

IMPERIAL

Investigation of the oligomeric state of the
mycobacterial lipid transporter, MmpL3:
biophysical and computational insights.

Sara Ciocco

2020 – 2024

Department of Chemistry, Imperial College London

Submitted in part fulfilment of the requirements for the degree of PhD
and the Diploma of Imperial College London

Statement of Originality

This thesis and the work presented are my own, and all external sources have been appropriately referenced and attributed to their originator in the text, bibliography or in footnotes. Whether experiments and relative analyses were carried out by, or in collaboration with, another party, this has been appropriately stated in the relevant Materials and Methods section of each chapter.

Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non-Commercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Abstract

The mycobacterial membrane protein large 3 (MmpL3) is responsible for the transport of key precursor lipids to the periplasm of Mycobacteria. Multiple structures of MmpL3 from *Mycobacterium tuberculosis* (*Mtb*) and *Mycobacterium smegmatis* (*Msmeg*) in various conformational states indicate that the protein is structurally and functionally monomeric ^{1–4}. However, all other resistance, nodulation and cell division transporters structurally characterised to date are either dimeric or trimeric ⁵. In this thesis, an in-depth biophysical and computational analysis of the oligomeric status of *Msmeg* and *Mtb* MmpL3 was carried out. MmpL3 was extracted from the membrane either into styrene maleic acid lipid particles (SMALPs) or purified in detergent and reconstituted into synthetic nanodiscs. Initial evidence on migration pattern in native and denaturing gels, as well as Western blot analyses showed the presence of two distinct MmpL3 populations, likely monomeric and dimeric protein. Analytical ultracentrifugation experiments on size exclusion chromatography fractions confirmed that the protein not only exists as a monomer but also arranges into homodimers, both in SMALPs and synthetic nanodiscs. Furthermore, sucrose gradient separation of MmpL3 populations from *Msmeg* reconstituted into synthetic nanodiscs allowed clear identification of the presence of both protein arrangements using laser induced liquid bead ion desorption (LILBID), a native mass spectrometry technique. Preliminary cryogenic electron microscopy analysis on the same sample confirmed that the protein forms physiological dimers in nanodisc particles. Untargeted lipidomics experiments on MmpL3 co-purified lipids revealed phosphatidylethanolamine and phosphatidylglycerol as the predominant lipid classes. Therefore, Molecular Dynamics simulations were performed on a dimer model embedded in a physiologically relevant lipid composition, revealing the potential protein-protein interface and suggesting that

lipids may play a role in stabilising the dimeric arrangement. In conclusion, all results presented in this work show that *Msmeg* and *Mtb* MmpL3 can exist as homodimers in a variety of membrane mimetic systems.

Acknowledgements

Thank you to the Engineering and Physical Sciences Research Council (EPSRC) and the National Institute for Health and Care Research (NIHR) who funded this PhD.

I would like to thank my supervisors Prof. Bernadette Byrne, Prof. Ian R. Gould and Dr Nick Brooks for their guidance and direction over the past four years. A particular mention to you Bernadette, for your constant support and kindness through the years that have been invaluable.

Thank you to Prof. Tim Dafforn, who welcomed me in his laboratory at Birmingham University. A heartfelt thanks to Dr Naomi Pollock, for having made my time in the lab so enjoyable. I will deeply treasure all the scientific as well as personal discussions that we had, especially the cake-related ones ;). Moreover, a special mention to Pooja Sridhar who kindly run all analytical ultracentrifugation experiments, most of the time at a very short notice. This collaboration turned out to be decisive in the direction of my PhD, thus I am truly grateful.

Thank you to Prof. Nina Morgner and her PhD student Alicia Just, from Frankfurt University for their precious help and advice with native mass spectrometry. A special mention to Dr Zoe Hall and her PhD student Julia Babuta from Imperial College London, who helped me to carry out all lipidomics experiments. A huge thank you to Dr Joseph D. Barritt, who volunteered his time to help me with the cryo-electron microscopy work. Your patience and kindness have been so appreciated, there is not an appropriate number of coffees that could repay your support ;).

To my colleagues in the Gould group, particularly Tony, Hisham and Zhili, who provided advice during my limited time spent at the Molecular Sciences Research Hub, thank you. Thank you to my colleagues/friends in the Byrne research group, starting with Dr Cristina Cecchetti, who helped me settling in the lab during my first year, in tough circumstances due to COVID-19 restrictions. I am glad we met, and I am truly grateful for the lovely friendship we have had since. A special thank you to Dr Nicole Deacon-Smith and Menbere Wubshet: the long days in the lab were much more enjoyable thanks to you, I had lots of fun and loved chatting and gossiping during our MANY coffee breaks. You have my utmost gratitude, and I am so thankful for our friendships. A heartfelt thanks to Dr Natacha Larburu, who basically adopted me and supported me throughout my PhD, by giving me her precious advice and by constantly

motivating me, you have my sincerest appreciation. How not to dedicate a special mention to my CDT friends, Annie, Caroline, Jenna, Will, Tom, with whom I shared this amazing journey and made it fun and enjoyable. To Annie, your immeasurable kindness, constant support and thoughtfulness mean the world to me. To Caroline, for your positive energy, strength and generosity, I am forever grateful. Break Girls forever! I am so fortunate to have you as my closest friends and my PhD would not have been the same without you.

To my dearest friends from Kymab, thank you. Cecilia, the passion and optimism you put into everything you do are inspirational. Thank you deeply to my “Raptors” Morgane and Rachael, you have been a precious source of encouragement and energy-boost over the years, even from afar. To all my other friends who always cheered for me and have been there regardless, you know who you are, so thank you.

To Marco, I cannot express enough how invaluable your unwavering support is to me. Thank you for standing by me through thick and thin, for believing in me without reserve and for always pushing me to give my best.

Finally, to my family, my solid and irreplaceable pillars who I thank from the bottom of my heart for their immense and unconditional love.

Table of contents

Statement of Originality	2
Copyright Declaration.....	2
Abstract.....	3
Acknowledgements	5
Table of contents	7
List of Figures.....	11
List of Tables.....	12
Abbreviations.....	14
1. Introduction.....	19
1.1. Tuberculosis disease and its causative agent.....	19
1.1.1 Mycobacteria phylogeny and its origin	19
1.1.2 Overview of TB infection	21
1.1.3 Global prevalence of TB.....	22
1.1.4 Current treatments against TB and rise of antibiotic-resistant strains.....	23
1.1.5 The cell envelope as essential virulence factor for Mycobacteria	25
1.2. MmpL3 as potential target for the development of novel anti-tubercular agents	28
1.2.1 The Resistance-Nodulation-Cell Division proteins superfamily.....	28
1.2.2 MmpL3 as a druggable target	29
1.2.3 MmpL3 structure and transport mechanism.....	29
1.3. Membrane proteins and their environment	33
1.3.1 Lipids and the biological membrane.....	35
1.3.2 Detergents.....	42
1.3.3 MP reconstitution into synthetic nanodisc systems	43
1.3.4 Polymers	46
1.4. Experimental approaches for MP characterisation	48
1.4.1 Separative Techniques.....	48
1.4.2 Structural determination of MPs	49

1.4.3	Biophysical characterisation.....	51
1.4.3.1	Study of MPs dynamics	51
1.4.3.2	Oligomeric arrangements of MPs	51
1.4.3.3	MPs stability.....	54
1.5.	Computational approaches for MP characterisation.....	55
1.6.	Summary and research aims	60
2.	MmpL3 protein expression and purification.....	62
2.1	Introduction	62
2.2	Materials and Methods	62
2.2.1	Cloning of full-length <i>Msmeg</i> and <i>Mtb</i> MmpL3.....	62
2.2.2	Bacterial expression of <i>Msmeg</i> and <i>Mtb</i> MmpL3	63
2.2.3	MmpL3 purification in detergent.....	63
2.2.4	Use of BL21(DE3) Δ acrABEF as new bacterial expression strain	64
2.2.5	Cloning of <i>Msmeg</i> and <i>Mtb</i> MmpL3 Δ C-terminal constructs	65
2.2.6	Membrane scaffold protein 1 E3D1 (MSP1E3D1) expression	66
2.2.7	Membrane scaffold protein 1 E3D1 (MSP1E3D1) purification	66
2.2.8	<i>Msmeg</i> and <i>Mtb</i> MmpL3 FL and Δ C reconstitution into nanodiscs	67
2.2.9	<i>Msmeg</i> and <i>Mtb</i> MmpL3 FL and Δ C into SMALPs.....	68
2.2.10	PAGE gels.....	69
2.2.11	Western Blot.....	69
2.2.12	Assessment of protein stability.....	70
2.2.13	Protein ID analysis through Mass Spectrometry (MS)	71
2.3	Results	72
2.3.1	<i>Msmeg</i> and <i>Mtb</i> MmpL3 purification in detergent	72
2.3.1.1	<i>Msmeg</i> MmpL3 full-length expression in <i>E. coli</i> BL21(DE3) WT	72
2.3.1.2	Optimisation of detergent-based purification	73
2.3.1.3	End product contamination with bacterial protein	76
2.3.1.4	Expression in <i>E. coli</i> BL21(DE3) Δ acrABEF	78
2.3.1.5	Generation of C-terminally truncated constructs.....	78
2.3.2	<i>Msmeg</i> and <i>Mtb</i> MmpL3 reconstitution into nanodiscs	80
2.3.2.1	Standard nanodiscs reconstitution process	80
2.3.2.2	Optimisation of nanodiscs reconstitution process.....	82
2.3.3	Protein stability assessment: MmpL3 in detergent micelles and NDs.....	83
2.3.3.1	Thermostability assay	83
2.3.3.2	Thermal denaturation analysis.....	85

2.3.4	<i>Msmeg</i> MmpL3 into SMALPs.....	86
2.4	Discussion	88
3.	Biophysical investigation on MmpL3 oligomeric state	93
3.1	Introduction	93
3.2	Materials and Methods	93
3.2.1	Sedimentation velocity analytical ultracentrifugation (SV-AUC).....	93
3.2.2	Nanodiscs separation through sucrose gradient ultracentrifugation	94
3.2.3	Mass photometry on <i>Msmeg</i> MmpL3 Δ C and empty NDs	94
3.2.4	Native MS.....	95
3.2.5	Cryo-EM data collection and processing.....	96
3.2.6	Lipidomics	96
3.3	Results	97
3.3.1	MmpL3 is present in two populations in SMALPs and MSP-based NDs	97
3.3.2	Mass photometry on <i>Msmeg</i> MmpL3 NDs	99
3.3.3	Native MS analysis confirms observations of <i>Msmeg</i> MmpL3 oligomers	104
3.3.4	Preliminary cryo-EM analysis shows <i>Msmeg</i> MmpL3 Δ C exists as a dimer	105
3.3.5	Lipidomics analysis show lipid species are retained during MmpL3 purification in DDM, MSP-NDs and SMALPs	108
3.4	Discussion	111
4.	Investigation of monomeric and dimeric <i>Msmeg</i> and <i>Mtb</i> MmpL3 using Molecular Dynamics (MD) simulations	115
4.1	Introduction	115
4.2	Materials and Methods	116
4.2.1	<i>Msmeg</i> MmpL3 crystal structures evaluation	116
4.2.2	<i>Msmeg</i> MmpL3 Δ C model building for MD	116
4.2.3	<i>Mtb</i> MmpL3 Δ C model building for MD.....	117
4.2.4	Oligomeric structure prediction for <i>Msmeg</i> and <i>Mtb</i> MmpL3 Δ C constructs	117
4.2.5	MD simulations settings	118
4.2.6	Analysis of MD trajectories.....	119
4.3	Results	120
4.3.1	Selection of reference crystal structures for <i>Msmeg</i> and <i>Mtb</i> MmpL3	120
4.3.2	<i>Msmeg</i> and <i>Mtb</i> MmpL3 monomeric models and MD simulation results.....	122
4.3.3	RMSd, B-factors and RMSf analyses on monomeric simulations	123

4.3.4	Periplasmic cavity size changes according to protein-ligand binding state.....	128
4.3.5	Analysis of protein-lipid substrate interactions within the central periplasmic cavity	131
4.3.6	Analysis of protein-lipid binding sites on TM domains.....	134
4.3.7	Investigation into MmpL3 oligomerisation state	137
4.3.8	Dimer interface analysis on GalaxyWeb models.....	138
4.3.9	MD simulations on dimeric <i>Msmeg</i> MmpL3	142
4.3.10	Protein-protein interactions	142
4.3.11	Periplasmic domain dynamics in dimer simulations	144
4.3.12	Protein-lipid hydrogen bonds analysis in dimer simulations.....	146
4.4	Discussion	148
4.4.1	Results on monomeric MmpL3.....	148
4.4.2	Results on dimeric MmpL3.....	151
4.4.3	Conclusions.....	152
5.	Final discussion	153
6.	Bibliography	157
7.	Appendix: Copyrights and permissions.....	180
7.1	Approval 1	181
7.2	Approval 2	182
7.3	Approval 3	183
7.4	Approval 4	184

List of Figures

Figure 1.1. <i>Mycobacterium tuberculosis</i> phases of infection.	20
Figure 1.2. Estimated number of incident TB cases in 2022..	24
Figure 1.3. The structural architecture of <i>Mtb</i> cell envelope and MmpL3 suggested mechanism of lipid transport.	24
Figure 1.4. <i>Msmeg</i> MmpL3 topology.	31
Figure 1.5. <i>Msmeg</i> MmpL3 structure.	32
Figure 1.6. Fluid mosaic model of the cell membrane.	35
Figure 1.7. Summary of approaches to preparing membrane protein sample suitable for downstream analysis.	36
Figure 1.8. Structure of a phospholipid.	37
Figure 1.9. Types, curvature and self-assembly of lipids.	38
Figure 1.10. Lipid interactions with membrane proteins.	39
Figure 1.11. Collective membrane physical properties.	40
Figure 1.12. Example structures of detergents, polymers and amphipols used to stabilise membrane proteins.	41
Figure 1.13. Graphical summary of some of the common experimental techniques used for the detection and characterisation of membrane proteins.	49
Figure 1.14. Different motions in proteins and their relative time scale.	56
Figure 1.15. Evolution of membrane models from the pioneering stage of 30 years ago up to the realistic complexity of current times.	58
Figure 1.16. Potential energy function at the core of AMBER force field.	59
Figure 2.1. MmpL3 expression and detergent-based purification work-flow.	64
Figure 2.2. MmpL3 DDM-based purification and NDs reconstitution processes.	68
Figure 2.3. Protein stability assessment.	71
Figure 2.4. Analysis of <i>Msmeg</i> MmpL3 ΔC produced in the $\Delta acrABEF$ strain at different time points post induction.	73
Figure 2.5. <i>Msmeg</i> MmpL3 FL purification using Ni^{2+} -NTA resin and varying Imidazole concentration for column washes.	74
Figure 2.6. <i>Mtb</i> MmpL3 FL purification using Ni^{2+} -NTA resin and varying imidazole concentration for column washes.	77
Figure 2.7. <i>Mtb</i> and <i>Msmeg</i> MmpL3 FL and ΔC -term purification in DDM and relative yields.	80
Figure 2.8. MmpL3 NDs reconstitution.	81
Figure 2.9. SDS-PAGE gel and anti-His WB of samples obtained during DDM-based purification and NDs reconstitution of FL and ΔC MmpL3 from <i>Msmeg</i> and <i>Mtb</i>	83
Figure 2.10. Thermostability test on <i>Msmeg</i> and <i>Mtb</i> MmpL3.	84
Figure 2.11. CPM assay results.	87
Figure 2.12. <i>Msmeg</i> MmpL3 FL and ΔC into SMALPs.	88
Figure 3.1. SV-AUC results for <i>Msmeg</i> MmpL3 ΔC in SMALPs and MSP-NDs.	98
Figure 3.2. <i>Msmeg</i> MmpL3 ΔC NDs and empty NDs purification using a Superose® 6 column.	100

Figure 3.3. Sucrose density gradient ultracentrifugation separation of <i>Msmeg</i> MmpL3 Δ C NDs and empty NDs.....	101
Figure 3.4. Mass photometry on <i>Msmeg</i> MmpL3 Δ C and empty NDs.	103
Figure 3.5. Native Mass Spectrometry for <i>Msmeg</i> MmpL3 Δ C NDs.	105
Figure 3.6. Second sucrose density gradient separation of <i>Msmeg</i> MmpL3 Δ C NDs and empty NDs.	106
Figure 3.7. 2D classifications of monomeric and dimeric assemblies of MmpL3.	107
Figure 3.8. Cryo-EM analysis workflow.	108
Figure 3.9. Lipidomics analysis on <i>Msmeg</i> and <i>Mtb</i> MmpL3 protein in detergent, MSP-NDs and SMALPs.	109
Figure 3.10. Abundance of lipid species in <i>Msmeg</i> MmpL3 Δ C in DDM, POPC NDs and SMALPs.	111
Figure 4.1. 6OR2 and 6AJF superposition.	121
Figure 4.3. Membrane assembling using CHARMM-GUI Membrane Builder.	124
Figure 4.4. RMSd plots for 1.16-6OR2 repeats.	125
Figure 4.5. B-factors and RMSf analysis for <i>Msmeg</i> MmpL3	126
Figure 4.6. B-factors and RMSf analysis for <i>Mtb</i> MmpL3.	127
Figure 4.7. 1.16-6OR2 residues with highest RMSf values.	128
Figure 4.8. <i>Msmeg</i> MmpL3 1.16 cavity size comparison to crystal structure.	129
Figure 4.9. PND – PCD centre of mass (c.o.m.) distance measurements during MD simulations..	130
Figure 4.10. 6AJF, 6OR2 and 7NVH structures with co-crystallised ligands.....	132
Figure 4.11. <i>Msmeg</i> MmpL3 simulations including POPC, POPE and SOPE as ligands: H-bonds analysis results.	133
Figure 4.12. MmpL3 lipid binding pockets.	135
Figure 4.13. <i>Mtb</i> MmpL3 protein-lipid hydrogen bonds analysis.	136
Figure 4.14. GalaxyHomomer predicted <i>Msmeg</i> MmpL3 dimeric structure.	139
Figure 4.15. Hydrophobic surface properties of <i>Msmeg</i> MmpL3 and <i>Burkholderia multivorans</i> HpnN	140
Figure 4.16. The interaction interface of the MmpL3 dimer.	141
Figure 4.17. Protein-protein interactions for <i>Msmeg</i> MmpL3 dimer.....	144
Figure 4.18. PD dynamics in <i>Msmeg</i> MmpL3 dimer simulations.	145
Figure 4.19. <i>Msmeg</i> MmpL3 dimer: protein-lipid interactions.	147
Figure 4.20. <i>Msmeg</i> MmpL3 dimer: protein-lipid interactions on top leaflet	147
Figure 4.21. <i>Msmeg</i> MmpL3 dimer: hydrogen bond analysis on protein-lipid interactions on top leaflet.	149
Figure 4.22. <i>Msmeg</i> MmpL3 dimer: protein-lipid interactions on bottom leaflet.	150

List of Tables

Table 1. MmpL3 oligomeric structures predicted by GalaxyHomomer server.....	138
---	-----

Table 2. $\Delta\Delta G_{\text{bind}}$ values obtained from computational alanine scanning for <i>Msmeg</i> MmpL3 dimer model	143
--	-----

Abbreviations

2D	–	Two-dimensional
3D	–	Three-dimensional
6-DDTre	–	6-n-dodecyl- α , α -trehalose
Δ C	–	C-terminally truncated
AcrB	–	Acriflavine resistance B
AcrD	–	Acriflavine resistance D
AFM	–	Atomic force microscopy
AMBER	–	Assisted Model Building with Energy Refinement
apo A-I	–	Apolipoprotein A-I
a.u.	–	Arbitrary Units
AUC	–	Analytical ultracentrifugation
BAM	–	β -barrel assembly machinery
BCG	–	Bacille Calmette-Guérin
BSA	–	Bovine Serum Albumin
Ca	–	Carbon- α
CADD	–	Computer Aided Drug Design
CDL	–	Cardiolipin
CG	–	Coarse grain
CHARMM	–	Chemistry at Harvard Molecular Mechanics
CHS	–	Cholesterol hemisuccinate
CMC	–	critical micelle concentration
c.o.m.	–	Centre of mass
COVID-19	–	Coronavirus disease 2019
CPM	–	7-diethylamino-3-(4'-maleimidylphenyl)-4-methylcoumarin
cryo-EM	–	Cryogenic electron microscopy
cryo-ET	–	Cryo-electron tomography
CTD	–	C-terminal domain
CTF	–	Contrast transfer function
CV	–	Column volumes
CYP450	–	Cytochrome P450
DDM	–	n-dodecyl- β -D-maltopyranoside
DEER	–	Double electron–electron resonance
DIBMA	–	Diisobutylene maleic anhydride
DM	–	n-decyl-beta-maltoside
DMNG	–	Decyl maltose neopentyl glycol
DMSO	–	Dimethylsulfoxide
DNA	–	Deoxyribonucleic acid
DNaseI	–	Deoxyribonuclease I
DOPC	–	1,2-dioleoyl-sn-glycero-3-phosphocholine
DPPC	–	dipalmitoyl-phosphatidylcholine
DSF	–	Differential scanning fluorimetry

EDTA	–	Ethylenediamine tetraacetic acid
emPAI	–	Exponentially modified protein abundance index
ENA	–	European Nucleotide Archive
EPSRC	–	Engineering and Physical Sciences Research Council
FcγRIIB	–	Fc gamma receptor IIB
FF	–	Force field
FL	–	Full length
FT	–	Flow-through
GDN	–	Glyco-diosgenin
GNG	–	Glucose-neopentyl glycol
GPCR	–	G-protein coupled receptor
GPU	–	Graphics processing unit
GUI	–	Graphical user interface
hENT1	–	Human equilibrative nucleoside transporter 1
HIV	–	Human immunodeficiency virus
HpnN	–	Hopanoid biosynthesis-associated RND transporter
HRP	–	Horseradish peroxidase
HSP	–	High sucrose percentage
ICA	–	Indolcarboxamide
IL-6	–	Interleukin 6
IM	–	Inner membrane
IMAC	–	Immobilised metal affinity chromatography
IMP	–	Integral membrane proteins
IPTG	–	Isopropyl β-d-1-thiogalactopyranoside
IS	–	Internal Standards
iSCAMS	–	Interferometric scattering mass spectrometry
KatG	–	Catalase-peroxidase
Kd	–	Dissociation constant
kDa	–	Kilodalton
keV	–	Kiloelectron volt
KOAc	–	Potassium acetate
kV	–	Kilovolt
LB	–	Luria-Bertani broth
LC	–	Liquid chromatography
LDAO	–	Lauryldimethylamine-N-oxide
LILBID	–	Laser-induced liquid bead ion desorption
LMNG	–	Lauryl maltose neopentyl glycol
LSP	–	Low sucrose percentage
LX-MS	–	Lipidomic lipid-exchange mass spectrometry
MA	–	Mycolic acid
mAGP	–	Mycolic arabinogalactan polysaccharides
MAS	–	Magic angle spinning
MCP-1	–	Monocyte chemoattractant protein-1

MD	–	Molecular dynamics
MDR	–	Multidrug resistance
MFS	–	Major facilitator superfamily
Mincle	–	Monocyte-inducible C-type lectin
MmpL3	–	Mycobacterial membrane protein Large 3
MP	–	Membrane protein
MS	–	Mass spectrometry
MscL	–	Mechanosensitive channel of large conductance
<i>Msmeg</i>	–	<i>Mycobacterium smegmatis</i>
MSP	–	Membrane scaffold protein
<i>Mtb</i>	–	<i>Mycobacterium tuberculosis</i>
MTBC	–	<i>Mycobacterium tuberculosis complex</i>
MW	–	Molecular weight
MWCO	–	Molecular weight cut off
N/A	–	Not available
NaCl	–	Sodium chloride
Nano-DSF	–	Nano-differential scanning fluorimetry
Nano-ESI	–	Nano-electrospray ionisation
ND	–	Nanodisc
NIHR	–	National Institute for Health and Care Research
Ni ²⁺ -NTA	–	Nickel-nitrilotriacetic acid
NMR	–	Nuclear magnetic resonance
nMS	–	Native mass spectrometry
NPC1	–	Human Niemann-Pick C1
NSP	–	Nanodisc scaffold peptide
NTM	–	Non-tuberculous mycobacteria
NTS-1	–	Neurotensin 1
OD ₆₀₀	–	Optical density at 600 nm
OG	–	Octyl-beta-glucoside
OM	–	Outer membrane
PA	–	Phosphatidic acid
PBC	–	Periodic boundary conditions
PBS	–	Phosphate buffered saline
PC	–	Phosphatidylcholine
PCD	–	C-terminal periplasmic domain
PCR	–	Polymerase chain reaction
PD	–	Periplasmic domain
PDB	–	Protein Data Bank
PE	–	Phosphatidylethanolamine
PG	–	Phosphatidylglycerol
pH	–	Potential of hydrogen
PI	–	Phosphatidylinositol
PMA	–	Polymethacrylate

PME	–	Particle mesh Ewald
PND	–	N-terminal periplasmic domain
POPC	–	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphatidylcholine
POPE	–	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine
PRR	–	Pattern recognition receptor
PS	–	Phosphatidylserine
RMSd	–	Root-Mean-Square deviations
RMSf	–	Root-Mean-Square fluctuations
RND	–	Resistance, nodulation and cell division
RR-TB	–	Rifampicin-resistant tuberculosis
SANS	–	Small-angle neutron scattering
SAXS	–	Small-angle X-ray scattering
SDS	–	Sodium dodecyl-sulfate
SDS-PAGE	–	Sodium dodecyl-sulfate polyacrylamide gel electrophoresis
SE	–	Sedimentation equilibrium
SEC	–	Size-exclusion chromatography
SEC-MALS	–	Size-exclusion chromatography multi-angle light scattering
SERCA	–	Sarco-endoplasmic reticulum calcium ATPase
SHP-1	–	Src homology region 2 domain-containing phosphatase 1
SM	–	Sphingomyelin
SMA	–	Styrene maleic acid
SMALP	–	Styrene maleic acid lipid particle
SMA-PAGE	–	Styrene maleic acid polyacrylamide gel electrophoresis
SOPE	–	1-stearoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine
STB	–	Smooth tubercle bacilli
SV	–	Sedimentation velocity
T6D	–	Trehalose 6-decanoate
TB	–	Tuberculosis
TDM	–	Trehalose dimycolates
TEM	–	Transmission electron microscope
TEV	–	Tobacco etch virus
T _m	–	Melting Temperature
TM	–	Transmembrane
TMC	–	C-terminal transmembrane domain
TMD	–	Transmembrane domain
TMM	–	Trehalose monomycolate
TMN	–	N-terminal transmembrane domain
TNF- α	–	Tumour necrosis factor alpha
Tris-HCl	–	Tris(hydroxymethyl)aminomethane hydrochloride
UV	–	Ultraviolet
WB	–	Western blot
WHO	–	World Health Organisation
WT	–	Wild type

XDR	–	Extensively drug-resistant
zDOPE	–	Discrete Optimised Protein Energy score

1. Introduction

1.1. Tuberculosis disease and its causative agent

Tuberculosis (TB) is an old disease that has affected humans for thousands of years ^{6,7}. Its aetiological agent remained unknown until Dr Robert Koch announced the discovery of the bacillus *Mycobacterium tuberculosis* (*Mtb*), in 1882 ⁸. *Mtb* is an intracellular obligate pathogen which mainly infects lung-resident macrophages causing the formation of granulomas (Figure 1.1). These are nodules composed of pulmonary tissue, bacteria and immune cells recruited to the infection site. After establishing the initial infection *Mtb* can trigger severe tissue damage and spread to other hosts by the airborne particles which are generated through coughing. However, *Mtb* can also remain latent in the granuloma stage for many years without showing any symptoms and stay undetected ^{6,8–10}. For this reason, it was estimated by a modelling study in 2016 that about a quarter of the world's population had been infected with *Mtb* ^{5,6}. However, recent analyses suggest that this number is much lower if we account for currently infected people, and if we consider that only about 5 – 10% of these will go on to develop TB disease at some point during their lifetime ^{13,14}. After being infected with TB, the first 2 year-window presents the highest risk of developing the disease, after which some people will clear the infection ¹⁵.

1.1.1 Mycobacteria phylogeny and its origin

The genus *Mycobacterium* includes over 170 different species and is thought to have existed since the Jurassic period ^{16,17}. Traditionally, *Mycobacterium* species have been divided into slow-growers and fast-growers, with the three major mycobacterial pathogens of humans (*Mycobacterium tuberculosis* complex or MTBC, *Mycobacterium leprae* and *Mycobacterium ulcerans*) belonging to the slow-growers group ¹⁸. In addition, immune-compromised individuals can be infected by several so-called non-tuberculous mycobacteria (NTMs) and develop the disease. These include *Mycobacterium abscessus* (a fast-grower) and *Mycobacterium avium*, *Mycobacterium marinum*, *Mycobacterium xenopi*, *Mycobacterium goodii* and *Mycobacterium kansasii* (all slow-growers). However, the human-adapted MTBC, which is believed to

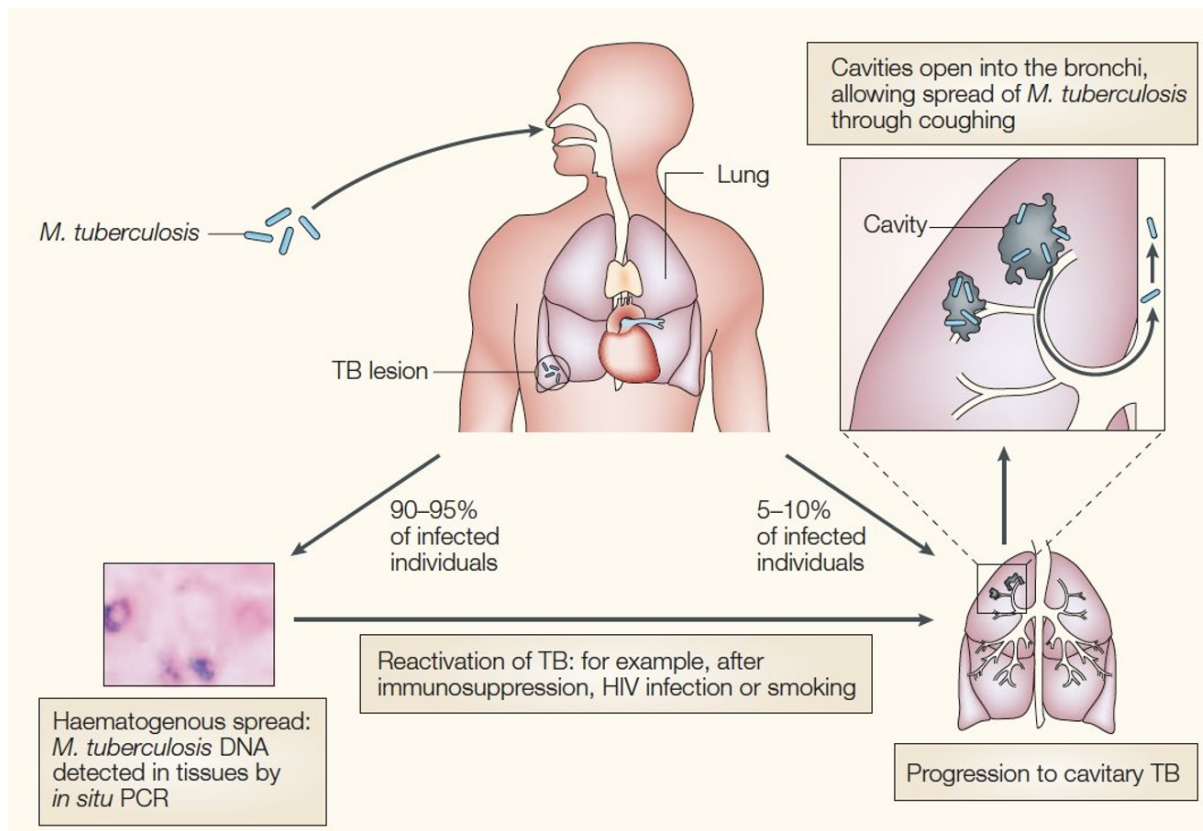


Figure 1.1. *Mycobacterium tuberculosis* phases of infection. Tuberculosis (TB) is spread by aerosols and can establish infection with or without a visible primary lesion. Such lesion tends to be sited towards the lung base and close to the pleura. In 90–95% of individuals, the infection remains latent, which is probably reflected by the pathogen’s DNA that can be detected in tissue by *in situ* PCR. In 5-10% of individuals, the disease progresses, most often in the lung apex, where the ratio of ventilation to blood perfusion is highest. Although in most infected people the disease remains latent for decades, it can be reactivated when an individual is immunosuppressed, particularly through HIV infection. Progressive disease is characterized by weight loss, toxicity of tumour-necrosis factor, fibrosis and cavitation. The cavities eventually open into the bronchi, allowing the spread of TB by aerosols during coughing. Reprinted with permission ⁵. Copyright 2005, Springer Nature.

have evolved from a *Mycobacterium canettii*-like ancestor into an obligate pathogen of humans, is unique as it has to cause pulmonary disease to transmit between individuals ^{19,20}. This is unlike many other pathogens, whose transmission is not directly linked to virulence ²¹. Furthermore, MTBC does not need environmental or animal reservoirs as a direct consequence of its local adaptation to specifically manipulate the human immune system to promote its replication and spread ¹⁶. Indeed, TB disease is mainly caused by a disproportionate inflammatory immune response to antigenic load initiated by the host. This results in lung tissue damage and granuloma necrosis, dramatically increasing the transmission potential of the pathogen ²². Recent studies have highlighted how, in contrast to many other pathogens that evade host immunity by antigenic variation, the large majority of known human T-cell epitopes in the MTBC are hyperconserved ^{23–25}. This suggests that the pathogen adopts a strategy of

immune subversion by eliciting activation of the host's immune responses to benefit its transmission potential.

Most evidence to date supports Africa as the continent of origin for the human-adapted MTBC, especially given that smooth tubercle bacilli (STB), the closest living relatives of the common ancestor of the MTBC, are almost exclusively found in the Horn of Africa ^{20,26}. Seven human-adapted MTBC lineages have been identified, with some occurring globally and others showing a strong geographical restriction. Africa is the only continent that harbours all seven of the human-adapted MTBC lineages and thus retains the largest diversity. This diversity decreases with increasing distance from the African continent ²⁷. Paleomicrobiology studies have suggested that TB was carried out of Africa through migrations during the Neolithic period, as well as dating the interaction between mycobacteria and humans to approximately 73,000 years ago ²⁸.

1.1.2 Overview of TB infection

The first entry points for *Mtb* are constituted by the nose or the mouth, from where the pathogen then travels through the upper and lower airways to arrive in the alveolar space ²⁸. Here, it infects the specialised innate immune cells known as alveolar macrophages that reside in the alveoli and that protect the host from the airborne particulates and pathogens these air sacs are often exposed to ²⁹. Furthermore, during primary infection *Mtb* also infects dendritic cells found within the interstitial space and that can also interact with airborne particles ³⁰. Thus, these innate immune cells serve as early reservoirs of infection and are used by the pathogen to activate an adaptive immune response. In particular, the infected macrophages in addition to infected dendritic cells can migrate to draining lymph nodes and recruit T and B cells that function to limit the infection ^{28,31,32}. Infected macrophages along with non-infected macrophages, inflammatory monocytes, neutrophils, and T cells recruited as a result of the associated inflammation and tissue damage then form the characteristic TB granuloma. Interestingly, despite the fact that a granuloma limits progression of the infection, it also ensures the bacteria are maintained in a latent state and thus avoid clearance by the immune system ^{28,33}. From this stage, active infection can develop through the granuloma liquefaction phenomenon ³⁴. When this occurs, a central necrotic material known as caseum is formed in tuberculous lesions which provides an even more favourable growth environment for *Mtb*. From these caseous foci that open up to an airway, *Mtb* can then spread beyond the lung and reach naive

hosts through airborne transmission ³⁵. During active TB infection and in the absence of treatment, a chronic and bloody cough is produced through pathogen-mediated stimulation of airway nociceptive neurons. This characteristic and persistent cough is also what allows *Mtb* to propagate infection and renew its life cycle in a new uninfected host ³⁶.

1.1.3 Global prevalence of TB

In May 2014, the World Health Assembly passed a resolution, approved with full support from its 194 Member States, the new post-2015 End TB Strategy, a target under the Sustainable Development Goals. Three pillars constitute the foundation for the success of the End TB Strategy: 1) integrated, patient-centred care and prevention; 2) bold policies and supportive systems; 3) intensified research and innovation. All three aspects are critical to break the trajectory of the TB epidemic and reach the ambitious targets of zero deaths, disease, and suffering due to TB ³⁷.

Every year, the World Health Organisation (WHO) publishes a detailed and up-to-date assessment of the TB epidemic and progress in prevention, diagnosis and treatment of the disease. The 2023 Global Tuberculosis Report ranked TB as one of the top 10 causes of death worldwide and as the second deadliest infectious disease caused by a single agent after Coronavirus disease 19 (COVID-19) ³⁸. According to the report, 10.6 million people fell ill with TB in 2022 globally, of which 55% of cases were men, 33% women and 12% were children aged 0 - 14. A 3.9% increase in incidence rate was observed between 2020 and 2022, compared to the average 2% decrease per year recorded over the past two decades. One of the main causes for this setback was the COVID-19 pandemic that had a significant impact on access and provision of healthcare services, delaying diagnosis and treatment of people who developed TB in those years. In 2022, 1.3 million people died from TB worldwide, including 167,000 HIV-positive people, and the total number of deaths was close to the value recorded pre-pandemic, in 2019 ³⁸.

Once infected with *Mtb*, the probability of developing the disease dramatically increases in HIV-positive people and among people affected by risk factors such as undernutrition, diabetes, smoking and alcohol consumption. In fact, TB is a disease tightly linked to poverty, which also brings stigma and discrimination for whoever suffers from it. Furthermore, the COVID-19 pandemic had an economic impact on TB disease burden, by worsening the overall social situation of underdeveloped countries,

at the same time by increasing undernourishment and by reducing the income per capita. In particular, about half of TB patients and their households would spend over 20% of their annual income only to face TB-related costs. This economic barrier has therefore catastrophic effects on the access to and completion of TB treatment, consequently increasing the incidence rates. Almost 90% of all TB cases in 2022 were reported in 30 high TB burden countries, with two-thirds of the total cases recorded in eight countries: India (27%), Indonesia (10%), China (7.1%), the Philippines (7.0%), Pakistan (5.7%), Nigeria (4.5%), Bangladesh (3.6%) and the Democratic Republic of the Congo (3.0%) (Figure 1.2) ³⁸.

Despite the efforts put into place in the fight against TB, its eradication is far from complete, requiring more investments, research and patient-centred solutions in the years to come to reach the ultimate WHO target of zero deaths ³⁸.

1.1.4 Current treatments against TB and rise of antibiotic-resistant strains

Due to the seriousness of the global TB situation, major efforts have focused on the development of both a new vaccine and new therapies to tackle TB once and for all. This is crucial as since the approval of the only available TB vaccine, the bacille Calmette-Guérin (BCG) vaccine, developed more than a century ago, and the first- and second-line antibiotics, including isoniazid, rifampicin, fluoroquinolones and ethionamide developed in the 1940s-1960s, only two new drugs, bedaquiline and delamanid, have been approved recently, in 2012 and 2014, respectively ^{6,16}.

Without treatment, the chances of dying from TB disease are high (about 50%) ³⁹. The latest WHO guidelines strongly suggest the use of a 6-month regimen of isoniazid, rifampicin, ethambutol and pyrazinamide for people with drug-susceptible TB (both pulmonary and extrapulmonary) with 88% treatment success rates recorded. Despite the high cure rate, the surge in drug resistance cases is alarming ^{38,40}.

Mtb shows low genetic diversity compared to other bacteria, mostly as a result of a finely evolved genome that has allowed the pathogen to adapt to its human host, a scenario where most emerging *de novo* mutations would likely be deleterious and are therefore lost each time *Mtb* transmits to a new host ¹⁶. However, despite a very low mutation rate (0.3 – 0.5 substitutions per genome per year), the lack of ongoing horizontal gene transfer and resistance plasmids, antibiotic resistance in *Mtb* can arise quickly ^{41,42}. In clinical settings, mutations conferring resistance to drugs that present little or no fitness cost tend to be preferentially selected ⁴³. When these mutations are

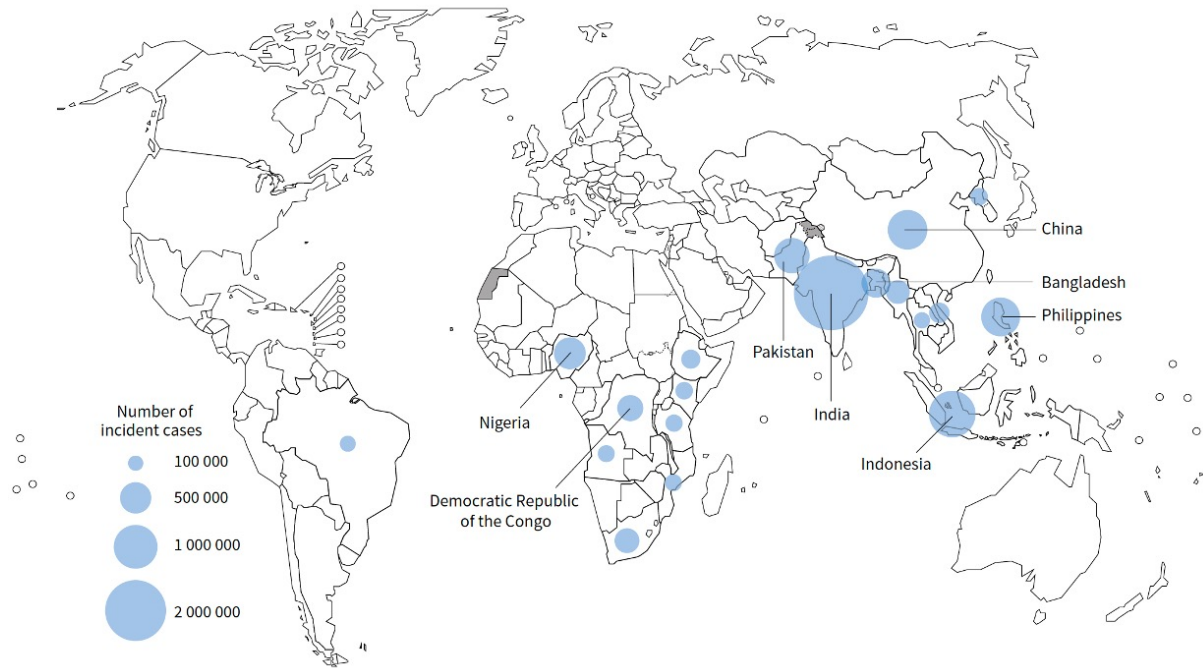


Figure 1.2. Estimated number of incident TB cases in 2022. Countries with at least 100,000 incident cases are indicated. The eight countries ranked in order from first to last in terms of numbers of cases, and that accounted for about two thirds of global cases in 2022, are India, Indonesia, China, the Philippines, Pakistan, Nigeria, Bangladesh and the Democratic Republic of Congo. Reprinted with permission ³³. Copyright 2023, WHO.

associated with a fitness cost for the pathogen (e.g. cell growth), secondary compensatory mutations can overcome the initial fitness deficits by interacting epistatically with the resistance mutation ⁴⁴. These evolutionary mechanisms highlight the remarkable capability of *Mtb* to potentially develop drug resistance against all efficacious drugs at low or no cost.

Drug resistance is usually conferred by the intrinsically impenetrable mycobacterial cell wall, discussed later in this Chapter, and by acquired chromosomal mutations ^{45,46}. These mutations cause drug target modification or overexpression, prevention of prodrug activation, as well as the upregulation of efflux pumps as a result of a specific environmental cue ^{47,48}. The number of drug-resistant *Mtb* strains against many classes of antibiotics is rapidly increasing, mainly as a result of the scarcity of novel antituberculosis compounds, the administration of multiple antibiotics in combination, as well as the prolonged use of these drugs at subinhibitory concentrations. These factors have indirectly facilitated the evolution of strains concomitantly resistant to several drugs used for treatment ⁴⁵.

Globally, 410,000 people developed rifampicin-resistant TB (RR-TB) or multidrug-resistant TB (MDR-TB, defined as resistance to isoniazid and rifampicin) in

2022, with three countries accounting for 42% of cases: India (27%), the Philippines (7.5%) and the Russian Federation (7.5%)³⁸. Treatment for people diagnosed with these variants consists of a new 6-month regimen including bedaquiline, pretomanid, linezolid and moxifloxacin, referred to as BPaLM. The same regimen but without moxifloxacin (BPaL) can be used for people with pre-extensively drug-resistant TB (pre-XDR-TB), defined as TB that is resistant to rifampicin and any fluoroquinolone⁴⁹. Both of these regimens recorded a global average of treatment success rate for drug-resistant TB of 63% in the most recent patient cohort. Nevertheless, as stated in the 2023 Global Tuberculosis Report, the number of people diagnosed and started on treatment was only equivalent to about 40% of those in need and still below pre-pandemic levels. To complicate the matter even further, treatment efficacy remains very low for XDR-TB, indicating resistance to rifampicin, any fluoroquinolone and at least one of bedaquiline or linezolid³⁸.

As of August 2023, there were 28 drug candidates and 16 vaccines in various stages of clinical trials, highlighting the increased attention given to this disease, especially over the last decade⁵⁰.

Alongside the search for new drugs, it is also important to screen for potential new targets, especially in the fight against antibiotic-resistant strains, which are becoming more common.

1.1.5 The cell envelope as essential virulence factor for *Mycobacteria*

The structural architecture of *Mtb*, and specifically its waxy and impenetrable outer membrane, distinguishes mycobacteria from other bacterial species (Figure 1.3). The cell envelope acts as a virulence factor for its crucial role in bacterial growth as well as for its interaction with the host immune system. Mycolic acids (MAs) are very long-chain (C₆₀-C₉₀) fatty acids which constitute the main components of the outer membrane (OM) of mycobacteria, accounting for around 40-60% of the cell's dry weight⁵¹⁻⁵³.

MAs are a mixture of structurally related species, synthesized from trehalose monomycolate (TMM) molecules, abundant glycolipids formed by the conjugation to the disaccharide trehalose via ester bonds. TMMs are synthesized in the cytoplasm, transported into the periplasm where they are converted into other molecular species, such as trehalose dimycolates (TDMs) and mycolic arabinogalactan polysaccharides (mAGP), by the Antigen 85 complex in a generally well-established pathway^{54,55}.

These species then constitute the OM of the pathogen. Of particular interest is the fact that the trehalose motif of TMM or TDM is recognised by the pattern recognition receptor (PRR) monocyte-inducible C-type lectin (Mincle), present in macrophages and dendritic cells ^{28,56,57}. This interaction triggers SH2-domain-containing phosphatase 1 (SHP-1) and Fc gamma receptor IIB (FcγRIIB), leading to inhibition of phagosome arrest ^{28,58}. Mincle also indirectly increases the production of tumour necrosis factor alpha (TNF-α), interleukin 6 (IL-6), and monocyte chemoattractant protein-1 (MCP-1) with subsequent recruitment of monocytes and neutrophils ^{59,60}. The pathogen-mediated immune response activated by the host when interacting with several MAs highlights the fundamental role that the mycobacterial cell envelope plays in establishing *Mtb* primary infection.

The MA biosynthetic pathway has been evolutionarily conserved and it is a highly regulated process, as the synthesis of such long and unique fatty acids requires a huge metabolic investment from the pathogen. There are many key components in this pathway, such as enzymes involved in fatty acid synthesis and activation, MA modifying enzymes, transporters and transferases ⁶¹. In recent years many new compounds have been discovered to target specific steps of this pathway. As a result of this, MAs biosynthesis has been validated as a druggable pathway, leading to the development of first- and second-line antibiotics, such as isoniazid and ethionamide, which are still successfully used in the clinic today for treatment of drug-sensitive TB ⁵¹.

Interestingly, both of these antibiotics are prodrugs that require metabolic activation by specific activating enzymes. Thus, it has been shown that mutations in the activating enzymes are the major reason for the development of clinical resistance ⁶¹. Isoniazid, which inhibits mycolic acid synthesis by targeting the enoyl-ACP reductase InhA, remains as the most important first-line anti-TB drug since the early 50s ⁶². However, resistance occurs frequently because the administered prodrug needs the enzyme catalase-peroxidase (KatG) to be transformed into its reactive free radical form. Loss of function mutations in the gene encoding for the activating enzyme often arise as KatG is a non-essential enzyme, thus leading to antibiotic-resistant strains ^{63–65}.

More recently, attention has focused on targeting the transport process whereby the MAs precursors are transferred from the inner membrane (IM), where they are synthesised, to their final cellular destination in the OM ⁵¹.

1.2. MmpL3 as potential target for the development of novel anti-tubercular agents

1.2.1 The Resistance-Nodulation-Cell Division proteins superfamily

MmpL3 belongs to the resistance, nodulation and cell division (RND) transporter superfamily, whose members function to pump exogenous agents to the outside of the cell, via a H⁺ dependent co-transport mechanism (antiport), driven by a proton motive force ^{66–68}. However, unlike other family members that exert an antibiotic-resistance function by pumping exogenous agents to the outside, MmpL3 in *Mtb* as well as its orthologues *Mycobacterium smegmatis* (*Msmeg*) MmpL3 and CmpL1 protein in *Corynebacterium glutamicum*, exert their efflux activity on endogenous compounds, such as the TMMs ^{67,69}. Intriguingly, all of the structures of MmpL3 obtained to date indicate that MmpL3 is structurally and functionally a monomer ^{1–4}. This is in contrast to most other RND transporters, the best characterised example of which is the *Escherichia coli* multidrug acriflavine resistance B (AcrB) protein pump ^{5,70}. This protein has been shown to oligomerise as a homotrimer both alone and in complex with the other components to form the tripartite efflux complex, AcrAB-TolC ⁵. In fact, AcrB oligomerisation is likely needed for association with AcrA and TolC, thus ensuring the formation of the complete complex. The asymmetric nature of the homotrimer is key for protein function, as each drug extrusion cycle is coordinated via the alternating conformations adopted by the individual protomers, in either access, binding and extrusion state, through a functionally rotating mechanism ^{71,72}. A trimer has also been reported as the most common oligomeric arrangement for the related aminoglycoside exporter, AcrD ^{73,74}. In their work, Zhang and colleagues showed that AcrD likely arranges into multiple quaternary structures, including dimeric and monomeric structures and these were isolated by cryo-EM ⁷⁴. However, it is worth noticing that the homotrimer was the most favourable and stable form ⁷⁴.

In contrast, the *Burkholderia multivorans* hopanoid biosynthesis-associated RND transporter HpnN is dimeric ⁷⁵. HpnN is responsible for movement of pentacyclic triterpenoid lipids that are structural and functional homologues of cholesterol, from the plasma membrane to the outer membrane of some gram-negative bacteria ⁷⁶. However, the human Niemann-Pick C1 (NPC1) protein which also plays a crucial role in cholesterol homeostasis appears to be monomeric ⁷⁷.

1.2.2 MmpL3 as a druggable target

The essential role played by MmpL3 has been confirmed by many researchers over the past decade ^{51,78–80}. In particular, the observation that *mmpL3* is the only *mmpL* gene that is impossible to knock out by allelic replacement and that it is conserved across the *Mycobacterium* genus suggests it provides a crucial contribution to virulence ^{51,67,78–80}. Indeed, knockout of the *mmpL3* gene which encodes MmpL3 leads to non-viable mutant strains ^{51,78–80}. Moreover, its conditional depletion and consequent loss of function also affects lipid metabolism and transport, with observed accumulation of TMM molecules on the inside of the cell ^{51,66,81}. However, strain viability is restored by introduction of an exogenous *mmpL3* gene from either the same strain or another mycobacterium strain. The analogy of function and the interchangeability across the *mmpL3* from *Mtb* and the one from the *Mycobacterium smegmatis* (*Msmeg*) strain in particular, validated the two as orthologs and justified the common use of the *Msmeg* strain for *in vitro* studies and protein characterisation ⁵¹. Nonetheless, the medical relevance of this target is also linked to the inhibitory effect shown by multiple classes of small molecules, some of which are in various stages of clinical trials ^{82–84}.

1.2.3 MmpL3 structure and transport mechanism

Almost all the studies aforementioned have been carried out *in vitro* and no high-resolution structure was available until recently.

In 2019, two independent research groups published two crystal structures of engineered versions of *Msmeg* MmpL3 (PDB 6AJF and 6OR2), isolated in detergent ^{1,2}. Both structures were generated by truncating the C-terminus and for the 6AJF structure by Zhang *et al* the sequence was further engineered by fusing a T4 lysozyme at the C-terminal domain to facilitate crystallisation ¹. The atomic models were resolved by X-ray diffraction and were found to be in agreement with a previously determined partial structure ^{1,2,85}. These structures revealed that MmpL3 presents 12-transmembrane (TM) alpha-helical segments, organised into two domains, TM 1a - 6 form the N-terminal domain (TMN) and TM 7a - 12 form the C-terminal domain (TMC) (Figure 1.4 and Figure 1.5, a and b). The two domains share an approximate 2-fold symmetry, despite exhibiting only about 40% sequence similarity (Figure 1.5, b). Two subdomains extend into the periplasmic space between TM 1b and TM 2 (PND) and

TM 7b and TM 8 (PCD), forming a central cavity thought to be crucial for TMM translocation (Figure 1.5, a) ^{1,2}. Furthermore, a disordered loop (aa 353 - 381) links the TMN to the TMC, while another alpha-helical structure (α 8) connects the latter to an unstructured C-terminal domain (CTD), unresolved (Figure 1.4).

The *Mtb* MmpL3 (PDB 7NVH) structure was also solved in detergent micelles, by cryo-EM and using 6OR2 as reference for refinement ⁴. Also in this case, the protein sequence was engineered by truncating the C-terminal domain. In addition, 35 residues (aa 343 - 377) connecting TM 6 to TM 7 were not solved ⁴.

The residue pairs considered to be responsible for proton translocation in the *Msmeg* MmpL3 protein are Asp256 and Tyr257 (Asp251 and Tyr252 in *Mtb*), located on TM 4, that respectively interact with Tyr646 and Asp645 (Tyr641 and Asp640 in *Mtb*), located on TM 10 (Figure 1.5, c). These residue pairs are suggested to delimit a channel that has also been identified as the possible binding site for many MmpL3 inhibitors ¹. High-throughput whole-cell phenotypic screening against compound libraries have led to the identification of many antitubercular agents, of different chemical classes and they all target MmpL3 proton relay network ⁸⁶. In particular, SQ109 is a promising anti-TB drug that entered a Phase 2b-3 clinical study ⁸⁷. This compound was identified from a combinatorial library screening based on the 1,2-ethylenediamine pharmacophore found in ethambutol, although the two structures are completely different, as well as their molecular targets ⁸⁸. Ethambutol, one of the first-line TB antibiotics, inhibits three arabinosyltransferases, ultimately preventing the biosynthesis of arabinogalactan and lipoarabinomannan, while SQ109 targets MmpL3, thus inhibiting MA transport to the cell wall ^{61,88-90}. Zhang and co-workers produced evidence in support of SQ109 specific binding to *Msmeg* MmpL3 in a microscale thermophoresis assay, measuring a dissociation constant (K_d) of 1.65 μ M. They also showed that *Msmeg* growth is susceptible to SQ109 exposure in a dose-dependent manner using a cell-based inhibition assay ¹. Upon inhibitor binding, the six C-terminal TM helices move away from the protein central axis and undergo a noticeable conformational change, creating a pocket to accommodate SQ109. Furthermore, SQ109 presented synergistic interactions with rifampicin and isoniazid, as well as showing promising results against MDR-TB and XDR-TB clinical strains ¹. Another compound, the adamantyl urea AU1235, was also found to bind in the same pocket as

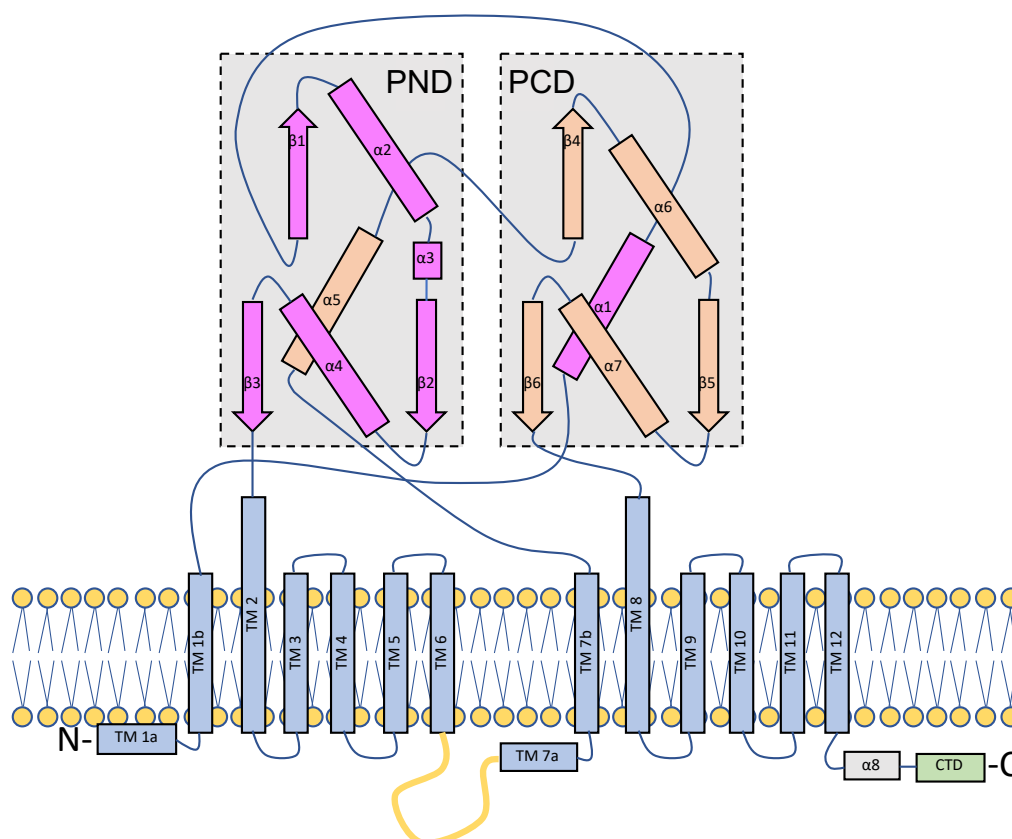


Figure 1.4. *Msmeg* MmpL3 topology. The full-length protein is made of 1013 residues, organised in 12 TM domains (in blue), 2 periplasmic domains (in grey boxes) and 1 C-terminal domain (CTD, in green) located in the cytoplasm. The protein can be divided into two parts, almost symmetrical. The N-terminal domain is divided into 6 TM domains, TM 1 - TM 6, with TM 1 subdivided into 1a and 1b and TM 2 being longer than the other helices. The PND domain starts from TM 1b, forming 4 α -helices and 3 β -strands, and ends in TM 2. The C-terminal domain presents the same number of TM domains, TM 7 - TM 12, with the only exception that TM 12 is linked to an α -helical structure α 8 (in grey), then connected to the CTD. The PCD domain starts from TM 7a and ends in TM 8, presenting 3 α -helices and 3 β -strands. The two periplasmic domains are intertwined as the α 1 helix from the N-terminal is part of the PCD, while the α 5 helix belonging to the C-terminal is part of the PND domain. Moreover, the loop connecting TM 6 to TM 7a is shown in yellow, with residues not present in the solved structures.

SQ109 but with a K_d 5.7-fold lower than that for SQ109⁵¹. Indolcarboxamides (ICAs) have been reported as a new class of anti-tuberculosis agents, with ICA38 recording the highest *in vitro* anti-*Mtb* activity in this group, binding to MmpL3 with a K_d of 0.16 μ M in the same pocket as the other two compounds described^{1,91}.

Interestingly, all these compounds as well as many others, regardless of their chemical scaffolds, have been suggested to specifically inhibit MmpL3 since the resistant mutations have been specifically mapped to the *mmpL3* gene¹. They work by disrupting the Asp-Tyr pairs implicated in the H^+ co-transport, thus blocking substrate translocation and ultimately inhibiting mycobacterial cell wall assembly. Therefore, it makes sense that ~90% of the 83 resistance positions identified reside within the transmembrane domain, with 60% of these within 10 Å from the drug binding

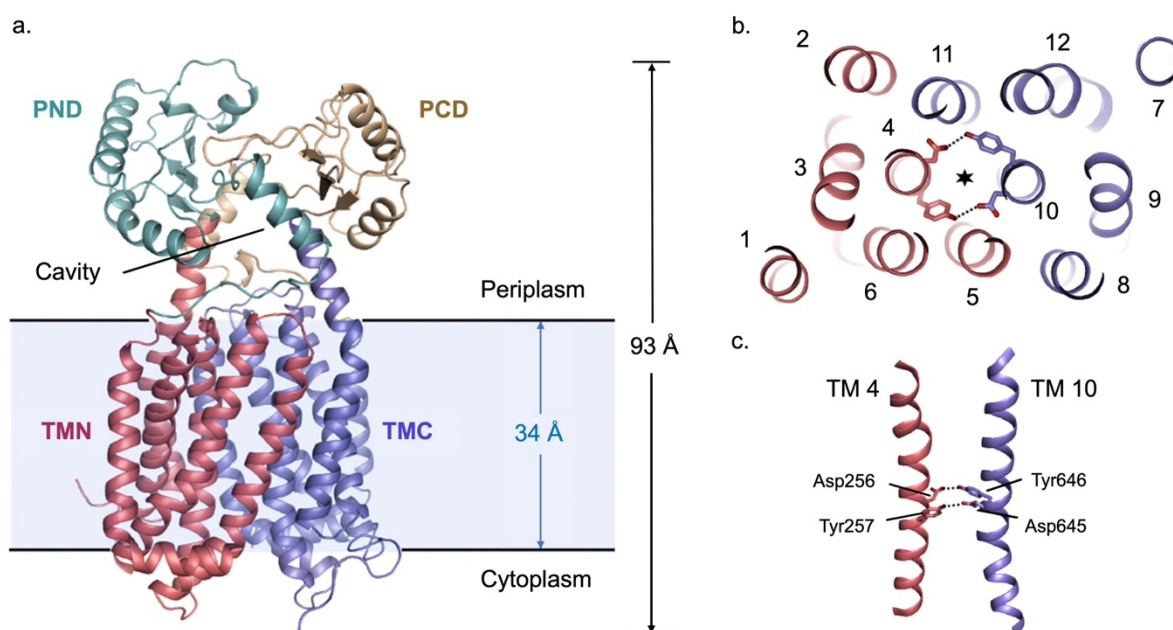


Figure 1.5. *Msmeg* MmpL3 structure. a. Cartoon representation of the MmpL3 crystal structure, with main domains in different colours: TMN in red, TMC in purple, PND in green and PCD in yellow, with these two sub-domains delimiting the periplasmic cavity. b. A clipped view of the transmembrane region viewed from the periplasm, with TM helices numbered. The star indicates the pseudo-two-fold axis. Asp and Tyr residues side-chains are shown as stick models and are involved in the generation of MmpL3 proton motive force. Hydrogen bonds are displayed as dashed lines. c. The two Asp-Tyr pairs are located in the middle of TM helices 4 and 10, lying at the core of the two TM bundles. Adapted with permission ⁶⁵. Copyright 2019, Elsevier.

site. However, it is also significant that ~40% of these mutations are located >10 Å away from the drug pocket, suggesting that the periplasmic cavity might also be involved in MmpL3 resistance mechanism ⁴. Nevertheless, when a large number of *Mtb* genomes mostly originating from clinical samples and obtained from the European Nucleotide Archive (ENA) were analysed, the prevalence of pre-existing resistance mutations in MmpL3 was estimated to be 0.1%. This low percentage provided an estimate for the level of environmental resistance that MmpL3 inhibitors might encounter once licensed ⁴. This was also related with the high conservation of *mmpL3* gene, supported by the very low ($>10^{-11}$) *in vitro* bacterial mutation rate observed for SQ109 ⁸⁷.

Crystal and cryo-EM structures of *Msmeg* MmpL3 have been obtained in complex with detergent molecules 6-n-dodecyl- α,α -trehalose (6-DDTre) and trehalose 6-decanoate (T6D), both homologues of TMM, as well as with n-dodecyl- β -D-maltopyranoside (DDM), the native substrate TMM, and a molecule of phosphatidylethanolamine (PE) ¹⁻³. A molecule of the detergent lauryl maltose neopentyl glycol (LMNG) was identified in complex with the *Mtb* MmpL3 structure ⁴. Taken together these structures reveal a large and flexible substrate binding site,

predicted to be accessible from the outer leaflet of the membrane, and allowing substrate transport via the central vestibule between the two PND and PCD lobes ^{2,3}. The periplasmic cavity has been suggested as the conduit through which TMM molecules are exported into the periplasm, a role that also been supported by site-directed mutagenesis experiments ^{2,3,69}. A combination of high-resolution structures and molecular dynamics simulations of the apo and inhibitor-bound states have suggested Ser423 on TM 7 and Asn524 on α 7 acting as an entrance gate for substrates ³. In support of this hypothesis, many structures present detergent or substrate molecules bound at this side pocket, with the ligand head group protruding into the periplasmic cavity while being uplifted from the membrane plane. Once the substrate has bound to this site, transport is then finely modulated through a coordinated combination of movements involving the TMDs, mediating proton translocation, and the opening of the periplasmic cavity, allowing substrate export ³.

Together, the solved structures marked an essential step forward in MmpL3 characterisation, but they also carry some limitations. As with almost all crystal and cryo-EM structures the protein is fixed in a single conformation, providing only static snapshots. In addition, the conditions used for crystallisation or cryo-EM experiments are not representative of the protein behaviour in its physiological environment. Moreover, in the case of *Msmeg* MmpL3 structures 6OR2 and 6AJF, as well as for *Mtb* MmpL3 7NVH, the C-terminal domain was removed to improve protein stability and increase the diffraction quality of the crystals for the *Msmeg* structures ^{1,2,4}. In addition, the cytoplasmic loop connecting the TMN to the TMC (aa 353 – 381 for *Msmeg* and aa 343 – 377 for *Mtb*) could not be resolved, probably due to its high flexibility ^{1,2,4}. Finally, we have seen that several molecules of detergents and/or phospholipids were co-crystallised with the structures, suggesting that the protein establishes important interactions with its surrounding environment which need to be further investigated by reproducing its native state ¹⁻⁴.

1.3. Membrane proteins and their environment

Membrane proteins (MPs) are notoriously difficult to work with, as extracting them from the hydrophobic environment of the biological membrane, while retaining a physiologically relevant and stable state, suitable for downstream analysis, remains a challenge. There is a clear lack of knowledge and understanding of these important

molecules and this is reflected in the fact that less than 2% of the 89,000 total unique protein sequences in the PDB database are membrane proteins (www.rcsb.org and <https://blanco.biomol.uci.edu/mpstruc/>). Although less abundant in nature than soluble proteins, MPs account for over 60% of all pharmacological targets (<http://www.proteinatlas.org>)⁹².

The biological membrane is a selectively permeable barrier, made of a complex mixture of phospholipids arranged in a bilayer and interspersed proteins (Figure 1.6). The membrane physically separates the internal and external environment of cells and organelles as well as exerting a wide range of essential cellular functions⁹³. The lipid components have traditionally been considered as the static scaffold of the membrane, while the vast majority of biological functions were thought to be carried out by the proteins. However, membrane lipids have been recently shown to be key players in MPs structure, function, trafficking and regulation^{94–96}. Therefore, rather than studying MPs and lipids individually, there is an urgent need to shed light on the complex MP-lipid interactions taking place at membrane level, in order to obtain a true picture of how MPs function.

