LIBRARY
Te Tumu Herenga
THE UNIVERSITY OF AUCKLAND

RESEARCHSPACE@AUCKLAND

Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand). This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

To request permissions please use the Feedback form on our webpage.
http://researchspace.auckland.ac.nz/feedback

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the Library Thesis Consent Form

STRUCTURAL GENOMICS

OF

MYCOBACTERIUM TUBERCULOSIS

By

Jodie Margaret Johnston

A thesis submitted in partial fulfilment of the
requirements for the degree of

Doctor of Philosophy

University of Auckland, 2004

ABSTRACT

In 1998 the genome sequence of *Mycobacterium tuberculosis* H37Rv was published[1]. *M. tuberculosis* is the primary causative agent of tuberculosis, a disease with a long history in humans, which still has a great impact on human mortality today. As part of the *M. tuberculosis* Structural Genomics Consortium we selected nine target genes (Rv0534c (*menA*); Rv0548c (*menB*); Rv0553 (*menC*); Rv0555 (*menD*); Rv0542c (*menE*); Rv3853 (*menG*); Rv0558 (*ubiE*); Rv0989c (grcC2) and Rv0990c) from *M. tuberculosis*, including all known members of the menaquinone biosynthesis pathway, for structural studies.

All nine genes were taken through the structural genomics "pipeline", either becoming stuck at various "bottlenecks" or continuing successfully to structure solution. At the initial bioinformatics analysis step, eight of the nine targeted genes were deemed suitable for further study. PCR amplification and cloning of these genes into several different expression vectors followed. Expression of the gene products for the seven successfully cloned genes was undertaken in an *E. coli* expression host, followed by experiments (refolding, lysis buffer and expression temperature screens) aimed at obtaining soluble protein in sufficient quantities for crystallisation. Of the seven proteins successfully overexpressed, five remain at this stage as they could not be obtained in soluble form. The remaining two, Rv3853 (MenG), solubilised by refolding, and MenB, solubilised by 24°C expression, were purified and both successfully produced diffracting crystals. The crystal structure of Rv3853 was determined by isomorphous replacement (SIRAS) and refined at 1.9 Å resolution (R = 19.0% and R_{free} = 22.0%). The structure of several different crystal forms of MenB, were determined by molecular replacement. Refinement of two of these structures, MenB_P4_32_12 at 2.15Å resolution (R = 20.3% and R_{free} = 23.1%) and MenB_C2-NCoA at 2.3 Å resolution (R = 19.7% and R_{free} = 22.5%), has been completed.

The structure of Rv3853, combined with the discovery that UbiE was more likely to catalyse the final, *S*-adenosylmethionine-dependent, methyltransfer step of menaquinone biosynthesis, led to the conclusion that Rv3853 had been misannotated as MenG. Combined with further bioinformatics analysis the Rv3853 structure has been useful in providing new ideas as to the real function of Rv3853. In contrast, the structure of MenB confirmed its place as a member of the crotonase superfamily although the C-terminus was located in a position not observed in other crotonase superfamily structures. Several flexible regions likely to be important in MenB function have been identified by examination of the various MenB structures.

ACKNOWLEDGEMENTS

I feel grateful to have worked with so many great people, in an environment so conducive to learning over the past few years. I would like to thank my supervisors Professor Ted Baker and Dr. Vickery Arcus for all their time, advice, support and encouragement over the years. I would especially like to thank Ted for providing such an interesting project to work on as well as a great laboratory in which to work on it. To Vic, I gratefully acknowledge all the time spent over the last four years showing me the "the ropes" and the infectious enthusiasm with which every new discovery along the way was greeted, thank-you.

To me the sharing of knowledge is one of the best gifts you can give, and in that vein, along with my supervisors, I would also like to acknowledge a number of people who have taught me a lot over the past four years. To, Mrs Heather Baker, Dr. Shaun Lott, Dr. Chris Squire, Dr. Brian Anderson, Dr. Graham Card, Dr. James Dickson, Dr. Clyde Smith, David Goldstone, Dr. Peter Haebel, Prof. Peter Metcalf, thank-you, your efforts have been much appreciated. Furthermore, I would like to thank Drs. Li-Wei Hung and Leonid Flaks, Professor Tom Terwilliger and the *M. tuberculosis* Structural Genomics Consortium as well as collaborators on the MenG project, Drs. Michael Parker and Craig Morton, for their contributions. Also thank-you to my colleagues; Julie, Nina, Miriam, Roberta, Moyra, Randall, Harriet, Caroline, Rochelle, Erica, Nayden, Shin-May, Elaine, Peter, Richard, Rachael, Ivan, Steve, Neil, Simon, Anthony, John, Jess, Eddie, Mel, Claire, Vern, Annette, Mattias, Jenny, and Andrew for their support and encouragement over the time I have known each of them. And in memory of a colleague, Rainer Knijff, who sadly is no longer with us.

None of this would have been possible for me without funding and I gratefully acknowledge the financial support I have received from the Foundation of Research, Science and Technology in the form of a Bright Futures Top Achiever Doctoral Scholarship and from the University of Auckland in the form of an Auckland University Doctoral Scholarship. I also want to acknowledge the funding for this work provided by the New Economy Research Fund of New Zealand and the Health Research Council of New Zealand.

And lastly I would like to thank my family, Roger, Diane, Jessica, Michael, and James, who are always there for me with encouragement, unconditional love and, on occasions, cheese scones, tea and technical support. Also to my family in-law, many of whom are teachers, I could not have found a nicer bunch.

