Molecular Assembly and Interactions of Membrane Proteins and Viral Glycoproteins using Mass Photometry



Anna Olerinyova
St John's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy in Theoretical and Physical Chemistry
20th May, 2022

Acknowledgements

Over the last couple of years, I have been lucky to be given the amazing support and guidance from many incredible people, whether through academic help and feedback, or kind reassurance. I have also been fortunate to receive funding and guidance from the EPSRC, DTC, and St. John's College, for which I am very grateful.

First and foremost, I would like to express my gratitude to my primary supervisor, Philipp Kukura, for his guidance, help, and constant support he kindly provided throughout my time in his fantastic group. I would also like to thank Carol Robinson, for her overarching support and guidance. Both the Kukura and Robinson groups have been amazing to work with, full of incredible people, many of whom I can call friends and colleagues. During my first rotation in the Kukura group, Nick kindly introduced me not only to the inner workings of the amazing technique being developed in this lab, but also to the excitement of observing single proteins landing on a simple piece of glass in real life. I would like to thank Adar being instrumental in guiding the early stages of my research. I would also like to thank Manish for his invaluable advice on many scientific and life problems, enabling me to grow both as a person and a scientist. I also need to thank Adam and Gavin for their guidance through the more technical side of experiments and patient explanations of the physical details of mass photometry. I would also like to thank Fabian, Weston, Roi, and the rest of the TeamBio for the endless discussions and help with the COVID project, without which we would not have been able to achieve these exciting results. These thanks also extend to the rest of the Kukura group for their amazing support, friendship, and many fun stories shared over pints after an exhausting work day, especially during those strange pandemic times. It has also been a great pleasure working with my numerous collaborators, and I am eternally grateful for their guidance and the opportunity to work on novel systems and proteins.

My time in Oxford was made amazing in large part due to a group of fantastic friends I have made. I would especially like to thank Alix and Natalia who have been on the same thesis writing and submission timeline, providing constant moral support and a space to vent. I am also incredibly grateful for the help and support from Jess and Cong, whose proofreading abilities have helped me avoid many grammar errors and awkward sentence structures, as well as many other friends

who were always there for me. Last but not least, I would like to express my eternal gratitude to my parents, brother, grandparents, as well as the rest of my family for their never-ending support, encouragement and for always believing in me. I am truly thankful for all you have done for me.

Originality Statement

The research described in this thesis was carried out by the author between April 2017 and December 2021 and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text. The use of the first person plural is strictly a matter of style in keeping with standard scientific convention.

Publications List

The work carried out by the author between April 2017 and December 2021 led to the following publications:

1. Quantitative Mass Imaging of Single Biological Macromolecules. *Science* 2018. Young, G., Hundt, N., Cole, D., Fineberg, A., Andrecka, J., Tyler, A., **Olerinyova, A.**, Ansari, A., Marklund, E. G., Collier, M. P., Chandler, S. A., Tkachenko, O., Allen, J., Crispin, M., Billington, N., Takagi, Y., Sellers, J. R., Eichmann, C., Selenko, P., Frey, L., Riek, R., Galpin, M. R., Struwe, W. B., Benesch, J.L.P. and Kukura, P.

2. Mass Photometry of Membrane Proteins. *Chem* 2021. **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.

3. Lipids Mediate Promiscuous Supramolecular Assemblies in the Bacterial Outer Membrane. M. N. Webby, A. O. Oluwole, P. G. Inns, N. G. Housden, **A. Olerinyova**, C. Pedebos, N. T. Reichmann, S. Khalid, P. Kukura, S. Mohammed, C. V. Robinson, C. Kleanthous. (*submitted*)

4. The Spike-ACE2 interaction underlying SARS-CoV-2 infection and inhibition is enhanced by intermolecular crosslinking. **A. Olerinyova**, R. Asor, M. S. Kushwah, F. Soltermann, L. S. P. Rudden, S. Vasiljevic, J. Brun, M. Hensen, M. Hill, M. T. Degiacomi, N. Zitzmann, W. B. Struwe, P. Kukura. (*submitted*)

Abstract

Integral membrane proteins (IMPs) and their interactions form the basis underlying many processes essential for life of all living organisms. Despite their pivotal importance, they are challenging to study due to the presence of a bilayer-embedded transmembrane domain requiring astute strategies to achieve solubilisation in aqueous environments. Here we present mass photometry as a novel technique to investigate IMPs and their interactions in a variety of membrane mimetic systems, uncovering the interactions underlying the bacterial outer membrane architecture. We then apply mass photometry to viral membrane glycoproteins, the HIV-1 envelope spike and SARS-CoV-2 spike, and their interactions with a soluble inhibiting protein and host cell receptor, respectively. In both cases, we describe how the oligomeric nature of the involved proteins facilitates intermolecular crosslinking, facilitating enhanced inhibition.

We begin by establishing the applicability of mass photometry to measure a variety of membrane mimetic systems, including detergents, by determining the molecular weight of different IMPs in detergents, amphipols, lipid nanodiscs, and native nanodiscs. We further show the heterogeneity stemming from native membrane protein extraction with native nanodiscs, as well as the utility of mass photometry in predicting IMP functionality in different lipid nanodisc assemblies. We apply these findings to characterise the architecture and interactions between the IMP of the Gram-negative bacterial outer membrane, providing important insights into the outer membrane island formation and contributions to antibiotic resistance.

Next, we focus on the interactions of the HIV-1 envelope spike glycoprotein with its natural inhibitor, banana lectin BanLec. Through careful mass photometry measurements, we determine the dynamics of this multivalent interaction, resulting in the formation of highly heterogeneous species through intermolecular crosslinking. This approach is then applied to investigate the mechanism behind SARS-CoV-2 infection. We measure the interactions between the viral spike and its human host cell receptor, ACE2, showing the importance of its dimeric nature for achieving its full neutralisation potential through spike crosslinking.

This work introduced mass photometry as an important tool for membrane protein research with broad applications in protein-protein interactions as well as in characterisation of complex heterogeneous reactions important in viral infections.

Contents

List of Figures **iv**

List of Abbreviations **vii**

1 Introduction **1**

2 Theoretical Background **5**
2.1 Integral Membrane Proteins 5
2.1.1 Lipid Bilayer Properties 8
2.2 Membrane Mimetic Systems 13
2.2.1 Detergents 14
2.2.2 Amphipols 21
2.2.3 Lipid Nanodiscs 28
2.2.4 Native Nanodiscs 38
2.3 Viral Glycoproteins 45
2.3.1 HIV Env Glycoprotein 45
2.3.2 SARS-COV-2 Spike Glycoprotein 47
2.4 Mass Photometry 56

3 Mass Photometry of Membrane Proteins **65**
3.1 Introduction 66
3.2 Materials and Methods 69
3.2.1 Mass Photometry Measurements 69
3.2.2 Mass Photometry Image Processing 71
3.2.3 Mass Photometry Calibration Procedure 72
3.2.4 Detergents 72
3.2.5 Amphipols 75
3.2.6 Lipid Nanodiscs 78
3.2.7 Native Nanodiscs 80
3.3 Results 82
3.3.1 Empty Protein Carriers 82
3.3.2 Detergents 85

3.3.3 Amphipols . . . 88
3.3.4 Lipid Nanodiscs . . . 90
3.3.5 Native Nanodiscs . . . 92
3.4 Discussion and Conclusions . . . 96

4 Bacterial Outer Membrane Protein Interactions 99
4.1 Introduction . . . 100
4.2 Materials and Methods . . . 103
4.2.1 BPA Incorporation and Extraction of the Outer Membrane Fractions . . . 104
4.2.2 OmpF and BtuB Purification . . . 104
4.2.3 SDS and Blue Native PAGE . . . 107
4.2.4 Protein Crosslinking with Glutaraldehyde . . . 108
4.2.5 Mass Photometry Measurements . . . 108
4.3 Results . . . 110
4.3.1 OmpF and BtuB Analysis . . . 111
4.3.2 OmpF V177^{BPA} Interacts with OM lipids, TBDTs, and BtuB 113
4.3.3 OmpF G198^{BPA} and BtuB are Self-Crosslinked by Glutaraldehyde . . . 119
4.4 Discussion . . . 123

5 HIV Envelope Protein - BanLec Interaction 127
5.1 Introduction . . . 128
5.2 Materials and Methods . . . 131
5.2.1 Interferometric Scattering Microscopy . . . 132
5.2.2 Sample Preparation and Landing Assay . . . 133
5.2.3 Data Analysis . . . 134
5.3 Results . . . 138
5.3.1 Env Interaction with BanLec . . . 138
5.3.2 BanLec-Induced Env Oligomerisation . . . 141
5.3.3 Quantifying the Env–BL Interaction . . . 144
5.4 Discussion . . . 147

6 Spike-ACE2 Interaction in SARS-CoV-2 151
6.1 Introduction . . . 152
6.2 Materials and Methods . . . 157
6.2.1 Protein Expression and Purification . . . 157
6.2.2 Mass Photometry . . . 160
6.2.3 Molecular Docking . . . 167
6.2.4 SARS-CoV-2 Neutralisation Assay . . . 169

6.3 Results 171
6.3.1 RBD Binds ACE2 Monomers and Dimers Equally 171
6.3.2 ACE2 Dimers Crosslink Spike into Large Oligomers 173
6.3.3 Modelling Possible ACE2–Spike Interactions 179
6.3.4 SARS-CoV-2 is Strongly Inhibited by ACE2 but not mACE2 183
6.4 Discussion and Conclusions 185

7 Summary and Outlook **188**
7.1 Summary 188
7.2 Outlook 191

References **194**

List of Figures

2.1 Phospholipid and cell membrane structure. 9
2.2 The main lipids of the Gram-negative bacterial outer membrane. . . 11
2.3 Different membrane mimetic systems ordered by the level of nativeness. 14
2.4 Detergent interactions and micelle shapes. 15
2.5 Detergent solubilisation of the membrane and membrane proteins. . 19
2.6 Molecular subunit structure of the A8-35 amphipol. 22
2.7 IMP solubilisation in amphipols. 24
2.8 Lipid nanodisc, MSP, and circularised nanodisc schematics. 30
2.9 The process of IMP insertion into lipid nanodiscs. 33
2.10 Membrane protein extraction using SMA. 38
2.11 The structure and properties of SMA and DIBMA copolymers. . . . 41
2.12 Simplified structure of the SARS-CoV-2 virion. 48
2.13 Effects of SARS-CoV-2 infection on transcription and translation. . 51
2.14 Overview of the renin-angiotensin system and its possible associations with SARS-CoV-2 infection and COVID-19 associated lung injury. . 54
2.15 The basis of signal detection in mass photometry. 57
2.16 Mass photometry (Interferometric Scattering Mass Spectrometry (iSCAMS)) precision and accuracy characterisation. 62
2.17 Calculated distributions of the specific volumes and refractive indices of different proteins. 63

3. Mass Photometry of Membrane Proteins **65**
3.1 Contrast to mass calibration. 73
3.2 Analysis of OmpF detergent-APol exchange. 77
3.3 Analysis of KcsA detergent-APol exchange. 77
3.4 MP of detergent micelles. 83
3.5 Mass photometry of detergent micelles and nanodisc membrane protein carriers. 84
3.6 MP Analysis of respiratory complex I in LMNG. 86
3.7 Analysis of bo_3 oxidase in LMNG using MP, MD and SEC-MALS. . 88

3.8 OmpF trimers and KcsA tetramers in APols at different protein:APol ratios. 90
3.9 KcsA lipid nanodisc characterisation data. 92
3.10 KcsA LN assembly screen. 93
3.11 Mass photometry of KcsA in native nanodiscs. 94
3.12 Lipid/protein analysis of *E. coli* KcsA native NDs. 95
3.13 Theoretical lipid/protein/SMA stoichiometry of KcsA native NDs. . 96

4. Bacterial Outer Membrane Protein Interactions 99
4.1 Cell wall structure of Gram positive and negative bacteria. 101
4.2 OmpF BPA incorporation sites selected for MP measurements. . . . 105
4.3 The structure and mechanism of BPA crosslinking. 105
4.4 The structure and mechanism of glutaraldehyde crosslinking. 109
4.5 Ion exchange and MP analysis of OmpF and BtuB. 112
4.6 SDS and bnPAGE analysis of OmpF V177^{BPA} interacting with TBDTs, OM lipids, and BtuB. 114
4.7 SEC and SDS-PAGE of OmpF V177^{BPA} mixed with BtuB. 116
4.8 Mass photometry of SEC fractions of OmpF V177^{BPA} mixed with BtuB. 117
4.9 Mass photometry of selected OmpF V177^{BPA} + BtuB fractions and controls. 118
4.10 SEC and SDS-PAGE analysis of glutaraldehyde-induced OmpF-BtuB crosslinking. 120
4.11 MP analysis of OmpF crosslinking to BtuB. 120
4.12 SEC and SDS-PAGE analysis of BtuB and OmpF G198^{BPA} crosslinking controls. 121
4.13 MP of BtuB and OmpF G198^{BPA} crosslinking controls. 122
4.14 OmpF, BtuB and glutaraldehyde crosslinking data comparison. . . 123

5. HIV Envelope Protein - BanLec Interaction 127
5.1 Env gp160 complex compared to the Env gp140 construct. 132
5.2 HSP16.5 and Env data used for mass calibration. 135
5.3 Mass distributions of Env titrated with 0-50 nM BL. 140
5.4 Env–BL oligomer formation increases with BL concentration. 142
5.5 Averaged histogram of 20 repeats of 10 nM Env with 20 nM BL. . . 145
5.6 Env interactions with 0.1-40 nM BL. 147
5.7 Proposed mechanism of BL crosslinking Env in solution and *in vivo*. 149

6. Spike-ACE2 Interaction in SARS-CoV-2 151

6.1 Possible interactions between membrane-anchored spike and ACE2. 154
6.2 Characterisation of spike, ACE2, and mACE2. 159
6.3 ACE2 distribution evolution over 120 min. 163
6.4 Analysis of the mACE2–RBD interaction. 172
6.5 Mass distributions and binding energy analysis of the ACE2–RBD interaction. 174
6.6 Mass photometry analysis of spike–ACE2 interaction. 176
6.7 Particle distribution and count analysis of the spike–ACE2 interaction. 178
6.8 MD analysis of possible spike interactions with a single ACE2. . . . 181
6.9 Analysis of the maximal occupancy of spike and its dimer. 182
6.10 SARS-CoV-2 infection is inhibited more strongly by ACE2 than mACE2. 184

List of Abbreviations

dn/dc	Refractive Index Increment Value
K_D	Dissociation Constant
k_{off}	Dissociation Rate Constant
k_{on}	Association Rate Constant
AA	Amino Acid
ACE2	Angiotensin Converting Enzyme 2
AFM	Atomic Force Microscopy
AIDS	Acquired Immune Deficiency Syndrome
AngI/II	Angiotensin I/II
AOD	Acousto-Optic Deflector
apo-AI	Apolipoprotein A-I
AT1R	Angiotensin II Receptor Type 1
AUC	Analytical Ultracentrifugation
BPA	Para-benzyolphenylalanine
BtuB	Vitamin B12 Transporter
CMC	Critical Micelle Concentration
CR3022	Anti SARS-CoV-2 Spike Antibody
cryoEM	Cryo-Electron Microscopy
CV	Column Volume
DDM	n-Dodecyl-β-D-Maltoside
DLS	Dynamic Light Scattering
DMPC	1,2-Dimyristoyl-sn-Glycero-3-Phosphocholine
DPC	n-Dodecylphosphocholine
DTT	1,4-Dithiothreitol
ESI	Electrospray Ionisation
FepA	Ferrienterobactin Receptor
FhuA	Ferrichrome Outer Membrane Transporter/Phage Receptor
FWHM	Full-Width at Half-Maximum
GPCR	G-Protein Coupled Receptor
HIV	Human Immunodeficiency Virus
HPLC	High-Performance Liquid Chromatography
IC_{50}	Half-Maximal Inhibitory Concentration

IFN	Interferon
IMP	Integral Membrane Protein
IPTG	Isopropyl β-D-1-thiogalactopyranoside
iSCAT	Interferometric Scattering Microscopy
KDE	Kernel Density Estimate
LMNG	Lauryl Maltose Neopentyl Glycol
LN	Lipid Nanodisc
LPS	Lipopolysaccharide
mAb	Monoclonal Antibody
MALS	Multi-Angle Light Scattering
MAP	Membrane-Associated Protein
MD	Molecular Dynamics
MP	Mass Photometry
MQ	Milli-Q H_2O
MS	Mass Spectrometry
MSP	Membrane Scaffold Protein
MW	Molecular Weight
MWCO	Molecular Weight Cut-Off
NA	Numerical Aperture
NMR	Nuclear Magnetic Resonance
NN	Native Nanodisc
NSP	Non-Structural Protein
OD	Optical Density
OG	n-Octyl-β-D-Glucoside
OM	Outer Membrane
OMP	Outer Membrane Protein
OmpF	Outer Membrane Porin F
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate-Buffered Saline
PDB	Protein Data Bank
PL	Phospholipid
POPC	1-Palmitoyl-2-Oleoyl-sn-Glycero-3-Phosphocholine
PSF	Point Spread Function
PTM	Post-Translational Modification
RBD	Receptor Binding Domain
RI	Refractive Index
rpm	Rotations per Minute
RT	Room Temperature
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus/2

SDS	Sodium Dodecyl Sulphate
SEC	Size-Exclusion Chromatography
SMA	Styrene-Maleic Acid
SMALP	Styrene Maleic Acid Co-polymer Lipid Particle
smFRET	Single Molecule Fluorescence/(Förster) Resonance Energy Transfer
SPR	Surface Plasmon Resonance
TBDT	TonB Dependent Transporters
TM	Transmembrane
w/v	weight/volume
WT	Wild-Type

1 Introduction

Biological membranes represent the boundary that confine all living organisms, separating the biological and chemical processes that are essential for life. Far from simple, these structures contain many associated and embedded proteins that carry out physiologically important functions. This is reflected in their genomic frequency, with an estimated 26% of human genes containing confirmed or predicted transmembrane (TM) domains, indicative of integral membrane proteins (IMPs).[1] Located at the border between the intra- and extracellular environments, IMPs serve a number of indispensable functions. For example, they facilitate transport of nutrients and waste across the membrane, ensure intercellular communication and signalling, as well as catalyse a variety of chemical and enzymatic reactions.

The functional importance of IMPs is undeniable. The majority of small molecule drugs target IMPs, with many more being developed.[2] At the same time, the presence of specific IMPs and lipids in the outer membrane of Gram-negative bacteria contributes to the rise in antibiotic resistance.[3] Furthermore, yet another type of IMPs facilitate the first step in most viral infections, forming bonds with host receptors, which can lead to cell attachment and fusion. As such, IMPs are

critical not only for cell physiology in health and disease, but also play a major role in global health threats posed by the different infections and diseases, spreading antibiotic resistance, and viral pandemics, one of which we are currently witnessing.

Continuing research into the structure, function, and interactions of IMPs is thus immensely important. The function of most IMPs depends on their large interaction networks, whether with soluble proteins or with other IMPs, lipids, co-factors, or small molecules. However, progress in IMP research has been slow compared to soluble proteins, due to challenges related to their structure. IMPs contain a TM domain embedded in the lipid bilayer where it is surrounded by the hydrophobic tails of phospholipids. This section renders most IMPs highly unstable in an aqueous environment, generally required for most structural and biophysical techniques. A large variety of strategies has been developed to aid in their solubilisation, with more native-like options becoming available recently. Nevertheless, whilst these can enable structural or functional characterisation, studying IMP–protein interactions, especially between different IMPs, remains exceedingly difficult.

Recent development of a label-free, single-molecule microscopic imagining technique, mass photometry, provided a novel approach for studying biomolecules and their interactions.[4,5] This method measures the molecular weight of individual biomolecules based on their scattering contrast, providing information regarding sample purity,[6] protein oligomeric assembly or interactions,[7] with applications extending to DNA and other biomolecules.[8] In this work, we extend the applicability of mass photometry to IMPs and membrane mimetic systems, which we then use to characterise the interactions between IMPs and lipids in the outer membrane of Gram-negative bacteria. We then focus on viral membrane glycoproteins, defining their interactions with soluble proteins and IMPs alike, and describe a novel mechanism for these interactions, with applications in viral infection prevention

and treatment.

Chapter 2 starts by providing an overview of the membrane environment and discusses the development, properties, and uses of different membrane mimetic systems in IMP research. We then focus on viral glycoproteins, specifically the HIV-1 envelope spike glycoprotein and the SARS-CoV-2 spike glycoprotein, outlining their structure, function, and host cell interactions. In addition, the theoretical basis of mass photometry is discussed, introducing the reader to this novel technique.

Chapter 3 first demonstrates the ability of mass photometry to detect different membrane mimetic systems, followed by characterising a variety of IMPs in selected membrane mimetics. We describe a simple dilution method for IMP measurements in detergents, allowing us to define an IMP monomer-dimer equilibrium. We also show the utility of mass photometry in predicting the outcome of lipid nanodisc assembly on IMP function, and characterise the heterogeneity of an IMP purified in native nanodiscs.

In Chapter 4, we apply these findings to study the macromolecular assemblies in the outer membrane of Gram-negative bacteria using mass photometry. Through the development of a photocrosslinking strategy, we describe the role of phospholipids and lipopolysaccharide in facilitating the interactions between the *E. coli* trimeric porin OmpF and the vitamin B transporter BtuB. We find that these lipids can act as a type of molecular 'glue', forming the basic unit of outer membrane protein islands.

In Chapter 5, we investigate the basis of HIV-1 inhibition by the tetrameric banana lectin BanLec through studying its interaction with the HIV-1 envelope spike glycoprotein, Env. Using a near-native solubilised Env construct gp140, we

define the reaction strength and describe BanLec-induced Env oligomerisation, contributing to the high efficiency of HIV-1 infection inhibition *via* Env clustering. Through this work, we illustrate the ability of mass photometry to characterise highly heterogeneous reactions involving intermolecular crosslinking, as well as the accurate detection of glycoproteins.

We further apply this knowledge in Chapter 6 to investigate the molecular details of the SARS-CoV-2 spike glycoprotein interaction with its host receptor, angiotensin-converting enzyme 2 (ACE2). Using mass photometry, we show that the solubilised trimeric spike and human dimeric ACE2 form large oligomers *in vitro*. We also find that ACE2 dimers inhibits SARS-CoV-2 infections two-orders of magnitude more potently than ACE2 monomers, which we propose occurs due to ACE2 dimer-mediated crosslinking of spike trimers, with important implications for viral infectivity and novel treatment options.

Finally, Chapter 7 provides a short summary of the contents and findings described in this Thesis, as well as a brief outlook into potential extensions of this work and future directions of IMP and viral glycoprotein research using mass photometry.

2
Theoretical Background

Contents

2.1 Integral Membrane Proteins **5**
2.1.1 Lipid Bilayer Properties 8
2.2 Membrane Mimetic Systems **13**
2.2.1 Detergents 14
2.2.2 Amphipols 21
2.2.3 Lipid Nanodiscs 28
2.2.4 Native Nanodiscs 38
2.3 Viral Glycoproteins **45**
2.3.1 HIV Env Glycoprotein 45
2.3.2 SARS-COV-2 Spike Glycoprotein 47
2.4 Mass Photometry **56**

2.1 Integral Membrane Proteins

Membrane proteins encompass a group of proteins that are either integrated within the membrane (IMPs), or attached to the membrane, whether permanently as integral monotopic membrane proteins, through an association with one of the membrane surfaces, or temporarily as membrane associated proteins (MAPs), through a transient interaction.[9] They are essential for every living organism, including viruses, facilitating communication and exchange of materials with

the extracellular environment, providing means of interaction for the cell whilst maintaining the protection afforded by the cellular membrane. They serve a wide variety of indispensable functions, including cell signalling transduction, ion and small molecule transport, reaction catalysis, or interactions with other cells, surfaces, or various substrates.[10]

Perhaps due to their necessity and functional variety, it has been approximated that genes encoding membrane proteins comprise ∼20-30% of most genomes, and up to 1/3 of the human genome,[11] which translates to ∼23 % of the human proteome.[12] Despite their comparatively low abundance, estimates suggest that membrane proteins constitute over 60% of all human drug targets, with G protein-coupled receptors (GPCRs) accounting for almost half.[13] Given their importance in the transduction of cell signalling and the large impact on downstream intracellular signalling cascades upon their activation, combined with pharmacological ease of access compared to intracellular and nuclear proteins, their prevalence as drug targets is unsurprising.

Whilst the medical and biological importance of membrane proteins is undeniable, studying the biochemical, structural, and functional properties of membrane proteins, especially IMPs, has proven challenging. IMPs contain a hydrophobic transmembrane (TM) region facilitating their membrane insertion and lipid interactions, both of which are essential for maintaining the correct structure and function of the protein. However, many experimental techniques used for structural or biophysical characterisation require water solubility, and are as such incompatible with native membrane-inserted IMPs. As a result, a variety of MMS has been developed, with the aim to enable membrane extraction, solubilisation, and purification of IMPs in water-based buffers. These include the relatively non-native detergents, which are the first and to this day the most commonly used MMS; a variety of amphipathic

polymers with different functions and properties; novel solubilisation methods aimed at recreating or capturing the native phospholipid (PL) bilayer environment, such as lipid or native nanodiscs; and small artificial or native lipid-composed MMS, such as liposomes or vesicles, which can be said to best resemble the native membrane environment of IMPs.[14] Section 2.2 discusses various MMS in more detail, specifically focusing on detergents, amphipols, lipid and native nanodiscs.

Despite the variety of MMS available, there are still significant obstacles associated with IMP research, including problems with IMP overexpression, membrane extraction, and successful solubilisation maintaining a functional, near-native state of the protein. As a result, many IMPs remain uncharacterised, constituting less than 2% of the reported unique protein structures in the Protein Data Bank (PDB).[15,16] Their immense pharmacological and biological significance make IMPs the subject of extensive research efforts, which inevitably face the experimental challenges related to the identification of an MMS suitable for a specific IMP and selected downstream applications, IMP extraction and purification complexity, as well as difficult questions about the nativeness of the selected MMS and its effects on the IMP structure and function.

Whilst staggering progress has been made over the last two decades in the field of IMP research, it often created more questions and sometimes even challenged previously accepted dogmas, with the emerging complexity of membrane-protein mixtures sometimes daunting. Better understanding of the importance and influence of the native membrane environment lead to a growing shift in the MMS used for IMP research, away from more disruptive and less native ones, like detergents, and towards more membrane-mimicking MMS, like lipid or native nanodiscs. Among others, it highlighted the key role of annular lipids for membrane protein structure and function, as well as the importance of the physical aspects of the IMP being

embedded in a lipid bilayer, such as lateral pressure, membrane fluidity, and nanodomain formation. In this Section, the key aspects of lipid bilayer properties and protein-lipid interactions, some of the membrane mimetic systems, well as some of the major techniques used for IMP characterisation will be briefly outlined and discussed, providing background for the experimental work presented in this Thesis.

2.1.1 Lipid Bilayer Properties

In many of the currently known living organisms, a thin polar membrane encircles and separates individual living cells from their often harsh extracellular environment.[17] Apart from containment, it also facilitates the exchange of critical ions, nutrients, metabolites, and other small molecule through its selective permeability obtained *via* a wide variety of numerous transporters, porins, channels, and pumps.[18] The lipid bilayer itself is highly impermeable to hydrophilic molecules, a property crucial for the maintenance of correct intracellular pH, ion concentration, and osmotic pressure.[19] Extensive biophysical and biochemical research has created a detailed model of the structure and composition of biological membranes, discovering highly dynamic levels of organisation, whether across the membrane in the form of lipid asymmetry, or laterally, with a variety of lipid nanodomains and lipid-protein organisation units described.

Most biological membranes contain a double leaflet of amphipathic PLs interspersed with other types of lipids, such as different sterols (*e.g.* cholesterol in vertebrates, or ergosterol in fungi), or a variety of glycolipids.[20] As for the amphipathic PLs, their hydrophilic, charged head groups are positioned on the outer surface, and the hydrophobic, apolar fatty acid chain tails face inwards, forming an oil-like environment in between the polar outer-facing surfaces on either side (Figure 2.1, right). The majority of PLs found in humans or in some model organisms, such as the Gram-negative bacterium *Escherichia coli*, contain two fatty

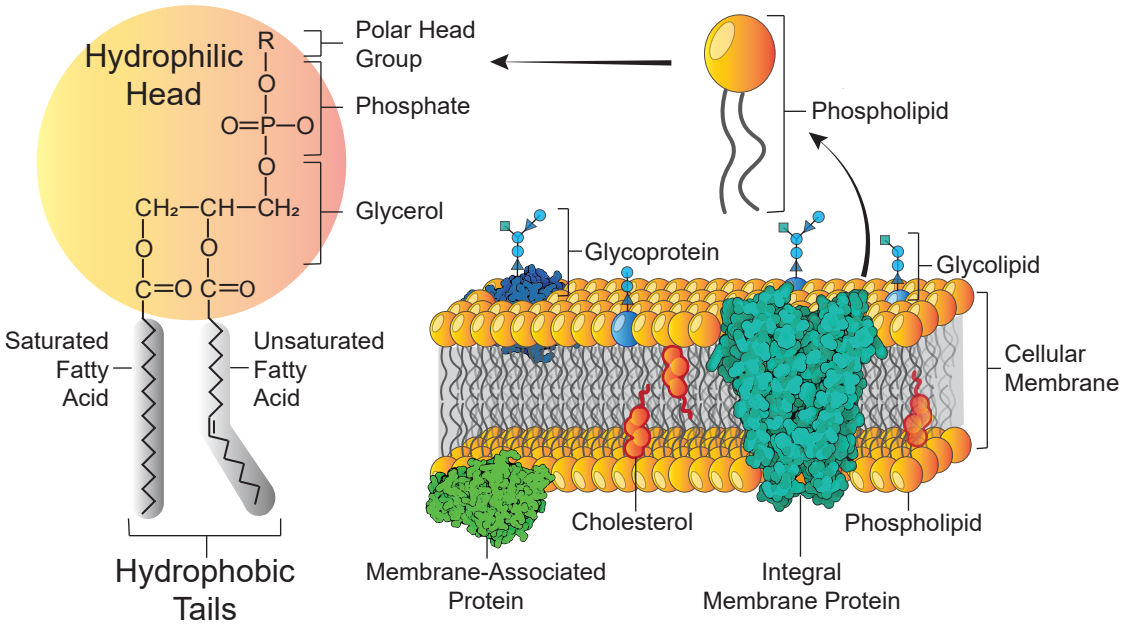


Figure 2.1: Phospholipid and cell membrane structure. Left, molecular structure of a typical PL, containing a hydrophobic head with a polar head group (R), a phosphate group, and glycerol, and hydrophobic tails composed of a saturated and unsaturated fatty acid. **Right**, simplified structure of a cell membrane, consisting of a PL bilayer with the addition of cholesterol and glycolipids, integral membrane proteins, membrane-associated proteins, and glycoproteins.

acid tails, which vary in length and the degree of saturation, and are connected *via* an alcohol moiety, usually glycerol, to a phosphate group and a polar head group (Figure 2.1, left).[21] PLs can spontaneously self-assemble into bilayers in aqueous solutions, thanks to the tendency of the hydrophobic tails to self-associate through hydrophobic interactions in order to exclude water (called the hydrophobic effect), whilst the hydrophilic heads interact not only with each other, but also with the aqueous environment. As such, a PL suspension in water often results in the spontaneous formation of micelles, liposomes, or a bilayer.[22]

In general, the PLs' polar head groups contribute to the chemical and physical properties of the membrane surface, such as its charge, whilst the fatty acids have a larger impact on the physical properties of the bilayer itself, such as its transition temperature, elasticity, and fluidity.[23] Specific head groups can strongly affect

the surface charge and function of the specific PL bilayer area, including lipids covalently glycosidically bonded to carbohydrate chains and sphingolipids serving a multitude of functions in cell-to-cell signalling and recognition, surface attachment, cell protection, and molecular anchoring, amongst others.[24]

On the other hand, the length and saturation of the hydrophobic tails impact the exact temperature when phase transition occurs. At lower temperatures, bilayers can be often found in a solid gel phase, which transitions at higher temperatures to a fluid phase. Generally speaking, longer fatty acid chains increase this temperature through increased Van der Waals' interactions, and as such decrease membrane fluidity, whilst a higher degree of unsaturation has the opposite effect due to increased steric hindrance preventing fatty acid packing, reducing Van der Waals interactions. The presence of sterols, such as cholesterol in animal membranes, increases lipid packing, providing mechanical strength and decreasing its permeability, as well as aiding in the function of certain IMPs.

The combination of a specific head group, fatty acid tails, and the molecular type of their linkage results in a staggering level of diversity of lipids present in biological membranes, with several thousand identified through lipidomics cataloguing.[26] The exact lipid composition appears to vary across species and organisms, different cell types and their organelles, the age of the cells and organism, but also across the bilayer leaflets, resulting in lipid asymmetry, and laterally, forming numerous lipid rafts and membrane domains of specific lipid and protein composition.[21] The variation in the chemical structure of the lipids ranges from relatively subtle differences, such as the position of a double-bond in a fatty acid tail or various stereoisomers, to more profound ones, including a variety of head groups, different levels of glycosylation, or the type of the lipid backbone. Biological membranes also vary compositionally, with the specific ratio of lipids affecting the overall

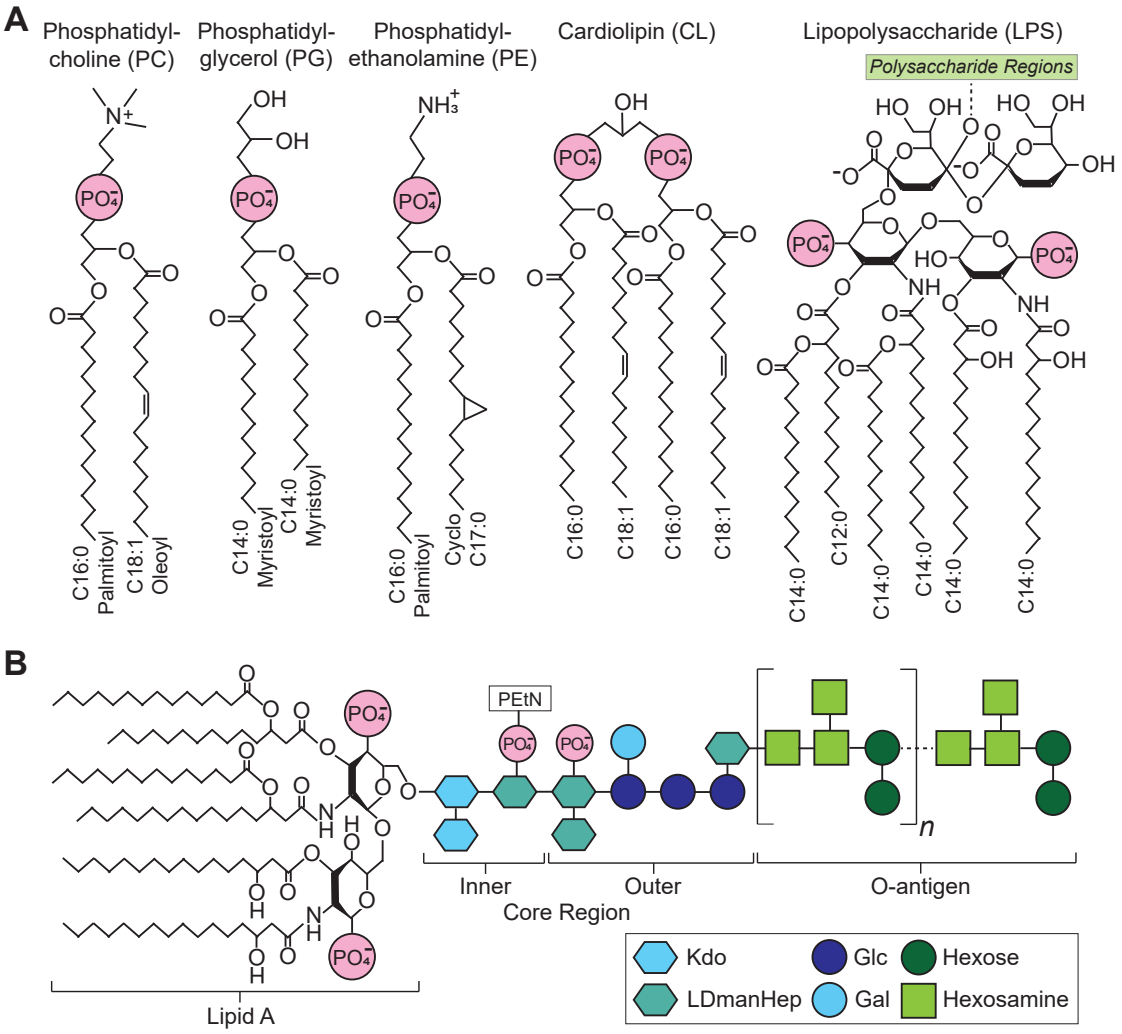


Figure 2.2: The main lipids of the Gram-negative bacterial outer membrane. **A.** Example chemical structures of phosphatidylcholine (PC), phosphatidylglycerol (PG), phosphatidylethanolamine (PE), cardiolipin (CL), and lipopolysaccharide (LPS), lipid A section. The outer membrane of *E. coli* generally contains PE, PG, and CL in the inner leaflet, and LPS in the outer leaflet. Even though PC is not commonly found in bacterial membranes, it is often used for *in vitro* outer membrane protein folding studies as it forms bilayers easily and is zwitterionic, *i.e.* carries no net charge. **B.** Schematic of an exemplary LPS structure, including the types of sugar moieties present. There is considerable variation between species in their specific LPS structures, varying in the number and length of the acyl chains in the lipid A region, as well as the types of sugars present and their ratio in the polysaccharide section. Some laboratory strains of *E. coli* do not contain any O-antigen, in which case the LPS is referred to as 'rough'. The O-antigen also varies considerably between species, with up to 40 glycan repeats present. The legend at the bottom right of the figure contains the types of sugars present in this example structure. Kdo - keto-deoxyoctulosonate; LDmanHep - L-glycero-D-manno-heptose; Glc - glucose; Gal - galactose; PEtN - phosphorylethanolamine. Adapted with permission from *Journal Biological Chemistry*, Role of the lipid bilayer in outer membrane protein folding in Gram-negative bacteria, J. E. Horne *et al.*, ©2020, Creative Commons CC-BY License 4.0.[25]

membrane behaviour and properties.[26] Even though the composition of a specific membrane changes dynamically, *e.g.* with age, metabolic state or cell division,[27] it is also controlled through homeostasis, with misregulation resulting in a variety of pathologies.[28] The dominant lipid types present in the *E. coli* outer membrane are outlined in Figure 2.2 to provide an example of the structural variety present in biological membranes. It also serves as a reference point for the work presented in Chapters 3 and 4 that uses some of these lipids during membrane protein reconstitution or analyses their interactions with surrounding membrane proteins.

Whilst the exact role of this high level of diversity remains largely elusive, it is well-documented that specific lipids serve different functions. Besides creating a barrier between the intracellular and extracellular environments, the presence of discrete organelles in eukaryotes aids in further compartmentalisation, leading to segregation of biochemical reactions enabling higher efficiency. Lipids also create the potential for membrane budding, fusion, fission, and tubulation, necessary for cell division, molecular trafficking and intercellular signalling, amongst others.[26] Additionally, about tens of different types of lipids function in intracellular signalling. The degradation of these amphipathic molecules allows for signal transmission within the membrane, through their hydrophobic section, as well as in the cytosol, through the liberated polar head group.

Whilst most of the lipids in a biological membrane are relatively mobile in the fluid phase, a shell of lipids sometimes surrounds IMPs and contributes to their correct structure, function, and lateral mobility and diffusion within the bilayer. Unlike the bulk membrane, these annular lipids can be partially immobilised by their respective protein-lipid interactions. The specific types and ratios of lipids in the bulk membrane and surrounding different IMPs can affect protein organisation,

either contributing to tight IMP packing and aggregation, or allowing for protein diffusion and dispersion in the bilayer.[26]

2.2 Membrane Mimetic Systems

MMS generally interact with the hydrophobic TM section of IMPs, shielding it from the surrounding aqueous solution, which facilitates solubility and prevents protein aggregation. MMS, mainly detergents, have been used for IMP solubilisation and extraction for over a hundred years.[29] In the last two decades, different MMS have been developed, improving the nativeness of the IMP environment. Some of the common MMS are depicted in Figure 2.3, ordered from the most native (membrane) to the least native (mimetic). One of the factors determining their nativeness is the presence of (native) lipids, which can be completely absent in the less native MMS like detergents and amphipols. Other factors are also important, such as the membrane curvature affecting lipid packing and distribution as well as exerting lateral forces on the inserted IMPs.[24] This is important to consider in certain MMS, such as small vesicles or liposomes, although the specific membrane curvature a given IMP might experience in a native membrane remains a field of active research.

The level of nativeness is not necessarily the only factor to consider in the choice of MMS. Most of the common MMS require the use of detergents for membrane solubilisation and IMP extraction, except the styrene-maleic acid copolymer lipid particles (SMALPs), also called native nanodiscs (NNs). This can expose the target IMP to a disruptive environment and complicates the experimental procedure, requiring detergent removal and second MMS insertion. The method of IMP characterisation should also be considered, as many MMS are unsuitable for specific techniques, due to *e.g.* the levels of background noise, polydispersity, low stability or size. Each MMS has its own advantages and disadvantages, and its nativeness is

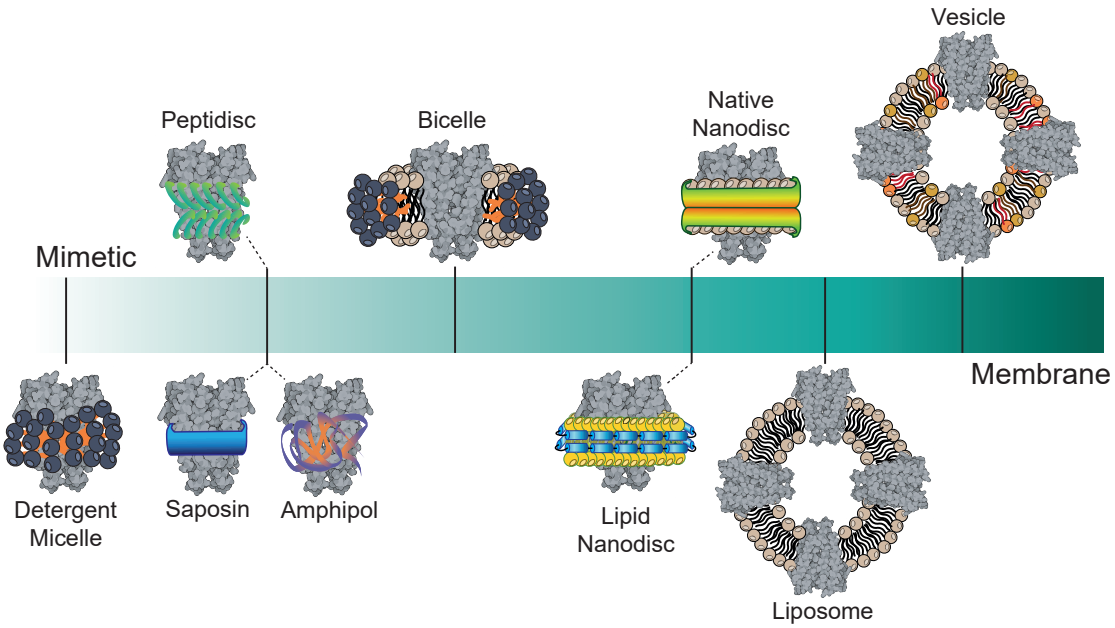


Figure 2.3: Different MMS ordered by the level of nativeness. Selected MMS are presented on the scale from the most native PL bilayer (membrane) to the least native (mimetic). Detergent micelles are considered the least native type of MMS, whilst native vesicles most closely resemble the conditions in the native lipid bilayer. Model protein shows a cartoon representation of KcsA (grey, PDB: 1BL8) created using the software *Illustrate.*[30] Adapted by permission from Springer Nature: *Nature Chemical Biology*, Importance of the Membrane for Biophysical Measurements, D. S. Chorev *et al.*, ©2020.[14]

just one factor to consider. Regardless, there has been a shift towards the use of more native MMS. For example, lipid nanodiscs and SMALPs balance maintaining the native function and interactions with their applicability to various techniques, high stability and ease of use. Four commonly used MMS were chosen for mass photometry investigation, aiming to cover both highly artificial (detergents and amphipols) and more native (lipid and native nanodiscs) MMS. Other MMS, whilst highly important and indispensable for many techniques, such as bicelles, saposins, liposomes, vesicles, or peptidiscs, will not be discussed in this section.

2.2.1 Detergents

Detergents, also called surfactants due to their surface activity, are small amphipathic molecules composed of a hydrophilic, charged section and a fatty, hydrophobic tail (Figure 2.4A). The polar head group confers some degree of water solubility, whilst

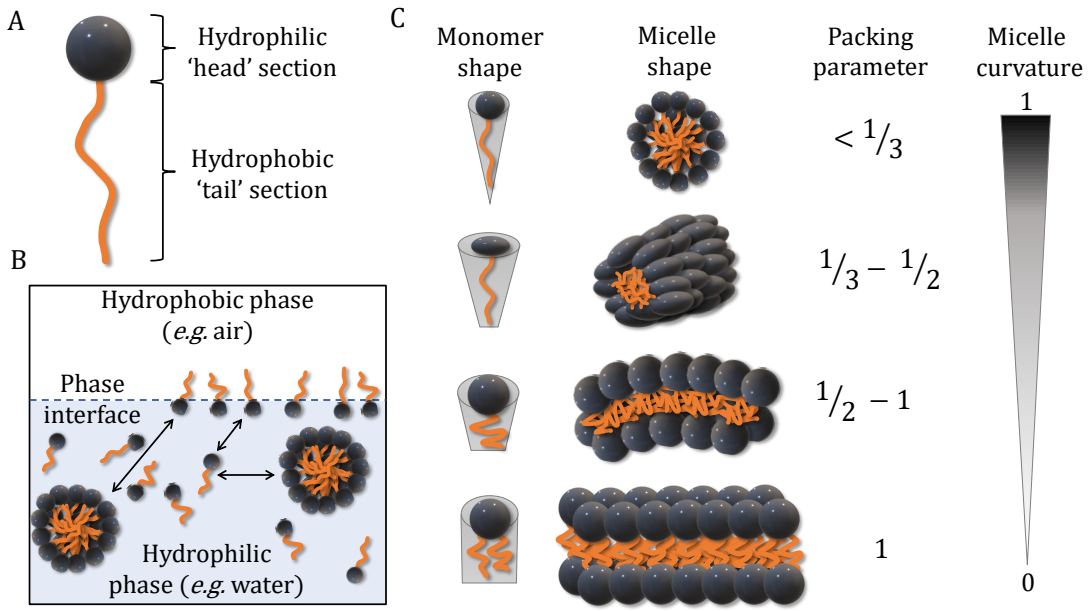


Figure 2.4: Detergent interactions and micelle shapes. A. The general structure of a detergent molecule. **B.** The behaviour of detergent monomers in a hydrophilic solution, such as water, containing an interface with a hydrophobic medium such as air. A detergent monolayer forms at the interface, with solubilised monomers and micelles present in the solution. **C.** The shape of a detergent monomer ranges from conical to cylindrical, affecting its packing parameter, usually between $< {}^{1}/_{3}$ (conical) and 1 (cylindrical). This impacts the micelle shape and curvature, resulting in a variety of shapes, including spherical, detergent tubes, vesicles, or planar bilayers. Adapted with permission from Springer Nature Customer Service Centre GmbH: **Springer International Publishing**, *The Next Generation in Membrane Protein Structure Determination*, Detergents in Membrane Protein Purification and Crystallisation by A. Anandan *et al.*, 13-26. ©2016.[32]

the fatty tail facilitates binding to hydrophobic surfaces. Detergent monomers in a polar solvent (such as water) tend to self-associate into micelles.[29] Although chemically relatively simple, they behave in a complex manner when interacting with themselves and other molecules, such as PLs or IMPs. The micellar and phase behaviour of detergents is strongly affected by their physicochemical characteristics and the monomer concentration, as well as the ionic strength and pH of the solution, the presence of lipids, IMPs, and other molecules.[31]

Initially, naturally occurring saponins and bile salts were used for denaturing proteins[33] and IMP extraction, although with a loss of activity.[34] Synthesis of

milder, non-ionic surfactants, such as Triton X-100 and Tween-20, enabled the isolation and characterisation of many IMPs as well as the viral membrane, its composition and function of the membrane glycoproteins.[35–37] Nowadays, the number of detergents available is easily in the thousands, with ongoing research into novel molecules with improved properties.

Chemical and Biophysical Properties of Detergents

The hydrophobic segment of a detergent varies in size, negatively affecting its water solubility, which is counteracted by the hydrophilic group.[29] Detergents are often described by their head group charge: (1) ionic (cationic/anionic); (2) zwitterionic, containing a positive and a negative charge, overall neutral; and (3) non-ionic, composed of hydrophilic groups that are not charged. Micelle formation is another defining property of detergents, occurring in aqueous solutions above the critical micelle concentration (CMC) and temperature (CMT). Below the CMC, detergents exist in an equilibrium between solubilised monomers and a monolayer at a phase interface, usually between water and air (Figure 2.4B). Above the CMC, detergents spontaneously aggregate into micelles. This process is driven predominantly by hydrophobic forces, where the fatty tails orient themselves inwards, approaching the liquid state of pure hydrocarbon chains.[29] This prevents them from interfering with the hydrogen bond formation between the polar groups and the surrounding water. As such, micelle formation can be described as a phase separation phenomenon.[29]

The shape of a detergent monomer affects the micelle packing efficiency and its shape, which ranges from spherical or cylindrical micelles to curved or planar bilayers (Figure 2.4C). It is described by the packing parameter, $P = v/la$, where v is the hydrocarbon chain volume, l is its length, and a is the equilibrium area per monomer at the aggregate (micelle) surface.[38] The packing parameter ranges tend to correlate with the shapes formed, although the pH and the ionic concentration of the

surrounding solution are also important due to the electrostatic interactions between the hydrophilic head group of a detergent monomer and the level of hydration it experiences. Whilst many of the micelles studied can be approximated as roughly spherical, variations in the shape, temperature, concentration, and the presence of lipids or proteins can lead to other shapes, such as ellipsoid or rod-like.[39,40]

The CMC is not a single, set value, but rather a narrow range of concentrations, dependent on the temperature, impurities and biological molecules, and the ionic concentration. It can be determined by measuring the changes in water surface tension or conductivity, or with organic fluorescent dyes. These absorb and emit light depending on their immediate molecular environment and partition into the hydrophobic micellar interior above the CMC.[41] Additionally, the CMC and the aggregate behaviour change with the temperature and detergent concentration. The intersection of the CMC and CMT is the Krafft point, describing the lowest concentration and temperature at which micelles form. Below the CMT and above the CMC, a hydrated crystalline solid phase can occur, sometimes forming solid detergent crystals. At certain concentrations and temperatures, phase separation can occur, with the micelles partitioning into a detergent-rich phase. This can also be achieved by changing the salt content, sometimes exploited for IMP isolation and purification.[42] In general, hydrophilic groups oppose micelle formation, bonding preferentially with water. Consequently, the CMC of ionic detergents is often higher than non-ionic.[29] The micelle size follows the opposite trend, with ionic detergents having lower aggregation number.[29]

The high CMC, small micelle size, and the strong denaturating effects on proteins of ionic detergents make them more suitable for other applications, such as sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). On the other hand, zwitterionic detergents, such as lauryl-dimethylamine-N-oxide (LDAO) tend

to be less denaturing, and are better at membrane solubilisation than non-ionic detergents, and are often considered the 'middle ground'.[43] Nevertheless, non-ionic detergents generally have the least impact on the protein structure and function, and are often the best suited for IMP extraction and study.

Despite some of their potential drawbacks, detergents have provided critical contributions to the field of membrane protein study. The vast majority of IMP research over the past 100 years has relied on the use of detergents for IMP extraction and solubilisation.[31] Whilst some of the novel solubilisation methods, such as lipid or native nanodiscs, are becoming increasingly popular, they were only invented over the last twenty years. Indeed, out of all the membrane protein structures deposited in the Protein Data Bank (PDB) between 2010–2019, the vast majority were produced using detergents.[44] Furthermore, categorising the structures deposited by their respective membrane protein classes reveals that only in the TRP channel class, which contains a few dozen structures, the number of structures determined through other membrane mimetic systems (lipid nanodiscs and amphipols, in this case) approaches half.[44] Additionally, recent developments of novel, non-ionic detergents, such as lauryl maltose neopentyl glycol (LMNG) or glyco-diosgenin (DGN), remove many of the theoretical downsides potentially associated with the use of some of the classical detergents, such as excessive delipidation, denaturation, and high background signal. Overall, it is important to note that not only have detergents been instrumental in progressing the field of IMP research, they also remain at the forefront of scientific discovery, and their importance to this research area cannot be understated.

Detergent Interactions with the Membrane and IMPs

Understanding detergent interactions with IMPs and PLs is important, but even a simple addition of lipids causes changes in the CMC, micelle shape and size, and

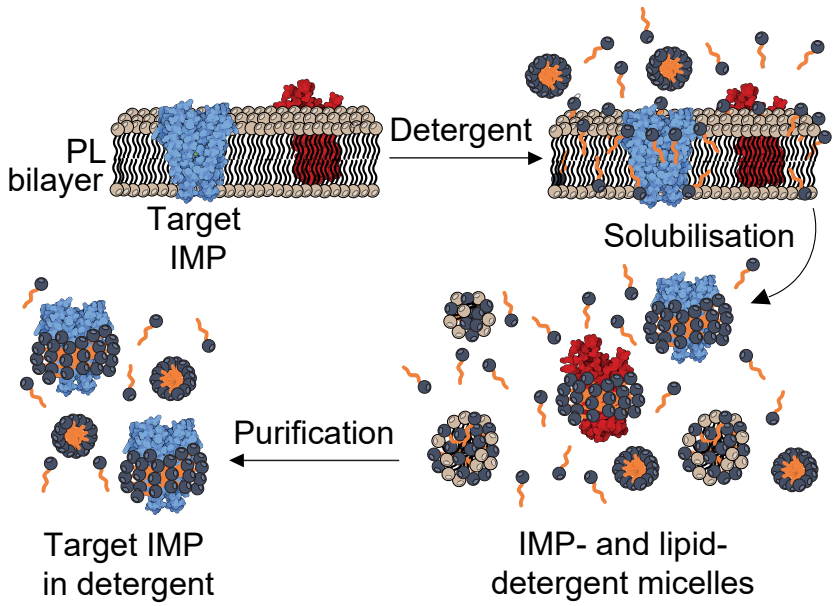


Figure 2.5: Detergent solubilisation of the membrane and membrane proteins. An excess of a detergent is added to the isolated PL bilayers containing IMPs. The detergent solubilises the membrane and IMPs by inserting into the bilayer, breaking it up and forming mixed lipid, IMP, and detergent micelles. The target IMP can then be purified for further experiments. Cartoon representations of KcsA (blue, PDB:1BL8) and OmpA (red, PDB:1BXW) were created using the Illustrate software.[30]

phase behaviour.[31,45] PL solubilisation can be described in three stages, accompanied by sharp phase transitions. In stage I, low concentration detergent monomers insert into the PL bilayer. Stage II occurs upon bilayer saturation, with mixed PL-detergent micelles likely present as well. At higher detergent concentrations in stage III, the PL bilayer is fully solubilised into mixed PL-detergent micelles.[31] The heterogeneity of a native membrane leads to a highly complicated system, where these phase transitions may not be present. The interactions between detergents and proteins vary with the type of protein. Soluble and membrane-associated proteins (MAPs) are often denatured, especially if interacting with stronger, ionic detergents,[29] whereas the TM region of IMPs binds the fatty detergent tails, facilitating water solubilisation. A general model for detergent solubilisation of the PL bilayer is outlined in Figure 2.5.

The interactions between detergents and IMPs are very difficult to predict due to their complex properties. A level of delipidation during IMP extraction is necessary for the IMP liberation from the lipid bilayer, ideally removing most of the bulk lipids for solubilisation whilst preserving annular lipids essential for IMP function.[46] IMPs also react differently to the same detergent. For example, Triton X-100 preserves the activity of some IMPs, such as D-alanine carboxypeptidase[47] or the cholinergic receptor,[48] but interferes with the function of cytochrome oxidase.[49]

High detergent concentrations common during IMP extraction often promote protein monomerisation as the detergent molecules replace the protein contacts between the IMP monomers, whereas low concentrations can lead to IMP aggregation due to lack of detergent molecules to cover the exposed hydrophobic TM sections of the IMP. As such, identification of the native IMP oligomeric state using detergents is non-trivial, as different methods, detergents and concentrations can produce inconsistent results. Still, other experimental factors heavily influence the success of IMP extraction, including the pH and ionic strength of the buffer, the presence of lyotropic salts, ligands, or chemical protective agents, such as glycerol or sugars.[46] As such, studies of novel IMP systems often require much trial-and-error due to high complexity of these systems, testing for maintenance of IMP structure and function, presence of native lipids, as well as native oligomerisation status.

The techniques used for IMP characterisation also affect the detergent choice.[50] High concentrations of micelles cause excessive background signal, problematic for many techniques. Detergent removal or exchange can be performed using size exclusion chromatography (SEC), dialysis,[51] detergent-adsorbing hydrophobic resins (such as BioBeads),[52] or ion exchange chromatography.[53] As the rate of protein-detergent dissociation can be similar to that of micelles, protein aggregation and precipitation often occur below the CMC. Apart from the background noise, high

detergent concentrations also impart changes to the solution properties, such as the viscosity, causing issues with *e.g.* solution NMR. Nevertheless, some novel detergents, such as LMNG, allow the removal of practically all micelles without losing IMP solubilisation,[54] due to the slow kinetic off-rate of the monomers from the IMP,[16,55] facilitating many successful structural and functional studies.

2.2.2 Amphipols

Amphipols (APols) are short, amhipathic polymers, originally formed by grafting hydrophobic molecules, such as octylamines, onto a strongly hydrophilic backbone, such as low MW polyacrylic acid.[56] The attachment is performed by creating an amide bond between varying proportions of the carboxylic groups of the polyacrylic acid precursors and the amino group of amines with a hydrocarbon chain. The most commonly used APol, A8-35, consists of 35% ungrafted carboxylic groups, with 25% and 40% of the rest of the carboxylic groups attached to octylamines and isopropylamines, respectively (Figure 2.6).[56] Recent analysis showed that A8-35 monomers weigh on average 4.3 kDa,[57] although the MW distribution appears to be quite broad.[58] This relatively short length increases the homogeneity of IMP-APol complexes and reduces intermolecular bridging.[58]

Since their development in 1996,[56] a variety of APol molecules has been developed, mainly substituting the hydrophobic groups, or using a different backbone entirely.[58] One of the main aims of novel APol synthesis has been to mitigate their downsides, namely the sensitivity to low pH and divalent cations, and reduce the potential interaction with the extracellular sections of the IMP, all caused by the negative charge of the backbone.[56,58] However, balancing the hydrophobic side chains with a polar but neutral backbone whilst maintaining water solubility and IMP solvating effects has proven challenging. Nevertheless, successful non-ionic APols (NAPols) have been developed, with glucose facilitating the hydrophilic

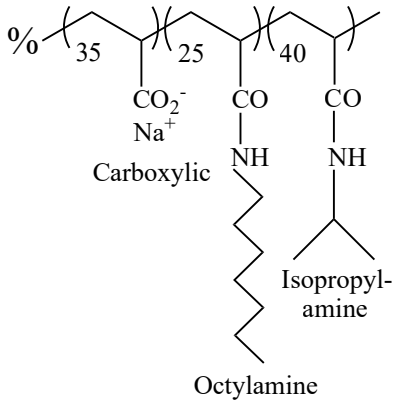


Figure 2.6: Molecular subunit structure of the A8-35 amphipol. The amphipathic polymer APol A8-35 is composed of a hydrophilic polyacrylic acid backbone with 35% ungrafted carboxylic acid groups, 25% amide-bonded octylamines and 40% amide-bonded isopropylamides.

interactions,[59] seemingly escaping the problems outlined above. They can also be used to solubilise IMPs during cell-free synthesis, which is strongly inhibited by ionic APols.[60] Alternatively, isotopic labelling was used to study APol distribution and the APol:IMP ratio,[61] whilst fluorescence labelling helped to map the TM region of an IMP with FRET.[62] Adding a biotin or a polyhistidine tag enables IMP-APol complex immobilisation for surface plasmon resonance (SPR), recently applied to GPCR ligand screening.[63]

Chemical and Biophysical Properties of Amphipols

All varieties of APol are highly soluble in water,[58] assembling into small particles that sequester the hydrophobic chains similarly to detergent micelles with a well-defined, narrow size range.[64,65] A8-35 particles are the most studied, forming roughly globular particles of ∼6.3 nm in diameter and 40 kDa MW, equivalent to ∼9 A8-35 molecules. Their critical aggregation concentration is very low (∼0.002 g/L),[66] suggesting that in solution and during IMP solubilisation, A8-35 is mainly present in this form.

The assembly of these structures has also been studied using molecular dynamics (MD). Upon suspension in an aqueous solution, the A8-35 monomers collapse onto themselves, followed by spontaneous assembly into globular aggregates.[67] However, their surface viscosity appears higher than that of detergent micelles or a PL bilayer,[68] potentially restricting conformational changes of APol-solubilised IMPs whilst also contributing to their stabilising effect on the protein structure.

The high solubility of classical APols is affected by their sensitivity to pH and divalent cations. Since the solubility depends on the interactions between the negatively charged carboxylic groups on the backbone and water molecules, pH below 7 or the addition of divalent cations prevent the formation of these hydrogen bonds, resulting in APol precipitation.[61,65,69] This major drawback, whilst useful in some circumstances, was the major driver for the development of alternatives to the traditional polyacrylic acid backbone APols, such as NAPols,[64] phosphocholine APols (PC-APols),[70] or sulphonated APols (SAPols),[71] which overcome these limitations.

Amphipol Interactions with the Membrane and IMPs

APols were designed as an alternative to detergents for IMP study. Due to their amphipathic nature, they can bind to the hydrophobic TM sections of IMPs, generally with higher affinity than detergents, shielding them from the surrounding aqueous solution.[72,73] APols circumvent some of the problems associated with detergents, mainly IMP inactivation due to extensive removal of often essential lipids or hydrophobic cofactors, and subunit dissociation.[74]

APols have also proven advantageous for a variety of techniques. The low CMC enables excess APol micelle removal without IMP aggregation, keeping the overall APol concentration minimal, beneficial for structural studies and NMR, where the

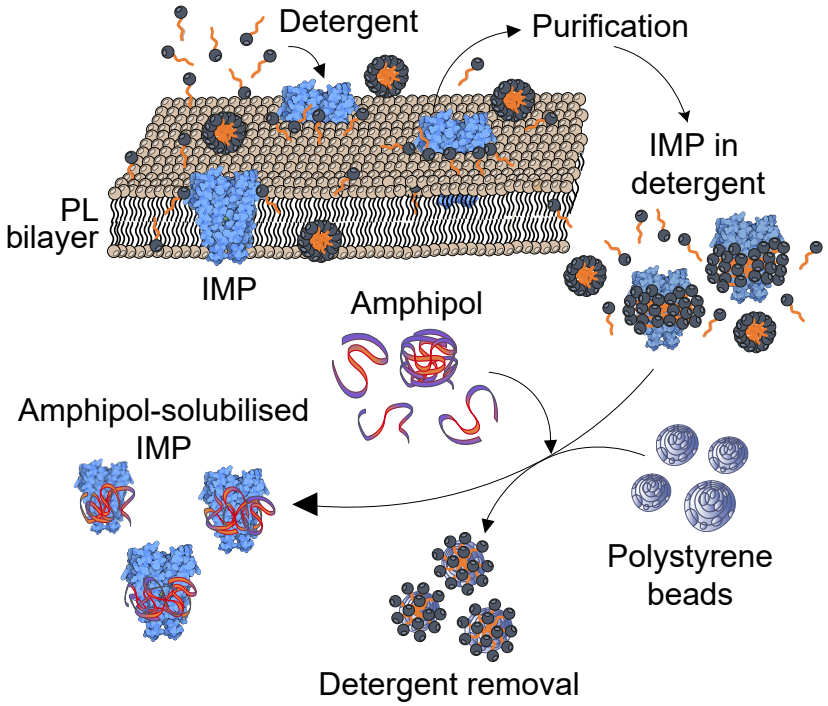


Figure 2.7: IMP solubilisation in amphipols. The isolated PL bilayers, containing the IMP of interest, are solubilised with an appropriate detergent. The IMP-containing detergent micelles are then extracted and purified, after which amphipol (APol) molecules are added and detergent molecules removed through porous adsorbent beads (*e.g.* polystyrene BioBeads) or other methods, such as dialysis. After the detergent removal, IMP is solubilised in small APol-IMP complexes, and excess APol particles are removed.

associated increase in solution viscosity at high detergent concentrations can be problematic.[56] The low IMP-APol kinetic off-rate, caused by multiple points of APol attachment, contributes to the efficient removal of the empty APol particles (see Table 2.1 for the specific values and comparison with detergents).[74] Additionally, their comparatively large size largely prevents the solubilisation of individual lipids and small hydrophobic co-factors, which instead associate with the IMPs, aiding in activity restoration.[74] On the other hand, this low affinity to lipids means that APols cannot solubilise IMPs directly from lipid bilayers.[75,76] As such, detergent-facilitated IMP solubilisation and purification is usually necessary prior to APol transfer. The simplified process of IMP solubilisation in APol is outlined in Figure 2.7.

	APol	**APol – $C_{12}E_8$**	**LMNG**	**DDM**
Off-rate (s^{-1})	0.033	> 0.69	143	1.43×10^6

Table 2.1: Off-rate comparison between APol and detergents. Comparison of the off-rates measured for a labelled APol A8-35–IMP complex exchanging with free, unlabelled A8-35 molecules or between APol and the detergent $C_{12}E_8$,[77] as well as between LMNG and the GPCR βA_2R, or DDM and βA_2R.[78] The off-rates are shown as K_{off} values in seconds^{-1}, with the APol values calculated from $t_{1/2}$ half-time decay data, provided by Zoonens *et al.*[77]

APols have now been used to study around 40 IMPs and IMP complexes,[58] containing hydrophobic areas ranging from a single TM domain,[74] regular α-helical bundles or β-barrels, to the extended hydrophobic surface area of the 1.7 MDa respirasome.[79] The IMP-APol complexes studied to date mainly use A8-35, characterising properties such as the area of contact between APol and the selected IMP, the shape, composition, dynamics, and stability of the complex, and the presence and identity of the bound lipids.[58] A8-35 adsorbs onto the hydrophobic TM region of the IMP, forming a thin layer of ∼1.5-2 nm.[61,79,80] Whilst no specific contacts between APols and the hydrophobic patches of the IMP extracellular (EC) loops have been observed or predicted with MD simulations, transient interactions between the basic EC residues of the bacterial outer membrane protein X (OmpX) and the carboxylic groups of the APol backbone have been predicted by MD,[80] and observed in the case of a soluble protein, lysozyme.[76]

The low IMP-APol off-rate significantly increases the stability of the IMP-APol complex, especially when compared to detergents, manifesting in *e.g.* high IMP denaturation temperature,[81,82] which correlates with improved IMP stability during long-term storage.[76,82–84] However, the extent and source of the stabilising effects vary between IMPs and APols, where the lower charge density of NAPols appears to be beneficial.[60,83] Native lipids added during the APol-detergent exchange that

bind to the IMP upon detergent removal also increase IMP stability, in addition to contributing to its native function reconstitution.[61,85]

The number of APol molecules bound per IMP can be difficult to measure with great accuracy, and is generally determined by AUC, chemical, spectroscopic, and SANS measurements, or MD simulations.[58] It appears to increase only slowly with larger hydrophobic TM regions, and does not decrease much below the size of the empty A8-35 APol particles for small IMPs (∼40 kDa). MP studies of the bacterial outer membrane protein F (OmpF), described in more detail in Chapter 3, have shown that the amount of bound APol appears to depend on the APol:IMP ratio used during APol-detergent exchange, with OmpF oligomerisation and aggregation observed below a critical APol:IMP ratio.[16]

Similar to detergents, some free APol particles appear necessary for the initial IMP solubilisation, although 1:1 weight/weight (w/w) ratio seems to be sufficient, amounting to less background compared to most detergents.[61,84,86] The absence of sufficient free APol particles can result in IMP oligomerisation, likely due to the equilibrium between solution APol particles and IMP-APol complexes, resulting in non-specific IMP-IMP interactions through their exposed TM region.[16,86] Nevertheless, once inserted into APols, IMPs tend to remain solubilised upon extensive dilution in APol-free buffer and after excessive washing of surface-immobilised APol-IMP complexes.[77,87] These observations could be potentially explained by the high energetic barrier of exposing the hydrophobic IMP sections to water, combined with the slow off-rate between IMP and APols.[66,87]

Since the interactions between APols and the IMP extracellular (EC) domain appear limited in most cases, APol solubilisation does not seem to affect the binding of small, soluble ligands to TM receptors and channels. This has been exemplified

by ligands binding to G protein-coupled receptors (GPCRs),[60,63,82,88] or calcium and adenosine triphosphate (ATP) binding to the sarcoplasmic/endoplasmic reticulum calcium ATPase (SERCA) 1a.[76] Similarly, interactions with antibodies[89] or GPCRs binding to solubilised G proteins are not impacted.[60] APols also appear to generally support IMP function, often showing the levels of functionality observed in a PL bilayer,[58,61,62] or at least comparable to detergents.[84] However, the function of some IMPs, such as SERCA1a, is reversibly inhibited by APols, likely due to the extensive level of conformational rearrangement necessary for its function prevented by the tight interactions with APols.[83] As such, whilst generally useful for examining the functional properties of IMP systems, there are some restrictions that need to be kept in mind.

