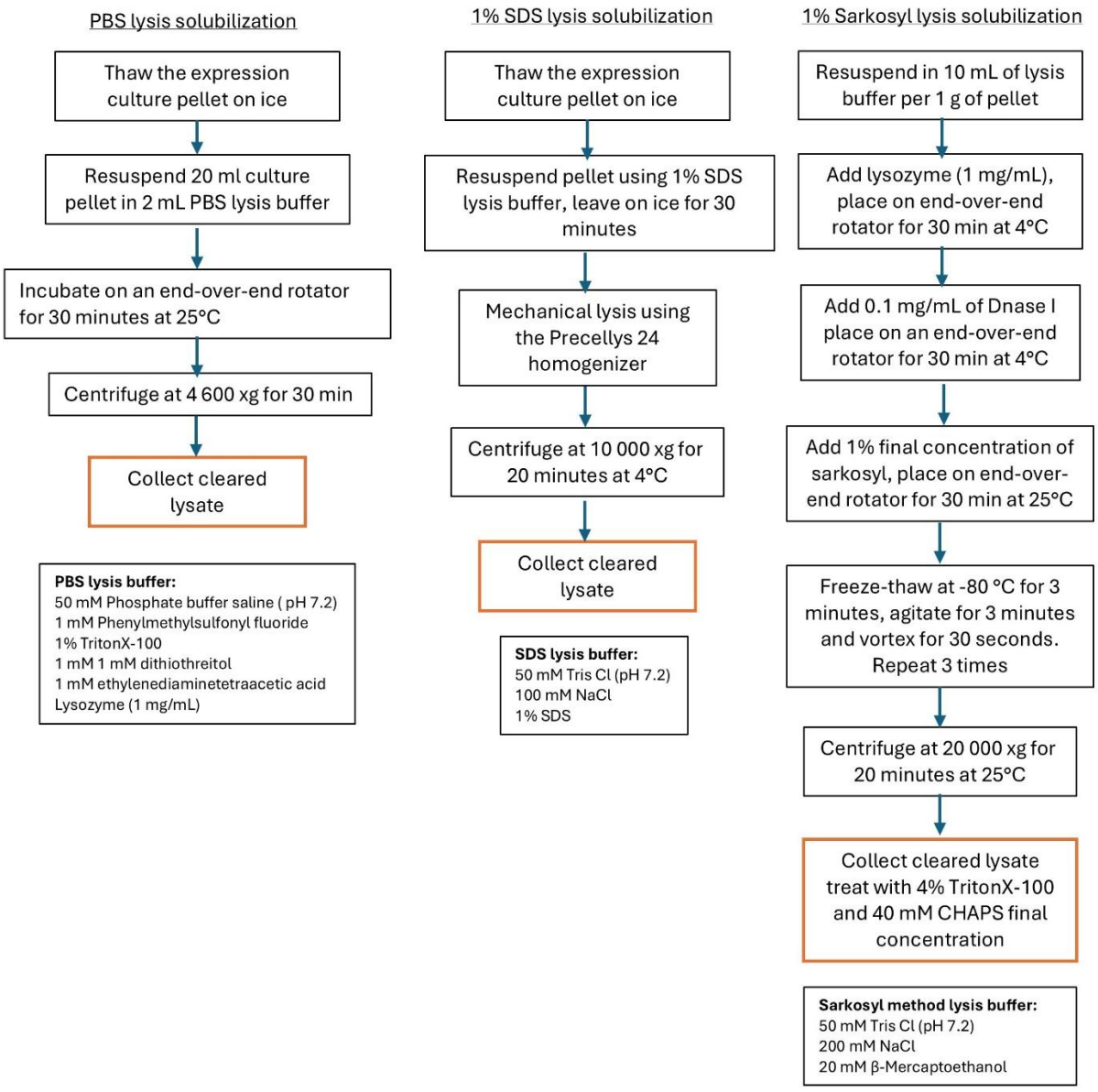


Figure 4.1. Solubilization of GST-affinity-tag proteins. The expressed GST- recombinant protein of interest is outlined, showing three methods associated with lysis and solubilization. The

PBS lysis method is an enzymatic method used to determine the solubility of the recombinant protein. The chemical lysis methods using SDS or Sarkosyl were enhanced by mechanical lysis.

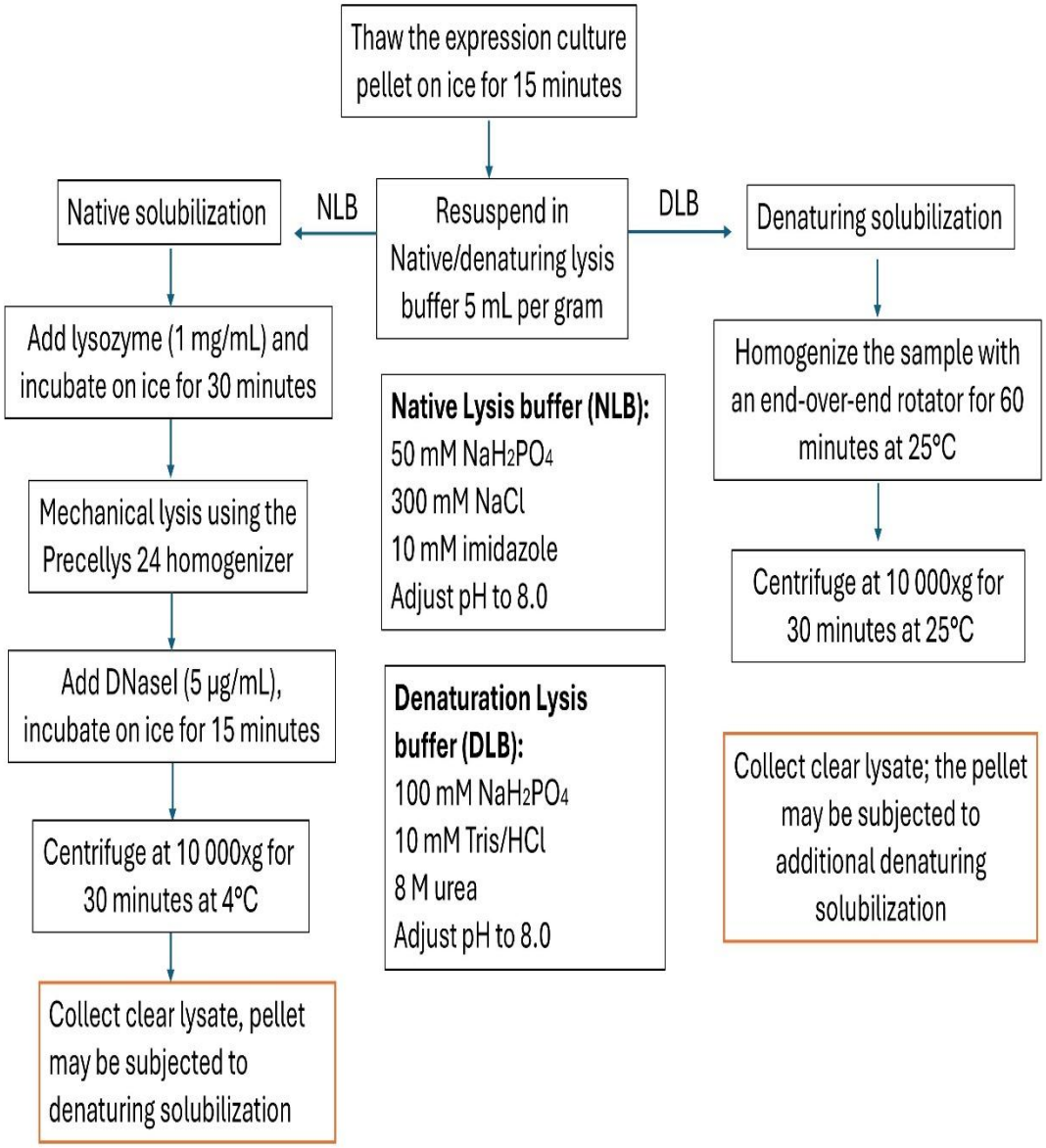


Figure 4.2. Solubilization of His-affinity-tag proteins. Expressed culture pellets were analysed to determine the solubility of the recombinant protein of interest. Solubilization was performed using two lysis buffer methods. Ideally, the native solubility is preferred; if the protein occurred in an inclusion body, a lysis buffer containing urea (8 M) was used to improve the solubility of the protein under denaturing conditions.

The GST-proteins were affinity purified by either gravity-flow chromatography using Poly-Prep gravity-flow columns (Bio-Rad Laboratories Ltd., CA, USA) or by the batch method using 15 mL centrifuge tubes (Table 4.3). Briefly, gravity-flow columns were packed with 1-2 mL of Pierce glutathione agarose (Thermo Fisher Scientific, MA, USA); the glutathione agarose contained buffer was eluted out the column for 5 minutes, and the column was washed with wash buffer (50 mM Tris, 150 mM NaCl, pH 8.0). Prepared clear lysate (1–3 mL) containing GST-affinity-tag proteins was allowed to bind to the column and placed on an end-over-end rotator (Dynal, Bainbio, MD, USA) for 1 hour at 25°C. The flow-through was collected and reapplied to bind to the column for 15 minutes. The column was washed thrice to remove unbound proteins. Three 1 mL eluted fractions were collected over 10 minute intervals by applying an elution buffer containing 10 mM reduced glutathione to the agarose to collect the GST-affinity tag recombinant protein of interest.

For the batch method purification, 3 mL of glutathione resin was added to the 15 mL tube, centrifuged at 700 xg (Eppendorf centrifuge 5702RG, Darmstadt, Germany) to separate the glutathione agarose contained supernatant buffer from the agarose, and mixed with wash buffer, then centrifuged and discarded the wash buffer. Clear lysate was added to the centrifuge tube and placed on an end-over-end rotator (Dynal, BainBio, MD, USA) for 2 hours at 25°C, followed by centrifugation at 700 x g for 2 minutes, then reapplied for 30 minutes. The samples were taken at various stages (total cell lysate, clear lysate, wash steps, and elution steps) of the purification process to monitor the concentration and purity of the rec-protein. The agarose was washed 3 times with 5 mL wash buffer, and 1 mL eluted samples were collected three times.

The His-affinity-tag proteins were purified using the batch method as described above and prepared as per the manufacturer's instructions, with HisPur Cobalt resin (Thermo Fisher, MA, USA). Native or denaturing conditions (using 8M urea) were used depending on the solubility of the recombinant proteins. Affinity-tag agarose and resin were regenerated for multiple uses.

Impurities were removed, and purified recombinant proteins were concentrated using the 30 kDa or 15 kDa Amicon Ultra Centrifugal Filters (Merck Millipore, Frankfurter, Germany). Briefly, the sample was passed through a filtered Amicon tube by centrifugation at 4,000 xg (Heraeus multifuge, Thermo Fisher Scientific, Kalkberg, Germany) for 10 minutes. Proteins ≥30 kDa were retained on the filter and collected as the retentate, while proteins ≤30 kDa were collected in the

flow-through according to the manufacturer's instructions. In addition, the His-affinity tag proteins were subjected to ethanol precipitation to remove elution buffer (containing 150 mM imidazole) and were resuspended in PBS with 0.01% SDS to prevent aggregation. The purified proteins were quantified using the Bicinchoninic Acid (BCA) assay kit (Sigma-Aldrich, MO, USA).

Table 4.3. Purification of GST/His-affinity tag proteins using various parameters to improve protein yields

Purification parameter	GST-affinity tag recombinant protein	His-affinity tag recombinant protein
Soluble sample	PBS lysis / SDS lysis with cooling and centrifugation/ Sarkosyl lysis with 4% Triton X-100 and 40 mM CHAPS	Native/ denaturation state: No urea added (native solubilization), addition of 8 M urea (denaturation solubilization)
Purification method	Gravity-flow chromatography (GFC), Batch method,	Batch method
Removal of protein impurities	GFC: based on flow rate by gravity flow, Batch: Centrifugation (700 x g)	Batch: Centrifugation (700 x g)
Purification binding agarose	Pierce™ Glutathione agarose (Thermo Fisher Scientific)	HisPur Cobalt Resin (Thermo Fisher Scientific)
Elution buffer	50 mM Tris, 150 mM NaCl; pH 8.0 containing 10 mM reduced glutathione	Native: 50 mM sodium phosphate, 300 mM sodium chloride, 150 mM imidazole; pH 7.4 Denaturing: Native with 8 M urea
Elution buffer regeneration	0.1 M Tris, 0.5 M NaCl, 0.1% SDS; pH 8.5	20 mM 2-(*N*-morpholine)-methanesulfonic acid, 0.1 M sodium chloride; pH 5.0
Removal of Elution buffer components	Amicon Ultra centrifugal tubes	Amicon Ultra centrifugal tubes, ethanol precipitation

4.3.9. Cleavage site of the glutathione-S-transferase affinity tag

The pGEX-6P-1 vector map includes a PreScission site, which is a recognition and cleavage site for Pierce™ HRV 3C Protease (Thermo Fisher, MA, USA) to separate the expressed protein from the GST-affinity tag. This was performed on GST-Rv2730 to remove the larger affinity tag for downstream serological analysis of the protein. The cleavage was performed using the HRV cleavage-in-solution method as per the manufacturer's instructions. Briefly, 2.5 µL of HRV 3C reaction buffer was added to 200 µg of recombinant protein. Two units (1 µL) of HRV 3C protease was added to the reaction tube containing ultrapure water in the 25 µL total reaction. A positive

control was compared using GST-Syk protein contained in the kit. The reaction outcomes were shown on 12% SDS-PAGE gel.

4.3.10. Visualization by sodium dodecyl sulfate polyacrylamide gel electrophoresis

Proteins were separated according to molecular weight on a mini-PROTEAN tetra electrophoresis system (Bio-Rad Laboratories Ltd., CA, USA) and analysed on the sodium dodecyl sulfate-polyacrylamide (SDS-PAGE) gel. Briefly, protein lysate samples were mixed with 2X Laemmli sample buffer (1:1 ratio, Bio-Rad Laboratories Ltd., CA, USA) in a 1.5 mL microcentrifuge tube and heated at 95°C for 5-10 minutes in a heating block to denature and linearize proteins, thereby improving migration on the SDS-PAGE gel. The prepared proteins (15 µL) were loaded into the wells electrophoresed in the gel (5% stacking gel, 10-12% resolving gel) for 120-140 minutes at 140 volts in cold Tris-Glycine-SDS (0.025 mM, 0.19 mM and 1% respectively) 1x running buffer. The gels were removed and saturated with Coomassie brilliant blue – 250 (Bio-Rad Laboratories Ltd., CA, USA) for 45 minutes while agitated to stain the proteins. Destaining solution (acetic acid, water, methanol; 1:5:4 ratio, Merck Millipore, Frankfurter, Germany) was used to remove unbound staining solution for 1 hour of agitation. Gels were placed in water overnight on an orbital shaker (Stuart Orbital Shaker, Hilton Instruments, Aberdeen, UK) to improve protein band visibility. The SDS-PAGE gel image was captured using a Syngene G: BOX (Syngene, Cambridge, UK) imaging system or color imaging with a mobile camera.

4.3.11 Detection of affinity tags using Western blot

Briefly, proteins separated by SDS-PAGE gel were transferred to a nitrocellulose membrane using a Trans-blot turbo transfer system (Bio-Rad Laboratories Ltd., CA, USA) [41]. The membrane was blocked with 5% skim milk (Merck Millipore, Frankfurter, Germany) in phosphate-buffered saline containing 0.5% Tween20 (PBS-T) for 15 minutes on the orbital shaker (Stuart Orbital Shaker, Hilton Instruments, Aberdeen, UK). The GST-proteins were detected by adding primary GST-monoclonal antibody (Dilution 1: 1 000, Invitrogen, CA, USA), and blocking buffer overnight at 4°C. The following day, the membrane was washed with 30 mL PBS-T wash buffer 4 times for 15 min each with agitation on the orbital shaker (Stuart Orbital Shaker, Hilton Instruments, Aberdeen, UK) , and a horseradish (HRP) conjugated goat anti-mouse IgG secondary antibody (Dilution 1: 5 000, Invitrogen, CA, USA). The His-affinity proteins were detected using Anti-Poly-Histidine-HRP conjugated antibody (Dilution 1: 10 000, Sigma-Aldrich, MO, USA).

The membrane was washed with 30 mL PBS-T four times with agitation on an orbital shaker (Stuart Orbital Shaker, Hilton Instruments, Aberdeen, UK) for 15 minutes per wash. Finally, Pierce 1-step Ultra TMB-Blotting solution (Thermo Fisher Scientific, MA, USA) was added for 3-5 minutes for chromogenic detection.

4.4. Results

4.4.1 Cloning of *Mycobacterium tuberculosis* gene sequences into pGEX-6P-1 (GE Healthcare) and pET-28a+ (Sigma-Aldrich)

The NEBcutter 2.0 server identified compatible restriction sites within the full-length gene sequence. The *BamHI* and *EcoRI* restriction enzymes were selected for the forward primers and reverse primers, respectively for most genes. However, Rv2730 analysis showed *XhoI* restriction site compatibility for the reverse primer. The GC-rich overhang design forms clamps that preserve the primer sequence and prevent degradation during gene amplification by PCR. The SignalP 6.0 server tool identified signal peptide sequences in Rv3717 and Rv3491, with predicted cleavage sites between amino acid residues 31–32 and 19–20, respectively (Figure 4.3). Removal of the signal peptide sequences improved downstream protein expression and purification. Secondary analysis using SecretomeP predicted that the CH proteins Rv3811 and Rv2730 are leaderless proteins translocated via the non-classical secretory pathway.

In silico cloning analysis using SnapGene software, confirmed correct insertion of the gene sequence with compatible restriction sites to the selected plasmid map and simulated successful downstream cloning and expression. The PCR amplification of the *M. tuberculosis* genomic DNA template successfully yielded PCR products of the correct nucleotide size by gel electrophoresis (B2: Supplementary Figure 3). Following ligation using a vector-to-insert molar ratio of 1:3 the presence of the gene insert in the transformed clones was confirmed by colony PCR (B2: Supplementary Figure 4). Secondary confirmation by restriction profiling of the extracted recombinant plasmid verified the correct insertion of the gene of interest (B2: Supplementary Figure 5).

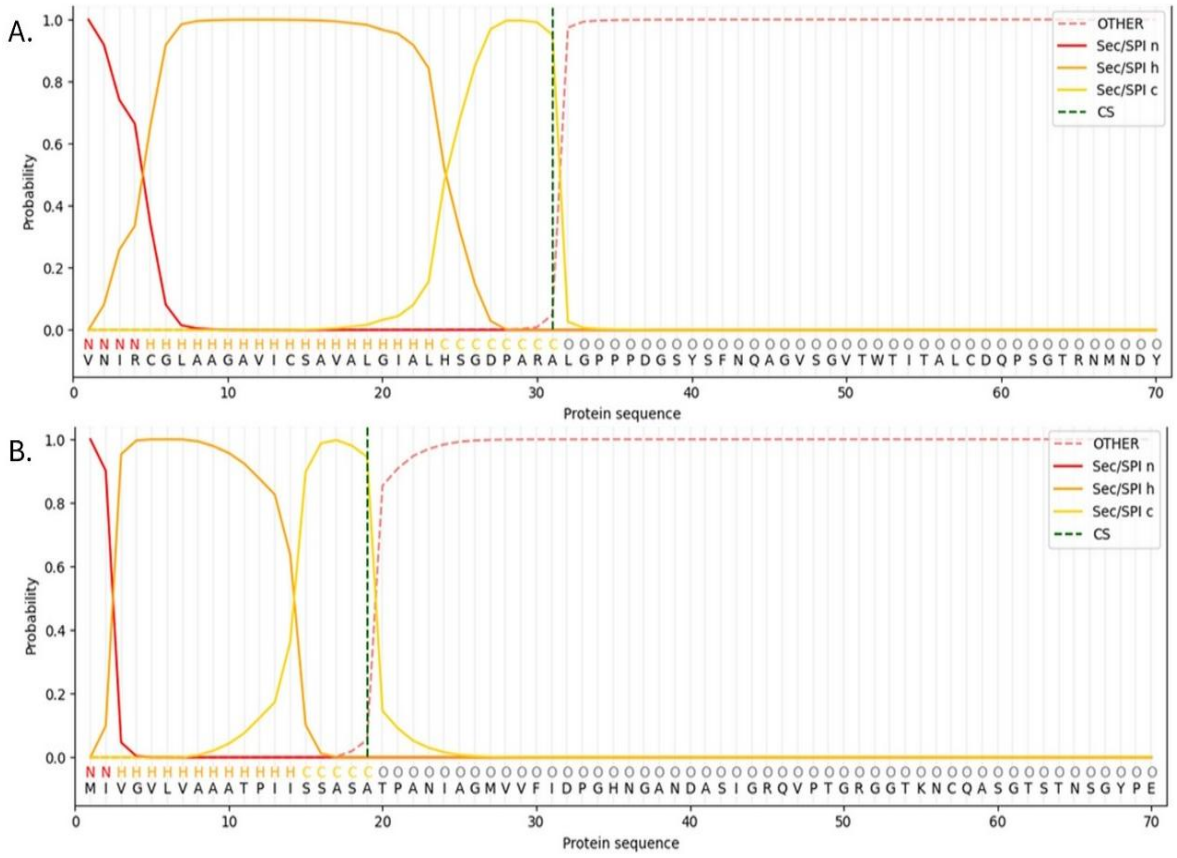


Figure 4.3. The detection of the signal peptide cleavage site (CS) of the conserved hypothetical *Mycobacterium tuberculosis* protein sequence, as shown by SignalP 6.0. A. *M. tuberculosis* amino acid sequence Rv3491 cleavage of signal peptide (Sec/SP1) between amino acid positions 31-32. **B**. *M. tuberculosis* amino acid sequence Rv3717 cleavage of signal peptide (Sec/SP1) between amino acid positions 19-20. (N: amino-terminal region, H: Hydrophobic region, C: Carboxy-terminal region, O: no signal peptide region)

4.4.2 Expression of recombinant GST/His-affinity-tag proteins

Recombinant proteins were expressed in *E. coli* BL21(DE3) transformed with pGEX-6P-1 using optimized conditions (Figure 4.4). Although the optimal growth temperature for *E. coli* is 37°C, expression of GST-affinity tag proteins occurred at lower temperatures (25–30°C). Upon visual interpretation of the SDS-PAGE gels, protein bands for Rv3811, Rv3491, Rv3209, Rv2730, and Rv2598 illustrated highest expression at 30°C, while Rv3717 showed optimal expression at 25°C (Table 4.2 and Figure 4.4). The majority of GST-proteins showed detectable expression within 4 hours of induction, whilst prolonged induction led to increased background from host protein expression. For Rv3209, the optimal induction time was 2 hours. Beyond this time (T3–T5),

protein aggregation was observed within the wells of the SDS-PAGE gel, resulting in lighter protein bands. All proteins were induced with 1.0 mM IPTG, with the exception of Rv3811 and Rv2730, which were expressed optimally at 0.25 mM IPTG (Figure 4.4).

The His-proteins were expressed at low temperature overnight (±18 hours) to slow the growth kinetics of the *E. coli* host proteins and improve the solubility of the induced His-affinity tag protein (Figure 4.5). Before induction, bacterial culture were grown to an OD $_{600nm}$ 0.4-0.6, as bacterial cells were placed on ice and IPTG induced expression was subsequently performed at 16°C. Although Rv2730 expressed well in the GST-affinity-tag system, the His-tagged fusion showed no detectable protein expression (B2: Supplementary Figures 6 and 7). Additionally, *E. coli* cultures containing the Rv2730-pET-28a recombinant plasmid exhibited slow or no growth at 16°C, 25°C, and 37°C. Furthermore, the Rv2272-pET-28a (+) recombinant plasmid did not induce overexpression of Rv2272, although expression was confirmed by Western blot (B2: Supplementary Figures 8 and 9).

4.4.3 Purification of glutathione-S-transferase/ 6x Histidine affinity-tag proteins

4.4.3.1 GST- affinity tag proteins

Initially, expressed pellets were resuspended in PBS cell lysis buffer to determine protein solubility by enzymatic and chemical treatment. The signal peptide increased the hydrophobicity of the GST-Rv3717 (50.3 kDa) and GST-Rv3491 (46.7 kDa). Recombinant protein GST-Rv3717 was not soluble in PBS lysis and 1% SDS lysis methods (B2: Supplementary Figure 10). However, GST-Rv3491 was able to be solubilized in 1% SDS but resulted in poor binding to the glutathione resin even though attempts to remove the SDS by cooling and centrifugation were performed (B2: Supplementary Figure 11). The GST-recombinant *M. tuberculosis* proteins were predominantly present in the pellet, indicating inclusion body formation, rather than in the clear lysate required for purification. In contrast, the GST native protein was solubilized in PBS lysis buffer and demonstrated a higher yield when purified using the batch method (B2: Supplementary Figure 12). The second method, using 1% SDS, was shown to improve solubilization of the GST-recombinant proteins in cleared lysate compared to PBS lysis method. The third method, using 1% sarkosyl, proved efficient in solubilizing the inclusion bodies at higher yields.

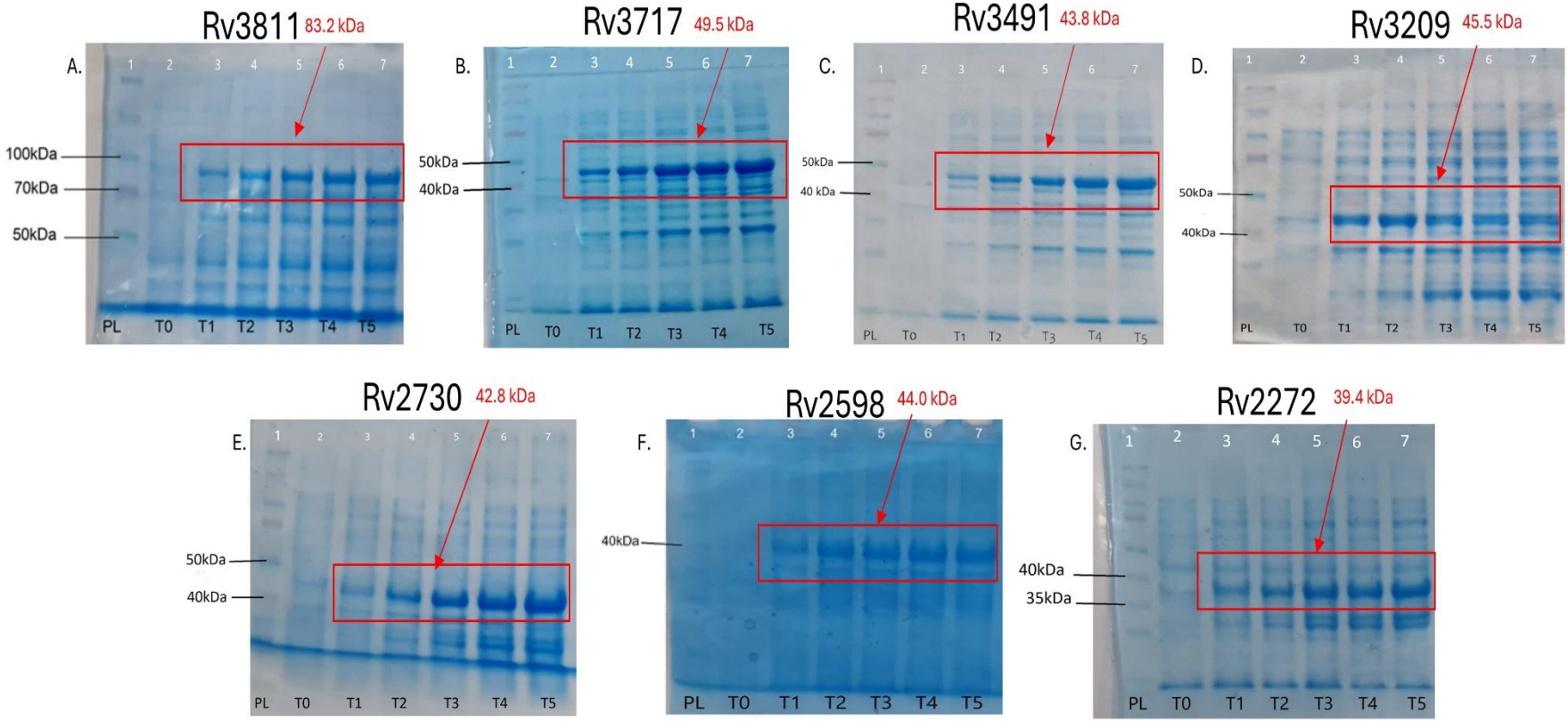


Figure 4.4. SDS-PAGE gels (10%) of pilot expressed GST-affinity-tag induced *Mycobacterium tuberculosis* conserved hypothetical proteins at various time intervals. A. GST-Rv3811 (83.6 kDa) at 0.25 mM IPTG, B. GST-Rv3717 (49.5 kDa) at 1.0 mM IPTG, C. GST-Rv3491 (43.8 kDa) at 1.0 mM IPTG, D. GST-Rv3209 (45.5 kDa) at 1.0 mM IPTG, E. GST-Rv2730 (42.9 kDa) at 0.25 mM IPTG F. GST-Rv2598 (44.0 kDa) at 1.0 mM IPTG, G. GST-Rv2272 (39.4 kDa) at 1.0 mM IPTG. Lane 1: Spectra multicolor Broad range protein ladder (PL, Thermo Fisher Scientific), Lane 2: Uninduced GST-recProtein (T_0), Lane 3: Induced GST-recProtein at 1 hour (T_1), Lane 4: Induced GST-recProtein at 2 hours (T_2), Lane 5: Induced GST-recProtein at 3 hours (T_3), Lane 6: Induced GST-recProtein at 4 hours (T_4), Lane 7: GST-recProtein at 5 hours (T_5).

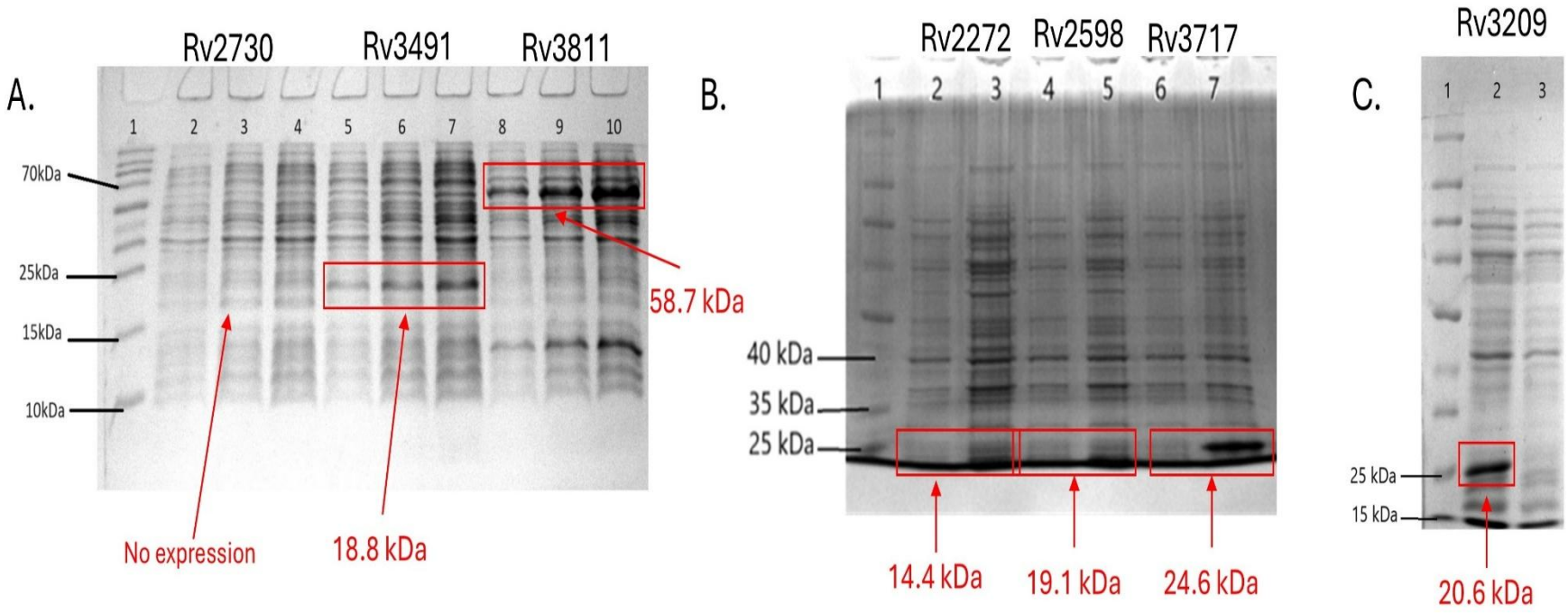


Figure 4.5. SDS-PAGE gels (12%) of pilot expressed His-affinity-tag induced *Mycobacterium tuberculosis* conserved hypothetical proteins at 16°C overnight. Optimized protein expression of A. His-Rv2730 (17.9 kDa), His-Rv3491 (18.8 kDa) and His-Rv3811 (58.7 kDa), B. His-Rv2272 (14.4 kDa), His-Rv2598(18.3 kDa), His-Rv3717 (24.6 kDa), C. His-Rv3209 (20.6 kDa) shown on SDS-PAGE gel (12% concentration). **A.** Lane 1: Spectra multicolor broad range protein ladder (Thermo Fisher Scientific), Lane 2-4: His-Rv2730 induced at 0.25 mM IPTG, Lane 5-7: recRv3491 induced at 1.0 mM IPTG, Lane 8-10: Rv3811 induced at 0.25 mM IPTG **B.** Lane 1: Broad spectrum protein ladder, Lane 2: Rv2272 uninduced, Lane 3: Rv2272 induced at 1.0 mM IPTG, Lane 4: Rv2598 uninduced, Lane 5: Rv2598 induced 1.0 mM IPTG, Lane 6: Rv3717 uninduced, Lane 7: Rv3717 induced at 1.0 mM IPTG **C**. Lane 1: Spectra multicolor Broad range protein ladder (Thermo Fisher Scientific), Lane 2: Rv3209 Induced at 1.0 mM IPTG, Lane 3: Rv3209 uninduced

The expressed pellet was lysed in PBS with an anionic detergent, 1% SDS [39]; however, this did not result in effective binding to the resin, and consequently, low protein yields were observed. Subsequently, the 1% sarkosyl solubilization method [33] was used to solubilize proteins from inclusion bodies with sufficient protein yields (Figure 4.6, A-G). The sarkosyl was neutralized using Triton X-100 and CHAPS before applying the clear lysate to the glutathione resin.

The gravity flow column purification method did not adequately retain the GST-tag protein to the resin, resulting in low yields. No improvement was observed despite extending the binding time of the clear cell lysate to the glutathione resin on the end-over-end rotator (Dynal, BainBio, MD, USA) and reapplication of the clear lysate to allow previously unbound recombinant protein to bind to the glutathione resin. The batch method purification improved purified protein yields in contrast to that of gravity-flow column purification.

4.4.3.2 Histidine affinity tag proteins

The solubility of His-affinity-tag conserved hypothetical *M. tuberculosis* proteins was analyzed using two distinct methods: native and denaturing. Solubilization was successful under denaturing conditions for Rv3811, Rv3717, Rv3491, Rv3209, and Rv2598. However, when proceeding with the purification, Rv3811 was able to be purified under native conditions (Figure 4.7A). Higher yields were obtained from the purified His-tagged protein eluents compared to the GST-tagged recombinant proteins (Figure 4.6). Subsequently, due to expression challenges, Rv2730 and Rv2272 could not be purified, although cloning was confirmed successful (B2: Supplementary Figure 6 and 8C, respectively). An effort to enzymatically cleave the PreScission site on the GST-Rv2730 for downstream applications was not successful (B2: Supplementary Figure 13). Ethanol precipitation was used to remove chemical detergents from the purification process. This method also increased protein yield, as eluents from the purification steps were pooled and the protein pellet was resuspended.

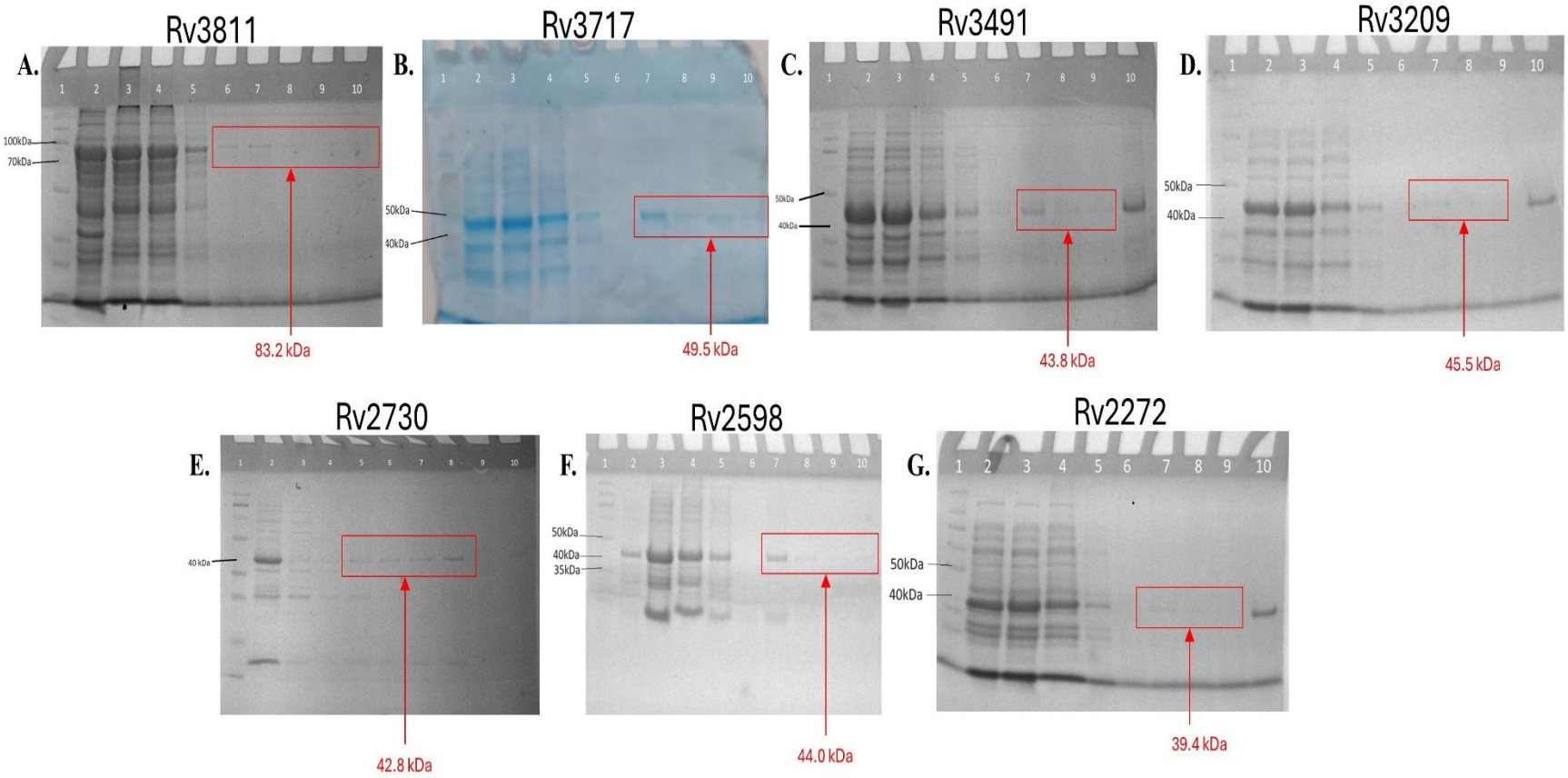


Figure 4.6. SDS-PAGE gel (12%) of purified GST-affinity tag proteins by Pierce glutathione agarose (Thermo Fisher Scientific) by batch method purification. A. GST-Rv3811 (83.6 kDa) at 0.25 mM IPTG, B. GST-Rv3717 (49.5 kDa) at 1.0 mM IPTG, C. GST-Rv3491 (43.8 kDa) at 1.0 mM IPTG, D. GST-Rv3209 (45.5 kDa) at 1.0 mM IPTG, E. GST-Rv2730 (42.9 kDa) at 0.25 mM IPTG F. GST-Rv2598 (44.0 kDa) at 1.0 mM IPTG, G. GST-Rv2272 (39.4 kDa) Lane 1: Spectra multicolor Broad range protein ladder (Thermo Fisher Scientific), Lane 2: Cell total lysate Lane 3: Soluble GST-protein (clear lysate), Lane 4: flow through, Lane 5: wash 1, Lane 6: wash 2, Lane 7: Elution 1, Lane 8: Elution 2, Lane 9: Elution 3, Lane 10: Glutathione resin (C, D, G)

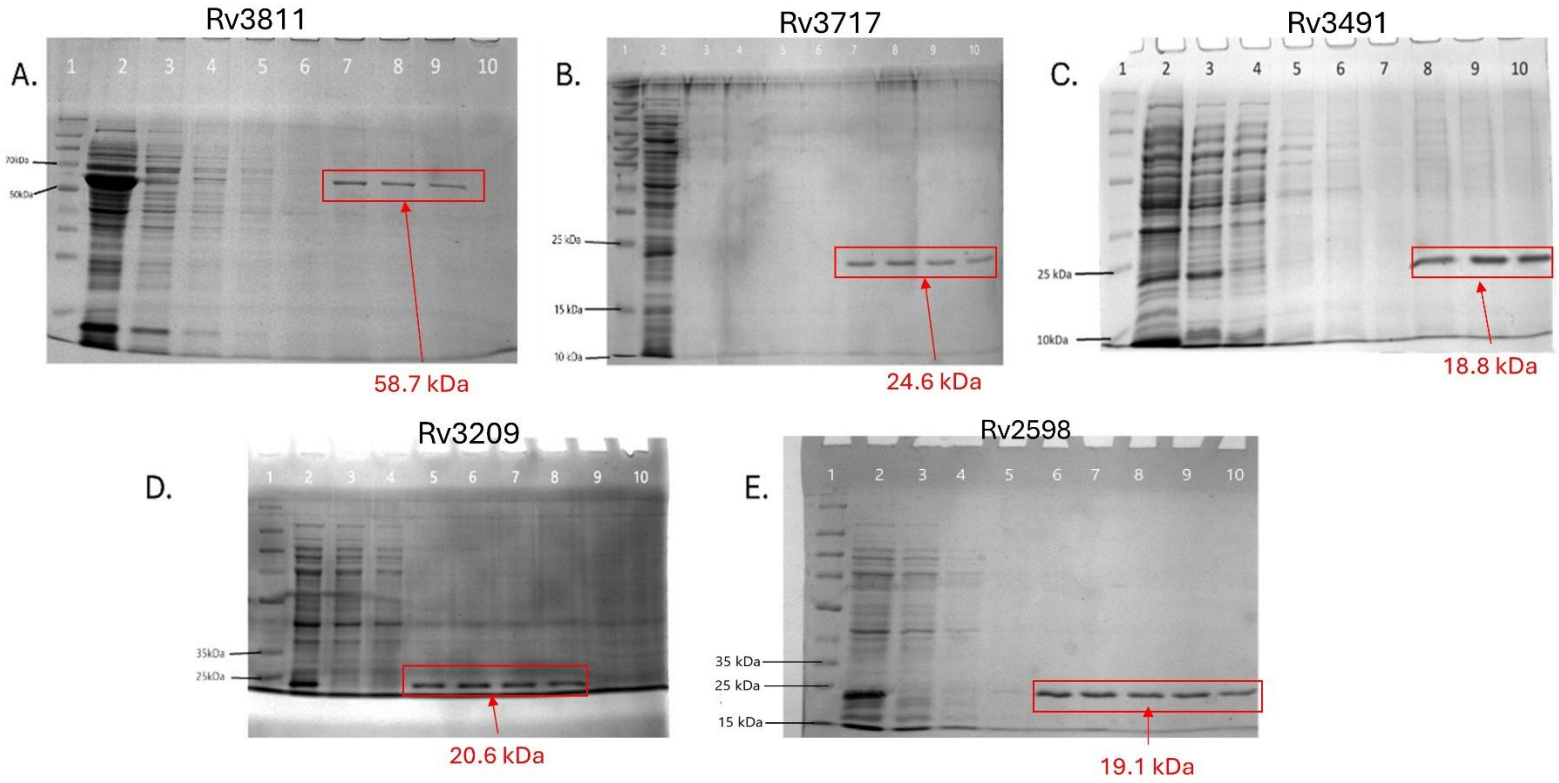


Figure 4.7. SDS-PAGE gel (12%) of purified His-affinity tag proteins using HisPur Cobalt Resin (Thermo Fisher Scientific) by batch method purification. Lane 1: Spectra multicolor Broad range protein ladder, Lane 2: Total lysate (denatured), Lane 3: Soluble protein (clear lysate), Lane 4: Flow through, A. **His-Rv3811(58.7 kDa)** Lane 5,6: wash, Lane 7-9: Elution B. His-**Rv3717 (24.6 kDa)** Lane 5,6: wash, Lane 7-10: Elution. C **His-Rv3491 (18.8 kDa)** Lane 5-7: wash, Lane 8-10: Elution **D. His-Rv3209 (20.6 kDa)** Lane 5: wash, Lane 6-8: Elution **E. Rv2598 (19.1 kDa)** Lane 5: Wash 1, Lane 6-10: elution.

4.5. Discussion

There is an urgent need for the development of novel diagnostic and therapeutic markers to reduce the burden of tuberculosis globally [11,42,43]. Recent advances in the field of bioinformatics and proteogenomic research have enabled the exploration of various components of the mycobacterial cell [44-46]. In particular, adhesin-related proteins on the cell surface of pathogenic mycobacteria have generated interest as they may act as targeted biomarkers for such applications [13,17]. In this study, a methodology was optimized for the cloning, expression, and purification of *M. tuberculosis* CH proteins that possess adhesin potential [10,17]. Two specialized vector systems, pGEX-6P-1 and pET-28a, were systematically compared for the recombinant synthesis of *M. tuberculosis* CH proteins.

The construction of the recombinant vector required systematic optimization of primer design, restriction digestion, and ligation into the cloning vector to induce high yielding expression of the affinity tag recombinant protein [47]. The *in silico* SnapGene package was used to perform DNA sequence analysis of the *M. tuberculosis* gene virtually cloned into the pGEX-6P-1 and pET-28a vectors prior to laboratory cloning [47]. While direct cloning of recombinant *M. tuberculosis* plasmids into expression strain *E. coli* (BL21) has been reported [48,49], the use of cloning-into optimized *E. coli* (DH5α) remains the recommended approach as this preserves the foreign recombinant plasmid and reduces mutations [50]. In the present study, confirmation of the recombinant plasmid was achieved by two methods, colony PCR [51,52] and restriction digestion profiling [53,54]. These methods together validated positive gene insertion for both recombinant pGEX-6P-1 and pET-28a prior to inducing expression of the *M. tuberculosis* CH protein. Lia *et al.*[32] reported that *M. tuberculosis* surface proteins containing signal peptide sequences are cleaved when translocated to the membrane. These signal peptide sequences have high hydrophobic regions that increase the hydrophobicity of protein sequence [32]. The removal of these hydrophobic signal sequences from the *M. tuberculosis* constructs reduced inclusion body aggregation [23,55], a modification that proved essential for downstream solubilization and purification, particularly, for Rv3717 and Rv3491 .

Protein expression techniques to achieve optimal yield were established by varying IPTG concentrations, induction time, and incubation temperature [56]. A previous study by Kusuma *et al.* [57] supported the abovementioned parameters for the expression of *M. tuberculosis* adhesin-related proteins. Extended induction times promoted inclusion body formation and increased

expression of *E. coli* BL21(DE3) host proteins including host-related optimal incubation temperatures. Sørensen and Mortensen [58] showed that *E. coli* growth in culture media at low temperatures reduced the formation of protein aggregates; therefore, low-temperature expression was performed for pET-28a clones. In the current study, *M. tuberculosis* proteins Rv2730 and Rv2272 were expressed well in the GST-affinity tag system in contrast to the poor expression in pET-28a. This led to the hypothesis that the larger GST-affinity tag (26 kDa) sequestered the protein into inclusion bodies, exerting a less toxic effect on the growth of the *E. coli* culture. In the His-affinity tag system, induced culture media showed poor to no growth of bacterial culture or expression of the His-Rv2730 clones (cloning confirmed in B2: Supplementary Figure 6) at various expression conditions, leading to toxic effect of membrane-associated proteins in *E. coli* [59]. The toxic recombinant plasmid interferes with the host expression thereby leading to cell death by decreased expression of the host proteins [60].

Recombinant protein solubility is considered critically important to purification, as the soluble fraction in clear lysate are able to directly interact with the affinity resin to improve purity and high yields [61]. In this study, solubilization was performed using various methods for affinity tag systems. Recombinant GST-proteins were insoluble using the PBS lysis method and remained in inclusion bodies. However, anionic detergents, using 1% SDS [39] or 1% sarkosyl [33] methods were explored to maintain the native conformation of the protein to retain functional integrity. Although both methods produced soluble protein in clear lysate, poor yields were seen at purification as SDS was shown to aggressively unfold proteins and prevented binding to the glutathione agarose. Schlanger *et al.* [39] reported that SDS disrupts molecular interactions through strong micellar binding to the protein. While it can be removed by lowering the temperature and centrifugation, the recombinant protein remains denatured with destabilized function [62]. In contrast, an analog of SDS, sarkosyl, demonstrates higher stability when forming micelles in water [63]. In related *M. tuberculosis* studies, sarkosyl was used as a non-denaturing detergent to effectively solubilize inclusion bodies of protein Rv1509 [64], shikimate dehydrogenase [65], phosphate specific transport (PstB) [66], and known adhesin SapM phosphatase [67], and cell wall anchoring motif (LysM) [26]. In a subsequent study by Massiah and Wright [33], sarkosyl was shown to be neutralized by the synergistic effect of CHAPS and Triton X-100, which improved the yield of GST-affinity tag *M. tuberculosis* CH proteins in the current study. A comparison of the

two expression systems revealed higher purification and yields with the sarkosyl method for GST-affinity-tagged proteins with a higher expectancy that the functional integrity would be maintained.

The His-affinity tag *M. tuberculosis* CH proteins could not be solubilized using the native method; therefore, a chaotropic agent, 8 M urea, was used to denature the inclusion bodies and fractionate His-proteins in clear lysate. Previous *M. tuberculosis* studies successfully purified the proteins, Rv1984c and Rv3452 [68], EspC, CFP7 and PPE57 [69], PPE17 [70], Ag85B and PstS1 [71], by solubilizing the inclusion bodies with 4-8 M urea. The denaturing method improved His-affinity-tagged protein solubility, and purification yields were comparatively higher than GST-affinity purification. Urea as a denaturing agent, did not affect the binding of the His-affinity tag to the HisPur resin. However, the urea within the buffer may destabilize the recombinant protein function and affect downstream applications [33].

Although recombinant GST-affinity proteins were comparatively low-yielding, pooling eluents and concentrating purified proteins using Amicon Ultra centrifugal tubes (15/30 kDa cutoff) was effective. The Amicon ultra centrifugal tubes were ideal for larger affinity tags as it isolated these recombinant GST-proteins. For smaller His-affinity-tagged proteins, ethanol precipitation was used to concentrate protein yields and, in particular, to facilitate the removal of urea and other chemical additives.

4.6. Conclusion

The intrinsic solubility of recombinant proteins places significant challenges relating to hydrophobicity, signal peptide sequences, inclusion body formation and host toxicity, as was experienced with the *M. tuberculosis* CH proteins with adhesin potential. This study demonstrated the use of GST and His-affinity tag systems outlining the process and application showing complementary advantages of both. Purified recombinant proteins were obtained in both expression systems. However, depending on the application, the GST-affinity tag may be adequate in the expression of membrane-bound proteins that may confer toxicity in *E. coli*. The His-affinity tag system produced higher yields of the *M. tuberculosis* CH proteins under denaturing conditions but may require further assays to determine the functional integrity of the recombinant protein (described in Chapter 6, fibronectin-binding assay). Additionally, due to its smaller molecular weight, the His-affinity tag is unlikely to interfere with downstream immunogenicity-related studies, particularly for non-toxic targets.

Limitations of the study and recommendations

Significant challenges were encountered with both expression systems. The unsuccessful expression of His-Rv2730 and inhibition of host growth kinetics, as well as the poor induction of His-Rv2272 warrants further investigation. Dumon-Seignovert *et al.*[72] proposed that toxicity and instability of recombinant proteins in *E. coli* BL21(DE3) strains may be overcome by expressing in C41/43 (DE3) strain. This is achieved by improving the stability of the recombinant plasmid by adding a *par* locus from pSC101 [72]. Additionally, eukaryotic expression systems may provide an alternative choice to express toxic *M. tuberculosis* proteins such *E. coli* BL21 (DE3) Rosetta strain [73] or yeast host *Pichia pastoris* expression systems [74]. Other parameters such as suppression of inducible promoters, cryptic promoters, and control of plasmid copy number have also been shown to inhibit the toxic effect, thereby improving expression of Rv2730 [60].

Protein insolubility, resulting from expressed aggregates of recombinant proteins forming inclusion bodies, remains a major challenge in recovery of high yields of purified product. Beyond the solubilization methods shown here such as modifying the lysis buffer and including mechanical lysis using the Precellys 24 homogenizer (Bertin Technologies, Paris, France), alternative methods for proteins with a hydrophobic physiochemistry may consider the use of chemical chaperones as folding additives, or suitable fusion tag systems suitable such as pCold1, pColdTF with chaperone pGro7 [30]. In addition to denaturing detergents SDS and sarkosyl, an alternative anionic detergent, sodium lauryl glutamate, would be recommended, [62].

Both expression systems encode cleavage sites within the plasmid construct. Downstream applications of purified proteins may be affected by the attachment of the affinity tag; therefore, cleavage of the tag is often performed. However, in this study, attempts to remove the GST-affinity tag using PreScission Protease (Thermo Fisher Scientific, MO, USA) at the cleavage site were challenging due to protein solubility issues (B2: Supplementary Figure 13). Therefore, vectors with smaller affinity tags, such as pET-28a, are recommended for immunological application.

Credit authorship contribution

RM: conceptualization, experimental work, data curation, and analysis, writing-original manuscript **SS:** conceptualization, methodology, writing-review and editing, supervision. **BP:**

Writing-review and editing, **MP:** conceptualization, methodology, writing-review and editing, funding acquisition, supervision.

Acknowledgement

This study was supported by the University of KwaZulu-Natal (UKZN) College of Health Sciences scholarship, UKZN APACHE Flagship funds, and the South African Medical Research Council Division of Research Capacity Development under the Bongani Mayosi Health Scholars Programme. The content hereof is the sole responsibility of the authors and does not necessarily represent the official views of the funding bodies.

Data availability

Data is available in Appendix B2: Chapter 4 Supplementary data of this thesis.

Reference

1. Naveed M, Chaudhry Z, Ali Z, Amjad M. Annotation and curation of hypothetical proteins: prioritizing targets for experimental study. Advancements in Life Sciences. 2018;5(3):73-87. https://doi.org/10.62940/als.v5i3.581

2. Heyckendorf J, Georghiou SB, Frahm N, *et al.* Tuberculosis treatment monitoring and outcome measures: new interest and new strategies. Clinical Microbiology Reviews. 2022;35(3):e00227-00221. https://doi.org/10.1128/cmr.00227-21

3. Jacobs R, Malherbe S, Loxton AG, *et al.* Identification of novel host biomarkers in plasma as candidates for the immunodiagnosis of tuberculosis disease and monitoring of tuberculosis treatment response. Oncotarget. 2016;7(36):57581. https://doi.org/10.18632/oncotarget.11420

4. Peloquin CA, Davies GR. The treatment of tuberculosis. Clinical Pharmacology & Therapeutics. 2021;110(6):1455-1466. https://doi.org/10.1002/cpt.2261

5. Migliori GB, Ong CWM, Petrone L, D'Ambrosio L, Centis R, Goletti D. The definition of tuberculosis infection based on the spectrum of tuberculosis disease. Breathe. 2021;17(3):210079. https://doi.org/10.1183/20734735.0079-2021

6. Vasiliu A, Martinez L, Gupta RK, *et al.* Tuberculosis prevention: current strategies and future directions. Clinical Microbiology and Infection. 2024;30(9):1123-1130. https://doi.org/10.1016/j.cmi.2023.10.023

7. World Health Organization. Global tuberculosis report 2025 Geneva: World Health Organization: WHO. 2025. Available from: https://iris.who.int/server/api/core/bitstreams/e97dd6f4-b567-4396-8680-717bac6869a9/content.

8. Li J, Wang Y, Yan L, *et al.* Novel serological biomarker panel using protein microarray can distinguish active TB from latent TB infection. Microbes and Infection. 2022;24(8):105002. https://doi.org/10.1016/j.micinf.2022.105002

9. Martin C, Aguilo N, Marinova D, Gonzalo-Asensio J. Update on TB vaccine pipeline. Applied Sciences. 2020;10(7):2632. https://doi.org/10.3390/app10072632

10. Maharajh R, Pillay M, Senzani S. A computational method for the prediction and functional analysis of potential *Mycobacterium tuberculosis* adhesin-related proteins. Expert Review of Proteomics. 2023;20(12):483-493. https://doi.org/10.1080/14789450.2023.2275678

11. Squeglia F, Ruggiero A, De Simone A, Berisio R. A structural overview of mycobacterial adhesins: Key biomarkers for diagnostics and therapeutics. Protein Science. 2018;27(2):369-380. https://doi.org/10.1002/pro.3346

12. Bisht D, Meena LS. Adhesion molecules facilitate host-pathogen interaction & mediate *Mycobacterium tuberculosis* pathogenesis. Indian Journal of Medical Research. 2019;150(1):23-32. https://doi.org/10.4103/ijmr.IJMR_2055_16

13. Silva CAM, Danelishvili L, McNamara M, *et al.* Interaction of mycobacterium leprae with human airway epithelial cells: adherence, entry, survival, and identification of potential adhesins by surface proteome analysis. infection and immunity. 2013;81(7):2645-2659. https://doi.org/10.1128/iai.00147-13

14. Simionatto S, Marchioro SB, Galli V, *et al.* Cloning and purification of recombinant proteins of *Mycoplasma hyopneumoniae* expressed in *Escherichia coli.* Protein Expression and Purification. 2010;69(2):132-136. https://doi.org/10.1016/j.pep.2009.09.001

15. Viljoen A, Dufrêne YF, Nigou J. Mycobacterial adhesion: from hydrophobic to receptor-ligand interactions. Microorganisms. 2022;10(2):454. https://doi.org/10.3390/microorganisms10020454

16. Vinod V, Vijayrajratnam S, Vasudevan AK, Biswas R. The cell surface adhesins of *Mycobacterium tuberculosis*. Microbiological Research. 2020;232:126392. https://doi.org/10.1016/j.micres.2019.126392

17. Yang H, Ma Y, Lei X, *et al.* Stopping Tuberculosis at the Gate: The Role of *M. tuberculosis* Adhesins in Infection and Intervention. Vaccines. 2025;13(7):676. https://doi.org/10.3390/vaccines13070676

18. Bo H, Moure UAE, Yang Y, *et al. Mycobacterium tuberculosis*-macrophage interaction: Molecular updates. Frontiers in Cellular and Infection Microbiology. 2023;13(1). https://doi.org/10.3389/fcimb.2023.1062963

19. McIntyre S, Warner J, Rush C, Vanderven HA. Antibodies as clinical tools for tuberculosis. Frontiers in Immunology. 2023;14(1). https://doi.org/10.3389/fimmu.2023.1278947

20. Tomlin H, Piccinini AM. A complex interplay between the extracellular matrix and the innate immune response to microbial pathogens. Immunology. 2018;155(2):186-201. https://doi.org/10.1111/imm.12972

21. Galperin MY, Koonin EV. 'Conserved hypothetical' proteins: prioritization of targets for experimental study. Nucleic Acids Research. 2004;32(18):5452-5463. https://doi.org/10.1093/nar/gkh885

22. Ren N, JinLi J, Chen Y, *et al.* Identification of new diagnostic biomarkers for *Mycobacterium tuberculosis* and the potential application in the serodiagnosis of human tuberculosis. Microbial Biotechnology. 2018;11(5):893-904. https://doi.org/10.1111/1751-7915.13291

23. Kumar S, Puniya BL, Parween S, Nahar P, Ramachandran S. Identification of novel adhesins of *m. tuberculosis* h37rv using integrated approach of multiple computational algorithms and experimental analysis. PLOS ONE. 2013;8(7):e69790. https://doi.org/10.1371/journal.pone.0069790

24. Kuo C-J, Gao J, Huang J-W, *et al.* Functional and structural investigations of fibronectin-binding protein Apa from *Mycobacterium tuberculosis*. Biochimica et Biophysica Acta (BBA) - General Subjects. 2019;1863(9):1351-1359. https://doi.org/10.1016/j.bbagen.2019.06.003

25. Viljoen A, Alsteens D, Dufrêne Y. mechanical forces between mycobacterial antigen 85 complex and fibronectin. Cells. 2020;9(3):716. https://doi.org/10.3390/cells9030716

26. Mustafa AD, Kalyanasundram J, Sabidi S, *et al.* Recovery of recombinant *Mycobacterium tuberculosis* antigens fused with cell wall-anchoring motif (LysM) from inclusion bodies using non-denaturing reagent (N-laurylsarcosine). BMC Biotechnology. 2019;19(1):27. https://doi.org/10.1186/s12896-019-0522-x

27. Tian S, Chen H, Sun T, *et al.* Expression, purification and characterization of Esx-1 secretion-associated protein EspL from *Mycobacterium tuberculosis*. Protein Expression and Purification. 2016;128:42-51. https://doi.org/10.1016/j.pep.2016.08.001

28. Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: advances and challenges. Frontiers Microbiology. 2014;5(2014). https://doi.org/10.3389/fmicb.2014.00172

29. Young CL, Britton ZT, Robinson AS. Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications. Biotechnology Journal. 2012;7(5):620-634. https://doi.org/10.1002/biot.201100155

30. Arakawa T, Akuta T, Ejima D, Tsumoto K. Solubilization and refolding of inclusion bodies by detergents. Protein Expression and Purification. 2025;236:106791. https://doi.org/10.1016/j.pep.2025.106791

31. Niederweis M, Danilchanka O, Huff J, Hoffmann C, Engelhardt H. Mycobacterial outer membranes: in search of proteins. Trends in Microbiology. 2010;18(3):109-116. https://doi.org/10.1016/j.tim.2009.12.005

32. Lai J-S, Cheng C-W, Sung T-Y, Hsu W-L. Computational comparative study of tuberculosis proteomes using a model learned from signal peptide structures. PLOS ONE. 2012;7(4):e35018. https://doi.org/10.1371/journal.pone.0035018

33. Massiah MA, Wright KM, Du H. Obtaining soluble folded proteins from inclusion bodies using sarkosyl, TritonX-100, and CHAPS: application to lb and m9 minimal media. Current Protocols in Protein Science. 2016;84(1):6.13.11-16.13.24. https://doi.org/10.1002/0471140864.ps0613s84

34. Ramsugit S, Pillay M. Identification of *Mycobacterium tuberculosis* adherence-mediating components: a review of key methods to confirm adhesin function. Iranian Journal of Basic Medical Science. 2016;19(6):579-584. https://dx.doi.org/10.22038/ijbms.2016.7124

35. Thermo Fisher Scientific. Peptide synthesis and proteotypic peptide analyzing tool 2020. Available from: https://www.thermofisher.com/za/en/home/life-science/protein-biology/peptides-proteins/custom-peptide-synthesis-services/peptide-analyzing-tool.html.

36. Larsen MH, Biermann K, Tandberg S, Hsu T, Jacobs Jr WR. Genetic manipulation of *Mycobacterium tuberculosis*. Current protocols in microbiology. 2007;6(1):10A. 12.11-10A. 12.21. https://doi.org/10.1002/9780471729259.mc10a02s6

37. Azevedo F, Pereira H, Johansson B. Colony PCR. In: Domingues L, editor. PCR: methods and protocols. New York, NY: Springer New York; 2017. 129-139. https://doi.org/10.1007/978-1-4939-7060-5_8

38. Goranson-Siekierke J, Erbe JL. Restriction analysis of recombinant plasmids. In: Casali N, Preston A, editors. *E coli* plasmid vectors: methods and applications. Totowa, NJ: Humana Press; 2003. p. 175-181. https://doi.org/10.1385/1-59259-409-3:175

39. Schlager B, Straessle A, Hafen E. Use of anionic denaturing detergents to purify insoluble proteins after overexpression. BMC Biotechnology. 2012;12(1):95. https://doi.org/10.1186/1472-6750-12-95

40. IBA Life Sciences. Expression and purification of proteins using 6xHistidine-tag: A comprehensive manual Germany: IBA Life Sciences. 2021 [accessed 2024 August 18]. Available from: https://www.iba-lifesciences.com/media/a8/ee/aa/1631860506/Manual-6xHistidine-tag.pdf&ved=2ahUKEwi4goPvo5eSAxUQVkEAHbVHOnIQFnoECBgQAQ&usg=AOvVaw2BPp0ybQQa1PyQ54mvhIYg.

41. Towbin H, Staehelin T, Gordon J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proceedings of the National Academy of Sciences. 1979;76(9):4350-4354. https://doi.org/10.1073/pnas.76.9.4350

42. Yong YK, Tan HY, Saeidi A, *et al.* Immune biomarkers for diagnosis and treatment monitoring of tuberculosis: current developments and future prospects. Frontiers in Microbiology. 2019;10: 2789. https://doi.org/10.3389/fmicb.2019.02789

43. Alonzi T, Repele F, Goletti D. Research tests for the diagnosis of tuberculosis infection. Expert Review of Molecular Diagnostics. 2023;23(9):783-795. https://doi.org/10.1080/14737159.2023.2240230

44. Dulberger CL, Rubin EJ, Boutte CC. The mycobacterial cell envelope — a moving target. Nature Reviews Microbiology. 2020;18(1):47-59. https://doi.org/10.1038/s41579-019-0273-7

45. Heiniger B, Schori C, Arefian M, *et al.* Proteogenomic discovery of novel small proteins in clinical *Mycobacterium tuberculosis* strains. bioRxiv. 2026. https://doi.org/10.64898/2026.01.27.701740

46. Chaudhuri R, Kulshreshtha D, Raghunandanan MV, Ramachandran S. Integrative immunoinformatics for mycobacterial diseases in R platform. Systems and Synthetic Biology. 2014;8(1):27-39. https://doi.org/10.1007/s11693-014-9135-9

47. Celie PHN, Parret AHA, Perrakis A. Recombinant cloning strategies for protein expression. Current Opinion in Structural Biology. 2016;38:145-154. https://doi.org/10.1016/j.sbi.2016.06.010

48. Ahmad A, Agus R, Massi MN, Handayani I, Karim H. Cloning and expression of recombinant protein CFP21 from *Mycobacterium tuberculosis* as a tuberculosis vaccine candidate. Journal of Physics: Conference Series. 2019;1341(3):032011. https://doi.org/10.1088/1742-6596/1341/3/032011

49. Agus R, Hidayah N, Sjafaraenan. Isolation and cloning of Rv3204 of *Mycobacterium tuberculosis* to *Escherichia coli* BL21 as vaccines tuberculosis : a preliminary study. Journal of Physics: Conference Series. 2019;1341(2):022010. https://doi.org/10.1088/1742-6596/1341/2/022010

50. Phue J-N, Lee SJ, Trinh L, Shiloach J. Modified *Escherichia coli* B (BL21), a superior producer of plasmid DNA compared with *Escherichia coli* K (DH5α). Biotechnology and Bioengineering. 2008;101(4):831-836. https://doi.org/10.1002/bit.21973

51. Jacobus AP, Gross J. Optimal cloning of PCR fragments by homologous recombination in *Escherichia coli*. PLOS ONE. 2015;10(3):e0119221. https://doi.org/10.1371/journal.pone.0119221

52. Kostylev M, Otwell AE, Richardson RE, Suzuki Y. Cloning should be simple: *escherichia coli* dh5α-mediated assembly of multiple dna fragments with short end homologies. PLOS ONE. 2015;10(9):e0137466. https://doi.org/10.1371/journal.pone.0137466

53. Hoseini SS, Sauer MG. Molecular cloning using polymerase chain reaction, an educational guide for cellular engineering. Journal of Biological Engineering. 2015;9(1):2. https://doi.org/10.1186/1754-1611-9-2

54. Bertero A, Brown S, Vallier L. Chapter 2 - Methods of Cloning. In: Jalali M, Saldanha FYL, Jalali M, editors. Basic Science Methods for Clinical Researchers. Boston: Academic Press; 2017. p. 19-39. https://doi.org/10.1016/B978-0-12-803077-6.00002-3

55. Gupta SK, Shukla P. Advanced technologies for improved expression of recombinant proteins in bacteria: perspectives and applications. Critical Reviews in Biotechnology. 2016;36(6):1089-1098. https://doi.org/10.3109/07388551.2015.1084264

56. Agrawal S, Padmaswari MH, Stokes AL, *et al.* Optimizing recombinant cas9 expression: insights from *E. coli* BL21(DE3) strains for enhanced protein purification and genome editing. Biomedicines [Internet]. 2024; 12(6):1226.

57. Kusuma SAF, Parwati I, Rostinawati T, *et al.* Optimization of culture conditions for mpt64 synthetic gene expression in *Escherichia coli* BL21 (DE3) using surface response methodology. Heliyon. 2019;5(11). https://doi.org/10.1016/j.heliyon.2019.e02741

58. Sørensen HP, Mortensen KK. Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli.* Microbial Cell Factories. 2005;4(1):1. https://doi.org/10.1186/1475-2859-4-1

59. Gialama D, Kostelidou K, Michou M, Delivoria DC, Kolisis FN, Skretas G. development of *Escherichia coli* strains that withstand membrane protein-induced toxicity and achieve high-level recombinant membrane protein production. ACS Synthetic Biology. 2017;6(2):284-300. https://doi.org/10.1021/acssynbio.6b00174

60. Saida F, Uzan M, Odaert B, Bontems F. Expression of highly toxic genes in *e. coli*: special strategies and genetic tools. Current Protein and Peptide Science. 2006;7(1):47-56. https://doi.org/10.2174/138920306775474095

61. Jeffery CJ. Expression, Solubilization, and purification of bacterial membrane proteins. Current Protocols in Protein Science. 2016;83(1):29.15.21-29.15.15. https://doi.org/10.1002/0471140864.ps2915s83

62. Arakawa T, Niikura T, Kita Y, Akuta T. Sodium dodecyl sulfate analogs as a potential molecular biology reagent. Current Issues in Molecular Biology. 2024;46(1):621-633. https://doi.org/10.3390/cimb46010040

63. Chisnall B, Johnson C, Kulaberoglu Y, Chen YW. Insoluble protein purification with sarkosyl: facts and precautions. In: Chen YW, editor. Structural genomics: general applications. Totowa, NJ: Humana Press; 2014. 179-186. https://doi.org/10.1007/978-1-62703-691-7_12

64. P M, Ahmad J, Samal J, *et al. Mycobacterium tuberculosis* Specific Protein Rv1509 Evokes Efficient Innate and Adaptive Immune Response Indicative of Protective Th1 Immune Signature. Frontiers in Immunology. 2021;122021. https://doi.org/10.3389/fimmu.2021.706081

65. Magalhães MLB, Pereira CP, Basso LA, Santos DS. Cloning and expression of functional shikimate dehydrogenase (EC 1.1.1.25) from *Mycobacterium tuberculosis* H37Rv. Protein Expression and Purification. 2002;26(1):59-64. https://doi.org/10.1016/S1046-5928(02)00509-0

66. Sarin J, Aggarwal S, Chaba R, Varshney GC, Chakraborti PK. B-subunit of phosphate-specific transporter from *Mycobacterium tuberculosis* Is a thermostable ATPase. Journal of Biological Chemistry. 2001;276(48):44590-44597. https://doi.org/10.1074/jbc.M105401200

67. Fernandez-Soto P, Cavet JS, Tabernero L. Expression and purification of soluble recombinant SapM from *Mycobacterium tuberculosis*. Protein Expression and Purification. 2020;174:105663. https://doi.org/10.1016/j.pep.2020.105663

68. Brust B, Lecoufle M, Tuaillon E, *et al. Mycobacterium tuberculosis* Lipolytic enzymes as potential biomarkers for the diagnosis of active tuberculosis. PLOS ONE. 2011;6(9):e25078. https://doi.org/10.1371/journal.pone.0025078

69. Arif S, Akhter M, Khaliq A, Nisa Zu, Khan IH, Akhtar MW. Serodiagnostic evaluation of fusion proteins from multiple antigens of *Mycobacterium tuberculosis* for active TB. Tuberculosis. 2021;127:102053. https://doi.org/10.1016/j.tube.2021.102053

70. Najafi A, Tafaghodi M, Sankian M, Amini Y, Ghazvini K. Cloning, expression, and refolding of ppe17 protein of *Mycobacterium tuberculosis* as a promising vaccine candidate. Iranian Journal of Medical Sciences. 2019;44(1):53-59. https://doi.org/10.30476/ijms.2019.40625

71. Dewi DNSS, Mertaniasih NM, Soedarsono, *et al.* Antibodies against native proteins of *Mycobacterium tuberculosis* can detect pulmonary tuberculosis patients. Scientific Reports. 2023;13(1):12685. https://doi.org/10.1038/s41598-023-39436-4

72. Dumon-Seignovert L, Cariot G, Vuillard L. The toxicity of recombinant proteins in *Escherichia coli*: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). Protein Expression and Purification. 2004;37(1):203-206. https://doi.org/10.1016/j.pep.2004.04.025

73. Rizzi C, Frazzon J, Ely F, *et al.* DAHP synthase from *Mycobacterium tuberculosis* H37Rv: cloning, expression, and purification of functional enzyme. Protein Expression and Purification. 2005;40(1):23-30. https://doi.org/10.1016/j.pep.2004.06.040

74. Benabdesselem C, Barbouche MR, Jarboui MA, Dellagi K, Ho JL, Fathallah DM. High level expression of recombinant *Mycobacterium tuberculosis* culture filtrate protein CFP32 in *Pichia pastoris*. Molecular Biotechnology. 2007;35(1):41-49. https://doi.org/10.1385/MB:35:1:41

Chapter 5: Validation of *Mycobacterium tuberculosis* Conserved Hypothetical Proteins with Adhesin Potential

This Chapter explores the integration of molecular docking to determine the key amino acid binding residues responsible for adhesin bonds. This study experimentally validates the extracellular matrix binding potential of the selected conserved hypothetical proteins. This manuscript was submitted to Proteomics Journal (Wiley publishing: Impact factor: 3.9) on the 18th of May 2026 and is under review.