TABLE OF CONTENTS

CHAPTER 1 1

INTRODUCTION 1

1.1 STRUCTURAL GENOMICS 1
1.1.1 GENOMIC PROJECTS: A WEALTH OF SEQUENCE INFORMATION 1
1.1.2 DEFINING STRUCTURAL GENOMICS 1
1.1.3 THREE MAIN OUTCOMES OF STRUCTURAL GENOMICS 2
1.1.3.1 Providing Structural Templates and Mapping Fold Space 2
1.1.3.2 Using Structural Information to Aid Functional Annotation 4
1.1.3.3 Structural Genomics and Drug Design 7
1.1.4 TECHNICAL CONSIDERATIONS 8
1.1.4.1 A Typical Structural Genomics Pipeline 9
1.1.5 CURRENT STRUCTURAL GENOMICS PROJECTS 10
1.1.6 THE *MYCOBACTERIUM TUBERCULOSIS* STRUCTURAL GENOMICS CONSORTIUM (MTBSGC) 12
1.1.6.1 Beginnings: The Genome Sequence 12
1.1.6.2 The Setup and Facilities 12
1.1.6.3 The Aims 12
1.2 TUBERCULOSIS 13
1.2.1 THE DISEASE: AN OVERVIEW 13
1.2.2 THE CAUSATIVE AGENT: *MYCOBACTERIUM TUBERCULOSIS* 14
1.2.3 THE DISEASE: ANCIENT ORIGINS 15
1.2.4 THE DISEASE: INFECTION AND PROGRESSION 15
1.2.5 THE DISEASE: THE LATENT/PERSISTENT/DORMANT STATE 17
1.2.5.1 Models of Latent Disease 18
1.2.6 THE DISEASE: A RECENT HISTORY 22
1.2.6.1 Tuberculosis and HIV 23
1.2.7 THE DISEASE: TREATMENT AND MULTIPLE DRUG RESISTANCE (MDR) 23
1.2.7.1 Treatment of Tuberculosis: A Brief History 23
1.2.7.2 Multiple Drug Resistance (MDR) 25
1.2.7.3 Drug Development Efforts in Recent Years 26
1.3 *MYCOBACTERIUM TUBERCULOSIS*, HYPOXIA AND MENAQUINONE 27
1.4 MENAQUINONE BIOSYNTHESIS 28
1.4.1 MENAQUINONE, UBIQUINONE AND ELECTRON TRANSPORT 28
1.4.1.1 Quinones: An Introduction 28
1.4.1.2 Roles of Ubiquinone and Menaquinone in Bacteria 31
1.4.2 MENAQUINONE AND ELECTRON TRANSPORT IN *MYCOBACTERIUM TUBERCULOSIS* 35
1.4.3 THE MENAQUINONE BIOSYNTHESIS PATHWAY 39
1.4.3.1 Formation of *O*-Succinyl Benzoic Acid (OSB): MenD, MenC and MenF 42
1.4.3.2 Formation of 1,4-Dihydroxy-2-Naphthoic Acid (DHNA): MenB, MenE and MenH 43
1.4.3.3 Prenylation: MenA 46
1.4.3.4 Methylation: UbiE/MenH/MenG or MenG 47
3.5 THE MENAQUINONE BIOSYNTHESIS PATHWAY OF *MYCOBACTERIUM TUBERCULOSIS* 48
1.5 AIMS OF THE PROJECT 49

CHAPTER 2 52

GENERAL METHODS 52

2.1 BIOINFORMATICS 52
2.1.1 PROTEIN AND NUCLEOTIDE INFORMATION 52
2.1.1.1 The *Mycobacterium tuberculosis* Structural Genomics Consortium Database 52
2.1.1.2 Other Databases 52
2.1.2 PRIMER DESIGN 53
2.1.3 SEARCHING FOR SEQUENCE HOMOLOGUES USING BLAST AND PSI-BLAST 53

2.1.4 SEQUENCE ALIGNMENTS 54
2.2 DNA-RELATED MATERIALS AND PROTOCOLS 54
2.2.1 GENERAL DNA MATERIALS AND METHODS 54
2.2.1.1 Primers 54
2.2.1.2 DNA Standards 56
2.2.1.3 DNA Buffers and Loading dye 56
2.2.1.4 Agarose Gels and DNA Electrophoresis 56
2.2.1.5 Restriction Enzymes 57
2.2.1.6 Plasmid Vectors 57
2.2.1.7 Antibiotics 58
2.2.1.8 Bacterial Cell Culture Media 58
2.2.1.9 Bacterial Cloning Strain, Bacterial Plate Storage and Glycerol Stocks 58
2.2.2 POLYMERASE CHAIN REACTION (PCR) AMPLIFICATION OF TARGET SEQUENCES 58
2.2.3 PURIFICATION OF DNA FRAGMENTS AND CONCENTRATION OF DNA 60
2.2.4 DIGESTION OF DNA WITH RESTRICTION ENDONUCLEASES 60
2.2.5 DNA LIGATION 61
2.2.6 PREPARATION OF ELECTRO-COMPETENT CELLS 62
2.2.7 TRANSFORMATION OF DNA 62
2.2.8 ANALYSIS OF TRANSFORMED COLONIES BY COLONY PCR 62
2.2.9 PURIFICATION OF PLASMID DNA (MINI, MIDI AND MAXI PREPS) 63
2.2.9.1 Further DNA Purification Steps for Digested or Plasmid DNA 63
2.2.10 DNA SEQUENCING 64
2.3 PROTEIN EXPRESSION AND PURIFICATION MATERIALS AND PROTOCOLS 65
2.3.1 SDS-POLYACRYLAMIDE GEL ELECTROPHORESIS (SDS-PAGE) 65
2.3.1.1 Gel Preparation 65
2.3.1.2 Buffers and Stains 66
2.3.1.3 Protein Standards 66
2.3.1.4 Sample Preparation, Running Gels and Gel Staining 67
2.3.2 NATIVE-POLYACRYLAMIDE GEL ELECTROPHORESIS (NATIVE-PAGE) 68
2.3.3 BACTERIAL CELL CULTURE FOR PROTEIN EXPRESSION 68
2.3.3.1 Bacterial Cell Culture Media 68
2.3.3.2 Bacterial Expression Strains 68
2.3.3.3 DNA Transformation into the Expression Strains 68
2.3.4 SMALL SCALE PROTEIN EXPRESSION TESTS AND SOLUBILITY TESTING 69
2.3.4.1 Protein Expression and Solubility Tests 69
2.3.4.2 Solubility Screen Lysis Buffers 70
2.3.5 LARGE SCALE EXPRESSION 72
2.3.6 SOLUBLE PROTEIN EXTRACTION 73
2.3.7 INSOLUBLE PROTEIN EXTRACTION 73
2.3.8 PROTEIN REFOLDING 74
2.3.8.1 Refolding Solutions 74
2.3.8.2 The Initial Small-Scale Refolding Screen 76
2.3.8.3 Large Scale Refolding 77
2.3.9 Ni^{2+} AFFINITY PROTEIN PURIFICATION 77
2.3.9.1 Small Scale Batch Ni^{2+} Affinity Binding Tests 77
2.3.9.2 Immobilised Metal Affinity Chromatography (IMAC) 78
2.3.10 GST AFFINITY PROTEIN PURIFICATION 80
2.3.10.1 GST Affinity Batch Purification and IMAC as Used for MenE 81
2.3.11 PROTEIN PURIFICATION BY SIZE EXCLUSION CHROMATOGRAPHY 82
2.3.12 PURIFICATION BY ANION EXCHANGE CHROMATOGRAPHY ON A MONO-Q COLUMN 82
2.3.13 CONCENTRATION OF PURIFIED PROTEIN 83
2.3.13.1 Protein Concentration Estimation 83
2.3.13.2 Concentration 83
2.3.13.3 Buffer Exchange Using a Concentrator 84
2.3.14 PROTEIN ANALYSIS 84
2.3.14.1 N-terminal sequencing 84

