

Structural & Biophysical Characterisation of *M. tuberculosis* MmpL3

Oliver Adams
Department of Biochemistry
St.Anne's College

Supervisors:
Prof. Simon Newstead
Dr. Philip Fowler

A Thesis Submitted For:
DPhil Interdisciplinary Bioscience

June 2021

Structural & Biophysical Characterisation of *M. tuberculosis* MmpL3

Submitted for the degree of Doctor of Philosophy

Oliver Adams, St.Anne's College

Abstract

Tuberculosis (TB) is the leading cause of death from a single infectious agent, with a staggering 10 million people estimated to have contracted the disease in 2019 alone. Whilst treatments for TB exist, continual emergence of drug-resistant strains necessitates urgent development of next-generation antituberculars. A promising new target is the essential glycolipid exporter mycobacterial membrane protein large 3 (MmpL3), a member of the resistance-nodulation-division (RND) superfamily. MmpL3 is required for construction of the distinctive *Mycobacterium tuberculosis* (*Mtb*) cell envelope; a complex lipid-rich assembly critical to mycobacterial viability. Despite discovery of numerous MmpL3 inhibitors, including one in clinical trials, a structural understanding of the transporter, and by extension the variants identified to confer resistance to said experimental agents, has remained elusive. To this end, the research herein presents a 3.0 Å reconstruction of *Mtb* MmpL3; the culmination of a project structurally pursuing the RND exporter through construct design & characterisation, *in vitro* binder selection, nanodisc reconstitution, *in meso* & *in surfo* crystallisation trials, and imaging by Cryo-EM. Unexpectedly monomeric, the *Mtb* MmpL3 atomic model provides insights into the evolutionary co-option of RND architecture for lipophile transit. Most notably, the protein houses an extensive periplasmic cavity with putative functional relevance as the glycolipid translocation conduit. A detergent molecule resolved within the cavity plausibly sheds light on interactions employed by MmpL3 to coordinate its native substrate. Additionally, comprehensive mapping of known resistance-conferring substitutions suggests the existence of at least two routes, operant within different domains, by which antituberculars inhibiting MmpL3 can be mutationally circumvented. Taken together, the data reported in this thesis set the stage for an era of structure-based drug design against MmpL3; guided by an enhanced appreciation of the underlying resistance landscape.

Acknowledgements

Without guidance, assistance, and encouragement from a raft of individuals this thesis, and the over four years of work underpinning it, simply would not have been possible. First and foremost, I would like to sincerely thank my supervisors — Prof. Simon Newstead & Dr. Philip Fowler. Collectively, their scientific insight, mentorship, support, and patience have proved invaluable during my DPhil, ensuring it always felt worthwhile to persevere, even during the most arduous periods. I also extend special thanks to Dr. Joanne Parker & Dr. Justin Deme, whose respective expertise in membrane protein biochemistry and Cryo-EM, hands-on training, direct experimental contributions, and frequent advice, were essential to the success of the project. I similarly express gratitude to Dr. Iwan Zimmermann for his aid with sybody selection, and for making me feel wholly welcome during a secondment at the University of Zürich. Moreover, I appreciate and acknowledge the efforts of the learned facilities staff at the Central Oxford Structural Molecular Imaging Centre (COSMIC; Dr. Errin Johnson & Adam Costin), the Molecular Biophysics Suite (Dr. David Staunton), and the Advanced Proteomics Facility (Dr. Sabrina Liberatori), whose specialist skills expedited various aspects of the research herein. For their camaraderie, fruitful discussions, and provision of a convivial lab environment, I thank members of the Newstead Group past and present, in particular Phil, Richard & Ashley. Away from the bench I am truly indebted to Lizzie, my friends, and my parents, for without them I surely could not have weathered the ups and downs of academic life. Finally, I am grateful to the Biotechnology and Biological Sciences Research Council (BBSRC) for kindly funding my DPhil.

Contents

Page

List of Figures v

List of Tables ix

List of Acronyms xi

1 Introduction 1

1.1 The Mycobacterial Cell Envelope 2

1.2 The MmpL Family of RND Transporters 8

1.3 MmpL3 in Focus 17

1.4 Aims & Objectives 21

2 Materials & Methods 23

2.1 Construct Cloning 23

2.2 Protein Expression & Purification 25

2.3 Membrane Protein Characterisation 30

2.4 Sybody Generation & Screening 33

2.5 Reconstitution into Detergent Alternatives 35

2.6 Crystallisation Trials 37

2.7 Cryo-EM Trials & Structure Determination 38

2.8 Structural & Sequence Analyses 41

3 Homologue Screening: Construct Design & Characterisation 45

3.1 The Screening Pipeline 46

3.2 The Initial Homologue Set 49

3.3 Other *M. tuberculosis* MmpLs 61

3.4 The MmpL3 Truncation 66

4 The MmpL3 Truncation: Initial Structural Attempts 75

4.1 Crystallisation Trials 75

4.2 Cryo-EM Trials 81

5 Synthetic Nanobodies 93

5.1 Nanobodies: Structural Biology Aids 93

5.2 Synthetic Nanobodies 97

5.3 Selection Against Truncated MmpL3 100

5.4 Screening Against Truncated MmpL3 111

5.5 Truncation-Sybody Complex Structural Trials 118

6 MSP Nanodisc Studies: From PepTSt to MmpL3 129

6.1 The MSP Nanodisc System 130

6.2 PepTSt Nanodiscs: A Model System? 133

6.3 MmpL3 Nanodiscs 147

7 The MmpL3 Truncation: Structure Determination & Analysis 157

7.1 An Aside on Structures of *Msmg* MmpL3 158

7.2 Structural Elucidation of Truncated MmpL3 160

7.3 Regions of Functional & Therapeutic Interest 168

7.4 The Mutational Landscape of MmpL3 178

7.5 Final Remarks & Future Perspectives 186

A Sequence Alignments 195

B Supplementary Figures & Data 203

C Publications 227

Bibliography 243

List of Figures

1.1 *Mtb* Cell Envelope Architecture & Major Mycomembrane Lipids 4
1.2 Structure-Annotated Phylogenetic Tree of the RND Superfamily 9
1.3 Variation in RND Transporter Oligomeric State 10
1.4 Overview of MmpL Function, Topology, Structure & Inhibitors 14

3.1 FSEC Traces of MmpL3, CmpL1, CmpL3 & NCgl-1194 52
3.2 Purification of MmpL3-TEV-GFP-His_8 in DDM 54
3.3 Purification of MmpL3-FLAG in DDM 55
3.4 Purification of CmpL1-TEV-GFP-His_8 in DDM 57
3.5 Purification of CmpL1-FLAG in DDM 58
3.6 Purification of CmpL3-TEV-GFP-His_8 in DDM 60
3.7 FSEC & Purification of MmpL11-TEV-GFP-His_8 Constructs in DDM . . 65
3.8 MmpL3 C-Terminal Truncation Rationale & FSEC Traces 67
3.9 Purification of $MmpL3_{1\text{-}753}$-TEV-GFP-His_8 in DDM & LMNG 69
3.10 Evaluation of $MmpL3_{1\text{-}753}$ Oligomeric State 71
3.11 CD Spectroscopy of $MmpL3_{1\text{-}753}$ in DDM & LMNG 73

4.1 Example Results of $MmpL3_{1\text{-}753}$ Vapour Diffusion Trials in LMNG . . . 77
4.2 Example Non-Protein Crystalline Species Seen in LCP Trials 80
4.3 Cryo-EM Screening of $MmpL3_{1\text{-}753}$ in DDM 83
4.4 Cryo-EM Screening of $MmpL3_{1\text{-}753}$ in LMNG 85
4.5 Cryo-EM Screening of $MmpL3_{1\text{-}753}$ in LMNG + Haem 87
4.6 Cryo-EM Screening of Cross-linked $MmpL3_{1\text{-}753}$ in LMNG + Haem . . 88
4.7 Cryo-EM Screening of A8-35 Reconstituted $MmpL3_{1\text{-}753}$ + Haem . . . 91

5.1 IgG, Antibody Fragment, HCAb, and Nb Structural Comparison 95
5.2 Exemplar Applications of Nbs in Structural Biology 96
5.3 An Overview of the Sybody Libraries & Selection Pipeline 99

5.4 MmpL3$_{1\text{-}753}$-Avi-TEV-GFP-His$_8$ DDM Purification & Biotinylation 101
5.5 ELISA of Sybodies Selected Against MmpL3$_{1\text{-}753}$ in DDM 104
5.6 Concave & Loop Sybody CDR Sequence Analysis 107
5.7 Convex Sybody CDR Sequence Analysis 110
5.8 SEC Traces of Small-Scale Batch Purified Sybodies 112
5.9 MmpL3$_{1\text{-}753}$-Sybody Pull-Down & Co-Elution Screen Optimisation . . 114
5.10 Collated MmpL3$_{1\text{-}753}$-Sybody SEC Co-Elution Results 115
5.11 BLI Sybody Off-Rate Ranking 116
5.12 Large-Scale Purification of Tagless Sybodies 119
5.13 MmpL3$_{1\text{-}753}$-Sybody Complexation for Crystallisation in DDM 121
5.14 MmpL3$_{1\text{-}753}$-Sybody Complexation for Cryo-EM in DDM & LMNG . . . 123
5.15 MmpL3$_{1\text{-}753}$-Sybody Pull-Down Comparison in DDM & LMNG 125
5.16 Sybody Re-ELISA in LMNG & Cryo-EM Imaging of MmpL3$_{1\text{-}753}$-MG2 . 126

6.1 Nanodiscs: Reconstitution, Characterisation & MSPs 131
6.2 Exemplar PepTSt Structures 134
6.3 Cryo-EM Reconstruction of PepTSt-MSP1D1-Asolectin Nanodiscs . . . 136
6.4 Potential PepTSt-MSP1D1 Contact Sites 138
6.5 Cryo-EM Screening of PepTSt-MSP1E3D1-Asolectin Nanodiscs 140
6.6 Screening PepTSt Reconstitution into MSP2N2-Asolectin Nanodiscs . . 141
6.7 Cryo-EM Screening of PepTSt-MSP2N2-Asolectin Nanodiscs 142
6.8 Native-PAGE Comparison of PepTSt-MSP-Asolectin Nanodiscs 143
6.9 Empty Asolectin-Nanodisc Neg-Stain Analysis 145
6.10 Screening MmpL3$_{1\text{-}753}$-MSP1D1-Asolectin Nanodisc Reconstitution . . 148
6.11 Further MmpL3$_{1\text{-}753}$-MSP-Asolectin Nanodisc Reconstitution Screens . 149
6.12 MmpL3$_{1\text{-}753}$-MSP1D1ΔH5-Asolectin Nanodisc Characterisation 151
6.13 Neg-Stain of MmpL3$_{1\text{-}753}$-MSP1D1ΔH5-Asolectin Nanodiscs 152
6.14 Cryo-EM Screening of MmpL3$_{1\text{-}753}$-MSP1D1-Asolectin Nanodiscs . . . 153
6.15 Cryo-EM Screening of MmpL3$_{1\text{-}753}$-MSP1D1ΔH5-Asolectin Nanodiscs 154
6.16 MmpL3$_{1\text{-}753}$-MSP1D1ΔH5-Asolectin Nanodisc Sybody Complexation . 155

7.1 MmpL3$_{1\text{-}753}$ LMNG Purification, Cryo-EM Workflow & Map Resolution 161
7.2 Cryo-EM Map & Atomic Model of MmpL3$_{1\text{-}753}$ in LMNG 163
7.3 Structural Comparison of *Mtb* & *Msmg* MmpL3 166
7.4 MmpL3$_{1\text{-}753}$ Regions of Functional & Therapeutic Importance 169
7.5 Structural Analysis of *Mtb* & *Msmg* MmpL3 Periplasmic Ligands 171
7.6 Comparison of MmpL3$_{1\text{-}753}$ to Other RND Lipophile Transporters . . . 174
7.7 *Mtb* MmpL3$_{1\text{-}753}$ Conservation of the *Msmg* Common Drug Pocket . . 176
7.8 Structural Mapping of MmpL3 Drug-Resistance Mutations 179
7.9 Stratification of Drug-Resistance Mutations by Inhibitor Class 183
7.10 Visualising the Wider Non-Synonymous Landscape of *Mtb* MmpL3 . . 185
7.11 Proposed Model for MmpL3-Mediated TMM Export 189

A.1 Sequence Alignment of *Mtb* MmpL3, *Mtb* MmpL11 and *Cgl* CmpL1 . . 195
A.2 Sequence Alignment of *Mtb* MmpL8 with *Cgl* CmpL3 197
A.3 Sequence Alignment of *Mtb* Rv1250 with *Cgl* NCgL-1194 198
A.4 Sequence Alignment of *Mtb* MmpL5 with *Mtb* MmpL7 199
A.5 Sequence Alignments of Concave and Loop Library Sybodies 200
A.6 Sequence Alignment of Convex Library Sybodies 201
A.7 Sequence Alignment of MmpL3 Orthologues 202

B.1 NCgl-1194 Expression, Phenotype & Attempted Purification 203
B.2 Native-PAGE & Cross-linking of MmpL3/CmpL1-FLAG in DDM 203
B.3 *M. smegmatis* Expression Trial Summary 204
B.4 MmpL3$_{1\text{-}753}$ Full CD Melt Spectra in DDM & LMNG 204
B.5 MmpL3$_{1\text{-}753}$ SEC Haem Association Assay 205
B.6 CmpL1-Avi-FLAG DDM Purification & Biotinylation 206
B.7 MmpL3$_{1\text{-}753}$-Avi-TEV-GFP-His$_8$ LMNG Purification & Biotinylation . . . 206
B.8 MmpL3$_{1\text{-}753}$-Sybody SEC Co-Elution Traces 207
B.9 MmpL3$_{1\text{-}753}$-Sybody BLI Sensorgrams 208
B.10 Epitope Binning Construct Design & Periplasmic Domain Purification . 211
B.11 Purification of MSP2N2 and MSP1D1ΔH5 212

B.12 Purification of PepTSt-TEV-GFP-His_8 in DDM/DM 212
B.13 Summary of Lipids Resolved in PepTSt LCP Structures 213
B.14 Additional MSP-Asolectin Nanodisc Reconstitution Screening 213
B.15 Full CD Melt Spectra of MSP1D1ΔH5-Asolectin Nanodiscs 214
B.16 Overview of the *Msmg* MmpL3 Crystal Structures 215
B.17 $MmpL3_{1\text{-}753}$ Model-to-Map Fit 216
B.18 MmpL3 Conservation Analysis 216
B.19 MmpL3 Inter-Repeat Polar Contact Analysis 217

List of Tables

3.1 A Summary of FSEC Screening Detergent Properties 48

3.2 Initial Homologue Set Details & Expression Counts 51

3.3 Additional MmpL Trial Set Details & Expression Counts 63

3.4 MmpL3 C-Terminal Truncation Details & Expression Counts 66

4.1 Summary of $MmpL3_{1\text{-}753}$ Vapour Diffusion Trials 76

4.2 Summary of $MmpL3_{1\text{-}753}$ LCP Trials 79

5.1 Sybody Library Enrichment Post-Phage Display 103

5.2 Summary of $MmpL3_{1\text{-}753}$-Sybody Vapour Diffusion Trials 121

5.3 Summary of $MmpL3_{1\text{-}753}$-Sybody LCP Trials 122

7.1 Cryo-EM Data Collection, Refinement & Model Validation Statistics . . 165

B.1 Summary of Sybody Screening Results 209

B.2 Curated MmpL3 Drug-Resistance Mutations 218

B.3 Genome-Mined Non-synonymous *Mtb* MmpL3 Substitutions 224

B.4 Lineage Breakdown of Most Abundant *Mtb* MmpL3 Variants 226

Acronyms & Abbreviations

2D-CAVG Two-Dimensional Class Average

6DDTre Lauryl-6-Trehaloside

AAA Aromatic Amino Acid

ABC ATP-Binding Cassette

A280/370/650 Absorbance 280, 370 & 650 nm

AcrB Acridine Resistance Protein B

AEC Anion-Exchange Chromatography

AG Arabinogalactan

AGM Arabinogalactan Mycolate

Amp Ampicillin

APol Amphipol

APPE Aryl Polyene Pigment Exporter

Apr Apramycin

Aso Asolectin

ATP Adenosine Triphosphate

AU Arbitrary or Absorbance Units

BirA *E. coli* Biotin Ligase

BLI Bio-Layer Interferometry

BN-PAGE Blue Native-Polyacrylamide Gel Electrophoresis

Cam Chloramphenicol

CD Circular Dichroism

cDNA Complementary DNA

CDR Complementarity-Determining Region

Cgl *Corynebacterium glutamicum*

CMC Critical Micelle Concentration

CmpL Corynebacterial Membrane Protein Large

COSMIC Central Oxford Structural Molecular Imaging Centre

Cryo-EM Cryogenic Electron Microscopy

CRyPTIC Comprehensive Resistance Prediction For Tuberculosis: An International Consortium

CTD C-terminal Domain

CTF Contrast Transfer Function

CV Column Volume

CYMAL-5 5-Cyclohexyl-1-Pentyl-β-D-Maltoside

DARPin Designed Ankyrin Repeat Proteins

DAT Di-acyltrehalose

DDM n-Dodecyl-β-D-Maltoside

DED Direct Electron Detector

DM n-Decyl-β-D-Maltoside

DMNG Decyl Maltose Neopentyl Glycol

DNA Deoxyribonucleic Acid

DSS Disuccinimidyl Suberate

ECL Enhanced Chemiluminescence

EDTA Ethylenediaminetetraacetic acid

ELISA Enzyme-Linked Immunosorbent Assay

EMDB Electron Microscopy Data Bank

ENA European Nucleotide Archive

ER Endoplasmic Reticulum

EST Eukaryotic Sterol Transporter

$\lambda_{\mathbf{Ex/Em}}$ Excitation/Emission Wavelength

Fab Antigen-Binding Fragment

FDA Food & Drug Administration

FRET Fluorescence Resonance Energy Transfer

FSC Fourier Shell Correlation

FSEC Fluorescence-Detection Size-Exclusion Chromatography

FT Flow-Through

FX Fragment Exchange

GFP Green Fluorescent Protein

GIF Gatan Imaging Filter

GPCR G Protein-Coupled Receptor

HAE Hydrophobe/Amphiphile Efflux

HCAb Heavy Chain Antibody

HME Heavy Metal Efflux

HRP Horse Radish Peroxidase

IgG Immunoglobulin-γ

IMAC Immobilised Metal Affinity Chromatography

IPTG Isopropyl-β-D-Thiogalactopyranoside

Kan Kanamycin

LB Luria Broth

LCP Lipidic Cubic Phase

LDAO Lauryldimethylamine Oxide

LMNG Lauryl Maltose Neopentyl Glycol

LTR Left To Right

MA Mycolic Acid

MAG Monoacylglycerol

mAGP Mycolyl-Arabinogalactan-Peptidoglycan

MATE Multidrug & Toxic Compound Extrusion

MBP Maltose-Binding Protein

MBT Mycobactin

MFP Membrane Fusion Protein

MFS Major Facillitator Superfamily

MmpL Mycobacterial Membrane Protein Large

MmpS Mycobacterial Membrane Protein Small

MNG Maltose-Neopentyl Glycol

MO Monoolein

mRNA Messenger Ribonucleic Acid

Msmg *Mycobacterium smegmatis*

MSP Membrane Scaffold Protein

MST Micro-Scale Thermophoresis

Mtb *Mycobacterium tuberculosis*

MW Molecular Weight

MWCO Molecular Weight Cut-Off

MWE Mycolate Wax Ester

Nb Nanobody

ND Nanodisc

Neg-Stain Negative Stain Electron Microscopy

NFE Nodulation Factor Exporter

NG n-Nonyl-β-D-Glucoside

Ni-NTA Nickel Nitrilotriacetic Acid

NMR Nuclear Magnetic Resonance

NUR Non-Uniform Refinement

OADC Oleic Acid, Albumin, Dextrose & Catalase

OD Optical Density

OM Outer Membrane

OMF Outer Membrane Factor

ORF Open Reading Frame

PAT Penta-acyltrehalose

PBS Phosphate Buffered Saline

PC C-terminal Periplasmic Loop

PCR Polymerase Chain Reaction

PD Periplasmic Domain

PDB Protein Data Bank

PDC Protein-Detergent Complex

PDIM Phthiocerol Dimycocerosate

$PD_{F/B/T}$ Front, Back & Top Periplasmic Apertures

PE Phosphatidylethanolamine

PEG Polyethylene Glycol

PG Peptidoglycan

PMF Proton Motive Force

PN N-terminal Periplasmic Loop

POT Proton-Dependent Oligopeptide Transporter
PVDF Polyvinylidene Difluoride
qPCR Quantitative Polymerase Chain Reaction
RBS Ribosome Binding Site
RELION Regularised Likelihood Optimisation
Rev-FT Reverse Flow-Through
RFU Relative Fluorescence Units
RMSD Root-Mean-Square Deviation
RND Resistance-Nodulation-Division
RP Resistance Position
RPM Revolutions Per Minute
RSR Real-Space Refinement
RT-PCR Reverse Transcription Polymerase Chain Reaction
S200 Superdex 200 Increase 10/300 GL
SAXS Small-Angle X-Ray Scattering
scFv Single-Chain Variable Fragment
SDM Site-Directed Mutagenesis
SDS-PAGE Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SEC Size-Exclusion Chromatography
SEC-MALS Size-Exclusion Chromatography-Multi-Angle Light Scattering
SIMPLE Single-Particle Image Processing Linux Engine
SL Sulfolipid
SMR Small Multidrug Resistance
STAR Self-Defining Text Archive & Retrieval
Sybody Synthetic Nanobody
TB Terrific Broth or Tuberculosis
TDM Trehalose Dimycolate
TEM Transmission Electron Microscopy
TEV Tobacco Etch Virus
THPP Tetrahydropyrazolo[1,5-*a*]pyrimidine-3-carboxamide
T_m Melt Temperature
TM Transmembrane Helix
TMB 3,3',5,5'-Tetramethylbenzidine
TMD Transmembrane Domain
TMM Trehalose Monomycolate
UN United Nations
v/v Volume Per Volume
VD Vapour Diffusion
VHH HCAb Variable Domain
w/v Weight Per Volume
w/w Weight Per Weight
WHO World Health Organization
XRC X-Ray Crystallography

Introduction

CHAPTER 1

Claiming ~1.4 million lives in 2019 alone, tuberculosis (TB) remains the leading cause of mortality attributable to a single pathogen (World Health Organization, 2020) — i.e. the bacillus *Mycobacterium tuberculosis* (*Mtb*), estimated to latently infect nearly a quarter of the global population (Houben & Dodd, 2016). Discovered as the aetiological agent by Koch (1882), *Mtb* is an aerosol-transmissible and anaerotolerant obligate human pathogen. Whilst capable of causing infection throughout the body, TB largely manifests as a pulmonary disease; *Mtb* principally colonising alveolar macrophages following phagocytic internalisation (Pai et al., 2016). Once intracellular, the bacterium acts to reprogramme these leukocytes to ensure continued subsistence within its replicatory niche — the phagosome. Subsequent disease progression sees formation of granulomas, hallmark lesions of TB lung pathology (Gengenbacher & Kaufmann, 2012). Organised immunological aggregates, comprising a core of mycobacteria-laden macrophages bound by a rim of recruited lymphocytes, granulomas serve to contain the invasion and, being generally successful in this endeavour, frequently lead to asymptomatic, quiescent, and non-communicable *Mtb* infections (Ehlers & Schaible, 2013). However, in ~10% of individuals, weeks to decades after initial exposure, granulomal containment fails; giving rise to systemic bacterial dissemination and rendering the patient contagious (i.e active TB). Left untreated, this proves fatal in approximately half of cases (Pai et al., 2016).

Although standard therapeutic regimes warrant lengthy (≥6 months), combinatorial, and ideally daily, administration of the four first-line antituberculars —

Isoniazid, Rifampicin, Ethambutol & Pyrazinamide — they achieve, by most recent estimations, an impressive 85% cure rate for pan-susceptible *Mtb* infections (Pai et al., 2016; Chakaya et al., 2021). Unfortunately, their serviceability has been curtailed by emergence and spread of drug-resistant strains (Seung et al., 2015). Such isolates, recalcitrant to treatment with one, multiple, or even all first- and second-line compounds (Velayati et al., 2009; Udwadia et al., 2012), represent a tangible threat to global health security — with next to 500,000 cases being reported annually (World Health Organization, 2020), and evidence that prominent resistant strains have been in circulation for at least 40–50 years (Eldholm et al., 2015; Cohen et al., 2015). Notably, unlike other bacterial pathogens which exploit horizontal gene transfer and mobile genetic elements, antibiotic resistance in *Mtb* evolves solely via *de novo* chromosomal mutation — often with worryingly little fitness cost (Eldholm & Balloux, 2016). Future control of TB, if not the near-eradication sought by various international initiatives (e.g. the WHO End TB Strategy, Uplekar et al. (2015); the UN Political Declaration on Tuberculosis, Sahu et al. (2019)), clearly necessitates discovery of next-generation antituberculars; preferably boasting novel mechanisms of action, better safety profiles, and reduced regime durations. Historically, *Mtb* cell wall biogenesis has proven a bounteous reservoir for antitubercular target discovery, a trend that, as it will be seen, looks set to continue (Batt et al., 2020).

1.1 The Mycobacterial Cell Envelope

Mycobacteria, alongside other families of the *Corynebacteriales* order, are typified by architecturally complex and compositionally distinct cell envelopes, intimately linked to their physiology, pathogenicity, and antibiotic susceptibility (Jackson et al., 2021). A thick multi-layered assemblage, with a core of covalently connected strata, the mycobacterial cell wall has a notably high lipid content — lipids accounting for up to 40–70% of dry cell mass (Guinn & Rubin, 2017; Batt et al., 2020). This renders the cell surface extremely hydrophobic, restricting envelope

permeability, and ultimately conferring mycobacteria with intrinsic resistance to many host bactericidal mechanisms and chemotherapeutics (Jankute et al., 2015).

1.1.1 Structural Organisation

Moving outwards from the plasma membrane, which is rich in cardiolipin and mannosylated phosphatidylinotisols, the cell wall comprises two successive glycopolymers — peptidoglycan (PG) and then arabinogalactan (AG) — followed by an atypical outer membrane (OM) and, lastly, a loosely attached capsular medium (Jackson et al. (2021); Figure 1.1[A]). Thus, despite being Gram-positive, mycobacteria are true diderms; benefiting from a defined periplasmic compartment (Jackson, 2014). The principal molecular components of these envelope layers are organised as follows.

Peptidoglycan

Whilst of the same basic construction as in other eubacteria — a glycan backbone of alternating *N*-acetylglucosamine and *N*-acetylmuramic acid residues cross-linked into a mesh via peptide sidechains — mycobacterial PG differs in a number of pertinent ways. Not only are a portion of its muramyl units oxidised, generating *N*-glycolyl moieties thought to partake in hydrogen bonding, it also boasts significantly more peptide cross-links (70–80% vs. 50% in *E. coli*; Batt et al. (2020)). Such modifications augment the mechanical strength of mycobacterial PG, ensuring it provides a robust foundation onto which the rest of the cell wall can be built (Alderwick et al., 2015). To this end, muramic acid C6-hydroxyls act as attachment sites for AG; covalently tethering the heteropolysaccharide through diglycosylphosphoryl bridges (Jankute et al., 2015).

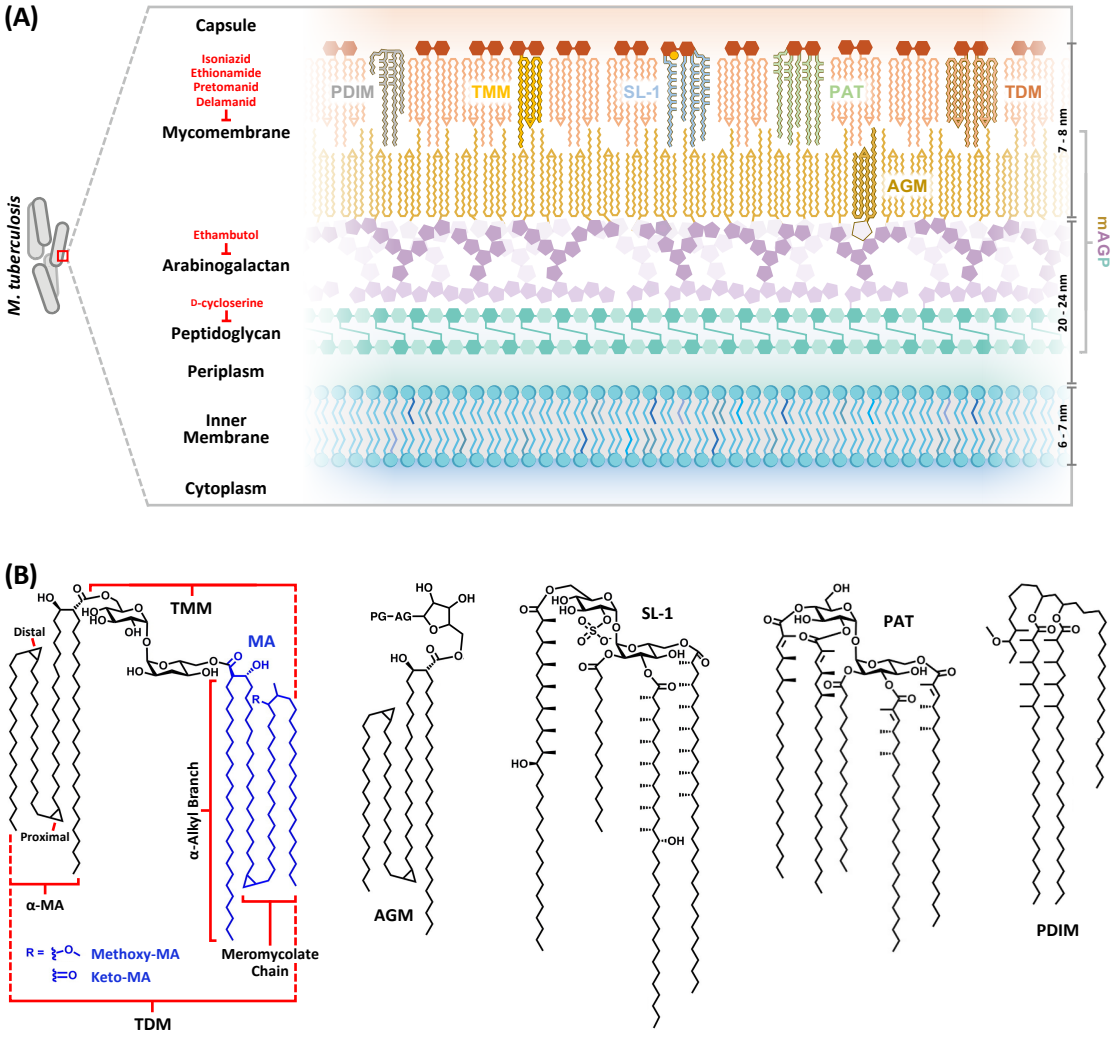


Figure 1.1: *Mtb* Cell Envelope Architecture & Major Mycomembrane Lipids

(A) Stylised representation of the *Mtb* cell envelope. Red text highlights licensed antituberculars known to inhibit biogenesis of specific strata. Cell wall dimensions sourced from Batt et al. (2020). Key mycomembrane lipids are annotated and colour-coded; trehalose monomycolate (TMM), trehalose dimycolate (TDM), arabinogalactan mycolate (AGM), phthiocerol dimycocerosate (PDIM), sulfolipid-1 (SL-1) & penta-acyltrehalose (PAT). Skeletal formulas for these are provided in **(B)**, adapted from Touchette & Seeliger (2017). Labels designate substructures within TDM corresponding to TMM and mycolic acids (MAs), as well as the location and nature of meromycolate functionalities defining the principal MA subclasses. All MAs are depicted as W-conformers.

Arabinogalactan

AG, a tree-like polymer representing ~35% of envelope mass, is predominantly fabricated from galactose and arabinose sugars — both in furanoid form (Alderwick et al., 2015). From a linear galactan trunk, circa 30 monosaccharide units long and affixed to PG, branch three arabinan chains. These, again composed

of ~30 monomers, are themselves highly branched, and end in hallmark hexa-arabinoside motifs whose non-reducing termini are esterified to, and so anchor, the major lipids of the OM inner leaflet — mycolic acids (MAs; Jankute et al. (2015)). Approximately two-thirds of AG termini are mycolated in this way. The resulting mycolyl-arabinogalactan-peptidoglycan (mAGP) complex acts a "covalent skeleton" for the envelope, and is central to its structural integrity and determination of cell shape (Batt et al., 2020).

The Mycomembrane & Mycolic Acids

The OM, or so-called mycomembrane, is an asymmetric bilayer (Figure 1.1[A]). As well as being essential for viability, it is the cell wall facet primarily limiting mycobacterial permeability (Jackson et al., 2021). Accompanying the aforenoted arabinogalactan mycolate (AGM) inner leaflet, the mycomembrane is completed by a menagerie of non-covalently associated outer leaflet (glyco)lipids — featuring both mycobacterially conserved and species-specific constituents (Jackson, 2014). In *Mtb* these include phthiocerol dimycocerosates (PDIMs), di- & penta-acyltrehaloses (DATs & PATs), sulfolipids (SLs), trehalose monomycolate (TMM), and trehalose dimycolate (TDM; Figure 1.1[B]); almost all of which are esters of MAs and/or trehalose (Jackson, 2014). TMM is especially critical to OM biogenesis. Catalysed by the Ag85 complex, a trio of secreted mycoloyltransferases, its mycolate group is transesterified onto AG, or a second TMM molecule, to form AGM and TDM respectively. TMM is thus a vital lipid feedstock for construction of both mycomembrane leaflets (Batt et al., 2020).

First isolated from *Mtb* over 80 years ago (Stodola et al., 1938), MAs have since been established as signature *Corynebacteriales* lipids and are renowned for being the longest biological fatty acids discovered to date (Daffé et al., 2017). Displaying genus-variable chain lengths, making them sought after chemotaxonomic markers (Goodfellow & Jones, 2015), MAs are α-alkyl β-hydroxy fatty acids and contain an astounding ~60–90 carbon atoms in mycobacteria (Figure 1.1[B]).

These are distributed between the meromycolate (C_{42}–C_{62}) and shorter α-branch (C_{20}–C_{26}) chains (Jackson et al., 2021). Differential enzymatic decoration of the former, altering the chemical moieties (e.g. double bonds, cyclopropyl groups, methyl-branches, oxygenated functions) present at specific proximal and distal sites, defines a multitude of MA subclasses (Marrakchi et al., 2014); of which α-, keto-, and methoxy-MAs are the major variants found in *Mtb* (Quémard, 2016).

As electron microscopy indicates the OM is at most 15% thicker than its cytoplasmic counterpart, the consensus is that mycolates must partially fold up to be accommodated into the mycomembrane; lipids in opposite leaflets likely also intercalating (Zuber et al., 2008; Hoffmann et al., 2008). Congruently, amassing evidence suggests the identity and stereochemistry of meromycolate functionalities strongly influence MA conformation and, in turn, OM fluidity (Marrakchi et al., 2014). For example, keto-MAs with an α-methyl *trans*-cyclopropane moiety have a predilection for the "W-conformer" in which the mero-chain undergoes hairpin bending at both functional groups; thereby achieving a compact structure with four hydrocarbon stretches arranged parallely (Villeneuve et al., 2007; Groenewald et al., 2019).

1.1.2 The Cell Wall, Virulence & Therapy

Alongside the capsule, a matrix of low complexity exopolysaccharides (Kalscheuer et al., 2019), the OM is the primary interface with the external environment. Consistently, in *Mtb* many of its components directly modulate host-pathogen interactions (Stanley & Cox, 2013). For instance, beyond its status as the most abundant outer leaflet glycolipid, TDM is a quintessential virulence factor with numerous evidenced roles, throughout the *Mtb* infection cycle, in both immunostimulation and immunoevasion (Hunter et al., 2006). Notable TDM bioactivities include facilitation of macrophage uptake by receptor-mediated phagocytosis, maintenance of intracellular survival through inhibition of phagosome maturation, promotion of tissue damage by elicitation of inflammatory cytokine production,

and inducement of the granulomatous response (Rajni et al., 2011; Jackson, 2014; Daffé et al., 2017). Appositely, the subset of mycomembrane lipids unique to pathogenic mycobacteria (e.g. PDIMs, SLs) are also potent virulence determinants; having been implicated in functions as diverse as phagosomal escape, repression of toll-like receptor signalling, and triggering of the mammalian cough reflex, to name but a few (Ly & Liu, 2020; Ruhl et al., 2020).

Given the centrality of the envelope to *Mtb* survival and pathogenesis, together with its hierarchal construction, it is unsurprising that inhibition of various aspects of cell wall biogenesis are proven routes to antitubercular intervention (Batt et al. (2020); Figure 1.1[A]). In addition to the first-line agents Isoniazid and Ethambutol, blocking steps in MA and AG synthesis respectively, multiple established second-line therapies (e.g. D-Cycloserine, Ethionamide), a slew of preclinical agents, and two out of the only three anti-TB compounds achieving FDA approval in the past half-century (i.e. Delamanid & Pretomanid; TB Alliance (2019)) act by perturbation of envelope assembly (Vilchèze, 2020). In particular, MA biosynthesis is considered the "Achilles' heel" of the tubercle bacillus (Nataraj et al., 2015) — in its absence the mycomembrane fails to form, compromising bacterial viability. MAs are cytoplasmically manufactured by concerted action of ~30 enzymes (Marrakchi et al., 2014). Together they catalyse a well characterised anabolic pathway that culminates in production of plasma membrane TMM and represents a veritable reservoir of essential proteins — all potential, and many validated, antitubercular targets (North et al., 2014). It follows that processes underlying subsequent export of intracellular TMM, for OM incorporation, should likewise be of therapeutic relevance. Accordingly, over the last decade the mycobacterial membrane protein large (MmpL) family transporter MmpL3 — a member of the resistance-nodulation-division (RND) superfamily and the focus of this thesis — has simultaneously emerged as the probable TMM exporter and a promising anti-TB target (Rayasam, 2014; Bolla, 2020). Before considering MmpLs in detail, an introductory synopsis of RND proteins is provided.

1.2 The MmpL Family of RND Transporters

1.2.1 The RND Superfamily

Although infamous as Gram-negative multidrug efflux pumps, members of the RND superfamily are found throughout all domains of life (i.e. archaea, eubacteria & eukarya). Defined by a signature topology — 12 transmembrane helices (TMs) together with substantial extracellular domains in the loops connecting TMs 1–2 & 7–8 — these secondary active antiporters harness ion gradients, predominantly the proton motive force (PMF), to energise export of variegated exogenous and endogenous species (Klenotic et al., 2021). Identified allocrites include heavy metals, antimicrobials, pigments, organic solvents, lipids, toxic metabolites, and sterols, among many others. Consistent with this versatility, RND proteins contribute to physiological phenomena as disparate as developmental signalling (Qi & Li, 2020), polypeptide translocation (Tsukazaki, 2018), quorum sensing (Alvarez-Ortega et al., 2013), and, of course, antibiotic resistance (Du et al., 2018). Phylogenetically, the superfamily comprises at least nine separable families (Figure 1.2); segregating RND members by organismal class and function (Yen et al., 2010). For example, whilst seven of the families are prokaryotic, only Gram-negative RND transporters populate the hydrophobe/amphiphile efflux (HAE)-1 and heavy metal efflux (HME) families; whilst those of Gram-positives cluster exclusively within the HAE2 clade. Eukaryotic RNDs, meanwhile, are strictly confined to the remaining two families — the eukaryotic sterol transporters (ESTs) and dispatched proteins (Nikaido, 2018).

With most families now having had representatives structurally characterised, it is increasingly evident that functional specialisation has evolved through elaboration of the size and fold of the RND extramembranous domains, as well as by modulation of oligomeric state (Zhang et al. (2018); Klenotic et al. (2021); Figures 1.2 & 1.3[A]). In contrast, the transmembrane domain (TMD) displays remarkable structural, if not sequence, conservation across the entire superfamily;

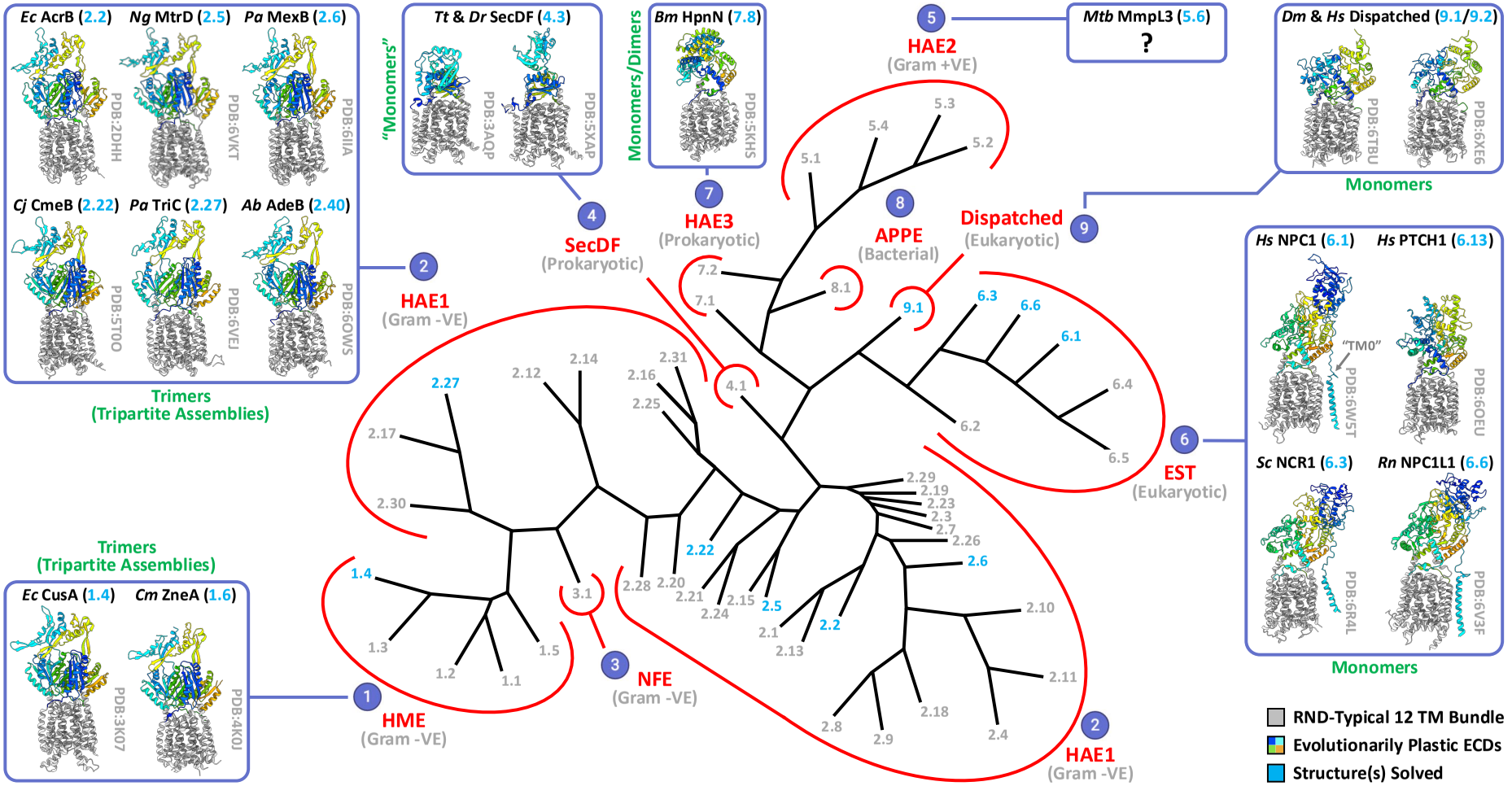


Figure 1.2: Structure-Annotated Phylogenetic Tree of the RND Superfamily

Superfamily phylogram, drawn from Yen et al. (2010), semi-exhaustively labelled with solved RND structures; each depicted as a single cartoon protomer coloured by domain organisation (TMs grey, extracellular domains (ECDs) rainbow). Family organismal restriction and oligomeric state are listed. Numbers denote Transporter Classification Database (Saier et al., 2021) identifiers when prepended with 2.A.6. Structures composed in UCSF ChimeraX v.1.1 (Pettersen et al., 2021).

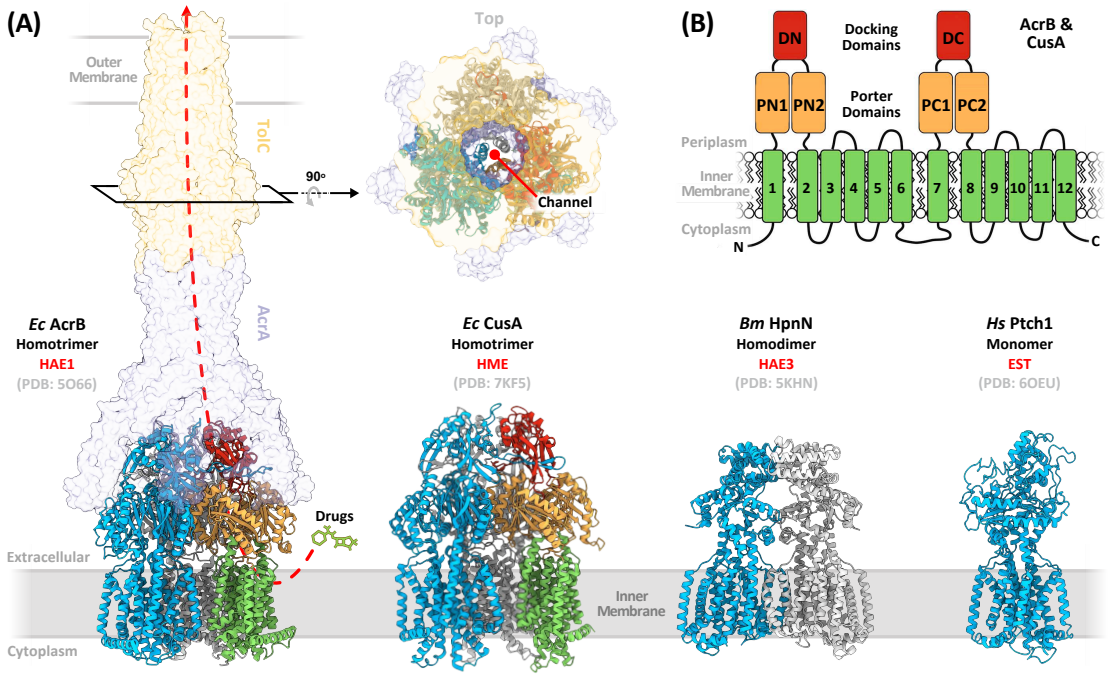


Figure 1.3: Variation in RND Transporter Oligomeric State

(A) A collection of RND structures of differing oligomeric state, each in cartoon form, contextualised within a common membrane (grey box). Protomers are coloured distinctly. For AcrB (PDB: 5O66, Wang et al. (2017)) and CusA (PDB: 7K5F, Moseng et al. (2021)) one monomer is shaded equivalently to the topology diagram presented in **(B)**; adapted from Chalut (2016). Moreover, AcrB is shown assembled with AcrA (lilac) and TolC (yellow); both rendered as transparent surfaces. The top view of this tripartite complex emphasises the export channel employed in drug-efflux (dashed red arrow). Structures formatted using UCSF ChimeraX v.1.1 (Pettersen et al., 2021).

reliably forming two six-helix bundles (TMs 1–6 & 7–12) related by a central two-fold pseudo-symmetry axis. Alongside incorporation of $\alpha + \beta$ sandwiches in both extracellular loops, said TMD configuration betrays the genesis of ancestral RND transporters by internal gene duplication events (Zhang et al., 2018) — an origin first suggested on early bioinformatic analysis of the superfamily (Saier et al., 1994; Tseng et al., 1999). That evolution is seemingly preoccupied with the architecture of the extracytoplasmic domains, which account for up to 60% of RND protein mass (Du et al., 2018), is congruent with their often sole provision of the substrate export route; the TMD instead serving as the conduit for proton influx (Nikaido, 2018). Marked separation of the pathways traversed by substrates and protons distinguishes RND pumps from other PMF-dependent

superfamilies (e.g. MFS, SMR & MATE), and, by in essence decoupling determinants of transporter energisation and substrate recognition, is likely paramount to construction of antiporters with readily diversifiable specificity (Nikaido, 2018).

In broad terms, the mechanism of RND export involves sequential binding and release of protons in the TMD; facilitating their transit from the external environment into the cytoplasm, and thus down their electrochemical gradient (Yamaguchi et al., 2015). In particular, a cluster of charged residues at the centre, and interface, of TMs 4 & 10 — typically featuring one or more aspartates — have been implicated as essential components of the proton-relay network in multiple RND families (Guan & Nakae, 2001; Franke et al., 2003; Tsukazaki et al., 2011; Bernut et al., 2016). Crucially, changes in TMD protonation state are remotely coupled to large conformational changes in the extracellular loops, and it is these that drive target extrusion (Eicher et al., 2014). Despite a smattering of reports hinting RND systems are capable of acquiring their substrates from the cytosol (Aires & Nikaido, 2005; Pak et al., 2013; Moseng et al., 2021), on the whole the extramembrane domains are thought to capture export substrates proximal to the surface of, or residing within, the plasma membrane outer leaflet (Klenotic et al., 2021).

By far the biggest contributor to structural, biochemical, and cellular comprehension of RND transporters has been investigation of acridine resistance protein B (AcrB) — an *E. coli* HAE1 polyspecific drug-efflux pump and, in many respects, the paradigmatic, model, superfamily member (Kobylka et al., 2020). Although the first RND protein to be crystallised (Murakami et al., 2002), AcrB has remained the focus of structural pursuits; accounting for well over 100 PDB depositions at the time of writing. Together these reveal the Gram-negative transporter to assemble into a jellyfish-esque homotrimer, in which the majority of inter-protomer contacts are formed within the periplasmic headpiece (Figure 1.3[A]). Both extracellular loops of each monomer fold into three subdomains — two homologous porter domains topped by an aptly named docking domain

(Klenotic et al. (2021); Figure 1.3[B]). The latter recruits AcrA, a hexameric membrane fusion protein (MFP), that appends an α-helical channel to top of the RND pump. This in turn serves as an adaptor for TolC, a trimeric outer membrane factor (OMF) that, continuous with the pore of AcrA, extends said channel across the Gram-negative OM (Kobylka et al., 2020). The resulting RND-MFP-OMF tripartite complex (3:6:3 stoichiometry), the standard assembly for members of the HAE1 and HME families, permits AcrB to realise transenvelope efflux (Figure 1.3[A]); exported substrates exiting into the AcrA-TolC lumen to be shuttled out of the cell (Alav et al., 2021). Extensive study has suggested AcrB employs a "functional rotation mechanism", in which, at any given moment, the three monomers are in non-equivalent conformations — rendering the homotrimer asymmetric whilst active. TMD-mediated PMF transduction consecutively, and cooperatively, cycles each protomer through three conformational states; in doing so facilitating substrate binding, within a porter domain, followed by its peristaltic export, in each individual monomer (Zwama & Yamaguchi, 2018). As will become evident, the overall applicability, to MmpLs, of insights garnered from AcrB and other prominent Gram-negative RND pumps remains to be determined.

1.2.2 MmpLs: Endogenous Lipid Transporters

Only discovered after sequencing of the *Mtb* genome, the genes of the MmpL family were initially annotated as "conserved large membrane proteins" of unknown function (Cole et al., 1998); the MmpL designation referencing both the predicted size of their multipass products, frequently >100 kDa, as well as a perceived phylogenetic restriction to mycobacteria. Follow up *in silico* analyses promptly noted topological similarities with the wider RND superfamily, and likewise identified close linkage between multiple *mmpL* genes and loci involved in fatty acid metabolism — ultimately driving speculation that MmpLs function in lipid transport (Tekaia et al., 1999). Preliminary experimental support for this hypothesis was forthcoming, with transposon mutagenesis studies implicating

MmpL7 in translocation of PDIMs the very same year (Cox et al., 1999). Over two decades on, MmpLs, now recognised to pervade the *Corynebacteriales* order, have been firmly established as HAE2 family RND transporters; predominantly responsible for export of endogenous lipids crucial to OM biogenesis (Jackson et al., 2021). Consonantly, the number of MmpLs per genome correlates with mycomembrane lipid diversity — as illustrated, for example, by the presence of 4, 13, or up to 31 *mmpL* genes in *Corynebacterium glutamicum* (*Cgl*; Yang et al. (2014)), the *Mtb* reference strain H37Rv (Domenech et al., 2005), and select isolates of *M. abscessus* (Melly & Purdy, 2019) respectively.

Of the thirteen *Mtb* MmpLs, seven have experimentally assigned functions (Figure 1.4[A]) — five in lipid transport (MmpL3/7–8/10–11) and two in siderophore egress (MmpL4–5; Melly & Purdy (2019)). Alongside the aforementioned roles of MmpL3 (expounded in Section 1.3) and MmpL7, MmpL8 exports SLs (Converse et al., 2003; Domenech et al., 2004), MmpL10 DATs (Belardinelli et al., 2014), and MmpL11 biofilm-related species such as mycolate wax esters (MWEs; Pacheco et al. (2013); Wright et al. (2017)). MmpL4 and MmpL5, meanwhile, are reported to have semi-redundant activities in the secretion of mycobactins (MBTs); small molecule scavengers of host iron ions (Wells et al., 2013). Testament to the physiological importance of the substrates they transport, individual knockout of all bar one of these characterised MmpLs attenuates *Mtb* virulence in mice (Domenech et al., 2005; Lamichhane et al., 2005). It is the indispensability of TMM to mycomembrane formation that renders MmpL3 the exception, it being incapable of genetic inactivation. This essentiality having been confirmed in numerous independent studies (Domenech et al., 2005; Grzegorzewicz et al., 2012; Varela et al., 2012; Li et al., 2016; DeJesus et al., 2017; Degiacomi et al., 2017; McNeil & Cook, 2019). Export of other OM lipids may be undertaken by some of the six *Mtb* MmpLs yet to have designated functions (MmpL1–2/6/9/12–13); although two have been conjectured defunct in modern *Mtb* strains on the basis of open reading frame truncation (MmpL6; Arumugam et al. (2019)) or splitting (MmpL13a/b; Melly & Purdy (2019)).

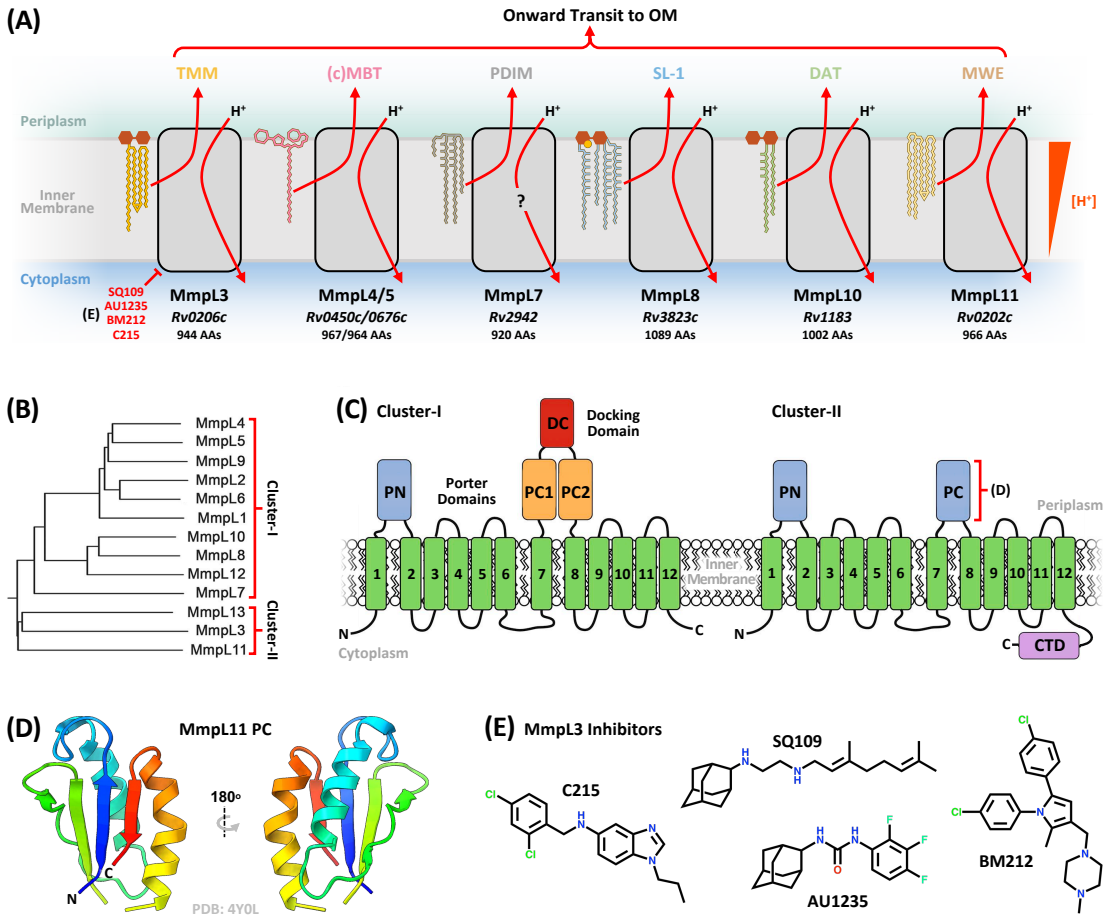


Figure 1.4: Overview of MmpL Function, Topology, Structure & Inhibitors

(A) Schematic summarising the primary roles of the seven *Mtb* MmpLs assigned functions. All, with possible exception of MmpL7, are expected to undertake proton-coupled antiport. The H37Rv gene identifier and length of each protein are also given. Allocrite abbreviations as in Figure 1.1, plus mycolate wax ester (MWE), di-acyltrehalose (DAT) & (carboxy)mycobactin ((c)MBT). **(B)** *Mtb* MmpL phylogenetic tree revealing two divergent clusters, adapted from Chim et al. (2015). **(C)** Predicted topologies of Cluster-I & -II MmpLs, modified from Chalut (2016). **(D)** Cartoon representations of the C-terminal porter domain of MmpL11 (PDB: 4Y0L, Chim et al. (2015)), coloured blue (N-terminus) to red (C-terminus). **(E)** Skeletal formulae of the first preclinical antituberculars identified to target MmpL3, prepared with ChemDraw Professional v.17.0.0.206. Structures rendered with UCSF ChimeraX v.1.1 (Pettersen et al., 2021).

MmpL Topology, Phylogeny & Interactors

In line with their RND classification, full-length MmpLs — comprising 920 (MmpL7) to 1146 (MmpL12) residues in *Mtb* — are predicted to contain 12 TMs interspersed by sizeable extracytoplasmic domains in the loops connecting TMs 1–2 & 7–8 (Chalut, 2016); a topology most thoroughly validated for MmpL3 (Carel et al., 2014; Belardinelli et al., 2016). Notably, MmpL periplas-

mic architecture is expected to deviate significantly from that espoused by stalwart RND transporters and likewise differentiates the MmpL family into two phylogenetic subgroupings (Melly & Purdy (2019); Figure 1.4[B]). Cluster-I MmpLs (*Mtb* MmpL1−2/4−10/12) host a much smaller extramembranous loop between TMs 1−2, relative to AcrB (~150 vs. 300 residues), whereas the sequence bridging TMs 7−8 (~350 residues) is thought to attain the canonical tri-subdomain (i.e. twin porter modules & a docking unit) configuration (Chim et al. (2015); Figure 1.4[C]). In comparison, both extracellular loops of Cluster-II MmpLs (*Mtb* MmpL3/11/13) are of the more diminutive variety; these proteins also featuring a cytoplasmic C-terminal domain (CTD, Chim et al. (2015)). The only high resolution insight into MmpL structure preceding the start of my DPhil, a crystal structure of the second periplasmic loop of *Mtb* MmpL11 (Figure 1.4[D]), revealed a fold equivalent to a single porter subdomain (Chim et al., 2015). By homology, all short form MmpL loops (i.e. Cluster-I & -II) can be anticipated to fold comparably.

Given the extracellular domains of prototypical RND pumps are key to transporter oligomerisation, substrate recognition, and OMF-coupling via MFPs (Nikaido, 2018), the divergence of periplasmic architecture in MmpLs indicates they are likely to differ mechanistically from the HAE1/HME families — despite being thought to similarly undertake substrate translocation via proton-coupled antiport (Bernut et al., 2016). Corroborating said disparity, there is no evidence MmpLs assemble into tripartite complexes spanning both mycobacterial membranes (Jackson et al., 2021). Nevertheless, MmpLs almost certainly act in unison with various binding partners to realise their export functions. Prominent among the many purported MmpL interactors are the analogously named mycobacterial membrane protein small (MmpS) family. Encoded in putative operons alongside four *Mtb* MmpLs, the correspondingly numbered MmpS proteins (MmpS1−2/4−5) contain but ~150 residues; representing a soluble periplasmic domain anchored by a single N-terminal TM (Chalut, 2016). Notwithstanding

the absence of homology, their genomic colocalisation with MmpLs has led to suggestions MmpSs may execute MFP-like roles. Whilst MmpS1 and MmpS2 remain uncharacterised, MmpS4 and MmpS5 have been shown to complex their cognate transporters (i.e. MmpLs 4 & 5), are required redundantly for siderophore secretion (**Wells et al., 2013**), and, in the case of MmpS5, have been implicated in MmpL5 homotrimerisation *in vivo* (**Yamamoto et al., 2021**).

As mycobacteria lack proteins of sequence similarity to *E. coli* TolC, and given the envelope membranes are separated by several hundred angstroms (**Zuber et al., 2008**), MmpLs must function in concert with periplasmic chaperones to successfully deliver lipidic cargo to the OM (**Jackson et al., 2021**). Lipoproteins are among the prime candidates for this role, a number of studies supporting that some of the ~100 encoded in *Mtb* (**Becker & Sander, 2016**) act as OM lipid shuttles (**Touchette & Seeliger, 2017**). Congruously, structures of the *Mtb* lipoproteins LppX (PDB: 2BYO, **Sulzenbacher et al. (2006)**), LprG (PDB: 3MHA, **Drage et al. (2010)**) and LpqN (PDB: 6E5F, **Melly et al. (2019)**) all reveal hydrophobic cavities occupied by lipids or lipid-esque molecules. These three are of particular relevance as LppX is required for OM localisation of PDIMs (**Sulzenbacher et al., 2006**), whilst both LprG and LpqN ostensibly interact with Ag85, MmpL3 and/or MmpL11 (**Touchette et al., 2017; Melly et al., 2019**). In common with the reported MmpL-MmpS interface (**Wells et al., 2013**), MmpLs appear to contact said lipoproteins via their extracellular loops (**Melly et al., 2019**) — suggesting, akin to other RND transporters, that the periplasmic domains of MmpLs are key mediators of protein-protein interactions.

Beyond lipoproteins and MmpSs, prevailing hypotheses posit MmpL interactions serve to nucleate multiprotein complexes coupling lipid biosynthesis to export (**Chalut, 2016**). Aside from frequent colocalisation of *mmpL* and lipid synthase genes, such clusters often syntenously conserved across mycobacterial species (**Viljoen et al., 2017**), this “scaffold” model stems from observations that *Mtb* mutants deficient in MmpL8 or MmpL10 intracellularly accumulate precur-

sors of SLs and DATs respectively (Touchette & Seeliger, 2017) — implying the final stages of lipid synthesis are coupled to translocation. Although conceptually appealing, biochemical evidence for these scaffolding complexes remains scant; essentially amounting to a topologically questionable interaction between MmpL7 and the PDIM polyketide synthase PpsE (Jain & Cox, 2005), as well as an *M. bovis* complex, identified by native-PAGE, containing MmpL8 and at least three genomically local liposynthetic enzymes (Zheng et al., 2011). Whilst *E. coli* two-hybrid screens recently hinted MmpL3 may have an elaborate interactome, almost none of the hits pertained to MA biosynthesis (Belardinelli et al., 2019).

To summarise, the MmpL family have evolved to be specialist transporters of specific endogenous lipophiles. They have co-opted RND architecture to export a diverse collection of complex, bioactive, OM lipids that are equally critical for mycomembrane construction and *Mtb* virulence. Consistent with their functional and topological divergence from archetypal Gram-negative RND pumps, MmpLs are increasingly understood to act with an array of similarly bespoke, albeit poorly characterised, binding partners. Ultimately, the MmpLs of *Mtb* represent a family, of structurally understudied integral membrane proteins, offering significant potential as novel antitubercular targets. Keeping this overview in mind, attention now returns to MmpL3 — a Cluster-II MmpL distinguished by its essentiality and proven druggability.

1.3 MmpL3 in Focus

Comprising 944 residues in *Mtb*, MmpL3 orthologues are present throughout the *Mycobacteriaceae* family; corroborating the indispensability of the protein to mycobacterial growth *in vitro* (Domenech et al., 2005), during infection in mice (Li et al., 2016), and within human macrophages (Degiacomi et al., 2017). TMM was first conjectured as a glycolipid carrier for MA export by Takayama et al. (2005), although translocation was erroneously proposed to be via an

ATP-binding cassette (ABC) transporter. Delineation of MmpL3 as the genuine TMM exporter took a further seven years, a flurry of near back-to-back publications (Grzegorzewicz et al., 2012; Tahlan et al., 2012; Stanley et al., 2012; La Rosa et al., 2012; Varela et al., 2012) uncovering its fundamental function and related therapeutic exploitability. Predominantly, these studies inferred the MmpL to be targeted by a handful of disparate experimental antituberculars — namely BM212 (La Rosa et al., 2012), AU1235 (Grzegorzewicz et al., 2012), SQ109 (Tahlan et al., 2012) & C215 (Stanley et al. (2012); Figure 1.4[E]) — owing to their exposure selecting resistance mutations in the *mmpL3* gene. More importantly, treatment with said mycobactericidal compounds was found to induce intracellular accumulation of TMM, abolition of OM mycolation (i.e. loss of AGM & TDM), and whole-cell ultrastructural changes; outcomes phenocopying conditional depletion of MmpL3 (Varela et al., 2012), and thus convincingly establishing it as the long-sought TMM transporter. Strikingly, predating its classification as an MmpL3 inhibitor SQ109 had entered clinical trials (Cole, 2012); it now having completed at least Phase I, and up to Phase IIb-III, trials in multiple territories (Butler & Paterson, 2020; Degiacomi et al., 2020). In exemplifying the aptitude of MmpL3 as a target for next-generation antituberculars, the aforementioned agents laid the foundation for ongoing pharmaceutical interest in the RND exporter. Consequently, today MmpL3 is the putative target for a multitude of preclinical drugs (Shao et al., 2020). These span more than nine chemical classes (see Chapter 7) and, encouragingly, a number synergise with existing licensed antituberculars (Nikonenko et al., 2007; Stec et al., 2016; Li et al., 2017). Even more appealingly, in offering a novel mode of action, MmpL3 inhibitors should be widely efficacious against the drug-resistant *Mtb* strains currently in circulation (Rayasam, 2014).

Given the substantial medicinal chemistry efforts against MmpL3, the protein has been subject to surprisingly few biochemical investigations. Robust characterisation presumably being stymied by both protein essentiality and difficulties

inherent in designing *in vitro* transport assays featuring substrates as unwieldy as TMM. In contrast to some Cluster-I MmpLs, *Mtb* MmpL3 lacks genomically local *mmpS* or lipid synthase genes. Its depletion is likewise inconsistent with a coupled biosynthesis and export model — the affects of genetic or small molecule inhibition of MmpL3 arising solely from mislocalisation of TMM, unaccompanied by changes to the total mycolate pool (**Tahlan et al., 2012**). Notably, of the many reported MmpL3 binding partners only two, both transmembrane proteins, have been shown essential for TMM export — that is TmaT (**Belardinelli et al., 2016, 2019**) & TftA (**Fay et al., 2019**). Whilst the function of the latter remains unknown, orthologues of the former have been demonstrated to be TMM mycolyl acetyltransferases in both corynebacteria (**Yamaryo-Botte et al., 2015**) and mycobacteria (**Belardinelli et al., 2016**); prompting suggestions transient TMM acetylation is a prerequisite, possibly a signal, for MmpL3-mediated export.

To date, the most tractable system for MmpL3 functional dissection has leveraged the viability of an *M. smegmatis* (*Msmg*) Δ*mmpL3* strain to be rescued by episomal expression of wild-type *Mtb* MmpL3. Using this setup, **Belardinelli et al. (2016)** assessed the functionality of ~60 *Mtb* MmpL3 mutants; in doing so identifying seven essential residues (D251, S288, G543, D640, Y641, D710 & R715) and, in combination with envelope mycolate quantification, a further ten positions compromising mycobacterial growth and/or MmpL3 activity on substitution for cysteine. A proportion of the latter, however, only displayed phenotypes on addition of sulfhydryl-reactive compounds. Absence of a full-length MmpL structure has precluded proper interpretation of these 17 sites, and for that matter the amassing catalogue of MmpL3 mutants conferring resistance to preclinical agents. Nevertheless, by analogy with better characterised RND transporters (i.e. AcrB), the cluster of essential residues mapping to the centres of TMs 4 (D251) & 10 (D640 & Y641) have reasonably, and repeatedly, been posited as probable PMF couplers (**Bernut et al., 2016; Belardinelli et al., 2016**). Along with neighbouring Y252, these residues form an Asp-Tyr tetrad conserved, with

the exception of MmpL7, in all full-length *Mtb* MmpLs — consonant with them having a key role in transporter energisation (Székely & Cole, 2016). Even so, the mechanism by which the PMF drives MmpL export remains unclear.

Through expression of C-terminally truncated constructs, episomal complementation also revealed the MmpL3 CTD to be superfluous to Δ*mmpL3* strain rescue and thus, presumably, TMM export (Belardinelli et al., 2016). Intriguingly, an earlier study expressing *Mtb* MmpL3-GFP fusions in *Msmg* revealed protein accumulation at the cell poles and septa — sites thought to coordinate envelope biogenesis with cell elongation and division (Carel et al., 2014). This localisation was lost on CTD deletion, strongly suggesting the sizeable cytoplasmic domain (~200 residues), whilst irrelevant to TMM translocation, acts to positionally regulate MmpL3 during the mycobacterial cell cycle.

The only other method yielding prominent insights into MmpL3 function executes innovative, yet technically involved, assays on *Msmg* spheroplasts. In combination with chemical and enzymatic tools probing the topology of plasma membrane TMM, Xu et al. (2017) used these envelope-less cells to show MmpL3 may possess a TMM flippase activity — i.e. facilitating redistribution of the glycolipid from the inner to the outer leaflet of the cytoplasmic membrane. Whether this functionality is distinct from TMM transport, and to what extent flipping is an artefactual consequence of removing the OM TMM sink, is yet to be determined. Nonetheless, the spheroplast system has since seen fruitful employment, as a transport assay proxy, in the evaluation of multiple early-stage MmpL3 inhibitors (Shetty et al., 2018; Dupont et al., 2019; Li et al., 2020b). Similarly reliant on metabolic labelling, to achieve bioorthogonally reactive mycolates, use of photoactivatable TMM analogues recently verified MmpL3 interacts with its transport substrate *in cellulo* (Kavunja et al., 2020). Outside of lipid export both MmpL3 and MmpL11 have also been mooted, somewhat controversially, to moonlight as *Mtb* haem importers; in part on the basis that their first periplasmic loop binds the metallated porphyrin when purified in isolation (Tullius et al., 2011; Owens et al., 2013).

Given almost all characterised RND proteins are ion-coupled antiporters (Nikaido, 2018), it is difficult to comprehend how the Cluster-II MmpLs could energise transport of chemically unrelated species in opposite directions. Accordingly, these auxiliary roles are now largely discounted (Mitra et al., 2019; Jackson et al., 2021). Other *Mtb* MmpLs are linked to more reputable secondary functions however, most notably the participation of MmpL5 in drug-efflux (Briffotaux et al., 2017).

1.4 Aims & Objectives

Undeniably, compounds targeting MmpL3 have the potential to make a significant impact on the treatment of TB — a disease that has claimed in excess of a million lives per year for over a decade, and whose containment has been repeatedly confounded by emergence and propagation of drug-resistant strains. However, expediting clinical arrival of a first-in-class MmpL3 inhibitor necessitates drastic improvements in the functional and pharmacological understanding of the essential transporter. Among the many pertinent questions remaining unanswered at the inception of my DPhil were:

- How does MmpL3 engage TMM, and by what mechanism is the PMF transduced to facilitate glycolipid export?
- How do accessory proteins interact with MmpL3, and what do they contribute to TMM translocation?
- How does the CTD mediate MmpL3 recruitment to envelope biogenesis foci?
- How is MmpL3 targeted by a chemically diverse array of novel antitubercular, and do they share a common binding pocket and/or mode of action?
- How do resistance-conferring MmpL3 variants circumvent preclinical agents inhibiting the transporter?

Adequately resolving these lines of inquiry requires experimental strategies spanning multiple disciplines. That said, the single most revelatory advance would be elucidation of an *Mtb* MmpL3 structure to high resolution. This is the primary aim of the thesis. Such an atomic model could guide future biochemical dissection of TMM export, will be vital to contextualisation of the accruing mutational data on MmpL3 and, crucially, can be expected to usher in an era of structure-based drug design. Moreover, in absence of any full-length HAE2 family structures at project outset, knowledge of MmpL architecture was also anticipated to be informative with regards to RND superfamily evolution. The following Chapters detail successful pursuit of an *Mtb* MmpL3 structure, as well as its subsequent analysis, achieved though hard-won and sequential completion of the objectives outlined below:

1. Identify an MmpL3, or homologue, construct capable of robust heterologous expression and purification to homogeneity (Chapter 3).
2. Biophysically characterise the purified protein to confirm suitability for structural studies (Chapter 3).
3. Delineate sample conditions permitting structure solution by X-ray crystallography or cryogenic electron microscopy (Chapters 4–7).
4. Evaluate the resultant structure with respect to the MmpL3, and wider RND, literature to provide new insights (Chapter 7).

Note, from approximately midway through the research detailed herein, crystal structures of *Msmg* MmpL3 alone and in complex with various small molecule inhibitors began to be reported (**Zhang et al., 2019; Su et al., 2019; Yang et al., 2020**). Discussion of these structures is presented alongside that of the *bona fide* antitubercular target, the *Mtb* orthologue, in Chapter 7.

Materials & Methods

CHAPTER 2

2.1 Construct Cloning

2.1.1 *E. coli* Vectors

E. coli codon-optimised genes for all targets (*Mtb* MmpL3/5/7/11, *Cgl* CmpL1/3 & NCgl-1194) were synthesised by Invitrogen GeneArt services; provided within the vector pMA-T or as DNA fragments. From these, open reading frames (ORFs) were PCR amplified (KOD Hot Start DNA Polymerase, Novagen) with primers adding the restriction sites (NdeI, BamHI, XhoI or KpnI) requisite for cloning. Once digested (FastDigest Restriction Enzymes, Thermo Fisher Scientific), amplicons were ligated (Quick Stick Ligase, Bioline) into the isopropyl-β-D-thiogalactopyranoside (IPTG)-inducible vectors pWaldo-GFPd/e (Waldo et al., 1999; Drew et al., 2001) and/or pET-11c-TEV-His; provisioning C-terminal TEV-GFP-His_8 and TEV-His_8 tags respectively.

Clone recovery and plasmid propagation were via transformation into OmniMAX *E. coli* cells (Invitrogen), plated on Luria Broth (LB) agar (37 °C) or inoculated in LB liquid media (37 °C, 180 RPM), grown overnight in the presence of selective antibiotic — kanamycin (Kan, 50 μg mL^{-1}) or ampicillin (Amp, 100 μg mL^{-1}). Successful gene insertion was verified by diagnostic digest and Sanger sequencing (Source BioScience). All miniprep and DNA clean-up steps (i.e. PCR purification & gel extraction) were performed with GeneJET kits (Thermo Fisher Scientific) according to the manufacturer's instructions.

With alternative oligos, otherwise identical procedures were employed to subclone C-terminally FLAG-tagged (MmpL3/CmpL1-FLAG), CTD truncated ($MmpL3_{1\text{-}753}$, $MmpL11_{1\text{-}714}$ & $MmpL11_{1\text{-}693}$), and periplasmic subdomain (MmpL3-$PD1_{32\text{-}187}$ & MmpL3-$PD2_{417\text{-}551}$) constructs. Individual subdomain deletions (MmpL3-$PD1\Delta_{1\text{-}36,182\text{-}753}$ & MmpL3-$PD2\Delta_{1\text{-}421,551\text{-}753}$) were obtained by first amplifying gene regions either side of the sequence to be removed, using primer pairs boasting compatible overhangs, and then fusing the resultant amplicons by overlap PCR. Fused products were restriction cloned into pWaldo-GFPe as above.

Constructs for sybody selection ($MmpL3_{1\text{-}753}$-Avi-GFP-TEV-His_8, CmpL1-Avi-FLAG) were generated by introduction of an Avi-tag downstream of the target C-terminus. This was achieved by site-directed mutagenesis of previously cloned constructs. Following PCR amplification of the entire plasmid, with partially complementary forward and reverse primers designed for Avi-tag insertion, template vector was removed with DpnI (FastDigest Restriction Enzymes, Thermo Fisher Scientific). The digested PCR product was then transformed into OmniMAX cells for clone recovery and verification as described previously.

2.1.2 *M. smegmatis* Vectors

Targets to be trialled for *Msmg* expression (MmpL3/5, CmpL1/3 & NCgl-1194) were cloned into the mycobacterial, *E. coli* replicable, and acetamide-inducible vector pACE (Addgene #110085, **Arnold et al. (2018)**) via the fragment exchange (FX) approach (**Geertsma & Dutzler, 2011**); thereby furnishing C-terminal 3C-GFP-His_{10} tags.

ORFs were first PCR amplified to add the necessary restriction sites, using primers designed with the fxcloning web tool. Next, in a one-pot reaction, plasmid and insert were digested with BspQI (FastDigest Restriction Enzymes, Thermo Fisher Scientific). Finally, post heat inactivation, the mixture was supplemented with ATP and T4 ligase prior to transformation into OmniMAX cells; subsequently

plated on LB agar with apramycin (Apr, 25 μg mL^{-1}). In exchanging a *ccdB* suicide cassette for the gene of interest, the FX strategy ensures all resultant colonies harbour the desired insert. Nevertheless, cloning success was confirmed by diagnostic PCR and Sanger sequencing. The *ccdB*-resistant *E. coli* strain DB3.1 (Invitrogen) was used for empty pACE propagation.

2.1.3 Sybody Vectors

Sybodies selected against MmpL3$_{1\text{-}753}$ by ribosome display were initially cloned into the phagemid pDx_init (Addgene #110101). From here, over the remainder of the selection cascade, binder ORFs were sequentially subcloned into the ELISA vector pSb_init (Addgene #110100) and then the MBP-fusion expression vector pBXNPHM3 (Addgene #110099). All steps employed FX cloning, and are described in detail elsewhere (Zimmermann et al., 2018, 2020).

2.2 Protein Expression & Purification

2.2.1 Membrane Protein Expression in *E. coli*

Expression of pWaldo and pET-11c constructs was in *E. coli* C43(DE3) cells (Lucigen, Miroux & Walker (1996)). Transformants, via intermediary starters, were used to seed multilitre Terrific Broth (TB) cultures supplemented with Amp or Kan (37 °C, 215 RPM). At an OD$_{600}$ of ~0.6–0.8, the temperature was dropped and expression induced overnight by addition of 0.4 mM IPTG (25 °C, 215 RPM). Expression level was determined by whole-cell GFP-fluorescence (λ_{Ex}: 488 nm, λ_{Em}: 512 nm), from 1 mL of culture resuspended in 200 μL of 1X PBS, measured on a SpectraMax M3 Microplate Reader (Molecular Devices). For those constructs lacking a GFP reporter, initial expression was verified by anti-tag western blot (Sigma-Aldrich Mouse α-His or α-FLAG, Agilent Dako Goat α-Mouse-HRP) using standard wet transfer (Immobilion-P PVDF Membranes, Millipore) and chemiluminescence detection (Amersham ECL Detection Reagent

& Hyperfilm, Cytiva) methods. Post-expression, cells were harvested (~9000 g, 15 min), resuspended in 1X PBS (5 mL g^{-1}), and lyophilized DNAse I (Sigma-Aldrich) added prior to freezing.

2.2.2 Membrane Preparation

Thawed *E. coli* cell pellets were lysed by single pass through a TS Series Continuous Cell Disruptor (Constant Systems Ltd), equipped with NESLAB ThermoFlex 1400 Chiller (Thermo Fisher Scientific), running at 25 kpsi and pre-cooled to 4 °C. After lysate clarification (30,000 g, 30 min, 4 °C), membranes were pelleted by ultracentrifugation (>200,000 g, 2 h, 4 °C) and then resuspended in 1X PBS (6 mL g^{-1}) by Dounce homogenisation, aliquoted, and snap-frozen.

2.2.3 Fluorescence Size Exclusion Chromatography

The protocol for detergent screening by fluorescence-detection size-exclusion chromatography (FSEC) was adapted from published methods (Kawate & Gouaux, 2006). For a given GFP-fusion construct, an aliquot of membrane was thawed, split into 200 μL fractions, and each solubilised in one of a panel of detergents (Anatrace) at 1% (w/v) and in a final volume of 400 μL 1X PBS. After ~1 h rotation at 4 °C, non-solubilised material was removed by ultracentrifugation (200,000 g, 20 min, 4 °C). GFP counts were measured pre- and post-centrifugation to estimate solubilisation efficiency. Per detergent, and in a successive fashion, 250 μL of recovered supernatant was injected onto a Superose 6 10/300 GL column (Cytiva); pre-equilibrated in FSEC buffer (20 mM Tris pH 7.5, 150 mM NaCl, 0.03% DDM) and in-line with a RF-20A Fluorescence Detector (Shimadzu) set to monitor GFP.

2.2.4 Membrane Protein Purification

Purification of TEV-GFP-His_8 fusions was as broadly described elsewhere (Drew et al., 2005; Newstead et al., 2011). All steps were undertaken at 4 °C. Thawed

membranes were first resuspended in 1X PBS supplemented with 150 mM NaCl, before solubilisation in 1% (w/v) DDM or LMNG (Anatrace) under gentle agitation for ~1–1.5 h. Non-solubilised material was removed by ultracentrifugation (>200,000 g, 1 h), and then imidazole to a final concentration of 15 mM added to the recovered supernatant. Protein in this was bound to HisPur-Ni-NTA resin (Thermo Fisher Scientific) in batch with stirring, and incubated until binding plateaued as determined by GFP counts. The resin, packed into a glass econo-column (Bio-Rad), was next sequentially washed with 10 column volumes (CVs) of purification buffer (1X PBS, 150 mM NaCl, 0.1% detergent) containing 15 mM imidazole, and 15 CVs supplemented with 30 mM imidazole. Target was then eluted in 5 CVs of purification buffer plus 250 mM imidazole. After addition of tobacco etch virus (TEV) protease, produced in-house, the protein was dialysed overnight (Specta-Por tubing, 3.5 kDa MWCO) in 3.5 L of gel filtration buffer (20 mM Tris pH 7.5, 150 mM NaCl, 0.03% DDM or 0.003% LMNG).

The next day, 0.22 μm syringe-filtered (Millipore) dialysate was passed through a 5 mL HisTrap column (Cytiva) to remove both the TEV protease and liberated GFP-His_8 affinity tag. The resulting pure protein was spin concentrated using a Vivaspin-20 (50 kDa MWCO, Sartorius) to a final volume of 500 μL, or a multiple thereof, prior to polishing on a Superdex 200 10/300 Increase GL column (Cytiva) pre-equilibrated in gel filtration buffer. If DM solubilised protein was sought, detergent exchange was achieved on-column at this stage. SEC fractions of interest were pooled, spin concentrated to the desired extent in a Vivaspin-2 or Vivaspin-500 (50 kDa MWCO, Sartorius), and the protein either snap-frozen as aliquots or used same day.

Purification of TEV-His_8 and FLAG-tagged membrane proteins proceeded similarly. The latter, in not requiring TEV cleavage, were purified in a single day using Anti-FLAG Affinity Resin (Pierce); protein being eluted with ≥5 × 1 CV 0.1 mg mL^{-1} 3XFLAG Peptide (Sigma-Aldrich) post-washing.

Purity was assessed throughout all purifications, of membrane proteins or otherwise, by analysis of samples loaded on self-cast SDS-PAGE gels (10-15% acrylamide) stained with Coomassie blue (InstantBlue, Expedeon). Where needed, in-gel GFP fluorescence was recorded pre-staining using an Fuji FLA Phosphor Imager, gel band protein identity was confirmed by mass spectrometry performed by Dr. Sabrina Liberatori of the Advanced Proteomics Facility (Dept. of Biochemistry, University of Oxford), and anion-exchange chromatography executed on a HiTrap Q HP column (Cytiva). Protein concentration was determined spectrophotometrically (NanoDrop ND-1000) using sequence derived extinction coefficients calculated with ExPASy ProtParam (Gasteiger et al., 2005).

2.2.5 Membrane Protein Expression in *M. smegmatis*

Targets cloned into pACE were miniprepped from *E. coli*, and then 1 μL of concentrated vector (>500 ng μL^{-1}) electroporated (2 mm gap, 2.5 kV) into 100 μL of electrocompetent mc^2155 *Msmg* cells (ATCC, Snapper et al. (1990)). These were left to recover for ~3.5 h at 37 °C in 900 μL of 7H9 OADC-supplemented media (BD Biosciences), prior to plating on 7H10 OADC-supplemented (Sigma-Aldrich) Apr plates. After three days at 37 °C, successfully transformed colonies were identified by colony-PCR and grown up for glycerol stock preparation.

For expression, said stocks were used to seed three night 7H9-OADC Apr starter cultures (37 °C, 180 RPM), featuring Tween-80 (0.05% v/v) as an anti-clumping agent, from which 10 mL expression cultures were subsequently inoculated. At an OD_{600} of ~0.6 cultures were induced with 0.05–0.4% (w/v) acetamide (Sigma-Aldrich) and left to grow overnight (37 °C, 180 RPM). Expression was confirmed by in-gel GFP fluorescence of SDS-PAGE separated crude lysate.

2.2.6 Sybody Expression & Purification

Sybodies cloned into pBXNPHM3, and thus possessing N-terminal His_{10}-MBP-3C tags, were expressed in *E. coli* MC1061 cells (Lucigen). Multilitre TB-Amp cultures

were grown until attaining an OD_{600} of ~0.5 (37 °C, 215 RPM), and then cooled for ~25 min prior to overnight induction with 0.02% (w/v) L-arabinose (25 °C, 215 RPM). The following morning cells were harvested (~9,000 g, 15 min) and pellets frozen, post-addition of DNAse I (Sigma-Aldrich) and resuspension in 20 mM Tris pH 7.5, 150 mM NaCl (5 mL g^{-1}).

On purification, thawed pellets were first broken by cell disruption at 25 kpsi (Constant Systems Ltd) and the resulting lysate centrifugally clarified (8,000 g, 20 min, 4 °C). The recovered supernatant, augmented with 15 mM imidazole, was incubated with HisPur-Ni-NTA beads (Thermo Fisher Scientific) for ~1.5 h under gentle agitation. The resin, once loaded onto a gravity flow column, was next washed with 15 CVs of resuspension buffer supplemented with 50 mM imidazole, before sybody elution in 5 CVs supplemented with 300 mM imidazole. After addition of human rhinovirus 3C protease, purified in-house, binders were dialysed overnight against 4 L of 20 mM Tris pH 7.5, 150 mM NaCl. Filtered dialysate, having had its imidazole concentration reset to 15 mM, was then pumped through a 5 mL HisTrap column (Cytiva). Tag-free sybody in the flow-through was spin concentrated using an Amicon Ultra (3 kDa MWCO, Milipore) and injected on a Sepax SRT-C SEC-300 column (Sepax Technologies) pre-equilibrated in 20 mM Tris pH 7.5, 150 mM NaCl. Pure fractions were pooled, concentrated, aliquoted, and snap-frozen. Small-scale batch purifications from pSb_init were undertaken using published protocols (Zimmermann et al., 2018, 2020).

2.2.7 Membrane Scaffold Protein Expression & Purification

N-terminally His-tagged, TEV cleavable, constructs of the membrane scaffold proteins (MSPs) 1D1ΔH5 (Addgene #71714), 1D1 (Addgene #20061), 1E3D1 (Addgene #20066) & 2N2 (Addgene #29520) were expressed in BL21-Gold *E. coli* cells (Invitrogen). On reaching an OD_{600} of ~0.6, 1 L TB-Kan cultures were induced with 0.4 mM IPTG and left to grow for an additional ~3.5 h (32 °C, 215

RPM) before harvesting (~9,000 g, 15 min). Cell pellets were resuspended in 1X PBS (5 mL g^{-1}), and solid DNAse I (Sigma-Aldrich) added, ahead of freezing.

For purification, thawed pellets were made up to 40 mL in 1X PBS, supplemented with 200 mM NaCl and 1% (v/v) Triton X-100, and then lysed by sonication. Once clarified (40,000 g, 30 min, 4 °C), lysate was stirred with HisPur-Ni-NTA resin (Thermo Fisher Scientific) for ≥1.5 h in the presence of 20 mM imidazole. After extensive washing, MSPs were recovered in 5 CVs of elution buffer (1X PBS, 200 mM NaCl, 400 mM imidazole) and subsequently dialysed overnight, with TEV protease, in 2 L of dialysis buffer (20 mM Tris pH 7.5, 150 mM NaCl). The next day, filtered dialysate underwent a second round of Ni-NTA bead capture, at 50 mM imidazole, to remove the protease and cleaved affinity tag. Pure MSPs were lastly spin concentrated, with concomitant buffer exchange to decrease imidazole content, using a Vivaspin-20 (10 kDa MWCO, Sartorius), aliquoted, and flash-frozen.

2.3 Membrane Protein Characterisation

2.3.1 Negative Stain Electron Microscopy

Purified protein was diluted in gel filtration buffer to 0.001–0.06 mg mL^{-1} and then 5–10 μL dispensed onto a freshly glow-discharged carbon-coated Cu grid (300 mesh, TAAB Laboratories Equipment Ltd). After 1 min incubation, grids were washed twice in 2 × 30 μL water droplets and then stained twice in 2 × 30 μL 2% (w/v) uranyl acetate, or uranyl formate, droplets. Grids were blotted before and after each step and were incubated for 30 s in the final stain pool. Once air-dried, samples were imaged on a Talos 200 kV Cryo-TEM (Thermo Fisher Scientific) housed within the Central Oxford Structural Molecular Imaging Centre (COSMIC). Some grids were prepared, and all imaged, by Dr. Justin Deme.

For nanodisc size analysis ~20 images were collected per sample and processed in the EMAN2 suite (Tang et al., 2007). Micrographs were swarm picked and the resultant ~850–1500 particles subject to reference-free 2D classification into 20 classes. From these, disc diameters were measured manually using a 1.43 Å pixel^{-1} conversion factor.

2.3.2 SEC-MALS

Size-exclusion chromatography with multi-angle light scattering (SEC-MALS) was carried out by Dr. David Staunton within the Molecular Biophysics Suite (Dept. of Biochemistry, University of Oxford). Protein at 1 mg mL^{-1} was injected onto a Superdex 200 10/300 Increase GL column (Cytiva) in-line with DAWN HELEOS-II Scattering and Optilab rEX Refractometry Detectors (Wyatt Technology); all pre-equilibrated in gel filtration buffer (20 mM Tris pH 7.5, 150 mM NaCl, 0.03% DDM). Data were analysed within ASTRA (Wyatt Technology).

2.3.3 Blue Native-PAGE

Native-PAGE was undertaken using precast 15-well 4–16% Bis-Tris gels (Invitrogen Novex) according to the manufacturer's instructions. Briefly, 3–4.5 μg of protein was loaded per lane and the gel run at 150 V for 1 h followed by 250 V for 1.5 h. Gels were then fixed for 15 min (100 mL 40% MeOH, 10% acetic acid v/v) and bands visualised by destaining for >1 h (100 mL 8% acetic acid).

2.3.4 Chemical Cross-Linking

For explorative cross-linking, purified protein was first dialysed overnight in a Slide-A-Lyzer MINI Dialysis Device (2 kDa MWCO, Thermo Fisher Scientific) against 100 mL 1X PBS, 0.02% DDM. Next, with Tris removed, the protein was split into 6–8 μg fractions and each incubated with one of a gradient of formaldehyde (0.005–1.5% v/v), glutaraldehyde (0.0005–0.1% v/v) or disuccinimidyl suberate (0.01–2.5 mM, Thermo Fisher Scientific) concentrations. After incuba-

tion at room temperature, for 20–30 min, the cross-linking reaction was quenched by addition of 1M Tris pH 8.0 and the results assessed by SDS-PAGE (8% acrylamide gel). $MmpL3_{1\text{-}753}$ glutaraldehyde cross-linking for Cryo-EM proceeded comparably, but was performed in LMNG and supplemented with a gel filtration step post-quenching.

2.3.5 Circular Dichroism Spectroscopy

Circular dichroism (CD) scans were performed on a J-815 Spectropolarimeter (Jasco) using 1 mm pathlength quartz cuvettes (Starna Scientific). For secondary structure assessment, concentrated protein was diluted to ~0.2 mg mL^{-1} in CD buffer (10 mM potassium phosphate pH 7.0, 50 mM sodium sulphate ± 3 × critical micelle concentration (CMC) detergent) and then absorption spectra collected over a wavelength range of 190–260 nm. To measure thermal stability the same sample was subject to successive scans over 195–240 nm, measured in 2 °C increments between 20–95 °C. Melt curves were generated by plotting CD signal at 209 nm against temperature; sigmoidally fit within QtiPlot to derive the T_m of thermal unfolding.

2.3.6 Haem Co-Elution Assay

Haem was prepared from hemin (Sigma-Aldrich) as described elsewhere (Owens et al., 2013). Approximately 100–120 μg of DM purified protein was incubated on ice for 2 h at haem-to-protein molar ratios spanning 0:1–2:1. These were then analysed on a Superdex 200 Increase 10/300 GL column (Cytiva), pre-equilibrated in gel filtration buffer (20 mM Tris pH 7.5, 150 mM NaCl, 0.3% DM) and in-line with RF-20A Fluorescence and SPD-20A UV-Vis Detectors (Shimadzu). Elution Trp-fluorescence (λ_{Ex}: 280 nm, λ_{Em}: 348 nm) as well as absorbance at 280 nm and 370 nm were monitoried; the latter acting as a read-out for haem. Binding of the porphyrin was identified by coincident A370 and A280 peaks, and was generally concomitant with Trp-fluorescence quenching relative to the

no haem control. For Cryo-EM, MmpL3$_{1\text{-}753}$-haem complexation in LMNG was achieved similarly.

2.4 Sybody Generation & Screening

2.4.1 Protein Biotinylation

Purified Avi-tagged proteins, at a final concentration of ≤40 μM, were biotinylated overnight at 4 °C using the BirA biotin-ligase reaction kit (Avidity) according to the manufacturer's guidelines. The following day, ligase and surplus biotin were removed by SEC on a Superdex 200 10/300 Increase GL column (Cytiva) pre-equilibrated in gel filtration buffer (20 mM Tris pH 7.5, 150 mM NaCl, 0.03% DDM or 0.003% LMNG). Biotinylation efficiency was then assessed by streptavidin shift assay. In short, a sample of biotinylated target was incubated with ~3 × molar excess streptavidin (Sigma-Aldrich), for ≥5 min, prior to analysis by SDS-PAGE; executed on ice at 120V to maintain biotin-streptavidin interactions in-gel.

2.4.2 Sybody Selection

Sybodies were selected against DDM purified, biotinylated, MmpL3$_{1\text{-}753}$ under the supervision of, and with assistance from, Dr. Iwan Zimmermann during a secondment to the University of Zürich. The *in vitro* selection cascade employed, graphically summarised in Figure 5.3, is described in detail — step-by-step — in Zimmermann et al. (2020). Later rescreening in LMNG, through repetition of just the enzyme-linked immunosorbent assay stage, was accomplished by largely identical means.

2.4.3 Sybody Pull-Downs

For binder evaluation by pull-down, 40 μg of detergent solubilised MmpL3$_{1\text{-}753}$ was incubated with a Myc-His$_6$-tagged sybody, in a 1:1.5 molar ratio, on ice for ≥30 min. The components were topped up to a final volume of 100 μL, with

buffer (20 mM Tris pH 7.5, 150 mM NaCl, 0.03% DDM or 0.003% LMNG), and supplemented with 15 mM imidazole. After taking a 7.5 μL "load" sample, the mixture was rotated with 15 μL of HisPur-Ni-NTA resin (Thermo Fisher Scientific) at 4 °C for 30 min. Beads were next washed twice with 400 μL of buffer featuring 30 mM imidazole; being spun down (500 g, 2 min) and the supernatant discarded between all steps. Bound protein was recovered in 30 μL buffer augmented with 300 mM imidazole, 7.5 μL of which was taken as an "elute" sample and ran adjacent to the corresponding "load" sample by SDS-PAGE (15% acrylamide gel). No sybody and off-target sybody controls were performed alongside. An increased stringency variant, using 300 or 600 mM NaCl buffers throughout, was also trialled. Nanodisc pull-downs were carried out equivalently, in detergent-free buffer (20 mM HEPES pH 7.5, 150 mM NaCl), but with all quantities scaled down by half to conserve protein.

2.4.4 Sybody Co-Elution

To identify binders capable of SEC-stable complexation, 40 μg aliquots of pure MmpL3$_{1\text{-}753}$ were first individually incubated with 74 different Myc-His$_6$-tagged sybodies in a 1:1.5 target-to-binder molar ratio; made up to a final volume of 100 μL with gel filtration buffer (20 mM Tris pH 7.5, 150 mM NaCl, 0.03% DDM). After ≥30 min on ice, free and bound sybody were separated from one another by injection onto a pre-equilibrated Sepax SRT-C SEC-300 column (Sepax Technologies). Initial trials employed a TSKgel G3000SWXL column (Tosoh Bioscience), but were discontinued due to adverse column interactions. Successful co-elution was determined by binder presence in MmpL3$_{1\text{-}753}$ peak fractions, analysed by SDS-PAGE (15% acrylamide gel) and stained using the ProteoSilver Plus Silver Stain Kit (Sigma-Aldrich). Nanodisc co-elution trials were achieved similarly, but used tagless sybodies and detergentless buffers (20 mM HEPES pH 7.5, 150 mM NaCl). For Cryo-EM and crystallisation, co-elution with tag-free sybodies was again as described — only differing by being scaled up, and by utilising a 1:2 MmpL3$_{1\text{-}753}$-to-sybody molar ratio pre-SEC .

2.4.5 Sybody Off-Rate Ranking

Sybodies were kinetically ranked by Bio-Layer Interferometry (BLI), performed on an Octet-RED384 instrument using Streptavidin (SA) Biosensors (ForteBio). Once equilibrated in buffer (20 mM Tris pH 7.5, 150 mM NaCl, 0.03% DDM), sensors were loaded by dipping into 50 μL of 0.1 mg mL^{-1} biotinylated MmpL3$_{1\text{-}753}$ for 300 s. After a 60 s buffer baseline, biosensors were next transferred into 50 μL of 500 nM Myc-His$_6$-tagged sybody and association recorded for 300 s. Dissociation was measured, again for 300 s, by returning sensors to buffer. Seven sybodies were processed in parallel per BLI experiment, with an eighth channel used as a reference tip (i.e. loaded with MmpL3$_{1\text{-}753}$ but unexposed to sybody). Data were analysed with Octet Systems Software (ForteBio); raw sensorgrams undergoing reference subtraction, y-axis alignment, inter-step correction, and Savitzky-Golay filtering prior to partial fitting for off-rate estimation.

2.5 Reconstitution into Detergent Alternatives

2.5.1 Liposome Reconstitution

DM solubilised MmpL3$_{1\text{-}753}$ was reconstituted into asolectin (soyabean phospholipid extract, Sigma-Aldrich) liposomes via the dilution method. Protein and extruded asolectin (Mini Extruder, Avanti Polar Lipids) were mixed at a 50:1 lipid-to-protein ratio (w/w) and incubated at room temperature for 30 min. After a further 15 min on ice, the mixture was diluted into 70 mL of 50 mM potassium phosphate buffer, pH 7.0, and ultracentrifuged (>200,000 g, 2 h, 4 °C). Finally, pelleted proteoliposomes were resuspended in the same buffer, to a final concentration of 0.5 μg mL^{-1} (assuming 100% reconstitution), and subject to two rounds of freeze-thaw before analysis by SDS-PAGE to gauge reconstitution efficiency.

2.5.2 Amphipol A8-35

The protocol employed to screen exchange of DDM solubilised MmpL3$_{1\text{-}753}$ into amphipol A8-35 (Generon) was adapted from elsewhere (Zoonens et al., 2014). Briefly, ~200 μg of pure protein was combined with A8-35 at amphipol-to-protein mass ratios spanning 0:1–10:1. After incubation on ice for 1–1.5 h, detergent was removed by rotation with ~1 mg of activated Bio-Beads (SM-2 Resin, Bio-Rad) for 2–2.5 h at 4 °C. Samples, cleared of polystyrene adsorbents, were then analysed by successive injection onto a Superdex 200 10/300 GL column (Cytiva) which had been pre-equilibrated in detergentless buffer (20 mM Tris pH 7.5, 150 mM NaCl). The 0:1 ratio served as a DDM removal control. For Cryo-EM the optimal amphipol-to-protein mass ratio, 3:1, was moderately scaled up.

2.5.3 MSP Nanodiscs

Empirical screening of target insertion into MSP nanodiscs (NDs) began with asolectin (Sigma-Aldrich) solubilisation in 3 × molar excess sodium cholate; realised by rotation at room temperature for 30 min. Next, in a final volume of 150 μL reconstitution buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 15 mM sodium cholate), a range of protein : MSP : lipid molar ratios were assembled and incubated on ice for ≥1.25 h post-mixing. Components were always added in a fixed order — buffer, solubilised asolectin, purified membrane protein, and lastly the MSP of choice. Final target concentration was 2 μM for PepTSt and 3.5 μM for MmpL3$_{1\text{-}753}$. Detergent removal was achieved in two consecutive Bio-Bead (SM-2 Resin, Bio-Rad) incubation steps, the first for 1.5 h and the second overnight. In each, sample was transferred to a fresh tube and activated beads, equivalent to ~30 μL, added prior to rotation at 4 °C. The following morning, Bio-Beads were extracted and the assemblies then ultracentrifuged (100,000 g, 30 min, 4 °C). Per reconstitution ratio, 130 μL of recovered supernatant was evaluated on a Superdex 200 Increase 10/300 GL column (Cytiva); pre-equilibrated in detergentless buffer (20 mM HEPES pH 7.5, 150 mM NaCl). Empty ND (e.g. 0:2:100) and

detergent removal (e.g. 1:2:0) controls were processed in parallel. Successful ND incorporation was confirmed by target and MSP copresence in peak SEC fractions analysed by SDS-PAGE. Bands were visualised using the ProteoSilver Plus Silver Stain Kit (Sigma-Aldrich) according to the manufacturer's instructions. Material for ND Cryo-EM, or characterisation, was generated by scaling up the best assembly ratios identified; generally by a factor of 20–30 fold.

2.6 Crystallisation Trials

2.6.1 Vapour Diffusion

Details of the MmpL3$_{1\text{-}753}$ *in surfo* crystallisation trials undertaken, both alone and in the presence of sybody, are compiled in Tables 4.1 & 5.2. All were setup in MRC 96-well sitting drop plates (SWISSCI), into which 70 μL of reservoir solution had been pre-dispensed, per well, from one of four Mem-series sparse matrix screens (MemGold, MemGold2, MemTrans & MemChannel, Molecular Dimensions Ltd; Newstead et al. (2008); Parker & Newstead (2012)). With aid of a Mosquito liquid handling robot (STP Labtech), 100 nL concentrated protein and 100 nL reservoir solution were mixed to form the crystallisation droplet. After sealing, plates were stored at 4 or 18 °C and automatically imaged, in both visible and Trp-fluorescence channels, within a Rigaku Minstrel HT UV hotel. Wells were monitored remotely via CrystalTrak.

2.6.2 Lipidic Cubic Phase

Specifics of the MmpL3$_{1\text{-}753}$ *in meso* crystallisation trials attempted, with or without binder complexation, are summarised in Tables 4.2 & 5.3. First, protein-laden mesophase was generated by homogenising concentrated MmpL3$_{1\text{-}753}$ with molten host lipid using a dual-syringe mixer (Art Robbins Instruments); employing defined lipid-to-protein ratios (w/w) — 60:40 for monoolein (Sigma-Aldrich), 50:50 for 7.8 monoacylglycerol (Avanti Polar Lipids). Using a Gryphon

LCP robot (Art Robbins Instruments), this was dispensed as 50 nL boli onto siliconised glass microscope slides (Thermo Fisher Scientific) which had been modified, with a perforated adhesive spacer (Saunders), to give a 96-well format. Next, each bolus was overlain with 800 nL of precipitant solution sourced from a commercial (MemMeso, Molecular Dimensions Ltd) or custom (Tokyo Set, Oxford Set) sparse matrix screen. Post-sealing, sandwich plates were kept at 4 or 20 °C and periodically inspected with a benchtop microscope equipped with a cross-polariser. Where desired, well Trp-fluorescence was imaged within a Rigaku Minstrel HT UV hotel. Crystalline species, to be shot with X-rays at the I24 microfocus beamline (Diamond Light Source, Harwell, UK), were harvested by Prof. Simon Newstead.

2.7 Cryo-EM Trials & Structure Determination

2.7.1 Grid Preparation, Imaging & Data Processing

Grid preparation and data processing were performed partially, and imaging almost entirely, by Dr. Justin Deme. Unless stated otherwise, micrographs were collected in counting mode on a Titan Krios 300 kV Cryo-TEM (Thermo Fisher Scientific) fitted with a GIF energy filter (Gatan) and a K2 Summit Detector (Gatan). A minority of samples were screened, without dataset collection, on a Talos Arctica 200 kV Cryo-TEM (Thermo Fisher Scientific) equipped with a Falcon 3 Direct Electron Detector (Thermo Fisher Scientific). All microscopes were part of the COSMIC facility.

