The Influence of Lipid Binding on the Oligomeric State of Membrane Protein Complexes



Joseph A C Donlan

Hertford College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2020

Acknowledgements

Personal

Like many a fresh-faced graduate about embark upon a PhD, if asked in October 2013 how I imagined the upcoming journey to a completed DPhil thesis might go, I'm sure my answer would have been hopelessly naive. The past six years at Oxford have been more challenging than I could ever have expected, but they have also been filled with meaningful experiences, opportunities to learn and grow as a scientist and as a person, and relationships which I know will stay with me for life. There were many moments when I thought I would never make it to this point, but it is certainly true that I would not have done so without the support of the friends and colleagues mentioned below.

First, I would like to thank my supervisor Carol Robinson, who gave me the opportunity to join a world-class lab, the freedom to explore my own ideas, and the confidence to pursue big scientific questions. I am also grateful to Kallol Gupta for all of his support and training, his endless enthusiasm, eclectic knowledge and his infectious love for all things LeuT!

Thanks are also due to many other members of the Robinson and Benesch groups who kindly shared their expertise, knowledge and good humour throughout my time in Oxford. Joe Gault taught me all there is to know about orbitraps and lipids, as well as how to avoid the temptations of 'cowboy science', and was always there for a chat when I needed it. Weston Struwe provided valuable guidance in the early days of the SemiSWEET project, plus good advice and endless carbohydrate expertise. Tim Allison was my molecular biology guru, helping an inexperienced chemist run his first agarose gel and build his first plasmid. And Jon Hopper was always a valuable source of mass spec advice, whether teaching me what a Taylor cone is, or saving the day when stubborn proteins and difficult instruments would not cooperate. Notable mentions also go to Mark, Idi, Cherine, Jani, Jason, Heidi and Georg, all of whom helped me out in various ways and generally made my time in the lab a more enjoyable and fulfilling one, as well as Todd, Helena and Andy for keeping the lab running smoothly over the years. Finally, thanks to my fellow DPhil inmates Shane, Stephen, Miranda, Reid and Gogs for their good conversation, solidarity, and friendship, which kept me sane when things got tough.

There is more to life than science, and my time at Oxford would not have been anything like as rewarding, fun and memorable without the following people from Hertford, the DTC and afar. One's best friends are those who stick around when things get tough, so particular thanks are due to those who were there for me as I dealt with with the entrance of chronic illness into my life in 2016, from which I have emerged a stronger and wiser person. Thank you, therefore, to: Julia, Ali C., Joe, Ross, Eleanor, Farah, Luiza, Michael, Jen, Alison, Vicky, Ali W., Marianne, Sam, Madalena, Jemma, Henry, Alice, Hannah S, Isi, Ros, Lydia, Faaris and Hannah T.

And finally, thanks to my family for their never-ending and unconditional support and love.

Institutional

I am grateful to the Engineering and Physical Sciences Research Council for funding my DPhil, and to my advisor David Gavaghan and all the staff at the Life Sciences Interface Doctoral Training Centre for their advice and support.

Abstract

A large fraction of membrane proteins assemble into long-lived, non-covalent complexes, in a process called oligomerisation. Oligomerisation confers many benefits on membrane proteins, from increased thermal stability and access to a greater diversity of potential functions, to greater opportunities for cellular regulation. Despite this ubiquity and importance, membrane protein oligomerisation, and the role of the lipid environment in determining oligomeric state are poorly understood.

Non-denaturing mass spectrometry (MS) is a versatile method for probing both oligomeric state and protein-lipid interactions and hence is an ideal technique for investigating the relationship between these two phenomena. In this thesis, we report the application of a novel MS-based methodology to examine the oligomerisation of two dimeric membrane proteins: the bioamine transporter LeuT and the sugar transporter SemiSWEET.

For SemiSWEET, we used both high-resolution and high-energy MS methods to characterise a small cohort of lipids which co-purify with the protein. Then, we performed subunit exchange experiments with dimeric SemiSWEET, establishing that it exists in a monomer-dimer equilibrium, the quaternary dynamics of which can be monitored in real time. Finally, by adding exogenous cardiolipin to SemiSWEET, we found that lipid binding shifts the monomer-dimer equilibrium towards the dimeric state.

For LeuT, we observed by MS that the dimeric form, but not the monomeric form, of the protein co-purifies with a larger cohort of lipids, including diacylphospholipids and cardiolipin. Using biochemical experiments and molecular dynamics simulations we ascertained that these lipids bind specifically at the LeuT dimer interface, via a bridging interaction between cardiolipin phosphate groups and cationic residues on the protein subunits, and that these protein-lipid interactions are essential for LeuT dimerisation.

Overall, the results show that the oligomeric states of membrane proteins can be strongly influenced by the local lipid environment, suggesting a potential mechanism for spatial and temporal regulation of protein function.

Contributions

This thesis reports the results of experiments carried out as part of larger collaborative projects within the Robinson Group at the University of Oxford. Whilst the vast majority of the experiments and analyses were conducted by me, I have described some work carried out by others in order to present an intellectually coherent narrative.

For stylistic reasons I have used the third-person pronoun 'we' throughout the thesis. This should be taken by default to mean 'I', unless specifically indicated in the text. I have also included a short section at the end of the two results chapters which summarises the contributions to the work in those specific chapters.

The entire thesis is written by me, in my own words.

Contents

List of Figures **ix**

Glossary of abbreviations **xi**

1 Membrane protein oligomerisation and the role of lipids **4**

1.1 Protein oligomerisation: ubiquity and importance 4
1.2 Protein oligomerisation in the membrane 6
1.2.1 Challenges in studying membrane protein oligomerisation . 9
1.3 Transient oligomers - weak and strong 12
1.4 Strong transient oligomers - what are the molecular triggers which modulate oligomeric state? 15
1.4.1 Ligand-induced oligomerisation 15
1.4.2 Post-translational modifications 17
1.5 The role of the lipid environment in determining membrane protein structure 18
1.5.1 How does the lipid environment affect oligomeric state? . . . 20
1.5.2 Other evidence for the role of lipids in oligomerisation 29
1.6 Oligomerisation dynamics in the membrane 32
1.7 Research focus 34

2 Non-denaturing Mass Spectrometry **36**

2.1 Development of non-denaturing mass spectrometry 37
2.2 The non-denaturing MS experiment 40
2.3 Applications to membrane proteins 47
2.3.1 Determining oligomeric states 47
2.3.2 Identifying endogenous lipids 48
2.3.3 Investigating the effects of protein-lipid interactions: exogenous lipids 51
2.3.4 Characterising dynamic processes 52
2.4 Expanding the mass spectrometry toolkit 53
2.5 High resolution MS 57
2.6 The orbitrap mass analyser 59

2.7 Resolution 59
2.8 Adapting the orbitrap to native MS 61
2.9 Characterising lipid-dependent oligomerisation by mass spectrometry 64

3 SemiSWEET: characterising a lipid-dependent monomer-dimer equilibrium 65
3.1 SWEETs and SemiSWEET 66
3.2 SemiSWEET co-purifies with endogenous lipids 68
3.3 High energy MS elucidates composition of SemiSWEET-lipid complex 69
3.4 High resolution MS characterises cardiolipin molecules directly bound to SemiSWEET 71
3.5 SemiSWEET binds cardiolipins with longer chain lengths than those in the bulk membrane bilayer 75
3.6 SemiSWEET exists in a dynamic equilibrium between monomeric and dimeric forms 77
3.7 Cardiolipin binding shifts the monomer-dimer equilibrium towards to the functional dimer 81
3.8 Discussion 84
3.9 Materials and Methods 88
3.9.1 Protein production and purification 88
3.9.2 Non-denaturing mass spectrometry 89
3.9.3 High energy non-denaturing mass spectrometry experiments 90
3.9.4 Characterising cardiolipin bound to SemiSWEET 91
3.9.5 Characterising cardiolipin in the bulk *E. coli* membrane . . . 91
3.9.6 Monitoring subunit exchange of SemiSWEET dimers 93
3.9.7 Cardiolipin titration experiment 93
3.9.8 Contributions 94

4 LeuT: A Dimer Dependant on Lipids 95
4.1 LeuT: insights from published structures 96
4.2 LeuT is a dimer which forms a non-covalent adduct with endogenous lipids 98
4.3 High energy native mass spectrometry identifies LeuT-bound lipids 101
4.4 Site-directed mutagenesis pinpoints lipid binding to the dimer interface 102
4.5 MD simulations reveal a cardiolipin binding mode which bridges the LeuT dimer interface 105
4.6 Cardiolipin binding is essential for LeuT dimerisation 109
4.7 Delipidation and Relipidation of LeuT *in vitro* Highlights Lipid-Dependent Dimerisation 110
4.8 Discussion 113

4.9 Materials and Methods 117
4.9.1 Molecular cloning and plasmid construction 117
4.9.2 Protein purification 118
4.9.3 Non-denaturing mass spectrometry 119
4.9.4 Chemical delipidation 121
4.9.5 Molecular dynamics simulations 121
4.9.6 Contributions 122

5 Discussion and Further Work **123**
5.1 Probing the local lipid environment 124
5.2 Characterising bound lipids 125
5.3 Measuring oligomeric states 128
5.3.1 Characterising oligomerisation dynamics 129
5.4 How many other oligomeric protein structures are dependant on protein-lipid interactions? 130
5.5 Next steps - transient oligomers 136

Bibliography **137**

List of Figures

1.1 Membrane protein oligomers of varying sizes 7
1.2 Transient and constitutive oligomers 11
1.3 Free energy profiles for transient oligomerisation 14
1.4 Survey of membrane protein structures 23
1.5 Survey of membrane protein structures, cont. 27
1.6 SM18 binding to the transmembrane domain of p24 30
1.7 A model of how lipid binding affects VDAC oligomerisation 31

2.1 Schematic of the Q-ToF Mass Spectrometry platform used for non-denaturing MS of membrane proteins 43
2.2 Schematic and voltage diagram of the modified high energy source of the Q-ToF mass spectrometer 55
2.3 The key stages of the HE-nMS experiment 56
2.4 Selective dissociation of protein-lipid complexes of MATE and exogenous phosphatidylglycerol by HE-nMS 58
2.5 Schematic of the orbitrap analyser 60
2.6 Schematic of the Q-Exactive instrument used for high resolution MS experiments. 63

3.1 Crystal structure of dimeric SemiSWEET 67
3.2 Mass spectrum of the SemiSWEET monomer bound to endogenous lipids 69
3.3 Structures of main *E. coli* lipids 70
3.4 HE-nMS analysis of the 6+ charge state of SemiSWEET monomer-lipid complex. 72
3.5 High resolution mass spectrum of SemiSWEET shows the monomer binding to cardiolipin at isotopic resolution. 73
3.6 Monoisotopic zero-charge mass spectrum of SemiSWEET bound to cardiolipin 74
3.7 Characterisation of cardiolipins extracted from the bulk *E. coli* membrane. 76

3.8 Comparison of the total acyl chain lengths of cardiolipin molecules bound directly to SemiSWEET to those extracted from the total *E. coli* membrane. 77
3.9 Mass spectrum of SemiSWEET, shows presence of SemiSWEET monomers and dimers. 78
3.10 SemiSWEET dimers undergo subunit exchange, implying a dynamic monomer-dimer equilibrium 80
3.11 Mass spectrum of SemiSWEET monomers and dimers, bound to exogenous cardiolipin 82
3.12 Plot showing the relative abundances of SemiSWEET monomers and dimers after addition of different concentrations of exogenous cardiolipin. 83

4.1 Crystal structure of dimeric LeuT 97
4.2 Mass spectrum of LeuT reveals monomers and mass-shifted dimers. 98
4.3 Mass spectrum of ELIC pentamers bound to endogenous phospholipids 100
4.4 HE-nMS analysis on the 23+ dimer charge state of LeuT shows dimer adduct is made up of diacylphospholipids and cardiolipin. 102
4.5 HE-nMS analysis of LeuT dimer adduct, dissociated with increasing collision energy. 103
4.6 Mass spectrum of the LeuT dimer interface mutant F488A/Y489A, showing LeuT monomers with no lipid bound. 104
4.7 Coarse-grained molecular dynamics simulations of LeuT in a model membrane identify lipid binding sites 107
4.8 Cardiolipin binds to positively charged residues on both sides of the LeuT dimer interface. 108
4.9 Mass spectrum of the LeuT mutant K376A/H377A, which disrupts the cardiolipin binding site, shows only LeuT monomers. 110
4.10 Mass spectrum of LeuT, expressed in a cardiolipin-deficient *E. coli* strain, shows LeuT monomers only. 111
4.11 Chemical delipidation of LeuT reversibly abrogates the dimer. . . . 112

5.1 Predicted stability of membrane protein oligomers across the PDB . 132
5.2 Mass spectra of NhaA and NapA Na^+/H^+ antiporters 133
5.3 HE-nMS analysis of the NhaA dimer 135

Glossary of abbreviations

AAC . . . . . The mitochondrial ATP/ADP carrier.

APT . . . . . The lipid aminopentanetetrol.

ASU . . . . . The asymmetric unit of a crystal.

C_8E_4 . . . . . The detergent tetraethylene glycol monooctyl ether.

CCS . . . . . Collision cross-section.

CGMD . . . . Coarse-grained molecular dynamics.

CID . . . . . Collision-induced dissociation.

CL . . . . . The lipid cardiolipin.

CMC . . . . . Critical micelle concentration.

DC . . . . . Direct current, one of the two components of the voltage applied to the quadrupole.

DDM . . . . . The detergent dodecyl maltoside.

EGF . . . . . Epidermal growth factor.

EGFR . . . . . Epidermal growth factor receptor, a receptor tyrosine kinase.

ELIC . . . . . A pentameric ligand-gated ion channel from *Erwinia chrysanthemi*.

ESI . . . . . Electrospray ionisation.

FTICR . . . . A Fourier transform ion cyclotron resonance mass spectrometer.

GFP . . . . . Green fluorescent protein.

GPCR . . . . . G-protein-coupled receptor.

HCD cell . . . Higher-energy, C-trap dissociation cell.

HE-nMS . . . High-energy, non-denaturing mass spectrometry.

IMAC . . . . . Immobilised metal-ion affinity chromatography.

IMMS . . . . . Ion mobility mass spectrometry.

LeuT . . . . . The leucine transporter, a dimeric membrane bioamine transporter from *Aquifex aeolicus*.

LCMS . . . . . Liquid chromatography mass spectrometry.

MALS . . . . . Multi-angle light scattering.

MATE . . . . . Monomeric transporter from the multidrug and toxic-compound exporter family, from *Pyrococcus furiosus*.

MCP . . . . . Micro-channel plate.

MD . . . . . Molecular dynamics.

MS . . . . . Mass spectrometry.

MscS . . . . . The heptameric mechanosensitive channel of small conductance from *Escherichia coli*.

NG . . . . . The detergent nonyl glucoside

NhaA . . . . . A dimeric Na^+/H^+ antiporter from *Escherichia coli*.

NSS . . . . . Neurotransmitter sodium symporter.

OG . . . . . The detergent octyl glucoside.

PA . . . . . The lipid phosphatidic acid.

PC . . . . . The lipid phosphatidylcholine.

PDB . . . . . The protein data bank (https://www.rcsb.org)

PG . . . . . The lipid phosphatidylglycerol.

PI . . . . . The lipid phosphatidylinositol.

PIP_2 . . . . . The lipid phosphatidylinositol-4,5-bisphosphate.

PS . . . . . The lipid phosphatidylserine.

PTM . . . . . Post-translational modifications

Q-ToF . . . . . Quadrupole time-of-flight mass spectrometry.

RF . . . . . Radio frequency, one of the two components of the voltage applied to the quadrupole.

SemiSWEET . A dimeric sugar transporter from *Vibrio splendidus*. Bacterial analogue of the eukaryotic SWEET transporters.

SERT . . . . . Serotonin transporter.

SM . . . . . The lipid sphingosylphosphorylcholine.

TMD . . . . . The transmembrane domain of a membrane protein.

VDAC . . . . . A voltage-dependent anion channel found in the inner mitochondrial membrane.

Introduction

Membrane proteins perform a range of essential functions in organisms across all kingdoms of life[1]. Acting in roles such as receptors, channels, transporters and enzymes, membrane proteins adopt a plethora of different shapes, sizes and topologies[2]. One mechanism by which they achieve such structural diversity is via assembly into non-covalent complexes, a process known as oligomerisation. Formation of these complexes can increase the stability of membrane proteins[3], allow them to carry out more complicated functions by bringing together different functional units, and provide an opportunity for cellular regulation of protein function[4].

Oligomers are ubiquitous in biology, with a large fraction of all proteins forming long-lived homomeric or heteromeric assemblies[5]. This is no less true of proteins in the membrane, and there are many known examples of membrane proteins which must form a specific oligomeric state to carry out their function[6;7;8]. However, whilst the literature now contains a large amount of information on membrane protein secondary and tertiary structure (as of August 2019 there are 938 unique membrane proteins for which structures have been reported[2]), studying the quaternary structure of membrane protein oligomers has remained challenging. This is partly because, *in vivo*, membrane proteins reside in the lipid bilayer, and hence a complete understanding of membrane protein quaternary structure must take into account protein-protein and protein-lipid interactions, as well as other properties of the whole membrane such as lateral pressure, fluidity and dynamics[9].

This thesis will focus on just one of these structural aspects: the role of protein-lipid interactions in determining membrane protein oligomeric state. We will show that, using a novel mass spectrometry (MS)-based methodology, it is possible to identify endogenous lipids that co-purify with membrane proteins, thereby gaining insight into the local lipid environment around a membrane protein, and then

assess the effect of the binding of these lipids on the protein's oligomeric state. To demonstrate the potential of this method, we have applied it to investigate the role of lipids in the oligomerisation of two dimeric membrane proteins: the sugar transporter SemiSWEET from *Vibrio splendidus*, and the bioamine transporter LeuT from *Aquifex aeolicus*. We conclude that both SemiSWEET and LeuT can be categorised as transient oligomers, proteins which adopt a different oligomeric state depending on the chemical environment they inhabit, in this case the immediate lipid environment.

Thesis Outline

In Chapter 1 we begin by providing an overview of the current literature on membrane protein oligomerisation, focusing on what is known about the factors which contribute to the stability of membrane protein oligomers and particularly the role of protein-lipid interactions. We introduce the concept of transiently oligomeric membrane proteins, and examine the existing evidence for proteins which adopt a different oligomeric state depending on their lipid environment. Finally we argue that to study the effect of lipid binding on oligomeric state, an approach is required which can determine simultaneously membrane protein oligomeric state and the presence and identity of bound lipids, and propose that a non-denaturing mass spectrometry-based methodology meets these criteria.

In Chapter 2 we outline the non-denaturing mass spectrometry method which forms the basis of most of the experiments reported in later chapters. After detailing our standard MS platform for studying membrane proteins, we focus on two recent modifications, the incorporation of a high-energy source and the employment of a high resolution orbitrap mass analyser, which have allowed us to characterise endogenous lipid binding to membrane proteins in greater detail.

In Chapter 3 we explore the oligomerisation and lipid binding properties of the sugar transporter SemiSWEET. First, we identify a small cohort of phospholipids which co-purify with SemiSWEET, including diacyl phospholipids and cardiolipin. We then use high resolution MS to characterise the distribution of acyl chain lengths of SemiSWEET-bound cardiolipin, and show that this distribution differs from that

of cardiolipin in the bulk *E. coli* membrane. In addition, we find that SemiSWEET exists in a monomer-dimer equilibrium which cardiolipin binding shifts towards the functional dimer. We categorise SemiSWEET as a weak transient oligomer.

In Chapter 4 we focus on the bioamine transporter LeuT. Using evidence from MS experiments and molecular dynamics simulations, we show that a specific cohort of endogenous lipids binds at the LeuT dimer interface. Furthermore, we find that the key protein-lipid interaction involved consists of a bound cardiolipin forming a bridge between the two subunits of the dimer, and that this interaction is essential for maintenance of the LeuT dimeric state. We conclude that LeuT is a strong transient oligomer, since it adopts a dimeric state only in the context of a specific local lipid environment.

In the discussion, we draw upon the findings presented in this thesis to conclude that protein-lipid interactions can play a key role in stabilising the quaternary structure of oligomeric membrane protein complexes, and that a MS-based approach can be employed to gain insight into this process. Furthermore, characterisation of endogenous lipids bound to membrane proteins can aid in identifying which lipids are involved in stabilising oligomers and the mechanisms by which stabilisation occurs. We consider a simple model of the strength of protein-protein interactions at oligomeric interfaces in order to rationalise the differing extents to which lipids are important in determining membrane protein oligomeric states. And, finally, we consider the biological relevance and significance of our conclusions, and suggest where future research might best be targeted.

1

Membrane protein oligomerisation and the role of lipids

Contents

1.1 Protein oligomerisation: ubiquity and importance . . . 4
1.2 Protein oligomerisation in the membrane 6
1.2.1 Challenges in studying membrane protein oligomerisation 9
1.3 Transient oligomers - weak and strong 12
1.4 Strong transient oligomers - what are the molecular triggers which modulate oligomeric state? 15
1.4.1 Ligand-induced oligomerisation 15
1.4.2 Post-translational modifications 17
1.5 The role of the lipid environment in determining membrane protein structure 18
1.5.1 How does the lipid environment affect oligomeric state? 20
1.5.2 Other evidence for the role of lipids in oligomerisation . 29
1.6 Oligomerisation dynamics in the membrane 32
1.7 Research focus 34

1.1 Protein oligomerisation: ubiquity and importance

The biological cell is a highly crowded environment, in which individual protein molecules are constantly coming into contact with one another[10]. Most of these

interactions are short-lived and non-specific, and are therefore of little functional significance[11]. However, in recent decades it has become clear that a large fraction of cellular proteins undergo some kind of long-lived interactions with specific partners, causing them to form assemblies which are both amenable to experimental characterisation and have well-defined biological functions[12]. These assemblies are termed oligomers, and the process of their formation oligomerisation.

Two notable features of protein oligomerisation are apparent from a broad view of the existing literature: its ubiquity and its functional significance. A recent snapshot of protein X-ray crystal structures, reported in a 2015 study[13], calculated that 60% of proteins of known structure self-assemble into homomers, complexes of a single type of subunit, or heteromers, complexes formed from multiple distinct subunit types. In fact, since analysis of protein quaternary structure and oligomer formation by crystallographic methods tends to produce false negatives, this is likely to be an underestimate of oligomer abundance. These numbers suggest that oligomerisation is a process with relevance to a significant majority of cellular proteins, and is therefore a topic worthy of dedicated study.

Protein oligomerisation is important for biological processes, since it expands the range of different functions and mechanisms of regulation available to the cell[4]. Heteromer formation, for example, involves several different functional units coming together into a single complex, and hence allows for multi-step functions to be carried out. For example, multi-enzyme complexes such as the Pyruvate Dehydrogenase Complex can pass a substrate between multiple active sites as it undergoes different enzyme-catalysed reactions, a process which is more efficient than requiring the substrate to diffuse into several different active sites on isolated enzymes[14].

Oligomer formation can also provide a more complex scaffold to support function than could be achieved within a single subunit, as has been observed in oligomeric enzymes with active sites at subunits interfaces[15;16]. Such novel functions can also emerge more quickly in evolutionary time in an oligomeric system, since it is much faster to evolve a new protein interface than a new global protein fold[17].

In homomeric assemblies, coupling of subunits in a protein oligomer can facilitate regulation of protein function, for example through allostery or cooperativity between subunits[18]. In systems where different oligomeric states differ in activity, controlling oligomer formation can therefore provide a mechanism by which the activity of a protein can be tuned.

More generally, generating new functions via self-assembly of several individual protein units is more efficient than via evolution of a larger single polypeptide since synthesis of multiple short polypeptides is much less error-prone.

Finally, oligomerisation can also increase the ability of a protein to resist unfolding, a theory supported by the observation that increasing oligomeric state is a strategy for improving protein stability adopted by thermophilic organisms[19].

Of course, the fact that a particular protein undergoes oligomerisation does not necessarily imply that oligomerisation provides a tangible functional benefit to the host organism[20;17]. In many cases, oligomerisation may have no significant effect on protein activity, but is simply neither selected for or against by evolutionary pressures. Indeed, we shall see later in this chapter that there are several examples of membrane proteins which are known to form higher-order oligomers, but for which the functional benefits of doing so remain unclear. Equally, one can imagine contexts in which oligomer formation is actively detrimental in a biological context. For example, large proteins diffuse slowly, and a hypothetical enzyme acting on a slowly diffusing substrate could be under evolutionary pressure to adopt a low oligomeric state to increase its activity.

1.2 Protein oligomerisation in the membrane

This thesis focuses on protein oligomerisation in one specific context, that of biological membranes. Here, proteins are embedded in the highly hydrophobic phospholipid bilayer, a radically different environment from that inhabited by soluble proteins.

Analyses of existing membrane protein structures suggest that, like soluble proteins, membrane proteins have a high propensity to oligomerise[21;22], and there

are numerous studies in the literature highlighting membrane proteins for which oligomer formation plays a key role in function or regulation[6;7;8]. As shown in figure 1.1, these oligomers can range in oligomeric state from dimers to complexes with more than 10 membrane-embedded subunits.

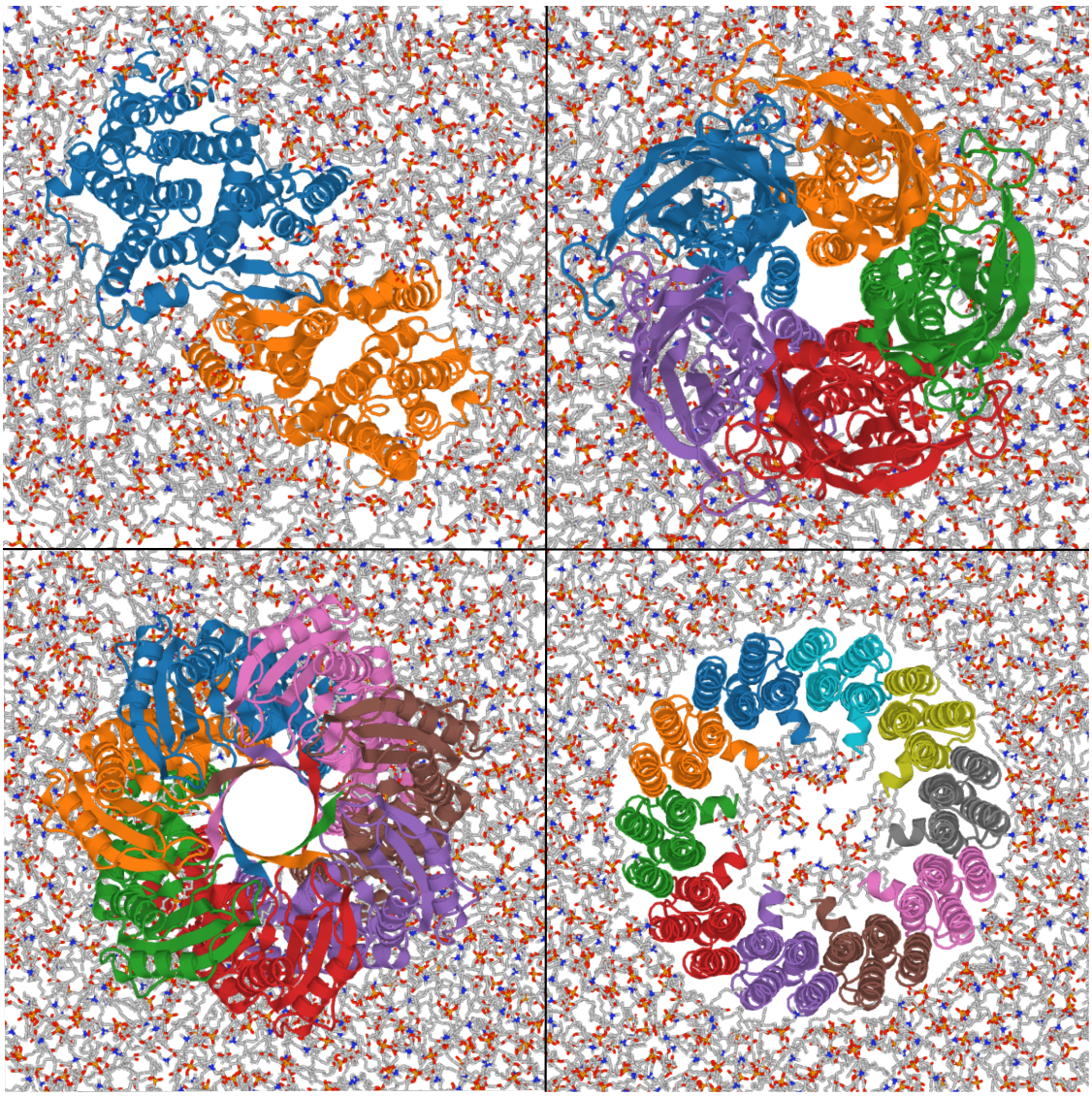


Figure 1.1: Illustrative structures of oligomeric membrane proteins with diverse oligomeric states, embedded in a model lipid bilayers. Examples shown are the Na^+/H^+ antiporter NhaA (dimer, top left), the ligand-gated ion channel ELIC (pentamer, top right), the mechano-sensitive channel MscS (heptamer, bottom left) and the rotor of the V-type Na^+-ATPase (decamer, bottom right). Lipid bilayer-embedded structures were downloaded from memprotmd.bioch.ox.ac.uk - PDB IDs: 4atv, 2yn6, Ynal, 2bl2.[23]

The majority of channel proteins, for example, use assembly of alpha-helical bundle subunits into a homomer in order to form an ion permeation pathway. These

pathways typically consist of residues contributed from each protomer, arranged symmetrically in the membrane around a central pore, through which ions can travel across the membrane[24;25;26].

For membrane-embedded transporter proteins, oligomerisation often plays a more subtle regulatory role, since in general these proteins contain all of the apparatus required for transport within a single protomer. This behaviour has been shown for the trimeric betaine transporter which, while still active in its monomeric form, must form a trimer in order to respond to osmotic stress[27].

Cooperativity has also been shown to be a consequence of oligomer formation for membrane transporters in the major facilitator superfamily such as the lactose transporter LacS[28](positive cooperativity between promoters in a dimer) and the phosphate transporter Pht1 (negative cooperativity between promoters in homomers and heteromers).[29]

A final example of a class of membrane proteins whose oligomerisation has been extensively studied are the G-protein-coupled receptors (GPCRs), a superfamily which is implicated in a large number of cellular signalling processes, and which constitutes the largest class of therapeutic targets. GPCR oligomerisation was suggested in the literature as early as the 1970s[30] and has been a controversial topic ever since[31]. That some GPCRs form homo- and hetero-oligomers in live cells is now broadly accepted[32;33] (indeed, there are now whole databases of examples of observed GPCR oligomers[34]) but it remains unclear in many specific cases what the oligomeric state is *in vivo*, and what the consequences of oligomerisation are for coupling to downstream signalling processes.

The three examples above are taken from a growing body of literature showing the prevalence and importance of oligomerisation for membrane proteins of different structures and biological functions. However, membrane protein oligomerisation remains poorly understood, despite the fact that there is now a substantial body of high resolution structural data published on membrane proteins.

1.2.1 Challenges in studying membrane protein oligomerisation

Studying membrane protein oligomerisation is challenging for a number of reasons. First, membrane proteins are difficult systems to work with biochemically: they generally express in poor yield, can be toxic to heterologous expression systems, are insoluble in water and hence require stabilisation with some kind of membrane mimic, and often require specific post-translational modifications or lipid environments in order to be fully folded and active[35]. The method of solubilisation can have a profound effect on the properties of the purified membrane protein complex, and often requires a trade-off to be made between accurately mimicking the native membrane and being amenable to experimental manipulation. Overall, producing pure, folded, active membrane proteins in large quantities is not a trivial task, although significant progress has been made in addressing many of these difficulties in recent decades[36].

Quaternary structure is a particularly difficult property to study for membrane proteins, since the presence of the solubilising agent (usually a detergent micelle) often interferes with conventional methods for determining stoichiometry or molecular weight. Methods such as analytical ultracentrifugation[37], size-exclusion chromatography[38] or native gel electrophoresis can produce misleading estimates of the stoichiometry of a membrane protein complex, since the complex will migrate, elute or sediment according to the combined mass/size of the protein-detergent complex. Detergents also cause problems in approaches such as multiple angle light scattering (MALS), since the detergent molecules also scatter light, though this can be mitigated by applying an appropriate correction if key properties of the detergent are known[39].

Even X-ray crystallography, the most widely used technique for generating high-resolution structures of proteins, is prone to artifactual measurements of oligomeric state. In many cases, individual subunits can pack together in the crystal in such a way that they form interfaces that resemble those found in 'real' biological oligomers, thereby leading to reports of protein oligomerisation which are

in fact false positives[40]. In addition, the highly artificial conditions under which membrane proteins crystallise (including the absence of key lipids) can cause them to adopt non-native conformations - a process which also may lead to the formation of non-biological oligomers. On the other hand, it is possible for oligomers which do form in the native membrane not to be detected by crystallographic methods, in cases where the oligomer does form in the crystal but the asymmetric unit contains only a monomeric unit, leading to a false negative measurement.[41]

A crucial caveat in interpreting any measurement of the oligomeric state for membrane proteins which have been extracted from the native membrane is that any influence of protein-lipid or lipid-lipid interactions is lost. Since the role of these interactions in mediating oligomeric state is not well understood, it is unclear how closely *in vitro* oligomeric state measurements resemble biological reality. However, there have been a number of reported examples of systems in which the oligomeric state has been found to vary depending on the detergent or membrane mimic used to solubilise the protein complex, suggesting that it cannot be taken for granted that such measurements represent the state of proteins in native membranes.

One set of approaches which, in principle, allow measurements of oligomeric states of proteins in a biologically representative context are fluorescence-based methods[42]. In these methods the membrane protein of interest is fused to a fluorescent protein, usually one of the many variants of green fluorescent protein (GFP). These fusion proteins can be expressed and their oligomeric state determined *in vivo*, or after reconstitution into artificial bilayer mimics. In step-wise photobleaching, this is achieved by careful laser excitation of individual oligomers, and measurement of the fluorescent emission over time. Prolonged excitation results in photobleaching of the fluorescent proteins, and so the intensity of the emitted light will decrease in a step-wise fashion as the individual fluorescent proteins in the oligomer bleach. Since there is a 1:1 ratio of fluorescent proteins to membrane protein subunits, counting the number of separated bleaching steps allows the oligomeric state to be measured without disrupting the protein structure or environment.

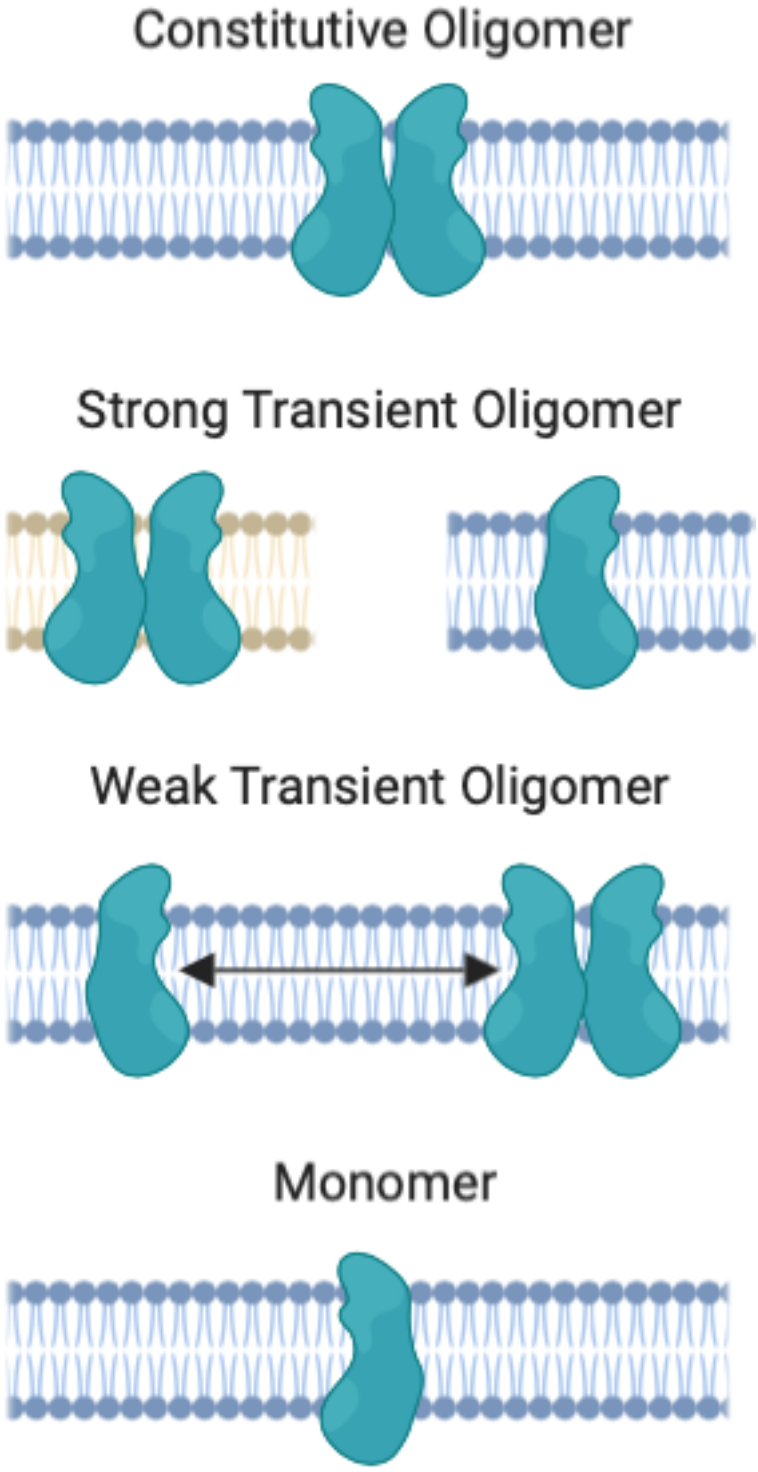


Figure 1.2: Schematic showing the different categories of transient oligomer for a hypothetical dimeric membrane protein. Constitutive oligomers adopt a single oligomeric state, whereas monomeric proteins do not form oligomers. Strong transient oligomers adopt a different oligomeric state depending on the molecular context, or in response to molecular triggers, and weak transient oligomers exist in a dynamic equilibrium between different oligomeric states.

1.3 Transient oligomers - weak and strong

So far we have discussed oligomerisation as a binary phenomenon, classifying proteins into those which form higher order oligomers and those which exist exclusively in a monomeric form. This classification is oversimplified, however, since it neglects a third category of proteins which do not adopt a single oligomeric state, but rather can be observed in different oligomeric states. These proteins are termed *transient oligomers*, of which there are two categories, weak and strong[43], as illustrated in figure 1.2. Strong transient oligomers are those which adopt a fixed oligomeric state, but can change oligomeric state in response to a molecular trigger, such as ligand binding or post-translational modification. Weak transient oligomers, on the other hand, adopt a distribution of oligomeric states which interconvert as part of a dynamic equilibrium. The ratios of the different oligomeric states, and the rate of interconversion between the different states will depend on the fundamental thermodynamics and kinetics of the equilibrium, which in turn can be influenced by the local biological environment.

Specifically, it is important to distinguish between stability and transience when discussing transient or dynamically interconverting oligomers. Like all chemical reactions, oligomerisation is an equilibrium process. The position of the equilibrium, say between monomer and oligomer, will depend on the relative stability (difference in free energy) of the two species i.e. the thermodynamics, whereas whether the two species will interconvert on a reasonable timescale will depend on the size of the activation barrier to interconversion i.e. the kinetics.

Figure 1.3 illustrates these principles in more detail for a hypothetical protein which can form a dimer. Three free energy profiles are shown for the dissociation reaction: ●● $\rightleftharpoons$ 2 ●, where ●● represents the dimer and ● the monomer. In 1.3a the dimeric state is much more stable than the monomeric state i.e. ΔG for the dissociation reaction is large and positive, meaning that at equilibrium almost all of the protein will be dimeric. In 1.3b, the monomeric and dimeric states are much closer in stability, and therefore it is possible in this case that non-negligible populations of both monomer and dimer would be present at equilibrium (with the

dimer as the larger population). However, the kinetic barriers to association and dissociation ($\Delta G^{\ddagger}_{assoc.}$ and $\Delta G^{\ddagger}_{dissoc.}$ respectively) are still large, implying that both dimer formation and dissociation are very slow. This case would correspond to non-interconverting oligomeric states. In the final scenario, 1.3c, the monomer and dimer are of comparable stability, and the kinetic barriers to their interconversion are small. We would expect such a profile to correspond to a system in which there are observable populations of both states, and that these states would be interconverting i.e. the dimer would be a truly transient oligomer.

Transient interactions are crucial for a large number of biological processes, many of which are dependent on hugely complex networks of protein-protein interactions[44]. Transient protein oligomerisation forms just one part of this large network of interactions, and is of interest because, by having an oligomeric state which can be modulated depending on the biological context, transiently oligomeric proteins provide an opportunity for the cell to respond to changes in its environment, crucial for example for processes such signal transduction, and to regulate function[43;20].

Of course, understanding how all of the steps from molecular trigger to a change in the oligomeric state distribution of a particular protein to an effect on either that protein's function or some downstream signalling process is no small task. For membrane proteins, we have already seen that accurately determining oligomeric states is challenging using conventional structural biology techniques, but these challenges are exacerbated when studying transient oligomers. In the case of strong oligomers, where the oligomeric state depends on the biological context, if the quaternary structure of the protein complex is measured under artificial 'non-physiological' conditions *in vitro* then the oligomeric state detected may bear no resemblance to that of the protein in its native cellular environment. For weak transient oligomers, the objective for experimentalists is to be able to resolve and quantify the different populations in order to determine accurately the distribution of oligomeric states, and also to be able to observe oligomerisation dynamics in

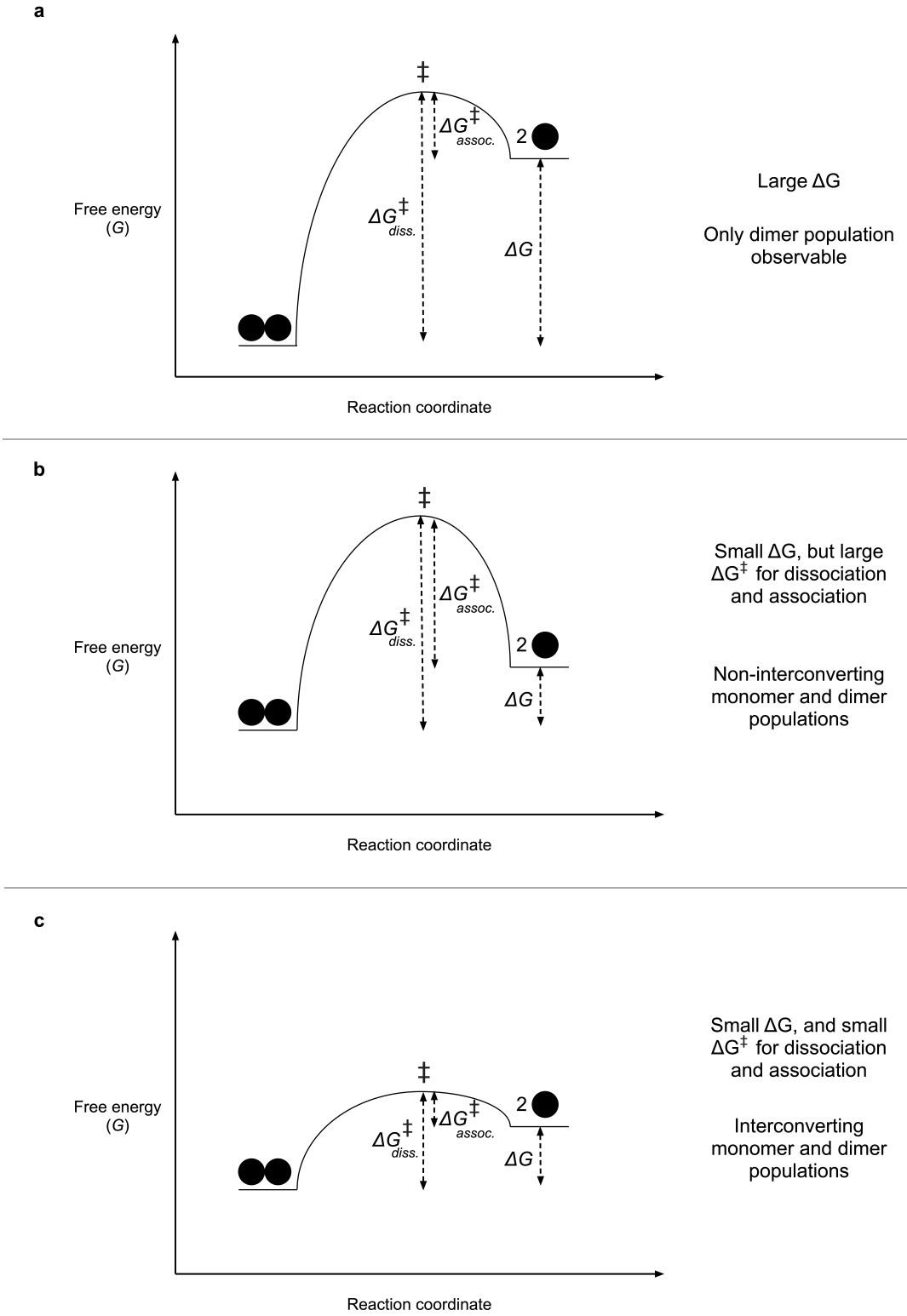


Figure 1.3: Free energy profiles for a hypothetical monomer-dimer equilibrium: ●● $\rightleftharpoons$ 2 ●, under different conditions. **a** Dimer much more stable than monomer, ($\Delta G >> 0$) therefore protein exists almost entirely as a dimer. **b** Dimer and monomer are of comparable stability, but kinetic barriers to association and dissociation, ($\Delta G^{\ddagger}_{assoc.}$ and $\Delta G^{\ddagger}_{dissoc.}$) are large. We would predict detectable populations of both states but not interconverting. **c** Dimer and monomer are of comparable stability and kinetic barriers to association and dissociation are low. Corresponds to interconverting monomer and dimer populations.

order to extract kinetic parameters for the oligomerisation equilibria - both of these objectives are challenging to achieve with conventional structural biology techniques.