2.3.14.2 Mass Spectrum Analysis 84
2.3.14.3 Dynamic Light Scattering 85
2.3.14.4 Small Scale Dialysis Experiments (Undertaken for MenB) 85
2.4 GENERAL PROTEIN CRYSTALLISATION MATERIALS AND PROTOCOLS 86
2.4.1 CRYSTOOL PRE-SCREEN 86
2.4.2 HANGING DROP METHOD 86
2.4.2.1 Preparation of Siliconised Coverslips 86
2.4.2.2 Materials and General Hanging Drop Method 86
2.4.2.3 Whisker Seeding in Hanging Drop Format 87
2.4.3 OIL BATCH METHOD 87
2.4.4 SITTING DROP METHOD 87
2.4.5 CRYSTAL SCORING SYSTEM 88
2.4.6 CRYSTALS SCREENS AND CRYSTAL OPTIMISATION 88
2.4.6.1 Crystal Screens 88
2.4.6.2 Crystal Optimisation Strategy 89
2.4.7 CRYSTAL SOAKS AND CO-CRYSTALLISATION EXPERIMENTS 90
2.4.7.1 Soaks 90
2.4.7.2 Co-crystallisation Experiments 90
2.5 CRYSTAL MOUNTING, FREEZING AND X-RAY DATA COLLECTION METHODS 90
2.5.1 ROOM TEMPERATURE MOUNTING 90
2.5.2 CRYO-MOUNTING 91
2.5.2.1 Cryoprotectant Testing 91
2.5.2.2 Flash-freezing Crystals 91
2.5.3 X-RAY DATA COLLECTION 92
2.5.3.1 Home Source Data Collection 92
2.5.3.2 Synchrotron Data Collection 92
2.6 GENERAL DATA PROCESSING AND STRUCTURE SOLUTION METHODS 93
2.6.1 DATA PROCESSING 93
2.6.1.1 General Data Collection and Processing Strategy 93
2.6.1.2 Integration and Scaling with DENZO and SCALEPACK 93
2.6.1.3 Matthews Analysis 95
2.6.2 STRUCTURE SOLUTION USING MIR/SIR/SIRAS 95
2.6.2.1 General Procedure 95
2.6.2.2 SOLVE and RESOLVE 96
2.6.3 STRUCTURE SOLUTION USING MOLECULAR REPLACEMENT (MR) 98
2.6.3.1 General Procedure 99
2.6.3.2 Molecular Replacement with MOLREP and AMoRe 100
2.6.3.3 Subsequent Treatment of the Molecular Replacement Model 100
2.6.4 MODEL BUILDING AND REFINEMENT 101
2.6.4.1 Autobuilding using ARP/wARP 101
2.6.4.2 Refinement in CNS 101
2.6.4.3 Manual Building Using O 103
2.6.5 MODEL AND STRUCTURE FACTOR CHECKING 104
2.6.6 LIGAND GENERATION: HIC-UP 104
2.6.7 STRUCTURAL ANALYSIS 105
2.6.7.1 Structural Overlays 105
2.6.7.2 Useful Structural Databases 105
2.6.7.3 Hydrogen Bonds, Contacts and Interface Interactions 105

CHAPTER 3 107

RV3853: "MENG" 107
3.1 INTRODUCTION 107
3.2 METHODS AND RESULTS 111
3.2.1 A NOTE ON METHODS 111
3.2.2 BIOINFORMATICS ANALYSIS 111