Where possible, SEC fractions destined for Cryo-EM were immediately pooled, spin concentrated as required, and the protein subsequently kept on ice until grid vitrification the same day. In all cases, 4 μL of sample was dispensed onto the surface of a freshly glow-discharged holey carbon-coated grid (Quantifoil 300 mesh, Cu or Au R1.2/1.3, Agar Scientific). After allowing 10 s for adsorption, grids were blotted for 2–4.5 s (blot force -5 or -15, 100% humidity, 4 or 8 °C)

and then flash-frozen by plunging into liquid ethane using a Vitrobot Mark IV (Thermo Fisher Scientific).

In general, datasets were preprocessed in SIMPLE (Reboul et al., 2018) and then RELION (Scheres, 2012) used for all operations downstream of particle picking. Technical details for most prominent datasets — those of LMNG solubilised $MmpL3_{1\text{-}753}$ — are provided in the following two sections.

2.7.2 Original MmpL3 LMNG Dataset

From $MmpL3_{1\text{-}753}$ grids (2 mg mL^{-1}) frozen shortly after protein purification, 1,326 movies were collected at a pixel size of 0.87 Å. A subset of 500 of these were first semi-manually swarm picked in EMAN2 (Tang et al., 2007). The resulting 37,798 particles were used to generate 80 references against which the entire micrograph stack were then autopicked; undertaken with SIMPLE (28-02-2018 build, Reboul et al. (2018)) using 256 × 256-pixel boxes. This yielded ~90,000 particles, which were next subject to reference-free 2D classification trials in both RELION (23-01-2018 build, Zivanov et al. (2018)) & SIMPLE. Subsequent efforts at 3D classification, employing the ~49,000 particles in the best 2D classes, failed to reconstruct an informative volume. A graphical summary of the processing pipeline is presented in Figure 4.4.

2.7.3 Final MmpL3 LMNG Dataset

From grids of freshly purified $MmpL3_{1\text{-}753}$ (2 mg mL^{-1}), data were gathered in counting super-resolution mode on a 300 kV Titan Krios G3 Cryo-TEM (Thermo Fisher Scientific) equipped with a BioQuantum Imaging Filter (Gatan) and a K3 Direct Electron Detector (Gatan). Using a pixel size 0.832 Å, 15,020 movies spanning a defocus range of -3.0 to -0.5 μm were collected; each with a total dose of 58.2 $e^{-}/Å^2$, spread across 40 fractions, over an exposure time of 2.80 s.

Patched (15 × 10) motion correction, dose weighting, contrast transfer function (CTF) estimation, and particle autopicking were all implemented on the fly using SIMPLE 3.0 (Caesar et al., 2020). This yielded 7,764,838 particles, extracted in 256 × 256-pixel boxes, which were then subject to two rounds of reference-free 2D classification; the first on the fly within SIMPLE 3.0, using cluster2D-stream, and the second in cryoSPARC (v.2.15.0, Punjani et al. (2017)). Discarding poor-quality classes after each round left 1,564,980 particles for downstream processing, which was performed predominantly in cryoSPARC.

The particle subset was next input into multi-class (k = 4) *ab initio* model generation, without symmetry imposition (i.e. in C_1). This generated a single featureful volume, displaying the expected $MmpL3_{1\text{-}753}$ domain organisation, composed of 758,680 particles. Subsequently passing these through two rounds of non-uniform refinement (Punjani et al., 2020), interspersed by bayesian particle polishing (Zivanov et al., 2019) undertaken in RELION 3.1.0 (Zivanov et al., 2020) within 320 × 320-pixel boxes, gave a 3.4 Å map. The requisite STAR file for this polishing was generated using the csparc2star.py script from UCSF pyem (Asarnow et al., 2019). Further heterogeneous refinement of the volume against the previous four *ab initio* classes recovered 562,667 particles, whose non-uniform refinement returned a 3.3 Å reconstruction. A final round of reference-free 2D classification was employed to remove any remaining suboptimal particles, resulting in a stack retaining the best 414,082. Successive non-uniform, beamtilt, and local refinement of these yielded the 3.0 Å map of $MmpL3_{1\text{-}753}$ deposited in the EMDB under accession code EMD-12604. Throughout processing all refinement reference maps were lowpass filtered to 8 Å. Gold-standard Fourier shell correlations (FSC) using the 0.143 criterion, as well as local resolution estimations, were calculated in RELION 3.1.0. A schematic summarising the processing workflow is provided in Figure 7.1.

2.7.4 Truncated MmpL3 Model Building & Refinement

Atomic model building began by threading the *Mtb* MmpL3$_{1\text{-}753}$ sequence onto a previously reported crystal structure of the *Msmg* orthologue (PDB: 6OR2, Su et al. (2019)), using the Phyre2 web server (Kelley et al., 2015). The threaded structure was then docked into the 3.0 Å Cryo-EM volume in UCSF Chimera (v.1.13.1, Pettersen et al. (2004)), prior to further rigid body fitting in Coot (v.0.9.2-pre, Emsley et al. (2010)). Within the latter programme the model underwent manual real-space refinement (RSR), regions lacking unambiguous density were removed (residues 343-377, 753), and an LMNG molecule was built into the splayed species occupying the periplasmic chamber.

To generate the final model, detailed in Table 7.1 and deposited in the PDB under accession code 7NVH, the MmpL3$_{1\text{-}753}$ complex underwent further rounds of RSR in PHENIX (v.1.18.2, Liebschner et al. (2019)) with secondary structure, rotamer, ligand, and Ramachandran restraints applied, alongside additional minor adjustments in Coot. The globally sharpened map, output from cryoSPARC (Punjani et al., 2017) local refinement, was used for all rounds of RSR in both Coot and PHENIX. Model validity was assessed by MolProbity (Williams et al., 2018), executed within the latter suite. The density fit for modelled transmembrane sidechains, and the LMNG ligand, are shown in Supplementary Figure B.17.

2.8 Structural & Sequence Analyses

2.8.1 Structural Analysis & Visualisation

Cavities in the structures of MmpL3$_{1\text{-}753}$ and other RND lipophile transporters were calculated with Caver Analyst 2.0 (Jurcik et al., 2018) using default parameters. Conservation analysis was performed on the ConSurf server (Ashkenazy et al., 2016). All structural figures were composed in UCSF Chimera (v.1.12

& v.1.13.1, Pettersen et al. (2004)) or UCSF ChimeraX (v.1.1, Pettersen et al. (2021)), with exception of the minority of stylised renders produced using Illustrate (Goodsell et al., 2019). Where present, accompanying antitubercular skeletal formulas were prepared in ChemDraw Professional (v.17.0.0.206).

2.8.2 Sequence Analysis & Visualisation

Sequence alignments were generated using T-Coffee (Madeira et al., 2019), manually adjusted as required in Jalview (v2.10.5, Clamp et al. (2004)), and typeset with TEXshade (Beitz, 2000b). Sybody CDR radial phylograms were prepared with Evolview (He et al., 2016), and sequence logos with Logomaker (Tareen & Kinney, 2020). Prior to structure solution, transmembrane topology, disorder, and fold were predicted from the *Mtb* MmpL3 sequence using the MEMSAT3 (Jones, 2007), IUPred2A (Dosztányi, 2018), and Phyre2 (Kelley et al., 2015) web servers respectively. For regions of *Mtb* MmpL3 unresolved by Cryo-EM, primary sequence depictions were created with TEXtopo (Beitz, 2000a).

2.8.3 *Mtb* MmpL3 Mutation Collation

MmpL3 missense variants observed to confer resistance to one or more preclinical agents were extracted from 29 publications and then manually curated to give Supplementary Table B.2, which includes references to the original studies. Mutations identified in mycobacterial species other than *Mtb* were back-mapped by means of sequence alignments between the relevant MmpL3 orthologues.

The broader mutational landscape of *Mtb* MmpL3 was sampled with assistance from Dr. Philip Fowler. All *Mtb* genomes deposited in the ENA at the time of study, together with those reported in Walker et al. (2015) & The CRyPTIC Consortium & 100000 Genomes Project (2018), were aggregated. Totalling 45,878, the majority of these genomes underwent *Mtb* lineage classification using Mykrobe (Hunt et al., 2019) and short reads were mapped onto version 3 of the H37Rv *Mtb* reference genome (NC_000962.3); variants being called

using Clockwork (v.0.8.3) as described elsewhere (The CRyPTIC Consortium & Fowler, 2021). The CRyPTIC data warehouse was then interrogated for all non-synonymous substitutions in the MmpL3 ORF. The resultant 7,475 mutations, and any lineage associations of the most abundant, are compiled in Supplementary Tables B.3 & B.4.

Homologue Screening

Construct Design & Characterisation

CHAPTER 3

Obtaining high resolution membrane protein structures, be it by X-ray crystallography (XRC) or cryogenic electron microscopy (Cryo-EM), is well documented as being technically challenging. The process is regularly frustrated by bottlenecks at multiple steps — e.g. expression, purification, crystallogenesis, and/or vitrification. Therefore, to increase one's chances of success it is essential to identify target proteins exhibiting favourable characteristics during the preliminary stages of any structural biology project. These include: **(i)** moderate-to-high expression of folded protein in the chosen heterologous host, **(ii)** the capacity to be purified to near-homogeneity (≥90-95% pure) in sufficient yields (ideally ≥1 mg L^{-1} of growth media), and **(iii)** display both thermal and biological (i.e proteolytic degradation resistant) stability once purified. Whilst the relative importance of target properties varies with the specifics of a given structural method, all those mentioned above are, to a reasonable extent, desirable for either XRC or Cryo-EM. For example, the greater the yield of pure membrane protein the more rapidly, and thoroughly, crystallisation space or grid vitrification parameters can be explored. Meanwhile, targets with better thermal stability are likely to be more rigid; which is closely associated with both an increased competency for crystal contact formation as well as a higher conformational uniformity within the ice (Loll, 2014). As such properties cannot be reliably predicted from inspection of protein sequence alone, they must be determined experimentally.

3.1 The Screening Pipeline

Conventionally, a set of homologous proteins to the target of interest — here *Mtb* MmpL3 — is chosen and then systematically passed through a screening pipeline; akin to the one employed herein. The set should encompass ample sequence diversity to allow the best chance of optimising the aforementioned characteristics, but still retain enough sequence similarity to ensure the structural elucidation of any member is informative to the biological questions being asked.

Being of prokaryotic origin, expression of the MmpL3 homologue set was first trialled in *E. coli* in the form of C-terminal TEV-GFP-His_8 fusions; the codon-optimised genes having been cloned into vectors of the IPTG-inducible pWaldo-GFP series (Waldo et al., 1999; Drew et al., 2001). Use of a GFP-fusion permits facile verification of folded, membrane-localised, target expression by whole-cell fluorescence measurements (Bird et al., 2016). The fluorescent reporter also allows for meaningful expression comparisons between constructs, protein tracking throughout subsequent purification steps (e.g. by in-gel fluorescence) and, notably, the use of fluorescence-detection size-exclusion chromatography (FSEC) to perform medium-throughput detergent screening (Kawate & Gouaux, 2006).

3.1.1 Detergent Choice & FSECs

Detergents, formed of a hydrophilic headgroup and hydrophobic tail, are the primary amphipathic surfactants used to extract membrane proteins from the lipid bilayer prior to downstream manipulation. Not only must the chosen detergent efficiently solubilise the target protein into aqueous solution, it must do so in a way that the lost attributes of the bilayer (e.g. lateral pressure, shielding of the TMD, protein-lipid interactions) are adequately compensated for within the newly formed protein-detergent complex (PDC); and thus the membrane protein remains correctly folded (Anandan & Vrielink, 2016). Detergent choice

is therefore of paramount importance. The suitability of a given detergent to produce stably reconstituted target protein is dependent on both detergent physicochemical and target intrinsic properties, necessitating empirical investigation on a per-protein basis. To this end, a small amount of GFP-fusion expressing *E .coli* crude membrane is solubilised in a panel of candidate detergents (at 1% w/v). Following centrifugal clarification, the supernatants are successively loaded onto a SEC column with an in-line fluorescence detector set to monitor GFP (λ_{Ex}: 488 nm, λ_{Em}: 512 nm). Monodispersity of resultant FSEC profiles is indicative of target stability in the solubilising detergent. Comparison of fluorescence traces can thus be used to guide purification detergent choice, and information concerning PDC size may be additionally inferred from the FSEC elution volume (**Kawate & Gouaux, 2006**).

Detergents are broadly classified on headgroup charge (ionic, zwitterionic or non-ionic) and chemical structure (**Seddon et al., 2004**). Non-ionic detergents featuring larger headgroups and longer alkyl tails (~10–12 carbons) are said to be "mild" — that is least denaturing. The most popular of these remains the alkyl maltosides DDM and DM which collectively account for ~50% of reported membrane protein crystal structures (**Parker & Newstead, 2016**). However, whilst "harsher" and more destabilising, the smaller micelles of shorter chained detergents, such as LDAO, are traditionally favoured in crystallography. Generally allowing increased protein surface exposure, they promote improved crystal contacts and therefore higher resolution diffraction (**Privé, 2007; Sonoda et al., 2011**). Given the large extramembranous loops of MmpL3, this may hold less precedence here. Nevertheless, the FSEC detergent panel employed (summarised in Table 3.1) includes zwitterionic and non-ionic detergents of variable chain length, as well as representatives of the more lipid-like maltose-neopentyl glycols (MNGs) which are increasingly finding use in Cryo-EM (**Chae et al., 2010**). If one or more viable detergents is identified by FSEC, the next step is to trial purification.

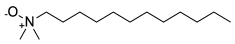
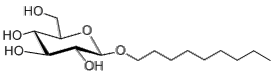
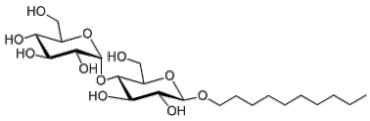
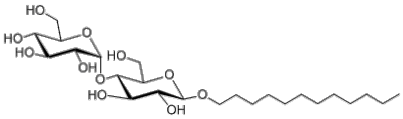
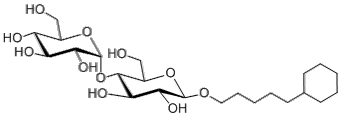
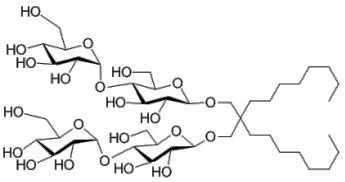
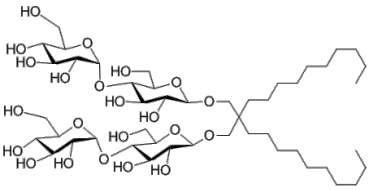
Name	Structure	Class	CMC (w/v)	Micelle Size
LDAO		Zwitterionic Amine Oxide	0.023%	21.5 kDa
NG		Non-ionic Glucoside	0.2%	90 kDa
DM		Non-ionic Maltoside	0.087%	40 kDa
DDM		Non-ionic Maltoside	0.0087%	72 kDa
CYMAL-5		Non-ionic Maltoside	0.12%	23 kDa
DMNG		Non-ionic Maltose-Neopentyl Glycol	0.0034%	-
LMNG		Non-ionic Maltose-Neopentyl Glycol	0.001%	91 kDa

Table 3.1: A Summary of FSEC Screening Detergent Properties

The classification, chemical structures and noteworthy properties of the seven FSEC screening detergents used throughout the project. See the Acronyms list for full nomenclature. Adapted from Stetsenko & Guskov (2017).

3.1.2 Purification, Characterisation & Iteration

Homologues proceeding to this stage are subject to an established purification protocol (see Drew et al. (2005)), that is used throughout the thesis. Briefly, post-solubilisation in the detergent of choice, TEV-GFP-His_8 fusions are purified by nickel immobilised metal affinity chromatography (IMAC) followed by site-specific cleavage with TEV protease during overnight dialysis. Native, untagged, target is then isolated from remaining undigested fusion, the liberated affinity tag, His_{10}-tagged TEV protease, and any other adventitious contaminants by a second "reverse" IMAC. Finally the protein is concentrated, prior to gel filtration on an S200 column. Purification progress is easily monitored by a combination of SDS-PAGE and GFP fluorescence measurements; the latter being especially

useful for estimating nickel resin binding efficiency. It is also informative to run a sample of the first IMAC elution down the gel filtration column, as undesirable fusion behaviour at this point (e.g. aggregation or major polydispersity) is often indicative of a poor outlook post-cleavage.

If amenable to robust purification, homologues advance to the last stage of the screening pipeline — "Characterisation". Here purified protein is subject to an ensemble of techniques aimed at probing aspects of fold, thermal stability, homogeneity, and oligomeric state. These include circular dichroism (CD) spectroscopy, negative stain electron microscopy (neg-stain), cross-linking, and size-exclusion chromatography with multi-angle light scattering (SEC-MALS). Only once a homologue displays satisfactory results for the entire pipeline is it worth investing significant efforts in structural attempts. Of course, and as will be demonstrated, screening is an iterative, non-linear, process; one may have to repeatedly trial new homologues, change vector or expression host, switch affinity tags, and modify construct design, before a viable combination is discovered. Keeping the above in mind, the MmpL3 homologue screening results are presented.

3.2 The Initial Homologue Set

Alongside *Mtb* MmpL3, three additional targets were chosen for the first round of screening — the HAE2 family RND proteins corynebacterial membrane protein large (CmpL)-1 and CmpL3, as well as the major facillitator superfamily (MFS) transporter NCgl-1194. All are from *Cgl* and were chosen primarily as they are shorter, and therefore potentially more experimentally tractable, than their mycobacterial counterparts.

CmpL1 is a clear MmpL3 orthologue (43% identity, 61% similarity), albeit without any equivalently substantive cytoplasmic CTD (Sequence Alignment Figure A.1). It is, however, documented to perform an analogous function in cell envelope biogenesis, having been linked to the export of trehalose monocoryno-

mycolates (Varela et al., 2012). Further to this, Belardinelli et al. (2016) had reported CmpL1 to be trimeric by neg-stain and it thus appeared an attractive target for Cryo-EM.

CmpL3 is another of the four *Cgl* MmpL homologues, most closely related to the *Mtb* SL transporter MmpL8 (37% identity, 58% similarity; Sequence Alignment Figure A.2). Genetically located within a carotenoid biosynthesis cluster (Krubasik et al., 2001), it too has been associated with cell wall construction (Yang et al., 2014), but otherwise remains largely uncharacterised. Although less functionally comparable to MmpL3, its relation to *Mtb* MmpLs more broadly still made CmpL3 a worthy target choice.

Finally NCgl-1194, whilst not an MmpL homologue, relates to the project via the theme of antimicrobial resistance. It is a *Cgl* homologue of *Mtb* Rv1250 (45% identity, 63% similarity; Sequence Alignment Figure A.3). A putative MFS drug-efflux pump, Rv1250 has been found over expressed in some Isoniazid resistant isolates (Li et al., 2015; Narang et al., 2017; Umar et al., 2019), as well as transcriptionally up-regulated in wild-type strains on exposure to subinhibitory doses of the first-line antitubercular (Garima et al., 2015).

3.2.1 Expression Tests & FSECs

All four targets were successfully expressed as TEV-GFP-His_8 fusions in Terrific Broth (TB), with whole-cell GFP counts ranging from ~16,000 (MmpL3) to ~52,000 (CmpL3) relative fluorescence units (RFU) per mL of culture (see Table 3.2). As a rule of thumb, counts above a threshold of ~10,000 RFU mL^{-1} are considered sufficient for further study. MmpL3 may have expressed least well as it is the largest by over 170 residues. More notably, NCgl-1194 expression appeared to retard *E. coli* growth; with harvested pellets always being both smaller and of altered colour (reddish-brown, not green; Supplementary Figure B.1[B]) relative to the other constructs. This likely indicates NCgl-1194 production is toxic to

Protein	Organism	Gene	UniProt	Residues	Mass (Da)	GFP Counts (RFU mL^{-1})
MmpL3	*M. tuberculosis*	*Rv0206c*	P9WJV5	944	100,904	16 – 18K
CmpL1	*C. glutamicum*	*Cgl2867*	Q8NLS1	772	83,056	38 – 41K
CmpL3	*C. glutamicum*	*Cgl0625*	Q8NSP3	730	77,330	48 – 52K
NCgl-1194	*C. glutamicum*	*Cgl1242*	Q8NR22	487	51,484	35 – 48K

Table 3.2: Initial Homologue Set Details & Expression Counts

Salient information on MmpL3, CmpL1, CmpL3, and NCgl-1194 curated from UniProt (Bateman, 2019), alongside the overnight TB expression range observed for each pWaldo-TEV-GFP-His_8 construct in *E. coli* C43(DE3) cells.

E. coli, possibly due to deleterious transport activity. Despite these caveats, the four proteins expressed more than adequately for FSEC screening.

The target set could be effectively solubilised in all seven FSEC detergents trialled (Table 3.1). Solubilisation efficiency was calculated as GFP count recovery after detergent incubation and ultracentrifugation. It was 90–100% for every protein-detergent pairing excluding MmpL3 and CmpL3 in NG, where solubilisation was ~75%. With exception of the two harshest detergents — LDAO and NG — monodisperse traces, indicative of stable reconstitution, were recorded for all FSECs (Figure 3.1). To control for the wait time between first and last sample injections, each batch of FSECs was begun and ended with a run in DDM; denoted DDM_{Start} and DDM_{Last} respectively. A reduction in peak height and/or monodispersity between these samples suggests a depreciation in protein quality on the experimental timescale (~8 hours), and was noted for both MmpL3 (Figure 3.1[A]) and CmpL3 (Figure 3.1[C]). The former also displayed a significant population of GFP-tagged degradation products, congruent with proteolytic attack of its CTD which is predicted to be unstructured. Only NCgl-1194 gave a favourable, yet small, peak in NG (Figure 3.1[D]); possibly hinting at a higher intrinsic stability relative to the other proteins. Lastly, all targets eluted earliest in DM, CYMAL-5, and LDAO. This is counter-intuitive as their short alkyl chains and/or diminutive headgroups would be expected to solubilise the constructs in some of the smallest micelles of the detergent set (Oliver et al., 2013). As

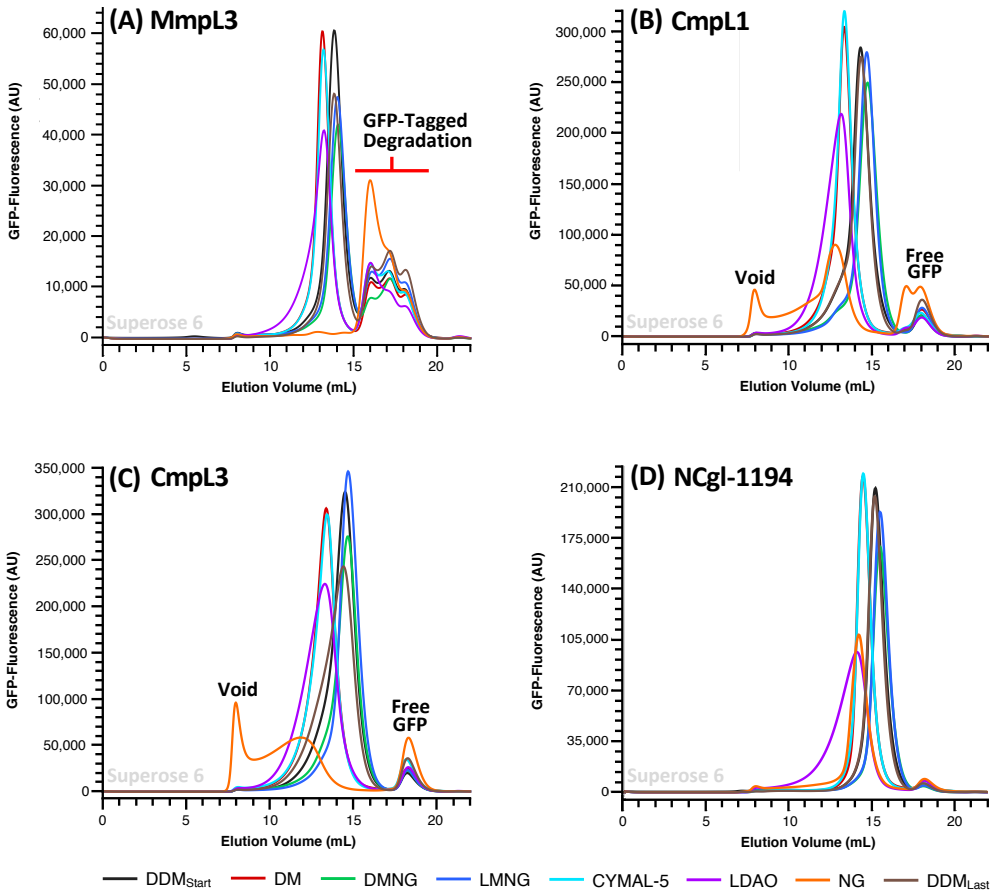


Figure 3.1: FSEC Traces of MmpL3, CmpL1, CmpL3 & NCgl-1194

FSEC traces of 1% (w/v) detergent solubilised *E. coli* membranes expressing **(A)** MmpL3–, **(B)** CmpL1–, **(C)** CmpL3–, and **(D)** NCgl-1194-TEV-GFP-His_8 fusions injected onto a Superose 6 column equilibrated in DDM. Void volume is ~8 mL and free GFP elutes at ~18 mL; both are labelled where prominent. Optimisation of detector sensitivity for construct expression level means GFP-Fluorescence (arbitrary units, AU) is only internally comparable per plot.

the phenomenon is seen for biologically unrelated transporters (i.e. RND & MFS members) it is likely artefactual. Speculatively, it may be a consequence of mixed micelle formation following partial exchange between DDM, in which the Superose 6 column is equilibrated, and the solublising detergent under test. Using a mild detergent in the mobile phase is unable to rescue destabilisation during solubilisation (Kawate & Gouaux, 2006), making it convenient to keep this constant for all samples. Having produced satisfactory FSEC traces in multiple detergents, the targets proceeded to initial purification and characterisation in DDM.

3.2.2 Purifications & Characterisation

MmpL3

Purification of *Mtb* MmpL3 in DDM proved difficult. Whilst detergent solubilisation remained comparable to the FSEC stage, binding efficiency to the nickel resin was poor, with just ~10–30% of solubilised protein ever bound over multiple purification attempts. Of the protein that could be recovered by IMAC, a significant proportion consisted of adventitious contaminants which were presumably pulled down by MmpL3. This was evident on subjecting the fusion to gel filtration (Figure 3.2[A]); which gave a bimodal profile, where only the first peak was robustly GFP fluorescent and also lead from the void — indicative of aggregation. The associated in-gel identified both higher order (i.e. SDS-resistant) and degredant GFP-tagged species (labelled in Figure 3.2[B]), the latter consistent with the FSECs. Bulk MmpL3-TEV-GFP-His_8 was recalcitrant to cleavage. The tag could be partially removed by adding TEV protease to peak SEC material generated from rerunning fusion fraction "A6" down the S200 column for a second time (Figure 3.2[A] inset). Such requirements suggest the presence of cleavage inhibitory contaminants or, more likely, that only a subpopulation of the fusion exists in a cleavable state. Once tag-free, the reverse IMAC gave pure (Figure 3.2[C]), SEC monodisperse (Figure 3.2[D]), MmpL3 whose identity was confirmed using mass spectrometry (gel band tryptic digest). However, the circuitous purification yielded little protein and that which it did appeared monomeric and aggregative by neg-stain (Figure 3.2[E]). For this, sample is embedded in a contrast enhancing heavy metal and then imaged by transmission electron microscopy (TEM); allowing rapid assessment of size and homogeneity (Ohi et al., 2004). The undesirable properties observed throughout purification — poor binding, contaminant co-elution, degradation, cleavage resistance, and aggregation — can be logically attributed, with the benefit of hindsight, to the disordered CTD of MmpL3; implying that its removal may be beneficial (see Section 3.4).

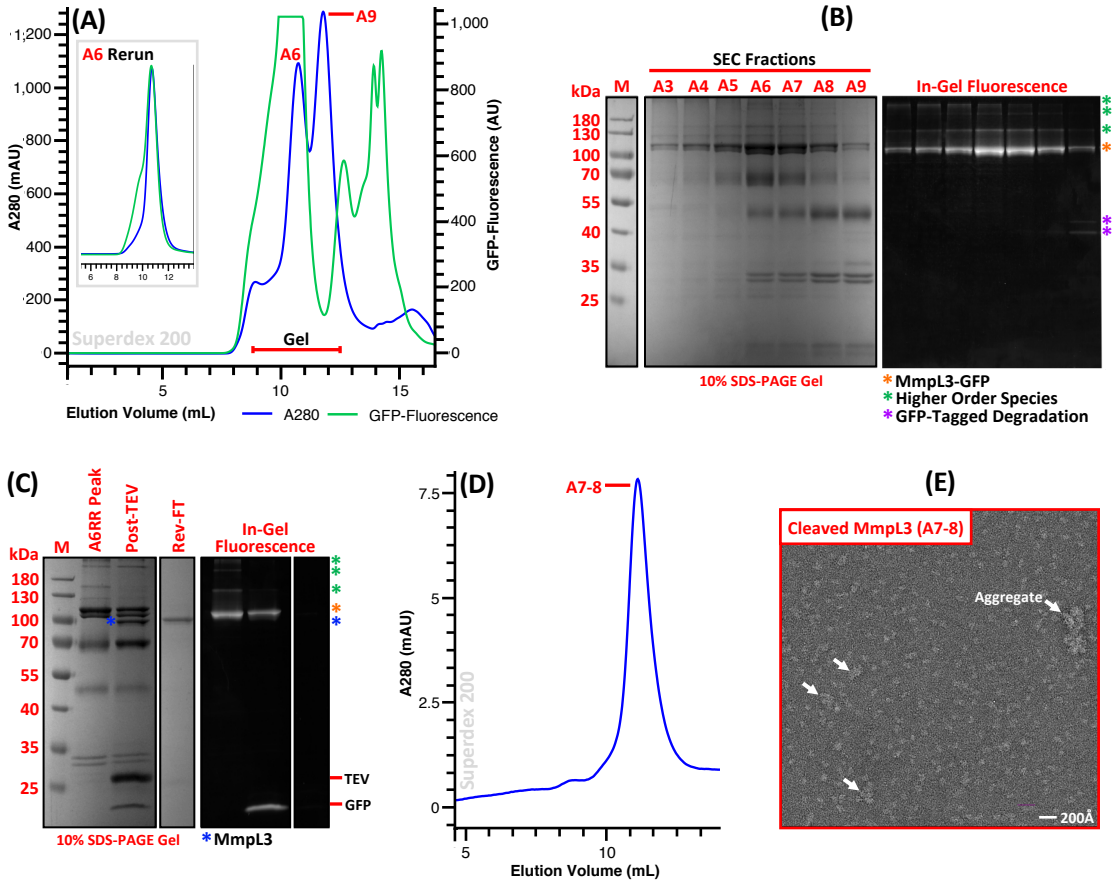


Figure 3.2: Purification of MmpL3-TEV-GFP-His$_8$ in DDM

(A) Absorbance 280 nm (A280, mAU) S200 profile of concentrated MmpL3-TEV-His-GFP$_8$ IMAC elution with in-line GFP fluorescence readings overlain. Peak fractions are highlighted. Inset: S200 trace of fraction A6 rerun. Red bar corresponds to fractions visualised on the SDS-PAGE gel of **(B)**, which is presented alongside the corresponding GFP in-gel. **(C)** SDS-PAGE gel, and in-gel, confirming partial TEV protease cleavage of fusion fraction "A6" rerun (A6RRPeak), as well as its purity after reverse IMAC (Rev-FT). **(D)** S200 profile of tag-free MmpL3. **(E)** Representative micrograph of cleaved MmpL3 negatively stained with uranyl formate; courtesy of Justin Deme. White arrows locate protein aggregates.

The monomerism of MmpL3 is at odds with the assertions of Belardinelli et al. (2016). They reported that *Cgl* expressed, strep-tag purified, *Mtb* MmpL3 was, like the majority of bacterial RND transporters studied to date, trimeric (Nikaido, 2018). This was concluded on the basis of both cross-linking and extrapolation from their homotrimeric CmpL1 neg-stain envelope. In the event that the C-terminal GFP of the fusion somehow precluded correct trimerisation, a C-terminal MmpL3-FLAG construct was cloned — again into pWaldo-GFP, but with a stop codon immediately after the newly added octapeptide (DYKDDDDK) tag. Compared to the original construct, MmpL3-FLAG could be purified in DDM without

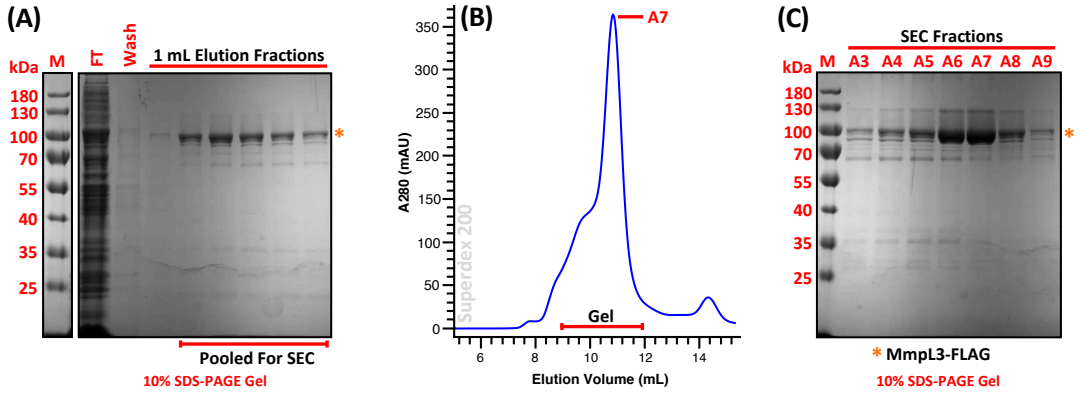


Figure 3.3: Purification of MmpL3-FLAG in DDM

(A) SDS-PAGE gel of MmpL3-FLAG purification. Anti-FLAG resin elution was with 0.1 mg mL^{-1} 3X FLAG peptide in 1 mL steps. Red bar indicates fractions pooled and concentrated for the S200 SEC run shown in **(B)**, fractions from which are analysed on the SDS-PAGE gel presented in **(C)**.

hassle (Figure 3.3); benefiting from the absence of a cleavage step, which afforded single day purification. Significantly more material could be recovered per litre of TB and, although shouldered, the SEC peak was near monodisperse (Figure 3.3[B]). Unfortunately, by SDS-PAGE MmpL3-FLAG ran as numerous "laddered" bands (Figures 3.3[A]&[C]); once again pointing to degradation. Nevertheless, the construct highlights the improvements that can be gained from simply changing affinity tag.

Inspecting the S200 elution volumes of cleaved MmpL3 and MmpL3-FLAG implies they are of similar size. To probe MmpL3-FLAG oligomeric state both blue native-PAGE (BN-PAGE) and cross-linking were attempted (Supplementary Figure B.2). The former, a non-denaturing electrophoretic technique aimed at maintaining quaternary structure interactions in-gel (Reisinger & Eichacker, 2008), showed MmpL3-FLAG to migrate as at least four species of ever increasing molecular weight (MW; Supplementary Figure B.2[A]). While semi-consistent with oligomerisation, it is at least equally congruent with aggregative self-association given that some of the species are in excess 700 kDa (assuming no gel artefacts). Cross-linking of the MmpL3-FLAG peak was performed with concentration gradients of the homobifunctional, primary amine-reactive, reagents glutaraldehyde

(5 Å spacer arm) and disuccinimidyl suberate (DSS, 11.4 Å spacer arm). Results were visualised by SDS-PAGE gel shift (Supplementary Figure B.2[C]). Glutaraldehyde showed little effect at most of the concentrations tested. DSS, meanwhile, cross-linked into a higher MW species (>180 kDa) only faintly present in the negative control and somewhat similar to that observed by Belardinelli et al. (2016). However, given the multitude of bands in the original sample it is difficult to interpret the results as evidence of *bona fide* trimerisation without additional study. As the screening pipeline would go onto identify more favourable MmpL3 constructs, this was never investigated further.

CmpL1

Up until cleavage, CmpL1 purified well in DDM; regularly binding to the nickel resin with ~40–55% efficiency and eluting cleanly (Figure 3.4[A]). Overnight dialysis with TEV protease unexpectedly, and repeatablely, yielded three post-cleavage products. These were found to co-elute within a broad, and thus heterogeneous, SEC peak accompanied by a prominent void (Figures 3.4[B]&[C]). Lacking endogenous canonical recognition motifs (ENLYFQ↓G/S), it initially seemed improbable that the added TEV protease was cleaving CmpL1 outside the tag-linker; but not impossible. Degenerate, so called "cryptic", TEV protease sites, with reduced cleavage efficiency, have been reported in comparable situations (e.g. White et al. (2004)). The largest product migrated at a MW consistent with full-length untagged CmpL1 (~83 kDa). The combined MWs of the two smaller products suggested they could be fragments arising from a single CmpL1 cleavage event, perpetrated by an unidentified protease. Supporting this, product ratio varied between purification attempts; the upper SDS-PAGE species being more abundant when the lower two were less abundant, and vice versa. The amount of the smaller products also seemed to mirror each other. One explanation was that GFP removal exposed an otherwise concealed protease site. However, similar pre- and post-cleavage behaviour of a straight CmpL1-TEV-His_8 construct (i.e. lacking GFP) argued against this.

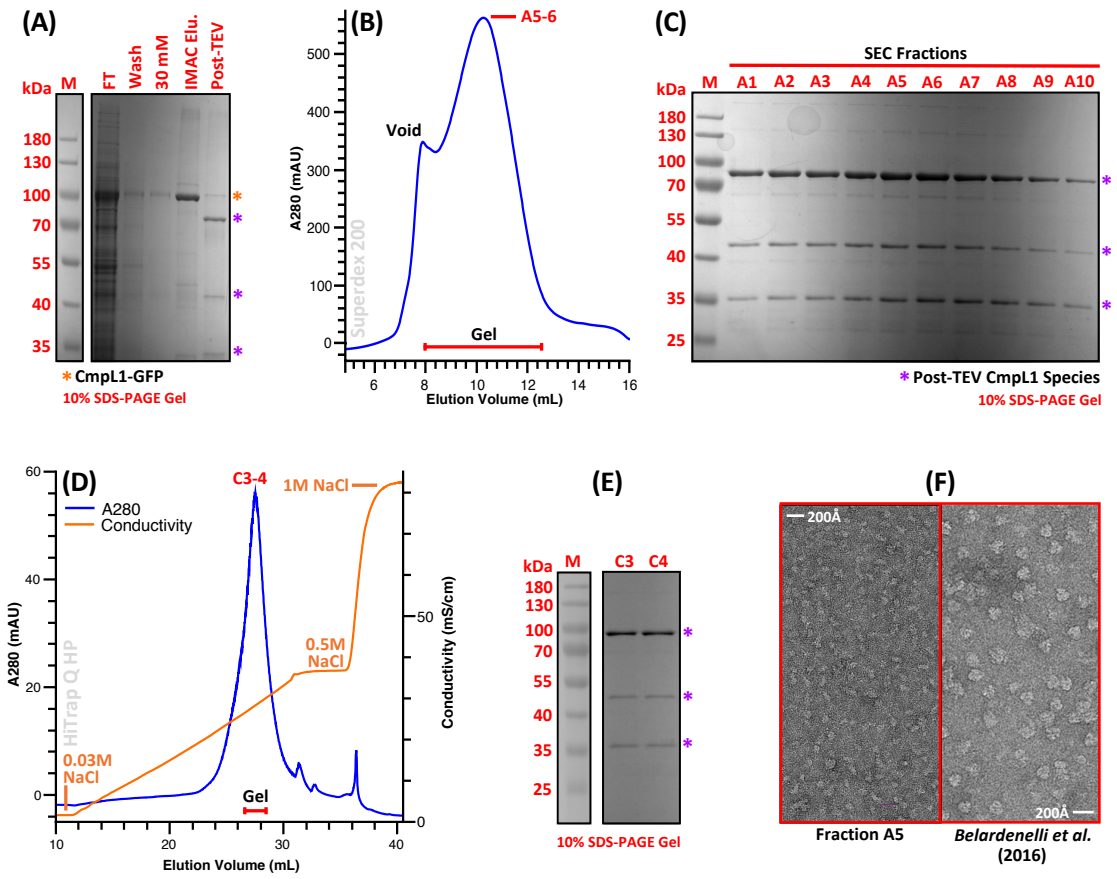


Figure 3.4: Purification of CmpL1-TEV-GFP-His$_8$ in DDM

(A) SDS-PAGE gel of CmpL1-TEV-GFP-His$_8$ purification, from IMAC flow-through (FT) to after TEV protease cleavage (Post-TEV). Notable species are marked with asterisks. **(B)** S200 SEC profile of concentrated CmpL1. Red bar corresponds to fractions visualised on the SDS-PAGE gel of **(C)**. **(D)** HiTrap Q HP anion-exchange chromatography (AEC) trace of CmpL1 overlain with the NaCl concentration (conductivity) changes used for elution. **(E)** SDS-PAGE gel of AEC peak fractions. **(F)** Representative micrograph of CmpL1 negatively stained with uranyl acetate (Left, courtesy of Justin Deme), compared to a CmpL1 micrograph collected by Belardinelli et al. (2016) (Right).

Scrutinising the CmpL1 sequence further, a non-transmembrane cryptic TEV site — broadly within the recognition sequence tolerance limits documented for the protease (Waugh, 2011) — can be reasonably proposed (IDIYVQ↓A, annotated in Sequence Alignment Figure A.1). Cleavage here would produce species of similar sizes (~35 & 48 kDa) to the observed smaller products, with divergence from the canonical motif ensuring incomplete proteolysis and hence the presence of full-length CmpL1. If correct, the hypothesis also explains the inability of anion-exchange chromatography (AEC) to separate the three species (Figures 3.4[D]&[E]); cryptic cleavage products presumably remaining associated

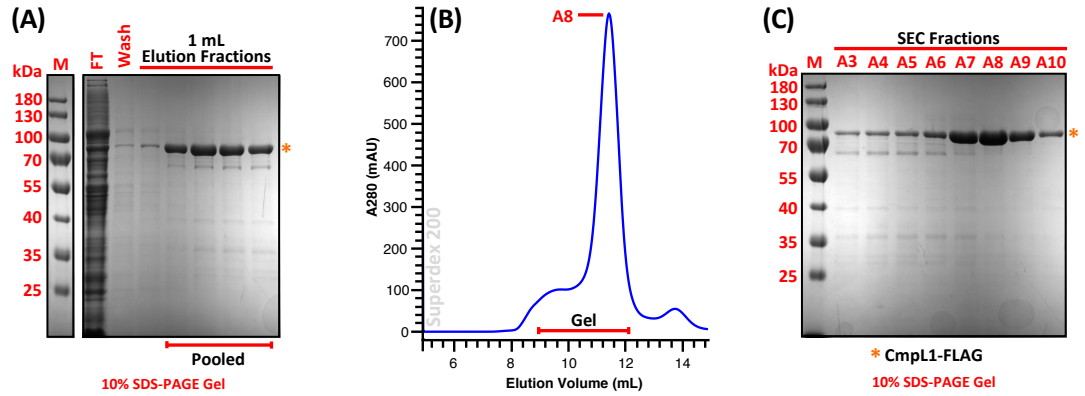


Figure 3.5: Purification of CmpL1-FLAG in DDM

(A) SDS-PAGE gel of CmpL1-FLAG purification with anti-FLAG resin. Red bar indicates elutions pooled and concentrated for the S200 SEC run shown in **(B)**, fractions from which are analysed on the SDS-PAGE gel presented in **(C)**.

within the same micelle. Mass spectrometry confirmed the presence of CmpL1 in all three SDS-PAGE bands, but did not readily validate the proposed cleavage model. Cross-linking was equally uninformative. Akin to MmpL3, neg-stain of CmpL1 failed to agree with the trimerism seen by Belardinelli et al. (2016) (Figure 3.4[F]). It cannot be ruled out that expression in *E. coli* has resulted in the absence of oligomerisation required factors (e.g. binding partners, specific lipids, substrates etc.) that are furnished in the native organism (i.e. *Cgl*). Switching to a FLAG tag had proved fruitful for MmpL3, so a C-terminal CmpL1-FLAG construct was next pursued.

Pleasingly, DDM purification of CmpL1-FLAG showed significant improvements. Near-pure protein was obtained following anti-FLAG resin elution (Figure 3.5[A]) and, post-concentration, yielded a monodisperse S200 profile (Figure 3.5[B]). With no need for TEV protease, the postulated cryptic site ceased to be of issue. Identically to MmpL3-FLAG, purified CmpL1-FLAG was subject to oligomeric state interrogation by BN-PAGE and cross-linking. The former showed a comparable number of species to MmpL3, thus not precluding multimerisation (Supplementary Figure B.2[A]). Cross-linking with both glutarladehdye and DSS produced multiple higher MW SDS-PAGE bands (Supplementary Figure B.2[B]) which may

represent true oligomers; the caveat being that aggregates, unable to enter the gel, were also formed. Despite potential, neg-stain of cross-linked CmpL1-FLAG did not convincingly recapitulate the findings of Belardinelli et al. (2016) that, on first glance, appeared so promising for Cryo-EM. Overall, CmpL1-FLAG would have been a more desirable construct had it been amenable to purification of the quality presented here more consistently. Not infrequently, a significant proportion of the protein would elute from the S200 column in a shoulder leading from the void. Unable to rectify such fickleness prior to identification of better targets, the protein was rarely revisited.

CmpL3

DDM solubilised CmpL3 bound well during the first IMAC step, boasting ~45–60% efficiency, and eluted with acceptable purity. Disconcertingly, evaluation of the GFP-fusion on the S200 column gave a multimodal, albeit entirely fluorescent, profile leading from the void (Figure 3.6[A]). Visualising peak contents by SDS-PAGE, accompanied by in-gel fluorescence, identified both higher order and degredant GFP-tagged species pervading the earlier fractions (Figure 3.6[B]). Unlike MmpL3 and CmpL1, untagged CmpL3 could be successfully recovered via overnight cleavage and subsequent reverse IMAC. Liberated CmpL3 retained the self-associative behaviour of the fusion however, giving a broad multi-peaked trace — the largest of which was in the void — when subject to gel filtration (Figure 3.6[C]). The protein populating these peaks was pure, with the earlier fractions once again displaying SDS-resistant higher MW species (Figure 3.6[D]). To assess the quality and oligomeric state of CmpL3 throughout the trace, a representative fraction from each non-void peak was analysed by neg-stain (Figure 3.6[E]). This showed only the latest eluting peak (fraction "A8") contained homogeneous, folded, and likely monomeric material. Earlier peaks were increasingly heterogeneous, featuring abundant aggregates and demonstrative of a protein inadequately stabilised by the purification approach employed. Owing to such unruly polydispersity, further work on CmpL3 was halted.

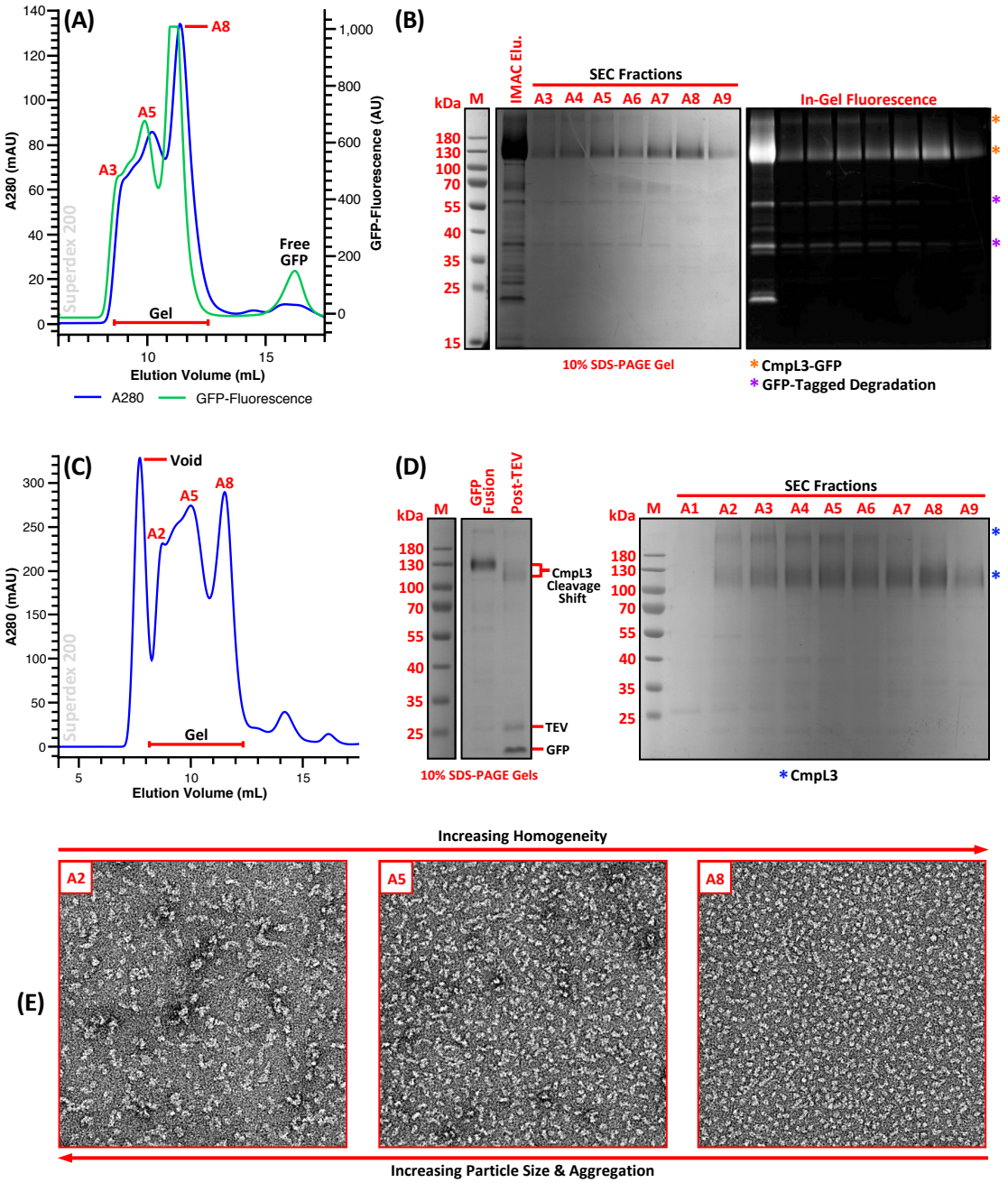


Figure 3.6: Purification of CmpL3-TEV-GFP-His$_8$ in DDM

(A) S200 profile of CmpL3-TEV-GFP-His$_8$ immediately following IMAC elution. In-line GFP fluorescence readings are overlain. Key peak fractions are labelled and the red bar corresponds to material visualised on the SDS-PAGE gel of **(B)**; presented alongside the corresponding GFP in-gel. Notable species are marked with asterisks. **(C)** S200 trace of concentrated CmpL3 post-cleavage. Fractions spanning the red bar are loaded on the SDS-PAGE gel of **(D)**. **(E)** Representative micrographs of cleaved CmpL3 negatively stained with uranyl formate, annotated with the peak fraction from which the protein was obtained. Images courtesy of Justin Deme.

NCgl-1194

Of the initial targets, least progress was made with NCgl-1194. As alluded, NCgl-1194-TEV-GFP-His_8 expression appeared toxic to *E. coli* and this was equally true of a NCgl-1194-TEV-His_8 construct also trialled. Expression of the latter could be confirmed by western blot (Supplementary Figure B.1[A]) and was likewise accompanied by bacterial pellet size reduction and colour change (as in Supplementary Figure B.1[B]). A significant amount of the GFP-fusion was lost during membrane preparation, probably indicating that only fraction of the protein is ever successfully inserted into the bilayer. Whilst the remaining NCgl-1194-TEV-GFP-His_8 solubilised well in DDM, as at FSEC, and showed workable IMAC binding efficiency (~50%), the protein recovered following nickel resin elution was both meagre and recalcitrant to TEV protease cleavage (Supplementary Figure B.1[C]). On this basis, pursuit of NCgl-1194 was largely abandoned.

To summarise, having progressed through the screening pipeline up until DDM test purification, constructs of MmpL3, CmpL1, CmpL3, and NCgl-1194 were all found to behave unsatisfactorily in one way or another; be it issues with aggregation, degradation, and/or affinity tag removal. That being said, clear progress had been made for some targets — especially MmpL3 and CmpL1 with respect to their FLAG-tagged variants. Prior to attempting to salvage these constructs for structural studies, a second set of proteins was screened: three more *Mtb* MmpLs, namely — MmpL5, MmpL7 and MmpL11.

3.3 Other *M. tuberculosis* MmpLs

As the only full-length Cluster-II *Mtb* MmpL besides MmpL3, MmpL11 (26% identity, 46% similarity; Sequence Alignment Figure A.1) was an obvious candidate for screening. It too has been implicated in the export of MA containing lipids (e.g. MWEs) required for cell wall biogenesis (**Pacheco et al., 2013; Wright et al., 2017**), as well as aspects of mycobacterial haem utilisation (**Owens et al., 2013**).

MmpL11 and MmpL3 are also syntenic — both being encoded in a conserved mycobacterial gene cluster. This is even retained in species having experienced substantive "genetic decay" (e.g. *M. leprae*, Varela et al. (2012)); congruent with physiological indispensability.

The Cluster-I *Mtb* MmpLs chosen — MmpL5 and MmpL7 (Sequence Alignment Figure A.4) — were decided upon on the basis of documented roles in antibiotic resistance or lipid transport. Additional to a function in siderophore export (Wells et al., 2013), over expression of MmpL5 in various mycobacterial species has been linked to efflux of, and thus resistance to, multiple compounds. These include bedaquiline (one of the newest licensed tuberculosis drugs) and clofazimine (an antileprotic and third-line antitubercular, Hartkoorn et al. (2014)) as well as azoles (Milano et al., 2009), thiacetazones (Halloum et al., 2017), mycobactericidal ionophores (Huang et al., 2017), and even a novel MmpL3 inhibitor (Li et al., 2020a).

As noted in the Introduction, MmpL7 was the first MmpL to be associated with lipid transport (Cox et al., 1999), in particular translocation of PDIMs; complex lipidic virulence factors restricted to *Mtb* and other mycobacterial pathogens (Camacho et al., 2001).

3.3.1 Expression Tests

Of the three MmpLs, only MmpL11 could be made to express as a TEV-GFP-His_8 fusion in *E. coli*; achieving whole-cell GFP counts of barely ~6,500 RFU mL^{-1} in TB (Table 3.3). MmpL5 was also unexpressible as a straight TEV-His_8 construct. Alongside the results of Section 3.2, one can tentatively posit that *E. coli* struggles to produce Cluster-I MmpLs but can better accommodate Cluster-II biogenesis.

Hoping a more native environment — of closer evolutionary relatedness to *Mtb* and *Cgl* — would permit MmpL5 expression, as well as improve the outlook for the initial homologue set, use of *Msmg* strain mc^2155 as an alternative het-

Protein	Organism	Gene	UniProt	Residues	Mass (Da)	GFP Counts (RFU mL^{-1})
MmpL5	*M. tuberculosis*	*Rv0676c*	P9WJV1	964	104,785	–
MmpL7	*M. tuberculosis*	*Rv2942*	P9WJU7	920	95,122	–
MmpL11	*M. tuberculosis*	*Rv0202c*	P9WJT9	966	103,502	6.5K
Truncations						
$MmpL11_{1\text{-}693}$	" "	" "	" "	693	73,426	6.0K
$MmpL11_{1\text{-}714}$	" "	" "	" "	714	76,147	8.1K

Table 3.3: Additional MmpL Trial Set Details & Expression Counts
Salient information concerning MmpL5, MmpL7 and MmpL11 curated from UniProt (Bateman, 2019), alongside details of the C-terminal MmpL11 truncations made. Overnight TB expression levels observed for each pWaldo-TEV-GFP-His_8 construct in *E. coli* C43(DE3) cells are presented. Dashes (–) correspond to null expression.

erologous host was briefly trialled. A model, non-pathogenic, "fast" replicating mycobacterium with a growing genetic toolkit, *Msmg* is increasingly finding use in mycobacterial structural biology (Bashiri & Baker, 2015). MmpL5 and the four initial targets were cloned into pACE-3C-GFP-His_{10}, an acetamide-inducible *Msmg* vector offering tag-free protein retrieval following cleavage with 3C protease (Arnold et al., 2018).

Unfortunately just CmpL3 and NCgl-1194 were semi-convincingly expressed, and at marginal levels relative to a positive control (determined by in-gel fluorescence of lysate, see Supplementary Figure B.3). This is at odds with the literature where *Mtb* MmpL3 has been expressed in *Msmg* in multiple studies (e.g. Belardinelli et al. (2016)), and may indicate the vector employed here is suboptimal. Given more material could be obtained more quickly from *E. coli*, use of *Msmg* as a "rescue" expression host was discontinued. Tests of MmpL7 and MmpL11 were not attempted.

In an effort to boost *E. coli* MmpL11 expression two C-terminal truncations were made, $MmpL11_{1\text{-}693}$ and $MmpL11_{1\text{-}714}$; designed to remove the entire, or the vast majority, of the CTD respectively. Unlike MmpL3, the CTD of MmpL11 is not predicted to be disordered and so may reasonably be thought less delete-

rious. Nevertheless, shorter constructs are often better expressing. Although MmpL11$_{1\text{-}693}$ reduced expression further, MmpL11$_{1\text{-}714}$ was modestly better than the full-length protein at ~8,100 RFU mL^{-1} (Table 3.3). Notwithstanding low GFP counts, all three MmpL11 variants proceeded to DDM FSEC trials.

3.3.2 MmpL11 FSECs & Purifications

MmpL11 FSEC behaviour (Figures 3.7[A]–[C]) mirrored the expression data, in so far that MmpL11$_{1\text{-}693}$ appeared the worst, with a prominent void, and MmpL11$_{1\text{-}714}$ best; being monodisperse and boasting the highest DDM solubilisation efficiency (~95%). Whilst the full-length gave a bimodal FSEC trace, it was free from aggregation and was thus taken forward to purification trials alongside MmpL11$_{1\text{-}714}$. Neither construct purified well in DDM. MmpL11$_{1\text{-}714}$ did bind significantly better to the nickel resin (~55% vs. ~20%), but the IMAC elutions of both showed abundant GFP-tagged species leading from the void on gel filtration (Figures 3.7[D]&[F]). Accompanying analysis by SDS-PAGE and in-gel fluorescence (Figures 3.7[E]&[G]) revealed a sizeable number of contaminants, as well as GFP-tagged degradants. MmpL11-TEV-GFP-His$_8$ migrated as a high MW species, only just entering the SDS-PAGE gel; again consistent with unfavourable self-association. Despite TEV protease treatment realising an increase in free GFP, neither cleaved full-length nor truncation could be recovered after reverse IMAC. This is suggestive of post-cleavage precipitation and/or adherence to the HisTrap column. Further work on MmpL11 was considered futile and a final effort employing an MmpL11-FLAG construct was likewise unsatisfactory.

Having now screened seven proteins, most in two expression hosts and often with various affinity tags, a truly tractable construct still remained to be found. Noting that CTD truncation of MmpL11 had yielded improved behaviour, though far from salvative, the same approach was applied to MmpL3.

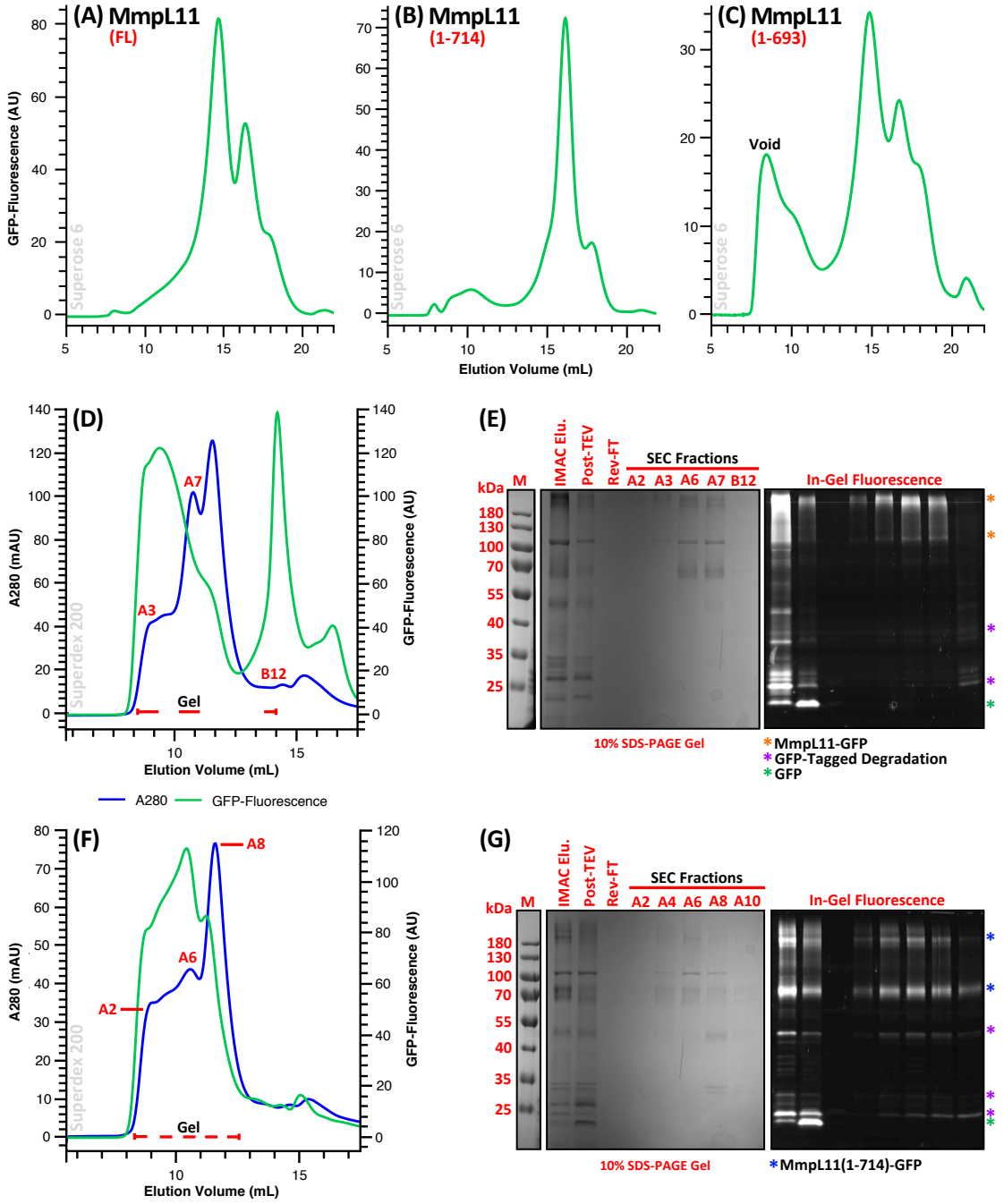


Figure 3.7: FSEC & Purification of MmpL11-TEV-GFP-His$_8$ Constructs in DDM

Superose 6 FSEC traces of 1% (w/v) DDM solubilised *E. coli* membranes expressing **(A)** full-length (FL) MmpL11–, **(B)** MmpL11$_{1\text{-}714}$– or **(C)** MmpL11$_{1\text{-}693}$-TEV-GFP-His$_8$ fusions. Void volume is ~8 mL and is annotated where prominent. **(D)** S200 profile of MmpL11-TEV-GFP-His$_8$ immediately following IMAC elution. In-line GFP fluorescence is overlain. Key peak fractions are labelled and the red bars correspond to material visualised on the SDS-PAGE gel of **(E)**; presented alongside the corresponding GFP in-gel. Notable species are marked with asterisks. **(F)** & **(G)** display the equivalent for the MmpL11$_{1\text{-}714}$-TEV-GFP-His$_8$ truncation.

3.4 The MmpL3 Truncation

Compared to MmpL11, the rationale for MmpL3 CTD truncation was compelling. Not only predicted to be largely disordered (Figure 3.8[A]) — a prime target for both *E. coli* proteases and adventitious interactors (as noted in Section 3.2) — the MmpL3 CTD is by far the least conserved domain of the transporter (Figure 3.8[B]). Consistent with lowered sequence constraints, within what is a mycobacteria universal and *Mtb* essential gene (Domenech et al., 2005; McNeil et al., 2017), the CTD is superfluous to core function. It is thought to regulate MmpL3 localisation during the mycobacterial cell cycle (Carel et al., 2014) and a number of viable truncations have been reported (annotated in Figure 3.8[A], Belardinelli et al. (2016)). Using these as guides, alongside transmembrane helix (see Sequence Alignment Figure A.1) and disorder predictors, a truncation at aspartate 753 (i.e. removal of the last 191 residues) was chosen.

From the outset MmpL3$_{1\text{-}753}$ displayed considerable improvements over the full-length protein, expressing with whole-cell GFP counts of up to ~30,000 RFU mL^{-1} in TB (Table 3.4). Excluding NG, the only detergent showing <100% solubilisation efficiency, the truncation FSECs were all monodisperse. Notably they lacked the substantive GFP-tagged degradation species seen for the original construct (Figure 3.8[C]); giving credence to the hypothesised vulnerability of the CTD to proteolysis. The truncation thus proceeded to purification trials.

Truncation	Organism	Gene	UniProt	Residues	Mass (Da)	GFP Counts (RFU mL^{-1})
MmpL3$_{1\text{-}753}$	*M. tuberculosis*	*Rv0206c*	P9WJV5	753	81,511	21 – 30K

Table 3.4: MmpL3 C-Terminal Truncation Details & Expression Counts
MmpL3$_{1\text{-}753}$ information, partly curated from UniProt (Bateman, 2019). Overnight TB expression range for the associated pWaldo-TEV-GFP-His$_8$ construct in *E. coli* C43(DE3) cells is given.

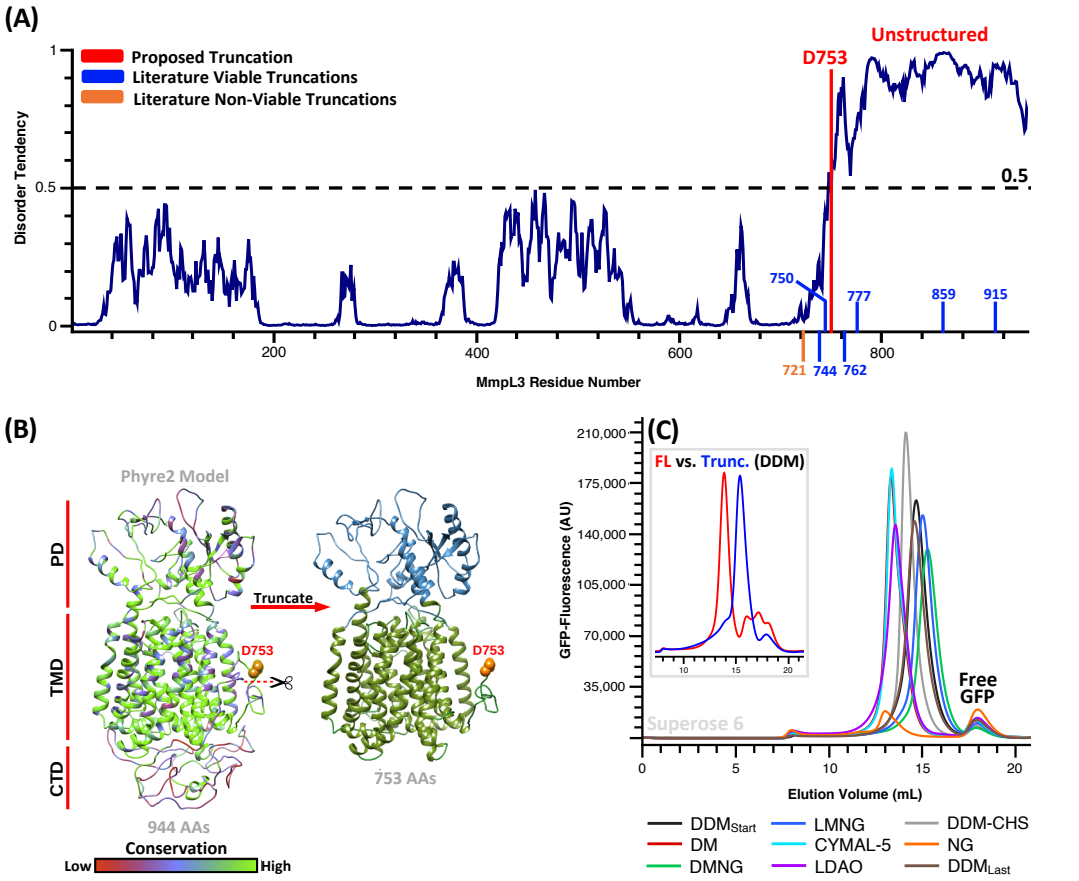


Figure 3.8: MmpL3 C-Terminal Truncation Rationale & FSEC Traces

(A) IUPred2A (Dosztányi, 2018) sequence-based prediction of full-length *Mtb* MmpL3 disorder, clearly identifying an unstructured C-terminus. Literature reported viable (blue) and non-viable (orange) truncations (Belardinelli et al., 2016), alongside the D753 truncation made here (red), are annotated. **(B)** Phyre2 (Kelley et al., 2015) "intensive" mode *Mtb* MmpL3 structural model; coloured by residue conservation determined from a T-Coffee (Madeira et al., 2019) multi-sequence alignment of >40 manually curated mycobacterial MmpL3 orthologues. Truncation at D753 is also depicted. Composed in UCSF Chimera v.1.12 (Pettersen et al., 2004). **(C)** Superose 6 FSEC traces of 1% (w/v) detergent solubilised *E. coli* membranes expressing $MmpL3_{1\text{-}753}$-TEV-GFP-His_8. DDM-CHS has a final cholesterol hemisuccinate (CHS) concentration of 0.2% (w/v). Column equilibrated in DDM. Inset: overlay of full-length and $MmpL3_{1\text{-}753}$ DDM FSECs.

3.4.1 Purifications

The favourable behaviour of $MmpL3_{1\text{-}753}$ reassuringly continued into purification. Nickel resin binding efficiency was seen to be ~30–60% over many repeats in DDM. IMAC elution yielded reasonably clean fusion protein, giving a entirely fluorescent dual-peaked profile when analysed by SEC — hinting at possible oligomerisation (Figure 3.9[A]). Visualisation of peak material by in-gel fluorescence identified one prominent band alongside a number of ancillary, SDS-

resistant, higher MW species. More importantly, no GFP-tagged degradants were detected (Figure 3.9[B]). Unlike the full-length protein, the truncation cleaved to completion on overnight dialysis with TEV protease. Pure tag-free MmpL3$_{1\text{-}753}$ could also be readily recovered following reverse IMAC (Figure 3.9[C]). After concentration, polishing by gel filtration would consistently give rise to a shouldered monodisperse trace (Figure 3.9[D]). To some degree, shoulder abundance and resolution would correlate with protein concentration at injection; more suggestive of undesirable self-association, as was seen for many of the constructs in Section 3.2, than oligomerisation. Material within the shoulder (i.e. fractions "A3–A7") displayed two additional higher order bands by SDS-PAGE (Figure 3.9[E]). These were confirmed, alongside the main band, to contain MmpL3$_{1\text{-}753}$ by mass spectrometry. Protein in the dominant SEC peak (i.e. fractions "A8–A10") retained its elution volume when rerun down the S200 column, arguing against a rapid equilibration with the shoulder species. As such, most subsequent work undertaken with DDM purified MmpL3$_{1\text{-}753}$ only used protein from fraction "A8/9" or later. By retaining just this material the yield of MmpL$_{1\text{-}753}$ was still ≥1 mg L^{-1} of TB.

Whilst not initially tested alongside DDM, it is salient to discuss the improvements to MmpL3$_{1\text{-}753}$ purification that were later achieved using LMNG. Also referred to as MNG-3, the structure of LMNG is equivalent to two DDM monomers having been fused at the second carbon of their alkyl chains. This forms a new quaternary centre, reducing detergent flexibility (Table 3.1, Chae et al. (2010)). More lipid mimetic than conventional detergents, boasting two hydrophobic tails of equal length, LMNG has been documented to stabilise multiple proteins relative to DDM; without detriment to oligomeric state, amenability to crystallisation, or activity (Breyton et al., 2019). In many cases the latter two characteristics also show improvement. Truncation IMAC binding efficiency was enhanced in LMNG, shifting to a narrower range of ~50–60%. As in DDM, complete fusion cleavage was realised on TEV protease addition and reverse IMAC yielded comparably pure

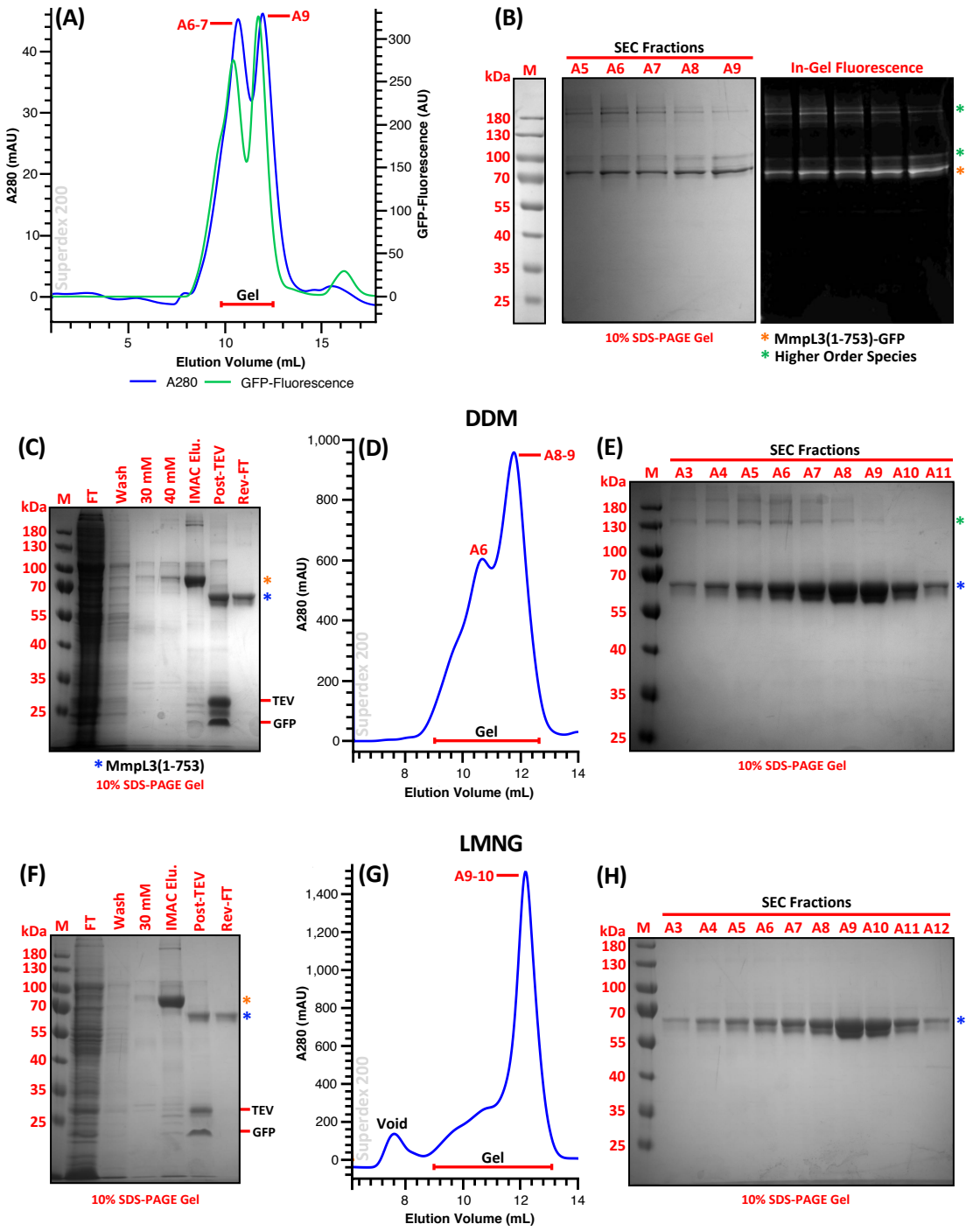


Figure 3.9: Purification of MmpL3$_{1\text{-}753}$-TEV-GFP-His$_8$ in DDM & LMNG

(A) S200 profile of MmpL3$_{1\text{-}753}$-TEV-GFP-His$_8$ IMAC elution in DDM. In-line GFP fluorescence is overlain and peak fractions are labelled. Red bar corresponds to material visualised on the SDS-PAGE gel of **(B)**; presented with the associated GFP in-gel. Notable species are marked with asterisks. **(C)** SDS-PAGE gel of truncation DDM purification, from nickel resin flow-through (FT) to reverse IMAC flow-through (Rev-FT). **(D)** S200 SEC trace of concentrated untagged MmpL3$_{1\text{-}753}$ in DDM. Red bar denotes fractions loaded on the SDS-PAGE gel of **(E)**. Equivalent data for the purification of MmpL3$_{1\text{-}753}$ in LMNG are shown in **(F)** – **(H)**.

protein (Figure 3.9[F]). Subjecting the LMNG solubilised tag-free $MmpL_{1\text{-}753}$ to SEC gave a sharp monodisperse peak, with minimal shoulder and, on occasion, a minor void species (Figure 3.9[G]). Indicative of LMNG discouraging shoulder formation relative to DDM, the associated higher MW bands were likewise reduced, or sometimes completely absent, on SDS-PAGE evaluation (Figure 3.9[H]). This is in agreement with reports of LMNG reducing protein aggregation when substituted for DDM (Breyton et al., 2019).

Clearly the MmpL3 CTD was the major determinant of undesirable full-length behaviour. $MmpL3_{1\text{-}753}$ is by far the best construct of all those tested, purifying well in DDM and even more favourably in LMNG. Prior to starting structural studies in earnest (see Chapter 4), aspects of truncation oligomeric state, fold and thermal stability were first checked.

3.4.2 Characterisation

Oligomeric State

In an effort to ascertain the oligomeric or aggregative status of DDM purified $MmpL3_{1\text{-}753}$ present in the gel filtration peak and shoulder, samples of each were assessed by SEC-MALS. This couples fractionation by SEC with in-line UV absorbance, light-scattering, and refractive index measurements to calculate species molar mass as they leave the column (Dobrin & Randall, 2006). The peak material (i.e. shoulder-free) showed a near-constant total mass across its SEC-MALS elution profile, corresponding reasonably to the expected size of an $MmpL3_{1\text{-}753}$ monomer (~82 kDa) in a DDM micelle. The additional species in the shoulder sample had a highly variable mass; more consistent with aggregates than true oligomers (Figure 3.10[A]).

As for full-length MmpL3, the truncation was also analysed by BN-PAGE — in both DDM and LMNG (Figure 3.10[B]). Again a multi-banded migration pattern was evident, but consistent with gel filtration results the LMNG sample exhibited

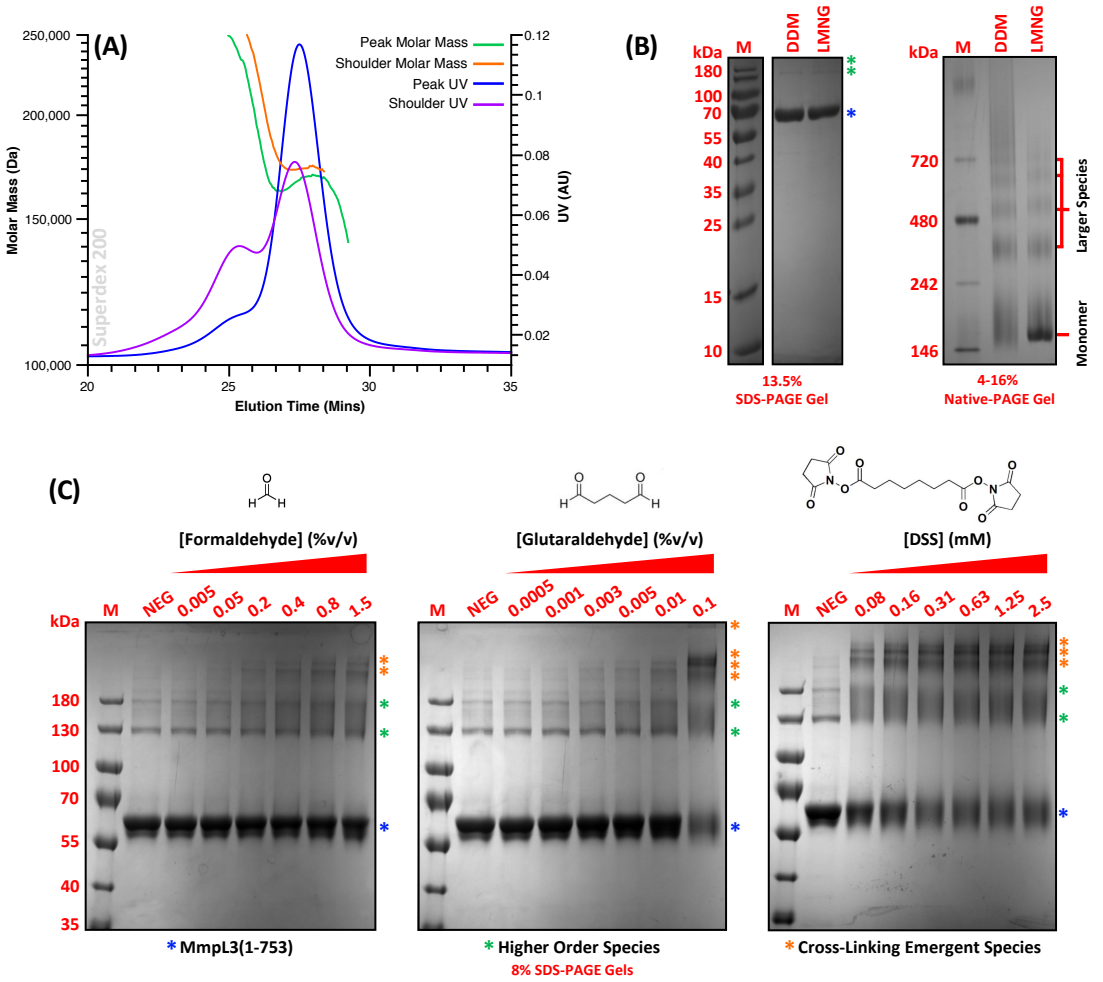


Figure 3.10: Evaluation of MmpL3$_{1\text{-}753}$ Oligomeric State

(A) Overlain SEC-MALS profiles of 1 mg mL^{-1} DDM purified MmpL3$_{1\text{-}753}$ material from the peak (blue & green) and shoulder (purple & orange) fractions of the gel filtration trace. Performed by David Staunton. **(B)** Novex Bis-Tris BN-PAGE gel comparison of MmpL3$_{1\text{-}753}$ solubilised in DDM and LMNG. The same samples run on an SDS-PAGE gel are presented for reference. **(C)** SDS-PAGE gels of MmpL3$_{1\text{-}753}$ in DDM and cross-linked with formaldehyde, glutaraldehyde, and DSS at the shown concentrations. Negative controls (NEG) lack cross-linker. Notable species are marked with asterisks.

less high MW smearing and a crisper, more intense, band of monomer compatible mass. Finally, to test whether the SDS-resistant shoulder species could still be oligomers, cross-linking was attempted (Figure 3.10[C]). The dominant MmpL3$_{1\text{-}753}$ band failed to "collapse" into the two higher order species for all cross-linkers employed, arguing against their oligomeric status. New bands only emerged above the 180 kDa marker and are, without further study, of unknowable relevance. They are probable aggregates. Taken together, the results indicate MmpL3$_{1\text{-}753}$ exists predominantly as monomer with differential tendency towards

self-association depending on purification detergent choice; LMNG repressing this attribute as compared to DDM.

Fold & Thermal Stability

To investigate the secondary structure and thermal stability of purified MmpL3$_{1\text{-}753}$, and identify any divergence in these traits between DDM and LMNG, circular dichrosim was used. Protein CD spectroscopy exploits the unequal absorption of left-handed and right-handed circularly polarised far UV light (~240–190 nm) by peptide bonds; contingent on the asymmetry provided by the greater protein fold (Pain, 2004; Miles & Wallace, 2016). As a result, characteristic CD spectra are generated for membrane proteins of differing secondary structural composition (see Figure 3.11[A]). This allows CD traces to function as a rapid readout for the presence, and class, of fold exhibited by novel proteins. Scans of MmpL3$_{1\text{-}753}$, in either DDM or LMNG, displayed a positive band at ~193 nm alongside negative bands at ~209 nm and ~222 nm (Figure 3.11[B]). These features are typical of proteins with abundant α-helical secondary structure (Greenfield, 2006), suggesting MmpL3$_{1\text{-}753}$ is folded in both detergents and that this fold is at least superficially consistent with the expected RND architecture.

Through combination of successive CD scans with an incremental temperature gradient, one can follow the loss of secondary structure as a function of temperature and therefore estimate protein thermal stability. Loss of CD at a single wavelength, here 209 nm, is then used to derive a melt temperature (T_m) — the temperature at which 50% of the starting CD signal has been lost during the unfolding transition. Comparing T_m values between conditions thus delineates stabilising factors of relevance to structural pursuits. As at all previous stages, LMNG showed desirability over DDM; stabilising MmpL3$_{1\text{-}753}$ by an additional ~6 °C (Figures 3.11[C]&[D]). Nevertheless, at ~48 °C the truncation is fairly stable in DDM. The sharp sigmoidal transition observed for LMNG is congruent with a single cooperative unfolding step. In comparison, the slightly broader

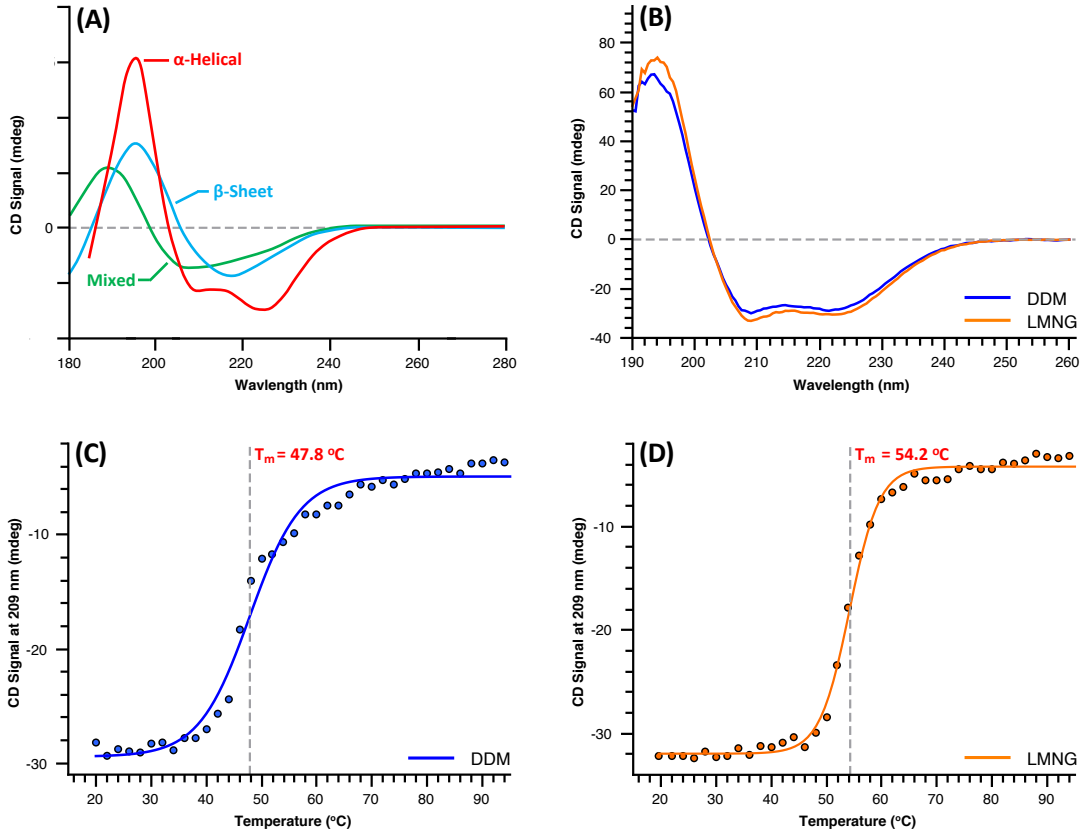


Figure 3.11: CD Spectroscopy of MmpL3$_{1\text{-}753}$ in DDM & LMNG

(A) Typical CD spectra recorded for membrane proteins of different secondary structural composition. Modified from Miles & Wallace (2016). **(B)** Overlain CD scans of MmpL3$_{1\text{-}753}$ purified in DDM (blue) and LMNG (orange). MmpL3$_{1\text{-}753}$ CD melt profiles at 209 nm in the same two detergents are shown in **(C)** & **(D)** respectively. Full spectra can be found in Supplementary Figure B.4. Temperature was increased from 20–95 °C in 2 °C increments. Melt temperatures (T_m) were determined from sigmoidal fits performed in QtiPlot.

transition seen for DDM may, speculatively, suggest greater heterogeneity of the starting sample and/or subsequent unfolding pathway. Overall, the CD findings show MmpL3$_{1\text{-}753}$ is suitably folded and stable for structural studies; be it in DDM or LMNG.

Additional Quality Indicators

Supplementing the screening pipeline, results from a collection of otherwise miscellaneous experiments, undertaken throughout the project, can also be leveraged as indicators of purified MmpL3$_{1\text{-}753}$ quality. Two are worth noting here. The first is the amenability of the truncation to reconstitution into asolectin

(soybean lipid extract) liposomes (data not shown); a prerequisite for *in vitro* transporter assay development. Albeit it with <100% efficiency, the ability of MmpL3$_{1\text{-}753}$ to be heterologously expressed, extracted into detergent, purified, and then successfully reintroduced into a liposomal bilayer reflects admirably on protein quality. The second concerns the putative, and somewhat controversial, role of MmpL3 in an *Mtb* haem scavenging pathway (Tullius et al., 2011; Mitra et al., 2019). The first periplasmic subdomain and CTD of *Mtb* MmpL3, when expressed in isolation, were reported to bind haem by Owens et al. (2013). To test if the truncation displayed similar activity by a comparable methodology, DM solubilised MmpL3$_{1\text{-}753}$ was incubated with haem over range of protein : haem molar ratios and then analysed by SEC. In-line absorbance at 370 nm, when overlain with protein absorbance at 280 nm, could be used to infer haem co-elution with MmpL3$_{1\text{-}753}$ but not the negative control (DM purified PepTSt, Supplementary Figure B.5). Corroborating the literature, this strengthens the case for the relevance of the truncation construct pursued here — even if the physiological role of such haem association remains far from clear.

3.4.3 Conclusion

After many iterations through the screening pipeline — necessitating exploration of two expression hosts, various affinity tags, multiple vectors, a range of detergents, and three C-terminal truncations — a viable construct of one of the seven screened proteins was identified. Satisfyingly, this was also the protein of primary interest; *Mtb* MmpL3. Removing the CTD, to give MmpL3$_{1\text{-}753}$, had proved transformative. The resulting truncation expressed well in *E. coli* as TEV-GFP-His$_8$ fusion, could be purified to yield significant amounts of folded homogeneous protein (≥1 mg L^{-1} of TB) in at least two detergents, and displayed favourable thermal stability. Unlike the majority of characterised RND transporters, it appeared monomeric under the purification conditions employed. MmpL3$_{1\text{-}753}$ clearly met the requirements for structural studies set out at the beginning of the chapter, and it would go on to become the main construct of the project.

The MmpL3 Truncation

Initial Structural Attempts

CHAPTER 4

The ample quantities of pure MmpL3$_{1\text{-}753}$ obtainable was favourable for commencement of structural studies, in particular crystallography. Successful generation of structure-grade membrane protein crystals is, however, a notorious bottleneck to high resolution structure determination. One must first locate, and then usually iteratively optimise, a set of conditions — within a dauntingly dimensive chemical space — able to induce target self-assembly into an ordered 3D array. Simultaneously, native fold must be maintained. The increasing commercial availability of rationally designed crystallisation screens, alongside robotisation, has expedited this search process, but it essentially remains semi-empirical. The following chapter details initial efforts to crystallise the truncation by two methods. One aims to directly crystallise the PDC (vapour diffusion), while the other exploits the polymorphic phase properties of specific lipids to promote crystallisation after reconstitution into a bilayer (lipidic cubic phase). Unable to achieve crystallogenesis with either, investigations of MmpL3$_{1\text{-}753}$ proceeded via Cryo-EM; yielding modest structural insights.

4.1 Crystallisation Trials

4.1.1 Vapour Diffusion

Attempting MmpL3$_{1\text{-}753}$ crystallisation by vapour diffusion (VD) was the logical place to begin. Not only being the most common method of membrane protein crystallisation (Delmar et al., 2015), it is the approach by which nearly all reported RND superfamily crystal structures have been obtained. This likely

stems from the advantages provided by their sizeable extramembranous domains; supplying a large hydrophilic surface, unobscured by detergent, for the robust inter-protein contacts required for type-II 3D crystals (Birch et al., 2018). These *in surfo* trials were performed with aid of a Mosquito crystallisation robot (STP Labtech), used to generate plates of 96 × 200 nL sitting drops of equal volume concentrated protein and precipitant well solution. The latter was sourced from four sparse matrix screens — namely MemGold (Newstead et al., 2008), MemGold2 (Parker & Newstead, 2012), MemTrans and MemChannel — each containing an array of conditions datamined from the PDB, and thus with a proven track record for α-helical membrane protein crystallogenesis. Post-sealing, plates were incubated at 4 or 18 °C on housing in a Rigaku Minstrel crystal imaging system; automated to periodically record visible light and tryptophan UV fluorescence images on a per well basis. Table 4.1 summarises the VD trials undertaken with $MmpL3_{1\text{-}753}$.

Ideally, as the airtight well system reaches equilibrium, protein in the sitting drop will slowly escape supersaturation via crystallisation rather than premature aggregation (Delmar et al., 2015). Although varying both purification detergent (DM, DDM & LMNG) and protein concentration (spanning ~4–20 mg mL^{-1}), in addition to temperature and crystallisation screen, near-imminent precipitation

Detergent	Conc. (mg mL^{-1})	Temp. (°C)	Screens
DM	~6 & 12	4 & 18	MG
DDM	~4 & 8	4 & 18	MG, MG-II, MT & MC
DDM	~10 & 20	4 & 18	MG, MG-II, MT & MC
LMNG	~9	18 / 4 & 18	$MG^{4,18}$, $MG\text{-}II^{18}$ & MT^{18}
LMNG	~20	18	MG & MG-II

Table 4.1: Summary of $MmpL3_{1\text{-}753}$ Vapour Diffusion Trials

Details of sitting drop crystallisation trials undertaken with the truncation. From LTR: Purification detergent, protein concentration (NanoDrop), plate incubation temperature, and the sparse matrix crystallisation screens trialled (MemGold (MG, Newstead et al. (2008)), MemGold2 (MG-II, Parker & Newstead (2012)), MemTrans (MT), & MemChannel (MC)) are tabulated. Superscript values denote breakdown of incubation temperatures where given.

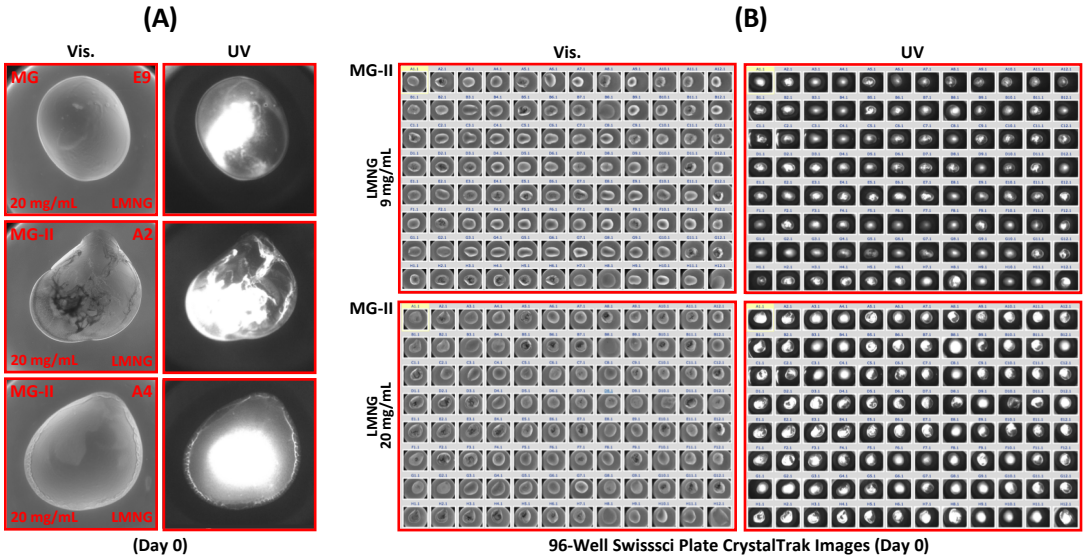


Figure 4.1: Example Results of MmpL3$_{1\text{-}753}$ Vapour Diffusion Trials in LMNG

(A) Representative wells, imaged by visible light (Vis.) and tryptophan UV fluorescence, soon after vapour diffusion setup with LMNG solubilised MmpL3$_{1\text{-}753}$. Annotated (clockwise) with crystallisation screen (see Table 4.1), well number, purification detergent, and protein concentration. **(B)** Comparison of entire MemGold2 plates prepared with truncation at 9 & 20 mg mL^{-1}, imaged imminently after housing. All images collected on a Rigaku Minstrel and accessed via CrystalTrak.

of the truncation was seen frequently, often in the majority of wells (Figure 4.1). Whilst this could be somewhat alleviated by lowering protein concentration, crystallisation remained stubbornly absent. Rationalising that the MmpL3$_{1\text{-}753}$-detergent complex was therefore probably too rapidly destabilised in the presence of the screened precipitants, crystallogenesis by lipidic cubic phase (LCP) was next tried — the requisite reintroduction into a bilayer having a known stabilising effect (Liu et al., 2010).

4.1.2 Lipidic Cubic Phase

In the LCP or *in meso* method, concentrated target protein is reconstituted into a quasi-solid bilayer environment through homogenisation with a LCP-competent host lipid — most commonly, and first demonstrated with, monoolein (MO, 9.9 monoacylglycerol (MAG); Landau & Rosenbusch (1996)). If combined, generally by dual-syringe mixer, in the correct ratio (~2:3 protein : lipid w/w for MO) and at a permissible temperature (~20 °C for MO), then the cubic phase

spontaneously forms (Birch et al., 2018). This is an intricate three-dimensional curved bilayer — in which incorporated proteins can laterally diffuse — bicontinuous with a network of interpenetrating aqueous channels (Caffrey & Cherezov, 2009). The protein-laden mesophase can then be screened against a panel of precipitant solutions to identify any favouring nucleation and subsequent bulk crystallisation. This is thought to arise by local precipitant-induced lamellarisation of the cubic phase, facilitating both the inter-protein contacts required for type-I crystallisation as well as conducting diffusive delivery of fresh protein to drive crystal growth (Caffrey, 2015).

In meso trials with MmpL3$_{1\text{-}753}$, begun in MO with DDM purified protein, were greatly expedited by the semi-automation provided by a Gryphon LCP robot (Art Robbins). In short, this was used to dispense 96 × 50 nL boli of truncation-laden cubic phase onto a glass sandwich plate, and then overlay these with 800 nL of an equivalent number of different precipitant solutions — sourced from commercial (e.g. MemMeso) or custom (e.g. the Tokyo Set (TS), Nureki Lab, University of Tokyo) sparse matrix crystallisation screens. Once sealed, plates were periodically monitored for evidence of crystallogenesis by a combination of visible light, cross-polarisation, and tryptophan UV fluorescence microscopy. A summary of the MmpL3$_{1\text{-}753}$ LCP trials is collated in Table 4.2.

Despite testing nine crystallisation screens — alongside parallel variations in host lipid (MO or the shorter 7.8 MAG), protein concentration (ranging ~18–42 mg mL^{-1}), upstream purification detergent (DDM or LMNG), and plate incubation temperature (4 or 20 °C) — successful MmpL3$_{1\text{-}753}$ crystallisation was never realised. Although crystalline species frequently appeared, these were readily shown to be false positives (e.g. salt) through absence of UV fluorescence, presence of equivalent objects outside the bolus, and/or on being harvested and shot with X-rays (Figure 4.2). Before moving on, it is worth briefly discussing the behaviour of MmpL3$_{1\text{-}753}$ during LCP setup and the factors potentially contributing to crystallisation failure.

Detergent	Conc. (mg mL^{-1})	Host Lipid	Temp. (^{o}C)	Screens
DDM	~18	MO	20 / 4 & 20	MM^{4,20} & TS1-6^{20}
DDM	~18	7.8 MAG	20	MM & TS1-6
DDM	~28	MO	20 / 4 & 20	MM^{4,20} & TS1-6^{20}
DDM	~28	7.8 MAG	20	TS1-6
DDM	~37	MO	4 / 4 & 20	MM^{4}, OX3/6/9^{4,20} & TS6^{4,20}
LMNG	~18	MO	20 / 4 & 20	MM^{4,20} & TS1-6^{20}
LMNG	~42	MO	20	MM & TS1-6

Table 4.2: Summary of MmpL3$_{1\text{-}753}$ LCP Trials

Specifics of the LCP crystallisation trials performed with the truncation. From LTR: Purification detergent, protein concentration (NanoDrop), cubic phase host lipid (Monoolein (MO) or 7.8 Monoacylglycerol (MAG)), plate incubation temperature, and the 96-condition sets screened (MemMeso (MM), Tokyo Set (TS, Nureki lab), & Oxford Set (OX)) are tabulated. Superscript values denote breakdown of incubation temperatures where given.

Successful formation of the cubic phase, on mixing of host lipid and protein, is typified by development of optical transparency (Caffrey & Cherezov, 2009). Visual inspection of all MmpL3$_{1\text{-}753}$ LCP batches displayed some degree of residual "haziness"; indicating that formation of the desired mesophase was likely suboptimal and thus a clear hindrance for crystallogenesis. This was most evident for samples purified in DDM and homogenised with 7.8 MAG, to the point where attempts at ~37 mg mL^{-1} protein induced mass precipitation. It was hoped setup at higher concentrations would more readily promote nucleation (Caffrey, 2015). Exploring a range of host lipids and working at the limits of tolerable protein concentration are nevertheless advisable. The former acknowledges that there is no "one-size-fits-all" hosting lipid given the compositional, and so physicochemical, diversity of the biological membranes within which target proteins naturally reside (Li et al., 2011). Screening of the truncation with further MAG variants is probably warranted. One could likewise consider doping the mesophase with mycobaterial-esque lipids; tailoring LCP microstructure with native components having proven successful for a number proteins (e.g. see Li et al. (2014)). Whilst still unsuccessful, introducing LMNG purified MmpL3$_{1\text{-}753}$ into MO consistently gave better, semi-transparent, LCP and protein concentrations up to ~42 mg mL^{-1}

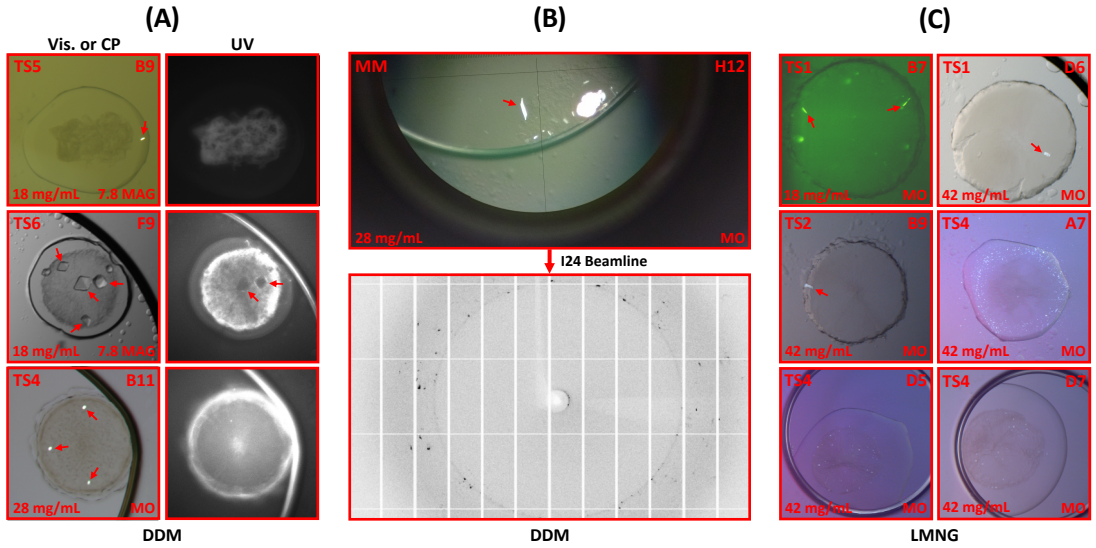


Figure 4.2: Example Non-Protein Crystalline Species Seen in LCP Trials

(A) Representative collection of crystalline, non-UV fluorescent (i.e. non-protein), species (red arrows) formed in the boli of MmpL3$_{1\text{-}753}$ LCP trials using DDM purified protein. Associated visible light (Vis.) or cross-polarised (CP) images are provided; annotated (clockwise) with crystallisation screen (see Table 4.2), well number, host lipid, and protein concentration. **(B)** Example crystalline LCP species (red arrow) shot at beamline I24 (Diamond Light Source), giving the non-proteinaceous diffraction shown. **(C)** Equivalent array of CP images highlighting crystalline objects emergent in the boli of MmpL3$_{1\text{-}753}$ LCP trials using LMNG purified protein. Note appearance of multiple birefringent loci outside of the bolus of Tokyo Set 4 (TS4)-A7, indicative of a high background of crystallising well component.

were easily trialled. This may be linked to the favourable behaviour of the truncation in this detergent (see Chapter 3), although could also represent differing effects of incorporating DDM and LMNG into the MO bilayer.

