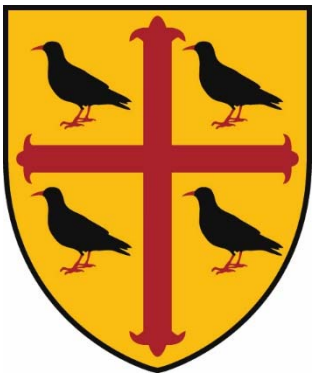
ELUCIDATING THE ROLE OF DETERGENTS, LIPIDS AND POST-TRANSLATIONAL MODIFICATIONS IN MEMBRANE PROTEIN STABILITY BY MASS SPECTROMETRY

Idlir Piro Liko

University of Oxford

St. Edmund Hall



This dissertation is submitted for the degree of Doctor of Philosophy in Physical and Theoretical Chemistry

2015

Elucidating the role of detergents, lipids and post-translational modifications in membrane protein stability by mass spectrometry

Idlir Piro Liko, St. Edmund Hall

This dissertation is submitted for the degree of Doctor of Philosophy in Physical and Theoretical Chemistry

Michaelmas term 2015

Abstract

Mass spectrometry in combination with ion mobility are utilised to characterise the stability of membrane proteins in the gas phase and to elucidate protein-lipid interactions. Detergents are most commonly used to transport membrane proteins from solution into the gas phase. Using a set of membrane proteins solubilised in several detergents, an ionisation mechanism of membrane protein-detergent complexes is proposed. This suggests that detergents serve as charge carriers in both positive and negative polarity. A better understanding of the underlying mechanism allows the identification of detergents that are removed at low energies and reduce Coulombic repulsion. The activation energy and charge density are found to be important parameters that dictate the stability of membrane proteins in the gas phase, allowing conditions to be found that preserve native-like conformations.

The stability of membrane proteins in the gas phase is also influenced by bound lipids. Methods to identify endogenous lipids directly from membrane protein-lipid complexes were developed. The stability provided by lipid binding is quantified

allowing the identification of lipids that enhance the stability of particular membrane proteins. Ionic protein-lipid interactions are investigated in the context of variability in pH. Protonation of acidic residues plays an important role in selectivity of lipid binding and enhances the stability at low pH of active conformations present in physiological conditions.

The methods developed are used to determine the oligomeric state of cytochrome *c* oxidase, pushing the boundaries of what is possible with non-denaturing mass spectrometry to the study of highly complex mammalian membrane proteins. By preserving non-covalent interactions in the gas phase evidence of equilibrium between monomeric and dimeric species of cytochrome *c* oxidase is found. Mass deviations can be attributed to a large lipid entity suggesting the role of lipids in stabilising the native oligomeric state. Using conventional mass spectrometry methods new lipids and phosphorylation sites are identified. Acetylation sites that have not been reported previously were located in close proximity to the dimeric interface, suggesting a potential role in stabilising this large dimeric interface.

Table of Contents

Abstract iii

Declaration viii

Abbreviations ix

Acknowledgements xii

Publications xiv

Chapter One: Mass spectrometry as a structural biology tool 1

1.1 Principles of mass spectrometry 2

1.1.1 Ion sources 2

1.1.2 Mass analysers 5

1.1.3 Detectors 11

1.2 Principles of ion mobility 13

1.2.1 RF-confined linear drift tube ion mobility 15

1.2.2 Travelling wave ion mobility 16

1.2.3 Determination of the rotationally averaged collision cross section 18

1.3 Present instruments 19

1.4 Mass spectrometry of protein complexes 22

1.4.1 Instrumental modification for transmittance of large ions 24

1.4.2 Structure elucidation in the gas phase 28

1.4.3 Investigation of protein dynamics in the gas-phase 32

1.4.4 Using IM-MS to monitor protein structure in the gas-phase 35

1.5 Membrane protein complexes 36

1.5.1. Preserving intact membrane protein complexes in the gas phase 39

1.5.2 Expressing membrane proteins and making them MS compatible 41

1.5.3 Using IM-MS to monitor the gas-phase dynamics of membrane proteins....44

1.6 Aims of this dissertation.... 47

Chapter Two: Probing the mechanism of membrane protein ionisation and detergent release in the gas phase.... 49

2.1 The mechanisms of electrospray ionisation.... 50

2.2 Detergent screening for optimisation of mass spectral quality.... 54

2.3 The effect of detergent on charge state distributions.... 61

2.4 The effect of polarity on charge state distribution.... 77

2.5 Preserving the native oligomeric state by reducing Coulombic repulsion .. 81

2.6 Probing the activation energy required to release membrane proteins from detergent micelles.... 86

2.7 Monitoring the stability of membrane proteins by collision-induced unfolding.... 94

2.8 Probing the stabilising effect of lipids on membrane proteins.... 103

2.9 Conclusion.... 108

Chapter Three: Tandem mass spectrometry of membrane protein complexes.... 113

3.1 Factors affecting gas-phase dissociation.... 115

3.1.1 Collision-induced dissociation....117

3.1.2 Infrared multiphoton dissociation....120

3.2 Tandem MS of membrane proteins.... 121

3.3 Infrared multiphoton dissociation of membrane proteins.... 132

3.4 Direct lipid identification.... 136

3.5 Conclusion.... 148

Chapter Four: Binding of lipids to membrane proteins is influenced by chargeable residues.... 151

4.1 Interaction of chargeable residues with charged lipids.... 152

4.2 Investigating the density of chargeable residues in the lipid head group binding region in membrane proteins.... 158

4.3 Preserving lipid binding by tailoring polarity 162

4.4 OmpF voltage-gated stability enhanced at low pH in presence of POPG. 171

4.5 Lipid binding enhances peptide binding to OmpF 176

4.6 Molecular dynamics simulations show differences in preferential PG binding at low pH 177

4.7 Conclusion 182

Chapter Five: The role of post-translational modifications and lipids in stabilisation of the oligomeric state of mammalian cytochrome *c* oxidase 185

5.1 The complexity of the mammalian cytochrome *c* oxidase 186

5.2 Cytochrome *c* oxidase is dimeric 190

5.3 Identification of co-purified lipids from cytochrome *c* oxidase 196

5.4 Post-translational modification of cytochrome *c* oxidase 201

5.5 Conclusion 205

Chapter Six: Conclusions 209

Chapter Seven: Experimental Procedures 215

7.1 General methods 215

7.2 Methods for Chapter Two 218

7.3 Methods for Chapter Three 226

7.4 Methods for Chapter Four: 229

7.5 Methods for Chapter Five 234

References 239

Declaration

This dissertation is a summary of research performed at Department of Chemistry, University of Oxford between September 2012 and September 2015. The work described here is wholly my own and not the outcome of a collaboration except where is specifically indicated in the text. It has not, neither as a whole, nor in part been submitted for a degree, diploma or any other academic qualification at any other university or academic institution. The length of this dissertation does not exceed the word limit.

Idlir Piro Liko

Oxford, October 2015

Abbreviations

AmtB	Ammonium channel
AqpZ	Aquaporin z
CcO	Cytochrome *c* oxidase
CCS	Collisional cross section
CDL	Cardiolipin
CID	Collision-induced dissociation
CIU	Collision-induced unfolding
CMC	Critical micelle concentration
CRM	Charged residue model
CW	Continuous wave
Da	Dalton
DDM	n-Dodecyl β-D-maltoside
DNA	Deoxyribonucleic acid
ECD	Electron capture dissociation
EHSS	Exact hard sphere scattering
E_{LAB}	Laboratory frame energy
ELIC	Pentameric ligand-gated ion channel
EM	Electron microscopy
ESI	Electrospray ionisation
ETD	Electron transfer dissociation
GFP	Green fluorescence protein
HCD	High-energy collision dissociation
HPLC	High performance liquid chromatography
ID	Internal diameter
IEM	Ion evaporation model

IM	Ion mobility
IMAC	Immobilized metal ion affinity chromatography
IRMPD	Infra-red multiphoton dissociation
IRMPD	Infra-red multiphoton dissociation
ITC	Isothermal calorimetry
LC	Liquid chromatography
LDAO	Lauryldimethylamine N-oxide
MALDI	Matrix assisted laser desorption ionisation
MATE	Multi-antimicrobial extrusion
MS	Mass spectrometry
MscL	Mechanosensitive channel of large conductance
MWCO	Molecular weight cut off
NG	Neopentyl glycol
NMR	Nuclear magnetic resonance
OG	Octyl glucoside
OGNG	Octyl glucose neopentyl glycol
OmpF	Outer membrane protein F
OTG	n-Octyl beta-D-thioglucopyranoside
PA	Phosphatidic acids
PA	Projection approximation
PC	Phosphatidylcholine
PDB	Protein data bank
PE	Phosphatidylethanolamines
PG	Phosphatidylglycerol
PI	Phosphatidylinositol
PS	Phosphatidylserine

Q	Quadrupole
RF	Radio frequency
RT	Retention time
SASA	Solvent accessible surface area
SEC	Size exclusion chromatography
SID	Surface-induced dissociation
SPR	Surface plasmon resonance
Th	Thomson
TM	Trajectory method
TOF	Time of flight

Acknowledgements

First and foremost, I would like to thank Carol for all her support and for the great opportunity to be part of her group. She has always been eager to help, at the right time, and always at the moment when her guidance has been needed the most. I can proudly say, I have learnt an enormous amount working with her; what I learnt most from Carol is that nature is not boring, and if you find something exciting, follow it!

Thank you to Justin and Andy for motivating me to critically think about my data, how to improve its presentation (Benesching), and that often there is no better solution than a bottle of champagne.

I also would like to thank Todd for fixing all the broken instruments, and Helena for her underappreciated support in the lab.

Thank you to previous members of the Robinson, Benesch and Baldwin groups for their help during the early days. I especially would like to thank Min, Zoe, Argyris (come again), Toni (Montana), Dale (I am still thinking positive about negative), Georg (Nina and Ima), Eamonn (always ahead of me in the publication list), Julien (for holding the security rope during climbing) and Art, as well as other beer buddies.

I also owe thanks to current group members Victoria, Dom, Olya, Henrik, Erik, Anna, Erika, Jun, Fran, Shahid, Nisha, and especially Heidi and Iva for showing me that what is more important is to be happy even if you come from south east Europe. Thank you to Gupta that used to be the next Nobel laureate in structural biology. I would like to thank Jon for always providing a positive spin and Joe for spinning the other way around. Thanks to Carla for always providing help and supervision. Thank you to Jason

for showing me that we can learn things even from measurements in solution. To Matteo, that with MD simulation you can do everything as long as you push hard. To Weston for making me feel proud to be American, although that I am not, and Cherine for French. To Michael, you can find something but if you do not publish fast you won't be first and to Victor for showing me that is the second mouse that eat the cheese.

New DPhil students and postdocs Stephen, Joe D., Miranda, Shane, Reid, Jingwen, Kin Hoi, Sami, Jani, Di, Dror, and Tim N. for showing me that one can be excellent from day one.

In addition, I would like to thank my collaborators Sejeong, Tom, Phil, Bonnie, Hagan, Nick, Luke and many others for providing excellent samples and many constructive discussions.

I would like to thank Tim, for teaching me everything from how to handle DNA to how to fix my bike. It feels like I have learnt more from Tim than from Google.

I had a great time, without you all, everything would have been simple, easy and wrong!

I would also like to thank my family, Mom, Dad, Besa and Ani as well as the newest members of our family, Daniel, Silvana and Koli for providing much needed support.

Finally I thank Enva for putting up with me, making me feel proud, and giving a meaning to everything.

Publications

Some of the results of this dissertation are present in the following publications:

Chapter One

1) Hopper J. T.S., Yu Y.T., Li D., Raymond A., Bostock M., **Liko I.**, Mikhailov V., Laganowsky A., Benesch J.P. L., Caffrey M., Nietlispach D., Robinson C. V., Detergent-free mass spectrometry of membrane protein complexes. *Nature Methods*, 10, 1206, (2013)
2) Zhou M., Politis A., Davies R., **Liko I.**, Stewart A., Stock D., Robinson C.V., ATP binding reduces conformational flexibility in a rotary ATPase – Evidence from ion mobility mass spectrometry. *Nature Chemistry*, 6:3, 208, (2014)

Chapter Two

3) Mehmood S., Marcoux J., Hopper J.T.S., Allison T. M., **Liko I.**, Borysik A. J., Robinson C. V., Charge reduction stabilizes intact membrane protein complexes for mass spectrometry. *Journal of the American Chemical Society*, 136, 17010, (2014)
4) Reading E.*, **Liko I.***, Allison T.M., Benesch J.P.L., Laganowsky A., Robinson C.V., The role of the detergent micelle in preserving the structure of membrane proteins in the gas phase. *Angewandte Chemie*, 54, 4577, (2015)

5) Allison T.M., Reading E., **Liko I.**, Baldwin A.J., Laganowsky A., Robinson C.V., Quantifying the stabilizing effects of protein-ligand interactions in the gas-phase. *Nature Communications*,6, 8551, (2015)
6) **Liko I.**, Hopper J.T.S., Allison T.M, Benesch J.P.L., Robinson C.V., Negative ions enhance survival of membrane protein complexes, *Journal of American Society for Mass Spectrometry. Accepted*.

Chapter Three

7) Landreh M., **Liko I.**, Uzdavinys P., Coincon M., Hipper J.T.S., Drew D., Robinson C.V., Controlling release unfolding and dissociation of membrane protein complexes in the gas phase through collisional cooling. *Chemical Communications*, 51, 15582, (2015)
8) Mikhailov V., **Liko I.**, Mize T.H., Benesch J.P.L, Robinson C.V., Infrared laser activation of intact soluble and membrane protein complexes in the gas-phase reveals structural insights and bound lipids. *Analytical Chemistry*. *Accepted* .
9) Hopper J.T.S.*, **Liko I.***, Giles K., Mize T.H., Benesch J.P.L., Robinson C.V., High energy source modifications allow tandem mass spectrometry insights into membrane protein assemblies. *Manuscript in preparation*.

Chapter Four

10) **Liko I.**, Allison T.M., Degiacomi M., Lee S., Gault J., Housden N.G., Stansfeld P.R., Benesch J.L.P., Kleanthous C., Wallace B., Samson M.S.P., Bayley H., Hopper J.T.S., Robinson C.V., The selectivity of lipid biding to

membrane proteins is influenced by chargeable residues. *Manuscript in preparation.*

Chapter Five

11) **Liko I.**, Degiacomi M., Mohammed Sh., Yoshikawa Sh., Schmidt C., Robinson C.V., Dimer interface of bovine cytochrome c oxidase is influenced by local post-translational modifications and lipid binding. *Proceedings of the National Academy of Science. Accepted.*

*****Equal contributions

Chapter One: Mass spectrometry as a structural biology tool

The very first mass spectrometer, then called a parabola spectrograph, measured the mass of molecules smaller than five atoms in size[1]; a hundred years later a mass spectrum of a piece of complex biological machinery[2] was recorded. This progression has made mass spectrometry (MS) an indispensable tool in structural biology owning to its unparalleled sensitivity, mass resolution and accuracy.

Initially MS was limited to identification of small molecules in the low thousands of Daltons (Da) and with a simple fragmentation pattern; however with the invention of soft ionisation techniques, improved mass analysers and detectors, MS gained a leading role in identifying the primary structure of proteins and their post-translation modifications.[3] Later developments demonstrated that even specific non-covalent interactions can be detected[4,5] making a small step for MS but a giant leap towards structural elucidation of protein-ligand complexes and protein-protein interactions. A notable achievement was coupling ion mobility (IM) with MS,[6] allowing not only determination of stoichiometry but also their topologies, providing structural insight into the quaternary structure of protein complexes.[7] A major advantage MS has relative to other analytical techniques is its versatility of sample requirement, allowing interrogation of protein complexes isolated directly from the cell or direct identification of therapeutics from tissues.[8] Recently MS has been shown to provide structural information in the formidable case of membrane proteins,[9] providing insight to their stability and interactions of lipids, peptides and drugs.[10]

In this introduction, the focus is first placed on the principles of IM-MS and key instrumental developments such as soft ionisation, collisional cooling, and optimised quadrupoles for high-mass transmission, while modified Orbitrap technology and time of

flight (TOF) mass analysers are highlighted. These improvements helped guide MS into its role in structural biology.[11,12]

1.1 Principles of mass spectrometry

At the core of mass spectrometry is the analysis of gas-phase ions and reporting them based on the ratio of mass-to-charge (*m/z*). In the most simplistic form mass spectrometers can be divided in three physical parts: the ion source, analyser and detector.

1.1.1 Ion sources

The sample is ionized and transferred to the gas phase in the ion source. Ions are produced via injection or capture of electrons, protonation or deprotonation, by adduct formation with other species or gas-phase chemical reactions. A variety of ionisation techniques, including liquid-phase and solid-phase ion sources, have been developed giving the opportunity to generate ions from a variety of different sample types.

Typically ion sources are classified based on the energy that is transferred to the analyte during the ionisation process. Electron, chemical are considered hard ionisation techniques, usually generating small fragmented ions.[13] To preserve intact species soft ionisation techniques have been developed which can ionise samples from a moderate mass range (up to 10,000 Da), such as fast atomic bombardment[14] and to high mass range, up to tens of megadaltons as in the case of electrospray ionisation (ESI)[15,16] and matrix-assisted laser desorption ionisation (MALDI).[17,18] Other soft ionisation techniques include laser-induced liquid beam ion desorption (LILBID)[19] and desorption electrospray ionisation (DESI).[20,21]

MALDI is performed in two steps, first the sample is dissolved in a matrix, an organic solvent, typically sinapinic acid serving as a chromophore that adsorbs the laser irradiation.

Subsequently, under vacuum conditions, the dried sample is irradiated by an intense laser pulse. The mechanism of ion generation via MALDI is yet to be elucidated, however it is accepted that the energy generated by the laser is adsorbed by the matrix causing local sublimation and ionisation of the sample. Usually, MALDI generates singly charged ions allowing for direct mass measurements of complex samples.[22]

In contrast to MALDI, during ESI the analyte is introduced to the mass spectrometer in solution. Typically, volatile buffers such as ammonium acetate or ammonium bicarbonate are employed to increase the ionisation rate and minimise salt adduct formation. The solution is passed through a capillary with a weak flux and strong electric field generated by a potential difference between the capillary and cone (Figure 1.1). These are positioned orthogonal to each other, giving rise to the so-called z-spray source arrangement, which minimises orifice saturation by the solvent while allowing ions to be directed to the cone. Charges accumulate in solution on the surface at the tip of the capillary and subsequently highly charged droplets are formed.[23,24] As the surface tension of the droplet is broken the droplet shape forms the so-called Taylor cone[25,26] allowing the formation of daughter droplets that carry a charge dependent on the surface tension of the solvent and droplet diameter, as defined by the Rayleigh limit equation. This mechanism is cascaded, the smaller droplets form with an estimated volume of 2% of the parent droplet but with 15% of their charge.[27] In the final stages, as a consequence of further evaporation and desorption these highly charged droplets give rise to gas phase ions. The ion formation mechanism is not completely elucidated. The current proposed models are discussed in detail in the second chapter.

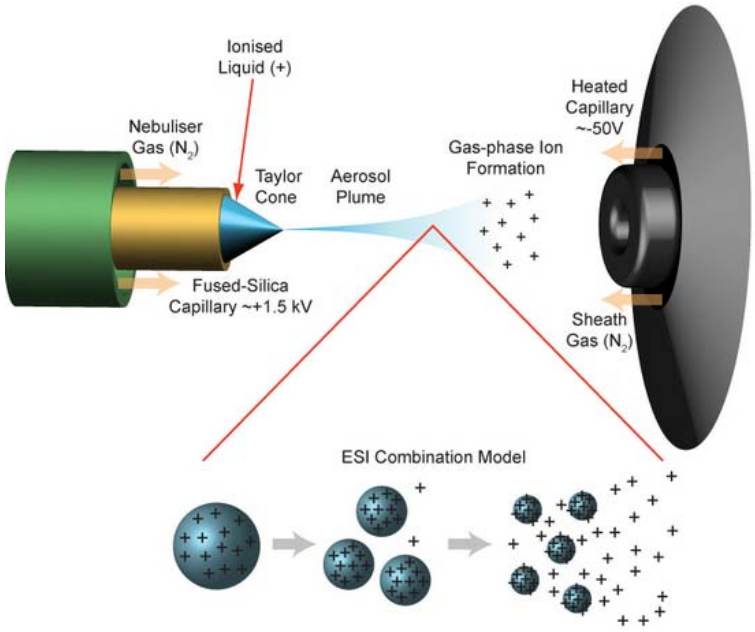


Figure 1.1: The electrospray ionisation process. The sample is introduced via a conductive capillary under an applied voltage. Initially a Taylor cone is formed at the tip of the capillary which generates a charged aerosol plume. Ion formation is facilitated via evaporation and droplet splitting. Figure from http://www.lamondlab.com.

ESI produces multiply charged ions allowing analysis of large complex species with utilisation of low nominal mass analysers.[23] Typically, Gaussian distributed consecutive ions are detected and the mass can be deconvolved by considering the *m/z* values of two adjacent peaks. In the case ions generated via protonation the mass and charge can be calculated using equation 1.1,[28] where z_1 is the charge of peak one, m_1 and m_2 are the *m/z* values of the first and second peaks respectively, m_p is the mass of a proton and i is the number of ions separating the first and the second peaks.

$$mass = \frac{i(m_2+m_p)(m_1+m_p)}{(m_2-m_1)} \tag{1.1a}$$

$$z_1 = \frac{i(m_2+m_p)}{(m_2-m_1)} \tag{1.1b}$$

1.1.2 Mass analysers

At the heart of MS is separation of ions according to their mass and change. Once the ions are generated they are directed to mass analysers via voltage controlled ion lenses and skimmers. Several mass analysers have been developed which separate ions based on their kinetic energy, momentum, flight time, trajectory stability and resonance frequency. The performance of mass analysers is measured by their accuracy and resolution. Other desirable characteristics are high speed of analysis, unlimited mass range and high ion transmission. Accuracy is the difference between an observed *m/z* value and the theoretical value. Current state-of-the-art instruments have high accuracy, in the low parts per million (ppm), typically 1–5 ppm. Resolution is defined as the peak width (Δm) at 50 percent of the peak height; the more narrow the peak width the higher the resolution.

The most widely used mass analysers are quadrupoles (Figure 1.2). They are composed of four parallel hyperbolic or cylindrical rods and separate ions by the means of ion trajectory stability in an applied electric field.[29,30] Two consecutive rods have opposite voltage sign which then alternates as a function of a direct potential (U) and the amplitude (V) of a radio frequency (RF) voltage. An ion will be attracted by an oppositely charged rod until the rod switches polarity, at which point it will move in the opposite direction due to charge repulsion. If the ion reaches the rod before it switches sign it will be filtered out by the quadrupole. Hence, an oscillating ion will have a stable trajectory if the amplitude of the oscillation remains less than the distance between rods (r), otherwise the ion will touch the rod and be discharged. For a given quadrupole with fixed r and constant angular frequency (ω) the stability area will be dependent on U and V as deduced from the Mathieu equation (Equation 1.2),[28] where a_u and q_u are proportionality constants of an ion in either the x or y direction.[30]

$$U = a_u \frac{m\omega^2 r^2}{8ez} \tag{1.2a}$$

$$V = q_u \frac{m\omega^2 r^2}{4ez} \tag{1.2b}$$

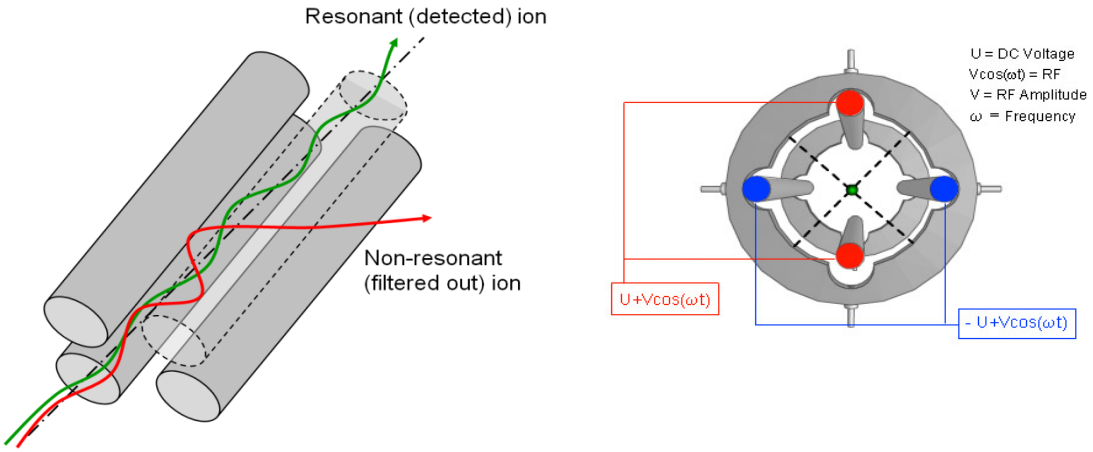


Figure 1.2: Quadrupole ion paths and rod energies. Adjacent rods are out of phase by 180°. Figure from: http://chemwiki.ucdavis.edu.

Maintaining a constant U/V ratio allows for scanning of ions with consecutive masses (Figure 1.3). In addition to this scanning mode quadrupoles can also operate in a static mode where only a particular ion is selected and allowed to pass through. Operating in this mode is desirable for interrogation of ions from a complex mixture as the selected ion can be further fragmented later in its path through the spectrometer and mass identified. The U/V ratio will define the resolution of the quadrupole, hence maintaining a fixed U/V ratio will result in a fixed resolution through the spectrum.[31] However, this mode of operation will result in peak width broadening with increased *m/z* values. As a result, a more desirable mode of operation is varying the U/V ratio so that the peak width is constant. Establishing a high U/V ratio correlates with a high resolving power, however when the quadrupole operates in only RF mode (U=0), the resolution becomes zero. In only-RF mode the quadrupole no longer functions as a mass filter, nevertheless, this mode is sufficient to focus ions towards the

centre of the quadrupole. For a given RF voltage, ions with masses lower than the stability area will oscillate with unstable trajectories whereas ions with larger masses will be focused. However, the focusing property is inversely proportion to *m/z* and ions with high *m/z* values with be poorly focused. The maximum *m/z* value that will be transmitted by a quadrupole is given by Equation 1.3, where the operating frequency is denoted by Ω.

$$[m/z]_{max} = \frac{4eV}{0.706\, r^2 \Omega^2} \tag{1.3}$$

To achieve transmission and focusing of ions with large masses, quadrupoles that operate with a high RF amplitude and low frequency have been developed.[32]

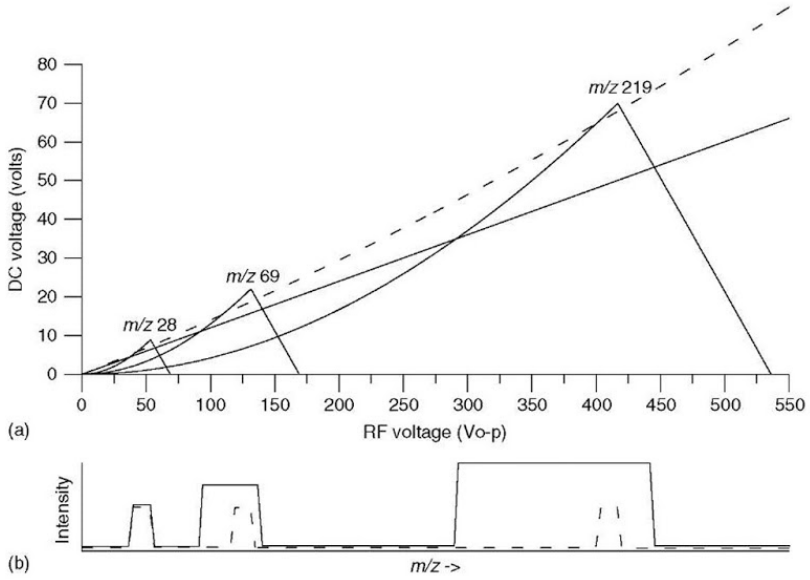


Figure 1.3: Ion stability diagrams based on the Mathieu equation. (a) The dashed line and straight lines represent the mass scan line for achieving stable trajectories of ions with *m/z* of 219 and 28 respectively. (b) An intensity profile of each line is shown, depicting that the dashed line will focus larger ions; however it will result in weak intensity for the smaller ions. Figure from Henchman and Still.[31]

Several instrumental designs use quadrupoles in tandem by taking advantage of both modes of operation, as a mass filter with a non-zero DC potential or as an ion guide by

applying a RF voltage only. For instance, in Waters Synapt G2 instruments, ions can be scanned or mass selected in the first quadrupole then activated in a collision cell, which is a second quadrupole operating in RF mode only and filled with a noble gas.

Another common analyser is the time of flight (TOF) mass analyser.[33] As the name suggests TOFs record the time it takes each ion to reach the detector (Figure 1.4). They are composed of two regions, the accelerating region or pusher, and field-free region, which ends at the detector. A packet of ions is accelerated in the direction of the flight tube by applying a potential between an electrode and the extraction grid. The accelerated ions are separated in the field-free region based on their velocities prior to being detected.[34] The time that each ion spends in the field-free region is recorded and converted to an *m/z* value by calculating its kinetic energy from the electrical potential (Equation 1.4)[28] where acceleration potential is denoted by *V* and the length of the field-free region by *L*.

$$m/z = \frac{2eVt^2}{L^2} \tag{1.4}$$

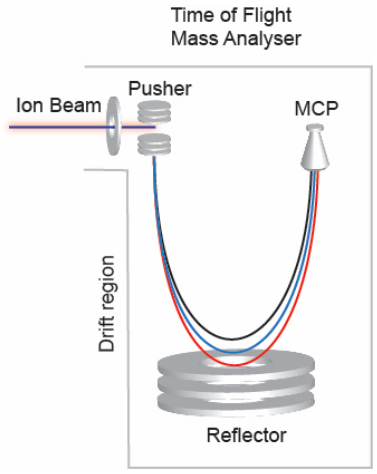


Figure 1.4: A schematic of a reflector time of flight mass analyser. Ions are accelerated then separated in the drift region. The reflector generates a unique ion trajectory separating the ions based on their mass with the low *m/z* ions arriving earlier than the higher *m/z* ions.

Since the time of flight is proportional to *m/z*, that is larger ions arrive at the detector at a later time than smaller ions, there is no theoretical upper mass limit.[35] In practice the analysis of high-mass ions in limited by a decrease in transmission due to dispersion of the ion beam. The resolution can be improved by increasing the length of the field-free region or lowering the acceleration voltage. However, lowering the acceleration voltage reduces sensitivity and linear extension of the flight tube increases ion scattering by gas molecules even though the TOF operates at 10^{-6} Torr. The resolution is also reduced due to the initial kinetic energy of the ion packet. This makes ions with the same *m/z* value arrive at the detector at different times.

To increase the resolution by narrowing the range of the initial kinetic energies of an ion packet two modifications have been made. First, a pulsed frequency is introduced at the accelerating region; ions with the same *m/z* value are brought to the same initial kinetic energy. In addition a reflectron is introduced at the end of the flight tube.[36] Serving as an ion mirror it redirects ions back through the fight tube, effectively doubling its length. Ions with high kinetic energy will penetrate the reflectron more before returning in the opposite direction, effectively cancelling the initial kinetic energy difference and improving the resolution.

Since the TOF measurement is time dependent and performed by a single packet of ions at a time, it is ideal to be coupled with pulsed ion sources such as MALDI. In the case of ESI, which generates ions continuously, an orthogonal position[37] of the TOF relative to the ion beam has been designed allowing for ion mass analysis in a continuous manner. As ions arrive they are trapped in the pusher region, then ions are orthogonally accelerated reaching the field-free region and mirrored by the reflectron where ions are separated. This

configuration minimises the initial ion velocities resulting in improved resolution without compromise in sensitivity.

One of the present challenges in biology is the elucidation of an organism's full proteome, for which mass spectrometry is useful.[38] For such an endeavour high accuracy with isotopic resolution mass analysers are required. Among the highest resolution mass analyses are the Fourier transform ion cyclotron resonance (FTICR) and the Orbitrap.[39-41] FTICR measures the frequency at which the ion oscillates in a given magnetic field, yielding high accuracy sufficient to mass separate large complex mixtures. Orbitraps offer comparable mass accuracy and resolution with a simpler instrumental architecture and are more user friendly than FTICR analyses. Hence, today Orbitraps are the analyser of chose in proteomics and metabolomics.[42]

The Orbitrap uses a static electric field to trap ions. The trap is constructed by an inner electrode in a spindle shape with an outer part shaped as a barrel that covers the entire inner electrode (Figure 1.5). Ions are introduced perpendicular to the z-axis by a C-trap and orbit around the inner electrode, oscillating in a harmonic motion in the z-direction. The oscillating ions will induce a current which is then Fourier transformed to individual frequencies. The frequency of the harmonic oscillation (ω) along the z-axis can be related to *m/z* value of a given ion and is independent of the kinetic energy, as shown by Equation 1.5,[28] where *k* is field curvature.

$$\omega = \sqrt{(z/m)k} \tag{1.5}$$

The oscillating frequency of an ion can be measured several times, significantly increasing its mass accuracy and sensitivity.

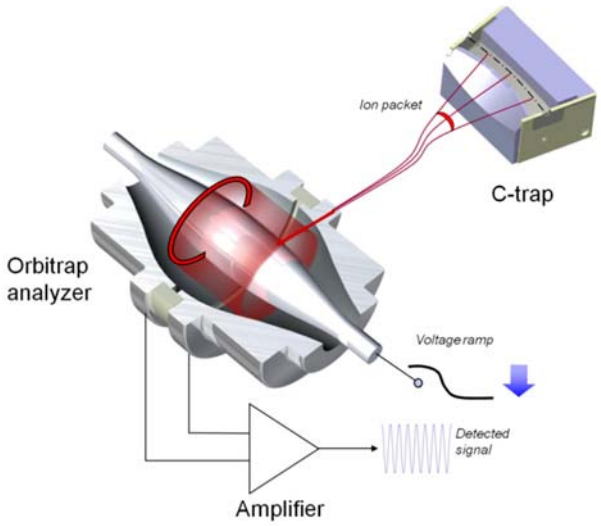


Figure 1.5: A schematic of an Orbitrap mass analyser. The ion packet is focused in the C-trap and accelerated towards the Orbitrap perpendicular to the *z*-axis. The oscillation harmonics of each ion is recorded and amplified prior to detection. Figure from Thermo Fisher Scientific.

1.1.3 Detectors

The generated and mass analysed ions are useful only if they are detected. As in the case of ion sources and mass analysers several types of detectors have been developed. These can be ranked based on the collection efficiency, signal-to-noise ratio, signal amplification, the dynamic range and the frequency of the response. They can be classified based on the mode of operation where some detectors consider the charge of the ions whereas others are based on image charge detection or measurement of the kinetic energy that an ion transfers to an electronic plate. In addition, point ion collectors which count a single mass ion at a time and array collectors which are capable or measuring multiple masses simultaneously, have been developed. It is worth mentioning that in Orbitrap and FTICR a non-destructive detector is positioned within the mass analysers in close proximity to the ion trajectory and detect ions based on the signal averaging of the oscillating ion packet.

Early detectors used a simple photographic plate at the end of the analyser. Ions with the same *m/z* value were deposited on the same spot on the plate and the intensity of the spot was correlated to ion abundance. This original design was important for early MS however it is now superseded by electron multipliers, which are the most used detectors. Electron multiplier detectors offer enhanced detection efficiency, with signal amplification benefitting from the kinetic energy transfer of a colliding ion. They are composed of either a single large dynode or of a series of smaller discrete dynodes. The dynodes operate at high potential to accelerate the mass analysed ions, effectively increasing the transfer of kinetic energy from the ion to the dynode. As an ion hits the detector plate other secondary particles with opposite charge are emitted. These secondary particles are then converted to electrons which are amplified exponentially, as each electron generates more electrons as it collide with the next dynode.

A discrete dynode detector is constructed by approximately twenty individual dynodes, whereas continuous dynode detectors have curved or cylindrical shapes, as in channeltron and microchannel plate (MCP) detectors.[43] A typical microchannel plate is composed of an ensemble of parallel cylindrical channels, each amplifying an electronic signal (Figure 1.6). A plate is capable of amplifying the number of electrons by a factor of 10^5 and in a typical detector several plates are positioned together. The yield of secondary electron emission (γ) is dependent on the total number of secondary electrons generated (N_e) and directly correlates to detector efficiency (ε) as given by Equation 1.6.[28]

$$\varepsilon = 1 - e^{-\gamma} \qquad (1.6)$$

By measuring the performance of an MCP with an inductive detector it has been shown the yield of secondary electron emission is inversely correlated with the square root of ion mass, as shown in Equation 1.7.[28] *U* is the acceleration voltage and z is ion charge.

$$\gamma = 7.4 \times 10^{-18} (Uez)^{3/2} m^{-1/2} \quad (1.7)$$

The efficiency of an MCP coupled to a TOF mass analyser decreases by 11 percent for ions with mass in the low kilodalton range.[44] Since the efficiency of a detector reduces with increased ion mass, new detectors such as inductive or cryogenic detectors that have high efficiency for large ions are under development.[45]

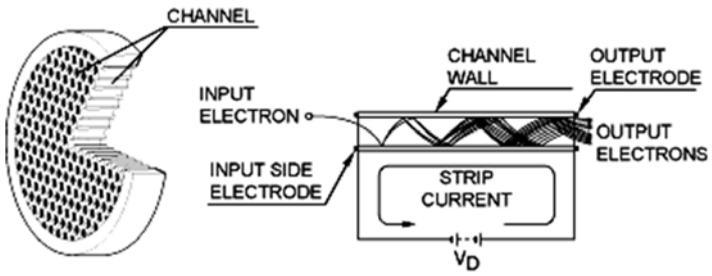


Figure 1.6: A schematic of a microchannel plate (left panel) and an individual channel (right). An electron cascade effect is illustrated inside the channel that leads to signal amplification. Figure from Hoffmann et al.[28]

1.2 Principles of ion mobility

Ion mobility (IM) separates ions based on their shape and charge, and can provide complementary information to MS.[7,46,47] IM-MS performs two dimensional separation of ions and can report their mass, which leads to elemental composition, and at the same time their solvent-free size that can be correlated to their structure.[48,49] IM separation involves passing ions through a chamber, called the drift tube, which is filled with a neutral gas. Small ions have a lower probability to collide with the gas molecules; hence they arrive earlier to the end

of the chamber than large ions. The latter interact more with the neutral gas molecules, and therefore take a longer time to traverse the drift tube (Figure 1.7). Only the surface of an ion interacts with the gas molecules hence ions will be separated based on their rotationally averaged collisional cross sections (CCS). Stable trajectories are established by applying an electrical field, as a result ions with a higher charge pass through the drift tube faster than ions with lower charge.

Three types of drift tubes are commonly employed; RF-confined linear drift tubes, traveling wave, and field asymmetric drift tubes. The latter has a high resolving power however the mobility of ions is correlated to their charge and applied field strength and not directly to their structural features, hence the CCS cannot be determined.

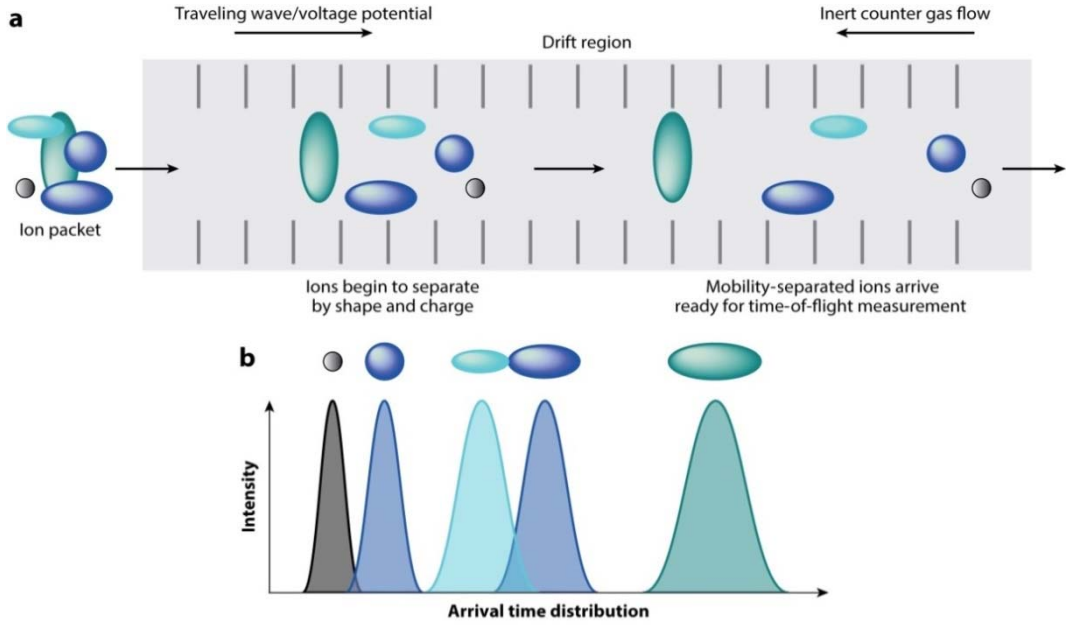


Figure 1.7: RF confined drift tube separation. (a) Separation of the ion packet is shown. Ions are separated under the influence of a counter gas based on their shape and charge. (b) An arrival time distribution diagram is depicted, showing that small ions arrive earlier than large ions. Figure from Mehmood, Allison and Robinson[50]

1.2.1 RF-confined linear drift tube ion mobility

The RF-confined linear drift tube IM separation is the one of the simplest forms of gas phase separation. It is composed of a chamber filled with a bath of neutral gas, usually nitrogen or a noble gas such as helium or argon, and operated under the influence of a static uniform electric field.[51] Ions having the same initial velocities are introduced to a RF-confined drift tube and the time taken by each ion to transit through the tube is recorded and reported once the ion reaches the detector. With a given constant electric field intensity (E), at a certain temperature (T) and pressure, the drift time (t_D) can be used to directly measure the ion CCS as shown by Mason-Schamp (Equation 1.8),[52] where m_I and m_N are the mass of ions and bath gas respectively, ρ is gas bath density, L is the length of the tube and k_b is the Boltzmann constant.

$$CCS_{AVE} = \frac{(18\pi)^{1/2} ze}{16\,(k_b T)^{1/2}} \left[\frac{1}{m_I} + \frac{1}{m_N}\right]^{1/2} \frac{t_D E}{L\rho} \tag{1.8}$$

This relationship between drift time and CCS is only valid when the ratio of electric field intensity to neutral gas density is low[53] (i.e. $E/\rho \leq 2\text{x}10^{-17}$ V cm^2) as the drag provided by the gas molecules relative to ion velocities is insufficient for ion separation based on their shape.

One of the major advantages of linear field drift tube separation is that the value of the drift time measured is absolute and no instrumental calibration is required. As a result, drift tube CCS measurements serve as a reference to calibrate other IM techniques to allow the calculation of CCS values from their respective arrival times.[54] This modality of separation does however suffer from low resolution and low duty cycle as the number of ions detected is low compared to the number generated.

1.2.2 Travelling wave ion mobility

Travelling wave ion mobility separates the ions by using a sequence of waves with symmetric potential that continuously passes through the tube. The device is made up by a series of stacked ring ion guides.[55] It operates under RF voltage where two sequential electrodes are out of phase by 180°. The RF voltage focuses the ions in the centre of the tube while a DC potential applied between each electrode sequentially creates a travelling wave that propagates the ions. The amplitude and frequency of the DC potential dictates the wave height and velocity respectively. By optimising the wave height and velocity ions are separated based on their mobility (Figure 1.8). For instance, an ion with a high mobility will be carried by the travelling wave and arrive at the detector faster than an ion with low mobility that will roll over the wave. Due to the roll-over effect, the path travelled by different ions is dependent on their mobility, hence having the same length of the tube and applying a travelling wave results in different effective ion path lengths, increasing the separation capabilities.

The mobility of an ion cannot be directly derived from the arrival time however it can be determined from empirically measured CCS data generated using drift tube ion mobility.[54] Once the wave height and velocity are optimised for separation the arrival time of an ion is measured. The wave height and velocity are then kept unchanged and a series of proteins with known CCS are recorded, in the same experimental conditions as the measurement performed on the sample of interest. The arrival time of the sample is then converted to CCS by generating a calibration curve using the proteins of known CCS.[46] As in the case of drift tube IM the measured CCS also depends on the choice of the neutral gas, hence the calibration and measurements are performed using the same gas. The major advantage of the travelling wave is its improved sensitivity however this mode of separation is hampered by a

modest resolution, and an increased potential for ion heating which can induce a gas-phase conformational change.

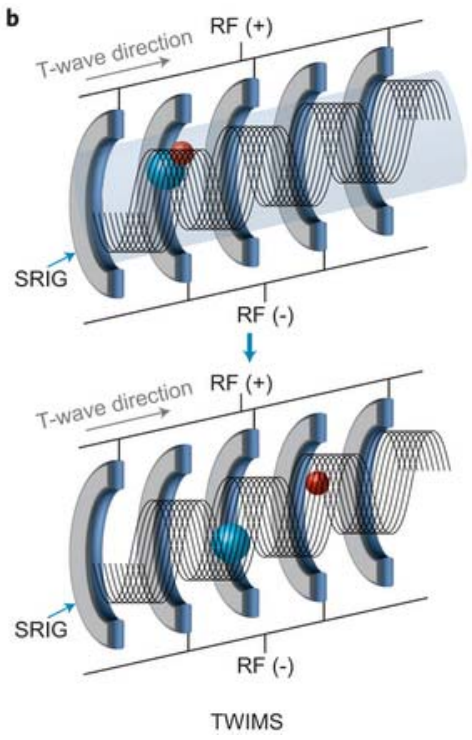


Figure 1.8: Traveling wave ion mobility separation. The applied potential at each electrode is alternated to generate a wave. Ions are separated as they propagate through the waves. Small ions will be carried over by the wave whereas the large ions have a longer mean free path due to roll over effect and exit the drift tube later. Figure from Lanucara et al. [7].

The coupling of IM to MS has become a powerful tool in structural biology providing information on complex mixtures, with the benefit of requiring only a very small amount of sample in the nanogram range. The coupled method provides high selectivity, an important criterion for characterisation of samples with low purity, helpful for samples that are challenging for structural interrogation by high-resolution techniques such as nuclear magnetic resonance (NMR) or X-ray crystallography.

IM-MS measurements are relatively fast compared to other techniques allowing for high-throughput screening.[56] The measurement is performed on the millisecond time scale and can be coupled with online chromatographic separation. An impressive application was performing liquid chromatography coupled with travelling wave IM-MS on a tryptic digest of *Escherichia coli* whole cell lysate.[57] This study showed an unprecedented increase (up to 60 percent) of peptides and proteins that could be identified. The limit of detection is another advantage for using IM-MS for sample characterisation. MS can detect at low attomolar concentrations and coupling to ion mobility improves the separation capabilities and increases the limit of detection 10-fold.[58]

1.2.3 Determination of the rotationally averaged collision cross section

The measured experimental CCS for proteins with known structures can be compared to theoretical CCS values. This comparison allows a better understanding of the chemistry of solvent free species and to correlate them to solution measurements. The theoretical CCS values can be derived from structures elucidated from X-ray, NMR, electron microscopy, small-angle X-ray scattering or molecular dynamics simulations. To calculate the CCS, from a known structural shape, several methods have been developed where projection approximation (PA)[59], exact hard sphere scattering (EHSS)[60] and trajectory method (TM)[61] are among the ones most frequently applied. The TM method takes into account collision of ions with drift gas as well as long range interactions. This method provides accurate CCS calculations, however performing the calculations is very computationally expensive, limiting the application of this method to small systems. The EHSS method on the other hand calculates CCS by considering only the short-range scattering and collision of ions with gas molecules. This method gives a good approximation for large ions such as proteins and is significantly faster than the TM. The PA method calculates the CCS by taking into

consideration an average geometric projection of all possible orientations. This method is fast and computationally inexpensive however it does not take into account the scattering of ions and the interactions with gas molecules. This usually results in an underestimation of the CCS (Ω). Modifications of the PA method to correct for the underestimation have been found empirically[54] (Equation 1.9).[54] M_{exp} is the experimental mass and M_{str} is the mass of the species with known structure.

$$\Omega_{PA,corrected} = 1.14\Omega_{PA}(M_{exp}/M_{str})^{\frac{2}{3}} \quad (1.9)$$

The corrected PA can be applied to large systems with missing residues, as these can be modelled as perfect spherical entities.[62] Another modification to the PA method is to include the superposition of atomic potentials and shape factors. This so-called projection superposition[63] approximation considers van der Waals interactions between ions and gas molecules providing accurate calculations with reduced computational time.

1.3 Present instruments

Having introduced the basic principles of MS and IM, the focus is now placed on the architecture of the instruments used in this study. For direct stoichiometric elucidation of protein complexes a modified QqTOF instrument was used (Q stands for a quadrupole operating in normal mode and q for a quadrupole operating in RF mode only) (Figure 1.9).

Electrospray generated ions from a *z*-spray ion source are introduced into the vacuum of a mass spectrometer by passing through a set of ion lenses optimised to focus the ion beam. As the ions pass from atmospheric pressure to vacuum a backing pressure is applied to reduce their kinetic energy. Ions are then either mass selected or scanned by a quadrupole operating with both DC and AC potentials. Subsequently, ions pass to the next quadrupole that serves

as a collision cell. It operates in RF-only mode to focus the ion beam and activate the ions. The collision cell is filled with gas molecules, usually nitrogen or argon, to facilitate ion activation. Finally the parent ion or ion fragments are mass analysed by a reflector TOF analyser positioned orthogonal to the ion beam and detected via an MCP-type detector.

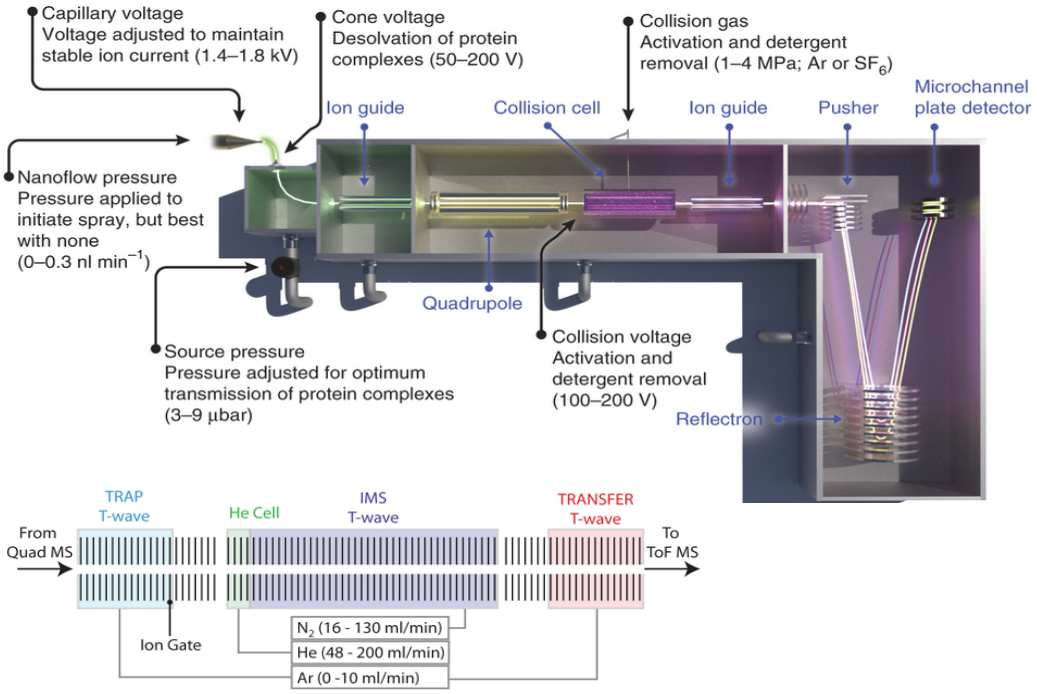


Figure 1.9: A schematic of the ion pathway in a QqTOF instrument. The inset shows the ion mobility cell, which in a Waters Synapt instrument is located in the position of the collision cell. Figure from Laganowsky et al.[7,64] and Waters.

For IM separation measurements a similar instrument design was used however instead of the collision cell an IM cell is utilised (Figure 1.9 inset). The IM cell is composed of three regions: a trap, the IMS, and transfer. As the name implies ions are initially trapped in the trap region and are pushed into the IMS region at a specific time interval. To improve the transmission of ions from the high pressure of the trap to the low pressure IMS region, a helium cell is introduced.[7] As ion packets enter the IMS region, ion mobility measurement is performed by using a travelling wave or drift tube, depending on the instrument. Shape and

charge separated ions are further focused in the transfer region and then mass analysed by the TOF. Both trap and transfer regions can serve as a collision cell, hence ions can be made to undergo conformational change or fragmented prior to or after the IM separation. If activation is applied in the transfer region the arrival time distribution of parent ion and fragment ions is the same. This type of activation allows for direct correlation of fragments to parent ions in a complex mixture.[65]

High resolution measurements are performed using the Q-Exactive instrument shown in Figure 1.10. In this instrument ions are initially directly electrosprayed to the cone. They are efficiently focused by an s-lens made of a series of electrodes and transmitted with increased sensitivity to a bent flatapole that operates in RF-only mode. Here the ion focus is improved and the noise is reduced by an active beam guide. Then ions pass through a quadrupole for mass selection followed by trapping and activation in a higher-energy collision dissociation (HCD) cell. From the HCD cell ions are further focused in a C-trap to form a highly compact ion packet and injected to the Orbitrap mass analyser.