3.2.3 Primer Design, PCR Amplification and Cloning 113
3.2.4 Protein Expression and Solubility Tests for His-tagged Rv3853 113
3.2.5 Protein Expression and Solubility for GST-fusion Rv3853 115
3.2.6 His-tagged Rv3853: Small Scale Refolding Tests 115
3.2.6.1 Small Scale Denaturing Purification 115
3.2.6.2 Initial Refolding Tests 116
3.2.6.3 Larger-scale Refolding Tests 116
3.2.7 His-tagged Rv3853: Preparative Refolding and Protein Purification 118
3.2.7.1 Large Scale Denaturing Purification and Preparative Refolding 118
3.2.7.2 Further Purification of Refolded Protein 119
3.2.8 Crystallisation 120
3.2.8.1 Initial Crystallisation Trials 120
3.2.8.2 Crystal Optimisation 121
3.2.9 Preparation of Heavy Atoms Derivatives 122
3.2.9.1 Native gels: Phastgel Heavy Metal Screening Analysis 122
3.2.9.2 Heavy Metal Soaks 122
3.2.10 X-Ray Data Collection 123
3.2.11 Data Processing 123
3.2.11.1 Processing of Native Data 123
3.2.11.2 Unit Cell Dimensions and Number of Molecules in the Asymmetric Unit 124
3.2.11.3 Processing of Derivative Data 124
3.2.11.4 Summary of Data Collection and Processing Statistics 125
3.2.12 Phase Determination and Improvement: SOLVE and RESOLVE 126
3.2.13 Autobuilding Using ARP/wARP 126
3.2.14 Refinement and Further Model Building 127
3.2.15 Model Completeness and Quality 129
3.3 Structural Analysis 130
3.3.1 Overview of the Structure 130
3.3.2 Monomer Fold 131
3.3.2.1 Structural Comparisons of the Monomer Fold 132
3.3.3 Quaternary Structure 133
3.3.4 Sequence Conservation in Rv3853 Homologues 134
3.3.5 Small Molecule Binding Sites 136
3.3.6 Binding Sites Suggested from *in silico* Analysis of the Rv3853 Structure 138
3.4 Discussion 140
3.4.1 Is Rv3853 the Methyltransferase MenG? 140
3.4.2 Clues to the Biochemical Function 142
3.4.2.1 Genetic Studies 142
3.4.2.2 Clues from Sequence Homology 142
3.4.2.3 Clues from the Structure: Mapping Conserved Resides and Investigating the Binding Sites ... 146
3.4.2.4 Clues from the Structure: Structural Homologues 148
3.4.3 The Structure of *E. coli* RraA: A Key Piece to the Puzzle 149
3.4.3.1 The Annotation of *E. coli* MenG as RraA: A Background 149
3.4.3.2 Structural Comparison of RraA and Rv3853: A New Functional Annotation! 149
3.4.4 Summary 152
3.4.5 The Next Step: Future Directions 152

CHAPTER 4 154

MENB 154
4.1 Introduction 154
4.2 Methods and Results 160
4.2.1 A Note About Methods 160
4.2.2 Bioinformatics Analysis 160
4.2.3 Primer Design, PCR amplification and Cloning 160
4.2.4 Protein Expression and Solubility Tests for His-tagged MenB 161

4.2.5 LARGE SCALE EXPRESSION AND INITIAL PROTEIN PURIFICATION 163
4.2.6 FURTHER PROTEIN PURIFICATION 164
4.2.6.1 Method A 164
4.2.6.2 Method B 166
4.2.6.3 Native Gel Analysis 167
4.2.6.4 Method C 169
4.2.7 PROTEIN PURIFICATION AND CHARACTERIZATION SUMMARY 171
4.2.8 PROTEIN MICRODIALYSIS EXPERIMENTS 172
4.2.9 N-TERMINAL SEQUENCING AND MASS SPECTRAL ANALYSIS 176
4.2.10 CRYSTALLISATION EXPERIMENTS 176
4.2.10.1 A General Overview 176
4.2.10.2 Identification and Refinement of Crystallization Conditions 177
4.2.10.3 A Summary of Crystal Morphology 178
4.2.10.4 Soaking and Co-crystallisation Experiments 179
4.2.11 X-RAY DATA COLLECTION 181
4.2.11.1 Data Collection from Method A Crystals 181
4.2.11.2 Data Collection from Method B "Diamond" Crystals 181
4.2.11.3 Data Collection from Method C Crystals 182
4.2.11.4 Summary of the Various Crystal Forms of MenB 183
4.2.12 DATA PROCESSING FROM METHOD A AND C CRYSTALS 183
4.2.13 METHOD A AND C CRYSTALS: STRUCTURE SOLUTION AND REFINEMENT 185
4.2.13.1 An Overview 185
4.2.13.2 The MenB_C2-Initial Model (2.7Å) 186
4.2.13.3 Higher Resolution MenB_P4_32_12 Model (2.15 Å) 187
4.2.13.4 The MenB_C2-Imz Model (2.0 Å) 190
4.2.13.5 The MenB_C2-NCoA Model (2.3 Å) 191
4.2.14 MODEL COMPLETENESS AND QUALITY: 193
4.2.14.1 The MenB_C2-Initial Model 193
4.2.14.2 The MenB_P4_32_12 Model 193
4.2.14.3 The MenB_C2-Imz Model 195
4.2.14.4 The MenB_C2-NCoA Model 195
4.2.15 METHOD B DIAMOND CRYSTALS: DATA PROCESSING AND STRUCTURE SOLUTION 197
4.2.15.1 Data Processing Overview 197
4.2.15.2 Processing in Cubic Space Groups P432 and P4_132 198
4.2.15.3 The MenB_Diamond Models: Model Building and Refinement in P4_132 199
4.2.15.4 Processing and Molecular Replacement in Other Space Groups 200
4.2.15.5 The MenB_Diamond Model: Model Completeness 202
4.3 OVERVIEW OF THE STRUCTURE AND DISCUSSION 202
4.3.1 THE MONOMER FOLD 202
4.3.1.1 Description Based on the MenB_P4_32_12 Model and MenB_C2-NCoA Model 202
4.3.1.2 Comparison of the Monomer Structure in Each Model 206
4.3.1.3 Comparisons with Other Crotonase Superfamily Members 209
4.3.1.4 Comparison With Another *M. tuberculosis* MenB (Rv0548c) Structure 212
4.3.2 THE QUATERNARY STRUCTURE 214
4.3.2.1 The Trimer 214
4.3.2.2 The Hexamer: A Conventional Dimer of Trimers 217
4.3.2.3 Diamond Crystal Form: An Alternative Hexameric arrangement? 221
4.3.3 THE EPPS MOLECULE 223
4.3.4 SEQUENCE CONSERVATION 226
4.3.5 THE ACTIVE SITE 229
4.3.5.1 Locating the Active Site and Identifying Potential Substrate Binding Residues 229
4.3.5.2 The CoA Ligand 232
4.3.5.3 Comparison with the Published MenB_1Q51 Structure 235
4.4 CONCLUDING DISCUSSION 239