Finally, unlike the crystal contact advantages it is expected to offer *in surfo*, one could argue that the periplasmic domain (PD) of MmpL3$_{1\text{-}753}$ has the potential to frustrate efforts *in meso*. Predicted to represent ~290 residues of extramembranous bulk (~32 kDa, annotated in Sequence Alignment Figure A.1), the PD would slow — or outright prevent — lateral diffusion within the LCP if incompatible with aqueous channel dimensions (~40–50 Å diameter for MO, Zabara & Mezzenga (2012); Caffrey (2015)). Testament to the importance of these considerations, multiple routes to "swelling" the mesophase (i.e. widening the aqueous channels) have been advanced for crystallisation of proteins with substantial extracellular protrusions (Cherezov et al., 2006; Zabara et al., 2018) — generally larger

than those of the truncation however. Nonetheless, such speculation is congruent with the paucity of LCP structures reported for RND family members. A rare exception is SecDF, part of the bacterial protein translocation machinery, recently solved in MO (Furukawa et al., 2018). Interestingly, the crystallisation condition contained 50% PEG400. This is a documented "spongifier" — a compound inducing transition to the sponge phase; which again boasts a more spacious aqueous network than LCP (Caffrey, 2015). As the TS3 set employed here screens 96 × PEG400 containing conditions, other avenues for LCP swelling were not prioritised and thus their applicability to $MmpL3_{1\text{-}753}$ crystallisation remains to be explored.

4.2 Cryo-EM Trials

Having proven recalcitrant to crystallogenesis by both *in meso* and *in surfo* approaches, a structural understanding of $MmpL3_{1\text{-}753}$ was next sought using single-particle Cryo-EM. The unexpected monomerism of the truncation precludes the methodological benefits afforded to oligomeric RND superfamily transporters — i.e. increased mass and exploitable internal symmetry — and it is thus a conventionally unattractive target for the technique. At only ~82 kDa, $MmpL3_{1\text{-}753}$ is a small particle and would be predicted to display a suboptimal signal-to-noise ratio within vitreous ice, frustrating downstream processing and warranting data collection at greater defocus. The latter boosts image contrast, but at the expense of achievable resolution (Cheng et al., 2015). Nevertheless, persistent exploitation of the technical advances that brought Cryo-EM to prominence — a trinity of automation, algorithmic improvement, and introduction of direct electron detectors (DEDs) (Cheng, 2018) — has seen the size limit for near-atomic resolution structure determination (often contentiously defined at <4 Å, Wlodawer & Dauter (2017)) continually fall (Lyumkis, 2019). With future developments this trend will likely extend to proteins ever closer to the theorised reconstruction limit, estimated at ~38 kDa (Henderson, 1995).

Whilst still uncommon, multiple sub-100 kDa soluble proteins have been successfully resolved by Cryo-EM to better than 4 Å — including isocitrate dehydrogenase (93 kDa, 3.8 Å, Merk et al. (2016)), haemoglobin (64 kDa, 3.2 Å, Khoshouei et al. (2017)), streptavidin (52 kDa, 3.2 Å, Fan et al. (2019)), and alcohol dehydrogenase (82 kDa, 2.7 Å, Herzik et al. (2019)). Evidence of similar progress below the 100 kDa threshold for integral membrane proteins, in the absence of alignment assistive Fab fragments, remains exceedingly rare. Notable exceptions are recent maps of the ER-resident immunoregulator STING (~80 kDa, 4 Å, Shang et al. (2019)), the human zinc-proton antiporter ZnT8 (~70 kDa, 3.8 Å, Xue et al. (2020)), and a SARS-CoV-2 viroporin (~60 kDa, 2.9 Å, Kern et al. (2020)). Although the majority of these examples capitalise on internal symmetry, often alongside contrast enhancement with a Volta phase plate (Danev et al., 2014), they represent an encouraging backdrop for the pursuit of MmpL3$_{1\text{-}753}$ by Cryo-EM; especially given in-house access to a 300 kV Titan Krios equipped with a Gatan K2 DED. The extramembranous domains of MmpL3 could also be reasonably expected, on protruding from the detergent micelle, to help particle classification. Even if, as was most probable, a high resolution structure of the truncation could not be obtained, an intermediate Cryo-EM envelope was still deemed valuable to inform on the overall architecture of the MmpL family. Being the first detergent trialled for purification, DDM solubilised MmpL3$_{1\text{-}753}$ was likewise the initial sample screened in the ice.

4.2.1 DDM

Taking care to avoid undesirable material (see Chapter 3) within the SEC shoulder of DDM purified truncation (Figure 4.3[A]), grids were prepared from peak protein at a range of concentrations (≤5 mg mL^{-1}) with aid of a Vitrobot plunge freezer. Specifics concerning vitrification parameters can be found in Chapter 2. Efforts to optimise particle density via concentration variation were largely futile, given the tendency of the samples towards in-ice aggregation (Figure 4.3[B]).

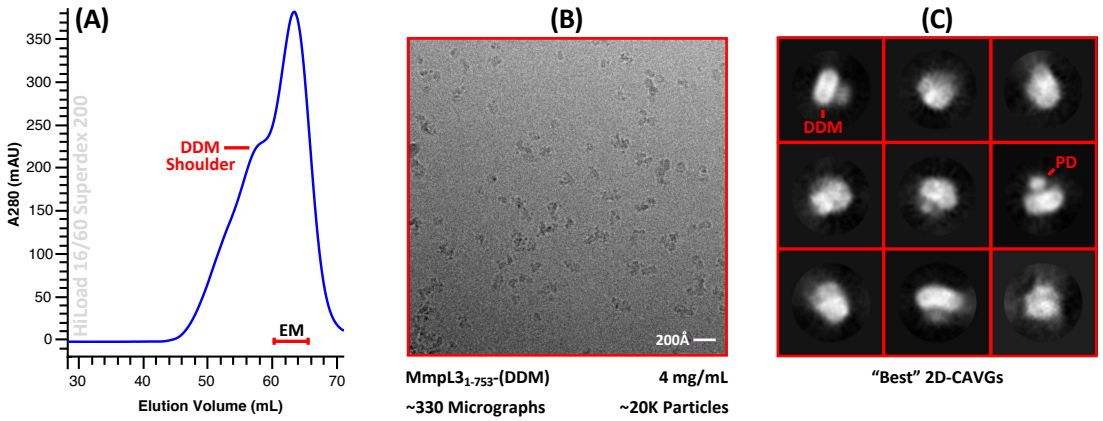


Figure 4.3: Cryo-EM Screening of MmpL3$_{1-753}$ in DDM

(A) HiLoad S200 gel filtration profile of the MmpL3$_{1-753}$ DDM purification which provided material for Cryo-EM screening. Red bar corresponds to fractions concentrated for grid preparation. **(B)** Representative Titan Krios micrograph of DDM solubilised truncation at 4 mg mL^{-1}. The large heterogeneous particles observed are indicative of aggregation. Approximate number of micrographs collected, and subsequent particles extracted, are noted. **(C)** Best 2D class averages (2D-CAVGs), generated using RELION (Scheres, 2012). The DDM micelle and periplasmic domain (PD) are annotated. Imaging and processing courtesy of Justin Deme.

This is consistent with the aforementioned propensity of the protein to self-association, only later found to be more prominent in DDM than LMNG. Equally, it may represent the inability of the conditions (e.g. detergent, buffer composition, blot time etc.) to prevent MmpL3$_{1-753}$ migration to — and denaturation at — the air-water interface prior to freezing (D'Imprima et al., 2019). The aggregation predictably curtailed the number of useful particles recovered. The resulting RELION (Scheres, 2012) 2D class averages (2D-CAVGs), generated from a modest test collection (~330 micrographs, 20K particles), were mostly blurred with ill-defined features (Figure 4.3[C]). However, a subset reassuringly had discernible protrusions from the micelle; density assignable to the MmpL3 PD. Confident, therefore, that a subpopulation of truncation remained folded post-vitrification, efforts were invested in modifying sample preparation to optimise in-ice behaviour. This began with, and was the motivation behind, trialling MmpL3$_{1-753}$ purification in LMNG.

4.2.2 LMNG

The improved purification and thermal stability of LMNG solubilised $MmpL3_{1\text{-}753}$ was pleasingly carried forward into the ice. Vitrification of SEC peak material (fraction "A10", Figure 4.4[A]) allowed realisation of much higher particle density, at a lower protein concentration (~2 mg mL^{-1}) and with less aggregation, relative to DDM (Figure 4.4[C]). The sample also likely benefited from a reduction in freeze-thaw cycles; grids being prepared as soon after gel filtration as possible. Imaging of protein from the minor shoulder (fractions "A4" & "A6"), however, was more reminiscent of DDM and showed increased in-ice aggregation — highlighting the necessity of only using the best obtainable material for Cryo-EM. Given the vast improvement in micrograph quality, a moderate sized dataset was collected (~1300 micrographs). This was processed as summarised in Figure 4.4[D].

To steer extraction towards the most favourable particles (i.e those in isolation, free of neighbour contacts), a subset of 500 micrographs were first semi-manually "swarm" picked in EMAN2 (Tang et al., 2007). The resulting ~38K particles underwent 2D classification, producing 80 references with which to perform automated template-based picking of the entire image stack. This recovered a total of ~90K particles, subsequently processed in both SIMPLE (Reboul et al., 2018) and RELION (Scheres, 2012) to generate 2D-CAVGs (Figures 4.4[E]&[F]). The classes displayed significantly more detail over that seen for DDM. The micellar protrusion was now resolved to be bilobed, providing individual definition to each MmpL3 periplasmic subdomain. Density within the micelle of side views was additionally ascribable to the expected bundle of transmembrane helices. Unquestionably, and in agreement with the biochemical analyses of Chapter 3, $MmpL3_{1\text{-}753}$ is monomeric.

Comparing the best SIMPLE and RELION 2D-CAVGs suggests the former better classifies truncation side views. The latter outputted a single such class with splayed extramembranous signal; congruent with either PD motional freedom

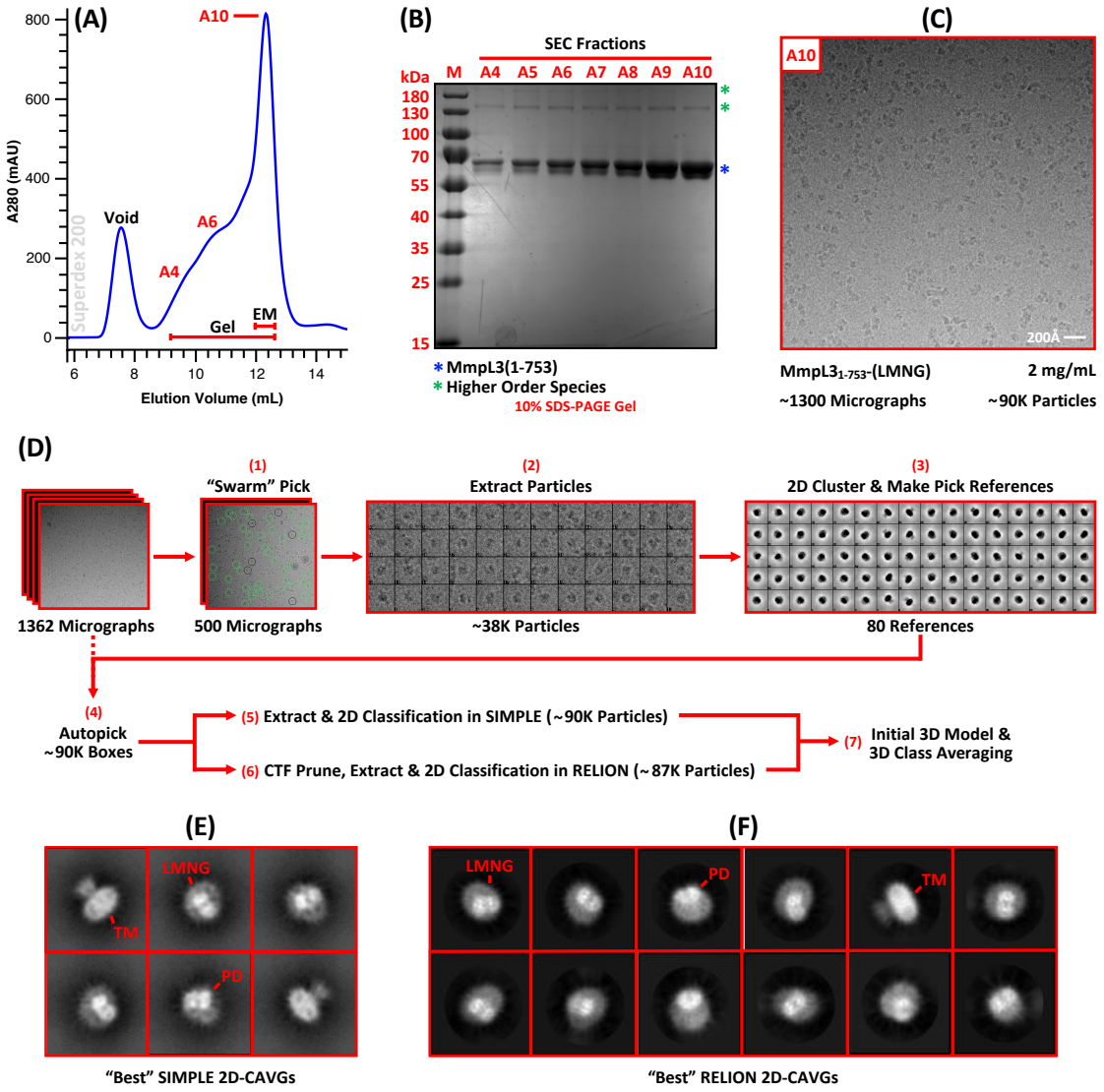


Figure 4.4: Cryo-EM Screening of MmpL3$_{1\text{-}753}$ in LMNG

(A) S200 SEC trace of MmpL3$_{1\text{-}753}$ purified in LMNG. The void peak and fractions referenced in the text are labelled. Red bars correspond to the material loaded on the SDS-PAGE gel of **(B)** or concentrated for in-ice screening on the Titan Krios; a representative micrograph from which is shown in **(C)**. Sample concentration, collected micrograph number and particle set size are also provided. **(D)** Schematic of the Cryo-EM data processing pipeline employed. Semi-manual "swarm" picking was performed in EMAN2 (Tang et al., 2007). The best 2D-CAVGs generated with either SIMPLE (Reboul et al., 2018) or RELION (Scheres, 2012) are shown in **(E)** and **(F)** respectively. The LMNG micelle, periplasmic domain (PD) and density consistent with transmembrane helices (TM) are annotated in both. Processing guidance from Justin Deme.

or collapse of multiple projections into a single aberrant average. Nevertheless, the RELION 2D-CAVGs were, on aggregate, superior. Buoyed by 2D-CAVG interpretability, 3D classification was attempted in both processing suites. Unfortunately, neither could be made to give convincing reconstructions. Alongside small particle size, gathering a sufficient distribution of accurate angular assignments

— i.e. robust 3D class averaging — may have been hampered by $MmpL3_{1\text{-}753}$ pseudo-symmetry, conformational heterogeneity, and/or preferential orientation within the ice. The dominance of periplasmic top views at the 2D stage lends credence to said projection bias; although appearance of side views optimistically means such anisotropy may not be irreconcilable. In an attempt to limit effects of any conformational freedom, further Cryo-EM screening of LMNG solubilised truncation was performed in the presence of haem — with or without mild cross-linking.

Haem Addition

Exploitation of ligands — be they substrates, inhibitors or co-factors — is a popular avenue to explore when trying to reduce conformational heterogeneity (Cheng et al., 2015). If fortuitous, added small molecules can stabilise the target in a rigid, potentially functionally informative, bound state; one more amenable to Cryo-EM. Noted previously, the literature suggests *Mtb* MmpL3 binds haem as part of an ill-defined iron uptake pathway (Owens et al., 2013) — an activity seemingly conserved by the truncation (Supplementary Figure B.5). Easier to purchase, and likely handle, than its primary lipid substrate (i.e. TMM), haem was an obvious choice of structural additive for $MmpL3_{1\text{-}753}$.

As in DM, in-line absorbance at 370 nm confirmed haem co-elution with LMNG purified truncation by SEC (Figure 4.5[A]). Concentrating peak material for grid preparation saw development of an increasingly brown tint, again pointing to haem association. A similar number of micrographs were collected (~1400), and then processed with the same pick references (extracting ~90K particles) and pipeline, as for the original LMNG dataset. Whilst still vitrified at ~2 mg mL^{-1}, micrographs showed a greater degree of particle clustering — verging on aggregation in places (Figure 4.5[B]). Whether this is a consequence of prep variation or directly related to haem addition is difficult to definitively say. No improvement in micrograph quality was achieved on repetition however, hinting

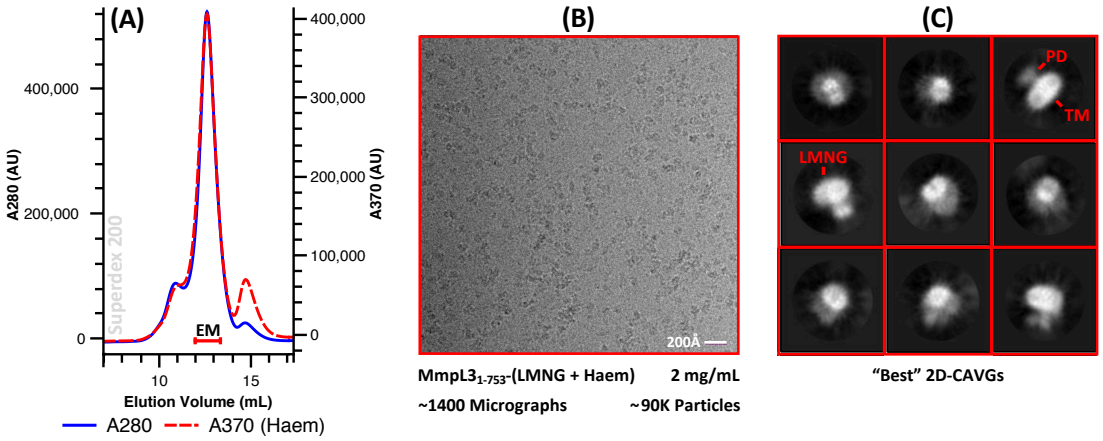


Figure 4.5: Cryo-EM Screening of MmpL3$_{1-753}$ in LMNG + Haem

(A) S200 SEC profile of LMNG purified MmpL3$_{1-753}$ incubated with haem in a 2:1 (haem:protein) molar ratio, for >1 hour, prior to reinjection. Absorbance 280 nm (A280, protein) is overlain with in-line absorbance 370 nm (A370); the latter used to monitor haem elution. Red bar corresponds to fractions concentrated for Cryo-EM. **(B)** Representative Titan Krios micrograph, presented with associated collection metadata. **(C)** Best 2D-CAVGs generated in RELION (Scheres, 2012). The LMNG micelle, periplasmic domain (PD) and transmembrane helix bundle (TM) are labelled. Processing guidance from Justin Deme.

that inclusion of the iron co-factor may well be detrimental. Despite showing a comparable domain organisation, the resulting 2D-CAVGs were unsurprisingly inferior to those generated in LMNG alone (Figure 4.5[C]).

The first MmpL3 periplasmic subdomain is one of two purported haem binding sites (Owens et al., 2013). Interestingly, the extramembranous signal of side views computed in RELION does appear marginally less splayed; although the attributability of such observations to haem is at best conjecture. Either way, the PD remains more poorly resolved than in classes produced by SIMPLE on processing the initial dataset. In the absence of enhanced 2D-CAVGs, further analysis was abandoned prior to 3D reconstruction.

Cross-linking

Far from physiological — i.e. at risk of inducing artefactual structural features — mild chemical cross-linking is nevertheless a common route taken to rigidify targets (Cheng et al., 2015). Ideally one restrains, without deleterious aggregation, domains which are motionally free, thereby minimising conformational hetero-

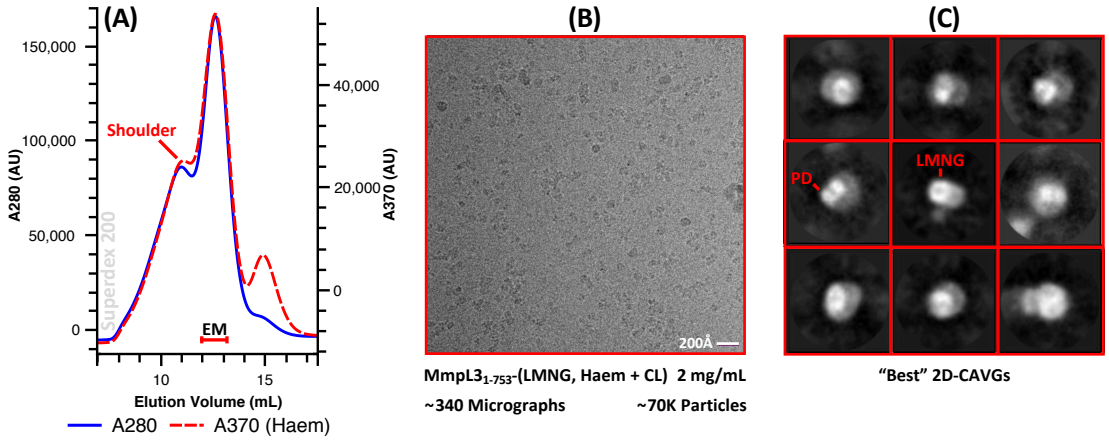


Figure 4.6: Cryo-EM Screening of Cross-linked MmpL3$_{1\text{-}753}$ in LMNG + Haem
(A) S200 gel filtration trace of glutaraldehyde cross-linked (0.05% v/v), LMNG solubilised, MmpL3$_{1\text{-}753}$ post-incubation with haem in a 3:1 (haem:protein) molar ratio. In-line absorbance at 280 nm (A280, protein) is overlain with that at 370 nm (A370); the latter employed for tracking haem. Red bar corresponds to material concentrated for grid preparation. The prominent SEC shoulder emergent on cross-linking is labelled. **(B)** Representative Titan Krios micrograph. Dataset size is noted. **(C)** Best 2D-CAVGs computed in RELION (Scheres, 2012). The LMNG micelle and periplasmic domain (PD) are annotated. Processing guidance from Justin Deme.

geneity. Glutaraldehyde is frequently the reagent of choice. This is highlighted by the moderate traction gained, within the Cryo-EM community, by specialised density gradient cross-linking protocols employing it (e.g. GraFix, Kastner et al. (2008)). Before attempting one of these more involved approaches, gentle glutaraldehyde cross-linking (0.05% v/v) in combination with haem addition was trialled as a means to the improve the 2D-CAVGs of LMNG solubilised MmpL3$_{1\text{-}753}$.

Preliminary experiments suggested cross-linked truncation retained haem binding activity, albeit with possibly reduced efficiency (as determined by SEC co-elution). To combat this, samples were incubated at an increased haem : protein molar ratio (3:1, up from 2:1) prior to preparative gel filtration. The resulting profile showed an emergent shoulder consistent with undesirable cross-linking by-products (Figure 4.6[A]). Avoiding such material, protein eluting in the main peak was again concentrated to ~2 mg mL^{-1} for vitrification. Cross-linked micrographs, akin to those with haem alone, failed to achieve the quality seen for the initial LMNG dataset (Figure 4.6[B]); and so only a small collection was undertaken (~340 im-

ages). Despite picking parameters being tweaked in order to maximise extracted particle count (~70K), the subsequent 2D-CAVGs remained worse than previous (Figure 4.6[C]). Ironically, side view clarity seemed to suffer most. Clearly motional freedom is not the primary factor preventing betterment of $MmpL3_{1\text{-}753}$ imaging by Cryo-EM, or the methods attempted here are too crude or ineffective to realise any improvement for such a small particle. To see if reconstitution of the truncation into a detergent-free environment might yield benefits, amphipols were trialled next.

4.2.3 Amphipol: A8-35

Amphipols (APols) are a novel class of short amphipathic polymer, rationally designed to adsorb to the hydrophobic surface of integral membrane proteins with high affinity — facilitating their stable solubilisation in absence of free surfactant molecules (Popot et al., 2011). Despite first being described over two decades ago (Tribet et al., 1996), APols — notably A8-35 — are now experiencing somewhat of renaissance as detergent substitutes in single-particle Cryo-EM. Driving this resurgence are documented improvements in protein stability, reduced conformational dynamics, and better in-ice particle contrast and/or distribution following exchange into the polymers (Popot, 2018). These advantages largely derive from the near-irreversibility of protein-APol association, in absence of detergents at least. This is achieved through avidity, multiple alkyl tails on each APol molecule interacting with the TMD of a single target protein (Le Bon et al., 2018). APols therefore allow use of detergent-free buffers. The comparative reduction in surface tension in turn permitting easier attainment of thin ice — ideal for Cryo-EM imaging (Cheng et al., 2015). Finally, unlike bulky detergent coronas, APols have been postulated to form thinner surfactant layers around the proteins they solubilise (Zoonens & Popot, 2014); enhancing the orientability of smaller particles one may speculate. Taken together, $MmpL3_{1\text{-}753}$ reconstitution into A8-35 was thus an attractive prospect.

The best characterised APol by far, A8-35 is formed of a hydrophilic polyacrylate backbone with a proportion of its carboxylates derivatised with the hydrophobes octylamine (~25%) and isopropylamine (~40%). Prior to in-ice screening, the optimal protein : A8-35 mass ratio for reconstitution into homogeneous protein-APol particles must be empirically determined (Zoonens & Popot, 2014). Briefly, ~200 μg of DDM purified truncation was first incubated with a range of A8-35 concentrations (1:0–1:10 protein : APol mass ratio). Newly formed MmpL3$_{1\text{-}753}$-DDM-APol ternary complexes were then depleted of detergent by addition of excess polystyrene Bio-Beads, before monodispersity analysis by SEC in a surfactant-free mobile phase (summarised in Figure 4.7[A]). Absence of protein in the 1:0 (protein : APol) negative control confirmed successful DDM removal, and comparison of the other S200 traces suggested MmpL3$_{1\text{-}753}$ reconstitution into A8-35 was optimal — i.e. most homogeneous — at a mass ratio of ~1:3 (Figures 4.7[B]&[C]). The moderately broad gel filtration profiles generated are consistent with reports of the slightly greater size dispersity of protein-APol systems as compared to their protein-detergent equivalents (Le Bon et al., 2018).

Scaling up APol reconstitution, this time with additive haem, gave a satisfactory SEC peak from which material was concentrated for vitrification (~3 mg mL^{-1}, Figure 4.7[D]). ~1700 micrographs were collected, and appeared more particulate — almost granular — relative to MmpL3$_{1\text{-}753}$ in detergent (Figure 4.7[E]). Processing this sizeable dataset yielded ~300K particles. The resulting 2D-CAVGs were certainly distinct from all preceding detergent samples; displaying reduced micellar bulk and an inversed orientation bias of the best averages — side views now dominating (Figure 4.7[F]). Whilst not splayed, the periplasmic subdomains lacked the detail recovered in the top views of previous RELION outputs. Overall, A8-35 2D-CAVG quality was, despite the numerous literature suggested benefits, surpassed by the LMNG data. Further investigation of APols was thus discontinued. These observations do, however, emphasise the capacity to tune in-ice particle behaviour via modification of sample preparation parameters.

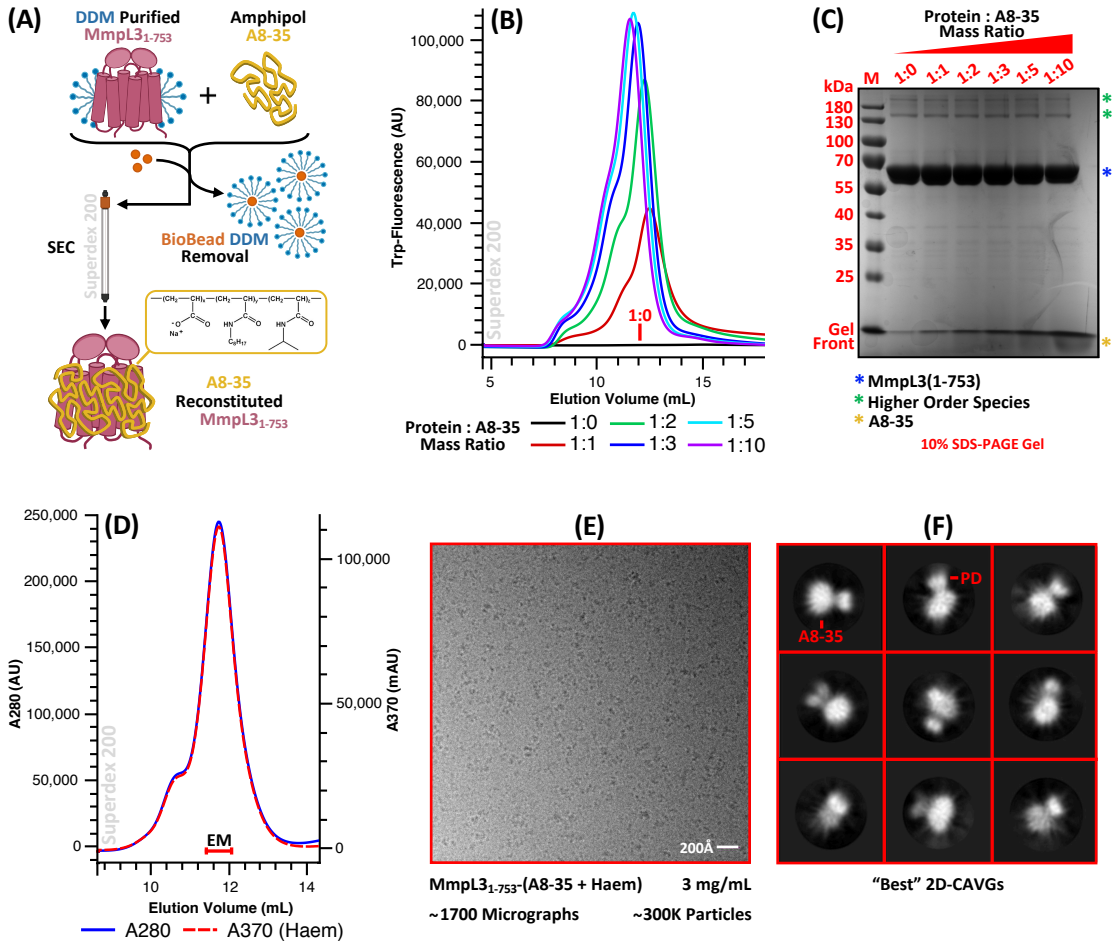


Figure 4.7: Cryo-EM Screening of A8-35 Reconstituted MmpL3$_{1\text{-}753}$ + Haem

(A) Diagrammatic summary of the protocol used to exchange DDM purified MmpL3$_{1\text{-}753}$ into amphipol (APol) A8-35. Chemical structure of the latter is also shown (x~0.35, y~0.25 & z~0.40). **(B)** Overlain S200 SEC traces used to find optimal protein:APol mass ratio for reconstitution. The detergent removal control, without added APol, is labelled. The accompanying SDS-PAGE gel of **(C)** confirms equal protein input for each ratio screened. **(D)** S200 profile of A8-35 reconstituted MmpL3$_{1\text{-}753}$ (1:3 protein:APol mass ratio) post-incubation with haem (2:1 haem:protein molar ratio). In-line absorbance at 370 nm (A370), monitoring haem, is plotted together with absorbance at 280 nm (A280, protein). Red bar corresponds to peak protein concentrated for Cryo-EM. **(E)** Representative Titan Krios micrograph; collection metadata is provided. **(F)** Best RELION (Scheres, 2012) 2D-CAVGs. A8-35 bulk and periplasmic domain (PD) are annotated. Processing courtesy of Justin Deme.

4.2.4 Conclusion

To summarise, initial structural attempts with MmpL3$_{1\text{-}753}$ were pursued by crystallography — first by VD, and soon after by the LCP approach. In either case, the truncation could not be coaxed to crystallise; in spite of screening a reasonable number of sparse matrix precipitant solutions, temperatures, protein concentrations, purification detergents, and/or host lipids. This, in part,

appears to have been a consequence of intolerably rapid protein destabilisation *in surfo* and suboptimal formation of the cubic phase *in meso*. Of course, more conditions can always be screened. In hope of gleaning preliminary structural information, despite reservations concerning particle size, Cryo-EM was trialled in DDM, LMNG, and APol A8-35. As at purification, the MNG class detergent proved most promising. $MmpL3_{1\text{-}753}$ was confirmed to be monomeric, displaying a domain organisation consistent with RND superfamily architecture on 2D-CAVG inspection. Efforts at 3D reconstruction failed, likely impeded by in-ice preferential orientation and/or a degree of PD motional freedom; both only serving to exacerbate the difficulties inherent in imaging an ~82 kDa protein. Cross-linking and use of haem as a structural additive were explored, but yielded no appreciable benefits. Aiming to surmount the apparent intractability of the truncation to structural studies, ancillary technologies — in the form of synthetic nanobodies (Chapter 5) and via reconstitution into the MSP nanodisc system (Chapter 6) — were next sought.

Synthetic Nanobodies

CHAPTER 5

Having so far been unable to obtain a high resolution structure of the protein in isolation, this chapter documents the selection, characterisation, and application of binders targeting MmpL3$_{1\text{-}753}$; aiming to leverage complexation as a means to augment truncation crystallisability and/or the quality of Cryo-EM imaging. After careful consideration of the features desirable in a structural chaperone, *in vitro* generation of synthetic nanobodies, achieved via a selection cascade specifically optimised for membrane proteins, was decided upon. As will be seen, despite recovery of a sizeable cohort of binders, passed through a rigorous bioinformatic and experimental evaluation pipeline, growth of diffracting MmpL3$_{1\text{-}753}$-nanobody complex crystals failed — both *in surfo* and *in meso*. Similarly, the potential value of the raised binders as truncation Cryo-EM fiducials could not be realised, in spite of nanobody association being evident in 2D-CAVGs in one case. Finally, unforeseen implications of detergent choice on binder utility, a consequence of the differential behaviour of MmpL3$_{1\text{-}753}$ when solubilised in DDM versus LMNG, are also presented.

5.1 Nanobodies: Structural Biology Aids

Nanobodies (Nbs) — recombinant single domain binders derived from the heavy chain antibodies (HCAbs) of *Camelidae* (e.g. camels, llamas, alpacas) — are increasingly the affinity reagents of choice for numerous applications in fundamental bioscience, and have simultaneously emerged as a novel biologic class (Arbabi-Ghahroudi, 2017). Initially reported over 25 years ago as unique mammalian immunoglobulin-γ (IgG) variants, HCAbs deviate from the classi-

cal heterotetrameric paring of heavy and light chains (Figure 5.1[A], Hamers-Casterman et al. (1993)). Instead, devoid of the latter, HCAbs are homodimers; each heavy chain protomer also missing the first constant domain (CH1, Figure 5.1[B]). This architecture confines the paratope to a single variable domain (VHH) encoded by only one exon, enabling facile manipulation with molecular biology (Muyldermans, 2013). Heterologous VHH expression yields a Nb (~110 residues, 15 kDa) retaining the antigen-binding properties of its parental HCAb. These are functionally equivalent to, yet a fraction of the size of, Fab fragments (2 Chains, ~440 residues, 55 kDa) obtained from conventional antibodies (Muyldermans, 2013).

Expectedly, Nbs exhibit an Ig-fold formed of two antiparallel β-sheets packed against one another and stabilised by a conserved inter-sheet disulphide (Figure 5.1[C]). The binding determinants of the resulting prolate molecule — 3 sequence hypervariable loops termed complementarity-determining regions (CDRs 1–3) — are clustered to give a continuous N-terminal paratope (Ingram et al., 2018). Unable to benefit from the additional CDR trio of the light chain, reducing exploitable sequence space, Nbs nevertheless achieve cognate antigen binding affinities, and specificities, comparable to standard IgG and its derivatives (i.e. Fab, scFv; Desmyter et al. (2015)). Systematic analyses of Nb-antigen complexes indicates light chain loss has been compensated, primarily, by a diversification in attainable CDR backbone conformations (Sircar et al., 2011; Mitchell & Colwell, 2018a,b). This structural plasticity is most evident in Nb CDR3, which is more sequence divergent, length variable, and on average longer (by ~3–4 residues), than corresponding canonical CDRs (Mitchell & Colwell, 2018b). Likewise, unconstrained by the requirement for light chain dimerisation, CDR3 is free to occupy a wider range of configurations relative to the Nb framework; improving shape-complementarity to the target (Mitchell & Colwell, 2018b). Finally, the proclivity of HCAb CDR3 to form extended finger-like protrusions permits Nbs, in conjunction with their modest size, to probe antigen clefts and cavities (e.g. ac-

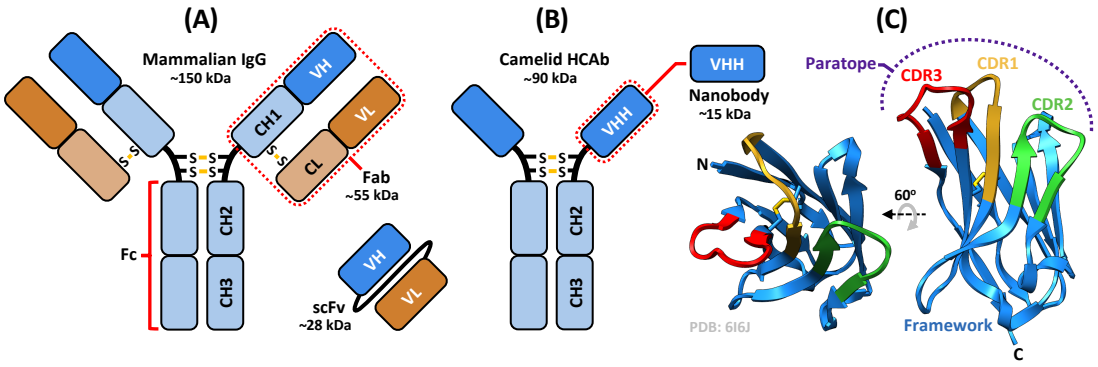


Figure 5.1: IgG, Antibody Fragment, HCAb, and Nb Structural Comparison

(A) Schematic of IgG domain organisation, coloured by chain (heavy (H) = blue, light (L) = brown) and domain variability (constant (C) = lighter, variable (V) = darker). Inter-chain disulphides are depicted. Commonly derived antibody fragments, Fab & scFv, are shown in relation. An equivalent summary of camelid HCAb is given in **(B)**, with Nb derivation highlighted. **(C)** Cartoon representation of typical Nb fold (PDB: 6I6J, Bräuer et al. (2019)). CDRs are colour-coded and the invariant framework blue. Figure layout inspired by Harmansa & Affolter (2018), and structures composed with UCSF Chimera v.1.13.1 (Pettersen et al., 2004).

tive sites) sterically inaccessible to larger binders (Henry & MacKenzie, 2018). Ability to utilise these cryptic epitopes is highly advantageous when pursuing membrane protein interactors, as the majority of the target exterior is often bilayer buried with only minimal hydrophilic outcrops (e.g. GPCRs, Steyaert & Kobilka (2011)) — albeit not the case for $MmpL3_{1\text{-}753}$ given its PD.

More generally, binders — be they Nbs, Fab fragments, DARPins etc. — are traditionally sought as crystallisation chaperones in structural biology. That is, ancillary proteins improving the probability, and ideally quality, of target crystallogenesis (Koide, 2009). On complexation, productive binders augment the hydrophilic surface available for crystal contact formation (Figure 5.2[A]); frequently producing more ordered lattices, diffracting to better resolution (Desmyter et al., 2015). In parallel, and no less important, chaperones can stabilise the target in a specific state; reducing conformational heterogeneity in solution and thus lowering the entropic cost of crystallisation (Lieberman et al., 2011). Such interactors, if selected judiciously, may trap targets into biologically insightful, but otherwise structurally unobtainable, conformations (Figure 5.2[B]; Uchański et al. (2020)). Subsequent to data collection a binder can further facilitate

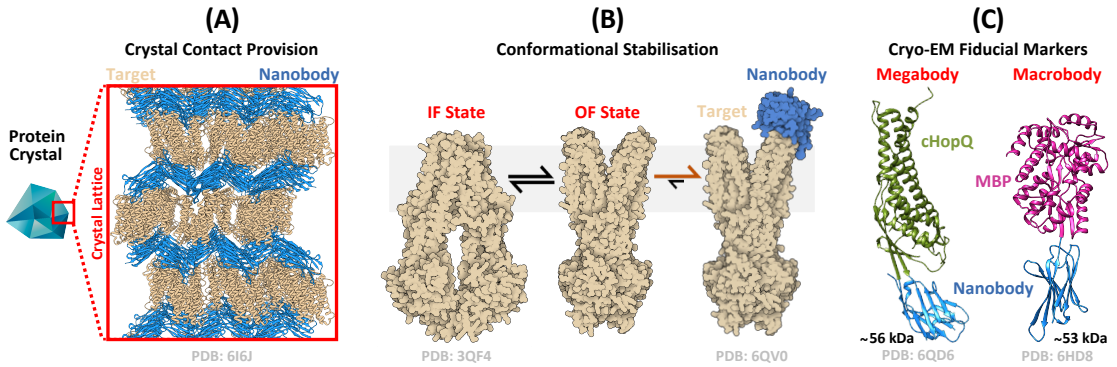


Figure 5.2: Exemplar Applications of Nbs in Structural Biology

(A) *In meso* crystal lattice packing of the KDEL receptor complexed with a synthetic Nb (PDB: 6I6J, Bräuer et al. (2019)). The alternating strata of target and chaperone exemplifies the importance of the latter in crystal contact provision. **(B)** Use of a conformationally discerning synthetic Nb to stabilise, and make structurally tractable, the outward-facing (OF) state of ABC exporter TM287/288 (PDB: 6QV0, Hutter et al. (2019)); which otherwise crystallises as the inward-facing (IF) form (PDB: 3QF4, Hohl et al. (2012)). Bilayer position approximated by the grey box. **(C)** Structures of the Megabody (PDB: 6QD6, Uchański et al. (2021)) and Macrobody (PDB: 6HD8, Brunner et al. (2020)), Nb chimeras engineered as fiducial aids for Cryo-EM. Proteins rendered with Illustrate (Goodsell et al., 2019) or UCSF Chimera v.1.13.1 (Pettersen et al., 2004).

phasing (Koide, 2009), acting as either an initial search model for molecular replacement, or a well characterised chassis in which to engineer heavy-atom sites (e.g. see Hansen et al. (2017)).

Chaperones, in particular Fabs, also find utility in Cryo-EM. Beyond conformational homogenisation, the binder acts a distinct fiducial marker against which to "*in silico* purify" the dataset; aiding image alignment, supplementing particle size, and breaking any target-intrinsic pseudo-symmetry (Wu et al., 2012; Hofmann et al., 2019). This technology has proved fruitful in resolving a number of small membrane proteins, including the serotonin transporter SERT (70 kDa, 3.6 Å, Coleman et al. (2019)), ER-resident glucosyltransferase ALG6 (64 kDa, 3.0 Å, Bloch et al. (2020)), and malarial chloroquine transporter PfCRT (49 kDa, 3.2 Å, Kim et al. (2019)). Although size means they are often overlooked in favour of larger binders, Cryo-EM applications of Nbs will only increase with recently developed chimeras — fusion constructs enlarging single domain antibodies to the approximate mass of a Fab (Figure 5.2[C], Brunner & Schenck (2020)). Providing access to all the above structural benefits — packaged into an easy-to-

handle, thermostable scaffold (T_m~60–80 °C) capable of robust production in *E. coli* (Dumoulin et al., 2009; Desmyter et al., 2015) — generating Nbs against $MmpL3_{1\text{-}753}$ was thus highly desirable. A critical decision remained, to opt for an initial *in vivo*, or entirely *in vitro*, selection campaign.

5.2 Synthetic Nanobodies

Historically, raising Nbs begins with repeated injection of purified target into a camelid, staggered over multiple months. This triggers a sustained immune response in which the HCAb population is expanded and then affinity matured (Pardon et al., 2014). Peripheral blood lymphocytes are next recovered and their mRNA extracted as template for reverse transcription. From the resulting complementary DNA (cDNA), exons containing the VHH repertoire are cloned into display vectors (e.g. phagemids, yeast display plasmids) for downstream enrichment of Nbs possessing the requisite properties (e.g. K_D, specificity, biological activity; Ingram et al. (2018)). Reliant on immunisation, *in vivo* selection suffers notable drawbacks beyond the truisms of expense, inconvenience, and timescale with respect to animal use. The approach is incompatible with delicate targets, being denatured by adjuvants and/or the elevated camelid body temperature (~38 °C), as well as proteins exhibiting suboptimal immunogenicity, or toxicity, within the host (Zimmermann et al., 2020). From a structural perspective, one is unable to select Nbs specific to ligand-induced conformations unless extremely high affinity compounds are to hand (Zimmermann et al., 2018). In contrast, a carefully implemented *in vitro* pipeline overcomes all these issues by affording complete control of the selection environment (e.g. temperature, buffer composition, ligand concentration etc.); enabling Nb pursuit against optimally stabilised targets presented, if so desired, in defined conformations (Bradbury et al., 2011). Mindful of such, the synthetic nanobody (sybody) platform — a 3-week *in vitro* protocol tailored for membrane proteins and pioneered by the Seeger Lab (University of Zürich, Zimmermann et al. (2018, 2020)) — was

chosen for MmpL3$_{1\text{-}753}$ binder selection. It has a proven track record with challenging targets, including the mycobacterial siderophore importer IrtAB (Arnold et al., 2020), the ER-resident retrieval protein KDELR from chicken (Bräuer et al., 2019), and the human neuronal glycine transporter GlyT1 (Shahsavar et al., 2021).

In place of a camelid HCAb response the sybody pipeline employs three extensive mRNA libraries encoding CDR-randomised VHH scaffolds. Purposefully varying in CDR3 length, each expresses a different paratope shape after which the library is aptly named — concave (small), loop (medium), and convex (large, Figure 5.3[A]; Zimmermann et al. (2018)). Together they encapsulate the diversity of binding modes observed in natural Nbs. Zimmermann et al. have likewise put prodigious effort into optimising amino acid frequencies within the sybody CDRs, avoiding inclusion of problematic residues (i.e. proline & cysteine) whilst also tempering hydrophobicity; the latter minimising emergence of "sticky" binders commonly enriched against membrane proteins (Seeger et al., 2012). Designed as a selection cascade, the sybody pipeline pans the three libraries using an enzymatically biotinylated target (Figure 5.3[B]); beginning with an initial ribosome display (Figure 5.3[C]). Interactors surviving this are then subject to two successive rounds of phage display (Figure 5.3[D]), prior to an enzyme-linked immunosorbent assay (ELISA) to identify any sybodies (Figure 5.3[E]) warranting further characterisation (Zimmermann et al., 2018). By varying the surface chemistry of the target immobilisation support used at each stage, as well as the display format, the protocol discourages selection of background binders unspecific to the protein of interest. Moreover, the sizeable starting diversity achievable via ribosome display (~10^{12} unique sequences per library), orders of magnitude greater than transformation-dependent technologies (i.e. phage display), is key to recovering high affinity binders (~500 pM–500 nM K_D) in absence of *in vivo* affinity maturation (Zimmermann et al., 2020).

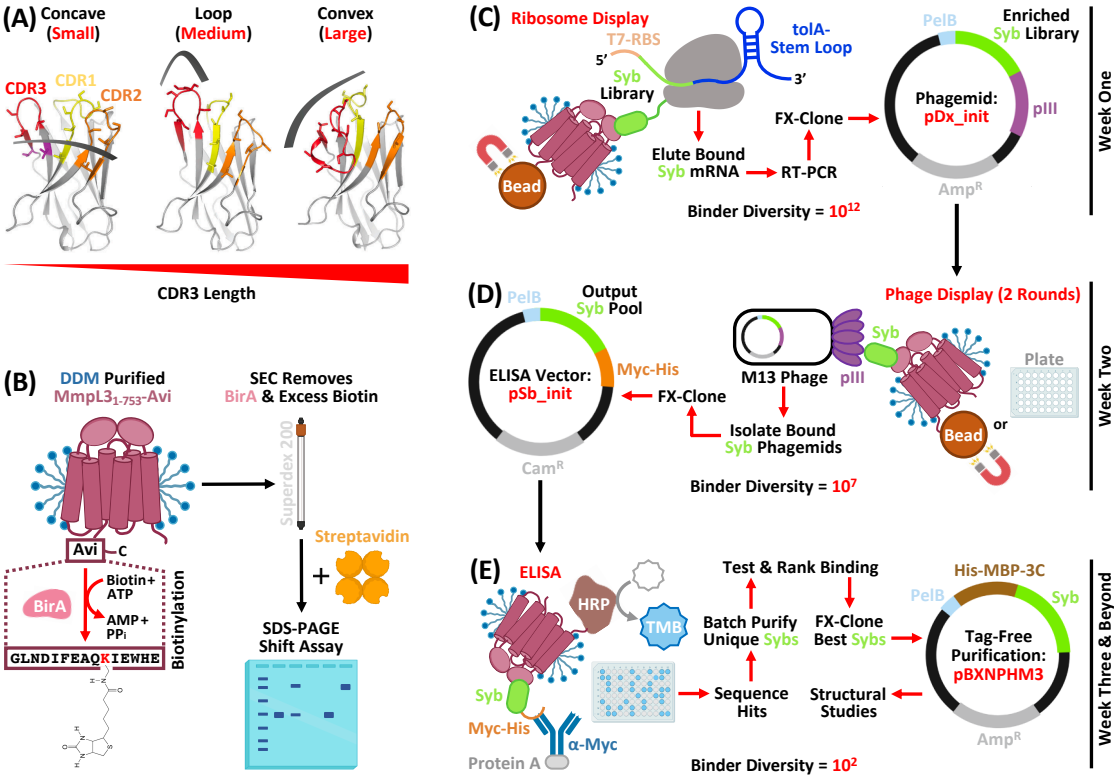


Figure 5.3: An Overview of the Sybody Libraries & Selection Pipeline

(A) Cartoon representation of the sybody library scaffolds. Paratope shape is indicated by a black line, the CDRs are colour-coded, and randomised positions depicted as sticks. LTR, CDR3 length increases from 9 (concave) to 12 (loop) to 16 (convex) residues. Adapted from Zimmermann et al. (2018). **(B)** Graphical summary of the protocol used to enzymatically biotinylate C-terminally Avi-tagged MmpL3$_{1-753}$ with *E. coli* biotin ligase BirA; exploited for avidin-based immobilisation within the pipeline. The extent of target biotinylation is judged by SDS-PAGE streptavidin shift assay. **(C)** – **(E)** Detailed schematic of the sybody selection cascade, spanning ribosome display to the ELISA; approximate diversity at each stage is noted. Ribosome display **(C)**: Ternary complexes of a library mRNA, the corresponding translated sybody, and a stalled ribosome are solution panned against the target, prior to pull-down onto streptavidin-coupled magnetic Dynabeads (T1 variety). Enriched binders are transferred into a phagemid (pDx_init) via reverse transcription PCR (RT-PCR) followed by fragment exchange (FX) cloning. Phage display **(D)**: M13 bacteriophages monovalently presenting sybodies, fused to coat protein pIII, are panned in two rounds. First via immobilisation on a neutravidin-coated MaxiSorp plate, and secondly in solution; again to be adsorbed onto magnetised streptavidin Dynabeads (C1 variety). Recovered interactors are subcloned into pSb_init, adding a C-terminal Myc-His$_6$ tag. ELISA **(E)**: Individual sybody clones are expressed, periplasmically prepped, and then captured within a MaxiSorp plate well pre-lined with α-Myc antibody. Post-target incubation, productive binders are identified by streptavidin-conjugated horseradish peroxidase (HRP) acting on chromogenic substrate 3,3',5,5'-tetramethylbenzidine (TMB). Hits are sequenced and unique sybodies batch purified for downstream characterisation; the most promising candidates being moved into vectors permitting large-scale tag-free production (e.g. pBXNPHM3). Additional details of the sybody selection methodology performed can be found in Zimmermann et al. (2020). Figure partly composed with BioRender.

5.3 Selection Against Truncated MmpL3

At the outset of sybody generation the preferability of $MmpL3_{1\text{-}753}$ purification in LMNG as opposed to DDM, regarding protein quality, had yet to be fully appreciated. Consequently, the truncation passaged through the pipeline was solubilised in the more economical alkyl maltoside; a detergent choice which would go on to have repercussions for binder utility (see Section 5.5.3). As alluded, the platform necessitates target biotinylation for selective immobilisation during the biopanning steps as well as development of the ELISA. This was achieved via introduction of an Avi-tag into the original pWaldo-GFP construct, inserted between the truncation C-terminus and adjoining TEV cleavage site.

Only 15 amino acids in length, the Avi-tag is an engineered minimal substrate of *E. coli* biotin ligase (BirA); the enzyme catalysing biotinylation of its singular lysine residue in an ATP-dependent manner (Figure 5.3[B]; Beckett et al. (2008)). As a highly efficient and site-specific reaction, one can easily attain the equimolar (1:1) protein-to-biotin ratios needed to optimally exploit the femtomolar (K_D ~4×10^{-14} M, Michael Green (1990)) biotin-avidin interaction for target display — assuming Avi-tag fusion, and subsequent biotin conjugation, are tolerated without compromising protein quality (Kuhn et al., 2020). This was evidentially true for $MmpL3_{1\text{-}753}$-Avi-TEV-GFP-His_8, which expressed (~18,000–27,000 RFU mL^{-1} in TB) and purified comparably to the base construct (Figures 5.4[A]&[B]). Following overnight incubation with recombinant BirA, biotinylated truncation was separated from both the ligase and excess biotin via a second gel filtration step; the retained monodispersity of the resulting trace discounting substantial biotin-induced destabilisation of $MmpL3_{1\text{-}753}$-Avi (Figure 5.4[C]). For comparison, a CmpL1-Avi-FLAG construct, cloned with a view to performing sybody selection in parallel with MmpL3, reproducibly exhibited a significantly broader, multi-modal, SEC profile post-biotinylation (Supplementary Figure B.6) and was thus abandoned. Complete $MmpL3_{1\text{-}753}$-Avi biotinylation was confirmed by streptavidin

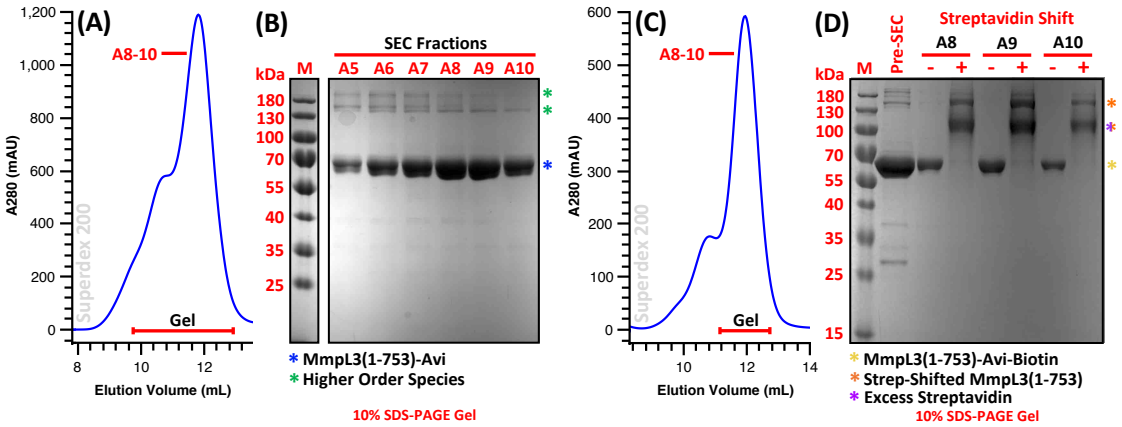


Figure 5.4: MmpL3$_{1\text{-}753}$-Avi-TEV-GFP-His$_8$ DDM Purification & Biotinylation

(A) S200 profile of concentrated MmpL3$_{1\text{-}753}$-Avi, post-cleavage and reverse IMAC, purified in DDM. Red bar indicates material visualised on the SDS-PAGE gel of **(B)**. Labelled peak fractions (A8–10) were pooled, concentrated, BirA-biotinylated overnight, and then re-injected to give the S200 trace presented in **(C)**; the red bar here corresponding to SEC fractions tested in the streptavidin shift assay shown in **(D)**. Adjacent lanes have sample incubated with (+) or without (–) 3 × molar excess streptavidin prior to SDS-PAGE. Asterisks highlight notable species.

shift assay (Figure 5.4[D]), in which the sample is saturated with streptavidin prior to SDS-PAGE analysis. Remaining folded and biotin bound in-gel (Viens et al., 2004), the avidin retards migration of biotinylated species — this mobility shift, relative to a control lane, providing a simple readout to quantify the extent of target biotinylation.

5.3.1 Ribosome Display

Formulated as mRNA, the sybody libraries are stabilised by stem-looped termini and encode binder open reading frames (ORFs) devoid of a stop codon. By design therefore, they assemble into ternary complexes of an mRNA, a stalled ribosome, and the corresponding nascent sybody — displayed from the ribosomal exit tunnel — on incubation with an *in vitro* translation kit (Figure 5.3[C]). These complexes, braced by added magnesium, were panned against biotinylated MmpL3$_{1\text{-}753}$-Avi in solution prior to magnetic pull-down onto streptavidin-coated beads. Notably, solution — as opposed to pre-immobilised surface — panning is favoured throughout the sybody selection cascade as it provides unrestricted epitope access, whilst minimising the target denaturation and avidity phenomena apparent at high

density protein surfaces (Zimmermann et al., 2020; Kuhn et al., 2020). After washing, mRNAs persisting in truncation-bound ternary assemblies were eluted from the beads with EDTA; this focused output library acting as template for a reverse transcription PCR (RT-PCR) bestowing restriction sites required downstream. The integrity of the, now cDNA, libraries was confirmed by quantitative PCR (qPCR) before transfer into phagemid (pDx_init) by fragment exchange (FX) cloning. A reliable one-pot reaction, combining type IIS endonuclease BspQI and counterselection marker *ccdB* (Geertsma & Dutzler, 2011), FX is employed for facile subcloning throughout the sybody vector suite.

5.3.2 Phage Display

Within pDx_init, sybody ORFs are N-terminally flanked by PelB, a leader peptide enforcing the periplasmic secretion necessary for Nb disulphide formation, and C-terminally fused to minor coat protein pIII. The latter enabled overnight production of virions monovalently displaying the binder library (Figure 5.3[D]); induced by helper phage superinfection of exponential phagemid cultures, and subsequently prepped for target screening via the standard precipitation method. For the primary phage display, biotinylated truncation was adsorbed onto wells of a neutravidin-coated MaxiSorp plate following incubation with the library in solution. Once washed, surviving phage-sybody interactors were eluted and used to infect a fresh *E. coli* culture; from which virions for the next round could be amplified. Akin to ribosome display, the second phage panning involved capture onto magnetised streptavidin beads — but was supplemented with an off-rate selection step, post-immobilisation. Competition with fiftyfold excess unbiotinylated $MmpL3_{1\text{-}753}$ aimed to bias recovery towards slower off-rate, and thus higher affinity, binders. To assess enrichment for target-specific sybodies, the libraries from both rounds were additionally screened against a negative control; namely biotinylated AcrB, the stalwart RND transporter, purified in DDM. Contrasting the number of phage, quantified by qPCR, eluted on panning with the truncation

	Phage Display Library Enrichment	
Sybody Library	**Round One**	**Round Two**
Small (Concave)	0.7	389.1
Medium (Loop)	1.4	34.3
Large (Convex)	1.8	415.8

Table 5.1: Sybody Library Enrichment Post-Phage Display

Tabulated enrichment factors of each sybody library for $MmpL3_{1\text{-}753}$ over off-target control AcrB, after one or two rounds of phage display. Enrichment is calculated as the number of phage eluted on selection with the target, divided by the equivalent number of phage recovered on panning with the control protein; measured by qPCR quantification of pDx_init.

relative to panning with AcrB — to which the libraries had had no prior exposure — allowed determination of the enrichment factors presented in Table 5.1. Spanning ~34–415-fold, these established all three libraries as being robustly enriched on completion of phage display. This is likely attributable to the extramembranous domains of $MmpL3_{1\text{-}753}$ providing multiple, readily accessible, hydrophilic epitopes — especially, it would seem, for the binding modes utilised by concave and convex sybodies. For reference, Zimmermann et al. (2020) report stable membrane protein targets to frequently give enrichment factors of >100-fold, yet nevertheless insist enrichments of >5-fold as warranting advancement to ELISA binder identification.

5.3.3 ELISA

From the enriched phagemid output, sybody ORFs were FX subcloned, in bulk, into the arabinose-inducible ELISA vector pSb_init; thereby maintaining their PelB leader, but exchanging C-terminal pIII fusion for a Myc-His_6 affinity tag. Per library (i.e. concave, loop & convex), 95 clones were individually picked, propagated, expressed, and then subject to osmo-enzymatic rupture of the outer membrane, to obtain sybody-laden periplasmic lysate. These extracts, in turn, were applied to discrete ELISA plate wells, pre-lined with α-Myc antibody affixed by Protein A, to capture and present binders (Figure 5.3[E]). Two such 96-well plates, each identical and with a positive control, were prepared per library; one

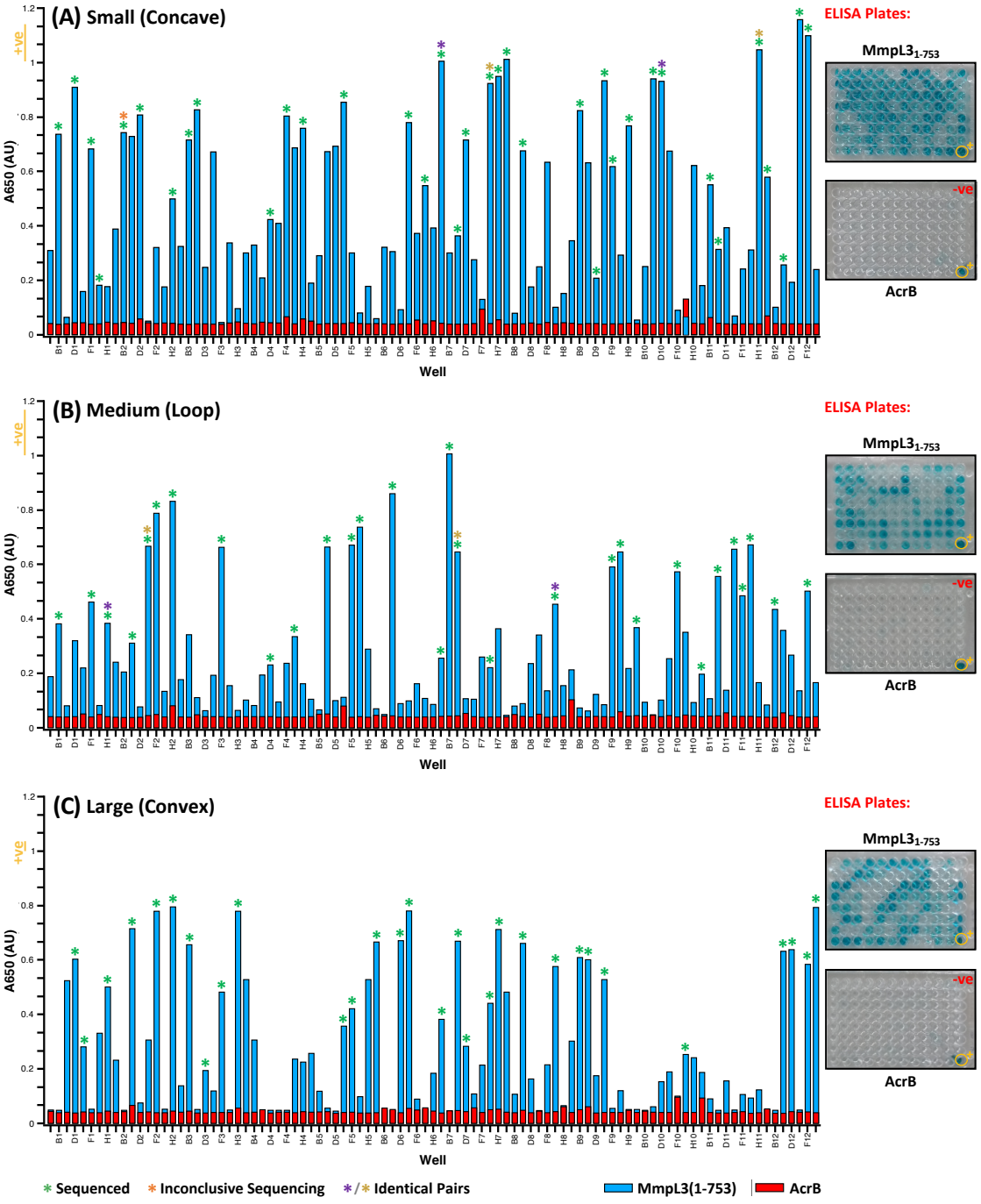


Figure 5.5: ELISA of Sybodies Selected Against MmpL3$_{1\text{-}753}$ in DDM

(A) Overlay of MmpL3$_{1\text{-}753}$ (blue) and AcrB control plate (red) ELISA results, plotted by well coordinate, for sybody clones of the concave library. Individual bars are non-cumulative. Photos of plates post-development are provided for reference. Positive control wells (H12), featuring interaction between ABC exporter TM287/288 and cognate binder Nb_TM#2 (Hutter et al., 2019), are circled in orange; their corresponding ELISA signals also annotated on the absorbance 650 nm (A650, AU) axis. Asterisks denote sybody hits selected for sequencing. Equivalent data for the loop and convex libraries is presented in **(B)** & **(C)** respectively.

incubated with biotinylated truncation and the other, controlling for off-target interactors, biotinylated AcrB. After extensive washing, wells retaining bound $MmpL3_{1\text{-}753}$, or AcrB, were visualised by action of streptavidin-conjugated HRP on chromogenic substrate TMB. This identified a multitude of target specific hits for all three libraries — defined as a sybody giving >1.5 × the ELISA signal for the truncation relative to AcrB background.

In order of abundance, and thus incongruent with the phage display enrichment scores, the concave library was ~92% (Figure 5.5[A]), the loop ~88% (Figure 5.5[B]), and the convex ~68% (Figure 5.5[C]) ELISA positive. Quickly evident by visual inspection alone, concave hits were not only more numerous but, on average, gave higher ELISA signals as compared to the other libraries. Although prompting speculation that shorter CDR3s may be most complementary to the $MmpL3_{1\text{-}753}$ exterior, Zimmermann et al. (2020) caution that sybody affinity is only poorly correlated with ELISA readout. Consequently, the subset of hits chosen for sequencing — totalling 96 across the three libraries (annotated in Figure 5.5) — encompassed wells exhibiting disparate signal intensities. Discarding for duplicates, as well as inconclusive sequence reads, left 91 binders; 33 concave, 28 loop, and 30 convex sybodies, all carried forward to small-scale batch purification (see Section 5.4).

5.3.4 Sybody Sequence Analysis

In advance of considering downstream characterisation of the ELISA output, it is pertinent to highlight insights obtained by simply analysing the CDR sequence diversity recovered within, and between, hits from each library (Figures 5.6 & 5.7). Whilst a Nb paratope may utilise the entire CDR trio, and even some framework positions, structural meta-analyses confirm CDR3 to be the major specificity determinant; the loop universally contributing residues to the binding interface, and frequently participating in over half of Nb-antigen pairwise interactions (Mitchell & Colwell, 2018a). As such, binder families — sequence similar Nbs presumed

to target the same epitope — are commonly inferred by CDR3 homology (Pardon et al., 2014); a convention maintained in the ordering of sequences presented in the following alignments. From here onwards, individual sybodies are referred to by their ELISA well coordinate prepended by a numeral denoting the parent library (4 - concave, 5 - loop & 6 - convex).

Concave Sybodies

CDR analysis of small library hits was perhaps most illuminating, indicating ~60% of sybodies (20/33, sequence block 4H2–4D7 inclusive, Figure 5.6[A]) could be classified into a single family, typified by recurring motifs in all three CDRs. This emphatically suggests the majority of concave binders were selected against a single, immunodominant, $MmpL3_{1\text{-}753}$ epitope — consistent with the aforenoted favourable phage enrichment and ELISA results. Starting with CDR1, all but one family member lack randomisation; that is, unexpectedly, they still possess the SSST-T sequence of the concave progenitor — the sybody template used for PCR-based library assembly (Zimmermann et al., 2018). One may speculate basal CDR1 is thus serendipitously pre-optimised against the truncation, and is therefore selectively maintained. However, it seems at least equally probable non-randomisation signals absence of selection pressure, implying CDR1 does not contribute to the family paratope. Although neither argument is wholly satisfactory, it is salient to note the binder inspiring concave library design, an anti-GFP Nb (PDB: 3K1K, Kirchhofer et al. (2010)), does partially engage CDR1 in antigen recognition. Even so, nearly a third of reported Nb-antigen structures lack any CDR1- or CDR2-mediated contacts (Mitchell & Colwell, 2018a).

The "4H2–4D7" family CDR2 is the least conserved of the loops, but nevertheless displays a striking predilection for aromatic amino acids (AAAs) at permuted positions 3 (19/20) and 10 (20/20). Tyrosine dominates the former (12/20), while the latter is invariably tryptophan — such stringency pointing to an essential function in $MmpL3_{1\text{-}753}$ association. Concordantly, both AAAs are known to be

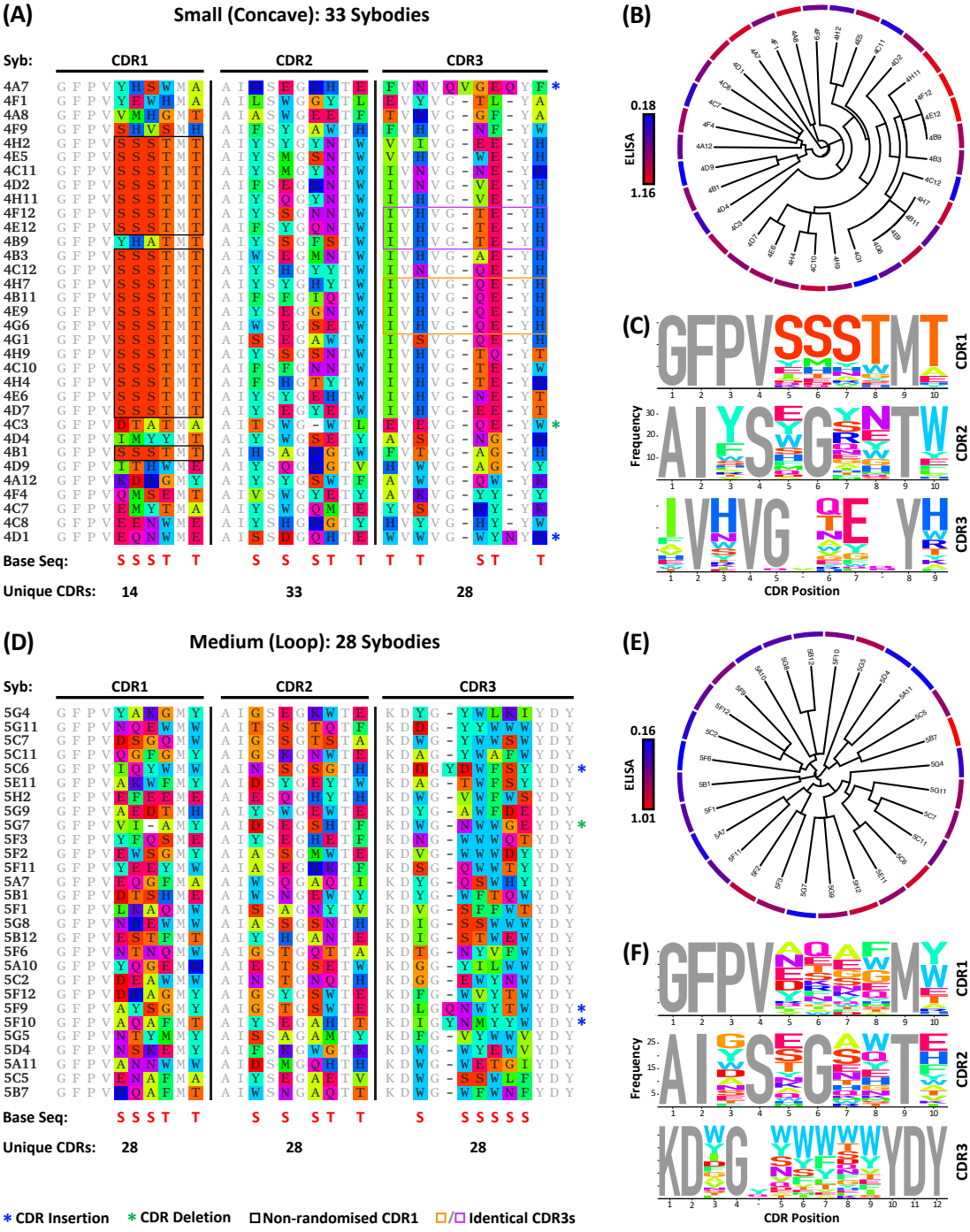


Figure 5.6: Concave & Loop Sybody CDR Sequence Analysis

(A) CDR sequence alignment of the 33 concave sybodies, ordered by CDR3 similarity. Invariable residues are grey, and randomised positions coloured by the Taylor scheme (Taylor, 1997); shading amino acids by size and polarity. Boxes annotate sybodies sharing identical and/or non-randomised CDRs, and asterisks CDRs containing indels. **(B)** Phylogenetic tree of small sybody CDR3s, colour coded by ELISA signal intensity at the perimeter. **(C)** Concave CDR sequence logos, coloured as for the alignment and numbered excluding insertional gaps. Equivalent CDR analysis for the 28 loop sybodies is given in **(D)**–**(F)**. CDR boundaries are defined as in Zimmermann et al. (2018). Alignments calculated with T-Coffee (Madeira et al., 2019), manually adjusted in Jalview (Clamp et al., 2004), and typeset within TEXshade (Beitz, 2000b). Radial phylograms generated with Evolview (He et al., 2016), and sequence logos with Logomaker (Tareen & Kinney, 2020). For full-length sequence alignments see Figure A.5.

important determinants of protein-protein interface energetics (Bogan & Thorn, 1998). In particular, the physicochemical properties of tyrosine — amphipathic, aromatic, bulky, and rigid — allow it to form non-polar, hydrogen bond, and cation-π interactions; a versatility explaining its natural enrichment in antibody CDRs (Koide & Sidhu, 2009), as well as success of minimalist synthetic Fab libraries based on a binary Tyr-Ser code (Fellouse et al., 2005).

Ultimately, the family is defined by its minimal deviation from CDR3 consensus I-H--Q/T-E-H, with selection strongest for the isoleucine (18/20) and glutamate (19/20) residues. The only familially tolerated alternatives at these sites are the structural analogues valine and glutamine respectively. The two canonical choices for randomised position six, Q or T, in turn delineate CDR3-identical subgroups. Of these, the theronine sporting cluster, 4F12–4B9, includes 4E12 — the sybody generating the highest ELISA signal of any hit, regardless of parent library. The remaining 13 concave sybodies fall into a number of ancillary, sparsely populated, and unrelated families; evident on CDR3 phylogrammatic analysis (Figure 5.6[B]), and reinforced by dominance of the principal family in CDR sequence logo plots (Figure 5.6[C]).

Loop Sybodies

In contrast to the concave library, the 28 loop sequences (Figure 5.6[D]) neither cluster into a predominant CDR3 lineage (Figure 5.6[E]) nor retain non-randomised CDR1s. In fact, not even a single pair of medium sybodies share identical sequences in any of the three CDRs — possibly predictable given the only modest enrichment achieved after phage display. Such diversity hopefully indicates hits target an array of different $MmpL3_{1\text{-}753}$ epitopes, rather than the loop binding mode being generally ill-suited to the truncation. Overridingly apparent, especially in otherwise noisy CDR sequence logos (Figure 5.6[F]), is the prevalence of AAAs, most frequently tryptophan, at all six CDR3 variable positions. This suggests hydrophobe burial is a common mechanism of inter-

action within the cohort, despite the presumed epitope mix, and is accordant with aromatic biases noted in paratopes of camelid derived Nbs (Zavrtanik et al., 2018). Most disconcerting are the subset of binders, roughly one-third (9/28), featuring sequence contiguous mulit-tryptophan motifs (WW or WWW) in CDR3; previously demonstrated to drive non-specificity in synthetic scFv libraries (Kelly et al., 2018). Unable to benefit from the protein quality control intrinsic to *in vivo* selection, ~20–60% of recovered sybodies are expected to possess similarly undesirable biophysical properties (Zimmermann et al., 2020).

Convex Sybodies

The bulk of the convex hits, analogous to the small library, can be assigned to one family (22/30, sequence block 6E6–6F12 inclusive, Figure 5.7[A]), described by both CDR1 non-randomisation (15/22) and a degenerate CDR3 consensus of the form -XS-WXY--XNAW-E-; "X" representing sites at which no single residue accounts ≥50% of familial diversity. Within this consensus, ubiquity of the tryptophans at positions 5 (22/22) and 13 (21/22) supports them playing a prominent role in the *modus operandi* of binding. Comparatively, displaying little in the way of positional conservation, the importance of CDR2 within the grouping remains ambiguous. Widespread preservation of basal CDR1 again brings into question its contributions. However, the elongated CDR3 of the large library makes the concept of CDR1 paratope exclusion more palatable. Clear from the CDR3 phylogram (Figure 5.7[B]), at least 5 of the residual 8 convex binders additionally cluster into a secondary set (sequence block 6F5–6G7 inclusive) founded on a -XN-YXX--WNAW-X- consensus. In sharing with the primary family both an AAA and the NAW motif at CDR3 consensual positions 5 & 11–13 respectively (Figure 5.7[C]), they conceivably represent a subsidiary of said family, and thus a staggering 90% (27/30) of convex binders can be hypothesised as contacting a single immunodominant epitope. One of course cannot delineate, from sequence data alone, whether this overlaps with the $MmpL3_{1\text{-}753}$ site speculatively targeted by most concave sybodies.

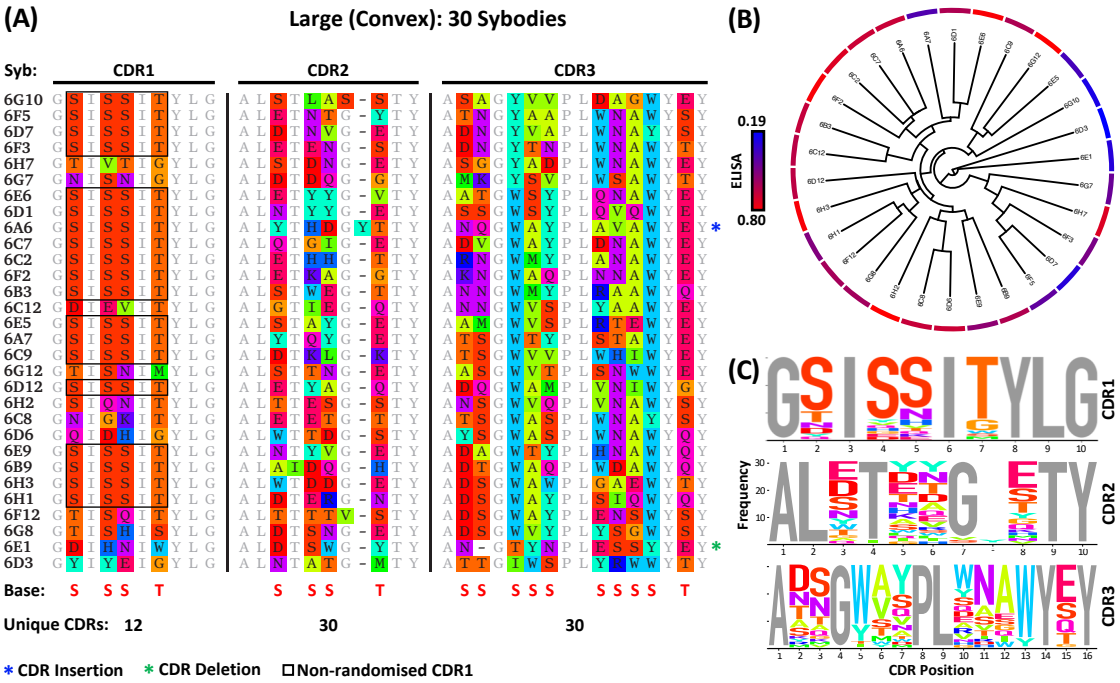


Figure 5.7: Convex Sybody CDR Sequence Analysis

(A) CDR sequence alignment of the 30 convex sybodies, clustered by homology of CDR3 residues. **(B)** Phylogram of large sybody CDR3s; branch termini colour indicate associated ELISA signals. **(C)** Sequence logos for convex sybody CDRs 1–3. Figure composed, coloured, and annotated as for Figure 5.6. A full-length convex sybody sequence alignment is shown in Figure A.6.

To summarise, CDR analysis of the 91 binders confirmed logical correlation between higher phage display enrichment — as exemplified by the concave and convex libraries — and a more restricted sequence diversity; the latter facilitating easier elucidation of familial relationships. Accordingly, the collective 63 small and large sybodies appear strongly selected against only a handful of epitopes, whereas the 28 medium hits feasibly target considerably more. Unanticipated, and still moderately perplexing, was the significant retainment of non-randomised CDR1s in the largest concave and convex families. Finally, the presence of tryptophan at one or more CDR2 or CDR3 positions, in all three libraries, implies the residue is often pivotal to sybody-truncation interactions. Keeping this sequence-gleaned information in mind, efforts to purify, experimentally validate, biophysically characterise, and then rank the sizeable sybody cohort — by far the most demanding phase of $MmpL3_{1\text{-}753}$ binder selection — are next presented.

5.4 Screening Against Truncated MmpL3

5.4.1 Sybody Batch Purification

The 91 sequence-unique candidates, still within pSb_init, were expressed small-scale in 50 mL TB overnight, periplasmically prepped, and then subject to high-throughput purification using a 96-well IMAC filter plate. This format permitted sample processing in batch, as well as convenient centrifugal recovery of pure sybody eluate into a second collection plate. From here binders were analysed by gel filtration, being sequentially injected onto a Sepax SRT-10C SEC-100 column with aid of an auto-sampler. Examination of the resultant traces indicated 83 sybodies — 32/33 concave (Figure 5.8[A]), 23/28 loop (Figure 5.8[B]), and 28/30 convex members (Figure 5.8[C]) — could be successfully produced. Excluding a minority of medium library hits, all gave monodisperse peaks free from substantial aggregates or oligomerisation. Only two, 4H7 & 4H12, displayed the delayed elution volumes symptomatic of adventitious column interactors. Pooled fractions of interest, realising sybody concentrations of ~1–38 μM, were kept as input material for downstream truncation binding screens.

Reappraising sequences of the 8 non-expressing and 2 late eluting sybodies identified, in some cases, individual traits or recurrent patterns hinting at plausible rationales for their aberrant behaviours. For example, the leucine-to-proline framework mutation within 4F9 (L15P, see Sequence Alignment Figure A.5) could conceivably be destabilising, and thus responsible for its poor expression. Equally, three of the five abandoned loop binders — 5G5, 5G8 & 5G11 — are the only library members to feature contiguous tryptophan doublets, or triplets, spanning CDR3 variable positions 7-8, suggesting situational intolerance of the motif. Lastly, due to the significant similarity between 4H7 & 4C12 one can cautiously infer sequence correlates of delayed Sepax elution; namely a tyrosine-rich CDR2 containing residues with cyclic sidechains at all permuted sites.

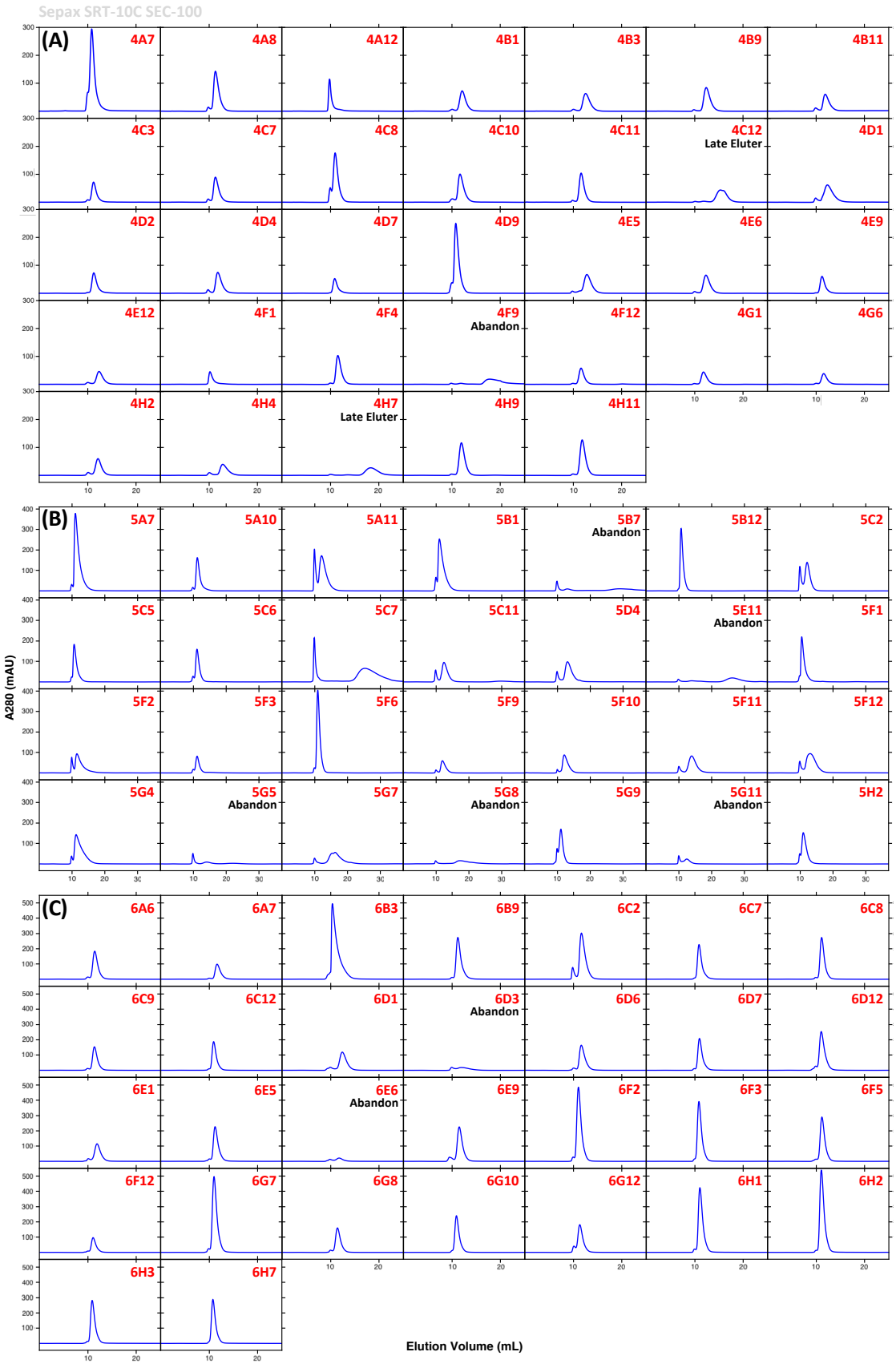


Figure 5.8: SEC Traces of Small-Scale Batch Purified Sybodies

(A) Shared axis plot of concave sybody gel chromatographs; obtained by sequential injection onto a Sepax SRT-10C SEC-100 column, post-IMAC batch purification. Profiles are labelled by binder alphanumeric identifier and, where relevant, delayed elution. Non-expressers were abandoned. Equivalent data for the loop and convex libraries is presented in **(B)** & **(C)** respectively.

5.4.2 Sybody Pull-Down & SEC Co-Elution Screens

The 83 purifiable hits next needed to be whittled down to those sybodies best suited to structural studies — a fundamental requirement being retained truncation binding outside the heterogeneous context of an ELISA. Initial attempts to validate interaction were by small-scale affinity pull-down. $MmpL3_{1\text{-}753}$ was mixed with 1.5-fold molar excess sybody, prior to nickel resin capture of the latter via its Myc-His_6 terminus. If interacting well, binder adsorption simultaneously immobilises untagged target; the presence, and amount, of which are easily readout by SDS-PAGE of bead eluate — superior library members recovering more truncation. Whilst this proved effective for a test panel of five sybodies (Figure 5.9[A]), with improved discrimination attainable by increasing salt concentration, it lacked the throughput and automation desired to process the entire cohort. For these reasons, screening by SEC co-elution was opted for. Samples were prepared identically to the pull-downs, but instead subject to analytical gel filtration. Essential to the assay is an ability to resolve target-complexed sybody from otherwise free, unbound, binder. Although evidentially not achieved in preliminary trials employing a TSKgel G3000SWXL column — from which sybodies eluted as broad protracted traces suggestive of column interaction (Figure 5.9[B]) — this was subsequently realised using a Sepax SRT-C SEC-300 column (Figure 5.9[C]). Having molecular weights too small to reliably induce $MmpL3_{1\text{-}753}$ profile shifts, binders forming SEC-stable complexes were identified by their detection in truncation peak fractions; evaluated by silver staining. An example of a co-elution negative sybody, 6E5, is given in Figures 5.9[D]&[E].

In total, ~90% of candidates progressing from batch purification were screened by SEC co-elution; amounting to 24/32 concave, 22/23 loop, and all 28 convex sybodies (Figure 5.10[A], Supplementary Figure B.8). The vast majority (63/74) formed, to varying extents, SEC-stable complexes with $MmpL3_{1\text{-}753}$, and non-binders predominantly belonged to the large scaffold. Co-elution positive hits were additionally differentiated through a three-tiered qualitative scoring system

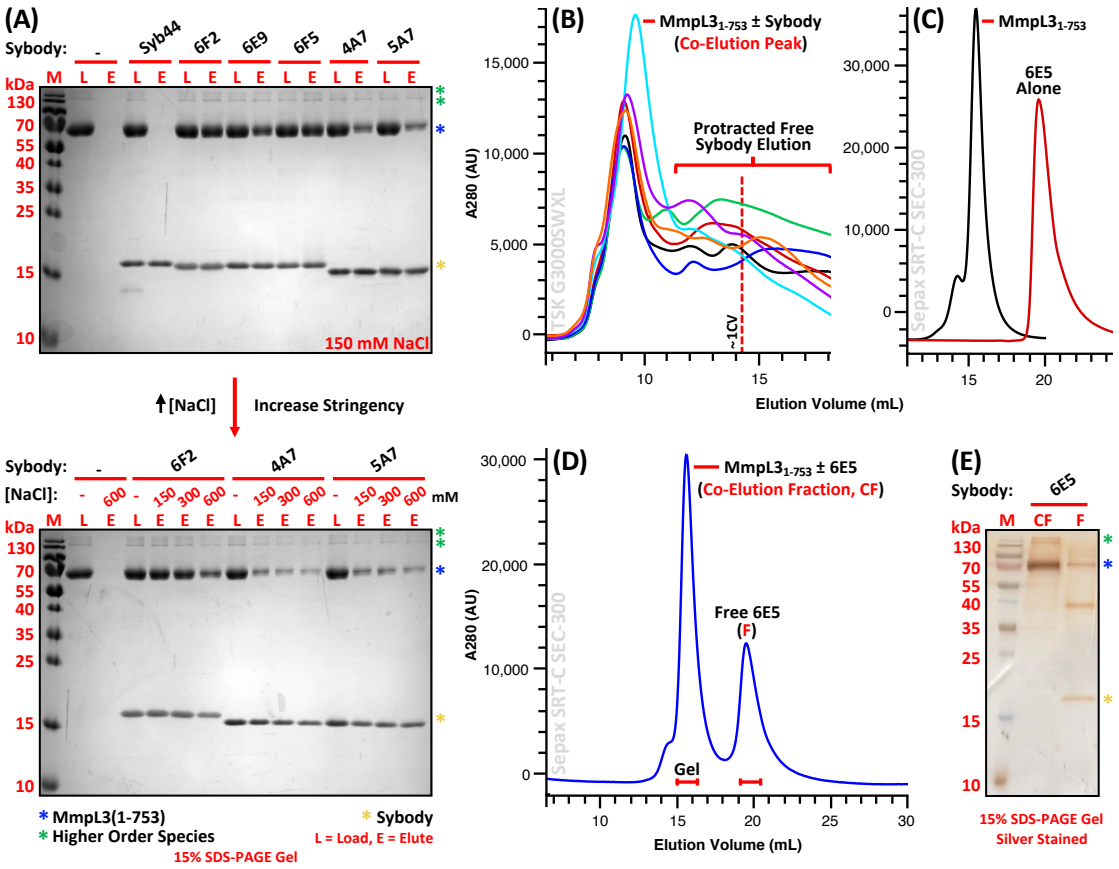


Figure 5.9: MmpL3$_{1\text{-}753}$-Sybody Pull-Down & Co-Elution Screen Optimisation

(A) SDS-PAGE gel reporting nickel resin pull-down of MmpL3$_{1\text{-}753}$ by a panel of Myc-His$_6$-tagged sybodies. Adjacent lanes show sample composition pre-bead loading (L) and post-bead elution (E). No sybody (–) and non-specific sybody (Syb44, selected against the KDEL receptor by Bräuer et al. (2019)) binding controls are included. Results from pull-down repetition at elevated salt concentrations are presented in the lower panel. **(B)** Traces obtained from truncation co-elution trials with six convex sybodies, run on a TSKgel G3000SWXL column. **(C)** Overlay of chromatographs for MmpL3$_{1\text{-}753}$ and sybody 6E5, individually injected onto a Sepax SRT-C SEC-300 column. The Sepax co-elution trial of both is given in **(D)**; red bars labelling fractions visualised on the silver stained SDS-PAGE gel of **(E)**, which confirms 6E5 does not form a SEC-stable complex with MmpL3$_{1\text{-}753}$. For all screens, pull-down or co-elution, ~20–40 μg of truncation was incubated with binder in 1:1.5 molar ratio prior to assaying complexation.

— present, good or robust — based on sybody band intensity (Figure 5.10[B]). Tellingly, the 23 "robust" binders — those recovered most abundantly within the truncation peak, and possibly at stoichiometric levels — populate only the small (14/23) and medium (9/23) libraries. Although the former are principally sybodies of the main concave family, including the ELISA adept 4E12, they also feature two familially peripheral binders, 4C3 & 4D1. This indicates intra-library selection of strong interactors against multiple epitopes, despite the apparent im-

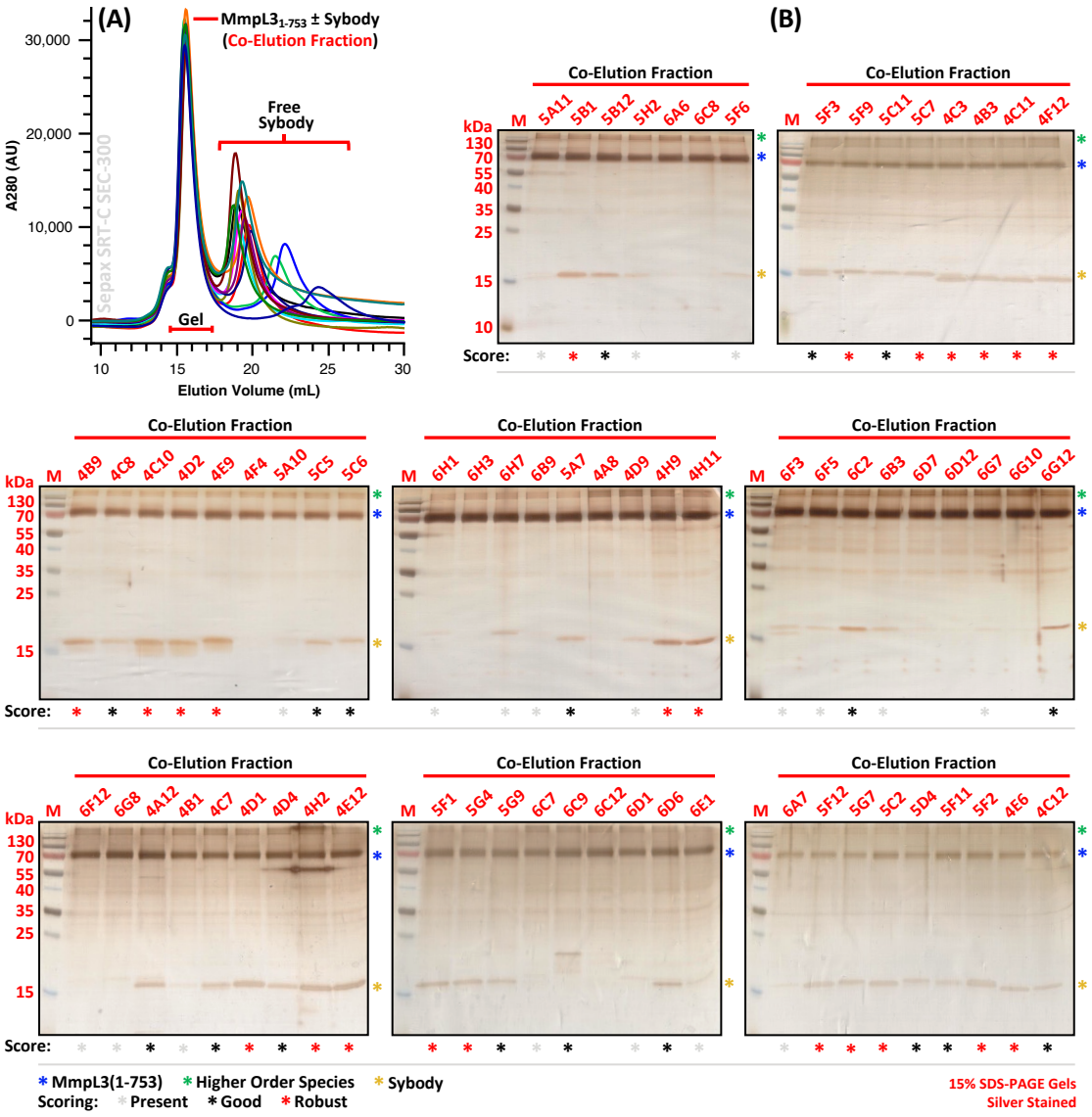


Figure 5.10: Collated MmpL3$_{1-753}$-Sybody SEC Co-Elution Results

(A) Exemplar overlay of 14 MmpL3$_{1-753}$-sybody co-elution traces, generated by consecutive injection onto a Sepax SRT-C SEC-300 column. Labels demarcate a peak containing the truncation-binder complex, or MmpL3$_{1-753}$ alone if this is not SEC-stable, and the free sybody peak. Material from the former, denoted by the red bar, for 69 such co-elution trials is analysed on the 8 silver stained SDS-PAGE gels of **(B)**. Asterisks annotate protein species or report a qualitative co-elution score per lane; based on sybody band presence and intensity. For individual co-elution profiles see Supplementary Figure B.8.

munodominance of one. In remaining firmly bound throughout gel filtration such "robust" hits are best suited to structural applications, likely permitting homogeneous purification of truncation-sybody complexes in which neither component is in excess. From experience, Zimmermann et al. (2020) note competent SEC co-eluters can be predicted to have a K_D<500 nM.

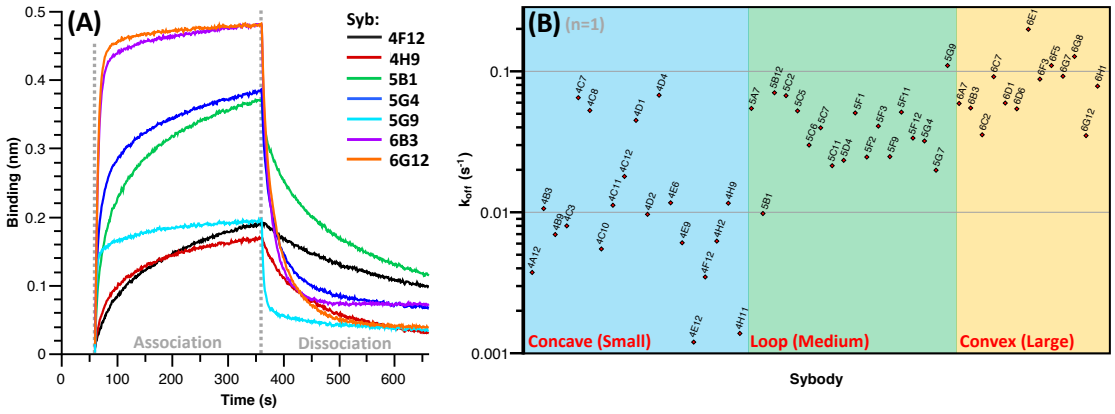


Figure 5.11: BLI Sybody Off-Rate Ranking

(A) Example overlay of Octet-RED384 BLI sensorgrams recording interaction of 7 sybodies with biotinylated $MmpL3_{1\text{-}753}$-Avi, the latter pre-loaded onto streptavidin-conjugated biosensor dipsticks. Dotted lines delimit the 300 second association and dissociation phases. **(B)** Scatter plot summary of BLI determined off-rates for 50 sybodies, grouped by library and plotted alphanumerically from LTR. For individual BLI traces see Supplementary Figure B.9.

5.4.3 Sybody Off-Rate Ranking

To further distinguish binders, 50/63 co-eluting sybodies — including all those rated "robust" — underwent kinetic analysis; specifically high-throughput off-rate (k_{off}) estimation, measured by Bio-Layer Interferometry (BLI) and performed on a Octet-RED384 instrument. Briefly, a layer of biotinylated $MmpL3_{1\text{-}753}$-Avi was first immobilised onto the tips of streptavidin-conjugated, fibre-optic, dipstick biosensors. By interpreting interference pattern shifts in reflected light incident to said protein layer, BLI reports changes in its optical thickness as a function of experimental time (Concepcion et al., 2009). Thus, by sequentially immersing truncation-loaded sensors into sybody, followed by buffer-only, microtiter plate wells, one records the real-time evolution of bio-layer thickness as binders adsorb, and subsequently desorb, from $MmpL3_{1\text{-}753}$. Plotting these time courses yields BLI characteristic biphasic sensorgrams (Figure 5.11[A]); from which truncation-sybody association and dissociation rates can be derived on curve fitting. The latter, reporting fractional complex decay per second, is the more convenient kinetic differentiator as it is a concentration-independent correlate of affinity.

Considered together, the 50 k_{off} values identify multiple trends within, and between, the sybody libraries (Figure 5.11[B]); mainly consistent with the SEC co-elution results, but offering additional quantitative nuance. The 19 concave binders displayed the broadest off-rate distribution, with the 10 slowest boasting smaller k_{off} estimates then any loop or convex sybody screened. Interestingly, of these top-ranked concave candidates, 80% belong to the dominant family; the best — achieving a dissociation rate of ~1.2×10^{-3} s^{-1} — being the CDR3 consensual, premier ELISA hit, and robust co-eluter, 4E12. In contrast, the 31 assayed medium and large sybodies cluster in a noticeably narrower, less favourable, higher k_{off} range. None, with the exception of loop binder 5B1 (ranked 11^{th} overall), attained off-rates surpassing 1×10^{-2} s^{-1}; a subset, mostly the poorest convex binders, even failing to better a k_{off} of 1×10^{-1} s^{-1}. To generalise, sybody biophysical suitability appears to deteriorate as one moves from the concave, to the loop, and finally to the convex library hits; and thus, with CDR3 extension.

At this juncture it was decided the cohort had been sufficiently characterised — through batch purification, SEC co-elution, and BLI screening (data is collated, per sybody, in Supplementary Table B.1) — as to allow rational choice of binders progressing to crystallisation, or Cryo-EM trials, in complex with $MmpL3_{1\text{-}753}$. The only lamentable omission, epitope binning, may have usefully delineated mutually exclusive sybody combinations; in doing so, likely refining the CDR-based familial classification system, and related sequence speculations. Efforts towards binding site definition were pursued however, with design and cloning of four MmpL3 epitope binning constructs (summarised in Supplementary Figure B.10), destined for use in pull-downs. These aimed, by individual deletion or subcloning, to dissect the epitope contribution of each periplasmic subdomain. Unfortunately, for reasons only fully appreciable later (see Chapter 7), the truncation likely cannot tolerate singular periplasmic loop removal due to the interdependent folding of both extramembranous lobes. Consequently, only able to express and purify the second periplasmic subdomain in isolation, epitope binning attempts were halted.

5.5 Truncation-Sybody Complex Structural Trials

5.5.1 Tag-Free Sybody Purification

Sybody structural endeavours aimed to individually complex $MmpL3_{1\text{-}753}$ with nine binders — three from each library (4C3/C10/E12, 5B1/F1/F12 & 6B3/D6/F2). Where successful, complexes would be screened for inducement of crystallogenesis, *in surfo* or *in meso*, and/or for improvements in imaging by Cryo-EM. To obtain ample structure-grade material, the candidates were first FX subcloned from pSb_init into the purification vector pBXNPHM3. The latter, still arabinose-inducible and retaining the PelB leader, exchanges the C-terminal Myc-His_6 tag of the former for an N-terminal, and 3C protease cleavable, His_{10}-MBP fusion. This not only facilitates generation of binders lacking a flexible affinity tag, ideal for crystallisation, but Zimmermann et al. (2020) posit MBP addition also leads to more consistent expression between clones. The number of crystallisation or vitrification trials realised for each of these sybodies, selected primarily with respect to Section 5.4 and hence including six robust co-eluters, of which two are the top BLI-ranked small (4E12) and medium (5B1) binders, was necessarily tempered by the timing of pBXNPHM3 subcloning success as well as factors only apparent on purification scale up.