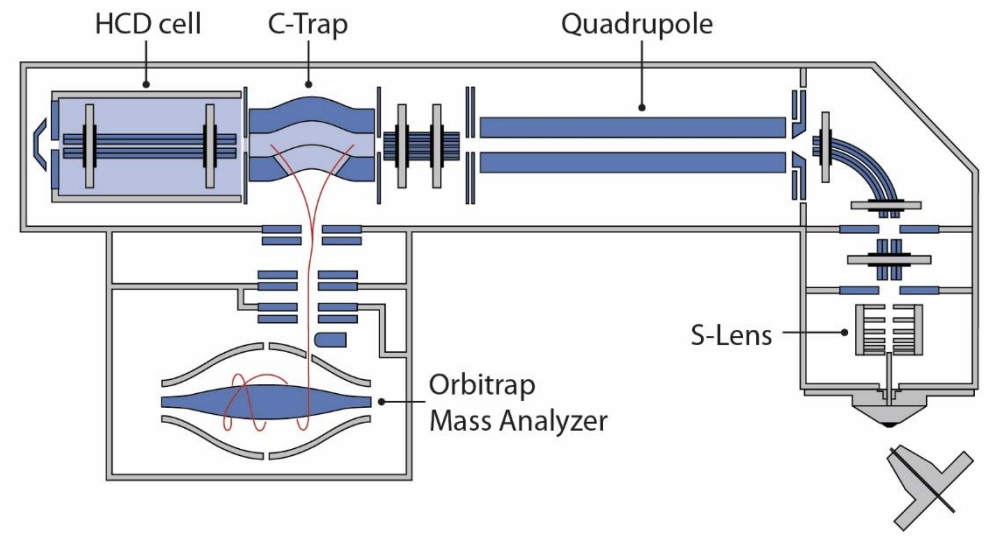


Figure 1.10: A commercially available Q-Exactive Orbitrap mass spectrometer. Figure courtesy of Thermo Fisher Scientific.

Another Orbitrap (LTQ) fitted with a linear ion trap is utilised for proteomics, lipidomics and post-translational modification analysis. The front end of the instrument is similar to the Q-Exactive with the exception that instead of the bent flatapole it has a linear quadrupole followed by an octapole for improved ion focusing. Ions are trapped in a linear ion trap and can be directed to HCD or collision-induced dissociation (CID) cell offering tandem MS capabilities.

1.4 Mass spectrometry of protein complexes

The vast majority of studies using MS as an analytical tool for structural elucidation of proteins rely on proteomics-based approaches.[42] In these experiments the protein is denatured and either enzymatically digested to small peptide fragments in solution, as in bottom-up methods, or analysed as one polypeptide chain that is sequentially fragmented inside the mass spectrometer, known as the top-down approach. The generated fragments are sequenced and matched with a database known fragments for protein identification. These methods also allow for identification of post-translational modifications.[66] However, as the proteins are denatured their overall fold and non-covalent interactions are not preserved, preventing the technique from providing information regarding subunit, protein-protein and non-covalent ligand-protein interactions as well as the elucidation of complex stoichiometry and oligomerisation without covalent sample labelling.

In the last two decades non-denaturing MS has evolved as an outstanding technique to examine intact protein complexes in the gas phase while retaining their non-covalent interactions. Despite instrumental challenges, the first question to answer was whether biological complexes will be preserved under vacuum during the course of the experiment, or the absence of water and the required ionisation energy would be too limiting. Early studies

found that salt- or solvent-bound adducts were preserved, hinting that non-covalent interactions can be maintained under the vacuum in a mass spectrometer.[67] A major breakthrough was finding that specific ligand interactions can be maintained and observed.[68] In a series of studies it was shown that specific receptor-ligand binding and enzyme substrate binding can be monitored my MS. For instance, Gamen, Li and Henion found that the binding of tacrolimus and rapamycin to the FKBP receptor can be detected.[5] The same group also observed the first noncovalent binding of lysozyme with hexasaccharide as a substrate.[69] Katta and Chait presented haem binding to myoglobin under vacuum condtions,[4] whereas Robinson and co-workers showed acyl-CoA binding with a series of substrates.[70] These achievements benefited from early instrumental and methodological developments such as the invention of soft ionisation techniques and were followed by other applications with increasing sample complexity and biological relevance[71] (Figure 11). In addition they served as a catalyst for further instrumental optimisation.

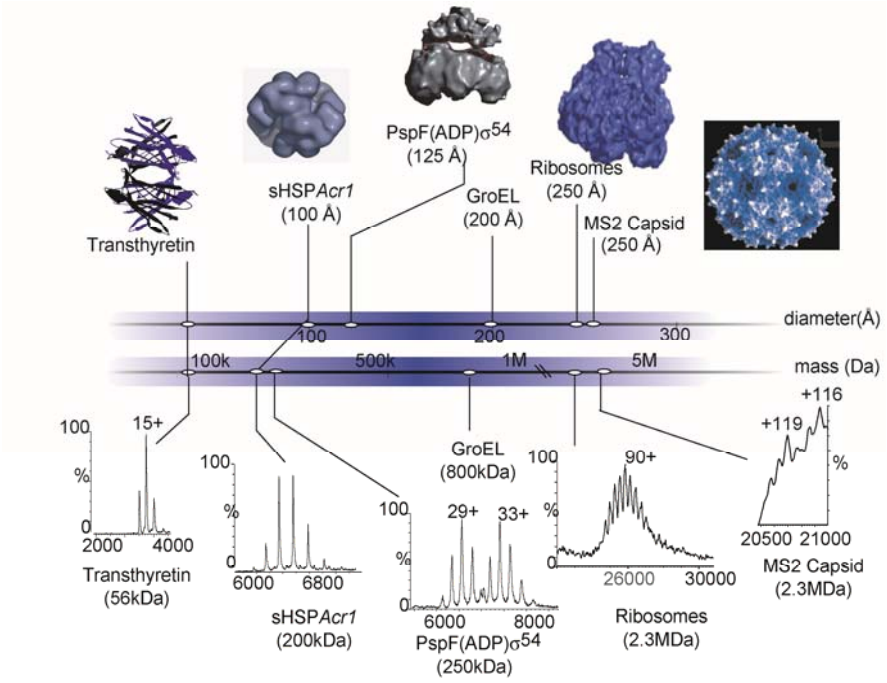


Figure 1.11: A summary of selected protein complexes analysed by mass spectrometry with increasing mass and complexity. Figure from Sharon and Robinson.[72]

1.4.1 Instrumental modification for transmittance of large ions

All early studies of large samples by MS employed soft ionisation techniques, namely ESI and MALDI, to lower the energy required for sample ionisation. ESI is performed in solution with minimal sample pre-treatment, however a low percentage of organic solvents or acids are usually added. In light of the increasing complexity of biological samples further instrument improvements were necessary. The high degree of solvation and heterogeneity, low transmittance and detection of large ions and the lack of high-resolution mass measurements were among the instrumental challenges that mass spectrometrists faced. As is typical in other scientific endeavours where success is not measured by solving a problem but by how many other problems the solution generates, ESI created additional challenges. For instance, relatively large volumes were required and even though the measurements were performed at low micromolar concentrations, characterisation of proteins with low levels of expression was difficult.

Originally ionisation was favoured by addition of acids and organic solvent. Elevated temperatures were applied to facilitate better desolvation. Despite this the ion efficiency is also relatively low, as only around 10 percent of species of interest are detected. To overcome some of these challenges a miniature ESI was developed. The so-called nanoelectrospray (nESI) can utilise a much lower sample flowrate (nanoliter per minute) as a result of a smaller micrometer range orifice capillary.[73] This adjustment not only requires less sample, usually 1 to 6 μL, but also reduces the size of the initial droplets leading to better desolvation.[74] As a result higher salt concentrations can be tolerated without the use of organic solvents or acids. As shown in Figure 1.12, data generated by nESI have fewer salt adducts and are better charge resolved.

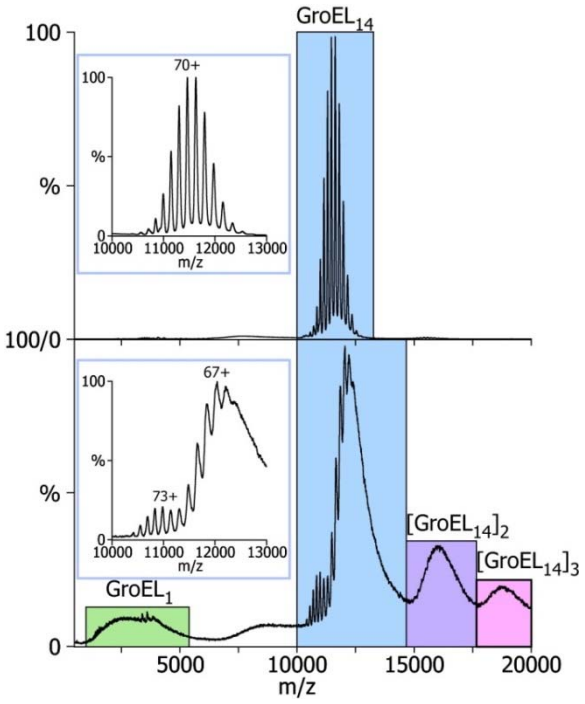


Figure 1.12: The effect of nESI on spectra quality. Top: spectrum of GroEL introduced by nESI which shows only the native stoichiometry. Each charge state is fully resolved allowing for a mass calculation with a high level of confidence. Bottom: spectrum from conventional ESI of the same sample. Non-specific peaks corresponding to higher oligomeric orders are detected. Peaks are not well resolved due to salt adducts and poor desolvation. Figure from Benesch et al.[32]

By spraying the same sample, the GroEL chaperone using both normal ESI and nESI, another major advantage of nESI is evident (Figure 1.12). In addition to better resolved spectra, the formation of artificial oligomers is also minimised. Due to a smaller initial droplet the probability of having two species trapped in the same droplet is lower, compared to conventional ESI where larger droplets lead to the formation of nonspecific adducts. Even with nESI at high concentrations the presence of artificial nonspecific oligomers can hinder analysis (Figure 1.13). As a consequence typically low micro-molar concentrations of sample are used. Taken together, nESI is less sample demanding, and allows characterisation from solution without denaturing substances. Based on these advantages all non-denatured spectra presented in this dissertation were recorded using nESI.

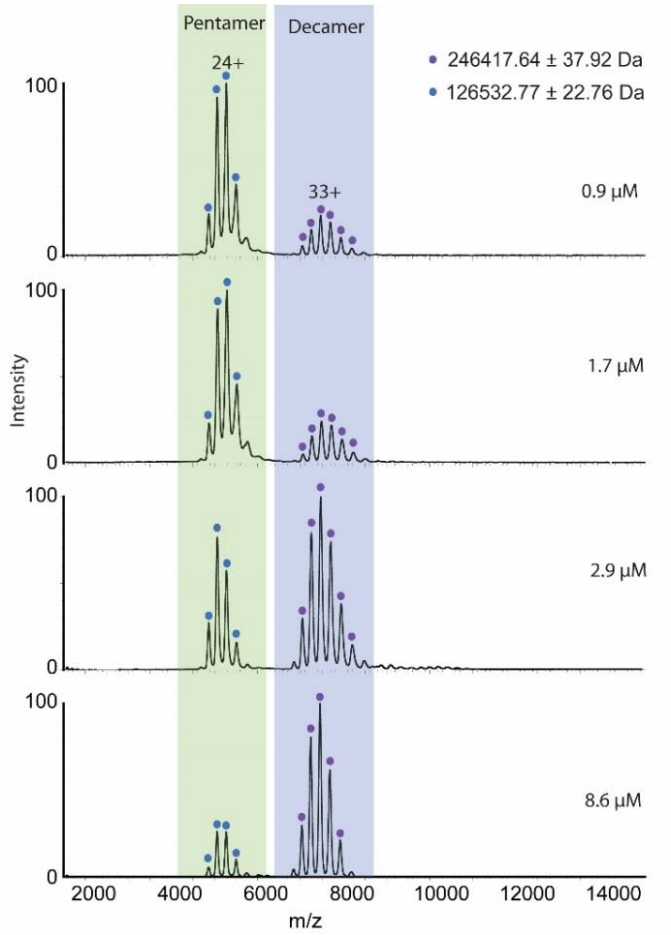


Figure 1.13: The effect of concentration on oligomeric state. Serum amyloid P component is known to form a pentamer in physiological conditions. A concentration series shows that predominantly pentamer is observed at low protein concentration and the amount of non-specific adduct formation increases with increase in concentration. Measurements were performed in 200mM ammonium acetate on a QToF instrument.

A major obstacle in MS of large analytes such as proteins is low transmittance and poor focusing of large ions. Ions are generated at atmospheric pressure whereas ion focusing by the first lens is performed at significantly lower pressure. This immediate change in pressure leads to ion acceleration, thereby increasing the ion kinetic energy. This increase in kinetic energy is very advantageous for the analysis of small ions as they are then better desolvated, however in the case of large ions the increase in kinetic energy makes them difficult to focus by the initial lenses. To overcome this problem instrumental modifications in the source

pressure led to improved transmittance of large ions.[75] Figure 14 shows a simulation of ion trajectories of a relatively large protein complex alcohol dehydrogenase (ADH, 146.8 kDa) at different source pressures applied on a quadruple operating in RF-only mode.[67]

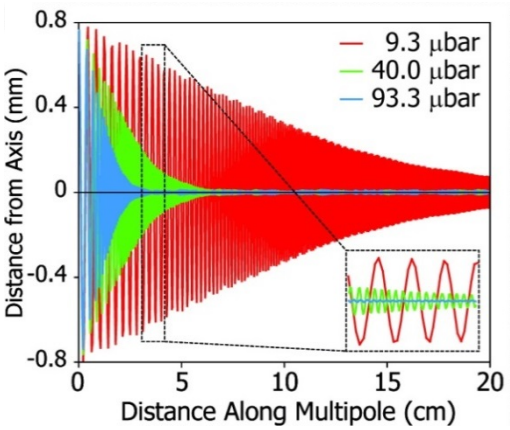


Figure 1.14: A simulation of a large ion being focused by a quadrupole as a function of pressure. Alcohol dehydrogenase ions are poorly focused at low pressures (red line); increasing the pressure allows for better focusing of ions (green and blue lines). Figure from Benesch et al.[32]

At low pressure the ADH ion trajectory has a large oscillation magnitude and ions are poorly focused within the length of the quadruple. Increasing the backing pressure results in a significantly narrower ion beam and better focusing of the ion packet. It is rationalised that higher pressure in the source region increases the number of collisions of large ions with gas molecules.[76] These collisions allow for some of the high mass ion's kinetic energy to be passed to the gas molecules, effectively lowering the kinetic energy of the large ion. With a reduced kinetic energy the large ions can then by better focused, leading to higher transmittance and a greater detection. The increase in the initial ion guide pressure is achieved by reducing the pumping efficiency of vacuum lines or adding an alternative gas in the ion source. Since ion velocity is decreased by collision this method is commonly known as collisional cooling, other scholars have also proposed collision focusing as a better term since by collisions the total energy is increased leading to ion heating rather than cooling.[67]

This discovery was instrumental for the ability of MS to be used for the structural elucidation of large ions over 1 megadalton in size.[76]

Mass analysers also need to be modified to detect large ions. It derives from Equation 1.3 that lower radio frequency will allow for analysis of larger *m/z* value ions, as will higher applied voltage or lower distance between the rods. However, operating at high voltage or lower distance between the rods leads to quadrupole discharge between rods with opposite phase. Conventional quadrupoles operating at high radio frequency limit the acquisition range to 4,000 *m/z*,[77] optimal for analysis of small ions. Lowering the frequency leads to better focusing of large ions and mass filtering of the small one, extending the acquisition range to 35,000 *m/z*.[74] Operating at this low frequency does come with a cost, as the resolving power of the quadrupole is significantly compromised. Due to this limitation of the quadrupoles, hybrid mass analyser systems are preferred for analysis of large ions.[78] For large mass ion analysis orthogonal TOF is found to perform better than a linear configurations. [76]

1.4.2 Structure elucidation in the gas phase

To find the identity of a compound, ions can be activated in a collision cell, producing fragments that can be mass analysed. These fragments can be thought of as being parts of a puzzle, which once placed in the right position give the whole picture; the identity of the compound in our case. The same approach is applied to large protein complexes where the fragments are usually stripped subunits. Maintaining the noncovalent interactions is important to know the mass of the complex as a whole, while activating the complex supplies information regarding subunit composition and connectivity, leading to complex stoichiometry.[79]

Several activation methods have been developed by taking advantage of different methods in which energy can be delivered to gas-phase ions. These include collision methods either with gas molecules or with a surface as in collision-induced dissociation (CID)[80] or surface-induced dissociation (SID)[81] respectively, radical formation via interactions with electrons as in electron capture dissociation (ECD),[82] or electron transfer dissociation (ETD)[83], and also by absorption of photons as in blackbody infrared radiative dissociation (BIRD)[84] or infrared radiation multiphoton dissociation (IRMPD).[85] Here, both CID and IRMPD are used to activate ions as is explained in more detail in Chapter 3.

The most frequently used method of activation is CID as it can offer sequential activation and is easy to implement. CID can be performed in the source region of mass spectrometers where ions are accelerated in the first ion guide, which is usually operating at higher pressures. A more desirable way, as it offers higher activation control, is performing CID in a designated collision cell, implemented as a quadrupole operating in RF-only mode filled with a neutral gas.[86]

Key mechanistic aspects of CID are well documented.[87] As the total ion energy increases, initially salt adducts are dissociated, then gas-phase rearrangement such as collapse or unfolding may occur, leading to dissociation by breaking of non-covalent interactions and ejection of subunits. At even higher energy, backbone fragmentation producing small ion fragments will occur. Typical of CID is asymmetric charge dissociation[88] (Figure 1.15). During CID a subunit is initially unfolded, and with further activation as it strips from the parent ion it carries a disproportionate amount of charge relative to its mass, and as a result highly charged monomers are observed in the low *m/z* region. This mechanism is dependent on the energy dissipation pathway. For instance, increasing the total charge density, by including a supercharging agent in the sample can alter the dissociation pathway resulting in

dimer or trimer formation (rather than monomer) and therefore a more symmetric charge distribution.[89]

This ability to fragment intact protein complexes in the gas-phase has been used to obtain information for a variety of systems. One of the most complex systems, which is crucial for understanding the mechanism of protein expression in eukaryotic cells is the translation initiator factor eIF3.[90] This thirteen subunit complex was analysed intact in the mass spectrometer enabling full characterisation of its architecture and stoichiometry. Since at present no atomic model for the full complex is known these results, obtained from MS alone although at low resolution, provide valuable insight into the number of subunits and their relative position. By activation of the complex it was found that this initiation factor is formed by three subcomplexes, and all subunit interactions were found. This allowed for a full map of subunit interactions to be created.

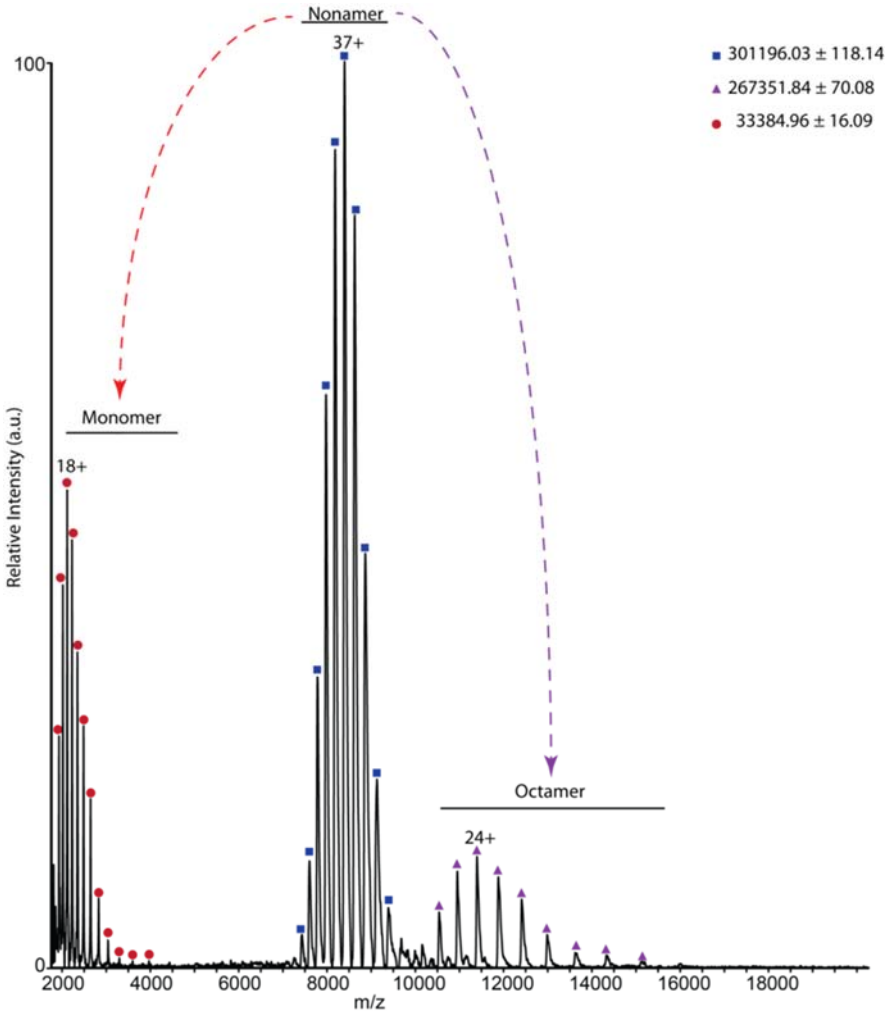


Figure 1.15: Collison-induced dissociation of the lysenin pore (13μM) solubilised in DDM micelles using 200 V. An asymmetric charge distribution is observed. Ejected monomers with one ninth the mass of the intact nonameric protein carry almost half the charge.

αβ-Crystallin represents another application where activation of an intact complex leads to a full characterisation of subunit interactions of a large oligomeric protein complex.[91] Due to a large degree of polydispersity the intact protein complex yields very poorly resolved spectra. Nevertheless the mass spectrum of the complex provided the first hint that even number oligomers are preferred over odd, and that oligomers range in size from ten to 40 subunits.[92] By activating the complex sequentially and performing tandem MS it was found that the building block is a dimer which interacts at the terminus to form larger oligomeric

states. These findings and many others have established MS as an analytical tool that provides insights into subunit organisation of protein complexes that are otherwise difficult to study.

1.4.3 Investigation of protein dynamics in the gas-phase

The ability to preserve non-covalent interactions creates a new opportunity to study the dynamics of protein complexes. Proteins are very responsive in nature; they interact with other proteins, peptides, drugs and other ligands. Characteristic of these interactions is the establishment of a binding equilibrium depending on the concentration and the time of incubation.[93] Both these aspects can be monitored by non-denaturing mass spectrometry.[94] For instance, a protein can be incubated with a known amount of a particular ligand and mass spectrometry performed at several time intervals (Figure 1.16).

During the course of time-resolved experiments the intensity of the ligand bound to the protein relative to the intensity of the unbound protein can be monitored to yield information about the kinetics of the interaction. Similarly, a protein can be incubated with different concentrations of a particular ligand and the intensity of the species present in the spectra can be examined to gain information regarding binding constants. For proteins with several subunits the association or dissociation constants can be derived by monitoring the intensity of peaks with mass corresponding to the sum of individual sub-complexes. In the case of homomeric systems chemically labelled monomers can be utilised.[95] Methods routinely applied for kinetic measurement by MS are summarised in Figure 1.16.

One of the major advantages of using MS to study protein interactions is its high sensitivity. Coupled with high accuracy this allows for the identification of all major interactions in a system. Individual binding events at low picomolar concentration can be

monitored, providing a substantial improvement relative to NMR or isothermal titration calorimetry (ITC), where for selected systems the concentrations used to perform the experiments may not be biologically relevant.[96] A major limitation however arises from using ESI. While other techniques are very tolerant towards buffer compositions, in the case of electrospray only highly volatile buffers, such as ammonium acetate can be used.[78] In addition, harsh gas-phase activation that yields better resolved spectra may not preserve the binding interactions from solution. Small ligands bound to large protein complexes also present a challenge. Not only does the resolution decrease with increasing *m/z* value but large ions also have higher charge making small ligands difficult to resolve in *m/z* scale.

To overcome these challenges several approaches have been developed. For example, the utilisation of a proxy protein with known affinity for the ligand, which appears in the low *m/z* region, has been reported.[96] The variation of the intensity in the ligand bound to the proxy protein peak can provide relative information to the binding of a ligand with a large complex. Alternatively, by performing tandem MS on the poorly resolved ligand bound peak, individual ligand ions can be released from the complex.[97] The intensity of the ligand peak can then be monitored as a function of time or concentration. If the dissociation of the ligand from the complex is uncontrolled then the dissociation constant (K_d) will be overestimated. It is worth mentioning that these alternative techniques should be supported by controlled experiments to insure that solution interactions are well preserved in the gas phase. One of the most striking examples of monitoring small molecules bound to a large protein complex is the binding adenosine triphosphate (ATP) with a mass of 507.2 Da to the 800 kDa molecular chaperone GroEL.[98]

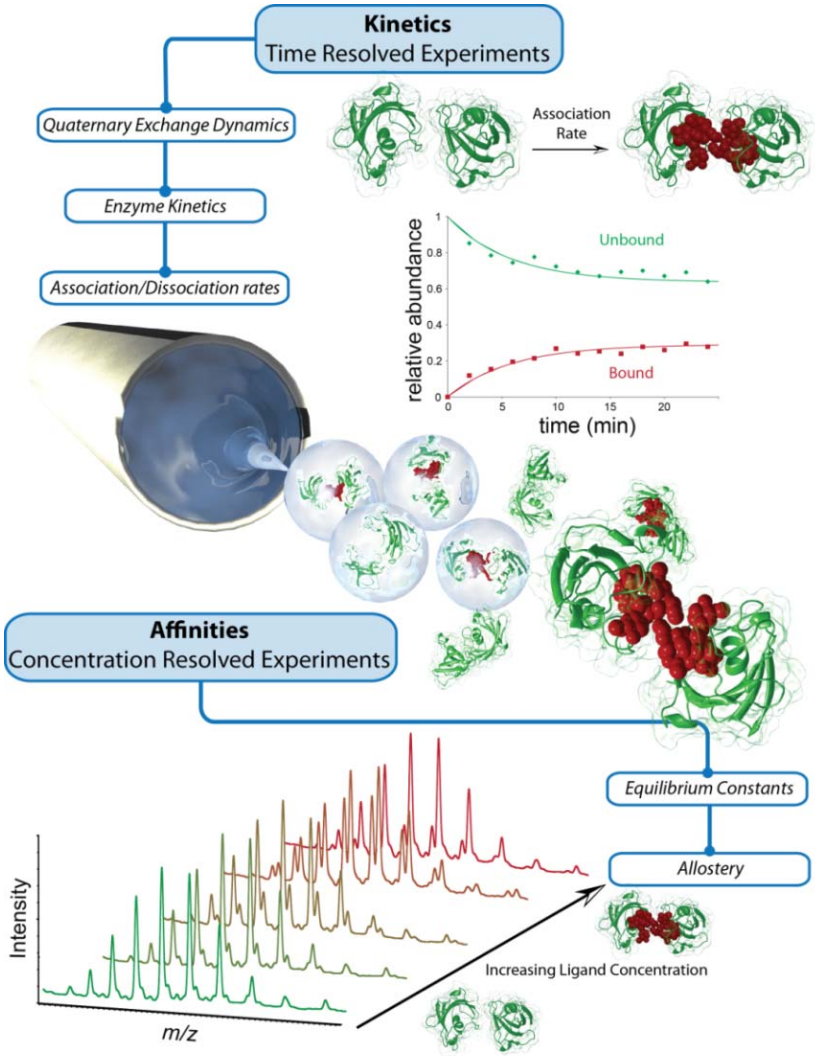


Figure 1.16: Experimental approaches for measuring protein dynamics by mass spectrometry. Time resolved experiments allow for enzyme kinetics, on/off rates or subunit exchange to be monitored. Equilibrium constants can be derived from concentration driven approaches. Figure from Hopper and Robinson[99]

MS-based methods are particularly useful for measuring kinetics where multiple proteins and ligands interact. Each complex can be identified and multiple K_d values can be measured simultaneously. The activity of HSP90, one of the most abundant heat shock proteins, is highly dependent on binding with other chaperones such as HSP70, Hoo and FKBP52. By adding each individual chaperone sequentially the full interaction cycle of HSP90 binding to co-chaperones was elucidated.[100] Due to the presence of multiple interactions highly complex

spectra were recorded that could yield adequate information after deconvolution. The K_d values were measured for all species in equilibrium resulting in 14 different K_d values for the total complex. Taken together this study demonstrates the power of MS methods to study the interactions of protein complexes in a more convoluted but realistic scenario where many interactions are intertwined.

1.4.4 Using IM-MS to monitor protein structure in the gas-phase

A key criterion of a structural biology tool is the preservation of protein structures and complexes as they occur in the native environment. A breakthrough in MS was finding that during a typical experiment the structure of proteins is generally preserved even in the absence of water.[97] This was determined by using IM-MS to measure the CCS of proteins which were found to be in parity with the values calculated from atomic coordinates. Furthermore, that protein structure is preserved in the gas phase agrees with other indirect measurements such as the preservation of activity of enzymes and viruses, or preservation of global structure visualised by EM, of proteins that have previously been analysed by MS.[101,102] These observations show that within the timescale of a typical IM-MS experiment the folded protein structure can be preserved, at least to a good approximately and in a reversible manner, in isolation from an aqueous environment.

IM-MS allows for different conformations to be observed on the timescale of a MS experiment providing an opportunity to measure conformational dynamics. By activating ions prior to IM measurement the gas-phase unfolding pathways can be monitored. Since ions are usually activated by collision with neutral gas molecules, collision-induced unfolding (CIU)[103] can monitor the conformations of proteins in the gas phase, to determine the number of intermediate states and identify specific ligands that can induce conformational changes or

exhibit a resistance to unfolding. The ability to monitor the conformational landscape provides information regarding the intrinsic physiochemical properties of protein complexes. By being able to isolate a protein from its native environment while maintaining its 'native' folded-like state, the effect of hydration, salts and other cofactors present in the environment on conformational dynamics can be reported. The structural dynamics can be monitored on a millisecond timescale, providing insights into the conformational heterogeneity of the sample.[104]

Unlike many analytical techniques where only an average of all present states can be measured, in a typical IM-MS experiment several conformations can be separated and analysed simultaneously.[7] This provides a unique opportunity to explore the effect of temperature, pH and other stress conditions on the stability of structural conformers. In addition to conformational dynamics, each unfolded intermediate can also be explored. Since these intermediates are intrinsically transient, they are challenging for other analytical techniques. Importantly, using IM-MS, information regarding protein folding or even the number of domains can be derived from recording the number and behaviour of gas-phase intermediates.[105]

1.5 Membrane protein complexes

Proteins that are present within biological membranes are known as membrane proteins. They regulate the membrane composition and dynamics and at the same time play a crucial role in traffic of solutes.[106] Around 30 percent of the genome encodes for membrane proteins and they are drug targets of more than 50 percent of present therapeutics.[107] A characteristic feature of membrane proteins is the transmembrane domain, formed predominately by outward-facing hydrophobic residues. Unlike soluble proteins where hydrophobic residues

are located in the centre of the protein and polar residues are in direct contact with the aqueous environment, in the case of membrane proteins the transmembrane region is in direct contact with the lipid environment, with polar and charged residues predominantly positioned in the regions of the protein outside of the membrane.

Membrane proteins are classified as integral or peripheral depending on their interaction with the membrane. Integral membrane proteins are permanently embedded in the membrane; hence, once taken out from the hydrophobic environment they are very unstable and prone to unfolding or aggregation. They can be single hydrophobic peptides embedded in the membrane up to multiple transmembrane α-helices or β-barrels.[108] Peripheral membrane proteins on the other hand interact with the biological membrane temporally, typically to enhance their stability or induce a signal. In terms of function, membrane proteins are classified as transporters, receptors, channels and enzymes.[107]

On the quest to elucidate the chemistry of life a crucially important contribution has been the structural determination of proteins at atomic resolution. In the last fifty years a rapid expansion of known structures has been possible due to major advances in crystallographic techniques, and more recently by advanced microscopy techniques. However not all protein families have equally progressed towards structure elucidation. From a structural point of view membrane proteins should be easy targets for structure determination since they have well-ordered regions with similar chemistry. Nevertheless, the known structures of membrane proteins are far less than the soluble ones (Figure 1.17). Only 556 structures of unique membrane proteins have been deposited to date.

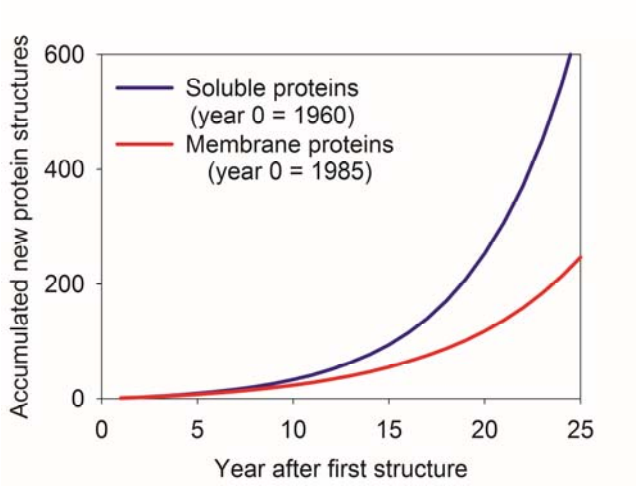


Figure 1.17: A time course of X-ray crystallography structure elucidation of membrane proteins. The red line shows the number of unique membrane proteins with known structures since the first membrane protein structure was deposited in the PDB. The blue line shows the equivalent numbers for soluble proteins. It is evident that the rate of structural characterisation has been faster for soluble proteins than membrane proteins. Data were taken from the membrane protein databank (http://blanco.biomol.uci.edu/mpstruc/).

Two reasons why progress towards structure elucidation of membrane proteins has been slow are low expression yield and the need for complex solubilisation environments. While overexpression of soluble proteins commonly yields large quantities, in the case of membrane proteins, protein folding happens in the membrane, hence, the amount of membrane proteins than can be produced is limited by the membrane itself and also by other pathways that are initiated for transport to the membrane and facilitation of correct folding, which are limited in the case of over expression.[109]

As part of a hydrophobic environment, isolation of homogeneous membrane protein samples is sometimes not feasible. To solubilise the membrane and isolate a particular membrane protein detergents are commonly used. Detergents have similar chemical properties to lipids, as they have both a hydrophobic chain and a polar head group. Due to their conical geometry they form a micelle rather than a bilayer, as in the case of lipids,

which is highly concentration dependent. The micelle is in equilibrium with individual detergent molecules, and is usually monodisperse in solution with a fixed aggregation number. Large quantities of detergent micelles are usually required to compete with bound lipids. Detergents will surround a membrane protein by placing their hydrophobic chains towards the transmembrane region hence protecting it from contact with the aqueous environment.

Although detergent solubilisation will significantly lower the complexity of a system, in practice the sample is far from being homogeneous. Frequently lipids co-purify with membrane proteins and increase the degree of heterogeneity, *vida infra*. In addition, although detergent can provide a stable environment for membrane proteins they often interfere with analytical methods therefore making membrane proteins difficult to study. Furthermore, each observed phenomena represents the kinetic and thermodynamics of detergent-solubilised membrane protein interactions rather than those that may occur in a more native lipidic environment.

1.5.1. Preserving intact membrane protein complexes in the gas phase

Initial attempts to preserve the native-like structure of membrane proteins were guided by predictions that the low dielectric constant of the lipid environment resembles the vacuum; hence membrane proteins may retain their folded structures once brought into the gas phase.[110] However, the degree of preservation of specific interactions as well as the means of delivering membrane proteins into the gas phase remained elusive. In addition, the effect of detergent upon ionisation was found to alter the charge state distribution for soluble proteins, suggesting membrane proteins solubilised in detergent may be difficult to analyse.[111]

The first method to deliver membrane proteins into the gas phase was using large unilamellar vesicles.[112] The results showed selectivity for anionic lipid binding to the bacterial K+ channel; however the subunit interactions were not preserved.[113] Since vesicles provided a challenge for ionisation other vehicles such as detergents were tried. The initial problem was the generation of poorly resolved spectra due to the presence of a high concentration of detergent. In addition, the ionisation efficiency was relatively low as the protein complexes had to compete for charging with empty detergent micelles. Lowering the micelle concentration prior to MS was found to be useful, as well measurement using non-ionic detergents.[114]

The first membrane protein complex sprayed at a detergent concentration lower than the critical micelle concentration (CMC) was EmrE.[115] Although the protein could be introduced into the gas phase and the spectra quality was enhanced, the preservation of the transmembrane domain in solution at low concentration of detergent was feasible. At the time a major emphasis was placed on spectra quality rather than stabilisation of the membrane protein complex itself.

A major breakthrough was to spray membrane protein complexes at above the CMC and to then remove the micelle once the complex is in the gas phase.[9] Barrera and co-workers showed that the ABC transporter $BtuC_2D_2$ can be maintained intact as a tetramer, with the preservation of both transmembrane and soluble domain non-covalent interactions. It was later rationalised that the micelle itself stabilises and protects a membrane protein complex on its transition from solution to the gas phase.[116]

Membrane protein samples are ionised as a membrane protein-detergent complex but without further activation an unresolved peak will normally be observed. Detergent is

removed from the protein in the gas phase by collisional activation yielding highly resolved spectra. This method has made MS a suitable technique to study membrane proteins and means it can provide valuable information regarding their structure and dynamics, as it does for soluble proteins (Figure 1.18).

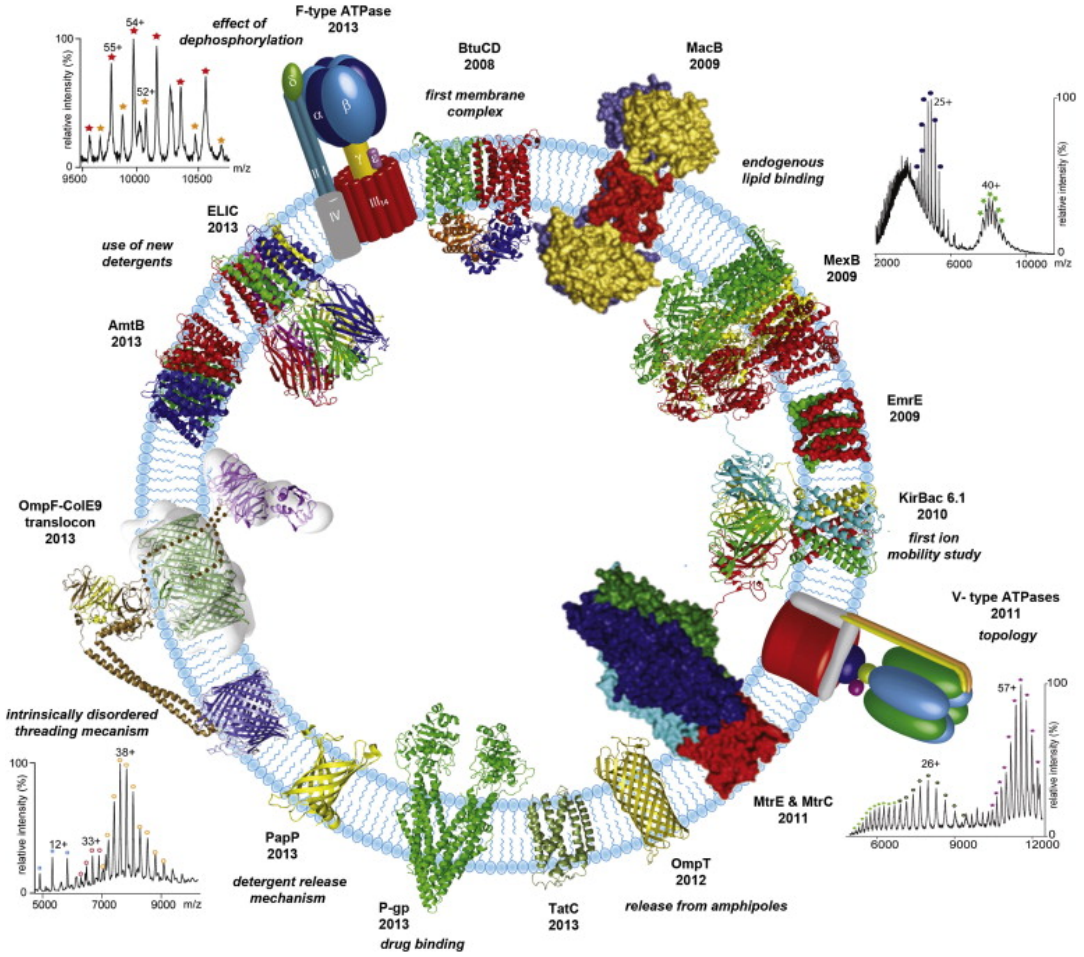


Figure 1.18: Membrane protein mass spectrometry in the last six years. Since the first spectra of membrane proteins an increasing number of membrane proteins with high complexity have been studied by mass spectrometry. Figure from Marcoux and Robinson[117].

1.5.2 Expressing membrane proteins and making them MS compatible

The original results from the MS of membrane proteins prompted the need for establishing expression and purification protocols of membrane proteins tailored for MS

characterisation. It was evident from very early that MS provided a high resolution view of sample purity. Samples that were characterised as crystallographically pure, frequently from inspection of size-exclusion chromatography (SEC) peak shapes, were often found to be heterogeneous by MS, due to the presence of lipids or other low mass cofactors that cannot be directly identified by other techniques.

Due to the often low expression yields of membrane proteins many established overexpression protocols are focusing on producing the highest possible quantities. For instance, many outer membrane proteins are overexpressed via inclusion bodies which yield large quantities; however the oligomeric state of refolded proteins were usually lower than the membrane embedded ones. For mass spectrometry, where the oligomeric state of the proteins is measured, expression and purification protocols utilising protein isolated from the membrane is preferred. Drew and co-workers have developed protocols for overexpression of membrane proteins from bacterial cell lines that can be adequately adjusted for application to MS.[118,119] For better visualisation of expression green florescent protein (GFP) fused constructs are frequently used to optimise the expression parameters and at the same time monitor the amount of protein present in the membrane.

The optimisation of membrane protein purification protocols, as well as MS operation conditions, are usually required to achieve well-resolved spectra of membrane proteins.[64] Detergent screening is performed to choose which detergent to use to solubilise and isolate membrane proteins from lipid bilayers. Based on the degree of delipidation in resulting mass spectra detergents are classified as harsh, which usually strip the membrane proteins free of lipids, or mild where a protein-lipid complex is isolated. Harsh detergents are usually preferred for MS of membrane proteins because they are associated with fewer lipid bound species.

Depending on the stability of membrane proteins delipidation can lead to disruption of the native oligomeric state and unfolding. Hence, the strength of detergent as well as the concentration are two crucial parameters that need to be optimised for each membrane protein studied. For cases where no detergent can preserve the native fold other vehicles that better mimic the native environment have recently been shown to be MS compatible. In this case membrane proteins are solubilised in nanodiscs, native-discs, amphipols and bicelles (Figure 1.19). As in the case of detergents, the membrane proteins are released via collisional activation. Although the energy required is usually higher than for detergents, by preserving the lipid environment the native oligomeric state can be maintained as shown for the case of DgkA.[120] The utilisation of nanodiscs and native-discs provides a better framework for characterisation of lipid protein interactions especially in the case of native-discs where endogenous lipids are co-purified. Another factor that plays an important role in spectra quality is membrane protein compatibility with the volatile buffers necessary for nESI.

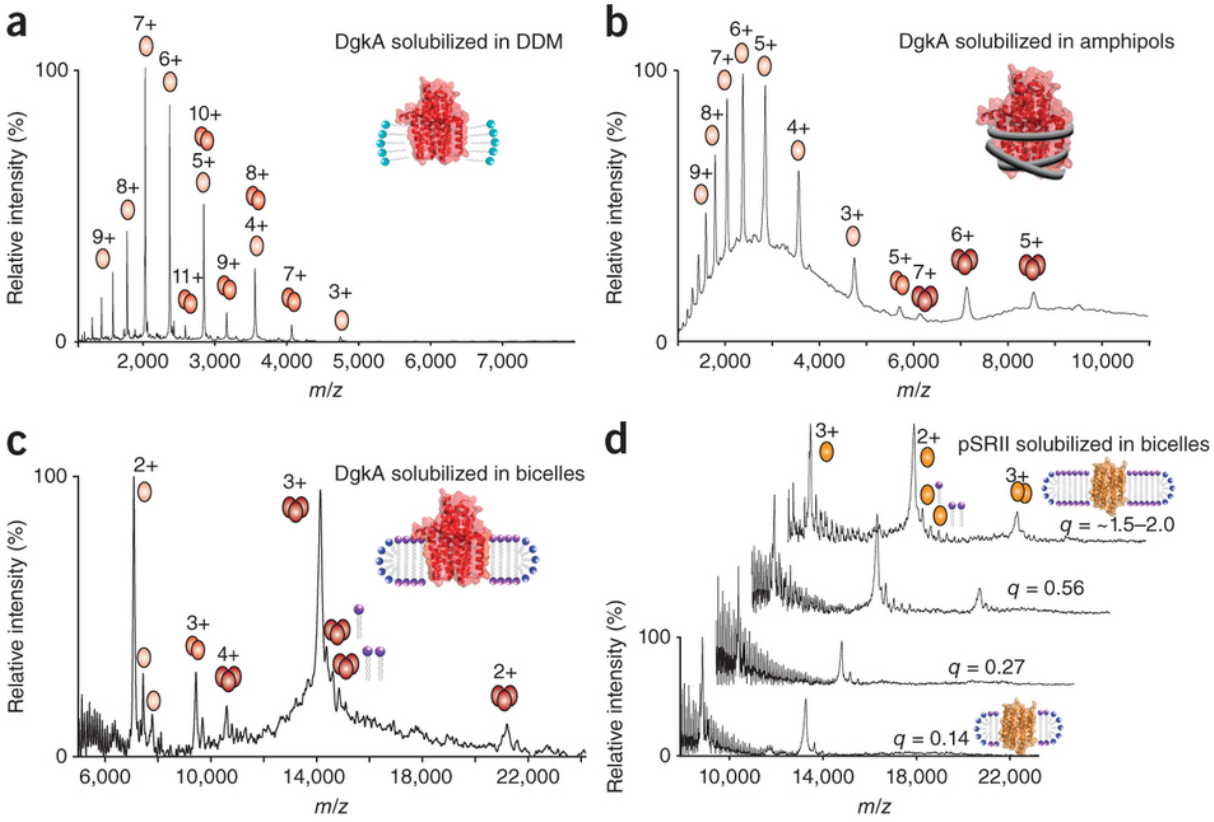


Figure 1.19: Vehicles that can be used to transfer intact membrane proteins from solution to the gas phase. (a) Mass spectrum of DgkA solubilised in DDM. Due to the activation energy for detergent removal, monomers are the predominantly observed species. (b) Amphipols are also compatible with MS characterisation. (c) More native-like environments such as bicelles can be used, resulting in high intensity spectra for the physiological oligomeric state of DgkA and pSRII. (d) The longer the length of the bicelles (q) the better the preservation of the physiological quaternary structure. Figure from Hopper et al.[120]

1.5.3 Using IM-MS to monitor the gas-phase dynamics of membrane proteins

IM-MS of membrane proteins can provide unprecedented information regarding protein conformational dynamics. Membrane proteins can be highly flexible, and can exhibit several conformations within the lipid environment. High resolution techniques often provide information either on the average conformation or only the most stable one, whereas IM-MS potentially measure all gas-phase conformations and information about the solution phase can be deducted. Furthermore, these conformations can also be reported upon stimulation, such as

with the addition of small molecules, or by change in temperature or pH. An early example of the effects of therapeutics on membrane protein conformation was shown by the binding of cyclosporine A (CsA), nucleotides and lipids to a multidrug efflux pump P-glycoprotein (P-gp).[121] Initially it was established that both inward facing and outward facing conformations can be distinguished by ion mobility. However, in a typical membrane protein MS experiment the large amount of energy used to remove the micelle can potentially lead to protein unfolding. Consequently distinguishing between a large conformer and unfolded species is crucial in interpretation of IM spectra.

Experiments where the equilibrium can be influenced by binding are desirable. For instance in the case of P-gp it was found that adding ATP or CsA separately has no effect on the equilibrium between the two conformations, whereas the addition of both CsA and ATP shifts the equilibrium towards the outward facing conformation. A similar effect is also achieved by combined addition of cardiolipin and CsA. By monitoring the entire conformational landscape a possible mechanism of drug efflux was proposed where nucleotide binding stabilises the outward conformation upon drug binding, leading to drug release.

Crucial for measurements is the realisation that different conformations can be monitored by IM-MS if the equilibrium between different conformations is on the millisecond time scale. Conformational change due to lipid binding happens, usually, on the second time scale making them possible to be probed my IM even for large complexes such as ATPase. This large protein complex is composed of a large soluble head group and a membrane rotary ring where ATP is synthesised by taking advantage of the proton motive force. The membrane-embedded complex is formed by three sub-complexes I, C and L. Non-denaturing mass spectrometry reported that subunit I is displaced at low ATP concentrations. By measuring

the arrival time distributions of the ICL complex and comparing with CL complexes in the presence of endogenous lipids, it was found that the gaps left by movement of the I subunit are filled by lipids. Furthermore, the I subunit is connected with the peripheral stack which controls the assembly/disassembly of the complex. A hybrid methodology using IM-MS and molecular dynamics (MD) simulations showed that upon ADP binding the I subunit is more flexible leading to complex disassembly, whereas in the presence of a proton motive force the flexibility of the I subunit is restricted. As a result the assembly of the ATPase is modulated.[104]

PTMs play an important role in activity and function of membrane proteins. Many signalling pathways are regulated by phosphorylation or acetylation of specific residues. Applying a hybrid approach where the modified sites are identified by bottom-up proteomics and global conformational changes are monitored by non-denaturing mass spectrometry has yielded astonishing results. Schmidt and co-workers identified phosphorylated residues on eukaryotic F-type ATPase.[122] Of interest were the modified sites on the subunit interfaces which potentially can influence the oligomeric state. Dephosphorylation of these residues revealed that the affinity for nucleotide binding was lower relative to phosphorylated conditions. In addition, major structural differences between these interfaces were found in both regimes. In combination with comparative crosslinking it was further shown that the structure of subunit ε, which connects the soluble head region of the ATPase with the rotary membrane ring, is altered by phosphorylation. Taken together these results demonstrate the ability of MS to explore the effects of PTMs and lipids on structure changes on membrane proteins. Here similar approaches are used to probe the effect of PTMs on the oligomeric state of cytochrome *c* oxidase.

1.6 Aims of this dissertation

A breakthrough in gas-phase structural biology was the realisation that protein structure is preserved in the absence of water during a MS experiment.[97] In light of this finding and instrumental development the range of proteins studied by MS dramatically increased. The ability to perform measurements in near physiological conditions and observe the binding of small molecules to large biological complexes has made MS a desirable method to study protein complexes. In tandem with IM, structural information such as gas phase conformation, oligomeric state and subunit connectivity can be determined.[123] Many of the developments for soluble proteins have readily translated to the study of membrane proteins by native MS. However, study of this special class of proteins can be complicated, which is also the case when they are studied by MS.

Not all membrane proteins can be easily studied by native MS. Whilst detergent solubilised membrane proteins are most often used, depending on the detergent there is a large variability in spectral quality and the types of MS experiments that may be able to be performed. Therefore the initial aim was to determine the effect detergents have on the gas-phase properties of membrane proteins; specifically the role detergents play in determining the charge state of membrane proteins and the activation energy required for detergent removal.

Identifying and quantifying the effects of lipid binding to membrane proteins may facilitate a better understanding of this tightly-coupled relationship. A powerful method used in MS for the characterisation of bound ligands is mass selection and activation of the selected ions, yet commercial instrument designs limit its applicability to soluble proteins. The second aim was to apply new methods to enable mass selection of membrane proteins for

identification of bound lipids, as well as to quantify the protein stabilising effects of lipids in the gas phase.

Cellular membranes, and their membrane protein constituents, regulate many important biological processes, but can be prone to experiencing environmental stress. Elucidating how the chemistry behind membrane proteins and their interactions with lipids may be modulated by changes in environmental conditions was the motivation of the third aim. How lipid-protein interactions may respond to pH stress and identifying if specific lipids could stabilise the active conformation of membrane proteins at low pH was investigated.