CHAPTER 5 ... 241

OTHER TARGETS ... 241

5.1 GENERAL INTRODUCTION ... 241
5.2 A POTENTIAL ISOPRENYL DIPHOSPHATE SYNTHASE GRCC2 (RV0989C) ... 241
5.2.1 INTRODUCTION ... 241
5.2.1.1 Isoprenyl Diphosphate Synthases in *Mycobacterium tuberculosis* H37Rv ... 244
5.2.1.2 Why Target Rv0989c? ... 244
5.2.2 METHODS AND RESULTS ... 245
5.2.2.1 Bioinformatics ... 245
5.2.2.2 Primer Design, PCR Amplification and Cloning ... 247
5.2.2.3 His-tagged Rv0989c: Protein Expression and Solubility ... 248
5.2.2.4 GST-fusion Rv0989c: Protein Expression and Solubility ... 251
5.2.2.5 His-tagged Rv0989c: Refolding ... 252
5.2.3 DISCUSSION ... 253
5.3 AN UNKNOWN: RV0990C ... 254
5.3.1 INTRODUCTION ... 254
5.3.2 METHODS AND RESULTS ... 256
5.3.2.1 Bioinformatics ... 256
5.3.2.2 Primer Design, PCR Amplification and Cloning ... 256
5.3.2.3 pProEX Hta Rv0990c: Protein Expression and Solubility ... 257
5.3.2.4 GST-Fusion Rv0990c: Protein Expression and Solubility ... 257
5.3.3 DISCUSSION ... 257
5.4 MENA (RV0534C) AND MENC (RV0553) ... 259
5.5 MEND (RV0555) ... 260
5.5.1 INTRODUCTION ... 260
5.5.2 METHODS AND RESULTS ... 261
5.5.2.1 Bioinformatics Analysis ... 261
5.5.2.2 Primer Design, PCR Amplification and Cloning ... 262
5.5.2.3 His-tagged MenD: Protein Expression and Solubility ... 263
5.5.2.4 His-tagged MenD: Refolding ... 264
5.5.3 DISCUSSION ... 265
5.6 MENE (RV0542C) ... 266
5.6.1 INTRODUCTION ... 266
5.6.2 METHODS AND RESULTS ... 267
5.6.2.1 Bioinformatics Analysis ... 267
5.6.2.2 Primer Design, PCR Amplification and Cloning ... 268
5.6.2.3 His-tagged MenE: Protein Expression and Solubility ... 269
5.6.2.4 GST-fusion MenE: Protein Expression and Solubility ... 269
5.6.3 DISCUSSION ... 272
5.7 UBIE (RV0558) ... 273
5.7.1 INTRODUCTION ... 273
5.7.2 METHODS AND RESULTS ... 275
5.7.2.1 Bioinformatics Analysis ... 275
5.7.2.2 Primer Design, PCR Amplification and Cloning ... 275
5.7.2.3 His-tagged UbiE: Protein Expression and Solubility ... 276
5.7.2.4 GST-fusion UbiE: Protein and Solubility ... 276
5.7.2.5 His-tagged UbiE: Refolding ... 277
5.7.3 DISCUSSION ... 277
5.8 THE SOLUBILITY BOTTLENECK ... 278