Structural and functional analyses of MPs have historically relied on the extraction of membrane proteins from the lipid environment using a small number of detergents with relatively simple architectures⁹⁷. Their amphiphilic nature allowed both disruption of the lipid bilayer and stabilisation of the protein of interest in aqueous solution but the number and type of membrane proteins that could be worked with was limited. Although novel detergents with improved properties have been recently developed, providing a wide range of molecules suitable for individual proteins and specific applications, these agents harbour some intrinsic limitations, as they along with all detergents potentially destabilise the protein and remove important lipid molecules during the solubilisation process^{98–100}.

Much effort has gone into the development of optimised approaches to maintain MPs in a physiologically relevant state as this is crucial both to provide relevant biological insights as well as to allow effective and rational drug design (Figure 1.7)¹⁰⁰.

Several membrane mimetic systems have been developed that encapsulate MPs in lipid based nanodiscs, either directly from the native membrane using polymers or via reconstitution of detergent-purified protein into nanodiscs of defined lipid composition (Figure 1.7, d and f)^{100–102}.

In the next sections, the use of detergents, synthetic and native nanodiscs for

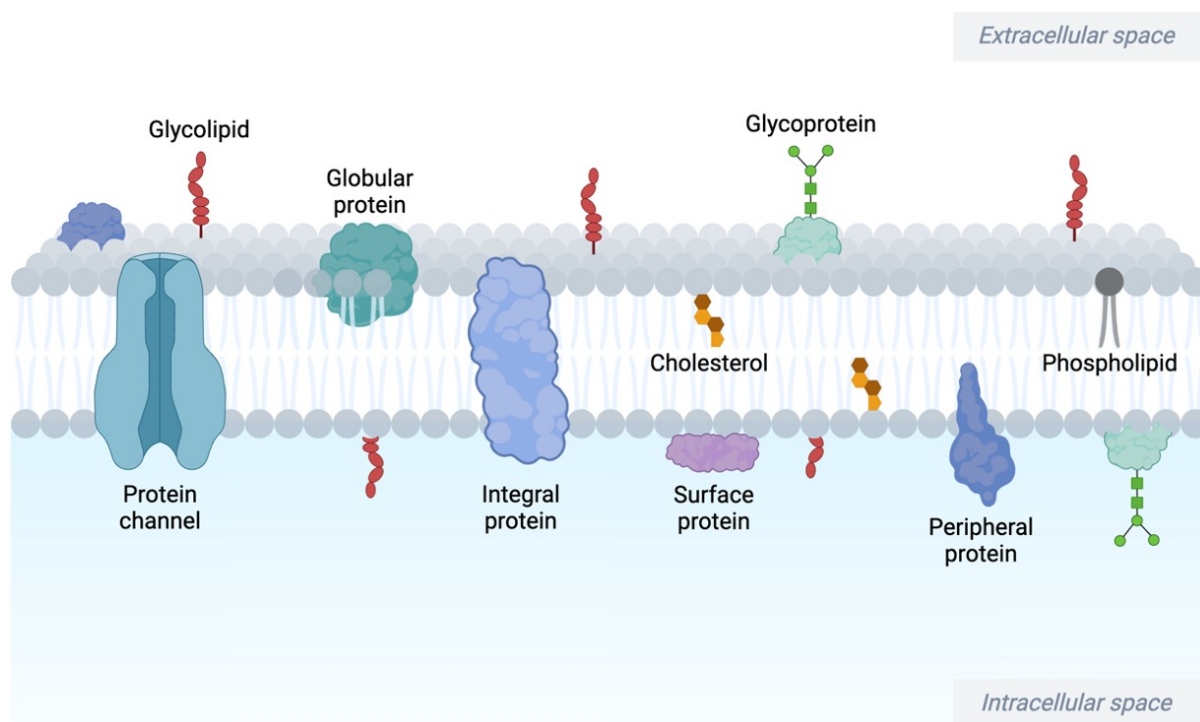


Figure 1.6. Fluid mosaic model of the cell membrane. Different classes of lipid species and membrane proteins form the complex composition of the cell membrane. The figure was created with BioRender.com and adapted from the template “Phospholipid Bilayer Membrane”.

MP manipulation as effective alternatives to the physiological lipid environment will be discussed in more detail. Sections 1.3.2, 1.3.3 and 1.3.4 feature much of that reviewed by Woubshete *et al* ¹⁰³.

1.3.1 Lipids and the biological membrane

In a membrane, there are three major classes of membrane lipid molecules – phospholipids, sterols, and glycolipids ⁹³. Of these, the most abundant lipids are known as glycerophospholipids (Figure 1.8). These are amphiphilic molecules, made of a polar glycerol-phosphate head group and hydrophobic fatty acid tails. Phospholipids can also be classified by the overall charge of the head group into zwitterionic phospholipids (e.g. phosphatidylethanolamine or PE and phosphatidylcholine or PC) and anionic phospholipids (e.g. phosphatidylinositol or PI and phosphatidylglycerol or PG). Each class contains a range of species, where the fatty acid chains vary in length and saturation ⁹³. In biological systems, phospholipid molecules self-assemble into bilayers with the head groups facing the water solution while the tails cluster together forming the inner core of the membrane, so as to minimise their contact with the surrounding aqueous environment. The driving force of this self-assembly is known as

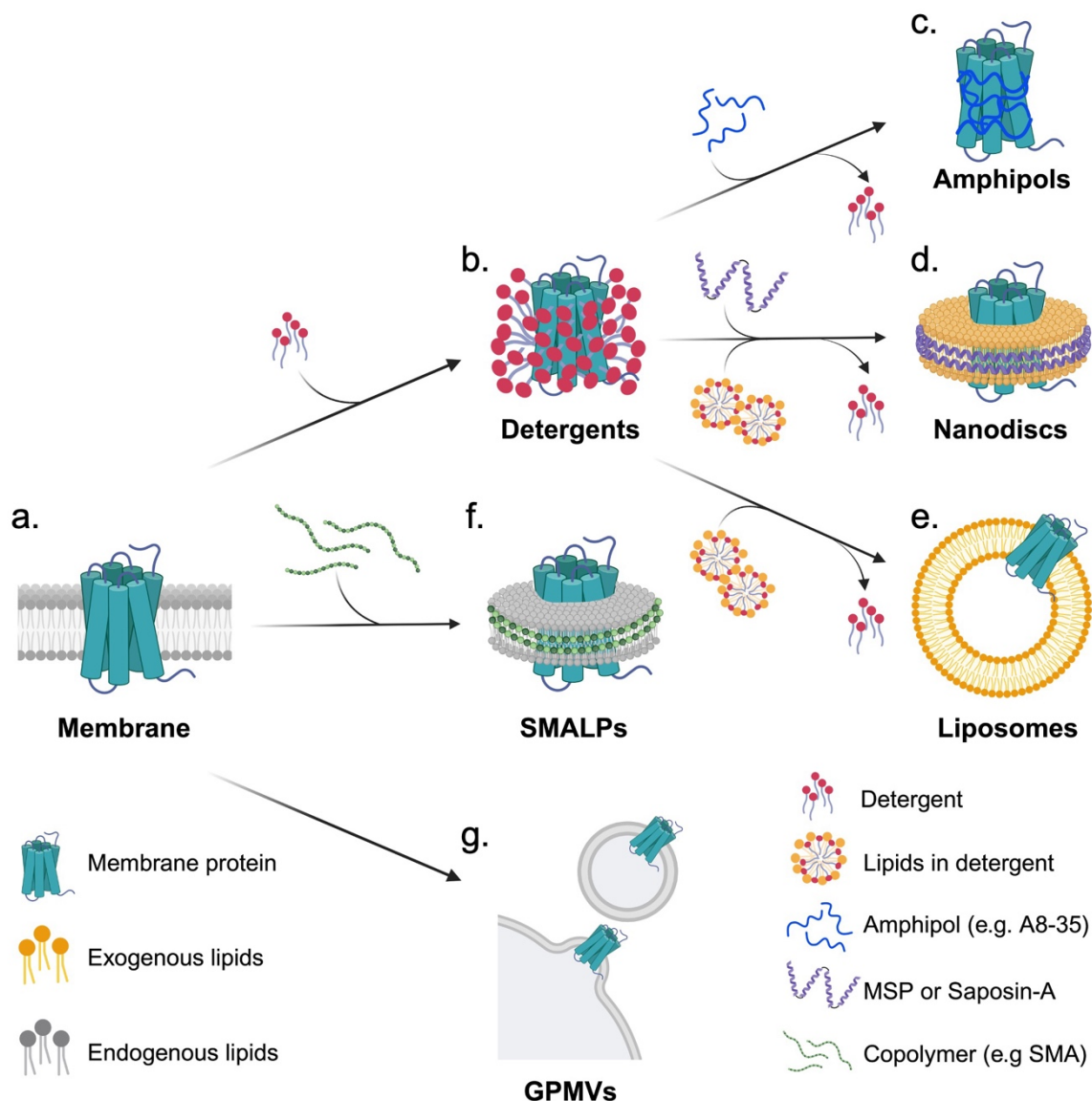


Figure 1.7. Summary of approaches to preparing membrane protein sample suitable for downstream analysis. a. The integral membrane proteins are located within the native lipid bilayer. b. The most common method of extraction utilizes detergents and the resultant protein-detergent complexes can then be used for isolation of the target protein followed by direct analysis in detergent-based solution or the protein can be exchanged into c, amphipols, or reconstituted into d, nanodiscs or e, liposomes. f. Alternatively, the protein can be extracted from the membrane into native nanodiscs containing the protein of interest and associated membrane lipids using polymer-based systems, or g, vesicles can be generated from membranes and used directly for analysis. SMALPs: styrene maleic acid lipid particles. GPMVs: giant plasma membrane vesicles. MSP: membrane scaffold protein. Reprinted with permission ¹⁰³. Copyright 2024, John Wiley and Sons.

the hydrophobic effect ^{104,105}. Lipids aggregate into various phases depending on different factors, including the lipid types and composition, the water content, and the temperature and pressure. Membrane phospholipids usually have an overall cylindrical shape (Type 0) when the cross-sectional areas of the head group and the acyl chains are similar. Such molecules are called bilayer prone lipids, and they aggregate to form

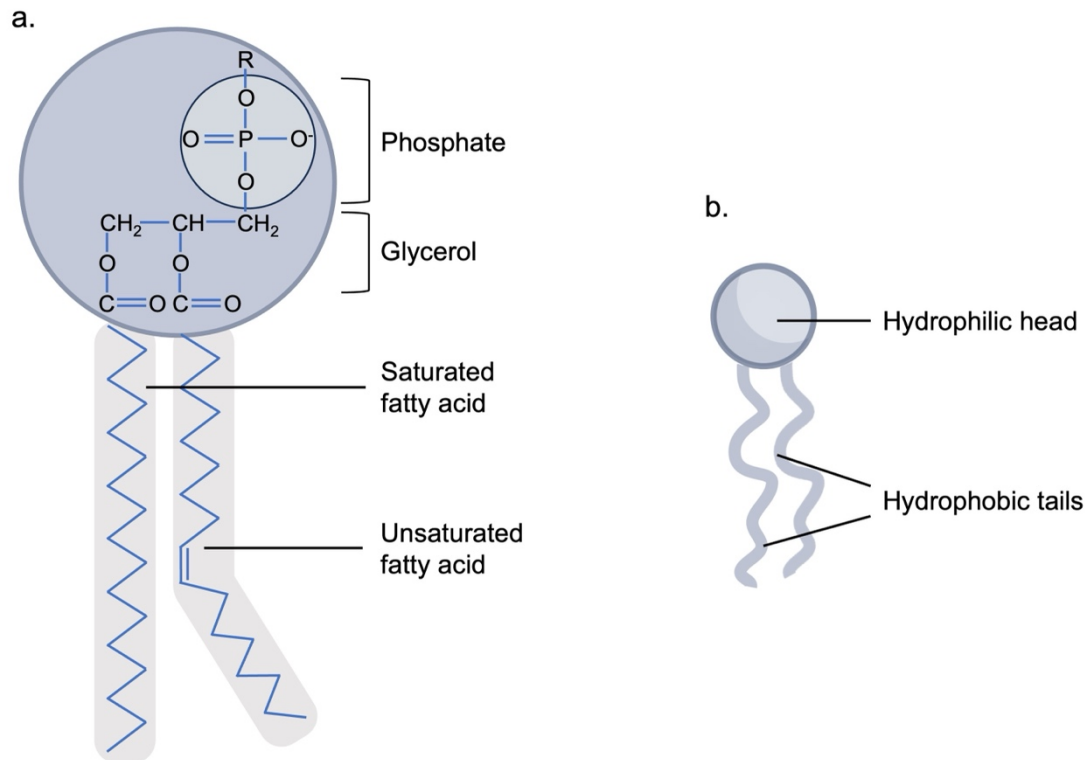


Figure 1.8. Structure of a phospholipid. Each molecule is composed of two fatty acids forming the hydrophobic tails and of a modified phosphate group attached to a glycerol backbone forming the hydrophilic head group. Representation of a phospholipid from its structural formula (a) to its usual schematic representation (b).

a fluid lamellar phase (zero curvature) in biological membranes (Figure 1.9). If the cross-sectional areas of head group and acyl chains are different from each other, there are two possible scenarios. In both cases these are categorised as non-bilayer prone lipids. If the head group is larger than its acyl chains, the overall shape of the lipid is conical (Type I) and generates an intrinsic positive curvature that leads to the formation of micelles in water ^{106,107}. An inverse conical (Type II) molecular shape instead is seen in lipids that have smaller head groups compared to their acyl chains. These lipids generate negative curvature and form a reverse hexagonal phase or inverted micelle (Figure 1.9) ¹⁰⁷. Thus, in a biological membrane adjusting the bilayer/non-bilayer prone ratio is extremely important since the overall contribution of their molecular forces affects the average curvature and its frustration, essential factors for proper functioning of the bilayer ¹⁰⁸.

In addition to the structural variation between different lipid shape, lipid bilayers also show diversity in composition. Various proportions of different lipids are found in the membranes of different biological species, different cell types and different organelle within a single cell ¹⁰⁹. Furthermore, biological membranes are asymmetric

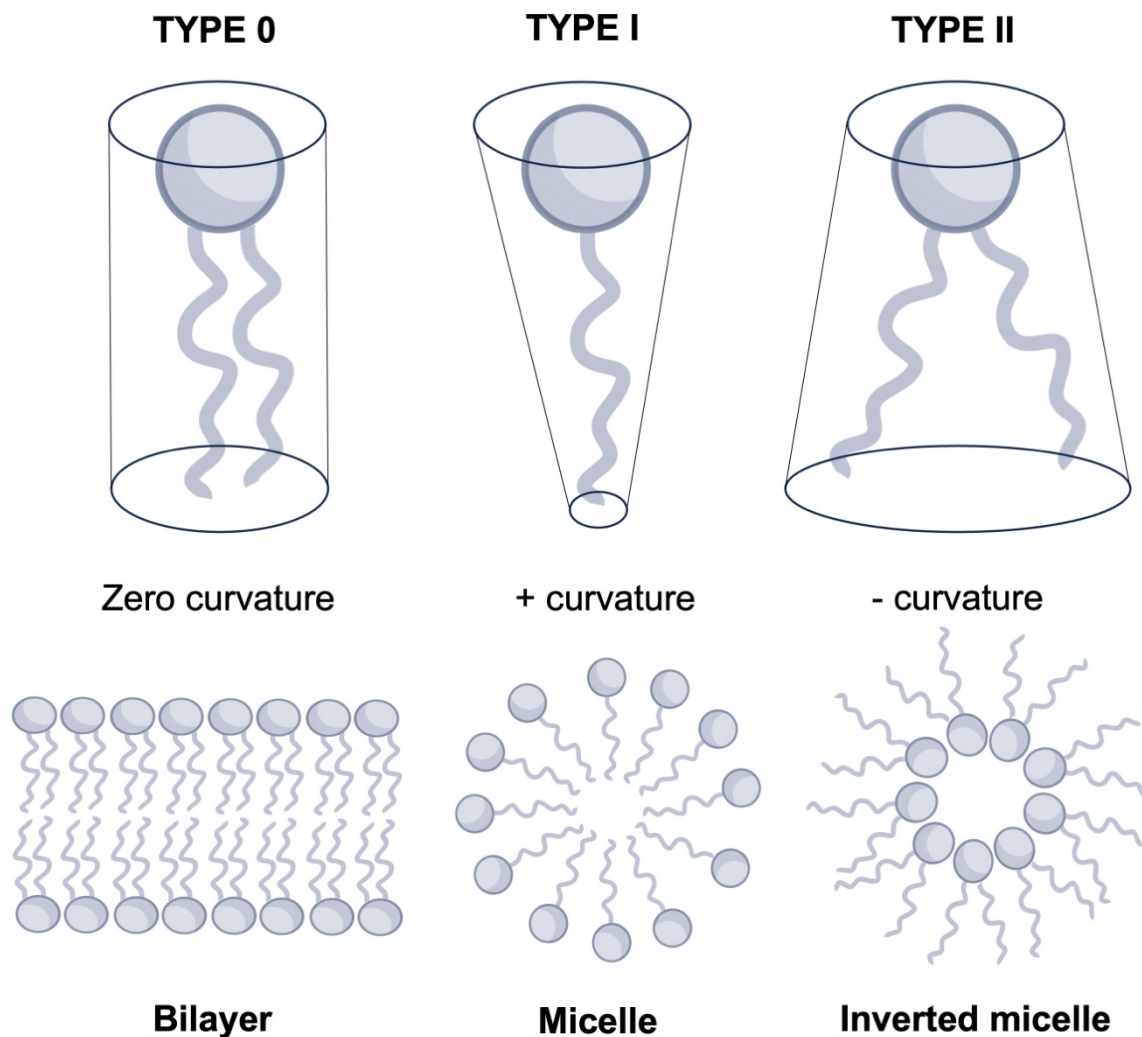


Figure 1.9. Types, curvature and self-assembly of lipids. Type 0 lipids present zero curvature, are cylindrical in shape and form bilayers when aligned. Lipids with positive (Type I) and negative (Type II) curvature are grouped as non-bilayer prone lipids. Non-bilayer prone lipids with positive curvatures tend to form micelles, while those with negative curvature form inverted micelles.

with different lipid compositions in the inner and outer leaflets ¹¹⁰. It has been shown that the translocation of lipids via a flip-flop mechanism as well as the maintenance of bilayer asymmetry play a fundamental role in a number of key processes such as cell apoptosis, signalling, and modulation of protein channel opening ^{111–113}.

The fluidity of biological membranes is fundamental to their function as it allows proteins and lipids to explore the cell surface and find interaction partners. The intrinsic capacity for cells to autonomously maintain their membrane fluidity has been recognised since the 1970s ^{109,114}. Bacteria can regulate their lipidome by initiating synthesis of new lipids but also modifying the existing species in order to maintain the required membrane properties ¹¹⁵.

As previously mentioned, functional interactions between membrane proteins

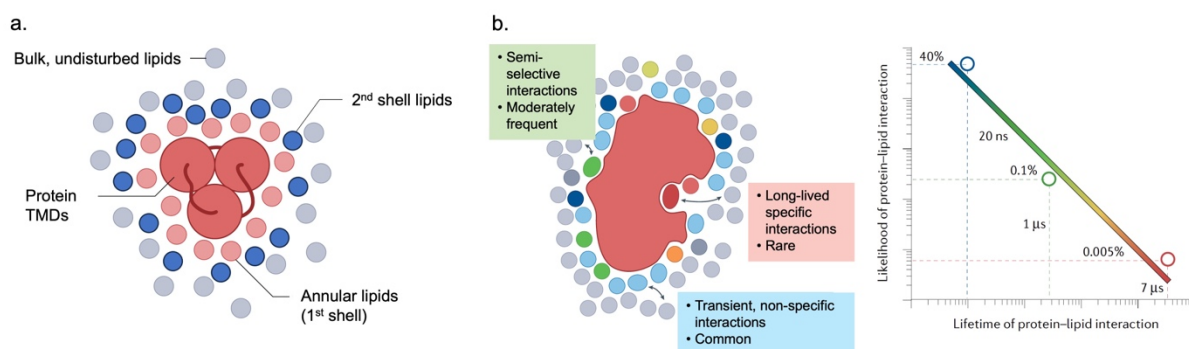


Figure 1.10. Lipid interactions with membrane proteins. a. The 1st shell of annular lipids stably associates with transmembrane protein domains (TMDs) and presents a slower exchange rate than lipids forming the bulk, which are undisturbed by the presence of the protein. b. Different lifetime and frequency of interactions are established between integral membrane proteins (IMPs) and individual lipid molecules (coloured), in competition with those from bulk lipids (grey). Protein–lipid interactions span a range of lifetimes (represented by the different colours) as suggested by simulations. There are high affinity interactions, which are rare, as well as very common short-lived interactions (in the nanosecond range), which indicate rapid lipid exchange that is not influenced by protein binding. These observations imply that specific, semi-selective and more solvent-like interactions could potentially be simultaneously relevant for an IMP. Figure adapted with permission ¹⁰⁹. Copyright 2023, Springer Nature.

and other biomolecules are fundamental to molecular cell biology. Integral membrane proteins (IMP) are solvated by complex mixtures of lipids that influence their structures, dynamics and functions. Thus, according to their proximity to the membrane proteins lipids can be classified as 1st and 2nd shell lipids or bulk lipids (Figure 1.10) ¹⁰⁸. Lipids included in the 1st shell form interactions with IMPs and are thus motion-restricted compared to bulk lipids. In fact, their exchange rate with bulk lipids (10^7 s^{-1} at 37 °C) has been measured as one or two factors slower than the time for lipid exchange in the absence of protein ^{116,117}. They are also termed annular lipids as they form a ring around the protein. If the protein-lipid interaction is strong, leading to slow off-rates, the protein residue involved will be considered as a hotspot for lipid binding, with the lipid potentially having regulative effects on the IMP function. The most common examples of function-regulating annular lipids are found to exist in the structures of potassium channels (Human Kir2.2 and bacterial KcsA) and the large-conductance mechanosensitive channel (MscL) ^{118–121}. In the crystal structure of the homotetrameric potassium channel KcsA a lipid binding site with specificity for anionic lipids was resolved at each monomer-monomer interface, with occupation of this site being necessary for channel opening ¹¹⁹. The 1st shell of lipids also includes interfacial lipids, that tend to bind proteins in their indented surfaces consisting of pockets, cavities and clefts ⁹⁴. These lipids are often structurally resolved and frequently reside at the contact sites between the adjacent monomeric units within oligomeric assemblies. One

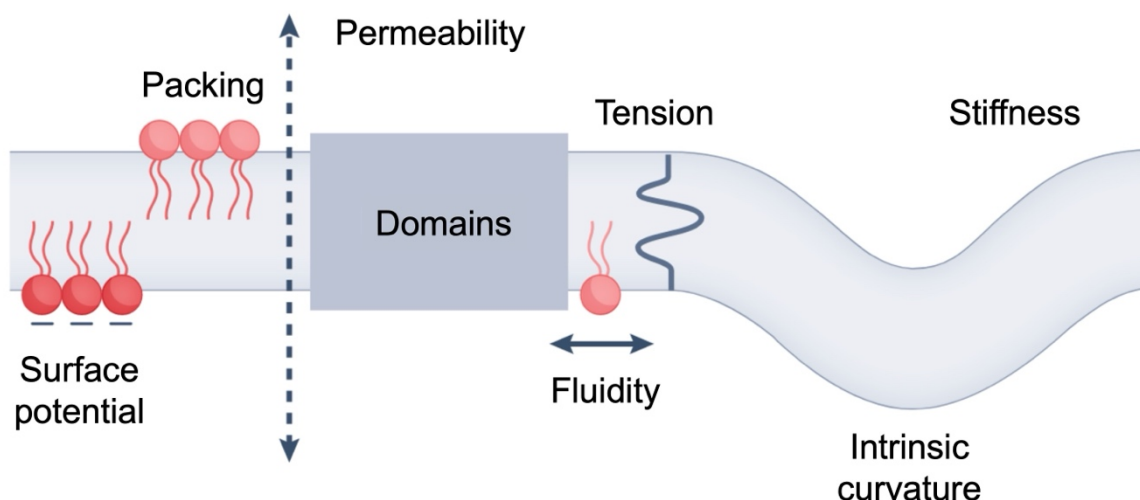


Figure 1.11. Collective membrane physical properties. Individual lipid molecules can serve as specific cofactors for integral membrane proteins (IMPs), but also collectively contribute to the biophysical properties of the lipid bilayer. Thus, not only membrane lipids can functionally regulate IMPs, but also IMPs can transduce signals regulating the lipidome. Adapted with permission ¹⁰⁹. Copyright 2023, Springer Nature.

example of this type of interaction was observed in the 2.9 Å X-ray structure of bacteriorhodopsin, where 3 molecules of haloarchaeal glycolipid S-TGA-1 (3-HSO₃-Galpβ1-6Manpα1-2Glc pα-1-archaeol) were solved at the interface between adjacent protomers, acting as a molecular “glue” in bacteriorhodopsin trimerisation ^{122,123}. Interfacial lipids were also observed in the structures of ion channels, proton pumps, receptors and transporters ^{94,124–127}. Finally, another class of lipids found in the 1st shell are the non-annular. They are usually buried deep within small pockets/clefts formed by proteins and adopt unusual positions (e.g. head group below the membrane plane and/or non perpendicular to the bilayer) and have much lower off-rates compared to annular lipids ^{119,122,128}. Some non-annular lipids have major roles in the regulation of membrane protein function by contributing to their folding and self-assembly in the membrane, with phosphatidylinositol (PI) the most recognised example ^{122,129}. PI was observed in the structure of the cytochrome *bc*₁ complex residing in an interhelical, thus unusual, position with its inositol head group stabilised by the formation of several hydrogen bonds with the side chains of catalytic subunits. When the lipid is bound the torsion forces generated by the fast movement of the extrinsic domain of the Rieske protein dissipate, ensuring the movement of the catalytic domain that is essential for the activity of the enzyme. Hence, PI acts as a regulator molecule for the cytochrome *bc*₁ complex and could also be important for the assembly of the complex as it is located at a position where four TM subunits of the monomer come in contact with each other ^{122,130}.

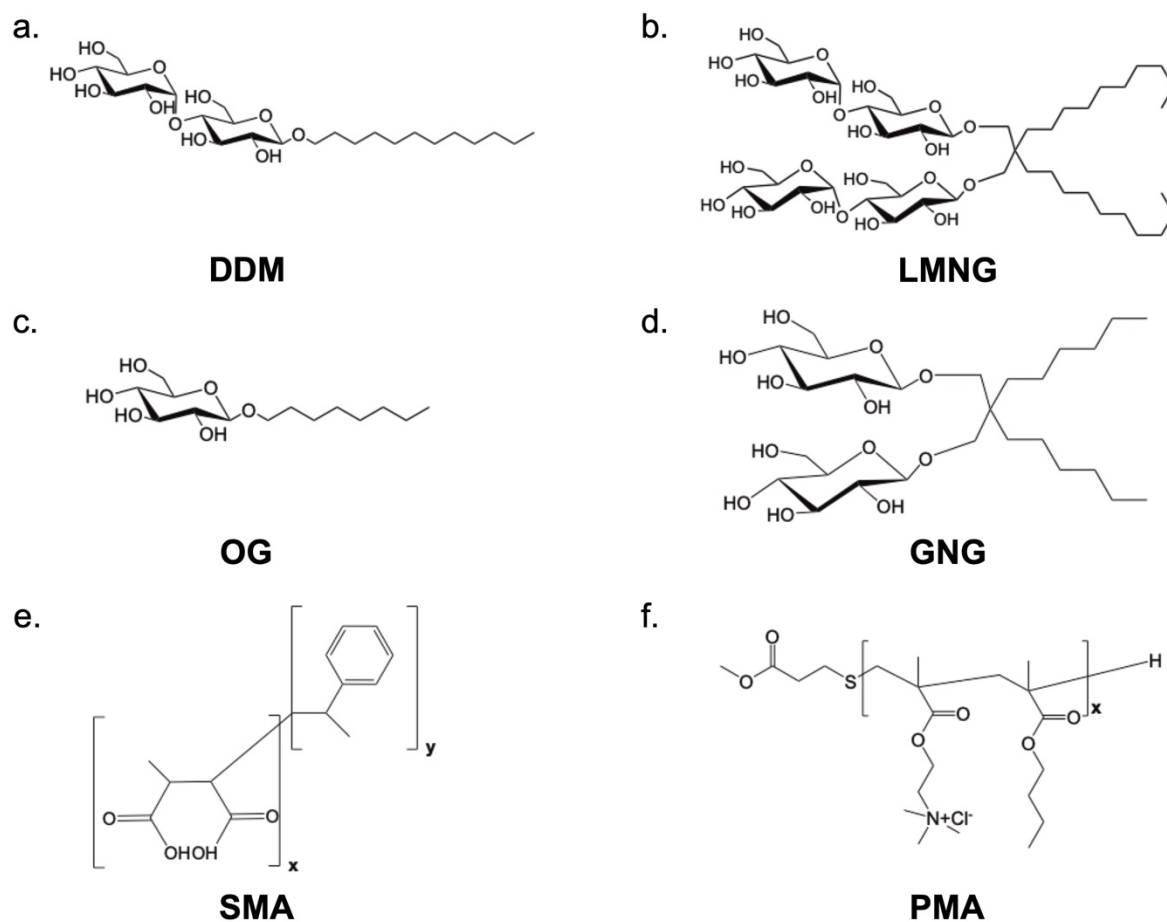


Figure 1.12. Example structures of detergents, polymers and amphipols used to stabilise membrane proteins. a, Dodecyl- β -maltoside (DDM) and b, lauryl maltose pentyl glycol (LMNG), are detergents that contain 1 or 2 maltose head groups and 1 or 2 C12 alkyl chains, respectively. c, Octylglucoside (OG) and d, glucose neopentyl glycol (GNG), are examples of glucoside detergents that contain glucose head groups and C8 alkyl chains. e. Styrene maleic acid (SMA) is a polymer that contains different ratios of styrene to maleic acid groups. f. Polymethacrylate (PMA) is a polymer that consists of a hydrophobic butyl methacrylate and a cationic methacrylcholine chloride group. Reproduced with permission ¹⁰³. Copyright 2024, John Wiley and Sons.

On the other hand, properties of bulk lipids are not significantly affected by the presence of a membrane protein, as these do not directly interact with IMP ¹²⁸. However, bulk lipids can affect membrane protein function as they contribute to the macroscopic collective properties of the membrane (Figure 1.11). These properties include viscosity (fluidity), lateral stress and spontaneous membrane curvature ¹⁰⁸.

The interplay between membrane proteins and the lipid bilayer is remarkable and entangles at many different levels. For instance, individual lipid molecules can act as specific protein cofactors or actively contribute to the regulation of collective membrane properties like its thickness, fluidity and level of lipid packing which in turn affect protein conformational states and function. In this context, several factors need

to be carefully considered when investigating the behaviour of a membrane protein in its native environment. Thus, the choice of the mimetic system is crucial to ensure the preservation of correct protein folding during extraction, to study MPs function, structure and dynamics in their physiological conditions.

1.3.2 Detergents

Detergents can be broadly grouped as anionic, ionic or zwitterionic, based on their chemical and charged nature. Standard detergents have a linear structure with a hydrophilic head group, often sugar based, and a long hydrophobic alkyl chain (Figure 1.12, a - d) ¹³¹. The dual hydrophobic-hydrophilic character of the detergent molecules mirrors that of the membrane. At a concentration specific to the individual detergent, the critical micelle concentration (CMC), individual detergent molecules self-assemble into a micelle structure with the hydrophobic tails buried in the core and the hydrophilic head groups exposed on the surface (Figure 1.7, b), effectively replacing the membrane by forming a mixture of detergent-protein-lipid, detergent-protein and detergent-lipid complexes ⁹⁷. In particular, individual detergent CMC is an important factor to consider, as this is directly correlated to micelle stability and size. As an example, standard detergents that have been widely used include the maltose-based DDM and n-decyl-beta-maltoside (DM), that share the same head group but present different alkyl chain lengths (12 carbon-chain for DDM and 10 carbon-chain for DM), and octylglucoside (OG), with a glucose head group and a C8 alkyl chain ⁹⁷. Detergents with shorter alkyl chains tend to have higher CMCs. In particular, the longer the hydrophobic chains the larger the micelle size, meaning that fewer molecules are needed to form a micelle. Micelle size can impact the stability of the target protein in the detergent-based solution and influence the efficiency of extraction from the membrane ⁹⁷. Furthermore, choice of detergent is crucial for downstream analysis. While DDM and OG have been used successfully for structural studies, there are only a very limited number of membrane proteins that are stable in the small micelles formed by the latter. In fact, DDM is good for isolation of many membrane proteins, as its comparatively large micelles tend to effectively shield the hydrophobic regions of the protein, whereas the much smaller micelles of OG expose much more of a membrane protein including the hydrophobic regions, often leading to aggregation and precipitation during the membrane solubilisation step ¹³¹. However, it is the exposure of larger regions of the protein in the smaller OG micelles that favours crystal lattice

formation. Therefore, researchers often purify in one detergent and then exchange into another for crystallisation purposes ¹³².

Novel variants of these standard detergents include LMNG and glucose neopentyl glycol (GNG), which contain a more complex double hydrophilic head group and double hydrophobic tail region based on two molecules of DDM/OG linked via a quaternary carbon atom ^{133,134}. These variants present a lower CMC than the closest equivalent standard detergent. So for example, LMNG has a CMC ~0.01 mM while DDM has a CMC ~1.7 mM ⁹⁷. The lower the CMC the greater the alkyl chain density, that also translates into stronger hydrophobic interactions with any encapsulated MP and perhaps more importantly with other detergent molecules, ultimately leading to effective self-assembly of more stable micelles ¹⁰⁰. This is clearly achieved in the case of LMNG by the doubling of the maltose head group alongside the doubling of the alkyl chain groups ^{97,133}.

To date, DDM and LMNG account for a substantial proportion of high resolution membrane protein structures deposited in the PDB database ^{100,135}. Nevertheless, substantial effort to develop and characterise novel detergents has not only made improved and tailored tools available to researchers but has provided insights into what detergent properties are crucial for enhanced stability of encapsulated membrane proteins.

Importantly, irrespective of the detergent used, it is likely that at the extraction step as well as during the purification process many, if not all, of the lipids which originally surrounded the target protein are lost. Therefore, detergent-based purification of MPs does not represent the ideal choice if the main research aim involves the investigation of the specific interactions taking place between target protein and the surrounding lipid molecules.

1.3.3 MP reconstitution into synthetic nanodisc systems

Detergent-purified membrane proteins can be reconstituted into synthetic nanodiscs ¹⁰². A nanodisc is comprised of a number of lipid molecules of defined and easily modifiable composition and an amphipathic protein molecule that wraps around the outside of the lipid patch maintaining it in solution (Figure 1.7, d). The first use of such synthetic nanodiscs employed engineered versions of the amphipathic protein human apolipoprotein A-I (apo A-I) as membrane scaffold protein (MSP) ¹³⁶. This protein, whose native function involves packaging lipids in the bloodstream and

regulating cellular cholesterol homeostasis, self-assembles into lipid/protein particles in the presence of exogenous phospholipids ¹³⁷. Synthetic derivatives of PC such as 1-palmitoyl-2-oleoyl-sn-glycero-3-PC (POPC) and dipalmitoyl-PC (DPPC), are typically used for nanodisc preparation although lipid mixtures can also be used ¹³⁸. Addition of a membrane protein in a detergent based solution to MSP and lipids, in combination with adsorbent beads to deplete the detergent, results in incorporation of the protein into the nanodisc ^{139–142}.

A wide range of downstream analyses, including but not limited to high resolution structural studies, can be performed on protein-carrying nanodiscs ¹⁰¹. The cryogenic electron microscopy (cryo-EM) structure obtained for the heat and capsaicin-activated ion channel (TRPV1), for example, allowed the identification of key protein-lipid interactions, playing a crucial role in the allosteric regulation of the channel ¹⁴³. Cryo-EM structures of nanodisc-embedded lipid flippase MsbA, an ATP-binding cassette transporter, uncovered the structural basis for lipopolysaccharide (LPS) recognition and mechanism of transport by delineating the conformational transitions of MsbA to flip LPS ¹⁴⁴. Nanodiscs are also suitable for exploring membrane protein complexes, such as the M2 muscarinic receptor, a GPCR, in complex with β -arrestin, revealing the importance of ligand-receptor-lipid environment interactions for complex formation ¹⁴⁵. Hydrogen-deuterium exchange mass spectrometry (HDX-MS) was successfully used to gain insights into conformation dynamics of outward- and inward-facing ensembles of the transporters LeuT and P-glycoprotein in nanodiscs ^{146,147}. Kapsalis *et al.* used a double electron–electron resonance (DEER) spectroscopy method to explore species-specific in-lipid structural differences of bacterial homologues of the mechanosensitive channel of large conductance (MscL) reconstituted into nanodiscs ^{148–150}. The same approach was used to explore the conformational dynamics of the bacterial multidrug transporter MdfA, reconstituted into nanodiscs ¹⁵¹. A substrate-responsive lateral gate was identified, suggesting that this protein functions via a modified model to the canonical alternating access identified for other members of the major facilitator superfamily (MFS) ¹⁵¹. Very recently, nanodiscs have been used in combination with a technique called lipidomic lipid-exchange MS (LX-MS) in order to explore lipid specific affinity for membrane proteins, such as the *Escherichia coli* ammonium transporter AmtB and aquaporin AqpZ ¹⁵².

The particle size achieved with the use of apo A-I derivative based nanodiscs is fixed and this is determined by the length of the protein sequence ^{139,142}. Therefore, the MSP needs to be carefully selected depending on the target protein and the optimal particle size has to be determined empirically, a time-consuming process. An alternative is the use of Saposin-A based nanoparticles (Salipro), that are self-tunable and are able to adapt to accommodate a wide range of MPs including molecules containing up to 56 TM regions ^{153,154}. Proteins encapsulated in Salipro nanoparticles are more thermostable and exhibit higher levels of function than the equivalent protein in DDM-based solutions ¹⁵⁴. Salipro nanoparticles have been successfully used for structural characterisation of the sarco-endoplasmic reticulum calcium ATPase (SERCA), the lipid flippase MsbA, the β -barrel assembly machinery (BAM) complex ^{155–157}.

Another alternative synthetic nanodisc approach is the use of peptidiscs, derived from an engineered version of an apo A-I derived nanodisc scaffold peptide (NSP). Peptidiscs are highly flexible and present improved hydrophobic interaction with lipid molecules due to the addition of Phe residues in their sequence ^{158–160}. Unlike the other nanodisc-based reconstitution systems described above, peptidiscs do not require addition of exogenous lipids to form and adapt to the size of the TM domain of the MP ^{158,160}. Although their use for structural studies has been limited to date, there are recent reports on the structure of the membrane-bound form of the HIV envelope glycoprotein, the structure of *Acinetobacter baumannii* ATP-ase solved by Demmer and co-workers, as well as a cryo-EM structure of *Msmeg* MmpL3 at a resolution of 3.23 Å, all obtained in this membrane mimetic system ^{161–163}. A further advantage of the peptidiscs is that by modifying their sequence to include an His tag, they can be used for extraction of untagged membrane proteins directly from the membrane, and even if present at a very low abundance. The proteins can then be characterised using MS, providing useful insights into the membrane proteome present in different cell types and under different conditions ¹⁶⁴.

There are advantages to the use of synthetic nanodiscs as they provide a more physiological environment than detergents and, depending on the system used, they allow tuning of the lipid environment surrounding a given protein, providing the means of investigating the role of specific lipids in membrane protein structure and function.

1.3.4 Polymers

Protein sample isolated in detergent-based solution is unlikely to fully represent the state of the molecule in the native membrane, and a more physiologically relevant MP extraction method utilises polymers (Figure 1.7, f) ¹⁶⁵. The original polymer, styrene maleic acid (SMA), is extensively used in the plastics industry and contains different ratios (usually 2:1 or 3:1) of hydrophobic (styrene) to hydrophilic (maleic acid) groups (Figure 1.12, e) ¹⁶⁶. SMA spontaneously forms disc-like particles in the presence of lipid membranes in a pH-dependent manner ¹⁶⁷. In this disc-like arrangements, SMA has been suggested to either punch holes in the membrane or to interact with the membrane in a coordinated manner to destabilise the bilayer structure ¹⁶⁸. Importantly, SMA encapsulates both the MP and the associated native lipid molecules, producing SMA-lipid particles (SMALPs) that are also defined as native nanodiscs. These can be isolated in the same way as detergent-extracted protein but should contain MPs that are in a more physiologically relevant state than when in detergent micelles. Given that SMALPs retain the annulus of membrane lipids that originally associated with the protein within the biological membrane, precise and detailed information about the lipidome can be obtained. This detergent-free method of MP extraction allows isolating of membrane protein complexes which are too unstable to withstand the rigours of a detergent-based extraction as well as to explore the role that specific lipids play in both protein state and function ¹⁶⁹. Many successful studies have been published where SMALPs have been used for functional and structural analyses of a variety of MPs. In particular, the functional stability of the human adenosine A2 receptor was observed to increase following protein reconstitution into native nanodiscs ¹⁷⁰; protein function was retained when the bacterial outer membrane enzyme PagP was reconstituted into SMALPs ¹⁶⁶; it was possible to identify the lipids co-purifying with the human equilibrative nucleoside transporter 1 (hENT1) ¹⁷¹; the analysis of the native lipid environment isolated with bacterial Sec translocase, SecYEG-SecA complex revealed an enrichment of negatively charged lipids, proved to be specific ¹⁷²; and the extensive network of lipid molecules interacting with the SMALP-reconstituted multidrug efflux pump AcrB was revealed by cryo-EM studies ¹⁷³.

Due to the low level of polymer material present in the nanodiscs, SMA absorbance in the UV range does not affect optical readings, therefore facilitating the use of native nanodiscs in circular dichroism and fluorescence spectroscopy analyses

¹⁷⁴. Furthermore, as the SMA polymer is negatively charged at the typical working pH range, SMALPs migrate intact within a native gel electrophoresis system (SMA-PAGE) ¹⁷⁵. Using this methodology, particle separation on the gel is consistent with the expected mass of membrane proteins in different oligomeric states or with the formation of large protein complexes, while retaining native lipids. Moreover, Pollock and colleagues also blotted protein-SMALPs from native gels onto grids for negative stain analyses and successfully imaged intact particles ¹⁷⁵.

However, SMA presents some incompatibility issues with certain divalent cations, particularly Mg^{2+} and Ca^{2+} , making certain assays relying on these ions difficult to perform, such as the ATP hydrolysis assay, dependent on Mg^{2+} ¹⁷⁶. SMA can also interact with Ni^{2+} -based immobilised metal affinity chromatography resins used for isolation of many membrane proteins in such a way as to limit the purification yield ¹⁷⁷. Furthermore, SMA is pH sensitive and the fixed size of the discs formed (typically in the range of 10 – 12 nm diameter) may rule out the successful encapsulation of membrane proteins and membrane protein complexes that extend beyond the specific range ¹⁷⁸.

Optimised SMA derivatives, such as a polymethacrylate (PMA) polymer, have been developed to overcome these limitations (Figure 1.12, f) ¹⁷⁹. Lavington and Watts extracted the GPCR Neurotensin 1 (NTS-1) using PMA, revealing that this polymer was less sensitive to divalent cations than SMA, thus allowing the isolation of the protein using Ni^{2+} -NTA resin ¹⁸⁰. Although yields of NTS-1 were lower than the control condition of DDM + cholesterol hemisuccinate (CHS), the GPCR in PMA exhibited similar activity to that showed by the receptor prepared in DDM + CHS at a higher protein concentration, indicating that NTS-1 encapsulated in PMA may be more functional. Another SMA-derivative developed to increase the tolerance to divalent cations is the polyanionic diisobutylene/maleic anhydride (DIBMA) copolymer ¹⁸¹. DIBMA forms nanodiscs with a diameter ranging between 12 and 29 nm, thus providing a more accommodating environment for larger MPs and complexes ¹⁸¹. DIBMA has been successfully used to extract the β_2 -adrenergic receptor (β_2AR) from membranes, and the extracted protein showed similar ligand-binding characteristics in both DIBMA and DDM but a higher thermal stability in the polymer ¹⁸². Finally, inulins are an additional group of nanodisc-forming polymers, synthesised using naturally recurring fructo-oligosaccharides as starting material ¹⁸³. The advantage of the inulin polymers is that they are non-ionic, thus causing minimal interference to native interactions of a

given membrane protein. For their non-ionic nature, inulins allowed successful reconstitution of two coexisting oppositely-charged cytochrome P450s ¹⁸⁴.

Although polymers represent a huge step forward for the structural and functional characterisation of MPs, they do not seem to be fully representative of a membrane environment. A recent study indicates that membrane lipids are more packed in a polymer environment than in a biological membrane, therefore potentially affecting membrane curvature and consequently physiological protein folding and conformation ¹⁸⁵. However, the ability to encapsulate and allow the study of native membrane protein-lipid complexes makes these an attractive alternative to detergent-extracted proteins, including detergent-based protein that is ultimately reconstituted into an artificial and user-defined lipid bilayer.

1.4. Experimental approaches for MP characterisation

A great variety of techniques and methods are employed in the investigation of membrane protein structure, function, dynamics and oligomeric arrangement (Figure 1.13). MPs reconstituted in the different mimetic systems described in the previous sections can be analysed by an array of experimental techniques, including but not limited to electrophoresis, X-ray crystallography, cryogenic electron microscopy (cryo-EM), nuclear magnetic resonance spectroscopy and other biophysical techniques, that will be presented in the sections below. Depending on the scope of the study, each of these diverse methodologies presents advantages and potential synergies but also limitations to their use.

1.4.1 Separative Techniques

A simple way to separate membrane proteins based on their molecular weight (MW) is to use Sodium Dodecyl-Sulfate Polyacrylamide Gel Electrophoresis or SDS-PAGE ^{186,187}. SDS is an anionic and denaturing detergent that coats the unfolded proteins with negative charges along their length. Hence, charge and shape do not influence the migration of the proteins, with the MW being the sole differentiating factor ^{186,187}. With stains such as colloidal Coomassie blue, the migration can be observed and compared with a reference molecular ladder to determine MW ¹⁸⁸. This technique is particularly useful when studying a new target protein, to determine its size and to get preliminary information on its migration pattern on a gel. However, it needs to be noted that this technique is mainly tailored for soluble proteins and it can be misleading

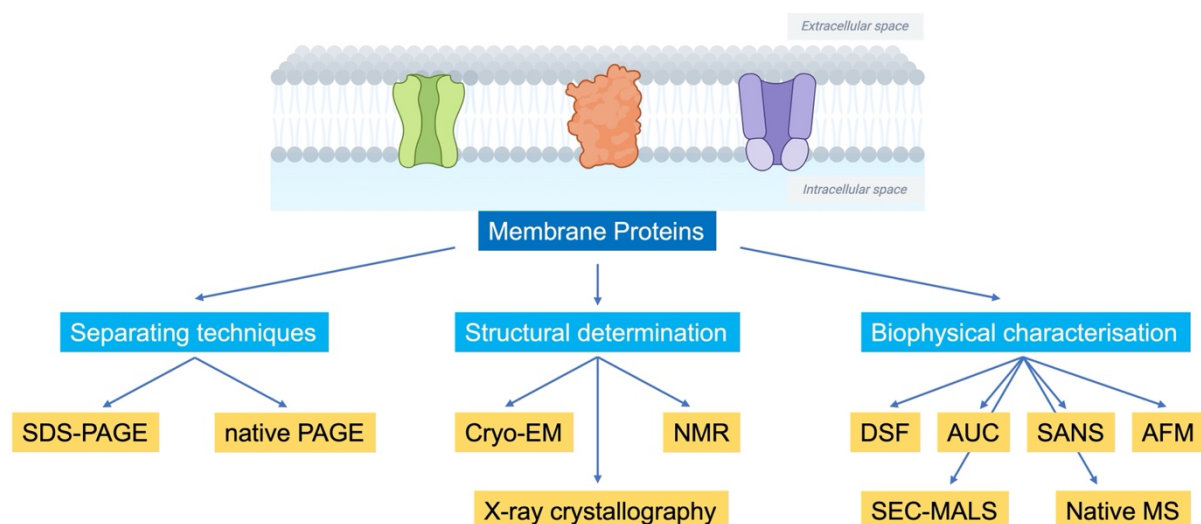


Figure 1.13. Graphical summary of some of the common experimental techniques used for the detection and characterisation of membrane proteins. SDS-PAGE: Sodium Dodecyl-Sulfate Polyacrylamide Gel Electrophoresis; cryo-EM: cryogenic electron microscopy; NMR: nuclear magnetic resonance; DSF: differential scanning fluorimetry; AUC: Analytical ultracentrifugation; SANS: Small-angle neutron scattering; AFM: Atomic Force Microscopy; SEC-MALS: Size exclusion chromatography multi-angle light scattering; native MS: native mass spectrometry. Membrane protein schematic was generated through BioRender.com.

when used for MPs, due to the different migration patterns often observed for these samples¹⁸⁹. In contrast, native PAGE is a strategy used to study proteins in their native biological state. This technique has value in studying membrane proteins, as it allows the analysis of samples, without subjecting them to denaturing agents and without altering their biological composition¹⁷⁵. At the pH used for Tris-glycine PAGE, the SMA polymer in SMALPs as well as the MSPs in synthetic nanodiscs are negatively charged, thus migration of the particles on the gel can be achieved by addition of a glycerol-based loading dye. Intact particles separate according to their MW, thus allowing the investigation of oligomeric arrangements and protein complexes. Furthermore, downstream analyses, such as lipidomics or cryo-EM, can be carried out directly on the gel-excised samples, as protein folding as well as original lipids are retained during electrophoresis¹⁷⁵.

1.4.2 Structural determination of MPs

X-ray crystallography has been used for decades and represents the primary procedure used to obtain protein structures at near atomic or even atomic resolution^{187,190}. It relies on the generation of electron density maps that illustrate the shape and structure of the crystallised macromolecules. However, it still suffers from many limitations especially in the crystallisation phase, as membrane proteins often require

an extensive optimisation of conditions before obtaining well-ordered crystals ¹⁹¹. Furthermore, the X-ray radiation that the crystals are exposed to is an important challenge in X-ray crystallography, compromising the quality of the collected data. Another limitation is the failure to experimentally visualise hydrogen atoms, essential for accurate atomic modelling of structures, allowing to gain insights into the protein's hydration shell as well as promoting the rational design of ligands ¹⁸⁷.

Cryo-electron microscopy (cryo-EM) enables direct visualization of a vitrified protein sample on a thin ice layer with a low-dose transmission electron microscope (TEM) operating at liquid nitrogen temperature ¹⁹². Cryo-EM can be used to determine the structure of isolated biomolecular complexes, covering a wide range of molecular masses, spanning from small proteins to large viruses and cells ¹⁹³. Despite cryo-EM's ongoing challenges in visualising small proteins with a MW below 100 kDa, the structural analysis of GPCRs (~40–50 kDa) remains at the limit of this technique. In recent years, cryo-EM has undergone a “resolution revolution” with researchers able for example to acquire the structure of Apoferritin at 1.25 Å resolution ^{194,195}. Cryo-EM has been particularly useful in determining the structural details and distinct configurations of bacterial RND transporters, offering prospects for designing structure-guided drugs intended for combating both MDR and XDR bacteria ¹⁹⁶. Another cryogenic technique called cryo-electron tomography (Cryo-ET) can be used to visualise membrane proteins at a 3D level within the context of intact cellular membranes, although resolution is still low ¹⁹⁷.

Nuclear Magnetic Resonance (NMR) is another technique that allows the acquisition of MPs 3D structures in their native lipid bilayer, or in reconstructed near-native composition of the lipid bilayer at an atomic resolution ^{198,199}. It is based on the detection of different energy levels after applying a magnetic field on an atom which has non-zero nuclear spins, such as ¹H NMR and ¹³C NMR, but there are also other types such as ³¹P NMR and ¹⁹F NMR ¹⁸⁷. Unlike other methods, such as cryo-EM and X-ray crystallography that only gives static structures of proteins and encounter limitations when analysing intrinsic disordered proteins, NMR provides dynamic information ¹⁸⁷. There are 2 main types of NMR: solution and solid-state NMR. The first approach is particularly useful for studying small to medium-sized proteins, as it has a limiting protein size of 100 kDa ²⁰⁰. Solid-state NMR on the other hand does not have size limitations and preserves the anisotropic nature of nuclear spin interactions, crucial for studying proteins embedded in a lipid bilayer ²⁰¹. Recently, through the

combination of NMR with the magic angle spinning (MAS) technique, it has been possible to study a protein's structure within its native environment, capturing lipid-protein interactions, alongside the investigation of dynamic processes like conformational changes and ligand binding with enhanced resolution and sensitivity²⁰².

1.4.3 Biophysical characterisation

1.4.3.1 Study of MPs dynamics

Atomic Force Microscopy (AFM) is a powerful tool if the user is interested in the visualisation of large protein complexes in various conformational states^{187,203}. AFM has been used, for instance, to examine the conformational changes of MPs upon ligand binding and provide details about their allosteric processes²⁰⁴. Further, investigating the dynamic behaviour and stability of membrane protein complexes has been possible with the use of AFM-based single molecule force spectroscopy, probing the protein-protein and protein-lipid interaction network²⁰⁵. Moreover, AFM can also provide valuable information on the nanomechanical properties of MPs, such as elasticity, flexibility, and response to mechanical stress²⁰⁶.

Insights into the dynamics of MPs and how they behave in their natural lipid context can also be obtained with the use of neutron scattering²⁰⁷. For instance, the general shape, size, and oligomeric state of membrane proteins can be determined using small-angle neutron scattering (SANS), providing an understanding on higher order assembly and structural organisation¹⁸⁷.

1.4.3.2 Oligomeric arrangements of MPs

As many membrane proteins in their natural cellular environment function as oligomeric complexes, it is important to not only assess the quality and integrity of the reconstituted MPs, but also their oligomeric state in the presence of detergent/lipid solutions before undertaking any functional, biophysical, or structural studies^{208–210}. Size-exclusion chromatography (SEC) has been the most common method in providing information regarding the MW and homogeneity of protein samples. However, the method is based on retention volumes usually calibrated for globular water-soluble proteins. In fact, for membrane proteins, the relative size determined from SEC relates to the protein-detergent or nanodisc particles²⁰⁸. SEC multi-angle

light scattering (SEC-MALS) is an accurate and versatile biophysical technique that is able to determine the composition, mass, and oligomeric state of membrane proteins in solution ^{189,211}. The SEC column is connected in-line to UV, light scattering, and refractive index detectors. The use of a SEC column is essential to achieve a physical separation of aggregates, complexes, and free micelles / empty nanodiscs. This allows for the different mass components to be independently analysed by the detectors ²¹¹.

The development of 'soft ionisation' methods like electron spray ionisation has facilitated the application of mass spectrometry (MS) for the study of intact complex biomolecules ²¹². Native MS uses non-denaturing ionisation conditions, usually with nano-electrospray ionisation (nano-ESI), to preserve non covalent complexes upon transfer from solution to the gas phase for mass analysis ²¹². In native MS experiments, volatile, aqueous and near physiological pH buffers are used in combination with mild temperatures and voltages. By preserving non covalent interactions this technique probes the investigation of protein–protein and protein–ligand interactions, as well as subunit architecture and complex stoichiometry ²¹³. Among the advantages of native MS is that it is compatible with a wide mass range, it does not require labelling and it is a rapid and highly sensitive technique, requiring only a few microliters of sample at high nanomolar concentration ²¹³. ESI-MS was an enabling technology for shotgun lipidomics, allowing identification and quantification of hundreds of lipid species without the need for chromatographic separation or fragmentation ²¹⁴. For what concerns studies on membrane proteins, the high concentrations of detergent used for solubilisation purposes can form heterogeneous detergent micelles that can interfere with the analysis. Thus, membrane protein complexes initially eluded analysis by MS methods until the realisation that they could be transferred to the gas-phase while still encapsulated in detergent micelles ²¹⁵. More recently, a nano-electrospray methodology followed by removal of detergent through collisional activation was applied to membrane proteins, revealing that intact folded proteins could be stripped of their detergent and analysed for oligomerisation, lipid binding and, most importantly, the structural and functional consequences of lipid–protein interactions ^{215,216}.

MS of intact proteins, including MPs in SMALPs and in MSP-based nanodiscs, can also be performed using laser-based ionisation techniques, such as laser-induced liquid bead ion desorption (LILBID) ^{217–220}. This method involves the generation of ions from aqueous droplets using an IR laser, whose power is tuned to modulate the degree of activation, replacing the collisional activation step in nano-ESI ²¹³. Peetz and co-

workers compared the performance of LILBID to nano-ESI for native MS of a range of membrane proteins ²²¹. Compared to nano-ESI, LILBID presented a lower mass resolution than nano-ESI, likely due to incomplete desolvation. However, LILBID was more tolerant to non-volatile buffers and salt than nano-ESI, and more importantly caused less unintended oligomer dissociation ^{213,221}.

The distinct oligomeric arrangements adopted by many MPs can also be discriminated in respect to their sedimentation velocity during analytical ultracentrifugation (AUC) ²²². AUC is a physical biochemistry method that allows for the real-time observation of the sedimentation of protein species in solution following the application of a strong centrifugal field ²²³. While in sedimentation velocity (SV), the evolution of the sedimentation process is monitored as a function of time, in sedimentation equilibrium (SE) experiments equilibrium concentration gradients at lower centrifugal fields are examined. SV and SE are complementary methodologies for the characterisation of protein behaviour in solution ²²⁴. Because AUC relies on the principle of mass conservation and the fundamental laws of gravitation, it has broad applicability and can be used for a variety of macromolecules in solution to study their mass and size, hydrodynamic shape, interactions and the formation of multi-protein complexes in their native state under biologically-relevant solution conditions ²²⁵.

In a sedimentation velocity experiment, solute molecules migrate toward the bottom of the cell given the application of a sufficiently large centrifugal force field. The rate of sedimentation in a centrifugal field, is described by the Svedberg equation (Eq. 1.1):

$$s = v / \omega_2 r = M (1 - \bar{v} \rho) / Nf \quad (\text{Eq. 1.1})$$

where s is the sedimentation coefficient (in Svedberg units, where $1 \text{ S} = 10^{-13} \text{ sec}$), v is the velocity of the molecule, $\omega_2 r$ is the strength of the centrifugal field, M is the molecular mass, f is the frictional coefficient (directly related to macromolecular shape and size), ρ is the solvent density, N is Avogadro's number, and $\bar{v}$ is the partial specific volume of the solute. For proteins, $\bar{v}$ in most cases can be calculated accurately from the amino acid composition. The solvent parameters (density and viscosity) can be either measured experimentally or calculated from the solvent composition using tabulated data ²²³.

The fundamental requirements for the sample are: 1) distinct optical properties between solute and solution components, 2) sedimentation or floating at a reasonable rate at an experimentally achievable gravitational field and 3) chemical compatibility with the sample cell ²²⁵. A great advantage of AUC is that, unlike other biophysical techniques, macromolecular properties can be probed without the need for standards, labels, assumptions, or physical matrices that could interfere with the solution behaviour of particles, that can range from single kilodaltons to gigadalton structures ²²².

1.4.3.3 *MPs stability*

The fluorescence dye-based differential scanning fluorimetry (DSF) method, also known as protein thermal-shift assay, is a well-established approach to assess the stability of soluble and membrane proteins in the presence of different buffer formulations, detergents, and small molecules ^{208,226–228}. Protein stability decreases proportionally to temperature increase and as a result the protein starts to unfold, allowing the binding of hydrophobic fluorescent dyes. The DSF method is based on the relationship between protein stability and its Gibbs free energy of unfolding (ΔG_u). The fluorescence intensity is plotted as a function of the temperature. As a result, a sigmoidal curve is generated where the inflection point of the curve is the “melting point” temperature (T_m) value, corresponding to presence of equal amounts of folded and unfolded protein ($\Delta G_u = 0$). It is by comparing T_m values at the different tested conditions, that valuable insights into the stability of the protein can be obtained ^{208,229}. SYPRO orange is the most commonly used fluorescent dye when working with soluble proteins, as it binds to the usually buried hydrophobic regions of soluble proteins that become solvent-exposed as they unfold ^{208,229}. For this reason and the fact that this dye also binds to detergents and lipids, it is not suitable to use for MP samples, as it leads to a non-specific increase in fluorescence background. To overcome this problem, a highly reactive thiol-specific fluorochrome 7-diethylamino-3-(4'-maleimidylphenyl)-4-methylcoumarin (CPM) can be used ²³⁰. The CPM dye has a high preference for thiols rather than for nucleophiles, which preferentially and specifically react to cysteine side chains. As a result of protein unfolding, cysteine residues in the core of the protein become solvent-exposed and react with the CPM molecules ²³⁰.

Recent advancements have allowed the development of a label-free DSF (nano-DSF) based on the changes of tryptophan fluorescence intensity upon protein unfolding (as tryptophan fluorescence is strongly related to its chemical environment). In addition, the technology is also able to measure tryptophan fluorescence intensity at two different wavelengths (330 and 350 nm). The ratio between the two wavelengths allows for a more accurate calculation of T_m values. The method is applicable to both soluble and membrane proteins ^{231,232}.

1.5. Computational approaches for MP characterisation

A complementary perspective to the experimental approaches used for MP characterisation could be achieved through molecular dynamics (MD) simulations, potentially revealing new atomistic insights into protein behaviour. MD simulations can be compared to a “computational microscope”, as described by Marrink and colleagues ^{233,234}. They provide important information about molecular motions, interactions and dynamics over a given time period and within solvent and membrane environments. The atomic interactions for each of the components in the simulated system are described by a force field (FF) ²³⁵. A FF defines each molecule’s interactions through solving Newton’s equations of motion using a potential energy function shown in Equation 1.2:

$$E_{potential} = E_{bond} + E_{angle} + E_{dihedrals} + E_{VdW} + E_{electrostatic} \quad (\text{Eq. 1.2})$$

where $E_{potential}$ is the potential energy of the system, calculated as the sum of the bonded interaction terms (E_{bond} , E_{angle} and $E_{dihedral}$) and the non-bonded interaction terms (E_{VdW} and $E_{electrostatic}$) ^{235,236}.

The quality of a simulation is reliant on how well this set of parameters represents the atomic system. Knowing the initial atomic position and velocity at time = 0, the potential energy function is calculated for each atom pair. As the forces on each atom equal to the negative gradient of the potential energy, their corresponding accelerations can be calculated through numerical integration of Newton’s law of motion. New positions and velocities can consequently be derived for a small interval of time dt , usually 2 femtoseconds. At this point the MD cycle can start again to calculate new atomic positions and velocities at increasing intervals of time (e.g. $2dt$, $3dt$ etc.). The result is

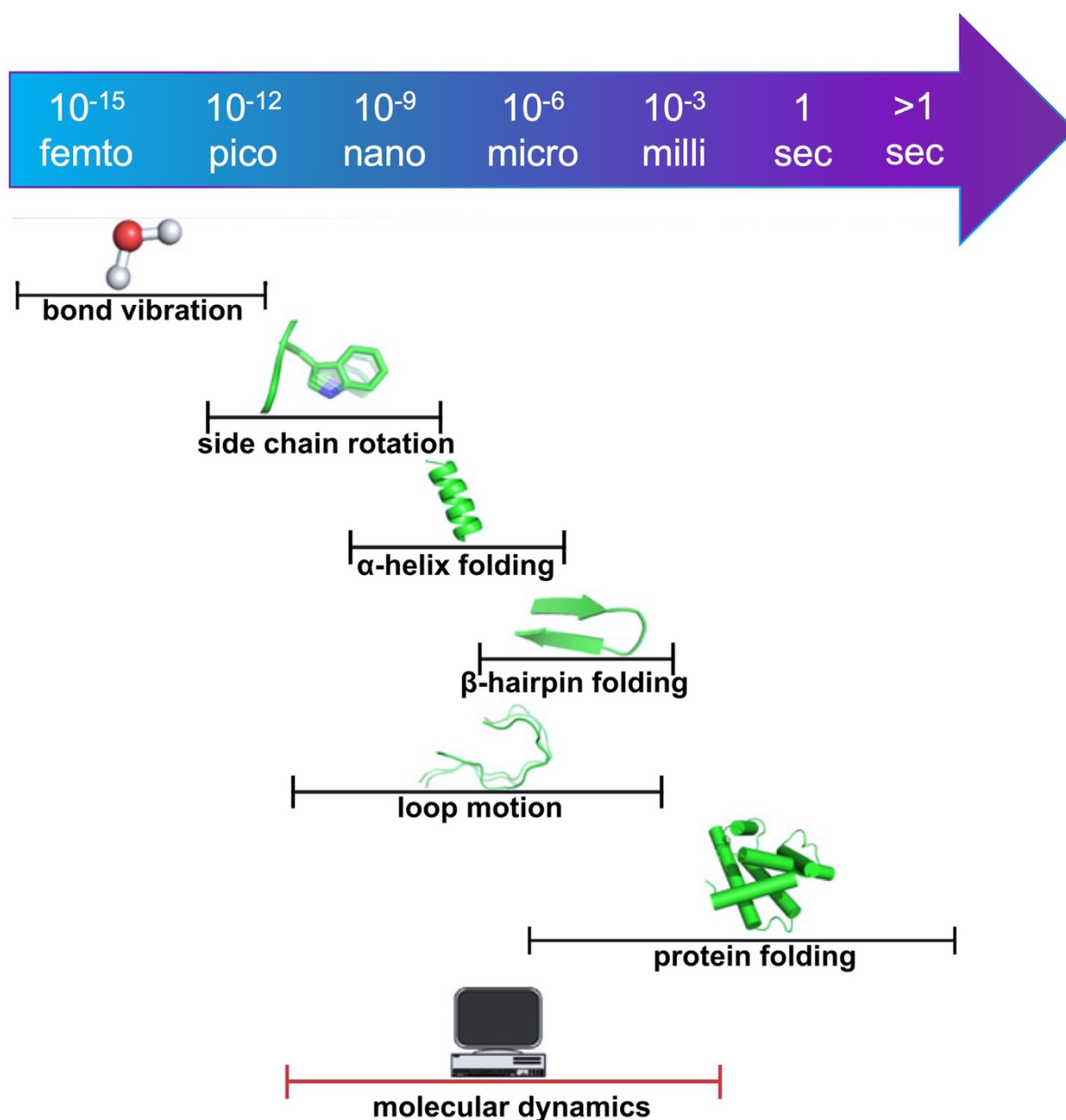


Figure 1.14. Different motions in proteins and their relative time scale. The time scale for the motions depends on the size of the structural elements and the environment. While bond vibration is in the femto- to picoseconds range, side chains rotation is between the pico- and nanoseconds time scale. The wide time frame can be explained by differentiating side chains on the surface of a protein that move faster than side chains buried in the protein, thus forming many contacts to other side chains. The α -helix, β -hairpin and protein folding also largely depends on size. It can take seconds especially for full protein folding events. MD simulations are generally used for a small part of the dynamic range in a protein, stretching up to the milliseconds range depending on the resolution used. Adapted with permission ²³⁶. Copyright 2012, Elsevier.

a trajectory which is an ensemble of each atom's position as a function of time ²³⁶. The time step in numerical integration is one of the limiting factors to simulate systems for relevant time scales, as biological processes usually extend over many orders of magnitude, ranging from picoseconds for side chain rotational motions up to seconds for protein folding (Figure 1.14) ²³⁷. Other factors to consider are the number of total

atoms in the system and the level of resolution required by the user to investigate a specific phenomenon, since both play roles in determining the computational cost. In some cases, an extended simulation time scale, without enormously compromising on resolution, can be achieved using a coarse-grain (CG) approach, where atoms or groups of atoms are represented as beads, to reduce the degrees of freedom that are extremely costly to simulate ^{233,238}.