Title: Validation of novel *Mycobacterium tuberculosis* adhesins with associated extracellular matrix components: a molecular docking and phenotypic approach

*Rivesh Maharajh[1], Manormoney Pillay[1], Ndumiso Mhlongo[2] and Sibusiso Senzani[1]**

[1]Discipline of Medical Microbiology, School of Laboratory Medicine and Medical Sciences, College of Health Sciences, University of KwaZulu-Natal, Durban 4013, South Africa

[2] Discipline of Medical Biochemistry, School of Laboratory Medicine and Medical Sciences, College of Health Sciences, University of KwaZulu-Natal, Durban 4013, South Africa

***Corresponding Author: Dr. S Senzani (PhD), Discipline of Medical Microbiology, University of KwaZulu-Natal, Durban, 4000, South Africa. Email address:** SenzaniS@ukzn.ac.za

5.1. Abstract

Adhesins are cell-surface proteins that mediate binding to host cell receptors and extracellular matrix (ECM) components. Despite enhancing the virulence of pathogenic bacteria, adhesins may have potential benefits as therapeutic biomarkers. The mycobacterial proteome contains numerous conserved hypothetical (CH) proteins, uncategorized and functionally uncharacterized gene products preserved across phylogenetic lineages. In this study, we aimed to identify and validate novel *Mycobacterium tuberculosis* adhesins from a group of CH proteins. Computational analysis was used to select *M. tuberculosis* CH proteins with potential adhesin functions. Molecular docking was performed to predict adhesin-ECM interactions to identify interactive binding sites. A total of seven recombinant (rec) proteins were cloned, expressed, and purified. The adhesion capacity of the purified rec-proteins to host ECMs (laminin, fibronectin, and collagen) was experimentally determined by indirect enzyme-linked immunosorbent assay. Computational analysis revealed high adhesin potential and predicted functional characteristics of these CH proteins. Furthermore, adhesin-ECM docking interactions exhibited high binding affinity among residues, as indicated by their docking scores and key binding residues. Experimental validation demonstrated high binding affinity of recRv3491 and recRv2730 to all ECM proteins, whereas recRv3811 showed a high binding affinity for fibronectin. In contrast, recRv3209, recRv2598, and recRv2272 showed poor binding affinity to all three ECMs. The study identified three novel *M. tuberculosis* adhesins, namely Rv3811, Rv3491, and Rv2730, and elucidated their potential structural binding sites to host cellular targets. This study revealed how molecular docking complements experimental validation in adhesin characterization, advancing our understanding of multifunctional surface proteins in pathogen-host interactions.

KEYWORDS: Conserved hypothetical, molecular docking, extracellular matrix, tuberculosis, adhesin

5.2. Introduction

According to the World Health Organization, *Mycobacterium tuberculosis*, the causative bacterial agent of the global tuberculosis epidemic, was accountable for 1.25 million deaths in 2023[1]. Currently, the only commercially available vaccine against tuberculosis is the Bacille Calmette-Guérin (BCG), a live attenuated injectable derived from *Mycobacterium bovis* [2]. The BCG vaccine has been shown to be moderately protective among children under the age of 15 years, in contrast to the diminished efficacy in adults [3,4]. The evolutionary demand for novel *M. tuberculosis* vaccine candidates, in a post- COVID-19 age, has focused on potential therapeutic vaccines at various phases of clinical trials [3,5]. Subsequently, there is still a need for developing effective novel therapeutic biomarkers that provide long-term protection against *M. tuberculosis*. In addition, the serological potential of protein-specific biomarkers may also provide a pipeline for point-of-care diagnostic development [6].

The invasive mechanism of cell-surface proteins on bacterial pathogens, called adhesins, establishes an initial interaction with host cell molecules [7,8]. These adhesin proteins have ligand-binding sites that identify host cell receptors and extracellular matrix (ECM) components [9]. Despite enhancing the virulence of pathogenic bacteria, adhesins may be beneficial as potential therapeutic biomarkers. For instance, the surface fimbriae protein FimCH complex has been successfully introduced as a vaccine for the treatment of urinary tract infections enacted by *Escherichia coli* [10]. Therefore, therapeutic development outcomes should be adequately directed within the adhesin-related protein pipeline.

Currently, various *M. tuberculosis* vaccine candidates that are in the developmental pipeline have a known adhesin protein component. These include subunit vaccines namely, 1) subunit viral delivery : TB/Flu04L (ESAT-6 and Ag85a), CHADOxA MV85A (Ag85A) and Ad Ag85A (Ag85a), 2) Adjuvant directed: GLA-SE (Rv3620,Rv3619, Rv2608 and Rv1813), IC31 (ESAT-6, Rv2660 and Ag85B), AS01E (Rv0125 and Rv1196) and DEAE (ESAT-6, CFP-10, and Ag85A) [11]. These adhesin-related proteins have emerged as promising vaccine candidates due to their adaptive responsiveness to T-lymphocyte activation and promotion of T-helper (TH)-17 differentiation [12]. The immunogenic profile of these adhesin-associated proteins supports their integration in the design of vaccine candidates.

Bioinformatic analyses can provide insight into the prediction of novel and potentially effective biomarkers. Compositional computational analysis can potentially predict adhesin potential

proteins from an array of available proteomic databases [13]. A widely used adhesin identification tool called SPAAN (a software program for the prediction of adhesin and adhesin-like proteins using neural networks), which predicts adhesin and adhesin-like proteins from the bacterial proteome sequence, has been adapted to produce a pipeline of novel *M. tuberculosis* proteins [14,15]. Molecular docking offers a predicted protein-protein interaction association between proteins of interest, particularly among potential adhesins and host cell surfaces [16]. A bioinformatic approach based on a systematic selection process enhances protein detection from datasets [15]. This strategy significantly reduces initial experimental costs and time constraints typically required for identifying novel adhesin biomarkers for vaccine or therapeutic development. In the current study, a selection of *M. tuberculosis* CH proteins [15] were explored through computational protein-protein interaction modelling to determine the binding affinity of adhesin peptides. Subsequently, potential adhesin protein candidates were molecular engineered with a glutathione S-transferase (GST) affinity tag, expressed and purified as recombinant proteins of interest. These recombinant proteins were evaluated *in vitro* using an adherence assay to determine binding to human-derived laminin, fibronectin, and collagen moieties.

5.3. Materials and Methods

5.3.1. Ethical consideration

The study was approved by the Biomedical Research Ethics Council, University of KwaZulu-Natal (approval reference: BREC/00001970/2020).

5.2.2. Bioinformatic Analyses

5.3.2.1. Adhesin specific in silico analysis

Novel adhesin proteins were selected by screening various conserved hypothetical proteins in the *M. tuberculosis H37Rv* (strain ATCC 25618) proteome as described by Maharajh *et al.* [15]. Briefly, adhesin and adhesin-like proteins were predicted using a neural network-based software, SPAAN [14]. Structural validation of the selected *M. tuberculosis* protein candidates was performed using Phyre2 homology model prediction and Pfam. Subcellular localization was assessed using tools Loctree3, Subloc, and Mycosub to identify the locality of the selected proteins as extracellular, membrane-bound, or intracellular, while secretion profiles were evaluated using

the SecretomeP tool [17]. Thereafter, CH proteins, Rv2272, Rv3811, Rv3491, Rv2730, Rv2598, and Rv3209 were selected for experimental validation. The *M. tuberculosis* protein Rv3717, which was previously demonstrated by Kumar *et al.* [18] to have adhesin-associated properties, was used as a positive control to validate assays. Hydrophobicity of the amino acid sequence of each protein was assessed by the application of the Thermo Fisher Scientific peptide synthesis and proteotypic peptide analyzing tool [19].

5.3.2.2. Protein-protein interactions

Molecular docking was performed using template-based modelling software, Hybrid Docking (HDOCK) server (http://hdock.phys.hust.edu.cn/) as an initial analytic tool to simulate the binding efficiency of protein-protein interactions by obtaining the most significant interaction based on the docking scores [20]. The HDOCK server optimizes energy minimization to determine the most stable conformation by achieving the lowest energy state for optimal orientation and simulated annealing complexes. The tertiary structure of laminin from human placenta (Sigma Aldrich, Uniprot accession no. P07924), fibronectin from human plasma-pure (Sigma Aldrich, Uniprot accession no. P02751) and collagen type 1 from human placenta (Sigma Aldrich, Uniprot accession no. P53420, Q01955, and Q14031) were determined as follows using Uniprot. The AlphaFold2 protein structures that were retrieved from the UniProt database accurately predicted the tertiary protein structure derived from the critical assessment of structural prediction (CASP14) and included energy minimization using AMBER force [21].

The ECM (host receptor) and *M. tuberculosis* protein (ligand) interaction was simulated using the HDOCK tool to generate a protein-protein complex and data of interfacial residue interactions, docking score, and docking confidence for multiple models. The resultant PDB transcript was then modified to illustrate the individual proteins (chain A and B) when viewing on the output tool. ChimeraX (version 1.8) was used to generate visual data by annotating geometric fit among proteins and the molecular hydrophobic potential of each *M. tuberculosis* protein. Subsequently, ChimeraX was used to analyze the interaction between proteins of interest and ECM components to determine sites of interaction and structural associations [22]. The PDBsum protein database generator (https://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html) was used to show interfacial residue interactions at the atomic level. Comparative analysis was further verified with phenotypic adhesin assays.

5.3.3. Phenotypic adhesin validation

5.2.3.1. Preparation of bacterial strains and plasmid vector

Mycobacterium tuberculosis (strain*: H37Rv*) whole genomic template DNA was extracted using the Cetyltrimethylammonium bromide lysozyme (Sigma-Aldrich, MO, USA) method [23]. *E. coli* (Dh5α and BL21: Invitrogen, CA, USA) were obtained from the discipline of Microbiology, School of Life Sciences at UKZN. Bacterial cells were subjected to chemical-competent heat-shock transformation to improve uptake of the plasmid vector as discussed in Chapter 4 of this thesis [24]. Plasmid vector pGEX-6p-1 (GE Healthcare, New Jersey, USA) containing the glutathione S transferase (GST-) affinity tag was obtained from National Bioproducts Institute (Durban, South Africa) and propagated at 37°C overnight in Luria Bertani Miller (Difco, Detroit, USA) media supplemented with ampicillin (100 µl/ml) as an antibiotic selective marker and extracted using GeneJet Plasmid Miniprep Kit as per manufactures instructions (Thermo Fisher Scientific, MA, USA).

5.3.3.2. Primer design

The selection of potential adhesin gene sequences was extracted from Mycobrowser using *M. tuberculosis H37Rv* reference genome (https://mycobrowser.epfl.ch/). The signal peptide nucleotide sequences were predicted using SignalP 6.0, and the Rv3717 and Rv3491 signal peptides were removed to enhance the solubility of the protein for downstream protein purification (Chapter 5, Results section, Figure 4.3). NebCutter 2.0 was used to identify restriction sites for compatible restriction enzyme selection, and Snapgene was used to perform *in silico* cloning to simulate accurate gene sequence insertion into the pGEX-6p-1 vector for downstream applications. The synthesized primer sequences (Inqaba Biotechnical Industries, Pretoria, South Africa) were designed with compatible restriction sites (refer to Chapter 4, Table 4.1).

5.3.3.3. Extracellular matrix binding enzyme-linked immunosorbent adhesin assay

The *M. tuberculosis* CH proteins were, cloned, expressed and purified and verified in Chapter 4 of this thesis. An optimized indirect ELISA was performed to determine the adhesin potential of the CH proteins by binding to three ECM components, laminin derived from human placenta (Uniprot accession no. P07924, Sigma-Aldrich, MO, USA), fibronectin from human plasma-pure (Uniprot

accession no. P02751, Sigma-Aldrich, USA) and collagen type 1 from human placenta (Uniprot accession no. P53420, Q01955, and Q14031, Sigma-Aldrich, MO, USA). The ELISA was optimized using four different concentrations (1 µg, 3 µg, 5 µg, and 10 µg) of the ECM coating protein to determine the highest sensitivity of detection.

Briefly, Covalink NH F8 module flat-bottom 96-well plates (Thermo Fisher Scientific, Roskilde, Denmark) were coated with in 100 µL of an optimized amount of 10 µg of each ECM product PBS (pH 7.2), 10 µg of Bovine Serum Albumin (BSA) control, or 50 µg of GST/GST-fused protein (reactive protein), respectively, into each designated well overnight at 4°C. Covalink plate wells were washed with 0.1% PBS-tween20 (Sigma-Aldrich, MO, USA) and blocked for 2 hours at 37°C with 1%s kim milk (Merck Millipore, Darmstadt, Germany) in 0.1% PBST. A total of standardized 50 µg of GST/GST fusion proteins were added to each well and incubated for 1 hour at 37°C and overnight at 4°C. The plate wells were washed with 0.1% PBST to remove unbound proteins and were incubated with GST tag monoclonal antibody (Thermo Fisher Scientific, MA, USA) for 1 hour at 37°C, followed by HRP-conjugated goat anti-mouse IgG secondary antibody (Dilution, 1: 5000 Invitrogen, CA, USA). Each wash step was performed with 0.1% PBST (Figure 5.1). The reaction was developed with 1-step ultra TMB-ELISA substrate (Thermo Fisher Scientific, MA, USA) and stopped with 50 µL of H_2SO_4 (2 Molar, Sigma-Aldrich, MO, USA), which also served as a blank in separate triplicate wells. The absorbance signals of each well were read at 450 nm on the iMark microplate reader (Bio-Rad Industries, CA, USA) and analyzed on Microplate Manager version 6.3 (Bio-Rad Industries, CA, USA). Analysis of the data was performed relative to the detection of the recombinant protein coating (reactive protein).

$$Binding\ intensity\ (\%) = \frac{MTB{-}ECM\ interaction\ (Abs\ at\ 450nm)}{Reactive\ Protein\ (Abs\ at\ 450nm)}\ X\ 100$$

The absorbance signals of each *M. tuberculosis*-ECM interaction were determined as a binding intensity percentage in comparison to the coated reactive protein and therefore, translated as very strong (≥ 70%), strong (50-69%), moderate (21-49%), and poor (≤ 20%).

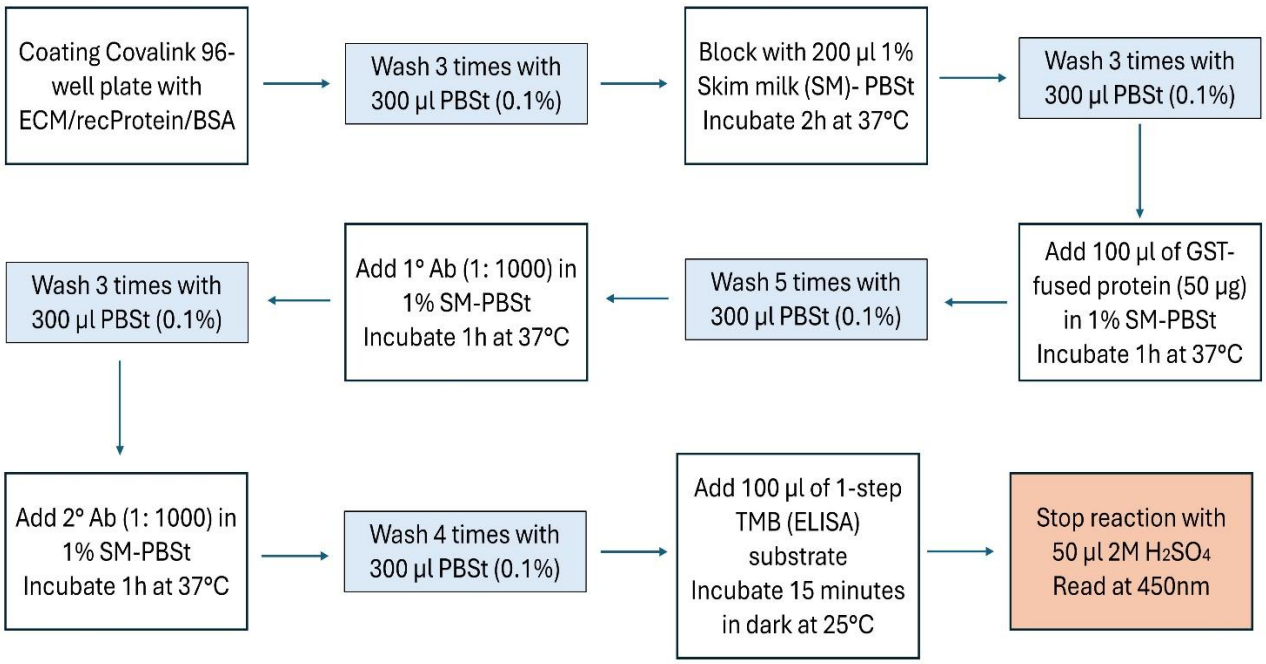


Figure 5.1. Flow diagram of the ELISA for adhesin binding detection. A schematic representation of the adhesin assay to determine experimental binding of GST-affinity tag *M. tuberculosis* proteins to extracellular matrix components.

5.3.4. Statistical analysis

Data for the bioinformatic output were analyzed using SPSS (IBM Statistical Program for Social Science for Windows, Version 24.0, Armonk, NY, IBM Corporation). The adhesin assay was performed in triplicate technical repeats to determine variability and reproducibility. Microsoft Excel was used to generate graphs with mean values and standard deviations for the adhesin comparative analysis.

5.4. Results

5.4.1. in Silico Analysis

5.4.1.1. Function and structural prediction of Adhesins

The SPAAN tool analysis was used to identify potential adhesin and adhesin-like proteins within the selected probability of adhesin ($P_{ad} \geq 0.51$). According to the compositional attributes of the SPAAN analysis for the selected conserved hypothetical proteins, Rv3717 selected with known ECM binding function [18], Rv3209 and Rv3491 exhibited high P_{ad} potential (P_{ad}: 0.777-0.795). Three proteins, Rv3811, Rv2272, and Rv2730, demonstrated moderate P_{ad} potential (P_{ad}: 0.615-

0.684). In contrast, the potential adhesin protein, Rv2598, showed a low P_{ad} potential (P_{ad}=0.552), which was still within the eligible selection criteria.

Homology modelling performed by Phyre2 analysis revealed the predictive functions of the selected CH proteins (Table 5.1). The protein Rv3209 was predicted to be a membrane protein, which was corroborated by the subcellular localization prediction. Protein Rv3491 was predicted to facilitate protein binding with the presence of a signal peptide sequence relating to the translocation of the protein to cell surface moieties. Similar to Rv3717, the potential adhesin protein, Rv3811, was predicted as an N-acetylmuramoyl-L-alanine amidase; however, it predicted hydrolase activity, in contrast to sugar-binding protein activity of the Rv3717 containing a signal peptide sequence. Although Rv3811 does not possess a signal peptide sequence, SecretomeP revealed that the protein may be translocated via a non-classical pathway in *M. tuberculosis*. The adhesin protein Rv2272 displayed structurally similar transport protein attributes by homology modelling; however, subcellular localization predicted the protein to be expressed in the cytoplasm with no translocation sequences present. The potential adhesin Rv2730 was predicted to function as a carbohydrate-binding protein with a very low Phyre2 confidence score (≤5%) with the highest-ranking protein template. However, Rv2730 was predicted to be a secreted protein via the non-classical pathway using SecretomeP prediction, similar to Rv2598.

The hydrophobicity prediction was notably high among all proteins, ranging from 58% to 92% (Table 4.1). This elevated hydrophobic characteristic was attributed to the specific amino acid composition of the *M. tuberculosis* protein, influenced by the hydrophobic index rating. Furthermore, the localization of these hydrophobic proteins enables integration in the lipid membrane regions, increasing the virulence of the *M. tuberculosis* surface complex. The molecular hydrophobic potential of each *M. tuberculosis* protein is further illustrated within the molecular docking models and reflects their predicted hydrophobicity.

5.4.1.2. Prediction and validation of adhesin-extracellular matrix docking models

An *in silico* molecular docking approach demonstrated the potential interactive model and binding affinity of extracellular matrix proteins (receptor), namely collagen, laminin, and fibronectin, against *Mycobacterium tuberculosis* adhesins (ligand). The HDOCK data output selected the highest stability of protein-protein interaction based on docking scores of the first-ranked simulation.

Table 5.1. Computational characteristic analysis of putative adhesin proteins

P_{ad} score	Gene loci	Protein size (kDa)	No. of AA	Uniprot/ Pfam annotation	Phyre2 alignment	Subcellular localization	SignalP/ SecretomeP	Hydropho-bicity (%)
0.838	Rv3717	24.84	241	N-acetylmuramoyl-L-alanine amidase Rv3717	sugar-binding protein	Cytoplasmic	Signal peptide	82.32
0.795	Rv3209	19.12	186	Conserved hypothetical threonine and proline-rich protein	membrane protein	Membrane Bound	Unknown	66.89
0.777	Rv3491	20.33	192	Uncharacterized protein	protein binding	Cytoplasmic	Signal peptide	70.44
0.684	Rv3811	57.22	539	N-acetylmuramoyl-L-alanine amidase	hydrolase	Membrane Bound	Non-classical pathway	92.09
0.661	Rv2272	12.96	122	Uncharacterized protein	Transport protein	Cytoplasmic	Unknown	82.2
0.615	Rv2730	16.48	158	Uncharacterized protein	Carbohydrate-binding protein	Secreted	Non-classical pathway	58.59
0.552	Rv2598	17.58	164	Uncharacterized protein	transferase	Cytoplasmic	Non-classical pathway	66.28

In addition to the highest confidence for binding affinity between the proteins to form highly stable interactions among receptor-ligand complexes (Table 5.2). Based on the computational efficacy of HDOCK, the docking models with the most suitable docking scores were selected to illustrate the interactions between the potential adhesin and selected ECM proteins.

However, there were 100 simulated models per docking interaction performed by HDOCK; therefore, a top 10 scoring model was assessed (supplementary information). The first docking model from the top 10 ranked complexes was selected for each *M. tuberculosis*-ECM interaction. The binding affinity ranged from (-365.94 kcal/mol) to (-240.62 kcal/mol), which was predicted to achieve the most efficient molecular interactions simulated for each receptor-ligand model. The highest binding affinity for collagen was predicted among Rv3491 (-365.94 kcal/mol, p=0.987), fibronectin with Rv2598 (-340.04 kcal/mol, p= 0.987); and laminin with Rv3717 (-337.80 kcal/mol, p=0.977); and these represented the most stable binding efficiency for each docking interaction.

Protein Rv3491 showed the highest stability for binding as the protein required the least binding energy required to form an interaction with all three of the ECMs. In contrast, the GST native protein required the highest binding energy across all ECM interactions to form stable

bonds, which translated into the weakest binding affinity when compared to the docking score data output.

Table 5.2. Representation of the most stable receptor-ligand binding affinity model according to docking scores (kcal/mol).

Protein	Receptor-ligand docking score kcal/mol (*P*-value)		
	Collagen	**Fibronectin**	**Laminin**
GST Native	-240.62 (0.860)	-269.84 (0.917)	-263.99 (0.907)
Rv2272	-346.49 (0.981)	-321.18(0.968)	-319.44 (0.967)
Rv2598	-302.19 (0.955)	-340.04 (0.978)	-303.90 (0.956)
Rv2730	-272.84 (0.921)	-316.25 (0.965)	-278.71 (0.929)
Rv3209	-314.81 (0.964)	-315.84(0.965)	-300.89 (0.953)
Rv3491	-365.94 (0.987)	-301.01(0.953)	-304.07 (0.956)
Rv3717	-303.22 (0.955)	-301.51(0.948)	-337.80 (0.977)
Rv3811	-272.73 (0.921)	-273.30 (0.922)	-284.68 (0.937)

5.4.1.3. Extracellular matrix binding complex

The crystalline structure of collagen, comprising 1690 amino acids (AA), exhibited a chicken-wire mesh-like tertiary conformation encapsulating the potential adhesin proteins of *M. tuberculosis* (Figure 5.2). The AA composition ranging from L1470 to T1690 was shown to be the dominant binding subunit region of the collagen moiety, with the exception of Rv3209 (range: D429-G680). GST-Collagen showed the lowest number (83 receptor: ligand sites) of interactive binding residues, of which four highly interactive AA and position (Y627, P73, G674, Q1674) interacted with 29 AA GST residues (Figure 5.2.A). Rv3491-Collagen simulated the highest number (110 receptor: ligand binding sites) of interactive binding residues, in which seven highly interactive AA and position (F915, P917, L1176, G1465, F1466, Q1574, M1617) interacted with 45 Rv3491 AA residues (Figure 5.2.F). In addition, Rv2730, Rv3209, and Rv3717 had greater than 100 interactive binding residues with the collagen complex.

The largest of the three ECM components, fibronectin (2477 AA), formed a tertiary structure as a homodimer composed of non-identical polypeptide chains. The common active binding site among predicted *M. tuberculosis* adhesins interacted with the fibronectin AA sequences ranging from D330-R938 and H1135-Q1594, which suggests a broad range of interactive binding residues within the complex. The *M. tuberculosis* proteins were found to bind differently to unique sites on the fibronectin protein, demonstrating the complexity of the structure with multiple binding recognition sites (Figure 5.3). Similar to collagen, Fn-GST displayed the lowest number of interactive binding residues (80 receptor: ligand sites), whereby more than half the interactive sites occur with seven fibronectin AAs (range: K25-R59). The highest receptor to ligand binding residue interaction was predicted among Rv3491 (109 bonds) and Rv3717 (107 bonds) to fibronectin. A total of 7 fibronectin AAs interacted with 37 AA within the Rv3491 sequence and Rv3717 individually, thus demonstrating a high number of interchangeable recognition binding sites to enhance adhesion.

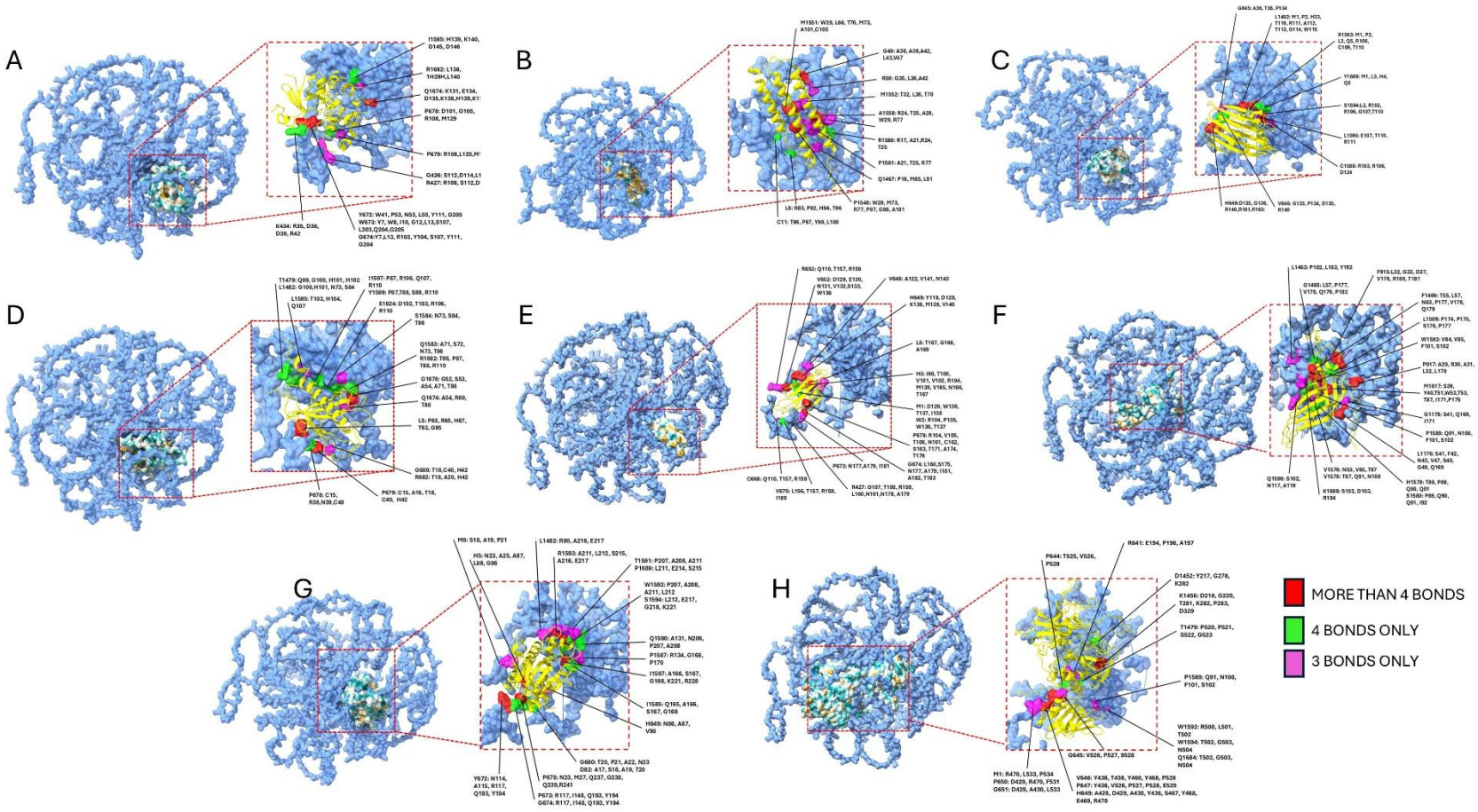


Figure 5.2. Predicted molecular binding interaction between collagen (receptor) and *Mycobacterium tuberculosis* adhesin proteins (ligand) within an ECM range of L1470-T1690. The number of receptor:ligand binding residues is indicated in brackets
A. GST (83) B. Rv2272 (97) C. Rv2598 (73) D. Rv2730 (101) E. Rv3209 (100) F. Rv3491 (110) G. Rv3717 (106) H. Rv3811 (97)

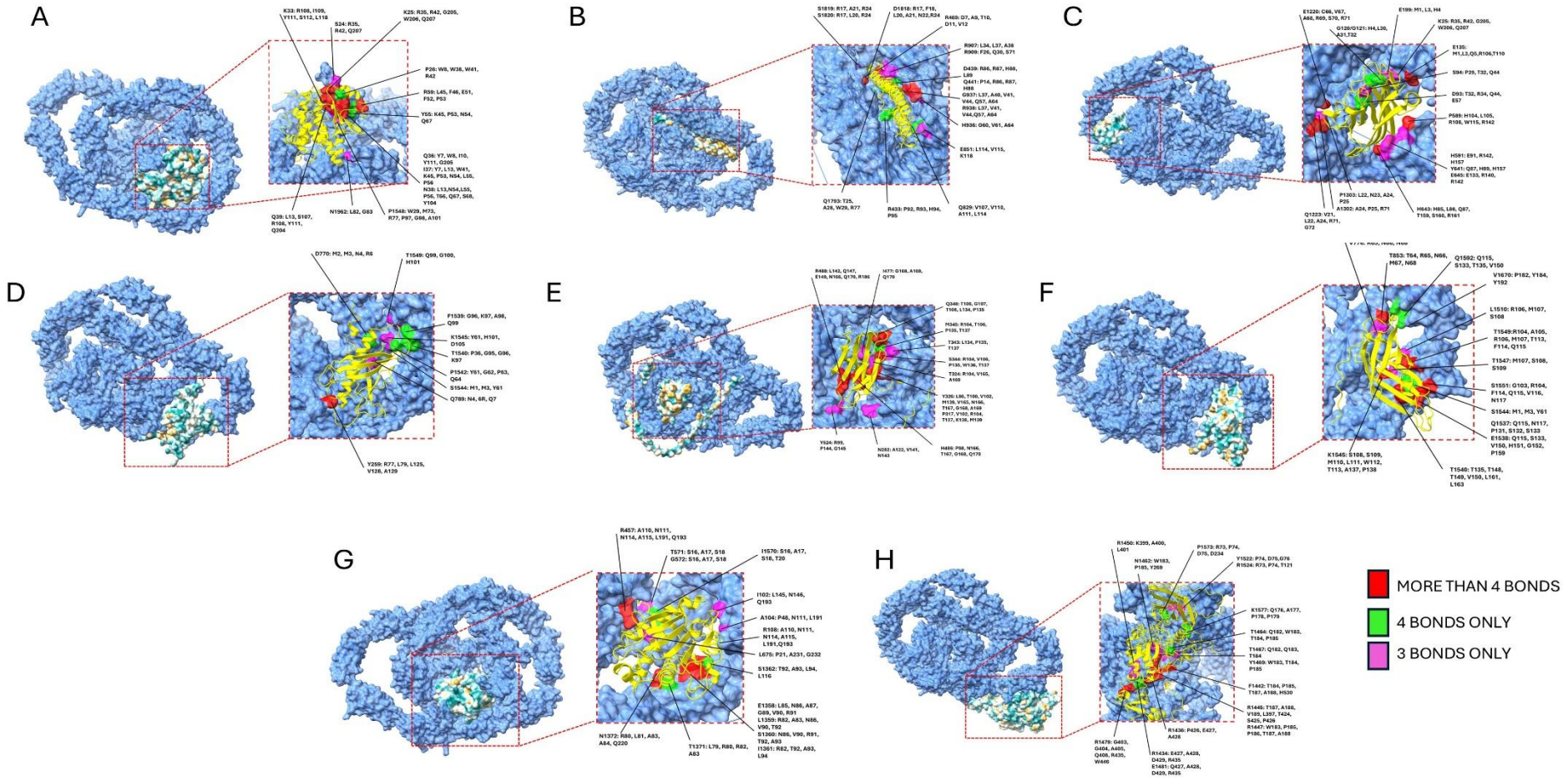


Figure 5.3. Predicted molecular interaction between fibronectin (receptor) and *Mycobacterium tuberculosis* adhesin proteins (ligand) within an ECM range of D330-R938 and H1135-Q1594. The number of receptor:ligand binding residues is indicated in brackets A. GST (80) B. Rv2272 (133) C. Rv2598 (111) D. Rv2730 (81) E. Rv3209 (104) F. Rv3491 (109) G. Rv3717 (107) H. Rv3811 (102)

The tertiary structure of laminin (1786 AA) was largely composed of heterotrimeric chains (α5β1γ1) closely associated with laminin-511. Notably, the elongated active β1 chain presented a prominent binding subunit region, whereby the selected *M. tuberculosis* proteins were shown to be encapsulated. This interaction occurred within a defined protein loop conformation, suggesting a potential mechanism of *M. tuberculosis* protein anchorage and stabilization within the laminin interface (Range: E496- V1152). The protein Rv2598 showed the lowest (73 receptor: ligand sites) number of interactive binding residues within laminin, including three laminin Aas and position (V829, F843, Q844) interacting with a total of 20 Rv2598 AA residues. However, Rv3717 was shown to have the highest number (133 receptor: ligand sites) of interactive binding residues with laminin, whereby seven laminin AA residues were capable of interacting with a total of 40 AA within the Rv3717 sequence. The active region of binding was comparable to all the ligand proteins interacting within the fold of the elongated region of the laminin moiety (Figure 5.4).

5.4.1.4. Mycobacterium tuberculosis protein binding complex

The interactive binding activity of the selected *M. tuberculosis* adhesins identified a large number of key binding residues specific to the ECM components investigated. Adhesin AA residues that interacted with more than four AA within individual ECM proteins were considered to be key binding residues. These key binding residues (Table 4) enhanced the stability of the adhesin-ECM interaction.

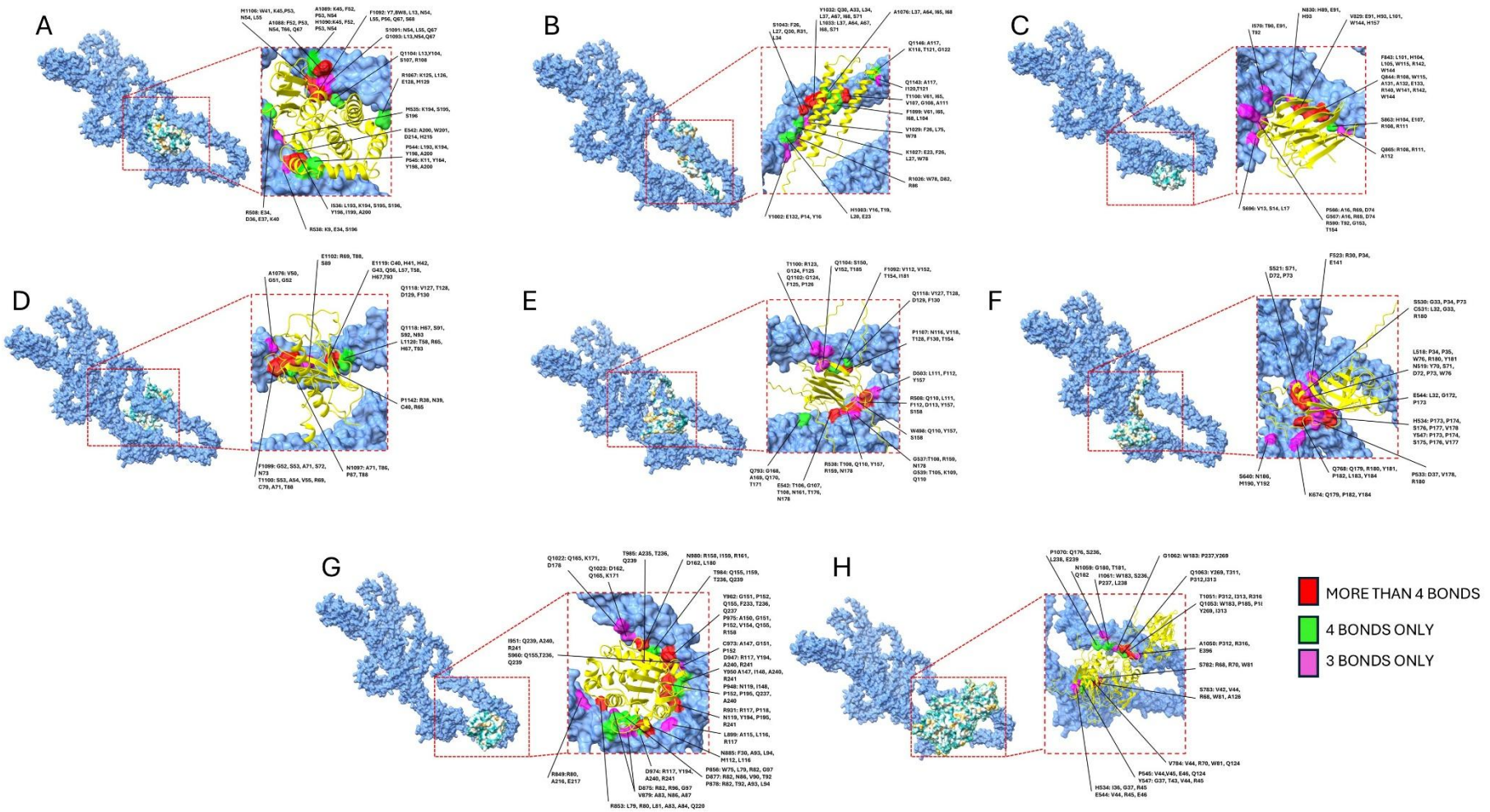


Figure 5.4. Predicted molecular interaction between laminin (receptor) and *Mycobacterium tuberculosis* adhesin proteins (ligand) within an ECM range of E496- V1152. The number of receptor:ligand binding residues is indicated in brackets A. GST (121) B. Rv2272 (86) C. Rv2598 (73) D. Rv2730 (90) E. Rv3209 (87) F. Rv3491 (78) G. Rv3717 (133) H. Rv3811 (83)

Table 5.3. Identification of *Mycobacterium tuberculosis* adhesin key binding residues interacting with extracellular matrix (ECM) proteins.

M. tuberculosis Protein (Range)	*M. tuberculosis* protein key binding residues to ECMs		
	Collagen	**Fibronectin**	**Laminin**
GST native (W8-Q207)	K138, H139, Q207	N54, R108, Y111,	K45, P53, N54, K125, K194, S196, Y198, A200,
Rv2272 (D9-T121)	T25, W29, V47, R77, R87, R93, T96, L100,	T8, D11, R17, R24, L34, L37, R87, R93	E23, F26, I68, I120, T121
Rv2598 (M1-R161)	M1, Q5, R103, E107, T110, R111	M1, Q5, T32, R34, S70, R71, R140, R142	A16, T92, H104, R108, E133, R142
Rv2730 (M1-R158)	H42, T88, G100, T103, Q107	M1, R6, P63, K97, H101, R152, Q156, R158	R38, C40, T88, R158
Rv3209 (P91-R186)	R104, D129, R159	P91, P92, R99, R104, P135, T137, Q147, N166, G168, Q170, R186	T108, Q110, F112, V118, F130, R159, N178
Rv3491 (R30-Y192)	T87, Q90, Q91, S102, Q169, P175, S176, P177, V178	R65, R104, R106, S108, S109, Q115, G179, P182, Y192	R30, L32, N68, P177, G179, R180, Y184, Y192
Rv3717 (S16-R241)	A17, S18, T20, P21, N23, R117, G168, A208, E217, A240	S16, A17, R82, A83, N86, A87, R91, T92	R82, R117, G151, P152, Q155, R158, D162, T236, Q239, R241
Rv3811 (H20-P534)	H20, L501, T502, N504, R519, S522, P528	R73, P74, Q182, W183, T184, P185, A188, D429, H530	V42, V44, R45, W81, W183, P185, P312, E396

5.4.1.5. Atomic bond interactions

The PDBsum generator analysis of the HDOCK output demonstrated non-contact bonds (Van der Waals, VDW) to be the most common atomic bond interactions among AA residues between the ECM and *M. tuberculosis* protein (Table 4.4). These bonds represented the closest atomic interactions between interface residue amino acids to form stable, albeit variable interactions within the ECM-*M. tuberculosis* protein. Although Salt bridges represented the strongest bond interactions, these remained in low numbers, followed by hydrogen bonds. The strongest atomic bond interaction among salt bridges was observed within the GST-laminin complex, which, in contrast, displayed the lowest docking score energy (-263.99kcal/mol) among all proteins to laminin. The VDW bonds were the highest among Rv2730-collagen, Rv2272-fibronectin, and Rv3717-laminin, corresponding to high binding affinity docking scores.

Table 5.4. Molecular interactive forces between *Mycobacterium tuberculosis* protein and ECM based on the PDBsum simulation.

Protein	**Number of binding interactions between protein chains**								
	Collagen			**Fibronectin**			**Laminin**		
Bond type	**Salt bridge**	**Hydrogen**	**VDW**	**Salt bridge**	**Hydrogen**	**VDW**	**Salt bridge**	**Hydrogen**	**VDW**
GST Native	1	9	166	1	8	212	6	6	258
Rv2272	3	6	173	2	9	314	2	2	192
Rv2598	1	5	201	4	10	239	-	3	169
Rv2730	1	9	238	3	8	182	2	12	232
Rv3209	-	8	236	2	8	211	1	4	211
Rv3491	-	11	219	1	16	269	-	4	155
Rv3717	1	4	134	4	13	248	2	13	293
Rv3811	2	9	213	1	10	239	2	7	182

5.4.2. Experimental Validation

5.4.2.1. Cloning, Expression, and Purification of GST fusion M. tuberculosis proteins

The cloning expression and purification of the proteins are comprehensively addressed in Chapter 4 of this thesis. The GST native and GST-fused proteins were purified by the batch method in moderate concentrations (range: 1.712-4.156 mg/mL) while maintaining their solubility and native folding conformation. Rv2598, Rv2730, and Rv3209 expressed GST dimers at 26kDa Figure 5.5. A) and were confirmed by western blot (Figure 5.5. B).

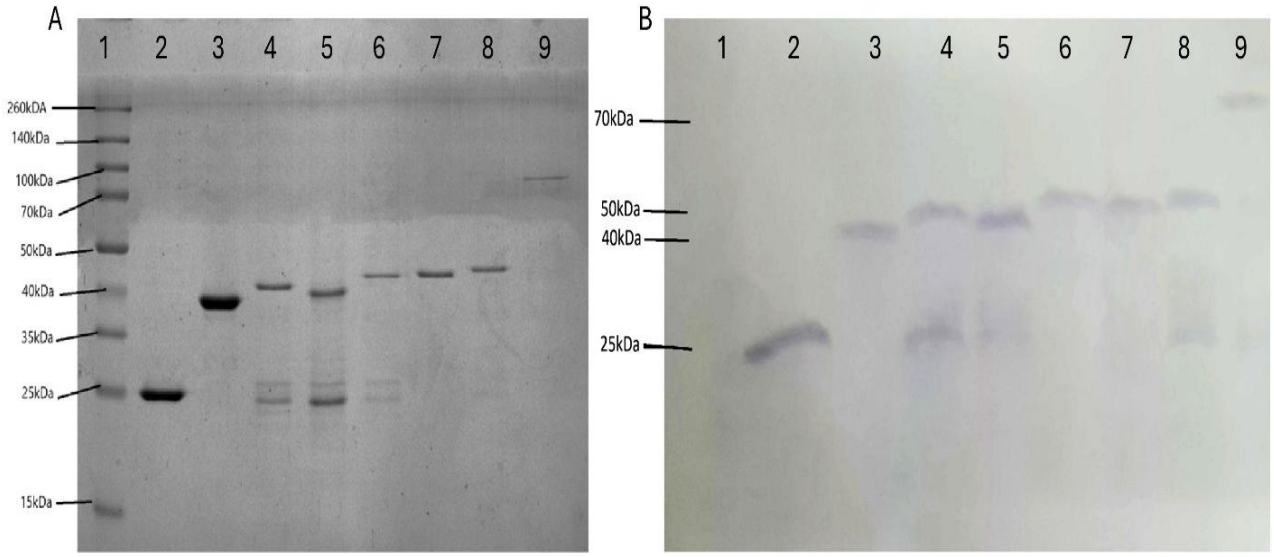


Figure 5.5. A. SDS-PAGE (15% gel) of batch method purified GST-fused proteins using sarkosyl purification method, and B. Identification of GST-fused proteins by western blot. Lane 1: Spectra multicolor broad range protein ladder (Thermo Fisher Scientific, MA, USA) Scientific, USA), Lane 2: GST protein (26.4 kDa), Lane 3: GST-Rv2272 (39.4 kDa), Lane 4: GST-Rv2598 (44.0 kDa), Lane 5: GST-Rv2730 (42.9 kDa), Lane 6: GST-Rv3209 (45.5 kDa), Lane 7: GST-Rv3491 (43.8 kDa), Lane 8:GST-Rv3717 (49.5 kDa), Lane 9: GST- Rv3811 (83.6 kDa).

5.4.2.2. Identification of extracellular matrix-Mycobacterium tuberculosis binding interaction (ELISA)

An optimized ELISA facilitated the functional characterization of adhesin candidates and enabled quantitative assessment of their binding affinities to extracellular matrix (ECM) components: fibronectin, laminin, and collagen, respectively (Figure 5.6). The binding reactivity of each interaction was normalized against a blank and compared to the negative control, which confirmed

the absence of nonspecific binding to the plate coated with BSA. Furthermore, each GST-fused protein (50 μg) served as a positive control to establish the maximum signal intensity, representing the reactivity of the corresponding GST-fused protein (RP: Reactive Protein). The native GST protein was employed solely as an affinity tag control to evaluate whether the GST moiety itself exhibited any binding activity toward the selected ECM components (Figure 5.6). The resulting absorbance at 450 nm was minimal (reactivity < 0.1), indicating negligible to no binding to the individual ECM components.

Proteins Rv2730 and Rv3491 exhibited the highest signal reactivity in the interactions with laminin, leading to the very strong binding intensities of 89% and 92%, respectively. Similarly, Rv2730 and Rv3491 proteins demonstrated strong to very strong binding intensities with collagen. In contrast, Rv2272, Rv2598, Rv3209, and Rv3811 performed poorly with respect to collagen and laminin binding. Fibronectin binding interactions demonstrated a stronger association with Rv2730, Rv3491, Rv3717, and Rv3811, indicating favorable interactions towards fibronectin in comparison to the other ECMs. Notably, Rv3717, a previously characterized adhesin, demonstrated binding properties comparable to those of newly characterized Rv3811. Overall, adhesin binding affinity was demonstrated among novel *M. tuberculosis* CH proteins Rv2730, Rv3491, and Rv3811, with the highest adhesin properties associated with Rv2730 and Rv3491 in association with all three ECM components (Table 5.5).

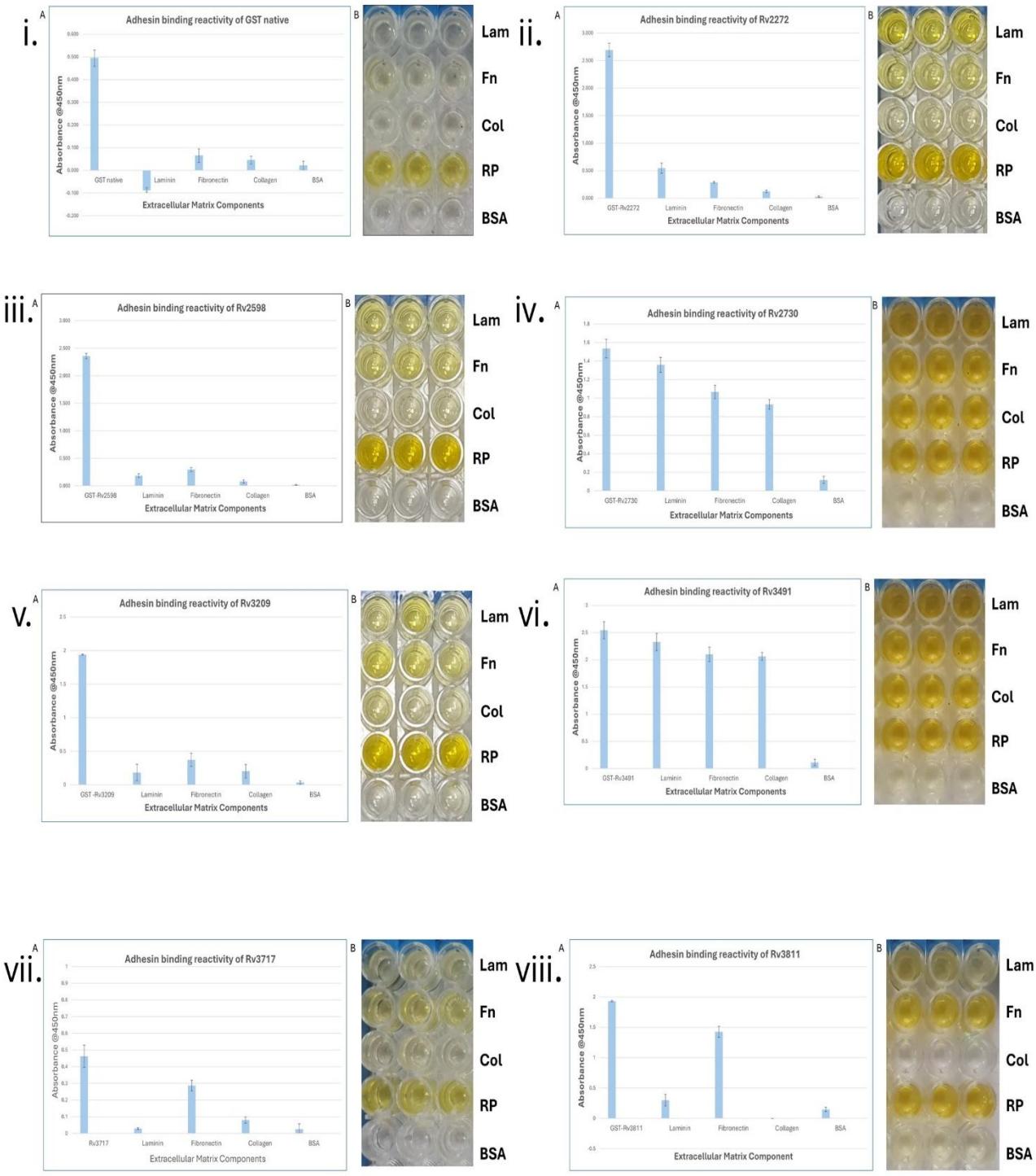


Figure 5.6. Adhesin-binding potential of conserved hypothetical *Mycobacterium tuberculosis* GST-recombinant proteins among host extracellular matrix components. The association among GST and GST-fused proteins with the respective ECM is represented (i-viii) in A. Graphical representation of adhesin detection of ECM to a purified GST-fused protein, and B. ELISA plate imaging of the binding reactivity. RP (Reactive Protein), BSA (Bovine Serum Albumin), Lam (laminin), Fn(fibronectin), and Col (collagen).

Table 5.5. Comparison of the binding intensity of *Mycobacterium tuberculosis* conserved hypothetical proteins to extracellular matrix proteins.

M. tuberculosis Protein	Extracellular Matrix Component (% binding intensity)		
	Laminin	**Fibronectin**	**Collagen**
GST Native	Poor (0)	Poor (13)	Poor (9)
Rv2272	Moderate (20)	Poor (10)	Poor (5)
Rv2598	Poor (8)	Poor (13)	Poor (3)
Rv2730	Very Strong (89)	Very Strong (70)	Strong (61)
Rv3209	Poor (9)	Poor (19)	Poor (10)
Rv3491	Very Strong (92)	Very Strong (83)	Very Strong (81)
Rv3717	Poor (6)	Strong (62)	Poor (18)
Rv3811	Poor (16)	Very Strong (74)	Poor (0)

Very strong: ≥70%, Strong: 50-69%, moderate: 21-49% and poor: ≤ 19%

5.4.2.3. Correlation between in silico analysis and phenotypic enzyme-linked immunosorbent assay findings

The comparison between the experimental outcome and the molecular interactions demonstrates strong associations with Rv3491, as this protein interacted consistently among all three extracellular matrix components. Fibronectin and Rv3491 established the strongest binding affinities presented by the docking scores and intermolecular forces supporting the interaction between these two proteins, substantiating the high adhesin potential according to the SPAAN prediction (P_{ad} = 0.777). Despite SPAAN predicting Rv3209 to have a high adhesin probability (P_{ad} = 0.795) and a high number of key binding residues identified in computational predictions, the ELISA demonstrated very poor binding affinity to collagen and laminin, with a slight reactivity to fibronectin. The molecular interactions predicted for Rv3717 showed very high binding affinity (-301.51 kcal/mol to -337.80 kcal/mol) for all ECMs, whereby laminin was shown to have the highest stability, similar to Rv3811. In contrast, phenotypic assessment of the *M. tuberculosis* protein Rv3811 exhibited a high binding intensity to fibronectin only, comparable to that of the known experimental control Rv3717. Notably, Pfam annotations functionally predicted both proteins to have amidase enzymatic activity.

The detection marker for the adhesin assay, GST- affinity tag fused to the recombinant protein, as a stand-alone native protein, demonstrated poor to low binding. These findings provide substantial evidence that GST does not contribute to adhesion-related molecular interactions, thereby validating its role as a non-adhesin control in binding assays (-240.62 kcal/mol to -263.99 kcal/mol), and this was confirmed by the low adherence to all ECM substrates in the experimental outcomes. Furthermore, *M. tuberculosis* proteins Rv2272 and Rv2598, with a P_{ad} score less than 0.661 exhibited poor binding to all ECMs in the adhesin assay, even though their docking scores demonstrated strong binding affinities to the ECM components. Subcellular localization predictions indicated that both proteins are located within the cytoplasmic membrane, potentially restricting the need to target ECM substrates. In contrast, predicted secreted Rv2730 (P_{ad} =0.615), demonstrated a comparable strong to moderate binding affinity with laminin, fibronectin, and collagen.

5.5. Discussion

This study investigated a selection of conserved hypothetical proteins from *Mycobacterium tuberculosis* for their functional roles as mycobacterial adhesins. There currently exists a paucity of literature addressing the structural or functional role of these *M. tuberculosis* conserved hypothetical proteins. However, studies have reported that (i) Rv2598 is induced during biofilm growth[27] and expressed upstream from ECM-binding Rv2599 [18], (ii) Rv3811 is predicted to have an amidase-2 domain which is non-essential for amidase activity [28], (iii) Rv3209 is computationally identified to interact with human proteins [29], (iv) Rv3491 is predicted as a parent glycopeptide similar to known *M. tuberculosis* adhesins PstS1 and Apa [30], (v) Rv2272 is a transmembrane protein that is regulated by sigma factor E regulon which is responsible for virulence and known to be more prominent in drug resistance[31], and (vi) Rv2730 is potentially involved in cell wall biosynthesis or assembly similar to *Mycobacterium bovis* bcg2743 gene [32]. These criteria motivated the exploration of the adhesin potential in the current study, which confirmed that the bioinformatic analysis correlated with several observed phenotypic findings.

The benchmark SPAAN tool demonstrated previous success for the prediction of adhesin gene function [14,15,18]. Subsequently, Kumar *et al.* [18] provided experimental evidence for the high binding affinity of *M. tuberculosis* globular proteins to ECMs, including the validation of Rv3717 [18]. In the present study, the conventional indirect ELISA confirmed the strong binding

affinity of Rv3717 to fibronectin. The P_{ad} scores reflected strong binding affinities for proteins, Rv3717 and Rv3209. However, Rv2272, Rv2598, and Rv3209 were not successful in binding to any of the three ECM proteins tested, in contrast to the computational protein-protein interaction prediction. These findings advocate that a robust systematic pipeline is necessary for a stronger and more accurate selection of adhesin proteins, not only based on the SPAAN output, but also identification of signal peptide, transmembrane, and subcellular sequence predictions as recommended by Maharajh *et al.* [15].

This current study integrated protein-protein interactions analysis with the addition of a molecular docking-based adhesin prediction approach to demonstrate the binding affinity relationship between CH proteins and ECMs. The HDOCK server proved to be a multi-analytical tool that demonstrated high accuracy in identifying key binding amino acid residues on large protein surfaces in several studies [20,33]. Similarly, the present study also identified key binding residues of the *M. tuberculosis* potential adhesin and host ECM components, respectively. The molecular docking approach has been widely used for identifying mycobacterial epitope regions by interacting with cellular response mechanisms, which include major histocompatibility complexes related to vaccine potential [34,35].

Subsequently, studies have shown the use of molecular docking in drug therapy and resistance [36,37]. However, this docking approach to elucidate adhesin potential has not previously been employed until now [36]. The inclusion of a previously identified adhesin, Rv3717 [18], validated the high binding affinities among all ECMs and further confirmed the experimental outcome. The majority of CH proteins were identified as high fibronectin-binding proteins, a class of adhesins that have been widely explored [38]. Collagen forms a high abundance in the host and can associate with pathogenic proteins called anchored collagen adhesins [16]. However, these proteins showed poor collagen-binding strength with the exception of Rv2730 and Rv3491. The docking model adapted for *Streptococcus mutans* collagen-binding proteins was established so that multiple binding residues were present to support binding to collagen [39]. This reflected in this study outcome.

Collagen binding and degradation [40] play an initial role in *M. tuberculosis* pathogenesis as collagen I is a major constituent of lung parenchyma and promotes the formation of granulomas during initial infection [41], therefore, these proteins may very well contribute intrinsically to pathogenesis [42]. A large number of known adhesins bind to laminin, namely early secretory

antigenic target 6 (ESAT6) [43], curli pili (MTP) [44], malate synthase [45] and protein kinase D (PknD) [46] and potentially have the ability to modulate the inflammatory response among lung surfactants in alveolar cells [47]. The current study showed that predicted secreted Rv2730 and cytoplasmic-bound Rv3491, even though both possess signal peptide sequences for translocation to the cell surface, were prominent laminin-binding proteins, thereby potentially increasing the virulence of the *M. tuberculosis* pathogen adherence to lung laminin-saturated epithelial cells.

Mycobacterium tuberculosis membrane-associated proteins are structurally composed of hydrophobic regions that account for 73% of surface proteins [48]. The synthesis and purification of the *M. tuberculosis* recombinant proteins posed significant challenges, as the majority of these proteins were shown to contain hydrophobic constituents [9,46,49]. The ChimeraX visualization tool [50] generated areas of lipophilicity potential that supported the hydrophobic nature of these probable surface proteins. This hydrophobicity posed significant challenges for protein solubility, which proved to be a critical determinant in achieving adequate yields of purified protein for downstream applications. To address the insolubility, inclusion bodies [51] were subjected to solubilization under anionic denaturing conditions using SDS detergent [52]. However, this approach resulted in markedly low recovery yields, underscoring the limitations of SDS-based solubilization for these hydrophobic protein targets. To overcome these challenges, successful yields were obtained by the sarkosyl method proposed by Massiah *et al.* [25]. This involved neutralization of the sarkosyl detergent by Triton X-100 prior to the purification of the GST-recombinant protein. This procedure ensured the retention of the biological function as demonstrated by the adherence properties of each refolded recombinant protein.

Although approximately thirty-four *M. tuberculosis* adhesin proteins have been identified to date [15], the presence of numerous CH proteins suggests that the full selection of adhesins may be significantly undetermined. These uncharacterized proteins may play pivotal roles in host-pathogen interactions and warrant further investigation to elucidate their adhesin contributions to *M. tuberculosis* adhesion and virulence [8,15,38]. In addition, immunogenic adhesin biomarkers are rather favorable for diagnostic and therapeutic development [11,53]. Furthermore, this study has established a robust pipeline that integrates a series of functional and structural bioinformatics that may improve the experimental validation of novel *M. tuberculosis* adhesin candidates.

While this study provides valuable insights into the adhesin potential of selected *M. tuberculosis* CH proteins, several limitations remain. Firstly, further in silico validation of protein-protein interactions employing molecular dynamics simulations is necessary to assess the stability of these interactions under varying physiological conditions, including fluctuations in pH, temperature, and solvent composition, particularly related to host environmental conditions. Additionally, multiple docking models yielded sub-optimal scores, indicating the need for deeper exploration to facilitate optimal binding affinity of newly identified key binding residues on the selected ECMs. Moreover, the structural integrity of *M. tuberculosis* adhesin interactions with host ECM components and cellular receptors has not yet been fully characterized within a simulated host environment.

In addition, experimental techniques such as isothermal titration calorimetry could be employed to quantify the thermodynamic parameters of these interactions and validate computational predictions. Finally, the identified adhesin proteins and their key binding residues should be further evaluated as potential epitope regions capable of eliciting an immune response, thereby supporting their development as serological biomarkers or therapeutic targets. [34] In this study, poor reacting *M. tuberculosis* conserved hypothetical proteins may still need further investigation to elucidate their biological significance and potential involvement in host-pathogen interactions. Studies have shown that in addition to the currently explored ECMs, there is an array of host substrates, such as thrombospondin, elastin, vitronectin, and elastin, which could act as adhesin-related substrates [15,54]. Subsequently, these predicted adhesins may also bind to host cellular receptor components [8,55].

5.6. Conclusion

In summary, *M. tuberculosis* recombinant CH proteins Rv3811, Rv3491, and Rv2730, have been experimentally validated to have definitive fibronectin-binding potential. In addition, Rv3491 and Rv2730 demonstrated versatile adhesin characteristics as evidenced by their strong binding affinities to collagen and laminin. This was the first extracellular matrix adhesin study that simulated the adhesin-ECM interaction by employing a molecular docking approach. Subsequently, the docking simulations provided insight into the key amino acid binding residue interaction between ECM and *M. tuberculosis* conserved hypothetical proteins to support experimental outcomes. This study supports the use of consolidated bioinformatic tools for the

experimental validation of novel adhesin and adhesin- related proteins using the SPAAN tool as a foundation.

Funding

The project running costs and partial doctoral scholarship were supported by the University of KwaZulu-Natal (UKZN) APACHE Flagship funds and publication funds (Prof. M. Pillay). Mr. R. Maharajh's doctoral scholarship was further supported by the College of Health Sciences scholarship and the SAMRC Division of Research Capacity Development under the Bongani Mayosi Health Scholars Programme. The content hereof is the sole responsibility of the authors and does not necessarily represent the official views of the funding bodies.

Notes

The authors declare no competing interests.

This study was in part presented at: the SAMRC 16th early scientists conference (5 November 2024), The UKZN College of Health Sciences symposium (26 August 2025).

Supplementary data

Available in Appendix B3: Chapter 5 Supplementary data of this thesis

References

1. World Health Organization. Global tuberculosis report 2025. Geneva: World Health Organization; 2025.

2. Evans TG, Schrager L, Thole JJV. Status of vaccine research and development of vaccines for tuberculosis. Vaccine. 2016;34(26):2911-2914. https://doi.org/10.1016/j.vaccine.2016.02.079

3. Kaufmann SH, Lange C, Rao M, *et al.* Progress in tuberculosis vaccine development and host-directed therapies—a state of the art review. The Lancet Respiratory Medicine. 2014;2(4):301-320. https://doi.org/10.1016/s2213-2600(14)70033-5

4. Harris RC, Dodd PJ, White RG. The potential impact of BCG vaccine supply shortages on global paediatric tuberculosis mortality. BMC Medicine. 2016;14(1):138. https://doi.org/10.1186/s12916-016-0685-4

5. Schrager LK, Vekemens J, Drager N, Lewinsohn DM, Olesen OF. The status of tuberculosis vaccine development. The Lancet Infectious Diseases. 2020;20(3):e28-e37. https://doi.org/10.1016/S1473-3099(19)30625-5

6. Dheda K, Ruhwald M, Theron G, Peter J, Yam WC. Point-of-care diagnosis of tuberculosis: Past, present and future. Respirology. 2013;18(2):217-232. https://doi.org/10.1111/resp.12022

7. Reedoy K, Loots D, Beukes D, Reenen Mv, Pillay B, Pillay M. *Mycobacterium tuberculosis* curli pili (MTP) is associated with significant host metabolic pathways in an A549 epithelial cell infection model and contributes to the pathogenicity of *Mycobacterium tuberculosis*. Metabolomics. 2020;16:1-14. https://doi.org/10.1007/s11306-020-01736-5

8. Vinod V, Vijayrajratnam S, Vasudevan AK, Biswas R. The cell surface adhesins of *Mycobacterium tuberculosis*. Microbiological Research. 2020;232:126392. https://doi.org/10.1016/j.micres.2019.126392

9. Viljoen A, Dufrêne YF, Nigou J. Mycobacterial adhesion: from hydrophobic to receptor-ligand interactions. Microorganisms. 2022;10(2):454. https://doi.org/10.3390/microorganisms10020454

10. Silmon de Monerri NC, Che Y, Lees JA, *et al.* Structure-based design of an immunogenic, conformationally stabilized FimH antigen for a urinary tract infection vaccine. PLOS Pathogens. 2025;21(2):e1012325. https://10.1371/journal.ppat.1012325

11. Martin C, Aguilo N, Marinova D, Gonzalo-Asensio J. Update on TB vaccine pipeline. Applied Sciences. 2020;10(7):2632. https://doi.org/10.3390/app10072632

12. Sharma B, Dua B, Joshi B. Investigating role of *Mycobacterium tuberculosis* secretory antigens in altering activation of t cell signaling events in Jurkat T cells. The International Journal of Mycobacteriology. 2020;9(4):405-410. https://doi.org/10.4103/ijmy.ijmy_172_20

13. O'Rourke KF, Gorman SD, Boehr DD. Biophysical and computational methods to analyze amino acid interaction networks in proteins. Computational and Structural Biotechnology Journal. 2016;14:245-251. https://doi.org/10.1016/j.csbj.2016.06.002

14. Sachdeva G, Kumar K, Jain P, Ramachandran S. SPAAN: a software program for prediction of adhesins and adhesin-like proteins using neural networks. Bioinformatics. 2005;21(4):483-491. https://doi.org/10.1093/bioinformatics/bti028

15. Maharajh R, Pillay M, Senzani S. A computational method for the prediction and functional analysis of potential *Mycobacterium tuberculosis* adhesin-related proteins. Expert Review of Proteomics. 2023;20(12):483-493. https://doi.org/10.1080/14789450.2023.2275678

16. Madani A, Garakani K, Mofrad MR. Molecular mechanics of *Staphylococcus aureus* adhesin, CNA, and the inhibition of bacterial adhesion by stretching collagen. PloS one. 2017;12(6):e0179601. https://doi.org/10.1371/journal.pone.0179601

17. Chiliza TE, Pillay M, Pillay B. Identification of unique essential proteins from a *Mycobacterium tuberculosis* F15/LAM4/KZN phage secretome library. Pathogens and Disease. 2017;75(1):ftx001. https://doi.org/10.1093/femspd/ftx001

18. Kumar S, Puniya BL, Parween S, Nahar P, Ramachandran S. Identification of novel adhesins of *m. tuberculosis* h37rv using integrated approach of multiple computational algorithms and experimental analysis. PLOS ONE. 2013;8(7):e69790. https://doi.org/10.1371/journal.pone.0069790

19. Thermo Fisher Scientific. Peptide Synthesis and Proteotypic Peptide Analyzing Tool 2020 [cited 2024 18 September]. Available from: https://www.thermofisher.com/za/en/home/life-science/protein-biology/peptides-proteins/custom-peptide-synthesis-services/peptide-analyzing-tool.html.