Finally, aspects of the methods previously developed were combined to perform non-denaturing MS on a large complex sample. The aim was to determine conditions that preserve the non-covalent interactions of the native oligomeric state of the mammalian membrane protein cytochrome *c* oxidase, to identify the bound endogenous lipids, and to quantify the effect of these lipids and post-translational modifications on the protein stability.

In summary, the research presented here aims to elucidate the effect detergents have on membrane proteins in the gas phase, to identify bound lipids and the PTMs of membrane proteins, and the effects they have on oligomeric state and structural stability.

Chapter Two: Probing the mechanism of membrane protein ionisation and detergent release in the gas phase

The ability to maintain non-covalent interactions in the absence of water has meant mass spectrometry can be used as a tool in ways that were perhaps unimaginable not so long ago. Of particular relevance, that the structural integrity of proteins can be preserved is an important prerequisite when applying mass spectrometry as structural biology tool. To conserve the native-like folded conformation, in the presence of charged particles, while in a state with an increased total energy, a better understanding of charge-driven unfolding and collisional activation unfolding mechanisms is necessary. Although during collision-induced dissociation (CID) process these two unfolding pathways are intertwined, as ion acceleration is charge dependent, in the present study the unfolding effects arising from each pathway are decoupled.

Charge driven unfolding, also known as Coulombic unfolding, is directly affected by the ionisation mechanism. For membrane proteins, which are introduced to mass spectrometer in detergent clusters, an additional level of complexity is introduced. The total amount of charge will not only depend on the characteristics of complex but also on the detergent-protein interactions. In addition, removal of detergent is performed at elevated energies, relative to the energy typically used for soluble proteins, hence a balance between detergent removal and protein unfolding must be established.

In this chapter the focus is placed on the proposed mechanisms of electrospray and their applicability to membrane protein native mass spectrometry. A series of membrane proteins, covering a large mass range is used to explore the effect of detergents on charge distributions and the activation energy required for detergent removal. A potential mechanism of

membrane protein electrospray ionisation and gas-phase unfolding is proposed. Based on this mechanism, conditions to maintain membrane proteins in a native-like state were found, and used to probe the effect of lipid on membrane protein gas-phase stability.

2.1 The mechanisms of electrospray ionisation

The macroscopic features of electrospray are well known,[25] however, a detailed electrospray ionisation (ESI) mechanism remains elusive. Currently, a few ionisation modules exist,[68] that mechanistically describe the transition from charged droplets to protein ions. Two competing mechanisms are generally more accepted, known as the ion evaporation model (IEM),[124] and the charged residue model (CRM).[16] A third mechanism, the chain ejection model (CEM),[125] has also been proposed, however, as the general consensus is that this model applies to denatured proteins, it is not discussed in detail in this dissertation. Recently, there has been a growing consensus that combining aspects of both mechanisms could lead to a better description of the electrospray process.[126]

Characteristic of ESI mass spectra is the presence of multiple charged ions.[15] In the early stages of ESI, droplets with a large diameter can accommodate an excess amount of charge on their surface. Droplet evaporation and Coulombic fission leads to smaller charged droplets and eventually charged proteins. IEM proposes that as the droplet evaporates and the charge density increases, Coulombic repulsion becomes stronger than solvation forces, resulting in emission of charged particles and eventually ejection of the charged macromolecule from the droplet (Figure 2.1).[127] While this mechanism can account for the charge state distribution of low molecular weight ions, it cannot adequately describe ion formation of large macromolecular assembles.[128]

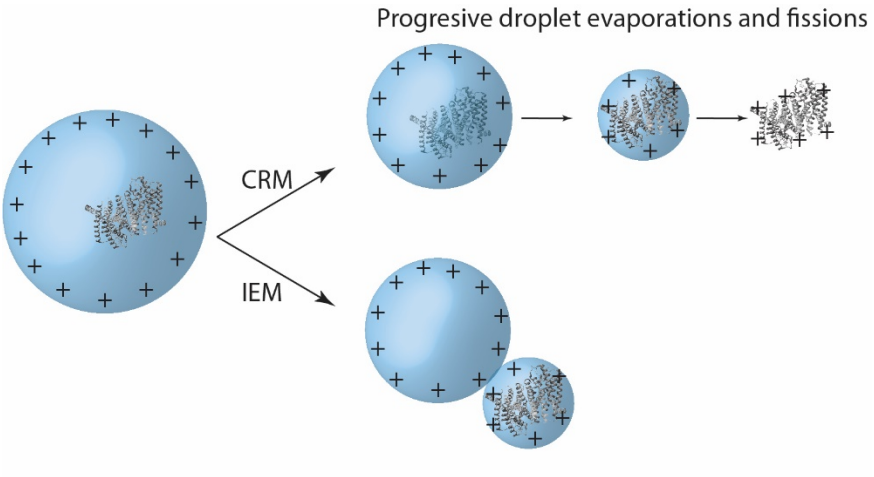


Figure 2.1: Proposed models of electrospray ionisation. The charged residue model proposes that solvent (blue area) evaporates as neutral and the charge is then passed on to the protein (grey) whereas following the ion evaporation model charged macromolecular ions are ejected from the droplet. Positive ions are denoted with +.

On the other hand the CRM proposes that solvent evaporates as neutral until the charge of the droplet matches the Rayleigh limit, where the Coulombic repulsion overcomes the surface tension, leading to droplet fission and the generation of smaller droplets.[24] When a single analyte is present in the droplet, complete neutral solvent evaporation will lead to deposition of the charge from the droplet surface to the protein (Figure 2.1). It is evident that following the CRM the maximum change (q_{max}) will depend on the Rayleigh limit,[129] where D is droplet diameter, γ is surface tension and ε_0 is the electrical permittivity of a vacuum:

$$q_{max} = \pi(8\varepsilon_0\gamma D^3)^{\frac{1}{2}} \quad (2.1)$$

If ion generation follows the CRM then the charge distribution will be directly correlated to droplet diameter. Indeed, de la Mora showed that this can be the case, as average charge states of proteins are consistent with the predicted charge state based on spherical droplets as a function of protein diameter.[126] Subsequently, this relationship has been found to also be true regardless of the presence of basic or acidic residues present on the protein.[130] Clearly

protein surface area dictates the amount of charge on a protein rather than the chemical characteristics of its primary structure. On the basis of this ionisation mechanism the surface area of proteins with unknown structure could potentially be derived by MS measurements.[131] However, this model does not account for cases where charge reducing or increasing agents are present in the sample.

With the development of non-denaturing mass spectrometry it became apparent that folded proteins generate ions with less average charge and a narrower charge state distribution than proteins denatured in solution (Figure 2.2). This difference under non-denaturing conditions arises because different ionisation mechanisms perhaps operate in each case. For instance the CEM for denatured proteins and the CRM for folded ones;[67] providing evidence that possibly both ionisation mechanisms are relevant.

A charged residue-field emission model has been proposed by Hogan and co-workers which combines aspects of both the CRM and IEM.[132] In this model, excess charge is extracted from the droplet by solvated ions that serve as charge carriers. Once the ions have fully evaporated the remaining charge is transferred to the protein, generating multiple charged ions. In this mechanism, what dictates the charge state of a protein is the extent of ion emission from the droplet. Ion emission happens only at electric fields above a certain threshold. Once an ion has been emitted, the electric field decreases. Subsequently, neutral evaporation will re-establish the high electric field on the droplet surface leading to the next emission event. This process occurs on a fast time scale and as such the electric field is effectively maintained at the threshold value. Hence, the final droplet will still have a charge state that is dependent on the surface area of the analyte. The charged residue-field emission model provides an adequate prediction of charge state distributions and can be applicable to all analytes regardless of their mass and folded state as shown by IM measurements.[133] The

combined charged residue-field emission model is useful in the case of membrane proteins, where it can be used to explain the effect of detergent on charge state distributions.

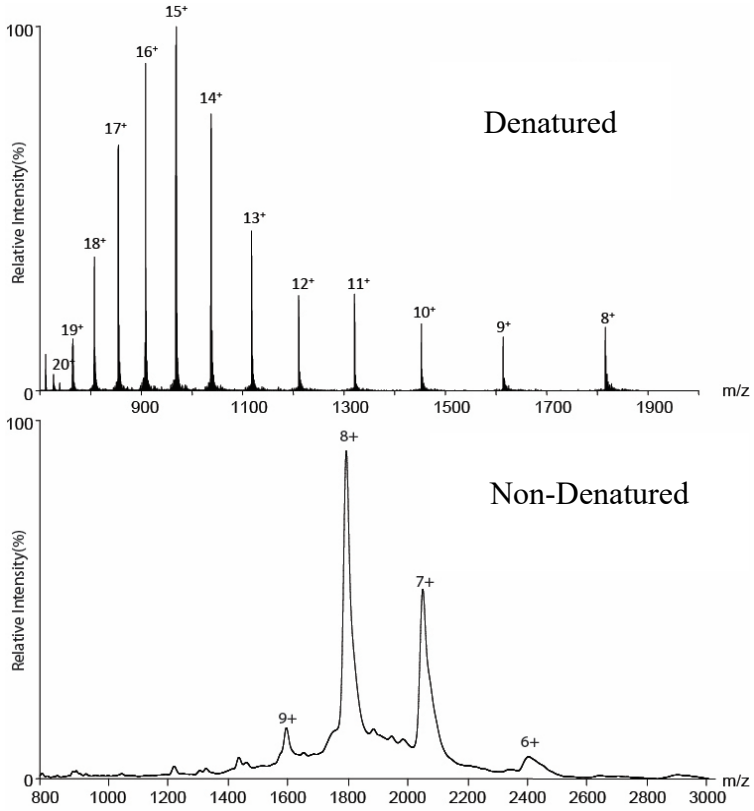


Figure 2.2: A comparison between the charge state distribution under denaturing conditions (top spectrum) and non-denaturing conditions (bottom spectrum). GlnK protein at 10µM was either denatured under acidic conditions or maintained in the folded state in native-like conditions using 200mM ammonium acetate and applying a low capilary, cone and collisional voltage of 20 V, 0.8kV and 10V respectively.

Membrane proteins are introduced to the mass spectrometer, typically, in the presence of detergents, hence, the effect that detergents have on charge states and gas phase stability of membrane proteins was investigated. Initially the ability of detergent to serve as a charge carrier was investigated by recording the mass spectra of the membrane protein *Pf*MATE solubilised in different detergents. Following established protocols for detergent screening,[64] optimised conditions, i.e. highly resolved spectra at low activation energy, were found.

Another factor that influences the charge state distribution of membrane proteins is solvent accessible area. To investigate the combined effect of detergents and solvent accessible area a set of membrane proteins, expressed in-house, were solubilised in a series of detergents from different classes and analysed by mass spectrometry. The activation energy required to remove detergents was measured for the membrane proteins *Pf*MATE and AqpZ, identifying detergents that can be removed at low activation energy and preserve the native-like conformation of these membrane proteins.

2.2 Detergent screening for optimisation of mass spectral quality

While detergents facilitate generation of gas-phase membrane protein ions[9] they severely reduce the spectral quality.[134] A prerequisite to reveal the membrane protein ionisation mechanism and probe the stability of membrane protein ions in the gas phase is generation of well-resolved spectra where the effect of each detergent bound can be monitored. To improve spectra quality, while maintaining the membrane proteins in folded-like conformations, a detergent screening protocol was developed.[64] Application of this protocol for spectra optimization of *Pf*MATE is shown. *Pf*MATE was expressed in-house in bacterial cell lines. After expression and isolation of membranes, detergents were utilised to solubilise the membrane protein.

The identity and the concentration of the detergent used for solubilisation are two crucial parameters in addition to time and temperature of extraction.[135] Detergents introduced at high concentration will compete with bound lipids and liberate the membrane protein from the lipid bilayer.[136] Low amounts of detergent are associated with a smaller extent of delipidation yielding heterogeneous populations of lipid-bound species that may not be resolved by MS.[137] However, in addition to solvation, lipids have been shown to also have a structural

role for many membrane proteins.[138] At high detergent concentration removal of structural lipids may lead to structural changes of membrane proteins and to protein unfolding. Hence, extraction of membrane proteins from lipid bilayers is done using detergents at a critical solubilisation concentration.

To maintain the native fold, the chemical properties of detergent and their ability to mimic the lipid bilayer are crucial.[139] Detergents that are better at out-competing lipids are classified as harsh.[135] Usually, these detergents are charged, have small head groups, and are typically short chain with <8 alkyl repeats. Although they perform better than mild detergent for sample delipidation, they mimic the native environment to a lesser extent meaning that many proteins are not stable when solubilised in these detergents. For instance, it has been estimated that only 20 percent of membrane proteins are stable in LDAO,[140] a zwitterionic harsh detergent.

Detergent with a neutral and large head group and long alkyl chains are known as mild detergents. They assist membrane proteins by providing more native environments however lipids are usually co-purified.[135] The most used detergent for MS studies is n-dodecyl β-D-maltoside (DDM). As a mild detergent, typically membrane proteins are solubilised at 20× the critical micelle concertation (CMC) whereas MS spectra are recorded at 2× CMC. Usually, although highly sample dependent, adduct peaks, potentially lipid bound species, are present in the spectra, as shown in Figure 2.3. For example, *Pf*MATE fused with green fluorescence protein (GFP) was found to co-purify with adducts with an average mass of 1,390 Da, possibly corresponding to cardiolipin (CDL). In addition, elevated energy in the collision cell was required to remove the micelle (190 V); hence, the native folded protein in solution most likely adopted an unfolding conformation during collisional activation. An impurity with a mass of 36 kDa was also co-purified.

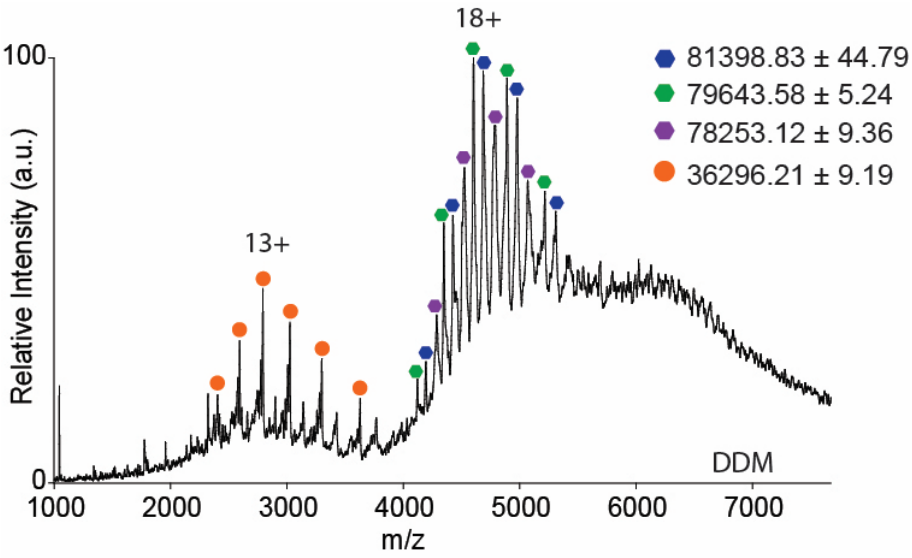


Figure 2.3: Mass spectrum of *Pf*MATE-GFP solubilised in DDM. The presence of lipid bound, potentially cardiolipin, is observed and impurities are also co-purified. The spectra were recorded at 190 V collision voltage. Prior to MS measurements the protein was buffer exchange to 200 mM ammonium acetate and 2xCMC of DDM.

To improve sample homogeneity and spectra quality a detergent screen was performed for *Pf*MATE. There are three steps where detergent can be exchanged: at extraction, during immobilized-metal affinity chromatography (IMAC) purification or during size-exclusion chromatography (SEC) purification. To eliminate the effect of detergent on spectra quality and only monitor the extent of delipidation during extraction, the first detergent screen was performed at extraction, with samples exchanged to DDM, using a SEC column, before MS analysis.

Figure 2.4 shows the spectra from a detergent screen at the extraction level. Six detergents were chosen from different detergent families. In all cases bound lipids are present and spectra are relatively poorly resolved, with the exception of foscholine-16 where spectral resolution allows for identification of adducts with a mass of 735 Da and 1,399 Da. These masses agree with the average mass of a phospholipid and cardiolipin respectively. Hence,

for *Pf*MATE, foscholine-16 can better preserve the binding of phospholipids, relative to DDM where only cardiolipin was found to be bound. For other detergents the poorly resolved spectra suggest that the protein is bound to a large lipid entity due to poor delipidation and these lipids cannot be unambiguously identified due to the low resolution. In all cases where the proteins were sprayed in the presence of DDM, a high activation voltage, up to 200 V in the collision cell, was required for micelle removal.

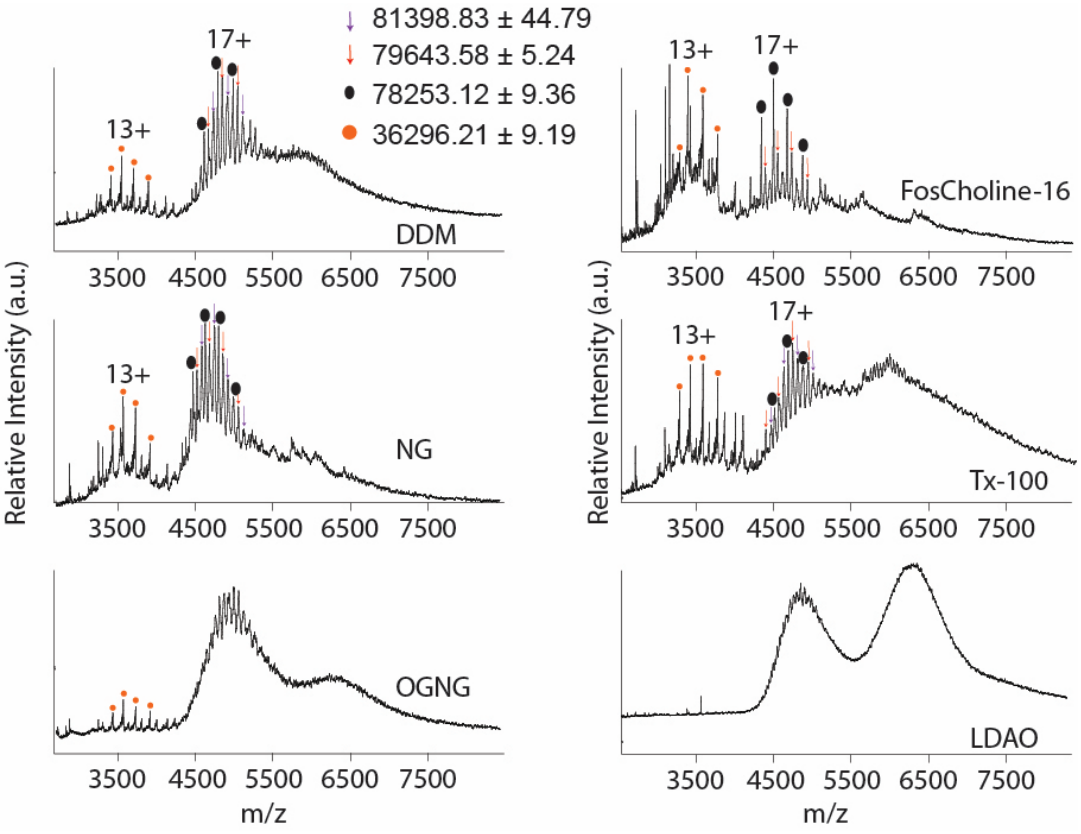


Figure 2.4: Mass spectra of *Pf*MATE-GFP demonstrating the detergent screening at extraction level. After extraction all samples were exchanged, using a SEC column, to DDM and 200mM ammonium acetate. The best resolved spectra were achieved at a collision energy of 190V and 200V for DDM, OGNG, LDAO and NG, TX-100,FosCholine-16 respectively.

Taken together these results demonstrate that detergent chemistry plays an important role in delipidation during extraction. The degree of delipidation can be monitored by non-

denaturing mass spectrometry, allowing for identification of detergents that can solubilise the proteins with minimal amount of lipid bound.

To improve spectral resolution and find detergents that can be removed at lower activation energies another detergent screen was performed. This time the same detergent was used for extraction and for MS analysis (Figure 2.5). A substantial improvement in the spectral quality was observed for OG, NG, OGNG and LDAO, and to a lesser extent for C8E4, as compared to DDM or Cymal 6.

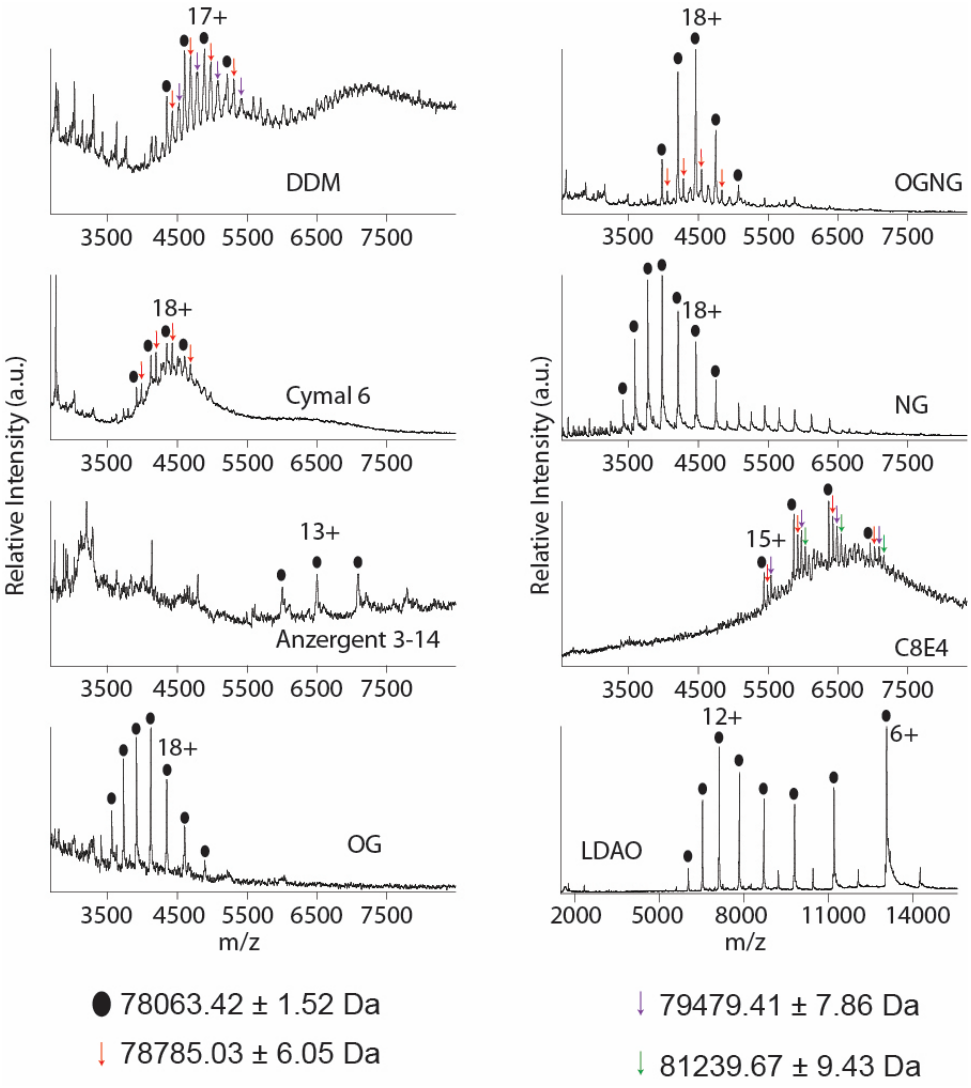


Figure 2.5: Mass spectra of *Pf*MATE-GFP showing the dependence on spectra quality of detergent screening by maintaining the same detergent for both extraction and MS experiments. Measurements in the presence of LDAO resulted in clear and highly resolved spectra. The collision voltage applied was 200V for DDM, Cymal6, Anzergent3-14, OG, OGNG, NG. In the case of C8E4 and LDAO a collision voltage of 120V was applied.

The activation energy needed for detergent removal was variable based on the detergent. For instance, OG, NG, OGNG, DDM and Cymal 6 required high activation energy (≥170 V) for the charge state series to be resolved whereas for LDAO and C8E4 an activation voltage of 100 V was sufficient for detergent removal. This result provides the first hint that the activation energy needed for detergent removal is dependent on detergent chemistry.

The population of adduct bound species is smaller for OGNG relative to DDM, Cymal 6 and C8E4 whereas clear spectra are observed for OG, NG, LDAO. Both OG and NG spectra show the presence of detergent micelle in the low *m/z* region, unlike for LDAO and C8E4. It is evident that LDAO provided the highest resolution spectrum and no lipid adducts or other impurities were observed.

Where the protein adopts only one conformation, the charge residue-field emission model predicts a single Gaussian charge state distribution,[132] however in the case of MS analysis of *Pf*MATE solubilised in LDAO detergent a deviation from this is observed suggesting the possibility that multiple conformations are present: an unfolded one with higher average charge state and a folded one with a lower average charge state. In the high *m/z* region a dimer is observed which can also affect the charge state distribution. As LDAO is considered to be a harsh detergent,[135] potentially a population of the protein unfolds either during extraction, due to a high concentration of LDAO, or during collisional activation. However, as detergent removal in the case of LDAO is performed at significantly lower energy than for other detergents, it can be hypothesised that *Pf*MATE unfolds during extraction from the membrane. To test whether *Pf*MATE unfolds at the extraction step OGNG was used for extraction, then the sample was exchanged to LDAO at a low concentration of 0.094% (w/v) corresponding to 2× CMC, using a SEC column. In this case a Gaussian charge state distribution was observed, confirming the hypothesis that LDAO unfolds the protein only at high concentration during the extraction stage. The solubilised *Pf*MATE in OGNG was also exchanged to C8E4 detergent yielding highly resolved spectra with a normal distribution of charge states (Figure 2.6).

Taken together these results demonstrate the significance of detergent screening during extraction, purification and spraying as a crucial step for MS analysis of membrane protein.

As expected, harsh detergents at least for *Pf*MATE generated highly resolved spectra with less lipids bound than mild ones. However they also potentially lead to protein unfolding during at high concentration during extraction, hence a combination of several detergents is preferred, where harsh detergents are only used in low concentration immediately prior to MS analysis.

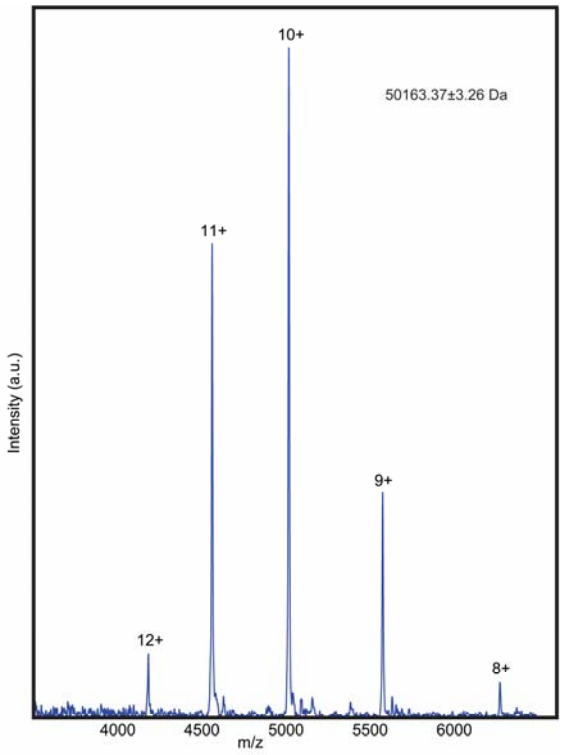


Figure 2.6: A highly resolved mass spectrum of *Pf*MATE solubilised in OGNG and sprayed in C8E4 in 200mM ammonium acetate. The sample was concentrated to 8µM and recoded on a Waters Synapt G1 instrument. The applied cone, capillary and collsion voltage were 160V, 1.8kV and 140V respectively. The sample is fully dilipidated and no adduct impurities are bound.

2.3 The effect of detergent on charge state distributions

In addition to a different degree of delipidation, MS analysis of *Pf*MATE in different detergents provided the first hint that the charge state distribution is detergent dependent. Previous studies performed on soluble proteins also indicated that detergent may have an

effect on the overall charge state,[111] however since this study was done under denatured conditions, directly correlated structural effects on membrane protein-detergent interactions were not possible. Maintaining a folded conformation and being able to manipulate the charge state distribution without compromise in spectra quality is crucial for minimising Coulombic unfolding leading to preservation of native-like conformation in the gas phase. MS analysis of *Pf*MATE fused with GFP sprayed in different detergents showed that OG, NG, OGNG, Cymal 6 and DDM detergents had similar charge state distributions at around 17+. Interestingly, C8E4 and LDAO gave much lower charge state distributions. The same trend was also observed after GFP cleavage. For instance in the case of OG an average charge state of 13+ was recorded whereas in C8E4 the charge state distribution was centred at 9+. A closer investigation on the chemistry allowed each detergent to be classified in two different categories: detergents with a saccharide group, such as OG, NG, DDM, Cymal 6, OTG and DM, and harsh detergents that have instead a polyethylene glycol head group. The effect on the charge state distribution of LDAO are similar to the latter group hence for simplicity LDAO is classified in the same group as polyethylene glycols. To account for protein specificity, another membrane protein AqpZ was selected and MS measurements in the presence of detergents from both classes were collected (Figure 2.7).

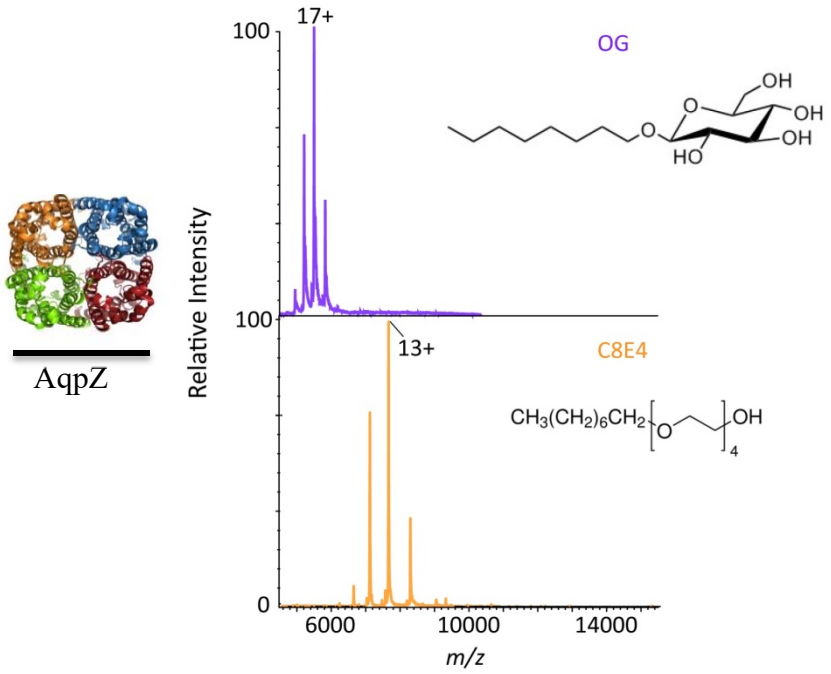


Figure 2.7: Mass spectra of AqpZ showing that th charge state distributions is dependent on detergent class. A higher charge state distribution is observed for OG (purple) detergent as compared to C8E4 (orange). The insets show the chemical structure of both detergents and AqpZ (PDB 2AMB).Measurements were performed in 200mM ammonium acetate and 2xCMC of each detergent with a protein concertation of 13.5µM. The applied collision voltage for OG was 200V whereas for C8E4 was 120V.

As in the case of *Pf*MATE a higher charge state distribution was observed for OG, centred at 17+, compared with C8E4 (13+). To further elucidate the relationship between detergent class and charge state distribution as a function of protein mass a combination of ten membrane proteins spanning a mass range of 50 kDa to 426 kDa (Table 2.1) solubilized in twelve different detergent were analysed (Table 2.2).

Table 2.1: Proteins used in this study. The native oligomeric state and its mass are denoted.

Protein	Mass (kDa)	Oligomeric state
*Pf*Mate	50	1
MscL	85	5
AqpZ	99	4
AmtB	127	3
FocA	160	5
ELIC	184	5
AqpZ-GFP	211	4
AmtB-GFP	214	3
MscL-GFP	226	5
GlpF-GFP	235	4
Foca-GFP	298	5
Lysenin	300	9
MdtF	338	3
AcrB	344	3
MscS-GFP	413	7
MdtF-GFP	419	3
AcrB-GFP	426	3

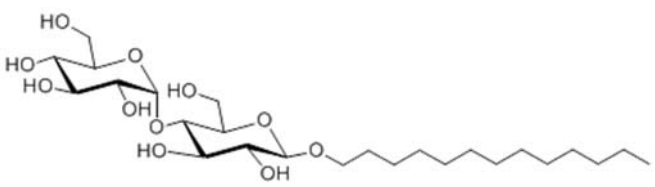
Not all membrane proteins were stable in all detergents; hence charge state distributions could only be calculated for a subset of 49 detergent-protein combinations (Figure 2.8 and Table 2.2). Saccharide detergents were consistently associated with higher charge states, whereas polyethylene glycol detergents resulted in lower average charge state distributions across the mass range. The results were also plotted against the expected maximum charge state derived from the Rayleigh limit of soluble proteins, in ammonium acetate, as function of mass.

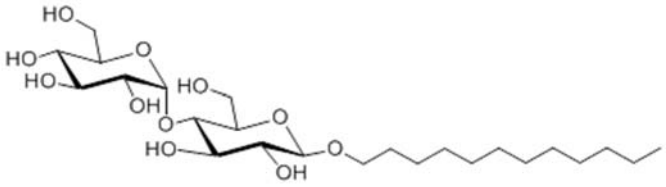
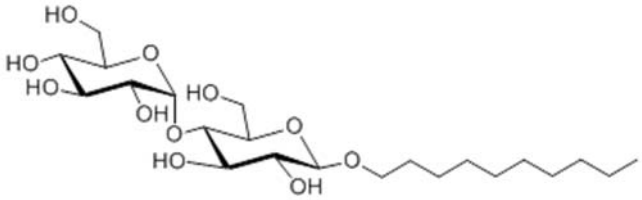
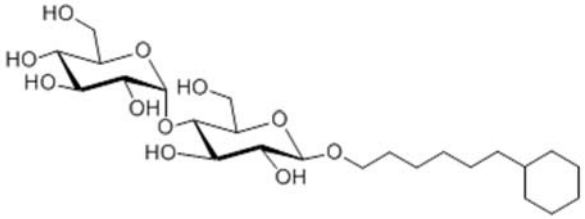
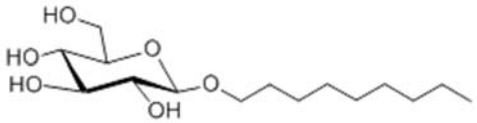
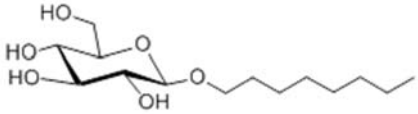
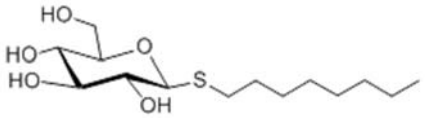
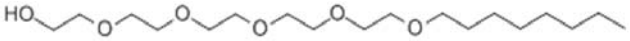
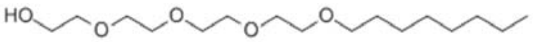
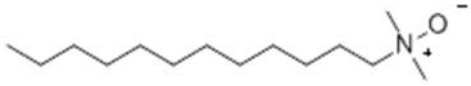
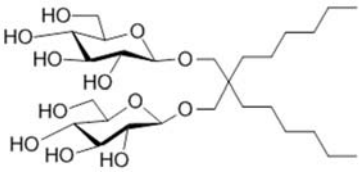
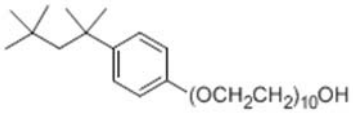
The Rayleigh equality correlates the maximum charge state with the diameter of the droplet, as shown in Equation 2.1, and can be rearranged in a form of a power law ($q_{max} = a\ M^{\ b}$) to show a relationship between the maximum charge state of a macromolecule with its mass, shown in Equation 2.2 where, maximum charge is denoted by q_{max}, mass by M, γ is the surface tension, N_A Avogadro's number and ε_0 electrical permittivity of vacuum:

$$q_{max} = 4\left(\frac{\pi\gamma\varepsilon_0}{\rho e^2 N_A} \times M\right)^{\frac{1}{2}} \tag{2.2}$$

By making an approximation for the macromolecule density as 1 g cm^{-3} and using the measured surface tension of ammonium acetate as 70 m Nm^{-1}, the constants *a* and *b* of the power law are calculated to be 0.078 and 0.5 respectively. The data for each detergent class were also fitted to a power law deriving the constant *a* to be 0.0249 and 0.0036 for saccharides and polyethylene glycol detergents respectively. A derivation from soluble proteins was found for the constant *b* for polyethylene glycols (*b* = 0.71) whereas in saccharide detergents the *b* constant was lower, calculated to be 0.58.

Table 2.2: Properties and structures of detergents used in this study. Molecular weights and critical micelle concertations were taken from Anatrace. For TX-100 an average value of molecular weight is shown.

Detergent	Molecular weight (g/mol)	Critical micelle concentration (mM)	Structure
TDM	524.6	0.033	

DDM	510.6	0.17	
DM	482.6	1.8	
Cymal6	508.5	0.56	
NG	306.4	6.5	
OG	292.4	20	
OTG	308.4	9	
C8E5	350.5	7.1	
C8E4	306.5	8	
LDAO	229.4	2	
OGNG	568.7	1.02	
TX-100	647	0.24	

These calculations support the observation that the presence saccharide detergents gives slightly lower charge states, due to a lower *a* constant, than soluble proteins, whereas in the case of polyethylene glycols a reduction of as much as 20 percent of the charge state distribution relative to the Rayleigh limit is observed. A larger deviation from the expected charge state distribution is measured if the mass of the micelle is added to protein complexes (Figure 2.8 inset). This deviation can be rationalized by considering that the detergent micelle protects the membrane region from charging and that the water exposed surface area plays a more dominant role in charging.[141]

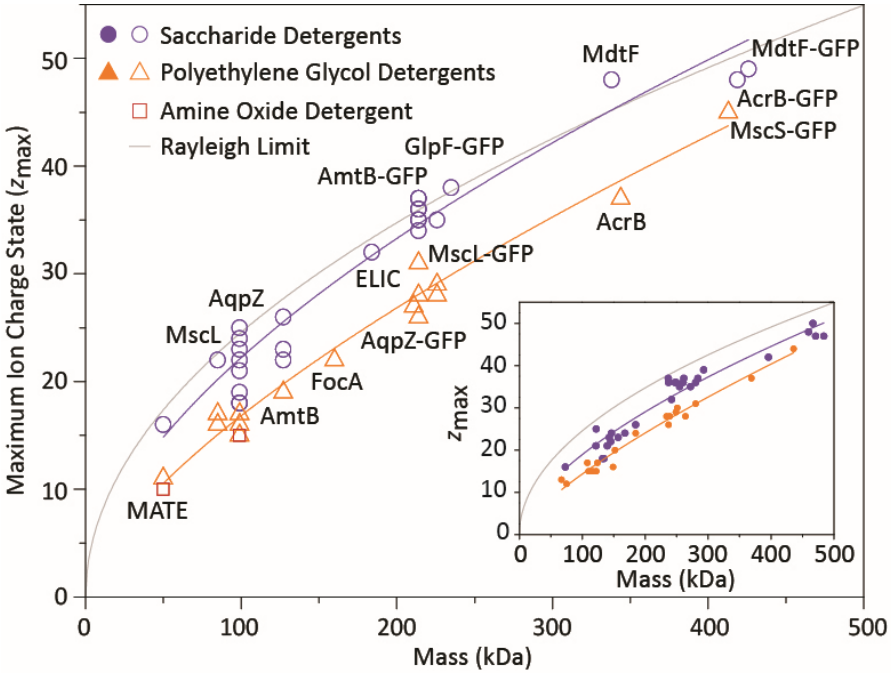


Figure 2.8: Maximum ion charge state (z_{max}) post-detergent release as a function of protein mass (M) and protein-micelle mass (inset). The theoretical Rayleigh limit (grey line) is for 200 mM ammonium acetate droplets. Data are fitted to the power-law relationship $z = a M^{b}$. Proteins were solubilised at 2xCMC of each detergent and sprayed at 10µM. Data were collected in partnership with Eamonn Reading.

Table 2.3: Maximum observed charge and average charge state of selected proteins solubilised in saccharide detergents.

Protein	z_{ave}	Z_{max}
OG		
AmtB-GFP	31.2	37
AqpZ	17.0	21
MATE	13.3	16
NG		
AmtB-GFP	30.4	35
AmtB	21.1	23
AqpZ	18.0	21
AcrB-GFP	43.5	49
MdtF-GFP	41.8	48
OTG		
AmtB-GFP	31.9	37
AqpZ	18.0	24
DM		
AmtB-GFP	31.4	36
AqpZ	16.0	20
DDM		
AmtB-GFP	30.4	35
AmtB	24.0	26
AqpZ	20.2	23
MscL	19.4	22
MscL-GFP	31.7	35
AcrB-GFP	42.7	47
MscS-GFP	43.1	47
ELIC	28.8	32
MdtF	39.5	44
GlpF-GFP	34.3	38
UDM		
AmtB-GFP	30.3	36
AqpZ	16.2	21
Cymal-5		

Protein	z_{ave}	Z_{max}
AmtB-GFP	31.3	36
AqpZ	19.2	25
Cymal-6		
AmtB-GFP	30.5	36
AqpZ	20.0	22
TDM		
AmtB-GFP	30.0	36
OGNG		
AmtB-GFP	27.9	34
AmtB	19.9	22
AqpZ	16.2	19

Table 2.4: Average charge and maximum charge states observed for the selected proteins solubilised in polyethylene glycol detergents, TX-100 and LDAO detergents

Protein	z_{avg}	Z_{max}
TX-100		
AmtB-GFP	21.6	28
AqpZ	13.1	16
C8E4		
AmtB-GFP	21.3	26
AmtB	16.3	19
AqpZ	13.6	17
MscL	11.8	16
MscL-GFP	24.8	29
FocA	20.3	22
AcrB	32.6	37
MATE	9.3	11
C8E5		
AmtB-GFP	22.8	28
AqpZ	11.9	15
AqpZ-GFP	22.8	27
MscL	12.5	17

Protein	z_{avg}	Z_{max}
MscL-GFP	23.6	28
MscS-GFP	36.9	45
C12E8		
AmtB-GFP	25.8	31
C16E20		
AmtB-GFP	25.8	31
LDAO		
AqpZ	12.7	15
MATE	7.5	10

To better understand why polyethylene glycol detergents give spectra with lower charge state distributions two possible scenarios where charge reduction can happened were considered: whether charge stripping occurs during detergent activation or where ion formation with less overall charge is the dominant mechanism. First the hypothesis that detergent leaves as charged species during activation was considered. This hypothesis fits with the rationale that detergents simply solubilise membrane proteins, hence the protein detergent complex behaves as a soluble protein. The charge state of the detergent protein complex will be equal to the sum of charges carried by the protein and the micelle charge. Since the micelle is removed inside the mass spectrometer the measured charge does not reflect the charge during ionisation, but rather only part of this charge as detergent leaves carrying charge. To test this hypothesis mass spectra of the membrane protein *Pf*MATE, solubilised in C8E4 and DDM, were recorded. Instrumental conditions were optimised to allow for direct visualisation of bound detergent molecules by operating at low activation energies. Detergent removal was monitored as a function of increasing collision energy and the average charge state was recorded for each level of energy (Figure 2.9). As the collision energy increases sequential detergent removal is observed. However, the charge state

distribution remains the same, demonstrating that detergent leaves as neutral species and as a result does not account for charge reduction in the case of polyethylene glycol detergents.

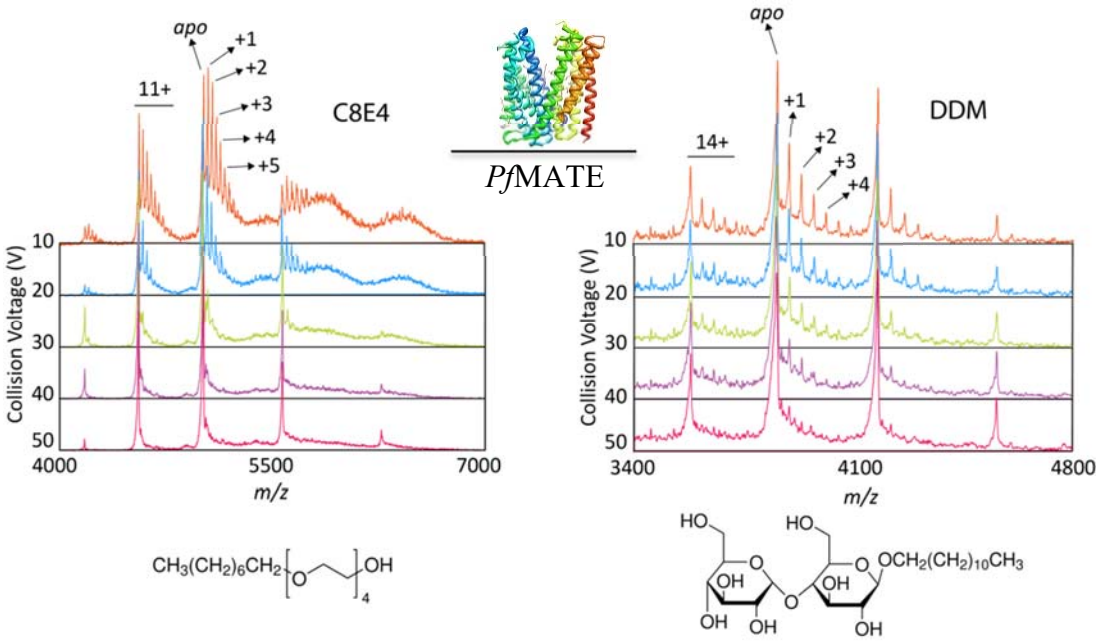


Figure 2.9: Detergent molecules are lost as neutrals upon collisional activation, with the detergent chemistry influencing both the ionization and release of membrane proteins in the gas phase. Mass spectra of *Pf*MATE, at a final concertation of 12µM in 200mM ammonium acetate and solubilised in 2xCMC of DDM (right), C8E4 (left), show detergent removal upon collisional activation. Inset is the crystal structure of *P. furiosus* MATE (PDB: 3VVN). Arrows show the number of bound detergents at low activation energy.

The second hypothesis was then considered, where the presence of detergent during ion formation is responsible for lowering the total charge in the droplet. Charge reduction can result from either a change in surface tension or in the light of the charge residue-field emission mechanism, detergent serving as a charge carrier, leaving the final protein ion with less charge. The fact that surface tension can alter the charge state distribution is stated in the Rayleigh limit equation (Equations 2.1 and 2.2), hence the presence of detergent may alter the droplet surface tension and thus shift the charge state distribution. Using drop shape analysis, the surface tension of each detergent was measured and plotted against the average

charge state or maximum charge state for AmtB-GFP and AqpZ (Figure 2.10). In both cases no correlation between surface tension and charge state distribution was observed. Therefore surface tension variations from the presence of different detergents does not account for the observed charge state differences based on detergent class.

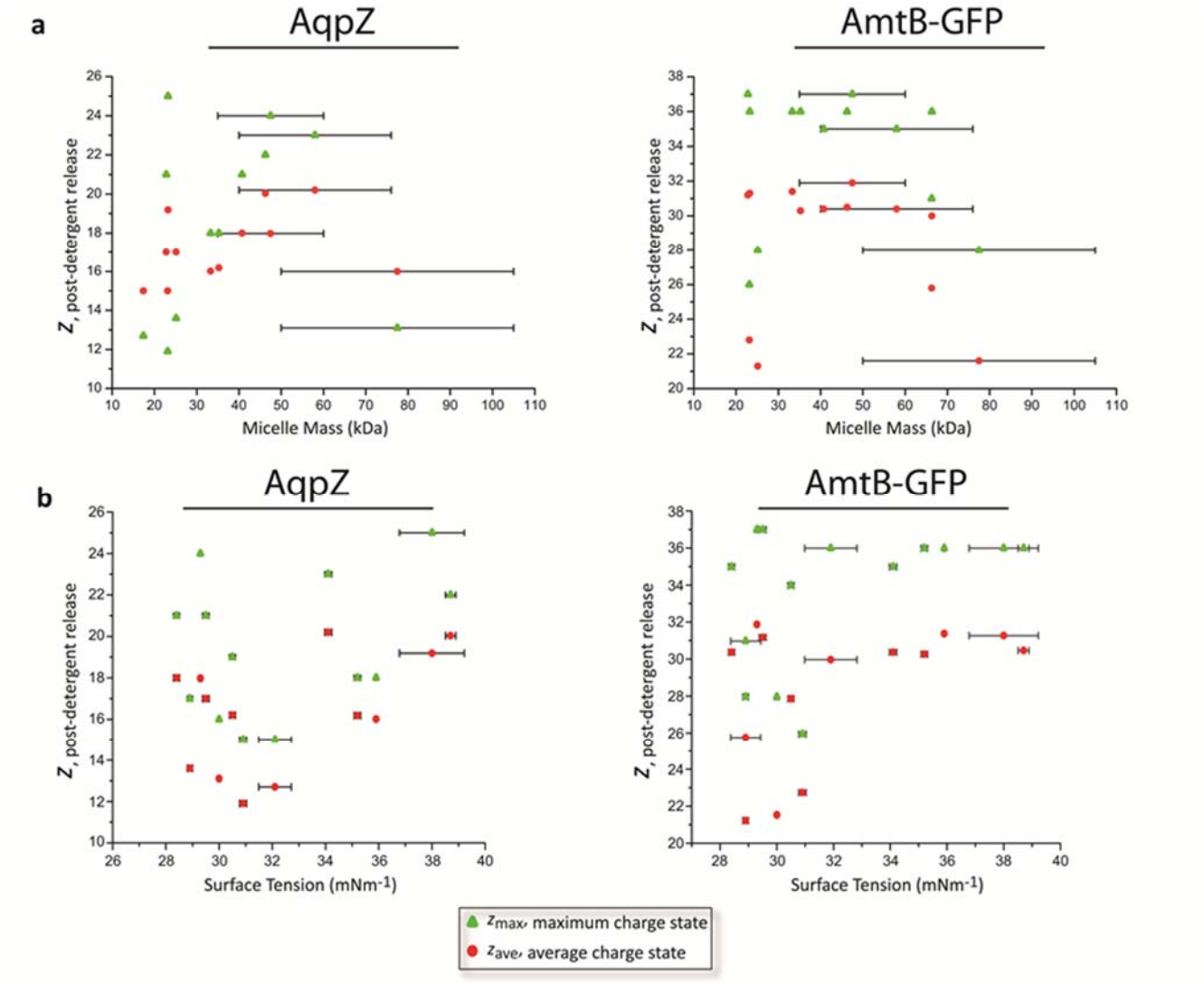


Figure 2.10: Plots of micelle mass (a) and surface tension (b) of detergent against z_{ave} (red circles) and z_{max} (green triangles) for both AqpZ(left) and AmtB (right) membrane proteins. The micelle mass was calculated from reported aggregation number taken from Affymetrix. The surface measurements were performed at 2xCMC of each detergent in 200mM ammonium acetate. The error bars for micelle mass reflect the uncertainty in published micelle sizes. The error bars for surface tension reflect the standard deviation of repeated independent measurements (n = 5). Micelle mass calculations were provided by Eamonn Reading.

As surface tension is not responsible for the lower charge states in the presence of polyethylene glycol detergent, it can be hypothesised that this detergent may serve as a charge carrier, unlike saccharide detergents which emit as neutral species. This hypothesis

was tested by considering a soluble protein, transthyretin (TTR), to decouple the effect of detergent bound to the protein and free detergent in the droplet (Figure 2.11). As TTR is a soluble protein, the charge state distribution measured in the presence of detergent can be compared to measurements in ammonium acetate only without detergent present.