Despite these challenges, there has been some progress in identifying individual examples of strong and weak transient oligomers which has been reported in the literature.

1.4 Strong transient oligomers - what are the molecular triggers which modulate oligomeric state?

In searching the literature for examples of strong transient oligomers, we are considering proteins which exhibit different oligomeric states in different molecular contexts or change oligomeric state in response to a molecular trigger. Here we will review three different molecular triggers which have been shown to affect oligomeric state of membrane proteins: ligand binding, post-translational modification and protein-lipid interactions, the latter of which will form the focus of the experiments in this thesis.

1.4.1 Ligand-induced oligomerisation

Ligand-induced oligomerisation, where a protein adopts a different oligomeric state in response to the binding of a small molecule, has long been of interest to biochemists because of its relevance for cellular signalling. In multi-cellular organisms, a common method by which a cell receives an external signal is by the interactions of an extracellular ligand with receptors in the plasma membrane. The current understanding is that binding of this ligand induces some structural change in the receptor which then affects interactions between the intracellular portion of the receptor and intracellular proteins which, in turn, leads to various downstream signalling processes within the target cell. Elucidating the quaternary structure of membrane protein receptors, and their structural dynamics in response to ligand binding, is therefore crucial for understanding molecular signalling and has also been a significant focus of efforts in the drug discovery field. One aspect of membrane protein structure which can be modulated upon ligand binding is oligomeric state

- indeed there are decades-old reports of ligand-induced dimerisation potentially having some role in intercellular signalling in the context of the immune response[45].

One paradigmatic example of a system for which the mechanism of receptor activation is well-understood is the epidermal growth factor receptor (EGFR)[46]. This tyrosine kinase receptor, which consists of an extracellular ligand-binding domain, a transmembrane domain and cytoplasmic kinase domain, exists in a monomeric state in the membrane in the absence of EGF. This configuration represents a deactivated state which does not interact productively with intracellular signalling proteins. Binding of EGF causes conformational changes in the extracellular domain of EGFR, such that it can dimerise with another EGF-bound receptor. Homo-dimer formation causes the intracellular domains on the two subunits to be brought into close proximity, resulting in cross-phosphorylation of tyrosine residues on the two kinase domains. The phosphorylated dimeric state is the activated state of the receptor, since the phosphorylated residues also serve as docking sites for SH2 and PTB domains of signalling proteins that are recruited and go on to activate downstream signalling processes. Once the extracellular signal has stopped, and therefore the concentration of the extracellular ligand decreases, the receptor returns to monomeric state and is dephosphorylated, resulting in deactivation of signal propagation.

For other types of receptor, for example in GPCRs, the role of ligand binding in modulating oligomeric state and in particular the functional consequences of oligomerisation are highly controversial. There is certainly evidence that many different GPCRs form higher order oligomers, but it is often unclear whether this oligomerisation is required or even relevant to ligand binding and signal propagation[47].

It also remains unclear to what extent other classes of membrane protein undergo changes in oligomeric state when interacting with ligands, but it is likely that the examples from the receptor class are just the tip of the iceberg.

1.4.2 Post-translational modifications

Post-translational modifications (PTMs) are important in modulating protein-protein interaction networks and are therefore a crucial mechanism by which various cellular processes can be regulated. Indeed, proteins which undergo PTMs have been shown to play a more central and connected role in these networks than those that do not[48], since their affinity for various interaction partners can be tuned depending a specific pattern of modifications. As well as affecting interactions with other proteins, PTMs can affect the affinity of a protein subunit for itself, and therefore its oligomeric state.

Examples of systems whose propensity to oligomerise is altered by PTMs abound in the soluble protein literature. These include cases in which a single PTM causes a specific change in oligomeric state such as the transcription factor STAT5[49] which forms dimers upon phosphorylation, and the *E. coli* acetyl transferase PatZ, which switches from a tetrameric to an octameric form upon acetylation[50]. PTM-driven oligomerisation may also be implicated in disease processes linked to protein aggregation, as suggested by studies showing that several different alpha-synuclein PTMs increase the propensity of the protein to form toxic oligomers[51].

PTM-linked oligomerisation is less well-documented for membrane proteins, but nevertheless there have been some elegant examples reported recently. The HIV viral protein Vpu, for example, forms a range of higher-order oligomers, but it has been observed that the degree of oligomerisation is inversely correlated with the degree of phosphorylation of two serine residues[52]. In the case of the plant plasma membrane H^+-ATPase, on the other hand, phosphorylation of a C-terminal threonine residue causes the dimeric complex to recruit an additional regulatory protein and assemble into a hexameric complex, a process which results in increased ATPase activity[53]. Finally, there is also strong evidence to suggest that certain class C GPCRs form stable dimers which are stabilised by disulfide bonds across the dimer interface[47], though it is debatable whether these covalently bonded dimers would fall into the class of transient oligomers once the disulfide bond has formed.

1.5 The role of the lipid environment in determining membrane protein structure

The examples of ligand binding and post-translational modification outlined above illustrate an important principle: the oligomeric state of a membrane protein is highly dependent on its molecular context. One hugely important aspect of a membrane protein's molecular context is its immediate lipid environment. Being situated in the biological membrane means that membrane proteins experience multiple interactions with lipids, over different length-scales, all of which have the potential to affect the secondary, tertiary and quaternary structure of the protein[54]. There is also a huge diversity of lipid chemical structures[55]- a typical animal cell has around 1,000 major lipid types, but tens of thousands of minor lipids - the classification of which has required the development of a new field of lipidomics[56]. These facts mean that the complete network of protein-lipid interactions in which each membrane protein participates is a dauntingly complex one.

When categorising protein-lipid interactions, it is instructive to divide membrane lipids into 3 classes, depending on their proximity to the protein of interest. The first of these classes is the so-called structural lipids, which are defined as those which form specific interactions with membrane proteins, often with high affinity, and usually in well-defined binding pockets or grooves. Binding of structural lipids has been shown in multiple examples to be crucial for protein function. For example, gating of the potassium channel Kir2.2 requires binding of the lipid PIP_2 in a pocket defined by specific positively charged residues, a process which is therefore important for the regulation of resting membrane potential[57]. In contrast, the TRAAK potassium channel has a substrate binding cavity which can be blocked by a lipid - it is hypothesised that the protein has a lipid-bound non-conductive state and a non-lipid-bound conductive state, showing that protein-lipid interactions can directly impact function[58]. Tightly bound lipids can also enhance the stability of membrane proteins, as shown for example in the case of the human β_2-adrenergic receptor, which has specific binding site for cholesterol. Cholesterol binding at this site confers a significant increase in stability on this receptor[59].

The remainder of the protein's membrane-embedded surface is solvated by a ring of lipids known as the lipid annulus. Annular lipids are thought to form weaker interactions with membrane proteins than structural lipids do, and in some cases freely exchange with lipids in the rest of the membrane[60]. Determining the lipid composition of the annulus is challenging, since the relatively weaker protein-lipid interactions here often to not survive the disruption required to extract membrane proteins from the membrane for structure determination experiments. Nevertheless, it has been found that the composition of the lipid annulus can be important for membrane protein function. The *E. coli* lactose permease LacY, for example, has been shown to be unable to facilitate uphill transport of lactose when it is assembled in membranes which lack phsophatidylethanolamine (PE)[61]. There is no specific PE binding site on LacY - this effect is due to a misorientation in a transmembrane helix which occurs when the lipid annulus carries too high a negative charge. Increasing the PE content of the membrane reduces the overall negative charge (since the other predominant lipids are phosphatidyl glycerol (PG) and cardiolipin (CL), both of which are anionic), reorients the helix correctly, and restores function. Identifying whether the structural consequences of adding a particular lipid to a membrane protein are a result of interactions in specific binding sites or a change in annular composition can be achieved by testing how the magnitude of the effect scales with the quantity of added lipid, as well as how sensitive it is to lipid class. For example, connexin channels are inactive in purely phosphatidylcholine (PC) membranes, but gain activity as the proportion of anionic phosphatidylserine (PS) lipids in the membrane increases. This effect increases up until PS makes up 60% of the membrane, and is reproducible when either phosphatidylinositol (PI) or phosphatidic acid (PA), both also ionic lipids, are added, suggesting that the activity is influenced by formation of an annulus of anionic lipids[62].

Finally, bulk lipids are those which make up the rest of the membrane and therefore do not form direct interactions with the membrane protein of interest. However, this does not mean that bulk lipids have no effect on the structure and function of proteins since the lipid composition of a membrane determines its

global properties, which in turn affect protein structure. For example, membrane properties such as lateral tension, curvature, fluidity and hydrophobic thickness have all been shown to affect membrane protein structure[63;64;65].

It is worth caveating the model of structural-annular-bulk lipids, with a brief consideration of the relative numbers of lipids and proteins in a typical membrane. The framework above assumes a classical 'fluid-mosaic' description of the membrane, in which sparsely distributed proteins float in a sea of lipids. However, if we consider that a typical *E. coli* membrane is around 60 % protein, 40 % lipid, and take reasonable estimates of a typical membrane protein mass (30 kDa) and lipid mass (600 Da), we can estimate that there will only be around 50 lipids per protein in the membrane[66]. If it takes 20-30 lipids to cover the surface of a membrane protein i.e. to form an annulus, then there are not enough lipids even to form two full layers around each protein. This phenomenon is similar for mammalian cell membranes, whose protein content can range from 20 - 75 % depending on the cell or organelle type[67]. In short, this means that in practice membrane proteins are very closely spaced in the membrane, they will come into contact with one another very frequently just by random diffusion in the membrane, and they are not surrounded by multiple layers of lipids. Oligomerisation of membrane proteins also therefore takes place against a background of competing non-productive protein-protein interactions, as well as protein-lipid interactions.

1.5.1 How does the lipid environment affect oligomeric state?

Having established that protein-lipid interactions affect various aspects of membrane protein structure and function, we will now focus on the relationship between these interactions and membrane protein oligomeric state. This relationship is of particular interest, since it can provide a mechanism by which the activity of cellular proteins can be regulated. It is well-established that different membranes within an organism can have markedly different lipid compositions[55]. More recently, evidence has also emerged that the relative abundance of different lipids can also vary across a

contiguous membrane, and that membranes within a single cell can have complex sub-structure, with micro-domains of varying lipid composition[68;69]. In addition, the lipid profile of a membrane can vary significantly over time, for example throughout the cell cycle, or in response to circadian rhythms[70;71]. Hence one can speculate that proteins whose oligomeric state is determined by the local lipid environment have the potential to be regulated both spatially and temporally, providing an opportunity for the cell to tune the activity of these proteins depending on their molecular context.

Evidence from membrane protein crystal structures

Since the early days of membrane protein crystallography, many examples have emerged of structures which contain resolvable lipid molecules[72]. These structures depict bound lipids in all three of the 'layers' described in section 5.1: in highly specific binding sites (structural), solvating the membrane protein surface (annular), and not in direct contact with the protein (bulk). The structures also highlight a diversity of protein-lipid interactions, involving a wide range of amino acid residues. Before we focus on structural evidence for lipids being involved in oligomerisation, in order to provide context we will first present several representative examples of high resolution structures from across the protein data bank which feature lipids bound to membrane proteins, highlighting the residues which bind lipids, the chemical interactions involved, and the potential impact of these interactions on the protein structure/function if known.

Perhaps unsurprisingly, structural lipids are over-represented in published structures, since these lipids are more likely to remain attached to the protein through the purification and crystallisation process and also have low conformational flexibility so are more likely to be resolved. One example of a protein complex which is thought to contain a highly specific lipid binding site is the Na^+/K^+ ATPase. A recent high resolution structure (shown in figure 1.4a) of the phosphorylated, ouabain-bound form of the protein revealed a cholesterol binding within the transmembrane region, interacting with residues from both the α and γ subunits[73]. The cholesterol molecule is well-defined in the electron density maps, suggesting a

tight, specific interaction. The binding pocket is defined by a series of hydrophobic residues (Trp, Phe, Val, Tyr, Pro) which form hydrophobic interactions with the hydrocarbon rings on the cholesterol, with the hydroxyl group on ring A able to hydrogen bond to the backbones of a Val and Gly residue. The authors show that cholesterol binding stabilises the activity of this enzyme, and that this stabilisation is not as effective when the structurally related sterol ergesterol is added, providing additional evidence that the protein-lipid interaction is highly specific.

Another membrane protein structure which contains highly resolved cholesterol molecules is that of the human A_{2A} adenosine receptor (A_{2A}AR) (PDB ID 4EIY[75]). The GPCR structure contains three cholesterol binding sites on the extracellular half of the receptor, two of which are at crystal contacts. The third however is more interesting, in that it occupies a hydrophobic groove along one of the transmembrane helices, forming extensive contacts with the aromatic sidechain of a phenylalanine residue. This Phe residue is conserved as a hydrophobic residue (Ile, Val or Phe) in all receptors of the adenosine family, suggesting this hydrophobic lipid pocket has some structural significance. Hydrogen bonding interactions between the cholesterol hydroxyl group and the main chain carbonyl of a serine residue also help to secure the cholesterol molecule in the binding groove. Binding of the cholesterol in the groove is thought to stabilise helix VI, fixing a functionally important asparagine residue in the ligand binding pocket of the receptor, as well as contributing to overall thermostability of the protein.

In some cases, where a specific lipid can be doped into a membrane protein sample for crystallography, it is possible to observe the effect of specific protein-lipid interactions on protein structure by comparing the structure with and without exogenous lipids. This was illustrated in an elegant structural study by Wharton *et al.* on the GIRK2 potassium channel. In this structure (PDB ID 3SYA), one molecule of the lipid PIP_2 is found bound to each subunit of the protein, with its headgroup positioned at the interface between the transmembrane domain and the extracellular domain. The acyl chains are in contact with the transmembrane helices, forming hydrophobic interactions with non-polar residues. The three phosphate

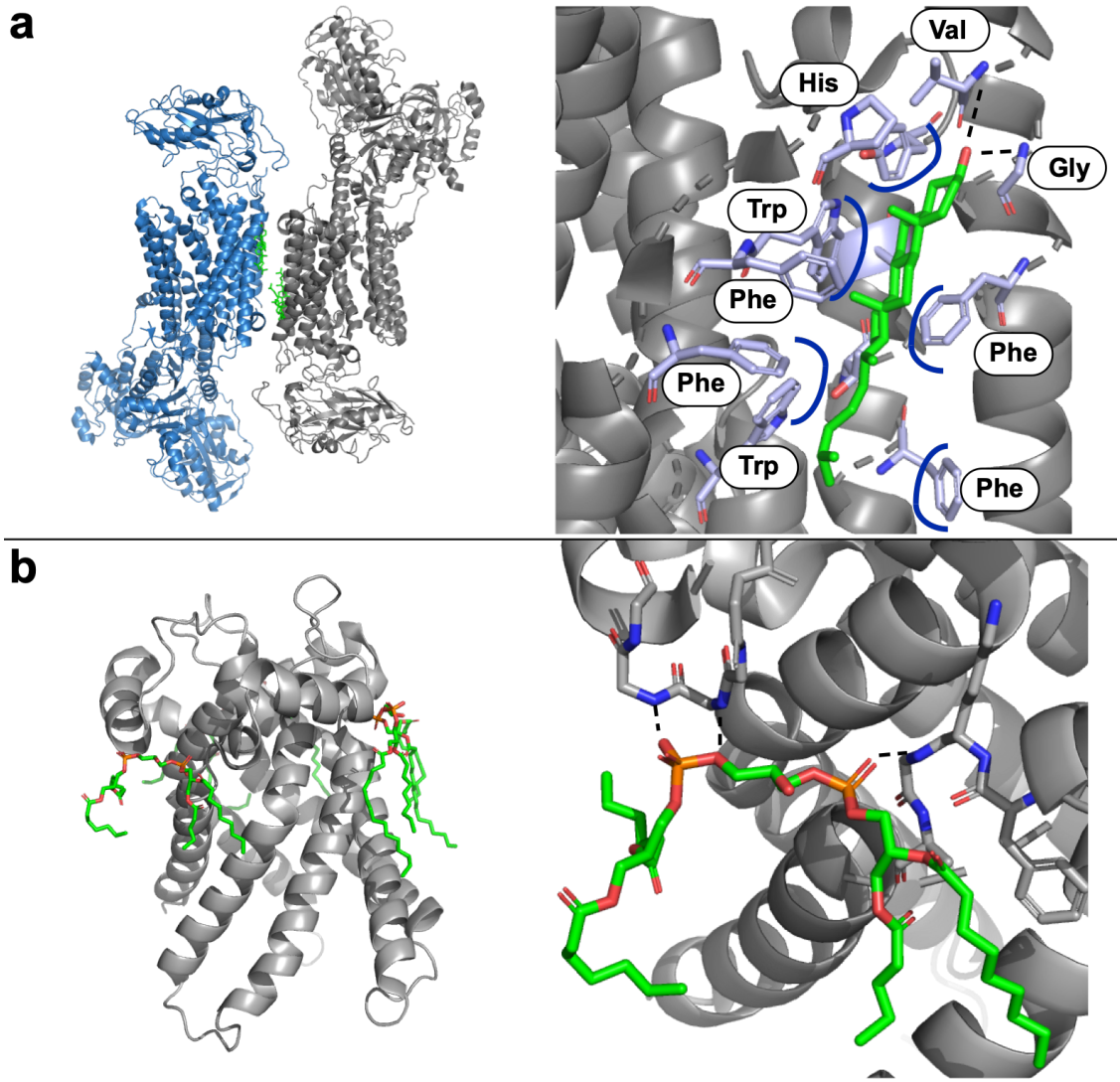


Figure 1.4: Illustrative crystal structures of membrane proteins containing resolved bound lipids. **a** Structure of Na^+/K^+ ATPase bound to ouabain, PDB ID 4HYT[73]. The structure consists of two heterotrimeric proteins (coloured blue and grey), each of which is bound to two cholesterol molecules (green). Zooming in on one of the cholesterol binding sites (right panel) shows a well-defined network of interactions which form a tight pocket. Key interactions are hydrophobic interactions (blue semicircles) with the hydrocarbon rings and tail, as well as hydrogen bonding between backbone nitrogen atoms and the cholesterol hydroxyl group. **b** Structure of the monomeric mitochondrial ATP/ADP carrier (AAC), PDB ID 2C3E[74]. The AAC monomer is bound to 3 cardiolipin molecules, which bind principally via their headgroups. Key molecular interaction is hydrogen bonding between backbone nitrogen atoms and the two negatively charged cardiolipin phosphates.

groups in the headgroup mean that this region of the lipid is highly negatively charged - this is complemented by a series of positively charged lysine residues in the binding pocket which form charge-charge interactions with the phosphates. There are also hydrogen bonds between the polar hydroxyl groups on the inositol ring and backbone amide nitrogens. By comparing the *apo*- and PIP_2-bound structures, the authors show that PIP_2 binding leads to a small rotation of some inner helices near the channel gate, though alone this binding is not sufficient to effect a change in the open/closed state of the channel.

Not all membrane protein structures contain lipids which bind to the protein via both their head groups and hydrocarbon tails. In the case of the case of the mitochondrial ATP/ADP carrier (AAC), the structure of which is shown in figure 1.4b[74], the protein forms high-affinity protein-lipid interactions with lipid head groups only. Three cardiolipin molecules bind the monomeric protein, which normally resides in the inner mitochondrial membrane (a membrane environment which is enriched in cardiolipin). As will become important later, cardiolipin has two phosphate groups connected by a three-carbon linker. In the AAC structure, each of these phosphate groups is coordinated via hydrogen bonding interactions with backbone nitrogens. In contrast to other structures with specific hydrophobic binding pockets, in this structure the four hydrophobic tails of the cardiolipins splay out away from the protein surface, or only make minimal contact with surface residues. The structural and functional significance of these protein-lipid interactions is as yet unclear.

Conversely, in some structures it is the acyl tails of the lipid which appear to form more well-defined interactions with a membrane protein than do the headgroups. An example which demonstrates this is structure of the mitochondrial respiratory complex II (PDB ID 1ZOY[76]). In this structure, two phospholipids are identified whose acyl chains extend along the axis of the transmembrane domain of the protein. However, whilst the acyl chains have a very defined structure in the electron density maps, binding tightly to a series of hydrophobic residues, the headgroup binding was poorly resolved. The authors tentatively assigned the lipid as PE, based on

the low-resolution density being most consistent with PE out of the most abundant mitochondrial membrane lipids, but the lack of definition in the headgroup region suggests that protein-lipid interactions in this system are dominated by hydrophobic interactions with the acyl lipid tails.

The cytochrome bc_1 complex, part of the electron transport chain in eukaryotes, is another system in which a phospholipid binds in a highly specific binding site within the transmembrane region, as shown in figure 1.5a. In a structure of the complex from *S. cerevisiae*, two PE lipid binding sites were observed within each bc1 subcomplex; the figure illustrates the interactions involved in binding of PE to one of the sites. The positively charged quaternary amine within the PE headgroup forms hydrogen bonds with the carboxyl group of a glutamic acid side chain, whilst the the phosphate group is coordinated via hydrogen bonds to two tyrosine hydroxyl groups, and to an ordered water molecule. A tryptophan then coordinates this water molecule, completing an ordered network of interactions around the head group. A series of aromatic amino acid side chains (Phe, Tyr) then form hydrophobic interactions with the PE's two hydrophobic tails. The authors suggest that role of the PE lipids in this structure is to form a mediating layer between the membrane-spanning part of the protein, and therefore to aid in orienting the complex vertically in the bilayer.

Another component of the electron transport chain, cytochrome c oxidase, also has several solved structures which contain resolved lipids. In one notable structure, of a bovine heart cytochrome c oxidase (PDB ID 2DYR[77]) 13 different lipids were detected, each fitting neatly into a binding site on the protein surface. The structure most clearly shows specific protein-lipid interactions for 2 different lipids which bind to subunit III, a PG and acardiolipin. For PG, the glycerol hydroxyls in the head group act as hydrogen bond donor with a histidine residue and an ordered water molecule, which is in turn coordinated to the backbone carbonyl of an alanine. The negatively charged phosphates are coordinated by backbone nitrogen H-bonds from argininie and methionine residues, and even the carbonyl part of the ester groups in the fatty acid tails are hydrogen bonded to

a tryptophan sidechain and a leucine backbone nitrogen. For the cardiolipin, as seen in other structures, the phosphate groups define the head group binding - in this structure they act as H-bond acceptors from glutamine, histidine, and an ordered water, as well as being involved in charge-charge interactions with a lysine residue. Interestingly, the acyl chains of the PG lipids are found to adopt a nearly superimposable conformation in cytochrome C oxidase structures across species, suggesting that these are conserved interactions where lipids intercalate within helices rather than merely sticking to regions on the protein surface.

PE lipid binding was also observed in a recent crystal structure of the photosynthetic reaction centre from *Thermochromatium tepidum* (PDB ID 1EYS[78]). In this structure PE binds in a distinct binding site on the transmembrane domain. The binding site consists of positively charged residues (an arginine and a lysine) which can form salt bridges with the negatively charged PE phosphate group. There is also a tyrosine hydroxyl group and glycine backbone nitrogen which are appropriately positioned to act as hydrogen bond donors to the phosphate. The acyl tails then slot neatly into a cleft between a transmembrane helix from one subunit and a helical bundle from two other subunits - this also allows it to form hydrophobic interactions with the tail groups of two chlorophyll a molecules which protrude from the reaction centre complex.

One remarkable structure in which low-affinity resolved around a membrane protein complex is a 2005 structure of the *E. coli* aquaporin Aqp0, depicted in figure 1.5b[79]. In this structure, Aqp0 forms a tetramer, whose surface is solvated by a complete annulus of DMPC lipids. The authors were able to identify 9 lipids per protein monomer, of which 7 appeared to form interactions with amino acid residues on the protein. Whilst these lipids form extensive protein contacts, resolution was poor at the C2 position of the glycerol within the DMPC headgroup, suggesting a lack of selectivity for the biological enantiomer. MD simulations of Aqp0 in a model bilayer showed that the lipids at these sites have short residence times, implying that the sites represent preferential attachment sites rather than tight binding pockets, and that the lipids here are in frequent exchange with those in the bulk membrane[80].

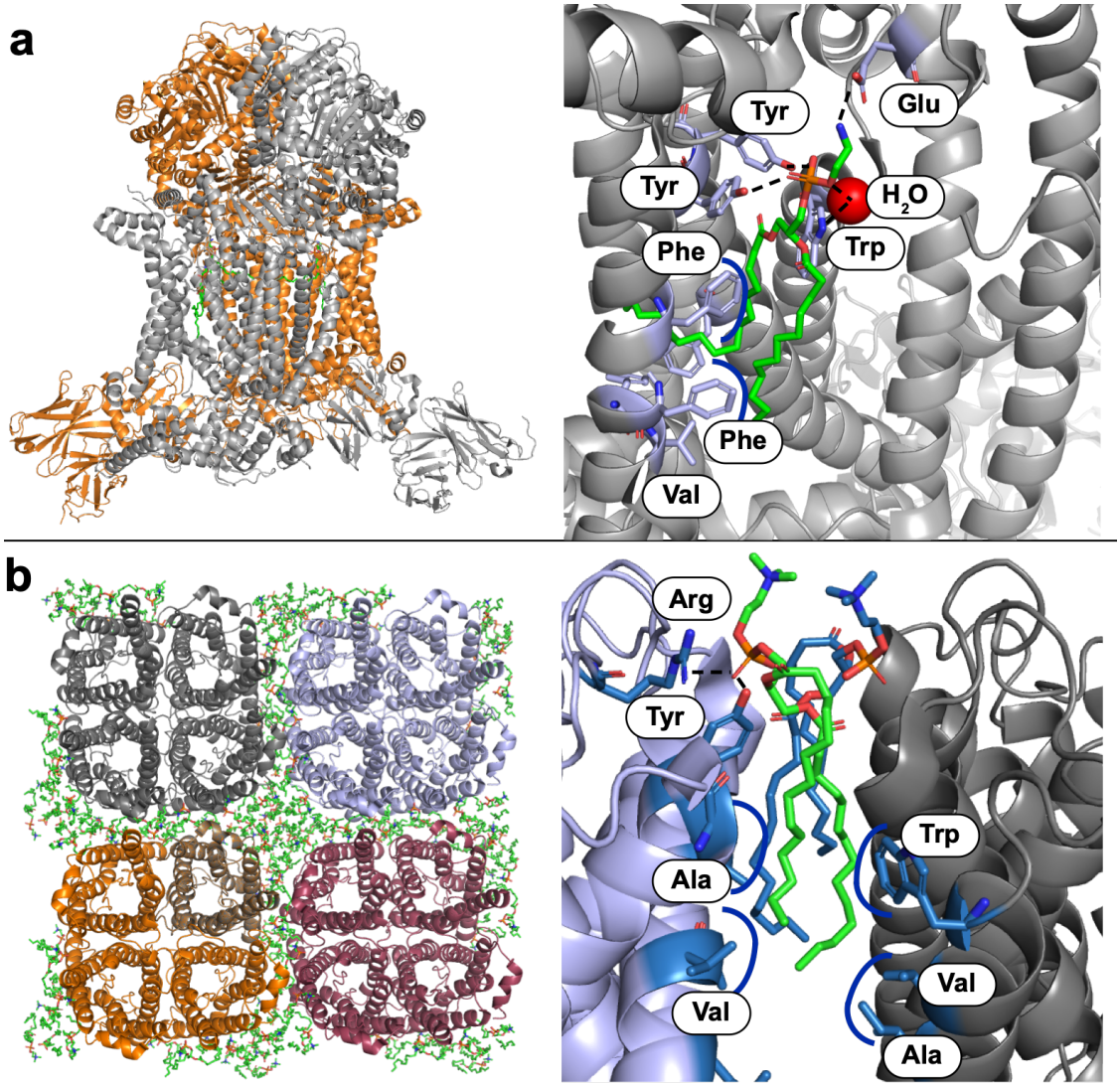


Figure 1.5: Illustrative crystal structures of membrane proteins containing resolved bound lipids. **a** Structure of the cytochrome bc_1 complex from *S. cerevisiae*, two interacting copies of which are shown here (grey and orange), along with two bound PE lipids (green). Zoom-in shows one of the PE binding sites, which consists of an ordered network of interactions around the head and tail groups. The amine of the ethanolamine headgroup forms hydrogen bonds (black dotted lines) with a glutamine, whilst the phosphate is coordinated by two tyrosines and an ordered water molecule (which itself is then hydrogen bonded to a tryptophan on the protein). Hydrophobic interactions (blue semicircles) between aromatic or branched hydrocarbon sidechains and the aliphatic lipid tails are also observable in the structure. **b** Structure of *E. coli* aquaporin Aqp0, surrounded by an annular shell of DMPC lipids; PDB ID 2B6O[79]. Each symmetry equivalent monomer is associated with 9 DMPC lipids, which are associated to varying degrees with the protein subunits. Zoom-in shows one DMPC lipid (green) and its protein-lipid and lipid-lipid interactions. The phosphate in the head group forms a salt bridge with a positively charged arginine residue, and also accepts a hydrogen bond from a nearby tyrosine. The hydrophobic tails pack against a symmetry-equivalent lipid (dark blue), as well as forming hydrophobic interactions with numerous hydrophobic residues from from two different Aqp0 tetramers.

Despite the lower affinity of binding compared to the structural lipids described above, it is still possible to observe well-resolved protein-lipid interactions in this structure, formed as the lipids adapt to the surface of the protein to form a tight seal around the transmembrane region. Figure 1.5b illustrates one such set of interactions - the negatively charged phosphate of the DMPC interacts with an arginine and tyrosine via a salt bridge (Arg) and hydrogen bond (Tyr), and the hydrophobic tails pack against other lipids, as well as forming hydrophobic interactions with residues from two neighbouring Aqp0 molecules (Leu, Val, Ala, Trp, Ile).

Focusing on lipids in interfaces

The ten examples discussed above illustrate the wide range of membrane protein structures which have now been solved that contain lipids, and the various different binding modes and chemical interactions that can be involved. Most important for this thesis, however, are structures in which lipids were observed in sites corresponding to the interfaces between subunits in oligomeric membrane proteins. Observation of these lipids in multiple structures prompted researchers to speculate that 'interfacial' lipids might play some role in stabilising the quaternary structure of membrane proteins. Examples have been reported from membrane proteins of varied oligomeric states, functions and from various organisms and include the trimeric betaine transporter[81], the dimeric β_2-adrenergic receptor[82], the *Anabaena* sensory rhodopsin[83], and a 13-subunit cytochrome C oxidase[84]. Merely observing lipids at sites close to subunit interfaces is not sufficient evidence to conclude that such lipids are important for oligomerisation, however. If the protein-protein interactions between subunits are strong, then the oligomeric complex may be highly stable even in the absence of specific lipids, with the observed lipids in the crystal structure simply filling in the available space in grooves in the interface. As mentioned earlier, lipids can also be observed at artifactual interfaces in crystal structures which do not exist in a biological context. A notable example of this occurred for the mitochondrial ATP/ADP carrier (AAC), described above. The AAC structure contained cardiolipin molecules at what appeared to

be plausible interfaces between protein subunits, leading to the hypothesis that cardiolipin binding could be important for AAC oligomerisation[74]. Later work showed that AAC is in fact monomeric, meaning the 'interfacial lipids' were merely a product of crystal packing[85].

1.5.2 Other evidence for the role of lipids in oligomerisation

An alternative approach to assessing the role of lipids in mediating oligomerisation is to directly measure the oligomeric state of purified membrane proteins in an experimental set-up in which the lipid composition can be controlled. This approach was adopted in two elegant studies by Contreras *et al.* and *Betaneli et al.*[86;87].

Contreras *et al.* investigated the lipid interactions protein p24, a protein involved in COPI vesicle biogenesis. Using Förster resonance energy transfer (FRET) experiments, they found a tight interaction between p24 and a sphingolipid (SM 18), observing FRET between a Trp residue on the protein and a fluorescently labelled SM 18 analogue. The FRET efficiency was sensitive to changes in both the head-group and the hydrophobic chain length, and was not observed for a close analogue of the protein, p23, suggesting that the protein-lipid interaction is highly specific. The authors went on to map the lipid binding site using alanine scanning, and were able to characterise a well-defined binding pocket for SM 18, see figure 1.6. Subsequent cross-linking experiments then showed that p24 adopted a biologically active dimeric state when reconstituted in vesicles containing SM 18, whereas no dimers were detected when reconstituted in vesicles lacking this lipid. Molecular dynamics simulations on SM 18-bound p24 in a model membrane suggested that the most plausible dimer interface is not close to the SM 18 binding site, implying that this could be an allosteric mechanism of lipid induced dimerisation.

Betaneli and colleagues studied the oligomerisation of the voltage-dependent anion channel VDAC, found in the inner mitochondrial membrane, the oligomerisation of which has been found to be linked to cellular apoptosis. The authors incorporated red- and green-labelled VDAC into vesicles containing different lipid compositions and, using fluorescence cross correlation spectroscopy estimated the extent of VDAC

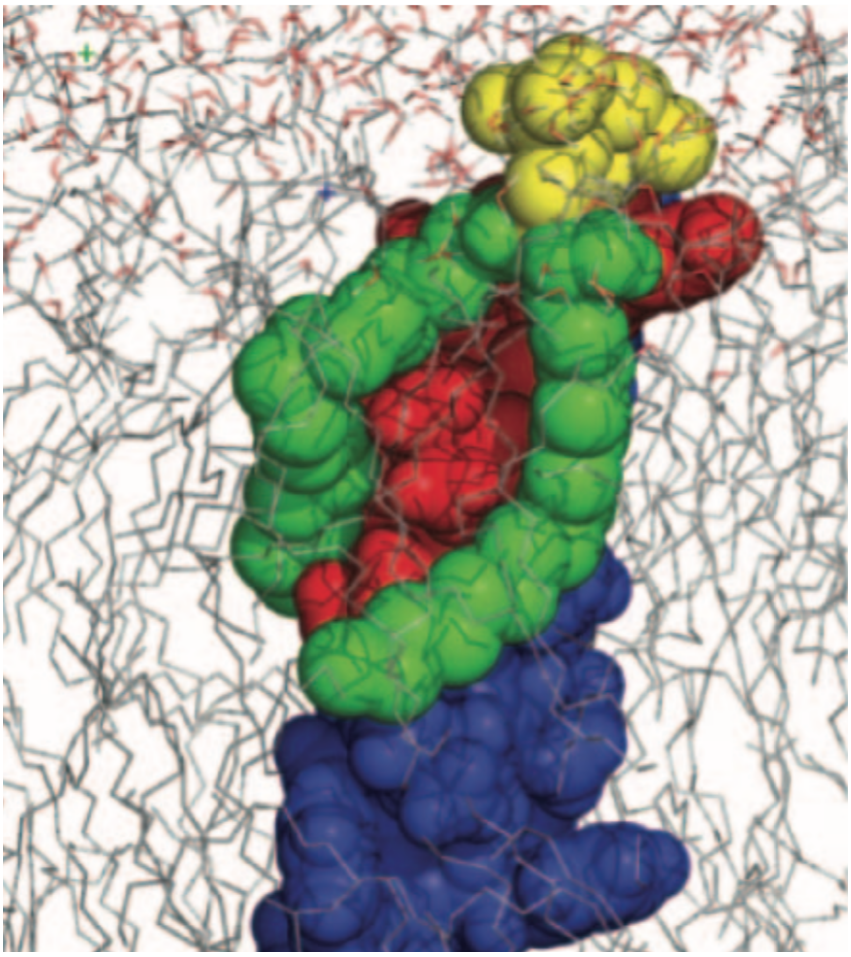


Figure 1.6: Snapshot from molecular dynamics simulations highlighting interaction between the lipid SM 18 and the transmembrane domain (TMD) of p24. SM 18 (head group - yellow, backbone/acyl tails - green) fits into a well-defined binding site on the TMD (binding pocket - red, rest of protein - blue). Binding promotes oligomerisation, despite the binding pocket being distant from the predicted oligomeric interface. Figure adapted from Contreras *et al.*[86])

oligomerisation. They found that increasing the PG-content of the vesicles favoured VDAC oligomerisation whereas a higher CL-content disfavours oligomerisation. This is an interesting observation as these two lipids have very similar head-group chemistry, suggesting that the markedly different effect on oligomeric state is a consequence of the lipid hydrophobic tails. They suggest a model, illustrated in figure 1.7 in which negatively charged PG binds at the oligomeric interface, acting as a bridge between positively charge residues on the VDAC surface. CL also has a negatively properties, but the authors suggest that the 4 hydrophobic tails sterically hinder formation of protein-protein interactions required for oligomer formation.

The relationship between PG and CL abundance and VDAC oligomeric state is notable as during apoptosis the PG content of the mitochondrial inner membrane is known to increase markedly, whilst the CL concentration decreases - therefore this lipid-induced oligomerisation mechanism could provide an explanation for how the cell can regulate apoptosis by changing its lipid composition.

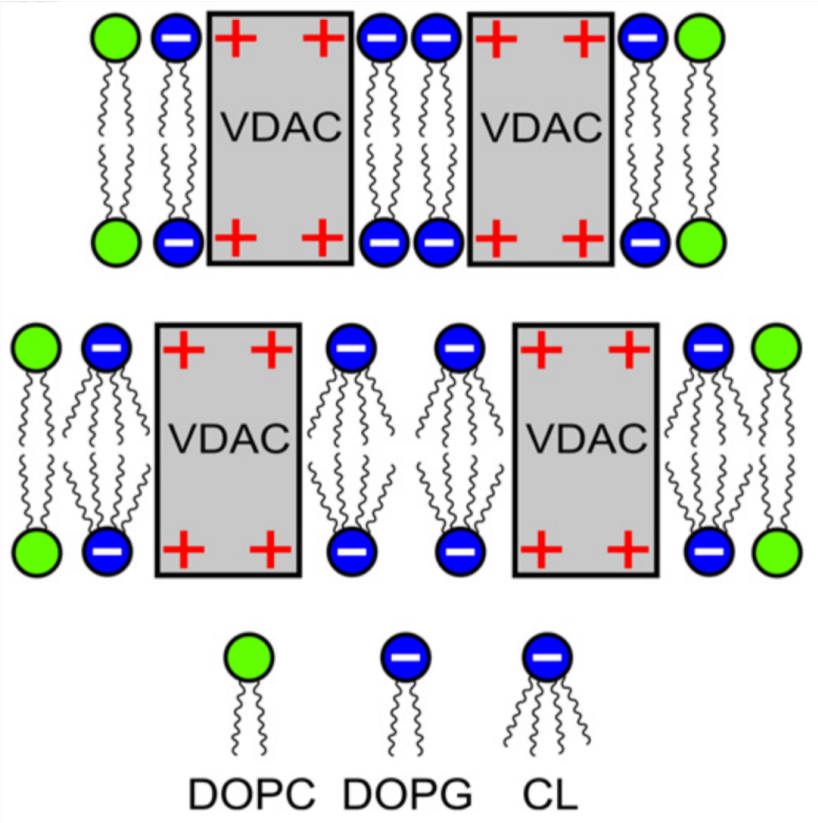


Figure 1.7: A model proposed by Betaneli *et al.* to explain how PG and CL binding influence oligomerisation of the VDAC ion channel. Positively charged residues on the VDAC surface can interact with negatively charged PG and CL head-groups. However, the steric bulk of the acyl chains of these two lipids determines whether these interactions favour or disfavour VDAC association. Whilst the more compact PG can act as a bridge between VDAC subunits, promoting oligomerisation, the bulkier CL with 4 acyl tails sterically hinders protein-protein interactions and therefore disfavours formation of higher-order oligomers. Figure adapted from Betaneli *et al.*[87])

The two examples above both illustrate how a change in lipid composition of a membrane can lead to changes in the oligomeric state of a membrane protein. What is striking is that this that both studies propose different mechanisms by which

these changes are induced. In the case of p24 the SM 18 lipid binds distant to the oligomeric interface, suggesting that the binding induces some conformational change in the protein which favours dimerisation (analogous to an allosteric mechanism). In contrast, the VDAC model suggests that oligomerisation is favoured or disfavoured as a result of lipid binding at the interface.

What the two examples do have in common is that they both involve specific lipids binding, similar to the structural lipids discussed above. There is also evidence for less specific lipid binding events inducing oligomerisation, or at least clustering of membrane proteins. For example, Raja et al. have demonstrated that anionic lipids can stabilise the tetrameric structure of the potassium channel KcsA[88] - this is likely to be a consequence of general properties of the lipid annulus rather than binding in specific sites. Alternatively, even global properties of the membrane can influence the propensity of proteins to oligomerise. Botelho *et al.* have found evidence that oligomerisation of rhodopsin is influenced by bilayer thickness, possibly as a consequence of changes in membrane curvature[89].

One weakness of experiments involving manipulating the lipid environment of a membrane protein and then measuring the impact on oligomeric state is that it requires the experimenter to choose which lipids to investigate. When the membrane protein of interest is purified, it is typically isolated from the native membrane using highly delipidating detergent and therefore most of the native protein-lipid interactions are lost. Deciding which exogenous lipids then to add to the solution in order to create a controlled lipid environment is partly a matter of informed guesswork. As mentioned above, even if the overall lipid composition of the host membrane is known, this may not represent the local lipid environment around the protein. Hence, it can be difficult to know how accurately such experiments reflect the regulation of membrane protein oligomeric state in a biological context.

1.6 Oligomerisation dynamics in the membrane

So far we have considered oligomeric state as a parameter which, for a given protein in a specific molecular context, takes a single value. For those membrane

proteins which form oligomers with very strong protein-protein interactions this is a reasonable assumption - once such oligomers form it is safe to assume that they will persist with a single oligomeric state. However, oligomerisation, like any other chemical reaction, is really an equilibrium process which can in theory be described by an equilibrium constant, and kinetic parameters such as a rate of dissociation and association of subunits. When the position of this equilibrium lies far towards the oligomeric form we have constitutive and strongly transitive oligomers, and when it lies far towards the dissociated subunits we have monomeric proteins. This naturally raises the question of whether there are systems where the equilibrium constant lies somewhere in the middle of these two extremes, such that there are measurable populations of more than one oligomeric state which are dynamically interconverting. These systems would be the weak transient oligomers that were introduced earlier in this chapter.

Observing dynamic oligomerisation equilibria in action requires an experimental technique with specific capabilities. First, it must be a so-called 'single-molecule' technique, as opposed to an ensemble method, since it must be able to quantify the abundance of different oligomeric states within the same sample. In addition, it must be able to characterise dynamic processes in order to show that the different states observed are interconverting. Up to now, single-molecule fluorescence-based methods have been most successfully employed for this purpose, principally for characterising GPCR dynamics. Studies of this type began to appear in the literature in 2010, and showed that receptors such as the M1 muscarinic receptor[90] and the corticotropin-releasing factor receptor[91] exist in the plasma membrane and/or the endoplasmic reticular membrane in at least two different oligomeric states, and that these states interconvert. Kasai *et al.* were even able to carry out a full thermodynamic characterisation of the monomer-dimer equilibrium of the N-formyl peptide receptor[92] in live cells, determining a 2D equilibrium constant and association and dissociation rate constants both in the presence and absence of ligand.

The biological significance of these oligomerisation equilibria is still not fully understood. It has been suggested that the existence of a dynamic equilibrium between oligomeric states with differing activities could provide an opportunity for the activity of a membrane protein to be tuned. This would be a more subtle form of tuning than in the case of the strong transient oligomers described above, potentially involving small changes in the equilibrium position.

1.7 Research focus

Having reviewed the current state of the literature on the relationship between the lipid environment and the oligomerisation of membrane proteins, we will now outline the specific research focus of this thesis.

In order to fully understand how the immediate lipid environment of a membrane protein influences its oligomeric state, three steps are required. First, one must characterise the immediate lipid environment. Are there any structural lipids in well-defined binding pockets, or in oligomeric interfaces? Is there a distinct lipid annulus around the protein? How do these structural, interfacial and annular lipids differ from the bulk membrane lipids in terms of lipid class, chain length or other properties. Secondly, one must characterise the oligomeric state of the membrane protein in a molecular context as close to that of the native membrane as possible. What is the oligomeric state and does the system exist in more than one oligomeric state simultaneously? Do these different oligomeric states interconvert and at what rate? Finally, in order to probe the relationship between the protein-lipid interactions and the oligomeric state, these interactions must be perturbed, disrupted and replaced.

This approach requires an experimental technique which can simultaneously measure a membrane protein's oligomeric state and characterise its bound lipids. In this thesis, we use a native mass spectrometry-based methodology to achieve this, the details of which are outlined in the next chapter. We then apply this methodology to investigate the oligomerisation of two transient oligomeric membrane proteins, SemiSWEET and LeuT. These are bacterial homodimeric proteins and hence serve as tractable model systems for understanding membrane protein oligomerisation. In

chapter 3, we consider SemiSWEET, which we show binds structural lipids which have an aliphatic chain-length distribution which is distinct from that of the bulk membrane. Furthermore we find that SemiSWEET exists in a monomer-dimer equilibrium and we observe interconversion between these two states in real time. Finally we show that the position of this equilibrium can be perturbed upon the addition of specific lipids. These observations allow us to classify SemiSWEET as a weak transient oligomer. For LeuT, see chapter 4, we find a set of tightly bound structural lipids which bind the dimeric state only, and go on to show that these lipids bind at the dimer interface and are required for dimerisation. This allows us to confidently classify LeuT as a strong transient oligomer in which a specific lipid environment dictates the oligomeric state. Overall our results show that mass spectrometry can be used to classify the oligomerisation behaviour of membrane proteins and understand the role of protein-lipid interactions in determining oligomeric state.