20. Yan Y, Tao H, He J, Huang S-Y. The HDOCK server for integrated protein–protein docking. Nature Protocols. 2020;15(5):1829-1852. https://doi.org/10.1038/s41596-020-0312-x

21. Gordon CH, Hendrix E, He Y, Walker MC. alphafold accurately predicts the structure of ribosomally synthesized and post-translationally modified peptide biosynthetic enzymes. Biomolecules. 2023;13(8):1243. https://doi.org/10.3390/biom13081243

22. Meng EC, Goddard TD, Pettersen EF, *et al.* UCSF ChimeraX: Tools for structure building and analysis. Protein Science. 2023;32(11):e4792. https://doi.org/10.1002/pro.4792

23. Larsen MH, Biermann K, Tandberg S, Hsu T, Jacobs Jr WR. Genetic manipulation of *Mycobacterium tuberculosis*. Current protocols in microbiology. 2007;6(1):10A. 12.11-10A. 12.21. https://doi.org/10.1002/9780471729259.mc10a02s6

24. Chang AY, Chau V, Landas JA, Pang Y. Preparation of calcium competent *Escherichia coli* and heat-shock transformation. JEMI Methods. 2017;1(22-25). https://doi.org/10.1385/0-89603-383-x:129

25. Massiah MA, Wright KM, Du H. Obtaining soluble folded proteins from inclusion bodies using sarkosyl, Triton X-100, and CHAPS: application to LB and M9 minimal media. Current Protocols in Protein Science. 2016;84(1):6.13.11-16.13.24. https://doi.org/10.1002/0471140864.ps0613s84

26. Towbin H, Staehelin T, Gordon J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proceedings of the National Academy of Sciences. 1979;76(9):4350-4354. https://doi.org/10.1073/pnas.76.9.4350

27. Islam MS, Richards JP, Ojha AK. Targeting drug tolerance in mycobacteria: a perspective from mycobacterial biofilms. Expert Review of Anti-infective Therapy. 2012;10(9):1055-1066. https://doi.org/10.1586/eri.12.88

28. Machowski EE, Senzani S, Ealand C, Kana BD. Comparative genomics for mycobacterial peptidoglycan remodelling enzymes reveals extensive genetic multiplicity. BMC Microbiology. 2014;14(1):75. https://doi.org/10.1186/1471-2180-14-75

29. Verma RN, Malik MZ, Singh GP, Subbarao N. Identification of key proteins in host–pathogen interactions between *Mycobacterium tuberculosis* and *Homo sapiens*: A systematic network theoretical approach. Healthcare Analytics. 2022;2:100052. https://doi.org/10.1016/j.health.2022.100052

30. Alonso H, Parra J, Malaga W, *et al.* Protein O-mannosylation deficiency increases LprG-associated lipoarabinomannan release by *Mycobacterium tuberculosis* and enhances the TLR2-associated inflammatory response. Scientific Reports. 2017;7(1):7913. https://doi.org/10.1038/s41598-017-08489-7

31. Fontán PA, Aris V, Alvarez ME, *et al. Mycobacterium tuberculosis* sigma factor e regulon modulates the host inflammatory response. The Journal of Infectious Diseases. 2008;198(6):877-885. https://doi.org/10.1086/591098

32. Danilchanka O, Mailaender C, Niederweis M. Identification of a Novel Multidrug Efflux Pump of *Mycobacterium tuberculosis*. Antimicrobial Agents and Chemotherapy. 2008;52(7):2503-2511. https://doi.org/10.1128/aac.00298-08

33. Lensink MF, Brysbaert G, Nadzirin N, *et al.* Blind prediction of homo- and hetero-protein complexes: The CASP13-CAPRI experiment. Proteins: Structure, Function, and Bioinformatics. 2019;87(12):1200-1221. https://doi.org/10.1002/prot.25838

34. Hossain MS, Azad AK, Chowdhury PA, Wakayama M. Computational identification and characterization of a promiscuous t-cell epitope on the extracellular protein 85b of *Mycobacterium spp*. for peptide-based subunit vaccine design. BioMed Research International. 2017;2017(1):4826030. https://doi.org/10.1155/2017/4826030

35. Chundru D, Bhattrai S, Timilsina M, *et al.* In silico design, optimization, and evaluation of a multi-epitope vaccine targeting the clostridium perfringens collagen adhesin protein. Microorganisms. 2025;13(5):1147. https://doi.org/10.3390/microorganisms13051147

36. Periz ÇD, Ulusoy S, Kaya Kinaytürk N. Molecular docking and in vitro evaluation of violacein-alginate beads for targeted inhibition of staphylococcus aureus biofilm formation. ACS Omega. 2025;10(29):31827-31839. https://doi.org/10.1021/acsomega.5c03106

37. Lata M, Sharma D, Kumar B, *et al.* Proteome Analysis of ofloxacin and moxifloxacin induced *Mycobacterium tuberculosis* isolates by proteomic approach. Protein & Peptide Letters. 2015;22(4):362-371. https://doi.org/10.2174/0929866522666150209113708

38. Yang H, Ma Y, Lei X, *et al.* Stopping tuberculosis at the gate: the role of *M. tuberculosis* adhesins in infection and intervention. Vaccines. 2025;13(7). https://doi.org/10.3390/vaccines13070676

39. Mieher JL, Schormann N, Wu R, *et al. Streptococcus mutans* collagen-binding protein Cnm is a multifunctional adhesin: a structural investigation. Molecular Oral Microbiology. 2025:e70005. https://doi.org/10.1111/omi.70005

40. Squeglia F, Ruggiero A, Berisio R. Collagen degradation in tuberculosis pathogenesis: the biochemical consequences of hosting an undesired guest. Biochemical Journal. 2018;475(19):3123-3140. https://doi.org/10.1042/bcj20180482

41. González-Avila G, Sandoval C, Herrera MT, *et al. Mycobacterium tuberculosis* effects on fibroblast collagen metabolism. Respiration. 2008;77(2):195-202. https://doi.org/10.1159/000163064

42. Al Shammari B, Shiomi T, Tezera L, *et al.* The extracellular matrix regulates granuloma necrosis in tuberculosis. The Journal of Infectious Diseases. 2015;212(3):463-473. https://doi.org/10.1093/infdis/jiv076

43. Kinhikar AG, Verma I, Chandra D, *et al.* Potential role for ESAT6 in dissemination of *M. tuberculosis* via human lung epithelial cells. Molecular Microbiology. 2010;75(1):92-106. https://doi.org/10.1111/j.1365-2958.2009.06959.x

44. Ramsugit S, Pillay B, Pillay M. Evaluation of the role of *Mycobacterium tuberculosis* pili (MTP) as an adhesin, invasin, and cytokine inducer of epithelial cells. Brazilian Journal of Infectious Diseases. 2016;20(2):160-165. https://doi.org/10.1016/j.bjid.2015.11.002

45. Kinhikar AG, Vargas D, Li H, *et al. Mycobacterium tuberculosis* malate synthase is a laminin-binding adhesin. Molecular Microbiology. 2006;60(4):999-1013. https://doi.org/10.1111/j.1365-2958.2006.05151.x

46. Be NA, Bishai WR, Jain SK. Role of *Mycobacterium tuberculosis* pknD in the pathogenesis of central nervous system tuberculosis. BMC Microbiology. 2012;12(1):7. https://doi.org/10.1186/1471-2180-12-7

47. Ryndak MB, Chandra D, Laal S. Understanding dissemination of *Mycobacterium tuberculosis* from the lungs during primary infection. Journal of Medical Microbiology. 2016;65(5):362-369. https://doi.org/10.1099/jmm.0.000238

48. Sinha S, Kosalai K, Arora S, *et al.* Immunogenic membrane-associated proteins of *Mycobacterium tuberculosis* revealed by proteomics. Microbiology. 2005;151(7):2411-2419. https://doi.org/10.1099/mic.0.27799-0

49. Jankute M, Nataraj V, Lee OY-C, *et al.* The role of hydrophobicity in tuberculosis evolution and pathogenicity. Scientific Reports. 2017;7(1):1315. https://doi.org/10.1038/s41598-017-01501-0

50. Pettersen EF, Goddard TD, Huang CC, *et al.* UCSF chimerax: structure visualization for researchers, educators, and developers. Protein science. 2021;30(1):70-82. https://doi.org/10.1002/pro.3943

51. Palmer I, Wingfield PT. Preparation and extraction of insoluble (inclusion-body) proteins from *Escherichia coli*. Current Protocols in Protein Science. 2012;70(1):6.3.1-6.3.20. https://doi.org/10.1002/0471140864.ps0603s70

52. Schlager B, Straessle A, Hafen E. Use of anionic denaturing detergents to purify insoluble proteins after overexpression. BMC Biotechnology. 2012;12(1):95. https://doi.org/10.1186/1472-6750-12-95

53. Zhang Y, Xu J-c, Hu Z-d, Fan X-y. Advances in protein subunit vaccines against tuberculosis. Frontiers in immunology. 2023;14:1238586. https://doi.org/10.3389/fimmu.2023.1238586

54. Halper J. Basic components of connective tissues and extracellular matrix: fibronectin, fibrinogen, laminin, elastin, fibrillins, fibulins, matrilins, tenascins and thrombospondins. Progress in Heritable Soft Connective Tissue Diseases. 2021:105-126. https://doi.org/10.1007/978-3-030-80614-9_4

55. Bisht D, Meena LS. Adhesion molecules facilitate host-pathogen interaction & mediate *Mycobacterium tuberculosis* pathogenesis. Indian Journal of Medical Research. 2019;150(1):23-32. https://doi.org/10.4103/ijmr.IJMR_2055_16

Chapter 6: Immunological Evaluation of Conserved Hypothetical Proteins of *Mycobacterium tuberculosis*

This Chapter examines the immunological properties of the *Mycobacterium tuberculosis* conserved hypothetical proteins. The biological activity of Histidine affinity-tagged proteins synthesized in Chapter 4 was confirmed, as shown by the fibronectin-binding potential in Chapter 5. Immunoinformatic analysis revealed the predicted epitopes for antigenic stimulus in the humoral and cellular response in the host. Immunoinformatics analysis yielded putative epitopes that can induce an immune response in the host. In addition, serological analysis of these conserved hypothetical proteins was conducted to identify the presence of Immunoglobulin G-specific antibodies produced in response to the conserved hypotheticals in a South African tuberculosis cohort. This manuscript was formatted for the *Tuberculosis and Respiratory Disease Journal.*

Title: Evaluation of novel adhesin-related *Mycobacterium tuberculosis* conserved hypothetical protein biomarkers: an immunoinformatic and serological approach

*Rivesh Maharajh[1], Sibusiso Senzani[1], Veron Ramsuran[1], Balakrishna Pillay[2] and Manormoney Pillay[1]**

[1]Discipline of Medical Microbiology, School of Medicine, College of Health Sciences, University of KwaZulu-Natal, Durban 4013, South Africa

[2] Discipline of Microbiology, School of Life Sciences, College of Sciences, University of KwaZulu-Natal, Durban 4013, South Africa

***Corresponding Author: Prof. Manormoney Pillay (PhD), Discipline of Medical Microbiology,**

University of KwaZulu-Natal, Durban, 4000, South Africa. Email address: Pillayc@ukzn.ac.za

6.1. Abstract

The burden of tuberculosis (TB) remains a global health challenge. Despite current efforts, reliable serological biomarkers and effective vaccine candidates are still needed. In this study, *Mycobacterium tuberculosis* conserved hypothetical (CH) proteins were systematically evaluated as serological diagnostic and immunomodulating vaccine therapeutic targets. Immuno-informatic analysis was performed on five *M. tuberculosis* proteins (Rv3811, Rv3717, Rv3491, Rv3209, Rv2598) to determine antigenicity and simulate host immune response. 6xHistidine affinity tag recombinant proteins were expressed and purified. Fibronectin binding activity was assessed using an enzyme-linked immunosorbent assay (ELISA). A total of 100 plasma samples from a South African cohort (GeneXpert-confirmed: 50 TB-positive and 50 TB-negative) were used to identify infection by measuring *M. tuberculosis* antigen-specific immunoglobulin G antibody levels in an indirect ELISA. The impact of risk factors and antigen combinations on the sensitivity and specificity of detection was statistically analyzed. Epitope prediction identified surface-exposed discontinuous regions with the highest predicted antigen score for Rv2598 (0.743). Strong fibronectin binding was observed for Rv3811, Rv3717, and Rv3491. Immune simulation predicted strong interferon-γ stimulation across all antigens. The immune simulation revealed that most tested *M. tuberculosis* antigens, particularly Rv3209, elicited robust IgM and IgG1 antibody responses and a strong IFN-γ-driven cellular response with active IL-2 signaling, whereas PstS1 showed minimal humoral stimulation and regulatory cytokines (TGF-β, IL-10, IL-12) remained low. The ELISA analysis revealed significantly higher IgG immunoreactivity by Rv3811 (p=0.036) and Rv2598 (p=0.012) in TB-positive compared with TB-negative groups. In contrast, Rv3491 exhibited an inverse association (p<0.001), with higher immunoreactivity in TB-negative samples. The receiver operating characteristic analysis demonstrated modest diagnostic accuracy for Rv2598 (AUC=0.646, p=0.008) and Rv3811 (AUC=0.622, p=0.030). Human immunodeficiency virus co-infection impaired Rv2598 detection, and no antigen combination significantly improved diagnostic performance. The *M. tuberculosis* CH proteins showed limited serological diagnostic potential in the cohort evaluated. However, the ability of these proteins to elicit a detectable IgG response and their predicted antigenic profiles warrant further investigation as candidates for immunomodulatory therapeutic or vaccine interventions.

KEYWORDS: tuberculosis, fibronectin-binding proteins, serological biomarkers, antigenic determinants

6.2. Introduction

The global burden of tuberculosis (TB) remains the leading cause of mortality among communicable diseases, with an estimated 1.23 million deaths in 2024 [1]. Approximately 8.2 million individuals were newly diagnosed with TB, indicating an increase in case notifications, and highlighting the ongoing challenges for TB control and prevention [1,2]. Therefore, rapid and accurate diagnostic tools to enable early detection and effective monitoring are needed to reduce the transmission of TB and to mitigate the emergence of drug resistance [3]. In addition, immunomodulating biomarkers may have therapeutic vaccine potential, as these antigens may induce the host immune system to augment protective responses that suppress or eliminate the pathogen [4].

The World Health Organization (WHO) emphasizes the importance of rapid detection methods for individuals at high risk of developing active TB and drug resistance [2]. Currently, approved immunological detection tests for *M. tuberculosis* include antigen-based skin tests, interferon-gamma (IFN-γ) release assays, and tuberculin skin tests [2]. However, these have limitations, including long incubation time and false-positive results in Bacille Calmette-Guérin (BCG) vaccinated populations [5,6]. Moreover, newer technological advances required skilled laboratory personnel and expensive, complex equipment [6,7]. Significant advances have been achieved beyond conventional diagnostic approaches for TB, which historically relied on labor-intensive culture methods and microscopic examination that were both time-consuming and limited in sensitivity [5,7,8].

Currently, nucleic acid amplification tests (NAATs) such as the Xpert MTB/XDR (Cepheid) are strongly recommended as the frontline diagnostic approach for initial detection of active TB and drug-resistant tuberculosis across diverse clinical contexts [6,7]. In addition, NAATs have been promoted as the preferred diagnostic tools, enabling earlier detection, faster resistance profiling, and improved patient outcomes compared to conventional methods [7]. Other methods include line probe assays Genotype MTBDR*plus/sl* and Genoscholar NTM+MDRTBII/PZA-TBII) and targeted next-generation sequencing (Deeplex Myc-TB, Ampore-TB and Tbseq) for drug resistance testing [2,5,7]. Antigenic proteins have been extensively investigated for the development of immunoassay-based tests for the detection of *M. tuberculosis* [9-12]. These assays are designed to improve the sensitivity and specificity of TB

diagnostics by leveraging host immune responses to pathogen-specific antigens, including adhesin-derived antigens that are prominently displayed on the mycobacterial cell surface [13,14].

Mycobacterial adhesins, particularly those associated with fibronectin binding potential [15-17], have previously been explored for their serological and therapeutic importance as biomarkers for detecting TB, with varying sensitivity and specificity exhibited in diverse populations [4,18-21]. During the onset of TB infection, B-cell proliferation results in the production of antibodies that play a central role in host defense by restricting microbial replication [22]. Immunoglobulin (Ig) isotypes such as IgM enhance granuloma formation [22], and later shifts towards IgG production during the intermediate and chronic phase of infection [23]. Elevated levels of IgG have been reported to occur within latent or long-standing active infections [24].

Serological diagnostic approaches largely focus on detecting IgM and IgG isotypes directed against *M. tuberculosis* antigenic epitopes [25,26]. Furthermore, detection of pathogen-specific antibodies may provide correlates of protection for vaccine development [27]. Antibody profiling not only provides insight into the magnitude and quality of the humoral immune response but also enables the identification of immune signatures associated with protective efficacy [28]. Such correlates can be instrumental in bridging preclinical findings with clinical outcomes, guiding the rational design of vaccine candidates, and informing the selection of immunological endpoints in clinical trials [29]. Despite the progress thus far, there remains a pressing need for novel and reliable serological biomarkers of *M. tuberculosis.*

In the current study, the immunogenic characteristics of the proteins were explored using an immunoinformatic pipeline. Furthermore, an enzyme-linked immunosorbent assay (ELISA) was used to assess the serological diagnostic and immunomodulatory vaccine potential of five conserved hypothetical recombinant (rec-) proteins with predicted adhesin potential, as described by Maharajh *et al.* [30]. In addition, the fibronectin-binding potential of these rec-proteins, namely Rv3811, Rv3717, Rv3491, Rv3209, and Rv2598, was investigated to determine the biological functionality of the synthesized proteins.

6.3. Materials and Methods

6.3.1 Ethical declaration

The study was approved by the Biomedical Research Ethics Committee (BREC), University of KwaZulu-Natal (BREC approval reference: BREC/00001970/2020). The TB-positive and TB-negative plasma samples had been collected with informed consent in a previous study (BREC approval reference: (BREC/00005530/2023) by Prof V. Ramsuran, and was retrieved from his repository at the Discipline of Medical Microbiology, Howard College, UKZN.

6.3.2 Immunoinformatic characterization of *Mycobacterium tuberculosis* conserved hypothetical proteins.

The SPAAN tool (https://sourceforge.net/projects/adhesin/files/SPAAN/) was used to determine the predicted adhesin potential of Rv3811, Rv3717, Rv3491, Rv3209, and Rv2598 *Mycobacterium tuberculosis* (H37Rv) conserved hypothetical proteins with a probability of adhesin (*P*ad) threshold ≥0.51. The respective protein sequences and Uniprot accession numbers were retrieved from the Mycobrowser tool (https://mycobrowser.epfl.ch/) for immunoinformatic analysis to determine antigenic epitope regions. The Vaxijen 2.0 (https://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) was used as a baseline identification tool to determine if the predicted adhesins were probable protective antigens above a threshold of 0.4. The epitope prediction was performed on multiple epitope prediction platforms to identify immunogenic B-cell epitope peptides within the selected adhesin proteins.

Antigenic peptide linear sequences were identified and compared using the Bepipred linear epitope prediction 2.0 tool (https://tools.iedb.org/bcell/) and the ABCpred artificial neural network-based B-cell epitope prediction server (https://webs.iiitd.edu.in/raghava/abcpred/AC_submission.html). In addition, the predicted antigenic peptides tool using Kolaskar and Tongaonkar method [31] (http://imed.med.ucm.es/Tools/antigenic.html#Software) has shown the average antigenic propensity < 1 to have antigenic properties. AlphaFold2 protein structures retrieved from the UniProt database were used to predict the tertiary protein structure derived from the critical assessment of structural prediction (CASP14), and energy minimization using AMBER force to improve prediction accuracy [32]. The protein models were examined to identify exposed epitope

regions (discontinuous) using structure-based epitope prediction tools, Discotope 2.0 (https://services.healthtech.dtu.dk/services/DiscoTope-2.0/) and Ellipro (https://tools.iedb.org/ellipro/). The linear-sequence prediction tools were then compared to discontinuous prediction outcomes to identify the most active B-cell epitope regions for antibody recognition.

6.3.3 Fibronectin-binding potential of recombinant *Mycobacterium tuberculosis* proteins

A selection of five *M. tuberculosis* CH genes, namely, Rv3811, Rv3717, Rv3491, Rv3209, and Rv2598, were cloned, expressed, and purified into 6x histidine (His)-affinity tag recombinant proteins using methods described in Chapter 4. Protein size and purity were confirmed by 12% sodium dodecyl sulfate-polyacrylamide electrophoresis and western blot (Supplementary Figure 1). The biological adhesin properties of the conserved hypothetical proteins were confirmed by the fibronectin-binding potential. An optimized ELISA (Figure 1) was performed to determine binding of the rec- His-tagged proteins to 10 µg Fibronectin from human plasma-pure (Uniprot accession no. P02751, Sigma-Aldrich, USA) in PBS (pH 7.2). A negative control containing 10 µg Bovine Serum Albumin (BSA) and a positive control containing 50 µg, His-affinity tag proteins were included in technical triplicate.

The method discussed in this chapter is the modified methodology shown in Chapter 5. Briefly, each protein component was coated onto the wells by passive adsorption using Covalink NH F8 module flat-bottom 96-well plates (Thermo Fisher Scientific, Roskilde, Denmark) and incubated overnight at 4°C. The ELISA plates were washed thrice with 300 µL phosphate buffered saline with 0.1% tween80 (PBST) and blocked with 1% skim milk- PBST for 2 hours at 37°C. After three wash steps, 50 µg purified rec- protein was added to each well containing the fibronectin and BSA control, and incubated overnight at 4°C. Thereafter, the wells were washed five times with 300 µL PBST. The presence of the recombinant His-affinity tag proteins was detected by the addition of 6x-His tag monoclonal antibody, Horse-radish peroxidase (HRP) conjugated (Thermo Fisher Scientific, MA, USA) in 5% SM-PBST to each well and incubated for 45 minutes at 37°C. Lastly, the wells were washed three times with 300 µL PBS-tween (0.1%) and once with PBS (pH 7.2) only. The reaction was developed with 1-step ultra TMB-ELISA substrate (Thermo Fisher Scientific, MA, USA) and stopped with 50 µL of H_2SO_4 (2 Molar, Sigma-Aldrich, MO, USA), which also served as a blank in separate triplicate wells. The absorbance signals of

each well were read at 450 nm on the iMark microplate reader (Bio-Rad Industries, CA, USA) and analyzed on Microplate Manager version 6.3 (Bio-Rad Industries, CA, USA). The absorbance readings corresponded to the binding of the His-affinity tag rec- protein to the wells coated with the fibronectin. The reactive protein (positive control) measured the total absorbance for the detection of the recombinant His-affinity tag protein added to the plate. The BSA negative control measured the absorbance of potential non-specific binding of the His-protein.

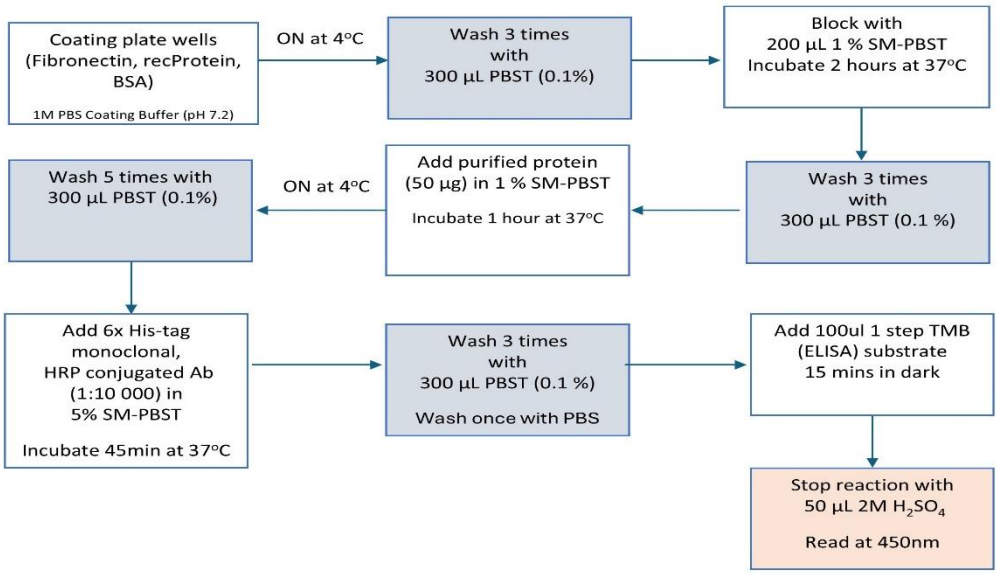


Figure 6.1. Flow diagram of the ELISA detection of the fibronectin binding potential of His-affinity tag recombinant *Mycobacterium tuberculosis* conserved hypothetical proteins. (BSA-bovine serum albumin, PBST- Phosphate buffer saline with Tween80, TMB-Tetramethylbenzidine)

6.3.4 Enzyme-linked immunosorbent assay (ELISA) detection of antibodies against *Mycobacterium tuberculosis* conserved hypothetical proteins in patients' plasma samples.

6.3.4.1 Study population

Plasma samples of 100 South African patients were obtained from a biorepository maintained by the collaborator. Blinded samples were derived from two groups: 50 patients with pulmonary tuberculosis confirmed by GeneXpert MTB/RIF assay, and those who were undergoing treatment at a public health TB clinic. Fifty patients with a GeneXpert MTB/RIF confirmed TB negative status were recruited from a medical outpatient department in the eThekwini municipality area,

South Africa. Clinical data were collected for all patients, including demographic data, co-morbidities, and HIV status, which was subsequently laboratory confirmed (Table 6.1).

Table 6.1. Demographic data of participants with GeneXpert confirmed TB status (N=100)

Demographic	TB positive (n=50)	TB negative groups (n=50)	*p*-value
Gender			*<0.001*
Male	37	20	
Female	11	26	
Unknown	2	4	
Age			*0.290*
18-25	3	5	
26-35	22	17	
36-45	10	12	
46-55	8	5	
≥56	3	9	
unknown	4	2	
Ethnicity			*0.500*
Black	44	47	
Indian	1	1	
Mixed-race	5	2	
Smoking status			*0.154*
Yes	21	14	
No	25	31	
Unknown	4	5	
Co-Morbidity			*0.233*
HIV	20	14	
Hypertension	1	4	
Diabetes	2	4	
None	29	31	

6.3.4.2 Antigen-specific plasma immunoglobulin G antibody detection by enzyme-linked immunosorbent assay.

Two sets of unblinded Confirmed TB-positive and TB-negative samples were used to optimize various parameters by determining the optimal positive-to-negative TB ratio (Appendix B4.2 serological optimization supplementary data). The Nunc Maxisorp plate demonstrated more effective binding of the *M. tuberculosis* antigen to the bottom of wells compared to the Nunc

Covalink plate. Phosphate-buffered saline elicited a lower background signal compared to carbonate-bicarbonate (CB) buffer when coated on Nunc Maxisorp plate wells. Subsequently, the PBS antigen coating buffer showed a higher and more consistent positive-to-negative TB ratio compared to the CB buffer. Various antigen coating concentrations (1.562 µg/mL, 3.125 µg/mL, 6.25 µg/mL, 12.50 µg/mL, 25 µg/mL) were explored using two antigen coating buffers to determine optimal reactivity. Larger proteins, namely Rv3811 and PstS1, showed improved immunoreactivity at higher concentrations (6.25 µg/mL); however, smaller proteins such as Rv3717, Rv3491, Rv3209, and Rv2598 were optimal at 3.125 µg/mL. Plasma samples were investigated to determine the optimal dilution for *M. tuberculosis* conserved hypothetical protein-specific IgG antibody binding to the antigen. This was determined to be a dilution of 1:300. Two secondary antibodies for detecting bound *M. tuberculosis* conserved hypothetical antigen-specific IgG antibody in blood plasma were examined for optimal detection of IgG (Table 2). However, mouse anti-human IgG Fc secondary, HRP conjugated (Thermo Fisher Scientific, MA, USA) showed improved positive-to-negative TB ratio at 1:1000 compared to monoclonal mouse anti-human IgG-Fc specific peroxidase (Merck, Darmstadt, Germany) at 1:10 000.

An indirect ELISA was performed to measure the immunoreactivity of *M. tuberculosis* CH proteins by detecting antigen-specific IgG in plasma samples from TB-positive and TB-negative patients. Briefly, 100 µL of purified proteins (3.125-6.25 µg/mL) in phosphate-buffered saline (pH 7.4) were coated onto the bottom of Nunc Maxisorp 96-well plates (Merck, Darmstadt, Germany) and incubated overnight at 4°C. Plates were allowed to acclimatize at 37°C for 1 hour.

Following three washes with PBST (pH 7.2 with 0.05% Tween 80), wells were blocked with 1% skim milk-PBST and incubated at 37°C for 1 hour. Washing was repeated three times with 0.05% PBST, and the patient plasma sample (1:300 dilution) was added and incubated at 37°C for 2 hours. A wash step was repeated six times with 0.05% PBST, and the conjugated secondary antibody, mouse anti-human IgG Fc secondary, HRP conjugated (Thermo Fisher Scientific, MA, USA) was added and incubated at 37°C for 1 hour. The final wash was performed nine times, and 100 µL of 1-step Ultra-TMB substrate (Thermo Fisher Scientific, MA, USA) was added and incubated for 20 minutes at 25°C in the dark. The reaction was stopped with 50 µL of H_2SO_4 (2 M), read at 450 nm on the iMark microplate reader (Bio-Rad Industries, CA, USA) and analyzed using Microplate Manager version 6.3 (Bio-Rad Industries, CA, USA).

Table 6.2. Optimized parameters for improved ELISA performance and detection of TB-positive and TB-negative blood plasma samples.

Parameters	Comparisons	Optimal conditions
96-well Immunoplate	Nunc Maxisorp, Nunc Covalink	Nunc Maxisorp
Antigen coating buffer	Phosphate Buffer Saline (PBS, pH 7.2) Carbonate-bicarbonate (pH 9.6)	PBS (pH 7.2)
Antigen concentration	1.6, 3.125, 6.25, 12.5, 25 µg/mL,	6.25 µg/mL (Rv3811, PstS1); 3.125 µg/mL (Rv3717, Rv3491, Rv3209, Rv2598)
Plasma sample dilution	1:50, 1:100, 1:300, 1:900, 1:1800	1: 300
Selection of secondary antibody and dilution	Monoclonal mouse anti-human IgG-Fc specific peroxide (Merck): 1:5,000, 1:10,000, 1:15,000, 1:20,000 mouse Anti-human IgG Fc secondary-HRP (Thermo Fisher Scientific): 1:500, 1:1,000	Thermo Fisher Scientific: 1:1,000 dilution.

6.3.5. Statistical analysis

Data was captured in Microsoft Excel and exported to SPSS statistical package version 30.0 (IBM Corp, Armonk, NY, USA). GraphPad Prism version 8.0 (GraphPad Software, CA, USA) was used to illustrate statistical outputs. The means, medians, interquartile ranges, standard deviation (SD), and coefficient of variation (COV) were analyzed using technical replicates of the samples to assess the reproducibility of the ELISA assay. Descriptive statistics were performed for univariate analysis to describe the epidemiological data presented. Bivariate analysis was conducted to determine significant associations among categorical groups using Pearson's chi-squared test. Demographic regression analysis was performed with missing variable data therefore samples were reconciled and excluded during the analysis (N=86). The analysis of the ELISA outcomes showed that the data followed a non-parametric distribution according to the one-sample Kolmogorov-Smirnov test ($p<0.001$). The Mann-Whitney U test was used to determine differences between two independent variable groups. Multivariable logistic regression analysis was employed to identify co-morbidity risk factors (HIV) associated with TB status by pairwise group comparison matrices. Receiver operating characteristic (ROC) curve analysis was applied to assess diagnostic accuracy. The area under curve (AUC) further evaluated the sensitivity and specificity for TB detection. Classifier evaluation metrics determined the optimal cutoff value, and Youden's index evaluated the maximum potential effectiveness of the serological biomarker. In addition, the positive likelihood ratios and positive predictive values were determined to provide insight as a

serodiagnostic marker. A binary logistic regression model was used to assess the association between the immunoreactivity (absorbance) of each *M. tuberculosis* protein detected by ELISA and tuberculosis (TB) disease status. Binary logistic regression with interaction terms was used to assess whether antigen combinations (two, three, four, and five) improved TB outcome detection, using antigen combinations (ELISA-based correct/incorrect classification indicators via optimized cutoffs) in interaction with confirmed clinical TB status (positive/negative). Statistical significance was set at $p < 0.05$ with 95% confidence intervals.

6.4. Results

6.4.1 Immunoinformatic profiling of the *Mycobacterium tuberculosis* conserved hypothetical proteins.

6.4.1.1. Characteristics of conserved hypothetical proteins

Among the five *M. tuberculosis* conserved hypothetical proteins investigated, Rv3717, previously identified by Kumar and colleagues [33], was predicted to have the highest probability of adhesin (P_{ad}, score = 0.838). Subsequently, the predicted probable protective antigen level of Rv3717 amidase protein (PPA = 0.633) was determined by whole protein antigenicity prediction (Table 6.3). However, antigen Rv2598 with the lowest predicted adhesin (P_{ad} score: 0.552 was determined to have the strongest antigenic prediction of 0.743.

When comparing the probable protective antigen level among the amidase proteins, Rv3811 showed low antigenic probability (PPA = 0.399). The second-highest probability of adhesin (P_{ad} = 0.795) was observed for Rv3209, despite a low protein antigenicity prediction (0.570), which was within the probable protective antigen threshold. Both Rv3717 and Rv3491 were predicted to contain signal peptide sequences that allow translocation of the protein to the transmembrane region of the mycobacterial cell [30]. This supports the association of high P_{ad} prediction and moderate antigenicity (PPA = 0.675) for Rv3491. *M. tuberculosis* protein Rv3811 was observed to possess the largest number (n=25) antigenic determinants, compared with six to 10 antigenic determinants for the remaining *M. tuberculosis* antigens. This corresponds to the highest number of amino acids (n = 539) within the Rv3811 peptide sequence. While all proteins exhibited an average antigenic propensity of greater than 1, Rv2598 and Rv3209 exhibited the highest strength of antigenicity based on fewer antigenic determinants.

Table 6.3. Characteristics of *Mycobacterium tuberculosis* conserved hypothetical proteins.

Characteristics	*Rv3811*	*Rv3717*	*Rv3491*	*Rv3209*	*Rv2598*
Probability of adhesin (P_{ad})	0.684	0.838	0.777	0.795	0.552
Molecular weight of protein (kDa)	57.22	24.84	20.33	19.12	17.58
Number of amino acids	539	241	192	186	164
Uniprot function	N-acetylmuramoyl-L-alanine amidase	(Zinc-dependent peptidoglycan amidase	Secreted protein	Conserved hypothetical threonine and proline rich protein	Uncharacterized protein
Signal peptide sequence	No	Yes	Yes	No	No
Probable protective antigen	0.399	0.633	0.675	0.570	0.743
Number of antigenic determinants (Average antigenic propensity)	25 (1.017)	8 (1.018)	6 (1.042)	10 (1.051)	10 (1.042)

6.4.1.2. Epitope prediction analysis

The Immune Epitope Database (IEDB) analysis showed the absence of any recorded epitope datasets for the five novel, previously uncharacterized proteins. Linear B-cell epitope prediction using Bepipred 2.0, Predicted Antigen Peptides, and ABCpred identified continuous epitope regions spanning the full-length sequences of all five *M. tuberculosis* CH proteins and the known immunogenic protein 38-kDa phosphate-binding periplasmic lipoprotein, PstS1 (B4: Supplementary Figure 2). ABCpred was the most sensitive detection of prediction compared to the others as it predicted the highest number of epitope regions across all proteins. However, Bepipred 2.0 identified longer continuous epitope regions, while the predicted antigenic peptide tool identified short antigenic peptide sequences (B4: Supplementary Table 1).

In contrast, structure-based prediction tools, Ellipro and Discotope, revealed discontinuous B-cell epitope regions localized on the protein surfaces (Figures 6.2). The Discotope tool identified solvent-accessible surface areas based on amino acid composition, spatial, and surface exposure. However, Ellipro identified flexible protruding regions that may reduce the specificity of identifying epitope regions, as observed among *M. tuberculosis* CH proteins (Figure 6.2). Notably, for the larger and smaller proteins Rv3811 (57 kDa) and Rv3209 (19.12 kDa) respectively, both structure-based approaches were comparable and consistently identified a higher number of epitope regions among larger proteins compared to smaller protein sequences (Figure 6.3. 1A: 82 residues, 1D:78 residues, and 2A: 300 residues, 2D: 90 residues). Analysis of the amino acid residues within Rv3811 revealed a higher number of key epitope regions with the Ellipro analysis (300 residues) compared to the Discotope output (82 residues). This finding was similar for all proteins, suggesting that Ellipro is more sensitive in capturing flexible surface regions that do not meet the criteria of Discotope recognition (B4: Supplementary Table 2).

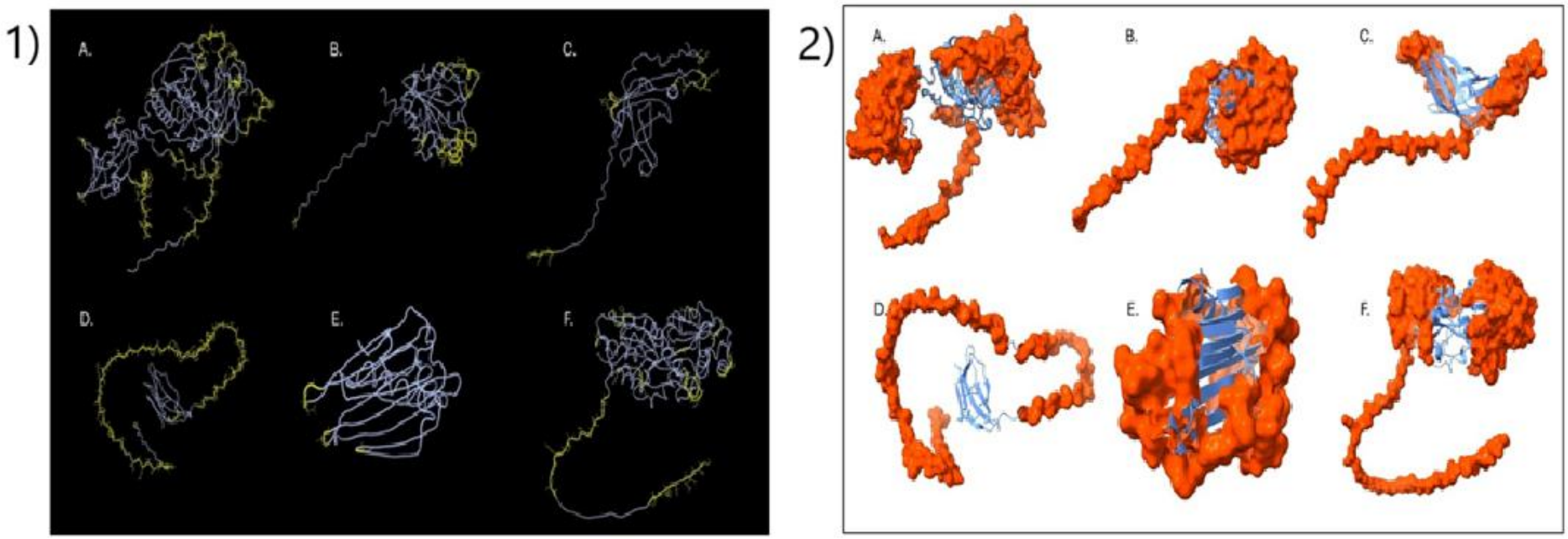


Figure 6.2. Exposed epitope regions of *Mycobacterium tuberculosis* conserved hypothetical proteins are shown as discontinuous regions determined by 1) Discotope 2.0 (visualized using Jmol) and 2) Ellipro (visualized using ChimeraX). The exposed epitope regions are shown in yellow for the Discotope prediction and orange in the Ellipro prediction. A. Rv3811, B. Rv3717, C. Rv3491, D. Rv3209, E. Rv2598, F. Rv0934 (PstS1)

6.4.1.3 Simulation of the host immune system

The immune system simulation evaluated whether the newly identified B-cell epitope regions could activate IgM and IgG antibody-producing cells (Figure 3). The IgM isotypes for Rv3811, Rv3717, Rv3491, and Rv2598 *M. tuberculosis* antigens peaked after 13 days post-immunization (≥ 6000 antibody titers) and declined thereafter. In contrast, PstS1(Rv0934) exhibited very low stimulation of IgM antibody production (approximately 2000 antibody titers). However, exceptionally high IgM and IgG1 antibody production was simulated for Rv3209, >15,000 and >10,000 Ab titers, respectively. Both IgG (1+2) were shown to induce significantly less (1000-4000 Ab titer, with the exception of Rv3209) than the IgM counterpart, while Rv3717 and Rv3209 showed no induction of the IgG2 only. Subsequently, IgG1 was more stimulated in the simulation compared to IgG2, notably due to IgG1 primarily targeting such proteins and polypeptide antigens, and latter polysaccharide antigens [34]. No stimulation of the control antigen PstS1 was observed towards IgG1 and IgG2 production, and the humoral response remained low among IgM only (Figure 3F).

The simulated cellular immune response of each antigen showed a significantly high upregulation (≥400,000 ng/mL) of Interferon-gamma (IFN- γ). Three low-level regulated cytokines, namely transforming growth factor- beta (TGF-β) (50,000 – 175,000 ng/mL), followed by interleukin-10 (≤ 50,000 ng/mL) and interleukin-12 (≤ 25,000 ng/mL), which are critical in inducing granuloma and lung fibrosis upon TB infection, were observed (Figure 4B). An inset measuring the response to danger signals, such as pathogen/damage-associated molecular patterns, shows low exposure to cell damage but a consistent upregulation of interleukin-2, indicating an active immune response (Figure 3B).

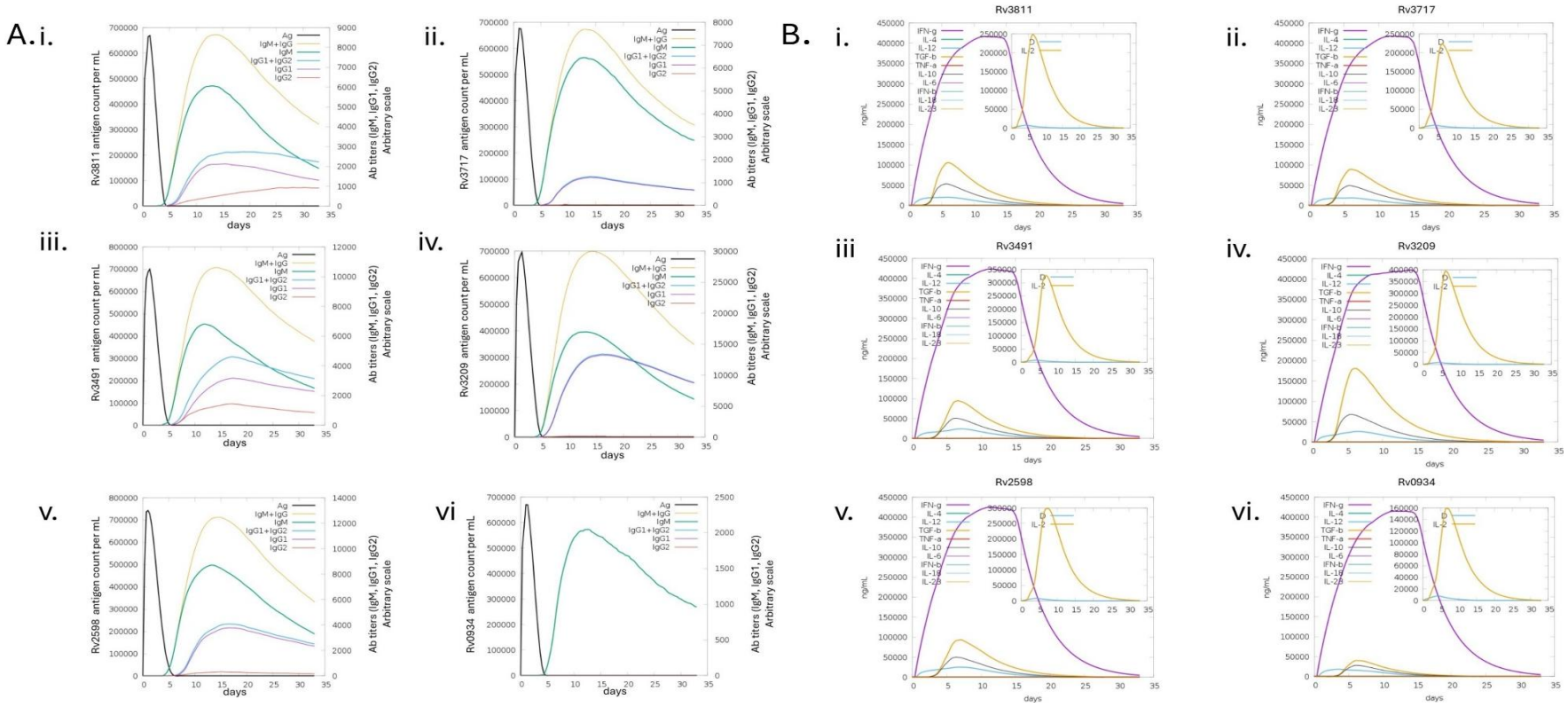


Figure 6.3. Simulation of the (A) Humoral response following simulated immunization and (B) levels of cytokine response to *Mycobacterium tuberculosis* antigens. i. Rv3811 ii. Rv3717 iii. Rv3491 iv. Rv3209 v. Rv2598 vi. Rv0934 (PstS1) was performed using the C-ImmSim server. IgM production peaked at day 13, with Rv3209 as the most immunogenic protein. IgG1 response was higher compared to IgG2 response across all antigens. Rv0934 stimulated a weak IgM response with no IgG production. All antigens have induced a strong upregulation of IFN-γ, with low upregulation of TGF-β, IL-10, and IL-12. Inset data shows danger signal and upregulation of IL2 (Figure 6.3B).

6.4.2 Fibronectin binding affinity of *Mycobacterium tuberculosis* conserved hypothetical proteins.

The recombinant His-affinity tag protein retained its functional integrity, as evidenced by its binding affinity for fibronectin (Fn), which was consistent with the results reported in Chapter 5. The positive control (reactive protein) represented the total reactivity to the detection antibody of the His-affinity tag proteins coated onto the ELISA (Nunc 96-plate) wells This exhibited a maximum absorbance (Abs_{450nm} range: 0.41-0.653) for each recombinant protein. Bovine serum albumin (BSA), the negative control, consistently yielded minimal binding signal (Abs_{450nm}: ≤0.036), thereby confirming the absence of non-specific binding of the recombinant protein to the ELISA plate wells (Table 6.4).

Table 6.4. Fibronectin (Fn) binding potential of *Mycobacterium tuberculosis* Conserved Hypothetical proteins at $Absorbance_{450nm}$.

M. tuberculosis Protein	Reactive protein (±SD)	Fn-binding (±SD)	BSA (±SD)
Rv3811	0.653 (0.055)	0.585 (0.048)	-0.00033 (0.000)
Rv3717	0.579 (0.058)	0.547 (0.077)	0.036 (0.000)
Rv3491	0.566 (0.015)	0.504 (0.066)	0.021 (0.005)
Rv3209	0.534 (0.080)	0.085 (0.030)	0.024 (0.002)
Rv2598	0.410 (0.026)	-0.026 (0.014)	0.034 (0.006)

SD: standard deviation

Three recombinant proteins, His-Rv3811(Abs_{450nm} = 0.585 ± 0.048), His-Rv3717(Abs_{450nm} = 0.547 ± 0.077), and His-Rv3491(Abs_{450nm} = 0.504 ± 0.066) showed very strong Fn-binding affinity comparable to their respective reactive protein positive control (Figure 4). In contrast, His-Rv3209 showed very poor Fn-binding potential (Abs_{450nm}=0.085 ± 0.030), and His-Rv2598 did not bind to Fn (Abs_{450nm}=-0.026 ± 0.014). When comparing Fibronectin (Fn) binding to the total reactive protein, the level of binding of the recombinant protein to the Fn could be observed.

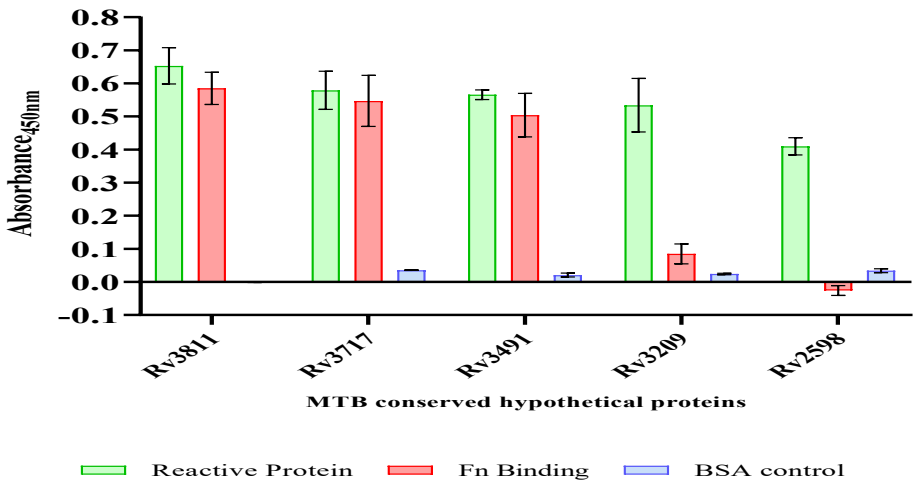


Figure 6.4. Fibronectin (Fn) binding activity at OD absorbance at 450nm of 6xHis-affinity tag *Mycobacterium tuberculosis* conserved hypothetical proteins. The positive control represents the maximum antibody-detected signal. Fn binding shows the binding potential of the recombinant protein. Bovine serum albumin, negative control, demonstrated non-specific interactions of the rec-protein. Recombinant proteins Rv3811, Rv3717, and Rv3491 exhibited Fn-binding affinity comparable to the positive control, whereas recRv3209 and recRv2598 showed minimal or no binding to Fn.

6.4.3 Serological detection of *Mycobacterium tuberculosis* conserved hypothetical Proteins.

6.4.3.1 Demographic analysis

The study group consisted of 100 plasma samples obtained from patients with GeneXpert confirmed TB-positive (n=50) and TB-negative (n=50). The majority of participants were male (n = 57) and exhibited a significantly higher TB infection rate compared to females in this cohort ($p < 0.001$). The highest proportion of patients fell within the 26–35 age group; however, the distribution across TB-positive and TB-negative groups was not significantly different (p = 0.290). Similarly, the highest proportion of ethnicity was among black participants, but not significantly different among groups. No significant differences were observed between GeneXpert TB-positive and TB-negative groups for smoking status (p = 0.154). Regarding co-morbidity prevalence, participants reported multiple illnesses, which showed a non-significant outcome across groups (p= 0.233). A multivariable logistic regression analysis (N=86) showed that the male gender (OR = 19.84, 95% CI: 3.76-104.74, p<0.001) and people living with HIV status (OR = 6.82, 95% CI: 1.57-29.63, p=0.010) were strongly associated with a GeneXpert TB-positive outcome. However,

demographic factors such as age, ethnicity, and smoking status were not significant predictors of TB outcome (Table 6.5).

Table 6.5. Multivariable logistic regression model of demographic factors associated with GeneXpert tuberculosis status (N=86).

Demographic characteristics	*Coefficient (B)*	Odds ratio	95% confidence Interval		*P*-value
			Lower	Upper	
Age					
18-25		1.00			0.336
26-35	1.216	1.435	0.461	24.653	.231
36-45	.305	.072	0.147	12.549	.788
46-55	.225	.038	0.131	11.945	.845
≥56	-.298	.060	0.068	8.123	.807
Ethnicity					
Black		1.00			0.260
Colored	2.530	12.550	0.601	262.051	0.103
Indian	-.052	.950	0.046	19.440	0.973
Gender					
Female		1.00			
Male	2.988	19.839	3.758	104.743	<0.001
Smoking status					
No		1.00			
Yes	-0.640	0.527	0.140	1.987	0.344
HIV status					
Negative		1.00			
Positive	1.920	6.682	1.571	29.631	0.010

6.4.3.2 Enzyme-linked immunosorbent assay for the detection of antibodies against the conserved hypothetical proteins among TB-positive and TB-negative

Descriptive statistics and comparative analysis using the Mann-Whitney U test showed Immunoglobulin G (IgG) immunoreactivity levels for six *M. tuberculosis* conserved hypothetical proteins (B4.1: Supplementary Table 3, Figure 6.5). Rv2598 was detected significantly higher Rv2598-IgG specific antibodies in TB-positive compared to TB-negative plasma (median: 0.354 vs. 0.227; $p = 0.012$), with minimal overlap in ranges (TB positive: 2.450; TB negative: 0.631).

Similarly, Rv3811 also demonstrated significantly elevated IgG-specific antibody levels in TB-positive samples (median: 0.504 vs. 0.375; p = 0.036). Conversely, Rv3491 showed significantly lower IgG-specific antibody levels in TB-positive patients (mean: 0.323 vs. 0.504; p < 0.001), an inverse correlation finding. Antigen Rv3717 (p = 0.873), Rv3209 (p = 0.535), and the control protein PstS1 (p = 0.847) showed no significant differences between TB positive and TB negative groups.

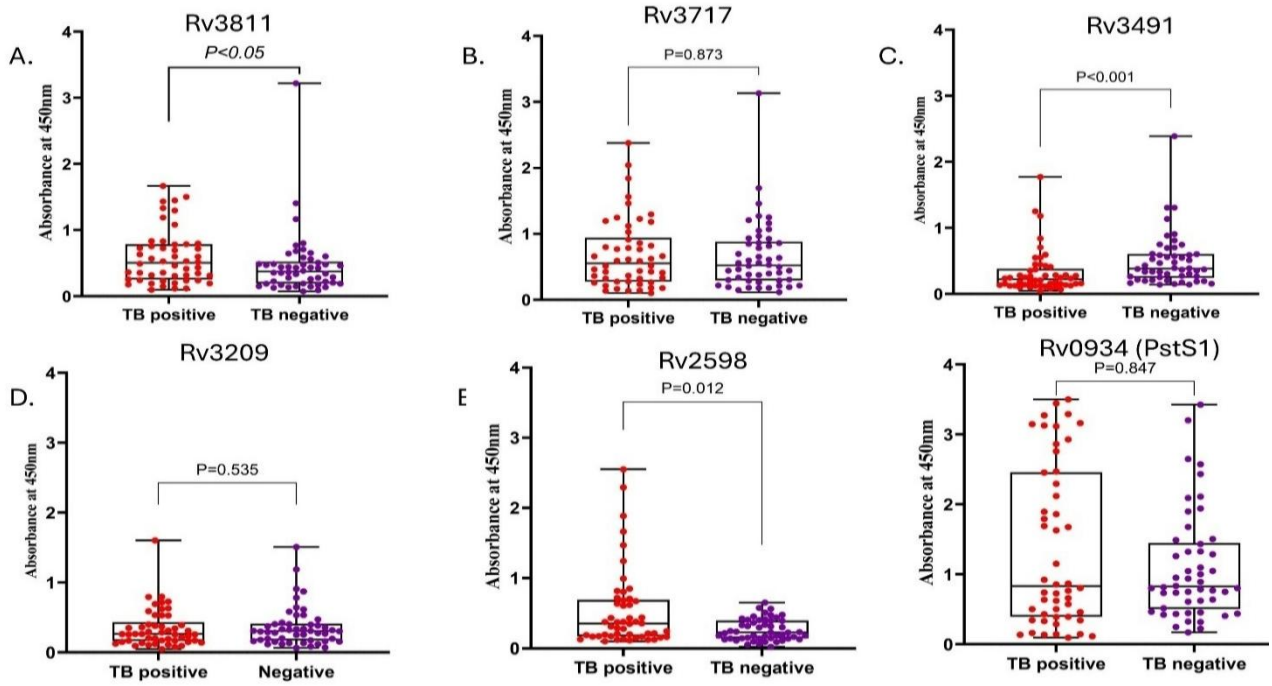


Figure 6.5. ELISA detection of IgG antibodies against *Mycobacterium tuberculosis* Conserved hypothetical proteins among TB-positive (n=50) and TB-negative (n=50) patient plasma samples.

6.4.3.3. Performance characteristics of the enzyme-linked immunosorbent assay for the detection of anti-IgG antibodies against Mycobacterium tuberculosis conserved hypothetical proteins.

The receiver operating characteristic (ROC) curve analysis (Figure 6.6) demonstrated high discrimination between TB-positive and TB-negative in the AUC for both Rv3811 (AUC = 0.622, p = 0.030) and Rv2598 (AUC =0.646, p = 0.008). Antigens Rv3717, Rv3209, and PstS1 showed poor probability of distinguishing the TB outcome (AUC = 0.466-0.511, p > 0.050). However, Rv3491 showed a significant difference between high reactivity of TB-negative compared to low reactivity to TB-positive samples (p <0.001), and the ELISA performance was below the point of reference (AUC = 0.279).

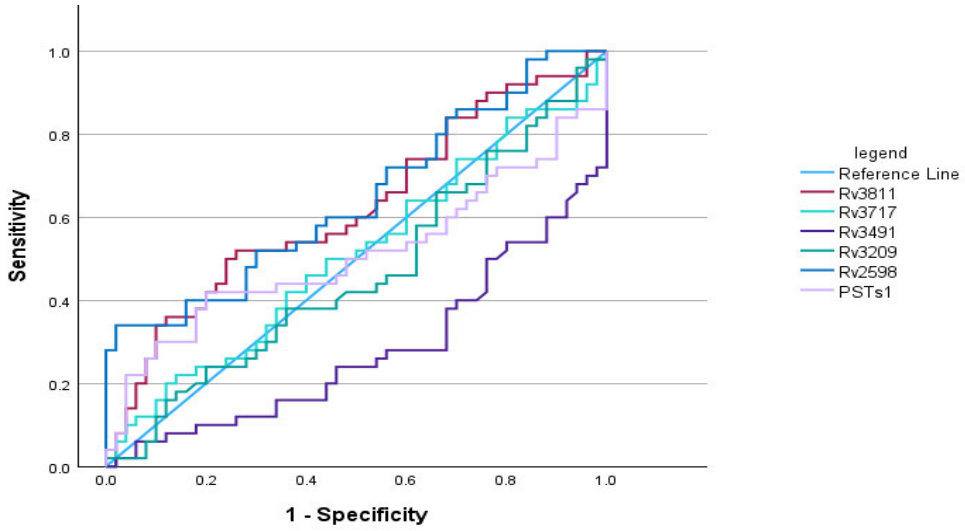


Figure 6.6. Comparison of ROC curves showing detection efficiency of the ELISA for *Mycobacterium tuberculosis* antigens among TB-positive (n=50) and TB-negative (n=50) patient plasma samples.

Using Youden's index from the classifier evaluation matrix for the ROC curve analysis, the ideal ELISA cutoff values were established ($Absorbance_{450nm}$: 0.280-1.567). The ELISA assay demonstrated low sensitivity and higher specificity for clinical use (Table 6.6). Rv2598 demonstrated high specificity (98.0%) and was more than 16 times more likely to detect the disease in a person with TB compared to a person without (positive likelihood ratio, LR+ = 16). However, sensitivity was very low for Rv2598 (32.0%). Rv3811 demonstrated modest discriminatory power between TB-positive and TB-negative detection (sensitivity 50.0%, specificity 74.0%) when compared to the remaining antigens.

Antigen Rv3491 performed significantly below chance of detection (AUC = 0.279, $p < 0.001$), as the immunoreactivity of the antigen was substantially higher in TB-negative samples compared to TB-positive samples. Antigen Rv3717, Rv3209, and PstS1 showed no statistically significant difference therefore showed poor diagnostic value ($p > 0.05$). However, PstS1 showed the highest ELISA cutoff (Classifier evaluation matrix, CEM = 1.567), relating to a high immunoreactivity of detecting known TB-positive samples in the cohort, with 2.1 odds that the test will detect TB in a positive sample compared to a sample without TB. Positive predictive values varied across all proteins (39.1-94.1%), whereby the highest probability that a person who tested positive for the disease was for Rv2598 (PPV=94.1%), although the range reflects limited sensitivity in this sample population.

Table 6.6. Performance characteristics of the ELISA to determine optimal sensitivity and specificity of *Mycobacterium tuberculosis* conserved hypothetical proteins in TB-positive and TB-negative plasma samples.

***M. tuberculosis* Protein**	**AUC**	***P*-value**	**ELISA cutoff value (CEM)**	**Youden's index**	**Sensitivity (%)**	**95% confidence Interval**	**Specificity (%)**	**95% confidence Interval**	**Positive likelihood (LR+)**	**PPV (%)**
Rv3811	0.622	0.030	0.513	0.260	50.0	35.5-64.5	74.0	59.6-85.4	1.92	65.8
Rv3717	0.509	0.872	0.613	0.060	46.0	31.8-60.7	58.0	43.2-71.8	1.10	52.3
Rv3491	0.279	<0.001	0.208	-0.260	52.0	37.4-66.3	22.0	11.5-36.0	0.67	40.0
Rv3209	0.466	0.554	0.618	0.400	16.0	7.2-29.1	88.0	75.7-95.5	1.3	57.1
Rv2598	0.646	0.008	0.593	0.320	32.0	19.5-46.7	98.0	89.4-99.9	16	94.1
PstS1 (control)	0.511	0.852	1.567	0.220	42.0	28.2-56.8	80.0	66.3-89.9	2.1	67.7

AUC=Area Under the Curve, CEM= Classifier Evaluation Matrix, PPV= Positive Predictive Value, LR+: Positive Likelihood Ratio

6.4.3.4.Immunoreactivity for TB detection among groups

The binary logistic regression model revealed that antigen Rv2598 shows a strong, statistically significant positive association with disease outcome (OR = 187.4, 95% CI: 4.5 – 7806.5, p = 0.006); however, the wide confidence interval depicts low precision in outcome due to the sample size of the cohort (Table 6.7). Antigen Rv3491 shows a strong, statistically significant inverse association (OR = 0.003, 95% CI: 0.000 – 0.087, p < 0.001), suggesting that Rv3491 is downregulated during infection or serves as a marker of protection, since the immune response clears or controls infection; therefore, low detection in TB-positive samples.

Table 6.7. The association of immunoreactivity of the *Mycobacterium tuberculosis* protein detection in ELISA in association with the TB outcome using a binary regression model.

***M. tuberculosis* Protein**	***Coefficient (B)***	**Odds ratio**	**95% confidence Interval**		***p*-value**
			Lower	**Upper**	
Rv3811	1.675	5.34	0.993	28.688	0.050
Rv3717	0.494	1.64	0.113	23.767	0.717
Rv3491	-5.689	0.003	0.000	0.087	<0.001
Rv3209	0.483	1.62	0.177	14.806	0.669
Rv2598	5.223	187.42	4.500	7806.54	0.006
PSTs1	0.0466	0.63	0.243	1.617	0.334

At the point estimate, Rv3811 suggests a positive association, although it is unstable (OR= 5.34, 95% CI: 0.993 – 28.688, p = 0.050); however, a larger sample size would be required to prove a true association. Antigens Rv3717, Rv3209, and PstS1 show no significant association, with wide CIs that can be further explored with a higher sample size.

6.4.3.5. Immunoreactivity of antigen combination analysis

A series of binary logistic regression models was used to evaluate whether two-antigen combinations improved TB detection, with each model including main effects and an interaction term for each antigen pair (Table 6.8). For the Rv3811 and Rv2598 combination, the model showed evidence of a statistical relationship (omnibus score test p < 0.001). Although the majority converged successfully, this model evaluating Rv3811 and Rv2598 exhibited complete separation,

as the combination of predictors adequately predicted the TB outcome ($p > 0.05$, OR = 0.000). Consequently, neither Rv3811 nor Rv2598 provided an improvement in TB detection (correctly predicted; Rv3811 = 63, Rv2598 = 66). Of the 15 two-antigen combinations analysed, only the interaction of Rv3811 and Rv3209 showed a significant outcome ($p = 0.019$). When both the ELISA for Rv3811 and Rv3209 correctly predicted the TB outcome, the odds of true TB positivity were reduced by 96% (OR = 0.038, CI: 0.002-0.590). This suggests a synergistic protective effect for ruling out TB rather than improved detection. No combination demonstrated improved detection (OR > 1). Given these findings, extending the analysis to three-antigen interactions would likely result in statistical overfitting and is not recommended with the current sample size (N = 100).

Table 6.8. Binary logistic regression analysis for predicting TB outcomes, showing the main and interaction effects for all combinations.

Antigen combination	*B*-coefficient	Odds ratio	95% Confidence Interval		*P*
			Lower	Upper	
Rv3811	-1.253	0.286	0.081	1.009	0.052
Rv3717	-0.049	0.952	0.149	6.099	0.959
Interaction term (Rv3811*Rv3717)	0.172	1.187	0.132	10.649	0.878
Rv3811	-2.068	0.126	0.040	0.402	<.001
Rv3491	1.453	4.278	0.462	39.604	0.201
Interaction term (Rv3811*Rv3491)	0.869	2.385	0.192	29.670	0.499
Rv3811	1.609	5.000	0.537	46.528	0.157
Rv3209	-1.995	0.136	0.028	0.653	0.013
Interaction term (Rv3811*Rv3209)	-3.274	0.038	0.002	0.590	0.019
Rv3811	18.025	67311457.082	0.000	.	0.999
Rv2598	-5.576	0.004	0.000	0.066	<0.001
Interaction term (Rv3811*Rv2598)	-16.492	0.000	0.000	.	0.999
Rv3811	1.232	3.429	0.372	31.589	0.277
PstS1	-0.560	0.571	0.106	3.092	0.516
Interaction term (Rv3811*PstS1)	-2.298	0.101	0.006	1.555	0.100
Rv3717	-1.199	0.302	0.099	0.918	0.035
Rv3491	2.251	9.500	1.091	82.725	0.041
Interaction term (Rv3717*Rv3491)	-0.363	0.696	0.058	8.326	0.775
Rv3717	1.322	3.750	0.402	34.955	0.246
Rv3209	-3.178	0.042	0.009	0.200	<0.001
Interaction term (Rv3717*Rv3209)	-1.470	0.230	0.015	3.504	0.290

Rv3717	17.907	59832404.640	0.000	.	0.999

Table 6.8. Binary logistic regression analysis for predicting TB outcomes, showing the main and interaction effects for all combinations. Continued…

Antigen combination	*B*-coefficient	Odds ratio	95% Confidence Interval	*P*	Antigen combination
Rv2598	-24.499	0.000	0.000	.	0.998
Interaction term (Rv3717*Rv2598)	2.728	15.300	0.000	.	1.000
Rv3717	20.327	673114489.047	0.000	.	0.999
PstS1	-2.079	0.125	0.028	0.553	0.006
Interaction term (Rv3717*PstS1)	-19.634	0.000	0.000	.	0.999
Rv3491	19.999	484642458.978	0.000	.	0.998
Rv3209	-3.344	0.035	0.009	0.140	<0.001
Interaction term (Rv3491*Rv3209)	-18.775	0.000	0.000	.	0.998
Rv3491	18.258	85025016.561	0.000	.	0.999
Rv2598	-4.999	0.007	0.001	0.062	<0.001
Interaction term (Rv3491*Rv2598)	-16.022	0.000	0.000	.	0.999
Rv3491	20.615	897486265.773	0.000	.	0.999
PstS1	-2.197	0.111	0.035	0.358	<0.001
Interaction term (Rv3491*PstS1)	-18.600	0.000	0.000	.	0.999
Rv3209	-19.817	0.000	0.000	.	0.998
Rv2598	-20.430	0.000	0.000	.	0.998
Interaction term (Rv3209*Rv2598)	16.669	17339626.257	0.000	.	0.998
Rv3209	-2.565	0.077	0.013	0.438	0.004
PstS1	0.208	1.231	0.202	7.506	0.822
Interaction term (Rv3209*PstS1)	-1.543	0.214	0.018	2.506	0.219
Rv2598	-4.800	0.008	0.001	0.102	<0.001
PstS1	17.907	59832394.589	0.000	.	0.999
Interaction term (Rv2598*PstS1)	-17.384	0.000	0.000	.	0.999

6.4.3.6. Co-morbidity analysis

Human Immunodeficiency Virus (HIV) status is known to influence TB outcome, therefore, a pairwise group comparison matrix was adopted to evaluate statistically significant groups (B4: Supplementary Table 4). Approximately one third of the sample cohort (n=34) were confirmed HIV-positive, of which one fifth of samples (n=20) represented TB and HIV coinfection. The Mann–Whitney U test revealed overall that HIV status did not significantly differentiate antibody

detection for all comparison groups ($p > 0.05$). However, antigen Rv2598 significantly distinguished TB-positive from TB-negative individuals among HIV-negative samples ($p = 0.038$), but not among HIV-positive samples ($p = 0.158$). For antigen Rv3491, significant differences were observed between coinfected and uninfected samples ($p < 0.001$), as well as between HIV-positive ($p = 0.036$) and HIV-negative ($p = 0.005$) groups (Table 6.9).

Table 6.9. Comparison of *Mycobacterium tuberculosis* antigen-specific antibody detection among HIV/ TB groups (p<0.05 significance)

Group 1 (n)	Group 2 (n)	Rv3811	Rv3717	Rv3491	Rv3209	Rv2598	PstS1
TB+/HIV+ (20)	**TB-/HIV- (36)**	0.128	0.758	**<0.001**	0.090	0.094	0.824
TB+/HIV+ (20)	**TB+/HIV- (30)**	0.722	0.649	0.635	0.113	0.828	0.707
TB+/HIV+ (20)	**TB- /HIV+(14)**	0.500	0.717	**0.036**	0.569	0.158	0.569
TB+HIV- (30)	**TB-/HIV- (36)**	0.058	0.913	**0.005**	0.938	**0.038**	0.471
TB+HIV- (30)	**TB-/HIV+ (14)**	0.226	0.940	0.91	0.364	0.107	0.960
TB-/HIV+ (14)	**TB-/HIV- (36)**	0.437	0.771	0.545	0.266	0.897	0.476

Consistent with earlier overall observations, detection of Rv3491-specific antibodies exhibited a negative correlation, demonstrating higher immunoreactivity in the TB-negative group. Associations with hypertension and diabetes mellitus were not explored due to non-statistically significant sample sizes.

6.5. Discussion

The *M. tuberculosis* genome consists of four thousand genes; however, 25% of these genes have not been functionally characterized or explored immunologically but have remained conserved within phylogenetic lineages [30]. Referred to as conserved hypothetical proteins, they can be further explored for their novelty as immunogenic biomarkers, particularly those with adhesin properties that enhance the pathogenicity of *M. tuberculosis* [21]. The application of current

conventional diagnostic methods requires labor-intensive molecular-based techniques, necessitating exploration for novel serological biomarkers as more practical diagnostic alternatives [2]. In addition, the vaccine pipeline includes various known *M. tuberculosis* adhesin-related constructs as viable immune-stimulating candidates [35]. In this study, the immunoinformatic characteristics, functional integrity relating to fibronectin binding potential, and detection of IgG-specific antibodies to *M. tuberculosis* conserved hypothetical proteins, namely, Rv3811, Rv3717, Rv3491, Rv3209, and Rv2598, were explored.

Adhesin properties and Antigenicity of conserved hypothetical Mycobacterium tuberculosis proteins

The SPAAN tool has been shown to predict probable adhesin potential based on the protein sequence [30,33,36]. A study conducted by Kumar *et al.* in 2013 remains the only one to explore *M. tuberculosis* proteins for their adhesin potential using the SPAAN tool [33]. The present study predicted the potential enzymatic function associated with the adhesin potential of conserved hypothetical proteins.

All proteins were selected based on high average antigenic propensity as determined by the Kolaskar and Tongaonkar method for empirical identification of antigenic determinants [31]. Both predicted amidases, Rv3811 and Rv3717, are associated with cell wall glycopeptides [37,38], were shown to have low protective antigenic properties by the Vaxijen tool. These cell wall glycopeptides are embedded within the cell wall and are associated with hydrolase activity; therefore, their antigenicity may be inhibited [37,39]. Antigen Rv2598 was shown to have the highest probable protective antigen score; however, no functional annotations were represented in Uniprot. Notably, a previous study reported that Rv2598 is potentially induced during the maturation phase of mycobacterial biofilm growth, suggesting a role in drug tolerance [40].

Epitope prediction analysis of immunogenic Mycobacterium tuberculosis proteins

The recognition of antibodies against an immunogenic protein is dependent on the B-cell epitope region [41], making it an ideal candidate for vaccine or diagnostic markers. In this study, both linear and discontinuous epitopes were explored. The linear (continuous) epitope prediction tools were shown to be very sensitive in the detection of epitope regions within the full length of the

sequence. However, in the tertiary protein structure in their native folded state, the discontinuous epitope prediction method was more appropriate, as this provided a visual analysis of surface-exposed epitope regions [42]. The structure-based prediction tools revealed that Ellipro was very sensitive to predicting epitope regions when compared to Discotope. The Ellipro tool modelled the tertiary structures as an ellipsoid, thereby detecting protruding globular surfaces potentially interacting with host antibodies, allowing for the prediction of a greater number of epitopes distributed across all surfaces accessible regions [43].

Predicted Cellular and humoral response of potential immunogenic Mycobacterium tuberculosis proteins.