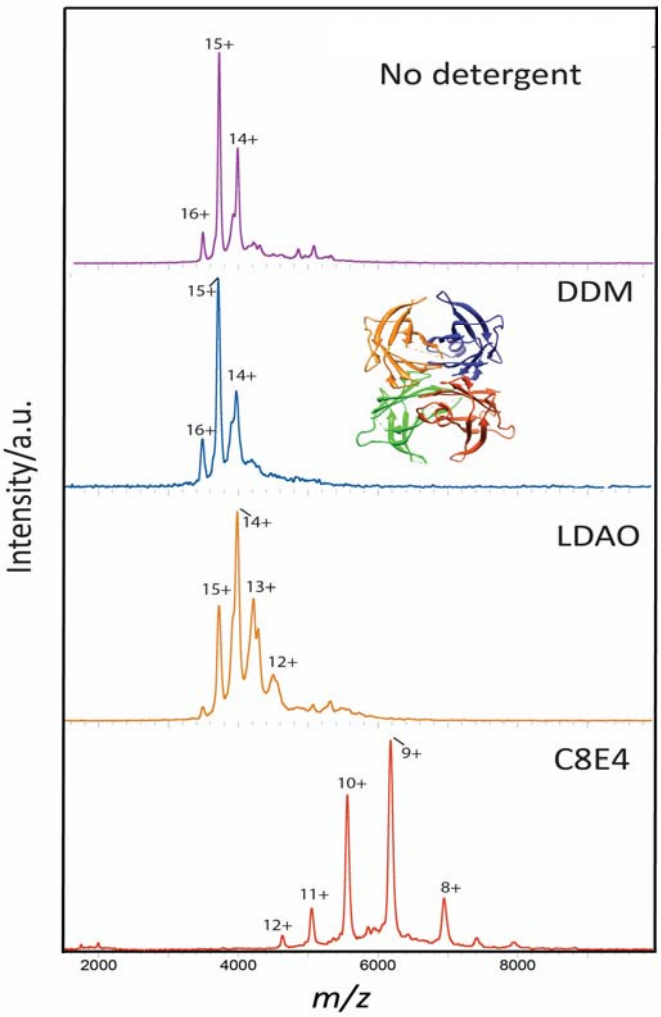


Figure 2.11: Mass spectra of TTR showing that charge state dependence on the presence of charge carrier species. Mass spectra of showed the same charge state distribution in the absence of detergent (200mM ammonium acetate) and DDM purple and blue spectra respectively, whereas a slight charge decrease is observed for LDAO (orange) and a significant reduction in the case of C8E4 (red). TTR was maintained at 8.6μM and detergent concentrations were at 2xCMC. A low collision voltage of 30V was applied for DDM, LDAO and no detergent whereas in C8E4 the applied collision voltage was 60V.

The average charge state in ammonium acetate was found to be 14.7+. There was no significant difference in average charge state between the measurement in ammonium acetate and in the presence of DDM (z_{ave} = 14.5+). However a large reduction in charge was found in the case of C8E4 (z_{ave} = 9.4+) and a noticeable, but smaller reduction in the case of LDAO (z_{ave} = 13.8+). These results provide evidence in favour of the hypothesis that charge reduction is a direct consequence of detergent removal as a charge carrier.

To provide insight into if the micelle as an entity provides a role in charge reduction or only the chemical characteristics of the detergent head group are responsible for the observed variation of the charge state distribution the head group of C8E4, C4, was added to the TTR sample. A charge reduction to only an average of 13.2+ was observed demonstrating that synergetic effects of the detergent clusters are important for the observed charge states in the presence of detergents.

An interesting effect was observed for membrane proteins sprayed in the presence of LDAO. Relative to saccharide detergents, the presence of LDAO in the sample yielded charge reduced states similar to C8E4. However, collisional activation of the protein leads to a further reduction of the charge state, shown in Figure 2.12 for the case of *Pf*MATE. For instance at a collision voltage of 100 V the charge state distribution centres at 9+, increasing the energy to 240 V the charge state distribution shifts to 6+. This observation can be explained due to proton transfer from the protein to detergent molecules at high collisional activation.

Molecular dynamics (MD) simulations have shown that as the protein detergent complex is transferred into the gas phase, the absence of water may allow some of the detergent molecules to reverse in orientation, with the polar head group in close proximity to the

protein.[142] High levels of activation may promote this rearrangement, and as a result the polar head group could pick up a charge from the protein leaving the detergent-stripped protein with less charge. Although this cannot be fully explained, it is evident that charge formation in the case of LDAO follows a mechanism other than exclusively charge carrier emission.

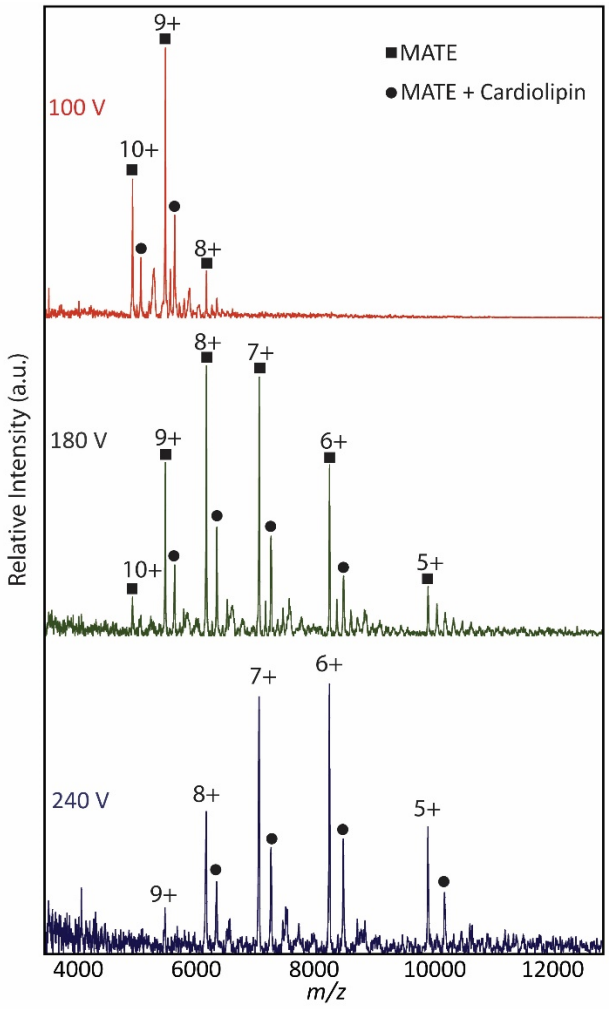


Figure 2.12: The dependence of charge state distribution of membrane protein *Pf*MATE, solubilised in LDAO, on collisional voltage. Increase in the collision voltage from 100 V (red) to 240 V (blue) shows a reduction in the charge state distribution centre from 9+ to 6+. The population of lipid bound is unaffected. The measurements were performed at a protein concertation of 6.2µM.

2.4 The effect of polarity on charge state distribution

Previous studies have shown that in negative polarity, soluble proteins are associated with a lower charge state distribution.[133] By employing nESI to transfer proteins from solution to the gas phase while preserving their non-covalent interactions and performing IM in both positive and negative polarity, it was found that the measured collision cross sections were in parity between polarities. As the surface area of the proteins remained the same regardless of polarity it was suggested that the CRM cannot explain the observed charge state differences. Furthermore, by taking into consideration that the proteins were sprayed in identical conditions in both polarities, the observed differences in the ion charge were attributed to charge carrier effects in accordance with the charge residue-field emission model. It was then hypothesised that in negative polarity, anions had lower charge carrier energy than the corresponding cation in positive polarity.

In the light of these measurements performed in negative polarity it is interesting to investigate if the same phenomena are manifest in the presence of detergents. This has two direct implications. First, if charge reduction in negative polarity is independent of detergent chemistry than at least in negative polarity polyethylene glycol detergents do not serve as charge carriers, hinting that a mechanism other than the charge residue-field emission model applies. However, if a charge reduction in negative polarity is observed only for saccharide detergents, then these experiments provide more evidence that the charge residue-field emission model applies in the presence of detergents and polyethylene glycol detergents are ejected from the droplet as charge carriers. Second, combining both charge reducing effects arising from polyethylene glycol detergents and negative polarity may lead to accessing even lower charge states, which may better preserve native oligomeric states. However, if these two effects cannot be combined, accessing lower charge states for saccharide detergents in

negative polarity can be an alternative way to preserve the oligomeric states of proteins that are not stable in polyethylene glycol detergents.

To investigate the dependence of charge state distribution on polarity in the presence of detergents, a set of membrane proteins that were expressed in-house and solubilised in both detergent classes was used. Noticeably for saccharide surfactants a charge reduction in negative polarity relative to measurements performed in positive polarity was observed. For polyethylene glycol detergents there was minimal charge state difference between both polarities (Figure 2.13). For instance, the measurements performed for AmtB solubilised in C8E4 detergent show no significant difference in charge state distribution between both polarities; the centre of the charge state distributions were 18.9+ and 19.6- for positive and negative polarity respectively. The same measurements performed in OG detergent shows a reduction in the average charge state from 22.1+ to 18.2- from positive to negative polarity. In a similar way the ligand-gated ion channel (ELIC) a pentameric membrane protein was subject to MS measurement in DDM detergent, a saccharide type detergent, in both positive and negative polarity (Figure 2.14). In positive polarity the pentamer charge state distribution is centred at 30+. Maintaining the same sample conditions but performing the measurement in negative polarity the charge state distribution shifted to 23-. Importantly, this reduction in the charge state stabilises the native pentameric oligomeric state (*vida infra*).

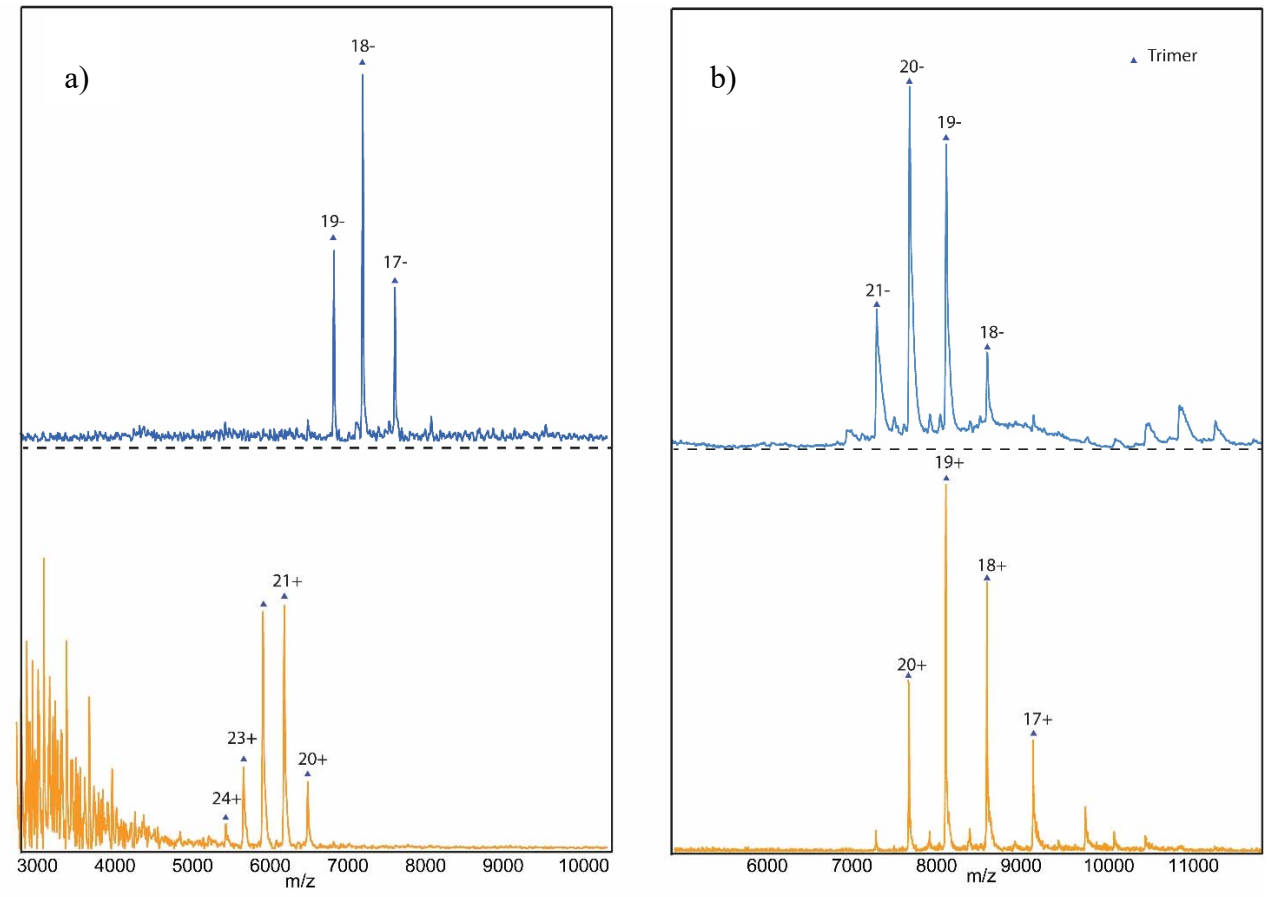


Figure 2.13: The effect of polarity in charge state distribution is dependent in detergent class. (a) Mass spectra of *Ec*AmtB solubilised in OG at 2xCMC showed a charge reduction from positive to negative polarity (orange and blue spectra respectively). (b) Mass spectra of *Mt*AmtB solubilised in C8E4 shows no charge dependence on polarity. In both cases the protein concertation was 12µM and the applied collision energy was 180V for OG and 90V for C8E4. The cone and capillary were 160V and 1.8kV in positive mode and 100V and 1.1kV in negative mode.

To examine the dependence of polarity on average charge state relative to solvent accessible surface area, nine membrane proteins in a mass range from 50 kDa to 300 kDa were considered and measurements were performed in both polyethylene glycol and saccharide surfactants. An increase in mass results in greater charge state difference between positive and negative polarity in the presence of saccharides (Figure 2.14). However, performing the same measurements in polyethylene glycol detergents resulted in no difference in charge state distributions between both polarities. In the light of the combined

charged-residue ion-evaporation model this difference in average charge distribution that depends on both polarity and surfactant class can be attributed to charge carrier effect, achieved by buffer particles and polyethylene glycol surfactant in negative and positive polarities respectively.

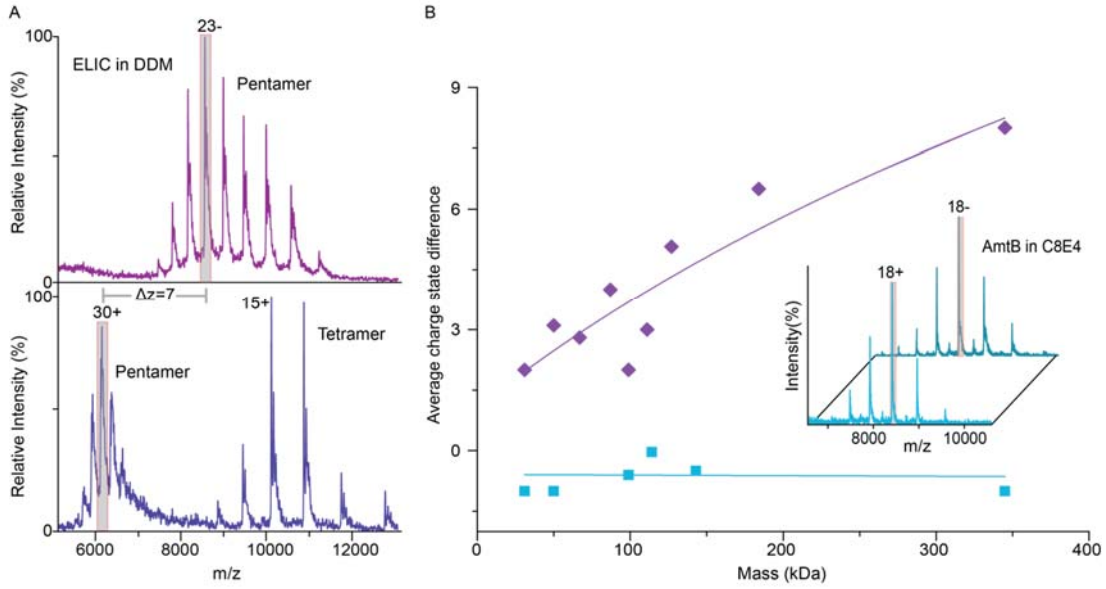


Figure 2.14: The effect of negative polarity on charge state in the presence of detergents. Mass spectra of ELIC pentameric species showed a charge reduction, in the presence of DDM, from positive (blue) to negative polarity (purple). The native oligomeric state is better preserved in negative polarity, whereas in positive polarity a large amount of dissociation of the pentamer to tetramer is observed. The protein contraction was 16µM and the applied voltage was 180V in both polarities. (b) Average charge state difference between two polarities solubilised in saccharides detergents (purple) and C8E4 (blue) of 9 membrane proteins (MATE, McsL, AqpZ, AmtB, *Mt*AmtB, OmpF, FocA, ELIC, AcrB). Proteins were solubilised at 2xCMC each detergents and sprayed at 10µM. The inset show the mass spectra of AmtB in C8E4 in both polarities, no change is charge state was observed.

2.5 Preserving the native oligomeric state by reducing Coulombic repulsion

The detergent is removed inside the mass spectrometer at elevated activation energies via a CID process during which ions are accelerated in an applied potential. As a result ions with high charge will experience higher activation leading to better desolvation and removal of non-specific adducts. In addition, high charge will impose a greater extent of Coulombic repulsion on the protein, which combined with high applied energy, as in the case of membrane proteins, results in disruption of non-covalent interactions and loss of native oligomeric state. On the other hand, having a lower charge state distribution allows for lower activation energy to be applied to the membrane protein ions, hence better preserving their native oligomeric state. As a result detergents with charge reducing properties can be exploited to better preserve the native oligomeric state. Figure 2.15 shows mass spectra of AmtB in DDM and C8E4. In the case of DDM the native trimer oligomer has a higher charge state distribution and the energy needed to remove the detergent is also sufficient to strip a high charge monomer from the trimeric complex leaving a low charge state dimer due to asymmetric charge dissociation during CID. Hence, the total intensity of the trimer is diminished. In the presence of C8E4, lower charge states are accessible and the trimer state is the only present species after detergent removal.

Many proteins are not stable in the presence of harsh charge reducing detergents. In order to reduce Coulombic repulsion in these cases, as common for soluble proteins, charge reduction agents are used. For instance by using 10mM of imidazole the average charge state of AmtB in OG detergent is reduced from 21+ to 18+. The same extent of charge reduction is also observed in the presence of acetonitrile vapours in the ionisation chamber. Interestingly these charge reducing effects can be combined, providing insights that charge reduction

mechanism of imidazole and acetonitrile vapours are a result of complementary mechanisms. In the case where both charge reducing agents are used the average charge state distribution is shifted from 21+ to 16+ and only the native trimer oligomeric species is present in the spectra.

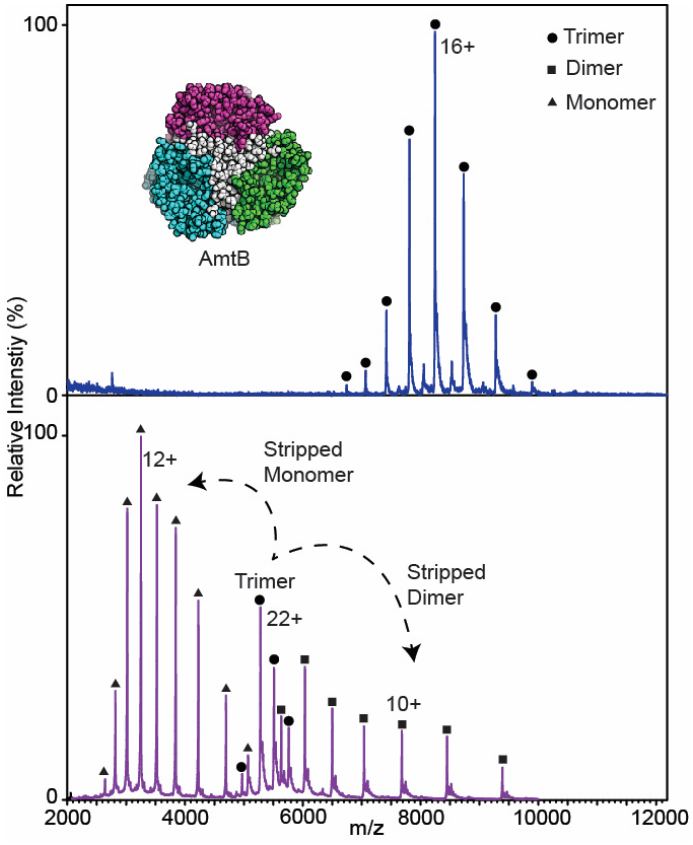


Figure 2.15: Polyethylene glycol detergents stabilise the native oligomeric state. Mass spectra of AmtB in C8E4 (blue) show only the trimeric state whereas the energy needed to liberate the complex from DDM detergent (purple) is sufficiently high to fragment the complex resulting in monomers and dimers. The protein was maintain at 9μM in both cases and the applied voltage was 200V for DDM and 140V for C8E4.

In the case of relatively large proteins (≥80 kDa) where only saccharide detergents can preserve the native stoichiometry and addition of charge reducing agents leads to protein unfolding or a compromise in spectra quality, negative polarity can be exploited to achieve lower charge state distributions. Unlike for soluble proteins, where no significant advantage

of negative polarity in preserving the native-like structures was found,[133] in the case of membrane proteins, where high activation energies are applied for detergent removal, accessing a lower charge state distribution is highly desirable to maintain the native oligomeric state. Figure 2.14 and 2.16 shows the mass spectra of two pentameric membrane proteins, ELIC in the presence OG detergent and the mechanosensitive channel for large conductance (MscL), solubilised in NG detergent, respectively. Due to the high activation required to eject NG and OG detergents from MscL and ELIC respectively, the native oligomeric state cannot be unambiguously assigned, as both proteins dissociate via CID into tetramer and monomer. However, using the same sample conditions in negative polarity, the native oligomeric state at lower charge states was preserved without generating CID products. For instance, in case of ELIC a charge shift of seven, by simply changing polarity, was sufficient to completely eliminate CID fragmentation and only the native oligomeric state was observed. Therefore negative polarity can be used to preserve the native oligomeric states of membrane proteins that are solubilised in saccharide detergents.

The case of MscL illustrates another advantage of using negative polarity in non-denaturing conditions for mass measurement of the oligomeric state. In this case the energy needed to remove the NG micelle is substantially higher than the maximum energy tolerable to preserve the non-covalent interactions. Consequently due to incomplete micelle removal the pentameric oligomeric state cannot be resolved. The spectra show the presence of CID products; the tetramer is well-resolved whereas the monomer intensity cannot be easily visualised due to the presence of the detergent micelle in the low *m/z* region. Switching the polarity to negative resulted in preservation of the pentameric species with the charge state distribution centred at 14-.

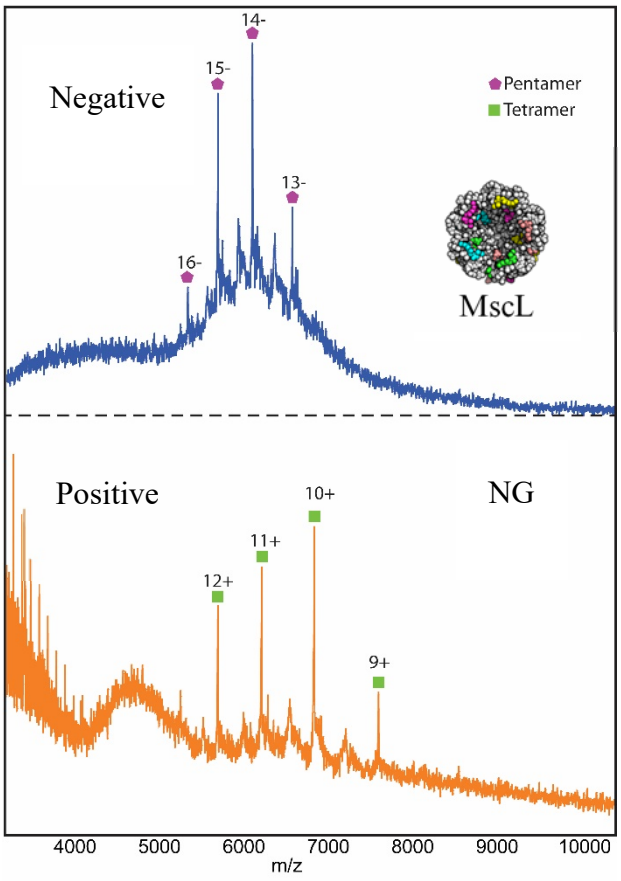


Figure 2.16: Charge reduction in negative polarity stabilises the native oligomeric state. MscL solubilised in NG detergent in positive polarity (orange) shows only the dissociated tetramer, whereas the native oligomeric state pentamer cannot be resolved. The same sample in negative polarity (blue) resulted in pentameric species only. The applied voltage were 170V and 180V for negative and positive polarity respectively.

When Coulombic repulsion plays a crucial role in destabilising the non-covalent interactions, as is typical for large multi-subunit protein complexes, charge reduction in negative polarity can be combined with addition of charge stripping to access even lower charges. For example lysenin forms a large pore with a mass of 301 kDa and the average charge state distribution in positive mode with the protein solubilised in DDM is 37.2+. As in the case of AmtB, ELIC and MscL the protein dissociates by expelling an unfolded monomer with high charge state to lower the total Coulombic repulsion. Performing these measurements in negative polarity, as expected, the average charge state distribution drops to

28.8-, approximately eight charges lower. Adding imidazole means in negative polarity an additional reduction by six charges on average is achieved, bringing the average charge state to 23.4-. Interestingly, imidazole addition resulted in charge reduction in both positive and negative polarity. It has previously been hypothesised that charge altering species will have the opposite effect in different polarities,[143] however as shown in Figure 2.17, this claim is not supported by the data present here. Hence, the charging mechanism in the presence of charge reducing agents remains elusive.

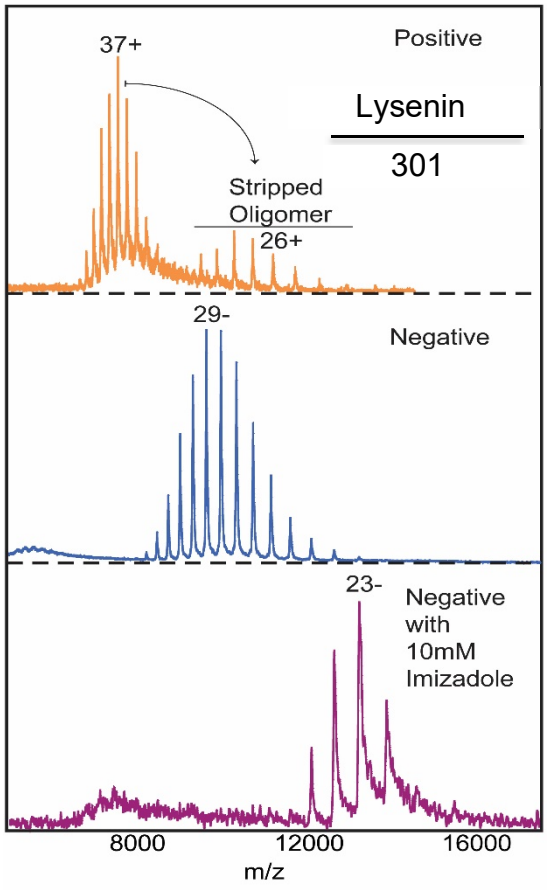


Figure 2.17: The charge reduction effect of negative polarity and charge reduction agents can be combined to access lower charge states. (Top) Mass spectrum of the lysenin pore in DDM. (Middle) Negative polarity (blue) of the same sample shows a reduction in charge state. (Bottom) Addition of 5 mM imidazole further reduces the charge state (purple).The protein was maintained at 16µM and buffer exchange to 200mM ammonium acetate and 2xCMC DDM prior to measurements.

These results demonstrate that charge reduction by minimising Coulombic repulsion may lead to preservation of the native oligomeric state. Polyethylene glycol detergents or negative polarity of saccharide detergents are two useful methods for accessing lower charge states.

2.6 Probing the activation energy required to release membrane proteins from detergent micelles

The activation needed to eject the membrane protein ELIC from the OG detergent micelle in negative polarity was higher than the supplied voltage on positive polarity, 160 V and 145 V, respectively, yet more dissociation was observed in positive polarity due to higher charge. The total energy experienced by each ion, commonly expressed as the laboratory frame energy (E_{LAB}) can be calculated by multiplying the charge state with activation energy, as shown in Equation 2.4, where z is the ion charge and V is the voltage applied in the collision cell:

$$E_{LAB} = z \times V \tag{2.4}$$

In the case of ELIC in positive mode the 30+ charge state experiences an E_{LAB} of 4.35 keV, in contrast to only 3.68 keV in the case of negative polarity for the 23- charge state. This difference in the total energy of the ions was crucial for preservation of the native oligomeric state.

The activation energy required to remove detergent is one of the components of E_{LAB}, hence, to better characterize the effect of detergent on preservation of native folded structures the energy required to remove all detergent was measured. Here, as in the case of charge state distribution calculations, mass spectra of AqpZ and *Pf*MATE were collected in the presence of seven different detergents from both saccharide and polyethylene glycol detergent classes

and the detergent removal was reported as a function of collision energy. As energy increases fewer detergent clusters are present bound to the proteins. Since *Pf*MATE is a small membrane protein, in both detergent classes observation of the number of detergent bound is possible, however for larger membrane proteins it is common to only observe an unresolved 'hump' at low collision energies in the mass spectrum due to a large number of polydisperse detergent clusters bound (Figure 2.18). As the energy increases the hump shifts to lower *m/z* values since detergents leaves as neutral and the total mass of the detergent protein complex decreases. Eventually charge states corresponding to the expected *m/z* values of the membrane protein appear in the spectra. Further increase in the collision energy may result in removal of remaining detergent molecules but also in subunit dissociation due to loss of non-covalent interactions.

Relatively well-resolved charge states are observed for C8E4 and to a lesser extent for C8E5 at activation energy of 60 V, however, the same activation voltage in the case of OG resulted in an unresolved hump in the mass spectrum. When the activation voltage supplied is the same between OG and C8E4, and since OG spectra are associated with higher charge states the E_{LAB} is higher for OG, however fewer detergents are removed. These findings provide the first hint that saccharide detergents require a greater total energy to be removed from the protein-detergent complex.

To better characterise this effect a large set of saccharide detergents including OG, DM, UDM and OTG was considered and the energy required to remove their micelles from membrane proteins was compared with the energy required for removal of C8E5, C8E5 and LDAO (Figure 2.19). In the case of AqpZ the activation energy was increased until full dissociation of the native tetrameric state was observed whereas for *Pf*MATE the entire possible range of collision voltage values was used up to the maximum of 240 V.

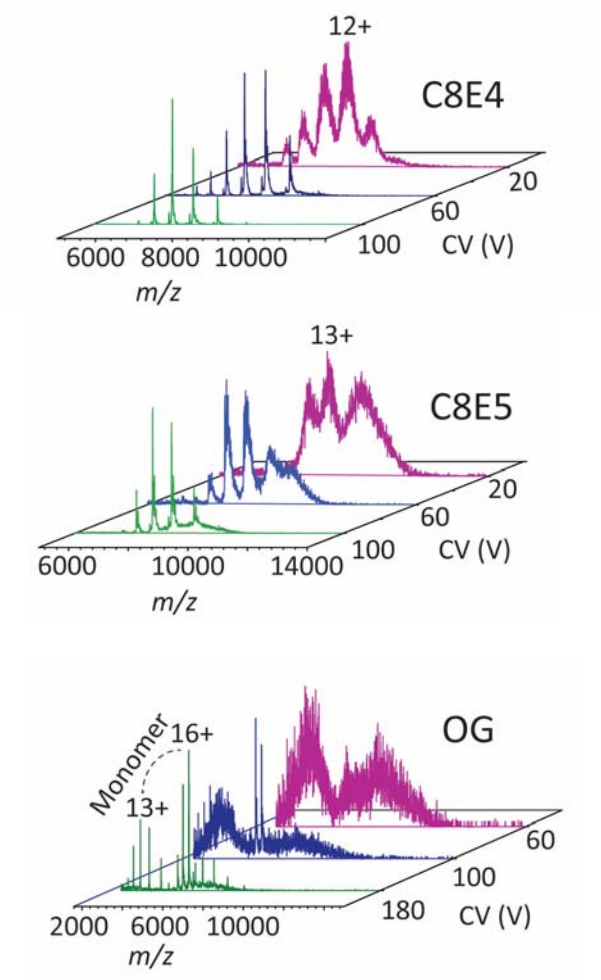


Figure 2.18: The effect of activation energy for detergent removal. Mass spectra of AqpZ acquired at low (purple), medium (blue) and high (green) collisional voltage in three different detergents, C8E4 (top) C8E5 (middle), and OG (bottom). The spectra shows that less energy is needed to remove C8E4 and C8E5 in comparison to OG.

To monitor the energy required for detergent release E_{LAB} was plotted against resolving power and average number of detergent bound. To calculate the E_{LAB} the average charge state distribution was used and the resolution calculated from the full-width at half-maximum. The number of detergent bound was either calculated directly where spectra were sufficient to

resolve each detergent bound or by deconvolution of the added mass and dividing it by the mass of detergent.

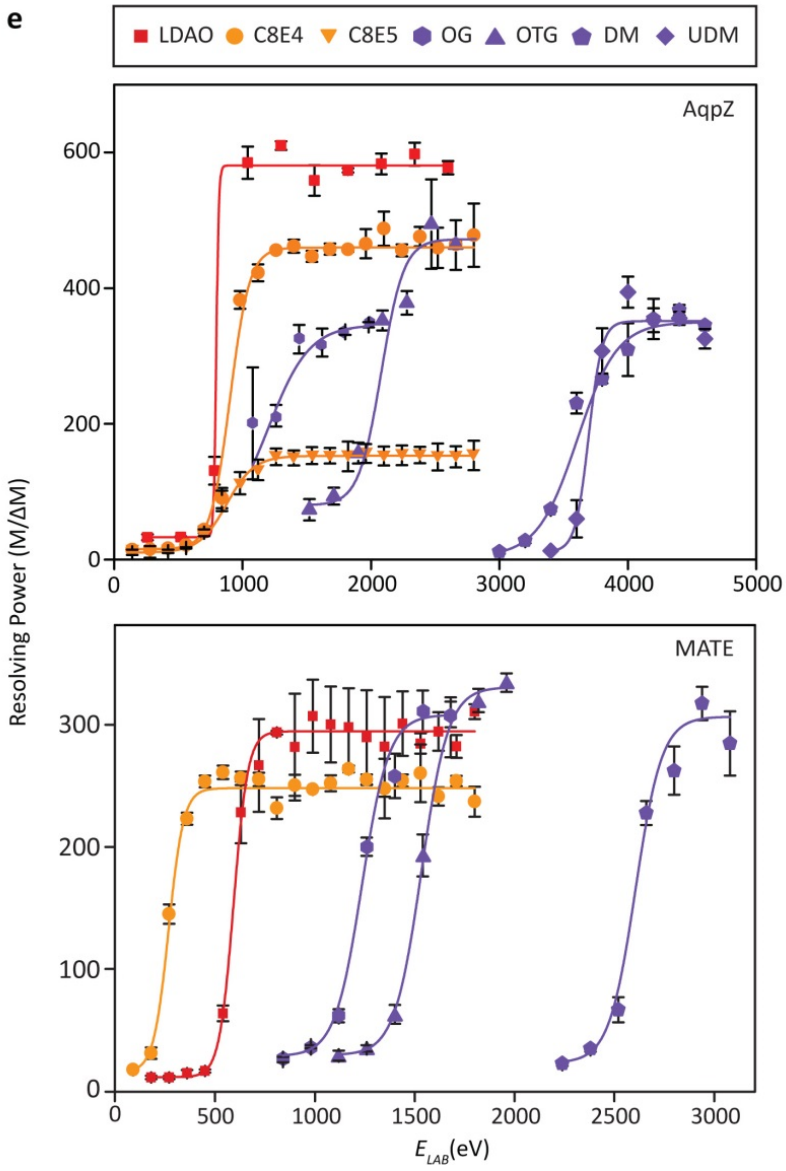


Figure 2.19: Detergent class dictates the release energy. The *m/z* resolving power at E_{LAB} for AqpZ (top) and *Pf*MATE (bottom) is shown. The number of bound detergent molecules is proportional to the *m/z* peak width. Detergents are coloured by class (saccharide, purple; PEG, orange; LDAO, red). Data are fitted with a sigmoidal distribution, and error bars represent the standard deviation from three replicates.

The rationale to account for resolving power as a degree of detergent release will only hold if detergent binding is nonspecific[144] and similar to salt adducts where the energy needed for their removal is linearly correlated with the amount of salt molecules being dissociated. As shown in Figure 2.20 there is a linear relationship between peak width and added mass for both AqpZ and *Pf*MATE. The same relationship was also observed in the case of the average number of detergent bound and resolution (Figure 2.21). By combining these two linear relationships the heterogeneity of detergent bound was calculated as peak width in mass divided by the mass of detergent, and was linearly correlated with average number of detergent bound providing evidence that collisional activation of protein-detergent complexes is consistent with Poissonian statistics of adduct removal. Hence, detergent binding in the gas phase is non-specific, behaving similarly to salts for soluble proteins.

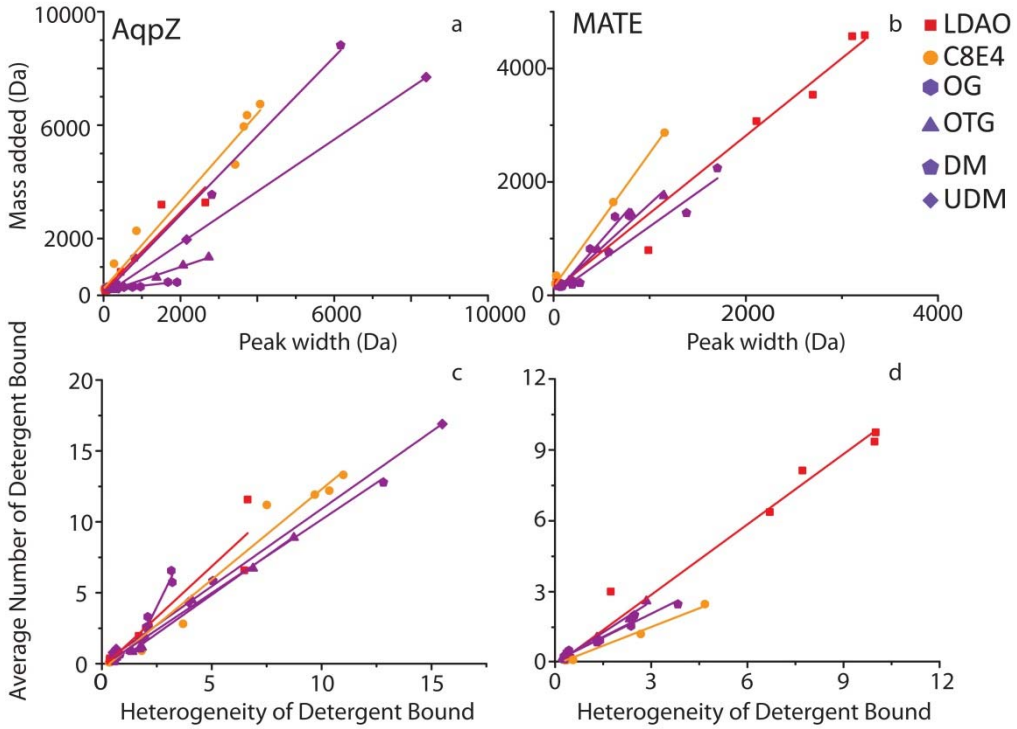


Figure 2.20: Detergent is removed from membrane proteins in a concerted manner upon collisional activation. The average number of detergents bound to AqpZ (a) and *Pf*MATE (b) ions, upper panels. The number of bound detergent molecules is proportional to the inverse of *m/z* resolving power (ΔM/M), lower panels. Detergents are coloured by class: saccharide purple, polyethylene glycol orange, and amine oxide red. Data were collected up to 200 V accelerating potential voltage due to instrumental limitations. Data are fitted with a sigmoidal distribution. Error bars represent one standard deviation from repeated measurements (n = 3).

Noticeably, better resolved peaks are achieved with increase in energy due to detergent removal. By comparing different detergent classes it is evident that C8E4 and LDAO yielded highly resolved spectra at relatively low activation energy as compared to saccharide detergents. In the case of AqpZ, for both detergents total activation energy, an E_{LAB} of 800 eV, was sufficient to completely remove the detergent micelle from the protein complex. Higher E_{LAB}, up to 2,500 eV, were needed to remove the detergent micelle for OG and OTG. Although the charge state distributions for OG and OTG are similar to DM and UDM; highly resolved spectra for DM and UDM were only possible at the maximum collisional voltage

provided by the instrument (Figure 2.21). Even with high activation detergent molecules were still bound as energy was dissipated via subunit dissociation prior to detergent release. For instance, at E_{LAB} of 4,800 eV one detergent was found to be bound compared to complete detergent removal at 1,200 eV for C8E4 and LDAO.

Interestingly, the number of bound detergent, even at low energies does not correspond with the aggregation number of the micelle found in solution. It was estimated that the ratio of detergent micelle to membrane protein complex is 1:1, hence 78 OG detergent molecules are expected to be bound in solution.[145] The measurements presented here show that at mild activation of 10 V an average of seven OG detergent molecules can be estimated to be bound in the unresolved peak. For C8E4 and LDAO form an aggregation number in solution of 82 and 76 respectively, only 13 detergent molecules were found to be present in the spectra at an E_{LAB} of 100 eV. This behaviour can in explained by considering that prior to activation in the collision cell the complex is also activated in the source, hence the majority of the micelle is removed during the initial desolvation stage. In addition, detergent micelles in the gas phase are found to be highly polydisperse.[146] Possibly, protein-detergent complexes with low aggregation number can be better focused and filtered by the front optics, as a result they are detected with a higher signal intensity.

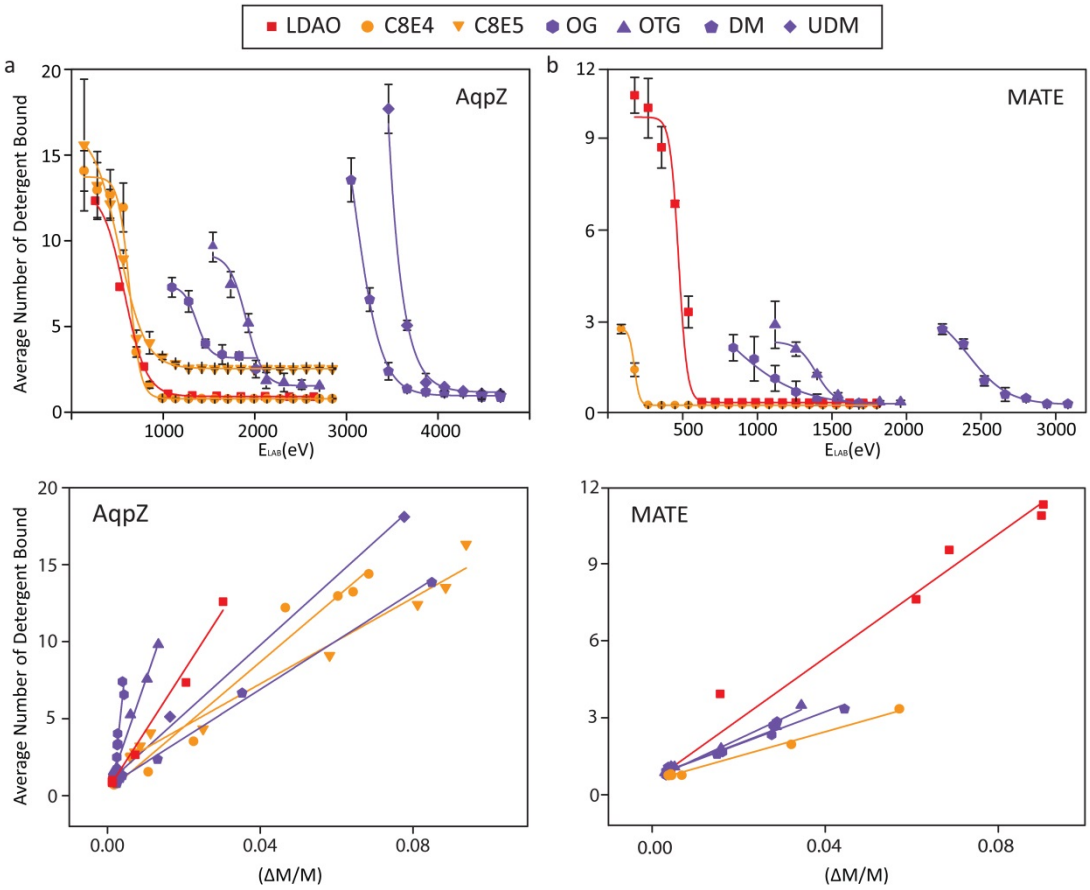


Figure 2.21: Detergent is removed from membrane proteins in a concerted manner upon collisional activation. The average number of detergents bound to AqpZ (a) and *Pf*MATE (b) ions, upper panels. The number of bound detergent molecules is proportional to the inverse of *m/z* resolving power (ΔM/M), lower panels. Detergents are coloured by class (saccharide purple, polyethylene glycol orange, and amine oxide red). Data were only collected up to 200 V accelerating potential voltage due to instrumental limitations. Data are fitted with a sigmoidal distribution. Error bars represent one standard deviation from repeated measurements (n = 3).

As expected, the spectra recorded in the presence of C8E5 shared many of the characteristics of C8E4 spectra, such as detergent removal at low activation energies and yield highly resolved spectra at low E_{LAB}, however bound detergents were found to be more resistant than C8E4, providing evidence that increasing polar content in the detergent head group may increase the affinity for detergents to membrane proteins. This would be in agreement with the binding of ligands to soluble proteins where increased electrostatic

interactions enhance binding.[147,148] Furthermore, it can be speculated that the different energy regimes needed to remove OG relative to UDM could be attributed to differences in the number of hydrogen bonds formed between these detergents and the membrane proteins. Relatively insignificant differences were found in the release energy between UDM and DM, hence increase in chain length does not significantly stabilise binding.

By comparing both detergent classes, it is clear that mild detergents require higher activation energy to be removed. In the light of the fact that lipids remain bound even at the elevated energies needed to remove detergent it can be derived that detergent that mimic lipids more closely are associated with higher activation energy for removal. These findings can be used to explain why mild detergents perform better in solubilizing folded membrane proteins in solution, hinting that maximizing the interactions between membrane proteins and detergent leads to better preservation of the native-like folded structure. Nevertheless, in the gas phase, the high activation energies needed for detergent removal make the use of saccharide detergents undesirable to preserve the native-like folded states. As a consequence polyethylene glycol detergents are preferred to interrogate the native-like folded structure and monitor the change in gas-phase structures by IM.

2.7 Monitoring the stability of membrane proteins by collision-induced unfolding

Collisional activation can induce conformational changes by partial unfolding with retention of the oligomeric state.[149,150] In a Synapt instrument, ions are activated in the trap region, prior to IM measurement; hence the effect this activation has on the size of ions and whether they are folded or collision-induced unfolded can be measured. IM can be used to monitor the effect of charge and activation energy on each gas-phase structure by reporting

the arrival time distribution for each species. Folded species will interact less with gas particles hence arrive faster than unfolding or unfolded species which will have a larger CCS to interact with gas particles. For proteins that have cavities in the native structure, in the absence of water and at low activation energies, the collapsed state may also be observed by IM as their arrival time distributions are lower than expected. In addition, when using a RF confining drift cell the acceleration experienced by each ion is charge dependent. As a result, ions with high charge spend less time in the drift cell, hence for ions with the same shape the arrival time distribution linearly correlates with *m/z* values at different charge states.

The effect of activation energy on protein conformation is shown in Figure 2.22. Mass spectra of *Pf*MATE at 20 V and 200 V in the trap, show virtually the same charge state distributions, whereas IM provides evidence of the presence of different gas-phase conformations, obviously deriving the gas-phase structure of a protein by simple investigation of the charge state distribution is potentially misleading. A narrow charge state distribution may suggest the folded conformation is preserved[151] in solution but this information cannot be directly correlated to the behaviour of ions in the gas phase.

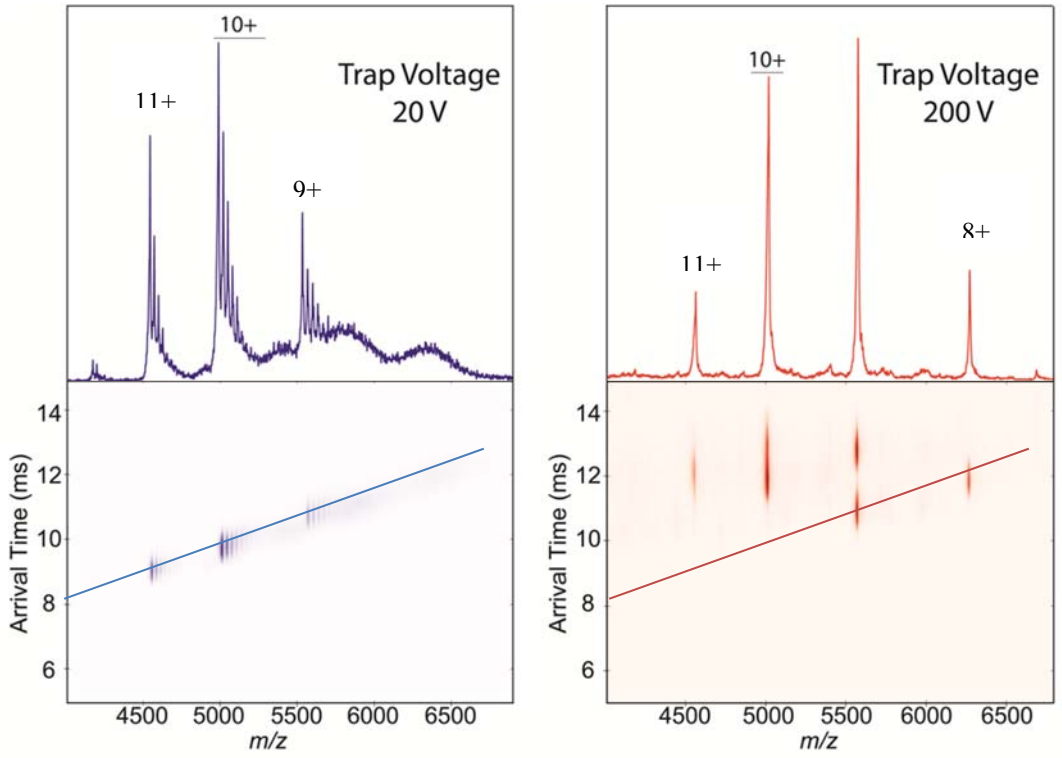


Figure 2.22: The effect of activation voltage on protein conformation. Mass spectra and ion mobility recording top and bottom respectively, for *Pf*MATE at 20 V (blue) and 200 V (red). The data suggest that at low activation energy all charge states are folded and native-like in structure. Increase in collision energy leads to protein unfolding. The protein was solubilised in 2xCMC C8E4 in 20mM ammonium acetate and sprayed at a concertation of 8.5µM. In both cases the cone of capillary vulgates were 160V and 1.8kV respectively. Measurements were done on a RF-confined linear drift tube instrument modified for large mass measurements.

The arrival time distribution of *Pf*MATE at 20 V shows the characteristics of a folded conformation (Figure 2.22). Since in RF-confined drift tubes the charge effect in arrival time distributing is linear, the charge state distribution of ions with the same shape in mobility space should be linear. Furthermore at this low activation the measured arrival time distributions are in parity with the expected arrival time distributions back-calculated from the computed CCS of the atomic coordinates. Therefore at 20 V, all ions, regardless of their charge state, are in the native-like folded state. At 200 V a deviation from the folded pattern in IM space is observed. It is obvious that higher charge states, for example 11+ and 10+,

have greater arrival times compared to the same charge states at 20 V but this is also the case for lower charge states in the same spectrum. Since these ions have higher than expected arrival time their solution shape is not preserved. Interestingly, in the case of charge state 9+, the arrival time distribution is composed of two well-separated peaks, providing evidence that both folded and unfolded conformations are present. In other words, the total activation energy (E_{LAB}) is not high enough to fully unfold the entire ion population hence ions are observed in two states.

To monitor the effect of activation energy on protein gas-phase conformation the CCS of for each charge state was plotted as function of collision voltage from 10 V to 240 V (Figure 2.23).