Many MD simulations software packages have been developed for membrane-protein modelling, such as CHARMM, AMBER and GROMACS, and FFs are constantly updated with more advanced algorithms ^{239–241}. For example, membrane models have increased in complexity over the years thanks to technology advancements. From the pioneering stage of 30 years ago, basic properties of 1- or 2-component membranes were explored around the millennium. From then on, the complexity of the simulated membrane systems as well as length of simulations were gradually increased, culminating in the current era of more realistic membrane models (Figure 1.15) ²³³. AMBER stands for Assisted Model Building with Energy Refinement, and it comprises a series of programs for biomolecular simulations, using specific and publicly available molecular FFs. The equation defining the AMBER force field terms is described in more detail in Figure 1.16 ²⁴⁰.

The simulation workflow is constituted by three main steps: system preparation, simulation and analysis of trajectories, which are carried out by separate programs, thus resulting in an easier manipulation of individual parts of the program suite. Within the AMBER package, LEaP is used to prepare the system prior to running the simulation program *sander*. Given a PDB file, LEaP generates a *prmcrcd* file, containing the initial Cartesian coordinates, and a *prmtop* (parameter-topology) file that lists the FFs and their relative parameters needed for the simulation. The latest FF to use for proteins is ff19SB, while lipid21 and TIP3P are used to model the lipid counterpart and water, respectively ^{242–244}. These files are necessary to define topology and coordinates that will be fed into the initial energy minimisation step carried out prior to performing any MD simulation in order to identify a local minimum and find a structure of lower energy ²⁴⁰.

A range of different temperatures and pressures can then be applied to the initial model to observe any variation in coordinate trajectories over time, long enough to detect any low-frequency changes.

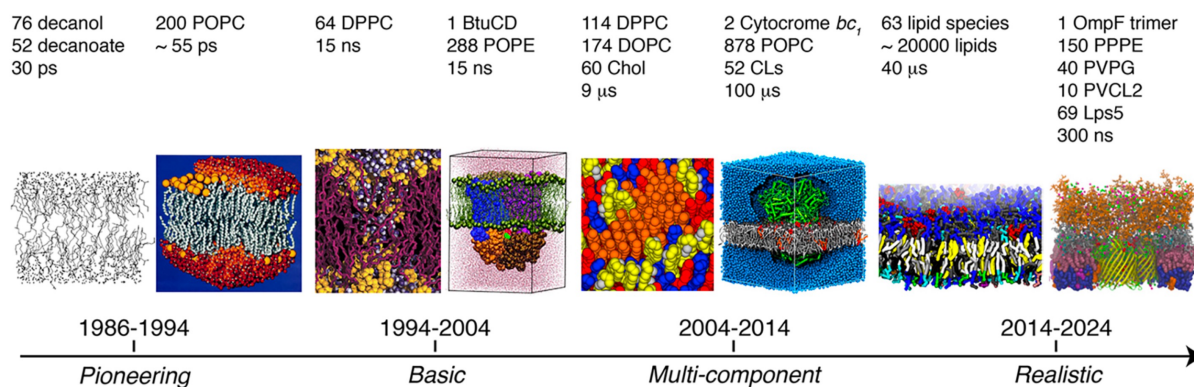


Figure 1.15. Evolution of membrane models from the pioneering stage of 30 years ago up to the realistic complexity of current times. POPC, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine; DPPC, 1,2-dipalmitoyl-sn-glycero-3-phosphocholine; POPE, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine; DOPC, 1,2-dioleoyl-sn-glycero-3-phosphocholine; Chol, cholesterol; CLs, cardiolipins; PPPE, 1-palmitoyl-2-palmitoleoyl-phosphatidylethanolamine; PVP, 1-palmitoyl-2-vacenoyleoyl-phosphatidylglycerol; PVCL2, 1,10-palmitoyl-2,20-vacenoyleoyl cardiolipin; Lps5, *E. coli* R1 lipopolysaccharide core with repeating units of O6-antigen. Reprinted with permission ²³². Copyright 2019, American Chemical Society.

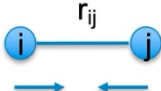
To date, many biomolecules have been studied in MD simulations, including proteins, nucleic acids, glycans and membranes for which recently developed FFs are able to reflect the huge variety of lipid structures found in nature, thus providing more realistic simulations ²³³.

Of particular interest is the great number of bacterial membrane proteins that have been investigated by MD including many multidrug efflux pumps, such as AcrB, MtrD, AdeB and MexB ²⁴⁵. MD simulations helped elucidating the molecular mechanism of substrate binding and transport, thus shedding a light on the mechanism of antibiotic resistance exerted by these proteins ^{246–248}. The protein-substrate interactions and their dynamics can be investigated through a computational lens, allowing a more rational antibiotic design as well as the development of inhibitors targeting these pumps ²⁴⁹.

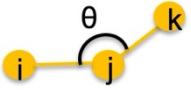
Regarding bacterial RND family members, with a particular focus on *Mtb* MmpL proteins, Sandhu and Akther performed MD simulations on the homology modelled MmpL5 assembled in a trimer and embedded it in a 1-palmitoyl, 2-oleoyl- sn-glycero-3-phosphocholine (POPC) lipid bilayer ²⁵⁰. They identified two drug binding sites in this RND protein positioned at opposite ends of the continuous transport channel that lead substrates from the cytoplasmic to the periplasmic space. This provided the first evidence regarding the potential transport mechanism of these efflux pumps in *Mtb* ²⁵⁰. More recently, Su and colleagues used targeted MD simulations to elucidate the potential mechanism of TMM transport mediated by *Msmeg* MmpL3 ³. They used the

$$V_{\text{AMBER}} = \sum_i^{n_{\text{bonds}}} b_i (r_i - r_{i,\text{eq}})^2 + \sum_i^{n_{\text{angles}}} a_i (\theta_i - \theta_{i,\text{eq}})^2 + \sum_i^{n_{\text{dihedrals}}} \sum_n^{n_{i,\text{max}}} (V_{i,n}/2) [1 + \cos(n\phi_i - \gamma_{i,n})] \\ + \sum_{i < j}^{n_{\text{atoms}}} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + \sum_{i < j}^{n_{\text{atoms}}} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}},$$

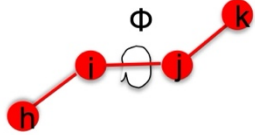
1. Bond stretching



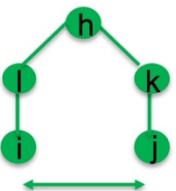
2. Angle bending



3. Torsion



4. Van der Waals



5. Electrostatic

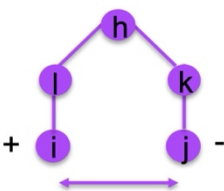


Figure 1.16. Potential energy function at the core of AMBER force field. Numbers indicate the different terms of the equation, including bonded interactions, such as bond stretching (1), angle bending (2) and dihedrals (3) as well as non-bonded interactions, such as Van der Waals (4) and electrostatic (5) terms.

structural information on TM lipid binding sites and by using targeted MD suggested a possible path for TMM molecules to get transported from the outer leaflet of the inner membrane to the periplasm passing through MmpL3 central periplasmic cavity³. In a recent paper, Carbone and colleagues investigated the allosteric inhibition mechanism that SQ109 exerts on TMM translocation by comparing MD results obtained with MmpL3 in its apo and inhibitor-bound state⁹⁰.

Over the last few years, many pharmaceutical companies have shown an increasing interest in implementing computational approaches in drug discovery to potentially save time and money, the two main factors currently limiting the process. Investments of huge sums of capital and resources have been employed with the main aim of replacing a wet-lab drug screening approach with an accurate and higher-throughput Computer Aided Drug Design (CADD) method for *de novo* identification or repurposing of drugs²⁵¹. Alongside multinational corporations, many start-ups and academic research groups have also joined the challenge of developing *in silico* drug discovery platforms. The advancement of computer technologies, especially with the advent of powerful quantum computers, as well as the invention of new efficient ways

to store and manage data online (e.g. cloud storage) have accelerated progress in the computational field ²⁵².

It is evident that in this scenario MD simulations, molecular docking and machine learning algorithms represent important tools to computationally describe complex biological systems and protein-protein interactions, thus helping in the discovery of new therapeutics ²⁵². Also, considering the increasing focus of membrane proteins as drug targets and the intrinsic difficulty in expressing and purifying them in a stable form, MD simulations could provide a relevant model for protein-ligand interactions (e.g. drugs). However, it needs to be noted that all computational tools require the availability of high-resolution 3D structures of protein and ligands, knowledge of protein function, as well as of the intermolecular forces involved in establishing the interactions between binding partners. Therefore, like any other approaches, *in silico* methods are not infallible, thus are more valuable when employed in combination with other drug discovery tools and with laboratory-based techniques ²⁵².

1.6. Summary and research aims

Substantial research has been carried out over the last 20 years or so in the development of new tools for membrane protein manipulation and characterisation. As it is important to note that there is still no one-fits-all approach that is suitable for all membrane proteins, the wide range of available tools can disorient the user and can make it challenging to know where to start with a given protein of interest. When investigating a new MP is useful to perform some detergent-based screening at first. Even if detergents are not ultimately chosen for the target molecule, this will provide preliminary information on the behaviour of the membrane protein in aqueous solution and will form the basis for further comparisons in native/synthetic nanodiscs and other membrane mimetic systems. Once a high-resolution structure for the target MP is available, computational approaches can be used to investigate specific phenomena at atomistic detail, such as the exploration of protein-protein or protein-lipid interactions, the determination of substrate/drug binding sites as well as the investigation of molecular mechanisms of protein function.

Results generated in this project will show the immense value of a combined experimental-computational approach as a powerful tool for the study of biological

systems. In particular, this thesis will focus on three main aims: 1) the expression and isolation of high purity MmpL3 protein in detergent micelles, SMALPs and MSP-based nanodiscs; 2) the investigation of the oligomerisation state of the mycobacterial protein in the aforementioned membrane-mimetic systems using several biophysical and structural methods; 3) the investigation of MmpL3 behaviour in a lipid bilayer as well as its likely protomer-protomer interface through atomistic MD simulations.

2. MmpL3 protein expression and purification

2.1 Introduction

The starting point for any experimental analysis of a membrane protein is the production of that protein, typically in a recombinant expression system. As it was not feasible to express the *Msmeg* and *Mtb* MmpL3s in the native organisms, the proteins were expressed in *E. coli*. Starting from standard protocols for bacterial overexpression, some optimisation was necessary to produce high purity, high homogeneity protein. As extensively described in the literature, extraction of MPs into aqueous based solution represents a challenge as the native lipid environment plays a crucial role in retaining protein function and correct folding^{131,253}. Several experimental approaches were used to purify *Msmeg* and *Mtb* MmpL3 and these will be discussed in this Chapter. A detergent-based purification was the first method implemented, and was followed by reconstitution into synthetic nanodiscs, to better mimic the lipid bilayer and increase protein stability^{97,98,100,141,142,254,255}. In addition, polymer-based extraction was also performed. In particular, the styrene-maleic acid copolymer acts like a detergent molecule and is added directly to membrane samples for solubilisation^{174,176,178}. The advantage is that the target protein is encapsulated directly in its native environment, therefore retaining the original lipid composition of the expression host¹⁷⁶. The three approaches were used to successfully purify and reconstitute MmpL3, allowing to carry out downstream analyses to characterise the protein stability and behaviour in different membrane mimetic systems.

2.2 Materials and Methods

2.2.1 Cloning of full-length *Msmeg* and *Mtb* MmpL3

Plasmid pET22b/42a encoding the C-terminal His₈-tagged full-length (FL) *Msmeg* MmpL3 (UniProtKB entry: A0QP27) was obtained from Dr Zhujun Xu, from Shu-Sin Chng's group (National University of Singapore). *Mtb* MmpL3 FL (UniProtKB entry: P9WJV5) in pMA-T vector was kindly donated by Dr Oliver Adams, from Professor Simon Newstead's group (University of Oxford). The full-length *Mtb* MmpL3 gene was amplified by Polymerase Chain Reaction (PCR) using the primers 5'-

AAGGAGATATACATATGTTTGCATGGTGGGGTCGTAC-3' (FW *Mtb* NdeI) and 5'-GGTGGTGGTGCTCGAGCAGACGACCTTCACGACGC-3' (RV *Mtb* XhoI) incorporating NdeI and XhoI as restriction enzyme sites. The gene was then cloned into the expression vector pET22b/42a which incorporates a C-terminal His₈ tag after a 2 aa-linker (LE). Both vectors were initially transformed into NEB[®] 10-beta competent cells for DNA amplification. The sequence was verified, and the DNA transformed into BL21(DE3) wild type (WT) *E. coli* competent cells, and glycerol stocks were prepared from single colonies.

2.2.2 Bacterial expression of *Msmeg* and *Mtb* MmpL3

For protein expression, bacterial culture was grown in Luria-Bertani (LB) broth under Ampicillin selection (100 µg/mL) at 37 °C and 220 rpm until the optical density at 600 nm (OD₆₀₀) reached 0.5-0.7. Isopropyl β-d-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM and the culture harvested after 4 hours (h) by centrifugation at 4,000 *g* for 10 minutes at 4 °C. Cell pellets were resuspended in Buffer A, containing 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 100 µg/mL lysozyme, 50 µg/mL deoxyribonuclease I (DNaseI) and EDTA-free protease inhibitor (SIGMAFAST[™], Sigma-Aldrich). Cells were lysed at 27 kpsi, twice, and the lysate centrifuged at 12,000 *g* for 10 minutes at 4 °C to remove unbroken cells. The supernatant was then ultracentrifuged for 2 h at 200,000 *g*. Membranes were homogenised in Buffer B (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 10% v/v glycerol and protease inhibitor) and stored at -80 °C.

2.2.3 MmpL3 purification in detergent

Membranes were solubilised by constant stirring for 1.5 h at 4 °C in n-Dodecyl-β-D-maltopyranoside (DDM, Anatrace) at a final 1% w/v concentration. The extracted membrane proteins were then separated from non-solubilised material by ultracentrifugation at 200,000 *g* for 1 h, at 4 °C. A 10 mL aliquot of nickel-nitrilotriacetic acid (Ni²⁺-NTA) Superflow resin (Qiagen) was pre-equilibrated with Affinity Buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 10% v/v glycerol, 0.03% w/v DDM and 10 mM imidazole) and mixed with the membranes for 2 h, at 4 °C. The resin mixture was then loaded onto a poly-prep/glass econo-column chromatography column (Bio-Rad) and

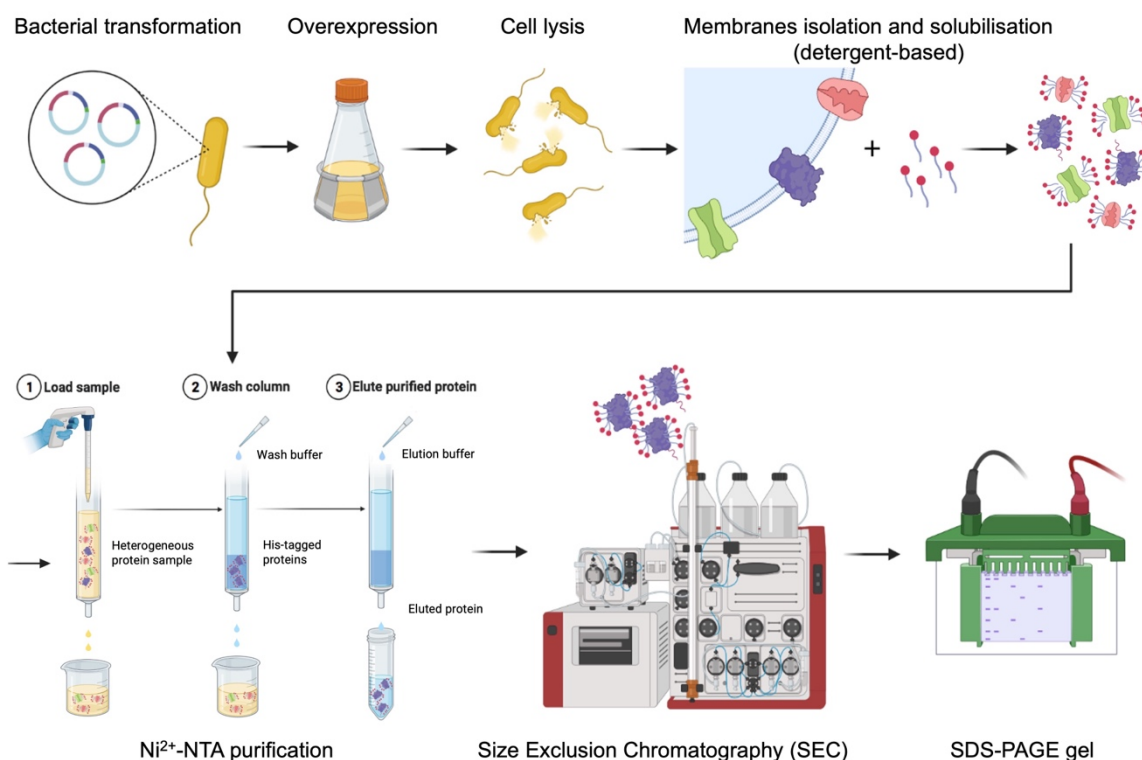


Figure 2.1. MmpL3 expression and detergent-based purification workflow. Briefly, bacteria were transformed with plasmid carrying the target gene and proteins were recombinantly expressed at 6 or 12 L scale. After cell lysis, membranes were isolated and membrane proteins solubilised using the detergent of choice. The His-tagged protein was isolated using a two-step protocol (Ni²⁺-IMAC followed by SEC). Protein fractions from main SEC peak were run on a SDS-PAGE gel to verify purity and the MW. Figure was generated using BioRender.com.

allowed to drain by gravity. The resin was then washed with 10x column volumes (CVs) of Wash Buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 10% v/v glycerol, 0.03% w/v DDM and 30 mM imidazole). MmpL3 was then eluted from the column by addition of Elution Buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 10% v/v glycerol, 0.03% w/v DDM and 300 mM imidazole) and the eluted protein concentrated to ~500µL using 100 kDa MWCO centrifugal filters (Millipore Amicon™ Ultra, Merck). The concentrated sample was then loaded onto a Superdex® 200 Increase 10/300 gel filtration column pre-equilibrated with SEC buffer A (20 mM Tris HCl pH 8.0, 150 mM NaCl, 10% glycerol, 0.03% w/v DDM). Fractions corresponding to the main protein peak were concentrated using 100 kDa MWCO centrifugal filters and either used fresh or stored at -80 °C for future experiments (Figure 2.1).

2.2.4 Use of BL21(DE3) Δ acrABEF as new bacterial expression strain

Initial attempts from original cell strain successfully isolated MmpL3s but MS analyses indicated that both samples were contaminated with the bacterial efflux pump, acriflavine resistance B or AcrB. The extent of contamination depended on the

construct expressed. Whilst this was a minor contaminant in the case of *Msmeg* MmpL3 FL, it was the main product of isolation of *Mtb* MmpL3 FL. Therefore, an alternative bacterial cell strain generated by Prof. Yamanaka (Hiroshima International University) carrying deletions for *acrA*, *acrB*, *acrE* and *acrF*, named BL21(DE3) Δ *acrABEF*, was used. Cells were first made chemically competent with a salt treatment that permeabilises the cell membrane and allows plasmid entry ²⁵⁶. In detail, after streaking cells onto an LB agar plate, a colony was picked and grown overnight in 3 mL LB media. The following day 100 mL LB were inoculated with 1 mL of the overnight culture and grown until OD₆₀₀ = 0.5. Culture was then aliquoted in sterile 50 mL tubes and incubated on ice for 10 minutes, before being centrifuged at 4,000 g for 10 minutes at 4 °C. Pellets were resuspended in 1/6th culture volume of CCMB 80 transformation buffer (10 mM KOAc, 80 mM CaCl₂*2H₂O, 20 mM MnCl₂*4H₂O, 10 mM MgCl₂*6H₂O, 10% glycerol, pH 6.4) and incubated on ice for 20 minutes, followed by another round of centrifugation. Finally, cell pellets were resuspended in 1/24th culture volume of the same transformation buffer and 100 μ L aliquots prepared in cold 1.5 mL centrifuge tubes, before snap-freeze and store them at -80 °C. Finally, competence was assessed by transforming 100 pg of control pUC19 DNA into 100 μ L cells. Following a 45-minute incubation at 37 °C cells were plated at 1:1, 1:10 and 1:100 dilutions, and number of colonies obtained confirmed cells as chemically competent. These cells were then transformed and used as expression strain for all experiments presented in this Chapter.

2.2.5 Cloning of *Msmeg* and *Mtb* MmpL3 Δ C-terminal constructs

Δ C-terminal constructs were generated by PCR amplification of the region of the genes encoding amino acid residues 1-773 for *Msmeg* MmpL3 (UniProtKB entry: A0QP27) and 1-753 for *Mtb* MmpL3 (UniProtKB entry: P9WJV5) and cloned into the pET22b/42a expression vector, carrying the C-terminal His₈-tag, using NdeI and XhoI as restriction enzymes. The primer pairs used for this purpose were: 5'-GAAGGAGATATACATATGTTTCGCCTGGTGGGGTTCGG-3' (FW *Msmeg* NdeI) and 5'-GTGGTGCTCGAGCCGCTGATCGGTCTCGGACTC-3' (RV *Msmeg* 773 XhoI) for the *Msmeg* MmpL3 C-terminally truncated construct; 5'-AAGGAGATATACATATGTTTGCATGGTGGGGTTCGTAC-3' (FW *Mtb* NdeI) and 5'-GTGGTGCTCGAGATCAGGCAGATGAATTTACCCAGGCC-3' (RV *Mtb* 753 XhoI)

for the *Mtb* MmpL3 Δ C construct. The resultant *Msmeg* Δ C construct lacked C-terminal 774-1013 residues, while the *Mtb* Δ C construct lacked amino acids 754-944. Both constructs were then chemically transformed into NEB[®] 10-beta cells first, and into BL21(DE3) Δ *acrABEF* once they had been sequence-verified.

2.2.6 Membrane scaffold protein 1 E3D1 (MSP1E3D1) expression

The pET-28a(+) plasmid carrying the gene for the His-tagged membrane scaffold protein MSP1D1E3 was transformed into *E. coli* BL21(DE3) strain and a glycerol stock was stored at -80 °C. When needed, the bacteria were streaked on LB-agar plates containing 50 μ g/mL kanamycin and incubated overnight at 37 °C. A single colony was inoculated into a 10 mL starter culture in LB supplemented with 50 μ g/mL kanamycin and incubated overnight (37 °C, 220 rpm), before being diluted in 10 x 1 L of terrific broth (TB) media supplemented with antibiotic. Cultures were incubated until the OD₆₀₀ reached 2.2-2.5 (37 °C, 220 rpm), before being induced by addition of 1 mM IPTG. After 4 hours incubation at the same temperature and speed, cells were harvested by centrifugation at 4,000 *g* for 10 minutes at 4 °C. The cell pellet was resuspended in lysis buffer (20 mM sodium phosphate pH 7.4, 1% v/v Triton X-100) supplemented with one EDTA-free protease inhibitor (SIGMAFAST[™], Sigma-Aldrich) and 50 μ g/mL DNaseI, and stored at -80 °C until further use.

2.2.7 Membrane scaffold protein 1 E3D1 (MSP1E3D1) purification

On the first day of purification, cell pellets were thawed before loading onto a Constant Systems cell disruptor twice at 26 kpsi. The lysate obtained was ultra-centrifuged at 4 °C, 40,000 *g* for 30 minutes. Supernatant was loaded onto a 5 mL His-Trap column (GE Healthcare) pre-equilibrated with 40 mM phosphate buffer, pH 7.4 and the flow-through was discarded. The column was washed with 5 CV of MSP wash buffer 1 (40 mM Tris-HCl pH 8.0, 300 mM NaCl, 1% v/v Triton X-100), 5 CV of MSP wash buffer 2 (40 mM Tris-HCl pH 8.0, 300 mM NaCl, 20 mM imidazole, 50 mM sodium cholate) and 5 CV of MSP wash buffer 3 (40 mM Tris-HCl pH 8.0, 300 mM NaCl, 50 mM imidazole). The protein was then eluted with 30 mL of MSP elution buffer (40 mM Tris-HCl pH 8.0, 300 mM NaCl, 400 mM imidazole). After measuring protein concentration using a NanoDrop[™] (+His MW = 32599, ϵ = 29910 M⁻¹cm⁻¹), Tobacco Etch Virus (TEV) protease was added to the eluate (80 μ g TEV : 1 mg MSP1E3D1) to

remove the His tag, and the mixture was dialysed (12 kDa MWCO) against dialysis buffer (20 mM Tris-HCl pH 7.4, 100 mM NaCl, 0.5 mM EDTA) for 16 hours at 4 °C.

On the second day, the dialysed sample was recovered and centrifuged for 5 minutes at 3,000 *g*, 4 °C before being filtered through a 0.2 µm low protein binding syringe filter. After adjusting imidazole concentration to 20 mM, the sample was then passed through a 5 mL His-Trap column (GE Healthcare) pre-equilibrated with post-dialysis equilibration buffer (40 mM Tris-HCl pH 7.4, 300 mM NaCl, 20 mM imidazole) and the flow-through was concentrated using 30 kDa MWCO centrifugal filters. Protein concentration was measured and sample concentrated to 12 – 15 mg/mL (-His, MW = 29981, $\epsilon = 26930 \text{ M}^{-1}\text{cm}^{-1}$). The isolated protein was aliquoted, flash frozen in liquid nitrogen and stored at -80 °C until use.

2.2.8 *Msmeg* and *Mtb* MmpL3 FL and ΔC reconstitution into nanodiscs

The concentrated MmpL3 sample after elution from the Ni²⁺-NTA column was mixed with MSP1E3D1 and detergent-solubilised 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC, from Avanti Lipids) lipid in a protein:MSPs:lipids of 1:10:1000, 1:6.6:660, 1:5:500 molar ratios, depending on the experiment. Following an overnight incubation at 4 °C with rocking, the mix was added to activated Biobeads SM2 (400 mg : 350 µg of protein) for detergent removal and incubated for 2 hours at 4 °C, with rocking. The nanodisc mix was transferred with the use of a 25G hypodermic needle (BD Microlance - 0.5x16mm) into a 100 kDa MWCO centrifugal filter and concentrated before loading it onto a Superdex® 200 Increase 10/300 gel filtration column pre-equilibrated with SEC buffer B (20 mM Tris-HCl pH 8.0, 150 mM NaCl). Fractions corresponding to main peak were then pooled and concentrated using 100 kDa MWCO centrifugal filters and either used fresh or stored at -80 °C for future experiments. MmpL3 DDM-based purification and NDs reconstitution processes are shown in Figure 2.2: once protein is in a detergent micelle, it can either be purified by gel filtration and concentrated (Figure 2.2, a), or it can be reconstituted into NDs (Figure 2.2, b).

In an attempt to reduce the empty ND population, as discussed in section 2.3.2.2, an optimised protocol was developed. This method aimed to differentiate between protein-NDs and empty NDs, taking advantage from the lack of His tag in the latter. After incubation with the adsorbent beads, the sample was run through a His-

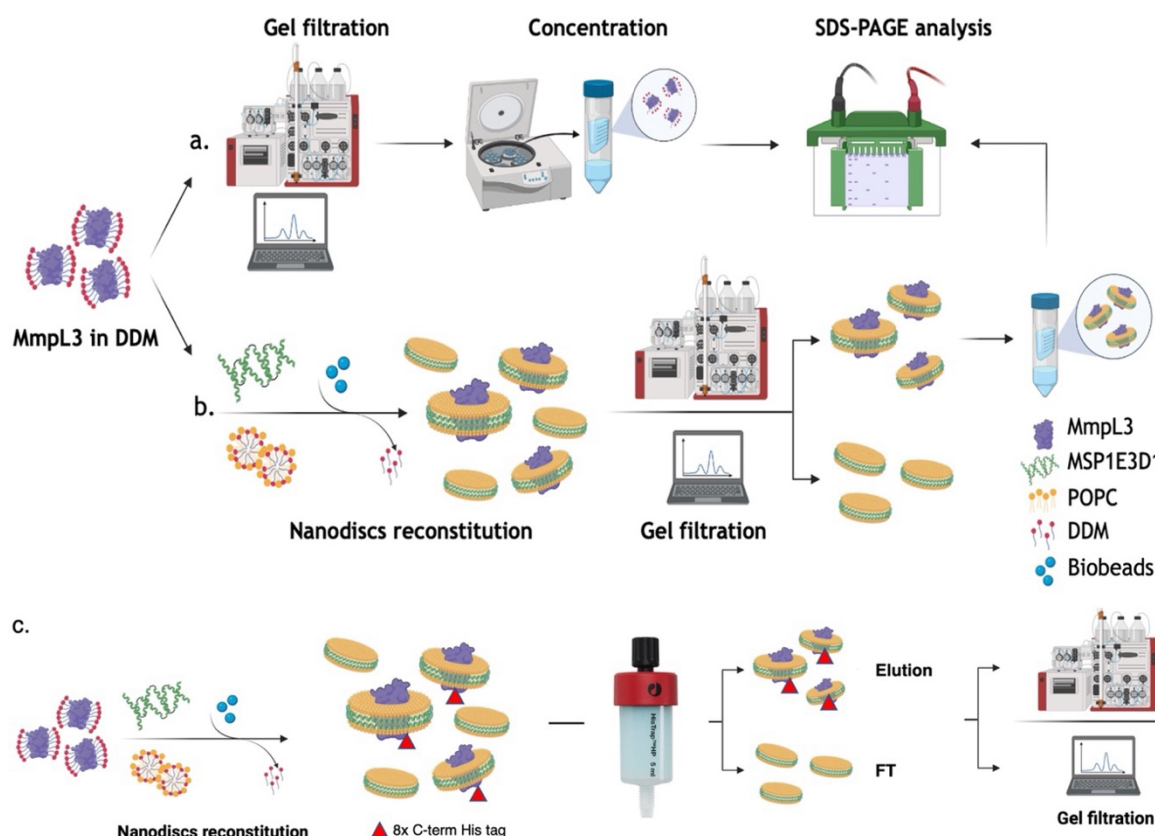


Figure 2.2. MmpL3 DDM-based purification and NDs reconstitution processes. First, MmpL3 was purified in DDM using a Ni^{2+} -NTA resin. Once in detergent, it either followed route a. and was further purified through a gel filtration step before being concentrated, or it followed route b. and was reconstituted into NDs. Better separation between empty- and MmpL3-NDs was achieved when route described in c was followed. Only MmpL3-NDs could bind to His-trap column as they carry a C-terminal His-tag and were collected in the elution step, while empty NDs were collected in the flow-through (FT). Figure was generated using BioRender.com.

trap column (HisTrap™ HP, Cytiva), pre-equilibrated with SEC buffer B supplemented with 10 mM imidazole. As a result of the absence of His tag, empty NDs could not bind to the Ni^{2+} -resin and were therefore collected in the flow-through (FT), that was subsequently concentrated using a 50 kDa MWCO centrifugal filter. The column, retaining the protein-NDs, was then washed with 5 CVs of SEC buffer B containing 30 mM imidazole. Finally, MmpL3-carrying NDs were eluted with 300 mM imidazole buffer and concentrated using a 100 kDa MWCO centrifugal filter. Both samples were then run on a Superdex® 200 Increase 10/300 GL gel filtration column (Cytiva) pre-equilibrated with SEC buffer B. Fractions corresponding to the main peak were pooled and concentrated using 100 kDa MWCO centrifugal filters and either used fresh or stored at -80 °C for future experiments (Figure 2.2, c).

2.2.9 *Msmeg* and *Mtb* MmpL3 FL and ΔC into SMALPs

For SMALPs preparation, membranes obtained as in section 2.2.2 were

resuspended in SEC buffer B and wet pellet weight diluted to 60 mg/mL. Membranes were then solubilised with gentle shaking for 2 h at room temperature in a buffer containing 20 mM Tris-HCl pH 8.0, 150 mM NaCl and 2.5% (w/v) SMA2000P (Cray Valley), kindly provided by Prof. Timothy Dafforn (University of Birmingham, UK). After centrifugation at 200,000 *g* for 1 h at 4 °C, supernatant containing the soluble fraction was incubated with Ni²⁺-NTA resin overnight 4 °C, with rocking. The sample was then loaded on a poly prep/glass econo-column chromatography column (Bio-Rad) and allowed to drain by gravity. The resin was initially washed with 20 CVs of imidazole-free buffer with the same composition as SEC buffer B, followed by 10 CVs washes with the same buffer, containing 10 mM imidazole and finally SMALP-MmpL3 was eluted washing the resin with 5 CVs of SEC buffer B, containing 500 mM imidazole. The eluted sample was concentrated using 100 kDa MWCO centrifugal filters to a final volume of ~500 µL prior to loading onto a Superdex® 200 Increase 10/300 gel filtration column pre-equilibrated with SEC buffer B. Fractions over the peak were then concentrated using 100 kDa MWCO centrifugal filters and either used fresh or stored at -80 °C for future experiments.

2.2.10 PAGE gels

For denaturing gels, protein fractions were mixed with SDS loading buffer in a 1:1 ratio and loaded on a NuPage™ 4 – 12% Bis-Tris PAGE Gel (Invitrogen™), using Novex™ Sharp Pre-stained Protein Standard as MW reference (range from 3.5 to 260 kDa) and separated at a constant voltage of 180 V for 40 minutes in NuPAGE MES SDS running buffer. For native gels, protein fractions were mixed with SDS-free loading buffer in a 3:1 ratio and loaded on a Novex™ 4 – 20% Tris-Glycine Gels using NativeMark™ Unstained Protein Standard (Invitrogen™) as MW reference (range from 20 to 1236 kDa) and separated at a constant voltage of 150 V in NativePAGE Running Buffer, for 1 h or until the dye front reached the bottom of the gel cassette. Denaturing or native gels were either Coomassie-stained and visualised using the GelDoc imaging system by Bio-Rad or rinsed in ultrapure water prior to being used for Western Blot analyses (see section 2.2.11).

2.2.11 Western Blot

Native PAGE or SDS PAGE gels were transferred to a nitrocellulose

membrane using an iBlot2 (20 V, 7 min; Invitrogen™). The membrane was incubated with blocking buffer (PBS, 0.2% Tween® 20 and 5% w/v skim milk) for 45 minutes at room temperature, with gentle shaking. Primary anti-His antibody (mouse monoclonal IgG2b, abm®), diluted to 1:3000 or 1:4000 according to the experiment, was then added and membrane was incubated for 1 h at room temperature. A horseradish peroxidase (HRP)-linked secondary antibody (anti-mIgG2, Cell Signaling Technology), diluted to 1:3000 or 1:4000 according to the experiment, was subsequently added after extensively washing the membrane with PBS containing 0.2% of Tween® 20. After performing three more wash cycles, the signal was developed using Clarity™ Western ECL substrates (Bio-Rad) following the kit manufacturer's recommendations and membrane was imaged using a GelDoc imaging system (Bio-Rad).

2.2.12 Assessment of protein stability

In order to test the stability of *Msmeg* and *Mtb* MmpL3 in a range of different detergents, thermal denaturation analysis using CPM dye (Invitrogen™) was performed. Specifically, 1 µL of 10 mg/mL detergent- purified protein in SEC buffer A was mixed with selected detergents at 3x critical micelle concentration (CMC) in a 96-well optical bottom plate to give a final volume of 147 µL. Detergents used for this assay were purchased from Anatrace and were: DDM, n-decyl-β-D-maltopyranoside (DM), lauryl maltose neopentyl glycol (LMNG), decyl maltose neopentyl glycol (DMNG) and lauryldimethylamine-N-oxide (LDAO). 1 µL of 7 mg/mL nanodisc-reconstituted protein or empty NDs in SEC buffer B was added to a 96-well optical bottom plate to give a final volume of 147 µL. 3 µL CPM dye at 40 µg/mL were added to each well in the dark. Fluorescence was measured using a SpectraMax plate reader (Molecular Devices) with λ excitation of 387 nm and λ emission of 463 nm every 5 minutes for 2 hours at 40 °C. CPM is nonfluorescent until it reacts with either thiol groups found in cysteines which become exposed as the protein unfolds or with expanding hydrophobic cores, as more recently described in the literature ^{230,257}. Therefore, CPM fluorescence signal directly correlates with protein unfolding. The percentage of folded protein (Percentage_{fold}) at each time point t was calculated using equation 2.1 (see next page).

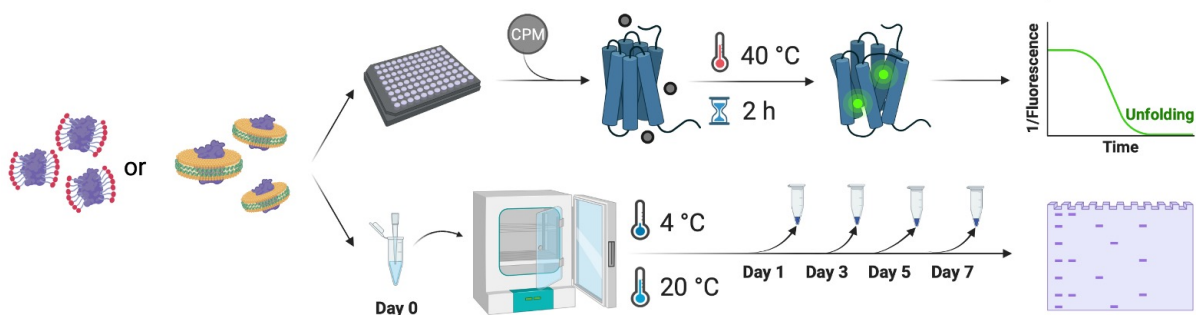


Figure 2.3. Protein stability assessment. Concentrated protein in detergent or ND-reconstituted was used to perform CPM-assay (top) and thermostability test (bottom).

Temperature- and time- dependent long term stability assessments by SDS-PAGE analysis were also carried out. Detergent- and nanodisc- purified MmpL3 at high concentration (e.g. 10 mg/mL for DDM samples) were incubated at 4 °C and 20 °C for 7 days. Protein samples were collected on day 1 and then on alternate days, a SDS PAGE gel was run following the procedure stated in section 2.1.7 for denaturing gels. The gel was then Coomassie-stained and imaged using a GelDoc imaging system (Bio-Rad). Both methods are illustrated in Figure 2.3.

$$Percentage_{fold} = \left[1 - \left(\frac{F_t - F_b}{F_{max}} \right) \right] \times 100$$

Equation 2.1. Equation to determine the relative percentage of folded protein over time in a thermal denaturation assay. For each of the measured time points, F_t is the fluorescence reading of the protein-containing wells, F_b represents the fluorescence of empty nanodiscs wells subtracted only for MmpL3-NDs signals, while F_{max} is the maximum fluorescence measured across all readings per tested construct.

2.2.13 Protein ID analysis through Mass Spectrometry (MS)

In order to confirm protein identity, different MW gel bands were excised and analysed by the BSRC Mass Spectrometry & Proteomics Facility at St Andrews University. The peptides resulting from digestion with trypsin were analysed by coupling liquid chromatography (LC) and electrospray ionisation MS (ESI-MS). The generated spectra were searched against NCBIprot, Swissprot or their internal database BMS using Mascot. Protein hits were ranked based on a protein score, reflecting the percentage of sequence coverage obtained with matching peptides. Each of these peptides in turn provide an expected ions score $-10\log(P)$, where P measures the probability for the observed match to be a random event, with scores >46 indicating identity or extensive homology. Furthermore, exponentially modified protein abundance index (emPAI) was used as an estimate of absolute protein amount as it takes into account the number of sequenced peptides per protein.

2.3 Results

2.3.1 *Msmeg* and *Mtb* MmpL3 purification in detergent

2.3.1.1 *Msmeg* MmpL3 full-length expression in *E. coli* BL21(DE3) WT

Initially, the plasmid carrying the full-length sequence of *Msmeg* MmpL3 was obtained and a chemical transformation into *E. coli* BL21(DE3) WT was performed in order to generate glycerol stocks. According to the methods published by Xu and colleagues, expression of MmpL3 was induced with 1 mM IPTG, with the cells harvested 2 h after induction ⁶⁶. In this work, the incubation period following IPTG induction was extended to 4 h to maximise cell growth and at the same time carry out expression within a single working day. A time-course expression was performed using the *Msmeg* MmpL3 Δ C expressed in *E. coli* BL21(DE3) Δ acrABEF cells (see section 2.2.4 and 2.2.5 for details on bacterial strain or on construct, respectively).

As shown in Figure 2.4, a small-scale culture (30 mL) was set up to test different time-points for harvesting cells post-IPTG induction. OD₆₀₀ values were also recorded at each time point to monitor cell growth. Cells collected at each time point were then lysed and the membranes isolated using the procedure described in section 2.2.2. Membranes were separated on an SDS-PAGE gel and an anti-His WB performed to assess protein band intensity and MW (Figure 2.4, a). As shown in the WB, MmpL3 overexpression was observed already 1 h post-induction, with similar expression level for all the time-points tested. However, a smeared signal was observed for the overnight incubation, suggesting that lengthening the expression might have detrimental effects on the quality of the final product. Nevertheless, bacterial growth was observed to increase over time although reaching a state with little growth rate if compared to uninduced cultures collected at the 4 h time-point, with the only exception being the overnight condition that presented the highest OD₆₀₀ value (Figure 2.4, b). However, given that the overnight incubation might impact sample quality, the 4 h time-point was confirmed as the preferred time-point to harvest cells at, allowing for high protein quality and possibly highest yields within a single working day.

The expression protocol was therefore finalised and used for all experiments described in this work.

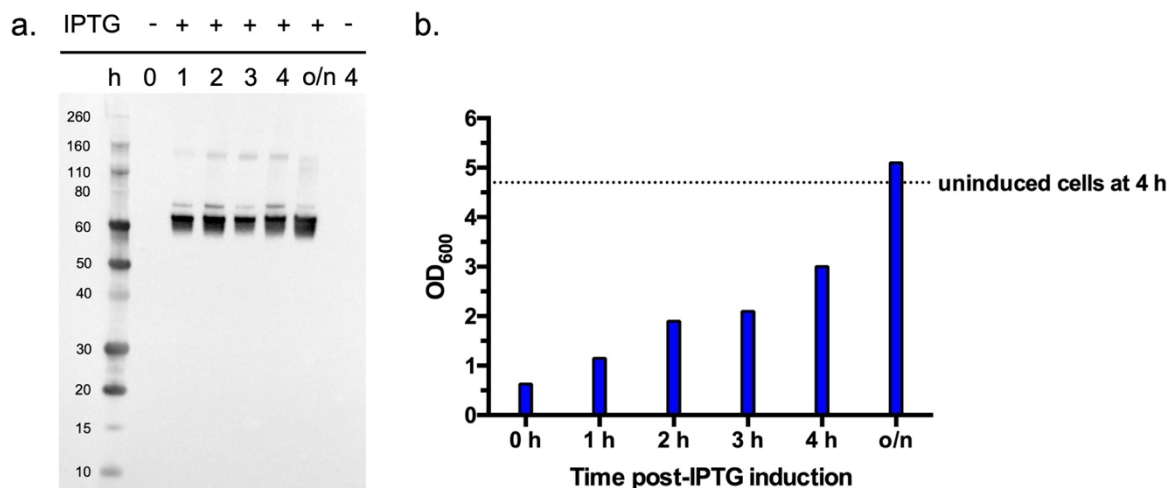


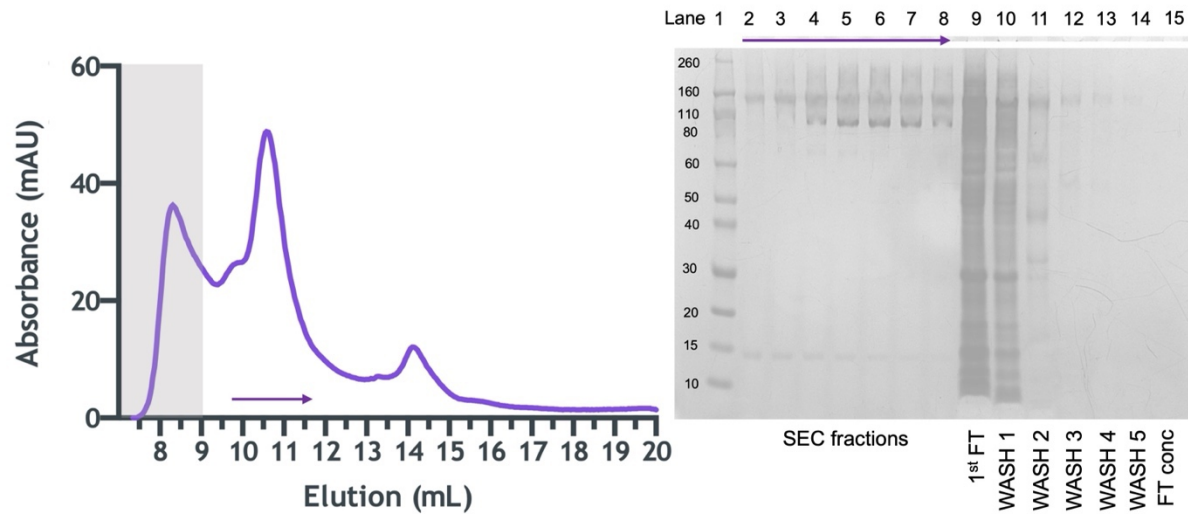
Figure 2.4. Analysis of *Msmeg* MmpL3 Δ C produced in the Δ acrABEF strain at different time points post induction. a. anti-His tag WB performed on isolated membrane samples collected prior to IPTG addition (0 h), and post-IPTG induction (1 h, 2 h, 3 h, 4 h, incubation at 37 °C and 220 rpm) as well as overnight (incubation at 20 °C and 180 rpm), following IPTG induction. While at 1 - 4 h post-IPTG induction two protein bands are present, at ~60 kDa (main band) and at ~140 kDa, in the overnight incubation (o/n) a smear signal is present for the higher MW band. No MmpL3 band is detected prior to induction (0 h) or after 4 h additional incubation in the absence of inducer. b. A plot of the OD₆₀₀ values recorded at each time point. Values between 0 and 4 h increased over time but growth rate was significantly lower than uninduced cells collected at the 4 h time-point. Overnight incubation gave the highest OD₆₀₀ value, at 5.1.

2.3.1.2 Optimisation of detergent-based purification

Once the expression protocol was established, the next step was to optimise the purification process. In the literature published up to that point, the main scientific groups working on MmpL3 all used DDM as the extraction detergent, therefore the same was used here ^{1,2,66}. However, the type of affinity resin used differed. While Xu and colleagues used a TALON resin, Zhang and co-workers opted for a Ni²⁺-NTA resin ^{1,66}. These are both metal affinity resins, but TALON contains cobalt ions, while Ni²⁺-NTA, as suggested by the name, is nickel-based.

Msmeg MmpL3 FL was used to optimise the purification process. First, TALON resin was tested and washed with 10 CVs of wash buffer containing 50 mM imidazole as reported by Xu and colleagues ⁶⁶. Using a medium-high concentration of imidazole is helpful to remove weakly resin-bound contaminants present in the sample, therefore achieving a much purer eluate. Following the TALON purification step, the protein was passed through a gel filtration column to increase product purity. However, in the SEC profile almost no protein was detected at the expected range of elution volumes, between 10 - 12 mL. Aliquots from the flow-through as well as from the washing steps were loaded onto a SDS-PAGE gel in an attempt to identify at which step the protein

a. 5x 20 mL washes at 50 mM imidazole concentration



b. 4x 25 mL washes at 30 mM imidazole concentration

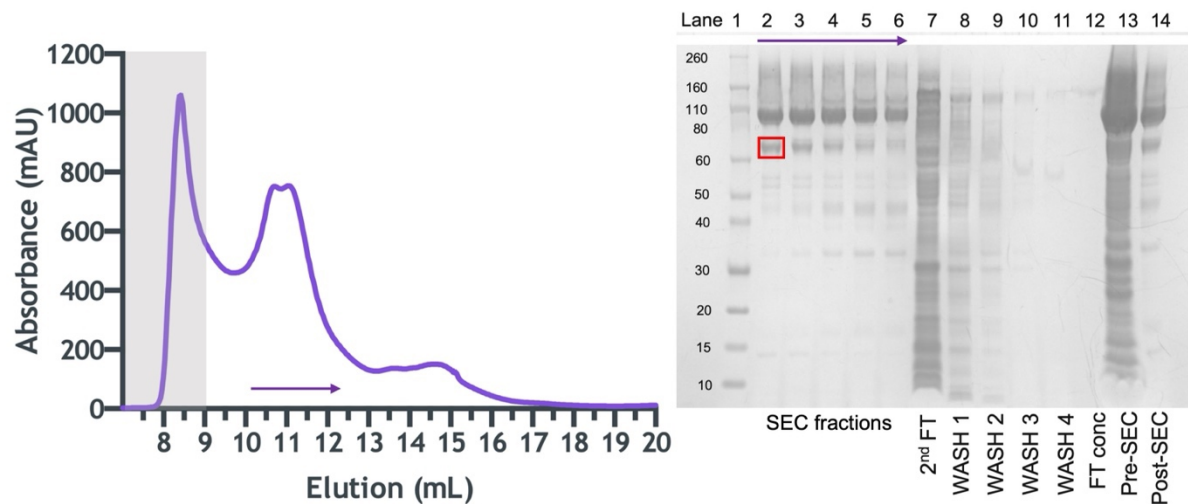


Figure 2.5. *Msmeg* MmpL3 FL purification using Ni^{2+} -NTA resin and varying imidazole concentration for column washes. a. 5x 20 mL washes at 50 mM imidazole concentration were used during purification. On the left, SEC profile reporting elution volume (mL) on x axis and absorbance at 280nm (UV) on the y axis for the first condition. Protein main peak is located between 10 - 12 mL elution volume, corresponding to fractions indicated by the purple arrow; on the right, SDS-PAGE gel with SEC fractions and aliquots of flow-through (FT), washes 1-5, FT from the concentrator. Two bands visible at ~ 110 kDa and 150 kDa. When pooled together protein yield was low, ~100 μg . b. 4x 25 mL washes at 30 mM imidazole concentration were used during purification. On the left, SEC profile with main protein peak between 10-12 mL with ~800 mAU as maximum. On the right, SDS-PAGE gel with samples from the purification as indicated. Two bands, at ~110 kDa and 150 kDa, were sent for MS sequencing and identified as MmpL3 as the top hit. In band in the red box band at 70 kDa was also sent for MS sequencing and was identified as a contaminant product from *E. coli* (ArnA protein). SEC void volume in grey area.

was lost. The Coomassie-stained gel revealed that the flow-through and the first wash steps were responsible for the highest percentage of protein loss, possibly suggesting that the binding between the His-tag and the cobalt resin was too weak. This might be due to the position of the His-tag at the end of the C-terminus (aa 773 - 1013), which has already been described as a disordered domain⁶⁹. Therefore, the interaction of

the tag with the resin might not be favourable, mainly because of the strict spatial requirements of the cobalt-containing reactive core (data now shown).

As an alternative, a nickel-based resin Ni^{2+} -NTA was tested, as used by Zhang and colleagues to carry out the purification of their hybrid construct composed of *Msmeg* MmpL3 (aa 1 - 748) fused to a T4 lysozyme and followed by a C-terminal His₁₀-tag ¹. For this experiment, the DDM-solubilised membrane sample was incubated with the nickel resin while all other conditions were kept the same as for TALON resin. As expected, the main peak corresponding to MmpL3 was detected between 10 and 12 mL of SEC elution volume and the corresponding fractions showed the presence of two bands on the gel, one just under 110 kDa and the other at around 150 kDa (Figure 2.5, a). However, although good progress had been made, the amount of end product (~100 µg) was insufficient for further analysis. As in the previous experiment the washing steps accounted for the majority of protein loss, possibly as a consequence of a weak binding between the resin and the His-tag. Several other conditions were tested in order to increase the yield, considering both the CV and the imidazole concentration used. Finally, a successful purification was achieved by using nickel-based resin and 10 CVs washes at 30 mM imidazole concentration (Figure 2.5, b). Roughly 1 mg of pure protein was obtained from a 3 L bacterial culture. When the fractions from the SEC were analysed by SDS-PAGE, the same two bands already identified in the previous experiment were detected, showing the same migration pattern. Bands at 110 kDa and 150 kDa were sent for MS sequencing and both were identified as *Msmeg* MmpL3 as the top hit. Peptides covering between 43 - 54% of the full-length protein were detected with a high confidence score for both samples, including the C-terminal domain and the His-tag, giving strong evidence that the full-length protein was successfully purified. A third band located at around 70 kDa was also sent for sequencing in order to investigate its identity. Although it cannot be excluded that this band is a degradation product of MmpL3, the analyses revealed that it is more likely to be a contamination product. *E. coli* protein ArnA, bifunctional polymyxin resistance protein, was detected as the top hit, with a 66% of sequence coverage, while only 28% of peptides matched MmpL3, possibly due to sample smearing from the high intensity band at 110 kDa.

At the time of these experiments, there was no mention in the literature of contaminating species present in the main MmpL3 band, although Xu and colleagues had used a different expression strain carrying a deletion for the bacterial efflux pump

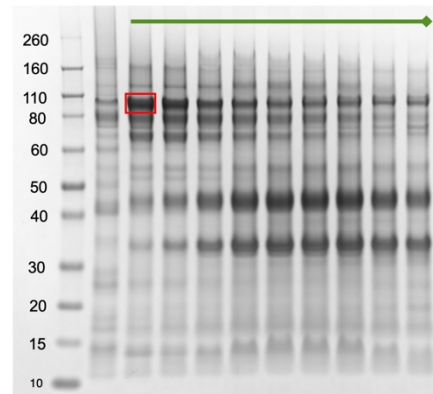
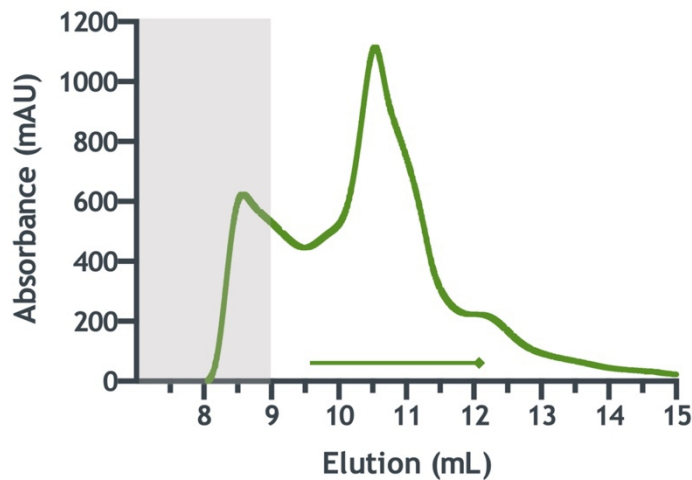
acriflavine resistance B (AcrB) protein. After carrying out a more detailed analysis of the MS products detected for the 110 kDa band, AcrB was indeed detected, with an abundance between 2 and 6 times lower than the *Msmeg* MmpL3 protein, identified as first hit. However, as described in the next section, the contamination significantly affected the purity of *Mtb* MmpL3 FL, therefore urging some troubleshooting.

2.3.1.3 End product contamination with bacterial protein

The optimised expression and purification protocols described in sections 2.2.2 and 2.2.3 were used to purify the *Mtb* MmpL3 FL construct. Unfortunately, the yields obtained with the *Mtb* protein were significantly lower compared to the *Msmeg* sample. Therefore, in order to increase the amount of end product the wash steps were performed at 10 mM imidazole concentration, giving a high UV signal during gel filtration (Figure 2.6, a). However, when fractions were analysed by SDS-PAGE, the sample was very heterogeneous containing several different protein bands. The main product band at ~110 kDa was excised and sent for MS sequencing and it was identified as AcrB with a high confidence score and abundance, with a 32 - 52% coverage, depending on the reference database used. AcrB has already been described as a major contaminant of purified His-tagged proteins whenever *E. coli* is used as the expression system. The protein is a trimer belonging to the same RND transporter family as MmpL3, and each monomer contains at the C-terminus 2 His residues that are in close proximity within the trimer, resulting in a natural His tag²⁵⁸. Thus, WT AcrB has high affinity for Ni²⁺-immobilised resin. This bacterial protein represents a real problem, especially for X-ray crystallography as it can form stable crystals from very low concentrations (<0.1 mg/mL down to picograms) and even when representing less than 5% of the final purified product^{259,260}. Furthermore, electron microscopy (EM) experiments are potentially vulnerable to the same problem, as small amounts of contaminating AcrB can impact protein identification²⁵⁸.

As decreasing the imidazole concentration significantly impacted the purity of the sample, purification was carried out by including extensive washing with higher imidazole concentrations, and at the same time doubling the starting material of what was used in previous experiments (Figure 2.6, b). This decision was also justified because a low bacterial growth was observed in terms of pellet size and relative OD₆₀₀. Although a slight increase in protein yield was achieved, and contaminant

a. 10 mM imidazole wash



Band identified as AcrB by MS sequencing

b. 30 mM imidazole wash

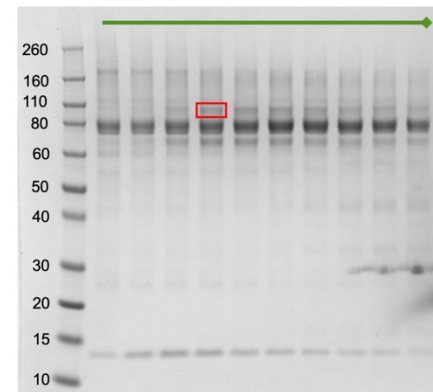
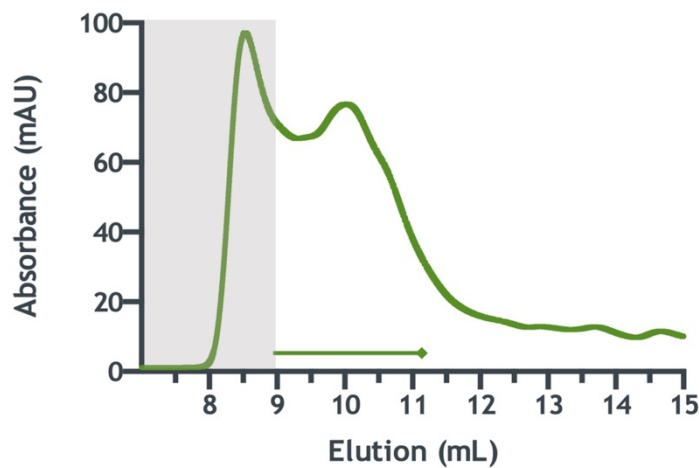


Figure 2.6. *Mtb* MmpL3 FL purification using Ni^{2+} -NTA resin and varying imidazole concentration for column washes. a. 10 mM imidazole washes were used. On the left, SEC profile reporting elution volume (mL) on the x axis and absorbance at 280nm (UV) on the y axis. The main protein peak is located between 10-12 mL elution volume, corresponding to fractions indicated with green arrow; on the right, SDS-PAGE gel with SEC fractions. Multiple protein species were present with band at ~110 kDa (red box) as main product and identified as AcrB by MS sequencing. b. 30 mM imidazole washes were used during purification. On the left, SEC profile with main protein peak between 10-12 mL with ~80 mAU as maximum. On the right, SDS-PAGE gel with SEC fractions. Two bands at 80 kDa and 170 kDa were sent for sequencing and gave *Mtb* MmpL3 as first hit. Although at a lower intensity compared to what observed in a, band at ~110 kDa (red box) was identified as AcrB by MS sequencing. Therefore, only a partial reduction in contamination concentration was achieved but not its complete removal. SEC void volume is indicated in the grey area.

concentration was indeed reduced, it became obvious that any downstream application results would be questioned because a complete removal of AcrB could not be achieved (Figure 2.6, b). For this reason, an alternative bacterial cell strain was selected.

2.3.1.4 Expression in *E. coli* BL21(DE3) Δ acrABEF

In many bacterial species the efflux pumps belonging to the RND protein family form part of tripartite systems; in particular, in *E. coli* the inner membrane protein (e.g. AcrB or AcrF) interacts with the outer membrane protein channel (e.g. TolC) through an accessory periplasmic protein (e.g. AcrA or AcrE) to form the complete efflux pump. In order to avoid AcrB protein contamination, a specific bacterial strain generated by Prof. Yamanaka (Hiroshima International University) was obtained ²⁶¹. The strain carries deletions of the *acrA*, *acrB*, *acrE* and *acrF* genes and will be referred to as BL21(DE3) Δ acrABEF going forward. Following the procedure stated in section 2.2.3, chemically competent cells were created and transformed with the plasmid DNA encoding for *Msmeg* and *Mtb* MmpL3. This bacterial strain was then used as the expression system for the production of protein used to generate all further experimental results presented in this work.

Both FL constructs were successfully expressed and purified following the optimised protocol described in section 2.2.3. As shown in Figure 2.7, SDS-PAGE analysis indicated that MmpL3 was the main product for each construct. Protein bands corresponding to monomers were detected at ~80 and ~110 kDa for *Mtb* and *Msmeg*, respectively. Moreover, bands migrating at twice the MWs values mentioned above were also observed. MS analyses were performed on *Msmeg* proteins bands migrating at ~160 and ~220 kDa. Sequencing results revealed: 1) MmpL3 as the first hit with high confidence score and abundance; 2) the absence of AcrB contamination, therefore giving high confidence in the end product quality. Furthermore, both bands were identified as MmpL3, potentially suggesting the existence of a dimeric protein state.

2.3.1.5 Generation of C-terminally truncated constructs

As mentioned in the previous section, C-terminally truncated constructs were generated alongside the FL version of both *Msmeg* and *Mtb* proteins. The main reason behind this decision was the ability to directly compare results to the published literature. Zhang and colleagues carried out structural studies on the truncated version of *Msmeg* MmpL3, where the residues 1 – 748 were fused to a T4 lysozyme (160 aa), followed by a His₁₀ tag at the C-terminus ¹. In their work, Su *et al.* generated a *Msmeg* MmpL3 construct carrying residues 1 – 773 in frame with a C-terminal His₆ tag ². The

truncated sequence of *Mtb* MmpL3 encoding for aa 1 – 753 followed by a TEV cleavable C-terminal GFP-His₈ affinity-tag was instead used by Adams and co-workers to solve the cryo-EM structure of the protein in LMNG ⁴. The *Msmeg* MmpL3 Δ C construct lacking aa 774 – 1013 (UniProtKB entry: A0QP27) was generated as described in the paper published by Su and colleagues in 2019 ². The *Mtb* Δ C construct lacking aa 754 – 944 (UniProtKB entry: P9WJV5) was based on the work of Adams *et al.* ⁴. It had already been shown that the soluble C-terminal domain is not essential for bacterial growth nor for MmpL3's functional activity ⁶⁹. Nevertheless, its presence is known to negatively affect protein stability and diffraction quality of the crystals, given its unstructured and highly flexible nature. Thus, all structural studies have been carried out with engineered versions of the protein, lacking the C-terminus to improve resolution without affecting protein folding ^{1,2,4}.

For this work, *Msmeg* and *Mtb* MmpL3, both Δ C and FL constructs, were purified in different membrane-mimetic systems in order to compare protein behaviour. This was done to exclude any bias that could arise from either the presence or lack of the C-terminal domain, as well as improve protein yield and stability. A significant increase in yield was observed for the purified *Msmeg* MmpL3 Δ C construct compared to the FL. When detergent-purified proteins were injected onto the SEC column, two broad multi-peaks were observed, with a ~3x higher A₂₈₀ signal observed for *Msmeg* MmpL3 Δ C (9.5 – 11.5 mL elution volume) compared to that for the FL protein (9 – 11 mL). Furthermore, the FL construct presented a higher level of aggregation (peak between 8 – 9 mL) than the Δ C construct. The final yield obtained after concentrating SEC fractions was ~3 mg for the *Msmeg* MmpL3 Δ C and ~1 mg for the FL (Figure 2.7, c and d).

In the case of the *Mtb* construct, truncation also improved the final yield, although both constructs exhibited relatively low expression (Figure 2.7, d). For both constructs, doubling the starting culture volume from 6 L to 12 L, yielded only 400 μ g of protein for FL and 900 μ g for the Δ C construct. Marked aggregation was also observed, as shown by the peaks corresponding to the column void volume in the SEC traces, between 8 and 9 mL (Figure 2.7, a and b). Despite these issues encountered with both *Mtb* constructs, there was sufficient isolated material for all four protein constructs to perform some downstream analyses. However, for some work I had to limit the analysis to the *Msmeg* proteins.

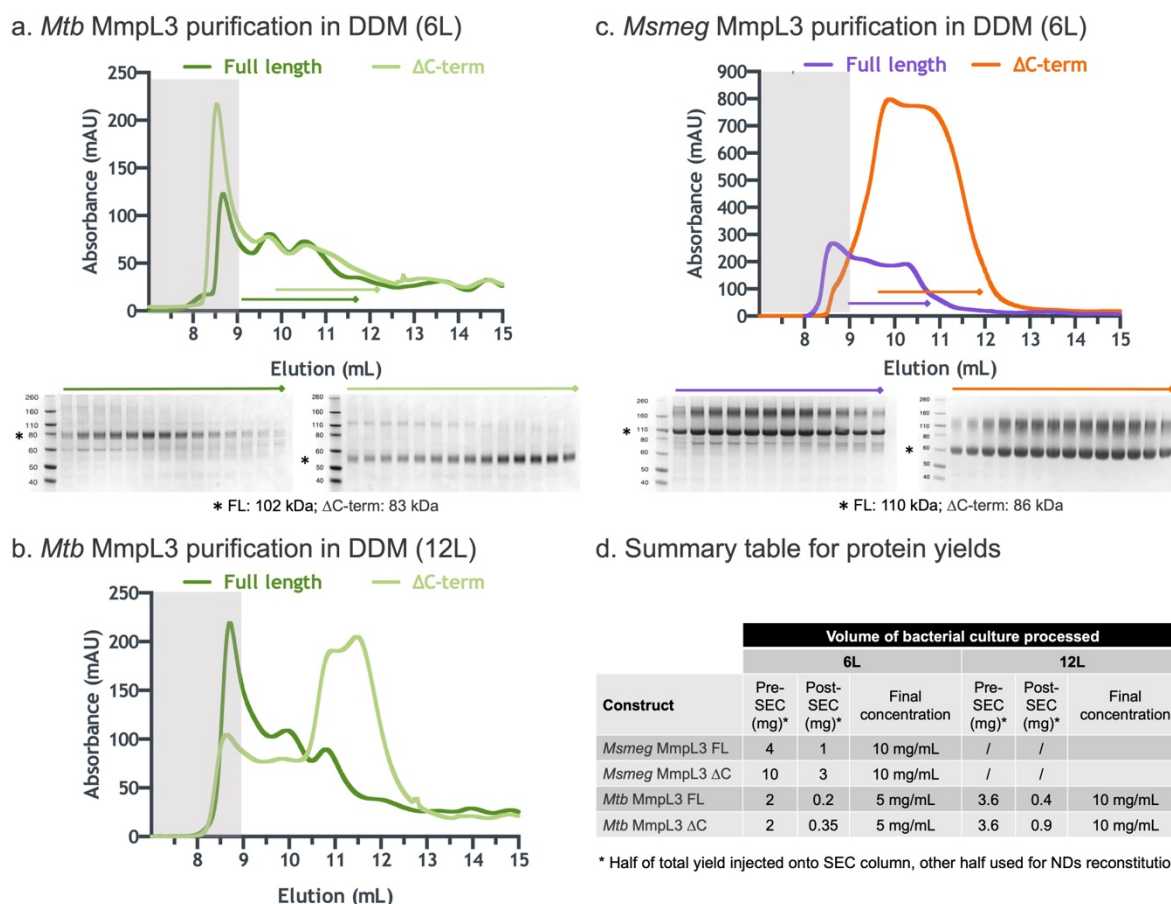


Figure 2.7. *Mtb* and *Msmeg* MmpL3 FL and ΔC-term purification in DDM and relative yields. a. Top, SEC profile reporting elution volume (mL) on x axis and absorbance at 280nm (UV) on the y axis for *Mtb* MmpL3 FL (dark green) and ΔC (light green) constructs obtained from 6 L cultures. Below, SDS-PAGE gels for protein fractions corresponding to the regions highlighted by the arrows in SEC trace, with stars indicating MWs for monomeric *Mtb* MmpL3s (FL: 102 kDa; ΔC-term: 83 kDa). b. SEC profile of *Mtb* MmpL3 FL (dark green) and ΔC (light green) constructs obtained from 12 L of cultures. c. SEC profile for *Msmeg* MmpL3 FL (purple) and ΔC (orange) constructs obtained from 6L cultures. Below, SDS-PAGE gels for protein fractions corresponding to arrows in SEC trace, with stars indicating MWs for monomeric *Msmeg* MmpL3s (FL: 110 kDa; ΔC-term: 86 kDa). d. Summary table for protein yields obtained after DDM based purification of all constructs. Pre-SEC (mg) column indicates milligrams injected onto the SEC column, corresponding to half of the total mgs obtained after elution from Ni²⁺-NTA resin. The other half was used for ND reconstitution. Post-SEC (mg) column indicates the final protein amount, obtained after SEC fractions were pooled together and concentrated down to the value indicated in 'Final concentration' column.