The humoral immune response has an essential role in clearing *M. tuberculosis* infection within the host [44,45]. Using the C-ImmSim server to simulate the antibody response specific to *M. tuberculosis* CH proteins revealed the IgG and IgM response to the predicted epitope regions. Antigen Rv3209, which was predicted to be involved in host–pathogen interspecies interactions [46], stimulated the strongest humoral response. All proteins exhibited a moderate antibody stimulus with the exception of Rv0934 (PstS1). Contrary to a low IgM and no IgG response in the simulation, previous studies have shown that PstS1 has been associated with high detection rates in patient sera samples relating to the immunoglobulin isotypes in varying populations [28,47,48]. In this study, the predicted cytokine response reflected an upregulation of IFN-γ and interleukin-2. It has been shown that IFN-γ has an antimicrobial role in restricting replication and eliminating *M. tuberculosis* and non-tuberculous mycobacteria, signaling phagosome maturation, granuloma formation, autophagy, and apoptosis [49]. The upregulation of interleukin-2 demonstrates an active immune response to the antigen presented in the simulation, promoting cellular immunity through upregulating the host T-cell response [50].

Fibronectin-binding potential of Mycobacterium tuberculosis conserved hypothetical proteins.

The host extracellular matrix glycoprotein fibronectin (Fn) interacts with a multitude of cellular components, providing major biological functions including wound healing, cell proliferation and differentiation, as well as the regulation of phagocytosis and cell migration [51-53]. Mycobacterial cell-surface adhesins utilize host ECMs for initial attachment, facilitating invasion and spread of infection [17]. Therefore, Fn may have serological biomarker potential as these are largely

interactive with the host [15]. In the current study, Fn-binding potential was explored to demonstrate the functional integrity of the His-affinity tag recombinant protein, thus corroborating the adhesin outcome of the GST-recombinant protein (Chapter 4 in the thesis). Two potential adhesin vaccine candidates are listed by the MycobacRv selection [54], Rv3811 and Rv3491 had a moderate probability of adhesin index predicted by the SPAAN tool and demonstrated a high binding affinity to Fn during the adhesin assay. Conversely, Rv3209 showed poor Fn-binding affinity despite a high predicted probability of adhesin index. Recombinant protein Rv2598 showed no binding, attributable to the low probability of adhesin near the threshold. Recombinant protein Rv3717 yielded a very similar outcome, a high Fn-binding potential, as shown in an earlier study conducted in 2013 [33]. In support of identifying Fn-binding function, studies have shown that fibronectin-binding proteins (FnBPs) of *M. tuberculosis* have a major role in pathogenesis, particularly in increased attachment to alveolar macrophages [55-57]. In addition, FnBPs such as mycolyl transferase (Antigen 85 complex) have been shown to have serological and therapeutic benefits [56,58,59].

IgG antibodies against Mycobacterium tuberculosis conserved hypothetical proteins are detectable in patients' plasma.

In the current scope for TB diagnosis, particularly in high-burden countries, the sensitivity of TB detection is limited and is unable to differentiate the progression of disease within the spectrum [60]. Subsequently, the antibody response during the various stages of infection may provide a clear indication of TB. In a 2024 update, the WHO has therefore endorsed the need for rapid point-of-care diagnostics by assessing the target product profile with a prescribed minimum sensitivity of 70% for non-sputum detection [61]. However, this remains a challenge as the heterogeneous antibody response may vary in different population settings [2,62,63]. Nevertheless, detecting *M. tuberculosis* specific antibodies in patient blood remains an explorative strategy, as protein biomarkers are shown to elicit the host immune response during TB infection, thereby potentially expanding the diagnostic and therapeutic repertoire.

In this study, *M. tuberculosis* CH protein-specific antibodies were detected in GeneXpert TB-positive and GeneXpert TB-negative plasma using an optimized ELISA for detection. Optimization remained a crucial process as all parameters for the ELISA were tested for the most effective detection method. The addition of the control antigen PstS1 was shown to have high

sensitivity and specificity in the detection of TB in previous studies [28,47,48,64], therefore, was included to measure the performance of the ELISA among this particular South African cohort.

Antigen-specific antibody response among TB-positive and TB-negative

In the current study, GeneXpert [65] confirmed samples were obtained of active TB infection or TB-negative status, enabling a comparative assessment of antibody response. Immunglobulin G-specific antibodies to the *M. tuberculosis* CH proteins were measured among TB-positive and TB-negative groups. Rv3811 showed significant differences between groups, resulting in a higher detection of IgG in TB-positive plasma samples relative to Rv3717, even though they both are functionally characterized as amidases [37,38]. Conversely, IgG antibodies specific to Rv3491 were detected at a higher rate within the TB-negative group. This may be attributed to the possibility that the samples have a cross-reactivity with the conserved epitope regions within non-tuberculous mycobacteria [66,67].

In addition, samples from patients with latent TB infection, incipient or subclinical TB phase of infection [60,68] were not tested, and therefore the assay was unable to distinguish TB-negative from these states. Antibodies specific to Rv3491 may be upregulated during the initial stage of infection but downregulated in active TB infection as the host immune response accelerates the containment of infection by shifting the response to recognize another antigenic stimulus. Subsequently, the transition from latent to active replication may alter the gene expression profile of *M. tuberculosis*, thereby downregulating the antibody production towards these specific antigens [9]. This may be attributable to early correlates of clearance, as antigen-based diagnostic tests such as IGRA and tuberculin skin tests only detect TB infection as the adaptive response progresses [69]. Therefore, the downregulation of the Rv3491 antigen may act as a relevant biomarker of poor bacterial control.

Serological performance of the Mycobacterium tuberculosis conserved hypothetical proteins.

The ROC curve showed Rv3811 (AUC=0.622, p= 0.030, cutoff value >0.513) and Rv2598 (AUC=0.646, p=0.008, Cutoff value>0.593) to have the highest diagnostic discriminatory potential compared to all the antigens, including the control, PstS1 (AUC=0.511, p=0.852, cutoff value >1.567). A study conducted by Dewi *et al.* [70] in 2023, elucidated the response of patient

IgG antibodies against native PstS1 purified directly from *M. tuberculosis* and compared it to recombinant PstS1. In comparison to the current study, recombinant PstS1 produced by Dewi *et al.* [70] similarly showed poor ROC curve analysis (AUC = 0.527, p = 0.723, cutoff value > 1.277). This study indicated a very weak performance of the ELISA in detecting antibodies specific to recombinant PstS1 for distinguishing TB-positive from TB-negative samples. In the current study, Youden's index and the classifier evaluation matrix were used to determine the optimal cutoff values for the sensitivity and specificity of detecting TB interpreted by the ROC curve. *M. tuberculosis* antigen Rv2598 showed the highest specificity (98%, CI 89.4–99.9), followed by Rv3209 (88%, CI 75.7-95.5) and PstS1 (80%, CI 66.3-89.9) with low sensitivity (16-42%). Comparatively, Rv3811 was revealed to have moderate sensitivity (50%, CI 35.5-64.5) and specificity (74%, CI 59.6-85.4). Although this study produced poor serological performance, similar studies compared known *M. tuberculosis* adhesins, such as Ag85 complex, which showed high sensitivity (84.1%) in contrast to ESAT-6 (64.9%) and CFP-10 (66%), with high specificity (≥85%) across all antigens [71]. However, multi-epitope fusion constructs have been shown to improve detection with higher sensitivity and specificity [62,72,73].

Combination analysis by logistic regression

A binary logistic regression model showed the relationship between TB detection among various combinations of individual antigens. The outcome corroborated with the Mann-Whitney U test in this current study, showing that Rv2598 and Rv3811 were able to significantly distinguish between TB groups and demonstrated that antibodies against these antigens are present. These findings provided a baseline for combination analysis. Therefore, an interaction-term model was adopted to assess whether combinations of antigens could improve the correct classification of TB-positive and TB-negative samples [74,75]. A previous ELISA-based study showed that the combination of antigens can improve the detection of TB for the detection of IgG, IgM, and mucosal IgA antigen-specific antibodies [76]. In the present study, however, it was shown that combinations of the highest performing proteins did not significantly improve the detection of TB in the sample cohort. Thus, statistically increasing the number of antigen combinations did not significantly improve the correct identification of true-positive and true-negative among the samples. Furthermore, assessing these antigens as biological combinations or as fusion constructs may represent an improved

approach for TB serological detection. Indeed, antigen combination studies have been shown to improve detection in TB-positive and TB-negative cohorts [77-81].

Co-morbidity analysis

The onset of immunoregulatory diseases directly influences TB progression and immune detection [78,82]. Co-infection with HIV and TB remains a global health concern, particularly in settings with high prevalence of HIV [1,83]. This poses challenges in the detection and treatment of co-infection [84]. Tuberculosis disease burden in South Africa is considered highly endemic, especially in settings with a high HIV burden, where effective diagnosis remains challenging [64,85]. This study cohort reflects this burden of co-infection, with higher infection rates observed among males. Furthermore, this finding is concerning, as this demographic is associated with a higher risk of mortality [86-88].

In the present study, a pairwise comparison was designed and performed to evaluate the detection performance between HIV-positive and HIV-negative samples among the TB groups in the cohort. Antigen Rv2598 showed an improved TB detection in HIV-negative samples, suggesting HIV impairs the antibody response in the detection of TB of this antigen in this population. No significant differences in detection were observed for the remaining antigens. However, Rv3491 exhibited a protective effect, showing higher immunoreactivity in uninfected samples than in TB-HIV co-infected samples. A similar trend was observed for Rv3491 when comparing HIV-positive and HIV-negative individuals within TB groups.

Antigen-based diagnostic assays such as QuantiFERON TB Gold In-tube and tuberculin skin test have moderate sensitivity to detecting TB in coinfected patients, similar to that of C-tb (Statens Serum Institute, Copenhagen, Denmark), a novel skin test that employs antigens ESAT-6 and CFP-10, similar to the QuantiFERON TB gold assay [89]. Latent TB detection among people living with HIV is a global priority [1]. In addition to antigen-based assays, lateral flow urine lipoarabinomannan has shown good sensitivity in detecting TB in patients co-infected with HIV [2,7].

6.6. Conclusion

This study sought to characterize the immunogenic properties of conserved hypothetical *M. tuberculosis* proteins with adhesin function. Bioinformatics revealed the presence of surface-exposed antigenic regions within the protein sequence and a simulated host immune response to the selected M. tuberculosis antigens. This study emphasizes the diversity among antibody responses to detect TB by the variation between patient groups. Not all patients in the cohort exhibited detection of antigen-specific antibodies to the same antigens. Amidase protein, Rv3811 detected a moderate IgG response among confirmed TB-positive samples and low response to confirmed TB-negative , however functionally similar Rv3717, performed poorly. Additionally, the physiochemistry of *M. tuberculosis* antigen Rv3491 requires further investigation, as this protein was shown to be downregulated in active TB cases, suggesting immunoregulatory potential as a serodiagnostic marker. In addition, future studies should explore these antigens in relation to the spectrum of disease relating to latent TB and the different stages of progression [60], including downstream application for treatment monitoring [3,34]. Importantly, the detectable antigen-specific antibody responses indicate that these *M. tuberculosis* proteins are recognized by the host immune system, supporting their continued investigation as potential therapeutic biomarkers.

6.6. Limitations/recommendations

In relation to the immunoinformatic analysis, specific upregulation of IFN-γ induced by these conserved hypothetical *M. tuberculosis* antigens suggest potential diagnostic utility within the context of IFN-gamma release assays [90]. Moreover, a study has shown that a combination of IL-2 with IFN-γ as host biomarkers for co-stimulation of CFP-10 and ESAT-6 may enhance the detection of latent TB infection [91]. As a proof-of-concept, *M. tuberculosis* antigen-specific IgG antibodies were detected in a small cohort of South African plasma samples (N=100). However, the ability to detect active TB disease was poor across all antigens in this specific cohort. There exists a diversity among populations with heterologous antibody production. Variation among HIV co-infected samples led to variation in the antibody response, as shown by Rv2598. This indicates that immunosuppression directly affects IgG antibody detection.

This study only investigated active TB detection, however, no confirmation assays of latent TB infection among individuals were performed on the cohort. It is recommended that future

studies provide confirmation of latent infection by performing an IGRA test [80] and detect mycobacteria other than TB (MOTT) by molecular-based methods [92,93]. Further studies are required to address these findings in larger population groups. The analysis based on a combination of multiple antigens in this study was performed statistically with individual ELISA outcomes. Alternatively, a cocktail of purified proteins may have provided a biological comparative analysis, thereby improving IgG response in heterogeneous serological samples. Despite these limitations, there remains potential in further investigating Rv3811 and Rv2598 epitope regions in combination or incorporating them into a fusion construct to improve the detection of these antigen-specific IgG antibodies in a larger cohort.

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Koobashnee Pillay of the discipline of Medical Microbiology, UKZN for her contribution of the purified His-affinity tag Rv0934 (PstS1) protein used as a control for the serological evaluation of the conserved hypothetical *M. tuberculosis* proteins.

FUNDING

The project running costs and partial doctoral scholarship were supported by the University of KwaZulu-Natal (UKZN) APACHE Flagship funds and publication funds (Prof. M. Pillay). Mr. R. Maharajh's doctoral scholarship was further supported by the College of Health Sciences scholarship and the SAMRC Division of Research Capacity Development under the Bongani Mayosi Health Scholars Programme.

AUTHOR CONTRIBUTIONS

RM: conceptualization, methodology, experimental work, data curation, computational and statistical analysis, writing-original manuscript, **SS:** conceptualization, methodology, writing-review and editing, supervision. **VR:** Patient sample curation, statistical analysis **BP:** Writing-review and editing, **MP:** conceptualization, methodology, writing-review and editing, funding acquisition, supervision. All authors: approval of the final manuscript

SUPPLEMENTARY DATA

Available in Appendices B4: Chapter 4 of this thesis

REFERENCE

1. World Health Organization. Global tuberculosis report 2025 Geneva: World Health Organization: World Health Organization. 2025. Available from: https://iris.who.int/server/api/core/bitstreams/e97dd6f4-b567-4396-8680-717bac6869a9/content.

2. World Health Organization. WHO consolidated guidelines on tuberculosis. Module 3: diagnosis–rapid diagnostics for tuberculosis detection Geneva: World Health Organization. 2024. Available from: https://www.who.int/publications/i/item/9789240089488.

3. Heyckendorf J, Georghiou SB, Frahm N, *et al.* Tuberculosis treatment monitoring and outcome measures: new interest and new strategies. Clinical Microbiology Reviews. 2022;35(3):e00227-00221. https://doi.org/10.1128/cmr.00227-21

4. Zhuang L, Yang L, Li L, Ye Z, Gong W. *Mycobacterium tuberculosis*: immune response, biomarkers, and therapeutic intervention. MedComm. 2024;5(1):e419. https://doi.org/10.1002/mco2.419

5. Dong B, He Z, Li Y, Xu X, Wang C, Zeng J. Improved conventional and new approaches in the diagnosis of tuberculosis. Frontiers in Microbiololgy. 2022;13(2022): 4410. https://doi.org/10.3389/fmicb.2022.924410

6. Kontsevaya I, Cabibbe AM, Cirillo DM, *et al.* Update on the diagnosis of tuberculosis. Clinical Microbiology and Infection. 2024;30(9):1115-1122. https://doi.org/10.1016/j.cmi.2023.07.014

7. Zaporojan N, Negrean RA, Hodișan R, Zaporojan C, Csep A, Zaha DC. Evolution of laboratory diagnosis of tuberculosis. Clinics and Practice. 2024;14(2):388-416. https://doi.org/10.3390/clinpract14020030

8. Huang Y, Ai L, Wang X, Sun Z, Wang F. Review and updates on the diagnosis of tuberculosis. Journal of Clinical Medicine. 2022;11(19):5826. https://doi.org/10.3390/jcm11195826

9. Panda S, Kearns K, Cheng C, Lindestam Arlehamn CS. From antigens to immune responses: Shaping the future of TB detection and prevention. International Journal of Infectious Diseases. 2024;141(1):106983. https://doi.org/10.1016/j.ijid.2024.106983

10. Gong W, Wu X. Differential diagnosis of latent tuberculosis infection and active tuberculosis: a key to a successful tuberculosis control strategy. Frontiers in Microbiology. 2021;12(2021). https://doi.org/10.3389/fmicb.2021.745592

11. Flores J, Cancino JC, Chavez-Galan L. Lipoarabinomannan as a point-of-care assay for diagnosis of tuberculosis: how far are we to use it? Front Microbiol. 2021;12:638047. https://doi.org/10.3389/fmicb.2021.638047

12. Nogueira BMF, Krishnan S, Barreto-Duarte B, *et al.* Diagnostic biomarkers for active tuberculosis: progress and challenges. EMBO Molecular Medicine. 2022;14(12):e14088. https://doi.org/10.15252/emmm.202114088

13. Bisht D, Meena LS. Adhesion molecules facilitate host-pathogen interaction & mediate *Mycobacterium tuberculosis* pathogenesis. Indian Journal of Medical Research. 2019;150(1):23-32. https://doi.org/10.4103/ijmr.IJMR_2055_16

14. Yang H, Lei X, Chai S, Zhang S, Su G, Du L. A multi-epitope vaccine incorporating adhesin-derived antigens protects against *Mycobacterium tuberculosis* infection and dissemination. Frontiers in Immunology. 2025;16(2025):1707471. https://doi.org/10.3389/fimmu.2025.1707471

15. Espitia C, Rodríguez E, Ramón-Luing L, Echeverría-Valencia G, Vallecillo AJ. Host-pathogen interactions in tuberculosis. In: Cardona P-J, editor. Understanding Tuberculosis - Analyzing the origin of *Mycobacterium tuberculosis* pathogenicity. London: IntechOpen; 2012. p. 43-76. https://doi.org/10.5772/30310

16. Kuo C-J, Gao J, Huang J-W, *et al.* Functional and structural investigations of fibronectin-binding protein Apa from *Mycobacterium tuberculosis*. Biochimica et Biophysica Acta (BBA) - General Subjects. 2019;1863(9):1351-1359. https://doi.org/10.1016/j.bbagen.2019.06.003

17. Viljoen A, Alsteens D, Dufrêne Y. Mechanical forces between mycobacterial antigen 85 complex and fibronectin. Cells. 2020;9(3):716. https://doi.org/10.3390/cells9030716

18. Jacobs R, Malherbe S, Loxton AG, *et al.* Identification of novel host biomarkers in plasma as candidates for the immunodiagnosis of tuberculosis and monitoring of tuberculosis treatment response. Oncotarget. 2016;7(36):57581. https://doi.org/10.18632/oncotarget.11420

19. Squeglia F, Ruggiero A, De Simone A, Berisio R. A structural overview of mycobacterial adhesins: Key biomarkers for diagnostics and therapeutics. Protein Science. 2018;27(2):369-380. https://doi.org/10.1002/pro.3346

20. Govender VS, Ramsugit S, Pillay M. *Mycobacterium tuberculosis* adhesins: potential biomarkers as anti-tuberculosis therapeutic and diagnostic targets. Microbiology. 2014;160(9):1821-1831. https://doi.org/10.1099/mic.0.082206-0

21. Yang H, Ma Y, Lei X, *et al.* Stopping Tuberculosis at the Gate: The Role of *M. tuberculosis* Adhesins in Infection and Intervention. Vaccines. 2025;13(7):676. https://doi.org/10.3390/vaccines13070676

22. Achkar JM, Chan J, Casadevall A. B cells and antibodies in the defense against *Mycobacterium tuberculosis* infection. Immunological reviews. 2015;264(1):167-181. https://doi.org/10.1111/imr.12276

23. Flores-Gonzalez J, Urbán-Solano A, Ramón-Luing LA, *et al.* Active tuberculosis patients have high systemic IgG levels and B-cell fingerprinting, characterized by a reduced capacity to produce IFN-γ or IL-10 as a response to *M. tuberculosis* antigens. Frontiers in Immunology. 2023;14:1263458. https://doi.org/10.3389/fimmu.2023.1263458

24. Maekura R, Kitada S, Osada-Oka M, *et al.* Serum antibody profiles in individuals with latent *Mycobacterium tuberculosis* infection. Microbiology and immunology. 2019;63(3-4):130-138. https://doi.org/10.1111/1348-0421.12674

25. Singh M, Bhatt P, Sharma M, Varma-Basil M, Chaudhry A, Sharma S. Immunogenicity of late stage specific peptide antigens of *Mycobacterium tuberculosis*. Infection, Genetics and Evolution. 2019;74:103930. https://doi.org/10.1016/j.meegid.2019.103930

26. Stewart P, Patel S, Comer A, *et al.* Role of B cells in *Mycobacterium tuberculosis* infection. Vaccines. 2023;11(5):955. https://doi.org/10.3390/vaccines1105095

27. Bhatt K, Verma S, Ellner JJ, Salgame P. Quest for correlates of protection against tuberculosis. Clinical and Vaccine Immunology. 2015;22(3):258-266. https://doi.org/10.1128/CVI.00721-14

28. Irvine EB, Darrah PA, Wang S, *et al.* Humoral correlates of protection against *Mycobacterium tuberculosis* following intravenous BCG vaccination in rhesus macaques. iScience. 2024;27(12). https://doi.org/10.1016/j.isci.2024.111128

29. Wang J, Fan X-Y, Hu Z. Immune correlates of protection as a game changer in tuberculosis vaccine development. npj Vaccines. 2024;9(1):208. https://doi.org/10.1038/s41541-024-01004-w

30. Maharajh R, Pillay M, Senzani S. A computational method for the prediction and functional analysis of potential *Mycobacterium tuberculosis* adhesin-related proteins. Expert Review of Proteomics. 2023;20(12):483-493. https://doi.org/10.1080/14789450.2023.2275678

31. Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Letters. 1990;276(1-2):172-174. https://doi.org/10.1016/0014-5793(90)80535-q

32. Gordon CH, Hendrix E, He Y, Walker MC. AlphaFold accurately predicts the structure of ribosomally synthesized and post-translationally modified peptide biosynthetic enzymes. Biomolecules. 2023;13(8):1243. https://doi.org/10.3390/biom13081243

33. Kumar S, Puniya BL, Parween S, Nahar P, Ramachandran S. Identification of novel adhesins of *m. tuberculosis* h37rv using integrated approach of multiple computational algorithms and experimental analysis. PLOS ONE. 2013;8(7):e69790. https://doi.org/10.1371/journal.pone.0069790

34. Yong YK, Tan HY, Saeidi A, *et al.* Immune biomarkers for diagnosis and treatment monitoring of tuberculosis: current developments and future prospects. Frontiers in Microbiology. 2019;10(2019):1-18. https://doi.org/10.3389/fmicb.2019.02789

35. Stop TB Partnership. TB vaccine clinical pipeline 2026 [accessed 2026 February 14]. Available from: https://newtbvaccines.org/tb-vaccine-pipeline/clinical-phase/.

36. Sachdeva G, Kumar K, Jain P, Ramachandran S. SPAAN: a software program for prediction of adhesins and adhesin-like proteins using neural networks. Bioinformatics. 2005;21(4):483-491. https://doi.org/10.1093/bioinformatics/bti028

37. Vermassen A, Leroy S, Talon R, Provot C, Popowska M, Desvaux M. Cell wall hydrolases in bacteria: insight on the diversity of cell wall amidases, glycosidases and peptidases toward peptidoglycan. Frontiers in Microbiology. 2019;10:331-331. https://doi.org/10.3389/fmicb.2019.00331

38. Machowski EE, Senzani S, Ealand C, Kana BD. Comparative genomics for mycobacterial peptidoglycan remodelling enzymes reveals extensive genetic multiplicity. BMC Microbiology. 2014;14(1):75. https://doi.org/10.1186/1471-2180-14-75

39. Daffé M, Marrakchi H. Unraveling the Structure of the mycobacterial envelope. Microbiology Spectrum. 2019;7(4):1-11. https://doi.org/10.1128/microbiolspec.gpp3-0027-2018

40. Islam MS, Richards JP, Ojha AK. Targeting drug tolerance in mycobacteria: a perspective from mycobacterial biofilms. Expert Review of Anti-infective Therapy. 2012;10(9):1055-1066. https://doi.org/10.1586/eri.12.88

41. Galanis KA, Nastou KC, Papandreou NC, Petichakis GN, Pigis DG, Iconomidou VA. Linear B-Cell epitope prediction for in silico vaccine design: a performance review of methods available via command-line interface. International Journal of Molecular Sciences. 2021;22(6):3210. https://doi.org/10.1101/833418

42. Zheng D, Liang S, Zhang C. B-Cell Epitope Predictions Using Computational Methods. In: Tsumoto K, Kuroda D, editors. Computer-Aided Antibody Design. New York, NY: Springer US; 2023. p. 239-254. https://doi.org/10.1007/978-1-0716-2609-2_12

43. Ponomarenko J, Bui H-H, Li W, *et al.* ElliPro: a new structure-based tool for the prediction of antibody epitopes. BMC Bioinformatics. 2008;9(1):514. https://doi.org/10.1186/1471-2105-9-514

44. Chandra P, Grigsby SJ, Philips JA. Immune evasion and provocation by *Mycobacterium tuberculosis*. Nature Reviews Microbiology. 2022;20(12):750-766. https://doi.org/10.1038/s41579-022-00763-4

45. Liu D, Wang J, Xu Z, Chen X, Jiao Xa. Phagocytosis: strategies for macrophages to hunt *Mycobacterium tuberculosis*. One Health Advances. 2024;2(1):32. https://doi.org/10.1186/s44280-024-00065-9

46. Verma RN, Malik MZ, Singh GP, Subbarao N. Identification of key proteins in host–pathogen interactions between *Mycobacterium tuberculosis* and *Homo sapiens*: A systematic network theoretical approach. Healthcare Analytics. 2022;2:100052. https://doi.org/10.1016/j.health.2022.100052

47. Ma Z, Ji X, Yang H, *et al.* Screening and evaluation of *Mycobacterium tuberculosis* diagnostic antigens. European Journal of Clinical Microbiology & Infectious Diseases. 2020;39(10):1959-1970. https://doi.org/10.1007/s10096-020-03951-3

48. Han Y, Ma S, Ma R, *et al.* Enhancing detection efficiency of RD-proteins in *Mycobacterium tuberculosis* culture filtrate by biomimetic affinity chromatography coupled with LC-MS/MS. Journal of Chromatography B. 2026;1278:125081. https://doi.org/10.1016/j.jchromb.2026.125081

49. Ghanavi J, Farnia P, Farnia P, Velayati AA. The role of interferon-gamma and interferon-gamma receptor in tuberculosis and nontuberculous mycobacterial infections. The International Journal of Mycobacteriology. 2021;10(4):349-357. https://doi.org/10.4103/ijmy.ijmy_186_21

50. Qiu X, Wang H, Tang Y, *et al.* Is interleukin-2 an optimal marker for diagnosing tuberculosis infection? A systematic review and meta-analysis. Annals of Medicine. 2020;52(7):376-385. https://doi.org/10.1080/07853890.2020.1800073

51. Mosher DF. Physiology of fibronectin. Annual Review of Medicine. 1984;35(1):561-575. https://doi.org/10.1146/annurev.me.35.020184.003021

52. Berdiaki A, Neagu M, Tzanakakis P, Spyridaki I, Pérez S, Nikitovic D. Extracellular matrix components and mechanosensing pathways in health and disease. Biomolecules. 2024;14(9):1186. https://doi.org/10.3390/biom14091186

53. Halper J. Basic components of connective tissues and extracellular matrix: fibronectin, fibrinogen, laminin, elastin, fibrillins, fibulins, matrilins, tenascins and thrombospondins. in: halper j, editor. progress in heritable soft connective tissue diseases. Cham: Springer International Publishing; 2021. p. 105-126. https://doi.org/10.1007/978-3-030-80614-9_4

54. Ramsugit S, Pillay M. Identification of *Mycobacterium tuberculosis* adherence-mediating components: a review of key methods to confirm adhesin function. Iranian Journal of Basic Medical Science. 2016;19(6):579-584. https://dx.doi.org/10.22038/ijbms.2016.7124

55. Pasula R, Wisniowski P, Martin WJ. fibronectin facilitates *Mycobacterium tuberculosis* attachment to murine alveolar macrophages. Infection and Immunity. 2002;70(3):1287-1292. https://doi.org/10.1128/iai.70.3.1287-1292.2002

56. Karbalaei Zadeh Babaki M, Soleimanpour S, Rezaee SA. Antigen 85 complex as a powerful *Mycobacterium tuberculosis* immunogene: biology, immune-pathogenicity, applications in diagnosis, and vaccine design. Microbial Pathogenesis. 2017;112:20-29. https://doi.org/10.1016/j.micpath.2017.08.040

57. Steingart Karen R, Dendukuri N, Henry M, *et al.* Performance of Purified Antigens for Serodiagnosis of Pulmonary Tuberculosis: a Meta-Analysis. Clinical and Vaccine Immunology. 2009;16(2):260-276. https://doi.org/10.1128/CVI.00355-08

58. Mufida DC, Aziz AM, Misturiansyah NI. Potential of B-cell epitopes protein ag85 complex *mycobacterium tuberculosis* as serodiagnostic antigen of tuberculosis by in silico study. Journal of Biomedicine and Translational Research 2023. 2023:7. https://doi.org/10.14710/jbtr.v9i1.16379

59. Ptak CP, Kuo CJ, Chang YF. *Mycobacterium tuberculosis* antigen 85 family proteins: mycolyl transferases and matrix-binding adhesins. moonlighting proteins. 2017:357-369. https://doi.org/10.1002/9781118951149.ch20

60. Migliori GB, Ong CWM, Petrone L, D'Ambrosio L, Centis R, Goletti D. The definition of tuberculosis infection based on the spectrum of tuberculosis disease. Breathe. 2021;17(3):210079. https://doi.org/10.1183/20734735.0079-2021

61. World Health Organization. Target product profiles for tuberculosis diagnosis and detection of drug resistance Geneva: World Health Organization. 2024 [accessed 2026 May 30]. Available from: https://www.who.int/publications/i/item/9789240097698.

62. Arif S, Akhter M, Khaliq A, Nisa Zu, Khan IH, Akhtar MW. Serodiagnostic evaluation of fusion proteins from multiple antigens of *Mycobacterium tuberculosis* for active TB. Tuberculosis. 2021;127(1):102053. https://doi.org/10.1016/j.tube.2021.102053

63. Lyashchenko K, Colangeli R, Houde M, Jahdali HA, Menzies D, Gennaro ML. Heterogeneous antibody responses in tuberculosis. Infection and Immunity. 1998;66(8):3936-3940. https://doi.org/10.1128/iai.66.8.3936-3940.1998

64. Hwang W-H, Lee W-K, Ryoo SW, Yoo K-Y, Tae G-S. Expression, purification and improved antigenicity of the *Mycobacterium tuberculosis* PstS1 antigen for serodiagnosis. Protein Expression and Purification. 2014;95:77-83. https://doi.org/10.1016/j.pep.2013.11.011

65. da Silva MP, Cassim N, Ndlovu S, *et al.* More than a decade of GeneXpert® *Mycobacterium tuberculosis*/Rifampicin (Ultra) testing in South Africa: laboratory insights from twenty-three million tests. Diagnostics. 2023;13(20):3253. https://doi.org/10.3390/diagnostics13203253

66. Lindestam-Arlehamn CS, Paul S, Mele F, *et al.* Immunological consequences of intragenus conservation of *Mycobacterium tuberculosis* T-cell epitopes. Proceedings of the National Academy of Sciences. 2015;112(2):E147-E155. https://doi.org/10.1073/pnas.1416537112

67. Wang X, Li M, Fan X, *et al.* Study of gene homologous antigen, B cell antigen epitope and cross reaction with antibody IgG between two common non-tuberculous Mycobacteria and *Mycobacterium tuberculosis*. Disease Surveillance. 2023;38(3):358-362. https://doi.org/10.3784/jbjc.202301030560

68. Hur Y-G, Kim A, Kang YA, *et al.* Evaluation of antigen-specific immunoglobulin g responses in pulmonary tuberculosis patients and contacts. Journal of Clinical Microbiology. 2015;53(3):904-909. https://doi.org/10.1128/jcm.03050-14

69. Verrall AJ, G. Netea M, Alisjahbana B, Hill PC, van Crevel R. Early clearance of *Mycobacterium tuberculosis*: A new frontier in prevention. Immunology. 2014;141(4):506-513. https://doi.org/10.1111/imm.12223

70. Dewi DNSS, Mertaniasih NM, Soedarsono, *et al.* Antibodies against native proteins of *Mycobacterium tuberculosis* can detect pulmonary tuberculosis patients. Scientific Reports. 2023;13(1):12685. https://doi.org/10.1038/s41598-023-39436-4

71. Kumar G, Dagur PK, Singh PK, *et al.* Serodiagnostic Efficacy of *Mycobacterium tuberculosis* 30/32-kDa mycolyl transferase complex, esat-6, and cfp-10 in patients with active tuberculosis. Archivum Immunologiae et Therapiae Experimentalis. 2010;58(1):57-65. https://doi.org/10.1007/s00005-009-0055-4

72. Brust B, Lecoufle M, Tuaillon E, *et al. Mycobacterium tuberculosis* lipolytic enzymes as potential biomarkers for the diagnosis of active tuberculosis. PLOS ONE. 2011;6(9):e25078. https://doi.org/10.1371/journal.pone.0025078

73. Arif S, Akhter M, Khaliq A, Akhtar MW. Fusion peptide constructs from antigens of *M. tuberculosis* producing high T-cell mediated immune response. PLOS ONE. 2022;17(9):e0271126. https://doi.org/10.1371/journal.pone.0271126

74. Ramsuran V, Kulkarni H, He W, *et al.* Duffy-Null–Associated Low Neutrophil Counts Influence HIV-1 Susceptibility in High-Risk South African Black Women. Clinical Infectious Diseases. 2011;52(10):1248-1256. https://doi.org/10.1093/cid/cir119

75. Cao SH, Chen YQ, Sun Y, *et al.* Screening of serum biomarkers for distinguishing between latent and active tuberculosis using proteome microarray. Biomedical and Environmental Sciences. 2018;31(7):515-526. https://doi.org/10.3967/bes2018.069

76. Raja A, Ranganathan UD, Bethunaickan R. Improved diagnosis of pulmonary tuberculosis by detection of antibodies against multiple *Mycobacterium tuberculosis* antigens. Diagnostic Microbiology and Infectious Disease. 2008;60(4):361-368. https://doi.org/10.1016/j.diagmicrobio.2007.11.006

77. Sun L, Chen Y, Yi P, *et al.* Serological detection of *Mycobacterium tuberculosis* complex infection in multiple hosts by One Universal ELISA. PLOS ONE. 2021;16(10):e0257920. https://doi.org/10.1371/journal.pone.0257920

78. Alonzi T, Repele F, Goletti D. Research tests for the diagnosis of tuberculosis infection. Expert Review of Molecular Diagnostics. 2023;23(9):783-795. https://doi.org/10.1080/14737159.2023.2240230

79. Khademi F, Yousefi-Avarvand A, Derakhshan M, *et al. Mycobacterium tuberculosis* HspX/EsxS fusion protein: gene cloning, protein expression, and purification in *Escherichia coli*. Reports of Biochemistry and Molecular Biology. 2017;6(1):15-21. https://europepmc.org/article/MED/29090225

80. Li J, Wang Y, Yan L, *et al.* Novel serological biomarker panel using protein microarray can distinguish active TB from latent TB infection. Microbes and Infection. 2022;24(8):105002. https://doi.org/10.1016/j.micinf.2022.105002

81. Wu X, Yang Y, Zhang J, *et al.* Humoral Immune Responses against the *Mycobacterium tuberculosis* 38-kilodalton, MTB48, and CFP-10/ESAT-6 antigens in tuberculosis. Clinical and Vaccine Immunology. 2010;17(3):372-375. https://doi.org/10.1128/CVI.00287-09

82. McLean MR, Lu LL, Kent SJ, Chung AW. An Inflammatory story: antibodies in tuberculosis comorbidities. Frontiers in Immunology. 2019;10(2019): 2846. https://doi.org/10.3389/fimmu.2019.02846

83. Bruchfeld J, Correia-Neves M, Källenius G. Tuberculosis and HIV Coinfection. Cold Spring Harbor Perspective in Medicine. 2015;5(7):a017871. https://doi.org/10.1101/cshperspect.a017871

84. Navasardyan I, Miwalian R, Petrosyan A, Yeganyan S, Venketaraman V. HIV–TB coinfection: Current therapeutic approaches and drug interactions. Viruses. 2024;16(3):321. https://doi.org/10.3390/v16030321

85. Bhatt A, Syed ZQ, Singh H. Converging epidemics: A narrative review of tuberculosis (TB) and human immunodeficiency virus (HIV) coinfection. Cureus. 2023;15(10): e47624. https://doi.org/10.7759/cureus.47624

86. Kubjane M, Osman M, Boulle A, Johnson LF. The impact of HIV and tuberculosis interventions on South African adult tuberculosis trends, 1990-2019: a mathematical modeling analysis. International Journal of Infectious Diseases. 2022;122:811-819. https://doi.org/10.1016/j.ijid.2022.07.047

87. Osman M, van Schalkwyk C, Naidoo P, *et al.* Mortality during tuberculosis treatment in South Africa using an 8-year analysis of the national tuberculosis treatment register. Scientific Reports. 2021;11(1):15894. https://doi.org/10.1038/s41598-021-95331-w

88. Nicholson TJ, Hoddinott G, Seddon JA, *et al.* A systematic review of risk factors for mortality among tuberculosis patients in South Africa. Systematic Reviews. 2023;12(1):23. https://doi.org/10.1186/s13643-023-02175-8

89. Hoff ST, Peter JG, Theron G, *et al.* Sensitivity of C-Tb: a novel RD-1-specific skin test for the diagnosis of tuberculosis infection. European Respiratory Journal. 2016;47(3):919-928. https://doi.org/10.1183/13993003.01464-2015

90. Goletti D, Delogu G, Matteelli A, Migliori GB. The role of IGRA in the diagnosis of tuberculosis infection, differentiating from active tuberculosis, and decision making for initiating treatment or preventive therapy of tuberculosis infection. International Journal of Infectious Diseases. 2022;124:S12-S19. https://doi.org/10.1016/j.ijid.2022.02.047

91. Rapulana AM, Mpotje T, Baiyegunhi OO, *et al.* Combined analysis of host IFN-γ, IL-2 and IP-10 as potential LTBI biomarkers in ESAT-6/CFP-10 stimulated blood. Frontiers in Molecular Medicine. 2024;4 (2024). https://doi.org/10.3389/fmmed.2024.1345510

92. Yadav R, Arora J, Bhalla M. Laboratory diagnosis of tuberculosis: conventional and newer methods. International Journal of Scientific Research. 2020;9(10):319-324. https://doi.org/10.21275/SR20930230327

93. Indriani SI, Pratiwi A, Simatupang ETM, Simanjuntak AM. current knowledge of mycobacterium other than tuberculosis (mott) in this current era: definition, taxonomy, and diagnose. Respiratory Science. 2024;4(3):197-208. https://doi.org/10.36497/respirsci.v4i3.122

Chapter 7: Synthesis of Research Findings

7.1. Overview of literature

Communicable diseases, in particular tuberculosis, are responsible for approximately 1.23 million mortalities, an increasing number of active tuberculosis (TB) cases, and place a significant burden on the health care system globally [1]. Strategic efforts to control and reduce the burden of disease have improved over the past decade, while the global plans to stop TB have been strategized for 2023-2030 [2,3]. Thirty high-burden countries account for almost 90% of cases [1,2]. High-burden countries such as South Africa have associated challenges, such as high human immunodeficiency virus (HIV) prevalence, drug resistance, poor access to treatment, and treatment monitoring [1,2,4-6]. However, governmental efforts persist in overcoming these challenges by introducing various end TB campaigns [7].

Effective strategies to improve early diagnosis remain the cornerstone for successful treatment. There exists a need for point-of-care testing, which has the potential to accelerate diagnosis and reduce the need for extensive laboratory-based testing [8]. In addition, current TB vaccine strategies have been under clinical development [9]. To date, there remains a single commercially available Bacillus Calmette-Guérin (BCG) vaccine that has been effective in children, but the protection diminishes as adulthood is reached [10]. Therefore, rapid diagnostic tests and novel vaccines would be ideal to combat various stages within the spectrum of TB progression, targeting the severity of the disease [11].

Recent research has reported that surface-associated molecules called adhesins have shown some promise as diagnostic and therapeutic biomarkers [12], whereby the detection of these biomarkers may be related to the modulation of the host response. Particularly, the exposure of these adhesins with the potential to bind host cellular components, such as the extracellular matrix (ECM), exhibits biomarker characteristics [13-15]. This chapter 7 outlined the importance of *M. tuberculosis* CH proteins with adhesin function and their application in mycobacterial proteomics to further explore the use of these key immunogenic proteins with diagnostic and therapeutic benefits. The current study aimed to identify and validate potential novel immunogenic adhesin/adhesin-related proteins among the CH group of *Mycobacterium tuberculosis* proteins for evaluation in serodiagnostic and vaccine development pipelines.

7.2. Identification of adhesin-related conserved hypothetical proteins of *Mycobacterium tuberculosis* (Chapter 3)

A Computational method for the prediction and functional analysis of potential *Mycobacterium tuberculosis* adhesin-related proteins. Maharajh R, Pillay M, Senzani S. Published in the ***Expert Review of Proteomics*.** 2023;20(12):483-493.

Chapter 3 focused on the identification of *M. tuberculosis* conserved hypotheticals with potential adhesin and adhesin-related functions.

Adhesins enhance the pathogenicity of *M. tuberculosis* as these proteins are the first point of contact with the host cellular environment, and promote attachment [16]. These molecules, therefore, have beneficial therapeutic potential based on their proteomic composition. This physiology differs from that of the host and potentially capable of eliciting an immunological response. This study provided a systematic bioinformatics-based pipeline to identify novel adhesins within the *M. tuberculosis* proteome. Firstly, the software program for the prediction of adhesin and adhesin-like proteins using neural networks (SPAAN) provided a foundation to identify potential adhesins [17]. The dataset of potential adhesins for the *M. tuberculosis* conserved hypothetical group was assessed based on the probability of adhesin (P_{ad}) index within a threshold of $P_{ad} \geq 0.51$.

This study identified 34 known *M. tuberculosis* adhesins within the P_{ad} threshold. A list of 167 *M. tuberculosis* CH proteins was further categorized by predicting the biological function by homology modelling and comparing it to the level of adhesin potential [18]. Subcellular localization enhanced the selection by the prediction of the proteins to be translocated to a specific cell site, by comparing three suitable tools (PSORTb, LocTree3, and MycoSub) [19]. This analysis was corroborated with the signal peptide analysis by secretory prediction methods using SignalP and, in cases of proteins with leaderless translocation sequences, SecretomeP [20,21]. The study provides a dataset of potential adhesins from the unknown proteomic category of CH proteins, thereby addressing the gap in the need to identify novel diagnostic and therapeutic biomarkers.

7.3. The cloning, expression, and purification of recombinant proteins of *Mycobacterium tuberculosis* (Chapter 4)

The use of GST/His affinity tag expression systems for recombinant production of solubilized *Mycobacterium tuberculosis* conserved hypothetical proteins with adhesin potential. Maharajh R, Senzani S, Pillay B, Pillay M. Manuscript formatted to ***Molecular Microbiology*** journal requirements.

Chapter 4 focused on the methodological approach for the cloning, expression, and purification of conserved hypothetical proteins by comparing glutathione-s-transferase (GST) /6x Histidine (His)-affinity tag expression systems. The synthesis of targeted purified proteins of interest holds scientific importance for elucidating their structural and biological roles and potentially serves as a source of essential biomarkers with significant diagnostic and therapeutic implications, particularly related to the pathogenicity of tuberculosis. In this study, the seven *M. tuberculosis* conserved hypothetical proteins discussed in Chapters 5 and 6 were recombinantly produced in two expression systems, GST- and His- affinity tags expressed with pGEX-6-P-1 and pET28a vectors, respectively [22].

Molecular optimization included removal of the signal peptide from the gene sequence during primer design, optimal temperature induction, and isopropyl β-D-1-thiogalactopyranoside (IPTG) concentrations. Purification was shown to significantly improve after the removal of the signal peptide sequence from Rv3717 and Rv3491 by reducing the hydrophobicity of the protein's physiochemistry [23]. In this study, the recombinant plasmids were propagated in *E. coli* Dh5α for improved cloning of the vector and subsequently expressed in *Escherichia coli* BL21 (DE3) for protein expression at optimal induced conditions [24].

Solubilization of the recombinant protein lysate remains the key process to improve purified protein yields. This was the most challenging aspect of synthesis, as the expressed *M. tuberculosis* conserved hypothetical proteins were shown to aggregate within inclusion bodies [25]. Among GST-affinity tag proteins, phosphate buffered saline lysis revealed that the recombinant proteins resided in inclusion bodies found in the pellet fraction. Subsequently, 1% sodium dodecyl sulfate (SDS) method promoted solubilization in the clear lysate fraction [26]. However, the harsh denaturing properties of SDS detergent prevented binding of the GST-affinity tag component to the glutathione agarose. Therefore, the 1% sarkosyl method proposed by Massiah *et al.*[22] for solubilization significantly improved protein yields by neutralizing the anion

detergent with Triton X-100 and 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS). The His-affinity tag proteins were solubilized using the denaturing method, which included a chaotropic agent, 8 M urea, in the lysis buffer, and were subsequently removed by the ethanol precipitation method.

There were several challenges encountered when expressing His-Rv2730 and His-Rv2272 in *E. coli* BL21 (DE3). Recombinant protein His-Rv2730, despite being a confirmed recombinant plasmid, inhibited the growth of *E. coli* when induced with IPTG. This led to the proposition that the His-Rv2730 expression system exposed the toxicity and instability of the membrane-bound protein [27]. Additionally, in spite of attempts to cleave the PreScission site, the GST moiety could not be removed. Furthermore, His-Rv2272, predicted to be associated with cell-wall processes, did not overexpress when induced; however, the expression was detected by western blot. Although all seven GST-affinity-tagged proteins were successfully purified, the application was limited to an adhesin validation study described in Chapter 5, as the 26 kDa GST-affinity tag would sterically interfere with serological detection of anti-*M. tuberculosis* antibodies. Therefore, the His-affinity tag proteins were purified for use in detecting IgG-specific antibodies to Rv3811, Rv3717, Rv3491, Rv3209, and Rv2598, as presented in Chapter 6.

7.4 Validation of *Mycobacterium tuberculosis* conserved hypothetical proteins with adhesin potential (Chapter 5)

Validation of novel *Mycobacterium tuberculosis* adhesins with associated extracellular matrix components: A molecular docking and phenotypic approach. Maharajh R, Pillay M, Mhlongo N, Senzani S. Currently under journal review in ***Proteomics.*** Submitted in May 2026.

Chapter 5 focused on the bioinformatic and phenotypic approach in identifying the adherence potential of selected conserved hypothetical proteins to human-derived extracellular matrix (ECM) proteins, laminin, fibronectin, and collagen.

Seven predicted adhesin-related CH (Rv3811, Rv3717, Rv3491, Rv3209, Rv2598, Rv2272) based on the P_{ad} index described in Chapter 3, were selected to determine the binding affinity to human-derived ECMs. The bioinformatic component of this study focused on the molecular docking potential of the ECM-*M. tuberculosis* interaction using HDOCK [28]. The *M. tuberculosis* CH and ECM structures were predicted using AlphaFold2, and molecular docking was performed to

determine the stability of binding, with their respective docking scores indicating key binding residues. This provided a structural-level understanding of their potential binding sites and the atomic bond interactions between the ECM-*M. tuberculosis* amino acid residues.

The experimental validation of adhesin function required the production of GST-affinity recombinant proteins using the pGEX6-P-1 vector described in Chapter 4. The recombinant proteins were subjected to an enzyme-linked immunosorbent assay designed to determine adhesin potential among the three ECMs. The GST native protein was used to confirm the absence of non-specific binding of the tag to any of the ECMs. Three *M. tuberculosis* conserved hypothetical proteins, Rv3811, Rv3491, and Rv2730, showed a very strong binding intensity to fibronectin, whilst Rv3491 and Rv2730 showed strong to very strong binding intensity to all three ECMs. A previously identified conserved hypothetical protein, Rv3717, by Kumar *et al.* [29], which showed comparable binding to fibronectin, served as a biological control for the adhesin assay.

This study evaluated the SPAAN prediction from Chapter 3 by assessing the molecular docking binding affinity and validated these findings towards the phenotypic adhesin binding assay. Overall, this study confirmed the functional integrity of the recombinant GST- affinity tag proteins by revealing their binding potential to human-derived ECMs.

7.5 Immunological evaluation of conserved hypothetical proteins of *Mycobacterium tuberculosis* (Chapter 6)

Evaluation of novel adhesin-related *Mycobacterium tuberculosis* conserved hypothetical protein biomarkers: an immunoinformatic and serological approach. Maharajh R, Senzani S, Ransuran V, Pillay B, Pillay M. Manuscript formatted to *Tuberculosis and Respiratory Diseases* journal requirements.

Chapter 6 outlined the immunogenic characteristics of the *M. tuberculosis* conserved hypothetical proteins by immunoinformatic profiling, the functional integrity by fibronectin binding potential, and the detection of IgG-specific antibodies in patient plasma samples. The immunogenic characteristics of the conserved hypothetical proteins were explored through immunoinformatic analysis to determine the antigenic epitope regions. The prediction of the *M. tuberculosis* conserved hypothetical proteins with probable protective antigen potential evaluated using Vaxijen 2.0 revealed that all proteins were potential candidates. Linear epitope prediction tools were compared across three platforms (ABCpred, predicted antigenic peptides and Bepipred 2.0) to

identify immunogenic B-cell epitopes [30]. This revealed that linear predictions were highly sensitive in detecting antigenic regions within the protein sequences. When compared to discontinuous epitope predictions (tertiary structure-based), Ellipro and Discotope 2.0, exposed surface amino acids showed favorable antigenic sites to interact with host antibodies. However, Discotope 2.0 was shown to be less sensitive by identifying solvent- accessible surface areas specific to the amino acid composition, spatial and surface exposure.

Furthermore, the immunogenic profiling was ascertained by simulation of the *M. tuberculosis* conserved hypothetical proteins within the host immune system using the C-ImmSim server [31]. The humoral response to IgG and IgM was simulated, showing that all the conserved hypothetical proteins exhibited a moderate response, with the strongest by Rv3209. However, a well-established antigenic adhesin, PstS1, showed low IgM and no IgG response, in contrast to high detection rates of anti-PstS1 IgG antibodies in other studies [32-34]. The cellular response was reported to be significantly high when measuring IFN- γ among all the antigenic stimuli. The functional integrity of the His-affinity tag proteins was explored in this study to determine if the adhesin potential was retained, similar to that described in Chapter 5 with the GST-affinity expression system. The fibronectin binding potential showed comparable outcomes, exhibiting high binding potential of His-Rv3811, His-Rv3717, and His-Rv3491 to human-derived fibronectin similar to that of the GST-affinity expression system.

The serological detection of IgG antibodies specific to the *M. tuberculosis* conserved hypothetical proteins was evaluated among GeneXpert-confirmed TB-positive (n=50) and TB-negative (n=50) cohorts in a proof-of-concept study design. This study used an extensively optimized ELISA to improve the detection of *M. tuberculosis* antigen-specific antibodies in plasma samples from a cohort in South Africa. The study showed a significantly elevated IgG-specific antibody detection level in TB-positive compared to TB-negative samples for Rv3811 and Rv2598 ($p< 0.5$). Conversely, Rv3491 showed elevated levels of IgG reactivity among TB-negative samples compared to TB-positive ($p < 0.001$). There exists a possibility that (i) Rv3491 may be expressed with similar B-cell epitopes among non-tuberculous mycobacteria [35,36], (ii) the spectrum of disease progression was not measured from a latent to subclinical TB phase [11,37] in TB-negative patients, and (iii) an altered gene expression profile that downregulates the Rv3491-specific IgG antibodies in active TB infection while shifting the immune response to

recognize other antigenic stimuli. These outcomes may be attributed to early correlates of clearance and, therefore, a relevant biomarker for measuring bacterial control.

There may exist a cross reactivity within antigens that showed no differences in differentiation of TB status but detected antigen specific IgG antibodies in the ELISA. However, the study showed that the antigen Rv3491 was strongly recognized within the TB negative sera and raises the important question of cross reactivity. These often have high sequence homology across species notably MTB, BCG and NTM. None of the five MTB CH proteins is located within the RD1 region (Rv3871-Rv3879c). As a South African population that is >90% vaccinated with BCG, it should be similar in both groups and not elevated in TB negative group only.

The performance of the ELISA to accurately detect TB-positive compared to TB-negative samples was measured using the ROC curve. The area under the curve (AUC) showed a moderate discriminatory potential for Rv3811 (AUC=0.622, p= 0.030) and Rv2598 (AUC=0.646, p=0.008). However, low sensitivity (50% and 32%) and high specificity (74% and 98%) were exhibited for Rv3811 and Rv2598, respectively. The control PstS1 (AUC= 0.511, p= 0.852) showed poor discriminatory performance within the ELISA and the highest ELISA absorbance cutoff value (1.567) among all the proteins. This demonstrated high reactivity within the cohort to detect PstS1. This cohort produced a similar poor performance in a study that compared purified *M. tuberculosis* PstS1 to recombinant PstS1 [38], with the demonstrating an AUC= 0.527 (p = 0.723, cutoff value: 1.277). Two-combination analysis revealed no improvement in detecting IgG in TB groups, thereby negating the need for three or four antigen combinations. In a population where HIV co-infection directly influences TB disease progression, the study aimed to perform a pairwise comparison to evaluate the detection performance among HIV status and TB status within the cohort. Only Rv2598 showed an improved TB detection among HIV negative groups. The study showed diversity among patients to detect single antigens within the South African cohort. However, detection of *M. tuberculosis* antigen-specific IgG in patient plasma indicates the ability to induce host immune response, suggesting that these proteins may still have therapeutic potential.

7.6 Limitations and recommendations

This study has several limitations at various phases of the project. Computational analysis using the SPAAN tool has demonstrated success in many studies [19,29,39]. However, it was shown that certain known *M. tuberculosis* adhesins were discovered below the P_{ad} threshold, particularly

those *M. tuberculosis* proteins binding to plasminogen [40,41]. The presence of these sequences in other *M. tuberculosis* lineages is unknown. It would be important to ascertain their presence in diverse strains for diagnostic and vaccine design. Therefore, future studies should include identification of these sequences in as many lineage strains as possible, followed by phenotypical measurements of their adherence characteristics. While a larger selection of *M. tuberculosis* potential adhesins from the dataset would have strengthened the experimental validation, the selection was based on the following: P_{ad} index, predicted antigenic peptide, and including those *M. tuberculosis* conserved hypothetical proteins from the MycobacRv list of potential vaccines. This study explored three ECMs for adhesin potential, whilst other ECMs, such as fibrillin, latent TGF-β binding proteins, fibulins, matrilins, tenascins, and thrombospondins, have not been explored in the context of *M. tuberculosis* binding affinity [42]. Furthermore, various adhesin receptor sites, such as mannose-binding receptors, macrophage receptors, and epithelial cells could also be explored further [43,44].

The expression and purification of recombinant *M. tuberculosis* proteins remain challenging, particularly issues relating to the solubilization of inclusion bodies of aggregated hydrophobic proteins. There are different host expression systems, such as *M. smegmatis* [45], *E. coli* BL21 (DE3) Rosetta strain [46], or yeast *Pichia pastoris* [47] that should be explored to improve the solubility of inclusion bodies and membrane-bound proteins that present high toxicity to host expression systems. In this study, Rv2730 showed promising adhesin binding potential and was within the MycobacRv list of potential vaccine candidates [44]. However, due to high toxicity when cloned in the pET28a vector and unsuccessful cleavage of the GST-affinity tag, this could not be explored in the serological evaluation study. The serological evaluation study was designed as proof-of-concept; therefore, was limited by sample size and a diverse population selection. However, the performance in detecting antigen-specific antibodies was poor among this TB cohort. A larger cohort, within the spectrum of disease progression phases and treatment monitoring, would expand the detection criteria for these proteins. Patient data collection should include the stage of infection, bacterial load, treatment administered, diagnostic tests carried out, and HIV viral load, to improve the interpretation of outcomes of ELISA-based studies. In a South African setting with a high burden of disease, whilst GeneXpert-confirmed samples confirm active TB infection, latent TB infection may still be prevalent in GeneXpert-confirmed TB-negative patients. Therefore, more robust diagnostic testing is required across all TB profiles. Despite this,

Rv3811 and Rv2598 can be further investigated for therapeutic development by designing biological combinations either as a cocktail of antigens or a fusion construct. There now exists a dataset of unexplored *M. tuberculosis* CH proteins with adhesin and adhesin-related potential that still requires further evaluation as diagnostic and therapeutic biomarkers.

7.7 Conclusion

This study established a robust computational pipeline to develop a comprehensive dataset of putative adhesins from a group of conserved hypothetical proteins of *M. tuberculosis* . This study provided bioinformatics and experimental evidence for the adhesin-binding potential of a selection of seven conserved hypothetical proteins. Three new *M. tuberculosis* adhesins (Rv3811, Rv3491, and Rv2730) have been identified and functionally characterized, demonstrating strong fibronectin binding affinity and are grouped as fibronectin-binding proteins. *M. tuberculosis* CH proteins, Rv3491 and Rv2730 have, in addition, demonstrated versatile adhesin properties, binding to both collagen and laminin, suggesting a significant role as highly virulent proteins enhancing mycobacterial pathogenicity. Furthermore, the molecular synthesis of recombinant *M. tuberculosis* conserved hypothetical proteins has been extensively investigated in this thesis, exploring two affinity tag expression systems to produce sufficient yields of purified proteins. Understanding the hydrophobicity of *M. tuberculosis* proteins and determining the solubilization method without compromising the protein conformation, protein yield, and purity are a critical aspect of synthesizing recombinant proteins. The choice of expression systems, GST-affinity tag or His-affinity tag in this study, was shown to be strongly dependent on the intended downstream application, namely, exploring biological function or serological evaluation, respectively. In the serological assessment, newly identified adhesin Rv3811 elicited a moderate IgG response among the TB cohort, while the non-adhesin and uncharacterized protein, Rv2598, demonstrated the highest ELISA performance in this study. However, the serodiagnostic performance of these conserved hypothetical proteins did not meet the minimum sensitivity thresholds recommended by the WHO for point-of-care testing in the cohort evaluated. Despite this, the detectable antigen-specific antibody responses indicate that these *M. tuberculosis* proteins are recognized by the host immune system, supporting their continued investigation as potential therapeutic biomarkers or immunomodulatory candidates.

7.8 Mycobacterium tuberculosis conserved hypothetical candidate-level evidence summary. Collectively, this study evidence demonstrates that Rv3811, Rv3491, and Rv2730 are valid novel adhesins based on SPAAN prediction, phenotypic ECM binding, and molecular docking. Rv3811 and Rv2598 show moderate immunogenic potential but limited diagnostic performance (sensitivity < 50%). The inverse correlation of Rv3491 with TB positivity may reflect early clearance, cross-reactivity, or differential expression during infection progression. These findings reveal Rv3811 and Rv2730 as potential candidates for therapeutic targeting based on adhesin function, while Rv2598 and Rv3811 warrant further investigation as vaccine or immunomodulatory candidates rather than serological diagnostic biomarkers.

References

1. World Health Organization. Global tuberculosis report 2025 Geneva: World Health Organization: World Health Organization. 2025. Available from: https://iris.who.int/server/api/core/bitstreams/e97dd6f4-b567-4396-8680-717bac6869a9/content.

2. World Health Organization. WHO consolidated guidelines on tuberculosis. Module 3: diagnosis–rapid diagnostics for tuberculosis detection. Geneva: World Health Organization. 2024. Available from: https://www.who.int/publications/i/item/9789240089488.

3. Pandey BD, Ngwe Tun MM, Shah Y, Suzuki Y, Morita K. Ending tuberculosis by 2030: understanding the transmission. The Lancet Regional Health – Western Pacific. 2023;38. https://doi.org/10.1016/j.lanwpc.2023.100851

4. Churchyard G, Mametja L, Mvusi L, *et al.* Tuberculosis control in South Africa: Successes, challenges and recommendations. South African Medical Journal. 2014;104(3):244-248. https://doi.org/10.7196/samj.7689

5. Heidary M, Shirani M, Moradi M, *et al.* Tuberculosis challenges: Resistance, co-infection, diagnosis, and treatment. European Journal of Microbiology and Immunology. 2022;12(1):1-17. https://doi.org/10.1556/1886.2021.00021

6. Statistics South Africa. Mortality and causes of death in South Africa: Findings from death notification Pretoria, South Africa 2022 [accessed 2026 January 6]. Available from: https://www.statssa.gov.za/publications/P03093/P030932022.pdf&ved=2ahUKEwjW9Nih5LOSAxVRUEEAHa1FMYwQFnoECBkQAQ&usg=AOvVaw26V1pll-Mxsl3wqhmmPmUT.

7. South African Department of Health. national TB recovery plan 4.0: April 2025 – March 2026 Pretoria: South African Department of Health. 2025. Available from:

https://www.health.gov.za/wp-content/uploads/2025/05/TB-Recovery-Plan-4_final_250526-1.pdf&ved=2ahUKEwjF3seOsLOSAxW4W0EAHc8bKyYQFnoECBkQAQ&usg=AOvVaw2cz-LtaKFCB2QSKUijiMrG.

8. Zaporojan N, Negrean RA, Hodișan R, Zaporojan C, Csep A, Zaha DC. Evolution of Laboratory Diagnosis of Tuberculosis. Clinics and Practice. 2024;14(2):388-416. https://doi.org/10.3390/clinpract14020030

9. Stop TB Partnership. TB vaccine clinical pipeline 2026 [cited 2026 14 February]. Available from: https://newtbvaccines.org/tb-vaccine-pipeline/clinical-phase/.

10. Fatima S, Kumari A, Das G, Dwivedi VP. Tuberculosis vaccine: A journey from BCG to present. Life Sciences. 2020;252:117594. https://doi.org/10.1016/j.lfs.2020.117594

11. Migliori GB, Ong CWM, Petrone L, D'Ambrosio L, Centis R, Goletti D. The definition of tuberculosis infection based on the spectrum of tuberculosis disease. Breathe. 2021;17(3):210079. https://doi.org/10.1183/20734735.0079-2021

12. Vinod V, Vijayrajratnam S, Vasudevan AK, Biswas R. The cell surface adhesins of *Mycobacterium tuberculosis*. Microbiological Research. 2020;232:126392. https://doi.org/10.1016/j.micres.2019.126392

13. Bisht D, Meena LS. Adhesion molecules facilitate host-pathogen interaction & mediate *Mycobacterium tuberculosis* pathogenesis. Indian Journal of Medical Research. 2019;150(1):23-32. https://doi.org/10.4103/ijmr.IJMR_2055_16

14. Viljoen A, Dufrêne YF, Nigou J. Mycobacterial adhesion: from hydrophobic to receptor-ligand interactions. Microorganisms. 2022;10(2):454. https://doi.org/10.3390/microorganisms10020454

15. Yong YK, Tan HY, Saeidi A, *et al.* Immune biomarkers for diagnosis and treatment monitoring of tuberculosis: current developments and future prospects. Frontiers in Microbiology. 2019;10(2019): 2789. https://doi.org/10.3389/fmicb.2019.02789

16. Bo H, Moure UAE, Yang Y, *et al. Mycobacterium tuberculosis*-macrophage interaction: Molecular updates. Frontiers in Cellular and Infection Microbiology. 2023;13(2023): 2963. https://doi.org/10.3389/fcimb.2023.1062963

17. Sachdeva G, Kumar K, Jain P, Ramachandran S. SPAAN: a software program for prediction of adhesins and adhesin-like proteins using neural networks. Bioinformatics. 2005;21(4):483-491. https://doi.org/10.1093/bioinformatics/bti028

18. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocol. 2015;10(6):845-858. https://doi.org/10.1038/nprot.2015.053

19. Yang H, Ma Y, Lei X, *et al.* Stopping tuberculosis at the gate: the role of *M. tuberculosis* adhesins in infection and intervention. Vaccines. 2025;13(7):676. https://doi.org/10.3390/vaccines13070676

20. Chiliza TE, Pillay M, Pillay B. Identification of unique essential proteins from a *Mycobacterium tuberculosis* F15/LAM4/KZN phage secretome library. Pathogens and disease. 2017;75(1):ftx001. https://doi.org/10.1093/femspd/ftx001

21. Cornejo-Granados F, Zatarain-Barrón ZL, Cantu-Robles VA, *et al.* Secretome prediction of two *M. tuberculosis* clinical isolates reveals their high antigenic density and potential drug targets. Frontiers in Microbiology. 2017;8:128-128. https://doi.org/10.3389/fmicb.2017.00128

22. Massiah MA, Wright KM, Du H. Obtaining soluble folded proteins from inclusion bodies using sarkosyl, Triton X-100, and CHAPS: Application to LB and M9 minimal media. Current Protocols in Protein Science. 2016;84(1):6.13.11-16.13.24. https://doi.org/10.1002/0471140864.ps0613s84

23. Lai J-S, Cheng C-W, Sung T-Y, Hsu W-L. Computational comparative study of tuberculosis proteomes using a model learned from signal peptide structures. PLOS ONE. 2012;7(4):e35018. https://doi.org/10.1371/journal.pone.0035018

24. Phue J-N, Lee SJ, Trinh L, Shiloach J. Modified *Escherichia coli* B (BL21), a superior producer of plasmid DNA compared with *Escherichia coli* K (DH5α). Biotechnology and Bioengineering. 2008;101(4):831-836. https://doi.org/10.1002/bit.21973

25. Jeffery CJ. Expression, Solubilization, and purification of bacterial membrane proteins. current Protocols in Protein Science. 2016;83(1):29.15.21-29.15.15. https://doi.org/10.1002/0471140864.ps2915s83

26. Schlager B, Straessle A, Hafen E. Use of anionic denaturing detergents to purify insoluble proteins after overexpression. BMC Biotechnology. 2012;12(1):95. https://doi.org/10.1186/1472-6750-12-95

27. Dumon-Seignovert L, Cariot G, Vuillard L. The toxicity of recombinant proteins in *Escherichia coli:* a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). Protein Expression and Purification. 2004;37(1):203-206. https://doi.org/10.1016/j.pep.2004.04.025

28. Yan Y, Tao H, He J, Huang S-Y. The HDOCK server for integrated protein–protein docking. Nature Protocols. 2020;15(5):1829-1852. https://doi.org/10.1038/s41596-020-0312-x

29. Kumar S, Puniya BL, Parween S, Nahar P, Ramachandran S. Identification of novel adhesins of *m. tuberculosis* h37rv using integrated approach of multiple computational algorithms and

experimental analysis. PLOS ONE. 2013;8(7):e69790. https://doi.org/10.1371/journal.pone.0069790

30. Galanis KA, Nastou KC, Papandreou NC, Petichakis GN, Pigis DG, Iconomidou VA. Linear B-cell epitope prediction for in silico vaccine design: a performance review of methods available via command-line interface. International Journal of Molecular Sciences. 2021;22(6):3210. https://doi.org/10.1101/833418

31. Rapin N, Lund O, Castiglione F. Immune system simulation online. Bioinformatics. 2011;27(14):2013-2014. https://doi.org/10.1093/bioinformatics/btr335

32. Ma Z, Ji X, Yang H, *et al.* Screening and evaluation of *Mycobacterium tuberculosis* diagnostic antigens. European Journal of Clinical Microbiology & Infectious Diseases. 2020;39(10):1959-1970. https://doi.org/10.1007/s10096-020-03951-3

33. Irvine EB, Darrah PA, Wang S, *et al.* Humoral correlates of protection against *Mycobacterium tuberculosis* following intravenous BCG vaccination in *Rhesus macaques*. iScience. 2024;27(12): 111128. https://doi.org/10.1016/j.isci.2024.111128

34. Han Y, Ma S, Ma R, *et al.* Enhancing detection efficiency of RD-proteins in *Mycobacterium tuberculosis* culture filtrate by biomimetic affinity chromatography coupled with LC-MS/MS. Journal of Chromatography B. 2026;1278:125081. https://doi.org/10.1016/j.jchromb.2026.125081

35. Lindestam Arlehamn CS, Paul S, Mele F, *et al.* Immunological consequences of intragenus conservation of *Mycobacterium tuberculosis* T-cell epitopes. Proceedings of the National Academy of Sciences. 2015;112(2):E147-E155. https://doi.org/10.1073/pnas.1416537112

36. Wang X, Li M, Fan X, *et al.* Study of gene homologous antigen, B cell antigen epitope and cross reaction with antibody IgG between two common non-tuberculous mycobacteria and *Mycobacterium tuberculosis*. Disease Surveillance. 2023;38(3):358-362. https://doi.org/10.3784/jbjc.202301030560

37. Hur Y-G, Kim A, Kang YA, *et al.* Evaluation of Antigen-Specific Immunoglobulin G Responses in Pulmonary Tuberculosis Patients and Contacts. Journal of Clinical Microbiology. 2015;53(3):904-909. https://doi.org/10.1128/jcm.03050-14

38. Dewi DNSS, Mertaniasih NM, Soedarsono, *et al.* Antibodies against native proteins of *Mycobacterium tuberculosis* can detect pulmonary tuberculosis patients. Scientific Reports. 2023;13(1):12685. https://doi.org/10.1038/s41598-023-39436-4

39. Kamaruzaman INA, Staton GJ, Ainsworth S, Carter SD, Evans NJ. Characterisation of putative outer membrane proteins from *Leptospira borgpetersenii* serovar Hardjo-bovis identifies novel

adhesins and diversity in adhesion across genomospecies orthologs. Microorganisms. 2024;12(2):245.

40. Ayón-Núñez DA, Fragoso G, Bobes RJ, Laclette JP. Plasminogen-binding proteins as an evasion mechanism of the host's innate immunity in infectious diseases. Bioscience Reports. 2018;38(5). https://doi.org/10.1042/bsr20180705

41. Xolalpa W, Vallecillo AJ, Lara M, *et al.* Identification of novel bacterial plasminogen-binding proteins in the human pathogen. Proteomics. 2007;7(18):3332-3341. https://doi.org/10.1002/pmic.200600876

42. Halper J. Basic components of connective tissues and extracellular matrix: fibronectin, fibrinogen, laminin, elastin, fibrillins, fibulins, matrilins, tenascins and thrombospondins. in: halper j, editor. progress in heritable soft connective tissue diseases. Cham: Springer International Publishing; 2021. p. 105-126. https://doi.org/10.1007/978-3-030-80614-9_4

43. Squeglia F, Ruggiero A, De Simone A, Berisio R. A structural overview of mycobacterial adhesins: Key biomarkers for diagnostics and therapeutics. Protein Science. 2018;27(2):369-380. https://doi.org/10.1002/pro.3346

44. Ramsugit S, Pillay M. Identification of *Mycobacterium tuberculosis* adherence-mediating components: a review of key methods to confirm adhesin function. Iranian Journal of Basic Medical Science. 2016;19(6):579-584. https://dx.doi.org/10.22038/ijbms.2016.7124

45. Bashiri G, Baker EN. Production of recombinant proteins in *Mycobacterium smegmatis* for structural and functional studies. Protein Science. 2015;24(1):1-10. https://doi.org/10.1002/pro.2584

46. Rizzi C, Frazzon J, Ely F, *et al.* DAHP synthase from *Mycobacterium tuberculosis* H37Rv: cloning, expression, and purification of functional enzyme. Protein Expression and Purification. 2005;40(1):23-30. https://doi.org/10.1016/j.pep.2004.06.040

47. Benabdesselem C, Barbouche MR, Jarboui MA, Dellagi K, Ho JL, Fathallah DM. High level expression of recombinant *Mycobacterium tuberculosis* culture filtrate protein CFP32 in *Pichia pastoris*. Molecular Biotechnology. 2007;35(1):41-49. https://doi.org/10.1385/MB:35:1:41

Appendices

Appendix A: Research approval

A1. Biomedical Research Ethics Committee approval letter (UKZN)

UNIVERSITY OF KWAZULU-NATAL
INYUVESI YAKWAZULU-NATALI

11 October 2020

Mr Rivesh Maharajh (205514541)
School of Lab Med & Medical Sc
Medical School

Dear Mr Maharajh,

Protocol reference number: BREC/00001970/2020
Project title: Identification of Mycobacterium tuberculosis immunogenic adhesin proteins for the application as a potential vaccine and a diagnostic candidate
Degree: PhD

EXPEDITED APPLICATION: APPROVAL LETTER

A sub-committee of the Biomedical Research Ethics Committee has considered and noted your application.

The conditions have been met and the study is given full ethics approval and may begin as from 11 October 2020. Please ensure that outstanding site permissions are obtained and forwarded to BREC for approval before commencing research at a site.

This approval is subject to national and UKZN lockdown regulations dated 26th August 2020, see (http://research.ukzn.ac.za/Libraries/BREC/BREC_Lockdown_Level_2_Guidelines.sflb.ashx). Based on feedback from some sites, we urge PIs to show sensitivity and exercise appropriate consideration at sites where personnel and service users appear stressed or overloaded.