Large-scale purification — proceeding successively through IMAC, overnight dialysis with proteolytic tag removal, reverse IMAC (Figure 5.12[A]), and finally SEC polishing (Figures 5.12[B]&[C]) — gave homogeneous protein for all nine binders (Figure 5.12[D]). For many, however, the gel filtration trace was bimodal (e.g. 4E12, 5B1 & 5F12). Sybody in the first peak co-eluting with various contaminants, on occasion the liberated MBP tag, and thus limiting usable material to fractions of the second. Nevertheless, from only 3–4 L of TB ~1–8 mg of pure tag-free binder could be invariably recovered, more than sufficient to begin complexation for structural studies. Among sybodies at the lower yielding end of this range are 5F1 and 6B3, a direct consequence of their repeated precipitation

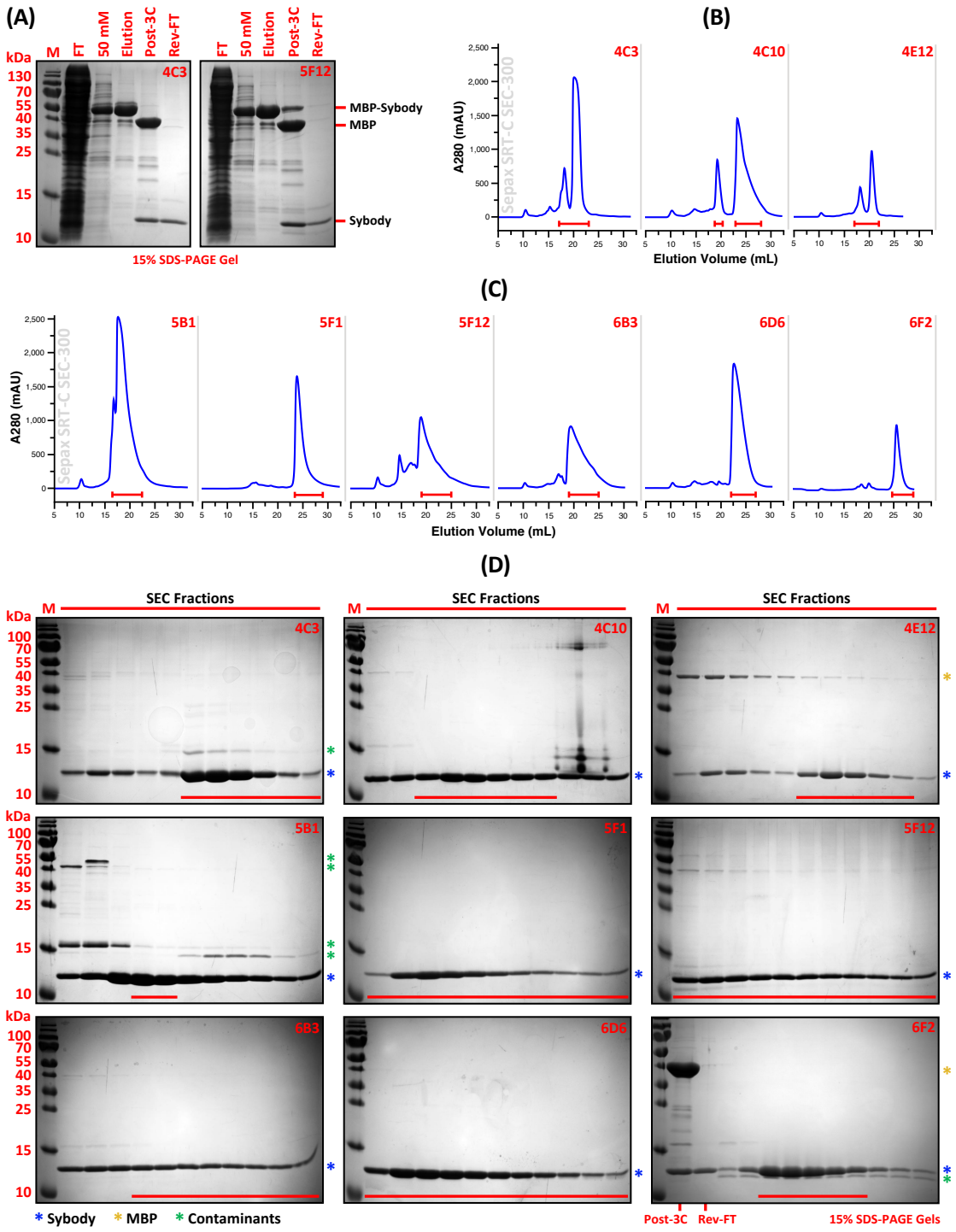


Figure 5.12: Large-Scale Purification of Tagless Sybodies

(A) SDS-PAGE gels summarising purification of His_{10}-MBP-3C fusions of 4C3 & 5F12; spanning nickel resin flow-through, protease cleavage, and reverse IMAC (Rev-FT). Results are representative of all 9 binders scaled up, the Sepax SRT-C SEC-300 traces for which are presented in **(B)** & **(C)**; annotated by sybody alphanumeric designator. Red bars identify peak material visualised on the correspondingly labelled SDS-PAGE gels shown in **(D)**. Fractions of interest, pooled and concentrated for structural studies, are highlighted by the red line.

on concentration; both prior to gel filtration and again on SEC fraction pooling. That similar behaviour was absent during small-scale batch purification is undoubtedly a result of the lower protein concentrations achieved, thereby not surpassing the suboptimal solubility thresholds of 5F1 and 6B3 — in turn clearly discouraging their use in crystallography. In advancing so far along the sybody pipeline, yet still being biophysically inadequate, 5F1 and 6B3 are another pertinent reminder of the *in vivo* quality control benefits afforded to conventionally raised Nbs.

5.5.2 Crystallisation Trials

For crystallisation, MmpL3$_{1\text{-}753}$-sybody samples were assembled in DDM at 1:2 molar ratio, incubated on ice to allow complex formation, and then excess tag-free sybody separated by gel filtration (Figure 5.13[A]). Central peak fractions displaying the most equi-stoichiometric ratio of MmpL3$_{1\text{-}753}$ to sybody, although rarely convincingly 1:1 (Figures 5.13[B]&[C]), were then pooled and concentrated for either VD or LCP plate setup; executed as detailed in Section 4.1. In total five such complexes, including at least one sybody from each library (4C3/C10, 5B1/F12 & 6D6), were trialled *in surfo* as summarised in Table 5.2; generally at two temperatures, two concentrations, and against both MemGold precipitant screens (Newstead et al., 2008; Parker & Newstead, 2012). Despite never exceeding ~8 mg mL^{-1}, non-productive and rapid precipitation, as documented in the VD trials of MmpL3$_{1\text{-}753}$ alone, remained common for sybody complexes; with no evidence the binders fruitfully augmented truncation crystallisability. Disappointingly, this was equally true *in meso*, where most of the same complexes, reconstituted into MO at around 20 mg mL^{-1}, were screened against seven precipitant panels; summarised in Table 5.3. The few noteworthy crystalline species emergent in complex-laden LCP predominately occurred as single entities, and were often confirmed as salt on shooting with X-rays. The only compelling exception was observed in a MmpL3$_{1\text{-}753}$-5B1 bolus (screen TS4, well B5), in

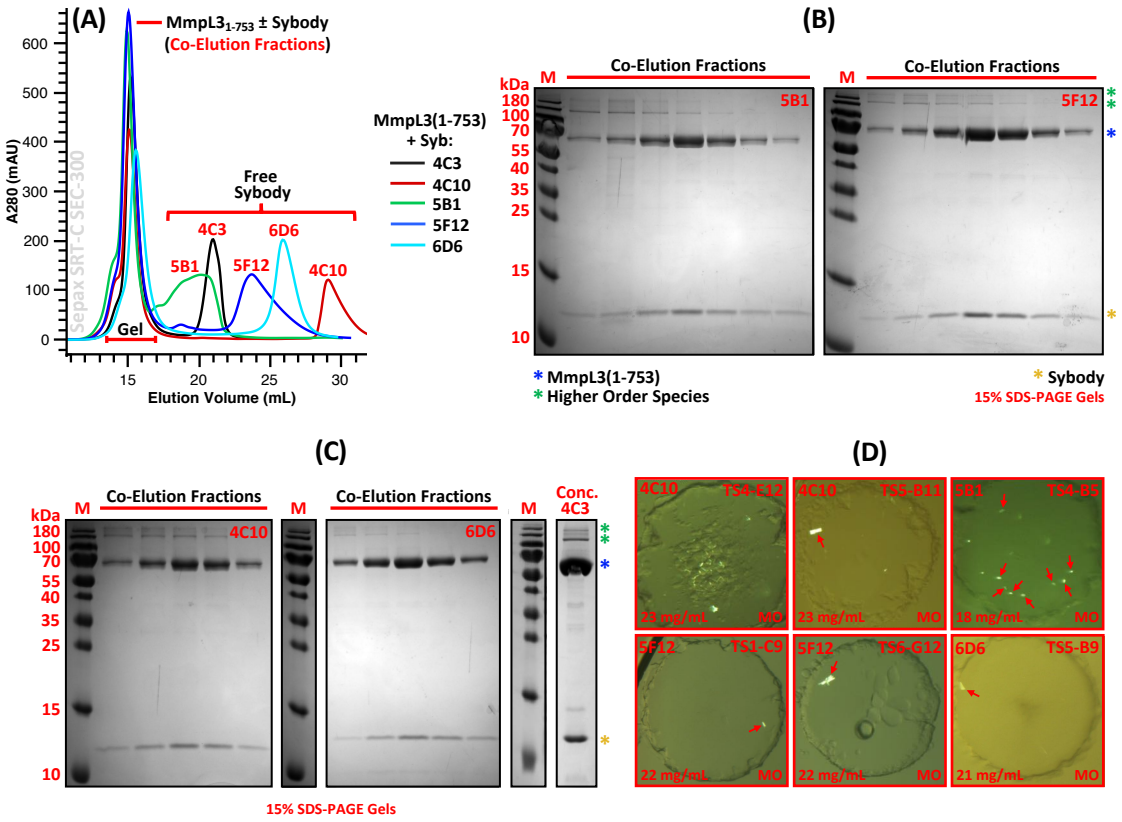


Figure 5.13: MmpL3$_{1-753}$-Sybody Complexation for Crystallisation in DDM

(A) Overlay of Sepax SRT-C SEC-300 traces obtained by individually incubating DDM purified MmpL3$_{1-753}$ with five different tag-free sybodies, in a 1:2 molar ratio, prior to injection. Annotations identify complex and free sybody peaks; the red bar denoting material from the former visualised on the SDS-PAGE gels of **(B)** & **(C)**. Central co-elution fractions were next concentrated for VD (see Table 5.2) and/or LCP (see Table 5.3). Example sample composition post-concentration is given in the gel lane analysing the truncation-4C3 complex. **(D)** Representative cross-polarised photos of the crystalline species (red arrows) observed in boli of MmpL3$_{1-753}$-sybody LCP trials. Images are annotated (clockwise) with binder alphanumeric designator, crystallisation screen & well number, host lipid, and complex concentration.

Sybody	Conc. (mg mL^{-1})	Temp. (^{o}C)	Screens
4C3	~2.5 & 5	4 & 18	MG & MG-II
4C10	~4 & 8	18	MG-II
5B1	~5	4 & 18	MG & MG-II
5F12	~5	4 & 18	MG & MG-II
6D6	~2.5 & 5	4 & 18	MG & MG-II

Table 5.2: Summary of MmpL3$_{1-753}$-Sybody Vapour Diffusion Trials

Specifics of sitting drop crystallisation trials undertaken with truncation-sybody complexes in DDM. Tabulated from LTR: Sybody name, complex concentration (NanoDrop), plate incubation temperature, and the sparse matrix screens tested (MemGold (MG, Newstead et al. (2008)), MemGold2 (MG-II, Parker & Newstead (2012))).

Sybody	Conc. (mg mL^{-1})	Host Lipid	Temp. (°C)	Screens
4C10	~23	MO	20	MM & TS1-6
5B1	~18	MO	20	MM & TS1-6
5F12	~23	MO	20	MM & TS1-6
6D6	~21	MO	20	MM & TS1-6

Table 5.3: Summary of MmpL3$_{1\text{-}753}$-Sybody LCP Trials

Details of the LCP crystallisation trials setup with truncation-sybody complexes in DDM. Tabulated from LTR: Sybody name, complex concentration (NanoDrop), cubic phase host lipid (Monoolein (MO)), plate incubation temperature, and the 96-condition sets screened (MemMeso (MM), Tokyo Set (TS, Nureki lab)).

which grew a constellation of straight-edged rhomboid objects (Figure 5.13[D]). Failing to diffract however, the condition was not pursued further. On aggregate, and at least for subset tested, these VD and LCP results bring into the question the suitability, if not viability, of sybodies as truncation crystallisation chaperones. To determine their potential utility as Cryo-EM fiducials, MmpL3$_{1\text{-}753}$-sybody vitrification trials were next attempted.

5.5.3 Cryo-EM Trials

Besides using only freshly DDM purified truncation, to minimise protein freeze-thaw, MmpL3$_{1\text{-}753}$-sybody Cryo-EM samples were prepared identically as for crystallisation (Figure 5.14[A]). Once concentrated, complexation was checked by SDS-PAGE (Figure 5.14[B]), prior to grid vitrification and appraisal of particle density and distribution by imaging on a 200 kV Talos Arctica Cryo-TEM (Figure 5.14[C]). Of the seven tag-free sybodies trialled, including at least two from each scaffold (4C3/C12/E12, 5F1/F12 & 6D6/F2), six co-eluted commensurately with the behaviour of their previously characterised Myc-His$_6$ derivatives. However, 4E12 — the "best" candidate by the metrics of ELISA signal and k_{off} — surprisingly failed to form a SEC-stable complex with MmpL3$_{1\text{-}753}$ (Figure 5.14[B]). As 4C10 co-eluted well, also belongs to the dominant concave family, and differs from 4E12 at just three permuted positions (two in CDR2, one in CDR3), this anomaly was not investigated further. Yet even speculatively

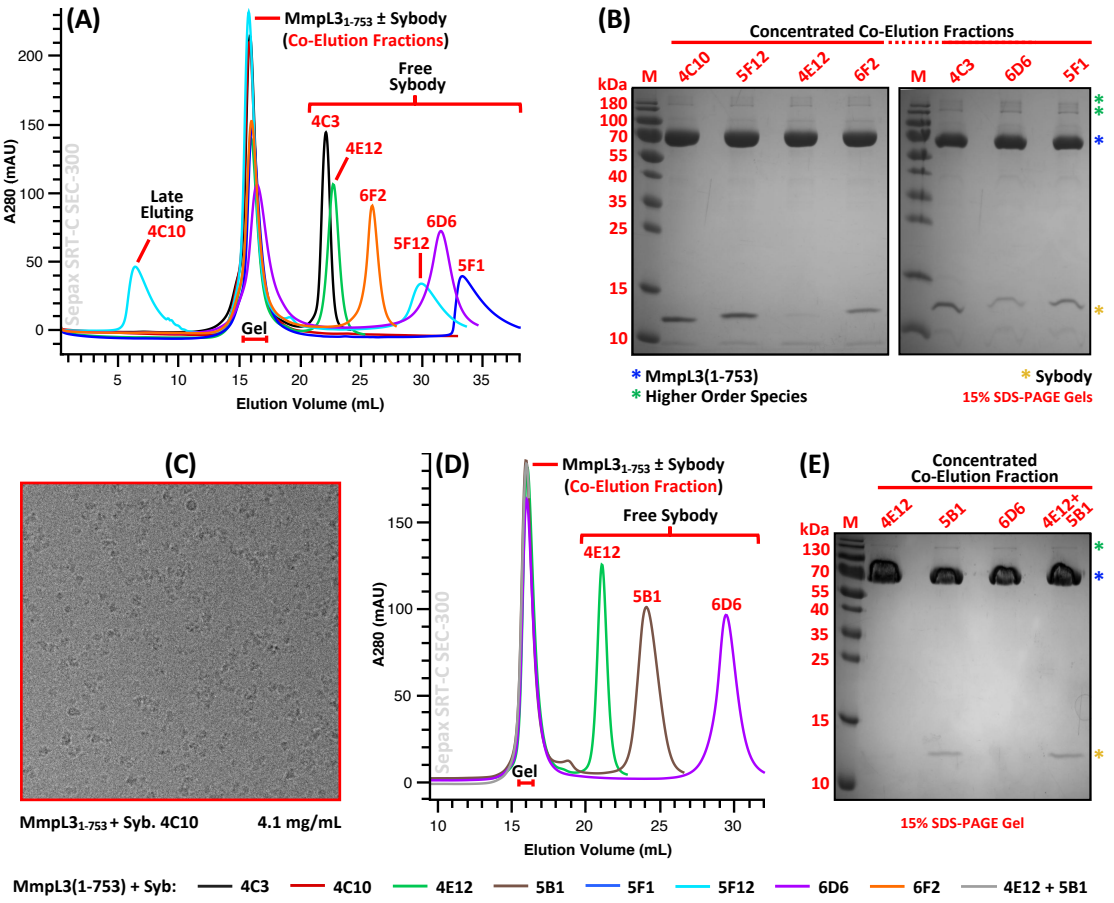


Figure 5.14: MmpL3$_{1\text{-}753}$-Sybody Complexation for Cryo-EM in DDM & LMNG

(A) Overlain Sepax SRT-C SEC-300 profiles reporting consecutive injection of DDM purified truncation, each incubated with one of seven tag-free binders, predominately in a 1:2 molar ratio, prior to gel filtration. Labels delineate MmpL3$_{1\text{-}753}$-sybody complex and free sybody peaks. The red bar identifies complex material concentrated for vitrification screening and analysed on on the SDS-PAGE gels of **(B)**. Note the differing stoichiometries apparent in each truncation-sybody sample. **(C)** Exemplar Arctica micrograph of the MmpL3$_{1\text{-}753}$-4C10 complex imaged at the stated concentration. Equivalent data documenting attempts to assemble sybody complexes with LMNG solubilised MmpL3$_{1\text{-}753}$ are shown in **(D)** & **(E)**.

coming up with satisfactory reasons for the deterioration of tagless 4E12 binding, beyond freeze-thaw intolerance or a peculiarity specific to the prep, is difficult.

As preliminary micrographs showed complexes of the remaining sybodies were moderately clumpy in DDM (Figure 5.14[C]), re-complexation in LMNG was sought (Figure 5.14[D]) before undertaking substantive data collection. This was in hope of leveraging the increased stability, and related particle improvements, already evidenced for MmpL3$_{1\text{-}753}$ alone in the MNG class detergent (see Sections 3.4.2 & 4.2.2). Unexpectedly, albeit now not for 4E12, the initial panel of

sybodies tested in LMNG either outright failed to form SEC-stable complexes or, in the case of 5B1, co-eluted with MmpL3$_{1\text{-}753}$ so substoichimetrically as to be not worth imaging (Figure 5.14[E]) — hinting at incompatibility between the binders, originally selected in DDM, and target presented in LMNG. To ascertain the pervasiveness of this putative detergent conditionality, nickel resin pull-down of the truncation by six Myc-His$_6$ tagged sybodies was compared in DDM and LMNG (Figure 5.15). All but 6F2 displayed a marked reduction in MmpL3$_{1\text{-}753}$ capture in the latter, suggesting LMNG may compromise affinity of a significant proportion of the sybody cohort; explaining the difficulty of isolating workable truncation complexes in the detergent by gel filtration. One must presume, therefore, that the MmpL3$_{1\text{-}753}$ epitopes targeted by many of the binders are either sterically inaccessible in LMNG, a product of increased micellar size (Stetsenko & Guskov, 2017), or, more likely given its aforenoted affects on protein purification, stability, and imaging, that the lipid-esque detergent alters truncation conformational dynamics and/or fold in a manner reducing relevant epitope availability. That sybodies have been gainfully selected in LMNG against other targets (Zimmermann et al., 2020; Shahsavar et al., 2021) strongly argues the observed phenomenon does not stem from LMNG incompatibility with the wider platform, but is, as postulated, an unforeseen consequence of detergent mismatch between binder selection and downstream application.

Rescreening for an LMNG Compatible Fiducial

As the extent to which LMNG perturbed sybody association, although reasonably severe, appeared variable between binders (Figure 5.15), it followed that there may exist a subpopulation in the alkyl maltoside-selected libraries competent at robust, SEC-stable, and stoichiometric complexation in the desired detergent. Therefore, as an alternative to restarting the entire selection campaign in LMNG, glycerol stocks of the second phage display outputs were revived and again FX subcloned, into pSb_init, for re-ELISA against biotinylated MmpL3$_{1\text{-}753}$-Avi purified in the MNG (Supplementary Figure B.7). Unable to facilely recover ≥95

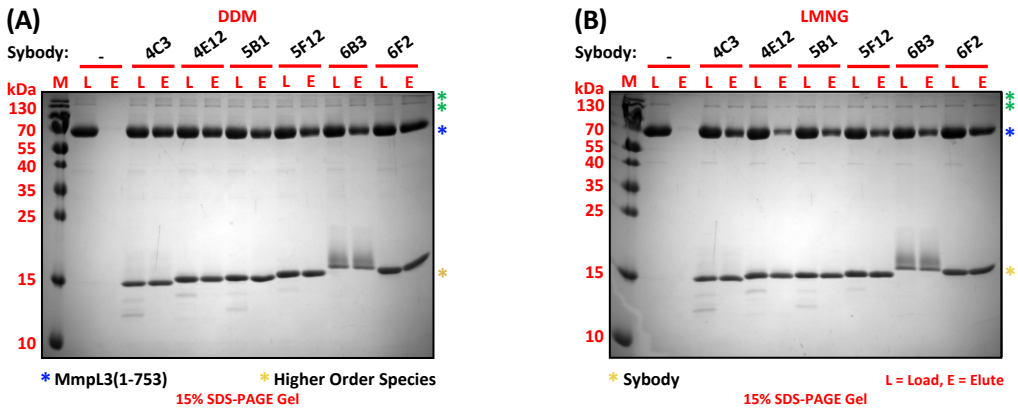


Figure 5.15: $MmpL3_{1\text{-}753}$-Sybody Pull-Down Comparison in DDM & LMNG

(A) & **(B)** SDS-PAGE gels contrasting the nickel resin pull-down of $MmpL3_{1\text{-}753}$ by the same panel of six Myc-His_6-tagged sybodies in DDM versus LMNG. Adjacent lanes show sample composition pre-bead loading (L) and post-bead elution (E). A no sybody (–) control is included. For both detergents, ~40 μg of truncation was incubated with binder in 1:1.5 molar ratio prior to pull-down.

new clones for the concave scaffold, only the loop and convex libraries were advanced. In each case the same binders were simultaneously assayed with respect to DDM solubilised truncation on a second plate — serving as both a comparator and a control.

Evident from visual inspection alone, the scarcity of ELISA signal developed in LMNG, contrasted with its recurrent abundance in DDM, is consistent with considerable differences in epitope presentation between the detergents (Figure 5.16[A]). It leads one to consider whether DDM may permit partial unfolding of $MmpL3_{1\text{-}753}$, to expose artificially immunodominant, possibly semi-linearised, epitopes. Nonetheless, a few wells exhibited weak signal in LMNG and one, a medium sybody henceforth denoted MG2, was carried forward for characterisation. Identical at 5/16 variable positions, including three in CDR3, MG2 is most similar to 5B7 of the first cohort (Figure 5.16[A]). Unlike the latter sybody however, abandoned due to non-expression, small-scale purification of MG2 yielded homogeneous, gel filtration monodisperse, Myc-His_6 tagged protein (Figures 5.16[B]&[C]). More importantly, MG2 displayed pull-down of (Figure 5.16[D]), and SEC co-elution with (Figures 5.16[E]&[F]), the truncation in

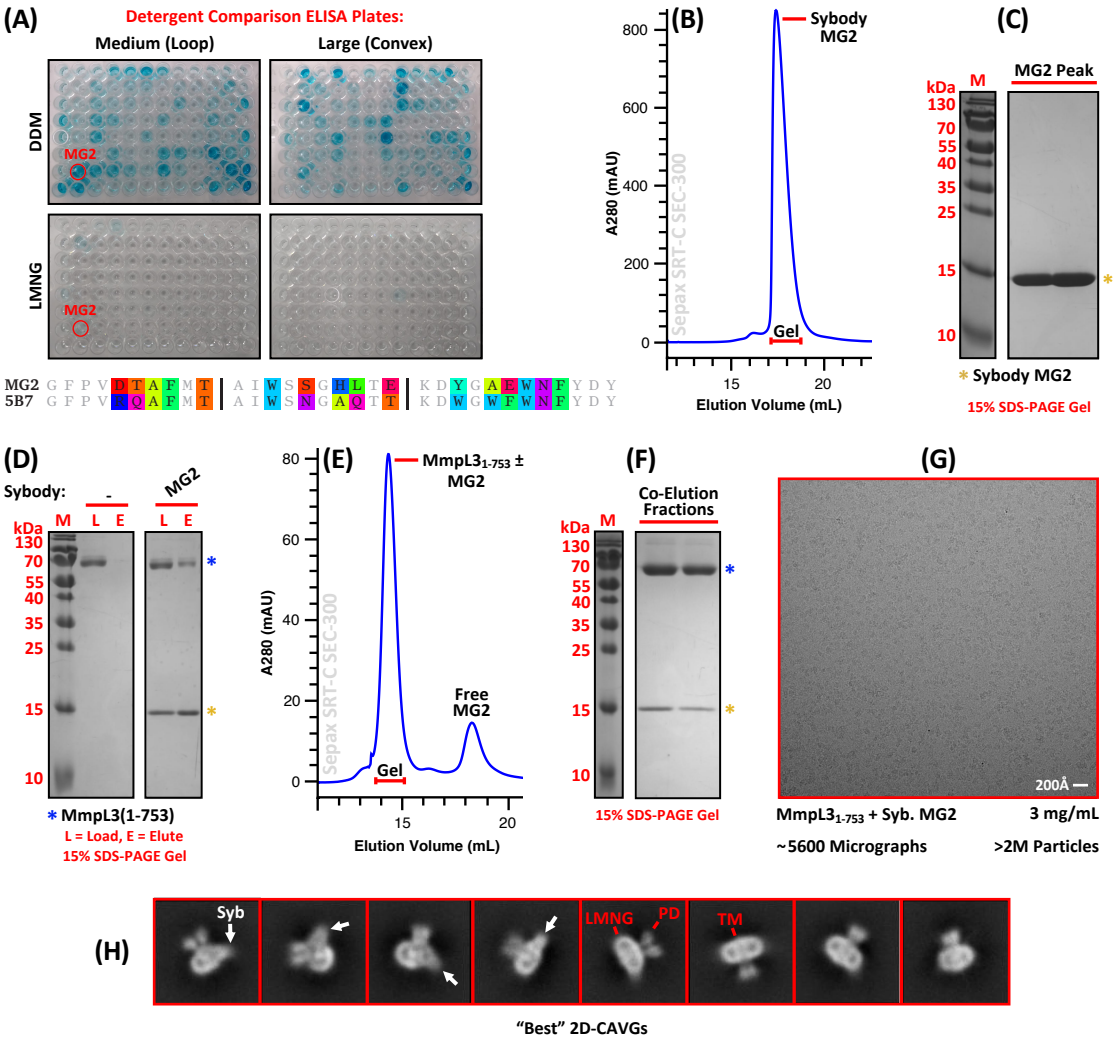


Figure 5.16: Sybody Re-ELISA in LMNG & Cryo-EM Imaging of MmpL3$_{1\text{-}753}$-MG2

(A) Post-development ELISA plate photos documenting re-screening of the phage display output from the medium and large libraries in DDM versus LMNG. The same 95 pSb_init clones, per library, are assayed in each detergent. The remaining well (H12, bottom right-hand corner) hosts 4E12, the sybody producing the strongest signal in the initial ELISA, as a positive control. Note the stark difference between detergents, highlighting the general incompatibility of DDM-selected binders with LMNG purified MmpL3$_{1\text{-}753}$. Loop sybody MG2 (circled), giving a modest read-out in the latter detergent, was advanced in hope of its viability as an LMNG compatible Cryo-EM fiducial. MG2 CDRs 1–3 are provided, delimited by black lines, coloured as in Figure 5.6, and aligned to 5B7; the most sequence similar sybody from the first cohort. **(B)** Sepax SRT-C SEC 300 trace of Myc-His$_6$-tagged MG2; purified small-scale by IMAC. The red bar denotes protein loaded on the SDS-PAGE gel of **(C)**. This purified material was used to perform the truncation nickel-resin pull-down presented in **(D)** and SEC co-elution trial shown in **(E)** & **(F)**, both undertaken in LMNG. The former includes a no sybody (–) control, with adjacent lanes analysing samples pre-bead loading (L) and post-bead elution (E). SEC co-elution was undertaken with ~250 μg of truncation incubated with MG2 in 1:1.5 molar ratio prior to injection. **(G)** Representative section of a Titan Krios micrograph generated on imaging of the MmpL3$_{1\text{-}753}$-MG2 complex in LMNG, assembled at a 1:1.2 truncation:sybody molar ratio; without further SEC polishing. Collection metadata is given. **(H)** Best SIMPLE (Reboul et al., 2018) 2D-CAVGs. The LMNG micelle, periplasmic domain (PD), transmembrane bundle (TM), and putative MG2 density (Syb, white arrows) are annotated. Processing courtesy of Justin Deme.

LMNG. Consequently, a sample of the MmpL3$_{1\text{-}753}$-MG2 complex was imaged by Cryo-EM. Good particle density was realised at ~3 mg mL^{-1}, and recent installation of a K3 DED expedited collection of a Titan Krios dataset exceeding 2 million particles (Figure 5.16[G]). In an effort to maximise the proportion that were sybody-bound, the LMNG MmpL3$_{1\text{-}753}$-MG2 sample was assembled at a 1:1.2 molar ratio and then directly concentrated and vitrified — i.e. without intervening complex isolation by SEC. Despite this, but undeniably still progress, only a small fraction of resultant 2D-CAVGs featured extra-density ascribable to MG2 (Figure 5.16[H]). These showed the sybody to engage not, as had been anticipated, the readily accessible PD of MmpL3$_{1\text{-}753}$, but instead to bind to, or at least in the vicinity of, the LMNG micelle; feasibly contacting one or more loops connecting helices of the transmembrane bundle. As the non-complexed 2D-CAVGs, representing the majority of the particle stack, were not of comparable quality to those of original LMNG dataset, additional processing was deemed futile. At this point, headway being made in Cryo-EM imaging of MmpL3$_{1\text{-}753}$ alone (see Chapter 7) indicated a sybody fiducial may not be essential, and so further work on MG2 was halted. If ever revisited, the most effective route to a sybody useful for structural studies of MmpL3$_{1\text{-}753}$ is likely to begin by repeating selection afresh, presenting the target in LMNG from ribosome display onwards. Failing that, one could pursue *in vitro* affinity maturation of MG2 with view to achieving a uniform field of complexes in-ice.

5.5.4 Conclusion

To recap, as binders are a proven method to aid structural solution of challenging membrane protein targets, often furnishing advantages for both XRC and Cryo-EM, *in vitro* selection of sybodies against MmpL3$_{1\text{-}753}$ was pursued. After multiple rounds of ribosome and phage display, 91 unique ELISA positive hits were taken forward; spanning the three paratope scaffolds. In-depth familial classification through sequence analysis revealed prevalent non-randomisation of

CDR1, frequent selection for AAAs in CDRs 2–3, and indications that the majority of concave and convex candidates may target but a few, immunodominant, truncation epitopes. In spite of then screening the cohort via an extensive experimental characterisation pipeline — proceeding through small-scale batch purification, nickel resin pull-down and/or SEC co-elution, and finally off-rate ranking by BLI — sybodies promoting either crystallogenesis or improved in-ice imaging of DDM solubilised $MmpL3_{1\text{-}753}$ could not be found. Attempts to use sybodies in combination with LMNG purified truncation quickly uncovered a general incompatibility with the detergent; convincingly attributable to differential epitope presentation relative to DDM. Although further panning did eventually identify a LMNG competent binder, confirmed to complex with $MmpL3_{1\text{-}753}$ by Cryo-EM, its serviceability as fiducial remains to be determined. Overall, the sybody screening results highlight the paramount importance, in order to preserve binder utility downstream, of thoroughly optimising membrane protein preparation, especially with respect to detergent choice, before beginning a selection cascade. To assess any Cryo-EM benefits emergent on reconstitution back into a bilayer, Chapter 6 next reports outcomes of $MmpL3_{1\text{-}753}$ incorporation into MSP nanodiscs.

MSP Nanodisc Studies

From PepTSt to MmpL3

CHAPTER 6

Seeking to structurally explore *Mtb* MmpL3 in a more native setting, reconstitution of the truncation into membrane scaffold protein (MSP) nanodiscs was next attempted. It was hoped this would improve — possibly with fiducial aid from one or more of the sybodies generated in Chapter 5 — the Cryo-EM tractability of $MmpL3_{1\text{-}753}$. In turn, nanodiscs could be expected to offer an avenue to dissect pertinent protein–lipid interactions only emergent within a bilayer environment. These are likely to be of utmost importance considering the role of MmpL3 as a TMM exporter. Before discussing the truncation, whose nanodisc reconstitution proved successful but otherwise failed to provide post-vitrification benefits, the chapter details investigations undertaken with PepTSt — a proton-coupled peptide transporter from the bacterium *Streptococcus thermophilus*. A much valued "workhorse" within the Newstead Group, and on the back of over a decade of study, PepTSt has been thoroughly characterised; both structurally and functionally. It thus represented a perfect candidate with which to acquire competency in MSP nanodisc methodology, as well as Cryo-EM data processing, prior to further MmpL3 pursuits. Readily compliant, PepTSt could be reconstituted into nanodiscs bounded by a variety of MSP constructs, and a "near-atomic" resolution (~3.6 Å) structure in MSP1D1 is presented — one of the only nanodisc-inserted MFS transporter resolved by Cryo-EM to date. Idiosyncrasies observed in PepTSt nanodisc size are also examined; being of note to practitioners more generally given their disagreement with the "longer MSP, larger nanodisc" orthodoxy reported throughout much of the literature.

6.1 The MSP Nanodisc System

Nanodiscs (NDs) — optimised puck-like lipoproteins of controllable size and lipid composition — are arguably the most well characterised, and near native, detergent-alternative technology available for membrane protein reconstitution; others of note include saposins, the aforementioned APols, and styrene maleic acid lipid particles (i.e. SMALPs; Denisov & Sligar (2017)). The discoidal ND bilayer is stabilised, and ultimately delimited, by a toroidal pair of antiparallel MSPs, secured to one another by an array of salt bridges (Figure 6.1[A]; Bibow et al. (2017)). These proteinaceous "belts" are of modular design, having been rationally engineered from human apolipoprotein A1 — a core constituent of the high-density lipoprotein responsible for circulatory fat transport — over fifteen years ago (Bayburt et al., 2002; Bayburt & Sligar, 2009; Denisov et al., 2004). Formed of series of short amphipathic α-helices punctuated by glycine or proline pivots, the suite of available MSPs vary in length; each construct defining their own upper limit of bilayer diameter and thus, in theory, the maximal transmembrane footprint they can accommodate (Bayburt & Sligar, 2010). The MSP variants employed here are detailed in Figure 6.1[B]. All were expressed as N-terminal His_6-TEV fusions in *E. coli* and first purified by IMAC. Proteolytic affinity tag removal was achieved during overnight dialysis, with acceptably pure MSP recovered following reverse IMAC (Supplementary Figure B.11).

The advantages of the MSP ND system stem from the capability to reconstitute a target protein into a experimentally tunable bilayer (Denisov & Sligar, 2017). It is increasingly recognised that lipids are far from an inert support matrix, but instead act as active modulators of function, fold, stability, dynamics, and/or oligomeric state of the membrane proteins they host (Lee, 2004; Gupta et al., 2017; Martens et al., 2018). Additional to the bulk properties of biological membranes (e.g. lateral pressure, electrostatic potential, thickness etc.), this is a result of the lipid annulus contacting the transmembrane domain. Said lipids may associate so strongly with the protein exterior, in specific binding sites, as

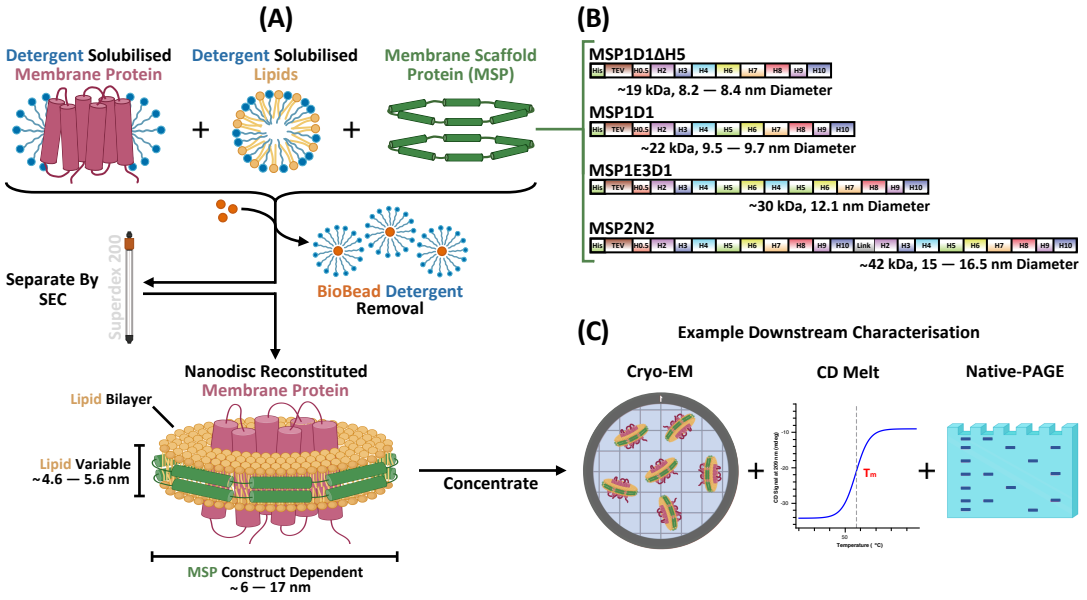


Figure 6.1: Nanodiscs: Reconstitution, Characterisation & MSPs

(A) Schematic of the ND reconstitution protocol used. Protein solubilisation detergent varied (DDM, DM or LMNG), whilst lipid — in all cases asolectin — was solubilised in 3 × molar excess sodium cholate. **(B)** Modular organisation of the four MSP constructs employed, presented in ascending size order. Excluding the N-terminal His_6-TEV tag and linker (Link), each coloured box represents a helical unit of 11 (H0.5, H3, & H9), 22 (H2 & H4–8) or 24 (H10) residues. Depiction inspired by Sigma-Aldrich. Approximate MSP ND diameter collated from Ritchie et al. (2009) & Yokogawa et al. (2019); derived from SEC, EM and/or SAXS. **(C)** Example ND characterisation techniques used in this chapter. Figure partly composed with BioRender.

to survive detergent delipidation and thus become resolvable by subsequent structural efforts (Yeagle, 2014). Unlike detergent micelles, the miniaturised and solution-dispersed bilayers provide by NDs are an almost ideal vehicle to explore the nuances of such protein-lipid interactions; be it by biochemical, biophysical, or structural methods (Denisov & Sligar, 2016). Unsurprisingly, they have found significant utility in Cryo-EM. Here NDs are exploited for the improved contrast afforded by a detergentless buffer, as well as the target stabilising effects, and physiological relevance, of reintroduction into lipid (Efremov et al., 2017). A prime example is the MSP2N2-reconstituted structure of TRPV1, a cation channel gated by various stimuli (e.g. temperature), which resolved both annular and allosteric lipids (Gao et al., 2016). NDs were attractive for $MmpL3_{1\text{-}753}$ as they might not only permit interrogation of interactions with its lipid substrate (i.e. TMM), but also with the multiple hydrophobic — lipid-partitioning — preclinical antituberculars in development (e.g. SQ109, AU1235, SPIRO).

As for detergent solubilised samples, the suitability of NDs for structural studies is dependent on obtaining a homogeneous preparation; that is, a ND population of uniform size, lacking variance in the number of target proteins reconstituted per disc, and free of aggregates (Ritchie et al., 2009). Achieving this is well documented to require careful optimisation of the protein : MSP : lipid molar ratio used during assembly (McLean et al., 2018). One often begins, if appropriate target information is at hand (e.g. a crystal structure, homology model etc.), by selecting an MSP long enough to entrap the transmembrane region in at least two concentric rings of bilayered lipids (Efremov et al., 2017). The latter can be compositionally tailored, to mimic native membranes for example, as desired.

In practice, small test assemblies are undertaken to empirically screen a range of reconstitution conditions (Ritchie et al., 2009), and were performed in the current work as follows (Figure 6.1[A]). Sodium cholate solubilised asolectin — a commercial phospholipid extract from soybean, applied fruitfully in several ND structures (e.g. TRPV1 Gao et al. (2016), PKD2 Shen et al. (2016), & OSCA1.2 Jojoa-Cruz et al. (2018)) and already shown to be competent as an $MmpL3_{1\text{-}753}$ liposome host — was mixed with purified target and MSP at a series of molar ratios, and in a final volume of ~150 μL. After equilibration on ice (>1 hour), detergent removal, triggering ND self-assembly, was accomplished by incubation with excess Bio-Beads in two consecutive rounds; the final being overnight, and both at 4 °C. The next day, post-ultracentrifugation, ND quality was assessed by SEC in a detergentless mobile phase. Optimal ratios give a sharp monodisperse trace with minimal empty discs or lipid-rich aggregates. ND formation was further verified by target and MSP presence in peak fractions analysed by SDS-PAGE and visualised by silver staining. Once found, viable reconstitution parameters were scaled up by ~20–30× for Cryo-EM trials, ND characterisation (Figure 6.1[C]), and, where applicable, sybody complexation attempts. In advance of tackling the relative unknowns of MmpL3, ND methodology was gainfully learnt through experimentation on PepTSt.

6.2 PepTSt Nanodiscs: A Model System?

6.2.1 An Introduction to PepTSt

An MFS symporter using the inwardly directed proton electrochemical gradient to energise uptake of environmental di- and tripeptides, PepTSt is a bacterial stalwart of the wider proton-dependent oligopeptide transporter (POT), also known simply as the peptide transporter, family (Newstead, 2015). Much interest in this experimentally tractable protein derives from its homology to, and thus ability to inform on, two human POTs — PepT1 (SLC15A1) and PepT2 (SLC15A2) (Solcan et al., 2012). The former drives bulk absorption of dietary protein in the small intestine, whereas the latter minimises urinary loss of nutritional peptides in the kidneys (Daniel et al., 2006). An astounding and physiologically efficient feature of many POTs is their substrate promiscuity. For example, PepT1 and PepT2 are predicted to transport >7000 unique di- and tripeptides (Newstead, 2015). Likewise, peptidomimetic compounds — i.e. boasting steric similarity to a peptide — can hijack these polyspecific systems; a mechanism underpinning the favourable pharmacokinetic profiles of multiple drug classes (e.g. β-lactam antibiotics; Rubio-Aliaga & Daniel (2002)). A detailed structural understanding of PepTSt, and by extension human POTs, able to delineate the molecular basis of substrate recognition and subsequent uptake, may therefore facilitate rational design of peptide-derivatised prodrugs exploiting PepT1/2 for improved oral bioavailability (Brandsch, 2009).

The pharmacological relevance and sheer crystallisability of PepTSt has resulted in a slew of *apo* and substrate-bound structures — 21 having been deposited in the PDB at the time of writing. Most were solved *in meso* (Figure 6.2[A]; e.g. Lyons et al. (2014); Martinez Molledo et al. (2018a)), but the first was crystallised by VD in DDM (Figure 6.2[B]; Solcan et al. (2012)). They show PepTSt to display a broadly canonical MFS fold, formed of two helical bundles arranged pseudo-symmetrically to give a "V"-shaped transporter consistently captured in

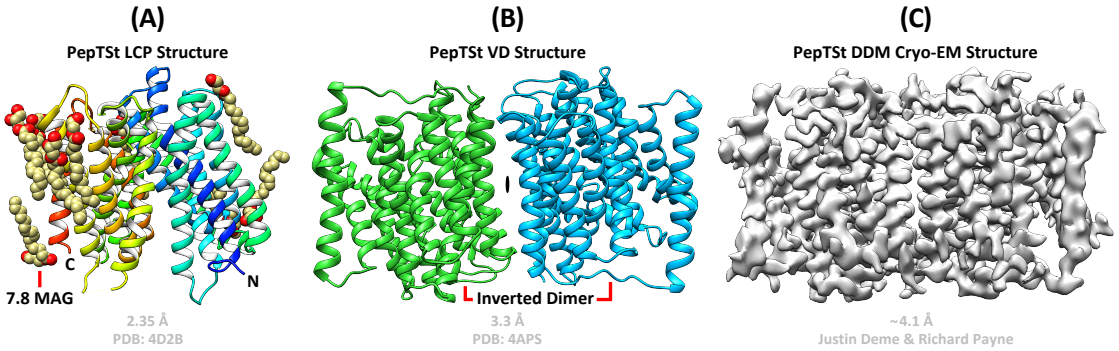


Figure 6.2: Exemplar PepTSt Structures

(A) Monomeric asymmetric unit of the first lipidic cubic phase (LCP) structure of PepTSt (PDB: 4D2B, Lyons et al. (2014)). Shown in cartoon form, coloured blue (N-terminus) to red (C-terminus), and with co-crystallised 7.8 MAG lipids shown as spheres. **(B)** Inverted dimer asymmetric unit of vapour diffusion (VD) crystallised PepTSt in DDM (PDB: 4APS, Solcan et al. (2012)). Cartoon coloured by chain, with C_2 symmetry axis highlighted. **(C)** Cryo-EM map of DDM solubilised PepTSt, displaying identical quaternary arrangement to the VD asymmetric unit, solved by Richard Payne & Justin Deme (unpublished). All structures are in the inward-open conformation and were composed with UCSF Chimera v.1.13.1 (Pettersen et al., 2004).

the inward-open conformation (i.e. with the peptide-binding site accessible to the cytoplasm). Each domain features six transmembrane helices with a "3+3" inverted repeat topology. Akin to other bacterial POTs, the inter-domain loop hosts two additional membrane-spanning helices (HA & HB) to bring the total to fourteen (Newstead, 2015).

Whilst these structures, elaborated upon by functional studies (e.g. see Parker et al. (2014)), begin to tease apart rules governing PepTSt substrate engagement and proton-coupling, much work remains to elucidate how POTs, and MFS transporters more generally, operate within native bilayers. Their alternating-access cycles warrant extensive conformational change and the proteins are therefore highly likely to be modulated by the surrounding membrane (Yan, 2015). Exemplifying this, a recent single-molecule FRET study reported distinct conformational modulation of DtpA, an *E. coli* POT, on reconstitution into saposin nanoparticles of differing lipid composition (Male et al., 2020). Similarly, others have leveraged NDs, hydrogen-deuterium exchange mass spectrometry, and molecular dynamics simulations, to demonstrate specific lipid interactions shifting the conformational equilibria of three MFS members (Martens et al., 2018). In hopes of analogous

insights, obtaining a structural description of PepTSt reconstituted into NDs was sought. The bacterial POT had previously proven amenable to Cryo-EM in-house, having been resolved to ~4.1 Å in detergent (Figure 6.2[C]); despite its small size (~53 kDa monomer) and lack of alignment assistive extramembranous domains. The increased mass and symmetry afforded by dimerisation in DDM, replicating the asymmetric unit of the original PepTSt structure (Figure 6.2[B]), clearly helped; albeit oligomerisation of dubious biological significance given the antiparallel protomer arrangement (Quistgaard et al., 2017). On joining the ND project, colleagues — Joanne Parker and Justin Deme — had already begun preliminary Cryo-EM screening of PepTSt reconstituted into MSP1D1, and it is thus from this starting point results are presented.

6.2.2 The PepTSt-MSP1D1-Asolectin Structure

Detergent purified PepTSt (Supplementary Figure B.12) could be optimally reconstituted into MSP1D1-asolectin NDs at a protein : MSP : lipid molar ratio of 1:3:100. Scaled up, grids prepared at ~0.9 mg mL^{-1} showed viable in-ice density with clear particle separation (Figure 6.3[A]) and so a substantial dataset was collected (~4200 micrographs). On the fly autopicking, employing a previous detergent solubilised PepTSt volume as reference, recovered ~630K particles. Processed in SIMPLE (Reboul et al., 2018), the resulting 2D-CAVGs (Figure 6.3[B]) identified a pronounced PepTSt dimer trapped in a tight ellipsoidal ND of ~120 Å in its longest dimension. Particles in reasonable 2D-CAVGs (~200K) were selected as input for the generation of a trio of C_2-symmetric 3D classes in RELION (Scheres (2012); Figure 6.3[C]). The most populated and detailed of these, “Class two” (40.4%), confirmed the dimer to be of the same inverted arrangement seen in the *in surfo* and Cryo-EM structures, both solved in DDM. This was relatively surprising. Reintroduction into a lipid bilayer had been hypothesised to most likely trigger monomerisation, in a manner comparable to the asymmetric units of all PepTSt LCP structures. The

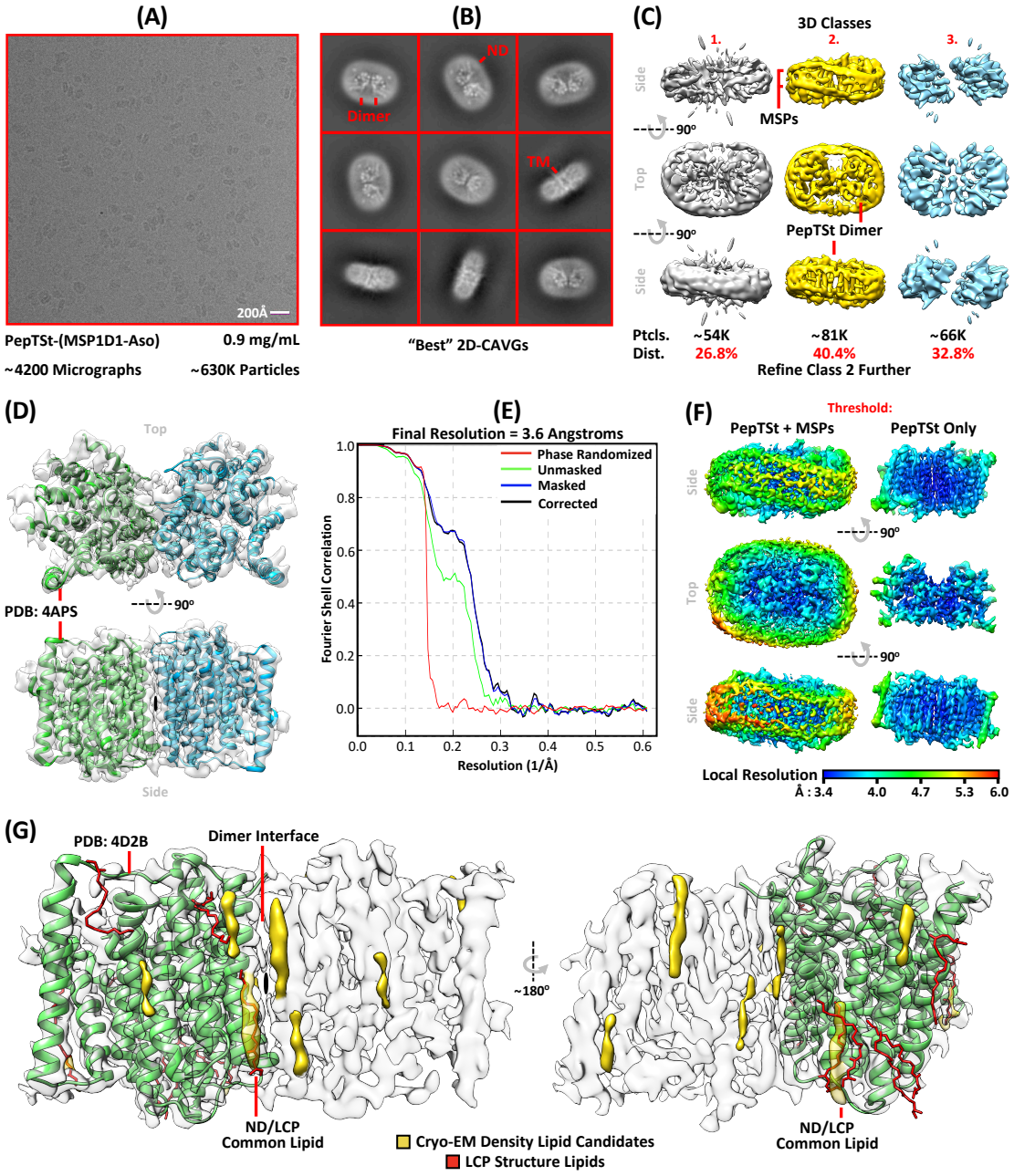


Figure 6.3: Cryo-EM Reconstruction of PepTSt-MSP1D1-Asolectin Nanodiscs

(A) Representative Titan Krios micrograph of PepTSt-MSP1D1-asolectin NDs prepared at a 1:3:100 protein:MSP:lipid molar ratio. Sample concentration, collected micrograph number, and particle set size are noted. **(B)** Best SIMPLE (Reboul et al., 2018) 2D-CAVGs. The top-viewed PepTSt dimer, ND exterior, and side-viewed transmembrane bundle (TM) are annotated. **(C)** Initial 3D classes generated in RELION (Scheres, 2012) with enforced C_2 symmetry. Particle distribution is given. Class two, boasting evident PepTSt and MSP density, was chosen for further refinement. **(D)** Final RELION post-processed map, thresholded to remove ND density for clarity, with the original PepTSt *in surfo* structure, coloured by chain, docked (PDB: 4APS, Solcan et al. (2012)). Associated gold-standard FSC curves are shown in **(E)**, with resolution stated at the 0.143 cut-off. **(F)** Maps coloured by local resolution computed in RELION, thresholded to include or hide the ND. **(G)** First PepTSt *in meso* structure (PDB: 42DB, Lyons et al. (2014)) docked into the Cryo-EM envelope. Non-protein, putative lipid, density is coloured yellow for comparison to the co-crystallising 7.8 MAG molecules in red. Structures composed in UCSF Chimera v.1.13.1 (Pettersen et al., 2004). ND sample preparation by Joanne Parker, image processing led by Justin Deme.

benefits of dimerisation for Cryo-EM were thus realised without the predicted need to cross-link PepTSt, by engineering interfacial disulphides for example, prior to ND reconstitution.

Strikingly, two parallel bands are also resolved in class two — presumably each a copy of MSP1D1 — wrapping around the dimer in a contorted fashion, and consequently offset by different angles with respect to the transporter on opposing faces of the ND. The boundary traced by the buckled MSPs, a plausible guide to bilayer position, is reminiscent in shape to the micellar bulk of the PepTSt-DDM volume. This suggests the observed departure from planarity may be a result of POT, rather than MSP, induced "membrane" deformation; a notion in keeping with a meta-analysis concluding detergent and ND belts recovered in Cryo-EM reconstructions, of the same protein, are often alike (**Zampieri et al., 2020**). However, it should equally be noted that ND phospholipids are documented to be slightly perturbed relative to native bilayers (**Skar-Gislinge et al., 2010**). Refining class two further gave a final map of ~3.6 Å global resolution (Figures 6.3[D]&[E]). Whilst an improvement on the detergent Cryo-EM structure, it is not solely attributable to ND incorporation due to dissimilar processing pipelines between datasets, executed in different versions of RELION. Subsequent local analysis showed a logical resolution gradient, gradually worsening as one moves outwards from a well packed dimer interface (~3.4 Å) to the more variable ND periphery (~4.7–6.0 Å, Figure 6.3[F]).

A handful of non-proteinaceous tubular densities, putative lipids, were found to decorate the PepTSt surface; although are not indistinguishable from tightly bound detergent, having survived relipidation, at this resolution. Cross-referencing these with 7.8 MAG molecules co-crystallised with the first *in meso* PepTSt structure identifies moderate overlap (Figure 6.3[G]). The majority of the remaining densities are positionally consistent with lipids resolved in at least one of the other 18 LCP structures reported to date (Supplementary Figure B.13). Such

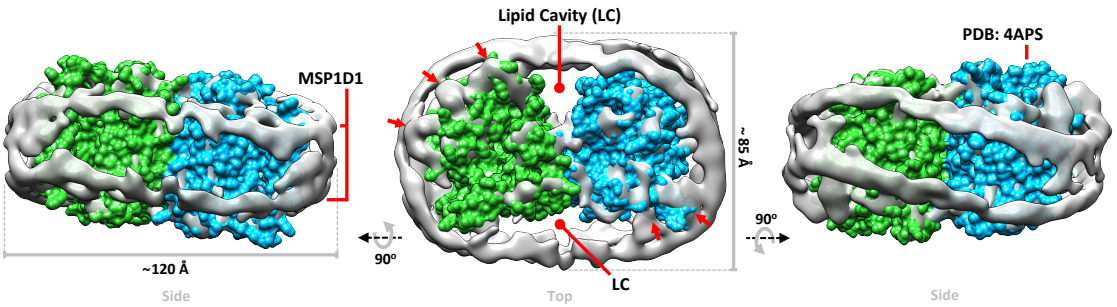


Figure 6.4: Potential PepTSt-MSP1D1 Contact Sites

Surface representation of PepTSt *in surfo* structure (PDB: 4APS, Solcan et al. (2012)), coloured by chain, docked into PepTSt-MSP1D1-asolectin class two (grey). For clarity, smaller "blobs" of inter-MSP density are hidden. Red arrows non-exhaustively identify putative MSP1D1-PepTSt interaction sites; many are associated with loops or chain termini not modelled in the crystal structure. Cavities adjacent to the dimer interface, presumably lipid occupied, and dimensions of the elliptical ND are annotated. Prepared with UCSF Chimera v.1.13.1 (Pettersen et al., 2004).

consistency points to the existence of specific association sites, with annular lipid affinities greater than the rest of the PepTSt exterior. The functional relevance for these remains to be determined. Nevertheless, it is tempting to conjecture that the six ND "lipids" seen in the vicinity of the dimer interface somehow aid oligomer stabilisation; a role for interfacial lipids becoming increasingly evident through systematic study (e.g. Gupta et al. (2017)) and accruing individual examples (e.g. BetP Koshy et al. (2013), UapA Pyle et al. (2018), & Vrg4 Parker et al. (2019)).

The ability to resolve the MSPs at all requires they achieve at least a partially fixed register relative to PepTSt. Coupled with ND tightness, it is thus probable that a number of direct contacts exist between MSP1D1 and the transporter; likely contributing to, and stabilising, the askew perimeter (Figure 6.4). Such interactions are congruent with reports of MSP-membrane protein reconstitution in absence of exogenously added lipids — so called "minimal NDs" (Salvador et al., 2019) — as well as an ensemble of structures of the archeal sodium-aspartate symporter Glt_{Tk} recently solved in MSP2N2-lipid NDs (Arkhipova et al., 2020). The latter describes a homotrimer entrapped in a compact circular bilayer appressed by two clearly resolved MSP belts. The authors note a ND diameter

of only ~11 nm, aberrantly small for MSP2N2 (characterised at ~15–16.5 nm; Ritchie et al. (2009)), but posit that any direct MSP-Glt_{Tk} interactions cannot be deleterious given that their protein retains conformational freedom. Indeed, a tight ND may be speculated to have actually improved Cryo-EM image alignment for PepTSt, by minimising inter-particle heterogeneity arising from differential transporter "float" — translation and azimuthal rotation — within the discoidal bilayer (Denisov & Sligar, 2017). Nonetheless, and with knowledge that PepTSt remains dimeric post-reconstitution, the longer scaffolds MSP1E3D1 and MSP2N2 were next pursued. It was hoped the theoretically larger bilayers provisioned by these constructs would abolish direct MSP-PepTSt contacts, whilst simultaneously permitting an increased number of protein-lipid interactions to be resolved.

6.2.3 Reconstituting PepTSt into Larger MSPs For Cryo-EM

MSP1E3D1

As for MSP1D1, PepTSt was readily reconstituted in MSP1E3D1 — a scaffold including an additional three helical segments — to give a monodisperse S200 profile accompanied by a minor shoulder of empty NDs (Figures 6.5[A]&[B]). This was achieved at a protein : MSP : asolectin molar ratio of 1:5:100; confounding expectations that more lipid per scaffold would be required relative to the shorter MSP, but perhaps also risking under-lipidation. Given the twisted boundary of the PepTSt-MSP1D1 volume, it is of pertinence that lipid insufficiency during lipoprotein assembly is linked to deviations from bilayer planarity; tending to produce saddle-shaped NDs (Denisov & Sligar, 2017). In any case, Cryo-EM screening at ~1 mg mL^{-1} appeared viable (Figure 6.5[C]) and so ~3500 micrographs of PepTSt-MSP1E3D1-asolectin were collected, yielding an initial dataset of ~520K particles. Processing gave 2D-CAVGs very alike to PepTSt-MSP1D1 (Figure 6.5[D]), and the best C_2-symmetric 3D class displayed identical ellipsoidal morphology to the original structure (Figure 6.5[E]) — i.e. without desired increase in ND dimensions — and so further refinement was abandoned.

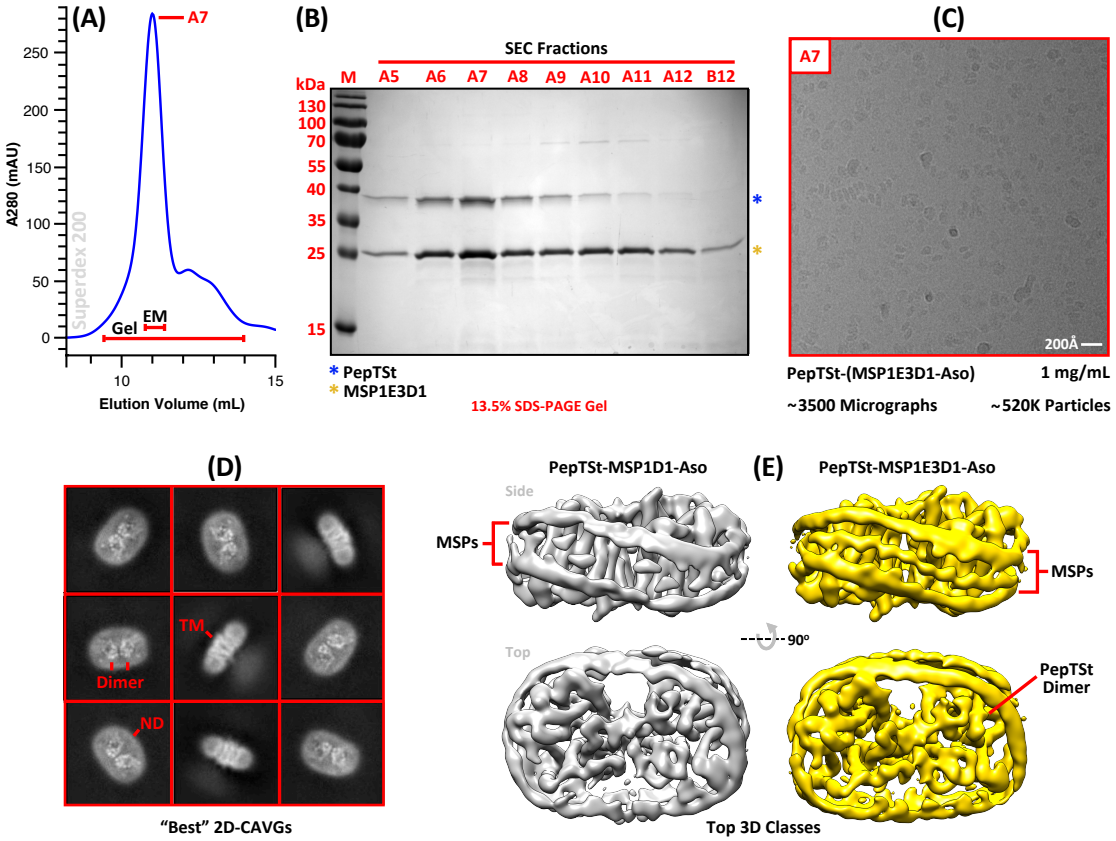


Figure 6.5: Cryo-EM Screening of PepTSt-MSP1E3D1-Asolectin Nanodiscs

(A) SEC trace of PepTSt-MSP1E3D1-asolectin NDs reconstituted at 1:5:100 protein:MSP:lipid molar ratio. Red bars show fractions loaded on the SDS-PAGE gel of **(B)** or concentrated for Cryo-EM; a representative Titan Krios micrograph from which is presented in **(C)**. Collection metadata is given. **(D)** Best SIMPLE 2D-CAVGs (Reboul et al., 2018), labelled as for previous ND samples. **(E)** Side by side comparison of the top RELION (Scheres, 2012) C_2-symmetric 3D classes of MSP1D1– (grey) & MSP1E3D1-asolectin (yellow) incorporated PepTSt. Cryo-EM processing by Justin Deme. Volumes rendered in UCSF Chimera v.1.13.1 (Pettersen et al., 2004).

If discounting consistent under-lipidation, map identity suggests the aforementioned MSP-PepTSt interactions may be sufficiently favourable as to template ND size regardless of MSP length; at least when reconstituting into asolectin. If correct, then an "extra" portion of MSP1E3D1 could be surmised to protrude from the ND boundary and, as unobserved, exhibit significant flexibility. Similar overhangs have been advanced for MSPs with unremoved affinity tags (Skar-Gislinge et al., 2010). Speculation aside, PepTSt incorporation into MSP2N2, nearly the longest scaffold available, was trialled next.

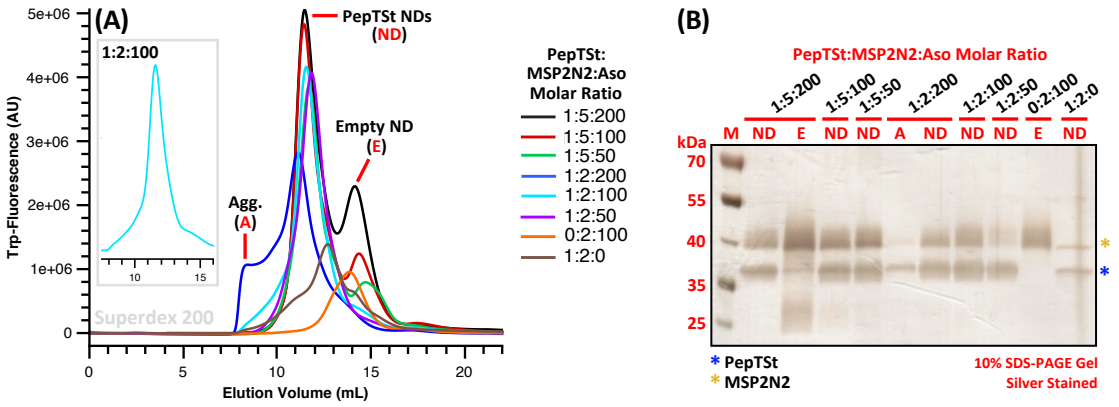


Figure 6.6: Screening PepTSt Reconstitution into MSP2N2-Asolectin Nanodiscs

(A) S200 traces screening PepTSt reconstitution into MSP2N2-asolectin NDs over a range of protein:MSP:lipid molar ratios. Inset: Monodisperse profile of "best" ratio, chosen for subsequent Cryo-EM scale up. Peak labels delineate aggregate (Agg.), NDs with PepTSt incorporated (ND), & empty NDs (E). The same descriptors identify peak fractions visualised on the silver stained SDS-PAGE gel of **(B)**. The 1:2:0 control shows PepTSt-MSP2N2 reconstitution in absence of lipids.

MSP2N2

Constructed from two MSP1D1 molecules fused by a short glycine-threonine linker, with one copy having had the first amphipathic module deleted, MSP2N2 has a total of 19 helical segments (Grinkova et al., 2010). To ensure thorough exploration of reconstitution space, six protein : MSP : lipid molar ratios were screened for PepTSt incorporation into MSP2N2-asolectin; alongside empty-ND (0:2:100) and lipidless (1:2:0) controls. An overlay of the resulting S200 profiles (Figure 6.6[A]) provides a robust example of the need to optimise conditions. Excess lipid (e.g. 1:2:200) caused aggregate accumulation, whereas surplus MSP (e.g. 1:5:200) encouraged empty-ND formation. Optimal reconstitution was found at a ratio of 1:2:100, producing a monodisperse trace lacking both aggregates and unpopulated discs. Success was further verified by peak fraction SDS-PAGE (Figure 6.6[B]), which also highlighted PepTSt can form lipid-free complexes with MSP2N2 (see 1:2:0), albeit less efficiently. The delayed SEC elution of these "minimal NDs", relative to lipid-furnished assemblies, argues against the above concerns surrounding severe under-lipidation.

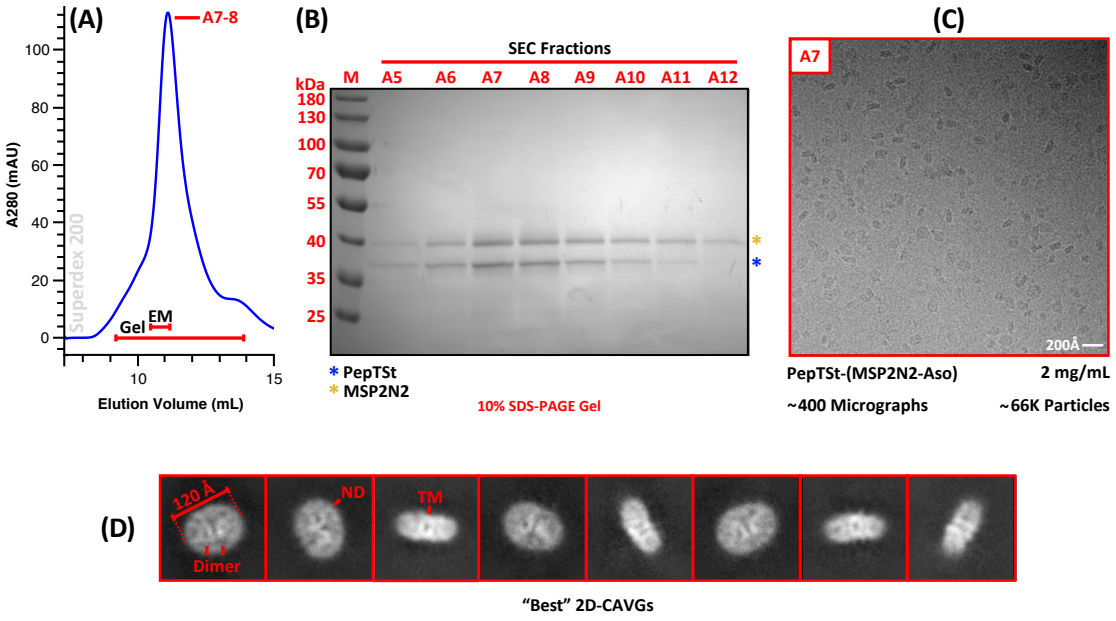


Figure 6.7: Cryo-EM Screening of PepTSt-MSP2N2-Asolectin Nanodiscs

(A) SEC profile of PepTSt-MSP2N2-asolectin NDs achieved using a 1:2:100 protein:MSP:lipid molar ratio. Red bars identify material visualised on the SDS-PAGE gel of **(B)** or concentrated prior Cryo-EM screening; a representative Titan Krios micrograph from which is presented in **(C)**, alongside collection metadata. **(D)** Best SIMPLE (Reboul et al., 2018) 2D-CAVGs, annotated as in earlier PepTSt datasets. Longest ND dimension is given. Image processing led by Justin Deme.

The 1:2:100 PepTSt-MSP2N2-asolectin system scaled well, generating a comparably favourable chromatogram to screening (Figures 6.7[A]&[B]) when preparing sample for Cryo-EM. In light of the surprising uniformity displayed by the PepTSt-MSP1D1 and PepTSt-MSP1E3D1 volumes, only a small exploratory dataset (~400 micrographs) was collected for PepTSt-MSP2N2; a workable particle density being realised at ~2 mg mL^{-1} (Figure 6.7[C]). Processing the amassed ~66K particles yielded elliptical 2D-CAVGs, once again with a longest ND axis of ~120 Å (Figure 6.7[D]). Anticipating, therefore, a 3D reconstruction undifferentiable from previous, collecting more images was deemed unnecessary.

Unprecedentedly, PepTSt thus appears to template dimensionally indistinguishable asolectin-NDs with all three MSP constructs trialled. The mechanism underlying this remains unclear, but, as noted, direct scaffold-transporter interactions could be important. If not overhanging, rationalising where the additional residues of MSP1E3D1 and, especially, MSP2N2 localise within the final lipopro-

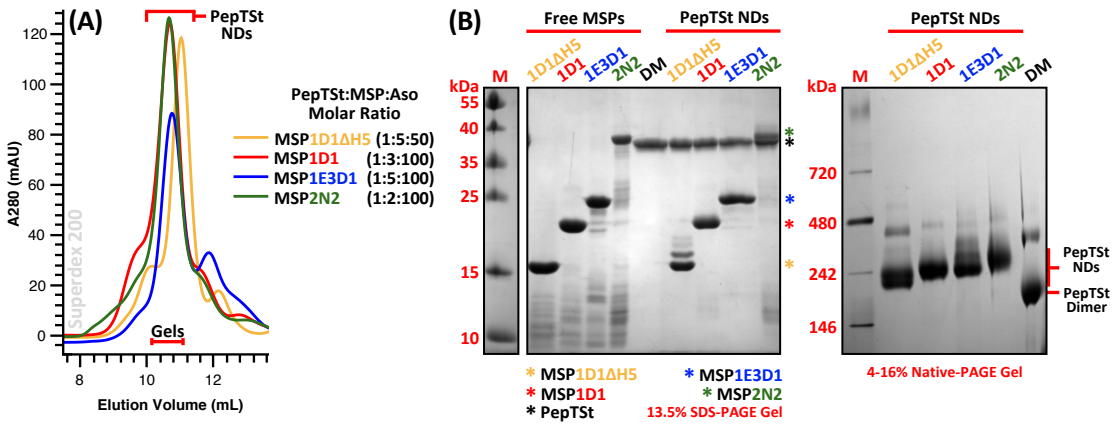


Figure 6.8: Native-PAGE Comparison of PepTSt-MSP-Asolectin Nanodiscs

(A) Overlain S200 traces of PepTSt incorporated into asolectin NDs bounded by four different MSPs; reconstituted at the stated protein:MSP:lipid molar ratios. Red bar denotes peak material visualised on the SDS-PAGE and Novex Bis-Tris BN-PAGE gels displayed in **(B)**. Lanes loaded with free MSPs or DM solubilised PepTSt are provided for internal reference.

tein assemblies is difficult. For the latter it is conceivable, although unlikely, that a single MSP2N2 molecule could encircle PepTSt twice. Of relevance, a recent ND structure of the V_o subcomplex of yeast V-ATPase provides limited precedence for MSP stoichiometry deviations; it resolving three MSP1E3D1 belts, attendant with direct scaffold-membrane protein contacts (Roh et al., 2020). Following up from, and supplementing, the preceding structural characterisation, comparison of different PepTSt-MSP-asolectin NDs by BN-PAGE was attempted.

6.2.4 PepTSt Nanodisc Native-PAGE

Alongside MSP1D1, 1E3D1 & 2N2, PepTSt was also inserted into MSP1D1ΔH5-asolectin NDs prior to BN-PAGE analysis. One of the shortest scaffolds — originally engineered to provide faster tumbling discs for improved resolution NMR spectra — MSP1D1ΔH5 (~8.2–8.4 nm diameter; Hagn et al. (2013)) would not logically be thought able to accommodate the PepTSt dimer. The enigmatic nature of the system prevailing, reconstitution could nevertheless be accomplished at a protein : MSP : lipid molar ratio of 1:5:50 (Supplementary Figures B.14[A]&[B]). All four freshly prepped PepTSt-MSP-asolectin ND variants gave acceptable SEC profiles (Figure 6.8[A]), and peak fractions were then concomitantly visualised

by SDS- and BN-PAGE (Figure 6.8[B]). The latter showed a series of reasonably sharp ND bands migrating with the expected incremental increase in mass from PepTSt-MSP1D1ΔH5 to PepTSt-MSP2N2. All also ran above the dimeric DM solubilised PepTSt reference, reaffirming physical association with the MSPs; each clearly detected in the samples separated by SDS-PAGE. Assuming the comparable proportions of the three structurally studied NDs implies equally comparable lipid carriage, the observed differences in mass are primarily attributable to differential MSP length.

6.2.5 Empty Nanodisc Negative-Stain EM

In an effort to ascertain whether the unchanging dimensions of the PepTSt-NDs are artefactual, empty asolectin-NDs were assembled from the four MSPs and subject to neg-stain for size comparison. The vacant discs, reconstituted in absence of PepTSt and with the lipid : MSP molar ratio increased by 50% to compensate, gave moderately broad S200 profiles (Figures 6.9[A]&[B]). Peak material was then diluted to ~0.02 mg mL^{-1} prior to grid application, negatively stained with uranyl acetate, and ~20 micrographs of each sample collected on a 200 kV Talos instrument. As a comparator, PepTSt-MSP1D1ΔH5-asolectin NDs were also imaged.

On visual inspection, representative micrographs (Figure 6.9[C]) are generally in good agreement with the literature — MSP1D1ΔH5 bounding the smallest empty NDs, MSP2N2 the largest, and MSP1D1 & 1E3D1 indistinguishably intermediate between the two. This was similarly supported on estimating variation in ND longest axis from 2D-CAVGs generated using EMAN2 (Tang et al., 2007) (Figure 6.9[D]). Although an inherently imperfect approach, given the various confounding effects of staining (e.g. contributions from grain size, unequal stain deposition, stain-induced aggregation), the ranges obtained are semi-consistent with previous analyses of empty NDs by TEM (see Hagn et al. (2013)) and likewise agree with reports identifying size heterogeneity in peak ND SEC

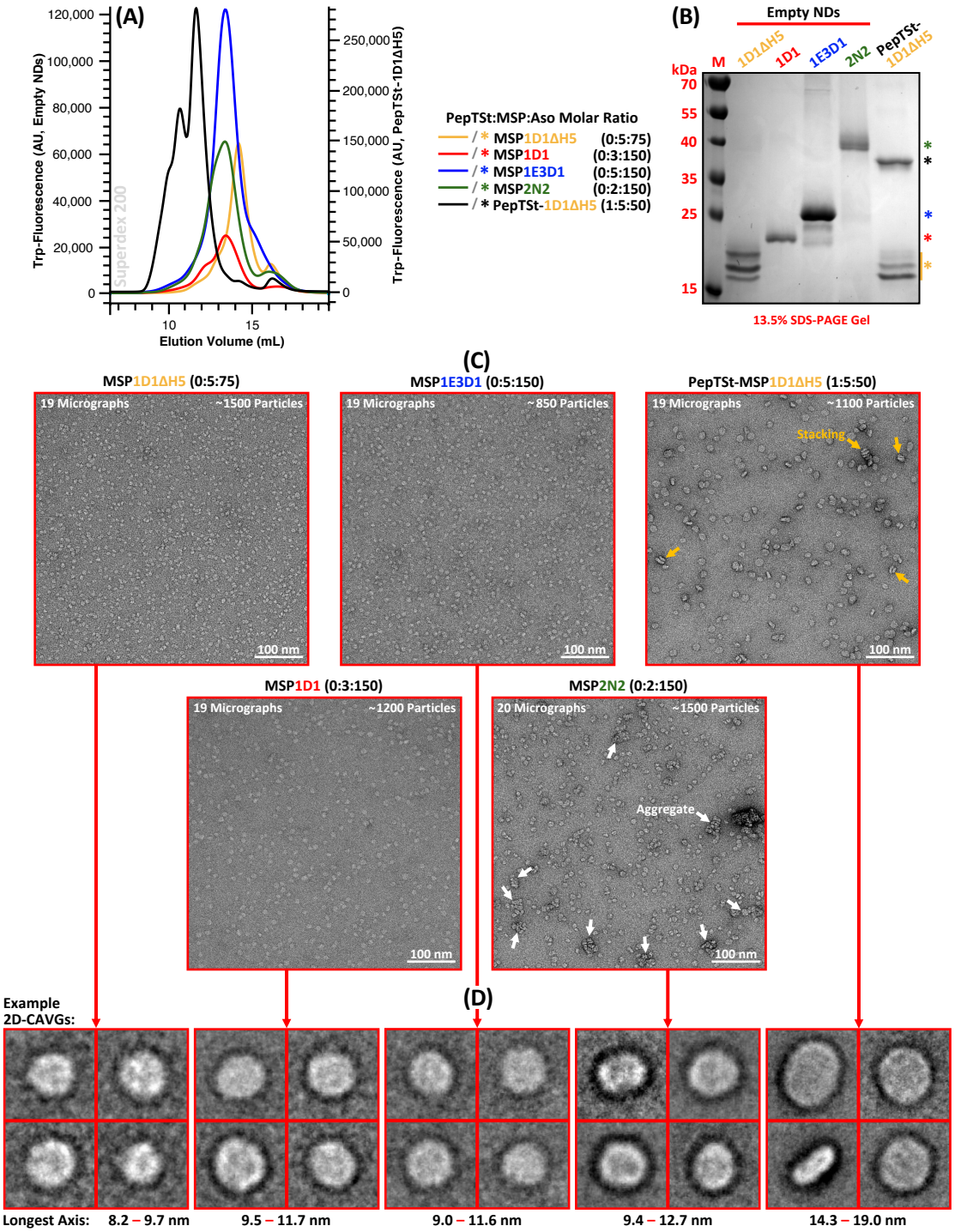


Figure 6.9: Empty Asolectin-Nanodisc Neg-Stain Analysis

(A) Superimposed SEC profiles of empty NDs, constructed from the four MSPs, as well as PepTSt-MSP1D1ΔH5-asolectin; prepared at the stated protein:MSP:lipid molar ratios. Peak contents is analysed on the SDS-PAGE gel of **(B)**, and was diluted to ~0.02 mg mL^{-1} prior to negative-staining with uranyl acetate. Representative micrographs are presented in **(C)**. Images are labelled with collection metadata, and annotated to highlight aggregation or ND stacking where appropriate. **(D)** Example 2D-CAVGs from each dataset generated within EMAN2 (Tang et al., 2007). The range in longest axis, estimated from measures of five 2D-CAVGs, is also given. Image acquisition and processing guidance from Justin Deme.

fractions (Skar-Gislinge et al., 2018). PepTSt-MSP1D1ΔH5-asolectin NDs were largest by far and had noticeably altered on-grid behaviour relative to the unpopulated NDs; displaying reduced preferential orientation as well as emergence of so called "rouleaux" structures — extraneous stacks of NDs thought to be bridged by cationic stain interactions (McLean et al., 2018). Overall, the empty asolectin-ND results are reassuringly compatible with accepted notions of MSP-determined disc diameter; again suggesting PepTSt is somehow responsible for the dimensional uniformity of NDs it is reconstituted into.

6.2.6 PepTSt Nanodiscs: A Summary

Whilst studies of the PepTSt-MSP-asolectin ND system proved structurally fruitful, the generated volumes regrettably failed to provide new biological insights into the role of bilayer interactions in the POT family; structures being near-identical to those previously solved in DDM. The PepTSt-MSP1D1 map does additionally contain putative lipid densities, and these are positionally consistent with MAGs co-crystallised in a collection of the many published *in meso* PepTSt structures. Nevertheless, from a methodological standpoint the work synergises well with a recent structure of another POT — $PepTSo_2$ — obtained in saposin-lipid nanoparticles (Nagamura et al., 2019). Together the studies highlight the value of exploring membrane mimetic technologies when trying to resolve small, oligomeric, MFS transporters by Cryo-EM. The constancy of PepTSt ND size on reconstituting into different MSPs is also a potent reminder, alongside other examples (e.g. Glt_{Tk}-MSP2N2 Arkhipova et al. (2020), yeast V_o-MSP1E3D1 Roh et al. (2020)), that the ND platform does not always behave as predicted. Edifying detour into PepTSt complete, discussion now returns to MmpL3 and efforts made in applying NDs to the less well characterised target.

6.3 MmpL3 Nanodiscs

6.3.1 Reconstitution Screening

As for PepTSt, attempts began by screening for incorporation of DDM purified $MmpL3_{1\text{-}753}$ into MSP1D1-asolectin NDs. Unlike the POT however, and despite trialling ten protein : MSP : lipid molar ratios alongside the standard controls (i.e. 1:2:0 & 0:2:100), optimal conditions — a SEC monodisperse reconstitution unpolluted by neither aggregates nor empty discs — could not be found. Exploring ratios of the form 1:2:X gave polydisperse traces with abundant lipid-rich species leading from, or entirely within, the void (Figure 6.10[A]). Raising the proportion of MSP to assemble at ratios of 1:5:X did yield improvements, but also concomitant augmentation of peaks presumed to be vacant NDs or lipid-free scaffold (Figure 6.10[B]). All 1:5:X profiles showed a pair of peaks eluting between ~11-14 mL, containing both the truncation and MSP1D1 by SDS-PAGE (Figure 6.10[C]). This indicated formation of a heterogeneous ND population. Concerned, therefore, that the first peak featured material reminiscent of the $MmpL3_{1\text{-}753}$ purification shoulder (see Chapter 3), the ratio maximising reconstitution into the second species — 1:5:100 — was chosen for subsequent Cryo-EM endeavours. The silver stained gel likewise detected truncation in the peak assigned to "empty" NDs, albeit at reduced levels and with more MSP as compared to earlier fractions; a likely result of insufficient column resolution between target-containing and truly unpopulated discs.

After downstream experiments with $MmpL3_{1\text{-}753}$-MSP1D1-asolectin NDs, but salient to outline now, reconstitution into additional scaffolds and/or using target purified in different detergents was pursued. Screening insertion into MSP1D1ΔH5 (Figure 6.11[A]) or MSP2N2 (Supplementary Figures B.14[C]&[D]) from DDM solubilised protein produced multimodal traces with doublet ND peaks analogous to the original MSP1D1 trials. The "best" protein : MSP : lipid molar ratios were 1:5:25 and probably 1:5:100, for the shorter and longer MSPs

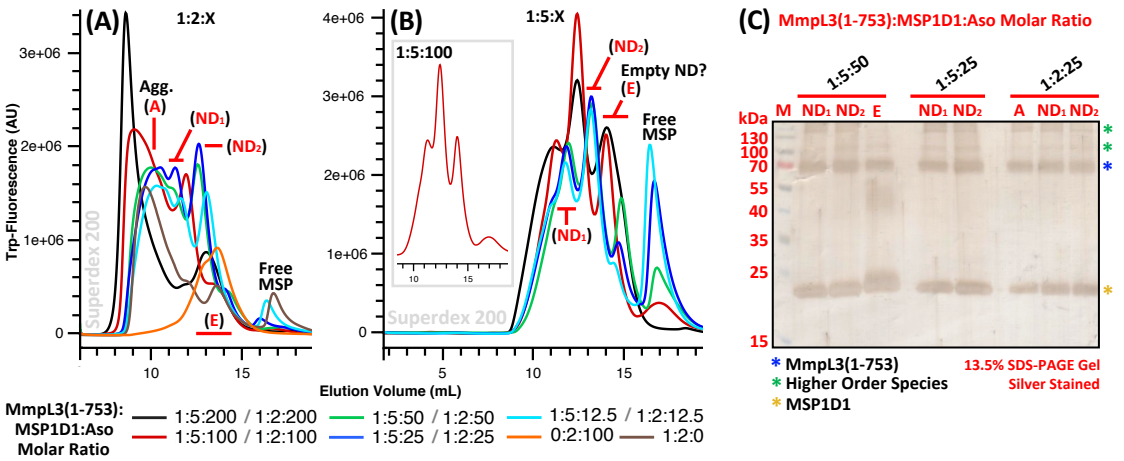


Figure 6.10: Screening MmpL3$_{1\text{-}753}$-MSP1D1-Asolectin Nanodisc Reconstitution
(A) Overlain SEC profiles screening reconstitution of DDM purified MmpL3$_{1\text{-}753}$ into MSP1D1-asolectin NDs at a range of protein:MSP:lipid molar ratios of the form 1:2:X. Labels identify peaks presumed to be eluting aggregates (Agg.), target-containing NDs (ND_1 & ND_2) or empty NDs (E). Peaks assigned to free MSP are also annotated. Equivalent traces summarising screening at molar ratios of 1:5:X are presented in **(B)**. Inset: S200 profile of "best" ratio, scaled up for downstream investigations. **(C)** Silver stained SDS-PAGE gel of material taken from choice screening peaks, identified with the same descriptors as in the chromatograms.

respectively. A similar trend of worse quality reconstitutions at ratios spanning 1:2:X was evident.

Given the mediocre SEC profiles of NDs assembled from DDM purified MmpL3$_{1\text{-}753}$, and recalling the greatly improved protein recovered in LMNG (see Chapter 3), trialling incorporation from the MNG class detergent seemed worthwhile. Although its low CMC (0.001% w/v) means it is highly unlikely to be fully removed by Bio-Bead incubation, leading some to rightly question LMNG compatibility with the MSP platform (Autzen et al., 2019), a number of successful ND reconstitutions involving the detergent have been reported (e.g. Kraft et al. (2015); Laverty et al. (2019)). Exploratory tests employing MSP1D1 and MSP1D1ΔH5 both returned bimodal traces (Figure 6.11[B]). Robust formation of the first of two sharp peaks appeared lipid dependent on comparison to the 1:2:0 control; lending some confidence, alongside SDS-PAGE contents (Figure 6.11[C]), to at least semi-typical ND assembly. The second peak eluted at a volume most akin to free MSP, but was more prominent and shifted leftward relative to screening

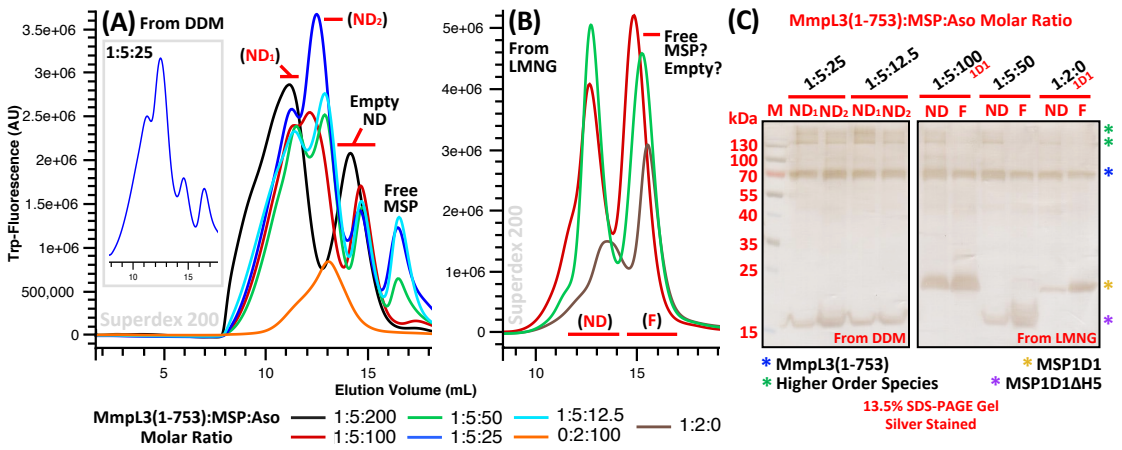


Figure 6.11: Further MmpL3$_{1\text{-}753}$-MSP-Asolectin Nanodisc Reconstitution Screens
(A) S200 traces screening MmpL3$_{1\text{-}753}$ incorporation into MSP1D1ΔH5-asolectin NDs at the given protein:MSP:lipid molar ratios, using target solubilised in DDM. Inset: "Best" ratio, later used for CD and sybody complexation tests. Annotations denote peaks assumed to be target-containing NDs (ND$_1$ & ND$_2$), empty NDs, or free MSP (F). **(B)** Overlay of SEC profiles trialling LMNG purified MmpL3$_{1\text{-}753}$ reconstitution into MSP1D1 (red/brown) or MSP1D1ΔH5 (green) asolectin NDs. **(C)** Silver stained SDS-PAGE gel of various peak fractions, labelled as in the chromatograms.

from DDM — potentially hinting at MSP-LMNG interactions, in view of the lipid-like structure of the detergent. Confounding matters, this peak also contained MmpL3$_{1\text{-}753}$ by silver stain. Such peculiarities could be effects of residual LMNG and/or the aforementioned issues of column resolution. As only the earlier species was abundant on scaling up the 1:5:50 MSP1D1ΔH5 condition, further conjecture is avoided.

MmpL3$_{1\text{-}753}$ had proven amenable to asolectin-ND incorporation using a variety of MSP constructs and from two purification detergents, but it remained to be determined whether lipoprotein reconstitution would ameliorate target structural intractability. Hence, the final sections of the chapter detail investigation of truncation ND stability, homogeneity, Cryo-EM viability, and sybody binding competence.

6.3.2 Characterisation

CD-Melts & Native-PAGE

Rationalising that the tighter packing of the smallest MmpL3$_{1\text{-}753}$ NDs may result in greater stabilisation and thus Cryo-EM utility, truncation MSP1D1ΔH5-asolectin discs — reconstituted from both DDM and LMNG — were chosen for characterisation by CD and BN-PAGE. Scaling up of "optimal" ratios yielded chromatograms somewhat altered relative to screening; assembly from the alkyl maltoside giving dual ND peaks but with abundant aggregates, and from the MNG detergent a near monodisperse profile shouldered by an additional minor species (Figure 6.12[A]). Subsequent BN-PAGE gave semi-sharp ND bands displaying the expected MSP-lipid mass increase relative to detergent solubilised truncation (Figure 6.12[B]). Earlier eluting material in the lesser ND peak (fraction "A7"), from DDM or LMNG, showed a dominant band with significantly retarded migration as compared to later eluting assemblies (fraction "Peak"). Whether this represents adventitious entrapment of multiple truncation monomers into a single disc as opposed to *bona fide* — physiologically relevant and possibly lipid-induced — MmpL3$_{1\text{-}753}$ oligomerisation is unclear without further study. It would undoubtedly be enlightening to pursue ND incorporation with a mycobacterial lipid extract should the platform be revisited in the future. In any case, contrasting the MmpL3$_{1\text{-}753}$ ND and PepTSt ND BN-PAGE results indicates, concordant with SEC, much greater heterogeneity in former; mirroring the properties of each protein at purification.

Similar to protein-detergent complexes, CD melt-determined thermal stability has been advocated as a useful ND quality metric (Hagn et al., 2018). However, it is reliant on the α-helical MSP unfolding transition being sufficiently higher than that of the encircled target protein as to ensure facile deconvolution of each. As NDs have been reported to melt at >80 °C (Hagn et al., 2018) it was hoped the technique would be readily applicable to the truncation, which in detergent was previously seen to unfold at ~48 °C (DDM) & ~54 °C (LMNG, see Chapter 3).

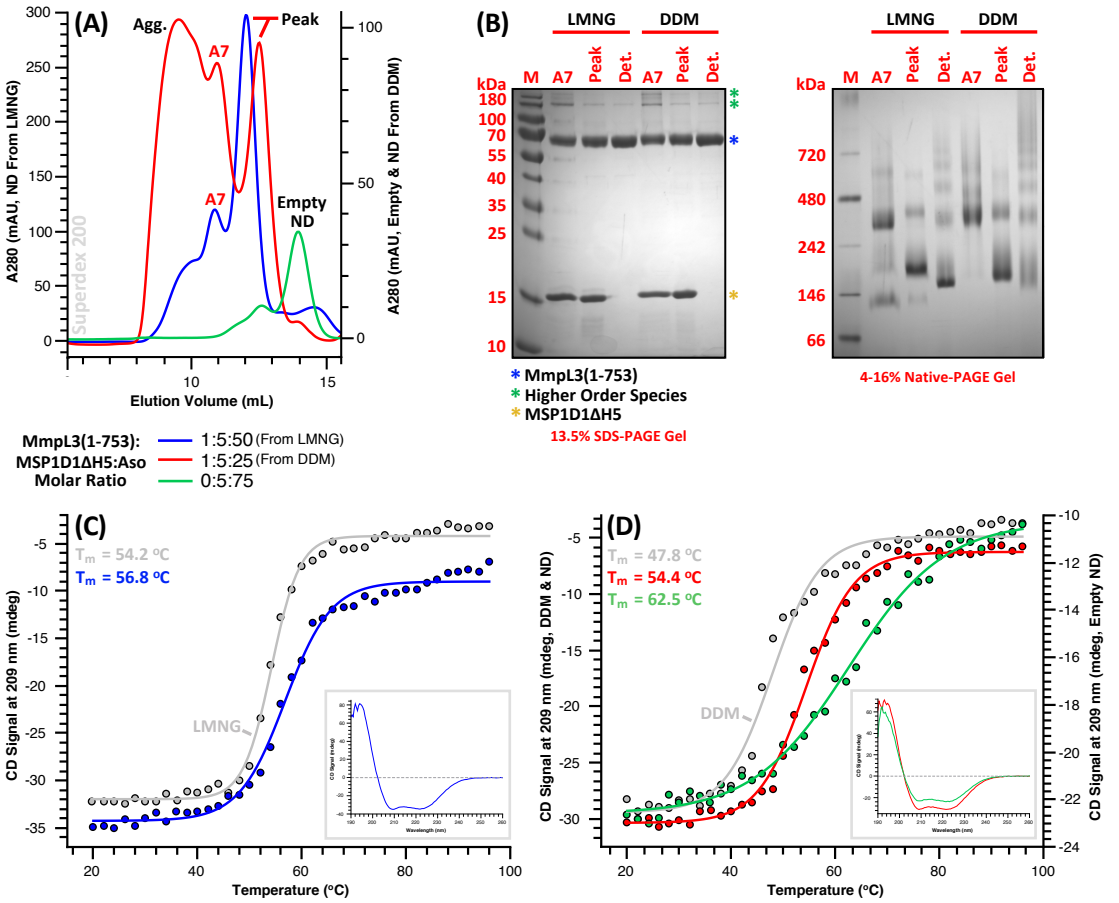


Figure 6.12: MmpL3$_{1\text{-}753}$-MSP1D1ΔH5-Asolectin Nanodisc Characterisation

(A) SEC profiles of MmpL3$_{1\text{-}753}$-MSP1D1ΔH5-asolectin NDs reconstituted from DDM or LMNG purified protein, assembled at protein:MSP:lipid molar ratios of 1:5:25 and 1:5:50 respectively. Abundant aggregate (Agg.) is annotated where present and an empty ND trace, formed at 0:5:75, is overlain. Chromatogram labels denote material analysed in the remainder of the figure. **(B)** Novex Bis-Tris BN-PAGE comparison of detergent solubilised (Det.) and MSP1D1ΔH5 ND incorporated truncation. The same samples separated by SDS-PAGE are also shown. **(C)** CD melt of peak MmpL3$_{1\text{-}753}$-MSP1D1ΔH5 NDs, produced from LMNG purified protein, on heating from 20-95 °C in 2 °C increments (blue). Melt temperature (T_m) derived from sigmoidal fit in QtiPlot at 209 nm. The detergent melt profile, from Figure 3.11, is plotted as reference (grey). Inset: ND CD scan pre-melt. Equivalent data for NDs assembled from DDM solubilised protein (red), as well as the empty MSP1D1ΔH5-asolectin discs (green), is presented in **(D)**. For full CD melt spectra see Supplementary Figure B.15.

To check, an empty ND control was melted in parallel with the MmpL3$_{1\text{-}753}$-MSP1D1ΔH5-asolectin assemblies (Figures 6.12[C]&[D]). The vacant discs gave a broad transition with a T_m of ~63 °C, providing an unfortunate partial overlap, and likely interference, with target unfolding signal. Melt temperatures recorded for truncation NDs reconstituted from DDM and LMNG purified protein were ~54 °C & ~57 °C respectively. Tentatively this suggests ND incorporation sta-

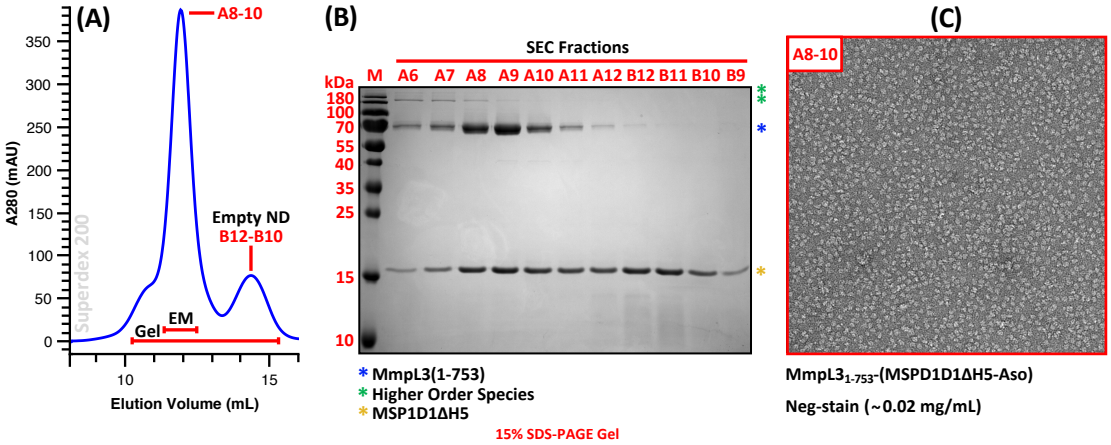


Figure 6.13: Neg-Stain of MmpL3$_{1-753}$-MSP1D1ΔH5-Asolectin Nanodiscs

(A) S200 profile of MmpL3$_{1-753}$-MSP1D1-asolectin NDs reconstituted from LMNG purified truncation, at a protein:MSP:lipid molar ratio of 1:5:50. Red bars denote fractions ran on the SDS-PAGE gel of **(B)** or diluted to ~0.02 mg mL^{-1} for uranyl acetate negative-staining; a representative micrograph from which is presented in **(C)**. Imaged by Justin Deme.

bilises MmpL3$_{1-753}$, especially relative to the less stabilising alkyl maltoside, but given the nature of the MSP1D1ΔH5 transition such conclusions remain infirm. Furthermore, the observation that NDs derived from LMNG solubilised truncation exhibit slightly higher thermal stability, than those reconstituted from DDM, again brings into question the influence of residual detergent on the system.

Negative-Stain EM

To obtain an indication of the on-grid homogeneity achievable with MmpL3$_{1-753}$ NDs, and thus ascertain whether any stability gains could even be structurally exploited, truncation MSP1D1ΔH5-asolectin lipoproteins assembled from LMNG were assessed by neg-stain. Highlighting batch-to-batch ND variation, preparative reconstitution produced a near-ideal monodisperse trace accompanied by only a small population of empty discs (Figures 6.13[A]&[B]). Visualisation of peak material stained with uranyl acetate showed a field of compact, uniform, particles (Figure 6.13[C]) and so proceeding to vitrification trials was justified.

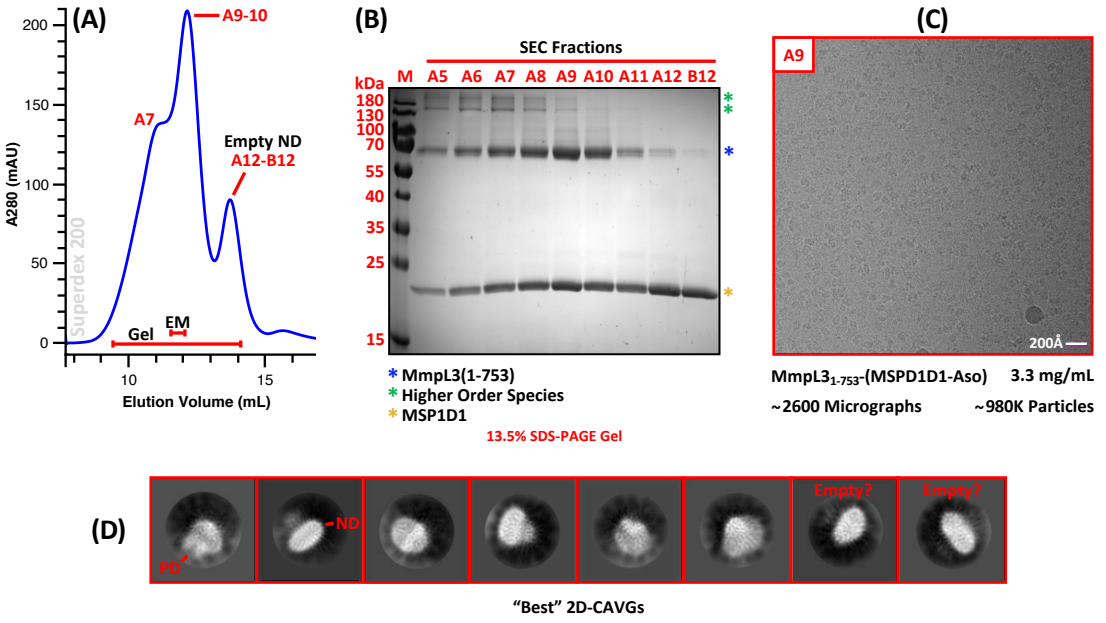


Figure 6.14: Cryo-EM Screening of MmpL3$_{1\text{-}753}$-MSP1D1-Asolectin Nanodiscs

(A) S200 chromatogram of MmpL3$_{1\text{-}753}$-MSP1D1-asolectin NDs reconstituted, from DDM purified truncation, at a protein:MSP:lipid molar ratio of 1:5:100. Red bars correspond to fractions loaded on the SDS-PAGE gel of **(B)** or concentrated for Cryo-EM; an exemplar Titan Krios micrograph from which is presented in **(C)**, together with dataset specifics. **(D)** Best RELION (Scheres, 2012) 2D-CAVGs. Density corresponding to the periplasmic domain (PD), or ND bulk, is annotated. Potentially empty discs are also highlighted. Processing guidance from Justin Deme.