The use of APols has been validated for numerous structural techniques, such as NMR or cryo-EM. APols may be particularly well-suited for cryo-EM studies due to the sample homogeneity, IMP stabilisation and the absence of free APol aggregates. On the other hand, APol complex structure determination with X-ray crystallography has been described as 'highly frustrating',[58,90] likely as a result of the strong electrostatic repulsion between the charged hydrophilic backbones, preventing tight packing of the complexes in a crystal.[90]

APols have also been used widely for studies with MS, whether in the form of matrix-assisted laser-desorption ionisation (MALDI),[85] or electrospray ionisation (ESI).[58,91] Generally speaking, MS can detect most of the IMP subunits, sometimes with the associated lipids as well.[85] Additionally, APols have been shown to be better at stabilising IMPs in the gas phase during non-covalent (native) MS compared to detergents, resulting in less unfolding and structural collapse.

APols have also been found to facilitate the refolding of denatured IMPs. These are often isolated from overexpresion-induced inclusion bodies solubilised by high concentrations of chaotropic agents, such as urea or SDS.[60,82] APols often result in 60-90% yields, further increased by the addition of native lipids.[73,92] This process allows for IMP purification with APols without the need for detergents, simplifying the process and preventing the potential negative effects of detergent exposure on sensitive IMPs.[60,71,82] Additionally, studying this process can also provide some important clues towards the general minimal conditions necessary for IMP folding *in vitro*.[58] Alternatively, NAPols have been used successfully to aid IMP folding in cell-free protein synthesis,[60] during which minimal RNA production and protein expression complexes are isolated *in vitro*, usually from a cell lysate, and used for protein production.[93,94]

The higher protein stability conferred by APols, together with the low amounts of empty APol aggregates necessary for IMP solubilisation and lesser changes in the solution physicochemical properties make APols an attractive MMS with wide adaptations across many structural and functional techniques. Nevertheless, there are multiple disadvantages associated with APols. The necessity of solubilising the cell membranes and extracting the target IMP with detergents, followed by APol insertion (Figure 2.7) complicates the use of APols, and exposes potentially sensitive IMPs to detergents. Despite the commonly observed near-native preservation of IMP function, APols create a highly restrictive environment, dampening or preventing the natural conformational changes associated with IMP function,[58,83] and are chemically and structurally far removed from a native, or even native-like IMP environment.

2.2.3 Lipid Nanodiscs

The development of a method to maintain IMPs in a state close to the native bilayer environment whilst retaining water solubility has been transformative for IMP

research. Empty (devoid of any proteins) lipid nanodiscs (LNs) were first described in 2002, created by mixing genetically engineered amphipathic helical proteins, called membrane scaffold proteins (MSPs), with synthetic saturated or unsaturated PLs and detergents.[95] Upon detergent removal, the MSPs and PLs spontaneously self-assembled into discoidal particles with a total diameter (R_2, Figure 2.8A) of ∼9.1-9.9 nm, forming nanoparticles containing two copies of the MSP and, in the case of the first described MSP (MSP1), either ∼124 or ∼156 molecules of the unsaturated lipid 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine (POPC) or the saturated lipid 1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC) across the two PL membrane leaflets, respectively.[95,96] The MSPs used were derived from apolipoprotein A-I (apo-AI) through the removal of its globular N-terminal domain,[95] the absence of which is critical for LN homogeneity.[96]

The discoidal shapes of the immature HDL particles and LNs have been attributed to the interactions between the hydrophobic side of the amphipathic apo-AI or MSP α-helices and the hydrophobic alkyl chains of the PLs, with two copies of the protein encapsulating the edges of the two PL bilayer leaflets, stabilising the structure (Figure 2.8A).[97–99] Whilst there has been some discussion about the exact orientation of the α-helices, labelled polyethylene glycol precipitated nanodiscs studied by solid-state NMR provided clear evidence for the parallel orientation of the MSP to the PL bilayer surface (the 'belt' model, shown in Figure 2.8A), instead of the perpendicular orientation (the 'picket fence' model).[100]

Since the MSPs encapsulate the PL bilayer, their length is the main determinant of the overall LN diameter.[96,102,103] To better suit the variation in the TM region size of different IMPs, both shorter and longer MSP versions have been developed, mainly by altering the most commonly used MSP1D1. This construct consists of 10 amphipathic α-helices, an 11 amino acid deletion in the helix 1, an N-terminal

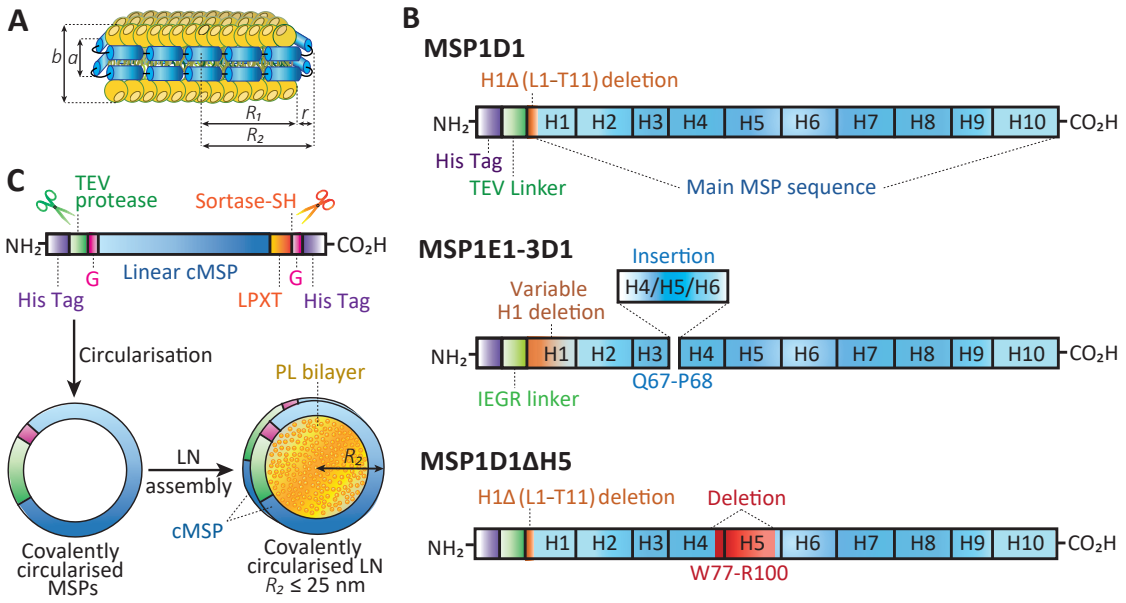


Figure 2.8: Lipid nanodisc, MSP, and circularised nanodisc schematics. A. A cartoon outline of an empty lipid nanodisc (LN), consisting of a PL bilayer (yellow) encircled by two membrane scaffold proteins (MSP, blue). The LN can be described by the average MSP α-helix diameter (r), the MSP belt thickness (a), the LN width (b), and its inner (R_1) and outer radius (R_2). **B.** A schematic of some commonly used MSP constructs, including the standard length MSP1D1 (top), the elongated MSP1E1-3D1 (middle), and the truncated MSPD1ΔH5 (bottom), showing any tags, linkers, insertions and deletions. **C.** A schematic of the circularised MSP (cMSP) construct in the linear and circularised forms, indicating its enzymatic cleavage sites. During circularisation, the N-terminal His-Tag and tobacco etch virus (TEV) linker are cleaved off, leaving an N-terminal glycine (G). Sortase excises the C-terminal His-Tag and G, ligating the N-terminal G to the LPXT sequence. The cMSPs are then used for the assembly of LN of up to 25 nm. Section **C** was adapted with permission from K. M. Padmanabha Das *et al.*, Large Nanodiscs: A Potential Game Changer in Structural Biology of Membrane Protein Complexes and Virus Entry, *Frontiers in Bioengineering and Biotechnology*, 2020, **8**, 539.[101] Licenced under the Creative Commons Attribution License (CC BY).

His-Tag, and a linker (Figure 2.8B, top). The longer MSPs were initially made by inserting α-helices into the middle of the MSP1D1 sequence, creating MSP1E1-3D1 containing 1-3 extra α-helices (Figure 2.8B, middle), which increased the LN diameter to ∼12.9 nm for the MSP1E3D1 version.[96] The shorter MSPs were first created by deleting the 22 N-terminal amino acids in the first MSP1D1 helix,[96] although the more commonly used truncated MSPD1ΔH5 contains deletions in the helix 5 instead (Figure 2.8B, bottom).[104]

Despite these manipulations, the LN size range remained limited. More recent MSP constructs created LNs as small as 6.3 nm, useful for studying smaller IMPs, such as OmpX (18.6 kDa) or bacteriorhodopsin (28 kDa),[104] particularly applicable to NMR.[105] This MSP engineering approach also extended the upper limit, yielding LNs of up to ∼16-17 nm by linking two copies of MSP1D1. Nevertheless, this size remained unsuitable for investigating large membrane supramolecular complexes and assemblies.[102]

The upper limits of the LN size were significantly extended by the development of covalently circularised MSPs (cMSP) through a C-terminus sortase recognition sequence (LPGTG) and a glycine (G) residue near the N-terminus, increasing the stability and homogeneity of the circularised LNs (cLN) (Figure 2.8C).[106] This enabled significant extensions to the MSP backbone, producing cLNs of up to 80 nm through cMSP linking.[101,106] Unfortunately, the already low cMSP yield decreased with its increasing length. Together with the high complexity of the circularisation protocol, their applications remained limited. Recently, a novel one-step protein ligation approach, the SpyCatcher-SpyTag technology,[107] was applied to circularise MSPs and produce cLNs of up to 100 nm.[108] The cMSPs circularise spontaneously during protein expression, significantly improving their stability and solubility, preventing aggregation and resulting cellular degradation. This method achieved 10 × higher yields of the larger cMSPs and substantially simplified their production.[108] However, their large size precludes purification through SEC, resulting in a broad size distribution, with the hopes of alternative or novel purification methods resolving this issue.[108]

Since the development of LNs, significant efforts have been directed at understanding and improving the MSPs, demonstrated by the extensive research aimed

at creating large LNs. However, the lipid component of the LNs has received less attention. Since the activity and function of IMPs generally depend on its interactions with the surrounding lipid bilayer, understanding the structure, dynamics, and biophysical properties of the LN PL bilayer is crucial for their optimisation and appropriate result interpretation.

Even though it might appear reasonable to assume similar or identical lipid behaviour in LNs and lipid bilayers, vesicles, or liposomes, the physical constraints of the MSP belt exert significant effects on the lipids, creating a distinct, non-uniform lipid environment.[109–113] SANS experiments using DMPC apo-AI LNs indicated that the area occupied by a single DMPC head group ($\sim$50 Å^2) was significantly smaller than in vesicles ($\sim$56.7 Å^2), suggesting tighter head group packing.[110] Similarly, comparing the fluidity of the LN bilayers with that of DMPC large unilamellar vesicles showed higher acyl chain order in the LNs, associated with tighter PL packing.[110] Tighter PL packing in LNs compared to liposomes was also demonstrated through sets of lipid spin labels probing different bilayer regions using EPR.[114] These observations correlate with the higher phase transition temperature observed in LNs than in a PL bilayer, with important implications for the LN assembly, which only occurs in the liquid crystalline phase.[115]

Incorporating Integral Membrane Proteins into Lipid Nanodiscs

Like APols, MSPs are incapable of inserting into and solubilising cell membranes. Instead, detergents are generally used for the extraction and purification of the target IMP, followed by mixing with the selected PLs and MSPs at a specific molar ratio, and incubation.[95] The detergent is then gradually removed, facilitating the LN self-assembly, which is initiated by the lipid bilayer formation when the detergent concentration falls below the CMC, with the LN assembly occurring upon further detergent removal (outlined in Figure 2.9).[109] Finally, SEC or affinity

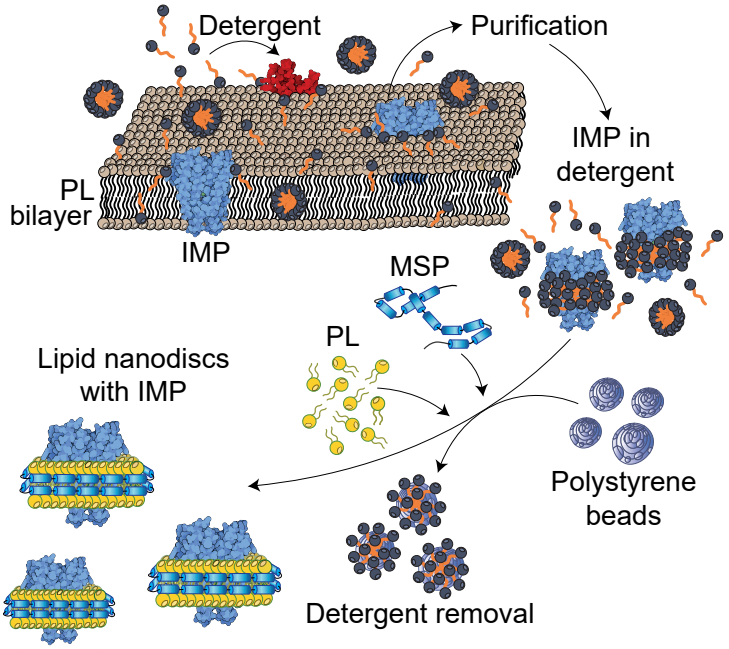


Figure 2.9: The process of IMP insertion into lipid nanodiscs. The isolated PL bilayers, containing the IMP of interest, are solubilised with an appropriate detergent. IMP-detergent micelles are extracted and purified, and mixed at a specific molar ratio with selected PLs and membrane scaffold proteins (MSPs). After incubation, the detergent is gradually removed by the addition of an adsorbing material, such as polystyrene beads, or by dialysis. This facilitates protein solubilisation in the PLs and the LN self-assembly. Figure adapted with permission from A. Olerinyova and A. Sonn-Segev *et al.*, Mass Photometry of Membrane Proteins, *chem*, 2021, **7**, 224-236.[16] Licenced under the Creative Commons CC-BY-NC-ND license.

chromatography are usually used to remove the empty LNs and achieve a clean, homogenous sample.

IMP incorporation into LNs is a multistep process, requiring a number of prior estimations and calculations, and often extensive optimisation. IMP aggregation is generally prevented through an excess of MSPs and lipids whilst maintaining the correct IMP:lipid:MSP ratio. The type of detergent and lipids used, together with the identity of the target IMP and the rate of detergent removal critically affect the overall success rate.[103,116] Using previous data and the surface area of the PL used, the optimal lipid:MSP stoichiometry likely to result in full, but not overpacked LNs can be estimated,[96] deducting the hydrophobic cross-sectional area of the target

IMP from the total bilayer surface area.[116] The MSP:IMP ratio and the MSP length can be selected to favour the incorporation of predominantly monomers, dimers, or higher oligomers of the target IMP.[103,116] Whilst providing precise control over the overall IMP PL bilayer environment, the type of PLs used can impact both the yield and homogeneity of the LNs,[103] especially important for studies using *e.g.* NMR or cryo-EM. The choice of PL also strongly impacts the IMP structure and function, potentially aiding in the reconstitution of its native activity.[105]

The type of detergent used in the LN reconstitution process needs to balance efficient solubilisation of cellular membranes without affecting the IMP properties. Unfortunately, this is not always possible for sensitive proteins.[117,118] To mitigate the negative effects of the detergent, the isolated cell membranes can be mixed directly with the detergent, MSPs, and lipids, followed by rapid detergent removal after membrane solubilisation, minimising its effects.[103] This method also enables the purification and solubilisation of an entire library of IMPs expressed in a specific tissue, as exemplified by the extraction and solubilisation of the synaptic membrane proteome in LNs, facilitating the study of their interactions with the Alzheimer's disease-associated amyloid β-oligomers.[119]

Lipid Nanodiscs in Membrane Protein Research

LNs have been employed for IMP solubilisation in studies using a large variety of structural, functional, spectroscopic, and single-molecule techniques. Thanks to the generally high sample homogeneity, they are well-suited for ensemble structural methods like cryo-EM. This technique was used to study *e.g.* the construction of membrane-inserted toxins,[120] and the interactions and progression of a nascent polypeptide chain from the ribosomal peptidyltransferase into the membrane through the SecYEG translocon complex.[121] Cryo-EM was also able to determine the structure of MAPs upon membrane attachment, exemplified by Factor VIII binding

to the negatively-charged LNs of varying lipid composition.[122] Compared to their cryo-EM applications, LNs have long resisted most crystallisation attempts, possibly due to the LN bilayer inhomogeneity resulting in partial resistance to the gel-phase transition,[112] or the electrostatic repulsion between the outside of the MSPs. As with APols, IMP crystallisation in LNs was recently achieved *via* the protein transfer into a mesophase.[123]

Thanks to the solubilisation capabilities of LNs, experiments at high protein concentrations are possible without IMP aggregation, useful for *e.g.* nuclear magnetic resonance (NMR) or electroparamagnetic resonance (EPR). High-resolution NMR structures of IMPs embedded in LNs acquired through NMR have been published,[124] recently achieving an atomic-level detail of the studied IMPs in a near-native lipid bilayer environment.[104,125,126] Unfortunately, the size limitations of NMR prevent its application to medium or large LNs. EPR has been used to determine individual intramolecular distances, evaluating conformational changes and ligand binding dynamics.

The typical sample monodispersity and stability also make LNs well-adapted to MS. The use of hydrogen/deuterium exchange (H/DX) MS with LNs are complicated by the presence of MSPs and lipids, requiring removal,[127] and the large size of many IMPs, increasing the time needed for protein digestion post labelling and consequently the proportion of the bound deuterium ions diffusing out.[128] On the other hand, studying LNs using MALDI-TOF MS opens the door to many exciting applications, such as studying IMP protein interactions, or detecting post-translational modifications.[105] IMPs interactions with soluble proteins and MAPs can also be studied with proteomic MS, *e.g.* by combining stable isotope labelling of amino acids in cell culture (SILAC) and LC-MS/MS.[129]

In the last decade, native MS has emerged as a powerful tool to study native IMPs in intact LNs,[130] providing information about their stoichiometry, topology, and interactions with ligands, other proteins, and lipids.[131] The complexity of the LN spectra proved to be a challenge initially,[105,130] which was only overcome relatively recently,[132] although the analysis of the LN lipid mixtures most similar to the native membranes remains exceedingly complex.[133] LNs are also amenable to the study of IMP interactions within the membrane using FRET, partly due to the control over their lipid bilayer composition.[134]

LNs are also easily adapted to single molecule studies, where the PLs or the MSPs are tagged with an affinity tag, affixing the LN onto a functionalised surface to immobilise it, enabling access of substrates and ligands to both sides of the IMP. Immobilised LNs can be studied with *e.g.* single-molecule fluorescence spectroscopy,[135] atomic force microscopy,[103] or SPR.[136] Novel single-molecule in-solution techniques, such as mass photometry, enabled a new way of LNs characterisation, revealing information about LN composition and homogeneity in relation to the functional state of the inserted IMP.[5,16]

Despite their undeniable utility and widespread use, LNs face difficulties with IMP-IMP interaction studies. Including two or more IMPs in the LN assembly mixture results in a non-uniform, random incorporation and varying stoichiometries of the IMPs.[105] Even after successful IMP complex insertion and purification into a homogenous sample, several purification steps are often required, resulting in poor yield. The IMP orientation is also likely to be mismatched due to the bidirectionality of IMP incorporation, rendering a significant portion of the sample unusable.[105] As such, different MMS are generally better suited for probing these interactions.

Overall, LNs have proved irreplaceable in numerous structural and functional studies, providing a near-native environment for the study of IMP properties and interactions. Recent developments, such as the circularised MSPs synthesis,[106] enabled further LNs stabilisation and size expansion. Nevertheless, there are some notable drawbacks associated with LNs, such as their complicated assembly requiring extensive optimisation. Although some IMPs insert with $> 90\%$ efficiency, the success rate is often $< 10\%$, especially for GPCRs and large protein complexes.[109] On the other hand, the assembly provides control over the lipid composition, which can impact the structure and function of the target IMP. However, single PLs or simple PL mixtures do not replicate the complexity of a native PL bilayer. Whilst appropriate cell membrane lipid extracts can be used, the LN assembly process incorporates the lipids bidirectionally, resulting in symmetric membrane leaflets, in stark contrast to essentially all biological membranes.

As such, the constricted, simplified PL bilayer of LNs cannot fully replicate the compositional complexity, asymmetry or fluidity of native biological membranes, with the diffusion of lipids and signalling molecules, IMP-IMP interactions, membrane nanodomains or IMP oligomerisation underrepresented or completely absent in LNs. Despite these downsides, LNs have helped pushed the boundaries of knowledge and understanding of the structure and function of numerous IMPs, their conformational changes, reactivity, interactions with soluble proteins and the critical roles of lipids. They have enabled the investigation of complex biological phenomena in a near-native environment, greatly expanding membrane biology research in general. Their increasing popularity for IMP research is hence unsurprising, with the number of publications using LNs increasing exponentially.[109]

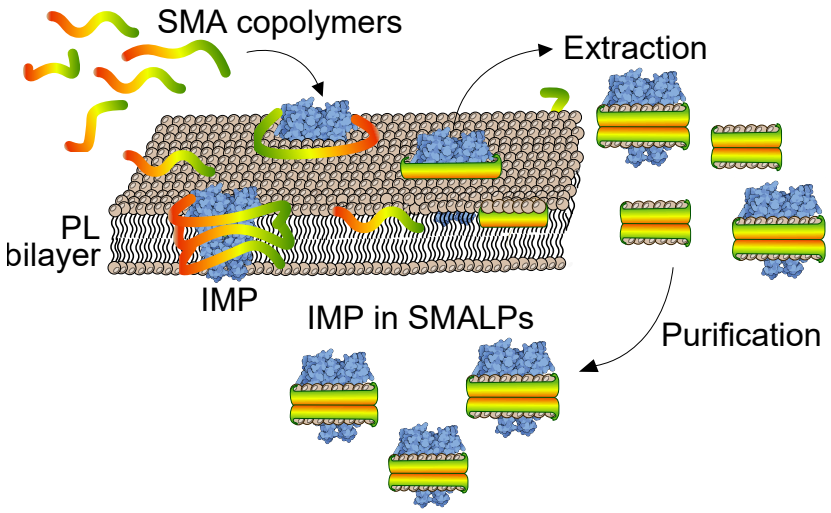


Figure 2.10: Membrane protein extraction using SMA. The isolated PL bilayers with the IMP of interest are solubilised with an excess of styrene-maleic acid (SMA) copolymers, which insert into the membrane and cut out nanodisc-like sections, some containing the IMP of interest. The SMA lipid particles (SMALPs) are then purified, yielding SMALPs with the inserted target IMP. Figure adapted with permission from A. Olerinyova and A. Sonn-Segev *et al.*, Mass Photometry of Membrane Proteins, *chem*, 2021, **7**, 224-236.[16] Licenced under the Creative Commons CC-BY-NC-ND license.

2.2.4 Native Nanodiscs

Native nanodiscs (NNs) are a novel MMS that generally refers to amphipathic polymer-encased nanoparticles containing sections of native membranes, used to extract and solubilise IMPs. These amphipathic polymers are characterised by alternating hydrophilic and hydrophobic groups, akin to those of APols. Unlike APols, however, they can adsorb to and disrupt liposomes and lipid bilayers, with some polymer variations able to insert themselves into the bilayer and wrap around the sections of PLs in a 'doughnut'-like manner, producing nanoparticles resembling HDL nanodiscs, illustrated in Figure 2.10.[137] This distinguishing feature was first described in the early 1980s intedended for the delivery of hydrophobic pharmaceuticals, and is thought to occur through a pH-induced change from a linear, extended chain conformation to an uncharged, coil-like secondary structure.[138]

Despite the membrane-solubilising and nanodisc-forming properties of these

amphipathic polymers being known for decades, they were only applied to IMP research in 2009. The first polymer shown to solubilise IMPs whilst retaining their function was the styrene-maleic acid (SMA) copolymer, solubilising PagP and bacteriorhodopsin from liposomes.[139] Besides the IMP, SMA/lipid particles (SMALPs) were also found to contain ∼10 surrounding lipids per protein,[139] also occurring during native membrane IMP extraction.[140–143] Considering the previously discussed function of these lipids, this property provides NNs with a significant advantage over other MMS that often remove most or all of the lipids, artificially adding potentially important ones to the IMP later.

The simplicity of the IMP extraction for NNs, illustrated in Figure 2.10, is another advantage of this MMS. Briefly, the isolated membranes with the target IMP are mixed with the selected polymer for ∼1 h at RT with gentle shaking, facilitating membrane solubilisation and NN formation. These are then separated by ultracentrifugation and purified with affinity chromatography or SEC.[143,144] Their high stability and the lack of micelle formation means that no further polymer needs to be added after the membrane solubilisation, enabling characterisation by techniques sensitive to background noise.[16,144,145]

As the SMA copolymer is synthesised by free-radical polymerisation, it is inherently polydisperse and random, with a wide variety of sizes present.[145] Roughly purified mixtures of specific sizes and maleic acid contents have been produced and tested for their IMP extraction and stabilisation abilities, and the SMA 2000 variant, with an average MW of ∼7.5 kDa and ∼33 % maleic acid content, produced the best overall results.[144] Generally, the SMA copolymer with a 2:1 or 3:1 styrene:maleic acid ratio has been widely used for IMP extraction and purification.[145,146] The size of NNs produced by SMA 2000 was found to be ∼8-9 nm, but interestingly, increasing the average MW of the copolymer did not result in the expected size

change, instead decreasing the size to ∼5 nm.[144] It remains unclear what governs the size of SMALPs, and although extraction of larger IMPs with up to 36 TM helices has been demonstrated,[147] the use these polymers for larger IMPs is restricted, including many multi-protein membrane complexes.

Besides the limited size range, there are issues associated with the hydrophobic styrene group and the hydrophilic maleic group of the SMA copolymer, with the structure shown in the top left of Figure 2.11. The styrene aromatic ring strongly absorbs UV light, hampering many spectroscopic applications, and can also non-specifically interact with certain column resins.[143,148,149] It might also have a more disruptive effect on the acyl-chain order of the lipids within the NN, creating a less native IMP environment.[150] Di-isobutylene maleic acid (DIBMA) copolymer, developed in 2017, overcomes these issues by replacing the aromatic styrene ring with an aliphatic chain (Figure 2.11, top right), resulting in more native lipid properties.[150] DIBMA also appears to form larger NNs, up to ∼29 nm, albeit at the expense of less efficient extraction of certain IMPs.[146]

Perhaps the most significant drawback of SMALPs is their inherent sensitivity to low pH and divalent cations, both caused by the binding of the positively charged ions to the negatively charged OH^- of the maleic acid groups.[143,144] This charge neutralisation leads to polymer aggregation, nanodisc disruption, and polymer and IMP precipitation.[151] Whilst DIBMA NNs appear to tolerate moderately higher amounts of divalent cations than SMA,[146,150] other novel amphipathic polymers have been developed to overcome these issues. Most newly developed amphipathic polymers focus on chemically adapting the hydrophilic group, aiming for an overall neutral or positive charge.[152,153] The carboxyl acids groups have been replaced with the zwitterionic phosphatidylcholine groups (termed zwitterionic SMA, or zSMA), which appear to tolerate low pH and divalent cations, as well as enabling

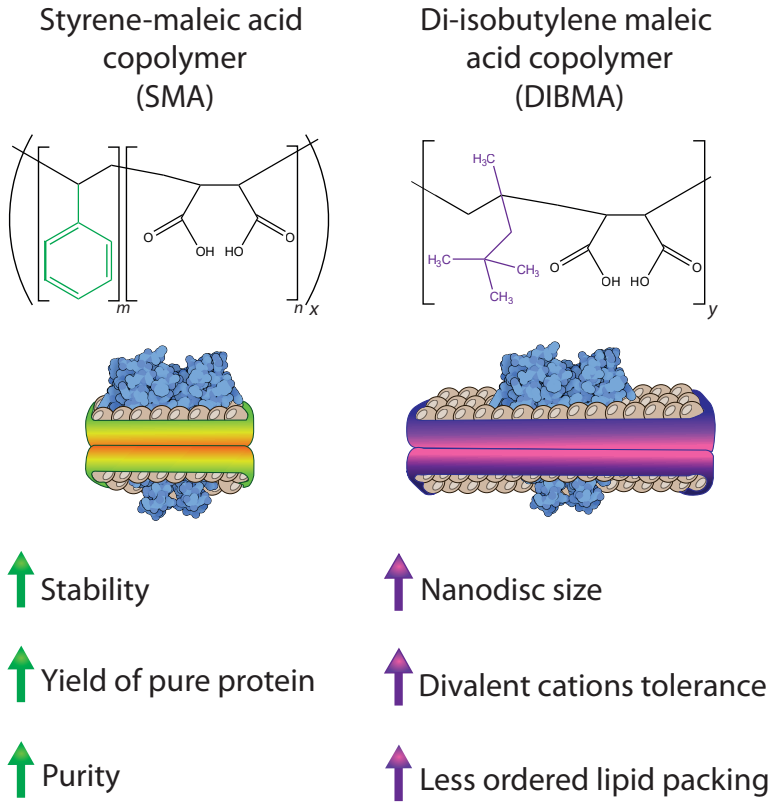


Figure 2.11: The structure and properties of SMA and DIBMA copolymers. The chemical structures of the styrene-maleic acid (SMA) and di-isobutylene maleic acid (DIBMA) copolymers, with the hydrophobic sections highlighted in green (SMA) or purple (DIBMA). m and n indicate the ratio of the hydrophobic and hydrophilic SMA sections, whilst x and y represent the number of repeats in the SMA or DIBMA copolymer, respectively. Illustrations of their respective lipid particles are below, followed by a brief outline of their individual advantages. Adapted from *Biochimica et Biophysica Acta (BBA)- Biomembranes*, **1862**, A. A. Gulamhussein *et al.*, A Comparison of SMA (Styrene Maleic Acid) and DIBMA (Di-isobutylene Maleic Acid) for Membrane Protein Purification, 183281, ©2020, with permission from Elsevier.

control over the NN size by tuning the copolymer unit size.[154] The zSMA was shown to extract different IMPs at a similar or better efficiency than SMA.[155] Nevertheless, the highly-charged nature of most of these polymers needs to be considered when used for an IMP containing a charged extracellular domain, as the charged polymer groups have been shown to interact with them, even causing IMP inactivation.[156] Alternatively, grafting cysteamine onto the maleic acid groups of the SMA copolymer enables easy NN backbone functionalisation with *e.g.* fluorophores, fluorescent dyes or NN surface immobilisation.[157] However, whilst useful for specific

applications, these SMA-SH NNs retain the same problems as SMALPs, limiting their general applicability. Recently, a biologically inspired variation on the DIBMA backbone, termed Glyco-DIBMA, has been reported. This glycopolymer appears to outperform established amphipathic copolymers in vesicle solubilisation efficiency, displaying increased hydrophobicity and reduced charge density whilst maintaining high solubility in aqueous solutions, overcoming many of the aforementioned issues associated with high charge density.[158]

Despite these downsides, SMALPs remain the most used type of NNs. They have been used for IMP extraction from a wide variety of membranes, which they mostly solubilise highly efficiently. However, highly ordered and protein-dense membranes, such as the photosynthetic membrane, can present an obstacle.[159] As the SMA copolymer solubilises the membranes by intercalating its styrene group with the lipid tails,[147,160] it is likely that the high protein density and lower lipid content could interfere with this process, whilst the large complex size could prevent SMA encircling.[159]

Native Nanodiscs in Membrane Protein Research

Thanks to their enhanced stability, monodispersity, and ease of concentration, NNs are in theory well-adapted for many structural techniques. Sub-nm resolution of IMPs in NNs has only been achieved recently using cryo-EM, as demonstrated by the structures of AcrB and alternative complex III, in a supercomplex with cytochrome oxidase.[161,162] Like LNs and APols, NNs cannot be crystallised directly due to the repulsion between the outer polymers, but the IMPs can be transferred directly into the lipidic cubic phase, resulting in high-resolution structures.[163] Solution NMR is also likely incompatible with NNs, as their relatively large size would limit the tumbling speed of the IMPs.[145] Nevertheless, solid-state NMR has recently shown potential for high-resolution IMP structures in NNs, exemplified by the zinc

diffusion facilitator CzcD.[164]

NNs can facilitate IMP functional studies, usually with no or minor modifications required.[145] A radioligand binding assay of the GPCR $A_{2A}R$ demonstrated comparable activity in SMALPs to native membranes.[142] Using a fluorescence-based assay, light-binding studies of the ghrelin and melatonin receptors also showed normal activity and downstream effector recruitment in SMALPs compared to their native environments,[141] confirming the nativeness of NNs. IMP function in NNs has also been observed using FRET and other fluorescence-based assays.[145]

The identity of the co-purified native lipids is also of great interest, providing crucial information regarding their function in protein structure and function.[145] They can be analysed using a variety of techniques, including gas chromatography, thin layer chromatography (TLC), LC-MS, ESI-MS, or phosphorous assays, offering varying levels of detail regarding the amount, head group, and tail length of the lipids. The estimated number of lipids per IMP in NNs varies between proteins, from ∼11 for a single β-barrel protein PagP,[139] to ∼40 for the trimeric, 36 TM helices protein AcrB,[165] although one study reported up to ∼150 lipids per bacterial IMP complex.[166] This variation could be a result of the exact lipid:protein:SMA ratio and the experimental protocol used.

Analysing the lipid identity often shows selective enrichment of specific lipids.[145] KcsA and the SecYEG complex both co-purified with anionic lipids at higher proportions compared to the bulk membrane.[145,149] Detailed analysis by MS can reveal differences in the lipid tail enrichment, with some proteins preferring specific long-chain unsaturated or mono-unsaturated lipids, excluding different polyunsaturated PLs altogether.[145] Nevertheless, NNs appear to be highly dynamic, with rapid lipid transfer and exchange between the individual SMALPs shown by time-resolved

FRET combined with small-angle neutron scattering (SANS).[167]

NNs present an exciting novel MMS, combining experimental simplicity of protein extraction and near-native lipid environment with applicability to many structural and functional techniques. NNs could potentially be applied to large-scale membrane proteome analysis and lipidomics to determine the key protein-lipid interactions, although the results could be skewed by the low solubility of highly protein-dense membrane regions.[159] NNs also display fast dynamics of lipid exchange compared to LNs or vesicles, which occurs not only by monomer diffusion but also through fast collisional transfer, suggesting the flexible nature of the SMA belt around the lipid core facilitates a 'hydrocarbon continuum', enabling this rapid lipid exchange.[167] These properties have profound implications for NN research interpretation, suggesting the lipidome of IMP-containing SMALPs is not simply a snapshot of the IMP's immediate lipid surroundings at the time of extractions, but is also affected by the specific lipid-protein interactions and exchange. Further research and characterisation of different amphipathic polymers would also be beneficial, identifying their best applications.

The recent development of cycloalkane-modified amphipols (cyclAPols) combines the desirable properties of APols with the direct membrane-solubilising ability of NNs. The linear aliphatic side groups of the polyacrylic acid APol backbone were replaced with aliphatic cyclic groups, providing the ability to solubilise membranes, supporting the hypothesis that hydrophobic cyclic side groups are crucial for membrane solubilisation.[168] Their extraction properties were demonstrated with bacterial IMPs, including the solubilisation-resistant purple membrane of *Halobacterium salinarum*.[168] They were also able to extract and solubilise AcrB for cryoEM analysis, achieving a high-resolution structure comparable to those in other MMS. These initial studies also suggested cyclAPols might be applicable

to *e.g.* native MS, identifying co-purified annular lipids.[169] However, the large cyclic group might perturb the lipid conformation, similarly to the SMA styrene group. More research into these different polymers will improve our understanding of the limitations of this novel, exciting MMS.

2.3 Viral Glycoproteins

2.3.1 HIV Env Glycoprotein

The human immunodeficiency virus (HIV) has been discovered more than 40 years ago. If untreated, the infection develops into acquired immunodeficiency syndrome (AIDS), which has since caused over 32 million deaths. It is estimated that over 38 million people are infected with this virus at this moment, in 2022.[170] Despite decreasing mortality thanks to the development of antiretroviral treatments, the development of improved therapies and prevention methods remains crucial. Unfortunately, despite decades-long vaccine development efforts, an effective vaccine against HIV remains elusive.

One of the major challenges is the high frequency of mutations of the viral genome, leading to a high diversity of potential vaccine targets, as well as alterations in the fundamentals of the viral biological processes.[171] As such, this high mutation rate enables the virus to evade host immune system as well as novel antiviral drugs. An important factor in the immune evasion is the viral fusion protein, HIV-1 envelope spike glycoprotein (Env). It is presented on the surface of the virion in high density, providing it with highly heterogeneous and extensive glycosylation of the virion's surface.[172] Also known as the glycan shield, this type of glycosylation is thought to be one of the main obstacles to an efficient immune response.[173] Since the glycans are produced within host cells through their native glycosylation pathways, the immune system considers the glycosylated virus as immunologically

non-threatening and even native.[172] The glycan shield is also constantly changing as a result of the high mutation rate of the virus, making the development of universal drugs targetting the virus, or an efficient immune response by the host challenging.[171]

Nevertheless, it has been found that some areas of this shield remain relatively more constant, which enabled some infected individuals to develop broadly neutralising antibodies, showing neutralisation efficacy against different HIV strains. Even though they cannot eliminate the virus completely, they form a good starting point for the development of a vaccine by enabling immune system priming.[174] Another potential path to vaccine development would be immunogens derived from Env that could result in an immune response, inducing the expression of broadly neutralising antibodies,[175] although so far this has not been achieved. It has been assumed that Env exists as a mixture of different conformation states, with only specific ones being able to initiate a positive immune response. Engineering suitable mutants would help with this avenue of HIV treatment and prevention. Ideally, this construct would be an artificially stabilised, immunogenically relevant state capable of raising an immune response.

Env exists as a non-covalent trimer of heterodimers containing gp120 and gp41. When present on virions in an unbound conformation, it is assumed to exist in a specific, pre-triggered conformation, often called state 1.[176] When bound to the host cell through the CD4 receptor, its target receptor, it transiently enters an intermediate state 2 before transitioning to an open conformation, state 3.[177] Out of these, it is expected that state 1 is the most relevant for eliciting immune response. Recently, a stabilised Env trimer was engineered, resembling the conformation in state 1.[178] Termed BG505 SOSIP.664 gp140, it interacted with high affinity with all broadly neutralising antibodies, targetted against a variety of HIV-1 Env epitopes, suggesting its potential use as a vaccine as well as its applicability as a model

for structural and biophysical characterisations, or interaction studies with other, potentially neutralising, small molecules or inhibitory proteins.

2.3.2 SARS-COV-2 Spike Glycoprotein

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) belongs to the betacoronavirus lineage,[179] and was first identified as the causative agent of a cluster of pneumonia cases in Wuhan, Hubei Province, China, in December 2019.[180,181] Since then, the virus has spread worldwide, causing almost 250 million confirmed cases and resulting in over 5 million deaths as of November 2021.[182] It is of likely zoonotic origin, with 96% genome sequence identity with bat coronavirus BatCov RaTG13,[180] with pangolins identified as a possible intermediate link,[183] with one study identifying up to 92.4% sequence similarity between betacoronavirus isolated from pangolins and SARS-CoV-2.[184] It is also the third case of a coronavirus infecting humans this century alone, following the Severe Acute Respiratory Syndrome (SARS) and Middle-East Respiratory Syndrome (MERS) coronaviruses outbreaks in the past 20 years.[185]

SARS-CoV-2 is a single-stranded, positive-sense ribonucleic acid virus that's enveloped. Its genome spans 29.9 kilobase pairs, and encodes four structural proteins - small envelope (E), membrane (M), nucleocapsid (N) and spike (S) glycoproteins, which are used for viral particle assembly (E), host receptor recognition (S), virion shape (M) and encapsulating RNA genome (N) (Figure 2.12). Additionally, it also encodes for 16 non-structural proteins (NSP1-16) with a variety of functions, with a number of accessory proteins identified involved in RNA replication, processing, and survival of the virus in the host cell.[187]

The spike glycoprotein mediates viral entry into host cells for coronaviruses.[188,189] It is a homotrimeric transmembrane protein, with a single TM helix anchoring it

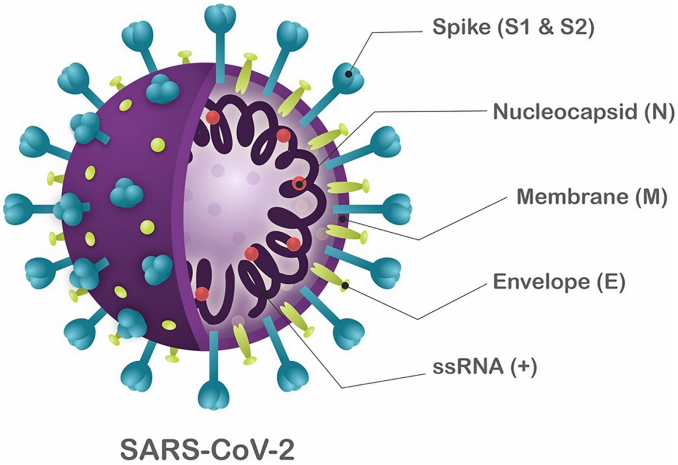


Figure 2.12: Simplified structure of the SARS-CoV-2 virion. The structure of the virus is mainly determined by four structural proteins - spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins. The viral envelope is a lipid bilayer, and contains the S, E, and M proteins. The viral positive sense single stranded RNA (ssRNA (+) is positioned inside the virion, and is encapsulated by the N protein. The S protein is composed of two subunits, S1 and S2. Figure reproduced with permission from I. A. Santos *et al.*, Antivirals Against Coronaviruses: Candidate Drugs for SARS-CoV-2 Treatment?, *Frontiers in Microbiology*, 2020, **11**, 1818.[186] Licenced under the Creative Commons Attribution License (CC BY).

to the viral envelope. It is composed of two subunits, S1 and S2, which perform separate functions, host cell receptor recognition and binding and host cell membrane fusion, respectively, being a type 1 fusion machine.[190] The S1 subunit comprises the N-terminal domain (NTD) and receptor binding domain (RBD), which exists in two conformations - the 'up' and 'down' positions, which can transition in a hinge-like movement. Upon moving to the 'up' conformation, the RBD strongly binds its host cell receptor, angiotensin converting enzyme 2 (ACE2); however, this position is thought of as less stable.[191] The proportion of RBDs in the 'up' position is still a matter of discussion. A number of techniques have been used to determine this ratio, including cryo-electron microscopy (cryoEM) and single-molecule fluorescence/Förster resonance energy transfer (smFRET), reporting results varying from 50% of spike particles in 1 RBD 'up' positions only (with the rest being all 'down' positions) in the absence of ACE2, as determined by cryoEM[192] to 22% of all particles displaying all RBD 'up' positions in absence of ACE2 and

48% all 'up' in the presence of dimeric ACE2, as determined by smFRET.[193] It is worth noting that RBD position appears to be very dynamic, and as such difficult to capture. It also seems to be strongly dependent on other factors, such as buffer pH and salt composition, with one study reporting 40% of all spikes in the 'up' position in presence of ACE2 at cytosolic pH (7.4), dropping to 32% at endosomal pH (5.5).[194]

The S2 is composed of the fusion peptide, heptad repeat 1 (HR1), central helix (CH), connector domain (CD), heptad repeat 2 (HR2), TM domain, and cytoplasmic domain. The area between the S1/S2 subunits contains a protease cleavage site, where host protease cleaves off the S1 subunit, leading to S2 activation and elongation, priming spike for fusion with the host cell membrane through newly exposed hydrophobic region at the top of S2.[190] Besides mediating viral membrane fusion with the host cellular plasma membrane (PM), the S2 subunit is also responsible for spike flexibility through three hinges present in the stalk region of the spike glycoprotein, giving the S1 subunit an unexpected degree of rotational freedom, as found by combination of cryo-electron tomography and MD simulations.[195] This enables multiple spike proteins on the curved surface of the virion to reach higher number of ACE2 receptors on the comparatively flat host cell surface. The entire spike protein is also extensively glycosylated, containing 22 glycosylation sites per protomer.[192] These are critical for correct protein folding and viral shielding from the immune system, but also facilitate its interaction with ACE2, as shown by molecular dynamics (MD) simulations and directed glycan site mutagenesis experiments.[196]

SARS-CoV-2 viral infection in a cell has two effects on host cell protein expression. Firstly, the virus recruits host proteins to destroy all host messenger RNA (mRNA), sparing only mRNA with a viral tag.[197] In addition, NSP16 disrupts global mRNA splicing by binding pre-mRNA recognition domains U1 and U2 of the small nucleolar

RNA.[198] (Figure 2.13, top) Secondly, it reduces overall protein translation by 70%. This is because NSP1 physically blocks the entrance of ribosomes by binding to the 18S ribosomal subunit of the ribosome entry channel, preventing the entry of host mRNA (Figure 2.13, middle). Most remaining translational capacity is dedicated to viral RNA, which proceeds efficiently, as the viral genomic mRNA 5' untranslated region stimulates translation.[199,200] In addition, NSP1 also blocks nuclear exit pores, preventing newly transcribed immune response mRNA, which would normally alert the immune system to new infection, from ever being translated and activating the immune system (Figure 2.13, bottom).[197] To the same effect, NSP8 and NSP9 disrupt protein trafficking by binding to 7SL RNA in the signal recognition particle, suppressing the natural cell interferon response to viral infection.[198] Such form of immune evasion is common amongst viruses, but SARS-CoV-2 is particularly efficient - and some of the new variants, such as the alpha variant, are becoming ever more effective in this form of immune evasion.

This level of efficiency of immune evasion is particularly dangerous, as when the immune system eventually gets alerted to the virus, it is often already abundant in tissues, causing out-of-proportion immune response, such as cytokine storm and hyperinflammation, which can lead to multiple organ damage and death. Some approved treatments are based around dampening the immune response and reduce inflammation.[201]

Angiotensin-Converting Enzyme 2

ACE2 is a homologue of the ACE receptor, identified in 2000.[202,203] ACE2 gene is positioned on the X chromosome and is composed of 20 introns and 18 exons, that can produce 6 variants thanks to alternative splicing.[204] Once expressed, it is 805 amino acids (AAs) long with a C-terminal domain membrane anchor and a

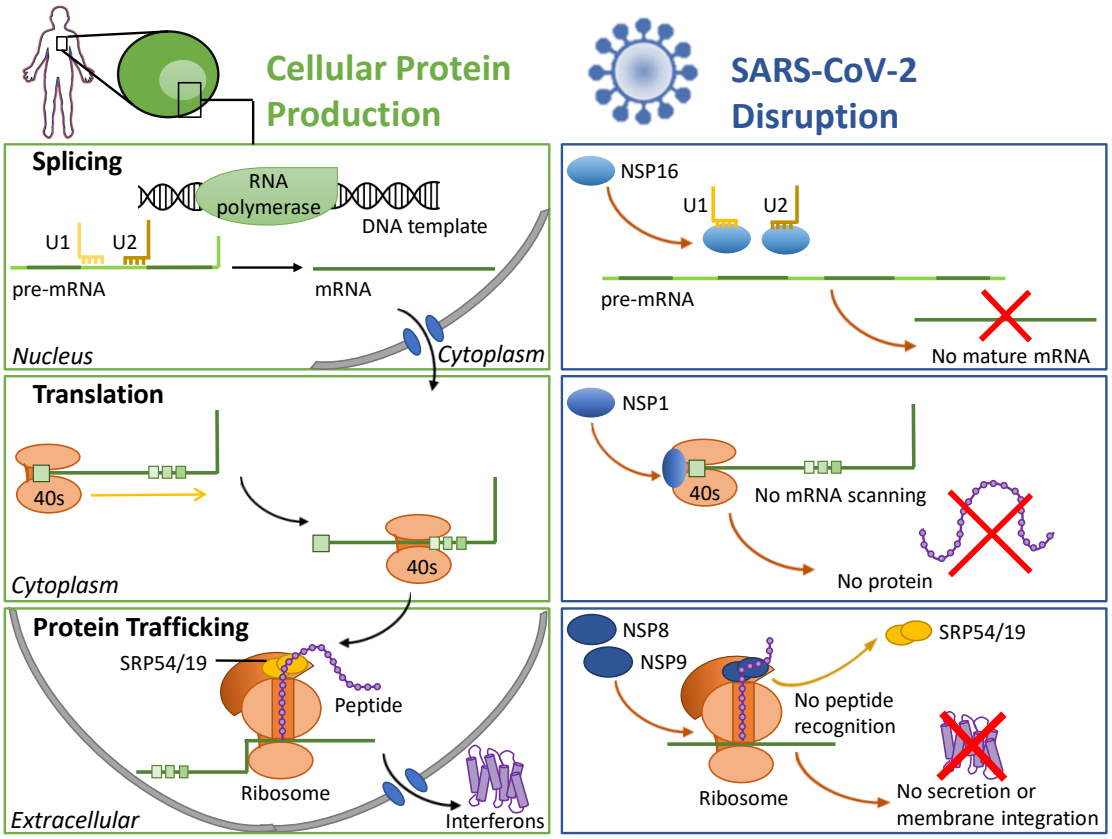


Figure 2.13: Effects of SARS-CoV-2 infection on transcription and translation. Top, simplified outline of DNA transcription and resulting pre-mRNA splicing, and its disruption by NSP16 binding to U1 and U2 of the small nucleolar RNA (snRNA) facilitating the splicing. Middle, translation initiation schematic and its disruption by NSP1 binding to the mRNA entry channel, preventing mRNA scanning and protein synthesis. Bottom, the signal recognition particle (SRP) recognises signals targeting proteins for the endoplasmic reticulum (ER), and hence for membrane insertion or secretion. NSP8 and NSP9 bind to 7SL RNA, preventing signal recognition and interfering with protein secretion and membrane integration, effectively disrupting the interferon (IFN) secretory pathway for immune system activation. Figure adapted with permission from A. K. Banerjee *et al.*, SARS-CoV-2 Disrupts Splicing, Translation, and Protein Trafficking to Suppress Host Defenses, *Cell*, 2020, **183**, 1325-1339.[198] Licenced under the Creative Commons Attribution License (CC BY) by Elsevier, ©2020.

single extracellular N-terminal domain, which consists of the active catalytic site domain, a dimerisation neck domain as well as a SARS-CoV-2 and SARS-CoV binding sequence. ACE2 functions as a carboxypeptidase, meaning it removes single AAs from the C-terminal ends of its substrate molecules.[205]

Full length, wild-type (WT) ACE2 contains a dimerisation 'neck' domain located between the TM helix and the carboxypeptidase domain (CPD).[206] Whilst CPD plays a crucial role in the vasoprotective axis of the renin-angiotensin system through catalysing the hydrolysis of Angiotensin I (Ang-I) to Ang-(1-9) and Ang-II to Ang-(1-7),[202,207] it also contains the binding motif for the receptor-binding domain (RBD) of spike, facilitating viral cell entry and infection.[191,192,206] As part of its function, the extracellular portion of ACE2 is cleaved by the membrane-bound disintegrin and metallopeptidase domain 17 protein (ADAM17) and transmembrane serine proteases 2 (TMPRSS2),[208,209] resulting in 'shedding' of the soluble ACE2 ectodomain (sACE2). The neck domain residues 697–716 have been shown to be cut during the cleavage,[208] suggesting reduced dimer formation of sACE2.

Due to its role in SARS-CoV-2 infection and COVID-19 disease, ACE2 has been at the forefront of scientific attention. In order to identify organs most susceptible to the infection and to explain the broad spectrum of symptoms affecting people suffering from COVID-19, significant research has been conducted on ACE2 expression across different tissues, both at the level of protein and mRNA expression.[210] Counterintuitively, it appears both from human transcriptome and experimental models that ACE2 expression is actually rather low in the lung, and appears to be limited to type II alveolar epithelial cells.[211,212] Nevertheless, it has been proposed that the release of inflammatory cytokines in response to SARS-CoV-2 presence, such as interferons (IFNs), can potentiate the infection by elevating ACE2 expression.[213,214] Contrary to this research, it has been shown recently that

stimulation by IFNs due to viral presence leads to expression of a truncated ACE2 form (referred to as δACE2), missing 365 NTD AAs, meaning it is incapable of binding SARS-CoV-2 spike glycoprotein, and hence cannot contribute to infection enhancement in COVID-19.[215]

Apart from the respiratory tract, other organs that seem most affected by COVID-19 appear to be the gastrointestinal tract and kidneys, with the disease causing frequent dysfunction and even failure of these organ systems upon SARS-CoV-2 infection.[216,217] Additionally, study of ACE2 protein and mRNA expression in different experimental models, as well as analysis of three databases have shown that most significant levels of ACE2 expression is found in the testis, heart muscle and thyroid gland.[211]

ACE2 is generally embedded in the PM of cells, but can be solubilised by proteolytic cleavage by a number of proteases, a processed termed "shedding" (Figure 2.14). Transmembrane protease, serine 2 (TMPRSS2), disintegrin and metalloproteinase domain-containing protein (ADAM)10 and ADAM17 are examples of proteases that can cause ACE2 shedding into the interstitium.[210] Research has shown that ACE2 binding to the receptor-binding domain (RBD) of the SARS-CoV-2 spike glycoprotein initiates such cleavage by ADAM17/tumour necrosis factor-converting enzyme (TACE),[208] producing a solubilised, catalytically active form.[209] It appears that inhibiting ADAM17/TACE can result in a decrease of viral entry, both *in vitro* and *in vivo*, suggesting their importance for SARS-CoV infectivity and potential use in SARS-CoV-2 treatment.[219]

Nevertheless, the role of ACE2 cleavage and shedding remains unclear, although understanding the mechanism behind ACE2 shedding, function and plasma levels of solubilised ACE2 (sACE2) can lead to discovery of better treatments and improved diagnosis, both of which are of high importance during the COVID-19 pandemic.

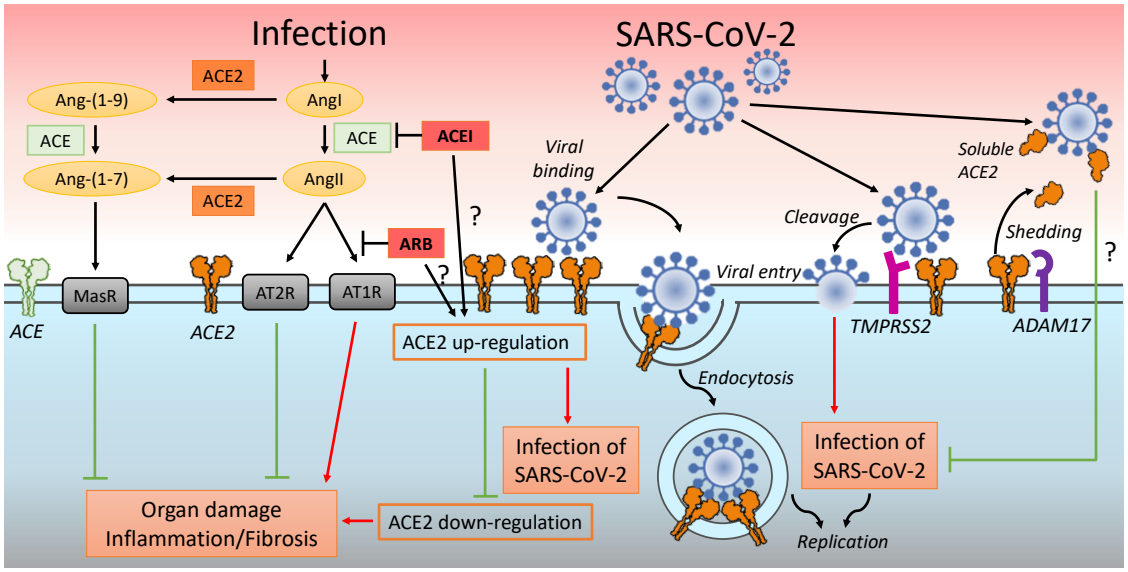


Figure 2.14: Overview of the renin-angiotensin system and its possible associations with SARS-CoV-2 infection and COVID-19 associated lung injury. Angiotensin-converting enzyme (ACE) converts Angiotensin (Ang) I to Ang II. Ang II signalling can lead to various organ damage, such as lung injury, through inflammation promotion and fibrosis *via* the Ang II type 1 receptor (AT1R), whilst the AngII type 2 receptor (AT2R) has opposing effects. ACE2 catalyses the conversion of Ang II to Ang-(1–7) and Ang I to Ang-(1–9). Ang-(1–7) signalling downsizes the effects of fibrosis and inflammation through the Mas receptor (Mas-R). ACE2 upregulation by some ACE inhibitors (ACEIs) and AT1R blockers (ARBs) has been suggested previously. ACE2 also acts as the entry point for SARS-CoV-2. ACE2-binding spike glycoprotein on the virion gets internalized by endocytosis, downregulating membrane-bound ACE2 on the cell surface. This can encourage organ damage through ACE/Ang II/AT1R pathway activation and ACE2/Ang-(1–7)/Mas-R pathway deactivation. Another proposed mechanism of SARS-CoV-2 entry is through the transmembrane protease, serine 2 (TMPRSS2)-mediated cleavage of the spike, occurring together with ACE2 release. ACE2 can also be released from the membrane by ADAM17 forming soluble sACE2. sACE2 may in turn prevent the virus from binding to membrane-anchored ACE2. Figure adapted with authors' permissions from Furuhashi *et al.,* Potential differential effects of renin-angiotensin system inhibitors on SARS-CoV-2 infection and lung injury in COVID-19, 2020, **43**, 1-4.[218]

As demonstrated above, lessons can be learned from research done previously on SARS-CoV, since the mechanisms of cell entry appear related. To illustrate this, it has been shown that SARS-CoV spike binding to the ACE2 receptor leads to ADAM17-dependent ACE2 cleavage and shedding, which promotes viral entry and uptake into cells.[220] TMPRSS2, on the other hand, uses different targets in the ACE2 protein sequence to induce proteolytic cleavage, acting on the intracellular cytoplasmic domain.[221] Its cleavage also does not result in a soluble, catalytically

active isoform of ACE2.[208] Despite these differences, it appears to be necessary for SARS-CoV and SARS-CoV-2 cellular entry.[221,222]

sACE2 levels can also be elevated by certain proinflammatory cytokines, such as tumour necrosis factor-α, or interleukin-1β.[223] Whilst these elevated sACE2 levels could, in theory, result in an infection improvement due to extra sACE2 acting as a viral spike decoy, binding to virion surface before it manages to latch onto cell membrane, the associated increase in Angiotensin II (AngII) levels and resulting activation of the Ang II/AT1R axis can lead to stronger inflammatory response.[210] Indeed, presence of elevated levels of sACE2 from excessive ACE2 shedding is correlated with worse disease progression and prognosis, potentially as a result of elevated levels of AngII.[224,225]

In addition to sACE2 levels being indicative of disease prognosis, it has also been postulated that different levels of sACE2 serum levels, ACE2 expression, and ACE2 polymorphisms could have an effect on the range of phenotypes of COVID-19 disease within the population. Since specific residues on ACE2 receptor are utilised by the SARS-CoV-2 virus for recognition by the RBD, more specifically the ones on the carboxypeptidase domain (CPD) of the ACE2, by a combination of polar residue interactions, hydrogen bonds (H-bonds), and van der Waals forces,[206] it follows that natural genetic variations within this region could result in some degree of populational variation of symptoms caused by COVID-19. Indeed, specific variances have been discovered amongst different populations,[226–228] although the specific effects on SARS-CoV-2 spike binding remain yet to be quantified experimentally. Similarly, natural variations in sACE2 levels have been proposed to be in part responsible for variable reaction to SARS-CoV-2 infections, although the evidence appears to be pointing both ways. On one hand, higher circulating levels of sACE2 could offer protection by acting as decoy receptors,[229] as discussed above, and could

even explain lower susceptibility of women and children to the infection.[230] On the other hand, elevated levels of sACE2 could point to higher levels of either ACE2 expression or ADAM17 activity (or both), which could increase susceptibility to the infection.[231,232]

2.4 Mass Photometry

Mass photometry (MP) is based on interferometric scattering microscopy (iSCAT), and enables optical detection, differentiation and counting of single proteins and other biomolecules based on their interferometric scattering contrast, proportional to their molecular weight.[5] It enables sample purity determination, detection of the component molecular weight distributions, but also investigations into biomolecular interactions and their dynamics.

During a mass photometry measurement, the sample diluted in an aqueous buffer is placed on a flat, solid, transparent material, usually a glass coverslip, which is positioned on top of a mass photometer. Resembling an inverted microscope, the coverslip is put on top of an oil immersion lens objective, which focuses light directly at the glass-water interface. Whilst most of the incident light (E_I) is transmitted, a small portion gets reflected (E_R) and even less of it gets scattered (E_S) at the glass surface by the investigated molecules (Figure 2.15). These reflected and scattered portions of the light are then collected by an objective with a high numerical aperture, creating the signal which we process later.

The incident light used to illuminate the sample is generated by a diode laser of a specific wavelength, *e.g.* 445 nm. Given that the interferometric contrast is dependent on the wavelength through $1/\lambda^2$, it becomes apparent that smaller laser wavelengths are preferred. In practice, a variety of laser wavelengths can

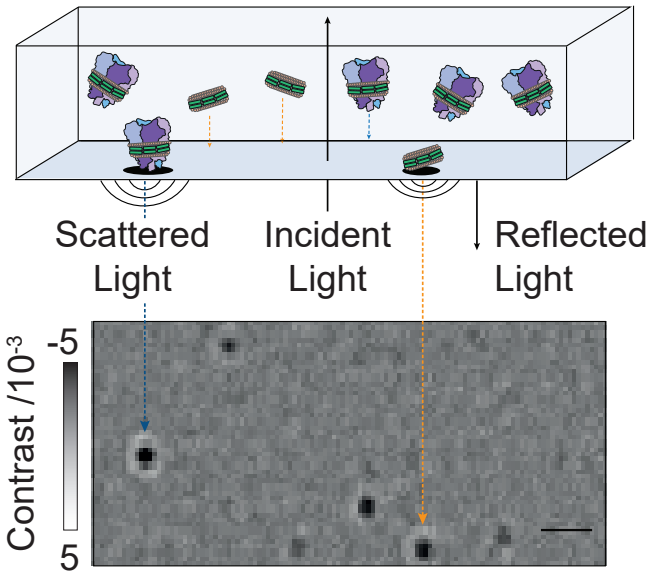


Figure 2.15: The basis of signal detection in mass photometry. The incident light (E_I) is focused by a microscope objective on the sample, collecting the reflected light (E_R) from the glass coverslip as well as the light scattered (E_S) by present particles, in this case illustrated by empty and full lipid nanodiscs. Figure adapted with permission from A. Olerinyova and A. Sonn-Segev *et al.*, Mass Photometry of Membrane Proteins, *chem*, 2021, **7**, 224-236.[16] Licenced under the Creative Commons CC-BY-NC-ND license.

be used, although the choice is often limited by the availability of lasers with specific wavelength and sufficient power (*i.e.* around 2 Watts. As such, the laser wavelengths of the instruments used within this work were 445 nm (blue), used for a subset of experiments within Chapter 3, and 520 nm (green), used for the majority of experiments presented in this work. Whilst the blue set-up provided the highest contrast resolution and the lowest detection limit, it was at the expense of higher protein unbinding from the glass coverslip surface and the risk of burning of the objective lens and other optics components in the set-up due to the high energy associated with the lower wavelength. Other wavelengths have also been used, such as 660 nm (red), with the specific application of combining iSCAT with single-molecule fluorescence imaging.[233] This wavelength carries less energy, and as such has worse contrast resolution and higher detection limit; nevertheless, it was chosen for this application due to the undesirable cross-excitation of the most frequently green fluorescently-tagged proteins by the blue and green laser

wavelengths.

After collimation, the beam gets deflected by two acousto-optic deflectors (AODs), oriented orthogonally, also allowing us to scan the field of view later. Following AODs, the beam passes through a polarising beam splitter (PBSp), followed by a quarter wave plate (QWP), allowing efficient separation of the reflected light from the incident light. The oil immersion lens objective is then used to focus the beam on the glass-water interface, when the AODs generate beam deflection and rapid scanning of the beam (at ∼100 kHz, which is 100 × faster than the camera frame rate), allowing for an evenly illuminated field of view (FOV). Rayleigh scattering can be used to describe the interaction between the nanoscopic objects being imaged and the incident light, and is the basic principle underlying iSCAT and MP. It postulates that the particles of size much below the wavelength of light being used can be excited as a result of their electric polarisability, which turns them into radiating dipoles that emit light in essentially all directions. In the reflection mode of iSCAT, the back-scattered light is collected by the objective, and arises from both the glass surface and analyte molecules, combined with the reflected light from the water-glass interface.

The light is then pass through the PBSp in the back-focal objective plane, separating incident and collected light. We can then write the detected intensity, I_{det} as:

$$I_{det} = I_{inc}[r^2 + |s|^2 + 2r|s|cos(\phi)] \tag{2.1}$$

where I_{inc} refers to the incident light intensity, r^2 to the glass-water interface reflectivity, s to the scattering amplitude of the analyte object, and ϕ refers to the phase difference between the scattered and the reflected fields at the detector.