This approval is valid for one year from 11 October 2020. To ensure uninterrupted approval of this study beyond the approval expiry date, an application for recertification must be submitted to BREC on the appropriate BREC form 2-3 months before the expiry date.

Any amendments to this study, unless urgently required to ensure safety of participants, must be approved by BREC prior to implementation.

Your acceptance of this approval denotes your compliance with South African National Research Ethics Guidelines (2015), South African National Good Clinical Practice Guidelines (2006) (if applicable) and with UKZN BREC ethics requirements as contained in the UKZN BREC Terms of Reference and Standard Operating Procedures, all available at http://research.ukzn.ac.za/Research-Ethics/Biomedical-Research-Ethics.aspx.

BREC is registered with the South African National Health Research Ethics Council (REC-290408-009). BREC has US Office for Human Research Protections (OHRP) Federal-wide Assurance (FWA 678).

The sub-committee's decision will be noted by a full Committee at its next meeting taking place on 10 November 2020.

Yours sincerely,

[redacted]

Prof D Wassenaar
Chair: Biomedical Research Ethics Committee

Biomedical Research Ethics Committee
Chair: Professor D R Wassenaar
UKZN Research Ethics Office Westville Campus, Govan Mbeki Building
Postal Address: Private Bag X54001, Durban 4000
Email: BREC@ukzn.ac.za
Website: http://research.ukzn.ac.za/Research-Ethics/Biomedical-Research-Ethics.aspx

Founding Campuses: Edgewood Howard College Medical School Pietermaritzburg Westville

INSPIRING GREATNESS

A2. Gatekeepers' permission letter for plasma sample analysis

UNIVERSITY OF KWAZULU-NATAL
INYUVESI YAKWAZULU-NATALI

25 June 2024

Prof. Veron Ramsuran
Nelson Mandela School of Medicine, University of KwaZulu-Natal (UKZN)
719 Umbilo Road
Durban, 4001, South Africa

Research Office
Biomedical Research Ethics Administration
Westville Campus, Govan Mbeki Building
University of KwaZulu-Natal
Private Bag X5401
Durban
4000
KwaZulu-Natal, South Africa
Tel: +27312604769
Email: BREC@ukzn.ac.za

To Whom It May Concern:

RE: Gate Keepers permission for sample use for study: Identification of Mycobacterium tuberculosis immunogenic adhesin proteins for the application as a potential vaccine and a diagnostic candidate (BREC/00001970/2020), R Maharajh

I hereby grant permission for the use of clinical specimens collected under the approved study (reference number BREC/00005530/2023) in the abovementioned research project. The research proposed does fall within the scope of the original ethical approval, i.e. testing of TB proteins from plasma samples. The stored samples will remain within my labs repository until required.

Please do not hesitate to contact me if you have any queries.

Kind regards

[redacted]

Yours sincerely

Prof. Veron Ramsuran

Professor: Nelson R. Mandela School of Medicine, College of Health Sciences (CHS)
Research Fellow: UK Royal Society and AAS - Future Leader of African Independent Research (FLAIR)
Guest Researcher: Leidos Biomedical Research Inc. - Contractor to NCI (NIH)
Associate Scientist: Centre for the AIDS Programme of Research in South Africa (CAPRISA)
Affiliate Member: African Academy of Sciences (AAS)

Physical: 719 Umbilo Road, Umbilo, Durban, 4001
Tel: +27 (0) 260 1929
Email: RamsuranV@ukzn.ac.za

UNIVERSITY OF KWAZULU-NATAL INYUVESI YAKWAZULU-NATALI
INSPIRING GREATNESS

Appendix B: Supplementary data

B1. Chapter three supplementary data

Supplementary table 1: Phyre2 analysis of categorized *Mycobacterium tuberculosis* known adhesin gene loci (n=34).

Gene loci	P_{ad} score	Adhesin	Phyre2 alignment	Phyre2 Significance	Phyre2 Coverage
Cell wall and cell wall processes (n=8)					
Rv3312a	0.672	MTP	endoglycosidase	48.3	9
Rv3875	0.937	ESAT-6	Ferritin-like ESAT6	99	99
Rv3763	0.759	LpqH	Unknown (LpgH)	99	100
Rv3194c	0.637	Rv3194c	Structural protein	99.3	81
Rv0475	0.682	HBHA	Lipid transport	70	32
Rv0309	0.876	Rv0309	hydrolase	99.6	64
Rv0934	0.845	PstS1	Periplasmic binding	100	89
Rv0418	0.703	LpqL	Hydrolase (Zn-dependent exopeptidase)	100	54
Rv1860	0.582	Apa	Lyase (exopolygalacturonate lyase)	0.8	3
PE/PPE (n=2)					
Rv1759c	0.907	Wag22	Structural protein (contractile)	99.7	27
Rv1818c	0.902	PE-PGRS33	Structural protein	99.8	76
Regulatory Proteins (n=2)					
Rv3804c	0.877	Ag85a	hydrolase	100	85
Rv0931c	0.582	PknD	Protein kinase	100	63
Lipid Metabolism (n=4)					
Rv0129c	0.865	Ag85c	hydrolases	100	82
Rv1980c	0.827	Mpt64	unknown	100	89
Rv1886c	0.764	Ag85b	hydrolases	100	87
Rv3803c	0.816	Mpt51	hydrolase	100	89
Intermediary metabolism/respiration (n=8)					
Rv0363c	0.712	Fba	Hydrolase	100	89
Rv0462	0.611	LpdC	Oxidoreductase (dihydrolipoyldehydrogenase)	100	100
Rv0467	0.532	IcL	Lyase/ phosphoenoylpyruvate mutase	100	100
Rv2220	0.396	Glutamine synthase A1	Ligase (glutamine synthetase)	100	100
Rv2110c	0.357	PrcB	Hydrolase (proteasome)	100	85
Rv3248c	0.349	SahH	hydrolase	100	98
Rv1392	0.290	MetK	lyase (adenosylmethionine synthase)	100	93
Rv1837c	0.270	Malate synthase	Malate synthase	100	95
Information Pathway (n=2)					
Rv2986c	0.641	Hup	DNA binding protein	99	46
Rv0685	0.339	Ef-tu	Ribosomal (EFTU)	100	76
Virulence/detoxification/adaptation(n=5)					
Rv0619	0.792	Mce-1	Transport protein MCE	100	24
Rv0440	0.420	GroEL2	chaperone	100	92
Rv3417c	0.418	GroEL1	Chaperone	100	83
Rv3418c	0.399	GroES	Chaperone (cpn60.2)	100	70
Rv0350	0.371	DnaK	Chaperone (DnaK)	100	56
Conserved Hypotheticals (n=3)					
Rv3656c	0.536	Type IV pili	Cell cycle/lipid binding	58.3	34
Rv2599	0.913	Rv2599	Alpha-lytic protease	46	28
Rv3717	0.838	Rv3717	Hydrolase (peptidoglycan amidase)	100	72

d values below the selection threshold (≥0.51)

Supplementary table 2: List of *Mycobacterium tuberculosis* Conserved hypothetical proteins within the P_{ad} index of ≥0.51 (n=167)

P_{ad} Scor	Gene Loci	Uniprot/ Pfam Annotation	Phyre2 alignment	phyre2 confidence	phyre2 coverage	PsortB Analysis	Mycosub Analysis	Loctree3 Analysis	Subcellular Localization Output	SignalP/ SecretomeP
0.909	Rv0991c	Conserved serine-rich protein	metal binding protein	97	36	Unknown	Secretory	Secreted extracellular	Secreted	Non-classical pathway
0.898	Rv0387c	PPE-SVP domain-containing protein	DNA-binding protein	23	11	Unknown	Secretory	Secreted extracellular	Secreted	Unknown
0.886	Rv2599	Uncharacterized protein Rv2599	lytic protease	46	28	Cytoplasmic membrane	Cytoplasmic	Plasma membrane	Cytoplasmic	TAT Signal peptide
0.879	Rv1116a	Uncharacterized protein	RNA-binding protein	26	28	Unknown	Cytoplasmic	Cytoplasm	Cytoplasmic	Unknown
0.879	Rv2575	Uncharacterized protein Rv2575	Hydrolase	98	46	Unknown	Cytoplasmic	Inner membrane	Cytoplasmic	Unknown
0.872	Rv3207c	Conserved protein	zincin metalloprotease	98	45	Unknown	Cytoplasmic	Secreted extracellular	Cytoplasmic	Signal peptide
0.871	Rv3572	Uncharacterized protein	Transferase	38	31	Unknown	Cytoplasmic	Secreted extracellular	Secreted	Signal peptide
0.858	Rv0061c	Uncharacterized protein	Lyase	30	6	Periplasmic	Cytoplasmic	Secreted extracellular	Secreted	Signal peptide
0.857	Rv1906c	Conserved protein	motor protein	18	28	Unknown	Secretory	Secreted extracellular	Secreted	Signal peptide
0.838	Rv3717	N-acetylmuramoyl-L-alanine amidase Rv3717	sugar-binding protein	100	90	Unknown	Cytoplasmic	Secreted extracellular	Cytoplasmic	Signal peptide
0.814	Rv2077a	Uncharacterized protein	methyl transferase	18	21	Unknown	Secretory	Secreted extracellular	Secreted	Non-classical pathway
0.807	Rv3337	Uncharacterized protein	Hydrolase	98	41	Unknown	Secretory	Secreted extracellular	Secreted	Unknown
0.806	Rv1810	Conserved protein	Oxidoreductase	23	22	Unknown	Membrane	Plasma membrane	Membrane Bound	Signal peptide
0.796	Rv3103c	Hypothetical proline-rich protein	membrane protein	78	20	Unknown	Secretory	Secreted extracellular	Secreted	Unknown
0.795	Rv3209	Conserved hypothetical threonine and proline-rich protein	membrane protein	100	41	Unknown	Membrane	Secreted extracellular	Membrane Bound	Unknown
0.792	Rv3705c	Conserved protein	protein transport	100	88	Unknown	Cytoplasmic	Secreted extracellular	Cytoplasmic	Signal peptide
0.791	Rv3822	acyltransferase Chp1	Hydrolase	100	48	Unknown	Membrane	Secreted extracellular	Membrane Bound	Unknown
0.787	Rv0378	Conserved hypothetical glycine-rich protein	Transferase	62	25	Unknown	Secretory	Secreted extracellular	Secreted	Unknown
0.785	Rv1137c	Uncharacterized protein	Transferase	9	8	Unknown	Secretory	Secreted extracellular	Secreted	Unknown

P_{ad} Score	Gene Loci	Uniprot/ Pfam Annotation	Phyre2 alignment	phyre2 confidence	phyre2 coverage	PsortB Analysis	Mycosub Analysis	Loctree3 Analysis	Subcellular Localization Output	SignalP/ SecretomeP
0.778	Rv2641	Cadmium-induced protein CadI	Glyoxalase	99	90	Unknown	Cytoplasmic	Cytoplasm	Cytoplasmic	Unknown
0.777	Rv3369	Conserved protein	Unknown	100	95	Cytoplasmic	Cytoplasmic	Cytoplasm	Cytoplasmic	Non-classical pathway
0.777	Rv3491	Uncharacterized protein	protein binding	43	12	Cytoplasmic membrane	Cytoplasmic	Inner membrane	Cytoplasmic	Signal peptide
0.772	Rv1891	Conserved protein	sugar-binding protein glucomylase	15	21	Cytoplasmic membrane	Cytoplasmic	Secreted extracellular	Secreted	Signal peptide
0.764	Rv3897c	Uncharacterized protein	structural protein	100	99	Unknown	Cytoplasmic	Secreted extracellular	Secreted	Signal peptide
0.761	Rv0455c	Conserved protein	oxidoreductase	28	24	Unknown	Secretory	Plasma membrane	Secreted	Signal peptide
0.761	Rv2078	Uncharacterized protein Rv2078	RNA-binding protein	13	18	Unknown	Secretory	Secreted extracellular	Secreted	Non-classical pathway
0.752	Rv0963c	Uncharacterized protein Rv0963c	alpha/beta hydrolase	98	39	Unknown	Membrane	Cytoplasm	Membrane Bound	Non-classical pathway
0.747	Rv0577	Putative glyoxylase CFP32	Glyoxalase	100	99	Unknown	Secretory	Secreted extracellular	Secreted	Unknown
0.746	Rv0614	Uncharacterized protein	protein binding	43	12	Unknown	Cytoplasmic	Secreted extracellular	Secreted	Unknown
0.746	Rv1815	Uncharacterized protein Rv1815	Hydrolase	100	80	Unknown	Membrane	Secreted extracellular	Secreted	Signal peptide
0.743	Rv2998	Uncharacterized protein	Diaminopimelate epimerase	13	38	Unknown	Cytoplasmic	Secreted extracellular	Secreted	Non-classical pathway
0.736	Rv2257c	Conserved protein	Hydrolase	100	100	Unknown	Membrane	Plasma membrane	Membrane Bound	Signal peptide
0.735	Rv0679c	Conserved threonine-rich protein	transmembrane helix	34	12	Extracellular	Membrane	Secreted extracellular	Secreted	Lipoprotein Signal peptide
0.729	Rv2307c	Hypothetical glycine-rich protein	Hydrolase	100	79	Cytoplasmic membrane	Membrane	Inner membrane	Membrane Bound	Unknown
0.723	Rv3747	Conserved protein	transport protein	71	17	Cytoplasmic membrane	Secretory	Cytoplasm	Membrane Bound	Non-classical pathway
0.716	Rv2114	Conserved protein	hydrolase	41	12	Unknown	Membrane	Cytoplasm	Membrane Bound	Non-classical pathway
0.713	Rv1754c	Conserved protein	hydrolase	100	52	Cytoplasmic membrane	Membrane	Secreted extracellular	Membrane Bound	Unknown
0.705	Rv1158c	Conserved hypothetical	antimicrobial protein	8	6	Cytoplasmic	Secretory	Secreted extracellular	Secreted	Signal peptide
0.704	Rv2744c	PspA protein	lipid binding protein	98	77	Cytoplasmic	Cytoplasmic	Cytoplasm	Cytoplasmic	Unknown
0.703	Rv3035	Uncharacterized protein Rv3035	oxidoreductase	100	99	Cytoplasmic membrane	Cytoplasmic	Secreted extracellular	Membrane Bound	Non-classical pathway
0.702	Rv1268	Uncharacterized protein Rv1268c	transport protein	100	78	Unknown	Cytoplasmic	Secreted extracellular	Cytoplasmic	Signal peptide
0.701	Rv1804	Conserved protein	oxidoreductase	18	25	Unknown	Cytoplasmic	Plasma membrane	Cytoplasmic	Signal peptide

*P*ad Score	Gene Loci	Uniprot/ Pfam Annotation	Phyre2 alignment	phyre2 confidence	phyre2 coverage	PsortB Analysis	Mycosub Analysis	Loctree3 Analysis	Subcellular Localization Output	SignalP/ SecretomeP
0.699	Rv1975	SCP domain-containing protein	unknown extracellular protein	100	54	Unknown	Membrane	Secreted extracellular	Secreted	Signal peptide
0.699	Rv0712	Formylglycine-generating enzyme	membrane protein	100	100	Cytoplasmic	Cytoplasmic	Cytoplasm	Cytoplasmic	Unknown
0.694	Rv3090	PHB domain-containing protein	membrane protein	99	54	Unknown	Cytoplasmic	Inner membrane	Membrane Bound	Non-classical pathway
0.694	Rv1203c	Uncharacterized protein	transferase	97	97	Cytoplasmic	Secretory	Cytoplasm	Cytoplasmic	Non-classical pathway
0.694	Rv1134	Uncharacterized protein	ferredoxin protein	99	65	Unknown	Cytoplasmic	Cytoplasm	Cytoplasmic	Non-classical pathway
0.693	Rv3707c	DUF4185 domain-containing protein	hydrolase	100	83	Unknown	Cytoplasmic	Secreted extracellular	Secreted	Unknown
0.685	Rv2264c	Conserved hypothetical proline rich protein	chaperone	57	100	Unknown	Membrane	Secreted extracellular	Membrane Bound	Signal peptide
0.684	Rv3811	PGRP domain-containing protein	hydrolase	100	32	Extracellular	Membrane	Secreted extracellular	Membrane Bound	Non-classical pathway
0.680	Rv1100	Conserved protein	lipid binding protein	38	56	Unknown	Secretory	Secreted extracellular	Secreted	Non-classical pathway
0.679	Rv2660c	Uncharacterized protein	cell adhesin/calcium binding	12	24	Unknown	Cytoplasmic	Secreted extracellular	Cytoplasmic	Non-classical pathway
0.679	Rv2240	Uncharacterized protein Rv2240c	hydrolase	43	1	Unknown	Cytoplasmic	Secreted extracellular	Secreted	Lipoprotein Signal peptide
0.678	Rv2960c	Uncharacterized protein	methyl transferase	20	17	Unknown	Secretory	Secreted extracellular	Secreted	Non-classical pathway
0.675	Rv1951c	Uncharacterized protein	hydrolase	8	36	Unknown	Secretory	Secreted extracellular	Secreted	Non-classical pathway
0.674	Rv3353c	Uncharacterized protein	oxidoreductase	99	63	Cytoplasmic membrane	Secretory	Secreted extracellular	Secreted	Lipoprotein Signal peptide
0.672	Rv2706c	Uncharacterized protein	toxin	27	31	Cytoplasmic membrane	Secretory	Secreted extracellular	Secreted	Non-classical pathway
0.667	Rv1057	Seven-bladed beta-propeller protein	apoptotic protease	100	98	Unknown	Membrane	Secreted extracellular	Membrane Bound	Non-classical pathway
0.666	Rv0381c	Uncharacterized protein	hydrolase	25	23	Unknown	Cytoplasmic	Secreted extracellular	Cytoplasmic	Lipoprotein Signal peptide
0.659	Rv3354	domain-containing protein	oxidoreductase	19	19	Cytoplasmic membrane	Membrane	Secreted extracellular	Membrane Bound	Signal peptide
0.658	Rv3067	Uncharacterized protein	lyase -tryptophan synthase	58	38	Cytoplasmic membrane	Secretory	Secreted extracellular	Secreted	Signal peptide
0.656	Rv2283	Uncharacterized protein	oxidoreductase	6	32	Unknown	Secretory	Secreted extracellular	Secreted	Unknown
0.654	Rv3718	Conserved protein	unknown	100	99	Unknown	Cytoplasmic	Cytoplasm	Cytoplasmic	Unknown
0.654	Rv1109	Uncharacterized protein	hydrolase	60	9	Unknown	Secretory	Secreted extracellular	Secreted	No signal Peptide present

P_{ad} Score	Gene Loci	Uniprot/ Pfam Annotation	Phyre2 alignment	phyre2 confidence	phyre2 coverage	PsortB Analysis	Mycosub Analysis	Loctree3 Analysis	Subcellular Localization Output	SignalP/ SecretomeP
0.650	Rv0454	Uncharacterized protein	RuBisCO small subunit	56	20	Unknown	Secretory	Secreted extracellular	Secreted	Unknown
0.648	Rv2820c	CRISPR system Cms protein Csm4	RNA-binding protein	100	94	Unknown	Membrane	Cytoplasm	Membrane Bound	Unknown
0.647	Rv0340	Conserved protein	hydrolase	99	53	Unknown	Cytoplasmic	Secreted extracellular	Secreted	Non-classical pathway
0.643	Rv2302	Uncharacterized protein	transferase	82	45	Unknown	Secretory	Cytoplasm	Cytoplasmic	Unknown
0.640	Rv3136A	Conserved protein	oxidoreductase	28	20	Cytoplasmic	Secretory	Inner membrane	Membrane Bound	Unknown
0.640	Rv2292	Uncharacterized protein	hydrolase	100	86	Unknown	Cytoplasmic	Secreted extracellular	Cytoplasmic	Non-classical pathway
0.640	Rv0725c	S-adenosyl-L-methionine-dependent methyltransferase	transferase	100	67	Cytoplasmic	Membrane	Cytoplasm	Cytoplasmic	Unknown
0.639	Rv3898c	Uncharacterized protein	cell invasion	24	10	Unknown	Secretory	Secreted extracellular	Secreted	Non-classical pathway
0.636	Rv0397a	Conserved protein	hydrolase	98	41	Unknown	Secretory	Inner membrane	Secreted	Signal peptide
0.633	Rv3706c	Conserved hypothetical proline rich protein	membrane protein	82	39	Unknown	Membrane	Secreted extracellular	Membrane Bound	Non-classical pathway
0.628	Rv2525	Putative peptidoglycan hydrolase	hydrolase/unknown	100	86	Extracellular	Membrane	Secreted extracellular	Membrane Bound	Unknown
0.622	Rv3033	Uncharacterized protein	hydrolase	37	15	Unknown	Cytoplasmic	Secreted extracellular	Cytoplasmic	Lipoprotein Signal peptide
0.619	Rv1998c	Uncharacterized protein Rv1998c	isomerase	100	95	Cytoplasmic	Cytoplasmic	Cytoplasm	Cytoplasmic	Non-classical pathway
0.619	Rv3267	Conserved protein (CPSA-related protein)	membrane protein	100	61	Cytoplasmic membrane	Membrane	Inner membrane	Membrane Bound	Signal peptide
0.615	Rv0471c	prenyltransferase	transferase	99	48	Unknown	Cytoplasmic	Cytoplasm	Cytoplasmic	Non-classical pathway
0.615	Rv2730	Uncharacterized protein	membrane protein	50	5	Extracellular	Cytoplasmic	Secreted extracellular	Secreted	Non-classical pathway
0.609	Rv2082	Uncharacterized protein	signal protein	99	29	Cytoplasmic membrane	Membrane	Secreted extracellular	Membrane Bound	Unknown
0.608	Rv1157c	Conserved ala-protein	hydrolase	0	1	Cytoplasmic	Secretory	Inner membrane	Secreted	TAT Signal peptide
0.607	Rv1875	Conserved protein	oxidoreductase	100	95	Unknown	Secretory	Cytoplasm	Cytoplasmic	Non-classical pathway
0.602	Rv2929	Uncharacterized protein	transport protein	16	7	Unknown	Secretory	Secreted extracellular	Secreted	Unknown
0.600	Rv2422	Uncharacterized protein	hydrolase	43	31	Unknown	Cytoplasmic	Secreted extracellular	Cytoplasmic	Unknown
0.596	Rv2083	Uncharacterized protein	ferredoxin like	4	3	Unknown	Cytoplasmic	Secreted extracellular	Secreted	Non-classical pathway

P_{ad} Score	Gene Loci	Uniprot/ Pfam Annotation	Phyre2 alignment	phyre2 confidence	phyre2 coverage	PsortB Analysis	Mycosub Analysis	Loctree3 Analysis	Subcellular Localization Output	SignalP/ SecretomeP
0.596	Rv0274	glyoxalase	unknown	38	40	Unknown	Cytoplasmic	Cytoplasm	Cytoplasmic	Unknown
0.595	Rv2229c	hydrolase Acylphosphatase	hydrolase/unknown	100	96	Cytoplasmic	Cytoplasmic	Cytoplasm	Cytoplasmic	Non-classical pathway
0.594	Rv0021c	Nitronatate monooxygenase	oxidoreductase	100	95	Cytoplasmic	Cytoplasmic	Cytoplasm	Cytoplasmic	Unknown
0.591	Rv3705a	Conserved hypothetical proline-rich protein	oxidoreductase	82	57	Cytoplasmic membrane	Secretory	Secreted extracellular	Secreted	Non-classical pathway
0.590	Rv2577	Uncharacterized protein	hydrolase	100	82	Unknown	Cytoplasmic	Secreted extracellular	Cytoplasmic	TAT Signal peptide
0.589	Rv3683	Conserved protein	hydrolase	100	60	Unknown	Membrane	Cytoplasm	Membrane Bound	Unknown
0.589	Rv3716	Nucleoid-associated protein Rv3716c	DNA-binding protein	100	100	Unknown	Cytoplasmic	Cytoplasm	Cytoplasmic	Non-classical pathway
0.586	Rv0968	Uncharacterized protein	transferase	38	22	Unknown	Secretory	mitochondria	Secreted	Unknown
0.586	Rv3745c	Uncharacterized protein	unknown	100	99	Unknown	Secretory	Secreted extracellular	Secreted	Unknown
0.584	Rv0911	Conserved protein	oxidoreductase	100	93	Cytoplasmic	Cytoplasmic	Secreted extracellular	Cytoplasmic	Unknown
0.584	Rv3796	Conserved protein	hydrolase	100	91	Unknown	Membrane	Plasma membrane	Membrane Bound	Lipoprotein Signal peptide
0.581	Rv3541c	3-oxo-4,17-pregnadiene-20-carboxyl-CoA hydratase beta subunit	lyase	100	98	Unknown	Cytoplasmic	Cytoplasm	Cytoplasmic	Unknown
0.580	Rv2256c	Uncharacterized protein	hydrolase	63	28	Unknown	Cytoplasmic	Cytoplasm	Cytoplasmic	Non-classical pathway
0.578	Rv1419	Uncharacterized protein	hydrolase	100	79	Unknown	Secretory	Secreted extracellular	Secreted	Signal peptide
0.577	Rv1961	Uncharacterized protein	viral protein receptor binding	55	7	Unknown	Cytoplasmic	Cytoplasm	Cytoplasmic	Non-classical pathway
0.576	Rv0500a	Putative DNA-binding protein	DNA-binding protein	100	99	Unknown	Cytoplasmic	Cytoplasm	Secreted	Non-classical pathway
0.576	Rv3655	Apoptosis inhibitor	glycosidase	16	15	Unknown	Secretory	Inner membrane	Secreted	Signal peptide
0.573	Rv0979c	Uncharacterized protein	cell adhesin-integrin IIB	7	14	Cytoplasmic membrane	Secretory	Secreted extracellular	Secreted	Unknown
0.572	Rv3484	Possible exported protein	membrane protein	100	53	Unknown	Cytoplasmic	Secreted extracellular	Cytoplasmic	Non-classical pathway
0.569	Rv1519	Uncharacterized protein	transferase	92	44	Unknown	Secretory	Secreted extracellular	Secreted	Non-classical pathway
0.567	Rv3282	Nucleoside triphosphate pyrophosphatase	hydrolase	100	90	Cytoplasmic	Membrane	Cytoplasm	Cytoplasmic	Unknown

P_{ad} Score	Gene Loci	Uniprot/ Pfam Annotation	Phyre2 alignment	phyre2 confidence	phyre2 coverage	PsortB Analysis	Mycosub Analysis	Loctree3 Analysis	Subcellular Localization Output	SignalP/ SecretomeP
0.566	Rv3748	Uncharacterized protein	transport protein	65	18	Unknown	Secretory	Cytoplasm	Secreted	Unknown
0.565	Rv3678a	Uncharacterized protein	unknown	38	19	Unknown	Secretory	Secreted extracellular	Secreted	Unknown
0.565	Rv2619c	Conserved protein	double-stranded beta helix	100	83	Cytoplasmic	Membrane	Cytoplasm	Cytoplasmic	Non-classical pathway
0.563	Rv3847	Uncharacterized protein	DNA-binding protein	36	16	Unknown	Membrane	Secreted extracellular	Membrane Bound	Non-classical pathway
0.557	Rv1734c	Uncharacterized protein Rv1734c	acyl transferase	100	99	Unknown	Cytoplasmic	Secreted extracellular	Cytoplasmic	Unknown
0.557	Rv0699	Uncharacterized protein	translation elongator protein	7	8	Extracellular	Secretory	Cytoplasm	Secreted	Lipoprotein Signal peptide
0.556	Rv3182	Putative toxin HigB3	toxin	100	82	Unknown	Cytoplasmic	Cytoplasm	Cytoplasmic	Unknown
0.555	Rv1669	Uncharacterized protein	DNA binding	46	28	Unknown	Cytoplasmic	Secreted extracellular	Secreted	TAT Signal peptide
0.555	Rv1632c	DUF402 domain-containing protein	hydrolase	100	78	Cytoplasmic	Cytoplasmic	Cytoplasm	Cytoplasmic	Unknown
0.555	Rv1914c	Uncharacterized protein	transport protein	48	21	Cytoplasmic membrane	Secretory	Secreted extracellular	Secreted	Unknown
0.552	Rv0807	SCP_3 domain-containing protein	transport protein	90	48	Cytoplasmic	Cytoplasmic	Cytoplasm	Cytoplasmic	Unknown
0.552	Rv2598	Uncharacterized protein	transferase	43	5	Cytoplasmic	Cytoplasmic	Plasma membrane	Cytoplasmic	Non-classical pathway
0.551	Rv2216	Epimerase family protein	hydrolase	100	98	Cytoplasmic membrane	Cytoplasmic	Cytoplasm	Cytoplasmic	Signal peptide
0.550	Rv2728c	Conserved alanine rich protein	oxidoreductase	100	99	Cytoplasmic membrane	Secretory	Cytoplasm	Cytoplasmic	Non-classical pathway
0.549	Rv3212	Conserved alanine valine rich protein	oxidoreductase	100	86	Unknown	Cytoplasmic	fimbrium	Cytoplasmic	Lipoprotein Signal peptide
0.549	Rv1083	Uncharacterized protein	membrane protein	78	19	Unknown	Secretory	Inner membrane	Membrane Bound	Non-classical pathway
0.549	Rv2774c	Uncharacterized protein	oxidoreductase	19	10	Cytoplasmic	Membrane	Cytoplasm	Cytoplasmic	Non-classical pathway
0.547	Rv1728c	Uncharacterized protein	hydrolase	95	36	Unknown	Membrane	Secreted extracellular	Secreted	Non-classical pathway
0.545	Rv0799c	ferredoxin	ferredoxin like	100	91	Cytoplasmic	Membrane	Secreted extracellular	Secreted	Non-classical pathway
0.544	Rv1499	Uncharacterized protein	hydrolase	25	11	Unknown	Cytoplasmic	Secreted extracellular	Secreted	Unknown
P_{ad} Score	Gene Loci	Uniprot/ Pfam Annotation	Phyre2 alignment	phyre2 confidence	phyre2 coverage	PsortB Analysis	Mycosub Analysis	Loctree3 Analysis	Subcellular Localization Output	SignalP/ SecretomeP
0.544	Rv1154	Uncharacterized protein	hydrolase	82	15	Cytoplasmic	Membrane	Cytoplasm	Cytoplasmic	Unknown

0.543	Rv2798c	Uncharacterized protein	cell invasion	96	75	Unknown	Secretory	Secreted extracellular	Secreted	Unknown
0.541	Rv3005c	Uncharacterized protein	transport protein	49	31	Cytoplasmic membrane	Membrane	Inner membrane	Membrane Bound	Non-classical pathway
0.540	Rv0965c	Uncharacterized protein	cell invasion	87	55	Unknown	Cytoplasmic	Cytoplasm	Cytoplasmic	Unknown
0.539	Rv2387	Conserved protein	transport protein	100	78	Cytoplasmic membrane	Membrane	Inner membrane	Membrane Bound	Unknown
0.539	Rv1078	Proline-rich antigen homolog	cell adhesin	23	17	Cytoplasmic membrane	Secretory	Secreted extracellular	Secreted	Non-classical pathway
0.539	Rv3114	CMP/dCMP-type deaminase domain-containing protein	hydrolase	100	88	Unknown	Cytoplasmic	Cytoplasm	Cytoplasmic	Non-classical pathway
0.539	Rv0877	Uncharacterized protein	transferase	58	21	Cytoplasmic	Membrane	Cytoplasm	Cytoplasmic	Non-classical pathway
0.538	Rv3333c	Hypothetical proline rich protein	synthetase	58	40	Unknown	Cytoplasmic	Secreted extracellular	Secreted	Signal peptide
0.538	Rv1889c	Uncharacterized protein	transferase	100	95	Unknown	Membrane	Cytoplasm	Membrane Bound	Lipoprotein Signal peptide
0.538	Rv1117	Conserved protein	oxidoreductase	100	86	Unknown	Cytoplasmic	Cytoplasm	Cytoplasmic	Unknown
0.536	Rv3831	Uncharacterized protein	gene regulation	45	21	Cytoplasmic membrane	Membrane	Inner membrane	Membrane Bound	Unknown
0.536	Rv3656c	Type IV pili (flp)-adhesin binds to macrophage and epithelial cells	cell cycle	58	34	Cytoplasmic membrane	Secretory	Secreted extracellular	Secreted	Signal peptide
0.535	Rv1950c	Uncharacterized protein	cell invasion	91	73	Unknown	Secretory	Secreted extracellular	Secreted	Unknown
0.534	Rv1762c	Uncharacterized protein	membrane protein	96	34	Cytoplasmic	Cytoplasmic	Cytoplasm	Cytoplasmic	Unknown
0.534	Rv2771c	Uncharacterized protein	flavodoxin oxidoreductase	100	97	Cytoplasmic	Cytoplasmic	Cytoplasm	Cytoplasmic	Non-classical pathway
0.534	Rv3627	Carboxypeptidase	beta-lactamase	100	80	Unknown	Cytoplasmic	Secreted extracellular	Cytoplasmic	Unknown
0.533	Rv2414c	Competence domain-containing protein	transferase	58	8	Cytoplasmic membrane	Membrane	Inner membrane	Membrane Bound	Unknown
0.530	Rv0464c	Carboxymuconolactone decarboxylase	lyase	95	100	Unknown	Secretory	Inner membrane	Secreted	Non-classical pathway
0.526	Rv0762c	Uncharacterized protein	isomerase	100	67	Cytoplasmic	Cytoplasmic	Cytoplasm	Membrane Bound	Unknown
0.526	Rv2738c	Uncharacterized protein	peptide binding protein	51	38	Cytoplasmic	Cytoplasmic	Cytoplasm	Cytoplasmic	Unknown
0.526	Rv2821c	CRISPR system Cms endoribonuclease Csm3	RNA-binding protein	100	90	Cytoplasmic	Cytoplasmic	Cytoplasm	Cytoplasmic	Non-classical pathway
0.526	Rv3098c	Uncharacterized protein	DNA-binding protein	9	8	Unknown	Secretory	Secreted extracellular	Secreted	Unknown
0.525	Rv3678c	Conserved protein	sugar-binding protein	100	100	Unknown	Cytoplasmic	Cytoplasm	Cytoplasmic	Signal peptide

P_{ad} Score	Gene Loci	Uniprot/ Pfam Annotation	Phyre2 alignment	phyre2 confidence	phyre2 coverage	PsortB Analysis	Mycosub Analysis	Loctree3 Analysis	Subcellular Localization Output	SignalP/ SecretomeP
0.525	Rv0185	Uncharacterized protein	DNA-binding protein	93	36	Unknown	Secretory	Cytoplasm	Secreted	Non-classical pathway
0.524	Rv3686c	Uncharacterized protein	membrane protein	35	14	Unknown	Membrane	Cytoplasm	Membrane Bound	Non-classical pathway
0.523	Rv0299	Toxin	lyase	58	38	Unknown	Secretory	Secreted extracellular	Secreted	Non-classical pathway
0.522	Rv2312	Uncharacterized protein	DNA-binding	32	18	Unknown	Secretory	Cytoplasm	Secreted	Non-classical pathway
0.521	Rv3440c	Uncharacterized protein	cell invasion	95	82	Unknown	Cytoplasmic	Secreted extracellular	Cytoplasmic	Non-classical pathway
0.520	Rv1489	Conserved protein	membrane protein electron transport	26	32	Cytoplasmic membrane	Cytoplasmic	Inner membrane	Membrane Bound	unknown
0.518	Rv1716	Uncharacterized protein	hydrolase	100	80	Cytoplasmic	Cytoplasmic	Cytoplasm	Cytoplasmic	Non-classical pathway
0.518	Rv0057	DUF5650 unknown	transferase	26	22	Unknown	Cytoplasmic	Secreted extracellular	Secreted	TAT Signal peptide
0.517	Rv3900c	Conserved hypothetical alanine-rich protein	biosynthetic protein	32	25	Cytoplasmic	Cytoplasmic	Secreted extracellular	Cytoplasmic	Non-classical pathway
0.517	Rv2517c	Uncharacterized protein	hydrolase	23	43	Unknown	Secretory	Secreted extracellular	Secreted	Unknown
0.516	Rv0333	SnoaL-like domain-containing protein	unknown	97	80	Unknown	Secretory	Cytoplasm	Cytoplasmic	Unknown
0.516	Rv0039	Duf308 unknown	apoptosis	21	16	Cytoplasmic membrane	Cytoplasmic	Inner membrane	Membrane Bound	Signal peptide
0.514	Rv1706A	Uncharacterized protein	oxidoreductase	11	7	Cytoplasmic membrane	Secretory	Inner membrane	Secreted	Unknown
0.513	Rv0616c	Uncharacterized protein	transferase	9	10	Cytoplasmic membrane	Membrane	fimbrium	Membrane Bound	Non-classical pathway
0.511	Rv2797c	Uncharacterized protein	alpha/beta hydrolase	98	19	Unknown	Cytoplasmic	Cytoplasm	Cytoplasmic	Non-classical pathway
0.511	Rv0787	Uncharacterized protein	transport protein	95	46	Extracellular	Cytoplasmic	Secreted extracellular	Cytoplasmic	Non-classical pathway

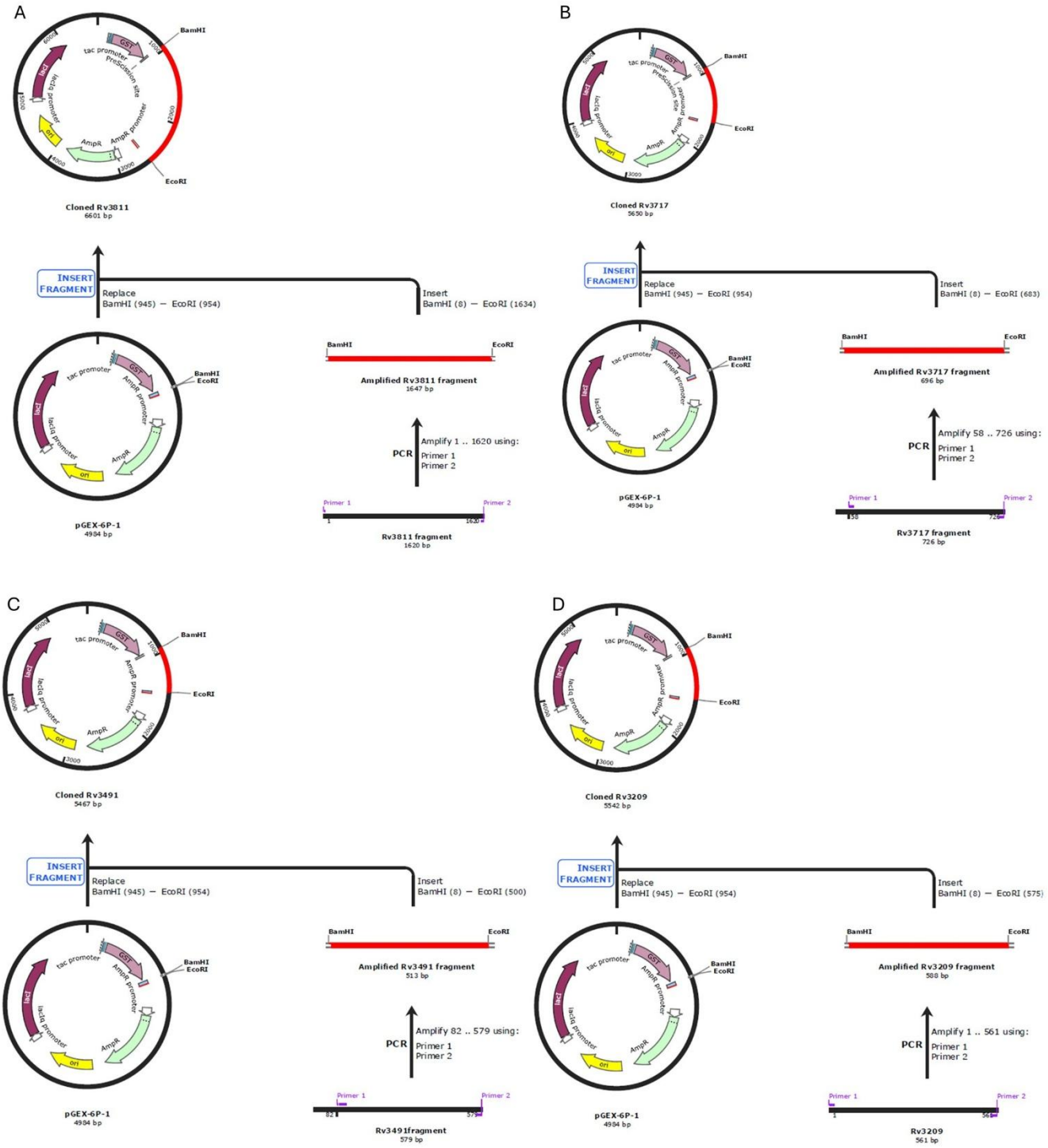
Supplementary table 3: List of predicted adhesin and adhesin-like genes from the *Mycobacterium tuberculosis* H37Rv genome using SPAAN analysis (N=776)**.**

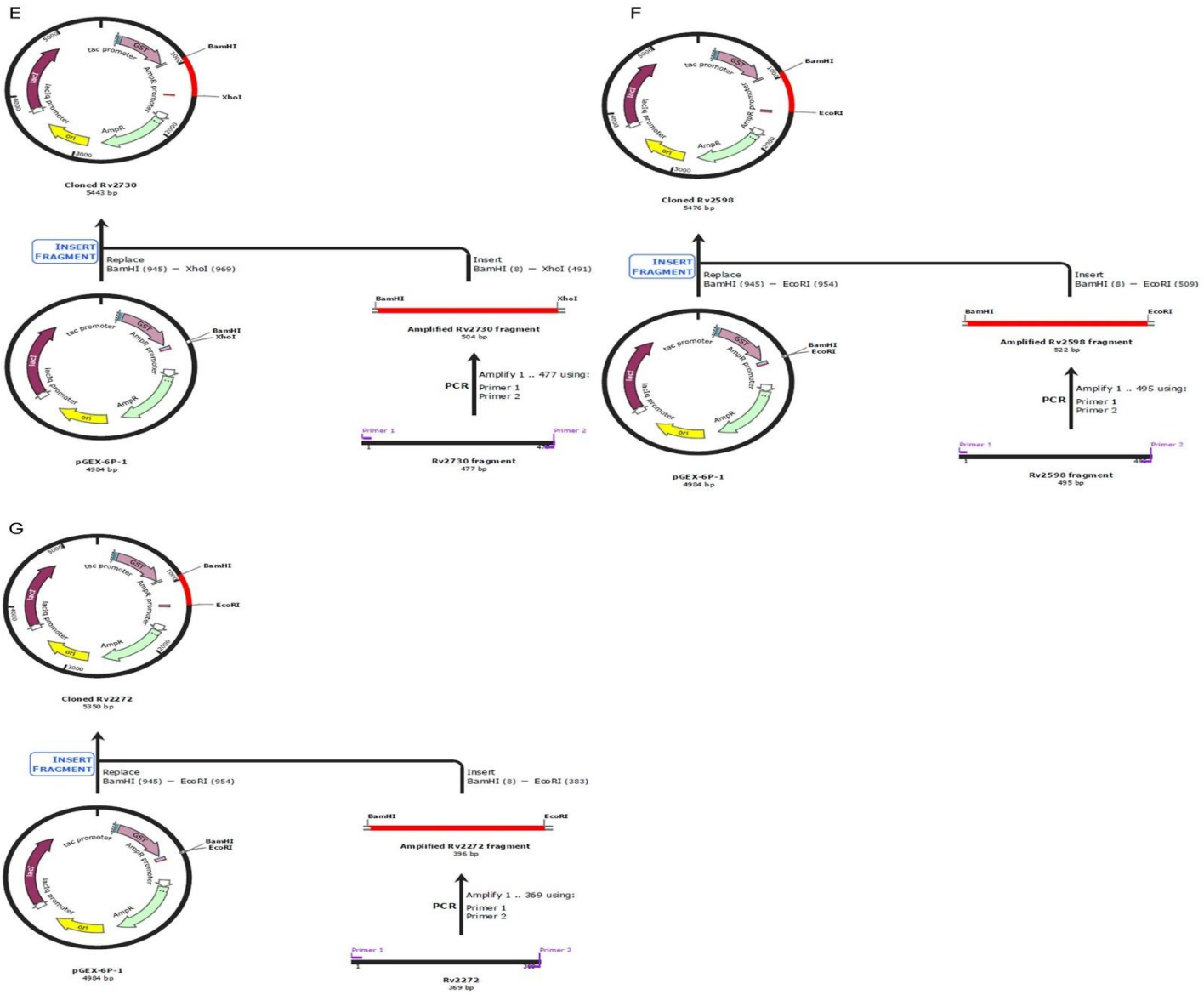
Functional Category (Mycobrowser)	Potential Adhesin gene			Total
	Higher P_{ad} Index	Moderate P_{ad} Index	Lower P_{ad} Index	
PE/PPE	*Rv0109, Rv0124, Rv0159c, Rv0160c, Rv0279c, Rv0304c, Rv0354c, Rv0355c, Rv0442c, Rv0532, Rv0578c, Rv0742, Rv0746, Rv0747, Rv0755c, Rv0832, Rv0833, Rv0834c, Rv0872c, Rv0977, Rv0978c, Rv1040c, Rv1067c, Rv1068c, Rv1087, Rv1089, Rv1091, Rv1135c, Rv1172c, Rv1195, Rv1196, Rv1243c, Rv1325c, Rv1361c, Rv1396c, Rv1441c, Rv1450c, Rv1452c, Rv1468c, Rv1548c, Rv1646, Rv1651c, Rv1706c(PPE23), Rv1705c, Rv1753c, Rv1759c (Wag22), Rv1768, Rv1788, Rv1789, Rv1803c, Rv1807, Rv1818c (Pe-PGRS33), Rv1840c, Rv1917c, Rv1918c, Rv2162c, Rv2328, Rv2340c, Rv2352c, Rv2353c, Rv2356c, Rv2396, Rv2487c, Rv2490c, Rv2519, Rv2591, Rv2615c, Rv3097c, Rv3136, Rv3159c, Rv2634c, Rv2741, Rv2768c, Rv2769c, Rv2770c,Rv2853, Rv3343c, Rv3345c, Rv3347c, Rv3350c, Rv3367, Rv3388, Rv3477, Rv3478, Rv3507, Rv3508, Rv3512, Rv3512, Rv3514, Rv3514, Rv3532, Rv3533c, Rv3558, Rv3590c, Rv3595c, Rv3621c, Rv3622c, Rv3653, Rv3812*	*Rv0096, Rv0151c, Rv0256c, Rv0278c, Rv0285, Rv0305c,* Rv*0388c, Rv0754, Rv0878c, Rv0915c, Rv0916c, Rv0980c, Rv1039c, Rv1168c, Rv1386, Rv1430,* R*v1790, Rv1801, Rv1802, Rv1806, Rv1808, Rv1809, Rv2107, Rv2108, Rv2123, Rv2408, Rv2892c, Rv3022a, Rv3125c, Rv3135, Rv3144c, Rv3344c, Rv3429, Rv3539, Rv3739c, Rv3746c, Rv3872, Rv3873, Rv3892c, Rv3893c*	*Rv0285,* R*v0453, Rv1088, Rv1169c, Rv1214c, Rv1387, Rv1706c (socB), Rv1787, Rv2431c, Rv2608, Rv3021c, Rv3022c, Rv3652, Rv3738c*	**153**
Cell wall and Cell wall processes	*Rv0287, Rv0309, Rv0344c, Rv0774c, Rv0928, Rv0932c, Rv0934 (PstS1), Rv0954, Rv1115, Rv1217c, Rv1269c, Rv1270c, Rv1639c, Rv1690, Rv1881c, Rv1980c(Mpt64), Rv2253, Rv2341, Rv2376c, Rv2389c, Rv2875, Rv3020c, Rv3036c, Rv3576, Rv3810, Rv3865, Rv3875(ESAT-6), Rv3878, Rv3890c, Rv3891c, Rv3904c*	*Rv0009, Rv0011c, Rv0040c, Rv0054, Rv0130, Rv0178, Rv0179c, Rv0229a, Rv0288, Rv0341, Rv0398c, Rv0418 (LpqL), Rv0426c, Rv0431, Rv0461, Rv0475 (HBHA), Rv0483, Rv0538, Rv0545c, Rv0559c, Rv0583c, Rv0584, Rv0603, Rv0838, Rv0888, Rv0962c, Rv0988, Rv1004, Rv1029, Rv1037c, Rv1097c, Rv1174c, Rv1184c, Rv1198, Rv1228, Rv1230, Rv1252c, Rv1275, Rv1291c, Rv1362c, Rv1368, Rv1411c, Rv1418, Rv1541c, Rv1672c, Rv1672c, Rv1758, Rv1793, Rv1799, Rv1857, Rv1861, Rv1884c, Rv1910, Rv1911c, Rv1921c, Rv1926c, Rv1970, Rv1973, Rv1984c, Rv1987, Rv2046, Rv2077c, Rv2081, Rv2091, Rv2116, Rv2138, Rv2198c, Rv2270, Rv2272, Rv2273, Rv2290, Rv2301, Rv2318, Rv2346c, Rv2400c, Rv2403c, Rv2450c, Rv2518c, Rv2544, Rv2576, Rv2623, Rv2668,*	*Rv0016c, Rv0116c, Rv0180, Rv0188, Rv0200, Rv0203, Rv0205, Rv0227, Rv0236c, Rv0237, Rv0261c, Rv0314c, Rv0320, Rv0399c, Rv0403c, Rv0451c, Rv0463, Rv0477, Rv0479, Rv0482, Rv0513, Rv0544, Rv0604, Rv0622, Rv0638, Rv0658c, Rv0677c, Rv0680c, Rv0847, Rv0875, Rv0901, Rv0924c, Rv0961, Rv1016c, Rv1018c, Rv1022, Rv1024, Rv1064c, Rv1200, Rv1224, Rv1303, Rv1433, Rv1476, Rv1481, Rv1488, Rv1508a, Rv1510, Rv1607, Rv1616, Rv1672c, Rv1698, Rv1730, Rv1744c, Rv1795, Rv1811, Rv1845c, Rv1860 (Apa), Rv1888c, Rv1922, Rv2145c, Rv2156c, Rv2164c, Rv2203, Rv2223c, Rv2254, Rv2281, Rv2347c, Rv2507, Rv2543, Rv2560, Rv2563, Rv2597, Rv2686c, Rv2700, Rv2721, Rv2790c,*	**245**

		Rv2784c, Rv2796c, Rv2799, Rv2843, Rv2873, Rv2878, Rv2911, Rv2945c, Rv3016, Rv3017c, Rv3019c, Rv3065,	*Rv2806, Rv2846c, Rv2920c, Rv2938, Rv2972c, Rv3006, Rv3070, Rv3092c, Rv3330, Rv3335c, Rv3390,*	
Functional Category (Mycobrowser)	**Potential Adhesin gene**			**Total**
	Higher P_{ad} Index	**Moderate P_{ad} Index**	**Lower P_{ad} Index**	
		Rv3194c, Rv3289c, Rv3298c, Rv3310, Rv3312a (MTP), Rv3451, Rv3452, Rv3578c, Rv3584, Rv3613, Rv3616c, Rv3619c, Rv3682, Rv3724, Rv3763 (LpqH), Rv3843c, Rv3857, Rv3874, Rv3880c	*Rv3395a, Rv3448, Rv3454, Rv3479, Rv3593, Rv3612, Rv3657c, Rv3666c, Rv3695, Rv3732, Rv3806c, Rv383, Rv3864, Rv3901, Rv3905c*	
Regulatory Proteins	*None*	*Rv0020c, Rv1566c, Rv3840*	*Rv0002, Rv0031, Rv0033, Rv0062, Rv0072, Rv0086, Rv0113, Rv0128, Rv0148, Rv0173, Rv0377, Rv0931c (PknD), Rv3193, Rv3744, Rv3855*	**18**
Lipid metabolism	*Rv0129c(Ag85c), Rv3803c(Mpt51), Rv3804c (Ag85a)*	*Rv1679, Rv1714, Rv1886c(Ag85b), Rv2500c, Rv3233c, Rv3281, Rv3374, Rv3540c, Rv3668c, Rv3671*	*Rv0242c, Rv0533c, Rv0972c, Rv1070c, Rv1135a, Rv1323, Rv1483, Rv1627, Rv2002, Rv2262c, Rv2881c, Rv3802c*	**25**
Intermediatory metabolism/ respiratory	*None*	*Rv0063, Rv0125, Rv0137c, Rv0363c (Fba), Rv0374c, Rv0462 (LpdC), Rv0751c, Rv0814, Rv1090, Rv1295, Rv1305, Rv1385, Rv1454c, Rv1470, Rv1471, Rv1623c, Rv1856, Rv1885, Rv2043c, Rv2199c, Rv2277, Rv2322c, Rv2351c, Rv2421c, Rv2702, Rv3118, Rv3198a, Rv3352c, Rv3500c*	*Rv0132c, Rv0224, Rv0249c, Rv0291, Rv0315, Rv0467 (IcL), Rv0478, Rv0500, Rv0524, Rv0534c, Rv0635, Rv0636, Rv0663, Rv0711, Rv0773c, Rv1240, Rv1245c, Rv1296, Rv1306, Rv1333, Rv1383, Rv1451, Rv1554, Rv1570, Rv1604, Rv1657, Rv1817, Rv2007c, Rv2193, Rv2194, Rv2202c, Rv2349c, Rv2350c, Rv2552c, Rv2672, Rv2674, Rv2726c, Rv2775, Rv2780, Rv2970c, Rv3086, Rv3154, Rv3175, Rv3214, Rv3247c, Rv3449, Rv3502c, Rv3567c, Rv3581c, Rv3787c, Rv3914*	**80**
Conserved hypothetical proteins	*Rv0061c, Rv0387c, Rv0991c, Rv1116a, Rv1810, Rv1906c, Rv2077a, Rv2575, Rv2599, Rv3207c, Rv3209, Rv3337, Rv3572, Rv3717*	*Rv0340, Rv0378, Rv0381c, Rv0397a, Rv0454,Rv0455c, Rv0471c, Rv0577, Rv0614, Rv0679c, Rv0712, Rv0963c, Rv1057, Rv1100, Rv1109, Rv1134, Rv1137c, Rv1158c, Rv1203c, Rv1157c, Rv1268c, Rv1754c, Rv1804, Rv1815, Rv1875c, Rv1891, Rv1951c, Rv1975, Rv1998c,Rv2078, Rv2082, Rv2114, Rv2240c, Rv2257c, Rv2264c, Rv2283, Rv2292c, Rv2302, Rv2307c, Rv2422, Rv2525c, Rv2641, Rv2660c, Rv2706c, Rv2730, Rv2744c, Rv2820c, Rv2960c, Rv2929, Rv2998,Rv3033, Rv3035, Rv3067,Rv3103c, Rv3136a, Rv3090, Rv3267, Rv3353c, Rv3354, Rv3369, Rv3491, Rv3627c,Rv3706c, Rv3707c, Rv3718c, Rv3747, Rv3811, Rv3822, Rv3897c, Rv3898c*	*Rv0021c, Rv0039c, Rv0057, Rv0185, Rv0274, Rv0299, Rv0333, Rv0464c,Rv0500a, Rv0616c, Rv0699, Rv0725c, Rv0762c, Rv0787,Rv0799c, Rv0807,Rv0877, Rv0911, Rv0965c, Rv0968c, Rv0968, Rv0979c, Rv1078, Rv1083, Rv1117, Rv1669, Rv1706a, Rv1716, Rv2797c, Rv1154c, Rv1419, Rv1489, Rv1499, Rv1519, Rv1632c, Rv1728c, Rv1734, Rv1762c, Rv1889, Rv1914c, Rv1950c, Rv1961, Rv2083, Rv2216, Rv2229c, Rv2256c, Rv2312, Rv2387,Rv2414c, Rv2517c, Rv2577, Rv2598, Rv2619c, Rv2728c, Rv2738c, Rv2771c, Rv2774c, Rv2798c, Rv2821c, Rv3005c, Rv3098c,Rv3114, Rv3182, Rv3212, Rv3282, Rv3333c, Rv3440c,Rv3484c, Rv3541c, Rv3655c, Rv3656c (type IV pili), Rv3678a Rv3678c,Rv3683, Rv3686c, Rv3705a, Rv3705c, Rv3716c, Rv3745c,Rv3748, Rv3796, Rv3831, Rv3847, Rv3900c,*	**167**
Information pathways	*None*	*Rv0444c, Rv0639, Rv0640, Rv2534c, Rv2582, Rv2890c, Rv2986c (Hup), Rv3053c*	*Rv0429c, Rv0641, Rv0651, Rv0702, Rv0703, Rv0706, Rv1080c, Rv1107c, Rv1340, Rv1653, Rv2055c, Rv2441c, Rv2442c, Rv2593c, Rv2594c, Rv2755c, Rv2909c, Rv3366, Rv3912*	**27**
Insertion sequences	*None*	*Rv1575, Rv1576c, Rv1581, Rv2650c*	*Rv0094c, Rv0397, Rv1579c, Rv2087, Rv2087, Rv2182c, Rv2654c, Rv2666, Rv2810c*	**13**

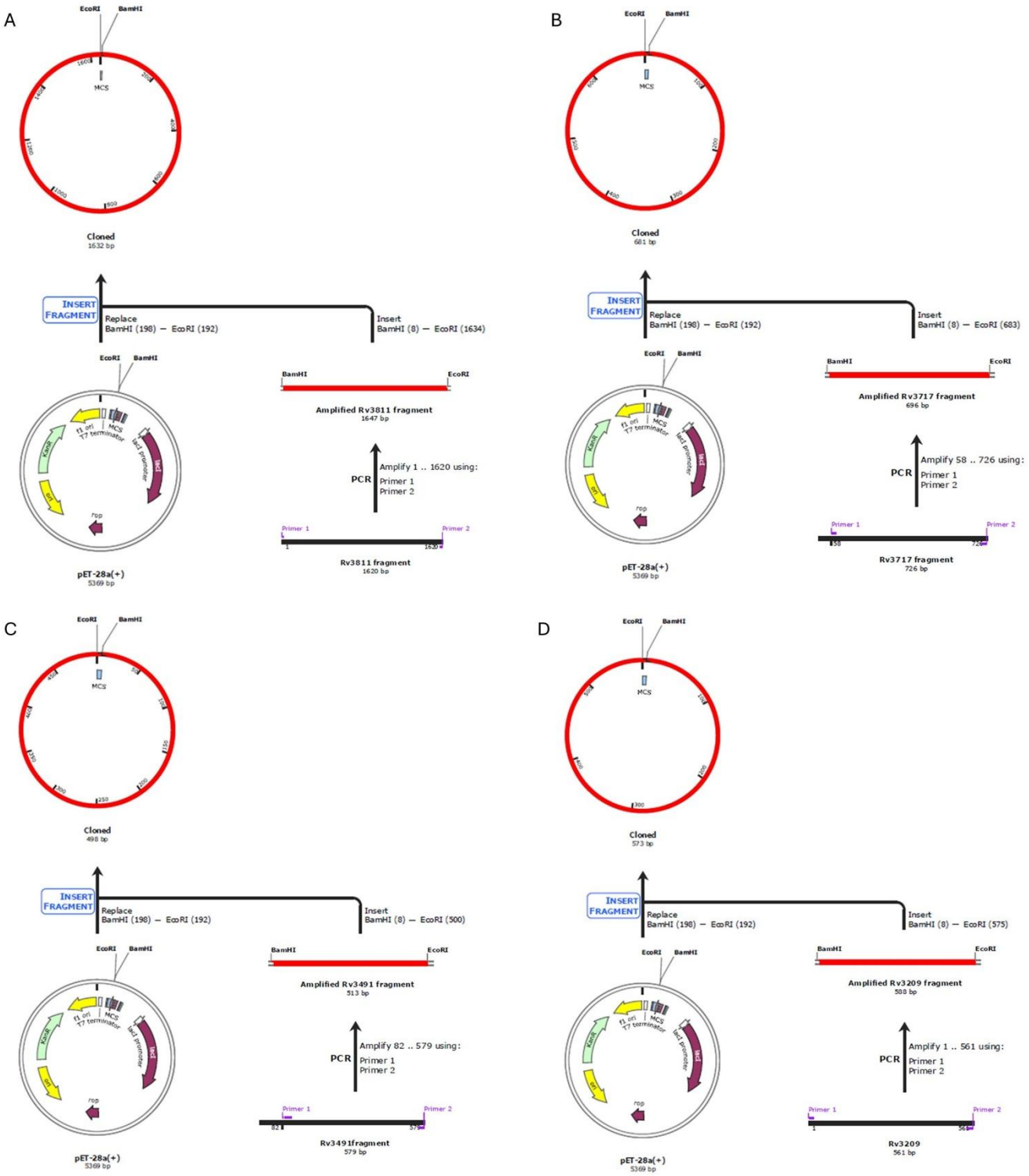
Virulence/ Detoxification/ Adaptation	*None*	*Rv0168, Rv0169, Rv0171, Rv0174, Rv0432, Rv0589, Rv0590, Rv0591, Rv0592, Rv0594, Rv0619 (Mce-1), Rv1085c, Rv1478, Rv1495, Rv1964, Rv1965, Rv1966, Rv1982c, Rv1983, Rv2190c, Rv2238c, Rv2429, Rv2758c, Rv3180c, Rv3494c, Rv3499c, Rv3639, Rv3648c, Rv3846*	*Rv0170, Rv0172, Rv0554, Rv0587, Rv0588, Rv1102c, Rv1932, Rv1953, Rv1968, Rv1969, Rv2209, Rv2231a, Rv2602, Rv2830c, Rv3177, Rv3269,* Rv*3496c, Rv3498c, Rv3749c*	**48**

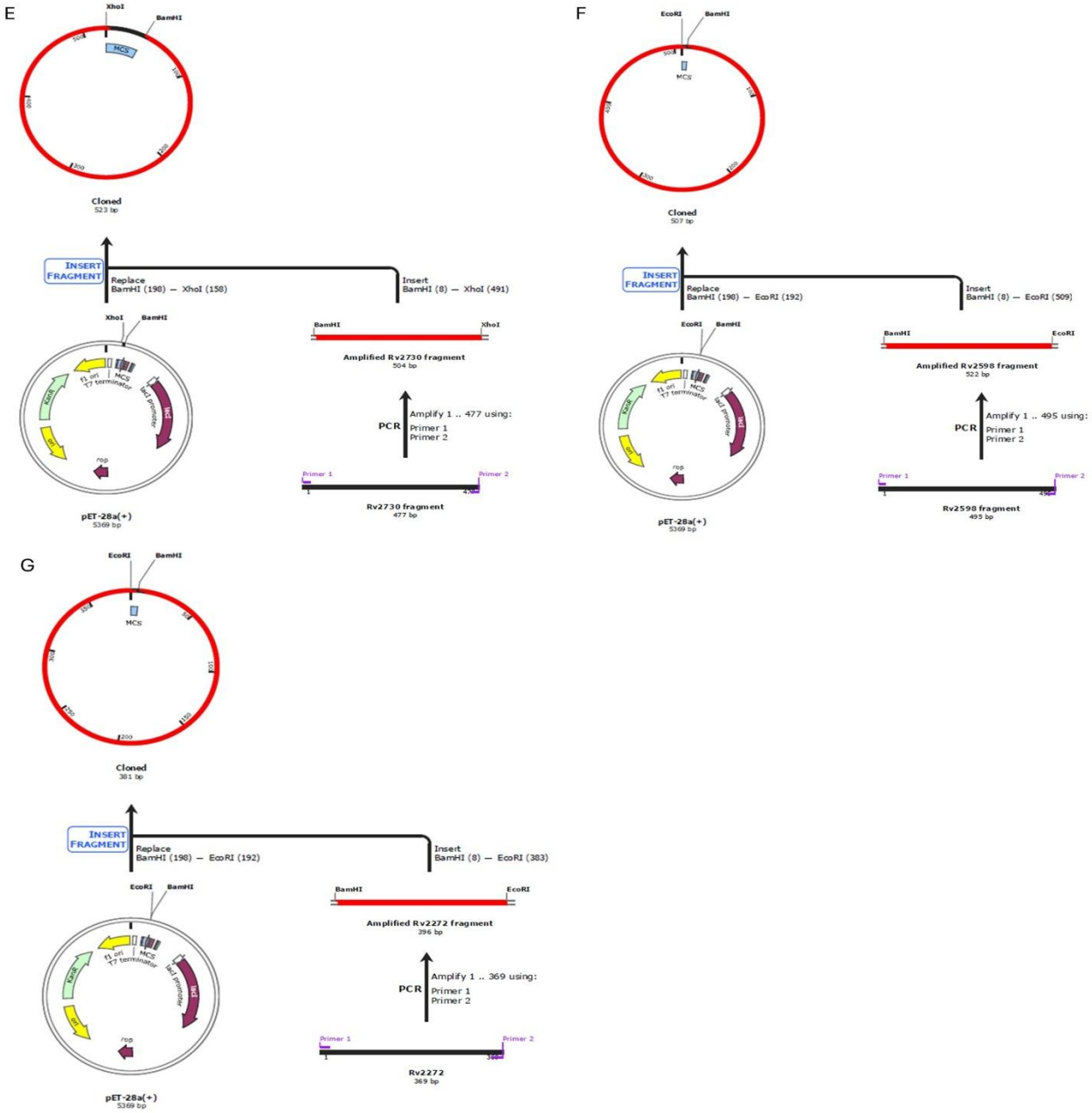
B2. Chapter four supplementary data



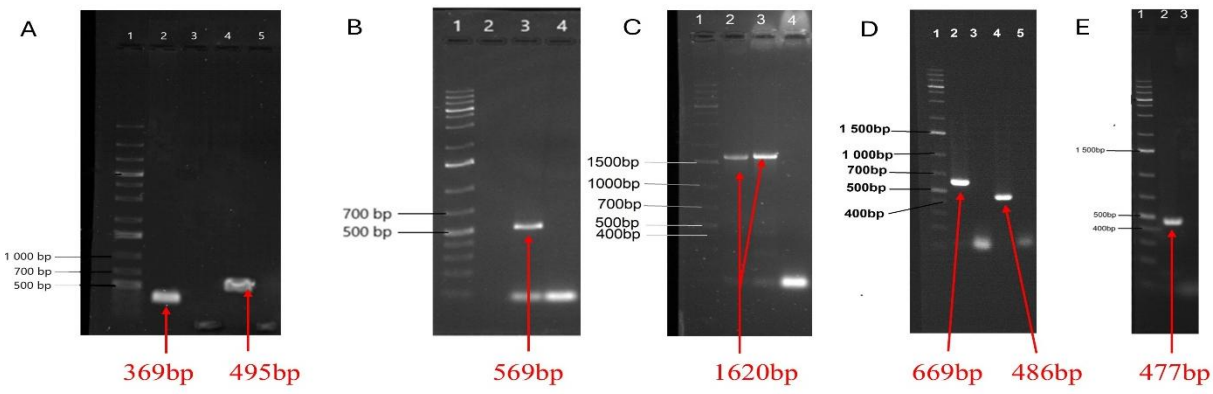


Supplementary Figure 1. in Silico cloning of the *Mycobacterium tuberculosis* conserved hypothetical gene insert into the plasmid vector for pGEX-6P-1 (4984bp) using SnapGene (www.snapgene.com). In the pGEX-6P-1 vector the origin of replication (Ori) resulting in a high copy plasmid. The GST affinity tag (25.5 kDa) is bound to the cleavable PreScission site (0.9 kDa), containing and the multiple cloning site (MCS) with restriction sites for insertion of the restricted amplified PCR product to be expressed by inducing the *lacI* operon and *Tac* promoter. The plasmids provide a selective marker, Ampicillin resistance (*AmpR*) gene for detecting transformants.

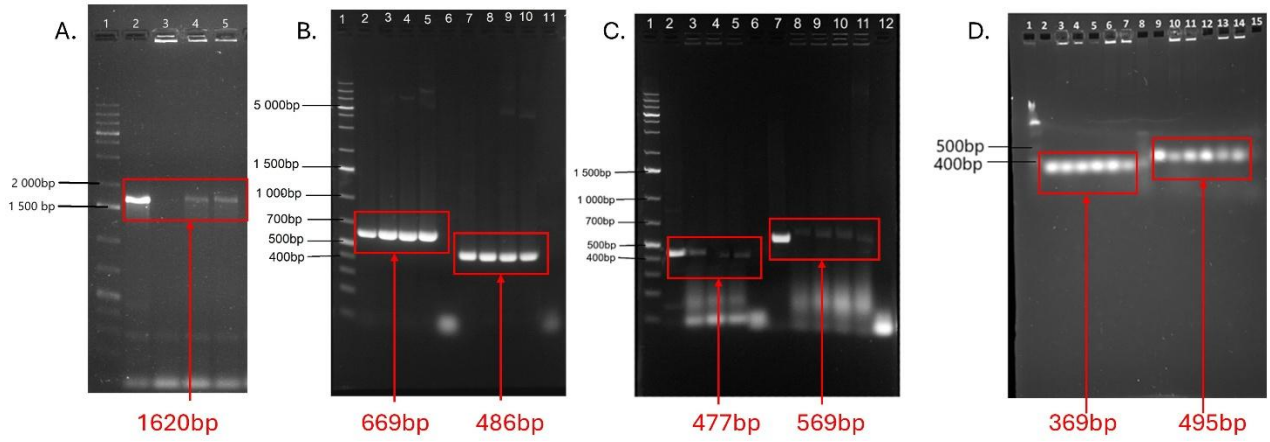

A
EcoRI
BamHI
MCS
Cloned
1632 bp
INSERT FRAGMENT
Replace
BamHI (198) – EcoRI (192)
Insert
BamHI (8) – EcoRI (1634)
Amplified Rv3811 fragment
1647 bp
PCR
Amplify 1 .. 1620 using:
Primer 1
Primer 2
Primer 1
Primer 2
Rv3811 fragment
1620 bp
pET-28a(+)
5369 bp
B
Cloned
681 bp
Replace
BamHI (198) – EcoRI (192)
Insert
BamHI (8) – EcoRI (683)
Amplified Rv3717 fragment
696 bp
Amplify 58 .. 726 using:
Rv3717 fragment
726 bp
C
Cloned
498 bp
Replace
BamHI (198) – EcoRI (192)
Insert
BamHI (8) – EcoRI (500)
Amplified Rv3491 fragment
513 bp
Amplify 82 .. 579 using:
Rv3491fragment
579 bp
D
Cloned
573 bp
Replace
BamHI (198) – EcoRI (192)
Insert
BamHI (8) – EcoRI (575)
Amplified Rv3209 fragment
588 bp
Amplify 1 .. 561 using:
Rv3209
561 bp



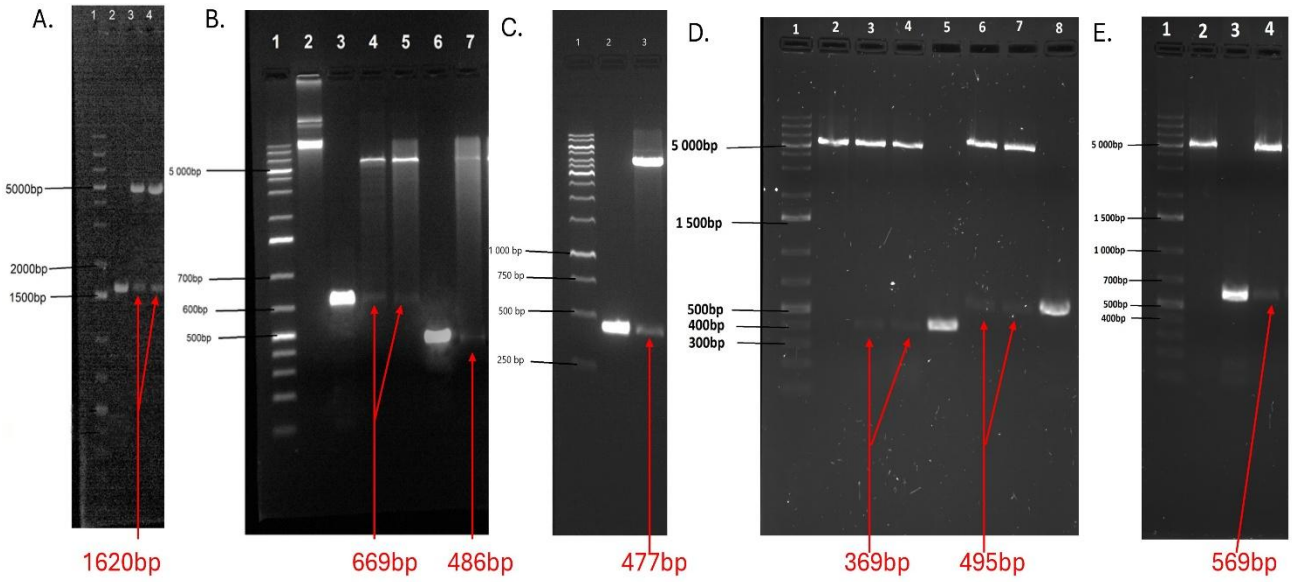
Supplementary Figure 2. in Silico cloning of the *Mycobacterium tuberculosis* conserved hypothetical gene insert into the plasmid vector for pET-28a (5369bp) using SnapGene (www.snapgene.com). In the pET-28a vector the origin of replication (Ori) resulting in a high copy plasmid, with the repressor of primer (Rop) protein controlling host bacterial plasmid copy numbers. The His-affinity tag (0.8 kDa) bound to the cleavage thrombin site (0.6 kDa), containing and the multiple cloning site (MCS) with restriction sites for insertion of the restricted amplified PCR product to be expressed by inducing the *lacI* operon with IPTG. The plasmids provide a selective marker, Kanamycin resistance (*KanR*) gene for detecting transformants.