2.3.2 *Msmeg* and *Mtb* MmpL3 reconstitution into nanodiscs

2.3.2.1 Standard nanodiscs reconstitution process

In an attempt to prepare MmpL3 in a more physiological membrane environment, the 4 constructs were initially extracted and isolated with the use of detergent and subsequently individually reconstituted into MSP nanodiscs (NDs), following the procedure illustrated in Figure 2.2, b²⁵⁴. The size of each disc can be

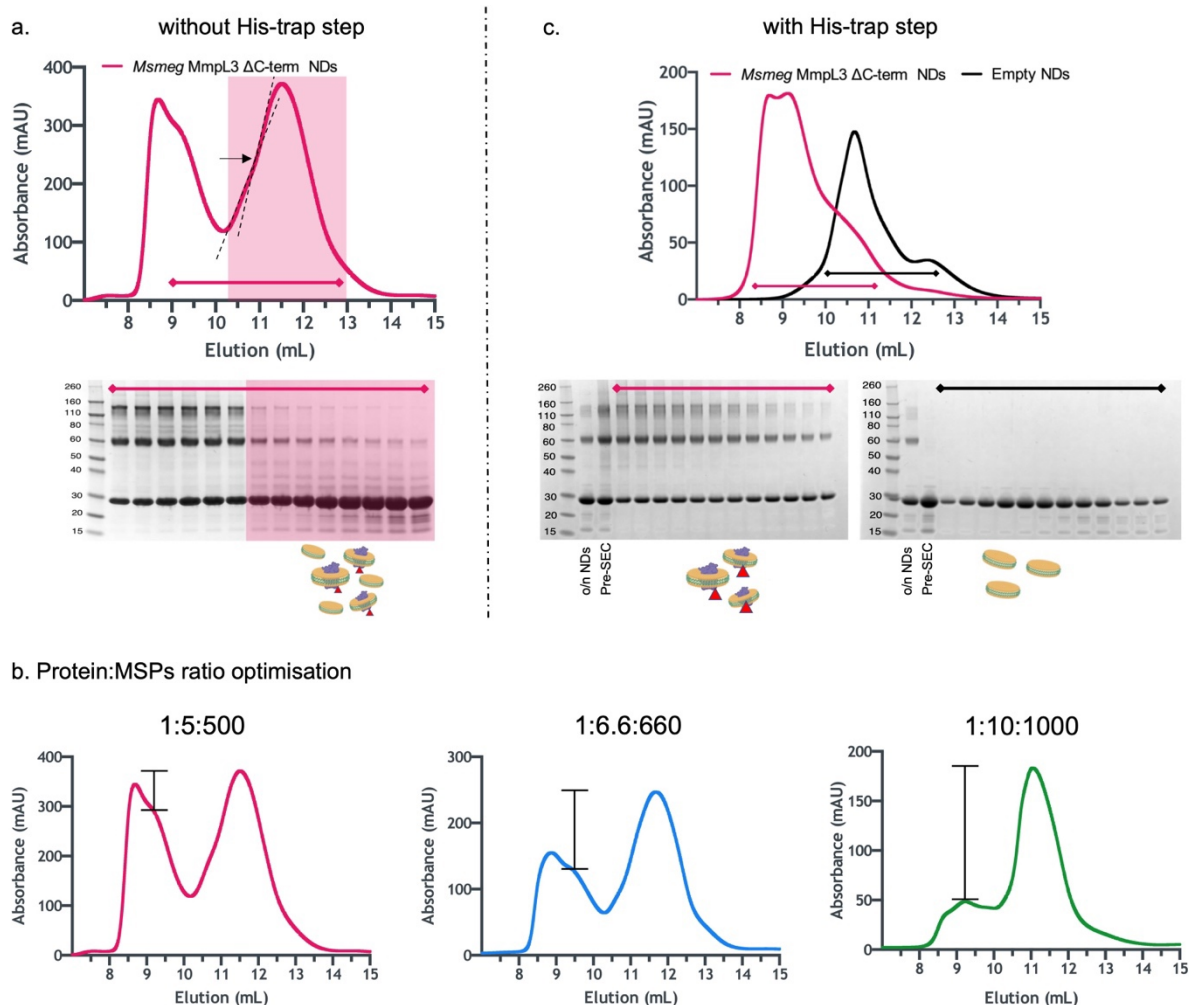


Figure 2.8. MmpL3 NDs reconstitution. a. Standard protocol without His-trap step used for *Msmeg* MmpL3 ΔC reconstitution into POPC NDs. Top, SEC trace presenting two main peaks (9 – 10.5 mL and 10.5 – 13 mL, in pink shaded area). Black arrow indicates change in slope at ~11 mL, where dashed lines meet. Fractions run on SDS-PAGE gel revealed a mixture of MmpL3-NDs and empty NDs in the 10.5 – 13 mL elution range. b. SEC profiles obtained using different protein:MSPs ratios with standard protocol, from 1:5:500 (magenta) to 1:6.6:660 (blue) up to 1:10:1000 (green). Signal for second peak, between 10.5 and 13 mL, increased in intensity compared to first peak (9 – 10.5 mL) going from lowest to highest ratio used, meaning that a higher proportion of empty NDs was present in final product. c. Optimised protocol for NDs reconstitution included an extra His-trap step which separated empty NDs in flow-through (FT) and MmpL3-NDs in elution. Different species are separated in the gel filtration step, with a partial overlap between monodispersed MmpL3-NDs (magenta) and empty NDs (black) between 10.5 – 13 mL elution volume. SDS-PAGE gel confirmed lack of protein in the empty NDs fractions, while MmpL3-NDs fractions presented presence of monomeric and dimeric species, keeping the right protein:MSP proportion.

controlled by selecting the appropriate membrane scaffold protein, or MSP. A range of engineered versions of the amphipathic protein, human apolipoprotein A-1, have been generated, meaning that the particle's diameter can be adjusted according to the target protein^{141,254}. In this study, MSP1E3D1 was used in combination with POPC lipid molecules, to generate particles homogeneous in size, with a diameter of ~12.9 nm. As an example, SEC analysis of *Msmeg* MmpL3 ΔC in NDs is shown in Figure

2.8, a. All constructs presented similar SEC traces, with at least two main populations observed. The slope of the second peak changed at around 11 mL elution volume, suggesting the presence of a mixed population. In fact, when fractions were run on an SDS-PAGE gel, only the fractions up to 10.5 mL elution volume showed the expected ratio of MmpL3 and MSPs (1:2). Band intensity for the later fractions (between 10.5 mL and 13 mL), instead, appeared to be higher for MSPs, indicating the presence of smaller and free discs as the most abundant species, while MmpL3-NDs represent only a minor fraction (Figure 2.8, a).

Although successful reconstitution of MmpL3 in POPC NDs was achieved for all constructs, some optimisation was carried out in an attempt to reduce the amount of free NDs in favour of MmpL3-NDs. Different protein:MSPs:lipids ratios were tested as shown in Figure 2.8, b. Generally, using a molar ratio of 1:10:1000 yielded a peak eluting between 10.5 mL and 13 mL (containing free nanodiscs as main species) that was at least 3x higher compared to the earlier peak eluting between 9 and 10 mL (MmpL3-NDs). However, the two peaks (containing free and reconstituted nanodiscs, respectively) showed the same height when a ratio of 1:5:500 was used (Figure 2.8, b). Therefore, by reducing the protein:MSP ratio while keeping the MSP:lipid ratio the same, the empty ND population was also reduced, with the best results obtained with a 1:5 protein:MSPs molar ratio. Nevertheless, a complete removal of free nanodiscs could not be achieved using this experimental procedure.

2.3.2.2 Optimisation of nanodiscs reconstitution process

As the presence of empty NDs in the final product could represent an issue for downstream applications, a new protocol was implemented, as illustrated in Figure 2.2, c. Following sample incubation with biobeads, NDs were loaded onto a His-trap column. During this step empty NDs, that lack the His-tag, did not bind to the Ni²⁺-based resin and were collected in the flow-through and concentrated. After performing extensive column washing, MmpL3-NDs were eluted from the column and concentrated before loading them on the SEC column. By using this method, it was possible to separate empty POPC NDs from the protein-carrying ones, significantly reducing the chances of cross-contamination and increasing sample purity (Figure 2.8, c). At the same time, two main populations were identified for the MmpL3-NDs, corresponding to two broad peaks on the SEC traces. When fractions were run on a Coomassie-stained SDS-PAGE gel, the MWs observed were in line with the ones of

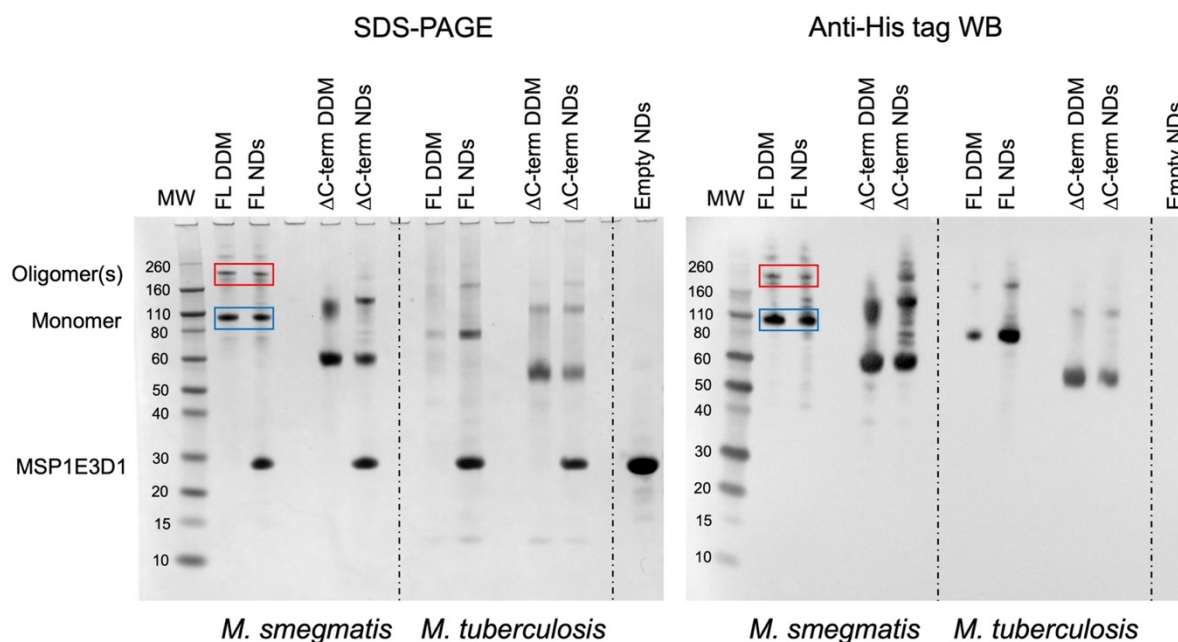


Figure 2.9. SDS-PAGE gel (left) and anti-His WB (right) of samples obtained during DDM-based purification and NDs reconstitution of FL and Δ C MmpL3 from *Msmeg* and *Mtb*. MW = Molecular weight markers. Individual purified proteins in either DDM or nanodiscs (NDs) are indicated above the relevant lanes. The blue and red boxes indicate the likely monomeric and oligomeric forms of the full-length version of MmpL3 from *Msmeg*, respectively. A similar pattern of two bands is seen for all the other protein constructs. MSP1E3D1 band is located at ~30kDa and it is not detected by anti-His WB.

the detergent-extracted proteins, including the dimeric species. Protein identity was also confirmed by WB, where no signal was detected for the scaffold protein MSP1E3D1, lacking the His tag (Figure 2.9). Nevertheless, if fractions from the two peaks were concentrated individually, a different MWs distribution was observed. Early-eluting fractions showed a higher proportion of the larger MW species, possibly indicating oligomeric arrangements, while the smaller species predominantly found in late-eluting fractions were monomeric protein. These observations, together with further findings discussed in section 2.3.3, guided the decision to further investigate the oligomerisation state of MmpL3 using experimental and computational approaches.

2.3.3 Protein stability assessment: MmpL3 in detergent micelles and NDs

2.3.3.1 Thermostability assay

DDM-purified and NDs samples were used for follow-up experiments to assess thermostability. For each of the 4 constructs, freshly purified material at a concentration of 10 mg/mL for detergent-extracted and ~7 mg/mL for NDs, was incubated for 7 days at 4 °C and 20 °C and aliquots were collected at day 1 and at

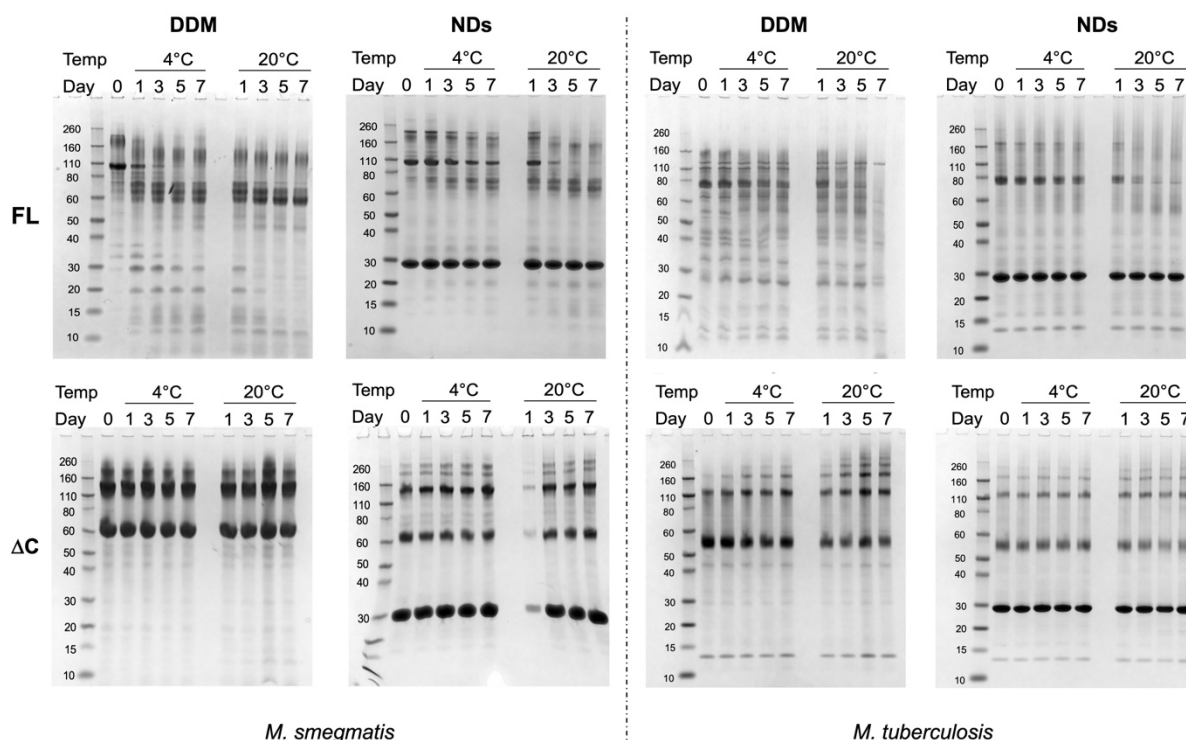


Figure 2.10. Thermostability test on *Msmeg* and *Mtb* MmpL3. Left panel, results for *Msmeg* MmpL3 FL (top row) and ΔC construct (bottom row) for DDM-purified and NDs-reconstituted protein. Right panel, results for *Mtb* MmpL3 FL (top row) and ΔC (bottom row) for both DDM-purified and NDs-reconstituted protein.

each alternate day and then all samples were submitted to SDS-PAGE analysis. As shown in Figure 2.10, the *Msmeg* MmpL3 FL in DDM appeared to be quite unstable. This was observed at 4 °C, with the main protein band at ~110 kDa showing lower signal intensity on day 1 compared to day 0, and an associated appearance of lower molecular weight bands in the 10 - 30 kDa range and between 60 - 80 kDa, suggesting a high level of degradation. Moreover, the same degradation products were observed when the protein was incubated at the higher temperature already at day 1, suggesting that in order to keep the protein stable it would be preferable to carry out experiments using freshly purified material and to keep the temperature at 4 °C or lower. However, when NDs carrying *Msmeg* MmpL3 FL were tested, the protein appeared to be stable at 4 °C throughout the whole test duration, while stability improved, although only slightly, when the sample was incubated at 20 °C, with the main protein band at 110 kDa starting to disappear from day 3 with concomitant appearance of the lower MW species. Therefore, it can be concluded that reconstitution into NDs for this construct had a beneficial effect on protein stability at the lower temperature. When *Msmeg* MmpL3 ΔC in DDM was tested, the protein appeared to be quite stable at both temperatures and throughout the whole duration of the experiment supporting earlier

studies on this protein. This high level of stability was also observed in the NDs sample, confirming the results obtained with the DDM-purified protein (Figure 2.10, left panel).

Regarding the *Mtb* samples, DDM-extracted FL construct appeared to be quite stable at 4 °C, while it started showing degradation after 1 day incubation at 20 °C (Figure 2.10, right panel). The same trend was observed when *Mtb* MmpL3 FL NDs were tested, with high stability observed at 4 °C during the whole week, but only for 1 day when incubated at 20 °C. When *Mtb* MmpL3 ΔC in DDM was tested, instead, degradation did not occur at both temperatures. However, aggregation products were present as higher molecular species appeared starting from day 3 at 4 °C and 20 °C, with bands increasing in MW size and intensity for the higher temperature condition. These observations suggested that in order for this protein to be stable, experiments needed to be carried out at 4 °C and preferably with 1-3 days from purification. Results obtained with *Mtb* MmpL3 ΔC in NDs, instead, showed a high level of stability at both temperatures throughout the whole duration of the assay, confirming once again that in general NDs preparations improved protein thermostability compared to the DDM-purified counterpart, especially for the FL constructs (Figure 2.10, right panel).

2.3.3.2 Thermal denaturation analysis

Using a thermal denaturation analysis, we also assessed 1) protein stability in various detergents and 2) further compared detergent-purified samples to NDs preparation. The protein isolated in DDM was diluted into a range of other detergents at 3x their CMC and stability tested in a thermal denaturation assay using the thiol-reactive dye CPM. The unfolding of the protein was calculated over time as inversely proportional to the CPM fluorescence signal. DDM, DM, LMNG and DMNG are all non-ionic sugar derivatives, while LDAO is a zwitterionic amine-oxide detergent. Non-ionic detergents are generally preferred for membrane protein solubilisation and manipulation, as in contrast to zwitterionic surfactants, they are non-denaturing and do not affect a protein's functional state. As shown in Figure 2.11 (left panel) in the assay performed on *Msmeg* MmpL3 FL, although LDAO did not cause the highest level of protein unfolding overall, it produced the curve with the steepest descent, meaning that MmpL3 immediately started to unfold, showing a sudden ~10% reduction of folding at the first time-point compared to the other detergents. The percentage of protein folded started to plateau at around 30% after 20 minutes, possibly more as a

consequence of increased protein aggregation that prevented dye binding rather than as a real stability reading. Proceeding from the least to the most stable, LDAO was followed by DM, DDM, DMNG and finally LMNG, keeping the highest percentage of protein folded for longer. The same trend was also observed for the *Msmeg* MmpL3 Δ C, although the protein stayed folded for longer compared to FL construct and in the most stabilising detergents DMNG and LMNG, retained >30% folded material after 2 h at 40 °C.

These results seem to contradict the previous choice of detergent and suggest the use of either LMNG or DMNG to achieve higher protein stability. However, these detergents form very stable micelles and are extremely hard to exchange, therefore limiting some of the downstream applications which will be discussed later in this thesis. In light of these considerations, DDM was used as the solubilisation agent of choice for the purification of membrane protein MmpL3.

When detergent-purified samples were compared to NDs, the latter showed stability only second to the best performing detergent, LMNG, for both FL and Δ C constructs. It is worth noting that when empty NDs were tested individually, a gradually increasing CPM fluorescence was detected, meaning that these particles undergo some level of unfolding. Although the MSP protein does not contain cysteines in its sequence, it cannot be excluded that incubating empty NDs for 2 h at 40 °C could affect their stability. Under these conditions some hydrophobic residues, usually interacting with lipids, can become exposed allowing the dye to bind. Therefore, the empty NDs signal was subtracted from MmpL3-NDs samples, to only report the specific unfolding behaviour of protein-carrying discs. From the results obtained with NDs samples, the Δ C constructs proved to be more stable than FL constructs, for both *Msmeg* and *Mtb* MmpL3 (Figure 2.11, right panel).

Together, these findings confirmed what was observed in the thermostability test, with NDs carrying Δ C constructs showing a higher stability than FL ones.

2.3.4 *Msmeg* MmpL3 into SMALPs

SMALPs differ from the synthetic nanodiscs described so far as these are lipid particles formed by extracting the protein with associated native membrane lipids, so-called native nanodiscs. SMA therefore acts like a detergent, but the extracted particles retain the host's original lipid composition (see section 1.3.4, Chapter 1). To

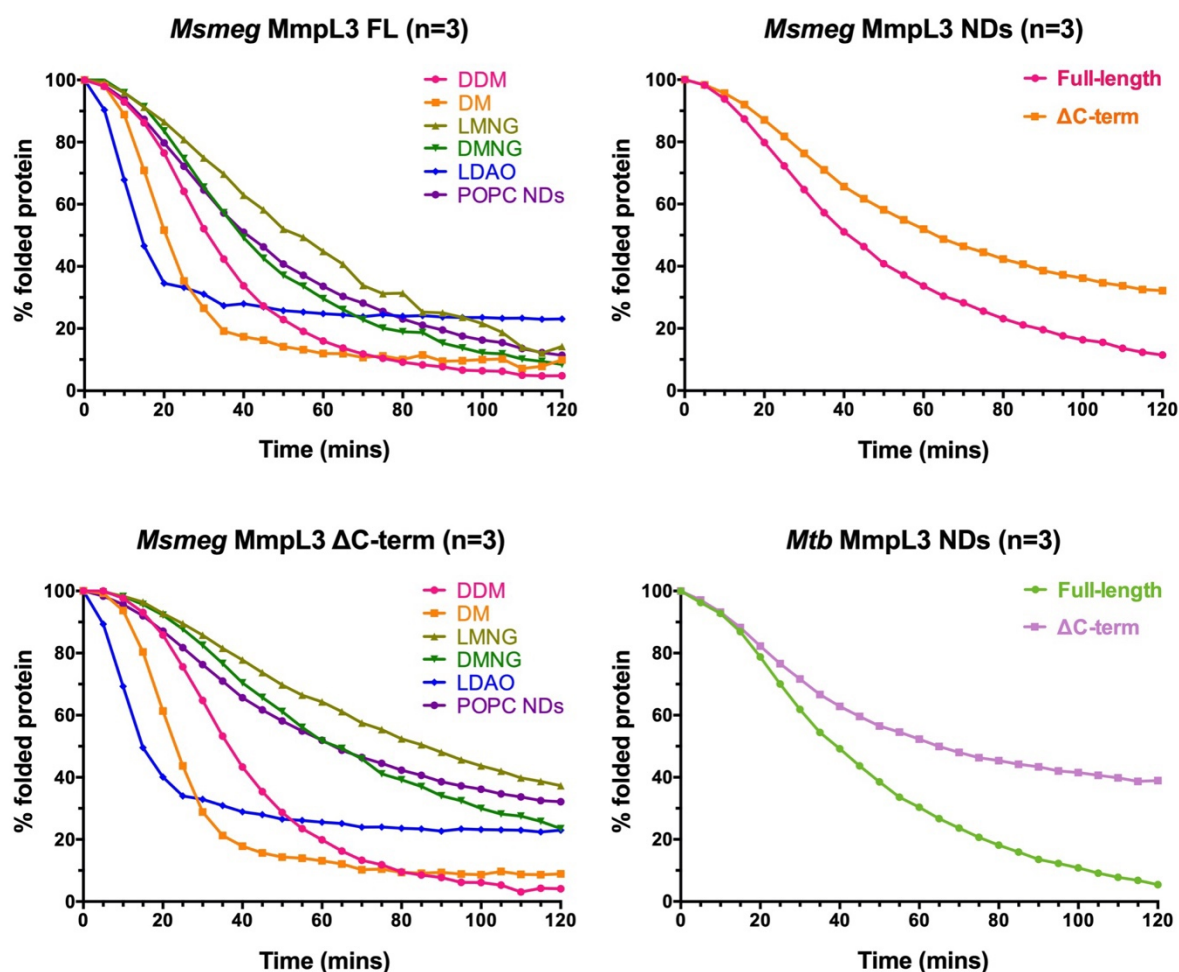


Figure 2.11. CPM assay results. Left panel for average results for *Msmeg* MmpL3 FL and ΔC in detergents and NDs. Right panel for *Msmeg* and *Mtb* MmpL3 into NDs. Empty NDs signal was subtracted from MmpL3-NDs wells. All conditions were tested in three different experiments (n=3).

carry out experiments with SMALPs, a collaboration with Professor Timothy Dafforn from University of Birmingham was established. For the purpose of this thesis, only results for *Msmeg* MmpL3 reconstituted in native nanodiscs are shown. As can be observed from the SEC trace, two overlapping peaks were detected for both FL and ΔC species (Figure 2.12, a and b). When individual fractions from both peaks were submitted to SDS-PAGE analysis, the main band corresponding to MmpL3 migrated at the expected MW (60 kDa for ΔC construct and 100 kDa for FL one), meaning that the protein had been successfully reconstituted. Furthermore, exploiting SMA's net negative charge individual fractions were run on a native gel, without the use of a reducing agent, to investigate the native MWs of the two populations from the SEC. The ΔC construct early-eluting fractions (9 – 11 mL elution volume) showed a MW of ~170 kDa while a smeary band of ~80 kDa was observed for the later ones. A similar

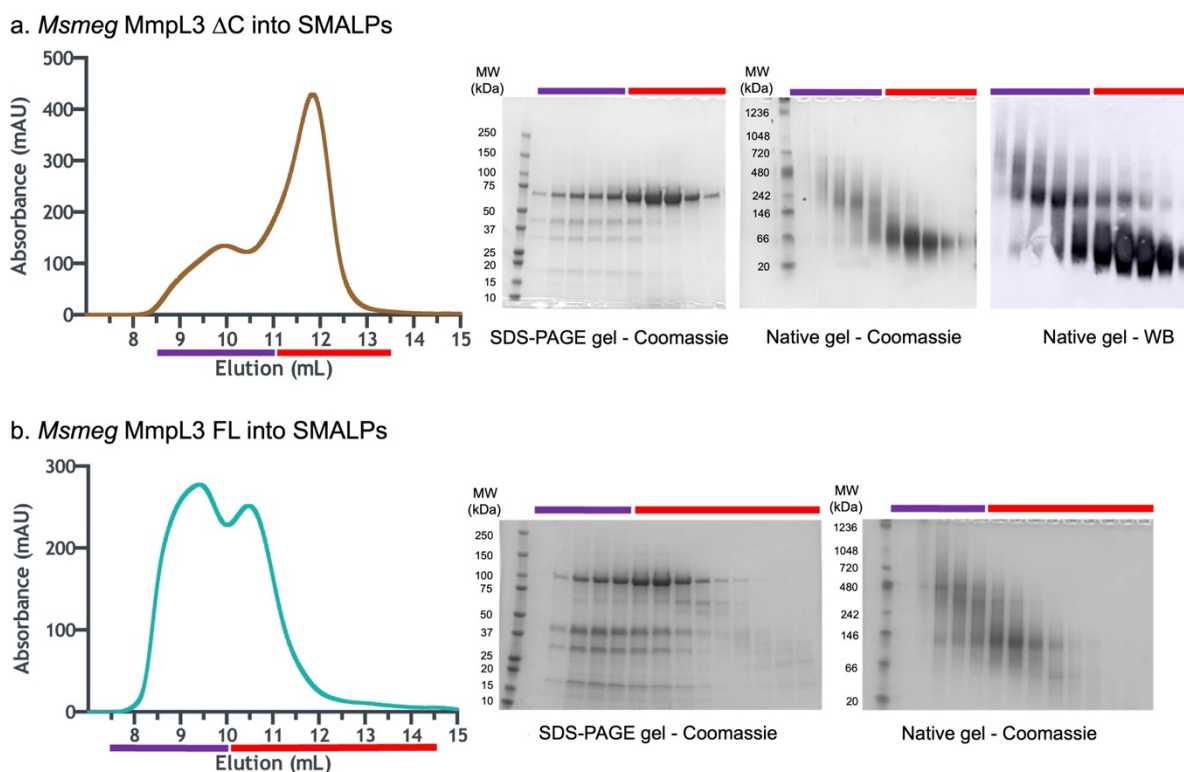


Figure 2.12. *Msmeg* MmpL3 FL and Δ C into SMALPs. a. Results for the Δ C construct: from left to right, SEC profile, Coomassie-stained SDS-PAGE and native gels for early-eluted fractions (purple) and late-fractions (red). On the right anti-His WB on native gel performed on same fractions. b. Results for FL construct: from left to right, SEC profile, Coomassie-stained SDS-PAGE and native gels for early-eluted fractions (purple) and late-fractions (red).

trend was also seen for the FL construct, although the separation of the two populations was less clear cut. For the Δ C construct, protein identity was also confirmed by anti-His WB analysis (Figure 2.12, a).

Results obtained with SMALPs represented further proof of the existence of multiple MmpL3 populations, likely to represent different oligomeric states, after having observed the same pattern using a detergent-based and a synthetic NDs approach (Figure 2.9).

2.4 Discussion

In this Chapter, MmpL3 expression and purification was optimised, starting from several published methods ^{1,2,66}. BL21(DE3) is the most common host *E. coli* strain used in the bacteriophage T7 promoter expression system, where the λ -lysogen DE3 regulates the production of T7 RNA polymerase, whose expression is under the control of the IPTG-inducible lacUV5 promoter. A protocol for bacterial overexpression in this strain was slightly modified in terms of incubation time post-IPTG induction, in

order to achieve the maximum expression within a single working day. Regarding the detergent-based purification protocol, the choice of affinity resin and imidazole concentration for wash steps were also optimised to increase quality and yield of the isolated material.

Specifically, cobalt- and nickel-based resins were tested. Cobalt-based resins are highly selective for His-tagged proteins as the cobalt ions are positioned at the core of a spatially restricted pocket where one metal-ion only interacts with histidine residues when they are adjacent. Moreover, each positively charged cobalt ion remains tightly bound to the resin pocket, which contains three carboxyl groups and one nitrogen atom, therefore reducing metal-ion leaching. In the nickel-based resins, such as for Ni^{2+} -NTA, the spatial requirements for binding to histidine residues are less strict, meaning that these resins could also bind His residues found naturally within proteins, decreasing the binding specificity and potentially the purity of the eluate. Furthermore, the nickel ions are less tightly bound in the resin pocket as two coordination complexes can be formed, one of which has a planar structure, resulting in a weaker interaction and allowing for metal-ion leaching from the resin.

When both resins were tested for the purification of *Msmeg* MmpL3 FL in DDM, results indicated a significant product loss during wash steps when cobalt-based resin was used, achieving very low protein yields. This could be the consequence of a weak interaction established between resin and protein C-terminal His tag, possibly due to the aforementioned strict spatial requirements for binding. In *Msmeg* MmpL3 FL the tag could be partially buried or less accessible, significantly reducing the amount of protein interacting with free resin. When Ni^{2+} -NTA resin was used, and the imidazole concentration for wash steps were adjusted, yields improved, and the protocol was finalised. However, initial attempts from the original cell strain successfully isolated MmpL3s but MS analyses indicated that both samples were contaminated with the bacterial multidrug efflux pump, AcrB. The extent of contamination depended on the construct expressed. Whilst this was a minor contaminant in the case of *Msmeg* MmpL3 FL, it was the main product of isolation of *Mtb* MmpL3 FL, especially when the imidazole concentration used during wash steps was also reduced. Therefore, although protein yields improved, the use Ni^{2+} -NTA resin increased the concentration of contaminants present in the eluate, as a consequence of the less strict criteria of binding to His residues.

Further expression was therefore carried out in a knockout bacterial strain lacking the *acr* genes, effectively eliminating the issue of contamination with AcrB, likely to interfere with our downstream analyses.

When detergent-based purification was repeated for the *Msmeg* MmpL3 FL construct and protein stability was tested in a thermostability assay, it was evident that the ΔC version underwent less degradation when incubated for a week at 4°C and 20°C, as already suggested in the literature ^{1,69,2}. The same distinction could be drawn for the FL and ΔC versions of *Mtb* MmpL3. It is worth mentioning, however, that the *Mtb* MmpL3 constructs always showed low yields for both versions, limiting the analyses that were performed on these constructs.

DDM was initially selected as detergent of choice as it is one of the most commonly used for the purification of membrane proteins and because it had been used in three reference papers for MmpL3 purification ^{1,2,66}. However, protein stability in different detergents was assessed using CPM assay, with molecules selected based on their usage in the purification of membrane proteins but also on their main properties, such as micelle size and CMC. Using DDM and DM as examples, the latter has a hydrophobic tail 2 carbons shorter than DDM that reduces the micelle size (MW ~40 kDa against 72 kDa for DDM) but at the same time increases the CMC of ~10 folds (0.087% or 1.8 mM for DM against ~0.0087% or 0.17 mM observed for DDM) ⁹⁷. LMNG and DMNG differ from the other maltose derivatives, DDM and DM, as they each contain two maltopyranoside headgroups as well as two alkyl chains of different lengths (dioctylpropane for DMNG and didecylpropane for LMNG), thus resulting in a much lower CMC than DDM / DM and a more stable micelle. The results obtained in the CPM assay confirmed the five detergents' expected ranking in conferring protein stability. Although CMC values follow the scale LMNG < DMNG < DDM < LDAO < DM, protein denaturation and aggregation is usually associated with LDAO usage, since this is a relatively harsh detergent. According to these results, LMNG was the best performing detergent. However, its highly stable micelles can impact downstream applications, especially if detergent exchange is needed as in the case of MSP-based nanodiscs. Therefore, DDM was confirmed as the preferred choice for all purifications, representing the right compromise between protein stability and its downstream use. Furthermore, in accordance with published data, stability analyses in detergent revealed that the truncated proteins showed a lower level of unfolding when heated.

In this Chapter, it was proved possible to extract MmpL3 from the membrane both in DDM and as SMALPs, as well as reconstitute the DDM isolated material into MSP based nanodiscs. Intriguingly in all cases, monomeric and dimeric species were identified, as observed by SDS-PAGE and WB analyses. This was supported by a relatively broad gel filtration peak, with two main protein populations, and suggests that oligomer formation is not dependent on the C-terminal domain.

In particular, this behaviour was initially seen in DDM-purified samples. In order to investigate protein behaviour in a more physiological environment which is not provided in detergent micelles, synthetic POPC NDs carrying MmpL3 were generated. MSP1E3D1 was used to form the belt of the disc, differently from the cNW9 scaffold protein used by Su and colleagues ³. The two mentioned MSPs differ in length and therefore form discs that are different in diameter, around 9 nm for cNW9 and ~13 nm for MSP1E3D1. The choice of MSP is dictated by the size of the target protein, also considering its multimeric state. In their paper, Su *et al.* generated cNW9 NDs and isolated particles by cryo-EM carrying the monomeric form of MmpL3 ³. However, cNW9 NDs would be too small to accommodate a dimer, as calculated by the Membrane Builder tool on the CHARMM-GUI web-based platform ^{262,263}.

Moreover, to retain correct protein folding and conformation, an adequately sized lipid bilayer area is required to not incur into membrane distortions ¹⁴¹. POPC was selected for its cylindrical shape, as it self-assembles in bilayers with no curvature when added to an aqueous solution and can therefore simulate a membrane patch. The resulting bilayer thickness is ~5.5 nm, considering that each POPC molecule measures 28.3 Å in length ²⁶⁴. In this work, MmpL3-NDs showed a significant increase in protein stability compared to the DDM-purified samples, both in a thermal and a CPM-based assays, also confirming the pattern seen for detergent-purified samples in that ΔC constructs were more stable than FL ones. Moreover, two distinct bands were identified for each construct reconstituted into MSP-based POPC NDs, when samples were analysed by SDS-PAGE and WB, in agreement with results obtained for DDM-purified proteins.

However, as the starting material in the MSP-based NDs is represented by the DDM-purified protein, there could be some bias in the results obtained. Specifically, if protein artificially forms oligomeric species when purified in detergent, the same higher MW species could be retained once the detergent is exchanged and NDs are formed.

Furthermore, given the addition of SDS, and the heating that occurs during electrophoresis, it was not possible to rule out that these higher molecular weight bands were artefacts of the gel running process.

Therefore, MmpL3 was also extracted directly from the membrane using SMA to form SMALPs to compare protein behaviour in three different systems, the latter being a more physiological representation of the protein native environment. The main component of these lipid particles is a polymer that presents hydrophilic maleic acid groups and hydrophobic styrene groups, which spontaneously rearrange into a disc-like shape when in close proximity to lipid membranes. The target protein is solubilised by using SMA chains that encapsulate the membrane section where it is located, therefore avoiding any stress the protein undergoes in a detergent-based extraction and maintaining it in a more physiologically relevant state. SMA functions as a detergent at membrane-level, meaning that the purification process needs fewer steps compared to the reconstitution into MSP-based nanodiscs. More importantly, the native lipid composition is retained during extraction. The diameter of each SMALP is around 10 - 12 nm and the MW of the SMA component is estimated to be between 7.5 - 10 kDa. Furthermore, polymers are found to establish an optimal interaction with lipid membranes at a pH ranging between 7 and 8, as a balance between electrostatic interactions and hydrophobic effect can be achieved, therefore maximising membrane solubilisation ¹⁷⁴.

To date, there is no publication available on SMALPs-MmpL3, therefore constituting a novel approach for the biophysical characterisation of this membrane protein. When *Msmeg* MmpL3 was successfully isolated into these native nanodiscs, two main protein populations were identified following gel filtration. Furthermore, the presence of distinct species was also confirmed through native gel and WB analyses. The MWs observed reasonably corresponded to a monomeric and a dimeric form of MmpL3, as already observed in DDM-purified and NDs-reconstituted proteins. To a certain extent, the importance of these results was higher compared to previous ones, as they were obtained using a more gentle and physiological purification process which should better preserve protein native environment and oligomeric state.

These results paved the way for further investigations of MmpL3 oligomeric state by carrying out a thorough biophysical characterisation of the two distinct protein populations identified using several complementary methodologies, as described in Chapter 3.

3. Biophysical investigation on MmpL3 oligomeric state

3.1 Introduction

As described in Chapter 2, MmpL3 was observed to form oligomeric arrangements during SDS-PAGE analysis, not only when isolated in DDM but also following reconstitution into MSP-based nanodiscs. Further evidence of the existence of MmpL3 dimers was obtained when MmpL3 was reconstituted into SMALPs. Two distinct protein populations were observed on native PAGE for early- and late-eluting SEC fractions, with same trend detected across the *Msmeg* and *Mtb* constructs. Following up from the aforementioned results, MmpL3 oligomeric state was further investigated, and results will be presented in this Chapter.

Several experimental techniques have been used by the scientific community for exploring the physiological oligomeric arrangement of membrane proteins. Among these, SEC multi-angle light scattering (SEC-MALS) is commonly used to determine the mass of membrane proteins and their higher order structures in detergent solution. However, it does require chromatographic peaks of protein species to be well resolved for achieving mass accuracy. Within separating techniques, blue native PAGE is an electrophoresis approach used to study proteins in their native state, without using denaturing agents that alter the biological composition. For structural analysis, cryo-electron microscopy (Cryo-EM) is a technique enabling direct visualisation of isolated biomolecular complexes following vitrification of the sample in a thin and amorphous ice layer^{74,208,265}.

In this Chapter, other biophysical techniques, such as sedimentation velocity analytical ultracentrifugation (SV-AUC), mass photometry, native MS (nMS), as well as the already mentioned cryo-EM were used to better characterise the two distinct populations observed for the *Msmeg* MmpL3 ΔC protein^{74,175,218,266–268}. Together, the evidence collected strongly suggests that MmpL3 can indeed exist as a dimer.

3.2 Materials and Methods

3.2.1 Sedimentation velocity analytical ultracentrifugation (SV-AUC)

DDM-solubilised protein, NDs and SMALPs samples were generated following

experimental procedures described in Chapter 2, sections 2.2.3, 2.2.8 and 2.2.9, respectively. Samples were prepared to a concentration of 0.1 - 0.5 mg/mL in 20 mM Tris-HCl, 150 mM NaCl, pH 8.0 buffer. Sedimentation velocity (SV) experiments were performed by Pooja Sridhar at the Birmingham Biophysical Characterisation Facility, University of Birmingham. Experiments were performed at 20 °C in a Beckman Coulter XL-I Analytical Ultracentrifuge, using a Ti 50 rotor. Samples were centrifuged for 16 h at either 81,000 *g* or 144,000 *g*, depending on the predicted sample MW to allow for full sedimentation, and were monitored by absorbance at 280 nm. SEDNTERP was used to calculate protein partial specific volume, solvent density and viscosity²⁶⁹. Data was then analysed applying a continuous *c(s)* distribution using the Lamm equation in the program SEDFIT²⁷⁰. During the analyses the frictional ratio parameter was allowed to float and the best fit absorbance *c(s)* distributions of sedimented species were plotted using GUSI²⁷¹.

3.2.2 Nanodiscs separation through sucrose gradient ultracentrifugation

MmpL3-NDs were generated using the optimised protocol as described in Chapter 2, section 2.2.8 by using a 1:5:250 molar ratio of protein to MSPs to POPC lipids. Empty NDs were separated from protein-carrying NDs and both samples were purified by SEC, on a Superose® 6 Increase 10/300 GL column (Cytiva). Post-SEC concentrated pooled fractions for *Msmeg* MmpL3 Δ C and empty NDs samples in 200 μ L of SEC Buffer B (20 mM Tris-HCl, pH 8.0 and 150 mM NaCl) were layered on top of 4 mL linear 10 – 30% sucrose gradients prepared in the same buffer. Sucrose gradient centrifugations were conducted at 204,000 *g* in a SW60 Ti rotor (Beckman Coulter) for 18 h at 4 °C with slow acceleration and no brake. Twenty-eight fractions of 150 μ L were manually collected and alternate aliquots were analysed by native and SDS–PAGE gels (4 – 12% Bis-Tris) and either Coomassie-stained or used for WB analysis.

3.2.3 Mass photometry on *Msmeg* MmpL3 Δ C and empty NDs

Mass photometry experiments were conducted using a Refeyn TwoMP mass photometer (Refeyn Ltd., Oxford, UK). MassGlass UC sample carrier slides and six-well sample cassettes (Refeyn Ltd.) were used for sample loading. In each measurement, 19 μ L of pre-filtered buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl) were placed in the well for focusing, followed by adding and mixing 1 μ L of the sample to

achieve a final dilution of 10 nM in a 20 μ L droplet. Movies were recorded for 60 seconds under standard settings ^{272–275}. The mass photometry measurements were calibrated using β -amylase, corresponding to molecular weights of 56 kDa (monomer), 112 kDa (dimer), and 224 kDa (tetramer).

Data acquisition was performed at room temperature using the AcquireMP software package. Data were processed using the DiscoverMP software package (Refeyn Ltd.). Peaks for each mass species were manually identified and fitted with a Gaussian function. Subsequently, all histograms were plotted using 10 kDa bin widths using DiscoverMP software ^{272,273,275}.

3.2.4 Native MS

Laser induced liquid bead ion desorption (LILBID) is a soft mass spectrometry ionisation method, which allows the investigation of non-covalently bound complexes including their stoichiometry under almost native conditions ²²¹. All LILBID experiments were carried out by Alicia Just, PhD student in the group of Prof. Nina Morgner (Goethe University, Frankfurt). Prior to analysis by LILBID-MS all analyte samples were buffer-exchanged into 20 mM Tris-HCl, pH 8.0 using Zeba Micro Spin desalting columns (Thermo Scientific, USA, 7 kDa MWCO). For each measurement 4 μ L of buffer exchanged and degassed sample were used. LILBID involves soft transfer of the analyte in solution to the gas phase. Microdroplets of 50 μ m diameter of the analyte were produced by a piezo-driven droplet generator (MD-K-130, Microdrop Technologies GmbH, Norderstedt, Germany) at a frequency of 10 Hz. The microdroplets were transferred to a high vacuum and irradiated by a pulsed IR laser operating at the vibrational absorption wavelength of water (2.8 μ m). The laser energy, adjusted to 10 – 23 mJ, is absorbed by the water molecules, resulting in an explosive expansion of the microdroplets and a release of the analyte ions to the gas phase. The analyte ions were accelerated by a pulsed electric field and analysed with a homebuilt time-of-flight mass spectrometer. LILBID settings have previously been published in detail ^{217,221}.

Ion detection was performed in negative mode and data acquisition was done using the homebuilt software *Massign* based on LabView ²⁷⁶. Each mass spectrum shown is the result of an averaged signal of 1,000 microdroplets. Data was calibrated based on measurements of an aqueous 10 μ M bovine serum albumin (BSA) solution, smoothed and background corrected.

3.2.5 Cryo-EM data collection and processing

Cryo-EM experiments were carried out with Joseph D. Barritt, postdoctoral researcher in Dr Sarah Rouse's group (Imperial College London). For the *Msmeg* MmpL3 Δ C ND sample, a 3 μ L of 0.6 mg/ml aliquot were applied to glow-discharged (Quantifoil Cu R2/2, 300 mesh) holey carbon grids, blotted for 2x 5 seconds at 4 °C and 100% humidity. After 30 seconds wait, they were plunge-frozen in liquid ethane using a Vitrobot mark IV (Thermo Fisher, Waltham, Massachusetts) before transfer into cartridges. Cryo-EM data was collected in the Electron Microscopy Centre at Imperial College London, equipped with a 200 kV Glacios™ 2 Cryo Transmission Electron Microscope (Cryo-TEM) (Thermo Scientific™) with a Selectris energy filter. Cryo-EM images were recorded at -0.7 to 2.5 μ m defocus with 2.7 mm spherical aberration and nominal 79 K magnification, corresponding to a 1.5 Å/pixel sampling interval. The total specimen dose for each EER fractionated movie was 40 e⁻/Å².

The micrographs of MmpL3 were aligned using a patch-based motion correction for beam-induced motion using cryoSPARC ²⁷⁷. The contrast transfer function (CTF) parameters of the micrographs were determined using Patch CTF ²⁷⁷. After discarding poor micrographs, with low estimated CTF resolution and thick or crystalline ice, MmpL3 particles were automatically picked using a CHARM-GUI model of dimer in 13 nm NDs as a template. Initially, 1,180,857 particles were selected from 8,946 micrographs after template autopicking in cryoSPARC, using an extract bin x 2 particles for faster processing. The images were subjected to several rounds of two-dimensional (2D) classification into 200 classes, optimising window function to obtain the best protein alignment. False picks and classes with unclear features were removed and resulted in 249,442 particles that were used to generate a reference free *ab initio* three-dimensional (3D) reconstruction with C1 symmetry.

3.2.6 Lipidomics

Each protein sample was subjected to a lipid extraction as described by the Folch method ²⁷⁸. Briefly, LC-MS grade water was added to each sample to a final volume of 400 μ L. 1 mL of 2:1 chloroform : methanol was then added, and sample was vigorously mixed before a 13,800 g centrifugation step for 10 minutes. The organic layer (bottom) was carefully removed and transferred into a LC-MS glass vial, before adding 10 μ L of Internal Standards (IS) mix at 100 μ g/mL. IS for Cardiolipin (14:0-CDL

ammonium salt, #710332), Phosphatidylethanolamine (15:0-18:1-d7-PE, #791638), Phosphatidylcholine (15:0-18:1-d7-PC, #791637), Phosphatidylserine (15:0-18:1-d7-PS, #791639), Phosphatidylglycerol (15:0-18:1-d7-PG, #791640) and Sphingomyelin (18:1-d9 SM, #791649) were purchased from Avanti Lipids in 1 mg/mL chloroform solution. All standards were mixed together, and LC-MS grade methanol was added to reach a final concentration for each standard of 100 µg/mL. Each sample was then dried under a nitrogen stream and reconstituted in 100 µL of 2:1:1 isopropanol : acetonitrile : water. The lipid extracts for each sample were separated by LC using a Dionex UltiMate3000 RS Autosampler (Thermo Scientific) coupled to an LTQ Velos Pro Orbitrap (Thermo Scientific) for MS. Experiments were run by Julia Babuta, a PhD student supervised by Dr Zoe Hall (Imperial College London). In detail, 10 µL lipid extract were separated at a flow rate of 0.3 mL/min on an Acuity C18 BEH column (Waters, 50x2.1 mm, 1.7 µm) at 55 °C. Lipids were initially separated with 60:40 mobile phase A (acetonitrile : water 60:40, with 10 mM ammonium acetate) : mobile phase B (isopropanol : acetonitrile 90:10, with 10 mM ammonium acetate). This mobile phase gradient was linearly changed to 1:99 A:B over 8 minutes and kept constant for 30 seconds before switching back to 60:40 mobile phase A:B in 10 seconds. These conditions were maintained for a further 2.5 minutes. Lipids were analysed by MS in negative ion mode (spray voltage 2.8 kV, desolvation temperature 380 °C, desolvation gas 40 arbitrary units) with range 100 - 2000 m/z and mass resolution 60,000. Lipids were annotated by accurate mass through the LIPID MAPS database²⁷⁹. Tandem MS (collision induced dissociation) was performed on membrane samples to fragment intact lipids, in order to identify the fatty acyl chain composition using their diagnostic ions.

3.3 Results

3.3.1 MmpL3 is present in two populations in SMALPs and MSP-based NDs

From the SEC profile obtained for *Msmeg* MmpL3 ΔC reconstitution into SMALPs, two individual fractions were selected and analysed by SV-AUC as representative of the two different MmpL3 populations (Figure 3.1, a). Fraction 1 eluted between 9.5 and 10 mL during SEC purification, while fraction 2 between 11.5 and 12 mL. SV-AUC results confirmed the presence of two clear protein populations, as seen

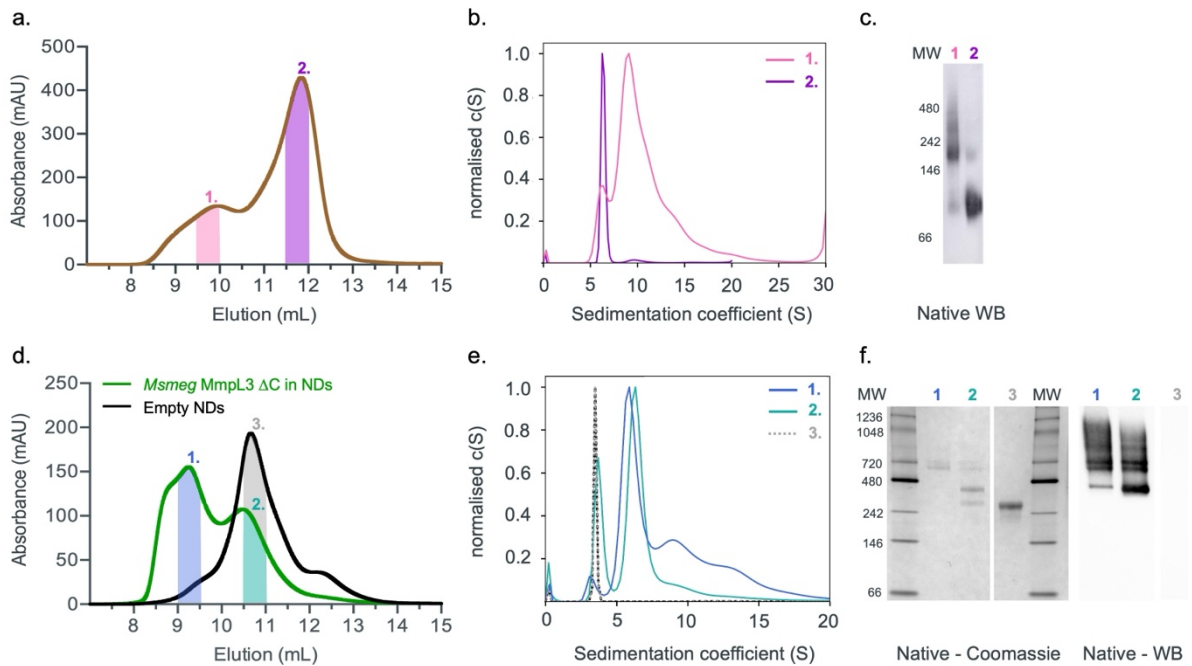


Figure 3.4. SV-AUC results for *Msmeg* MmpL3 Δ C in SMALPs and MSP-NDs. a. Size exclusion profile of *Msmeg* MmpL3 Δ C isolated in SMALPs. The high (peak 1) and low (peak 2) molecular weight peaks of the MmpL3 protein are indicated by the pink and purple shaded areas. b. The same colouring is used for the SV-AUC analysis with best fit absorbance $c(s)$ distributions of sedimented species obtained for protein in peaks 1 and 2 shown in pink and purple, respectively. c. Native Western blot analysis of the SMALP isolated samples using the anti-His tag antibody. d. Size exclusion profile of *Msmeg* MmpL3 Δ C isolated in DDM and then reconstituted into MSP-nanodiscs (dark green). Peaks 1 and 2 are indicated by the blue and cyan shaded areas. The size exclusion profile of empty nanodiscs (black) is shown with the peak (peak 3) indicated by the grey shaded area. e. SV-AUC absorbance $c(s)$ distributions of the two populations of *Msmeg* MmpL3 Δ C reconstituted in nanodiscs and empty nanodiscs coloured as in d. f. Native PAGE and native Western blot analysis of the *Msmeg* MmpL3 Δ C reconstituted in nanodiscs and empty nanodiscs. Lanes are labelled with peak numbers.

by the two distinct peaks centred around different sedimentation coefficients (S) values (Figure 3.1, b). For fraction 2 (purple), the predominant species was estimated to sediment at a $S = 6.26$ (126 kDa), likely indicating the monomeric form of MmpL3. The sedimentation coefficient for the predominant AUC peak observed for fraction 1 (pink), instead, was calculated to be $S = 9.09$ (213 kDa). This higher S value suggests the presence of a heavier species, possibly a MmpL3 dimer. Fraction 1 also presented a minor percentage of monomeric form of the protein, as indicated by the smaller peak centred at $S = 6.36$. The two fractions were also analysed by WB on a native PAGE gel, and showed two main bands migrating at ~ 180 kDa and at ~ 90 kDa for fraction 1 and 2, respectively (Figure 3.1, c). Therefore, the observed MWs seemed to agree with the hypothesis that MmpL3 can indeed arrange in oligomers, specifically into dimers when reconstituted into SMALPs. It should be noted that SV-AUC of SMALPs does not yield the direct mass of the protein. Instead it provides the mass of the protein

alongside the lipid and polymer that makes up the particle^{169,176,280}. This explains why the masses calculated using SV-AUC are elevated compared to the expected mass.

Intriguingly, when the same SV-AUC experiment was carried out using protein isolated in detergent and then reconstituted into MSP1E3D1-based nanodiscs, two populations of protein were also detected on top of the empty NDs peak, included as a control. As for the SMALPs, two individual SEC fractions (blue and cyan) were selected and concentrated for the *Msmeg* MmpL3 Δ C protein (dark green SEC trace), and one fraction (grey) was also selected as representative of the empty NDs sample (black SEC trace), as shown in Figure 3.4, d. SV-AUC results obtained for these fractions are shown in Figure 3.1, e. The difference between the two protein populations was less marked than in the SMALPs, either because the two fractions were closer to each other on the SEC trace, one contaminating the other, or because of the influence of the MSPs on the sedimentation of the samples. The low resolution of the peaks did not allow for an accurate estimate of the MWs for the tested MmpL3 species. Nevertheless, fraction 1 presented a higher dimer : monomer ratio than fraction 2, as observed by the different height of the peak centred at $S = 9.18$. Furthermore, it was evident that in fraction 2 the monomeric form was the most abundant species (at $S = 6.31$), although a significant proportion of the sample was also constituted by empty NDs, as shown by the smaller peak centred at $S = 3.78$. The presence of mixed ND populations was also observed when the three fractions were analysed by native PAGE gel and WB (Figure 3.1, f). Multiple bands were detected for fraction 1, especially visible in the WB, with the lower band possibly corresponding to MmpL3 monomeric form, while the higher MW bands strongly suggesting the presence of a dimer. Fraction 2 also presented multiple bands, although this time the monomer was the most abundant species. Furthermore, as observed in the SV-AUC results, fraction 2 also included a proportion of empty NDs, with a band migrating at ~ 250 kDa in the Coomassie-stained native gel, that is however not detected in the WB as the MSPs do not carry a His tag.

3.3.2 Mass photometry on *Msmeg* MmpL3 NDs

The two protein species identified by SV-AUC paved the way to investigate MmpL3 oligomeric state further. For this reason, the *Msmeg* MmpL3 ND reconstitution protocol was implemented to include a sucrose gradient ultracentrifugation step, with the aim of improving the resolution of separation of the two protein population peaks.

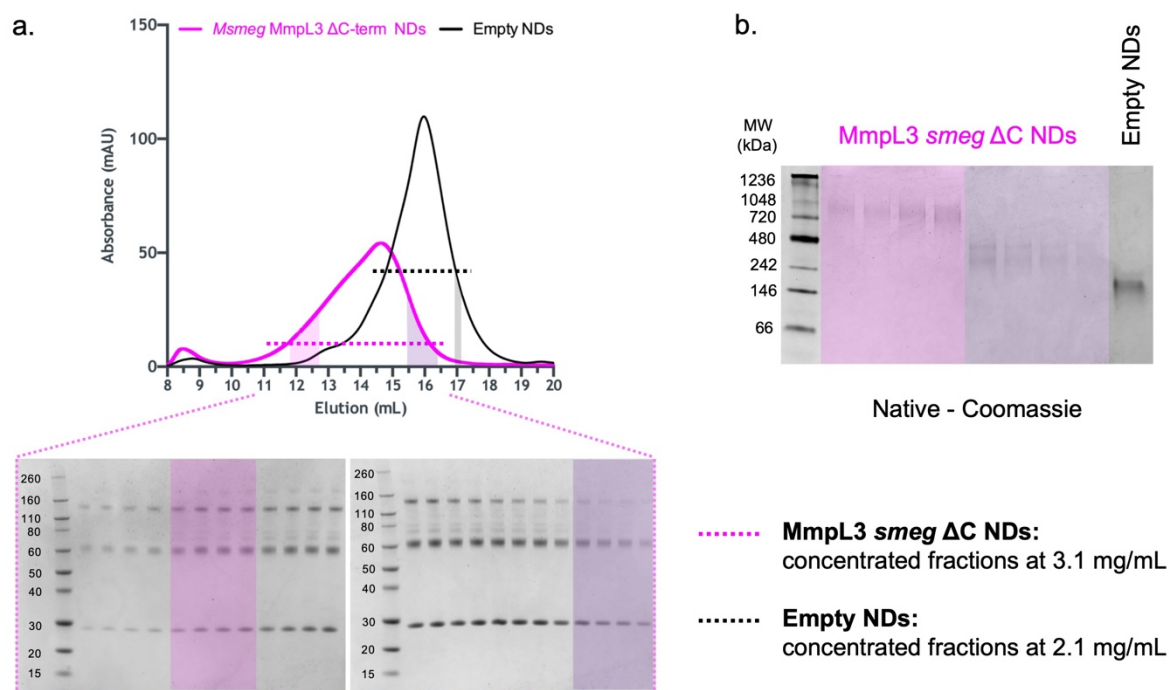


Figure 3.2. *Msmeg* MmpL3 Δ C NDs and empty NDs purification using a Superose® 6 column. a. Size exclusion profile of empty NDs (black line) and *Msmeg* MmpL3 Δ C isolated in NDs (magenta line) performed on a Superose® 6 Increase 10/300 GL column. Individual fractions for MmpL3-NDs between 12 and 16.5 mL were run on an SDS-PAGE gel to check MW. Dashed lines indicate pooled fractions that were concentrated using 100 kDa MWCO centrifugal filters. b. Highlighted fractions in a representing early-eluted fractions (pink shaded area), late-eluted fractions (purple shaded area) and empty NDs (grey shaded area) were also analysed by native PAGE to show differences in MW depending on elution volume and species.

First, the *Msmeg* MmpL3 Δ C protein was reconstituted into NDs (see protocol described in Chapter two, section 2.2.8). Empty NDs were separated from the protein-carrying ones through an additional Ni^{2+} -NTA immobilised metal affinity chromatography (IMAC) step, before performing a SEC purification (Figure 3.2). Then, pooled and concentrated fractions post-SEC for empty as well as protein-NDs were individually layered on a 10 – 30 % sucrose gradient and ultracentrifuged to allow separation of species by density (Figure 3.3, a). As the density increases linearly from top to bottom, different size proteins are fractionated along the gradient, with heavier and aggregated species settling at the bottom of the tube. It was not possible to achieve clear separation of the different populations in the *Msmeg* MmpL3 Δ C sample, as seen by the broad distribution peak obtained over the 14 – 24% sucrose gradient range (Figure 3.3, b). Nevertheless, the empty NDs gave a markedly different separation profile compared to the transporter protein sample, with a peak centred around 12 – 13% sucrose, therefore confirming their smaller size compared to MmpL3-NDs. Native PAGE and SDS-PAGE analysis (Figure 3.3, c and d) of both samples

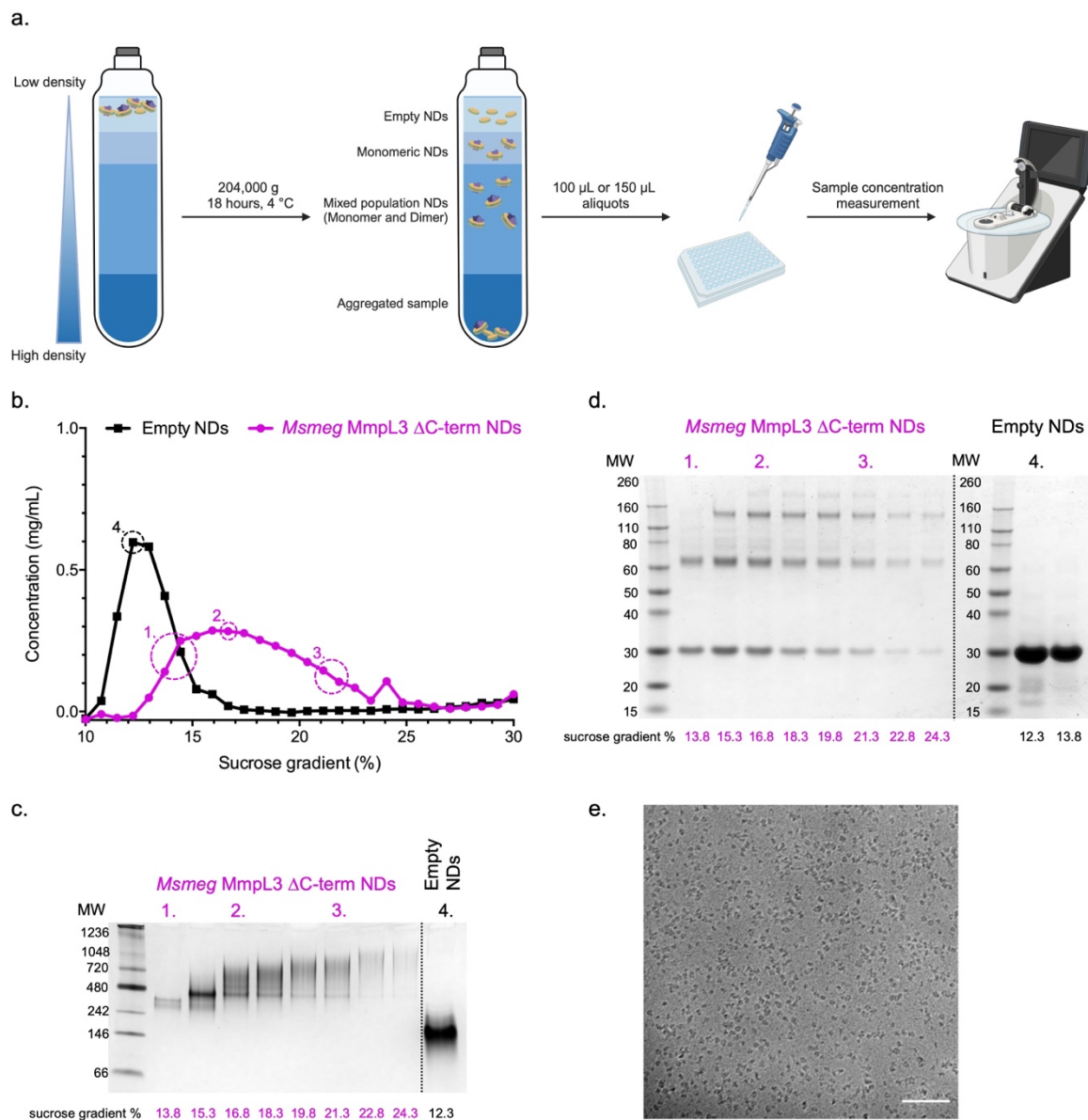


Figure 3.3. Sucrose density gradient ultracentrifugation separation of *Msmeg* MmpL3 Δ C NDs and empty NDs. a. Schematic representing the experimental procedure used to carry out sucrose density gradient ultracentrifugation. b. *Msmeg* MmpL3 Δ C (magenta) and empty NDs (black) separation profile over a 10-30% sucrose concentration range. c, d. Samples separating at different concentrations of sucrose as indicated were submitted to native PAGE (c) and SDS-PAGE (d) analysis. The data shown is representative of $n = 2$ independent experiments. e. Raw cryo-EM micrograph for aliquot 2 (in b). Particles size appears homogeneous and sample quality and concentration are suitable to carry out further analyses. Scale bar is 100 nm.

obtained at different sucrose concentrations confirmed that empty NDs were present at 12% sucrose, mainly monomeric forms of the protein at 14 – 15% sucrose and a greater amount of the oligomeric form of the protein observable at 17% sucrose. The 17% sucrose aliquot was also subjected to preliminary cryo-EM screening to check for sample quality. As shown in the raw micrograph (Figure 3.3, e), particles size distribution resulted homogeneous and sample concentration optimal, thus providing

preliminary information on sample behaviour during blotting. Results of more in-depth analyses using this technique will be described in section 3.3.4.