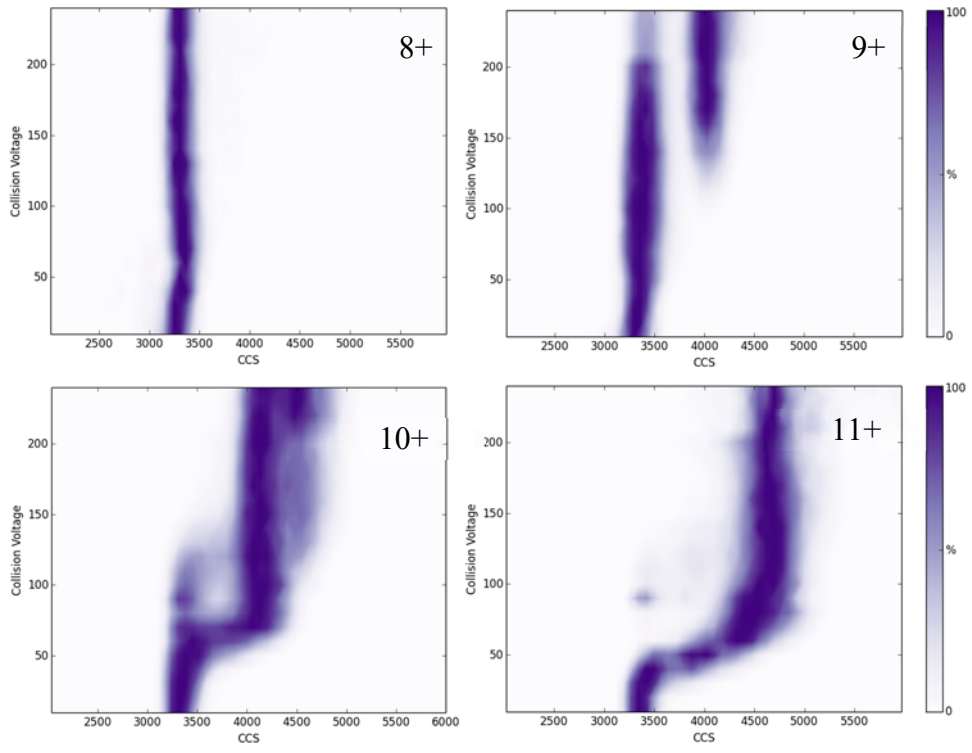


Figure 2.23: Monitoring the effect of collusion induced dissociation on proteins gas-phase shape. Unfolding plots of *Pf*MATE as a function of collision voltage for each charge state are shown. The arrival time distribution is converted to CCS to monitor the unfolding from the native conformation to an unfolded state. The protein was solublised in 2xCMC C8E4 in 200mM ammonium acetate.

For the lower charge state, 8+, the native-like conformation of *Pf*MATE was preserved at all accessible levels of collisional activation. Having an additional charge makes the complex unstable at collisional voltages ≥160 V and the presence of unfolded species with increased CCS is observed. Further increase in charge to 10+ and 11+ lowers the required energy, from 80 V and 50 V respectively, that the native-like conformation is observed to unfold. C8E4 detergent was crucial to access the lower the charge states, otherwise a higher charge state, as in the case of saccharide detergents, at the energy needed for detergent removal, would lead to protein unfolding.

The effect of charge and activation energy should be considered simultaneously to identify conditions where the native-like conformation is preserved.[105] Figure 2.24 shows a plot of charge, activation energy and CCS for *Pf*MATE. At low charge states and low energies the native-like folded conformation is maintained. Increasing the charge state or the activation energy leads to protein unfolding. Hence to preserve the native-like structure detergents that can be removed at low activation energies and which possess charge reducing properties are highly desirable. Both of these properties are provided by C8E4 and LDAO making these detergents highly compatible for preservation of native-like structure.

In the case of *Pf*MATE charge state 8+ remains folded even at maximum energy provided by the instrument indicating that accessing low charge states leads to preservation of native-like conformation. This is highly protein specific and access of stable charge states where a native-like conformation can be maintained at all voltages is typically not attainable. For instance, the outer membrane protein F (OmpF) is composed of three β-barrel channels all of which have a cavity in the middle. Performing measurements in C8E4 allows access to low charge states and detergent can be removed at low activation energies, however the native conformation is not maintained. Rather the structure collapses by 18 percent as indicated by

the experimental CCS compared to theoretical CCS (Figure 2.25). When the energy in the trap is increased to 140 V the unfolded intermediates, which have CCS values closer to the native-like conformation are observed. Notably, these species do not resemble the same conformation as OmpF in solution but rather represent an unfolded species relative to the collapsed state. Since collision-induced unfolding usually is initiated by unfolding of one subunit,[105] in case of OmpF the species with a coincidence CCS to the native conformation may have two subunits collapsed and an unfolded one. The collapse of proteins with cavities at low collision voltage has also been reported for soluble proteins.[89]

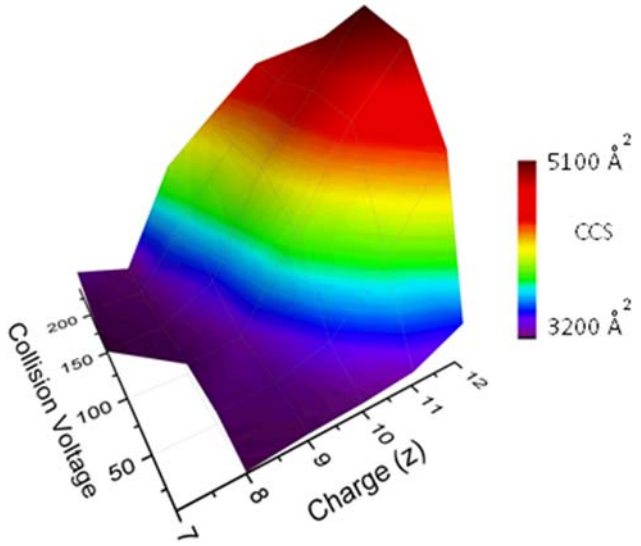


Figure 2.24: The effect of activation voltage and charge states on protein structural state. The data show that at low activation energy all charge states are native-like and folded. Similar to low charge states, 8+ or 7+ in this case. Increase in collision energy leads to protein unfolding and charge state induces protein unfolding, monitored here by recording the CCS for *Pf*MATE solubilised in C8E4.

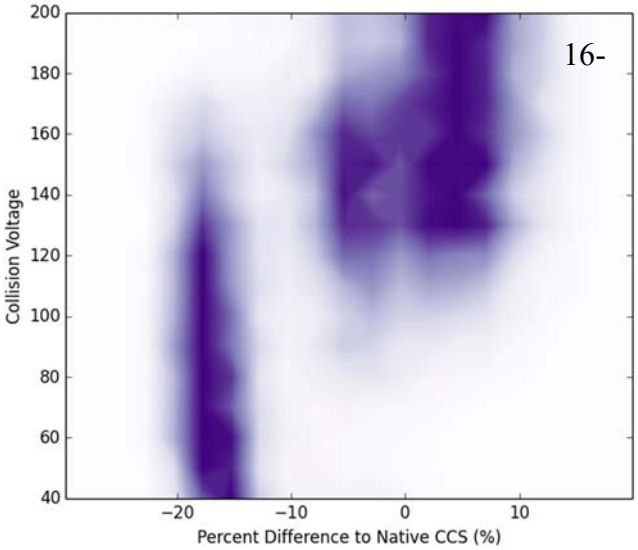


Figure 2.25 The collision cross section (CCS) of OmpF in OG, charge state 16-, as a function of activation voltage. CCS is shown as percent change from the native CCS. At low activation energy a collapse state is observed further increasing the energy lead to two unfolding states.

Having identified that polyethylene glycol detergent are highly desirable for preservation of the native conformation of membrane proteins in the gas phase, this detergent class was then selected to investigate the effect of activation energy and charge state on membrane proteins with variable mass and area of transmembrane and soluble domains. Four membrane proteins, AmtB, MscL, FocA and AcrB were selected due to their surface variability in membrane and soluble domains. Gas-phase unfolding trajectories as a function of activation energy were assembled for each charge state. The minimum E_{LAB} needed to maintain 50 percent of the native-like population was calculated and plotted as a function of charge. $E_{LAB,50\%}$ can be used as an indicator to report on protein stability upon charging and activation.[152]

For each membrane protein, a linear relationship between $E_{LAB,50\%}$ and charge was observed (Figure 2.26). An increase in total charge lowers the minimum energy needed to unfold the protein, hence the slope of this relationship, the 'Coulombic unfolding potential',

is negative for each membrane protein. Since during ESI the amount of charge is proportional to surface available for charging,[153] the solvent accessible surface area was calculated from previously reported atomic coordinates. Interestingly, the Coulombic unfolding potential was found to correlate with SASA (R^2 = 0.85). This empirical correlation can be explained by shielding of the membrane domain by detergent molecules from charge carriers present in the droplet, hence making the membrane protein less accessible for ionisation. The larger the lipophilic area, as a percentage of the total surface of the protein, the lower the Coulombic unfolding potential. Membrane proteins with large soluble domains appear to be more prone to unfolding in the gas phase whereas proteins with a large percentage embedded in the membrane are more resistant in response to addition of charge.

Knowing the charge and solvent accessible surface area for each protein is sufficient to calculate the surface charge density (σ) as given by Equation 2.5:

$$\sigma = z \times SASA^{-1} \tag{2.5}$$

For proteins with similar charge density, as in the case of AmtB and FocA, different numbers of unfolding intermediates were observed. Consequently, gas-phase unfolding trajectories do not only depend on charge and activation but the unfolding states are a characteristic of protein's inherent structural features. Taken together these results demonstrate that in the gas phase membrane proteins are inherently more stable than soluble proteins. In addition surface charge density can be correlated to protein stability upon gas-phase collision unfolding whereas the number of intermediates resembles the structural characteristic of the proteins. Similar measurements for soluble proteins have found that the number of unfolding intermediates, at least for monomeric soluble proteins, can be correlated to the number of domains.[105] In the case of membrane proteins studied here a direct

correlation between unfolding intermediates and number of domains was not observed, but is equally difficult to un-tease due to the frequently oligomeric nature of the proteins.

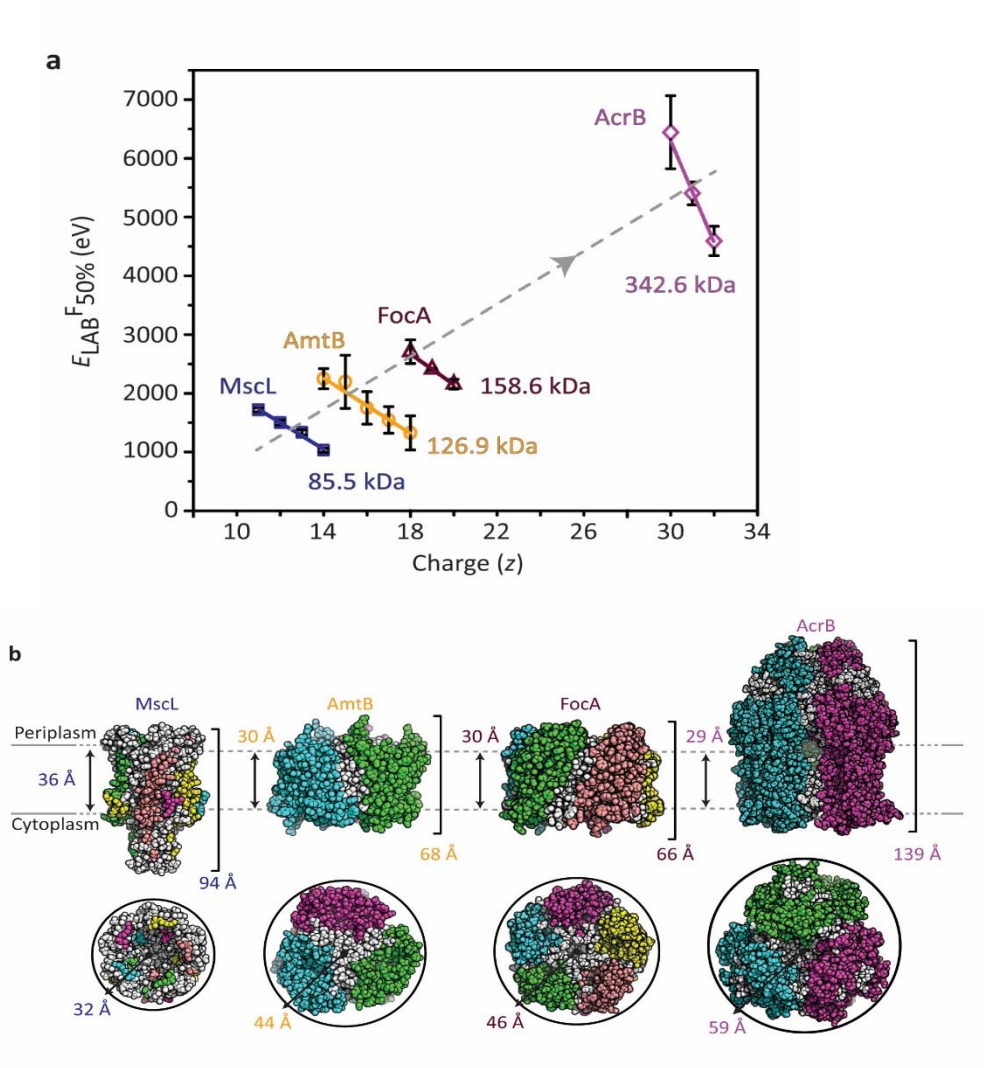


Figure 2.26: Coulombic repulsion affects the stability of membrane proteins in the gas phase. (a) Linear fits of $E_{LAB,50\%}$ versus charge state (z). The error bars represent the standard deviation from triplicate measurements. Data analysis was performed by Eamonn Reading. As the size of the membrane protein increases the energy required to unfold the structure increases due to the ergodic collisional activation process (grey dashed line). The extent of unfolding as a function of charge (m, gradient from linear fits) is m = −229 ± 17, −236 ± 16, −259 ± 15, and −865 ± 78 for MscL, AmtB, FocA, and AcrB, respectively. The errors are the standard error of the linear fits. (b) Crystal structures of *M. tuberculosis* MscL (PDB 2OAR), *E. coli* AmtB (PDB 1U7G), *E. coli* FocA (PDB 3 KCU), and *E. coli* AcrB (PDB 2GIF). The oligomer subunits are uniquely coloured, with the residues at the subunit interfaces coloured light grey. Transmembrane regions were calculated by Matteo Degiacomi.

2.8 Probing the stabilising effect of lipids on membrane proteins

In solution several biophysical techniques have been developed where a particular ligand is allowed to bind with a protein and the change in spectroscopic properties or the change in enthalpy of the system are measured. The two most common techniques that can characterize the effect of ligand binding in solution are isothermal titration calorimetry (ITC) and surface plasmon resonance (SPR). ITC simultaneously measures the binding affinity (K_d), stoichiometry, enthalpy and entropy changes of ligand binding to a protein. Although all thermodynamic properties of binding are determined in a single experiment, without fluorescence labelling or immobilisation, ITC requires relatively large sample amounts. SPR on the other hand uses minute amounts of sample however the protein or ligand is immobilized on a surface.[154] An advantage of SPR is that the same sample can be used for screening several ligands. However both ITC and SPR measure average binding events where individual contributions from free and bound states are not decoupled.

Recently by combining IM and MS, methods that can quantify the effect of ligand binding have been developed. Hyung and co-workers have established a method where the resistance to unfolding of different binding states was tracked as a function of collisional activation.[155] Other methods have also used collision-induced unfolding to report on the effect of ligand binding by monitoring the mid-point between smallest and largest species on the spectrum or quantifying different patterns in the unfolding trajectory.[156,157]

In the method applied here the unfolding pathways of the unbound form and each bound state are followed simultaneously. To quantify the ligand stabilisation effect the unfolding pathway should be similar between the unbound and bound states, which should have the same number of intermediates. Fortuitously, this has always been the case for lipid binding to

membrane proteins.[152] In addition the protein should initially be in its native conformation, this is a demanding but crucial criteria, to fully capture the stabilisation effect from native conformation to unfolded state. Hence, optimisation of mass spectra is an important prerequisite. A balance between the activation energy applied for adduct removal and minimising what is applied to preserve native-like conformations is necessary. In addition, detergents not only compromise the spectra quality, but when the difference in the ligand mass relative to mass of detergent is small (≤50 Da) it is difficult to mass separate ligand bound and detergent bound species. Furthermore, detergent removal is an energetically expensive process. To maintain the native-like conformation the total ion activation should be maintained at a minimum across all instrumental regions. For instance, low activation in the source by applying a high backing pressure and minimum voltage on both cone and capillary are necessary. The transfer region of the IM cell is also maintained at low voltages typically 50 V to ensure that no further structural changes are imposed after drift time measurements.

Unfolding plots are then generated by varying the activation energy in the collision cell. Recording the arrival time distribution of each bound species at several levels of activation and converting it to CCS requires manipulation of a large data set. A novel software was developed by Allison and co-workers[152] that not only can rigorously analyse a large volume of IM-MS data in a short amount of time but also fits the collision-induced unfolded data to a model that accurately defines the change in energy needed to unfold ligand-bound states relative the unbound.

A description of data analysis by this software is given in Figure 2.27. Initially the MS conditions necessary to achieve a native-like conformation is empirically optimised, then a set of unfolding data are generated by increasing the voltage in the collision cell while recording the arrival time of each ion. For each ion the arrival time distribution is extracted

and stacked to create a gas-phase unfolding plot. Folded, intermediate and unfolded states are defined and a model is fitted at each state to effectively derive $E_{LAB,50\%}$ for both bound and unbound states. An increase in the $E_{LAB,50\%}$ of the ligand bound states relative to the unbound is attributed to stabilisation effects provided by the binding event. The model uses a multiple state system where the folded, partially unfolded intermediates and the most unfolded form are assumed to be in equilibrium at $E_{LAB,50\%}$ as described by the equilibrium constant K_{eq} for each transition. The stability of each species, folded (F), intermediate (I) and unfolded (U) can be derived by the change in free energy by correlating the equilibrium constant for species i and j ($K_{eq}^{i\text{-}j}$), to stability for each state ($G_{i\text{-}j}$) where T stands for temperature in Kelvin and R is the ideal gas constant, as given by Equation 2.6:

$$\Delta G_{F-I} = -RT\ lnK_{eq}{}^{F-I}$$

$$\Delta G_{I-U} = -RT\ lnK_{eq}{}^{I-U} \tag{2.6}$$

The change in free energy is assumed to linearly correlate with the collisional voltage (V) and the fraction of each species can be quantified by introducing proportionality coefficient (m), as shown in Equation 2.7, where ΔG_0 stands for free energy at zero collision voltage:

$$\Delta G = \Delta G_o - m(V) \tag{2.7}$$

By obtaining the proportionality coefficient and free energy at zero accelerating voltage the $E_{LAB,50\%}$ can be then derived by Equation 2.8:

$$E_{LAB,50\%} = \Delta G_0 \times m^{-1} \tag{2.8}$$

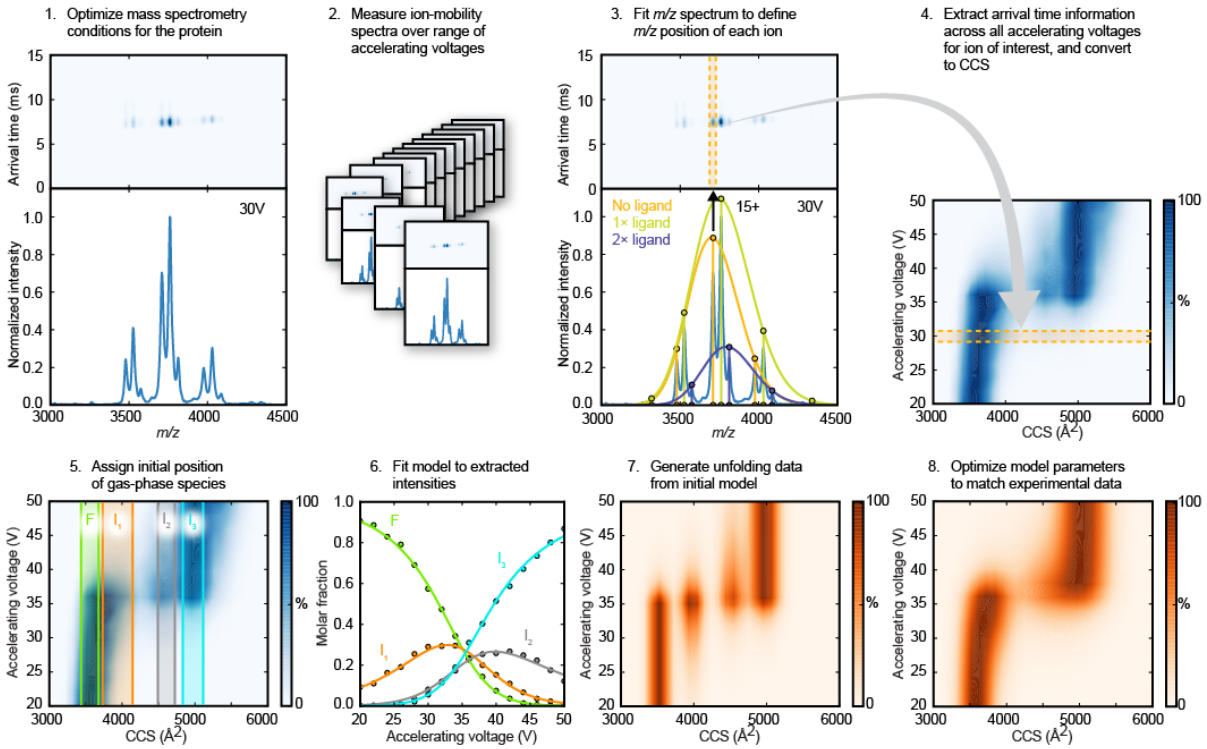


Figure 2.27: A schematic of the workflow performed by PULSAR software. Mass spectra are optimised for separation in both ion mobility and mass dimensions. Subsequently, measurements at increasing activation energy are recorded. The data are mass assigned and the position of each peak in the arrival time distribution is correlated to the *m/z* spectrum. The arrival time distribution is converted to CCS and the data are stacked to form an unfolding plot. Each unfolding state is defined, including the native conformation and the data are fitted to a model. The minimum energy needed to unfold half of the population of each species is reported. Figure taken from Allison *et al.*[152]

Having a method to measure the effect of ligand binding in the gas-phase the attention was then placed on the stabilisation effect of lipids upon binding to membrane proteins. Two membrane proteins, MATE composed of 12 transmembrane α-helices, and OmpF composed of three subunits each composed by 16-strand β-barrel, both solubilised in C8E4 were used. The MATE construct used here was from *P. furiosus* an archaea that has a different lipid composition in the membrane than the host expression cell, *E. coli*. Hence, *Pf*MATE was initially fully stripped from the native *E. coli.* lipids and synthetic lipids were added. Whereas in the case of OmpF endogenous lipid bound were preserved. Since OmpF is an outer

membrane protein the bound lipids can potentially be rough lipopolysaccharide (LPS), as suggested by mass measurement. Due to the presence of large cavities in each β-barrel, OmpF collapses even under charge reduced conditions at low activation energy, nevertheless the stability of lipid binding can still be measured since in both unbound and lipid bound form the protein starts from the same collapsed state, albeit this measurement has low biological relevance. A stabilisation of the protein complex upon LPS binding was observed (Figure 2.28). This increase in stabilisation upon lipid binding may be due to the lipid head group that stabilises the charge interactions whereas the lipid chain, composed of six individual carbon chains, shields the transmembrane region and protects from unfolding.

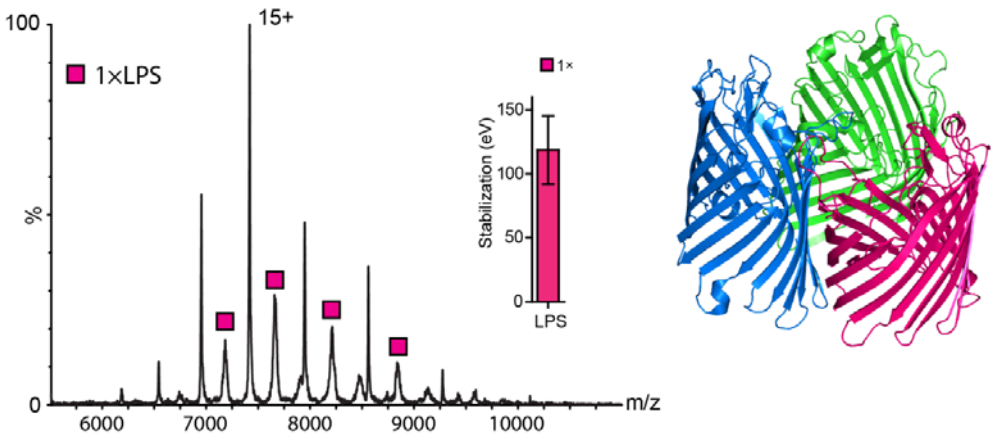


Figure 2.28: The effect of lipopolysaccharide lipid (LPS) on OmpF stability. OmpF (shown in the inset) was co-purified with LPS. The sample was subtilized in C8E4 detergent and mesuments were done at a protein concertation of 2.9μM.

The fully delipidated *Pf*MATE was incubated with CDL and phosphatidyglycerol (PG). The binding of two CDL molecules was observed giving rise to three populations in the spectra (Figure 2.29). Performing unfolding experiments and comparing the energy needed to unfold the lipid bound states relative to the unbound shows that lipid binding stabilises the protein complex in the gas phase, in addition these effects are additive. Hence two CDL bound increases the resistance to unfolding relative to one lipid bound only. Interestingly, in

the case of PG a slight destabilisation effect was observed in the case of one PG binding no significant stabilisation for two phospholipids bound.

These observations provide insights to structural stability upon lipid binding, while lipids that enhance gas-phase stabilisation, CDL in this case, may be structurally significant, other lipids like PG had no structural stability role but simply solvate the membrane protein.

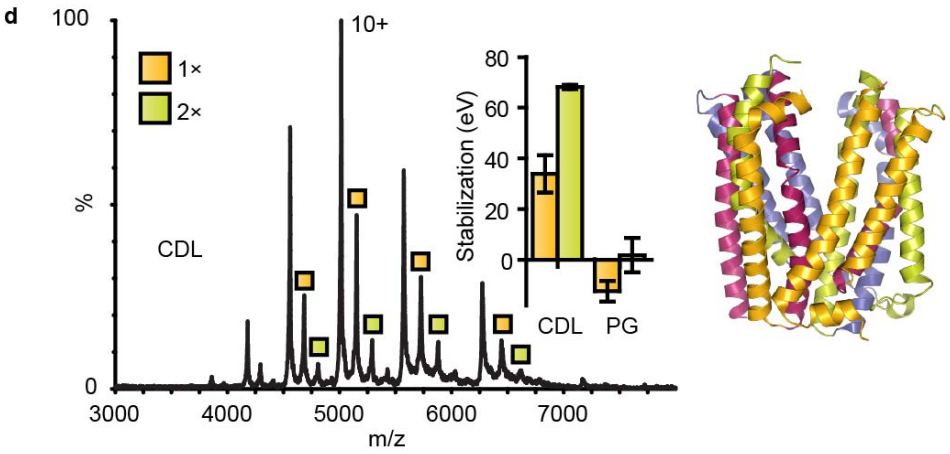


Figure 2.29: The effect of lipid stabilisation on *Pf*MATE solubilised in C8E4. Membrane protein *Pf*MATE (inset) was incubated with cardiolipin (CDL) at 7.5µM and phosphatidylglycerol (PG) 14µM. The protein concertation was 3.8 µM. The resistance to unfolding for up to two lipid molecules in each case was recorded. One lipid bound is denoted as yellow, and two as green.

2.9 Conclusion

In this chapter the fundamentals of electrospray ionisation in the presence of detergents are described in detail. The major focus was not only placed in finding conditions to maintain the native conformation but also to reveal crucial factors that dictate the membrane protein conformational landscape in the gas phase. Interestingly, in the case of membrane proteins detergent class has a significant role in the dependence of charge state distribution on polarity. For instance saccharide detergents are associated with a lower average charge state

distribution in negative polarity compared to positive, and this is positively correlated with increase in mass. However, in the presence of polyethylene glycol surfactants there is no significant change in the average charge state distribution between the two polarities along a broad mass range. Furthermore, achieving a lower charge state in negative polarity in the presence of soft surfactants is highly advantageous for preserving the native oligomeric state, while maintaining mild buffer conditions without the presence of charge reducing agents. These results are also important to reveal the mechanism of electrospray ionisation when surfactants are present in solution.

To minimize the high collision energy on membrane protein ions detergents with charge reducing properties or charge reducing organic solvent are utilized. However it has also been shown that the choice of surfactant perturbs the oligomeric state whereas organic charge reducing agents may lead to membrane protein unfolding. Harsh detergents such as C8E4 and LDAO are found to be highly MS compatible yielding highly resolved peaks with fewer adducts. They are associated with lower charge states, hence minimise Coulombic unfolding and at the same time less activation energy is required for their removal, significantly diminishing collisional unfolding (Figure 2.30).

Lowering the charge state is correlated with accessing energy levels which permit detergent removal while maintaining a native conformation. Saccharide detergents give lower charge state distributions in negative polarity, making this mode of operation preferable in assessing the native conformation and oligomeric states where saccharide detergents are necessary.

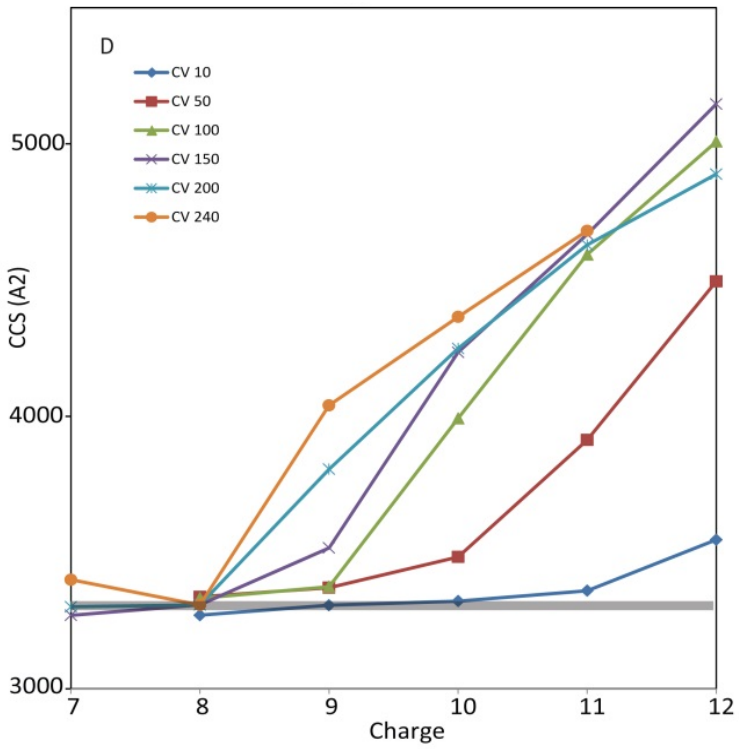


Figure 2.30: Charge reduction plays an important role in preservation of native folded state. Collision cross section of *Pf*MATE in C8E4, as a function of charge and energy showed that a low charge state and low activation energy are needed to preserve the native folded state. The CCS of the native structure is show in grey.

The surface charge density and E_{LAB} were found to dictate fundamental properties of the state of proteins in the gas phase (Figure 2.31). Higher charge state density is less compatible in preserving the native oligomer state, whereas when having a low surface charge density a high activation energy can be applied while preserving the native oligomeric state. This criterion is crucial for understanding that large proteins have low charge density which is why they can sustain the high activation needed for saccharide detergent removal. For instance ATPase is shown to be preserved in the native-like folded state even after high activation for DDM removal.[104] However, for small proteins charge reduction is necessary to access low charge state density and at the same time the E_{LAB} needed to remove the detergent needs to be sufficiently low to maintain their native-like state.

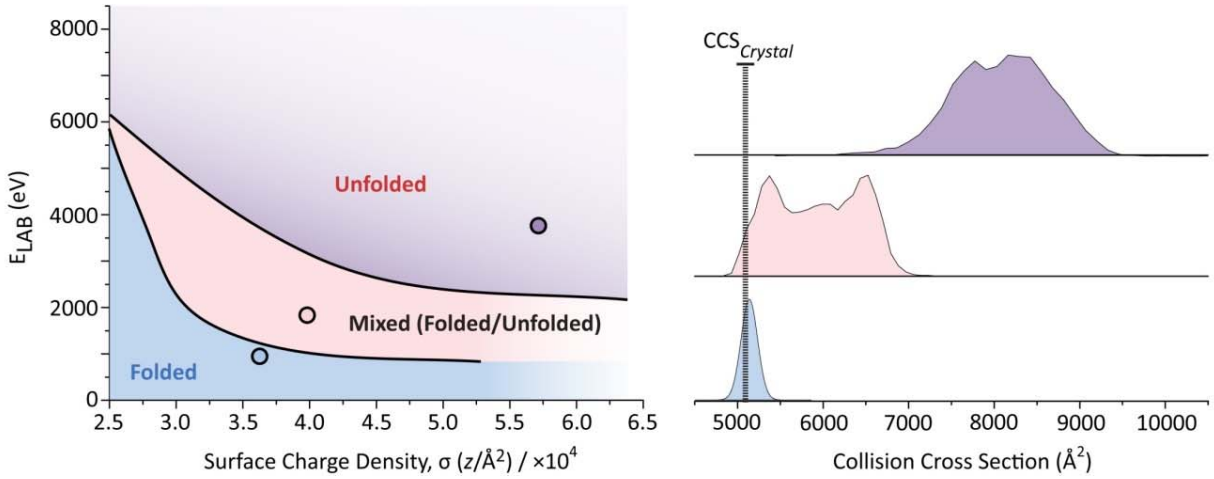


Figure 2.31: Protein stability diagram in the gas-phase as a function of E_{LAB} and surface charge density. Low E_{LAB} and charge density allows for preservation of folded state (blue). Increasing charge density and E_{LAB} leads to a mixture of folded and unfolded states (rose). Further increase in both parameters results in protein unfolding (purple). In the left a depiction of collision cross section for each regime is shown. The native-like CCS is indicated by the dashed line.

Finally by constructing an unfolding plot as a function of activation energy and comparing the energy required to unfold the bound relative to unbound species and ranking bound ligands based on their change in stabilisation, ligands that stabilise membrane proteins in the gas phase can be identified and their stabilizing effect can be quantified. This method is sensitive to low levels of stabilisation between bound and unbound states, identifying ligands that are structurally important.

Chapter Three: Tandem mass spectrometry of membrane protein complexes

The ability to fragment and precisely measure the mass of each component has been a useful feature of mass spectrometry as a tool for structure determination and identification of compounds.[158] By taking advantage of the current instrumental architecture ions are desolvated in the source region, mass selected in the quadrupole then activated and fragmented in the collision cell. This method of operation, often denoted as tandem MS or MS/MS, has been fully optimised for soluble protein complexes. Instruments are designed to handle soluble samples ionised in denaturing conditions, as is typical for proteomics experiments or the identification of small organic compounds.[42]

Three methods are usually used to derive ion genealogy. The most used method is precursor ion fragmentation. A particular ion is mass filtered in the first mass analyser. Then it is fragmented in a collision cell and all fragments are measured by the second mass analyser. In the second method all parent ions are screened in the first mass analyser and only those which generate a particular fragment ion are reported. The third method involves screening all parent ions and their fragments and reporting only particular parent-fragment ion pairs that have a neutral loss.[159]

To perform ion fragmentation elevated energies are needed. Typically ions can be activated in two instrumental regions; in the source region or in a designated collision cell. The energy in the source region is optimised for salt adduct or water removal, whereas the activating voltage applied in the collision cell allows for backbone fragmentation of denatured proteins. The quadrupole is positioned before the collision cell in the ion path, which allows mass selection to be the first step, followed by activation.

Although this instrumental configuration is suitable for soluble proteins, in the case of membrane proteins due to the increased complexity provided by the presence of detergents, the ability to perform tandem MS is hampered. As detergent is stripped in the collision cell the mass-to-charge ratio of the protein-detergent complex is unknown earlier in the ion path, as a result mass selection prior to activation is challenging. Furthermore, the total energy provided by the instrument is often not sufficient even for complete detergent removal, less so for subunit dissociation and/or fragmentation. To perform tandem MS on membrane proteins, the bound detergent must be removed before the quadrupole, allowing for mass selection of each species present.

To overcome these problems, two instrumental modifications were designed. First, the ability to remove bound detergent in the source region was developed. Applying elevated voltages in the source region or controlling the inlet pressure to reduce collisional cooling could be a sufficient method of detergent removal, giving access to the ability to mass isolate in the quadrupole. An orthogonal activation in addition to collision-induced dissociation (CID) is also developed. By taking advantage of elevated energies provided by infra-red radiation a laser beam was delivered to the collision cell and as a result it allowed even higher activation energies to be reached. The ability to use the quadrupole for mass isolation of membrane proteins is highly desirable, as it will make it possible to identify co-purifying lipids that are bound to these proteins. In addition, combining different activation methods allows different dissociation pathways of the same protein complex to be explored, providing, potentially, a complete map of subunit interactions.

In this chapter the focus is placed on activation methods, highlighting what governs the ability to reach different fragmentation pathways. Both CID in the source region and infra-red

multiphoton dissociation (IRMPD) in the collision cell are investigated with application to membrane proteins and lipid binding.

3.1 Factors affecting gas-phase dissociation

Several activation methods have been developed that are usually tailored to providing a specific dissociation pathway. Although the underlining atomistic mechanism that dictates the dissociation of multiply charged particles remains elusive, crucial feature that govern which pathway is followed have been identified.[160] Two of the most important criteria are the amount of energy that can be deposited and the trapping time during the activation. Other factors include the distribution of deposited energy and the energy dissipation, the form of the deposited energy and the kinetics of the dissociation reaction. Figure 3.1 shows the time scale of the most common forms of activation. The current methods cover a large time scale ranging from minutes to femtoseconds and as a result a variety of dissociation pathways can be observed. Three time regimes are distinguished, fast, slow and very slow; the latter is also known as slow heating.[161] Typically methods where activation happens in a long time scale are associated with higher overall activation energy relative to fast activation methods. For instance the total collision energy delivered to a protein is higher in the case of multiple collisions that happen in a microsecond time scale, whereas a single collision activation provides less overall activation energy however at a faster time in a range of picoseconds. These differences in the amount of energy and time of activation dictate the formation of different fragments, for example CID usually follows an asymmetric charge dissociation pathway whereas SID results in a symmetric one.[162-165]

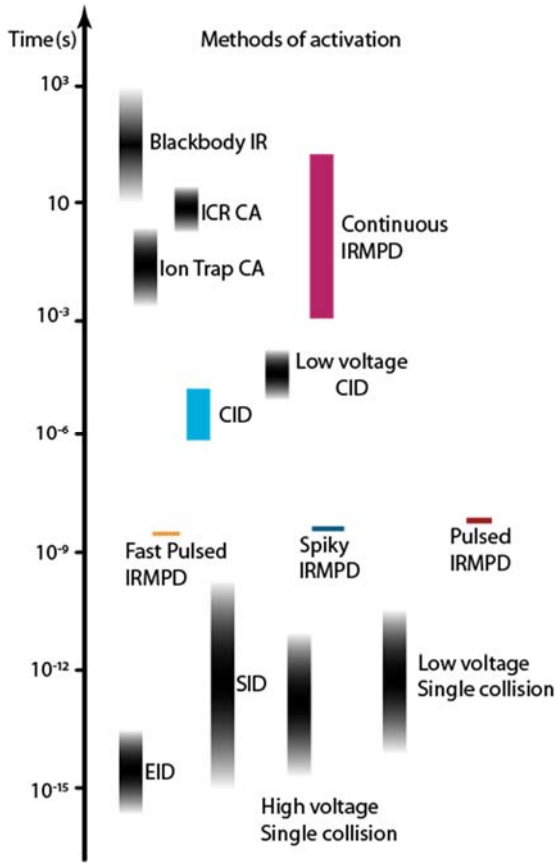


Figure 3.1: The time scale of the most common activation techniques. The methods used here are presented in colour, for comparison, other techniques are shown (black). Data were taken from McLuckey and Goeringer[160]

Other than the energy supplied, a crucial factor for ion activation is the rate of energy dissipation from the protein.[166] Typically, slow activation techniques suffer from a greater extent of energy dissipation, typically in a form of vibrational redistributions, or if trapped in the presence of bath gas, via collisional cooling. As a result in the case of slow activation methods the fragments usually derive from the paths requiring the lowest activation energy.[149] However if the energy is supplied faster than intramolecular vibrational redistributions an energetically higher dissociation pathway may be triggered. Here two slow heating methods CID and IRMPD are employed. In addition the time of activation of the IRMPD is varied by applying a pulsed and a continuous laser irradiation regime.

3.1.1 Collision-induced dissociation

Collisional activation is one the first dissociation techniques developed for ion dissociation or fragmentation due to its ability to be implemented in different instrumental designs and its simplicity of operation.[86,87] Ions are typically activated in a quadrupole or hexapole that is filled with a bath gas, usually nitrogen or argon. As ions are accelerated in the quadrupole they collide with the gas molecules leading to an increase in their total energy, and eventually dissociation.[149,167]The amount of energy that can be deposited to macromolecular ions depends on the relative kinetic energy, the number of collisions and the colliding protein-gas pair. The relative kinetic energy can be derived from first principles by considering the centre of mass collision energy (K_μ).[168,169] This method takes into consideration the relative mass differences between the macromolecular ion and the gas particle and expresses K_μ as a fraction of total kinetic energy (E_{LAB}) by assuming that neutral gas particles are statics. This is expressed in Equation 3.1, where m is the mass of an individual gas molecule and M is the mass of the protein ion:

$$K_{\mu} = (\frac{m}{m+M})E_{LAB} \tag{3.1}$$

Since the ratio of $m/(m+M)$ for a large ion colliding with a small gas particle will be small, this suggests CID will be a very inefficient process. However, carefully designed experiments have showed that the process can be highly efficient, reaching up to 90 percent of the total available energy to be transferred to the protein.[170] This discrepancy in theoretical derivation and experimental data is resolved by considering a direct momentum transfer between the complex and neutral gas particles, effectively removing the assumption that the velocity of small gas particles is zero.

Another factor that affects the efficiency of energy transfer is the number of collisions (n).[162] For a collision cell with a given length (l) operating at a certain pressure (p) and temperature (T), the number of collisions can be derived as:

$$n = c(lp\Omega T^{-1}) \tag{3.2}$$

Equation 3.2 shows that the number of collisions is dependent on collision cross section (Ω). The collision cross section can be approximated from the radius of protein ions and gas particles and as a consequence the number of collisions is directly dependent on ion and neutral particle mass. It can be derived that for large ions with a mass greater than 100 kDa the number of collisions reaches tens of thousands, whereas for small proteins is typically in the range of hundred to a thousand collisions. Since the length of the collision cell is fixed, at a given acceleration voltage and pressure, the time it takes large ions to pass through the collision cell was found to be in the range of hundreds of microseconds and increases as the mass of the protein is reduced. For low mass proteins the time spent in the collision cell can reach the low millisecond time scale. Hence, for large ions the high efficiency of energy transfer to macromolecular complexes from small gas particles can be attributed to the large number of collisions, whereas in the case of small ions the increase in time dominates. For a known number of collisions at a particular accelerating voltage the increase in internal energy can be calculated as:

$$\Delta E = zV\left(1 - \left(\frac{(m^2+M^2)}{(m+M)^2}\right)^n\right) \tag{3.3}$$

Equation 3.3 shows that the amount of increase in internal energy is directly correlated to charge (z), making highly charged proteins more prone to dissociation. Calculation of the total internal energy is crucial for activation and ion transition efficiency, especially for large ions. Due to the increase in internal energy, large ions reach significantly higher velocities

that can then be difficult to focus. On the other hand, Equation 3.3 does not take into consideration energy dissipation which lowers the internal kinetic energy and improves transmission.

Notable the change in internal energy of a given ion is directly proportional to the dissociating pathway. In the case of native MS the dissociation of non-covalent bonds is statistically more favourable than covalent bond cleavage. In addition, charge separation follows an asymmetric pathway where monomers are stripped with a high charge, as expected from their mass. For instance an octamer will not dissociate into two tetramers with equal charge or any other combination other than a monomer with significantly more charge than 1/8 of the charge of the complex, and the remaining heptamer with significantly lower charge. Upon further increase in accelerating voltage the same dissociation pathway will be reinitiated in which case the stripped heptamer serves as a precursor ion dissociating into a monomer and hexamer.

Since the ejected high order oligomers have a significant reduction in the charge state the dissociation steps are usually limited to a small number, typically one or two. This dissociation pathway requires the lowest amount of increase in internal energy. Further, if the charge state of the complex is increased by addition of charge increasing agents, such as sulfoxane, a more symmetric dissociation pathway is observed as the increase in Coulombic repulsion allows for higher energy dissociation pathways to be accessed.[89] The asymmetric dissociation pathway can be explained by considering that the ejected monomer initially unfolds to accommodate more charge before dissociating from the complex. This observation is supported by ion mobility where larger than expected rotational averaged collision cross sections are measured for monomers, and also by chemical cross-linking experiments where

monomers dissociate symmetrical due to their inability to unfold in the presence of other subunits that are chemically linked.

Taken together the general consensus is that the CID mechanism is initiated by acceleration of ions based on charge and applied voltage. Thousands of protein ion collisions with gas particles will increase the total internal energy of the protein. Subsequently, internal vibrational forces will induce local unfolding within the monomer. Both the secondary structure and inter subunit interactions are maintained by the same type of bonding. Although in secondary structure multiple bonds are formed they are in close proximity to each other and bond formation is highly cooperative, unlike subunit-subunit interactions that are less in total number but a further distance apart. As a result breaking one interaction has an insignificant effect on other interactions. Hence, once a subunit starts to unravel less energy is needed to completely unfold it compared with the energy needed to break every non-covalent interaction between subunits. Subunit unfolding increases the surface area making it possible to accommodate an increased amount of change. The Coulombic repulsion also contributes to a larger extent of unfolding which leads to a reduction of non-covalent interactions between subunits and eventually complete separation of the unfolded monomer from the protein.

3.1.2 Infrared multiphoton dissociation

IRMPD takes advantage of infra-red laser irradiation to activate ions which leads to their dissociation. The most commonly used modality of the technique is continuous wave (CW) CO_2 lasers with a 10.6 μm wavelength of radiation. Similar to CID this method is known to be a slow heating method. It requires hundreds to thousands of absorbed IR photons for the dissociation to take place. However, unlike CID where the change in internal energy is directly dependent on the charge state of the ion, during IRMPD activation the amount of

energy delivered to an ion does not depend on the charge state and can be increased by increasing the intensity of the IR radiation without compromising ion trap performance or inducing ion losses. IRMPD has been successfully employed in mass spectrometry for sequencing of peptide and protein backbones.[171-176] Usually long (≥10 ms) periods of irradiation are necessary to provide sufficient numbers of absorbed photons in IRMPD. This requires confining the ions in an ion trap for the period of irradiation. Elevated gas pressures are essential for trapping ions of intact protein assemblies,[32,177] and therefore the energy-dissipating effect of collisional cooling is unavoidable during their activation by continuous IR radiation. Collisional cooling, however, leads to a significant reduction of IRMPD efficiency for trapped ions.

Pulsed IR lasers can provide high intensity output in the MW range *versus* 10–100 W power range of continuous IR lasers, accompanied by equally significant reduction in activation time (from ≥10 ms to ≤100 ns). Figure 3.1 shows a comparison of activation time of CID, CW IRMPD and the pulsed IRMPD that are employed in the investigations of this chapter. While both CID and IRMPD are in the millisecond to microsecond time scale respectively, pulsed IRMPD is in the nanosecond time scale where 50 ns and 100 ns pulsed lasers are used. In the pulsed laser regime the energy can be delivered to the gas-phase ions faster than the rate of radiative cooling, typically in 1–30 s^{-1}, thus the interference from these two energy dissipating mechanisms can be avoided.[178,179]

3.2 Tandem MS of membrane proteins

Collisional activation is typically applied in a designated collision cell. In a Synapt instrument this is performed in the trap cell which is positioned after the quadrupole. Since this configuration does not allow for mass selection prior to detergent removal, in this chapter

the possibility to remove the detergent in the source region was investigated. Previous studies[180] have shown that ion activation by increasing the acceleration voltage in the cone provides similar fragments to those observed by CID in a collision cell. However, the front optics, originally designed for salt adduct removal and desolvation cannot supply a high enough energy suitable for saccharide detergent stripping. In the case of polyethylene glycol detergents where a low energy is needed for detergent removal the energy supplied by the front end optics is sufficient for their release from the detergent micelle, however often spectra quality is compromised by poor desolvation and the presence of adducts.

Rather than increasing the acceleration voltage an increase ion activation can also be provided by lowering the source pressure.[181] As ions enter from atmospheric pressure to the mass spectrometer, they experience a pressure difference sufficient to accelerated them and increasing their total kinetic energy, and as a result large ions are difficult to focus by the initial ion guides. For better transmittance of large ions applying a backing pressure lowers the total ion kinetic energy via collisional cooling, which makes the ions easier to focus. Many membrane proteins are relatively small and still relatively easy to focus even at high kinetic energies.

To perform MS/MS the aim is to activate the ion prior to quadrupole selection, hence a fine balance between transmission and activation has to be established. Here differences in source pressure were applied and the efficiency of detergent removal is compared with its reciprocal activation in the collision cell. To minimize low transmission at low pressures a relatively low mass protein was selected. Measurements were performed using NapA, a dimer in its native oligomeric state, in the presence of bound exogenous PG lipids (Figure 3.2).

The activation energy in the collision cell was maintained at 50 V and the pressure was sequentially lowered from 7.5 mbar to 3.0 mbar resulted in consecutive removal of detergent and eventually a release of the intact dimeric species. Comparing these results with similar activation in the collision cell showed that pressure activation is almost identical to CID in the trap and the lowest pressure is sufficient to provide energy as high as 110 V in the collision cell. In addition to monitor the conformational change at different pressure regimes, the 12+ ion was selected and a "pressure-induced" unfolding plot was generated by recording the ions arrival time distributions at different pressures. Comparing this plot with collision-induced unfolding performed in the ion trap shows similar unfolding features. Initially a reduction in arrival time is observed in both cases followed by the first unfolding step at 3.0 mbar or 110 V.

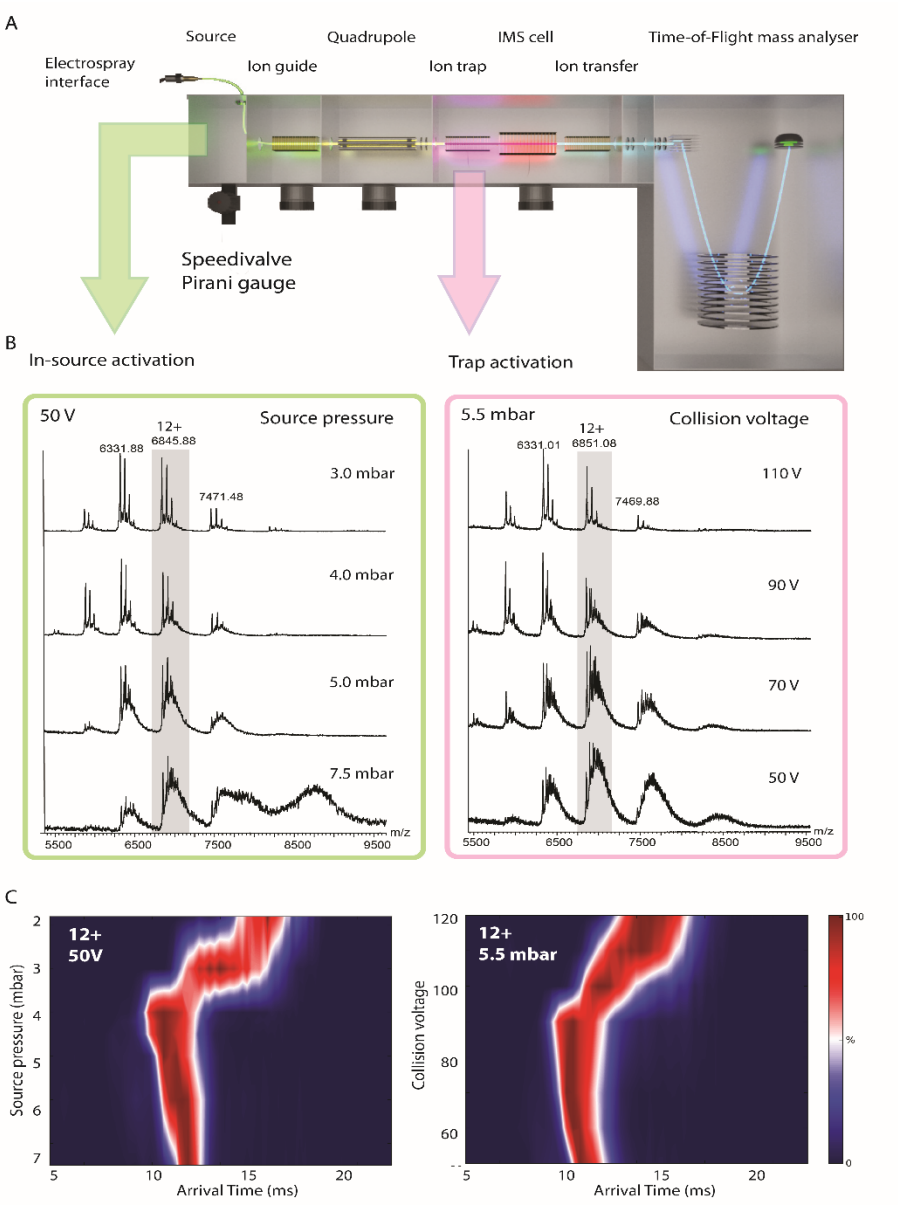


Figure 3.2: Lowering the source pressure is sufficient for detergent removal. (a) Instrumental configuration depicting both activation regions, the source (green) and ion trap (magenta). (b) Mass spectra of NapA showing a comparison of detergent removal between activation by trap voltage or source pressure. (c) Arrival time distribution comparison between the two activation methods of charge state 12+ highlighted in grey in (b). The protein was solubilised in C8E4 detergent using a SEC column. Measurements were obtained by Michael Landreh.