Given a small scatterer, s, its scattering amplitude can be written according to the Mie theory as:

$$s = \eta\epsilon_m\pi\frac{D^3}{2}\frac{\epsilon_p - \epsilon_m}{\epsilon_p + 2\epsilon_m} \tag{2.2}$$

where $\epsilon_{p/m}$ are the dielectric functions of the particle and surrounding medium, respectively, D is used to refer to the diameter of the spherical particle, and η refers to a constant dependent on the detection efficiency of the setup. Given this relationship, we can see that the scattering signal is dependent and scales with the polarizability of s, hence a function of the refractive index and proportional to the volume of the particle. It also becomes apparent that the pure scattering term $|s|^2$ in Equation 2.1 decreases with the size of the object. We can also see that $|s|^2$ tends to become negligible with small scatterers, when the scatterers' dielectric constant is similar to the surrounding medium, and also $r >> |s|$. If we distinguish between the light signal in the presence and absence of the scatterer, I_{det} and I_{bgd}, respectively, we can define scattering contrast c directly as follows:

$$c = \frac{I_{det}}{I_{bgd}} = 1 + \frac{2|s|cos\phi}{r} \tag{2.3}$$

A variety of other modifications and approaches are often employed to increase the signal of a small scatterer, *e.g.* by placing a partial reflector in the beam path centre, as described by Cole *et al.*[4] In order to visualise biomolecules using this approach, background removal is necessary, which is performed during the image analysis. Using this approach, biomolecules become visible as black or white dots, with their contrast, corresponding to their intensity.[5] This is performed sequentially throughout the movie length, resulting in apparent particle landing. While the particle does land on the glass coverslip, this happens at a much faster rate than the camera's frame rate, and the appearance of particle contrast getting stronger is a direct consequence of the background removal process.

Since the output information of a mass photometer is the interferometric scattering contrast of each particle landing event, a conversion is necessary to determine the MW distribution of an unknown sample. This is performed by measuring a set of biomolecules with a known, stable MW distribution, referred to as the mass calibration standard. The interferometric scattering contrast corresponding to the MW for each component of the standard is plotted, showing interferometric scattering contrast as a function of MW, and fitted with a linear mass calibration function. This has the form of $y = b \cdot x + c$, where y = interferometric scattering contrast, x = molecular weight, c = the y-axis intercept and b = the instrument calibration slope. Both c and b define the instrument calibration function, with substantial variations observed between different instruments, and slight variations occurring when the same instrument is used across a number of days. Given the generally low values of the y-axis intercept c, often in the range of 10^{-5}, it is sometimes approximated to 0 for easier fitting and mass calibration simplification.

The specific mass calibration standards are generally selected based on a few parameters. Firstly, the type of biomolecule investigated (*e.g.* DNA, protein, complex glycan) will affect the mass calibration due to differences in the refractive index of the biomolecules measured. As such, a mass standard of matching biophysical properties or biological composition should be selected. Next, the range of MW covered by the standards needs to be considered, as the expected MW of the sample should fall inside this range. Even though the contrast to mass calibration curve appears linear for the mass ranges investigated in this work, this does not necessarily apply at the lower limit of detection.[5] Similarly, when measuring very large biomolecules, the mass standard should cover the majority of the MW range investigated, as small variations in the low MW range of a mass standard could have a profound effects on the measured MW of a sample with a very high MW (> MDa). Lastly, any types of protein modifications present (such

as glycosylation) should be considered and ideally reflected in the mass standard, as they may alter the overall refractive index of the sample and hence the measured contrast. Depending on the degree of instrument focus drift and length of the measurement, it might be necessary to repeat the mass calibration measurement multiple times during the day to ensure accurate MW calibration.

Looking at Equations 2.1 and 2.2 in more detail, they detail the quantification of the ratiometric contrast of different biomolecules at the single molecule level. It becomes apparent that for objects of much smaller sizes when compared to the wavelength of light that are characterised by relatively insignificant differences in their densities and refractive indices, their scattering signals ought to be proportional to their molecular mass. This was indeed shown for MP, then called Interferometric Scattering Mass Spectrometry (iSCAMS), for a range of biomolecules including soluble proteins, lipid nanodiscs, and highly glycosylated proteins (Figure 2.16).[5] This finding corresponds well with the calculation of specific volumes and refractive indices of different amino acids, performed by Young and Hundt *et al.* (2018),[5] which were found to differ by approximately 1% (Figure 2.17.

Despite this apparent agreement, many open questions remain in relation to MP and iSCAT. Various dynamic ranges, such as the concentration limits, kinetics, and stoichiometry measurements remain largely undefined for MP. For instance, Soltermann *et al.* (2020)[7] has demonstrated the applicability of MP to the study of protein-protein interactions, their dissociation constants, and kinetics of interactions. This works builds on this, expanding on it towards complex heterogeneous interactions of membrane proteins and viral glycoproteins with a protein ligand or native host binding partner. Additionally, Soltermann *et al.* also introduced a method for expanding the concentration range suitable for MP experiments to pM region, using a sophisticated continuous flow pump system.[7,234]

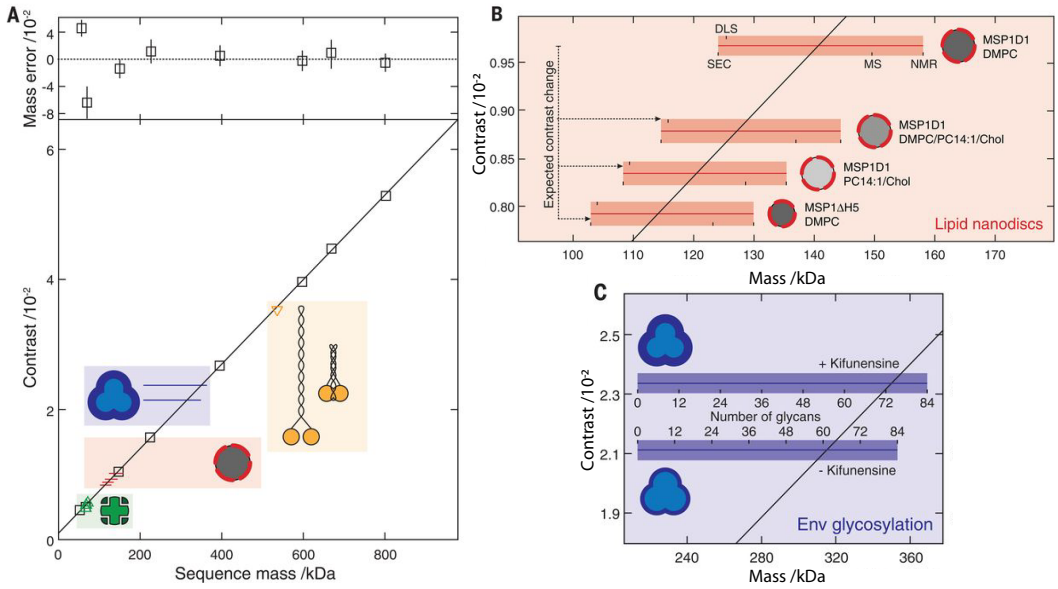


Figure 2.16: Mass photometry (Interferometric Scattering Mass Spectrometry (iSCAMS)) precision and accuracy characterisation. A. Interferometric scattering contrast plotted against the biomolecular molecular mass, showing soluble proteins used for mass calibration (black), analysis of dependence on protein shape (yellow), binding of a ligand to a protein (green), the composition of lipid nanodiscs (red), and the effects of glycosylation (blue). The mass calibration data is fitted with the line of best fit. **B.** Mass measurements of lipid nanodiscs with different lipid compositions encircled by different MSPs. The MSP1D1/DMPC lipid nanodisc masses found *via* alternative methods are marked on the bar and extrapolated to nanodiscs with different compositions, with the horizontal lines showing the expected mass range based on the characterisation technique. The thin bars show the measured contrast, and the thick bars represent the uncertainty of the measurement (standard error of the mean - SEM) for all repeated experiments. The upper text of each sample denotes the MSP present, and the lower text identifies the lipids present in the nanodisc. DLS - dynamic light scattering; SEC - size-exclusion chromatography; MS - mass spectrometry; NMR - nuclear magnetic resonance. **C.** The differential contrast of Env, as measured by MP for Env that was expressed in the absence or presence of kifunensine and the mass ranges associated, as expected for the different levels of glycosylation. In **B.** and **C.**, the line shown indicates the calibration obtained in **A.** Figure adapted with permission from the American Association for the Advancement of Science, from G. Young and N. Hundt *et al.*, Quantitative mass imaging of single biological macro-molecules, *Science*, **360**(6387), 423-427 (2018).[5]

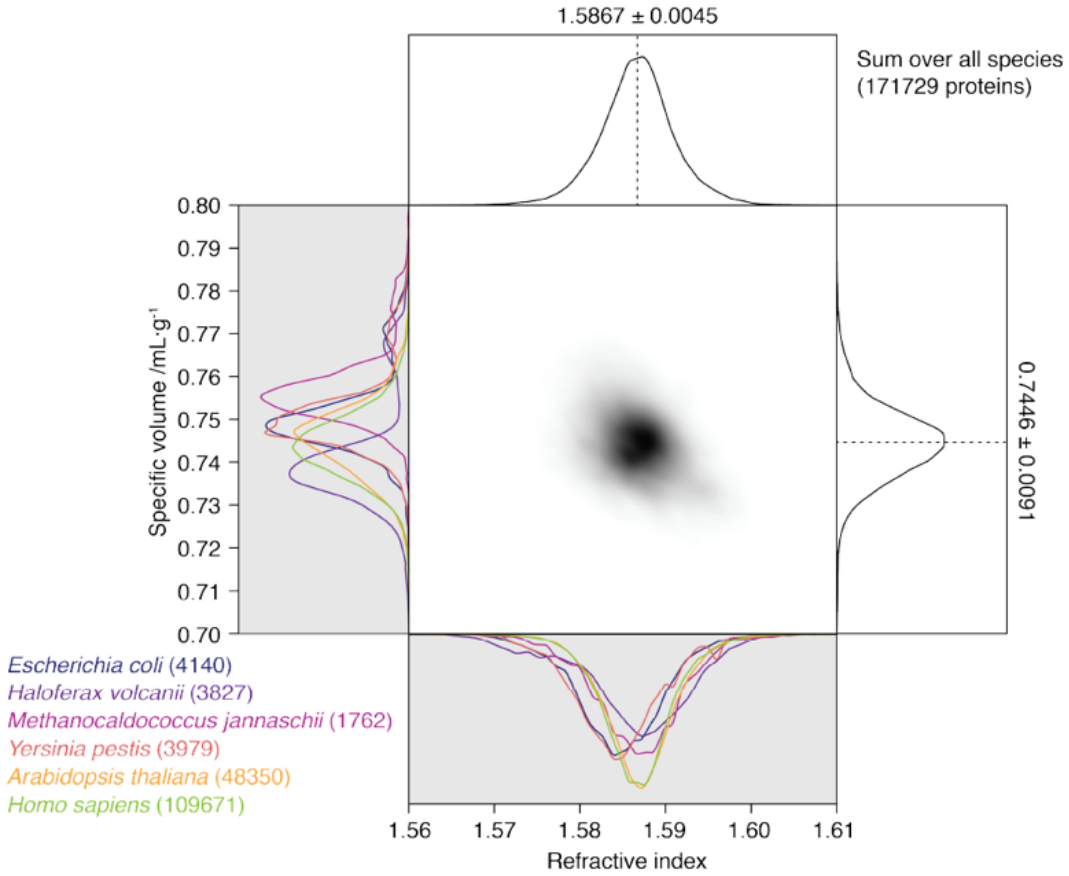


Figure 2.17: Calculated distributions of the specific volumes and refractive indices of different proteins. Distributions of the refractive indices (top - combined data, bottom - individual) and specific volumes (left - combined data, right - average) for all known proteins in the six genomes, indicated on the bottom left. The values were calculated from the sequences of amino acids, and the two-dimensional distributions of both, with the combined data shown in the middle. The individual distributions on the left and bottom were renormalised for identical areas. Figure copied with permission from the American Association for the Advancement of Science, from G. Young and N. Hundt *et al.*, Quantitative mass imaging of single biological macro-molecules, *Science*, **360**(6387), 423-427 (2018).[5]

However, the upper concentration limit for MP measurements remains unconquered, with the main limitations posed by the camera frame rate and the rapid particle landing rate preventing identification of distinct particles as they interact with the glass coverslip. Similarly, whilst the MW measured for the lipid nanodiscs and glycosylated proteins appears to roughly match those detected by complementary methods, a more careful evaluation is desirable and of immense importance for further applications of this methodology to more complex biological systems and

questions. This work briefly attempts to define the contribution of glycans to the MW measured by MP in Chapter 5. However, since the main focus is on the interactions of complex, heterogenous systems of membrane proteins and viral glycoproteins, a more thorough analysis of the contributions of various biomolecules to the contrast detected by MP remains beyond the scope of this work.

3

Mass Photometry of Membrane Proteins

Contents

3.1 Introduction **66**
3.2 Materials and Methods **69**
3.2.1 Mass Photometry Measurements 69
3.2.2 Mass Photometry Image Processing 71
3.2.3 Mass Photometry Calibration Procedure 72
3.2.4 Detergents 72
3.2.5 Amphipols 75
3.2.6 Lipid Nanodiscs 78
3.2.7 Native Nanodiscs 80
3.3 Results **82**
3.3.1 Empty Protein Carriers 82
3.3.2 Detergents 85
3.3.3 Amphipols 88
3.3.4 Lipid Nanodiscs 90
3.3.5 Native Nanodiscs 92
3.4 Discussion and Conclusions **96**

The work presented in this chapter is adapted with permission from:

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.

KcsA (*S. lividans*) was produced, inserted into lipid nanodiscs and measured with mass photometry (MP) and 2D[^{15}N, ^{1}H]-HMQC nuclear magnetic resonance (NMR) by Cédric Eichmann (ETH Zürich, CH). The bo_3 oxidase plasmid (*E. coli*) was kindly gifted by Prof. Christoph von Ballmoos (University of Bern, CH). The protein was produced and measured with SEC-MALS by Johannes Schimpf (Albert-Ludwigs Universität, DE), and with MP by Adar Sonn-Segev (University of Oxford, UK). The monomer-dimer LMNG equilibrium was modelled by Lucas S. P. Rudden (Durham University, UK). The respiratory complex I (*E. coli*) was provided by Brenda A. Schulman (Tennessee, US), and measured with MP by Adar Sonn-Segev. OmpF (*E. coli*) was provided by Joseph Gault (University of Oxford). SMA anhydride copolymer Xiran 30010 was gifted by Polyscope Polysciences (Geleen, NL). DMPC SMALPs were assembled by Adrian H. Kopf (Utrecht University, NL), who also produced KcsA (*E. coli*), purified it in SMALPs, and analysed their lipid content. Except for these, all the mass photometry measurements and analysis were performed by the author, together with the negative-stain TEM experiments, and the APol insertion of KcsA and APol.

3.1 Introduction

Integral membrane proteins (IMPs) constitute 20-30% of encoded gene products and have diverse functions, from signalling and transport across membranes to catalysis and mediation of enzymatic reactions.[1] They represent the majority of small molecule drug targets, encompassing G protein-coupled receptors, almost 50% of which are druggable.[2] IMPs are generally more complex to study than soluble proteins because their hydrophobic transmembrane region is intrinsically unstable in aqueous solution, requiring astute strategies to solubilise the protein or maintain a

native-like lipid environment. As a consequence, only 1.6% of the reported structures in the protein data bank (PDB) correspond to membrane proteins, calculated as the proportion of IMPs out of all unique (below 98% sequence identity) protein structures. The protein uniqueness was assessed by the author through sequence comparison,[15] and found 999 unique IMP structures out of total 61,301 unique protein structures deposited to the Protein Data Bank by the end of January 2020.

To overcome the difficulties involved in studying IMP structure and function, a diverse range of membrane mimetic systems (MMS) are used as protein carriers, varying in their usability and capacity to preserve membrane properties. They include detergent micelles, amphipols, nanodiscs and lipid particles, among others.[235] Most often, detergents are used to solubilize and purify IMPs because of their comparative ease of use, making it possible to efficiently extract proteins directly from isolated membranes or intact cells. Detergent micelles, however, do not necessarily mimic the planar structure of the native lipid bilayer or its compositional complexity, which can sometimes affect the function of proteins stabilized in them. For example, the structure or function of some detergent-solubilised IMPs can be affected by excessive delipidation and absence of crucial IMP-lipid interactions, lack of diverse biophysical forces acting *via* the planar PL bilayer, or removal of important protein co-factors.[78] Amphipols, whilst more stabilizing than detergents, rely on multiple purification steps, increasing the associated experimental complexity. Nanodisc (ND) mimetic assemblies maintain a lipid bilayer *via* annular amphipathic membrane scaffold proteins (MSPs, for lipid nanodiscs (LNs)) or styrene-maleic acid (SMA) polymers that encapsulate IMPs in a more native-like environment. SMA lipid particles (SMALPs) and MSP LNs exhibit the improved lateral stability of biological membranes; furthermore, unlike micelles, they do not require excess detergent, and so have a reduced background signal in structural studies using

certain methods, such as cryo-EM.[236]

MMS are inherently heterogeneous, with varying numbers of lipids and other solubilizing agents. Coupled with the heterogeneity of IMPs, for example in terms of oligomeric size, studies of IMP structure and function involve challenging and varying combinations of proteins and carriers. This complexity can significantly impact the feasibility, efficacy and ultimate success of IMP studies. As a result, careful purification and characterization of IMP preparations is essential for both functional and structural studies. Current protein purification and characterization approaches usually rely on a combination of chromatographic techniques, SDS-PAGE, and further analysis by size-exclusion chromatography (SEC), analytical ultracentrifugation, multi-angle light scattering (MALS), negative stain EM or native mass spectrometry (MS) (where applicable). Though well established, these workflows are time-consuming and pose challenges. For instance, MALS characterization requires a considerable amount of material ($>$ 100 ng) and high-resolution separation by SEC. The success of negative stain EM depends to some degree on prior knowledge of likely heterogeneity and sufficient protein size, while native MS of IMPs requires significant expertise and specialized instrumentation. Both of the latter methods necessarily operate under non-native conditions. A method capable of detailed, rapid and accurate sample characterization in solution could thus dramatically accelerate and improve structural and functional *in vitro* studies of IMPs.

We have recently introduced mass photometry (MP) and demonstrated its capabilities in terms of determining molecular mass, resolving different oligomeric states and detecting ligand binding to soluble proteins in a label-free, single molecule sensitive fashion in solution.[5] Given the importance of sample quality, we thus set out to explore the applicability and performance of MP for studying IMPs. Of particular

appeal in this context is the universal nature of MP, relying on the detection of changes in the reflectivity of a glass-water interface caused by interference between scattered and reflected light when individual objects bind to the interface.[4] The reflectivity change upon the molecule landing can then be used to determine the object mass using appropriate calibrants. MP reveals true equilibrium distributions, reports on populations rather than ensembles due to its single-molecule nature, requires minimal sample volumes (μL) and concentrations ($\ll \mu$M), and does not rely on indirect absorption measurements to determine molecular mass. Moreover, MP naturally extends towards characterizing structural heterogeneity,[6] and protein-protein interactions,[7] resulting in a universal platform for studying biomolecules.

3.2 Materials and Methods

All materials were sourced from Sigma Aldrich, unless stated otherwise. The majority of mass photometry measurements were performed by the author, unless specified. Samples were provided by a variety of groups, listed in the specific sections. The NMR KcsA activity measurements following reconstitution in LNs using different ratios were performed by the Rieker group (ETH Zürich).

3.2.1 Mass Photometry Measurements

Mass photometry data was acquired in microscope flow chambers. All microscope coverslips (No. 1.5, 24 × 50 and 24 × 24 mm^2, VMR) were cleaned by sonication with 50% isopropanol (HPLC grade)/Milli-Q H_2O (MQ), followed by sonication in MQ (5 min each). Microscope coverslips were dried by either a clean nitrogen stream or in the oven at 110 °C for 1 h. A small proportion of coverslips was cleaned by applying a layer of First Contact Polymer Optics Cleaner onto the surface, letting it dry for 15 min, peeling off the solidified layer, followed by rinsing in ethanol (HPLC grade)/MQ and drying with a clean nitrogen stream. Flow chambers were assembled from the clean coverslip immediately after the cleaning process, using

double-sided-sticky tape (3M),[5] and stored prior to use for up to three weeks. MP measurements were performed at a range of concentrations, approximately 10-50 nM. Where the protein stock concentration was higher, diluting to nM was done immediately prior to the measurement (unless stated otherwise). For each MP measurement, a buffer solution was added to the flow chamber and the focus position identified and secured for the entire measurement using a focus feedback loop based on total internal reflection of a reference laser beam.[4] Each measurement was taken for either 60 or 90 s after $\sim$15 μL of the diluted sample was introduced into the flowchamber.

All measurements were performed on three similar mass photometers. Most of the data was acquired using a home-built mass photometer as described previously.[5] Briefly, the output of a 520 nm laser diode (Lasertack) was collimated and sent through a pair of acousto-optic deflectors (AODs, AA Optoelectronic DTSXY-400). A 4f telecentric lens system images the deflection by the AODs into the back focal plane of the microscope objective (Olympus UApo N, 100 $\times$, 1.49 NA). The objective collects light reflected at the interface between a glass coverslip and some of the light scattered by the sample, with efficient separation of illumination and detection achieved through the combination of a polarizing beam splitter and a quarter-wave plate (Thorlabs). The same telecentric lens system images the back focal plane of the objective onto a partial reflector made from a thin layer of silver of 2.5 mm diameter deposited onto a window, which selectively attenuates the reflected light compared to light from point scatterers at the surface by a factor of about 1000. A final lens images the sample onto a CMOS camera (Ximea, MC023MG-SY) with 277.8 $\times$ magnification, resulting in a final pixel size of 21.1 nm/pixel. Before saving each movie file, areas of 4 $\times$ 4 pixels were binned for an effective pixel size of 84.4 nm/pixel, and frames were averaged 5- fold in time. The entire setup was constructed on a thick (50 mm) metal aluminium plate, and fully enclosed to

minimise the influence of air currents.

The bo_3 oxidase and respiratory complex I data were acquired with a similar, but commercial mass photometer, One^{MP} (Refeyn LTD, Oxford, UK), with effective pixel size and frame rate as described above. Empty SMALP and OmpF–APol data were acquired on a home-built mass photometer which uses 445 nm laser diode. Here, the pixel size is 23.4 nm and frame rate is 1 kHz. Prior to saving the images, a 5-fold time average and 3 × 3 pixel binning were applied, resulting in an effective frame rate of 200 Hz and an effective pixel size of 70.2 nm. Data acquisition was performed using either custom software written in Labview (for the home-built mass photometers) or Acquire^{MP} (Refeyn LTD, v1.2.1) for the commercial instrument.

3.2.2 Mass Photometry Image Processing

The videos of proteins binding to the glass surface were analysed with Discover^{MP} (Refeyn Ltd). The software detects binding events and determines the respective interferometric scattering contrasts. The user can choose how many frames are averaged for continuous background removal (n_{avg}) and can set the thresholds T1 and T2 for the two image filters, which are used to detect the binding events. Filter 1 is based on t-tests of the pixel intensity fluctuations. As a particle (*i.e.* IMP) binds to the glass, the pixel intensity changes suddenly. This change is associated with an increase of the filter 1 score calculated as $-ln(p)$, where p is the p-value of the t-test comparing pixel values at n_{avg} frames before and n_{avg} frames after the event. The smallest intensity jump amplitude that exceeds random noise fluctuations and is associated with a binding event is controlled by the value of threshold T1.

The signatures of the binding events in interferometric images are radially symmetric. Filter 2 measures the radial symmetry of all pixel neighbourhoods of the interferometric images.[237] The lowest symmetry score expected at the centre of

a peak is defined as threshold T2. Pixel clusters that exceed both thresholds T1 and T2 are used for peak fitting. The amplitude of the peak fit provides an estimate for the interferometric peak contrast. The peak signature (point spread function) is modelled as a superposition of two Sombrero functions multiplied by two Gaussians.[5]

3.2.3 Mass Photometry Calibration Procedure

The calibration protocol for the majority of experiments in this chapter included the measurement of several different protein oligomer solutions, with known masses. Each MP calibration was analysed using Discover^{MP}, where mean contrast values of all peaks were determined in the software using Gaussian fitting (Figure 3.1A). The mean contrast values were then plotted against the known mass of the proteins (Figure 3.1B) and fitted to a line, $y = b \cdot x$, with y = contrast, x = mass and b = calibration factor. For the data shown in Figures 3.5F, 3.9C, and 3.10B, calibration was performed using a mixture of protein standards of known molecular weight (MW), namely bovine serum albumin, alcohol dehydrogenase, apo-ferritin and thyroglobulin.

3.2.4 Detergents

Detergents n-dodecyl-β-D-maltoside (DDM), n-octyl-β-D-glucoside (OG), 2,2-didecyl-propane-1,3-bis-β-D-maltopyranoside (LMNG) were purchased from Anatrace. Detergent n-dodecylphosphocholine (DPC) was provided by the Rieker group. The critical micelle concentrations, as well as concentrations of stock solutions and stock solution preparation conditions, in addition to dilution, are provided in Table 3.1.

Potassium channel KcsA from *Streptomyces lividans* was expressed, purified and measured by MP in DPC by the Rieker group (ETH Zürich), as detailed in

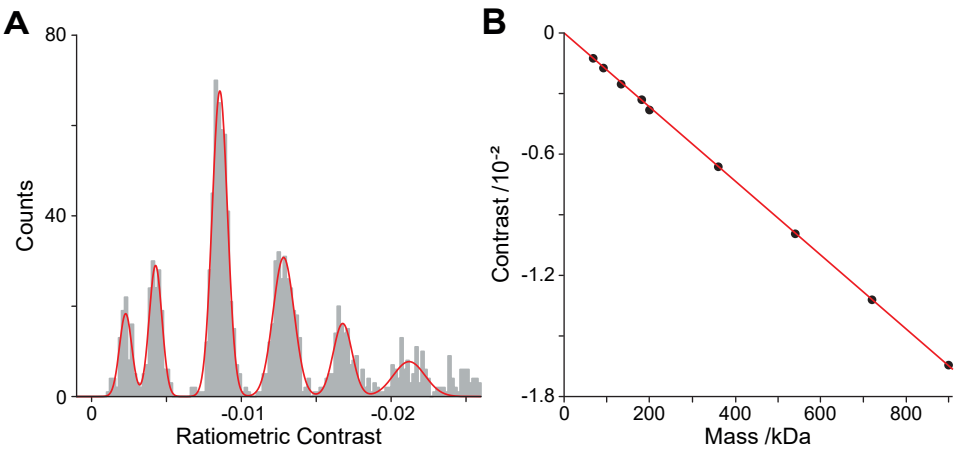


Figure 3.1: Contrast to mass calibration. A. Example histogram of a mixture of two mass calibrants as measured by MP. **B.** The contrast of two proteins of known MWs with different oligomeric states, bovine-serum albumin (BSA) and MS1000 mass standard were plotted against their mass (66, 132, and 198 kDa for BSA; 90, 180, 360, 540, 720, and 900 kDa for MS1000). The red line is the linear fit to the data according to $y = b \cdot x$, with b = -1.83 $\times$ 10^{-5} being the calibration factor.

Detergent	CMC (conditions)	Stock conc.	Stock preparation	Measurement conc.
DDM	0.0087%[238] (0.2M NaCl)	1%	12 h, 4 °C, shaking	0.02%
OG	0.53%[239] (H_2O)	1%	-	1%
LMNG	0.001%[240] (H_2O)	1-10%	12 h, 50 °C, shaking	0.0025%
DPC	0.047%[241] (H_2O)	0.09%	-	0.09%

Table 3.1: Experimental details of detergents used. All concentrations are shown as a weight/volume (w/v) percentage. Conc. - concentration.

Eichmann *et al.*[242] Briefly, KcsA was produced using a pET-28a vector containing the coding sequence of the wild-type protein fused to a thrombin-cleavable N-terminal His-tag (monomer 19.9 kDa). KcsA was expressed in the *E. coli* strain BL21 Star (DE3) (Invitrogen) in Luria-Bertani (LB) media. Cells were grown at 37 °C until OD_{600} = 0.8 and protein production was initiated by addition of 0.5 mM IPTG (Invitrogen). The culture was incubated for an additional 5 h at 37 °C. After protein expression, the cells were lysed by two microfluidizer (Microfluidics)

cycles. KcsA was extracted from the membrane with 20 mM (DPC) detergent by gentle stirring at 4 °C overnight. The cleared extraction mixture was loaded on a Ni^{2+} Sepharose 6 Fast Flow resin (GE Healthcare). After washing the resin with 3.3 mM DPC, KcsA was eluted with 300 mM imidazole and 3.3 mM DPC. Fractions containing the protein were pooled together, and the buffer was exchanged to 20 mM Tris-HCl pH 7.4 and 3 mM DPC using a PD-10 (GE Healthcare) desalting column. Expressed protein was stored at 166 μM at 4 °C in a buffer (20 mM DPC, 20 mM Bis-Tris, 150 mM NaCl, pH 7).

bo_3 oxidase from *E. coli* was provided by the Friedrich group. Briefly, bo_3 oxidase was expressed in the *E. coli* strain BL21(DE3) star/pETcyo hiscyoC. Plasmid pETcyo hiscyoC was a kind gift from Prof. Christoph von Ballmoos (Bern, Switzerland). Cells were grown at 37 °C in LB medium supplemented with 0.5% glycerol, 2 mM $MgSO_4$, 0.03 mM $FeSO_4$ and 0.01 mM $CuSO_4$. At an optical density at 600 nm of 1.5, gene expression was induced by an addition of 0.5 mM IPTG and cells were grown for a further 2 h. Cytoplasmic membranes were suspended in buffer A (50 mM 3-(N-morpholino)propanesulfonic acid (MOPS) and 20 mM NaCl, pH 7.5) and membrane proteins were solubilized by adding LMNG drop wise to a final concentration of 2% (weight/volume). Solubilized proteins were loaded onto a Probond Ni^{2+}-IDA column (25 mL) equilibrated in buffer A containing 20 mM imidazole and 0.005% LMNG. After washing with 92 mM imidazole, bound proteins were eluted in a single step to 284 mM imidazole. The eluate was concentrated by ultrafiltration (Amicon Ultra-15, 100 kDa MWCO) and subjected to SEC on Superose 6 (300 mL) in buffer A containing 0.005% LMNG. The main peak eluting after 177 mL was concentrated to 100 μM, aliquots were flash frozen in liquid nitrogen and stored at -80 °C until further use in 50 mM MOPS and 20 mM NaCl, pH 7.5 with 0.005% LMNG. SEC–multi-angle light scattering (SEC-MALS) validation was performed by the Friedrich group and the molecular dynamics simulation was done by Prof. Matteo Degiacomi

and Dr. Lucas S. P. Rudden (University of Durham, UK).

The respiratory complex I from *E. coli* was provided by Brenda A. Schulman (Tennessee, US). The complex was prepared as described previously,[243] with slight modifications. After affinity chromatography on a Probond Ni^{2+}-IDA column (35 mL), the complex was subjected to a Superose 6 (24 mL) SEC column. Peak fractions were concentrated to 30 μM (Amicon Ultra-15, 100 kDa MWCO) and stored in 50 mM MES/NaOH, 50 mM NaCl, 5 mM $MgCl_2$, pH 6.0 with 0.005% LMNG at -80 °C. MP measurements of complex I were performed at 15 nM. Dilution to measured concentrations was performed with buffer A either without any detergents or with 0.001% LMNG, and the complex was stored in 50 mM MES/NaOH, 50 mM NaCl, 5 mM $MgCl_2$, pH 6.0 with 0.005% LMNG at -80 °C. The MP measurements of bo_3 oxidase and the respiratory complex I were performed together with Dr. Adar Sonn-Segev (Kukura Group, University of Oxford).

3.2.5 Amphipols

A8-35 amphipol (APol) was purchased from Anatrace. 10% *w/v* stock solution was prepared by mixing with Milli-Q H_2O (MQ) and incubating at 4 °C for 12 h, with agitation.

E. coli outer membrane protein F (OmpF) was expressed and purified as described previously,[244] in 20 mM phosphate buffer (pH 6.5, 1% OG), and concentrated to 11 µM. Bio-Beads (Bio-Rad) were used for detergent adsorption for the exchange from detergent to APol, according to Zoonens *et al.*[58] The Bio-Beads were washed with methanol (HPLC grade), MQ, and twice with OmpF phosphate buffer not containing OG. Each step was done at room temperature with stirring, with MQ rinsing in between steps. A8-35 APol stock solution was diluted 10 × in OmpF

buffer. A range of APol to integral membrane protein ratios was calculated, from 0:1 (control) to 5:1, to determine the lowest ratio for effective protein extraction, in order to avoid excess APol content in the sample. All reaction components were added to a final exchange volume of 200 μL. An excess of washed Bio-Beads was added to each ratio, which was then incubated at 4 °C for a total of 16 h, with a change of Bio-Beads after 2 h to eliminate Bio-Beads saturated with detergent. After detergent removal, the solution was removed using fine tipped pipette tips and the sample was buffer-exchanged into OmpF Tris buffer (20 mM Tris-HCl, 100 mM NaCl, pH 8.0) using 0.5 mL molecular concentrators with MWCO of 30 kDa (Amicon). The sample was centrifuged (10 000 $\times$ g, 4 °C, 2 min), the flow through was discarded and 300 μL of fresh buffer was added, repeated 5 times, whilst making sure no protein precipitation occurred after each spin cycle. Following the buffer exchange, light absorbance at 280 nm was measured using a Nanodrop device to determine protein concentration. The samples were subsequently ultra-centrifuged (150 000 $\times$ g, 20 min), after which the protein concentration of supernatant of each sample was measured again, in order to compare the amount of protein retained, with the protein that did not get integrated into the APol particles aggregating and settling at the bottom in the pellet during ultracentrifugation (Figure 3.2). This method was used to determine the minimal amount of amphipol needed to extract protein without precipitation. The samples were then stored at 4 °C for up to a few weeks.

Potassium channel KcsA solubilised in DPC micelles from the Rieker group (ETH Zürich) was exchanged to APol using the same protocol described above for OmpF, with a few minor differences. KcsA buffer (20 mM Bis-Tris, 150 mM NaCl, pH 7) was used throughout the whole protocol and a second exchange of Bio-Beads was added after 4 h of incubation in the APol preparation. Figure 3.3 shows the

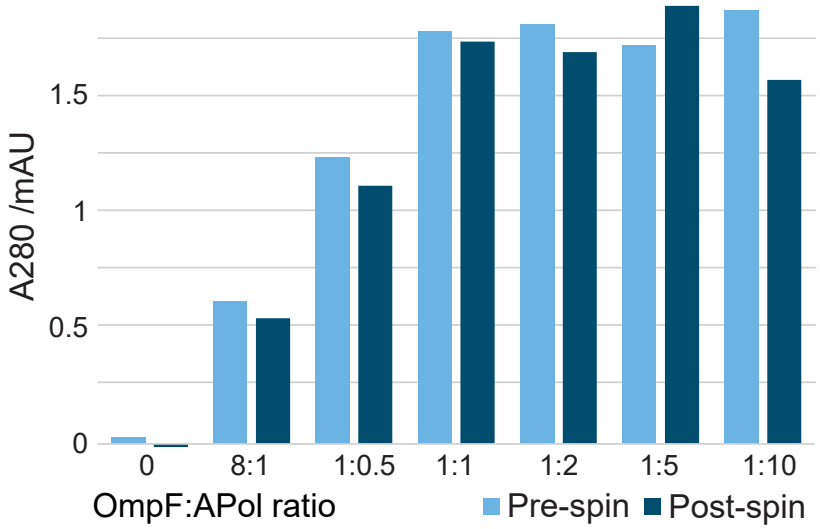


Figure 3.2: Analysis of OmpF detergent-APol exchange. UV-VIS A280 measurement of OmpF prior to ultra-centrifugation step in the amphipol insertion procedure (light blue) compared to the absorbance of the supernatant post-ultracentrifugation (dark blue).

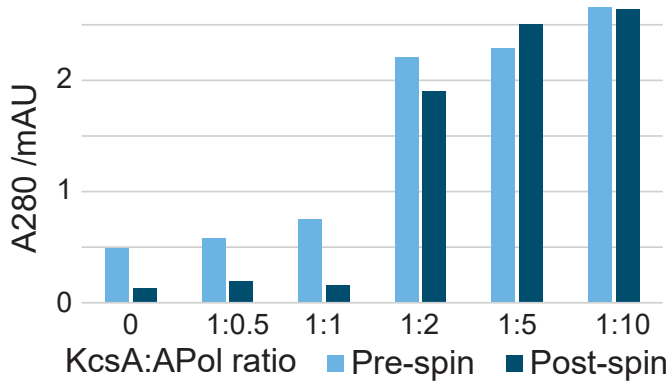


Figure 3.3: Analysis of KcsA detergent-APol exchange. **A** UV-VIS A280 measurement of KcsA prior to ultra-centrifugation step in the APol insertion procedure (light blue) compared to the absorbance of the supernatant post-ultracentrifugation (dark blue).

absorbance values for the different ratio for KcsA-APol.

All samples were measured using MP at 15-25 nM, diluted straight before measurement, unless specified otherwise.

3.2.6 Lipid Nanodiscs

The short length empty lipid nanodiscs (LNs) MSPΔH5 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC) and the regular length empty LN MSP1D1 DMPC were provided by the Rieker group (ETH Zürich). The empty LNs were measured at 10-25 nM.

Empty MSPΔH5 LNs were prepared as described previously.[245] Briefly, MSPΔH5 membrane scaffold protein (MSP) was overexpressed in BL21 Star (DE3) cells (Thermo Fisher) at 28 °C for 5 h and was purified after harvest by His-tag affinity chromatography involving a number of washing steps. Then, 200 μM MSPΔH5 were assembled with DMPC 14:0 lipids (Avanti Polar Lipids) in a molar ratio of 1:50 by shaking them in a glass vial for 8 h at 27 °C. Subsequently, 1 g/mL Bio-Beads SM-2 (Biorad) were added for 4 h and the MSPΔH5 LNs were purified by SEC using a Superdex 200 Increase 10/300GL column (GE Healthcare).

For the expression of MSP1D1 LNs, the MSP was fused to a Tobacco Etch Virus (TEV) protease cleavable N-terminal His-tag encoded in a pET-28a vector. The MSP1D1 variant of MSP1 deletes the first 11 amino acids of the original MSP1 sequence.[96] Expression and purification of the protein was carried out as described previously.[246] The His-tag on MSP1D1 was cleaved by the addition of TEV protease for 16 h at room temperature. MSP1D1 without His-tag has a MW of 22.0 kDa.

Reconstitution of KcsA into DMPC LNs and formation of empty LNs was performed following the protocol of Shenkarev *et al.*[246] KcsA was incorporated into saturated DMPC 14:0 MSP1D1 LNs, as well as DMPC MSPΔH5 LNs. For MSP1D1 LNs, DMPC lipids were first solubilized in sodium cholate (cholate/DMPC 2:1 molar ratio). The purified KcsA sample was mixed with scaffolding protein

MSP1D1, DMPC, and sodium cholate at a molar ratio of 1:20:800:1600 in a buffer containing 20 mM Tris-HCl pH 7.4 and 150 mM NaCl. Our choice of the KcsA tetramer:MSP:DMPC ratio was based on published reports from membrane protein groups that have extensively studied KcsA in LNs.[246,247] The final concentration of KcsA was 15 μM. The mixture was incubated overnight at 27 °C while shaking at 150 rpm. Incorporation of KcsA into LNs was initiated by addition of 80% w/v Bio-Beads for 2 h at 27 °C while shaking at 150 rpm. Empty nanodiscs were separated from LNs containing N-terminal His-tagged KcsA by Ni^{2+}-NTA affinity chromatography. The fraction of LNs containing KcsA was eluted with 300 mM imidazole and used for mass photometry measurements. For the MSPΔH5 LNs, DMPC lipids were solubilized in 200 mM sodium cholate at a concentration of 100 mM. LN assembly mixtures containing membrane protein KcsA (tetramer), scaffolding protein MSP1D1 and DMPC lipids were premixed at two different ratios (Table 3.2) to a volume of 500 μL and incubated at room temperature overnight. LN reconstitution was triggered by the addition of 50% mass/volume BioBeads to the assembly mixtures over 4 h. Assembled LNs were purified with SEC using Superdex 200 chromatographic column (GE Healthcare).

	KcsA tetramer : MSP1D1 : DMPC (molar ratio)	**KcsA tetramer /μM**	**MSP1D1 /μM**	**DMPC lipids /mM**
Assembly A	1:20:800	20	400	16
Assembly B	1:4:100	50	200	5

Table 3.2: KcsA:MSP1D1:DMPC LN assembly. Two assembly ratios screened for the reconstitution of the potassium channel KcsA into the lipid nanodiscs.

For small scale reconstitution of KcsA into LNs, the assembly mixtures containing membrane protein KcsA (tetramer), scaffolding protein MSP1D1 and DMPC lipids were premixed at varying ratios to a volume of 50 μL. Concentration of DMPC lipids

was kept at 16 mM, concentration of MSP was kept at 400 μM and concentration of KcsA (monomer) was varied in range from 20 μM up to 350 μM. After the overnight incubation at room temperature, LN reconstitution was triggered by the addition of 50% m/v Bio-Beads to the assembly mixtures over 4 h. Assembled LNs were purified with batch Ni^{2+}-NTA chromatography.

3.2.7 Native Nanodiscs

Styrene-maleic acid (SMA) copolymers, SMA-lipid particles (SMALPs) and KcsA native nanodiscs were provided by the Killier group (Utrecht University, Netherlands). Briefly, commercially available styrene-maleic anhydride (SMAnh) copolymer, Xiran30010 (number-average molecular weight (Mn) ∼2.5 kDa, polydispersity index (PDI) ∼2.6), was a kind gift from Polyscope Polysciences (Geleen, Netherlands). Conversion of the SMAnh polymers into its acid form (SMA) was achieved by hydrolysis under base-catalysed conditions as detailed previously.[248] SMA stock solutions were prepared at final concentrations of 5% w/v.

For preparation of *E. coli* membranes overexpressing KcsA, total membrane fractions of *E. coli* cells (strain BL21(DE3)) producing KcsA were obtained as described previously.[148] Briefly, cells were transformed with an N-terminal His-tagged pT7-KcsA vector, containing the KcsA gene. Membrane preparations were obtained by differential centrifugation after cell wall lysis and mechanical disruption through a French press. Membrane pellets were resuspended in buffer (Tris-HCl 5 mM, NaCl 300 mM, KCl 15 mM, pH 8) to an OD_{600} of ∼4. After lipid extraction according to the method of Bligh and Dyer,[249] the total phosphate content was determined to be 10 mM using the method of Rouser *et al.*[250] Membrane suspensions were stored at -20 °C until further use.

For SMA-mediated solubilization of KcsA from *E. coli* membranes, membrane stocks (4.8 mL) were thawed on ice and diluted with solubilization buffer (Tris-HCl 50 mM, NaCl 300 mM, KCl 15 mM, imidazole 10 mM, pH 8) to a total volume of 33 mL, resulting in phosphate concentration of ∼1.5 mM and protein concentration of ∼1 mg/mL. Polymer was added to the suspensions at a final concentration of 1% w/v. The mixture was incubated on a rotary disc at 4 °C overnight, during which the suspension cleared up significantly. To remove any unsolubilised material, the mixture was centrifuged at ∼40 000 × *g* for 1 h at 4 °C to pellet the non-soluble fraction. The solubilized fraction (supernatant) was carefully removed and used for His-tag purification of the nanodiscs. For purification of KcsA native nanodiscs, the solubilized fraction containing the nanodiscs was added to 6 mL of HisPure Ni^{2+}-NTA agarose beads (Thermo Scientific) and incubated on a rotary disc at 4 °C overnight. The beads were then loaded on a gravity-flow column for the affinity purification. First, the flow-through was collected, followed by washing with increasing amounts of imidazole (10 mM, then 50 mM) and finally elution with a high concentration of imidazole (300 mM). Elution fractions containing pure KcsA were combined, and the pooled fractions were concentrated using centrifuge spin filters (Amicon, 15 mL, 10 kDa MWCO). The imidazole was removed by washing the concentrate three times with buffer (Tris-HCl 50 mM, NaCl 300 mM, KCl 15 mM, pH 8).

The isolated KcsA native nanodiscs were further purified and analysed by SEC on an ÄKTA pure system (GE Healthcare), using a Superdex 200 10/300 GL SEC column (GE Healthcare). Protein was eluted using buffer (Tris-HCl 50 mM, NaCl 150 mM, KCl 15 mM, NaN_3 1 mM, pH 8). Fractions were collected on an autosampler and UV detection was performed at $\lambda = 280$ nm. Collected fractions were frozen and stored at -80 °C until further use. To determine sample purity using SDS-PAGE, the samples were incubated with Laemilli buffer without any reducing

agent, loaded onto 13% SDS-PAGE gel, and run at 175 V for 1 h. The gels were stained with Coomasie blue and scanned for further analysis. In SMALPs, the lipids were at 2.5 mM, and polymers at 1:1 or 1:4 ratio. For the MP measurements, the KcsA native nanodiscs were diluted from SEC fractions 4-10 times, and measured together with Dr. Adar Sonn-Segev.

3.3 Results

3.3.1 Empty Protein Carriers

MP is subject to background signatures from nanoscopic empty carriers, because they produce a scattering signal comparable to that of small proteins. To assess the extent to which these could hamper MP measurements, and to help interpret downstream measurements of complex systems, we first studied commonly used membrane mimetic systems in the absence of protein. We began with the detergent LMNG due to its popularity in structural studies based on enhanced stability of membrane proteins solubilized in this manner.[251]

In detergent micelles, using an abundance of detergent above the critical micelle concentration (CMC) aids in solubilizing hydrophobic, transmembrane regions of the IMP, thereby maintaining solubility. To assess the effect of an abundance of detergent micelles, we compared LMNG at 2.5 × CMC (0.0025% w/v) in buffer (50 mM MOPS, 20 mM NaCl, pH 7.5) (Figure 3.4, top frame) compared to a buffer blank (described in section 2.4). We observed considerable non-specific detergent interactions due to the effective concentration of LMNG micelles at 2.5 × CMC (25 μM)[251] being an order of magnitude larger than the current optimal concentration range for MP (high pM to mid nM). This caused surface saturation and rapid binding and unbinding on the glass interface, effectively increasing the imaging background, reducing mass resolution and raising the lower detection limit.

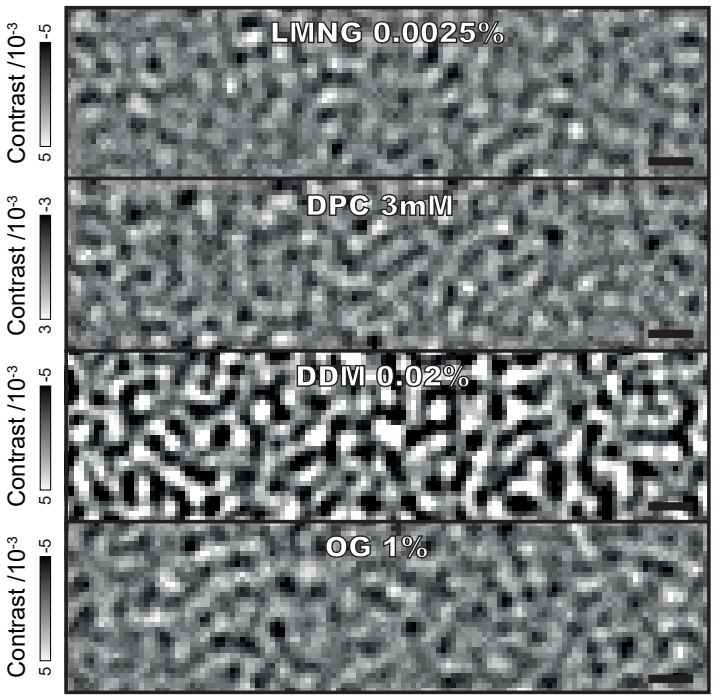


Figure 3.4: MP of detergent micelles. Images of detergents LMNG, DPC, DDM and OG at or above the critical micelle concentration (CMC), with the concentrations indicated in the top right corner. Scale bar represents 1 μm. The contrast scale bar is shown on the left of the frames.

Other detergents produced similar signatures at concentrations above the CMC (Figure 3.4).

Lipid nanodisc-based systems offer CMC-independent stability, a bilayer-like environment, and potentially lower levels of carrier heterogeneity, as evidenced in the original interferometric scattering mass spectrometry (iSCAMS) study.[5] We therefore imaged MSP1D1 DMPC LNs and SMALPs (SMA polymer to DMPC lipid ratio 1:1 w/w) using MP. NDs and SMALPs differed from LMNG in that distinct, homogeneous particles were detected (Figures 3.5A and B, respectively). Both unoccupied LNs (∼105 kDa, Figure 3.5C) and unoccupied SMALPs (∼115 kDa) produced mostly homogeneous mass distributions (Figure 3.5D).

To confirm that we detected unoccupied SMALPs and not disassembled polymer

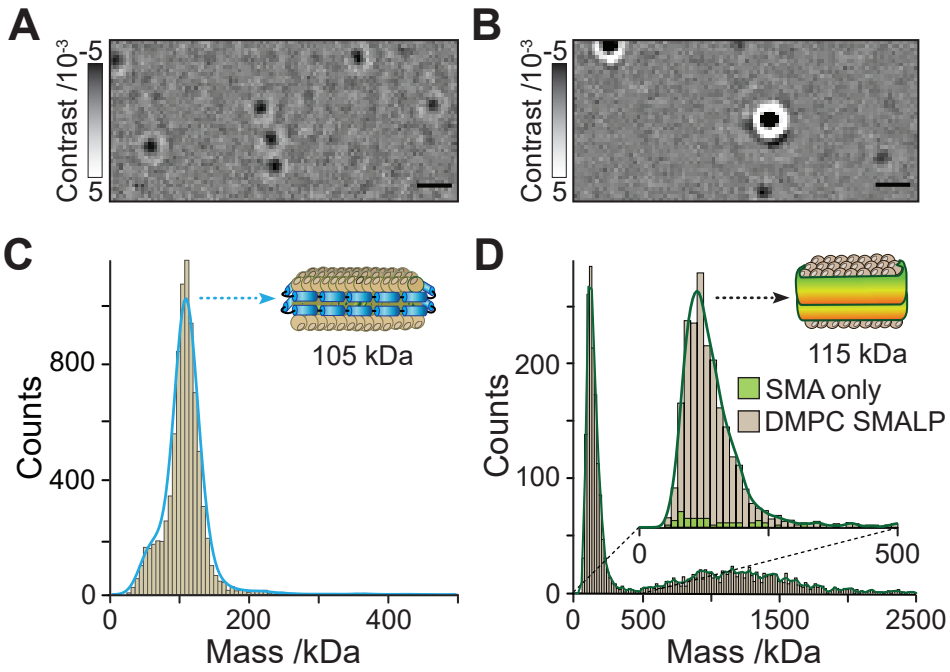


Figure 3.5: Mass photometry of detergent micelles and nanodisc membrane protein carriers. A. Representative MP movie frame of empty MSP1D1 DMPC lipid nanodiscs. **B.** Representative MP movie frames of empty DMPC SMALPs at a polymer to lipid ratio of 1:1 (weight/weight). Scale bars in **A.** and **B.** represent 1 μm. **C.** The resulting MP histograms and kernel density estimates for empty LNs. **D.** Corresponding distribution for unoccupied SMALPs (light brown) and SMA polymer aggregates (light green).

aggregates,[252] we measured SMA polymer at equal concentration as the polymer in unoccupied SMALPs and found a negligible number of particles (Figure 3.5D, inset, light green histogram). Particle distribution analysis revealed that the standard deviation (SD) of the ND peak fit was 23 kDa, or approximately 19%. Compared to standard soluble protein SD of approximately 8-10% of the MW,[5] LNs and membrane mimetics tend to have broader mass distribution due to variations in the number of constituent molecules present. For SMALPs, the SD of the main peak was even larger at 38 kDa, likely arising from varying numbers of lipids and SMA polymers per disc.

Unlike LNs, which are limited in size due to the length of the annular belt protein, SMA polymers can in principle form discs of unlimited size, but it is unclear what

governs the size of assembled SMALPs.[149] A recent report estimated 140 DMPC lipids per unoccupied SMALP,[253] which would correspond to approximately 20 kDa of SMA polymer and 95 kDa of DMPC lipids on average as detected by MP. In addition to the main SMALP peak, 40% of all observed particles were polydisperse and at a higher MW, approximately 500-2000 kDa, indicating some degree of ND oligomerisation. In line with this, it has been reported that SMALPs can form oligomeric 'rouleaux' stacks,[254] although this was attributed to a transmission election microscopy (TEM) artefact. This inhomogeneity can also result from an unoptimized SMALP assembly and purification process. MP can therefore guide sample production and purification in SMALP experiments.

3.3.2 Detergents

While the presence of unoccupied micelles increases the imaging background, limiting MP performance, we can still distinguish assembled complexes given sufficient differences in the protein and micelle object masses. To illustrate this, we investigated the large *E. coli* protein NADH:ubiquinone oxidoreductase (respiratory complex I, 770 kDa including detergent) in LMNG at 1 × CMC (Figure 3.6). We could clearly identify assembled protein particles, despite the presence of detergents, and determine their mass, with the observed MW of the assembled protein-detergent particles (770 kDa) being in an excellent agreement with previous studies using analytical ultracentrifugation.[255]

The major remaining limitations were a lack of specificity and detection sensitivity at low mass in the range of micellar size due to interfering detergent background, as well as a loss of mass resolution. Exploiting the slow off-rate of LMNG molecules from IMPs, gradient-based detergent removal (GRaDeR) has shown great potential to effectively remove empty micelles, while leaving membrane proteins solubilized.[55] We tested a simple 2,000 × drop dilution of complex I from

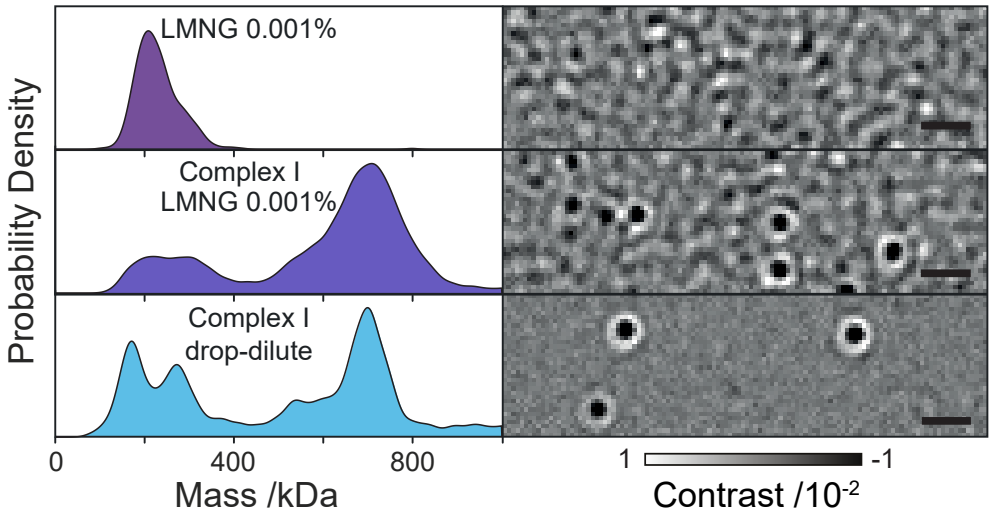


Figure 3.6: MP Analysis of respiratory complex I in LMNG. MP of LMNG at 1 × CMC (top) and distributions and typical ratiometric image of respiratory complex I in 1 × CMC LMNG (middle) and 2000 × LMNG dilution (bottom), after drop dilution. Scale bars represent 1 μm, and the contrast scale bar is shown on the bottom right.

5 × CMC into buffer without LMNG immediately prior to MP measurement. This dilution greatly reduced the detergent background, improving the measurement resolution and enabling us to resolve the low mass peaks. Importantly, the accuracy of the mass measurement was unaffected by LMNG concentration, showing that IMPs can be studied in the presence of LMNG by simply diluting out excess micelles.

We repeated this drop-dilution approach on the *E. coli* bo_3 oxidase (144 kDa monomer), where the presence of bo_3 dimers *in vitro* is debated, with some reports declaring the complete absence of dimers in detergent,[256] and others affirming their existence (although this occurred at the limit of detection, preventing further analysis).[257] We found monomers and dimers at approximately 80% and 20% relative abundances respectively (Figure 3.7). Here, the measured monomer mass of 290 kDa suggests approximately 146 LMNG detergent molecules (146 kDa) bound to bo_3, making the simplifying assumption of a similar contrast-to-mass conversion for LMNG to that of protein. Similarly, the measured dimer mass (506 kDa) corresponds to approximately 218 LMNG molecules and a 288 kDa protein contribution.

In the aforementioned study affirming the dimers' existence,[257] *bo*$_3$ dimers were suggested to be non-specific artefacts arising from detergent solubilization. We cannot exclude the possibility that dimer formation is caused by drop dilution from a high LMNG environment, although we did not observe dissociation or significant changes in the relative amount of *bo*$_3$ dimers at low concentrations over the course of 360 min (Figure 3.7D). Furthermore, SEC-MALS analysis of *bo*$_3$ at higher concentration in 0.003% LMNG (> CMC) was decidedly consistent with our MP analysis, both in terms of molar mass and the relative abundance of monomers and dimers, supporting the occurrence of *bo*$_3$ dimers *in vitro* (Figure 3.7C). Specifically, the conjugate *bo*$_3$-LMNG micelle molar mass was 293 kDa (monomer) and 464 kDa (dimer) in 72% and 28% relative abundance, respectively, in good agreement with our results.

To test the feasibility of dimer occurrence, we further explored the *bo*$_3$ oxidase dimer structure using protein-protein docking in conjunction with molecular dynamics (MD).[258] We simulated the monomer (Protein Data Bank (PDB): 1FFT)[259] in a POPC bilayer for 60 ns, and used this equilibrated structure to generate a pool of possible dimeric arrangements (the best candidate is shown in Figure 3.7B). We then calculated the transmembrane solvent-accessible surface area (SASA) of monomeric *bo*$_3$ (21,800 Å^2). Given the mass of LMNG determined by MP (146 kDa), the ratio of SASA per kDa of detergent is 151.4 Å^2 kDa^{-1}, which would amount to the presence of 144 lipids. Assuming this ratio is constant between monomeric and dimeric *bo*$_3$ oxidase, we derived the mass of LMNG bound to the dimer using, as a consensus value, the mean of dimeric model transmembrane SASA (33,500 Å^2). We obtained an LMNG mass of 221 ± 3 kDa, a prediction remarkably close to the MD predicted value of 218 kDa. These data support the presence of the *bo*$_3$ oxidase

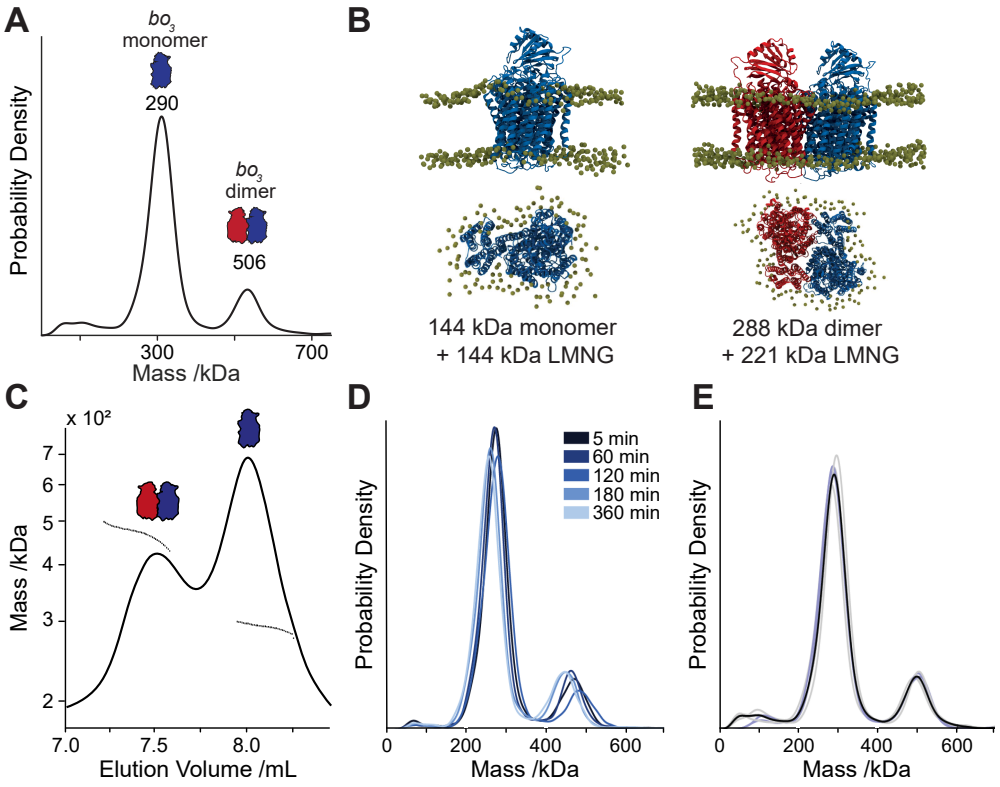


Figure 3.7: Analysis of bo_3 oxidase in LMNG using MP, MD and SEC-MALS. **A.** Detection of bo_3 oxidase monomers and dimers isolated from LMNG detergent micelles by drop dilution. **B.** Molecular docking and quantitation of bound detergents of bo_3 oxidase monomers (left) and dimers (right). **C.** SEC-MALS of bo_3 oxidase in LMNG. UV absorbance chromatogram and MALS traces (black dotted lines) of bo_3-LMNG micelle complexes in 0.003% LMNG containing running buffer. The conjugate protein-detergent molar mass was 293 kDa (monomer) and 464 kDa (dimer). The molar mass of individual components were 140 kDa bo_3 oxidase plus 153 kDa LMNG (monomer) and 259 kDa bo_3 oxidase plus 205 kDa LMNG (dimer). **D.** MP measurements of 20 nM bo_3 oxidase at different time point after drop-diluting in detergent-free buffer. Final LMNG concentration is ∼0.001 × CMC. Colours correspond to different waiting times after dilution, as indicated in the figure legend. **E.** Repeats of MP measurements of 10 nM (blue lines) and 20 nM (grey lines) bo_3 oxidase, showing the reproducibility of the measurement. The black line corresponds to the average overall repeats.

dimer *in vitro*, and illustrate the accuracy of MP in quantifying not only the mass of the polypeptide but also the detergent mass of solubilized membrane proteins.

3.3.3 Amphipols

APols represent an alternative to detergents for amphipathic IMP solubilization. APols are small polymers (∼4.3 kDa) that function similar to detergents, adsorbing

at hydrophobic patches on IMPs.[56] However, APols are more stabilizing than detergents and excess APol is not necessarily required to solubilize IMPs, meaning background signals from free APol can be minimized for structural and biophysical methods.[58] Nevertheless, their use requires an initial detergent solubilization step, adding experimental complexity compared with detergent solubilization alone. MP observations of APols alone were similar to observations of detergent micelles. To examine the applicability of MP to APols, we chose the *E. coli* OmpF, which has previously been studied by SANS in APols.[260] OmpF assembles into trimers of β-barrel monomers and functions as a small-molecule pore with an overall mass of 111 kDa.[261]

MP revealed a ratio-dependent assembly of OmpF trimers solubilized in A8-35 APols from 1:0.125 to 1:5 OmpF to APol ratios (Figure 3.8A). At or above 1:1 OmpF to APol, MP yielded single peaks of OmpF trimers, suggesting predominantly homogeneous samples. As we increased the APol concentration, we observed a clear mass increase corresponding to larger numbers (approximately 10–18) of bound APols. At sub-stoichiometric ratios, however, we found evidence for dimers and trimers of OmpF oligomers (1:0.5), consistent with published SANS results.[260] These results suggest an onset of aggregation correlated with loss of APol from OmpF-APol trimers, leading to hydrophobic protein-protein interactions of OmpF trimers. At the extreme 1:0.125 (OmpF:APol) ratio, we could no longer detect any solubilized protein, likely due to large-scale aggregation. We also observed a notable decrease in recovery at the 1:0.25 ratio compared with 1:0.125, but this is likely due to experimental error. Importantly, loss of OmpF during exchange from detergent to amphipols was minor from 1:0.5 to 1:5 protein:APol ratio based on UV-VIS measurements (Figure 3.2). By contrast, a similar titration for the *S. lividans* potassium channel, KcsA, did not reveal signs of further oligomerisation (Figure 3.8B). Instead, we observed loss of KcsA during transfer from detergent to

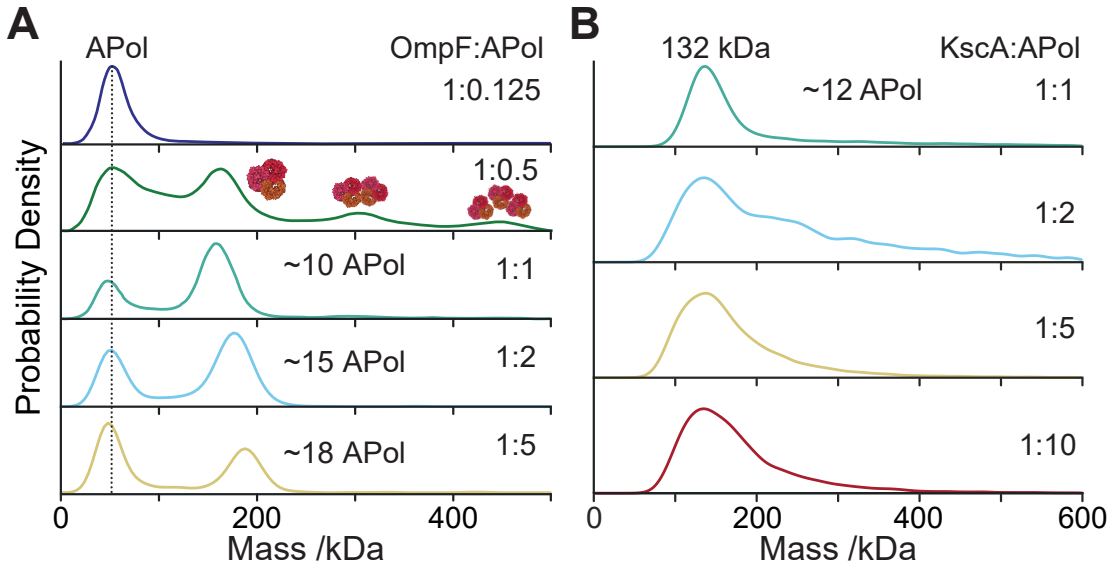


Figure 3.8: OmpF trimers and KcsA tetramers in APols at different protein:APol ratios. A. MP measurements of OmpF trimers at various OmpF:APol ratios used during APol insertion from detergent. Cartoons of OmpF trimers indicate the position of OmpF trimer, dimer of trimers, and trimer of trimers. For the ratios of 1:1 and above, the approximate number of APol molecules associated with each OmpF trimer was calculated and is displayed in the figure. The first peak, at ~40-50 kDa, is likely the free APol particles present in the solution. The rest OmpF trimer peak in each sample corresponds to ~160, 185, and 195 kDa. **B.** KcsA tetramer inserted into APol at different KcsA:APol ratios were measured by MP. The samples were diluted 200 × before the MP measurement, except for the 1:1 ratio, which was measured without dilution.

APol below 1:2 protein to APol ratios, as observed by UV-VIS measurements, which can be rationalized by partial disassembly of KcsA tetramers and protein aggregation (Figure 3.3). The free APol particles are not observed during this measurement due to the instrument resolution decreasing between the measurements of OmpF APol particles and KcsA APol particles.