Representative aliquots from the sucrose gradient experiment, numbered 1 to 4 in Figure 3.3, b, were submitted to a mass photometry experiment. With this technique, the MW of single biomolecules in solution is measured through the interference contrast between reflected and scattering light at the glass-liquid interface, also known as interferometric scattering mass spectrometry (iSCAMS) ²⁷³. As the resulting interference scales linearly with the mass of the biomolecule, a mass distribution is obtained from all measurements. It is worth noticing that mass photometry measures the particle's mass independent of its shape or configuration and it is not affected by hydrodynamic radius or steric hindrance. Movies are recorded and the software detects binding events and determines the respective interferometric scattering contrasts. The contrast information is then converted through a proprietary algorithm into mass distributions by comparison with known mass standard calibrants. Compared with soluble proteins, mass photometry provides a larger error to absolute mass results in the case of NDs, for two main reasons: 1) NDs are composed by a mixture of species (e.g. amino acid and lipids) having a different refractive index but the calibration is performed using a protein, bovine serum albumin in this case; 2) mass distributions for membrane mimetics tend to be broader as a result of varying number of constituent molecules ²⁶⁶.

As an example of movies obtained, Figure 3.4, a and b, shows selected movie frames where particles are visible as circular spots, usually presenting a dark inner part surrounded by a lighter outer circle. Already upon visual inspection, the empty NDs particles (aliquot number 4 in Figure 3.3, b) appeared smaller and with lower contrast compared to the *Msmeg* MmpL3 Δ C sample (17% sucrose, aliquot number 2 in Figure 3.3, b). Empty NDs produced an homogeneous mass distribution with 87% of all binding events captured within the same peak and an average MW of 120 kDa, with a standard deviation of 28 kDa (Figure 3.4, c). This MW value was significantly lower than the masses obtained for all MmpL3-carrying NDs sample, confirming the previous observation on particle size difference during movie inspection (Figure 3.4, a and b). The MW obtained for unoccupied NDs correlated well with the 1:50 MSP to lipid ratio used for the experiment, as the MSP-subtracted mass obtained of 60 kDa corresponded to ~80 POPC lipid molecules.

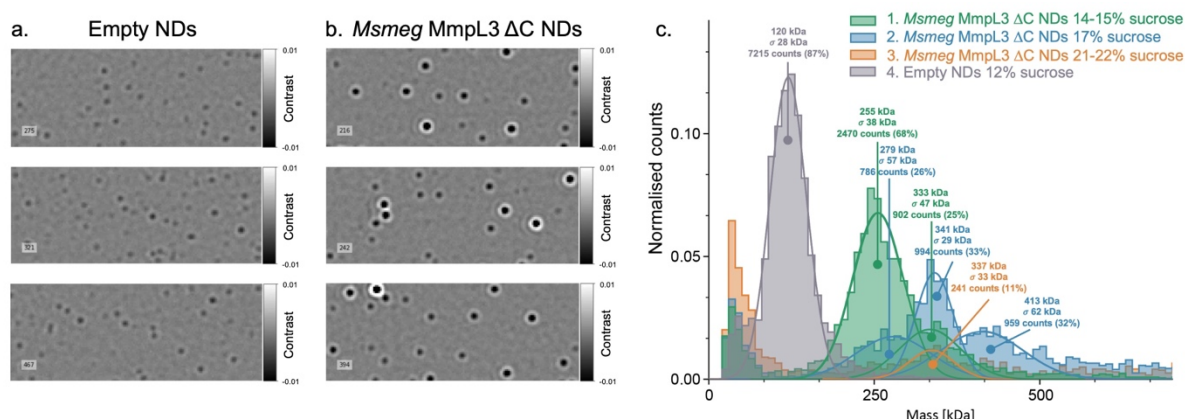


Figure 3.4. Mass photometry on *Msmeg* MmpL3 Δ C and empty NDs. a, b. Extracted frames from movies recorded for empty NDs (a) and *Msmeg* MmpL3 Δ C NDs separated at 17% sucrose density (b) showing particles at different contrast. c. Particle mass distributions for *Msmeg* MmpL3 Δ C NDs at 14 – 15% sucrose density (aliquot 1, green), 17% sucrose density (aliquot 2, blue) and at 21 – 22% sucrose density (aliquot 3, orange). Empty NDs at 12% sucrose density (aliquot 4) are in grey.

Interestingly, the 14 – 15% sucrose aliquot, mainly monomeric form of the protein, showed a polydisperse distribution, with a major peak (68% of total counts) centred at ~255 kDa, and a smaller peak representing the 25% of total events at ~333 kDa (Figure 3.4, c). These two MWs, also considering the presence of 2 copies of MSP1E3D1 and POPC lipids per ND particle, are compatible with the monomeric and dimeric form of the protein, respectively. The 17% sucrose aliquot produced a main peak including 33% of binding events with a mean MW value of 341 kDa, compatible with a protein dimer (Figure 3.4, c). Two smaller peaks were present at ~279 kDa and at ~413 kDa, although the spread of values for these distributions was quite large. The 21 – 22% sucrose aliquot mainly produced low MW reads, while 11% of particles showed a distribution centred at ~337 kDa, once again corresponding to the mass of a dimer (Figure 3.4, c). Therefore, the 21 – 22% sucrose sample is unlikely to contain species larger than dimers. Thus, the results obtained for the 21 – 22% sucrose sample implicitly exclude that the heavier species detected at 413 kDa for the 17% sucrose sample could be trimers, as the same oligomeric arrangements would be otherwise present at 21 – 22% sucrose.

The overall results from mass photometry confirmed the presence of two main protein populations in the MmpL3-NDs, as observed by SV-AUC. Although the sucrose gradient ultracentrifugation step did not fully separate protein species, it was possible to observe a different MW distribution for the 14 – 15% sucrose sample compared to the 17% sucrose aliquot. The sample at a lower sucrose density showed

a higher proportion of monomer, while the dimeric form of the protein was the main population of the higher density aliquot.

3.3.3 Native MS analysis confirms observations of *Msmeg* MmpL3 oligomers

The same aliquots separated by sucrose density gradient ultracentrifugation were also submitted to native MS analysis. The LILBID analysis clearly showed that *Msmeg* MmpL3 Δ C protein that separated to 14 – 15% sucrose, renamed low sucrose percentage (LSP), was monomeric in form with various different charge states when subjected to both soft (10 mJ) and harsh (23 mJ) laser intensities (Figure 3.5, a). At soft laser intensity, the main singly charged species between 150 and 170 kDa agreed with the expected MW of a ND including 2 copies of MSP1E3D1 (~30 kDa each) and 1 copy of MmpL3 (~86 kDa). In contrast, the protein sample that separated at higher concentrations of sucrose (17%), renamed high sucrose percentage (HSP), showed two relevant peaks for the dimeric form of the protein again with a variety of charge states both at soft and harsh laser intensities (Figure 3.5, b). This sample also contained small populations of monomeric forms of MmpL3, as shown by the peak at ~150 kDa in Figure 3.5, b. This may be due to residual monomeric protein in the fraction even after the sucrose gradient separation. The theoretical masses of complexes including two MSPs and either monomer or dimer of *Msmeg* MmpL3 Δ C are 145,835 kDa and 231,705 kDa, respectively (indicated by dotted lines in the peak insets in Figure 3.5, left panel). The proteo-NDs also contain lipids, of which some are retained during the LILBID process. The broad peaks of the LILBID spectra are due to a statistical distribution of bound lipids in the proteo-NDs species. The experimental MWs of the peak insets represent the NDs without any lipids and were determined as 147 kDa and 232 kDa, in very close agreement with the theoretical values and strongly supporting the presence of monomeric and dimeric species in our samples. Both the LSP and HSP samples were also analysed under high energy laser (23 mJ), inducing complex dissociation. Interestingly, when the HSP sample was analysed under these conditions not only a peak inset of the MW of the protein dimer (~170 kDa) was observed, but the intact MmpL3-NDs (peak inset at ~230 kDa) was also present, corroborating the hypothesis that MmpL3 can indeed form dimers, and these species are also relatively stable (Figure 3.5, b). Using harsh laser conditions on LSP produced peaks matching MWs of monomeric species (peak inset at ~87 kDa) as well as partly dissociated monomeric MmpL3-NDs (Figure 3.5, a).

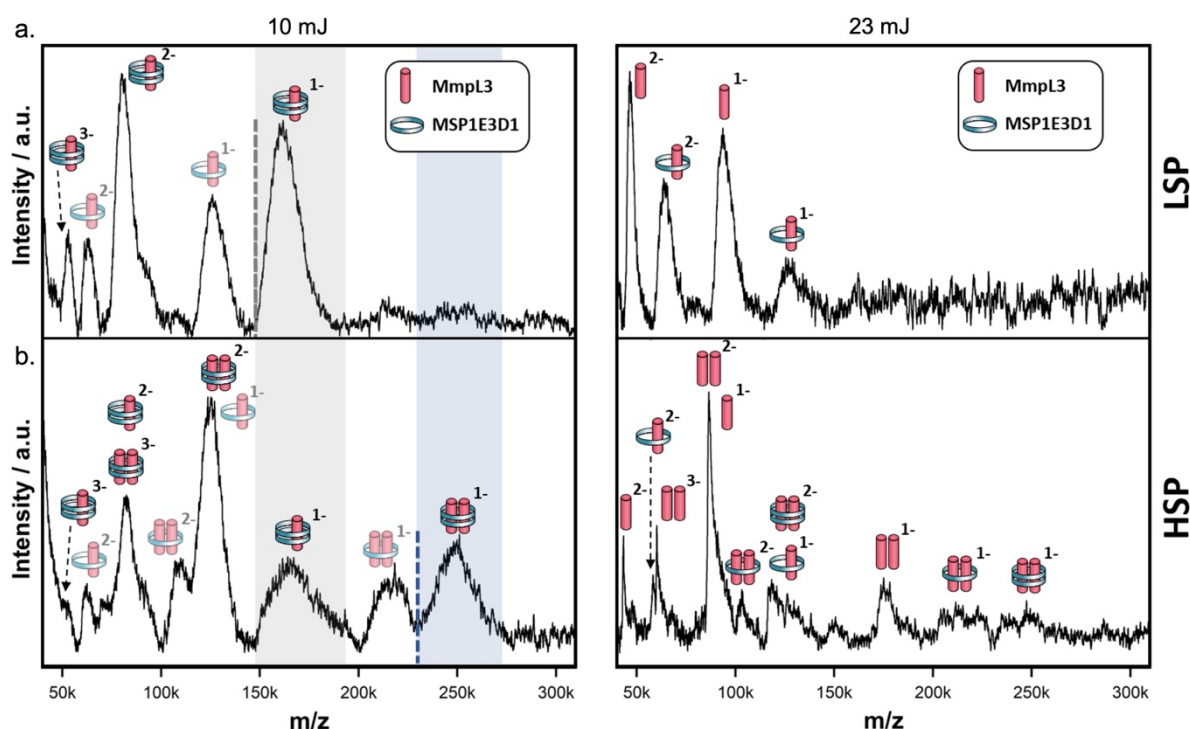


Figure 3.5. Native Mass Spectrometry for *Msmeg* MmpL3 Δ C NDs. a. nMS spectra of the *Msmeg* MmpL3 Δ C protein in MSP-NDs separating at low sucrose percentage (LSP) measured at 10 mJ (left panel) and 23 mJ (right panel) laser intensity. b. nMS spectra of the *Msmeg* MmpL3 Δ C protein in MSP-NDs separating at high sucrose percentage (HSP) measured at 10 mJ (left panel) and 23 mJ (right panel) laser intensity. The different protein-MSP combinations detected in each sample and relative charged species are indicated with pictograms. The translucent pictograms are species for which one MSP copy dissociated during the laser desorption process. Dotted lines in left panel (10 mJ) indicate the theoretical mass of NDs particles made of two MSPs copies and either one (grey) or two (blue) *Msmeg* MmpL3 Δ C proteins, respectively. The shaded areas to the right of these dotted lines cover the broad peaks, which stem from the inhomogeneous distribution of lipids within the respective proteo-NDs. The grey-shaded area indicates the presence of monomeric MmpL3-NDs while the blue-shaded area indicates dimeric MmpL3-NDs. The data shown is representative of $n = 2$ independent experiments.

3.3.4 Preliminary cryo-EM analysis shows *Msmeg* MmpL3 Δ C exists as a dimer

The HSP aliquot selected from a further *Msmeg* MmpL3 Δ C purification and sucrose gradient ultracentrifugation was prepared (Figure 3.6, a and b). Aliquots numbered 1 to 4 were analysed by SDS- and native PAGE gels, and protein bands detected by either Coomassie stain or anti-His WB analysis. There was good reproducibility of the separation results with the previous sucrose-based NDs separation (Figure 3.3), as indicated by the overlay of the SEC traces and the density gradient profiles (Figure 3.6, c). Aliquots of protein for cryo-EM were selected after comparison with aliquots submitted to mass photometry and nMS, to increase the chances of detecting MmpL3 in its dimeric form.

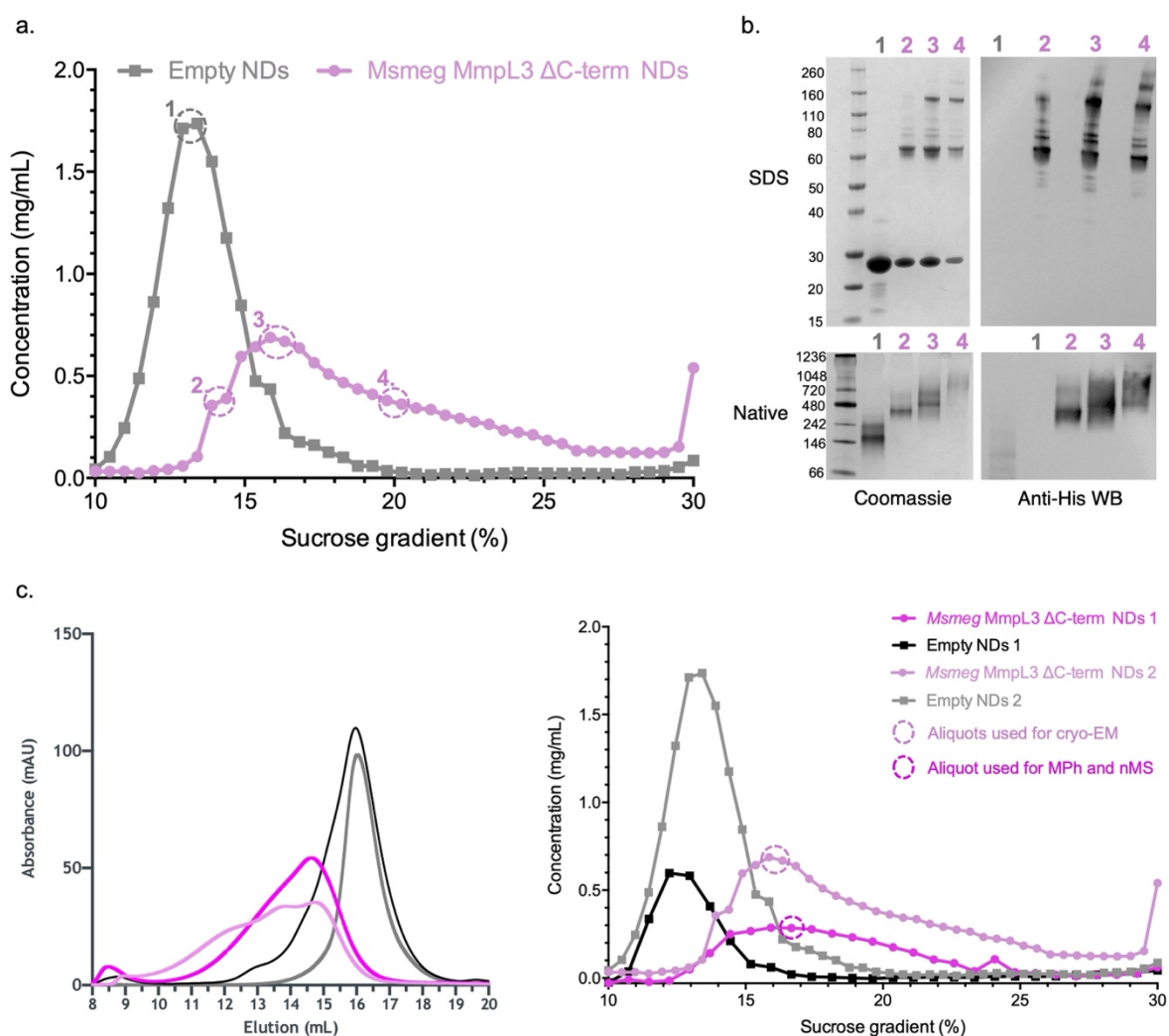


Figure 3.6. Second sucrose density gradient separation of *Msmeg* MmpL3 Δ C NDs and empty NDs. a. *Msmeg* MmpL3 Δ C (pink) and empty NDs (grey) separation profile over a 10 – 30% sucrose concentration range. b. Samples separating at different concentrations of sucrose, as indicated in a, were submitted to SDS- and native PAGE analysis. Gels were either stained using Coomassie or used to perform WB analysis. c. Comparison with sucrose gradient experiment shown in Figure 3.3. Left panel, SEC purification profile for *Msmeg* MmpL3 Δ C NDs and empty NDs from the two independent experiments on a Superose® 6 Increase 10/300 GL column. Right panel for sucrose gradient ultracentrifugation profiles of NDs samples over $n = 2$ independent experiments, with magenta dashed circle indicating aliquot from first experiment used for mass photometry (MPh) and nMS, while pink circle indicates combined aliquots used for cryo-EM analysis obtained from second experiment.

Aliquot 3 from Figure 3.6, a was analysed by cryo-EM and monodispersed samples of monomeric and dimeric populations of protein were observed. In order to facilitate selection of the monomeric versus dimeric protein, synthetic models of both forms of the protein in nanodiscs were generated using CHARMM-GUI and are shown in the upper panels of Figure 3.7, b and c, respectively. The experimental 2D class averages obtained after performing dimer template picking and optimisation of window size indicated that in some face-on views there was clear differentiation between monomer (Figure 3.7, b, middle panel) and dimer (Figure 3.7, c, middle panel). Views

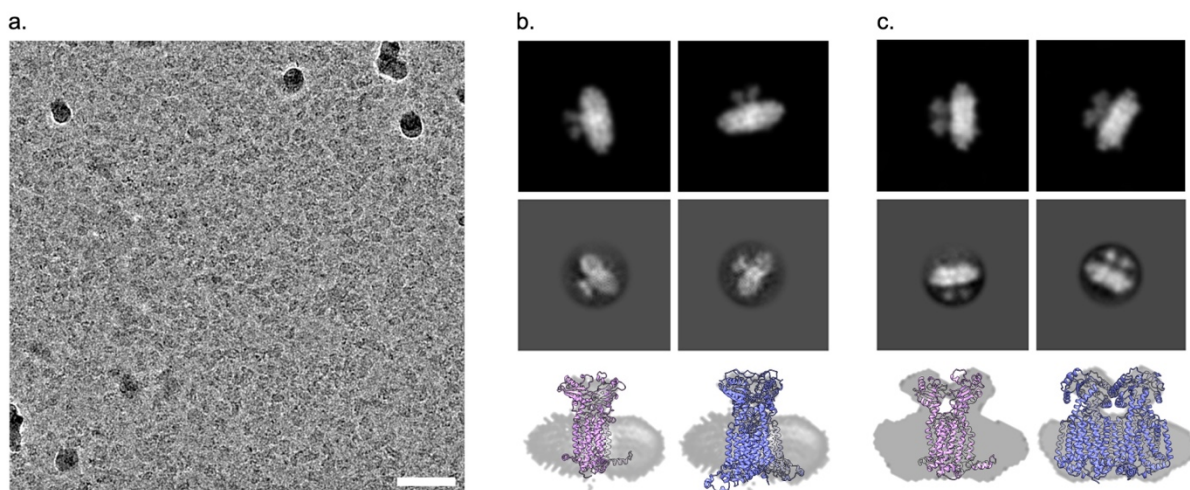
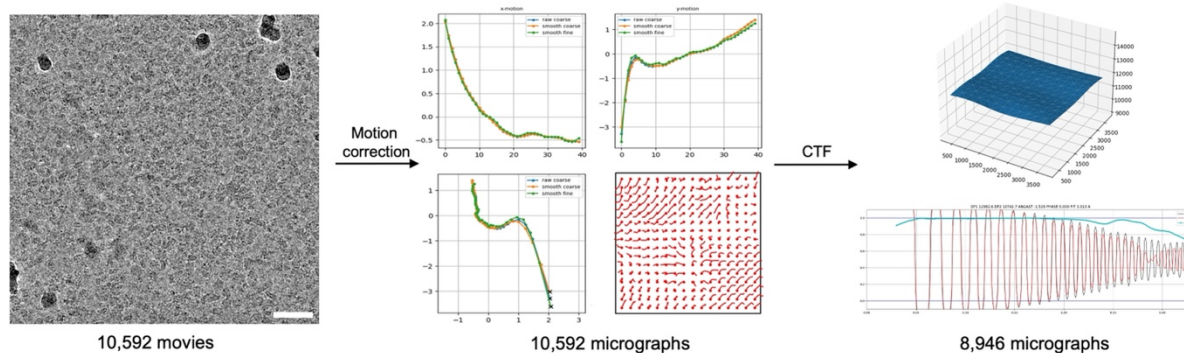


Figure 3.7. 2D classifications of monomeric and dimeric assemblies of MmpL3. a. Electron cryo micrograph of *Msmeg* MmpL3 Δ C nanodisc sample, scale bar 200 nm. b, c. Simulated (top panel) and real (middle panel) 2D classifications of MmpL3 monomer and dimer shown respectively in b and c, with bottom panels showing a molecular model fit to side views from 2D classification. Monomeric MmpL3 is shown in pink ribbon representation, while dimeric model in blue.

from the side are harder to differentiate since both the monomer and dimer looked similar with just slightly greater levels of density looking through the protein, especially visible in the TM domains. In particular, as shown in bottom panel of Figure 3.7, b and c, monomer and dimer models were fitted onto their representative 2D projections. When the monomer-like 2D projection was considered, both MmpL3 monomer and dimer fit if front-facing structures are rotated by 90° along the membrane plane so that one protomer covers the other (Figure 3.7, b, bottom panel). However, the same was not true for the monomer when the 2D projection of the dimer was used as reference. In this case, the experimental density observed for the two periplasmic lobes was significantly larger than each of the periplasmic domains in the MmpL3 monomeric structure. The two lobes, however, were fully occupied when the front view of the dimer model was fitted in (Figure 3.7, c, bottom panel). Despite the promising results obtained and the clear identification of front- and possible side-views of MmpL3 dimers, the lack of top-down views hindered the generation of accurate 3D maps from *ab initio* modelling. Details on the cryo-EM analysis workflow is shown in Figure 3.8.

Although these preliminary results were sufficient to prove the existence of dimeric forms of MmpL3, additional experiments would be necessary to improve dimer/monomer separation and to collect larger data sets to sample a higher number of different protein orientations. This would allow generation of more accurate 3D maps by *ab initio* reconstruction, potentially leading to a high resolution refined MmpL3 dimer structure.

a. Pre-processing



b. Particle-picking strategy

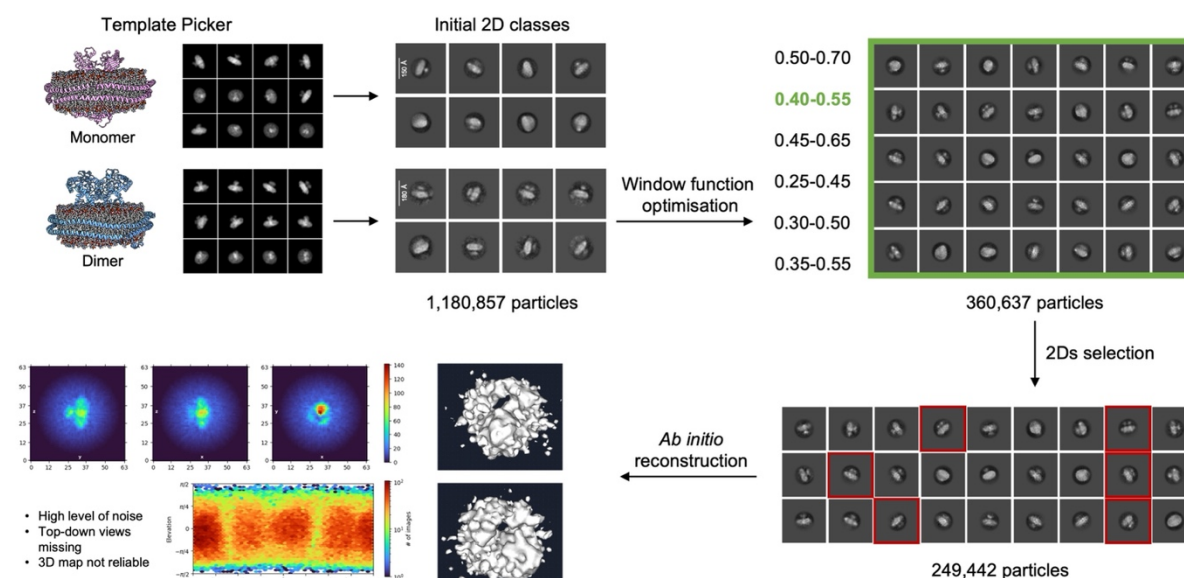


Figure 3.8. Cryo-EM analysis workflow. a. Pre-processing of movies and extracted micrographs, with motion correction and CTF performed. b. Particle-picking strategy, with synthetic models of NDs carrying either monomeric or dimeric MmpL3 used as template to pick particles and obtain initial 2Ds. Percentage of box size was optimized to adjust focus on single particles and perform several rounds of 2D classifications. Preliminary *ab initio* reconstruction was attempted but high level of noise and lack of top-down views gave not reliable 3D maps.

3.3.5 Lipidomics analysis show lipid species are retained during MmpL3 purification in DDM, MSP-NDs and SMALPs

Being embedded in the inner membrane of Mycobacteria, MmpL3 naturally interacts with the surrounding lipid molecules like any other integral membrane protein. However, we also know that MmpL3's function is to transport trehalose monomycolate (TMM) molecules into the periplasm as precursors for the assembly of the pathogen's outer membrane. Therefore, it is important to study the composition of lipid molecules co-purified with MmpL3 and to investigate potential variations occurring using different methods for its reconstitution into membrane mimetics systems.

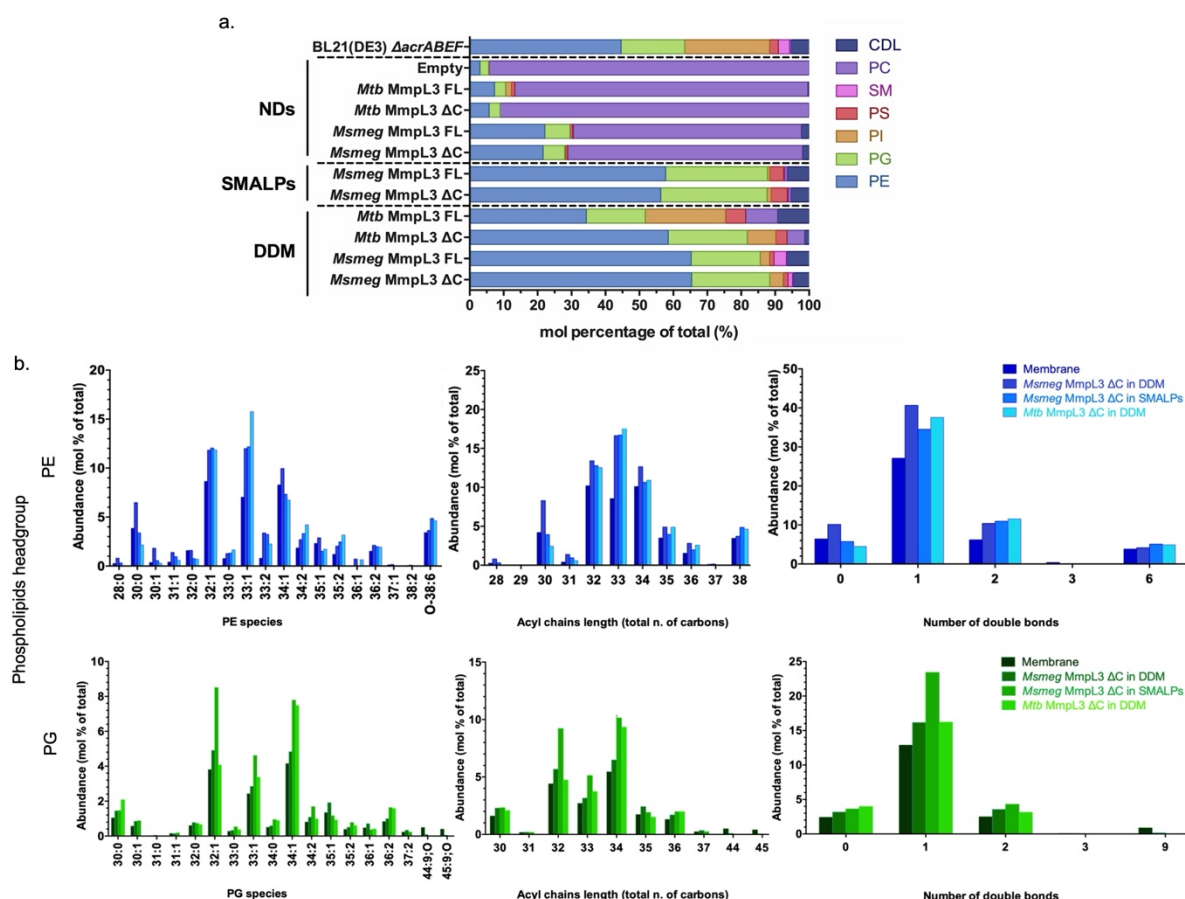


Figure 3.9. Lipidomics analysis on *Msmeg* and *Mtb* MmpL3 protein in detergent, MSP-NDs and SMALPs. a. For each sample, abundance of lipid species was plotted as stacked bars. Main lipid species identified in MmpL3 samples, empty NDs and *E. coli* membrane selecting a 2 - 7 min range as retention time on chromatogram. For each sample, abundance of lipid species was calculated as relative mol % over total lipid moles and was plotted as stacked bars, with individual lipids coloured according to the key. b. Abundance of specific PE (top row) and PG (bottom row) lipid species identified in *E. coli* membrane, *Msmeg* MmpL3 ΔC in DDM and SMALPs as well as *Mtb* MmpL3 ΔC in DDM (left panel). Acyl chains length (middle panel) and number of double bonds (right panel) were also analysed.

For this purpose, *Msmeg* MmpL3 FL and ΔC in DDM, MSP-NDs and SMALPs as well as *Mtb* MmpL3 FL and ΔC in DDM and MSP-NDs were submitted to lipidomics analysis (Figure 3.9, a). A range of 59 – 109 lipids was identified per sample, including species from the PE, PC, SM, PG and CDL lipid classes. Isolated membranes from *E. coli* BL21(DE3) $\Delta acrABEF$ were used as reference for the expression system's lipid composition. Furthermore, empty POPC NDs were used as background for MSP-based NDs, as MSP1E3D1 proteins for their amphipathic nature could retain some lipids during purification. Next, the average relative abundance of the main lipid classes across the different constructs / isolation methods was calculated and the lipid composition varied according to the isolation method (Figure 3.9, a). Lipid species eluting between 2 – 7 min retention time and within selected m/z range were identified

on MS spectra and their relative abundance calculated relatively to the internal standards (Figure 3.10, a). Lipids associated with MmpL3 that had been isolated in DDM or SMALPs were mainly from the PE and PG lipid classes. In contrast, when isolated in DDM and reconstituted in NDs, the major lipid class identified was PC. This is expected since NDs are composed of 16:0-18:1 PC lipids (also known as POPC). Thus, with the only exception of *Mtb* MmpL3 FL in DDM, all DDM-purified MmpL3 and SMALPs resulted in a very similar lipid composition, made of ~60 - 65% PE, ~25 - 30% PG and the other 10% composed of a mixture of PI, PS, SM, PC and CDL (Figure 3.9, a). By comparing the percentages for each lipid species with BL21(DE3) Δ acrABEF membrane composition, an enrichment in PE (+15%) and PG (+7.7%) and a depletion of ~22% of PI was observed on average in these samples. Another observation worth noting was that the enrichment of PE and PG observed in DDM and SMALPs samples was also detectable in the MSP-NDs, although with a lower mol % of lipids over total due to the very high percentage of PC used to generate the NDs. Taking the lipid composition of empty NDs as reference, 3% of total lipid moles was recorded by PE, 2.5% by PG, and an expected large 94.3% by POPC. MmpL3-NDs samples recorded on average a 14.2% of PE (21.9% if only *Msmeg* considered) and a 5.1% of PG (6.9% if only *Msmeg* considered), with POPC present at 78.4% (68.1% for *Msmeg* only), despite the latter being the only exogenous lipid species added during ND reconstitution, following detergent extraction.

The abundance of the main PE and PG lipid species identified in the membrane, in *Msmeg* MmpL3 Δ C (DDM and SMALPs) as well as in DDM-purified *Mtb* MmpL3 Δ C was plotted (Figure 3.9, b). Acyl chain length as well as the number of double bonds were also investigated, with the majority of lipid tails mostly monounsaturated and ranging between 30 and 36 total carbons in length shared between the two alkyl chains.

3 PE and 3 PG species were particularly abundant in all DDM and SMALPs samples (Figure 3.10, a and b). These were PE 32:1, 33:1 and 34:1, with the first two species showing an abundance of ~12%, while PE 34:1 was present at ~8.6 %. For some lipids, MS/MS fragmentation provided information about the acyl chains constituting the individual species. In particular, as shown in Figure 3.10, c and d, both PE 33:1 and PG 33:1 had a palmitic acid chain (16:0) as sn1 and a cis-9-heptadecenoic acid chain (9Z-17:1) as sn2. Some of the other species identified are reported in the table (Figure 3.10, b).

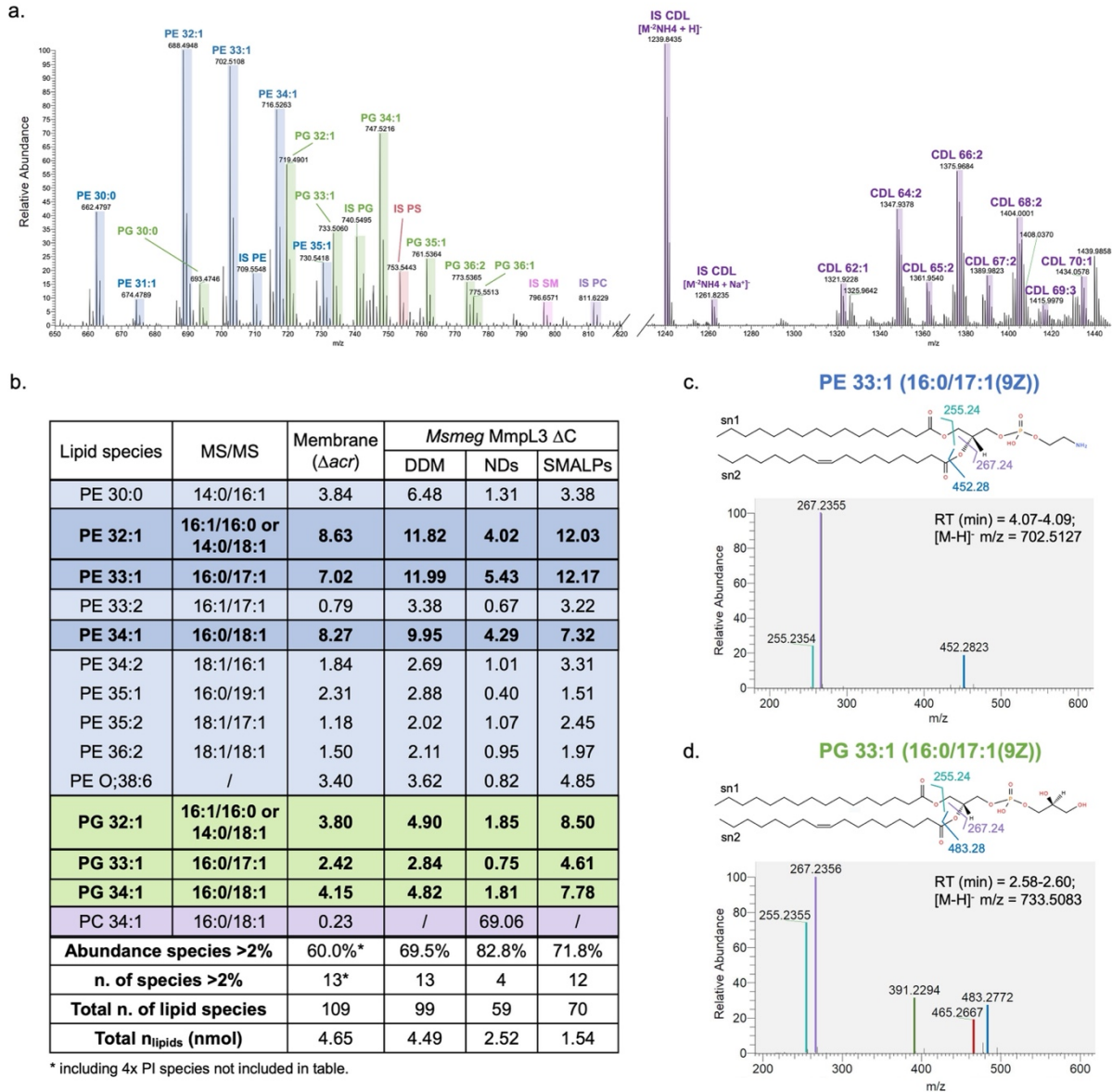


Figure 3.10. Abundance of lipid species in *Msmeg* MmpL3 Δ C in DDM, POPC NDs and SMALPs.
a. MS spectra for MmpL3 Δ C in DDM showing relative abundance of lipid species, within selected m/z range, identified between 2 – 7 min retention time (RT). Lipid species and relative internal standards (IS) coloured according to key in Figure 3.9, a. b. Table with relative abundance of lipid species >2% as mol % of total lipids identified in MmpL3 samples and *E. coli* membrane, used as reference. MS/MS column indicates if acyl chains were identified with certainty using fragmentation protocol. c, d. MS/MS fragmentation spectra for PE 33:1 (c) and PG 33:1 (d) lipid species. Main m/z values for acyl chains fragments were identified for both species. A m/z = 255.24 indicated a sn1 RCOO⁻ ion, while m/z = 267.24 indicated a sn2 RCOO⁻ ion. Loss of sn2 as ketene from [M-H]⁻ gave m/z values of 452.28 and 483.28 for PE and PG, respectively. Other fragments identified for PG 33:1 are reported in the spectrum with relative m/z values. Species-specific RT and relative [M-H]⁻ m/z values are also reported below each spectrum.

3.4 Discussion

Several published structures have all indicated that MmpL3 was monomeric, despite being the only known member to date in the RND superfamily not to arrange

into oligomers¹⁻⁴. Findings presented in Chapter 2 had indicated that MmpL3 was forming higher order oligomers, so here we used a range of biophysical analyses to explore this in more detail.

Specifically, selected fractions for MmpL3-SMALPs, representative of the two protein populations identified by SEC, were analysed by SV-AUC and sedimentation distributions confirmed the difference in size hypothesised between early- and late-eluting SEC fractions. Compared to the 12 mL fraction, heavier species were identified for the 9 – 10 mL fraction, likely indicating a dimeric form of the protein. The sedimentation coefficient value for the main peak observed for this fraction was less than double the value observed for the monomeric MmpL3, strongly suggesting the presence of 2 protein copies in one disc, rather than monomeric MmpL3-SMALPs doublets.

Interestingly, when selected MmpL3-NDs fractions were analysed by SV-AUC, two distinct *c(s)* distribution profiles were also observed with similar *S* values to SMALPs, although with a larger frictional ratio for NDs (1.4 vs 1.2 for SMALPs), indicating a more elongated particles shape. Moreover, the early-eluting fraction showing higher sedimentation values gave a low-resolution peak, possibly suggesting the presence of a mixture of protein populations. The cross-contamination between species was also observed by native PAGE analysis, especially for dimer/monomer and for monomer/empty NDs in the early- and late- eluting SEC fractions, respectively.

The SV-AUC results indicated that interactions supporting oligomer formation might be relatively weak compared to the stability shown by the monomer. The higher MW species showed a dual *c(s)* distribution compared to the monodispersed peak observed for the monomer, possibly implying a rearrangement from dimer into more stable particles carrying 1x protein copy. Nevertheless, dimeric species were constantly observed regardless of the method adopted for protein purification, either by detergent-extraction followed by MSP-based POPC NDs reconstitution or by SMA polymer extraction of MmpL3 directly from the membrane to form SMALPs. Both systems benefit from membrane lipid molecules that are retained in the final nanodiscs and this results in a more physiological arrangement of a protein or protein complex.

Further proof in support of the oligomeric status of the protein came from mass photometry and nMS. These techniques were used on representative *Msmeg* MmpL3 ΔC NDs aliquots separated on a sucrose density gradient. For both techniques, the protein that separated to 14 – 15% sucrose (LSP) was monomeric in form, with a MW

of ~147 kDa determined by nMS at the peak inset and a higher MW of 255 kDa from mass photometry. The discrepancy between these values could be due to the retention of lipid molecules in the mass photometry measurements that might contribute to the increment of the particle contrast, overall leading to a higher MW. The protein sample that separated at higher concentrations of sucrose (HSP) was largely dimeric in form as determined by mass photometry, although also presenting smaller and higher MW species. Small populations of monomeric forms of the protein were also observed by nMS measured at soft laser intensity. LILBID-MS is a soft mass spectrometry ionisation method, which allows the investigation of non-covalently bound complexes including their stoichiometry under almost native conditions ²²¹. By tuning the laser intensity, complex stability and lipid-protein interactions are probed. The use of elevated laser intensity causes the explosive expansion of the sample droplet and results in partially dissociated ND particles. Under these conditions, when ND samples were analysed by Henrich and colleagues, the scaffold proteins were among the first proteins to dissociate from the NDs and consequently lipids were also lost ²¹⁸. This is likely to account for the signals representing MmpL3-NDs with only 1x MSP copy as well as for protein-only signals detected in both LSP and HSP samples. Of particular interest is that in the HSP sample MmpL3 oligomeric species dissociated from the nanodiscs without losing their integrity, indicating that the protein indeed formed complexes and that was not present as two separate entities in the NDs.

The ultimate evidence that MmpL3 forms dimers was obtained when the HSP sample for *Msmeg* MmpL3 Δ C into NDs was submitted for cryo-EM analysis. The resulting experimental 2D class averages indicated a clear differentiation between monomer and dimer, especially from front views of the membrane-plane. In these orientations, two lobes were visible that stood out from the nanodisc level. Each lobe could be overlaid with a monomeric MmpL3 carrying two periplasmic domains, therefore only a dimer model could fit with the observed density. Unfortunately, the lack of top-down views possibly due to misalignment with the protein template did not allow us to obtain a 3D map, necessary to obtain a preliminary structure for further refinement. Although these results were promising and visually proved the existence of MmpL3 as a dimer into NDs, more optimisation is needed on two main fronts. First, a better separation between monomer and dimer forms of the protein would allow the preparation of cleaner, more homogeneous samples for cryo-EM data collection. Fewer SEC fractions and preferentially from small elution volumes should be used to

perform sucrose gradient separation, facilitating the selection of dimer particles. In addition, a tailored pipeline for data processing should be developed and several conditions trialled in order to sample 2Ds showing the protein in many orientations and ultimately generate a reliable 3D map.

However, it is important to state that the dimer was only a proportion of the protein that we were able to obtain even in our most stringent isolation and separation procedures. This taken together with the findings from the earlier structural studies indicates that the separation between monomer and dimer of MmpL3 is not particularly straightforward. The reasons for this are unclear but they might relate to either the role of the dimeric form or lipid/membrane requirements of the protein. For this second hypothesis, it has already been shown in the literature that several lipid and detergent molecules co-purify with MmpL3 and were identified during protein structural studies¹⁻⁴. Furthermore, Su and colleagues showed that *Msmeg* MmpL3 can promiscuously but actively bind to several lipid species, such as PE, PG and CDL². In our hands, the lipidomics analysis performed on *Msmeg* and *Mtb* MmpL3 in different membrane-mimetics systems showed a lipid composition that overall reflected the *E. coli* strain used for expression, with a significant enrichment in PE and PG species compared to the source membrane. In particular, PE constituted ~60 – 65% of total, while PG accounted for ~20 – 25% of total lipids quantified.

Our biophysical analysis has provided strong evidence that the protein is dimeric, however a high-resolution structure would be necessary in order to obtain information on the dimer interface. To address this key biological question and investigate protein-protein interactions at atomistic level, MD simulations were performed on a model of MmpL3 dimer. The experimental evidence gathered on membrane lipid composition was used to generate the simulated lipid bilayer, and protein-lipids interactions were also investigated. MD simulations results will be discussed in the next Chapter.

4. Investigation of monomeric and dimeric *Msmeg* and *Mtb* MmpL3 using Molecular Dynamics (MD) simulations

4.1 Introduction

The experimental evidence presented in Chapter 3 demonstrated the existence of MmpL3 as a dimer. However, as it was not possible to solve a high-resolution structure for this oligomeric species, the investigation of the protein-protein interactions could not be carried out. The co-purified lipid species identified through lipidomics analysis in the previous Chapter confirmed that MmpL3 actively interacts with the surrounding environment. Nevertheless, the protein-lipid binding sites could not be identified.

MD simulations represent a valid complementary approach to experimental procedures to answer key biological questions. Complex multicomponent systems can be investigated in atomistic detail, providing insights on protein, protein-protein and protein-lipid dynamics and interactions. One of the greatest advantages of MD is that the user has full control over the system to simulate, as well as on the environmental parameters to use.

In this Chapter, MD results obtained from monomeric *Msmeg* and *Mtb* MmpL3s are first discussed. MmpL3's published structures were embedded in a POPC lipid bilayer to replicate NDs lipid composition and investigate the protein behaviour within the membrane environment. Then, dimer models were also generated using the web-based GalaxyHomomer server starting from the protein structures. MD simulations performed on the *Msmeg* MmpL3 dimer allowed investigation of the protomer-protomer interactions providing useful insights into the interfacial residues involved. Furthermore, specific lipid binding sites were identified suggesting the role that lipids might play in stabilising the dimer. These findings could represent the starting point for future experiments investigating MmpL3 oligomeric state and how this arrangement might affect its function.

4.2 Materials and Methods

4.2.1 *Msmeg* MmpL3 crystal structures evaluation

Two crystal structures of the membrane protein MmpL3 were obtained from the RCSB Protein Data-bank database (<http://www.rcsb.org>)²⁸¹. The MmpL3 obtained from the native host *Mycobacterium smegmatis* (strain ATCC 700084/mc(2)155) solved in 6-n-dodecyl- α,α -trehalose (6-DDTre, a homologue of the native TMM substrate) was reported in Zhang *et al.* (6JAF – Uniprot ID I7G2R2), while MmpL3 from *M. smegmatis* heterologously expressed in *Escherichia coli* solved in DDM was reported in Su *et al.* (6OR2 – Uniprot ID A0QP27)^{1,2}. Both structures were solved by X-ray diffraction at resolutions of 2.70 Å for 6AJF and 2.59 Å for 6OR2. Deposited structure validation reports were reviewed to assess the overall quality of the structures. The parameters assessed were resolution, R-free values, Ramachandran and RSRZ outliers²⁸². Sequences were aligned using the Align function within UniProt, which uses the Clustal Omega algorithm and structurally superimposed using UCSF Chimera MatchMaker using the Needleman-Wunsch alignment algorithm and BLOSUM-62 matrix^{283–285}. Root-Mean-Square deviations (RMSd) of atomic positions were calculated using the Match -> Align function and unmodelled non-terminal residues present in the model were identified²⁸⁵. Following an extensive evaluation of both crystal structures, as presented in the Results section, 6OR2 was used to carry out all the work related to *Msmeg* MmpL3 in this thesis.

4.2.2 *Msmeg* MmpL3 Δ C model building for MD

The 2.6 Å crystal structure of C-terminally truncated *Msmeg* MmpL3 (UniProtKB A0QP27, aa 1 - 761, PDB 6OR2) was used as the reference structure². Missing non-terminal residues (aa 353 - 381) were modelled through the model loop function of Modeller software, available as a built-in tool in UCSF Chimera²⁸⁶. A standard loop modelling protocol was used to model the non-terminal missing structure, allowing for 1 residue adjacent to the missing regions to move. 30 models were generated and ranked based on the GA341 score, which determines the statistical probability of having the correct fold, and a normalised Discrete Optimised Protein Energy score (zDOPE), where more negative values indicate more favourable models^{287,288}. Alongside these factors, the specific amino acids composition was calculated using a

web-based tool for Peptide Hydrophobicity/Hydrophilicity Analysis (peptide2.com). Finally, secondary structure prediction was performed using JPred4 and SOPMA web tools ^{289,290}. The model renamed '1.16' was selected based on the aforementioned criteria and the PDB file was processed through the CHARMM (Chemistry at HARvard Molecular Mechanics) - GUI Membrane Builder web-based tool (<http://www.charmm-gui.org/>) ²⁶². This tool was used to generate the model of 1.16-6OR2 embedded in a POPC symmetric bilayer. The PDB was manually oriented with respect to the membrane planes, and the lipid bilayer was then generated by a replacement method using a 1:1 ratio of POPC molecules in the upper- and lower-leaflet. Furthermore, 0.15 M KCl ions were placed using the Monte Carlo method ²⁹¹. All the components were finally assembled, including water molecules as per the specified thickness of 22.5 Å, and a PDB file was generated.

In order to simulate the presence of a ligand within the periplasmic cavity, 3 different models were generated by substituting coordinates of the PE molecule present in the 6OR2 PDB with a molecule of POPC, POPE or 1-stearoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (SOPE), respectively ². The selected lipid molecules were selected to provide a range of different alkyl chain lengths as well as headgroups.

4.2.3 *Mtb* MmpL3 ΔC model building for MD

The 3 Å cryo-EM structure of C-terminally truncated *Mtb* MmpL3 (UniProtKB P9WJV5, aa 1 - 753, PDB 7NVH) heterologously expressed in *Escherichia coli* and isolated in Lauryl Maltose Neopentyl Glycol (LMNG) was used as the reference structure ⁴. Missing non-terminal residues (aa 343 - 377) were modelled through the model loop function of Modeller software, as already described in section 4.2.2 ²⁸⁶. The same criteria, as described in section 4.2.2, were applied to select the best model. The PDB file obtained was then processed through the CHARMM - GUI Membrane Builder web-based tool following the same process used for the *Msmeg* MmpL3 ΔC model ²⁶².

4.2.4 Oligomeric structure prediction for *Msmeg* and *Mtb* MmpL3 ΔC constructs

Msmeg MmpL3 (PDB ID: 7N6B) and *Mtb* MmpL3 (PDB ID: 7NVH) oligomeric structure prediction were made using the GalaxyHomomer server, available on <http://galaxy.seoklab.org/homomer> website ^{3,4,292}. The algorithm used templates

selected from the protein structure database to predict homomeric structures taking the structure as input. The oligomeric arrangement was imposed to be dimeric. In particular, the MmpL3 models obtained using the Resistance-Nodulation-Cell Division (RND) transporter HpnN (PDB ID: 5KHN and 5KHS) as template were used for further analyses^{75,293}. PDBePISA was used to detect salt bridges and hydrogen bonds²⁹⁴. Computational alanine scanning was performed using the Robetta webserver^{295,296}.

4.2.5 MD simulations settings

AMBER 16 and AmberTools18, including *sander* and the GPU-accelerated *pmemd* code, were used to run MD simulations^{297–299}. The PDB file obtained from CHARMM – GUI was initially processed through a Python script (*charmm lipid2amber.py*) to convert the lipid molecules to be compatible with AMBER^{240,262}. All hydrogens were then removed using *pdb4amber*, for the file to be correctly read by the preparatory program *tLeap*²⁹⁸. Force field ff19SB, lipid17 and tip3p were sourced into *tLeap* to define protein, lipid and water components respectively^{242,300,301}. Finally, topology and coordinate files were successfully generated. MD atomistic simulations of 5 repeats for each model were carried out over a 1.5 μ s time frame, after running two minimisations, two heating steps and a 10 ns equilibration. Explicit solvent and 0.15 M KCl were used to solvate the system. Protein was centred within the simulation box and the particle mesh Ewald (PME) periodic boundary condition (PBC) applied with a cut off of 10 Å for long-term electrostatic calculations²⁹⁹. The two rounds of minimisation were performed using *sander*. For each round, a total 2,500 cycles were carried out starting with deepest descent algorithm and switching to conjugate gradient for the last 1,000 cycles. In the first minimisation step, movements from the protein, excluding modelled residues 353 - 381, were restrained using a harmonic potential with a force-constant of 100 kCal mol⁻¹ Å⁻². This restraint was removed for the second minimisation step, allowing the entire system to move freely. Both minimization steps were performed keeping the periodic boundaries volume constant with no pressure scaling and no SHAKE algorithm applied³⁰².

GPU-accelerated *pmemd* code was used to run all the other steps and the SHAKE algorithm was applied to restrain bonds involving hydrogens, therefore allowing an increase of the time step used to 2 fs. The minimised system was heated through two sequential runs to a 300 K final T. First, the system was heated to 100 K

over 20 ps, using a Langevin thermostat with random seed and a collision frequency of 1.0 ps^{-1} , in a constant volume ensemble (NVE) ³⁰³. During this first heating step a force constant restraint of $10 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ was applied on the protein excluding modelled residues 353 - 381. The system was then heated to 300 K over 100 ps, by using the same thermostat but this time applying an anisotropic Berendsen weak-coupling barostat to also equilibrate the pressure at 1 bar, thus reproducing a constant T and P ensemble (NPT ensemble). A 10 ns equilibration step followed and was carried out at 1 bar and 300 K. Finally, a 1.5 μs MD production run was performed under the same conditions, with the exception of the use of an increased time step, from 2 fs to 4 fs, as Hydrogen Mass Repartitioning (HMR) was applied to the solute counterpart ³⁰⁴.

4.2.6 Analysis of MD trajectories

All trajectories were visualised using the MD movie function in UCSF Chimera, while *cpptraj* was used to perform calculations of RMSd, Root-Mean-Square fluctuations (RMSf) and B-factors ^{284,305,306}. Mass-weighted RMSd were calculated on the protein backbone, including Carbon- α ($\text{C}\alpha$), C, O and N as the selection mask, over the whole trajectory using the first frame of simulation as reference. RMSf and B-factors were calculated over the last 1,000 ns of production run, using the first frame of simulation of the specified timeframe as reference. RMSf and B-factor values across the five repeats for each model were also averaged and plotted using matplotlib Python library. PDB structures for the last 100 ns of simulation were produced for all the repeats and averaged.

MD simulations were also analysed to investigate the changes in the periplasmic cavity size, building on the initial results described in my MRes report ³⁰⁷. $\text{C}\alpha$ distances (in $\text{\AA}$) between relevant residues located within the periplasmic cavity were measured on the averaged PDBs obtained from the last 100 ns and over the whole trajectory using the *cpptraj* function *distance* and compared to those obtained for the same residue pairs in the crystal structures 6OR2 and 6AJF. Violin plots were also generated using matplotlib Python library for comparing distance distributions of 1.16-6OR2 in its apo form to the values obtained from *Msmeg* MmpL3 simulations with bound ligands ³⁰⁸. Furthermore, PND – PCD centre of mass distance distributions were

also calculated for all simulated models and compared to 6OR2 and 7NVH reference structures.

MD trajectories were analysed to identify lipid binding sites on MmpL3. Volumetric maps were generated using the VolMap tool within VMD by calculating lipid headgroup occupancy within 6 Å of the protein over the whole trajectory³⁰⁹. Moreover, cpptraj commands *hbond* and *lifetime* were used to investigate hydrogen bonds between lipid headgroups and protein³¹⁰. Heatmaps were generated using GraphPad Prism version 10.0.0 for Mac OS X (GraphPad Software, Boston, Massachusetts USA, www.graphpad.com).

4.3 Results

4.3.1 Selection of reference crystal structures for *Msmeg* and *Mtb* MmpL3

In 2019, two independent research groups published two crystal structures of engineered versions of *Msmeg* MmpL3, PDB IDs: 6AJF and 6OR2^{1,2}. The atomic models of the two *Msmeg* protein structures were resolved to a similar resolution by X-ray diffraction and were found to be in agreement with a previously determined partial structure⁸⁵. The two structures share 100% identity at sequence level up to aa 748. However, in order to obtain good diffracting crystals both sequences were engineered to remove the C-terminal domain; residues 749 to 1013 were removed for 6AJF, while residues 762 to 1013 were removed for 6OR2. Zhang and colleagues further altered the protein by adding the Enterobacteria phage T4 endolysin sequence (Uniprot ID D9IEF7) at the C-terminus as well as mutating Cys residues 816 and 869¹. Moreover, Zhang *et al.* structure presented a 2 aa linker connecting the MmpL3 sequence to endolysin and two tags, both at the N- and C-termini, covering positions 1 - 13 and 924 - 943. Furthermore, in both structures non-terminal residues 356 - 387 and 353 - 381 were not solved in the Zhang *et al.* and in the Su *et al.* structure, respectively^{1,2}. These residues connect TM 6 to TM 7a and the inability to resolve them was likely due to the high flexibility and/or disordered nature of this region, discussed later in this chapter. Some differences between the two structures emerged also in terms of expression system. In particular, Su *et al.* used *E. coli* as expression system, while 6AJF was recombinantly expressed in *M. smegmatis* strain mc(2)155 cells, possibly favouring a more physiological protein folding as the native organism was used². When superposed, the two crystal structures presented an overall RMSd

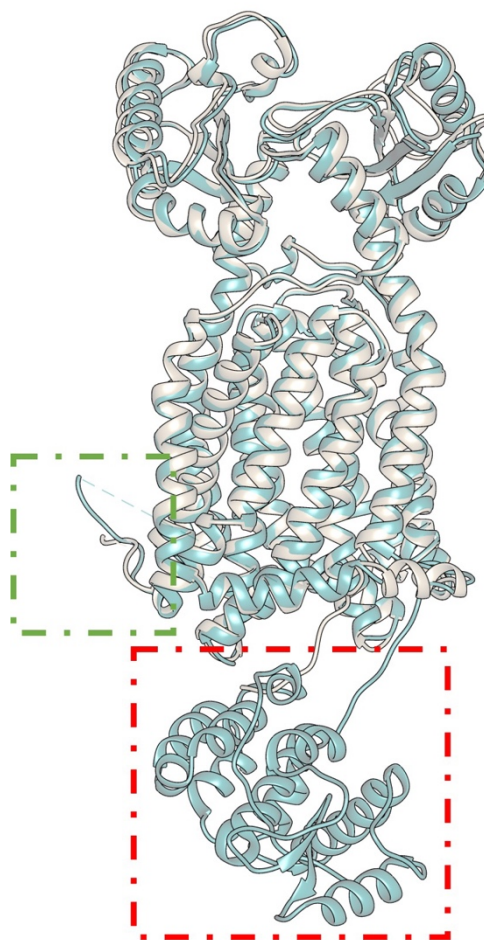


Figure 4.1. 6OR2 and 6AJF superposition. Superposition of 6OR2 and 6AJF with ribbon representation for the two structures in tan and dark cyan, respectively. The red box indicates the T4 endolysin sequence found in 6AJF, while the green box marks the position of the residues missing in both structures (dashed lines).

of 1.095 Å over 709 C α atoms (Figure 4.1). Due to the aforementioned sequence modifications of the Zhang *et al.* structure and for the better resolution obtained by Su *et al.*, 6OR2 was selected to perform all the monomeric simulations of *Msmeg* MmpL3 and analyses presented in this work ^{1,2}.

For *Mtb* MmpL3, a 3 Å cryo-EM structure was published in 2021, PDB ID: 7NVH ⁴. The protein sequence was also engineered and C-terminal domain residues 754 to 944 were truncated. As for the 6OR2 structure, the protein was heterologously expressed in *E. coli* and 35 non-terminal residues (aa 343 - 377) from the loop connecting TM 6 to TM 7 were not solved. 7NVH is the only available structure for *Mtb* protein, therefore it was used for all MD simulations.

4.3.2 *Msmeg* and *Mtb* MmpL3 monomeric models and MD simulation results

The MmpL3 6OR2 structure published by Su and co-workers was used as the reference for the *Msmeg* MmpL3 protein. The first task focused on modelling the 29 residues, located between positions 353 and 381, that were missing from the original PDB. The distance between the last modelled residues at the N'- and C'- termini (Thr352 and Arg382) measured 21.59 Å, a distance unlikely to accommodate a linear peptide, which would measure 95.32 Å. In terms of amino acid composition, which includes 48.28% of hydrophobic residues, the loop is more similar to the TM 1a (1 - 11), TM 7a (383 - 399) and α 8 (740 - 749) regions that do not cross the membrane, presenting altogether a lower percentage of hydrophobic residues compared to the other TM helices (averages of 47.5% and 67.5%, respectively). According to these results, a physiological position for the loop would be at the interface with the cytoplasmic face of the membrane. Secondary structure prediction methods suggest the presence of only short helical regions, connecting residues Pro357 - Leu359 and Gln363 - Thr376, while the rest of the sequence is disordered²⁸⁹. Thirty structures were generated using Modeller within UCSF Chimera and ranked for zDOPE, with scores ranging between -1.32 and -1.59 (average -1.46 ± 0.058)^{284,286}. A more compact loop structure and a shorter distance from the protein core were considered as energetically more favourable models. Furthermore, amino acid specific functions in protein structures were evaluated. Model 1.16 met all these criteria for *Msmeg* MmpL3 and also had Pro357 - Leu359 in a helical conformation, with the proline involved in the alpha-turn as typically found in nature.

Taking all the aforementioned considerations into account, the missing intracellular loop connecting TMN to TMC (aa 353 - 381) was generated through the Modeller software within the UCSF Chimera visualisation tool and the selected model was named '1.16-6OR2' (Figure 4.2, a)^{284,286}.

A similar situation was also observed when the reference structure for *Mtb* MmpL3, 7NVH, was analysed. 35 non-terminal residues in the loop between TMs 6 and 7 (aa 343 to aa 377) were missing in the PDB. The distance between the last modelled residues at the N'- and C'- termini (Leu343 and Arg377) measured 22.73 Å, a distance unlikely to accommodate a linear peptide. Therefore, this unstructured loop was modelled using the same methodology as for *Msmeg* MmpL3 and the best model was selected to carry out MD simulations (Figure 4.2, b)^{284,286}.

a. *Msmeg* model based on 6OR2

b. *Mtb* model based on 7NVH

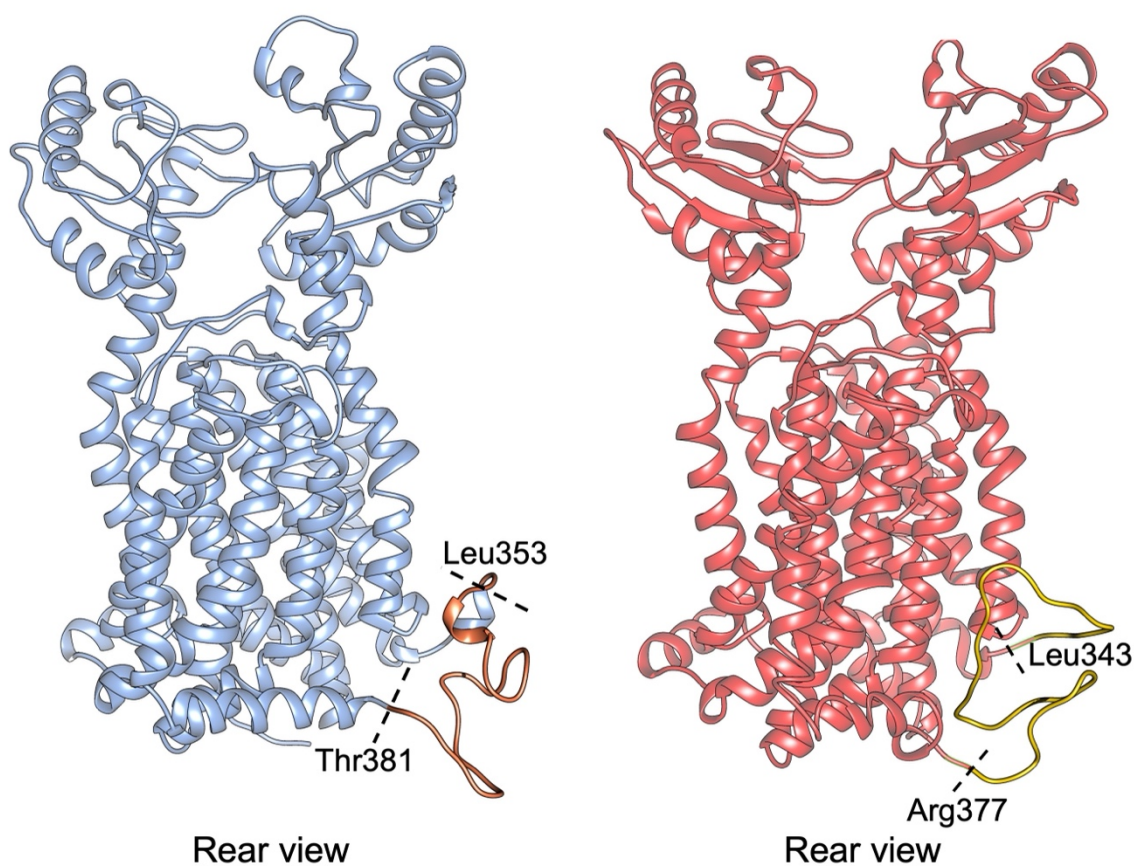


Figure 4.2. Missing residues modelling for *Msmeg* and *Mtb* MmpL3 structures. a. Rear view of *Msmeg* MmpL3 1.16 model (light blue), based on 6OR2 reference structure. Missing residues Leu353 – Thr381 were modelled through the Modeller plugin within UCSF Chimera and are shown in orange. Distance between aa 352 – 382 is 21.59 Å, not allowing for a linear peptide to be inserted in. b. Rear view of *Mtb* MmpL3 model (red), based on 7NVH reference structure. Missing residues Leu343 – Arg377 were modelled as in a, shown in gold. Also in this case, distance between aa 342 – 378 is 22.73 Å, not allowing for a linear peptide to be inserted in.

These PDBs were used to generate a system where the protein was embedded in a POPC lipid bilayer using CHARMM - GUI Membrane builder (Figure 4.3)²⁶². Five independent 1.5 μ s MD runs were performed, for both constructs, according to conditions described in section 4.2.5.