CHAPTER 6 ... 281

CONCLUDING DISCUSSION ... 281

LIST OF FIGURES

Number *Page*

FIGURE 1.1: A TYPICAL X-RAY CRYSTALLOGRAPHY STRUCTURAL GENOMICS PIPELINE. 10
FIGURE 1.2: SUMMARY OF THE *M. TUBERCULOSIS* INFECTION PROCESS. 17
FIGURE 1.3: RELEVANT QUINONES 30
FIGURE 1.4: THE ORIGINS OF THE MENAQUINONE SKELETON. 40
FIGURE 1.5: THE MENAQUINONE BIOSYNTHESIS PATHWAY 41
FIGURE 1.6: THE *MEN* CLUSTERS FROM *E. COLI* AND *B. SUBTILIS* 45
FIGURE 1.7: FIVE OF THE CLOSELY PLACED *MEN* GENES IN THE *M. TUBERCULOSIS* CHROMOSOME. 49
FIGURE 2.1: PROTEIN STANDARDS 67
FIGURE 3.1: PARTIAL REACTION SCHEME SHOWING THE ROLE OF MENG IN MENAQUINONE BIOSYNTHESIS 108
FIGURE 3.2: EXPRESSION AND SOLUBILITY TESTS OF HIS-TAGGED RV3853 115
FIGURE 3.3: NATIVE-PAGE ANALYSIS OF THE HIS-TAGGED RV3853 REFOLDING 118
FIGURE 3.4: PURIFICATION OF HIS-TAGGED RV3853. 120
FIGURE 3.5: CRYSTALS OF HIS-TAGGED RV3853 121
FIGURE 3.6: RAMACHANDRAN PLOT FOR RV3853 129
FIGURE 3.7: THE RV3853 MONOMER FOLD (A) AND TOPOLOGY DIAGRAM FOR THE MONOMER (B) 131
FIGURE 3.8: A RIBBON DIAGRAM (A) AND SURFACE REPRESENTATION (B) OF THE RV3853 TRIMER 134
FIGURE 3.9: MULTIPLE SEQUENCE ALIGNMENT OF 13 REPRESENTATIVE RV3853 SEQUENCE HOMOLOGUES. 135
FIGURE 3.10: STEREOVIEW OF THE TARTRATE ELECTRON DENSITY 136
FIGURE 3.11: STEREODIAGRAM OF THE BINDING SITE FOR THE TARTRATE MOLECULE 137
FIGURE 3.12: ELECTRON DENSITY (A) AND BINDING SITE (B) FOR THE PUTATIVE GLYOXALATE MOLECULE 138
FIGURE 3.13: MULTIPLE SEQUENCE ALIGNMENT OF 5 RV3853 AND 5 ALDOLASE SEQUENCE HOMOLOGUES 145
FIGURE 3.14: TARTRATE, ALDOLASE SUBSTRATE AND ALDOLASE CATALYSED REACTION 147
FIGURE 3.15: DIAGRAM OF BACKBONE OVERLAY OF THE RV3853 A CHAIN AND THE RRAA C CHAIN 151
FIGURE 3.16: OVERLAY OF RV3853 TARTRATE BINDING SITE AND *E. COLI* RRAA 151
FIGURE 4.1: PARTIAL REACTION SCHEME SHOWING THE ROLE OF MENB IN MENAQUINONE BIOSYNTHESIS 155
FIGURE 4.2: SUMMARY OF PROPOSED MENB-CATALYSED REACTION MECHANISM 156
FIGURE 4.3: THE STRUCTURE OF COENZYME A 157
FIGURE 4.4: SELECTED REACTIONS OF MEMBERS OF THE CROTONASE SUPERFAMILY 158
FIGURE 4.5: SOLUBILITY AND EXPRESSION TESTS UNDERTAKEN AT 25°C FOR HIS-TAGGED MENB 163
FIGURE 4.6: METHOD A PURIFICATION PROTOCOL. 165
FIGURE 4.7: COMPARISON OF PROTEIN BY SDS-PAGE BEFORE AND AFTER BUFFER EXCHANGE 167
FIGURE 4.8: NATIVE-PAGE ANALYSIS OF PROTEIN PREPARED BY METHOD A 168
FIGURE 4.9: NATIVE-PAGE ANALYSIS OF PROTEIN PREPARED BY METHOD B 168
FIGURE 4.10: NATIVE-PAGE ANALYSIS OF PROTEIN PREPARED BY METHOD B OR METHOD C 170
FIGURE 4.11: A METHOD C PURIFICATION 171
FIGURE 4.12: MICRODIALYSIS EXPERIMENT VARYING THE CONCENTRATION OF IMIDAZOLE 174
FIGURE 4.13: MICRODIALYSIS EXPERIMENT TO TEST SOLUBILITY 175
FIGURE 4.14: COMPOUNDS USED IN SOAKING AND CO-CRYSTALLIZATION EXPERIMENTS 179
FIGURE 4.15: MODELS BUILT AND REFINED USING DATA COLLECTED FROM METHOD A AND C CRYSTALS 185
FIGURE 4.16: RAMACHANDRAN PLOT FOR THE MENB_P43212 MODEL 194
FIGURE 4.17: RAMACHANDRAN PLOT FOR THE MENB_C2-NCOA MODEL 196
FIGURE 4.18: ARRANGEMENT OF SECONDARY STRUCTURAL ELEMENTS IN THE MENB MONOMER 204
FIGURE 4.19: STEREOVIEW OF THE MENB MONOMER 205
FIGURE 4.20: PYMOL[259] REPRESENTATION OF THE MENB MONOMER FOLD. 205
FIGURE 4.21: STEREOVIEW OF BACKBONE OVERLAYS OF VARIOUS MENB MODELS 208
FIGURE 4.22: BACKBONE OVERLAY OF THE MENB_C2-NCOA (MAGENTA) B CHAIN WITH ENOYL-COA HYDRATASE (YELLOW) AND 4-CHLOROBENZOYL-COA DEHALOGENASE (GREEN) 211
FIGURE 4.23: BACKBONE OVERLAY OF THE MENB_C2-NCOA (MAGENTA) B CHAIN WITH METHYLMALONYL-COA DECARBOXYLASE (ORANGE) 212
FIGURE 4.24: BACKBONE OVERLAY OF MENB_C2-NCOA AND MENB_1Q51 B CHAINS 214
FIGURE 4.25: THE MENB TRIMER 215
FIGURE 4.26: THE MENB TRIMER WITH SECONDARY STRUCTURE HIGHLIGHTED 216
FIGURE 4.27: STEREOVIEW OF THE A, B, A', B' HEXAMER INTERFACE 217
FIGURE 4.28: THE MENB HEXAMER: THE ARRANGEMENT OF A, B, C, A', B', C' 218
FIGURE 4.29: DIFFERING ARRANGEMENTS OF THE C-TERMINUS IN CROTONASE SUPERFAMILY MEMBERS 219
FIGURE 4.30: THE "STAGGERED HEXAMER" 222
FIGURE 4.31: ELECTRON DENSITY FOR THE EPPS MOLECULE IN THE MENB_$P4_32_12$ MODEL 224