3.3.4 Lipid Nanodiscs

The high resolution of MP also enabled us to screen and optimize the conditions for reconstitution of KcsA from *S. lividans* into LNs. After selecting the lipid and MSP types, there are three main variables in the preparation of KcsA embedded nanodiscs: the absolute and relative concentrations of KcsA, MSPs, and the lipids (Figure 3.9A). In the initial screen (condition B in Table 3.2), a KcsA:MSP1D1:DMPC lipid ratio

of 1:4:100 was selected on the basis of an excellent SEC profile (Figure 3.9B, black). The MP profile clearly indicated three different species with masses of 118, 280, and 391 kDa corresponding, respectively, to an empty LN, a LN with two KcsA tetramers, and two fused LNs (Figure 3.9C, black). By contrast, a KcsA:MSP1D1:DMPC lipid ratio of 1:20:800 exhibited a slightly poorer SEC profile (Figure 3.9B, blue), but a much more homogeneous MP profile (Figure 3.9C, blue), with a mass of 161 kDa corresponding to a single KcsA tetramer embedded in a nanodisc consisting of 2 × MSP1D1 proteins and approximately 25 lipids. The latter preparation exhibited $^{15}NH^{4+}$ ion binding upon addition of 50 mM $^{15}NH^{4+}$, as revealed by the presence of 2 strong and 2 very weak NH^{4+} cross peaks in a 2D [^{15}N, ^{1}H]-HMQC NMR spectrum assigned to the 4 ion binding sites in the selectivity filter of KcsA (Figure 3.9D). By contrast, we could not detect any signatures of ion binding in the former sample preparation shown to be highly heterogeneous and lacking KcsA tetramers by MP, demonstrating the lack of functional KcsA (Figure 3.9D, inset).

To further evaluate which of the possible assembly ratios is essential for ensuring assembly and purification of KcsA tetramers, we carried out a screen to determine the optimal IMP to scaffold ratio with a fixed scaffold:lipid ratio of 1:40 at a lipid concentration of 16 mM. We monitored sample absorbance at 280 nm, indicative of KcsA content, for the initial assembly mixture, the assembly mixture after overnight incubation, the washing fraction from batch Ni^{2+}-NTA chromatography, and the elution fraction of purified KcsA LNs (Figure 3.10A).

Characterizing all resulting KcsA MSP LN samples by MP revealed homogeneous samples with a consistent mass around 161 kDa (Figure 3.10B). These results suggest that the IMP:scaffold molar ratio is not crucial for correct sample assembly. Instead, the key parameter controlling the functionality of KcsA embedded in LNs is the

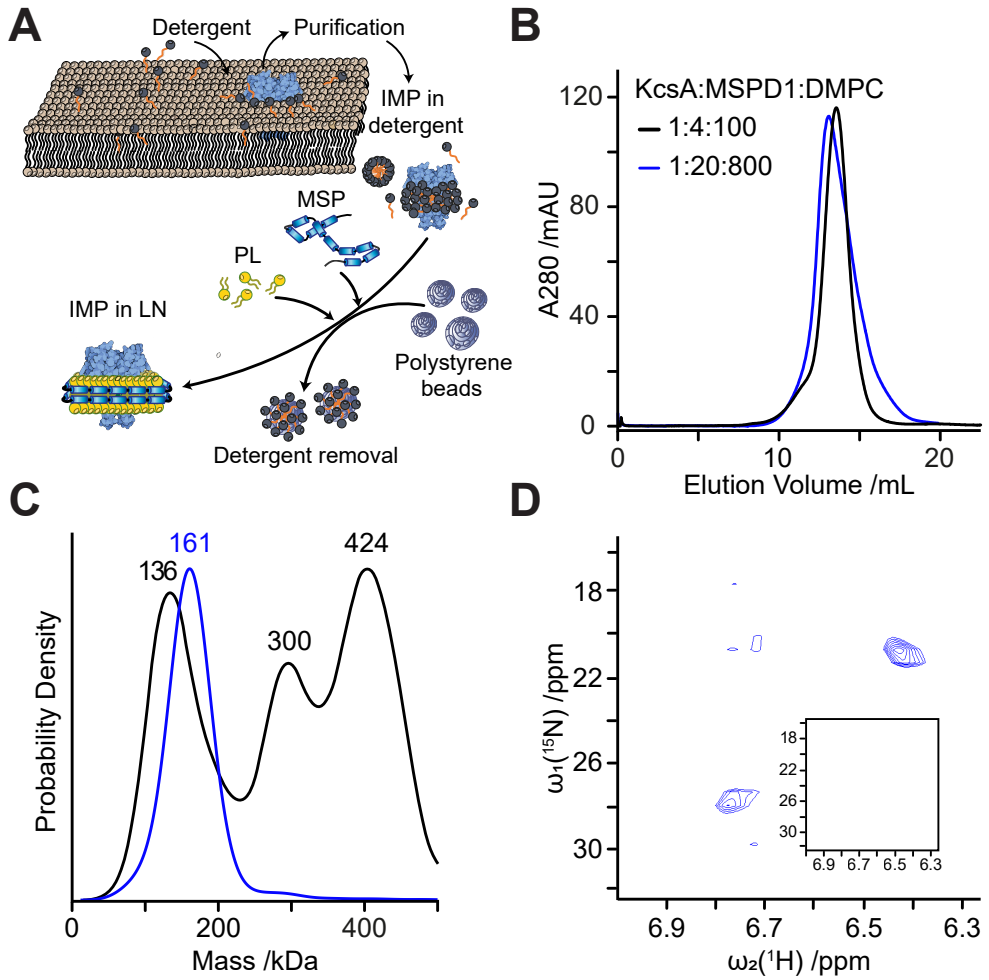


Figure 3.9: KcsA lipid nanodisc characterisation data. A. Schematic showing KcsA solubilization and insertion into lipid nanodiscs. **B.** SEC chromatograms of KcsA MSP LNs assembled at different KcsA:MSP1D1:DMPC ratios. **C.** Corresponding MP analysis. **D.** 2D[^{15}N, ^{1}H]-HMQC spectra revealing NH^{4+} binding by KcsA in native NDs prepared in a 1:20:800 KcsA:MSP1D1:DMPC ratio, but no signal was detected from 1:4:100 ratios demonstrating protein inactivity (inset).

scaffold:lipid ratio, contrary to previously published protocols,[262] highlighting the potential of MP in future LN studies.

3.3.5 Native Nanodiscs

SMA polymers have attracted considerable interest because they can spontaneously solubilize lipid membranes without the need for detergent (Figure 3.11A).[148] In short, SMA polymers solubilize IMPs directly from their native environment, retaining natural lipids that may be important for protein stability, assembly, function,

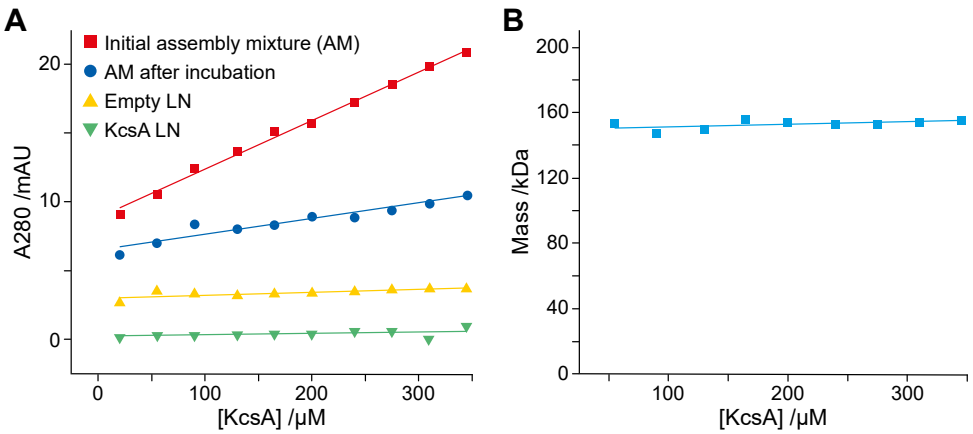


Figure 3.10: KcsA LN assembly screen. A. A series of small-scale assembly mixtures (AM), 50 μL each, of KcsA into MSP LNs was carried out to screen for optimal membrane protein to scaffold ratio. Sample absorbance at 280 nm (A280) was monitored for the initial assembly mixture, assembly mixture after overnight incubation, washing fraction from batch Ni^{2+}-NTA chromatography and elution fraction or purified KcsA LNs. Recorded absorbances are depicted as a function of the input KcsA concentration. **B.** The average mass of the final LN particle distributions was measured with MP. All the test assemblies resulted in a homogeneous sample with a consistent mass, showing that membrane protein to scaffold ratio of up to 1:4.5 does not compromise the sample fitness and therefore is optimal for maximum LN yield per scaffold.

and interactions. The heterogeneity inherent in native membranes, *i.e.*, lipid populations, dynamics, oligomerisation, and proximity of different proteins, may be better reflected in extracted native NDs. This is opposed to MSP LNs, where IMPs are solubilised in detergent prior to assembly with known lipids and protein belts, and where their limited diameter may restrict applicability to larger complexes.

To explore MP's potential for characterizing native ND preparations and aiding in structural and functional studies, we expressed *E. coli* KcsA tetramers (80 kDa) and solubilized the membrane using SMA (Xiran 30010), as outlined in Figure 3.11A. KcsA native NDs were isolated by affinity chromatography and further purified by SEC. Analysis by SDS-PAGE gel (Figure 3.11B) indicated relatively pure protein preparations, in particular in SEC fractions 13 and 14. The SEC profile exhibited two main peaks (Figure 3.11C) and MP analysis of individual fractions taken across both distributions revealed a significant variability in sample heterogeneity

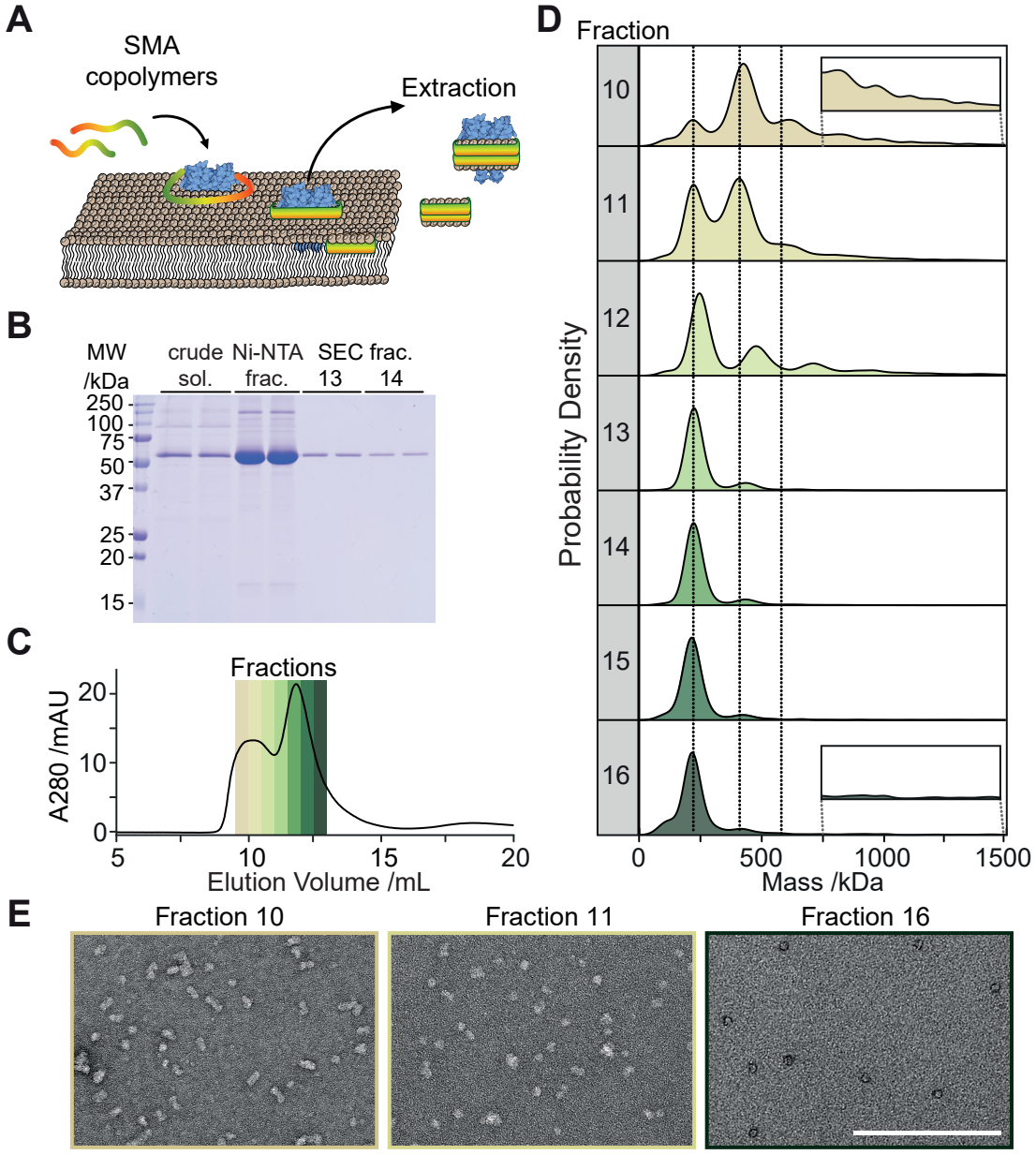


Figure 3.11: Mass photometry of KcsA in native nanodiscs. A. Schematic of native ND formation and extraction of proteins from membranes by SMA. **B.** SDS-PAGE of KcsA native NDs (Xiran 30010 SMA) during purification. **C.** SEC chromatogram of KcsA native NDs, showing the relative protein absorbance at 280 nm (A280). **D.** MP analysis of Xiran 30010 SMA SEC fractions 10–16. Insets in fractions 10 and 16 show magnified view of the high MW histogram sections. **E.** Negative staining EM of fractions 10, 11, and 16 illustrating variability in KcsA native ND assembly. Scale bar, 200 nm.

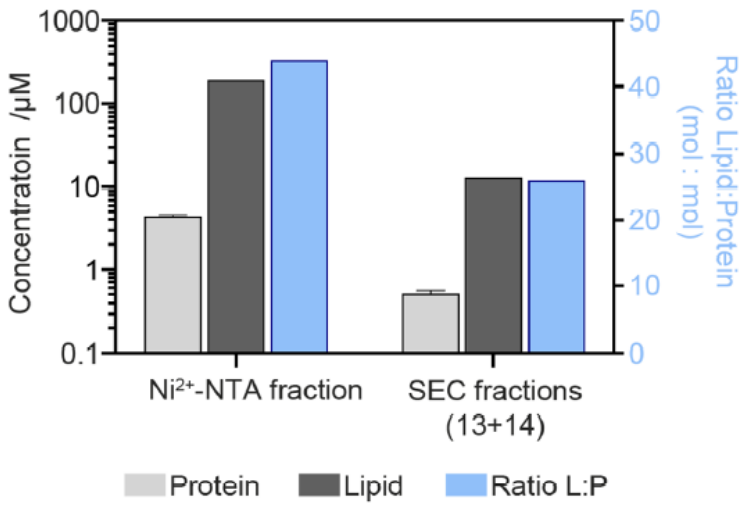


Figure 3.12: Lipid/protein analysis of *E. coli* KcsA native NDs. Experimentally derived protein and lipid concentrations and their respective molar ratios for the Ni^{2+}-NTA fraction compared to combined fractions 13 and 14 from size-exclusion chromatography. Protein concentration is plotted with ± standard error of the mean (SEM).

(Figure 3.11D). Fraction 10 revealed a main peak of ∼400 kDa, with additional peaks at ∼250 and ∼600 kDa, indicating a variation in sample size, which may in part be a result of clustering of native nanodiscs or extraction of higher KcsA oligomers, as well as possible Ni^{2+}-binding contaminants. Successive fractions showed a progressive decrease in the number of species, with the most homogeneous fraction 14 eluting at the apex of the second major SEC peak. Fractions 15 and 16, however, were more heterogeneous with a smaller mass shoulder (∼125 kDa) next to the main peak (∼250 kDa), consistent with minor quantities of empty SMALPs.

To better understand the oligomerisation and composition of the KcsA native NDs, we analysed our preparation's lipid-to-protein ratio. After affinity purification of the His-tagged KcsA in native NDs, we found that ∼44 lipids were present per KcsA tetramer (Figure 3.12). Considering a particle size of ∼250 kDa, and assuming a polymer contribution of ∼30 kDa,[263] the remaining lipid-to-protein ratio would be ∼200 if the SMALPs consisted of a single tetramer/particle, but ∼42 if the SMALPs were composed of clusters of protein, *i.e.* dimers of tetramers

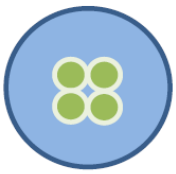
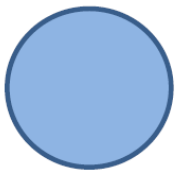
Theoretical assemblies				
~Ratio of lipid-to-protein (mol/mol)	200	42	0	300 [lipids]
Supramolecular organisation	Single tetramer	Dimer of tetramers	Trimer of tetramers	Empty SMALP
Protein (*)	80 kDa (1 mol)	160 kDa (2 mol)	240 kDa (3 mol)	0 kDa (0 mol)
Lipid (**)	140 kDa (195 mol)	60 kDa (84 mol)	0 kDa (0 mol)	220 kDa (306 mol)
Polymer (***)	30 kDa (10 mol)	30 kDa (10 mol)	30 kDa (10 mol)	30 kDa (10 mol)
Total mass (†)	250 kDa	250 kDa	270 kDa	250 kDa
*Protein: KcsA tetramer = 80 kDa. **Lipid: POPE = 0.718 kDa. ***Polymer: M_n = 3 kDa. †Total in one nanodisc.				

Figure 3.13: Theoretical lipid/protein/SMA stoichiometry of KcsA native NDs.

(Figure 3.13). Though the SEC fractions contained only a limited amount of material, we were able to estimate a lipid-protein ratio of ∼20–50 from fractions 13 and 14, consistent with KcsA being present as dimers of tetramers, in agreement with signatures of clustering reported previously.[264–266] Additionally, we qualitatively confirmed the observed transition from heterogeneous to homogeneous sample composition by measuring fractions 10, 11, and 16 with negative stain electron microscopy (Figure 3.11E), which showed larger species in fractions 10 and 11, and a homogeneous distribution of smaller species in fraction 16.

3.4 Discussion and Conclusions

These results demonstrate that MP can accurately characterize membrane-protein carriers and IMPs at the single-particle level in solution. Given its universal applicability, speed, and ease of operation, MP is likely to have a significant impact on in vitro studies of IMPs. While many of the results from both empty and filled amphipathic carriers can be obtained with alternative techniques at the ensemble

level, none can match the speed (60 s per experiment), low sample requirements (< pmoles), and single-particle sensitivity and resolution of MP. For SMALPs, we showed the dramatic difference in sample information as revealed by SEC compared with MP. Such sub-chromatographic resolution characterization of samples during purification holds great promise to improve and accelerate sample preparation for subsequent structural or functional studies. For example, in Figure 3.11, fraction 15 would be immediately suitable for subsequent characterization by cryoEM, providing a route to experimentally test whether detergent and native environments are indeed indistinguishable,[267] although the fact that our SMALPs consist almost exclusively of dimers of tetramers already points toward non-negligible differences between the two. The stark differences between the information content from SEC and MP are further illustrated by our results on KcsA in LNs. Two almost indistinguishable SEC profiles were shown to consist of completely different assemblies, with only one, the one identified by MP, being functional (Figure 3.9).

Despite MP's advantages, limitations remain: (1) MP's concentration range is currently limited to the < 100 nM regime, which represents a challenge for the majority of detergents. This can be addressed by utilising already proven approaches for translating single-molecule techniques from the nM to the μM or even mM regime,[268] or surface passivation approaches. (2) The output of MP is not protein specific. Gaining insight on protein versus lipid mass thus requires some *a priori* knowledge of the protein mass. (3) The masses reported herein are derived from a calibration based on globular polypeptides. Given that lipids have a different polarizability to polypeptides, and one that indeed may also vary depending on their nanoscopic location, such as within a nanodisc compared with contact with solution as in a detergent, our mass determination for lipid-polypeptide constructs are unlikely to be highly accurate. This could be improved if well-defined mass standards—consisting of lipids only, or mixtures of lipids and polypeptides, were

available. As a result, we have no means to strictly evaluate the accuracy of our mass measurements. Nevertheless, we have previously found close correlation between MP-measured and predicted mass changes as a consequence of lipid changes alone,[5] and seen mass accuracy of a polypeptide calibration on the order of 20%, even when compared with DNA, which is structurally completely different to globular proteins,[8] suggesting that the achievable mass accuracy is reasonable. In the future, a more quantitative scale could be provided by obtaining robust standards, or by explicit comparison with native MS measurements, assuming, where possible, that no lipids or weakly associated species are lost during the measurement.

However, our results should not be viewed exclusively in comparison with existing sample characterization techniques. Existing techniques are very powerful and have enabled and continue to yield detailed information on sample composition and heterogeneity. Key to MP is the combination of speed and simplicity, as well its broader capabilities, which include detection of small ligand binding,[5] quantitative evaluation of binding affinities and kinetics in a surface free manner,[7,269] nucleic acid interactions,[8] and more general capabilities for characterizing sample heterogeneity.[6] All of these measurements are performed in the same way as the ones presented herein: by adding small volumes (μL) of low concentration ($< \mu$M) unlabelled samples in a buffer of choice to a microscope cover slide. This combination of speed and ease of use with a broad feature palette of additional capabilities will make MP a powerful, universal method to study IMP structure, function, and interactions.

4

Bacterial Outer Membrane Protein Interactions

Contents

4.1 Introduction **100**
4.2 Materials and Methods **103**
4.2.1 BPA Incorporation and Extraction of the Outer Membrane Fractions 104
4.2.2 OmpF and BtuB Purification 104
4.2.3 SDS and Blue Native PAGE 107
4.2.4 Protein Crosslinking with Glutaraldehyde 108
4.2.5 Mass Photometry Measurements 108
4.3 Results **110**
4.3.1 OmpF and BtuB Analysis 111
4.3.2 OmpF V177^{BPA} Interacts with OM lipids, TBDTs, and BtuB 113
4.3.3 OmpF G198^{BPA} and BtuB are Self-Crosslinked by Glutaraldehyde 119
4.4 Discussion **123**

The work presented in this Chapter forms a part of a submitted manuscript: Melissa N. Webby, Abraham O. Oluwole, Patrick G. Inns, Nicholas G. Housden, Anna Olerinyova, Conrado Pedebos, Nathalie T. Reichmann, Syma Khalid, Philipp Kukura, Shabaz Mohammed, Carol V. Robinson, Colin Kleanthous.

Lipids Mediate Promiscuous Supramolecular Assemblies in the Bacterial Outer Membrane, *submitted.*

BPA OmpF insertion, BtuB and OmpF production, glutaraldehyde crosslinking, amphipol insertion, and SDS- and native PAGE analysis were performed by Melissa N. Webby (University of Oxford). All the mass photometry experiments and analysis were designed and executed by the author. All of the MP figures were created by the author.

4.1 Introduction

Gram-negative bacteria are becoming increasingly resistant to antibiotics, posing a great threat to the health of humans and domesticated animals alike.[3,270] One of the most important factors contributing to the rise of antibiotic resistance is the presence of the outer membrane (OM) in Gram-negative bacteria, which prevents the entry of certain antibiotics that are otherwise active against Gram-positive bacteria, such as the glycopeptide vancomycin.[270,271] The asymmetry of the OM is a direct cause of its impermeability, with the outer leaflet consisting of lipopolysaccharides (LPS).[272] Lipid A is the innermost of the three regions of LPS, and forms a divalent cation-chelated gel-like barrier, whereas phospholipids (PLs), mainly phosphatidyl ethanolamine (PE), phosphatidyl glycerol (PG), and cardiolipin (CL) are the main components of the inner leaflet of the OM (Figure 4.1).[273] In addition to contributing to antibiotic resistance, this combination of OM components also confers their ability to survive hostile environments.

The insertion of a large number of outer membrane proteins (OMPs) in the OM significantly contributes to the protection it provides. In *E. coli*, OMPs are encoded by ∼1.5% of its genome, facilitating the flow of metabolites and small nutrients and the intake of essential molecules, such as metal ions and vitamins.[274] They also

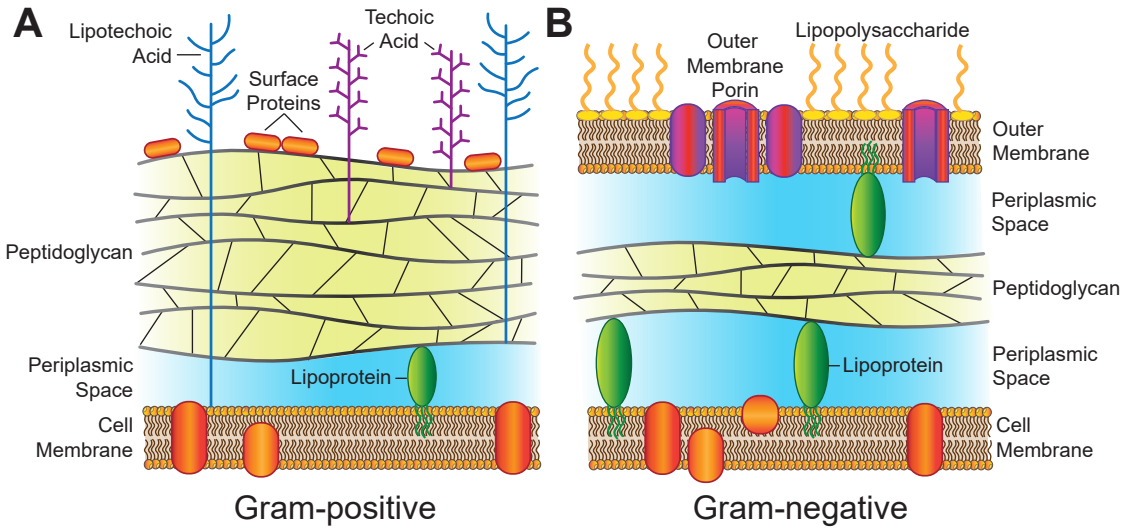


Figure 4.1: Cell wall structure of Gram positive and negative bacteria. A. A schematic view of the cell wall structure of Gram-positive bacteria, showing a thick peptidoglycan layer interspersed with lipoteichoic and techoic acids and surface proteins, on top of a cell membrane containing integral and peripheral membrane proteins, separated by a periplasmic space populated with lipoproteins. **B.** A schematic view of the cell wall structure of Gram-negative bacteria, including a cell membrane containing integral membrane proteins and membrane-associated proteins, lipoproteins in the periplasmic space with a thin layer of peptidoglycan in the centre. An outer membrane sits on the top, with interspersed outer membrane porins and other integral membrane proteins, decorated with lipopolysaccharides on the outer edge.

stabilise and maintain the OM and offer protection by hydrolysing antimicrobial peptides, as well as performing structural functions, such as surface adhesion during pathogenesis and biofilm formation.[25] Even though OMPs serve numerous functions, they are almost all composed of a common β-barrel structure, 8-36 β-strands long.[25,275] Interestingly, both OMPs and LPS are estimated to be similarly abundant in the OM, with $\sim 10^6$ copies per cell.[276,277] They are incorporated into the OM by generally essential β-barrel proteins, with their insertion catalysed by BamA and the lipoproteins BamBCDE for OMPs,[278–280] and LptD and the lipoprotein LptE for LPS.[281]

The long-standing theory of OM organisation has been that it contains evenly distributed and mixed OMPs and LPS, but an increasing number of recent studies

suggested some degree of OM organisation and demixing.[282,283] The limited amounts of diffusion of LPS and OMPs in the OM of *E. coli* means these molecules can be seen as essentially immobile, remaining broadly in their place of membrane insertion.[282] Using high-resolution atomic force microscopy (AFM), the OM has been shown to be phase-separated, with regions or patches that are rich in OMPs or LPS detected.[283] As cells elongate, these LPS patches also experience morphological changes, such as fusion or fission.[282]

Different OMPs also cluster differently in the OM. The trimeric porin OmpF, present in high amounts in *E. coli*, appears to form somewhat hexagonal arrays, whereas some monomeric OMPs (such as the TonB-dependent transports (TBDT), like BtuB and Cir), which are present in lower amounts, seem to form islands within the OM.[282,283] There also appears to be an element of spatio-temporal organisation within the OM, with the new OMPs being inserted at the middle of the cells and the older OMPs sequestering at the poles of the elongating cells.[284] It is important to note that bacteria in general lack methods for OMP extraction and degradation once present in the OM, meaning that cell division, together with binary partitioning of the OMPs, causes rapid turnover of the OMPs in an expanding bacterial population.

Structurally, in Gram-negative bacteria, OMPs are mainly monomers or trimers. In this chapter, we investigated the near-neighbour contacts of the trimeric porin OmpF, composed of three closely positioned 16-stranded β-barrels. Each of the three subunits of OmpF forms a non-specific pore allowing the diffusion of small molecules, with molecular weight (MW) of under $\sim$600 Da, which includes some antibiotics. Together with OmpC, which is its structural homologue, they are among the most numerous OMPs in *E. coli*, representing over 40% of the OM proteome.[276] Apart from its lipid contacts, we focus on OmpF interaction with BtuB, which is

a monomeric 22-stranded β-barrel protein belonging to the proton-motive force-dependent TBDT family, forming a central channel that is occluded with a plug. BtuB occurs in much smaller amounts than OmpF, with under ∼2,000 copies present in a single bacterial cell.[276] Regardless, it serves a critical role, facilitating active transport of vitamin B12, which is a scarce nutrient important for bacterial survival and proliferation.

It remains unknown what forces within the OM drive the differential OMP associations. To investigate this issue at the molecular level, we designed a photoactivatable crosslinker mutants of the OmpF, capable of capturing the OM near-neighbour contacts in *E. coli in vivo*. These complexes are then isolated and defined, with the higher-order complexes of the OMPs replicated *in vitro* and analysed using size-exclusion chromatography (SEC), ion exchange chromatography, sodium dodecyl sulphate (SDS) and blue native (bn) polyacrylamide gel electrophoresis (PAGE), and mass photometry (MP). We are able to show that the lipid asymmetry, contributing to the relative selective impermeability of the OM, is also responsible for the formation of promiscuous homologous and heterologous OMP associations, potentially aiding in the nucleation of the OMP islands present in the OM.

4.2 Materials and Methods

BPA incorporation, extraction of OM fractions, OmpF and BtuB purification and OMP amphipol insertion were kindly performed by Dr. Melissa Webby (Kleanthous Group, University of Oxford). Figures 4.2, 4.6, 4.7, 4.10, 4.12, and the ion exclusion chromatography data presented in Figure 4.5A were created by the author using data provided by Dr. Webby. Some figures were adapted with permission from a manuscript created as a result of this collaborative effort, indicated in the Figure descriptions.

4.2.1 BPA Incorporation and Extraction of the Outer Membrane Fractions

Site-directed mutagenesis was used to modify an OmpF-encoding plasmid to include a TAG amber stop codon at valine (V) 177 and glycine (G) 198, at positions indicated in Figure 4.2, with the BPA crosslinking mechanism outlined in Figure 4.3. A chemically competent cell line BZB1107 (ΔOmpF, ΔOmpC, ΔLamB) was transformed with the protein-encoding plasmid and pVol plasmid. The transformed cells were then grown in Luria Broth (Gibco, Thermo Fisher Scientific) at pH 6.4, containing appropriate antibiotics, 0.5 mM para-benzoylphenylalanine (BPA) dissolved in 0.8 M NaOH, with 0.8 M HCl maintaining the pH, until the optical density at 600 nm (OD_{600}) reached 0.5-0.7. To induce protein expression, 0.15% arabinose was added, and the cells were incubated at 37 °C for 2 h. BtuB (wild-type, WT) was expressed using the same protocol, with minor alterations. The cell line RK5016 (*metE70 argH BtuB recA*) was used, and the transformed cells were incubated at 18 °C overnight after the induction of protein expression. The cells were harvested through centrifugation (4,200 rpm, 20 °C, 12 min) and the cell pellet resuspended in 2 mL of lysis buffer (10 mM Tris-HCl, pH 8.0, 0.25% lithium 3,5-diiodosalicylic acid), with the addition of 1 mM phenylmethylsulfonyl fluoride. Following resuspension, 5-8 mL of cells were placed in a 10 × 10 cm petri dish on ice, in the CL-1000 Ultraviolet Crosslinker, and exposed to UV light (395 nm, 90 min, mixed at 45 min). The OM fractions containing OmpF or BtuB were then extracted according to a protocol described previously.[285]

4.2.2 OmpF and BtuB Purification

The OMP of interest was extracted in the solubilised OM fraction following a series of ultra-centrifugation steps to facilitate protein solubilisation in detergent (1% n-octyl-β-D-glucopyranoside (OG), Anatrace), after which it was syringe filtered with

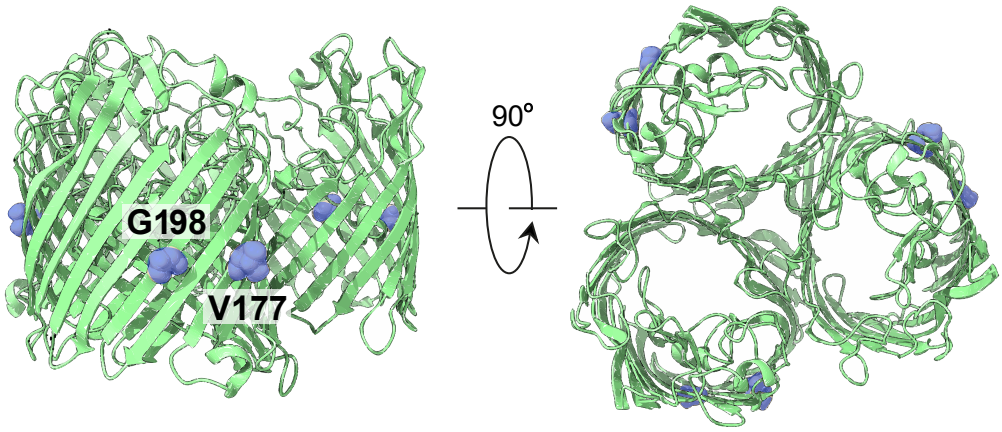


Figure 4.2: OmpF BPA incorporation sites selected for MP measurements. The two sites for BPA incorporation in OmpF (PDB: 3K19), valine (V) 177 and glycine (G) 198, are located approximately in the centre of the membrane-facing side of the β-barrels. The 90° rotation highlights the location of the sites relative to the trimer interface. Adapted with permission from M. N. Webby *et al.*, Lipids Mediate Promiscuous Supramolecular Assemblies in the Bacterial Outer Membrane, 2022, *submitted*.[286]

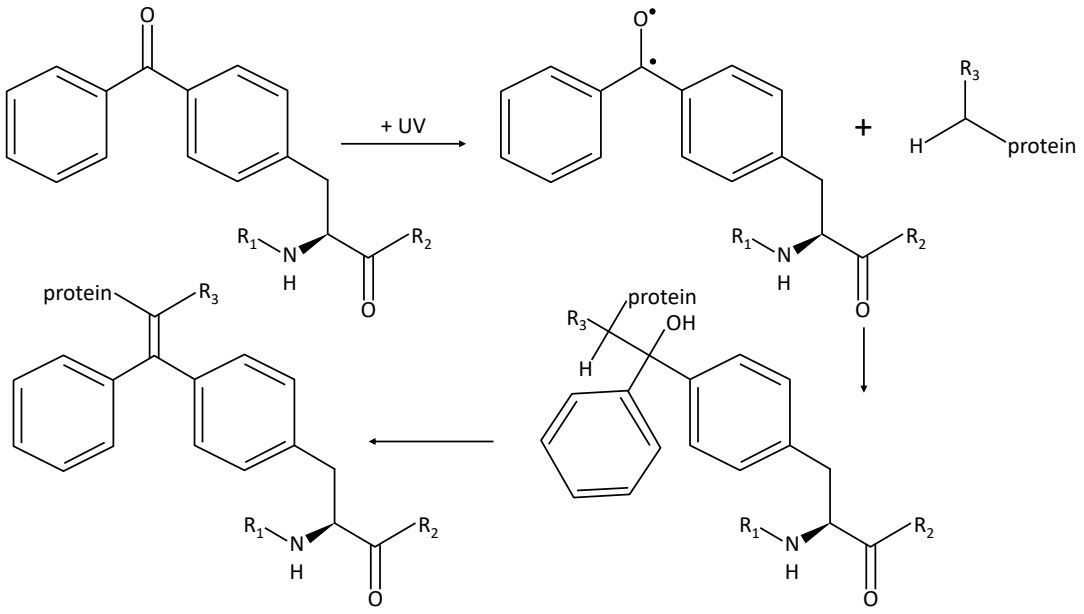


Figure 4.3: The structure and mechanism of BPA crosslinking. The chemical structure of para-benzoylphenylalanine (BPA, left) incorporated as an amino acid at a specific position in a amino acid chain, with the rest of the BPA-modified protein marked as R_1 and R_2, respectively. Upon UV light exposure, a radical is formed that subsequently reacts with a second protein, marked by 'protein' and R_3. The resulting crosslink undergoes a water elimination reaction to form a stable chemical crosslink, spanning approximately 4 Å.

a 0.45 μm filter. The OM fraction containing OmpF was then injected onto a new Q Sepharose resin equilibrated in SEC buffer (10 mM Tris-HCl, pH 8.0, 5 mM EDTA, 1% OG), whilst the BtuB OM fraction was applied to a new diethylaminoethanol (DEAE) Sepharose resin in a XK 16/20 column, which had been equilibrated with a 10% high salt buffer (50 mM Tris-HCl, pH 7.5, 5 mM EDTA, 0.54% OG, 1 M LiCl) and 90% low salt buffer (50 mM Tris-HCl, pH 7.5, 5 mM EDTA, 0.54% OG).

For OmpF, the loaded protein was washed with 2 column volumes (CV, ∼20 mL) and eluted with a linear salt gradient (0-500 mM LiCl) across 3.6 CV. The sample was then applied to a MonoQ 4.6/100 PE column equilibrated in the SEC buffer, facilitating ion exchange. The injected sample was washed with 4 CV of buffer, and eluted using a 0-400 mL LiCl linear salt gradient over 14 CV. OMP identity in the individual peaks was confirmed using SDS-PAGE, together with in-gel digestion mass spectrometry peptide fingerprint analysis. The peaks were pooled separately, concentrated with a 50 kDa MW cut-off filter (Vivaspin), and desalted using a HiTrap 5 mL desalting column. The BtuB sample, loaded onto the equilibrated column, was washed with 10% high salt buffer until a stable baseline was achieved (∼2 CV), with any other contaminants washed out with ∼3.5 CV of a linear gradient of 10-50% high salt buffer. Following stepping to 100% high salt buffer, BtuB was eluted into a 96 well plate in 1 mL fractions. The protein was then precipitated with an overnight incubation at 4 °C, pelleted by centrifugation (12,000 × g, 4 °C, 15 min), and resuspended in the SEC buffer. The sample was then applied to a SEC buffer equilibrated Superdex 200 16/60 column, and fractions containing BtuB were pooled. Finally, MonoQ 4.6/100 PE column was used for ion exchange, and HiTrap 5 mL desalting column was used for desalting, applied as required.

4.2.3 SDS and Blue Native PAGE

Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE) was used to confirm the MW of the proteins present in the samples and fractions obtained during protein purification and protein combination and crosslinking experiments. 4 × loading buffer (200 mM Tris-HCl, pH 6.8, 8% weight/volume (w/v) SDS, 0.4% w/v bromophenol blue, 40% volume/volume (v/v) glycerol, 400 mM 2-mercaptoethanol) was added to the samples to result in 1 × loading buffer. The combined samples were then heated for 3 min at 98 °C, cooled, and loaded into the wells of a 10% bis-acrylamide gel, inserted in the SE250 Mighty Small II Mini Vertical Protein Electrophoresis Unit (Hoefer). The electrophoresis was performed in the running buffer (25 mM Tris-HCl, 192 mM glycine, 3.5 mM SDS) using 30 mA constant current, for 35 min. Subsequently, the gels were stained with the coomassie stain buffer (10% acetic acid, 50% ethanol, 0.2% Coomassie Brilliant Blue R-250 Dye) for 20-40 min, followed by destaining with a mixture of 10% acetic acid and 10% ethanol.

Blue native (bn) PAGE was performed to visualise the formation of non-covalent complexes between OmpF crosslinked to LPS, FepA or FhuA, and between OmpF crosslinked to PLs mixed with BtuB. For each sample or fraction, 9 μL of $\sim$30 μM protein in 1% OG was combined with 0.45 μL of the NativePAGE® 5% G-250 Sample Additive and 3 μL of the sample loading dye, according to the manufacturer's protocol. For the crosslinking experiments, NativePAGE® Sample Buffer (4 ×) was added to 12 μL final volume. 12 μL of the prepared sample was then loaded onto a NativePAGE™ 4-16% Bis-Tris Mini Protein Gel (Novex Life Technologies) and placed in the XCell SureLock® Mini-Cell gel running tank, filled with 1 × NativePAGE™ Running Buffer for the anode, and a 1 × NativePAGE™ Cathode Additive added to the Running Buffer for the cathode. As a MW indicator, 10 μL

of NativeMark^{TM} Unstained Protein Standard was loaded into the first well. The gels were run at 150 V for 1 h and 250 V for another hour, at 4 °C, after which they were stained in the coomasie stain buffer (1 h), and destained as described above.

4.2.4 Protein Crosslinking with Glutaraldehyde

The OmpF G198^{BPA} and BtuB WT protein samples were diluted to 30 μM in crosslinking buffer (20 mM potassium phosphate, 100 mM NaCl, 1% OG), and mixed at 2:1 ratio. The glutaraldehyde crosslinking was performed on ice, with the final glutaraldehyde concentration of 0.5%. The schematic of glutaraldehyde crosslinking is shown in Figure 4.4. After thorough mixing, the sample was incubated on ice for 30 min, followed by reaction quenching with 1 M Tris-HCl pH 8.0, mixed with the sample to a final concentration of ∼100 mM. The sample was then spun in a table-top centrifuge at 20,000 × g for 5 min at room temperature, and the supernatant loaded onto a Superose 6 increase 10/300 column, equilibrated in the crosslinking buffer. The column was then run at 0.5 mL/min, with 0.5 mL eluted fractions collected.

4.2.5 Mass Photometry Measurements

To facilitate complex formation, OmpF V177^{BPA} was mixed with BtuB in SEC buffer at 2:1 molar ratio and incubated overnight at 4 °C. The mixed sample, as well as individual proteins, were transferred into amphipol (APol) for MP analysis, according to a previously described protocol.[287] After APol insertion from detergent, the samples were injected onto a Superose 6 increase 10/300 column to allow for protein separation, and 0.25 mL eluted fractions were collected. The approximate protein concentration across the fractions of the individual proteins and mixed sample was estimated using the 280 nm absorbance during the elution of fractions and calculated extinction coefficients, based on the sequence of the proteins.

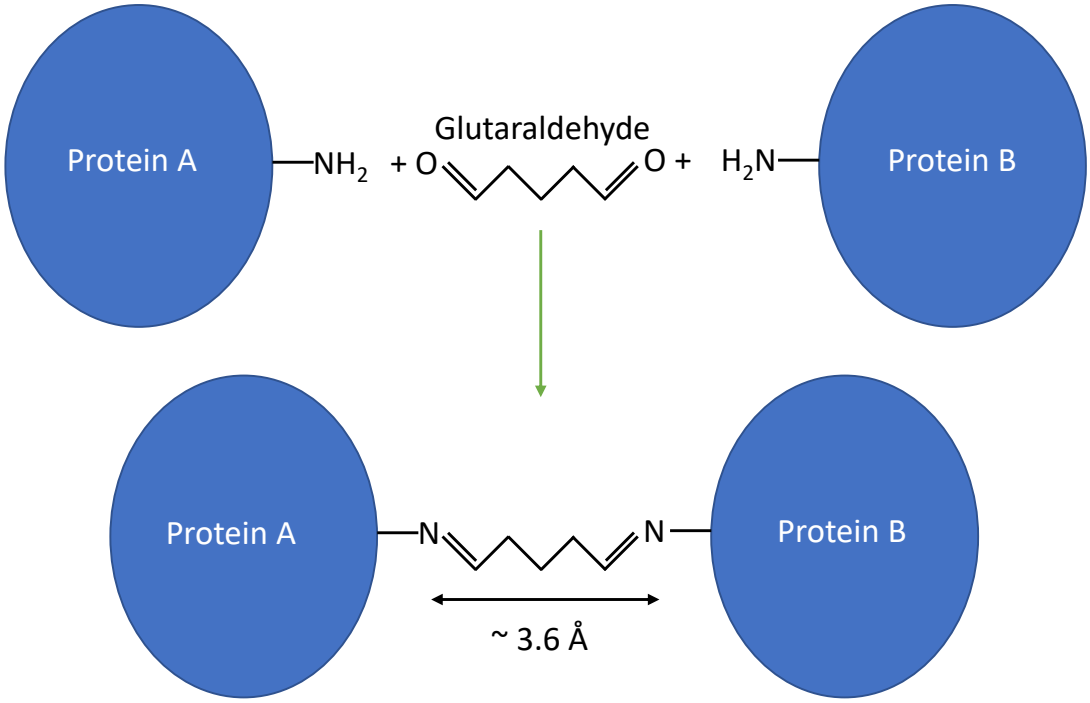


Figure 4.4: The structure and mechanism of glutaraldehyde crosslinking. The chemical structure of glutaraldehyde as it reacts with two amino groups on nearby proteins. The reaction forms a stable chemical crosslink, spanning approximately 4 Å.

For mass photometry (MP) measurements, the cleaning of the coverslips (No. 1.5, 24 × 50, VWR) was facilitated through sonication in a 1:1 mix of 2-propanol (HPLC grade) and Milli-Q H_2O (MQ), followed by sonication in MQ, each for ∼5 min. The cleaned coverslips were oven-dried (110 °C, 1 hour), stored in a sealed box, and used within 1 week. 2 × 2 microwell gaskets (Grace Bio-Labs reusable CultureWell^{TM} gaskets, well capacity of 3-10 μL) were placed on a clean coverslip and on the objective of the MP instrument. Prior to each measurement, 3 μL of MP buffer (10 mM Tris-HCl, pH 8.0, 5 mM EDTA) was inserted into the gasket to identify the focus position, which was then automatically secured for the duration of the measurement. 27 μL of freshly diluted sample (∼30-60 nM) was then added to the gasket, and results recorded for 60 seconds with a 4 × 4 pixel binning and 5-fold frame binning applied. The recording was performed using a commercial mass photometer (One^{MP}, Refeyn Ltd, UK), or a similar mass photometer built within the

Kukura lab, described previously.[5] The applied binning resulted in a 70.2 nm/pixel or a 84.4 nm/pixel size for the commercial and home-built instruments, respectively, and a 200 Hz final frame rate. Each experiment was done in a separate gasket, with at least triplicate repeats performed in different gasket wells.

The recordings were analysed with Discover^{MP} (version 2.3, Refeyn Ltd), as described previously in Section 3.2, with the averaged number of frames used for background removal set to 5. The list of fitted particles were exported, and data analysis and fitting performed with MATLAB (version 2020b), using the curve fitting toolbox, which fits a specified number of Gaussians to the data with gradient descent. MW calibration and the MW conversion of the interferometric scattering contrast values of detected particles was performed using an internal standard, MS1000, of known MW and oligomeric distribution. MS1000 was measured at the start and at the end of the experiments on a given day, and analysed as above. The centres of the histogram peaks were then fitted by the Gaussian fitting function embedded within Discover^{MP}, the contrasts were plotted against the MS1000 MWs, and the data points fitted to a linear function $y = bx$, where y is the particle contrast, x is the corresponding MW, and b is the instrument calibration factor.

4.3 Results

In order to investigate close associations under ∼4 Å of complex OMP communities within the OM, a protocol for a photoactivatable crosslinking was developed and established. The incorporation of BPA at two membrane-facing, transmembrane (TM) region amino acids of the OmpF β-barrels facilitated their crosslinking to C-H bonds upon UV light exposure. An *E. coli* strain, lacking the *ompf* gene, was transformed with an *ompf* gene with a single BPA site inserted, and grown with BPA present in the growth medium. To test whether the inserted BPA-containing OmpF

was functional, colicin N-mediated killing ability, usually enabled by OmpF, was evaluated. Both V177^{BPA} and G198^{BPA} OmpF mutant-expressing *E. coli* cells were found to be susceptible to colicin N toxicity to the same extent as WT, suggesting correct OmpF folding, OM insertion, and function. The mutant OmpF expressing cells were subsequently exposed to UV light of 365 nm wavelength for ∼90 min. The OMs of the UV-treated cells were then solubilised using 1% OG, with OmpF BPA mutants extracted, purified, and analysed.

4.3.1 OmpF and BtuB Analysis

The protein-protein and protein-lipid interactions within the OM were tested using the OmpF WT and the V177^{BPA} mutant, as well as the monomeric β-barrel protein BtuB. The proteins were purified in 1% OG and, in the case of OmpF, further separated into fractions using ion exchange chromatography, producing spectra containing a main peak 1 and peak 2. The second peak was strongly pronounced in the V177^{BPA} mutant, likely as a result of BPA-crosslinked LPS molecules altering the surface charge of the complex (Figure 4.5A). SDS-PAGE and lipid-binding analysis of peak 1 and 2 from the V177 and G198 BPA mutants indicated equivalent lipid and protein binding, with the main differences between the separate peaks rather than the mutants.[286] As such, whilst the use of each specific mutant is indicated throughout this chapter, they are treated as functionally equivalent.

The separated peaks from the OmpF WT and V177^{BPA} mutant were collected, transferred to APol, and peak 1 was analysed using mass photometry (Figure 4.5B and C for the WT and V177^{BPA}, respectively). The OmpF WT sample produces a narrow peak at a lower MW than that of V177^{BPA} peak 1, with the mean fitted mass of ∼179 kDa, and standard deviation of 15 kDa. This corresponds to an OmpF trimer with an added APol mass of ∼68 kDa, equivalent to about 15 APol molecules, which is in good agreement with previously determined values.[16] There is also a

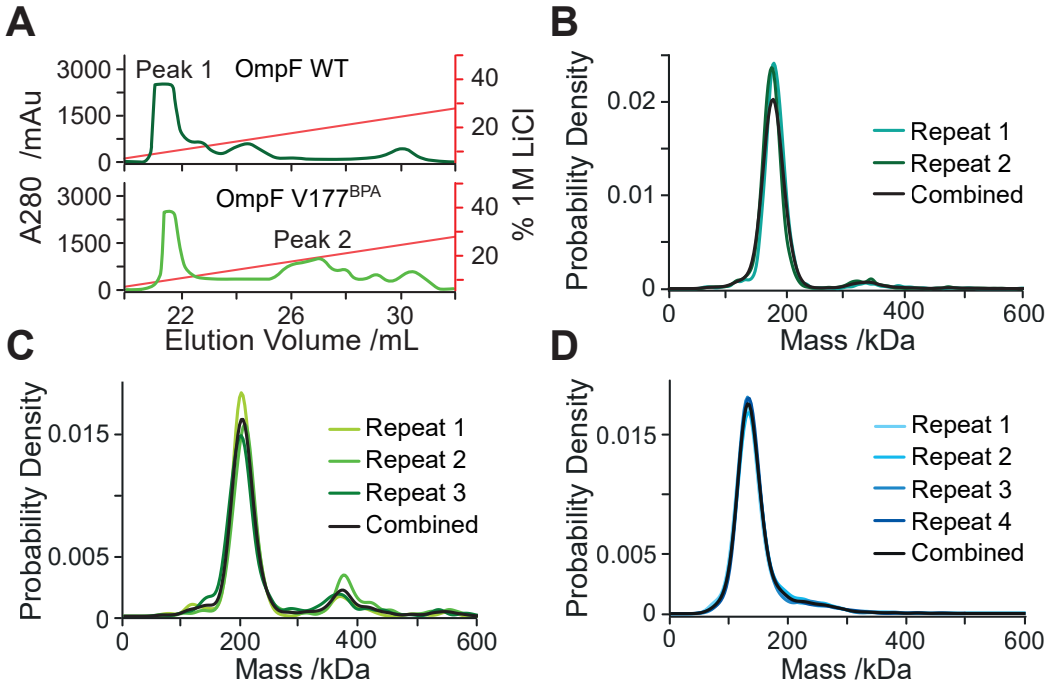


Figure 4.5: Ion exchange and MP analysis of OmpF and BtuB. A. Ion exchange chromatography traces of OmpF WT (top) and V177^{BPA} (bottom), eluted over LiCl gradient. The traces show the presence of peak 1, composed of predominantly OmpF trimers, and peak 2, which is much more prominent in the mutant sample due to the presence of the crosslinked LPS aiding in co-purification and interactions with other OMPs. **B.** Mass photometry (MP) analysis of OmpF WT peak 2, plotted as kernel density estimate (KDE) functions. The majority of the protein is present in the form of a trimer. **C.** MP analysis of the OmpF V177^{BPA} mutant (peak 1), composed of three repeats and plotted as KDE functions. The dimer of trimers peaks is pronounced in comparison to OmpF WT. **D.** MP analysis of BtuB WT, composed of four repeats. The protein is almost entirely present as a monomer.

small amount of dimer of OmpF trimers visible in the WT data, suggesting weak self-association between OmpF trimers. In contrast to OmpF WT, the V177^{BPA} peak 1 sample displays a much higher degree of oligomerisation or complex formation, as evidenced by the increased size of the dimer of trimers peak (373 kDa), as well as the presence of a small peak at ∼550 kDa. The main OmpF trimer peak is also shifted to a higher MW, by over 20 kDa (202 kDa), suggesting crosslinking between the V177^{BPA} group on the OmpF β-barrels and the OM phospholipids (PLs). MP analysis of BtuB WT transferred to APol from OG following purification is shown in Figure 4.5D. It forms a large peak at ∼132 kDa, with very limited amounts of

protein dimerisation present, similar in frequency to that of the OmpF WT sample.

4.3.2 OmpF V177^{BPA} Interacts with OM lipids, TBDTs, and BtuB

To test the effects of the BPA insertion, OmpF V177^{BPA} mutant was crosslinked *in vivo* in the OM through UV light exposure. We could identify crosslink formation between OmpF, PLs and LPS, eluting separately in peak 1 and peak 2 during ion exchange chromatography (Figure 4.5A, bottom). These interactions were then confirmed through SDS-PAGE, PL analysis and LPS staining.[286] Due to its large size, LPS crosslinked to OmpF can be identified as a separate band in SDS-PAGE analysis, located directly above the non-crosslinked OmpF band (Figure 4.6A, top). Crosslinked LPS also appears to significantly increase the quantity of TBDTs co-purified with OmpF in peak 2, compared to peak 1 or the control sample without UV exposure, although this is significantly confounded by comparatively stronger dye absorbance and staining of these TBDTs. Additionally, small amounts of preserved LPS-crosslinked OmpF-FepA/FhuA complex can be identified through bnPAGE (Figure 4.6A, bottom), absent in the control or peak 1 sample.

SDS and bnPAGE analysis was then applied to probe the *in vitro* complex formation between OmpF and another monomeric β-barrel protein, BtuB, by mixing the OmpF V177^{BPA} peak 1 sample, crosslinked *in vivo* as above, with BtuB WT at 1:1.5 molar ratio. As can be seen *via* SDS-PAGE analysis (Figure 4.6B, top), whilst the PL-crosslinked (peak 1) OmpF sample co-purifies with FepA, no higher order complexes of this interaction can be identified by bnPAGE (Figure 4.6B, bottom left), although it is important to remember that the aforementioned strong dye absorbance of FepA exaggerates the apparent quantities of the protein present in the sample, potentially contributing to the lack of visible complex in the bnPAGE.

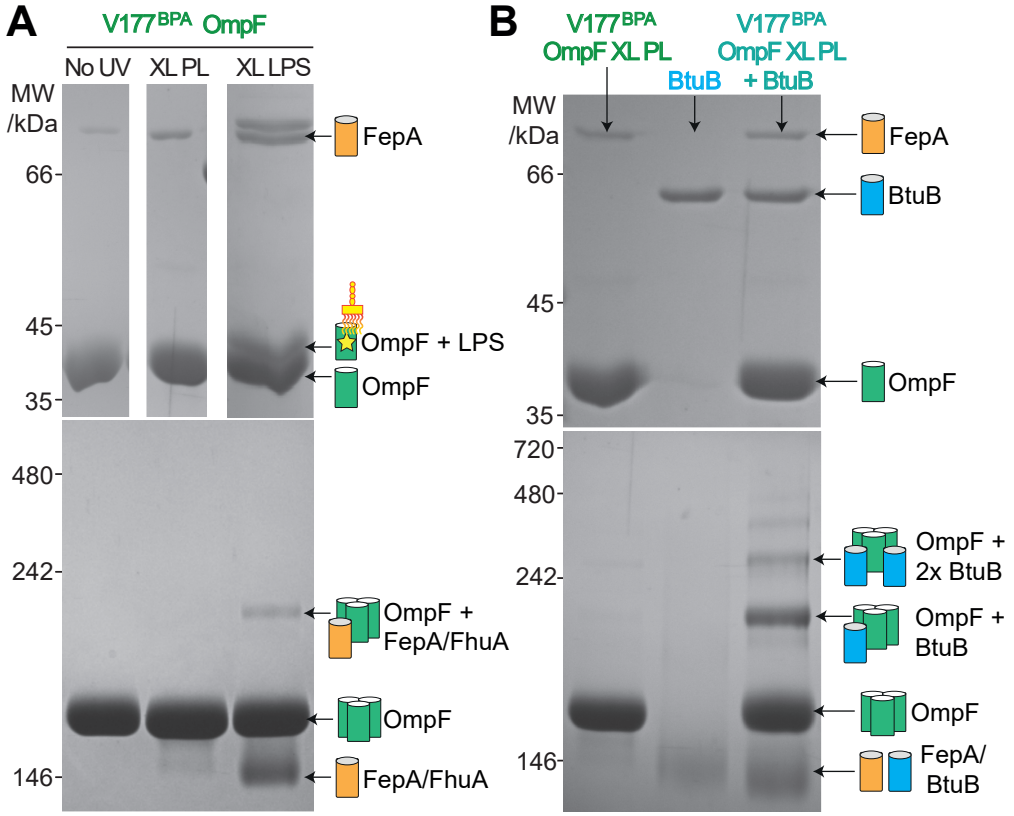


Figure 4.6: SDS and bnPAGE analysis of OmpF V177^{BPA}, interacting with TBDTs, OM lipids, and BtuB. A. Denaturing SDS-PAGE (top) and bnPAGE (bottom) of UV-activated OmpF V177^{BPA} sample extracted from the outer membrane of *E. coli*, separated by ion exchange into peak 1 (crosslinked (XL) to phospholipids (PL)) and peak 2 (crosslinked (XL) to LPS), showing promiscuous complex formation between LPS-crosslinked OmpF the co-purified TonB-dependent transporters (TBDTs) FepA and FhuA. The sequences of both TBDTs in the bnPAGE OmpF-FepA/FhuA complex band were also identified by in-gel digestion mass spectrometry peptide mass fingerprinting analysis.[286] **B.** 15 μM of purified BtuB WT was mixed with 10 μM of OmpF V177^{BPA} peak 1 sample and analysed by denaturing SDS-PAGE (top) and bnPAGE (bottom). BtuB addition, together with one molecule of LPS bound to BtuB, causes the appearance of higher molecular weight complexes between OmpF and BtuB, which are absent in the BtuB and OmpF V177^{BPA} controls.

In comparison, the addition of 15 μM BtuB to 10 μM (trimeric) OmpF results in the formation of numerous higher MW species, visualised through bnPAGE in Figure 4.6B, bottom right. These can be identified as OmpF-BtuB complexes, partly due to their absence in the OmpF V177^{BPA} PL-crosslinked (peak 1) and BtuB WT controls (Figure 4.5B and D), as well as through in-gel protein digestion and MS peptide mass fingerprinting analysis.[286] These results strongly suggest

that *in vitro* mixing of PL-crosslinked OmpF V177^{BPA} with BtuB facilitates the formation of higher oligomeric complexes between the two proteins, likely through the interactions between the crosslinked PLs.

We then further characterised these heterologous OmpF-BtuB complexes with MP. To test the complex formation, the mixture was prepared as above and exchanged into APol and separated into fractions using SEC, which were then measured with MP. The 280 nm absorbance (A280) of the fractions, corresponding approximately to their protein content, is shown for OmpF V177^{BPA}, BtuB WT, and their 1:1.5 mix in Figure 4.7A, top. The mixed trace appears to be almost completely matched to the individual proteins' traces, suggesting little interaction between the two proteins. A magnified view of the OmpF-BtuB SEC trace is shown in the bottom panel of Figure 4.7A, with the fractions chosen for further analysis by MP highlighted in turquoise. However, the SDS-PAGE of selected OmpF-BtuB SEC fractions shows varying proportions of both BtuB and OmpF across the fractions (Figure 4.7B), indicating failed protein separation during gel filtration based on their different MW, as would be expected for non-interacting proteins.

Measuring the SEC fractions of the OmpF-BtuB mixed sample with MP (Figure 4.8) shows that whilst BtuB and OmpF remain the dominant peaks in most of the fractions, there also appears to be a large number of high MW particles forming peaks in the 300-700 kDa range. These are notably absent in the BtuB WT control spectra (Figure 4.5D), and different in appearance as well as stronger than the OmpF V177^{BPA} peak 1 dimers of trimers peak (Figure 4.5C). The proportion of these higher MW complex peaks is especially more prominent in fractions D9-D11, with a notable shift to a lower MW apparent in fractions D11 and D12. These fractions are shown in Figure 4.9, overlayed with the control OmpF V177^{BPA} and BtuB WT distributions. The magnified view of the 300-700 kDa region shows

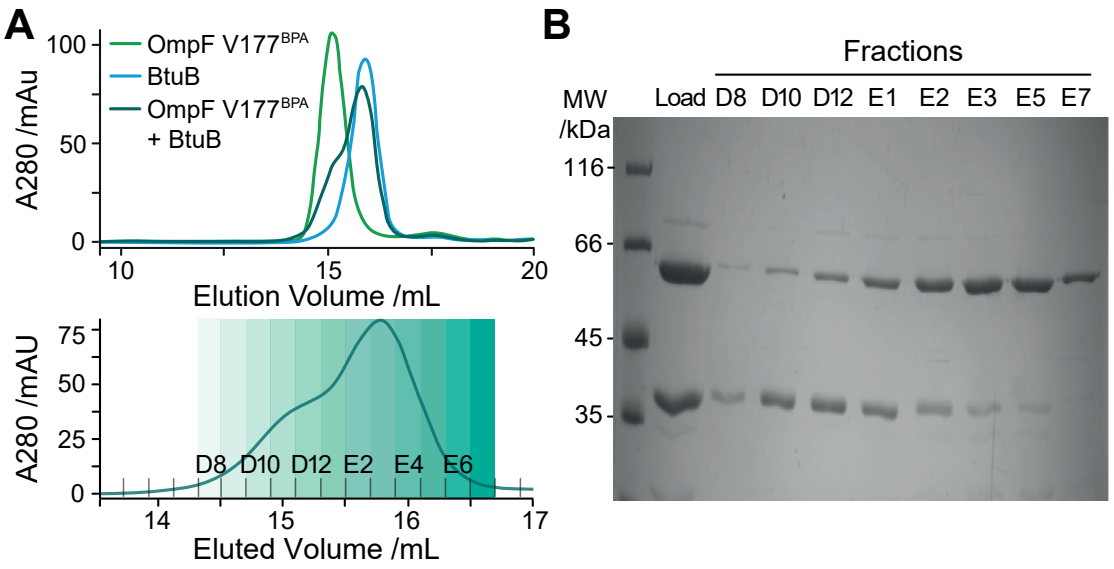


Figure 4.7: SEC and SDS-PAGE of OmpF V177^{BPA} mixed with BtuB. A. Overlayed gel filtration chromatographs showing the sample absorbance at 280 nm (A280) of OmpF V177^{BPA}, BtuB WT, and OmpF V177^{BPA} mixed with BtuB WT at 2:1 ratio in 1% OG. **B.** SEC A280 trace of OmpF V177^{BPA} peak 1 mixed with BtuB WT. Highlighted in shades of purple are the fractions selected to measure by mass photometry. **C.** SDS-PAGE of some of the selected fractions, showing varying presence of both proteins, compared to the overall column load, showed on the left of the gel.

OmpF forming oligomeric complexes with BtuB. The approximate MW of the identified peaks broadly agrees with an OmpF trimer binding up to three BtuB monomers. Given the presence of the BPA mutation on every OmpF monomer potentially crosslinking to molecules within 4 Å *in vivo*, the presence of these high-order oligomers seems feasible, and agrees with the bnPAGE results presented in Figure 4.9B, bottom.

However, since the peak 1 sample of the OmpF V177^{BPA} mutant was found to display increased frequency of self-association resulting in more dimers of trimers, measured as a taller peak at ∼375 kDa visible in Figure 4.5C, and OmpF complexed to two BtuB monomers would appear at ∼400 kDa, there is the possibility of an overlap between the two peaks. Unfortunately, the current limits of MP resolution make it difficult to distinguish between the two peaks. In order to approximate what proportion of the 400 kDa peak can be assigned to 2 × OmpF, we can compare the

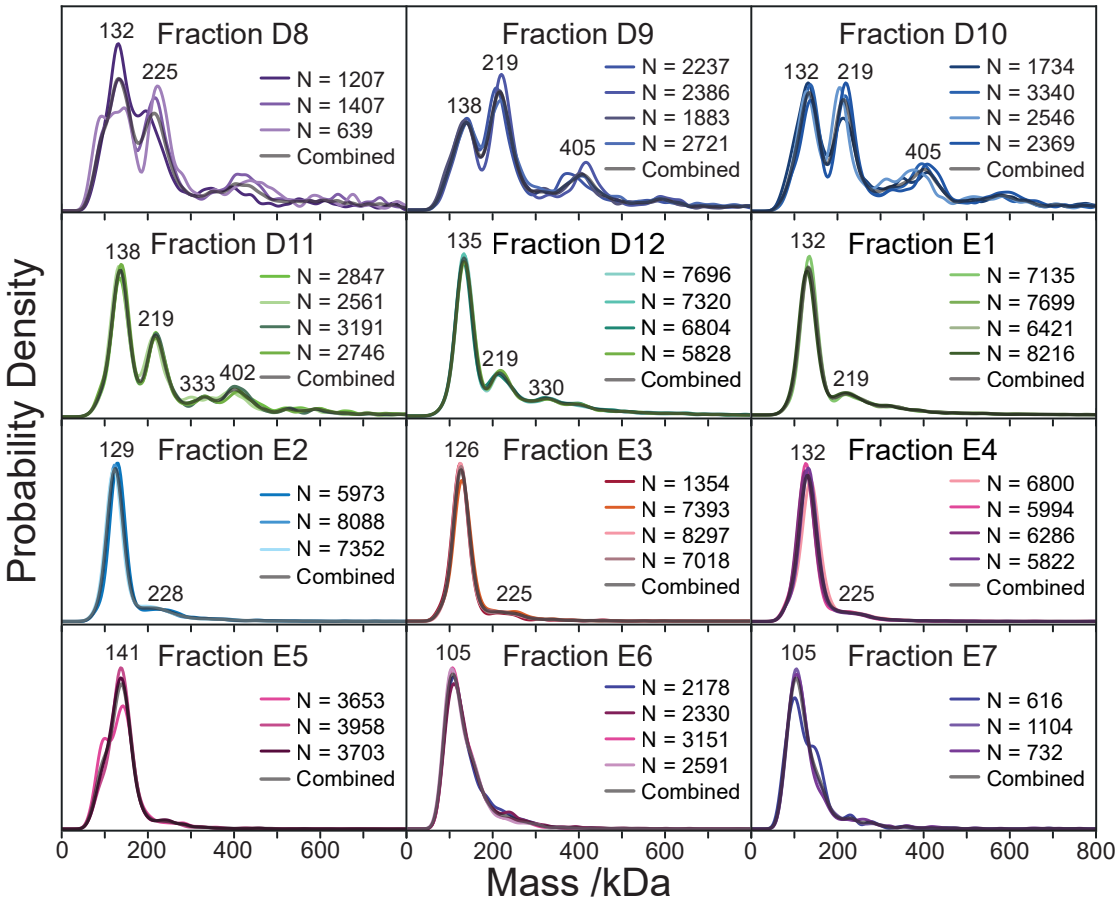


Figure 4.8: Mass photometry of SEC fractions of OmpF V177^{BPA} mixed with BtuB. Mass photometry measurements of SEC fractions of BtuB WT mixed with OmpF V177^{BPA} at 1:1.5 ratio in 1% OG. Shown are the KDE of the individual and combined data of at least 3 repeats for each fraction D8-E7, with the number of particles detected in each repeat indicated in the legend, and the measured MW above each peak, in kDa. The combined data is shown by the grey line in each plot.