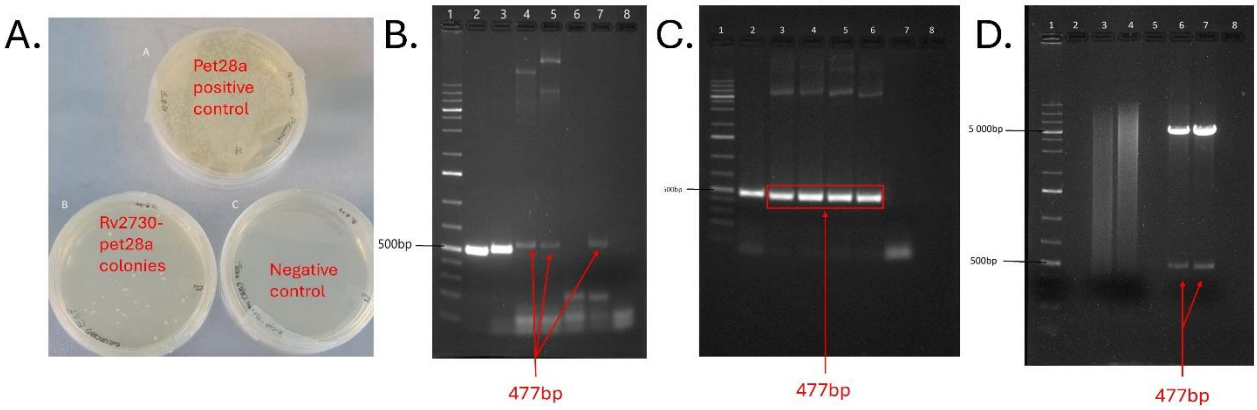
Supplementary Figure 3. Agarose gel (1.5%) of amplified *Mycobacterium tuberculosis* genes of interest performed by polymerase chain reaction. A. Lane 2: Rv2272 (369bp), Lane 3: Rv2272 Negative control , Lane 4: Rv2598 (495 bp), Lane 5: Rv2598 negative control Lane **B.** Lane 3: Rv3209 (569bp), Lane 4: Rv3209 negative control, **C.** Lane 2,3: Rv3811 (1620bp), Lane 4: Rv3811 negative control. **D.** Lane 2: Rv3717 (669 bp), Lane 3: Rv3717 negative control, Lane 4: Rv3491 (486 bp), Lane 5: Rv3491 negative control **E.** Lane 2: Rv2730 (477 bp), Lane 3: Rv2730 negative control. All gels lane 1: Gene ruler 1KB plus DNA ladder (Thermo Fisher Scientific)



Supplementary Figure 4: Gel electrophoresis (1.5%) of Colony PCR confirmation for the confirmation of recombinant pGEX-6P1plasmid in *E.coli* (Dh5a) transformed colonies shown . A. Lane 2: Rv3811 (1620 bp) positive control Lane 4,5: Rv3811 colony PCR **B.** Lane 2: Rv3717 (669 bp) positive control, Lane 3-5: Rv3717 colony PCR, Lane 6: negative control, Lane 7: Rv3491 (486 bp) positive control. Lane 8-10: Rv3491 colony PCR, Lane 11: negative control **C.** Lane 2: Rv2730 (477 bp) positive control, Lane 3-5: Rv2730 colony PCR, Lane 6: negative control, Lane 7: Rv3209 (569bp) positive control, Lane8-11: Rv3209 colony PCR, Lane 12: negative control **D.** Lane 2: Rv2272 (369 bp) positive control, Lane 3-7: Rv2272 colony PCR, Lane 8: negative control, Lane 9: Rv2598 (495bp) colony PCR Lane 10-14: Rv2598 colony PCR, Lane 15: Negative control. All gels lane 1: Gene ruler 1KB plus DNA ladder (Thermo Fisher Scientific)

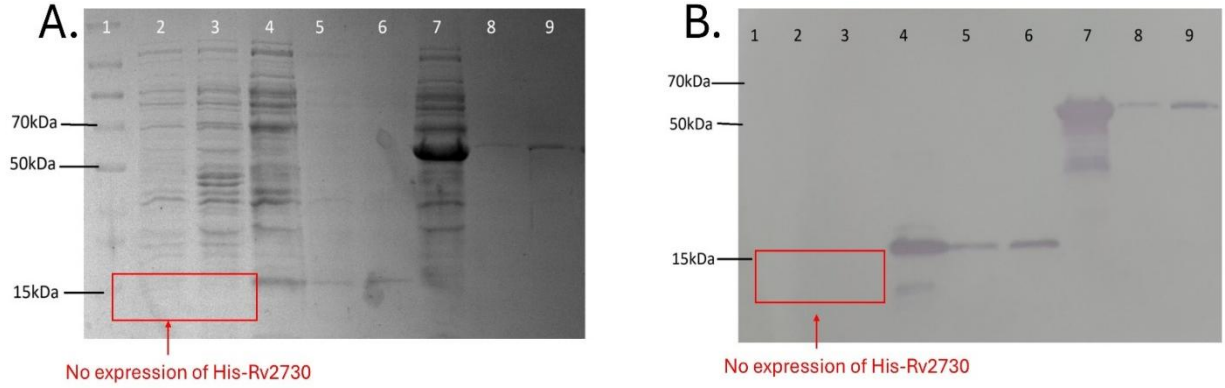


Supplementary Figure 5: Gel electrophoresis (1.5%) of Restriction profiling to confirm recombinant pGEX-6P-1. A. Lane 2: Rv3811 (1620 bp) Lane 3-4: Restricted Rv3811-Pgex-6p1 **B.** Lane 2: unrestricted pGEX-6p-1, Lane 3: Rv3717 (669 bp), Lane 4,5: Restricted Rv3717-pGEX-6P-1, Lane 6: Rv3491 (486 bp). Lane 7, : Restricted Rv3491-pGEX-6P-1 **C.** Lane 2: Rv2730 (477 bp), Lane 3: Restricted Rv2730-pGEX-6P-1 **D.** Lane 2: Pgex6p1 restricted plasmid Lane 3,4: Restricted Rv2272-Pgex6p1 Lane 5: Rv2272 (369 bp), Lane 6,7: Restricted Rv2598-Pgex6P1 Lane 8 : Rv2598 (495bp) **E.** Lane 3: Rv3209(569bp), Lane 4: Restricted Rv3209-pGEX-6P-1. All gels lane 1: Gene ruler 1KB plus DNA ladder (Thermo Fisher Scientific)

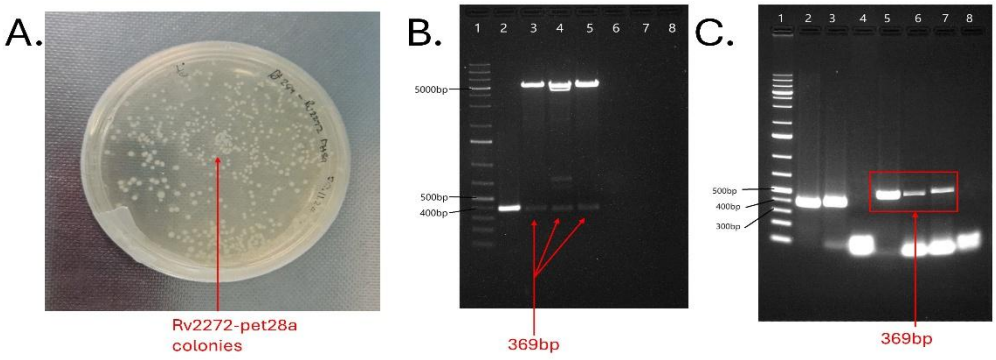


Supplementary Figure 6. Cloning confirmation of *Mycobacterium tuberculosis* conserved hypothetical protein pET-28a-Rv2730 A. Transformation of Rv2730 in PET-28a vector growing on LB-Kan agar plates. pET-28a plasmid positive control (top), pET-28a-Rv2730 transformed cells (Left), pET-28a-Rv2730 Negative control (Right), **B. Gel electrophoresis (1.5%) of Colony PCR confirmation pET-28a-Rv2730.** Lane 1: Gene ruler 1KB plus DNA ladder (Thermo Fisher Scientific), Lane 2: Rv2730 (477 bp) PCR positive control, Lane 3: pET-28a-Rv2730 colony 3 , Lane 4: pET-28a-Rv2730 colony 4, Lane 5: pET-28a-Rv2730 colony 5, lane 6: pET-28a-Rv2730 colony 6, Lane 7: PCR negative control **C.**

Gel electrophoresis (1.5%) of pET-28a plasmid vector PCR confirmation of Rv2730 (477bp), amplified PCR products. Lane 1: Gene ruler 1KB plus DNA ladder (Thermo Fisher Scientific), Lane 2: Rv2730 PCR positive control, Lane 3: Rv2730 colony 3, Lane 4: Rv2730 colony 4, Lane 5: Rv2730 colony 5, lane 6: Rv2730 colony 6,Lane 7: PCR negative control, Lane 8: blank **D. Gel electrophoresis (1.5%) of Restriction digestion Confirmation of transformed pET-28a-Rv2730- plasmid.** Lane 1: Gene ruler 1KB plus DNA ladder (Thermo Fisher Scientific), Lane2: Blank, Lane 3: pET-28a-Rv2730 DH5a colony 3, Lane 4: pET-28a-Rv2730 DH5a colony 4, Lane 5: Blank, Lane 6: pET-28a-Rv2730 BL21 colony 3, Lane 7: pET-28a-Rv2730 BL21 colony 4

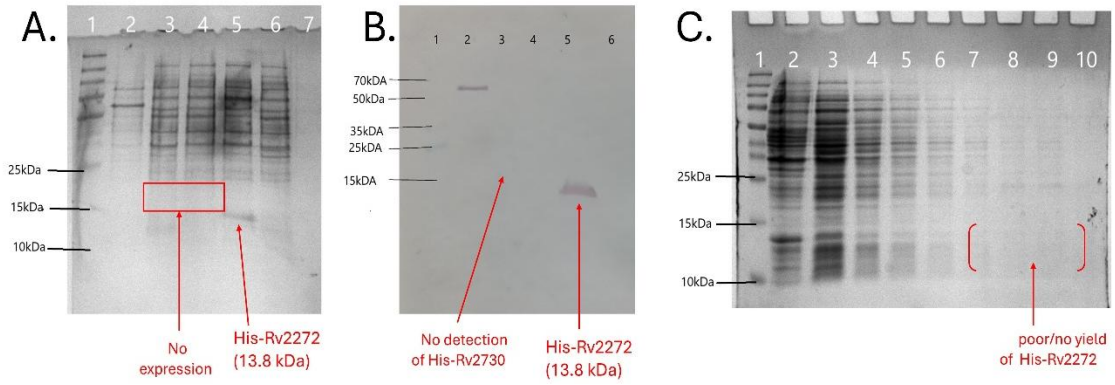


Supplementary Figure 7. Detection of purified His-affinity tag proteins using HisPur Cobalt resin Batch purification (SDS-PAGE gel 12% concentration).A. shown by SDS-PAGE and B. Western Blot Lane 1: Spectra multicolor Broad range protein ladder (Thermo Fisher Scientific), **Lane 2: Rv2730 uninduced, Lane 3: Rv2730 induced (at 37ºC),** Lane 4: Rv3491 clear lysate, Lane 5: Rv3491 purified protein, Lane 6: Rv3491 purified protein, Lane 7: Rv3811 clear lysate, Lane 8: Rv3811 purified protein, Lane 9: Rv3811 purified protein

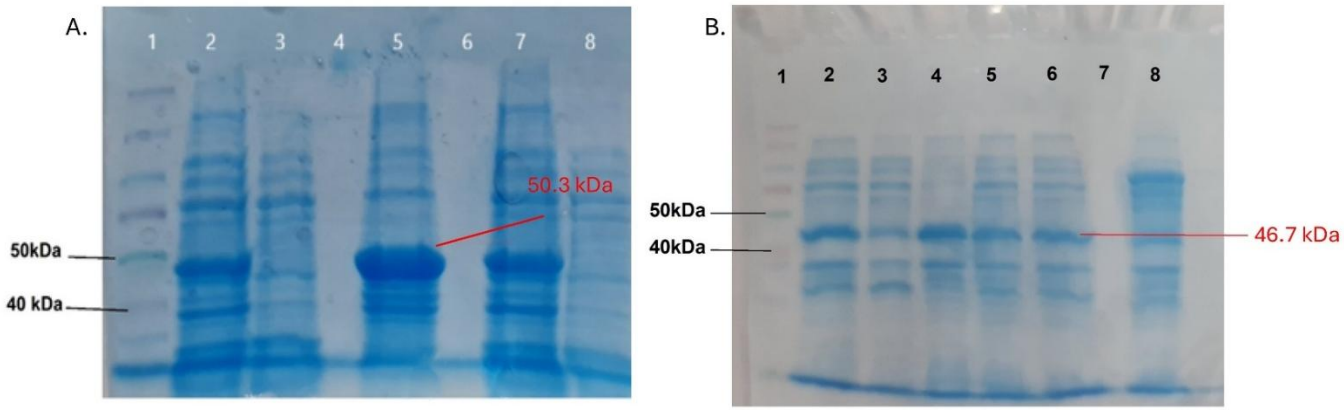


Supplementary Figure 8. Cloning confirmation of *Mycobacterium tuberculosis* conserved hypothetical protein pET-28a-Rv2272 A. Transformation of Rv2272 in pET-28a vector growing on LB-Kan agar plates. B. Gel electrophoresis (1.5%) of Restriction digestion confirmation on recRv2272. Lane 1: Gene ruler 1KB plus DNA ladder (Thermo Fisher Scientific), Lane 2: Rv2272 positive control, Lane 3: RecRv2272 (col 1), Lane 4: RecRv2272 (col 2), Lane 5: RecRv2272 (col 3), Lane 6-8: blank. **C. Gel electrophoresis (1.5%) of Plasmid PCR for the identification of gene Rv2272 (369bp) and Rv2730(477bp)** Lane 1: Gene ruler 1KB plus DNA ladder (Thermo Fisher Scientific), Lane 2: Rv2272

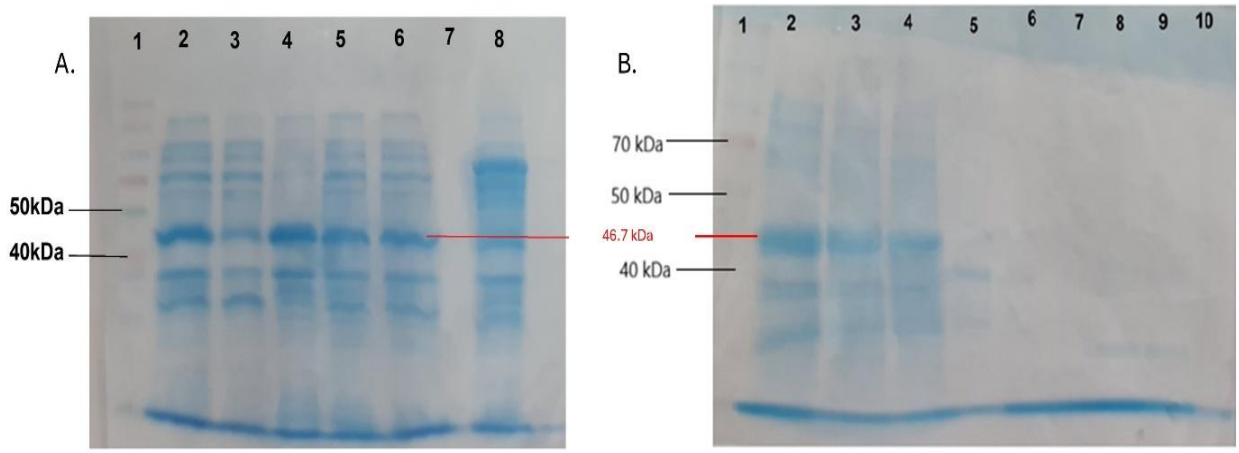
positive control, Lane 3: RecRv2272 (Colony 4), Lane 4: Rv2272 negative control Lane 5: Rv2730 positive control, Lane 6: RecRv2730 Colony 4, Lane 7: RecRv2730 Colony 5, Lane 8: Rv2730 negative control



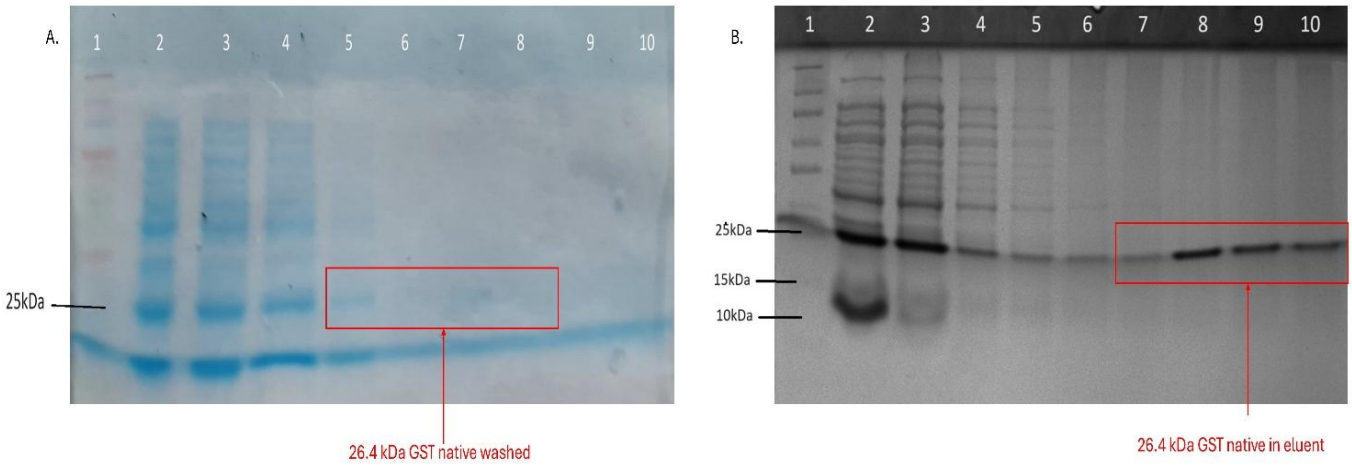
Supplementary Figure 9. Detection of expressed His-Rv2272 and poor yield purification using HisPur Cobalt resin Batch purification shown by A. SDS-PAGE and B. Western Blot Lane 1: Spectra multicolor Broad range protein ladder (Thermo Fisher Scientific), Lane 2: His-Rv3811, Lane 3: His-Rv2730, Lane 4: His-Rv2730,**Lane 5: His-Rv2272,** Lane 6: His-Rv2272, Lane 7: Blank, Lane 8: Rv2272-pCOLDTF/pGRO7 whole lysate, Lane 9: Rv2272-pCOLDTF/pGRO7 elution 2, lane 10:Blank and **C. SDS-PAGE gel (12%) for the purification of His-Rv2272 (13.8kDa).** Lane 1: Spectra Multicolor broad range ladder (Thermo Fisher Scientific), Lane 2: Rv2272 induced pellet, Lane 3: RecRv2272 Soluble lysate, Lane 4: RecRv2272 lysate flowthrough, Lane 5: wash 1, Lane 6: wash 2, Lane 7: wash 3, Lane 8: RecRv2272 Elution 1, Lane 9: RecRv2272 elution 2, Lane 10: RecRv2272 elution 3.



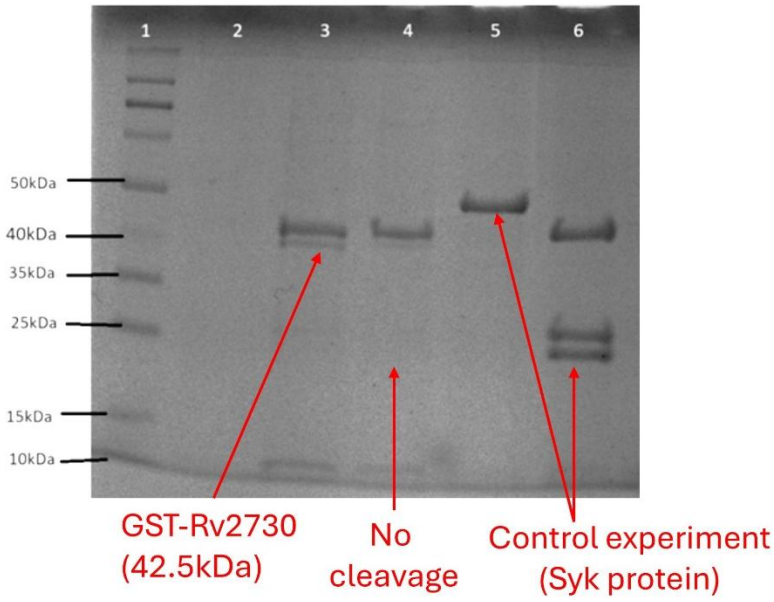
Supplementary Figure 10. SDS-PAGE gel (12%) of A. Solubilization of GST-Rv3717 with signal peptide (50.3 kDa), Lane 1: Spectra Multicolor Broad range protein ladder (Thermo Fisher Scientific), Lane 2: Total cell lysate, Lane 3: Soluble clear lysate (PBS lysis method), Lane 4: blank, Lane 5: Pellet fraction, Lane 6: Blank, Lane 7: Total lysate, Lane 8: solubilized protein lysate (1% SDS method) **B. Solubilization of GST-Rv3491 with signal peptide (46.7 kDa).** Lane 1: Spectra Multicolor Broad range protein ladder (Thermo Fisher Scientific), Lane 2: Total Lysate, Lane 3: clear lysate Lane 4: Pellet, Lane 5: total lysate prep (1% SDS method), Lane 6: clear lysate (solubilized protein), Lane 7: blank, Lane 8: pellet



Supplementary Figure 11. SDS-PAGE gel (12%) of A. Solubilization of GST-Rv3491 with signal peptide (46.7 kDa), Lane 1: Spectra Multicolor Broad range protein ladder (Thermo Fisher Scientific), Lane 2: Total Cell Lysate, Lane 3: clear lysate (PBS lysis method) Lane 4: Pellet fraction, Lane 5: Precelley total lysate prep, Lane 6: Soluble clear lysate (1% SDS method), Lane 7: blank, Lane 8: pellet fraction, **B. Purification attempt of GST-Rv3491 with signal peptide.** Lane 1: Spectra Multicolor Broad range protein ladder (Thermo Fisher Scientific), Lane 2: Total cell lysate, Lane 3: Soluble protein (clear lysate), Lane 4: Before wash, Lane 5-7: Wash step, Lane 8-10: Elution



Supplementary Figure 12. SDS-PAGE gel (12%) of purified of GST native (26.4 kDa) by Pierce glutathione agarose (Thermo Fisher Scientific) by batch method purification. **A. Purification using the 1% SDS method B. Purification using PBS lysis buffer.** Lane 1: Broad range protein ladder, Lane 2: GST only total lysate Lane 3: GST only Soluble protein (clear lysate), Lane 4: GST only flowthrough Lane 5: GST only wash 1, Lane 6: GST only wash 2, Lane 7: Elution 1, Lane 8: Elution 2, Lane 9: Elution 2, Lane 10: Elution 3.



Supplementary Figure 13. SDS-PAGE gel (12%) of in solution HRV 3C Protease Cleavage of GST-Rv2730 and positive control (GST-Syk). Lane 1: Spectra multicolor Broad range protein ladder, Lane 2: Blank, Lane 3: GST-Rv2730, Lane 4: GST-Rv2730 cleavage, Lane 5: GST-Syk protein, Lane 6: Positive control (cleaved GST-Syk protein)

B3. Chapter five supplementary data

Supplementary table 1. Reactivity of GST-fused proteins among host-derived extracellular matrix components to determine adhesin properties by Adhesin ELISA assay at absorbance (abs) of 450nm. (reactive protein=RP)

ELISA Coated protein	Abs 1	Abs 2	Abs 3	Mean	SD (±)	Reactivity %
GST Native						
Laminin	-0.086	-0.097	-0.077	-0.087	0.010	0
Fibronectin	0.098	0.054	0.042	0.065	0.029	13
Collagen	0.026	0.049	0.06	0.045	0.017	9
GST native (RP)	0.526	0.503	0.455	0.494	0.036	100
BSA negative control	0.011	0.043	0.007	0.020	0.020	4
Rv2272						
Laminin	0.631	0.447	0.562	0.547	0.093	20
Fibronectin	0.302	0.273	0.290	0.289	0.015	10
Collagen	0.101	0.149	0.124	0.125	0.024	5
GST-Rv2272 (RP)	2.554	2.736	2.784	2.692	0.121	100
BSA negative control	0.015	0.043	0.027	0.029	0.014	1
Rv2598						
Laminin	0.218	0.184	0.150	0.184	0.034	8
Fibronectin	0.290	0.336	0.265	0.297	0.036	13
Collagen	0.053	0.084	0.105	0.081	0.026	3
GST-Rv2598 (RP)	2.338	2.321	2.412	2.357	0.048	100
BSA negative control	0.023	0.008	0.017	0.016	0.008	<1
Rv2730						
Laminin	1.292	1.451	1.334	1.359	0.082	89
Fibronectin	1.150	1.038	1.014	1.067	0.073	70
Collagen	0.939	0.980	0.879	0.932	0.051	61
GST-Rv2730 (RP)	1.498	1.459	1.650	1.535	0.101	100
BSA negative control	0.098	0.115	0.137	0.117	0.019	8
Rv3209						
Laminin	0.140	0.221	0.187	0.183	0.041	9
Fibronectin	0.373	0.369	0.374	0.372	0.002	19
Collagen	0.189	0.209	0.211	0.203	0.012	10
GST-Rv3209 (RP)	1.934	1.948	1.936	1.940	0.008	100
BSA negative control	0.036	0.023	0.042	0.034	0.009	2
Rv3491						
Laminin	2.384	2.451	2.149	2.328	0.159	92
Fibronectin	2.117	2.224	1.959	2.100	0.133	83
Collagen	2.050	1.998	2.139	2.062	0.071	81
GST-Rv3491 (RP)	2.411	2.504	2.719	2.544	0.158	100
BSA negative control	0.124	0.121	0.086	0.110	0.021	4

ELISA Coated protein	Abs 1	Abs 2	Abs 3	Mean	SD (±)	Reactivity %
Rv3717						
Laminin	0.024	0.035	0.030	0.029	0.006	6
Fibronectin	0.296	0.252	0.314	0.287	0.032	62
Collagen	0.077	0.102	0.064	0.081	0.019	18
GST-Rv3717 (RP)	0.438	0.411	0.539	0.462	0.067	100
BSA negative control	0.015	0.028	0.0435	0.026	0.010	6
Rv3811						
Laminin	0.317	0.298	0.287	0.300	0.031	16
Fibronectin	1.531	1.388	1.360	1.426	0.092	74
Collagen	0.000	-0.006	0.001	-0.002	0.004	0
GST-Rv3811 (RP)	1.934	1.921	1.944	1.933	0.012	100
BSA negative control	0.186	0.144	0.116	0.148	0.035	8

Supplementary Table 2. Laminin binding to *Mycobacterium tuberculosis* protein interactive Receptor to ligand binding residues showing the groups of interactive bonds with RMSD scores.

Laminin:GST (121 interactive Receptor: Ligand binding residues)				
Single interaction (28 Bonds)	**Double interactions (24 Bonds)**	**Triple interactions (12 Bonds)**	**Quadruple interactions (32 Bonds)**	**More than 4 interactions (25 Bonds)**
W498 - E34 4.359	Q528 - S196 2.211	M535 - K194 4.361	R508 - E34 4.555	I536 - L193 4.027
E525 - E29 3.006	Q528 - K197 3.882	M535 - S195 4.356	R508 - D36 4.515	I536 - K194 4.922
S526 - E30 4.450	G537 - S196 3.602	M535 - S196 4.599	R508 - E37 2.838	I536 - S195 3.709
C529 - S196 3.016	G537 - Y198 3.763	R538 - K9 2.381	R508 - K40 4.921	I536 - S196 4.834
S530 - S196 4.692	R756 - H215 3.695	R538 - E34 3.701	E542 - A200 4.346	I536 - Y198 3.672
C531 - S196 2.872	R756 - K218 2.879	R538 - S196 4.767	E542 - W201 4.122	I536 - I199 4.757
R532 - K194 4.555	L781 - P167 2.732	S1091 - N54 1.423	E542 - D214 3.190	I536 - A200 2.303
P533 - K194 4.716	L781 - M168 1.090	S1091 - L55 4.942	E542 - H215 3.859	A1088 - F 52 3.313
H534 - K194 3.099	S782 - L163 4.895	S1091 - Q67 4.681	E544 - L193 2.526	A1088 - P53 4.598
C540 - S196 4.480	S782 - P167 3.027	G1093 - L13 4.884	E544 - K194 3.326	A1088 - N54 3.570
V543 - A200 4.505	R795 - M168 3.337	G1093 - N54 4.231	E544 - Y198 4.208	A1088 - T66 4.062
Y547 - K194 4.255	R795 - D171 2.936	G1093 - Q67 3.143	E544 - A200 4.956	A1088 - Q67 3.683
M751 - K 218 3.828	D1066 - K125 3.411		P545 - K11 4.916	F1092 - Y7 4.379
N757 - H215 4.908	D1066 - E128 4.433		P545 - Y164 3.517	F1092 - W8 3.395
S780 - M168 4.277	P1070 - G97 4.693		P545 - Y198 3.675	F1092 - L13 3.112
V784 - Y164 4.381	P1070 - D101 3.545		P545 - A200 3.518	F1092 - N54 3.404
D786 - K218 2.809	P1094 - Q67 3.167		R1067 - K125 1.753	F1092 - L55 3.041
N788 - K218 3.119	P1094 - Y104 4.884		R1067 - L126 4.576	F1092 - P56 4.675
Q793 - P167 3.566	E1098 - R108 2.978		R1067 - E128 4.028	F1092 - Q67 4.109
N1087 - N54 4.644	E1098 - K125 4.805		R1067 - M129 3.589	F1092 - S68 4.618
S1095 - Y104 2.839	P1107 - W8 4.146		A1089 - K45 3.090	M1106 - W41 3.626
C1096 - R108 2.213	P1107 - R42 3.732		A1089 - F52 3.584	M1106 - K45 2.769
N1097 - R108 2.472	F1109 - K45 3.278		A1089 - P53 1.760	M1106 - P53 3.404
F1099 - K125 4.669	F1109 - F46 4.789		A1089 - N54 3.673	M1106 - N54 2.485
G1108 - R42 4.940			H1090 - K45 3.369	M1106 - L55 3.321
V1127 - F46 4.110			H1090 - F46 4.386	
E1128 - N43 4.728			H1090 - P53 3.962	
C1129 - R42 4.926			H1090 - N54 4.413	
			Q1104 - L13 4.625	
			Q1104 - Y104 4.860	
			Q1104 - S107 2.931	
			Q1104 - R108 4.072	

Range:W498- C1129

Laminin: Rv2272 (86 interactive Receptor: Ligand binding residues)				
Single interaction (17 Bonds)	**Double interactions (12 Bonds)**	**Triple interactions (15 Bonds)**	**Quadruple interactions (20 Bonds)**	**More than 4 (22 Bonds)**
L1001 - Y16 4.236	N1031 - I68 4.806	Y1002 - E13 3.012	H1003 - Y16 2.566	Y1032 - Q30 3.032
Y1016 - E23 1.936	N1031 - S71 3.823	Y1002 - P14 4.521	H1003 - T19 3.517	Y1032 - A33 2.725
Q1022 - E13 4.100	D1044 - E23 4.932	Y1002 - Y16 3.053	H1003 - L20 4.261	Y1032 - L34 2.714
Q1023 - E13 2.582	D1044 - L27 3.259	R1026 - W78 4.599	H1003 - E23 3.328	Y1032 - L37 3.288
D1024 - R86 2.869	N1064 - L75 3.286	R1026 - D82 4.924	K1027 - E23 3.620	Y1032 - A67 3.545
C1028 - F26 4.577	N1064 - Q79 4.017	R1026 - R86 2.189	K1027 - F26 3.807	Y1032 - I68 4.562
G1042 - L34 3.873	N1097 - L104 4.497	V1029 - F26 3.899	K1027 - L27 2.430	Y1032 - S71 3.314
W1073 - V61 4.040	N1097 - V107 3.249	V1029 - L75 1.673	K1027 - W78 3.874	L1033 - L37 2.912
Q1074 - I68 4.716	C1144 - I120 4.939	V1029 - W78 3.311	A1076 - L37 3.214	L1033 - A64 4.664
L1075 - I68 3.232	C1144 - T121 3.110	E1119 - A113 2.858	A1076 - A64 2.861	L1033 - A67 4.001
S1077 - I68 4.853	D1145 - I120 4.807	E1119 - L114 3.992	A1076 - I65 3.975	L1033 - I68 2.613
Q1102 - L114 3.726	D1145 - T121 3.093	E1119 - A117 2.906	A1076 - I68 2.159	L1033 - S71 4.931
E1116 - K118 2.729		Q1143 - A117 3.710	F1099 - V61 4.841	S1043 - F26 4.468
F1121 - T121 4.410		Q1143 - I120 3.978	F1099 - I65 2.964	S1043 - L27 3.015
T1141 - I120 3.762		Q1143 - T121 2.699	F1099 - I68 2.632	S1043 - Q30 3.445
S1147 - T121 4.189			F1099 - L104 2.936	S1043 - R31 4.822
V1152 - I1R20 4.357			Q1146 - A117 3.465	S1043 - L34 4.297
			Q1146 - K118 3.278	T1100 - V61 4.772
			Q1146 - T121 1.340	T1100 - I65 3.156
			Q1146 - G122 3.668	T1100 - V107 3.071
				T1100 - G108 4.419
				T1100 - A111 3.785

Range: L1001- V1152

Laminin: Rv2598 (73 interactive Receptor: Ligand binding residues)				
Single interaction (16 Bonds)	**Double interactions (12 Bonds)**	**Triple interactions (21 Bonds)**	**Quadruple interactions (4 Bonds)**	**More than 4 (20 Bonds)**
V568 - A16 4.705	R573 - T92 4.912	P566 - A16 4.897	S863 - H104 4.973	V829 - E91 3.059
S569 - T90 3.444	R573 - D94 4.851	P566 - R69 3.277	S863 - E107 3.963	V829 - H93 3.334
V571 - T92 1.414	E695 - P10 4.293	P566 - D74 2.542	S863 - R108 2.787	V829 - L101 4.050
E572 - T92 4.396	E695 - V13 2.132	G567 - A16 2.495	S863 - R111 4.643	V829 - W144 3.753
P592 - A16 4.088	Q826 - R100 4.667	G567 - R69 3.865		V829 - H157 4.174
Q619 - S151 4.038	Q826 - H104 2.455	G567 - D74 3.787		F843 - L101 2.556
Y698 - S149 3.514	G845 - R108 3.871	I570 - T90 3.098		F843 - H104 3.109
L700 - T92 4.884	G845 - R142 4.944	I570 - E91 4.300		F843 - L105 3.619
L825 - R100 3.025	Q862 - H104 3.504	I570 - T92 4.789		F843 - W115 4.935
G827 - H104 4.662	Q862 - R108 3.671	R590 - T92 3.472		F843 - R142 2.658
S828 - L101 4.813	C864 - R108 2.207	R590 - G153 4.701		F843 - W144 3.051
H841 - H157 4.447	C864 - A112 4.481	R590 - T154 2.546		Q844 - R108 2.731
C842 - R142 2.318		S696 - V13 3.767		Q844 - W115 3.308
L855 - E133 4.820		S696 - S14 3.069		Q844 - A131 2.262
H858 - E133 4.582		S696 - L17 4.017		Q844 - A132 3.458
P866 - E113 3.577		N830 - H89 4.494		Q844 - E133 3.476
		N830 - E91 2.509		Q844 - R140 3.758
		N830 - H93 3.682		Q844 - W141 4.134
		Q865 - R108 4.323		Q844 - R142 2.723
		Q865 - R111 2.876		Q844 - W144 4.039
		Q865 - A112 3.599		

Range: P566-P866

Laminin: Rv2730 (90 interactive Receptor: Ligand binding residues)				
Single interaction (27 Bonds)	**Double interactions (20 Bonds)**	**Triple interactions (6 Bonds)**	**Quadruple interactions (16 Bonds)**	**More than 4 interactions (21 Bonds)**
R508 - Q99 3.864	R579 - R152 3.757	A1076 - V50 3.574	N1097 - A71 3.132	F1099 - G52 3.470
T551 - R158 4.460	R579 - Q156 4.661	A1076 - G51 2.147	N1097 - T86 4.027	F1099 - S53 4.574
L556 - R158 4.319	I580 - R152 3.674	A1076 - G52 2.335	N1097 - P87 4.391	F1099 - A71 3.450
E558 - R158 1.956	I580 - Q156 2.239	E1102 - R69 2.559	N1097 - T88 1.216	F1099 - S72 3.394
Q577 - R152 4.175	P750 - M2 3.543	E1102 - T88 4.036	Q1118 - H67 3.676	F1099 - N73 3.171
D578 - R152 3.268	P750 - L155 3.547	E1102 - S89 3.141	Q1118 - S91 2.181	T1100 - S53 4.303
P581 - Q156 3.937	N788 - G157 4.963		Q1118 - S92 4.137	T1100 - A54 1.846
T584 - R158 3.677	N788 - R158 4.500		Q1118 - N93 4.306	T1100 - V55 4.817
G585 - R158 3.147	Q791 - G157 2.701		L1120 - T58 4.532	T1100 - R69 3.179
M751 - L155 4.413	Q791 - R158 2.757		L1120 - R65 2.672	T1100 - C70 4.252
D786 - G154 3.733	Y1032 - S75 2.616		L1120 - H67 2.891	T1100 - A71 3.162
G789 - R158 3.299	Y1032 - V76 4.344		L1120 - T93 3.539	T1100 - T88 3.214
G790 - R158 3.891	C1103 - T88 4.822		P1142 - R38 4.770	E1119 - C40 4.351
G800 - R158 3.812	C1103 - S89 4.133		P1142 - N39 4.813	E1119 - H41 4.314
R801 - R158 4.271	I1139 - R38 3.468		P1142 - C40 2.800	E1119 - H42 2.641
R805 - R158 3.810	I1139 - R65 4.826		P1142 - R65 2.849	E1119 - G43 4.733
N1031 - S75 3.441	E1140 - A16 3.219			E1119 - Q56 3.035
L1033 - S75 3.211	E1140 - C40 3.040			E1119 - L57 4.319
Q1074 - G51 4.475	T1141 - C40 2.844			E1119 - T58 2.829
S1077 - G51 3.684	T1141 - H42 3.052			E1119 - H67 3.173
Q1104 - T88 4.408				E1119 - T93 4.631
R1112 - R69 4.054				
F1121 - T93 2.683				
P1136 - R38 2.477				
R1137 - R38 3.943				
G1138 - R38 4.361				
Q1143 - H42 2.818				

Range: R508-Q1143

Laminin: Rv3209 (87 interactive Receptor: Ligand binding residues)				
Single interaction (16 Bonds)	**Double interactions (16 Bonds)**	**Triple interactions (21 Bonds)**	**Quadruple interactions (12 Bonds)**	**More than 4 interactions (22 Bonds)**
G499 - Y157 4.673	D505 - L111 4.360	W498 - Q110 2.623	Q793 - G168 4.747	R508 - Q110 2.501
L500 - Y157 4.733	D505 - F112 4.715	W498 - Y157 3.052	Q793 - A169 2.718	R508 - L111 2.804
S501 - Y157 2.982	Q528 - R159 4.779	W498 - S158 4.875	Q793 - Q170 3.406	R508 - F112 3.458
L504 - F112 2.967	Q528 - N178 4.976	D503 - L111 3.798	Q793 - T171 4.643	R508 - D113 2.579
P509 - Q110 4.017	I536 - T176 3.148	D503 - F112 3.217	F1092 - V118 4.663	R508 - Y157 3.143
D511 - Q110 4.321	I536 - N178 3.688	D503 - Y157 4.537	F1092 - V152 3.468	R508 - S158 4.800
S526 - R159 2.772	N541 - G107 4.353	G537 - T108 4.848	F1092 - T154 3.529	R538 - T108 4.074
G527 - R159 4.824	N541 - T108 3.113	G537 - R159 4.866	F1092 - I181 4.746	R538 - Q110 1.863
P545 - T176 3.427	N1097 - G124 3.395	G537 - N178 2.606	Q1118 - V127 4.438	R538 - Y157 4.218
D578 - P92 4.364	N1097 - T148 2.557	Q539 - T108 1.834	Q1118 - T128 2.066	R538 - R159 2.653
D753 - P135 4.479	F1099 - G145 3.445	Q539 - K109 4.952	Q1118 - D129 4.763	R538 - N178 4.189
L781 - Q170 3.190	F1099 - Q147 4.856	Q539 - Q110 4.981	Q1118 - F130 2.943	E542 - T106 3.730
E1098 - Q147 4.922	C1105 - V118 3.751	T1100 - R123 3.627		E542 - G107 3.000
C1103 - P126 3.590	C1105 - P126 4.963	T1100 - G124 3.963		E542 - T108 4.779
M1106 - V118 4.887	G1108 - T128 4.748	T1100 - F125 3.024		E542 - N161 2.373
F1121 - F130 2.441	G1108 - F130 3.121	Q1102 - G124 4.600		E542 - T176 4.067
		Q1102 - F125 3.763		E542 - N178 3.569
		Q1102 - P126 2.555		P1107 - N116 3.339
		Q1104 - S150 3.037		P1107 - V118 4.532
		Q1104 - V152 4.257		P1107 - T128 4.249
		Q1104 - T185 3.728		P1107 - F130 3.594
				P1107 - T154 3.494

Range: G499-F1121

Laminin: Rv3491 (78 interactive Receptor: Ligand binding residues)

Single interaction (13 Bonds)	Double interactions (12 Bonds)	Triple interactions (27 Bonds)	Quadruple interactions (No Bonds)	More than 4 interactions (26 Bonds)
E496 - S71 3.812	R532 - R180 4.342	S521 - S71 2.588	NIL	L518 - P34 3.380
E525 - R30 4.385	R532 - R181 3.436	S521 - D72 4.838		L518 - P35 4.183
S526 - R30 2.646	P605 - Q179 4.945	S521 - P73 4.210		L518 - W76 2.299
Q528 - R30 4.826	P605 - Y184 3.451	F523 - R30 3.922		L518 - R180 2.809
C529 - G33 4.195	Y606 - Y184 2.292	F523 - P34 4.067		L518 - Y181 2.785
M535 - L32 4.620	Y606 - Y192 3.777	F523 - E141 4.217		N519 - Y70 4.092
I536 - L32 3.941	S641 - M190 4.586	S530 - G33 3.759		N519 - S71 3.084
S607 - Y192 2.924	S641 - Y192 4.275	S530 - P34 3.396		N519 - D72 4.283
T639 - M190 4.352	R642 - N186 4.899	S530 - P73 4.844		N519 - P73 3.814
L765 - Y184 3.896	R642 - Y192 2.939	C531 - L32 3.375		N519 - W76 4.237
L766 - Y184 2.411	H767 - Q179 3.717	C531 - G33 3.776		H534 - P173 3.450
G770 - P177 3.471	H767 - Y184 3.889	C531 - R180 3.032		H534 - P174 3.159
A772 - P177 4.338		P533 - D37 2.841		H534 - S176 3.682
		P533 - V178 3.687		H534 - P177 3.419
		P533 - R180 3.312		H534 - V178 4.563
		E544 - L32 4.781		Y547 - P173 4.452
		E544 - G172 4.535		Y547 - P174 3.270
		E544 - P173 2.250		Y547 - P175 4.361
		S640 - N186 4.001		Y547 - S176 4.434
		S640 - M190 1.657		Y547 - P177 3.945
		S640 - Y192 4.010		Q768 - Q179 3.415
		K674 - Q179 2.588		Q768 - R180 4.452
		K674 - P182 4.260		Q768 - Y181 1.644
		K674 - Y184 2.811		Q768 - P182 3.726
		T769 - P177 3.465		Q768 - L183 4.421
		T769 - V178 4.103		Q768 - Y184 2.798
		T769 - Q179 4.665		

Range: E496- A772

Laminin: Rv3717 (133 interactive Receptor: Ligand binding residues)

Single interaction (21 Bonds)	Double interactions (14 Bonds)	Triple interactions (30 Bonds)	Quadruple interactions (28 Bonds)	More than 4 interactions (40 Bonds)
F811 - E217 1.916	Y847 - L79 2.990	R849 - R80 3.159	P856 - W75 4.546	R853 - L79 3.127
K819 - E217 3.792	Y847 - R82 4.088	R849 - A216 3.107	P856 - L79 4.838	R853 - R80 2.494
T837 - A216 4.738	A900 - L116 4.652	R849 - E217 3.164	P856 - R82 4.577	R853 - L81 4.516
C876 - R82 4.272	A900 - R117 2.876	D875 - R82 4.706	P856 - G97 4.687	R853 - A83 2.750
A848 - R80 3.988	G949 - Q237 3.985	D875 - R96 4.877	D877 - R82 4.914	R853 - A84 4.721
E882 - R91 2.379	G949 - A240 3.015	D875 - G97 3.546	D877 - N86 1.808	R853 - Q220 3.297
M891 - R91 4.211	A959 - Q239 4.116	V879 - A83 4.664	D877 - V90 4.261	N885 - F30 3.983
G892 - R91 3.973	A959 - A240 2.545	V879 - N86 3.535	D877 - T92 3.797	N885 - A93 3.284
G901 - R117 4.426	Q974 - A150 4.613	V879 - A87 4.257	P878 - R82 4.019	N885 - L94 2.983
Q928 - R117 4.532	Q974 - G151 3.916	L899 - A115 3.184	P878 - A83 4.696	N885 - M112 4.431
A930 - R117 2.842	H979 - R161 3.341	L899 - L116 4.979	P878 - N86 2.432	N885 - L116 3.841
S932 - R117 4.800	H979 - D162 4.906	L899 - R117 4.825	P878 - T92 4.969	R931 - R117 2.815
V945 - R241 3.066	L1021 - K171 2.713	I951 - Q239 4.850	L884 - R82 3.732	R931 - P118 2.329
C946 - R241 2.924	L1021 - D178 2.495	I951 - A240 3.290	L884 - T92 1.522	R931 - N119 2.929
G952 - R241 4.287		I951 - R241 2.341	L884 - A93 3.521	R931 - Y194 3.097
G961 - Q155 4.192		S960 - Q155 4.966	L884 - L94 4.777	R931 - P195 4.030
S972 - A147 3.376		S960 - T236 4.868	D947 - R117 4.316	R931 - R241 4.451
C976 - R158 1.961		S960 - Q239 4.661	D947 - Y194 4.263	P948 - N119 4.741
Q977 - R158 3.046		C973 - A147 4.559	D947 - A240 3.990	P948 - I148 4.533
C978 - R158 4.215		C973 - G151 4.145	D947 - R241 4.818	P948 - P152 4.172
Y1002 - D162 4.535		C973 - P152 4.209	Y950 - A147 4.136	P948 - P195 4.866
		T985 - A235 4.942	Y950 - I148 4.724	P948 - Q237 2.037
		T985 - T236 3.756	Y950 - A240 1.671	P948 - A240 4.100
		T985 - Q239 3.141	Y950 - R241 3.563	Y962 - G151 3.776
		Q1022 - Q165 2.806	T984 - Q155 3.221	Y962 - P152 3.091
		Q1022 - K171 3.761	T984 - I159 3.672	Y962 - Q155 2.463
		Q1022 - D178 4.224	T984 - T236 3.286	Y962 - F233 3.989
		Q1023 - D162 4.295	T984 - Q239 4.810	Y962 - T236 4.760
		Q1023 - Q165 2.948		Y962 - Q237 3.183
		Q1023 - K171 4.826		P975 - A150 3.503
				P975 - G151 2.494
				P975 - P152 4.235
				P975 - V154 3.814
				P975 - Q155 3.760
				P975 - R158 4.280
				N980 - R158 2.552
				N980 - I159 4.890
				N980 - R161 1.581
				N980 - D162 2.413
				N980 - L180 4.198

Range: F811-Q1023

Laminin-Rv3811 (83 interactive Receptor: Ligand binding residues)				
Single interaction (15 Bonds)	**Double interactions (12 Bonds)**	**Triple interactions (18 Bonds)**	**Quadruple interactions (24 Bonds)**	**More than 4 (14 Bonds)**
G546 - V44 3.497	F1011 - K399 3.140	H534 - I36 4.448	P545 - V44 2.392	S783 - V42 4.658
A772 - V42 4.590	F1011 - E402 2.935	H534 - G37 3.922	P545 - R45 3.641	S783 - V44 2.767
E774 - V42 4.438	D1048 - R316 4.533	H534 - R45 2.737	P545 - E46 3.095	S783 - R68 3.031
C775 - V42 4.738	D1048 - E396 2.374	E544 - V44 4.708	P545 - Q124 3.408	S783 - W81 2.639
P777 - R68 3.832	K1049 - E396 3.905	E544 - R45 2.386	Y547 - G37 3.477	S783 - A126 3.431
L781 - W81 4.330	K1049 - K399 4.860	E544 - E46 4.880	Y547 - T43 2.941	T1051 - P312 3.562
C1012 - K399 4.022	L1055 - P185 4.038	S782 - R68 4.229	Y547 - V44 3.931	T1051 - I313 3.391
Y1017 - P312 4.180	L1055 - T187 2.325	S782 - R70 2.773	Y547 - R45 2.046	T1051 - R316 3.004
C1054 - P185 3.585	C1056 - W183 3.872	S782 - W81 3.826	V784 - V44 3.683	T1051 - E396 3.958
L1057 - W183 4.571	C1056 - P185 3.547	A1050 - P312 3.285	V784 - R70 3.183	Q1053 - W183 3.667
P1058 - Q182 2.844	A1069 - T181 4.347	A1050 - R316 1.870	V784 - W81 4.813	Q1053 - P185 3.724
V1060 - W183 3.821	A1069 - L238 2.603	A1050 - E396 2.027	V784 - Q124 2.935	Q1053 - P186 4.718
R1067 - D234 2.648		N1059 - G180 3.532	I1061 - W183 3.831	Q1053 - Y269 4.445
C1068 - S236 4.948		N1059 - T181 3.554	I1061 - S236 2.085	Q1053 - I313 4.683
N1071 - Q176 2.858		N1059 - Q182 3.064	I1061 - P237 2.787	
		G1062 - W183 3.603	I1061 - L238 3.576	
		G1062 - P237 3.365	Q1063 - Y269 3.222	
		G1062 - Y269 3.094	Q1063 - T311 3.663	
			Q1063 - P312 4.109	
			Q1063 - I313 4.572	
			P1070 - Q176 4.580	
			P1070 - S236 4.447	
			P1070 - L238 3.828	
			P1070 - E239 4.402	

Range: G546- N1071

Supplementary table 3. Fibronectin binding to *Mycobacterium tuberculosis* protein interactive Receptor to ligand binding residues showing the groups of interactive bonds with RMSD scores.

Fibronectin:GST (80 interactive Receptor: Ligand binding residues)					
Single interaction (12 Bonds)	**Double interactions (10 Bonds)**	**Triple interactions (9 Bonds)**	**Quadruple interactions (8 Bonds)**	**More than 4 interactions (41 Bonds)**	
V21 - K113 4.884	H34 - Y111 2.902	S24 - R35 4.654	P26 - W8 2.358	K25 - R35 3.602	I37 - P56 4.483
Q23 - Q207 2.779	H34 - Q207 4.323	S24 - R42 4.681	P26 - G38 3.898	K25 - R42 4.001	N38 - L13 3.253
Y29 - Q207 4.465	Y35 - S107 4.628	S24 - Q207 3.385	P26 - W41 3.333	K25 - G205 4.133	N38 - N54 1.585
G32 - Y111 4.902	Y35 - R108 4.223	Q40 - L100 4.972	P26 - R42 3.788	K25 - W206 2.179	N38 - L55 3.019
W41 - R108 3.218	G56 - R45 3.451	Q40 - Y104 2.808	Y55 - K45 2.986	K25 - Q207 2.611	N38 - P56 3.144
E42 - R108 2.605	G56 - N54 4.442	Q40 - R108 4.989	Y55 - P53 3.468	K33 - R108 3.045	N38 - T66 4.168
T53 - Q67 3.767	G57 - R45 3.153	N1962 - L82 4.811	Y55 - N54 3.689	K33 - I109 4.481	N38 - Q67 3.040
C54 - N54 3.003	G57 - F46 4.305	N1962 - G83 4.892	Y55 - Q67 3.546	K33 - Y111 4.654	N38 - S68 3.958
S58 - F46 4.088	T81 - K87 3.612	N1962 - R89 2.471		K33 - S112 3.967	N38 - Y104 4.344
G79 - K87 2.316	T81 - K91 4.376			K33 - L118 4.072	Q39 - L13 3.583
N80 - K87 3.085				Q36 - Y7 4.688	Q39 - S107 2.554
P1961 - D77 3.554				Q36 - W8 3.201	Q39 - R108 4.213
				Q36 - I10 4.447	Q39 - Y111 3.404
				Q36 - Y111 3.061	Q39 - Q204 4.974
				Q36 - G205 3.372	R59 - L45 4.509
				I37 - Y7 4.623	R59 - F46 4.109
				I37 - L13 3.447	R59 - E51 3.598
				I37 - W41 2.798	R59 - F52 2.927
				I37 - K45 4.593	R59 - P53 2.354
				I37 - P53 4.784	
				I37 - N54 0.785	
				I37 - L55 2.493	

Range: V21- T81, P1961-N1962

Fibronectin: Rv2272 (133 interactive Receptor: Ligand binding residues)

Single interaction (40 Bonds)		Double interactions (24 Bonds)	Triple interactions (24 Bonds)	Quadruple interactions (24 Bonds)	More than 4 (21 Bonds)
T330 - R93 1.945	P856 - Q57 4.361	E428 - R83 2.667	M447 - D11 3.589	R433 - P92 4.644	R469 - D7 3.020
Q334 - F18 2.313	S858 - Q57 4.116	E428 - R87 3.666	M447 - V12 3.712	R433 - R93 2.327	R469 - A9 3.766
G336 - R93 2.808	T889 - I68 4.778	M431 - R87 4.518	M447 - H88 2.264	R433 - H94 2.941	R469 - T10 2.549
H337 - R93 4.887	L908 - L34 4.401	M431 - H94 4.468	E851 - L114 3.028	R433 - P95 3.273	R469 - D11 4.434
L338 - R93 4.495	V934 - L37 4.202	W438 - R87 2.543	E851 - V115 4.265	D439 - R86 4.099	R469 - V12 2.571
H358 - R93 3.402	D1716 - L36 4.584	W438 - L89 3.213	E851 - K118 4.768	D439 - R87 1.785	G937 - L37 3.001
T389 - Y99 3.968	R1738 - T32 2.827	K440 - R86 4.314	R892 - L34 3.403	D439 - H88 3.528	G937 - A40 4.143
S390 - Y99 4.354	D1753 - T51 3.087	K440 - R87 3.589	R892 - L37 4.267	D439 - L89 3.532	G937 - V41 3.394
E391 - Y99 2.934	T1788 - R24 4.905	S802 - L104 4.366	R892 - A38 4.873	Q441 - P14 4.871	G937 - A64 2.606
Y432 - 9H4 3.670	P1789 - R24 1.270	S802 - V107 2.242	R907 - L34 1.872	Q441 - R86 2.960	G937 - I68 4.792
Q437 - L89 4.366	N1791 - R31 1.634	V803 - I65 3.245	R907 - G35 4.748	Q441 - R87 4.235	R938 - L37 4.772
M444 - D11 4.802	V1792 - R31 4.100	V803 - V107 4.402	R907 - A38 4.672	Q441 - H88 4.099	R938 - V41 4.176
G445 - D11 1.737	Y1797 - R24 3.352	Q849 - V107 4.400	P909 - F26 3.983	Q829 - V107 3.813	R938 - V44 3.564
H446 - D11 3.331	L1815 - R17 4.920	Q849 - V110 2.508	P909 - Q30 3.563	Q829 - V110 1.825	R938 - Q57 3.218
Y465 - A9 4.415	S1821 - R17 2.621	E939 - L37 3.064	P909 - S71 4.841	Q829 - A111 2.787	R938 - A64 4.610
S466 - A9 4.713	V1822 - R17 3.590	E939 - V41 3.141	H936 - G60 4.687	Q829 - L114 2.679	D1818 - R17 3.682
D470 - D7 3.812	V1823 - R17 4.535	E1715 - T32 4.229	H936 - V61 4.408	I910 - Q30 4.500	D1818 - F18 2.622
E804 - I68 4.872		E1715 - L36 4.567	H936 - A64 3.236	I910 - L34 4.493	D1818 - L20 4.743
N831 - V107 3.914		P1817 - A21 4.578	S1819 - R17 4.271	I910 - L37 3.006	D1818 - A21 1.712
V1941 - T8 4.290		P1817 - R24 3.982	S1819 - A21 4.061	I910 - S71 4.842	D1818 - N22 4.476
D1943 - T8 3.403		K1925 - D7 4.645	S1819 - R24 3.567	Q1793 - T25 3.548	D1818 - R24 2.015
Q850 - L114 4.702		K1925 - T8 1.587	S1820 - R17 4.064	Q1793 - A28 4.297	
G854 - K118 4.795		Y1927 - T8 4.714	S1820 - L20 3.673	Q1793 - W29 3.136	
		Y1927 - T10 4.104	S1820 - R24 3.306	Q1793 - R77 4.636	

Range: T330-R938, D1818-D1943

Fibronectin: Rv2598 (111 interactive Receptor: Ligand binding residues)

Single interaction (20 Bonds)	Double interactions (30 Bonds)	Triple interactions (18 Bonds)	Quadruple interactions (16 Bonds)	More than 4 (27 Bonds)
W97 - R34 4.628	K92 - R34 4.733	S94 - P29 3.729	D93 - T32 4.454	E135 - M1 3.523
P133 - Q5 3.327	K92 - E57 2.982	S94 - T32 3.529	D93 - R34 2.697	E135 - L3 4.009
H134 - Q5 3.771	M95 - R34 3.257	S94 - Q44 3.936	D93 - Q44 2.647	E135 - Q5 3.307
W153 - M1 2.922	M95 - Q44 4.866	E119 - M1 4.020	D93 - E57 2.179	E135 - R106 2.608
C155 - M1 4.404	R132 - M1 1.866	E119 - L3 4.913	G120 - H4 2.656	E135 - T110 2.407
P565 - R140 4.992	R132 - L3 3.939	E119 - H4 3.758	G120 - L30 3.231	P589 - H104 3.397
P587 - R108 3.009	T136 - Q5 4.971	H591 - E91 4.768	G120 - A31 4.397	P589 - L105 4.017
I592 - R142 4.146	T136 - R106 3.614	H591 - R142 3.445	G120 - T32 4.993	P589 - R108 3.665
G642 - Q87 4.109	G137 - Q5 4.606	H591 - H157 4.525	G121 - H4 4.950	P589 - W115 2.643
Q644 - R161 4.724	G137 - R106 4.715	Y641 - Q87 2.062	G121 - L30 4.568	P589 - R142 3.943
T647 - R140 4.682	Q588 - H104 2.706	Y641 - H89 2.417	G121 - A31 4.702	H643 - H85 2.522
H1135 - V13 4.717	Q588 - R108 3.616	Y641 - H157 4.118	G121 - T32 4.531	H643 - L86 4.700
E1137 - V13 4.737	P589 - W144 3.110	E645 - E133 3.666	P1303 - L22 4.662	H643 - Q87 2.872
P1140 - S70 4.194	P589 - H157 4.851	E645 - R140 3.018	P1303 - N23 3.578	H643 - T159 3.536
P1222 - S70 4.946	S590 - R142 1.861	E645 - R142 4.038	P1303 - A24 3.951	H643 - S160 4.599
A1259 - G58 4.021	S590 - W144 4.892	A1302 - A24 4.081	P1303 - P25 3.223	H643 - R161 3.790
G1260 - G58 4.141	Q639 - R142 2.221	A1302 - P25 4.444		E1220 - C66 4.905
D1287 - R59 4.815	Q639 - T159 3.064	A1302 - R71 4.190		E1220 - V67 2.595
T1307 - R59 3.242	V646 - D135 4.961			E1220 - A68 4.426
Q1389 - D39 4.497	V646 - R140 3.972			E1220 - R69 4.872
	V1221 - S70 4.963			E1220 - S70 0.559
	V1221 - R71 4.688			E1220 - R71 2.787
	T1225 - R71 4.926			Q1223 - V21 3.459
	T1225 - G72 4.423			Q1223 - L22 4.502
	D1289 - G58 3.202			Q1223 - A24 4.723
	D1289 - R59 4.552			Q1223 - R71 3.256
	T1305 - L22 4.748			Q1223 - G72 4.818
	T1305 - N23 2.749			
	L1306 - N23 4.232			
	L1306 - P82 3.331			

Range: K92-C155, H591-T647, H1135-Q1389

Fibronectin: Rv2730 (81 interactive Receptor: Ligand binding residues)				
Single interaction (30 Bonds)	**Double interactions (18 Bonds)**	**Triple interactions (12 Bonds)**	**Quadruple interactions (16 Bonds)**	**More than 4 interactions (5 Bonds)**
P258 - R77 2.267	P769 - N4 3.086	Q789 - N4 4.644 SN	D770 - M2 4.263	Y259 - R77 2.873
G260 - R77 2.946	P769 - R6 2.926	Q789 - 6R 2.600 RG	D770 - M3 4.747	Y259 - L79 4.515
H261 - E126 4.882	T771 - T151 4.905	Q789 - Q7 2.573 GLN	D770 - N4 4.215	Y259 - L125 3.099
S688 - R144 4.462	T771 - R152 3.395	S1544 - M1 3.728	D770 - R6 4.767	Y259 - V128 3.027
L735 - Q7 4.690	T772 - T151 4.643	S1544 - M3 3.632	F1539 - G96 4.529	Y259 - A129 3.270
D766 - R6 3.490	T772 - L155 2.982	S1544 - Y61 2.980	F1539 - K97 2.470	
A767 - R6 2.937	V773 - L155 3.116	K1545 - Y61 4.599	F1539 - A98 4.659	
P768 - R6 4.090	V773 - Q156 2.724	K1545 - H101 2.85	F1539 - Q99 4.417	
R787 - N4 3.583	S786 - M2 4.538	K1545 - D105 2.528	T1540 - P63 4.181	
G827 - R158 4.762	S786 - N4 4.470	T1549 - Q99 2.009	T1540 - G95 3.014	
V846 - R152 2.678	I848 - R152 2.200	T1549 - G100 2.636	T1540 - G96 3.876	
V847 - R152 3.005	I848 - Q156 2.264	T1549 - H101 3.629	T1540 - K97 4.277	
Q849 - Q156 3.653	Q850 - L155 4.992		P1542 - Y61 4.077	
E851 - R158 2.955	Q850 - Q156 3.763		P1542 - G62 4.965	
T852 - R158 3.305	I1550 - Q99 3.949		P1542 - P63 4.365	
T853 - R158 4.621	I1550 - G100 4.712		P1542 - Q64 4.277	
G854 - R158 4.796	S1551 - Q99 3.000			
A1515 - M1 3.564	S1551 - G100 2.769			
P1516 - M1 2.726				
A1517 - M1 3.373				
V1518 - M1 4.582				
R1521 - P63 4.563				
Y1522 - P63 4.702				
Y1527 - Q99 1.648				
Q1537 - K97 2.289				
E1538 - K97 2.771				
G1543 - M3 4.736				
S1546 - M1 4.617				
T1547 - H101 4.416				
A1548 - H101 4.737				

Range: P258-H261, D770-G854, A1515-A1548

Fibronectin: Rv3209 (104 interactive Receptor: Ligand binding residues)					
Single interaction (19 Bonds)	**Double interactions (30 Bonds)**	**Triple interactions (15 Bonds)**	**Quadruple interactions (4 Bonds)**	**More than 4 interactions (36 Bonds)**	
V315 - T137 3.390	G240 - T185 4.596	N282 - A122 4.869	N345 - R104 4.930	P317 - V102 2.491	H486 - P98 4.868
F318 - R104 3.025	G240 - R186 3.635	N282 - V141 4.459	N345 - T106 4.996	P317 - R104 3.884	H486 - N166
T319 - R104 1.348	R241 - Q147 3.381	N282 - N143 2.294	N345 - P135 3.641	P317 - T137 2.887	4.390
F325 - R104 4.991	R241 - R186 4.982	T324 - R104 2.771	N345 - T137 4.784	P317 - K138 4.748	H486 - T167
S327 - P93 4.288	L316 - T137 4.563	T324 - V165 4.457		P317 - M139 3.424	4.551
D349 - P135 4.413	L316 - M139 3.608	T324 - A169 3.403		Y326 - L96 2.648	H486 - G168
H358 - P91 4.094	V360 - P91 2.345	T343 - L134 4.467		Y326 - T100 4.461	4.362
V362 - P91 4.018	V360 - P92 4.065	T343 - P135 2.662		Y326 - V102 3.183	H486 - Q170
D387 - P91 3.052	I434 - P91 3.959	T343 - T137 4.693		Y326 - M139	4.452
P416 - P92 4.861	I434 - P92 3.404	I477 - G168 4.011		3.682	R488 - L142
A418 - P92 3.129	Y479 - G168 3.480	I477 - A169 3.519		Y326 - V165 4.097	4.979
H420 - P92 3.508	Y479 - Q170 3.280	I477 - Q170 4.472		Y326 - N166 2.339	R488 - Q147
E421 - P92 4.627	K487 - N166 4.740	Y524 - R99 3.493		Y326 - T167 3.576	4.898
V474 - Q170 3.602	K487 - Q170 2.824	Y524 - P144 4.210		Y326 - G168 4.654	R488 - E149
D475 - Q170 3.150	E490 - V172 4.923	Y524 - G145 3.540		Y326 - A169 3.996	3.759
F485 - Q170 2.340	E490 - R186 2.944			S344 - R104 3.358	R488 - N166
H489 - R186 3.380	E491 - Q147 4.553			S344 - V105 4.252	3.772
G521 - Q147 4.547	E491 - R186 4.697			S344 - T106 4.719	R488 - Q170
S529 - P144 4.363	G492 - Q147 4.928			S344 - P135 3.001	4.069
	G492 - R186 4.893			S344 - W136	R488 - R186
	M494 - P98 2.022			2.839	3.866
	M494 - R99 4.461			S344 - T137 1.344	
	T522 - G145 4.194			Q348 - T106 3.327	
	T522 - Q147 2.618			Q348 - G107 2.959	
	F523 - R99 4.256			Q348 - T108 4.987	
	F523 - Q147 4.491			Q348 - L134 4.972	
	Q525 - P98 4.198			Q348 - P135 2.992	
	Q525 - R99 3.187				
	D528 - R99 0.979				
	D528 - P144 2.850				

Range: G240-S529

Fibronectin: Rv3491 (109 interactive Receptor: Ligand binding residues)					
Single interaction (11 Bonds)	**Double interactions (16 Bonds)**	**Triple interactions (21 Bonds)**	**Quadruple interactions (24 Bonds)**	**More than 4 interactions (37 Bonds)**	
D777 - N66 4.136	Q850 - N68 4.049	V776 - R65 4.125	F1539 - Q115	T853 - T64 4.878	T1540 - L161 1.829
D778 - R65 2.262	Q850 - D69 4.705	V776 - N66 3.695	2.700	T853 - R65 1.481	T1540 - L163 4.589
E851 - N68 2.823	T1505 - R104 3.151	V776 - N68 4.282	F1539 - S133	T853 - N66 3.497	K1545 - S108 3.136
T852 - N68 2.936	T1505 - R106 2.461	L1510 - R106 4.206	2.881	T853 - M67 4.456	K1545 - S109 3.143
G854 - R65 3.263	S1508 - R104 3.812	L1510 - M107 4.182	F1539 - T135	T853 - N68 3.358	K1545 - M110 4.798
T855 - R65 3.498	S1508 - R106 4.129	L1510 - S108 4.556	3.126	Q1537 - Q115 4.640	K1545 - L111 1.856
A1503 - R106 4.637	Y1522 - L161 3.704	T1547 - M107 4.667	F1539 - V150	Q1537 - N117 3.114	K1545 - W112
A1504 - R106 4.059	Y1522 - L163 4.100	T1547 - S108 1.339	3.501	Q1537 - P131 4.850	4.597
P1506 - R104 3.952	Y1527 - Q115 2.715	T1547 - S109 3.019	Q1592 - S176	Q1537 - S132 4.248	K1545 - T113 3.123
T1507 - R104 3.693	Y1527 - N117 2.525	V1670 - P182 4.115	3.783	Q1537 - S133 4.953	K1545 - A137 3.494
R1521 - K163 3.849	K1589 - P182 3.185	V1670 - Y184 3.721	Q1592 - P177	E1538 - Q115 4.322	K1545 - P138 4.849
V1541 - K163	K1589 - L183 3.128	V1670 - Y192 2.473	3.299	E1538 - S133 2.546	T1549 - R104 3.618
4.594	S1546 - S108 4.595		Q1592 - V178	E1538 - V150 3.075	T1549 - A105 4.541
P1542 - K163 3.556	S1546 - S109 3.257		2.123	E1538 - H151 3.103	T1549 - R106 2.646
G1543 - K163	A1548 - S108 3.298		Q1592 - Q179	E1538 - G152 3.364	T1549 - M107 4.832
4.847	A1548 - T113 4.738		4.093	E1538 - P159 4.334	T1549 - T113 3.190
S1544 - S109 3.851	P1668 - L183 4.847		Q1594 - S176	T1540 - T135 4.108	T1549 - F114 4.032
I1550 - Q115 3.846	P1668 - Y192 3.510		4.535	T1540 - T148 2.407	T1549 - Q115 3.742
G1552 - Q115	L1669 - P182 4.495		Q1594 - P177	T1540 - T149 4.894	S1551 - G103 4.940
4.205	L1669 - Y192 4.700		1.968	T1540 - V150 3.117	S1551 - R104 3.493
M1593 - Q179	Q1671 - Q179 3.363		Q1594 - V178		S1551 - F114 3.491
2.216	Q1671 - Y184 2.427		4.110		S1551 - Q115 2.316
T1672 - Y192 4.567			Q1594 - Q179		S1551 - V116 4.347
			4.156		S1551 - N117 4.631

Range: V776-T855, A1503-Q1594

Fibronectin: Rv3717 (107 interactive Receptor: Ligand binding residues)				
Single interaction (24 Bonds)	**Double interactions (20 Bonds)**	**Triple interactions (14 Bonds)**	**Quadruple interactions (12 Bonds)**	**More than 4 interactions (37 Bonds)**
L9- G67 4.557	L10- S66 3.302	I102- L145 4.462	I570- S16 3.940	R108- A110 4.384
E68- T54 4.917	L10- G67 3.723	I102- N146 4.344	I570- A17 3.270	R108- N111 2.512
G105- N111 4.647	L11- S66 3.952	I102- Q193 4.756	I570- S18 2.719	R108- N114 2.724
I109- Q193 4.620	L11- G67 4.197	A104- P48 4.956	I570- T20 4.686	R108- A115 3.723
T112- A147 3.602	L12- N65 4.114	A104- N111 3.133	S1362- T92 3.973	R108- L191 4.827
A114- A147 4.039	L12- S66 4.982	A104- L191 3.119	S1362- A93 1.556	R108- Q193 3.072
E459- R158 2.822	G103- G50 4.277	T571- S16 3.514	S1362- L94 4.028	R457- R158 4.911
R546- S16 3.948	G103- L191 3.199	T571- A17 3.949	S1362- L116 3.595	R457- R161 3.564
F569- A17 4.075	S110- N146 4.176	T571- S18 4.353	T1371- L79 2.540	R457- D162 1.763
D650- T20 4.214	S110- Q193 4.759	G572- S16 2.892	T1371- R80 3.791	R457- Q163 3.967
P674- P21 3.865	F672- A87 3.783	G572- A17 4.745	T1371- R82 4.797	R457- Q165 3.053
A697- S16 3.032	F672- L88 3.425	G572- S18 1.513	T1371- A83 2.252	R457- A166 3.177
S698- S16 3.423	V1356- N86 4.700	L675- P21 4.898		E1358- L85 4.764
D699- S16 3.887	V1356- A87 4.750	L675- A231 4.145		E1358- N86 2.659
E839- R134 4.303	A1357- N86 3.293	L675- G232 4.746		E1358- A87 4.476
D1326- R80 2.525	A1357- A87 3.369	N840- R134 2.743		E1358- G89 4.262
T1341- R91 4.641	D1365- R82 3.912	N840- G168 3.576		E1358- V90 3.307
N1342- R91 2.166	D1365- G97 4.689	N840- P170 4.048		E1358- R91 4.199
L1344- R91 4.406	V1369- L79 4.647	Y1347- A83 3.657		L1359- R82 3.475
P1363- L116 4.922	V1369- R82 2.858	Y1347- N86 4.238		L1359- A83 4.660
V1368- R82 3.256		Y1347- A87 4.571		L1359- N86 0.384
L1370- R82 4.878				L1359- V90 4.828
G1388- R91 4.211				L1359- T92 3.659
T1407- E217 3.574				S1360- N86 4.139
				S1360- V90 4.821
				S1360- R91 3.003
				S1360- T92 1.552
				S1360- A93 3.772
				I1361- R82 4.993
				I1361- T92 3.631
				I1361- A93 4.629
				I1361- L94 4.998
				N1372- R80 2.653
				N1372- L81 4.770
				N1372- A83 3.485
				N1372- A84 3.342
				N1372- Q220 2.572

Range: L9-A114, R457-E839, D1326-T1407

Fibronectin-Rv3811 (102 interactive Receptor: Ligand binding residues)										
Single interaction (18 Bonds)		**Double interactions (20 Bonds)**		**Triple interactions (21 Bonds)**		**Quadruple interactions (20 Bonds)**		**More than 4 (23 Bonds)**	
R3 - P534	3.729	H1438 - T187	4.624	R1436 - P426	4.396	R1434 - E427	3.517	F1442 - T184	3.827
G4 - P534	4.479	H1438 - H530	2.423	R1436 - E427	4.314	R1434 - A428	4.726	F1442 - P185	4.570
H1437 - T187	4.034	P1439 - T184	2.653	R1436 - A428	4.541	R1434 - D429	4.989	F1442 - T187	2.376
S1443 - P527	4.184	P1439 - T187	4.507	R1450 - K399	4.727	R1434 - R435	2.649	F1442 - A188	4.065
G1448 - P426	3.437	G1444 - A188	4.700	R1450 - A400	2.758	T1464 - Q182	4.471	F1442 - H530	4.987
P1465 - Q182	4.254	G1444 - H530	4.179	R1450 - L401	4.860	T1464 - W183	1.338	R1445 - T187	3.120
G1466 - Q182	2.384	P1446 - T187	3.728	N1462 - W183	3.670	T1464 - T184	3.076	R1445 - A188	3.354
E1468 - Q182	4.903	P1446 - H530	3.657	N1462 - P185	2.920	T1464 - P185	3.913	R1445 - V189	4.117
V1470 - T532	4.938	L1463 - W183	4.121	N1462 - Y269	4.932	E1481 - Q427	4.244	R1445 - L397	3.498
V1474 - D429	4.711	L1463 - P185	4.303	T1467 - Q182	3.473	E1481 - A428	4.690	R1445 - T424	2.622
S1482 - D429	4.385	L1484 - F531	4.646	T1467 - W183	3.973	E1481 - D429	2.462	R1445 - S425	2.696
R1521 - D75	4.252	L1484 - L533	4.121	T1467 - T184	4.769	E1481 - R435	2.296	R1445 - P426	1.458
V1536 - D50	3.139	I1486 - T532	3.001	Y1469 - W183	4.179	P1573 - R73	3.519	R1447 - W183	4.845
T1540 - P74	3.998	I1486 - L533	4.162	Y1469 - T184	4.418	P1573 - P74	4.728	R1447 - P185	2.024
V1566 - P74	3.451	E1538 - K72	4.374	Y1469 - P185	2.882	P1573 - D75	3.398	R1447 - P186	3.117
D1571 - R73	3.096	E1538 - P74	4.865	Y1522 - P74	2.978	P1573 - D234	4.742	R1447 - T187	2.913
S1575 - P74	2.531	Y1564 - P52	3.304	Y1522 - D75	2.287	K1577 - Q176	2.700	R1447 - A188	4.978
P1578 - Q176	3.987	Y1564 - T121	4.721	Y1522 - G76	4.792	K1577 - A177	2.679	R1479 - G403	3.915
		S1572 - R73	3.373	R1524 - R73	4.696	K1577 - P178	4.446	R1479 - G404	3.299
		S1572 - D234	2.848	R1524 - P74	3.670	K1577 - P179	4.639	R1479 - A405	2.467
				R1524 - T121	4.799			R1479 - Q408	4.728
								R1479 - R435	3.590
								R1479 - W446	3.097

Range: R1436- P1578

Supplementary table 4. Collagen binding to *Mycobacterium tuberculosis* protein interactive Receptor to ligand binding residues showing the groups of interactive bonds with RMSD scores.