2

Non-denaturing Mass Spectrometry

Contents

2.1 Development of non-denaturing mass spectrometry . . 37
2.2 The non-denaturing MS experiment 40
2.3 Applications to membrane proteins 47
2.3.1 Determining oligomeric states 47
2.3.2 Identifying endogenous lipids 48
2.3.3 Investigating the effects of protein-lipid interactions: exogenous lipids 51
2.3.4 Characterising dynamic processes 52
2.4 Expanding the mass spectrometry toolkit 53
2.5 High resolution MS 57
2.6 The orbitrap mass analyser 59
2.7 Resolution 59
2.8 Adapting the orbitrap to native MS 61
2.9 Characterising lipid-dependent oligomerisation by mass spectrometry 64

In the previous chapter we provided an introduction to membrane protein oligomerisation and discussed what is currently known about the role of the local lipid environment in influencing membrane protein oligomeric states. We concluded by noting that, in order to explore this phenomenon, an experimental technique is required that can simultaneously measure the oligomeric state and lipid binding properties of membrane protein complexes. Here, we will argue that a mass

spectrometry-based methodology is capable of performing these measurements, and is therefore an ideal method for studying membrane protein oligomerisation.

We will begin by briefly summarising the history and development of mass spectrometry, with a focus on how it has been adapted for the study of large non-covalent membrane protein complexes. This will be followed by a detailed outline of the current 'state-of-the-art' MS platform which is employed by our laboratory, and examples of prominent studies in which this method has been applied to gain insight into the structure, function, and interactions of oligomeric membrane proteins. Finally, we will discuss two recent modifications to our standard MS platform, the introduction of a high energy source and adoption of a high resolution orbitrap mass analyser, developed immediately before the outset of this study, which have allowed us to characterise protein-lipid complexes in more detail and which form the basis for key experiments described in later chapters.

2.1 Development of non-denaturing mass spectrometry

Whilst mass spectrometry (MS) has been the workhorse of analytical chemistry since the middle of the 20th century, it was not until the development of so-called 'soft' ionisation techniques in the 1980s that MS was first successfully applied to the analysis of intact biological macromolecules[93]. Using hard ionisation methods, which had worked well for small molecules, it was very difficult to generate gas phase ions of large biomolecules without the undesirable side-effect of significant fragmentation. In contrast, using the softer electrospray ionisation (ESI) method, J. B. Fenn and colleagues were able to analyse a full-length intact protein molecule by MS for the first time[94], opening up the possibility of studying, in Fenn's words, 'large, fragile molecules of biochemical importance'.

Not long after these initial successes with denatured proteins it was observed that not only could disruption of covalent bonds within proteins be avoided using ESI but, with careful control of experimental conditions, non-covalent interactions, for example between protein and ligand,[95;96] or between protein subunits in an

oligomeric complex[97;98], could be preserved in the gas phase[99]. Furthermore, examples of studies in which ESI-MS was used to address biological questions, such as resolving debates on oligomeric states of protein complexes, soon emerged[100]. These were the founding studies of the field of non-denaturing MS (also known as native MS), and rapidly changed the technique in the eyes of biologists from a technical curiosity to a method which had the potential to provide biologically relevant insight into the structure and dynamics of biomolecules[101]. For the purposes of this thesis, we define non-denaturing MS to encompass mass spectrometry approaches involving electrospray ionisation of biological analytes from non-denaturing solvents[102]. The word non-denaturing strictly refers to the state of the analyte before ionisation (i.e. in solution), rather than upon detection, although in many cases the goal of the non-denaturing MS experiment is to preserve the native-like structure of the analyte ions for the full duration of the measurement.

By the turn of the millennium, analysis of non-covalent biomolecular complexes by MS had become routine, with the capabilities of the method being expanded to tackle megadalton sized oligomeric complexes[103], organelles such as the ribosome[104], and even intact viruses[105]. Alongside the expansion of the scope of possible targets to which the new MS approaches could be applied, significant progress was made in establishing that proteins can maintain a global structure in the gas phase which broadly resembles their solution-phase structure[106], providing important validation of non-denaturing MS as a method which could yield biological insight. For example, in contrast to the spectra of denatured proteins, which exhibit a very broad charge state distribution as a large number of residues are exposed to the solvent, non-denaturing MS spectra consist of narrow charge state distributions consistent with a folded protein's smaller solvent-exposed surface area. Stoichiometries measured by MS are also consistent with those measured by solution-phase techniques, as are other quantitative measurements such as dissociation constants[107]. It has even been shown that viral particles transmitted through a mass spectrometer and then collected on a metal plate retain their ability to infect cells, in addition to their overall rod-like structure when

examined by electron microscopy[108]. Finally, a compelling body of evidence for physiologically relevant gas-phase protein structure has emerged from experiments using ion-mobility spectrometry (IMS) coupled to mass spectrometry. In IMS, the time taken for a protein molecule to travel through a chamber filled with neutral gas molecules is measured. This time turns out to be dependent on the protein's collision cross-section (CCS) of the molecule, a rotationally-averaged cross-sectional area. Collision cross-sections can also be calculated from X-ray crystal structures, and compared to those measured by MS. The current evidence suggests that the gas phase CCS measurements are similar to those predicted from crystal structures[109]. Taken together, these studies constitute a compelling and growing body of evidence that protein complexes retain their global structure in the gas phase, on the timescale of the MS experiment.

With soluble protein complexes of varying sizes established as tractable targets for study by non-denaturing MS, attention naturally turned to membrane protein complexes given that they were, and remain, difficult to study by other methods. Moreover, in contrast to what was originally predicted for soluble proteins, it was thought that membrane proteins would likely maintain their fold in the gas phase as the non-polar environment of the vacuum would mimic that of biological membranes[110].

A significant barrier to obtaining interpretable mass spectra of membrane protein complexes was that the standard method of purification involved solubilising them in detergent micelles. Early analyses of membrane proteins produced spectra dominated by a broad 'hump', a consequence of the heterogeneity in the number of detergent molecules. Attempts to remove as much detergent as possible from the solution were unsuccessful: either it was impossible to remove enough detergent, or the proteins became unstable once the detergent concentration became too low. A breakthrough came in 2008 however, when Barrera et al successfully recorded the mass spectrum of the $BtuC_2D_2$ heterotetramer[111]. By keeping the complex in detergent in solution, but removing the detergent micelle via

collisions with gas molecules inside the mass spectrometer, they were able to detect the 'naked' intact tetramer.

In the 10 years following this study, our lab has analysed a significant number of membrane protein using non-denaturing MS, and, through study of the detergent release mechanism, optimised solution conditions and instrumental parameters[112]. Our standard protocol for MS analysis of membrane proteins has been reported in detail elsewhere[113], but will be outlined here in order to facilitate explanation of the modifications made to this protocol later in this chapter.

2.2 The non-denaturing MS experiment

A typical membrane protein sample is prepared for analysis by non-denaturing MS by solubilising it in a buffer consisting of a volatile salt (usually ammonium acetate) and an appropriate detergent (at a concentration greater than the critical micelle concentration (CMC)). Under these conditions, the membrane protein complex will be encased in a detergent micelle, a belt of detergent molecules which surround the hydrophobic transmembrane region of the protein, separating it from the solvent. The micelle will likely also contain lipid molecules which co-purified with the protein. Removing as much non-volatile salt and other small molecules from the solution is important as these molecules will bind to the complex in the late stages of the electrospray process, resulting in peak broadening in the final spectrum. These contaminants are usually removed by buffer-exchanging the sample into a buffer consisting of a volatile salt such as ammonium acetate, which evaporates easily during desolvation[114].

Choosing the right detergent is also an important consideration, since different detergents impart different stabilities to proteins in solution, ionise with differing efficiencies, and will have different affinities for the rest of the complex in the gas phase. An ideal detergent is one which allows the membrane protein to maintain its fold and oligomeric state in solution but which can also be dissociated with relative facility in the gas phase, thereby giving clean, well-resolved mass spectra. For example, the detergent dodecyl maltoside (DDM), a favourite of structural

biologists since it imparts stability to a wide range of membrane proteins, can sometimes be unsuitable for non-denaturing MS experiments. This is because the high level of gas-phase activation required to dissociate it in the collision cell cause disruption of the quaternary structure of the protein[113].

Choosing the most effective detergent remains largely a matter of screening a range of detergents known to be 'MS-friendly', although some progress has been made in understanding the effect of different properties of detergent on the final mass spectrum.[115]

Once the sample is prepared, it is ready to be introduced into the mass spectrometer. An overview of the quadrupole-time-of-flight (Q-ToF) platform routinely used by our lab for non-denaturing MS of membrane proteins is shown in figure 2.1 - this set-up has been the standard for analysis of large biomolecular complexes for the last 15 years[116]. The sample is transferred to a gold-coated glass capillary positioned close to the source of the mass spectrometer.

The experiment begins with electrospray ionisation of the solution. Ionisation is initiated by applying a potential difference between the capillary and the sample inlet, which draws the liquid through the narrow capillary tip. Accumulation of charge in the nascent drop causes it to elongate until it eventually deforms into a characteristic shape known as a Taylor cone. At this point the drop disperses into a fine spray of charged droplets, which are emitted from the capillary and pulled towards the sample inlet down the pressure and potential difference gradients. The capillary is often oriented perpendicular to the sample inlet, in a so-called 'Z-spray' arrangement, so that ionised species enter the sample cone while the majority of neutral molecules fly past, though some neutral molecules still enter the instrument as a consequence of the pressure gradient. Before the droplets enter the instrument, they undergo further solvent evaporation until their Coulombic repulsion exceeds their surface tension. Once this threshold, referred to as the Rayleigh limit, is reached, droplet fission occurs producing several smaller 'daughter' droplets. These smaller droplets then undergo further cycles of evaporation and fission until all

of the solvent has evaporated, leaving the charge which was on the surface of the droplet behind, spread across the desolvated protein-detergent-lipid complex[117].

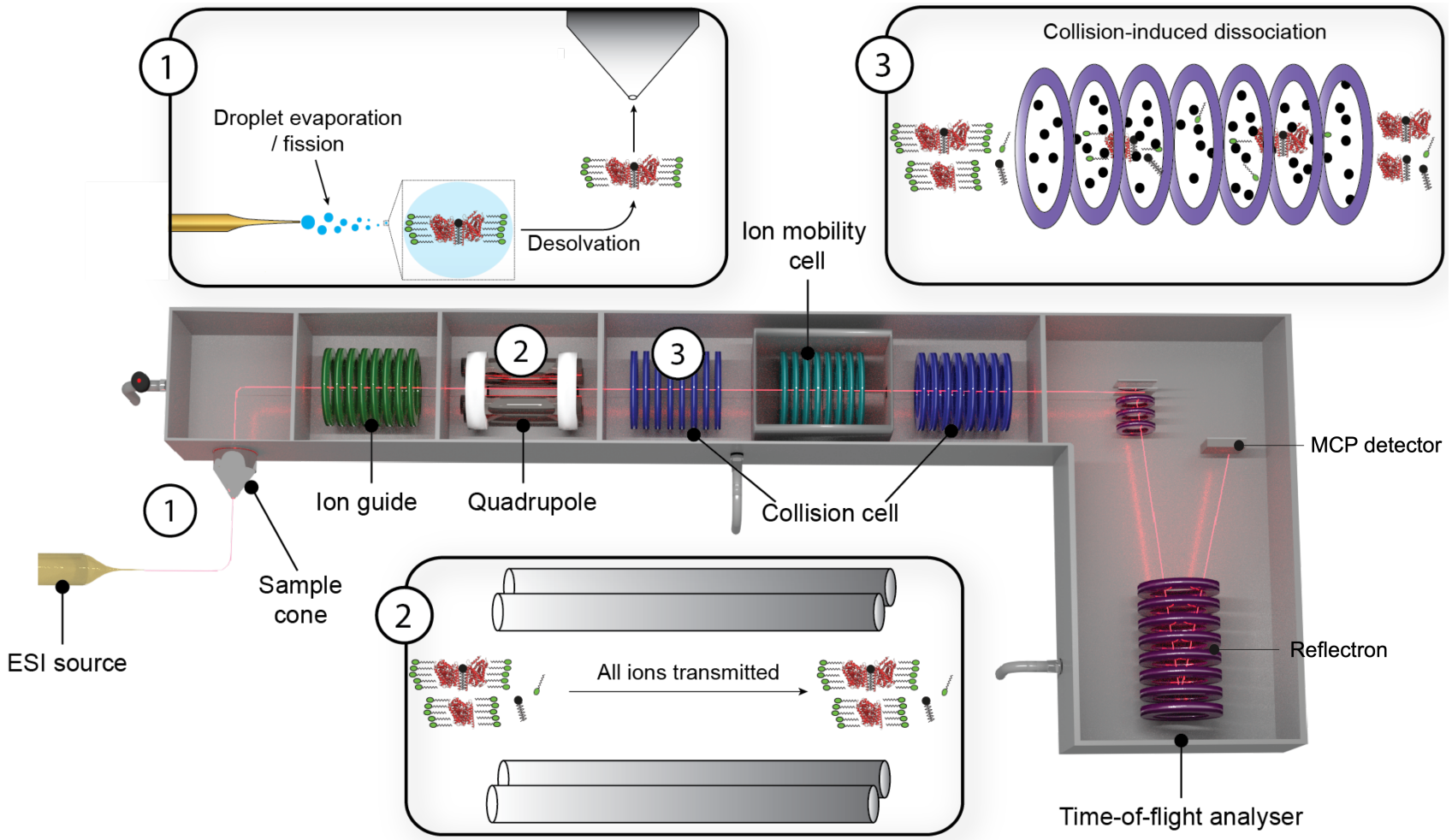


Figure 2.1: Schematic of the Q-ToF platform used for non-denaturing MS characterisation of membrane protein complexes. Membrane protein-lipid complexes are ionised at the nano-ESI source, and desolvated as they enter the instrument through the sample inlet (1). After passing through a series of ion guides, all of the ions travel through the quadrupole (2), which is used as a broadband mass filter. The ions then enter the collision cell (3), where the protein-lipid complexes are liberated from their detergent micelles via collisions with neutral gas molecules. The 'naked' complexes then enter the ToF analyser, where their mass:charge ratio is measured and the mass spectrum outputted.

Desolvation of the analyte is vital for good spectral quality, and has been vastly improved by the adoption of very narrow capillaries (with orifice diameters of 1 - 5 μm), an innovation commonly referred to as 'nano-electrospray ionisation'. Nano-ESI significantly reduces sample consumption, allows desolvation of analytes from non-volatile solvents such as water, and removes the need for other desolvating sample conditions such as high temperature, low pH or organic solvent, all of which lead to denaturation of proteins.

Once inside the instrument, the ions enter the vacuum and traverse a series of different regions between the source and the detector, moving down a pressure and potential gradient. At all stages, the voltages and pressures applied in the different regions are controlled to maintain a careful balance between outcomes such as ion transmission and signal to noise, which determine spectral quality, and keeping the overall structure of the protein complex as close to 'native' as possible.

In the source region, where the pressure is typically maintained at 3 - 9 μbar, the protein-detergent-lipid complex undergoes multiple collisions with neutral gas molecules. This collisional cooling process thermalises the complex, reducing its kinetic energy. This reduction in kinetic energy allows the ion beam to be focused more efficiently, improving transmission of large ions. Collisional cooling is essential for transmission of high molecular-weight analytes, and is one of the key instrumental developments which has enabled analysis of large non-covalent complexes[116].

Further focusing of the ion beam occur as it passes through several ion guides before entering the quadrupole. The quadrupole consists of a series of four parallel cylindrical or hyperbolic rods, connected in two pairs, and arranged around a central axis. A potential with constant (DC) and alternating (RF) components is then applied to the rods, such that there is a positively charged pair and a negatively charged pair of rods which swap polarities at a specified frequency. Whether a particular ion can maintain a stable trajectory within the quadrupole depends on its mass/charge ratio (m/z) and on the particular RF and DC voltages chosen. It is therefore possible to set the RF and DC voltages such that only ions within a desired m/z range are transmitted and hence the quadrupole can be used as a

mass filter. This mass filter has many uses in non-denaturing MS experiments, but cannot be usefully employed in using this set-up for membrane proteins, since at this stage in the instrument the protein complex of interest remains encapsulated in the detergent micelle. In a standard membrane protein MS experiment the quadrupole is operated in RF-only mode, so it acts as a broadband mass filter, the result being that further focusing of the ions occurs but no mass selection. Mass selection of membrane proteins is possible, however, using the high energy source modification discussed later in this chapter.

Upon exiting the quadrupole, the protein complex ions are then accelerated into the collision cell, where they collide with neutral gas molecules (typically Ar or SF_6. These collisions dissociate the detergent molecules, leaving the naked protein-lipid complex to enter the ToF mass analyser. This process is known as collision-induced dissociation (CID), and has many useful applications in non-denaturing MS[118]. For example, CID can permit elucidation of composition and connectivity of complicated hetero-oligomeric or polydisperse protein complexes[119] by careful dissociation of full complexes into smaller subcomplexes with spectra which are easier to analyse. In addition, when coupled with ion mobility spectrometry, CID can be used to quantify the stability of proteins against dissociation or unfolding[120;121].

Careful control of CID is crucial in any membrane protein MS experiment, as the amount of activation applied must be set high enough to fully dissociate the detergent micelle, but not so high as to disrupt the structure of the protein complex or fragment the individual polypeptide chains. The amount of activation applied to the ions is normally tuned by adjusting the collision voltage, the potential difference to which the ions are exposed as they enter the collision cell, but the extent to which this activation dissociates the detergent micelle and dissociates or unfolds the protein complex also depends on many other factors such as the gas pressure inside the collision cell, which collision gas is used, the size and shape of the protein complex, the charge state, whether any lipids or other ligands are bound, and which detergent is used. Hence, the ideal parameters for liberating a protein complex

from its detergent micelle are highly protein- and detergent-dependent, and, to some extent, must be determined empirically.

After the collision cell, modern Q-ToF instruments may contain an ion-mobility cell, in which the collision cross section of the protein complex can be measured using ion-mobility mass spectrometry (IMMS). IMMS was not used in this investigation, so it will not be described in detail here.

In the final stage of the membrane protein MS experiment the ions enter the time-of-flight analyser. Here, at the pusher region, an orthogonal positive potential is applied, causing a discrete packet of ions to accelerate with a fixed kinetic energy along the full length of the flight tube. At the reflectron the ions are reversed back towards a micro-channel plate (MCP) detector, overall following a V-shaped trajectory. Inclusion of the reflectron effectively doubles the length of the flight path taken by the ions, thereby improving the resolution. Finally, at the detector the electron pulse of each ion is detected and the total flight time through the ToF analyser recorded. This can be related to the mass to charge ratio as follows:

$$ezV = \frac{1}{2}mv^2 \tag{2.1}$$

where e is the charge on the electron (1.602×10^{-19} C), z is the charge of the ion, V is the potential energy 'push' applied to the ions, m is the mass of the ion, and v its velocity. Assuming a linear flight tube, the velocity of the ions can be replaced by the ratio of the length of the flight path (L) to the flight time (t):

$$v = \frac{L}{t} \tag{2.2}$$

$$ezV = \frac{mL^2}{2t^2} \tag{2.3}$$

$$\frac{m}{z} = \frac{2eVt^2}{L^2} \tag{2.4}$$

From which it can be seen that the mass-to-charge ratio m/z is proportional to the square of the flight time. Since all of the other parameters are known for a given experiment, the instrument calculates the m/z for each ion based on the flight time and generates the mass spectrum, a plot of m/z against relative ion abundance.

2.3 Applications to membrane proteins

The Robinson lab has studied many different membrane proteins, purified from a variety of different organisms, using the non-denaturing MS platform described above[112]. Whilst these studies have explored various aspects of membrane protein structure and interactions, for the purposes of this study we shall outline 4 key measurements which MS has been shown to be capable of in recent years which are particularly relevant to the study of membrane protein oligomerisation and which form the foundations of the work reported in the rest of this thesis.

2.3.1 Determining oligomeric states

The most straightforward property of non-covalent membrane protein complexes which can be measured by MS is the oligomeric state. For a homomeric complex, made up of copies of only one protein subunit, if the amino acid sequence of the monomeric form of a given protein is known, the oligomeric state can often simply be inferred from the MS measurement of the mass of the intact complex. MS measurements of oligomeric states of homomeric complexes have been carried out for proteins of a wide range of oligomer sizes and molecular weights from dimers to heptamers[122;123;124;125]. Assigning subunit stoichiometry is more challenging for heteromeric complexes, but can be achieved by analysis of subcomplexes which are often also present in the spectra. Examples of cases in which large non-covalent heteromeric complexes were preserved in the gas phase and their oligomeric states measured include a pioneering study by Zhou *et al.* of the 26-subunit V-type ATPase from *Thermus thermophilus*[126], and a study by Housden *et al.* on the *E. coli* colicin E9 translocon, a complex of two membrane proteins and soluble protein located on both sides of the membrane[127].

Three key advantages of MS for determining oligomeric states, relative to other biophysical methods, are its speed (a MS measurement can be recorded in a few minutes once the sample is prepared), high mass accuracy, and the fact that the protein complex mass can be determined in the absence of detergent in the gas phase. Methods which have traditionally been used for determining the molecular

weight and therefore oligomeric state such as analytical ultracentrifugation, size-exclusion chromatography and native gel electrophoresis and multi-angle light scattering all suffer from inherent problems due to the presence of detergent, which changes the molecular weight, size, sedimentation coefficient and light scattering of the complex in solution.

Similarly, as discussed in chapter 1, the dominant technique of structural biology, X-ray crystallography also performs poorly when determining oligomeric states. This is in part because the artificial conditions of the crystal packed environment can lead to both false positives and false negatives in oligomeric state determination. Perhaps more importantly, there is no direct relationship between the oligomeric macromolecular complex and the asymmetric unit (ASU) of the crystal - a complex may span several ASUs or be contained within a single ASU, or several complexes can exist within a single ASU. These possibilities mean that even if a biological oligomer has formed in the crystal, distinguishing this biological oligomer from artifactual crystal packing oligomers is non-trivial, a fact which has lead to the development of computational algorithms to solve this problem[41].

2.3.2 Identifying endogenous lipids

A common feature of mass spectra of non-denaturing MS experiments on purified membrane proteins is the presence of co-purified lipids bound to the protein complex. It is assumed that these endogenous lipids represent those which are directly in contact with membrane proteins in the lipid bilayer of the host organism, and therefore MS experiments which can detect and characterise endogenous lipids have the potential to provide insight into the local lipid environment around a membrane protein. This is of interest to membrane protein biologists, as it remains unclear in many cases the extent to which the lipids immediately adjacent to membrane proteins (sometimes called the lipid annulus) differ from those lipids in the bulk membrane. Furthermore, by the time that a non-covalent interaction between a protein and an endogenous lipid has been detected in a mass spectrum, the interaction has persisted through the protein purification steps required to

extract the protein from the host organism, and also the various manipulations in the vacuum of the mass spectrometer. This suggests that the small number of endogenous lipids which are detected in mass spectra are those which form strong protein-lipid interactions, and hence could be acting as 'structural' lipids.

Endogenous lipids were first reported in membrane protein mass spectra in a 2009 study by Barrera *et al.*[122], in which the authors observed up to two molecules of phosphatidylethanolamine binding to the homodimeric ABC transporter MacB and cardiolipin to the heterodimer LmrCD. These lipids were assumed to fall into the category of structural lipids, residing in small well-defined binding sites, since they were relatively small in number and had persisted attached to the protein even after incubation in high concentrations of delipidating detergent.

In contrast, in later studies on multi-subunit complexes in which large cavities were present[126;128], larger cohorts of lipids were observed forming lipid 'plugs'. In the V-type ATPase from *Enterococcus hirae*, for example, a single cardiolipin molecule was found associated with each of the 10 subunits of the membrane-embedded 'K-ring' which forms part of the main rotor. Using the stoichiometry of the lipid binding and electron densities from an X-ray crystal structure of the rotor, the authors were able to 'dock' the 10 cardiolipins such that they formed a hydrophobic lining inside the K-ring, and significantly reduced the diameter of the cavity formed in the centre of the ring.

Identifying and characterising lipids which form an annulus around a protein is more challenging than looking at structural lipids, since this involves a larger number of lipid molecules which are likely forming weaker interactions with the membrane protein than those in well-defined binding sites. However, two recent studies from the Robinson group have shown the potential for MS techniques to investigate the lipid annulus. A challenge in obtaining well-resolved spectra of membrane proteins bound to large numbers of lipids is that the more lipids are bound to a protein, the more heterogeneity and therefore the more broadening of the spectrum occurs.

Bechara *et al.* managed to address this problem by careful partial delipidation of the dimeric *E. coli*-expressed ABC transporter *TmrAB* using prolonged incubation

with detergent. By monitoring the extent of the lipidation of the protein by analysing it using non-denaturing MS, and conventional lipid extraction followed by LCMS, the authors could investigate the different 'layers' of lipids around the protein. Interestingly, when comparing a partially delipidated sample of *TmrAB* (13 bound lipids per dimer, after 24 hours of incubation) to one which had been almost fully delipidated (5 lipids per dimer, after 48 hours of incubation), they found a striking difference in the class of the bound lipids. Whilst the partially delipidated sample contained 8 PE and 5 PG molecules per dimer, the more fully delipidated form consisted of 5 PG molecules only, suggesting that the anionic PG lipids form stronger interactions with *TmrAB*, and perhaps therefore form the inner layer of lipids around the dimer.

An alternative approach was adopted in a 2016 study by Marty *et al*[129], in which the authors analysed membrane proteins in nano-discs, self-assembled lipid bilayers surrounded by two amphipathic membrane scaffold proteins. Nanodiscs allow membrane proteins to be encapsulated in an environment which mimics the natural membrane but with a pre-defined lipid composition. By using highly activating conditions inside the mass spectrometer, the authors were able to release two different proteins (the tetrameric aquaporin AqpZ and trimeric ammonia transporter AmtB, both from *E. coli*) from nanodiscs in a manner which retained the oligomeric structure of the protein complexes as well as up to 120 bound lipids. By comparing the distributions of lipids observed in the mass spectra with estimates for the numbers of lipids present in the different local environments around the protein in MD simulations of AqpZ and AmtB in a lipid bilayer, it was possible to infer which layers of lipids were being detected in the MS experiments. When conditions were more mild i.e. just activating enough to release the proteins from the nanodiscs, a lipid distribution was observed that was consistent with the presence of a 'head group shell', plus a more weakly bound annular belt. When the collision energy is raised, the more weakly-bound belt is lost, and only the lipids corresponding to the head group shell are retained. This example illustrates the potential of MS to distinguish between lipids in different shells around a membrane

protein, and gain insight into the interactions involved between the protein and lipids in different parts of the protein's local environment.

2.3.3 Investigating the effects of protein-lipid interactions: exogenous lipids

The studies described in the previous section have involved experiments investigating endogenous or co-purified lipids which have allowed researchers to learn about the lipids which are directly associated with the protein in the membrane. An alternative approach to gain insight into protein-lipid interactions and their impact on protein structure and function in a more systematic and controlled way is to prepare protein samples under delipidating conditions and then subsequently add specific exogenous lipids. This approach is less 'physiological', in that the lipids involved in these experiments are chosen according to the judgement of the investigator rather than which lipids survive the purification process. However, it does simplify the analysis of protein interactions, since properties such as lipid class, aliphatic chain length, protein:lipid ratio can be controlled in the experiment.

For example, in a key study reported in a 2014 paper, Laganowsky *et al.* quantified the effect of exogenous lipids on the stability of 3 oligomeric membrane proteins[130]. The authors used ion mobility mass spectrometry to measure the CCS of the membrane proteins as they were subjected to increasing collisional activation. As the CCS represents a rotationally averaged 'size' of a protein complex, based on how long the complex takes to traverse a gas-filled chamber, it is a useful measure of protein compactness or 'foldedness'. By measuring the CCS changes across the collision energy range of the instrument the authors could observe different unfolding transitions and calculate a measure of stability relative to unfolding for each complex.

It was found that different lipids imparted differing levels of resistance to unfolding to the different complexes. For example, phosphatidylglycerol was shown to provide significant stabilisation to the ammonia transporter AmtB (to a much greater extent than lipids with 5 other head-group chemistries). The authors then used this information to carry out crystallography experiments on AmtB, in which

they solved the crystal structure of AmtB in the presence of PG. Comparing this structure with a previously published structure solved in the absence of lipids, a distinct conformational change could be seen, in which a 12-residue loop moves to form a specific PG binding site. This illustrates the importance of protein-lipid interactions for membrane protein tertiary structure, and shows the potential of MS-based characterisation of these interactions to inform other biophysical experiments.

2.3.4 Characterising dynamic processes

As discussed in the previous chapter, whilst some membrane proteins form tightly bound oligomers, many exist in a dynamic equilibrium between different oligomeric states. Hence, a complete characterisation of membrane protein oligomerisation must involve a consideration of kinetics as well as thermodynamics. Mass spectrometry is ideally placed to carry out measurements of quaternary structure dynamics for several reasons[131]. It can capture transitions over a wide range of timescales, permits rapid analysis with high sensitivity, and is also a non-ensemble method, meaning it can quantify the abundance of co-existing states.

Whilst before this work was carried out there were no reports in the literature of real-time monitoring of membrane protein oligomerisation by MS, such experiments are now routine for soluble proteins. The most common method by which this is achieved is via so-called subunit exchange experiments. Generally, these involve mixing the wild-type form of an oligomeric protein with an isotopically labelled (for example with ^{13}C or ^{15}N) form with a higher mass. After mixing, the composition of the reaction is then monitored by MS. If the oligomers in the solution are in a dynamic equilibrium one or more hetero-oligomer populations (consisting of oligomers with both 'light' and 'heavy' subunits) will emerge over time. Rate constants can be calculated for the subunit exchange, allowing comparison of the rates of oligomerisation of different proteins, and the effect of mutations or other experimental conditions. This approach has been employed successfully to study the dynamics of protein complex assembly in systems such as small heat shock

proteins[132], and the tetrameric protein transthyretin which is strongly associated with systemic amyloid disease[133].

2.4 Expanding the mass spectrometry toolkit

As we described in chapter 1, one of the key aims of our study is to be able to identify the components of endogenous lipid-protein complexes in order to gain insight into the local lipid environment around a membrane protein. However, whilst the standard MS platform described above is well suited to determining membrane protein oligomeric states and in characterising exogenous lipid binding, it has limitations when it comes to the study of binding of endogenous lipids.

Let us consider the example of a hypothetical oligomeric membrane protein which, when examined by MS, is found to bind a series of small molecules which are suspected to be phospholipids. Lipid binding generally manifests itself in mass spectra as a series of satellite peaks adjacent to a more intense *apo*-protein peak for each charge state. In the case where the oligomer binds only equivalents of a single lipid, therefore, it is generally possible to tentatively identify this molecule as a phospholipid by calculating the mass difference between the lipid-bound peak and the *apo* peak. However, in many cases, the scenario is not so simple. The protein may bind lipids of different classes which have overlapping masses, a common example being the case of cardiolipin which has a mass similar to twice that of a typical diacylphospholipid. Alternatively, if the protein is highly lipidated, there may not be a prominent *apo*-protein charge state distribution from which a mass can be extracted to compare with the lipid-bound species. Furthermore, since the mass resolution of the measurement is generally lower at high m/z it is very difficult to obtain a highly accurate lipid mass in order to fully identify the lipid.

One solution to this problem might be simply to significantly increase the amount of activation applied in the collision cell of the instrument in order to dissociate any bound lipids from the protein oligomer. This approach would allow a lipid-free protein mass to be measured, against which the lipid-bound protein mass could be compared, and would also allow detection of free lipids in the low m/z region

of the mass spectrum. There are two problems with this strategy, however. The first is that there is often not enough activating power available in the collision cell to fully dissociate membrane protein-lipid complexes. This is partly due to the fact that some of the applied energy must be used up in liberating the protein-lipid complex from the detergent micelle, and hence in some cases little activating potential is left to disrupt protein-lipid interactions. Secondly, using this approach it is often not possible to unambiguously discern the origin of dissociation products detected in the low m/z region. Since the hypothetical dimer will have come from a solution which also contains free detergents and lipids, detergent-lipid micelles, monomers in micelles, it will be impossible to determine which lipids detected under highly activating conditions have come from any of these other sources, rather than specifically from dissociation of the oligomer.

To get around this problem, we adopted an approach known as tandem MS or MS/MS which has long been routinely applied to soluble protein-ligand complexes. Tandem MS has been particularly powerful when applied to heterogenous or polydisperse systems containing oligomers of different sizes and in determining the composition and topology of large hetero-oligomeric complexes with multiple subunits[119;134]. In the tandem MS experiment, a protein complex of interest is first identified in an initial MS scan. Next, the voltages applied at the quadrupole are tuned so that only the region of m/z space corresponding to a single charge state of the complex are transmitted to the next stage of the instrument. The complex can then be dissociated in the collision cell, and the masses of the dissociation products measured with the knowledge that all species detected have come from the isolated complex.

Applying this approach to membrane protein complexes is complicated by the fact that mass selection at the quadrupole is not possible due to the presence of the detergent micelle. Since the micelle is not dissociated from the protein complex until the collision cell, the complex passes through the quadrupole as a heterogeneous species bound to varying numbers of detergent molecules, precluding selective transmission of a narrow m/z window. To carry out a membrane protein

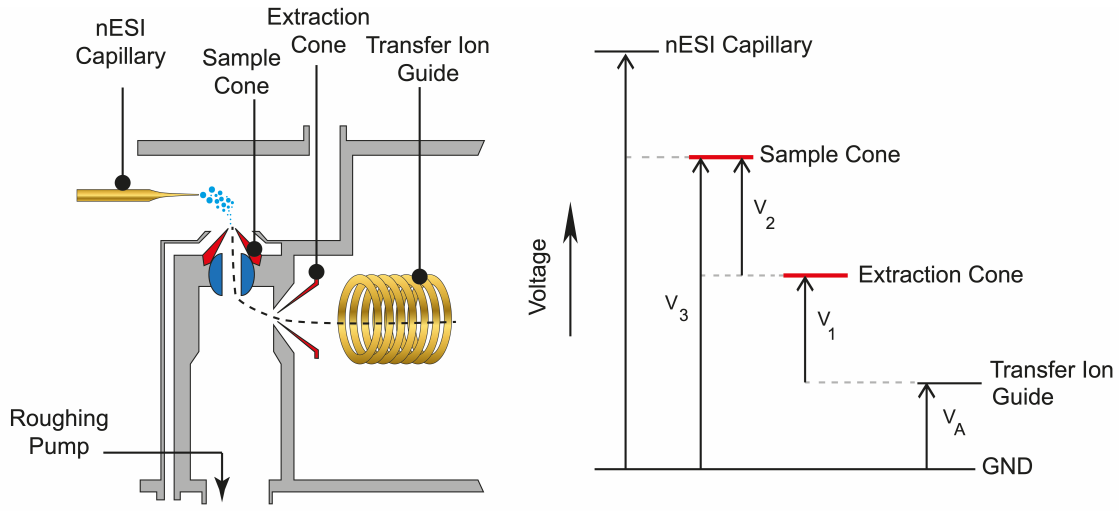


Figure 2.2: Schematic of the source region of a Q-ToF mass spectrometer, modified in order to apply a greater amount of in-source activation to membrane protein complex ions, and its associated voltage diagram. Additional activation relative to the standard platform is achieved by an extension of the range of the extraction voltage V_3, the potential difference between the extraction cone and the transfer ion guide, to 200 V. This increased activation allows liberation of membrane protein complexes from detergent micelles in the source region of the instrument. Figure adapted from Gupta *et al*[135].

tandem MS experiment, therefore, we used a version of the standard platform which had been recently modified to significantly increase the amount of activation which could be applied in the source region of the mass spectrometer, a schematic of which is shown in figure 2.2. This modification was achieved by introducing the capacity to drive the sample cone voltage from an external supply, allowing the sample cone voltage to be increased to up to 200 V.

This increase in activating power at the source meant that it was now possible to completely dissociate protein-lipid complexes from the detergent micelle, select individual charge states of the complex in the quadrupole and then dissociate them using the full collision energy range available in the collision cell. This approach, the key stages of which are illustrated in figure 2.3 has been termed high-energy non-denaturing mass spectrometry (HE-nMS).

An example of the HE-nMS workflow in action is shown in figure 2.4. In this proof-of-concept experiment, the monomeric multidrug exporter MATE from *Pyrococcus furiosus* was mixed with phosphatidylglycerol (PG), and the mixture analysed by MS. Initially, the quadrupole voltages are set so that all species are

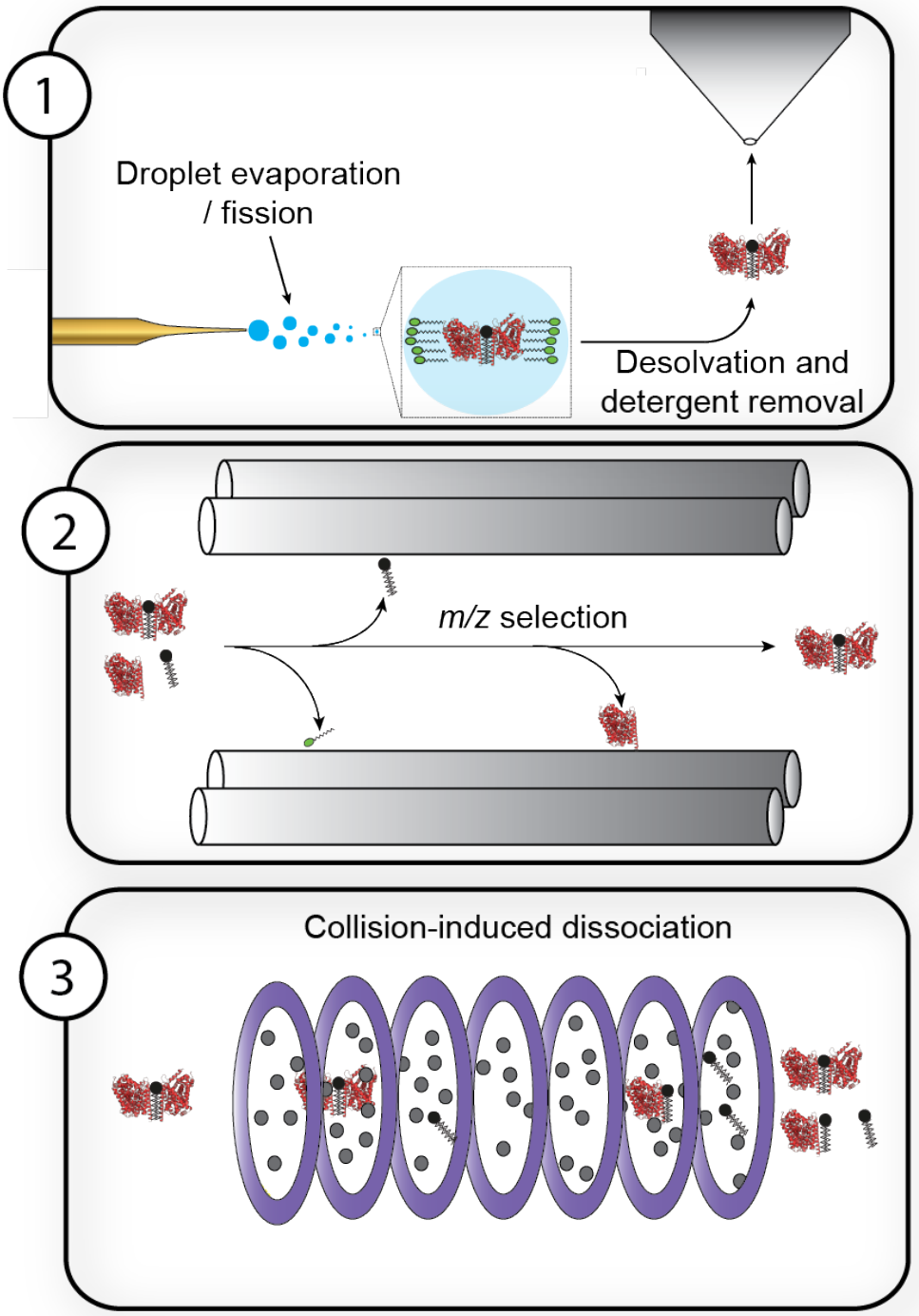


Figure 2.3: Snapshots of key stages of the HE-nMS experiment. At the source (1), the increased level of activation relative to the standard experiment causes the protein complex to be desolvated and released from the detergent micelle. A specific oligomeric state and lipidation state of the demicellated protein complex (in this schematic a lipid-bound dimer) can then be selectively transmitted through the quadrupole (2), while free lipids, detergents and other oligomeric states are not transmitted. In the collision cell (3), the energy from collisions with neutral gas molecules is used to disrupt protein-protein and protein-lipid interactions within the protein-lipid complex, in the knowledge that any dissociation products detected must have come from the dimer.

transmitted, and the collision voltages are set to 0. The cone voltage is then increased in order to dissociate the protein-lipid complex from the detergent micelle. This adjustment is achieved by monitoring the mass spectrum, and increasing the cone voltage until the spectrum changes from a broad hump to a set of well-resolved charge state distributions representing MATE bound to up to 4 PG molecules, shown 2.4a. The spectrum at this point should be identical to that produced from dissociation of the micelle in the collision cell using the standard MS platform described earlier. A single charge state of the species of interest is then chosen from the mass spectrum, in this case the 10+ charge state of MATE bound to a single PG, and the voltages on the quadrupole set so that only a m/z window corresponding to this charge state is transmitted, as shown in figure 2.4b. Finally, the collision cell voltage can be increased in order to dissociate the isolated species. In this example, activation of the MATE-PG complex results in loss of either a neutral or a positively charged species of mass 771 Da, corresponding to PG, and leaving *apo*-MATE with 10 or 9 charges respectively.

2.5 High resolution MS

The high energy source method described above provides one potential approach to characterising protein-lipid complexes in the case where it is challenging to extract accurate masses of the different components from the spectrum of the lipid-bound protein oligomer. By isolating and then dissociating the protein-lipid complex, the individual components can be measured separately and hence the overall composition of the complex can be determined.

An alternative approach to detecting and elucidating the composition of protein-lipid complexes is to carry out mass measurements of these complexes at higher resolution. In mass spectrometry, resolution (strictly speaking, resolving power) is the ability of an instrument to yield distinct signals for two ions with a small difference in m/z, and hence a high resolution MS experiment is one in which species of very similar masses can be detected and quantified separately. In the case of a membrane protein binding to different combinations of endogenous lipids,

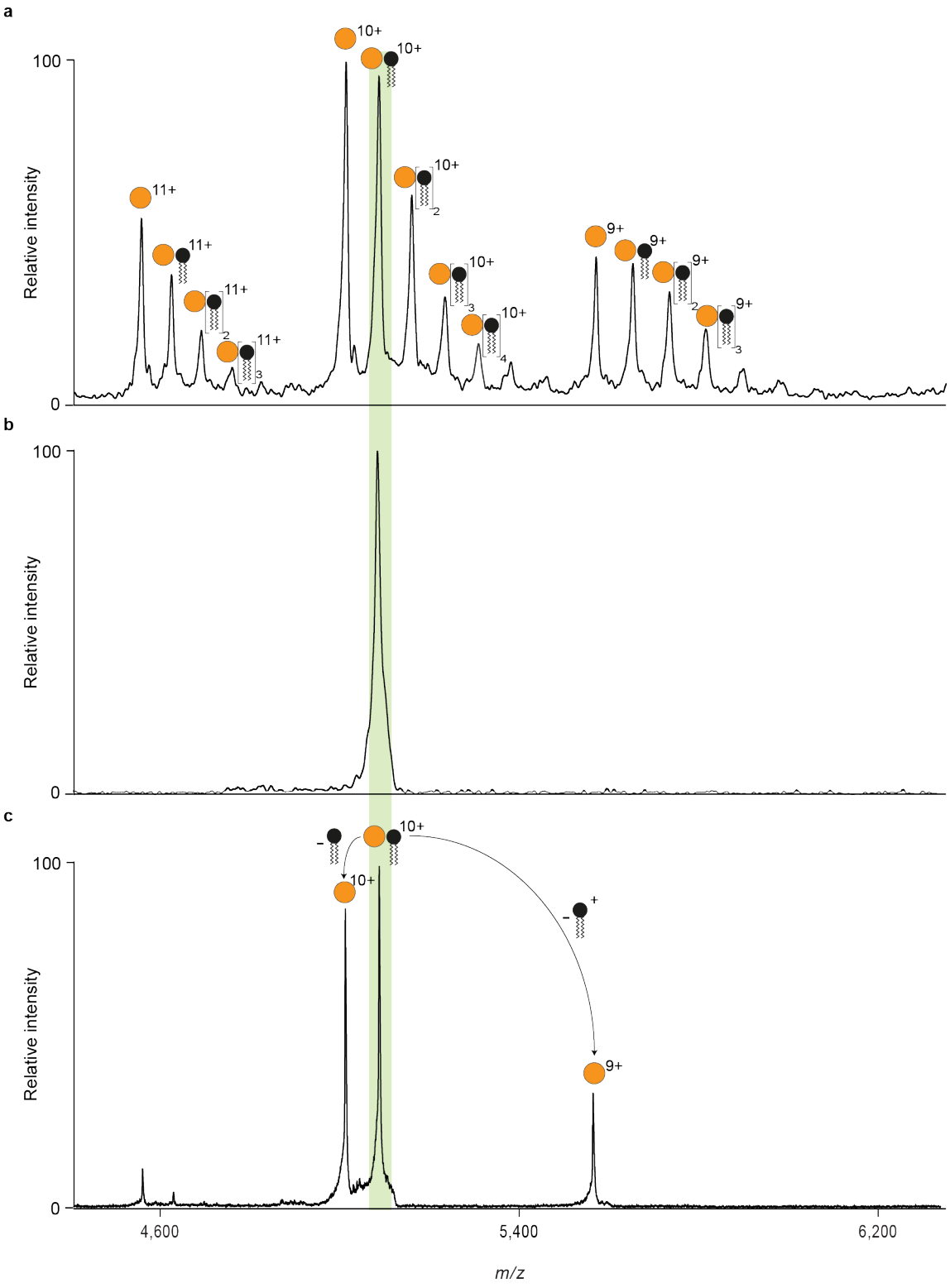


Figure 2.4: Selective dissociation of protein-lipid complexes of MATE and exogenous phosphatidylglycerol (PG) by HE-nMS. **a** Mass spectrum of MATE, supplemented with exogenous PG, released from DDM micelles using only in-source activation shows MATE monomers (●, 51.0 kDa) bound to up to 4 PG molecules (⚲, 771 Da). **b** Mass spectrum of MATE bound to a single PG molecule after mass selection of the 10+ charge state of this species at the quadrupole. **c** Dissociation of the 10+ MATE-PG complex results in loss of neutral or positively charged, to produce 10+ and 9+ *apo*-MATE respectively.

therefore, we reasoned that using high resolution MS would enable us to both separate lipid-bound species of similar mass, and also gain greater insight into other properties of the bound endogenous lipids such as acyl chain length.