6.3.3 Cryo-EM Screening

MmpL3$_{1\text{-}753}$-MSP1D1-Asolectin NDs from DDM

Being the first MmpL3$_{1\text{-}753}$ ND assembled, the MSP1D1-asolectin discs reconstituted from DDM purified protein were also the first imaged by Cryo-EM. The scaled up S200 profile remained comparable to screening, though the earlier eluting peak became less resolved and presented as a shoulder to the main species (Figures 6.14[A]&[B]). Excellent in-ice particle density was realised at ~3.3 mg mL^{-1}, and on collecting ~2600 micrographs an initial dataset of almost 1 million particles was accumulated (Figure 6.14[C]). Processing these generated rather dismal 2D-CAVGs (Figure 6.14[D]), undeniably worse than the original LMNG attempt (see Chapter 4). Whilst the periplasmic subdomains of a number classes betray successful ND insertion of MmpL3$_{1\text{-}753}$, many lack discernible features and are thus indistinguishable from empty discs. Pollution of the sample by such vacant

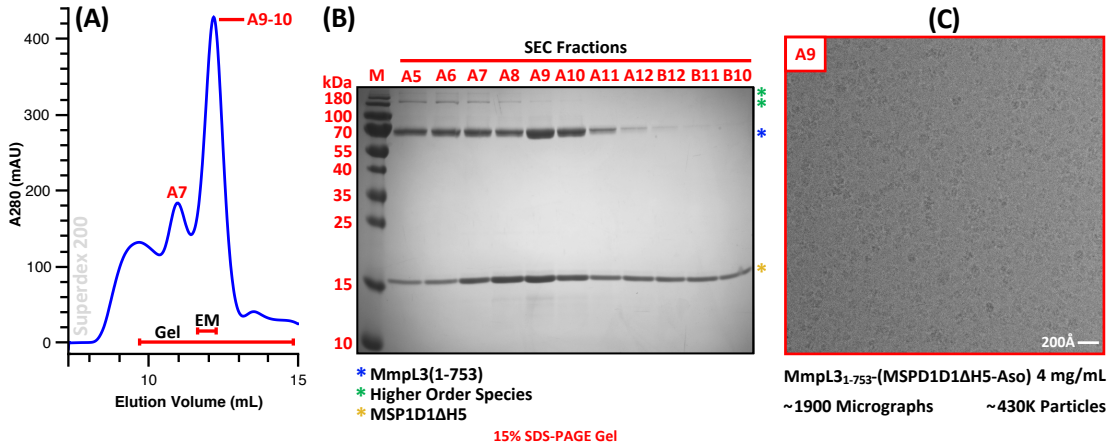


Figure 6.15: Cryo-EM Screening of MmpL3$_{1\text{-}753}$-MSP1D1ΔH5-Asolectin Nanodiscs
(A) SEC trace of MmpL3$_{1\text{-}753}$-MSP1D1ΔH5-asolectin NDs assembled, using LMNG purified target, at a 1:5:50 protein:MSP:lipid molar ratio. Red bars denote material visualised on the SDS-PAGE gel of **(B)** or prepped for Cryo-EM; a representative Titan Krios micrograph from which is shown in **(C)**, alongside collection metadata.

species may have frustrated image alignment, and would account for the poor processing results. Although not tested here, an avenue to alleviate this would be to reconstitute a tagged (e.g. FLAG, hexa-histidine) MmpL3$_{1\text{-}753}$ construct and then employ affinity purification to isolate only target-containing NDs (Efremov et al., 2017). Instead, Cryo-EM of a MSP1D1ΔH5-asolectin sample — reconstituted from LMNG — was prioritised, given the aforenoted monodispersity and homogeneous neg-stain micrographs.

MmpL3$_{1\text{-}753}$-MSP1D1ΔH5-Asolectin NDs from LMNG

On forming MSP1D1ΔH5 NDs for vitrification the post-assembly chromatogram departed from the ideality seen for the neg-stain preparation, more resembling the reconstitution undertaken for characterisation by CD and BN-PAGE (Figures 6.15[A]&[B]). Nonetheless, the S200 profile displayed less heterogeneity than the MSP1D1 discs already screened in-ice and grids frozen at ~4 mg mL^{-1} proved viable for dataset collection (Figure 6.15[C]). Despite appreciably better SEC dispersity, and potentially higher thermal stability, of the MSP1D1ΔH5 NDs from LMNG, the ~430K particles recovered failed to produce intelligible 2D-CAVGs. Whilst suboptimal ice thickness was thought to have been a compounding

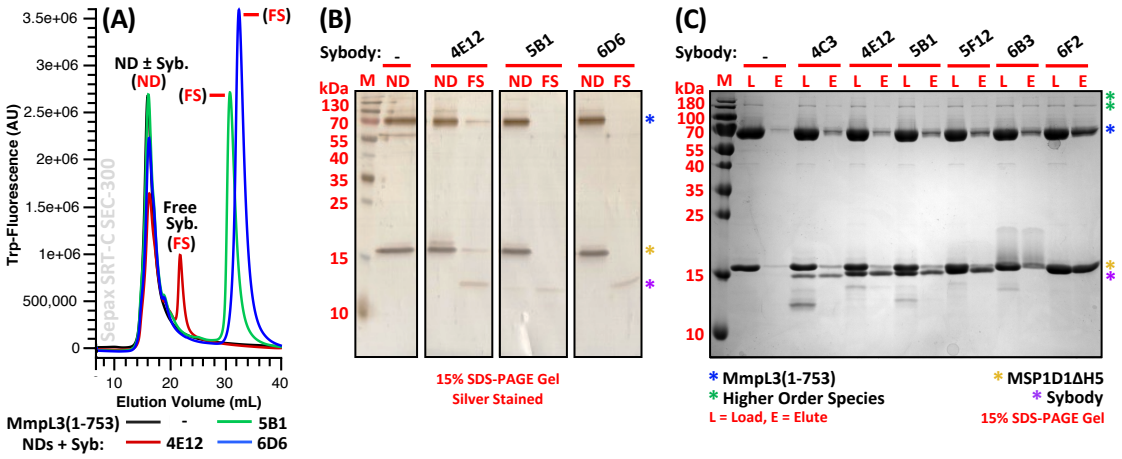


Figure 6.16: MmpL3$_{1\text{-}753}$-MSP1D1ΔH5-Asolectin Nanodisc Sybody Complexation

(A) Sepax SRT-C SEC-300 traces of MmpL3$_{1\text{-}753}$-MSP1D1ΔH5-asolectin ND co-elution trials with three sybodies, one from each library. NDs assembled from DDM purified protein at a 1:5:25 protein:MSP:lipid molar ratio, and peak SEC material incubated with tag-free sybody in a 1:1.5 ND:sybody molar ratio prior to reinjection. Chromatograms annotations delineate ND ± sybody and free sybody (FS) peaks. Material from both is visualised on the silver stained SDS-PAGE gel of **(B)**. Nickel resin pull-down of truncation NDs, reconstituted from LMNG purified protein in Figure 6.13, by six sybody-Myc-His$_6$ constructs is shown on the SDS-PAGE gel of **(C)**. In some lanes MSP and sybody co-migrate. Two sybodies per library were tested, incubated at a 1:1.5 ND:sybody molar ratio before bead addition.

factor, alongside the MSP1D1 results it became evident that NDs were unlikely to provide a route to truncation structural studies. Before abandoning MSPs completely, sybody complexation with ND reconstituted MmpL3$_{1\text{-}753}$ was explored as means to provide both extra bulk and fiducial aid to particles; equivalent to the original attempts with detergent solubilised target.

6.3.4 Sybody Complexation

As in Chapter 5, sybody–ND association was assayed by SEC co-elution coupled with peak silver staining (Figures 6.16[A]&[B]), as well as nickel resin pull-downs (Figure 6.16[C]). The outcome, predating discovery of MG2 (see Section 5.5.3), was equivalent to the original binding attempts with LMNG purified MmpL3$_{1\text{-}753}$, in that tight binders to ND reconstituted truncation, capable of co-elution, could not be found. Pull-downs confirmed sybodies retained some affinity for lipoprotein incorporated target, but at a greatly reduced level relative to DDM, and

possibly even LMNG, solubilised protein (contrast to pull-downs with the same sybody set seen in Figure 5.15). The only exception was convex library sybody "6F2" which, nevertheless failing to co-elute, displayed the best pull-down behaviour; comparable between NDs and the MNG detergent. Presentation of epitopes bound by the DDM-selected sybodies therefore appears similarly curtailed in NDs and LMNG. Analogously, this may be ascribed to the MSP-lipid environment sterically hindering epitope access, modulating the conformational dynamics of reconstituted $MmpL3_{1\text{-}753}$ in manner decreasing cognate epitope availability, and/or dissuading partial unfolding events required for linear epitope formation. Regardless of underlying cause, unable to form SEC stable sybody complexes, further ND experimentation was halted.

6.3.5 Conclusion

Overall, application of ND technology to $MmpL3_{1\text{-}753}$ proved ineffectual in improving the Cryo-EM viability of the target, in turn prohibiting realisation of the more ambitious goal of structurally describing any protein-lipid interactions. Progress was ultimately stymied by a combination of sample heterogeneity, possibly related to empty ND contamination, and sybody incompatibility. Whilst the truncation was successfully reconstituted into ND bilayers encircled by three MSP constructs, the host lipid — asolectin — remained constant throughout. Not exchanging this for a lipid mixture more biologically mimetic of the *Mtb* plasma membrane is one of only a few ND avenues left unexhausted. Unless inducing some form oligomerisation or significant protein stabilisation however, it is unlikely to be redemptive. Taking an overview of $MmpL3_{1\text{-}753}$ reconstitutions clearly highlights the influence of upstream purification detergent on subsequent ND assembly, and should be of particular consideration when using low CMC detergents where residual surfactant may remain. In a final effort to make structural inroads, Chapter 7 next describes successful attempts to leverage the methodological developments and hardware improvements emergent since collection of the original LMNG Cryo-EM dataset.

The MmpL3 Truncation

Structure Determination & Analysis

CHAPTER 7

Alluded to at the end of the Introduction, during the course of my DPhil significant insights into the architecture of MmpL3, as well as its interactions with multiple inhibitors, were provided by elegant studies reporting crystal structures of the *Msmg* orthologue (61% sequence identity to *Mtb*). The first of these — Zhang et al. (2019) — was published ~1.75 years into the project, and was subsequently followed up by two further publications — Su et al. (2019) & Yang et al. (2020). Nevertheless, given the severity of the TB pandemic, the evident druggability of MmpL3, and the numerous medicinal chemistry campaigns against it, solving the structure of the *bona fide* antitubercular target remained a high-priority. Consequently, $MmpL3_{1\text{-}753}$ continued to be my primary research focus.

By now attempts to structurally characterise the truncation had encompassed crystallographic (VD & LCP) and/or Cryo-EM screening of the protein in various detergents (DM, DDM & LMNG; Chapter 4), in a detergent-alternative (Amphipol A8-35; Chapter 4), in the presence of additives with or without cross-linking (Haem ± Glutaraldehyde; Chapter 4), in complex with *in vitro* selected binders (Sybodies; Chapter 5), and post-reconstitution into state-of-the-art membrane mimetics (MSP NDs; Chapter 6). Out of all of these, the greatest structural information had been garnered from 2D-CAVGs of $MmpL3_{1\text{-}753}$ imaged in LMNG. Hoping a combination of microscopic and algorithmic advancements, largely emergent in the >2.4 years since collection of the original dataset, could be leveraged to resolve the sample further, fresh protein was purified in LMNG and the most substantive dataset of the thesis collected.

Persistence paid off. This final chapter reports the Cryo-EM structure of *Mtb* MmpL3$_{1\text{-}753}$ at a global resolution of 3.0 Å. At only ~82 kDa the truncation is among the smallest membrane proteins, neither benefiting from exploitable intrinsic symmetry nor the fiducial aid of a binder, solved by the technique to date (Nygaard et al., 2020). The resulting atomic model reveals an extensive cavity network of putative functional relevance as the TMM conduit. An observation further corroborated by comparison with structures of other lipophile transporting RND superfamily members, including the *Msmg* homologue. Examination of contacts made by an LMNG molecule, resolved within said cavity, potentially sheds light on the interactions employed by the transporter to coordinate its native substrate. The chapter goes on to detail the most comprehensive structural mapping of MmpL3 missense variants ever undertaken — analysing over 100 unique resistance-conferring substitutions selected by exposure to preclinical agents, in addition to non-synonymous mutations mined from >45,000 whole-genome sequenced *Mtb* clinical, research, and environmental isolates. In conjunction with the *Msmg* models, this work provides a robust foundation for both functional dissection of, and structure-based drug design against, *Mtb* MmpL3; guided by an enhanced appreciation of the underlying mutational landscape.

7.1 An Aside on Structures of *Msmg* MmpL3

Before discussing the Cryo-EM reconstruction of MmpL3$_{1\text{-}753}$ in earnest, a brief background on the *Msmg* studies is given, focussing on insights derived by contrasting the methods employed (e.g. aspects of construct design, protein preparation, crystallisation etc.) with those advanced for *Mtb* MmpL3 in previous Chapters. In-depth analysis of the *Msmg* findings is presented by comparison with the *Mtb* structure in later Sections (see 7.2 & 7.3). In total eight *in surfo* structures of *Msmg* MmpL3 have been reported (Supplementary Figure B.16), spanning resolutions of 2.6–3.1 Å (Zhang et al., 2019; Su et al., 2019; Yang et al., 2020). Two are in *apo* states, and the other six complexed with different

preclinical inhibitors belonging to five scaffold classes (in brackets) — Rimonabant (Pyrazole), SQ109 (Ethylenediamine), ICA38 (Indolecarboxamide), NITD-349 (Indolecarboxamide), AU1235 (Urea Derivative) & SPIRO (Spirocycle).

Because its status as a novel antitubercular target is the *raison d'être* of MmpL3 research, one can assume the authors of the *Msmg* structures also found *Mtb* MmpL3 recalcitrant to crystallisation before pursuing a non-pathogenic homologue as substitute. Nonetheless, both proteins must present comparable experimental challenges as all the *Msmg* constructs bear strikingly similar C-terminal truncations to MmpL3$_{1\text{-}753}$, either alone (*Msmg* residues 1–773, 1/8 structures, Su et al. (2019)) or supplemented with a C-terminal T4-lysozyme fusion (*Msmg* residues 1–748, 7/8 structures, Zhang et al. (2019); Yang et al. (2020); Supplementary Figure B.16); in each case rationalised on analogous grounds to Section 3.4 (i.e. full-length instability & degradation). Although the least conserved region of the transporter (Supplementary Figure B.18), retainment of CTD disorder between *Mtb* and *Msmg* MmpL3 suggests the attribute is of functional relevance.

The *Msmg* truncations were purified and crystallised in DDM (2/8 structures, Zhang et al. (2019); Su et al. (2019)) or Lauryl-6-Trehaloside (6DDTre, 6/8 structures, Zhang et al. (2019); Yang et al. (2020)). The latter, equivalent to the former maltoside but with the headgroup exchanged for trehalose, is a partial TMM mimetic, and therefore a shrewd detergent choice; one that in hindsight should have been considered for MmpL3$_{1\text{-}753}$. However, that DDM solubilised *Mtb* MmpL3 — even when further truncated and fused with the muramidase (MmpL3$_{1\text{-}743}$-Lysozyme) to replicate the Zhang et al. (2019) construct — failed to crystallise in screens reproducing the published *Msmg* conditions (data not shown), and protein concentrations (8–20 mg mL^{-1}), strongly argues the orthologues are biophysically dissimilar. Hence, MmpL3$_{1\text{-}753}$ may likewise not crystallise in 6DDTre. Moreover, that one *Msmg* structure was solved from protein overproduced in *E. coli* (Su et al., 2019) implies use of the same non-mycobacterial expression host is unlikely to have compromised efforts to crys-

tallise MmpL3$_{1\text{-}753}$. The differential crystallisation propensity of *Mtb* and *Msmg* MmpL3 is thus almost certainly a result of intrinsic differences in orthologue properties (e.g. dynamics, stability, surface charge distribution etc.), highlighting the utility of undertaking a thorough homologue screen at the outset of any structural project. For reference, protein for the other seven *Msmg* structures was over expressed in *Msmg* mc^{2}155 cells (Zhang et al., 2019; Yang et al., 2020), as attempted for *Mtb* MmpL3 in Section 3.3.1.

7.2 Structural Elucidation of Truncated MmpL3

7.2.1 Breakthrough in Truncation Cryo-EM Studies

Initial imaging of MmpL3$_{1\text{-}753}$ in LMNG had given promising 2D-CAVGs (Section 4.2.2), but ultimately faltered on attempts at 3D reconstruction. Buoyed by the increased frequency of Cryo-EM publications reporting sub-100 kDa membrane protein structures in the intervening period (Nygaard et al., 2020), alongside successes in the Newstead Group with other small targets (e.g. Parker et al. (2021)), the sample was revisited. The truncation was prepared and vitrified identically to previous (Figure 7.1[A]&[B]), but was then subject to a systemically updated data collection and processing pipeline; summarised in Figure 7.1[C] and thoroughly detailed in Chapter 2.

The improved frame rate and field of view provisioned by upgrade to a K3 DED, exploited in concert with a beam-image shift acquisition format, significantly expedited data collection; permitting streamlined recovery of >15,000 micrographs. These were preprocessed on the fly in SIMPLE 3.0 (Caesar et al., 2020) to yield in excess of 7.7 million particles, an approximate 86-fold increase on the original dataset. Absent from the workflows enacted in Chapter 4, the movies crucially underwent patched motion correction; ensuring finer correction for beam-induced particle drift from the outset. The benefits of this were evident in early 2D-CAVGs (Figure 7.1[C]), which resolved a host of MmpL3$_{1\text{-}753}$ projections

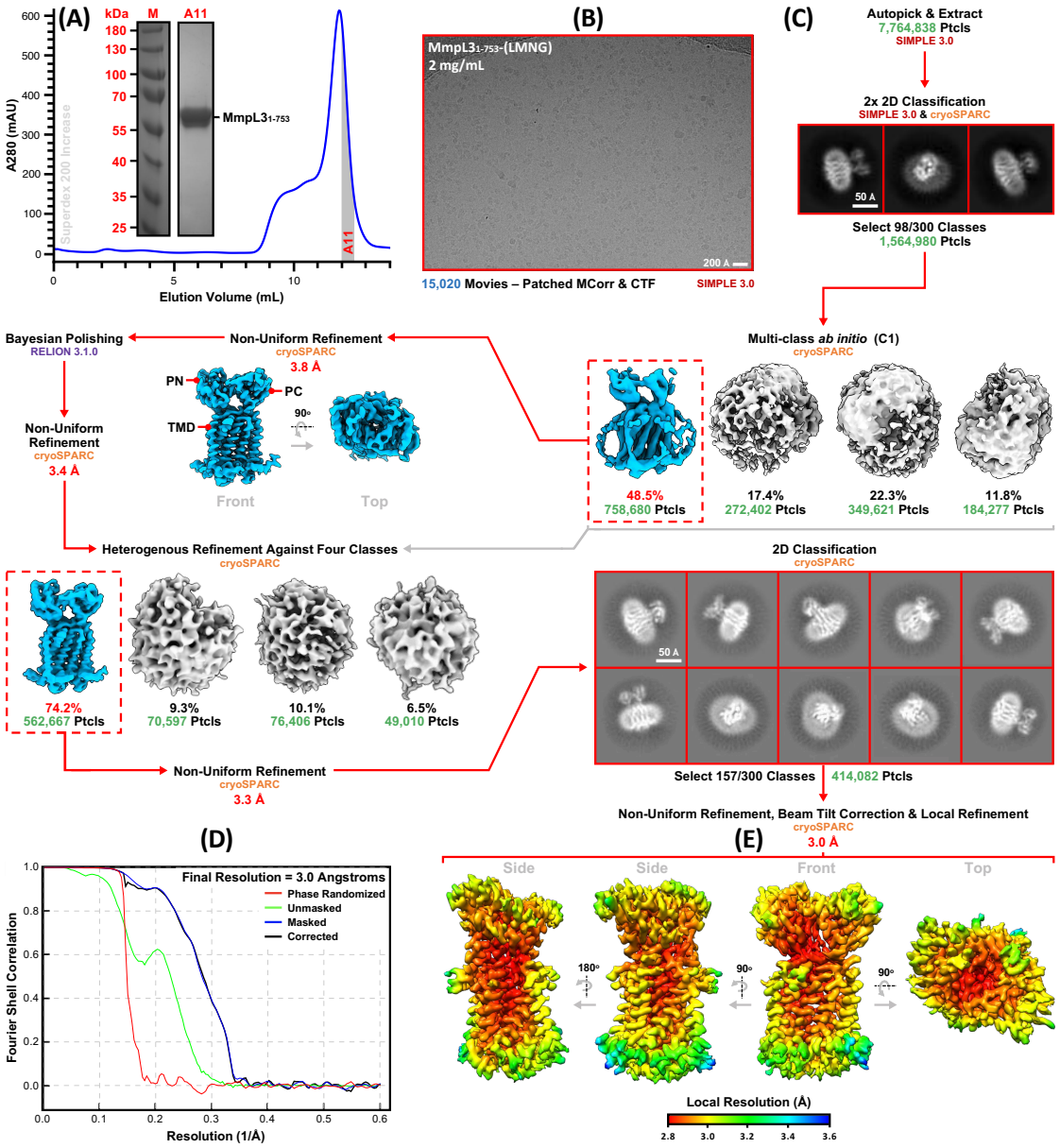


Figure 7.1: $MmpL3_{1\text{-}753}$ LMNG Purification, Cryo-EM Workflow & Map Resolution

(A) S200 profile of LMNG solubilised truncation purified for Cryo-EM studies. The grey bar denotes the SEC fraction ("A11") analysed on the inset 10% SDS-PAGE gel and subsequently concentrated for imaging; a representative Titan Krios micrograph from which is shown in **(B)**. Sample concentration and dataset size are stated. **(C)** Schematic of the Cryo-EM data processing pipeline, spanning autopicking to refinement. The software package(s) employed at each step — SIMPLE 3.0 (Caesar et al., 2020), RELION 3.1.0 (Zivanov et al., 2020), and/or cryoSPARC v.2.15.0 (Punjani et al., 2017) — are noted. The number of particles (Ptcls) retained, and resolution achieved, at various intermediate stages are given. **(D)** FSC plot for the final $MmpL3_{1\text{-}753}$ reconstruction. Global resolution (3.0 Å) is reported at the gold-standard 0.143 cut-off. **(E)** Local resolution estimates for the sharpened truncation map, with micellar density hidden, viewed from four directions. As for the FSC curves, this was computed in RELION 3.1.0. Volumes rendered in UCSF Chimera v.1.13.1 (Pettersen et al., 2004) & ChimeraX v.1.1 (Pettersen et al., 2021). Processing courtesy of Justin Deme.

with discernible secondary structure elements — most notably the separation of individual transmembrane helices (TMs) within the LMNG micelle. Earlier concerns regarding sample in-ice preferential orientation and PD motional freedom are thus certainly unfounded.

Consistent with 2D-CAVG quality, subsequent multi-class *ab initio* model generation within cryoSPARC (Punjani et al., 2017) gave a single featureful volume composed of ~760K particles, displaying the expected MmpL3 domain organisation, and with defined TMs (Figure 7.1[C]). The majority of the downstream processing steps were likewise undertaken in the cryoSPARC suite, largely due to its implementation of non-uniform refinement (NUR) (Punjani et al., 2020) — an approach that actively accounts for spatial variations in map disorder (e.g. as contributed by a detergent corona) to aid particle alignment. In particular, it has an impressive record for improving the resolution of small membrane protein reconstructions (e.g. Kim et al. (2019)). Passing the *ab initio* output through only a single round of NUR recovered a 3.8 Å map, boasting clear helical pitch and the beginnings of explicable sidechain density (Figure 7.1[C]). After additional NUR jobs, interspersed with bayesian polishing (per-particle motion correction) in RELION 3.1.0 (Zivanov et al., 2019, 2020) and various operations to further clean the dataset (Figure 7.1[C]), ~414K particles (5.3% of the starting stack) remained. Final beam-tilt correction and local refinement of these gave a 3.0 Å reconstruction of $MmpL3_{1\text{-}753}$ (Figure 7.1[D]), deposited in the EMDB under accession code EMD-12604; local resolution being estimated to vary between ~2.8–3.6 Å across the proteinaceous density (Figure 7.1[E]).

$MmpL3_{1\text{-}753}$ Monomerism

Congruent with earlier construct characterisation (Section 3.4.2) and in-ice screening (Section 4.2.2), the Cryo-EM map reaffirmed the monomeric nature of $MmpL3_{1\text{-}753}$ (Figure 7.2[A]); agreeing with the structures, and accompanying full-length protein biochemical analyses (BN-PAGE, SAXS & native mass spectrom-

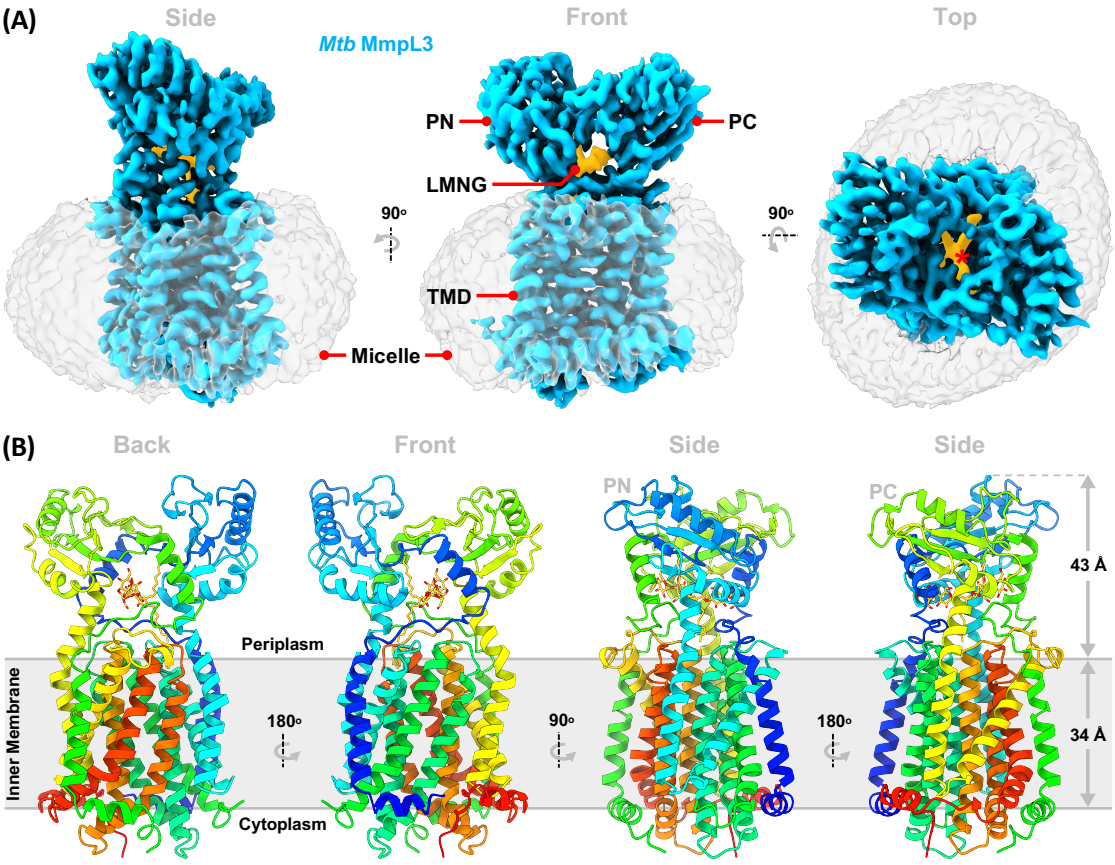


Figure 7.2: Cryo-EM Map & Atomic Model of MmpL3$_{1\text{-}753}$ in LMNG

(A) 3.0 Å Cryo-EM volume of LMNG solubilised MmpL3$_{1\text{-}753}$ as viewed from the membrane plane (side, front), or perpendicular to it from the periplasm (top). Proteinaceous density is coloured sky blue, the surrounding detergent corona transparent grey, and a fortuitously resolved PD LMNG ligand gold. Annotations highlight the transmembrane domain (TMD), as well as the N-terminal (PN) and C-terminal (PC) lobes of the PD. The red asterisk approximates the two-fold pseudo-symmetry axis running down the centre of the transporter. For clarity elements have been differentially contoured, and are taken from both the sharpened (protein) and unsharpened (LMNG, micelle) maps, with dust hidden. **(B)** Cartoon representation of the MmpL3$_{1\text{-}753}$ atomic model (comprising residues 1–342, 378–752), depicted in chainbow colouring from blue (N-terminus) to red (C-terminus), and presented from four orthogonal directions. The position of the membrane bilayer, estimated from the Cryo-EM volume detergent micelle, is indicated by the grey bar. The PD-bound LMNG is shown as sticks, coloured gold. Map and model statistics are given in Table 7.1. Structures and volumes prepared with UCSF ChimeraX v.1.1 (Pettersen et al., 2021).

etry), presented for the *Msmg* homologue (Su et al., 2019; Zhang et al., 2019). Divergence from the prototypical homotrimeric assembly seen for Gram-negative RND efflux-pumps, first established by stalwart AcrB (Murakami et al., 2002), is exceedingly rare for prokaryotic superfamily members of known structure. One of the only other examples is *Burkholderia sp.* HpnN, a hopanoid transporter solved in both monomeric (Su et al., 2021) and dimeric (Kumar et al., 2017)

states. In fact, the monomeric nature of MmpL3 is most reminiscent of RND proteins belonging to the eukaryotic sterol transporter (EST) family. For example, structures of the mammalian EST cholesterol translocators NPC1 (Gong et al., 2016), NPC1L1 (Huang et al., 2020), and in some instances PTCH1 (Gong et al., 2018; Qi et al., 2019), all identify single-chain architectures. Given MmpL3, HpnN, and ESTs are all endogenous lipophile transporters, it is tempting to speculate monomerism is somehow a prerequisite to co-opt the RND fold for such tasks.

Of course, it is possible to countenance that detergent extraction and CTD removal may be confounding factors, and therefore one cannot entirely discount homo-oligomerisation of full-length *Mtb* MmpL3 when embedded in its native membrane, with access to any necessary ancillary proteins. Of pertinence, scant evidence has linked the CTD to a role in multimerisation, or at least self-interaction (Belardinelli et al., 2019). MmpL5 homotrimerisation, dependent on MmpS5 coexpression, was also recently inferred *in vivo* by fluorescence photobleaching experiments (Yamamoto et al., 2021). However, as it is a Cluster-I MmpL (i.e. unequal periplasmic loop sizes and lacking a CTD) the relevance of MmpL5 oligomerisation with respect to MmpL3 is at best ambiguous.

7.2.2 Overall Architecture of MmpL3$_{1\text{-}753}$

At 3.0 Å the electron microscopy volume provided ample, sufficiently interpretable, sidechain density (Supplementary Figure B.17), allowing confident modelling of almost the entirety of MmpL3$_{1\text{-}753}$ (residues 1–342, 378–752; Figure 7.2[B]). This was achieved by first threading the *Mtb* MmpL3 sequence onto the most appropriate *Msmg* model (i.e. PDB: 6OR2, the *apo* state structure without C-terminal lysozyme fusion, Su et al. (2019)) using the Phyre2 web server (Kelley et al., 2015). The threaded structure was next docked into the sharpened Cryo-EM map and underwent iterative rounds of real-space refinement in both Coot (Emsley et al., 2010) and PHENIX (Liebschner et al., 2019) to give the

Mtb MmpL3$_{1\text{-}753}$ (EMDB: EMD-12604, PDB: 7NVH)	
Data Collection & Processing	
Magnification	105,000
Voltage (kV)	300
Electron Exposure (e^-/Å^2)	58.2
Defocus Range (μm)	-3 to -0.5
Pixel Size (Å)	0.832
Symmetry Imposed	C1
Initial Particle Images (no.)	7,764,838
Final Particle Images (no.)	414,082
Map Resolution (Å)	3.0
FSC Threshold	0.143
Map Resolution Range (Å)	2.8 – 4.0
Refinement	
Initial Model Used (PDB Code)	6OR2
Model Resolution (Å)	3.0
FSC Threshold	0.143
Model Resolution Range (Å)	2.8 – 4.0
Map Sharpening *B* factor (Å^2)	-86.7
Model Composition	
Non-hydrogen Atoms	5529
Protein Residues	717
Ligands	1
B Factors (Å^2)	
Protein	53.19
Ligand	63.15
Root-mean-square Deviations	
Bond Lengths(Å)	0.004
Bond Angles (°)	0.641
Validation	
MolProbity Score	1.78
Clashscore	7.31
Poor Rotamers (%)	0.69
Ramachandran Plot	
Favoured (%)	94.39
Allowed (%)	5.61
Disallowed (%)	0

Table 7.1: Cryo-EM Data Collection, Refinement & Model Validation Statistics

final model (Figure 7.2[B]), deposited in the PDB under accession code 7NVH. For further details see Chapter 2. Cryo-EM data collection and model validation statistics are summarised in Table 7.1.

Akin to the *Msmg* homologue (1.49 Å RMSD over all atoms; Figure 7.3[A]), *Mtb* MmpL3$_{1\text{-}753}$ adopts a transmembrane domain (TMD) fold archetypal of RND proteins. That is, 12 TMs arranged as two sequence-contiguous bundles (TMs 1–6 & 7–8). These are related to one another by a central two-fold pseudo-symmetry axis, running perpendicular to the membrane plane (Figures 7.2[A] & 7.3[B]). The TMD is further braced against the inner leaflet of the cytoplasmic membrane

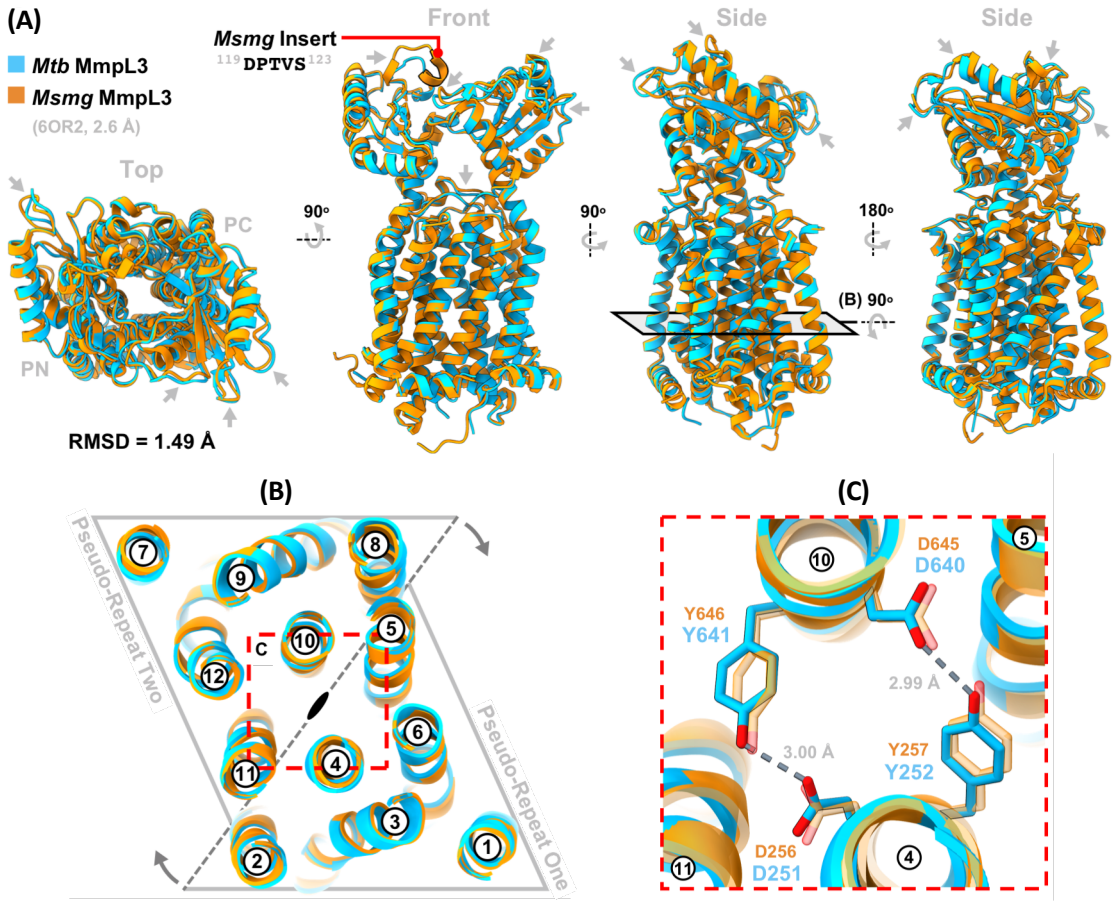


Figure 7.3: Structural Comparison of *Mtb* & *Msmg* MmpL3

(A) Alignment of the *Mtb* MmpL3$_{1\text{-}753}$ (sky blue) and *Msmg* MmpL3$_{1\text{-}773}$ (orange, PDB: 6OR2, Su et al. (2019)) models (1.49 Å RMSD over all atoms), shown in cartoon representation, and viewed from the membrane plane (front, side) or from the periplasm (top). Grey arrows emphasise deviations in loop architecture between the orthologues. A black outline, and accompanying annotation, identify the PN pentapeptide (DPTVS) present in *Msmg* but not *Mtb* MmpL3. **(B)** Clipped view of the overlayed TMDs, sliced as indicated in (A). The black oval designates the central two-fold pseudo-symmetry axis, relating the two six TM repeats (enclosed in grey triangles). Helices are numbered. **(C)** Detailed close-up of region boxed in (B). Sticks depict sidechains of the pair of Asp-Tyr dyads connecting TM4 and TM10 through hydrogen bonds (dashed lines). Structures aligned and rendered in UCSF ChimeraX v.1.1 (Pettersen et al., 2021).

by three short, lateral, amphipathic α-helices. Two precede TM1 and TM7, and the third follows TM12 and thus leads into the CTD. As expected, each six TM-bundle hosts a sizeable periplasmic loop, adjoining TMs 1–2 in the N-terminal half of MmpL3$_{1\text{-}753}$ (PN, residues 37–166) and TMs 7–8 (PC, residues 415–544) in the C-terminal half of the molecule (Figure 7.2[B]). Both display an α-β-α-β-α-β topology, with the first α-helix of each contributing to the tertiary structure of the other, adjacent, loop. This partial interdigitation of secondary elements results

in PN and PC clasping one another, co-folding to give a singular PD capping the TMD and sharing its pseudo-symmetry axis. Being significantly longer than the rest of the TMs, the N-termini of TM2 and TM8 are seen to penetrate up into, and thus form a composite part of, the PD; itself protruding over 40 Å into the periplasmic space (Figure 7.2[B]). As the distance between the centres of the inner and outer mycobacterial membranes is in excess of 300 Å (Zuber et al., 2008), the structure unequivocally supports theories MmpL3 functions in concert with accessory proteins to deliver TMM to the mycomembrane (Belardinelli et al., 2019; Melly et al., 2019). The interdependence of PN and PC is also consistent with aforementioned difficulties in advancing single subdomain deletions for sybody epitope binning (see Section 5.4.3).

The internal structural repeat in $MmpL3_{1\text{-}753}$ is consonant with the gene duplication events accepted to have driven emergence of modern-day RND transporters; an evolutionary origin first proposed in the seminal phylogenetic analyses defining the superfamily (Saier et al., 1994; Tseng et al., 1999). The sequence bridging the related N- and C-terminal halves of the truncation (residues 343–377), connecting TM6 to the lateral helix preceding TM7 (labelled in Supplementary Figure B.18[A]), is presumably unstructured and/or highly mobile given the lack of density for it in the Cryo-EM map. Disorder of this inter-repeat linker seems common for RND proteins as structures from disparate superfamily members frequently leave the region unmodelled — these include AcrB (Murakami et al., 2002), CusA (Long et al., 2010), HpnN (Kumar et al., 2017), PTCH1 (Zhang et al., 2018), and Dispatched (Cannac et al., 2020). If minimally structured the linker may be “sticky”, in which case it is a plausible candidate for the immunodominant epitope hypothesised to be the target of many of the sybodies selected in DDM in Chapter 5. Equally, considering its location, the inter-repeat linker could be the site bound by MG2, which would explain why the LMNG-competent sybody appeared as a “micelle binder” in 2D-CAVGs (see Section 5.5.3).

Finally, whilst the *Mtb* MmpL3$_{1\text{-}753}$ fold is commensurate with that of the *Msmg* homologue, multiple subtle deviations in PD architecture are observable (grey arrows in Figure 7.3[A]). Most apparent is the altered conformation of the uppermost PN coil (residues 112–118), a consequence of *Msmg* MmpL3, like many other mycobacterial orthologues (Sequence Alignment Figure A.7), featuring a five residue insert at the position relative to *Mtb* (black outline in Figure 7.3[A]); albeit of unknown significance. Interestingly, in the wider RND superfamily the equivalent regions in both extramembranous loops are recognised evolutionary hotspots, frequently hosting insertion of whole domains linked to member functional specialisation (Zhang et al., 2018).

Having now solved the structure of *Mtb* MmpL3$_{1\text{-}753}$ attention next turned to what the model could disclose, in conjunction with the *Msmg* studies, about the mechanism of TMM export and the mode action of, and routes to resistance to, preclinical antituberculars targeting the protein. Pursuit of these functional and therapeutic questions is the basis of both Sections 7.3 & 7.4, and leads to discussion of an extensive cavity network in the PD and a potentially "universal" drug pocket in the TMD. Notably, the latter incorporates the Asp-Tyr interaction network between TM4 and TM10 (Figure 7.3[C]). Almost universally conserved in MmpLs, this tetrad is thought to be essential to proton motive force (PMF) coupling within the family (Bernut et al., 2016; Zhang et al., 2019).

7.3 Regions of Functional & Therapeutic Interest

7.3.1 The Periplasmic Cavity: A Versatile Lipid Binding Space

Similar to the *Msmg* structures (Zhang et al., 2019; Su et al., 2019), MmpL3$_{1\text{-}753}$ possesses a cavity formed of a central vestibule, between the PN and PC lobes, out of which branch two descending channels; each constricted in the vicinity of the PD-TMD "neck" (Figure 7.4[A]). On opposing faces of the protein these channels exit into the outer leaflet of the inner membrane, via grooves defined

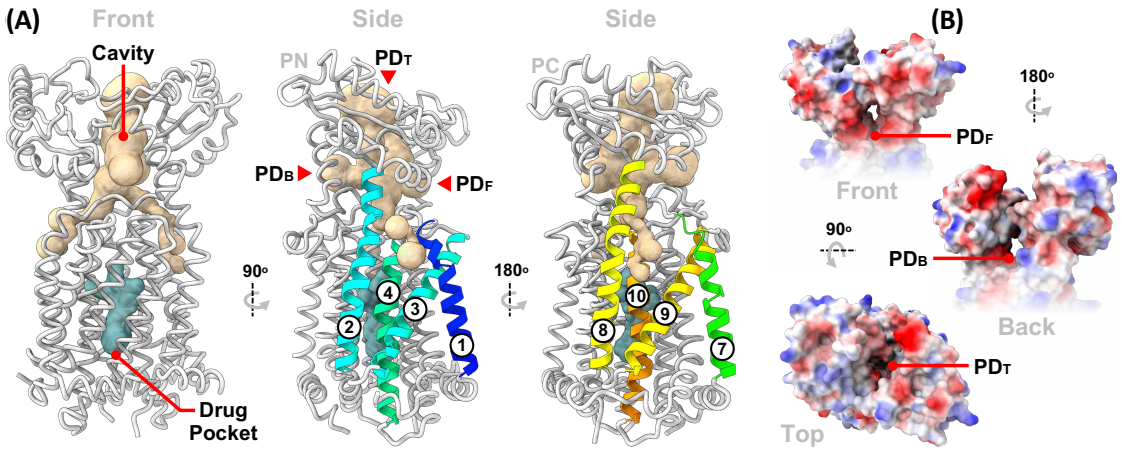


Figure 7.4: MmpL3$_{1\text{-}753}$ Regions of Functional & Therapeutic Importance

(A) Licorice representation of MmpL3$_{1\text{-}753}$ (grey) with surface depictions emphasizing the PD cavity network (wheat) and TMD drug binding pocket (teal). The latter is a composite derived from structures of the *Msmg* orthologue (PDBs: 6AJG to 6AJJ, 7C2N & 7C2M, Zhang et al. (2019); Yang et al. (2020)), resolving the compounds boxed in Figures 7.7[A] & 7.8[A]. Calculated using Caver Analyst 2.0 (Jurcik et al., 2018), the PD cavity is formed of a central vestibule, enclosed by PN and PC, and opens into the periplasm via apertures at the front (PD$_F$), back (PD$_B$), and top (PD$_T$) of the molecule. Two descending channels, putative inner membrane TMM inlets, fork from the main chamber and exit into the outer leaflet via grooves on either side of MmpL3$_{1\text{-}753}$. The TMs bounding these portals (1–4 & 7–10) are accentuated in cartoon form in the two side views, coloured as in Figure 7.2[B]. **(B)** Electrostatic surface renderings of the truncation PD, highlighting the hydrophilic residues gating the labelled PD$_F$, PD$_B$, and PD$_T$ openings. Structures visualised with UCSF ChimeraX v.1.1 (Pettersen et al., 2021).

by TMs 1–4 or TMs 7–10. The upper PD chamber likewise possesses three apertures, individually opening into the periplasm from the front (PD$_F$), back (PD$_B$), and top (PD$_T$) of MmpL3$_{1\text{-}753}$ and each gated by a combination of charged and hydroxyl-containing sidechains (Figure 7.4[B]). Given its position in the molecule, this cavity network is highly likely to represent the conduit through which *Mtb* MmpL3 abstracts TMM from the outer leaflet and then translocates it towards the periplasm. A process presumably driven by conformational changes energised by TMD-transduction of the PMF. Once within the PD, TMM would next need to be handed-off, conceivably through PD$_T$, to a as yet undefined acceptor, potentially lipoprotein LpqN (Melly et al., 2019), for onward transit to the mycomembrane.

Consistent with a crucial functional role, cysteine site-directed mutagenesis (SDM) of various *Mtb* MmpL3 cavity-lining residues has been shown to compromise

mycobacterial viability (G543) or at least impair MmpL3 activity (Q304, D550 & P625, labelled later in Figure 7.6[A]; Belardinelli et al. (2016)). Tellingly, screening the same cysteine SDM library for intolerance to thiol-reactive agents identified residues defining, or vicinal to, the constrictions of both the channels descending from the primary vestibule (Q40, Y235 & P625, see Figure 7.6[A]). That chemical occlusion of these induced growth inhibition, again attendant with altered MmpL3 function (Belardinelli et al., 2016), strongly suggests they represent the entry routes used to subduct outer leaflet TMM into the protein. Such cellular phenotypes are undoubtedly linked to faltering cell wall biogenesis, downstream of suboptimal TMM export by the described MmpL3 mutants.

Periplasmic Ligands

The Cryo-EM volume fortuitously captured $MmpL3_{1\text{-}753}$ in complex with a non-proteinaceous, lipid-like, species splayed between PN and PC, and thus inhabiting the central periplasmic chamber (gold density in Figure 7.2[A]). Although incongruous in its separation from the bulk detergent corona, the density could readily accommodate a single LMNG molecule, which was subsequently built in (Figure 7.2[A], Supplementary Figure B.17). As LMNG is in essence a synthetic lipid, having been rationally designed to contain two alkyl chains (Chae et al., 2010; Breyton et al., 2019), and is found along the suggested TMM transport route (Klenotic et al., 2021), analysis of its interactions with the protein are likely to be mechanistically informative regarding native substrate engagement. Nestled within the PD interior, the ligand is able to realise a substantial interface; packing against a total of 35 residues (within 4.5 Å). These are predominantly contributed by PN, PC and TM2, but also feature sidechains at the N-terminus of TM8 and in the loops connecting TMs 3–4, and TMs 11–12 (Figure 7.5[A]). Whilst the LMNG tails point in divergent directions, one halfway down into the channel exiting via TMs 1–4 and the other up towards PD_T, both are accommodated in pockets partially scaffolded by TM2 aliphatics (L166, V169, L173 & I177) and end in aromatic residues (Y235 & F236 or F440 & Y447) provided by nearby

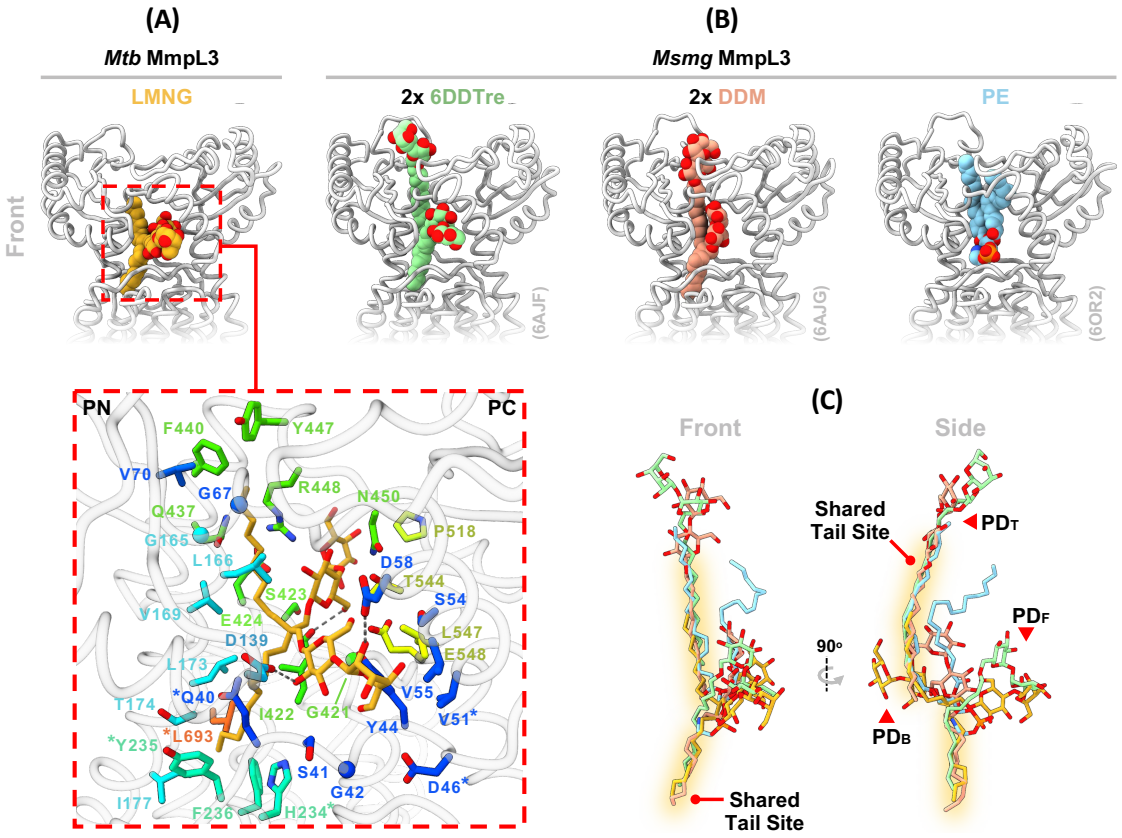


Figure 7.5: Structural Analysis of *Mtb* & *Msmg* MmpL3 Periplasmic Ligands

(A) Detailed close-up of the $MmpL3_{1\text{-}753}$ PD LMNG-binding site. Sidechains of residues within 4.5 Å of LMNG, as well as the detergent itself (gold), are shown as sticks; except for four glycines denoted by balls. Residues are coloured by primary sequence order, as in Figure 7.2[B]. Dashed lines denote truncation-LMNG hydrogen bonds. Asterisks highlight positions conferring inhibitor resistance when mutated. **(B)** Comparison of the *Mtb* LMNG complex with the three lipophilic species co-crystallised within the *Msmg* MmpL3 PD: lauryl-6-trehaloside (6DDTre, pale green, PDB: 6AJF, Zhang et al. (2019)), dodecyl maltoside (DDM, salmon, PDB: 6AJG, Zhang et al. (2019)), and phosphatidylethanolamine (PE, light blue, PDB: 6OR2, Su et al. (2019)). The ligands are shown as spheres, and proteins (grey) in licorice representation. **(C)** Overlays juxtaposing the poses of the four PD ligands, depicted as sticks with retained colouring. Common alkyl chain sites, spotlighted by soft yellow borders, are annotated. The approximate positions of the PD apertures are indicated. Composed in UCSF ChimeraX v.1.1 (Pettersen et al., 2021).

secondary structure elements. In contrast, the two LMNG maltose headgroups bifurcate to be individually coordinated by residues of the PD_F and PD_B apertures. In doing so their hydroxyls are hydrogen-bonded by the sidechains of D58 and D139, as well as the backbone carbonyl of I422, leaving the terminal saccharide of each headgroup to be partly solvent exposed as it emerges through its chosen PD opening (Figure 7.5[A]).

Surveying the *Msmg* MmpL3 structures shows the homologue has been serendipitously co-crystallised with three unique PD occupants (Figure 7.5[B]), namely pairs of the purification detergents 6DDTre and DDM (Zhang et al., 2019), as well as a single phosphatidylethanolamine (PE) lipid (Su et al., 2019). Not only does this confirm the PD cavity to be a versatile amphiphile interaction space, but structural overlay with $MmpL3_{1\text{-}753}$ identifies striking commonalities in the binding poses achieved by all four lipid, or lipid-like, ligands (Figure 7.5[C]). The pockets embracing the LMNG alkyl chains are comparably exploited by 6DDTre and DDM to accommodate their respective hydrophobic tails. Likewise PE, whilst displaying a binding mode overall most dissimilar to the others, places its shorter heptadecanoate acyl chain into the upper of these two tail sites; leaving its peripheral carbons to extend through PD_T. To a lesser extent, disaccharide headgroup recognition is also conserved among the MmpL3-detergent complexes. The trehalose headgroups of the two 6DDTre molecules, mimicking TMM, are separately processed by PD_F and PD_T in the *Msmg* structures. Residues lining the former aperture hydrogen bond the trehalose unit, analogous to the contacts between $MmpL3_{1\text{-}753}$ and LMNG, whereas the latter opening leaves the second headgroup largely solvent exposed, akin to the association apparent between one DDM headgroup and PD_T.

Taken together, the interactions made between the MmpL3 orthologues and LMNG, DDM, 6DDTre, and PE, are suggestive of a "division of labour" model for PD lipid handling — wherein the central vestibule sequesters client alkyl chains away from the bulk periplasm, potentially into specific pockets (Figure 7.5[C]), whilst the proximate hydrophilic openings (PD_F, PD_B & PD_T) dock the polar headgroup, coordinating it via hydrogen bonds and providing it a degree of solvent access. Modularisation of binding would permit accommodation of lipidic cargo in, and subsequent manipulation into, diverse conformations. This flexibility is plausibly essential for translocation of substrates as unwieldly as TMM (MW >1.4 kDa, Fujita et al. (2005)), and likely underlies the lipid binding promiscuity observed for *Msmg* MmpL3 by native mass spectrometry (Su et al., 2019).

Re-evaluating the behaviour of $MmpL3_{1\text{-}753}$ in DDM and LMNG (see Chapter 3) in light of the lipophile binding poses is also informative. Assuming the *Mtb* transporter complexes DDM comparably to *Msmg* MmpL3 (Zhang et al., 2019), one can infer the greater stabilisation imparted by LMNG is, likely at least partially, due to its ability brace the PD. As the two headgroups and two alkyl chains of the MNG simultaneously straddle four lipid binding modules (PD_F, PD_B & both tail sites), the detergent conceivably acts as a strut, rigidifying the extramembranous domain. In comparison, in only occupying three binding modules (PD_T & both tail sites), split over two uncoupled molecules, DDM would appear less effective at bolstering PD structural integrity.

Cavities in Other RND Lipophile Transporters

Beyond corroboratory mutagenesis phenotypes (Belardinelli et al., 2016) and structural analysis of the PD ligands, support for the functionality of the $MmpL3_{1\text{-}753}$ cavity can be amassed by comparison with other RND superfamily members — in particular the aforenoted hopanoid transporter, HpnN, and well characterised EST cholesterol traffickers NPC1, NPC1L1, and PTCH1. Aptly, the HAE2 (incl. the MmpLs) and HAE3 (incl. HpnN) families are among the prokaryotic RND transporters clustering phylogenetically closest to the EST proteins (see Figure 1.2 ; Yen et al. (2010)). Not only does MmpL3 share putative monomerism with HpnN and ESTs, they are similarly thought to execute ion-coupled translocation of their lipophilic substrates via tunnels (Kumar et al., 2017; Zhang et al., 2018; Qian et al., 2020; Huang et al., 2020). As for the TMM exporter, these are seen to traverse the length of their extramembrane domains, emptying into the membrane leaflet via grooves within the RND-conserved TMD fold (Figures 7.6[A]&[B]). Equivalent to MmpL3, the relevance of these conduits has been validated by both mutagenesis (Kumar et al. (2017); Zhang et al. (2018); Qi et al. (2018); Long et al. (2020); Figure 7.6[B]) and, in the case of the ESTs, evaluation of the multiple ligands resolved along their routes (Figure 7.6[C]).

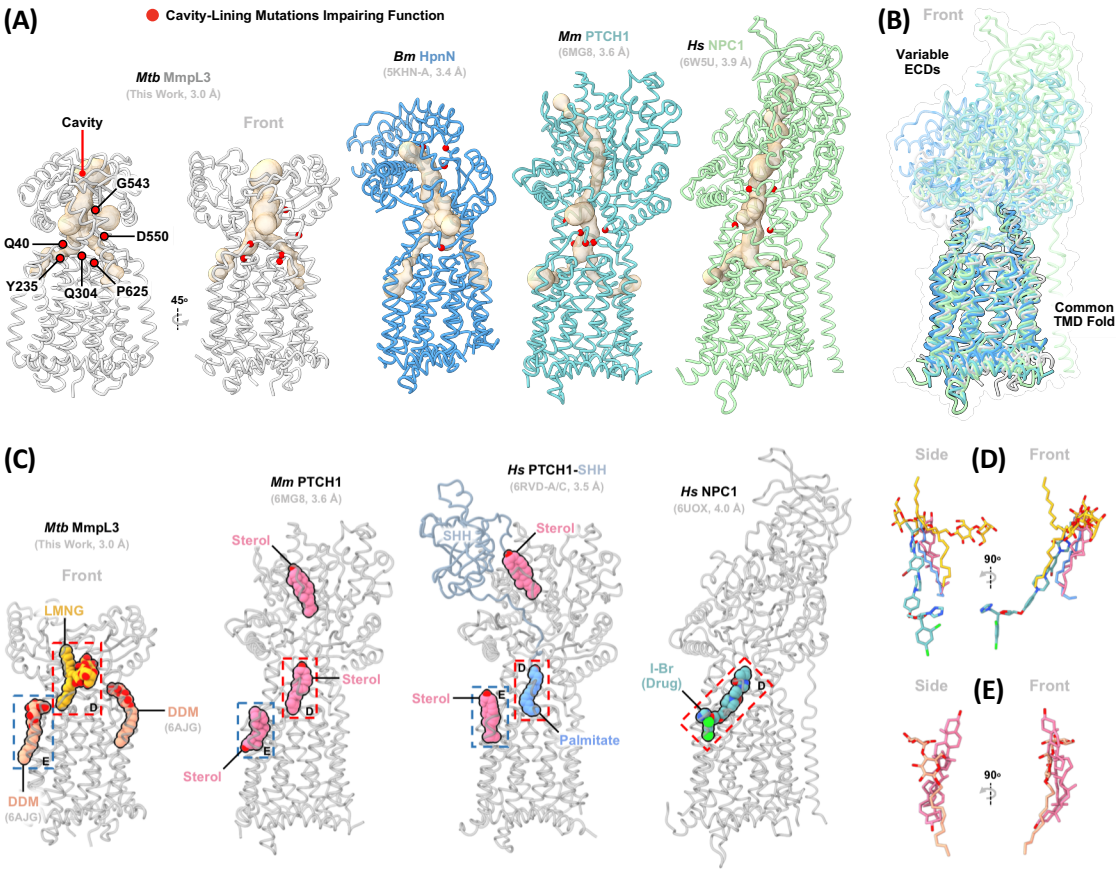


Figure 7.6: Comparison of MmpL3$_{1\text{-}753}$ to Other RND Lipophile Transporters

(A) Extramembranous cavities of *Mtb* MmpL3$_{1\text{-}753}$ (grey), *Burkholderia multivorans* (*Bm*) hopanoid transporter HpnN (blue, PDB: 5KHN, Chain A, Kumar et al. (2017)), *Mus musculus* (*Mm*) cholesterol translocator PTCH1 (teal, PDB: 6MG8, Zhang et al. (2018)), and *Homo sapiens* (*Hs*) lysosomal cholesterol exporter NPC1 (pale green, PDB: 6W5U, Qian et al. (2020)). Proteins rendered in licorice representation, and cavities (transparent wheat surfaces) calculated with Caver Analyst 2.0 (Jurcik et al., 2018). Red balls highlight examples of cavity-lining residues reported to impair protein function when mutated. All-atom alignment of the same four proteins is given in **(B)**, accentuating their shared RND superfamily TMD fold. **(C)** Comparison of ligands resolved within the cavities and TM grooves of MmpL3, *Mm* PTCH1, the *Hs* sonic hedgehog (SHH) PTCH1 complex (PDB: 6RVD, Chains A/C, Rudolf et al. (2019)), and *Hs* NPC1 (PDB: 6UOX, Long et al. (2020)). DDM molecules presented with *Mtb* MmpL3$_{1\text{-}753}$ are superimposed from a structure of the *Msmg* orthologue (PDB: 6AJG, Zhang et al. (2019)). All ligands — LMNG (gold), DDM (salmon), Sterol (pink), Palmitate (blue), and a brominated derivative of the antifungal Itraconazole (I-Br, teal) — are shown as spheres. Species occupying the central vestibule (red box) and TMs 1–4 groove (blue box) are overlain in **(D)** & **(E)** respectively; depicted as sticks with retained colouring. Structures prepared in UCSF ChimeraX v.1.1 (Pettersen et al., 2021).

Compellingly, overlay with structures of PTCH1 and NPC1 shows the MmpL3$_{1\text{-}753}$ LMNG occupies a location analogous to the binding site accommodating substrative sterol (Zhang et al., 2018), the palmitate terminus of the lipidated regulator sonic hedgehog (Rudolf et al., 2019), and/or antifungal compounds (Long et al.,

2020) in these transporters (Figure 7.6[D]). Likewise, sterols frequently resolved within the groove formed by EST TMs 1–4, advanced as the portal for sterol passage between protein and membrane, colocalise with one of the DDM molecules co-crystallised with *Msmg* MmpL3 (Zhang et al. (2019); Figure 7.6[E]). Given the physiological significance, and therapeutic value of small molecule occluders (Huang et al., 2020; Long et al., 2020), of EST tunnels, one can posit the role of the MmpL3 cavity in TMM export comparably renders it an attractive, but as yet unexplored, target for novel antitubercular development. One that would complement the multitude of compounds, explored in the following sections, inhibiting MmpL3 by binding its TMD pocket.

7.3.2 The Universal Drug Pocket

Unexpected given their chemical disparity, the six preclinical antituberculars resolved in complex with *Msmg* MmpL3 (Zhang et al., 2019; Yang et al., 2020), spanning five scaffold series, all occupy the same TMD site at the interface of TMs 4–5 & 10–11; intersecting the central two-fold pseudo-symmetry axis of the protein (Figures 7.4[A] & 7.7[A]). The structures thus identify a potentially universal drug pocket, expanded upon and corroborated by *in silico* docking (Zhang et al., 2019; Dupont et al., 2019; Li et al., 2020b), engaged by most, if not all, anti-MmpL3 agents currently in development. It is defined by twenty-five residues (Figure 7.7[B]), each directly interacting with one or more of the co-crystallised inhibitors, inclusive of the previously mentioned Asp-Tyr tetrad (D251, Y252, D640 & Y641). Connecting TM4 to TM10 by a pair of hydrogen bonds, these four residues are purported to have a critical role in MmpL3 PMF transduction by analogy to equivalently located ionisable clusters in well characterised RND systems (e.g. the AcrB Asp-Asp-Lys triad, Eicher et al. (2014)), as well as on the basis of mutational analyses (Bernut et al., 2016; Belardinelli et al., 2016).

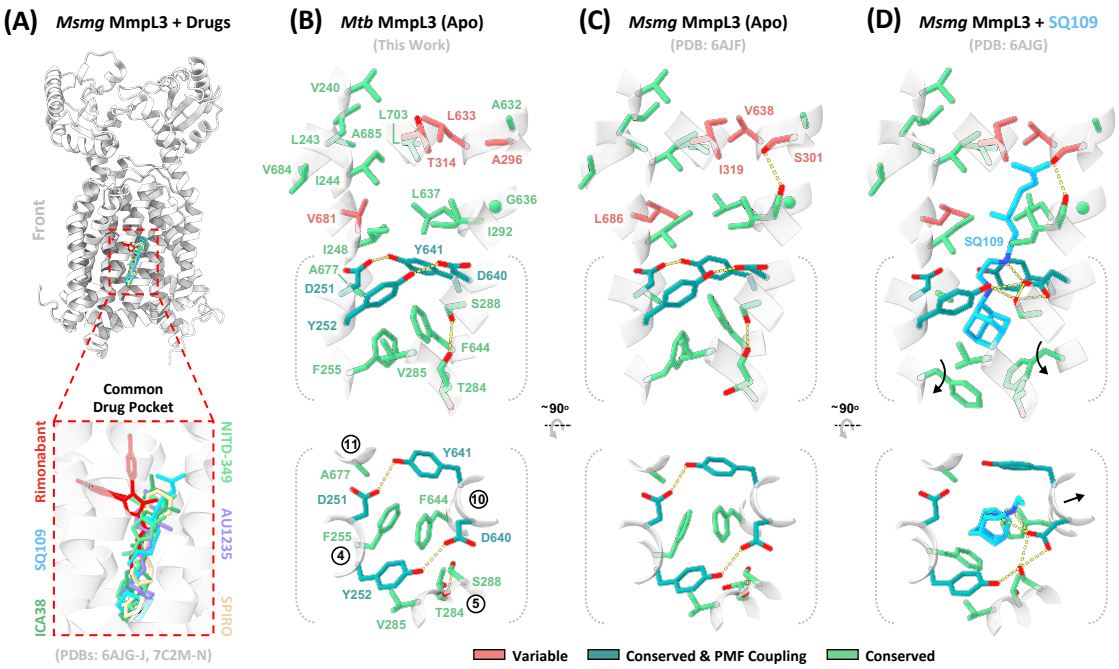


Figure 7.7: *Mtb* MmpL3$_{1\text{-}753}$ Conservation of the *Msmg* Common Drug Pocket

(A) Cartoon representation of *Msmg* MmpL3 (grey) superimposed with all six co-crystallised preclinical inhibitors (PDBs: 6AJG to 6AJJ, 7C2N & 7C2M, Zhang et al. (2019); Yang et al. (2020)). These occupy a shared TMD pocket (red box) and are shown as sticks colour-coded by drug class (as in Figure 7.9). For clarity the C-terminally fused T4 lysozyme is hidden. **(B)** Detailed close-up of the common drug pocket as resolved in *apo Mtb* MmpL3$_{1\text{-}753}$. Sidechains of the 25 residues interacting with at least one of the inhibitors in the *Msmg* structures are portrayed as sticks (balls for Gly), coloured by conservation between the two orthologues. The base of the pocket, enclosed in grey brackets, is additionally shown from an orthogonal viewpoint with TMs numbered. Dashed lines denote hydrogen bonds. Equivalent depictions of *Msmg* MmpL3 in *apo* and SQ109-bound states (PDBs: 6AJF & 6AJG, Zhang et al. (2019)) are given in **(C)** & **(D)** respectively. Here only variable residues are labelled, using *Msmg* numbering. Black arrows highlight inhibitor-induced structural rearrangements. Note the disruption the Asp-Tyr hydrogen bond network, thought to be involved with proton motive force (PMF) coupling, on drug complexation. Composed with UCSF ChimeraX v.1.1 (Pettersen et al., 2021).

All but four of the drug-contacting residues (A296, T314, L633 & V681) are identical between the *Msmg* and *Mtb* orthologues (red in Figures 7.7[B]&[C]); unsurprising considering the pocket resides within the most conserved region of MmpL3 (Supplementary Figure B.18). Not only does this imply the *Msmg* findings are likely generalisable to the *bona fide* antitubercular target, it lends credence to the idea MmpL3 inhibitors may be pan-mycobactericidal (Bolla, 2020). Expounded by Zhang et al. (2019) & Yang et al. (2020), structural comparison of the pocket in *apo* (Figures 7.7[B]&[C]) and inhibitor bound (e.g. to

SQ109, Figure 7.7[D]) states delineates a common mechanism of action for these compounds, executed by a shared pharmacophore — two bulky hydrophobes bridged by a more polar nitrogenous moiety. The large hydrophobic groups (e.g. adamantyls, geranyls, halogenated phenyls etc.) are seen to wedge into apolar subsites above and below the Asp-Tyr plane. In doing so, they prime their intervening nitrogen atom(s) (e.g. part of an indole, amide, piperidine etc.) to interact with D640, interrupting its hydrogen bond to Y252. Combined with the outward breathing of TMs 6–12 in steric accommodation of the drugs, this triggers reconfiguration of the entire Asp-Tyr hydrogen bond network (Figure 7.7[D]), in theory uncoupling MmpL3 from the PMF and so fatally de-energising TMM extrusion.

Unnoted in the *Msmg* publications, helical displacement following drug binding also results in severance of a second set of interactions, salt-bridges, connecting TMs 4 & 10. These are distal to the pocket at the C-termini of the helices, on the cytoplasmic face of the protein, but again are formed between a tetrad of conserved residues (R259, E262, R648 & E651; Supplementary Figure B.19) arranged around the two-fold pseudo-symmetry axis. Intriguingly, cysteine SDM of R259 has been shown to impair MmpL3 activity, suggesting this Arg-Glu network is of functional importance (**Belardinelli et al., 2016**). Alongside additional breakage of polar contacts joining TMs 5 & 8 (Supplementary Figure B.19), such observations prompt conjecture that the efficacy of MmpL3-targeting agents stems from an ability to disengage the two TM repeats (i.e. TMs 1–6 & 7–12) from one another. How this decoupling may curtail function, and its relevance to compromise of PMF transduction, awaits proper description of the MmpL3 transport cycle.

To recap, from a therapeutic perspective, the structure of $MmpL3_{1\text{-}753}$ thus wholly validates the *Msmg* studies, confirming the *Mtb* protein to possess a comparable TMD drug pocket. Keeping the proposed *modus operandi* of MmpL3 inhibitors in mind, the next section explores their mutational evasion.

7.4 The Mutational Landscape of MmpL3

7.4.1 The Missense Landscape of MmpL3 Drug-Resistance

Since its initial description as the protein inhibited by BM212 (La Rosa et al., 2012) MmpL3 has been identified as the target for numerous preclinical antitubercular derived from a myriad of different scaffolds (Degiacomi et al., 2020). Indeed the list is seemingly ever-growing, and includes Pyrroles and Pyrazoles (La Rosa et al., 2012; Poce et al., 2019), Ethylenediamines (Tahlan et al., 2012; Foss et al., 2016), Indolecarboxamides (Lun et al., 2013; Rao et al., 2013), Piperidinol Derivatives (Dupont et al., 2016, 2019), Benzimidazoles (Stanley et al., 2012; Korycka-Machała et al., 2019), Urea Derivatives (Grzegorzewicz et al., 2012), Spirocycles (Remuiñán et al., 2013; Tantry et al., 2015), and Benzothioazole Amides (De Groote et al., 2018), among others. Predominantly such agents have been discovered by high-throughput whole-cell phenotypic screening of mycobacteria against sizeable chemical libraries, with targets for active compounds inferred by their mutation in whole-genome sequenced resistant isolates (Goldman, 2013). Consequently there exists a substantial repository of drug-resistant MmpL3 missense variants in the literature. These were exhaustively collated, and then plotted onto the MmpL3$_{1\text{-}753}$ model to provide valuable insights into the mechanisms of resistance.

In toto, 112 unique resistance-conferring mutations were aggregated, occurring at 83 different positions in MmpL3 (Supplementary Table B.2); 21 of which differ between the *Mtb* and *Msmg* homologues. Of the 112 mutations, 109 variants at 80 positions were mappable onto the MmpL3$_{1\text{-}753}$ structure (Figure 7.8[A]). The remaining substitutions populated the unresolved inter-repeat linker (residues 343–377). Most mutations were described in *Mtb* (57 positions), but to provide a more complete picture of the resistance landscape variants originally identified in *Msmg* (11 positions), *M. bovis* (13 positions), and *M. abscessus* (19 positions) were also back-mapped, justifiable given the extent of MmpL3 orthologue conservation (Sequence Alignment Figure A.7).

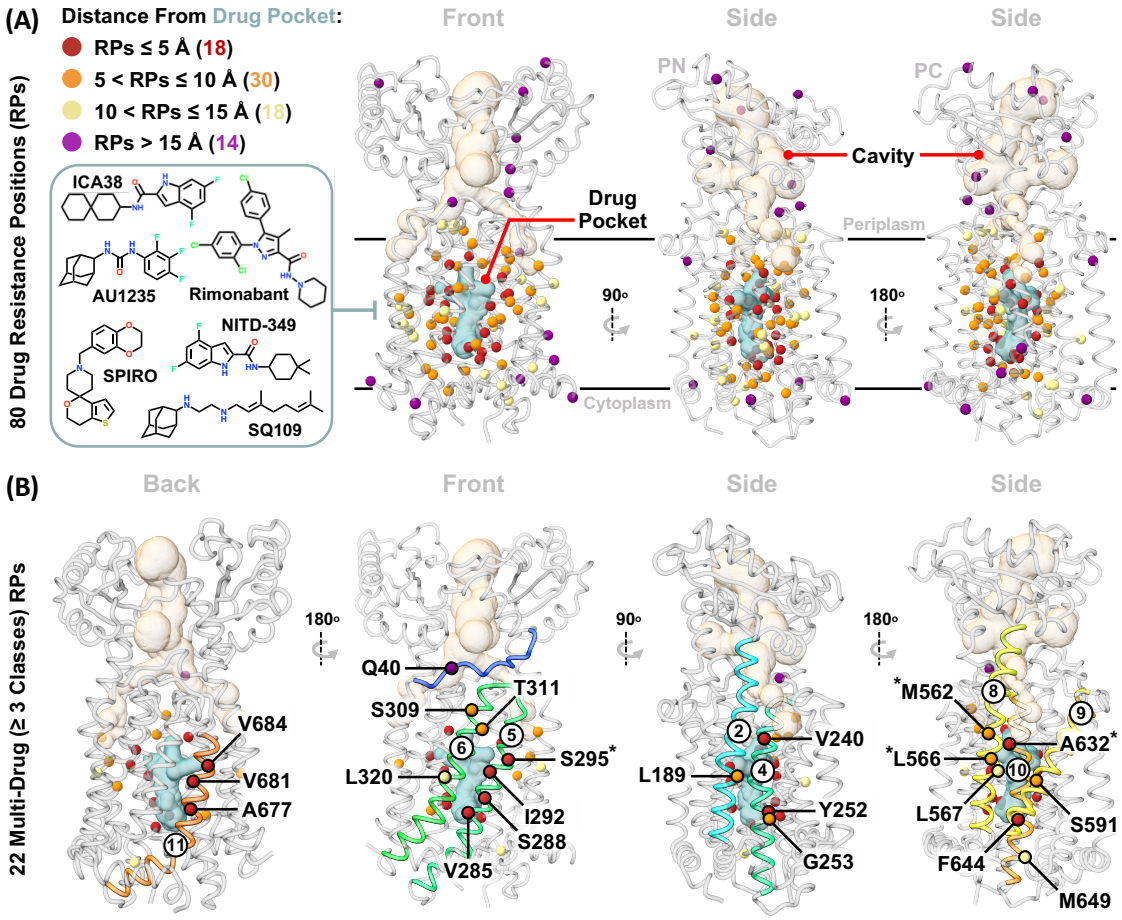


Figure 7.8: Structural Mapping of MmpL3 Drug-Resistance Mutations

(A) 80 resistance positions (RPs) plotted onto the $MmpL3_{1\text{-}753}$ structure as balls and colour-coded by distance from the universal drug pocket (transparent teal); a composite of the inset inhibitors (teal box), generated as in as in Figure 7.4[A]. At each RP one or more missense mutations (collated in Table B.2) are reported to confer resistance to at least one preclinical antitubercular targeting MmpL3. The truncation, in licorice representation (transparent grey), is displayed in multiple orientations viewed from the membrane plane, and is contextualised within a bilayer (black lines). The periplasmic cavity is also shown (transparent wheat surface). The subset of RPs causing cross-resistance to ≥3 MmpL3 inhibitor classes are similarly mapped in **(B)**. Residues at these 22 cross-RPs are annotated, and their associated secondary structure elements coloured as in Figure 7.2[B]. Black asterisks denote cross-RPs thus far only observed in studies of the *M. abscessus* MmpL3 orthologue. Structures visualised in UCSF ChimeraX v.1.1 (Pettersen et al., 2021), and skeletal formulas prepared with ChemDraw Professional v.17.0.0.206.

Analysing the distribution of the mappable resistance positions (RPs) reveals ~90% reside in the TMD, with 60% within 10 Å of the universal drug pocket (Figure 7.8[A]). Such clustering implies the dominant route to resistance is a weakening of MmpL3-drug interactions, through mutation of residues defining, or in the vicinity of, the transmembrane site. This conclusion is consistent with micro-scale thermophoresis (MST) results demonstrating a handful of pocket-local

variants compromise *Msmg* MmpL3 inhibitor affinity *in vitro* (Zhang et al., 2019). Likewise, the distribution is congruous with suggestions that the magnitude and spectrum of the resistance phenotype bestowed by a given MmpL3 missense mutant correlates with its proximity to said drug pocket (Williams et al., 2019). However, as a significant proportion of RPs are located >10 Å away, including some within the LMNG binding site (at Q40, D46, V51, H234, Y235 & L693; asterisks in Figure 7.5[A]), there is conceivably at least one additional, and as yet uncharacterised, alternative MmpL3 resistance mechanism — likely involving the PD and its cavity. Future studies investigating the role of the MmpL3 PD will be required to uncover how these mutations circumvent action of drugs targeting the TMD. As has been argued by others, they may somehow allow MmpL3 to realise viable levels of TMM export despite inhibitor complexation (Li et al., 2019).

Multi-Class Resistance Positions

A greater appreciation of variants capable of conferring resistance to multiple preclinical scaffolds is critical to ongoing efforts to design compounds that are less easy to mutationally evade. Filtering the RPs for those associated with resistance to at least three classes of MmpL3 inhibitor left only twenty-two (Figure 7.8[B]). All, except Q40, fell within the TMD and 50% are within 5 Å of the common pocket. Notably, ten are drug-contacting residues as defined by the *Msmg* structures (Zhang et al. (2019); Yang et al. (2020); compare Figures 7.7[B] & 7.8[B]). TMs 4–6, 8, and 10–11 each host three or more cross-RPs and are thus putative multi-class resistance hotspots.

Residing in the linker joining TM1 to the first PN helix, and therefore incongruous in its separation (>15 Å) from the drug binding site, mutations at Q40 (to R or H; Supplementary Table B.2) almost certainly act by the aforementioned "alternative" resistance route. Consistently, the Q40R variant failed to compromise *Msmg* MmpL3 affinity for inhibitors as measured by MST (Zhang et al., 2019). In also being highly conserved, lining the MmpL3$_{1\text{-}753}$ channel exiting via TMs 1–4

within the LMNG binding site, and coordinating the ethanolamine headgroup of PE in one of the *Msmg* structures (Su et al., 2019), Q40 represents an important mechanistic site, at the crossroads of MmpL3 transport and resistance, and therefore warrants further biochemical study.

Structural interpretation of the physicochemical changes introduced by variants at the other twenty-one cross-RPs identifies a number of plausible explanations for their phenotypes. For example, a facile route to resistance, with minimal fold-disruption, would be conservative binding site mutations that primarily alter sidechain size, thereby impairing steric complementarity between drug and pocket. This approach is apparent at three *Mtb* drug-interacting valines (V240, V285 & V684) that all mutate to smaller residues (A or G; Supplementary Table B.2) to achieve multi-class resistance. The reverse mechanism, enlarging sidechains to promote steric clashes with complexed inhibitors, is employed at two drug-interacting alanines (A632 & A677); which confer cross-resistance when replaced by bulkier aliphatics or aromatics (Supplementary Table B.2). Although more difficult to structurally rationalise, modulation of residue hydrogen bond capacity appears another path to resistance, either through loss (T311I, S591I), gain (I292S/T, L566S), or exchange (S288T) of hydroxyl-containing sidechains (Supplementary Table B.2).

Further away from the shared drug pocket many of the substitutions at cross-RPs result in more dramatic chemical perturbations, in line with both their presumed action at a distance, and a diminished capacity to collaterally incapacitate the proton-relay system. These less conservative variants either introduce additional charge into the TMD (L189R & G253E) or supply intra-TM prolines (S295P, L320P & L567P; Supplementary Table B.2). Combining a restricted dihedral range with an inability to backbone hydrogen bond, mutation to the latter imino acid can be expected to be especially conformationally disruptive.

Interestingly, four of the twenty-two cross-RPs have so far only been observed in, and were hence back-mapped from, *M. abscessus* (asterisks in Figure 7.8[B]). This hints MmpL3 orthologues may preferentially mutate at different residues, or at least with differential frequency at equivalent positions, to facilitate drug escape; an important caveat if, as hoped, antitubercular MmpL3 inhibitors are to be repurposed for treatment of non-tuberculous mycobacterioses in the future (Li et al., 2018; Sethiya et al., 2020).

Resistance Position Stratification by Scaffold

Finally, to further dissect the resistance landscape, RPs were stratified by the nine established MmpL3 inhibitor classes, as well as a tenth miscellaneous grouping of orphan scaffolds (Figure 7.9). All retained a degree of RP clustering around the common drug pocket, corroborating its use by the four classes yet to have a member structurally characterised in complex with MmpL3 (i.e. Piperidinol Derivatives, Benzimidazoles, THPP Derivatives & Benzothioazole Amides). Where structures do exist, finer analysis of the RPs with respect to drug binding pose indicated class-specific differences in RP distribution can often be attributed to subtleties in the interactions each makes with MmpL3. The most obvious example is for the Pyrrole/Pyrazole family (Figure 7.9[A]), where RPs are enriched around a subsite (residues L243, V681, V684, A685 & L703) unused in the binding of the other structurally resolved scaffolds (Zhang et al., 2019). Equally, the RP differences apparent between Indolecarboxamides (Figure 7.9[C]) and Spirocycles (Figure 7.9[G]) agrees with observations that SPIRO, an example of the latter, makes fewer interactions at the top of the shared inhibitor pocket, but more at the bottom, relative to NITD-349, a member of the former family (Yang et al., 2020). Thus, despite significant overlap, there is clear differentiation in the pattern of RPs selected by each drug class.

To summarise, structural mapping of the MmpL3 resistance landscape showed mutational disruption of the TMD drug pocket to be the primary route to mono- and

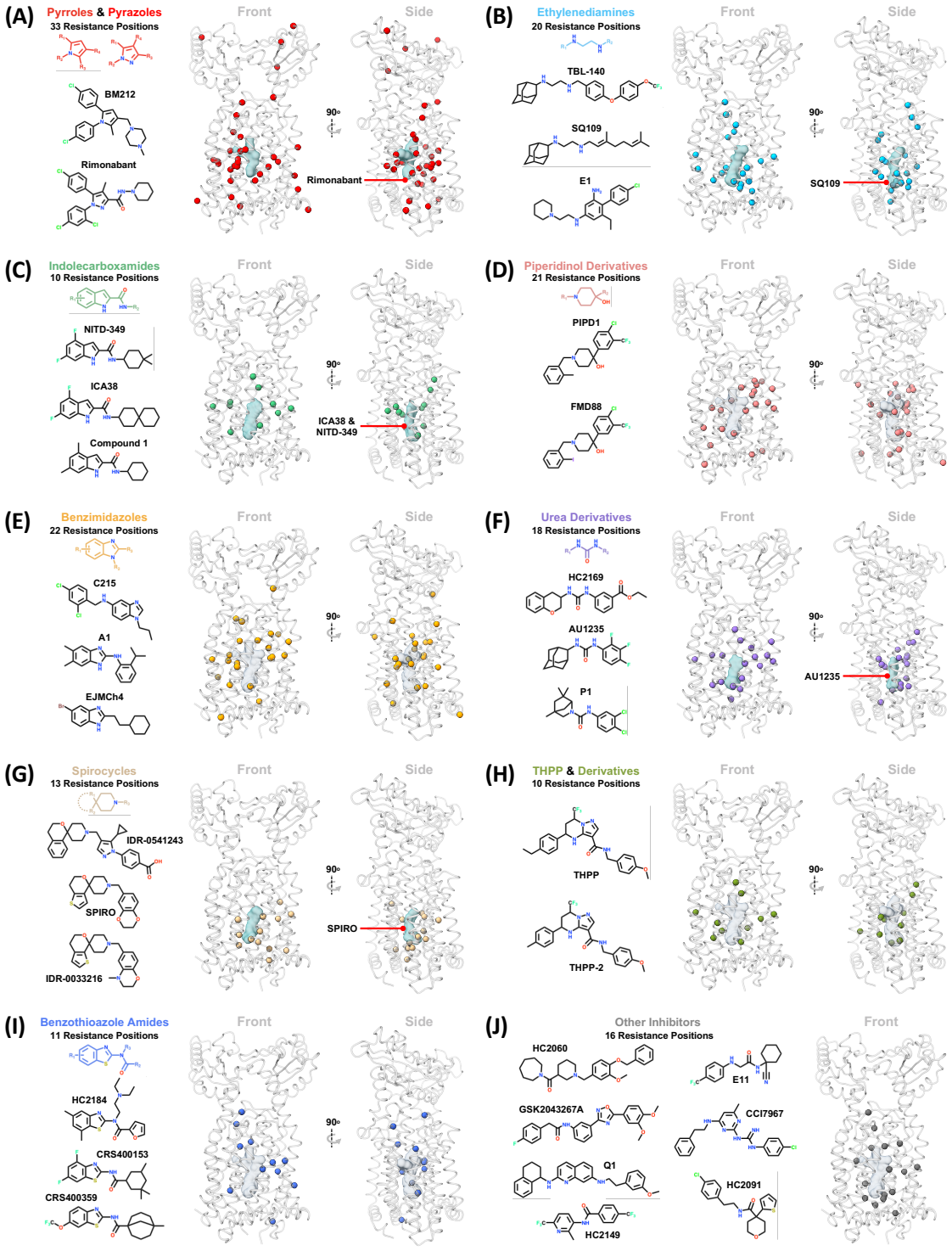


Figure 7.9: Stratification of Drug-Resistance Mutations by Inhibitor Class

(A) – (J) Discrete mapping of MmpL3 missense variants conferring resistance to nine preclinical inhibitor classes (A – I), alongside a tenth grouping of orphan scaffolds (J). Resistance positions (RPs) are depicted by balls, and the protein (transparent grey, licorice) viewed from the membrane plane. For class members co-crystallised with *Msmg* MmpL3 (PDBs: 6AJG to 6AJJ, 7C2N & 7C2M, Zhang et al. (2019); Yang et al. (2020)) a surface representation is superimposed (transparent teal) and annotated, otherwise the composite drug pocket is shown (transparent grey, generated as in Figure 7.4[A]). Exemplar agents are given for each family, and where informative a skeletal formula of the core class scaffold provided. Mutations are collated in Table B.2. Composed with UCSF ChimeraX v.1.1 (Pettersen et al., 2021) and ChemDraw Professional v.17.0.0.206.

multi-class resistance, implicated disparate physicochemical changes in resistance conferral (i.e. steric effects, altered hydrogen bonding potential, intra-TM proline introduction etc.), and indicated existence of a probable subsidiary resistance mechanism related to the PD cavity. Going forward, these insights, alongside active curation of a variant database resembling Supplementary Table B.2, have the potential to guide both structure-based drug design (e.g. compound optimisation to limit contacts with cross-RPs) and, one day, targeted prescription of MmpL3 inhibitors in the clinic.

7.4.2 The Wider Non-Synonymous Landscape of *Mtb* MmpL3

In the context of the sizeable collection of *Mtb* genomes accessible from the European Nucleotide Archive (ENA), the preclinical status of MmpL3-targeting antituberculars presents a unique opportunity. As the ENA-deposited isolates largely originate from clinical samples, unexposed to any experimental agent, these sequences can be cross-referenced with the resistance catalogue (Supplementary Table B.2) to gauge the extent of any pre-existing resistance MmpL3 inhibitors may encounter once licensed. To this end, 45,878 *Mtb* genomes were retrieved from the ENA and mined for non-synonymous mutations in the *Rv0206c* (i.e. MmpL3) ORF. Concordant with its essentiality, no MmpL3 variants were detected in the majority (91.5%) of samples. The remaining 3,890 genomes harboured 7,475 non-synonymous mutations (collated in Supplementary Table B.3), distributed over 597 distinct amino acid substitutions at 439 different positions in MmpL3 (Figure 7.10). Although the bulk of variants (422/597, 71%) were identified only three or fewer times, the most abundant 25, each observed ≥30 times, account for almost three-quarters of the dataset (5,374/7,475, 72%). Considering just the 25 prevalent mutations to begin with, they are dispersed throughout MmpL3 (Figure 7.10) and, reassuringly, none have been reported in the resistance literature to date (Supplementary Table B.2). Whilst it cannot be ruled out that some are resistance-conferring, one can surmise that it is

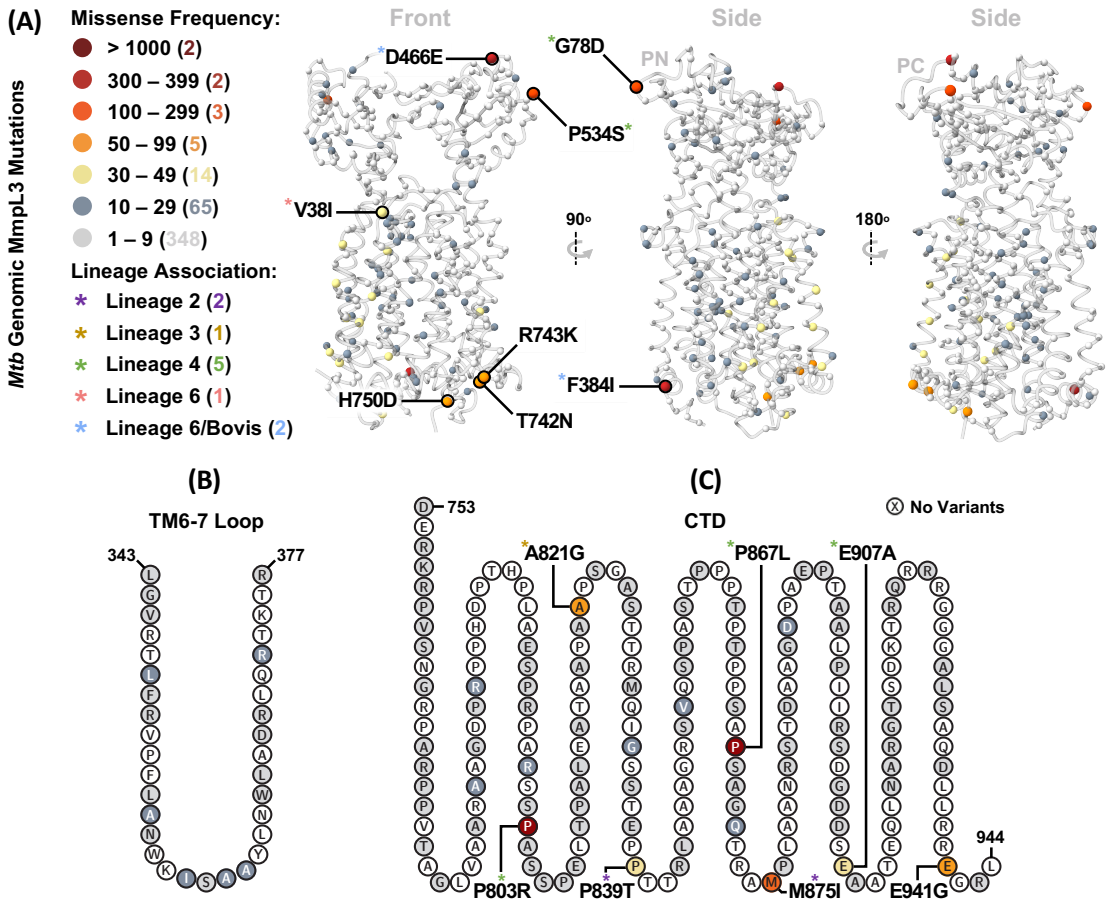


Figure 7.10: Visualising the Wider Non-Synonymous Landscape of *Mtb* MmpL3

(A) Structural mapping of MmpL3 missense variant positions mined from >45,000 *Mtb* genomes (collated in Table B.3), depicted by balls and colour-coded by mutational frequency. $MmpL3_{1\text{-}753}$ is shown in licorice representation (transparent grey). The most abundant substitutions, and/or those with apparent *Mtb* lineage-association (coloured asterisks, see Table B.4), are labelled. Incidence of non-synonymous mutations in the unresolved TM6–7 loop (residues 343–377) and removed CTD (residues 753–944) are presented in the equivalently coloured primary sequence representations of **(B)** & **(C)** respectively. Structures prepared with UCSF ChimeraX v.1.1 (Pettersen et al., 2021), and topology depictions with TEXtopo (Beitz, 2000a). Data provided courtesy of Philip Fowler, on behalf of the CRyPTIC Consortium.

unlikely. As may be expected, the chemical perturbation introduced by these variants tends to be smaller in the TMD than at the protein periphery (i.e. PD & CTD), mirroring orthologue domain conservation (Supplementary Figure B.18). Similarly rationalisable, the only mutations recorded >1,000 times are found in the CTD (1,979 instances of P803R & 1,506 instances of P867L). For many samples *Mtb* lineage was also genomically inferred (Supplementary Table B.4). This revealed a subset of the 25 to be ostensibly lineage-specific (asterisks in Figure 7.10), suggesting they are ancestral mutations (Comas et al., 2013).

Conversely, the lack of abundant non-lineage defining variants implies there is currently little selection pressure on MmpL3 and/or non-synonymous changes tend to be deleterious.

The most frequent resistance-conferring mutation in the genome-mined dataset, V210A (in TM3), occurs in just 21 samples. Although 16 of these are from but three studies — Guerra-Assunção et al. (2015), Colangeli et al. (2018) & Pankhurst et al. (2016) — all three sequence clinical isolates, unpolluted by exposure to novel MmpL3 inhibitors. A further 12 resistance-conferring substitutions are present in an additional 32 genomes in the ENA (highlighted red in Supplementary Table B.3), of which only three (G628A, Y252C, I585V) are recorded >5 times. Simply counting these, and keeping in mind that the minority of genomes derived from research isolates may be over-enriched for resistance, one arrives at an estimate of 0.1% (53/45,878) as the upper limit for the environmental prevalence of MmpL3 resistance mutations. This low percentage provides evidence that MmpL3 inhibitors, on debut in the clinic, should initially be efficacious against almost all circulating *Mtb* strains.

7.5 Final Remarks & Future Perspectives

At the outset of my DPhil *Mtb* MmpL3 had already been established as a promising antitubercular target, SQ109 having entered Phase II clinical trials many years earlier (Sacksteder et al., 2012), yet the essential RND transporter remained structurally uncharacterised. In fact, significant information on the architecture of any MmpL protein was lacking; only minimal insights gleanable from a crystal structure of an MmpL11 periplasmic subdomain (Chim et al., 2015), as well as a neg-stain envelope of the *Cgl* orthologue CmpL1 (Belardinelli et al., 2016). Accordingly, a project seeking to structurally resolve a substantive MmpL construct, inclusive of its TMD and ideally of *Mtb* MmpL3, was embarked upon. It was hoped this would shed light on the mechanisms underlying MmpL3 TMM export,

a crucial step in the biogenesis of the *Mycobacteriaceae* cell wall, and in turn help delineate the mode of action of the numerous MmpL3 inhibitors under development (Degiacomi et al., 2020).

Through extensive homologue screening, exploring the nexus of variables relevant to recombinant membrane protein production (e.g. expression host, affinity tag, vector & detergent), a C-terminally truncated construct displaying the requisite properties for structural pursuits — *Mtb* MmpL3$_{1\text{-}753}$ — was identified (Chapter 3). Despite prodigious efforts, the truncation proved recalcitrant to both crystallisation and Cryo-EM reconstruction in a multitude of conditions (Chapter 4), including when augmented by synthetic binders (Chapter 5) or reconstituted back into a lipid bilayer (Chapter 6). Reinforced by recurrence throughout these Chapters is the influence, a thus importance, of detergent choice. In each the dissimilar attributes of MmpL3$_{1\text{-}753}$ in DDM versus LMNG proved prominent; be it in terms of differential SEC monodispersity, thermal stability, in-ice behaviour, epitope presentation and sybody compatibility, or outcomes of ND assembly.

Persevering, late in the DPhil Cryo-EM of LMNG solubilised MmpL3$_{1\text{-}753}$ was successfully revisited and culminated in a 3.0 Å map. Although notably reliant on technical advancements emergent since project inception, both expediting micrograph collection (e.g. the K3 DED) and transforming data processing capabilities (e.g. bayesian polishing, Zivanov et al. (2019); NUR, Punjani et al. (2020)), the reconstruction is nevertheless an impressive feat given the truncation presents as monomeric, psuedo-symmetric, particles with less than 82 kDa of ordered mass. The resultant atomic model revealed MmpL3$_{1\text{-}753}$ to house a sizeable periplasmic cavity, whose outer leaflet connectivity and serendipitous LMNG entrapment clearly implicate it as the TMM translocation conduit. By this point the last of the eight *in surfo* structures of *Msmg* MmpL3 had been reported (Zhang et al., 2019; Su et al., 2019; Yang et al., 2020). In resolving the protein in complex with six different inhibitors, from nearly as many classes, these studies brought

about a seismic change in the understanding of MmpL3-drug interactions. They ultimately revealed a "universal" TMD drug pocket; defining severance of the Asp-Tyr network, and thus transporter de-energisation by obstruction of proton conductance, as a common mechanism of action. In needing to resort to the non-pathogenic orthologue however, the *Msmg* publications implicitly reaffirm the challenges presented by the *bona fide* antitubercular target. By drawing together the now nine structures of MmpL3, their comparative analysis in the current Chapter, the limited MmpL3 biochemical literature, and inferences made by analogy with better characterised RND systems (i.e. AcrB, ESTs), the following model of TMM export can be reasonably proposed — summarised in Figure 7.11; individual steps are referenced in-text by the corresponding numeral.

7.5.1 A Putative Model for TMM Export

Newly synthesised TMM is first flipped from the inner to the outer leaflet of the plasma membrane (①); conceivably by MmpL3 guiding the lipid along one of its shallow TMD grooves (Xu et al., 2017; Bolla, 2020), in manner similar to that conjectured for PTCH1 and cholesterol (Zhang et al., 2018). Possibly concomitant with this inter-leaflet transferal, or soon there after, outer leaflet TMM next binds either of said indentations (i.e. TMs 1–4 or 7–10); imitating the DDM and 6DDTre molecules resolved in two of the *Msmg* structures (PDBs: 6AJG & 6AJI, Zhang et al. (2019)). From here, TMM is withdrawn from the membrane, into MmpL3, through one of the pair of channels branching downward from the primary PD vestibule (②). This, and subsequent steps in the transport cycle, necessitate conformational change in MmpL3, almost certainly driven by influx of protons down their electrochemical gradient, through the TMD (③), and considered later in greater detail. Now in a "post-extraction" state (④), TMM is initially accommodated as a splayed species resembling the *Mtb* MmpL3$_{1\text{-}753}$-LMNG complex. Within the PD, in line with the hypothesised "division of labour" framework for lipid handling, the alkyl chains of TMM bifurcate to be engaged by

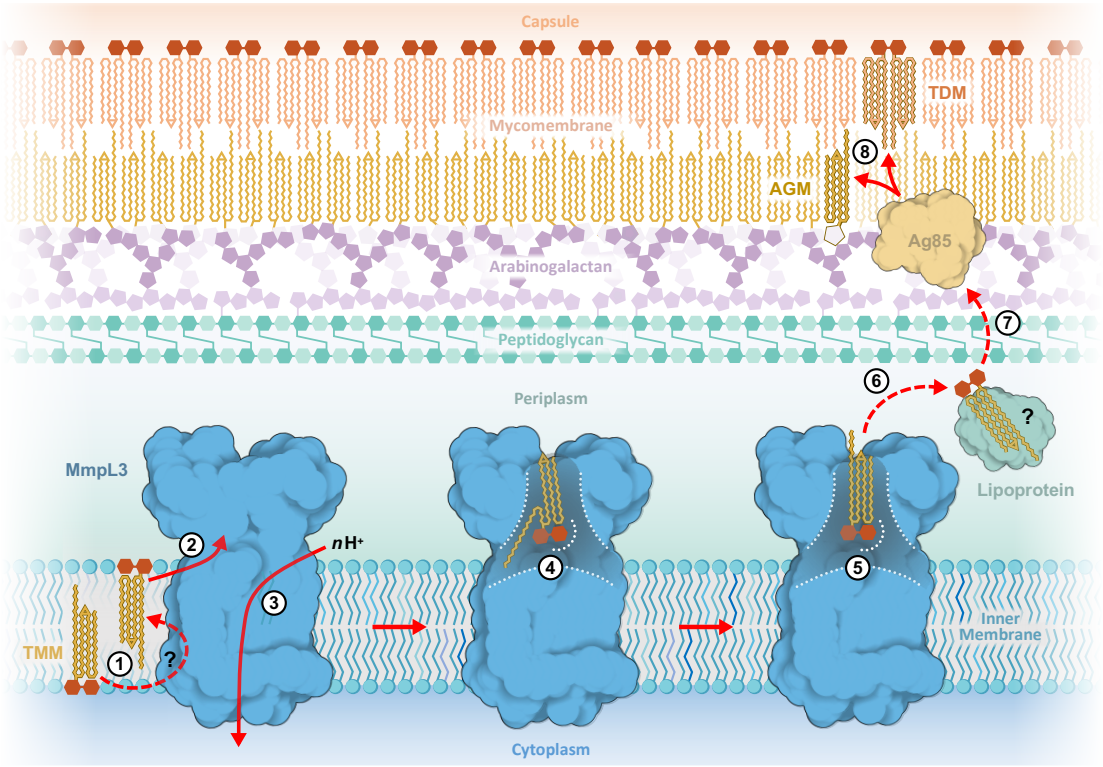


Figure 7.11: Proposed Model for MmpL3-Mediated TMM Export

A simplified model of TMM export, formulated by synthesising MmpL3 structural insights with the wider literature. **(①)** Flipping of TMM from the inner to the outer plasma membrane leaflet. **(②)** Uptake of outer leaflet TMM into the PD cavity via groove-bound inlets. **(③)** TMD transduction of the PMF, energising the conformational changes underlying the transport cycle. **(④)** A splayed TMM transport intermediate, modularly bound by PD tail pockets and aperture-based headgroup coordination. **(⑤)** TMM manipulation into a pre-handover state, with alkyl chains extending up towards and partially through PD_T. **(⑥)** TMM hand-off to a periplasmic chaperone via PD_T. **(⑦)** Lipoprotein transit of TMM across the cell envelope. **(⑧)** Transesterification of TMM by the mycosyltransferases of the Ag85 complex, to give the mycomembrane constituents AGM or TDM. The MmpL3 CTD and binding partners are not depicted. Darkened regions, delimited by dotted borders, designate the PD cavity network and its openings. Question marks denote protein activities or identities yet to be firmly established. Figure partly composed with Illustrate (Goodsell et al., 2019).

defined tail pockets, one up towards PD_T and the other trailing into the entrance channel, sequestering them away from the bulk periplasm. Meanwhile, the trehalose headgroup is encircled by the PD_F or PD_B aperture, which coordinates it via hydrogen bonds and provides it a degree of solvent access. The penultimate export step sees MmpL3 manipulate its lipidic cargo into a "pre-handover" state (⑤), reminiscent of the *Msmg* MmpL3-PE complex (PDB: 6OR2, Su et al. (2019)). Potentially employing the aperture-headgroup interaction as a fulcrum, MmpL3

reconfigures TMM so both its tails point upwards, partially extending through PD_T. If not already pre-emptively docked, one of a number of speculatively redundant periplasmic chaperones (Melly et al., 2019) arrives to receive the lipid as it is extruded by MmpL3, tails first, through PD_T (⑥). Once the lipoprotein departs, shuttling TMM (⑦) to the Ag85 complex (⑧) for mycomembrane incorporation as AGM or TDM (Jankute et al., 2015), MmpL3 resets to begin the transport cycle anew.

The suite of MmpL3 structures, in being resolved in near identical conformations, offer little insight into how the PMF is mechanically transduced to execute the above TMM efflux model. However, given that the TMD fold is the most conserved facet of RND architecture, one can assume that the conformational changes it undergoes are likely to be comparable, at least superficially, between superfamily members. As such, by analogy with the well documented transport mechanism of AcrB (Eicher et al., 2014), the following can be surmised. MmpL3 D251 & D640, of the Asp-Tyr cluster bridging TMs 4 & 10, undergo sequential protonation and deprotonation events. This is achieved by an alternating access mechanism in which they are first exposed to the periplasm, for proton loading, and then the cytoplasm, for proton release; in each case receiving or dispatching protons with the bulk solvent by means of "proton wires", composed of suitably labile sidechains and/or transient water channels. The former likely include the smattering of charged TMD residues, beyond D251 & D640, implicated in MmpL3 function by SDM (i.e. R259, D710 & R715, Belardinelli et al. (2016)). These cyclical changes in Asp-Tyr protonation state, in turn reconfiguring its interaction network, induce the structural repeats of the TMD to shift relative to one another. Such movements presumably have direct consequences for TMM transport due to their expected relay, up into the PD, by the extramembrane extensions of TMs 2 & 8. Pertinently, residues coupling these extended helices to TMs of the opposing repeat (i.e Q304 & S691) have been intimated, again by SDM (Belardinelli et al.,

2016), in MmpL3 activity. Also noteworthy, any lateral motion in TMs 2 or 8 can be anticipated to modulate the openness the adjacent TMM entry channel.