4.3.3 RMSd, B-factors and RMSf analyses on monomeric simulations

Protein C α RMSd values were calculated over the whole trajectory for all repeats to assess protein stability in the system and values compared to the starting structure (first frame). As shown in Figure 4.4 for the 1.16-6OR2 construct, this analysis revealed that RMSd rapidly increased at the beginning of the simulation, meaning that the protein was quickly moving away from the rigid initial pose, as expected. After about 0.5 μ s into the simulation, an equilibrium was reached and

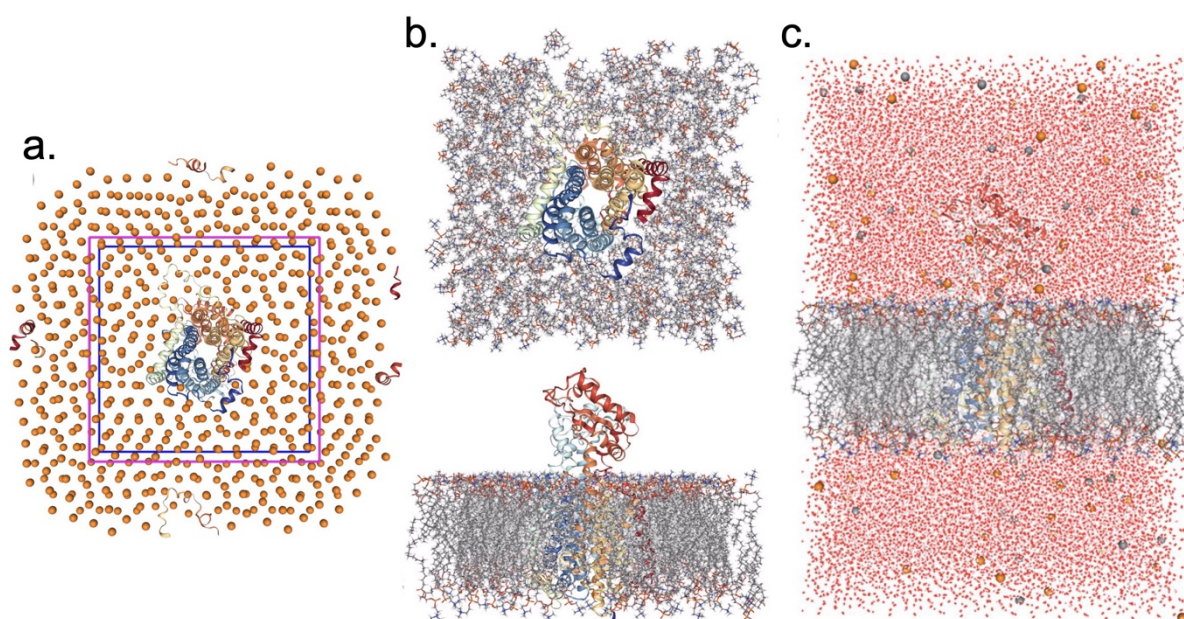


Figure 4.3. Membrane assembling using CHARMM-GUI Membrane Builder. a. Lipid headgroups are represented as orange spheres for the user to have a preliminary idea of lipid packing and box size. Spheres are then replaced in b. by proper lipids, reflecting the membrane composition selected by the user. Finally in c. all components, including water molecules (red) and ions (floating spheres) are assembled to generate a PDB file.

curves plateaued at around 2.5 - 3.5 Å, suggesting that the system had achieved a stable state (Figure 4.4). This allowed meaningful analyses on the rest of the trajectory. First, B-factors were calculated for each of the 761 $C\alpha$ of *Msmeg* MmpL3 apo over the last 1 μ s of simulation for all repeats and plotted alongside the values deposited for the reference PDB file. As B-factors measure the amplitude of atomic oscillations around their equilibrium position, higher values indicate that the electron density map is more difficult to calculate³¹¹. As shown in Figure 4.5 a, B-factors values for the 1.16-6OR2 MD runs were consistently lower than the PDB ones. Larger values were only recorded for the periplasmic domains (PND and PCD) as well as for the modelled loop, all regions that have a high freedom of movement. Overall, the atomic positions were identified with much higher accuracy by the MD simulations compared to the starting crystal structure, especially for the TM regions. Furthermore, RMSf analyses for the same construct confirmed the results obtained from B-factors analysis for the specific protein residues that fluctuated the most from their initial position during MD simulations. Through this analysis the modelled loop and the two periplasmic regions were once again identified as highly dynamic regions, showing the highest RMSf values (Figure 4.5, b).

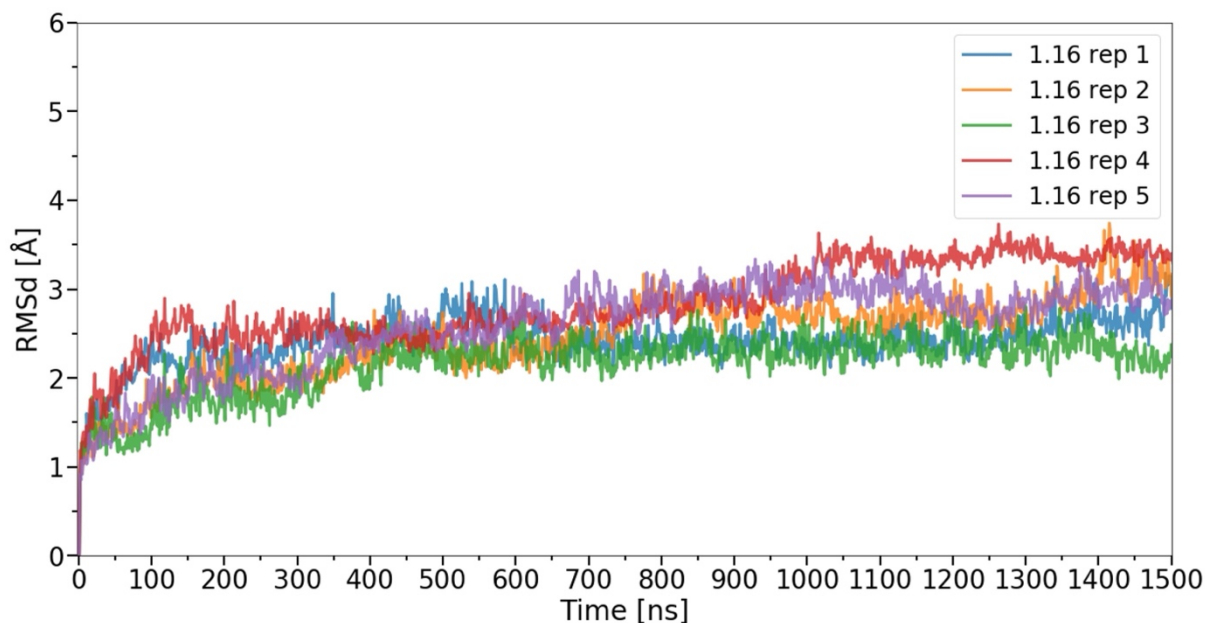


Figure 4.4. RMSd plots for 1.16-6OR2 repeats. Graph shows results from RMSd analyses performed over 1.5 μ s of trajectory for the 1.16-6OR2 apo model. Different colored lines are representative of the five repeats. Every frame in the production run is compared to the first frame of trajectory and RMSd are only calculated for protein backbone, including C α , C, O and N atoms. After an initial steep curve over the first 100 ns, where the protein is still finding its stability within the membrane, the repeats get to an equilibrium at around 500 ns, with only minimal RMSd value oscillation over the last 1 μ s. MD simulation time in nanoseconds (ns) is reported on the x axis, while RMSd values in Angstroms (Å) are shown on the y axis.

The same analyses were also performed on the *Mtb* MmpL3 apo simulations, and similar conclusions to those previously reported could be drawn: 1) B-factors for the simulated model were consistently lower than reference PDB, with the only exceptions being the modelled loop and the periplasmic regions; 2) when RMSf were performed across the 5 repeats, residues within the modelled loop and the two periplasmic regions recorded the highest oscillation values (Figure 4.6, a and b).

In particular, residues 76 - 78, 116 - 120 and 162 - 165 presented the highest RMSf peaks for the PND domain in the 1.16-6OR2 model. Finally, RMSf values were the highest for residues 464 - 470, 490 - 496, 504 - 510 and 537 - 542, located in the PCD domain. Large fluctuations were also observed for aa 750 - 761 but were ignored as these atomic movements were probably caused by random forces acting on the artificially generated C-terminus. All repeats consistently identified the most flexible protein residues within these regions, suggesting highly reproducible protein dynamics.

After structural analysis, it was possible to conclude that all the PND and PCD residues showing high RMSf values were mostly unstructured connecting loops. Given

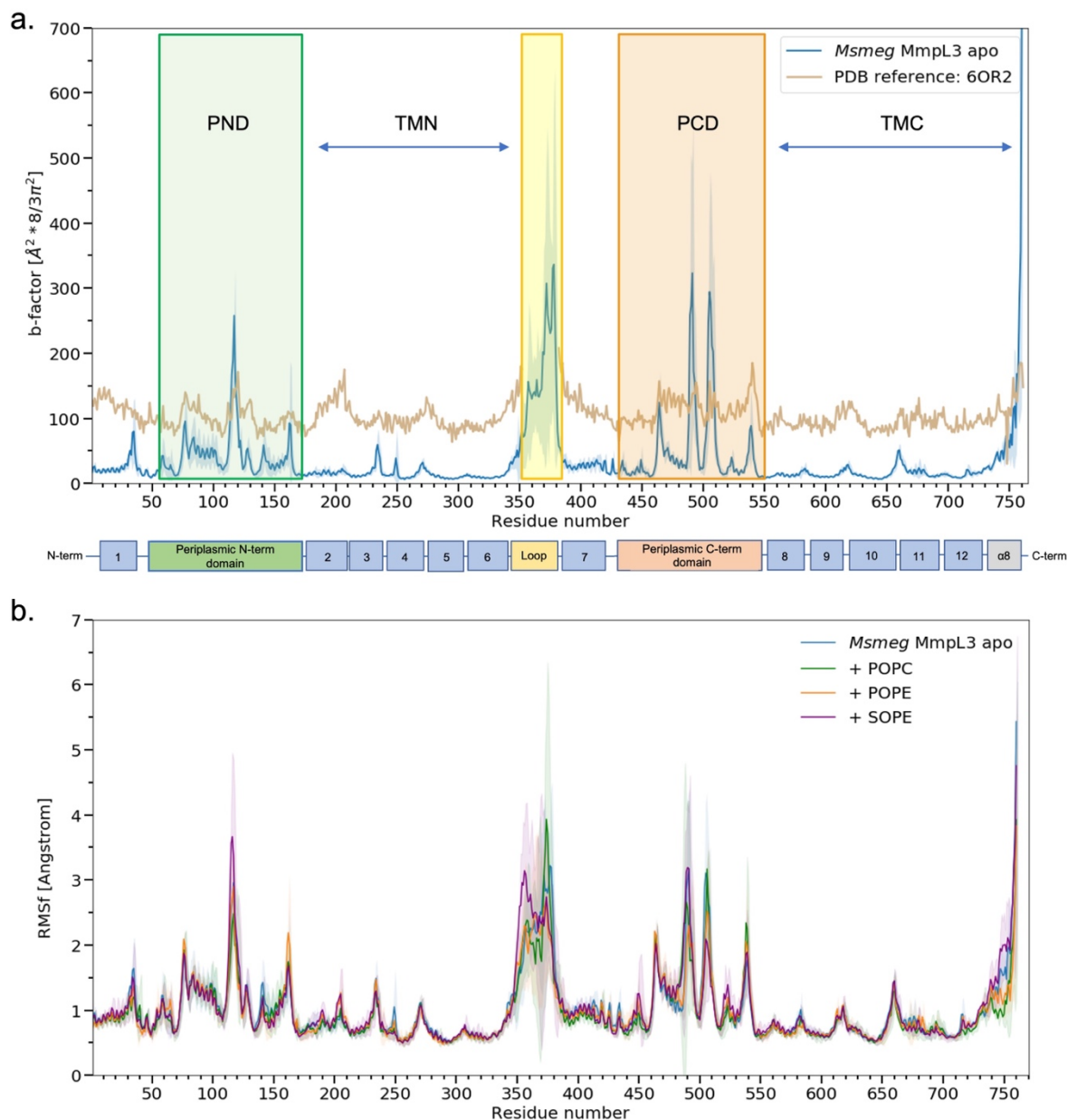


Figure 4.5. B-factors and RMSf analysis for *Msmeg* MmpL3 a. B-factors calculation over 5 repeats for simulated apo *Msmeg* protein (blue line) and for the 6OR2 reference structure (orange line). B-factors values for aa 353 – 381 are not available for 6OR2 (line gap) as these residues were not solved in the PDB. B-factors were calculated on the protein backbone (761 C α) over the last 1 μ s of simulation. Average values are indicated by the blue solid line, while the range across all 5 repeats is represented by blue shaded area. Highest values were recorded for the periplasmic regions (in green and orange boxes) and for the modelled loop (yellow box). Transmembrane (TM) regions represented with blue arrows including TM numbering. TM 1 - 6 indicate the N-terminal TM domain (TMN), while TM 7 - 12 the C-terminal TM domain (TMC). b. RMSf comparison between *Msmeg* apo (blue line) and ligand-bound protein over the last 1 μ s of simulation. This analysis identified the same regions as the most dynamic and flexible during MD simulations as shown in a. The highest fluctuations values are given in $\text{\AA}$.

the 2-fold protein symmetry, the identified flexible regions on one periplasmic domain were mirrored on the other, as shown in Figure 4.7. Of particular interest, given their potential role in substrate translocation, are PND residues 116 - 120 located in close proximity to regions on the PCD domain between aa 490 - 496 and 504 - 510,

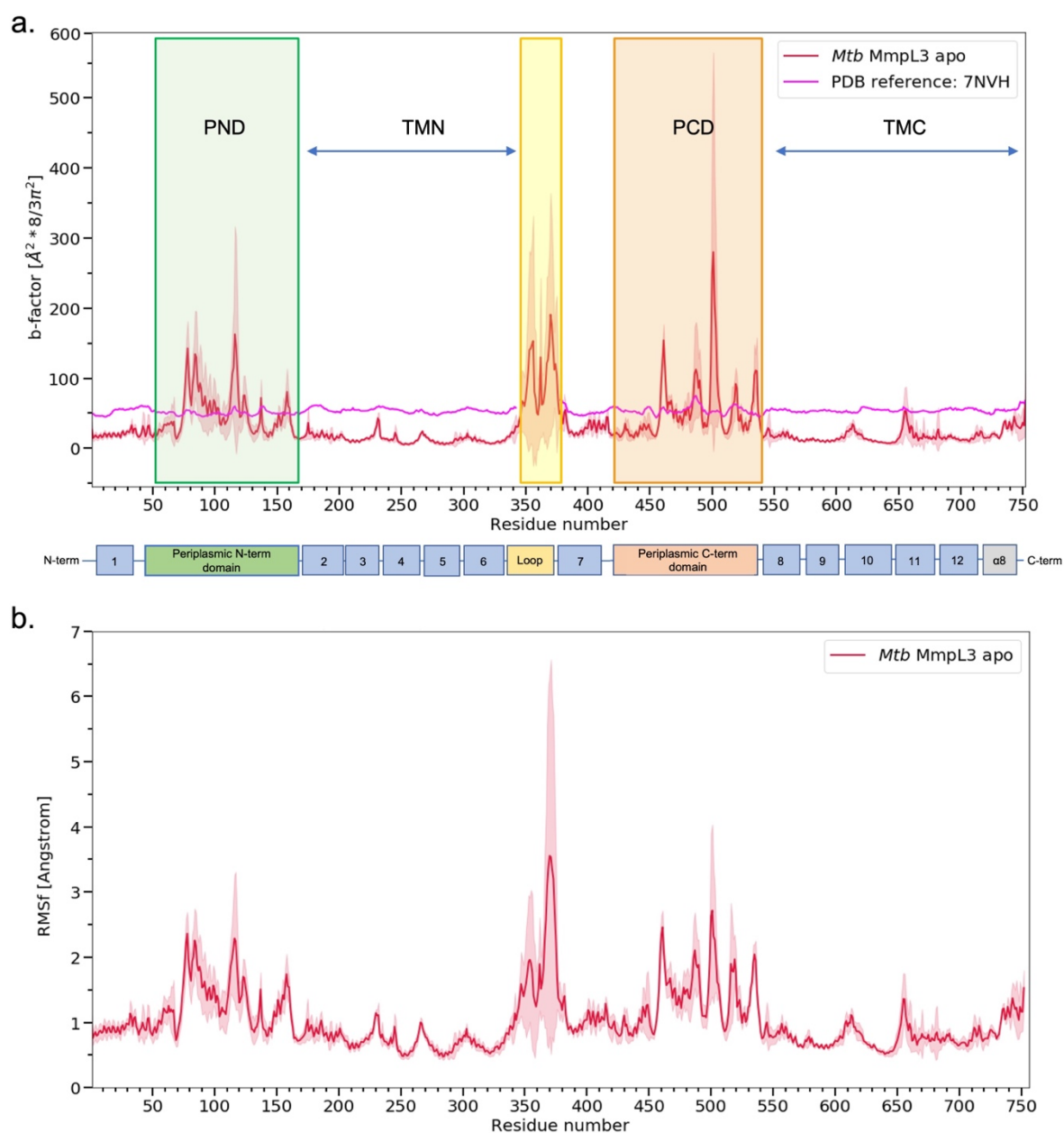


Figure 4.6. B-factors and RMSf analysis for *Mtb* MmpL3. a. B-factors analysis performed on simulated apo *Mtb* MmpL3 reported as average values (magenta solid line) and range across all 5 repeats in shaded area. Values for reference PDB 7NVH are shown with purple line (aa 343 – 377 missing). b. RMSf values calculated for the simulated apo *Mtb* protein over the last 1 μ s of simulation, with same colouring as in a.

potentially controlling access of the TMM substrate molecules to the central cavity. Su and colleagues copurified a molecule of PE 16:0 / 17:1(9Z) visible in their X-ray crystal structure bound within the same region of the protein, confirming the importance of studying the structure and behaviour of the central cavity ².

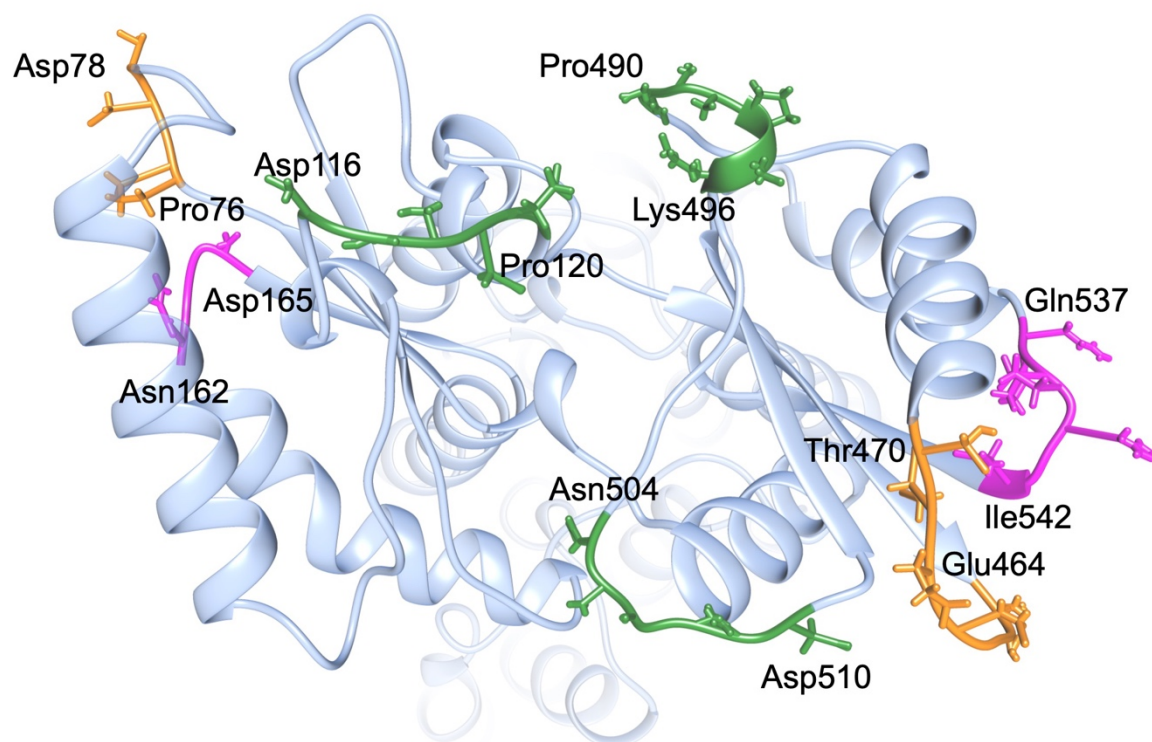


Figure 4.7. 1.16-6OR2 residues with highest RMSf values. Top view of 6OR2 structure. Residues 76 - 78 connect β 1 to α 2 in the PND domain, like residues 464 - 470 link β 4 to α 6 in the PCD region (orange). Residues 162 - 165 in the PND region bridge α 4 and β 3, while aa 537 - 542 α 7 and β 6 in the PCD domain (magenta). All the regions mentioned above are located on the opposite side of the two periplasmic domains thus they do not come into contact during the simulation. PND residues 116 - 120 bridging α 2 and α 3, however, are located in close proximity to the flexible PCD regions between 490 - 496 and 504 - 510, which connect α 6 to β 5 (green).

Although RMSf calculations provide useful insights into fluctuations per residue over the full length of the simulation, they do not provide any details on the direction of movement of specific regions of the protein. However, by visual inspection a possible closure of the central funnel-shaped pore was observed, with the two periplasmic regions that appeared to be moving towards each other, reducing the space available for drugs or other substrates to reach the cavity. In contrast with the crystal structure, where a molecule of PE was co-crystallised in the cavity therefore representing the bound or holo state of MmpL3, the simulations were performed without ligand. Therefore, MD results seem to reflect the protein changing conformation from its bound to its apo state. In the apo state the two periplasmic domains reduce the central cavity space as no substrate needs to be accommodated.

4.3.4 Periplasmic cavity size changes according to protein-ligand binding state

The size of the cavity pore was further investigated by comparing the apo 1.16-

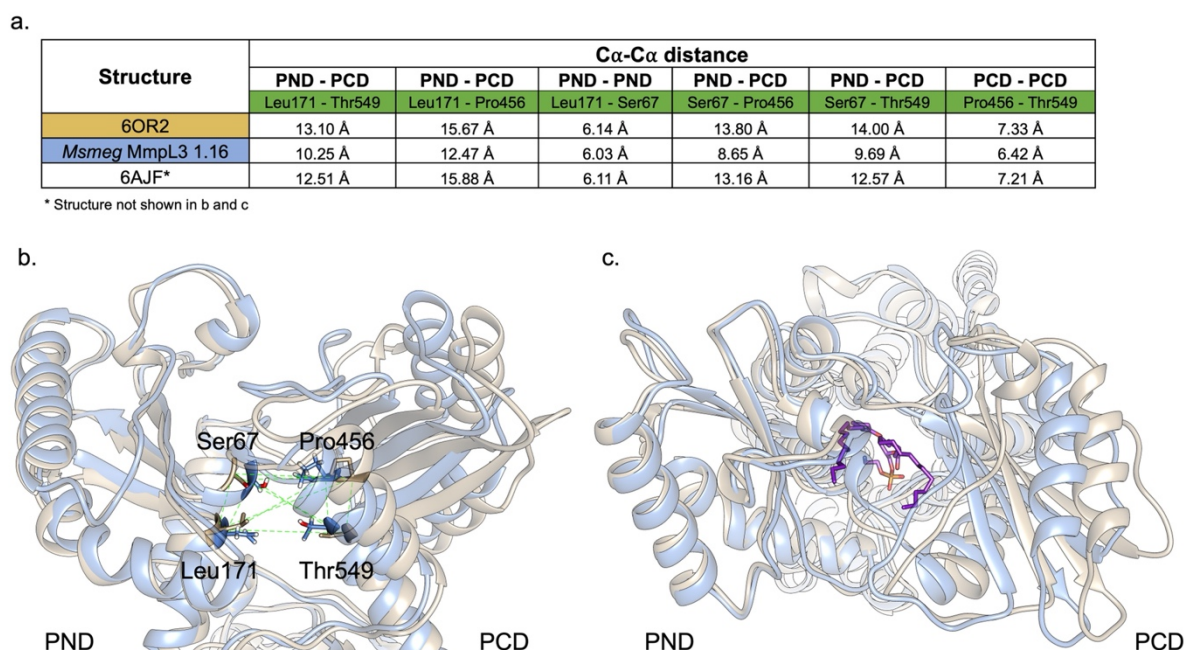


Figure 4.8. *Msmeg* MmpL3 1.16 cavity size comparison to crystal structure. a. Table summarising C α - C α distance measurements for 4 residues pairs located within the periplasmic cavity of crystal structures 6OR2 and 6AJF, and 1.16 average model PDB over the last 100 ns of simulation. Ser67 and Leu171 are located on the PND domain, while Pro456 and Thr549 on the PCD domain. PND - PCD distances showed lower values compared to the reference structure, while PND - PND and PCD - PCD distances remained almost unaltered. b. Front-view superposition of 6OR2 (tan) and 1.16 model average PDB over last 100 ns of simulation (light blue), with atoms shown (sticks) for residues used for distance measurements (dashed green lines), with values reported in a. c. Top-view superposition as in b, including PE molecule (purple) co-crystallised with protein in 6OR2, which cannot be accommodated in simulated model without clashes. As evident from all results shown, cavity size decreased in *Msmeg* MmpL3 simulated model, with the PCD domain moving towards the PND domain suggesting a closure of the central cavity pore. These results could also be the consequence of the absence of ligands in the simulations, showing the protein conformation in its apo form.

6OR2 protein to simulations with ligands inserted in the periplasmic cavity. Four residues located inside the central cavity and at the interface between the two periplasmic domains were selected, 2 in the PND (Ser67 and Leu171) and 2 in the PCD (Pro456 and Thr549). C α distances were calculated for the 6OR2 and 6AJF structures as well as for the PDB obtained averaging the last 100 ns of the 5 repeats for the *Msmeg* MmpL3 1.16 apo model (Figure 4.8, a). Results revealed a reduction in the linear distance between opposite-sided residues compared to the bound reference structure, while the PND - PND and PCD - PCD pair distances remained unaltered. Compared to 6OR2 and 6AJF reference structures, a smaller cavity size was measured as a possible consequence of the closure movement of the PND and PCD regions. This was also observed when the average PDB for *Msmeg* MmpL3 1.16 was superposed to the reference structure 6OR2, as the PCD domain appeared to move closer to the PND, reducing the size of the central cavity (Figure 4.8, b).

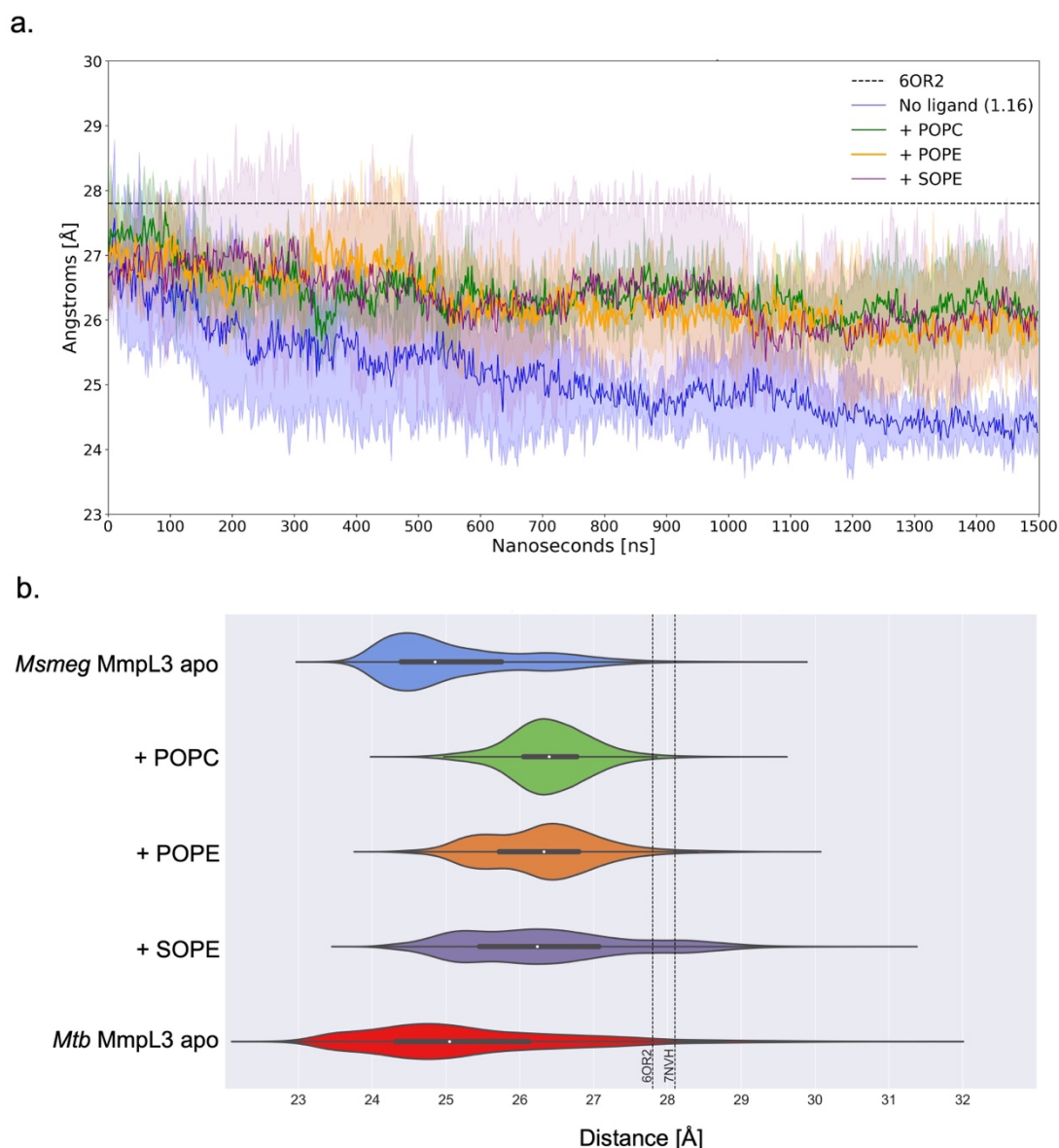


Figure 4.9. PND – PCD centre of mass (c.o.m.) distance measurements during MD simulations. a. PND – PCD c.o.m. distance measurement along trajectory files, compared to 6OR2 (dashed line). C.o.m. distance average values represented in solid lines with range over 5 repeats in shaded area. *Msmeg* MmpL3 apo in blue, *Msmeg* MmpL3 with POPC in green, with POPE in orange and with SOPE in purple. b. PND – PCD c.o.m. distance distribution for all simulated systems represented with violin plots. 6OR2 and 7NVH reference measurements were 27.8 Å and 28.1 Å, respectively (dashed lines). Colours as in a, and *Mtb* MmpL3 with no ligand bound in red. White dots in each violin plot represent median values.

Therefore, the PE molecule co-crystallised in 6OR2 could not be accommodated within the simulated unbound model without causing steric clashes (Figure 4.8, c).

However, when the protein was simulated in its holo state, the closure movement of the two periplasmic domains was partially reduced and the central cavity appeared to be in an ‘open state’. First, the centre of mass (c.o.m.) of each periplasmic domain was calculated and the distance between these two artificial points was then calculated along the whole trajectory and compared to the crystal structure (Figure

4.9, a). PND – PCD c.o.m. distance for the MD simulations including ligands was consistently higher than that recorded for the unbound state, although lower than the distance between the two regions in the crystal structure. The first observation confirmed the hypothesis of a conformational change reducing the distance between the two periplasmic domains of the protein in the unbound/closed- vs bound/open-state.

In contrast, the reduced PND – PCD c.o.m. distance observed in all simulated ligand-bound systems compared to 6OR2 could be the consequence of the different parameters used in MD simulations and for X-ray crystallography. For example, the temperature used for MD runs (300 K) reflects a more physiological condition than the cryogenic settings used in crystallographic studies (100 K), where overall atomic motions are reduced affecting protein thermal flexibility and dynamics.

Density distributions for PND – PCD c.o.m. measurements are shown in Figure 4.9, b as violin plots. The reduction in central cavity size observed for *Msmeg* MmpL3 apo was also observed for the *Mtb* MmpL3 protein in its apo form, with a PND – PCD c.o.m. median distance distribution over the 5 repeats of ~25 Å. Moreover, this same simulated model presented a lower value compared to the reference structure, 7NVH, as already observed for the *Msmeg* protein.

From these results, the reduction in size of the central cavity was confirmed in a quantitative manner both when the averaged structures of the last 100 ns of the simulated models were considered as well as along the whole trajectory. Altogether these observations confirmed that MD simulations were capable of sampling different conformational states of MmpL3, reproducing either an unbound or a ligand-bound state.

4.3.5 Analysis of protein-lipid substrate interactions within the central periplasmic cavity

All solved structures for *Msmeg* MmpL3 showed the presence of specific electron densities relative to detergents or lipid molecules in the periplasmic cavity. In the 6AJF structure 2 molecules of 6-DDTre were co-crystallised with MmpL3, while a molecule of 16:0-17:1(9Z) PE was identified in the 6OR2 structure (Figure 4.10, a and b) ^{1,2}. More recently, in the 7N6B structure published by Su and colleagues a molecule of TMM was also observed, confirming the ability of the periplasmic cavity to actively

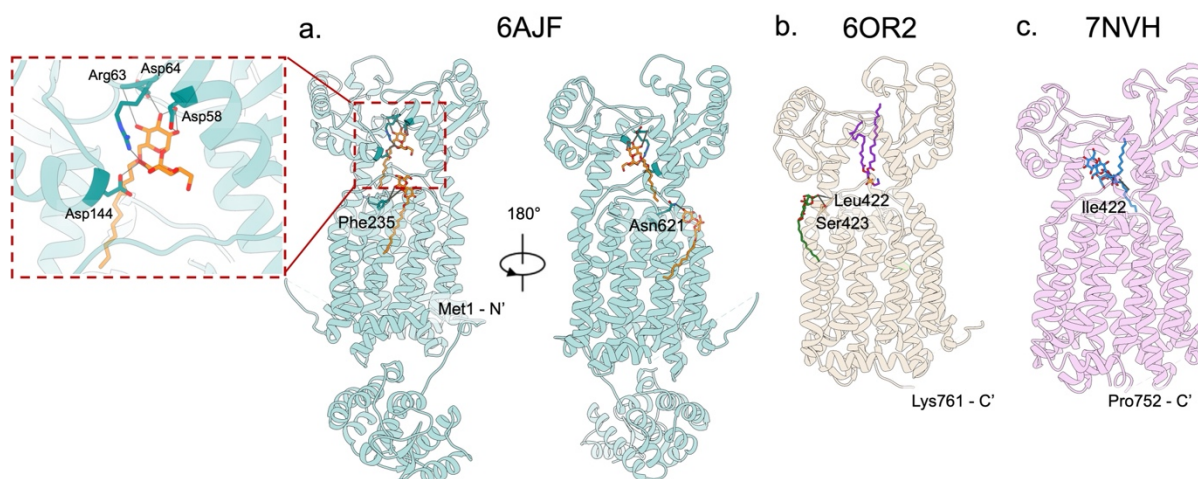


Figure 4.10. 6AJF, 6OR2 and 7NVH structures with co-crystallised ligands. a, 6AJF (*Msmeg* MmpL3), ribbon representation (cyan). Left, front view with 2 molecules of 6-DDTre (orange) identified in the periplasmic cavity (with zoomed area in red box) and 1 molecule of 6-DDTre interacting with Phe235 in binding pocket formed by TM 3 and 6. Right, rear view showing 1 molecule of 6-DDTre interacting with Asn621. b, Rear view of 6OR2 (*Msmeg* MmpL3) in ribbon representation (tan), with 1 PE molecule (purple) found in the central cavity and 1 molecule of DDM (dark green) in pocket formed by TM 7 and TM 10 forming hydrogen bonds with Leu422 and Ser423. c, Rear view of 7NVH (*Mtb* MmpL3) in ribbon representation (pink), with 1 LMNG molecule (blue), interacting with bottom of periplasmic cavity residue Ile422.

bind ligands³. The same was observed for the *Mtb* MmpL3 cryo-EM structure, 7NVH, where a molecule of LMNG was observed bound within the cavity of the periplasmic domain (Figure 4.10, c)⁴. Analyses carried out on these structures revealed that a few residues within the pore form hydrogen bonds with the ligands and some are common across the different structures: Asp58 and Asp144 (Asp139 in 7NVH) for 6AJF and 7NVH; Gln40 for 6OR2 and 7N6B; Ile422 in 7NVH; Arg63 and Asp64 in 6AJF only; Tyr44, Gln442 and Glu455 in 7N6B only. Interestingly, some of these residues were also observed to form H-bonds in the simulations performed on *Msmeg* MmpL3 (1.16-6OR2 as starting structure) with the 3 different lipid molecules. In particular, Gln40, Tyr44, Arg63 consistently interacted with PC and PE lipid headgroups. Furthermore, Ser41, 54 and 67 on the PND domain and Arg501 and Thr549 on the PCD domain were observed to form stable and long-term interactions with the three ligands (Figure 4.11, a). Together, these residues create a network that contribute to the retention of the lipid molecules within the periplasmic cavity throughout the full length of the simulation (Figure 4.11, b). It is worth noticing that during the simulations all ligands interact with a wide range of residues located at different heights within the periplasmic cavity that extends by ~40 Å into the periplasmic space. This not only confirms that MmpL3 is able to bind different lipid species but also that the central cavity potentially acts as a channel through which substrates get transported into the periplasm.

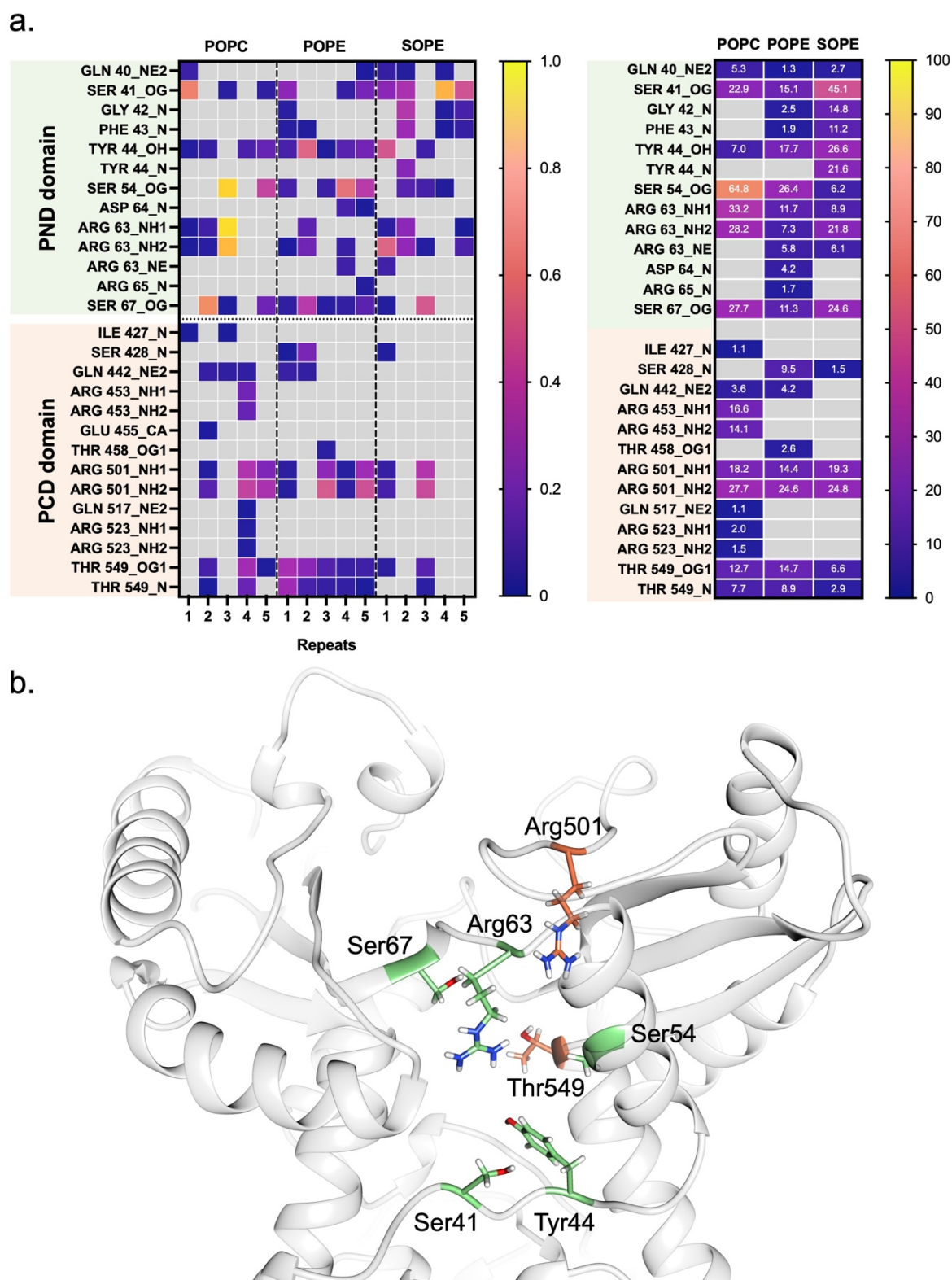


Figure 4.11. *Msmeg* MmpL3 simulations including POPC, POPE and SOPE as ligands: H-bonds analysis results. a. Left, heatmap reporting ligand-protein residues interaction per repeat during 1.5 μ s simulation as fraction of total frames. Right, heatmap with averaged values of ligand-protein residues interactions as % of total trajectory frames, for each ligand. Protein residues from the PND and PCD domains are highlighted in the green and orange boxes, respectively. Ser41, Tyr44, Ser54, Arg63 and Ser67 of the PND domain and Arg501 and Thr549 of the PCD domain form stable H bonds with all three ligands. b, Front view of 1.16-6OR2 structure with main residues from PND (green) and PCD (orange) domains involved in H-bonds formation with the side chains in stick representation.

4.3.6 Analysis of protein-lipid binding sites on TM domains

In addition to the molecules identified in the protein central cavity, a few detergent molecules were also found at the interface with the periplasm. The hydrophobic chains mainly interact with the TM regions, while the headgroups interact with residues in the periplasmic domain. In particular, a molecule of DDM / TMM was resolved in the pocket formed by TM 7 – TM 10 in two Su *et al* structures, interacting with two leucine residues, Leu422 and Leu424, and Ser423 (Figure 4.10, b) ^{2,3}. Furthermore, in the 6AJF structure one molecule of 6-DDTre formed hydrogen bonds with Asn621 located on the loop connecting TM 9 to TM 10 while protruding towards the binding pocket formed between TM 2 and TM 11, while another detergent molecule interacted with Phe235 in the pocket between TM 3 and TM 6 (Figure 4.10, a).

The same lipid binding sites were identified in the simulated models of monomeric *Msmeg* and *Mtb* MmpL3, together with some additional sites, that can be subdivided into 6 pockets (Figure 4.12, a). The interaction between lipid molecules on the upper leaflet of the bilayer and protein residues on TM helices involved the formation of hydrogen bonds that were consistent among repeats for *Msmeg* and *Mtb* MmpL3 (Figure 4.12, b and c; Figure 4.13, a and b). Only protein residues forming H-bonds with lipids located in the upper leaflet for a percentage of total frames > 10% were considered as relevant and used for the results presented in this section.

In particular, in the *Msmeg* MmpL3 apo MD simulations, for the lipid substrate binding pocket between TM 7 – TM 10 found in the 6OR2 and 7N6B structure (zone 1), Ser423 consistently formed H-bonds with one PC headgroup across 4/5 repeats, with an average of 40.3% of the total frames ^{2,3}. In the same binding pocket, Leu424 and Gln554 were also observed to interact with PC residues in two separate repeats although with a lower percentage of total frames, of 12.5 and 18.8, respectively. During MD simulations, the PC headgroup moved towards the periplasmic space, sticking out from the membrane level, visible in the occupancy map calculated over the 5 repeats (Figure 4.12, d). Gly616, Ser617 and Gly618, with a minor contribution of Trp609, formed the 2nd binding pocket and interacted with a specific PC headgroup (PC 766) in 4 of the 5 repeats, covering 64.2% of the total trajectory. In binding pocket 3, protein-lipid hydrogen bonds involved Asn621 in all repeats over 26% of total frames, and Arg188 with a % of total frames ranging between 23.1 and 36.9. Residues in binding pocket 4, including Ile32, Gln186 and Gln243, recorded a low % of total interaction

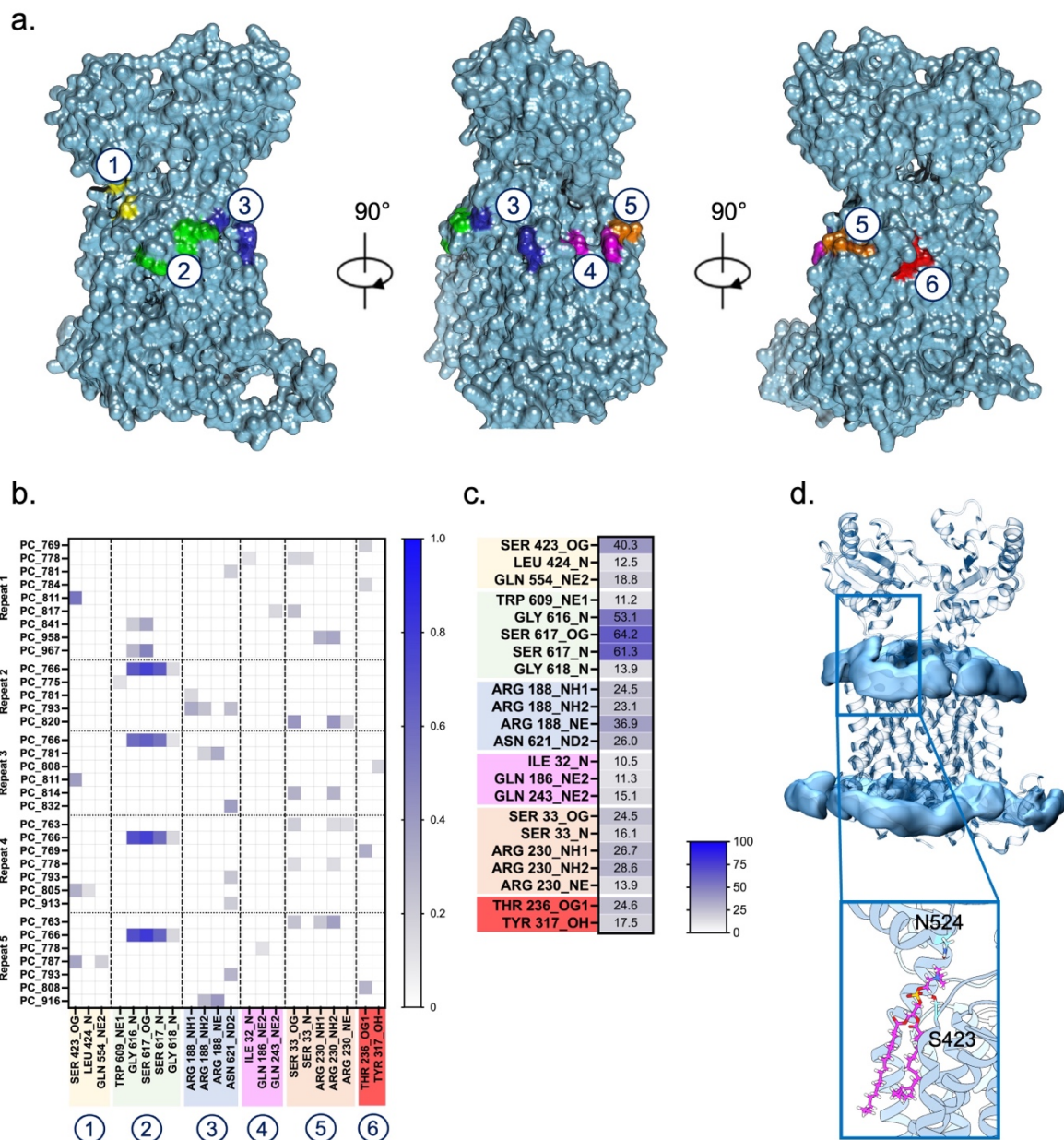


Figure 4.12. MmpL3 lipid binding pockets. a. Main lipid binding pockets identified 1 – 6 starting from the rear of the protein (left) to front view (right). b-d. *Msmeg* MmpL3 protein-lipid hydrogen bonds analysis. b. Heatmap with H-bonds results visualised per repeat with PC molecules on y axis and protein residues on x axis and same colouring scheme for binding pockets as in a. Values range from 0 to 1, with 1 indicating the bond is present over the whole simulation length. c. H-bonds average results summary with max value of 100 % of total frames bond is present. d. Averaged lipid occupancy map (rear view) generated in VMD selecting PC molecules within 6 Å from protein along the whole trajectory. Zoomed in box shows protein residues forming a 'gate' for lipid entrance in the cavity, with POPE molecule in purple shifting upwards from membrane plane.

and only in 2 repeats. Ser33 and Arg230, forming binding pocket 5, also showed H-bond formation over the 5 repeats with a % of total frames reaching 28.6%. Finally, in the 6th binding pocket identified, H-bonds were observed between PC residues and Thr236 and Tyr317 (Figure 4.12, b and c). The H-bonds analyses performed on this model revealed important protein residues involved in lipid-protein interaction that

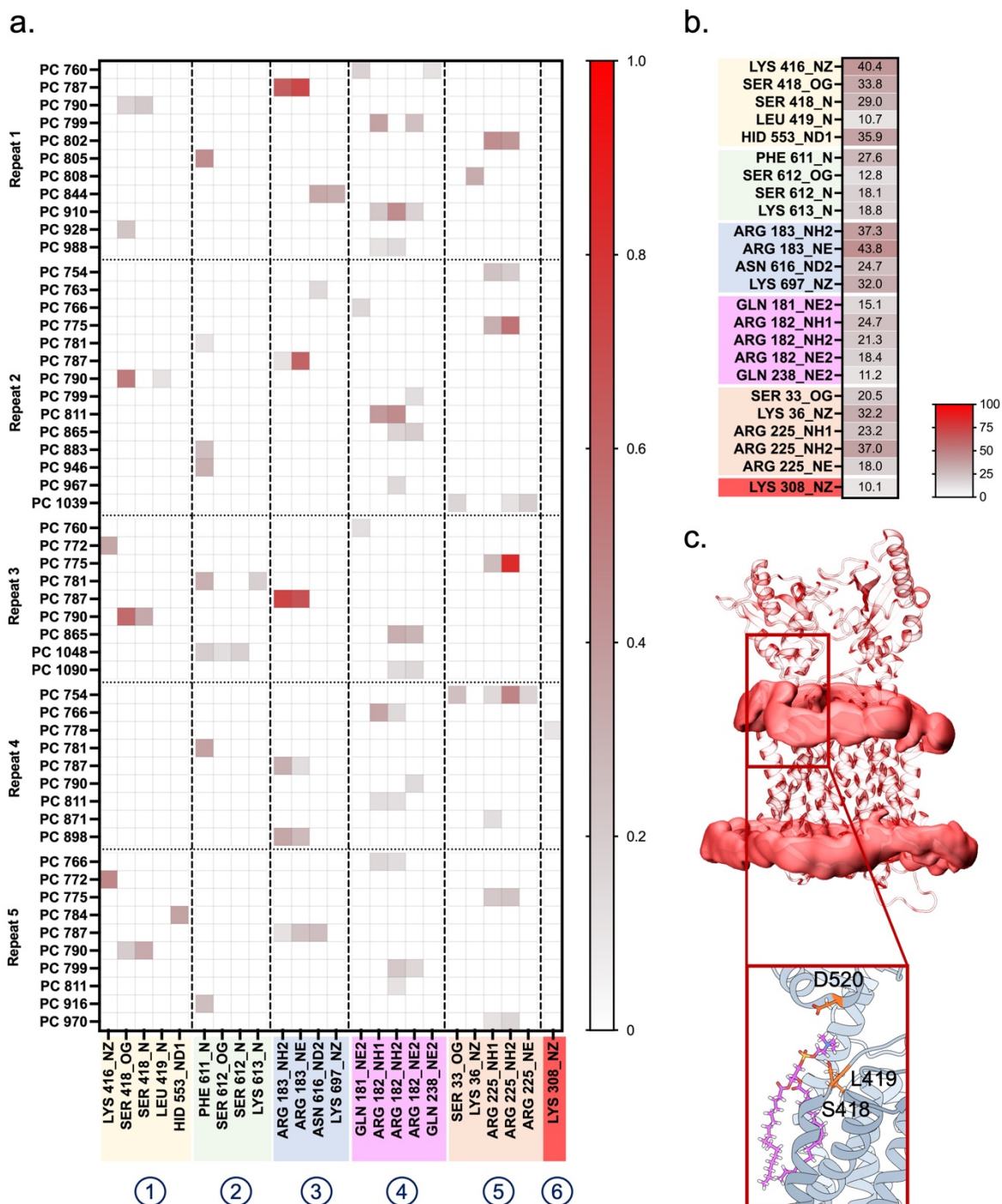


Figure 4.13. *Mtb* MmpL3 protein-lipid hydrogen bonds analysis. a. Heatmap with H-bonds results visualised per repeat with PC molecules on y axis and protein residues on x axis and same colouring scheme for binding pockets as in Figure 4.12, a. Values range from 0 to 1, with 1 indicating the bond is present over the whole simulation length. b. H-bonds average results summary with max value of 100 % of total frames bond is present. c. Averaged lipid occupancy map (rear view) generated in VMD selecting PC molecules within 6 Å from protein along the whole trajectory. Zoomed in box shows protein residues forming a 'gate' for lipid entrance in the cavity, with POPC molecule in purple shifting upwards from membrane plane.

matched the sites occupied by the 6-DDTre detergent molecules in the Zhang *et al* structure (zone 3 and zone 6) ¹.

When these analyses were performed on the *Mtb* protein, the same lipid binding pockets were identified, with some residues matching the *Msmeg* ones. In particular, Ser418 (Ser423 in the *Msmeg* model) formed hydrogen bonds in 4 of the 5 repeats, reaching on average a 31.4% of total frames. The same behaviour of the PC headgroup protruding out of the membrane plane already observed for *Msmeg* was also present in the *Mtb* simulation, as evident from the lipid occupancy map calculated for this construct (Figure 4.13, c). In binding pocket 1 Lys416, Leu419 (Leu424 in the *Msmeg* model) and His553 formed interactions. In the 2nd binding pocket mainly Phe611, but also Ser612 and Lys613 (Gly616, Ser617 and Gly618 in the *Msmeg* model) were involved, although with a much lower total % of frames compared to the *Msmeg* model. Nevertheless, similar results between the two proteins were observed when binding pocket 3 was analysed. Arg183, Asn616 (Arg188 and Asn621 in *Msmeg* MmpL3) and Lys697 all contributed to the formation of stable hydrogen bonds with different PC groups across all repeats. Three glutamine residues, Gln181, Gln182 and Gln238 (Gln186 and Gln243 in *Msmeg* model) also significantly contributed with a percentage of total frames ranging between 18.0 and 37.0 % and forming another lipid binding pocket (zone 4). Furthermore, Ser33 (same residue in *Msmeg*) and Arg225 (Arg230) constituted pocket number 5, while Lys308 was the only residue identified as part of pocket 6 (Figure 4.13, a and b).

Altogether, the MD simulations results showed in this section were in agreement with the lipid binding sites observed by MmpL3 experimental structural determination. Furthermore, several new potential TM binding pockets were also identified in this study that could be important for protein stability in the membrane environment.

4.3.7 Investigation into MmpL3 oligomerisation state

As discussed in section 4.2.4, protein structures for *Msmeg* (PDB ID: 7N6B) and *Mtb* (PDB ID: 7NVH) MmpL3 were used as input to generate template-based oligomeric models through the GalaxyHomomer server. As the experimental evidence suggested the dimer as the likely MmpL3 oligomerisation state, a 2-mer arrangement was imposed, results shown in Table 1. For both protein species, the predicted model used the dimeric RND hopanoid transporter HpnN from *Burkholderia multivorans* as template (PDB IDs: 5KHS and 5KHN)⁷⁵. For each construct, the model with the largest

Table 1. MmpL3 oligomeric structures predicted by GalaxyHomomer server

<i>M. smegmatis</i> MmpL3 (PDB 7N6B)						
Model no.	Template (PDB)	Oligomer Template	No. of subunits	Interface area (Å ²)	Sequence identity	Structure similarity (TM-score)
1	5KHS	HpnN	2-mer	2720.7	11.2	0.7135
2	5KHN	HpnN	2-mer	2444.6	11.6	0.7092

<i>M. tuberculosis</i> MmpL3 (PDB 7NVH)						
Model no.	Template (PDB)	Oligomer Template	No. of subunits	Interface area (Å ²)	Sequence identity	Structure similarity (TM-score)
1	5KHS	HpnN	2-mer	1920.5	11.6	0.7196
2	5KHN	HpnN	2-mer	2501.8	11.6	0.7151

interaction interface area was selected and was superimposed to the respective template PDB. Superimposition of *Msmeg* MmpL3 and 5KHS gave a RMSd value 2.29 Å over 474 atoms (Figure 4.14). When the *Mtb* model was superimposed to 5KHN, a RMSd value of 2.36 Å over 465 atoms was obtained. While the TM domains of HpnN and MmpL3 showed the highest structural alignment, the architecture of the periplasmic domains differ substantially, as HpnN has 2 additional domains extending further away from the membrane plane towards the periplasmic space. Furthermore, the dimer interface identified for the MmpL3 proteins differ from HpnN, as discussed in the next section. Nevertheless, valuable insights could still be gained by comparing the two proteins. As the two MmpL3 proteins have been shown to exert the same lipid transport function in *Msmeg* and *Mtb* and share a 72% sequence identity over the truncated protein sequences (aa 1 – 773 for *Msmeg* MmpL3 and aa 1 – 753 for *Mtb* MmpL3) structurally determined, MD simulations were only run using *Msmeg* dimeric model as reference. This choice seemed sensible in light of the huge computational cost associated with the simulation of such complex systems, including multiple components and a large number of atoms.

4.3.8 Dimer interface analysis on GalaxyWeb models

Starting from the dimer model obtained using GalaxyWeb, an extensive analysis of the interface was conducted. In the HpnN structure, the transmembrane

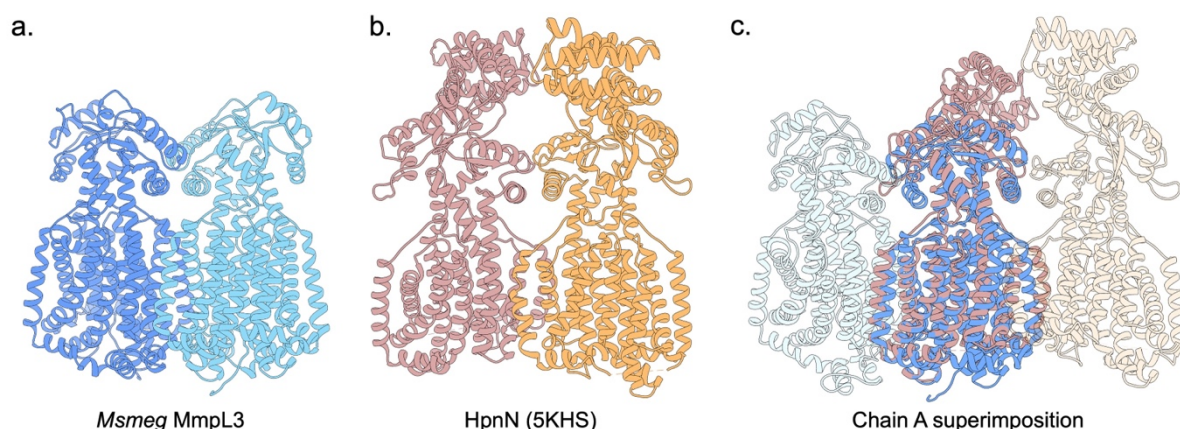


Figure 4.14. GalaxyHomomer predicted *Msmeg* MmpL3 dimeric structure. a. Front view of *Msmeg* MmpL3 model 1 in ribbon representation and coloured by chain. b. Front view of template protein HpnN (PDB: 5KHS). c. Superimposition of chain A from both structures reveals regions involved in dimer interface differ for the two structures.

helices TM 7 and TM 9 and the periplasmic helices α 6 and α 7 were identified as important regions for protomer interactions by the MRes student Chris Lee. The interface area between the monomers in the MmpL3 models was found to be 2501.8 – 2720.7 Å² across the two species, with calculated solvation-free energy gain upon formation of the interface (Δ^iG) of -31.6 kcal/mol and -41.4 kcal/mol for *Msmeg* MmpL3 and *Mtb* MmpL3, respectively. The large interface area and the negative Δ^iG values indicated stable and strong interactions between the two protomers.

When *Msmeg* MmpL3 mode of dimerisation was analysed in more detail, it was observed that hydrophobic residues located on TM 1 – 3 were involved in establishing van der Waals interactions between the protomers, in a similar manner to the template protein HpnN although involving different TM domains and periplasmic regions ⁷⁵. Specifically, TM 7 of HpnN (residues 453 – 474) and TM 1b of *Msmeg* MmpL3 (residues 13 – 36) contain several leucine, isoleucine and valine residues that interact with TM located on the other protomer, specifically with TM 8 in the HpnN structure and with TM 2 – 3 in MmpL3 (Figure 4.15). Although MmpL3 only shares 11.2% sequence identity with HpnN, and the regions involved in dimer interface differ, the hydrophobic interactions occurring at the TM dimer interface suggest that the dimerisation mode is similar in both protein species. In addition to the hydrophobic interactions involving the aforementioned TM regions, several hydrogen bond-forming residues were identified on TM domains as well as in the periplasmic α 3, α 6 and α 7 in HpnN and in α 2 and α 4 in MmpL3, as shown in Figure 4.16, a. The main residues involved in hydrogen bond formation on the periplasmic domains were Arg65, Asp103,

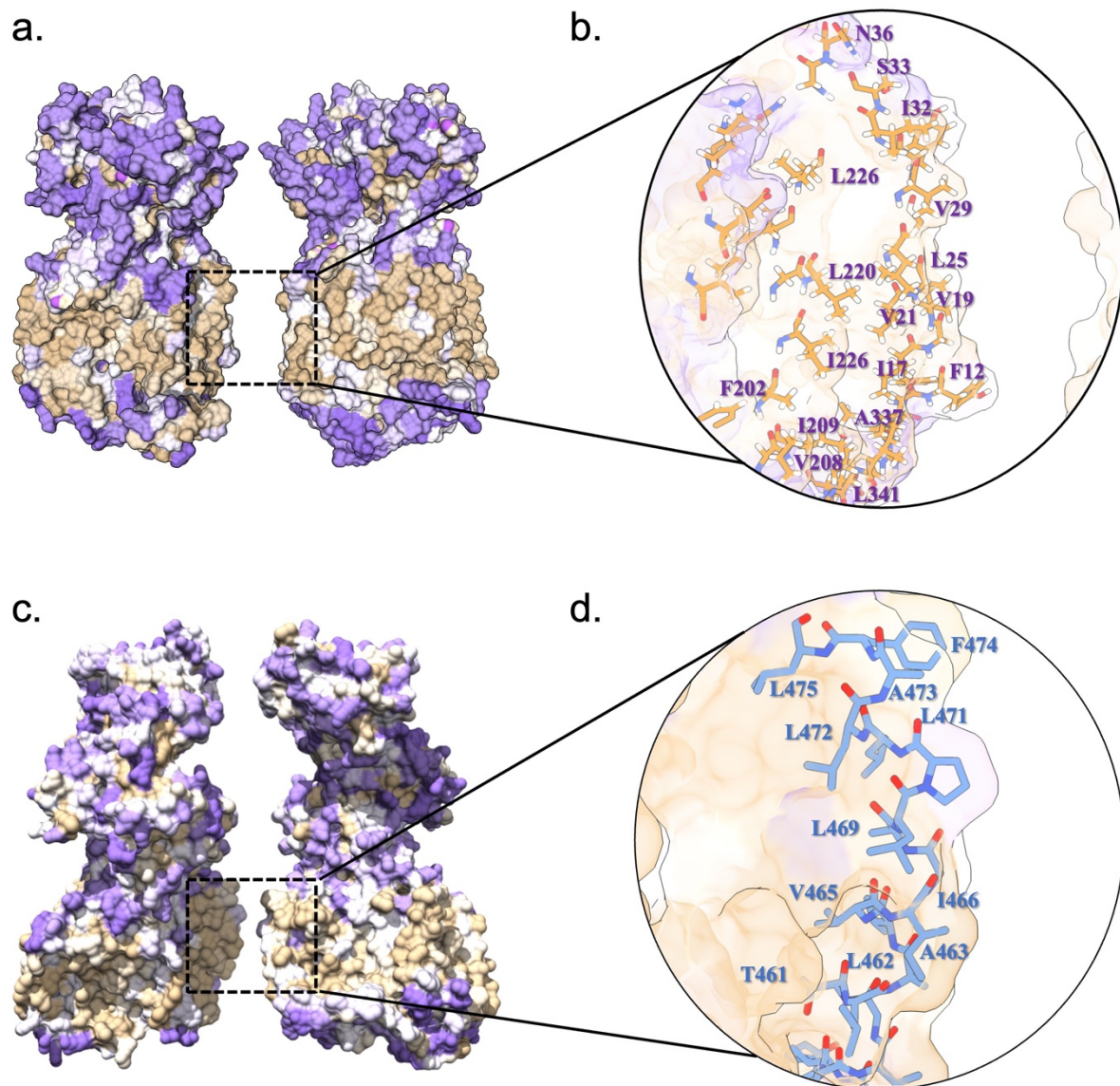


Figure 4.15. Hydrophobic surface properties of *Msmeg* MmpL3 (a) and *Burkholderia multivorans* HpnN (c). Residues are coloured based on their increasing hydrophobicity, from purple (hydrophilic), white (neutral) to beige (hydrophobic). The regions in contact with the neighbouring monomer are highlighted in the black box. Zoom-in views of the contact region, consisting of hydrophobic residues for *Msmeg* MmpL3 (b) and HpnN (d) are shown. Figures for HpnN (c and d) were created by the MRes student Chris Lee.