2.6 The orbitrap mass analyser

For the high resolution MS experiments reported in later chapters, we used a mass spectrometer with an orbitrap mass analyser.

The orbitrap mass analyser consists of a central, spindle-shaped electrode and an outer electrode shaped like a barrel divided into two parts, see figure 2.5. If the instrument is operated in positive ion mode, a negative potential is applied to the central electrode and the positive analyte ions are injected into the orbitrap with kinetic energy such that they stably orbit the central electrode. A positive potential is applied to the outer electrodes and hence the orbiting analyte ions are also repelled back and forth along the longitudinal axis. The shapes of electrodes and the form of the potentials applied are chosen so that the axial motion of the ions is harmonic, and so the frequency of the axial oscillation is directly linked to the mass/charge ratio and independent of the kinetic energy with which the ions were injected[136;137]:

$$\omega = \sqrt{\frac{kz}{m}} \tag{2.5}$$

The frequency of the oscillations is detected by measuring the image current induced by the ions in the two linked outer electrodes. A Fourier transform is then applied to convert the image current measurement from the time domain to the frequency domain, from which the mass spectrum can be generated.

2.7 Resolution

When using a ToF analyser, the resolution of the mass measurement is dependent on the length of the flight path along which the ions travel, and is independent of the absolute m/z of the individual species. This means that, in theory, there is no upper limit on the resolving power of the instrument, since any two species

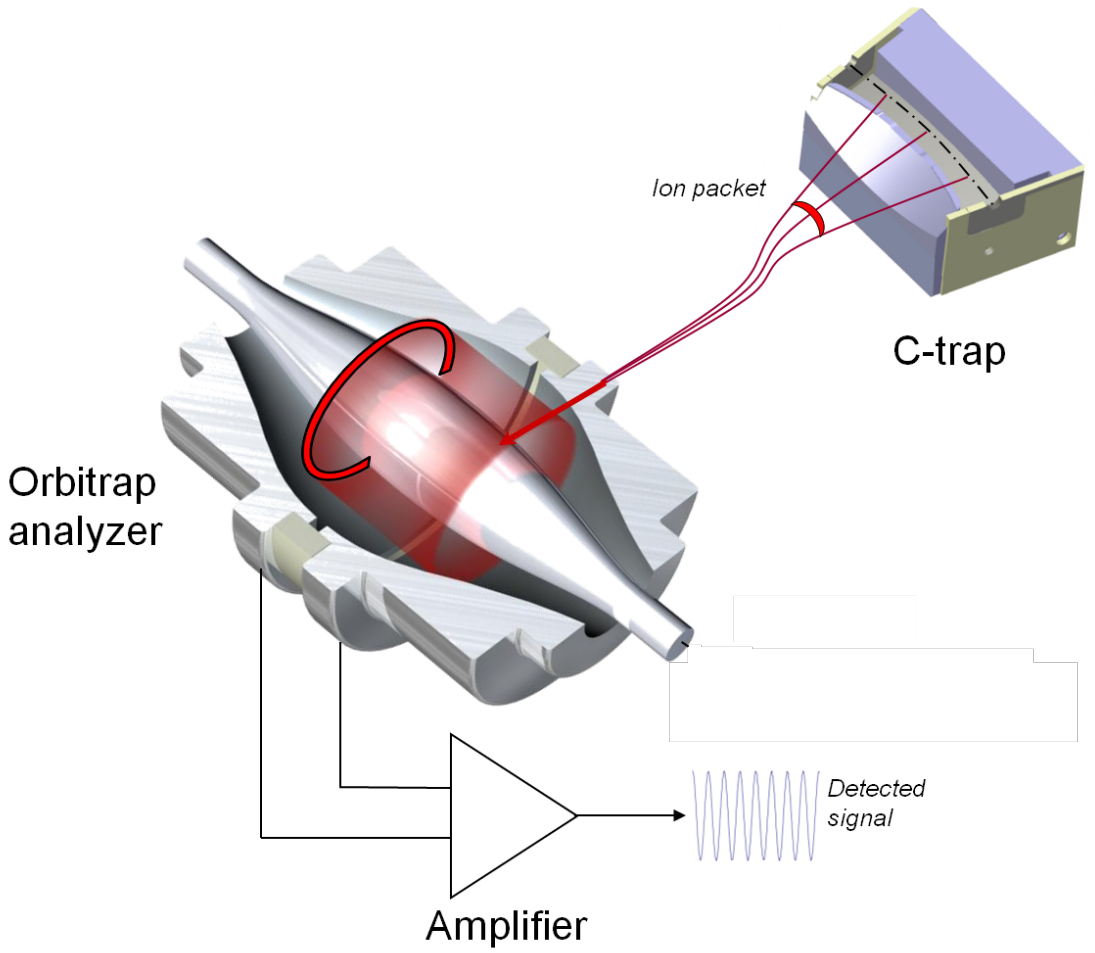


Figure 2.5: Schematic of the orbitrap analyser and the flight path of ions passing through it. Ions are injected from the C-trap in a compact packet into the orbitrap. They then orbit the central spindle electrode, oscillating axially. The axial oscillations induce an image current in the outer electrodes, the Fourier transform of which allows calculation of the mass/charge ratio and generation of the mass spectrum. Artwork by Thermo Fisher Scientific, made available under a CC-share-alike licence.

can be resolved in the mass spectrum given a long enough ToF flight tube. In practice, however, extremely large flight tubes are not practical in a laboratory context, and the benefits of increased resolution must be offset against decreased sensitivity as there will be greater loss of ions by scattering after collisions with gas molecules or by angular dispersion of the ion beam.

The resolving power of the orbitrap analyser depends on the detection time, the amount of time for which the image current is recorded for a particular packet of

ions, and the period of the oscillating ions, which is in turn dependent on m/z.

$$R \propto \frac{T_{det}}{m/z} \tag{2.6}$$

Just as the resolution of the ToF can in theory be increased indefinitely by using an increasingly long flight tube, the resolution of the orbitrap analyser can be increased indefinitely by extending the length of the detection time. Again, however, the more time the ions spend orbiting the central electrode, the more likely they are to collide with other gas molecules and be lost. This manifests itself as a reduction in signal to noise at higher transient times. Despite this upper bound on the range of transient times that can be used, in practice the orbitrap analyser can achieve much higher resolution than the ToF analyser.

2.8 Adapting the orbitrap to native MS

The orbitrap was first developed in the 1990s by Alexander Makarov[136], and was first incorporated into a commercial instrument, the LTQ Orbitrap, in 2005[137]. Orbitrap instruments quickly gained popularity in the proteomics community as they allowed high resolution analysis without the huge cost and space requirements of FTICR instruments. A number of key extensions to the family of orbitrap instruments, including the incorporation of a collision (HCD) cell, development of a stand-alone orbitrap instrument (the Exactive), and of a quadrupole-orbitrap hybrid (the Q-Exactive), occurred in the following decade, expanding the versatility of the orbitrap analyser. Following these developments, a question naturally arose as to whether the orbitrap might be able to be used for high resolution analysis of large non-covalent complexes. Guided by the principles which had been used in the 1990s to modify ToF instruments for the analysis of high molecular weight, the lab of Albert Heck were able to extend the mass range of a modified Exactive Plus so that it could record high resolution spectra of protein complexes as large as 800 kDa[138]. In their 2012 paper, Rose *et al* showed that the orbitrap analyser permitted a resolution (full width half height) of 25,000 at m/z of 5,000 and 16,000 at m/z of 10,000, significantly better than that of a ToF instrument which achieved

a resolution of 5,000 across this m/z range. This resolution allowed the separation of different glycoforms of a 146 kDa antibody and detection and distinction of ATP and ADP binding to the 801 kDa GroEL complex, illustrating the potential of the orbitrap analyser to measure small mass changes on high-molecular-weight species indicative of small molecule binding.

With the potential of the orbitrap to analyse large soluble protein complexes established, the question naturally arose as to whether this approach would also be successful for membrane protein analysis. This was of particular interest to our lab, as previous studies on membrane protein interactions with lipids, drugs and other ligands had been restricted to those where the mass differences between proteins and substrates were greater than 200-300 Da[139]. Gault *et al* therefore modified a Q-Exactive instrument, the architecture of which is shown in figure 2.6, in order to be able to carry out high resolution mass measurements on intact membrane protein complexes[140].

The path taken by membrane protein complexes through the Q-Exactive is broadly analogous to their path through a Q-ToF instrument described above. Gas-phase ions are generated by an ESI source and then proceed down a potential and pressure gradient through a series of focusing lenses and the quadrupole. The ions then enter a linear ion trap, the C-trap, before being transferred to the HCD collision cell. In the HCD cell, collisional activation is applied to dissociate the detergent micelle, after which the ions return to the C-trap. After further collisional cooling and focusing in the C-trap, the ions are then finally accelerated into the orbitrap for mass analysis.

The key modifications required to maintain non-covalent interactions in the gas phase and optimal transmission of high mass ions consisted of raising the source pressure and establishing a gentle voltage gradient across the lenses in this region. Once these modifications had been carried out, Gault *et al.* found that the Q-Exactive was capable of transmitting and analysing membrane protein complexes with molecular weights up to at least 200 kDa, whilst maintaining non-covalent interactions between protein subunits. Furthermore, the modified

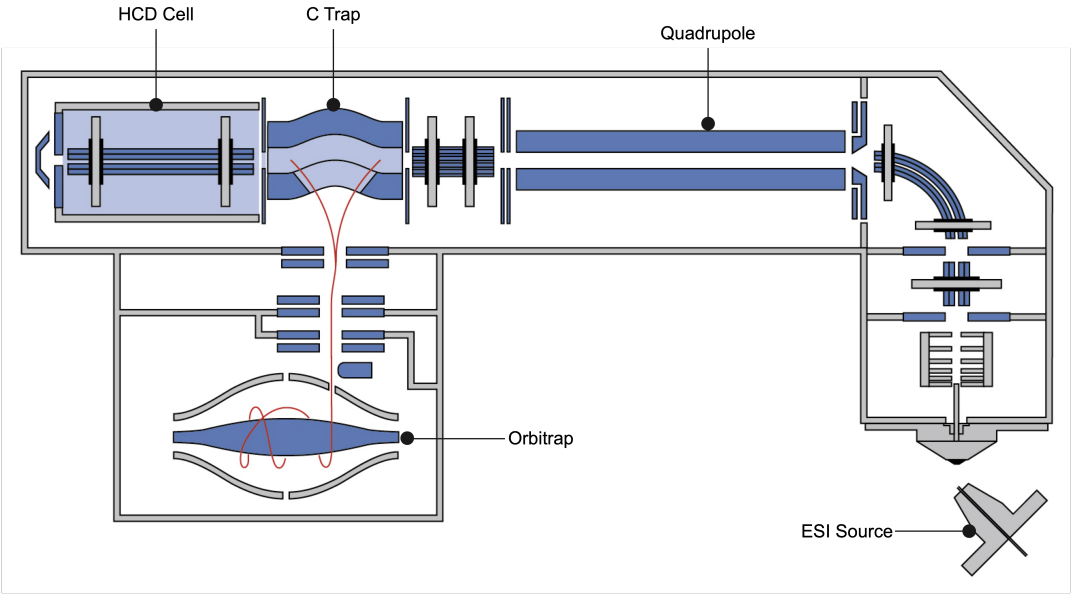


Figure 2.6: Schematic of the Q-Exactive instrument used for high resolution MS experiments in this study. Membrane protein-lipid complexes are ionised at the nano-ESI source, and desolvated as they enter the instrument. The ions are then focused through a series of lenses, before passing through the quadrupole. Next, the ions are held in the C-trap, before being transmitted into the HCD cell, where they undergo collisions with neutral gas molecules in order to liberate the membrane proteins from their detergent micelles. Finally the 'naked' protein complexes then enter the Orbitrap analyser, where their mass:charge ratio is measured and the mass spectrum outputted. Figure adapted from Gault *et al*[140].

instrument was able to perform mass measurements with greater precision than had previously been reported on a range of different membrane protein complexes, measurements which were significantly closer to the theoretical oligomer masses, calculated from the protein sequence.

With the significantly increased resolution provided by the Q-Exactive instrument, small mass differences between high molecular weight species are now resolvable, and hence a number of different experiments become feasible that are not possible on conventional Q-ToF instruments.

Firstly, it is possible to resolve small ligands bound to large protein complexes. For example using the Q-Exactive an endogenous lipid bound to the 186 kDa

pentameric ion channel ELIC can be observed as a set of satellite peaks to the main *apo*-pentamer peaks, whereas these satellite peaks are not resolved in the Q-ToF spectrum[140].

Secondly, multiple binding events involving similar mass species can be dissected. This was demonstrated by Gault *et al.*, who used the instrument to investigate binding of the 111 kDa trimeric porin OmpF to the OBS1 peptide, in the presence of a phosphatidylglycerol lipid. Using a Q-ToF instrument, when the OmpF/peptide/lipid mixture is analysed, there is a detected species which could be assigned to the either the trimer bound to a single OBS1 or to 2 PG molecules, but distinguishing between these two options is not possible because their theoretical masses differ by only 55.5 Da. Using the orbitrap instrument, however, the small mass difference was resolvable and the relative abundances of both species could then be quantified.

Whilst this second example is rather contrived, it neatly illustrates the potential of the instrument to deconvolute the binding of multiple species to a non-covalent protein complex, something which will be important in characterising protein-endogenous-lipid complexes in later chapters.

2.9 Characterising lipid-dependent oligomerisation by mass spectrometry

In this chapter we have introduced non-denaturing MS as a powerful method for elucidating the quaternary structure and lipid interactions of oligomeric membrane proteins. With the high energy source and orbitrap modifications, at the outset of this work we had access to an expanded MS toolkit, which opened up the possibility to characterise protein-lipid complexes in more detail by systematic dissociation and high resolution mass measurement. In the rest of this thesis, we will describe how these tools can be used to investigate the relationship between membrane protein lipid interactions and oligomeric state, and gain biological insight which would not be achievable with other structural biology techniques.

3

SemiSWEET: characterising a lipid-dependent monomer-dimer equilibrium

Contents

3.1 SWEETs and SemiSWEET 66
3.2 SemiSWEET co-purifies with endogenous lipids 68
3.3 High energy MS elucidates composition of SemiSWEET-lipid complex 69
3.4 High resolution MS characterises cardiolipin molecules directly bound to SemiSWEET 71
3.5 SemiSWEET binds cardiolipins with longer chain lengths than those in the bulk membrane bilayer 75
3.6 SemiSWEET exists in a dynamic equilibrium between monomeric and dimeric forms 77
3.7 Cardiolipin binding shifts the monomer-dimer equilibrium towards to the functional dimer 81
3.8 Discussion 84
3.9 Materials and Methods 88
3.9.1 Protein production and purification 88
3.9.2 Non-denaturing mass spectrometry 89
3.9.3 High energy non-denaturing mass spectrometry experiments 90
3.9.4 Characterising cardiolipin bound to SemiSWEET 91
3.9.5 Characterising cardiolipin in the bulk *E. coli* membrane 91
3.9.6 Monitoring subunit exchange of SemiSWEET dimers . . 93
3.9.7 Cardiolipin titration experiment 93
3.9.8 Contributions 94

In the following two chapters we will discuss the applications of the new mass spectrometry approaches outlined in chapter 2 to study the endogenous lipid binding and oligomerisation of two dimeric transporters, the sugar transporter SemiSWEET from *Vibrio splendidus* and the bioamine transporter LeuT from *Aquifex aeolicus*. In each case, we begin by identifying the lipids which co-purify with the membrane protein, by selective dissociation of the protein-lipid complex, then investigate the role of these lipids in mediating the oligomeric state of the protein, either by manipulating endogenous protein-lipid interactions or by adding in exogenous lipids.

3.1 SWEETs and SemiSWEET

This chapter focuses on our investigations into the sugar transporter SemiSWEET from *Vibrio splendidus*. SemiSWEETs are the prokaryotic members of the SWEET family of monosaccharide and disaccharide transporters, which are found in organisms from Archaea to plants to humans. The first SWEETs were identified in 2010[141], and were predicted to form a 7 transmembrane helix bundle consisting of two 3TM repeats separated by a linker helix. The prokaryotic SemiSWEETs, identified in 2013[142], were found to consist of a single 3TM unit, suggesting that the SWEETs evolved from a duplication of the ancestral SemiSWEET gene, and that SemiSWEETs dimerise to form the functional pore for sugar transport. Structural information on the SemiSWEETs soon followed, with the publication of the structures of two SemiSWEETs in 2014[143]. The structural study reported a plausible dimeric structure for SemiSWEET, as shown in figure 3.1 in which the translocation pore forms at the dimer interface, and provided convincing evidence that the physiological and functional oligomeric state of SemiSWEET is dimeric.

This was the state of the literature on SemiSWEET as we began our study into membrane protein oligomerisation and how it is influenced by protein-lipid interactions. We identified the *Vibrio sp.* SemiSWEET as a suitable candidate system with which to test our MS method for investigating the role of protein-lipid

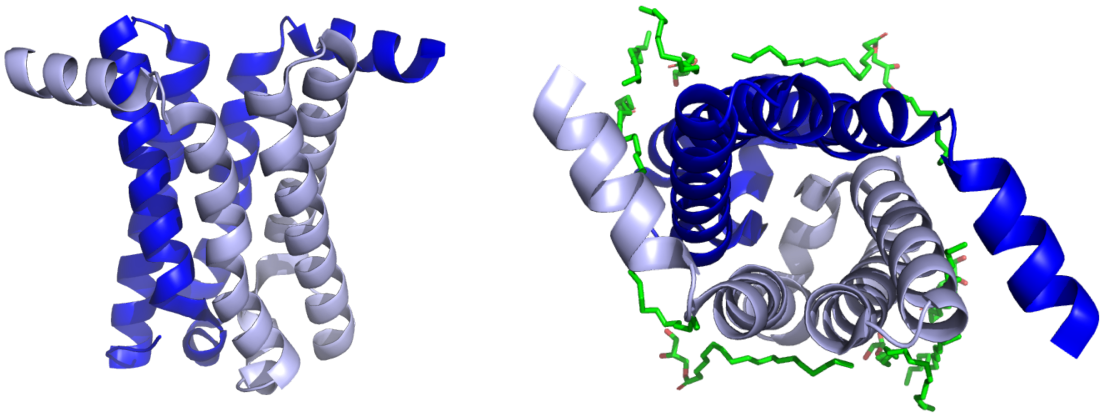


Figure 3.1: Crystal structure of the dimeric *Vibrio sp.* SemiSWEET, viewed from the side of the membrane (left) and the extracellular face (right). The structure also contains a small number of long-chain non-lipid molecules (green) which stick to the protein surface. Structures adapted from PDB ID 4QND[143].

interactions in mediating oligomerisation for several key reasons. SemiSWEET is one of the smallest reported membrane proteins, with a monomeric mass of only 11 kDa, and therefore we predicted that we would be able to generate high-quality and high-resolution spectra as a proof of concept for our MS-based method. The SWEETs as a family also merit detailed study since they are also of significant broader importance in both plants (where sucrose transport is a crucial factor in determining crop yields[144]) and humans, where glucose transporters are implicated in glucose homeostasis and mutations in these transporters can lead to conditions such as diabetes, glucose malabsorption and epilepsy[145]. Finally, in a more mundane practical sense, the previously published work on SemiSWEET had shown that the protein could be produced reliably by recombinant expression in *E. coli* in good yields and also provided a structural model against which we could rationalise our results. Working with a system in which a specific oligomeric state has been shown to be functional is also beneficial in that it allows some connection to be drawn between conclusions reached about oligomerisation behaviour and consequences for protein function.

3.2 SemiSWEET co-purifies with endogenous lipids

We began our study by expressing and purifying SemiSWEET from *E. coli*, according to the previously published protocol[143], before carrying out a non-denaturing MS analysis on the purified, detergent-solubilised protein complex using our standard membrane protein MS platform. Figure 3.2 shows the mass spectrum of SemiSWEET solubilised in DDM, the detergent which was used in the purification process and one commonly used for solubilisation of purified membrane protein complexes. The spectrum is dominated by a charge state distribution corresponding to the mass of the SemiSWEET monomer with satellite peaks (most obviously on the 6+ charge state) indicative of endogenous lipid binding. Unfortunately, no SemiSWEET dimer (the proposed functional oligomeric state) could be detected under these conditions, implying that DDM, whilst useful for purification, was not a suitable detergent for MS experiments studying the oligomerisation of SemiSWEET. The absence of dimer is most likely due to the high levels of gas phase activation required to remove the DDM micelle in the collision cell, which would result in dissociation of the dimer, although the possibility that SemiSWEET is monomeric in DDM in solution cannot be ruled out.

Despite dissociation of the dimer, it was possible to gain some insight into the protein-lipid interactions of SemiSWEET by analysing the DDM-solubilised sample. *E. coli* has a relatively simple lipid composition, with only three major lipid classes: phosphatidylethanolamine (PE, 75%), phosphatidylglycerol (PG, 20%) and cardiolipin (CL, 5%)[146]. The chemical structures of these lipids are shown in figure 3.3.

In the spectrum in figure 3.2, the well-resolved satellite peaks correspond to masses of around + 750 Da and + 1,470 Da relative to the monomeric protein mass. We assigned the + 750 Da peak as a bound (diacyl-)phospholipid, but assigning the + 1,470 peak was more problematic because of the similarity in mass between 2 phospholipids and 1 cardiolipin, a lipid which is abundant in the *E. coli* membrane.

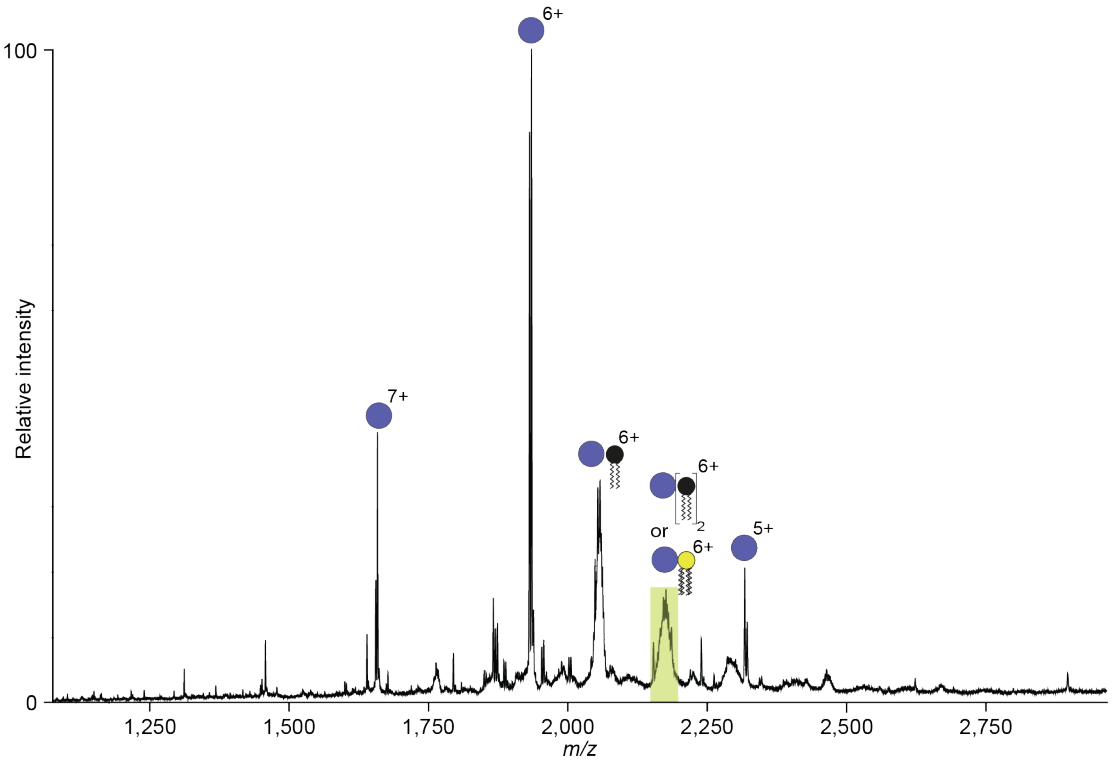


Figure 3.2: Mass spectrum of SemiSWEET (, 13.1 kDa), released from DDM micelles, shows the SemiSWEET monomer binding to a single diacylphospholipid (, 750 Da) and either two diacylphospholipids or a cardiolipin (, ≈ 1,470 Da).

3.3 High energy MS elucidates composition of SemiSWEET-lipid complex

In order to unpick the relative contributions of phospholipid and cardiolipin to the + 1,470 Da monomer adduct, we turned to the HE-nMS method which was introduced in Chapter 2. Whilst the standard membrane protein MS approach employed above involves liberating the monomer from the DDM micelle in the collision cell, in the HE-nMS method we were able to dissociate the micelle in the source region, and then selectively transmit only the m/z window corresponding to the monomer + 1,470 Da adduct. Further activation could be then applied to the protein-lipid complex in order to elucidate its composition. The results of this experiment, in which increasing collision voltages are applied to the protein-lipid complex, are shown in figure 3.4. Upon increasing activation, 3 new species become visible in the mass spectrum. First we see species emerging which correspond

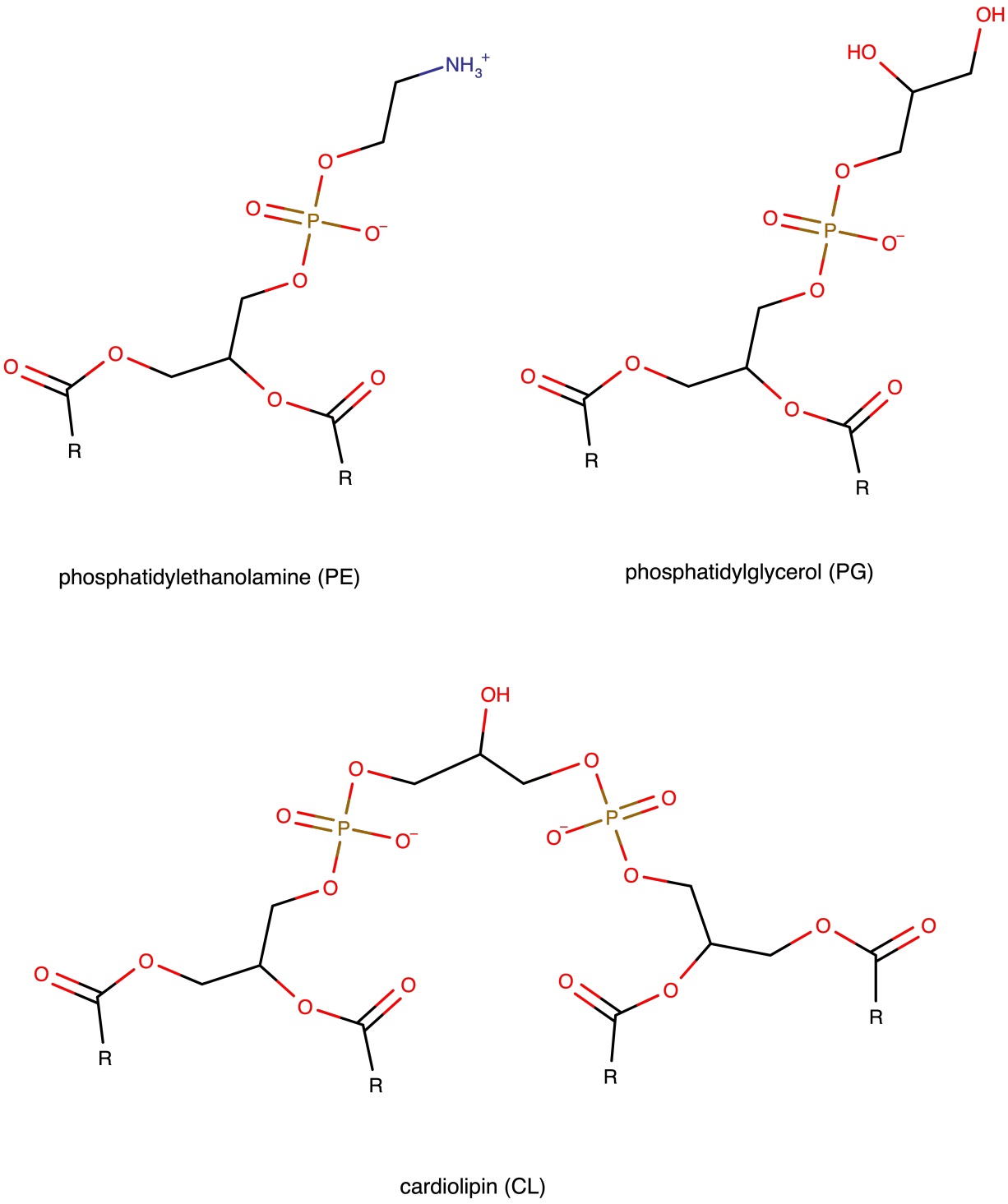


Figure 3.3: Chemical structures of the major lipid classes in *E. coli*.

to the 5+ and 6+ charge states of the monomer plus one diacylphospholipid - these represent a loss of a positively charged and neutral phospholipid respectively. At higher collision voltages a delipidated monomer species, generated by loss of a further neutral diacylphospholipid emerges and then dominates the spectrum. Finally at collision voltages of 80 V and above, direct conversion of the + 1,470 Da adduct to the delipidated monomer occurs, implying loss of a single cardiolipin. Taking together all of these dissociation steps, from the HE-nMS experiment we can deduce that the SemiSWEET monomer is co-purifying with at least 2 diacylphospholipids and 1 cardiolipin.

3.4 High resolution MS characterises cardiolipin molecules directly bound to SemiSWEET

In chapter 1 we discussed how it remains challenging to characterise the different layers of lipids around a membrane protein: bulk lipids, annular lipids and tightly-bound structural lipids. Conventional lipid identification methods normally involve chemical extraction of lipids from a purified membrane protein or cellular sample using organic solvent. These extracted lipids are then separated by a chromatographic procedure and then identified and characterised using MS-based lipidomics. The problem with this approach is that any information on the location and binding modes of the lipids is lost. Considering a solution of purified membrane protein, solubilised in a detergent micelle, as an example, we can surmise that the solution contains lipids in several different chemical environments: bound tightly to the protein; forming an annulus around the protein surface; inside the proteomicelle but not in direct contact with the protein; inside 'empty' lipid-detergent micelles etc. Lipids from all these different environments are mixed together during extraction with organic solvent and hence contextual information on the lipids' immediate environment is lost.

From the initial experiments on SemiSWEET we deduced earlier that the purified protein remains tightly bound to diacyl phospholipids and cardiolipin. Taking the bound cardiolipin as an example, we hypothesised that, if it were possible to measure

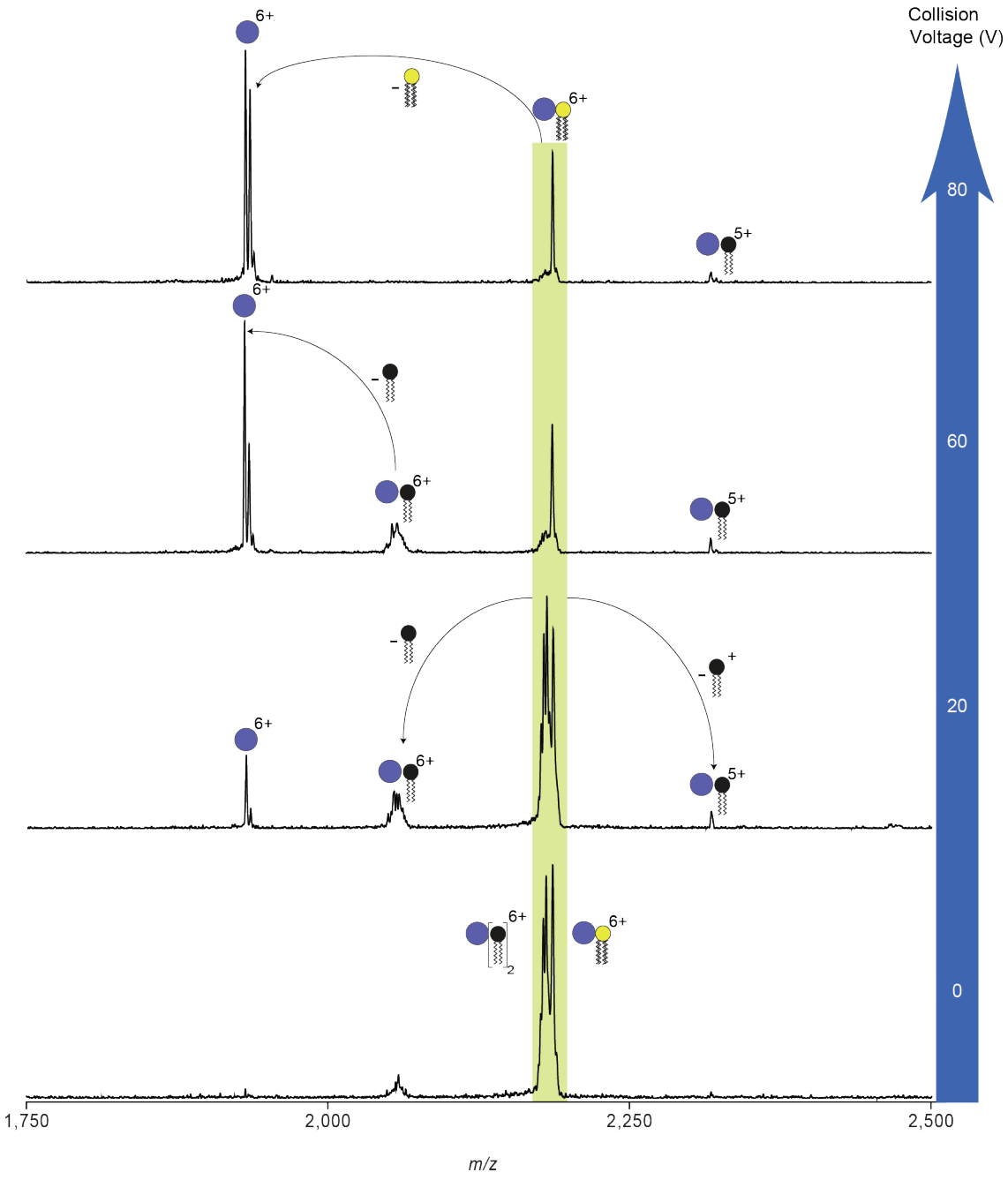


Figure 3.4: HE-nMS analysis of the 6+ charge state of SemiSWEET monomer-lipid complex, released from DDM micelles, as a function of collision voltage. The detergent-free monomer-lipid complex (6+ charge state highlighted in green) is selected in the quadrupole, and then dissociated via collisions with neutral gas molecules in the trap collision cell. The resulting mass spectra show sequential loss of up to 2 diacylphospholipids (, 750 Da), and 1 cardiolipin (, ≈ 1,470 Da), to leave the apo-SemiSWEET monomer (, 13.1 kDa).

the monomer + cardiolipin mass at very high resolution, it should be possible to analyse the fine structure of the spectrum and therefore gain insight into the subset of cardiolipin molecules which are tightly bound to SemiSWEET.

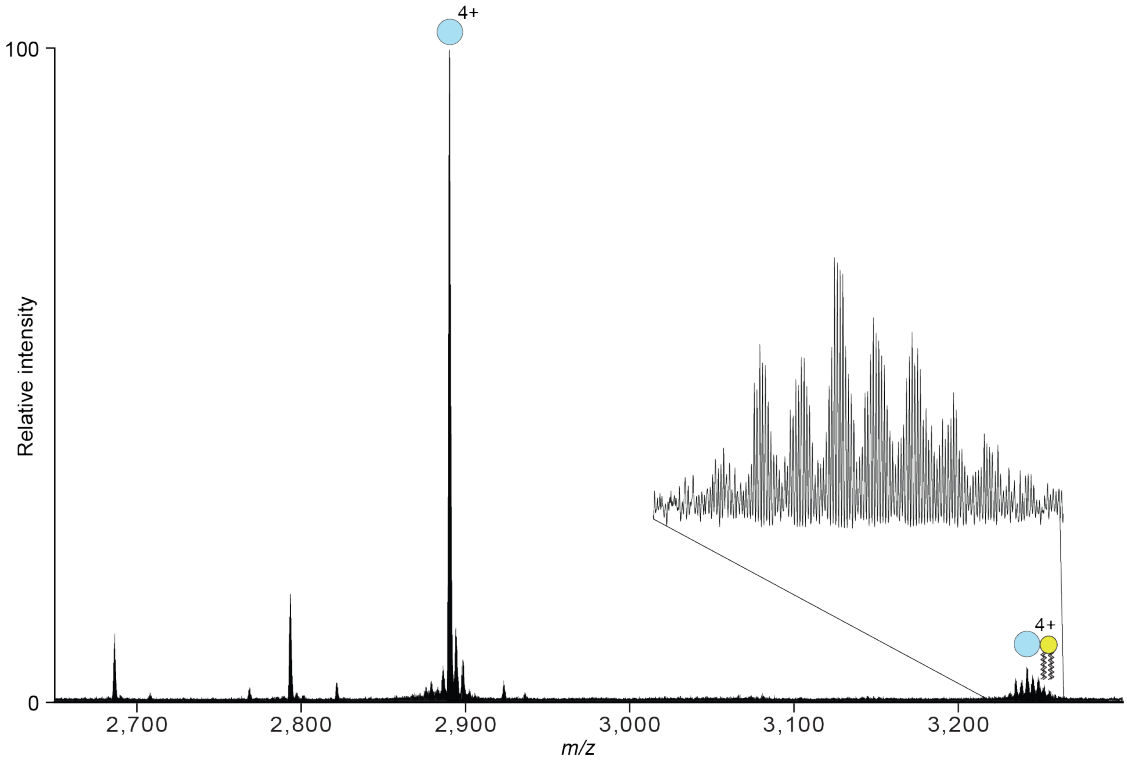


Figure 3.5: High resolution mass spectrum of SemiSWEET (○, 11.6 kDa) released from DDM micelles, shows SemiSWEET monomer binding to cardiolipin ( ) at isotopic resolution. Inset shows expansion of cardiolipin-bound peak, revealing a well-resolved distribution of a homologous series of lipids

.

To carry out a high resolution analysis, we first modified the purification of SemiSWEET in order to introduce a further purification step (cleavage of the C-terminal poly-histidine tag followed by reverse IMAC). This adjusted protocol produced noticeably better quality spectra, with sharper peaks and improved signal:noise, but at the cost of a significant amount of yield from the purification. An unintended consequence of this change to the purification protocol was that the amount of diacylphospholipid retained at the end of the purification was significantly reduced, whereas cardiolipin binding was retained. This removed much of the heterogeneity in the monomer + 1,470 Da peak and hence simplified the analysis. Using the modified Q-Exactive instrument, discussed in chapter 2, with its high

resolution orbitrap mass analyser, we then re-recorded the mass spectrum of DDM-solubilised SemiSWEET at the maximum resolution available to the instrument. The resulting spectrum is shown in figure 3.5, focusing on the +4 charge state. Immediately apparent from the spectrum is fine structure in the lipid-bound peaks which was not resolvable in our earlier experiments using the time-of-flight analyser. Zooming in on the cardiolipin-bound monomer region, we observed a series of Gaussian distributions which differed by a mass of 14 Da. We deduced that these Gaussians must represent different acyl chain lengths of the bound cardiolipin, with the 14 Da representing a CH_2 unit. The distributions corresponding to different chain lengths had their own fine structure, resolvable down to the level of different isotopes. This is the first reported example of a non-covalent complex between a membrane protein and a lipid to be characterised at isotopic resolution.

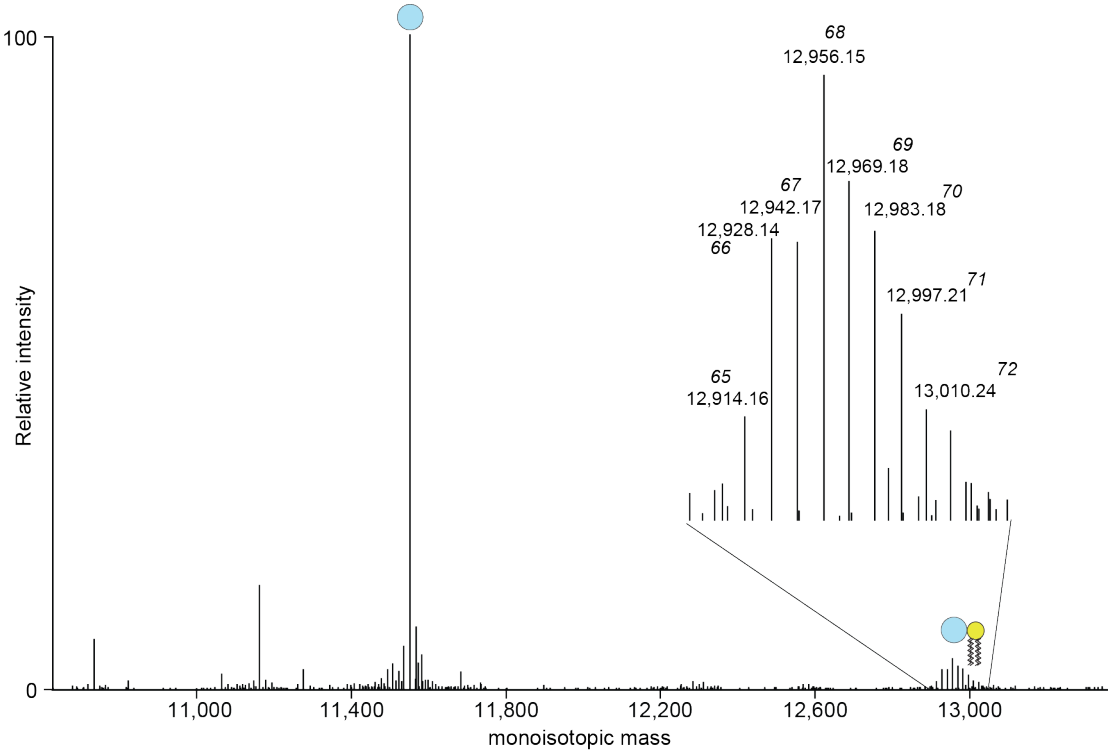


Figure 3.6: Monoisotopic zero-charge mass spectrum of SemiSWEET, released from DDM micelles, shows the SemiSWEET monomer (●, 11.6 kDa) binding to cardiolipin (⚲). Inset shows expanded SemiSWEET-cardiolipin region, with different peaks corresponding to different total acyl-chain lengths of the bound cardiolipin, labelled with mass and total number of acyl chain carbon atoms (italics).

In order to analyse the distribution of acyl chain lengths in SemiSWEET-bound cardiolipin, we applied the Xtract deconvolution algorithm to the isotopically resolved mass spectrum in order to generate the monoisotopic, zero-charge spectrum, shown in figure 3.6. Xtract is based on the THRASH algorithm developed by Horn *et al.*[147], and consists of the following steps. First, the algorithm examines the spacing within clusters of isotopically resolved peaks to obtain an estimate for charge state and therefore the mass of the components. It then fits the observed distribution of intensities to that of a poly-'averagine',[148] a hypothetical protein with isotopic abundances calculated from statistical occurrences of amino acids in proteins. This allows determination of the monoisotopic mass which best fits the experimental profile. Finally, the results from all charge states are combined to produce a single monoisotopic mass for each component, in this case each chain length of lipid.

The relative abundance of each chain length of lipid can be quantified easily, since the deconvolved, monoisotopic mass spectrum combines into a single peak the contributions from all of the different isotopes and charge states. We assigned the different chain lengths by subtracting the monoisotopic protein mass from each lipid-bound peak and then comparing the resulting cardiolipin masses to a database of reported *E. coli* lipids[149].