Future delineation of the order and interplay of conformational changes occurrent in the TMD and PD is essential to a proper understanding of the MmpL3 transport cycle. Validation, elaboration and/or refutation of the proposed TMM/H^+-antiport model awaits further structural endeavours; ideally resolving MmpL3 in complex with TMM, in absence of lipophilic PD ligands (i.e truly *apo*), at a range of pHs to dissect conformation as function of protonation state, and, to a similar end, with residues of the Asp-Tyr tetrad mutated to inert alternatives (i.e. N, F & A). Certainly equally important will be use of *in silico* and biochemical techniques to expand upon, develop, and test hypotheses emergent from the current, and hopefully continually amassing, collection of MmpL3 structures. For example, to confirm the functional relevance and/or redundancy of the PD cavity, one could assess the *in vivo* viability of constructs in which its openings are selectively, or combinatorially, occluded by introduction of artificial salt-bridges. Computationally, simulations should be fruitfully employed in calculation of sidechain pK_as to disclose the probable proton-relay route, as well as analyse protein dynamics in various Asp-Tyr protonation states. Beyond the minutiae of the MmpL3 mechanism, a number of fundamental and conceptual questions concerning the protein remain; a selection of which are considered with brevity next.

7.5.2 Outstanding Fundamental Questions

What are the Benefits of Monomerisim to MmpL3?

The weight of evidence points to MmpL3 functioning as a monomer, representing a stark departure from the homotrimerisim displayed by almost all structurally characterised prokaryotic RND systems. In likewise sharing a function in endogenous lipophile transport with the vast majorty of other known monomeric RND proteins (i.e. HpnN & ESTs), one must speculate single-chain architecture

bestows significant advantages for the execution of such tasks by the RND fold. The nature of these theorised benefits remains to be determined, and may only fully come to light on greater structural and mechanistic understanding of the entire superfamily.

What is the Role of Symmetry in MmpL3?

The striking two-fold pseudo-symmetry of MmpL3, a consequence of evolution by internal gene duplication (Tseng et al., 1999), leads one to question whether the two halves of the protein undergo equivalent conformational changes or, maybe more likely, have become specialised to perform distinct, complementary, roles in TMM export. Such lines of inquiry extend to the comparably symmetric PD cavity — are both or only one of the entrance channels emptying into the outer leaflet used for TMM uptake? Of possible relevance, the grooves hosting the channel inlets are chemically dissimilar; the side defined by TMs 7–10 being the more hydrophobic of the pair for example. Moreover, structurally resolved PD ligands are only seen to occupy the channel exiting via TMs 1–4; lending credence to its preferential use in TMM abstraction which, if correct, parallels the lipophile transit asymmetry proposed for the tunnels of HpnN (Kumar et al., 2017) and some ESTs (Zhang et al., 2018; Winkler et al., 2019).

What is the Purpose of the MmpL3 CTD?

The enigmatic and non-essential MmpL3 CTD, its intrinsic disorder proving to be a hindrance to structural pursuits of both orthologues, has been acutely under studied. Repeatablely linked to localisation of the transporter to foci of envelope biogenesis (i.e. the cell poles & septum; Carel et al. (2014), Belardinelli et al. (2019)), how the ~200 residue, highly charged, and proline-rich cytoplasmic domain, which is poorly conserved with exception of its extreme C-terminus (Supplementary Figure B.18), achieves such recruitment is far from clear. It may well exploit its unstructured character to act as a protein-protein interaction hub, a role similarly suggested for the disordered intracellular termini of some human

membrane proteins (Bürgi et al., 2016). Intriguingly, multiple CTD sites have been reproducibly identified in mycobacterial phosphoproteomic studies (Prisic et al., 2010; Fortuin et al., 2015; Carette et al., 2018; Zeng et al., 2020; Le et al., 2020); hinting MmpL3 is likely subject to some form of phospho-regulation. Whether post-translational modification affects transporter localisation, kinetics, turnover, binding partner affinity, or acts by other means, is unknown.

What do MmpL3 Binding Partners Contribute to TMM Export?

The list of MmpL3 protein interactors is steadily increasing, with over twenty having now been identified using an array of approaches (e.g. two-hybrid screens, pull-downs, surface plasmon resonance). Of these, most interesting with respect to lipid handling are the putative TMM lipoprotein LpqN (Melly et al., 2019), the TMM acetyltransferase TmaT (Belardinelli et al., 2019), and TftA (Fay et al., 2019); a protein of unknown function. The last two are both essential transmembrane proteins evidenced to be necessary for TMM transport (Belardinelli et al., 2016; Fay et al., 2019). How such accessory proteins interact with MmpL3, and how their interactions coordinate lipid translocation, must be resolved if a more holistic, and thus accurate, model of TMM export is to be realised.

Overall, the structural analysis of *Mtb* MmpL3 reported herein represents substantial progress towards elucidating the mechanism of TMM transport, with implications for the understanding of the wider MmpL family. Synergising with studies of the *Msmg* orthologue, the research paves the way for an era of structure-based drug design — prearmed with knowledge of the underlying resistance landscape — against one of the most important antitubercular targets identified in recent decades. Going forward, this body of work should provide a solid foundation for pursuit of outstanding questions relating to MmpL3 physiology, which on resolution will undoubtedly reveal new therapeutic avenues. Ultimately the future for MmpL3 inhibitors is promising, being poised to make a sizeable impact in combating a pathogen that has plagued humanity for many thousands of years.

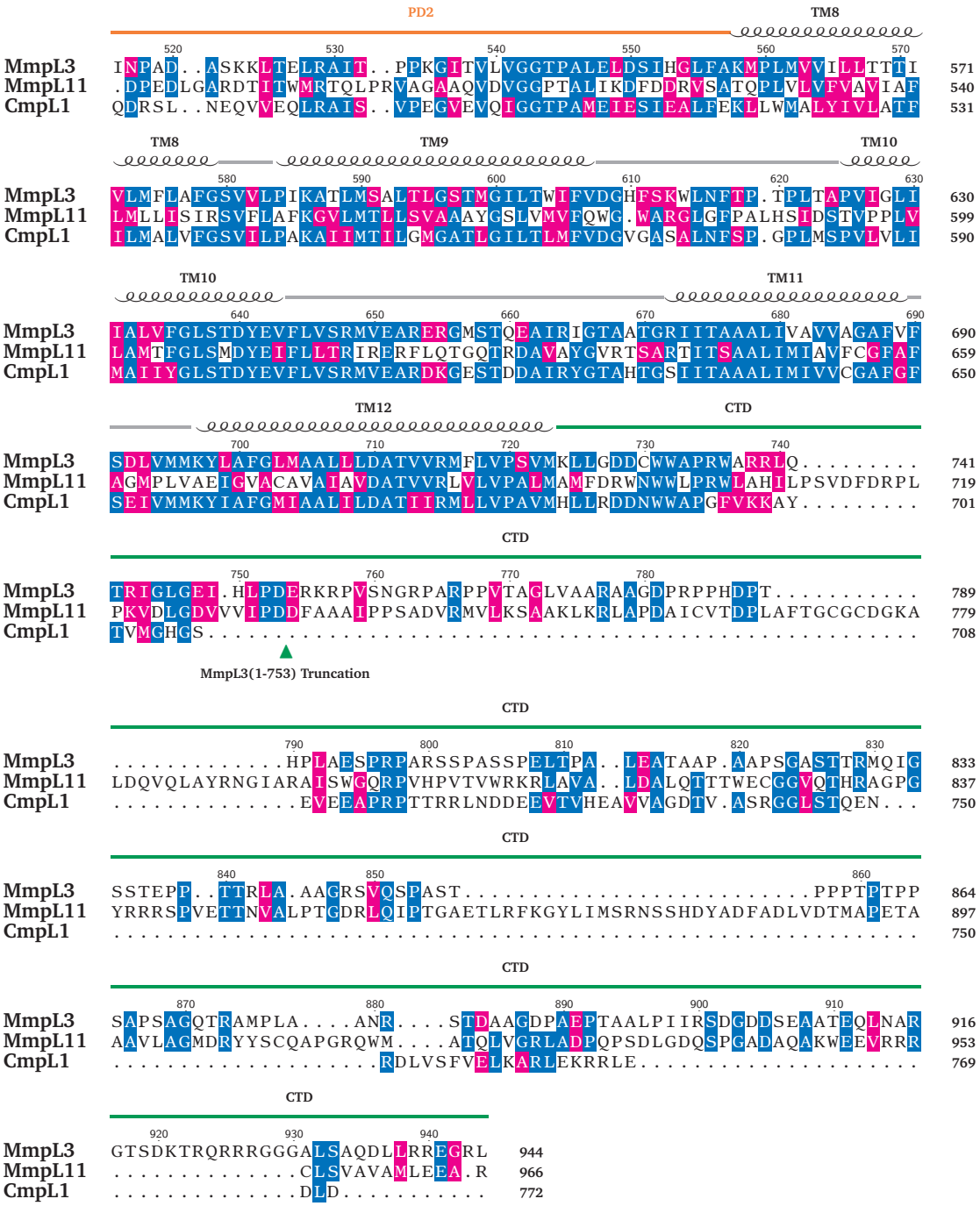
Appendices

Sequence Alignments

APPENDIX A

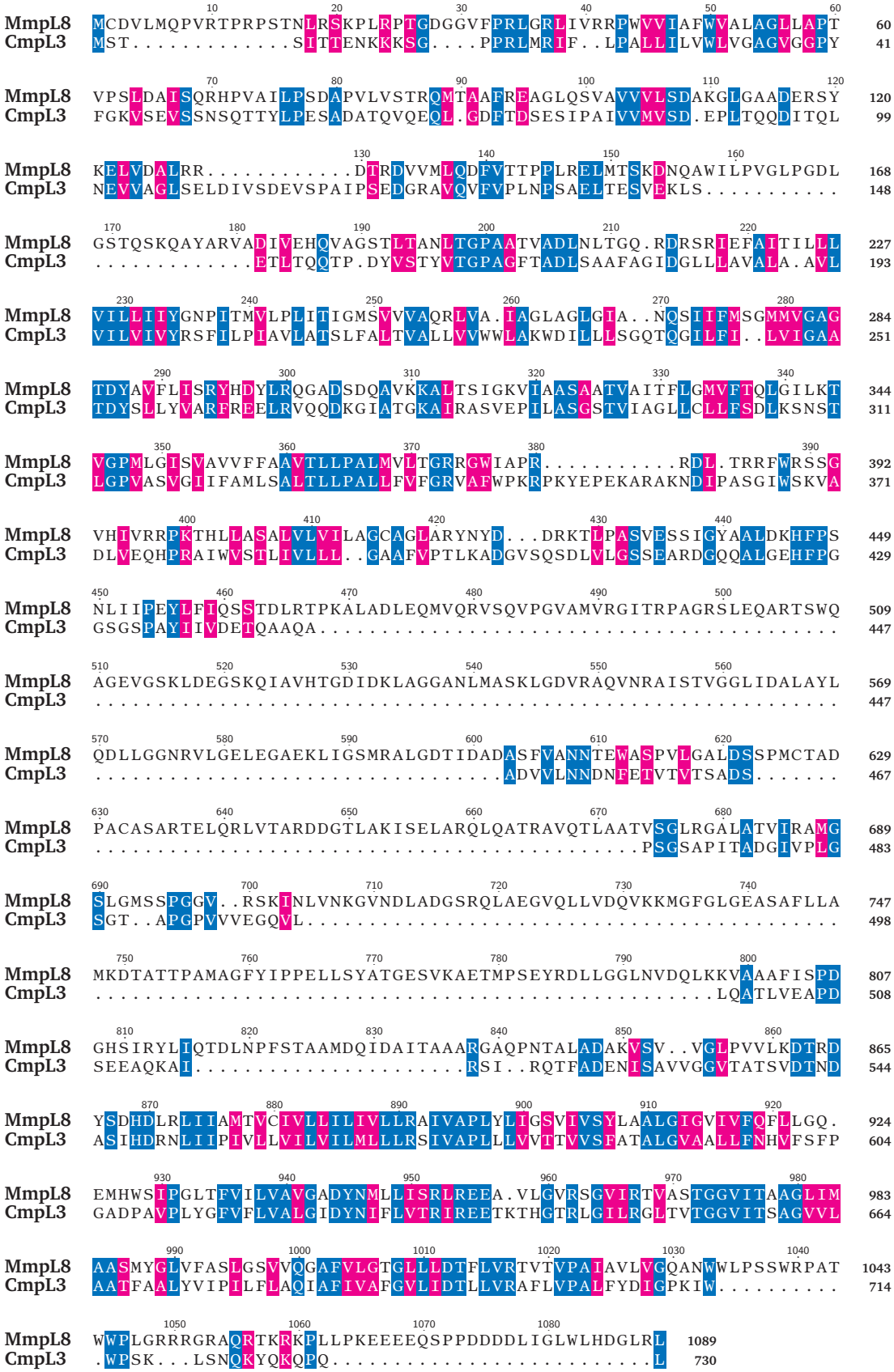
Figure A.1: Sequence Alignment of *Mtb* MmpL3, *Mtb* MmpL11 and *Cgl* CmpL1





MmpL3—MmpL11—CmpL1 alignment generated with T-Coffee (Madeira et al., 2019), coloured by residue identity (blue) and similarity (magenta) within TEXshade (Beitz, 2000b), and annotated with MmpL3 transmembrane helix (TM, spirals) positions predicted with MEMSAT3 (Jones, 2007). Approximate locations of the MmpL3 periplasmic (PD, orange) and C-terminal (CTD, green) domains are highlighted, as is the site of MmpL3 truncation (D753) employed in the current study. The proposed cryptic TEV protease cleavage site of CmpL1, IDIYVQ↓A, is also labelled. MmpL3–CmpL1: 43% identity, 61% similarity. MmpL3–MmpL11: 26% identity, 46% similarity. MmpL11–CmpL1: 27% identity, 46% similarity.

Figure A.2: Sequence Alignment of *Mtb* MmpL8 with *Cgl* CmpL3



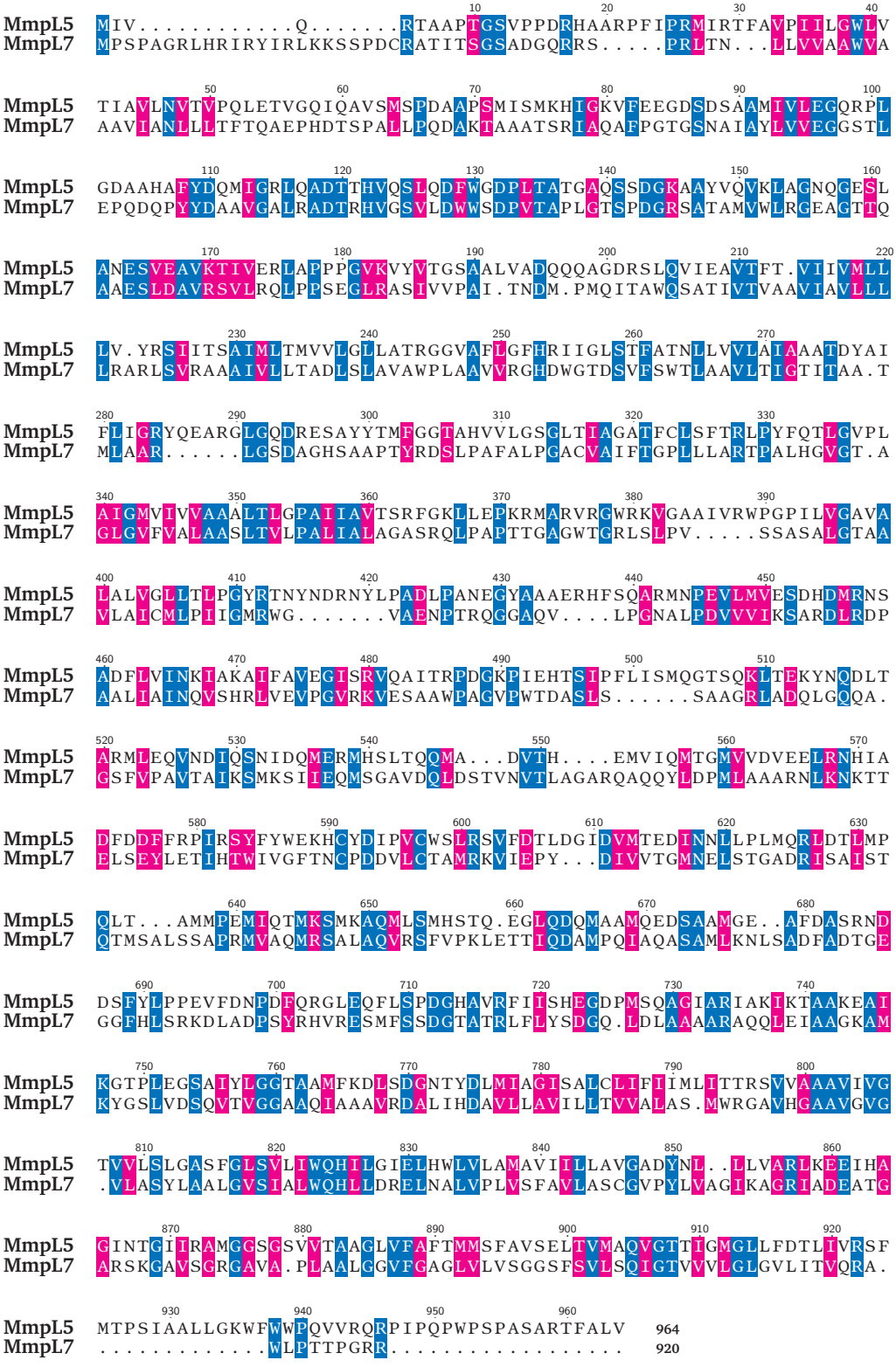
MmpL8—CmpL3 alignment generated with T-Coffee (Madeira et al., 2019) and coloured by residue identity (blue, 37%) and similarity (magenta, 58%) within TEXshade (Beitz, 2000b).

Figure A.3: Sequence Alignment of *Mtb* Rv1250 with *Cgl* NCgL-1194



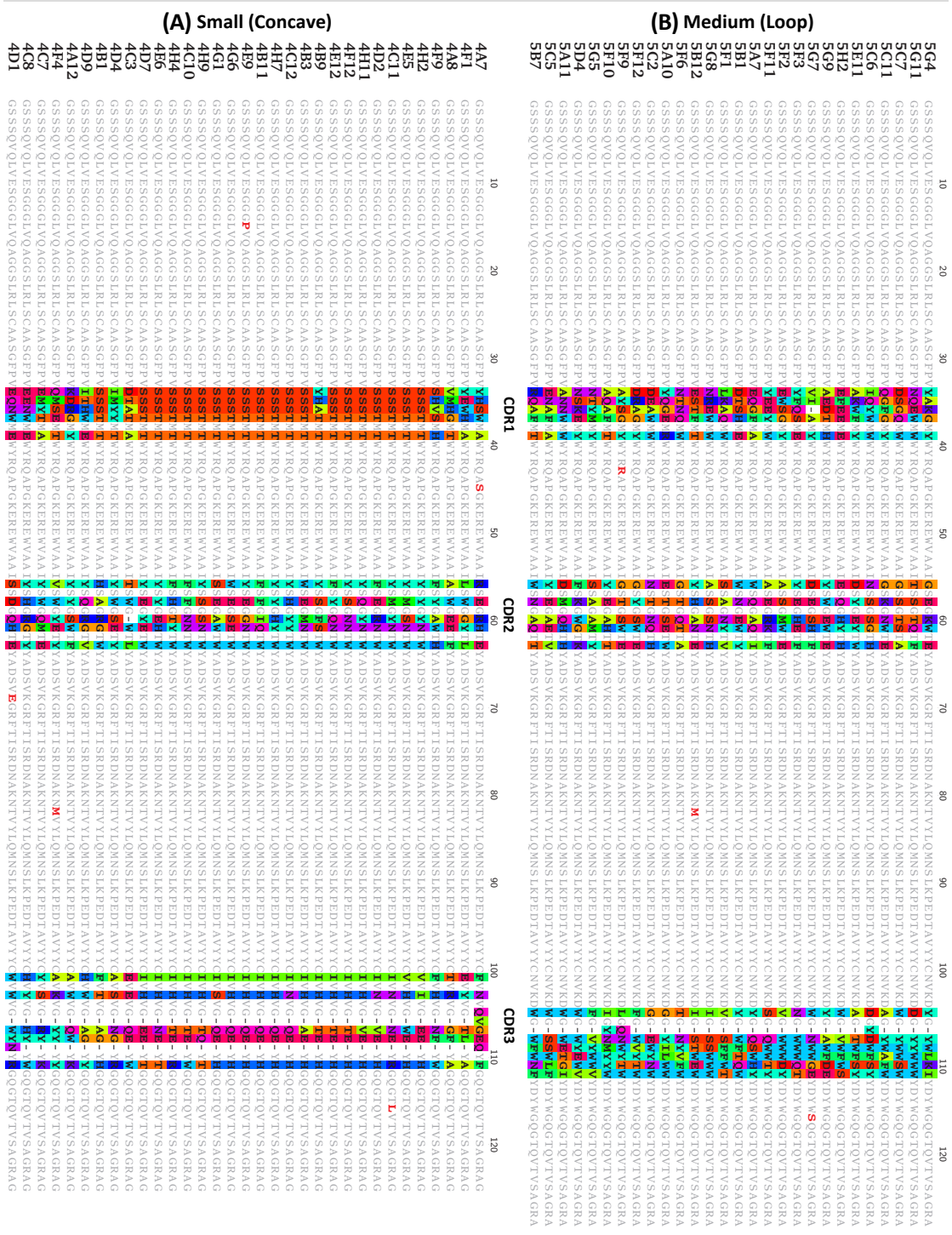
Rv1250—NCgL-1194 alignment generated with T-Coffee (Madeira et al., 2019) and coloured by residue identity (blue, 45%) and similarity (magenta, 63%) within TEXshade (Beitz, 2000b).

Figure A.4: Sequence Alignment of *Mtb* MmpL5 with *Mtb* MmpL7



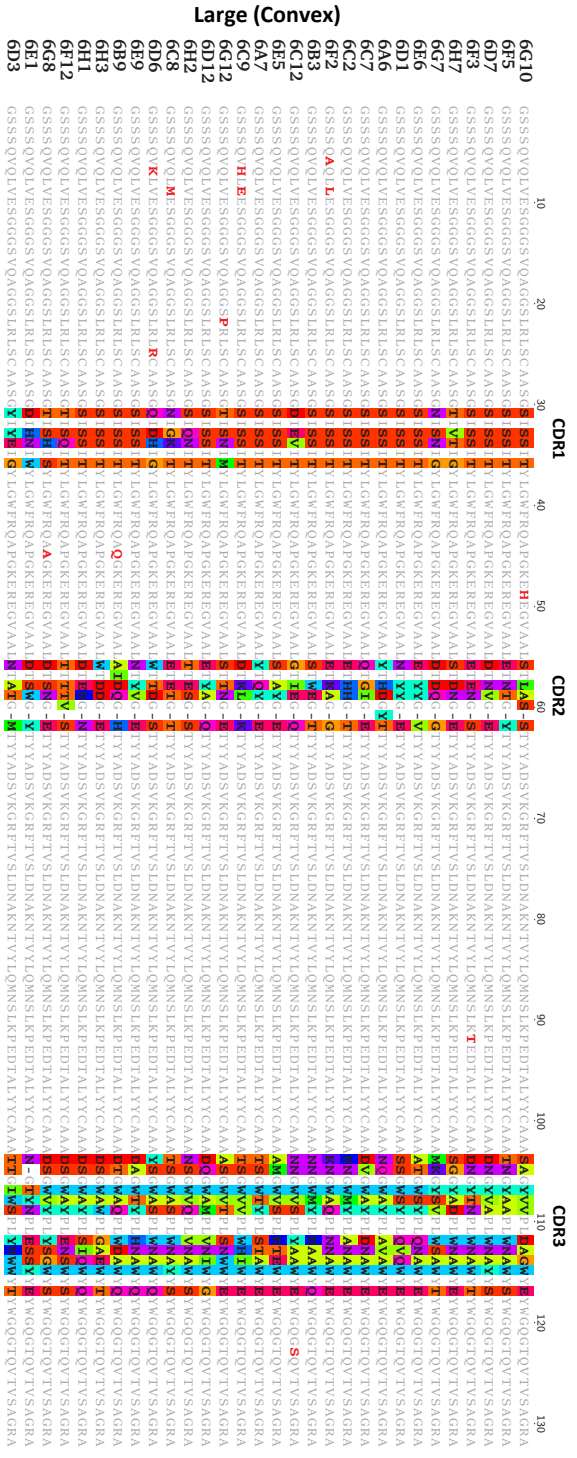
MmpL5—MmpL7 alignment generated with T-Coffee (Madeira et al., 2019) and coloured by residue identity (blue, 24%) and similarity (magenta, 44%) within TEXshade (Beitz, 2000b).

Figure A.5: Sequence Alignments of Concave and Loop Library Sybodies



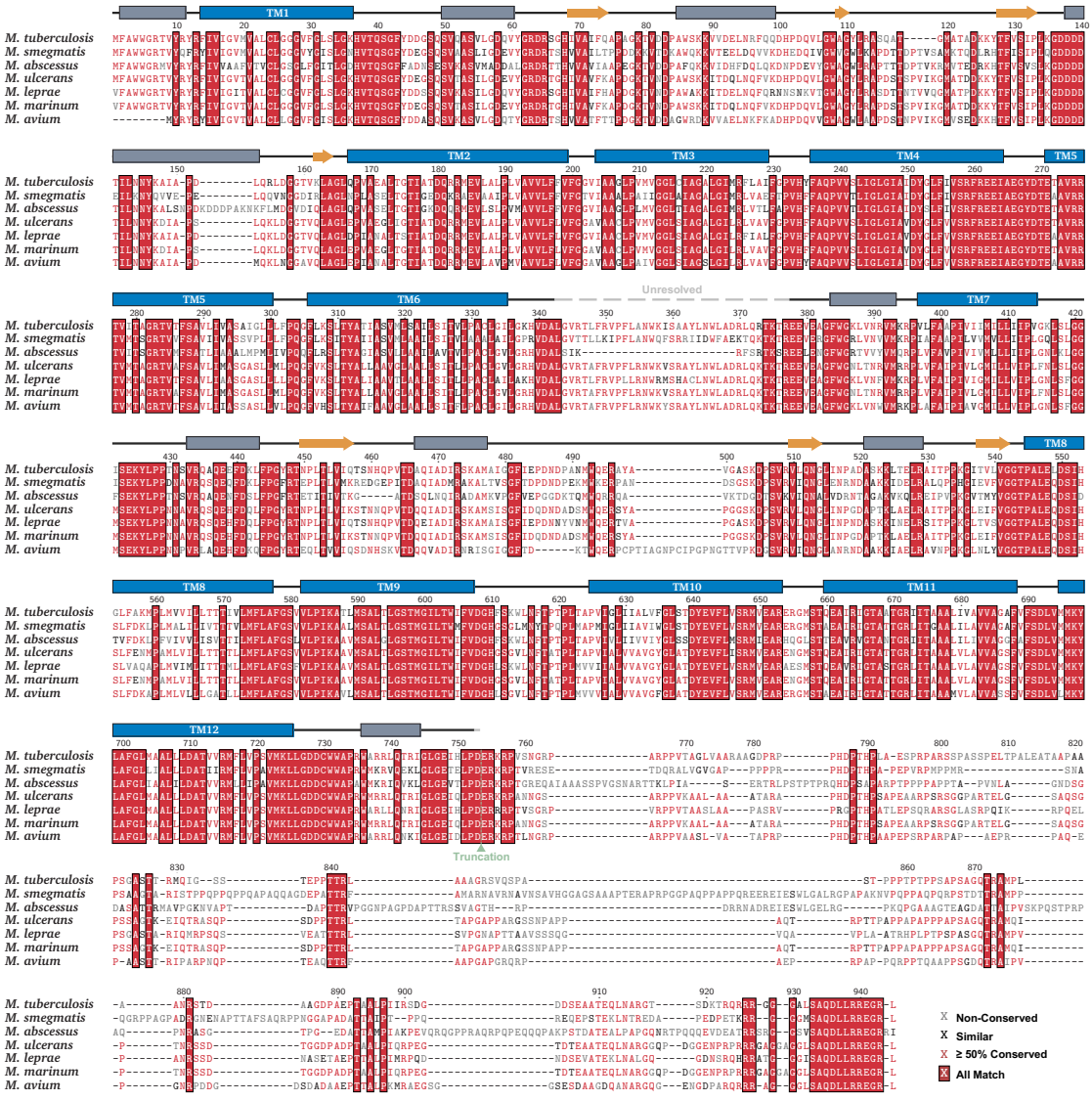
Sequence alignment of **(A)** concave and **(B)** loop sybodies selected against MmpL3$_{1\text{-}753}$ in DDM; ordered by CDR3 homology. Invariant framework positions are grey, CDR randomised residues shaded according to the Taylor scheme (Taylor, 1997), and framework point mutations emboldened in red. Aligned via T-Coffee (Madeira et al., 2019), manually edited in Jalview (Clamp et al., 2004), and typeset with TEXshade (Beitz, 2000b).

Figure A.6: Sequence Alignment of Convex Library Sybodies



Sequence alignment of convex sybodies selected against $MmpL3_{1\text{-}753}$ in DDM; listed by CDR3 similarity. Non-variable residues are grey, CDR randomised positions coloured using the Taylor scheme (Taylor, 1997), and framework point mutations emboldened in red. Aligned via T-Coffee (Madeira et al., 2019), adjusted in Jalview (Clamp et al., 2004), and composed with TEXshade (Beitz, 2000b).

Figure A.7: Sequence Alignment of MmpL3 Orthologues



Sequence alignment of *Mtb* MmpL3 with orthologues from *Msmg* (61% identity), *M. abscessus* (56% identity), *M. ulcerans* (76% identity), *M. leprae* (76% identity), *M. marinum* (76% identity), and *M. avium* (70% identity). Positions of secondary structure elements, derived from the $\text{MmpL3}_{1\text{-}753}$ model, are indicated. Grey and blue rectangles denote non-transmembrane and transmembrane helices (TMs) respectively, and yellow arrows label β-strands. Unresolved regions are shown by a dashed grey line and the C-terminal truncation at D753 is annotated. Residues conserved in all presented orthologues are enclosed in red boxes. Aligned using T-Coffee (Madeira et al., 2019) and typset with TEXshade (Beitz, 2000b).

Appendices

Supplementary Figures & Data

APPENDIX B

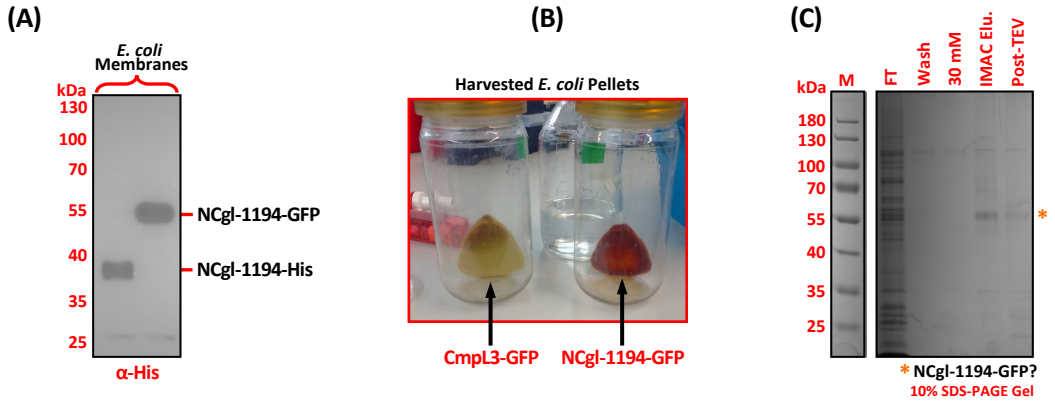


Figure B.1: NCgl-1194 Expression, Phenotype & Attempted Purification

(A) Anti-His western blot showing NCgl-1194-TEV-His_8 and NCgl-1194-TEV-GFP-His_8 in *E. coli* membranes, expressed from pET-11C-TEV-His_8 and pWaldo-TEV-GFP-His_8 respectively. **(B)** Colour comparison of harvested *E. coli* pellets expressing CmpL3 and NCgl-1194 GFP-fusions. **(C)** SDS-PAGE gel of NCgl-1194-TEV-GFP-His_8 purification in DDM, from IMAC flow-through (FT) to after attempted TEV protease cleavage (Post-TEV).

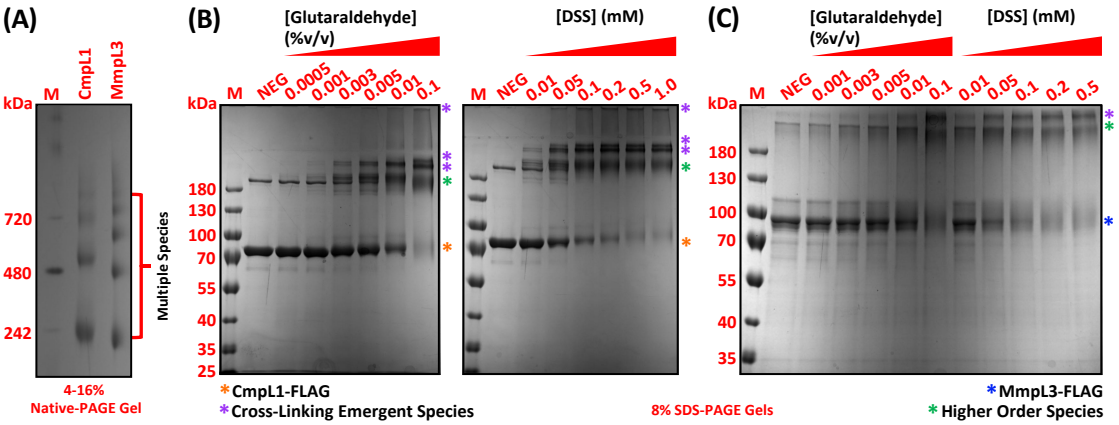


Figure B.2: Native-PAGE & Cross-linking of MmpL3/CmpL1-FLAG in DDM

(A) Novex Bis-Tris BN-PAGE gel of CmpL1-FLAG and MmpL3-FLAG peak fractions. **(B)** SDS-PAGE gels of CmpL1-FLAG cross-linking with glutaraldehyde and DSS at the shown concentrations. Negative controls (NEG) lack cross-linker. **(C)** Equivalent SDS-PAGE gel of MmpL3-FLAG cross-linking. Notable species are marked with asterisks.

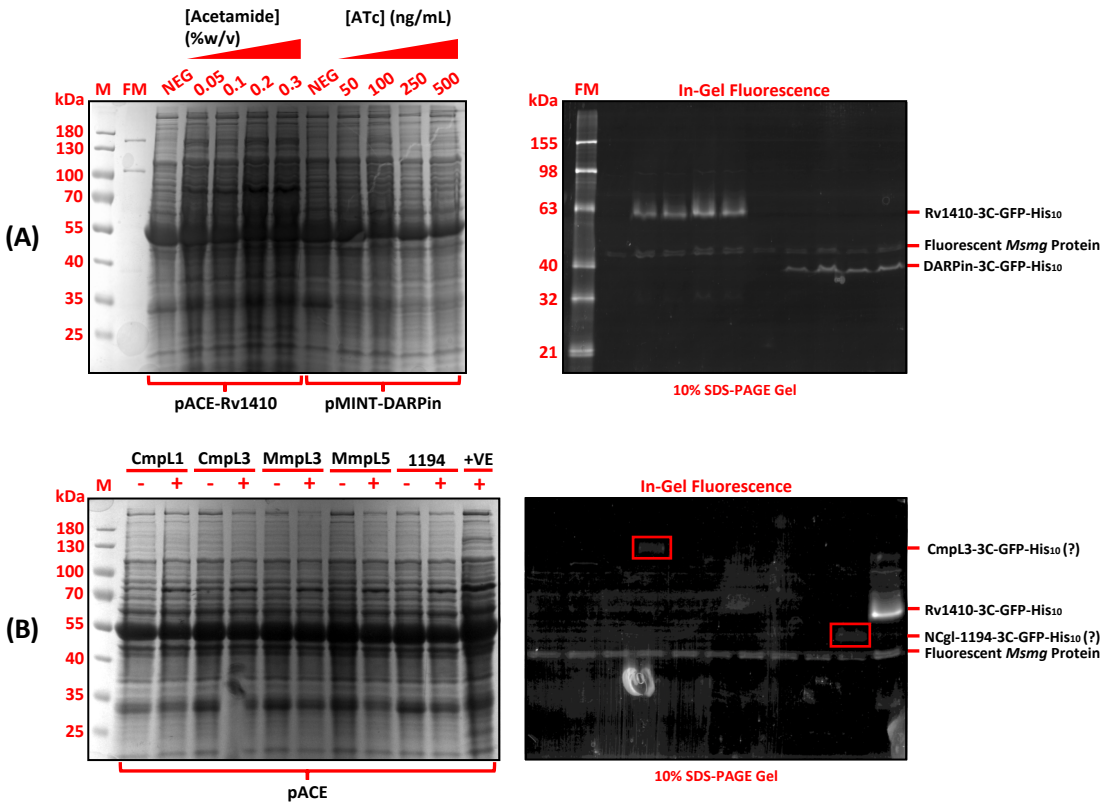


Figure B.3: *M. smegmatis* Expression Trial Summary

(A) SDS-PAGE gel and GFP in-gel of lysate from *Msmg* mc^{2}155 cells induced to express from two positive control vectors — pACE-Rv1410-3C-GFP-His$_{10}$ (an *Mtb* MFS transporter) and pMINT-DARPin-3C-GFP-His$_{10}$ (a synthetic binder). The former is acetamide, and the latter anhydrotetracycline (ATc), inducible. An autofluorescent host protein is also seen. Vectors courtesy of Markus Seeger. **(B)** Equivalent SDS-PAGE and in-gel presenting *Msmg* expression trials with CmpL1, CmpL3, MmpL3, MmpL5 & NCgl-1194 cloned into pACE-3C-GFP-His$_{10}$ and induced with 0.4% (w/v) acetamide (+). Negative controls lack inducer (–), postive control (+VE) is again Rv1410. Red boxes highlight putative expression of CmpL3 & NCgl-1194 at marginal levels.

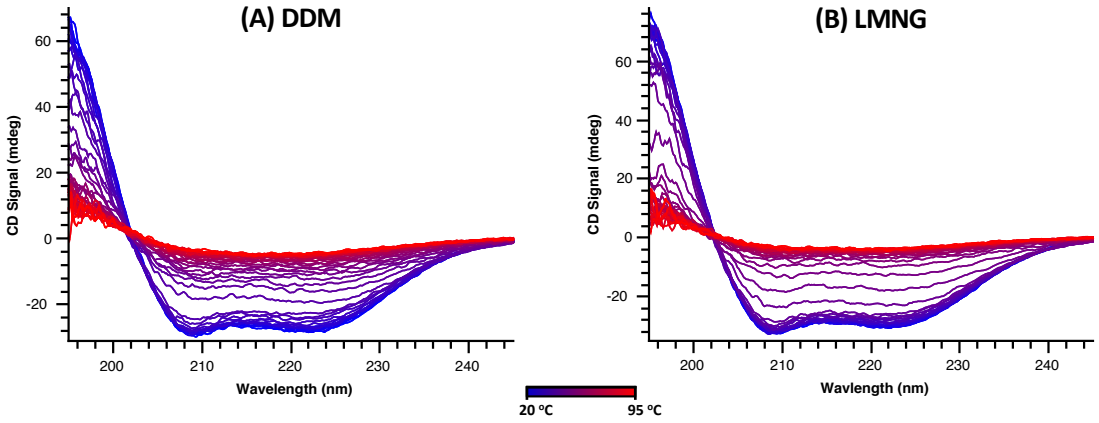


Figure B.4: MmpL3$_{1\text{-}753}$ Full CD Melt Spectra in DDM & LMNG

(A) Full CD scans of DDM solubilised MmpL3$_{1\text{-}753}$ between 20–95 °C, taken in 2 °C increments. Equivalent spectra for LMNG purified truncation are shown in **(B)**.

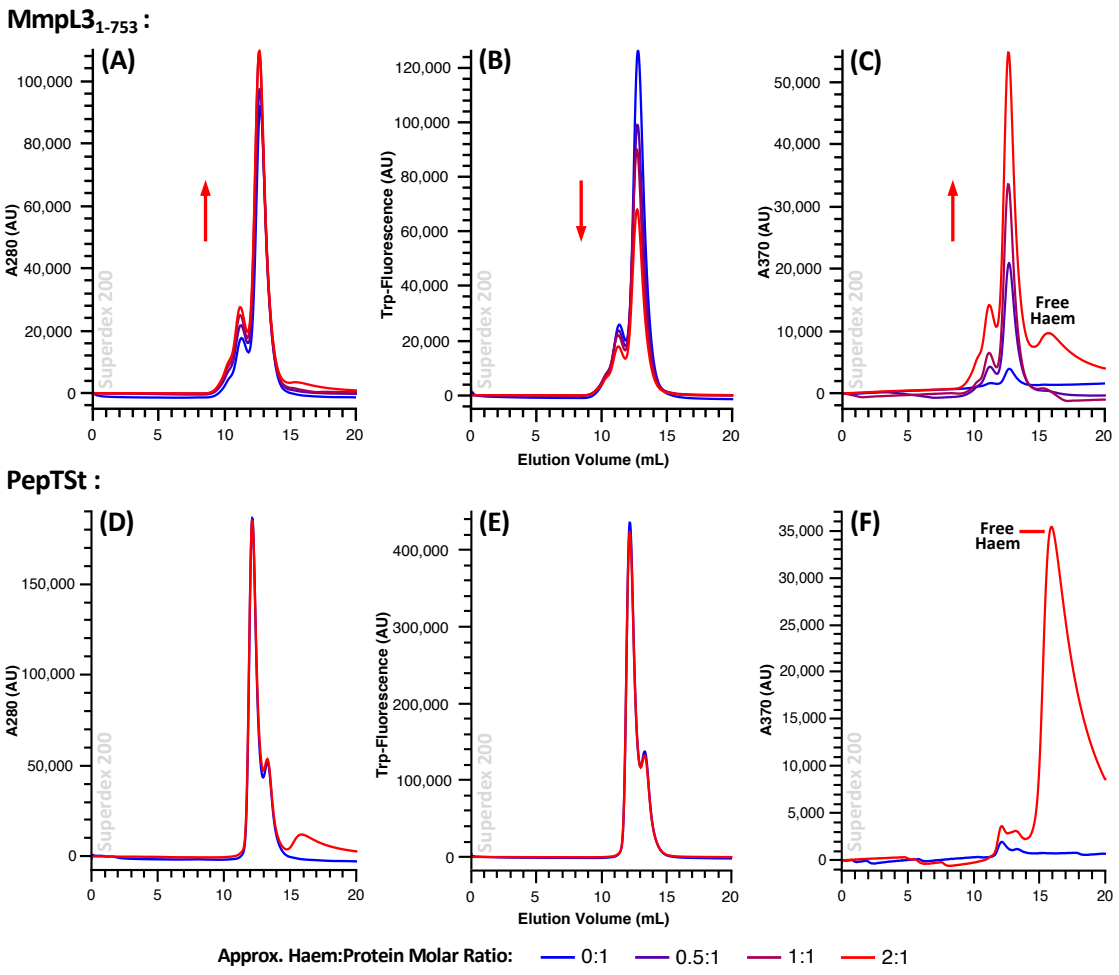


Figure B.5: MmpL3$_{1\text{-}753}$ SEC Haem Association Assay

~120 μg DM solubilised MmpL3$_{1\text{-}753}$ was incubated with "haem" (prepared from hemin as in Owens et al. (2013)) on ice for 2 hours, at various haem:protein molar ratios, prior to gel filtration on an S200 column. Elution **(A)** absorbance 280 nm (A280), **(B)** tryptophan (Trp) fluorescence (λ_{Ex}: 280 nm, λ_{Em}: 348 nm), and **(C)** absorbance 370 nm (A370) recorded by in-line detectors are shown. The latter tracks haem. Red arrows indicate trends with increasing haem:protein molar ratio. Greater MmpL3$_{1\text{-}753}$ coincident A370 and Trp-fluorescence quenching at higher haem-to-protein molar ratios is consistent with haem association. **(D)**–**(F)** show the equivalent control experiment traces obtained using ~100 μg DM purified PepTSt per run (courtesy of Richard Payne). Analogous protein coincident A370 and Trp-fluorescence quenching increases are absent, indicative of no significant haem binding.

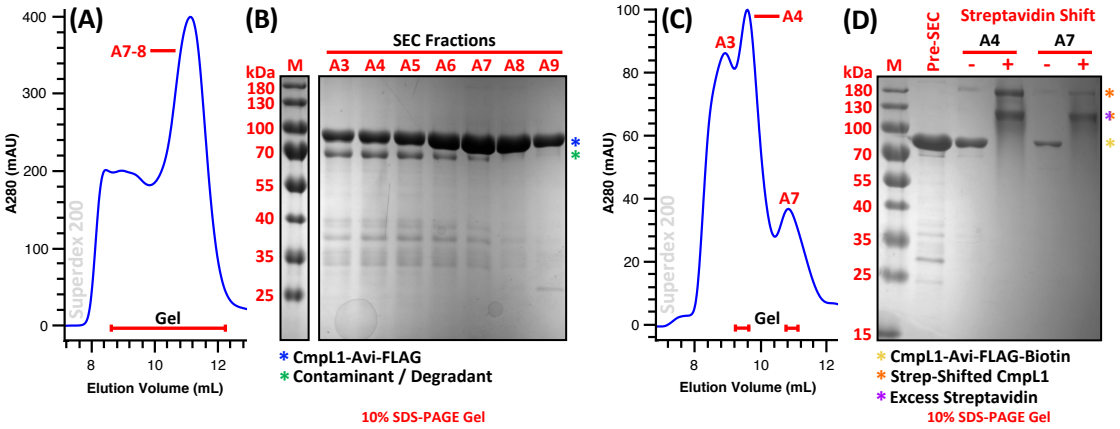


Figure B.6: CmpL1-Avi-FLAG DDM Purification & Biotinylation

(A) S200 trace of concentrated CmpL1-Avi-FLAG purified in DDM. Red bar denotes protein visualised on the SDS-PAGE gel of **(B)**. Peak fractions (A7–10) were pooled, concentrated, BirA-biotinylated overnight, and then re-injected to give the S200 profile displayed in **(C)**; the red bars here corresponding to material tested in the streptavidin shift assay presented in **(D)**. Adjacent lanes have sample incubated with (+) or without (–) 3 × molar excess streptavidin prior to SDS-PAGE. The heterogeneous post-biotinylation chromatogram, leading from the void, is congruent with biotin conjugation destabilising CmpL1-Avi-FLAG, but could also be consequence of target intrinsic instability. Asterisks identify notable species.

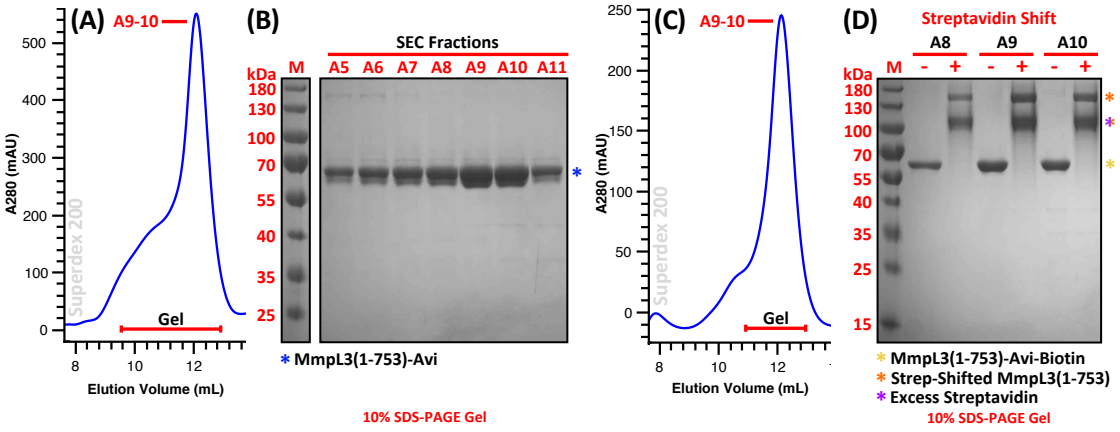


Figure B.7: $MmpL3_{1\text{-}753}$-Avi-TEV-GFP-His_8 LMNG Purification & Biotinylation

(A) S200 profile of concentrated $MmpL3_{1\text{-}753}$-Avi, post-cleavage and reverse IMAC, purified in LMNG. Red bar indicates protein analysed on the SDS-PAGE gel of **(B)**. Peak fractions (A9–10) were pooled, concentrated, BirA-biotinylated overnight, and then re-injected to give the S200 trace depicted in **(C)**; the red bar here denoting material subject to the streptavidin shift assay reported in **(D)**, performed as described in Supplementary Figure B.6[D].

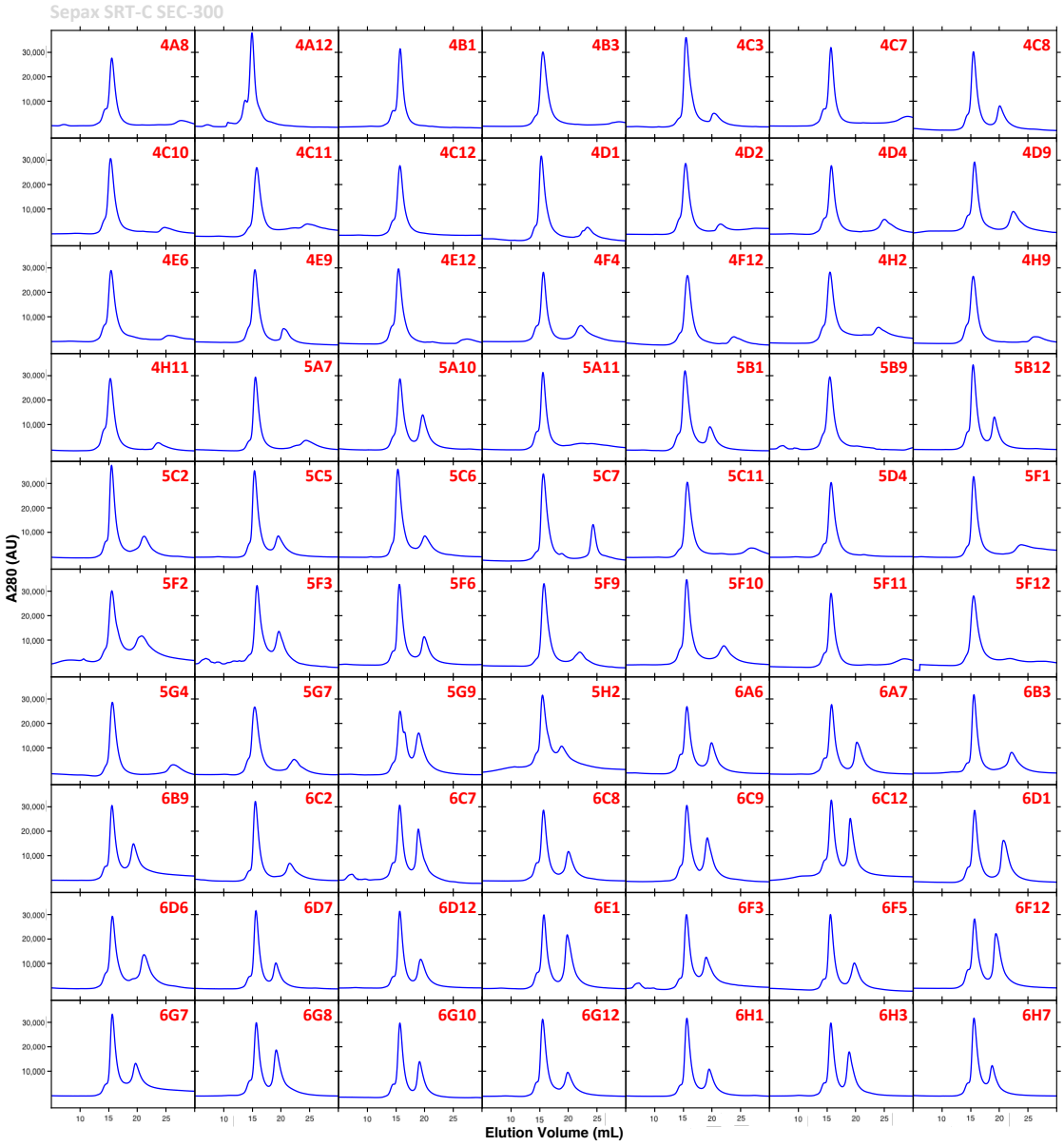


Figure B.8: MmpL3$_{1\text{-}753}$-Sybody SEC Co-Elution Traces

Shared axis plot of 70 MmpL3$_{1\text{-}753}$-sybody co-elution profiles, attained by successive injection onto a Sepax SRT-C SEC-300 column and labelled by sybody alphanumeric designator. In each, the earlier eluting species is either the truncation-binder complex or, if SEC-unstable, MmpL3$_{1\text{-}753}$ alone. Composition of this peak is analysed on the SDS-PAGE gels of Figure 5.10, excluding 5F10 due to sample loss. The second peak, where present, corresponds to free sybody.

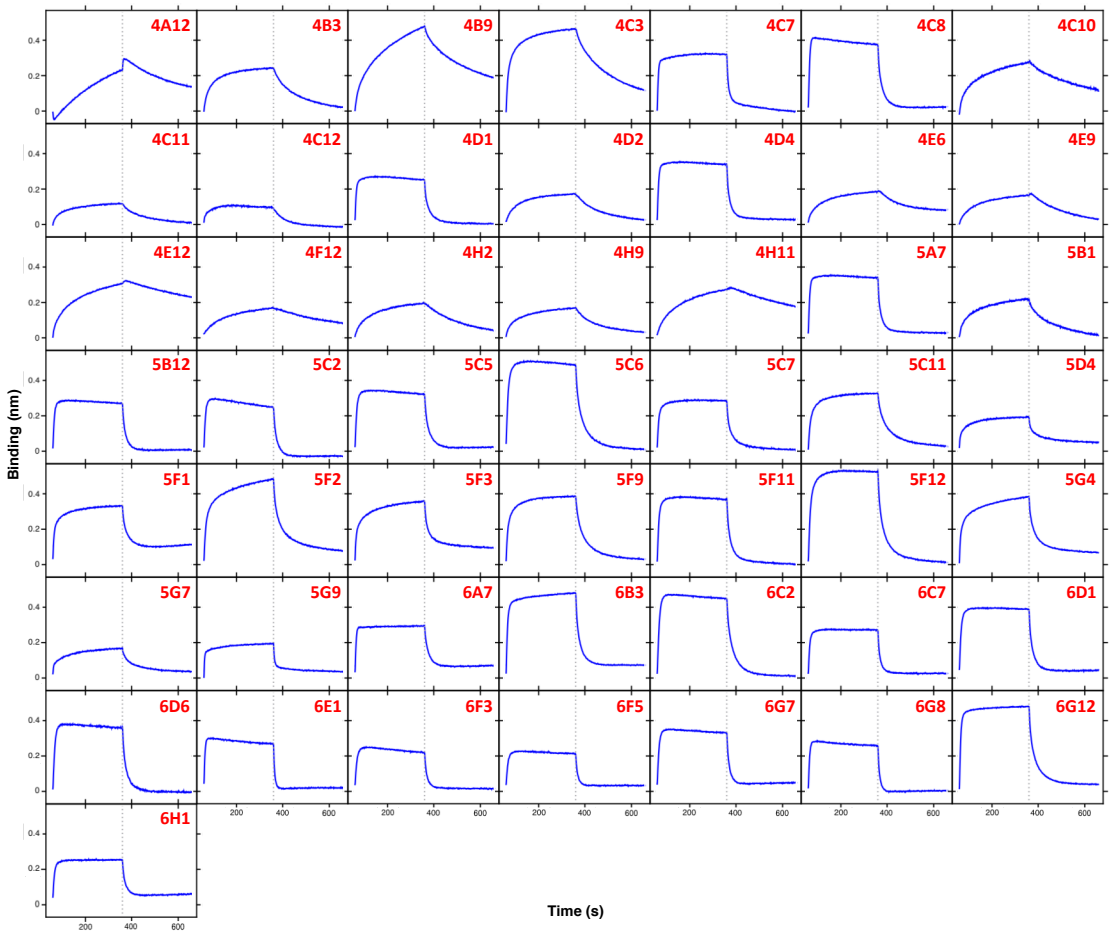


Figure B.9: **MmpL3$_{1\text{-}753}$-Sybody BLI Sensorgrams**

Shared axis plot of 50 Octet-RED384 BLI sensorgrams annotated by sybody alphanumeric identifier. Each charts interaction of the named binder with biotinylated MmpL3$_{1\text{-}753}$-Avi, the latter having been pre-immobilised onto a streptavidin-functionalised dipstick biosensor. The single dotted line demarcates the association and dissociation phases, each 300 seconds long.

Table B.1: Summary of Sybody Screening Results

Library	Sybody	Batch Purification	SEC Co-Elution	Estimated k_{off} (s^{-1})
Small (Concave)	4A7	✓	+	
	4A8	✓	✗	
	4A12	✓	++	3.7×10^{-3}
	4B1	✓	+	
	4B3	✓	+++	1.1×10^{-2}
	4B9	✓	+++	7.0×10^{-3}
	4B11	✓		
	4C3	✓	+++	8.0×10^{-3}
	4C7	✓	++	6.5×10^{-2}
	4C8	✓	++	5.3×10^{-2}
	4C10	✓	+++	5.5×10^{-3}
	4C11	✓	+++	1.1×10^{-2}
	4C12	✓	++	1.8×10^{-2}
	4D1	✓	+++	4.5×10^{-2}
	4D2	✓	+++	9.7×10^{-3}
	4D4	✓	++	6.8×10^{-2}
	4D7	✓		
	4D9	✓	+	
	4E5	✓		
	4E6	✓	+++	1.2×10^{-2}
	4E9	✓	+++	6.1×10^{-3}
	4E12	✓	+++	1.2×10^{-3}
	4F1	✓		
	4F4	✓	✗	
	4F9	✗		
	4F12	✓	+++	3.5×10^{-3}
	4G1	✓		
	4G6	✓		
	4H2	✓	+++	6.3×10^{-3}
	4H4	✓		
	4H7	✓		
	4H9	✓	+++	1.2×10^{-2}
	4H11	✓	+++	1.4×10^{-3}
Medium (Loop)	5A7	✓	++	5.5×10^{-2}
	5A10	✓	+	
	5A11	✓	+	
	5B1	✓	+++	9.6×10^{-3}
	5B7	✗		
	5B12	✓	++	7.0×10^{-2}
	5C2	✓	+++	6.7×10^{-2}

Medium (Loop)	5C5	✓	✢✢	5.2×10^{-2}
	5C6	✓	✢✢	3.0×10^{-2}
	5C7	✓	✢✢✢	4.0×10^{-2}
	5C11	✓	✢✢	2.1×10^{-2}
	5D4	✓	✢✢	2.3×10^{-2}
	5E11	✗		
	5F1	✓	✢✢✢	5.1×10^{-2}
	5F2	✓	✢✢✢	2.5×10^{-2}
	5F3	✓	✢✢	4.1×10^{-2}
	5F6	✓	✢	
	5F9	✓	✢✢✢	2.5×10^{-2}
	5F10	✓		
	5F11	✓	✢✢	5.1×10^{-2}
	5F12	✓	✢✢✢	3.4×10^{-2}
	5G4	✓	✢✢✢	3.2×10^{-2}
	5G5	✗		
	5G7	✓	✢✢✢	2.0×10^{-2}
	5G8	✗		
	5G9	✓	✢✢	1.1×10^{-1}
	5G11	✗		
	5H2	✓	✢	
Large (Convex)	6A6	✓	✗	
	6A7	✓	✢	5.9×10^{-2}
	6B3	✓	✢	5.5×10^{-2}
	6B9	✓	✢	
	6C2	✓	✢✢	3.5×10^{-2}
	6C7	✓	✢	9.2×10^{-2}
	6C8	✓	✗	
	6C9	✓	✢✢	
	6C12	✓	✗	
	6D1	✓	✢	5.9×10^{-2}
	6D3	✗		
	6D6	✓	✢✢	5.4×10^{-2}
	6D7	✓	✗	
	6D12	✓	✗	
	6E1	✓	✢	2.0×10^{-1}
	6E5	✓	✗	
	6E6	✗		
	6E9	✓	✗	
	6F2	✓	✢	
	6F3	✓	✢	8.8×10^{-2}
	6F5	✓	✢	1.1×10^{-1}
	6F12	✓	✢	

Large (Convex)	6G7	✓	✢	9.2×10^{-2}
	6G8	✓	✢	1.3×10^{-1}
	6G10	✓	✗	
	6G12	✓	✢✢	3.5×10^{-2}
	6H1	✓	✢	7.9×10^{-2}
	6H2	✓	✗	
	6H3	✓	✗	
	6H7	✓	✢	

Tabulated summary of the batch purification status (Figure 5.8), SEC co-elution behaviour (Figure 5.10), and BLI estimated off-rate (Figure 5.11) for all 91 sequence unique sybodies selected against $MmpL3_{1\text{-}753}$ in DDM. Empty cells indicate experiments not undertaken. Crosses highlight binders failing to purify or co-elute, after which the candidate was abandoned. The qualitative co-elution scores "present", "good", and "robust" are denoted by ✢, ✢✢ & ✢✢✢ respectively. Off-rates are given to two significant figures. Associated k_{off} errors, although not tabulated, are at least ~1-2 orders of magnitude smaller than the stated value in all cases.

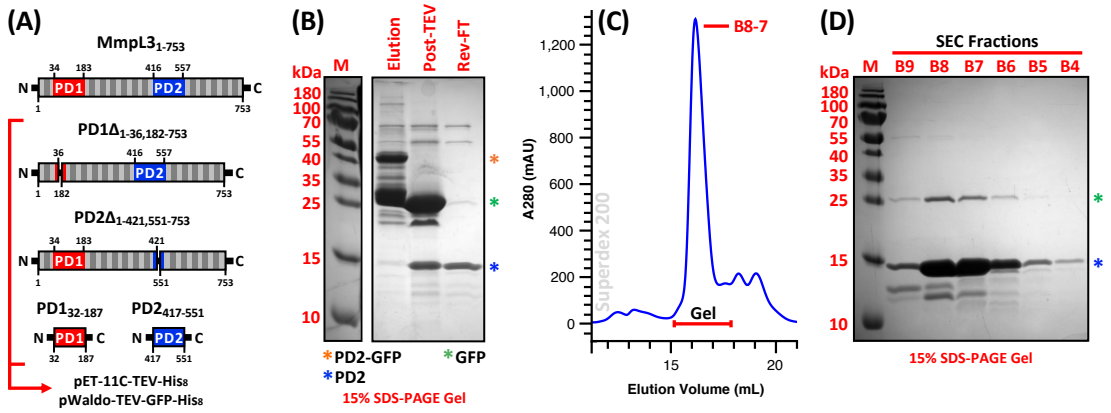


Figure B.10: Epitope Binning Construct Design & Periplasmic Domain Purification
(A) Schematic of the four MmpL3 constructs designed for epitope binning of the sybody cohort, cloned into pET-11C-TEV-His_8 and/or pWaldo-TEV-GFP-His_8. As reference, the domain organisation of $MmpL3_{1\text{-}753}$ is provided. Dark grey rectangles signify TMs, and coloured boxes the first (PD1, red) and second (PD2, blue) periplasmic regions. Numeric annotations define residues at domain boundaries. The constructs either individually delete (PDΔ1 & PDΔ2), or subclone in isolation, each periplasmic unit. Only $PD2_{417\text{-}551}$ expressed convincingly. **(B)** SDS-PAGE gel summarising $PD2_{417\text{-}551}$-TEV-GFP-His_8 purification. From LTR, lanes visualise nickel resin eluate, results of overnight TEV cleavage, and reverse IMAC flow-through (Rev-FT). **(C)** S200 SEC trace of concentrated, tag-free, PD2. Red bar corresponds to material loaded on the SDS-PAGE gel of **(D)**. Note apparent contamination of PD2 peak fractions with co-eluting GFP-His_8.

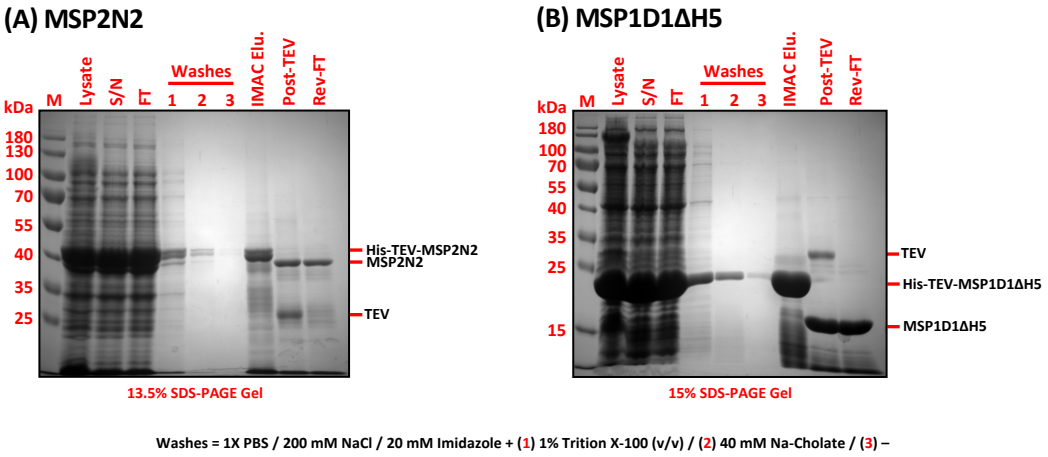


Figure B.11: Purification of MSP2N2 and MSP1D1ΔH5

(A) SDS-PAGE gel summarising His_6-TEV-MSP2N2 purification from BL21-Gold *E. coli*; samples, from LTR, span lysis (by sonication), centrifugal clarification, IMAC, overnight dialysis with TEV protease cleavage, and reverse IMAC flow-through (Rev-FT). The latter was concentrated to ~500 μM, aliquoted, and flash frozen for later use. Wash buffer composition is noted. Triton X-100 is included to remove any endogenous lipid pulled along with the MSPs. **(B)** Equivalent gel detailing His_6-TEV-MSP1D1ΔH5 purification.

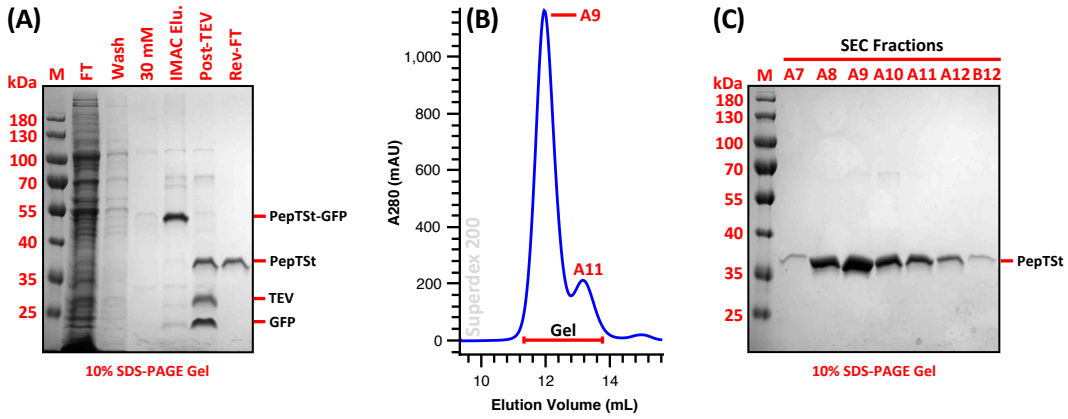


Figure B.12: Purification of PepTSt-TEV-GFP-His_8 in DDM/DM

(A) SDS-PAGE gel of PepTSt-TEV-GFP-His_8 purification in DDM from nickel resin flow-through (FT) to reverse IMAC flow-through (Rev-FT). **(B)** S200 SEC profile of concentrated, tag-free, PepTSt exchanged into DM on-column. The bimodal trace highlights partial separation of the dominant dimer (fraction "A9") from monomeric PepTSt (fraction "A11"). Red bar denotes fractions loaded on the SDS-PAGE gel of **(C)**.

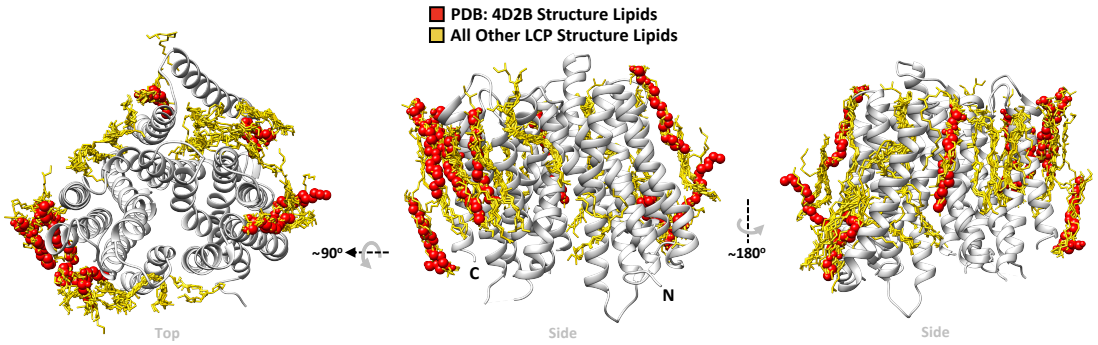


Figure B.13: Summary of Lipids Resolved in PepTSt LCP Structures

Cartoon representation of PepTSt (grey) from the original *in meso* structure (PDB: 4D2B, Lyons et al. (2014)) with resolved 7.8 MAGs shown as spheres (red). 7.8 or 9.7 MAG host lipids co-crystallised in the other 18 PepTSt LCP structures are depicted as sticks (yellow); collated from PDBs: 4D2C, 4D2D (Lyons et al., 2014), 4XNI, 4XNJ (Huang et al., 2015), 5D58, 5D59 (Huang et al., 2016), 5D6K (Ma et al., 2017), 5OXK, 5OXL, 5OXM, 5OXN, 5OXO, 5OXP, 5OXQ, 6EIA (Martinez Molledo et al., 2018a), 6FMR, 6FMY (Huang et al., 2018), & 6GHJ (Martinez Molledo et al., 2018b). Almost all MAG association sites are occupied by molecules from multiple depositions, suggesting the PepTSt exterior displays a non-uniform affinity for annular lipids. Composed with UCSF Chimera v.1.13.1 (Pettersen et al., 2004)

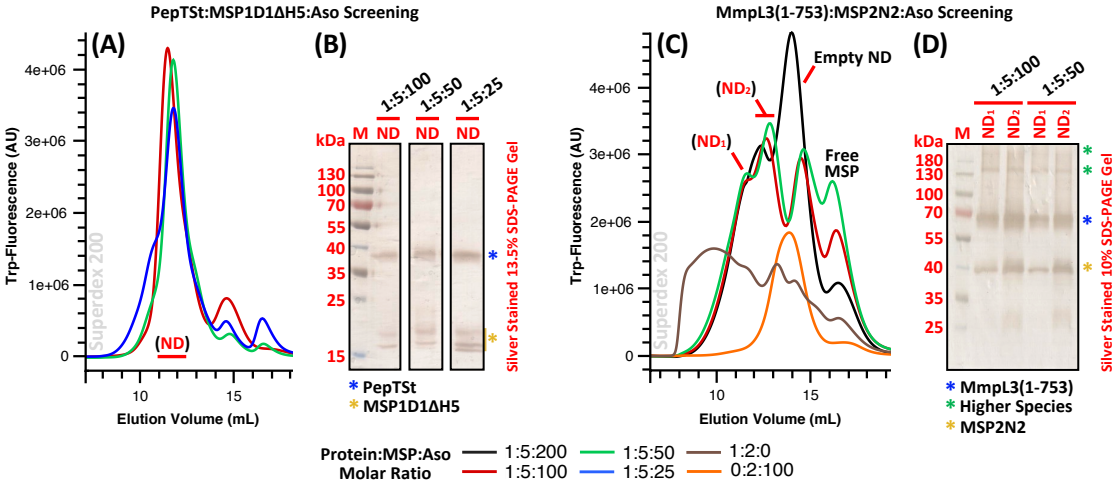


Figure B.14: Additional MSP-Asolectin Nanodisc Reconstitution Screening

(A) Overlain S200 traces screening PepTSt insertion into MSP1D1ΔH5-asolectin NDs at the stated protein:MSP:lipid molar ratios. Peak fractions are visualised on the silver stained SDS-PAGE gel of **(B)**. Equivalent chromatograms produced trialling incorporation of DDM purified $MmpL3_{1\text{-}753}$ into MSP2N2-asolectin NDs are shown in **(C)**, accompanied by gel **(D)**. Annotations denote peaks assigned as target-containing discs (ND_1 & ND_2), empty NDs, or free MSP.

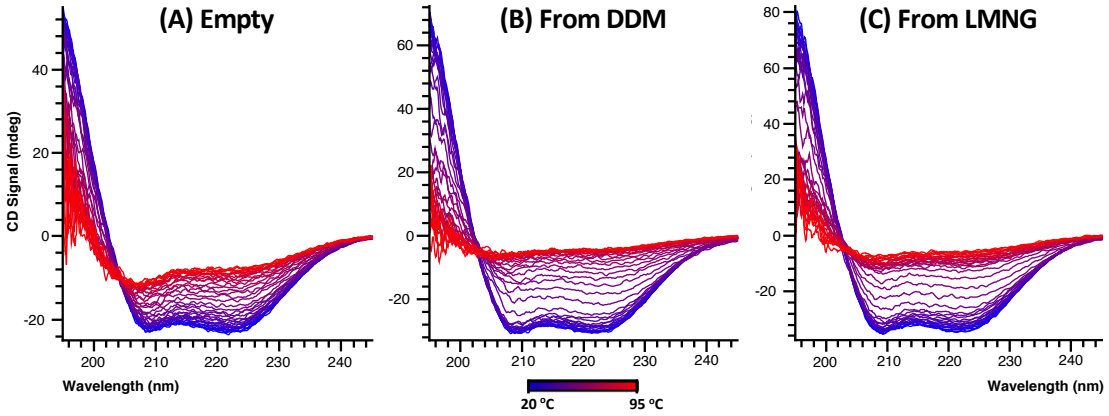


Figure B.15: Full CD Melt Spectra of MSP1D1ΔH5-Asolectin Nanodiscs

(A) Full CD scans of empty MSP1D1ΔH5-asolectin NDs between 20–95 °C, collected in 2 °C increments. Analogous spectra for $MmpL3_{1\text{-}753}$-MSP1D1ΔH5-asolectin discs reconstituted from DDM and LMNG purified truncation are presented in **(B)** & **(C)** respectively.

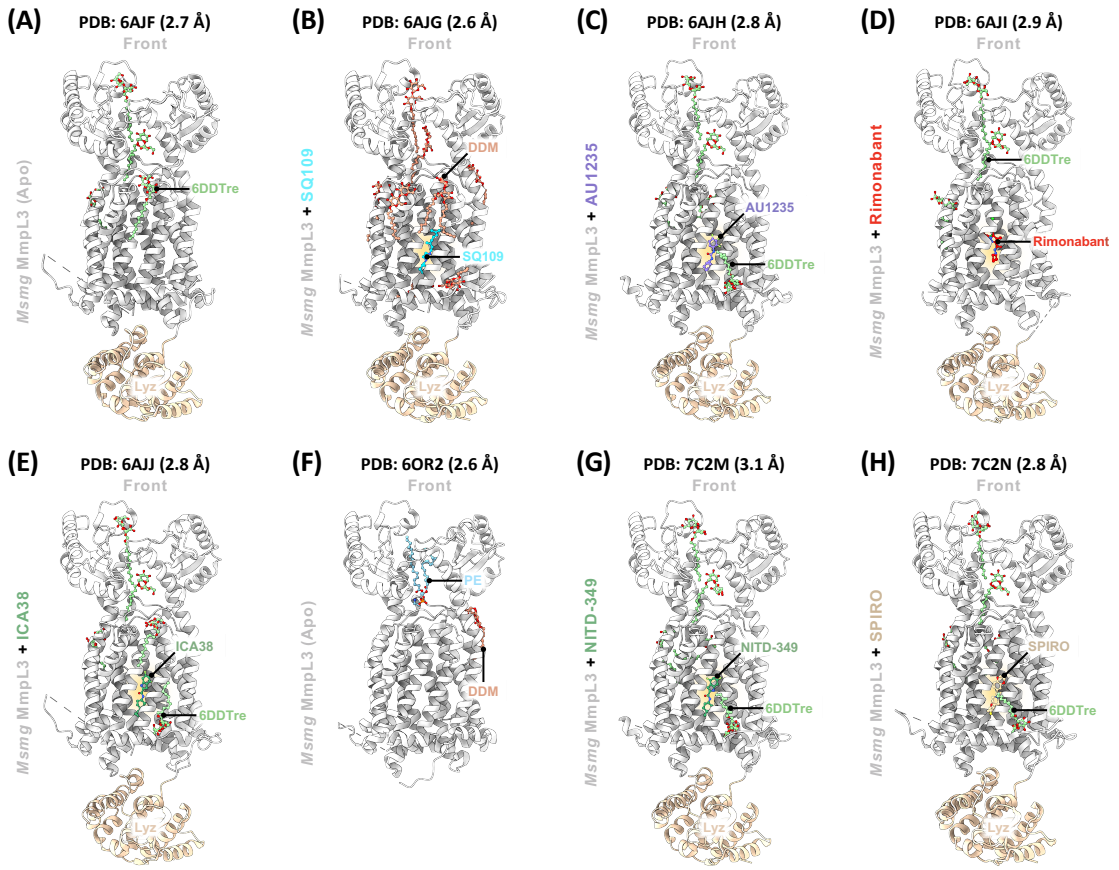


Figure B.16: Overview of the *Msmg* MmpL3 Crystal Structures

(A) – (H) The eight *in surfo* crystal structures of *Msmg* MmpL3 reported at the time of writing; shown in cartoon representation, viewed from the membrane plane, and presented in order of publication (PDBs: 6AJF to 6AJJ, Zhang et al. (2019), 6OR2, Su et al. (2019), 7C2M & 7C2N, Yang et al. (2020)). The resolution of each is stated, and spans 2.6–3.1 Å. All structures are of C-terminal truncations, having been solved from crystals of either *Msmg* MmpL3$_{1\text{-}773}$ (PDB: 6OR2, (F)) or a C-terminal T4 lysozyme (Lyz, Wheat) fusion of *Msmg* MmpL3$_{1\text{-}748}$ (PDBs: 6AJF to 6AJJ, 7C2M & 7C2M;(A) – (E) & (G) – (H)). Dashed lines denote unresolved residues. Ligands are in ball and stick depiction, annotated, and are colour-coded as in Figure 7.5[B] (Detergents & Lipids) and Figure 7.9 (Drugs). Where complexed with a preclinical inhibitor, the universal transmembrane drug pocket is spotlighted by a soft yellow background. 6DDTre - lauryl-6-trehaloside, PE - phosphatidylethanolamine. Structures prepared with UCSF ChimeraX v.1.1 (Pettersen et al., 2021).

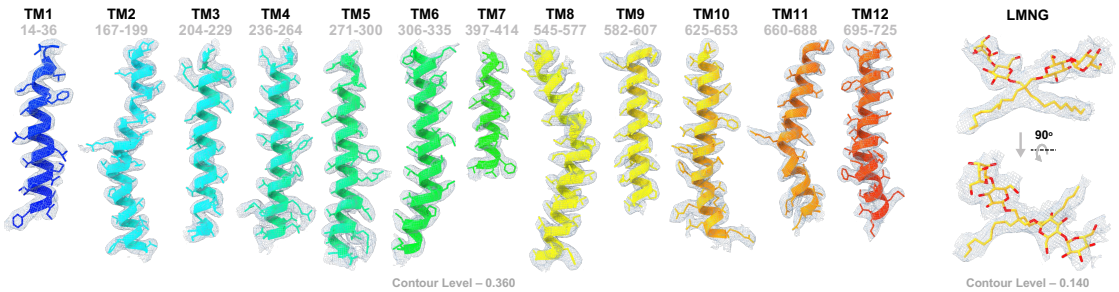


Figure B.17: MmpL3$_{1\text{-}753}$ Model-to-Map Fit

Density fit for the 12 modelled MmpL3$_{1\text{-}753}$ transmembrane helices (TMs) as well as the LMNG molecule occupying the periplasmic vestibule, depicted in cartoon and/or stick representation and coloured as in Figure 7.2[B]. Structures and maps (grey mesh) rendered with UCSF ChimeraX v.1.1 (Pettersen et al., 2021).

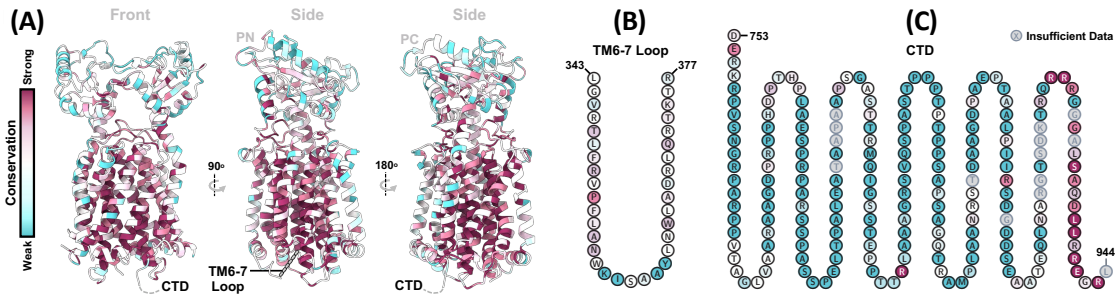


Figure B.18: MmpL3 Conservation Analysis

(A) Cartoon representation of MmpL3$_{1\text{-}753}$ coloured by homologue conservation; calculated using the ConSurf server (Ashkenazy et al., 2016). Locations of the unresolved TM6–7 loop (residues 343–377) and the removed C-terminal domain (CTD, residues 753–944) are annotated. Conservation results for these missing regions are presented in the equivalently colour-coded primary sequence representations of **(B)** & **(C)** respectively. Structures composed with UCSF ChimeraX v.1.1 (Pettersen et al., 2021), and topology depictions with TEXtopo (Beitz, 2000a).

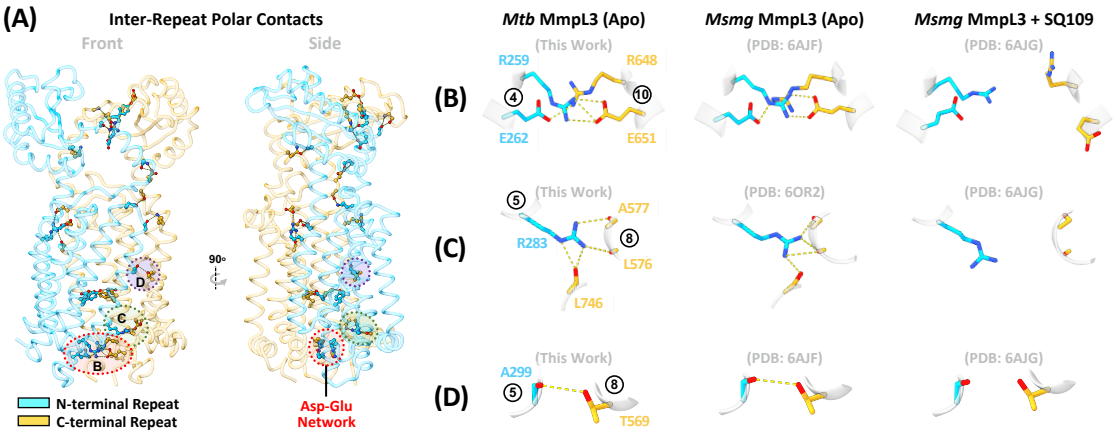


Figure B.19: MmpL3 Inter-Repeat Polar Contact Analysis

(A) Licorice representation of $MmpL3_{1\text{-}753}$ coloured by structural repeat (amino acids 1–342 blue, 378–752 gold). Ball-and-sticks accentuate residues making inter-repeat polar interactions (hydrogen bonds & salt-bridges), visualised as dashed lines. Dotted ovals denote the subset of contacts, other than the Asp-Tyr network presented in Figure 7.7, severed on drug binding. The corresponding matrix of close-ups, **(B)**–**(D)**, compares these interaction sites between the *apo* ($MmpL3_{1\text{-}753}$, PDBs: 6AJF & 6OR2, Zhang et al. (2019); Su et al. (2019)) and SQ109-bound (PDB: 6AJG, Zhang et al. (2019)) MmpL3 structures. TMs are numbered. Figure prepared in UCSF ChimeraX v.1.1 (Pettersen et al., 2021).

Table B.2: Curated MmpL3 Drug-Resistance Mutations

Residue	Mutation(s)	Drugs	Origin Species	Reference(s)
R11	R11P	PIPD1, BMC-2i	*Mab* (R11P)	Dupont et al. (2016) Raynaud et al. (2020)
Q40	Q40H	HC2149, HC2184	*Mtb*	Williams et al. (2019)
	Q40R	SQ109	*Mtb*	Tahlan et al. (2012)
	Q40R	BM212, Compounds 5/8	*Mtb*	Poce et al. (2013)
D46	D46G (+G253E)	BM212, Compounds 5/8	*Mtb*	Poce et al. (2013)
V51	V51A	C215	*Mtb*	Stanley et al. (2012)
S87	S87P (+V681I)	BM212, Compounds 5/8	*Mtb*	Poce et al. (2013)
L189	L189R	A1	*Mtb*	Dal Molin et al. (2019)
	L189R	TBL-140	*Mtb*	Foss et al. (2016)
	L189R	NITD Series	*Mtb*	Rao et al. (2013)
	L189R	BMC-2i	*Mab* (L203R)	Raynaud et al. (2020)
	L189R	THPP, SQ109, AU1235, NITD-304/349	*Mtb**	Li et al. (2019)
	L189A	THPP	*Mtb**	Li et al. (2019)
L191	L191P	BM212	*Msmg* (L196P)	La Rosa et al. (2012)
V192	V192M	BM212	*Msmg* (V197M)	La Rosa et al. (2012)
V195	V195A	EJMCh4	*Mtb*	Korycka-Machała et al. (2019)
P209	P209A	PIPD1	*Mtb*	Dupont et al. (2019)
V210	V210A	TBL-140	*Mtb*	Foss et al. (2016)
L215	L215S	BM212	*Mtb*	La Rosa et al. (2012)
V233	V233G	EJMCh6, BMC-2i	*Mab* (V247G)	Raynaud et al. (2020)
H234	H234D	E1	*Mtb*	Dal Molin et al. (2019)
Y235	Y235L	BM212	*Msmg* (F240L)	La Rosa et al. (2012)
P239	P239H	Compound 22	*Mtb*	Poce et al. (2019)
V240	V240A	HC2184	*Mtb*	Williams et al. (2019)
	V240A	BM212	*Mbov* (V240A)	La Rosa et al. (2012)
	V240A	Compound 6	*Mtb*	Poce et al. (2019)
	V240A	EJMCh6	*Mtb*	Korycka-Machała et al. (2019)
	V240M (+R735C)	BM212, Compounds 5/8	*Mtb*	Poce et al. (2013)
I244	I244T	HC2169	*Mtb*	Williams et al. (2019)
	I244T	BM212	*Msmg* (I249T)	La Rosa et al. (2012)
L246	L246P	PIPD1, EJMCh6, BMC-2i	*Mab* (L260P)	Dupont et al. (2016) Raynaud et al. (2020)
A249	A249P	THPP	*Mtb*	Remuiñán et al. (2013)
	A249V (+L291L)	BM212	*Msmg* (A254V+I296L)	La Rosa et al. (2012)
I250	I250M (+L320P)	BM212	*Mbov* (I250M+L320P)	La Rosa et al. (2012)
Y252	Y252C	IDR-0033216/0334448	*Mtb*	McNeil et al. (2020)
	Y252C	SPIRO	*Mtb*	Remuiñán et al. (2013)
	Y252C	HC2169	*Mtb*	Williams et al. (2019)
	Y252C	AU1235	*Msmg* (Y257C)	McNeil et al. (2017)
	Y252S	IDR-0033216	*Mtb*	McNeil et al. (2020)

Y252	Y252D	PIPD1, AU1235	*Mtb*	Dupont et al. (2019)
G253	G253E	HC2169, HC2184	*Mtb*	Williams et al. (2019)
	G253E	AU1235	*Mtb*	Grzegorzewicz et al. (2012)
	G253E	P1	*Mtb*	Dal Molin et al. (2019)
	G253E	EJMCh4	*Mtb*	Korycka-Machała et al. (2019)
	G253E	NITD Series	*Mtb*	Rao et al. (2013)
	G253E	THPP, SQ109, AU1235, NITD-304/349	*Mtb**	Li et al. (2019)
	G253E (+D46G)	BM212, Compounds 5/8	*Mtb*	Poce et al. (2013)
	G253E (+I516T)	BM212, Compounds 5/8	*Mtb*	Poce et al. (2013)
	G253R	NITD-304/349	*Mtb**	Li et al. (2019)
	G253W	AU1235, NITD-304/349	*Mtb**	Li et al. (2019)
F255	F255L	IDR-0033216/0334448	*Mtb*	McNeil et al. (2020)
	F255L	SPIRO	*Mtb*	Remuiñán et al. (2013)
	F255L	GSK2043267A	*Mbov* (F255L)	Trofimov et al. (2018)
	F255L	CCI7967	*Mtb*	Grover et al. (2021)
	F255L (+V285A)	IDR-0334448, SQ109	*Mtb*	McNeil et al. (2020)
	F255L (+L567P)	IDR-0334448/0541243, SQ109	*Mtb*	McNeil et al. (2020)
	F255L (+V646M)	IDR-0334448/0541243, SQ109	*Mtb*	McNeil et al. (2020)
	F255L (+M649T)	IDR-0334448, SQ109	*Mtb*	McNeil et al. (2020)
	F255L (+M723T)	IDR-0334448	*Mtb*	McNeil et al. (2020)
	F255L (+L567P-F644L)	IDR-0541243, SQ109, AU1235	*Mtb*	McNeil et al. (2020)
	F255L (+L567P-F644I)	IDR-0541243, SQ109, AU1235	*Mtb*	McNeil et al. (2020)
	F255V	CCI7967	*Mtb*	Grover et al. (2021)
	F255V	SPIRO	*Mtb*	Remuiñán et al. (2013)
V257	V257L	PIPD1	*Mtb*	Dupont et al. (2019)
S258	S258G (+L320P)	BM212, Compounds 5/8	*Mbov* (S258G+L320P)	Poce et al. (2013)
T277	T277M (+L320P)	BM212	*Mbov* (T277M+L320P)	La Rosa et al. (2012)
T284	T284A	CCI7967	*Mtb*	Grover et al. (2021)
	T284A	TBL-140	*Mtb*	Foss et al. (2016)
V285	V285A	EJMCh4	*Mtb*	Korycka-Machała et al. (2019)
	V285A	PIPD1	*Mtb*	Dupont et al. (2019)
	V285A	EJMCh6, BMC-2i	*Mab* (V299A)	Raynaud et al. (2020)
	V285A (+F255L)	IDR-0334448, SQ109	*Mtb*	McNeil et al. (2020)
	V285G	HC2169	*Mtb*	Williams et al. (2019)
	V285G	PIPD1	*Mtb*	Dupont et al. (2019)
	V285G	E1	*Mtb*	Dal Molin et al. (2019)
T286	T286K (+L320P)	BM212	*Mbov* (T286K+L320P)	La Rosa et al. (2012)
S288	S288T	Indole-2-carboxamides, SQ109, AU1235	*Mtb*	Lun et al. (2013)
	S288T	Compound 1 & Derivatives	*Mtb*	Li et al. (2020b)
	S288T	SQ109, AU1235	*Msmg* (S293T)	McNeil et al. (2017)
	S288T	SQ109, BM212, AU1235, NITD-304/349	*Mtb**	Li et al. (2019)
	S288A	SQ109, AU1235	*Msmg* (S293A)	McNeil et al. (2017)
V290	V290A	TBL-140	*Mtb*	Foss et al. (2016)
L291	L291L (+A249V)	BM212	*Msmg* (I296L+A254V)	La Rosa et al. (2012)

I292	I292T	SPIRO	*Mtb*	Remuiñán et al. (2013)
	I292T	PIPD1, AU1235	*Mtb*	Dupont et al. (2019)
	I292T	E11-7	*Mbov* (I292T)	Shetty et al. (2018)
	I292S	SPIRO	*Mtb*	Remuiñán et al. (2013)
	I292S	CRS400153/359	*Mtb*	De Groote et al. (2018)
	I292L	Compound 1	*Mtb*	Tantry et al. (2015)
	I292F	AU1235	*Msmg* (I297F)	McNeil et al. (2017)
A294	A294T (+L320P)	BM212	*Mbov* (A294T+L320P)	La Rosa et al. (2012)
S295	S295P	PIPD1	*Mab* (A309P)	Dupont et al. (2016)
	S295P	CRS400153/359	*Mab* (A309P)	De Groote et al. (2018)
	S295P	Compounds 5/25	*Mab* (A309P)	Pandya et al. (2019)
	S295P	Compounds 6/12	*Mab* (A309P)	Kozikowski et al. (2017)
	S295P	EJMCh6, BMC-2i, AU1235, Compound 12	*Mab* (A309P)	Raynaud et al. (2020)
L299	L299Q	HC2149	*Mtb*	Williams et al. (2019)
	L299I (+I631F)	PIPD1	*Mab* (M313I+I616F)	Dupont et al. (2016)
S309	S309C	THPP, SQ109, AU1235, NITD-304/349	*Mtb**	Li et al. (2019)
T311	T311I	HC2149	*Mtb*	Williams et al. (2019)
	T311I	PIPD1	*Mab* (T325I)	Dupont et al. (2016)
	T311I	EJMCh6, BMC-2i, AU1235	*Mab* (T325I)	Raynaud et al. (2020)
	T311I	NITD Series	*Mtb*	Rao et al. (2013)
	T311I	THPP, AU1235	*Mtb**	Li et al. (2019)
A316	A316T (+L320P)	BM212	*Mbov* (A316T+L320P)	La Rosa et al. (2012)
L320	L320P	BM212	*Mbov* (L320P)	La Rosa et al. (2012)
	L320P	Compound 1 & Derivatives	*Mtb*	Li et al. (2020b)
	L320P	C215	*Mtb*	Stanley et al. (2012)
	L320P (+I250M)	BM212	*Mbov* (L320P+I250M)	La Rosa et al. (2012)
	L320P (+S258G)	BM212, Compounds 5/8	*Mbov* (L320P+S258G)	Poce et al. (2013)
	L320P (+T277M)	BM212	*Mbov* (L320P+T277M)	La Rosa et al. (2012)
	L320P (+T286K)	BM212	*Mbov* (L320P+T286K)	La Rosa et al. (2012)
	L320P (+A294T)	BM212	*Mbov* (L320P+A294T)	La Rosa et al. (2012)
	L320P (+A316T)	BM212	*Mbov* (L320P+A316T)	La Rosa et al. (2012)
	L320R (+D466K-I585V)	BM212, Compounds 5/8	*Mbov* (L320R+ E466K-I585V)	Poce et al. (2013)
S321	S321T	BM212	*Msmg* (A326T)	La Rosa et al. (2012)
A342	A342V	BM212	*Msmg* (A347V)	La Rosa et al. (2012)
N365	N365S (+V581A-V681I)	BM212, Compounds 5/8	*Mtb*	Poce et al. (2013)
Q372	Q372R	NITD-304/349	*Mtb**	Li et al. (2019)
R373	R373W	HC2149	*Mtb*	Williams et al. (2019)
M393	M393I	PIPD1	*Mab* (M384I)	Dupont et al. (2016)
P413	P413L	PIPD1, EJMCh6, BMC-2i	*Mab* (P404L)	Dupont et al. (2016) Raynaud et al. (2020)
D466	D466K (+L320R-I585V)	BM212, Compounds 5/8	*Mbov* (E466K+ L320R-I585V)	Poce et al. (2013)
M492	M492T (+V564A-V681I)	BM212, Compounds 5/8	*Mtb*	Poce et al. (2013)
I516	I516T (+G253E)	BM212, Compounds 5/8	*Mtb*	Poce et al. (2013)

M562	M562G	PIPD1, EJMCh6, BMC-2i, AU1235	*Mab* (V547G)	Dupont et al. (2016) Raynaud et al. (2020)
V564	V564A (+M492T-V681I)	BM212, Compounds 5/8	*Mtb*	Poce et al. (2013)
L566	L566S	PIPD1	*Mab* (L551S)	Dupont et al. (2016)
	L566S	CRS400153/359	*Mab* (L551S)	De Groote et al. (2018)
	L566S	EJMCh6, BMC-2i	*Mab* (L551S)	Raynaud et al. (2020)
	L566F	PIPD1	*Mab* (L551F)	Dupont et al. (2016)
	L566F	EJMCh6, BMC-2i, AU1235	*Mab* (L551F)	Raynaud et al. (2020)
L567	L567P	HC2060, HC2091	*Mtb*	Williams et al. (2019) Zheng et al. (2018)
	L567P	SQ109	*Mtb*	Tahlan et al. (2012)
	L567P	PIPD1, AU1235	*Mtb*	Dupont et al. (2019)
	L567P	TBL-140	*Mtb*	Foss et al. (2016)
	L567P	THPP, NITD-304/349, SQ109	*Mtb**	Li et al. (2019)
	L567P (+F255L)	IDR-0334448/0541243, SQ109	*Mtb*	McNeil et al. (2020)
	L567P (+F255L-F644L)	IDR-0541243, SQ109, AU1235	*Mtb*	McNeil et al. (2020)
	L567P (+F255L-F644I)	IDR-0541243, SQ109, AU1235	*Mtb*	McNeil et al. (2020)
V581	V581A (+N365S-V681I)	BM212, Compounds 5/8	*Mtb*	Poce et al. (2013)
I585	I585S	HC2184	*Mtb*	Williams et al. (2019)
	I585V (+L320R-D466K)	BM212, Compounds 5/8	*Mbov* (I585V+ L320R-E466K)	Poce et al. (2013)
S591	S591I	HC2184	*Mtb*	Williams et al. (2019)
	S591I	SPIRO	*Mtb*	Remuiñán et al. (2013)
	S591I	P1, Q1	*Mtb*	Dal Molin et al. (2019)
	S591I	Compound 1 & Derivatives	*Mtb*	Li et al. (2020b)
	S591I	NITD Series	*Mtb*	Rao et al. (2013)
	S591I	CCI7967	*Mtb*	Grover et al. (2021)
	S591I	THPP, SQ109, BM212, AU1235, NITD-304/349	*Mtb**	Li et al. (2019)
	S591T	HC2169	*Mtb*	Williams et al. (2019)
G596	G596R	IDR-0033216	*Mtb*	McNeil et al. (2020)
G628	G628A	PIPD1, EJMCh6, BMC-2i	*Mab* (V613A)	Dupont et al. (2016) Raynaud et al. (2020)
I631	I631F (+L299I)	PIPD1	*Mab* (I616F+M313I)	Dupont et al. (2016)
A632	A632F	PIPD1, EJMCh6, BMC-2i, AU1235	*Mab* (V617F)	Dupont et al. (2016) Raynaud et al. (2020)
F635	F635C	EJMCh6, BMC-2i	*Mab* (Y620C)	Raynaud et al. (2020)
V643	V643M	HC2060, HC2091	*Mtb*	Williams et al. (2019) Zheng et al. (2018)
	V643M	TBL-140	*Mtb*	Foss et al. (2016)
	V643M	E11/11-9	*Mbov* (V643M)	Shetty et al. (2018)
F644	F644L	HC2060, HC2091	*Mtb*	Williams et al. (2019) Zheng et al. (2018)
	F644L	Compound 2	*Mtb*	Ioerger et al. (2013)
	F644L	TBL-140	*Mtb*	Foss et al. (2016)
	F644L (+F255L-L567P)	IDR-0541243, SQ109, AU1235	*Mtb*	McNeil et al. (2020)
	F644C	THPP	*Mtb*	Remuiñán et al. (2013)
	F644C	Compound 2	*Mtb*	Ioerger et al. (2013)

F644	F644C	CCI7967	*Mtb*	Grover et al. (2021)
	F644C	THPP, SQ109	*Mtb**	Li et al. (2019)
	F644N	HC2060	*Mtb*	Williams et al. (2019)
	F644I (+F255L-L567P)	IDR-0541243, SQ109, AU1235	*Mtb*	McNeil et al. (2020)
V646	V646M (+F255L)	IDR-0334448/0541243, SQ109	*Mtb*	McNeil et al. (2020)
M649	M649L	HC2091	*Mtb*	Zheng et al. (2018)
	M649L	PIPD1	*Mtb*	Dupont et al. (2019)
	M649T	HC2149	*Mtb*	Williams et al. (2019)
	M649T (+F225L)	IDR-0334448, SQ109	*Mtb*	McNeil et al. (2020)
A662	A662E	HC2184	*Mtb*	Williams et al. (2019)
T667	T667A	C215	*Mtb*	Stanley et al. (2012)
T670	T670N	HC2091	*Mtb*	Zheng et al. (2018)
	T670I	P1	*Mtb*	Dal Molin et al. (2019)
A677	A677V	THPP	*Mtb*	Remuiñán et al. (2013)
	A677V	TBL-140	*Mtb*	Foss et al. (2016)
	A677V	Compound 2	*Mtb*	Ioerger et al. (2013)
	A677T	CCI7967	*Mtb*	Grover et al. (2021)
A678	A678P	TBL-140	*Mtb*	Foss et al. (2016)
V681	V681I	PIPD1	*Mtb*	Dupont et al. (2019)
	V681I	BM212, Compounds 5/8	*Mtb*	Poce et al. (2013)
	V681I (+S87P)	BM212, Compounds 5/8	*Mtb*	Poce et al. (2013)
	V681I (+N365S-V581A)	BM212, Compounds 5/8	*Mtb*	Poce et al. (2013)
	V681I (+M492T-V564A)	BM212, Compounds 5/8	*Mtb*	Poce et al. (2013)
	V681M	A1	*Mtb*	Dal Molin et al. (2019)
V683	V683G (+V684G)	NITD Series	*Mtb*	Rao et al. (2013)
V684	V684G	BM212	*Msmg* (V689G)	La Rosa et al. (2012)
	V684G	NITD Series	*Mtb*	Rao et al. (2013)
	V684G	A1	*Mtb*	Dal Molin et al. (2019)
	V684G	AU1235, NITD-304/349	*Mtb**	Li et al. (2019)
	V684G (+V683G)	NITD Series	*Mtb*	Rao et al. (2013)
	V684A	C215	*Mtb*	Stanley et al. (2012)
	V684A	Q1	*Mtb*	Dal Molin et al. (2019)
	V684A	EJMCh6	*Mtb*	Korycka-Machała et al. (2019)
	V684A	AU1235, NITD-304/349	*Mtb**	Li et al. (2019)
	V684W	NITD-304/349	*Mtb**	Li et al. (2019)
L693	L693P	HC2184	*Mtb*	Williams et al. (2019)
	L693R	EJMCh6, BMC-2i	*Mab* (L678R)	Raynaud et al. (2020)
M695	M695T	HC2060	*Mtb*	Williams et al. (2019)
L699	L699M	HC2184	*Mtb*	Williams et al. (2019)
	L699M	TBL-140	*Mtb*	Foss et al. (2016)
A700	A700T	HC2091	*Mtb*	Williams et al. (2019) Zheng et al. (2018)
	A700T	SQ109	*Mtb*	Tahlan et al. (2012)
	A700T	TBL-140	*Mtb*	Foss et al. (2016)
M704	M704T	PIPD1	*Mtb*	Dupont et al. (2019)

V713	V713M	THPP	*Mtb*	Remuiñán et al. (2013)
M723	M723T (+F255L)	IDR-0334448	*Mtb*	McNeil et al. (2020)
R735	R735C (+V240M)	BM212, Compounds 5/8	*Mtb*	Poce et al. (2013)

Table collating resistance-conferring MmpL3 mutations extracted from the literature, *in toto* aggregating 112 unique single amino acid substitutions at 83 residues. The preclinical agents affected by each missense variant are listed, along with the mycobacterial species the mutation was identified in, and a reference to the source study. Variants pooled from species other than *Mtb* are back-mapped, with the original mutation(s) listed in brackets. In only a single case did back-mapping produce a synonymous mutation (see L291). For completeness, double and triple mutants are listed reciprocally (i.e. at all involved residues). Drugs are colour-coded by class as in Figure 7.9 (i.e. Pyrroles & Pyrazoles, Ethylenediamines, Indolecarboxamides, Piperidinol Derivatives, Benzimidazoles, Urea Derivatives, Spirocycles, THPP & Derivatives, Benzothioazole Amides, and Miscellaneous Scaffolds). Species are abbreviated as follows: *Mtb* - *M. tuberculosis*, *Msmg* - *M. smegmatis*, *Mbov* - *M. bovis*, and *Mab* - *M. abscessus*. *Mtb** denotes variants identified by reverse genetics studies of *Mtb* MmpL3 expressed in an *Msmg* (Δ*mmpL3*) strain.