Gln104, Gln140, Asp142, and Glu146 for *Msmeg* MmpL3. At the TM level, both *Msmeg* protomers formed interchain hydrogen bonds in the lower TM region closer to the cytoplasm. In particular, between residues located on TM 1a (Arg13) and amino acids on TM 6 – TM 7 and their connecting loop (Arg344 and Thr352). Most of the residues involved in *Msmeg* dimerisation were also identified to play a crucial role in *Mtb* dimerisation (Figure 4.16, b). In addition to the hydrophobic interactions involving TM 1b and TM 2 – TM 3, several hydrogen bonds were identified at the interface of *Mtb* MmpL3 dimer involving residues Arg65, Asp103, Gln104, Asp137 and Asp138 in the

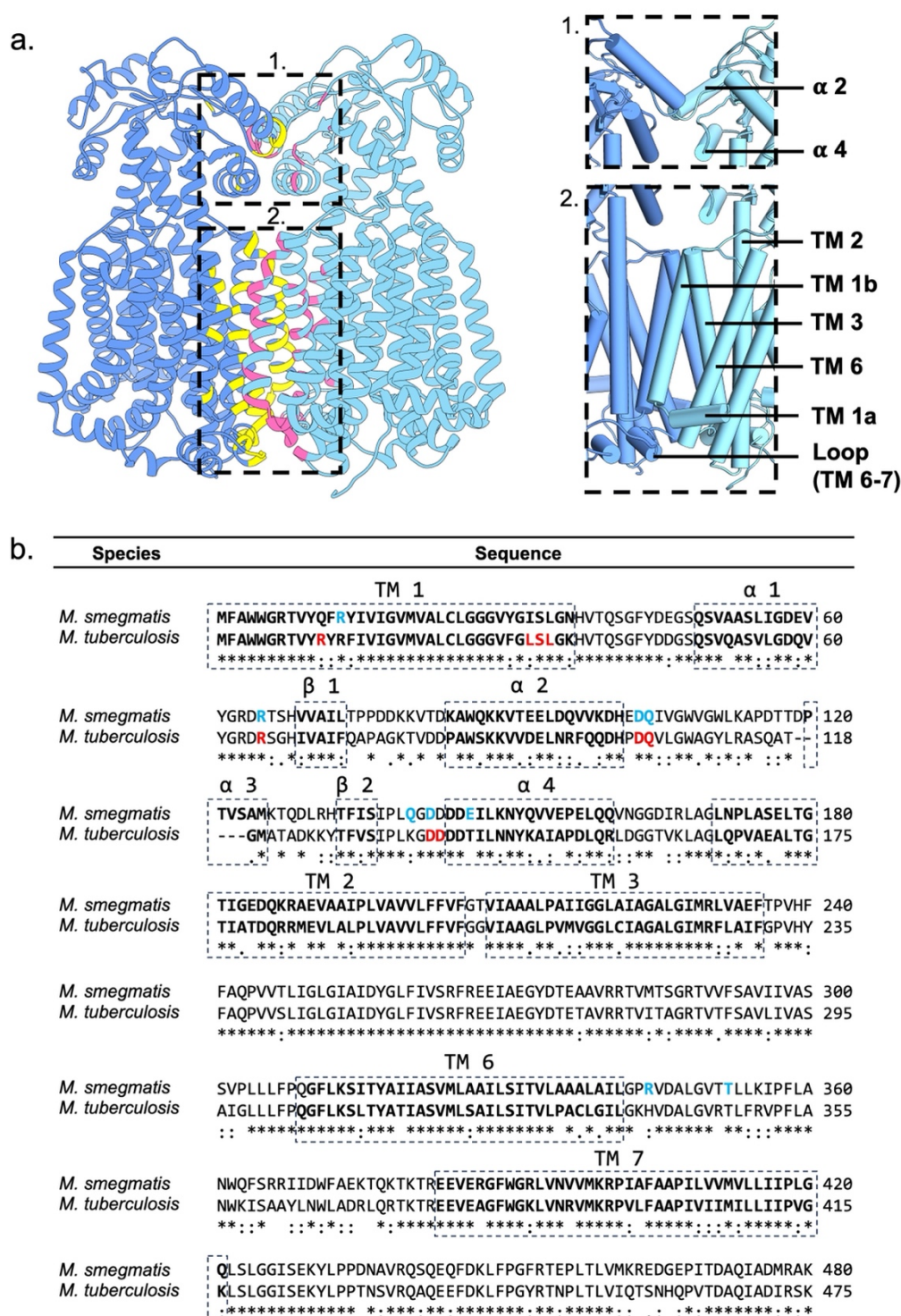


Figure 4.16. The interaction interface of the MmpL3 dimer. a. *Msmeg* MmpL3 dimer model generated using GalaxyHomomer server starting from PDB 7N6B (left hand panel). The individual protomers are coloured dark and pale blue with the helices shown in ribbon representation. The regions highlighted in the dashed boxes indicate those parts of the protein (both PD regions and TM domains) involved in the predicted dimer interface. A zoomed in image of the relevant secondary structure regions involved in the predicted dimer interface shown in cylinder representation (right hand panel). b. Sequence alignment between *Msmeg* and *Mtb* proteins for regions important for protein dimerization (main TM domains and PD regions within dashed boxes). Residues identified to form interchain hydrogen bonds in *Msmeg* model are in blue, while residues found in *Mtb* model are in red. A star symbol indicates identical residues between the two species. Semicolon indicates conservation of residues with strongly similar properties, while a dot indicates residues with weak similarities.

periplasmic regions, as well as Arg11 located on TM 1a and Val345 on the TM 6 – TM 7 connecting loop. Furthermore, only in the *Mtb* model were interactions in the upper TM region also identified, mainly involving Leu32, Ser33, Lys36, therefore resulting in an extended interface between the protomers. These observations suggest a crucial role for these interactions in the dimerisation and stabilisation of MmpL3. Nevertheless, it is essential to note that because these are predicted and rigid models, not all the interactions present in the native proteins could be captured.

To further corroborate the results obtained, computational alanine scanning was performed using the Robetta webserver on the *Msmeg* MmpL3 dimer, where changes in binding energy ($\Delta\Delta G_{\text{bind}}$) upon residue mutation into alanine were calculated (Table 2). Residues with $\Delta\Delta G_{\text{bind}}$ value greater than 1 kcal/mol were identified as hot-spots, potentially destabilising the dimer if mutated. As expected, mutation of hydrogen bond forming interfacial residues consistently showed high values of $\Delta\Delta G_{\text{bind}}$. The alanine scanning results confirmed that *Msmeg* MmpL3 is likely to dimerise via both the TM and periplasmic protein regions, forming stable hydrogen bonds interactions between the two chains.

4.3.9 MD simulations on dimeric *Msmeg* MmpL3

The dimeric model (model 1) of *Msmeg* MmpL3 obtained through GalaxyHomomer was embedded in a 60% POPE, 20% POPG and 20% POPC membrane based on our lipidomics analysis, and MD simulations were performed. Five repeats of 1.5 μs each were run and subsequently analysed to assess protein stability and behaviour in the membrane. Results obtained were then compared to the predicted model. In particular, hydrogen bond analyses were performed to gain insights into protein-protein and protein-lipid interactions. Lipid density maps were also generated to examine the arrangement and interaction of molecules with the dimers and to study the different species distribution around the protein.

4.3.10 Protein-protein interactions

Interchain hydrogen bonds forming throughout the whole trajectory and across all repeats were calculated and averaged. Only interactions > 10% total frames were considered relevant (Figure 4.17, a). With only a few exceptions, the residue pairs identified were consistent across both chains. In particular, in the PND domains, Arg65

Table 2. $\Delta\Delta G_{\text{bind}}$ values obtained from computational alanine scanning for *Msmeg* MmpL3 dimer model.

<i>Msmeg</i> MmpL3		
Model 1		
Residue	$\Delta\Delta G_{\text{bind}}$ (kcal/mol)	Protein region
Tyr10	3.14	TM 1a
Arg13	2.98	TM 1a
Leu25	1.60	TM 1b
Ile32	1.72	TM 1b
Arg65	1.55	α 1 – β 1
Asp103	2.30	α 2 – α 3
Gln104	3.70	α 2 – α 3
Glu146	2.19	α 4
Phe202	1.89	TM 2 – TM 3
Val208	1.44	TM 3
Ile209	1.49	TM 3
Ile340	1.87	TM 6
Arg344	2.87	TM 6 – TM 7
Thr352	2.86	TM 6 – TM 7

interacted with Asp103 and Glu102. Arg63 also formed H-bonds with Asp103, while Glu102 with Ser506 as well. Lys149 consistently interacted with Glu146, Gln104 with Asp142 while Gln140 mainly with its equivalent on the other chain (Figure 4.17, b). At the TM level, Ile32 interacted with the same residue on the associated chain, while Lys187 established interactions with Val29 (Figure 4.17, c). Finally, in the cytoplasmic side of the TM domains, Gln11, Arg13 and Tyr14 formed interactions with Thr352 and Leu353, while on both chains Ile340 accepted hydrogens from Arg344 (Figure 4.17, d).

In the *Mtb* model, most of the residues involved in establishing protein-protein interactions in the *Msmeg* dimer are conserved, with a few exceptions. Glu102 is substituted with a proline, Gln140 is substituted with a lysine therefore maintaining similar properties, while Glu146 is a threonine. At the TM level, some substitutions are

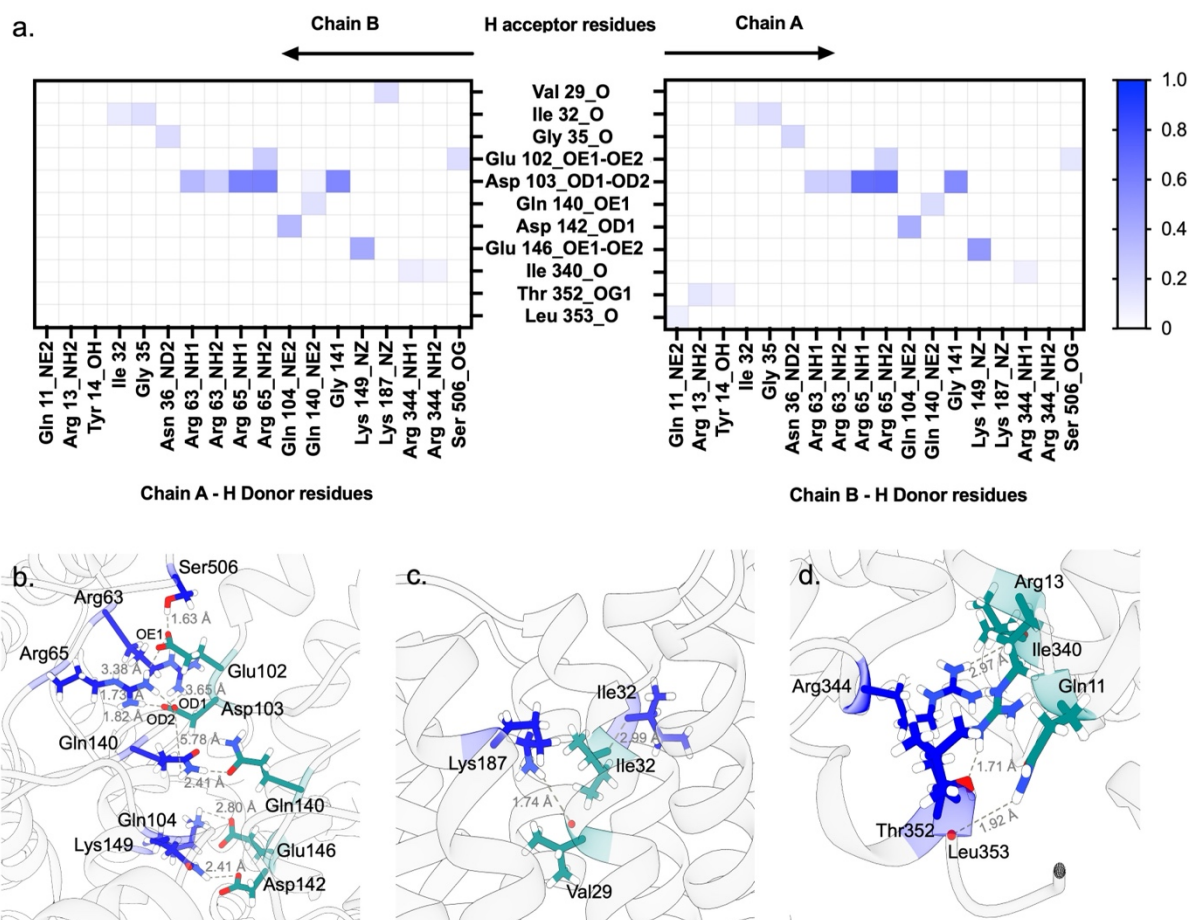


Figure 4.17. Protein-protein interactions for *Msmeg* MmpL3 dimer. a. Heatmaps for *Msmeg* MmpL3 protein-protein hydrogen bonds using both chain A residues as H donors and chain B residues as acceptors and the reverse. Max value on scale is 1, indicating a 100% of total simulation frames where bond is present. Main aa involved in protein dimerisation are located on periplasmic loops connecting $\alpha 1 - \beta 1$ (aa 63-65), $\alpha 2 - \alpha 3$ (aa 102-104) and $\beta 2 - \alpha 4$ (140-149) (b.); TM 1b (aa 29-32) and TM 2 (Lys187) in the region closest to periplasm (c.); TM 1a (aa 11-13), TM 6 (Ile340) and loop connecting TM 6 to TM 7 (aa 344-353) in region closest to cytoplasm (d.). Residues on chain A are in blue while residues on chain B are in dark cyan.

present but the residues have similar properties to the ones identified in the *Msmeg* model.

4.3.11 Periplasmic domain dynamics in dimer simulations

During monomer simulations it was observed that when MmpL3 is in its apo form the periplasmic central cavity adopts a closed state. When the PDB obtained averaging the last 100 ns of simulations was superimposed to the reference structure, one periplasmic domain was observed to move towards the other, thus restricting the central pore for substrates to bind (section 4.3.4). When the periplasmic domain behaviour was investigated for the *Msmeg* MmpL3 dimer, the same closure movement

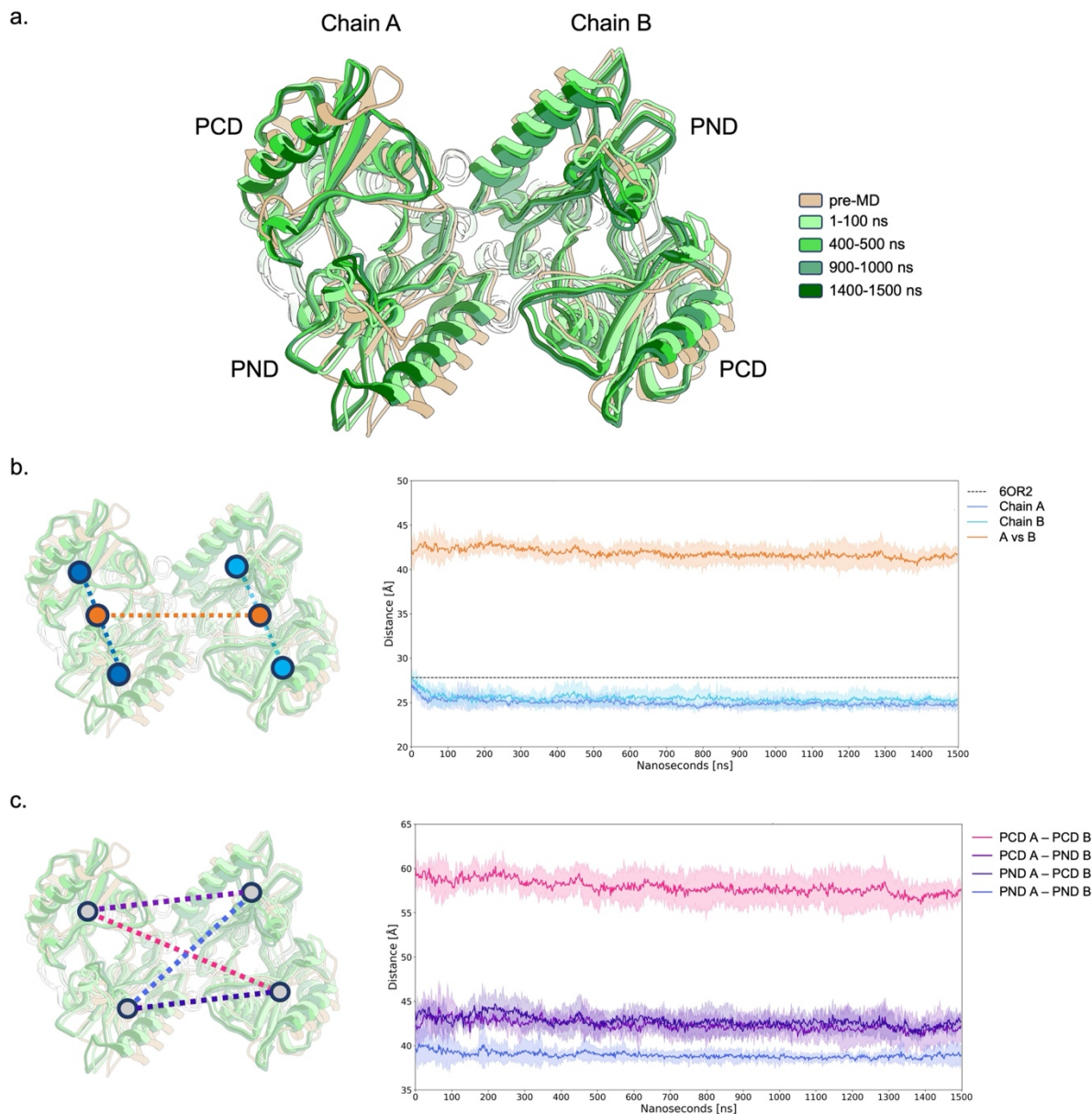


Figure 4.18. PD dynamics in *MsmeG* MmpL3 dimer simulations. a. Protein top view showing superimposition of averaged PDB structures at different simulation time-points. b. PND – PCD c.o.m. distance measurements along trajectory, compared to 6OR2 (dashed line). C.o.m. distance average values represented in solid lines with range over 5 repeats in shaded areas. Distance values for chain A are in blue, for chain B are in cyan, while interchain values are in orange. c. Interchain c.o.m. distance measurements along trajectory for each periplasmic subdomain pair. C.o.m. distance average values represented in solid lines with range over 5 repeats in shaded areas.

of the central cavity was observed for each individual protomer, and within the first 100 ns of the simulation (Figure 4.18, a and b). The PND – PCD distance for each chain dropped from 27.8 Å, reference measurement, to ~25 Å, same value observed in monomer simulations, and then remained stable over the remaining length of the trajectory. Interestingly, when interchain PND and PCD distances were measured, no

significant changes were observed, thus suggesting that the two protomers do not move towards each other (Figure 4.18, c).

4.3.12 Protein-lipid hydrogen bonds analysis in dimer simulations

As MmpL3 is an integral membrane protein natively embedded in a lipid bilayer, it is possible that lipids surrounding MmpL3 are involved in stabilising the dimeric structure by interacting with the protein. MmpL3 interaction with the surrounding membrane environment was therefore analysed specifically to calculate the formation of hydrogen bonds between protein residues and lipid headgroups. Averaged lipid density maps generated for both models revealed the presence of lipid molecules all around the dimer, with some densities intercalating within the protein-protein interface, supporting a stabilising role (Figure 4.19, a). The composition of the membrane (60% PE, 20% PG and 20% PC) was reflected in the lipid species identified in the occupancy map. This also confirmed that MmpL3 can bind to multiple lipid species not only through hydrophobic interactions but also through the active formation of hydrogen bonds. The hydrogen bond analysis carried out revealed that the protein residues establishing long-term interactions with the lipid headgroups were mainly located on the outer surface of the dimer, where the protein is most exposed to the membrane environment it is embedded in (Figure 4.19, b).

Residues in contact with the upper leaflet identified as important in the monomeric structures were also observed to form hydrogen bonds in the dimer simulations (Figure 4.20). These included Gln421, Ser423 and Leu424 (binding pocket 1 or lipid substrate entrance gate), aa 609, 616-618 (binding pocket 2), Asn621 (binding pocket 3) and Thr236 / Tyr317 (binding pocket 6). H-bonds were formed between all three lipid headgroups and the protein residues listed above, as shown by the heatmaps in Figure 4.21. The binding site hypothesised by Su and colleagues as a lipid entrance gate was observed to be occupied by one lipid molecule on both chains at the same time, suggesting each protomer can bind to its substrate independently³. The lipid molecule was positioned slightly above the membrane plane with a movement that suggested its partial translocation into the cavity, by interacting with a network of residues located within the TM and PCD domains (Figure 4.20, b).

Interestingly, a few lipid molecules were also observed to bind the protein residues located at the cleft formed between the two protomers, involving residues

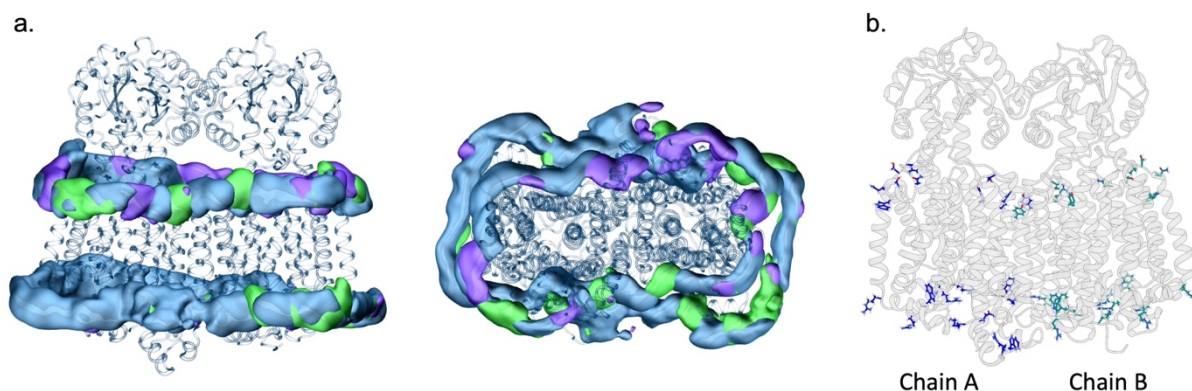


Figure 4.19. *Msmeg* MmpL3 dimer: protein-lipid interactions. a. Front (left) and top (right) view of averaged lipid occupancy map generated in VMD using a selection of PE (blue), PC (purple) and PG (green) molecules within 6 Å from protein along the whole trajectory. b. Main protein residues per chain involved in hydrogen bond formation with lipid molecules on top and bottom leaflets (front view).

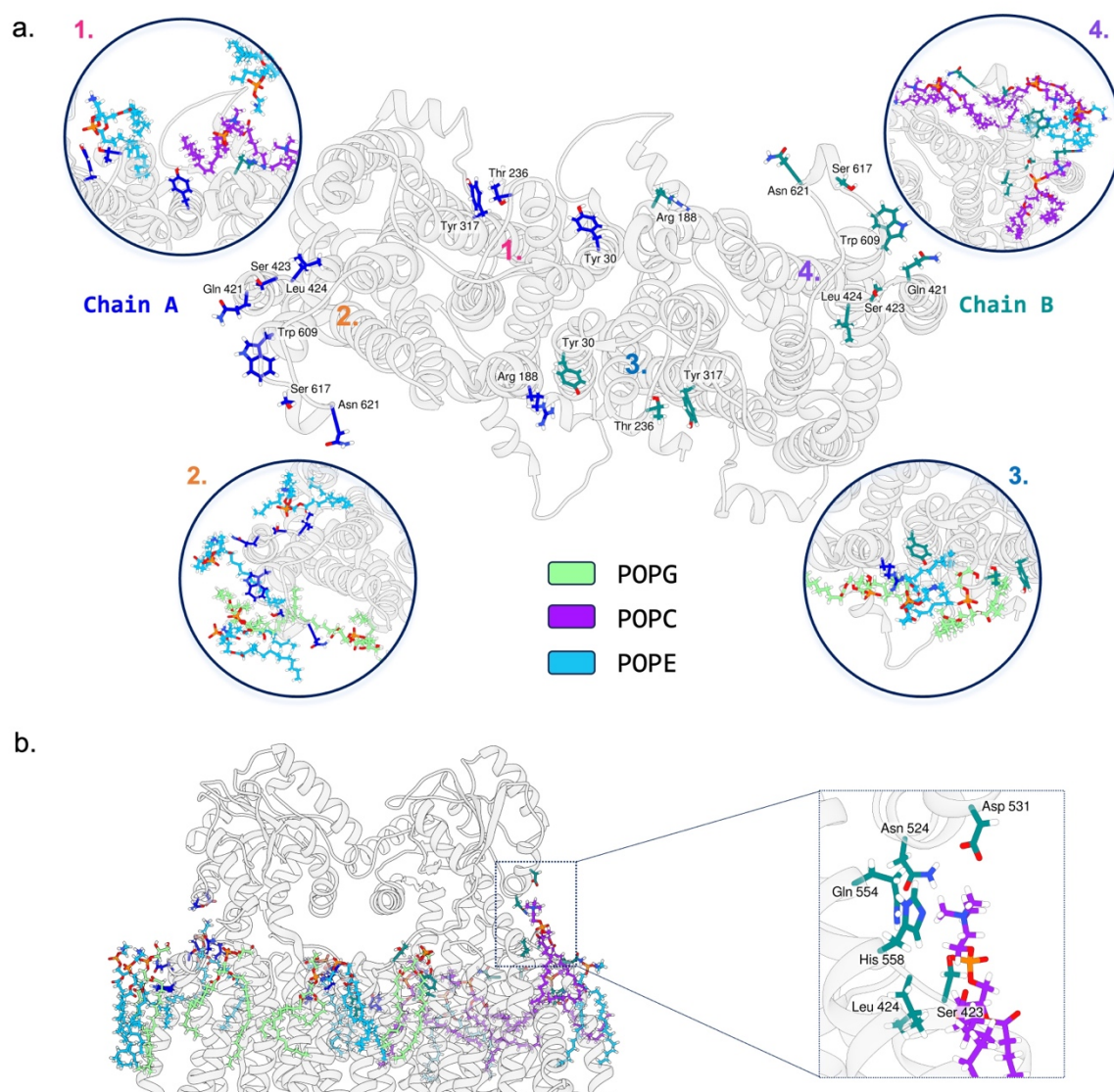


Figure 4.20. *Msmeg* MmpL3 dimer: protein-lipid interactions on top leaflet. a. Protein top view with main protein residues forming H bonds labelled and their side chains shown in stick representation. Inlets show these residues interacting with lipid species during simulation. b. Protein front view showing lipid species binding to MmpL3 residues, with focus on lipid entrance gate (inlet).

Arg188 and Tyr30 on both chains (Figure 4.20, a).

When the same analysis was performed on residues in contact with the lower leaflet, Trp4 / Arg7 (on TM 1a) and Trp362 / Arg366 (on cytoplasmic loop) from opposite chains were observed to bind several lipid molecules, that intercalated between the two protomers (Figure 4.22).

4.4 Discussion

4.4.1 Results on monomeric MmpL3

The structures solved by Su *et al.* and Adams and colleagues marked an essential step forward in MmpL3 structure characterisation²⁻⁴. Nevertheless, as with almost all solved structures, they carry some intrinsic limitations. First, the protein is fixed in a single conformation, providing only a static snapshot. In addition, the conditions used for structural studies of membrane proteins are not representative of the physiological behaviour in the context of the lipid bilayer. Furthermore, the use of detergents and other membrane mimetic systems to keep the protein in solution can alter its conformation and stability.

There is a tendency to deposit crystal structures in the PDB database with high B-factor values, and 6OR2, reference PDB for *Msmeg* MmpL3, is no exception. It was estimated by Carugo *et al.* that the average B-factor value for structures yielding a resolution lower than 3.3 Å should not be more than 80 Å². 6OR2 was solved at a 2.6 Å resolution but presents an average B-factor of 104.8 ± 21.4 Å² over 732 C α ³¹¹. This confirms there is a high level of uncertainty in determining the exact position of the backbone atoms, suggesting that dynamic studies are necessary in order to fully understand the physiological behaviour of the protein at atomic resolution. MD simulations have the advantage of allowing the study of a model structure mimicking the physiological conditions of temperature (T) and pressure (P). For example, compared to the 100 K at which diffraction data was collected for 6OR2, in this work the system was heated to a physiologically relevant T of 300 K and pressure was controlled during the entire length of the simulation². The energy produced allowed the entire system to move over time, thus letting the atoms explore their surroundings and adopt more energetically favourable conformations and interactions, that might differ from the reference structure. For example, a detailed distance analysis between

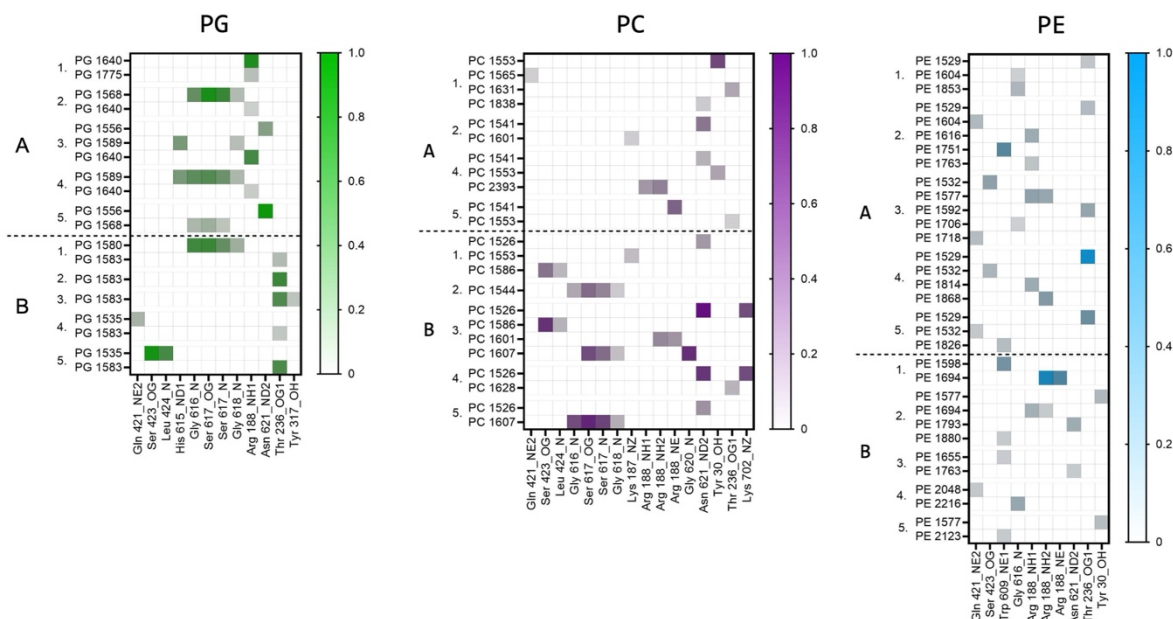


Figure 4.21. *Msmeg* MmpL3 dimer: hydrogen bond analysis on protein-lipid interactions on top leaflet. Heatmaps for protein-lipids (PG in green, PC in purple and PE in blue) hydrogen bonds by chain and simulation repeat. Protein residues and relative atoms involved in H bond are reported on x axis, while lipid headgroup molecules are on y axis. Max value on scale is 1, indicating a 100% of total simulation frames where bond is present. Only interactions >10% of total frames were plotted.

atoms located within the periplasmic domains of *Msmeg* MmpL3 highlighted a consistent trend in the dynamic behaviour of the periplasmic cavity. A ‘closed-state’ of the pore was observed in the unbound state of the protein in agreement with what had been showed for other RND transporters when no ligand is present ⁵. In contrast, an ‘open-state’ was observed when several lipid molecules were added to the system using the same atomic coordinates of the PE molecule identified in the reference structure for *Msmeg* MmpL3, 6OR2 ².

Another advantage of running MD simulations is that membrane proteins can be embedded in a lipid bilayer and the interactions between the membrane lipids and the protein can be studied at atomic resolution. Here, a particular focus was placed on protein residues located on the TM domains in closer contact with the periplasmic space, thus interacting with lipids on the upper leaflet of the bilayer. The reason behind this choice was that several solved *Msmeg* MmpL3 structures (6AJF, 6OR2 and 7N6B) were obtained in complex with detergent molecules that had been co-purified in the pockets located in the periplasmic end of TM helices 7 / 10 (zone 1), 9 / 11 (zone 3) and 3 / 6 (zone 6) ¹⁻³. Furthermore, it had been suggested by Su and colleagues that lipid binding pocket 1 could play a crucial role in the lipid transport mechanism of MmpL3 ². In particular, it has been hypothesised that Ser423 of TM7 could form a gate

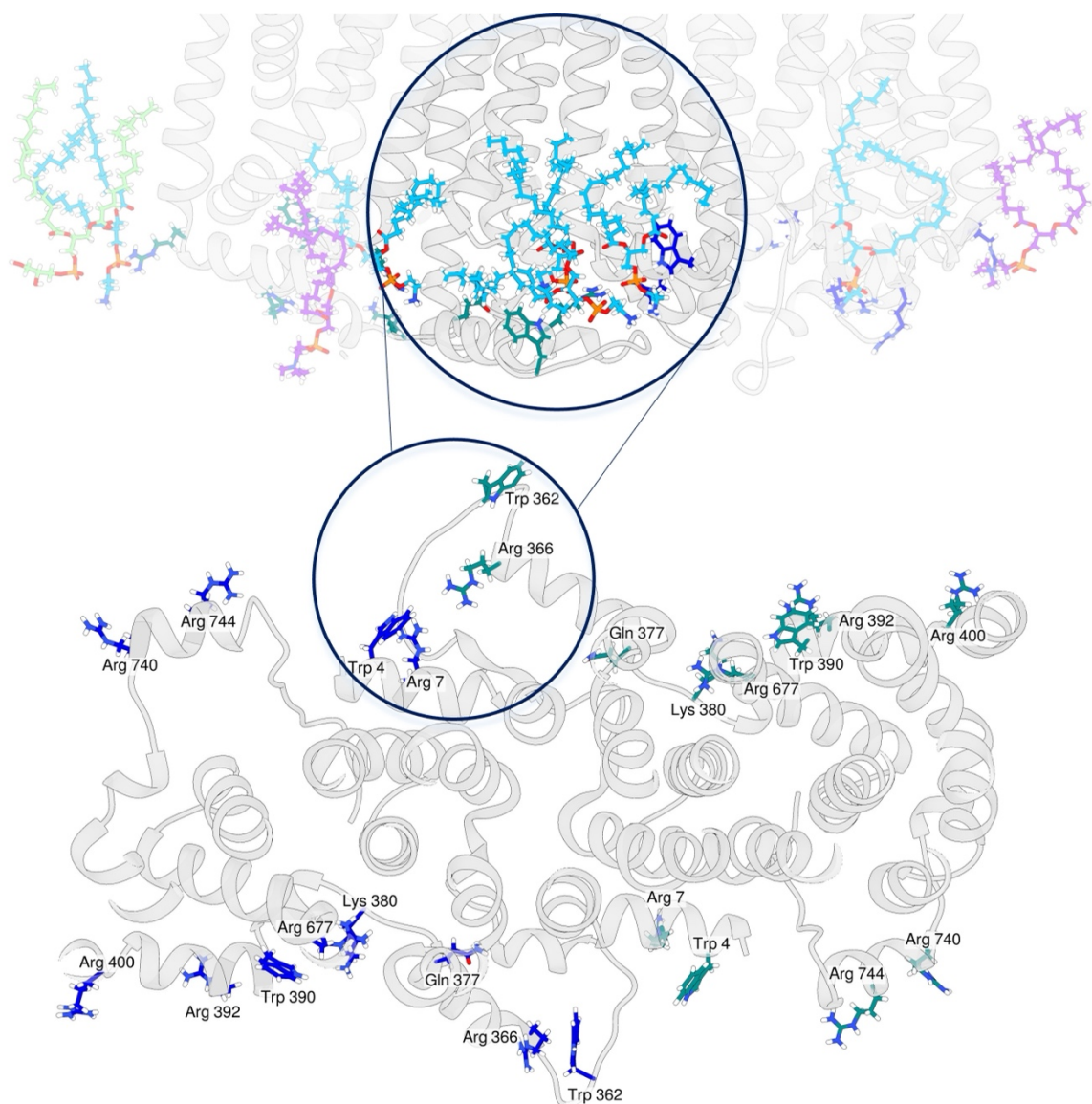


Figure 4.22. *Msmeg* MmpL3 dimer: protein-lipid interactions on bottom leaflet. Protein cytoplasmic view with main protein residues forming H bonds labelled and their side chains shown in stick representation. Inlet shows the main residues on opposite chains interacting with lipid species during simulation.

with Asn524 part of α 7 of the PCD, through which lipid molecules enter the central cavity³. During my simulations of monomeric MmpL3, PC molecules formed hydrogen bond interactions with Ser423 in the *Msmeg* MmpL3 model. In the dimer model, Gln421, Ser423 and Leu424 were also identified to form long-term interactions with PE, PG and PC molecules. Therefore, results suggest that the protein can bind to a wide variety of lipid species, not only to TMMs, although the binding affinity for TMMs has been shown to be higher and this may act as a regulatory mechanism controlling transport². When bound to this pocket, the lipid is lifted above the membrane plane, hinting at a possible transport initiation via movement of the substrate through the

‘gate’ and into the periplasmic cavity. Although no hydrogen bonds were observed between the lipid headgroups and Asn524, it cannot be excluded that a different type of interaction might take place or that alternative residues located on the PCD domain are involved in lipid transport.

Overall, results obtained for monomeric *Msmeg* and *Mtb* MmpL3 simulations not only consistently identified the same lipid binding pockets observed during protein structural determination, but also revealed several new TM binding sites. These findings highlighted the role that the lipid molecules potentially play in establishing or stabilising functional interactions with the protein. It was encouraging to achieve such close agreement between the computational and the experimental approaches, where one validated but also complemented the other in the investigation of MmpL3 in the context of a membrane.

4.4.2 Results on dimeric MmpL3

From the dimer models generated using HpnN, the only known RND protein that functions as a dimer as template, MmpL3 dimerisation could be investigated in more detail ^{75,293}. At structural level, extensive hydrophobic interactions between the protomers were established through several valine, leucine and isoleucine residues located on interface-forming TM helices, such as TM 1b, TM 2 and TM 3.

Although it was shown that Van der Waals interactions occurred at the level of the TM domains, it was unclear if and how the periplasmic domains were involved in MmpL3 dimerisation. MmpL3 has just two periplasmic domains, PND and PCD, while HpnN has an extra 2 domains (PD 3 and PD 4) extending further away into the periplasm ⁷⁵. Nevertheless, in the MmpL3 dimer the PND domains from each protomer are in close contact with each other, similar to PD 1 and PD 2 in the HpnN structure, suggesting a significant role in dimer interface formation. Multiple interchain hydrogen bonds at the dimer interface formed between PND regions of the two protomers were observed, mainly through the presence of polar and charged residues. Most of the key interface-forming residues involved in MmpL3 dimerisation were shared between *Msmeg* and *Mtb* proteins, as expected by the high percentage of sequence identity. Results obtained from simulations and from the interchain hydrogen bond analyses were consistent with findings obtained from the analysis of the GalaxyHomomer models, further supporting the significance of the key interface-forming residues.

Furthermore, in the MD simulations several new hydrogen bond-forming residues were identified both at PD and TM level compared to the pre-MD models, highlighting how dynamics studies are essential to capture a larger number of atomistic interactions that might not be present in rigid structures.

Computational alanine scanning provided additional support for the role of specific residues in dimer formation. Experimental studies with mutants would be useful to confirm their role in MmpL3 dimer stabilisation.

Overall, the MD simulations provided valuable insights into the dimerisation mode of MmpL3 by mainly identifying the likely dimer interface and the interactions stabilising dimer formation. As the simulated model was embedded in a membrane, we also hypothesised that lipid molecules might play a role in stabilising the dimer within the bilayer. This was also supported by the lipid density maps obtained for the two protein species, giving an annulus of lipids all around the two protomers with some lipid density intercalating within the dimer interface. The main long-term protein-lipid interactions were established on the outer surface of the dimer. Although the lipid-protein interactions at the interface presented a low frequency and duration, several lipid molecules were found to be bound at the cleft between the two monomers, suggesting their potential involvement in protein stabilisation.

4.4.3 Conclusions

Finally, it is worth noting that due to the high computational cost of running MD, the simulated systems were only able to capture a limited timeframe of the protein dynamics and of its interaction with the membrane. Nevertheless, as already observed with the monomer, the dimer maintained a high level of stability throughout the simulation, validating the results obtained and providing strong agreement with the predicted models regarding the key interface-forming residues. Overall, this Chapter provided valuable preliminary insights into the oligomerisation mode of MmpL3. However, these findings would need to be confirmed experimentally and a high-resolution structure of the dimer would be essential to investigate the protein-protein interactions. If solved, the dimeric 3D structure could be used as input file for MD simulations, thus providing a solid starting point to achieve high confidence results in regard to oligomer dynamics as well as to MmpL3 interactions with the surrounding membrane, potentially gaining more insights on its mechanism of lipid transport.

5. Final discussion

In this thesis, I have demonstrated how experimental procedures and computational approaches can synergistically be used to answer key biological questions. The greatest advantage of multidisciplinary projects is the chance to use a wide range of techniques and processes simultaneously, tackling the problem in a multi-faceted way. In this work, preliminary data suggesting the existence of MmpL3 oligomers was first obtained using SDS-PAGE and WB analysis on detergent-purified MmpL3 and the same detergent-purified MmpL3 reconstituted into MSP-based nanodiscs. When higher MW species were also observed by native PAGE analysis of MmpL3 in SMALPs, this gave us the confidence to investigate MmpL3 oligomeric state further. Several biophysical techniques were then used to explore our hypothesis that MmpL3 exists as an oligomer. Specifically, results obtained by SV-AUC, mass photometry, native MS and cryo-EM all confirmed the existence of MmpL3 as a dimer.

However, the data from these experimental techniques only provided limited information on the structure of the dimer and we were unable to obtain molecular level detail for the dimer interface. MD simulations filled in the gaps, revealing the structure of the dimer interface and the molecular basis of interactions between the individual protomers, involving TM and periplasmic domains, as well as detailed insights into MmpL3-membrane lipid interactions over time.

The highly effective complementary use of experimental and computational approaches was a key strength of this PhD project. In this case, combined experimental–computational methods provided new insights into the oligomeric arrangement of MmpL3 and how the protein is likely to behave when embedded in a native-like lipid bilayer. For example, we have seen how the lipidomics analysis carried out on MmpL3 in different membrane-mimetic systems has provided important information on the lipid composition that was then fed into the MD simulations performed on the MmpL3 dimer model. At the same time, the dimer model generated using GalaxyHomomer served as a synthetic template for the selection of protein-carrying particles during cryo-EM analysis. Furthermore, a clear differentiation between nanodisc particles carrying 1 or 2 protein molecules was achieved by

overlaying the monomer structure or the dimer model with the projected 2D images of the most representative 2D classes.

In some cases, the same phenomena may be observed across combined experimental–computational studies that would have been overlooked using experiments or simulations alone. Although a direct comparison between *in silico* and laboratory-based results always requires mindful consideration, the data presented in this thesis shows several instances where the two approaches pointed at the same conclusions ^{234,312}. As an example, we have seen how the TM pockets where substrate-like molecules co-purified during earlier MmpL3 structural determination were also clearly and reproducibly present persistently in MD simulations ^{1–4}. Hydrogen bonds were formed between lipid head groups and specific protein residues located all around the outer surface of the molecule over the whole length of the trajectory. Therefore, these lipid binding sites assume a relevance as possible mediators of both protein stability as well as playing a role in lipid transport. Of particular interest for lipid transport is the lipid entrance gate between TM 7 and TM 10 which was occupied by a TMM-analogue in the Su *et al.* structure ³. This site had already been hypothesised to be crucial for initiation of substrate transport through the central cavity and into the periplasm ². During MD, we observed that a lipid molecule occupied the same binding site, and it was also lifted up relative to the membrane plane, thus hinting at a translocating movement towards the central pore of the protein. Another example of a good level of agreement between experimental and computational results was provided by the analysis of the periplasmic cavity behaviour in response to ligands. What was observed during MD simulations was in line with the experimental evidence collected on other members of the RND protein superfamily ⁵. In the trimeric *Campylobacter jejuni* CmeB, several transient conformational states were observed for the periplasmic domain, likely reflecting the different stages the pump must go through during the transport cycle ³¹³. In one of the crystal structures one protomer is in the binding state with the periplasmic cleft open, while the other two protomers present closed conformations of the periplasmic cleft, either adopting an extrusion or resting state ³¹³. Therefore, a PD in an open conformation is usually associated with a substrate-bound state. This open-state was observed in the reference structure, with a PE molecule co-crystallised within the cavity, but also in the simulations including ligands ². The conformational opening movement of the PD is thought to take place to accommodate the substrate within the cavity during its

membrane to periplasm translocation. In contrast, when the protein was simulated in its apo state the periplasmic domains moved towards each other, reducing the size of the central pore as it was also observed for the apo form of the *E. coli* CusA protein³¹⁴.

While experimental procedures often make use of well-established protocols that have been widely used within the scientific community sometimes over decades, computational approaches are prone to rapidly evolve. It is often very hard to keep up with the advancement of computational speed and technology, thus complicating matters where it comes to reproducibility of results. In the MD field it is common to have yearly release of new software packages, providing up to date tools that allow simulation of systems with increasing complexity as well as more sophisticated post-MD analyses. In this work, the AMBER software package as well as its homonymous molecular mechanical force fields were used for the simulation of biomolecules^{240,315,316}. Within the AmberTools suite, a collection of programs and commands is available for the user to generate coordinates and topology files prior to simulations, as well as for downstream trajectory analysis²⁴⁰. The Assisted Model Building with Energy Refinement (AMBER) is also a family of force fields originally developed by Peter Kollman's group at the University of California (San Francisco)³¹⁵. Force fields represent the core of the MD software as they are used to describe bonded and non-bonded interactions between atoms within biomolecules. The specific set of parameters and how well each system's component is defined directly affect the simulations accuracy and reliability. The latest model of the SB protein force fields dates back to 2019²⁴². It was developed by the Simmerling Lab at the Stony Brook University with the name of ff19SB. In terms of lipid force fields, AMBER included the development of a modular framework and initial parameterisation within the Lipid11 force field, with significant advancements included in the 2014 version, Lipid14^{317,318}. At the time of this PhD project, the latest release was Lipid17, built upon the modularity of the previous version but with an extension of phospholipid residues to include anionic head groups and polyunsaturated tails. Therefore, Lipid17 was the force field used in the first MD simulations performed on the monomeric form of MmpL3. Although during the course of my PhD a newer version of the lipid force field was released, named Lipid21, a rational choice was made to continue using Lipid17 throughout the whole length of the project, to allow for a fairer comparison between simulated systems²⁴³. In fact, all systems were built using the same force fields,

repeats were individually minimised and MD simulations strictly run using the same environmental conditions and over the time frame of 1.5 μ s. In this way, by averaging the results obtained across the five different repeats for each system it was possible to achieve a great level of data significance and reproducibility which is usually a prerogative of experimental procedures.

The wide range of multidisciplinary approaches used throughout this thesis supported the hypothesis that MmpL3 exists as a homodimer. However, more work is required in order to more efficiently separate and isolate the dimeric species and achieving the ultimate aim of solving a high-resolution structure. Having a 3D structure of the dimer would allow for an in-depth characterisation of the protein-protein interface as well as the study of the protein-lipid interaction network, replacing the synthetic model used in this work. For example, mutants could be generated *in silico* as well as in the laboratory (e.g. through alanine scanning) and could be tested to identify the crucial interface-forming residues as well as important amino acids involved in protein-lipid interactions. Furthermore, a lipid screen could be performed to investigate whether the protein quaternary structure is regulated (e.g. stabilisation, functional activation) by specific lipids at membrane level. These molecules might act as regulators of protein oligomerisation and complex assembly, but they could also stabilise the different conformational states adopted by the protein or could be modulators of ligand binding ¹²⁶. As a final suggestion, it would be beneficial to carry out recombinant expression of MmpL3 in *Msmeg* or *Mtb* cells and to extract the protein in its native environment (e.g. in SMALPs). This would reduce any bias that could arise from protein manipulation during detergent-based purification, while replicating physiologically relevant conditions.

In summary, this investigation of the oligomeric state of MmpL3 successfully exploited the synergy between biophysical techniques and computational modelling. This represents a novel and original finding that contributes to our understanding of this important mycobacterial lipid transporter, a credible target in the continued fight against TB.

6. Bibliography

- 1 B. Zhang, J. Li, X. Yang, L. Wu, J. Zhang, Y. Yang, Y. Zhao, L. Zhang, X. Yang, X. Yang, X. Cheng, Z. Liu, B. Jiang, H. Jiang, L. W. Guddat, H. Yang and Z. Rao, Crystal Structures of Membrane Transporter MmpL3, an Anti-TB Drug Target, *Cell*, 2019, **176**, 636-648.e13.
- 2 C.-C. Su, P. A. Klenotic, J. R. Bolla, G. E. Purdy, C. V. Robinson and E. W. Yu, MmpL3 is a lipid transporter that binds trehalose monomycolate and phosphatidylethanolamine, *PNAS*, 2019, **116**, 11241–11246.
- 3 C.-C. Su, P. A. Klenotic, M. Cui, M. Lyu, C. E. Morgan and E. W. Yu, Structures of the mycobacterial membrane protein MmpL3 reveal its mechanism of lipid transport, *PLOS Biology*, 2021, **19**, e3001370.
- 4 O. Adams, J. C. Deme, J. L. Parker, P. W. Fowler, S. M. Lea and S. Newstead, Cryo-EM structure and resistance landscape of *M. tuberculosis* MmpL3: An emergent therapeutic target, *Structure*, 2021, S0969212621002173.
- 5 P. A. Klenotic, M. A. Moseng, C. E. Morgan and E. W. Yu, Structural and Functional Diversity of Resistance–Nodulation–Cell Division Transporters, *Chem. Rev.*, 2021, **121**, 5378–5416.
- 6 K. Dheda, C. E. Barry and G. Maartens, Tuberculosis, *The Lancet*, 2016, **387**, 1211–1226.
- 7 I. Hershkovitz, H. D. Donoghue, D. E. Minnikin, H. May, O. Y.-C. Lee, M. Feldman, E. Galili, M. Spigelman, B. M. Rothschild and G. K. Bar-Gal, Tuberculosis origin: The Neolithic scenario, *Tuberculosis (Edinb)*, 2015, **95 Suppl 1**, S122-126.
- 8 [The etiology of tuberculosis by Dr. Robert Koch. From the Berliner Klinische Wochenschrift, Volume 19 (1882)], *Zentralbl Bakteriol Mikrobiol Hyg A*, 1982, **251**, 287–296.
- 9 C. L. Dulberger, E. J. Rubin and C. C. Boutte, The mycobacterial cell envelope — a moving target, *Nature Reviews Microbiology*, 2019, 1–13.
- 10 G. A. W. Rook, K. Dheda and A. Zumla, Immune responses to tuberculosis in developing countries: implications for new vaccines, *Nature Reviews Immunology*, 2005, **5**, 661–667.
- 11 T. Jagielski, A. Minias, J. van Ingen, N. Rastogi, A. Brzostek, A. Żaczek and J. Dziadek, Methodological and Clinical Aspects of the Molecular Epidemiology of *Mycobacterium tuberculosis* and Other *Mycobacteria*, *Clin Microbiol Rev*, 2016, **29**, 239–290.
- 12 R. M. G. J. Houben and P. J. Dodd, The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling, *PLoS Med*, 2016, 13(10): e1002152. DOI:10.1371/journal.pmed.1002152.
- 13 J. C. Emery, A. S. Richards, K. D. Dale, C. F. McQuaid, R. G. White, J. T. Denholm and R. M. G. J. Houben, Self-clearance of *Mycobacterium tuberculosis* infection: implications for lifetime risk and population at-risk of tuberculosis disease, *Proc Biol Sci*, 2021, **288**, 20201635.

- 14 M. A. Behr, P. H. Edelstein and L. Ramakrishnan, Is *Mycobacterium tuberculosis* infection life long?, *BMJ*, 2019, **367**, l5770.
- 15 E. Vynnycky and P. E. Fine, Lifetime risks, incubation period, and serial interval of tuberculosis, *Am J Epidemiol*, 2000, **152**, 247–263.
- 16 S. Gagneux, Ecology and evolution of *Mycobacterium tuberculosis*, *Nat Rev Microbiol*, 2018, **16**, 202–213.
- 17 J. Hayman, MYCOBACTERIUM ULCERANS: AN INFECTION FROM JURASSIC TIME?, *The Lancet*, 1984, **324**, 1015–1016.
- 18 T. Rogall, J. Wolters, T. Flohr and E. C. Böttger, Towards a phylogeny and definition of species at the molecular level within the genus *Mycobacterium*, *Int J Syst Bacteriol*, 1990, **40**, 323–330.
- 19 D. Brites and S. Gagneux, Old and new selective pressures on *Mycobacterium tuberculosis*, *Infect Genet Evol*, 2012, **12**, 678–685.
- 20 P. Supply, M. Marceau, S. Mangenot, D. Roche, C. Rouanet, V. Khanna, L. Majlessi, A. Criscuolo, J. Tap, A. Pawlik, L. Fiette, M. Orgeur, M. Fabre, C. Parmentier, W. Frigui, R. Simeone, E. C. Boritsch, A.-S. Debie, E. Willery, D. Walker, M. A. Quail, L. Ma, C. Bouchier, G. Salvignol, F. Sayes, A. Cascioferro, T. Seemann, V. Barbe, C. Locht, M.-C. Gutierrez, C. Leclerc, S. D. Bentley, T. P. Stinear, S. Brisse, C. Médigue, J. Parkhill, S. Cruveiller and R. Brosch, Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of *Mycobacterium tuberculosis*, *Nat Genet*, 2013, **45**, 172–179.
- 21 D. Ebert and J. J. Bull, Challenging the trade-off model for the evolution of virulence: is virulence management feasible?, *Trends Microbiol*, 2003, **11**, 15–20.
- 22 J. D. Ernst, The immunological life cycle of tuberculosis, *Nat Rev Immunol*, 2012, **12**, 581–591.
- 23 I. Comas, J. Chakravarti, P. M. Small, J. Galagan, S. Niemann, K. Kremer, J. D. Ernst and S. Gagneux, Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved, *Nat Genet*, 2010, **42**, 498–503.
- 24 M. Coscolla, R. Copin, J. Sutherland, F. Gehre, B. de Jong, O. Owolabi, G. Mbayo, F. Giardina, J. D. Ernst and S. Gagneux, M. tuberculosis T Cell Epitope Analysis Reveals Paucity of Antigenic Variation and Identifies Rare Variable TB Antigens, *Cell Host Microbe*, 2015, **18**, 538–548.
- 25 I. Yruela, B. Contreras-Moreira, C. Magalhães, N. S. Osório and J. Gonzalo-Asensio, *Mycobacterium tuberculosis* Complex Exhibits Lineage-Specific Variations Affecting Protein Ductility and Epitope Recognition, *Genome Biology and Evolution*, 2016, **8**, 3751.
- 26 M. C. Gutierrez, S. Brisse, R. Brosch, M. Fabre, B. Omaïs, M. Marmiesse, P. Supply and V. Vincent, Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*, *PLoS Pathog*, 2005, **1**, e5.
- 27 I. Comas, E. Hailu, T. Kiros, S. Bekele, W. Mekonnen, B. Gumi, R. Tschopp, G. Ameni, R. G. Hewinson, B. D. Robertson, G. A. Goig, D. Stucki, S. Gagneux, A. Aseffa, D. Young and S. Berg, Population Genomics of *Mycobacterium tuberculosis* in Ethiopia Contradicts the Virgin Soil Hypothesis for Human Tuberculosis in Sub-Saharan Africa, *Curr Biol*, 2015, **25**, 3260–3266.

- 28 K. C. Rahlwes, B. R. S. Dias, P. C. Campos, S. Alvarez-Arguedas and M. U. Shiloh, Pathogenicity and virulence of *Mycobacterium tuberculosis*, *Virulence*, **14**, 2150449.
- 29 S. B. Cohen, B. H. Gern, J. L. Delahaye, K. N. Adams, C. R. Plumlee, J. K. Winkler, D. R. Sherman, M. Y. Gerner and K. B. Urdahl, Alveolar Macrophages Provide an Early *Mycobacterium tuberculosis* Niche and Initiate Dissemination, *Cell Host Microbe*, 2018, **24**, 439-446.e4.
- 30 T. S. Rodrigues, B. J. Conti, T. F. de C. Fraga-Silva, F. Almeida and V. L. D. Bonato, Interplay between alveolar epithelial and dendritic cells and *Mycobacterium tuberculosis*, *J Leukoc Biol*, 2020, **108**, 1139–1156.
- 31 K. B. Urdahl, S. Shafiani and J. D. Ernst, Initiation and regulation of T-cell responses in tuberculosis, *Mucosal Immunol*, 2011, **4**, 288–293.
- 32 S. M. Carpenter and L. L. Lu, Leveraging Antibody, B Cell and Fc Receptor Interactions to Understand Heterogeneous Immune Responses in Tuberculosis, *Front Immunol*, 2022, **13**, 830482.
- 33 H.-D. Park, K. M. Guinn, M. I. Harrell, R. Liao, M. I. Voskuil, M. Tompa, G. K. Schoolnik and D. R. Sherman, Rv3133c/dosR is a transcription factor that mediates the hypoxic response of *Mycobacterium tuberculosis*, *Mol Microbiol*, 2003, **48**, 833–843.
- 34 C. W. M. Ong, P. T. Elkington, S. Brilha, C. Ugarte-Gil, M. T. Tome-Esteban, L. B. Tezera, P. J. Pabisiak, R. C. Moores, T. Sathyamoorthy, V. Patel, R. H. Gilman, J. C. Porter and J. S. Friedland, Neutrophil-Derived MMP-8 Drives AMPK-Dependent Matrix Destruction in Human Pulmonary Tuberculosis, *PLoS Pathog*, 2015, **11**, e1004917.
- 35 M. U. Shiloh, Mechanisms of mycobacterial transmission: how does *Mycobacterium tuberculosis* enter and escape from the human host, *Future Microbiol*, 2016, **11**, 1503–1506.
- 36 C. R. Ruhl, B. L. Pasko, H. S. Khan, L. M. Kindt, C. E. Stamm, L. H. Franco, C. C. Hsia, M. Zhou, C. R. Davis, T. Qin, L. Gautron, M. D. Burton, G. L. Mejia, D. K. Naik, G. Dussor, T. J. Price and M. U. Shiloh, *Mycobacterium tuberculosis* Sulfolipid-1 Activates Nociceptive Neurons and Induces Cough, *Cell*, 2020, **181**, 293-305.e11.
- 37 The end TB strategy, <https://www.who.int/publications-detail-redirect/WHO-HTM-TB-2015.19>, (accessed 15 May 2024).
- 38 Global Tuberculosis Report 2023, <https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2023>, (accessed 15 May 2024).
- 39 E. W. Tiemersma, M. J. van der Werf, M. W. Borgdorff, B. G. Williams and N. J. D. Nagelkerke, Natural history of tuberculosis: duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: a systematic review, *PLoS One*, 2011, **6**, e17601.
- 40 WHO consolidated guidelines on tuberculosis. Module 4: Treatment – drug-susceptible tuberculosis treatment. Geneva: World Health Organization; 2022 (<https://iris.who.int/handle/10665/353829>) (accessed 15 May 2024).
- 41 V. Eldholm and F. Balloux, Antimicrobial Resistance in *Mycobacterium tuberculosis*: The Odd One Out, *Trends Microbiol*, 2016, **24**, 637–648.

- 42 G. V. Bloemberg, P. M. Keller, D. Stucki, A. Trauner, S. Borrell, T. Latshang, M. Coscolla, T. Rothe, R. Hömke, C. Ritter, J. Feldmann, B. Schulthess, S. Gagneux and E. C. Böttger, Acquired Resistance to Bedaquiline and Delamanid in Therapy for Tuberculosis, *N Engl J Med*, 2015, **373**, 1986–1988.
- 43 S. Gagneux, C. D. Long, P. M. Small, T. Van, G. K. Schoolnik and B. J. M. Bohannon, The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis*, *Science*, 2006, **312**, 1944–1946.
- 44 S. Borrell, Y. Teo, F. Giardina, E. M. Streicher, M. Klopper, J. Feldmann, B. Müller, T. C. Victor and S. Gagneux, Epistasis between antibiotic resistance mutations drives the evolution of extensively drug-resistant tuberculosis, *Evol Med Public Health*, 2013, **2013**, 65–74.
- 45 T. Smith, K. A. Wolff and L. Nguyen, Molecular Biology of Drug Resistance in *Mycobacterium tuberculosis*, *Curr Top Microbiol Immunol*, 2013, **374**, 53–80.
- 46 L. Nguyen, Antibiotic resistance mechanisms in *M. tuberculosis*: an update, *Arch Toxicol*, 2016, **90**, 1585–1604.
- 47 S. N. Goossens, S. L. Sampson and A. Van Rie, Mechanisms of Drug-Induced Tolerance in *Mycobacterium tuberculosis*, *Clin Microbiol Rev*, 2020, **34**, e00141-20.
- 48 N. Dookie, S. Rambaran, N. Padayatchi, S. Mahomed and K. Naidoo, Evolution of drug resistance in *Mycobacterium tuberculosis*: a review on the molecular determinants of resistance and implications for personalized care, *J Antimicrob Chemother*, 2018, **73**, 1138–1151.
- 49 WHO consolidated guidelines on tuberculosis. Module 4: Treatment – drug-resistant tuberculosis treatment. Geneva: World Health Organization; 2022 update (<https://iris.who.int/handle/10665/365308>) (accessed 15 May 2024).
- 50 WHO launches the TB research tracker, an online platform to track progress in TB research, <https://www.who.int/news/item/05-06-2023-who-launches-the-tb-research-tracker--an-online-platform-to-track-progress-in-tb-research>, (accessed 15 May 2024).
- 51 A. E. Grzegorzewicz, H. Pham, V. A. K. B. Gundi, M. S. Scherman, E. J. North, T. Hess, V. Jones, V. Gruppo, S. E. M. Born, J. Korduláková, S. S. Chavadi, C. Morisseau, A. J. Lenaerts, R. E. Lee, M. R. McNeil and M. Jackson, Inhibition of mycolic acid transport across the *Mycobacterium tuberculosis* plasma membrane, *Nature Chemical Biology*, 2012, **8**, 334–341.
- 52 C. E. Barry, R. E. Lee, K. Mdluli, A. E. Sampson, B. G. Schroeder, R. A. Slayden and Y. Yuan, Mycolic acids: structure, biosynthesis and physiological functions, *Prog Lipid Res*, 1998, **37**, 143–179.
- 53 M. Daffé and P. Draper, The envelope layers of mycobacteria with reference to their pathogenicity, *Adv Microb Physiol*, 1998, **39**, 131–203.
- 54 E. K. Schroeder, N. de Souza, D. S. Santos, J. S. Blanchard and L. A. Basso, Drugs that inhibit mycolic acid biosynthesis in *Mycobacterium tuberculosis*, *Curr Pharm Biotechnol*, 2002, **3**, 197–225.
- 55 K. Takayama, C. Wang and G. S. Besra, Pathway to Synthesis and Processing of Mycolic Acids in *Mycobacterium tuberculosis*, *Clin Microbiol Rev*, 2005, **18**, 81–101.
- 56 H. Feinberg, S. A. F. Jégouzo, T. J. W. Rowntree, Y. Guan, M. A. Brash, M. E. Taylor, W. I. Weis and K. Drickamer, Mechanism for recognition of an unusual

- mycobacterial glycolipid by the macrophage receptor mincle, *J Biol Chem*, 2013, **288**, 28457–28465.
- 57 I. Matsunaga and D. B. Moody, Mincle is a long sought receptor for mycobacterial cord factor, *J Exp Med*, 2009, **206**, 2865–2868.
 - 58 E. C. Patin, A. C. Geffken, S. Willcocks, C. Leschczyk, A. Haas, F. Nimmerjahn, R. Lang, T. H. Ward and U. E. Schaible, Trehalose dimycolate interferes with FcγR-mediated phagosome maturation through Mincle, SHP-1 and FcγRIIB signalling, *PLoS ONE*, 2017, **12**(4): e0174973. DOI:10.1371/journal.pone.0174973.
 - 59 W.-B. Lee, J.-S. Kang, J.-J. Yan, M. S. Lee, B.-Y. Jeon, S.-N. Cho and Y.-J. Kim, Neutrophils Promote Mycobacterial Trehalose Dimycolate-Induced Lung Inflammation via the Mincle Pathway, *PLoS Pathog*, 2012, **8**, e1002614.
 - 60 M. Hansen, J. Peltier, B. Killy, B. Amin, B. Bodendorfer, A. Härtlova, S. Uebel, M. Bosmann, J. Hofmann, C. Büttner, A. B. Ekici, M. Kuttke, H. Franzyk, C. Foged, S. Beer-Hammer, G. Schabbauer, M. Trost and R. Lang, Macrophage Phosphoproteome Analysis Reveals MINCLE-dependent and -independent Mycobacterial Cord Factor Signaling, *Mol Cell Proteomics*, 2019, **18**, 669–685.
 - 61 E. J. North, M. Jackson and R. E. Lee, New Approaches to Target the Mycolic Acid Biosynthesis Pathway for the Development of Tuberculosis Therapeutics, *Curr Pharm Des*, 2014, **20**, 4357–4378.
 - 62 J. Bernstein, W. A. Lott, B. A. Steinberg and H. L. Yale, Chemotherapy of experimental tuberculosis. V. Isonicotinic acid hydrazide (nydrazid) and related compounds, *Am Rev Tuberc*, 1952, **65**, 357–364.
 - 63 B. Heym, P. M. Alzari, N. Honoré and S. T. Cole, Missense mutations in the catalase-peroxidase gene, *katG*, are associated with isoniazid resistance in *Mycobacterium tuberculosis*, *Mol Microbiol*, 1995, **15**, 235–245.
 - 64 H. E. Jenkins, M. Zignol and T. Cohen, Quantifying the burden and trends of isoniazid resistant tuberculosis, 1994-2009, *PLoS One*, 2011, **6**, e22927.
 - 65 Y. Zhang, B. Heym, B. Allen, D. Young and S. Cole, The catalase-peroxidase gene and isoniazid resistance of *Mycobacterium tuberculosis*, *Nature*, 1992, **358**, 591–593.
 - 66 Z. Xu, V. A. Meshcheryakov, G. Poce and S.-S. Chng, MmpL3 is the flippase for mycolic acids in mycobacteria, *Proc Natl Acad Sci U S A*, 2017, **114**, 7993–7998.
 - 67 A. Viljoen, V. Dubois, F. Girard-Misguich, M. Blaise, J.-L. Herrmann and L. Kremer, The diverse family of MmpL transporters in mycobacteria: from regulation to antimicrobial developments, *Molecular Microbiology*, 2017, **104**, 889–904.
 - 68 H. Nikaido, RND transporters in the living world, *Research in Microbiology*, 2018, **169**, 363–371.
 - 69 J. M. Belardinelli, A. Yazidi, L. Yang, L. Fabre, W. Li, B. Jacques, S. kumar Angala, I. Rouiller, H. I. Zgurskaya, J. Sygusch and M. Jackson, Structure-function profile of MmpL3, the essential mycolic acid transporter from *Mycobacterium tuberculosis*, *ACS Infect Dis*, 2016, **2**, 702–713.
 - 70 S. Murakami, R. Nakashima, E. Yamashita and A. Yamaguchi, Crystal structure of bacterial multidrug efflux transporter AcrB, *Nature*, 2002, **419**, 587–593.