3.5 SemiSWEET binds cardiolipins with longer chain lengths than those in the bulk membrane bilayer

In order to provide a point of comparison, we also performed a conventional lipid extraction on a sample of the *E. coli* cells used to express SemiSWEET (taking the sample immediately before cell harvesting). We then used LCMS to separate the different lipids by their hydrophobicity, record the masses of the different lipids, and confirm their identity by fragmentation. By summing the region of the chromatogram where cardiolipins are known to elute, we generated a mass spectrum of the cardiolipin in the total membrane. We could then follow a similar process to that used to analyse the SemiSWEET-bound cardiolipins, and hence

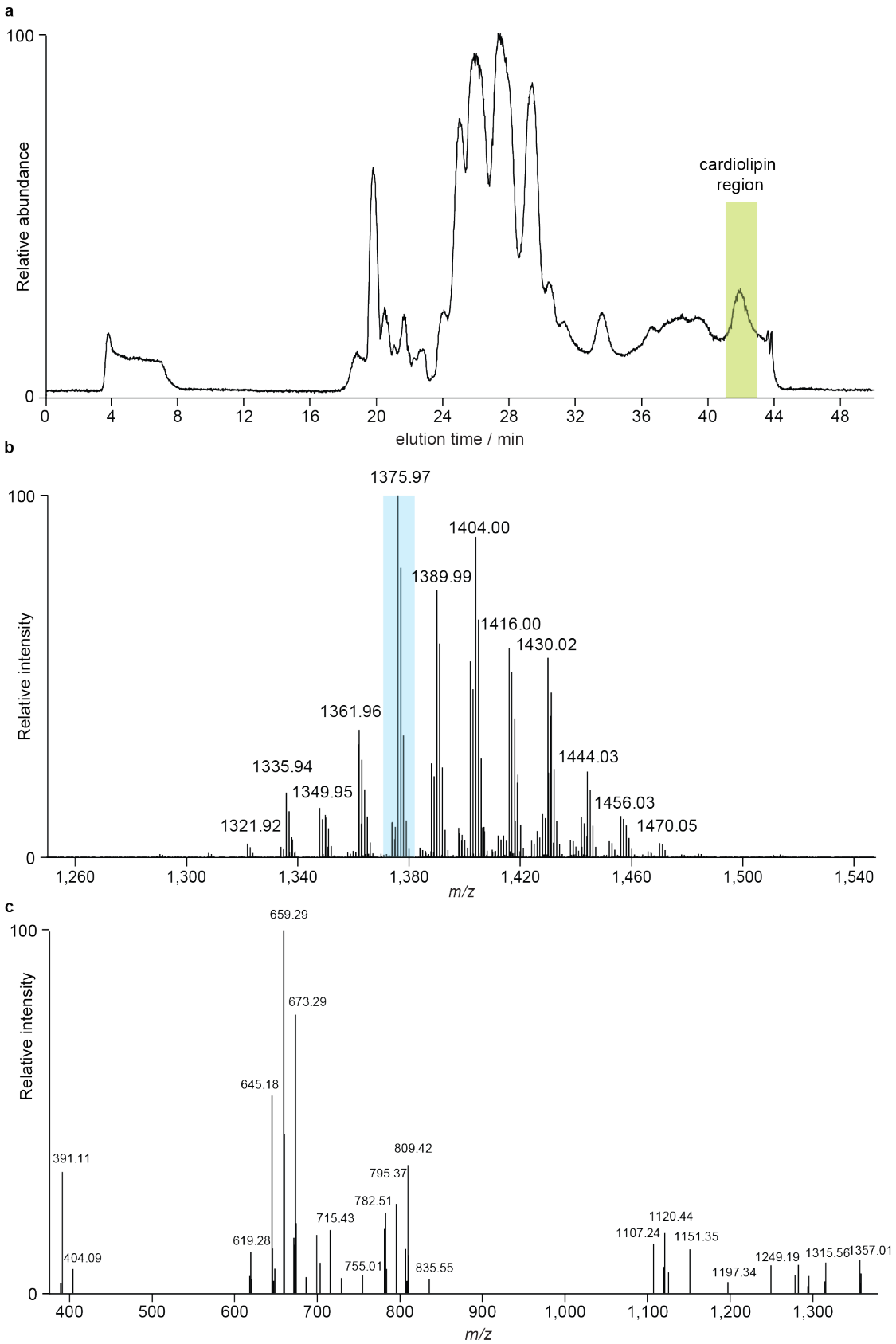


Figure 3.7: **a** Total ion count of total lipid extract from *E. coli*, analysed by reverse-phase LCMS. **b** Summed MS scans corresponding to the region where cardiolipin elutes, showing CL with different chain lengths, **c** Example MS/MS spectrum generated by fragmentation of the peak at $m/z = 1375.97$, identifying the lipid as a 66:2 CL by similarity to previously reported CL spectra[149].

obtain the distribution of chain lengths of cardiolipin present in the total cell membrane, as shown in figure 3.7.

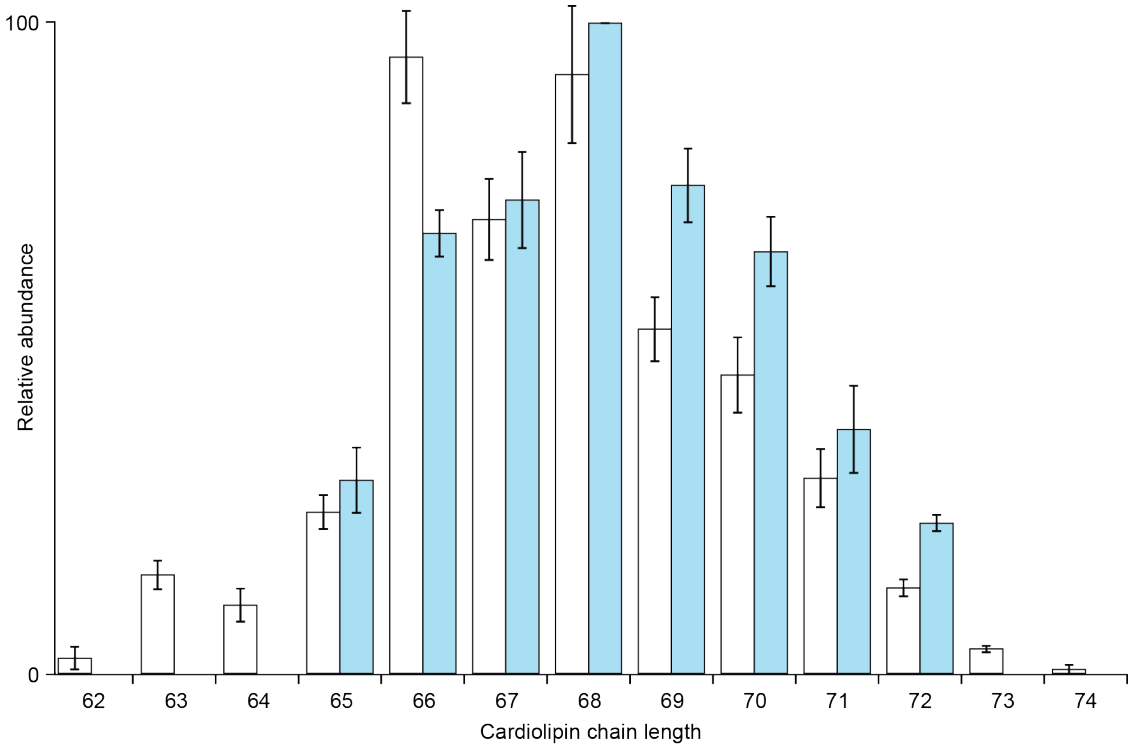


Figure 3.8: Comparison of the total acyl chain lengths of cardiolipin molecules bound directly to SemiSWEET (blue bars) to those extracted from the total *E. coli* membrane (open bars). Error bars represent 1 standard deviation from the mean.

Figure 3.8 is a plot of the relative abundance of the different total acyl chain lengths of cardiolipin molecules in the total membrane and also bound directly to SemiSWEET. Strikingly, we discerned a small but marked shift to higher chain lengths for the SemiSWEET-bound cardiolipins compared to those in the bulk membrane.

3.6 SemiSWEET exists in a dynamic equilibrium between monomeric and dimeric forms

Whilst we were able to gain insight into the interactions between SemiSWEET and endogenous lipids using the DDM-solubilised form of the protein, in order to investigate SemiSWEET oligomerisation we needed to find a detergent in which the functional dimer form was visible by MS. We therefore carried out a screen

of MS-compatible detergents as described previously[113], with the aim of finding a more labile detergent i.e. one which could be dissociated from SemiSWEET under gentle conditions thereby preserving the dimer. Following the screen, we settled upon tetraethylene glycol monooctyl ether (C_8E_4) as it gave high quality spectra with minimal levels of collisional activation and hence permitted detection of SemiSWEET in its biological oligomeric state. Figure 3.9 shows the mass spectrum of SemiSWEET in C_8E_4, in which both dimeric and monomeric forms are clearly visible. It is interesting to note that endogenous lipids, which were observed binding to the SemiSWEET monomer in the DDM spectrum, are absent here; evidently the process of detergent exchange from DDM to C_8E_4 resulted either in delipidation of the protein in solution, or in a weakening of protein-lipid interactions such that they are unable to persist in the gas phase.

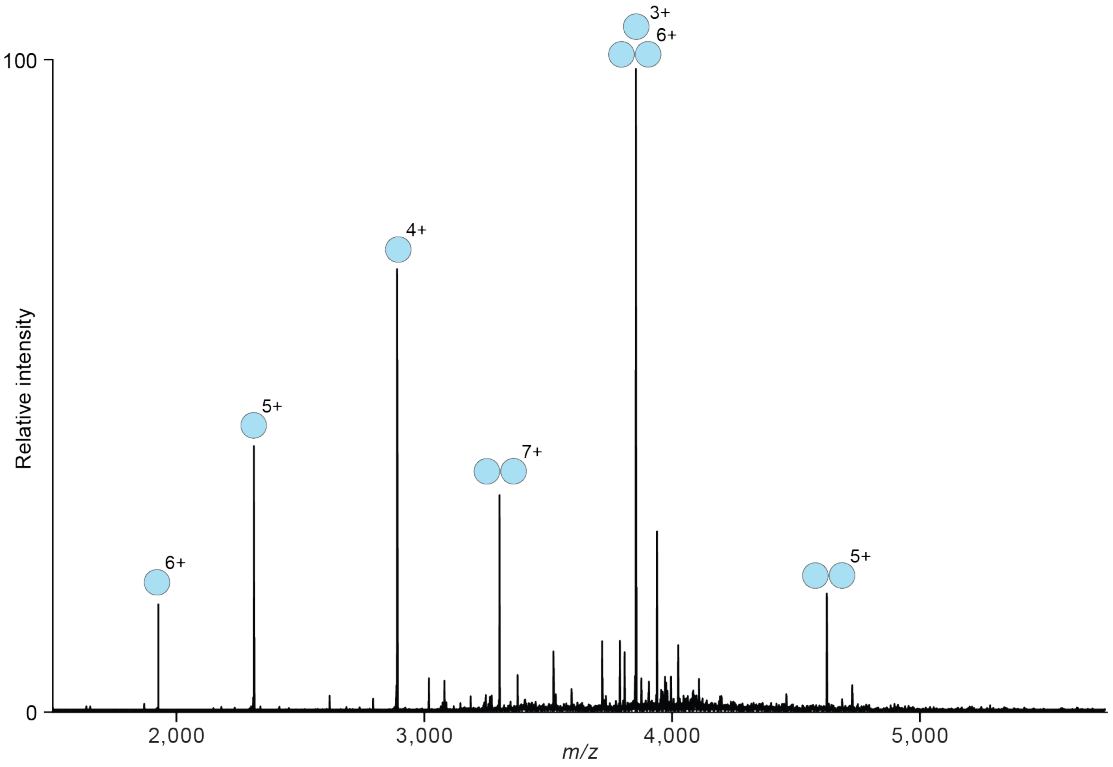


Figure 3.9: Mass spectrum of SemiSWEET, released from C_8E_4 micelles, shows SemiSWEET monomers (○, 11.6 kDa) and dimers (○○, 23.1 kDa).

The presence of significant populations of both monomeric and dimeric forms of SemiSWEET, under conditions in which we would expect gas-phase dissociation to

be minimal, in the mass spectrum prompted us to speculate whether there might be a dynamic solution-phase equilibrium between these two oligomeric states. In order to test this theory we prepared 'heavy' and 'light' forms of SemiSWEET (with and without the deca-His tag respectively) in C_8E_4, spectra shown in figures 3.10a and 3.9. We then mixed the heavy and light forms in an equimolar ratio, hypothesising that, if SemiSWEET monomers and dimers were freely equilibrating, there should be gradual formation of a 'mixed' dimer population consisting of dimers with one heavy and one light subunit, see schematic in figure 3.10b. Over time, the heterodimer would become the dominant population, since statistically we would predict a 1:2:1 abundance ratio of light:mixed:heavy dimers. These predictions were borne out by experiment: after mixing equimolar amounts of purified light and heavy SemiSWEET and allowing them to equilibrate for 10 minutes, we recorded a mass spectrum of the mixture. Figure 3.10c shows the dimer region of the spectrum, from which it can be seen that there is a prominent heterodimer charge state distribution. We were also able to monitor this subunit exchange process in real time by transferring the heavy/light mixture to the electrospray capillary immediately after mixing and then continuously recording the mass spectrum for ten minutes. By summing the spectrum across a small scan window at 1 minute intervals, and then quantifying the relative abundance of the different dimer populations, we generated time courses for the subunit exchange experiments, a typical example is shown in figure 3.10d. The time courses showed that subunit exchange can occur on the minute timescale, and provide powerful evidence that SemiSWEET dimers can freely exchange with monomers as part of a dynamic monomer-dimer equilibrium.

Whilst these experiments elegantly illustrate that MS can be used to monitor monomer-dimer equilibria, there are a number of practical challenges which prevented us from extracting reliable quantitative parameters such as equilibrium constants or rate constants for these processes. Whilst the deca-His tag was a convenient way to produce a 'heavy' SemiSWEET without needing to use molecular cloning to generate a new expression construct, it is a fairly large tag given the small size of SemiSWEET (equivalent to 13% of the protein mass) and it could

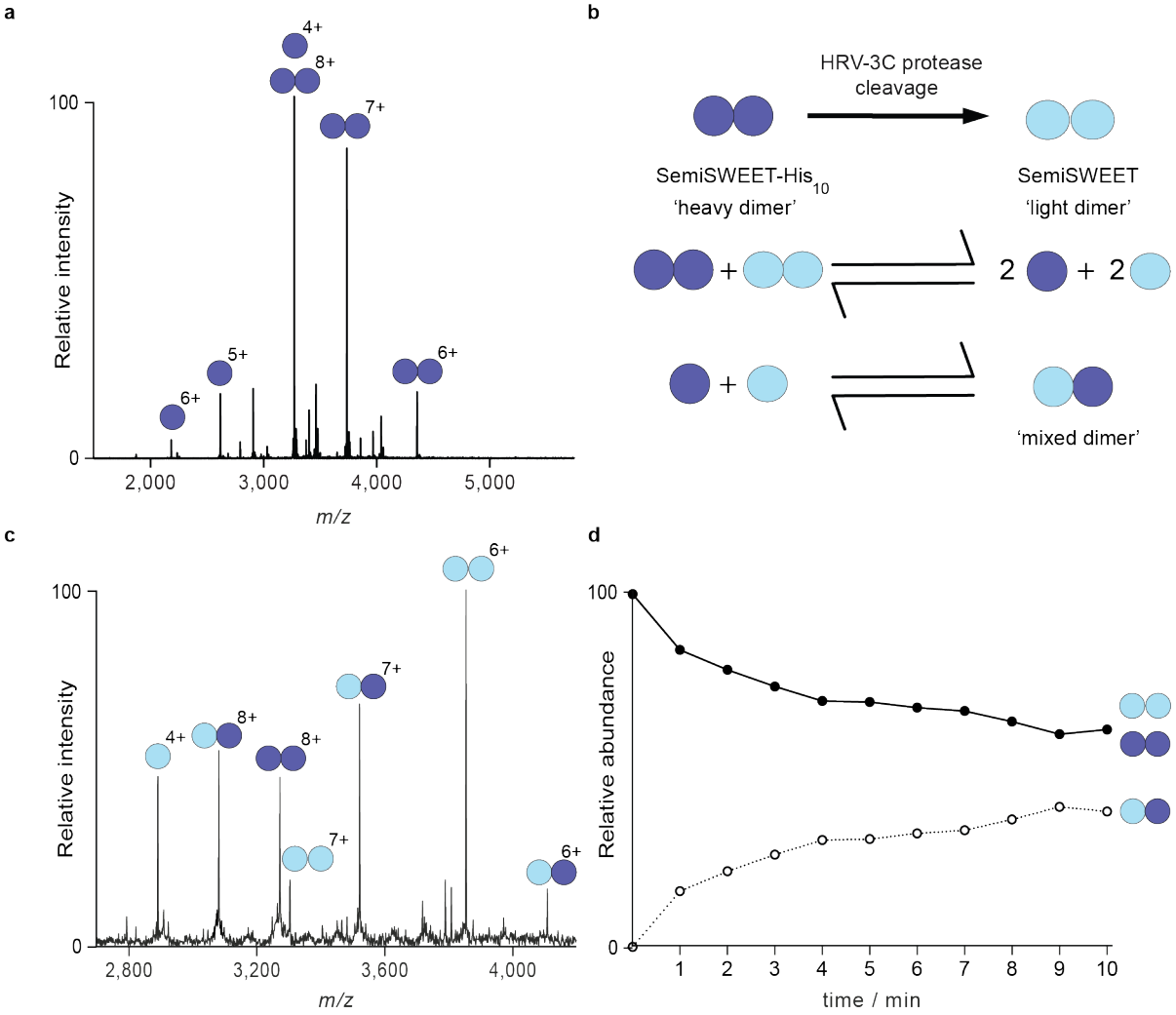


Figure 3.10: SemiSWEET dimers undergo subunit exchange, implying a dynamic monomer-dimer equilibrium. **a** Mass spectrum of SemiSWEET-His10, released from C_8E_4 micelles, shows SemiSWEET monomers (●, 13.1 kDa) and dimers (●●, 26.2 kDa). **b** Schematic of subunit exchange experiment. 'Light' dimers are generated by cleaving the deca-His tag from 'heavy' dimers. Upon mixing of light and heavy dimers, if the system is in dynamic equilibrium then the dimers will dissociate to form free monomers and then reform heavy, light or 'mixed' dimers. **c** Mass spectrum of SemiSWEET recorded 10 minutes after mixing equimolar ratios of heavy and light dimers. Present in the spectrum are SemiSWEET light dimers, heavy dimers and mixed dimers (●●, 24.6 kDa). **d** Time course of subunit exchange between SemiSWEET dimers shows the appearance of a heterodimer population on a timescale of minutes.

therefore affect various properties of the protein. When mixing solutions of light and heavy SemiSWEET that we measured as equal concentration in solution, we observed more heavy subunits in the mass spectrum. This implies that the heavy dimers either had a greater ionisation efficiency than the light dimers, or that a small amount of dissociation of light dimers was occurring which affected the heavy dimers to a lesser extent. Both of these potential pitfalls could be avoided by using a different tag - the high resolution of this experiment would mean that only a very small mass difference between the light and heavy subunit is required. Alternatively, or the protein mass could be shifted without changing the chemistry at all by, for example, producing the protein from *E. coli* grown in ^{15}N or ^{13}C media. With these optimisations, future work could be targeted towards assessing how the rate of subunit exchange is affected by altering conditions, adding small molecules ligands such as drug molecules, or modifying the hydrophobic environment of the proteins by adding different lipids.

3.7 Cardiolipin binding shifts the monomer-dimer equilibrium towards to the functional dimer

Now that we had established that SemiSWEET dimers are in a dynamic equilibrium in solution, in which dimers dissociate to form monomers and then reform, we next investigated if this equilibrium could be perturbed by modification of the lipid environment around the protein. Our earlier experiments analysing the endogenous lipids bound to SemiSWEET had identified cardiolipin as a lipid which can form strong interactions with SemiSWEET and that these interactions can persist throughout protein purification and the various manipulations inside the mass spectrometer. Whilst the detergent conditions in which we were able to monitor SemiSWEET dimerisation unfortunately also removed endogenous lipids, we supposed that adding exogenous cardiolipin to a solution of SemiSWEET could result in the same interactions forming, as long as the lipid molecules were given enough time to integrate into the detergent micelles. We began by testing to see if exogenous cardiolipin binding to SemiSWEET could be detected by MS and found

that the CL binding to both the monomer and dimer could be observed in the spectrum. A typical spectrum of SemiSWEET after incubation with exogenously added cardiolipin is shown in figure 3.11.

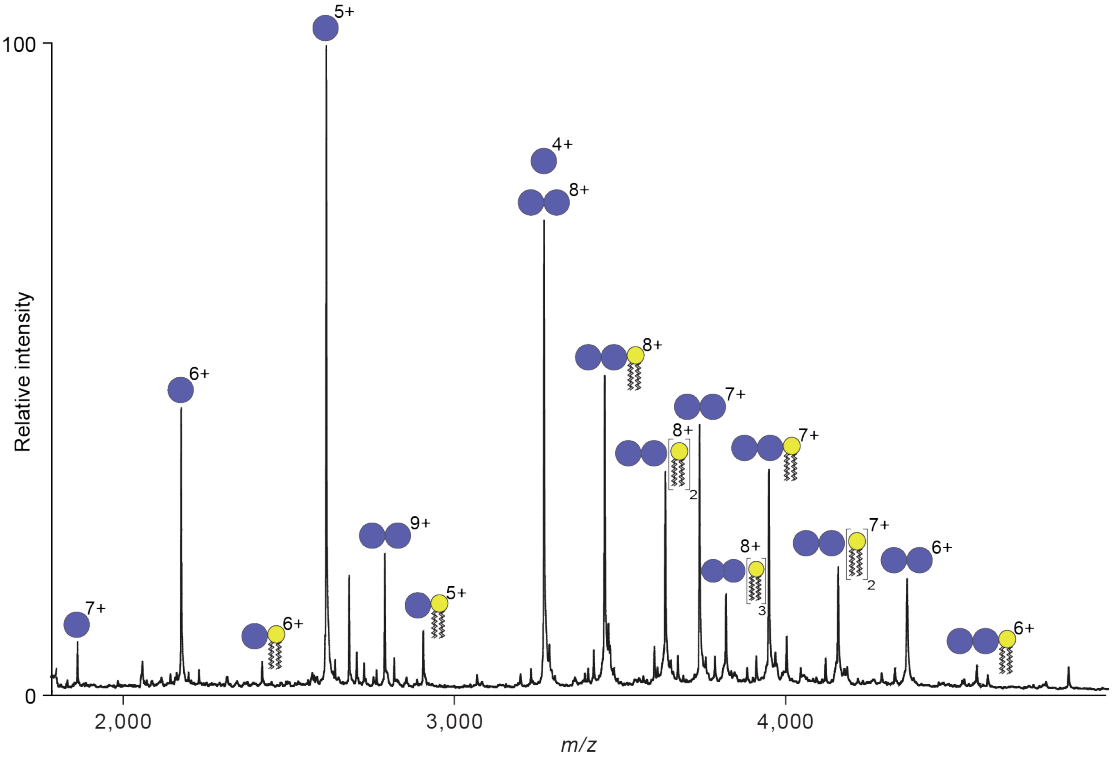


Figure 3.11: Mass spectrum of SemiSWEET, released from C8E4 micelles, after addition of exogenous cardiolipin. Both SemiSWEET monomers (●, 13.1 kDa) and dimers (●●, 26.2 kDa) bind cardiolipin (⚷, ≈ 1,470 Da), but the dimer does so to a greater extent.

With exogenous cardiolipin binding established, we then carried out an experiment in which we added different amounts of cardiolipin to a fixed concentration of SemiSWEET in solution and then measured the effect on the relative abundance of the monomer and dimer populations. The results are shown in figure 3.12. There is an obvious trend in the data: the relative abundance of the dimer increases as more cardiolipin is added to the solution. This suggests that cardiolipin binding stabilises the dimer relative to the monomer, and could also indicate a shift in the equilibrium position in solution towards the dimer. One alternative interpretation might be that some SemiSWEET dimer is being dissociated to monomer in the gas phase, and cardiolipin binding to SemiSWEET simply stablises the individual subunits with respect to unfolding, thereby preventing gas-phase dissociation of

the dimer. This would result in an increased dimer:monomer ratio in the mass spectrum, without a corresponding change in the solution populations of the two forms. We sought to reduce the contribution from this effect by using as gentle conditions as possible throughout the lipid concentration series, setting the collision energy to just high enough to dissociate the C_8E_4 micelle, and hence minimising gas-phase dissociation of the dimer.

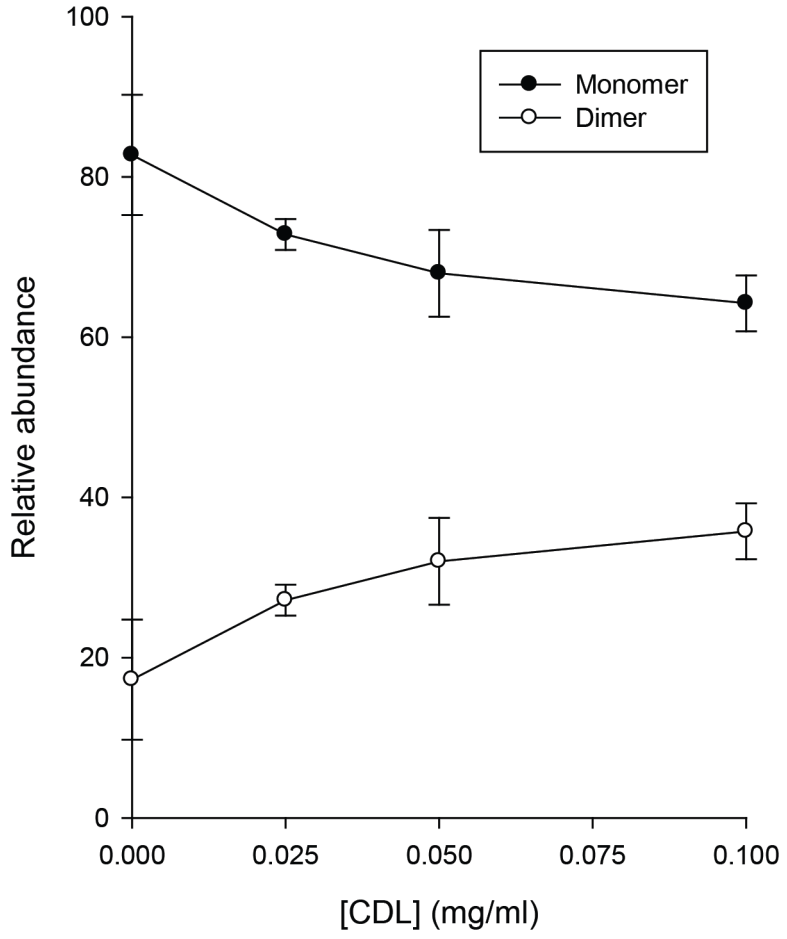


Figure 3.12: Plot showing the relative abundances of SemiSWEET monomers and dimers, extracted from mass spectra, after addition of different concentrations of exogenous cardiolipin. Error bars represent 1 standard deviation from the mean.

3.8 Discussion

In this chapter we have demonstrated that, using non-denaturing MS, it is possible to gain insight into the local lipid environment around the sugar transporter SemiSWEET, its oligomerisation behaviour and the relationship between these two phenomena.

Firstly, by detecting endogenous phospholipids and cardiolipin molecules that co-purify with SemiSWEET, we examined the lipids which are in direct contact with the protein. Whilst endogenous lipids have been detected in mass spectra in the past[122;126], in this study we were able to use recent methodological advances to probe the composition of protein-lipid complexes in greater detail. With HE-nMS we have shown that an MS/MS approach can be employed to selectively dissociate membrane protein-lipid complexes and hence resolve ambiguities in assignment when different lipid combinations have very similar masses. For SemiSWEET this allowed us to dissociate a complex of SemiSWEET and a lipid species of around 1,450 Da and find that the lipid component had contributions from both 2 (diacyl)phospholipids and a single cardiolipin. In addition we used the high resolving power of the orbitrap mass analyser to generate the first isotopically resolved spectrum of a non-covalent protein lipid complex, from which we could infer the distribution of acyl chain lengths of SemiSWEET-bound cardiolipin, as distinct from cardiolipin in the bulk *E. coli* membrane. This high resolution experiment illustrates the power of MS as a single molecule technique which is able to describe the heterogeneity in a system: since the mass of each cardiolipin molecule is measured separately, a distribution of chain lengths can be extracted rather than an averaged value across the ensemble of lipids. Analysis of the chain length distribution of the SemiSWEET-bound cardiolipins showed a slight preference for longer chain lengths relative to those we measured in the bulk membrane, suggesting that SemiSWEET could remodel the membrane by selecting particular lipids to form the annulus around itself. If so, SemiSWEET would be an elegant example of the symbiotic relationship between lipids and membrane proteins, in which membrane proteins can affect the lipids around them whilst their own properties are in turn being influenced by lipid binding.

A MS method for interrogating endogenous lipids in direct contact with membrane proteins has significant potential utility, as it opens up the possibility of gaining insight specifically into the subset of lipids which form the innermost layer around the protein, known as the lipid annulus. These are the lipids which form protein-lipid interactions which have been shown to be vital for protein structure[61], stability[130], and activity[150]. Furthermore, it is challenging to identify the lipids directly associated with a membrane protein by other biophysical techniques. X-ray crystal structures can resolve some structural lipids[72], but are time-consuming and challenging to solve, often lack the resolution to assign chain lengths and head group moieties. There are very crystal structures in which a full lipid annulus around a membrane protein can be resolved, with a notable exception being the 2D X-ray crystal structure of Aqp0[151], in part because crystallisation conditions involve high concentrations of detergents which outcompete non-selective lipid binding. Molecular dynamics simulations have been applied to a wide range of membrane proteins with the aim of gaining insight into low-selectivity protein lipid interactions, and have been shown to predict lipid positions consistent with those seen in crystal structures[80]. However, these simulations depend upon a pre-specified lipid composition, and so would be unable to identify effects such as a preference for longer lipid acyl chain lengths around a protein without incorporating a biologically representative heterogeneity in chain lengths into the lipids when the simulation is set up.

One aspect of this MS approach to identifying endogenous lipids which remains poorly understood is the role of the small number of endogenous lipids it identifies. Evidently these lipids form relatively strong interactions with membrane proteins, since they are retained throughout extensive purification steps and the various gas-phase manipulations of the MS experiment, but it is not yet clear whether they are structural lipids which bind in well-defined binding sites or whether they represent part of an annulus which covers the whole protein lipophilic surface. A more detailed understanding of the different layers of lipids around membrane proteins could be provided by combining a non-denaturing MS characterisation of tightly bound lipids with conventional lipidomics experiments which focus on a

larger number of lipids in the local environment. For example, a recently reported approach developed in our lab involves subjecting a purified membrane protein to repeated rounds of delipidation using detergent and then using lipidomics to characterise the lipids associated with the membrane protein after each round[152;135]. Alternatively, detergent extraction could be avoided altogether by extracting the membrane protein and its associated lipids directly from the membrane using methods like styrene-maleic acid lipid particles (SMALPs), also known as native nano-discs. The co-polymer liberates the protein and its lipids from the membrane, and, once purified, the encapsulated lipids can be identified and quantified by lipidomics and hence information on the local environment obtained. We predict that only with a combination of these different MS approaches, supported by molecular dynamics simulations and crystallography data will it be possible for a full picture of membrane protein-lipid interactions to be built.

The second key aspect of our studies on SemiSWEET were our investigations into its dimerisation. We have shown that SemiSWEET exists in a dynamic equilibrium between monomeric and dimeric forms, that this equilibrium can be monitored in real-time using MS, and that perturbations of the equilibrium position can be measured. According to classification outlined in the introduction, we therefore predict that SemiSWEET is therefore a weak transient oligomer, as it can exist in a dynamic equilibrium between two oligomeric states.

Membrane protein subunit exchange had been inferred during previous studies of the multi-drug resistance transporter EmrE carried out by our lab[122], but the work reported here is the first case of it being monitored in real time. The experiments reported here are a proof of concept that time-dependent measurements of oligomerisation are achievable with an MS technique, but they also open up the possibility of a thorough quantitative characterisation of membrane protein oligomerisation in future, something which few other techniques are able to achieve. A rigorous, quantitative picture of membrane protein oligomerisation in which an equilibrium dissociation constant, association rate constant, and dissociation rate constant are calculated, would be a powerful way of characterising the oligomerisation behaviour

of transiently forming oligomers. In these systems, a particular membrane protein will adopt a specific distribution of oligomeric states in its 'rest' state, but will then form a new distribution in response to a biochemical stimulus, examples of which could include ligand binding, post translational modification or a change in the local lipid environment. At the moment, the lack of quantitative evaluations of oligomeric equilibria severely limits the ability of biochemists to understand the influence of these stimuli on the oligomeric state of proteins such as G-protein coupled receptors, and hence predict correctly the time dependent consequences of these stimuli on downstream signalling processes.

Two key questions arose from our observations that SemiSWEET exhibits a monomer-dimer equilibrium and that the equilibrium can be perturbed by adding cardiolipin. Firstly, what is the biological significance of this equilibrium for the host organism? From consideration of the crystal structure it seems highly likely that dimerisation is essential for formation of the translocation pore, and so the monomer-dimer equilibrium represents an equilibrium between functional and non-functional forms. The equilibrium therefore provides an opportunity for regulation of function - in theory by adjusting the position of the equilibrium, the level of sugar transport could be controlled by the cell. A number of different mechanisms by which oligomerisation can be induced or inhibited have been been reported previously, for example by post-translational modification[153] or ligand binding[154]. Control of oligomeric state could therefore provide a method for the host organism to modulate the amount of SemiSWEET-mediated sugar transport in response to an external stimulus or change in environmental conditions. Secondly, could changes in the lipid environment around SemiSWEET be a method by which its oligomeric state and therefore function are modulated? It has been shown that the lipid composition of the cell can vary both spatially[155] and temporally[70], and one can therefore speculate that lipid-mediated oligomerisation could result in different levels of SemiSWEET activity at different times, under different conditions or stimuli or in different regions of the cell.

In conclusion, in this chapter we have seen that non-denaturing MS can be used to detect and probe the composition of endogenous lipids bound to the sugar transporter SemiSWEET, identify and monitor SemiSWEET dimerisation and investigate the effect of lipid binding on the oligomeric state. We saw that adding cardiolipin shifts the monomer-dimer equilibrium towards the functional dimer form, but that this effect is rather subtle. In the next chapter we will apply the same method to the bioamine transporter LeuT, a dimeric protein which we will show binds a larger cohort of lipids which, in contrast to SemiSWEET, are essential for dimerisation.

3.9 Materials and Methods

3.9.1 Protein production and purification

V. splendidus SemiSWEET was expressed from a pJexpress411 vector, containing a HRV 3C protease-cleavable 10 × His tag. The plasmid was transformed into BL21 DE3 *E. coli* (Novagen), and expressed and purified as reported previously[143]. Briefly, multiple colonies were used to inoculate 100 ml of LB media and grown overnight at 37 °C. 10 ml of this overnight culture was then used to inoculate each of 6 l of LB, which were then allowed to grow at 37 °C until the culture had attained an A_{600} of 0.8. Isopropyl β-d-thiogalactopyranoside (IPTG) was added, to a final concentration of 0.2 mM, and the culture grown for a further 15 hours at 22 °C. Cells were harvested by centrifugation at 5,000 g for 10 min at 4 °C, resuspended in lysis buffer (300 mM sodium chloride, 20 mM Tris; pH 7.4) supplemented with protease inhibitor cocktail (Roche) and lysed using a M-110 PS microfluidizer (Microfluidics), and the cell debris pelleted by centrifugation at 20,000 g for 25 min at 4 °C. Membranes were pelleted by centrifugation of the supernatant at 100,000 g for 2 h at 4 °C and subsequently resuspended in ice-cold resuspension buffer (100 mM sodium chloride, 20 mM Tris, 20% glycerol; pH 8.0) and homogenized using a Potter-Elvehjem Teflon pestle and glass tube. n-dodecyl β-d-maltopyranoside (DDM) was added to resuspended membranes to a final concentration of 2% (w/v) and the suspension incubated with gentle agitation for 2 h at 4 °C. Insoluble material was pelleted by centrifugation at 20,000 g for 30

min, and the supernatant filtered through 0.22 μm filters. SemiSWEET was purified by immobilised metal ion affinity chromatography (IMAC), using a HisTrap HP 5ml column (GE Healthcare) equilibrated with buffer A (150 mM sodium chloride, 20 mM Tris, 20 mM imidazole, 10% glycerol, 1 mM DTT, 0.02% DDM; pH 8.0) and eluted with buffer B (150 mM sodium chloride, 20 mM Tris, 500 mM imidazole, 10% glycerol, 1 mM DTT, 0.02% DDM; pH 8.0). The eluted protein was transferred to a dialysis cassette (100 kDa molecular weight cut-off (MWCO)) and left to dialyse overnight at 4 °C in dialysis buffer (150 mM sodium chloride, 20 mM Tris, 10% glycerol, 1 mM DTT, 0.02% DDM; pH 8.0). After overnight dialysis, dialysed protein was then concentrated using a 100 kDa MWCO concentrator and then injected onto a Superdex 200 Increase GL 10/300 column (GE Healthcare), equilibrated in buffer (150 mM sodium chloride, 20 mM Tris, 10% glycerol, 1 mM DTT, 0.02% DDM; pH 8.0). Peak fractions corresponding to DDM-solubilised SemiSWEET were collected, concentrated and used for further study. For generation of SemiSWEET without the deca-His tag, HRV 3C protease was added to the dialysis cassette, and a reverse IMAC step was added to the procedure immediately after dialysis.

C8E4-solubilised SemiSWEET was prepared by injecting concentrated DDM-solubilised SemiSWEET onto a Superdex 200 Increase GL 10/300 column equilibrated with 150 mM sodium chloride, 20 mM Tris, 10% glycerol, 1 mM DTT, 0.5% C8E4 at pH 8.0.

3.9.2 Non-denaturing mass spectrometry

Initial SemiSWEET spectra and cardiolipin titration experiments were acquired using a Synapt G1(Waters) with a Z-spray source, using nanoelectrospray capillaries prepared in-house[114]. The source pressure was set to 4-7 mBar, with a capillary voltage of 1.4-1.7 kV, capillary nanoflow of 0.05-0.2 mBar and argon as collision (trap) gas at flow rate of 1.5-8.0 ml min^{-1}. Other parameters, including the sample and extraction cone and trap bias voltages, collision voltages and quadrupole profile were optimized for maximal ion intensity and minimal dissociation of the target membrane protein complex[113]. Data were processed using MassLynx software.

Characterisation of SemiSWEET-bound cardiolipin experiments and subunit exchange experiments were carried out using a Q-Exactive orbitrap mass spectrometer[140], modified and optimised for non-denaturing mass spectrometry measurements of intact membrane protein complexes. Spectra were acquired in 'Native Mode' with maximum radio frequency applied to all ion optics, -3.2 kV to the central electrode of the Orbitrap and with ion trapping in the higher-energy collisional dissociation (HCD) cell. Ions were generated in positive ion mode from a static nanospray source using gold-coated nanospray capillaries prepared in-house. The AGC target was set to 1×10^6, the noise-level parameter to 3. The collision gas was argon and pressure in the HCD cell was maintained at approximately 1×10^{-9} mbar.

3.9.3 High energy non-denaturing mass spectrometry experiments

High energy source experiments were carried out using a Synapt G2 (Waters), modified to allow the use of higher voltages at the extraction cone in the source region of the instrument. This modification required implementation of the capability to drive the sample cone voltage from an external supply. This took the form of a patch cable introduced between the instrument lens control PCB and the source ion block. The ion block contains both heater elements and a thermocouple in addition to supporting the extraction cone and these were decoupled with the patch cable to prevent possible electrical breakdown with use of higher cone voltages. The extraction cone was patched directly through from the lens PCB while the sample cone voltage was decoupled and a new wire connection made to an external power supply. Finally, the configuration file for the extraction cone was modified to increase the maximum voltage setting from 10 V to 200 V.

HE-nMS experiments were then carried out as follows. Source pressure, capillary voltage, nanoflow and trap pressure were maintained as for standard non-denaturing MS experiments. Initially the trap and transfer collision voltages are set to 0 and the quadrupole profile set to transfer all ions. After the sample is loaded, and a stable spray established, the sample cone voltage is then increased until the detergent

micelle has been fully dissociated, and well-resolved detergent-free charge states of the oligomer of interest are observed. The m/z region of interest is identified and the quadrupole profile is set to transmit only this m/z window. The trap and transfer collision voltages can then be increased to induce dissociation of the isolated complex.

3.9.4 Characterising cardiolipin bound to SemiSWEET

Three separate preparations of DDM-solubilised SemiSWEET were prepared for MS analysis by buffer-exchange into mass spectrometry buffer (200 mM ammonium acetate, 0.02% DDM, 1 mM DTT, pH 7.4) using a centrifugal buffer exchange device (Micro Bio-Spin, Biorad). The samples were then diluted to an oligomeric protein concentration of 10 μM, 3 μl of which was used for each MS measurement. Q-Exactive spectra were acquired with 10 microscans and summed over 50-100 scans. An HCD cell voltage of 200 V was applied, and a transient time of 140 ms to achieve isotopic resolution. No in-source activation was applied. Data were analysed using Xcalibur software as follows. Spectra were first deconvolved using the Xtract algorithm to generate a monoisotopic spectrum. SemiSWEET-bound lipid peaks were then assigned chain lengths (where chain length n represents a lipid with n total acyl chain carbon atoms) by subtracting the measured monoisotopic *apo*-protein mass and then comparing masses to previously published *E. coli* cardiolipin spectra. Intensities were normalised against the most abundant lipid species and then averaged across the three biological repeats. Averages for each chain length species were then plotted as a percentage of the total bound cardiolipin (error bars are 1 standard deviation).

3.9.5 Characterising cardiolipin in the bulk *E. coli* membrane

5 ml samples of *E. coli* cells were extracted from 4 randomly chosen flasks of a 12-litre SemiSWEET expression, immediately before cell harvesting. Cells were pelleted, the supernatant removed, and then chemically lysed by vigorous vortexing with lysis buffer (20 mM Tris-HCL, 300 mM NaCl, pH 7.4) followed by snap freezing.

The cells were then thawed and total lipid extracted by a modified Bligh and Dyer procedure[156]. Briefly, 375 μl of chloroform:methanol 2:1 was added to 100 μl of cell suspension, followed by 125 μl chloroform and 125 μl water. Samples were then vortexed and allowed to separate. The organic layer was removed and evaporated to dryness, before being resuspended in 1,000 μl of a solution consisting of 68% solution A (ACN:H_2O 60:40, 10 mM ammonium formate and 0.1% formic acid) and 32% solution B (IPA:ACN 90:10, 10 mM ammonium formate and 0.1% formic acid).

Lipid samples were analysed by reverse phase liquid chromatography coupled to tandem mass spectrometry (RP LC-MS/MS), using a Dionex UltiMate 3000 RSLC Nano system coupled to an LTQ Orbitrap XL hybrid mass spectrometer (Thermo Fisher Scientific). Lipid extract prepared as described above was loaded into a 1 μl loop using the autosampler and loading pump, then transferred to a C18 column (Acclaim PepMap 100, C18 54 μm $\times$ 15 cm, Thermo Fisher Scientific), with a flow rate of 30 nl/min. After 10 min, the proportion of solvent B was increased to 65% over 1 min, increased to 80% over 6 min, held at 80% for 10 min, increased to 99% over 6 min, and held at 99% for 10 min. The eluent was delivered to the LTQ orbitrap mass spectrometer via a nanospray source. Spectra were acquired in negative ion mode, with a spray voltage of -1.6 kV and a capillary temperature of 160 °C. After a full MS scan had been acquired in the orbitrap analyser, tandem MS was performed using a top 5 method. Ions were selected, fragmented, and then detected in the ion trap. The AGC target was set to 1×10^4, CID energy to 38% and the activation time to 30 ms.

Cardiolipin chain length data was extracted by summing the region of the chromatogram known to correspond to cardiolipin (based on known elution time). Spectra were extracted and then deconvolved using the Xtract algorithm. Intensities were extracted, normalised and plotted as described above for the SemiSWEET-bound cardiolipins.

3.9.6 Monitoring subunit exchange of SemiSWEET dimers

Purified samples of C8E4-solubilised SemiSWEET and SemiSWEET-His_{10} were buffer exchanged into MS buffer (200 mM ammonium acetate, 0.5% C8E4, 1 mM DTT; pH 7.4) and the oligomer concentration adjusted to 20 μM. 3 μl of each of SemiSWEET and SemiSWEET-His_{10} were then mixed briefly on ice and 3 μl loaded into a nanoelectrospray capillary for immediate mass spectrometry data acquisition. Data were acquired with 1 microscan and an HCD cell voltage of 50 V - no in-source activation was applied. Data was acquired continuously for 10 min. Data were processed using Xcalibur software as follows: spectra were extracted from summation of the chromatogram in 30 second scan windows centred on each minute (for example, 0.75-1.25 min for the 1 min time point). Relative abundances of the SemiSWEET and SemiSWEET-His_{10} homodimers and the SemiSWEET.SemiSWEET-His_{10} heterodimer were calculated using the UniDec software package[157]. Plot of relative abundance of the homodimer and heterodimer populations was generated using SigmaPlot.

3.9.7 Cardiolipin titration experiment

Purified SemiSWEET-His_{10} was buffer exchanged into MS buffer (200 mM ammonium acetate, 0.5% C8E4, 1 mM DTT; pH 7.4) and then diluted to an oligomer concentration of 20 μM. A cardiolipin stock solution, 10 mg/ml in 200 mM ammonium acetate, was prepared as described previously[113]; aliquots of this stock were then diluted in MS buffer to a cardiolipin concentration 2 $\times$ that required for each MS experiment. SemiSWEET-His_{10} and cardiolipin were mixed in a 1:1 volume ratio and incubated on ice for 5 minutes (for the measurement in the absence of cardiolipin, SemiSWEET was mixed with MS buffer). 2 μl of the protein-lipid mixture was then used for each of 3 MS measurements at each of 4 lipid concentrations.

MS parameters were set as described above, with the source pressure set to 4.2 mbar, capillary nanoflow 0.1 mBar, trap collision voltage 50 V, transfer collision voltage 10 V, and collision gas flow rate 1.8 ml min^{-1}. Data were acquired for

100 scans, which were summed using MassLynx software and then processed using UniDec deconvolution software[157]. Relative monomer and dimer abundances were calculated by taking the sum of the deconvoluted intensities in all lipidation states, normalised to the total intensity of all SemiSWEET species and averaged over the 3 repeats. A plot of the relative monomer and dimer abundances was generated using SigmaPlot (error bars ±1 standard deviation).