OmpF dimer of trimer counts between the individual measurements.

Focusing on the fractions displaying a higher percentage of the apparent complex, we can determine the approximate concentrations of the unbound OmpF trimer by comparing the number of particles in the main OmpF trimer peak in these fractions with the OmpF V177^{BPA} peak 1 control. Since the concentration of OmpF used for mixing with BtuB and APol exchange was broadly the same as the control OmpF concentration during its APol transfer, and thanks to the slow off-rate between APol molecules and the TM region of membrane proteins, the amounts of OmpF dimer

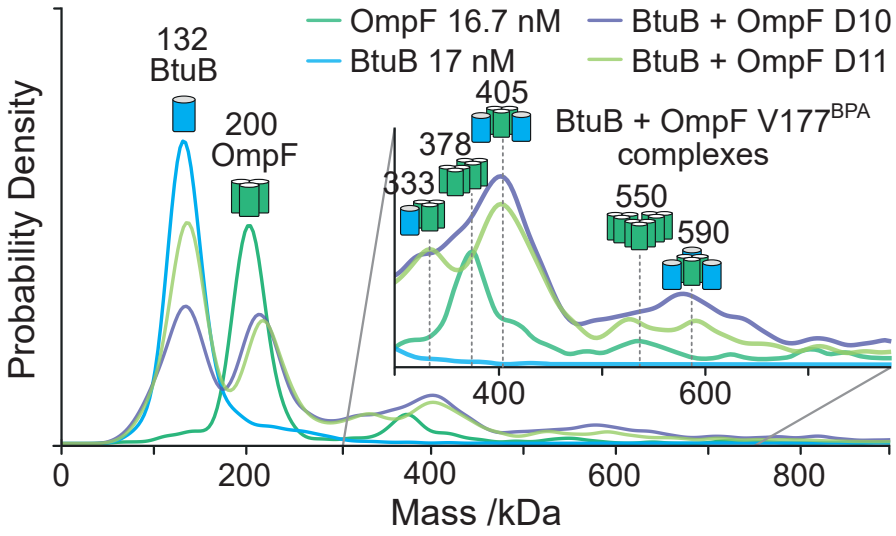


Figure 4.9: Mass photometry of selected OmpF V177^{BPA} + BtuB fractions and controls. KDE plots of fractions D10 and D11 of the OmpF V177^{BPA}:BtuB mix compared to the BtuB WT and OmpF V177^{BPA} peak 1 controls, measured in amphipol. The inset magnifying the higher molecular weight section (300-700 kDa) shows the formation of several higher MW species, representing heterologous OmpF-BtuB complexes and small amounts of OmpF dimers of trimers.

of trimers formed in the mixture can be expected to be proportionately similar to the amounts measured in the OmpF control. This allows us to estimate the expected number of dimers of trimers in the mix, which can then be subtracted from the total counts of the 400 kDa peak. Applying this analysis to the data from fractions D9-D12, we can calculate that the OmpF + 2 × BtuB complex contributes approximately 71.3 ± 4.3% of the particles detected in this peak, ranging from 59.6 to 78.8% across the fractions. Together with the different MW position of the two peaks, we can conclude that whilst OmpF V177^{BPA} dimers of trimers do contribute to some of the higher MW complexes observed, they form a small proportion, with the majority of the particles detected coming from the heterologous OmpF-BtuB complexes.

4.3.3 OmpF G198^{BPA} and BtuB are Self-Crosslinked by Glutaraldehyde

The OmpF-BtuB interaction was further probed by chemical crosslinking, in order to capture and stabilise the complexes formed. The PL-crosslinked peak 1 OmpF G198^{BPA} mutant and BtuB WT were purified separately in 1% OG and mixed at 2:1 molar ratio in the presence of a crosslinking agent, glutaraldehyde. After a 30 min incubation period, the mix was transferred to APol and separated through gel filtration, with the A280 traces of the mixture, BtuB, and OmpF G198^{BPA} peak 1 controls shown in Figure 4.12A, top. The mixed peak is magnified below, with the fractions collected and analysed by MP highlighted in turquoise. Figure 4.12B shows the SDS-PAGE of the collected fractions, as well as the crosslinking time checks, where the mixture containing glutaraldehyde was quenched after 5 and 30 min of crosslinking. Crosslinking for 30 min appeared to cause a strong increase in the presence of large MW species of varying sizes, absent in the BtuB or OmpF controls shown in the top section of Figure 4.6.

Initial analysis of the mixture by SEC, SDS-PAGE, and MP suggested strong complex formation across multiple fractions, especially fractions E4-E10, shown in Figure 4.11. However, throughout this analysis, the possibility of BtuB and OmpF self-crosslinking was not considered. As such, further controls of BtuB and OmpF G198^{BPA} peak 1 alone, mixed with the same quantity of the crosslinking agent and incubated for the same amount of time, were performed.

The A280 SEC traces of the OmpF G198^{BPA} peak 1 and BtuB crosslinking controls are shown in Figure 4.12A and C, respectively, with the fractions collected for further MP analysis highlighted in pink and purple. As can be seen from Figure 4.12B and D, the SDS-PAGE analysis of the fractions shows strong crosslinking in the control samples, forming numerous high MW species. To further confirm

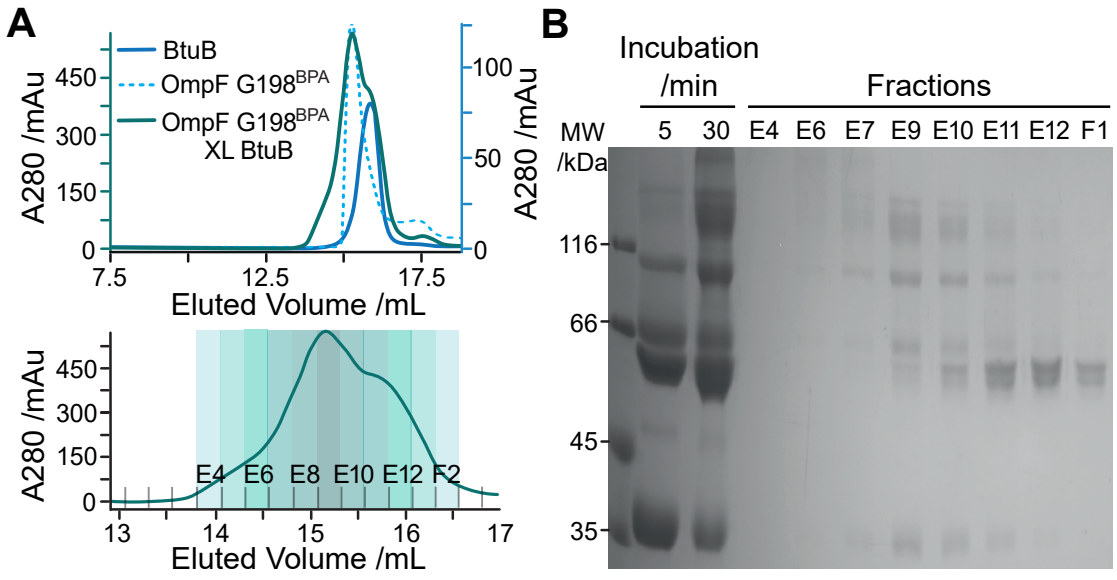


Figure 4.10: SEC and SDS-PAGE analysis of glutaraldehyde-induced OmpF-BtuB crosslinking. A. Top, 280 nm absorbance traces from SEC of BtuB WT and OmpF G198^{BPA} peak 1 controls, and OmpF crosslinked (XL) to BtuB in 1% OG. Bottom, magnification of the crosslinked trace. The individual fractions analysed by SDS-PAGE and mass photometry are highlighted. **B.** SDS-PAGE analysis of the individual fractions of OmpF crosslinked to BtuB. Crosslinking time checks were performed at 5 and 30 minutes, prior to loading onto the SEC column.

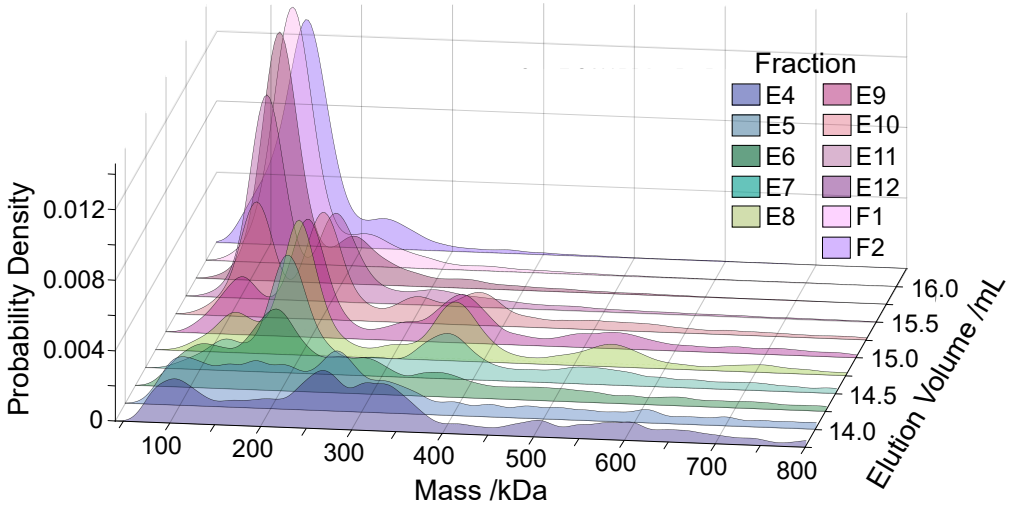


Figure 4.11: MP analysis of OmpF crosslinking to BtuB. OmpF G198^{BPA} was crosslinked (XL) in 1% OG to BtuB at approximately 2:1 ratio, transferred to amphipol and separated by SEC. Individual fractions E4-F2 were then analysed by mass photometry. The combined data of at least 3 repeats are shown for each fraction, plotted as kernel probability density in a waterfall plot. The predominant peak in Fractions E10-F2 is centred around 198 kDa, the predominant peak in Fractions E6-E9 is centred around 363 kDa, whilst the spectra of Fractions E4 and E5 are heterogeneous, with broad peaks that are difficult to assign.

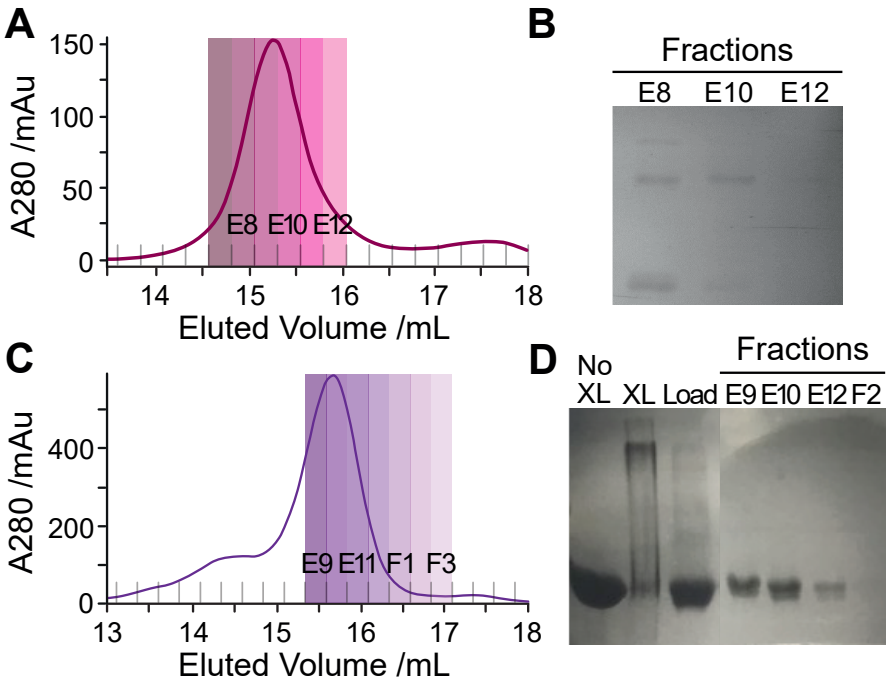


Figure 4.12: SEC and SDS-PAGE analysis of BtuB and OmpF G198^{BPA} crosslinking controls. A. SEC 280 nm absorbance trace (A280) of OmpF G198^{BPA} peak 1 sample crosslinked (XL) with glutaraldehyde for 30 min. The fractions analysed further by SDS-PAGE and mass photometry are highlighted in pink. **B.** SDS-PAGE analysis of the selected self-crosslinked OmpF fractions. **C.** SEC A280 trace of glutaraldehyde-crosslinked BtuB. The fractions selected for further analysis by SDS-PAGE and mass photometry are highlighted in purple. **D.** SDS-PAGE analysis of the selected fractions of self-crosslinked BtuB. The left side of the gel shows samples before and after crosslinking, indicating formation of higher MW species.

this phenomenon, selected fractions were analysed by MP, again showing a large number of different peaks at high MW (Figure 4.13), absent or less pronounced in the non-crosslinked BtuB and OmpF spectra shown earlier.

To confirm that the high MW peaks detected in the mixed, crosslinked sample, can be attributed to BtuB and OmpF forming homologous complexes, data from selected fractions of the three crosslinked samples were overlayed in a single plot, shown in Figure 4.14. The high MW peaks in the mixed spectra are also present, at very similar amplitude, in either the OmpF G198^{BPA} peak 1 or BtuB self-crosslinked data, as indicated by the OmpF and BtuB-representing cartoons. Interestingly, the BtuB self-crosslinked data appears to be present at a slightly higher MW compared

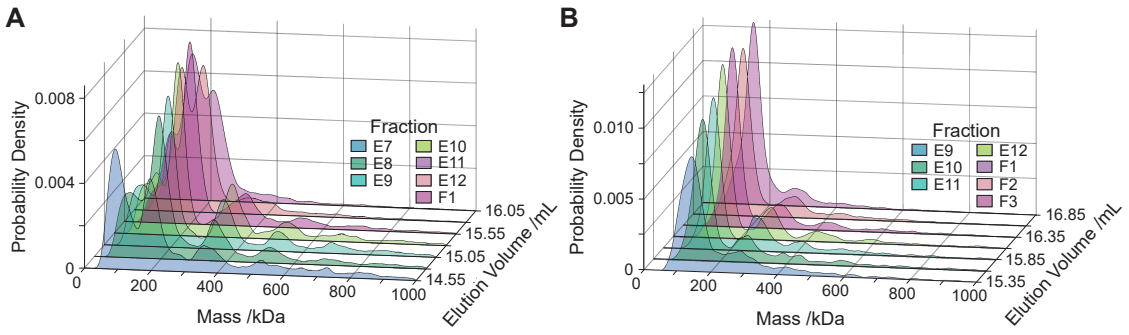


Figure 4.13: MP of BtuB and OmpF G198^{BPA} crosslinking controls. For the mass photometry measurements, SEC fractions of BtuB and OmpF G198^{BPA} crosslinking (XL) controls were selected based on their approximate molecular size, estimated from the elution volume of the fractions during gel filtration and SDS-PAGE migration, together with the protein content, as measured by the 280 nm absorbance during SEC. **A.** Selected OmpF G198^{BPA} peak 1 crosslinking control fractions E7-F1 were measured using MP at least 3 times. The combined averaged data was plotted as KDE functions, stacked in a waterfall plot. **B.** Fractions E9-F3 of the BtuB crosslinking control were measured with MP at least 3 times. The combined KDE data was averaged and plotted as above.

to the mixed sample. This could be due to enhanced crosslinking of BtuB to lipids present in the sample, although theoretically the glutaraldehyde crosslinking process should not result in complex formation between the protein and lipids. As such, this shift is most likely the result of slight instrument calibration error or focus shift, as the data for each of the samples were recorded on different days.

Overall, even though it appears that OmpF does interact with BtuB through OM PLs crosslinked to V177^{BPA} peak 1 sample (Section 4.3.2), when a non-specific crosslinking agent is introduced, such as glutaraldehyde, these OMPs appear to preferentially interact with themselves. It is possible that unlike the association between different proteins, self association may not require lipids. As a result, the distance between homologous proteins would be smaller, resulting in more efficient crosslinking, decreasing chances of heterologous species observation. Additional experiments with other co-purifying, potentially crosslinking OMPs would be valuable, in order to determine whether this is an OmpF and BtuB-specific behaviour, or a more generally applicable observation.

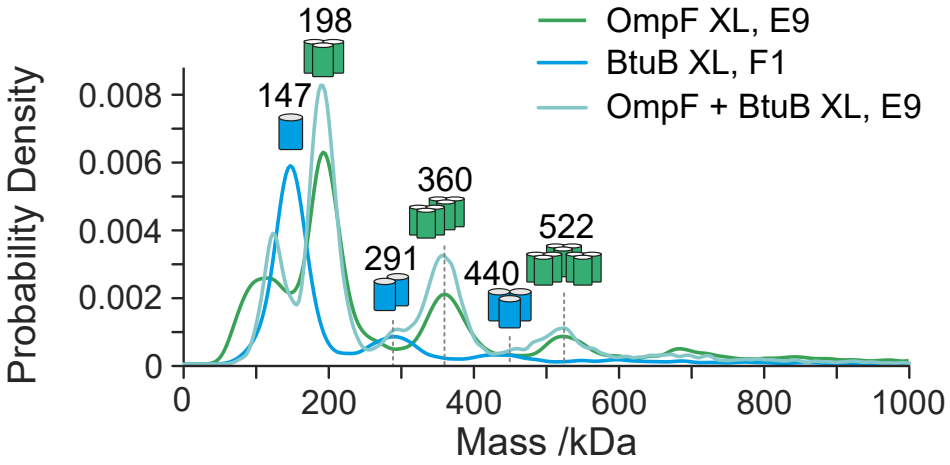


Figure 4.14: OmpF, BtuB and glutaraldehyde crosslinking data comparison. Representative glutaraldehyde-crosslinked (XL) fraction E9 of OmpF G198^{BPA} peak 1 mixed with BtuB (pale turquoise), and OmpF G198^{BPA} peak 1 (green) and BtuB (blue) controls were selected for comparable amounts of high MW species. Kernel density estimates were plotted for the combined datasets for each fraction and overlayed to highlight the matching peak positions of the mixed OmpF + BtuB sample and the self-crosslinked OmpF and BtuB data, indicated by dashed lines, suggesting little to no glutaraldehyde-induced crosslinking occurred between OmpF and BtuB.

4.4 Discussion

Overall, the crosslinking data obtained through a combination of MP, SEC, SDS and bnPAGE analysis show that mutual lipid interactions between OMPs can lead to the formation of homologous and heterologous complexes. Through the basis of these interactions, diverse supramolecular OMP islands could be formed based on the abundance of different OMPs, with the more frequent, trimeric OMPs like OmpF in *E. coli* performing the function of a fundamental organising unit. Taken together with their high frequency in the bacterial OM, these results suggest that the role of the trimeric OmpF and monomeric BtuB in the OM extends beyond their molecular function as a channel and transporter, and aids in the stability and organisation of the OM through a simple OmpF-lipid-OMP structural unit. These units would in turn result in a highly protein-packed membrane, causing lower permeability *via* diffusion of small molecules, ultimately causing increased protection against a variety of antimicrobial agents, as observed for the OM of many

Gram-negative bacteria.

Given the high frequency of specific lipid binding sites present in OMPs and integral membrane proteins in general, it would be expected for these lipids to play an important role in the OMP function and assembly, although the high concentration of these lipids in the OM could also play a role in facilitating these interactions.[288] The results presented suggest that these lipids, seemingly interspersed between OMPs, can act as a molecular binding agent, facilitating the formation and cohesion of OMP islands. The varying affinities of specific lipids, such as LPS, to different OMPs also likely contribute to the regulation and organisation of OMP islands, together with the variances observed in the OMP copy numbers.

Additionally, specific lipid interactions were previously shown to impact the function of different OMPs, such as LPS being required for BtuB vitamin B12 uptake,[289] as well as contributing to the open-closed state cycle of OmpF.[290] As such, other functions of lipids such as facilitating OMP oligomerisation through allosteric binding could be possible. Unfortunately, studying these specific OMP-lipid interactions can be challenging, partly due to the bulk lipids frequently co-purified with OMPs being difficult to distinguish from the lipids bound to specific sites on the target proteins, although it is important to consider that bulk lipids could bind specifically as well.[287] Through the development of the BPA-crosslinking strategy presented here, we were able to covalently stabilise and isolate specific OMP-lipid complexes, enabling efficient removal of these bulk lipids. Beyond the experiments presented here, this approach can allow for simpler analysis of OMP-lipid functions, including lipid roles in OMP oligomerisation, the allosteric binding of specific lipids, or conformational stabilisation.

Besides OMP island organisation study, the basic unit of OmpF-lipid-OMP could also be targeted by novel antibacterial agents for enhanced cell entry. This is the proposed mode of action for colicinE9 entry through the colicinE9 translocon, recently shown by cryoEM to form complexes with both OmpF and BtuB, together with a hypothetical LPS molecule suggested by an un-modelled electron density between the two β-barrels.[285] Further research supported the presence of LPS in this complex, including native MS, showing BtuB removed from the translocon complex associating with LPS, together with an OmpF orthologue OmpE36 shown to bind to two LPS molecules,[285,291] with the putative LP binding site in the cryoEM model aligning well with the OmpE36 model LPS binding site. The evidence presented here, together with previous research, suggests that specific bacteriocins, which require the presence of both a translocator and an OMP receptor for binding, utilise the described OMP-lipid organisation structure in order to recognise and bind to their receptors.

Evolutionarily, this OM island organisation could also serve to increase the efficacy of signalling pathways important for the survival of Gram-negative bacteria, allowing for OMP domain adaptation and improving the avidity of OMPs binding different epitopes.[292] Additionally, OMP clustering within these islands, distanced by a single lipid, would put them in sufficient proximity to allow for increased phage binding efficiency, which in other cases relies on the binding of a lipoprotein.[293] Finally, reducing TBDT steric crowding through their separation by trimeric porins could improve the efficiency of their coupling to the inner membrane and periplasmic Ton-system, providing energy for the import of small molecules facilitated by some of these TBDTs.

Research into the supramolecular organisation of the Gram-negative bacterial OM islands also helps improve our understanding of small molecule transport

by different biological systems. In addition to their crucial function in small molecule and antibiotic diffusion, we show that trimeric porins also fundamentally contribute to the bacterial OM organisation. This information improves our understanding of the OM biological function, potentially helping in the design process of new antibiotics, as well as improving the applicability and utility of current antibiotic treatments *via* directing attention on the specific OMP-lipid interactions fundamental for OM stability and organisation.

5

HIV Envelope Protein - BanLec Interaction

Contents

5.1 Introduction **128**
5.2 Materials and Methods **131**
5.2.1 Interferometric Scattering Microscopy 132
5.2.2 Sample Preparation and Landing Assay 133
5.2.3 Data Analysis 134
5.3 Results **138**
5.3.1 Env Interaction with BanLec 138
5.3.2 BanLec-Induced Env Oligomerisation 141
5.3.3 Quantifying the Env–BL Interaction 144
5.4 Discussion **147**

The data presented in this Chapter have been used for the following publication:

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*, **360**(6387), 423-427 (2018).[5]

HIV-1 Env and BanLec were produced by Weston B. Struwe (University of Oxford). HSP16.5 (*M. jannaschii*) was kindly provided by Olga Thachenko (University of Oxford). Figure 5.6 contains data produced jointly with Nikolas Hundt (University of Oxford).

5.1 Introduction

The behaviour and protein-protein interactions of the HIV envelope spike glycoprotein have been of considerable interest to the scientific and medical community for decades. HIV is a strain of infectious retrovirus that, without treatment, results in development of Acquired Immunodeficiency Syndrome (AIDS), which can then be fatal due to progressive failure of the immune system. The latest report by the United Nations World Health Organization showed that in 2020, approximately 30-45 million people worldwide were living with HIV infection.[170] Over 90% of HIV infections are due to HIV-1, group M (major) strain.[294] The increasing efficiency and availability of treatments, such as anti-retroviral therapy, has caused the number of AIDS-related deaths to decrease since 2010 by approximately 47%.[170] Unfortunately, the rate of new infections is not following this decline – over 1.5 million new infections were reported worldwide in 2020 alone.[170] As such, prevention, together with better education and improved access to treatment, is necessary to slow the spread and help with eradication of HIV. One of the recently developed approaches for preventing HIV-1 spread are topical microbicides inhibiting HIV viral replication and infectivity.[295]

HIV-1 is dependent on the interaction between its envelope protein complex (Env) with cell receptors on CD4(+) cells to infect host immune cells. Env is a trimer of heterodimers of the membrane-anchored gp41 and heavily glycosylated gp120, and is also referred to as the 'viral spike' of HIV-1.[296] Just as in other viral

systems, the host system extensively glycosylates the Env protein, providing a shield which prevents recognition and inhibition by most conventional neutralizing antibodies.[297] However, potent broadly-neutralising antibodies can be produced by some HIV-1 patients, generally targeting the glycosylated surface directly. Since up to 50% of the MW of gp120 consists of N-linked glycans,[298] the glycosylation shield presents an important target for the development of vaccinations and other methods of prevention that rely on inhibiting HIV-1 replication.

Recently, a potent HIV-1 replication inhibition by a plant lectin found in the ripened banana peel of the *Musa acuminata* or *Musa paradisiac* species, BanLec (BL), was described.[295] BL was first isolated in 1990,[299] and has since been of significant interest thanks to its high binding specificity for the sugar moiety mannose and mannose-containing oligosaccharides and glycoproteins.[300] Together with a number of other members of the lectin family, such as cyanovirin or griffithsin, BL can bind enveloped viral surface glycoproteins, and hence efficiently block viruses from entering host immune cells.[301] More specifically, for HIV-1 entry into cells, BL is one of the most potent natural inhibitors, with the half-maximal inhibitory concentration (IC_{50}) for preventing HIV-1 fusion with HeLa cells in the low nanomolar range.[295]

X-ray crystallography and molecular dynamics studies have shown that binding of BL to Env is mediated through two carbohydrate-binding sites on each monomeric subunit specific to disaccharides with α-1,3-mannosyl linkages, which are recognised within high mannose N-glycan epitopes on the surface of HIV-1.[299] Recent native MS experiments showed that to bind a single high-mannose glycan, both carbohydrate-binding sites on a monomer are necessary.[302] Through a combination of glycan depletion, high-performance liquid chromatography (HPLC) and HIV-1 depletion assays, it was found that BL binds specifically to areas containing 7-9 mannose

units,[301] which generally form over half of the glycan shield.[298]

Until recently, it was thought that BL was composed of a homodimer of two 15 kDa monomers. This structural information was obtained through a wide variety of techniques, including X-ray crystallography,[300,303,304] or SEC detection.[299] However, new results published by Hopper *et al.* showed BL adapts a tetrameric structure in solution without any other stoichiometries present, as detected by native MS and confirmed by SEC-MALS.[302] Further experiments aimed at interfering with the tetramer assembly through site-directed mutagenesis resulted in a significant decrease of BL neutralisation potency, implicating the importance of its tetrameric stoichiometry. Interestingly, this structural arrangement is common in other members of the jacalin family of lectins.[305] Together with the bidentate binding of single N-glycans, it appears to enhance the viral neutralisation properties by increasing binding avidity.[302] The structure of the HIV-1 glycan shield aids itself to multivalent glycan binding and is also characteristic of high-density glycans. Altogether, it helps overcome the typically weak glycan-protein interactions, which are generally in the milli- or micromolar affinity range.[306] Additionally, apart from the bidentate binding and avidity effects arising from its oligomeric structure, the presence of two other binding sites on BL, facing away from the bound Env suggests the possibility of BL binding to more than one Env, potentially crosslinking the proteins, further lowering the ability of HIV to infect the cells. However, due to the high complexity and variety of species arising from such interactions, these reactions have been challenging to probe.

Overall, BL has a strong potential for use in microbicides and as an HIV-1 neutralising therapeutic. It can also act as a starting point for designing synthetic lectins, with improved pharmacological properties and higher potency. During the last decade, our understanding of the interaction between HIV-1 Env and BL has

improved significantly. However, most of the studies done up to date were performed using either isolated N-glycans or viral neutralisation assays. Studying the direct interaction between the lectin and the glycans as attached to Env could significantly improve our understanding of this interaction. However, the extensive glycosylation makes this inherently difficult. Furthermore, as an integral membrane protein (IMP) complex, Env can be challenging to handle due to the lipid bilayer interactions necessary for correct folding and function. For this reason, we studied the Env–BL interaction using a stabilised, modified Env trimer, gp140, with interferometric scattering microscopy (iSCAT), which has been shown to be able to detect soluble and membrane proteins reliably.[5,16,307,308] Using a titration approach together with high repeat number studies of a single condition, we characterise the Env–BL interaction using iSCAT, demonstrating the crosslinking abilities of BL contributing to its strong inhibitory effect on HIV-1 infection.

5.2 Materials and Methods

All chemicals and materials used in this chapter were purchased from Sigma-Aldrich, unless stated otherwise. Env and BL proteins were produced and kindly provided by Dr. Weston Struwe (University of Oxford) and his research group, following protocols described previously.[178] Heat shock protein (HSP) 16.5 from *Methanocalcocuccus janaschii* was kindly provided by Dr. Olga Thachenko (University of Oxford).

HIV-1 Env is a heterohexameric IMP complex consisting of three copies of gp120 and gp41 each. However, the native membrane-inserted complex is challenging to express, purify, and prepare in aqueous solutions, likely due to its transmembrane region and necessary membrane interactions. Instead, throughout this chapter, Env is used to refer to a recombinant HIV-1 Env BG505 SOSIP.664 construct, gp140, outlined in Figure 5.1.[296] gp140 is an antigenically near-native fusion between

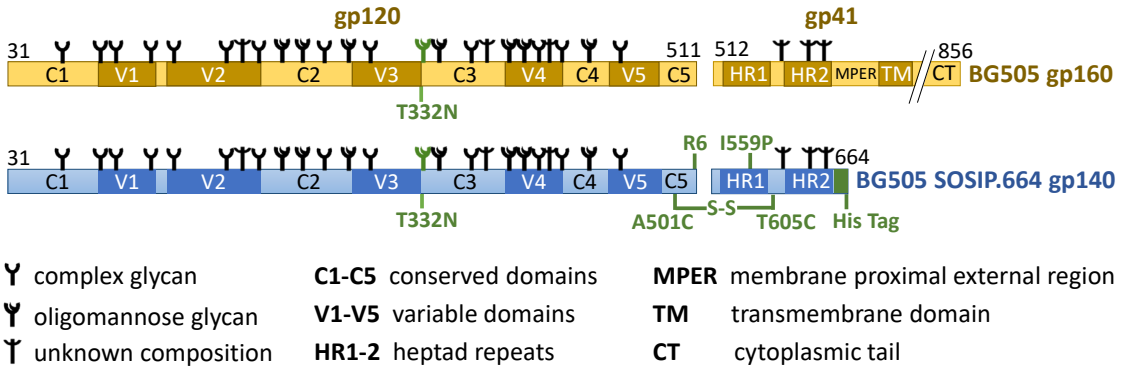


Figure 5.1: Env gp160 complex compared to the Env gp140 construct. A diagram of the sequence of the native Env BG505 gp160 membrane glycoprotein complex, consisting of gp120 and gp41, compared to the truncated, soluble, disulphide-bonded Env BG505 SOSIP.664 gp140 construct. The transmembrane domain (TM) and the cytoplasmic tail (CT) are truncated to enable protein solubility, whilst the C-terminal His Tag aids in protein purification. The cysteine disulphide bond is created by the A501C and T605C mutations. The approximate positions and broad types of all mapped glycans are indicated. Figure adapted with permission of American Association for the Advancement of Science, from Crystal Structure of a Soluble Cleaved HIV-1 Envelope Trimer, Figure 1A, Julien *et al.*, *Science* **342**, 2013;[296] permission conveyed through Copyright Clearance Center, Inc., with figure adaptations approved by the authors. The types of glycans present on the constructs were obtained from Sanders *et al.*[178]

gp120 and a truncated version of gp41 that forms a stable and soluble trimer. The removal of the transmembrane (TM) and cytoplasmic tail (CT) domains renders gp140 soluble, while the disulphide bond created by mutating alanine A501 and threonine T605 to cysteines (C) facilitates gp120 and gp41 fusion.[178,296] All the original N-glycosylation sites and their respective asparagine residues are preserved, with an additional glycosylation site created by mutating threonine T332 to an asparagine (N).[296] The amino acid MW of the monomer is ∼76 kDa, with 24 glycan sites per monomer of ∼1.5-2 kDa each, bringing the total expected glycoprotein MW of the Env gp140 trimer to 336-372 kDa.

5.2.1 Interferometric Scattering Microscopy

The Env–BL interaction was investigated using a home-built interferometric light scattering microscope (iSCAT), constructed by Dr. G. Young and D. Cole, as described in Young and Hundt *et al.*[5] Briefly, the autofocus detection and automatic

optimization of the z-plane position was performed with a custom-designed autofocus detection program, written by Dr. G. Young in the National Instruments^{TM} LabView program, using a 635 nm laser beam. A collimated 445 nm laser beam was scanned across the field of view causing sample illumination *via* two acousto-optic deflectors (AODs) in an independent, phase-locked fashion with frequencies of 79 and 83 kHz for the x- and y-axis, respectively. This set-up resulted in the laser beam passing through the imaged area thousands of times within the acquisition time of the camera, providing uniform illumination, avoiding interference fringes and reducing speckle noise.[5] The size of the image taken was 256 × 256 pixels^2, with conversion of 23.4 nm/pixel. Images were then binned during data acquisition into 3 × 3 sized pixel bins, resulting in a final 70.2 nm pixel size. The recording frame rate was 662 frames/sec, and the recordings were pre-averaged over 10 frames during data acquisition for shot noise reduction, resulting in a final frame rate of 66.2 frames/sec.

5.2.2 Sample Preparation and Landing Assay

The experiments in this chapter were performed using a landing assay, with particle recognition based on protein molecules attaching to the glass surface, causing a localised change in the refractive index (RI) with reference to the surrounding buffer, as explained in Section 2.4. Unless stated otherwise, the buffer used was Dulbecco's phosphate buffered saline (PBS, Gibco Life Technologies). Dr. Nikolas Hundt (University of Oxford) kindly supervised the conduct and the experimental design of most of the Env–BL interaction experiments.

Throughout this chapter, assembled coverslip chambers were used for iSCAT measurements. First, the 24 × 50 mm coverslips (No. 1.5,VWR) were sequentially rinsed in Milli-Q H_2O (MQ), ethanol, MQ, 2-propanol, MQ, ethanol, and MQ, whilst the 24 × 24 mm coverslips were rinsed in MQ, ethanol, MQ, ethanol and MQ, where the ethanol and 2-propanol used were HPLC-grade. The cleaned coverslips

were then dried using a clean nitrogen gas stream. The coverslip flow chambers were then constructed with a single piece of a 24 × 50 mm and a 24 × 24 mm glass coverslip, glued together with a double-sided transparent tape (Scotch), creating a channel of approximately 18-20 μL. BL and Env were retrieved from 4 °C, diluted to double of their desired concentrations with PBS, and kept on ice. 25 μL of Env and BL samples were then mixed and incubated on ice for ∼5 minutes before the measurements. The final concentration of Env was 10 nM, whilst the BL concentration varied from 1 to 50 nM. In order to convert ratiometric contrast into MW, proteins with a well-defined MW and distribution were used as protein standards. For most of the experiments, HSP16.5 was retrieved from -80 °C and diluted to 240 nM monomer concentration in PBS. HSP16.5 forms a 24-mer in solution with MW = 396 kDa.

At the start of the measurement, a fresh flow chamber was placed on the microscope objective, with a drop of objective immersion oil (IMMOIL-F30CC, Olympus) placed between the objective and the underside of the coverslip. The channel was then flushed with PBS or MQ, autofocus adjusted, and 20 μL of sample was inserted. The landing events were recorded for 120 seconds × 2 for each sample. In order to enable better detection of large Env clusters forming at high BL concentrations, the 20 nM BL and 10 nM Env condition was measured over 10 coverslips, giving access to a large number of repeats and high particle counts, aiding in the analysis of low-frequency high-MW events.

5.2.3 Data Analysis

The raw data was processed using Discover^{MP} (Refeyn Ltd., Version 2.1), with analysis frame binning set to 5. Analysis thresholds of T1 = 1.5 and T2 = 0.25 were used, with a final frame averaging of 50, and final frame rate of 13.24 frames/sec. Details of the data analysis program are described in Section 3.2. Prior to

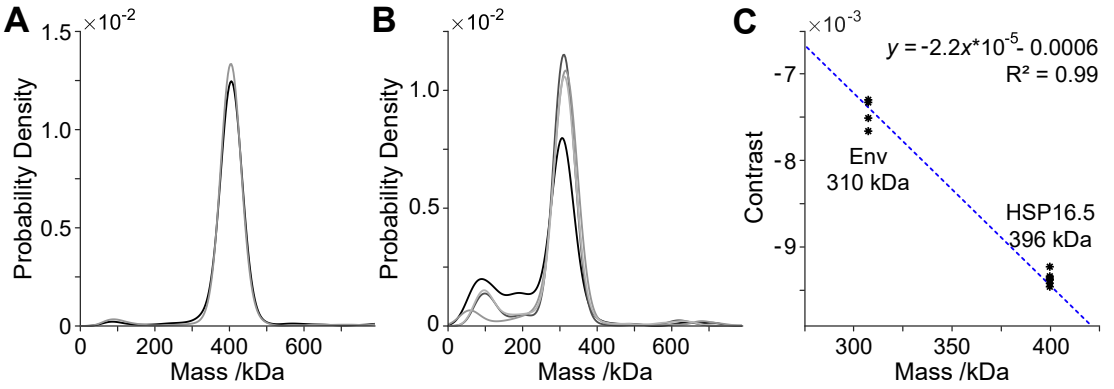


Figure 5.2: HSP16.5 and Env data used for mass calibration. A. Kernel density estimate (KDE) functions were fitted to the particles detected in the HSP16.5 movies (expected MW = 396 kDa), measured at 240 nM, twice. **B.** KDE functions were fitted to Env movies (expected MW = ∼310 kDa), measured at 10 nM trimeric concentration, four times. The kernel density size for plotting was 20 kDa. **C.** The mass calibration was performed *via* linear regression fitting of the mean contrast of the Env and HSP16.5 peaks to their expected MW. The graph shows combined data points for two individual days of measurements. The resulting equation and the goodness of fit R^2 value is shown on the top.

experimental data analysis and processing, the calibrant protein data were analysed using Discover^{MP}, and the resulting peaks fitted with the embedded Gaussian fitting function. A combination of HSP16.5 and Env (Figure 5.2A and B, respectively) were used, producing single peaks, with 2-4 repeats from the same coverslip recorded. The low probability density peak at ∼100 kDa in the Env spectra likely arises from either Env trimer disassembly or sample impurities. However, as the data analysis focuses on particles in the region above 300 kDa, the presence of this peak can be disregarded.

Following Gaussian peak fitting of the protein standards in Discover^{MP}, the contrasts were extracted and coupled with the expected MWs, and fitted with a linear function $y = bx + c$, where b corresponds to the calibrant slope and c corresponds to the y –axis intercept. The rest of the data was then analysed in Discover^{MP} with set calibration, and the MWs of individual particle landing events were extracted and analysed using MATLAB. The particles from each recording were sorted using

the histogram function as well as fitted to a kernel density function, with the kernel size set to 5-30 kDa. The consistency between repeats was checked by overlaying histograms before combining them, except for the 20 nM BL + 10 nM Env data, where the repeats from the same coverslip were combined.

The changes in the MW mean and standard deviation with increasing BL concentrations were fitted with mass action kinetics, according to Equation 5.2 derived from the reaction depicted in Equation 5.1, where k_{on} and k_{off} refer to the forward and reverse rate of association of the Env–BL complex, respectively, and n is the number of BL molecules binding to Env; in this case, $n = 1$.[302]

$$Env + n \times BL \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} Env \cdot BL_n \tag{5.1}$$

$$K_{BL} = \frac{k_{on}}{k_{off}} = \frac{[Env][BL]^n}{[Env \cdot BL_n]} \tag{5.2}$$

Through rearranging Equation 5.2, the proportion of Env bound to BL (Θ) can be expressed in terms of unbound BL concentration and the interaction K_{BL}, which is the equilibrium constant of the Env–BL interaction. Substituting for $n = 1$ gives Equation 5.3:

$$\Theta = \frac{\text{Occupied Protein}}{\text{Total Protein}} = \frac{[Env \cdot BL]}{[Env] + [Env \cdot BL]} = \frac{[BL]}{K_{BL} + [BL]} \tag{5.3}$$

From which we can further rearrange to calculate the K_{BL} for the Env–BL interaction:

$$K_{BL} = \frac{[BL]}{\Theta} - [BL] \tag{5.4}$$

For the combined data presented in Section 5.3.3, the data were analysed by modelling the Env–BL system as described in Young and Hundt *et al.*[5] Briefly, the following reactions and their respective equilibrium constants were modelled:

$$Env + Env \rightleftharpoons 2 \times Env \implies K_d = \frac{[Env]^2}{[Env_2]} \tag{5.5}$$

$$Env + BL \rightleftharpoons Env \cdot BL \implies K_{BL} = \frac{[Env][BL]}{[Env \cdot BL]} \tag{5.6}$$

$$Env_n \cdot BL_n + Env \rightleftharpoons Env_{n+1} \cdot BL_n \implies K_{Env} = \frac{[Env_n \cdot BL_n][Env]}{[Env_{n+1} \cdot BL_n]} \tag{5.7}$$

$$Env_{n+1} \cdot BL_n + BL \rightleftharpoons Env_{n+1} \cdot BL_{n+1} \implies K'_{BL} = \frac{[Env_{n+1} \cdot BL_n][BL]}{[Env_{n+1} \cdot BL_{n+1}]} \tag{5.8}$$

Where n refers to the number of Env and BL molecules in an oligomeric complex. Given the conservation of the total numbers of Env trimers and BL tetramers, we can express the initial concentrations of Env and BL ($[Env]_0$ and $[BL]_0$) as the sum of the number of species containing Env or BL, multiplied by the number of Env trimers or BL tetramers in each:

$$[Env]_0 = [Env] + 2 \times [Env_2] + [Env \cdot BL] + \sum_{m,n>0} m[Env_m \cdot BL_n] \tag{5.9}$$

$$[BL]_0 = [BL] + [Env \cdot BL] + \sum_{m,n>0} n[Env_m \cdot BL_n] \tag{5.10}$$

Using the equilibria and constants described in Equations 5.5–5.8, and taking $[Env]_0$ = 10 nM, and $[BL]_0$ based on the initial concentration of BL used (0.1-50 nM), it is possible to describe the concentration of each oligomer in terms of the equilibrium constants and $[Env]^n$ or $[BL]^n$, where $n > 1$. Substituting these into Equations 5.9 and 5.10, we can numerically solve two simultaneous equations to calculate the solution concentrations of Env and BL, *i.e.* $[Env]$ and $[BL]$, which can then be used to calculate the concentrations of the rest of the oligomeric species. Once the concentrations of $[Env]$, $[BL]$, and other oligomeric species are known, we can calculate K_d, K_{BL}, K_{Env}, and K'_{BL}, characterising the Env–BL interaction at a single-molecule, oligomeric level.

5.3 Results

5.3.1 Env Interaction with BanLec

When calculating the expected MW of the gp140 Env construct when measured by iSCAT, it is important to take into account the difference between the refractive index increment (*dn/dc*) values of glycans and polypeptides. iSCAT measures the changes in the local refractive index resulting in differential light scattering and interference (see Section 2.4 for more details). Together with the use of non-glycosylated protein standards for instrument calibration, the different protein and glycan *dn/dc* values will affect the expected measured contrast of heavily glycosylated proteins. Since 32.1–38.7% of the gp140 MW consists of glycans, some extent of MW underestimation would be expected.

The average generally accepted *dn/dc* values of glycans and polypeptides are approximately ∼0.15 and 0.18 mL/g, respectively.[234] Given that ∼50% of gp120 is composed of glycans, and assuming linear addition of refractive indexes of the glycan and protein sections,[309] iSCAT should underestimate the MW by ∼8.3%. Comparison of the gp120 MW obtained by iSCAT and native MS revealed that iSCAT underestimates the MW by ∼6%.[234] Due to the slightly lower proportion of glycans in the gp140 Env construct, we can deduce that iSCAT should underestimate its MW by 4.8–5.8%, resulting in expected mass between 316.5 and 354 kDa. Throughout the experiments presented here, the average measured Env MW is ∼310–315 kDa, which fits well with the lower estimated Env gp140 mass after iSCAT MW underestimation is taken into account. The measured value also suggests that the Env used here contains glycans at the lower end of the expected weight range (1.5–2 kDa), or that not all 24 glycan sites are occupied. Other techniques capable of high mass resolution and protein and glycan digestion would be necessary to determine the precise occupancy rate of the Env glycan sites and the average mass

of the glycans attached.

Characterisation of the Env–BL interaction using iSCAT could provide novel insights into the mechanism of HIV neutralisation by BL. Each BL monomer contains two sugar binding motifs that are both employed to bind a single glycan,[310] meaning that Env and BL interact in a 1:1 ratio. However, due to its conformation, BL dimer is capable of binding to two Env monomers in a trimer, increasing the binding affinity drastically as a result of the higher avidity of the interaction.[302] Since protein-glycan interactions are generally very weak, this arrangement enables efficient binding of BL to Env. Consequently, the highest number of BL tetramers expected to bind to a single Env trimer is 2, although the binding of the second BL would likely be much weaker compared to the first tetramer.

In order to probe these predictions, 10 nM Env was measured alone, and in combination with 1-50 nM BL. Figure 5.3 shows the respective mass distributions, with each condition containing 2 or 4 repeats. Increasing BL concentrations, especially after 10 nM BL, lead to notable Env trimer peak broadening, a shift to a higher MW, and an appearance of a distribution 'tail' of increasing MW. The MW increase is indicated by the shift of the Env control peak position (light blue line, 310 kDa) to ∼330–340 kDa (dark blue line). Whilst the MW of the main Env peak in presence of higher BL concentrations appears lower than expected upon complete Env–BL binding (∼360–370 kDa), this is likely the result of the superimposition of the Env and Env + BL peaks.

Alternatively, it is possible that at the relatively low BL concentrations present, BL exists in a dimer-tetramer equilibrium. The evidence for the BL tetrameric structure in solution was obtained using native MS and SEC-MALS, both of which measure proteins at μM concentration, an order of magnitude higher than iSCAT.[302]

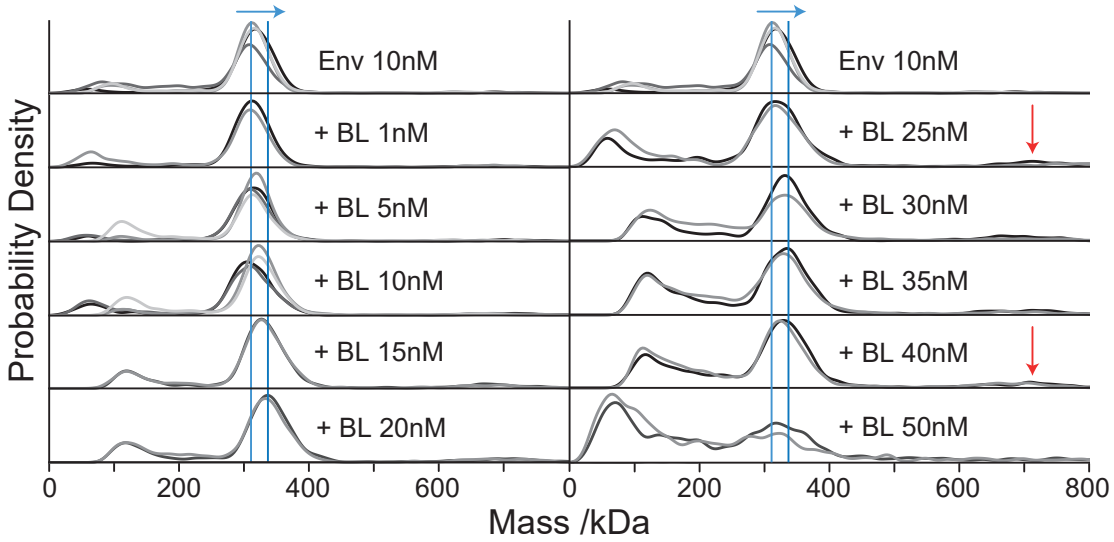


Figure 5.3: Mass distributions of Env titrated with 0-50 nM BL. Vertically stacked KDE plots of Env 10 nM control and Env 10 nM with a titration of progressively higher concentrations of BL, from 1 to 50 nM. A shift of the Env peak towards a higher MW is observable with increasing BL concentration, marked by the light blue line (Env blank MW) shifting to the dark blue line (Env + BL MW). Additionally, an increase in Env dimers of trimers occurs (examples marked by red arrows), as more Env is dimerised by BL.

At the same time, previous studies reported dimeric BL using techniques such as X-ray crystallography.[300,303,311] As such, it is likely that BL exists as a mixture of dimers and tetramers at the low nM concentrations measured. Unfortunately, the detection limits of iSCAT prevent the characterisation of this hypothetical equilibrium using this technique.

Despite this observable shift in the mean MW, it is challenging to quantitatively evaluate the Env–BL interaction due to the considerable level of fluctuations in the mean peak position and overall particle counts, whether between the same condition repeats or different BL concentrations. This is most likely the result of general experimental variation and instrument focus drift throughout the measurements, together with small changes in the observed MW of specific peaks caused by the titration being performed across multiple days. Additionally, as most of the conditions were repeated only twice, any statistical analysis regarding the overall

particle counts or peak position shift would be inaccurate.

5.3.2 BanLec-Induced Env Oligomerisation

In addition to BL interacting with the Env trimer, its tetrameric structure could facilitate Env oligomerisation through further Env binding to its two remaining glycan binding sites, which could then bind another BL and Env, leading to clustering. Such BL-induced Env aggregation on the viral surface *in vivo* could lead to strongly enhanced viral neutralisation. Indeed, the HIV neutralisation efficiency of a genetically modified BL dimer is much lower than that of the native tetramer, despite presumably nearly identical binding affinities to the Env trimer alone.[302]

In an iSCAT experiment, this enhanced Env–BL complex formation would be likely demonstrated through an increased proportion of particles in the higher MW regions. We can indeed observe the appearance of a broad peak at ∼740 kDa, corresponding to a complex containing two Env and zero to three BL proteins (Figure 5.3, examples highlighted by red arrows). The tetrameric structure of BL appears capable of acting as an intermolecular crosslinker, facilitating Env trimer dimerisation. A very small peak at ∼630 kDa is also present in the control Env sample (visible in Figure 5.2A, right), suggesting some dimerisation also in the absence of BL. Nevertheless, there is an increase in the proportion of particles in the dimer section, correlated with higher BL concentration, as well as a clear shift of the mean of the peak, indicating that increasing BL concentration leads to a higher level of Env dimerisation through BL crosslinking. As the dimers are also capable of binding two or three BL tetramers, the shift to a higher MW is more pronounced.

As expected, the presence of BL leads to further Env crosslinking, resulting in higher occurrence of high MW species. At the highest tested BL concentration

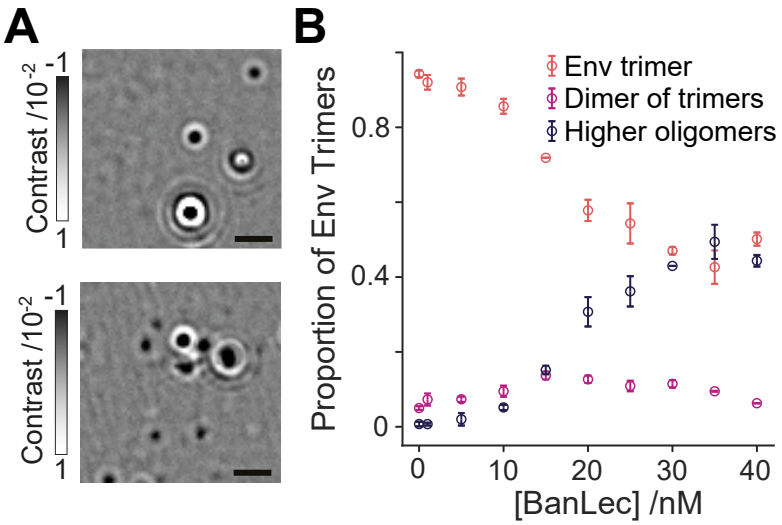


Figure 5.4: Env–BL oligomer formation increases with BL concentration. A. Frame 1249 from repeat 1 (top) and frame 52 from repeat 2 (bottom) of 10 nM Env + 50 nM BL were exported from Discover^{MP}, showing the presence of large oligomeric species and cluster formation, respectively. The contrast scale bar is indicated on the left of the frames, and the size scale bar is on the bottom right, representing 2 μm. **B.** Changes in the proportion of Env trimer particles present in the Env trimer peak (orange), Env dimer of trimers peak (magenta), and all higher oligomers detected (purple), over increasing BL concentration from 0 to 40 nM. The data represents the calculated mean proportion of Env trimers in each population, out of the total calculated Env trimers across 2–4 repeats. The error bars indicate ± standard deviation.

(50 nM), the frequency of these large oligomers complicated the measurements, resulting in broad, unclear peaks (Figure 5.3, bottom right plot). These complexes would often take up a large part of the field of view during the recording (Figure 5.4A, top), as well as 'float', rapidly alternating between binding and unbinding, appearing as a semi-constant objects for a significant part of the recording. The floating species would then sometimes obscure other landing events, forming particle clusters difficult to interpret using the particle detection algorithm (Figure 5.4A, bottom). However, it is important to bear in mind that blank control 0–50 nM BL measurements were not performed at the time of these experiments. As such, it is possible that these aggregates and floating particles are an artefact of the high BL concentration and contain predominantly BL clusters.

The Env–BL reaction was further evaluated in terms of the proportions of total

Env trimers (molar proportion) present in three separate populations - Env trimer, Env dimer of trimers, and higher oligomers. The evolution of these proportions of 10 nM Env over 0–40 nM increasing BL concentration is depicted in Figure 5.4B. Due to the aforementioned analysis difficulties present as a result of the high number of large oligomers, the 50 nM BL condition was not included. For the Env trimer and dimer of trimers populations, the number of Env trimers was estimated based on the number of particles detected through the Discover^{MP} built-in Gaussian function fitting for each repeat, multiplied by 2 for the dimer of trimers population. In the case of higher oligomers, all particles above ∼900 kDa were counted, their MWs summed, and the population mean calculated. The average MW was then calculated based on the mean and the total number of particles. Since there are essentially no higher oligomers present in the control Env 10 nM measurements, we can assume that this entire population is crosslinked by BL. Most Env molecules in the complex, except the peripheral trimers, will be bound to 2 × BL, and vice versa. As such, we can assume 1:1 Env:BL complex stoichiometry, and divide the mean oligomeric MW by the combined MW of Env and BL (∼370 kDa), yielding the average number of Env trimers in the complex.

Comparing the evolution of the different populations' proportions in Figure 5.4B, the amount of Env trimers in the Env trimer peak decreases in an apparent sigmoid fashion, mirrored by a similar increase in the amount of trimers present in the higher oligomers. Interestingly, whilst the proportion of dimers increases as well, from 5% in the Env control up to ∼13.5% with 15 nM BL, it subsequently plateaus, and even decreases slightly. This is likely caused by the BL tetramers binding to the kinetically less-favourable, single exposed monomers on the Env dimer, followed by a rapid interaction with more Env, leading to an increased formation of higher oligomers alongside Env trimer and dimer of trimers depletion. In this scenario, the Env dimer of trimers acts as a reaction intermediate between the

Env trimer and large Env–BL complexes, remaining at a relatively constant, low level regardless of BL concentration.

5.3.3 Quantifying the Env–BL Interaction

Whilst the BL titration enables some degree of analysis, especially of the higher oligomer formation, the variation within the main Env trimer peak position precludes more thorough evaluation of the Env trimer - BL tetramer reaction. As such, a single condition was selected, and measured over many repeats, on the same day, across multiple coverslips. The BL concentration selected was 20 nM, providing a middle ground between sufficient rates of the Env–BL reaction without excessively large amounts of high MW oligomers complicating the data analysis. This condition was repeated 20 times, over 10 separate coverslips. For the data analysis, the two sets of data from each coverslip were combined, in order to increase the number of particles within each histogram for more accurate Gaussian fitting. The averaged histogram of the 10 nM Env + 20 nM BL reaction is shown in Figure 5.5, with the SD and 95% confidence intervals shown as the darker and lighter green areas around the mean, respectively. The inset magnifies the section of the histogram showing the dimer of trimers and higher oligomers. For the dimer spectra, individual Env + BL species are resolvable, as labelled by the cartoons of Env and BL.

Similar to the titration, the appearance of larger oligomeric species can be quantified by counting all species above 450 kDa, indicating the amount of Env in dimer of trimers and higher oligomers. For all 10 coverslip repeats, the average number of particles above 450 kDa is 542.1 ± 48.26 (mean ± SD) particles. In comparison, for the two Env 10 nM controls, measured on different days, the number of particles above 450 kDa was on average 120 ± 1, so almost 4.5 × lower than with 20 nM BL. This difference is statistically significant, with a two-tailed t-test resulting in a p value of < 0.0001. These particles appear to not only contribute

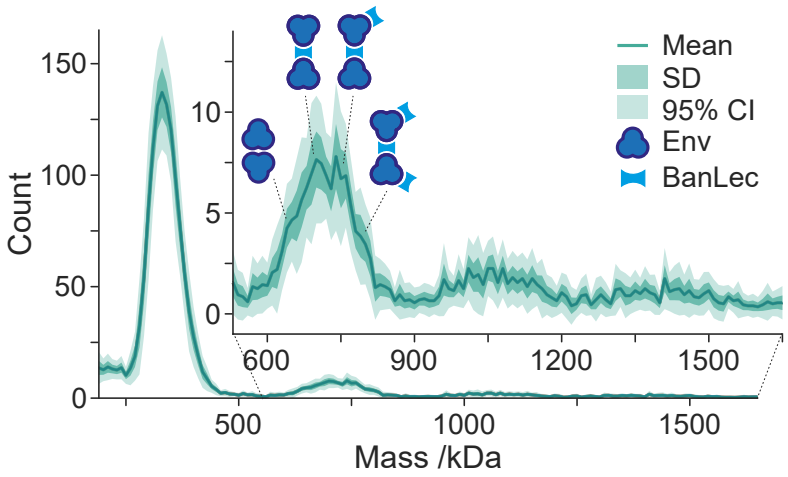


Figure 5.5: Averaged histogram of 20 repeats of 10 nM Env with 20 nM BL. Averaged histogram of 10 nM Env + 20 nM BL, measured across 20 repeats using 10 separate coverslips. The mean is plotted as the dark green line, with the standard deviation (SD) and the 95% confidence intervals (CI) shown in darker and lighter green shades, respectively. The histogram section showing Env dimers of trimers and high oligomers is magnified in the inset. For the Env dimer of trimers peak, illustrations indicate the positions of its composite species, *i.e.* 2 Env trimers bound to 0-3 BL tetramers.

to the Env dimer of trimers peak, but also to higher oligomeric species, as can be seen in Figure 5.5, inset. There appear to be broad peaks in the MW range corresponding to the dimer of trimers, trimer of trimers, tetramer of trimers and even pentamer of trimers of Env, with respective numbers of BL bound. Focusing on the dimer of trimers peak, it appears to have multiple components, suggesting different extent of BL occupancy. As described above, BL has four binding sites, with each binding site containing two glycan-binding motifs. These two binding motifs are required to bind to a single Env subunit at any given time; as such, it is reasonable to infer that BL can act as an oligomerisation tool for Env, binding multiple Env trimers at the same time, crosslinking them.

Using the information obtained through this large dataset, the coverslip repeats were combined, the Env monomer peak areas were plotted as histograms, and fitted to three Gaussian functions using MATLAB's Curve Fitting Tool, with the peaks fitted representing Env alone, Env + BL, and Env + 2×BL. In short, this tool

minimises a non-linear least squares loss using gradient descent. The mean MW positions and standard deviations of each peak were restricted to a defined range of values, and the resulting amplitudes of the optimised fits were used to compare the changes in peak proportions. Using Equation 5.4 derived in Section 5.2, it was possible to determine the equilibrium constant of BL binding to an Env trimer (K_{BL}), which was found to be 5.2 ± 1 nM, or 0.19 ± 0.037 nM^{-1}.

Further datasets repeating the Env–BL titration with additional BL concentrations were obtained by Dr. Nikolas Hundt, using the methods outlined in Section 5.2. Combined with the data analysed above, these experiments yield a clear picture of the molecular details of the Env–BL interaction. The mass histograms of the combined data are shown in Figure 5.6A, with the number of Env and BL units present in the complex indicated on top. Plotting the evolution of the Env oligomeric species molar fractions across the BL titration in Figure 5.6B confirms that the combined data is in good agreement with the analysis presented in Figure 5.4, showing similar decline in the proportion of Env monomers, plateauing Env dimer, trimer, and tetramer proportions, and increasing proportions of Env pentamers and higher oligomers.

The schematic on the top of Figure 5.6B suggests our proposed simplified mechanism for Env oligomerisation by BL, with the different equilibrium constants indicated. Using this additional data, it was possible to calculate more precise values for these constants: $K_{BL} = 0.12$ nM^{-1}, $K_{Env} = 8$ nM^{-1}, and $K'_{BL} = 0.4$ nM^{-1}, which agree well with the analysis presented above as well as bulk studies performed using 'sandwich' SPR, estimating K_{BL} to be 0.19 nM^{-1}, and revealing a secondary binding event with an equilibrium constant of 2.85 nM^{-1}.[301] Our resolution allowed us to demonstrate that the second Env binding to a BL–Env complex is stronger than to free BL, a feature typical of cooperative behaviour. Given that multivalency

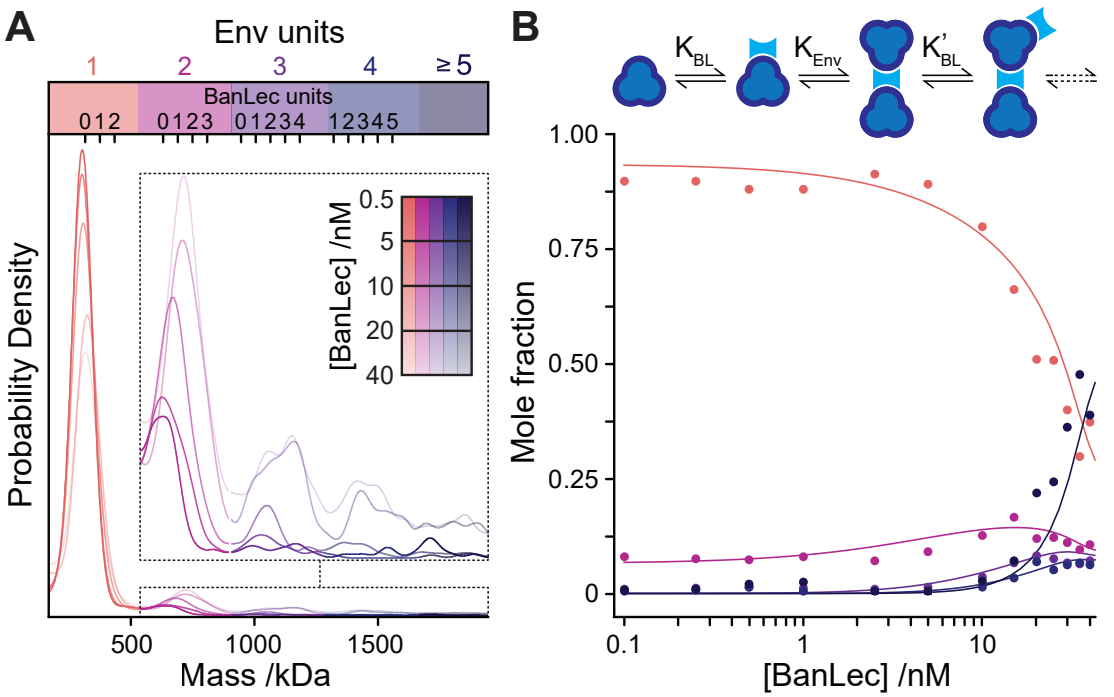


Figure 5.6: Env interactions with 0.1-40 nM BL. A. KDE plots of Env (10 nM) in the presence of 0.1-40 nM BL, with the expected positions for multiples of bound BL tetramers to multiple Env units indicated on the top. A zoomed-in view of the region above 500 kDa is shown in the inset, amplifying the low-abundance high MW species. The kernel bandwidth used for plotting was set to 30 kDa. **B.** Molar fractions of separate n·Env + m·BL species, coloured according to the Env units colour scheme in **A.**, plotted against BL concentration, where n is 1-4 or 5+ Env molecules, and m varies depending on n, from 0 to n+1. Modelled predictions are shown by the curves, created according to the model of which the simplified version is displayed on above the plot. Figure adapted with permission from the American Association for the Advancement of Science, from G. Young and N. Hundt *et al.*, Quantitative mass imaging of single biological macro-molecules, *Science*, **360**(6387), 423-427 (2018).[5]

of the Env–BL interaction and aggregation of Env on the viral surface have been suggested to strongly contribute to HIV-1 viral neutralisation by BL, these result provide key insights into the mechanism of BL anti-viral efficiency and could be used to improve our understanding of other viral systems.

5.4 Discussion

The aim of this Chapter was to improve our understanding of the Env–BL interaction at the single-molecule level in solution. It was shown that iSCAT is capable of

detecting glycoproteins, as well as characterising the interaction details between a soluble protein (BL) and a viral, medically important glycoprotein (Env) in a highly complex and oligomeric manner.

However, it is important to note that whilst there seemed to be linear increase of the mean MW of the small MW and Env dimer peaks with increasing BL concentration, this was associated with a large variation in the data for the individual concentrations. This proved challenging for in-depth analysis of the dataset presented in the Sections 5.3.1 and 5.3.2. At the higher end of BL concentrations, the mean MW appeared to be decreasing. This could be counter-intuitive, as with higher BL concentrations one would expect to detect more Env–BL complexes, resulting in a higher average MW. One of the possible explanations is that at high BL concentrations, a larger proportion of the protein forms aggregates with multiple species of both BL and Env in it. This crosslinking mechanism is illustrated in Figure 5.7A and B, showing the proposed Env–BL interactions in solution and on virions under conditions similar to *in vivo*, respectively.