These results demonstrate that pressure reduction in the source region is not only effective for complete detergent removal but it can be tightly controlled and the gain in

energy activation is comparable with voltage activation in the collision cell. Furthermore, as activation at low pressure is similar to CID the two effects can be combined allowing for an even greater increase in the ion's total internal energy. A benefit of this combination is direct observation of additional unfolding steps that are not attainable by one activation method (Figure 3.3). For instance, activation of the NapA dimer, by applying 200 V in the collision cell, led to observing species with arrival time distributions of 16.3 ms. Pressure reduction in the source region was sufficient to unfold the species further and species with an arrival time distribution of 18.1 ms were observed. In addition the combination of both activation methods led to dissociation of monomers from the complex. This example serves as a proof of principle, suggesting that by combining the activation energy in both regions sufficiently high energy is generated leading to subunit dissociation that can be used to construct an interaction map of subunits in a protein complex.

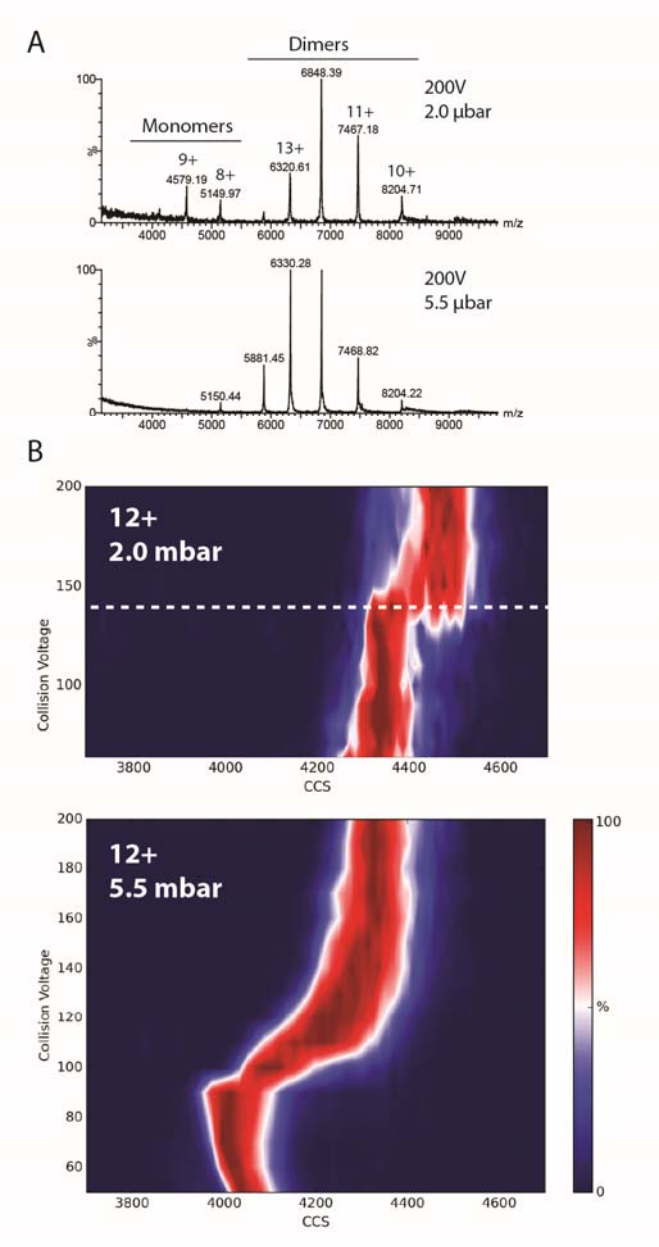


Figure 3.3: Combination of both activation in source at low pressure and CID in the trap is sufficient to dissociate the NapA dimer to monomer (a) Mass spectra recorded at 200V and 2.0µbar show both intact dimer and ejected monomer. (b) A high activation regime allows for direct visualisation of additional unfolding states. The protein was solubilised in C8E4 detergent. Measurements were performed by Michael Landreh.

The removal of bound detergents in the source region allows mass selection in the quadrupole. By applying 40 V in the trap and a pressure of 2.3 mbar in the source, ions with a

particular *m/z* value can be mass selected and the bound lipid species can be observed bound to the proteins or in the low *m/z* region after activation of the selected ion (Figure 3.4). As high kinetic energy leads to poor ion focusing and the kinetic energy is mass dependent, a larger complex, AmtB with a mass of 138 kDa was also used to investigate if the method can be applied to larger ions. The increase in total kinetic energy would lead to a more efficient detergent removal. Figure 3.5 shows a spectrum of AmtB at 5.5 mbar activated at 100 V in comparison to a pressure of 2.0 mbar and applying only 30 V in the collision cell. Using a high pressure of 5.5 mbar in the source region and an activation of 100 V in the trap was sufficient to resolve charge sates, however, detergent is still bound to the protein. In the case of low pressure and relatively low collision voltage, as expected, complete detergent removal was observed.

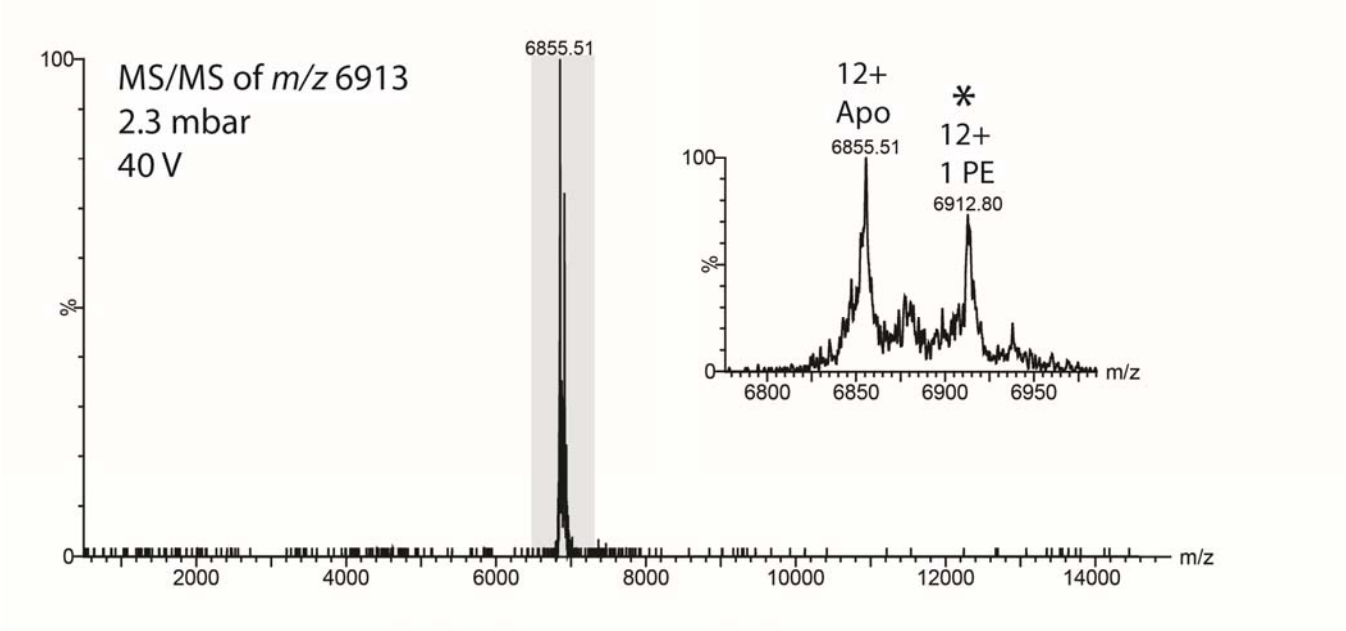


Figure 3.4: Tandem mass spectra of NapA in C8E4 show that low source pressure activation allows for MS/MS selection. Inset shows a zoomed region of slected 12+ charge state of NapA both *apo* and one PE bound.

Although activation at low pressure was sufficient to remove the bound detergent and preserve lipid binding the total ion intensity was low. MS measurements on AmtB at a pressure of 2.0 mbar gave poorly resolved spectra due to insufficient ion focusing. Surprisingly, further increase in the collisional voltage while keeping a low pressure regime in the source region resulted in complete lipid removal that was not previously feasible by using one method of activation. The improvement in intensity can be attributed to increased homogeneity of ions where after lipid removal the total amount of the unbound form was increased. Lipid removal was not observed in the case of high pressure even at the maximum energy provided by the instrument, demonstrating that measurements in the low pressure regime can be used for generating homogeneous spectra while preserving non covalent-interactions between protein subunits.

Taken together these measurements demonstrate that additional activation can be accessed by using a low pressure regime in the source region. The increased internal energy is sufficient to remove bound detergents from a membrane protein complex in the source and allows for tandem MS of membrane proteins. Further, lipids and subunits can be dissociated by combining activation provided by both high voltage in the collision cell and low pressure in the source.

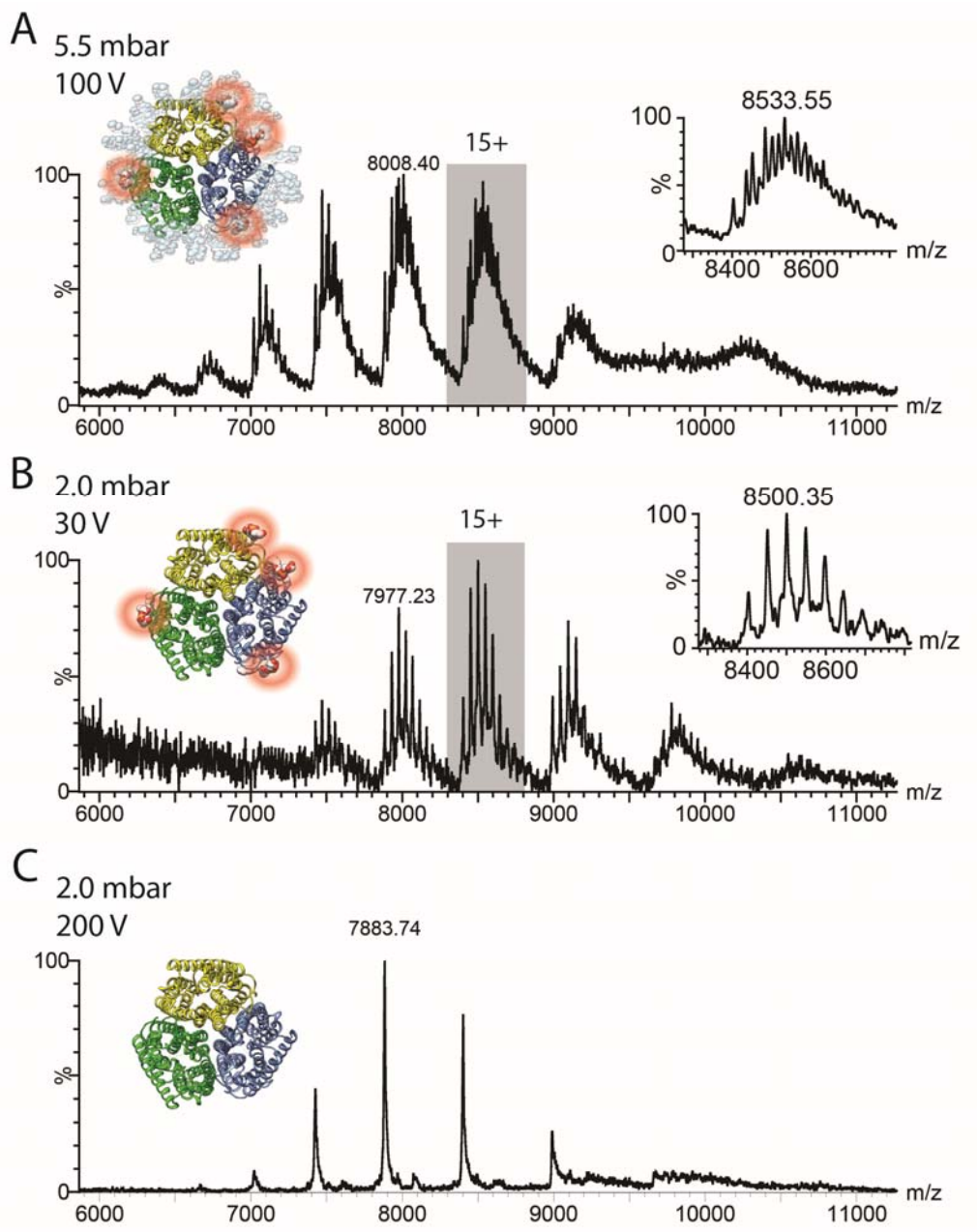


Figure 3.5: In-source activation of AmtB bound to PG lipids. (a) Detergent is removed by applying collisional activation in the trap. Only partial detergent removal was observed. (b) At minimum trap activation complete detergent removal is performed by activation due to low pressure in the source. The spectra quality is severely compromise due to poor focusing of the large ions. (c) Activation in both the trap and source region results in removal of lipid bound species and only the unbound form is resolved.

Although these low pressure measurements enable tandem MS of membrane proteins, they were performed using membrane proteins solubilised in polyethylene glycol detergents that can be easily removed in the gas phase. In addition, the large proteins gave low intensity spectra, hence an alternative approach was also investigated. Rather than lowering the pressure in the source, the total internal energy was increased by applying a higher voltage in the cone. At increased cone voltage and high pressure total ion energy is increased whereas the kinetic energy is maintained at low levels. As a result large ions can be better focused and at the same time ions are excited to a higher energy regime. The maximum cone voltage provided by commercial instruments is not sufficient for removal of saccharide detergents. Therefore the instrument was modified to increase the cone voltage from 200 V to 400 V. Modifications included adjustments in the electronic board and introduction of a tuneable power supply to control the extended voltages in the sample and extraction cones, as well as customisation of the control software. These modifications were designed by Jonathon Hopper and were made on a Synapt platform for enhanced selectivity and sensitivity.

To investigate if the increased energy is sufficient for detergent removal OmpF solubilised in OG detergent was used. Mass spectra collected at 300 V on the sample cone, which is twice the voltage typically used in conventional instruments, showed complete detergent removal and release of the intact trimeric oligomer. Although OmpF is a relatively large protein (111 kDa), a high signal intensity was achieved. Hence, these modifications allow for detergent removal from membrane proteins with high mass and solubilised in both detergent classes, making this method less sample demanding than application of low pressure.

The same measurements were also performed using both AqpZ and AmtB solubilised in C8E4 detergent. In both cases detergent removal was achieved without compromise in

spectra quality. Similar to applying low pressure in the source, removal of detergent in the front optics enables mass selection in the quadrupole, and in this case saccharide detergents can be used as well. Figure 3.6 shows the mass selection 24+ ion from OmpF solubilised in OG after removal of bound OG detergent. This method is used in section 3.2 to identify lipids bound to membrane proteins.

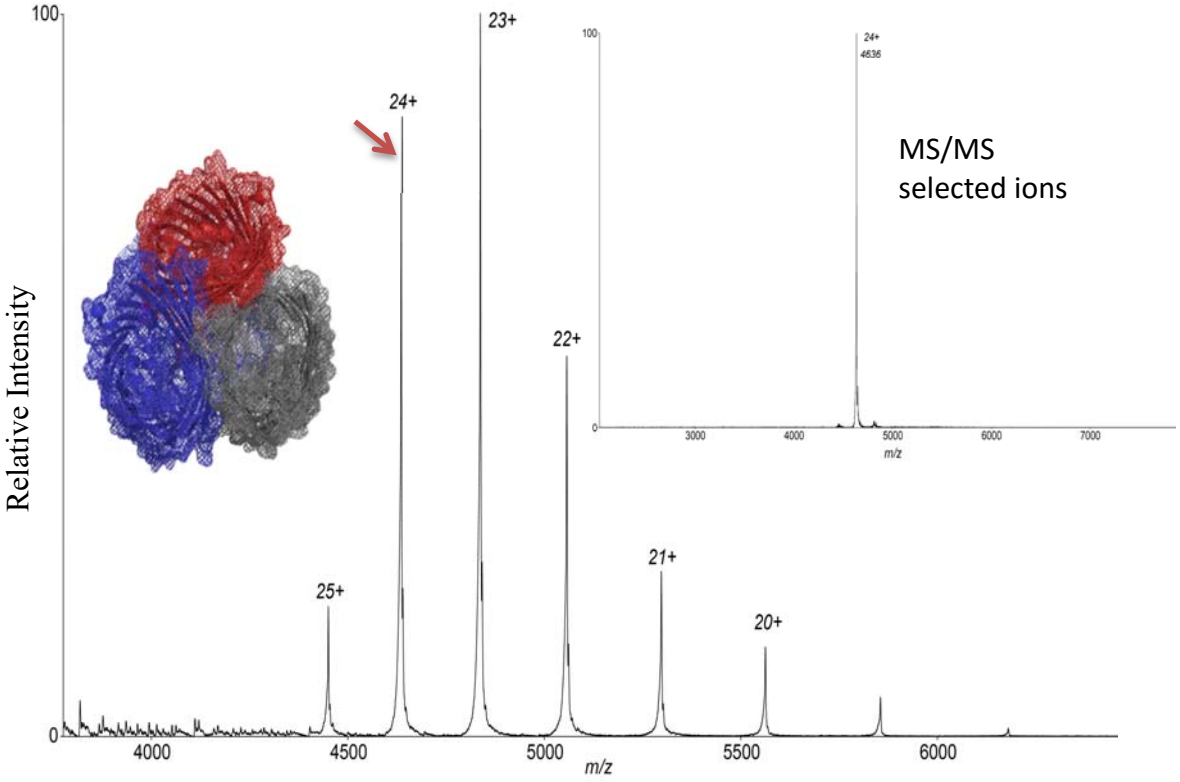


Figure 3.6: Mass spectrum of OmpF solubilised in OG after detergent removal in the source by applying 300 V on the sample cone. Inset shows the mass section of 24+ ion.

The results presented here show that MS/MS selection can be performed on membrane proteins by removing the bound detergents in the front end of the mass spectrometer. Without instrument modification but by simply lowering the pressure in the source detergents can be dissociated from small membrane proteins solubilised in C8E4. Instrument modifications to apply a higher voltage on the sample cone allows for tandem MS on larger membrane proteins solubilised in both polyethylene glycol and saccharide detergents.

3.3 Infrared multiphoton dissociation of membrane proteins

Here pulsed and continuous IR radiation as a means of releasing membrane protein complexes from gas-phase micelles for MS analysis has been explored. Several instrumental modifications were introduced in order to apply pulsed and CW IRMPD as shown in Figure 3.7. The Ultima MS instrument was modified for high-mass measurements. A long wavelength IR laser was introduced to the collision cell via gold-coated copper mirrors positioned in the TOF region and an IR window. The collision cell was converted to a linear ion trap by controlling the entrance and exit electrodes. The laser pulsing and trapping of ions were synchronised by a waveform generator. These instrumental modifications were designed by Victor Mikhailov.

MW-IR radiation from either pulsed or continuous lasers can effectively eject detergent molecules from a membrane protein detergent complex, and its charge states can then be resolved. Furthermore, releasing membrane protein complexes from detergent clusters by 100 ns radiation is possible for a range of detergents. Pulsed 50 ns activation was also able to release membrane proteins from gas-phase detergent clusters, but the efficiency was somehow lower, and a part of the broad ion signal from the protein in micelles was still observed. This could be due to the much lower pulse energy of 50 ns as compared with 100 ns of radiation, 40 mJ *versus* 550 mJ respectively.

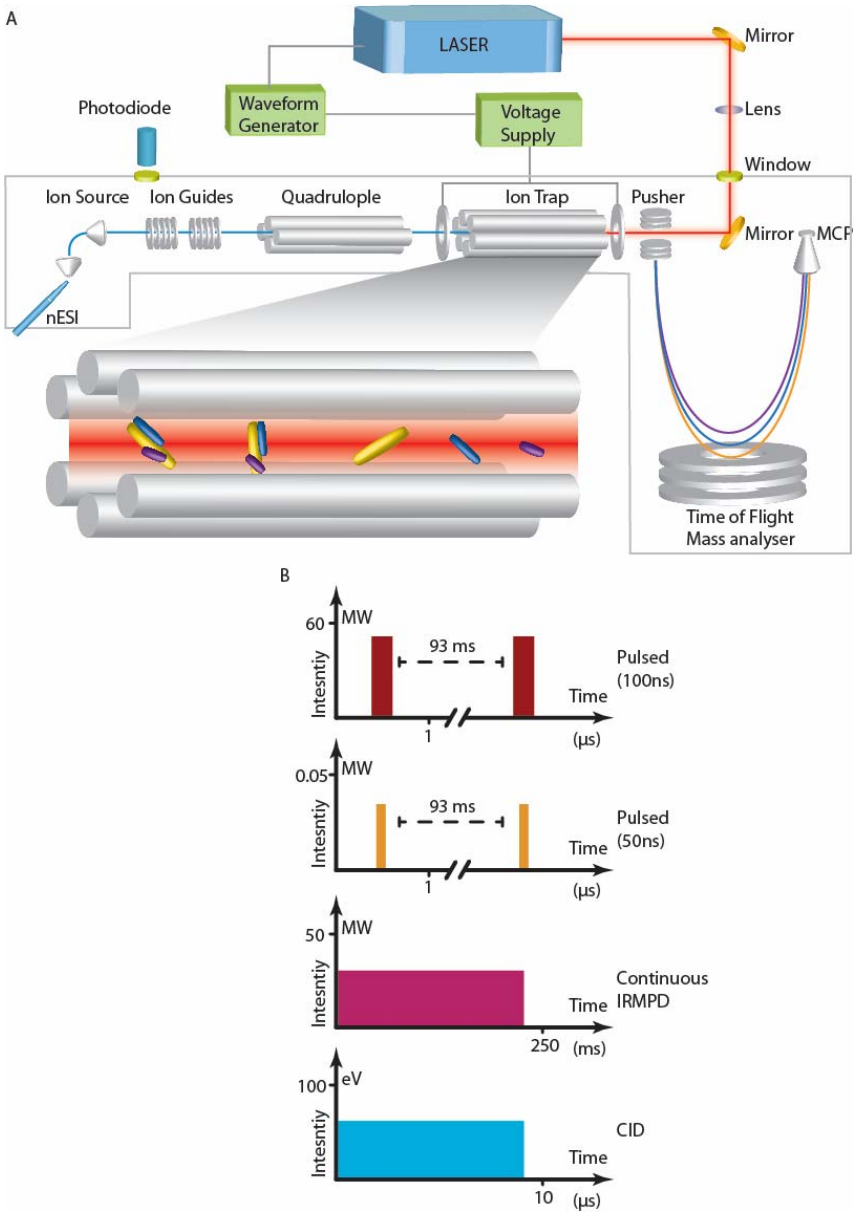


Figure 3.7: Schematic of the IRMPD setup and activation regimes. (a) A depiction of the modifications made to introduce a laser in the trap. (b) Intensity profile of the pulsed regions, continuous IRMPD and CID used in this chapter.

Mass selection of proteins in detergents with a large selection window was found to improve the mass spectra quality by reducing the amount of chemical noise from the detergent cluster and provide additional ion activation. Under IR activation partial dissociation of the membrane protein complex liberated from the detergents was observed,

but this effect depended on the protein, detergent, laser power and the region of the ion signal selected for activation (Figure 3.8). For instance, when the ions from the descending slope of a broad signal from V-type ATPase solubilised in DDM micelles was selected for IR activation, this polyprotein complex was liberated intact in relatively low charge states centred around z = 37+, and no noticeable dissociation was recorded**.** Mass selection of the ions from the ascending slope of the signal followed by IR activation resulted in higher charge states of the intact ATPase, around z = 45+ and stripped subunits were observed.

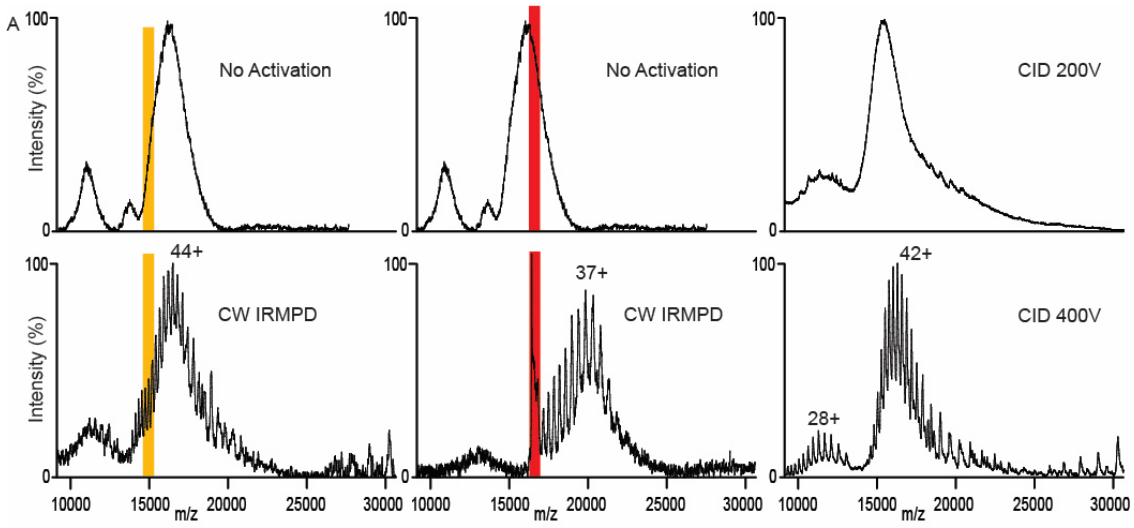


Figure 3.8: Mass selection results in different charge state distributions. ATPase spectra recorded using CW IRMPD and CID activation. By selecting a lower *m/z* value (yellow) high charge states are observed and whereas a high *m/z* mass provided low charge state species. CID measurements showed that conventional instruments are not able to resolve the complex. Only after modification to high activation the complex can be resolved. IRMPD data were obtained in partnership with Victor Mikhailov.

It was not possible to release the intact ATPase, a 26-subunit membrane protein complex, from DDM micelles using maximum collision voltage of 200 V, which is a standard upper limit in many instruments (Figure 3.8). In contrast, all IR activation mass spectra of ATPase were obtained using a standard continuous IR laser commonly used for IRMPD applications.

Mass spectra of detergent-free ATPase could only be produced by CID using collision voltages in the range of 300–400 V on a specially modified Q-TOF2 instrument, additional evidence of the efficiency of IR activation in liberating membrane protein complexes from gas-phase detergent clusters.

The ability of IR activation to remove the bound detergents but retain lipid bound species was investigated. As shown in chapter two, using controllable collisional activation, it is possible to liberate an intact membrane protein complex from gas-phase detergent clusters and retain the bound lipids at the same time. However, since IR radiation is absorbed by detergent molecules as well as by lipids it can be expected that in the case of IR activation the lipids could be lost when the protein is released from bound detergents. To test the ability of IR activation to preserve bound lipid species the membrane proteins AqpZ was incubated with exogenous PG lipids and activated by IR radiation. The result showed that IR activation is sufficiently gentle, and the lipid-bound protein complexes can be observed in the mass spectrum (Figure 3.9).

In all cases of both 100 ns and 50 ns pulsed regimes, after IR activation the membrane protein complexes underwent some degree of dissociation after micelle ejection. Asymmetric fragments were formed similar to CID of protein complexes and therefore determined by the internal vibrational redistribution mechanism. Taken together the present results demonstrate the ability to release proteins from detergent clusters in the gas-phase by IRMPD while preserving lipid-protein non-covalent interactions.

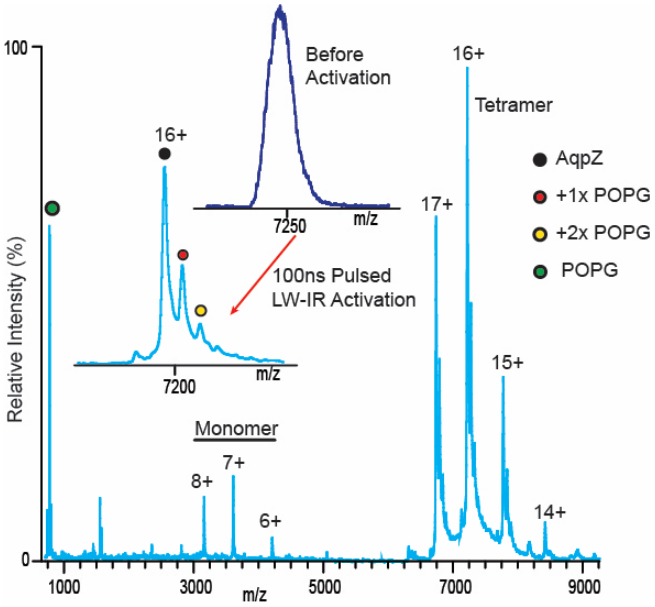


Figure 3.9: Mass spectra of Aqpz bound to POPG lipids and activation by IRMPD. IR-Laser activation was found to preserve lipid binding. The protein to lipid ratio was maintained 1:5 and the protein concertation was 7µM.

3.4 Direct lipid identification

Lipid-protein interactions are important for modulating the structure and function of proteins associated with cellular membranes. Prerequisite in probing theses interactions is identification of the lipids that are bound membrane proteins. The most widely used method for lipid identification is lipidomics.[182] In this method the purified protein-lipid complex is denatured in acidic solution, and the lipids are extracted using organic solvent. They are then separated and identified using liquid chromatography (LC) coupled to MS. The method performs well in identifying a large variety of lipids however it has two major drawbacks. First sample pre-treatments are sometimes associated with lipid loss since extraction of the entire lipid envelope is case dependent. Second, the lipid identified could not be in directly bound to the protein but simply co-purified with the detergent micelle.

To overcome these problems a non-denatured platform for identification of lipid-bound species was developed. By applying high a voltage on the sample cone, each peak bound to a membrane protein can be mass selected and fragmented to allow direct identification of the bound species. The process involves the following steps. First a membrane protein complex is sprayed in non-denaturing conditions and the presence of additional bound peaks is observed. Subsequently, detergent is then removed in the source by applying elevated voltages on the cone. The adduct peak is mass selected in the quadrupole and fragmented in the trap cell releasing the lipid from the protein. Finally, high energy is applied in the transfer region to fragment the lipid further, for full identification. The ability to fragment in the transfer allows the parent ion to be matched with its fragments as both parent and fragment ions will have the same arrival time distributions.

To test the feasibility of this method a small membrane protein, *Pf*MATE, was selected. The protein was extracted from *E. coli* membranes in which it had been over-expressed with excess amount of detergent for complete delipidation. The sample was then incubated with 13 μM of POPC and electrosprayed in the presence of C8E4. The detergent was completely removed in the source and the resulting mass spectrum is shown in Figure 3.10. Up to two lipid bound peaks could be observed in the high *m/z* region whereas is the low *m/z* region the spectrum is highly populated and POPC bound to a sodium ion is the most dominate peak.

Mass selection of the charge 10+ adduct peak with a small selection window shows that this peak could be mass filtered and separated from the unbound peak. Without trap activation no peak was observed in the low *m/z* region. Increasing the voltage to 70 V, the unbound form at charge 10+ and 9+ are observed but the intensities of the unbound and the adduct peaks for charge 10+ are almost the same. Further increase in the collision voltage to 110 V significantly dissociates the adduct peak giving rise to increased intensity of both the

unbound 10+ ion and the charge stripped 9+. These measurements show that at low activation energy lipids leave predominately as neutral species, and as the energy increases they are more likely to become charged. As expected in the low *m/z* region a singly charged peak with a mass of 761 Da corresponding to POPC is observed, showing that the method is feasible for lipid identification by dissociating the lipid-protein complex.

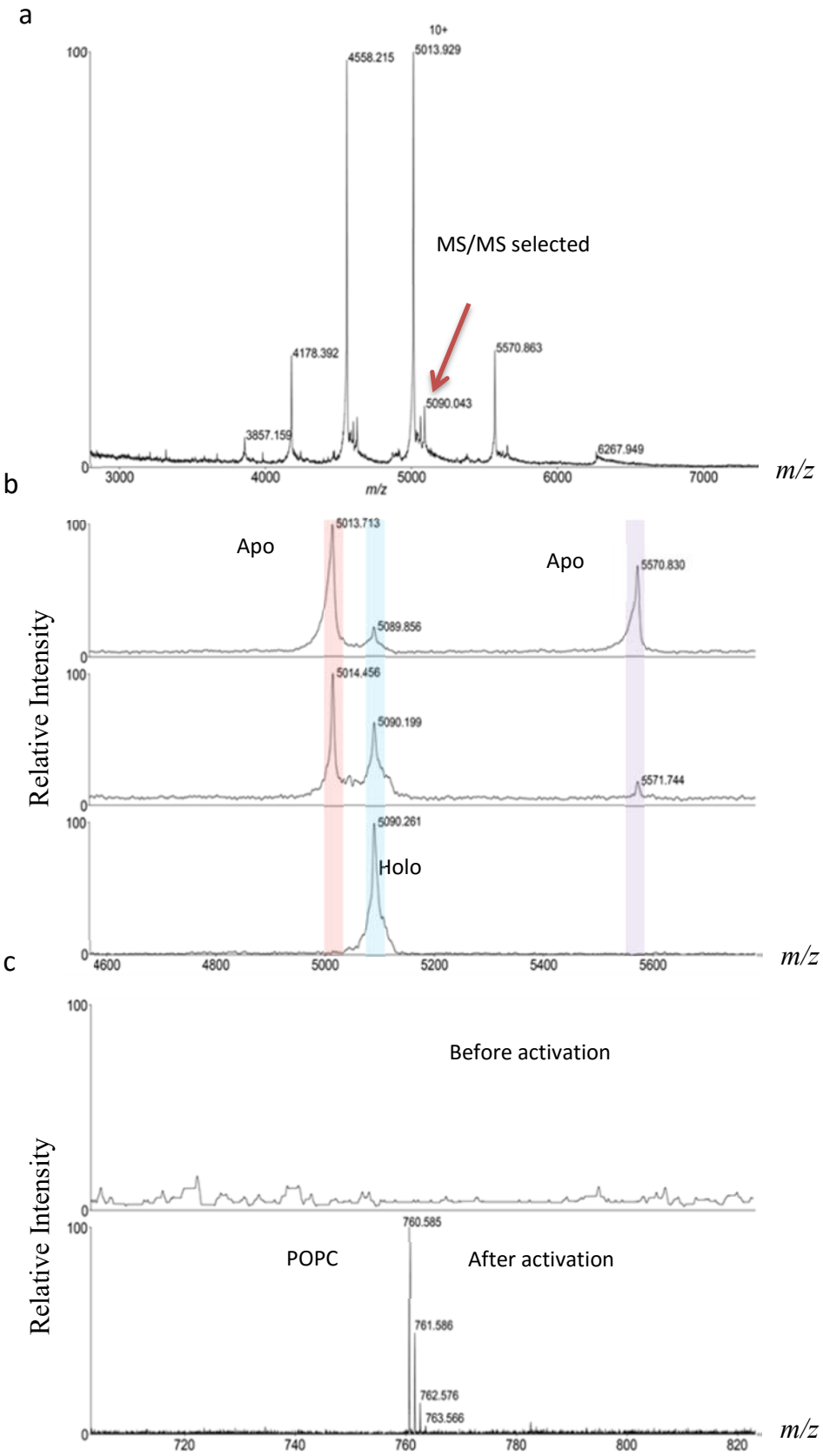

a
10+
5013.929
4558.215
MS/MS selected
4178.392
5570.863
5090.043
3857.159
6267.949
m/z
b
Apo
Apo
5013.713
5570.830
5089.856
5014.456
5090.199
5571.744
5090.261
Holo
Relative Intensity
m/z
c
Before activation
760.585
POPC
After activation
761.586
762.576
763.566
Relative Intensity
m/z

Figure 3.10: Identification of exogeneity bound lipids (POPC) to membrane protein *Pf*MATE by tandem MS. (a) The protein was solubilised in C8E4 and the detergent was removed by activation in source. After detergent removal lipids adduct peaks are observed, and mass selected (b). Increasing the energy from 20V (bottom spectra in (b)) to 80V (middle spectra) results in neutral loss of lipids highlighted in red and charge stripped species (purple). The bound lipid species (blue) are further activated (top spectra) with 120V for an addition removal of lipid bound speies. The mass filtered ions are activated in the trap and charged lipids are observed(bottm) in the low mass region (c) with isotopic resolution. The protein to lipid ratio was 1:2 and protein concertation was 2.8µM

Notably the mass addition to the unbound protein in the high *m/z* region cannot be measured precisely, for instance in this case the mass measured was 760 ± 20 Da, hence it is difficult to identify the lipid, whereas the measurement in the low *m/z* region is isotopically resolved and can be assigned to POPC with a high level of confidence.

In cellular membranes, membrane proteins are solvated by several lipids, and the ones that are tightly bound or structurally important are usually co-purified. Therefore it was of interest to see if it was possible to simultaneously identify different lipids present in a mixture. To show that the method is applicable to different samples and capable of identifying each bound-lipid in a mixture, AqpZ was used and incubated with POPC and DPPC (Figure 3.11). The mass difference between these two lipids is 26 Da resulting in only 1.5 Th at charge state 17+. Although the two lipids are bound to the protein in the high *m/z* region they cannot be resolved: four adduct peak are detected with a mass of 752 ± 35 Da, corresponding to approximately the average mass (748 Da) of the two lipids.

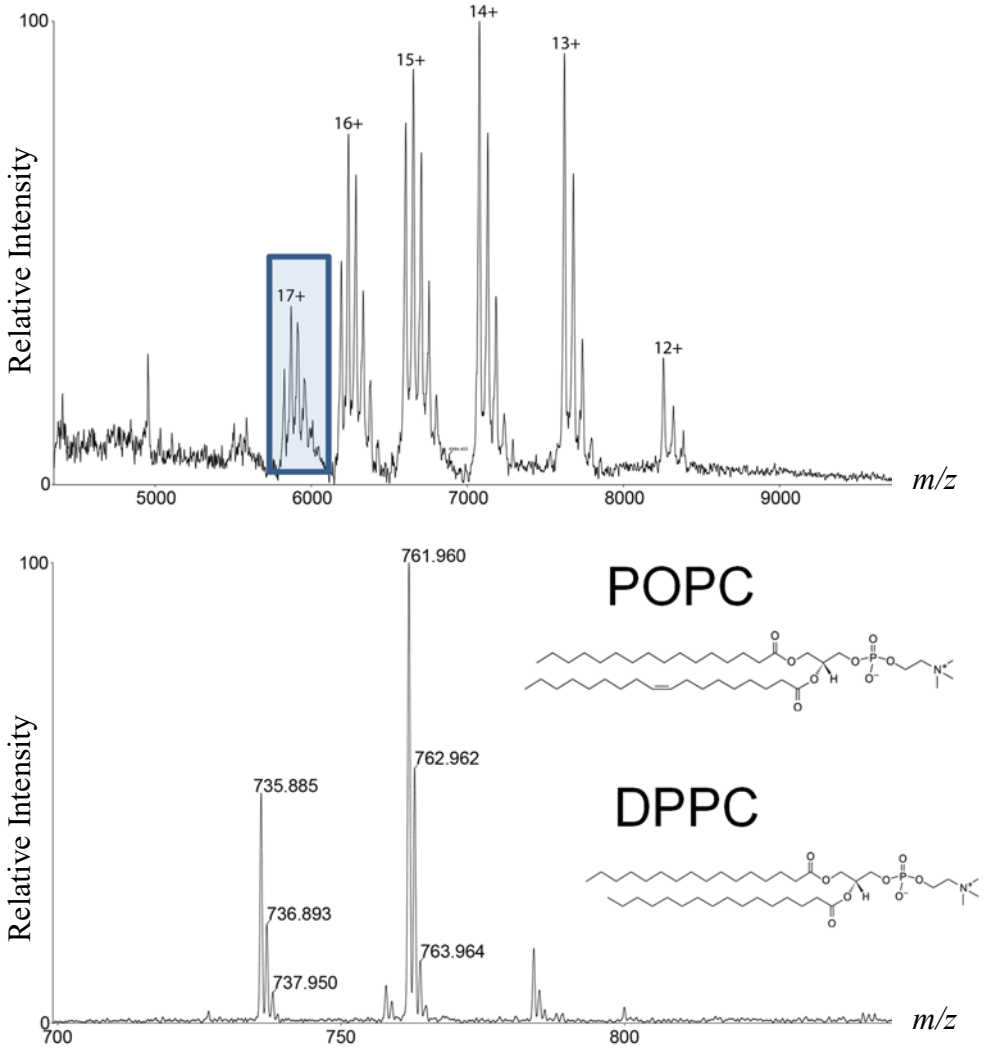


Figure 3.11: Identification of mixture of POPC and DPPC lipids bound to AqpZ. (a) Mass spectrum shows that the mixture cannot be mass resolved. (b) High-voltage in-source activation, mass selection of the 17+ ion (highlighted in blue) in the quadrupole and high activation in the trap are used to fragment unresolved bound lipids and isotopically resolve them in the low *m/z* region. A mixture of 1:1 POPC and DPPC was used and the protein concertation was maintain at 4.7µM. Prior to measures the protein was solubilised in C8E4

The first adduct peak at charge state 17+ was mass selected and dissociated in the trap region of the mass spectrometer. The same pattern for AqpZ was observed as for *Pf*MATE, where neutral dissociation of lipids dominated removal of charged lipids. Surprisingly, when an energy of 150 V was applied the peak corresponding to the unbound form with charge 17+

was completely depleted and only the unbound form of charge stripped lipids with charge 16+ was observed. Since AqpZ is a tetramer this observation can be explained by dissociation of the unbound-tetramer, with charge 17+, to monomer and trimer. As expected an increased intensity of the monomer was recorded. Importantly, in the low *m/z* region peaks corresponding to both POPC and DPPC were observed. Singly charged ions with a mass of 735 Da can be assigned to DPPC whereas ions with a mass of 761 can be assigned to POPC. This demonstrates that even a complex mixture of lipids that cannot be mass resolved in the high *m/z* region can be dissociated and analysed in the low *m/z* region with high resolution, allowing for identification of all bound adducts by a highly precise mass measurement.

To fully identify each lipid by its fragmentation pattern ions were activated in the transfer in an attempt to separate the fragments based on their arrival time distribution.[65] The intensity of the fragment ions was very low since the instrument is modified for high mass measurement and detection of low mass ions is severely reduced. These low intensities made measurements in the IM dimension impractical; nevertheless, the high resolution provided by the precursor lipid ion is sufficient for direct mass identification.

In addition to combining in-source CID with collisional activation in the trap the ability to activate ions by IR irradiation was also studied (Figure 3.12). Direct IR activation of the ions from the broad ion signal also produces low-mass ions: protein monomers and lipids, making this activation combination suitable for lipid identification.

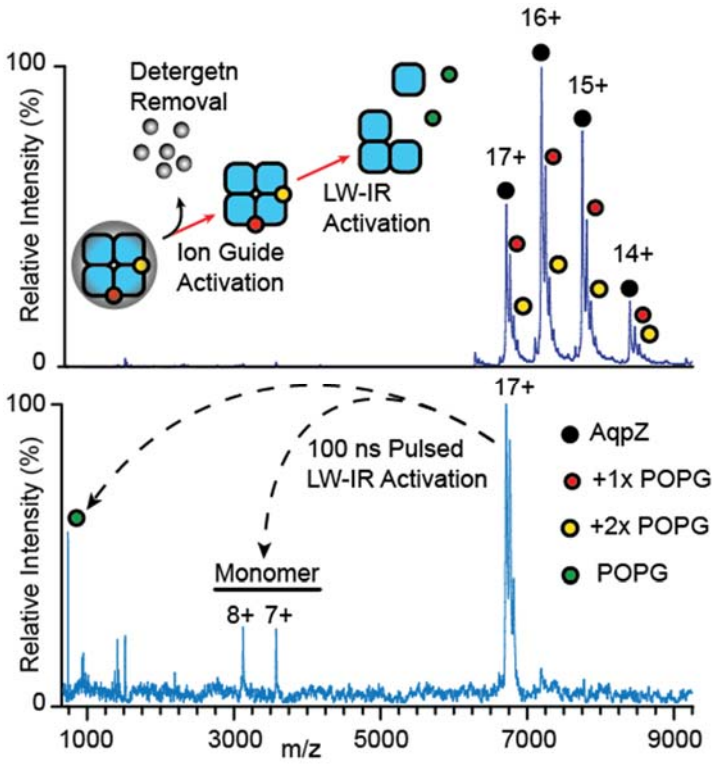


Figure 3.12: Combination of in-source CID and IRMPD in the trap for lipid dissociation and identification. Top panel shows the removal of C8E4 micelle by in-source activation and direct visualisation of bound POPG peaks. The charge 17+ of the lipid bound peaks of AqpZ was mass selected and fragmented to directly visualise POPG in the low *m/z* region, bottom panel. The protein at 12μM was incubated 20μM of POPG lipids prior to measurements.

As in the case of collisional activation, lipids can be identified from the peaks in the low *m/z* region, produced by singly charged lipids, with higher accuracy than from the mass peaks produced by multiply-charged complexes with lipids bound. However, some of the observed lipids could also be contained in micelles and not bound to the protein, which are then released by IR activation. To identify the lipids that are bound to the protein it is important to first isolate the protein with lipids from the gas-phase detergents, and then dissociate the detergent-free complex releasing the bound lipids for mass analysis. This approach would be particularly useful for the analysis of samples with *a priori* unknown lipid content, *vide infra*. The first step in this approach was achieved by increasing the acceleration voltage to 250 V

between the front optics of the ion guide in the mass spectrometer, thus promoting collisional activation of the complexes in micelles before the mass selection in the quadrupole. AqpZ solubilised in C8E4 was subjected to detergent removal, and its 17+ charge state containing both unbound and bound forms (bound PG) was then mass selected and further dissociated by 100 ns pulsed IRMPD, releasing the monomer subunits and the lipids from the complex. These results showed that two modes of activation can be combined to allow for tandem MS of membrane proteins and direct identification of lipids bound to the protein complex.

Having established that in source CID is capable at mass selection of a mixture of lipids and at the same time the mass resolution in the low *m/z* region is sufficiently high for lipid identification, the attention was placed on identifying co-purified lipids, revealing lipids from the cellular membrane that are potentially tightly bound to the protein. The purification protocol of AqpZ was modified, by using DDM a mild detergent for extraction, to preserve endogenous bound lipids. Further, the non-denaturing MS measurements were performed in both positive and negative polarity as the charge may affect lipid binding (this effect is elaborated on further in Chapter 4).

In both polarities lipid adduct peaks are observed, however the intensity of adduct peaks is higher in negative polarity, hence this mode was selected for further analysis. The adduct peak of the charge state 16- was selected (Figure 3.13). Although the intensity of this peak is lower than 15- selecting a higher charge state is preferred as a larger population of charged lipids are dissociated from the complex. In a low charge state lipid will leave as neutral and as a result will not be detectable. These results can be explained by a charge density threshold, above which charged lipids could dissociate from the complex. In this case activating charge state 16- lead to both an increase in the unbound form signal with charges of 16- as well as 15- corresponding to neutral loss and charged lipid loss respectively. In the

low *m/z* region distinct lipid peaks were observed centred around a mass of 731 Da. Other peaks had a characteristic mass difference of addition or removal of two CH_2 chain length units, with a mass of 28 Da. From this precise mass measurement it can be derived that this lipid belongs to the PG class with chain lengths of odd carbon unit repeats with a total number of 31, 33 and 35 CH_2.

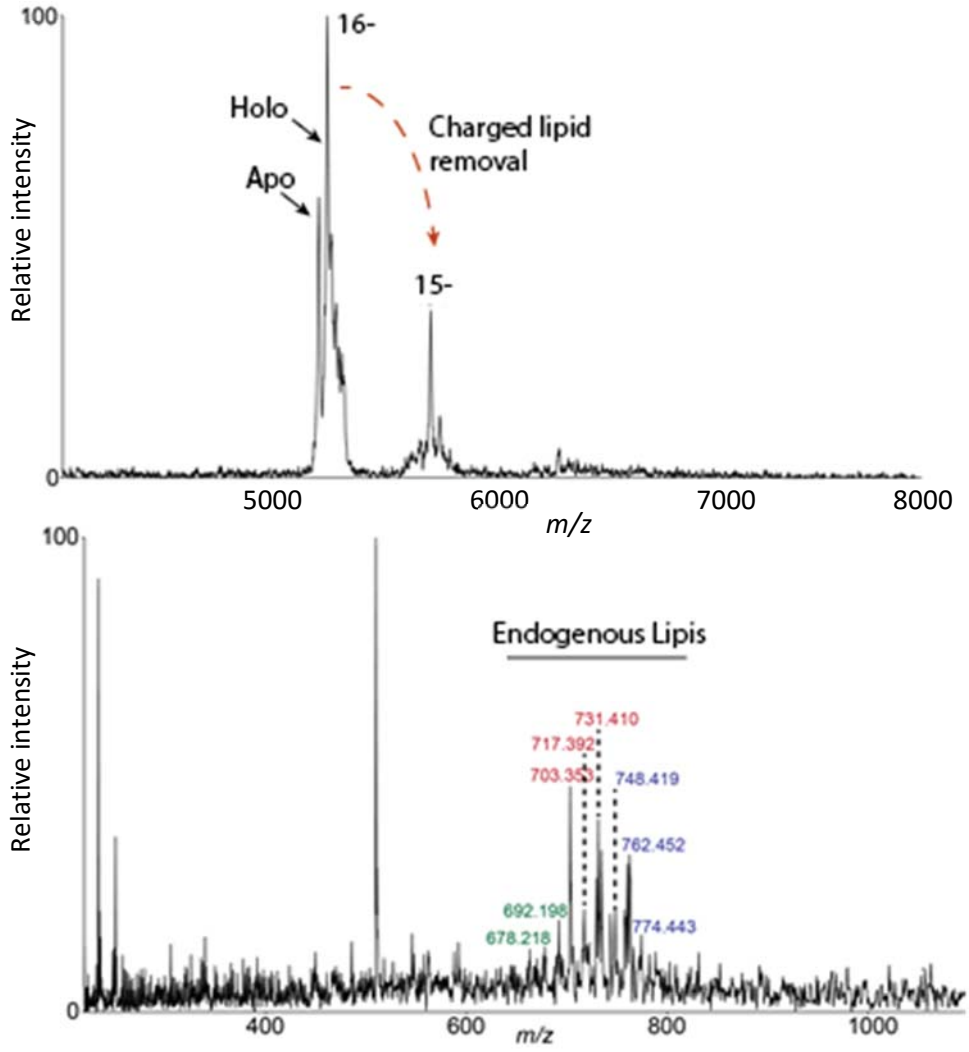


Figure 3.13: Identification of endogenously co-purified lipids by tandem MS. AqpZ was purified with bound lipids which resulted in adduct peaks in the high *m/z* region. These peaks with charge state 16- were mass selected and removed from the complex by activation in the trap (top). High resolution mass measurements allowed identification of these lipids as PG with variable chain length in the range of a total of 31–35 carbon unit repeats (bottom). The protein was extracted with 10xCMC OG and sprayed in 2xCMC C8E4 in negative polarity.

In the case of co-purified lipids usually a mixture of lipids is present which introduces an additional level of complexity. To show that the present method can be applied for direct lipid identification on other membrane protein classes neurotensin receptor 1 (NTSR1), a G-protein-coupled-receptor was chosen. This class of proteins are notoriously difficult to overexpress and purify due to their low stability in the absence of the membrane.[183] Hence purification protocols that preserve the annular lipid belt by using mild detergents are preferred. In addition, lipids may play a role in their function hence making direct lipid identification necessary for full characterisation of their chemistry.

NTSR1 was chimerically expressed from bacterial cell lines reducing the level of complexity due to the absence of post-translational modifications. Although the lipid composition of bacterial membranes is significantly different from that present in mammalian cell lines, the co-purified lipids in bacteria may resemble the same chemistry as mammalian lipids, or at least they provide evidence if lipids are structurally significant. NTSR1 sprayed as both monomer and dimer with additional satellite peaks that were subject to MS/MS analysis (Figure 3.14). The intensity of adduct peaks is relatively low hence the peak with the highest intensity, corresponding to charge state of 10+, was selected. Prior to selection the detergent was removed in the source region and the mass selected ions were further activated, resulting in fragments of the unbound form with charge states of 10+, 9+ and 8+, meaning that lipids dissociated as charge species.