Figure 4.32: Binding site of the EPPS molecule 225
Figure 4.33: Multiple sequence alignment of 10 MenB protein sequences 227
Figure 4.34: Alignment of MenB and other crotonase superfamily protein sequences 228
Figure 4.35: Segments likely to be involved in active site formation in MenB 231
Figure 4.36: Stereoview of the electron density for the CoA ligand 232
Figure 4.37: Location of the CoA ligand within the B chain of the MenB_C2-NCoA model 234
Figure 4.38: Location of the CoA Molecule in the MenB_C2-NCoA Hexamer 234
Figure 4.39: Stereoview of the CoA binding region 235
Figure 4.40: Active site region of MenB_C2-NCoA chain A showing the NCoA binding pocket 238
Figure 5.1: The polyisoprenyl tail of menaquinone comes from a polyisoprenyl diphosphate 242
Figure 5.2: Alignment of the aspartate-rich motifs for 9 isoprenyl diphosphate synthases 246
Figure 5.3: A 1% agarose gel of a PCR amplification reaction. 248
Figure 5.4: Example of a double diagnostic digest for pET42a rTEV Rv0989c and Rv0990c. 248
Figure 5.5: A 37ºC expression and solubility test for His-tagged Rv0989c. 249
Figure 5.6: Ni^{2+} binding tests using soluble cell lysate from an 18ºC expression culture 251
Figure 5.7: Expression and solubility tests of GST-fusion Rv0990c. 257
Figure 5.8: Role of MenA and MenC in menaquinone biosynthesis 259
Figure 5.9: Scheme highlighting the role of MenD in menaquinone biosynthesis 261
Figure 5.10: PCR amplification of *menD* using set two primers (1% Agarose Gel). 263
Figure 5.11: A solubility and expression test for His-tagged MenD at 18ºC. 264
Figure 5.12: Refolding experiments for His-tagged MenD. 265
Figure 5.13: Partial reaction scheme showing the role of MenE in menaquinone biosynthesis 267
Figure 5.14: Solubility and expression test for GST-fusion MenE at 24ºC. 270
Figure 5.15: Medium scale GST purification of GST-fusion MenE. 272
Figure 5.16: Scheme showing the likely role of UbiE in menaquinone biosynthesis 274
Figure 5.17: Expression and solubility test for GST-fusion UbiE at 18ºC. 277

TABLE OF TABLES

Number *Page*

TABLE 1.1: A SELECTION OF STRUCTURAL GENOMICS PROJECTS CURRENTLY BEING UNDERTAKEN 11
TABLE 1.2: *IN VIVO* MODELS FOR LATENT TB 19
TABLE 1.3: THE *IN VITRO* WAYNE HYPOXIC MODEL FOR LATENT TB 21
TABLE 1.4: THE *IN VITRO* NUTRIENT STARVATION MODEL FOR LATENT TB 22
TABLE 1.5: SUMMARY OF THE MOST COMMONLY USED ANTI-TB DRUGS 25
TABLE 1.6: *C. GLUTAMICUM* GENES WHOSE PROTEINS PASS ELECTRONS TO MENAQUINONE 34
TABLE 1.7: *C. GLUTAMICUM* GENES WHOSE PROTEINS ACCEPT ELECTRONS FROM MENAQUINOL 34
TABLE 1.8: *M. TUBERCULOSIS* PROTEINS THAT MAY INTERACT WITH MENAQUINONE 37
TABLE 1.9: 11 GOALS OF THIS PROJECT 50
TABLE 2.1: BLAST AND PSI-BLAST RELATED TERMS 54
TABLE 2.2: PRIMER SEQUENCES 56
TABLE 2.3: PROPERTIES OF RESTRICTION ENZYMES USED 57
TABLE 2.4: ANTIBIOTIC SOLUTIONS USED 58
TABLE 2.5: TYPICAL REAGENTS FOR A 50 µL POLYMERASE CHAIN REACTION 59
TABLE 2.6: A TYPICAL PCR REACTION SEQUENCE 59
TABLE 2.7: TYPICAL REACTION FOR A PCR FRAGMENT DOUBLE DIGEST 60
TABLE 2.8: TYPICAL REACTION FOR A DIAGNOSTIC DOUBLE DIGEST 61
TABLE 2.9: TYPICAL REACTION FOR PREPARATIVE PLASMID DIGESTION 61
TABLE 2.10: A TYPICAL LIGATION REACTION 61
TABLE 2.11: STACKING GEL REAGENTS FOR A 10 GEL MIXTURE 65
TABLE 2.12: RESOLVING GEL REAGENTS FOR A 10 GEL MIXTURE 65
TABLE 2.13: GENERAL LYSIS BUFFER SCREEN 71
TABLE 2.14: PH LYSIS BUFFER SCREENS 71
TABLE 2.15: DETERGENT LYSIS BUFFER SCREENS 72
TABLE 2.16: REFOLDING SCREEN SOLUTIONS (ADAPTED FROM THE HAMPTON REFOLDING SCREEN) 76
TABLE 2.17: SOLUTIONS USED IN MICRODIALYSIS EXPERIMENTS 85
TABLE 2.18: CRYSTAL SCORING SYSTEM 88
TABLE 2.19: CRYSTALLISATION SCREENS 89
TABLE 3.1: ALDOLASES ALIGNING TO THE RV3853 PROTEIN SEQUENCE 113
TABLE 3.2: SOLUBILITY TESTS UNDERTAKEN FOR HIS-TAGGED RV3853 114
TABLE 3.3: HEAVY METAL SOAKS OF HIS-TAGGED RV3853 CRYSTALS 123
TABLE 3.4: DATA COLLECTION AND PROCESSING STATISTICS 125
TABLE 3.5: PROGRESS OF MODEL BUILDING AND REFINEMENT FOR THE RV3853 STRUCTURE 128
TABLE 3.6: SUMMARY OF REFINEMENT AND MODEL DETAILS 130
TABLE 3.7: STRUCTURAL HOMOLOGUES OF RV3853 FOUND USING DALI OR SSM 132
TABLE 4.1: SEQUENCE HOMOLOGUES OF RV0542C IN THE PDB (FOUND BY BLAST ANALYSIS) IN 2002 160
TABLE 4.2: EXPRESSION AND SOLUBILITY TESTS FOR HIS-TAGGED MENB 162
TABLE 4.3: PROTEIN PURIFICATION AND CHARACTERISATION SUMMARY 172
TABLE 4.4: INITIAL SUCCESSFUL CRYSTALLISATION CONDITIONS 177
TABLE 4.5: A SUMMARY OF CRYSTAL TYPES OBSERVED 178
TABLE 4.6: CRYSTAL FORMS OF MENB FROM WHICH DIFFRACTION DATA WERE COLLECTED 183
TABLE 4.7: DATA PROCESSING STATISTICS FOR METHOD A AND METHOD C CRYSTALS 184
TABLE 4.8: MODEL BUILDING AND REFINEMENT OF THE MENB_C2-INITIAL MODEL 187
TABLE 4.9: MODEL BUILDING AND REFINEMENT FOR THE MENB_$P4_32_12$ MODEL 189
TABLE 4.10: MODEL BUILDING AND REFINEMENT FOR THE MENB_C2-IMZ MODEL 190
TABLE 4.11: MODEL BUILDING AND REFINEMENT FOR MENB_C2-NCOA MODEL 192
TABLE 4.12: REFINEMENT AND MODEL DETAILS FOR THE MENB_$P4_32_12$ MODEL 194
TABLE 4.13: REFINEMENT AND MODEL STATISTICS FOR THE MENB_C2-NCOA MODEL 197
TABLE 4.14: PROCESSING STATISTICS FOR DIAMOND CRYSTALS IN P432/$P4_132$ 198
TABLE 4.15: ABSENT REFLECTIONS WITH HIGH I/Σ_I VALUES FOR THE MENADIONE SOAK CRYSTAL DATA 199
TABLE 4.16: R FACTORS AFTER PARTIAL REFINEMENT OF THE MENB_DIAMOND MODELS 200
TABLE 4.17: SUMMARY OF SSM OVERLAYS FOR MODELS OF MENB 206
TABLE 4.18: RESULTS OF SEARCHES FOR STRUCTURAL HOMOLOGUES OF MENB 210
TABLE 4.19: SUMMARY OF THE RESIDUES INVOLVED IN THE TRIMERISATION 216
TABLE 4.20: SUMMARY OF TRIMER-TRIMER INTERACTIONS IN THE MENB_C2-NCOA HEXAMER 220
TABLE 4.21: ACTIVE SITE TEMPLATE SEGMENTS FOR CROTONASE SUPERFAMILY STRUCTURES 231
TABLE 5.1: CLASSES OF PRENYLTRANSFERASES THAT MAKE ALL-*E*-POLYISOPRENYL DIPHOSPHATES 243