Whilst this should theoretically be possible to test by analysing the data over a large range of MW values, in practice the many obstacles encountered during experiments at higher BL concentrations are prohibitive to this type of analysis, even though other properties at higher BL concentrations, such as the rate of formation of clusters, could be analysed. Similar to the extended experiments with 20 nM BL, it would be desirable to test the individual interactions more thoroughly by observing a larger number of particles over multiple experiments. Unfortunately, this can be time and material-consuming, with the possibility of no meaningful results. When observing interactions at concentrations above 20 nM, frequently the high amounts of BL protein reacted with Env to form large clusters in solution that would land and stick on the coverslip surface for the duration of the recording (Figure 5.4A),

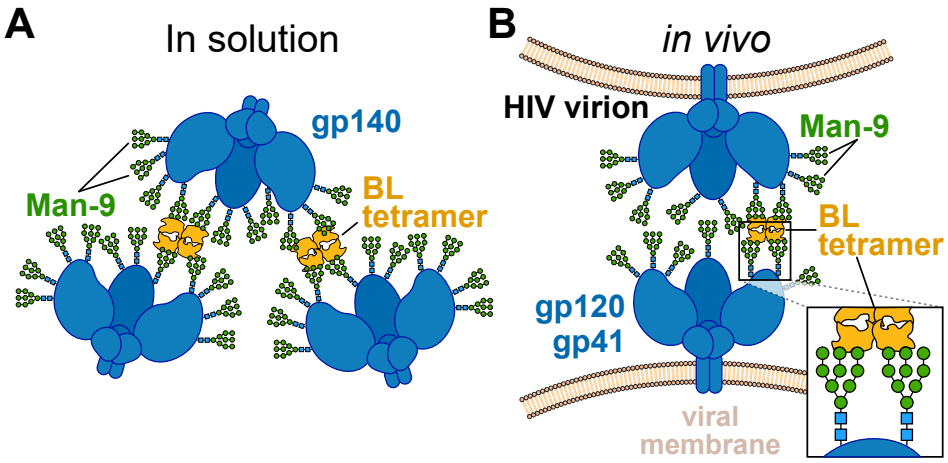


Figure 5.7: Proposed mechanism of BL crosslinking Env in solution and *in vivo*. **A.** Each BL monomer contains two carbohydrate-binding motifs, in total 8 per tetramer. The BL tetramer contains two opposite-facing dimers, each of which can bind to a different Env trimer, resulting in intermolecular crosslinking in solution. **B.** *In vivo*, BL has been shown to be able to crosslink individual virions through binding of two Env trimers anchored to the viral membranes of two different virions, resulting in viral particle aggregation.[301] Section **B.** of this Figure was adapted with permission from S. Lusvarghi *et al.*, Binding Site Geometry and Subdomain Valency Control Effects of Neutralizing Lectins on HIV-1 Viral Particles, *ACS Infectious Diseases*, **2**(11), 882-891 (2016).[301] Copyright © 2022, American Chemical Society.

complicating landing events observations and producing noise. Furthermore, these clusters likely bound large amounts of Env trimers, decreasing the frequency of landing events, regardless of a large cluster directly visible in the field of view. As such, the single-molecule sensitivity of iSCAT appears to limit the investigations of complex, aggregating systems. Experimental optimisations, such as limiting the variation in the time between sample insertion and the start of the recording, careful stock dilutions and recent pipette calibration could be implemented in future experiments to enable thorough particle count analysis. This could provide information about the total numbers of observable molecules present in the sample and their distribution amongst the different oligomeric species, enabling analysis of the loss of protein to large, hard-to-observe protein aggregates.

Using mass action law kinetics, the Env–BL interaction was investigated. The

K_{BL} values estimated from both the partial and complete datasets were consistent with recently reported values (0.19 nM^{-1} and 0.12 nM^{-1}, respectively, compared to 0.19 nM^{-1}, as determined by SPR[301]). Nevertheless, it is important to note that multiple time points after equilibration were not measured, and as such it is impossible to determine the K_{on} and K_{off} rates, and as such whether the system had reached true equilibrium. Regardless, the proximity of this value to the one determined previously by SPR suggests that the 5 min incubation was sufficient to reach equilibrium. Similar to previous studies, these values are approximately 10 × higher compared to the HIV-1 viral neutralisation IC_{50} values reported in the literature.[301,302,311–313] The ability of BL to induce inter-virion crosslinking, as illustrated in Figure 5.7B and observed recently,[301] could explain this difference in the equilibrium constant for the Env–BL interaction measured invariably in an isolated-component system *in vitro* and the IC_{50} values obtained using Env anchored to virions, where this type of virus crosslinking is possible, enhancing viral neutralisation.

The results of this project provide further directions into investigation of viral glycoprotein interactions and intermolecular dynamics. The results confirm the unique nature of BL–Env interaction at low nanomolar concentrations, with hopes for further development and use of BL as an HIV-1 prevention tool. With further testing and research, it could be possible for BL to be used similarly to a recently tested HIV-1 anti-retroviral agents embedded within a vaginal ring,[314] especially targeted for women in areas with high risk of HIV-1 infection and limited access to regular treatment or other prevention methods. Additionally, the results of this research are also encouraging for further development of other glycoprotein detection and analysis tools using iSCAT, which would enable large advances in this challenging but medically important field of research.

6

Spike-ACE2 Interaction in SARS-CoV-2

Contents

6.1 Introduction **152**
6.2 Materials and Methods **157**
6.2.1 Protein Expression and Purification 157
6.2.2 Mass Photometry 160
6.2.3 Molecular Docking 167
6.2.4 SARS-CoV-2 Neutralisation Assay 169
6.3 Results **171**
6.3.1 RBD Binds ACE2 Monomers and Dimers Equally . . . 171
6.3.2 ACE2 Dimers Crosslink Spike into Large Oligomers . . 173
6.3.3 Modelling Possible ACE2–Spike Interactions 179
6.3.4 SARS-CoV-2 is Strongly Inhibited by ACE2 but not mACE2 183
6.4 Discussion and Conclusions **185**

The work presented in this Chapter is a part of the following manuscript:

Anna Olerinyova, Roi Asor, Manish S. Kushwah, Fabian Soltermann, Lucas S. P. Rudden, Snežana Vasiljevic, Juliane Brun, Mario Hensen, Michelle Hill, Matteo T. Degiacomi, Nicole Zitzmann, Weston B. Struwe, Philipp Kukura.

The Spike-ACE2 Interaction Underlying SARS-CoV-2 Infection and Inhibition is Enhanced by Intermolecular Crosslinking, *in preparation.*

SARS-CoV-2 spike construct was kindly provided by Prof. Max Crispin (University of Southampton, UK). Spike glycoprotein (SARS-CoV-2), ACE2 and mACE2 (*H. sapiens*) were produced and purified by Dr. Mario Hensen and Dr. Snezana Vasiljevic (University of Oxford). Michelle Hill and Juliane Brun conducted the viral neutralisation assay and created Figure 6.10. Dr. Manish S. Kushwah (University of Oxford) kindly helped with experimental design, data collection and analysis. Standard free energy change minimisation was performed by Dr. Roi Asor (University of Oxford), who also created Figures 6.5 and 6.6A and B, adapted with permission. Molecular modelling was done by Dr. Lucas S. P. Rudden, producing data shown in Figures 6.8 and 6.9. We also thank Dr. Fabian Soltermann and Dr. Weston B. Struwe (University of Oxford) for their input in experimental design and planning.

6.1 Introduction

The first step in SARS-CoV-2 cell entry and COVID-19 infection involves binding of the viral envelope spike glycoprotein (further referred to as 'spike'), present on the surface of the virion, to the angiotensin-converting enzyme 2 (ACE2) receptor, attached through a single-pass transmembrane (TM) domain to the host cell membrane (Figure 6.1).[180,206,221] Due to its importance, this interaction has been studied extensively using a variety of biophysical and structural methods. At the same time, concerted global efforts resulted in the development of a number of COVID-19 vaccines at an unprecedented pace. The astonishing progress made in understanding SARS-CoV-2 infection mechanism and COVID-19 prevention, diagnosis, and treatment offers hopes for the end of the pandemic. Nevertheless, several challenges remain, including the low vaccination rates in developing countries and incomplete vaccine uptake in some developed countries. While almost two-thirds of the global population have been at least partly vaccinated by May 2022,[315]

the continuous evolution of the virus means that until near-complete eradication or wide-spread immunity is achieved, further waves of the COVID-19 pandemic remain a distinct possibility.

Like other RNA viruses, SARS-CoV-2 experiences highly frequent genetic mutations due to the low replication fidelity of its RNA-dependent RNA polymerase.[316] The neutral (non-deleterious) and selectively beneficial mutations tend to accumulate within the gene pool, facilitating the evolution of different viral strains. Occasionally, these changes might result in higher viral transmissibility, virulence, and pathogenicity, or confer resistance to existing vaccines and treatments.[317] For example, within a month, the Omicron variant of concern dominated the pandemic and outcompeted the previous Delta variant.[318] Its high transmissibility, rapid doubling time,[319] extensive resistance to neutralising antibodies and evasion of convalescent and vaccine immunity continue to raise concerns,[320–323] although the variant does not appear to increase disease severity in most cases.[324,325] Nevertheless, the disparity in regional vaccination rates could lead to the emergence of a more pathogenic, highly transmissible variant soon, reversing the progress achieved in combating this pandemic and pose an immense threat to human health.[317]

Consequently, further investigations of the mechanisms and regulatory details involved in SARS-CoV-2 infection remain important. Given its pivotal role in the COVID-19 infection, improved understanding of the spike–ACE2 interaction could lead to development of more efficient treatment options, potentially resistant to the aforementioned viral mutational escape. To date, most biophysical and structural studies of the spike–ACE2 interaction applied some form of simplification, often using a truncated, solubilised monomeric ACE2 (mACE2) and solubilised spike trimer or its isolated monomeric receptor-binding domain (RBD).[191,192,206,326] As such, this interaction has often been considered as a 1:1 reaction, despite the

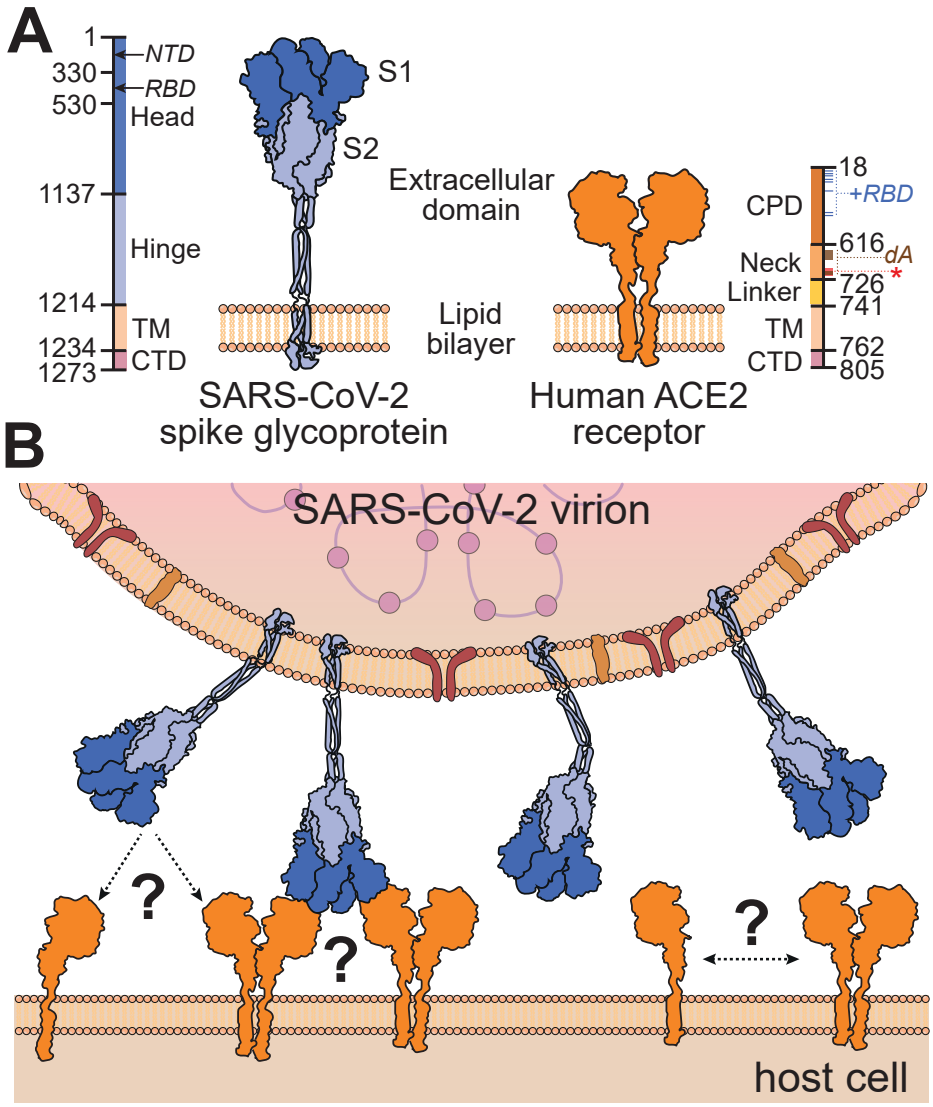


Figure 6.1: Possible interactions between membrane-anchored spike and ACE2. A. Cartoon structures of membrane-inserted SARS-CoV-2 spike glycoprotein (**left**) and human angiotensin-converting enzyme 2 (ACE2, **right**). The left block diagram shows the amino-acid positions of spike domains. *NTD* - N-terminal domain; *RBD* - receptor binding domain; TM - transmembrane domain; CTD - C-terminal domain. The right block diagram represents the amino-acids sequence of ACE2, showing the positions of its domains and linker. CPD - carboxypeptidase domain; Neck - ACE2 dimerisation domain; TM - transmembrane domain; CTD - C-terminal domain. Short blue lines indicate the approximate positions of RBD-binding residues in the CPD.[206] Red asterix shows the ADAM17 and TMPRSS2 peptidase cleavage region.[208] *dA* refers to the positions of the two α-helices in the neck domain responsible for ACE2 dimerisation.[206] Note the co-localisation of the peptidase cleavage site and the second *dA* section. **B.** Schematic of a SARS-CoV-2 virion approaching a host cell, showing potential interactions between spike and ACE2 proteins. Simplified contents of the virion and other proteins present in its membrane are depicted.

oligomeric nature of both binding partners leading to other possible reactions, indicated in Figure 6.1. Whilst this approach can be attributed to the technical and analytical challenges associated with the likely heterogeneity inherent to hetero-oligomeric reactions, the molecular details are of particular interest for refining our understanding of the viral infection process, potentially leading to improved prevention and treatment of COVID-19.

Different strategies have been employed to treat COVID-19 patients, including the development of various antivirals, neutralising antibodies, and immunomodulatory agents. However, the rapid evolution of novel SARS-CoV-2 strains can render most spike-targetting neutralising antibody therapies ineffective due to the high frequency of mutations in the RBD and N-terminal domain (NTD) as a consequence of the pivotal importance of ACE2 binding and the selective pressure of immune evasion.[327,328] Omicron is the most heavily mutated variant of concern detected so far, with over 30 spike amino acid alterations, including 15 in the RBD, resulting in higher ACE2 affinity and improved cell attachment,[329–332] together with evasion of vaccine-induced humoral immunity and perturbation of most antigenic sites recognised by neutralising antibodies.[330,333,334] Whilst various antivirals and immunomodulators are undergoing trials, an alternative approach has been to utilise the naturally strong spike–ACE2 binding through decoy ACE2 receptors neutralising the virus.[335–337]

Unsurprisingly, dimeric ACE2 (dACE2) was selected over its naturally secreted monomeric form as it exhibits much higher efficacy in suppressing SARS-CoV-2 infection, showing two orders of magnitude lower half-maximal inhibitory concentration (IC_{50}) than the monomer. Accordingly, wild-type (WT) ACE2 has demonstrated much tighter binding affinity to spike *in vitro* than engineered ACE2 lacking the dimerisation 'neck' domain (mACE2), exhibiting dissociation constant

(K_D) values of < 100 pM, compared to $\sim$20 nM for mACE2.[206] Given that ACE2 interacts with the isolated receptor binding domain (RBD) of spike with similar strength irrespective of its oligomeric state,[336] it is possible that the combination of dimeric ACE2 and trimeric spike is essential for the tight binding necessary for efficient infection. Oligomerisation of ACE2 has thus become one of the avenues in the development of ever more potent SARS-CoV-2 inhibitors. This strategy often targets multiple ACE2 binding sites per spike trimer, reaching pM affinities using mutagenesis and geometrically optimised oligomerisation to trimers.[336–339]

Whilst some of the engineered ACE2 decoys utilise ACE2–spike avidity to improve their spike binding affinity, often *via* connecting two mACE2 proteins with a linker, a single molecule of naturally dimeric ACE2 has not been reported to bind two RBDs on the same spike trimer. Separating the ACE2 monomers by a large enough linker appears necessary for this type of ACE2–spike avidity, perhaps due to the distance between individual RBDs. An alternative explanation for the exceptionally strong WT ACE2–spike interaction thus depends on ACE2-mediated spike crosslinking, a well-established concept in glycobiology as the cluster glycoside effect.[340,341] This type of interaction regime would be highly beneficial; pathologically, it would lead not only lead to radically enhanced binding affinity, but also receptor clustering on both membranes, promoting endocytosis. On the other hand, from a therapeutical perspective, a tightly binding dimeric decoy receptor that cannot bind to the same oligomeric target would rapidly lead to receptor clustering on the surface of the virus, creating an essentially irreversible lock on spike availability, strongly inhibiting infectivity. Interestingly, ACE2–spike crosslinking will only occur in the absence of ACE2–spike avidity, contrary to the strategy pursued recently for the ACE2 decoy receptor development.

However, characterising or even detecting such interactions is extremely challenging due to the associated heterogeneity. Crosslinking a trimeric protein by a dimeric binding partner can lead to a highly heterogeneous mixture of oligomeric states, making structural characterisation, which rests on identifying highly abundant species, near impossible. Ensemble-based affinity studies will reveal enhanced interaction strengths in such scenarios, but cannot resolve the underlying molecular mechanisms. As shown previously, mass photometry (MP) can be exceedingly proficient in quantifying the interactions and dynamics of protein assemblies without the need for protein labelling.[5–7,16] The potential of MP for measuring complex interactions has also been demonstrated in Chapter 5, involving the HIV spike glycoprotein (Env) being crosslinked by the banana lectin BanLec, where we could quantify successive interaction steps and stoichiometries that were impossible to resolve by other methods. Here, we use MP to resolve the molecular mechanisms and multivalent nature by which ACE2 dimers interact with the SARS-CoV-2 spike trimers, with important implications for the functional role of receptor assembly in viral infectivity.

6.2 Materials and Methods

6.2.1 Protein Expression and Purification

Cloning, protein expression and purification was performed by Dr. Mario Hensen and Dr. Snezana Vasiljevic (Zitzmann Lab, Glycobiology Institute, University of Oxford).

SARS-CoV-2 Spike Glycoprotein

The SARS-CoV-2 spike construct was generously provided by Prof. Jason McLellan (University of Texas, US), *via* the Crispin Lab (University of Southampton, UK) and is originally described in Wrapp *et al.*[191] Briefly, the gene encoding the spike ectodomain (residues 1-1208, GenBank: MN908947) was modified to include two

proline substitutions at the positions 986 and 987 and the removal of the furin cleavage site at residues 682-685 (a "GSAS" substitution), a C-terminal T4 fibritin trimerisation motif, and HRV3C protease cleavage site, a TwinStrepTag and an octahistidine tag. FreeStyle human embryonic kidney (HEK) 293F cells (Thermo Fisher) were then transiently transfected and grown in FreeStyle^{TM} MAX Reagent and Gibco OptiPRO^{TM} serum-free media following the manufacturers' protocol. After 4-5 days, the supernatant containing secreted spike was filtered using 0.45 and 0.22 μm filter and purified first by nickel affinity (Ni^{2+}-NTA) chromatography followed by immunoaffinity chromatography using an immobilised CR3022 column prepared in-house (cyanogen bromide activated Sepharose 4 resin, General Electrics (GE) Healthcare). The spike-bound column was washed with high-salt wash buffer (20 mM Tris-HCl, 500 mM NaCl, pH 8), and the protein was eluted using 3 M $MgCl_2$ (pH 7.2). The purified spike was immediately buffer exchanged into 10 mM Tris-HCl and 75 mM NaCl (pH 8) and further concentrated using a 30 kDa molecular weight cut-off (MWCO) centrifugal filter (GE Healthcare) at 3200 $\times$ *g*, 4 °C. Spike purity was assessed by SDS-PAGE (Figure 6.2A) and aliquots were flash-frozen in liquid nitrogen and stored at -80 °C.

Human ACE2 and Monomeric ACE2

The DNA sequence encoding for the *Homo sapiens* angiotensin converting enzyme 2 (ACE2, NCBI Ref.: NM_021804.3) amino acids 2-725 and 2-611, representing ACE2 and monomeric ACE2 (mACE2) respectively, were cloned into the expression vector pHLsec in the frame to the vector's N-terminal secretion leader peptide and a C-terminal hexahistidine tag using Gibson assembly (New England Biolabs). The expression construct was used to heat-shock transform competent *E. coli* DH5α cells (New England Biolabs), which were grown under antibiotic selection (1:2000 v/v kanamycin) at 37 °C for 16–18 h. Single colonies were picked, and the final

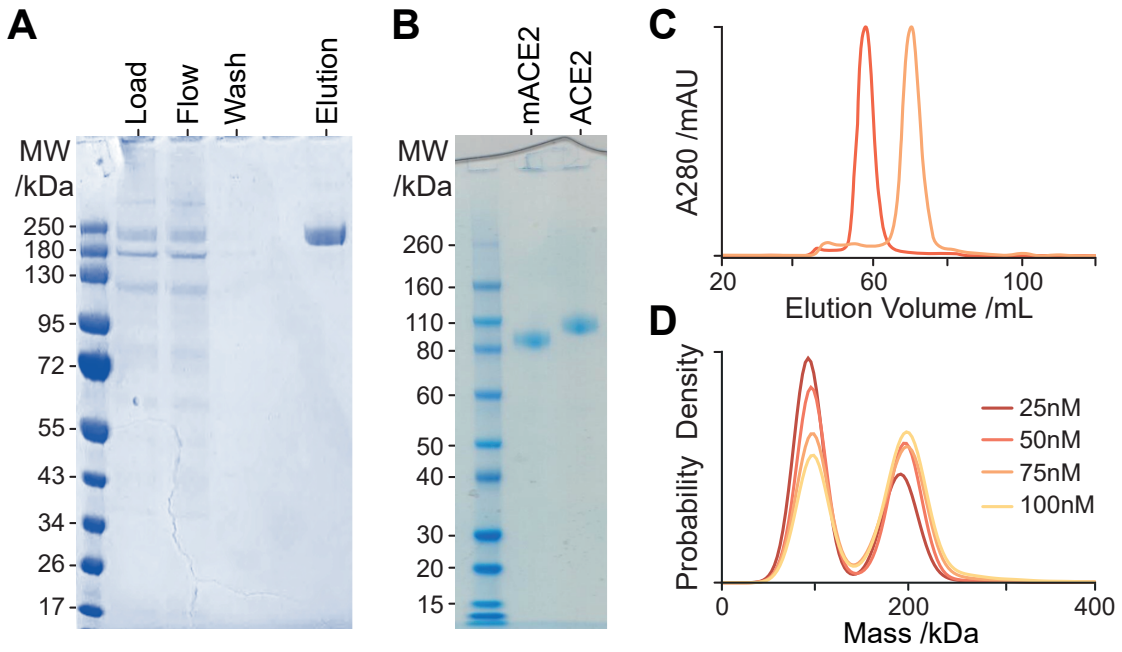


Figure 6.2: Characterisation of spike, ACE2, and mACE2. A. SDS-PAGE gel of samples collected during the final purification of spike, including column load, flow-through, wash sample and the final elute containing spike. **B.** SDS-PAGE gel of purified mACE2 and ACE2, showing clear difference in MW due to the cleaved neck region in mACE2 contruct. **C.** Size-exclusion chromatography 280 nm absorbance trace collected during ACE2 (dark orange) and mACE2 (light orange) purification, showing different elution volumes due to different oligomerisation states of ACE2 (dimer) and mACE2 (monomer). **D.** ACE2 was measured at 25-100 nM using mass photometry, with the resulting combined KDE traces plotted for each mass distribution. The data show increasing proportion of the ACE2 dimer with increasing concentration.

expression construct confirmed by DNA sequencing (Source Biosciences).

The pHLsec-ACE2 constructs designed above were used to transfect Freestyle HEK 293F cells for protein expression. HEK 293F cell culture with a cell density of 1.2×10^6 cells per mL was transfected with a total of 1 μg/mL plasmid DNA. Cell transfection was carried out using the transfection reagent FreeStyle MAX (Thermo Fisher Scientific) and OptiPRO serum-free medium (Thermo Fisher Scientific) following the manufacturer's protocol. Cells were incubated in vented Erlenmeyer flasks, shaking at 37 °C and 8% CO_2. The cell cultures' supernatant was harvested 4 days post transfection using centrifugation at 3000 $\times$ g for 30 min and the cell

pellet discarded. Phosphate buffered saline (PBS) was added to a final concentration of 1 × PBS and the pH adjusted to 7.4 using NaOH. The final sample was 0.45 and 0.22 μm filtered before further processing.

5 mL HisTrap excel column (GE Healthcare) was equilibrated with buffer A (1 × PBS, pH 7.4) and the filtered supernatant applied at a low flow rate. Subsequently, the column was washed with 10 column volumes (CV) of 2% buffer B (1 × PBS, pH 7.4, 500 mM imidazole) followed by a step-elution at 70% buffer B over 20 CVs. The elution was collected and fractions showing absorbance at 280 nm were analysed by SDS-PAGE to identify peaks containing ACE2. After SDS-PAGE, the desired fractions were pooled and the sample concentrated to a final volume of $<$ 5 mL using a 30 kDa MWCO Ultra-15 spin concentrator (Amicon) at 3200 × *g*, 4 °C. The concentrated sample was then applied onto a Superdex 200 pg 16/600 column (GE Healthcare), equilibrated with buffer A. Elution peak fractions containing the target protein, as determined by SDS-PAGE (Figure 6.2B), were pooled and the sample concentrated as described before. SEC elution profiles for mACE2 and ACE2 show distinct elution volumes indicating different protein volumes, and as such MW (Figure 6.2C). Following concentration, aliquots were flash-frozen in liquid nitrogen and stored at -80 °C for up to 1 year. The purified ACE2 was tested using mass photometry (MP) for purity and monomer-dimer equilibrium at different nM concentrations (Figure 6.2D).

6.2.2 Mass Photometry

Data Acquisition and Processing

Mass photometry (MP) measurements were conducted using glass coverslips and gaskets (Grace Bio-Labs reusable CultureWell^{TM} gaskets, well capacity of 3-10 μL). The glass coverslips (No. 1.5, 24 × 50, VWR) were prepared as described in Section 4.2.5. Prior to the start of data acquisition, the focus position was identified

and secured for the measurement. This was done with a focus feedback circuit which relies on total internal reflection using a reference beam.[4] For the spike–ACE2 interaction measurements, proteins were either mixed with buffer A (control) or together at μM concentration, and incubated at room temperature (RT) for a specified amount of time. The sample was then diluted to a desired nM concentration by rapid mixing in buffer, and 30 μL of sample was added to a gasket. Data acquisition was started within 14–18 sec after dilution, in order to prevent complex disassociation at nM concentrations. Data was recorded for 60 sec, and each experiment was repeated 2-4 times. For the RBD–ACE2 interaction experiments, proteins were diluted to their desired concentrations, mixed, and incubated at RT for 10 min prior to measuring. MP experiments were run on two similar mass photometer set-ups. For most of the experiments, a home-built mass-photometer was used, as described previously.[5] For some experiments, a commercial mass photometer, One^{MP} (Refyen Ltd), was used. The two set-ups are described in detail in Section 3.2. Mass calibration was performed using MS1000, kindly provided by Dr. Manish S. Kushwah (Kukura group, University of Oxford, UK). The contrast-to-mass calibration was conducted using the protocol described in Section 4.2.5. In some cases, separate calibrations were performed for morning and afternoon measurements due to contrast shift of the mass photometer used.

The movies acquired during protein landing assays were analysed with Discover^{MP} (Refeyn, version 2.1), as described in Section 3.2. The file containing a list of particles' position, time, contrast and MW was imported into MATLAB (version 2021b) and binding events were plotted as histograms or kernel density estimate (KDE) functions. For the count analysis of the individual spike oligomers and spike–ACE2 complexes, the embedded Gaussian fitting function in Discover^{MP} was used to extract the particle counts of individual peaks. Where peak fitting was prohibited by low particle numbers and peak width, the particle counts were estimated from the MW

range calculated for each oligomer. For higher oligomers, the total particle counts above spike tetramer were taken, and divided by the average MW of the particles in this range. For spike–ACE2 complexes, the oligomers were taken to contain spike and ACE2 in 1:1 ratio. Significance testing was performed using two-tailed t-test.

ACE2 Monomer-Dimer Equilibrium

At the nM concentration range measured, ACE2 is present as a mixture of monomers and dimers (Figure 6.2D). Prior to recording, ACE2 was diluted to its final measuring concentration and incubated for ∼5 min at RT, allowing the protein to reach presumed oligomeric equilibrium, although further experiments investigating the ACE2 monomer-dimer equilibrium over 120 min revealed continues decrease in dimer proportion after dilution to 40 nM (Figure 6.3). Given that for the majority of the following experiments, the samples were incubated for 5-10 min only, it is important to bear in mind that the ACE2 dimer was likely not in true equilibrium during these measurements.

Due to the respective MWs of ACE2 monomer (∼100 kDa) and dimer (∼200 kDa), they appear as well-resolved, separate peaks in mass histograms. As such, we can apply the in-built Gaussian fitting in Discover^{MP} to extract the peak counts, and convert these to ACE2 monomer and dimer concentrations, $[mA]$ and $[dA]$, respectively. Since significant increase in particle unbinding, mainly ACE2 monomers, was observed at higher ACE2 concentrations, the number of unbinding monomers and dimers in each recording was subtracted from their respective peak counts during the concentration conversion, which was performed *via* a conversion factor, f_A, as described for MP analysis by Soltermann *et al.*,[7] according to the following Equations:

$$f_A = \frac{[A]_t}{N_{mA} - uN_{mA} + 2 \times (N_{dA} - uN_{dA})}, \tag{6.1}$$

$$[mA] = f_A{*}(N_{mA} - uN_{mA}) \ , \quad [dA] = f_A * (N_{dA} - uN_{dA}) \tag{6.2}$$

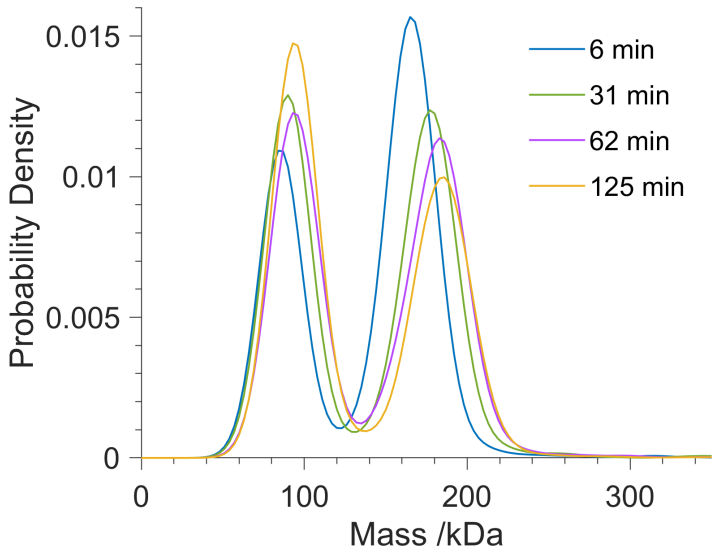


Figure 6.3: ACE2 distribution evolution over 120 min. ACE2 was measured at 40nM and plotted as a kernel probability density distribution after 6-125 min after dilution, during which the samples were kept on ice. The plotted lines are the average of two repeats. The offset in the measured MW is due to a strong focus drift of the instrument on the day of the measurement.

where $[A]_t$ refers to the total monomeric concentration of ACE2 measured, N_{mA} and N_{dA} represent the detected binding monomer and dimer counts, whilst uN_{mA} and uN_{dA} refer to the unbinding monomer and dimer counts, respectively. Using the calculated concentrations of ACE2 monomer and dimer, we can then estimate the ACE2 monomer-dimer dissociation constant, K_{dA}, with Equation 6.3:

$$mA + mA \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} dA \quad \Longrightarrow \quad K_{dA} = \frac{k_{off}}{k_{on}} = \frac{[mA]^2}{[dA]} \tag{6.3}$$

mACE2–RBD Interaction

For the mACE2–RBD reaction experiments, 10 nM mACE2 was mixed with 1, 5, 10, 20, or 30 nM RBD, and the sample was incubated at RT for ∼10 min, allowing the reactions to reach equilibrium. Due to the small MW shift caused by RBD binding (∼30 kDa), the Discover^{MP} Gaussian fitting was not applicable. Instead, peak fitting was performed as described in Section 5.2.3, using the Curve Fitting Tool (cftool, MATLAB, version 2021b), which minimises non-linear least

squares loss using gradient descent. To account for the rate of unbinding events increasing with higher RBD concentrations, binding events in the 0–250 range and unbinding events in the -250–0 kDa range were plotted separately as histograms using 5 kDa bin width, after which the unbinding histogram was subtracted from the binding particles' histogram. The resulting unbinding-adjusted histograms were then fitted to a sum of two Gaussian functions, with the exact function fitted by cftool shown in Equation 6.4:

$$f(x) = A_1 * e^{-\frac{(x-b_1)^2}{c_1^2}} + A_2 * e^{-\frac{(x-b_2)^2}{c_2^2}} \tag{6.4}$$

where A_1 and A_2 represent the fitted amplitudes, b_1 and b_2 are the fitted centroid positions (peak means, μ), and c_1 and c_2 are fitted parameters related to the peak standard deviation, σ, through a reciprocal of $\sqrt{2}$. During the fitting, the σ values were restricted to $\sim$13 $\pm$ 4 kDa, typical for soluble proteins of this MW measured by MP. The relative abundances of the different species were then calculated from the area fractions (AF) of their respective Gaussian fits, using the following relationship:

$$AF = \frac{A * \sigma}{1/2\pi} \tag{6.5}$$

Since the data fitted is plotted as histograms with 5 kDa bins, dividing the AF by 5 translates the AF values into fitted particle counts. From the counts, we can estimate the concentrations of free (apo) mACE2, $[mA]_a$, and the mACE2 $\cdot$ RBD complex, $[mA \cdot R]$, by calculating the proportion of bound or free mACE2 and multiplying it by the total mACE2 concentration measured. Subtracting the complex concentration from the total RBD concentration used in each condition yields the free RBD concentration, $[R]_a$. These can then be used to estimate the dissociation constant, K_{mAR}:

$$mA + RBD \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} mA \cdot RBD \quad \Longrightarrow \quad K_{mAR} = \frac{k_{off}}{k_{on}} = \frac{[mA]_a[R]_a}{[mA \cdot R]} \tag{6.6}$$

Next, the calculated fraction of RBD · mACE2 (complex fraction Θ) was plotted against the RBD concentration (Figure 6.4C). Given that mACE2–RBD interaction occurs between a single protein binding site (mACE2) and one ligand (RBD), the RBD titration data can be fitted with a binding curve, where Θ is described in terms of the total RBD concentration, $[R]_t$, and the dissociation constant, K_Θ, as indicated in Equation 6.7:

$$\Theta = \frac{[R]_t}{K_\Theta + [R]_t} \tag{6.7}$$

Note that whilst K_Θ describes the same interaction as K_{mAR}, it is determined using an alternative method.

Standard Free Energy Change Minimisation

The Gaussian fitting, standard free energy change minimisation and the resulting dissociation constant calculations described below were performed by Dr. Roi Asor (Kukura Group, University of Oxford). The results of these fittings and calculations are shown in Figure 6.5, which was created by Dr. Asor and adapted with permission.

Since the mACE2 + RBD, dACE2 + RBD, and dACE2 · RBD + RBD reactions involve identical binding sites, we can simplify the system to two interaction parameters, defined by Equation 6.3 (K_{dA}) and the following Equations:

$$mA + RBD \rightleftharpoons mA \cdot RBD \quad \Longrightarrow \quad K_R = K_{mAR} = \frac{[mA]_a[R]_a}{[mA \cdot R]} \tag{6.8}$$

$$dA + RBD \rightleftharpoons dA \cdot RBD \quad \Longrightarrow \quad \frac{K_R}{2} = K_{AR1} = \frac{[dA]_a[R]_a}{[dA \cdot R]} \tag{6.9}$$

$$dA \cdot RBD + RBD \rightleftharpoons dA \cdot 2RBD \quad \Longrightarrow \quad K_R = K_{AR2} = \frac{[dA \cdot R][R]_a}{[dA \cdot 2R]} \tag{6.10}$$

where the dissociation constant K_R is divided by 2 in Equation 6.9 to account for the presence of two RBD binding sites on an unoccupied ACE2 dimer. The standard (Gibbs) free energy change of the ACE2 monomer reaction with RBD, ΔG^0_R, and of the ACE2 dimer formation, ΔG^0_A, can then be used to describe the

reactions between mACE2, ACE2 and RBD. The standard free energy change at equilibrium, when $\Delta G = 0$, is defined in Equation 6.11:

$$\Delta G^0 = -\bar{R}\ T\ ln(K_{eq}C_0) \tag{6.11}$$

where $\bar{R}$ is the molar gas constant (8.314 J K^{-1} mol^{-1}), T is the absolute temperature during the measurement (293.15 K), K_{eq} is the equilibrium constant of the reaction, and C_0 is the reference state for the concentrations of the reactants and products, here chosen as 1 M. Rearranging for the ACE2 dimer formation and the RBD–ACE2 reaction dissociation constants (K_{dA} and K_R, respectively) yields the following Equations:

$$K_{dA} = 1M \times 10^9 \frac{[nM]}{[M]} e^{-\frac{\Delta G^0_A}{k_B T}}, \quad K_R = 1M \times 10^9 \frac{[nM]}{[M]} e^{-\frac{\Delta G^0_R}{k_B T}} \tag{6.12}$$

where $\times\ 10^9$ converts units from molar to nanomolar, and k_B represents the Boltzmann constant (1.38×10^{-23} J K^{-1}), relating the molar gas constant $\bar{R}$ to the number of molecules in 1 mol.

Each guess for the interaction parameters, given the total starting concentrations, results in a set of predicted concentrations for each resulting species, *i.e.* $[mA]_a$, $[mA \cdot R]$, $[dA]_a$, $[dA \cdot R]$, and $[dA \cdot 2R]$. These concentrations are then used to determine the species' relative proportions, from which Gaussian functions for each species can be determined, centred at their expected MW ± ∼10 kDa to account for instrument focus drift. The standard deviation is also restricted to a narrow range of feasible values, centred around ∼11 kDa. Plotting the resulting Gaussians and their combined curve allows fit comparison to the experimentally determined histograms for each set of RBD and ACE2 concentrations. The best fit is then found by minimising the reduced χ^2 values of the sum of the Gaussians for all conditions and repeats using the SciPy fsolve function, divided by the number of parameters, in

this case ΔG_R^0 and ΔG_A^0. After determining the best fits, the dissociation constants for the individual reactions can be calculated with Equations 6.3 and 6.8–6.10.

6.2.3 Molecular Docking

Analysis of ACE2 Binding to Two Distinct Spike Trimers

All molecular docking experiments described in this Chapter were conducted by Dr. Lucas S. P. Rudden and Prof. Matteo Degiacomi (University of Durham, UK). This approach was first used to test the possibility of two spike glycoproteins, anchored to the viral membrane, binding to the same ACE2 receptor, inserted in the host cell membrane. This is dependent on three major factors: (1) the spike conformation, (2) the positioning of the SARS-CoV-2 virion over the ACE2 receptor and host cell membrane, and (3) the shape of the viral membrane in the immediate vicinity of the spike proteins. A crystal structure of the SARS-CoV-2 spike RBD bound to the soluble domain of mACE2 (PDB: 6M0J) was used to generate this model. This structure was duplicated and aligned *via* the mACE2 atoms to the chains of an ACE2 homodimer cryoEM structure (PDB: 6M18, in complex with B0AT1), providing a model of the complete dimeric ACE2 receptor with two bound RBD domains. The flexibility of SARS-CoV-2 spike in this model was represented by adapting a 4.2 μsec long simulation (produced and made available by the Amaro group, University of California, US). This simulation features a fully glycosylated spike with one RBD in the open state, inserted in a lipid bilayer.[196] Each of these 4200 nsec long simulated spike conformations was duplicated and aligned through the open RBD domain (chain A) to the template RBD domains bound to the ACE2 homodimer.

The relative positions of the ACE2 receptor and the two bound spikes enabled inferring the relative positions of the virion and the membrane curvature between the transmembrane (TM) regions of these two spikes for each spike conformation.

The reference frame was defined as the ACE2 host cell membrane, with the z-axis representing the membrane plane normal, and the origin being the mean position of the membrane entry and exit atoms of the ACE2 receptor (Cα-atoms of residues 741–761). The viral membrane orientation at the TM regions of the two spike was then defined by the axis connecting the mean positions of the spikes' membrane entry and exit residues (Cα atoms of the residues 1214 and 1234 for every chain of the spikes). Correcting these axes by the observed tilt of the TM region yielded normal vectors, enabling the definition of the planes parallel to the simulated membrane of the virus. Lastly, the viral membrane shape was defined according to the arc bridging the two virion membrane planes, assuming a constant membrane curvature between the two TM regions. From these definitions, the attack angle of the virion was derived based on the axis connecting the origin (TM entry and exit of the ACE2 receptor) to the midpoint of this arc between the two spike proteins, described in spherical coordinates (azimuth Θ and elevation ψ). The membrane curvature was represented through a scalar that defined the magnitude and direction of a vector along this axis, negative for a concave and positive for a convex viral membrane. The length of the arc connecting the two spikes' TM regions enabled measurements of the distance between them. Further mathematical details outlining this model are described by L. S. P. Rudden.[342]

Modelling of Multiple ACE2 Bound to Crosslinked Spikes

The open chain of the SARS-CoV-2 spike 6M18 structure was aligned onto the remaining closed chains, however this yielded no conformations permitting interactions with more than one ACE2. As such, in order to model the binding of multiple ACE2 receptors to a spike trimer and crosslinked dimer of trimers, a SARS-CoV-1 spike structure with two chains in the open RBD state (PDB: 6CS1) was used, showing 63% homology to the SARS-CoV-2 spike structure. A coordinate transfer of the open state chain A onto the closed chains was performed, and the structure

was mutated into the SARS-CoV-2 spike using the Modeller program.[343] Motifs of up to 10 residues long that were missing were patched, while the structure was frozen, preventing any optimisation further from the desired conformational state. This new structure allowed the alignment of one ACE2 per open RBD without clashes, resulting in a model of a spike bound to three ACE2 dimers.

The plausibility of oligomeric states including cross-linked spikes was examined by aligning this homology model onto a SARS-CoV-2 spike dimer (PDB: 7JJJ). Using the methods described above, it was possible to align five ACE2 receptors onto the spike dimer free binding sites, producing no clashes. The required cellular membrane curvature was calculated by measuring the curvature between two ACE2 bound to the same spike protein, performed as outlined in the previous Section. As the curvature between the two ACE2 receptors was only 0.01 nm^{-1}, the two ACE2 TM stalks were approximated as parallel and their average principal axes centred at the midpoint were taken as the starting point to calculate the host cell membrane curvature. When calculating the curvature of the 2 spike $\cdot$ 5 ACE2 oligomer, the principal axis of the central ACE2 TM stalk was defined as the normal to the intersection point, rather than calculating the intersection of the two planes. This calculation was conducted on one frame and not a simulation, resulting in the absence of a standard variation of the reported value.

6.2.4 SARS-CoV-2 Neutralisation Assay

Focus-forming assay was used to test anti-viral effects of solubilised ACE2 and mACE2 against SARS-CoV-2.[344] SARS-CoV-2 was generated by infecting the African green monkey kidney cell line Vero E6 cells at a multiplicity of infection of 0.01 in virus Dulbecco's Modified Eagle Medium (Thermo Fisher Scientific) containing 2% heat-inactivated fetal bovine serum (Thermo Fisher Scientific). Viral particles were harvested 72 h post infection. Viral supernatant for analysis was

prepared in half-logarithmic dilutions in viral propagation medium, and 20 μL of every dilution was loaded into a 96 well plate (Greiner Bio-one), with 4 $\times$ repeats. Next, 100 μL Vero E6 cells (0.45×10^6 cells/mL) in virus propagation medium were plated into each well. After incubation for 2 h, 100 μL 2% carboxymethyl cellulose overlay (1:1 mix of 4% carboxymethyl cellulose in H_2O and virus propagation medium) was added to the cells followed by incubation for 24 h at 37 °C, 5% CO_2. Next, the overlay was extracted, and the cells washed with PBS (Gibco), followed by fixing using 4% paraformaldehyde (Alfa Aesar). Following 30 min incubation, the paraformaldehyde was removed, and the reaction was suppressed with 1% ethanolamine in PBS (Merck).

For cell permeabilisation, 2% Triton-X100 in PBS (Merck) was applied for 30 min at 37 °C. The plates were rinsed 3 $\times$ with wash buffer (0.1% Tween-20 in PBS), after which EY2A anti-N human monoclonal antibody (mAb) at 10 pM in wash buffer was added, kindly gifted by Arthur Huang (Chang Gung Memorial Hospital, Taiwan) *via* Alain Townsend and Jack Tan (Weatherall Institute of Molecular Medicine, University of Oxford). The plates were then incubated with rocking at RT for 1 h, after which they were rinsed 3 $\times$ with wash buffer. Secondary antibody anti-Human IgG (Fc-specific)-peroxidase-conjugate produced in goat (Merck) was then diluted 1:5000 in wash buffer and added to the plate, after which the plates were incubated with shaking at RT for 1 h. The plates were then washed 3 $\times$ with wash buffer. Afterwards, TrueBlue peroxidase substrate (Insight Biotechnology Ltd) was introduced to the wells and rocked at RT for 10 min. Next, the substrate was discarded, and the plates rinsed with double deionized H_2O for 10 min. The water was discarded, and the plates allowed to air-dry. The foci were quantified with an ELISPOT classic reader system (AID GmbH).

6.3 Results

6.3.1 RBD Binds ACE2 Monomers and Dimers Equally

In order to investigate the dynamics of the spike–ACE2 interaction, we first characterised the basic building blocks involved in this reaction using two engineered forms of ACE2. The first one, further referred to as 'ACE2', lacks the CTD and TM section (ACE2Δ727_F805, see Figure 6.1A, right), yielding solubilised ACE2 capable of dimer formation. In the second ACE2 version, the dimer-forming neck domain was also removed (ACE2Δ612_F805), creating a soluble, monomeric ACE2 form (referred to as 'mACE2'). Both mACE2 and ACE2 eluted as single peaks by size exclusion chromatography (SEC) with ACE2 arriving at a later point (Figure 6.2C), indicating the presence of purely monomeric and dimeric species at μM concentrations present during SEC. MP measurements of ACE2 at nM concentrations confirmed the presence and concentration-dependent proportion of monomers (∼92 kDa) and dimers (∼185 kDa), indicative of nM binding affinity (Figure 6.2D). Applying the single-shot dissociation constant determination method (SSKD) developed by Soltermann *et al.*,[7] we can determine the ACE2 monomer-dimer dissociation constant, K_{dA}, to be 12.2 ± 3.4 nM. Interestingly, whilst the related ACE has been reported to form dimers in cells,[345] and a full-length ACE2 dimer structure has been determined in complex with B^0AT1,[206] to the best of our knowledge, comparative estimations of the binding strength of this interaction have not been determined before.

Next, we explored the binding between mACE2 and the receptor binding domain (RBD) of the spike. Upon mixing mACE2 and RBD at different ratios, we found a clear shoulder that grew with increasing RBD concentrations and disappeared at lower ratios (Figure 6.4A). By fitting the mass distributions to a sum of Gaussian peaks (Figure 6.4B), we determined the dissociation constant K_{mAR}

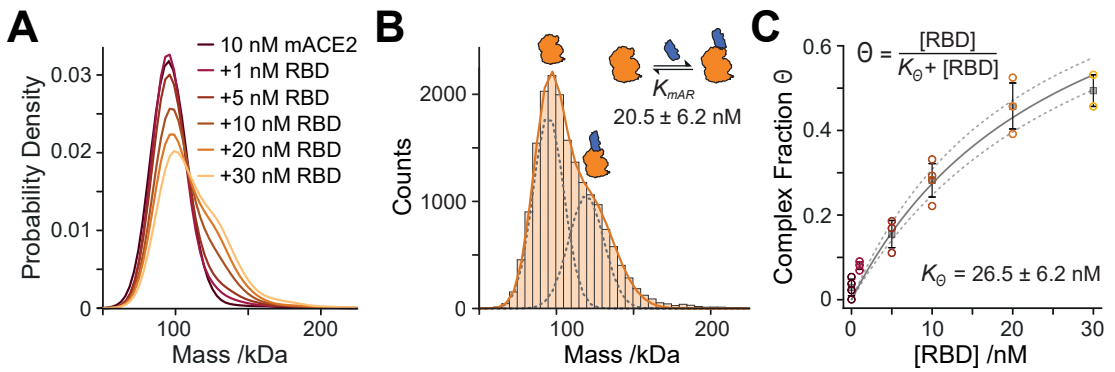


Figure 6.4: Analysis of the mACE2–RBD interaction. A. 10 nM mACE2 was measured with a range of RBD concentrations (1-30 nM), resulting in an appearance of a ∼30 kDa higher MW shoulder indicating more frequent RBD–ACE2 complex formation with increasing RBD concentration. **B.** The sum of two Gaussians (dotted lines) was fitted to the combined 10 nM mACE2 + 20 nM RBD mass histogram (orange) using gradient descent, with the K_{mAR} for the interaction calculated using the single-shot K_D method.[7] **C.** The mACE2–RBD complex fraction Θ increases with RBD concentration. The data were fitted to a standard binding curve, with K_Θ referring to the RBD–mACE2 dissociation constant. The mean ± standard deviation of the fitted curve are represented by the solid and dashed lines, respectively.

of this interaction to be 20.5 ± 6.2 nM using the SSKD method described for K_{dA}. Additionally, standard binding curves were fitted to the occupied protein fraction (complex fraction Θ) plotted against the RBD concentration, yielding a dissociation constant K_Θ for the mACE2–RBD interaction of 26.5 ± 6.2 nM. Comparing K_{mAR} and K_Θ, we can conclude that both dissociation constant determination methods yield approximately equivalent values, in agreement with previous results.[7] Furthermore, this value is in good agreement with previous estimations for the mACE2–RBD binding obtained by surface plasmon resonance (14.7 nM) and bio-layer interferometry (1.2 nM).[191,192]

In order to fully characterise the ACE2–RBD interaction, the number of reactions involved would be four, including the ACE2 monomer-dimer equilibrium, the monomer–RBD reaction, and the ACE2 dimer binding to the first and second RBD, respectively (see Equations 6.3 and 6.8–6.10). However, since the binding sites involved in all the RBD–ACE2 interactions are identical, we can simplify the

system and characterise it through fitting two interaction parameters. Through simultaneous standard free energy change minimisation for the ACE2 monomer-dimer and the mACE2–RBD reactions, labelled as ΔG^0_A and ΔG^0_R, respectively, we can predict the resulting species' concentrations given the starting concentrations of the reaction components and mass conservation. The experimental data and the resulting best fits for all measured RBD concentrations are shown in Figure 6.5A, determined by reduced χ^2 minimisation (Figure 6.5B).

Using this method, we can determine the dissociation constants for each of the four reactions (Figure 6.5C). The ACE2 monomer–dimer dissociation constant K_dA is 11.7 ± 1 nM, in excellent agreement with the SSKD-determined value of 12.2 ± 3.4 nM. The dissociation constants for the mACE2–RBD and dACE2 · RBD–RBD reactions, K_{mAR} and K_{AR2}, respectively, appear to be identical (28.2 nM) and in excellent agreement with the K_{mAR} values determined by the binding curve fitting (26.5 nM) and SSKD (20.5 nM) methods, whilst the dACE2–RBD K_{AR1} is halved, in accordance with twice the number of RBD binding sites available. Based on these experiments and calculations, we can conclude that the dissociation constants for both ACE2 dimerisation and ACE2–RBD interactions are indistinguishable within the experimental error, indicating independent binding of RBD to ACE2 subunits. The confirmation of these results through three different K_D determination methods further solidifies these findings and suggests the validity of these approaches.

6.3.2 ACE2 Dimers Crosslink Spike into Large Oligomers

Having characterised the basics of the system, we proceeded to measure the more complex spike trimer. Purified spike at 82 nM solution monomeric concentration exhibits a major peak of ∼550 kDa, as predicted for the glycosylated assembly (Figure 6.6A, top). Additionally, we observed oligomers of trimers, in line with cryoEM reports.[346] Addition of 50 nM mACE2 redistributes the original single

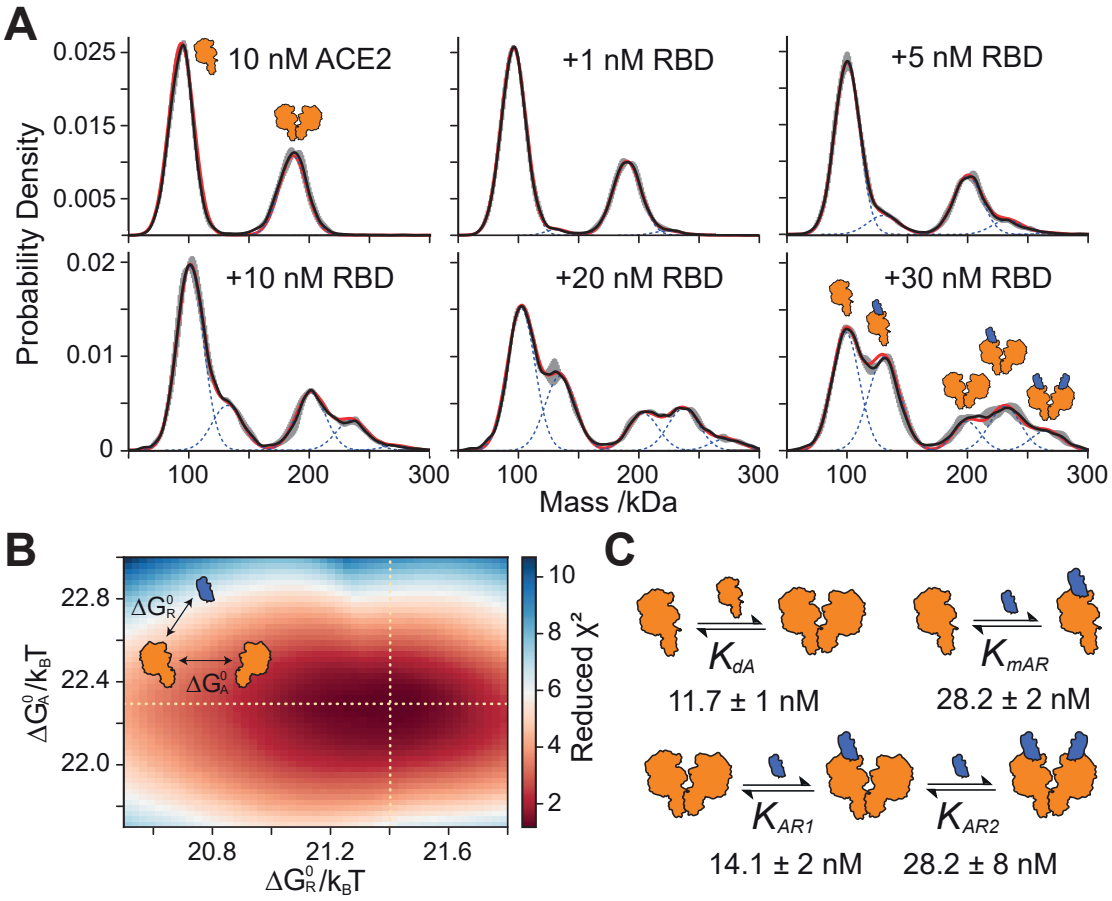


Figure 6.5: Mass distributions and binding energy analysis of the ACE2–RBD interaction. A. 10 nM ACE2 was measured by MP with a range of RBD concentrations (1-30 nM), resulting in the appearance of a higher MW shoulder in the monomer section of the mass distribution, indicative of RBD binding. Upon RBD binding to the ACE2 dimer, the first higher MW shoulder appears at lower RBD concentration, followed by the appearance of a second higher MW peak, representing the binding of a second RBD molecule to the ACE2 dimer. The peaks of 10 nM ACE2 control and + 30 nM RBD condition are labelled with cartoon representations of the ACE2 and RBD proteins. The black line represents the combined averaged KDE plots, with the data variation shown by the grey shaded areas around the curve. The Gaussian fits are shown by dashed blue lines, and are combined in the red line. **B.** A heat map of reduced χ^2 minimisation as a function of different $\Delta G^0_A/k_BT$ and $\Delta G^0_A/k_BT$ values, with the lowest value indicated by the intersection of the yellow dashed lines. **C.** The dissociation constant values determined for different interactions from the binding energy minimisation performed in part **B.** Displayed are the dissociation constants for the ACE2 monomer-dimer equilibrium (K_{dA}), the mACE2 + RBD interaction (K_{mAR}), the dACE2 + RBD binding (K_{AR1}), and the dACE2·RBD + RBD interaction (K_{AR2}).

spike feature into four peaks representing the apo spike, spike + mACE2, spike + 2 mACE2, and spike + 3 mACE2 species (Figure 6.6A, middle inset). We observed similar behaviour for the spike dimer of trimers, where we could identify binding of up to 5 mACE2 molecules (Figure 6.6A, middle inset). Upon the addition of 50 nM ACE2 to 82 nM spike, we observed spike binding predominantly to one ACE2 dimer, with smaller amounts of ACE2 monomer and 2 × ACE2 dimer observed, as indicated in Figure 6.6A, bottom inset.

To further investigate the spike–ACE2 interaction, spike 50nM was measured with 10, 50, and 100 nM ACE2 (Figure 6.6B). Whilst the 10 and 50 nM ACE2 data resemble that of 82 nM spike with 50 nM ACE2, the 100 nM ACE2 data displays a small peak at ∼1150 kDa, potentially indicating 3 ACE2 dimers binding to a single spike trimer. Interestingly, even though this interaction has been previously documented to occur with mACE2,[180,326] and was detected in our mACE2–spike MP experiments (Figure 6.6A, middle inset), the existence of this species would suggest the absence of intra-spike avidity, meaning that ACE2 dimer cannot engage two RBDs on the same spike simultaneously, an important implication for understanding the molecular dynamics and strength of the spike–ACE2 interaction.

Additionally, during the ACE2–spike MP measurements, the appearance of large, MDa-sized species was observed, absent in the spike or spike + mACE2 controls. Particles above the MW of the aforementioned spike oligomers (∼3 MDa) were rarely observed in the spike 82 nM recordings (Figure 6.6C), in contrast to the spike + ACE2 movies containing species sometimes above 10-15 MDa (Figure 6.6D). This behaviour is reminiscent of the BL-induced Env oligomerisation described in Chapter 5, suggesting that the large observed particles are oligomeric ACE2–spike complexes formed as a result of ACE2–mediated spike crosslinking.

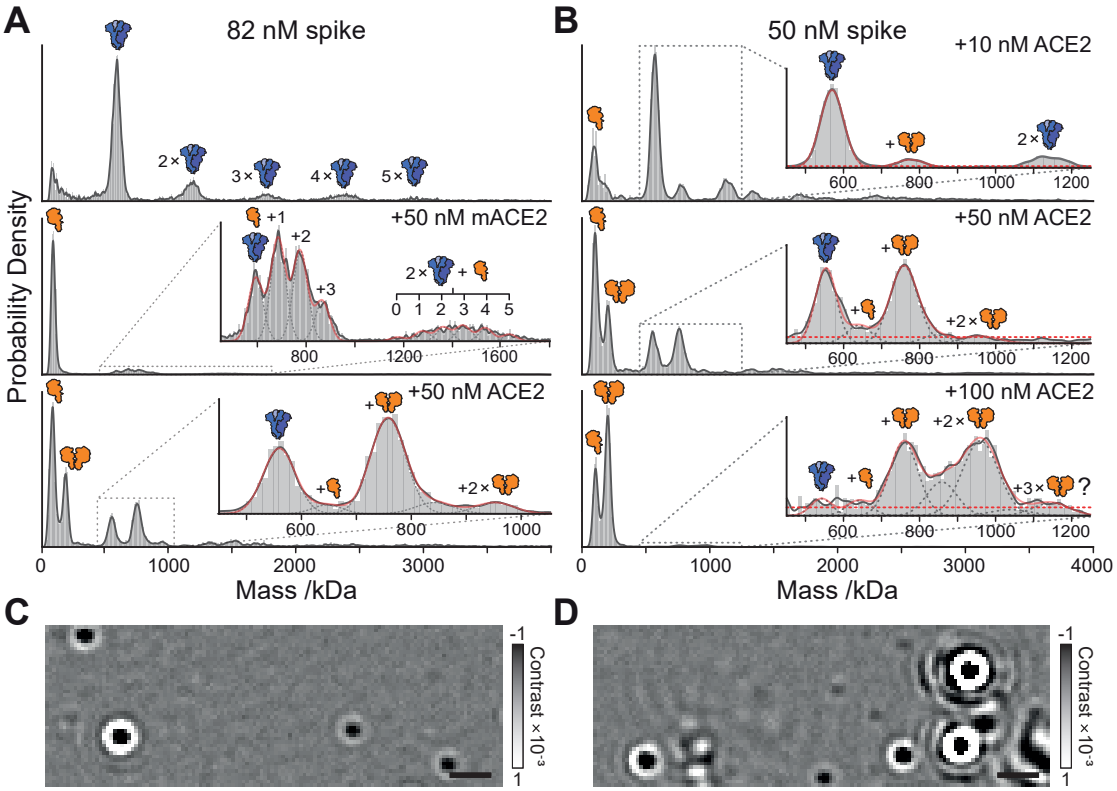


Figure 6.6: Mass photometry analysis of spike–ACE2 interaction. A. MP measurements of 82 nM spike, either alone (top), with 50 nM mACE2 (middle) or ACE2 (bottom). The MW values of the peaks of spike 82 nM are 550, 1090, 1600, 2130, and 2690 kDa. The MW of the spike + 0-3 mACE2 are580, 675, 760, and 845 kDa. The MW of the spike + ACE2 peaks are 565, 760, and 960 kDa. The dark grey curves indicate the averaged values from 2-4 repeats. Cartoon representations of spike, mACE2, and ACE2 are used to label the different species present. The middle inset magnifies the 400–1800 kDa section, highlighting up to 3 mACE2 binding to a spike trimer and up to 5 mACE2 binding to a spike dimer of trimers. The bottom inset magnifies the 450–1050 kDa region, showing the binding of 1–2 ACE2 monomers and up to two ACE2 dimers to a spike trimer. The dashed curves show the fitted Gaussian functions. **B.** Mass histograms of 50 nM spike with 0-100 nM ACE2. Insets show magnified regions (450–1250 kDa), with spike and spike + n·ACE2 labelled using cartoon representations. The dark grey lines indicate the averaged, combined data, whilst the dashed grey lines indicate the positions of the Gaussian fits to the individual species. The faint red line shows the combined fitted functions, and the dashed red line indicates the baseline particle levels present between peaks. **C.** An example frame from an MP recording of 50 nM spike. The particles shown are ∼2.5 and 0.5 MDa. **D.** An example frame from an MP recording of 50 nM spike + 100 nM ACE2, showing protein aggregation and large particles of ∼16 and 19 MDa.

In addition to the large particles described above, a significant drop in the overall number of smaller spike-containing complexes was observed. Figure 6.7A and B illustrates the significant decline in the counts of species containing single spike trimer or a dimer of trimer, respectively, upon 50 nM ACE2 addition to 82 nM spike, as compared to spike alone or the 50 nM mACE2 control. Interestingly, this decline does not translate to spike trimers and tetramers of trimers, perhaps as a result of low counts and relatively high variation between repeats. Alternatively, the loss of comparatively smaller spike oligomers to larger complexes *via* crosslinking could be compensated by an associated increase in spike trimer and tetramer of trimers-containing species formation through the same mechanism. Unfortunately, whilst the single-molecule resolution achieved with MP enables the analysis of the overall oligomeric distribution, it does not allow the characterisation of the molecular arrangements or bonds present in these complexes.

In order to quantify the observed high MW oligomers and their distribution, the mass distributions of complexes heavier than spike tetramer of trimers (∼2200 kDa) and equivalent spike–mACE2 and spike–ACE2 complexes are shown as violin plots in Figure 6.7C. The addition of both ACE2 and mACE2 resulted in a clear shift towards higher MW due to the added MW of mACE2 and ACE2 molecules in complexes. Accounting for the added MW of mACE2 at ∼1:1 spike–mACE2 ratio, the calculated mean and median values of the spike control and spike + mACE2 distributions are in good agreement. However, when the same analysis is applied to the spike–ACE2 distributions, the adjusted mean and median remain significantly higher. Finally, the overall proportions of spike in the form of a trimer decreases significantly upon ACE2 addition, whilst the proportion in the form of oligomers larger than tetramer of trimers increases compared to the spike or + mACE2 controls (Figure 6.7D), correlating the loss of spike trimers to higher oligomeric species. We can thus conclude that the formation of these large particles is a direct result of the

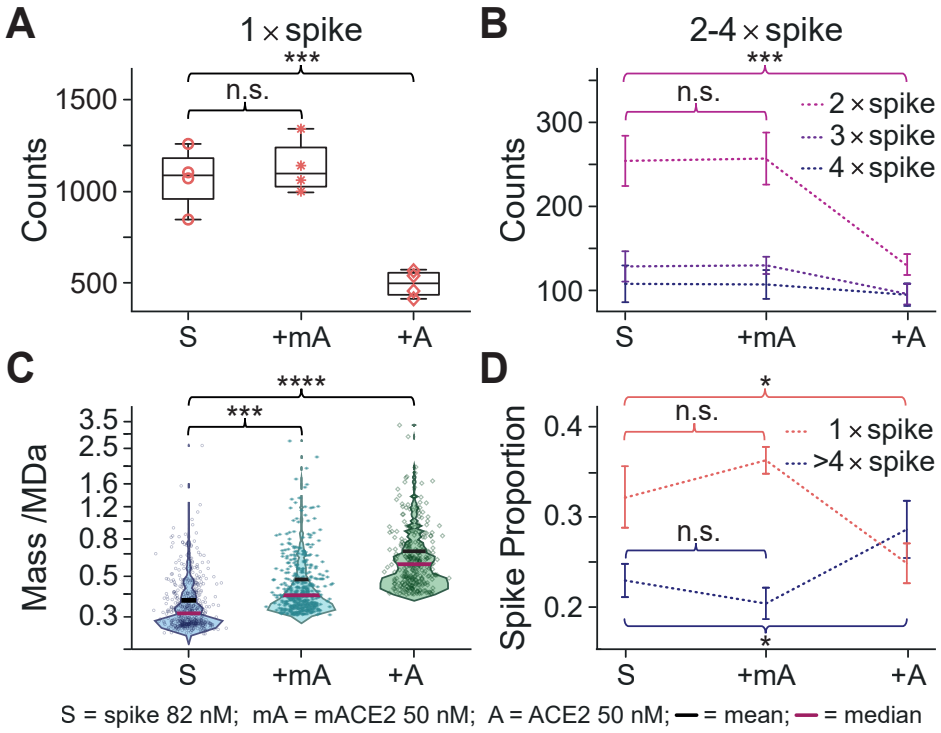


Figure 6.7: Particle distribution and count analysis of the spike–ACE2 interaction. Comparison of oligomeric species counts, high MW particle distributions, and spike proportions between spike (82 nM, S), spike + 50 nM mACE2 (+mA) or + 50 nM ACE2 (+A). **A.** Counts of species containing spike trimer decrease significantly upon ACE2 addition, but remain unchanged when mACE2 is added under the same conditions. **B.** Count comparison of species containing higher oligomers of spike (dimers, trimers, and tetramers of trimers) showing a significant decrease of spike dimers of trimers with ACE2 as compared to mACE2, and no change in the counts of the other evaluated oligomers. **C.** The distributions of high MW particles differ significantly between each tested condition. Detected particles with MW above the spike tetramer of trimers (or equivalent complexes containing mACE2 or ACE2) were plotted as violin plots, displaying each distribution overlayed with a scatter plot of the included particles. The mean and median MW for each distribution is indicated. **D.** The proportions of spike trimers and higher oligomers (above tetramer of trimers) out of the total detected particles were evaluated. Whilst there is no change between the spike and + mACE2 controls, the proportion of spike trimers is significantly lower upon ACE2 addition, whilst the proportion of higher oligomers is higher. Significance testing for all plots was performed using two-tailed student t-test. n.s. - not significant; $* \ p \leq 0.05$; $** \ p \leq 0.01$; $*** \ p \leq 0.001$; $**** \ p \leq 0.0001$.

dimeric nature of ACE2 and is likely related to the observed depletion of the low MW spike complexes, potentially through ACE2-mediated inter-spike crosslinking.