3.9.8 Contributions

- Protein purification was carried out by JACD.
- Standard non-denaturing MS experiments, including cardiolipin titrations and subunit exchange, were carried out by JACD.
- High resolution MS experiments, including characterisation of SemiSWEET-bound and bulk membrane cardiolipins, were carried out by JACD, with assistance from Joseph Gault.
- HE-nMS measurements were performed by Kallol Gupta.
- Data analysis was done by JACD, with assistance from Joseph Gault.

4

LeuT: A Dimer Dependant on Lipids

Contents

4.1 LeuT: insights from published structures 96
4.2 LeuT is a dimer which forms a non-covalent adduct with endogenous lipids 98
4.3 High energy native mass spectrometry identifies LeuT-bound lipids 101
4.4 Site-directed mutagenesis pinpoints lipid binding to the dimer interface 102
4.5 MD simulations reveal a cardiolipin binding mode which bridges the LeuT dimer interface 105
4.6 Cardiolipin binding is essential for LeuT dimerisation 109
4.7 Delipidation and Relipidation of LeuT *in vitro* Highlights Lipid-Dependent Dimerisation 110
4.8 Discussion 113
4.9 Materials and Methods 117
4.9.1 Molecular cloning and plasmid construction 117
4.9.2 Protein purification 118
4.9.3 Non-denaturing mass spectrometry 119
4.9.4 Chemical delipidation 121
4.9.5 Molecular dynamics simulations 121
4.9.6 Contributions 122

In the previous chapter, we saw that non-denaturing MS can be used to identify and characterise a small cohort of lipids that co-purify with the sugar transporter SemiSWEET, and used this information to investigate its oligomerisation and the

influence of cardiolipin on this process. We found that, whilst lipid binding could influence the position of the SemiSWEET monomer-dimer equilibrium, the dimer was stable in the absence of any bound lipids. In this chapter, we will turn to the bioamine transporter LeuT, and show that, in contrast to SemiSWEET, the LeuT dimer requires lipids bound in the dimer interface in order to be stable.

Furthermore, we find that is possible to pinpoint the location of lipid binding, the residues involved in protein-lipid interactions, and hence gain insight into the mechanism by which lipid binding influences the oligomeric state.

4.1 LeuT: insights from published structures

LeuT is a member of the neurotransmitter sodium symporter (NSS) family, a group of related proteins which use energy from sodium and chloride electrochemical gradients to transport a wide range substrates across the membrane[158]. In humans, NSSs such as the serotonin and dopamine transporters play a key role in neurotransmission,[159] and are therefore targets for pharmaceutical therapeutics[160], as well as many psychoactive drugs. LeuT initially became of interest to the structural biology community as a prokaryotic homologue of the eukaryotic NSSs. Whilst eukaryotic NSSs were considered very challenging targets in the early years of membrane protein crystallography, other prokaryotic homologues were predicted to be more tractable for structural elucidation, LeuT in particular since it comes from a thermophilic organism and is therefore highly thermally stable. These predictions proved to be accurate: LeuT was the first NSS structure to be solved in 2005[161]. The structure, shown in figure 4.1, established the 'LeuT fold' as a new fold for membrane proteins, elucidated the sites of sodium and leucine binding, and explained the mechanism of gating. Although LeuT crystallised as a monomer in the asymmetric unit in this first structure, a plausible dimer structure was generated via rotation of the monomer about a 2-fold symmetry axis and the prediction that LeuT is dimeric was confirmed by later structures which contained dimers in the asymmetric unit.

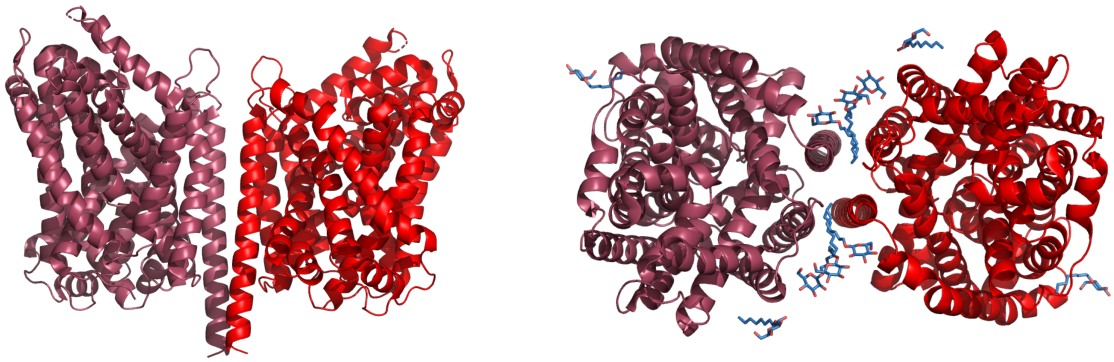


Figure 4.1: Crystal structure of the dimeric leucine transporter LeuT from *Aquifex aeolicus*, viewed from the side of the membrane (left) and the extracellular face (right). The structure also contains several octyglucoside detergent molecules (blue) in the dimer interface, as well as a small number of other medium chain non-lipidic molecules around the protein surface. Structures adapted from PDB ID 2A65[161].

When examining the dimer interface across different LeuT structures, we noted a number of interesting features. The interface consists of relatively few protein-protein contacts compared to other membrane protein interfaces, principally a set of π-stacking interactions between a small number of aromatic residues. In the 2005 structure solved by Yamashita *et al.* there is also electron density in the dimer interface corresponding to long alkyl chains, which the authors assigned to detergent molecules. The authors did not comment on the significance of these detergent molecules in the interface in their paper, but we noted that this density could equally represent interfacial lipid molecules. Difficulties in assigning electron density corresponding to long-chain aliphatic tails were often encountered when solving early membrane protein crystal structures, since lipid-like molecules often have greater conformational flexibility than the protein components of the structure and hence are less well-resolved, though this problem has been somewhat mitigated in the higher resolution structures which have appeared in the last 5 years[72]. Alternatively, the lipid-like molecules seen in the LeuT interface could be correctly be assigned as detergent molecules, but these detergent molecules could have displaced biological lipids during the purification or crystallisation process. Regardless, the presence of lipid-like molecules in the interface and the relatively sparse protein-protein

contacts suggested that LeuT might be a system in which lipids could mediate oligomerisation by binding at the oligomeric interface.

We therefore set out to investigate: whether we could detect endogenous lipids associated with the LeuT dimer; whether these endogenous lipids were localised to the dimer interface; and finally to what extent interfacial lipids were important for LeuT oligomerisation.

4.2 LeuT is a dimer which forms a non-covalent adduct with endogenous lipids

We began by expressing LeuT in *E. coli*, and purifying it according to a previously reported protocol[161]. In order to find an optimal detergent for LeuT MS experiments, we carried out a screen of MS-compatible detergents[113], identifying octyl glucoside (OG) as the detergent which gave the highest quality spectra whilst preserving the reported biological oligomeric state.

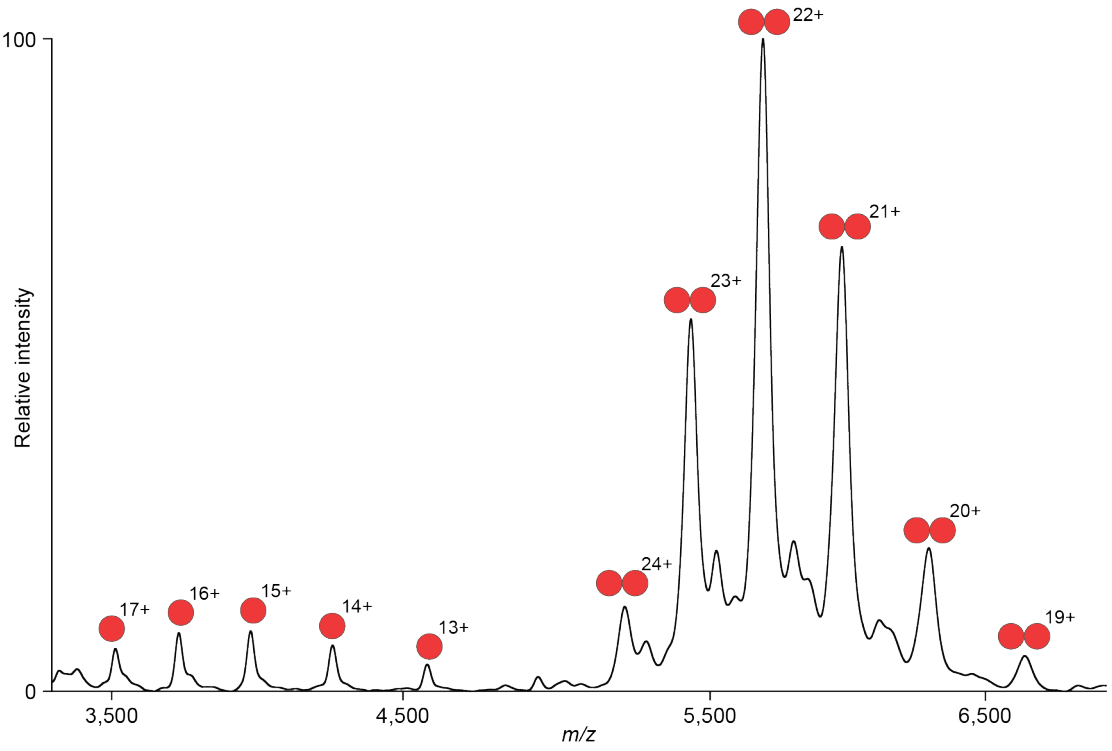


Figure 4.2: Mass spectrum of LeuT, liberated from OG micelles, reveals LeuT monomers (●, 59.3 kDa) and mass-shifted dimers (●●, 126.0 kDa).

Figure 4.2 shows the mass spectrum of OG-solubilised LeuT, which contains two well-resolved charge-state distributions: a high-intensity distribution centred around an *m/z* of 5,700 and a low-intensity distribution centred around an *m/z* of 4,000. The low-intensity distribution corresponds to a mass of 59.3 kDa, consistent with the mass of the LeuT monomer as predicted from the amino acid sequence. Assigning the high-intensity distribution was less straightforward, since it corresponds to a mass of 126.0 kDa, 7.4 kDa higher than would be predicted for a LeuT dimer (126.0 kDa = 2 x 59.3 + 7.4 kDa). We provisionally assigned the 126.0 kDa series as representing LeuT dimers forming a non-covalent adduct with some other species. We ruled out formation of both a covalent adduct (since in this case the monomers would also be mass-shifted) and a weakly-bound, non-specific adduct (since the adduct was stable across the full collision energy range of the instrument). Non-covalent interactions between membrane proteins and small molecules are commonly observed in membrane protein mass spectra, with the most likely small molecule candidates in this case being detergent or endogenous lipids. However, in the majority of membrane protein mass spectra reported before this study, detergent/lipid binding typically manifested as satellite peaks of larger *apo*-protein peaks[162]. This pattern is illustrated in figure 4.3 in which the pentameric ion channel ELIC is shown to bind up to 2 lipids as a series of satellite peaks of the *apo*-pentamer. However, in the LeuT spectrum there is no *apo*-dimer distribution and the spectrum is dominated by a distribution corresponding to a single adduct state.

Whilst assignment of the monomer charge state distribution is straightforward, it is important to determine if monomers seen in the mass spectrum represent solution-phase monomers or those formed from gas-phase dissociation of higher order oligomers. Some dissociation of higher order oligomers is possible (and occasionally unavoidable) in membrane protein mass spectra, since a significant amount of activation must be applied via collision with neutral gas molecules in order to liberate membrane protein oligomers from the detergent micelle. However, if the complex is activated too much the individual subunits begin to unfold and then dissociate, in a process known as collision-induced dissociation (CID), and

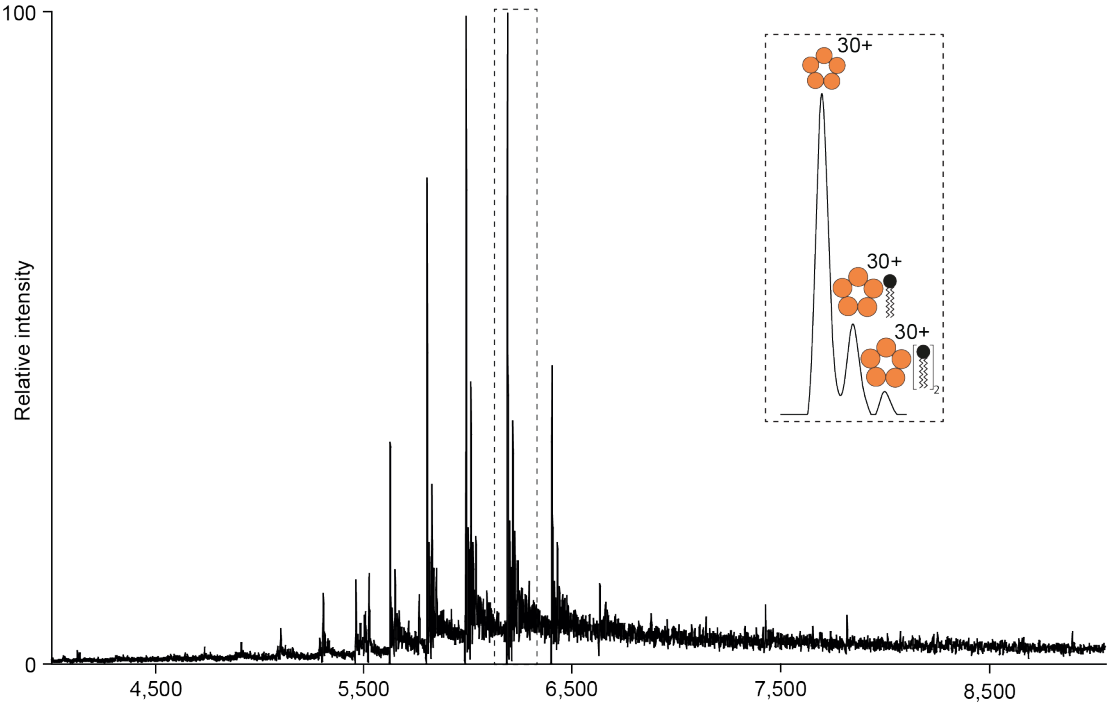


Figure 4.3: Mass spectrum of ELIC, liberated from DDM micelles, shows ELIC pentamers (, 186.4 kDa) bound to up to 2 phospholipids (, 750 Da). Inset shows expanded and smoothed 30 + charge state. Figure adapted from Gault *et al*, 2016 [140]

this produces mass spectra with oligomer distributions which do not correspond to those in solution [118]. CID is a process which produces distinctive signatures in mass spectra: the unfolded monomers take a disproportionate amount of charge and therefore appear as a broad distribution at low m/z whilst the remaining stripped oligomer retains proportionally much less charge at therefore appears at higher m/z. In the LeuT case this effect is clearly not seen in the spectrum: the monomer charge state distribution is no broader than that of the dimer, the average charge state is consistent with a folded monomeric species (with a slightly larger exposed surface area:volume ratio than the dimer), and there is no evidence of asymmetric dissociation into highly-charged and charge-reduced forms. Hence we concluded that both the monomer and dimer species seen in the spectrum are likely present in solution.

4.3 High energy native mass spectrometry identifies LeuT-bound lipids

To test our hypothesis that the LeuT dimer adduct is a protein-lipid complex, we sought to dissociate it in order to elucidate its composition. As discussed in chapter 2, simply repeating the MS analysis above with higher collision energy is often an ineffective way to probe composition of membrane protein-lipid complexes because there is often insufficient energy available for complete dissociation, and because without mass selection it is impossible to assign the origin of dissociation products seen in the mass spectra. Using a mass spectrometer modified to allow increased in-source activation of protein complexes, however, circumvents both of these problems, and hence we chose to analyse the LeuT dimer complex using the HE-nMS method outlined in chapter 2. Using this set-up, we were able to apply a much higher voltage at the extraction cone, and hence fully dissociate the OG micelle from LeuT in the source region. We then used the quadrupole to select narrow m/z windows corresponding to individual charge states of the dimer adduct, and then dissociated these adducts in the collision cell. Figure 4.4 shows the results of dissociating the 23+ dimer adduct charge state. Visible in the spectrum are LeuT monomers bound to a series of adducts with masses approximately 700, 1,400 and 2,100 Da, consistent with up to 3 phospholipid molecules binding to each monomer. Cardiolipin (CL), mass approximately 1,400 Da, is also an abundant component of the *E. coli* membrane, and so we reasoned that CL binding could be contributing to the + 1,400 Da or + 2,100 Da species, respectively alone or in combination with 1 other phospholipid.

Figure 4.5 shows the results of recording HE-nMS spectra over a range of collision voltages. From these spectra, we observed that, as the collision voltage was increased, there was a decrease in the abundance of the lipid-bound monomer peaks as the lipids are dissociated from the complex. The + 2,100 Da series disappeared first, but the + 1,400 Da series persisted, even as the intensity of the + 700 Da series decreased. This is consistent with a contribution from CL to the + 1,400 Da peak and possibly also to the + 2,100 peak. We therefore concluded that the LeuT

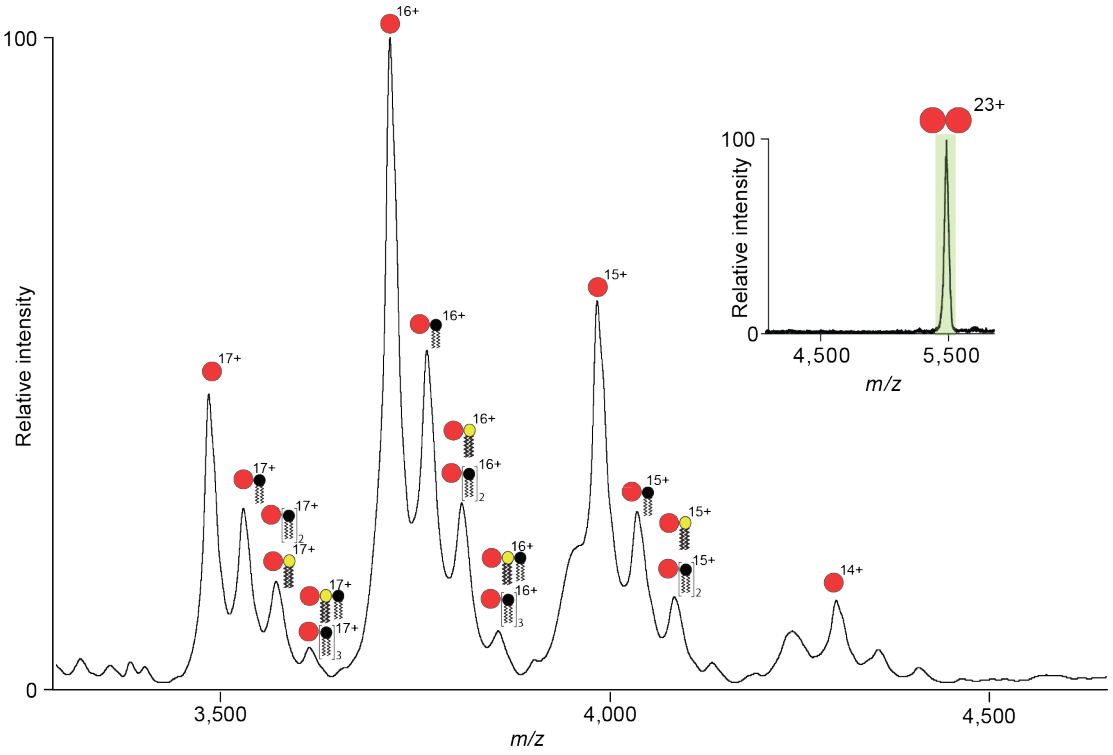


Figure 4.4: Mass spectrum of LeuT monomers (●, 59.3 kDa) bound to phospholipids (, ≈ 750 Da) and cardiolipin (, ≈ 1,450 Da) following dissociation of the 23+ charge state of the LeuT dimer adduct (●●, 126.0 kDa) by HE-nMS. Inset shows mass selection of the 23+ charge state prior to dissociation.

dimer forms an adduct with up to 3 phospholipids and 1 CL molecule per monomer. No well-resolved protein-detergent adducts were observed in any of the spectra, but we could not rule out the possibility that the 7.4 kDa mass shift may be a consequence of specific phospholipid and CL binding plus some weaker interactions with lipids, detergents or other small molecules which are completely dissociated from the monomers during the HE-nMS experiment and therefore not detected.

4.4 Site-directed mutagenesis pinpoints lipid binding to the dimer interface

Having identified the mass-shifted LeuT dimer as an adduct with lipid, we next considered where these lipids might bind and with what consequences for the overall structure. Given that the lipids bind to the dimer and not the monomer, we supposed that the most likely location for lipid binding would be at the dimer

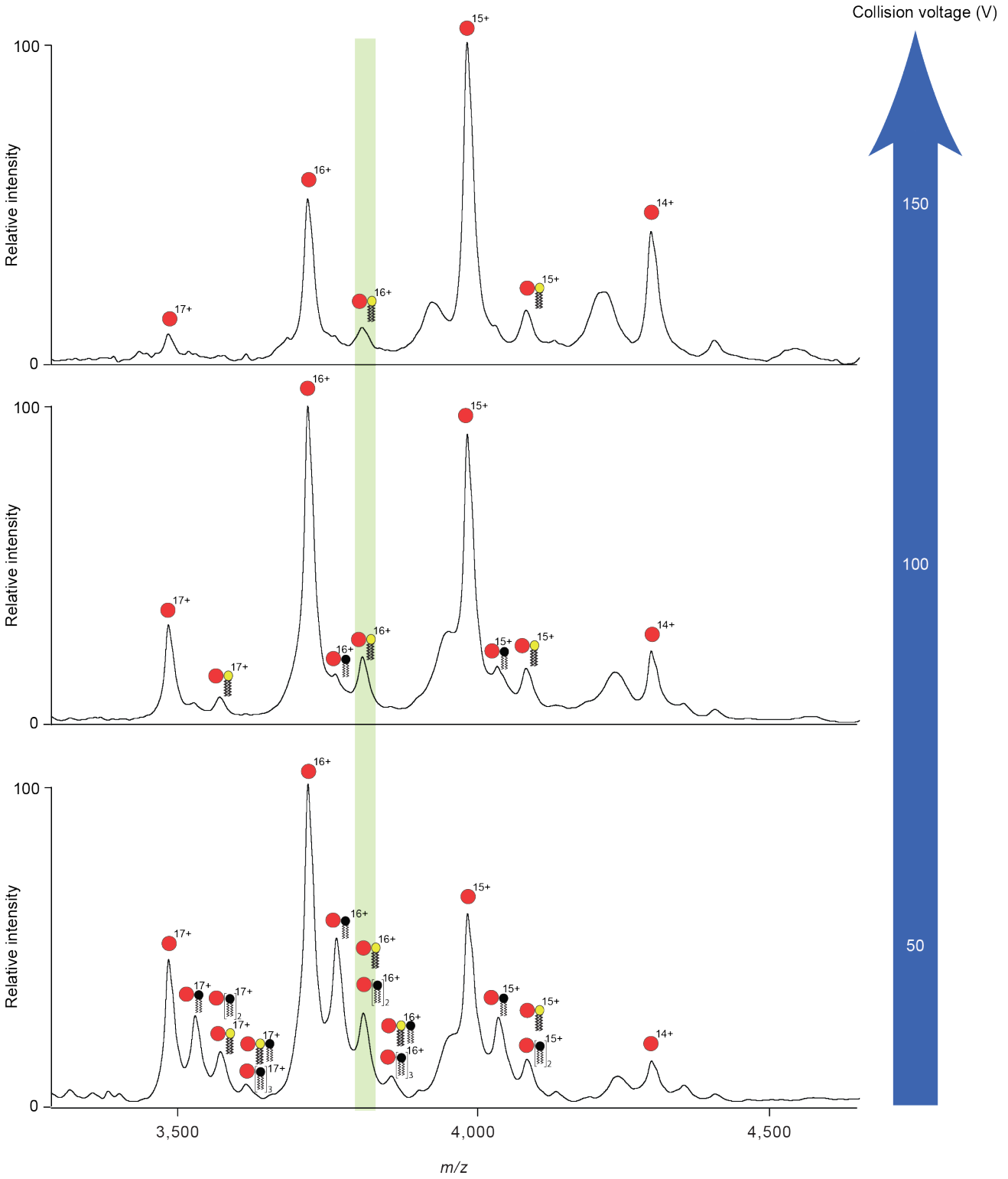


Figure 4.5: Mass spectrum of LeuT monomers (●, 59.3 kDa) following dissociation of LeuT dimer-lipid adduct (●●, 126.0 kDa) by HE-nMS at different trap collision voltages. At low collision voltage, the monomer is bound to phospholipids (, ≈ 750 Da) and cardiolipin (, ≈ 1,450 Da), but at higher collision voltages only cardiolipin binding is retained.

interface, consistent with the observation of lipid-like electron density in the dimer interface of the LeuT crystal structure. To test the hypothesis that lipids bind at the dimer interface, we used site-directed mutagenesis to generate a LeuT mutant in which the π-stacking interactions between tyrosine and phenylalanine residues (LeuT F488A/Y489A) have been deleted. These π-stacking interactions make up the only protein-protein interactions between the two subunits, and hence we hypothesised that the mutant would be unable to dimerise. Following expression and purification of the dimer interfaces mutant, we analysed it by MS. Figure 4.6 shows the mass spectrum of the F488A/Y489A mutant; only monomers without bound lipid are observed. This implies that lipid binding cannot occur without the dimer interface formed from the 2 LeuT subunits, and so lipid binding must be occurring at the dimer interface.

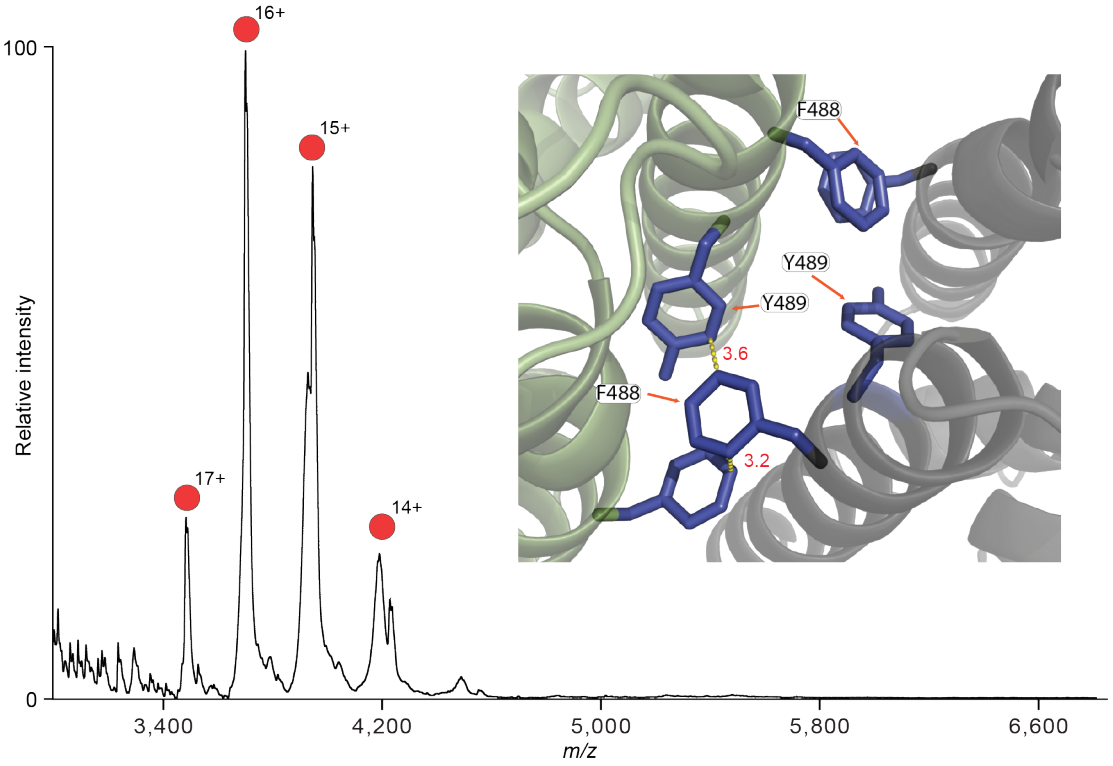


Figure 4.6: Mass spectrum of the LeuT mutant F488A/Y489A, showing LeuT monomers (●, 59.3 kDa) with no lipid bound. Inset shows the LeuT dimer interface residues involved in face-face π-stacking interactions (yellow dotted lines, distances in Ångstroms shown in red).

4.5 MD simulations reveal a cardiolipin binding mode which bridges the LeuT dimer interface

Using MS we had elucidated that there are lipids binding at the LeuT dimer interface and gained some idea of how many and which lipids might be involved. However, MS methods can not provide certain information on protein-lipid interactions such as the precise location of the binding site, which protein residues and which functional groups of the lipids are involved in binding, and the types chemistry involved. In order therefore to gain a better understanding of the details of protein-lipid interactions at the LeuT dimer interface, we turned to our collaborator Dr Phil Stansfeld, who used the recently developed MemProtMD software program to run Molecular Dynamics (MD) simulations on LeuT in a simulated membrane bilayer. The finer details of how this method works have been reported elsewhere[23], and the simulation parameters used in this work are covered in the Materials and Methods, but a brief overview is as follows.

The MemProtMD pipeline begins with an atomic resolution membrane protein structure, in this case the LeuT structure, which it orients using the MEMEMBED program so that the bilayer will form in the xy plane and transmembrane helices will be approximately parallel to the z axis[163]. The protein structure is then placed in a box with periodic boundary conditions, before being converted into a coarse-grained (CG) structure. CG simulations are used in this context since they are 2-3 orders of magnitude faster to run than simulations with atomistic structures. Amino acid residues and lipids are represented as virtual 'beads' in the CG structure, each of which corresponds to around 4 heavy atoms and is classified as either apolar, non-polar, polar or charged. Lipids are then added to the box, in a random orientation and in a ratio depending on the desired lipid composition of the bilayer, followed by water and sodium chloride. In this case, the lipids consisted of a 1:2:7 ratio of cardiolipin to phosphatidylglycerol (PG) to phosphatidylethanolamine (PE), since this is approximately the lipid composition of the *E. coli* membrane[164]. With all of the components now present in the box, the system is equilibrated for 100 ns at 323 K, using the MARTINI 2.1 force field. This allows the membrane

bilayer to form around the protein. Once the bilayer has formed, the simulation continues for a further 900 ns, and the resulting CG structure is converted back to atomistic detail to generate a final membrane-embedded structure.

The final 800 ns (post-equilibration) of the simulation can then be analysed to identify lipid binding sites. Identification of binding sites is carried out by analysing the relative frequency with which a CG lipid bead occupies a given region of the protein. Protein regions which are occupied by specific groups on particular lipids with a high frequency/probability are likely to be binding sites for the those lipids.

When the lipid binding site analysis was applied to LeuT, notable differences in the binding behaviour of the different lipids were detected, as shown in figure 4.7. For PE and PG, binding 'hotspots' were detected at locations evenly spread across the surface of the dimer structure, including in the dimer interface, implying that this lipid has no strong preference for binding in any particular site on LeuT. Similar behaviour was observed for PG, though binding at the interface was localised to a small region at the extracellular side of the structure. On the other hand, cardiolipin binding hotspots were highly localised to the dimer interface, implying that there is a well-defined specific binding site for cardiolipin in the interface.

Further analysis of cardiolipin binding in the CGMD simulations allowed us to elucidate the protein-lipid interactions formed between this lipid and LeuT. Figure 4.8a outlines the dimer interface binding site in greater detail by showing the sites where binding of cardiolipin: phosphate, glycerol and alkyl tails, occur. Furthermore, it was also possible to generate an atomistic structure of a representative lipid in the binding site from the coarse-grained structure, in order to visualise lipid binding on the level of individual amino acid residues, as shown in figure 4.8b. In this atomistic structure, a protein-lipid interaction involving the negatively charged phosphate groups of cardiolipin and positively charged residues (namely histidine, lysine and arginine) on the protein was clearly visible. Since each CL has 2 phosphate groups, a single CL molecule can form hydrogen bonding interactions with both protein subunits, forming a bridge between them which could therefore stabilise the dimer interface.

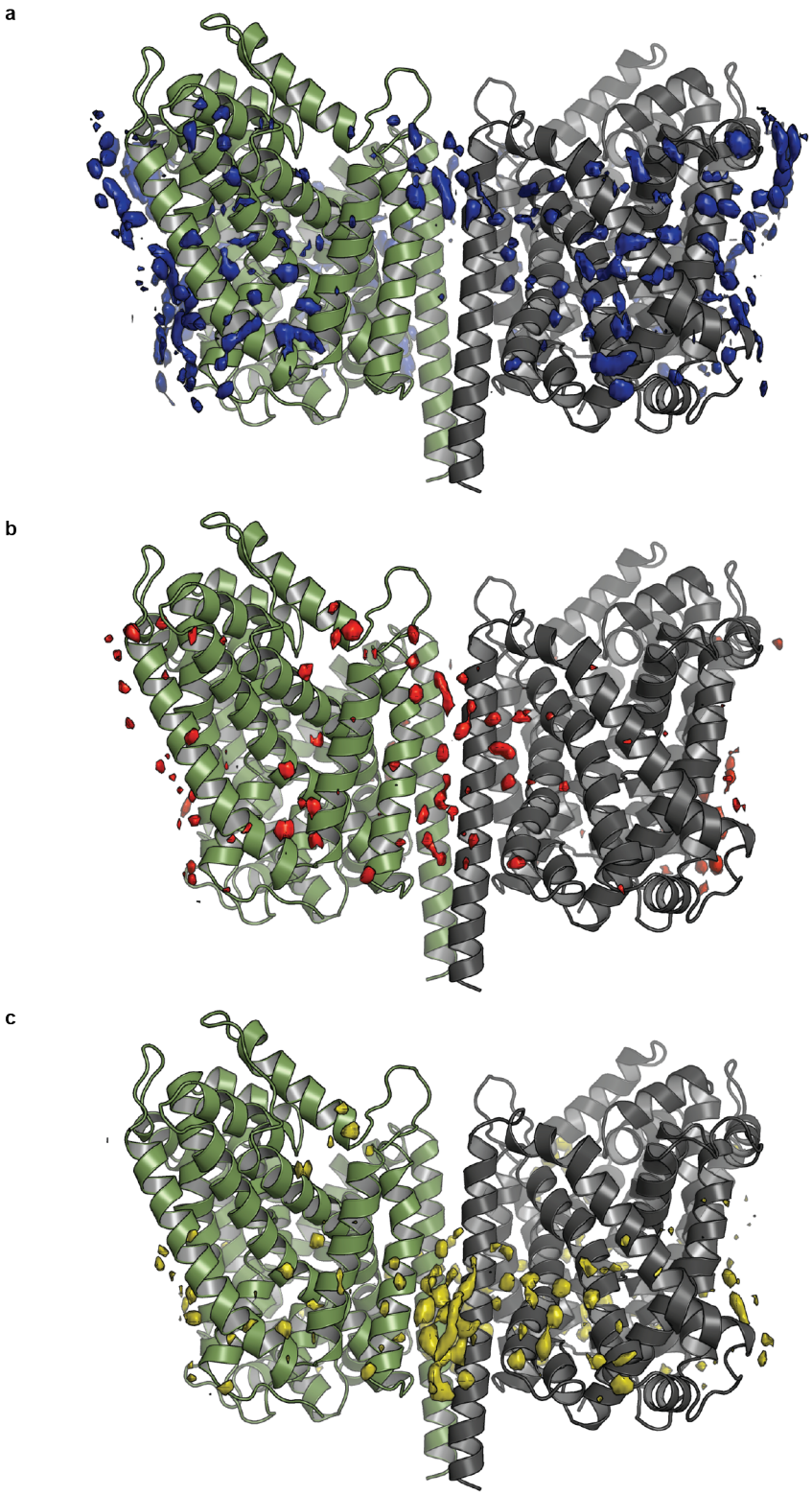


Figure 4.7: Lipid binding sites on LeuT, identified by coarse-grained molecular dynamics simulations. Surface densities represent the most occupied positions of phosphatidylglycerol (PG, **a**), phosphatidylethanolamine (PE, **b**) and cardiolipin (CL, **c**) across five repeats of 1-μs CGMD simulations.

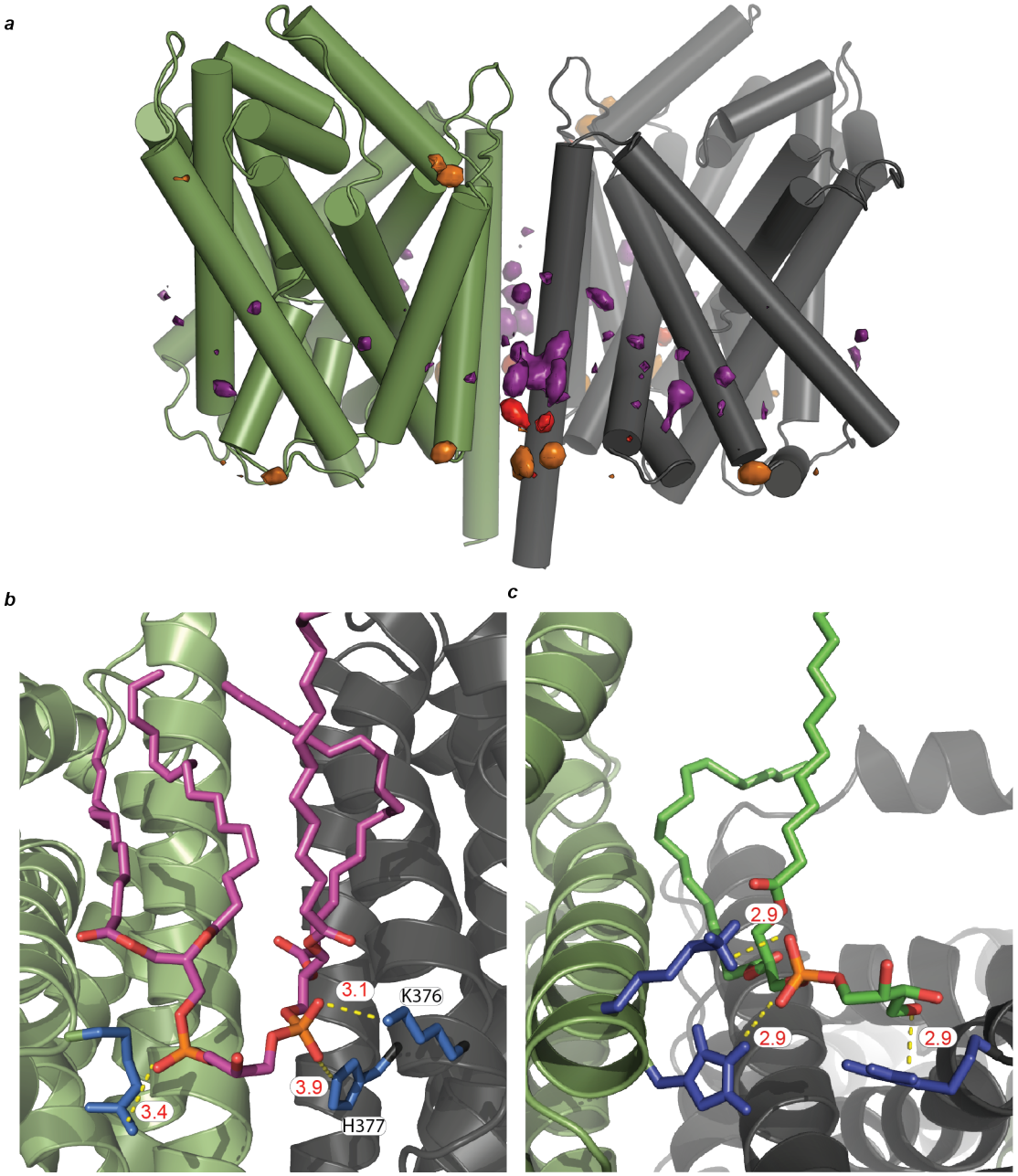


Figure 4.8: Molecular dynamics simulations reveal protein-lipid interactions which bridge the LeuT dimer interface. **a** Particle densities representing the most occupied sites of cardiolipin phosphate (orange), glycerol (red) and alkyl tails (purple) over 5 repeats of a 1-μs CGMD simulation show a clear binding site at the dimer interface. **b** All-atom snapshot of a representative cardiolipin molecule binding at the LeuT dimer interface. The two negatively charged cardiolipin phosphates are positioned favourably to interact with positively charged residues on the two LeuT subunits, potential H-bond interactions shown in yellow and distances (in Å) labelled in red, thereby bridging the interface. **c** LeuT dimer interface with aminopentanetetrol (APT, a lipid present in the host organism *A. aeolicus*) superimposed onto the position of an interfacial cardiolipin. Favourable distances between hydrogen bond acceptors on APT and the positively charged LeuT residues identified in the cardiolipin simulations show that APT could potentially form a similar bridging interaction across the interface.

Overall, these molecular dynamics results supported our hypothesis from the MS data that there are specific protein-lipid interactions occuring at the LeuT dimer interface. Moreover, by identifying specific cardiolipin binding sites, and the residues forming critical LeuT-cardiolipin interactions which bridge the dimer interface, we gained insight into the mechanism by which lipid binding might stabilise the LeuT dimer.

4.6 Cardiolipin binding is essential for LeuT dimerisation

The identification of a specific interaction between CL and positively charged residues in the dimer interface of LeuT provided us with an opportunity to test whether protein-lipid interactions in the interface are required for dimerisation of LeuT. We supposed that if this CL bridging interaction is essential for LeuT oligomerisation, then if we disrupt CL binding and measure the LeuT oligomeric state as before, we would observe that LeuT forms only monomers. To test this supposition, we chose two strategies to disrupt the CL binding interaction, site directed mutagenesis and expression in a CL-deficient *E. coli* cell line.

It has been shown that the final step in cardiolipin synthesis in *E. coli* involves a condensation reaction between two phospholipids which is catalysed by 1 of 3 different phospholipases. CL is not essential for *E. coli* survival, and hence it is possible to generate a viable *E. coli* cell line in which all 3 of the cardiolipin synthase enzymes have been knocked out. This triple knock-out strain, ΔclsABC, has a phospholipid membrane which contains no CL and therefore any biochemical structures or processes dependant on CL are disrupted. We expressed and purified LeuT in the ΔclsABC cell line, keeping the expression protocol as close to that for our standard LeuT expression as possible, and then analysed it by MS. Figure 4.10 shows the resulting mass spectrum which contains only monomeric LeuT, implying that in the absence of CL the LeuT dimer does not form, or, if it does, it is not stable enough to be maintained in the gas phase.

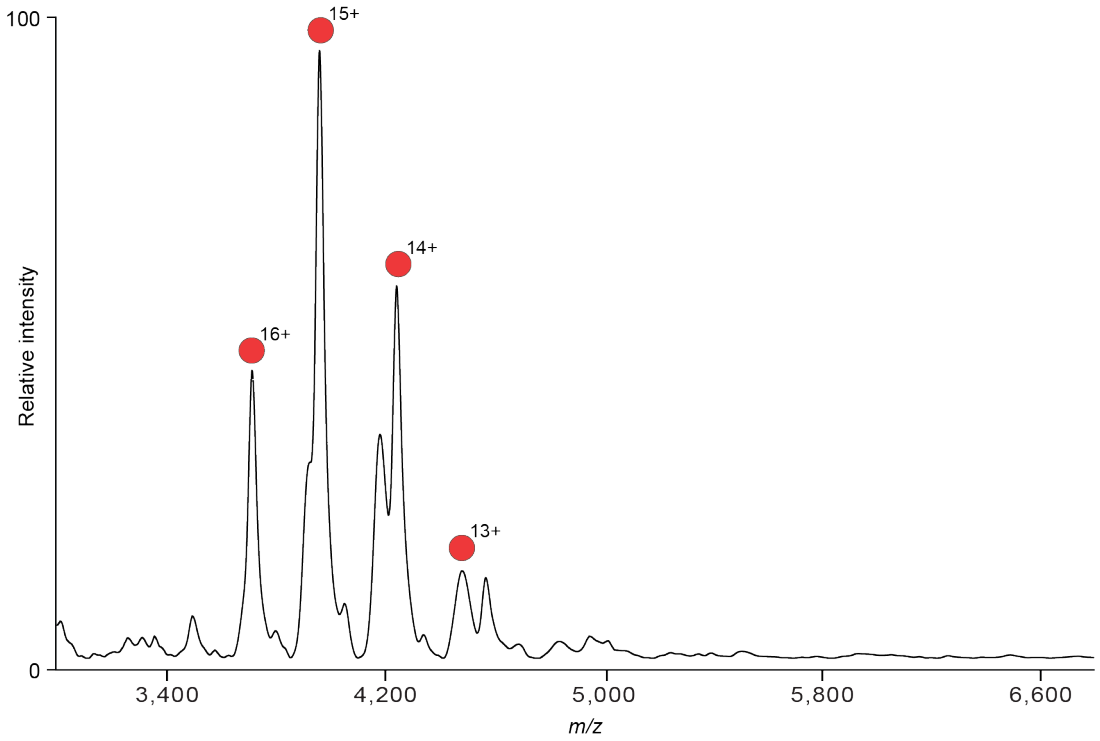


Figure 4.9: Mass spectrum of the LeuT mutant K376A/H377A, released from OG micelles. Mutation to alanine of two of the positively charged residues which are predicted to interact with cardiolipin at the dimer interface yields exclusively monomeric LeuT, implying that the cardiolipin binding interaction is required for LeuT dimerisation

We observed similar results when we used site-directed mutagenesis to delete the CL-LeuT interaction. Examining the structure of CL-bound LeuT generated by MemProtMD described above, we designed the LeuT mutant K376A/H377A in which a positively charged lysine and histidine, proposed to form ionic interactions with the CL phosphate groups, were replaced with alanine. Expressing and purifying this mutant also produced only monomeric LeuT, as deduced by MS and shown in Figure 4.9, again suggesting that LeuT cannot dimerise if it cannot form ionic interactions with the cardiolipin head group.