Table B.3: Genome-Mined Non-synonymous *Mtb* MmpL3 Substitutions

Mutation	Freq.	Mutation	Freq.	Mutation	Freq.	Mutation	Freq.	Mutation	Freq.
P803R	1979	S591G	15	A869T	7	N390T	4	E439D	2
P867L	1506	G47A	15	A502T	7	L419F	4	A148V	2
D466E	352	I680V	15	E171N	7	Q59E	4	P768L	2
F384I	344	T588V	15	F226L	7	I297V	4	L343P	2
P534S	242	I673L	14	L540Y	6	Q494R	4	G62D	2
G78D	131	A362S	14	V634G	6	I359T	4	E439K	2
M875I	112	K338R	14	A557D	6	L367I	4	D138G	2
A821G	66	Q871H	14	T538E	6	R738K	4	S721F	2
T742N	58	V293I	14	L549Q	6	A804T	4	K442Q	2
H750D	50	D471N	14	L354M	6	A818S	4	G231S	2
E941G	50	R800S	13	A490P	6	A355V	4	K36D	2
R743K	49	A778T	13	T569A	6	L227I	4	E378Q	2
V38I	38	I15M	12	G554S	6	A737M	4	P796S	2
F14Y	38	Q660A	12	I605M	6	T141S	3	I406V	2
P839T	38	G833R	12	S522A	6	Y12C	3	G298S	2
M20T	35	V689M	12	I571T	6	A193T	3	A85P	2
I256V	35	I149T	12	A497S	6	Y498F	3	V759I	2
T272A	34	V90I	11	F635Y	6	P102R	3	A478T	2
L32I	33	G628V	11	V500P	6	G887S	3	I149V	2
E907A	33	L246F	11	G628A	6	V70I	3	Q924K	2
M211I	31	D46E	11	Y498V	6	V392A	3	K135N	2
G26L	30	A676G	11	T526Q	6	V714I	3	E794K	2
C126S	30	S638A	10	A519G	6	P876S	3	A825E	2
M244L	30	K613G	10	K558N	6	T160K	3	S131T	2
I204V	30	L227V	10	I537L	6	K135T	3	V161I	2
R783C	28	L633V	10	I631V	6	S432P	3	L301V	2
Y235F	28	V91I	10	E439H	6	M319T	3	I631F	2
R391W	27	M704A	10	P892S	6	D152G	3	E907K	2
V203A	27	R155K	10	L561A	6	R735S	3	T810A	2
F198L	26	Y127H	10	A123G	5	D905H	3	V581F	2
V681L	26	T458S	9	S835W	5	Q435L	3	V414L	2
R373K	25	L227A	9	T860P	5	R757P	3	R916L	2
Q52K	24	R395L	9	K162E	5	S806P	3	I585T	2
D888E	23	G480S	9	L32M	5	E379A	3	P853L	2
M695L	23	I585V	8	A85T	5	W5L	3	R926C	2
I279M	23	S459T	8	T160A	5	Y61C	3	S33T	2
V850M	23	I469V	8	G158R	5	K416N	3	E794Q	2
M696L	22	I229V	8	I406T	5	K36N	3	S33R	2
A687S	22	W614L	8	A273T	5	V397M	3	A820T	2
A361R	22	S849F	8	D520E	5	P811L	3	L153M	2
T588A	22	P619A	8	K162Q	5	G596E	3	P429L	2
A355R	21	Q167E	8	I531V	5	A825V	3	T430N	2
V210A	21	Y252C	7	S823F	5	T430D	3	G903S	2
S242T	19	R655S	7	V169I	5	T538A	3	R735L	2
L348A	19	D157N	7	I422M	5	S507L	3	G207C	2
L589V	18	F611L	7	K613T	5	G772R	3	A821E	2
V60T	18	A478S	7	R473H	4	R350W	3	P168L	2
I359Y	18	M319I	7	F30Y	4	I264L	3	V539L	2
A220S	18	V563L	7	S597F	4	I359V	3	D152A	2
L189V	17	V572L	7	L25F	4	T141A	3	C332Y	2
R113K	17	F611G	7	M599L	4	V455M	3	G415D	2
A669T	17	I630V	7	N356T	4	T327I	2	A894V	2
L191M	16	S882P	7	G415V	4	L299M	2	A71T	2
T286M	16	A467Q	7	A915V	4	T160M	2	G337S	2

L843V	2	I744L	1	I483S	1	G762E	1	P485T	1
V689A	2	T286R	1	A170D	1	D753P	1	R743C	1
D103N	2	I627V	1	L813M	1	S321A	1	H68N	1
T532A	2	F73S	1	E654A	1	K308H	1	A178E	1
M492T	2	E808G	1	V414A	1	G383E	1	V634I	1
G887V	2	D884N	1	S317T	1	V581I	1	L813P	1
V293A	2	R900W	1	D486A	1	T277M	1	G245S	1
V581L	2	R377S	1	Y146D	1	Q59R	1	T742H	1
A150T	2	F635S	1	T453I	1	G903C	1	S933Y	1
H750R	2	A123E	1	E891K	1	P76Q	1	F129V	1
D904G	2	G207V	1	L191A	1	V340I	1	G542S	1
R881P	2	T286K	1	P413S	1	G26A	1	A228T	1
G415S	2	T284A	1	G421R	1	V650L	1	L243M	1
P151S	2	Q238H	1	P560S	1	L573F	1	P758H	1
E891A	2	I264V	1	T174N	1	S901A	1	M224T	1
C216T	2	V1M	1	P897A	1	V500I	1	R509H	1
V210L	2	V564L	1	E837D	1	F717V	1	P402L	1
R943H	2	L679A	1	V328M	1	L932M	1	P533Q	1
A401T	2	A812T	1	E484K	1	S795L	1	P485Q	1
G536E	2	L576S	1	E378D	1	T270I	1	A470S	1
E495G	2	L307V	1	G334R	1	I256S	1	Q49E	1
A77D	2	A490G	1	P782T	1	L291I	1	I256N	1
H461Y	2	S826N	1	G214V	1	T918M	1	R842H	1
W614V	2	G870S	1	I359W	1	I244V	1	A148T	1
G780E	1	L398V	1	E527D	1	A632V	1	P133Q	1
K394R	1	A75S	1	S852W	1	V683G	1	R757W	1
A546T	1	G26S	1	S325Y	1	E527A	1	M477T	1
A908T	1	Q52H	1	D884E	1	I516V	1	I422V	1
P402S	1	E794!	1	A249T	1	S432L	1	D488K	1
A662S	1	G762R	1	S868C	1	F349L	1	I132V	1
N145D	1	D888G	1	Q40R	1	I250V	1	S868F	1
A765S	1	I250T	1	F482L	1	A894E	1	T179S	1
P767S	1	I297A	1	C332G	1	Y268C	1	I405L	1
Q468P	1	A682S	1	G747R	1	I406G	1	S802R	1
T286V	1	W366C	1	A700T	1	L454Q	1	I411V	1
I229L	1	G344A	1	P518A	1	S826G	1	Q468L	1
D936N	1	V241I	1	M319L	1	R370L	1	G727R	1
L106S	1	L189Q	1	I256T	1	P857S	1	A778V	1
N450D	1	I571M	1	T272V	1	L348M	1	Q435H	1
T568N	1	K338H	1	V21I	1	A499S	1	S360T	1
P807L	1	A592V	1	S612A	1	G636S	1	D888N	1
S865L	1	G421A	1	K756T	1	R391L	1	M562T	1
R755Q	1	S901P	1	R655N	1	C216A	1	V60F	1
V60A	1	A322S	1	G26C	1	R881H	1	G298C	1
G207S	1	A401I	1	K442S	1	R391Q	1	L301F	1
R923P	1	R473C	1	L189R	1	K36M	1	D64G	1
M704V	1	P102H	1	A382H	1	S317A	1	T862P	1
V29I	1	F302L	1	M830I	1	I537M	1	N914T	1
A497T	1	G415F	1	A895V	1	R797Q	1	Q238E	1
P518T	1	S901L	1	T314M	1	M211V	1	A172V	1
M559V	1	A776E	1	G917S	1	K162M	1	R766H	1
V397L	1	V55L	1	M492I	1	I472V	1	L227T	1
F97L	1	A294V	1	E381D	1	A436S	1	L629M	1
P396L	1	E262G	1	A467S	1	I335T	1	A557V	1
I456V	1	R182L	1	L740P	1	I516M	1	M599I	1
A382T	1	I229M	1	R739H	1	V212L	1	S805R	1
T770I	1	G57A	1	V194I	1	T860A	1	R182K	1

Q457L	1	R155W	1	G536R	1	G219S	1	E438D	1
P445L	1	A812V	1	G207F	1	P84Q	1	G207A	1
V345M	1	I264M	1	R672H	1	D466N	1	D904E	1
V161A	1	A121V	1	S855P	1	L555I	1	V274L	1
A815V	1	G337C	1	G6D	1	P767H	1	V285M	1
A557S	1	F717L	1	L23I	1	S760G	1	G78S	1
K126N	1	F444S	1	D369H	1	K425N	1		
A931V	1	T598A	1	H339N	1	L32V	1		
F444L	1	V192A	1	T122I	1	A296G	1		

Table collating 7,475 non-synonymous mutations in *Mtb* MmpL3 distributed over 597 unique substitutions at 439 different residues, presented in order of frequency of occurrence (Freq.). These were mined from 45,878 *Mtb* genomes sourced from the European Nucleotide Archive (ENA), Walker et al. (2015), and The CRyPTIC Consortium & 100000 Genomes Project (2018). Data provided courtesy of Philip Fowler, on behalf of the CRyPTIC Consortium. Red text highlights resistance-conferring variants identified by cross-referencing with Table B.2.

	***Mtb* Lineage**							
Mutation	**Two**	**Three**	**Four**	**Six**	**Bovis**	**Caprae**	**Mixed**	**Unknown**
F14Y	-	-	2	-	-	-	1	33
M20T	-	-	2	-	1	-	1	30
L32I	-	-	1	-	-	-	1	30
V38I	-	-	-	38	-	-	-	-
G78D	-	-	131	-	-	-	-	-
M211I	-	-	6	-	-	-	2	23
I256V	-	-	1	-	-	-	1	32
T272A	-	-	1	-	-	-	1	32
F384I	-	-	-	96	178	9	-	61
D466E	-	-	6	96	178	9	-	63
P534S	-	-	242	-	-	-	-	-
T742N	-	-	-	-	-	-	1	55
R743K	-	-	-	-	-	-	1	46
H750D	-	-	-	-	-	-	1	47
P803R	-	-	1963	-	-	-	13	0
A821G	-	64	-	-	-	-	1	0
P839T	38	-	-	-	-	-	-	-
P867L	-	-	1493	-	-	-	12	-
M875I	112	-	-	-	-	-	-	-
E907A	-	-	33	-	-	-	-	-
E941G	-	-	-	-	-	-	-	50

Table B.4: Lineage Breakdown of Most Abundant *Mtb* MmpL3 Variants

Table detailing any *Mtb* lineage-associations for genome-mined variants occurring >30 times (see Table B.3). Lineage assigned using Mykrobe (Hunt et al., 2019). Data provided courtesy of Philip Fowler, on behalf of the CRyPTIC Consortium.

Appendices

Publications



Publications resulting from the research presented in this thesis:

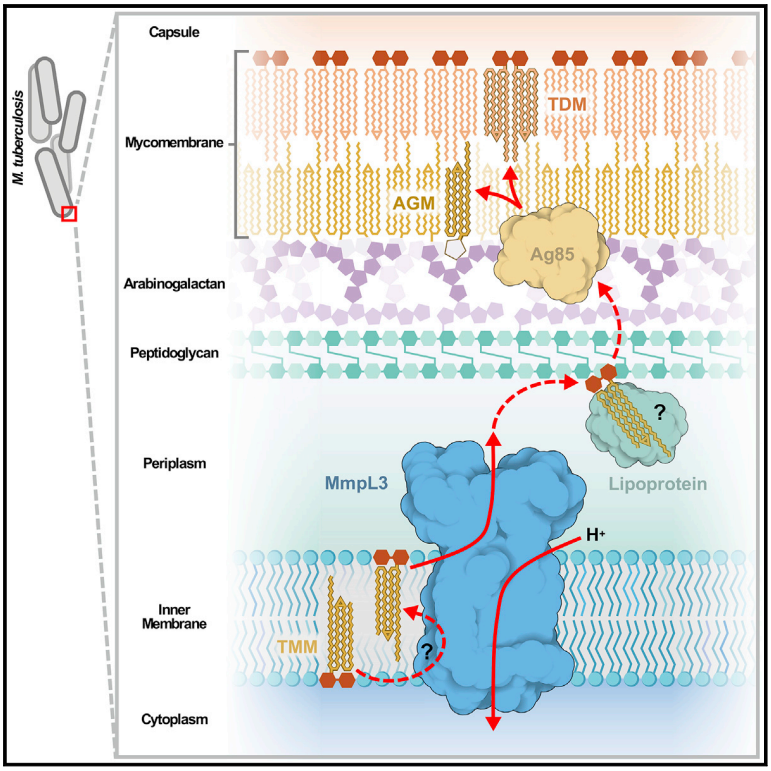
Adams O, Deme JC, Parker JL, the CRyPTIC Consortium, Fowler PW, Lea SM, Newstead S. Cryo-EM structure and resistance landscape of *M. tuberculosis* MmpL3: An emergent therapeutic target. *Structure*. 2021 jul; p. 1–10. , doi:10.1016/j.str.2021.06.013.

Structure

Short Article

Cryo-EM structure and resistance landscape of *M. tuberculosis* MmpL3: An emergent therapeutic target

Graphical abstract



Highlights

- Cryo-EM structure of *M. tuberculosis* MmpL3 determined at 3.0 Å resolution
- An LMNG molecule within the periplasmic cavity suggests the TMM export pathway
- Comprehensive structural mapping of resistance-conferring MmpL3 variants
- Genome-mined MmpL3 mutations indicate minimal pre-existing resistance

Authors

Oliver Adams, Justin C. Deme, Joanne L. Parker, the CRyPTIC Consortium, Philip W. Fowler, Susan M. Lea, Simon Newstead

Correspondence

susan.lea@nih.gov (S.M.L.), simon.newstead@bioch.ox.ac.uk (S.N.)

In brief

Adams et al. report a structure of the *M. tuberculosis* glycolipid transporter MmpL3, a protein essential to envelope biogenesis and an emergent therapeutic target. The structure reveals the probable lipid export pathway, and mapping of resistance-conferring mutations suggests the existence of multiple routes to preclinical inhibitor evasion.

Adams et al., 2021, Structure *29*, 1–10
November 4, 2021 © 2021 The Author(s). Published by Elsevier Ltd.
https://doi.org/10.1016/j.str.2021.06.013

CellPress

Please cite this article in press as: Adams et al., Cryo-EM structure and resistance landscape of *M. tuberculosis* MmpL3: An emergent therapeutic target, Structure (2021), https://doi.org/10.1016/j.str.2021.06.013

Structure

CellPress
OPEN ACCESS

Short Article

Cryo-EM structure and resistance landscape of *M. tuberculosis* MmpL3: An emergent therapeutic target

Oliver Adams,[1] Justin C. Deme,[2,3,4] Joanne L. Parker,[1] the CRyPTIC Consortium,[8] Philip W. Fowler,[5,6] Susan M. Lea,[2,3,4,*] and Simon Newstead[1,7,9,*]

[1]Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK
[2]The Sir William Dunn School of Pathology, University of Oxford, Oxford OX1 3RE, UK
[3]Central Oxford Structural Molecular Imaging Centre (COSMIC), University of Oxford, Oxford OX1 3RE, UK
[4]Center for Structural Biology, Center for Cancer Research, National Cancer Institute, Frederick, MD 21702-1201, USA
[5]Nuffield Department of Medicine, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DU, UK
[6]National Institute of Health Research (NIHR) Oxford Biomedical Research Centre, John Radcliffe, Oxford OX3 9DU, UK
[7]The Kavli Institute for Nanoscience Discovery, University of Oxford, Oxford OX1 3QU, UK
[8]A Corporate Author; for the List of Consortium Members see the section at the end of this article.
[9]Lead contact
*Correspondence: susan.lea@nih.gov (S.M.L.), simon.newstead@bioch.ox.ac.uk (S.N.)
https://doi.org/10.1016/j.str.2021.06.013

SUMMARY

Tuberculosis (TB) is the leading cause of death from a single infectious agent and in 2019 an estimated 10 million people worldwide contracted the disease. Although treatments for TB exist, continual emergence of drug-resistant variants necessitates urgent development of novel antituberculars. An important new target is the lipid transporter MmpL3, which is required for construction of the unique cell envelope that shields *Mycobacterium tuberculosis* (*Mtb*) from the immune system. However, a structural understanding of the mutations in *Mtb* MmpL3 that confer resistance to the many preclinical leads is lacking, hampering efforts to circumvent resistance mechanisms. Here, we present the cryoelectron microscopy structure of *Mtb* MmpL3 and use it to comprehensively analyze the mutational landscape of drug resistance. Our data provide a rational explanation for resistance variants local to the central drug binding site, and also highlight a potential alternative route to resistance operating within the periplasmic domain.

INTRODUCTION

Tuberculosis (TB) remains the leading cause of death attributable to a single pathogen (World Health Organization, 2020). Compounding this is the emergence and endemic nature of several drug-resistant *Mycobacterium tuberculosis* (*Mtb*) strains (Udwadia et al., 2012; Velayati et al., 2009). Mycobacterial membrane protein large 3 (MmpL3) is part of a vital biosynthetic pathway in *Mycobacterium* species and represents a promising new avenue for novel antitubercular drug development (Rayasam, 2014).

MmpL3 is an essential protein belonging to the resistance-nodulation-division (RND) superfamily of transporters (DeJesus et al., 2017; Domenech et al., 2005). Performing diverse functions throughout all domains of life, the RND superfamily were initially characterized as multi-drug efflux pumps in Gram-negative bacteria (Nikaido, 2018). In Gram-positive mycobacteria the MmpL family predominantly function as endogenous lipid transporters, with many, including MmpL3, involved in the biogenesis of the multi-layered cell envelope (Chalut, 2016; Viljoen et al., 2017). Specifically, MmpL3 exports intracellularly synthesized trehalose monomycolate (TMM) into the periplasm (Grzegorzewicz et al., 2012). TMM transport is essential for the formation of the unique *Mycobacterium* cell envelope that functions as the main physical barrier to drug entry and neutralization by the immune system (Dulberger et al., 2020; Jackson, 2014). The driving force for TMM export is thought to come from the proton-motive force (PMF) established across the inner membrane (Grzegorzewicz et al., 2012; Székely and Cole, 2016), although the precise mechanism by which the PMF drives lipid export remains unclear.

Disruption of envelope mycolation, as induced by small-molecule or genetic inhibition of MmpL3, has repeatedly been shown to be bactericidal (Degiacomi et al., 2017; Li et al., 2016; Shao et al., 2020; Varela et al., 2012). Indeed, the list of preclinical MmpL3 inhibitors is ever-growing, and features compounds with diverse chemical scaffolds (Dupont et al., 2016; De Groote et al., 2018; Grzegorzewicz et al., 2012; Lun et al., 2013; Remuiñán et al., 2013; La Rosa et al., 2012; Stanley et al., 2012; Tahlan et al., 2012). Encouragingly, some have been observed to synergize with existing antituberculars (Li et al., 2017; Nikonenko et al., 2007; Stec et al., 2016) and one, the ethylenediamine SQ109, has completed phase II clinical trials in various countries (Butler and Paterson, 2020; Degiacomi et al., 2020).

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Please cite this article in press as: Adams et al., Cryo-EM structure and resistance landscape of *M. tuberculosis* MmpL3: An emergent therapeutic target, Structure (2021), https://doi.org/10.1016/j.str.2021.06.013

Significant insights into the interactions between MmpL3 and its inhibitors were recently reported using the *Mycobacterium smegmatis* (*Msmg*) ortholog (Yang et al., 2020; Zhang et al., 2019), which shares 61% sequence identity to *Mtb*. Nevertheless, given the severity of the TB pandemic, the evident susceptibility of MmpL3 to small-molecule inhibitors, and the numerous medicinal chemistry campaigns against it, understanding the structure and biochemical properties of the *Mtb* transporter remains a high priority.

Here, we report the cryoelectron microscopy (cryo-EM) structure of *Mtb* MmpL3 at a resolution of 3.0 Å. We present the most comprehensive structural mapping of MmpL3 variants to date, analyzing over 100 unique resistance-conferring substitutions selected by exposure to preclinical agents. In addition, we examine non-synonymous mutations mined from >45,000 whole-genome sequenced *Mtb* isolates. In conjunction with the recently reported structures of MmpL3 from *Msmg*, our work provides a foundation for structure-based drug design against *Mtb* MmpL3, guided by an enhanced appreciation of the underlying mutational landscape.

RESULTS AND DISCUSSION

Cryo-EM structure of *Mtb* MmpL3

To obtain the structure of *Mtb* MmpL3 we generated a C-terminally truncated (residues 1–753) construct capable of being overexpressed in *E. coli* (Figure S1A). The intracellular C-terminal domain (CTD) is dispensable for transport, with multiple C-terminal *Mtb* MmpL3 truncations viable *in vivo* (Belardinelli et al., 2016). Despite extensive efforts, $MmpL3_{1\text{-}753}$ proved recalcitrant to crystallization. However, cryo-EM imaging of the protein purified in lauryl maltose neopentyl glycol (LMNG) permitted its three-dimensional reconstruction to a resolution of 3.0 Å (Figures S1B–S1E). Our map reveals $MmpL3_{1\text{-}753}$ to be monomeric (Figure 1A), consistent with the *Msmg* homolog (Su et al., 2019; Zhang et al., 2019) (Figure S2A). The cryo-EM map was of sufficient quality to confidently model almost the entirety (residues 1–342, 378–752) of the construct (Figures 1B and S1F; Table S1). *Mtb* $MmpL3_{1\text{-}753}$ adopts a transmembrane domain (TMD) fold archetypal of RND proteins, consisting of 12 transmembrane helices (TMs) arranged as two sequence-contiguous bundles (TMs 1–6 and 7–12). These are related by a central 2-fold pseudo-symmetry axis running perpendicular to the membrane plane (Figure S2B). The $MmpL3_{1\text{-}753}$ TMD is braced against the inner leaflet of the cytoplasmic membrane via three short, lateral, amphipathic α helices, two preceding TMs 1 and 7, with the third leading into the CTD. Characteristic of RND transporters, sizable periplasmic loops connect TMs 1–2 in the N-terminal half of $MmpL3_{1\text{-}753}$ (PN, residues 37–166) and TMs 7–8 (PC, residues 415–544) in the C-terminal half of the molecule. Both display an α-β-α-β-α-β topology, with the first α helix of each contributing to the tertiary structure of the adjacent loop. This partial interdigitation of secondary structural elements results in PN and PC clasping one another to give a singular periplasmic domain (PD), capping the TMD and sharing its pseudo-symmetry axis. The N termini of TM2 and TM8 are longer than the other helices and extend into the PD (Figure 1C), which itself protrudes over 40 Å into the periplasmic space.

Similar to structures of the *Msmg* MmpL3 homolog (Su et al., 2019; Zhang et al., 2019), the *Mtb* protein has an extensive cavity enclosed by the PD (Figure 1C). The cavity comprises a central vestibule between the PN and PC lobes, out of which branch two descending channels, each constricted in the vicinity of the PD-TMD "neck." On opposing faces of the protein these channels exit into the outer leaflet of the inner membrane via grooves bounded by TMs 1–4 or 7–10, respectively. The upper PD chamber likewise possesses three apertures opening into the periplasm from the front (PD_F), back (PD_B), and top (PD_T) of the molecule and each gated by a combination of charged and hydroxyl-containing side chains (Figure 1D). Although the role of this cavity has not yet been fully elucidated, its position within the molecule and relationship to the outer leaflet suggests it forms the conduit through which TMM is extracted into the periplasm, a role supported by site-directed mutagenesis (Belardinelli et al., 2016). Within this cavity a lipid-like species was captured in our structure, most likely the detergent LMNG, which adopts a splayed arrangement bound between the PN and PC subdomains (Figures 1A, 1B, and 2A). LMNG creates a substantial interface, packing against 35 residues (within 4.5 Å) and sitting within the suggested transport path for TMM (Klenotic et al., 2020). The interactions are predominantly contributed by PN, PC, and TM2, but also feature side chains at the N terminus of TM8 and in the loops connecting TMs 3–4 and 11–12 (Figure 2A). Although the LMNG tails point in divergent directions, one halfway down into the channel exiting via TMs 1–4 and the other up toward PD_T, both are accommodated in pockets partially scaffolded by TM2 aliphatic residues (L166, V169, L173, and I177), and end in aromatic residues (Y235 and F236 or F440 and Y447) provided by nearby secondary structure elements. In contrast, the two LMNG maltose head groups bifurcate to be individually coordinated by residues of the PD_F and PD_B apertures. In doing so their hydroxyls are hydrogen bonded to side chains of D58 and D139, as well as the backbone carbonyl of I422, leaving the terminal saccharide of each head group to be partly solvent exposed as it emerges through its respective PD opening.

A key question in the mechanism of MmpL3 is how this protein handles the large and complex TMM lipid. The previous *Msmg* MmpL3 structures were co-crystallized with three unique PD occupants (Figure 2B); pairs of the detergents lauryl-6-trehaloside (6DDTre, a structural analog of TMM) and dodecyl maltoside (DDM) (Zhang et al., 2019), as well as a single phosphatidylethanolamine (PE) lipid (Su et al., 2019). Structural comparison with the *Mtb* $MmpL3_{1\text{-}753}$-LMNG complex identifies striking similarities in the binding poses of all four ligands (Figure 2C). The pockets embracing the LMNG alkyl chains in our structure are also exploited by 6DDTre and DDM to accommodate their respective hydrophobic tails. PE, in contrast, sits with its shorter heptadecanoate acyl chain in the upper of these two tail sites, leaving its peripheral carbons to extend through PD_T. The comparison of *Mtb* MmpL3 and *Msmg* MmpL3 suggests a "division of labor" model for PD lipid handling, wherein the central vestibule sequesters the alkyl chains away from the bulk periplasm, potentially into specific pockets (Figure 2C), while the proximate hydrophilic openings (PD_F, PD_B, and PD_T) bind the polar head group, coordinating it through hydrogen bonds and providing it a degree of solvent access. Segmentation of binding would permit accommodation of lipidic cargo in, and subsequent

Please cite this article in press as: Adams et al., Cryo-EM structure and resistance landscape of *M. tuberculosis* MmpL3: An emergent therapeutic target, Structure (2021), https://doi.org/10.1016/j.str.2021.06.013

CellPress
OPEN ACCESS

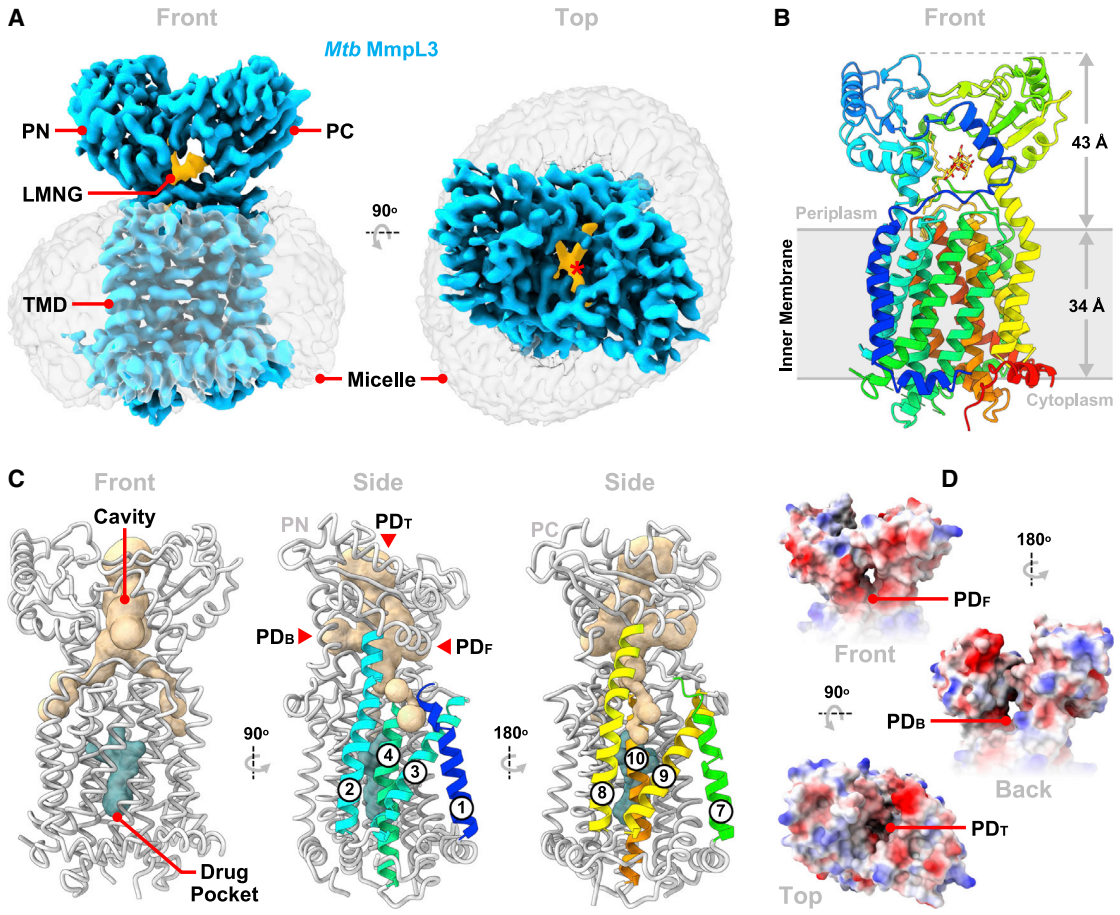


Figure 1. Cryo-EM map and structural analysis of *Mtb* MmpL3

(A) Cryo-EM density of *Mtb* MmpL3$_{1\text{-}753}$ as viewed from the membrane plane (front), or perpendicular to it from the periplasm (top). Proteinaceous density is blue (contour level 0.427, sharpened map), the surrounding detergent micelle gray, and a bound LMNG molecule gold (contour level 0.150, unsharpened map). Annotations highlight the transmembrane domain (TMD), and the N-terminal (PN) and C-terminal (PC) lobes of the periplasmic domain (PD).

(B) Cartoon representation of the *Mtb* MmpL3$_{1\text{-}753}$ atomic model, colored blue (N terminus) to red (C terminus). LMNG is shown as sticks (gold).

(C) Licorice representations of *Mtb* MmpL3$_{1\text{-}753}$ (gray) accompanied by surface depictions emphasizing its PD cavity network (wheat, calculated using Caver Analyst 2.0 [Jurcik et al., 2018]) and TMD drug binding pocket (teal). The TMs (1–4 and 7–10) lining cavity exits into the outer leaflet are indicated.

(D) Electrostatic surface renderings of the MmpL3$_{1\text{-}753}$ PD, highlighting the hydrophilic residues gating the labeled PD$_F$, PD$_B$, and PD$_T$ cavity openings.

See also Figures S1 and S2, and Table S1.

manipulation into, diverse conformations. This flexibility is plausibly essential for translocation of substrates as large as TMM (molecular weight >1.4 kDa, Fujita et al., 2005), and likely underlies the lipid binding promiscuity observed for *Msmg* MmpL3 by native mass spectrometry (Su et al., 2019).

Mapping MmpL3 mutations

MmpL3 is the target of many preclinical antituberculars (Degiacomi et al., 2020). These compounds target a conserved central pocket within the TMD that sits at the interface between TMs 4 and 5 and 10 and 11 (Figures 1C and S4B) and incorporates the Asp-Tyr interaction network (Figure S2C) proposed to function in coupling TMM transport to the PMF (Bernut et al., 2016; Yang et al., 2020; Zhang et al., 2019). Antitubercular agents have been predominantly discovered by high-throughput whole-cell phenotypic screening against sizable chemical libraries; with targets for active compounds inferred by their mutation in whole-genome-sequenced resistant isolates (Goldman, 2013). Consequently, there exists a substantial repository of drug-resistant MmpL3 variants in the literature, which we collated and mapped onto our MmpL3$_{1\text{-}753}$ model. We aggregated 112 unique resistance-conferring mutations occurring at 83 different positions in MmpL3, 21 of which differ between the *Msmg* and *Mtb* homologs (Table S2). Of the 112 mutations, 109 variants at 80 positions were mappable onto our *Mtb* MmpL3$_{1\text{-}753}$ structure (Figure 3A). The remaining positions populated the unresolved inter-repeat linker. Most mutations were described in *Mtb* (57 positions), but to provide a more complete picture of the resistance landscape we also back-mapped variants originally identified in *Msmg* (11 positions), *M. bovis* (13 positions), and *M. abscessus* (19 positions) by sequence alignment (Figure S4A).

Of the 83 resistance positions (RPs) identified, ~90% reside in the TMD and 60% are within 10 Å of the drug pocket (Figure 3A). Such clustering suggests that the dominant route to resistance is

Please cite this article in press as: Adams et al., Cryo-EM structure and resistance landscape of *M. tuberculosis* MmpL3: An emergent therapeutic target, Structure (2021), https://doi.org/10.1016/j.str.2021.06.013

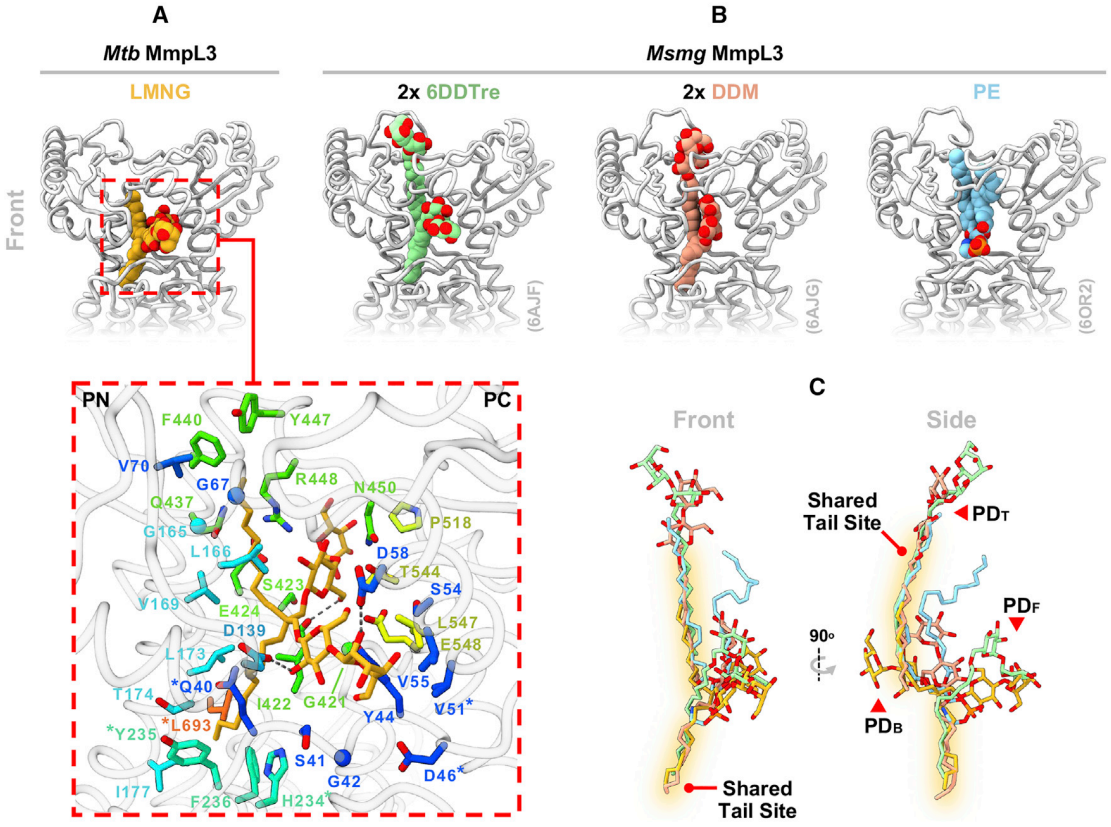


Figure 2. Comparison of MmpL3 periplasmic ligands

(A) Close-up of the *Mtb* $MmpL3_{1-753}$ PD LMNG. Residues within 4.5 Å, plus the detergent itself (gold), are shown as sticks; bar glycines shown as balls. Side chains are rainbow colored, as in Figure 1B. Dashed lines denote hydrogen bonds. Asterisks highlight residues conferring inhibitor resistance when mutated.

(B) Side-by-side comparison of the 6DDTre (green), DDM (salmon), and PE (blue) molecules co-crystallized within the *Msmg* MmpL3 PD. Ligands are shown as spheres and proteins (gray) in licorice representation. PDB codes are in brackets.

(C) Overlay of the *Mtb* and *Msmg* PD ligands, each depicted as sticks and colored as in (A) and (B). Common alkyl chain sites, and the positions of the PD apertures, are annotated.

through mutation of residues defining the 2-fold pseudo-symmetry axis relating the two six-helix bundles of the TMD. These mutations likely weaken the affinity of antitubercular drugs targeting the transmembrane site, as has been demonstrated for a handful of *Msmg* MmpL3 variants by MST (Zhang et al., 2019). This conclusion is also consistent with proposals that the magnitude and spectrum of the resistance phenotype given by an MmpL3 mutant correlates with its proximity to this pocket (Williams et al., 2019). However, our analysis reveals that a significant proportion of RPs are located >10 Å away from the drug pocket, including some within the LMNG binding site (asterisks in Figure 2A), suggesting that there is at least one additional, and as yet uncharacterized, MmpL3 resistance mechanism; likely involving the PD and its cavity. Future studies exploring the role of the PD within the mechanism of MmpL3 will be required to uncover how these mutations circumvent the action of drugs targeting the TMD.

To further dissect the resistance landscape, we categorized the RPs by the nine established MmpL3 inhibitor classes, as well as a tenth miscellaneous grouping of orphan scaffolds (Figure S3). All the scaffolds retained RP clustering around the common drug binding pocket, corroborating its use by classes yet to be structurally characterized in complex with MmpL3. Armed with our RP mapping data we undertook a finer analysis with respect to the drug binding poses resolved in the *Msmg* structures. This highlighted that class-specific differences in RP distribution can be attributed to subtleties in the interactions each makes with MmpL3. The most obvious example are the pyrrole/pyrazole class inhibitors, where RPs are enriched around a subsite (residues L243, V681, V684, A685, and L703) unused in the binding of other structurally characterized scaffolds (Figure S3A) (Zhang et al., 2019). Equally, the differences apparent between indolecarboxamides (Figure S3C) and spirocycles (Figure S3G) agrees with observations that SPIRO, an example of the latter, makes fewer interactions at the top of the shared inhibitor pocket, but more at the bottom, relative to NITD-349, a member of the former family (Yang et al., 2020). Thus, despite significant overlap, there is clear differentiation in the pattern of RPs selected by each drug class.

A greater understanding of the MmpL3 mutations capable of conferring resistance to multiple scaffolds is critical to design of agents less easily inducing resistance. Such knowledge, for example, could direct medicinal chemistry efforts toward minimizing lead interactions with the offending residues. We

Please cite this article in press as: Adams et al., Cryo-EM structure and resistance landscape of *M. tuberculosis* MmpL3: An emergent therapeutic target, Structure (2021), https://doi.org/10.1016/j.str.2021.06.013

CellPress
OPEN ACCESS

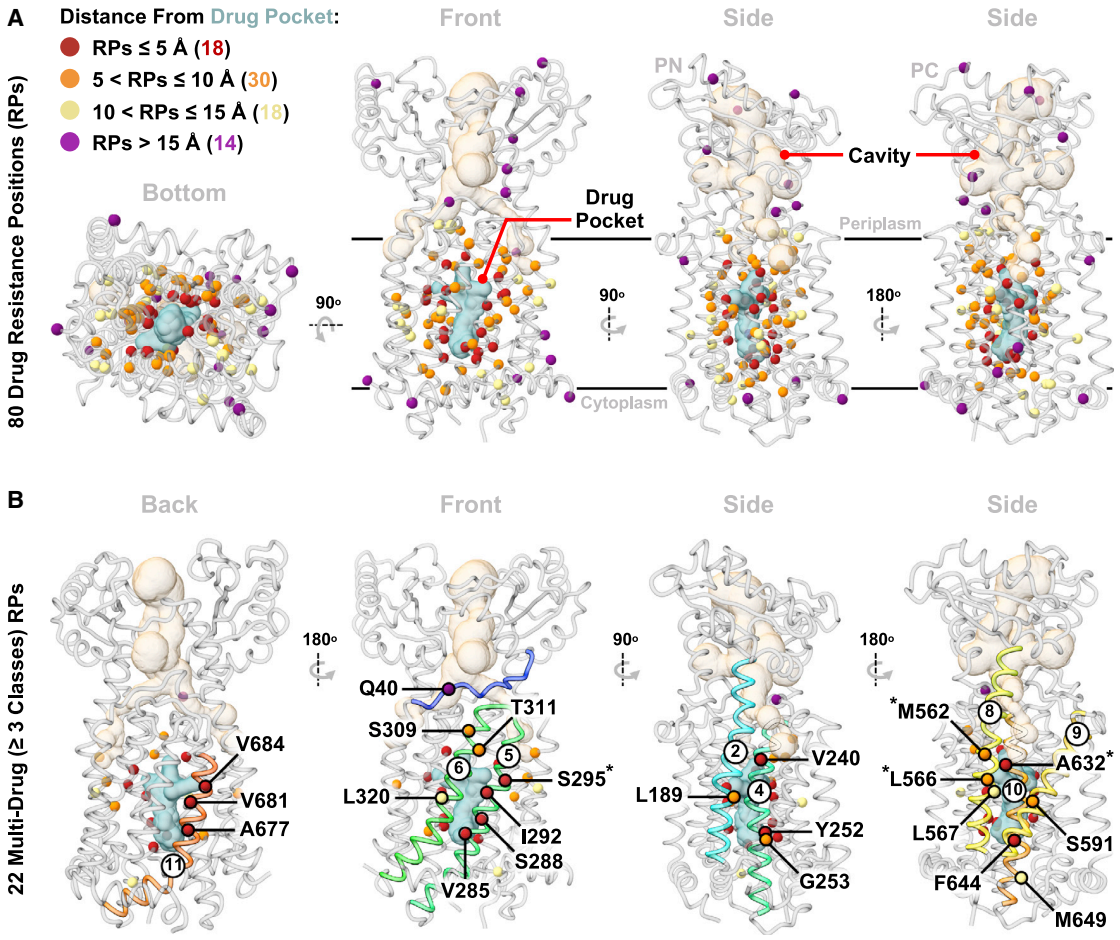


Figure 3. Mapping MmpL3 drug-resistance mutations

(A) Eighty resistance positions (RPs) plotted onto *Mtb* MmpL3$_{1-753}$ as balls, color-coded by distance from the universal drug pocket (teal). The PD cavity is also shown (wheat surface).

(B) Residues of the 22 RPs causing cross-resistance to a minimum of three classes of MmpL3 inhibitor. Black asterisks denote cross-RPs only observed in *M. abscessus* MmpL3.

See also Figures S3 and S4, and Table S2.

therefore filtered for RPs where variants have been associated with resistance to at least three MmpL3 inhibitor classes (Figure 3B). Twenty-two RPs were identified, all except Q40 fell within the TMD and 50% were within 5 Å of the common binding pocket. Mutations of Q40, which resides in the linker joining TM1 to the first PN helix and is >15 Å from the drug binding site, almost certainly act by the aforementioned "alternative" resistance mechanism. Consistently, the Q40R variant failed to alter *Msmg* MmpL3 affinity for inhibitors as measured by MST (Zhang et al., 2019). Q40 lines the MmpL3$_{1-753}$ channel exiting via TMs 1–4, within the LMNG binding site, and has thus far only been reported to confer resistance phenotypes on mutation to cationic amino acids (R or H, Table S2). It represents an important site for future biochemical studies.

Structural interpretation of the physiochemical changes introduced by variants at the other 21 cross-RPs identifies a number of plausible explanations for their phenotypes. A facile route to resistance, with minimal fold disruption, would be conservative binding site mutations that alter side chain size, thereby impairing steric complementarity. This mechanism is apparent for three *Mtb* drug-interacting valines (V240, V285, and V684) that mutate to smaller residues (A or G, Table S2) to achieve multi-class resistance. The reverse mechanism, enlarging side chains to promote steric clashes with complexed inhibitors, is employed at two drug-interacting sites, where alanine side chains (A632 and A677) are substituted for bulkier aliphatics or aromatics (Table S2). Modulation of residue hydrogen bond capacity is another route to resistance, either through loss (T311I, S591I), gain (I292S/T, L566S), or exchange (S288T) of hydroxyl-containing side chains (Table S2). Further away from the shared drug pocket many of the substitutions at cross-RPs result in dramatic chemical perturbations, in line with both their presumed action at a distance, and a diminished capacity to collaterally incapacitate the proton-relay system. These less-conservative variants either introduce additional charge into the TMD (L189R and G253E) or supply intra-TM prolines (S295P, L320P, and L567P, Table S2), which is conformationally disruptive. Interestingly, four of the 22 cross-RPs have so far only been observed in *M. abscessus*

Please cite this article in press as: Adams et al., Cryo-EM structure and resistance landscape of *M. tuberculosis* MmpL3: An emergent therapeutic target, Structure (2021), https://doi.org/10.1016/j.str.2021.06.013

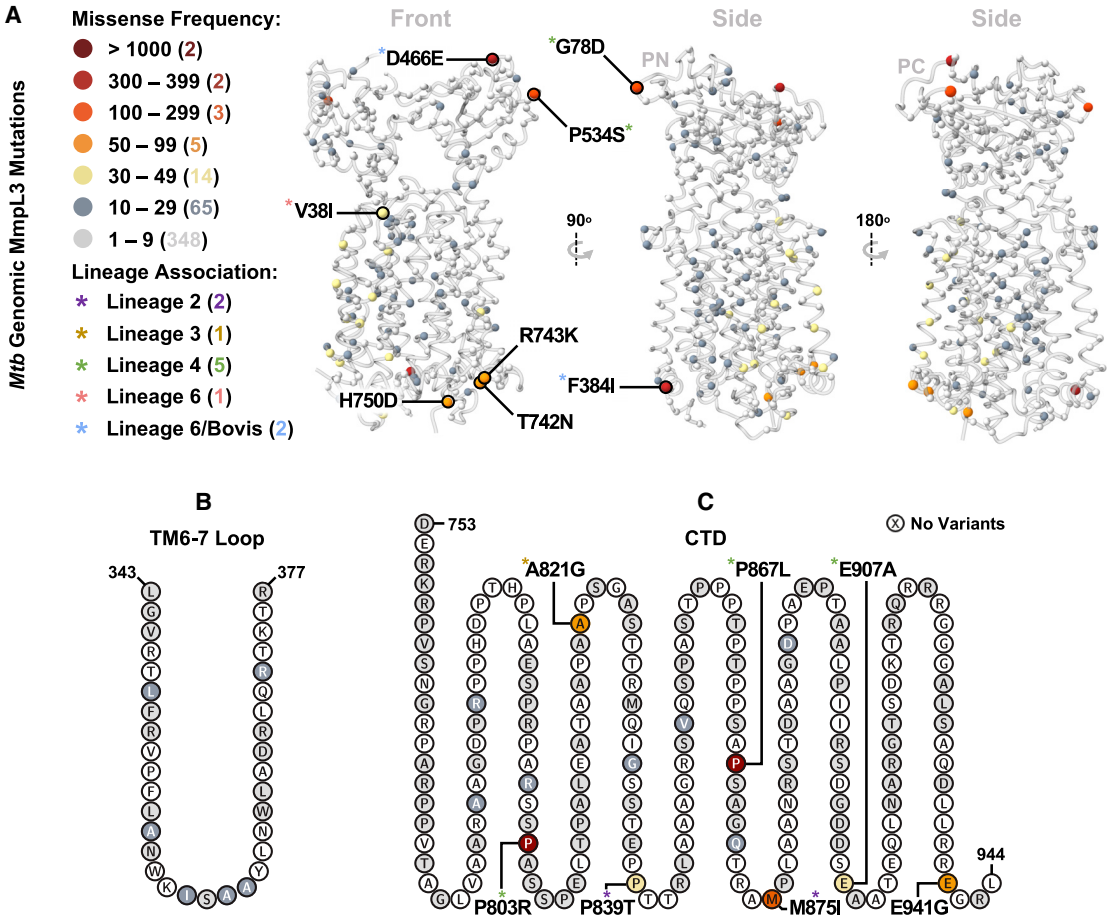


Figure 4. Wider non-synonymous landscape of MmpL3

(A) Structural mapping of MmpL3 variants mined from over 45,000 *Mtb* genomes, represented by balls and color-coded by mutational frequency. The most abundant substitutions, and/or those with apparent *Mtb* lineage association (colored asterisks) are labeled.

(B and C) (B) Incidence of non-synonymous mutations in the TM6-7 loop, and (C) C-terminal domain.

See also Figure S4 and Table S3.

(asterisks in Figure 3B). This suggests that MmpL3 orthologs may preferentially mutate at different residues to facilitate drug escape, an important caveat if future antitubercular MmpL3 inhibitors are to be repurposed for treatment of non-tuberculous mycobacterial infections (Li et al., 2018).

A large number of *Mtb* genomes (45,878), mostly originating from clinical samples, are available from the European Nucleotide Archive (ENA). By analyzing these with respect to our resistance catalog (Table S2) one can estimate how much environmental resistance any new antibiotics targeting MmpL3 are likely to encounter. Consistent with its essential nature, no MmpL3 mutations were detected in the majority of genomes; 3,890 (8.5%) out of the 45,878 samples harbored 7,475 non-synonymous mutations (Table S3). In total, 597 distinct amino acid substitutions were identified; however, the majority (422, 71%) were seen three or fewer times. Only 25 mutations were detected in 30 or more samples and accounted for 5,374 (72%) of the non-synonymous changes.

These 25 mutations are found throughout the protein (Figure 4) and none were identified by the *in vitro* resistance studies (Table S2). While we cannot conclude that they do not confer resistance, one can infer that it is unlikely. As may be expected, the chemical perturbation introduced by these variants tends to be smaller in the TMD than in the CTD, mirroring domain conservation (Figures S4B–S4D). A subset also appear specific to particular *Mtb* lineages (asterisks in Figure 4), suggesting that they are ancestral (Comas et al., 2013). Conversely, the lack of abundant non-lineage defining mutations (Table S3) implies there is currently little evolutionary pressure on the *mmpL3* gene and/or non-synonymous mutations tend to be deleterious.

The most prevalent resistance-conferring mutation in the genome-mined dataset, V210A, occurs in just 21 samples. Although 16 of these are found in but three studies (Colangeli et al., 2018; Guerra-Assunçaõ et al., 2015; Pankhurst et al., 2016), all three studies sequence clinical samples that can reasonably be expected not to have been exposed to any of the novel MmpL3 inhibitors under development. A further 12 mutations identified by experiment as conferring resistance are present in an additional 32 samples in the ENA (Table S3), but only three (G628A, Y252C, and I585V) are found in more than five

Please cite this article in press as: Adams et al., Cryo-EM structure and resistance landscape of *M. tuberculosis* MmpL3: An emergent therapeutic target, Structure (2021), https://doi.org/10.1016/j.str.2021.06.013

isolates. Simply counting these samples, and bearing in mind that the *Mtb* genomes deposited in the ENA are likely to be enriched for resistance, we arrive at 0.1% as an upper limit for the prevalence of pre-existing resistance mutations in MmpL3. This low percentage provides evidence that MmpL3 inhibitors, once licensed, are unlikely to encounter resistant variants in the environment.

In summary, our results provide a structural model of *Mtb* MmpL3, which, combined with our genetic analyses, provides a robust framework for antitubercular development against this essential protein.

MEMBERS OF THE CRyPTIC CONSORTIUM

Derrick W. Crook, Timothy E.A. Peto, A. Sarah Walker, Sarah J. Hoosdally, Ana L. Gibertoni Cruz, Joshua Carter, Alice Brankin, Sarah Earle, Samaneh Kouchaki, Alexander S. Lachapelle, Yang Yang, Timothy M. Walker, Philip W. Fowler, Daniel Wilson, and David A. Clifton, University of Oxford; Zamin Iqbal, Martin Hunt, Kerri M. Malone, Penelope Wintringer, Brice Letcher, and Jeff Knaggs, European Bioinformatics Institute; Daniela M. Cirillo, Emanuele Borroni, Simone Battaglia, Arash Ghodousi, Andrea Spitaleri, and Andrea Cabibbe, Emerging Bacterial Pathogens Unit, IRCCS San Raffaele Scientific Institute, Milan; Sabira Tahseen, National Tuberculosis Control Program Pakistan, Islamabad; Kayzad Nilgiriwala and Sanchi Shah, The Foundation for Medical Research, Mumbai; Camilla Rodrigues, Priti Kambli, Utkarsha Surve, and Rukhsar Khot, P.D. Hinduja National Hospital and Medical Research Center, Mumbai; Stefan Niemann, Thomas Kohl, and Matthias Merker, Research Center Borstel; Harald Hoffmann, Katharina Todt, and Sara Plesnik, Institute of Microbiology and Laboratory Medicine, IML red, Gauting; Nazir Ismail, Shaheed Vally Omar, Lavania Joseph Dumisani Ngcamu, Nana Okozi, and Shen Yuan Yao, National Institute for Communicable Diseases, Johannesburg; Guy Thwaites, Thuong Nguyen Thuy Thuong, Nhung Hoang Ngoc, and Vijay Srinivasan, Oxford University Clinical Research Unit, Ho Chi Minh City; David Moore, Jorge Coronel, and Walter Solano, London School of Hygiene and Tropical Medicine and Universidad Peruana Cayetano Heredia, Lima; George F. Gao, Guangxue He, Yanlin Zhao, Aijing Ma, and Chunfa Liu, China CDC, Beijing; Baoli Zhu, Institute of Microbiology, CAS, Beijing; Ian Laurenson and Pauline Claxton, Scottish Mycobacteria Reference Laboratory, Edinburgh; Robert J. Wilkinson, University of Cape Town, Imperial College London, and Francis Crick Institute; Anastasia Koch, University of Cape Town; Ajit Lalvani, Imperial College London; James Posey, CDC Atlanta; Jennifer Gardy, University of British Columbia; Jim Werngren, Public Health Agency of Sweden; Nicholas Paton, National University of Singapore; Ruwen Jou, Mei-Hua Wu, and Yu-Xin Xiao, CDC Taiwan; Lucilaine Ferrazoli, Rosangela Siqueira de Oliveira, and Juliana Maira Watanabe Pinhata, Institute Adolfo Lutz, São Paulo; James Millard, Africa Health Research Institute, Durban; Rob Warren, University of Stellenbosch, Cape Town; Annelies Van Rie, University of Antwerp; Simon Grandjean Lapierre, Marie-Sylvianne Rabodoarivelo, and Niaina Rakotosamimanana, Institut Pasteur de Madagascar; Camus Nimmo, University College London; Kimberlee Musser and Vincent Escuyer, Wadsworth Center, New York; Ted Cohen, Yale University; E. Grace Smith, Priti Rathod, Lisa Jarrett, and Daniela Matias, Public Health England, Birmingham.

SUPPORTING CITATIONS

The following references appear in the supplemental information: Dupont et al., 2019; Foss et al., 2016; Grover et al., 2021; Ioerger et al., 2013; Korycka-Machała et al., 2019; Kozikowski et al., 2017; Li et al., 2019; Li et al., 2020; McNeil et al., 2017; McNeil et al., 2020; Molin et al., 2019; Pandya et al., 2019; Poce et al., 2013; Poce et al., 2019; Rao et al., 2013; Raynaud et al., 2020; Shetty et al., 2018; Tantry et al., 2015; Trofimov et al., 2018; Zheng et al., 2018.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - *Mtb* MmpL3$_{1\text{-}753}$ molecular cloning & purification
  - Cryo-EM sample preparation & imaging
  - Cryo-EM data processing
  - Model building & refinement
  - Structural analysis & data Visualisation
  - *Mtb* MmpL3 mutation collation
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.str.2021.06.013.

ACKNOWLEDGMENTS

This work was supported by a Biotechnology and Biological Sciences Research Council (UKRI-BBSRC) studentship to O.A. (BB/M011224/1), Wellcome awards to S.M.L. (209194; 100298) and S.N. (215519; 219531), and through MRC grants to S.M.L. (MR/M011984/1) and J.L.P. (MR/S021043/1). Research in P.W.F.'s laboratory is supported by the NIHR Oxford Biomedical Research Center, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Oxford, UK, a Wellcome Enriching Engagement grant and Computational Biomedicine, an EU Horizon 2020 Center of Excellence (823712). We acknowledge use of the Central Oxford Structural Molecular Imaging Center (COSMIC), which is supported by Wellcome award (201536), the EPA Cephalosporin Trust, and a Royal Society/Wolfson Foundation Laboratory Refurbishment grant (WL160052). The CRyPTIC Consortium was supported by grants from the Bill and Melinda Gates Foundation (OPP1133541) and a Wellcome /Newton Fund-MRC Collaborative Award (200205/Z/15/Z). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health.

AUTHOR CONTRIBUTIONS

S.N., J.L.P., and P.W.F. conceived the project. O.A. performed all cloning and protein preparation. J.C.D. and S.M.L. performed all cryo-EM sample processing, data collection, and image analysis. O.A. and J.C.D. constructed the

Please cite this article in press as: Adams et al., Cryo-EM structure and resistance landscape of *M. tuberculosis* MmpL3: An emergent therapeutic target, Structure (2021), https://doi.org/10.1016/j.str.2021.06.013

atomic model. O.A. and P.W.F. conducted all bioinformatic and sequence-based analysis. O.A., P.W.F., J.L.P., and S.N. wrote the manuscript and prepared figures with contributions from J.C.D. and S.M.L.

DECLARATIONS OF INTERESTS

The authors declare no competing interests.

Received: April 12, 2021
Revised: May 24, 2021
Accepted: June 14, 2021
Published: November 4, 2021

REFERENCES

Asarnow, D., Palovcak, E., and Cheng, Y. (2019). asarnow/pyem: UCSF pyem v0.5. http://doi.org/10.5281/zenodo.3576630.

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. *44*, W344–W350.

Beitz, E. (2000a). TEXshade: shading and labeling of multiple sequence alignments using LATEX 2ε. Bioinformatics *16*, 135–139.

Beitz, E. (2000b). TEXtopo: shaded membrane protein topology plots in LATEX 2ε. Bioinformatics *16*, 1050–1051.

Belardinelli, J.M., Yazidi, A., Yang, L., Fabre, L., Li, W., Jacques, B., Angala, S.K., Rouiller, I., Zgurskaya, H.I., Sygusch, J., et al. (2016). Structure-function profile of MmpL3, the essential mycolic acid transporter from *Mycobacterium tuberculosis*. ACS Infect. Dis. *2*, 702–713.

Bernut, A., Viljoen, A., Dupont, C., Sapriel, G., Blaise, M., Bouchier, C., Brosch, R., de Chastellier, C., Herrmann, J.L., and Kremer, L. (2016). Insights into the smooth-to-rough transitioning in *Mycobacterium bolletii* unravels a functional Tyr residue conserved in all mycobacterial MmpL family members. Mol. Microbiol. *99*, 866–883.

Butler, M.S., and Paterson, D.L. (2020). Antibiotics in the clinical pipeline in October 2019. J. Antibiot. *73*, 329–364.

Caesar, J., Reboul, C.F., Machello, C., Kiesewetter, S., Tang, M.L., Deme, J.C., Johnson, S., Elmlund, D., Lea, S.M., and Elmlund, H. (2020). SIMPLE 3.0. Stream single-particle cryo-EM analysis in real time. J. Struct. Biol. X *4*, 100040.

Chalut, C. (2016). MmpL transporter-mediated export of cell-wall associated lipids and siderophores in mycobacteria. Tuberculosis *100*, 32–45.

Colangeli, R., Jedrey, H., Kim, S., Connell, R., Ma, S., Chippada Venkata, U.D., Chakravorty, S., Gupta, A., Sizemore, E.E., Diem, L., et al. (2018). Bacterial factors that predict relapse after tuberculosis therapy. New Engl. J. Med. *379*, 823–833.

Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K.E., Kato-Maeda, M., Parkhill, J., Malla, B., Berg, S., Thwaites, G., et al. (2013). Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. Nat. Genet. *45*, 1176–1182.

Degiacomi, G., Benjak, A., Madacki, J., Boldrin, F., Provvedi, R., Palù, G., Kordulakova, J., Cole, S.T., and Manganelli, R. (2017). Essentiality of *mmpL3* and impact of its silencing on *Mycobacterium tuberculosis* gene expression. Sci. Rep. *7*, 43495.

Degiacomi, G., Belardinelli, J.M., Pasca, M.R., de Rossi, E., Riccardi, G., and Chiarelli, L.R. (2020). Promiscuous targets for antitubercular drug discovery: the paradigm of DprE1 and MmpL3. Appl. Sci. *10*, 623.

DeJesus, M.A., Gerrick, E.R., Xu, W., Park, S.W., Long, J.E., Boutte, C.C., Rubin, E.J., Schnappinger, D., Ehrt, S., Fortune, S.M., et al. (2017). Comprehensive essentiality analysis of the *Mycobacterium tuberculosis* genome via saturating transposon mutagenesis. mBio *8*, 1–17.

Domenech, P., Reed, M.B., and Barry, C.E. (2005). Contribution of the *Mycobacterium tuberculosis* MmpL protein family to virulence and drug resistance. Infect. Immun. *73*, 3492–3501.

Drew, D.E., Heijne, G.V., Nordlund, P., and de Gier, J.L. (2001). Green fluorescent protein as an indicator to monitor membrane protein overexpression in *Escherichia coli*. FEBS Lett. *507*, 220–224.

Drew, D., Slotboom, D.-J., Friso, G., Reda, T., Genevaux, P., Rapp, M., Meindl-Beinker, N.M., Lambert, W., Lerch, M., Daley, D.O., et al. (2005). A scalable, GFP-based pipeline for membrane protein overexpression screening and purification. Protein Sci. *14*, 2011–2017.

Dulberger, C.L., Rubin, E.J., and Boutte, C.C. (2020). The mycobacterial cell envelope—a moving target. Nat. Rev. Microbiol. *18*, 47–59.

Dupont, C., Viljoen, A., Dubar, F., Blaise, M., Bernut, A., Pawlik, A., Bouchier, C., Brosch, R., Guérardel, Y., Lelièvre, J., et al. (2016). A new piperidinol derivative targeting mycolic acid transport in *Mycobacterium abscessus*. Mol. Microbiol. *101*, 515–529.

Dupont, C., Chen, Y., Xu, Z., Roquet-Banères, F., Blaise, M., Witt, A.-K., Dubar, F., Biot, C., Guérardel, Y., Maurer, F.P., et al. (2019). A piperidinol-containing molecule is active against *Mycobacterium tuberculosis* by inhibiting the mycolic acid flippase activity of MmpL3. J. Biol. Chem. *294*, 17512–17523.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D: Biol. Crystallogr. *66*, 486–501.

Foss, M.H., Pou, S., Davidson, P.M., Dunaj, J.L., Winter, R.W., Pou, S., Licon, M.H., Doh, J.K., Li, Y., Kelly, J.X., et al. (2016). Diphenylether-modified 1,2-diamines with improved drug properties for development against *Mycobacterium tuberculosis*. ACS Infect. Dis. *2*, 500–508.

Fujita, Y., Naka, T., Doi, T., and Yano, I. (2005). Direct molecular mass determination of trehalose monomycolate from 11 species of mycobacteria by MALDI-TOF mass spectrometry. Microbiology *151*, 1443–1452.

Goldman, R.C. (2013). Why are membrane targets discovered by phenotypic screens and genome sequencing in *Mycobacterium tuberculosis*? Tuberculosis *93*, 569–588.

Goodsell, D.S., Autin, L., and Olson, A.J. (2019). Illustrate: software for biomolecular illustration. Structure *27*, 1716–1720.e1.

De Groote, M.A., Jackson, M., Gonzalez-Juarrero, M., Li, W., Young, C.L., Wong, C., Graham, J., Day, J., Hoang, T., Jarvis, T.C., et al. (2018). Optimization and lead selection of benzothiazole amide analogs toward a novel antimycobacterial agent. Front. Microbiol. *9*, 2231.

Grover, S., Engelhart, C.A., Pérez-Herrán, E., Li, W., Abrahams, K.A., Papavinasasundaram, K., Bean, J.M., Sassetti, C.M., Mendoza-Losana, A., Besra, G.S., et al. (2021). Two-way regulation of MmpL3 expression identifies and validates inhibitors of MmpL3 function in *Mycobacterium tuberculosis*. ACS Infect. Dis. *7*, 141–152.

Grzegorzewicz, A.E., Pham, H., Gundi, V.A.K.B., Scherman, M.S., North, E.J., Hess, T., Jones, V., Gruppo, V., Born, S.E.M., Kordulakova, J., et al. (2012). Inhibition of mycolic acid transport across the *Mycobacterium tuberculosis* plasma membrane. Nat. Chem. Biol. *8*, 334–341.

Guerra-Assunçaõ, J.A., Houben, R.M.G.J., Crampin, A.C., Mzembe, T., Mallard, K., Coll, F., Khan, P., Banda, L., Chiwaya, A., Pereira, R.P.A., et al. (2015). Recurrence due to relapse or reinfection with *Mycobacterium tuberculosis*: a whole-genome sequencing approach in a large, population-based cohort with a high HIV infection prevalence and active follow-up. J. Infect. Dis. *211*, 1154–1163.

Hunt, M., Bradley, P., Lapierre, S.G., Heys, S., Thomsit, M., Hall, M.B., Malone, K.M., Wintringer, P., Walker, T.M., Cirillo, D.M., et al. (2019). Antibiotic resistance prediction for *Mycobacterium tuberculosis* from genome sequence data with Mykrobe. Wellcome Open Res. *4*, 191.

Ioerger, T.R., O'Malley, T., Liao, R., Guinn, K.M., Hickey, M.J., Mohaideen, N., Murphy, K.C., Boshoff, H.I.M., Mizrahi, V., Rubin, E.J., et al. (2013). Identification of new drug targets and resistance mechanisms in *Mycobacterium tuberculosis*. PLoS One *8*, e75245.

Jackson, M. (2014). The mycobacterial cell envelope—lipids. CSH Perspect. Med. *4*, 1–22.

Jurcik, A., Bednar, D., Byska, J., Marques, S.M., Furmanova, K., Daniel, L., Kokkonen, P., Brezovsky, J., Strnad, O., Stourac, J., et al. (2018). CAVER Analyst 2.0: analysis and visualization of channels and tunnels in protein structures and molecular dynamics trajectories. Bioinformatics *34*, 3586–3588.

Please cite this article in press as: Adams et al., Cryo-EM structure and resistance landscape of *M. tuberculosis* MmpL3: An emergent therapeutic target, Structure (2021), https://doi.org/10.1016/j.str.2021.06.013

Kelley, L.A., Mezulis, S., Yates, C., Wass, M., and Sternberg, M. (2015). The Phyre2 web portal for protein modelling, prediction, and analysis. Nat. Protoc. *10*, 845–858.

Klenotic, P.A., Moseng, M.A., Morgan, C.E., and Yu, E.W. (2020). Structural and functional diversity of resistance-nodulation-cell division transporters. Chem. Rev. *121*, 5378–5416.

Korycka-Machała, M., Viljoen, A., Pawełczyk, J., Borówka, P., Dziadek, B., Gobis, K., Brzostek, A., Kawka, M., Blaise, M., Strapagiel, D., et al. (2019). 1H-Benzo[*d*]Imidazole derivatives affect MmpL3 in *Mycobacterium tuberculosis*. Antimicrob. Agents Chemother. *63*, e00441–19.

Kozikowski, A.P., Onajole, O.K., Stec, J., Dupont, C., Viljoen, A., Richard, M., Chaira, T., Lun, S., Bishai, W., Raj, V.S., et al. (2017). Targeting mycolic acid transport by indole-2-carboxamides for the treatment of *Mycobacterium abscessus* infections. J. Med. Chem. *60*, 5876–5888.

Li, W., Obregón-Henao, A., Wallach, J.B., North, E.J., Lee, R.E., Gonzalez-Juarrero, M., Schnappinger, D., and Jackson, M. (2016). Therapeutic potential of the *Mycobacterium tuberculosis* mycolic acid transporter, Mmpl3. Antimicrob. Agents Chemother. *60*, 5198–5207.

Li, W., Sanchez-Hidalgo, A., Jones, V., de Moura, V.C.N., North, E.J., and Jackson, M. (2017). Synergistic interactions of MmpL3 inhibitors with antitubercular compounds *in vitro*. Antimicrob. Agents Chemother. *61*, 1–6.

Li, W., Yazidi, A., Pandya, A.N., Hegde, P., Tong, W., Calado Nogueira de Moura, V., North, E.J., Sygusch, J., and Jackson, M. (2018). MmpL3 as a target for the treatment of drug-resistant nontuberculous mycobacterial infections. Front. Microbiol. *9*, 1547.

Li, W., Stevens, C.M., Pandya, A.N., Darzynkiewicz, Z., Bhattarai, P., Tong, W., Gonzales-Juarrero, M., North, E.J., Zgurskaya, H.I., and Jackson, M.C. (2019). Direct inhibition of MmpL3 by novel antitubercular compounds. ACS Infect. Dis. *5*, 1001–1012.

Li, M., Phua, Z.Y., Xi, Y., Xu, Z., Nyantakyi, S.A., Li, W., Jackson, M., Wong, M.W., Lam, Y., Chng, S.S., et al. (2020). Potency increase of spiroketal analogs of membrane inserting indolyl mannich base antimycobacterials is due to acquisition of MmpL3 inhibition. ACS Infect. Dis. *6*, 1882–1893.

Liebschner, D., Afonine, P.v., Baker, M.L., Bunkoczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.W., Jain, S., McCoy, A.J., et al. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr. D Struct. Biol. *75*, 861–877.

Lun, S., Guo, H., Onajole, O.K., Pieroni, M., Gunosewoyo, H., Chen, G., Tipparaju, S.K., Ammerman, N.C., Kozikowski, A.P., and Bishai, W.R. (2013). Indoleamides are active against drug-resistant *Mycobacterium tuberculosis*. Nat. Commun. *4*, 1–8.

Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., et al. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. *47*, W636–W641.

McNeil, M.B., Dennison, D., and Parish, T. (2017). Mutations in MmpL3 alter membrane potential, hydrophobicity and antibiotic susceptibility in *Mycobacterium smegmatis*. Microbiology *163*, 1065–1070.

McNeil, M.B., Malley, T.O., Dennison, D., Shelton, C.D., and Sunde, B. (2020). Multiple mutations in *Mycobacterium tuberculosis* MmpL3 Increase resistance to MmpL3 inhibitors. mSphere *5*, 1–7.

Molin, M.D., Selchow, P., Schäfle, D., Tschumi, A., Ryckmans, T., Laage-witt, S., and Sander, P. (2019). Identification of novel scaffolds targeting *Mycobacterium tuberculosis*. J. Mol. Med. *97*, 1601–1613.

Nikaido, H. (2018). RND transporters in the living world. Res. Microbiol. *169*, 363–371.

Nikonenko, B.V., Protopopova, M., Samala, R., Einck, L., and Nacy, C.A. (2007). Drug therapy of experimental tuberculosis (TB): improved outcome by combining SQ109, a new diamine antibiotic, with existing TB drugs. Antimicrob. Agents Chemother. *51*, 1563–1565.

Pandya, A.N., Prathipati, P.K., Hegde, P., Li, W., Graham, K.F., Mandal, S., Drescher, K.M., Destache, C.J., Ordway, D., Jackson, M., et al. (2019). Indole-2-carboxamides are active against *Mycobacterium abscessus* in a mouse model of acute infection. Antimicrob. Agents Chemother. *63*, 1–16.

Pankhurst, L.J., del Ojo Elias, C., Votintseva, A.A., Walker, T.M., Cole, K., Davies, J., Fermont, J.M., Gascoyne-Binzi, D.M., Kohl, T.A., Kong, C., et al. (2016). Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study. Lancet Respir. Med. *4*, 49–58.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX: structure visualization for researchers, educators, and developers. Protein Sci. *30*, 70–82.

Poce, G., Bates, R.H., Alfonso, S., Cocozza, M., Porretta, G.C., Ballell, L., Rullas, J., Ortega, F., de Logu, A., Agus, E., et al. (2013). Improved BM212 MmpL3 inhibitor analogue shows efficacy in acute murine model of tuberculosis infection. PLoS One *8*, e56980.

Poce, G., Consalvi, S., Venditti, G., Alfonso, S., Desideri, N., Fernandez-Menendez, R., Bates, R.H., Ballell, L., Barros Aguirre, D., Rullas, J., et al. (2019). Novel pyrazole-containing compounds active against *Mycobacterium tuberculosis*. ACS Med. Chem. Lett. *10*, 1423–1429.

Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods *14*, 290–296.

Punjani, A., Zhang, H., and Fleet, D.J. (2020). Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. Nat. Methods *17*, 1214–1221.

Rao, S.P.S., Lakshminarayana, S.B., Kondreddi, R.R., Herve, M., Camacho, L.R., Bifani, P., Kalapala, S.K., Jiricek, J., Ma, N.L., Tan, B.H., et al. (2013). Indolcarboxamide is a preclinical candidate for treating multidrug-resistant tuberculosis. Sci. Transl. Med. *5*, 214ra168.

Rayasam, G.V. (2014). MmpL3 a potential new target for development of novel anti-tuberculosis drugs. Expert Opin. Ther. Targets *18*, 247–256.

Raynaud, C., Daher, W., Johansen, M.D., Roquet-Banères, F., Blaise, M., Onajole, O.K., Kozikowski, A.P., Herrmann, J.L., Dziadek, J., Gobis, K., et al. (2020). Active benzimidazole derivatives targeting the MmpL3 transporter in *Mycobacterium abscessus*. ACS Infect. Dis. *6*, 324–337.

Remuiñán, M.J., Pérez-Herrán, E., Rullás, J., Alemparte, C., Martínez-Hoyos, M., Dow, D.J., Afari, J., Mehta, N., Esquivias, J., Jiménez, E., et al. (2013). Tetrahydropyrazolo[1,5-a]pyrimidine-3-carboxamide and N-benzyl-6',7'-dihydrospiro[piperidine-4,4'-thieno[3,2-c]pyran] analogues with bactericidal efficacy against *Mycobacterium tuberculosis* targeting MmpL3. PLoS One *8*, e60933.

La Rosa, V., Poce, G., Canseco, J.O., Buroni, S., Pasca, M.R., Biava, M., Raju, R.M., Porretta, G.C., Alfonso, S., Battilocchio, C., et al. (2012). MmpL3 is the cellular target of the antitubercular pyrrole derivative BM212. Antimicrob. Agents Chemother. *56*, 324–331.

Shao, M., McNeil, M., Cook, G.M., and Lu, X. (2020). MmpL3 inhibitors as antituberculosis drugs. Eur. J. Med. Chem. *200*, 112390.

Shetty, A., Xu, Z., Lakshmanan, U., Hill, J., Choong, M.L., Chng, S.-S., Yamada, Y., Poulsen, A., Dick, T., and Gengenbacher, M. (2018). Novel acetamide indirectly targets mycobacterial transporter MmpL3 by proton motive force disruption. Front. Microbiol. *9*, 2960.

Stanley, S.A., Grant, S.S., Kawate, T., Iwase, N., Shimizu, M., Wivagg, C., Silvis, M., Kazyanskaya, E., Aquadro, J., Golas, A., et al. (2012). Identification of novel inhibitors of *M. tuberculosis* growth using whole cell based high-throughput screening. ACS Chem. Biol. *7*, 1377–1384.

Stec, J., Onajole, O.K., Lun, S., Guo, H., Merenbloom, B., Vistoli, G., Bishai, W.R., and Kozikowski, A.P. (2016). Indole-2-carboxamide-based MmpL3 inhibitors show exceptional antitubercular activity in an animal model of tuberculosis infection. J. Med. Chem. *59*, 6232–6247.

Su, C.C., Klenotic, P.A., Bolla, J.R., Purdy, G.E., Robinson, C.V., and Yu, E.W. (2019). MmpL3 is a lipid transporter that binds trehalose monomycolate and phosphatidylethanolamine. Proc. Natl. Acad. Sci. U S A *166*, 11241–11246.

Székely, R., and Cole, S.T. (2016). Mechanistic insight into mycobacterial MmpL protein function. Mol. Microbiol. *99*, 831–834.

Please cite this article in press as: Adams et al., Cryo-EM structure and resistance landscape of *M. tuberculosis* MmpL3: An emergent therapeutic target, Structure (2021), https://doi.org/10.1016/j.str.2021.06.013

Tahlan, K., Wilson, R., Kastrinsky, D.B., Arora, K., Nair, V., Fischer, E., Whitney Barnes, S., Walker, J.R., Alland, D., Barry, C.E., et al. (2012). SQ109 targets MmpL3, a membrane transporter of trehalose monomycolate involved in mycolic acid donation to the cell wall core of *Mycobacterium tuberculosis*. Antimicrob. Agents Chemother. *56*, 1797–1809.

Tantry, S.J., Degiacomi, G., Sharma, S., Jena, L.K., Narayan, A., Guptha, S., Shanbhag, G., Menasinakai, S., Mallya, M., Awasthy, D., et al. (2015). Whole cell screen based identification of spiropiperidines with potent antitubercular properties. Bioorg. Med. Chem. Lett. *25*, 3234–3245.

The CRyPTIC Consortium (2021). Epidemiological cutoff values.for a 96-well broth microdilution plate for high-throughput research antibiotic susceptibility testing of *M. tuberculosis*. medRxiv. https://doi.org/10.1101/2021.02.24.21252386.

The CRyPTIC Consortium and the 100,000 genomes project (2018). Prediction of susceptibility to first-line tuberculosis drugs by DNA sequencing. New Engl. J. Med. *379*, 1403–1415.

Trofimov, V., Kicka, S., Mucaria, S., Hanna, N., Ramon-Olayo, F., del Peral, L.V.-G., Lelièvre, J., Ballell, L., Scapozza, L., Besra, G.S., et al. (2018). Antimycobacterial drug discovery using mycobacteria-infected amoebae identifies anti-infectives and new molecular targets. Sci. Rep. *8*, 3939.

Udwadia, Z.F., Amale, R.A., Ajbani, K.K., and Rodrigues, C. (2012). Totally drug-resistant tuberculosis in India. Clin. Infect. Dis. *54*, 579–581.

Varela, C., Rittmann, D., Singh, A., Krumbach, K., Bhatt, K., Eggeling, L., Besra, G.S., and Bhatt, A. (2012). MmpL genes are associated with mycolic acid metabolism in mycobacteria and corynebacteria. Chem. Biol. *19*, 498–506.

Velayati, A.A., Masjedi, M.R., Farnia, P., Tabarsi, P., Ghanavi, J., ZiaZarifi, A.H., and Hoffner, S.E. (2009). Emergence of new forms of totally drug-resistant tuberculosis bacilli: super extensively drug-resistant tuberculosis or totally drug-resistant strains in Iran. Chest *136*, 420–425.

Viljoen, A., Dubois, V., Girard-Misguich, F., Blaise, M., Herrmann, J.L., and Kremer, L. (2017). The diverse family of MmpL transporters in mycobacteria: from regulation to antimicrobial developments. Mol. Microbiol. *104*, 889–904.

Waldo, G.S., Standish, B.M., Berendzen, J., and Terwilliger, T.C. (1999). Rapid protein-folding assay using green fluorescent protein. Nat. Biotechnol. *17*, 691–695.

Walker, T.M., Kohl, T.A., Omar, S.V., Hedge, J., del Ojo Elias, C., Bradley, P., Iqbal, Z., Feuerriegel, S., Niehaus, K.E., Wilson, D.J., et al. (2015). Whole-genome sequencing for prediction of *Mycobacterium tuberculosis* drug susceptibility and resistance: a retrospective cohort study. Lancet Infect. Dis. *15*, 1193–1202.

Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., et al. (2018). MolProbity: more and better reference data for improved all-atom structure validation. Protein Sci. *27*, 293–315.

Williams, J.T., Haiderer, E.R., Coulson, G.B., Conner, K.N., Ellsworth, E., Chen, C., Alvarez-Cabrera, N., Li, W., Jackson, M., Dick, T., et al. (2019). Identification of new MmpL3 inhibitors by untargeted and targeted mutant screens defines MmpL3 domains with differential resistance. Antimicrob. Agents Chemother. *63*, 1–18.

World Health Organization (2020). Global Tuberculosis Report 2020. Licence: CC BY-NC-SA 3.0. LGO (Geneva: World Health Organization).

Yang, X., Hu, T., Yang, X., Xu, W., Yang, H., Guddat, L.W., Zhang, B., and Rao, Z. (2020). Structural basis for the inhibition of mycobacterial MmpL3 by NITD-349 and SPIRO. J. Mol. Biol. *432*, 4426–4434.

Zhang, B., Li, J., Yang, X., Wu, L., Zhang, J., Yang, Y., Zhao, Y., Zhang, L., Yang, X., Yang, X., et al. (2019). Crystal structures of membrane transporter MmpL3, an anti-TB drug target. Cell *176*, 636–648.e13.

Zheng, H., Williams, J.T., Coulson, G.B., Haiderer, E.R., and Abramovitch, R.B. (2018). HC2091 kills *Mycobacterium tuberculosis* by targeting the MmpL3 mycolic acid transporter. Antimicrob. Agents Chemother. *62*, e02459–17.

Zivanov, J., Nakane, T., and Scheres, S.H.W. (2019). A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. IUCrJ *6*, 5–17.

Zivanov, J., Nakane, T., and Scheres, S.H.W. (2020). Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1. IUCrJ *7*, 253–267.

Please cite this article in press as: Adams et al., Cryo-EM structure and resistance landscape of *M. tuberculosis* MmpL3: An emergent therapeutic target, Structure (2021), https://doi.org/10.1016/j.str.2021.06.013

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
E. coli OmniMAX	Invitrogen	Cat# C854003
E. coli C43(DE3)	Lucigen	Cat# 60446-1
Chemicals, peptides, and recombinant proteins		
Lauryl Maltose Neopentyl Glycol (LMNG)	Anatrace	Cat# NG310
TEV Protease	Produced In-house	N/A
Critical commercial assays		
HisPur-Ni-NTA Resin	Thermo Fisher Scientific	Cat# 88221
HisTrap Column	Cytiva	Cat# 17-5255-01
Superdex 200 10/300 Increase GL Column	Cytiva	Cat# 28-9909-44
Deposited data		
Mtb MmpL3$_{1-753}$ Cryo-EM Volume	This Paper	EMDB code: EMD-12604
Mtb MmpL3$_{1-753}$ Atomic Coordinates	This Paper	PDB code: 7NVH
Msmg MmpL3$_{1-773}$ Atomic Coordinates	Su et al., 2019	PDB code: 6OR2
Oligonucleotides		
Primer: NdeI MmpL3$_{1-753}$ Forward GGAATTCCATATGTTTGCATGGT GGGGTCG	This Paper	N/A
Primer: BamHI MmpL3$_{1-753}$ Reverse ACTGGGATCCATCAGGCAGATGA ATTTCACCCAGGCC	This Paper	N/A
Recombinant DNA		
Plasmid: pMA-T-MmpL3$_{1-944}$	Invitrogen GeneArt	N/A
Plasmid: pWaldo-GFPd	Drew et al., 2001	N/A
Plasmid: pWaldo-GFPd-MmpL3$_{1-753}$	This Paper	N/A
Software and algorithms		
SIMPLE 3.0	Caesar et al., 2020	https://simplecryoem.com/
cryoSPARC (v.2.15.0)	Punjani et al., 2017	https://cryosparc.com/
UCSF pyem (v.0.5)	Asarnow et al., 2019	https://zenodo.org/record/3576630
RELION 3.1.0	Zivanov et al., 2020	https://www3.mrc-lmb.cam.ac.uk/relion/index.php/Main_Page
Phyre2 Server	Kelley et al., 2015	http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index
Coot (v.0.9.2-pre)	Emsley et al., 2010	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
PHENIX (v.1.18.2)	Liebschner et al., 2019	http://www.phenix-online.org/
UCSF Chimera (v.1.13.1)	Pettersen et al., 2004	https://www.cgl.ucsf.edu/chimera/
UCSF ChimeraX (v.1.1)	Pettersen et al., 2021	https://www.rbvi.ucsf.edu/chimerax/
CAVER Analyst 2.0	Jurcik et al., 2018	http://www.caver.cz/
ConSurf Server	Ashkenazy et al., 2016	https://consurf.tau.ac.il/
T-Coffee	Madeira et al., 2019	https://www.ebi.ac.uk/Tools/msa/tcoffee/
ChemDraw Professional (v.17.0.0.206)	PerkinElmer Informatics	https://perkinelmerinformatics.com/products/research/chemdraw/
Illustrate	Goodsell et al., 2019	https://ccsb.scripps.edu/illustrate/
T_EXshade	Beitz, 2000a	https://ctan.org/pkg/texshade

(*Continued on next page*)

Please cite this article in press as: Adams et al., Cryo-EM structure and resistance landscape of *M. tuberculosis* MmpL3: An emergent therapeutic target, Structure (2021), https://doi.org/10.1016/j.str.2021.06.013

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
T_EXtopo	Beitz, 2000b	https://ctan.org/pkg/textopo
Mykrobe	Hunt et al., 2019	https://www.mykrobe.com/
Other		
Vivaspin-20 (50 kDa MWCO) Concentrator	Sartorius	Cat# VS2032
Vivaspin-500 (50 kDa MWCO) Concentrator	Sartorius	Cat# VS0132
Quantifoil 300 Mesh Au R1.2/1.3 Grids	Agar Scientific	Cat# AGS143-8

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Simon Newstead (simon.newstead@bioch.ox.ac.uk).

Materials availability

All unique reagents generated in this study are available from the lead contact upon reasonable request.

Data and code availability

The cryo-EM volume has been deposited in the Electron Microscopy DataBank (EMDB) with accession code EMDB-12604, and the atomic coordinates have been deposited in the Protein DataBank (PDB) with accession code 7NVH. The mutational data underlying Figures 3 and 4 are provided in Tables S2 and S3 respectively. This paper does not report original code. Any additional information required to reanalyze the data reported is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Plasmid propagation and clone recovery were performed using OmniMAX *E. coli* cells (Invitrogen) plated on Luria Broth (LB) agar (37°C) or inoculated in LB liquid media (37°C, 180 RPM); both grown overnight in the presence of a selective antibiotic. Recombinant *Mtb* $MmpL3_{1\text{-}753}$ was produced in *E. coli* C43(DE3) cells (Lucigen), cultured in Terrific Broth supplemented with kanamycin (50 μg mL^{-1}) for ~18 hr (25°C, 215 RPM) following IPTG induction (final concentration 0.4 mM) at an OD_{600} ~0.6–0.8.

METHOD DETAILS

Mtb $MmpL3_{1\text{-}753}$ molecular cloning & purification

An *E. coli* codon-optimised version of the gene (*Rv0206c*) encoding full-length (residues 1–944) *Mtb* MmpL3 (Uniprot ID: P9WJV5) was synthesised by Invitrogen GeneArt services, provided in the vector pMA-T. Using the primers detailed in the Key Resources Table, a truncated open reading frame encoding only residues 1–753 of *Mtb* MmpL3 was amplified and then restriction sub-cloned into IPTG-inducible vector pWaldo-GFPd (Drew et al., 2001; Waldo et al., 1999); placing it in-frame with a C-terminal TEV-GFP-His_8 affinity-tag. $MmpL3_{1\text{-}753}$ expression in *E. coli* and purification to homogeneity, by Ni^{2+}-affinity chromatography followed by SEC, was achieved using standard protocols (Drew et al., 2005). All purification steps were undertaken at 4°C. Briefly, thawed membranes were first resuspended in 1X PBS supplemented with 150 mM NaCl, prior to solubilization in 1% (w/v) LMNG (Anatrace) under gentle agitation for 1.5 hr. Non-solubilized material was removed by ultracentrifugation (>200,000 g, 1 hr), and then imidazole to a final concentration of 15 mM was added to the recovered supernatant. Protein in this was bound to HisPur-Ni-NTA resin (Thermo Fisher Scientific) in batch, by incubation for ~2 hr 45 min with stirring. The resin, packed into a glass econo-column (BioRad), was next sequentially washed with 10 column volumes (CVs) of purification buffer (1x PBS, 150 mM NaCl, 0.1% (w/v) LMNG) containing 15 mM imidazole, and 15 CVs supplemented with 30 mM imidazole. $MmpL3_{1\text{-}753}$ was finally eluted in 5 CVs of purification buffer featuring 250 mM imidazole. After TEV protease addition, the protein was dialyzed overnight in gel filtration buffer (20 mM Tris pH 7.5, 150 mM NaCl, 0.003% (w/v) LMNG). The next day, dialysate was passed through a 5 mL HisTrap column (Cytiva) to remove both the TEV protease and liberated GFP-His_8 affinity-tag. The resulting pure $MmpL_{1\text{-}753}$ was spin concentrated using a Vivaspin-20 (50 kDa MWCO, Sartorius), prior to SEC polishing on a Superdex 200 10/300 Increase GL column (Cytiva) pre-equilibrated in gel filtration buffer. For cryo-EM, SEC fractions of interest were immediately pooled, spin concentrated to 2 mg mL^{-1} in a Vivaspin-500 (50 kDa MWCO, Sartorius), and the protein then kept on ice until grid preparation the same day. $MmpL3_{1\text{-}753}$ purity was assessed throughout purification by analysis of samples loaded on 10% SDS-PAGE gels stained with Coomassie blue (InstantBlue, Expedeon).

Cryo-EM sample preparation & imaging

A 4 μL aliquot of pure, LMNG-solubilized, $MmpL3_{1\text{-}753}$ (2 mg mL^{-1}) was dispensed onto the surface of a freshly glow-discharged holey carbon-coated grid (Quantifoil 300 mesh, Au R1.2/1.3, Agar Scientific). After allowing 10 s for adsorption, the grid was blotted

Please cite this article in press as: Adams et al., Cryo-EM structure and resistance landscape of *M. tuberculosis* MmpL3: An emergent therapeutic target, Structure (2021), https://doi.org/10.1016/j.str.2021.06.013

for 3 s (blot force −5, 100% humidity, 8°C) and then flash-frozen by plunging into liquid ethane using a Vitrobot Mark IV (Thermo Fisher Scientific). Data were gathered in counting super-resolution mode on a 300 kV Titan Krios G3 cryo-TEM (Thermo Fisher Scientific) equipped with a BioQuantum Imaging Filter (Gatan) and K3 Direct Electron Detector (Gatan). Using a pixel size 0.832 Å, 15,020 movies spanning a defocus range of −3.0 to −0.5 μm were collected; each with a total dose of 58.2 e−/Å^2, spread across 40 fractions, over an exposure time of 2.80 s. A representative micrograph is shown in Figure S1B.

Cryo-EM data processing

A schematic summarising the MmpL3$_{1\text{-}753}$ cryo-EM data processing workflow is provided in Figure S1C. Patched (15 x 10) motion correction, dose weighting, contrast transfer function (CTF) estimation, and particle autopicking were all implemented on-the-fly using SIMPLE 3.0 (Caesar et al., 2020). This yielded 7,764,838 particles, extracted in 256 × 256-pixel boxes, which were then subject to two rounds of reference-free 2D classification; the first on-the-fly within SIMPLE 3.0, using cluster2D-stream, and the second in cryoSPARC (v.2.15.0, Punjani et al., 2017). Discarding poor-quality classes after each round left 1,564,980 particles for downstream processing, which was performed predominantly in cryoSPARC. The particle subset was first subject to multi-class (k = 4) *ab initio* model generation, without symmetry imposition (i.e. in C1). This generated a single featureful volume, displaying the expected MmpL3$_{1\text{-}753}$ domain organization, composed of 758,680 particles. Subsequently passing these through two rounds of non-uniform refinement (Punjani et al., 2020), interspersed by Bayesian particle polishing (Zivanov et al., 2019) undertaken in RELION 3.1.0 (Zivanov et al., 2020) within 320 × 320-pixel boxes, gave a 3.4 Å map. The requisite STAR file for this polishing was generated using the csparc2star.py script from UCSF pyem (Asarnow et al., 2019). Further heterogeneous refinement of the volume against the previous four *ab initio* classes recovered 562,667 particles, whose non-uniform refinement returned a 3.3 Å reconstruction. A final round of reference-free 2D classification was employed to remove any remaining suboptimal particles, resulting in a stack retaining the best 414,082. Successive non-uniform, beamtilt, and local refinement of these yielded the 3.0 Å map of MmpL3$_{1\text{-}753}$ deposited in the EMDB under accession code EMD-12604. Throughout processing all refinement reference maps were lowpass filtered to 8 Å. Gold-standard Fourier shell correlations (FSC) using the 0.143 criterion, as well as local resolution estimations, were calculated in RELION 3.1.0 (Figures S1D and S1E).

Model building & refinement

Atomic model building began by threading the *Mtb* MmpL3$_{1\text{-}753}$ sequence onto a previously reported crystal structure of the *Msmg* orthologue (PDB code: 6OR2, Su et al., 2019), using the Phyre2 web server (Kelley et al., 2015). The threaded structure was then docked into the 3.0 Å cryo-EM volume in UCSF Chimera (v.1.13.1, Pettersen et al., 2004), prior to further rigid body fitting in Coot (v.0.9.2-pre, Emsley et al., 2010). Within the latter program the model underwent manual real-space refinement (RSR), regions lacking unambiguous density were removed (residues 343–377, 753), and an LMNG molecule was built into the splayed species occupying the PD chamber. To generate the final model, detailed in Table S1 and deposited in the PDB under accession code 7NVH, the MmpL3$_{1\text{-}753}$-detergent complex underwent further rounds of RSR in PHENIX (v.1.18.2, Liebschner et al., 2019) with secondary structure, rotamer, ligand, and Ramachandran restraints applied, alongside additional minor adjustments in Coot. The globally sharpened map, output from cryoSPARC local refinement, was used for all rounds of RSR in both Coot and PHENIX. Model validity was assessed by MolProbity (Williams et al., 2018) executed within the latter suite. The density fit for modeled TM side chains and the LMNG ligand are shown in Figure S1F.

Structural analysis & data Visualisation

The MmpL3$_{1\text{-}753}$ cavity was calculated with Caver Analyst 2.0 (Jurcik et al., 2018) using default parameters. Conservation analysis was performed on the ConSurf server (Ashkenazy et al., 2016). The MmpL3 ortholog sequence alignment was generated by T-Coffee (Madeira et al., 2019), and typeset with T$_E$Xshade (Beitz, 2000a). All structural figures were composed in UCSF Chimera (v.1.13.1, Pettersen et al., 2004) or UCSF ChimeraX (v.1.1, Pettersen et al., 2021). Primary sequence representations of the structurally unresolved regions of *Mtb* MmpL3 were made with T$_E$Xtopo (Beitz, 2000b). Antitubercular skeletal formulas were prepared in ChemDraw Professional (v.17.0.0.206), and the protein elements of the graphical abstract were produced using Illustrate (Goodsell et al., 2019).

Mtb MmpL3 mutation collation

Variants of *Mtb* MmpL3 observed to confer resistance to one or more preclinical agents were extracted from 29 publications. These were then manually curated to give Table S2, which includes references to all the original studies. Mutations identified in mycobacterial species other than *Mtb* were back-mapped by means of sequence alignments between the relevant MmpL3 orthologs. The broader mutational landscape of *Mtb* MmpL3 was sampled by retrieving all *Mtb* genomes deposited in the ENA at the time of study, together with those reported in Walker et al. (2015) and The CRyPTIC Consortium and the 100,000 Genomes Project, 2018. Totalling 45,878, the majority of these genomes underwent *Mtb* lineage classification using Mykrobe (Hunt et al., 2019) and the short reads were mapped onto version 3 of the H37Rv *Mtb* reference genome (NC_000962.3) and variants called using Clockwork v.0.8.3 as described elsewhere (The CRyPTIC Consortium, 2021). The CRyPTIC data warehouse was then interrogated for all non-synonymous substitutions in the MmpL3 open reading frame. The resultant 7,475 mutations, and any lineage associations of the most abundant, are compiled in Table S3.

Please cite this article in press as: Adams et al., Cryo-EM structure and resistance landscape of *M. tuberculosis* MmpL3: An emergent therapeutic target, Structure (2021), https://doi.org/10.1016/j.str.2021.06.013

QUANTIFICATION AND STATISTICAL ANALYSIS

Cryo-EM reconstruction and model building of *Mtb* MmpL3$_{1\text{-}753}$ was performed as described in the relevant 'Method details' sections, employing the software packages SIMPLE 3.0, cryoSPARC (v.2.15.0), RELION 3.1.0, Coot (v.0.9.2-pre), and PHENIX (v.1.18.2) as noted in the Key Resources Table. Final model statistics are given in Table S1.

Bibliography

Aires JR, Nikaido H. Aminoglycosides Are Captured from both Periplasm and Cytoplasm by the AcrD Multidrug Efflux Transporter of Escherichia coli. *Journal of Bacteriology*. 2005 mar; 187(6):1923–1929. , doi:10.1128/JB.187.6.1923-1929.2005.

Alav I, Kobylka J, Kuth MS, Pos KM, Picard M, Blair JMA, Bavro VN. Structure, Assembly, and Function of Tripartite Efflux and Type 1 Secretion Systems in Gram-Negative Bacteria. *Chemical Reviews*. 2021 may; 121(9):5479–5596. , doi:10.1021/acs.chemrev.1c00055.

Alderwick LJ, Harrison J, Lloyd GS, Birch HL. The Mycobacterial Cell Wall—Peptidoglycan and Arabinogalactan. *Cold Spring Harbor Perspectives in Medicine*. 2015 aug; 5(8):a021113. , doi:10.1101/cshperspect.a021113.

Alvarez-Ortega C, Olivares J, Martínez JL. RND multidrug efflux pumps: what are they good for? *Frontiers in Microbiology*. 2013; 4(FEB):1–11. , doi:10.3389/fmicb.2013.00007.

Anandan A, Vrielink A. Detergents in Membrane Protein Purification and Crystallisation. In: Moraes I, editor. *The next generation in membrane protein structure determination*, vol. 922; 2016.p. 13–28. , doi:10.1007/978-3-319-35072-1_2.

Arbabi-Ghahroudi M. Camelid Single-Domain Antibodies: Historical Perspective and Future Outlook. *Frontiers in Immunology*. 2017 nov; 8(NOV):1–8. , doi:10.3389/fimmu.2017.01589.

Arkhipova V, Guskov A, Slotboom DJ. Structural ensemble of a glutamate transporter homologue in lipid nanodisc environment. *Nature Communications*. 2020 dec; 11(1):998. , doi:10.1038/s41467-020-14834-8.

Arnold FM, Hohl M, Remm S, Koliwer-Brandl H, Adenau S, Chusri S, Sander P, Hilbi H, Seeger MA. A uniform cloning platform for mycobacterial genetics and protein production. *Scientific Reports*. 2018 dec; 8(1):9539. , doi:10.1038/s41598-018-27687-5.

Arnold FM, Weber MS, Gonda I, Gallenito MJ, Adenau S, Egloff P, Zimmermann I, Hutter CAJ, Hürlimann LM, Peters EE, Piel J, Meloni G, Medalia O, Seeger MA. The ABC exporter IrtAB imports and reduces mycobacterial siderophores. *Nature*. 2020 apr; 580(7803):413–417. , doi:10.1038/s41586-020-2136-9.

Arumugam P, Shankaran D, Bothra A, Gandotra S, Rao V. The MmpS6-MmpL6 Operon Is an Oxidative Stress Response System Providing Selective Advantage to Mycobacterium tuberculosis in Stress. *The Journal of Infectious Diseases*. 2019 jan; 219(3):459–469. , doi:10.1093/infdis/jiy526.

Asarnow D, Palovcak E, Cheng Y, asarnow/pyem: UCSF pyem v0.5; 2019. , doi:10.5281/ZENODO.3576630.

Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Research*. 2016 jul; 44(W1):W344–W350. , doi:10.1093/nar/gkw408.

Autzen HE, Julius D, Cheng Y. Membrane mimetic systems in CryoEM: keeping membrane proteins in their native environment. *Current Opinion in Structural Biology*. 2019 oct; 58:259–268. , doi:10.1016/j.sbi.2019.05.022.

Bashiri G, Baker EN. Production of recombinant proteins in Mycobacterium smegmatis for structural and functional studies. *Protein Science*. 2015 jan; 24(1):1–10. , doi:10.1002/pro.2584.

Bateman A. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*. 2019 jan; 47(D1):D506–D515. , doi:10.1093/nar/gky1049.

Batt SM, Burke CE, Moorey AR, Besra GS. Antibiotics and resistance: the two-sided coin of the mycobacterial cell wall. *The Cell Surface*. 2020 dec; 6(June):100044. , doi:10.1016/j.tcsw.2020.100044.

Bayburt TH, Grinkova YV, Sligar SG. Self-Assembly of Discoidal Phospholipid Bilayer Nanoparticles with Membrane Scaffold Proteins. *Nano Letters*. 2002 aug; 2(8):853–856. , doi:10.1021/nl025623k.

Bayburt TH, Sligar SG. Self-assembly of single integral membrane proteins into soluble nanoscale phospholipid bilayers. *Protein Science*. 2009 jan; 12(11):2476–2481. , doi:10.1110/ps.03267503.

Bayburt TH, Sligar SG. Membrane protein assembly into Nanodiscs. *FEBS Letters*. 2010 may; 584(9):1721–1727. , doi:10.1016/j.febslet.2009.10.024.

Becker K, Sander P. Mycobacterium tuberculosis lipoproteins in virulence and immunity - fighting with a double-edged sword. *FEBS Letters*. 2016 nov; 590(21):3800–3819. , doi:10.1002/1873-3468.12273.

Beckett D, Kovaleva E, Schatz PJ. A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Science*. 2008 dec; 8(4):921–929. , doi:10.1110/ps.8.4.921.

Beitz E. TEXtopo: Shaded membrane protein topology plots in LATEX2ϵ. *Bioinformatics*. 2000; 16(11):1050–1051. doi:10.1093/bioinformatics/16.11.1050.

Beitz E. TEXshade: Shading and labeling of multiple sequence alignments using LATEX 2ϵ. *Bioinformatics*. 2000; 16(2):135–139. doi:10.1093/bioinformatics/16.2.135.

Belardinelli JM, Larrouy-Maumus G, Jones V, Sorio de Carvalho LP, McNeil MR, Jackson M. Biosynthesis and Translocation of Unsulfated Acyltrehaloses in Mycobacterium tuberculosis. *Journal of Biological Chemistry*. 2014 oct; 289(40):27952–27965. , doi:10.1074/jbc.M114.581199.

Belardinelli JM, Stevens CM, Li W, Tan YZ, Jones V, Mancia F, Zgurskaya HI, Jackson M. The MmpL3 interactome reveals a complex crosstalk between cell envelope biosynthesis and cell elongation and division in mycobacteria. *Scientific Reports*. 2019 dec; 9(1):10728. , doi:10.1038/s41598-019-47159-8.

Belardinelli JM, Yazidi A, Yang L, Fabre L, Li W, Jacques B, Angala SK, Rouiller I, Zgurskaya HI, Sygusch J, Jackson M. Structure–Function Profile of MmpL3, the Essential Mycolic Acid Transporter from Mycobacterium tuberculosis. *ACS Infectious Diseases*. 2016 oct; 2(10):702–713. , doi:10.1021/acsinfecdis.6b00095.

Bernut A, Viljoen A, Dupont C, Sapriel G, Blaise M, Bouchier C, Brosch R, de Chastellier C, Herrmann JL, Kremer L. Insights into the smooth-to-rough transitioning in Mycobacterium bolletii unravels a functional Tyr residue conserved in all mycobacterial MmpL family members. *Molecular Microbiology*. 2016; 99(5):866–883. doi:10.1111/mmi.13283.

Bibow S, Polyhach Y, Eichmann C, Chi CN, Kowal J, Albiez S, McLeod RA, Stahlberg H, Jeschke G, Güntert P, Riek R. Solution structure of discoidal high-density lipoprotein particles with a shortened apolipoprotein A-I. *Nature Structural & Molecular Biology*. 2017 feb; 24(2):187–193. , doi:10.1038/nsmb.3345.

Birch J, Axford D, Foadi J, Meyer A, Eckhardt A, Thielmann Y, Moraes I. The fine art of integral membrane protein crystallisation. *Methods*. 2018 sep; 147(January):150–162. , doi:10.1016/j.ymeth.2018.05.014.

Bird LE, Nettleship JE, Järvinen V, Rada H, Verma A, Owens RJ. Expression Screening of Integral Membrane Proteins by Fusion to Fluorescent Reporters. In: Moraes I, editor. *The next generation in membrane protein structure determination*, vol. 922; 2016.p. 1–11. , doi:10.1007/978-3-319-35072-1.

Bloch JS, Pesciullesi G, Boilevin J, Nosol K, Irobalieva RN, Darbre T, Aebi M, Kossiakoff AA, Reymond JL, Locher KP. Structure and mechanism of the ER-based glucosyltransferase ALG6. *Nature*. 2020 mar; 579(7799):443–447. , doi:10.1038/s41586-020-2044-z.

Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. *Journal of Molecular Biology*. 1998 jul; 280(1):1–9. , doi:10.1006/jmbi.1998.1843.

Bolla JR. Targeting MmpL3 for anti-tuberculosis drug development. *Biochemical Society Transactions*. 2020 aug; 48(4):1463–1472. , doi:10.1042/BST20190950.

Bradbury ARM, Sidhu S, Dübel S, McCafferty J. Beyond natural antibodies: the power of in vitro display technologies. *Nature Biotechnology*. 2011 mar; 29(3):245–254. , doi:10.1038/nbt.1791.

Brandsch M. Transport of drugs by proton-coupled peptide transporters: pearls and pitfalls. *Expert Opinion on Drug Metabolism & Toxicology*. 2009 aug; 5(8):887–905. , doi:10.1517/17425250903042292.

Bräuer P, Parker JL, Gerondopoulos A, Zimmermann I, Seeger MA, Barr FA, Newstead S. Structural basis for pH-dependent retrieval of ER proteins from the Golgi by the KDEL receptor. *Science*. 2019 mar; 363(6431):1103–1107. , doi:10.1126/science.aaw2859.