6.3.3 Modelling Possible ACE2–Spike Interactions

For this type of crosslinking to occur, the geometric orientation of the proteins involved must enable the binding of a free monomeric subunit of a spike-bound ACE2 dimer to a different spike trimer. Furthermore, this spike–ACE2–spike interaction must be favoured over the free ACE2 subunit binding to the same spike through an adjacent RBD. However, this ACE2-spike avidity would only be possible if the position and orientation of the free ACE2 allows it to reach an RBD in an 'up' position on the same spike. To test this structural requirement, we built molecular models, leveraging an atomic structure of the RBD bound to ACE2 dimers (PDB: 6M18) and six available molecular dynamics simulation replicas (totalling 4.2 μs of simulation time) of spike trimers with their RBD in the 'up' state (PDB: 6VSB) embedded in a lipid bilayer.[196] We aligned the RBD of each simulated spike conformation onto that of the RBD bound to the ACE2 dimer. Specifically, we first removed the linkers connecting the spike trimer to the RBD subunit (P330–L335 and G526–K537). We then defined four distance constraints to minimise during docking using the Parallel Optimized Workbench (POW)[347] between F329 and C336, and C525 and C538 for the two RBD–spike intramolecular sites. The minimum distance identified by POW was 26 ± 3 Å for the 5 residue P330–L335 linker, and 42 ± 6 Å for the 11 residue G526—K537 linker (Figure 6.8A). Given the maximum span distance of these two linkers,[348] 12 Å and 30 Å respectively, we find that a position that facilitates dual binding is not possible without significant conformational flexibility. These results suggest that a single spike should be unable to bind both subunits of the same ACE2. Together with the absence of a structure showing this type of spike–ACE2 interaction arrangement,

we can conclude that this form of ACE2–spike avidity is unlikely.

We then investigated whether two different spikes could bind the same ACE2, aligning pairs of MD-generated conformations so that available RBDs bind one subunit of ACE2. For all available protein conformation, models could be built without clashes between spikes (Figure 6.8B). For each of our models, we confirmed whether such dual binding could also take place *in vivo*, accounting for viral membrane shape, and observed that the curvature between the two TM domains of spikes should range between 0.099 nm^{-1} and -0.074 nm^{-1} (Figure 6.8C). Thus, 14.3% of the simulation time is spent in conformations corresponding to the nominal shape of the virus,[349] with the remaining curvatures requiring plausible deformations in the local membrane topology. Distances between spikes along the membrane connecting them ranged from 16.2 nm to 58.9 nm, consistent with distances measured with electron tomography (10 nm and 80 nm).[350]

We then addressed whether a single spike can bind multiple ACE2 dimers simultaneously (Figure 6.9A, top) using the existing SARS-CoV-1 spike structure (PDB: 6CS1),[351] which has 63% homology with SARS-CoV-2 and features two RBD domains in the 'up' state at a greater angle, consistent with SARS-CoV-2 cryoEM data.[326] Using homology modelling and coordinate transfer, we created a SARS-CoV-2 spike with three open domains and aligned an ACE2 per RBD domain to produce a clash free model (Figure 6.9B). This structural assembly would require the three ACE2s to be bound to a membrane with a plausible local curvature of 0.003 nm^{-1}.[352] Given that two spikes can bind the same ACE2, a single spike can bind three separate ACE2s, and spikes are known to self-assemble on the cell surface,[346] there exist several possibilities for the formation of higher order oligomers.

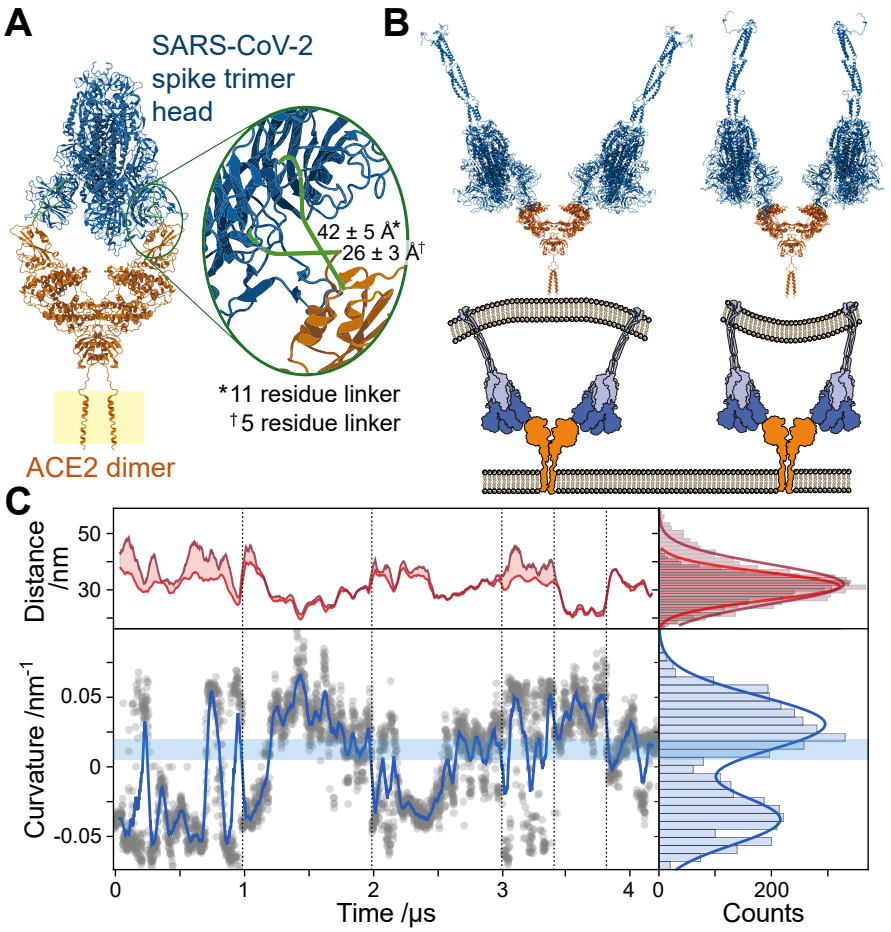


Figure 6.8: MD analysis of possible spike interactions with a single ACE2. A. Simulations of a spike trimer with two RBDs in the 'up' position (PDB: 6VSB) show it would be unable to engage both monomers of a single ACE2 dimer (PDB: 6M18), requiring a 26 or 42 Å long linker to facilitate this interaction. **B.** An ACE2 dimer is capable of binding two different spike trimers. The structures represent examples of simulated molecular orientations and positions, resulting in different viral membrane curvatures and distances between the spikes' TM regions. **C. Top**, the changes in the moving average distance measured between the two spike TM regions when bound to a single ACE2 dimer during the simulation time. The light red denotes the minimal, straight line distance and the dark red indicates the maximal distance, or arc length. **Bottom**, The evolution of the viral membrane curvature during the simulation. The solid blue line shows the moving average, and dots indicate individual data points. Dotted lines separate the simulation replicates. The light blue strip indicates curvatures which would correspond to a spherically-shaped virion within the reported radius range (50–200 nm).[349] Both the measured distance and membrane curvature are summarised on the right in histograms, fitted with Gaussian functions.

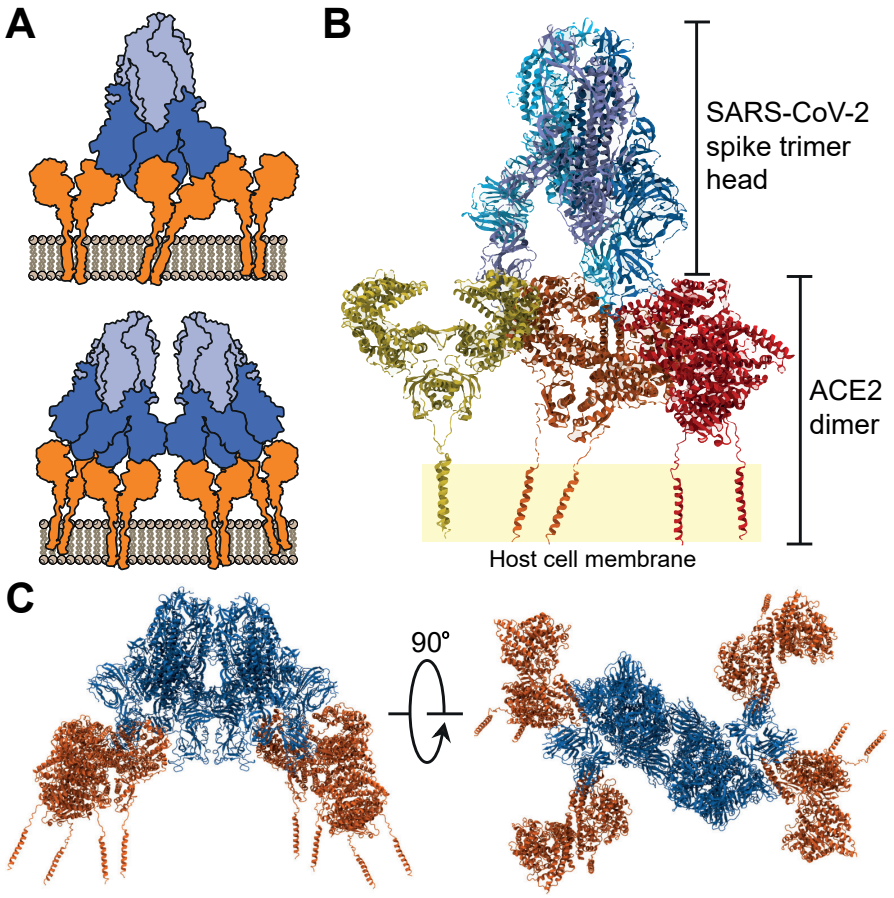


Figure 6.9: Analysis of the maximal occupancy of spike and its dimer. A. Cartoon schematics of the spike–ACE2 oligomers modelled in this Figure. **B.** MD model of a spike trimer with all three RBDs in the 'up' conformation (PDB: 6CS1 combined with homology modelling and coordinate transfer) binding to three ACE2 dimers, showing no clashes. **C.** Simulated structure of a spike dimer of trimers head sections (PDB: 7JJJ) binding up to 4 ACE2 dimers anchored to the host cell membrane. Whilst the attachment of another ACE2 dimer is theoretically possible, this would result in a significant host cell membrane distortion, and is thus not depicted.

We thus investigated the propensity of large and specific ACE2-spike complexes to form using the cryoEM structure establishing dimers of spike trimers (PDB: 7JJJ).[346] Attaching an ACE2 dimer to every available open RBD resulted in the two central ACE2s to clash. Thus, without invoking any flexibility of RBD domains, up to five ACE2 can bind to the spike dimer of trimers. *In vivo*, such an event would require the target cell to feature a local membrane curvature that, although within the range of known curvatures for cells on which ACE2 is expressed,[352] is significant (0.063 nm^{-1}). By omitting the binding of the central

ACE2, such curvature condition would relax onto a gentler one (0.002 nm^{-1}), resulting in the simulated oligomeric structure shown in Figure 6.9C, left. Adding another spike to the above model would close off one of the RBD sites to the ACE2, so we expect the number of ACE2 receptors to grow as $n + 2$, for an n-length oligomer, or $2n + 1$, if we allow higher curvatures, although we note that this may not hold for very large oligomers with significant torsion angles between terminal spikes. At the same time, the inherent flexibility of the ACE2 receptor in complex with the RBD has been highlighted, mirroring the range of motion exhibited by the spike protein.[353] We expect this flexibility to facilitate the variety of oligomeric states discussed here while decreasing any high curvature strains for both the virus and cell membranes.

6.3.4 SARS-CoV-2 is Strongly Inhibited by ACE2 but not mACE2

Previous work has shown that a viral spike crosslinking agent can have a dramatic effect on influenza and HIV infectivity.[354,355] One mechanistic reason given was that it impedes receptor or co-receptor binding, as crosslinking both occupies host recognition sites and prevents receptor clustering and binding avidity. Therefore, it is entirely plausible that competition exists between monomeric and dimeric forms of ACE2, with the dimer binding more strongly. We can test this by comparing the propensity of mACE2 or ACE2 to limit host-cell fusion via a focus forming assay with SARS-CoV-2 virions, using the workflow depicted in Figure 6.10A.

We incubated equal molar amounts of mACE2 or ACE2 with the England/02/2020 SARS-CoV-2 strain in a dilution series prior to the addition of Vero E6 cells and quantification of infectivity. With mACE2 we observed focus forming units (FFU) even at the highest concentration tested, but no infectivity was seen with ACE2, until we reached the concentration tested (25 nM) closest to the ACE2 monomer-

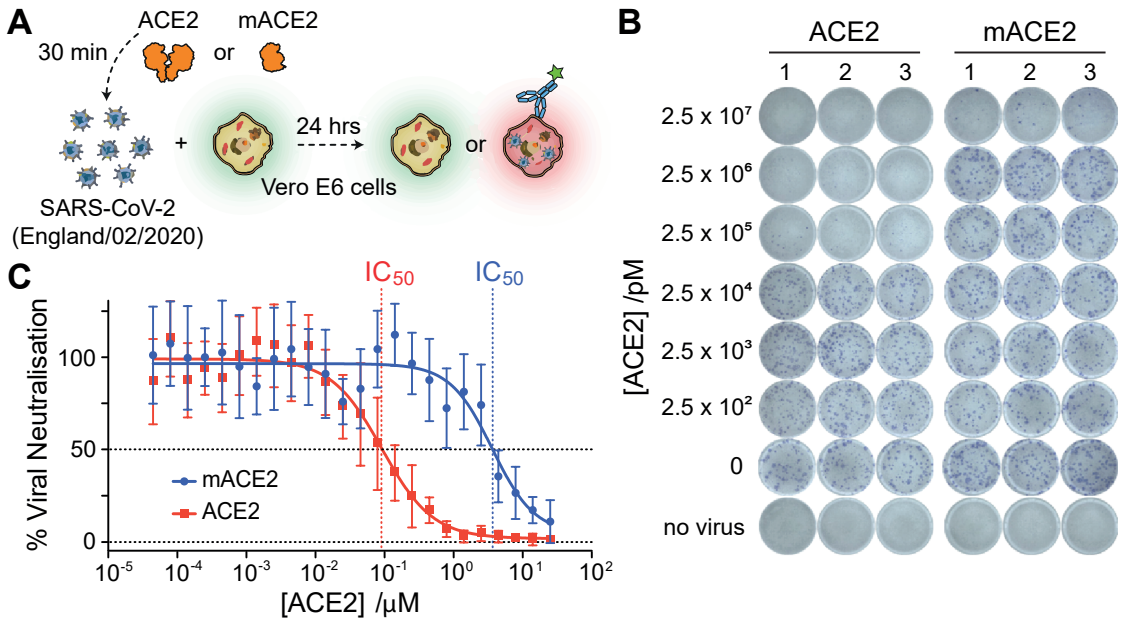


Figure 6.10: SARS-CoV-2 infection is inhibited more strongly by ACE2 than mACE2. **A.** Experimental schematic outlining the workflow of a viral infectivity inhibition testing focus-forming assay. A serial dilution of ACE2 or mACE2 is added to SARS-CoV-2 (England/02/2020) and incubated before adding Vero E6 cells. The infectivity is assessed by the presence of infected cells that are visualised by incubation with a primary, SARS-CoV-2-specific antibody (EY2A) followed by a secondary, peroxidase-conjugated antibody, after which peroxidase substrate is added and the infected foci are visualised using illumination. **B.** The differential ability to inhibit SARS-CoV-2 infection of ACE2 and mACE2 can be seen through the presence of focus-forming units (FFU) up to the highest mACE2 concentration tested, but only up to 2.5×10^4 pM ACE2. Shown are triplicate repeats of both conditions, together with a 0 pM ACE2/mACE2 and no virus controls. **C.** The ability of ACE2 and mACE2 to neutralise viral infection as a function of their concentration. Data are from quarter-log dilution series, enabling approximate half-maximal inhibition concentration (IC_{50}) determination.

dimer dissociation constant (Figure 6.10B). A more detailed quarter-log dilution series revealed that the half-maximal inhibitory concentrations (IC_{50}) were 0.09 μM for ACE2, and 8 μM for mACE2 (Figure 6.10C). Interestingly, when a similar concentration to the point at which ACE2 inhibited infection was measured by MP (75 nM, Figure 6.2D), the mass distributions show broadly similar counts of ACE2 monomers and dimers, at approximately 3:4 ratio. Together with the IC_{50} of mACE2 being almost two orders of magnitude higher than that of ACE2, these results indicate that the differences observed were not a consequence of mACE2 simply having half the RBD biding sites of ACE2 at μM or high nM concentrations.

6.4 Discussion and Conclusions

Since the start of the global COVID-19 pandemic, the spotlight of biophysical and structural studies has been on the spike–ACE2 interaction. Despite thorough characterisation using simplified reaction components, using both species in their native oligomeric forms has proven challenging. In this Chapter, we set out to investigate the molecular dynamics of this important interaction at different levels of complexity through a combination of mass photometry single-molecule measurements, molecular dynamics simulations, and viral neutralisation assay.

At the basic level of complexity, we show that while the ACE2 construct used here (ACE2Δ727_F805) exists in a monomer–dimer equilibrium (K_{dA} ∼12 nM), the ACE2–RBD interaction is not affected by ACE2 dimerisation. Whilst unsurprising since the dimerisation 'neck' domain is not involved in RBD binding,[206] this observation suggests that the stronger viral neutralisation potency of dimeric ACE2 as compared to the monomer is a directly due to the oligomeric nature of spike and ACE2. Furthermore, the appearance of large aggregates upon ACE2 addition to spike together with an overall particle count decrease in the spike MW range, both notably absent in the free spike or mACE2 controls further exemplify the importance of spike and ACE2 oligomeric states. Dimeric ACE2 appears to facilitate the formation of large oligomeric complexes through spike crosslinking, depleting the otherwise highly abundant spike trimers and dimers of trimers in the process. Whilst MP is not a structural technique and thus does not permit identification of the types of inter-molecular interactions involved in the formation of these complexes, ACE2–mediated crosslinking of spike trimers offers a reasonable explanation to the data presented.

Molecular dynamics simulations were employed to test the feasibility of the

interactions necessary for this type of crosslinking. We showed that both ACE2 dimer subunits cannot attach to two RBDs on the same spike, preventing intra-spike avidity, whilst the binding of a single ACE2 to two different spikes is possible and even likely, and has since been documented experimentally.[356] Given our crosslinking hypothesis, this interaction could act as a basic building block, bridging existing spike–ACE2 complexes together to form the observed large oligomers. Lastly, to extend these findings beyond single-molecule experiments and simulations, viral neutralisation testing confirmed two-orders of magnitude stronger SARS-CoV-2 infection inhibition by ACE2 compared to mACE2, providing evidence for the role of ACE2 dimerisation in inhibiting viral infection through a mechanism beyond simple ACE2–spike attachment.

At the same time, however, caution needs to be taken when interpreting these results. These experiments were conducted with solubilised protein constructs lacking their respective TM domains. As the oligomeric state of the native, membrane-inserted ACE2 remains to be determined, any firm conclusions regarding the role of ACE2 dimers *in vivo* are difficult to make. Additionally, solubilised proteins are capable of exhibiting much larger flexibility than their membrane-inserted counterparts. Even though the changes in membrane curvature during oligomer formation were taken into account, combined with extensive flexibility documented for both the spike hinge region and ACE2 linker section,[195,356] multiple ACE2 and spike proteins may not be proximal enough or bind strongly enough at the same time *in vivo* to facilitate this type of crosslinking. Furthermore, the spike construct used in these experiments is based on the sequence of the original SARS-CoV-2 variant. Over the last two years, the rapid evolution of this virus resulted in the appearance of five major variants of concern, each harbouring different sets of mutations, often concentrated in the RBD. Even though this should not in theory diminish the potential for this type of crosslinking, further experiments using the

spike from the latest Omicron variants would be necessary to confirm this.

Overall, we propose a novel mechanism for the spike–ACE2 interaction, a crucial first step of SARS-CoV-2 infection, through ACE2–mediated inter-spike crosslinking. Besides accounting for the observed stronger infection inhibition by ACE2 dimers, this mechanism would also enhance the strength of the viral attachment to the host cell membrane, resulting in the well-documented high infectivity and transmissibility of SARS-CoV-2. If shown to be generally applicable to spikes from different variants and membrane-anchored receptors, this mechanism could be further extended to develop viral infection treatments using covalently oligomerised decoy receptors, an antiviral therapy that would in theory be immune to viral mutational escape.

7 Summary and Outlook

Contents

7.1 Summary **188**
7.2 Outlook **191**

7.1 Summary

The main outcome of the work presented in this thesis was the extension of mass photometry as a single-molecule label-free technique to integral membrane proteins (IMPs) solubilised in a diverse range of membrane mimetic systems (MMS) and their interactions. Furthermore, we demonstrated the utility of mass photometry to characterise the heterogeneity of interactions between viral glycoproteins and their binding partners *in vitro*. We demonstrated these findings using different systems, including the bacterial outer membrane proteins and lipids, HIV-1 and SARS-CoV-2 membrane glycoproteins and their interactions with a soluble inhibitory protein and native host cell receptor, respectively.

We first explored the potential of mass photometry to detect and characterise

MMS, encountering technical limitations of the method with the most frequently used MMS, detergents. We were able to overcome this obstacle by modifying and simplifying a detergent-removal technique to successfully remove detergent background, enabling detection of respiratory complex I as well as characterisation of a monomer-dimer equilibrium of bo_3 oxidase. We then described the importance of protein:amphipol ratio for successful amphipol insertion and sample monodispersity. Applying mass photometry to native nanodisc KcsA samples revealed a large degree of heterogeneity across a size-exclusion chromatography purification profile, indicative of protein polydispersity in the membrane, with interesting implications for native membrane protein research. Lastly, testing different lipid nanodiscs KcsA assemblies showed that mass photometry can be used to predict protein functionality with a simple, rapid experiment needing only small quantities of protein, suggesting mass photometry as a tool for sample testing prior to structural characterisation.

Having characterised IMPs and MMS with mass photometry, we applied this knowledge to investigate the interactions between the bacterial outer membrane IMPs, focusing on the monomeric and trimeric β-barrels BtuB and OmpF, and the role of phospholipids and LPS in this interaction. We designed an *in vivo* photocrosslinking strategy through which the interacting outer membrane phospholipids and LPS were covalently attached to OmpF, resulting in complex formation with BtuB upon mixing of the purified proteins *in vitro*. This work contributed to the description of the outer membrane protein island architecture, composed of a trimeric and monomeric β-barrels linked through phospholipids and LPS, acting as a molecular 'glue'. These findings illustrate the ability of mass photometry to characterise IMP–IMP interactions, and have important implications for the design of novel antibacterial agents.

We next moved to viral membrane glycoproteins, starting with HIV-1. Whilst the HIV-inhibiting activity of the banana lectin BanLec has been well documented, the details of the mechanism behind its strong affinity and inhibition were less clear. We used mass photometry to characterise the interaction between BanLec and the HIV-1 envelope glycoprotein, Env, describing multivalent binding and intermolecular crosslinking of Env facilitated by BanLec, resulting in Env clustering. *In vivo*, this process would result in Env crosslinking on the surface of the virion, even between Env on two separate viral particles, potently neutralising viral infection by sequestering Env available for host cell attachment. Through this work, we demonstrated the strength of the single-molecule approach of mass photometry, enabling characterisation of complex, multivalent interactions.

Given the emergence of SARS-CoV-2 and the start of the global COVID-19 pandemic, we decided to apply our experience of working with viral glycoproteins and complex systems to the interaction between the SARS-CoV-2 spike glycoprotein and its host receptor, human ACE2. The initial characterisation of the simplified reaction components confirmed that on molecular level, monomeric and dimeric ACE2 interact with the receptor binding domain of spike equally. This changes when dimeric ACE2 or monomeric ACE2 are applied to the trimeric spike, resulting in the formation of large oligomers facilitated by ACE2 dimer, but not with ACE2 monomer. Using the quantitative aspect of mass photometry, we could confirm the associated decrease in the observed spike trimer counts as it was sequestered into these complexes by the ACE2 dimer. Whilst these experiments were performed using solubilised ACE2 and spike, we confirmed the feasibility of such crosslinking interactions with both binding partners anchored to the membrane through molecular dynamics simulations. Lastly, we confirmed the strong potency of viral neutralisation by ACE2 dimer compared to the monomer, likely due to the dimer-mediated spike crosslinking observed with mass photometry. This work

pushed the boundaries of complex reactions and their characterisation with mass photometry and provides important clues to the mechanism of viral infectivity and its potential exploitation for novel treatments of COVID-19.

7.2 Outlook

The experiments and analysis presented in this work pushed the boundaries of mass photometry as a technique and provided novel insights into multiple IMP systems and interaction mechanisms. The biophysical characterisation of different MMS in Chapter 3 could further be extended to novel MMS, such as peptidiscs, and other MMS, like bicelles. In addition, the drop-dilution method applied in this section could be tested on other detergents with low critical micelle concentration, such as GDN, providing a generalised approach to measuring IMPs in detergent with mass photometry. The effects of lipid and cholesterol content on the lipid nanodisc assembly and heterogeneity could also be tested, together with an ever-expanding repertoire of synthetic amphipathic polymers capable of native nanodisc formation and IMP extraction, an exciting novel avenue of IMP research. Through the use of these native membrane-disrupting polymers, entire membrane proteomes of specific cells could be extracted and their interactions analysed using mass photometry, following protein separation during the measurement through the use of *e.g.* electrophoretic microfluidic devices or specific protein tags. This bulk approach to membrane protein interactions and assembly could revolutionise how we perceive the membrane and reveal the true dynamics of this unique part of many cells.

Eventually, through the development and validation of supported lipid bilayers, it could be possible to move away from MMS in general and conduct experiments on fully-inserted IMPs in a near-native lipid bilayer environment. These experiments

could be conducted using a supported lipid bilayer, cushioned on the glass coverslip side with a layer of artificial polymer. Through this method, pioneered for membrane-associated proteins by Foley *et al.*,[357] both the mass and diffusion coefficient of IMPs could be measured simultaneously, allowing for unambiguous detection of protein-protein interaction through changes in both the MW measured and different diffusional speeds of the complex. Ultimately, perhaps through suspended lipid bilayers between two proximal coverslips, whole signalling cascades could be unravelled and investigated using MP in conjunction with a complementary technique, such as single-molecule FRET. For example, the extensively varied G protein signalling cascades have long been investigated through partial reactions and frozen complexes. Our understanding of these complicated systems could be vastly improved by using a single molecule, dynamic approach.

On the other hand, these developments could also be applied to outer membrane protein research as an extension of the work presented in Chapter 4. Further interactions between the outer membrane proteins and lipids could be characterised and quantified in more detail, aiding in our understanding of the outer membrane island formation and its relationship to antibiotic resistance emergence.

The intermolecular crosslinking of viral glycoproteins and their binding partners described in Chapters 5 and 6 could be further studied in relation to their implications for viral entry and infections rate. The applications of BanLec as a method of HIV-1 infection prevention could be further tested in high-risk areas, potentially aiding in decrease of infection rates. Regarding ACE2, there are multiple avenues of potential extension of the work presented here. Measurements of membrane-inserted ACE2 interactions with virion-anchored spike would provide mechanistic details of their interactions and verify the feasibility of such crosslinking interaction between two membrane-inserted proteins. The utility of solubilised, dimeric ACE2 as a

potent neutralising agent of SARS-CoV-2 infection could further be investigated, potentially resulting in COVID-19 treatment or prevention tool that would be immune from future viral mutational escape. Finally, the implications of this seemingly common crosslinking mechanism could be considered for other known and emerging viral infections, providing a relatively simple and effective method for quick viral neutralisation.

Overall, throughout this work we have demonstrated the multi-faceted applicability of mass photometry to studying IMPs, MMS, the interactions between different IMPs and lipids, as well as characterising complex, highly heterogeneous reactions implicated in viral entry and neutralisation. We believe this method could be further extended to investigate many other IMPs and their interaction networks, as well as further characterise the crosslinking mechanisms important for viral infections.

References

[1] L. Fagerberg, K. Jonasson, G. von Heijne, M. Uhlén and L. Berglund, Prediction of the Human Membrane Proteome, *Proteomics*, 2010, **10**, 1141–1149.

[2] A. S. Hauser, M. M. Attwood, M. Rask-Andersen, H. B. Schiöth and D. E. Gloriam, Trends in GPCR Drug Discovery: New Agents, Targets and Indications, *Nature Reviews Drug Discovery*, 2017, **16**, 829–842.

[3] Z. Breijyeh, B. Jubeh and R. Karaman, Resistance of Gram-Negative Bacteria to Current Antibacterial Agents and Approaches to Resolve It, *Molecules*, 2020, **25**, 1340.

[4] 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.

[5] 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.

[6] A. Sonn-Segev, K. Belacic, T. Bodrug, G. Young, R. T. VanderLinden, B. A. Schulman, J. Schimpf, T. Friedrich, P. V. Dip, T. U. Schwartz, B. Bauer, J.-M. Peters, W. B. Struwe, J. L. P. Benesch, N. G. Brown, D. Haselbach and P. Kukura, Quantifying the Heterogeneity of Macromolecular Machines by Mass Photometry, *Nature Communications*, 2020, **11**, 1772.

[7] 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, *Angewandte Chemie*, 2020, **132**, 10866–10871.

[8] Y. Li, W. B. Struwe and P. Kukura, Single Molecule Mass Photometry of Nucleic Acids, *Nucleic Acids Research*, 2020, **48**, 1–7.

[9] *Medical Cell Biology (Third Edition)*, ed. S. R. Goodman, Academic Press, San Diego, Third Edition edn., 2008, ch. 2, pp. 27–57.

[10] H. Yin and A. D. Flynn, Drugging Membrane Protein Interactions, *Annual Review of Biomedical Engineering*, 2016, **18**, 51–76.

[11] E. Wallin and G. von Heijne, Genome-Wide Analysis of Integral Membrane Proteins from Eubacterial, Archaean, and Eukaryotic Organisms, *Protein Science*, 1998, **7**, 1029–1038.

[12] M. Uhlén, L. Fagerberg, B. M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Åsa Sivertsson, C. Kampf, E. Sjöstedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A.-K. Szigyarto, J. Odeberg, D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P.-H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen and F. Pontén, Tissue-Based Map of the Human Proteome, *Science*, 2015, **347**, 1260419.

[13] J. P. Overington, B. Al-Lazikani and A. L. Hopkins, How Many Drug Targets Are There?, *Nature Reviews Drug Discovery*, 2006, **5**, 993–996.

[14] D. Chorev and C. Robinson, The Importance of the Membrane for Biophysical Measurements, *Nature Chemical Biology*, 2020, **16**, 1285–1292.

[15] G. Wang and R. L. J. Dunbrack, PISCES: a Protein Sequence Culling Server, *Bioinformatics*, 2003, **19**, 1589–1591.

[16] A. Olerinyova, A. Sonn-Segev, J. Gault, C. Eichmann, J. Schimpf, A. H. Kopf, L. S. 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.

[17] O. S. Andersen and R. E. Koeppe, Bilayer Thickness and Membrane Protein Function: An Energetic Perspective, *Annual Review of Biophysics and Biomolecular Structure*, 2007, **36**, 107–130.

[18] H. Hauser, M. C. Phillips and M. Stubbs, Ion Permeability of Phospholipid Bilayers, *Nature*, 1972, **239**, 342–344.

[19] D. Papahadjopoulos, S. Nir and S. Ohki, Permeability properties of phospholipid membranes: Effect of cholesterol and temperature, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1972, **266**, 561–583.

[20] E. J. Dufourc, Sterols and Membrane Dynamics, *Journal of Chemical Biology*, 2008, **1**, 63–77.

[21] T. Harayama and H. Riezman, Understanding the Diversity of Membrane Lipid Composition, *Nature Reviews Molecular Cell Biology*, 2018, **19**, 281–296.

[22] J. N. Israelachvili, D. J. Mitchell and B. W. Ninham, Theory of Self-Assembly of Lipid Bilayers and Vesicles, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1977, **470**, 185–201.

[23] L. H. Klausen, T. Fuhs and M. Dong, Mapping Surface Charge Density of Lipid Bilayers by Quantitative Surface Conductivity Microscopy, *Nature Communications*, 2016, **7**, 12447.

[24] S. Semrau and T. Schmidt, Membrane Heterogeneity - From Lipid Domains to Curvature Effects, *Soft Matter*, 2009, **5**, 3174–3186.

[25] J. E. Horne, D. J. Brockwell and S. E. Radford, Role of the Lipid Bilayer in Outer Membrane Protein Folding in Gram-Negative Bacteria, *Journal of Biological Chemistry*, 2020, **295**, 10340–10367.

[26] G. van Meer, D. R. Voelker and G. W. Feigenson, Membrane Lipids: Where They are and How They Behave, *Nature Reviews Molecular Cell Biology*, 2008, **9**, 112–124.

[27] G. Atilla-Gokcumen, E. Muro, J. Relat-Goberna, S. Sasse, A. Bedigian, M. Coughlin, S. Garcia-Manyes and U. Eggert, Dividing Cells Regulate Their Lipid Composition and Localization, *Cell*, 2014, **156**, 428–439.

[28] F. Lamari, F. Mochel, F. Sedel and J.-M. Saudubray, Disorders of Phospholipids, Sphingolipids and Fatty Acids Biosynthesis: Toward a New Category of Inherited Metabolic Diseases, *Journal of Inherited Metabolic Disease*, 2013, **36**, 411–425.

[29] A. Helenius and K. Simons, Solubilization of Membranes by Detergents, *Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes*, 1975, **415**, 29–79.

[30] D. Goodsell, L. Autin and A. Olson, Illustrate: Software for Biomolecular Illustration, *Structure*, 2019, **27**, 1716–1720.

[31] R. M. Garavito and S. Ferguson-Miller, Detergents as Tools in Membrane Biochemistry, *Journal of Biological Chemistry*, 2001, **276**, 32403–32406.

[32] A. Anandan and A. Vrielink, in *Detergents in Membrane Protein Purification and Crystallisation*, ed. I. Moraes, Springer International Publishing, 2016, pp. 13–28.

[33] M. L. Anson, The Denaturation Of Proteins By Synthetic Detergents And Bile Salts, *Journal of General Physiology*, 1939, **23**, 239–246.

[34] W. Kühne, *Chemische Vorgänge in der Netzhaut*, 1879, p. 264.

[35] B. Cartwright, P. Talbot and F. Brown, The Proteins of Biologically Active Subunits of Vesicular Stomatitis Virus, *Journal of General Virology*, 1970, **7**, 267–272.

[36] J. T. Seto, H. Becht and R. Rott, Isolation and Purification of Surface Antigens from Disrupted Paramyxoviruses, *Medical Microbiology and Immunology*, 1973, **159**, 1–12.

[37] A. Scheid, L. A. Caliguiri, R. W. Compans and P. W. Choppin, Isolation of Paramyxovirus Glycoproteins. Association of both Hemagglutinating and Neuraminidase Activities with the Larger SV5 Glycoprotein, *Virology*, 1972, **50**, 640–652.

[38] R. Nagarajan, Molecular Packing Parameter and Surfactant Self-Assembly: The Neglected Role of the Surfactant Tail, *Langmuir*, 2002, **18**, 31–38.

[39] P. G. Nilsson, H. Wennerstroem and B. Lindman, Structure of Micellar Solutions of Non-Ionic Surfactants. Nuclear Magnetic Resonance Self-Diffusion and Proton Relaxation Studies of Poly(ethylene Oxide) Alkyl Ethers, *Journal of Physical Chemistry*, 1983, **87**, 1377–1385.

[40] O. Lambert, D. Levy, J.-L. Ranck, G. Leblanc and J.-L. Rigaud, A New "Gel-like" Phase in Dodecyl Maltoside–Lipid Mixtures: Implications in Solubilization and Reconstitution Studies, *Biophysical Journal*, 1998, **74**, 918–930.

[41] T. Chakraborty, I. Chakraborty and S. Ghosh, The methods of determination of critical micellar concentrations of the amphiphilic systems in aqueous medium, *Arabian Journal of Chemistry*, 2011, **4**, 265–270.

[42] T. Arnold and D. Linke, Phase Separation in the Isolation and Purification of Membrane Proteins, *BioTechniques*, 2007, **43**, 427–440.

[43] S. Majeed, A. B. Ahmad, U. Sehar and E. R. Georgieva, Lipid Membrane Mimetics in Functional and Structural Studies of Integral Membrane Proteins, *Membranes*, 2021, **11**, 685.

[44] B. C. Choy, R. J. Cater, F. Mancia 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.

[45] H. Wennerström, Micelles. Physical Chemistry of Surfactant Association, *Physics Reports. A Review Section of Physics Letters*, 1979, **52**, 1–86.

[46] M. le Maire, P. Champeil and J. V. Møller, Interaction of Membrane Proteins and Lipids with Solubilizing Detergents, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2000, **1508**, 86–111.

[47] J. Umbreit and J. Strominger, Relation of Detergent HLB Number to Solubilization and Stabilization of D-Alanine Carboxypeptidase from *Bacillus subtilis* Membranes, *Proceedings of the National Academy of Sciences*, 1973, **70**, 2997–3001.

[48] J. Meunier, R. Olsen and J.-P. Changeux, Studies on the Cholinergic Receptor Protein from *Electrophorus electricus*: Effect of Detergents on some Hydrodynamic Properties of the Receptor Protein in Solution, *FEBS Letters*, 1972, **24**, 63–8.

[49] M. Rubin and A. Tzagoloff, Assembly of the Mitochondrial Membrane System: IX. Purification, Characterization, and Subunit Structure of Yeast and Beef Cytochrome Oxidase, *The Journal of Biological Chemistry*, 1973, **248**, 4269–74.

[50] D. Linke, in *Guide to Protein Purification, 2nd Edition*, ed. R. R. Burgess and M. P. Deutscher, Academic Press, 2009, vol. 463 of Methods in Enzymology, ch. 34, pp. 603–617.

[51] A. J. Furth, H. Bolton, J. Potter and J. D. Priddle, in *Part C: Enzyme Purification and Related Techniques*, Academic Press, 1984, vol. 104 of Methods in Enzymology, ch. 17, pp. 318–328.

[52] J. L. Rigaud, D. Levy, G. Mosser and O. Lambert, Detergent Removal by Non-Polar Polystyrene Beads - Applications to Membrane Protein Reconstitution and Two-Dimenstional Crystallization, *European Biophysics Journal - Biophysics Letters*, 1998, **27**, 305–319.

[53] T. Arnold and D. Linke, The Use of Detergents to Purify Membrane Proteins, *Current Protocols in Protein Science*, 2008, **53**, 4.8.1–4.8.30.

[54] 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, *Nature Methods*, 2010, **7**, 1003–1008.

[55] F. Hauer, C. Gerle, N. Fischer, A. Oshima, K. Shinzawa-Itoh, S. Shimada, K. Yokoyama, Y. Fujiyoshi and H. Stark, GraDeR: Membrane Protein Complex Preparation for Single-Particle Cryo-EM, *Structure*, 2015, **23**, 1769–1775.

[56] C. Tribet, R. Audebert and J.-L. Popot, Amphipols: Polymers that Keep Membrane Proteins Soluble in Aqueous Solutions, *Proceedings of the National Academy of Sciences*, 1996, **93**, 15047–15050.

[57] F. Giusti, J. Rieger, L. Catoire, S. Qian, A. Calabrese, T. Watkinson, M. Casiraghi, S. Radford, A. Ashcroft and J.-L. Popot, Synthesis, Characterization and Applications of a Perdeuterated Amphipol, *The Journal of Membrane Biology*, 2014, **247**, 909–924.

[58] M. Zoonens and J.-L. Popot, Amphipols for Each Season, *The Journal of Membrane Biology*, 2014, **247**, 759–796.

[59] K. S. Sharma, G. Durand, F. Giusti, B. Olivier, A.-S. Fabiano, P. Bazzacco, T. Dahmane, C. Ebel, J.-L. Popot and B. Pucci, Glucose-Based Amphiphilic Telomers Designed to Keep Membrane Proteins Soluble in Aqueous Solutions: Synthesis and Physicochemical Characterization, *Langmuir*, 2008, **24**, 13581–13590.

[60] P. Bazzacco, E. Billon-Denis, K. Sharma, L. Catoire, S. Mary, C. Bon, E. Point, J.-L. Banères, G. Durand, F. Zito, B. Pucci and J.-L. Popot, Nonionic Homopolymeric Amphipols: Application to Membrane Protein Folding, Cell-Free Synthesis, and Solution Nuclear Magnetic Resonance, *Biochemistry*, 2012, **51**, 1416–1430.

[61] Y. Gohon, T. Dahmane, R. Ruigrok, P. Schuck, D. Charvolin, F. Rappaport, P. Timmins, D. Engelman, C. Tribet, J.-L. Popot and C. Ebel, Bacteriorhodopsin/Amphipol Complexes: Structural and Functional Properties, *Biophysical Journal*, 2008, **94**, 3523–3537.

[62] M. Opacic, F. Giusti, J.-L. Popot and J. Broos, Isolation of *Escherichia coli* Mannitol Permease, EII^{mtl}, Trapped in Amphipol A8-35 and Fluorescein-Labeled A8-35, *The Journal of Membrane Biology*, 2014, **247**, 1019–1030.

[63] M. Bosco, M. Damian, V. Chauhan, M. Roche, P. Guillet, J. A. Fehrentz, F. Bonneté, A. Polidori, J. L. Banères and G. Durand, Biotinylated Non-Ionic Amphipols for GPCR Ligands Screening, *Methods*, 2020, **180**, 69–78.

[64] K. Sharma, G. Durand, F. Gabel, P. Bazzacco, C. Bon, E. Billon-Denis, L. Catoire, J.-L. Popot, C. Ebel and B. Pucci, Non-Ionic Amphiphilic Homopolymers: Synthesis, Solution Properties, and Biochemical Validation, *Langmuir: the ACS Journal of Surfaces and Colloids*, 2012, **28**, 4625–4639.

[65] Y. Gohon, F. Giusti, C. Prata, D. Charvolin, P. Timmins, C. Ebel, C. Tribet and J.-L. Popot, Well-Defined Nanoparticles Formed by Hydrophobic Assembly of a Short and Polydisperse Random Terpolymer, Amphipol A8-35, *Langmuir: the ACS Journal of Surfaces and Colloids*, 2006, **22**, 1281–1290.

[66] F. Giusti, J.-L. Popot and C. Tribet, Well-Defined Critical Association Concentration and Rapid Adsorption at the Air/Water Interface of a Short Amphiphilic Polymer, Amphipol A8-35: A Study by Förster Resonance Energy Transfer and Dynamic Surface Tension Measurements, *Langmuir: the ACS Journal of Surfaces and Colloids*, 2012, **28**, 10372–10380.

[67] J. Perlmutter, W. Drasler, W. Xie, J. Gao, J.-L. Popot and J. Sachs, All-Atom and Coarse-Grained Molecular Dynamics Simulations of a Membrane Protein Stabilizing Polymer, *Langmuir: the ACS Journal of Surfaces and Colloids*, 2011, **27**, 10523–10537.

[68] M. Tehei, J. Perlmutter, F. Giusti, J. Sachs, G. Zaccai and J.-L. Popot, Thermal Fluctuations in Amphipol A8-35 Particles: A Neutron Scattering

and Molecular Dynamics Study, *The Journal of Membrane Biology*, 2014, **247**, 897–908.

[69] Y. Gohon, G. Pavlov, P. Timmins, C. Tribet, J.-L. Popot and C. Ebel, Partial Specific Volume and Solvent Interactions of Amphipol A8-35, *Analytical Biochemistry*, 2004, **334**, 318–334.

[70] C. Diab, C. Tribet, Y. Gohon, J. L. Popot and F. Winnik, Complexation of Integral Membrane Proteins by Phosphorylcholine-based Amphipols, *Biochimica et Biophysica Acta (BBA)*, 2007, **1768**, 2737–2747.

[71] T. Dahmane, F. Giusti, L. Catoire and J.-L. Popot, Sulfonated Amphipols: Synthesis, Properties, and Applications, *Biopolymers*, 2011, **95**, 811–823.

[72] H. Autzen, D. Julius and Y. Cheng, Membrane Mimetic Systems in CryoEM: Keeping Membrane Proteins in their Native Environment, *Current Opinion in Structural Biology*, 2019, **58**, 259–268.

[73] J. H. Kleinschmidt and J.-L. Popot, Folding and Stability of Integral Membrane Proteins in Amphipols, *Archives of Biochemistry and Biophysics*, 2014, **564**, 327–343.

[74] J. L. Popot, E. Berry, D. Charvolin, C. Creuzenet, C. Ebel, D. Engelman, M. Floetenmeyer, F. Giusti, Y. Gohon, Q. Hong, J. Lakey, K. Leonard, H. A. Shuman, P. Timmins, D. Warschawski, F. Zito, M. Zoonens, B. Pucci and C. Tribet, Amphipols: Polymeric Surfactants for Membrane Biology Research, *Cellular and Molecular Life Sciences*, 2003, **60**, 1559–1574.

[75] P. Bazzacco, *Ph.D. thesis*, Non-ionic Amphipols: New Tools for *in vitro* Studies of Membrane Proteins. Validation and Development of Biochemical and Biophysical Applications, Université Paris, 2009.

[76] P. Champeil, T. Menguy, C. Tribet, J.-L. Popot and M. Maire, Interaction of Amphipols with Sarcoplasmic Reticulum Ca^{2+}-ATPase, *The Journal of Biological Chemistry*, 2000, **275**, 18623–18637.

[77] M. Zoonens, F. Giusti, F. Zito and J.-L. Popot, Dynamics of Membrane Protein/Amphipol Association Studied by Förster Resonance Energy Transfer: Implications for *in Vitro* Studies of Amphipol-Stabilized Membrane Proteins, *Biochemistry*, 2007, **46**, 10392–10404.

[78] K. Y. Chung, T. H. Kim, A. Manglik, R. Alvares, B. K. Kobilka and R. Prosser, Role of Detergents in Conformational Exchange of a G Protein-coupled Receptor, *Journal of Biological Chemistry*, 2012, **287**, 36305–36311.

[79] T. Althoff, D. J. Mills, J.-L. Popot and W. Kühlbrandt, Arrangement of Electron Transport Chain Components in Bovine Mitochondrial Supercomplex $I_1III_2IV_1$, *The EMBO Journal*, 2011, **30**, 4652–4664.

[80] J. Perlmutter, J. L. Popot and J. Sachs, Molecular Dynamics Simulations of a Membrane Protein/Amphipol Complex, *The Journal of Membrane Biology*, 2014, **247**, 883–895.

[81] C. Le Bon, A. Marconnet, S. Masscheleyn, J. L. Popot and M. Zoonens, Folding and Stabilizing Membrane Proteins in Amphipol A8-35, *Methods*, 2018, **147**, 95–105.

[82] T. Dahmane, M. Damian, S. Mary, J. L. Popot and J. L. Banères, Amphipol-Assisted *in Vitro* Folding of G Protein-Coupled Receptors, *Biochemistry*, 2009, **48**, 6516–6521.

[83] M. Picard, T. Dahmane, M. Garrigos, C. Gauron, F. Giusti, M. Maire, J.-L. Popot and P. Champeil, Protective and Inhibitory Effects of Various Types of Amphipols on the Ca^{2+}-ATPase from Sarcoplasmic Reticulum: A Comparative Study, *Biochemistry*, 2006, **45**, 1861–1869.

[84] D. Charvolin, M. Picard, L.-s. Huang, E. Berry and J.-L. Popot, Solution Behavior and Crystallization of Cytochrome bc_1 in the Presence of Amphipols, *The Journal of Membrane Biology*, 2014, **247**, 981–996.

[85] C. Bechara, G. Bolbach, P. Bazzacco, K. Sharma, G. Durand, J.-L. Popot, F. Zito and S. Sagan, MALDI-TOF Mass Spectrometry Analysis of Amphipol-Trapped Membrane Proteins, *Analytical Chemistry*, 2012, **84**, 6128–6135.

[86] W. Arunmanee, J. Harris and J. Lakey, Outer Membrane Protein F Stabilised with Minimal Amphipol Forms Linear Arrays and LPS-Dependent 2D Crystals, *The Journal of Membrane Biology*, 2014, **247**, 949–956.

[87] C. Tribet, C. Diab, T. Dahmane, M. Zoonens, J.-L. Popot and F. Winnik, Thermodynamic Characterization of the Exchange of Detergents and Amphipols at the Surfaces of Integral Membrane Proteins, *Langmuir: the ACS Journal of Surfaces and Colloids*, 2009, **25**, 12623–12634.

[88] J. L. Banères, J. L. Popot and B. Mouillac, New Advances in Production and Functional Folding of G-Protein-Coupled Receptors, *Trends in Biotechnology*, 2011, **29**, 314–322.

[89] D. Charvolin, J. B. Perez, F. Rouvière, F. Giusti, P. Bazzacco, A. Abdine, F. Rappaport, K. Martinez and J. L. Popot, The Use of Amphipols as Universal Molecular Adapters to Immobilize Membrane Proteins onto Solid Supports, *Proceedings of the National Academy of Sciences*, 2009, **106**, 405–410.

[90] J. L. Popot, in *Amphipols and Membrane Protein Crystallization*, 2018, pp. 497–531.

[91] J. Hopper, Y. Yu, D. Li, A. Raymond, M. Bostock, I. Liko, V. Mikhailov, A. Laganowsky, J. Benesch, M. Caffrey, D. Nietlispach and C. Robinson, Detergent-Free Mass Spectrometry of Membrane Protein Complexes, *Nature Methods*, 2013, **10**, 1206–1208.

[92] J. L. Popot, T. Althoff, D. Bagnard, J. L. Banères, P. Bazzacco, E. Billon-Denis, L. Catoire, P. Champeil, D. Charvolin, M. Cocco, G. Crémel, T. Dahmane, L. de la Maza, C. Ebel, F. Gabel, F. Giusti, Y. Gohon, E. Goormaghtigh, E. Guittet and M. Zoonens, Amphipols from A to Z, *Annual Review of Biophysics*, 2011, **40**, 379–408.

[93] G. Zubay, *In Vitro* Synthesis of Protein in Microbial Systems, *Annual Review of Genetics*, 1973, **7**, 267–287.

[94] D. Schwarz, V. Dötsch and F. Bernhard, Production of Membrane Proteins Using Cell-Free Expression Systems, *Proteomics*, 2008, **8**, 3933–3946.

[95] T. Bayburt, Y. Grinkova and S. Sligar, Self-Assembly of Discoidal Phospholipid Bilayer Nanoparticles with Membrane Scaffold Proteins, *Nano Letters*, 2002, **2**, 853–856.

[96] I. G. Denisov, Y. V. Grinkova, A. A. Lazarides and S. G. Sligar, Directed Self-Assembly of Monodisperse Phospholipid Bilayer Nanodiscs with Controlled Size, *Journal of The American Chemical Society*, 2004, **126**, 3477–3487.

[97] D. Atkinson, H. M. Smith, J. Dickson and J. P. Austin, Interaction of Apoprotein from Porcine High-Density Lipoprotein with Dimyristoyl Lecithin, *European Journal of Biochemistry*, 1976, **64**, 541–547.

[98] A. Wlodawer, J. P. Segrest, B. H. Chung, R. Chiovetti and J. N. Weinstein, High-Density Lipoprotein Recombinants: Evidence for a Bicycle Tire Micelle Structure Obtained by Neutron Scattering and Electron Microscopy, *FEBS Letters*, 1979, **104**, 231–235.

[99] V. Koppaka, L. Silvestro, J. Engler, C. Brouillette and P. Axelsen, The Structure of Human Lipoprotein A-I- Evidence for the “Belt" Model, *The Journal of Biological Chemistry*, 1999, **274**, 14541–14544.

[100] Y. Li, A. Z. Kijac, S. G. Sligar and C. M. Rienstra, Structural Analysis of Nanoscale Self-Assembled Discoidal Lipid Bilayers by Solid-State NMR Spectroscopy, *Biophysical Journal*, 2006, **91**, 3819–3828.

[101] K. M. Padmanabha Das, W. M. Shih, G. Wagner and M. L. Nasr, Large Nanodiscs: A Potential Game Changer in Structural Biology of Membrane Protein Complexes and Virus Entry, *Frontiers in Bioengineering and Biotechnology*, 2020, **8**, 539.

[102] Y. V. Grinkova, I. G. Denisov and S. G. Sligar, Engineering Extended Membrane Scaffold Proteins for Self-Assembly of Soluble Nanoscale Lipid Bilayers, *Protein Engineering, Design and Selection*, 2010, **23**, 843–848.

[103] I. G. Denisov and S. G. Sligar, Nanodiscs for Structural and Functional Studies of Membrane Proteins, *Nature Structural and Molecular Biology*, 2016, **23**, 481–486.

[104] F. Hagn, M. Etzkorn, T. Raschle and G. Wagner, Optimized Phospholipid Bilayer Nanodiscs Facilitate High-Resolution Structure Determination of Membrane Proteins, *Journal of the American Chemical Society*, 2013, **135**, 1919–1925.

[105] J. E. Rouck, J. E. Krapf, J. Roy, H. C. Huff and A. Das, Recent Advances in Nanodisc Technology for Membrane Protein Studies (2012–2017), *FEBS Letters*, 2017, **591**, 2057–2088.

[106] M. L. Nasr, D. Baptista, M. Strauss, Z. Y. Sun, S. Grigoriu, S. Huser, A. Plückthun, F. Hagn, T. Walz, J. M. Hogle and G. Wagner, Covalently Circularized Nanodiscs for Studying Membrane Proteins and Viral Entry, *Nature Methods*, 2017, **14**, 49–52.

[107] B. Zakeri, J. O. Fierer, E. Celik, E. C. Chittock, U. Schwarz-Linek, V. T. Moy and M. Howarth, Peptide Tag Forming a Rapid Covalent Bond to a Protein, through Engineering a Bacterial Adhesin, *Proceedings of the National Academy of Sciences*, 2012, **109**, E690–E697.

[108] S. Zhang, Q. Ren, S. Novick, T. Strutzenberg, P. Griffin and H. Bao, One-Step Construction of Circularized Nanodiscs using SpyCatcher-SpyTag, *Nature Communications*, 2021, **12**, 5451.

[109] S. G. Sligar and I. G. Denisov, Nanodiscs: A Toolkit for Membrane Protein Science, *Protein Science*, 2020, **30**, 297–315.

[110] M. Nakano, M. Fukuda, T. Kudo, M. Miyazaki, Y. Wada, N. Matsuzaki, H. Endo and T. Handa, Static and Dynamic Properties of Phospholipid Bilayer Nanodiscs, *Journal of the American Chemical Society*, 2009, **131**, 8308–8312.

[111] I. Schachter, C. Allolio, G. Khelashvili and D. Harries, Confinement in Nanodiscs Anisotropically Modifies Lipid Bilayer Elastic Properties, *The Journal of Physical Chemistry B*, 2020, **124**, 7166–7175.

[112] T. Bengtsen, V. L. Holm, L. R. Kjølbye, S. R. Midtgaard, N. T. Johansen, G. Tesei, S. Bottaro, B. Schiøtt, L. Arleth and K. Lindorff-Larsen, Structure and dynamics of a nanodisc by integrating NMR, SAXS and SANS experiments with molecular dynamics simulations, *eLife*, 2020, **9**, e56518.

[113] D. Martinez, M. Decossas, J. Kowal, L. Frey, H. Stahlberg, E. J. Dufourc, R. Riek, B. Habenstein, S. Bibow and A. Loquet, Lipid Internal Dynamics Probed in Nanodiscs, *ChemPhysChem*, 2017, **18**, 2651–2657.

[114] P. Stepien, A. Polit and A. Wisniewska-Becker, Comparative EPR Studies on Lipid Bilayer Properties in Nanodiscs and Liposomes, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2015, **1848**, 60–66.

[115] S. Inagaki, R. Ghirlando and R. Grisshammer, Biophysical Characterization of Membrane Proteins in Nanodiscs, *Methods*, 2013, **59**, 287–300.

[116] T. H. Bayburt and S. G. Sligar, Membrane Protein Assembly into Nanodiscs, *FEBS Letters*, 2010, **584**, 1721–1727.

[117] H.-X. Zhou and T. A. Cross, Influences of Membrane Mimetic Environments on Membrane Protein Structures, *Annual Review of Biophysics*, 2013, **42**, 361–392.

[118] T. A. Cross, M. Sharma, M. Yi and H. Zhou, Influence of Solubilizing Environments on Membrane Protein Structures, *Trends in Biochemical Sciences*, 2011, **36**, 117–125.

[119] K. Wilcox, M. Marunde, A. Das, P. Velasco, B. Kuhns, M. Marty, H. Jiang, C.-H. Luan, S. Sligar and W. Klein, Nanoscale Synaptic Membrane Mimetic Allows Unbiased High Throughput Screen That Targets Binding Sites for Alzheimer's-Associated Aβ Oligomers, *PLoS ONE*, 2015, **10**, e0125263.

[120] N. Akkaladevi, S. Mukherjee, H. Katayama, B. Janowiak, D. Patel, E. P. Gogol, B. L. Pentelute, R. J. Collier and M. T. Fisher, Following Natures Lead: On the Construction of Membrane-Inserted Toxins in Lipid Bilayer Nanodiscs, *The Journal of Membrane Biology*, 2015, **248**, 595–607.

[121] J. Frauenfeld, J. Gumbart, E. O. van der Sluis, S. Funes, M. Gartmann, B. Beatrix, T. Mielke, O. Berninghausen, T. Becker, K. Schulten and R. Beckmann, Cryo-EM Structure of the Ribosome-SecYE Complex in the Membrane Environment, *Nature Structural & Molecular Biology*, 2011, **18**, 614–621.

[122] K. Grushin, J. Miller, D. Dalm and S. Stoilova-McPhie, Factor VIII Organisation on Nanodiscs with Different Lipid Composition, *Thrombosis and Haemostasis*, 2015, **113**, 741–749.

[123] M. Nikolaev, E. Round, I. Gushchin, V. Polovinkin, T. Balandin, P. Kuzmichev, V. Shevchenko, V. Borshchevskiy, A. Kuklin, A. Round, F. Bernhard, D. Willbold, G. Büldt and V. Gordeliy, Integral Membrane Proteins Can Be Crystallized Directly from Nanodiscs, *Crystal Growth & Design*, 2017, **17**, 945–948.

[124] A. Kijac, Y. Li, S. Sligar and C. Rienstra, Magic-Angle Spinning Solid-State NMR Spectroscopy of Nanodisc-Embedded Human CYP3A4, *Biochemistry*, 2007, **46**, 13696–13703.

[125] K. S. Mineev, S. A. Goncharuk, P. K. Kuzmichev, M. Vilar and A. S. Arseniev, NMR Dynamics of Transmembrane and Intracellular Domains of p75NTR in Lipid-Protein Nanodiscs, *Biophysical Journal*, 2015, **109**, 772–782.

[126] L. Morgado, K. Zeth, B. Burmann, T. Maier and S. Hiller, Characterization of the Insertase BamA in Three Different Membrane Mimetics by Solution NMR Spectroscopy, *Journal of Biomolecular NMR*, 2015, **61**, 333–345.

[127] M. Redhair, A. F. Clouser and W. M. Atkins, Hydrogen-Deuterium Exchange Mass Spectrometry of Membrane Proteins in Lipid Nanodiscs, *Chemistry and Physics of Lipids*, 2019, **220**, 14–22.

[128] C. M. Hebling, C. R. Morgan, D. W. Stafford, J. W. Jorgenson, K. D. Rand and J. R. Engen, Conformational Analysis of Membrane Proteins in Phospholipid Bilayer Nanodiscs by Hydrogen Exchange Mass Spectrometry, *Analytical Chemistry*, 2010, **82**, 5415–5419.

[129] X. X. Zhang, C. S. Chan, H. Bao, Y. Fang, L. J. Foster and F. Duong, Nanodiscs and SILAC-Based Mass Spectrometry to Identify a Membrane Protein Interactome, *Journal of Proteome Research*, 2011, **11**, 1454–1459.

[130] M. T. Marty, H. Zhang, W. Cui, R. E. Blankenship, M. L. Gross and S. G. Sligar, Native Mass Spectrometry Characterization of Intact Nanodisc Lipoprotein Complexes, *Analytical Chemistry*, 2012, **84**, 8957–8960.

[131] G. Zhang, J. E. Keener and M. T. Marty, Measuring Remodeling of the Lipid Environment Surrounding Membrane Proteins with Lipid Exchange and Native Mass Spectrometry, *Analytical Chemistry*, 2020, **92**, 5666–5669.

[132] M. T. Marty, A. J. Baldwin, E. G. Marklund, G. K. A. Hochberg, J. L. P. Benesch and C. V. Robinson, Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary Interactions to Polydisperse Ensembles, *Analytical Chemistry*, 2015, **87**, 4370–4376.

[133] M. T. Marty, Nanodiscs and Mass Spectrometry: Making Membranes Fly, *International Journal of Mass Spectrometry*, 2020, **458**, 116436.

[134] D. R. McDougle, J. L. Baylon, D. D. Meling, A. Kambalyal, Y. V. Grinkova, J. Hammernik, E. Tajkhorshid and A. Das, Incorporation of Charged Residues in the CYP2J2 F-G Loop Disrupts CYP2J2–Lipid Bilayer Interactions, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2015, **1848**, 2460–2470.

[135] R. Lamichhane, J. J. Liu, G. Pljevaljcic, K. L. White, E. van der Schans, V. Katritch, R. C. Stevens, K. Wüthrich and D. P. Millar, Single-Molecule View of Basal Activity and Activation Mechanisms of the G Protein-Coupled Receptor β2AR, *Proceedings of the National Academy of Sciences*, 2015, **112**, 14254–14259.

[136] N. Bocquet, J. Kohler, M. N. Hug, E. A. Kusznir, A. C. Rufer, R. J. Dawson, M. Hennig, A. Ruf, W. Huber and S. Huber, Real-time Monitoring of Binding Events on a Thermostabilized Human A2A Receptor Embedded in a Lipid Bilayer by Surface Plasmon Resonance, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2015, **1848**, 1224–1233.

[137] S. R. Tonge and B. J. Tighe, *Lipid-Containing Compositions and Uses Thereof*, Aston University, U. S. Patent PCT/GB98/02546, Aug. 1998.

[138] S. R. Tonge and B. J. Tighe, Responsive Hydrophobically Associating Polymers: a Review of Structure and Properties, *Advanced Drug Delivery Reviews*, 2001, **53**, 109–122.

[139] 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, *Journal of The American Chemical Society*, 2009, **131**, 7484–7485.

[140] S. Gulati, M. Jamshad, T. Knowles, K. Morrison, R. Downing, N. Cant, R. Collins, J. Koenderink, R. Ford, M. Overduin, I. Kerr, T. Dafforn and A. Rothnie, Detergent-free purification of ABC (ATP-binding-cassette) transporters, *Biochemical Journal*, 2014, **461**, 269–278.

[141] C. Logez, M. Damian, C. Legros, C. Dupré, M. Guéry, S. Mary, R. Wagner, C. M'Kadmi, O. Nosjean, B. Fould, J. Marie, J. Fehrentz, J. Martinez, G. Ferry, J. A. Boutin and J. Banères, Detergent-free Isolation of Functional G Protein-Coupled Receptors into Nanometric Lipid Particles, *Biochemistry*, 2016, **55**, 38–48.

[142] M. Jamshad, J. Charlton, Y.-P. Lin, S. Routledge, Z. Bawa, T. Knowles, M. Overduin, N. Dekker, T. Dafforn, R. Bill, D. Poyner and M. Wheatley, G-Protein Coupled Receptor Solubilization and Purification for Biophysical Analysis and Functional Studies, in the Total Absence of Detergent, *Bioscience Reports*, 2015, **35**, e00188.

[143] S. C. Lee, T. J. Knowles, V. L. G. Postis, M. Jamshad, R. A. Parslow, Y.-P. Lin, A. Goldman, P. Sridhar, M. Overduin, S. P. Muench and T. R. Dafforn, A Method for Detergent-Free Isolation of Membrane Proteins in their Local Lipid Environment, *Nature Protocols*, 2016, **11**, 1149–1162.

[144] K. A. Morrison, A. Akram, A. Mathews, Z. A. Khan, J. H. Patel, C. Zhou, D. J. Hardy, C. Moore-Kelly, R. Patel, V. Odiba, T. J. Knowles, M.-u.-H. Javed, N. P. Chmel, T. R. Dafforn and A. J. Rothnie, Membrane Protein Extraction and Purification Using Styrene–Maleic Acid (SMA) Copolymer: Effect of Variations in Polymer Structure, *Biochemical Journal*, 2016, **473**, 4349–4360.

[145] 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.

[146] A. A. Gulamhussein, R. Uddin, B. J. Tighe, D. R. Poyner and A. J. Rothnie, A Comparison of SMA (Styrene Maleic Acid) and DIBMA (Di-Isobutylene Maleic Acid) for Membrane Protein Purification, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2020, **1862**, 183281.

[147] M. Jamshad, V. Grimard, I. Idini, T. J. Knowles, M. R. Dowle, N. Schofield, P. Sridhar, Y. Lin, R. Finka, M. Wheatley, O. R. T. Thomas, R. E. Palmer, M. Overduin, C. Govaerts, J.-M. Ruysschaert, K. J. Edler and T. R. Dafforn, Structural Analysis of a Nanoparticle Containing a Lipid Bilayer Used for Detergent-Free Extraction of Membrane Proteins, *Nano Research*, 2015, **8**, 774–789.

[148] J. M. Dörr, M. C. Koorengevel, M. Schäfer, A. V. Prokofyev, S. Scheidelaar, E. A. W. van der Cruijsen, T. R. Dafforn, M. Baldus and J. A. Killian, Detergent-Free Isolation, Characterization, and Functional Reconstitution of a Tetrameric K+ Channel: The Power of Native Nanodiscs, *Proceedings of the National Academy of Sciences*, 2014, **111**, 18607–18612.

[149] 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, *European Biophysics Journal*, 2015, **45**, 3–21.

[150] A. O. Oluwole, B. Danielczak, A. Meister, J. O. Babalola, C. Vargas and S. Keller, Solubilization of Membrane Proteins into Functional Lipid-Bilayer Nanodiscs Using a Diisobutylene/Maleic Acid Copolymer, *Angewandte Chemie International Edition*, 2017, **56**, 1919–1924.

[151] A. Chen, E. J. Majdinasab, M. C. Fiori, H. Liang and G. A. Altenberg, Polymer-Encased Nanodiscs and Polymer Nanodiscs: New Platforms for Membrane Protein Research and Applications, *Frontiers in Bioengineering and Biotechnology*, 2020, **8**, 1329.

[152] T. Ravula, N. Z. Hardin, S. K. Ramadugu and A. Ramamoorthy, pH Tunable and Divalent Metal Ion Tolerant Polymer Lipid Nanodiscs, *Langmuir*, 2017, **33**, 10655–10662.

[153] S. C. L. Hall, C. Tognoloni, J. Charlton, E. C. Bragginton, A. J. Rothnie, P. Sridhar, M. Wheatley, T. J. Knowles, T. Arnold, K. J. Edler and T. R. Dafforn, An Acid-Compatible Co-Polymer For the Solubilization of Membranes and Proteins Into Lipid Bilayer-Containing Nanoparticles, *Nanoscale*, 2018, **10**, 10609–10619.

[154] M. C. Fiori, Y. Jiang, G. A. Altenberg and H. Liang, Polymer-Encased Nanodiscs with Improved Buffer Compatibility, *Scientific Reports*, 2017, **7**, 7432.