4.7 Delipidation and Relipidation of LeuT *in vitro* Highlights Lipid-Dependent Dimerisation

In addition, we carried out *in vitro* delipidation experiments to see if LeuT could be reversibly de-oligomerised when key lipids were removed. It has been reported

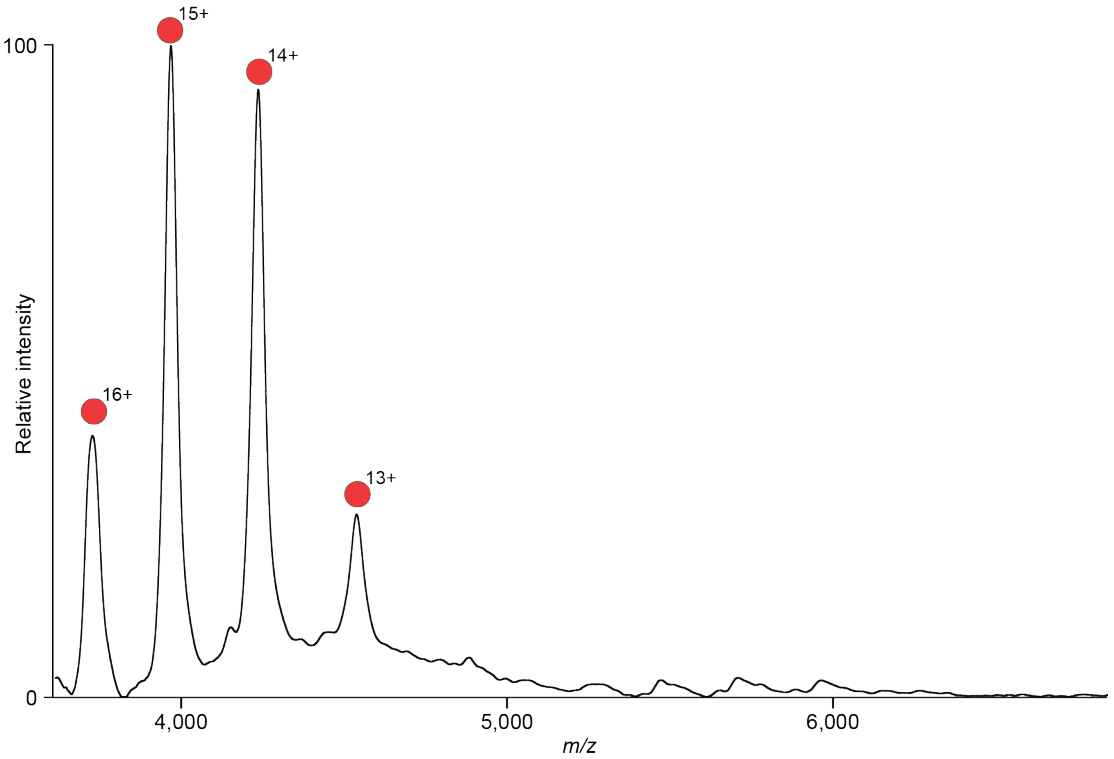


Figure 4.10: Mass spectrum of LeuT, expressed and purified from the cardiolipin-deficient BKT22 *E. coli* cell line and released from OG micelles, showing LeuT monomers (●, 59.3 kDa) with no lipid bound.

that solubilisation of membrane proteins in different detergents has varying effects on the extent to which the proteins are lipidated. The detergent NG is one of the most delipidating of those studied and so we incubated LeuT in NG overnight in order to delipidate the protein as much as possible. We then carried out an MS analysis on the NG solubilised LeuT (figure 4.11b), and on a sample obtained after a detergent exchange back to OG, using size-exclusion chromatography, in order to compare to our original LeuT spectrum in the same detergent (figure 4.11c). In both cases it can be seen that NG incubation results in a solution containing only delipidated monomers and no dimers. There is little qualitative difference between the OG- and NG-solubilised delipidated monomer spectra, apart from greater spectral quality in the OG case. To see if the dimer could be reformed upon the readdition of lipids, we incubated delipidated LeuT in OG, supplemented with *E. coli* polar lipids. Remarkably, when this sample was analysed by MS (figure 4.11d) we observed significant restoration of the dimer population. This provides

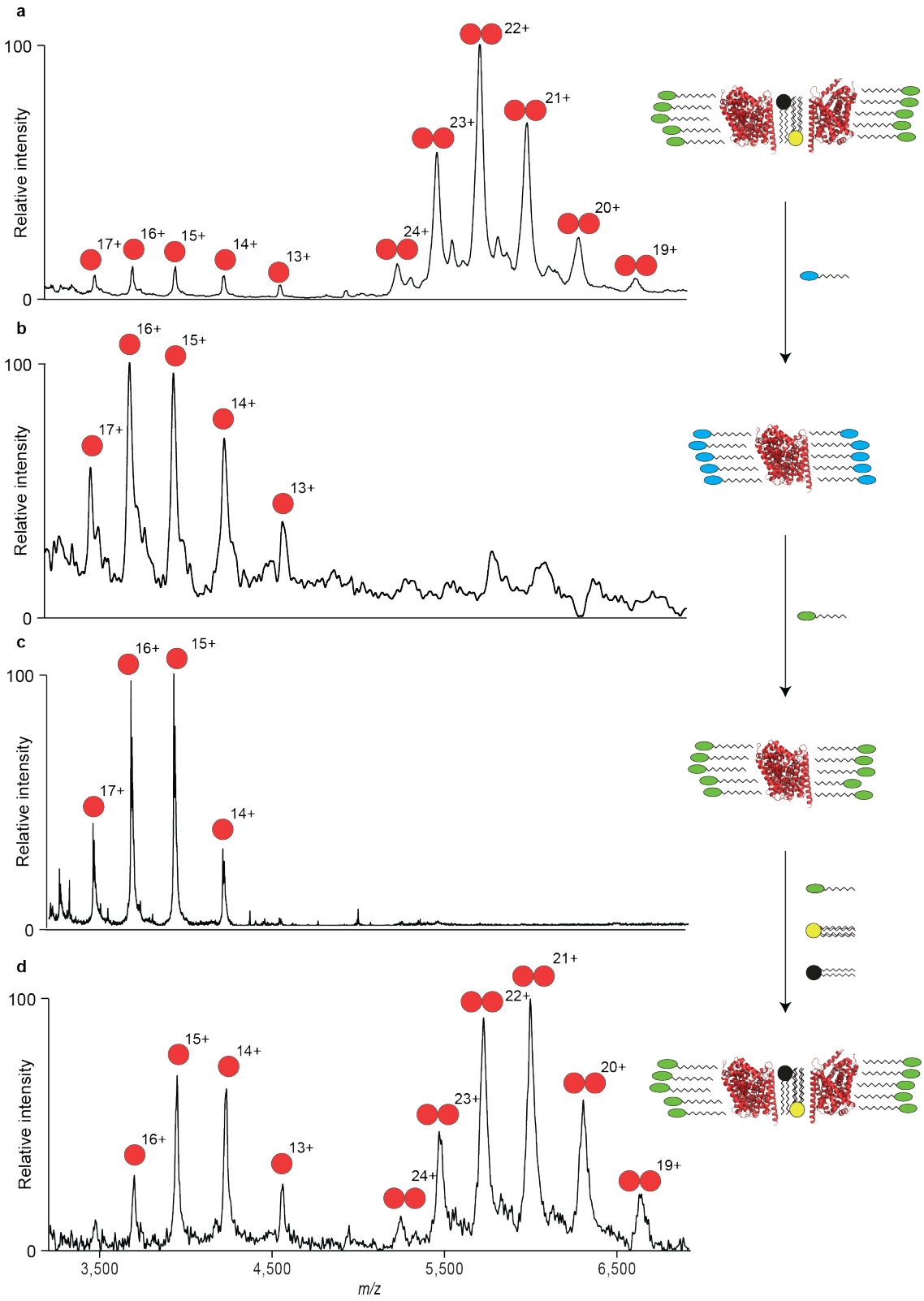


Figure 4.11: Mass spectra of LeuT following incubation with delipidating detergent and *E. coli* polar lipids. **a** Mass spectrum of LeuT released from OG micelles, showing lipid-bound dimer (●●, 126.0 kDa) and low abundance monomer (●, 59.3 kDa). **b** Mass spectrum of LeuT released from NG micelles after overnight incubation shows only delipidated monomers. **c** Mass spectrum of LeuT released from OG micelles after NG incubation. **d** Mass spectrum of LeuT released from OG micelles after incubation of delipidated monomers with *E. coli* polar lipids (,) shows restoration of the dimer population.

compelling evidence that lipid binding is essential for maintenance of the oligomeric state of the LeuT dimer. It is also evidence against a model of de-oligomerisation in which the individual LeuT subunits are simply unfolded or destabilised by NG solubilisation rather than dissociated, since the monomers would need to retain their overall structure in order to reform dimers once the lipids are re-added.

4.8 Discussion

In this chapter we have investigated the endogenous lipid binding behaviour of the bioamine transporter LeuT, thereby gaining insight into its local lipid environment. We found that LeuT is a dimer which co-purifies with a cohort of endogenous lipids that bind specifically to the dimeric state. Using a novel high energy MS method, we then demonstrated that this cohort includes up to 3 phospholipids (PE or PG) and 1 cardiolipin per LeuT monomer. Coarse-grained molecular dynamics simulations, supported by site directed mutagenesis experiments showed that these lipids bind at the dimer interface, and elucidated a specific bridging interaction between the two negatively charged phosphate groups of a single cardiolipin molecule and positively charged LeuT residues on both sides of the interface.

We also explored the relationship between LeuT lipid binding and its oligomeric state. Disruption of interfacial lipid binding by several different methods: chemical delipidation with detergent, expression in a CL-deficient cell line, and site directed mutagenesis of the lipid binding residues, resulted in LeuT being unable to dimerise, and hence we proposed that lipid binding at the dimer interface is required for LeuT oligomerisation.

These results are a further demonstration of the utility of MS for identifying endogenous lipids bound to purified membrane proteins, building on the evidence from the previous chapter on SemiSWEET. By elucidating the composition of the LeuT dimer adduct, we have identified lipids which are in direct contact with LeuT, as distinct from those in the bulk membrane. Furthermore, our finding that these lipids are important for the stability of the LeuT dimer elegantly illustrates that focusing on these 'structural' lipids allows us to gain insight

into which are lipids involved in key biological processes such as oligomerisation. By combining simulations and mutational studies, we were also able to extract positional information on lipid binding and obtain a residue-level picture of the lipid binding site.

Our analysis of LeuT revealed strikingly different lipid binding behaviour compared to SemiSWEET and most other membrane proteins that have been studied by our laboratory. For SemiSWEET we observed that the monomer and dimer bound to different numbers of lipids, both when we examined binding of co-purified endogenous lipids and exogenous lipids added to the protein. This pattern is consistent with what has been observed for endogenous lipid binding to other membrane proteins in previous studies In contrast, the LeuT dimer appears to bind a tightly defined homogenous cohort of lipids, represented in our original LeuT mass spectra by the 7.4 kDa mass shift. Dimeric LeuT was observed exclusively in this lipid binding state, and could not be detected by in the absence of these lipids. Our HE-nMS analysis showed that this cohort was made up of up to 3 phospholipids and 1 cardiolipin per monomer, and so it is surprising that we did not also observe LeuT dimers bound to different subsets of this lipid cohort. A possible explanation could be that the lipids bind in a cooperative manner: perhaps in the membrane there are LeuT dimers with a subset of the lipid cohort in the dimer interface, but, upon transfer to the gas phase only those dimers with the full lipid cohort can persist without dissociating into LeuT monomers and free lipids.

By identifying important structural lipids, the HE-nMS method could provide leads for future investigations into the interactions between lipids and proteins of interest. In a recent study from our laboratory, exogenous lipids were added to 3 purified membrane proteins in order to assess the effect on protein stability. Lipids which imparted significant stability were then used to direct crystallographic and functional studies on these proteins[130]. The study began with a blind screen of different lipids, and therefore involved a large number of experimental measurements. One possible way to improve the efficiency of this approach would be to add an initial HE-nMS step to the protocol, in which endogenous bound lipids are identified. The

identified lipids could then be candidates in a narrower more focused screen since they would already be known to interact strongly with the proteins. This approach could be of particular benefit when applied to proteins purified from host organisms with more complicated lipid compositions than *E. coli*, for example eukaryotes, where a blind screen would involve testing a large number of lipid candidates.

Turning to our findings on LeuT oligomerisation, we have demonstrated that lipid binding mediates the oligomeric state of LeuT, to the extent that protein-lipid interactions at the dimer interface are essential for the dimeric assembly to remain intact. This allows us to classify LeuT as a strong transient oligomer, since its oligomeric state is dependent on its molecular context i.e. its immediate lipid environment. As discussed in Chapter 1, there are several cases in the literature of membrane proteins in which the oligomeric state has been found to be influenced by the presence of specific lipids in their immediate environment[88;86;87]. In turn, there are a growing number of structures in which bound lipids can be resolved at the interface between subunits[72;81;57;126]. However there have been, up to now, no reported systems in which a link between these two phenomena, interfacial lipid binding and lipid-mediated oligomerisation has been explicitly demonstrated. Hence, this study provides important evidence for a mechanism of lipid-mediated oligomerisation in which specific lipids bind at oligomeric interfaces.

Assessing the functional significance of lipid-dependent oligomerisation for LeuT is challenging, in part because it is unclear what the functional benefits of LeuT dimer formation are. All of the components involved in the substrate transport cycle are contained within each monomeric unit, so it is unlikely that dimerisation is essential for function. It is possible that dimerisation could facilitate cooperativity between the two LeuT subunits, as seen for example for the dopamine transporter[165]. Given that we are now able to generate folded monomeric LeuT mutants, an interesting future line of enquiry might be to apply the experiments which have been used previously to measure LeuT transport activity (such as the scintillation proximity assay) to these monomeric constructs in order to see if there is an appreciable change in the 'per-monomer' transport rate[166;167]. Alternatively, one simple explanation

for why LeuT has evolved a dimeric quaternary structure could be that it provides increased stability against denaturation, which would be important for a protein from a thermophilic organism.

Oligomerisation has been shown to occur for many of the members of the NSS family, with examples including the serotonin transporter (SERT)[168], dopamine transporter (DAT)[169], and the GABA transporter (GAT)[170]. Evidence for functional implications of oligomer formation is less concrete, however. In general, it seems unlikely that oligomer formation is essential for the function of any of these proteins, but it could provide an opportunity for regulation of function if oligomerisation can be induced or inhibited by other cellular processes. One possible role for NSS oligomerisation which has been posited in the literature is that it could form part of a quality control process in protein synthesis/folding - for example, it has been shown that 'deoligomerisation' mutants of GAT are retained in the endoplasmic reticulum and not exported to the plasma membrane[171].

The case of SERT is particularly interesting, since recent studies have reported findings on SERT oligomerisation that are strikingly similar to those reported here for LeuT. SERT oligomers were detected *in vitro* almost 2 decades ago[168], and have subsequently been shown to form in live cells. In work published in 2014 and 2017[172;173], Anderluh et al found that SERT forms a distribution of different sized oligomers (from monomers up to pentamers) consistent with dynamic equilibration in the endoplasmic reticulum (ER) membrane. When SERT oligomers were measured in the plasma membrane, however, they find that the oligomeric distribution is fixed and independent of protein density. They propose that SERT oligomers pre-form in the ER and then are kinetically trapped as they enter the plasma membrane. This trapping is caused by binding of PIP2, a lipid which is abundant in the PM but absent in the ER membrane and which they suggest binds in the interface between protein subunits, via a bridging interaction between the anionic inositol rings on the PIP2 head group and positively charged residues on the protein surface. Hence, this study shows that lipid-mediated oligomerisation could be used as a mechanism for cells to set a protein's oligomeric state distribution independent of protein density

at the cell surface. Moreover it demonstrates that our work characterising lipid-mediated oligomerisation in LeuT is highly relevant to understanding regulation of quaternary structure of NSS family members in eukaryotic organisms.

Further evidence for the relevance of our LeuT data to eukaryotic NSSs came from structural studies on SERT, the crystal structure of which was published in 2016[174]. Whilst a dimeric structure could be formed from packing of two TM12 helices (the same helices that form the LeuT dimer interface) at the crystallographic interface, the authors were sceptical about whether this dimer would form in a 'real' membrane, noting also that they found SERT to be a monomer in detergent. The SERT structure contains a kink in TM12 which would prevent it from forming a dimer with an interface geometry like that of LeuT, though this could be a crystallographic artifact. Even so, later analyses[175], using molecular dynamics to assess the stability of different potential dimer structures showed that, even with the kink, TM12 dimers could be constructed which would be stable in the membrane.

Having identified LeuT as a system which requires protein-lipid interactions at its oligomeric interface in order to maintain its oligomeric state, we naturally went on to speculate on how general a phenomenon this might be. This question will be addressed in the general discussion in the next chapter.

4.9 Materials and Methods

4.9.1 Molecular cloning and plasmid construction

LeuT was expressed from a pET-16b vector, containing a thrombin-cleavable C-terminal 8 × His tag. The LeuT point mutants F488A/Y489A and K376A/H377A were generated using a Quikchange Lightning site-directed mutagenesis kit, according to the manufacturer's protocol. For expression in the cardiolipin-deficient BKT22 *E. coli* strain, the *LeuT* gene was amplified by PCR using a Phusion Flex Hot Start Polymerase (New England Biolabs) with primers designed for Infusion cloning (Clontech) using the manufacturer's online tool. The PCR product was purified using gel agarose electrophoresis, then used in an Infusion cloning reaction (Clontech) with a linearized pBAD vector containing a C-terminal 10 × His tag. The LeuT-eYFP

fusion protein was expressed from a pET-15b/pET-23b hybrid vector, containing the *LeuT* gene followed by a TEV protease cleavage site, the *eYFP* gene (with mutation A206K to abolish eYFP dimerisation) and a 6 × His tag. To construct this plasmid, the *LeuT* gene was amplified by PCR as above and used in an infusion reaction with a pET-15b/pET-23b hybrid vector[130], cut with appropriate restriction enzymes, and a synthetic gene block (Integrated DNA Techologies) containing the TEV site and eYFP gene. All plasmid constructs were confirmed by DNA sequencing (Source Bioscience).

4.9.2 Protein purification

C43 *E. coli* cells were transformed with the LeuT plasmid and wild-type LeuT was purified according to a previously published protocol[161]. In brief, multiple colonies were used to inoculate 100 ml of terrific broth (TB) and were grown overnight at 37 °C. A total of 10 ml of overnight culture was used to inoculate each of 6 l of TB, which were allowed to grow at 37 °C until the culture reached an A_{600} of 0.6. Isopropyl β-d-thiogalactopyranoside (IPTG) was added to a final concentration of 0.1 mM, and the cultures grown for 16 h at 20 °C. Cells were harvested by centrifugation at 5,000 g for 10 min at 4 °C, resuspended in lysis buffer solution (300 mM sodium chloride, 20 mM Tris; pH 7.4) supplemented with protease inhibitor cocktail (Roche) and lysed using a M-110 PS microfluidizer (Microfluidics), and the cell debris was pelleted by centrifugation at 20,000 g for 25 min at 4 °C. Membranes were pelleted by centrifugation of the supernatant at 100,000 g for 2 h at 4 °C and subsequently resuspended in ice-cold buffer solution (100 mM sodium chloride, 20 mM Tris, 20% glycerol; pH 8.0) and homogenized using a Potter-Elvehjem Teflon pestle and glass tube. n-dodecyl β-d-maltopyranoside (DDM) was added to resuspended membranes to a final concentration of 2% (w/v) and the suspension incubated with gentle agitation for 15 h at 4 °C. Insoluble material was pelleted by centrifugation at 20,000 g for 30 min and the supernatant filtered through 0.22-μm filters. LeuT was purified by immobilized-metal ion-affinity chromatography using a HisTrap HP 5 ml column (GE healthcare) equilibrated with buffer A (190 mM

sodium chloride, 10 mM potassium chloride, 20 mM Tris, 20 mM imidazole, 10% glycerol, 0.02% DDM; pH 8.0) and eluted with buffer B (190 mM sodium chloride, 10 mM potassium chloride, 20 mM Tris, 500 mM imidazole, 10% glycerol, 0.02% DDM; pH 8.0). The eluted protein was transferred to a dialysis cassette (100 kDa molecular weight cut-off) and dialysed against a dialysis buffer solution (190 mM sodium chloride, 10 mM potassium chloride, 20 mM Tris, 10% glycerol; pH 8.0) with 0.02% DDM overnight. A 100-kDa molecular weight cut-off (MWCO) concentrator was used to concentrate the dialysed protein. LeuT was then injected onto a Superdex 200 Increase GL 10/300 column (GE Healthcare), equilibrated in a buffer (190 mM sodium chloride, 10 mM potassium chloride, 20 mM Tris,10% glycerol; pH 8.0) with 1% octylglucoside. Peak fractions containing octylglucoside-solubilized LeuT were concentrated as above and either used for further study immediately or flash-frozen and stored at -80 °C. All protein concentration measurements were carried out using a UV/Vis spectrophotometer (DS-11 +, DeNovix).

F488A/Y489A and K376A/H377A LeuT mutants were purified as for the wild-type. For cardiolipin-deficient cell-line expression, BKTT22 cells[176] were transformed with the LeuT pBAD plasmid. Purification proceeded as for the wild-type, except that protein expression was induced at A_{600} of 0.6 by the addition of arabinose, to a final concentration of 0.02%.

4.9.3 Non-denaturing mass spectrometry

Samples were prepared for non-denaturing mass spectrometry by buffer-exchange into mass spectrometry buffer (200 mM ammonium acetate, 2 × critical micelle concentration of detergent of interest; pH 7.4) using a centrifugal buffer exchange device (Micro Bio-Spin, Biorad). The sample is then diluted to an oligomeric protein concentration of 10-20 μM, 3 μl of which is used for each acquisition.

LeuT mass spectrometry measurements were performed on a Synapt G1(Waters) with a Z-spray source, using nanoelectrospray capillaries prepared in-house[114]. The source pressure was set to 4-7 mBar, with a capillary voltage of 1.4-1.7 kV, capillary nanoflow of 0.05-0.2 mBar and argon as collision (trap) gas at flow rate

of 1.5-8.0 ml min^{-1}. Other parameters, including the sample and extraction cone and trap bias voltages, collision voltages and quadrupole profile were optimized for maximal ion intensity and minimal dissociation of the target membrane protein complex[113]. Data were processed using MassLynx software.

High-energy non-denaturing mass spectrometry

High energy source experiments were carried out using a Synapt G2 (Waters), modified to allow the use of higher voltages at the extraction cone in the source region of the instrument. This modification required implementation of the capability to drive the sample cone voltage from an external supply. This took the form of a patch cable introduced between the instrument lens control PCB and the source ion block. The ion block contains both heater elements and a thermocouple in addition to supporting the extraction cone and these were decoupled with the patch cable to prevent possible electrical breakdown with use of higher cone voltages. The extraction cone was patched directly through from the lens PCB while the sample cone voltage was decoupled and a new wire connection made to an external power supply. Finally, the configuration file for the extraction cone was modified to increase the maximum voltage setting from 10 V to 200 V.

HE-nMS experiments were then carried out as follows. Source pressure, capillary voltage, nanoflow and trap pressure were maintained as for standard non-denaturing MS experiments. Initially the trap and transfer collision voltages are set to 0 and the quadrupole profile set to transfer all ions. After the sample is loaded, and a stable spray established, the sample cone voltage is then increased until the detergent micelle has been fully dissociated, and well-resolved detergent-free charge states of the oligomer of interest are observed. The m/z region of interest is identified and the quadrupole profile is set to transmit only this m/z window. The trap and transfer collision voltages can then be increased to induce dissociation of the isolated complex.

4.9.4 Chemical delipidation

Purified, octylglucoside-solubilized LeuT was incubated in 2% nonylglucoside overnight at 4 °C. Subsequently, the sample was passed through a Superdex 200 Increase GL 10/300 column (GE Healthcare), equilibrated in 200 mM ammonium acetate with twice the critical micelle concentration of neopentyl glycol, in order to remove the excess octylglucoside and neopentyl glycol. The delipidated LeuT in neopentyl glycol was then re-exchanged back into octylglucoside using the above protocol with, the 200 mM ammonium acetate containing 1% octylglucoside. *E. coli* polar lipid stocks were made from powder (Avanti Polar Lipids Inc.) at a concentration of 10 mg ml^{-1},as described previously[113] and subsequently diluted 50 times in 200 mM ammonium acetate solution containing 1% octylglucoside.

4.9.5 Molecular dynamics simulations

All molecular dynamics simulations were performed using GROMACS v5.1.2[177]. The MemProtMD pipeline[23] was used with the Martini 2.2 force field[178] to run five repeats of a 1-μs CGMD simulation of the dimeric LeuT complex, taken from PDB entry 2A65. The last 800 ns of each of these simulation trajectories were considered for further analysis. The protein was centred within the simulation system to permit the assembly and equilibration of a membrane bilayer consisting of 10% cardiolipin, 20% 1-palmitoly,2-oleoyl,phosphatidylglycerol (POPG) and 70% 1-palmitoyl,2-oleoyl, phosphatidylethanolamine (POPE). Systems were neutralised with a 150 mM comncentration of NaCl. All simulations were performed at 323 K, with protein, lipids andf solved separately coupled to an external bath using the velocity-rescale thermostat[179]. Pressure was maintained at 1 bar with a semi-isotropic compressibility of 5×10^{-6} using the Berendsen barostat[180]. All bonds were constrained with the P-LINCS algorithm[181]. Electrostatics were measured using the 'reaction field' method,[182], while a Verlet cut-off scheme to permit GPU calculation of non-bonded contacts was used for Lennard-Jones parameters[183]. Simulations were performed with an integration time-step of 20 fs. Atomistic snapshots at the end-point of the CGMD simulations were created

using CG2AT[184] in combination with Alchembed[185]. The lipid densities and contacts with the protein during the molecular dynamics simulations were calculated using MDAnalysis[186], and locally written code. All images and animations were generated using Pymol[187].

4.9.6 Contributions

- All molecular cloning and site-directed mutagenesis experiments was carried out by JACD.
- Protein purification was carried out by JACD and Kallol Gupta.
- Standard non-denaturing MS experiments were performed by JACD with the assistance of Kallol Gupta.
- HE-nMS experiments were carried out by Kallol Gupta with assistance from Jonathan Hopper.
- Molecular dynamics simulations were performed by Phil Stansfeld.

5 Discussion and Further Work

Contents

5.1 Probing the local lipid environment 124
5.2 Characterising bound lipids 125
5.3 Measuring oligomeric states 128
5.3.1 Characterising oligomerisation dynamics 129
5.4 How many other oligomeric protein structures are dependant on protein-lipid interactions? 130
5.5 Next steps - transient oligomers 136

In this chapter we will draw upon the results and conclusions from chapters 3 and 4, and summarise the main conclusions of this thesis. The biological significance of the results in terms of LeuT and SemiSWEET biology have been explored at the end of their respective chapters, so here we will focus on the research questions outlined in chapter 1 and from a methodological perspective discuss the extent to which the methodology described in chapter 2 is suitable to address these research questions for membrane proteins more generally. This will include highlighting the limitations of our approach, and noting how some of these limitations have been overcome with more recent developments of the method.

The ultimate objective of this investigation was to understand the role of lipid binding and protein-lipid interactions in determining the oligomeric state of

membrane proteins. In particular, we noted in our introduction that the relationship between these two phenomena might be particularly significant in cases where membrane protein oligomers are transient, either because they adopt a different oligomeric state in different molecular contexts, or because they exist in a dynamic equilibrium between two oligomeric states. In order to identify systems for which protein-lipid interactions influence oligomeric state, we proposed a number of key steps in an ideal experimental characterisation. First, we proposed, we must be able to probe the lipid environment around a membrane protein oligomer, identifying structural and annular lipids, characterising them, identifying where they bind, and distinguishing their properties from those of the lipids in the membrane more generally. Next, we noted that the oligomeric state of the membrane protein must be determined in an environment as close to that of the native environment as possible, and any oligomerisation dynamics observed and characterised. Finally, the local lipid environment, or lipid-bound state, of the membrane protein should be perturbed/disrupted in order to understand the relationship between the lipid environment and the oligomeric state.

5.1 Probing the local lipid environment

For both LeuT and SemiSWEET, we were able to detect the presence of tightly bound lipids which remained attached to the membrane protein, even after extensive purification steps and transmission through the vacuum inside the mass spectrometer. For LeuT, this was a cohort of phospholipids and cardiolipins of fixed stoichiometry, attached to the dimeric state only. In contrast, for SemiSWEET we observed a heterogeneous distribution of lipid binding to the monomer. It is worth highlighting that the ability of non-denaturing MS to focus on a specific subset of membrane lipids which are in direct contact with a membrane protein of interest is shared by very few other structural biology methods.

A recurring challenge in interpretation these membrane protein mass spectra is relating the lipids in the mass spectra to the standard modes of lipid binding i.e. structural, annular and bulk. In the LeuT case, at least some of the lipids in the 7.4

kDa cohort are very likely to be structural, since they bind in a single stoichiometry and we could also localise the binding to the dimer interface using site-directed mutagenesis followed by further MS analysis. This hypothesis is also supported by the molecular dynamics simulations which highlight well-defined binding sites in the interface. In contrast, the SemiSWEET lipids are not detected binding with a single stoichiometry, and we have no evidence for a specific binding site, either at or away from the dimer interface. Molecular dynamics of SemiSWEET in a model bilayer also did not suggest a strong preference for lipid binding in any particular site or pocket, rather there were small lower-affinity hotspots distributed across the surface of the dimer. These observations perhaps suggest that the phospholipids and cardiolipin observed in the mass spectra are merely representatives of lipids in an annulus around SemiSWEET.

5.2 Characterising bound lipids

With the two modifications to the standard membrane protein MS method, in this study we overcame some of the limitations traditionally faced in analysing membrane protein mass spectra. Using the HE-nMS methodology, we were able to systematically dissociate multi-component protein-lipid complexes and divide the constituent lipids into broad classes, diacylphospholipids and cardiolipins. This tandem MS approach also permitted characterisation of spectra which contained species with overlapping m/z values, such as a protein bound to 2 diacylphospholipids or 1 cardiolipin which were previously unresolvable. One limitation of the HE-nMS method is that, in this case, we were unable to obtain more detailed information on the lipid fine structure, such as head group class for the diacylphospholipids, and acyl chain length and position of unsaturation. This is a practical limitation of the current experimental setup, rather than a theoretical one, since in principle it ought to have been possible with a HE-nMS-adapted instrument with two collision cells to: remove the micelle in the source; dissociate the LeuT dimer in the first collision cell; and then dissociate the lipid-bound monomer in the second collision cell. This would permit accurate mass measurement of the LeuT-associated lipids

in the low *m/z* region. As discussed in chapter 4, however, at the time this was not possible because of the challenge of establishing different pressure regimes in different regions of the instrument in order to facilitate transfer of both the high molecular weight protein-lipid complexes and low molecular weight free lipids. It is therefore impossible to say to what extent the properties of these structural lipids differ from the bulk lipids, and therefore how selective the protein-lipid interactions are which mediate LeuT oligomerisation. For SemiSWEET, in contrast we were able to carry out more extensive characterisation of one protein-lipid interaction, taking advantage of the relatively low mass of the semisweet monomer-cardiolipin complex, and high resolution of the orbitrap mass analyser. We used these to obtain a chain-length distribution for the lipids bound directly to SemiSWEET, and could therefore show that this one are delivered in had a chain-link distribution distinct from that of the rest of the bulk lipids.

One possibility for obtaining a full characterisation of the different layers of lipids around membrane protein, might be to combine the HE-nMS method, which works well for characterising small numbers of structural lipids, with more conventional lipidomics methods for identifying bulk and annular lipids. As outlined briefly in chapter 2, recent work published in our lab has established that, by carefully controlled incubation with delipidating detergent, it is possible to gradually displace increasing numbers of lipids from around a membrane protein complex. The remaining lipids can then be characterised by chemical extraction and then LC-MS/MS and the composition of the different lipid 'layers' can be determined. With this method, one can begin to build a picture of the lipid annulus, as distinct from the bulk lipids not in direct contact with the protein. The combination of these approaches has been recently incorporated into a protocol, published by Gupta *et al.* in 2018[135], which uses HE-nMS to probe structural lipids and sequential delipidation combined with lipidomics to characterise annular and bulk lipids. Testing the protocol on the ABC transporter MsbA, Gupta and colleagues showed that the ratio of PG to PE lipids increased as the dimeric complex was successively delipidated, with the predominantly delipidated complex being highly

enriched in PG (possibly suggesting a preference for anionic lipid close to the protein surface). However, tuning the process in order to leave a specific level of delipidation is a non-trivial and empirical process that is dependent on the properties of the specific protein, lipid, and detergents.

Some of the above issues with the HE-nMS method have been addressed in improvements in MS methodology in the short time since this work was carried out. In particular, it is now possible to combine the high energy source modification and the high resolution orbitrap analyser in a single instrument to allow a high resolution characterisation of lipids associated from protein lipid complexes. This was achieved on a Q-Exactive Plus which has been recently modified in our lab. The new instrument has a very similar architecture to the Q-Exactive instrument described in Chapter 2, but it is capable of accessing much higher voltages in the source region. These higher voltages mean that the detergent micelle can be dissociated from membrane protein complexes, and a particular charge state and lipid-bound state of the protein can be selected in the quadrupole. This complex can then be further activated in the HCD cell in order to dissociate the lipids. The key difference between this instrument and the modified high-energy Synapt used in this thesis is that the new instrument has a pressure regime which allows efficient transmission of the dissociated lipids and an orbitrap analyser to get highly accurate mass measurements for lipid identification. This approach has been applied to a number of systems, including an elegant study in which Yen *et al.* were able to identify endogenous lipids bound to the GPCR β-1-AR[188]. The authors found that a lipid-bound charge state of the monomer could be isolated and dissociated, and shown to be composed of β-1AR populations bound to PS and PIP. The ability to measure the dissociated lipid masses at high resolution also permitted assignment of the chain length and extent of unsaturation of these lipids as well as their class, all in one experiment.

5.3 Measuring oligomeric states

As outlined in chapter 2, one of the key strengths of non-denaturing mass spectrometry is that it can be used to determine the oligomeric states of non-covalent membrane protein complexes. In this context, this allowed us to detect LeuT and SemiSWEET dimers and monomers and hence measure the oligomeric states of these proteins under different conditions. The fact that MS is a single molecule method is important for these experiments, as it means that the abundance of different oligomeric states in a heterogenous mixture can be quantified in a single experiment. Furthermore, the distribution of oligomeric states in a sample can be followed over time. These properties were particularly important in this study focusing on transient oligomers. For LeuT, it allowed us to detect separate monomer and dimer populations, and see that only the monomer population was able to interact with lipids. And, for SemiSWEET, it allowed us to detect subunit exchange of monomers and hence infer that SemiSWEET exists in a dynamic monomer-dimer equilibrium.

As a caveat, it should be added that the relative abundance of different oligomeric states can be quite sensitive to the particular parameters used in any MS experiment, particularly as one seeks to find the right balance between conditions which result in clean, interpretable spectra and those which maintain the structure of the membrane protein in a form which best represents biology. It is not fully understood the extent to which the oligomeric state distribution in gas phase represents that of the protein in solution, and the extent to which the detergent-solubilised solution conditions mirror that of a native biological membrane. This means that, whilst drawing qualitative conclusions like the presence/absence of a particular stoichiometry or general trends such as increasing monomer:dimer ratio with lipid concentration may be valid, absolute values of measurements such as relative oligomeric state abundances or subunit exchange rates must be taken with a pinch of salt. We can mitigate these issues by aiming to work with reproducible parameters and conditions, so that we are comparing 'like-with-like' in our experiments. For example when considering the effect of a mutation, or different lipid composition on the oligomeric state, we can use identical parameters

such as protein concentration, and MS conditions such as voltages, pressures etc. in order to reduce the variability of the results.

One method by which these mass spectrometry results could be better validated, is by corroboration with other experimental methods, in particular those in which the membrane mimic better represents the native virtual membrane. For example one avenue which we pursued, but which was limited by time constraints, was to use step-wise photo bleaching of fluorescently-tagged LeuT expressed in *E. coli* and observed in live cells by fluorescence microscopy. Given more time, this would have allowed quantification of the distribution of LeuT oligomeric states *in vivo* and could have provided corroboration for our mass spectrometry results. Another option for relating gas-phase oligomeric state distributions to solution-phase distributions would be to use size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS)[39]. SEC-MALS is a low-resolution method, but it allows different oligomeric species to be separated in solution (using SEC) and then their approximate molecular weight and abundance measured based on how much they scatter light. This approach could be used to calibrate gas-phase measurements, since the measurements can be carried out using exactly the same buffers as are used in the MS experiments.

5.3.1 Characterising oligomerisation dynamics

The potential of our method to observe SemiSWEET subunit exchange in real time is an exciting one, and is the first reported example of membrane protein oligomerisation dynamics being followed by MS. It also opens up the possibility of quantitative characterisation of oligomerisation kinetics, which would allow comparison of the rates of dissociation and association of protein subunits in different contexts. However, there are a number of issues which must be addressed before reliable and biologically representative measurements are obtainable. First, optimisation of the protocol is required to make the measurements as reproducible as possible. In this study, samples are mixed manually by pipetting and then loaded into the electrospray capillary as quickly as possible. This inevitably introduces

some 'dead-time' into the measurements, since some exchange of subunits will occur before the MS measurement begins. Future method development could focus on optimising this process, for example by incorporating a stopped-flow system and a electrospray ionisation with an autosampler in order to improve the reproducibility of the mixing process and significantly reduce the dead-time. More fundamental, though, is that the mechanism of subunit exchange for membrane protein complexes in detergent micelles remains unclear. It is clear from our measurements that dimers are able to dissociate and reform, but, since the individual protein molecules are not all diffusing within the same membrane, there will be a contribution to the exchange rate from steps involving micelle dynamics, the physical chemistry of which is complex. This is an area which requires further mechanistic study, and calibration of MS measurements against kinetic measurements of oligomerisation by other methods, such as the *in vivo* fluorescence methods discussed in chapter 1.

5.4 How many other oligomeric protein structures are dependant on protein-lipid interactions?

In this thesis we have identified proteins which belong to the category of transient oligomers. Unlike most of the membrane protein systems which have previously been studied by native MS, which have been measured predominantly in a single oligomeric state[112], LeuT is dimeric in the presence of a specific lipid cohort, but monomeric when these lipids are absent, and SemiSWEET exists in a dynamic equilibrium between monomeric and dimeric forms. This raises the question as to why these proteins have this different oligomerisation behaviour, and whether one can predict which other proteins might behave in a similar way.

One simple hypothesis is that the dependence of the oligomeric state on protein-lipid interactions depends mainly on the relative strength of the protein-protein interactions between the different subunits. In simple terms, the rationale for this is that if there are already strong and numerous interactions between the protein subunits in the membrane, then there is no requirement for lipid binding to strengthen the oligomeric assembly. In contrast, when protein-protein interactions

are relatively weak, protein interactions can be used to reinforce the oligomer. Indeed, it was partly the observation that LeuT has very few protein-protein contacts at the dimer interface that motivated our initial investigation into its oligomerisation.

Given that there is now a large volume of structural data available in the PDB, it is possible to get an impression of how the strength of protein-protein interactions within oligomeric membrane proteins varies across the explored structural space. Deciding how to predict the strengths of such interactions is non-trivial, however, since the interactions between subunits will be a combination of polar or charge-based interactions like salt bridges and non-polar packing interactions between hydrophobic site chains. Quantifying the relative contributions of these different interactions to the stability of the oligomer is therefore highly complicated. Software does exist for predicting the stability of soluble protein assemblies, for example the PISA webserver[41], which calculates various parameters for known and hypothetical oligomers and combines these to calculate a dissociation energy (ΔG_{dissoc}). The ΔG_{dissoc} calculation is not applicable to membrane proteins, but it is possible to use some measures which can easily be extracted in a relatively high-throughput manner by the PISA software as a proxy for oligomeric stability. For example, one recent analysis carried out by Kallol Gupta in the Robinson Lab took buried surface area and number of salt bridges within oligomeric interfaces as proxy measures for oligomeric stability[189], and used PISA to calculate these parameters for a representative sample of all $\alpha-$helical membrane proteins of known structure. Grouping these by oligomeric state, one can plot the oligomeric strengths as shown in figure 5.1.

From the figure, it can be seen that the oligomers plotted span a wide range of buried surface areas. The plot also highlights 10 proteins for which non-denatured mass spectra had been recorded before this study. These generally lie towards the upper end of the buried surface area range, and it has been observed that their biological oligomeric states are stable in the gas phase in the absence of bound lipids. In contrast, LeuT has one of the lowest buried surface areas of the sample, and SemiSWEET lies roughly in the middle of the range. This picture

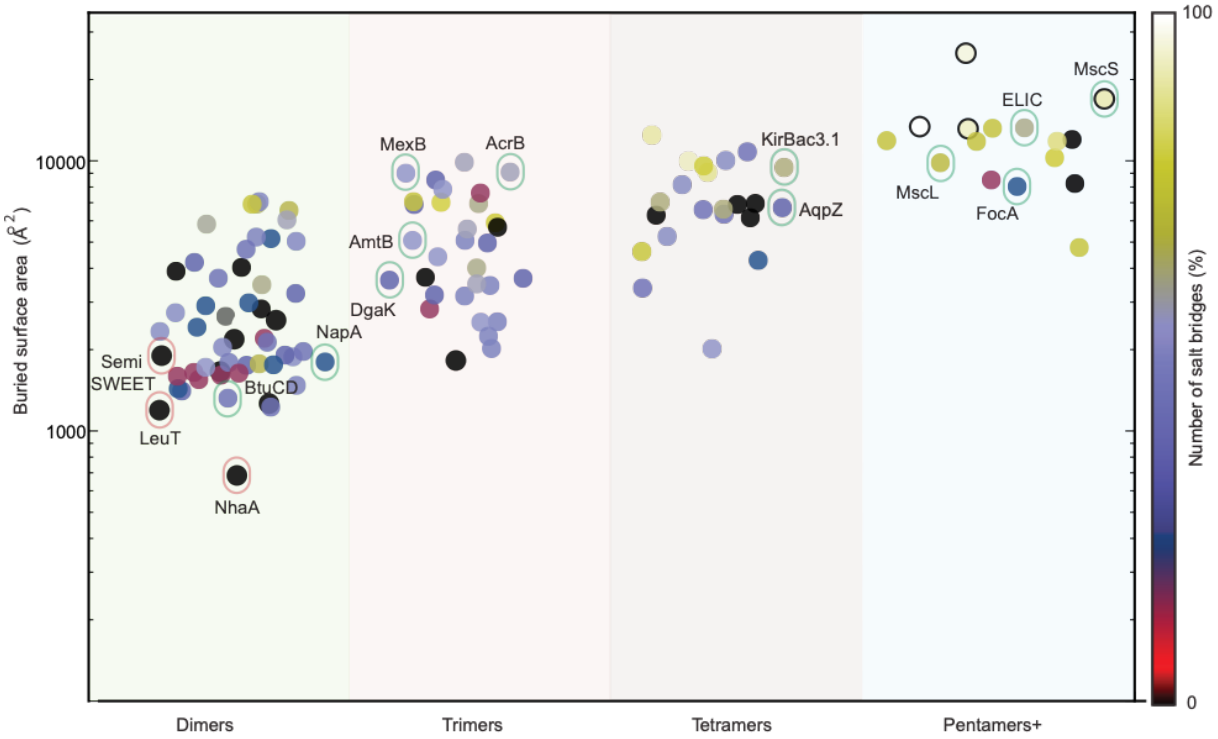


Figure 5.1: Buried surface area and number of salt bridges within representative oligomeric α-helical membrane proteins across the PDB. Protein oligomers are represented by circles colour-coded according to the number of salt bridges they form, where maximum number of salt bridges is designated to 100% and 0 is marked as 0%, and are grouped by oligomeric state (pentamers+ have an oligomeric state ≥ 5). A random horizontal jitter has been applied to all points to aid visualisation. Proteins for which mass spectra have been recorded in the literature are labelled. Figure adapted from Gupta *et al.*[189]

is broadly consistent with the LeuT dimer having a strict requirement for lipids and SemiSWEET having a looser dependence, where lipid binding can favour dimerisation but is not required.

Shortly after the conclusion of the work reported in this thesis, and the generation of the plot in figure 5.1, Gupta and colleagues in the Robinson group tested the model by analysing a pair of related proteins, the Na^+/H^+ antiporters NhaA and NapA from *E. coli* and *T. thermophilus* respectively. These proteins have broadly the same fold, the same function and were purified from the same type of membrane, but are predicted to have different oligomeric stability according to the plot in figure 5.1. They also have markedly different mass spectra, as shown in figure 5.2. The NapA spectrum clearly shows that the dimer is stable in the absence of lipids, with a prominent apo-dimer charge state distribution dominating the spectrum.