Collagen:GST (83 interactive Receptor: Ligand binding residues)									
Single interaction (14 Bonds)		**Double interactions (16 Bonds)**		**Triple interactions (12 Bonds)**		**Quadruple interactions (12 Bonds)**		**More than 4 interactions (29 Bonds)**	
H5 - M94	4.895	P428 - S112	4.727	G426 - S112B	4.772	K434 - R35B	3.998	Y672 - W41	3.347
S430 - Q207	4.759	P428 - K113	4.876	G426 - D114	3.158	K434 - D36B	3.236	Y672 - P53	4.820
A431- Q207	2.257	D429 - Y111	4.786	G426- L118	4.806	K434 - D39B	2.757	Y672 - N54	2.077
P432 - Q207	3.094	D429 - K113	4.089	R427 - R108	3.093	K434 - R42B	4.703	Y672 - L55	2.518
G433 - Q207	2.838	Y672 - Y7	4.999	R427 - S112	4.322	P678A - D101	2.620	Y672 - T111	3.572
V646 - P86	3.875	Y672 - W8	2.638	R427 - D114	3.866	P678A - G105	4.451	Y672 - G205	3.792
T665 - D114	4.634	G680 - K125	3.812	P679 - R108	4.758	P678A - R108	2.871	P673 - Y7	2.531
T671 - W8	4.254	G680 - E128	4.562	P679 - K125	2.857	P678A - M129	4.657	P673 - W8	3.852
F681 - K125	3.327	L1482 - G84	4.984	P679 - M129	4.682	I1585 - H139	3.485	P673 - I10	2.161
D682 - E128	2.732	L1482 - K85	4.703	R1682 - K138	4.227	I1585 - K140	4.749	P673 - G12	4.520
I1597 - D146	4.830	Y1599 - K138	4.764	R1682 - H139	1.795	I1585 - G145	4.865	P673 - L13	4.594
I1680 - H139	4.575	Y1599 - H139	2.532	R1682 - K140B	4.506	I1585 - D146	2.827	P673 - S107	4.768
Q1583 - H139	3.960	R1677 - K138	4.466					P673 - T111	2.335
Q1684 - D14	4.260	R1677 - H139	1.958					P673 - L203	4.938
		Q1678 - H139	3.270					P673 - Q204	3.315
		Q1678 - K175	3.669					P673 - G205	3.186
								G674 - Y7	4.597
								G674 - L13	3.483
								G674 - R103	3.981
								G674 - T104	4.045
								G674 - S107	2.282
								G674 - T111	3.750
								G674 - Q204	3.072
								Q1674 - L131	4.825
								Q1674 - E134	4.731
								Q1674 - D135	3.491
								Q1674 - K138	3.907
								Q1674 - H139	4.958
								Q1674 - K175	3.570

Range: S430-D682, I1585-Q1684

Collagen: Rv2272 (97 interactive Receptor: Ligand binding residues)

Single interaction (23 Bonds)	Double interactions (22) Bonds)	Triple interactions (18 Bonds)	Quadruple interactions (12 Bonds)	More than 4 (22 Bonds)
H5 - R87 3.291	L4 - R86 3.942	R50 - 35 3.470	L8 - R83 4.654	G49 - A38 4.424
I6 - R87 3.085	L4 - R87 2.422	R50 - 36 4.823	L8 - P92 4.166	G49 - A39 3.575
S12 - L100 2.765	V7 - R83 2.783	R50 - A42 4.439	L8 - H94 1.857	G49 - A42 3.044
F13 - L100 4.984	V7 - R87 4.241	Q1476 - F18 2.693	L8 - T96 3.966	G49 - L43 4.489
C47 - V47 4.286	M9 - H94 4.602	Q1476 - M85 4.863	C11 - T96 2.695	G49 - V47 4.846
D51 - L46 3.180	M9 - T96 3.858	Q1476 - L91 2.254	C11 - P97 4.086	P1548 - W29 4.468
L258 - V109 3.090	R10 - T96 4.346	M1552 - T32 1.489	C11 - Y99 3.675	P1548 - M73 3.844
E261 - V106 4.359	R10 - L100 3.724	M1552 - L36 4.053	C11 - L100 2.893	P1548 - R77 3.635
P263 - V106 4.772	P46 - V47 2.498	M1552 - T70 4.631	R1560 - R17 4.434	P1548 - P97 4.673
R563 - Q45 4.860	P46 - P48 3.723	E1557 - R24 3.405	R1560 - A21 3.478	P1548 - G98 3.815
P647 - A9 4.543	G48 - L43 4.277	E1557 - T25 3.762	R1560 - R24 2.263	P1548 - A101 4.554
H1471 - R93 4.724	G48 - V47 4.363	E1557 - A28 4.577	R1560 - T25 4.720	M1551 - W29 4.776
S1472 - R93 3.244	V260 - V102 4.064	P1561 - A21 4.829		M1551 - L66 3.809
Q1473 - R93 3.088	V260 - V106 2.526	P1561 - T25 2.297		M1551 - T70 4.563
A1546 - R93 4.309	T1474 - L91 3.869	P1561 - R77 2.445		M1551 - M73 4.153
A1547 - R77 3.753	T1474 - R93 4.431	Y1562 - T25 4.937		M1551 - A101 4.978
P1550 - W29 3.826	D1475 - F18 4.548	Y1562 - W29 2.296		M1551 - C105 1.983
L1554 - W29 4.580	D1475 - L91 2.556	Y1562 - R77 3.017		A1558 - R24 4.862
E1556 - R31 3.149	L1549 - W29 4.839			A1558 - T25 3.689
E1624 - R93 3.444	L1549 - M73 3.953			A1558 - A28 2.685
D1625 - R93 2.467	S1555 - A28 4.851			A1558 - W29 4.860
R1627 - R93 3.700	S1555 - R31 3.635			A1558 - R77 4.909
S1673 - Y99 4.382				

Range: L4-D51, L258-P263, H1471-S1673

Collagen: Rv2598 (73 interactive Receptor: Ligand binding residues)				
Single interaction (14 Bonds)	**Double interactions (16 Bonds)**	**Triple interactions (9 Bonds)**	**Quadruple interactions (8 Bonds)**	**More than 4 (26 Bonds)**
H643 - T113 2.658	P647 - P134A 4.415	G645 - A36 4.815	V646 - G133 3.278	H649 - D135 3.716
Y1462 - M1 3.850	P647 - 136A 4.906	G645 - T38 4.734	V646 - P134 2.678	H649 - G136 3.940
L1470 - R111 3.984	P1481 - T110A 3.461	G645 - P134 3.790	V646 - D135 4.681	H649 - R140 3.142
T1479 - R111 4.692	P1481 - R111A 4.011	C1588 - R103 4.501	V646 - R140 4.285	H649 - R161 4.869
C1480 - R111 4.658	I1585 - R100A 4.170	C1588 - R106 3.536	Y1689 - M1 3.611	H649 - R163 3.176
G1483 - M1 3.695	I1585 - R103A 2.604	C1588 - D124 3.157	Y1689 - L3 2.831	L1482 - M1 3.440
P1572 - M1 3.338	P1587 - R103A 3.410	L1595 - E107 3.753	Y1689 - H4 3.057	L1482 - P2 3.342
S1584 - R103 4.478	P1587 - D124A 3.774	L1595 - T110 2.802	Y1689 – Q5 3.151	L1482 - H33 4.930
P1586 - R103 3.602	Q1590 - G122A 4.430	L1595 - R111 2.249		L1482 - T110 3.425
P1589 - R106 4.115	Q1590 - D124A 4.667			L1482 - R111 3.985
T1591 - Q5 3.893	W1592 - Q5A 3.741			L1482 - A112 2.203
I1597 - E107 1.928	W1592 - R106A 2.083			L1482 - T113 3.098
Q1624 - R111 3.623	W1596 - E107A 3.320			L1482 - G114 4.364
V1687 - T110 4.931	W1596 - 111A 2.682			L1482 - W115 4.928
	Q1684 - R103A 2.284			R1593 - M1 2.358
	Q1684 - E107A 3.084			R1593 - P2 4.174
				R1593 - L3 2.765
				R1593 - Q5 4.372
				R1593 - R106 4.130
				R1593 - C109 4.944
				R1593 - T110 3.466
				S1594 - L3 4.241
				S1594 - R103 3.921
				S1594 - R106 3.720
				S1594 - G107 1.909
				S1594 - T110 2.363

Range: H643-H649, L1470- V1687

Collagen: Rv2730 (101 interactive Receptor: Ligand binding residues)				
Single interaction (17 Bonds)	**Double interactions (22 Bonds)**	**Triple interactions (15 Bonds)**	**Quadruple interactions (24 Bonds)**	**More than 4 interactions (15 Bonds)**
L4 - D94 4.990	M9 - H42 4.872	G680 – T18 3.609	P678 – C15 4.792	L5 – P63 4.560
L8 - S91 4.895	M9 - Q56 3.690	G680 – C40 4.266	P678 – R38 3.712	L5 – R65 2.394
422 - R19 3.101	H643 - Q99 4.372	G680 – H42 2.537	P678 – N39 4.425	L5 – H67 4.357
G423 - R19 4.995	H643 - G100 4.670	R682 - T18 4.605	P678 – C40 1.040	L5 – T93 2.694
I424 - R19 3.480	H649 - G62 4.519	R682 - A20 4.241	T1479 – Q99 4.727	L5 – G95 4.990
F681 - H42 3.629	H649 - P63 3.530	R682 - H42 4.789	T1479 – G100 3.181	P679 – C15 4.007
P691 - S23 4.127	G674 - T11 3.506	S1584 - N73 3.214 S1584 - S84 4.365	T1479 – H101 4.839	P679 – A16 4.319
L1470 - T103 4.231	G674 - K13 4.047	S1584 - T86 3.453	T1479 – H102 4.273	P679 – T18 4.526
P1478 - D102 4.885	K688 - S21 3.063	L1595 - T103 2.608	L1482 – G100 4.760	P679 – C40 3.114
C1480 - G100 4.373	K688 - S22 3.721	L1595 - H104 3.566	L1482 – H101 3.251	P679 – H42 3.919
P1481 - T103 3.696	I1585 - P87 3.845	L1595- Q107 3.356	L1582 – N73 4.027	G1678 – G52 4.954
R1593 - H104 4.666	I1585 - R110 4.951	Q1674 - A54 3.024	L1582 – S84 4.835	G1678 – S53 4.929
G1598 - R110 3.797	S1594 - H104 2.759	Q1674 - R69 2.050	Q1583 – A71 4.351	G1678 – A54 4.675
D1625 - R110 4.901	S1594 - Q107 1.378	Q1674 - T88 3.282	Q1583 – S72 3.828	G1678 – A71 3.120
K1656 - S75 4.818	W1596 - T103 3.828		Q1583 – N73 1.842	G1678 – T88 4.332
V1685 - Q107 4.101	W1596 - Q107 2.371		Q1583 – T86 2.211	
T1689 - G154 4.731	R1677 - T88 2.380		I1597 – P87 4.532	
	R1677 - S89 3.708		I1597 - R106 4.798	
	Q1684 - Q107 3.403		I1597 - Q107 3.176	
	Q1684 - R111 4.166		I1597 - R110 2.268	
	S1690 - G157 4.580		Y1599 – P 87 3.343	
	S1690 - R158 2.527		Y1599 – T88 2.718	
			Y1599 - S89 4.432	
			Y1599 - R110 4.783	
			E1624 - D102 4.228	
			E1624 - T103 4.617	
			E1624 – R106 3.224	
			E1624 - R110 3.024	
			R1682 - T86 3.006	
			R1682 - P87 2.489	
			R1682 - T88 4.673	
			R1682 - R110 2.928	

Range: L4-M9, 422-691, L1470-T1690

Collagen: Rv3209 (100 interactive Receptor: Ligand binding residues)				
Single interaction (7 Bonds)	**Double interactions (20 Bonds)**	**Triple interactions (18 Bonds)**	**Quadruple interactions (12 Bonds)**	**More than 4 interactions (43 Bonds)**
P428 - R159 3.938	I6 - R104 3.594	L4 - L96 3.779	M1 - D129 3.503	H5 - L96 4.703
D429 - R159 4.445	I6 - A169 4.994	L4 - V102 3.928	M1 - W136 3.817	H5 - T100 2.931
S430 - R159 4.055	V7 - G168 4.832	L4 - M139 2.871	M1 - T137 2.681	H5 - V101 4.234
A431 - R159 4.931	V7 - A169 2.582	L8 - T167 3.560	M1 - L138 2.756	H5 - V102 1.749
G657 - L111 2.971	M9 - G168 4.142	L8 - G168 3.175	W2 - R104 4.580	H5 - R104 2.903
L658 - L111 2.920	M9 - Q170 3.793	L8 - A169 3.822	W2 - P135 3.275	H5 - M139 3.512
G680 - R104 4.876	P650 - D129 4.495	V646 - A122 3.357	W2 - W136 3.570	H5 - V165 4.188
	P650 - L138 3.001	V646 - V141 4.667	W2 - T137 2.573	H5 - N166 3.043
	G651 - D129 3.740	V646 - N143 3.275	V670 - L156 4.229	H5 - T167 3.496
	G651 - K138 4.154	R653 - N131 3.292	V670 - T157 3.714	R427 - G107 3.951
	D656 - L111 4.140	R653 - V132 4.218	V670 - R159 3.530	R427 - T108 2.627
	D656 - L112 4.960	R653 - S133 2.592	V670 - I180 3.952	R427 - R159 2.396
	N669 - Y157 3.862	C668 - Q110 3.525		R427 - L160 4.173
	N669 - R159 2.644	C668 - T157 4.039		R427 - N161 4.278
	T671 - L156 4.531	C668 - R159 3.118		R427 - N178 2.435
	T671 - I180 2.592	P673 - N177B 4.887		R427 - A179 4.781
	Y672 - I180 2.869	P673 - A179B 3.965		H649 - Y119 3.052
	Y672 - I181 2.048	P673 - I181B 4.315		H649 - D129 4.331
	P679 - R104 4.552			H649 - K138 2.382
	P679 - S163 4.996			H649 - M139 3.323
				H649 - V140 2.331
				V652 - D129 4.195
				V652 - F130 2.986
				V652 - N131 3.415
				V652 - V132 3.293
				V652 - S133 3.155
				V652 - W136 2.636
				G674 - L160 4.874
				G674 - S175 3.539
				G674 - N177 3.015
				G674 - A179 3.534
				G674 - I181 2.226
				G674 - A182 2.995
				G674 - T183 3.827
				P678 - R104 3.717
				P678 - V105 4.824
				P678 - T106 2.810
				P678 - N161 4.443
				P678 - C162 2.973
				P678 - S163 2.517
				P678 - T171 4.477
				P678 - A174 4.587
				P678 - T176 4.775

Range: M1-L8, D429-G680

Collagen: Rv3491 (110 interactive Receptor: Ligand binding residues)

Single interaction (11 Bonds)		Double interactions (16 Bonds)		Triple interactions (18 Bonds)		Quadruple interactions (20 Bonds)		More than 4 interactions (45 Bonds)	
N1177 - Q169	3.046	F916 - G32	2.207	G1178 - S41	4.876	L1509 - P174	3.583	F915 - L32	2.710
L1506 - P175	4.121	F916 - P173	4.394	G1178 - Q169	4.968	L1509 - P175	1.569	F915 - G33	3.865
A1510 - P177	4.052	G1175 - N43	3.663	G1178 - I171	3.545	L1509 - S176	1.958	F915 - D37	4.852
A1575 - P175	4.593	G1175 - Q169	4.337	L1463 - P182	3.850	L1509 - P177	3.510	F915 - V178	4.741
A1577 - T87	4.110	L1467 - S176	3.452	L1463 - L183	2.807	H1579 - T88	3.827	F915 - R180	3.181
P1586 - Q91	4.316	L1467 - P177	3.817	L1463 - Y192	4.160	H1579 - P89	2.128	F915 - T181	4.919
P1587 - Q91	2.593	G1508 - P175	3.248	V1576 - N53	4.759	H1579 - Q90	3.390	P917 - A29	3.716
C1588 - Q91	4.839	G1508 - S176	2.304	V1576 - V85	3.513	H1579 - Q91	4.183	P917 - R30	2.684
T1591 - S102	3.786	Q1581 - P89	2.607	V1576 - T87	4.565	S1580 - P89	3.510	P917 - A31	4.908
S1618 - P175	3.152	Q1581 - Q90	2.553	V1578 - T87	3.921	S1580 - Q90	3.017	P917 - L32	4.143
N1660 - Q90	2.499	D1582 - Q90	4.994	V1578 - Q91	3.050	S1580 - Q91	4.379	P917 - L170	3.314
		D1582 - I92	4.237	V1578 - N100	4.296	S1580 - I92	3.690	L1176 - S41	2.997
		P1619 - P175	4.362	Q1590 - S102	4.468	P1589 - Q91	4.672	L1176 - F42	3.879
		P1619 - S176	4.934	Q1590 - N117	4.508	P1589 - N100	3.818	L1176 - N43	2.579
		V1655 - Q90	3.330	Q1590 - A119	4.144	P1589 - F101	4.287	L1176 - V47	4.227
		V1655 - I92	4.938	K1688 - S102	3.318	P1589 - S102	4.970	L1176 - S48	4.166
				K1688 - G103	3.868	W1592 - V84	4.807	L1176 - G49	4.216
				K1688 - R104	4.068	W1592 - V85	2.618	L1176 - Q169	3.178
						W1592 - F101	4.433	G1465 - L57	4.926
						W1592 - S102	2.331	G1465 - P177	3.942
								G1465 - V178	3.039
								G1465 - Q179	2.558
								G1465 - P182	4.094
								F1466 - T55	3.264
								F1466 - L57	3.253
								F1466 - N83	2.986
								F1466 - P177	4.697
								F1466 - V178	2.817
								F1466 - Q179	4.796
								Q1574 - T53	2.860
								Q1574 - I54	4.557
								Q1574 - T55	2.352
								Q1574 - N83	4.513
								Q1574 - V85	4.849
								Q1574 - P174	4.422
								Q1574 - P175	3.330
								Q1574 - V178	4.862
								M1617 - S39	2.387
								M1617 - Y40	4.484
								M1617 - T51	2.155
								M1617 - W52	4.141
								M1617 - T53	2.008
								M1617 - T87	4.400
								M1617 - I171	4.353
								M1617 - P175	4.298

Range: F915-F916, G1175-M1617

Collagen: Rv3717 (106 interactive Receptor: Ligand binding residues)				
Single interaction (14 Bonds)	**Double interactions (12 Bonds)**	**Triple interactions (18 Bonds)**	**Quadruple interactions (36 Bonds)**	**More than 4 interactions (26 Bonds)**
W2 - N23 4.542	I6 - T20 4.669	M9 - S18 4.763	P673 - R117 4.180	H5 - N23 3.558
L8 - P21 3.112	I6 - N23 4.461	M9 - A19 3.882	P673 - I148 4.410	H5 - A25 4.107
R10 - T20 4.948	V7 - T20 4.317	M9 - P21 4.231	P673 - Q193 3.828	H5 - A87 3.927
C11 - S18 2.327	V7 - P21 3.185	H649 - N86 2.764	P673 - Y194 3.690	H5 - L88 3.699
V646 - A83 4.759	S12 - S16 4.966	H649 - A87 4.763	G674 - R117 4.334	H5 - G89 4.576
T671 - R117 4.794	S12 - S18 3.874	H649 - V90 4.835	G674 - I148 4.165	Y672 - N114 4.323
F681 - T20 2.491	G686 - S16 4.021	L1482 - R80 3.465	G674 - Q193 4.804	Y672 - A115 4.862
P684 - A17 4.466	G686 - A17 4.935	L1482 - A216 3.457	G674 - Y194 3.659	Y672 - R117 3.027
P685 - A17 4.190	C1588 - R134 2.872	L1482 - E217 2.453	P679 - N23 4.865	Y672 - Q193 3.953
P1481 - P217 3.192	C1588 - A208 3.966	P1587 - R134 3.132	P679 - G238 2.718	Y672 - Y194 3.641
G1483 - E217 4.547	Q1684- G168 3.786	P1587 - G168 4.664	P679 - Q239 4.414	P678 - N23 3.714
P1589 - A208 3.952	Q1684 - L221 3.184	P1587 - P170 4.551	P679 - R241 3.772	P678 - M27 3.075
L1595 - E217 4.515		T1591 - P207 3.421	G680 - T20 4.060	P678 - Q237 4.655
R1682 - A166 4.473		T1591 - A208 3.961	G680 - P21 4.750	P678 - G238 3.162
		T1591 - A211 4.069	G680 - A22 4.513	P678 - Q239 4.947
		P1689 - L211 4.591	G680 - N23 4.489	P678 - R241 3.895
		P1689 - E214 4.477	D682 - A17 4.772	R1593 - A211 3.792
		P1689 - S215 4.919	D682 - S18 3.645	R1593 - L212 3.905
			D682 - A19 4.675	R1593 - S215 3.214
			D682 - T20 3.551	R1593 - A216 4.565
			I1585 - Q165 4.026	R1593 - E217 2.299
			I1585 - A166 2.626	I1597 - A166 4.219
			I1585 - S167 2.842	I1597 - S167 3.908
			I1585 - G168 2.446	I1597 - G168 4.906
			Q1590 - A131 4.844	I1597 - K221 3.061
			Q1590 - N206 3.772	I1597 - R228 4.577
			Q1590 - P207 3.754	
			Q1590 - A208 3.507	
			W1592 - P207 4.896	
			W1592 - A208 4.006	
			W1592 - A211 3.612	
			W1592 - L212 4.492	
			S1594 - L212 3.042	
			S1594 - E217 3.741	
			S1594 - G218 4.600	
			S1594 - K221 3.430	

Range: W2-C11, V646-P685, P1481-P1689

Collagen-Rv3811 (97 interactive Receptor: Ligand binding residues)				
Single interaction (22 Bonds)	**Double interactions (28 Bonds)**	**Triple interactions (9 Bonds)**	**Quadruple interactions (12 Bonds)**	**More than 4 (26 Bonds)**
S3 – P534 4.835	D188 – H20 3.633	M1 – R470 3.543	P644 – T525 3.622	V646 – Y436 3.948
H5 – S467 3.688	D188 – E21 2.430	M1 – L533 2.501	P644 – V526 4.247	V646 – T438 3.654
Q186 – H20 3.653	V627 – H20 4.021	M1 – P534 4.832	P644 – P527 3.858	V646 – Y460 3.480
P626 – H20 4.037	V627 – P22 4.631	R641 – E194 3.043	P644 – P528 4.012	V646 – Y468 3.031
G628 – H20 3.214	P1449 – E213 3.675	R641 – P196 4.786	D1452 – P215 4.214	V646 – P528 3.370
E640 – Q194 4.929	P1449 – T214 4.064	R641 – A197 4.363	D1452 – Y217 4.166	P647 – Y436 2.466
G642 – Q194 3.418	P1453 – Y217 2.783	G645 – V526 4.796	D1452 – G278 4.868	P647 – V526 4.767
H643 – T525 4.663	P1453 – L282 1.933	G645 – P527 4.381	D1452 – K282 4.176	P647 – P527 3.299
I1450 – T214 2.853	P1455 – T281 4.870	G645 – P528 2.679	T1479 – P520 4.379	P647 – P528 1.446
G1451 – T214 4.973	P1455 – K282 3.539	P650 – D429 3.214	T1479 – P521 3.345	P647 – E529 2.136
G1454 - L282 3.769	F1458 – T281 4.108	P650 – R470 4.665	T1479 – S522 3.603	H649 – A428 4.073
H1471 - R519 4.441	F1458 – M328 4.163	P650 – F531 4.392	T1479 – G523 4.293	H649 – D429 1.039
Q1473 - R519 4.857	S1472 – R519 3.009	G651 – D429 2.916		H649 – A430 2.949
Q1476 – S522 3.674	S1472 – S522 4.608	G651 – A430 4.360		H649 – A431 4.418
P1481 – P520 4.874	T1474 – R519 4.757	G651 – L533 4.573		H649 – Y436 2.661
L1482 – E461 3.260	T1474 – S522 4.913	W1592 – R500 4.783		H649 – S467 4.633
P1587 – T502 3.560	E1477 – S522 3.147	W1592 – L501 2.705		H649 – Y468 4.288
P1589 – L501 4.131	E1477 – G523 4.531	W1592 – T502 4.593		H649 – E469 1.456
Q1590 – L501 4.497	P1478 – R519 4.124	W1594 – T502 2.196		H649 – R470 3.165
T1591 – L501 4.959	P1478 – S522 3.267	S1594 – G503 3.720		K1456 – D218 4.943
L1595 – T518 3.685	I1585 – N504 4.458	S1594 – N504 3.697		K1456 – R219 4.107
E1624 – R519 2.790	I1585 – P515 4.427	Q1684 – T502 3.069		K1456 – G220 4.424
	C1588 – L501 3.654	Q1684 – G503 4.981		K1456 – T281 4.288
	C1588 – T502 3.740	Q1684- N504 2.661		K1456 – K282 4.462
	R1593 – L501 4.802			K1456 – P283 2.955
	R1593 – T502 4.043			K1456 – D329 4.171
	I1597 – N504 3.845			
	I1597 – A517 4.519			

Range: P626-H649, P1449-E1624

B4. Chapter six supplementary data

B4.1. Immunoinformatic supplementary data

Supplementary table 1. Predicted B-cell epitope (continuous) sequences of the Conserved hypothetical proteins using Bepipred, predicted antigenic peptide, and ABCpred prediction tools.

Protein	Predicted B-cell Epitope sequences		
	Bepipred 2.0	**Predicted antigenic peptides**	**ABCpred**
Rv3811	ANRPPASSHEPSPTPNTQLAEQPLIGLGGGVT VRELTQDT, PIDAPITQPPAGRPPNDLLDSGLGYRPATKEQP FGQNISAILISPPQAPPGTQWTPPTAVTMAGQ P, SRAEWGADESLRCETPEYD, NDYSPLE, GLTKPVEG, DVAPTP, LGMDDVDPRSMVDLQSAGSSYTTFPGGAI, NTDCPGNAA, AVMDEIRDIAAHFNDPPEELIKALEGGAIYQR WQALGGMNSALGAPTSPEADAADGA, ASQSYERGPLG, PTSAEIQEPLQIT, VDGITTPLATRPPSGPTVPPEHFTLPT	TVVIVAW, PASSHEP, LAEQPLIGLGGGVTVRE, TPFSLVALTG, STDPVFV, TTVQVAVTR, IDAPITQ, NISAILISPPQAP, PPTAVTM, ESLRCET, VRAAVVHH, SPLESAGIVKAIYTYHS, TLGWCDIAYNALVDKYGQVFE, LTKPVEG, DDVAPTPIQIRTVGR, SMVDLQSAG, AIARLPAIFTH, PGNAAYAVM, RDIAAHF, YWSPVTD, PITGAIYEA, GPLGLPTS, EIQEPLQI, NVTEVVDGITT, SGPTVPPEHFTLPT	PGTQWTPPTAVTMAGQ, TEVVDGITTPLATRPP, PEELIKALEGGAIYQR, LPAIFTHRDVGNTDCP, DPAGTDGSVELGGLNP, PGPRSTDPVFVGTTTT, TVVIVAWIANRPPASS, YQTEYETEPRDPAGTD, GAMYWSPVTDAQPITG, TKPVEGFHTGGFNRNT, ESLRCETPEYDRGVRA, HEPSPTPNTQLAEQPL, DAPITQPPAGRPPNDL, GAIYEAWASQSYERGP, KTLGWCDIAYNALVDK, VKAIYTYHSKTLGWCD, RGVRAAVVHHTAGSND, PPAIISRAEWGADESL, RAKRPDGDWGPWYQTE, GDLAGTSARVRAKRPD, YQRWQALGGMNSALGA, AILISPPQAPPGTQWT, AAHFNDPPEELIKALE, YAVMDEIRDIAAHFND, CPGNAAYAVMDEIRDI, LGYRPATKEQPFGQNI, WGVAMIGNFDDVAPTP, RAEWGADESLRCETPE, TTPLATRPPSGPTVPP, AGSSYTTFPGGAIARL, NSALGAPTSPEADAAD, GRLLGWRLGMDDVDPR, SAEIQEPLQITQNFQH,
Rv3717	AATPIISSASATPANI, GANDASIGRQVPTGRGGTKNCQASGTSTNSG YP, SRGNDNALGPCVDE, PLNAIQAGPS, Q, IPKANYIGQDGLYGR, ESAEGRQ	GVLVAAATPIISSAS, IAGMVVFI, LRLRAALNALGVRT, ALGPCVD, RPNAIVSLHA, FHVNYSAPPLNAIQAGPSVQF, DLAGLNLAQYPSILVE, YANALVRGVAGFLAT	TFTWETGLRLRAALNA, TSTNSGYPEHTFTWET, PGHNGANDASIGRQVP, AGMVVFIDPGHNGAND, LVAAATPIISSASATP, TKNCQASGTSTNSGYP, HVNYSAPPLNAIQAGP, IGRQVPTGRGGTKNCQ, AQYPSILVELGNMKNP, ISSASATPANIAGMVV, GGPASGRGFHVNYSAP, GQDGLYGRSDLAGLNL
Rv3491	DPARALGPPPD, SFNQAGVS, PSGTRNMNDYSDPIVW, TTPQQITRTDRLQNFSGR, ADGVACPDGSTAPSSETYAF, CGLQPKLSK, GPPPSPVQRYPL	RCGLAAGAVICSAVALGIALHS, PARALGP, GVSGVTWTITALCDQ, SDPIVWAFNCALNVVSTT, QADGVACPD, HTTVHGAVCGLQPKLSKQPFSLQLIGPPP SPVQRYPLYCNNI	SGTRNMNDYSDPIVWA, TLTGTHTTVHGAVCGL, LQLIGPPPSPVQRYPL, DGSYSFNQAGVSGVTW, TWTITALCDQPSGTRN, ARMSSMLWTFQVNQAD, PSPVQRYPLYCNNIAM, LGIALHSGDPARALGP, CPDGSTAPSSETYAFS
Rv3209	GDSTSTKAIVGAPAPRTVISTSPRPTAPTSTSP HPSPSTLRPQLPPETVTTVAPPGTGPTTVPTRT	GAVATAVII, TKAIVGA, APRTVIST, TSTSPHPSPSTLRPQLPPETVTTVAP, TAAPPQTAVPPPAP, NPRTVVYRV,	AAPPQTAVPPPAPLNP, RTVISTSPRPTAPTST, VAPPGTGPTTVPTRTP, APLNPRTVVYRVTGTK, SLPWTKMVVLNPGVQT, PETVTTVAPPGTGPTT,

Protein	Predicted B-cell Epitope sequences		
	Bepipred 2.0	**Predicted antigenic peptides**	**ABCpred**
	PTAAPPQTAVPPPAPLNP, RGF, GVQTE, QTVVA	QLFDLVNVVYT, RGFPVTDFNVSL, WTKMVVLNP, VQTESVVATSLYSRLNCSIVN	TSPHPSPSTLRPQLPP, TGTKQLFDLVNVVYTD, TAVIINSGDSTSTKAI, TLRPQLPPETVTTVAP, TKAIVGAPAPRTVIST, ESVVATSLYSRLNCSI, TVVASTNNAIIATCTR, VVYTDARGFPVTDFNV
Rv2598	PAPRPL, TG, SRGGDLPESTHAPGY, HTETHDE, FPGDDAA, AAEPDGTGW, SASGGTVV	HQLAIAPVDVSGALLGLVLNA, APRPLATHRL, SALQLGVLGASHVVTVEG, FCEEVSCVA, HAPGYHLQS, WLGGVFP, AALTALAA, TWHLYPS, SGGTVVHT	PESTHAPGYHLQSHTE, EPDGTGWRWRTWHLYP, GWLGGVFPGDDAALTA, RERCTRATGWLGGVFP, YHLQSHTETHDEAAFR,
Rv0934 (PstS1)	LHTLLA, AAAGCGSKPPSGSPETGAGAGTVATTPASSP VTLAETGST, NVTITAQGTGSGAGIAQAAAGTVNIGASDAY LSEGDMAAHKGLMNIALAISAQQVNYNLPG VSEHLKL, AMYQGTIKTWDDPQIAALNPGVNLPGT, VVPLHRSDGSGDTF, TQYLSKQDPEGWGKSPGFGT, ALGENGNGGMVTGCAETPGC, SFLDQASQRGLGEAQLGNSSGNFLLPDAQSIQ AAAAGFASKTPANQAISMI, IVNNRQKDAATAQTLQAFLHWAITDGNKASF , KLSDALIA	RLHTLLAVLTAAPLLLAAAGCGS, GAGTVATTPASSPVTL, GSTLLYPLFNLWG, RYPNVTI, AGIAQAAAGTVNI, MNIALAISAQQVNYNLPGVSEHLKLNGK VLAA, PQIAALNPGVNLPGTAVVPLH, FLFTQYLS, TTVDFPAVPG, MVTGCAETPGCVAYIGISFLDQ, GNFLLPDAQSIQAAAAG, PDGYPIINYEYAIV, TAQTLQAFLHW, SFLDQVHFQPLPPAVVKLSDALIA	AQSIQAAAAGFASKTP, PEGWGKSPGFGTTVDF, QGTIKTWDDPQIAALN, AGCGSKPPSGSPETGA, GSPETGAGAGTVATTP, TQYLSKQDPEGWGKSP,GIAQAAAGTVNIGASD, HWAITDGNKASFLDQV, NQAISMIDGPAPDGYP, TVNIGASDAYLSEGDM, SSPVTLAETGSTLLYP,GDMAAHKGLMNIALAI, YPIINYEYAIVNNRQK,HRSDGSGDTFLFTQYL, LGEAQLGNSSGNFLLP, PGTAVVPLHRSDGSGD, AFHERYPNVTITAQGT,AETGSTLLYPLFNLWG, GPAPDGYPIINYEYAI, AALNPGVNLPGTAVVP,AGTVATTPASSPVTLA,N GGMVTGCAETPGCVA,PGVSEHLKLNGKVLAA,A ATAQTLQAFLHWAIT,GTTVDFPAVPGALGEN, VLAAMYQGTIKTWDDP, AAAGFASKTPANQAIS, AISAQQVNYNLPGVSE, ALGENGNGGMVTGCAE, SFLDQASQRGLGEAQL, GNSSGNFLLPDAQSIQ, GCAETPGCVAYIGISF,GCVAYIGISFLDQASQ, VLTAAPLLLAAAGCGS, VHFQPLPPAVVKLSDA, VKIRLHTLLAVLTAAP, YAIVNNRQKDAATAQT, NKASFLDQVHFQPLPP

Supplementary table 2. B-cell epitope prediction of *Mycobacterium tuberculosis* conserved hypothetical proteins comparing Structure-based Discotope 2.0 and Ellipro.

Protein	Discotope	Ellipro
Rv3811	I11,A12, N13, R14, P15,P16, A17, S18, S19, H20, E21, P22, S23, P24, P74,D75, G76,D77, P93,A94,G95,T96, D97, G98, S99, V100, E101,L102,G103, G104, L105, N106, P107, I131,D132, A140, G141, R142, P143, P144, N145, D146, L147, L148, D149, S150, G151, ,T158, K159,E160,Q161, P162, Q165, P178,P179,G180,T181, Q182,W183,T184,P185, D330, N417, P484, D510, G511, P521, S522, G523, P524, T525, P527, E529, H530,L533, P534, T535, H536, P537, I538, T539 (82 residues)	M1, A2, A3, T4, V5, V6, I7, V8, A9, W10, I11, A12, N13, T532, L533, P534, T535, H536, P537, I538, T539, R14, P15, P16, A17, S18, S19, H20, E21, P22, S23, P24, T25, D146, L147, L148, D149, S150, G151, L152, G153, Y154, K399, L401, E402, G403, G404, A405, I406, Y407, Q408, R409, W410, Q411, A412, L413, G414, G415, N417, S418, A419, L420, G421, P426, E427, A428, D429, A430, A431, D432, G433, A434, R435, K441, M444, W446, S447, P448, V449, T450, D451, A452, Q453, P454, I455, T456, G457, A458, I459, E461, A462, W463, A464, S465, Q466, S467, Y468, E469, R470, G471, P472, L473, G474, L475, P476, T477, S478, A479, E480, I481, Q482, E483, P484, L485, Q486, I487, T488, Q489, N490, F491, Q492, H493, G494, T495, L496, N497, F498, E499, R500, L501, T502, G503, N504, V505, T506, E507, V508, V509, D510, G511, I512, T513, T514, P515, L516, A517, T518, Q33, P34, L35, I36, G37, L38, G39, G40, G41, V42, T43, V44, R45, E46, L47, T48, Q49, D50, T51, P52, F53, S54, G60, D61, L62, A63, G64, T65, R68, R70, A71, K72, R73, P74, D75, G76, D77, W78, G79, P80, W81, D92, P93, A94, G95, T96, D97, G98, S99, V100, E101, L102, G103, G104, L105, N106, P107, G108, P109, V117, G118, T119, T120, T121, T122, V123, Q124, A126, V127, T128, R129, P130, I131, D132, A133, P134, I135, T136, Q137, P138, P139, A140, G141, R142, P143, P144, N145, R155, P156, A157, T158, K159, E160, Q161, P162, F163, G164, Q165, N166, I167, R222, G290, G291, F292, D332, P333, S335, M336, V337, D338, L339, Q340, S341, A342, G343, S344, S345, Y346, T347, T348, F349, P350, G351, G352, A353, I354, A355, R356, L357, A359, V366, G367, R519, P520, P521, S522, G523, P524, A177, P178, P179, G180, T181, Q182 (300 residues)
Rv3717	M1, I2, I13, I14, S15, S16, A17, S18, A19, T20, P21, A38, D40, A41, R45, Q15, R51, G52, G53, T54, S60, N65, S66, G67, Y68, N100, A101, A115, G129, P130, A131, S132, A166, G168, N173, T174, I175, G176, Q177, D178, S215, E217, Q239, A240, R241 (45 residues)	I2, V3, G4, V5, L6, V7, A8, A9, A10, T11, P12, I13, I14, S15, S16, A17, S18, A19, T20, P21, A22, N23, I24, A25, G26, M27, A83, N86, A87, L88, G89, V90, R91, G238, Q239, A240, R241, T62, S63, T64, N65, S66, G67, Y68, P69, E76, D127, G128, G129, P130, A131, S132, G133, R134, G135, F136, S167, G168, I169, P170, K171, A172, N173, Y174, I175, G176, Q177, D178, G179, N203, M204, K205, N206, P207, A208, D209, S210, A211, L212, E214, S215, A216, E217, G218, R219, N36, G37, A38, N39, D40, A41, S42, G44, R45, Q46, V47, P48, T49, G50, R51, G52, G53, T54, K55, N56, C57, R96, G97, N98, D99, N100, A101, L102, G103, P104, C105, D107, E108, N111, N114, A115, L116, P118, A142, P143, P144, L145, N146, A147, I148, Q149, A150, G151, G182, S184, D185, N190, L191, A192, Q193, Y194, P195 (142 residues)
Rv3491	M1, N2, I3, R4, G63, T64, R65, N66, N68, D69, S71, P89, Q90, Q91, N187, I188 (16 residues)	M1, N2, I3, R94, A119, D120, G121, V122, A123, C124, P125, D126, G127, S128, T129, A130, P131, H151, G152, A153, V154, C155, G156, L157, Q158, P159, R4, C5, G6, L7, A8, A9, G10, A11, V12, I13, C14, S15, A16, V17, A18, L19, G20, I21, A22, L23, H24, S25, G26, D27, P28, T88, P89, Q90, Q91, I92, T93, D96, R97, N100, A31, L32, G33, D37, Q60, P61, S62, G63, T64, N66, M67, D69, Y70, S71, D72, P73, I74, V75, W76, I171, G172, P173, P174, P175, S176, P177, V178, Q179, P182, L183, Y184, C185, N186, N187, I188, A189, M190, C191, Y192 (99 residues)

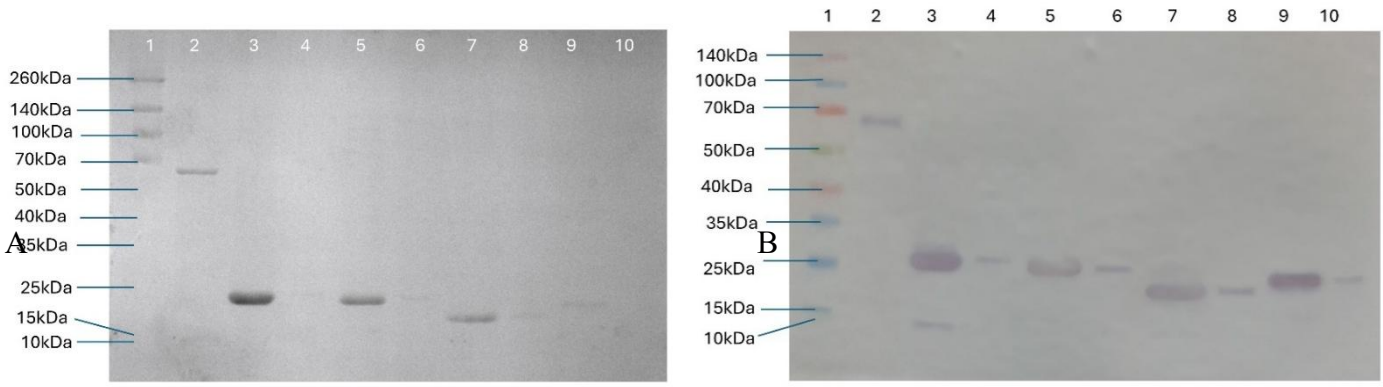
Protein	Discotope	Ellipro
Rv3209	M1, N13, S14, G15, D16, S17, T18, S19, T20, L21, A22, I23, G25, A26, P29, R30, T31, V32, I33, S34, T35, S36, P37, R38, P39, T40, A41, P42, T43, S44, T45, S46, P47, H48, P49, S50, P51, S52, T53, L54, R55, P56, Q57, L58, P59, P60, E61, T62, V63, T64, T65, V66, A67, P68, P69, G70, P73, T74, T75, V76, P77, T78, R79, T80, P81, T82, A83, A84, P85, P86, Q87, T88, A89, V90, P91, P92, P93, A94, P95 (78 residues)	M1, A2, L3, G4, A5, V6, A7, T8, A9, V10, I11, I12, N13, S14, G15, D16, S17, T18, S19, T20, K21, A22, I23, V24, G25, A26, P27, A28, P29, R30, T31, V32, I33, S34, T35, S36, P37, R38, P39, T40, A41, P42, T43, S44, T45, S46, P47, H48, P49, S50, P51, S52, T53, L54, R55, P56, P59, P60, E61, T62, V63, T64, T65, V66, A67, P68, P69, G70, T71, G72, P73, T74, T75, V76, P77, T78, R79, T80, P81, T82, A83, A84, P85, P86, Q87, T88, A89, V90, P91, P92, P93 (91 residue)
Rv2598	T38, D39, G40, P82, G136, T137 (6 residues)	Q5, L6, A7, I8, A9, P10, V11, D12, G49, H89, E91, T92, H93, D94, E95, A96, A97, R99, R100, F120, P121, G122, D123, D124, A125, A126, S149, A150, S151, G152, G153, S14, A16, L17, L18, G19, R69, G72, G73, D74, L75, P76, E77, S78, T79, H80, A81, P82, G83, H85, Q87, P134, D135, G136, T137, G138, R161, R163, P164, R34, A36, H37, T38, D39, G40, S41, A42, E57, S65, V67, S70, R71, H104, E107, R108, A112, T113, G114, N23, A24, P25, A26, P27, F60 (84 residues)
Rv0934 (PstS1)	M1, K2, I3, R4, L5, H6, G25, S26, K27, P28, P29, S30, G31, S32, P33, E34, T35, P204, E205, K209 (20 residues)	M1, K2, I3, R4, L5, H6, T7, L8, L9, A10, V11, L12, T13, A14, A15, P16, L17, L18, L19, A20, A21, A22, G23, C24, G25, S26, K27, P28, P29, S30, G31, S32, P33, E34, T35, G36, A37, G38, A39, G40, T41, V42, A43, T44, T45, P46, A47, S48, S49, P50, V51, T52, L53, A54, N66, L67, G69, P70, A71, F72, H73, E74, R75, Y76, P77, N78, V79, T80, I81, T82, A83, Q84, G85, Q94, A96, A97, G98, T99, V100, N101, A115, A116, H117, N321, R322, Q323, K324, D325, A326, A327, T328, A329, Q330, T331, Q333, A334, H337, W338, T341, D342, K345, A346, S347, F348, L349, D350, A370, I372, S373, S374, N135, L136, P137, G138, V139, S140, E141, H142, L143, K144, L145, N146, K148, V149, A152, M153, Q155, G156, T157, I158, K159, T160, W161, D162, D163, P164, Q165, I166, A167, A168, L169, N170, P171, G172, V173, N174, L175, P176, G177, T178, A179, V180, V181, P182, S186, D203, P204, E205, G206, W207, G208, K209, S210, P211, G212, V217, D218, F219, P220, A221, V222, P223, G224, A225, L226, G227, E228, T236, G237, A239, E240, T241, P242, G243, C244, R259, E263, A264, Q265, N268, S269, S270, G271, N272, F273, L274, L275, D302, G303, A305, P306, D307, G308, S257, Q258, G260, L261 (207 residues)

Supplementary table 3. Summary of ELISA outcomes of *Mycobacterium tuberculosis* conserved hypothetical proteins showing explorative data and performance as a detection biomarker between 50 TB-positive and 50 TB-negative plasma samples.

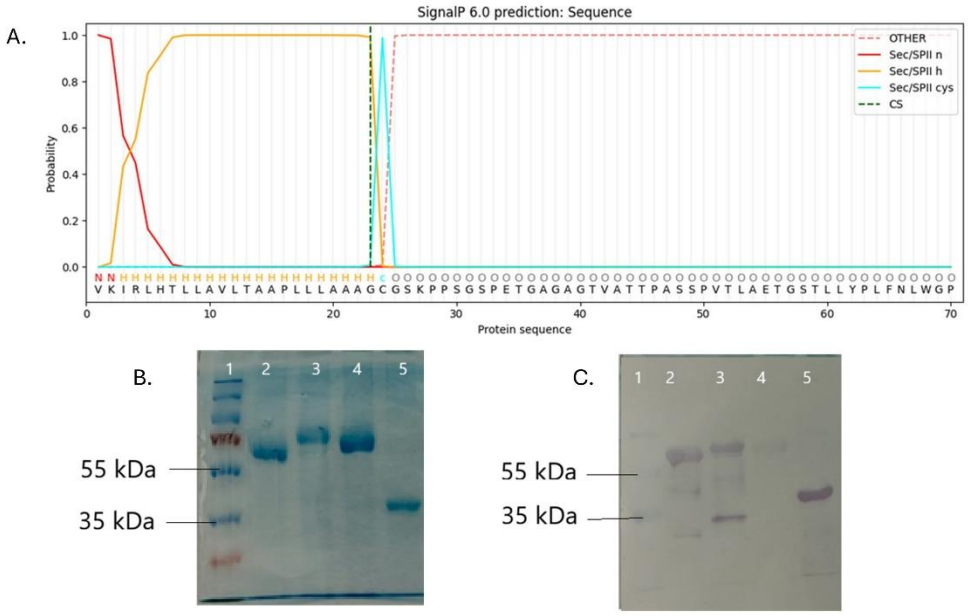
Descriptive analysis	Rv3811		Rv3717		Rv3491	
	TB +	TB-	TB+	TB-	TB+	TB-
Mean	0.590	0.463	0.691	0.655	0.323	0.504
Std. error of mean	0.058	0.067	0.074	0.074	0.047	0.056
95% CI	0.473-0.706	0.327-0.598	0.541-0.839	0.506-0.803	0.229-0.417	0.391-0.616
Median	0.504	0.375	0.554	0.522	0.221	0.383
Range	1.570	3.144	2.278	3.014	1.727	2.248
Minimum	0.098	0.074	0.101	0.116	0.045	0.141
Maximum	1.668	3.144	2.278	3.131	1.772	2.389
IQR	0.531	0.331	0.665	0.586	0.248	0.361
Std. deviation	0.408	0.476	0.525	0.522	0.329	0.396
Descriptive analysis	**Rv3209**		**Rv2598**		**PstS1**	
	TB+	**TB-**	**TB+**	**TB-**	**TB+**	**TB-**
Mean	0.344	0.366	0.542	0.266	1.374	1.102
Std. error of mean	0.038	0.040	0.079	0.022	0.163	0.109
95% CI	0.267-0.421	0.285-0.446	0.382-0.702	0.222-0.309	1.057-1.701	0.883-1.321
Median	0.263	0.304	0.354	0.227	0.829	0.825
Range	1.554	1.442	2.450	0.631	3.405	3.254
Minimum	0.048	0.066	0.102	0.022	0.095	0.170
Maximum	1.602	1.508	2.552	0.653	3.500	3.425
IQR	0.267	0.239	0.520	0.254	2.065	0.945
Std. deviation	0.273	0.283	0.562	0.154	1.149	0.773

Supplementary table 4. A pairwise group comparison designed to evaluate how HIV status affects TB ELISA detection performance.

Group 1	Group 2	outcome
Coinfected: TB+/HIV+ (20)	Uninfected: TB-/HIV- (36)	Is there a relationship between coinfection and uninfected individuals
Coinfected: TB+/HIV+ (20)	TB only: TB+/HIV- (30)	Does HIV affect detection among TB-positive individuals?
Coinfected: TB+/HIV+ (20)	HIV only: TB-/HIV+(14)	Does HIV affect TB detection?
TB only: TB+HIV- (30)	Uninfected: TB-/HIV- (36)	Is the ELISA able to distinguish between TB-positive and TB-negative samples among HIV negative samples?
TB only: TB+HIV- (30)	HIV only: TB-/HIV+ (14)	Is the ELISA able to distinguish between TB-positive and TB-negative with HIV positive samples?
HIV only: TB-/HIV+ (14)	Uninfected: TB-/HIV- (36)	Does HIV influence ELISA detection outcomes?



Supplementary Figure 5. Purified recombinant His-tagged *Mycobacterium tuberculosis* proteins shown by (A) 12% SDS-PAGE and (B)Western Blot. Lane 1: Spectra Multicolor Broad Range Ladder; Lane 2: RecRv3811; Lane 3: RecRv3717; Lane 4: RecRv3717; Lane 5: RecRv3209; Lane 6: RecRv3209; Lane 7: RecRv2598; Lane 8: RecRv2598; Lane 9: RecRv3491; Lane 10: RecRv3491. A total of five *Mycobacterium tuberculosis* conserved hypothetical proteins, namely Rv3811, Rv3491, Rv3209, and Rv2598, were selected to be cloned into a pET2a (+) vector system containing a 6xHis affinity tag used for the detection of expressed recombinant proteins. Due to the small affinity tag, slow log-phase expression was performed at 16°C overnight while agitating on a mechanical shaker to ensure adequate expressed *E.coli* pellet could be obtained. This also reduces the formation of inclusion bodies for downstream purification. In addition, to improve solubility, the signal peptides were removed from the sequences of Rv3717 and Rv3491. The expressed recombinant (rec-) proteins were then purified using the native method, without Urea (recRv3811). The remaining rec-proteins were purified using the denaturation method, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), and Western blot confirmed the size and purity of all five recombinant proteins with the absence of visible host *E. coli* proteins in the eluent.



Supplementary Figure 6. Confirmation of His-affinity tagged purified 36.7 kDa Periplasmic phosphate-binding lipoprotein PstS1 (Rv0934) of *Mycobacterium tuberculosis* shown by A. Signal peptide analysis for the removal of the signal peptide (cleavage position 23-24), B. 12% SDS-PAGE gel and C. Western Blot. Lane 1: PageRuler Plus Prestained Protein Ladder (Thermo Fisher Scientific), Lane 2-4: other *M.tuberculosis* proteins, Lane 5: His-PstS1 (Rv0934)

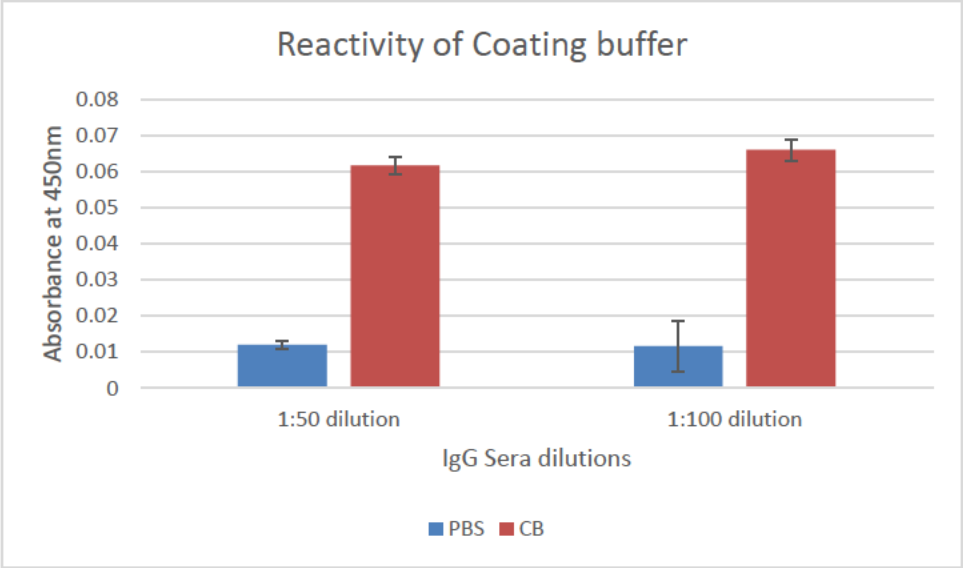
Supplementary table 7. Raw patient data with ELISA immunoreactivity outcomes (N=100)

sample		*Demographic data*				*confounding factors*		*TB status*	*ELISA outcome*					
No.	Patient ID	AGE	RACE	GENDER	SMOKER	CO-MORBIDITIES	HIV status	TB Outcome	Rv3811	Rv3717	Rv3491	Rv3209	Rv2598	PstS1
VR1	TB0002	49	Black	Female	No	Diabetes/Hypertension/Asthma/	Positive	Negative	0.562	0.967	0.510	0.403	0.317	1.942
VR2	TB0003	89	Black	Female	No	Hypertension	Negative	Negative	1.166	3.131	2.389	1.188	0.565	3.425
VR3	TB0004	53	Black	Female	No	HIV	Positive	Negative	0.683	1.460	0.883	0.307	0.475	3.203
VR4	TB0005	35	Black	Male	No	None	Negative	Negative	0.443	0.400	0.305	0.252	0.142	1.048
VR5	TB0006	43	Indian	Male	Yes	None	Negative	Negative	1.406	1.070	1.308	0.873	0.337	1.489
VR6	TB0007	60	Black	Female	No	Diabetes/HIV	Positive	Negative	0.586	0.927	0.377	0.354	0.577	2.434
VR7	TB0008	64	Black	Male	Yes	None	Negative	Negative	0.708	0.559	0.199	0.302	0.180	1.098
VR8	TB0009	52	Black	Male	No	Diabetes	Negative	Negative	3.219	1.209	1.307	0.909	0.238	1.898
VR9	TB0010	50	Black	Male	Yes	HIV	Positive	Negative	0.090	0.194	0.166	0.066	0.403	0.784
VR10	TB0011	34	Black	Male	Yes	None	Negative	Negative	0.148	0.178	0.175	0.134	0.054	0.622
VR11	TB0012	35	Black	Male	No	None	Negative	Negative	0.190	0.295	0.216	0.174	0.141	0.751
VR12	TB0013	34	Black	Male	Yes	None	Negative	Negative	0.479	0.848	0.388	0.397	0.442	1.433
VR13	TB0014	68	Black	Male	No	Hypertension	Negative	Negative	0.654	1.271	0.907	0.615	0.653	2.111
VR14	TB0015	56	Coloured	Female	No	None	Negative	Negative	0.342	0.612	0.431	0.319	0.165	1.260
VR15	TB0016	28	unknown	Female	unknown	Obesity	Negative	Negative	0.280	0.302	0.266	0.281	0.144	0.817
VR16	TB0017	20	Black	Male	Yes	None	Negative	Negative	0.204	0.213	0.155	0.182	0.112	0.750
VR17	TB0018	21	Black	Male	Yes	None	Negative	Negative	0.461	0.935	0.251	0.323	0.426	2.573
VR18	TB0019	44	Black	Male	Yes	HIV	Positive	Negative	0.171	0.116	0.142	0.070	0.092	1.002
VR19	TB0020	29	Black	Female	No	None	Negative	Negative	0.603	1.251	0.704	0.643	0.510	2.647
VR20	TB0021	30	Black	Male	No	None	Negative	Negative	0.261	0.339	0.304	0.583	0.196	1.326
VR21	TB0022	31	Black	Female	No	None	Negative	Negative	0.168	0.528	0.698	0.085	0.098	0.734
VR22	TB0023	28	Black	Female	No	HIV	Positive	Negative	0.434	0.865	0.751	0.337	0.233	0.892
VR23	TB0024	24	Black	unknown	No	None	Negative	Negative	0.288	0.308	0.548	0.162	0.217	0.248
VR24	TB0025	47	Black	Female	No	HIV	Positive	Negative	0.477	0.817	0.608	0.159	0.313	0.801
VR25	TB0026	44	Black	unknown	Yes	None	Negative	Negative	0.363	1.050	0.587	0.118	0.365	0.940
VR26	TB0027	34	Black	Female	No	None	Negative	Negative	0.222	0.599	0.590	0.473	0.130	0.422
VR27	TB0028	37	Black	Male	No	None	Negative	Negative	0.439	0.697	0.558	0.404	0.248	0.405
VR28	TB0029	29	Black	Female	No	HIV	Positive	Negative	0.770	1.695	1.136	0.783	0.393	1.677
VR29	TB0030	71	Coloured	Female	unknown	None	Negative	Negative	0.191	0.463	0.336	0.202	0.301	0.737
VR30	TB0031	85	unknown	unknown	unknown	None	Negative	Negative	0.470	0.500	0.478	0.534	0.454	0.434
VR31	TB0032	unknown	Black	Female	No	Diabetes/Hypertension	Negative	Negative	0.226	0.413	0.333	0.197	0.479	0.951
VR32	TB0033	34	Black	Female	No	None	Negative	Negative	0.190	0.377	0.467	0.117	0.175	0.447
VR33	TB0034	38	Black	Male	unknown	None	Negative	Negative	0.136	0.194	0.371	0.133	0.111	0.223
VR34	TB0035	60	Black	Female	No	None	Negative	Negative	0.384	0.516	0.572	0.334	0.228	0.769
VR35	TB0036	49	Black	Male	No	None	Negative	Negative	0.333	0.784	0.368	0.119	0.417	1.322
VR36	TB0037	39	Black	Female	No	None	Negative	Negative	0.424	0.654	0.602	0.255	0.392	0.833
VR37	TB0038	74	Black	Female	No	None	Negative	Negative	0.136	0.291	0.352	1.508	0.127	0.437
VR38	TB0039	43	Black	Female	No	HIV	Positive	Negative	0.145	0.218	0.528	0.155	0.090	0.170
VR39	TB0040	23	Black	Female	No	None	Negative	Negative	0.357	0.443	0.747	0.385	0.182	0.466
VR40	TB0041	unknown	Black	unknown	unknown		Negative	Negative	0.488	0.562	0.811	0.535	0.226	0.639
VR41	TB0042	42	Black	Female	No	HIV	Positive	Negative	0.353	0.149	0.193	0.152	0.022	0.515
VR42	TB0043	27	Black	Male	Yes	None	Negative	Negative	0.801	1.161	0.390	0.426	0.408	2.090
VR43	TB0044	62	Black	Female	No	Diabetes/HIV	Positive	Negative	0.645	0.385	0.265	0.278	0.183	0.802
VR44	TB0045	26	Black	Female	No	HIV	Positive	Negative	0.509	0.186	0.203	0.290	0.140	0.325
VR45	TB0046	32	Black	Female	Yes	None	Negative	Negative	0.429	0.182	0.152	0.268	0.154	0.319
VR46	TB0048	30	Black	Female	No	None	Negative	Negative	0.497	0.266	0.208	0.373	0.183	0.455
VR47	TB0049	42	Black	Male	Yes	HIV	Positive	Negative	0.488	0.636	0.188	0.334	0.299	1.285
VR48	TB0051	38	Black	Male	Yes	None	Negative	Negative	0.431	0.862	0.273	0.309	0.330	1.505
VR49	TB0052	22	Black	Male	Yes	None	Negative	Negative	0.074	0.219	0.141	0.196	0.064	0.610
VR50	TB0001	32	Black	Female	No	HIV	Positive	Negative	0.128	0.442	0.338	0.299	0.144	1.046
VR51	TB0053	unknown	Black	Female	No	None	Negative	Positive	0.191	0.788	0.595	0.390	0.189	2.119
VR52	TB0056	unknown	Black	Female	No	HIV	Positive	Positive	0.236	0.453	0.208	0.264	0.201	0.572
VR53	TB0058	39	Black	Female	No	HIV	Positive	Positive	0.601	2.379	1.251	0.589	0.638	3.290
VR54	TB0070	38	Black	Male	No	None	Negative	Positive	0.194	0.215	0.100	0.228	0.136	0.330
VR55	TB0071	37	Black	Female	No	HIV	Positive	Positive	0.182	0.142	0.080	0.205	0.126	0.161
VR56	TB0072	31	Black	Female	No	HIV	Positive	Positive	0.718	1.563	0.448	0.343	0.609	2.929
VR57	TB0073	27	Black	Female	No	HIV	Positive	Positive	0.243	0.237	0.235	0.048	0.130	0.740
VR58	TB0074	unknown	Black	unknown	unknown	None	Negative	Positive	0.317	0.178	0.145	0.125	0.102	0.144
VR59	TB0075	21	Black	Male	No	None	Negative	Positive	0.354	0.280	0.277	0.149	0.124	0.855
VR60	TB0076	55	Coloured	Male	unknown	None	Negative	Positive	0.414	0.266	0.246	0.204	1.517	1.389
VR61	TB0077	33	Black	Male	Yes	None	Negative	Positive	0.255	0.323	0.151	0.527	0.163	0.621
VR62	TB0078	36	Coloured	Male	Yes	None	Negative	Positive	0.797	1.463	0.706	0.797	0.994	2.470
VR63	TB0079	unknown	unknown	unknown	unknown	None	Negative	Positive	0.729	1.122	0.550	0.726	2.552	3.117
VR64	TB0080	35	Black	Male	Yes	None	Negative	Positive	0.176	0.325	0.133	0.535	0.168	0.135
VR65	TB0081	42	Black	Male	Yes	None	Negative	Positive	0.199	0.437	0.160	0.621	0.279	3.500
VR66	TB0082	29	Black	Male	Yes	HIV	Positive	Positive	0.148	0.367	0.129	0.692	0.195	0.115
VR67	TB0083	32	Black	Male	Yes	None	Negative	Positive	0.194	0.595	0.155	0.794	0.155	0.145
VR68	TB0084	34	Indian	Male	Yes	None	Negative	Positive	0.118	0.663	0.097	0.714	0.166	0.137
VR69	TB0085	46	Black	Male	Yes	None	Negative	Positive	0.346	0.427	0.068	0.095	0.440	0.868
VR70	TB0086	30	Black	Male	unknown	None	Negative	Positive	0.239	0.166	0.099	0.151	0.245	0.430
VR71	TB0087	54	Black	Male	Yes	None	Negative	Positive	0.364	0.260	0.182	0.313	0.379	0.502
VR72	TB0089	32	Black	Male	Yes	None	Negative	Positive	1.449	1.196	1.182	0.631	0.707	2.759
VR73	TB0090	27	Coloured	Male	Yes	None	Negative	Positive	0.230	0.101	0.108	0.193	0.147	0.390
VR74	TB0092	31	Black	Male	Yes	None	Negative	Positive	0.834	0.799	0.427	0.286	0.688	2.454
VR75	TB0093	28	Black	Male	Yes	None	Negative	Positive	0.312	0.135	0.126	0.173	0.174	0.502
VR76	TB0094	26	Black	Male	Yes	None	Negative	Positive	1.296	1.028	0.840	0.380	1.673	2.861

sample		Demographic data				confounding factors		TB status	ELISA outcome					
No.	Patient ID	AGE	RACE	GENDER	SMOKER	CO-MORBIDITIES	HIV status	TB Outcome	Rv3811	Rv3717	Rv3491	Rv3209	Rv2598	PstS1
VR77	TB0096	28	Black	Male	Yes	None	Negative	Positive	0.832	0.821	0.445	0.140	0.380	1.091
VR78	TB0097	29	Coloured	Male	Yes	None	Negative	Positive	0.695	0.517	0.552	0.530	0.374	0.661
VR79	TB0098	43	Coloured	Male	Yes	None	Negative	Positive	0.568	0.277	0.275	0.160	0.214	0.341
VR80	TB0099	30	Black	Male	Yes	None	Negative	Positive	0.724	0.424	0.370	0.168	0.291	0.765
VR81	TB0100	33	Black	Male	Yes	HIV	Positive	Positive	0.308	0.462	0.045	0.079	0.357	0.726
VR82	TB0101	41	Black	Female	No	HIV, Diabetes	Positive	Positive	0.517	0.363	0.353	0.258	0.175	0.287
VR83	TB0105	36	Black	Male	No	HIV	Positive	Positive	0.098	0.139	0.120	0.096	0.124	0.095
VR84	TB0107	48	Black	Male	No	HIV	Positive	Positive	0.357	0.512	0.136	0.133	0.376	0.638
VR85	TB0108	18	Black	Male	No	None	Negative	Positive	0.433	0.625	0.123	0.111	0.619	0.804
VR86	TB0109	30	Black	Female	No	HIV	Positive	Positive	0.599	0.769	0.247	0.121	0.673	1.152
VR87	TB0110	55	Black	Male	No	HIV	Positive	Positive	0.318	0.181	0.127	0.106	0.114	0.394
VR88	TB0112	78	Black	Male	No	Anemia	Negative	Positive	0.755	0.766	0.283	0.309	0.433	1.792
VR89	TB0113	31	Black	Male	No	None	Negative	Positive	1.433	1.227	0.207	0.296	0.806	3.274
VR90	TB0114	52	Black	Male	No	HIV, Diabetes	Positive	Positive	1.332	1.247	0.267	0.369	0.852	3.146
VR91	TB0115	39	Black	Male	No	HIV	Positive	Positive	0.464	0.332	0.234	0.257	0.209	0.343
VR92	TB0116	53	Black	Female	No	HIV, Hypertension	Positive	Positive	0.117	0.413	0.283	0.264	0.215	0.459
VR93	TB0118	46	Black	Male	No	None	Negative	Positive	1.190	1.296	0.251	0.350	1.246	3.162
VR94	TB0119	29	Black	Female	No	HIV	Positive	Positive	1.503	2.042	0.280	0.257	1.888	3.127
VR95	TB0120	56	Black	Male	Yes	HIV	Positive	Positive	1.668	1.843	0.181	0.361	1.665	3.445
VR96	TB0121	26	Black	Male	Yes	HIV	Positive	Positive	0.633	0.661	0.410	0.399	0.431	0.921
VR97	TB0123	28	Black	Male	No	None	Negative	Positive	1.083	1.184	1.772	1.602	2.294	1.230
VR98	TB0124	28	Black	Female	No	HIV, Asthma/ Respiratory illness	Positive	Positive	0.492	0.591	0.159	0.171	0.351	1.895
VR99	TB0125	37	Black	Male	No	HIV	Positive	Positive	0.783	0.916	0.184	0.218	0.820	1.675
VR100	TB0126	58	Black	Male	No	None	Negative	Positive	0.787	0.861	0.141	0.263	0.720	1.859

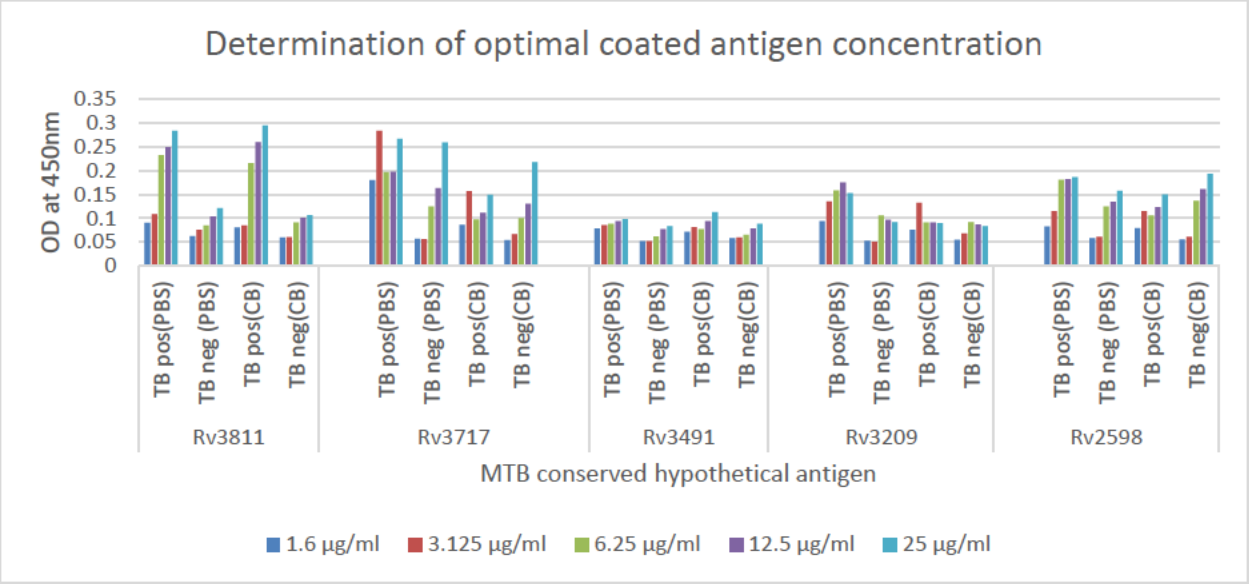
B4.2 Serological optimization supplementary data

(I) Coating buffer and blocking sensitivity

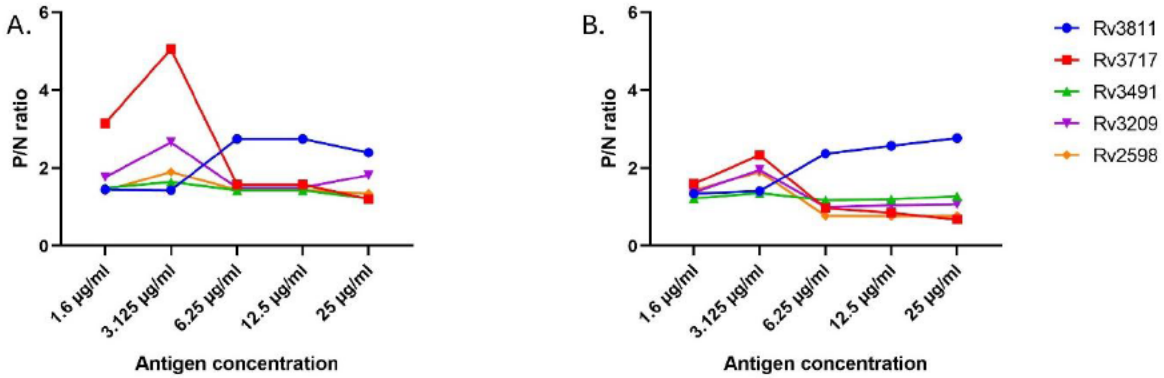


Supplementary Figure 1. Background signal interference based on the coating buffer used for positive blood plasma samples. Reactivity is better with the PBS buffer compared to the CB buffer. Subsequently, the best reactivity is seen in 25 ug/mL as the specificity is the highest, particularly at 1:50 serum dilution. There is very low reactivity among TB+ patients throughout.