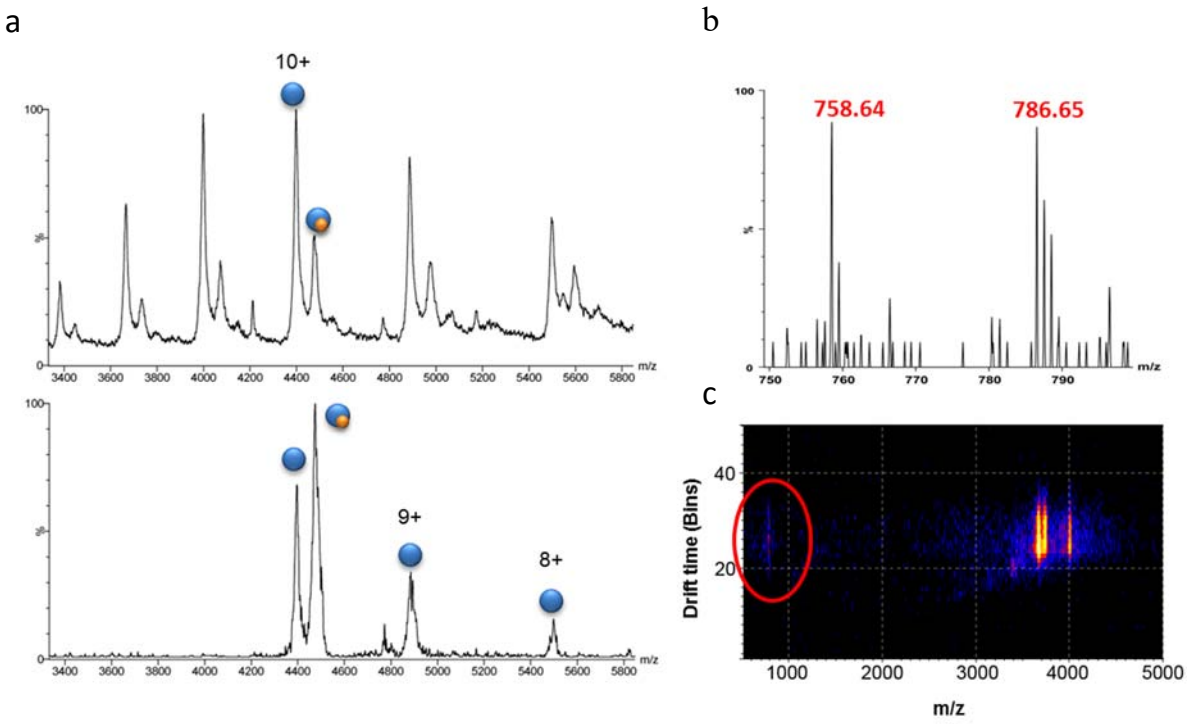


Figure 3.14: Identification of endogenous lipids bound to NTSR1. (a) Mass spectra recorded in foscholine-14 shows the presence of the apo protein (blue) and adducts peaks (orange) centred at 10+ cahrge state (top). The adduct peak of charge state 10+ was selected for MS/MS characterisation (bottom). Applying a high voltage of 130V resulted in removal of charged lipids and resulting apo protein with charge 9+ and 8+. (b) In the low *m/z* region two distinct lipid peaks were observed allowing identification of the lipid as PS. c) IM measurements provide evidence that this lipid (circled in red) is bound to the monomer species.

In the low *m/z* region two peaks with mass of 758 Da and 786 Da were observed. The difference between these two peaks is characteristic of two carbon chain length units hinting that these species are lipids. By measuring the precise mass the lipid was identified to be PS with a chain length of 24 and 36 carbon units for the high and low mass respectively. Since several oligomeric states are present and the 10+ charge can correspond to both monomer and dimer, dissociation in the transfer region was used to identify if the dissociated lipids were coming directly from the monomer or the dimer (Figure 3.14). When operating with the transfer region at high voltage dissociated ions will have the same arrival time as parent ions, since the IMS is performed prior to dissociation. The IM data shows that the arrival time

distribution of the lipids is the same as the arrival time distribution of monomeric units providing evidence that these lipids are fragmented from the monomer.

Taken together these results show that modification of the instrument in the source region is sufficient for micelle removal allowing for tandem MS of membrane proteins which can be used for direct lipid identification of endogenous co-purified lipids.

3.5 Conclusion

The ability to identify bound ligands has played a substantial role in further establishing MS as a useful structural biology tool. Although typically bound species are directly observed by non-denaturing MS, the resolution in the high *m/z* region is not sufficient for direct identification or separation of a bound mixture. Hence, protein complex dissociation in the gas phase by tandem MS has been developed as an alternative to directly mass identify ligands in the low *m/z* region.

The presence of detergents in membrane protein samples provides an addition layer of complexity making the conventional instrument architecture unsuitable for MS/MS analysis. In this chapter, three instrument modifications were used to overcome this problem. High energy in-source CID was found to liberate the proteins from detergents prior to quadrupole selection. Using both activation in the source and trap regions bound lipids could be dissociated and identified. The method is capable of identifying a mixture of lipids that could not otherwise be mass separated in the high *m/z* region. Application of low pressure in the source region was found to also be sufficient for detergent removal, however the sensitivity for large ions was severely diminished. Finally IRMPD in both CW and pulsed modality was shown to strip the detergents from membrane proteins in the collision cell and in combination with in-source CID was able to mass identify bound lipids that would dissociate from a given

protein. In summary, results presented in this chapter provide different approaches to activate proteins embedded in detergents, which allows access to the quadrupole for mass selection of bound species. These developments permit the ability to directly identify bound lipid species.

Chapter Four: Binding of lipids to membrane proteins is influenced by chargeable residues

Integral membrane proteins reside in a lipid bilayer environment; hence the interactions of membrane proteins with the surrounding lipids are important for their stability and conformation,[184] but also function.[185] Based on the affinity and specificity of these interactions the surrounding lipids are classified as bulk, annular and non-annular lipids.[185] Bulk lipids have no preference in binding, they simply solvate the membrane protein whereas annular lipids bind selectively around the protein without a specific binding site. Non-annular lipids on the other hand bind selectively to a specific lipid binding site on the protein, therefore their interaction is usually stronger compared to either bulk or annular lipids.

Several studies have demonstrated the importance of lipid binding to the conformation and function of membrane proteins[186-193], especially for non-annular lipids, however a comprehensive understanding of membrane protein-lipid interactions and how lipids mediate membrane protein behaviour is lacking due to the challenges of studying these interactions. Typically lipid-protein interactions are characterised by hydrophobic interactions, usually between non-polar residues and the lipid tail, and ionic interactions between chargeable residues and the lipid head group. In this chapter the focus is placed predominantly on the ionic interactions of annular lipids with residues in the lipid head group binding region of membrane proteins.

During a MS experiment both positively and negatively charged ions can be generated enabling the ability to selectively analyse the non-covalent binding of anionic, cationic or zwitterionic (neutral) lipids to membrane proteins. Previous studies have shown that under physiological conditions anionic lipids bind preferentially to basic residues[194-196]; here these

interactions are placed in the context of exogenous stimulus, such as a change in pH which can alter the charge of residues on a protein. These studies showed that, at low pH, neutralisation of acidic residues facilitates increased binding to basic residues.

4.1 Interaction of chargeable residues with charged lipids

Electrostatic interactions play an import role in biochemistry.[197] Almost all proteins have chargeable residues that are located across all regions of the protein; active sites, surface exposed areas or buried sites. The position of chargeable residues in active sites is typically associated with electrostatics-mediated ligand binding and functional regulation, whereas in non-active sites charged residues usually mediate protein-protein interactions, protein folding, and stability.[198] Other than having a direct role in function and stability of proteins chargeable residues are also important for practical reasons. They are utilized in many analytical tools due to their direct influence in solubility and aggregation. Proteins with a large number of the same charges tend to repel each other, for example hindering crystallisation efforts; on the other hand proteins with a large number of complementary charges attract each other, which can lead to aggregation or fibril formation.

Understanding electrostatic interactions has been the focus of several studies.[199] Simple small molecule systems in solution are well understood, however a full characterisation of the multiple charged interactions typically found in proteins remains elusive. These interactions are characterized by the distance between the two interacting charged particles spanning from local interactions or salt bridges, to long range interactions that in many cases extend beyond the boundaries of the protein.[197] Local interactions are typically restricted to a particular ion pair as the distance between the two ions cannot accommodate solvent molecules.

In solution, the charge on a protein is correlated with the number of amino acids in the protein that have charged side chains and the protonation state of the termini. There are five amino acids with charged side chains at physiological pH, arginine, histidine and lysine have positively charge side chains, and are denoted as basic residues, whereas aspartic acid and glutamic acid have negatively charged side chains and are referred to as acidic residues. Cysteine and tyrosine can also be charged due to protonation or deprotonation of their thiol and tyrosyl groups, respectively. The net charge of a protein is not only dependent on the number of chargeable residues in the amino acid sequence but is also influenced by pH, ionic strength and the dielectric constant of the medium.

At a given pH the protonation state of each chargeable residue is dependent on the pK_a of that residue in its surrounding environment (Table 4.1). For example, in water the average pK_a of aspartic acid is 4.0. At pH >4 the charge of this residue is -1, at pH 4 there is an equal population of neutral and negatively charged species, whereas at pH <4 the neutral form predominates. Most proteins have more than one chargeable residue present and the pK_a for each chargeable residue is altered due to the cooperativity of acid-base equilibria. For instance in T4 lysosome the pK_a of histidine at position 31 has been measured to be 9.1 due to a salt bridge formation with aspartic acid at positon 70. The measured pK_a values can also change due to variation in the dielectric constant of the environment. The same residues will have different pK_a values depending on their location in the protein and their degree of solvent accessibility. An example is the change in pK_a of aspartic acid at position 76 of RNAse T1 that due to its position in the buried region of the protein reaches as low as 0.5. Generally, the same residue may have different affinities for protonation depending on its position and interaction with other chargeable residues as well as the electrical properties of the environment.

Table 4.1: Average individual pK_a values of chargeable amino acids and lipids in water and the net charge state in positive and negative polarity. Data were taken from http://academics.keene.edu /rblatchly/Chem220/hand/npaa/aawpka.htm and Avantilipids.

Residue/Lipid	Average pK_a	Net charge in positive polarity	Net charge in negative polarity
Arginine	12.48	Positive	Neutral
Aspartic acid	3.90	Neutral	Negative
Cysteine	8.37	Neutral	Negative
Glutamic acid	4.07	Neutral	Negative
Histidine	6.04	Positive	Neutral
Lysine	10.54	Positive	Neutral
Tyrosine	10.46	Neutral	Negative
N-terminus	7.8	Positive	Neutral
C-terminus	3.6	Neutral	Negative
POPC	1.0	Positive	Neutral
POPG	3.5	Neutral	Positive
POPS	2.6/5.5/11.5	Positive	Negative

In the research described here another layer of complication is introduced as measurements in vacuum are compared with solution. In vacuum, charged particles can directly interact with each other depending on the polarity and separating distance. In the absence of solvent these interactions are typically enthalpically driven, however the opposite is observed in the presence of a polar solvent such as water. The free energy of interacting charged particles that are separated by distance that can freely accommodate several water molecules are usually dominated by the entropic term due to orientation of the water molecules around the charged residues.[197]

From a physical chemistry perspective, in this chapter the contribution of the lipid environment to the entropic term of binding as it solvates a membrane protein is addressed,

and this is correlated to the strength of the ionic interaction. This is analogous to the same effect water molecules have on soluble proteins with an additional complication that lipids can also be charged.

Similar to charged residues in proteins the net charge of lipids is also dictated by their apparent pK_a (Table 4.1). For the research described here POPC and POPG were used because their charges are complementary to each other (i.e. POPC can carry either zero or positive charge, and POPG either zero or negative charge). Figure 4.1 shows the net charge of these lipids as a function of pH. At low pH POPC is charged due to protonation of the phosphate group, whereas POPG is neutral, as pH increases POPC becomes neutral and POPG becomes negatively charged.

This charge alternation as a function of pH can then be paired with the charge of the protein to interrogate ionic interactions between lipids and proteins. Several studies have demonstrated the importance of ionic interactions for the binding of lipids to proteins[196,200] , concluding that lipids may, generally, play three major roles in regulating cellular functions: They can serve as pH sensors, provide the medium for protein clustering and stability, and favour different conformations of proteins.[194,200] For example, phosphatidic acid (PA) has a pK_a of 8.0 which allows this lipid to serve as a pH sensor in physiological pH ranges. Young and co-workers found that PA can regulate the expression of genes related with lipid metabolism and synthesis.[201] At physiological pH PA has a net negative charge which strengthens the binding and sequestering of the transcriptional repressor Opi1. At a pH <7 PA is protonated and its affinity to the positive amine groups of Opi1 is reduced, allowing Opi1 to repress the transcription of genes related to lipid synthesis. The finding that protonation state of PA due to cytoplasmic acidification suppresses gene expression was then correlated to the observation that in low glucose the medium becomes acidic, thus there being a direct

link between membrane biogenesis and cellular metabolism. It is remarkable that the key to this entire mechanism is the ionic interaction of PA lipids with charged residues of Opi1.

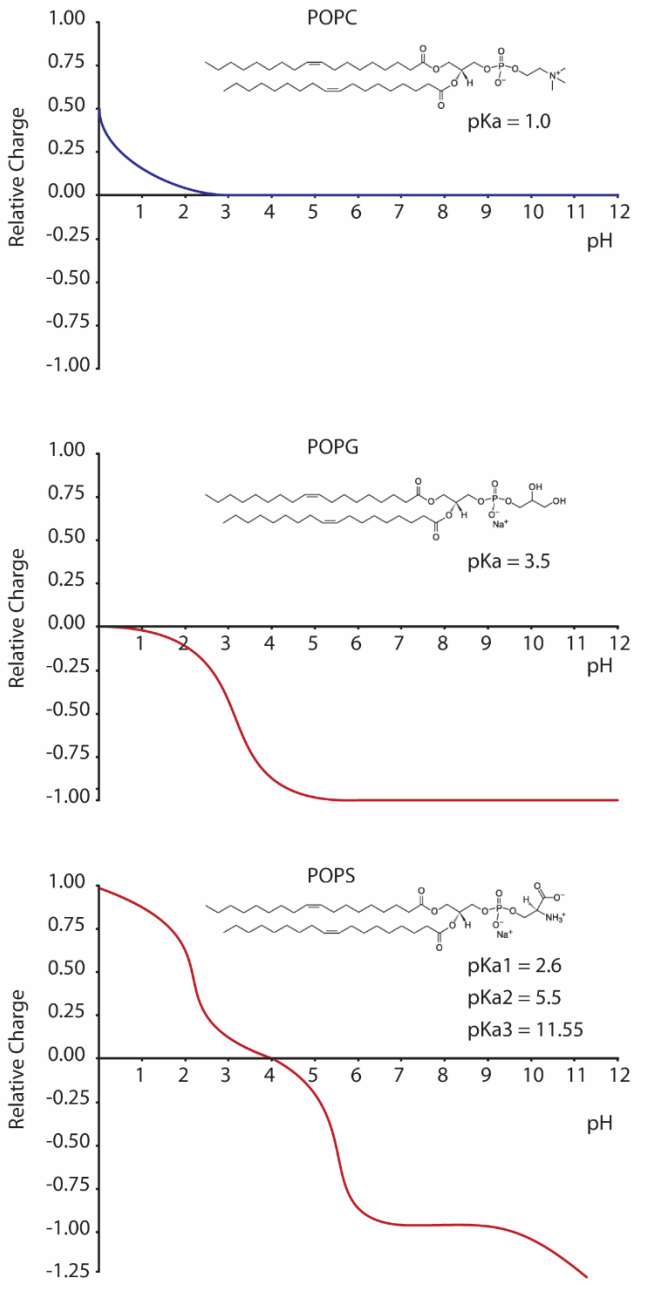


Figure 4.1: The relative charge of POPC, POPG and POPS as a function of pH at ionic strength of 0.1 (HCl). The data were taken from the CRC Handbook of lipid bilayers.[202]

The second major role that charged lipids can play is the ability to facilitate protein clustering and surface stability.[203] Acidic phospholipids will have a net negative charge, as

such they can assist proteins with positively charged residues to associate with the membrane. They can also facilitate the clustering of proteins that carry an abundance of residues with the same charge. N-ethylmaleimide-sensitive factor attachment protein receptor (SNARE) complex has a region composed of five basic residues (^{260}KARRKK^{265}) that are positioned in close proximity to the membrane.[204] Due to this large accumulation of positive charge SNARE complexes tend to repel each other. However, in the presence of acidic phospholipids this high positive charge is neutralised by negative charges provided by the lipid head group, allowing clustering. In addition to pH sensing and cluster formation, ionic protein-lipid interactions are also important in the regulation of protein conformation.[204] The inward-rectifier K^+ channel (Kir2) exists in two states, open and closed. In the absence of acidic lipids the closed conformation predominates.[204] However, by neutralising the charge of two basic residues in close proximity to the channel, the presence of PI lipids mean the open conformation becomes favoured. This allosteric regulation by lipids is crucial to the function of the protein, as is also indicated by the fact that both basic residues are well conserved among different species.

The number of cases where protein function and structure has been found to be regulated by lipid interactions has rapidly increased;[194] nevertheless the availability of analytical tools to study these interactions is limited. Here, in this chapter, a combination of non-denaturing MS, molecular dynamics and electrophysiology is used to reveal the allosteric effects of lipid binding to membrane proteins under different pH regimes.

4.2 Investigating the density of chargeable residues in the lipid head group binding region in membrane proteins

In the case of the membrane proteins chargeable residues are positioned predominately in the solvent-exposed regions whereas the transmembrane region is typically composed of non-polar residues. The position of lipids with respect to the membrane proteins is typically constrained to the interface between soluble and membrane regions. As a consequence the chargeable residues in this interface area are in contact with the lipid head groups. From the perspective of the lipids, this lipid head group binding region includes any residue within 3 Å of the lipid head groups.

For proteins in which lipid-protein interactions are not dominated by ionic forces the percentage of chargeable residues in the lipid head group binding region may be relatively small, since having charged residues in close proximity to the membrane region may carry a high enthalpic cost. On the other hand, if lipid-protein binding is dominated by electrostatic interactions a high number of chargeable residues in the lipid head group binding region may be expected to enhance the strength of these interactions. To identify membrane proteins with a high number of chargeable residues in the lipid head group binding region a calculation of the percentage of chargeable residues in the lipid head group binding region was performed for all membrane proteins in the PDBTM.

To perform this calculation it is first necessary to define the region of each protein which will bind the lipid head groups. In crystal structures of membrane proteins the membrane bilayer region is typically not well defined due to the absence of electron density for either lipids or detergents. Therefore MD simulations are generally used to position membrane proteins in hypothetical bilayers. They also provide the means to visualise the membrane

protein-lipid interactions. Initially membrane proteins were aligned to their expected position in a lipid bilayer. This allowed the position of the lipid head group binding region to be computed.

To calculate a metric of the chargeable residues in this region, the total surface area there the lipids are expected to bind was initially calculated, as defined by prediction where the membrane domain is located.[205] The area of this region contributed by either acidic or basic residues was then calculated, and reported as a fraction of the total area. Proteins were then ranked based on the fractional percentage of residues in the lipid head group binding region for both acidic and basic residues (Figure 4.2).

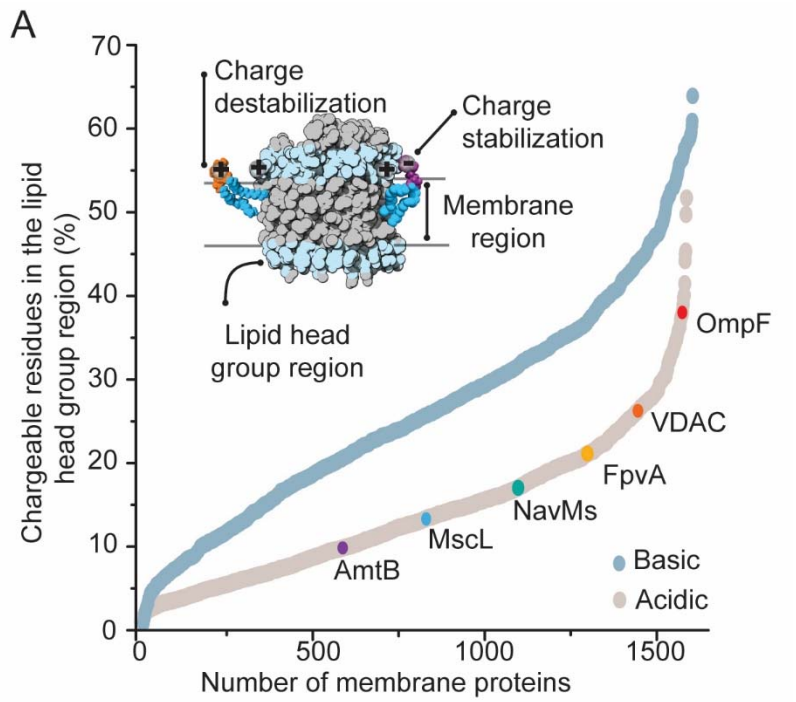


Figure 4.2: The variation of chargeable residues in the lipid head group binding regions of membrane proteins. Calculation of chargeable residues in the lipid head group binding region of all membrane proteins deposited in the Protein Data Bank with the exception of proteins with less than one single strand crossing the bilayer. On average a higher percentage area of positive residues (light blue) is observed relative to negative residues (beige). Inset depicts charge stabilisation and destabilisation by synchronisation of lipid head group region charge (blue region) and lipid charge (orange positive and magenta negative). Calculations were performed in partnership with Matteo Degiacomi

These calculations show that basic residues are more abundant in this region than acidic residues. The percentage of basic residues in the membrane proteins analysed is generally 10–15 percentage points higher, consistent with these residues playing an important role in membrane protein folding and insertion. In addition, the analysis showed that there is a large variability in the concentration of acidic or basic residue in this region. For approximately half of the membrane proteins analysed only 15 percent of the region is composed of acidic residues. The remainder of the proteins have a relatively high percentage (≥50 percent) of acidic residues in the lipid head group binding region. For proteins with a high percentage of chargeable residues in proximity to the expected positions of the lipid head groups it is

hypothesised that lipid-protein binding may be influenced by ionic interactions and modulated by the protonation state in the gas phase and the pH in solution.

The distribution of acidic residues positioned between the outer and inner membrane leaflets varies with the total percentage of these residues (Figure 4.3). For proteins with a smaller overall percentage of acidic residues in these regions the distribution is more symmetric. However, for proteins with a higher number of acidic residues, there is a greater propensity for the acidic residues to be distributed in an asymmetrical manner. These results provide evidence that whereas electrostatic interactions predominate in lipid binding they are typically localised in one of the membrane leaflets. Notably, annotated inner membrane proteins are associated with lower percentage of chargeable residues in the lipid head group binding region than for annotated outer membrane proteins.

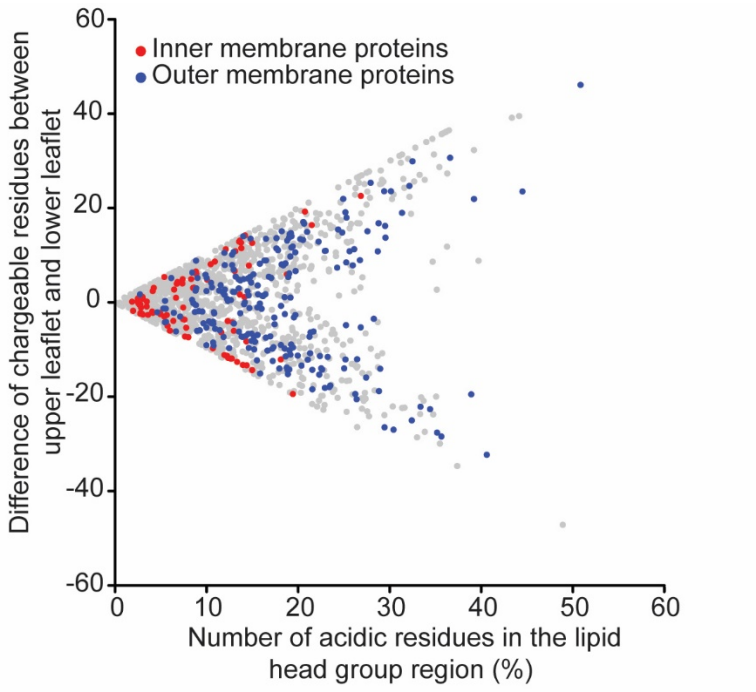


Figure 4.3: The distribution of acidic residues in the upper leaflet to the lower leaflet as a function of density in the lipid head group binding region. Inner membrane proteins (red) tend to have fewer acidic residues in this region than outer membrane proteins (blue). Un-annotated membrane proteins are coloured grey.

4.3 Preserving lipid binding by tailoring polarity

Using these calculations six membrane proteins with variability in the percentage of acidic residues in the lipid head group region were selected for further interrogation by native MS (Table 4.2). This was done to test the hypothesis that, depending on the protonation conditions of the environment, lipids may bind selectively to membrane proteins via coordination of the charge between the lipid head group binding region of the protein and the charge on the lipids.

Table 4.2: Calculation of the acidic residues in the lipid head group binding region of selected membrane proteins.

Protein	SASA	Percentage area of acidic residues in the head group region		
		Upper leaflet	Lower leaflet	Total
AmtB	32555	7.71	9.26	16.97
MscL	31169	12.17	10.32	22.49
NavMS	16779	10.78	17.13	27.91
VDAC	16713	13.37	24.11	37.48
OmpF	41039	12.90	33.35	46.25

The membrane proteins AmtB, MscL and VDAC were expressed and purified whereas OmpF and FpvA were kindly donated from the Kleanthous group (Oxford) and NavMS was a donation of Wallace group (Birkbeck).

MS measurements in both positive and negative polarities allows lipid binding to membrane proteins to be measured when both the protein and the lipid are protonated or deprotonated. As the mechanism of protonation/deprotonation during nESI is different to protonation/deprotonation in solution, as a function of pH, these MS measurements are only

used to draw analogy to what may occur in solution. A benefit of MS studies is that all species are intrinsically charged and the ionic interactions can be studied in the absence of solvent, hence in this case the predominant interactions are those of just the protein and the lipid.

Pairing of the same charge between the lipid head group binding region of the protein and the head group of a lipid can be expected to destabilise binding. To test this hypothesis the selected membrane proteins were analysed by mass spectrometry in positive and negative polarity after being incubated with POPC or POPG. To illustrate these experiments POPC and POPG binding to OmpF in both positive and negative polarity is shown in Figure 4.4.

As shown in Table 1, in positive polarity POPC is positive and OmpF has a charge state distribution centred at 19+, hence both the lipid and the protein are positively charged. The mass spectrum shows only the unbound form of OmpF present in positive mode, with non-lipid bound peaks observed. Maintaining the same lipid-to-protein ratio and switching the polarity to negative the lipid binding is preserved and up to three lipid molecules are observed to bind OmpF. In the case of measurements performed in negative polarity the protein has a net negative charge whereas the lipid is neutral. As lipid binding is observed in at least one polarity it can be assumed that in solution, prior to the electrospray process, POPC is bound to OmpF. However, during nESI in positive polarity, when repulsive ionic interactions are enhanced by protonation of both OmpF and POPC, with both species possessing a net positive charge, due to charge repulsions their binding is weakened resulting in dissociation of POPC from OmpF.

In the case of measurements performed in negative polarity POPC is neutral, hence the ionic interaction is not repulsive in nature and the lipid remains bound even at elevated

energies experienced during the course of the MS experiment. Extending this rationale it can be expected that POPG binding to OmpF will be observed in positive polarity and not in negative, as in these polarities POPG is neutral and negatively charged respectively. Indeed, this binding trend is observed in the mass spectra. Unlike POPC, the binding of POPG to OmpF can only be maintained in positive polarity.

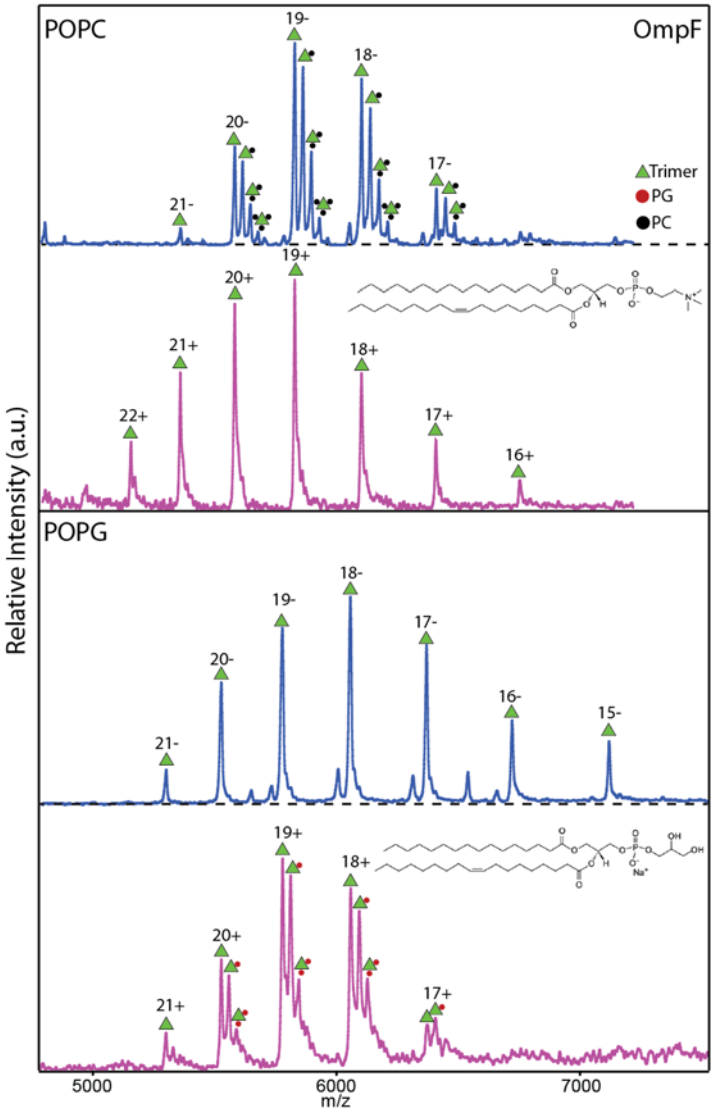


Figure 4.4: The effect of polarity on lipid binding. Mass spectra of POPC binding to OmpF showing that POPC binding to OmpF is preserved only in negative polarity. The binding of POPG shows that binding is only preserved in positive polarity. Blue spectra denote negative polarity and magenta positive polarity. The mass spectra were recorded in 200mM ammonium acetate and 2xCMC OG. Measurements were performed in a protein to lipid ratio of 1:5.

This explanation can also be tested by using a lipid that is charged in both polarities. In this case no lipid bound peaks would be expected. As shown in Figure 4.5 and Table 1, POPS has a net positive charge in protonated conditions and a net negative charge at high pH or negative polarity. Performing the MS measurements with OmpF and a lipid concentration of 14 µM and maintaining the same lipid-to-protein ratio as in the other lipid binding measurements, no binding was observed in both polarities. This is in agreement with the present hypothesis.

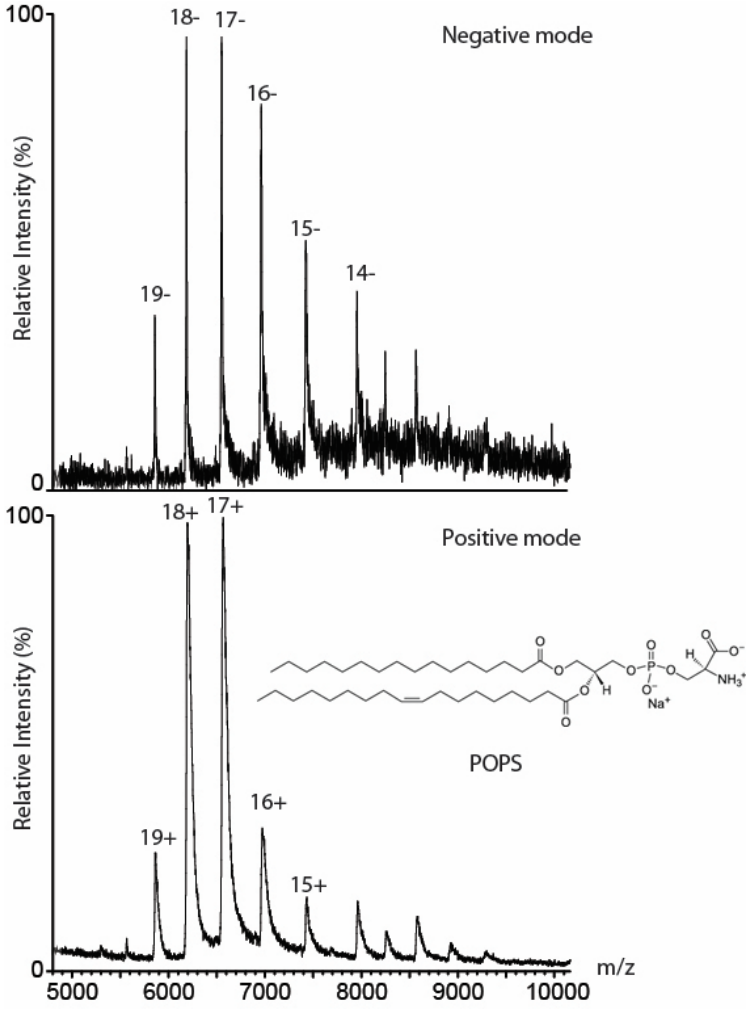


Figure 4.5: Mass spectra of OmpF incubated with POPS. No lipid binding was observed in both negative and positive polarity. In negative polarity POPS is expected to have a net negative charge, whereas in the positive polarity an overall positive charge is expected due to protonation of the phosphate and carboxylate groups.

The binding patterns presented here can be only attributed to the chemistry of the lipid head group since the lipid tail is the same for both POPG and POPC. To explore whether the lipid tail has any role in stabilising the binding two additional lipids DSPC and DPPC were analysed. These lipids differ in chain length, DSPC has a C_{18} chain length whereas DPPC has a C_{16} that is similar to POPC however without any degree of unsaturation. Depending on the transmembrane domain thickness it may be expected that longer chain lipids could bind more strongly to the membrane proteins since they can establish a larger number of non-covalent binding interactions with the membrane protein compared to shorter chain lipids.

The enhancement in stabilisation of lipid binding due to an increase in the lipid chain length was observed by MS measurements (Figure 4.6). While the general binding trends between the two polarities were similar to POPC binding, in the case of the longer chain lipid DSPC, a small population of lipid bound species was also observed in the positive mode. The presence of the lipid bound species in the positive mode can be attributed to an increase in binding affinity provided by the longer chain lipid. However, a comparison between lipid binding in positive and negative polarities, even in the case of DSPC, clearly shows that ionic interactions have a dominant effect in binding preservation. Analysing the same sample, up to four DSPC lipid molecules were observed bound in negative polarity compared to the low intensity of a single lipid bound in positive polarity. These results agree with previous measurements performed in solution showing that longer chain lipids that match the membrane domain have greater binding affinities than shorter chain lipids that may not cover the entire membrane. [206]

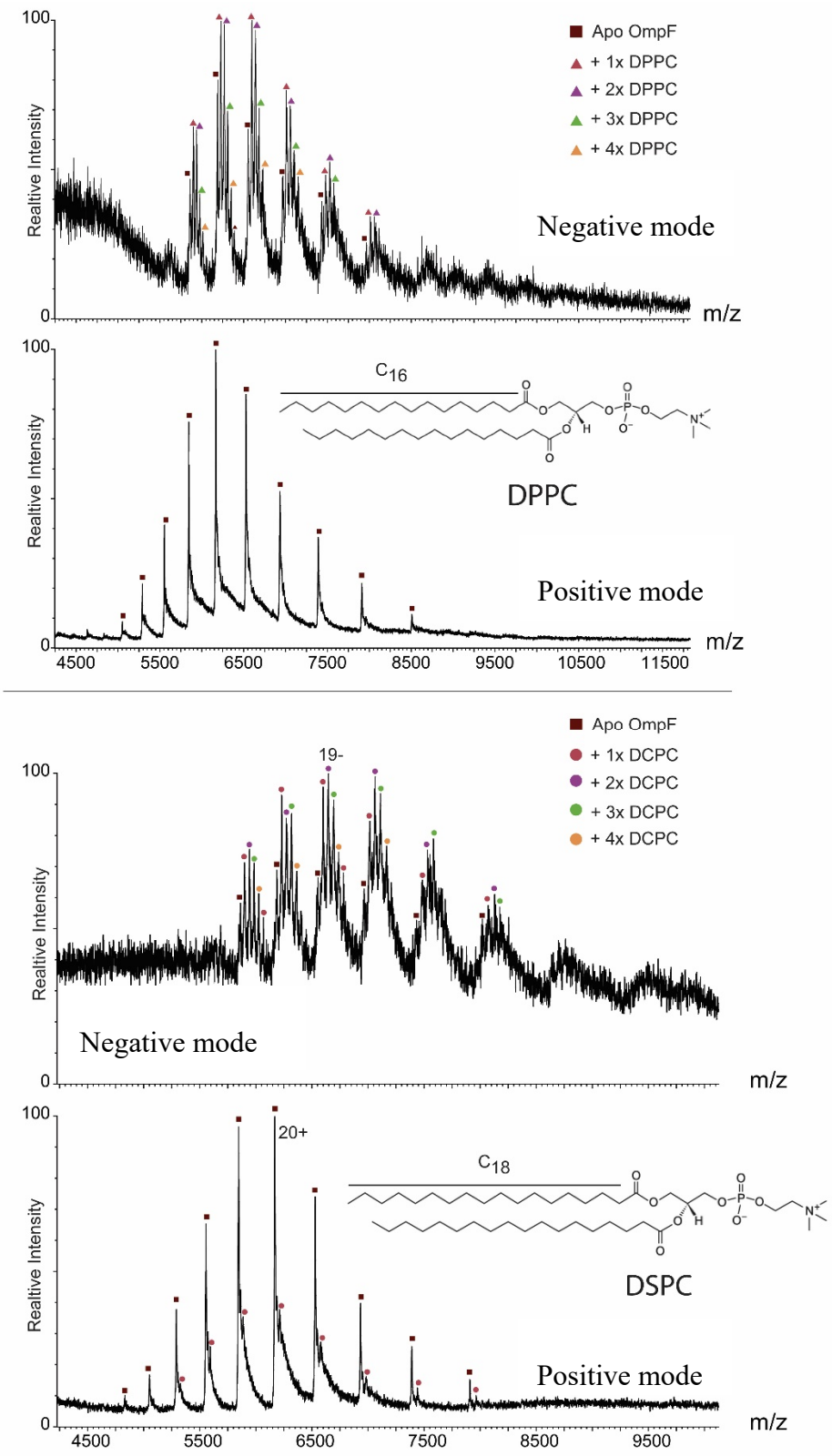


Figure 4.6: The dependence on the lipid chain length of lipid binding to OmpF in protonated/deprotonated conditions. The binding of DPPC (triangle) and DSPC (square) having a short and long chain, respectively, show that longer chain lipids stabilise lipid binding.

Having established that altering polarity to neutralise the charge on the lipids can lead to preservation of lipid binding to membrane proteins the next step was to investigate this behaviour as a function of the concentration of acidic residues. OmpF has a relatively high percentage area of acidic residues in the lipid head group binding region (33 percent in the outer membrane leaflet), and as a result a high charge density is likely to be present in this region. This will significantly affect the ionic interactions when binding to charged lipids. However, the majority of membrane proteins were found to have a much lower area of acidic residues in the region where lipid head groups bind. For instance, AmtB has only 9 percent of this surface area contributed by acidic residues in its lower membrane leaflet, and therefore the ionic interactions are likely to be less predominant. MS measurements of AmtB bound to POPG manifested a decrease in the effect of charge repulsion leading to the observation of lipid bound species regardless of polarity (Figure 4.7). In positive mode AmtB is positively charged and POPG is neutral, hence as in the case of OmpF, lipid bound peaks are observed. However, in negative polarity were both the AmtB and POPG are negatively charged the lipid-protein complex does not completely dissociate and remains in its bound state.

To investigate if this behaviour is protein dependent and can be attributed to other features of the protein other than the acidic residues the same measurements were performed with MscL, NavMS, FpvA and VDAC. MscL has a low percentage of acidic residues in the lipid head group binding region and the lipid affinity was not significantly affected by polarity or the charge of the lipid, similar to AmtB. For FpvA and VDAC only neutral lipids are preserved bound to charged proteins. Both these proteins have a high percentage of acidic residues, which is in agreement with the rationalisations made for OmpF.

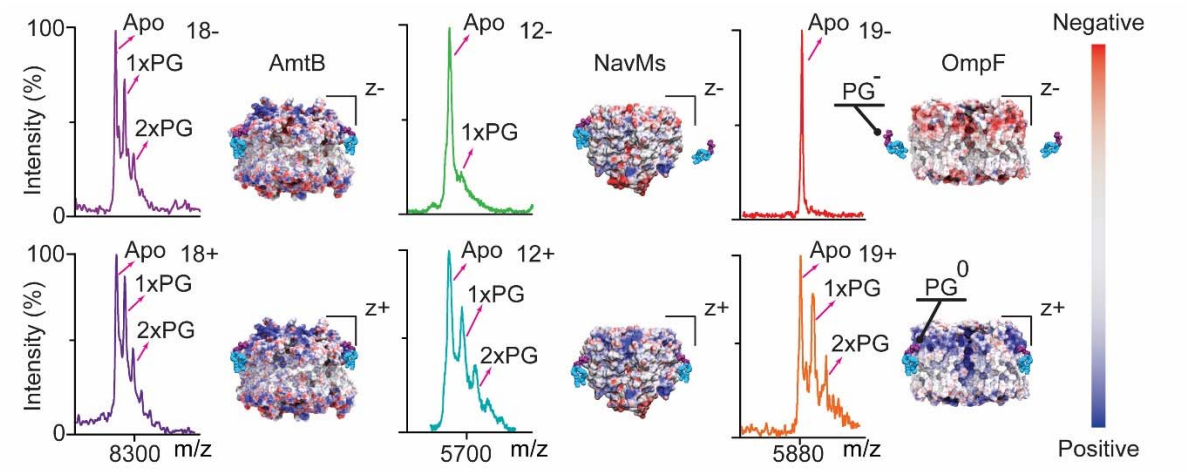


Figure 4.7: Lipid binding to membrane proteins correlates to protein electrostatic surface potential. POPG binding to AmtB (purple spectra, top negative polarity, bottom positive polarity) show that lipid binding is not affected by polarity. POPG binding to NavMS (green spectra) is preserved more in positive polarity. Binding of POPG to OmpF (orange spectra) is not preserved in negative polarity, but is observed in positive polarity. Inserted are the electrostatic surface potentials of the three proteins in protonated and deprotonated conditions (blue denotes positive charge, and red negative charge). An appreciable change in electrostatic potential of OmpF is observed in the different environments and insignificant differences are observed for AmtB.

An interesting case was that of lipid binding as a function of polarity for NavMS. This protein has a moderate percentage of acidic residues in the lipid head group binding region and a relatively high percentage of basic residues. Hence, in positive polarity the protein will likely establish a large positive potential in the vicinity of lipid binding whereas in negative polarity the charge accumulated by acidic residues will be less pronounced. As a result, positively charged lipids in positive polarity will experience a greater repulsion than negatively charged lipids in negative polarity. MS measurements showed that POPC binds only in negative polarity, whereas the binding of POPG was preserved to a greater extent in positive polarity compared to negative polarity (Figure 4.7). These results show charge synchronisation between the protein and lipid facilitate preservation of bound lipid species in the gas phase.

To further rationalise the effect of charge on membrane protein lipid binding, surface electrostatic calculations in deprotonated and protonated conditions were performed for the three representative cases, AmtB, NavMS and OmpF (Figure 4.7, insets). The protonated and deprotonated states resemble the positive and negative polarity conditions, respectively. Between these states, little difference is observed for AmtB, a moderate difference is observed for NavMS, and an appreciable difference is apparent for OmpF. Interestingly, the relative electrostatic differences correlate with the observations by mass spectrometry of lipid binding to these proteins. For AmtB, for which POPG and POPC bound in both polarities, there is no significant electrostatic difference between the deprotonated, neutral and protonated states. In the case of NavMs there is a moderate change in the surface electrostatic state, and the binding of the lipids (in both polarities) is slightly destabilised. For OmpF a large change in electrostatic state is observed between protonated and deprotonated environments. This correlates with the near complete destabilization of lipid binding for POPG in negative polarity and POPC in positive polarity.

Taken together, it can be concluded that the only difference between POPG and POPC lipids is the propensity to have a different charge in specific protonation states. As such, it seems reasonable that this may be a determinant of the lipid binding behaviours observed. By extension, the region of a membrane protein proximal to the lipid head group may also contribute to the preservation, or lack thereof, of lipid binding. It is believed that there is no direct correlation between charging in solution and the resulting charge state distribution by electrospray ionisation. However, in the case of membrane proteins encapsulated in a detergent micelle it is hypothesised that in the early stages of ionisation the soluble domains or more surface-exposed regions are more prone to protonation.[141] Therefore, a large portion of the charge on the protein as a whole is likely to be in the proximity of the lipid head group

binding region at this stage. This temporal charge localisation may therefore drive the preservation or lack thereof of lipid binding.

4.4 OmpF voltage-gated stability enhanced at low pH in presence of POPG

Having arrived at some conclusions as to how the observed MS phenomena may be occurring, it was interesting to then investigate if these same factors may affect lipid binding in solution. To probe if the results from both MD simulations and MS correlate with solution patch clamp experiments were performed. These experiments allow single channel measurements to be performed with a membrane protein embedded in a lipid bilayer, significantly improving the similarity of the *in vitro* experimental conditions to those found *in vivo*.

OmpF showed the most dramatic differences in lipid binding by mass spectrometry, and also the most significant differences in electrostatic surface. To assess whether the polarity dependent lipid binding effects observed in MS experiments is also translated to solution, electrophysiology experiments on OmpF in planar lipid bilayers of different lipid compositions, in different pH environments, were performed. Patch clamp experiments study the behaviour of ion channels under an applied voltage. The channel is initially inserted in a lipid membrane bilayer that has the same buffer composition on both sides. A patching pipette is positioned in contact with the membrane, as shown in Figure 4.8. Usually a vacuum is applied to seal the contact between the pipette and the membrane. In the electrode a voltage is kept constant and the ion current is monitored as a function of time.

Depending on the direction of the potential, ion movement across the channel is indicative of an open state. By the same token, no ion movement across the channel is

observed in closed states. When a protein is composed of multiple channels, as in the case of OmpF, the behaviour of each channel can be monitored separately. The ability to control the membrane or buffer composition makes patch clamp one of the most useful techniques to monitor the effect of lipid binding to membrane proteins in different environments.

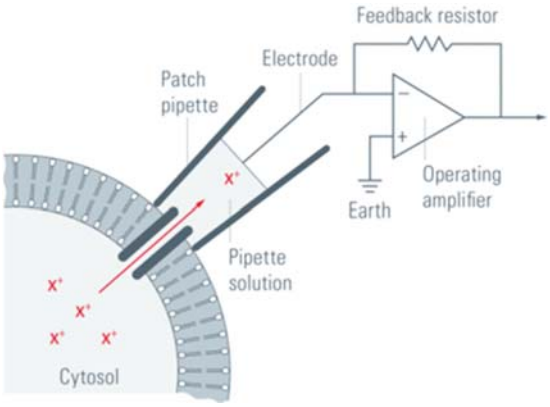


Figure 4.8: A diagram depicting electrophysiology measurements on a channel. The channel is initially constituted on a lipid bilayer and placed in direct contact to a buffer solution under a voltage via a patch palette. The generated current from ion passing through the channel is amplified and recoded as a function of time. Courtesy of Leica-microsystems

OmpF was inserted into a planar lipid bilayer composed of DPhPC and the channel states were measured by applying a constant voltage in both polarities. Initially, the behaviour of the OmpF channel in a high applied voltage was investigated. As evident from Figure 4.9, at pH 8/7.4 OmpF can resist an applied voltage of +200 mV for approximately 30 s followed by a full closure of all three channels. OmpF, which is trimeric, therefore has four possible states corresponding to the open or closed state of each of the constituent monomers: L0 (all channels open), L1 (one closed), L2 (two closed) and L3 (all channels closed). The same measurements were also performed at a low pH. A destabilising effect leading to faster channel closure was observed at low pH. Measurements performed at pH 4 showed that all channels fully close within a shorter period of time, consistent with previously published

results.[207,208] The effect of channel closure was then used to measure the stability of the protein under protonated conditions.

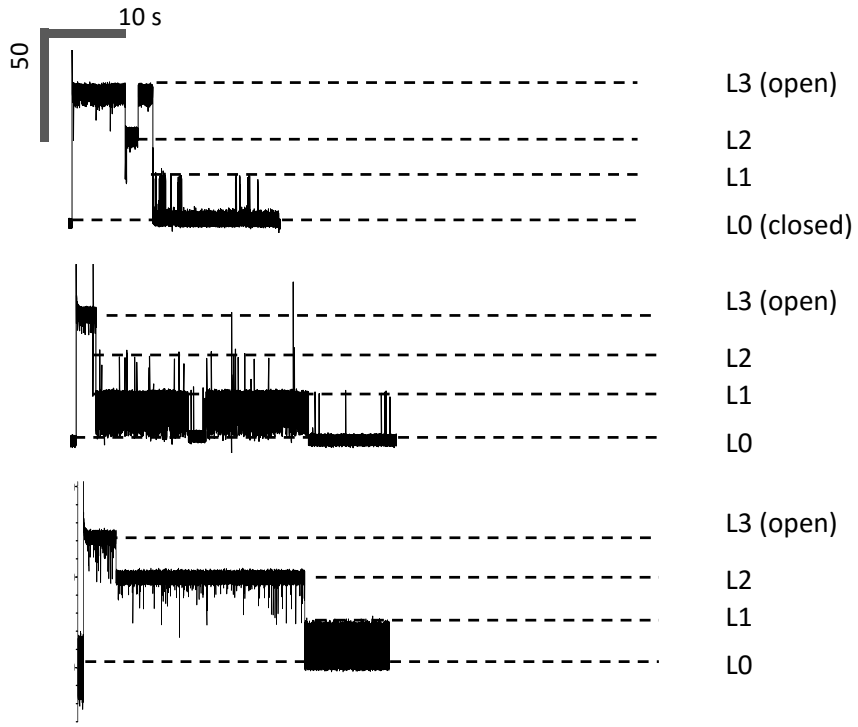


Figure 4.9: The dependence of OmpF channel stability as a function of pH. Measurements were recorded on planar lipid bilayer composed on DPhPC lipids in 20 mM acetate and 0.1 M KCl. The channel closes faster at pH 4 (top) then at pH 6 (middle) and it remains open for longer at pH 8 (bottom). Electrophysiology experiments were performed by Sejeong Lee.

From MS experiments it was found that POPG binding to OmpF is enhanced relative to POPC in protonated conditions. This suggests that upon POPG binding the open state of the channel may be stabilised even at low pH. Maintaining the same pore, and adding POPG, in order to end up with a final ratio of 5:1 DPhPC:POPG, it was found that the channel maintained an open state for a longer time compared to measurements performed in POPC-only lipid bilayers (Figure 4.10). In the absence of POPG in the bilayer at low pH under an applied voltage of 100 mV the channel closes within 20 ± 5 s whereas the same pore embedded in a DPhPC:POPG bilayer can maintain the open state for 60 ± 17 s. This effect

was not observed when the same experiments were performed at physiological pH (Figure 4.11), suggesting that the acidic residues may be affecting lipid binding, and thus channel stability of OmpF at low pH.

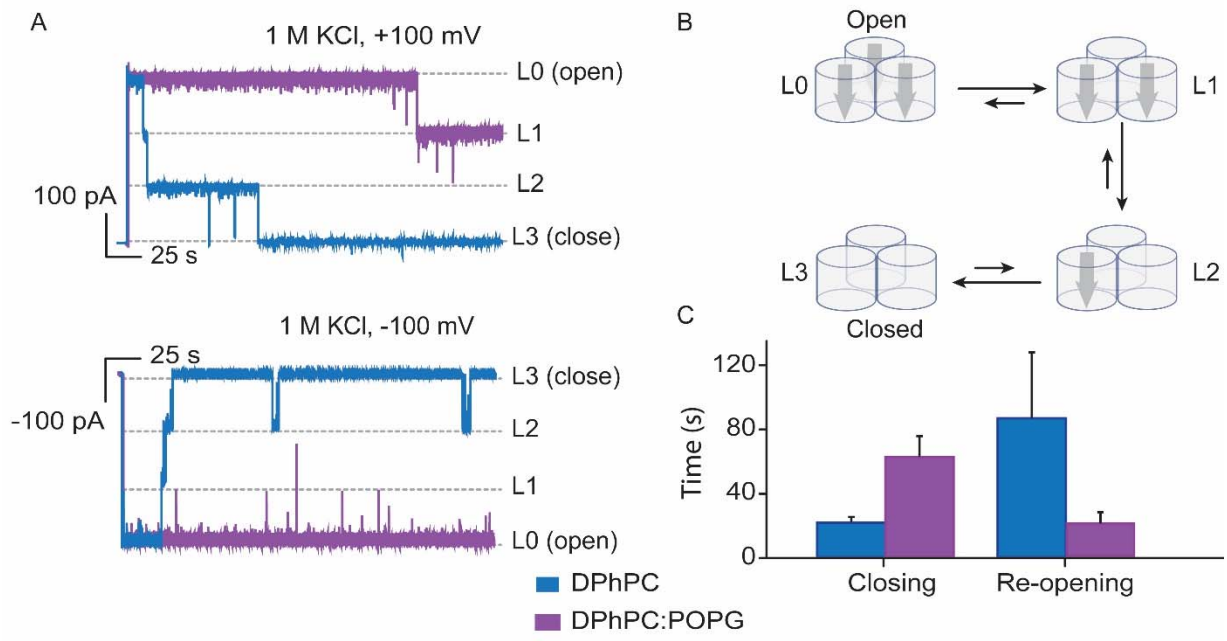


Figure 4.10: The open conformation of OmpF is stabilized by the presence of POPG. (a) Electrophysiology measurements of OmpF in 1 M KCl in protonated conditions (pH 4) in DPhPC (blue recording) and a mixture of 5:1 DPhPC:POPG (purple recording) in both positive and negative voltage regimes top and bottom channel recordings, respectively. (b) A diagram depicting all four possible states of the OmpF. Since the protein is a trimer each pore can close separately. (c) Histograms were constructed from an average of six independent measurements showing that on average the open state is stabilized by the presence of POPG. Error bars show 2× s.e.m. of six independent measurements.