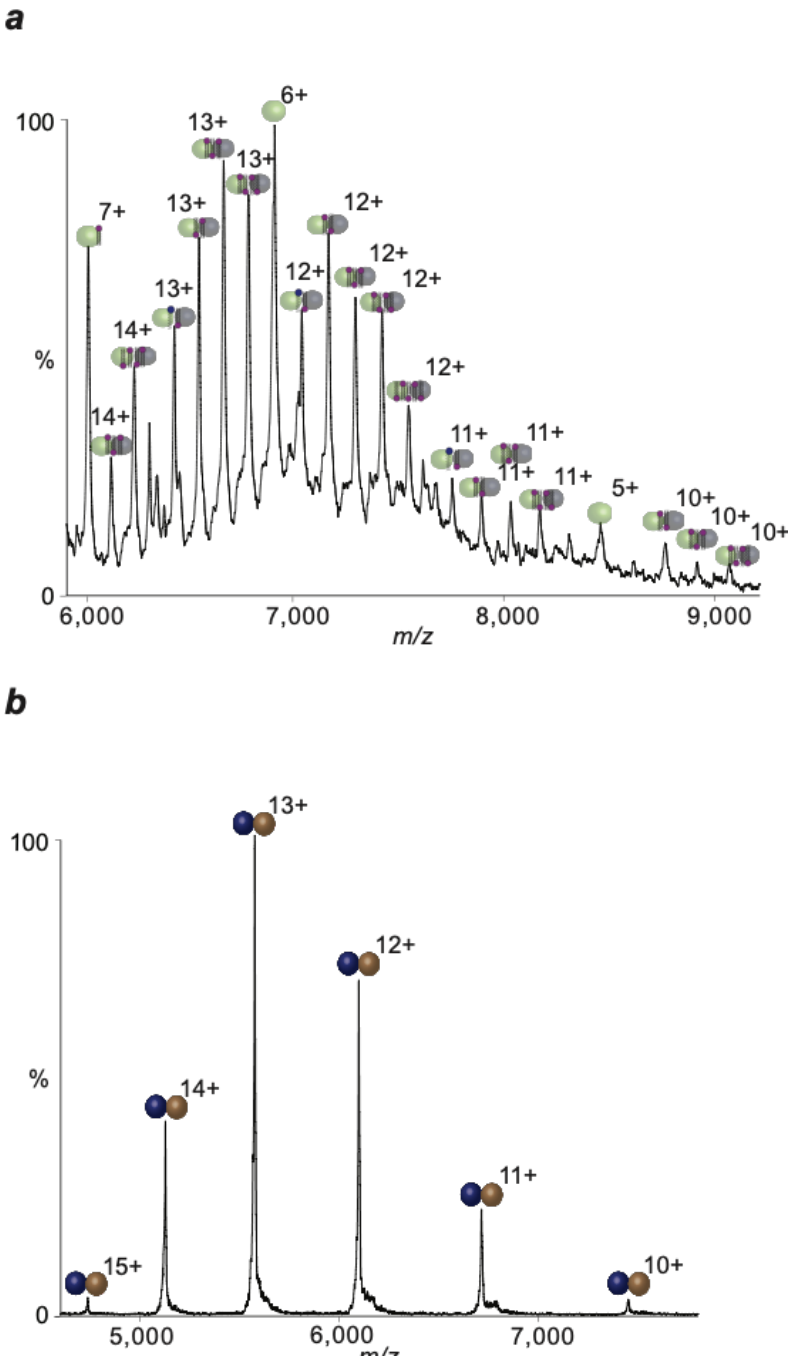


Figure 5.2: Mass spectra of NhaA and NapA Na^+/H^+ antiporters, released from C_8E_4 micelles. **a** Mass spectrum of NhaA shows monomers (green) and dimers (green/black) bound a heterogenous distribution of phospholipids (blue headgroups) and cardiolipins (red headgroups). **b** Mass spectrum of NapA shows a single distribution of dimers (blue/brown), without any bound lipids. Figure adapted from Gupta *et al.*[189]

In contrast, the NhaA spectrum is much more complicated, with a heterogeneous distribution of dimer states bound to different numbers of phospholipids and cardiolipin. Interestingly, whilst there were no apo-dimer peaks in the mass spectrum generated by the standard membrane protein MS analysis, when the HE-nMS method was applied to select and dissociate a cardiolipin-bound NhaA dimer, it was possible to find conditions in which a lipid-free dimer could be transmitted stably, see figure 5.3. This observation is inconsistent with the predictions from the buried surface area plot, since NhaA is predicted to have the least strong interface based on protein-protein interactions alone, and yet the mass spectrum suggests that it is better able to persist in the gas phase than the LeuT dimer.

In many respects, the fact that the predictions made by the buried surface area / salt bridge model are imperfect is unremarkable. There are several problems with this approach as a predictive tool, however. Firstly, the use of buried surface area and salt bridges is rather crude as a proxy for oligomeric stability, and does not take into account other thermodynamic properties such as solvation energies and the specific contributions of other interactions such as π-stacking or hydrogen bonds. Development of a rigorous and robust method for calculating the stability of oligomeric membrane proteins relative to their monomeric constituents is beyond the scope of this study, but it would be interesting to see how well such a method could predict requirement for lipids to stabilise oligomeric proteins. In addition, the inputs to this analysis are also dependent on X-ray crystal structures, which as we have discussed extensively in earlier chapters do not necessarily accurately reflect the structures of these proteins in real biological membranes. It is also worth pointing out that examples of weak oligomers which are likely to require lipid stabilisation are also those which are likely to be missing from this plot, since they are the most challenging to detect experimentally, and may have been mis-assigned as monomeric and therefore not included here.

In summary, we can see that, whilst there may be a general trend that membrane protein oligomers with smaller subunit interfaces have a greater tendency to have structures which are supported by protein-lipid interactions, oligomers with fairly

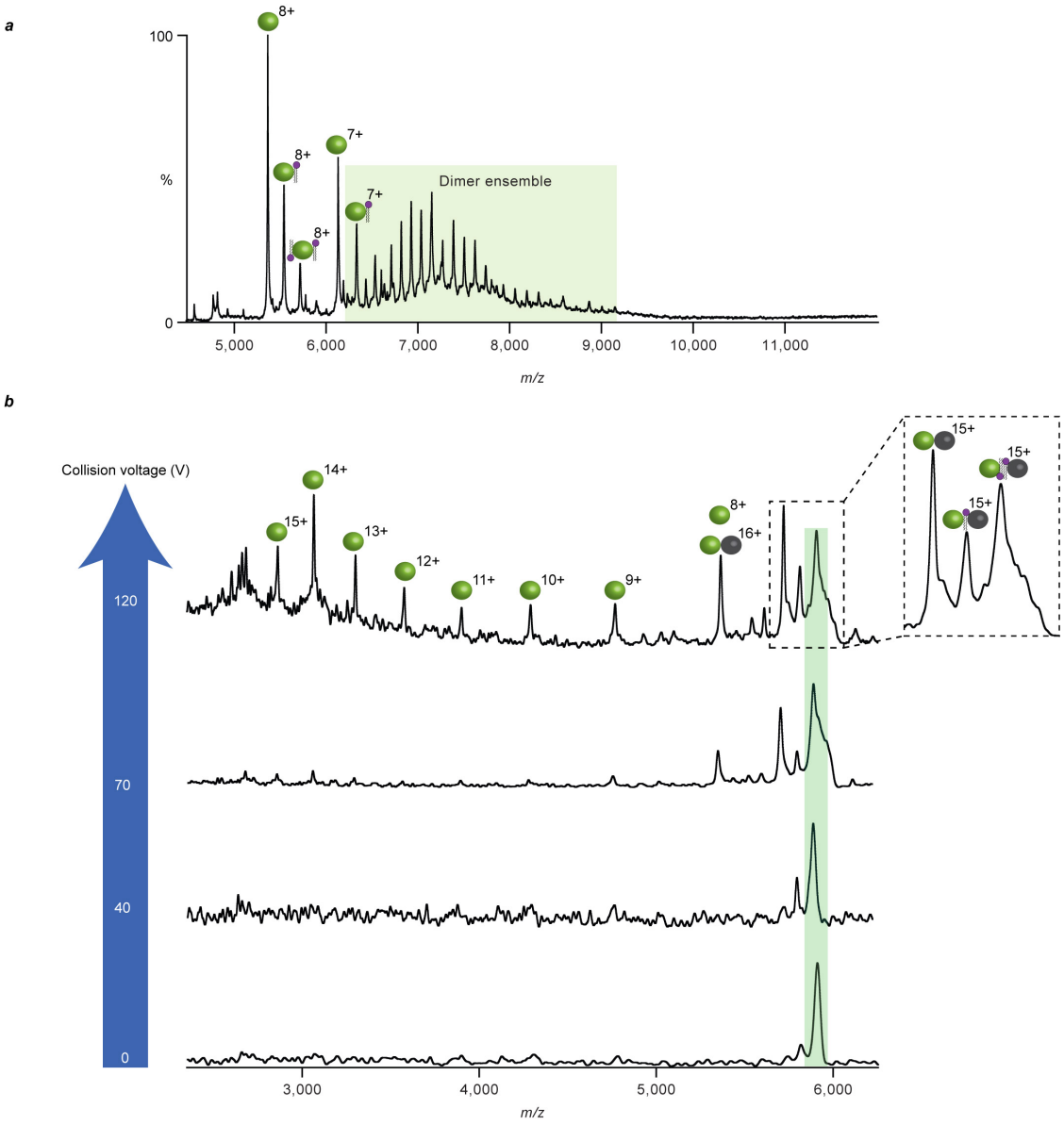


Figure 5.3: HE-nMS analysis of the NhaA dimer. **a** Mass spectrum of NhaA, released from C_8E_4 micelles, shows a broad ensemble of dimeric states, bound to a heterogeneous distribution of lipids. **b** HE-nMS analysis of the NhaA dimer bound to 2 cardiolipins (CL), dissociated with increasing collision energy. Mass selection of the 15 + charge state of the dimer, followed by application of increasing collision energy, leads to sequential dissociation of the 2 CL, leaving a transiently stable apo-dimer, followed by further dissociation of the dimer into monomers. Figure adapted from Gupta *et al.*[189]

similar predicted stabilities can have markedly different patterns of lipid binding and distributions of oligomeric states. In the absence of more sophisticated bioinformatics methods for identifying systems in which protein-lipid interactions strongly influence oligomeric states, robust experimental platforms such as those based on mass spectrometry will remain essential for investigating these phenomena.

5.5 Next steps - transient oligomers

Early in chapter 1, we stated that one of the principal motivating factors for studying membrane protein oligomerisation in detail, and specifically the role of protein-lipid interactions in this process, was the potential for insight into the behaviour of transient oligomers. In the course of our investigation into LeuT and SemiSWEET, we have gathered evidence to suggest that both of these proteins are transient oligomers, and hence this work has also acted as a proof of concept study which shows the potential value of a non-denaturing MS methodology for investigating the oligomerisation behaviour of other transient oligomers. Here, we have focused on two dimeric bacterial membrane proteins purified from *E. coli*, but the methods outlined in earlier chapters could theoretically be applied to much more complicated systems.

One avenue that might be fruitful to explore is examining the role of lipids in influencing membrane protein oligomerisation in cellular signalling. The early stages of signal propagation rely on receptors such as GPCRs, which bind an extracellular ligand and undergo a conformational change in order to trigger further downstream processes. Phospholipid binding to GPCRs has been shown to be involved in allosteric regulation of GPCR activity[190] and selectivity of G protein coupling[188]. In addition, as discussed in earlier chapters, many GPCRs also form higher order oligomers, with potential implications for function. It would be interesting to investigate the role of lipids in determining GPCR oligomeric state, to see if lipid-dependent changes in quaternary structure could also be implicated in regulation of GPCR signal propagation.

Bibliography

[1] Yeagle, P. L. Chapter 10 - Membrane Proteins. In *The Membranes of Cells*, 219 – 268 (Academic Press, Boston, 2016), 3rd edn.

[2] White, S. Membrane proteins of known 3D structure. URL `http://blanco.biomol.uci.edu/mpstruc/`.

[3] Miller, S., Lesk, A. M., Janin, J. & Chothia, C. The accessible surface area and stability of oligomeric proteins. *Nature* **328**, 834 – 836 (1987).

[4] Ali, M. H. & Imperiali, B. Protein oligomerization: how and why. *Bioorg. Med. Chem.* **13**, 5013 – 5020 (2005).

[5] Levy, E. D., Pereira-Leal, J. B., Chothia, C. & Teichmann, S. A. 3D complex: a structural classification of protein complexes. *PLOS Computational Biology* **2**, 1 – 12 (2006).

[6] Hilf, R. J. C. & Dutzler, R. X-ray structure of a prokaryotic pentameric ligand-gated ion channel. *Nature* **452**, 375 – 379 (2008).

[7] Bass, R. B., Strop, P., Barclay, M. & Rees, D. C. Crystal structure of *Escherichia coli* MscS, a voltage-modulated and mechanosensitive channel. *Science* **298**, 1582 – 1587 (2002).

[8] Murakami, S., Nakashima, R., Yamashita, E. & Yamaguchi, A. Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature* **419**, 587 – 593 (2002).

[9] Bogdanov, M., Mileykovskaya, E. & Dowhan, W. *Lipids in the Assembly of Membrane Proteins and Organization of Protein Supercomplexes: Implications for Lipid-linked Disorders*, 197–239 (Springer Netherlands, Dordrecht, 2008).

[10] Ellis, R. J. Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Current Opinion in Structural Biology* **11**, 114 – 119 (2001).

[11] Landry, C., Levy, E., Rabbo, D. A., Tarassov, K. & Michnick, S. Extracting insight from noisy cellular networks. *Cell* **155**, 983 – 989 (2013).

[12] Havugimana, P. *et al.* A census of human soluble protein complexes. *Cell* **150**, 1068 – 1081 (2012).

[13] Marsh, J. A. & Teichmann, S. A. Structure, dynamics, assembly, and evolution of protein complexes. *Annual Review of Biochemistry* **84**, 551 – 575 (2015).

[14] Zhou, Z. H., McCarthy, D. B., O'Connor, C. M., Reed, L. J. & Stoops, J. K. The remarkable structural and functional organization of the eukaryotic pyruvate dehydrogenase complexes. *PNAS* **98**, 14802 – 14807 (2001).

[15] Zhang, J., McCabe, K. A. & Bell, C. E. Crystal structures of λ exonuclease in complex with DNA suggest an electrostatic ratchet mechanism for processivity. *PNAS* **108**, 11872 – 11877 (2011).

[16] Pop, C., Timmer, J., Sperandio, S. & Salvesen, G. S. The apoptosome activates caspase-9 by dimerization. *Molecular Cell* **22**, 269 – 275 (2006).

[17] Goodsell, D. & Olson, A. Soluble proteins: size, shape and function. *Trends in Biochemical Sciences* **18**, 65 – 68 (1993).

[18] Cui, Q. & Karplus, M. Allostery and cooperativity revisited. *Protein Science* **17**, 1295 – 1307 (2008).

[19] Walden, H. *et al.* Tiny TIM: a small, tetrameric, hyperthermostable triosephosphate isomerase. *Journal of Molecular Biology* **306**, 745 – 757 (2001).

[20] Nooren, I. M. & Thornton, J. M. Structural characterisation and functional significance of transient protein-protein interactions. *Journal of Molecular Biology* **325**, 991 – 1018 (2003).

[21] Anderluh, A., Das, A. K. & Schütz, G. J. *Determining Oligomerization of Membrane Proteins by Single Molecule Methods*, 167–193 (Springer International Publishing, 2017).

[22] Tusnady, G. E., Dosztanyi, Z. & Simon, I. PDB_TM: selection and membrane localization of transmembrane proteins in the protein data bank. *Nucleic Acids Research* **33**, 275 – 278 (2005).

[23] Stansfeld, P. J. *et al.* MemProtMD: Automated insertion of membrane protein structures into explicit lipid membranes. *Structure* **23**, 1350 – 1361 (2015).

[24] Long, S. B., Campbell, E. B. & MacKinnon, R. Crystal structure of a mammalian voltage-dependent shaker family K^+ channel. *Science* **309**, 897 – 903 (2005).

[25] Gonen, T. & Walz, T. The structure of aquaporins. *Quarterly Reviews of Biophysics* **39**, 361 – 396 (2006).

[26] Calimet, N. *et al.* A gating mechanism of pentameric ligand-gated ion channels. *PNAS* **110**, 3987 – 3996 (2013).

[27] Perez, C., Khafizov, K., Forrest, L. R., Krämer, R. & Ziegler, C. The role of trimerization in the osmoregulated betaine transporter BetP. *EMBO reports* **12**, 804 – 810 (2011).

[28] Veenhoff, L. M., Heuberger, E. H. & Poolman, B. The lactose transport protein is a cooperative dimer with two sugar translocation pathways. *The EMBO Journal* **20**, 3056 – 3062 (2001).

[29] Fontenot, E. B., Ditusa, S. F., Kato, N., Dale, R. & Smith, A. P. Increased phosphate transport of *Arabidopsis thaliana* Pht1.1 by site-directed mutagenesis of tyrosine 312 may be attributed to the disruption of homomeric interactions. *Plant, Cell and Environment* **38** (2015).

[30] Limbird, L. & Lefkowitz, R. Negative cooperativity among β-adrenergic receptors in frog erythrocyte membranes. *JBC* **251**, 5007 – 5014 (1976).

[31] Ferré, S. *et al.* G protein-coupled receptor oligomerization revisited: functional and pharmacological perspectives. *Pharmacological Reviews* **66**, 413 – 434 (2014).

[32] Bouvier, M. Oligomerization of G-protein-coupled transmitter receptors. *Nature Reviews Neuroscience* **2**, 274 – 286 (2001).

[33] Kniazeff, J., Prazeau, L., Rondard, P., Pin, J.-P. & Goudet, C. Dimers and beyond: the functional puzzles of class C GPCRs. *Pharmacology & Therapeutics* **130**, 9 – 25 (2011).

[34] Khelashvili, G. *et al.* GPCR-OKB: the G Protein Coupled Receptor Oligomer Knowledge Base. *Bioinformatics* **26**, 1804 – 1805 (2010).

[35] Lacapere, J.-J., Pebay-Peyroula, E., Neumann, J.-M. & Etchebest, C. Determining membrane protein structures: still a challenge! *Trends in Biochemical Sciences* **32**, 259 – 270 (2007).

[36] Pandey, A., Shin, K., Patterson, R. E., Liu, X.-Q. & Rainey, J. K. Current strategies for protein production and purification enabling membrane protein structural biology. *Biochemistry and Cell Biology* **94**, 507 – 527 (2016).

[37] Howlett, G. J., Minton, A. P. & Rivas, G. Analytical ultracentrifugation for the study of protein association and assembly. *Current Opinion in Chemical Biology* **10**, 430 – 436 (2006).

[38] Hong, P., Koza, S. & Bouvier, E. S. P. A review: Size-exclusion chromatography for the analysis of protein biotherapeutics and their aggregates. *Journal of Liquid Chromatography & Related Technologies* **35**, 2923 – 2950 (2012).

[39] Slotboom, D. J., Duurkens, R. H., Olieman, K. & Erkens, G. B. Static light scattering to characterize membrane proteins in detergent solution. *Methods* **46**, 73 – 82 (2008).

[40] Bahadur, R. P., Chakrabarti, P., Rodier, F. & Janin, J. A dissection of specific and non-specific protein-protein interfaces. *Journal of Molecular Biology* **336**, 943 – 955 (2004).

[41] Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology* **372**, 774 – 797 (2007).

[42] Verdaasdonk, J. S., Lawrimore, J. & Bloom, K. Chapter 19 - Determining absolute protein numbers by quantitative fluorescence microscopy. In Waters, J. C. & Wittman, T. (eds.) *Quantitative Imaging in Cell Biology*, vol. 123 of *Methods in Cell Biology*, 347 – 365 (Academic Press, 2014).

[43] Nooren, I. M. & Thornton, J. M. Diversity of protein-protein interactions. *The EMBO Journal* **22**, 3486 – 3492 (2003).

[44] Perkins, J. R., Diboun, I., Dessailly, B. H., Lees, J. G. & Orengo, C. Transient protein-protein interactions: structural, functional, and network properties. *Structure* **18**, 1233 – 1243 (2010).

[45] Metzger, H. Transmembrane signalling: the joy of aggregation. *The Journal of Immunology* **149**, 1477 – 1487 (1992).

[46] Schlessinger, J. Ligand-induced, receptor-mediated dimerization and activation of EGF receptor. *Cell* **110**, 669 – 672 (2002).

[47] Gurevich, V. V. & Gurevich, E. V. How and why do GPCRs dimerize? *Trends in Pharmacological Sciences* **29**, 234 – 240 (2008).

[48] Duan, G. & Walther, D. The roles of post-translational modifications in the context of protein interaction networks. *PLOS Computational Biology* **11**, 1 – 23 (2015).

[49] Fahrenkamp, D. *et al.* Intramolecular hydrophobic interactions are critical mediators of STAT5 dimerization. *Scientific Reports* **6**, 35454 EP – (2016).

[50] de Diego Puente, T. *et al.* The protein acetyltransferase PatZ from *Escherichia coli* is regulated by autoacetylation-induced oligomerization. *Journal of Biological Chemistry* **290**, 23077 – 23093 (2015).

[51] Zhang, J., Li, X. & Li, J.-D. The roles of post-translational modifications on α-synuclein in the pathogenesis of Parkinson's diseases. *Frontiers in Neuroscience* **13**, 381 (2019).

[52] Chen, C.-P. *et al.* Membrane protein assembly: two cytoplasmic phosphorylated serine sites of Vpu from HIV-1 affect oligomerization. *Scientific Reports* **6**, 28866 EP (2016).

[53] Kanczewska, J. *et al.* Activation of the plant plasma membrane H^+-ATPase by phosphorylation and binding of 14-3-3 proteins converts a dimer into a hexamer. *PNAS* **102**, 11675 – 11680 (2005).

[54] Saliba, A.-E., Vonkova, I. & Gavin, A.-C. The systematic analysis of protein-lipid interactions comes of age. *Nature Reviews Molecular Cell Biology* **16**, 753 (2015).

[55] Harayama, T. & Riezman, H. Understanding the diversity of membrane lipid composition. *Nature Reviews Molecular Cell Biology* **19**, 281 EP – (2018).

[56] Fahy, E. *et al.* Update of the lipid maps comprehensive classification system for lipids. *Journal of Lipid Research* **50**, S9 – S14 (2009).

[57] Hansen, S. B., Tao, X. & MacKinnon, R. Structural basis of PIP_2 activation of the classical inward rectifier K^+ channel Kir2.2. *Nature* **477**, 495 – 498 (2011).

[58] Brohawn, S. G., Campbell, E. B. & MacKinnon, R. Physical mechanism for gating and mechanosensitivity of the human TRAAK K^+ channel. *Nature* **516**, 126 (2014).

[59] Hanson, M. A. *et al.* A specific cholesterol binding site is established by the 2.8 Å structure of the human β_2-adrenergic receptor. *Structure* **16**, 897 – 905 (2008).

[60] Contreras, F.-X., Ernst, A. M., Wieland, F. & Brügger, B. Specificity of intramembrane protein-lipid interactions. *Cold Spring Harbor Perspectives in Biology* **3** (2011).

[61] Dowhan, W. & Bogdanov, M. Lipid-protein interactions as determinants of membrane protein structure and function. *Biochemical Society Transactions* **39**, 767 – 774 (2011).

[62] Locke, D. & Harris, A. L. Connexin channels and phospholipids: association and modulation. *BMC Biology* **7**, 52 (2009).

[63] Simunovic, M. & Voth, G. A. Membrane tension controls the assembly of curvature-generating proteins. *Nature Communications* **6**, 7219 EP – (2015).

[64] Li, L. B., Vorobyov, I. & Allen, T. W. The role of membrane thickness in charged protein-lipid interactions. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1818**, 135 – 145 (2012).

[65] Maurya, S. R., Chaturvedi, D. & Mahalakshmi, R. Modulating lipid dynamics and membrane fluidity to drive rapid folding of a transmembrane barrel. *Scientific Reports* **3**, 1989 EP – (2013).

[66] Sundararaj, S. *et al.* The CyberCell Database (CCDB): a comprehensive, self?updating, relational database to coordinate and facilitate *in silico* modeling of *Escherichia coli*. *Nucleic Acids Research* **32**, D293–D295 (2004).

[67] Johansson, A. C. V. & Lindahl, E. Protein contents in biological membranes can explain abnormal solvation of charged and polar residues. *Proceedings of the National Academy of Sciences* **106**, 15684 – 15689 (2009).

[68] Sezgin, E., Levental, I., Mayor, S. & Eggeling, C. The mystery of membrane organization: composition, regulation and roles of lipid rafts. *Nature Reviews Molecular Cell Biology* .

[69] van Meer, G., Voelker, D. R. & Feigenson, G. W. Membrane lipids: where they are and how they behave. *Nature Reviews Molecular Cell Biology* **9**, 112 EP – (2008).

[70] Atilla-Gokcumen, G. E. *et al.* Dividing cells regulate their lipid composition and localization. *Cell* **156**, 428 – 439 (2014).

[71] Aviram, R. *et al.* Lipidomics analyses reveal temporal and spatial lipid organization and uncover daily oscillations in intracellular organelles. *Molecular Cell* **62**, 636 – 648 (2016).

[72] Yeagle, P. L. Non-covalent binding of membrane lipids to membrane proteins. *Biochimica et Biophysica Acta - Biomembranes* **1838**, 1548 – 1559 (2014).

[73] Laursen, M., Yatime, L., Nissen, P. & Fedosova, N. U. Crystal structure of the high-affinity na^+,k^+-atpase–ouabain complex with mg2+ bound in the cation binding site. *PNAS* (2013).

[74] Nury, H. *et al.* Structural basis for lipid-mediated interactions between mitochondrial ADP/ATP carrier monomers. *FEBS Letters* **579**, 6031 – 6036 (2005).

[75] Liu, W. *et al.* Structural basis for allosteric regulation of gpcrs by sodium ions. *Science* **337**, 232–236 (2012).

[76] Sun, F., Huo, X., Zhai, Y., Wang, A. & Rao, Z. Crystal structure of mitochondrial respiratory membrane protein complex ii. *Cell* **121**, 1043 – 1057 (2005).

[77] Shinzawa-Itoh, K. *et al.* Structures and physiological roles of 13 integral lipids of bovine heart cytochrome c oxidase. *The EMBO Journal* **26**, 1713 – 1725 (2007).

[78] Nogi, T., Fathir, I., Kobayashi, M., Nozawa, T. & Miki, K. Crystal structures of photosynthetic reaction center and high-potential iron-sulfur protein from *Thermochromatium tepidum*: Thermostability and electron transfer. *PNAS* **97**, 13561 – 13566 (2000).

[79] Gonen, T. *et al.* Lipid–protein interactions in double-layered two-dimensional aqp0 crystals. *Nature* **438**, 633 – 638 (2005).

[80] Stansfeld, P., Jefferys, E. & Sansom, M. Multiscale simulations reveal conserved patterns of lipid interactions with aquaporins. *Structure* **21**, 810 – 819 (2013).

[81] Koshy, C. *et al.* Structural evidence for functional lipid interactions in the betaine transporter BetP. *The EMBO Journal* **32**, 3096 – 3105 (2013).

[82] Cherezov, V. *et al.* High-resolution crystal structure of an engineered human β_2-adrenergic G protein-coupled receptor. *Science* **318**, 1258 – 1265 (2007).

[83] Vogeley, L. *et al.* *Anabaena* sensory rhodopsin: A photochromic color sensor at 2.0 Å. *Science* **306**, 1390–1393 (2004).

[84] Tsukihara, T. *et al.* The whole structure of the 13-subunit oxidized cytochrome C oxidase at 2.8 Å. *Science* **272**, 1136 – 1144 (1996).

[85] Kunji, E. R. *et al.* The transport mechanism of the mitochondrial ADP/ATP carrier. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1863**, 2379 – 2393 (2016).

[86] Contreras, F.-X. *et al.* Molecular recognition of a single sphingolipid species by a protein's transmembrane domain. *Nature* **481**, 525 – 529 (2012).

[87] Betaneli, V., Petrov, E. P. & Schwille, P. The role of lipids in VDAC oligomerization. *Biophysical Journal* **102**, 523 – 531 (2012).

[88] Raja, M. The role of phosphatidic acid and cardiolipin in stability of the tetrameric assembly of potassium channel KcsA. *Journal of Membrane Biology* **234**, 235 – 240 (2010).

[89] Botelho, A. V., Huber, T., Sakmar, T. P. & Brown, M. F. Curvature and hydrophobic forces drive oligomerization and modulate activity of rhodopsin in membranes. *Biophysical Journal* **91**, 4464 – 4477 (2006).

[90] Hern, J. A. *et al.* Formation and dissociation of M1 muscarinic receptor dimers seen by total internal reflection fluorescence imaging of single molecules. *PNAS* **107**, 2693 – 2698 (2010).

[91] Teichmann, A. *et al.* The specific monomer/dimer equilibrium of the corticotropin-releasing factor receptor type 1 is established in the endoplasmic reticulum. *Journal of Biological Chemistry* **289**, 24250 – 24262 (2014).

[92] Kasai, R. S. *et al.* Full characterization of GPCR monomer–dimer dynamic equilibrium by single molecule imaging. *The Journal of Cell Biology* **192**, 463 – 480 (2011).

[93] Fenn, J., Mann, M., Meng, C., Wong, S. & Whitehouse, C. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64 – 71 (1989).

[94] Meng, C. K., Mann, M. & Fenn, J. B. Of protons or proteins. *Zeitschrift für Physik D: Atoms, Molecules and Clusters* **10**, 361 – 368 (1988).

[95] Katta, V. & Chait, B. T. Observation of the heme-globin complex in native myoglobin by electrospray-ionization mass spectrometry. *JACS* **113**, 8534 – 8535 (1991).

[96] Ganem, B., Li, Y. T. & Henion, J. D. Observation of noncovalent enzyme-substrate and enzyme-product complexes by ion-spray mass spectrometry. *JACS* **113**, 7818 – 7819 (1991).

[97] Light-Wahl, K. J., Winger, B. E. & Smith, R. D. Observation of the multimeric forms of Concanavalin A by electrospray ionization mass spectrometry. *JACS* **115**, 5869 – 5870 (1993).

[98] Loo, J. A. Observation of large subunit protein complexes by electrospray ionization mass spectrometry. *Journal of Mass Spectrometry* **30**, 180 – 183 (1995).

[99] Loo, J. A. Studying non-covalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrometry Reviews* **16**, 1 – 23 (1997).

[100] Fitzgerald, M. C., Chernushevich, I. & Standing, K. G. Probing the oligomeric structure of an enzyme by electrospray ionization time-of-flight mass spectrometry. *PNAS* **93**, 6851 – 6856 (1996).

[101] Hilton, G. R. & Benesch, J. L. P. Two decades of studying non-covalent biomolecular assemblies by means of electrospray ionization mass spectrometry. *Journal of The Royal Society Interface* **9**, 801 – 816 (2012).

[102] Leney, A. C. & Heck, A. J. R. Native mass spectrometry: What is in the name? *JASMS* **28**, 5 – 13 (2017).

[103] Van Berkel, W. J. H., Van Den Heuvel, R. H. H., Versluis, C. & Heck, A. J. R. Detection of intact megadalton protein assemblies of vanillyl-alcohol oxidase by mass spectrometry. *Protein Science* **9**, 435 – 439 (2008).

[104] Rostom, A. A. *et al.* Detection and selective dissociation of intact ribosomes in a mass spectrometer. *PNAS* **97**, 5185 – 5190 (2000).

[105] Tito, M. A., Tars, K., Valegard, K., Hajdu, J. & Robinson, C. V. Electrospray time-of-flight mass spectrometry of the intact MS2 virus capsid. *JACS* **122**, 3550 – 3551 (2000).

[106] Breuker, K. & McLafferty, F. W. Stepwise evolution of protein native structure with electrospray into the gas phase, 10^{-12} to 10^{2} s. *PNAS* **105**, 18145 – 18152 (2008).

[107] Hopper, J. T. S. & Robinson, C. V. Mass spectrometry quantifies protein interactions: from molecular chaperones to membrane porins. *Angewandte Chemie International Edition* **53**, 14002 – 14015 (2014).

[108] Siuzdak, G. *et al.* Mass spectrometry and viral analysis. *Chemistry & Biology* **3**, 45 – 48 (1996).

[109] Scarff, C. A., Thalassinos, K., Hilton, G. R. & Scrivens, J. H. Travelling wave ion mobility mass spectrometry studies of protein structure: biological significance and comparison with X-ray crystallography and nuclear magnetic resonance spectroscopy measurements. *Rapid Communications in Mass Spectrometry* **22**, 3297–3304 (2008).

[110] Wolynes, P. G. Biomolecular folding in vacuo!!!(?). *PNAS* **92**, 2426 – 2427 (1995).

[111] Barrera, N. P., Di Bartolo, N., Booth, P. J. & Robinson, C. V. Micelles protect membrane complexes from solution to vacuum. *Science* **321**, 243 – 246 (2008).

[112] Marcoux, J. & Robinson, C. Twenty years of gas phase structural biology. *Structure* **21**, 1541 – 1550 (2013).

[113] Laganowsky, A., Reading, E., Hopper, J. T. S. & Robinson, C. V. Mass spectrometry of intact membrane protein complexes. *Nature Protocols* **8**, 639 – 651 (2013).

[114] Hernández, H. & Robinson, C. V. Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nature Protocols* **2**, 715 – 726 (2007).

[115] Reading, E. *et al.* The role of the detergent micelle in preserving the structure of membrane proteins in the gas phase. *Angewandte Chemie International Edition* **54**, 4577 – 4581 (2015).

[116] Sobott, F., Hernández, H., McCammon, M. G., Tito, M. A. & Robinson, C. V. A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal. Chem.* **74**, 1402 – 1407 (2002).

[117] de la Mora, J. F. Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Analytica Chimica Acta* **406**, 93 – 104 (2000).

[118] Benesch, J. L. P. Collisional activation of protein complexes: Picking up the pieces. *JASMS* **20**, 341 – 348 (2009).

[119] Aquilina, J. A., Benesch, J. L. P., Bateman, O. A., Slingsby, C. & Robinson, C. V. Polydispersity of a mammalian chaperone: Mass spectrometry reveals the population of oligomers in αB-crystallin. *PNAS* **100**, 10611 – 10616 (2003).

[120] Dixit, S. M., Polasky, D. A. & Ruotolo, B. T. Collision induced unfolding of isolated proteins in the gas phase: past, present, and future. *Current Opinion in Chemical Biology* **42**, 93 – 100 (2018).

[121] Hopper, J. T. S. & Oldham, N. J. Collision induced unfolding of protein ions in the gas phase studied by ion mobility-mass spectrometry: The effect of ligand binding on conformational stability. *JASMS* **20**, 1851 – 1858 (2009).

[122] Barrera, N. P. *et al.* Mass spectrometry of membrane transporters reveals subunit stoichiometry and interactions. *Nature Methods* **6**, 585 – 589 (2009).

[123] Wang, S. C. *et al.* Ion mobility mass spectrometry of two tetrameric membrane protein complexes reveals compact structures and differences in stability and packing. *J. Am. Chem. Soc.* **132**, 15468 – 15470 (2010).

[124] Reading, E. *et al.* The effect of detergent, temperature, and lipid on the oligomeric state of MscL constructs: Insights from mass spectrometry. *Chemistry & Biology* **22**, 593 – 603 (2015).

[125] Pliotas, C. *et al.* The role of lipids in mechanosensation. *Nat. Struct. & Mol. Biol.* **22**, 991 – 998 (2015).

[126] Zhou, M. *et al.* Mass spectrometry of intact V-type ATPases reveals bound lipids and the effects of nucleotide binding. *Science* **334**, 380 – 385 (2011).

[127] Housden, N. G. *et al.* Intrinsically disordered protein threads through the bacterial outer-membrane porin OmpF. *Science* **340**, 1570 – 1574 (2013).

[128] Schmidt, C. *et al.* Comparative cross-linking and mass spectrometry of an intact F-type ATPase suggest a role for phosphorylation. *Nature Communications* **4**, 1985 (2013).

[129] Marty, M. T., Hoi, K. K., Gault, J. & Robinson, C. V. Probing the lipid annular belt by gas-phase dissociation of membrane proteins in nanodiscs. *Angewandte Chemie International Edition* **55**, 550 – 554 (2016).

[130] Laganowsky, A. *et al.* Membrane proteins bind lipids selectively to modulate their structure and function. *Nature* **510**, 172 – 175 (2014).

[131] Ben-Nissan, G. & Sharon, M. Capturing protein structural kinetics by mass spectrometry. *Chem. Soc. Rev.* **40**, 3627 – 3637 (2011).

[132] Painter, A. J. *et al.* Real-time monitoring of protein complexes reveals their quaternary organization and dynamics. *Chemistry & Biology* **15**, 246 – 253 (2008).

[133] Keetch, C. A. *et al.* L55P Transthyretin accelerates subunit exchange and leads to rapid formation of hybrid tetramers. *Journal of Biological Chemistry* **280**, 41667 – 41674 (2005).

[134] Benesch, J. L. P., Aquilina, J. A., Ruotolo, B. T., Sobott, F. & Robinson, C. V. Tandem mass spectrometry reveals the quaternary organization of macromolecular assemblies. *Chemical Biology* **13**, 597 – 605 (2006).

[135] Gupta, K. *et al.* Identifying key membrane protein lipid interactions using mass spectrometry. *Nature Protocols* **13**, 1106 – 1120 (2018).

[136] Makarov, A. Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis. *Anal. Chem.* **72**, 1156 – 1162 (2000).

[137] Makarov, A. *et al.* Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal. Chem.* **78**, 2113 – 2120 (2006).

[138] Rose, R. J., Damoc, E., Denisov, E., Makarov, A. & Heck, A. J. R. High-sensitivity orbitrap mass analysis of intact macromolecular assemblies. *Nature Methods* **9**, 1084 – 1086 (2012).

[139] Marcoux, J. *et al.* Mass spectrometry reveals synergistic effects of nucleotides, lipids, and drugs binding to a multidrug resistance efflux pump. *PNAS* **110**, 9704 – 9709 (2013).

[140] Gault, J. *et al.* High-resolution mass spectrometry of small molecules bound to membrane proteins. *Nature Methods* **13**, 333 – 336 (2016).

[141] Chen, L.-Q. *et al.* Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature* **468**, 527 – 532 (2010).

[142] Xuan, Y. H. *et al.* Functional role of oligomerization for bacterial and plant SWEET sugar transporter family. *PNAS* **110**, 3685 – 3694 (2013).

[143] Xu, Y. *et al.* Structures of bacterial homologues of SWEET transporters in two distinct conformations. *Nature* **515**, 448 – 452 (2014).

[144] Chen, L.-Q. *et al.* Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. *Science* **335**, 207 – 211 (2012).

[145] Cura, A. J. & Carruthers, A. *Role of Monosaccharide Transport Proteins in Carbohydrate Assimilation, Distribution, Metabolism, and Homeostasis*, 863 – 914 (American Cancer Society, 2012).

[146] Sohlenkamp, C. & Geiger, O. Bacterial membrane lipids: diversity in structures and pathways. *FEMS Microbiology Reviews* **40**, 133 – 159 (2015).

[147] Horn, D. M., Zubarev, R. A. & McLafferty, F. W. Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *JASMS* **11**, 320 – 332 (2000).

[148] Senko, M. W., Beu, S. C. & McLafferty, F. W. Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. *JASMS* **6**, 229 – 233 (1995).

[149] Hsu, F.-F. *et al.* Structural characterization of cardiolipin by tandem quadrupole and multiple-stage quadrupole ion-trap mass spectrometry with electrospray ionization. *JASMS* **16**, 491 – 504 (2005).

[150] Bao, H., Dalal, K., Wang, V., Rouiller, I. & Duong, F. The maltose ABC transporter: action of membrane lipids on the transporter stability, coupling and ATPase activity. *Biochimica et Biophysica Acta - Biomembranes* **1828**, 1723 – 1730 (2013).

[151] Aponte-Santamaría, C., Briones, R., Schenk, A. D., Walz, T. & de Groot, B. L. Molecular driving forces defining lipid positions around aquaporin-0. *PNAS* **109**, 9887 – 9892 (2012).

[152] Bechara, C. *et al.* A subset of annular lipids is linked to the flippase activity of an ABC transporter. *Nat. Chem.* **7**, 255 – 262 (2015).

[153] Liko, I. *et al.* Dimer interface of bovine cytochrome C oxidase is influenced by local post-translational modifications and lipid binding. *PNAS* **113**, 8230 – 8235 (2016).

[154] Nadal, M. *et al.* Structure of the homodimeric androgen receptor ligand-binding domain. *Nature Communications* **8**, 14388 (2017).

[155] Mileykovskaya, E. & Dowhan, W. Cardiolipin membrane domains in prokaryotes and eukaryotes. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1788**, 2084 – 2091 (2009).

[156] Bligh, E. G. & Dyer, W. J. A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology* **37**, 911 – 917 (1959).

[157] Marty, M. T. *et al.* Bayesian deconvolution of mass and ion mobility spectra: From binary interactions to polydisperse ensembles. *Anal. Chem.* **87**, 4370 – 4376 (2015).

[158] Nelson, N. The family of Na^+/Cl^- neurotransmitter transporters. *Journal of Neurochemistry* **71**, 1785 – 1803 (2002).

[159] Masson, J., Sagné, C., Hamon, M. & Mestikawy, S. E. Neurotransmitter transporters in the central nervous system. *Pharmacological Reviews* **51**, 439 – 464 (1999).

[160] Iversen, L. Neurotransmitter transporters and their impact on the development of psychopharmacology. *British Journal of Pharmacology* **147**, S82 – S88 (2006).

[161] Yamashita, A., Singh, S. K., Kawate, T., Jin, Y. & Gouaux, E. Crystal structure of a bacterial homologue of Na^+/Cl^--dependent neurotransmitter transporters. *Nature* **437**, 215 – 223 (2005).

[162] Bechara, C. & Robinson, C. V. Different modes of lipid binding to membrane proteins probed by mass spectrometry. *JACS* **137**, 5240 – 5247 (2015).

[163] Nugent, T. & Jones, D. T. Membrane protein orientation and refinement using a knowledge-based statistical potential. *BMC Bioinformatics* **14**, 276 (2013).

[164] Raetz, C. R. Enzymology, genetics, and regulation of membrane phospholipid synthesis in *Escherichia coli*. *Microbiology and Molecular Biology Reviews* **42**, 614 – 659 (1978).

[165] Zhen, J. *et al.* Dopamine transporter oligomerization: impact of combining protomers with differential cocaine analog binding affinities. *Journal of Neurochemistry* **133**, 167 – 173 (2015).

[166] Quick, M. & Javitch, J. A. Monitoring the function of membrane transport proteins in detergent-solubilized form. *PNAS* **104**, 3603 – 3608 (2007).

[167] Harder, D. & Fotiadis, D. Measuring substrate binding and affinity of purified membrane transport proteins using the scintillation proximity assay. *Nature Protocols* **7**, 1569 – 1578 (2012).

[168] Kilic, F. & Rudnick, G. Oligomerization of serotonin transporter and its functional consequences. *PNAS* **97**, 3106 – 3111 (2000).

[169] Sorkina, T., Doolen, S., Galperin, E., Zahniser, N. R. & Sorkin, A. Oligomerization of dopamine transporters visualized in living cells by fluorescence resonance energy transfer microscopy. *Journal of Biological Chemistry* **278**, 28274 – 28283 (2003).

[170] Schmid, J. A. *et al.* Oligomerization of the human serotonin transporter and of the rat GABA transporter 1 visualized by fluorescence resonance energy transfer microscopy in living cells. *Journal of Biological Chemistry* **276**, 3805 – 3810 (2001).

[171] Scholze, P., Freissmuth, M. & Sitte, H. H. Mutations within an intramembrane leucine heptad repeat disrupt oligomer formation of the rat GABA transporter 1. *Journal of Biological Chemistry* **277**, 43682 – 43690 (2002).

[172] Anderluh, A. *et al.* Single molecule analysis reveals coexistence of stable serotonin transporter monomers and oligomers in the live cell plasma membrane. *Journal of Biological Chemistry* **289**, 4387 – 4394 (2014).

[173] Anderluh, A. *et al.* Direct PIP_2 binding mediates stable oligomer formation of the serotonin transporter. *Nature Communications* **8**, 14089 (2017).

[174] Coleman, J. A., Green, E. M. & Gouaux, E. X-ray structures and mechanism of the human serotonin transporter. *Nature* **532**, 334 (2016).

[175] Periole, X., Zeppelin, T. & Schiøtt, B. Dimer interface of the human serotonin transporter and effect of the membrane composition. *Scientific Reports* **8**, 5080 (2018).

[176] Tan, B. K. *et al.* Discovery of a cardiolipin synthase utilizing phosphatidylethanolamine and phosphatidylglycerol as substrates. *PNAS* **109**, 16504 – 16509 (2012).

[177] Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19 – 25 (2015).

[178] de Jong, D. H. *et al.* Improved parameters for the MARTINI coarse-grained protein force field. *Journal of Chemical Theory and Computation* **9**, 687 – 697 (2013).

[179] Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **126**, 014101 (2007).

[180] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81**, 3684 – 3690 (1984).

[181] Hess, B. P-LINCS: A parallel linear constraint solver for molecular simulation. *Journal of Chemical Theory and Computation* **4**, 116 – 122 (2008).

[182] Tironi, I. G., Sperb, R., Smith, P. E. & van Gunsteren, W. F. A generalized reaction field method for molecular dynamics simulations. *The Journal of Chemical Physics* **102**, 5451 – 5459 (1995).

[183] Páll, S. & Hess, B. A flexible algorithm for calculating pair interactions on SIMD architectures. *Computer Physics Communications* **184**, 2641 – 2650 (2013).

[184] Stansfeld, P. J. & Sansom, M. S. P. Molecular simulation approaches to membrane proteins. *Structure* **19**, 1562 – 1572 (2011).

[185] Jefferys, E., Sands, Z. A., Shi, J., Sansom, M. S. P. & Fowler, P. W. Alchembed: A computational method for incorporating multiple proteins into complex lipid geometries. *Journal of Chemical Theory and Computation* **11**, 2743 – 2754 (2015).

[186] Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry* **32**, 2319 – 2327 (2011).

[187] DeLano, W. L. The PyMOL molecular graphics system (2002).

[188] Yen, H.-Y. *et al.* PtdIns(4,5)P2 stabilizes active states of GPCRs and enhances selectivity of G-protein coupling. *Nature* **559**, 423 – 427 (2018).

[189] Gupta, K. *et al.* The role of interfacial lipids in stabilizing membrane protein oligomers. *Nature* **541**, 421 (2017).

[190] Dawaliby, R. *et al.* Allosteric regulation of G protein–coupled receptor activity by phospholipids. *Nature Chemical Biology* **12**, 35 – 39 (2015).