(II) Antigen concentration



Supplementary figure 2. Determination of optimal coating concentration for a conserved hypothetical *Mycobacterium tuberculosis* antigen by ELISA, comparing phosphate-buffered saline (PBS) and carbonate-bicarbonate (CB) buffers using TB-positive and TB-negative serum samples.



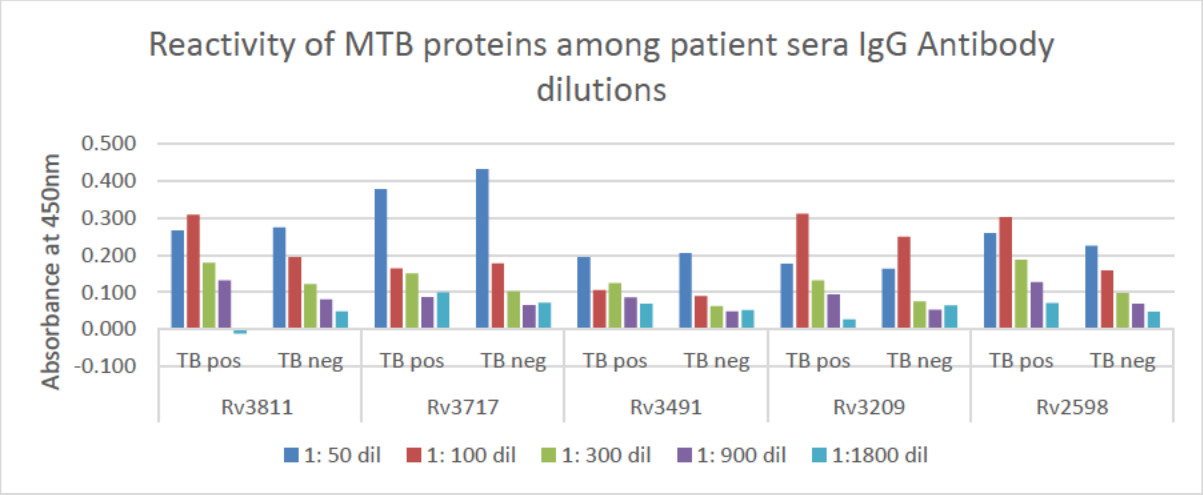
Supplementary figure 3. Positive-to-negative (P/N) ratios for determining the optimal concentration of a conserved hypothetical *Mycobacterium tuberculosis* antigen using TB-positive and TB-negative serum samples, comparing two coating buffers. A. Comparing *M. tuberculosis* antigen concentration coated with Phosphate-buffered saline (PBS), B. Comparing *M. tuberculosis* antigen concentration coated with Carbonate-bicarbonate (CB) buffer.

(III) Selection of Immunoplate, plasma dilutions, and coating buffer selection

Optimization parameters using: Maxisorp Immunoplate, PBS (pH 7.2) coating buffer, Patient plasma dilutions.

Supplementary table 1. Reactivity of TB patient blood plasma to *Mycobacterium tuberculosis* antigen with variable patient plasma dilution parameters for optimization using the Nunc Maxisorp Immunoplate coated with PBS (pH 7.2).

Antigen	Blood plasma	Dilution of patient blood plasma														
		1:50			1:100			1:300			1:900			1:1800		
Rv3811	TB positive	0.296	0.235	0.267	0.313	0.296	0.317	0.171	0.190	0.177	0.136	0.128	0.132	-0.007	-0.015	-0.012
	TB negative	0.285	0.272	0.266	0.195	0.195	0.195	0.122	0.118	0.124	0.077	0.086	0.076	0.048	0.048	0.048
Rv3717	TB positive	0.377	0.377	0.380	0.155	0.161	0.174	0.145	0.15	0.154	0.087	0.086	0.088	0.1	0.098	0.098
	TB negative	0.432	0.409	0.454	0.175	0.182	0.176	0.101	0.102	0.102	0.061	0.066	0.067	0.073	0.076	0.066
Rv3491	TB positive	0.202	0.185	0.194	0.1	0.104	0.111	0.125	0.119	0.129	0.082	0.085	0.087	0.097	0.055	0.053
	TB negative	0.215	0.211	0.19	0.087	0.087	0.095	0.058	0.065	0.062	0.051	0.046	0.048	0.064	0.043	0.046
Rv3209	TB positive	0.185	0.17	0.174	0.316	0.286	0.332	0.128	0.131	0.136	0.093	0.094	0.093	0.029	0.025	0.025
	TB negative	0.163	0.16	0.165	0.246	0.257	0.243	0.073	0.073	0.077	0.051	0.051	0.054	0.063	0.065	0.062
Rv2598	TB positive	0.254	0.26	0.262	0.3	0.316	0.291	0.189	0.19	0.182	0.126	0.134	0.119	0.071	0.067	0.073
	TB negative	0.239	0.224	0.213	0.156	0.163	0.157	0.098	0.1	0.093	0.067	0.072	0.067	0.045	0.047	0.049



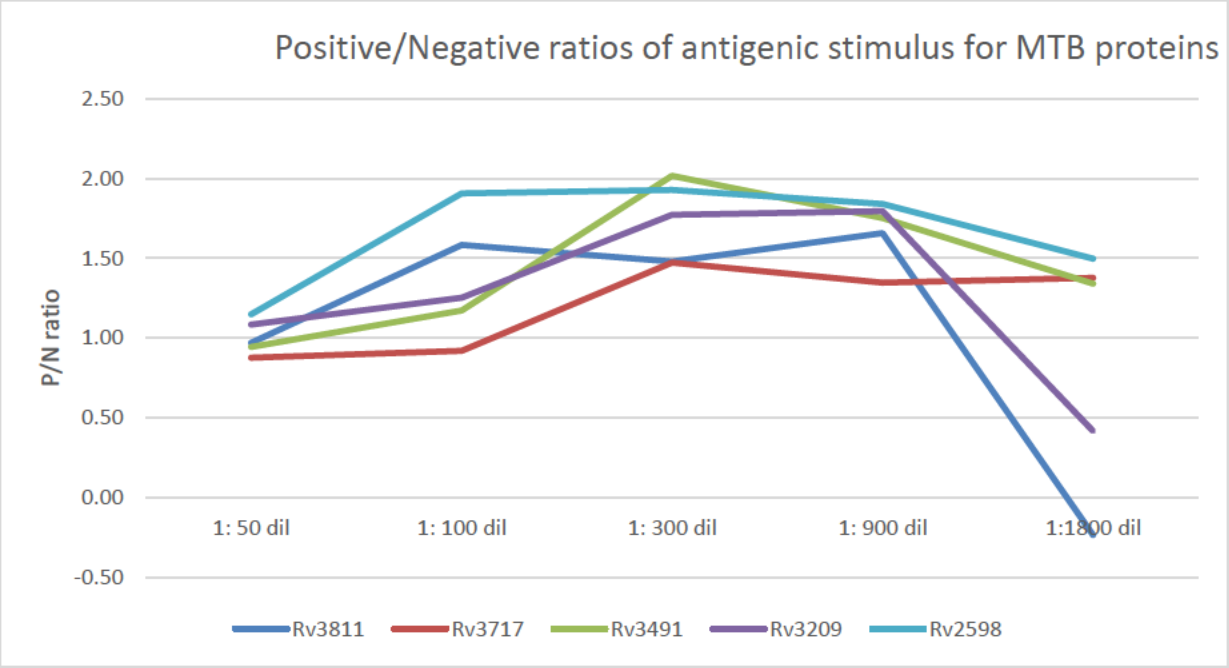
Supplementary figure 4. Graph of *Mycobacterium tuberculosis* protein stimulus to TB+/TB- South African patient blood plasma at variable dilution parameters using PBS (pH 7.2) coating buffers. (Secondary antibody dilution at 1:10 000)

Supplementary table 2. Reactivity of *Mycobacterium tuberculosis* antigenic protein among TB+/TB-patient blood plasma. CV<13%

Protein	*Blood plasma samples*	*Absorbance at 450nm (±SD)*			
		1:50	*CV%*	*1: 100*	*CV%*
Rv3811	*TB positive*	0.266 (0.031)	11.5	0.309 (0.011)	3.6
	TB negative	0.274 (0.010)	3.5	0.195 (0.000)	0.0
Rv3717	*TB positive*	0.378 (0.002)	0.5	0.163 (0.010)	5.9
	TB negative	0.432 (0.023)	5.2	0.178 (0.004)	2.1
Rv3491	*TB positive*	0.194 (0.0009)	4.4	0.105 (0.006)	5.3
	TB negative	0.205 (0.013)	6.5	0.090 (0.005)	5.2
Rv3209	*TB positive*	0.176 (0.008)	4.4	0.311 (0.023)	7.5
	TB negative	0.163 (0.003)	1.5	0.249 (0.007)	3.0
Rv2598	*TB positive*	0.259 (0.004)	1.6	0.302 (0.013)	4.2
	TB negative	0.225 (0.0013)	5.8	0.159 (0.004)	2.4
		1:300	*CV%*	**1:900**	*CV%*
Rv3811	*TB positive*	0.179 (0.010)	5.4	0.132 (0.004)	3.0
	TB negative	0.121 (0.003)	2.5	0.080 (0.006)	6.9
Rv3717	*TB positive*	0.150 (0.005)	3.0	0.087 (0.001)	1.1
	TB negative	0.102 (0.001)	0.6	0.065 (0.003)	5.0
Rv3491	*TB positive*	0.124 (0.005)	4.0	0.085 (0.003)	3.0
	TB negative	0.062 (0.004)	5.7	0.048 (0.003)	5.2
Rv3209	*TB positive*	0.132 (0.004)	3.1	0.093 (0.001)	0.6
	TB negative	0.074 (0.002)	3.1	0.052 (0.002)	3.3
Rv2598	*TB positive*	0.187 (0.004)	2.3	0.126 (0.008)	5.9
	TB negative	0.097 0.004)	3.7	0.069(0.003)	4.2
		1:1 800	SD	*CV%*	
Rv3811	*TB positive*	-0.011	0.004	-35.7	
	TB negative	0.048	0.000	0.0	
Rv3717	*TB positive*	0.099	0.001	1.2	
	TB negative	0.072	0.005	7.2	
Rv3491	*TB positive*	0.068	0.025	36.4	
	TB negative	0.051	0.011	22.3	
Rv3209	*TB positive*	0.026	0.002	8.8	
	TB negative	0.063	0.002	2.4	
Rv2598	*TB positive*	0.070	0.003	4.3	
	TB negative	0.047	0.002	4.3	

Supplementary table 3: Optimization of monoclonal mouse anti-human IgG-Fc specific peroxidase (Merck, Darmstadt, Germany) to determine the *Mycobacterium tuberculosis* detection in TB+/TB- patient blood plasma (P/N ratio).

M. tuberculosis protein	[patient blood plasma dilutions]				
	1: 50	**1: 100**	**1: 300**	**1: 900**	**1:1800**
Rv3811	0.97	1.58	1.48	1.66	-0.24
Rv3717	0.88	0.92	1.47	1.35	1.38
Rv3491	0.94	1.17	2.02	1.75	1.34
Rv3209	1.08	1.25	1.77	1.79	0.42
Rv2598	1.15	1.91	1.93	1.84	1.50

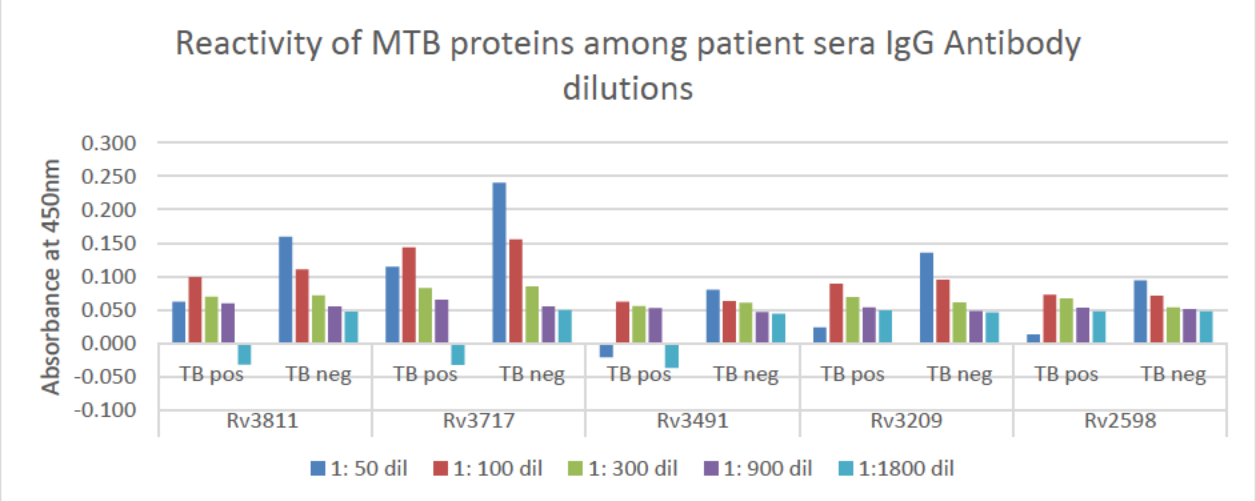


Supplementary figure 5. Positive-to-negative (P/N) ratios of antigenic stimulus of *Mycobacterium tuberculosis* proteins among TB+/TB- patient blood plasma at monoclonal mouse anti-human IgG-Fc specific peroxidase (Merck, Darmstadt, Germany) variable dilution parameters (P/N ratio).

(IV) Optimization parameters: Maxisorp Immunoplate, Carbonate bicarbonate (pH 9.6) coating buffer, patient plasma dilutions

Supplementary table 4. Reactivity of TB patient blood plasma to *Mycobacterium tuberculosis* antigen with variable patient blood plasma dilution parameters for optimization using the Nunc Maxisorp Immunoplate coated with PBS (pH 7.2).

Antigen	Blood plasma	Dilution of patient blood plasma														
		1:50			1:100			1:300			1:900			1:1800		
Rv3811	TB positive	0.068	0.059	0.059	0.101	0.096	0.099	0.071	0.069	0.068	0.071	0.054	0.053	-0.032	-0.035	-0.027
	TB negative	0.159	0.153	0.166	0.111	0.112	0.108	0.074	0.068	0.072	0.048	0.067	0.051	0.046	0.048	0.047
Rv3717	TB positive	0.114	0.108	0.121	0.14	0.142	0.147	0.081	0.084	0.082	0.061	0.071	0.064	-0.035	-0.027	-0.035
	TB negative	0.238	0.242	0.24	0.153	0.157	0.156	0.087	0.084	0.083	0.054	0.056	0.055	0.048	0.049	0.051
Rv3491	TB positive	-0.024	-0.014	-0.025	0.058	0.062	0.066	0.053	0.054	0.06	0.053	0.052	0.053	-0.031	-0.042	-0.038
	TB negative	0.08	0.08	0.08	0.063	0.064	0.062	0.085	0.047	0.049	0.049	0.044	0.046	0.044	0.044	0.044
Rv3209	TB positive	0.032	0.02	0.018	0.091	0.089	0.087	0.067	0.069	0.071	0.052	0.055	0.054	0.049	0.048	0.05
	TB negative	0.137	0.135	0.135	0.093	0.096	0.096	0.06	0.06	0.062	0.047	0.045	0.052	0.046	0.046	0.046
Rv2598	TB positive	0.018	0.011	0.01	0.072	0.072	0.073	0.063	0.077	0.061	0.055	0.05	0.054	0.048	0.047	0.048
	TB negative	0.094	0.097	0.091	0.07	0.071	0.072	0.054	0.052	0.055	0.049	0.055	0.048	0.046	0.046	0.05



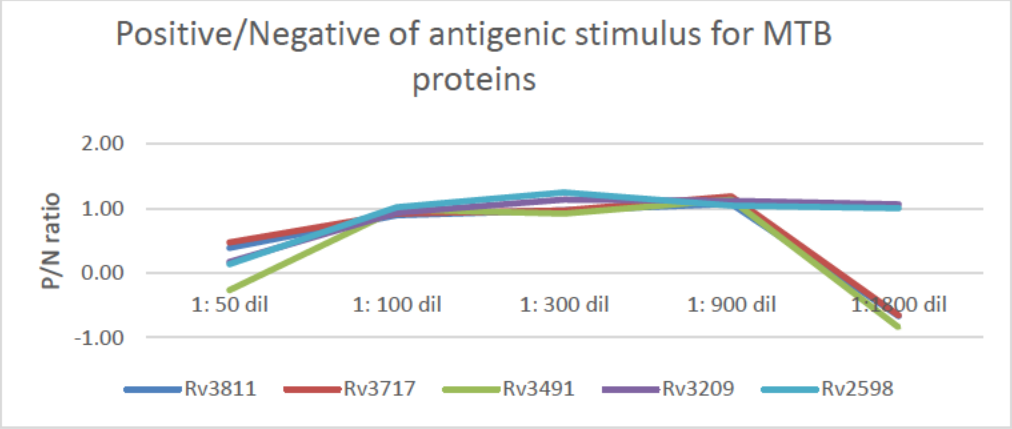
Supplementary figure 6. Graph of *Mycobacterium tuberculosis* protein stimulus to TB+/TB- South African patient blood plasma at variable dilution parameters using PBS (pH 7.2) coating buffers. (Merck Secondary antibody dilution at 1:10 000)

Supplementary Table 5. Reactivity of *Mycobacterium tuberculosis* antigenic protein among TB+/TB- patient blood plasma. CV<13%

Protein	*Blood plasma samples*	*Absorbance at 450nm (±SD)*			
		1:50	*CV%*	*1: 100*	*CV%*
Rv3811	*TB positive*	0.062 (0.005)	0.39	0.099 (0.003)	2.6
	TB negative	0.159 (0.007)	0.48	0.110 (0.002)	1.9
Rv3717	*TB positive*	0.114 (0.007)	-0.26	0.143 (0.004)	2.5
	TB negative	0.240 (0.002)	0.17	0.155 (0.002)	1.3
Rv3491	*TB positive*	-0.021 (0.006)	0.14	0.062 (0.004)	6.5
	TB negative	0.080 (0.000)	0.39	0.063 (0.001)	1.6
Rv3209	*TB positive*	0.023 (0.008)	0.48	0.089 (0.002)	2.2
	TB negative	0.136 (0.001)	-0.26	0.095 (0.002)	1.8
Rv2598	*TB positive*	0.013 (0.004)	0.17	0.072 (0.001)	0.8
	TB negative	0.094 (0.003)	0.14	0.071 (0.001)	1.4
		1:300	*CV%*	**1:900**	*CV%*
Rv3811	*TB positive*	0.069 (0.002)	2.2	0.059 (0.010)	17.0
	TB negative	0.071 (0.003)	4.3	0.055 (0.010)	18.5
Rv3717	*TB positive*	0.082 (0.002)	1.9	0.065 (0.005)	7.9
	TB negative	0.085 (0.002)	2.5	0.055 (0.001)	1.8
Rv3491	*TB positive*	0.056 (0.004)	6.8	0.053 (0.001)	1.1
	TB negative	0.060 (0.021)	35.4	0.046 (0.003)	5.4
Rv3209	*TB positive*	0.069 (0.002)	2.9	0.054 (0.002)	2.8
	TB negative	0.061 (0.001)	1.9	0.048 (0.004)	7.5
Rv2598	*TB positive*	0.067 (0.009)	13.0	0.053 (0.003)	5.0
	TB negative	0.054 (0.002)	2.8	0.051 (0.004)	7.5
		1:1 800		*CV%*	
Rv3811	*TB positive*	-0.031	(0.004)	-12.9	
	TB negative	0.047	(0.001)	2.1	
Rv3717	*TB positive*	-0.032	(0.005)	-14.3	
	TB negative	0.049	(0.002)	3.1	
Rv3491	*TB positive*	-0.037	(0.006)	-15.0	
	TB negative	0.044	(0.000)	0.0	
Rv3209	*TB positive*	0.049	(0.001)	2.0	
	TB negative	0.046	(0.000)	0.0	
Rv2598	*TB positive*	0.048	(0.001)	1.2	
	TB negative	0.047	(0.002)	4.9	

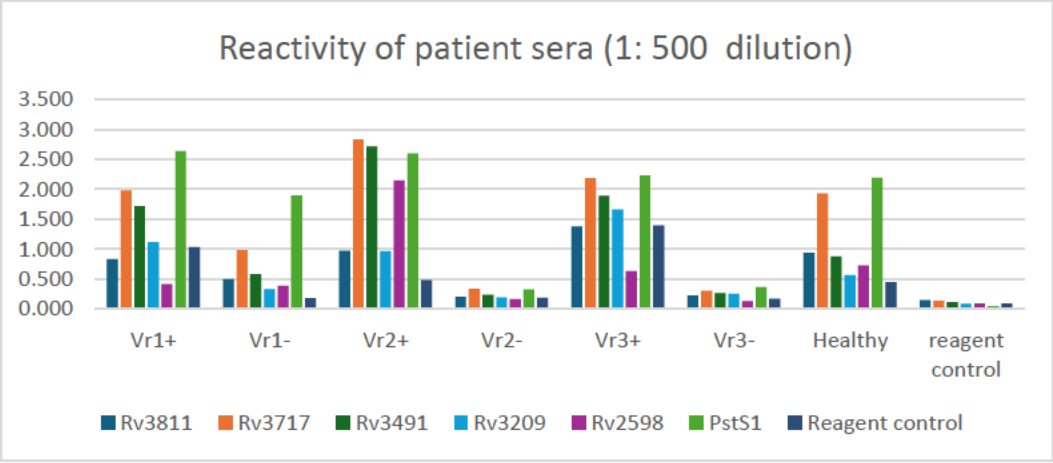
Supplementary table 6: optimization of secondary antibody to determine *Mycobacterium tuberculosis* detection in TB+/TB- patient blood plasma (P/N ratio).

M. tuberculosis protein	[patient blood plasma dilutions]				
	1: 50	1: 100	1: 300	1: 900	1:1800
Rv3811	0.39	0.89	0.97	1.07	-0.67
Rv3717	0.48	0.92	0.97	1.19	-0.66
Rv3491	-0.26	0.98	0.92	1.14	-0.84
Rv3209	0.17	0.94	1.14	1.12	1.07
Rv2598	0.14	1.02	1.25	1.05	1.01

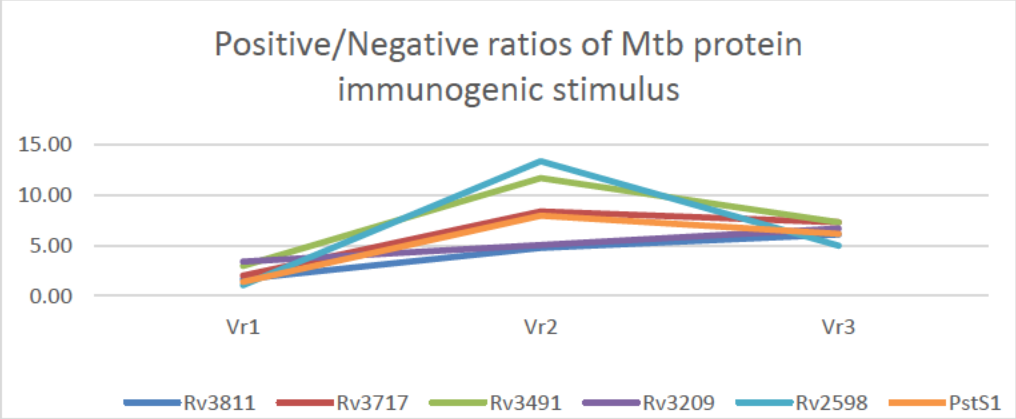


Supplementary figure 7. Positive-to-negative (P/N) ratios of antigenic stimulus of *Mycobacterium tuberculosis* proteins among TB+/TB- patient blood plasma at variable dilution parameters (P/N ratio).

Mouse IgG-Fc secondary, HRP conjugated (Thermo Fisher Scientific, 1: 500 dilution)



Supplementary figure 8. Graph of *Mycobacterium tuberculosis* conserved hypothetical proteins Antigenic stimulus to VR patient blood plasma (1:300) using PBS (pH 7.2) coating buffers. (Secondary antibody dilution at 1:500)

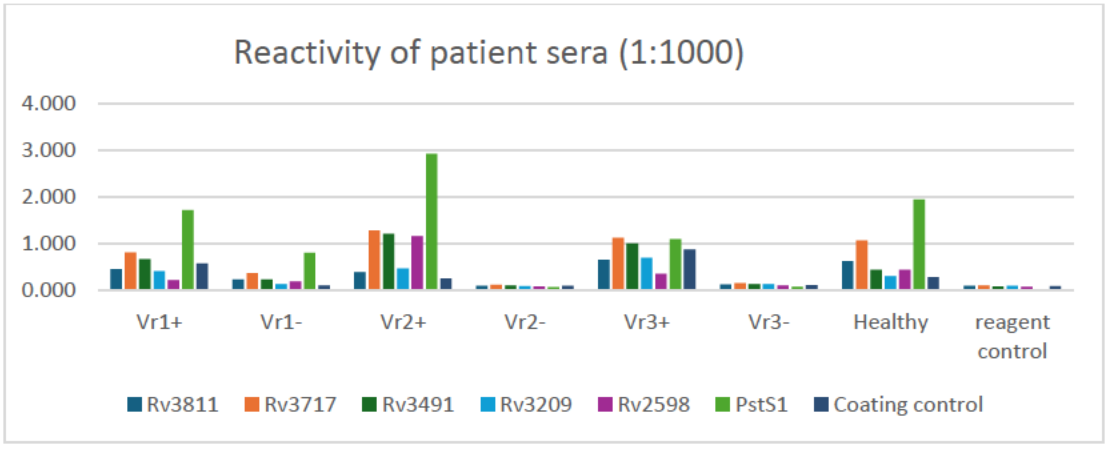


Supplementary figure 9. Positive-to-negative (P/N) ratios of antigenic stimulus of *Mycobacterium tuberculosis* conserved hypothetical proteins among TB+/TB- patient blood plasma dilution of 1:300 (P/N ratio).

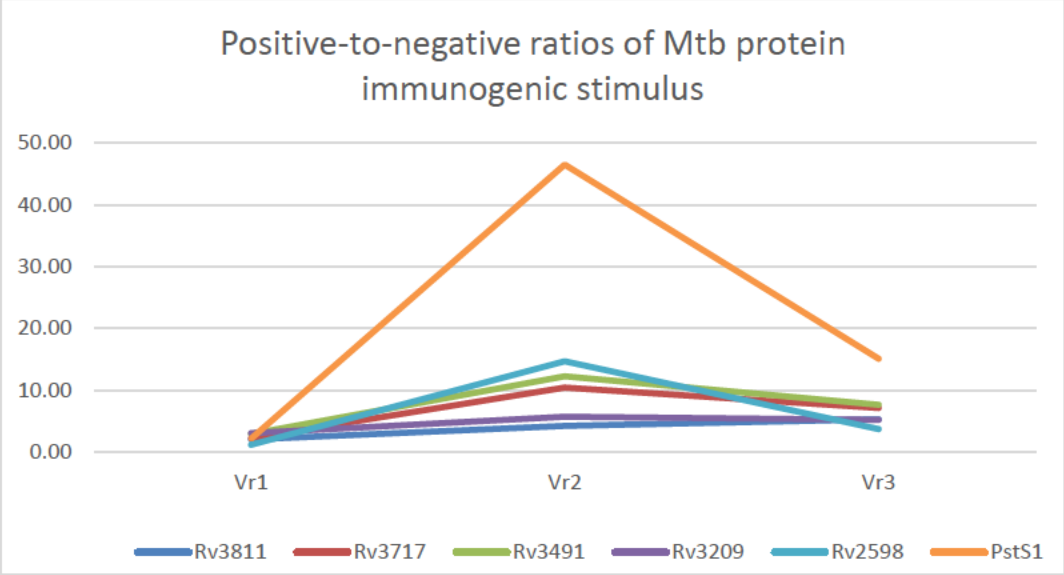
Supplementary Table 7. P/N ratios of patient plasma samples (secondary Antibody dilution 1: 500)

P/N ratios	VR1	VR2	VR3
Rv3811	1.66	4.7	6.1
Rv3717	2.01	8.3	7.3
Rv3491	2.97	11.6	7.3
Rv3209	3.39	5.0	6.7
Rv2598	1.07	13.3	5.0
PstS1	1.39	7.9	6.1

Mouse IgG-Fc secondary, HRP conjugated (Thermo Fisher Scientific, 1: 1 000)



Supplementary figure 10. Graph of *Mycobacterium tuberculosis* Conserved hypothetical proteins Antigenic stimulus to VR patient blood plasma (1:300) using PBS (pH 7.2) coating buffers. (Secondary antibody dilution at 1:1 000)

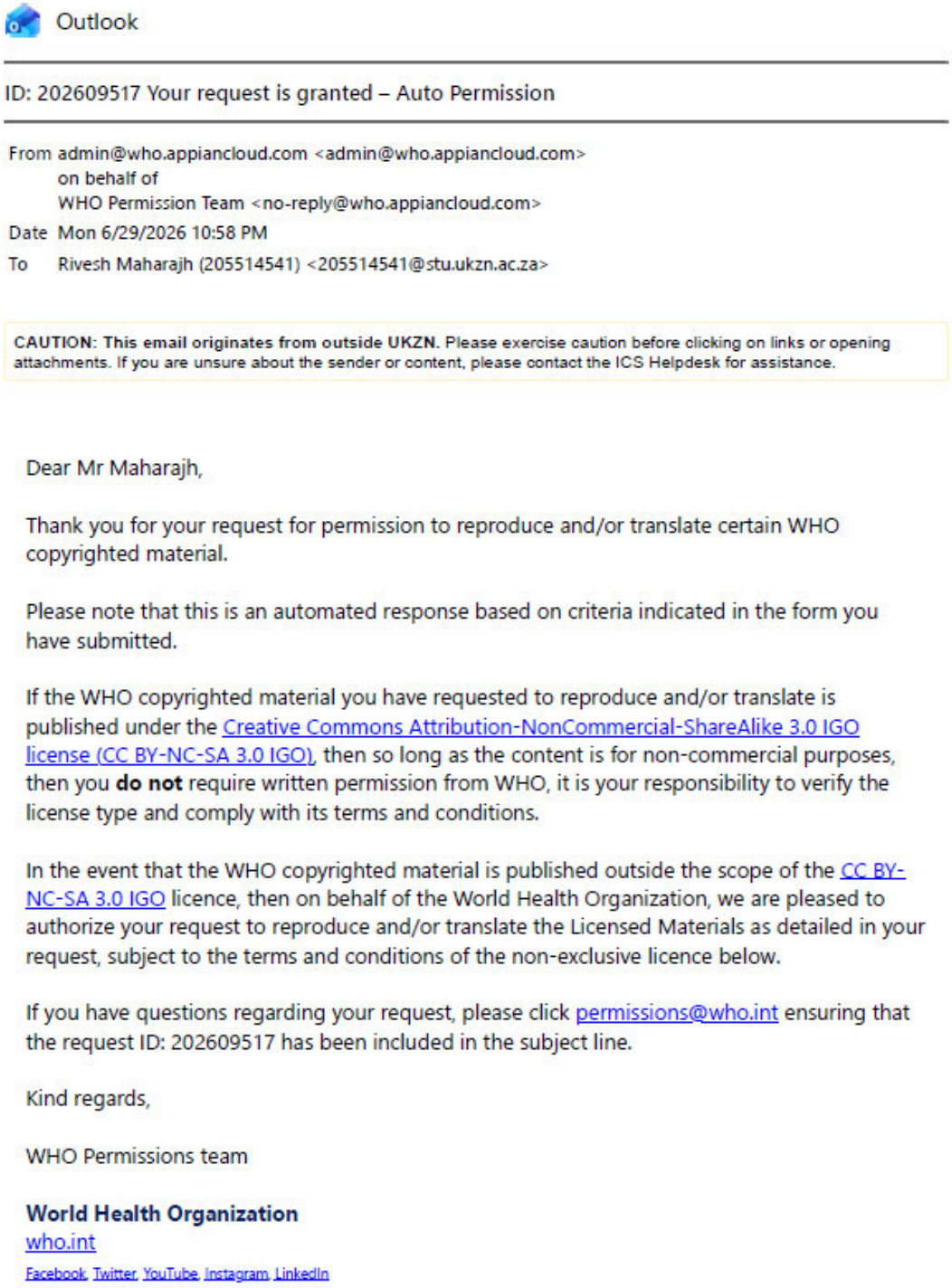


Supplementary figure 11. Positive-to-negative (P/N) ratios of antigenic stimulus of *Mycobacterium tuberculosis* conserved hypothetical proteins among TB+/TB- patient blood plasma dilution of 1:300 (P/N ratio).

Supplementary Table 8. Positive-to-negative (P/N) ratios of patient plasma samples (Secondary antibody dilution at 1:1 000)

P/N ratios	VR1	VR2	VR3
Rv3811	2.00	4.2	5.3
Rv3717	2.20	10.4	7.1
Rv3491	2.88	12.2	7.6
Rv3209	3.01	5.7	5.2
Rv2598	1.13	14.7	3.7
PstS1	2.12	46.5	15.1

Appendix C. Approved permission to use figures in chapter 2 literature review.



Outlook

ID: 202609517 Your request is granted – Auto Permission

From admin@who.appiancloud.com <admin@who.appiancloud.com>
on behalf of
WHO Permission Team <no-reply@who.appiancloud.com>
Date Mon 6/29/2026 10:58 PM
To Rivesh Maharajh (205514541) <205514541@stu.ukzn.ac.za>

CAUTION: This email originates from outside UKZN. Please exercise caution before clicking on links or opening attachments. If you are unsure about the sender or content, please contact the ICS Helpdesk for assistance.

Dear Mr Maharajh,

Thank you for your request for permission to reproduce and/or translate certain WHO copyrighted material.

Please note that this is an automated response based on criteria indicated in the form you have submitted.

If the WHO copyrighted material you have requested to reproduce and/or translate is published under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 IGO license (CC BY-NC-SA 3.0 IGO), then so long as the content is for non-commercial purposes, then you **do not** require written permission from WHO, it is your responsibility to verify the license type and comply with its terms and conditions.

In the event that the WHO copyrighted material is published outside the scope of the CC BY-NC-SA 3.0 IGO licence, then on behalf of the World Health Organization, we are pleased to authorize your request to reproduce and/or translate the Licensed Materials as detailed in your request, subject to the terms and conditions of the non-exclusive licence below.

If you have questions regarding your request, please click permissions@who.int ensuring that the request ID: 202609517 has been included in the subject line.

Kind regards,

WHO Permissions team

World Health Organization
who.int
Facebook, Twitter, YouTube, Instagram, LinkedIn

Signature Reviews

Pathogenicity and virulence of *Mycobacterium tuberculos*

Kathryn C. Rahlwes, Beatriz R.S. Dias, Priscila C. Campos, Samuel Alvarez-Arguedas & Michael U. Shiloh

Article: 2150449 | Received 18 Nov 2021, Accepted 17 Nov 2022, Accepted author version posted online: 23 Nov 2022, Published online: 04 Jan 2023

Cite this article https://doi.org/10.1080/21505594.2022.2150449 Check for updates

Rela

Peo

Virul
tube

Reprints and Permissions

This is an open access article distributed under the terms of the Creative Commons CC BY license, which permits unrestricted use, distribution, reproduction in any medium, provided the original work is properly cited.

You are not required to obtain permission to reuse this article in part or whole.

www.sigmaaldrich.com/ZA/en/life-science/legal/site-use-terms

Intellectual Property Rights

The Website and its entire contents, features, and functionality (including but not limited to all information, software, text, displays, images, video, and audio, and the design, selection, and arrangement thereof), are owned by us, our affiliates, licensors, or other providers of such material (including Marketplace Sellers) and are protected by applicable copyright, trademark, patent, trade secret, designs, database rights, and other intellectual property or proprietary rights laws.

You must not reproduce, distribute, modify, create derivative works of, publicly display, publicly perform, republish, download, store, or transmit any part of our Website or content on our Website without our written consent or if written consent is provided, in a manner that is inconsistent with purpose for which our written consent is provided except as specified in this section. You may download certain materials from the Website such as product images, charts, graphs, etc. for non-commercial personal use only. Research organizations whose primary goal is to conduct scientific research or to carry out educational activities, as well as accredited educational institutions such as universities and colleges may download and reproduce the materials mentioned in this section for the sole purpose of illustration for teaching or scientific research, so long as the source is indicated and only to the extent justified for the non-commercial purpose to be achieved. Materials specified in this section do not include the design or layout of the Website or its parts or any other design elements such as logos, icons, symbols, fonts, etc. of the Website. No right, title, or interest in or to the Website or any content on the Website is transferred to you, and all rights not expressly granted are reserved by us. Any use of the Website not expressly permitted by these Terms of Use is a breach of these Terms of Use and may violate copyright, trademark, and other laws.

Our company names, logos and all related product and service names, designs, and slogans are our trademarks or those of our affiliates or licensors. You must not use such marks without our prior written permission. All other names, logos, product and service names, designs, and slogans on this Website are the trademarks of their respective owners.

License Number	6298420339762
License date	Jun 29, 2026

Printable Details

Licensed Content

Licensed Content Publisher	Elsevier
Licensed Content Publication	Clinical Microbiology and Infection
Licensed Content Title	Tuberculosis prevention: current strategies and future directions
Licensed Content Author	Anca Vasiliu,Leonardo Martinez,Rishi K. Gupta,Yohhei Hamada,Tara Ness,Alexander Kay,Maryline Bonnet,Martina Sester,Stefan H.E. Kaufmann,Christoph Lange,Anna M. Mandalakas
Licensed Content Date	Sep 1, 2024
Licensed Content Volume	30
Licensed Content Issue	9
Licensed Content Pages	8

Order Details

Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/ illustrations	1
Format	electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No

About Your Work

Title of new work	The identification of novel Mycobacterium tuberculosis immunogenic adhesin proteins and the application as a potential vaccine and diagnostic candidate
Institution name	University of KwaZulu-Natal
Expected presentation date	Jun 2026

Additional Data

Portions	Figure 1
The Requesting Person / Organization to Appear on the License	Mr. Rivesh Maharajh

CCC Marketplace

This is a License Agreement between Rivesh Maharajh ("User") and Copyright Clearance Center, Inc. ("CCC") on behalf of the Rightsholder identified in the order details below. The license consists of the order details, the Marketplace Permissions General Terms and Conditions below, and any Rightsholder Terms and Conditions which are included below.

All payments must be made in full to CCC in accordance with the Marketplace Permissions General Terms and Conditions below.

Order Date	01-Jul-2026	**Type of Use**	Republish in a thesis/dissertation
Order License ID	1745485-1	**Publisher**	Wiley
ISBN-13	9781683670124	**Portion**	Image/photo/illustration

LICENSED CONTENT

Publication Title	Gram-Positive Pathogens	**Rightsholder**	John Wiley & Sons - Books
Article Title	Unraveling the Structure of the Mycobacterial Envelope	**Publication Type**	Book
		Start Page	1087
Date	10/22/2019	**End Page**	1095
Language	English		
Country	United States of America		

REQUEST DETAILS

Portion Type	Image/photo/illustration	**Distribution**	Other territories and/or countries
Number of Images / Photos / Illustrations	1	**Enter Territories/Countries**	South Africa
Format (select all that apply)	Electronic	**Translation**	Original language of publication
Who Will Republish the Content?	Academic institution	**Copies for the Disabled?**	No
Duration of Use	Life of current edition	**Minor Editing Privileges?**	No
Lifetime Unit Quantity	Up to 499	**Incidental Promotional Use?**	No
Rights Requested	Main product	**Currency**	USD

NEW WORK DETAILS

Title	The identification of novel Mycobacterium tuberculosis immunogenic adhesin proteins and the application as a potential vaccine and diagnostic candidate	**Institution Name**	University of KwaZulu-Natal
		Expected Presentation Date	2026-06-30
Instructor Name	Rivesh Maharajh		

ADDITIONAL DETAILS

The Requesting Person / Organization to Appear on the License	Rivesh Maharajh

REQUESTED CONTENT DETAILS

Title, Description or Numeric Reference of the Portion(s)	FIGURE 3 Model of the mycobacterial cell envelope	**Title of the Article / Chapter the Portion Is From**	Unraveling the Structure of the Mycobacterial Envelope
Editor of Portion(s)	Daffé, Mamadou; Marrakchi, Hedia	**Author of Portion(s)**	Daffé, Mamadou; Marrakchi, Hedia
Volume / Edition	3	**Publication Date of Portion**	2019-01-01
Page or Page Range of Portion	1087-1095		

Appendix D. Turnitin report

-Attached is the similarity report for the entire thesis submitted to Turnitin.com to determine the similarity index (14%). Chapter 3 was excluded from the assessment as the manuscript is the candidate's own publication.



88 "Handbook of Tuberculosis", Wiley, 2008
Publication

89 "Vaccine Design", Springer Science and Business Media LLC, 2022
Publication

90 "When HFE gene mutation doesn't make sense", Journal of Medical Genetics, 2002
Publication

91 Adrielle Zagmignan, Adeliane Castro da Costa, José Lima Viana, Lidio Gonçalves Lima N et al. "Identification of specific antibodies against the Ag85C-MPT51-HspX fusion protei (CMX) for serological screening of tuberculosis in endemic area", Expert Review of Clini Immunology, 2017
Publication

92 Carlos Martin, Nacho Aguilo, Dessislava Marinova, Jesus Gonzalo-Asensio. "Update on T Vaccine Pipeline", Applied Sciences, 2020
Publication

93 Freddy A. Nunez, Rafael C. Otoni, Marcos R. de Araujo-Silva, Vivian L. de Oliveira, Carlos A.O. Ramirez, Wendel A. Alves. "NiO/ZnO p–n heterojunction electrochemical immunosensor for label-free detection of the mycobacterium tuberculosis Ag85 compl Electrochimica Acta, 2026
Publication

94 J G Hakim. "Double blind randomised placebo controlled trial of adjunctive prednisolor in the treatment of effusive tuberculous pericarditis in HIV seropositive patients", Hear 2000
Publication

95 Jinzhong Guo, Zaixing Jia, Yourong Yang, Nan Wang et al. "Bioinformatics analysis, immunogenicity, and therapeutic efficacy evaluation of a novel multi-stage, multi-epito DNA vaccine for tuberculosis", International Immunopharmacology, 2025
Publication

96 research-repository.griffith.edu.au
Internet Source

97 scholar.ucu.ac.ug
Internet Source

98 Catherine F. Yang, Xiaodong Feng, Brent Luu. "Clinical Applications of Nanotechnology AI Integration", CRC Press, 2026
Publication

99 Fu-Rui Liang, Yue-Hui Hong, Cong-Cong Ye, Hailin Deng, Jian-Ping Yuan, Yun-Fang Hao, Jiang-Hai Wang. "Molecular characterization and gene expression of cathepsin L in Nile tilapia (Oreochromis niloticus)", Fish & Shellfish Immunology, 2017
Publication

100 Hafiz Mamoon Rehman, Muhammad Amjad Nawaz, Zahid Hussain Shah, Gyuhwa Chur Seung Hwan Yang. "Molecular Elucidation of Two Novel Seed Specific Flavonoid Glycosyltransferases in Soybean", Journal of Plant Biology, 2018
Publication

101 Huy Hoa Hoang, Pei-Chi Wang, Shih-Chu Chen. "The protective efficacy of recombinant hypoxic response protein 1 of Nocardia seriolae in largemouth bass (Micropterus salmoides)", Vaccine, 2020
Publication

102 Onesime Mbulayi, Saint-Jean Djungu, Loukia Aketi, Mohammed Amine Koulali et al. "Tuberculosis diagnosis using artificial intelligence: current trends and future prospects Frontiers in Medicine, 2026
Publication

103 assets.thermofisher.com
Internet Source

104 erepository.mku.ac.ke
Internet Source

105 prr.hec.gov.pk
Internet Source

106 www.agrisera.com
Internet Source

107 www.one-health.panafrican-med-journal.com
Internet Source

108 "Diagnosis of Mycobacterium", Springer Science and Business Media LLC, 2023
Publication

109 "Therapeutics of Mycobacterium", Springer Science and Business Media LLC, 2026
Publication

110 Cui-Cui Liu, Yan-Xia Ru, Hui-Ru Li, De-Jian Liu, Jia-Ye Liu, Shi-Long Zhao, Xuan Wu, You-Song Peng, Yao-Wang Li, Lei Deng. "Soluble influenza H3 trimer proteins enhance the breadth and potency of antibody response", International Journal of Biological Macromolecules, 2025
Publication

111 Linda Petrone, Alessandra Aiello, Saeid Najafi-Fard, Delia Goletti. "Impact of diabetes o TB infection diagnosis", International Journal of Infectious Diseases, 2026
Publication

112 Rahmi Amtha, Ferry Sandra, Rosalina Tjandrawinata, Indrayadi Gunardi, Anggraeny Put Sekar Palupi. "Current Research and Trends in Dental and Medical Technology", CRC Press, 2025
Publication

113 Sharmila Narayanan, Pallab Sanpui, Lingaraj Sahoo, Siddhartha Sankar Ghosh. "Tobacc phytaspase: Successful expression in a heterologous system", Bioengineered, 2017
Publication

114 accedacris.ulpgc.es
Internet Source

115 helvia.uco.es
Internet Source

116 qa-eprints.qut.edu.au
Internet Source

117 www.omicsonline.org
Internet Source

118 "Pulmonary Tuberculosis and Its Prevention", Springer Science and Business Media LLC 2022
Publication

118 c.coek.info
Internet Source

120 dspace.library.uu.nl
Internet Source

121 escholarship.org
Internet Source

122 findresearcher.sdu.dk
Internet Source

123 pharmacyeducation.fip.org
Internet Source

124 preprints.epiforecasts.io
Internet Source

125 ri.conicet.gov.ar
Internet Source

126 Kyu Hwan Shim, Mino Kang, Min Gon Kim, Bong Hyun Chung, Seong Soo A. An. "Detect of E. coli O157:H7 using its endogenous active membrane peroxidase", Toxicology and Environmental Health Sciences, 2011
Publication

127 Mohammad Hassan Sadeghi, Safa Radmehr, Neda Mohagheghzadeh, Javad Fathi, Yalda Malekzadegan, Hesam Zendehdel Moghadam. "Innovative electrochemical biosensors tuberculosis detection", Clinica Chimica Acta, 2025
Publication

128 blog.labtag.com
Internet Source

129 bmcinfectdis.biomedcentral.com
Internet Source

130 journals.umt.edu.pk
Internet Source

131 pure.eur.nl
Internet Source

132 www.teses.usp.br
Internet Source

133 C. Karavina, A. Gubba. " sp. and are natural hosts of in Zimbabwe ", Plant Disease, 2017
Publication

134 Guoyan Liu, Chunyan Chai. "Towards the development of a sensitive electrochemical sensor for the determination of chloramphenicol residues in milk", Analytical Methods, 2015
Publication

135 H. Zhang. "PPE protein (Rv3425) from DNA segment RD11 of Mycobacterium tuberculo a potential B-cell antigen used for serological diagnosis to distinguish vaccinated contr from tuberculosis patients", Clinical Microbiology and Infection, 2/2007
Publication

136 Jafar Amani. "In silico analysis of chimeric espA, eae and tir fragments of Escherichia co O157:H7 for oral immunogenic applications", Theoretical Biology and Medical Modellin 2009
Publication

137 Jen-Tsung Chen. "AI in Plant Science and Precision Agriculture", CRC Press, 2026
Publication

138 Julio Cesar Ayala, Elsa Pimienta, Caridad Rodriguez, Mónica Sarzo et al. "Assessment of ELISA for serodiagnosis of active pulmonary tuberculosis in a Cuban population", Asian Pacific Journal of Tropical Disease, 2015
Publication

139 P. Méndez-Samperio. "Development of tuberculosis vaccines in clinical trials: current status", Scandinavian Journal of Immunology, 2018
Publication

140 Potsiso L. Koma, Mashilo M. Matotoka, Raymond T. Makola, Thabiso Mphephu, Vinayak Singh, Peter Masoko. "Antioxidant, anti-inflammatory, antimycobacterial activities, and cytotoxic effects of medicinal plants and their potential mechanisms of action against Mycobacterium smegmatis", South African Journal of Botany, 2026
Publication

141 Pradeep Neupane, Ricardo G. Maggi, Manoj Basnet, Erin Lashnits, Gerard P. Andrews, Edward B. Breitschwerdt. "Bartonella henselae Recombinant Pap31 for the Diagnosis o Canine and Human Bartonelloses", Pathogens, 2022
Publication

142 Tyler Morrison, Danielle A. Gunn-Moore, Jayne C. Hope, Conor O'Halloran. "Cytokine an Chemokine Responses of Peripheral Blood Mononuclear Cells from Dogs Infected with Mycobacterium bovis", Pathogens, 2024
Publication

143 Yun Xu, Xiaochun Wang, Qiangsen Zhong. "Diagnostic potential of recombinant Mycobacterium tuberculosis PcaA antigen and its enhancement of protective efficacy a subunit vaccine booster following BCG priming", Journal of Microbiological Methods, 2(
Publication

144 api.mountainscholar.org
Internet Source

145 bmcpsychiatry.biomedcentral.com
Internet Source

146 books.ersjournals.com
Internet Source

147 cardiab.biomedcentral.com
Internet Source

148 cdr.lib.unc.edu
Internet Source

149 dash.harvard.edu
Internet Source

150 docslib.org
Internet Source

151 ophrp.org
Internet Source

152 repository.nwu.ac.za
Internet Source

153 tessera.spandidos-publications.com
Internet Source

154 tsukuba.repo.nii.ac.jp
Internet Source

155 virologyj.biomedcentral.com
Internet Source

156 www.ideals.illinois.edu
Internet Source

157 www.jove.com
Internet Source

158 www.tecnolab.com.ar
Internet Source

159 "Immunoinformatics", Springer Science and Business Media LLC, 2020
Publication

160 Arbind Kumar, Aashish Sharma, Gurkamaljit Kaur, Pooja Makkar, Jagdeep Kaur. " Functional characterization of hypothetical proteins of with possible esterase/lipas signature: a cumulative and approach ", Journal of Biomolecular Structure and Dyn 2016
Publication

161 Camila P. Sobrinho, Luana E. Araújo, Rodrigo L. Meira, Thainá Horta et al. "Novel pl biomarkers for the diagnosis of pulmonary tuberculosis", Tuberculosis, 2026
Publication

162 Chen, Tao, Jinfei Lin, Wei Wang, Joy Fleming, Liang Chen, Yunxia Wang, Haicheng Li, Guo, Jie Zhou, Xunxun Chen, Yuhui Chen, Qinghua Liao, Yang Shu, Yaoju Tan, Meili Guozhou Li, Lin Zhou, Qiu Zhong, Lijun Bi, Lina Guo, and Meigui Zhao. "Cytokine an Antibody Based Diagnostic Algorithms for Sputum Culture-Positive Pulmonary Tuberculosis", PLoS ONE, 2015.
Publication

163 Daniil M. Prigozhin, Daniela Mavrici, John P. Huizar, Hilary J. Vansell, Tom Alber. " Structural and Biochemical Analyses of -Acetylmuramyl-l-alanine Amidase Rv3717 F a Role in Peptidoglycan Fragment Recycling ", Journal of Biological Chemistry, 2013
Publication

164 Flavia Anna Mercurio, Daniela Marasco, Luciano Pirone, Emilia Maria Pedone, Maur Pellecchia, Marilisa Leone. "Solution Structure of the First Sam Domain of Odin and Binding Studies with the EphA2 Receptor", Biochemistry, 2012
Publication

165 G. Morsica. "Natural Killer-Cell Cytotoxicity in HIV-Positive and HIV-Negative Patient and Without Severe Course of Hepatitis B Virus Infection", Scandinavian Journal of Immunology, 9/2005
Publication

166 Manika Sharma, Devanshu Kumar, Ayushi Kaur Bedi, Pooja Yadav et al. "A novel multiepitope peptide vaccine with intrinsic adjuvant confers protective immunity in against tuberculosis", Microbial Pathogenesis, 2026
Publication



167 Mohamed Imdhiyas, Suvam Sen, Nagendra Barman, Lukumoni Buragohain, Yashpal Malik, Sachin Kumar. "Computational analysis of immunogenic epitopes in the p30 and p54 proteins of African swine fever virus", Journal of Biomolecular Structure and Dynamics, 2022
Publication

168 Mohammed A. Gafar, Calvin A. Omolo, Usri H. Ibrahim, Xylia Q. Peters, Eman A. Ismail, Rene Khan, Thirumala Govender. "Antimicrobial peptide-fucoidan nanoplexes: A novel multifunctional biomimetic nanocarrier for enhanced vancomycin delivery against bacterial infections and sepsis", International Journal of Pharmaceutics, 2025
Publication

169 S. Ozgur, M. Chekulaeva, G. Stoecklin. "Human Pat1b Connects Deadenylation with mR Decapping and Controls the Assembly of Processing Bodies", Molecular and Cellular Biology, 2010
Publication

170 Singh, Krishnak., Naresh Sharma, Diana Vargas, Zhentong Liu, JohnT. Belisle, Visalakshi Potharaju, Ajay Wanchu, Digambar Behera, and Suman Laal. "Peptides of a Novel *Mycobacterium tuberculosis*-Specific Cell Wall Protein for Immunodiagnosis of Tuberculosis", The Journal of Infectious Diseases, 2009.
Publication

171 Sunit K. Singh, Daniel Růžek. "Neuroviral Infections - General Principles and DNA Viruse CRC Press, 2013
Publication

172 Yoshinaga, M.. "Glutathione transferases with vanadium-binding activity isolated from vanadium-rich ascidian Ascidia sydneiensis samea", BBA - General Subjects, 200603
Publication

173 actaneurocomms.biomedcentral.com
Internet Source

174 coek.info
Internet Source

175 dspace.cuni.cz
Internet Source

176 public-pages-files-2025.por-journal.com
Internet Source

177 researchonline.lshtm.ac.uk
Internet Source

178 researchrepository.ucd.ie
Internet Source

ses.library.usyd.edu.au



179
Internet Source

180 univendspace.univen.ac.za
Internet Source

181 unsworks.unsw.edu.au
Internet Source

182 www.ghspjournal.org
Internet Source

183 www.jornaldepneumologia.com.br
Internet Source

184 www.redetb.hucff.ufrj.br
Internet Source

185 www.tdx.cat
Internet Source

186 "Tuberculosis Vaccines", Springer Science and Business Media LLC, 2025
Publication

187 Anuradha Rajamanickam, Evangeline Ann Daniel, Nikhil Gupte, Kannan Thiruvengadam al. "Plasma biomarkers CRP, iFABP, and zonulin as predictors of tuberculosis progressi in household contacts of pulmonary TB patients", Tuberculosis, 2026
Publication

188 Ashkan Safavi, Amirhosein Kefayat, Elham Mahdevar, Ardavan Abiri, Fatemeh Ghahremani. "Exploring the out of sight antigens of SARS-CoV-2 to design a candidate multi-epitope vaccine by utilizing immunoinformatics approaches", Vaccine, 2020
Publication

189 Fang Yang, Yingying Pan, Yazhou Chen, Shiming Tan, Mingfei Jin, Zirong Wu, Jing Huang "Expression and purification of Canis interferon α in Escherichia coli using different tag Protein Expression and Purification, 2015
Publication

190 Giggil Pushpamithran. "Modulatory effects of helminth coinfection on host immune responses to Mycobacterium tuberculosis : examining potential outcomes", Linkoping University Electronic Press, 2024
Publication

191 Gizeaddis Belay, Hailu Getachew, Tigist Birku, Aimro Tadese et al. "Rifampicin-resistant Mycobacterium tuberculosis and unsuccessful results from Xpert® MTB/Rif-Ultra assay Amhara Region, Ethiopia", Journal of Clinical Tuberculosis and Other Mycobacterial Diseases, 2025
Publication



192 Ian Robertson, Sacha Jensen, Penny Handford. "TB domain proteins: evolutionary insig into the multifaceted roles of fibrillins and LTBPs", Biochemical Journal, 2010
Publication

193 Irina Kontsevaya, Andrea Maurizio Cabibbe, Daniela Maria Cirillo, Andrew R. DiNardo e "Update on the diagnosis of tuberculosis", Clinical Microbiology and Infection, 2023
Publication

194 Ji, C.. "Expression and in vitro activation of Manduca sexta prophenoloxidase-activating proteinase-2 precursor (proPAP-2) from baculovirus-infected insect cells", Protein Expression and Purification, 200306
Publication

195 Kajal, Diksha Sharma, Rohit Rai. "Neoteric advancements in TB diagnostics and its futui frame", Indian Journal of Tuberculosis, 2021
Publication

196 Kathryn C. Rahlwes, Beatriz R.S. Dias, Priscila C. Campos, Samuel Alvarez-Arguedas, Michael U. Shiloh. " Pathogenicity and virulence of ", Virulence, 2023
Publication

197 Khaled Mohamed Sefow Al-arbi, Nombulelo P. Magula, Girish M. Mody. "Tuberculosis remains a major burden in systemic lupus erythematosus patients in Durban, South Africa", Frontiers in Medicine, 2023
Publication

198 Kim A. Woodrow, Isoken O. Airen, James R. Swartz. "Rapid Expression of Functional Genomic Libraries", Journal of Proteome Research, 2006
Publication

199 Louansha Nandlal, Rubeshan Perumal, Kogieleum Naidoo. "Rapid Molecular Assays for the Diagnosis of Drug-Resistant Tuberculosis", Infection and Drug Resistance, 2022
Publication

200 Lucinda Baatjies, Andre G. Loxton, Monique J. Williams. "Host and Bacterial Iron Homeostasis, an Underexplored Area in Tuberculosis Biomarker Research", Frontiers i Immunology, 2021
Publication

201 Marzieh Asadi, Sina Taghvimi, Ghazal Ghaznavi, Halime Parsaee et al. "Design and evaluation of a novel fusion antigen for diagnosing human strongyloidiasis: An immunoinformatics approach", PLOS One, 2026
Publication

202 Min Li, Haoxiang Wang, Bo Huang, Yuebo Jin, Qingwen Wang, Jing He. "Serum epitope-specific anti-recoverin autoantibodies as biomarkers for lupus retinopathy", Arthritis Research & Therapy, 2026
Publication



203 Miraj ud din, Xiaohui Liu, Hui Jiang, Sajjad ahmad, Xiangdong Lai, Xuemei Wang. "Advancing vaccine development in genomic era: A paradigm shift in vaccine discovery Progress in Biomedical Engineering, 2025
Publication

204 Nan Jia, Chaohong Wang, Xiaming Liu, Xiaolan Huang, Fei Xiao, Jin Fu, Chunrong Sun, Zheng Xu, Guirong Wang, Juan Zhou, Yi Wang. "A CRISPR-Cas12a-based platform for ultrasensitive rapid highly specific detection of Mycobacterium tuberculosis in clinical application", Frontiers in Cellular and Infection Microbiology, 2023
Publication

205 Pedro M. Alzari, H. Berglund, N. S. Berrow, E. Blagova et al. "Implementation of semi-automated cloning and prokaryotic expression screening: the impact of SPINE", Acta Crystallographica Section D Biological Crystallography, 2006
Publication

206 Pragati Agnihotri, Anil Kumar Shakya, Arjun K. Mishra, J. Venkatesh Pratap. "Crystal structure and characterization of nucleoside diphosphate kinase from Vibrio cholerae" Biochimie, 2021
Publication

207 Radha Gopalaswamy, Selvakumar Subbian. "Intricacies of Global Tuberculosis Management—EndTB-2035 on the Fence?", Journal of Respiration, 2025
Publication

208 Rivesh Maharajh, Manormoney Pillay, Sibusiso Senzani. " A computational method for prediction and functional analysis of potential adhesin-related proteins ", Expert Revie Proteomics, 2023
Publication

209 Saiyur Ramsugit, Sinenhlanhla Guma, Balakrishna Pillay, Paras Jain, Michelle R. Larsen, Siva Danaviah, Manormoney Pillay. "Pili contribute to biofilm formation in vitro in Mycobacterium tuberculosis", Antonie van Leeuwenhoek, 2013
Publication

210 Soumitra Mohanty, Lakshmanan Jagannathan, Geetanjali Ganguli, Avinash Padhi et al. Mycobacterial Phosphoribosyltransferase Promotes Bacillary Survival by Inhibiting Oxidative Stress and Autophagy Pathways in Macrophages and Zebrafish", Journal of Biological Chemistry, 2015
Publication

211 Sumera Zaib, Nehal Rana, Areeba, Imtiaz Khan et al. "Designing multi-epitope monkeyp virus-specific vaccine using immunoinformatics approach", Journal of Infection and Pul Health, 2022
Publication

212 aa-woonbiologie.nl
Internet Source



213 aacrjournals.org
Internet Source

214 bioresources.cnr.ncsu.edu
Internet Source

215 core.ac.uk
Internet Source

216 d197for5662m48.cloudfront.net
Internet Source

217 ebuah.uah.es
Internet Source

218 etd.aau.edu.et
Internet Source

219 gutpathogens.biomedcentral.com
Internet Source

220 hal.science
Internet Source

221 onlinelibrary.wiley.com
Internet Source

222 pearl-prod.plymouth.ac.uk
Internet Source

223 pure.manchester.ac.uk
Internet Source

224 pure.mpg.de
Internet Source

225 pure.tue.nl
Internet Source

226 qspace.library.queensu.ca
Internet Source

227 researchrepository.ru.ac.za
Internet Source

228 udspace.udel.edu
Internet Source

229 www.arca.fiocruz.br
Internet Source

www.besjournal.com



230
Internet Source

231 www.imrpress.com
Internet Source

232 www.scielo.br
Internet Source

233 www.science.org
Internet Source

234 www.tandfonline.com
Internet Source

235 www.zgflzz.cn
Internet Source

236 A.H. Alvarez, M.A. Flores-Valdez. " Can immunization with Bacillus Calmette-Guérin be improved for prevention or therapy and elimination of chronic infection? ", Expert Revi of Vaccines, 2019
Publication

237 Anzaan Dippenaar, Nabila Ismail, Melanie Grobbelaar, Selien Oostvogels et al. "Optimiz liquefaction and decontamination of sputum for DNA extraction from Mycobacterium tuberculosis", Tuberculosis, 2021
Publication

238 Bergey's Manual® of Systematic Bacteriology, 2001.
Publication

239 Claudia Carranza, Sigifredo Pedraza-Sanchez, Eleane de Oyarzabal-Mendez, Martha Torres. "Diagnosis for Latent Tuberculosis Infection: New Alternatives", Frontiers in Immunology, 2020
Publication

240 G. Sachdeva. "SPAAN: a software program for prediction of adhesins and adhesin-like proteins using neural networks", Bioinformatics, 09/16/2004
Publication

241 Haoyan Yang, Xinkui Lei, Siyu Chai, Sigen Zhang, Guimin Su, Lin Du. "A multi-epitope vaccine incorporating adhesin-derived antigens protects against Mycobacterium tuberculosis infection and dissemination", Frontiers in Immunology, 2025
Publication

242 Jalaledin Ghanavi, Poopak Farnia, Saeid Besharati, Parissa Farnia, Ali Akbar Velayati. "TI design for drug delivery using pili in Mycobacterium tuberculosis", Elsevier BV, 2024
Publication

243 Kayode T. Adu, Richard Wilson, Anthony L. Baker, John Bowman, Margaret L. Britz. " Prolonged Heat Stress of GCRL163 Improves Binding to Human Colorectal



Adenocarcinoma HT-29 Cells and Modulates the Relative Abundance of Secreted and C Surface-Located Proteins ", Journal of Proteome Research, 2020
Publication

244 Mahboubeh Zarei, Mohammad Reza Rahbar, Navid Nezafat, Manica Negahdaripour, Mohammad Hossein Morowvat, Younes Ghasemi. "Computational Analysis of Arginine Deiminase Sequences to Provide a Guideline for Protein Engineering", Current Proteomics, 2020
Publication

245 Olfa Frikha-Gargouri. "Evaluation of an in silico predicted specific and immunogenic antigen from the OmcB protein for the serodiagnosis of Chlamydia trachomatis infections", BMC Microbiology, 2008
Publication

246 Susan E Purchase, Joanna Brigden, James A Seddon, Neil A Martinson et al. "Risk of Tuberculosis Infection in Young Children Exposed to Multidrug-resistant Tuberculosis i the TB-CHAMP Multi-site Randomized Controlled Trial", Clinical Infectious Diseases, 20
Publication

247 Wallis, R.S.. "Biomarkers and diagnostics for tuberculosis: progress, needs, and translat into practice", The Lancet, 20100529/0604
Publication

248 Yingxiao Han, Songwei Ma, Rui Ma, Jingjing Li et al. "Enhancing detection efficiency of R proteins in Mycobacterium tuberculosis culture filtrate by biomimetic affinity chromatography coupled with LC-MS/MS", Journal of Chromatography B, 2026
Publication

249 "8th Course of the European School of Neuroimmunology", Journal of Neuroimmunolo 20081025
Publication

250 "Computational Vaccine Design", Springer Science and Business Media LLC, 2023
Publication

251 "In vitro Regulation of Activation with Human Telomerase Reverse Transcriptase Components Expressed in Escherichia coli and Human Telomerase RNA Component", Bulletin of the Korean Chemical Society, 2008
Publication

252 Edita Hansted, Angele Andriuskeviciene, Raimundas Sakalauskas, Rimantas Kevalas, Brigita Sitkauskiene. "T-cell-based diagnosis of tuberculosis infection in children in Lithuania: a country of high incidence despite a high coverage with bacille Calmette-Guerin vaccination", BMC Pulmonary Medicine, 2009
Publication

253 Esther van Woudenbergh, Edward B. Irvine, Leela Davies, Marwou de Kock et al. "HIV Is Associated with Modified Humoral Immune Responses in the Setting of HIV/TB



Coinfection", mSphere, 2020
Publication

254 Gonçalo F. Augusto, Sara S. Dias, Alexandre V. Abrantes, Maria R. O. Martins. "HIV/AIDS length of stay in Portugal under financial constraints: a longitudinal study for public hospitals, 2009–2014", BMC Health Services Research, 2019
Publication

255 Haoyan Yang, Yinuo Ma, Xinkui Lei, Siyu Chai, Sigen Zhang, Guimin Su, Songping Li, Lin "Stopping Tuberculosis at the Gate: The Role of M. tuberculosis Adhesins in Infection a Intervention", Vaccines, 2025
Publication

256 Jinjie Yu, Xueting Fan, Xiuli Luan, Ruihuan Wang et al. "A novel multi-component protei vaccine ECP001 containing a protein polypeptide antigen nPst51 riching in T-cell epitop showed good immunogenicity and protection in mice", Frontiers in Immunology, 2023
Publication

257 Lin-Sheng Li, Ling Yang, Li Zhuang, Zhao-Yang Ye, Wei-Guo Zhao, Wen-Ping Gong. "Fron immunology to artificial intelligence: revolutionizing latent tuberculosis infection diagn with machine learning", Military Medical Research, 2023
Publication

258 Sibusiso B. Maseko, Satheesh Natarajan, Vikas Sharma, Neelakshi Bhattacharyya et al. "Purification and characterization of naturally occurring HIV-1 (South African subtype C protease mutants from inclusion bodies", Protein Expression and Purification, 2016
Publication

259 Systems Biology of Tuberculosis, 2013.
Publication

260 openaccess.sgul.ac.uk
Internet Source