Breyton C, Javed W, Vermot A, Arnaud CA, Hajjar C, Dupuy J, Petit-Hartlein I, Le Roy A, Martel A, Thépaut M, Orelle C, Jault JM, Fieschi F, Porcar L, Ebel C. Assemblies of lauryl maltose neopentyl glycol (LMNG) and LMNG-solubilized membrane proteins. *Biochimica et Biophysica Acta (BBA) - Biomembranes*. 2019 may; 1861(5):939–957. , doi:10.1016/j.bbamem.2019.02.003.

Briffotaux J, Huang W, Wang X, Gicquel B. MmpS5/MmpL5 as an efflux pump in Mycobacterium species. *Tuberculosis*. 2017 dec; 107:13–19. , doi:10.1016/j.tube.2017.08.001.

Brunner JD, Jakob RP, Schulze T, Neldner Y, Moroni A, Thiel G, Maier T, Schenck S. Structural basis for ion selectivity in TMEM175 K+ channels. *eLife*. 2020 apr; 9:480863. , doi:10.7554/eLife.53683.

Brunner JD, Schenck S. Production and Application of Nanobodies for Membrane Protein Structural Biology. In: Perez C, Maier T, editors. *Expression , Purification, and Structural Biology of Membrane Proteins, Methods in Molecular Biology* Springer US; 2020.p. 167–184. doi:https://doi.org/10.1007/978-1-0716-0373-4_12.

Bürgi J, Xue B, Uversky VN, van der Goot FG. Intrinsic Disorder in Transmembrane Proteins: Roles in Signaling and Topology Prediction. *PLOS ONE*. 2016 jul; 11(7):e0158594. , doi:10.1371/journal.pone.0158594.

Butler MS, Paterson DL. Antibiotics in the clinical pipeline in October 2019. *The Journal of Antibiotics*. 2020 jun; 73(6):329–364. , doi:10.1038/s41429-020-0291-8.

Caesar J, Reboul CF, Machello C, Kiesewetter S, Tang ML, Deme JC, Johnson S, Elmlund D, Lea SM, Elmlund H. SIMPLE 3.0. Stream single-particle cryo-EM analysis in real time. *Journal of Structural Biology: X*. 2020; 4(November):100040. , doi:10.1016/j.yjsbx.2020.100040.

Caffrey M. A comprehensive review of the lipid cubic phase or in meso method for crystallizing membrane and soluble proteins and complexes. *Acta Crystallographica Section F Structural Biology Communications*. 2015 jan; 71(1):3–18. , doi:10.1107/S2053230X14026843.

Caffrey M, Cherezov V. Crystallizing membrane proteins using lipidic mesophases. *Nature Protocols*. 2009 may; 4(5):706–731. , doi:10.1038/nprot.2009.31.

Camacho LR, Constant P, Raynaud C, Lanéelle MA, Triccas JA, Gicquel B, Daffé M, Guilhot C. Analysis of the Phthiocerol Dimycocerosate Locus of Mycobacterium tuberculosis. *Journal of Biological Chemistry*. 2001 jan; 276(23):19845–19854. , doi:10.1074/jbc.M100662200.

Cannac F, Qi C, Falschlunger J, Hausmann G, Basler K, Korkhov VM. Cryo-EM structure of the Hedgehog release protein Dispatched. *Science Advances*. 2020 apr; 6(16):eaay7928. , doi:10.1126/sciadv.aay7928.

Carel C, Nukdee K, Cantaloube S, Bonne M, Diagne CT, Laval F, Daffé M, Zerbib D. Mycobacterium tuberculosis Proteins Involved in Mycolic Acid Synthesis and Transport Localize Dynamically to the Old Growing Pole and Septum. *PLoS ONE*. 2014 may; 9(5):e97148. , doi:10.1371/journal.pone.0097148.

Carette X, Platig J, Young DC, Helmel M, Young AT, Wang Z, Potluri LP, Moody CS, Zeng J, Prisic S, Paulson JN, Muntel J, Madduri AVR, Velarde J, Mayfield JA, Locher C, Wang T, Quackenbush J, Rhee KY, Moody DB, et al. Multisystem Analysis of Mycobacterium tuberculosis Reveals Kinase-Dependent Remodeling of the Pathogen-Environment Interface. *mBio*. 2018 may; 9(2):1–21. , doi:10.1128/mBio.02333-17.

Chae PS, Rasmussen SGF, Rana RR, Gotfryd K, Chandra R, Goren MA, Kruse AC, Nurva S, Loland CJ, Pierre Y, Drew D, Popot JL, Picot D, Fox BG, Guan L, Gether U, Byrne B, Kobilka B, Gellman SH. Maltose–neopentyl glycol (MNG) amphiphiles for solubilization, stabilization and crystallization of membrane proteins. *Nature Methods*. 2010 dec; 7(12):1003–1008. , doi:10.1038/nmeth.1526.

Chakaya J, Khan M, Ntoumi F, Aklillu E, Fatima R, Mwaba P, Kapata N, Mfinanga S, Hasnain SE, Katoto PDMC, Bulabula ANH, Sam-Agudu NA, Nachega JB, Tiberi S, McHugh TD, Abubakar I, Zumla A. Global Tuberculosis Report 2020 – Reflections on the Global TB burden, treatment and prevention efforts. *International Journal of Infectious Diseases*. 2021 mar; (xxxx):4–9. , doi:10.1016/j.ijid.2021.02.107.

Chalut C. MmpL transporter-mediated export of cell-wall associated lipids and siderophores in mycobacteria. *Tuberculosis*. 2016 sep; 100:32–45. , doi:10.1016/j.tube.2016.06.004.

Cheng Y. Single-particle cryo-EM—How did it get here and where will it go. *Science*. 2018 aug; 361(6405):876–880. , doi:10.1126/science.aat4346.

Cheng Y, Grigorieff N, Penczek PA, Walz T. A Primer to Single-Particle Cryo-Electron Microscopy. *Cell*. 2015 apr; 161(3):438–449. , doi:10.1016/j.cell.2015.03.050.

Cherezov V, Clogston J, Papiz MZ, Caffrey M. Room to Move: Crystallizing Membrane Proteins in Swollen Lipidic Mesophases. *Journal of Molecular Biology*. 2006 apr; 357(5):1605–1618. , doi:10.1016/j.jmb.2006.01.049.

Chim N, Torres R, Liu Y, Capri J, Batot G, Whitelegge JP, Goulding CW. The Structure and Interactions of Periplasmic Domains of Crucial MmpL Membrane Proteins from Mycobacterium tuberculosis. *Chemistry & Biology*. 2015 aug; 22(8):1098–1107. , doi:10.1016/j.chembiol.2015.07.013.

Clamp M, Cuff J, Searle SM, Barton GJ. The Jalview Java alignment editor. *Bioinformatics*. 2004 feb; 20(3):426–427. , doi:10.1093/bioinformatics/btg430.

Cohen KA, Abeel T, Manson McGuire A, Desjardins CA, Munsamy V, Shea TP, Walker BJ, Bantubani N, Almeida DV, Alvarado L, Chapman SB, Mvelase NR, Duffy EY, Fitzgerald MG, Govender P, Gujja S, Hamilton S, Howarth C, Larimer JD, Maharaj K, et al. Evolution of Extensively Drug-Resistant Tuberculosis over Four Decades: Whole Genome Sequencing and Dating Analysis of Mycobacterium tuberculosis Isolates from KwaZulu-Natal. *PLOS Medicine*. 2015 sep; 12(9):e1001880. , doi:10.1371/journal.pmed.1001880.

Colangeli R, Jedrey H, Kim S, Connell R, Ma S, Chippada Venkata UD, Chakravorty S, Gupta A, Sizemore EE, Diem L, Sherman DR, Okwera A, Dietze R, Boom WH, Johnson JL, Mac Kenzie WR, Alland D. Bacterial Factors That Predict Relapse after Tuberculosis Therapy. *New England Journal of Medicine*. 2018 aug; 379(9):823–833. , doi:10.1056/NEJMoa1715849.

Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, et al. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature*. 1998 jun; 393(6685):537–544. , doi:10.1038/31159.

Cole ST. Transporter targeted in tuberculosis. *Nature Chemical Biology*. 2012 apr; 8(4):326–327. , doi:10.1038/nchembio.918.

Coleman JA, Yang D, Zhao Z, Wen PC, Yoshioka C, Tajkhorshid E, Gouaux E. Serotonin transporter–ibogaine complexes illuminate mechanisms of inhibition and transport. *Nature*. 2019 may; 569(7754):141–145. , doi:10.1038/s41586-019-1135-1.

Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, Parkhill J, Malla B, Berg S, Thwaites G, Yeboah-Manu D, Bothamley G, Mei J, Wei L, Bentley S, Harris SR, Niemann S, Diel R, Aseffa A, Gao Q, et al. Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. *Nature Genetics*. 2013 oct; 45(10):1176–1182. , doi:10.1038/ng.2744.

Concepcion J, Witte K, Wartchow C, Choo S, Yao D, Persson H, Wei J, Li P, Heidecker B, Ma W, Varma R, Zhao LS, Perillat D, Carricato G, Recknor M, Du K, Ho H, Ellis T, Gamez J, Howes M, et al. Label-Free Detection of Biomolecular Interactions Using BioLayer Interferometry for Kinetic Characterization. *Combinatorial Chemistry & High Throughput Screening*. 2009 sep; 12(8):791–800. , doi:10.2174/138620709789104915.

Converse SE, Mougous JD, Leavell MD, Leary JA, Bertozzi CR, Cox JS. MmpL8 is required for sulfolipid-1 biosynthesis and Mycobacterium tuberculosis virulence. *Proceedings of the National Academy of Sciences*. 2003 may; 100(10):6121–6126. , doi:10.1073/pnas.1030024100.

Cox JS, Chen B, McNeil M, Jacobs WR. Complex lipid determines tissue-specific replication of Mycobacterium tuberculosis in mice. *Nature*. 1999 nov; 402(6757):79–83. , doi:10.1038/47042.

Daffé M, Quémard A, Marrakchi H. Mycolic Acids: From Chemistry to Biology. In: Geiger O, editor. *Biogenesis of Fatty Acids, Lipids and Membranes* Springer International Publishing; 2017.p. 1–36. doi:10.1007/978-3-319-43676-0.

Dal Molin M, Selchow P, Schäfle D, Tschumi A, Ryckmans T, Laage-Witt S, Sander P. Identification of novel scaffolds targeting Mycobacterium tuberculosis. *Journal of Molecular Medicine*. 2019 nov; 97(11):1601–1613. , doi:10.1007/s00109-019-01840-7.

Danev R, Buijsse B, Khoshouei M, Plitzko JM, Baumeister W. Volta potential phase plate for in-focus phase contrast transmission electron microscopy. *Proceedings of the National Academy of Sciences*. 2014 nov; 111(44):15635–15640. , doi:10.1073/pnas.1418377111.

Daniel H, Spanier B, Kottra G, Weitz D. From Bacteria to Man: Archaic Proton-Dependent Peptide Transporters at Work. *Physiology*. 2006 apr; 21(2):93–102. , doi:10.1152/physiol.00054.2005.

De Groote MA, Jarvis TC, Wong C, Graham J, Hoang T, Young CL, Ribble W, Day J, Li W, Jackson M, Gonzalez-Juarrero M, Sun X, Ochsner UA. Optimization and Lead Selection of Benzothiazole Amide Analogs Toward a Novel Antimycobacterial Agent. *Frontiers in Microbiology*. 2018 sep; 9(SEP):1–10. , doi:10.3389/fmicb.2018.02231.

Degiacomi G, Belardinelli JM, Pasca MR, De Rossi E, Riccardi G, Chiarelli LR. Promiscuous Targets for Antitubercular Drug Discovery: The Paradigm of DprE1 and MmpL3. *Applied Sciences*. 2020 jan; 10(2):623. , doi:10.3390/app10020623.

Degiacomi G, Benjak A, Madacki J, Boldrin F, Provvedi R, Palù G, Kordulakova J, Cole ST, Manganelli R. Essentiality of mmpL3 and impact of its silencing on Mycobacterium tuberculosis gene expression. *Scientific Reports*. 2017 apr; 7(1):43495. , doi:10.1038/srep43495.

DeJesus MA, Gerrick ER, Xu W, Park SW, Long JE, Boutte CC, Rubin EJ, Schnappinger D, Ehrt S, Fortune SM, Sassetti CM, Ioerger TR. Comprehensive Essentiality Analysis of the Mycobacterium tuberculosis Genome via Saturating Transposon Mutagenesis. *mBio*. 2017 mar; 8(1):1–17. , doi:10.1128/mBio.02133-16.

Delmar JA, Bolla JR, Su CC, Yu EW. Crystallization of Membrane Proteins by Vapor Diffusion. In: Shukla AK, editor. *Methods in Enzymology*, vol. 557, 1 ed. Elsevier Inc.; 2015.p. 363–392. , doi:10.1016/bs.mie.2014.12.018.

Denisov IG, Grinkova YV, Lazarides AA, Sligar SG. Directed Self-Assembly of Monodisperse Phospholipid Bilayer Nanodiscs with Controlled Size. *Journal of the American Chemical Society*. 2004 mar; 126(11):3477–3487. , doi:10.1021/ja0393574.

Denisov IG, Sligar SG. Nanodiscs for structural and functional studies of membrane proteins. *Nature Structural & Molecular Biology*. 2016 jun; 23(6):481–486. , doi:10.1038/nsmb.3195.

Denisov IG, Sligar SG. Nanodiscs in Membrane Biochemistry and Biophysics. *Chemical Reviews*. 2017; 117(6):4669–4713. doi:10.1021/acs.chemrev.6b00690.

Desmyter A, Spinelli S, Roussel A, Cambillau C. Camelid nanobodies: killing two birds with one stone. *Current Opinion in Structural Biology*. 2015 jun; 32:1–8. , doi:10.1016/j.sbi.2015.01.001.

D'Imprima E, Floris D, Joppe M, Sánchez R, Grininger M, Kühlbrandt W. Protein denaturation at the air-water interface and how to prevent it. *eLife*. 2019 apr; 8:1–18. , doi:10.7554/eLife.42747.

Dobrin N, Randall NW. New and Emerging Proteomic Techniques, vol. 328. Dobrin N, Randall NW, editors, New Jersey: Humana Press; 2006. , doi:10.1385/159745026X.

Domenech P, Reed MB, Barry CE. Contribution of the Mycobacterium tuberculosis MmpL Protein Family to Virulence and Drug Resistance. *Infection and Immunity*. 2005 jun; 73(6):3492–3501. , doi:10.1128/IAI.73.6.3492-3501.2005.

Domenech P, Reed MB, Dowd CS, Manca C, Kaplan G, Barry CE. The Role of MmpL8 in Sulfatide Biogenesis and Virulence of Mycobacterium tuberculosis. *Journal of Biological Chemistry*. 2004 may; 279(20):21257–21265. , doi:10.1074/jbc.M400324200.

Dosztányi Z. Prediction of protein disorder based on IUPred. *Protein Science*. 2018 jan; 27(1):331–340. , doi:10.1002/pro.3334.

Drage MG, Tsai HC, Pecora ND, Cheng TY, Arida AR, Shukla S, Rojas RE, Seshadri C, Moody DB, Boom WH, Sacchettini JC, Harding CV. Mycobacterium tuberculosis lipoprotein LprG (Rv1411c) binds triacylated glycolipid agonists of Toll-like receptor 2. *Nature Structural & Molecular Biology*. 2010 sep; 17(9):1088–1095. , doi:10.1038/nsmb.1869.

Drew D, Slotboom DJ, Friso G, Reda T, Genevaux P, Rapp M, Meindl-Beinker NM, Lambert W, Lerch M, Daley DO, Van Wijk KJ, Hirst J, Kunji E, De Gier JW. A scalable, GFP-based pipeline for membrane protein overexpression screening and purification. *Protein Science*. 2005 aug; 14(8):2011–2017. , doi:10.1110/ps.051466205.

Drew DE, von Heijne G, Nordlund P, de Gier JWL. Green fluorescent protein as an indicator to monitor membrane protein overexpression in Escherichia coli. *FEBS Letters*. 2001 oct; 507(2):220–224. , doi:10.1016/S0014-5793(01)02980-5.

Du D, Wang-Kan X, Neuberger A, van Veen HW, Pos KM, Piddock LJV, Luisi BF. Multidrug efflux pumps: structure, function and regulation. *Nature Reviews Microbiology*. 2018 sep; 16(9):523–539. , doi:10.1038/s41579-018-0048-6.

Dumoulin M, Conrath K, Van Meirhaeghe A, Meersman F, Heremans K, Frenken LGJ, Muyldermans S, Wyns L, Matagne A. Single-domain antibody fragments with high conformational stability. *Protein Science*. 2009 apr; 11(3):500–515. , doi:10.1110/ps.34602.

Dupont C, Chen Y, Xu Z, Roquet-Banères F, Blaise M, Witt AK, Dubar F, Biot C, Guérardel Y, Maurer FP, Chng SS, Kremer L. A piperidinol-containing molecule is active against Mycobacterium tuberculosis by inhibiting the mycolic acid flippase activity of MmpL3. *Journal of Biological Chemistry*. 2019 nov; 294(46):17512–17523. , doi:10.1074/jbc.RA119.010135.

Dupont C, Viljoen A, Dubar F, Blaise M, Bernut A, Pawlik A, Bouchier C, Brosch R, Guérardel Y, Lelièvre J, Ballell L, Herrmann JL, Biot C, Kremer L. A new piperidinol derivative targeting mycolic acid transport in Mycobacterium abscessus. *Molecular Microbiology*. 2016 aug; 101(3):515–529. , doi:10.1111/mmi.13406.

Efremov RG, Gatsogiannis C, Raunser S. Lipid Nanodiscs as a Tool for High-Resolution Structure Determination of Membrane Proteins by Single-Particle Cryo-EM. In: Ziegler C, editor. *Methods in Enzymology: A Structure-Function Toolbox for Membrane Transporter and Channels*, vol. 594, 1 ed. Elsevier Inc.; 2017.p. 1–30. , doi:10.1016/bs.mie.2017.05.007.

Ehlers S, Schaible UE. The Granuloma in Tuberculosis: Dynamics of a Host–Pathogen Collusion. *Frontiers in Immunology*. 2013; 3(JAN):1–9. , doi:10.3389/fimmu.2012.00411.

Eicher T, Seeger MA, Anselmi C, Zhou W, Brandstätter L, Verrey F, Diederichs K, Faraldo-Gómez JD, Pos KM. Coupling of remote alternating-access transport mechanisms for protons and substrates in the multidrug efflux pump AcrB. *eLife*. 2014 sep; 3:1–26. , doi:10.7554/eLife.03145.

Eldholm V, Balloux F. Antimicrobial Resistance in Mycobacterium tuberculosis : The Odd One Out. *Trends in Microbiology*. 2016 aug; 24(8):637–648. , doi:10.1016/j.tim.2016.03.007.

Eldholm V, Monteserin J, Rieux A, Lopez B, Sobkowiak B, Ritacco V, Balloux F. Four decades of transmission of a multidrug-resistant Mycobacterium tuberculosis outbreak strain. *Nature Communications*. 2015 nov; 6(1):7119. , doi:10.1038/ncomms8119.

Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallographica Section D Biological Crystallography*. 2010 apr; 66(4):486–501. , doi:10.1107/S0907444910007493.

Fan X, Wang J, Zhang X, Yang Z, Zhang JC, Zhao L, Peng HL, Lei J, Wang HW. Single particle cryo-EM reconstruction of 52 kDa streptavidin at 3.2 Angstrom resolution. *Nature Communications*. 2019 dec; 10(1):2386. , doi:10.1038/s41467-019-10368-w.

Fay A, Czudnochowski N, Rock JM, Johnson JR, Krogan NJ, Rosenberg O, Glickman MS. Two Accessory Proteins Govern MmpL3 Mycolic Acid Transport in Mycobacteria. *mBio*. 2019 jun; 10(3):1–17. , doi:10.1128/mBio.00850-19.

Fellouse FA, Li B, Compaan DM, Peden AA, Hymowitz SG, Sidhu SS. Molecular Recognition by a Binary Code. *Journal of Molecular Biology*. 2005 may; 348(5):1153–1162. , doi:10.1016/j.jmb.2005.03.041.

Fortuin S, Tomazella GG, Nagaraj N, Sampson SL, Gey van Pittius NC, Soares NC, Wiker HG, de Souza GA, Warren RM. Phosphoproteomics analysis of a clinical Mycobacterium tuberculosis Beijing isolate: expanding the mycobacterial phosphoproteome catalog. *Frontiers in Microbiology*. 2015 feb; 6(FEB):1–12. , doi:10.3389/fmicb.2015.00006.

Foss MH, Pou S, Davidson PM, Dunaj JL, Winter RW, Pou S, Licon MH, Doh JK, Li Y, Kelly JX, Dodean RA, Koop DR, Riscoe MK, Purdy GE. Diphenylether-Modified 1,2-Diamines with Improved Drug Properties for Development against Mycobacterium tuberculosis. *ACS Infectious Diseases*. 2016 jul; 2(7):500–508. , doi:10.1021/acsinfecdis.6b00052.

Franke S, Grass G, Rensing C, Nies DH. Molecular Analysis of the Copper-Transporting Efflux System CusCFBA of Escherichia coli. *Journal of Bacteriology*. 2003 jul; 185(13):3804–3812. , doi:10.1128/JB.185.13.3804-3812.2003.

Fujita Y, Naka T, Doi T, Yano I. Direct molecular mass determination of trehalose monomycolate from 11 species of mycobacteria by MALDI-TOF mass spectrometry. *Microbiology*. 2005 may; 151(5):1443–1452. , doi:10.1099/mic.0.27791-0.

Furukawa A, Nakayama S, Yoshikaie K, Tanaka Y, Tsukazaki T. Remote Coupled Drastic β-Barrel to β-Sheet Transition of the Protein Translocation Motor. *Structure*. 2018 mar; 26(3):485–489.e2. , doi:10.1016/j.str.2018.01.002.

Gao Y, Cao E, Julius D, Cheng Y. TRPV1 structures in nanodiscs reveal mechanisms of ligand and lipid action. *Nature*. 2016 jun; 534(7607):347–351. , doi:10.1038/nature17964.

Garima K, Pathak R, Tandon R, Rathor N, Sinha R, Bose M, Varma-Basil M. Differential expression of efflux pump genes of Mycobacterium tuberculosis in response to varied subinhibitory concentrations of antituberculosis agents. *Tuberculosis*. 2015 mar; 95(2):155–161. , doi:10.1016/j.tube.2015.01.005.

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM, editor. *The Proteomics Protocols Handbook* Totowa, NJ: Humana Press; 2005.p. 571–608. , doi:10.1385/1592598900.

Geertsma ER, Dutzler R. A Versatile and Efficient High-Throughput Cloning Tool for Structural Biology. *Biochemistry*. 2011 apr; 50(15):3272–3278. , doi:10.1021/bi200178z.

Gengenbacher M, Kaufmann SHE. Mycobacterium tuberculosis: success through dormancy. *FEMS Microbiology Reviews*. 2012 may; 36(3):514–532. , doi:10.1111/j.1574-6976.2012.00331.x.

Goldman RC. Why are membrane targets discovered by phenotypic screens and genome sequencing in Mycobacterium tuberculosis? *Tuberculosis*. 2013 nov; 93(6):569–588. , doi:10.1016/j.tube.2013.09.003.

Gong X, Qian H, Cao P, Zhao X, Zhou Q, Lei J, Yan N. Structural basis for the recognition of Sonic Hedgehog by human Patched1. *Science*. 2018 aug; 361(6402):eaas8935. , doi:10.1126/science.aas8935.

Gong X, Qian H, Zhou X, Wu J, Wan T, Cao P, Huang W, Zhao X, Wang X, Wang P, Shi Y, Gao GF, Zhou Q, Yan N. Structural Insights into the Niemann-Pick C1 (NPC1)-Mediated Cholesterol Transfer and Ebola Infection. *Cell*. 2016 jun; 165(6):1467–1478. , doi:10.1016/j.cell.2016.05.022.

Goodfellow M, Jones AL. Corynebacteriales ord. nov. In: *Bergey's Manual of Systematics of Archaea and Bacteria* Wiley; 2015.p. 1–14. , doi:10.1002/9781118960608.obm00009.

Goodsell DS, Autin L, Olson AJ. Illustrate: Software for Biomolecular Illustration. *Structure*. 2019 nov; 27(11):1716–1720.e1. , doi:10.1016/j.str.2019.08.011.

Greenfield NJ. Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols*. 2006 dec; 1(6):2876–2890. , doi:10.1038/nprot.2006.202.

Grinkova YV, Denisov IG, Sligar SG. Engineering extended membrane scaffold proteins for self-assembly of soluble nanoscale lipid bilayers. *Protein Engineering, Design and Selection*. 2010; 23(11):843–848. doi:10.1093/protein/gzq060.

Groenewald W, Parra-Cruz RA, Jäger CM, Croft AK. Revealing solvent-dependent folding behavior of mycolic acids from Mycobacterium tuberculosis by advanced simulation analysis. *Journal of Molecular Modeling*. 2019 mar; 25(3):68. , doi:10.1007/s00894-019-3943-5.

Grover S, Engelhart CA, Pérez-Herrán E, Li W, Abrahams KA, Papavinasasundaram K, Bean JM, Sassetti CM, Mendoza-Losana A, Besra GS, Jackson M, Schnappinger D. Two-Way Regulation of MmpL3 Expression Identifies and Validates Inhibitors of MmpL3 Function in Mycobacterium tuberculosis. *ACS Infectious Diseases*. 2021 jan; 7(1):141–152. , doi:10.1021/acsinfecdis.0c00675.

Grzegorzewicz AE, Pham H, Gundi VAKB, Scherman MS, North EJ, Hess T, Jones V, Gruppo V, Born SEM, Korduláková J, Chavadi SS, Morisseau C, Lenaerts AJ, Lee RE, McNeil MR, Jackson M. Inhibition of mycolic acid transport across the Mycobacterium tuberculosis plasma membrane. *Nature Chemical Biology*. 2012 apr; 8(4):334–341. , doi:10.1038/nchembio.794.

Guan L, Nakae T. Identification of Essential Charged Residues in Transmembrane Segments of the Multidrug Transporter MexB of Pseudomonas aeruginosa. *Journal of Bacteriology*. 2001 mar; 183(5):1734–1739. , doi:10.1128/JB.183.5.1734-1739.2001.

Guerra-Assunção JA, Houben RMGJ, Crampin AC, Mzembe T, Mallard K, Coll F, Khan P, Banda L, Chiwaya A, Pereira RPA, McNerney R, Harris D, Parkhill J, Clark TG, Glynn JR. Recurrence due to Relapse or Reinfection With Mycobacterium tuberculosis : A Whole-Genome Sequencing Approach in a Large, Population-Based Cohort With a High HIV Infection Prevalence and Active Follow-up. *Journal of Infectious Diseases*. 2015 apr; 211(7):1154–1163. , doi:10.1093/infdis/jiu574.

Guinn KM, Rubin EJ. Tuberculosis: Just the FAQs. *mBio*. 2017 dec; 8(6):1–14. , doi:10.1128/mBio.01910-17.

Gupta K, Donlan JAC, Hopper JTS, Uzdavinys P, Landreh M, Struwe WB, Drew D, Baldwin AJ, Stansfeld PJ, Robinson CV. The role of interfacial lipids in stabilizing membrane protein oligomers. *Nature*. 2017 jan; 541(7637):421–424. , doi:10.1038/nature20820.

Hagn F, Etzkorn M, Raschle T, Wagner G. Optimized Phospholipid Bilayer Nanodiscs Facilitate High-Resolution Structure Determination of Membrane Proteins. *Journal of the American Chemical Society*. 2013 feb; 135(5):1919–1925. , doi:10.1021/ja310901f.

Hagn F, Nasr ML, Wagner G. Assembly of phospholipid nanodiscs of controlled size for structural studies of membrane proteins by NMR. *Nature Protocols*. 2018 jan; 13(1):79–98. , doi:10.1038/nprot.2017.094.

Halloum I, Viljoen A, Khanna V, Craig D, Bouchier C, Brosch R, Coxon G, Kremer L. Resistance to Thiacetazone Derivatives Active against Mycobacterium abscessus Involves Mutations in the MmpL5 Transcriptional Repressor MAB_4384. *Antimicrobial Agents and Chemotherapy*. 2017 apr; 61(4):1–14. , doi:10.1128/AAC.02509-16.

Hamers-Casterman C, Atarhouch T, Muyldermans S, Robinson G, Hammers C, Songa EB, Bendahman N, Hammers R. Naturally occurring antibodies devoid of light chains. *Nature*. 1993 jun; 363(6428):446–448. , doi:10.1038/363446a0.

Hansen SB, Laursen NS, Andersen GR, Andersen KR. Introducing site-specific cysteines into nanobodies for mercury labelling allows de novo phasing of their crystal structures. *Acta Crystallographica Section D Structural Biology*. 2017 oct; 73(10):804–813. , doi:10.1107/S2059798317013171.

Harmansa S, Affolter M. Protein binders and their applications in developmental biology. *Development*. 2018 jan; 145(2). , doi:10.1242/dev.148874.

Hartkoorn RC, Uplekar S, Cole ST. Cross-Resistance between Clofazimine and Bedaquiline through Upregulation of MmpL5 in Mycobacterium tuberculosis. *Antimicrobial Agents and Chemotherapy*. 2014 may; 58(5):2979–2981. , doi:10.1128/AAC.00037-14.

He Z, Zhang H, Gao S, Lercher MJ, Chen WH, Hu S. Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Research*. 2016 jul; 44(W1):W236–W241. , doi:10.1093/nar/gkw370.

Henderson R. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Quarterly Reviews of Biophysics*. 1995 may; 28(2):171–193. , doi:10.1017/S003358350000305X.

Henry KA, MacKenzie CR. Antigen recognition by single-domain antibodies: structural latitudes and constraints. *mAbs*. 2018 aug; 10(6):815–826. , doi:10.1080/19420862.2018.1489633.

Herzik MA, Wu M, Lander GC. High-resolution structure determination of sub-100 kDa complexes using conventional cryo-EM. *Nature Communications*. 2019 dec; 10(1):1032. , doi:10.1038/s41467-019-08991-8.

Hoffmann C, Leis A, Niederweis M, Plitzko JM, Engelhardt H. Disclosure of the mycobacterial outer membrane: Cryo-electron tomography and vitreous sections reveal the lipid bilayer structure. *Proceedings of the National Academy of Sciences*. 2008 mar; 105(10):3963–3967. , doi:10.1073/pnas.0709530105.

Hofmann S, Januliene D, Mehdipour AR, Thomas C, Stefan E, Brüchert S, Kuhn BT, Geertsma ER, Hummer G, Tampé R, Moeller A. Conformation space of a heterodimeric ABC exporter under turnover conditions. *Nature*. 2019 jul; 571(7766):580–583. , doi:10.1038/s41586-019-1391-0.

Hohl M, Briand C, Grütter MG, Seeger MA. Crystal structure of a heterodimeric ABC transporter in its inward-facing conformation. *Nature Structural & Molecular Biology*. 2012 apr; 19(4):395–402. , doi:10.1038/nsmb.2267.

Houben RMGJ, Dodd PJ. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. *PLOS Medicine*. 2016 oct; 13(10):e1002152. , doi:10.1371/journal.pmed.1002152.

Huang CY, Olieric V, Howe N, Warshamanage R, Weinert T, Panepucci E, Vogeley L, Basu S, Diederichs K, Caffrey M, Wang M. In situ serial crystallography for rapid de novo membrane protein structure determination. *Communications Biology*. 2018 dec; 1(1):124. , doi:10.1038/s42003-018-0123-6.

Huang CY, Olieric V, Ma P, Howe N, Vogeley L, Liu X, Warshamanage R, Weinert T, Panepucci E, Kobilka B, Diederichs K, Wang M, Caffrey M. In meso in situ serial X-ray crystallography of soluble and membrane proteins at cryogenic temperatures. *Acta Crystallographica Section D Structural Biology*. 2016 jan; 72(1):93–112. , doi:10.1107/S2059798315021683.

Huang CY, Olieric V, Ma P, Panepucci E, Diederichs K, Wang M, Caffrey M. In meso in situ serial X-ray crystallography of soluble and membrane proteins. *Acta Crystallographica Section D: Biological Crystallography*. 2015; 71:1238–1256. doi:10.1107/S1399004715005210.

Huang CS, Yu X, Fordstrom P, Choi K, Chung BC, Roh SH, Chiu W, Zhou M, Min X, Wang Z. Cryo-EM structures of NPC1L1 reveal mechanisms of cholesterol transport and ezetimibe inhibition. *Science Advances*. 2020 jun; 6(25):eabb1989. , doi:10.1126/sciadv.abb1989.

Huang W, Briffotaux J, Wang X, Liu L, Hao P, Cimino M, Buchieri MV, Namouchi A, Ainsa JA, Gicquel B. Ionophore A23187 shows anti-tuberculosis activity and synergy with tebipenem. *Tuberculosis*. 2017 dec; 107:111–118. , doi:10.1016/j.tube.2017.09.001.

Hunt M, Bradley P, Lapierre SG, Heys S, Thomsit M, Hall MB, Malone KM, Wintringer P, Walker TM, Cirillo DM, Comas I, Farhat MR, Fowler P, Gardy J, Ismail N, Kohl TA, Mathys V, Merker M, Niemann S, Omar SV, et al. Antibiotic resistance prediction for Mycobacterium tuberculosis from genome sequence data with Mykrobe. *Wellcome Open Research*. 2019 dec; 4:191. , doi:10.12688/wellcomeopenres.15603.1.

Hunter RL, Olsen MR, Jagannath C, Actor JK. Multiple roles of cord factor in the pathogenesis of primary, secondary, and cavitary tuberculosis, including a revised description of the pathology of secondary disease. *Annals of Clinical and Laboratory Science*. 2006; 36(4):371–386.

Hutter CAJ, Timachi MH, Hürlimann LM, Zimmermann I, Egloff P, Göddeke H, Kucher S, Štefanić S, Karttunen M, Schäfer LV, Bordignon E, Seeger MA. The extracellular gate shapes the energy profile of an ABC exporter. *Nature Communications*. 2019 dec; 10(1):2260. , doi:10.1038/s41467-019-09892-6.

Ingram JR, Schmidt FI, Ploegh HL. Exploiting Nanobodies' Singular Traits. *Annual Review of Immunology*. 2018 apr; 36(1):695–715. , doi:10.1146/annurev-immunol-042617-053327.

Ioerger TR, O'Malley T, Liao R, Guinn KM, Hickey MJ, Mohaideen N, Murphy KC, Boshoff HIM, Mizrahi V, Rubin EJ, Sassetti CM, Barry CE, Sherman DR, Parish T, Sacchettini JC. Identification of New Drug Targets and Resistance Mechanisms in Mycobacterium tuberculosis. *PLoS ONE*. 2013 sep; 8(9):e75245. , doi:10.1371/journal.pone.0075245.

Jackson M. The Mycobacterial Cell Envelope - Lipids. *CSH Perspectives in Medicine*. 2014; p. 1–22. , doi:10.1086/592651.

Jackson M, Stevens CM, Zhang L, Zgurskaya HI, Niederweis M. Transporters Involved in the Biogenesis and Functionalization of the Mycobacterial Cell Envelope. *Chemical Reviews*. 2021 may; 121(9):5124–5157. , doi:10.1021/acs.chemrev.0c00869.

Jain M, Cox JS. Interaction between Polyketide Synthase and Transporter Suggests Coupled Synthesis and Export of Virulence Lipid in M. tuberculosis. *PLoS Pathogens*. 2005 sep; 1(1):e2. , doi:10.1371/journal.ppat.0010002.

Jankute M, Cox JAG, Harrison J, Besra GS. Assembly of the Mycobacterial Cell Wall. *Annual Review of Microbiology*. 2015 oct; 69(1):405–423. , doi:10.1146/annurev-micro-091014-104121.

Jojoa-Cruz S, Saotome K, Murthy SE, Tsui CCA, Sansom MSP, Patapoutian A, Ward AB. Cryo-EM structure of the mechanically activated ion channel OSCA1.2. *eLife*. 2018 nov; 7:1–17. , doi:10.7554/eLife.41845.

Jones DT. Improving the accuracy of transmembrane protein topology prediction using evolutionary information. *Bioinformatics*. 2007 mar; 23(5):538–544. , doi:10.1093/bioinformatics/btl677.

Jurcik A, Bednar D, Byska J, Marques SM, Furmanova K, Daniel L, Kokkonen P, Brezovsky J, Strnad O, Stourac J, Pavelka A, Manak M, Damborsky J, Kozlikova B. CAVER Analyst 2.0: analysis and visualization of channels and tunnels in protein structures and molecular dynamics trajectories. *Bioinformatics*. 2018 oct; 34(20):3586–3588. , doi:10.1093/bioinformatics/bty386.

Kalscheuer R, Palacios A, Anso I, Cifuente J, Anguita J, Jacobs WR, Guerin ME, Prados-Rosales R. The Mycobacterium tuberculosis capsule: a cell structure with key implications in pathogenesis. *Biochemical Journal*. 2019 jul; 476(14):1995–2016. , doi:10.1042/BCJ20190324.

Kastner B, Fischer N, Golas MM, Sander B, Dube P, Boehringer D, Hartmuth K, Deckert J, Hauer F, Wolf E, Uchtenhagen H, Urlaub H, Herzog F, Peters JM, Poerschke D, Lührmann R, Stark H. GraFix: sample preparation for single-particle electron cryomicroscopy. *Nature Methods*. 2008 jan; 5(1):53–55. , doi:10.1038/nmeth1139.

Kavunja HW, Biegas KJ, Banahene N, Stewart JA, Piligian BF, Groenevelt JM, Sein CE, Morita YS, Niederweis M, Siegrist MS, Swarts BM. Photoactivatable Glycolipid Probes for Identifying Mycolate–Protein Interactions in Live Mycobacteria. *Journal of the American Chemical Society*. 2020 apr; 142(17):7725–7731. , doi:10.1021/jacs.0c01065.

Kawate T, Gouaux E. Fluorescence-Detection Size-Exclusion Chromatography for Precrystallization Screening of Integral Membrane Proteins. *Structure*. 2006 apr; 14(4):673–681. , doi:10.1016/j.str.2006.01.013.

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

Kelly RL, Le D, Zhao J, Wittrup KD. Reduction of Nonspecificity Motifs in Synthetic Antibody Libraries. *Journal of Molecular Biology*. 2018 jan; 430(1):119–130. , doi:10.1016/j.jmb.2017.11.008.

Kern DM, Sorum B, Hoel CM, Sridharan S, Remis JP, Toso DB, Brohawn SG. Cryo-EM structure of the SARS-CoV-2 3a ion channel in lipid nanodiscs. *bioRxiv : the preprint server for biology*. 2020; doi:10.1101/2020.06.17.156554.

Khoshouei M, Radjainia M, Baumeister W, Danev R. Cryo-EM structure of haemoglobin at 3.2 Å determined with the Volta phase plate. *Nature Communications*. 2017 dec; 8(1):16099. , doi:10.1038/ncomms16099.

Kim J, Tan YZ, Wicht KJ, Erramilli SK, Dhingra SK, Okombo J, Vendome J, Hagenah LM, Giacometti SI, Warren AL, Nosol K, Roepe PD, Potter CS, Carragher B, Kossiakoff AA, Quick M, Fidock DA, Mancia F. Structure and drug resistance of the Plasmodium falciparum transporter PfCRT. *Nature*. 2019 dec; 576(7786):315–320. , doi:10.1038/s41586-019-1795-x.

Kirchhofer A, Helma J, Schmidthals K, Frauer C, Cui S, Karcher A, Pellis M, Muyldermans S, Casas-Delucchi CS, Cardoso MC, Leonhardt H, Hopfner KP, Rothbauer U. Modulation of protein properties in living cells using nanobodies. *Nature Structural & Molecular Biology*. 2010 jan; 17(1):133–138. , doi:10.1038/nsmb.1727.

Klenotic PA, Moseng MA, Morgan CE, Yu EW. Structural and Functional Diversity of Resistance–Nodulation–Cell Division Transporters. *Chemical Reviews*. 2021 may; 121(9):5378–5416. , doi:10.1021/acs.chemrev.0c00621.

Kobylka J, Kuth MS, Müller RT, Geertsma ER, Pos KM. AcrB: a mean, keen, drug efflux machine. *Annals of the New York Academy of Sciences*. 2020 jan; 1459(1):38–68. , doi:10.1111/nyas.14239.

Koch R. Die Aetiologie der Tuberculose. *Berliner Klinische Wochenschrift*. 1882; 19:221–230. doi:10.25646/5088.

Koide S. Engineering of recombinant crystallization chaperones. *Current Opinion in Structural Biology*. 2009 aug; 19(4):449–457. , doi:10.1016/j.sbi.2009.04.008.

Koide S, Sidhu SS. The Importance of Being Tyrosine: Lessons in Molecular Recognition from Minimalist Synthetic Binding Proteins. *ACS Chemical Biology*. 2009 may; 4(5):325–334. , doi:10.1021/cb800314v.

Korycka-Machała M, Viljoen A, Pawełczyk J, Borówka P, Dziadek B, Gobis K, Brzostek A, Kawka M, Blaise M, Strapagiel D, Kremer L, Dziadek J. 1H-Benzo[d]Imidazole Derivatives Affect MmpL3 in Mycobacterium tuberculosis. *Antimicrobial Agents and Chemotherapy*. 2019 oct; 63(10). , doi:10.1128/AAC.00441-19.

Koshy C, Schweikhard ES, Gärtner RM, Perez C, Yildiz Ö, Ziegler C. Structural evidence for functional lipid interactions in the betaine transporter BetP. *The EMBO Journal*. 2013 nov; 32(23):3096–3105. , doi:10.1038/emboj.2013.226.

Kozikowski AP, Onajole OK, Stec J, Dupont C, Viljoen A, Richard M, Chaira T, Lun S, Bishai W, Raj VS, Ordway D, Kremer L. Targeting Mycolic Acid Transport by Indole-2-carboxamides for the Treatment of Mycobacterium abscessus Infections. *Journal of Medicinal Chemistry*. 2017 jul; 60(13):5876–5888. , doi:10.1021/acs.jmedchem.7b00582.

Kraft TE, Hresko RC, Hruz PW. Expression, purification, and functional characterization of the insulin-responsive facilitative glucose transporter GLUT4. *Protein Science*. 2015 dec; 24(12):2008–2019. , doi:10.1002/pro.2812.

Krubasik P, Kobayashi M, Sandmann G. Expression and functional analysis of a gene cluster involved in the synthesis of decaprenoxanthin reveals the mechanisms for C50 carotenoid formation. *European Journal of Biochemistry*. 2001 jul; 268(13):3702–3708. , doi:10.1046/j.1432-1327.2001.02275.x.

Kuhn BT, Zimmermann I, Egloff P, Hürlimann LM, Hutter CAJ, Miscenic C, Dawson RJP, Seeger MA, Geertsma ER. Biotinylation of Membrane Proteins for Binder Selections. In: Perez C, Maier T, editors. *Expression , Purification, and Structural Biology of Membrane Proteins, Methods in Molecular Biology* Springer US; 2020.p. 151–165. , doi:10.1007/978-1-0716-0373-4_11.

Kumar N, Su CC, Chou TH, Radhakrishnan A, Delmar JA, Rajashankar KR, Yu EW. Crystal structures of the Burkholderia multivorans hopanoid transporter HpnN. *Proceedings of the National Academy of Sciences*. 2017 jun; 114(25):6557–6562. , doi:10.1073/pnas.1619660114.

La Rosa V, Poce G, Canseco JO, Buroni S, Pasca MR, Biava M, Raju RM, Porretta GC, Alfonso S, Battilocchio C, Javid B, Sorrentino F, Ioerger TR, Sacchettini JC, Manetti F, Botta M, De Logu A, Rubin EJ, De Rossi E. MmpL3 Is the Cellular Target of the Antitubercular Pyrrole Derivative BM212. *Antimicrobial Agents and Chemotherapy*. 2012 jan; 56(1):324–331. , doi:10.1128/AAC.05270-11.

Lamichhane G, Tyagi S, Bishai WR. Designer Arrays for Defined Mutant Analysis To Detect Genes Essential for Survival of Mycobacterium tuberculosis in Mouse Lungs. *Infection and Immunity*. 2005 apr; 73(4):2533–2540. , doi:10.1128/IAI.73.4.2533-2540.2005.

Landau EM, Rosenbusch JP. Lipidic cubic phases: A novel concept for the crystallization of membrane proteins. *Proceedings of the National Academy of Sciences*. 1996 dec; 93(25):14532–14535. , doi:10.1073/pnas.93.25.14532.

Laverty D, Desai R, Uchański T, Masiulis S, Stec WJ, Malinauskas T, Zivanov J, Pardon E, Steyaert J, Miller KW, Aricescu AR. Cryo-EM structure of the human $\alpha 1 \beta 3 \gamma 2$ GABAA receptor in a lipid bilayer. *Nature*. 2019 jan; 565(7740):516–520. , doi:10.1038/s41586-018-0833-4.

Le NH, Locard-Paulet M, Stella A, Tomas N, Molle V, Burlet-Schiltz O, Daffé M, Marrakchi H. The protein kinase PknB negatively regulates biosynthesis and trafficking of mycolic acids in mycobacteria. *Journal of Lipid Research*. 2020 aug; 61(8):1180–1191. , doi:10.1194/jlr.RA120000747.

Le Bon C, Marconnet A, Masscheleyn S, Popot JL, Zoonens M. Folding and stabilizing membrane proteins in amphipol A8-35. *Methods*. 2018 sep; 147(December 2017):95–105. , doi:10.1016/j.ymeth.2018.04.012.

Lee AG. How lipids affect the activities of integral membrane proteins. *Biochimica et Biophysica Acta (BBA) - Biomembranes*. 2004 nov; 1666(1-2):62–87. , doi:10.1016/j.bbamem.2004.05.012.

Li D, Howe N, Dukkipati A, Shah STA, Bax BD, Edge C, Bridges A, Hardwicke P, Singh OMP, Giblin G, Pautsch A, Pfau R, Schnapp G, Wang M, Olieric V, Caffrey M. Crystallizing Membrane Proteins in the Lipidic Mesophase. Experience with Human Prostaglandin E2 Synthase 1 and an Evolving Strategy. *Crystal Growth & Design*. 2014 apr; 14(4):2034–2047. , doi:10.1021/cg500157x.

Li D, Lee J, Caffrey M. Crystallizing Membrane Proteins in Lipidic Mesophases. A Host Lipid Screen. *Crystal Growth & Design*. 2011 feb; 11(2):530–537. , doi:10.1021/cg101378s.

Li G, Zhang J, Guo Q, Jiang Y, Wei J. Efflux Pump Gene Expression in Multidrug- Resistant Mycobacterium tuberculosis Clinical Isolates. . 2015; p. 1–12. doi:10.1371/journal.pone.0119013.

Li M, Nyantakyi SA, Go ML, Dick T. Resistance against Membrane-Inserting MmpL3 Inhibitor through Upregulation of MmpL5 in Mycobacterium tuberculosis. *Antimicrobial Agents and Chemotherapy*. 2020 nov; 64(12). , doi:10.1128/AAC.01100-20.

Li M, Phua ZY, Xi Y, Xu Z, Nyantakyi SA, Li W, Jackson M, Wong MW, Lam Y, Chng SS, Go ML, Dick T. Potency Increase of Spiroketal Analogs of Membrane Inserting Indolyl Mannich Base Antimycobacterials Is Due to Acquisition of MmpL3 Inhibition. *ACS Infectious Diseases*. 2020 jul; 6(7):1882–1893. , doi:10.1021/acsinfecdis.0c00121.

Li W, Obregón-Henao A, Wallach JB, North EJ, Lee RE, Gonzalez-Juarrero M, Schnappinger D, Jackson M. Therapeutic Potential of the Mycobacterium tuberculosis Mycolic Acid Transporter, MmpL3. *Antimicrobial Agents and Chemotherapy*. 2016 sep; 60(9):5198–5207. , doi:10.1128/AAC.00826-16.

Li W, Sanchez-Hidalgo A, Jones V, de Moura VCN, North EJ, Jackson M. Synergistic Interactions of MmpL3 Inhibitors with Antitubercular Compounds In Vitro. *Antimicrobial Agents and Chemotherapy*. 2017 apr; 61(4):1–6. , doi:10.1128/AAC.02399-16.

Li W, Stevens CM, Pandya AN, Darzynkiewicz Z, Bhattarai P, Tong W, Gonzalez-Juarrero M, North EJ, Zgurskaya HI, Jackson M. Direct Inhibition of MmpL3 by Novel Antitubercular Compounds. *ACS Infectious Diseases*. 2019 jun; 5(6):1001–1012. , doi:10.1021/acsinfecdis.9b00048.

Li W, Yazidi A, Pandya AN, Hegde P, Tong W, Calado Nogueira de Moura V, North EJ, Sygusch J, Jackson M. MmpL3 as a Target for the Treatment of Drug-Resistant Nontuberculous Mycobacterial Infections. *Frontiers in Microbiology*. 2018 jul; 9(July):1–9. , doi:10.3389/fmicb.2018.01547.

Lieberman RL, Culver JA, Entzminger KC, Pai JC, Maynard JA. Crystallization chaperone strategies for membrane proteins. *Methods*. 2011 dec; 55(4):293–302. , doi:10.1016/j.ymeth.2011.08.004.

Liebschner D, Afonine PV, Baker ML, Bunkóczi G, Chen VB, Croll TI, Hintze B, Hung LW, Jain S, McCoy AJ, Moriarty NW, Oeffner RD, Poon BK, Prisant MG, Read RJ, Richardson JS, Richardson DC, Sammito MD, Sobolev OV, Stockwell DH, et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallographica Section D Structural Biology*. 2019 oct; 75(10):861–877. , doi:10.1107/S2059798319011471.

Liu W, Hanson MA, Stevens RC, Cherezov V. LCP-Tm: An Assay to Measure and Understand Stability of Membrane Proteins in a Membrane Environment. *Biophysical Journal*. 2010 apr; 98(8):1539–1548. , doi:10.1016/j.bpj.2009.12.4296.

Loll PJ. Membrane proteins, detergents and crystals: what is the state of the art? *Acta Crystallographica Section F Structural Biology Communications*. 2014 dec; 70(12):1576–1583. , doi:10.1107/S2053230X14025035.

Long F, Su CC, Zimmermann MT, Boyken SE, Rajashankar KR, Jernigan RL, Yu EW. Crystal structures of the CusA efflux pump suggest methionine-mediated metal transport. *Nature*. 2010 sep; 467(7314):484–488. , doi:10.1038/nature09395.

Long T, Qi X, Hassan A, Liang Q, De Brabander JK, Li X. Structural basis for itraconazole-mediated NPC1 inhibition. *Nature Communications*. 2020 dec; 11(1):152. , doi:10.1038/s41467-019-13917-5.

Lun S, Guo H, Onajole OK, Pieroni M, Gunosewoyo H, Chen G, Tipparaju SK, Ammerman NC, Kozikowski AP, Bishai WR. Indoleamides are active against drug-resistant Mycobacterium tuberculosis. *Nature Communications*. 2013 dec; 4(1):2907. , doi:10.1038/ncomms3907.

Ly A, Liu J. Mycobacterial Virulence Factors: Surface-Exposed Lipids and Secreted Proteins. *International Journal of Molecular Sciences*. 2020 jun; 21(11):3985. , doi:10.3390/ijms21113985.

Lyons JA, Parker JL, Solcan N, Brinth A, Li D, Shah ST, Caffrey M, Newstead S. Structural basis for polyspecificity in the POT family of proton-coupled oligopeptide transporters. *EMBO reports*. 2014 aug; 15(8):886–893. , doi:10.15252/embr.201338403.

Lyumkis D. Challenges and opportunities in cryo-EM single-particle analysis. *Journal of Biological Chemistry*. 2019 mar; 294(13):5181–5197. , doi:10.1074/jbc.REV118.005602.

Ma P, Weichert D, Aleksandrov LA, Jensen TJ, Riordan JR, Liu X, Kobilka BK, Caffrey M. The cubicon method for concentrating membrane proteins in the cubic mesophase. *Nature Protocols*. 2017 sep; 12(9):1745–1762. , doi:10.1038/nprot.2017.057.

Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*. 2019 jul; 47(W1):W636–W641. , doi:10.1093/nar/gkz268.

Male TL, Bartels K, Wiggers F, Rosenblum G, Jungwirth J, Hofmann H, Löw C. The conformational distribution of a major facilitator superfamily peptide transporter is modulated by the membrane composition. *bioRxiv*. 2020; doi:10.1101/2020.01.09.900316.

Marrakchi H, Lanéelle MA, Daffé M. Mycolic Acids: Structures, Biosynthesis, and Beyond. *Chemistry & Biology*. 2014 jan; 21(1):67–85. , doi:10.1016/j.chembiol.2013.11.011.

Martens C, Shekhar M, Borysik AJ, Lau AM, Reading E, Tajkhorshid E, Booth PJ, Politis A. Direct protein-lipid interactions shape the conformational landscape of secondary transporters. *Nature Communications*. 2018; 9(1):1–12. , doi:10.1038/s41467-018-06704-1.

Martinez Molledo M, Quistgaard EM, Flayhan A, Pieprzyk J, Löw C. Multispecific Substrate Recognition in a Proton-Dependent Oligopeptide Transporter. *Structure*. 2018 mar; 26(3):467–476.e4. , doi:10.1016/j.str.2018.01.005.

Martinez Molledo M, Quistgaard EM, Löw C. Tripeptide binding in a proton-dependent oligopeptide transporter. *FEBS Letters*. 2018 oct; 592(19):3239–3247. , doi:10.1002/1873-3468.13246.

McLean MA, Gregory MC, Sligar SG. Nanodiscs: A Controlled Bilayer Surface for the Study of Membrane Proteins. *Annual Review of Biophysics*. 2018 may; 47(1):107–124. , doi:10.1146/annurev-biophys-070816-033620.

McNeil MB, Cook GM. Utilization of CRISPR Interference To Validate MmpL3 as a Drug Target in Mycobacterium tuberculosis. *Antimicrobial Agents and Chemotherapy*. 2019 aug; 63(8):1–7. , doi:10.1128/AAC.00629-19.

McNeil MB, Dennison D, Parish T. Mutations in MmpL3 alter membrane potential, hydrophobicity and antibiotic susceptibility in Mycobacterium smegmatis. *Microbiology*. 2017 jul; 163(7):1065–1070. , doi:10.1099/mic.0.000498.

McNeil MB, O'Malley T, Dennison D, Shelton CD, Sunde B, Parish T. Multiple Mutations in Mycobacterium tuberculosis MmpL3 Increase Resistance to MmpL3 Inhibitors. *mSphere*. 2020 oct; 5(5):1–7. , doi:10.1128/mSphere.00985-20.

Melly G, Purdy G. MmpL Proteins in Physiology and Pathogenesis of M. tuberculosis. *Microorganisms*. 2019 mar; 7(3):70. , doi:10.3390/microorganisms7030070.

Melly GC, Stokas H, Dunaj JL, Hsu FF, Rajavel M, Su CC, Yu EW, Purdy GE. Structural and functional evidence that lipoprotein LpqN supports cell envelope biogenesis in Mycobacterium tuberculosis. *Journal of Biological Chemistry*. 2019 oct; 294(43):15711–15723. , doi:10.1074/jbc.RA119.008781.

Merk A, Bartesaghi A, Banerjee S, Falconieri V, Rao P, Davis MI, Pragani R, Boxer MB, Earl LA, Milne JLS, Subramaniam S. Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery. *Cell*. 2016 jun; 165(7):1698–1707. , doi:10.1016/j.cell.2016.05.040.

Michael Green N. Avidin and streptavidin. In: *Methods in Enzymology*, vol. 184; 1990.p. 51–67. , doi:10.1016/0076-6879(90)84259-J.

Milano A, Pasca MR, Provvedi R, Lucarelli AP, Manina G, Luisa de Jesus Lopes Ribeiro A, Manganelli R, Riccardi G. Azole resistance in Mycobacterium tuberculosis is mediated by the MmpS5–MmpL5 efflux system. *Tuberculosis*. 2009 jan; 89(1):84–90. , doi:10.1016/j.tube.2008.08.003.

Miles AJ, Wallace BA. Circular dichroism spectroscopy of membrane proteins. *Chemical Society Reviews*. 2016; 45(18):4859–4872. , doi:10.1039/C5CS00084J.

Miroux B, Walker JE. Over-production of Proteins in Escherichia coli: Mutant Hosts that Allow Synthesis of some Membrane Proteins and Globular Proteins at High Levels. *Journal of Molecular Biology*. 1996 jul; 260(3):289–298. , doi:10.1006/jmbi.1996.0399.

Mitchell LS, Colwell LJ. Analysis of nanobody paratopes reveals greater diversity than classical antibodies. *Protein Engineering, Design and Selection*. 2018 jul; 31(7-8):267–275. , doi:10.1093/protein/gzy017.

Mitchell LS, Colwell LJ. Comparative analysis of nanobody sequence and structure data. *Proteins: Structure, Function, and Bioinformatics*. 2018 jul; 86(7):697–706. , doi:10.1002/prot.25497.

Mitra A, Ko YH, Cingolani G, Niederweis M. Heme and hemoglobin utilization by Mycobacterium tuberculosis. *Nature Communications*. 2019 dec; 10(1):4260. , doi:10.1038/s41467-019-12109-5.

Moseng MA, Lyu M, Pipatpolkai T, Glaza P, Emerson CC, Stewart PL, Stansfeld PJ, Yu EW. Cryo-EM Structures of CusA Reveal a Mechanism of Metal-Ion Export. *mBio*. 2021 apr; 12(2):1–17. , doi:10.1128/mBio.00452-21.

Murakami S, Nakashima R, Yamashita E, Yamaguchi A. Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature*. 2002 oct; 419(6907):587–593. , doi:10.1038/nature01050.

Muyldermans S. Nanobodies: Natural Single-Domain Antibodies. *Annual Review of Biochemistry*. 2013 jun; 82(1):775–797. , doi:10.1146/annurev-biochem-063011-092449.

Nagamura R, Fukuda M, Kawamoto A, Matoba K, Dohmae N, Ishitani R, Takagi J, Nureki O. Structural basis for oligomerization of the prokaryotic peptide transporter PepTSo2. *Acta Crystallographica Section F Structural Biology Communications*. 2019 may; 75(5):348–358. , doi:10.1107/S2053230X19003546.

Narang A, Giri A, Gupta S, Garima K, Bose M, Varma-Basil M. Contribution of putative efflux pump genes to isoniazid resistance in clinical isolates of Mycobacterium tuberculosis. *International Journal of Mycobacteriology*. 2017; 6(2):177. , doi:10.4103/ijmy.ijmy_26_17.

Nataraj V, Varela C, Javid A, Singh A, Besra GS, Bhatt A. Mycolic acids: deciphering and targeting the Achilles' heel of the tubercle bacillus. *Molecular Microbiology*. 2015 oct; 98(1):7–16. , doi:10.1111/mmi.13101.

Newstead S. Molecular insights into proton coupled peptide transport in the PTR family of oligopeptide transporters. *Biochimica et Biophysica Acta (BBA) - General Subjects*. 2015 mar; 1850(3):488–499. , doi:10.1016/j.bbagen.2014.05.011.

Newstead S, Drew D, Cameron AD, Postis VLG, Xia X, Fowler PW, Ingram JC, Carpenter EP, Sansom MSP, McPherson MJ, Baldwin SA, Iwata S. Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2. *The EMBO Journal*. 2011 jan; 30(2):417–426. , doi:10.1038/emboj.2010.309.

Newstead S, Ferrandon S, Iwata S. Rationalizing α-helical membrane protein crystallization. *Protein Science*. 2008 mar; 17(3):466–472. , doi:10.1110/ps.073263108.

Nikaido H. RND transporters in the living world. *Research in Microbiology*. 2018 sep; 169(7-8):363–371. , doi:10.1016/j.resmic.2018.03.001.

Nikonenko BV, Protopopova M, Samala R, Einck L, Nacy CA. Drug Therapy of Experimental Tuberculosis (TB): Improved Outcome by Combining SQ109, a New Diamine Antibiotic, with Existing TB Drugs. *Antimicrobial Agents and Chemotherapy*. 2007 apr; 51(4):1563–1565. , doi:10.1128/AAC.01326-06.

North E, Jackson M, Lee R. New Approaches to Target the Mycolic Acid Biosynthesis Pathway for the Development of Tuberculosis Therapeutics. *Current Pharmaceutical Design*. 2014 nov; 20(27):4357–4378. , doi:10.2174/1381612819666131118203641.

Nygaard R, Kim J, Mancia F. Cryo-electron microscopy analysis of small membrane proteins. *Current Opinion in Structural Biology*. 2020 oct; 64:26–33. , doi:10.1016/j.sbi.2020.05.009.

Ohi M, Li Y, Cheng Y, Walz T. Negative staining and image classification — powerful tools in modern electron microscopy. *Biological Procedures Online*. 2004 jan; 6(1):23–34. , doi:10.1251/bpo70.

Oliver RC, Lipfert J, Fox DA, Lo RH, Doniach S, Columbus L. Dependence of Micelle Size and Shape on Detergent Alkyl Chain Length and Head Group. *PLoS ONE*. 2013 may; 8(5):e62488. , doi:10.1371/journal.pone.0062488.

Owens CP, Chim N, Graves AB, Harmston CA, Iniguez A, Contreras H, Liptak MD, Goulding CW. The Mycobacterium tuberculosis Secreted Protein Rv0203 Transfers Heme to Membrane Proteins MmpL3 and MmpL11. *Journal of Biological Chemistry*. 2013 jul; 288(30):21714–21728. , doi:10.1074/jbc.M113.453076.

Pacheco SA, Hsu FF, Powers KM, Purdy GE. MmpL11 Protein Transports Mycolic Acid-containing Lipids to the Mycobacterial Cell Wall and Contributes to Biofilm Formation in Mycobacterium smegmatis. *Journal of Biological Chemistry*. 2013 aug; 288(33):24213–24222. , doi:10.1074/jbc.M113.473371.

Pai M, Behr MA, Dowdy D, Dheda K, Divangahi M, Boehme CC, Ginsberg A, Swaminathan S, Spigelman M, Getahun H, Menzies D, Raviglione M. Tuberculosis. *Nature Reviews Disease Primers*. 2016 dec; 2(1):16076. , doi:10.1038/nrdp.2016.76.

Pain R. Determining the CD Spectrum of a Protein. *Current Protocols in Protein Science*. 2004 nov; 38(1):7.6.1–7.6.24. , doi:10.1002/0471140864.ps0706s38.

Pak JE, Ekende EN, Kifle EG, O'Connell JD, De Angelis F, Tessema MB, Derfoufi KM, Robles-Colmenares Y, Robbins RA, Goormaghtigh E, Vandenbussche G, Stroud RM. Structures of intermediate transport states of ZneA, a Zn(II)/proton antiporter. *Proceedings of the National Academy of Sciences*. 2013 nov; 110(46):18484–18489. , doi:10.1073/pnas.1318705110.

Pandya AN, Prathipati PK, Hegde P, Li W, Graham KF, Mandal S, Drescher KM, Destache CJ, Ordway D, Jackson M, North EJ. Indole-2-Carboxamides Are Active against Mycobacterium abscessus in a Mouse Model of Acute Infection. *Antimicrobial Agents and Chemotherapy*. 2019 mar; 63(3):1–16. , doi:10.1128/AAC.02245-18.

Pankhurst LJ, del Ojo Elias C, Votintseva AA, Walker TM, Cole K, Davies J, Fermont JM, Gascoyne-Binzi DM, Kohl TA, Kong C, Lemaitre N, Niemann S, Paul J, Rogers TR, Roycroft E, Smith EG, Supply P, Tang P, Wilcox MH, Wordsworth S, et al. Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study. *The Lancet Respiratory Medicine*. 2016 jan; 4(1):49–58. , doi:10.1016/S2213-2600(15)00466-X.

Pardon E, Laeremans T, Triest S, Rasmussen SGF, Wohlkönig A, Ruf A, Muyldermans S, Hol WGJ, Kobilka BK, Steyaert J. A general protocol for the generation of Nanobodies for structural biology. *Nature Protocols*. 2014 mar; 9(3):674–693. , doi:10.1038/nprot.2014.039.

Parker JL, Corey RA, Stansfeld PJ, Newstead S. Structural basis for substrate specificity and regulation of nucleotide sugar transporters in the lipid bilayer. *Nature Communications*. 2019 dec; 10(1):4657. , doi:10.1038/s41467-019-12673-w.

Parker JL, Deme JC, Kuteyi G, Wu Z, Huo J, Goldman ID, Owens RJ, Biggin PC, Lea SM, Newstead S. Structural basis of antifolate recognition and transport by PCFT. *Nature*. 2021 may; , doi:10.1038/s41586-021-03579-z.

Parker JL, Mindell JA, Newstead S. Thermodynamic evidence for a dual transport mechanism in a POT peptide transporter. *eLife*. 2014 dec; 3:1–13. , doi:10.7554/eLife.04273.

Parker JL, Newstead S. Current trends in α-helical membrane protein crystallization: An update. *Protein Science*. 2012 sep; 21(9):1358–1365. , doi:10.1002/pro.2122.

Parker JL, Newstead S. The Next Generation in Membrane Protein Structure Determination, vol. 922 of Advances in Experimental Medicine and Biology. Moraes I, editor, Cham: Springer International Publishing; 2016. , doi:10.1007/978-3-319-35072-1.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*. 2004 oct; 25(13):1605–1612. , doi:10.1002/jcc.20084.

Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science*. 2021 jan; 30(1):70–82. , doi:10.1002/pro.3943.

Poce G, Bates RH, Alfonso S, Cocozza M, Porretta GC, Ballell L, Rullas J, Ortega F, De Logu A, Agus E, La Rosa V, Pasca MR, De Rossi E, Wae B, Franzblau SG, Manetti F, Botta M, Biava M. Improved BM212 MmpL3 Inhibitor Analogue Shows Efficacy in Acute Murine Model of Tuberculosis Infection. *PLoS ONE*. 2013 feb; 8(2):e56980. , doi:10.1371/journal.pone.0056980.

Poce G, Consalvi S, Venditti G, Alfonso S, Desideri N, Fernandez-Menendez R, Bates RH, Ballell L, Barros Aguirre D, Rullas J, De Logu A, Gardner M, Ioerger TR, Rubin EJ, Biava M. Novel Pyrazole-Containing Compounds Active against Mycobacterium tuberculosis. *ACS Medicinal Chemistry Letters*. 2019 oct; 10(10):1423–1429. , doi:10.1021/acsmedchemlett.9b00204.

Popot JL, Althoff T, Bagnard D, Banères JL, Bazzacco P, Billon-Denis E, Catoire LJ, Champeil P, Charvolin D, Cocco MJ, Crémel G, Dahmane T, de la Maza LM, Ebel C, Gabel F, Giusti F, Gohon Y, Goormaghtigh E, Guittet E, Kleinschmidt JH, et al. Amphipols From A to Z*. *Annual Review of Biophysics*. 2011 jun; 40(1):379–408. , doi:10.1146/annurev-biophys-042910-155219.

Popot JL. The Use of Amphipols for Electron Microscopy. In: Popot JL, editor. *Membrane Proteins in Aqueous Solutions From Detergents to Amphipols*, vol. 2000 Springer US; 2018.p. 533–590. , doi:10.1007/978-3-319-73148-3_12.

Prisic S, Dankwa S, Schwartz D, Chou MF, Locasale JW, Kang CM, Bemis G, Church GM, Steen H, Husson RN. Extensive phosphorylation with overlapping specificity by Mycobacterium tuberculosis serine/threonine protein kinases. *Proceedings of the National Academy of Sciences*. 2010 apr; 107(16):7521–7526. , doi:10.1073/pnas.0913482107.

Privé GG. Detergents for the stabilization and crystallization of membrane proteins. *Methods*. 2007 apr; 41(4):388–397. , doi:10.1016/j.ymeth.2007.01.007.

Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*. 2017 mar; 14(3):290–296. , doi:10.1038/nmeth.4169.

Punjani A, Zhang H, Fleet DJ. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nature Methods*. 2020 dec; 17(12):1214–1221. , doi:10.1038/s41592-020-00990-8.

Pyle E, Kalli AC, Amillis S, Hall Z, Lau AM, Hanyaloglu AC, Diallinas G, Byrne B, Politis A. Structural Lipids Enable the Formation of Functional Oligomers of the Eukaryotic Purine Symporter UapA. *Cell Chemical Biology*. 2018 jul; 25(7):840–848.e4. , doi:10.1016/j.chembiol.2018.03.011.

Qi C, Di Minin G, Vercellino I, Wutz A, Korkhov VM. Structural basis of sterol recognition by human hedgehog receptor PTCH1. *Science Advances*. 2019 sep; 5(9):eaaw6490. , doi:10.1126/sciadv.aaw6490.

Qi X, Li X. Mechanistic Insights into the Generation and Transduction of Hedgehog Signaling. *Trends in Biochemical Sciences*. 2020 may; 45(5):397–410. , doi:10.1016/j.tibs.2020.01.006.

Qi X, Schmiege P, Coutavas E, Li X. Two Patched molecules engage distinct sites on Hedgehog yielding a signaling-competent complex. *Science*. 2018 oct; 362(6410):eaas8843. , doi:10.1126/science.aas8843.

Qian H, Wu X, Du X, Yao X, Zhao X, Lee J, Yang H, Yan N. Structural Basis of Low-pH-Dependent Lysosomal Cholesterol Egress by NPC1 and NPC2. *Cell*. 2020 jul; 182(1):98–111.e18. , doi:10.1016/j.cell.2020.05.020.

Quémard A. New Insights into the Mycolate-Containing Compound Biosynthesis and Transport in Mycobacteria. *Trends in Microbiology*. 2016 sep; 24(9):725–738. , doi:10.1016/j.tim.2016.04.009.

Quistgaard EM, Martinez Molledo M, Löw C. Structure determination of a major facilitator peptide transporter: Inward facing PepTSt from Streptococcus thermophilus crystallized in space group P3121. *PLOS ONE*. 2017 mar; 12(3):e0173126. , doi:10.1371/journal.pone.0173126.

Rajni, Rao N, Meena LS. Biosynthesis and Virulent Behavior of Lipids Produced by Mycobacterium tuberculosis : LAM and Cord Factor: An Overview. *Biotechnology Research International*. 2011 dec; 2011:1–7. , doi:10.4061/2011/274693.

Rao SPS, Lakshminarayana SB, Kondreddi RR, Herve M, Camacho LR, Bifani P, Kalapala SK, Jiricek J, Ma NL, Tan BH, Ng SH, Nanjundappa M, Ravindran S, Seah PG, Thayalan P, Lim SH, Lee BH, Goh A, Barnes WS, Chen Z, et al. Indolcarboxamide Is a Preclinical Candidate for Treating Multidrug-Resistant Tuberculosis. *Science Translational Medicine*. 2013 dec; 5(214):214ra168–214ra168. , doi:10.1126/scitranslmed.3007355.

Rayasam GV. MmpL3 a potential new target for development of novel anti-tuberculosis drugs. *Expert Opinion on Therapeutic Targets*. 2014 mar; 18(3):247–256. , doi:10.1517/14728222.2014.859677.

Raynaud C, Daher W, Johansen MD, Roquet-Banères F, Blaise M, Onajole OK, Kozikowski AP, Herrmann JL, Dziadek J, Gobis K, Kremer L. Active Benzimidazole Derivatives Targeting the MmpL3 Transporter in Mycobacterium abscessus. *ACS Infectious Diseases*. 2020 feb; 6(2):324–337. , doi:10.1021/acsinfecdis.9b00389.

Reboul CF, Eager M, Elmlund D, Elmlund H. Single-particle cryo-EM—Improved ab initio 3D reconstruction with SIMPLE/PRIME. *Protein Science*. 2018 jan; 27(1):51–61. , doi:10.1002/pro.3266.

Reisinger V, Eichacker LA. Solubilization of membrane protein complexes for blue native PAGE. *Journal of Proteomics*. 2008 aug; 71(3):277–283. , doi:10.1016/j.jprot.2008.05.004.

Remuiñán MJ, Pérez-Herrán E, Rullás J, Alemparte C, Martínez-Hoyos M, Dow DJ, Afari J, Mehta N, Esquivias J, Jiménez E, Ortega-Muro F, Fraile-Gabaldón MT, Spivey VL, Loman NJ, Pallen MJ, Constantinidou C, Minick DJ, Cacho M, Rebollo-López MJ, González C, et al. Tetrahydropyrazolo[1,5-a]Pyrimidine-3-Carboxamide and N-Benzyl-6',7'-Dihydrospiro[Piperidine-4,4'-Thieno[3,2-c]Pyran] Analogues with Bactericidal Efficacy against Mycobacterium tuberculosis Targeting MmpL3. *PLoS ONE*. 2013 apr; 8(4):e60933. , doi:10.1371/journal.pone.0060933.

Ritchie TK, Grinkova YV, Bayburt TH, Denisov IG, Zolnerciks JK, Atkins WM, Sligar SG. Reconstitution of Membrane Proteins in Phospholipid Bilayer Nanodiscs. In: Düzgünes N, editor. *Methods in Enzymology*, vol. 464 Elsevier Masson SAS; 2009.p. 211–231. , doi:10.1016/S0076-6879(09)64011-8.

Roh SH, Shekhar M, Pintilie G, Chipot C, Wilkens S, Singharoy A, Chiu W. Cryo-EM and MD infer water-mediated proton transport and autoinhibition mechanisms of Vo complex. *Science Advances*. 2020 oct; 6(41):eabb9605. , doi:10.1126/sciadv.abb9605.

Rubio-Aliaga I, Daniel H. Mammalian peptide transporters as targets for drug delivery. *Trends in Pharmacological Sciences*. 2002 sep; 23(9):434–440. , doi:10.1016/S0165-6147(02)02072-2.

Rudolf AF, Kinnebrew M, Kowatsch C, Ansell TB, El Omari K, Bishop B, Pardon E, Schwab RA, Malinauskas T, Qian M, Duman R, Covey DF, Steyaert J, Wagner A, Sansom MSP, Rohatgi R, Siebold C. The morphogen Sonic hedgehog inhibits its receptor Patched by a pincer grasp mechanism. *Nature Chemical Biology*. 2019 oct; 15(10):975–982. , doi:10.1038/s41589-019-0370-y.

Ruhl CR, Pasko BL, Khan HS, Kindt LM, Stamm CE, Franco LH, Hsia CC, Zhou M, Davis CR, Qin T, Gautron L, Burton MD, Mejia GL, Naik DK, Dussor G, Price TJ, Shiloh MU. Mycobacterium tuberculosis Sulfolipid-1 Activates Nociceptive Neurons and Induces Cough. *Cell*. 2020 apr; 181(2):293–305.e11. , doi:10.1016/j.cell.2020.02.026.

Sacksteder KA, Protopopova M, Barry CE, Andries K, Nacy CA. Discovery and development of SQ109: a new antitubercular drug with a novel mechanism of action. *Future Microbiology*. 2012 jul; 7(7):823–837. , doi:10.2217/fmb.12.56.

Sahu S, Ditiu L, Zumla A. After the UNGA High-Level Meeting on Tuberculosis—what next and how? *The Lancet Global Health*. 2019 may; 7(5):e558–e560. , doi:10.1016/S2214-109X(19)30068-3.

Saier MH, Tam R, Reizer A, Reizer J. Two novel families of bacterial membrane proteins concerned with nodulation, cell division and transport. *Molecular Microbiology*. 1994 mar; 11(5):841–847. , doi:10.1111/j.1365-2958.1994.tb00362.x.

Saier MH, Reddy VS, Moreno-Hagelsieb G, Hendargo KJ, Zhang Y, Iddamsetty V, Lam KJK, Tian N, Russum S, Wang J, Medrano-Soto A. The Transporter Classification Database (TCDB): 2021 update. *Nucleic Acids Research*. 2021 jan; 49(D1):D461–D467. , doi:10.1093/nar/gkaa1004.

Salvador D, Glavier M, Schoehn G, Phan G, Taveau JC, Decossas M, Lecomte S, Mongrand S, Garnier C, Broutin I, Daury L, Lambert O. Minimal nanodisc without exogenous lipids for stabilizing membrane proteins in detergent-free buffer. *Biochimica et Biophysica Acta (BBA) - Biomembranes*. 2019 apr; 1861(4):852–860. , doi:10.1016/j.bbamem.2019.01.013.

Scheres SHW. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*. 2012 dec; 180(3):519–530. , doi:10.1016/j.jsb.2012.09.006.

Seddon AM, Curnow P, Booth PJ. Membrane proteins, lipids and detergents: not just a soap opera. *Biochimica et Biophysica Acta (BBA) - Biomembranes*. 2004 nov; 1666(1-2):105–117. , doi:10.1016/j.bbamem.2004.04.011.

Seeger MA, Mittal A, Velamakanni S, Hohl M, Schauer S, Salaa I, Grütter MG, van Veen HW. Tuning the Drug Efflux Activity of an ABC Transporter in vivo by in vitro Selected DARPin Binders. *PLoS ONE*. 2012 jun; 7(6):e37845. , doi:10.1371/journal.pone.0037845.

Sethiya JP, Sowards MA, Jackson M, North EJ. MmpL3 Inhibition: A New Approach to Treat Nontuberculous Mycobacterial Infections. *International Journal of Molecular Sciences*. 2020 aug; 21(17):6202. , doi:10.3390/ijms21176202.

Seung KJ, Keshavjee S, Rich ML. Multidrug-Resistant Tuberculosis and Extensively Drug-Resistant Tuberculosis. *Cold Spring Harbor Perspectives in Medicine*. 2015 sep; 5(9):a017863. , doi:10.1101/cshperspect.a017863.

Shahsavar A, Stohler P, Bourenkov G, Zimmermann I, Siegrist M, Guba W, Pinard E, Sinning S, Seeger MA, Schneider TR, Dawson RJP, Nissen P. Structural insights into the inhibition of glycine reuptake. *Nature*. 2021 mar; 591(7851):677–681. , doi:10.1038/s41586-021-03274-z.

Shang G, Zhang C, Chen ZJ, Bai Xc, Zhang X. Cryo-EM structures of STING reveal its mechanism of activation by cyclic GMP–AMP. *Nature*. 2019 mar; 567(7748):389–393. , doi:10.1038/s41586-019-0998-5.

Shao M, McNeil M, Cook GM, Lu X. MmpL3 inhibitors as antituberculosis drugs. *European Journal of Medicinal Chemistry*. 2020 aug; 200:112390. , doi:10.1016/j.ejmech.2020.112390.

Shen PS, Yang X, DeCaen PG, Liu X, Bulkley D, Clapham DE, Cao E. The Structure of the Polycystic Kidney Disease Channel PKD2 in Lipid Nanodiscs. *Cell*. 2016 oct; 167(3):763–773.e11. , doi:10.1016/j.cell.2016.09.048.

Shetty A, Xu Z, Lakshmanan U, Hill J, Choong ML, Chng SS, Yamada Y, Poulsen A, Dick T, Gengenbacher M. Novel Acetamide Indirectly Targets Mycobacterial Transporter MmpL3 by Proton Motive Force Disruption. *Frontiers in Microbiology*. 2018 dec; 9(December):1–12. , doi:10.3389/fmicb.2018.02960.

Sircar A, Sanni KA, Shi J, Gray JJ. Analysis and Modeling of the Variable Region of Camelid Single-Domain Antibodies. *The Journal of Immunology*. 2011 jun; 186(11):6357–6367. , doi:10.4049/jimmunol.1100116.

Skar-Gislinge N, Johansen NT, Høiberg-Nielsen R, Arleth L. Comprehensive Study of the Self-Assembly of Phospholipid Nanodiscs: What Determines Their Shape and Stoichiometry? *Langmuir*. 2018 oct; 34(42):12569–12582. , doi:10.1021/acs.langmuir.8b01503.

Skar-Gislinge N, Simonsen JB, Mortensen K, Feidenhans'l R, Sligar SG, Lindberg Møller B, Bjørnholm T, Arleth L. Elliptical Structure of Phospholipid Bilayer Nanodiscs Encapsulated by Scaffold Proteins: Casting the Roles of the Lipids and the Protein. *Journal of the American Chemical Society*. 2010 oct; 132(39):13713–13722. , doi:10.1021/ja1030613.

Snapper SB, Melton RE, Mustafa S, Kieser T, Jr WRJ. Isolation and characterization of efficient plasmid transformation mutants of Mycobacterium smegmatis. *Molecular Microbiology*. 1990 nov; 4(11):1911–1919. , doi:10.1111/j.1365-2958.1990.tb02040.x.

Solcan N, Kwok J, Fowler PW, Cameron AD, Drew D, Iwata S, Newstead S. Alternating access mechanism in the POT family of oligopeptide transporters. *The EMBO Journal*. 2012 aug; 31(16):3411–3421. , doi:10.1038/emboj.2012.157.

Sonoda Y, Newstead S, Hu NJ, Alguel Y, Nji E, Beis K, Yashiro S, Lee C, Leung J, Cameron AD, Byrne B, Iwata S, Drew D. Benchmarking Membrane Protein Detergent Stability for Improving Throughput of High-Resolution X-ray Structures. *Structure*. 2011 jan; 19(1):17–25. , doi:10.1016/j.str.2010.12.001.

Stanley SA, Cox JS. Host–Pathogen Interactions During Mycobacterium tuberculosis infections. In: Pieters J, McKinney JD, editors. *Pathogenesis of Mycobacterium tuberculosis and its Interaction with the Host Organism* Springer Berlin Heidelberg; 2013.p. 211–241. , doi:10.1007/82_2013_332.

Stanley SA, Grant SS, Kawate T, Iwase N, Shimizu M, Wivagg C, Silvis M, Kazyanskaya E, Aquadro J, Golas A, Fitzgerald M, Dai H, Zhang L, Hung DT. Identification of Novel Inhibitors of M. tuberculosis Growth Using Whole Cell Based High-Throughput Screening. *ACS Chemical Biology*. 2012 aug; 7(8):1377–1384. , doi:10.1021/cb300151m.

Stec J, Onajole OK, Lun S, Guo H, Merenbloom B, Vistoli G, Bishai WR, Kozikowski AP. Indole-2-carboxamide-based MmpL3 Inhibitors Show Exceptional Antitubercular Activity in an Animal Model of Tuberculosis Infection. *Journal of Medicinal Chemistry*. 2016 jul; 59(13):6232–6247. , doi:10.1021/acs.jmedchem.6b00415.

Stetsenko A, Guskov A. An Overview of the Top Ten Detergents Used for Membrane Protein Crystallization. *Crystals*. 2017 jul; 7(7):197. , doi:10.3390/cryst7070197.

Steyaert J, Kobilka BK. Nanobody stabilization of G protein-coupled receptor conformational states. *Current Opinion in Structural Biology*. 2011 aug; 21(4):567–572. , doi:10.1016/j.sbi.2011.06.011.

Stodola FH, Lesuk A, Anderson RJ. The Chemistry of the Lipids of Tubercle Bacilli. *Journal of Biological Chemistry*. 1938 dec; 126(2):505–513. , doi:10.1016/S0021-9258(18)73857-X.

Su CC, Klenotic PA, Bolla JR, Purdy GE, Robinson CV, Yu EW. MmpL3 is a lipid transporter that binds trehalose monomycolate and phosphatidylethanolamine. *Proceedings of the National Academy of Sciences*. 2019 jun; 116(23):11241–11246. , doi:10.1073/pnas.1901346116.

Su CC, Lyu M, Morgan CE, Bolla JR, Robinson CV, Yu EW. A 'Build and Retrieve' methodology to simultaneously solve cryo-EM structures of membrane proteins. *Nature Methods*. 2021 jan; 18(1):69–75. , doi:10.1038/s41592-020-01021-2.

Sulzenbacher G, Canaan S, Bordat Y, Neyrolles O, Stadthagen G, Roig-Zamboni V, Rauzier J, Maurin D, Laval F, Daffé M, Cambillau C, Gicquel B, Bourne Y, Jackson M. LppX is a lipoprotein required for the translocation of phthiocerol dimycocerosates to the surface of Mycobacterium tuberculosis. *The EMBO Journal*. 2006 apr; 25(7):1436–1444. , doi:10.1038/sj.emboj.7601048.

Székely R, Cole ST. Mechanistic insight into mycobacterial MmpL protein function. *Molecular Microbiology*. 2016 mar; 99(5):831–834. , doi:10.1111/mmi.13306.

Tahlan K, Wilson R, Kastrinsky DB, Arora K, Nair V, Fischer E, Barnes SW, Walker JR, Alland D, Barry CE, Boshoff HI. SQ109 Targets MmpL3, a Membrane Transporter of Trehalose Monomycolate Involved in Mycolic Acid Donation to the Cell Wall Core of Mycobacterium tuberculosis. *Antimicrobial Agents and Chemotherapy*. 2012 apr; 56(4):1797–1809. , doi:10.1128/AAC.05708-11.

Takayama K, Wang C, Besra GS. Pathway to Synthesis and Processing of Mycolic Acids in Mycobacterium tuberculosis. *Clinical Microbiology Reviews*. 2005 jan; 18(1):81–101. , doi:10.1128/CMR.18.1.81-101.2005.

Tang G, Peng L, Baldwin PR, Mann DS, Jiang W, Rees I, Ludtke SJ. EMAN2: An extensible image processing suite for electron microscopy. *Journal of Structural Biology*. 2007 jan; 157(1):38–46. , doi:10.1016/j.jsb.2006.05.009.

Tantry SJ, Degiacomi G, Sharma S, Jena LK, Narayan A, Guptha S, Shanbhag G, Menasinakai S, Mallya M, Awasthy D, Balakrishnan G, Kaur P, Bhattacharjee D, Narayan C, Reddy J, Naveen Kumar CN, Shandil R, Boldrin F, Ventura M, Manganelli R, et al. Whole cell screen based identification of spiropiperidines with potent antitubercular properties. *Bioorganic & Medicinal Chemistry Letters*. 2015 aug; 25(16):3234–3245. , doi:10.1016/j.bmcl.2015.05.087.

Tareen A, Kinney JB. Logomaker: beautiful sequence logos in Python. *Bioinformatics*. 2020 apr; 36(7):2272–2274. , doi:10.1093/bioinformatics/btz921.

Taylor WR. Residual colours: a proposal for aminochromography. *Protein Engineering Design and Selection*. 1997 jul; 10(7):743–746. , doi:10.1093/protein/10.7.743.

TB Alliance. Annual Report "Breaking Through" 2019; 2019.

Tekaia F, Gordon SV, Garnier T, Brosch R, Barrell BG, Cole ST. Analysis of the proteome of Mycobacterium tuberculosis in silico. *Tubercle and Lung Disease*. 1999 dec; 79(6):329–342. , doi:10.1054/tuld.1999.0220.

The CRyPTIC Consortium, 100000 Genomes Project. Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing. *New England Journal of Medicine*. 2018 oct; 379(15):1403–1415. , doi:10.1056/NEJMoa1800474.

The CRyPTIC Consortium, Fowler PW. Epidemiological cutoffs for a 96-well broth microdilution plate for high-throughput research antibiotic susceptibility testing of M. tuberculosis. *medRxiv*. 2021; doi:https://doi.org/10.1101/2021.02.24.21252386.

Touchette MH, Seeliger JC. Transport of outer membrane lipids in mycobacteria. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*. 2017 nov; 1862(11):1340–1354. , doi:10.1016/j.bbalip.2017.01.005.

Touchette MH, Van Vlack ER, Bai L, Kim J, Cognetta AB, Previti ML, Backus KM, Martin DW, Cravatt BF, Seeliger JC. A Screen for Protein–Protein Interactions in Live Mycobacteria Reveals a Functional Link between the Virulence-Associated Lipid Transporter LprG and the Mycolyltransferase Antigen 85A. *ACS Infectious Diseases*. 2017 may; 3(5):336–348. , doi:10.1021/acsinfecdis.6b00179.

Tribet C, Audebert R, Popot JL. Amphipols: Polymers that keep membrane proteins soluble in aqueous solutions. *Proceedings of the National Academy of Sciences*. 1996 dec; 93(26):15047–15050. , doi:10.1073/pnas.93.26.15047.

Trofimov V, Kicka S, Mucaria S, Hanna N, Ramon-Olayo F, Del Peral LVG, Lelièvre J, Ballell L, Scapozza L, Besra GS, Cox JAG, Soldati T. Antimycobacterial drug discovery using Mycobacteria-infected amoebae identifies anti-infectives and new molecular targets. *Scientific Reports*. 2018 dec; 8(1):3939. , doi:10.1038/s41598-018-22228-6.

Tseng TT, Gratwick KS, Kollman J, Park D, Nies DH, Goffeau A, Saier MH. The RND permease superfamily: an ancient, ubiquitous and diverse family that includes human disease and development proteins. *Journal of Molecular Microbiology and Biotechnology*. 1999 aug; 1(1):107–25. .

Tsukazaki T. Structure-based working model of SecDF, a proton-driven bacterial protein translocation factor. *FEMS Microbiology Letters*. 2018 jun; 365(12):1–9. , doi:10.1093/femsle/fny112.

Tsukazaki T, Mori H, Echizen Y, Ishitani R, Fukai S, Tanaka T, Perederina A, Vassylyev DG, Kohno T, Maturana AD, Ito K, Nureki O. Structure and function of a membrane component SecDF that enhances protein export. *Nature*. 2011 jun; 474(7350):235–238. , doi:10.1038/nature09980.

Tullius MV, Harmston CA, Owens CP, Chim N, Morse RP, McMath LM, Iniguez A, Kimmey JM, Sawaya MR, Whitelegge JP, Horwitz MA, Goulding CW. Discovery and characterization of a unique mycobacterial heme acquisition system. *Proceedings of the National Academy of Sciences*. 2011 mar; 108(12):5051–5056. , doi:10.1073/pnas.1009516108.

Uchański T, Masiulis S, Fischer B, Kalichuk V, López-Sánchez U, Zarkadas E, Weckener M, Sente A, Ward P, Wohlkönig A, Zögg T, Remaut H, Naismith JH, Nury H, Vranken W, Aricescu AR, Pardon E, Steyaert J. Megabodies expand the nanobody toolkit for protein structure determination by single-particle cryo-EM. *Nature Methods*. 2021 jan; 18(1):60–68. , doi:10.1038/s41592-020-01001-6.

Uchański T, Pardon E, Steyaert J. Nanobodies to study protein conformational states. *Current Opinion in Structural Biology*. 2020 feb; 60:117–123. , doi:10.1016/j.sbi.2020.01.003.

Udwadia ZF, Amale RA, Ajbani KK, Rodrigues C. Totally Drug-Resistant Tuberculosis in India. *Clinical Infectious Diseases*. 2012 feb; 54(4):579–581. , doi:10.1093/cid/cir889.

Umar F, Hatta M, Husain DR, Natzir R, Dwiyanti R, Junita AR, Primaguna MR. The effect of anti-tuberculosis drugs therapy on mRNA efflux pump gene expression of Rv1250 in Mycobacterium tuberculosis collected from tuberculosis patients. *New Microbes and New Infections*. 2019 nov; 32:100609. , doi:10.1016/j.nmni.2019.100609.

Uplekar M, Weil D, Lonnroth K, Jaramillo E, Lienhardt C, Dias HM, Falzon D, Floyd K, Gargioni G, Getahun H, Gilpin C, Glaziou P, Grzemska M, Mirzayev F, Nakatani H, Raviglione M. WHO's new End TB Strategy. *The Lancet*. 2015 may; 385(9979):1799–1801. , doi:10.1016/S0140-6736(15)60570-0.

Varela C, Rittmann D, Singh A, Krumbach K, Bhatt K, Eggeling L, Besra GS, Bhatt A. MmpL Genes Are Associated with Mycolic Acid Metabolism in Mycobacteria and Corynebacteria. *Chemistry & Biology*. 2012 apr; 19(4):498–506. , doi:10.1016/j.chembiol.2012.03.006.

Velayati AA, Masjedi MR, Farnia P, Tabarsi P, Ghanavi J, ZiaZarifi AH, Hoffner SE. Emergence of New Forms of Totally Drug-Resistant Tuberculosis Bacilli. *Chest*. 2009 aug; 136(2):420–425. , doi:10.1378/chest.08-2427.

Viens A, Mechold U, Lehrmann H, Harel-Bellan A, Ogryzko V. Use of protein biotinylation in vivo for chromatin immunoprecipitation. *Analytical Biochemistry*. 2004 feb; 325(1):68–76. , doi:10.1016/j.ab.2003.10.015.

Vilchèze C. Mycobacterial Cell Wall: A Source of Successful Targets for Old and New Drugs. *Applied Sciences*. 2020 mar; 10(7):2278. , doi:10.3390/app10072278.

Viljoen A, Dubois V, Girard-Misguich F, Blaise M, Herrmann JL, Kremer L. The diverse family of MmpL transporters in mycobacteria: from regulation to antimicrobial developments. *Molecular Microbiology*. 2017 jun; 104(6):889–904. , doi:10.1111/mmi.13675.

Villeneuve M, Kawai M, Watanabe M, Aoyagi Y, Hitotsuyanagi Y, Takeya K, Gouda H, Hirono S, Minnikin DE, Nakahara H. Conformational behavior of oxygenated mycobacterial mycolic acids from Mycobacterium bovis BCG. *Biochimica et Biophysica Acta (BBA) - Biomembranes*. 2007 jul; 1768(7):1717–1726. , doi:10.1016/j.bbamem.2007.04.003.

Waldo GS, Standish BM, Berendzen J, Terwilliger TC. Rapid protein-folding assay using green fluorescent protein. *Nature Biotechnology*. 1999 jul; 17(7):691–695. , doi:10.1038/10904.

Walker TM, Kohl TA, Omar SV, Hedge J, Del Ojo Elias C, Bradley P, Iqbal Z, Feuerriegel S, Niehaus KE, Wilson DJ, Clifton DA, Kapatai G, Ip CLC, Bowden R, Drobniewski FA, Allix-Béguec C, Gaudin C, Parkhill J, Diel R, Supply P, et al. Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: a retrospective cohort study. *The Lancet Infectious Diseases*. 2015 oct; 15(10):1193–1202. , doi:10.1016/S1473-3099(15)00062-6.

Wang Z, Fan G, Hryc CF, Blaza JN, Serysheva II, Schmid MF, Chiu W, Luisi BF, Du D. An allosteric transport mechanism for the AcrAB-TolC multidrug efflux pump. *eLife*. 2017 mar; 6:1–19. , doi:10.7554/eLife.24905.

Waugh DS. An overview of enzymatic reagents for the removal of affinity tags. *Protein Expression and Purification*. 2011 dec; 80(2):283–293. , doi:10.1016/j.pep.2011.08.005.

Wells RM, Jones CM, Xi Z, Speer A, Danilchanka O, Doornbos KS, Sun P, Wu F, Tian C, Niederweis M. Discovery of a Siderophore Export System Essential for Virulence of Mycobacterium tuberculosis. *PLoS Pathogens*. 2013 jan; 9(1):e1003120. , doi:10.1371/journal.ppat.1003120.

White JF, Trinh LB, Shiloach J, Grisshammer R. Automated large-scale purification of a G protein-coupled receptor for neurotensin. *FEBS Letters*. 2004 apr; 564(3):289–293. , doi:10.1016/S0014-5793(04)00195-4.

Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, Verma V, Keedy DA, Hintze BJ, Chen VB, Jain S, Lewis SM, Arendall WB, Snoeyink J, Adams PD, Lovell SC, Richardson JS, Richardson DC. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science*. 2018 jan; 27(1):293–315. , doi:10.1002/pro.3330.

Williams JT, Haiderer ER, Coulson GB, Conner KN, Ellsworth E, Chen C, Alvarez-Cabrera N, Li W, Jackson M, Dick T, Abramovitch RB. Identification of New MmpL3 Inhibitors by Untargeted and Targeted Mutant Screens Defines MmpL3 Domains with Differential Resistance. *Antimicrobial Agents and Chemotherapy*. 2019 oct; 63(10):1–18. , doi:10.1128/AAC.00547-19.

Winkler MBL, Kidmose RT, Szomek M, Thaysen K, Rawson S, Muench SP, Wüstner D, Pedersen BP. Structural Insight into Eukaryotic Sterol Transport through Niemann-Pick Type C Proteins. *Cell*. 2019 oct; 179(2):485–497.e18. , doi:10.1016/j.cell.2019.08.038.

Wlodawer A, Dauter Z. 'Atomic resolution': a badly abused term in structural biology. *Acta Crystallographica Section D Structural Biology*. 2017 apr; 73(4):379–380. , doi:10.1107/S205979831700225X.

World Health Organization. Global Tuberculosis Report 2020. Geneva; 2020.

Wright CC, Hsu FF, Arnett E, Dunaj JL, Davidson PM, Pacheco SA, Harriff MJ, Lewinsohn DM, Schlesinger LS, Purdy GE. The Mycobacterium tuberculosis MmpL11 Cell Wall Lipid Transporter Is Important for Biofilm Formation, Intracellular Growth, and Nonreplicating Persistence. *Infection and Immunity*. 2017 aug; 85(8):1–17. , doi:10.1128/IAI.00131-17.

Wu S, Avila-Sakar A, Kim J, Booth DS, Greenberg CH, Rossi A, Liao M, Li X, Alian A, Griner SL, Juge N, Yu Y, Mergel CM, Chaparro-Riggers J, Strop P, Tampé R, Edwards RH, Stroud RM, Craik CS, Cheng Y. Fabs Enable Single Particle cryoEM Studies of Small Proteins. *Structure*. 2012 apr; 20(4):582–592. , doi:10.1016/j.str.2012.02.017.

Xu Z, Meshcheryakov VA, Poce G, Chng SS. MmpL3 is the flippase for mycolic acids in mycobacteria. *Proceedings of the National Academy of Sciences*. 2017 jul; 114(30):7993–7998. , doi:10.1073/pnas.1700062114.

Xue J, Xie T, Zeng W, Jiang Y, Bai Xc. Cryo-EM structures of human ZnT8 in both outward- and inward-facing conformations. *eLife*. 2020 jul; 9:1–17. , doi:10.7554/eLife.58823.

Yamaguchi A, Nakashima R, Sakurai K. Structural basis of RND-type multidrug exporters. *Frontiers in Microbiology*. 2015 apr; 6(APR):1–19. , doi:10.3389/fmicb.2015.00327.

Yamamoto K, Nakata N, Mukai T, Kawagishi I, Ato M. Coexpression of MmpS5 and MmpL5 Contributes to Both Efflux Transporter MmpL5 Trimerization and Drug Resistance in Mycobacterium tuberculosis. *mSphere*. 2021 feb; 6(1):1–11. , doi:10.1128/mSphere.00518-20.

Yamaryo-Botte Y, Rainczuk AK, Lea-Smith DJ, Brammananth R, van der Peet PL, Meikle P, Ralton JE, Rupasinghe TWT, Williams SJ, Coppel RL, Crellin PK, McConville MJ. Acetylation of Trehalose Mycolates Is Required for Efficient MmpL-Mediated Membrane Transport in Corynebacterineae. *ACS Chemical Biology*. 2015 mar; 10(3):734–746. , doi:10.1021/cb5007689.

Yan N. Structural Biology of the Major Facilitator Superfamily Transporters. *Annual Review of Biophysics*. 2015 jun; 44(1):257–283. , doi:10.1146/annurev-biophys-060414-033901.

Yang L, Lu S, Belardinelli J, Huc-Claustre E, Jones V, Jackson M, Zgurskaya HI. RND transporters protect Corynebacterium glutamicum from antibiotics by assembling the outer membrane. *MicrobiologyOpen*. 2014 aug; 3(4):484–496. , doi:10.1002/mbo3.182.

Yang X, Hu T, Yang X, Xu W, Yang H, Guddat LW, Zhang B, Rao Z. Structural Basis for the Inhibition of Mycobacterial MmpL3 by NITD-349 and SPIRO. *Journal of Molecular Biology*. 2020 jul; 432(16):4426–4434. , doi:10.1016/j.jmb.2020.05.019.

Yeagle PL. Non-covalent binding of membrane lipids to membrane proteins. *Biochimica et Biophysica Acta (BBA) - Biomembranes*. 2014 jun; 1838(6):1548–1559. , doi:10.1016/j.bbamem.2013.11.009.

Yen MR, Chen JS, Marquez JL, Sun EI, Saier MH. Multidrug Resistance: Phylogenetic Characterization of Superfamilies of Secondary Carriers that Include Drug Exporters. In: Yan Q, editor. *Methods in Molecular Biology*, vol. 637; 2010.p. 47–64. , doi:10.1007/978-1-60761-700-6_3.

Yokogawa M, Fukuda M, Osawa M. Nanodiscs for Structural Biology in a Membranous Environment. *Chemical and Pharmaceutical Bulletin*. 2019 apr; 67(4):321–326. , doi:10.1248/cpb.c18-00941.

Zabara A, Chong JTY, Martiel I, Stark L, Cromer BA, Speziale C, Drummond CJ, Mezzenga R. Design of ultra-swollen lipidic mesophases for the crystallization of membrane proteins with large extracellular domains. *Nature Communications*. 2018 dec; 9(1):544. , doi:10.1038/s41467-018-02996-5.

Zabara A, Mezzenga R. Plenty of room to crystallize: swollen lipidic mesophases for improved and controlled in-meso protein crystallization. *Soft Matter*. 2012; 8(24):6535–6541. , doi:10.1039/c2sm25249j.

Zampieri V, Gobet A, Robert X, Falson P, Chaptal V. Nanodisc, amphipol or detergent belts in cryoEM reconstructions of membrane proteins are similar and correspond to a common ordered solvent layer. *bioRxiv*. 2020; 21(1):1–9. , doi:10.1101/2020.12.10.418871.

Zavrtanik U, Lukan J, Loris R, Lah J, Hadži S. Structural Basis of Epitope Recognition by Heavy-Chain Camelid Antibodies. *Journal of Molecular Biology*. 2018 oct; 430(21):4369–4386. , doi:10.1016/j.jmb.2018.09.002.

Zeng J, Platig J, Cheng TY, Ahmed S, Skaf Y, Potluri LP, Schwartz D, Steen H, Moody DB, Husson RN. Protein kinases PknA and PknB independently and coordinately regulate essential Mycobacterium tuberculosis physiologies and antimicrobial susceptibility. *PLOS Pathogens*. 2020 apr; 16(4):e1008452. , doi:10.1371/journal.ppat.1008452.

Zhang B, Li J, Yang X, Wu L, Zhang J, Yang Y, Zhao Y, Zhang L, Yang X, Yang X, Cheng X, Liu Z, Jiang B, Jiang H, Guddat LW, Yang H, Rao Z. Crystal Structures of Membrane Transporter MmpL3, an Anti-TB Drug Target. *Cell*. 2019 jan; 176(3):636–648.e13. , doi:10.1016/j.cell.2019.01.003.

Zhang Y, Bulkley DP, Xin Y, Roberts KJ, Asarnow DE, Sharma A, Myers BR, Cho W, Cheng Y, Beachy PA. Structural Basis for Cholesterol Transport-like Activity of the Hedgehog Receptor Patched. *Cell*. 2018 nov; 175(5):1352–1364.e14. , doi:10.1016/j.cell.2018.10.026.

Zheng H, Williams JT, Coulson GB, Haiderer ER, Abramovitch RB. HC2091 Kills Mycobacterium tuberculosis by Targeting the MmpL3 Mycolic Acid Transporter. *Antimicrobial Agents and Chemotherapy*. 2018 jul; 62(7):AAC.02459–17. , doi:10.1128/AAC.02459-17.

Zheng J, Wei C, Zhao L, Liu L, Leng W, Li W, Jin Q. Combining blue native polyacrylamide gel electrophoresis with liquid chromatography tandem mass spectrometry as an effective strategy for analyzing potential membrane protein complexes of Mycobacterium bovis bacillus Calmette-Guérin. *BMC Genomics*. 2011 dec; 12(1):40. , doi:10.1186/1471-2164-12-40.

Zimmermann I, Egloff P, Hutter CAJ, Kuhn BT, Bräuer P, Newstead S, Dawson RJP, Geertsma ER, Seeger MA. Generation of synthetic nanobodies against delicate proteins. *Nature Protocols*. 2020 may; 15(5):1707–1741. , doi:10.1038/s41596-020-0304-x.

Zimmermann I, Egloff P, Hutter CA, Arnold FM, Stohler P, Bocquet N, Hug MN, Huber S, Siegrist M, Hetemann L, Gera J, Gmür S, Spies P, Gygax D, Geertsma ER, Dawson RJ, Seeger MA. Synthetic single domain antibodies for the conformational trapping of membrane proteins. *eLife*. 2018 may; 7:e34317. , doi:10.7554/eLife.34317.

Zivanov J, Nakane T, Forsberg BO, Kimanius D, Hagen WJH, Lindahl E, Scheres SHW. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife*. 2018 nov; 7:1–22. , doi:10.7554/eLife.42166.

Zivanov J, Nakane T, Scheres SHW. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ*. 2019 jan; 6(1):5–17. , doi:10.1107/S205225251801463X.

Zivanov J, Nakane T, Scheres SHW. Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1. *IUCrJ*. 2020 mar; 7(2):253–267. , doi:10.1107/S2052252520000081.

Zoonens M, Popot JL. Amphipols for Each Season. *The Journal of Membrane Biology*. 2014 oct; 247(9-10):759–796. , doi:10.1007/s00232-014-9666-8.

Zoonens M, Zito F, Martinez K, Popot JL. Amphipols: A General Introduction and Some Protocols. Mus-Veteau I, editor, New York, NY: Springer New York; 2014. , doi:10.1007/978-1-4939-0662-8.

Zuber B, Chami M, Houssin C, Dubochet J, Griffiths G, Daffe M. Direct Visualization of the Outer Membrane of Mycobacteria and Corynebacteria in Their Native State. *Journal of Bacteriology*. 2008 aug; 190(16):5672–5680. , doi:10.1128/JB.01919-07.

Zwama M, Yamaguchi A. Molecular mechanisms of AcrB-mediated multidrug export. *Research in Microbiology*. 2018 sep; 169(7-8):372–383. , doi:10.1016/j.resmic.2018.05.005.