- 71 S. Murakami, R. Nakashima, E. Yamashita, T. Matsumoto and A. Yamaguchi, Crystal structures of a multidrug transporter reveal a functionally rotating mechanism, *Nature*, 2006, **443**, 173–179.
- 72 M. A. Seeger, A. Schiefner, T. Eicher, F. Verrey, K. Diederichs and K. M. Pos, Structural asymmetry of AcrB trimer suggests a peristaltic pump mechanism, *Science*, 2006, **313**, 1295–1298.
- 73 J. R. Aires and H. Nikaido, Aminoglycosides are captured from both periplasm and cytoplasm by the AcrD multidrug efflux transporter of *Escherichia coli*, *J Bacteriol*, 2005, **187**, 1923–1929.
- 74 Z. Zhang, C. E. Morgan, M. Cui and E. W. Yu, Cryo-EM Structures of AcrD Illuminate a Mechanism for Capturing Aminoglycosides from Its Central Cavity, *mBio*, 2023, **14**, e03383-22.
- 75 N. Kumar, C.-C. Su, T.-H. Chou, A. Radhakrishnan, J. A. Delmar, K. R. Rajashankar and E. W. Yu, Crystal structures of the *Burkholderia multivorans* hopanoid transporter HpnN, *Proceedings of the National Academy of Sciences*, 2017, **114**, 6557–6562.
- 76 D. M. Doughty, M. L. Coleman, R. C. Hunter, A. L. Sessions, R. E. Summons and D. K. Newman, The RND-family transporter, HpnN, is required for hopanoid localization to the outer membrane of *Rhodopseudomonas palustris* TIE-1, *Proc Natl Acad Sci U S A*, 2011, **108**, E1045-1051.
- 77 X. Li, F. Lu, M. N. Trinh, P. Schmiede, J. Seemann, J. Wang and G. Blobel, 3.3 Å structure of Niemann-Pick C1 protein reveals insights into the function of the C-terminal luminal domain in cholesterol transport, *Proc Natl Acad Sci U S A*, 2017, **114**, 9116–9121.
- 78 G. Degiacomi, A. Benjak, J. Madacki, F. Boldrin, R. Provvedi, G. Palù, J. Kordulakova, S. T. Cole and R. Manganelli, Essentiality of mmpL3 and impact of its silencing on *Mycobacterium tuberculosis* gene expression, *Scientific Reports*, 2017, **7**, 1–8.
- 79 P. Domenech, M. B. Reed and C. E. Barry, Contribution of the *Mycobacterium tuberculosis* MmpL Protein Family to Virulence and Drug Resistance, *Infection and Immunity*, 2005, **73**, 3492–3501.
- 80 W. Li, A. Obregón-Henao, J. B. Wallach, E. J. North, R. E. Lee, M. Gonzalez-Juarrero, D. Schnappinger and M. Jackson, Therapeutic Potential of the *Mycobacterium tuberculosis* Mycolic Acid Transporter, MmpL3, *Antimicrobial Agents and Chemotherapy*, 2016, **60**, 5198–5207.
- 81 C. Varela, D. Rittmann, A. Singh, K. Krumbach, K. Bhatt, L. Eggeling, G. S. Besra and A. Bhatt, MmpL Genes Are Associated with Mycolic Acid Metabolism in *Mycobacteria* and *Corynebacteria*, *Chem Biol*, 2012, **19**, 498–506.
- 82 W. Li, C. M. Stevens, A. N. Pandya, Z. Darzynkiewicz, P. Bhattarai, W. Tong, M. Gonzalez-Juarrero, E. J. North, H. I. Zgurskaya and M. Jackson, Direct Inhibition of MmpL3 by Novel Antitubercular Compounds, *ACS Infect. Dis.*, 2019, **5**, 1001–1012.
- 83 J. T. Williams, E. R. Haiderer, G. B. Coulson, K. N. Conner, E. Ellsworth, C. Chen, N. Alvarez-Cabrera, W. Li, M. Jackson, T. Dick and R. B. Abramovitch, Identification of New MmpL3 Inhibitors by Untargeted and Targeted Mutant Screens Defines MmpL3 Domains with Differential Resistance, *Antimicrobial Agents and Chemotherapy*, 2019, DOI:10.1128/AAC.00547-19.

- 84 M. B. McNeil, D. Dennison and T. Parish, Mutations in MmpL3 alter membrane potential, hydrophobicity and antibiotic susceptibility in *Mycobacterium smegmatis*, *Microbiology*, 2017, **163**, 1065–1070.
- 85 N. Chim, R. Torres, Y. Liu, J. Capri, G. Batot, J. P. Whitelegge and C. W. Goulding, The Structure and Interactions of Periplasmic Domains of Crucial MmpL Membrane Proteins from *Mycobacterium tuberculosis*, *Chemistry & Biology*, 2015, **22**, 1098–1107.
- 86 U. H. Manjunatha and P. W. Smith, Perspective: Challenges and opportunities in TB drug discovery from phenotypic screening, *Bioorganic & Medicinal Chemistry*, 2015, **23**, 5087–5097.
- 87 K. A. Sacksteder, M. Protopopova, C. E. Barry, K. Andries and C. A. Nacy, Discovery and development of SQ109: a new antitubercular drug with a novel mechanism of action, *Future Microbiol*, 2012, **7**, 823–837.
- 88 R. E. Lee, M. Protopopova, E. Crooks, R. A. Slayden, M. Terrot and C. E. Barry, Combinatorial lead optimization of [1,2]-diamines based on ethambutol as potential antituberculosis preclinical candidates, *J Comb Chem*, 2003, **5**, 172–187.
- 89 K. Tahlan, R. Wilson, D. B. Kastrinsky, K. Arora, V. Nair, E. Fischer, S. W. Barnes, J. R. Walker, D. Alland, C. E. Barry and H. I. Boshoff, 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*, 2012, **56**, 1797–1809.
- 90 J. Carbone, N. J. Paradis, L. Bennet, M. C. Alesiani, K. R. Hausman and C. Wu, Inhibition Mechanism of Anti-TB Drug SQ109: Allosteric Inhibition of TMM Translocation of *Mycobacterium Tuberculosis* MmpL3 Transporter, *J. Chem. Inf. Model.*, 2023, **63**, 5356–5374.
- 91 S. P. S. Rao, S. B. Lakshminarayana, R. R. Kondreddi, M. Herve, L. R. Camacho, P. Bifani, S. K. Kalapala, J. Jiricek, N. L. Ma, B. H. Tan, S. H. Ng, M. Nanjundappa, S. Ravindran, P. G. Seah, P. Thayalan, S. H. Lim, B. H. Lee, A. Goh, W. S. Barnes, Z. Chen, K. Gagaring, A. K. Chatterjee, K. Pethe, K. Kuhen, J. Walker, G. Feng, S. Babu, L. Zhang, F. Blasco, D. Beer, M. Weaver, V. Dartois, R. Glynn, T. Dick, P. W. Smith, T. T. Diagana and U. H. Manjunatha, Indolcarboxamide is a preclinical candidate for treating multidrug-resistant tuberculosis, *Sci Transl Med*, 2013, **5**, 214ra168.
- 92 R. Santos, O. Ursu, A. Gaulton, A. P. Bento, R. S. Donadi, C. G. Bologa, A. Karlsson, B. Al-Lazikani, A. Hersey, T. I. Oprea and J. P. Overington, A comprehensive map of molecular drug targets, *Nature Reviews Drug Discovery*, 2017, **16**, 19–34.
- 93 B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter, in *Molecular Biology of the Cell. 4th edition*, Garland Science, 2002.
- 94 K. Renard and B. Byrne, Insights into the Role of Membrane Lipids in the Structure, Function and Regulation of Integral Membrane Proteins, *International Journal of Molecular Sciences*, 2021, **22**, 9026.
- 95 V. Corradi, B. I. Sejdiu, H. Mesa-Galloso, H. Abdizadeh, S. Yu. Noskov, S. J. Marrink and D. P. Tieleman, Emerging Diversity in Lipid–Protein Interactions, *Chem. Rev.*, 2019, **119**, 5775–5848.

- 96 V. Corradi, E. Mendez-Villuendas, H. I. Ingólfsson, R.-X. Gu, I. Siuda, M. N. Melo, A. Moussatova, L. J. DeGagné, B. I. Sejdiu, G. Singh, T. A. Wassenaar, K. Delgado Magnero, S. J. Marrink and D. P. Tieleman, Lipid-Protein Interactions Are Unique Fingerprints for Membrane Proteins, *ACS Cent Sci*, 2018, **4**, 709–717.
- 97 A. Stetsenko and A. Guskov, An Overview of the Top Ten Detergents Used for Membrane Protein Crystallization, *Crystals*, 2017, **7**, 197.
- 98 G. G. Privé, Detergents for the stabilization and crystallization of membrane proteins, *Methods*, 2007, **41**, 388–397.
- 99 J. U. Bowie, Stabilizing membrane proteins, *Curr Opin Struct Biol*, 2001, **11**, 397–402.
- 100 H. J. Lee, H. S. Lee, T. Youn, B. Byrne and P. S. Chae, Impact of novel detergents on membrane protein studies, *Chem*, 2022, **8**, 980–1013.
- 101 J. W. Young, Recent advances in membrane mimetics for membrane protein research, *Biochemical Society Transactions*, 2023, **51**, 1405–1416.
- 102 J. M. Pettersen, Y. Yang and A. S. Robinson, Advances in nanodisc platforms for membrane protein purification, *Trends in Biotechnology*, 2023, **41**, 1041–1054.
- 103 M. Woubshete, S. Cioccolo and B. Byrne, Advances in Membrane Mimetic Systems for Manipulation and Analysis of Membrane Proteins: Detergents, Polymers, Lipids and Scaffolds, *ChemPlusChem*, **n/a**, e202300678.
- 104 C. Tanford, Interfacial free energy and the hydrophobic effect, *Proc Natl Acad Sci U S A*, 1979, **76**, 4175–4176.
- 105 D. Chandler, Interfaces and the driving force of hydrophobic assembly, *Nature*, 2005, **437**, 640–647.
- 106 E. van den Brink-van der Laan, J. Antoinette Killian and B. de Kruijff, Nonbilayer lipids affect peripheral and integral membrane proteins via changes in the lateral pressure profile, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2004, **1666**, 275–288.
- 107 M. F. Brown, Curvature forces in membrane lipid-protein interactions, *Biochemistry*, 2012, **51**, 9782–9795.
- 108 A. G. Lee, How lipids affect the activities of integral membrane proteins, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2004, **1666**, 62–87.
- 109 I. Levental and E. Lyman, Regulation of membrane protein structure and function by their lipid nano-environment, *Nat Rev Mol Cell Biol*, 2023, **24**, 107–122.
- 110 M. S. Bretscher, Asymmetrical Lipid Bilayer Structure for Biological Membranes, *Nature New Biology*, 1972, **236**, 11–12.
- 111 O. H. S. Ollila, M. Louhivuori, S. J. Marrink and I. Vattulainen, Protein Shape Change Has a Major Effect on the Gating Energy of a Mechanosensitive Channel, *Biophysical Journal*, 2011, **100**, 1651–1659.
- 112 R. A. Chaurio, C. Janko, L. E. Muñoz, B. Frey, M. Herrmann and U. S. Gaipf, Phospholipids: key players in apoptosis and immune regulation, *Molecules*, 2009, **14**, 4892–4914.
- 113 T. G. Pomorski and A. K. Menon, LIPID SOMERSAULTS: UNCOVERING THE MECHANISMS OF PROTEIN-MEDIATED LIPID FLIPPING, *Prog Lipid Res*, 2016, **64**, 69–84.

- 114 M. Sinensky, Homeoviscous Adaptation—A Homeostatic Process that Regulates the Viscosity of Membrane Lipids in *Escherichia coli*, *Proc Natl Acad Sci U S A*, 1974, **71**, 522–525.
- 115 Y.-M. Zhang and C. O. Rock, Membrane lipid homeostasis in bacteria, *Nat Rev Microbiol*, 2008, **6**, 222–233.
- 116 J. M. East, D. Melville and A. G. Lee, Exchange rates and numbers of annular lipids for the calcium and magnesium ion dependent adenosinetriphosphatase, *Biochemistry*, 1985, **24**, 2615–2623.
- 117 J. Carney, J. M. East and A. G. Lee, Penetration of lipid chains into transmembrane surfaces of membrane proteins: studies with MscL, *Biophys J*, 2007, **92**, 3556–3563.
- 118 S. B. Hansen, X. Tao and R. MacKinnon, Structural basis of PIP₂ activation of the classical inward rectifier K⁺ channel Kir2.2, *Nature*, 2011, **477**, 495–498.
- 119 P. Marius, S. J. Alvis, J. M. East and A. G. Lee, The interfacial lipid binding site on the potassium channel KcsA is specific for anionic phospholipids, *Biophys J*, 2005, **89**, 4081–4089.
- 120 A. M. Powl, J. M. East and A. G. Lee, Importance of direct interactions with lipids for the function of the mechanosensitive channel MscL, *Biochemistry*, 2008, **47**, 12175–12184.
- 121 A. M. Powl, J. M. East and A. G. Lee, Different Effects of Lipid Chain Length on the Two Sides of a Membrane and the Lipid Annulus of MscL, *Biophysical Journal*, 2007, **93**, 113–122.
- 122 H. Palsdottir and C. Hunte, Lipids in membrane protein structures, *Biochim Biophys Acta*, 2004, **1666**, 2–18.
- 123 M. K. Joshi, S. Dracheva, A. K. Mukhopadhyay, S. Bose and R. W. Hendler, Importance of specific native lipids in controlling the photocycle of bacteriorhodopsin, *Biochemistry*, 1998, **37**, 14463–14470.
- 124 C. Hunte and S. Richers, Lipids and membrane protein structures, *Curr Opin Struct Biol*, 2008, **18**, 406–411.
- 125 E. Pyle, A. C. Kalli, S. Amillis, Z. Hall, A. M. Lau, A. C. Hanyaloglu, G. Diallinas, B. Byrne and A. Politis, Structural Lipids Enable the Formation of Functional Oligomers of the Eukaryotic Purine Symporter UapA, *Cell Chemical Biology*, 2018, **25**, 840-848.e4.
- 126 L. Jodaitis, T. van Oene and C. Martens, Assessing the Role of Lipids in the Molecular Mechanism of Membrane Proteins, *Int J Mol Sci*, 2021, **22**, 7267.
- 127 K. Gupta, J. A. C. Donlan, J. T. S. Hopper, P. Uzdaviny, M. Landreh, W. B. Struwe, D. Drew, A. J. Baldwin, P. J. Stansfeld and C. V. Robinson, The role of interfacial lipids in stabilizing membrane protein oligomers, *Nature*, 2017, **541**, 421–424.
- 128 A. G. Lee, Biological membranes: the importance of molecular detail, *Trends in Biochemical Sciences*, 2011, **36**, 493–500.
- 129 R. Chadda, N. Bernhardt, E. G. Kelley, S. C. Teixeira, K. Griffith, A. Gil-Ley, T. N. Öztürk, L. E. Hughes, A. Forsythe, V. Krishnamani, J. D. Faraldo-Gómez and J. L. Robertson, Membrane transporter dimerization driven by differential lipid solvation energetics of dissociated and associated states, *eLife*, 2021, **10**, e63288.

- 130 C. Lange, J. H. Nett, B. L. Trumpower and C. Hunte, Specific roles of protein–phospholipid interactions in the yeast cytochrome bc₁ complex structure, *EMBO J*, 2001, **20**, 6591–6600.
- 131 A. M. Seddon, P. Curnow and P. J. Booth, Membrane proteins, lipids and detergents: not just a soap opera, *Biochim Biophys Acta*, 2004, **1666**, 105–117.
- 132 C. Ostermeier and H. Michel, Crystallization of membrane proteins, *Curr Opin Struct Biol*, 1997, **7**, 697–701.
- 133 P. S. Chae, S. G. F. Rasmussen, R. R. Rana, K. Gotfryd, R. Chandra, M. A. Goren, A. C. Kruse, S. Nurva, C. J. Loland, Y. Pierre, D. Drew, J.-L. Popot, D. Picot, B. G. Fox, L. Guan, U. Gether, B. Byrne, B. Kobilka and S. H. Gellman, Maltose-neopentyl glycol (MNG) amphiphiles for solubilization, stabilization and crystallization of membrane proteins, *Nat Methods*, 2010, **7**, 1003–1008.
- 134 P. S. Chae, R. R. Rana, K. Gotfryd, S. G. F. Rasmussen, A. C. Kruse, K. H. Cho, S. Capaldi, E. Carlsson, B. Kobilka, C. J. Loland, U. Gether, S. Banerjee, B. Byrne, J. K. Lee and S. H. Gellman, Glucose-Neopentyl Glycol (GNG) Amphiphiles for Membrane Protein Solubilization, Stabilization and Crystallization, *Chem Commun (Camb)*, 2013, **49**, 2287–2289.
- 135 J. L. Parker and S. Newstead, Current trends in α -helical membrane protein crystallization: An update, *Protein Sci*, 2012, **21**, 1358–1365.
- 136 I. G. Denisov, Y. V. Grinkova, A. A. Lazarides and S. G. Sligar, Directed Self-Assembly of Monodisperse Phospholipid Bilayer Nanodiscs with Controlled Size, *J. Am. Chem. Soc.*, 2004, **126**, 3477–3487.
- 137 J. P. Segrest, M. K. Jones, H. De Loof, C. G. Brouillette, Y. V. Venkatachalapathi and G. M. Anantharamaiah, The amphipathic helix in the exchangeable apolipoproteins: a review of secondary structure and function, *J Lipid Res*, 1992, **33**, 141–166.
- 138 M. J. Li, W. M. Atkins and W. D. McClary, Preparation of Lipid Nanodiscs with Lipid Mixtures, *Current Protocols in Protein Science*, 2019, **98**, e100.
- 139 N. Skar-Gislinge, J. B. Simonsen, K. Mortensen, R. Feidenhans'l, S. G. Sligar, B. Lindberg Møller, T. Bjørnholm and L. Arleth, Elliptical Structure of Phospholipid Bilayer Nanodiscs Encapsulated by Scaffold Proteins: Casting the Roles of the Lipids and the Protein, *J. Am. Chem. Soc.*, 2010, **132**, 13713–13722.
- 140 T. K. Ritchie, Y. V. Grinkova, T. H. Bayburt, I. G. Denisov, J. K. Zolnerchiks, W. M. Atkins and S. G. Sligar, in *Methods in Enzymology*, Elsevier, 2009, vol. 464, pp. 211–231.
- 141 I. G. Denisov and S. G. Sligar, Nanodiscs in Membrane Biochemistry and Biophysics, *Chem. Rev.*, 2017, **117**, 4669–4713.
- 142 N. Skar-Gislinge, N. T. Johansen, R. Høiberg-Nielsen and L. Arleth, Comprehensive Study of the Self-Assembly of Phospholipid Nanodiscs: What Determines Their Shape and Stoichiometry?, *Langmuir*, 2018, **34**, 12569–12582.
- 143 Y. Gao, E. Cao, D. Julius and Y. Cheng, TRPV1 structures in nanodiscs reveal mechanisms of ligand and lipid action, *Nature*, 2016, **534**, 347–351.
- 144 W. Mi, Y. Li, S. H. Yoon, R. K. Ernst, T. Walz and M. Liao, Structural basis of MsbA-mediated lipopolysaccharide transport, *Nature*, 2017, **549**, 233–237.
- 145 D. P. Staus, H. Hu, M. J. Robertson, A. L. W. Kleinhenz, L. M. Wingler, W. D. Capel, N. R. Latorraca, R. J. Lefkowitz and G. Skiniotis, Structure of the M2

- muscarinic receptor- β -arrestin complex in a lipid nanodisc, *Nature*, 2020, **579**, 297–302.
- 146 S. Adhikary, D. J. Deredge, A. Nagarajan, L. R. Forrest, P. L. Wintrode and S. K. Singh, Conformational dynamics of a neurotransmitter:sodium symporter in a lipid bilayer, *Proceedings of the National Academy of Sciences*, 2017, **114**, E1786–E1795.
 - 147 M. J. Li, M. Guttman and W. M. Atkins, Conformational dynamics of P-glycoprotein in lipid nanodiscs and detergent micelles reveal complex motions on a wide time scale, *J Biol Chem*, 2018, **293**, 6297–6307.
 - 148 M. F. Peter, K. Bountra, K. Beis and G. Hagelueken, PELDOR/DEER: An Electron Paramagnetic Resonance Method to Study Membrane Proteins in Lipid Bilayers, *Methods Mol Biol*, 2020, **2168**, 313–333.
 - 149 B. J. Lane and C. Pliotas, Approaches for the modulation of mechanosensitive MscL channel pores, *Front Chem*, 2023, **11**, 1162412.
 - 150 C. Kapsalis, Y. Ma, B. E. Bode and C. Pliotas, In-Lipid Structure of Pressure-Sensitive Domains Hints Mechanosensitive Channel Functional Diversity, *Biophysical Journal*, 2020, **119**, 448–459.
 - 151 E. H. Yardeni, S. Mishra, R. A. Stein, E. Bibi and H. S. Mchaourab, The Multidrug Transporter MdfA Deviates from the Canonical Model of Alternating Access of MFS Transporters, *Journal of Molecular Biology*, 2020, **432**, 5665–5680.
 - 152 G. Zhang, M. T. Odenkirk, C. M. Janczak, R. Lee, K. Richardson, Z. Wang, C. A. Aspinwall and M. T. Marty, Identifying Membrane Protein–Lipid Interactions with Lipidomic Lipid Exchange-Mass Spectrometry, *J. Am. Chem. Soc.*, 2023, **145**, 20859–20867.
 - 153 J. Frauenfeld, R. Löving, J.-P. Armache, A. F.-P. Sonnen, F. Guettou, P. Moberg, L. Zhu, C. Jegerschöld, A. Flayhan, J. A. G. Briggs, H. Garoff, C. Löw, Y. Cheng and P. Nordlund, A saposin-lipoprotein nanoparticle system for membrane proteins, *Nat Methods*, 2016, **13**, 345–351.
 - 154 A. Flayhan, H. D. T. Mertens, Y. Ural-Blimke, M. Martinez Molledo, D. I. Svergun and C. Löw, Saposin Lipid Nanoparticles: A Highly Versatile and Modular Tool for Membrane Protein Research, *Structure*, 2018, **26**, 345-355.e5.
 - 155 L. Xiao, L. Han, B. Li, M. Zhang, H. Zhou, Q. Luo, X. Zhang and Y. Huang, Structures of the β -barrel assembly machine recognizing outer membrane protein substrates, *The FASEB Journal*, 2021, **35**, e21207.
 - 156 D.-M. Kehlenbeck, D. A. K. Traore, I. Josts, S. Sander, M. Moulin, M. Haertlein, S. Prevost, V. T. Forsyth and H. Tidow, Cryo-EM structure of MsbA in saposin-lipid nanoparticles (Salipro) provides insights into nucleotide coordination, *The FEBS Journal*, 2022, **289**, 2959–2970.
 - 157 J. A. Lyons, A. Bøggild, P. Nissen and J. Frauenfeld, Saposin-Lipoprotein Scaffolds for Structure Determination of Membrane Transporters, *Methods Enzymol*, 2017, **594**, 85–99.
 - 158 M. L. Carlson, J. W. Young, Z. Zhao, L. Fabre, D. Jun, J. Li, J. Li, H. S. Dhupar, I. Wason, A. T. Mills, J. T. Beatty, J. S. Klassen, I. Rouiller and F. Duong, The Peptidisc, a simple method for stabilizing membrane proteins in detergent-free solution, *eLife*, 2018, **7**, e34085.
 - 159 H. Kariyazono, R. Nadai, R. Miyajima, Y. Takechi-Haraya, T. Baba, A. Shigenaga, K. Okuhira, A. Otaka and H. Saito, Formation of stable nanodiscs by

- bihelical apolipoprotein A-I mimetic peptide, *Journal of Peptide Science*, 2016, **22**, 116–122.
- 160 G. Angiulli, H. S. Dhupar, H. Suzuki, I. S. Wason, F. Duong Van Hoa and T. Walz, New approach for membrane protein reconstitution into peptidiscs and basis for their adaptability to different proteins, *eLife*, 2020, **9**, e53530.
 - 161 K. Rantalainen, Z. T. Berndsen, A. Antanasijevic, T. Schiffner, X. Zhang, W.-H. Lee, J. L. Torres, L. Zhang, A. Irimia, J. Copps, K. H. Zhou, Y. D. Kwon, W. H. Law, C. A. Schramm, R. Verardi, S. J. Krebs, P. D. Kwong, N. A. Doria-Rose, I. A. Wilson, M. B. Zwick, J. R. Yates, W. R. Schief and A. B. Ward, HIV-1 Envelope and MPER Antibody Structures in Lipid Assemblies, *Cell Rep*, 2020, **31**, 107583.
 - 162 J. K. Demmer, B. P. Phillips, O. L. Uhrig, A. Filloux, L. P. Allsopp, M. Bublitz and T. Meier, Structure of ATP synthase from ESKAPE pathogen *Acinetobacter baumannii*, *Science Advances*, 2022, **8**, eabl5966.
 - 163 J. Couston, Z. Guo, K. Wang, P. Gourdon and M. Blaise, Cryo-EM structure of the trehalose monomycolate transporter, MmpL3, reconstituted into peptidiscs, *Current Research in Structural Biology*, 2023, **6**, 100109.
 - 164 J. W. Young, I. S. Wason, Z. Zhao, D. G. Rattray, L. J. Foster and F. Duong Van Hoa, His-Tagged Peptidiscs Enable Affinity Purification of the Membrane Proteome for Downstream Mass Spectrometry Analysis, *J Proteome Res*, 2020, **19**, 2553–2562.
 - 165 B. Dirnberger, D. Korona, R. Popovic, M. J. Deery, H. Barber, S. Russell and K. S. Lilley, Enrichment of Membrane Proteins for Downstream Analysis Using Styrene Maleic Acid Lipid Particles (SMALPs) Extraction, *Bio Protoc*, 2023, **13**, e4728.
 - 166 M. Jamshad, Y.-P. Lin, T. J. Knowles, R. A. Parslow, C. Harris, M. Wheatley, D. R. Poyner, R. M. Bill, O. R. T. Thomas, M. Overduin and T. R. Dafforn, Surfactant-free purification of membrane proteins with intact native membrane environment, *Biochem Soc Trans*, 2011, **39**, 813–818.
 - 167 T. J. Knowles, R. Finka, C. Smith, Y.-P. Lin, T. Dafforn and M. Overduin, Membrane Proteins Solubilized Intact in Lipid Containing Nanoparticles Bounded by Styrene Maleic Acid Copolymer, *J. Am. Chem. Soc.*, 2009, **131**, 7484–7485.
 - 168 M. Xue, L. Cheng, I. Faustino, W. Guo and S. J. Marrink, Molecular Mechanism of Lipid Nanodisk Formation by Styrene-Maleic Acid Copolymers, *Biophys J*, 2018, **115**, 494–502.
 - 169 V. Postis, S. Rawson, J. K. Mitchell, S. C. Lee, R. A. Parslow, T. R. Dafforn, S. A. Baldwin and S. P. Muench, The use of SMALPs as a novel membrane protein scaffold for structure study by negative stain electron microscopy, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2015, **1848**, 496–501.
 - 170 M. Jamshad, J. Charlton, Y.-P. Lin, S. J. Routledge, Z. Bawa, T. J. Knowles, M. Overduin, N. Dekker, T. R. Dafforn, R. M. Bill, D. R. Poyner and M. Wheatley, G-protein coupled receptor solubilization and purification for biophysical analysis and functional studies, in the total absence of detergent, *Biosci Rep*, 2015, **35**, e00188.
 - 171 S. Rehan, V. O. Paavilainen and V.-P. Jaakola, Functional reconstitution of human equilibrative nucleoside transporter-1 into styrene maleic acid co-polymer lipid particles, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2017, **1859**, 1059–1065.

- 172 I. Prabudiansyah, I. Kusters, A. Caforio and A. J. M. Driessen, Characterization of the annular lipid shell of the Sec translocon, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2015, **1848**, 2050–2056.
- 173 W. Qiu, Z. Fu, G. G. Xu, R. A. Grassucci, Y. Zhang, J. Frank, W. A. Hendrickson and Y. Guo, Structure and activity of lipid bilayer within a membrane-protein transporter, *Proceedings of the National Academy of Sciences*, 2018, **115**, 12985–12990.
- 174 J. M. Dörr, S. Scheidelaar, M. C. Koorengevel, J. J. Dominguez, M. Schäfer, C. A. van Walree and J. A. Killian, The styrene–maleic acid copolymer: a versatile tool in membrane research, *Eur Biophys J*, 2016, **45**, 3–21.
- 175 N. L. Pollock, M. Rai, K. S. Simon, S. J. Hesketh, A. C. K. Teo, M. Parmar, P. Sridhar, R. Collins, S. C. Lee, Z. N. Stroud, S. E. Bakker, S. P. Muench, C. H. Barton, G. Hurlbut, D. I. Roper, C. J. I. Smith, T. J. Knowles, C. M. Spickett, J. M. East, V. L. G. Postis and T. R. Dafforn, SMA-PAGE: A new method to examine complexes of membrane proteins using SMALP nano-encapsulation and native gel electrophoresis, *Biochim Biophys Acta Biomembr*, 2019, **1861**, 1437–1445.
- 176 S. C. Lee, S. Khalid, N. L. Pollock, T. J. Knowles, K. Edler, A. J. Rothnie, O. R.T. Thomas and T. R. Dafforn, Encapsulated membrane proteins: A simplified system for molecular simulation, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2016, **1858**, 2549–2557.
- 177 D. J. K. Swainsbury, S. Scheidelaar, N. Foster, R. van Grondelle, J. A. Killian and M. R. Jones, The effectiveness of styrene-maleic acid (SMA) copolymers for solubilisation of integral membrane proteins from SMA-accessible and SMA-resistant membranes, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2017, **1859**, 2133–2143.
- 178 N. L. Pollock, S. C. Lee, J. H. Patel, A. A. Gulamhussein and A. J. Rothnie, Structure and function of membrane proteins encapsulated in a polymer-bound lipid bilayer, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2018, **1860**, 809–817.
- 179 K. Yasuhara, J. Arakida, T. Ravula, S. K. Ramadugu, B. Sahoo, J. Kikuchi and A. Ramamoorthy, Spontaneous Lipid Nanodisc Formation by Amphiphilic Polymethacrylate Copolymers, *J. Am. Chem. Soc.*, 2017, **139**, 18657–18663.
- 180 S. Lavington and A. Watts, Detergent-free solubilisation & purification of a G protein coupled receptor using a polymethacrylate polymer, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2021, **1863**, 183441.
- 181 A. O. Oluwole, J. Klingler, B. Danielczak, J. O. Babalola, C. Vargas, G. Pabst and S. Keller, Formation of Lipid-Bilayer Nanodiscs by Diisobutylene/Maleic Acid (DIBMA) Copolymer, *Langmuir*, 2017, **33**, 14378–14388.
- 182 Clare. R. Harwood, D. A. Sykes, B. L. Hoare, F. M. Heydenreich, R. Uddin, D. R. Poyner, S. J. Briddon and D. B. Veprintsev, Functional solubilization of the β 2-adrenoceptor using diisobutylene maleic acid, *iScience*, 2021, **24**, 103362.
- 183 T. Ravula and A. Ramamoorthy, Synthesis, Characterization, and Nanodisc Formation of Non-ionic Polymers**, *Angewandte Chemie*, 2021, **133**, 17022–17025.
- 184 B. Krishnarjuna, S.-C. Im, T. Ravula, J. Marte, R. J. Auchus and A. Ramamoorthy, Non-Ionic Inulin-Based Polymer Nanodiscs Enable Functional

- Reconstitution of a Redox Complex Composed of Oppositely Charged CYP450 and CPR in a Lipid Bilayer Membrane, *Anal. Chem.*, 2022, **94**, 11908–11915.
- 185 L. M. Real Hernandez and I. Levental, Lipid packing is disrupted in copolymeric nanodiscs compared with intact membranes, *Biophys J*, 2023, **122**, 2256–2266.
 - 186 H. Matsumoto, H. Haniu and N. Komori, Determination of Protein Molecular Weights on SDS-PAGE, *Methods Mol Biol*, 2019, **1855**, 101–105.
 - 187 I. Boulos, J. Jabbour, S. Khoury, N. Mikhael, V. Tishkova, N. Candoni, H. E. Ghadieh, S. Veessler, Y. Bassim, S. Azar and F. Harb, Exploring the World of Membrane Proteins: Techniques and Methods for Understanding Structure, Function, and Dynamics, *Molecules*, 2023, **28**, 7176.
 - 188 F. DI Girolamo, M. Ponzi, M. Crescenzi, J. Alessandrini and F. Guadagni, A simple and effective method to analyze membrane proteins by SDS-PAGE and MALDI mass spectrometry, *Anticancer Res*, 2010, **30**, 1121–1129.
 - 189 Y. Hayashi, H. Matsui and T. Takagi, Membrane protein molecular weight determined by low-angle laser light-scattering photometry coupled with high-performance gel chromatography, *Methods Enzymol*, 1989, **172**, 514–528.
 - 190 H. Zheng, K. B. Handing, M. D. Zimmerman, I. G. Shabalin, S. C. Almo and W. Minor, X-ray crystallography over the past decade for novel drug discovery – where are we heading next?, *Expert Opin Drug Discov*, 2015, **10**, 975–989.
 - 191 A. A. Kermani, A guide to membrane protein X-ray crystallography, *FEBS J*, 2021, **288**, 5788–5804.
 - 192 D. Lyumkis, Challenges and opportunities in cryo-EM single-particle analysis, *J Biol Chem*, 2019, **294**, 5181–5197.
 - 193 K. Murata and M. Wolf, Cryo-electron microscopy for structural analysis of dynamic biological macromolecules, *Biochim Biophys Acta Gen Subj*, 2018, **1862**, 324–334.
 - 194 K. Wentinck, C. Gogou and D. H. Meijer, Putting on molecular weight: Enabling cryo-EM structure determination of sub-100-kDa proteins, *Curr Res Struct Biol*, 2022, **4**, 332–337.
 - 195 K. M. Yip, N. Fischer, E. Paknia, A. Chari and H. Stark, Atomic-resolution protein structure determination by cryo-EM, *Nature*, 2020, **587**, 157–161.
 - 196 P. A. Klenotic, C. E. Morgan and E. W. Yu, Cryo-EM as a tool to study bacterial efflux systems and the membrane proteome, *Fac Rev*, 2021, **10**, 24.
 - 197 C.-Y. Huang, P. Draczowski, Y.-S. Wang, C.-Y. Chang, Y.-C. Chien, Y.-H. Cheng, Y.-M. Wu, C.-H. Wang, Y.-C. Chang, Y.-C. Chang, T.-J. Yang, Y.-X. Tsai, K.-H. Khoo, H.-W. Chang and S.-T. D. Hsu, In situ structure and dynamics of an alphacoronavirus spike protein by cryo-ET and cryo-EM, *Nat Commun*, 2022, **13**, 4877.
 - 198 S. G. Patching, Solid-state NMR structures of integral membrane proteins, *Mol Membr Biol*, 2015, **32**, 156–178.
 - 199 J. Thoma and B. M. Burmann, High-Resolution In Situ NMR Spectroscopy of Bacterial Envelope Proteins in Outer Membrane Vesicles, *Biochemistry*, 2020, **59**, 1656–1660.
 - 200 R. Puthenveetil and O. Vinogradova, Solution NMR: A powerful tool for structural and functional studies of membrane proteins in reconstituted environments, *J Biol Chem*, 2019, **294**, 15914–15931.

- 201 S. Sun, Y. Han, S. Paramasivam, S. Yan, A. E. Siglin, J. C. Williams, I.-J. L. Byeon, J. Ahn, A. M. Gronenborn and T. Polenova, Solid-State NMR Spectroscopy of Protein Complexes, *Methods Mol Biol*, 2012, **831**, 303–331.
- 202 Y. Nishiyama, G. Hou, V. Agarwal, Y. Su and A. Ramamoorthy, Ultrafast Magic Angle Spinning Solid-State NMR Spectroscopy: Advances in Methodology and Applications, *Chem Rev*, 2023, **123**, 918–988.
- 203 F. S. Ruggeri, T. Šneideris, M. Vendruscolo and T. P. J. Knowles, Atomic force microscopy for single molecule characterisation of protein aggregation, *Arch Biochem Biophys*, 2019, **664**, 134–148.
- 204 R. R. Sanganna Gari, J. J. Montalvo-Acosta, G. R. Heath, Y. Jiang, X. Gao, C. M. Nimigean, C. Chipot and S. Scheuring, Correlation of membrane protein conformational and functional dynamics, *Nat Commun*, 2021, **12**, 4363.
- 205 Z. N. Scholl and P. E. Marszalek, AFM-Based Single-Molecule Force Spectroscopy of Proteins, *Methods Mol Biol*, 2018, **1814**, 35–47.
- 206 I. D. Medalsy and D. J. Müller, Nanomechanical Properties of Proteins and Membranes Depend on Loading Rate and Electrostatic Interactions, *ACS Nano*, 2013, **7**, 2642–2650.
- 207 H. Nakagawa, T. Saio, M. Nagao, R. Inoue, M. Sugiyama, S. Ajito, T. Tominaga and Y. Kawakita, Conformational dynamics of a multidomain protein by neutron scattering and computational analysis, *Biophys J*, 2021, **120**, 3341–3354.
- 208 T. O. C. Kwan, R. Reis, G. Siligardi, R. Hussain, H. Cheruvara and I. Moraes, Selection of Biophysical Methods for Characterisation of Membrane Proteins, *Int J Mol Sci*, 2019, **20**, 2605.
- 209 C. Breyton, C. Tribet, J. Olive, J. P. Dubacq and J. L. Popot, Dimer to monomer conversion of the cytochrome b6 f complex. Causes and consequences, *J Biol Chem*, 1997, **272**, 21892–21900.
- 210 Y. Subburaj, K. Cosentino, M. Axmann, E. Pedrueza-Villalmanzo, E. Hermann, S. Bleicken, J. Spatz and A. J. García-Sáez, Bax monomers form dimer units in the membrane that further self-assemble into multiple oligomeric species, *Nat Commun*, 2015, **6**, 8042.
- 211 E. Sahin and C. J. Roberts, Size-exclusion chromatography with multi-angle light scattering for elucidating protein aggregation mechanisms, *Methods Mol Biol*, 2012, **899**, 403–423.
- 212 A. C. Leney and A. J. R. Heck, Native Mass Spectrometry: What is in the Name?, *J. Am. Soc. Mass Spectrom.*, 2017, **28**, 5–13.
- 213 J. E. Keener, G. Zhang and M. T. Marty, Native Mass Spectrometry of Membrane Proteins, *Anal. Chem.*, 2021, **93**, 583–597.
- 214 X. Han and R. W. Gross, Shotgun lipidomics: electrospray ionization mass spectrometric analysis and quantitation of cellular lipidomes directly from crude extracts of biological samples, *Mass Spectrom Rev*, 2005, **24**, 367–412.
- 215 A. Laganowsky, E. Reading, J. T. S. Hopper and C. V. Robinson, Mass Spectrometry of Intact Membrane Protein Complexes, *Nat Protoc*, 2013, **8**, 639–651.
- 216 A. Laganowsky, E. Reading, T. M. Allison, M. B. Ulmschneider, M. T. Degiacomi, A. J. Baldwin and C. V. Robinson, Membrane proteins bind lipids selectively to modulate their structure and function, *Nature*, 2014, **510**, 172–175.

- 217 N. Morgner, H.-D. Barth and B. Brutschy, A New Way To Detect Noncovalently Bonded Complexes of Biomolecules from Liquid Micro-Droplets by Laser Mass Spectrometry, *Aust. J. Chem.*, 2006, **59**, 109–114.
- 218 E. Henrich, O. Peetz, C. Hein, A. Laguerre, B. Hoffmann, J. Hoffmann, V. Dötsch, F. Bernhard and N. Morgner, Analyzing native membrane protein assembly in nanodiscs by combined non-covalent mass spectrometry and synthetic biology, *eLife*, 2017, **6**, e20954.
- 219 N. Hellwig, O. Peetz, Z. Ahdash, I. Tascón, P. J. Booth, V. Mikusevic, M. Diskowski, A. Politis, Y. Hellmich, I. Hänel, E. Reading and N. Morgner, Native mass spectrometry goes more native: investigation of membrane protein complexes directly from SMALPs †Electronic supplementary information (ESI) available: Experimental details, additional information on methods. See DOI: 10.1039/c8cc06284f, *Chem Commun (Camb)*, 2018, **54**, 13702–13705.
- 220 M. T. Marty, Nanodiscs and mass spectrometry: Making membranes fly, *International Journal of Mass Spectrometry*, 2020, **458**, 116436.
- 221 O. Peetz, N. Hellwig, E. Henrich, J. Mezhyrova, V. Dötsch, F. Bernhard and N. Morgner, LILBID and nESI: Different Native Mass Spectrometry Techniques as Tools in Structural Biology, *J Am Soc Mass Spectrom*, 2019, **30**, 181–191.
- 222 G. Edwards, U. Muthurajan, S. Bowerman and K. Luger, Analytical Ultracentrifugation (AUC): An Overview of the Application of Fluorescence and Absorbance AUC to the Study of Biological Macromolecules, *Current Protocols in Molecular Biology*, 2020, 133, e131. DOI:10.1002/cpmb.131.
- 223 J. Cole and J. Hansen, Analytical ultracentrifugation as a contemporary biomolecular research tool, *Journal of Biomolecular Techniques : JBT*, 1999, **10**, 163.
- 224 H. Zhao, C. A. Brautigam, R. Ghirlando and P. Schuck, Current Methods in Sedimentation Velocity and Sedimentation Equilibrium Analytical Ultracentrifugation, *Curr Protoc Protein Sci*, 2013, **20**, 10.1002/0471140864.ps2012s71.
- 225 J. L. Cole, J. W. Lary, T. Moody and T. M. Laue, Analytical Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium, *Methods Cell Biol*, 2008, **84**, 143–179.
- 226 U. B. Ericsson, B. M. Hallberg, G. T. DeTitta, N. Dekker and P. Nordlund, Thermofluor-based high-throughput stability optimization of proteins for structural studies, *Analytical Biochemistry*, 2006, **357**, 289–298.
- 227 M. Kohlstaedt, I. von der Hocht, F. Hilbers, Y. Thielmann and H. Michel, Development of a Thermofluor assay for stability determination of membrane proteins using the Na⁺/H⁺ antiporter NhaA and cytochrome c oxidase, *Acta Cryst D*, 2015, **71**, 1112–1122.
- 228 J. Fan, J. Heng, S. Dai, N. Shaw, B. Zhou, B. Huang, Z. He, Y. Wang, T. Jiang, X. Li, Z. Liu, X. Wang and X. C. Zhang, An efficient strategy for high throughput screening of recombinant integral membrane protein expression and stability, *Protein Expression and Purification*, 2011, **78**, 6–13.
- 229 F. H. Niesen, H. Berglund and M. Vedadi, The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability, *Nat Protoc*, 2007, **2**, 2212–2221.

- 230 A. I. Alexandrov, M. Mileni, E. Y. T. Chien, M. A. Hanson and R. C. Stevens, Microscale fluorescent thermal stability assay for membrane proteins, *Structure*, 2008, **16**, 351–359.
- 231 H. Svilenov, U. Markoja and G. Winter, Isothermal chemical denaturation as a complementary tool to overcome limitations of thermal differential scanning fluorimetry in predicting physical stability of protein formulations, *European Journal of Pharmaceutics and Biopharmaceutics*, 2018, **125**, 106–113.
- 232 C. G. Alexander, R. Wanner, C. M. Johnson, D. Breitsprecher, G. Winter, S. Duhr, P. Baaske and N. Ferguson, Novel microscale approaches for easy, rapid determination of protein stability in academic and commercial settings, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 2014, **1844**, 2241–2250.
- 233 S. J. Marrink, V. Corradi, P. C. T. Souza, H. I. Ingólfsson, D. P. Tieleman and M. S. P. Sansom, Computational Modeling of Realistic Cell Membranes, *Chem Rev*, 2019, **119**, 6184–6226.
- 234 S. A. Hollingsworth and R. O. Dror, Molecular dynamics simulation for all, *Neuron*, 2018, **99**, 1129–1143.
- 235 M. A. González, Force fields and molecular dynamics simulations, *JDN*, 2011, **12**, 169–200.
- 236 G. Wieczorek and D. Niedzialek, Molecular Dynamics. In *eLS*, John Wiley & Sons, Ltd, 2020, pp. 1–18.
- 237 T. Werner, M. B. Morris, S. Dastmalchi and W. B. Church, Structural modelling and dynamics of proteins for insights into drug interactions, *Advanced Drug Delivery Reviews*, 2012, **64**, 323–343.
- 238 S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid and A. Kolinski, Coarse-Grained Protein Models and Their Applications, *Chem Rev*, 2016, **116**, 7898–7936.
- 239 B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, CHARMM: The biomolecular simulation program, *Journal of Computational Chemistry*, 2009, **30**, 1545–1614.
- 240 D. A. CASE, T. E. CHEATHAM, T. DARDEN, H. GOHLKE, R. LUO, K. M. MERZ, A. ONUFRIEV, C. SIMMERLING, B. WANG and R. J. WOODS, The Amber Biomolecular Simulation Programs, *J Comput Chem*, 2005, **26**, 1668–1688.
- 241 D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, GROMACS: fast, flexible, and free, *J Comput Chem*, 2005, **26**, 1701–1718.
- 242 C. Tian, K. Kasavajhala, K. A. A. Belfon, L. Raguette, H. Huang, A. N. Miguels, J. Bickel, Y. Wang, J. Pincay, Q. Wu and C. Simmerling, ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution, *J. Chem. Theory Comput.*, 2020, **16**, 528–552.
- 243 C. J. Dickson, R. C. Walker and I. R. Gould, Lipid21: Complex Lipid Membrane Simulations with AMBER, *J. Chem. Theory Comput.*, 2022, **18**, 1726–1736.

- 244 L.-P. Wang, K. A. McKiernan, J. Gomes, K. A. Beauchamp, T. Head-Gordon, J. E. Rice, W. C. Swope, T. J. Martínez and V. S. Pande, Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15, *J Phys Chem B*, 2017, **121**, 4023–4039.
- 245 E. Rathi, A. Kumar and S. G. Kini, Computational approaches in efflux pump inhibitors: current status and prospects, *Drug Discovery Today*, 2020, **25**, 1883–1890.
- 246 S. Jamshidi, J. M. Sutton and K. M. Rahman, Computational Study Reveals the Molecular Mechanism of the Interaction between the Efflux Inhibitor PA β N and the AdeB Transporter from *Acinetobacter baumannii*, *ACS Omega*, 2017, **2**, 3002–3016.
- 247 A. V. Vargiu, V. K. Ramaswamy, G. Mallocci, I. Malvacio, A. Atzori and P. Ruggerone, Computer simulations of the activity of RND efflux pumps, *Research in Microbiology*, 2018, **169**, 384–392.
- 248 A. V. Vargiu and H. Nikaido, Multidrug binding properties of the AcrB efflux pump characterized by molecular dynamics simulations, *Proceedings of the National Academy of Sciences*, 2012, **109**, 20637–20642.
- 249 K. R. Pothula, C. J. F. Solano and U. Kleinekathöfer, Simulations of outer membrane channels and their permeability, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2016, **1858**, 1760–1771.
- 250 P. Sandhu and Y. Akhter, The drug binding sites and transport mechanism of the RND pumps from *Mycobacterium tuberculosis*: Insights from molecular dynamics simulations, *Archives of Biochemistry and Biophysics*, 2016, **592**, 38–49.
- 251 A. Ece, Computer-aided drug design, *BMC Chem*, 2023, **17**, 26.
- 252 S. J. Y. Macalino, S. Basith, N. A. B. Clavio, H. Chang, S. Kang and S. Choi, Evolution of In Silico Strategies for Protein-Protein Interaction Drug Discovery, *Molecules*, 2018, **23**, 1963.
- 253 B. C. Choy, R. J. Cater, F. Mancina and E. E. Pryor, A 10-year meta-analysis of membrane protein structural biology: Detergents, membrane mimetics, and structure determination techniques, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2021, **1863**, 183533.
- 254 S. G. Sligar and I. G. Denisov, Nanodiscs: A toolkit for membrane protein science, *Protein Sci*, 2021, **30**, 297–315.
- 255 I. G. Denisov and S. G. Sligar, Nanodiscs for structural and functional studies of membrane proteins, *Nat Struct Mol Biol*, 2016, **23**, 481–486.
- 256 D. Hanahan, J. Jessee and F. R. Bloom, in *Methods in Enzymology*, Academic Press, 1991, vol. 204, pp. 63–113.
- 257 Z. Wang, C. Ye, X. Zhang and Y. Wei, Cysteine residue is not essential for CPM protein thermal-stability assay, *Anal Bioanal Chem*, 2015, **407**, 3683–3691.
- 258 Christopher A. P. Glover, AcrB contamination in 2-D crystallization of membrane proteins: Lessons from a sodium channel and a putative monovalent cation/proton antiporter, *Journal of Structural Biology*, 2011, **176**, 419–424.
- 259 D. Veesler, S. Blangy, C. Cambillau and G. Sciara, There is a baby in the bath water: AcrB contamination is a major problem in membrane-protein crystallization, *Acta Crystallogr Sect F Struct Biol Cryst Commun*, 2008, **64**, 880–885.

- 260 G. Psakis, J. Polaczek and L.-O. Essen, AcrB et al.: Obstinate contaminants in a picogram scale. One more bottleneck in the membrane protein structure pipeline, *J Struct Biol*, 2009, **166**, 107–111.
- 261 H. Yamanaka, H. Kobayashi, E. Takahashi and K. Okamoto, MacAB Is Involved in the Secretion of Escherichia coli Heat-Stable Enterotoxin II, *J Bacteriol*, 2008, **190**, 7693–7698.
- 262 S. Jo, T. Kim, V. G. Iyer and W. Im, CHARMM-GUI: A web-based graphical user interface for CHARMM, *Journal of Computational Chemistry*, 2008, **29**, 1859–1865.
- 263 Y. Qi, J. Lee, J. B. Klauda and W. Im, CHARMM-GUI Nanodisc Builder for modeling and simulation of various nanodisc systems, *J Comput Chem*, 2019, **40**, 893–899.
- 264 E. Strandberg, D. Tiltak, S. Ehni, P. Wadhwani and A. S. Ulrich, Lipid shape is a key factor for membrane interactions of amphipathic helical peptides, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2012, **1818**, 1764–1776.
- 265 I. Boulos, J. Jabbour, S. Khoury, N. Mikhael, V. Tishkova, N. Candoni, H. E. Ghadieh, S. Veessler, Y. Bassim, S. Azar and F. Harb, Exploring the World of Membrane Proteins: Techniques and Methods for Understanding Structure, Function, and Dynamics, *Molecules*, 2023, **28**, 7176.
- 266 A. Olerinyova, A. Sonn-Segev, J. Gault, C. Eichmann, J. Schimpf, A. H. Kopf, L. S. P. Rudden, D. Ashkinadze, R. Bomba, L. Frey, J. Greenwald, M. T. Degiacomi, R. Steinhilper, J. A. Killian, T. Friedrich, R. Riek, W. B. Struwe and P. Kukura, Mass Photometry of Membrane Proteins, *Chem*, 2021, **7**, 224–236.
- 267 S. Inagaki and R. Ghirlando, Nanodisc characterization by Analytical ultracentrifugation, *Nanotechnology Reviews*, 2017; **6**(1): 3–14. DOI:10.1515/ntrev-2016-0082.
- 268 S. Inagaki, R. Ghirlando and R. Grisshammer, Biophysical characterization of membrane proteins in nanodiscs, *Methods*, 2013, **59**, 287–300.
- 269 T. Laue, B. Shah, T. Ridgeway and S. Pelletier, Computer-aided Interpretation of Sedimentation Data for Proteins, *Analytical ultracentrifugation in biochemistry and polymer science*.
- 270 P. Schuck, M. A. Perugini, N. R. Gonzales, G. J. Howlett and D. Schubert, Size-Distribution Analysis of Proteins by Analytical Ultracentrifugation: Strategies and Application to Model Systems, *Biophysical Journal*, 2002, **82**, 1096–1111.
- 271 C. A. Brautigam, in *Methods in Enzymology*, ed. J. L. Cole, Academic Press, 2015, vol. 562, pp. 109–133.
- 272 D. Wu and G. Piszczek, Standard Protocol for Mass Photometry Experiments, *Eur Biophys J*, 2021, **50**, 403–409.
- 273 G. Young, N. Hundt, D. Cole, A. Fineberg, J. Andrecka, A. Tyler, A. Olerinyova, A. Ansari, E. G. Marklund, M. P. Collier, S. A. Chandler, O. Tkachenko, J. Allen, M. Crispin, N. Billington, Y. Takagi, J. R. Sellers, C. Eichmann, P. Selenko, L. Frey, R. Riek, M. R. Galpin, W. B. Struwe, J. L. P. Benesch and P. Kukura, Quantitative mass imaging of single biological macromolecules, *Science*, 2018, **360**, 423–427.
- 274 D. Cole, G. Young, A. Weigel, A. Sebesta and P. Kukura, Label-Free Single-Molecule Imaging with Numerical-Aperture-Shaped Interferometric Scattering Microscopy, *ACS Photonics*, 2017, **4**, 211–216.

- 275 F. Soltermann, E. D. B. Foley, V. Pagnoni, M. Galpin, J. L. P. Benesch, P. Kukura and W. B. Struwe, Quantifying Protein–Protein Interactions by Molecular Counting with Mass Photometry, *Angew Chem Int Ed Engl*, 2020, **59**, 10774–10779.
- 276 N. Morgner and C. V. Robinson, Massign: An Assignment Strategy for Maximizing Information from the Mass Spectra of Heterogeneous Protein Assemblies, *Anal. Chem.*, 2012, **84**, 2939–2948.
- 277 A. Punjani, J. L. Rubinstein, D. J. Fleet and M. A. Brubaker, cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination, *Nat Methods*, 2017, **14**, 290–296.
- 278 J. Folch, M. Lees and G. H. Sloane Stanley, A simple method for the isolation and purification of total lipides from animal tissues, *J Biol Chem*, 1957, **226**, 497–509.
- 279 E. Fahy, M. Sud, D. Cotter and S. Subramaniam, LIPID MAPS online tools for lipid research, *Nucleic Acids Research*, 2007, **35**, W606–W612.
- 280 Z. Stroud, d_ph, University of Birmingham, 2020.
- 281 H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, The Protein Data Bank, *Nucleic Acids Res.*, 2000, **28**, 235–242.
- 282 S. Gore, E. Sanz García, P. M. S. Hendrickx, A. Gutmanas, J. D. Westbrook, H. Yang, Z. Feng, K. Baskaran, J. M. Berrisford, B. P. Hudson, Y. Ikegawa, N. Kobayashi, C. L. Lawson, S. Mading, L. Mak, A. Mukhopadhyay, T. J. Oldfield, A. Patwardhan, E. Peisach, G. Sahni, M. R. Sekharan, S. Sen, C. Shao, O. S. Smart, E. L. Ulrich, R. Yamashita, M. Quesada, J. Y. Young, H. Nakamura, J. L. Markley, H. M. Berman, S. K. Burley, S. Velankar and G. J. Kleywegt, Validation of Structures in the Protein Data Bank, *Structure*, 2017, **25**, 1916–1927.
- 283 F. Sievers and D. G. Higgins, Clustal Omega for making accurate alignments of many protein sequences, *Protein Sci.*, 2018, **27**, 135–145.
- 284 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, UCSF Chimera--a visualization system for exploratory research and analysis, *J Comput Chem*, 2004, **25**, 1605–1612.
- 285 E. C. Meng, E. F. Pettersen, G. S. Couch, C. C. Huang and T. E. Ferrin, Tools for integrated sequence-structure analysis with UCSF Chimera, *BMC Bioinformatics*, 2006, **7**, 339.
- 286 A. Sali and T. L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, *J. Mol. Biol.*, 1993, **234**, 779–815.
- 287 F. Melo, R. Sánchez and A. Sali, Statistical potentials for fold assessment, *Protein Sci*, 2002, **11**, 430–448.
- 288 M.-Y. Shen and A. Sali, Statistical potential for assessment and prediction of protein structures, *Protein Sci*, 2006, **15**, 2507–2524.
- 289 A. Drozdetskiy, C. Cole, J. Procter and G. J. Barton, JPred4: a protein secondary structure prediction server, *Nucleic Acids Res*, 2015, **43**, W389–W394.
- 290 C. Geourjon and G. Deleage, SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments, *Computer applications in the biosciences: CABIOS*, 1995, 11(6):681-4. doi: 10.1093/bioinformatics/11.6.681.

- 291 N. Metropolis and S. Ulam, The Monte Carlo Method, *Journal of the American Statistical Association*, 1949, **44**, 335–341.
- 292 M. Baek, T. Park, L. Heo, C. Park and C. Seok, GalaxyHomomer: a web server for protein homo-oligomer structure prediction from a monomer sequence or structure, *Nucleic Acids Res*, 2017, **45**, W320–W324.
- 293 R. Nakashima, K. Sakurai, S. Yamasaki, K. Hayashi, C. Nagata, K. Hoshino, Y. Onodera, K. Nishino and A. Yamaguchi, Structural basis for the inhibition of bacterial multidrug exporters, *Nature*, 2013, **500**, 102–106.
- 294 E. Krissinel and K. Henrick, Inference of macromolecular assemblies from crystalline state, *J Mol Biol*, 2007, **372**, 774–797.
- 295 T. Kortemme and D. Baker, A simple physical model for binding energy hot spots in protein–protein complexes, *Proceedings of the National Academy of Sciences*, 2002, **99**, 14116–14121.
- 296 T. Kortemme, D. E. Kim and D. Baker, Computational Alanine Scanning of Protein-Protein Interfaces, *Science's STKE*, 2004, **2004**, pl2–pl2.
- 297 D.A. Case, R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao and P.A. Kollman (2016), AMBER 2016, University of California, San Francisco.
- 298 D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. SalomonFerrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman (2018), AMBER 2018, University of California, San Francisco.
- 299 R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand and R. C. Walker, Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald, *J Chem Theory Comput*, 2013, **9**, 3878–3888.
- 300 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, Comparison of simple potential functions for simulating liquid water, *The Journal of chemical physics*, 1983. Vol. 79(2):926. DOI: 10.1063/1.445869.
- 301 IR Gould, AA Skjevik, CJ Dickson, BD Madej, RC Walker, Lipid17: A Comprehensive AMBER Force Field for the Simulation of Zwitterionic and Anionic Lipids, *in prep*.
- 302 S. Miyamoto and P. A. Kollman, Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models, *Journal of Computational Chemistry*, 1992, **13**, 952–962.
- 303 G. Bussi and M. Parrinello, Accurate sampling using Langevin dynamics, *Phys. Rev. E*, 2007, **75**, 056707.

- 304 C. W. Hopkins, S. Le Grand, R. C. Walker and A. E. Roitberg, Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning, *J. Chem. Theory Comput.*, 2015, **11**, 1864–1874.
- 305 D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. Salomon- Ferrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman, (2018) AMBER 2018, *University of California, San Francisco*.
- 306 D. R. Roe and T. E. Cheatham, PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data, *J. Chem. Theory Comput.*, 2013, **9**, 3084–3095.
- 307 S. Cioccolo, N. Brooks, B. Byrne and I. R. Gould, *Molecular Dynamics simulations as a novel approach for the characterisation of Mycobacterium smegmatis membrane pro-teins MmpL3*, Imperial College London, 2020.
- 308 J. D. Hunter, Matplotlib: A 2D Graphics Environment, *Computing in Science & Engineering*, 2007, **9**, 90–95.
- 309 VMD: Visual molecular dynamics - ScienceDirect, <https://www.sciencedirect.com/science/article/abs/pii/0263785596000185>, (accessed 10 June 2021).
- 310 D. R. Roe and T. E. Cheatham, PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data, *J. Chem. Theory Comput.*, 2013, **9**, 3084–3095.
- 311 O. Carugo, How large B-factors can be in protein crystal structures, *BMC Bioinformatics*, 2018, **19**, 61.
- 312 R. Cárdenas, J. Martínez-Seoane and C. Amero, Combining Experimental Data and Computational Methods for the Non-Computer Specialist, *Molecules*, 2020, **25**, 4783.
- 313 C.-C. Su, L. Yin, N. Kumar, L. Dai, A. Radhakrishnan, J. R. Bolla, H.-T. Lei, T.-H. Chou, J. A. Delmar, K. R. Rajashankar, Q. Zhang, Y.-K. Shin and E. W. Yu, Structures and transport dynamics of a Campylobacter jejuni multidrug efflux pump, *Nat Commun*, 2017, **8**, 171.
- 314 F. Long, C.-C. Su, M. T. Zimmermann, S. E. Boyken, K. R. Rajashankar, R. L. Jernigan and E. W. Yu, Crystal structures of the CusA efflux pump suggest methionine-mediated metal transport, *Nature*, 2010, **467**, 484–488.
- 315 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.
- 316 D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. DeBolt, D. Ferguson, G. Seibel and P. Kollman, AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules, *Computer Physics Communications*, 1995, **91**, 1–41.

- 317 Å. A. Skjevik, B. D. Madej, R. C. Walker and K. Teigen, LIPID11: A Modular Framework for Lipid Simulations Using Amber, *J. Phys. Chem. B*, 2012, **116**, 11124–11136.
- 318 C. J. Dickson, B. D. Madej, Å. A. Skjevik, R. M. Betz, K. Teigen, I. R. Gould and R. C. Walker, Lipid14: The Amber Lipid Force Field, *J. Chem. Theory Comput.*, 2014, **10**, 865–879.

7. Appendix: Copyrights and permissions

Figure number	Source	Copyright holder	Permission	Licence number	Approval number
1.1	https://doi.org/10.1038/nri1666	Springer Nature	Granted	5787211504765	1
1.2	https://www.who.int/publications/i/item/9789240083851	World Health Organization	Open Access (CC BY-NC-SA 3.0 IGO DEED)	N/A	N/A
1.3	https://doi.org/10.1016/j.str.2021.06.013	Elsevier	Open Access (CC BY 4.0)	N/A	N/A
1.5	https://doi.org/10.1016/j.cell.2019.01.003	Elsevier	Granted	5787230122307	2
1.7	https://doi.org/10.1002/cplu.202300678	John Wiley and Sons	Open Access (CC BY)	N/A	N/A
1.10, 1.11	https://doi.org/10.1038/s41580-022-00524-4	Springer Nature	Granted	5787540361094	3
1.12	https://doi.org/10.1002/cplu.202300678	John Wiley and Sons	Open Access (CC BY)	N/A	N/A
1.14	https://doi.org/10.1016/j.addr.2011.11.011	Elsevier	Granted	5787541073347	4
1.15, front cover image	https://doi.org/10.1021/acs.chemrev.8b00460	American Chemical Society	Open Access (CC-BY-NC-ND)	N/A	N/A

7.1 Approval 1

Order Completed

Thank you for your order.

This Agreement between Imperial College London – Sara Cioccolo ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number	5787211504765	Printable Details	
License date	May 13, 2024		
Licensed Content		Order Details	
Licensed Content Publisher	Springer Nature	Type of Use	Thesis/Dissertation
Licensed Content Publication	Nature Reviews Immunology	Requestor type	academic/university or research institute
Licensed Content Title	Immune responses to tuberculosis in developing countries: Implications for new vaccines	Format	print and electronic
Licensed Content Author	Graham A. W. Rook et al	Portion	figures/tables/illustrations
Licensed Content Date	Aug 1, 2005	Number of figures/tables/illustrations	1
		Would you like a high resolution image with your order?	no
		Will you be translating?	no
		Circulation/distribution	50000 or greater
		Author of this Springer Nature content	no
About Your Work		Additional Data	
Title of new work	Investigation of the oligomeric state of the mycobacterial lipid transporter, MmpL3: biophysical and computational insights.	Portions	Figure 1
Institution name	Imperial College London		
Expected presentation date	May 2024		
Requestor Location		Tax Details	
	Imperial College London Molecular Sciences Research Hub		
Requestor Location	London, W12 0BZ United Kingdom Attn: Imperial College London		

7.2 Approval 2

Order Completed

Thank you for your order.

This Agreement between Imperial College London – Sara Cioccoli ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number	5787230122307
License date	May 13, 2024

■ Licensed Content

Licensed Content Publisher	Elsevier
Licensed Content Publication	Cell
Licensed Content Title	Crystal Structures of Membrane Transporter Mmpl3, an Anti-TB Drug Target
Licensed Content Author	Bing Zhang,Jun Li,Xiaolin Yang,Lijie Wu,Jia Zhang,Yang Yang,Yao Zhao,Lu Zhang,Xiuna Yang,Xiaobao Yang,Xi Cheng,Zhijie Liu,Biao Jiang,Hualiang Jiang,Luke W. Guddat,Haitao Yang,Zihe Rao
Licensed Content Date	Jan 24, 2019
Licensed Content Volume	176
Licensed Content Issue	3
Licensed Content Pages	26

■ Order Details

Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No

■ About Your Work

Title of new work	Investigation of the oligomeric state of the mycobacterial lipid transporter, Mmpl3: biophysical and computational insights.
Institution name	Imperial College London
Expected presentation date	May 2024

■ Additional Data

Portions	Figure 1
----------	----------

📍 Requestor Location

Imperial College London
Molecular Sciences Research Hub

Requestor Location

London, W12 0BZ
United Kingdom
Attn: Imperial College London

■ Tax Details

Publisher Tax ID	GB 494 6272 12
------------------	----------------

7.3 Approval 3

Thank you for your order.

This Agreement between Imperial College London -- Sara Cioccolo ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number5787540361094

License dateMay 14, 2024

Licensed Content

Licensed Content PublisherSpringer Nature

Licensed Content PublicationNature Reviews Molecular Cell Biology

Licensed Content TitleRegulation of membrane protein structure and function by their lipid nano-environment

Licensed Content AuthorIlya Levental et al

Licensed Content DateSep 2, 2022

Order Details

Type of UseThesis/Dissertation

Requestor typeacademic/university or research institute

Formatprint and electronic

Portionfigures/tables/illustrations

Number of figures/tables/illustrations2

Would you like a high resolution image with your order?no

Will you be translating?no

Circulation/distribution50000 or greater

Author of this Springer Nature contentno

About Your Work

Title of new workInvestigation of the oligomeric state of the mycobacterial lipid transporter, MmpL3: biophysical and computational insights.

Institution nameImperial College London

Expected presentation dateMay 2024

Additional Data

PortionsFigure 1, Figure 4

Requestor Location

Imperial College London

Molecular Sciences Research Hub

Requestor Location

London, W12 0BZ

United Kingdom

Attn: Imperial College London

Tax Details

Printable Details

183

7.4 Approval 4

Order Completed

Thank you for your order.

This Agreement between Imperial College London -- Sara Cioccolo ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number	5787541073347		
License date	May 14, 2024		
Licensed Content		Order Details	
Licensed Content Publisher	Elsevier	Type of Use	reuse in a thesis/dissertation
Licensed Content Publication	Advanced Drug Delivery Reviews	Portion	figures/tables/illustrations
Licensed Content Title	Structural modelling and dynamics of proteins for insights into drug interactions	Number of figures/tables/illustrations	1
Licensed Content Author	Tim Werner,Michael B. Morris,Siavoush Dastmalchi,W. Bret Church	Format	both print and electronic
Licensed Content Date	Mar 30, 2012	Are you the author of this Elsevier article?	No
Licensed Content Volume	64	Will you be translating?	No
Licensed Content Issue	4		
Licensed Content Pages	21		
About Your Work		Additional Data	
Title of new work	Investigation of the oligomeric state of the mycobacterial lipid transporter, Mmpl3: biophysical and computational insights.	Portions	Figure 3
Institution name	Imperial College London		
Expected presentation date	May 2024		
Requestor Location		Tax Details	
	Imperial College London Molecular Sciences Research Hub	Publisher Tax ID	GB 494 6272 12
Requestor Location	London, W12 0BZ United Kingdom Attn: Imperial College London		