TABLE 5.2: LIKELY ALL-*E*-ISOPRENYL DIPHOSPHATE SYNTHASE GENES IN *M. TUBERCULOSIS* 244
TABLE 5.3: SOLUBILITY TESTS UNDERTAKEN FOR GST-FUSION MENE.. 269
TABLE 5.4: COMMON METHODS USED TO IMPROVE PROTEIN SOLUBILITY 280
TABLE 6.1: OUTPUT SUMMARY FROM THIS STUDY 282
TABLE 6.2: OUTPUT SUMMARY FROM THE *M. TUBERCULOSIS* STRUCTURAL GENOMICS CONSORTIUM 283

ABBREVIATIONS

AMP adenosine monophosphate

Amp. ampicillin

ATP adenosine triphosphate

BLAST basic local alignment search tool

CC correlation coefficient

C-terminal carboxy terminal

CAM chloramphenicol

CoA Coenzyme A

CV column volume

DNA deoxy-ribonucleic acid

dNTP deoxy-nucleotide triphosphate

DTT dithiothreitol

DNase deoxyribonuclease I

DHNA 1, 4 dihydroxy-2-naphthoic acid

EDTA ethylenediamine tetraacetic acid

EPPS 3-[4-(2-hydroxyethyl)-1-piperazinyl]propane sulfonic acid

GSH oxidised glutathione

GSSG reduced glutathione

HEPES N-2-hydroxyethylpiperazine-N'-2-ethanesulphonic acid

IPTG isopropyl β-D-thiogalactopyranoside

IMAC immobilised metal affinity chromatography

kb kilo-base pairs

kDa kilo-dalton

LB luria broth

MAD multiple wavelength anomalous diffraction

MES morpholine-ethanesulfonic acid

MIR multiple isomorphous replacement

MIRAS multiple isomorphous replacement with anomalous scattering

MR molecular replacement

MW molecular weight

MWCO molecular weight cut-off

mPEG methoxypolyethylene glycol

MPD 2-methyl 2,4-petanediol

N-terminal amino terminus

Native-PAGE non denaturing PAGE

NCoA naphthoyl-CoA (CoA ester of 1, 4 dihydroxy-2-naphthoic acid)

NCS non-crystallographic symmetry

NMR nuclear magnetic resonance

ORF open reading frame

OSB *O*-succinyl benzoic acid

PAGE polyacrylamide gel electrophoresis

PCR polymerase chain reaction

PDB protein data bank

PEG polyethylene glycol

PEG-MME polyethylene glycol monomethyl ether

PSI-BLAST position specific iterated BLAST

RNA ribonucleic acid

RNase ribonuclease A

r.t. room temperature

rTEV recombinant tobacco etch virus

SAD single wavelength anomalous diffraction

SDS sodium dodecyl sulphate

SDS-PAGE sodium dodecyl sulphate-PAGE

SIR single isomorphous replacement

SIRAS single isomorphous replacement with anomalous scattering

TAE tris-acetate-EDTA

TB tuberculosis

TMED N,N,N'N'-tetramethylethylenediamine

Tris tris(hydroxymethyl)aminomethane

UV ultraviolet