[155] M. C. Fiori, W. Zheng, E. Kamilar, G. Simiyu, G. A. Altenberg and H. Liang, Extraction and Reconstitution of Membrane Proteins into Lipid Nanodiscs Encased by Zwitterionic Styrene-Maleic Amide Copolymers, *Scientific Reports*, 2020, **10**, 9940.

[156] T. Ravula, N. Z. Hardin, J. Bai, S.-C. Im, L. Waskell and A. Ramamoorthy, Effect of Polymer Charge on Functional Reconstitution of Membrane Proteins in Polymer Nanodiscs, *Chemical Communications*, 2018, **54**, 9615–9618.

[157] S. Lindhoud, V. Carvalho, J. W. Pronk and M.-E. Aubin-Tam, SMA-SH: Modified Styrene-Maleic Acid Copolymer for Functionalization of Lipid Nanodiscs, *Biomacromolecules*, 2016, **17**, 1516–1522.

[158] B. Danielczak, M. Rasche, J. Lenz, E. Patallo-Perez, S. Weyrauch, F. Mahler, M. T. Agbadaola, A. Meister, J. O. Babalola, C. Vargas, C. Kolar and S. Keller, A Bioinspired Glycopolymer for Capturing Membrane Proteins in Native-Like Lipid-Bilayer Nanodiscs, *Nanoscale*, 2022, **14**, 1855–1867.

[159] D. J. 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.

[160] M. C. Orwick, P. J. Judge, J. Procek, L. Lindholm, A. Graziadei, A. Engel, G. Gröbner and A. Watts, Detergent-Free Formation and Physicochemical Characterization of Nanosized Lipid–Polymer Complexes: Lipodisq, *Angewandte Chemie International Edition*, 2012, **51**, 4653–4657.

[161] C. Sun, S. Benlekbir, P. Venkatakrishnan, Y. Wang, S. Hong, J. Hosler, E. Tajkhorshid, J. L. Rubinstein and R. B. Gennis, Structure of the Alternative Complex III in a Supercomplex with Cytochrome Oxidase, *Nature*, 2018, **557**, 123–126.

[162] M. Parmar, S. Rawson, C. A. Scarff, A. Goldman, T. R. Dafforn, S. P. Muench and V. L. Postis, Using a SMALP Platform to Determine a Sub-nm Single Particle Cryo-EM Membrane Protein Structure, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2018, **1860**, 378–383.

[163] J. Broecker, B. T. Eger and O. P. Ernst, Crystallogenesis of Membrane Proteins Mediated by Polymer-Bounded Lipid Nanodiscs, *Structure*, 2017, **25**, 384–392.

[164] B. Bersch, J. M. Dörr, A. Hessel, J. A. Killian and P. Schanda, Proton-Detected Solid-State NMR Spectroscopy of a Zinc Diffusion Facilitator Protein in Native Nanodiscs, *Angewandte Chemie International Edition*, 2017, **56**, 2508–2512.

[165] 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.

[166] D. J. K. Swainsbury, S. Scheidelaar, R. van Grondelle, J. A. Killian and M. R. Jones, Bacterial Reaction Centers Purified with Styrene Maleic Acid

Copolymer Retain Native Membrane Functional Properties and Display Enhanced Stability, *Angewandte Chemie International Edition*, 2014, **53**, 11803–11807.

[167] R. Cuevas Arenas, B. Danielczak, A. Martel, L. Porcar, C. Breyton, C. Ebel and S. Keller, Fast Collisional Lipid Transfer Among Polymer-Bounded Nanodiscs, *Scientific Reports*, 2017, **7**, 45875.

[168] A. Marconnet, B. Michon, C. Le Bon, F. Giusti, C. Tribet and M. Zoonens, Solubilization and Stabilization of Membrane Proteins by Cycloalkane-Modified Amphiphilic Polymers, *Biomacromolecules*, 2020, **21**, 3459–3467.

[169] A. J. Higgins, A. J. Flynn, A. Marconnet, L. J. Musgrove, V. L. G. Postis, J. D. Lippiat, C.-W. Chung, T. Ceska, M. Zoonens, F. Sobott and S. P. Muench, Cycloalkane-Modified Amphiphilic Polymers Provide Direct Extraction of Membrane Proteins for CryoEM Analysis, *Communications Biology*, 2021, 1337.

[170] Joint United Nations Programme on HIV/AIDS, *Fact Sheet*, 2021, www.unaids.org/en/resources/fact-sheet, accessed: 2021-09-03.

[171] J. M. Cuevas, R. Geller, R. Garijo, J. López-Aldeguer and R. Sanjuán, Extremely High Mutation Rate of HIV-1 *in vivo*, *PLoS Biology*, 2015, **13**, e1002251.

[172] C. Grundner, M. Pancera, J. Kang, M. Koch, J. Sodroski and R. Wyatt, Factors Limiting the Immunogenicity of HIV-1 gp120 Envelope Glycoproteins, *Virology*, 2004, **330**, 233–248.

[173] Z. T. Berndsen, S. Chakraborty, X. Wang, C. A. Cottrell, J. L. Torres, J. K. Diedrich, C. A. López, J. R. Yates III, M. J. van Gils, J. C. Paulson, S. Gnanakaran and A. B. Ward, Visualization of the HIV-1 Env Glycan Shield across Scales, *Proceedings of the National Academy of Sciences*, 2020, **117**, 28014–28025.

[174] J. Mascola, G. Stiegler, T. VanCott, H. Katinger, C. Carpenter, C. Hanson, H. Beary, D. Hayes, S. Frankel, D. Birx and M. Lewis, Protection of Macaques Against Vaginal Transmission of a Pathogenic HIV-1/SIV Chimeric Virus by Passive Infusion of Neutralizing Antibodies, *Nature Medicine*, 2000, **6**, 207–210.

[175] A. Trkola, Conformation of the Native HIV-1 Envelope Protein Raises Questions for Vaccine Ddesign, *Nature*, 2019, **568**, 321–322.

[176] M. Lu, X. Ma, L. Castillo-Menendez, J. Gorman, N. Alsahafi, U. Ermel, D. Terry, M. Chambers, D. Peng, B. Zhang, T. Zhou, N. Reichard, K. Wang, J. Grover, B. Carman, M. Gardner, I. Nikić, A. Sugawara, J. Arthos and W. Mothes, Associating HIV-1 Envelope Glycoprotein Structures with States on the Virus Observed by smFRET, *Nature*, 2019, **568**, 415–419.

[177] Q. Wang, A. Finzi and J. Sodroski, The Conformational States of the HIV-1 Envelope Glycoproteins, *Trends in Microbiology*, 2020, **28**, 655–667.

[178] R. W. Sanders, R. Derking, A. Cupo, J.-P. Julien, A. Yasmeen, N. de Val, H. J. Kim, C. Blattner, A. T. de la Peña, J. Korzun, M. Golabek, K. de los Reyes, T. J. Ketas, M. J. van Gils, C. R. King, I. A. Wilson, A. B. Ward, P. J. Klasse and J. P. Moore, A Next-Generation Cleaved, Soluble HIV-1 Env Trimer, BG505 SOSIP.664 gp140, Expresses Multiple Epitopes for Broadly Neutralizing but Not Non-Neutralizing Antibodies, *PLoS Pathogens*, 2013, **9**, e1003618.

[179] R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, Y. Bi, X. Ma, F. Zhan, L. Wang, T. Hu, H. Zhou, Z. Hu, W. Zhou, L. Zhao, J. Chen, Y. Meng, J. Wang, Y. Lin, J. Yuan, Z. Xie, J. Ma, W. J. Liu, D. Wang, W. Xu, E. C. Holmes, G. F. Gao, G. Wu, W. Chen, W. Shi and W. Tan, Genomic Characterisation and Epidemiology of 2019 Novel Coronavirus: Implications for Virus Origins and Receptor Binding, *The Lancet*, 2020, **395**, 565–574.

[180] P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, H.-D. Chen, J. Chen, Y. Luo, H. Guo, R.-D. Jiang, M.-Q. Liu, Y. Chen, X.-R. Shen, X. Wang, X.-S. Zheng, K. Zhao, Q.-J. Chen, F. Deng, L.-L. Liu, B. Yan, F.-X. Zhan, Y.-Y. Wang, G.-F. Xiao and Z.-L. Shi, A Pneumonia Outbreak Associated with a New Coronavirus of Probable Bat Origin, *Nature*, 2020, **579**, 270–273.

[181] F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, M.-L. Yuan, Y.-L. Zhang, F.-H. Dai, Y. Liu, Q.-M. Wang, J.-J. Zheng, L. Xu, E. C. Holmes and Y.-Z. Zhang, A New Coronavirus Associated with Human Respiratory Disease in China, *Nature*, 2020, **579**, 265–269.

[182] World Health Organization, *WHO Coronavirus (COVID-19) Dashboard*, www.covid19.who.int, 2021, accessed: 2021-09-10.

[183] C. Zhang, W. Zheng, X. Huang, E. W. Bell, X. Zhou and Y. Zhang, Protein Structure and Sequence Reanalysis of 2019-nCoV Genome Refutes Snakes as Its Intermediate Host and the Unique Similarity between Its Spike Protein Insertions and HIV-1, *Journal of Proteome Research*, 2020, **19**, 1351–1360.

[184] T. T.-Y. Lam, N. Jia, Y.-W. Zhang, M. H.-H. Shum, J.-F. Jiang, H.-C. Zhu, Y.-G. Tong, Y.-X. Shi, X.-B. Ni, Y.-S. Liao, W.-J. Li, B.-G. Jiang, W. Wei, T.-T. Yuan, K. Zheng, X.-M. Cui, J. Li, G.-Q. Pei, X. Qiang, W. Y.-M. Cheung, L.-F. Li, F.-F. Sun, S. Qin, J.-C. Huang, G. M. Leung, E. C. Holmes, Y.-L. Hu, Y. Guan and W.-C. Cao, Identifying SARS-CoV-2-Related Coronaviruses in Malayan Pangolins, *Nature*, 2020, **583**, 282–285.

[185] L. E. Gralinski and V. D. Menachery, Return of the Coronavirus: 2019-ncov, *Viruses*, 2020, **12**, 135.

[186] I. d. A. Santos, V. R. Grosche, F. R. G. Bergamini, R. Sabino-Silva and A. C. G. Jardim, Antivirals Against Coronaviruses: Candidate Drugs for SARS-CoV-2 Treatment?, *Frontiers in Microbiology*, 2020, **11**, 1818.

[187] M.-Y. Wang, R. Zhao, L.-J. Gao, X.-F. Gao, D.-P. Wang and J.-M. Cao, SARS-CoV-2: Structure, Biology, and Structure-Based Therapeutics Development, *Frontiers in Cellular and Infection Microbiology*, 2020, **10**, 1–17.

[188] W. Li, M. J. Moore, N. Vasilieva, J. Sui, S. K. Wong, M. A. Berne, M. Somasundaran, J. L. Sullivan, K. Luzuriaga, T. C. Greenough, H. Choe and M. Farzan, Angiotensin-Converting Enzyme 2 is a Functional Receptor for the SARS Coronavirus, *Nature*, 2003, **426**, 450–454.

[189] F. Li, W. Li, M. Farzan and S. C. Harrison, Structure of SARS Coronavirus Spike Receptor-Binding Domain Complexed with Receptor, *Science*, 2005, **309**, 1864–1868.

[190] F. Li, Structure, Function, and Evolution of Coronavirus Spike Proteins, *Annual Review of Virology*, 2016, **3**, 237–261.

[191] D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C.-L. Hsieh, O. Abiona, B. S. Graham and J. S. McLellan, Cryo-EM Structure of the 2019-nCoV Spike in the Prefusion Conformation, *Science*, 2020, **367**, 1260–1263.

[192] A. C. Walls, Y.-J. Park, M. A. Tortorici, A. Wall, A. T. McGuire and D. Veesler, Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein, *Cell*, 2020, **183**, 281–292.

[193] M. Lu, P. D. Uchil, W. Li, D. Zheng, D. S. Terry, J. Gorman, W. Shi, B. Zhang, T. Zhou, S. Ding, R. Gasser, J. Prévost, G. Beaudoin-Bussières, S. P. Anand, A. Laumaea, J. R. Grover, L. Liu, D. D. Ho, J. R. Mascola, A. Finzi, P. D. Kwong, S. C. Blanchard and W. Mothes, Real-Time Conformational Dynamics of SARS-CoV-2 Spikes on Virus Particles, *Cell Host and Microbe*, 2020, **28**, 880–891.e8.

[194] T. Zhou, Y. Tsybovsky, J. Gorman, M. Rapp, G. Cerutti, G.-Y. Chuang, P. S. Katsamba, J. M. Sampson, A. Schön, J. Bimela, J. C. Boyington, A. Nazzari, A. S. Olia, W. Shi, M. Sastry, T. Stephens, J. Stuckey, I.-T. Teng, P. Wang, S. Wang, B. Zhang, R. A. Friesner, D. D. Ho, J. R. Mascola, L. Shapiro and P. D. Kwong, Cryo-EM Structures of SARS-CoV-2 Spike without and with ACE2 Reveal a pH-Dependent Switch to Mediate Endosomal Positioning of Receptor-Binding Domains, *Cell Host and Microbe*, 2020, **28**, 867–879.e5.

[195] B. Turoňová, M. Sikora, C. Schürmann, W. J. H. Hagen, S. Welsch, F. E. C. Blanc, S. von Bülow, M. Gecht, K. Bagola, C. Hörner, G. van Zandbergen, J. Landry, N. T. D. de Azevedo, S. Mosalaganti, A. Schwarz, R. Covino, M. D. Mühlebach, G. Hummer, J. Krijnse Locker and M. Beck, In Situ Structural Analysis of SARS-CoV-2 Spike Reveals Flexibility Mediated by Three Hinges, *Science*, 2020, **370**, 203–208.

[196] L. Casalino, Z. Gaieb, J. A. Goldsmith, C. K. Hjorth, A. C. Dommer, A. M. Harbison, C. A. Fogarty, E. P. Barros, B. C. Taylor, J. S. McLellan, E. Fadda and R. E. Amaro, Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein, *ACS Central Science*, 2020, **6**, 1722–1734.

[197] Y. Finkel, A. Gluck, A. Nachshon, R. Winkler, T. Fisher, B. Rozman, O. Mizrahi, Y. Lubelsky, B. Zuckerman, B. Slobodin, Y. Yahalom-Ronen, H. Tamir, I. Ulitsky, T. Israely, N. Paran, M. Schwartz and N. Stern-Ginossar, SARS-CoV-2 Uses a Multipronged Strategy to Impede Host Protein Synthesis, *Nature*, 2021, **594**, 240–245.

[198] A. K. Banerjee, M. R. Blanco, E. A. Bruce, D. D. Honson, L. M. Chen, A. Chow, P. Bhat, N. Ollikainen, S. A. Quinodoz, C. Loney, J. Thai, Z. D. Miller, A. E. Lin, M. M. Schmidt, D. G. Stewart, D. Goldfarb, G. De Lorenzo, S. J. Rihn, R. M. Voorhees, J. W. Botten, D. Majumdar and M. Guttman, SARS-CoV-2 Disrupts Splicing, Translation, and Protein Trafficking to Suppress Host Defenses, *Cell*, 2020, **183**, 1325–1339.e21.

[199] K. Schubert, E. D. Karousis, A. Jomaa, A. Scaiola, B. Echeverria, L.-A. Gurzeler, M. Leibundgut, V. Thiel, O. Mühlemann and N. Ban, SARS-CoV-2 Nsp1 Binds the Ribosomal mRNA Channel to Inhibit Translation, *Nature Structural and Molecular Biology*, 2020, **27**, 959–966.

[200] M. Thoms, R. Buschauer, M. Ameismeier, L. Koepke, T. Denk, M. Hirschenberger, H. Kratzat, M. Hayn, T. Mackens-Kiani, J. Cheng, J. H. Straub, C. M. Stürzel, T. Fröhlich, O. Berninghausen, T. Becker, F. Kirchhoff, K. M. J. Sparrer and R. Beckmann, Structural Basis for Translational Shutdown and Immune Evasion by the Nsp1 Protein of SARS-CoV-2, *Science*, 2020, **369**, 1249–1255.

[201] E. Vainieri and H. Javed, Dexamethasone in Hospitalized Patients with Covid-19, *The New England Journal of Medicine*, 2021, **384**, 693–704.

[202] M. Donoghue, F. Hsieh, E. Baronas, K. Godbout, M. Gosselin, N. Stagliano, M. Donovan, B. Woolf, K. Robison, R. Jeyaseelan, R. E. Breitbart and S. Acton, A Novel Angiotensin-Converting Enzyme–Related Carboxypeptidase (ACE2) Converts Angiotensin I to Angiotensin 1-9, *Circulation Research*, 2000, **87**, E1–9.

[203] S. R. Tipnis, N. M. Hooper, R. Hyde, E. Karran, G. Christie and A. J. Turner, A Human Homolog of Angiotensin-Converting Enzyme, *Journal of Biological Chemistry*, 2000, **275**, 33238–33243.

[204] C. A. Devaux, J.-M. Rolain and D. Raoult, ACE2 Receptor Polymorphism: Susceptibility to SARS-CoV-2, Hypertension, Multi-Organ Failure, and COVID-19 Disease Outcome, *Journal of Microbiology, Immunology and Infection*, 2020, **53**, 425–435.

[205] A. J. Turner and N. M. Hooper, The Angiotensin–Converting Enzyme Gene Family: Genomics and Pharmacology, *Trends in Pharmacological Sciences*, 2002, **23**, 177–183.

[206] R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo and Q. Zhou, Structural Basis for the Recognition of SARS-CoV-2 by Full-Length Human ACE2, *Science*, 2020, **367**, 1444–1448.

[207] C. Vickers, P. Hales, V. Kaushik, L. Dick, J. Gavin, J. Tang, K. Godbout, T. Parsons, E. Baronas, F. Hsieh, S. Acton, M. Patane, A. Nichols and P. Tummino, Hydrolysis of Biological Peptides by Human Angiotensin-converting Enzyme-related Carboxypeptidase, *Journal of Biological Chemistry*, 2002, **277**, 14838–14843.

[208] A. Heurich, H. Hofmann-Winkler, S. Gierer, T. Liepold, O. Jahn and S. Pöhlmann, TMPRSS2 and ADAM17 Cleave ACE2 Differentially and Only Proteolysis by TMPRSS2 Augments Entry Driven by the Severe Acute Respiratory Syndrome Coronavirus Spike Protein, *Journal of Virology*, 2014, **88**, 1293–1307.

[209] D. W. Lambert, M. Yarski, F. J. Warner, P. Thornhill, E. T. Parkin, A. I. Smith, N. M. Hooper and A. J. Turner, Tumor Necrosis Factor-α Convertase (ADAM17) Mediates Regulated Ectodomain Shedding of the Severe-Acute Respiratory Syndrome-Coronavirus (SARS-CoV) Receptor, Angiotensin-Converting Enzyme-2 (ACE2), *Journal of Biological Chemistry*, 2005, **280**, 30113–30119.

[210] F. Scialo, A. Daniele, F. Amato, L. Pastore, M. G. Matera, M. Cazzola, G. Castaldo and A. Bianco, ACE2: The Major Cell Entry Receptor for SARS-CoV-2, *Lung*, 2020, **198**, 867–877.

[211] Y. Wang, Y. Wang, W. Luo, L. Huang, J. Xiao, F. Li, S. Qin, X. Song, Y. Wu, Q. Zeng, F. Jin and Y. Wang, A Comprehensive Investigation of the mRNA and Protein Level of ACE2, the Putative Receptor of SARS-CoV-2, in Human Tissues and Blood Cells, *International Journal of Medical Sciences*, 2020, **17**, 1522–1531.

[212] X. Zou, K. Chen, J. Zou, P. Han, J. Hao and Z. Han, Single-Cell RNA-Seq Data Analysis on the Receptor ACE2 Expression Reveals the Potential Risk of Different Human Organs Vulnerable to 2019-nCoV Infection, *Frontiers of Medicine*, 2020, **14**, 185–192.

[213] M. Zhuang, Y. Cheng, J. Zhang, X. Jiang, L. Wang, J. Deng and P. Wang, Increasing Host Cellular Receptor—Angiotensin-Converting Enzyme 2 Expression by Coronavirus May Facilitate 2019-nCoV (or SARS-CoV-2) Infection, *Journal of Medical Virology*, 2020, **92**, 2693–2701.

[214] C. Ziegler, S. J. Allon, S. K. Nyquist, I. Mbano, V. N. Miao, Y. Cao, A. S. Yousif, J. Bals, B. M. Hauser, J. Feldman, C. Muus, M. H. Wadsworth II,

S. Kazer, T. K. Hughes, B. Doran, G. J. Gatter, M. Vukovic, C. N. Tzouanas, F. Taliaferro, Z. Guo, J. P. Wang, D. F. Dwyer, K. M. Buchheit, J. Boyce, N. A. Barrett, T. M. Laidlaw, S. L. Carroll, L. Colonna, V. Tkachev, A. Yu, H. B. Zheng, H. P. Gideon, C. G. Winchell, P. L. Lin, B. Berger, A. Leslie, J. L. Flynn, S. M. Fortune, R. W. Finberg, L. Kean, M. Garber, A. Schmidt, D. Lingwood, A. K. Shalek, J. Ordovas-Montanes and H. Lung Biological Network, SARS-CoV-2 Receptor ACE2 is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Enriched in Specific Cell Subsets Across Tissues, *SSRN Journal*, 2020, **181**, 1016–1035.

[215] O. O. Onabajo, A. R. Banday, M. L. Stanifer, W. Yan, A. Obajemu, D. M. Santer, O. Florez-Vargas, H. Piontkivska, J. M. Vargas, T. J. Ring, C. Kee, P. Doldan, D. L. Tyrrell, J. L. Mendoza, S. Boulant and L. Prokunina-Olsson, Interferons and Viruses Induce a Novel Truncated ACE2 Isoform and not the Full-Length SARS-CoV-2 Receptor, *Nature Genetics*, 2020, **52**, 1283–1293.

[216] G. Cholankeril, A. Podboy, V. I. Aivaliotis, B. Tarlow, E. A. Pham, S. P. Spencer, D. Kim, A. Hsing and A. Ahmed, High Prevalence of Concurrent Gastrointestinal Manifestations in Patients With Severe Acute Respiratory Syndrome Coronavirus 2: Early Experience From California, *Gastroenterology*, 2020, **159**, 775–777.

[217] A. Werion, L. Belkhir, M. Perrot, G. Schmit, S. Aydin, Z. Chen, A. Penaloza, J. De Greef, H. Yildiz, L. Pothen, J. C. Yombi, J. Dewulf, A. Scohy, L. Gérard, X. Wittebole, P.-F. Laterre, S. E. Miller, O. Devuyst, M. Jadoul, J. Morelle, F. Aboubakar, S. Acid, N. Amini, S. Bailly, C. Beauloye, D. Castanares-Zapatero, E. Coche, C. Collienne, P. Cornette, I. De Brauwer, M. Dechamps, F. Dupriez, A. Froidure, Q. Garnir, B. Gerber, B. Ghaye, I. Gilard, S. Gohy, C. Grégoire, P. Hantson, L.-M. Jacquet, B. Kabamba, S. Kautbally, N. Lanthier, F. Larbaoui, G. Liistro, F. Maes, V. Montiel, B. Mwenge, S. Pierard, C. Pilette, A. C. Pouleur, A. Sogorb, P. Starkel, H. Rodriguez-Villalobos, M. Thoma, O. Van Caeneghem and D. Vancraeynest, SARS-CoV-2 Causes a Specific Dysfunction of the Kidney Proximal Tubule, *Kidney International*, 2020, **98**, 1296–1307.

[218] M. Furuhashi, N. Moniwa, H. Takizawa, N. Ura and K. Shimamoto, Potential Differential Effects of Renin-Angiotensin System Inhibitors on SARS-CoV-2 Infection and Lung Injury in COVID-19, *Hypertension Research : Official Journal of the Japanese Society of Hypertension*, 2020, **43**, 837–840.

[219] S. Haga, N. Nagata, T. Okamura, N. Yamamoto, T. Sata, N. Yamamoto, T. Sasazuki and Y. Ishizaka, TACE Antagonists Blocking ACE2 Shedding Caused by the Spike Protein of SARS-CoV are Candidate Antiviral Compounds, *Antiviral Research*, 2010, **85**, 551–555.

[220] S. Haga, N. Yamamoto, C. Nakai-Murakami, Y. Osawa, K. Tokunaga, T. Sata, N. Yamamoto, T. Sasazuki and Y. Ishizaka, Modulation of TNF–Converting Enzyme by the Spike Protein of SARS-CoV and ACE2 Induces

TNF- Production and Facilitates Viral Entry, *Proceedings of the National Academy of Sciences*, 2008, **105**, 7809–7814.

[221] M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T. S. Schiergens, G. Herrler, N.-H. Wu, A. Nitsche, M. A. Müller, C. Drosten and S. Pöhlmann, SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor, *Cell*, 2020, **181**, 271–280.e8.

[222] A. Shulla, T. Heald-Sargent, G. Subramanya, J. Zhao, S. Perlman and T. Gallagher, A Transmembrane Serine Protease Is Linked to the Severe Acute Respiratory Syndrome Coronavirus Receptor and Activates Virus Entry, *Journal of Virology*, 2010, **85**, 873–882.

[223] H. P. Jia, D. C. Look, P. Tan, L. Shi, M. Hickey, L. Gakhar, M. C. Chappell, C. Wohlford-Lenane and P. B. McCray, Ectodomain Shedding of Angiotensin-Converting Enzyme 2 in Human Airway Epithelia, *American Journal of Physiology- Lung Cellular and Molecular Physiology*, 2009, **297**, L84–L96.

[224] J. Ramchand, S. K. Patel, L. G. Kearney, G. Matalanis, O. Farouque, P. M. Srivastava and L. M. Burrell, Plasma ACE2 Activity Predicts Mortality in Aortic Stenosis and Is Associated With Severe Myocardial Fibrosis, *Journal of the American College of Cardiology: Cardiovascular Imaging*, 2020, **13**, 655–664.

[225] S. Epelman, W. W. Tang, S. Y. Chen, F. Van Lente, G. S. Francis and S. Sen, Detection of Soluble Angiotensin-Converting Enzyme 2 in Heart Failure, *Journal of the American College of Cardiology*, 2008, **52**, 750–754.

[226] Q. Li, Z. Cao and P. Rahman, Genetic Variability of Human Angiotensin-Converting Enzyme 2 (hACE2) among Various Ethnic Populations, *Molecular Genetics and Genomic Medicine*, 2020, **8**, 1344.

[227] E. Benetti, R. Tita, O. Spiga, A. Ciolfi, G. Birolo, A. Bruselles, G. Doddato, A. Giliberti, C. Marconi, F. Musacchia, T. Pippucci, A. Torella, A. Trezza, F. Valentino, M. Baldassarri, A. Brusco, R. Asselta, M. Bruttini, S. Furini and A. Ognibene, ACE2 Gene Variants May Underlie Interindividual Variability and Susceptibility to COVID-19 in the Italian Population, *European Journal of Human Genetics*, 2020, **28**, 1602–1614.

[228] X. Guo, Z. Chen, Y. Xia, W. Lin and H. Li, Investigation of the Genetic Variation in ACE2 on the Structural Recognition by the Novel Coronavirus (SARS-CoV-2), *Journal of Translational Medicine*, 2020, **18**, 321.

[229] D. Batlle, J. Wysocki and K. Satchell, Soluble Angiotensin-Converting Enzyme 2: A Potential Approach for Coronavirus Infection Therapy?, *Clinical Science*, 2020, **134**, 543–545.

[230] E. Ciaglia, C. Vecchione and A. A. Puca, COVID-19 Infection and Circulating ACE2 Levels: Protective Role in Women and Children, *Frontiers in Pediatrics*, 2020, **8**, 206.

[231] S. A. Kornilov, I. Lucas, K. Jade, C. L. Dai, J. C. Lovejoy and A. T. Magis, Plasma Levels of Soluble ACE2 are Associated with Sex, Metabolic Syndrome, and its Biomarkers in a Large Cohort, Pointing to a Possible Mechanism for Increased Severity in COVID-19, *Critical Care*, 2020, **24**, 452.

[232] P. Swärd, A. Edsfeldt, A. Reepalu, L. Jehpsson, B. E. Rosengren and M. K. Karlsson, Age and Sex Differences in Soluble ACE2 May Give Insights for COVID-19, *Critical Care*, 2020, **24**, 221.

[233] J. Ortega-Arroyo, D. Cole and P. Kukura, Interferometric Scattering Microscopy and its Combination with Single-Molecule Fluorescence Imaging, *Nature Protocols*, 2016, **11**, 617–633.

[234] F. Soltermann, *Ph.D. thesis*, Biophysical Characterization of Protein-Protein Interactions with Mass Photometry, University of Oxford, 2020.

[235] L. J. Catoire, X. L. Warnet and D. E. Warschawski, Micelles, Bicelles, Amphipols, Nanodiscs, Liposomes, or Intact Cells: The Hitchhiker's Guide to the Study of Membrane Proteins by NMR, *Membrane Proteins Production for Structural Analysis*, 2014, 315–345.

[236] T. Gewering, D. Januliene, A. B. Ries and A. Moeller, Know Your Detergents: A Case Study on Detergent Background in Negative Stain Electron Microscopy, *Journal of Structural Biology*, 2018, **203**, 242–246.

[237] G. Loy and A. Zelinsky, Fast Radial Symmetry for Detecting Points of Interest, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, **25**, 959–973.

[238] Anatrace, *D310-n-Dodecyl-β-D-Maltopyranoside, Anagrade*, 2021, https://www.anatrace.com/Products/Detergents/MALTOSIDES/D310, accessed: 2021-11-23.

[239] B. Lorber, J. B. Bishop and L. J. DeLucas, Purification of Octyl *β*-*D*-Glucopyranoside and Re-estimation of its MicellarSsize, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1990, **1023**, 254–265.

[240] 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. 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, *Nature Methods*, 2010, **7**, 1003–1008.

[241] Anatrace, *F308S - Fos-Choline-12, Sol-Grade, n-Dodecylphosphocholine*, 2022, https://www.anatrace.com/Products/Lipids/FOS-CHOLINE/F308S, accessed: 2022-07-31.

[242] C. Eichmann, L. Frey, I. Maslennikov and R. Riek, Probing Ion Binding in the Selectivity Filter of the KcsA Potassium Channel, *Journal of The American Chemical Society*, 2019, **141**, 7391–7398.

[243] S. Steimle, C. Bajzath, K. Dörner, M. Schulte, V. Bothe and T. Friedrich, Role of Subunit NuoL for Proton Translocation by Respiratory Complex I, *Biochemistry*, 2011, **50**, 3386–3393.

[244] N. G. Housden, J. A. Wojdyla, J. Korczynska, I. Grishkovskaya, N. Kirkpatrick, A. M. Brzozowski and C. Kleanthous, Directed Epitope Delivery across the *Escherichia coli* Outer Membrane through the Porin OmpF, *Proceedings of the National Academy of Sciences of the United States of America*, 2010, **107**, 21412—21417.

[245] S. Bibow, Y. Polyhach, C. Eichmann, C. N. Chi, J. Kowal, S. Albiez, R. A. McLeod, H. Stahlberg, G. Jeschke, P. Güntert and R. Riek, Solution Structure of Discoidal High-Density Lipoprotein Particles with a Shortened Apolipoprotein A-I, *Nature Structural and Molecular Biology*, 2017, **24**, 187–193.

[246] Z. O. Shenkarev, E. N. Lyukmanova, O. I. Solozhenkin, I. E. Gagnidze, O. V. Nekrasova, V. V. Chupin, A. A. Tagaev, Z. A. Yakimenko, T. V. Ovchinnikova, M. P. Kirpichnikov and A. S. Arseniev, Lipid-Protein Nanodiscs: Possible Application in High-Resolution NMR Investigations of Membrane Proteins and Membrane-Active Peptides, *Biochemistry Moscow*, 2009, **74**, 756–765.

[247] S. Imai, M. Osawa, K. Mita, S. Toyonaga, A. Machiyama, T. Ueda, K. Takeuchi, S. Oiki and I. Shimada, Functional Equilibrium of the KcsA Structure Revealed by NMR, *Journal of Biological Chemistry*, 2012, **287**, 39634–39641.

[248] A. H. Kopf, M. C. Koorengevel, C. A. van Walree, T. R. Dafforn and J. A. Killian, A Simple and Convenient Method for the Hydrolysis of Styrene-Maleic Anhydride Copolymers to Styrene-Maleic Acid Copolymers, *Chemistry and Physics of Lipids*, 2019, **218**, 85–90.

[249] E. G. Bligh and W. J. Dyer, A Rapid Method Of Total Lipid Extraction And Purification, *Canadian Journal of Biochemistry and Physiology*, 1959, **37**, 911–917.

[250] G. Rouser, S. Fleischer and A. Yamamoto, Two Dimensional Thin Layer Chromatographic Separation of Polar Lipids and Determination of Phospholipids by Phosphorus Analysis of Spots, *Lipids*, 1970, **5**, 494–496.

[251] C. Breyton, W. Javed, A. Vermot, C.-A. Arnaud, C. Hajjar, J. Dupuy, I. Petit-Hartlein, A. Le Roy, A. Martel, M. Thépaut, C. Orelle, J.-M. Jault, F. Fieschi, L. Porcar and C. Ebel, Assemblies of Lauryl Maltose Neopentyl Glycol (LMNG) and LMNG-Solubilized Membrane Proteins, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2019, **1861**, 939–957.

[252] S. Scheidelaar, M. C. Koorengevel, C. A. van Walree, J. J. Dominguez, J. M. Dörr and J. A. Killian, Effect of Polymer Composition and pH on Membrane Solubilization by Styrene-Maleic Acid Copolymers, *Biophysical Journal*, 2016, **111**, 1974–1986.

[253] M. Jamshad, V. Grimard, I. Idini, T. J. Knowles, M. R. Dowle, N. Schofield, P. Sridhar, Y. Lin, R. Finka, M. Wheatley, O. R. T. Thomas, R. E. Palmer, M. Overduin, C. Govaerts, J.-M. Ruysschaert, K. J. Edler and T. R. Dafforn, Structural Analysis of a Nanoparticle Containing a Lipid Bilayer Used for Detergent-Free Extraction of Membrane Proteins, *Nano Research*, 2014, **8**, 774–789.

[254] J. Dominguez Pardo, J. Dörr, M. Renne, T. Ould-Braham, M. Koorengevel, M. van Steenbergen and J. Killian, Thermotropic Properties of Phosphatidylcholine Nanodiscs bounded by Styrene-Maleic Acid Copolymers, *Chemistry and Physics of Lipids*, 2017, **208**, 58–64.

[255] B. Böttcher, D. Scheide, M. Hesterberg, L. Nagel-Steger and T. Friedrich, A Novel, Enzymatically Active Conformation of the Escherichia coli NADH:Ubiquinone Oxidoreductase (Complex I), *Journal of Biological Chemistry*, 2002, **277**, 17970–17977.

[256] A. Musatov, J. Ortega-Lopez, B. Demeler, J. P. Osborne, R. B. Gennis and N. C. Robinson, Detergent-Solubilized Escherichia coli Cytochrome bo3 Ubiquinol Oxidase: A Monomeric, not a Dimeric complex, *FEBS Letters*, 1999, **457**, 153–156.

[257] I. Palombo and D. O. Daley, Heme Incorporation into the Cytochrome bo3 Occurs at a Late Stage of Assembly, *FEBS Letters*, 2012, **586**, 4197–4202.

[258] L. S. P. Rudden and M. T. Degiacomi, Protein Docking Using a Single Representation for Protein Surface, Electrostatics, and Local Dynamics, *Journal of Chemical Theory and Computation*, 2019, **15**, 5135–5143.

[259] S. Iwata, M. Wikström, J. Abramson, S. Riistama, G. Larsson, A. Jasaitis, M. Svensson-Ek, L. Laakkonen and A. Puustinen, The Structure of the Ubiquinol Oxidase from Escherichia coli and its Ubiquinone Binding Site, *Nature Structural and Molecular Biology*, 2000, **7**, 910–917.

[260] W. Arunmanee, R. K. Heenan and J. H. Lakey, Determining the Amphipol Distribution within Membrane-Protein Fibre Samples Using Small-Angle Neutron Scattering, *Acta Crystallographica Section D: Structural Biology*, 2018, **74**, 1192–1199.

[261] J. Gault, J. A. C. Donlan, I. Liko, J. T. S. Hopper, K. Gupta, N. G. Housden, W. B. Struwe, M. T. Marty, T. Mize, C. Bechara, Y. Zhu, B. Wu, C. Kleanthous, M. Belov, E. Damoc, A. Makarov and C. V. Robinson, High-Resolution Mass Spectrometry of Small Molecules Bound to Membrane Proteins, *Nature Methods*, 2016, **13**, 333–336.

[262] F. Hagn, M. L. Nasr and G. Wagner, Assembly of Phospholipid Nanodiscs of Controlled Size for Structural Studies of Membrane Proteins by NMR, *Nature Protocols*, 2017, **13**, 79–98.

[263] Z. Stroud, S. C. Hall and T. R. Dafforn, Purification of Membrane Proteins Free from Conventional Detergents: SMA, New Polymers, New Opportunities and New Insights, *Methods - a Companion to Methods in Enzymology*, 2018, **147**, 106–117.

[264] K. M. Visscher, J. Medeiros-Silva, D. Mance, J. P. G. L. M. Rodrigues, M. Daniëls, A. M. J. J. Bonvin, M. Baldus and M. Weingarth, Supramolecular Organization and Functional Implications of K+ Channel Clusters in Membranes, *Angewandte Chemie International Edition*, 2017, **56**, 13222–13227.

[265] A. M. Giudici, M. L. Molina, J. L. Ayala, E. Montoya, M. L. Renart, A. M. Fernández, J. A. Encinar, A. V. Ferrer-Montiel, J. A. Poveda and J. M. González-Ros, Detergent-Labile, Supramolecular Assemblies of KcsA: Relative Abundance and Interactions Involved, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2013, **1828**, 193–200.

[266] M. L. Molina, F. N. Barrera, A. M. Fernández, J. A. Poveda, M. L. Renart, J. A. Encinar, G. Riquelme and J. M. González-Ros, Clustering and Coupled Gating Modulate the Activity in KcsA, a Potassium Channel Model, *Journal of Biological Chemistry*, 2006, **281**, 18837–18848.

[267] D. Matthies, C. Bae, G. E. Toombes, T. Fox, A. Bartesaghi, S. Subramaniam and K. J. Swartz, Single-Particle Cryo-EM Structure of a Voltage-Activated Potassium Channel in Lipid Nanodiscs, *eLife*, 2018, **7**, e37558.

[268] N. Zijlstra, F. Dingfelder, B. Wunderlich, F. Zosel, S. Benke, D. Nettels and B. Schuler, Rapid Microfluidic Dilution for Single-Molecule Spectroscopy of Low-Affinity Biomolecular Complexes, *Angewandte Chemie*, 2017, **129**, 7232–7235.

[269] D. Wu and G. Piszczek, Measuring the Affinity of Protein-Protein Interactions on a Single-Molecule Level by Mass Photometry, *Analytical Biochemistry*, 2020, **592**, 1–7.

[270] H. I. Zgurskaya, C. A. Löpez and S. Gnanakaran, Permeability Barrier of Gram-Negative Cell Envelopes and Approaches To Bypass It, *ACS Infectious Diseases*, 2015, **1**, 512–522.

[271] A. H. Delcour, Outer Membrane Permeability and Aantibiotic Resistance, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 2009, **1794**, 808–816.

[272] R. O'Shea and H. Moser, Physicochemical Properties of Antibacterial Compounds: Implications for Drug Discovery, *Journal of Medicinal Chemistry*, 2008, **51**, 2871–2878.

[273] H. Nikaido, Molecular Basis of Bacterial Outer Membrane Permeability Revisited, *Microbiology and Molecular Biology Reviews*, 2003, **67**, 593–656.

[274] W. C. Wimley, The Versatile β-Barrel Membrane Protein, *Current Opinion in Structural Biology*, 2003, **13**, 404–411.

[275] S. E. Rollauer, M. A. Sooreshjani, N. Noinaj and S. K. Buchanan, Outer Membrane Protein Biogenesis in Gram-Negative Bacteria, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2015, **370**, 20150023.

[276] G.-W. Li, D. Burkhardt, C. Gross and J. S. Weissman, Quantifying Absolute Protein Synthesis Rates Reveals Principles Underlying Allocation of Cellular Resources, *Cell*, 2014, **157**, 624–635.

[277] C. A. Schnaitman and J. D. Klena, Genetics of Lipopolysaccharide Biosynthesis in Enteric Bacteria, *Microbiological Reviews*, 1993, **57**, 655–682.

[278] T. Wu, J. Malinverni, N. Ruiz, S. Kim, T. J. Silhavy and D. Kahne, Identification of a Multicomponent Complex Required for Outer Membrane Biogenesis in *Escherichia coli*, *Cell*, 2005, **121**, 235–245.

[279] N. Noinaj, A. J. Kuszak, J. C. Gumbart, P. Lukacik, H. Chang, N. C. Easley, T. Lithgow and S. K. Buchanan, Structural Insight into the Biogenesis of β-Barrel Membrane Proteins, *Nature*, 2013, **501**, 385–390.

[280] R. Voulhoux, M. P. Bos, J. Geurtsen, M. Mols and J. Tommassen, Role of a Highly Conserved Bacterial Protein in Outer Membrane Protein Assembly, *Science*, 2003, **299**, 262–265.

[281] T. Wu, A. C. McCandlish, L. S. Gronenberg, S.-S. Chng, T. J. Silhavy and D. Kahne, Identification of a protein complex that assembles lipopolysaccharide in the outer membrane of *Escherichia coli*, *Proceedings of the National Academy of Sciences*, 2006, **103**, 11754–11759.

[282] P. Rassam, N. A. Copeland, O. Birkholz, C. Tóth, M. Chavent, A. L. Duncan, S. J. Cross, N. G. Housden, R. Kaminska, U. Seger, D. M. Quinn, T. J. Garrod, M. S. P. Sansom, J. Piehler, C. G. Baumann and C. Kleanthous, Supramolecular Assemblies Underpin Turnover of Outer Membrane Proteins in Bacteria, *Nature*, 2015, **523**, 333–336.

[283] G. Benn, I. V. Mikheyeva, P. George Inns, J. C. Forster, N. Ojkic, C. Bortolini, M. G. Ryadnov, C. Kleanthous, T. J. Silhavy and B. W. Hoogenboom, Phase Separation in the Outer Membrane of *Escherichia coli*, *Proceedings of the National Academy of Sciences*, 2021, **118**, e2112237118.

[284] P. Rassam, K. R. Long, R. Kaminska, D. J. Williams, G. Papadakos, C. G. Baumann and C. Kleanthous, Intermembrane Crosstalk Drives Inner-Membrane Protein Organization in *Escherichia coli*, *Nature Communications*, 2018, **9**, 1082.

[285] N. G. Housden, J. T. S. Hopper, N. Lukoyanova, D. Rodriguez-Larrea, J. A. Wojdyla, A. Klein, R. Kaminska, H. Bayley, H. R. Saibil, C. V. Robinson and C. Kleanthous, Intrinsically Disordered Protein Threads Through the Bacterial Outer-Membrane Porin OmpF, *Science*, 2013, **340**, 1570–1574.

[286] M. N. Webby, A. O. Oluwole, P. G. Inns, N. G. Housden, A. Olerinyova, C. Pedebos, N. T. Reichmann, S. Khalid, P. Kukura, S. Mohammed, C. V. Robinson and C. Kleanthous, Lipids Mediate Promiscuous Supramolecular Assemblies in the Bacterial Outer Membrane, *submitted*, 2022.

[287] M. L. R. Francis, M. N. Webby, N. G. Housden, R. Kaminska, E. Elliston, B. Chinthammit, N. Lukoyanova and C. Kleanthous, Porin Threading Drives Receptor Disengagement and Establishes Active Colicin Transport through *Escherichia coli* OmpF, *The EMBO Journal*, 2021, **40**, e108610.

[288] J. Shearer, D. Jefferies and S. Khalid, The Outer Membrane Proteins OmpA, FhuA, OmpF, EstA, BtuB and OmpX Have Unique Lipopolysaccharide Fingerprints, *Journal of Chemical Theory and Computation*, 2019, **15**, 2608–2619.

[289] C. Balusek and J. Gumbart, Role of the Native Outer-Membrane Environment on the Transporter BtuB, *Biophysical Journal*, 2016, **111**, 1409–1417.

[290] I. Liko, M. T. Degiacomi, S. Lee, T. D. Newport, J. Gault, E. Reading, J. T. S. Hopper, N. G. Housden, P. White, M. Colledge, A. Sula, B. A. Wallace, C. Kleanthous, P. J. Stansfeld, H. Bayley, J. L. P. Benesch, T. M. Allison and C. V. Robinson, Lipid Binding Attenuates Channel Closure of the Outer Membrane Protein OmpF, *Proceedings of the National Academy of Sciences*, 2018, **115**, 6691–6696.

[291] W. Arunmanee, M. Pathania, A. S. Solovyova, A. P. L. Brun, H. Ridley, A. Baslé, B. van den Berg and J. H. Lakey, Gram-Negative Trimeric Porins Have Specific LPS Binding Sites that are Essential for Porin Biogenesis, *Proceedings of the National Academy of Sciences*, 2016, **113**, E5034–E5043.

[292] R. Solan, J. Pereira, A. N. Lupas, R. Kolodny and N. Ben-Tal, Gram-Negative Outer-Membrane Proteins with Multiple β-Barrel Domains, *Proceedings of the National Academy of Sciences*, 2021, **118**, e2104059118.

[293] D. B. Datta, B. Arden and U. Henning, Major Proteins of *Escherichia coli* Outer Cell Envelope Membrane as Bacteriophage Receptors, *Journal of Bacteriology*, 1977, **131**, 821–829.

[294] P. M. Sharp and B. H. Hahn, Origins of HIV and the AIDS Pandemic, *Cold Spring Harbor Perspectives in Medicine*, 2011, **1**, a006841–a006841.

[295] M. D. Swanson, H. C. Winter, I. J. Goldstein and D. M. Markovitz, A Lectin Isolated from Bananas Is a Potent Inhibitor of HIV Replication, *Journal of Biological Chemistry*, 2010, **285**, 8646–8655.

[296] J.-P. Julien, A. Cupo, D. Sok, R. L. Stanfield, D. Lyumkis, M. C. Deller, P.-J. Klasse, D. R. Burton, R. W. Sanders, J. P. Moore, A. B. Ward and I. A. Wilson, Crystal Structure of a Soluble Cleaved HIV-1 Envelope Trimer, *Science*, 2013, **342**, 1477–1483.

[297] L. Dacheux, A. Moreau, Y. Ataman-Önal, F. Biron, B. Verrier and F. Barin, Evolutionary Dynamics of the Glycan Shield of the Human Immunodeficiency Virus Envelope during Natural Infection and Implications for Exposure of the 2G12 Epitope, *Journal of Virology*, 2004, **78**, 12625–12637.

[298] L. K. Pritchard, D. J. Harvey, C. Bonomelli, M. Crispin and K. J. Doores, Cell- and Protein-Directed Glycosylation of Native Cleaved HIV-1 Envelope, *Journal of Virology*, 2015, **89**, 8932–8944.

[299] V. L. Koshte, W. van Dijk, M. E. van der Stelt and R. C. Aalberse, Isolation and Characterization of BanLec-I, a Mannoside-Binding Lectin from Musa paradisiac (Banana), *Biochemical Journal*, 1990, **272**, 721–726.

[300] J. L. Meagher, H. C. Winter, P. Ezell, I. J. Goldstein and J. A. Stuckey, Crystal Structure of Banana Lectin Reveals a Novel Second Sugar Binding Site, *Vitamins and Coenzymes*, 2005, **15**, 1033–1042.

[301] S. Lusvarghi and C. Bewley, Griffithsin: An Antiviral Lectin with Outstanding Therapeutic Potential, *Viruses*, 2016, **8**, 296–314.

[302] J. T. Hopper, S. Ambrose, O. C. Grant, S. A. Krumm, T. M. Allison, M. T. Degiacomi, M. D. Tully, L. K. Pritchard, G. Ozorowski, A. B. Ward, M. Crispin, K. J. Doores, R. J. Woods, J. L. Benesch, C. V. Robinson and W. B. Struwe, The Tetrameric Plant Lectin BanLec Neutralizes HIV through Bidentate Binding to Specific Viral Glycans, *Structure*, 2017, **25**, 773–782.

[303] D. D. Singh, K. Saikrishnan, P. Kumar, Z. Dauter, K. Sekar, A. Surolia and M. Vijayan, Purification, Crystallization and Preliminary X-Ray Structure Analysis of the Banana Lectin from Musa paradisiaca, *Acta Crystallographica Section D: Structural Biology*, 2004, **60**, 2104–2106.

[304] D. Singh, K. Saikrishnan, P. Kumar, A. Surolia, K. Sekar and M. Vijayan, Unusual Sugar Specificity of Banana Lectin from Musa paradisiaca and

its Probable Evolutionary Origin. Crystallographic and Modelling Studies, *Vitamins and Coenzymes*, 2005, **15**, 1025–1032.

[305] J. Pratap, A. Jeyaprakash, P. Rani, K. Sekar, A. Surolia and M. Vijayan, Crystal Structures of Artocarpin, a Moraceae Lectin with Mannose Specificity, and its Complex with Methyl-α-D-Mannose: Implications to the Generation of Carbohydrate Specificity, *Journal of Molecular Biology*, 2002, **317**, 237–247.

[306] W. I. Weis and K. Drickamer, Structural Basis Of Lectin-Carbohydrate Recognition, *Annual Review of Biochemistry*, 1996, **65**, 441–473.

[307] J. Ortega Arroyo, J. Andrecka, K. M. Spillane, N. Billington, Y. Takagi, J. R. Sellers and P. Kukura, Label-Free, All-Optical Detection, Imaging, and Tracking of a Single Protein, *Nano Letters*, 2014, **14**, 2065–2070.

[308] M. Piliarik and V. Sandoghdar, Direct Optical Sensing of Single Unlabelled Proteins and Super-Resolution Imaging of Their Binding Sites, *Nature Communications*, 2014, **5**, 1–8.

[309] T. Arakawa and J. Wen, Determination of Carbohydrate Contents from Excess Light Scattering, *Analytical Biochemistry*, 2001, **299**, 158–161.

[310] M. Swanson, D. Boudreaux, L. Salmon, J. Chugh, H. Winter, J. Meagher, S. André, P. Murphy, S. Oscarson, R. Roy, S. King, M. Kaplan, I. Goldstein, E. Tarbet, B. Hurst, D. Smee, C. de la Fuente, H.-H. Hoffmann, Y. Xue, C. Rice, D. Schols, J. Garcia, J. Stuckey, H.-J. Gabius, H. Al-Hashimi and D. Markovitz, Engineering a Therapeutic Lectin by Uncoupling Mitogenicity from Antiviral Activity, *Cell*, 2015, **163**, 746–758.

[311] A. Sharma and M. Vijayan, Influence of Glycosidic Linkage on the Nature of Carbohydrate Binding in β-Prism I Fold Lectins: An X-Ray and Molecular Dynamics Investigation on Banana Lectin–Carbohydrate Complexes, *Vitamins and Coenzymes*, 2010, **21**, 23–33.

[312] G. Gupta, S. Vishveshwara and A. Surolia, Stability of Dimeric Interface in Banana Lectin: Insight From Molecular Dynamics Simulations, *International Union of Biochemistry and Molecular Biology Life*, 2009, **61**, 252–260.

[313] G. Ghosh and D. K. Mandal, Novel Unfolding Sequence of Banana Lectin: Folded, Unfolded and Natively Unfolded-Like Monomeric States in Guanidine Hydrochloride, *Biochimie*, 2014, **99**, 138–145.

[314] J. M. Baeten, T. Palanee-Phillips, E. R. Brown, K. Schwartz, L. E. Soto-Torres, V. Govender, N. M. Mgodi, F. Matovu Kiweewa, G. Nair, F. Mhlanga, S. Siva, L.-G. Bekker, N. Jeenarain, Z. Gaffoor, F. Martinson, B. Makanani, A. Pather, L. Naidoo, M. Husnik, B. A. Richardson, U. M. Parikh, J. W. Mellors, M. A. Marzinke, C. W. Hendrix, A. van der Straten, G. Ramjee, Z. M. Chirenje, C. Nakabiito, T. E. Taha, J. Jones, A. Mayo, R. Scheckter, J. Berthiaume, E. Livant, C. Jacobson, P. Ndase, R. White, K. Patterson, D. Germuga,

B. Galaska, K. Bunge, D. Singh, D. W. Szydlo, E. T. Montgomery, B. S. Mensch, K. Torjesen, C. I. Grossman, N. Chakhtoura, A. Nel, Z. Rosenberg, I. McGowan and S. Hillier, Use of a Vaginal Ring Containing Dapivirine for HIV-1 Prevention in Women, *The New England Journal of Medicine*, 2016, **375**, 2121–2132.

[315] E. Mathieu, H. Ritchie, E. Ortiz-Ospina, M. Roser, J. Hasell, C. Appel, C. Giattino and L. Rodés-Guirao, A Global Database of COVID-19 Vaccinations, *Nature Human Behaviour*, 2021, **5**, 947–953.

[316] J. M. Malpica, A. Fraile, I. Moreno, C. I. Obies, J. W. Drake and F. García-Arenal, The Rate and Character of Spontaneous Mutation in an RNA Virus, *Genetics*, 2002, **162**, 1505–1511.

[317] S. Thakur, S. Sasi, S. Pillai, A. Nag, D. Shukla, R. Singhal, S. Phalke and G. S. K. Velu, SARS-CoV-2 Mutations and Their Impact on Diagnostics, Therapeutics and Vaccines, *Frontiers in Medicine*, 2022, **9**, 815389.

[318] J. Hadfield, C. Megill, S. M. Bell, J. Huddleston, B. Potter, C. Callender, P. Sagulenko, T. Bedford and R. A. Neher, Nextstrain: Real-Time Tracking of Pathogen Evolution, *Bioinformatics*, 2018, **34**, 4121–4123.

[319] F. Grabowski, M. Kochańczyk and T. Lipniacki, The Spread of SARS-CoV-2 Variant Omicron with a Doubling Time of 2.0–3.3 Days Can Be Explained by Immune Evasion, *Viruses*, 2022, **14**, 294.

[320] Y. Cao, J. Wang, F. Jian, T. Xiao, W. Song, A. Yisimayi, W. Huang, Q. Li, P. Wang, R. An, Y. Wang, X. Niu, S. Yang, H. Liang, H. Sun, T. Li, Y. Yu, Q. Cui, S. Liu and X. S. Xie, Omicron Escapes the Majority of Existing SARS-CoV-2 Neutralizing Antibodies, *Nature*, 2022, **602**, 657–663.

[321] L. Liu, S. Iketani, Y. Guo, J. F. W. Chan, M. Wang, L. Liu, Y. Luo, H. Chu, Y. Huang, M. S. Nair, J. Yu, K. K. H. Chik, T. T. T. Yuen, C. Yoon, K. K. W. To, H. Chen, M. T. Yin, M. E. Sobieszczyk, Y. Huang, H. H. Wang, Z. Sheng, K.-Y. Yuen and D. D. Ho, Striking Antibody Evasion Manifested by the Omicron Variant of SARS-CoV-2, *Nature*, 2022, **602**, 676–681.

[322] M. Hoffmann, N. Krüger, S. Schulz, A. Cossmann, C. Rocha, A. Kempf, I. Nehlmeier, L. Graichen, A.-S. Moldenhauer, M. S. Winkler, M. Lier, A. Dopfer-Jablonka, H.-M. Jäck, G. M. Behrens and S. Pöhlmann, The Omicron Variant is Highly Resistant Against Antibody-Mediated Neutralization: Implications for Control of the COVID-19 Pandemic, *Cell*, 2022, **185**, 447–456.e11.

[323] S. Cele, L. Jackson, D. S. Khoury, K. Khan, T. Moyo-Gwete, H. Tegally, E. S. James, D. Cromer, C. Scheepers, D. G. Amoako, F. Karim, M. Bernstein, G. Lustig, D. Archary, M. Smith, Y. Ganga, Z. Jule, K. Reedoy, S.-H. Hwa, J. Giandhari, J. M. Blackburn, B. I. Gosnell, S. S. Abdool Karim, W. Hanekom,

NGS-SA, COMMIT-KZN Team, A. von Gottberg, J. N. Bhiman, R. J. Lessells, M.-Y. S. Moosa, M. P. Davenport, T. de Oliveira, P. L. Moore and A. Sigal, Omicron Extensively but Incompletely Escapes Pfizer BNT162b2 Neutralization, *Nature*, 2022, **602**, 654–656.

[324] A. Sigal, R. Milo and W. Jassat, Estimating Disease Severity of Omicron and Delta SARS-CoV-2 Infections, *Nature Reviews Immunology*, 2022, **22**, 267–269.

[325] R. Suzuki, D. Yamasoba, I. Kimura, L. Wang, M. Kishimoto, J. Ito, Y. Morioka, N. Nao, H. Nasser, K. Uriu, Y. Kosugi, M. Tsuda, Y. Orba, M. Sasaki, R. Shimizu, R. Kawabata, K. Yoshimatsu, H. Asakura, M. Nagashima, K. Sadamasu, K. Yoshimura, Genotype to Phenotype Japan (G2P-Japan) Consortium, H. Sawa, T. Ikeda, T. Irie, K. Matsuno, S. Tanaka, T. Fukuhara and K. Sato, Attenuated Fusogenicity and Pathogenicity of SARS-CoV-2 Omicron Variant, *Nature*, 2022, **603**, 700—705.

[326] D. J. Benton, A. G. Wrobel, P. Xu, C. Roustan, S. R. Martin, P. B. Rosenthal, J. J. Skehel and S. J. Gamblin, Receptor Binding and Priming of the Spike Protein of SARS-CoV-2 for Membrane Fusion, *Nature*, 2020, **588**, 327–330.

[327] I. Saha, N. Ghosh, N. Sharma and S. Nandi, Hotspot Mutations in SARS-CoV-2, *Frontiers in Genetics*, 2021, **12**, 753440.

[328] M. R. Rahbar, A. Jahangiri, S. Khalili, M. Zarei, K. Mehrabani-Zeinabad, B. Khalesi, N. Pourzardosht, A. Hessami, N. Nezafat, S. Sadraei and M. Negahdaripour, Hotspots for Mutations in the SARS-CoV-2 Spike Glycoprotein: a Correspondence Analysis, *Scientific Reports*, 2021, **11**, 23622.

[329] P. Han, L. Li, S. Liu, Q. Wang, D. Zhang, Z. Xu, P. Han, X. Li, Q. Peng, C. Su, B. Huang, D. Li, R. Zhang, M. Tian, L. Fu, Y. Gao, X. Zhao, K. Liu, J. Qi, G. F. Gao and P. Wang, Receptor Binding and Complex Structures of Human ACE2 to Spike RBD from Omicron and Delta SARS-CoV-2, *Cell*, 2022, **185**, 630–640.e10.

[330] M. Shah and H. G. Woo, Omicron: A Heavily Mutated SARS-CoV-2 Variant Exhibits Stronger Binding to ACE2 and Potently Escapes Approved COVID-19 Therapeutic Antibodies, *Frontiers in Immunology*, 2022, **12**, 830527.

[331] T. A. Shishir, T. Jannat and I. B. Naser, An *in-silico* Study of the Mutation-Associated Effects on the Spike Protein of SARS-CoV-2, Omicron Variant, *PLoS ONE*, 2022, **17**, e0266844.

[332] Q. Geng, K. Shi, G. Ye, W. Zhang, H. Aihara and F. Li, Structural Basis for Human Receptor Recognition by SARS-CoV-2 Omicron Variant BA.1, *Journal of Virology*, 2022, **96**, e00249–22.

[333] D. Mannar, J. W. Saville, X. Zhu, S. S. Srivastava, A. M. Berezuk, K. S. Tuttle, A. C. Marquez, I. Sekirov and S. Subramaniam, SARS-CoV-2 Omicron

Variant: Antibody Evasion and Cryo-EM Structure of Spike Protein–ACE2 Complex, *Science*, 2022, **375**, 760–764.

[334] Z. Cui, P. Liu, N. Wang, L. Wang, K. Fan, Q. Zhu, K. Wang, R. Chen, R. Feng, Z. Jia, M. Yang, G. Xu, B. Zhu, W. Fu, T. Chu, L. Feng, Y. Wang, X. Pei, P. Yang, X. S. Xie, L. Cao, Y. Cao and X. Wang, Structural and Functional Characterizations of Infectivity and Immune Evasion of SARS-CoV-2 Omicron, *Cell*, 2022, **185**, 860–871.e13.

[335] A. Zoufaly, M. Poglitsch, J. H. Aberle, W. Hoepler, T. Seitz, M. Traugott, A. Grieb, E. Pawelka, H. Laferl, C. Wenisch, S. Neuhold, D. Haider, K. Stiasny, A. Bergthaler, E. Puchhammer-Stoeckl, A. Mirazimi, N. Montserrat, H. Zhang, A. S. Slutsky and J. M. Penninger, Human Recombinant Soluble ACE2 in Severe COVID-19, *The Lancet Respiratory Medicine*, 2020, **8**, 1154–1158.

[336] K. K. Chan, D. Dorosky, P. Sharma, S. A. Abbasi, J. M. Dye, D. M. Kranz, A. S. Herbert and E. Procko, Engineering Human ACE2 to Optimize Binding to the Spike Protein of SARS Coronavirus 2, *Science*, 2020, **369**, 1261–1265.

[337] L. Zhang, S. Dutta, S. Xiong, M. Chan, K. K. Chan, T. M. Fan, K. L. Bailey, M. Lindeblad, L. M. Cooper, L. Rong, A. F. Gugliuzza, D. Shukla, E. Procko, J. Rehman and A. B. Malik, Engineered ACE2 Decoy Mitigates Lung Injury and Death Induced by SARS-CoV-2 Variants, *Nature Chemical Biology*, 2022, **18**, 342–351.

[338] K. K. Chan, T. J. C. Tan, K. K. Narayanan and E. Procko, An Engineered Decoy Receptor for SARS-CoV-2 Broadly Binds Protein S Sequence Variants, *Science Advances*, 2021, **7**, eabf1738.

[339] L. Zhang, K. K. Narayanan, L. Cooper, K. K. Chan, C. A. Devlin, A. Aguhob, K. Shirley, L. Rong, J. Rehman, A. B. Malik and E. Procko, An Engineered ACE2 Decoy Receptor can be Administered by Inhalation and Potently Targets the BA.1 and BA.2 Omicron Variants of SARS-CoV-2, *bioRxiv*, 2022, 2022.03.28.486075.

[340] R. T. Lee and Y. Lee, in *Complex Carbohydrates Part E*, Academic Press, 1987, vol. 138 of Methods in Enzymology, pp. 424–429.

[341] J. J. Lundquist and E. J. Toone, The Cluster Glycoside Effect, *Chemical Reviews*, 2002, **102**, 555–578.

[342] L. S. Powell-Rudden, *Ph.D. thesis*, The Impact of Dynamics in Protein Assembly, University of Durham, 2021.

[343] A. Fiser and A. Sali, ModLoop: Automated Modelling of Loops in Protein Structures, *Methods of Biochemical Analysis*, 2003, **19**, 2500–2501.

[344] A. Vanderheiden, V. V. Edara, K. Floyd, R. C. Kauffman, G. Mantus, E. Anderson, N. Rouphael, S. Edupuganti, P. Shi, V. D. Menachery,

J. Wrammert and M. S. Suthar, Development of a Rapid Focus Reduction Neutralization Test Assay for Measuring SARS-CoV-2 Neutralizing Antibodies, *Current Protocols in Immunology*, 2020, **131**, e116.

[345] K. Kohlstedt, C. Gershome, M. Friedrich, W. Muller-Esterl, F. Alhenc-Gelas, R. Busse and I. Fleming, Angiotensin-Converting Enzyme (ACE) Dimerization Is the Initial Step in the ACE Inhibitor-Induced ACE Signaling Cascade in Endothelial Cells, *Molecular Pharmacology*, 2006, **69**, 1725–1732.

[346] S. Bangaru, G. Ozorowski, H. L. Turner, A. Antanasijevic, D. Huang, X. Wang, J. L. Torres, J. K. Diedrich, J.-H. Tian, A. D. Portnoff, N. Patel, M. J. Massare, J. R. Yates, D. Nemazee, J. C. Paulson, G. Glenn, G. Smith and A. B. Ward, Structural Analysis of Full-Length SARS-CoV-2 Spike Protein from an Advanced Vaccine Candidate, *Science*, 2020, **370**, 1089–1094.

[347] M. T. Degiacomi and M. Dal Peraro, Macromolecular Symmetric Assembly Prediction Using Swarm Intelligence Dynamic Modeling, *Structure*, 2013, **21**, 1097–1106.

[348] Y. Choi, S. Agarwal and C. M. Deane, How Long is a Piece of Loop?, *PeerJ*, 2013, **1**, e1.

[349] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, Y. Wei, J. Xia, T. Yu, X. Zhang and L. Zhang, Epidemiological and Clinical Characteristics of 99 Cases of 2019 Novel Coronavirus Pneumonia in Wuhan, China: A Descriptive Study, *The Lancet*, 2020, **395**, 507–513.

[350] H. Yao, Y. Song, Y. Chen, N. Wu, J. Xu, C. Sun, J. Zhang, T. Weng, Z. Zhang, Z. Wu, L. Cheng, D. Shi, X. Lu, J. Lei, M. Crispin, Y. Shi, L. Li and S. Li, Molecular Architecture of the SARS-CoV-2 Virus, *Cell*, 2020, **183**, 730–738.e13.

[351] R. N. Kirchdoerfer, N. Wang, J. Pallesen, D. Wrapp, H. L. Turner, C. A. Cottrell, K. S. Corbett, B. S. Graham, J. S. McLellan and A. B. Ward, Stabilized Coronavirus Spikes are Resistant to Conformational Changes Induced by Receptor Recognition or Proteolysis, *Sci Rep*, 2018, **8**, 1–11.

[352] J. Zimmerberg and M. M. Kozlov, How Proteins Produce Cellular Membrane Curvature, *Nature Reviews Molecular cell biology*, 2005, **7**, 9–19.

[353] E. P. Barros, L. Casalino, Z. Gaieb, A. C. Dommer, Y. Wang, L. Fallon, L. Raguette, K. Belfon, C. Simmerling and R. E. Amaro, The Flexibility of ACE2 in the Context of SARS-CoV-2 Infection, *Biophysical Journal*, 2021, **120**, 1072–1084.

[354] E. Leikina, H. Delanoe-Ayari, K. Melikov, M.-S. Cho, A. Chen, A. J. Waring, W. Wang, Y. Xie, J. A. Loo, R. I. Lehrer and L. V. Chernomordik, Carbohydrate-Binding Molecules Inhibit Viral Fusion and Entry by Crosslinking Membrane Glycoproteins, *Nature Immunology*, 2005, **6**, 995–1001.

[355] C. Münk, G. Wei, O. O. Yang, A. J. Waring, W. Wang, T. Hong, R. I. Lehrer, N. R. Landau and A. M. Cole, The θ-Defensin, Retrocyclin, Inhibits HIV-1 Entry, *AIDS Research and Human Retroviruses*, 2003, **19**, 875–881.

[356] R. Yan, Y. Zhang, Y. Li, F. Ye, Y. Guo, X. Lu, X. Zhong, X. Chi and Q. Zhou, Structural Basis for the Different States of the Spike Protein of SARS-CoV-2 in Complex with ACE2, *Cell Research*, 2021, **31**, 717–719.

[357] E. Foley, M. Kushwah, G. Young and P. Kukura, Mass Photometry Enables Label-Free Tracking and Mass Measurement of Single Proteins on Lipid Bilayers, *Nature Methods*, 2021, **18**, 1247–1252.