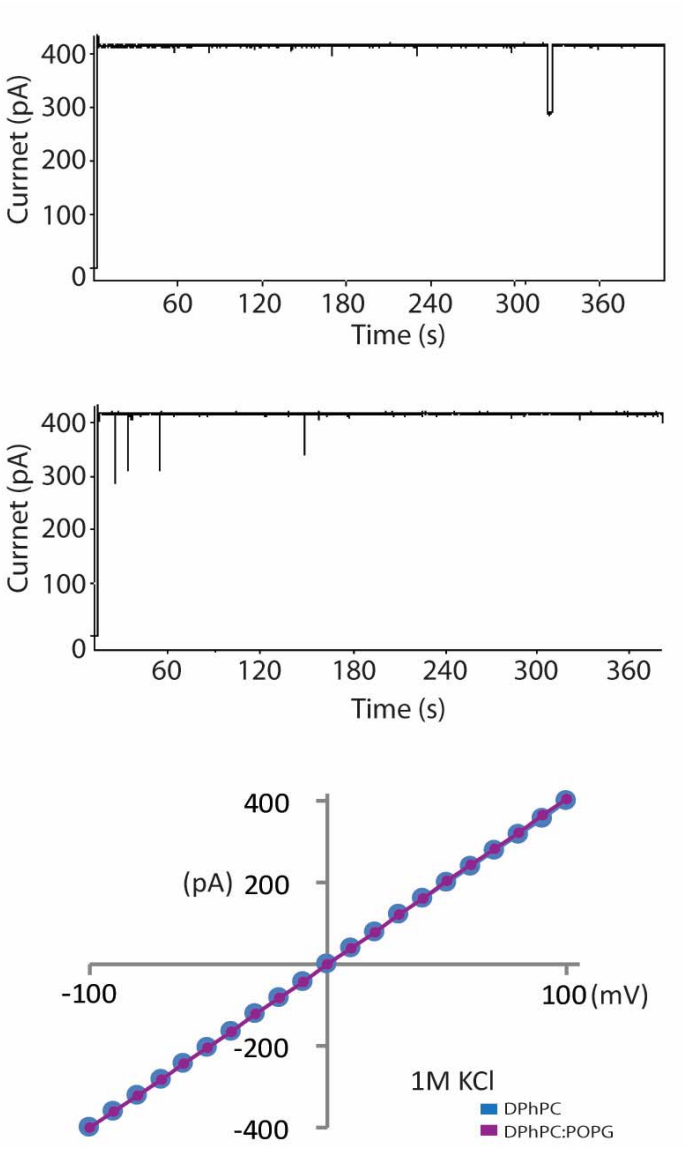


Figure 4.11: Electrophysiology measurements at pH 8 for OmpF reconstituted in planar lipid bilayers. Top channel recordings correspond to DPhPC only, whereas bottom channel recordings were obtained in the presence of POPG. Measurements were done at constant voltage of 100 mV. No significant change is observed between two measurements. The current-voltage curves (blue for DPhPC bilayer and purple for DPhPC:POPG bilayer) showed that the channel resistance does not change with addition of POPG.

The current-voltage (I-V) curve (Figure 4.11) shows no dependence of the current on the applied voltage in both bilayer compositions. Importantly, since the slope of the I-V curve remains unchanged after addition of POPG the channel resistance is unchanged in both cases.

This suggest that the lipid bilayer and OmpF structure is maintained the same regardless of the lipid composition hence the observed stabilisation effects can only be attributed to POPG binding rather than a conformation change of the bilayer by the presence of POPG at low pH.

4.5 Lipid binding enhances peptide binding to OmpF

OmpF is known to bind the peptide OmpF-Binding Site 1 (OBS1). As POPG appears to stabilize OmpF in the open state at low pH, it is of interest whether in a protonated environment, as proxied by positive polarity mass spectrometry, OBS1 binding could be influenced by the binding of POPG. Previous studies employing non-denaturing MS have discovered a biologically relevant mechanism of colicin entrance to the cell by recruiting OmpF. In these experiments it was found that up to three OBS1 peptides can bind to OmpF, one per each pore. Here, however, to simplify the spectra, a low concentration (5 μM) of the peptide was used. Although the ratio of the protein to lipid was 1:5, the major peak present in the spectrum corresponds to the unbound form (Figure 4.12). Up to two lipid-bound peaks are observed.

The effect of POPG on the binding of OBS1 to OmpF can be observed in a single measurement since both the bound and unbound lipid forms are present. Since the mass of the OBS1 peptide (1,552 Da) is very similar to the mass of two bound POPG lipids (1,540 Da) an Orbitrap mass analyser that can perform high resolution measures was used. Notably, the mass spectra showed that there is a higher population of OBS1 bound to POPG-bound OmpF than to OmpF alone (Figure 4.12). Taken together, these results show that upon exogenous stress, in the form of low pH, potential changes in lipid binding preference can affect the apparent stability and function of OmpF.

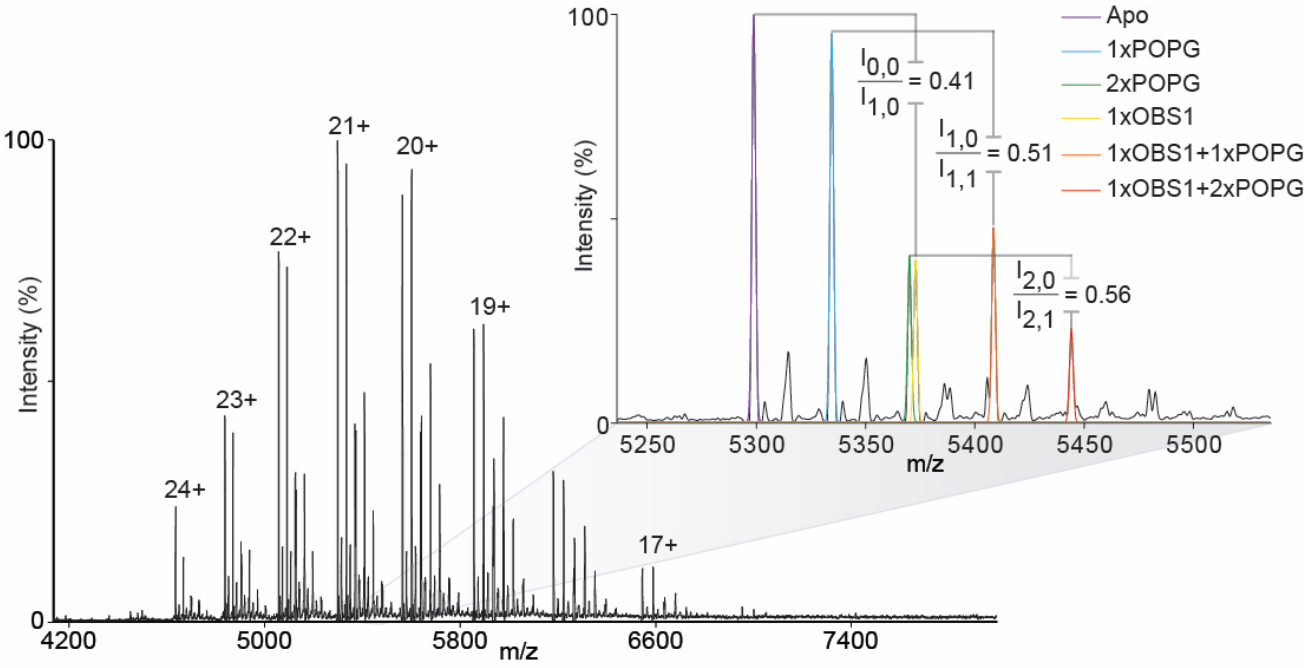


Figure 4.12: Mass spectra of OBS1 peptide binding to OmpF in the presence of POPG. The protein at 9.4μM was initially solubilised in OG at 2xCMC and incubated 50μM of POPG lipids that were solubilised in 200mM ammonium acetate. 10μM of OBS1 was added to the mixture. Measurements were recorded in Q-Exactive orbitrap applying 180V in the HCD cell. The inset shows a zoom region around 21+ charge state. A higher intensity of OBS1 binding to OmpF-POPG and OmpF-(2×POPG) is observed relative to peptide binding to apo OmpF.

4.6 Molecular dynamics simulations show differences in preferential PG binding at low pH

Having established by mass spectrometry that lipid binding correlates with the composition of the lipid head group binding region and POPG binding at low pH improves the stability and function of OmpF, the attention is then shifted to assessing ionic interactions in lipid binding as a function of pH, by means of molecular dynamics simulations. These simulations allow the ionic interactions, in conditions that mimic both solution and gas-phase measurements, to be probed.[209] During a typical MS experiment all observed species are charged hence this scenario can be simulated by protonation of lipids and acidic residues,

although the exact position of the charges remains elusive. In solution measurements, for instance during a patch clamp experiment at pH 4, only the acidic residues are protonated since the pK_a of the lipids is less than four. In addition, to measure the effect of charged lipids on binding, a hypothetical case where the lipids are protonated but not the protein was also considered.

Simulations were performed with two membrane proteins, AmtB and OmpF, that possess different percentages of acidic residues in the lipid head group binding region, and for which differences in lipid binding were observed by mass spectrometry. AmtB has a low percentage of acidic residues in this region, and OmpF has a high percentage of acidic residues. If acidic residues play a role in lipid binding then under protonated conditions, where the acidic residues and POPG become neutral, then a greater preference for binding POPG would be expected compared to higher pH conditions.

Due to the large ensemble of molecules and the low diffusion coefficient of the lipids, bilayer simulations were performed at the coarse-grained level. Each lipid head group was assigned to four beads whereas the lipid chain was simplified to nine beads. Measurements were performed using MARTINI force field[210] and each simulation ran for 1 μs. Prior to each simulation the membrane bilayer, composed of equal concentrations of POPC and POPG, was allowed to self-assemble. As the bilayer is composed of only two lipids the time that POPG spent bound to the protein was expressed as a preference for POPG lipid relative to POPC by comparing relative residence times of lipids in proximity to the acidic residues.

In protonated conditions OmpF has a greater preference for binding POPG, and there is no equivalent discernible change in lipid binding preference for AmtB (Figure 4.13). This suggests that acidic residues can influence lipid binding as pH lowers. Interestingly, in the

hypothetical case where lipid was the only protonated species a reduction in POPG binding was observed for both AmtB and OmpF, suggesting that charge synchronisation plays a significant role. Conversely, when pronating only the protein the opposite effect was observed; an increase in POPG binding that was more dominate for OmpF than AmtB.

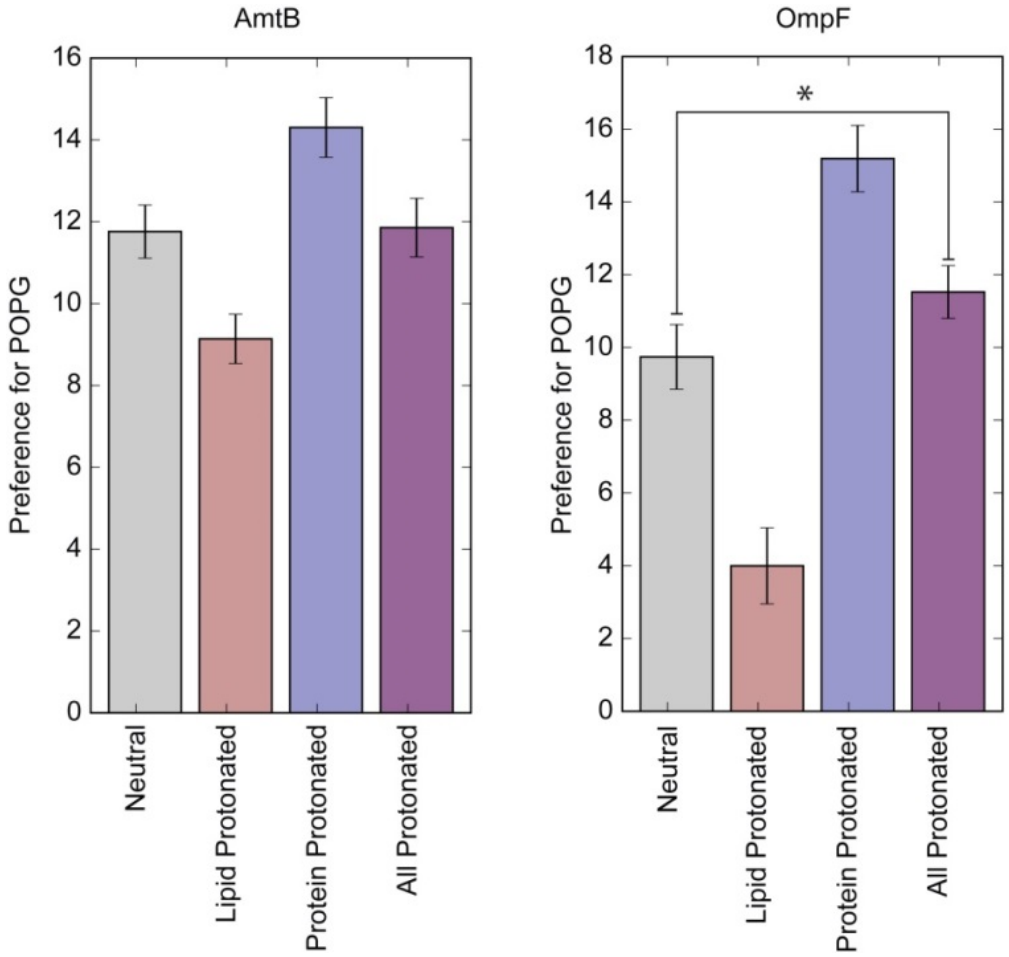


Figure 4.13: Molecular dynamics simulations of AmtB and OmpF. (a) POPG preference as compared to POPC lipids. Four protonation states are considered, no protonation of chargeable residues, only lipid protonation, only protein protonation and protonation of both lipid and protein. For both AmtB and OmpF protonation of lipids only causes a reduction in POPG preference, whereas protein protonation only demonstrates an increase in stability POPG binding. When both lipid and protein are protonated no significant preference for POPG is observed for AmtB (p value 0.74), in contras an increase in POPG affinity is observed for OmpF (* p value 6 x 10^{-7}). Error bars represent 2× s.e.m. MD simulations were performed by Tom Newport.

To better understand and identify which residues are directly affected the preference for POPG binding was investigated for all four charge conditions at the residue level (Figure 4.14). In the case of AmtB, lipid preference was majorly affected by arginine and histidine residues, whereas acidic residues had no direct influence in the change for preference in lipid binding under protonated conditions. As expected the opposite was observed for OmpF, where both aspartic acid and glutamic acid dictated the preference for POPG under protonated conditions, whereas basic residues had a net lower preference for POPG. Interestingly, tryptophan residues where found to also influence lipid binding, potentially due to their close proximity to acidic residues. The combination where only the protein is protonated may well resemble solution conditions at pH 4 as the acidic residues are protonated whereas the lipid charge is not affected. This suggests that under pH stress the protonation of acidic residues may favour the binding of acidic lipids to membrane proteins, and these interactions are predominately through direct binding to basic residues.

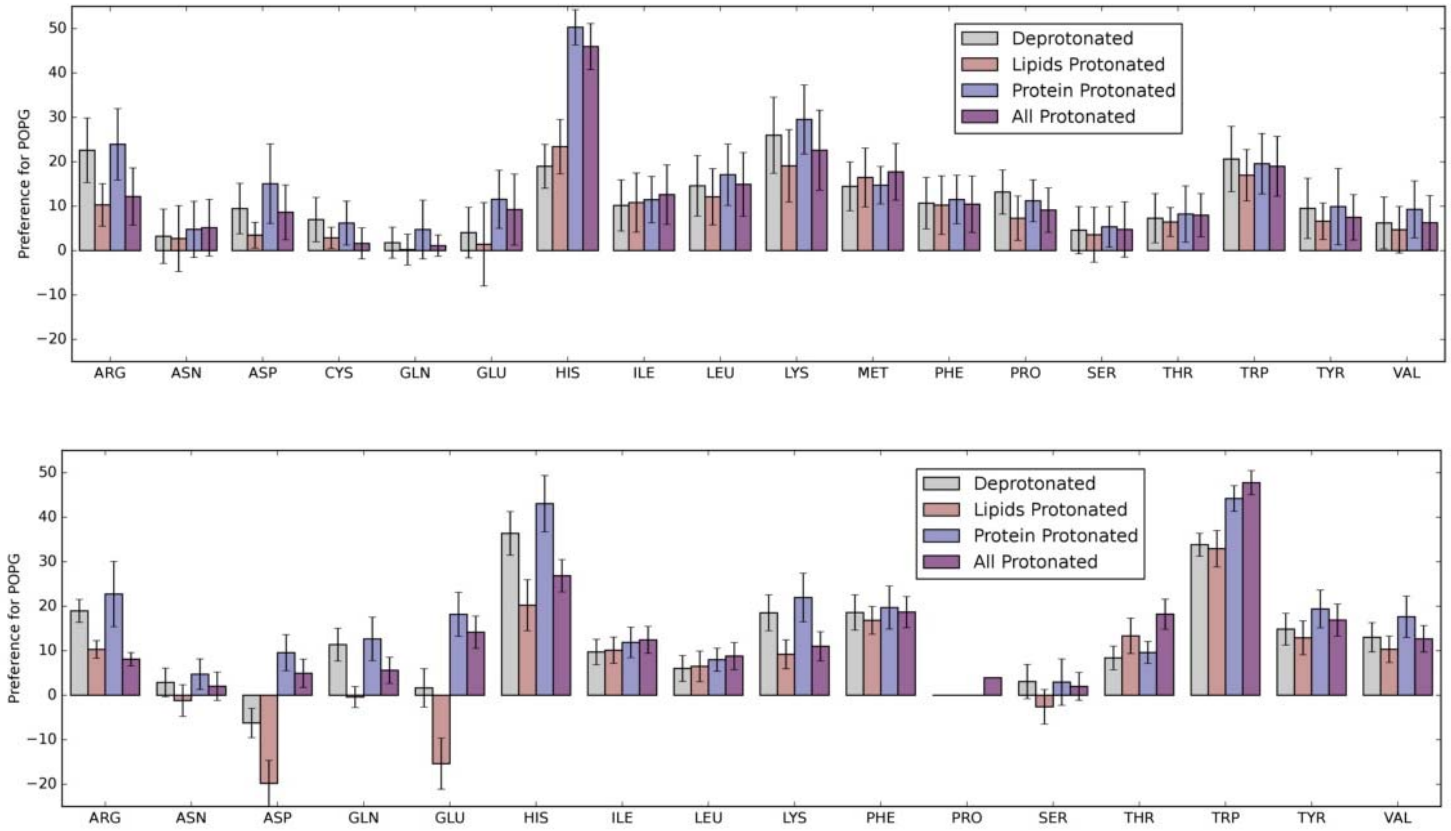


Figure 4.14: MD simulations at residue level of AmtB (upper panel) and OmpF (bottom panel) under normal and protonation conditions. All four modes of protonation are considered where calculations were performed with no alternation in protonation, only lipid being protonated, protein only protonated and both lipid and protein protonated. Data are shown relative to POPG preference as compared to POPC, a positive value shows preference for POPG and a negative value preference for POPC.

4.7 Conclusion

The presence of chargeable residues in the lipid head group binding region plays a significant role in the selectivity of lipid binding. The present data suggest that upon pH stress, different lipids are bound to membrane proteins to increase their stability. Calculating the chargeable residues and bonding distance to the lipid head group demonstrated that the vast majority of structurally known membrane proteins have a relatively low percentage area of chargeable residues in this region. Interestingly, around 10 percent of membrane proteins analysed had a high percentage area of charged (≥20%) residues in the lipid head group region, hence these proteins may bind lipids selectively based on the charge of the lipid head group in a protonated environment. Outer membrane proteins are expected to be more exposed to environmental changes. In addition, this class of protein has a higher proportional surface area of acidic residues in the lipid head group binding region relative to the inner membrane proteins, suggesting that for the outer membrane the electrostatic potential of proteins has a significant role in lipid binding and different lipids could be recruited in response to environmental changes.

OmpF at physiological pH is shown to preferentially bind PC over PG.[206] Here, by employing MD simulations and MS experiments, at low pH it was determined that the binding of PG is favoured due to synchronization of the lipid charge and the charge in the lipid head group binding region of the protein. These results suggest that the composition of annular lipids is dynamic, and lipid binding preference can change in response to environmental changes. Furthermore, at low pH, the protonation of acidic residues lowers the repulsion of acidic lipids and plays a crucial role in lipid binding.

The electrophysiology measurements performed in planar lipid bilayers that mimic cellular membranes showed that in the presence of PC lipid only OmpF closes rapidly at low pH. Previously it has been proposed that channel closure at high voltage and low pH is via a collapse of the outer domain.[207,208] In agreement with this model the results here show that binding of PG stabilizes the open conformation allowing OmpF to retain function in stressful environments for a longer time (Figure 4.15).

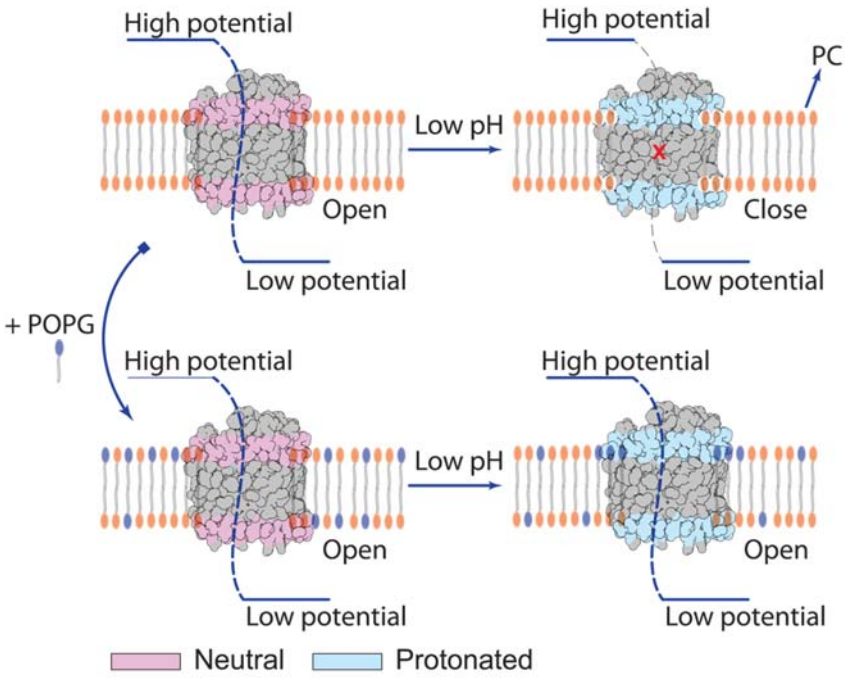


Figure 4.15: PG stabilises the open conformation of OmpF at low pH. In physiological pH the channel is in the open state. Lowering the pH shifts the channel to a closed form (denoted by the x). The addition of PG lipids stabilises the open state at low pH due to preferential binding of this lipid in the protonated environment relative to PC that is preferred at physiological pH.

Taken together these results show that annular lipid is a dynamic entity which can function as one unit with membrane proteins to increase the stability of a particular conformation and maintain function.

Chapter Five: The role of post-translational modifications and lipids in stabilisation of the oligomeric state of mammalian cytochrome *c* oxidase

The chemical composition of mammalian proteins in not limited to the 20 amino acids but rather vastly expanded by covalent modifications at specific residues. These post-translational modifications (PTM) have a great importance in biological function yet their study has been hampered by the lack of suitable methods.[211] Since each PTM changes the mass of the complex, conventional mass spectrometry has been the method of choice for initial characterisation, providing information on the position in the amino acid sequence and composition of the PTMs. Several PTMs have been identified, the most commonly seen being phosphorylation, acetylation and glycosylation. In a typical proteomics-based assay, the proteins are initially denatured and enzymatically cleaved.[42,212] The peptides produced are separated and their masses are measured with high accuracy to identify a potential chemical modification. Today, these experiments are performed routinely and parts of the procedure are fully automated. However, for non-denaturing mass spectrometry measurements the presence of PTMs in protein complexes provides an addition layer of complexity. As the mass of the complex is changed by each modification, the heterogeneity of the sample increases, resulting in multiple peaks in the mass spectra or even an unresolved hump. In the case of membrane proteins, the homogeneity of the sample can be further compromised due to the presence of bound lipids or detergent molecules. This complexity is increased further by the presence of multiple subunits within a complex, as in the case of cytochrome *c* oxidase (CcO).

CcO is an integral membrane protein complex composed of 13 protein subunits and it can bind a variety of lipids.[213] Despite the fact that the first crystal structure of bovine heart CcO was solved in the previous decade, the oligomeric state of CcO in solution remains elusive. Dimerisation of the complex has been previously proposed, however, definitive evidence for the dimer is lacking.[214] In this chapter non-denaturing mass spectrometry in tandem with conventional methods were used to investigate the oligomeric state of CcO and its stability.

Non-denaturing mass spectrometry of the intact protein complex revealed the presence of both the monomer and the dimer in solution and the stabilisation of the complex by a large lipid entity. Lipidomics-based assays were performed to identify lipid species that interact specifically with the enzyme. Furthermore, the phosphorylation and acetylation sites of cytochrome *c* oxidase, which were located in the peripheral subunits and the dimer interface, could potentially have a role in oligomeric stabilization. In addition to these findings, dimeric cytochrome *c* oxidase, which is composed of 26 protein subunits and associated lipids, represents the largest membrane-embedded protein complex that has been investigated in the gas-phase of a mass spectrometer. The importance of CcO and its biological implications and the challenges to study it provides to be studied by non-denaturing mass spectrometry are the major motivations behind this target.

5.1 The complexity of the mammalian cytochrome *c* oxidase

The electron transport chain in both bacteria and mitochondria consists of a series of membrane embedded protein complexes that transfer electrons to establish a proton gradient across the inner membrane.[215] This proton gradient is formed by four

complexes and is used by ATP synthase (complex V) to produce cellular energy (Figure 5.1). The electron transport chain is initiated in complex I (nicotinamide adenine dinucleotide – ubiquinone reductase (NAD-CoQ)). NADH is oxidised to NAD^+, which results in two protons and two electrons being transferred to the flavin mononucleotide (FMN) prosthetic group. While electrons are passed on to iron-sulphur clusters, the protons are actively transported to the outer side of the membrane. The electrons are cascaded to reduce ubiquinone (Q) and the reduced product (QH_2) diffuses through the membrane and binds to complex III. In a separate initiation point, additional electrons are transferred to Q from complex II (succinate dehydrogenase), via reduction of flavin adenosine dinucleotide (FAD). Q transfers the electrons to complex III (cytochrome *bc1* reductase) which interact with cytochrome *c*, a water soluble protein in the intermembrane space. Finally cytochrome *c* is oxidised by complex IV (cytochrome *c* oxidase).[216] The electrons are utilized to reduce dioxygen to water and transfer four protons across the membrane.

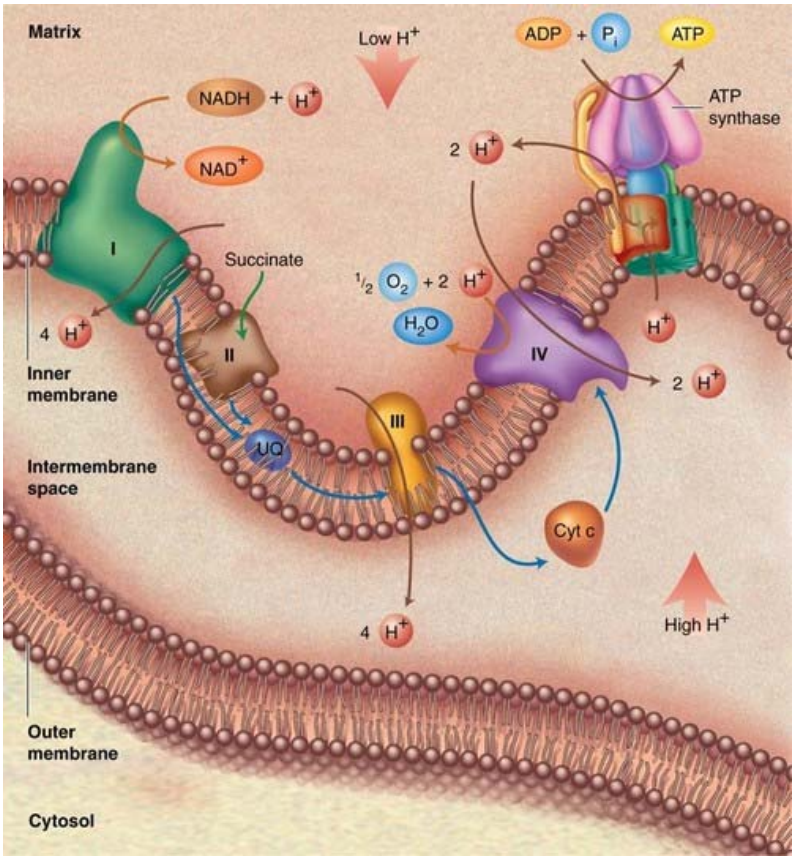


Figure 5.1: The electron transport chain in mitochondria. Figure taken from http://scienceknowledge.tumblr.com/post/70594638487

Cytochrome *c* oxidase is the last enzyme of the electron transport chain (complex IV). CcO of bovine heart is a large transmembrane complex and contains 13 different subunits.[217] The three largest subunits (I–III) are encoded by mitochondrial genes[218] and form the core of the enzyme;[219] subunits I and II contain metal binding sites and subunit III is a possible oxygen pathway.[220] The functional roles of the remaining ten protein subunits are yet to be determined, however, a stabilising function of these subunits has been suggested.[219] CcO was found to be dimeric in the crystal lattice[221] making it one of the largest transmembrane protein complexes studied by X-ray crystallography.

CcO contains a significant amount of intrinsic lipids[222] and several lipid classes have been observed in crystal structures of both bacterial and bovine CcO.[218,223,224] Two cardiolipins (CL), one phosphatidylcholine (PC), three phosphatidylethanolamines (PE), four phosphatidylglycerols (PG) and three triglycerides (TG) constitute the lipid content of bovine crystalline CcO.[225] Choline plasmalogens were also identified, however these were found to be identical to the corresponding PCs and likely are derivatives thereof.[225] Insights into protein-lipid interactions in CcO were obtained primarily by fitting the identified lipids into the crystal structure obtained at 1.8 Å.[220,226] Several lipids were thus identified to bridge the two CcO monomers suggesting that they may stabilise the dimer interface.

It is established that CcO is regulated by expression of specific isoforms that are dependent on the tissue, species, or developmental stage. CcO is also regulated by allosteric effectors such as ATP/ADP, palmitate or calcium, as well as by reversible phosphorylation.[227] To date, a total of 18 phosphorylation sites have been identified although more potential phosphorylation sites exist.[227] An interplay between allosteric ATP inhibition and reversible phosphorylation recently revealed inhibitory phosphorylation sites.[228]

Dissociation of tightly bound CcO subunits is also thought to regulate the enzyme.[227] Alkaline, urea or dodecyl maltoside treatment lead to selective loss of protein subunits,[229,230] which is often linked to inactivation of the enzyme.[231] Thermal denaturation revealed that tightly bound phospholipids have a stabilising role in both monomeric and dimeric CcO.[232] Subunit VIb has been proposed to increase the stability of dimeric CcO as suggested by its position in the crystal structure.[227] In addition to

dimeric CcO, a mitochondrial supercomplex $I_1III_2IV_1$ containing monomeric CcO has been detected.[233] However, as allosteric ATP inhibition requires cooperativity of the two monomers a biological relevance for the monomeric form continues to be debated. Consequently, it remains elusive if the dimer state of CcO is present in solution and required for activity. In addition, the role of PTMs and lipids in modulating structural stability and function regulation is not well understood.

The transferring of intact protein assemblies into the gas phase enabled the investigation of the subunit composition and determination of the oligomeric state in solution. To elucidate the stability of the oligomeric state of CcO, standard MS-based proteomics and lipidomics techniques were employed to identify PTMs and associated lipids. Thereby, in this chapter, all 13 subunits were characterised and it was found that both the monomer and the dimer are stabilised by a large lipid entity of a defined composition. Locating the identified PTMs in the atomic structure of CcO reveals hotspots for phosphorylation along the proposed dimer interfaces, suggesting stabilising roles in protein interfaces, and in regulating the enzyme.

5.2 Cytochrome *c* oxidase is dimeric

The sample complexity of CcO provides a challenge for direct characterisation by non-denaturing MS. To optimise the sample and increase MS compatibility, strategies developed previously for the analysis of membrane proteins, present in detail in chapters two and three were re-used, resulting in resolved spectra that show the complexity of the sample.

Initially MS measurements were performed using a DM-solubilised protein complex since this detergent was shown to maintain a stable conformation in solution and was used for protein extraction.[234] Under high activation energy, that is necessary to remove the bound detergent molecules by collisional activation in the gas phase, non-covalent interactions between the subunits of the protein could not be preserved. The protein complex dissociates to individual subunit that can be observed in the low *m/z* region of the spectrum (Figure 5.2). By matching the experimental mass with the theoretical one these subunits were assigned to VI, Va, VIc and VIIb. Based on their location in the crystal structure (PDB ID 2OCC) they were mostly found to correspond to soluble subunits on the periphery of the complex.

Importantly, individual subunits on the dimer interface were not observed indicating that they are tightly bound to each other. Moreover, three peak series in the high *m/z* range were also observed corresponding to stable sub-complexes of CcO. Due to the similar molecular weight of the protein subunits, multiple assignments are possible for these complexes; hence at this mass resolution their subunit composition cannot be identified. However, considering the mass of the smallest complex together with its position in the crystal structure it can be suggested that subunits I and IV form the core of these assemblies.

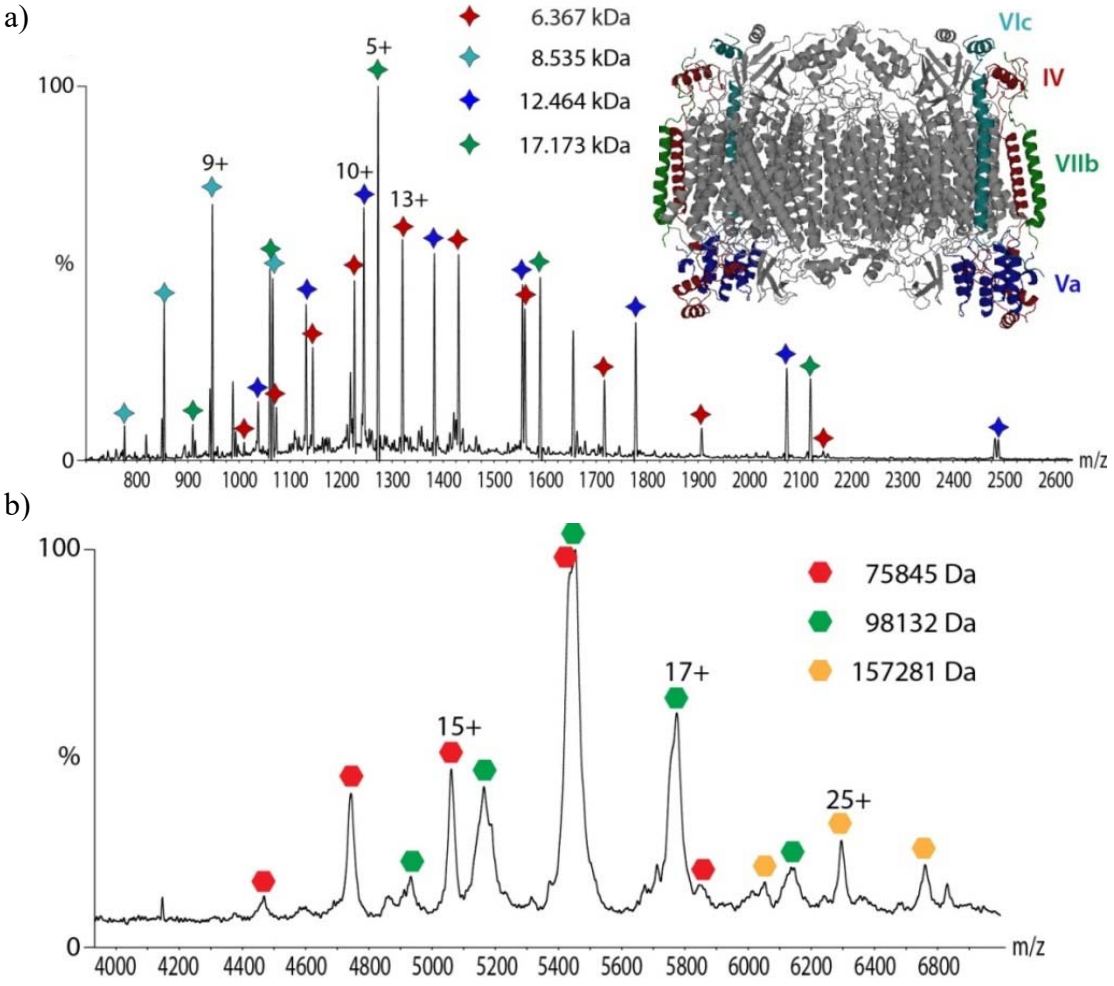


Figure 5.2: Subunit complexity of cytochrome *c* oxidase. (a) Intact CcO could not be observed in DM presumably due to the high activation energy required to release the protein from the detergent micelle. The low *m/z* region (<3,000 *m/z*) shows four peak series corresponding to peripheral subunits IV (red), Va (blue), VIc (cyan), and VIIb (green). (b) Three subcomplexes of CcO were observed by native MS in DM detergent. The smallest complex corresponds to the I–IV subcomplex.

These experiments reveal that CcO is not stable under the harsh mass spectrometry conditions required to release the protein complex from the DM detergent. In the light of methods developed in chapter 2, detergent screening using detergents from different classes was performed. This was to identify detergent compositions that are compatible with MS analysis and that would require lower activation energy to release the CcO

complex intact from the detergent micelle. In addition, measurements performed in the presence of DM detergent showed that low mass complexes were associated with a high charge, therefore, detergent that could reduce the charge state could possibly lead to better preservation of the intact complex. Detergents have a direct effect on the degree of lipidation and harsh detergents can be utilised to increase the degree of homogeneity of the sample and therefore signal intensity. C8E4 is considered a 'harsh' detergent, i.e. its ability to substitute for natural lipids is low compared with maltosides such as DM or DDM, but lower activation energy is required to release proteins from C8E4 micelles in the gas phase.[235]

After exchanging DM to C8E4 and recording a mass spectrum under controlled conditions, two charge state series at high *m/z* values were observed (Figure 5.3). Other detergent such as, OG, NG, LDAO did not improve the spectra quality. Considering the masses of each protein subunit obtained from denaturing LC-MS (Table 5.1), with exception of subunit I, III and VII_C that were calculated from their amino acid sequence, the expected masses of the intact CcO monomer and dimer correspond to 205.4 and 410.8 kDa respectively. The observed complexes have masses of 215.7 and 435.3 kDa, which are in close agreement to the expected values, but contain an additional 10 and 25 kDa of mass respectively. Due to the fact that CcO is a large transmembrane protein complex, these masses suggest that both the monomeric and dimeric complex are stabilised by other bound molecules, potentially a large lipid entity.

Subunit I, III and VII_C were not identified by denaturing LC-MS. Since these subunits are fully embedded in the membrane their high degree of hydrophobicity may

hinder the LC separation, suggesting that these subunits will bind strongly to lipids in solution.

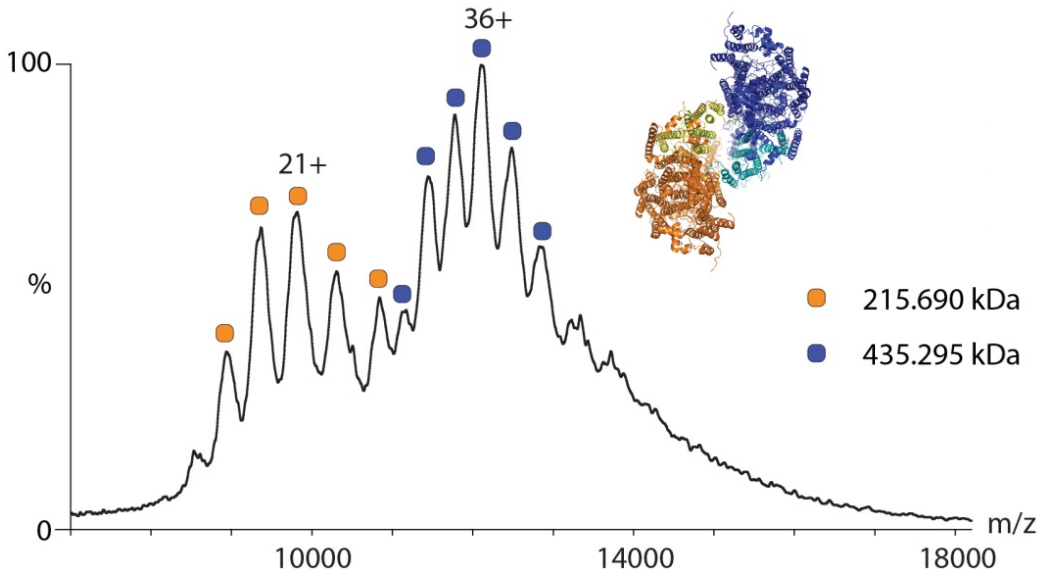


Figure 5.3: MS of the intact CcO complex reveals peripheral subunits and stabilisation of monomeric/dimeric CcO by associated lipids. After exchanging DM detergent for C8E4 the complex is intact and MS reveals monomeric (orange) and dimeric (blue) CcO. Observed masses of the monomer and dimer are higher than the theoretical calculated mass consistent with the association of lipids.

The mass spectrum in Figure 5.3 provides the first evidence of the presence of a dimeric CcO species in solution. It also suggests a possible equilibrium between both monomeric and dimeric species in solution, at least under *in vitro* conditions. The presence of the dimer in solution correlates well with the observed dimer from crystallographic measurements. Moreover, the dimeric interface is weak and only subunit VI_b is positioned within binding distance between the two monomeric species. Since this subunit is not conserved across different species[236] the dimeric interface

could potentially be stabilised by other factors. Next, the effect of lipids and PTMs in dimer interface stabilisation are considered.

Table 5.1: Exact protein masses as determined by denaturing LC-MS. The theoretical masses obtained from SWISS-PROT database, the retention time (RT) during LC-MS, the experimental masses and the mass differences are given for each protein subunit. For proteins that contain a transit peptide the masses with and without transit peptide are given. Subunits VIb and VIc include an initiator methionine which is removed. Note that some proteins eluted over several minutes and multiple mass spectra were obtained. For these proteins the average masses were calculated. Subunits I, III and VIIc were not identified by denaturing LC-MS.

	Theoretical mass [Da]				
Subunit	**w/ Transit peptide**	**w/o Transit peptide**	**RT [min]**	**Experimental mass [Da]**	**Δ Mass [Da]**
I	57,032	57,032	/	/	/
II	26,021	26,020	52.3–57.9	26,199	179
III	29,933	29,940	/	/	/
IV	19,572	17,150	35.8–37.8	17,297	147
Va	16,735	12,440	33.0–36.1	12,518	78
Vb	13,834	10,670	28.0–35.8	10,775	105
VIa	10,800	9,530	32.6–33.1	9,376	-154
VIb	10,156	10,030	27.0–29.6	10,094	64
VIc	8,610	8,480	29.6–31.5	8,615	135
VIIa	9,063	6,670	31.3–31.9	6,713	43
VIIb	9,065	6,360	/	/	/
VIIc	7,331	5,440	33.0–37.8	5,469	29
VIII	7,639	4,960	33.0–35.8	4,978	18
Total	225,791	204,722		205,366	644

5.3 Identification of co-purified lipids from cytochrome *c* oxidase

An established protocol was applied to analyse the lipid content of the CcO complex.[2] First, the protein-lipid complex was digested with trypsin and then the produced peptides were separated from the associated lipids by LC. The lipids were directly eluted into a high-resolution mass spectrometer and analysed by LC-MS/MS. Subsequently each lipid was identified by searching the precursor masses against a virtual lipid database containing TG, PC, phosphatidic acids (PA), phosphatidylserines (PS), PE, PG, phosphatidylinositols (PI) phosphatidylinositol phosphates (PIP) and CL with various lengths of acyl chains (www.lipidmaps.com). As this precursor ion search does not include the identification of the lipids by tandem-MS, the fragment spectra of abundant lipids were inspected by performing an MS/MS data search to verify the lipid species and identify the isomers. Following this workflow, prior to MS/MS fragmentation, hundreds of lipids from six different classes were identified to be present (Figure 5.4).

The large number of lipid species identified illustrates the complexity of membrane and the large combination of potential lipid-protein interactions. However, lipids potentially important for stability and function can also be identified from this large data set. To find lipids that are more abundant, the bound lipid species were ranked based on intensity. As these lipids have a higher affinity for the protein complex, it can be speculated that they are structurally and/or functionally important. The lipid series from each class covered a large range of intensities. The intensities of lipids were fitted to a Gaussian distribution and the 75th percentile was considered to identify the most abundant lipids. On average, PE and PG provided the highest intensity, whereas

cardiolipin species were the least abundant. This analysis showed at least 20 lipid molecules from each class bind to the protein complex with high affinity.

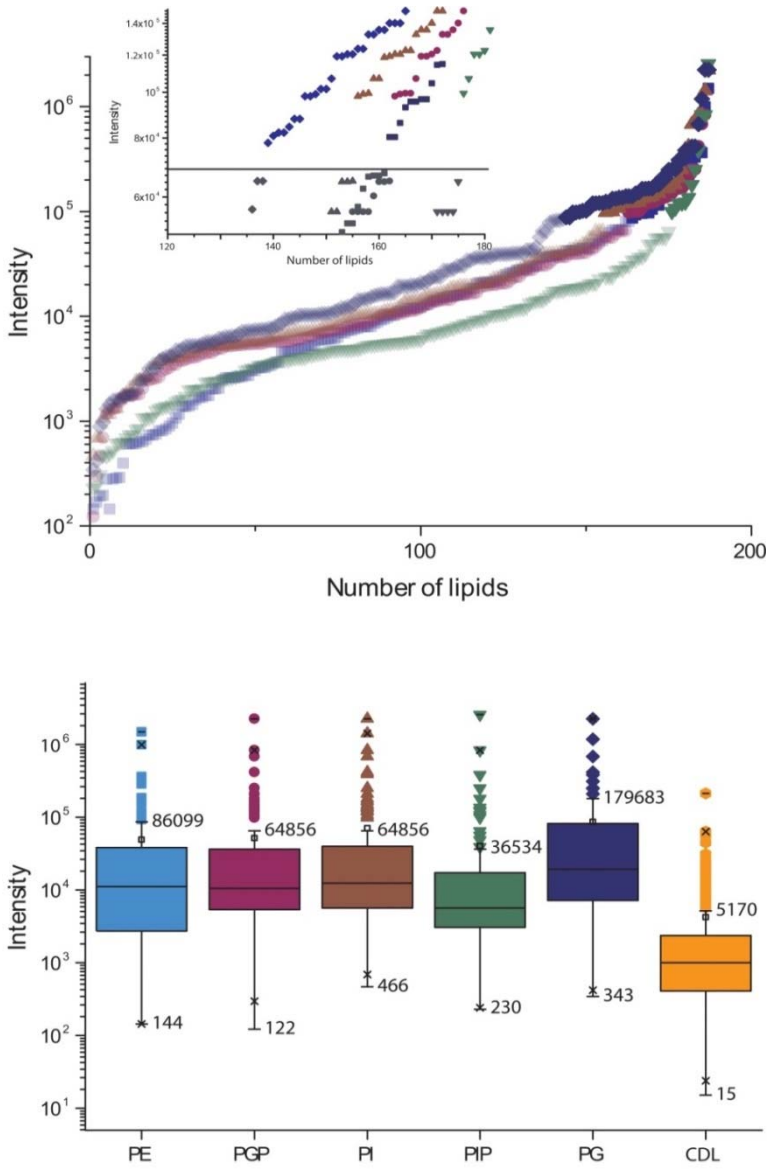


Figure 5.4: Lipids co-purified with cytochrome *c* oxidase. (a) The number of lipids ordered based on their intensity, high intensity lipids are shown in dark colour. PG (dark blue), PIP (green), PI (brown), PGP (magenta), PE (blue). Inset shows the cut off between the high and low intensity lipids. (b) Statistical analysis of the bound lipid species for each lipid class. The box plot presentation shows the 25, 50 and 75 percentile. The intensity of the 75 percentile is indicated.

By using MS/MS for lipid characterisation, four lipid species PG, PI, PS and PE were identified (Table 5.2). Manual inspection of the tandem mass spectra also revealed one species that was not included in the lipid database, phosphatidylglycerol phosphate (PGP), characterised by neutral loss of phosphate and co-elution with the corresponding PG. This implies that the identified PG isomers originate from their related PGP by a neutral loss of the phosphate group during activation in the mass spectrometer.

The lipid analysis further revealed preferences of fatty acids in these lipids; 16:0, 18:0, 18:1, 18:2 and 20:4 are clearly the favoured species. Most abundant lipid precursors are indicated in a combined mass spectrum over the LC elution time (Figure 5.5). The combined mass spectrum also reveals some singly charged species above 1,300 *m/z*, most likely corresponding to CL isomers. However it proved to be challenging to use this method to fragment, due to their low intensity and high mass. The low intensity of this lipid can be attributed to its low ionisability or loss during LC separation. A precursor ion search was conducted to reveal the most abundant chain length, which was found to be 18 carbon units long. While 16 and 20 carbon chain lengths were also found, no odd number of carbon chain lengths was identified, or these were excluded due to low intensity, suggesting that even number chain lengths are preferred.

In total, 35 lipid isomers belonging to six lipid classes were identified. Of these, PE is the most abundant species and showed 14 isomers and the highest precursor intensities. Compared to previous studies, lipid species that were not anticipated to be associated with CcO (e.g. PI and PGP) were also identified.

Table 5.2: Identified lipids in CcO. The lipid class, the identified *m/z* and retention time (RT), the area of extracted ion chromatograms and the identity of fatty acid side chains are given.

Lipid species	m/z	RT [min]	XIC [area]	Side chains	
Phosphatidylserine	786.53	17.06	5176354	18:1	18:1
	788.55	18.65	5434349	18:0	18:1
	810.53	16.55	1032265	18:0	20:4
	812.55	17.42	3990216	20:3	18:0
Phosphatidylethanolamine	698.52	16.10	86942694	18:2	15:1
	708.50	16.38	2500554	20:4	14:1
	714.15	17.24	7536574	16:0	18:2
	720.50	16:38	3252866	20:5	15:1
	722.52	17.59	36521846	15:1	20:4
	726.51	19.93	108115079	18:2	17:1
	736.50	15.57	8387826	16:0	20:5
	740.53	17:59	27057547	18:1	18:2
	742.54	19.40	177911109	18:0	18:2
	750.55	19.57	55999319	20:4	17:1
	752.53	17.94	13276776	20:4	17:0
	766.54	20.03	351505712	20:4	18:0
	768.56	19.75	50585719	18:0	20:3
Phosphoinositol	780.52	15.18	2266030	20:4	19:0
	885.55	14.03	54637801	18:0	20:4
	861.55	14.20	19990435	18:2	18:0
	863.55	16.52	3662266	18:1	18:0
	887.57	14.37	20786368	18:0	20:3
Phosphoglycerol	747.52	14.53	998032532	16:0	18:1
	749.54	15.35	428464867	16:0	18:0
	775.55	15.69	287891760	18:0	18:1
	745.51	14.03	240267363	18:2	16:0
	773.54	14.70	4843126	18:2	18:0
Phosphoglycerolphosphate	829.52	14.53	114242798	16:0	18:1
	831.54	15.35	30210395	16:0	18:0
	827.51	14.03	21788964	18:2	16:0

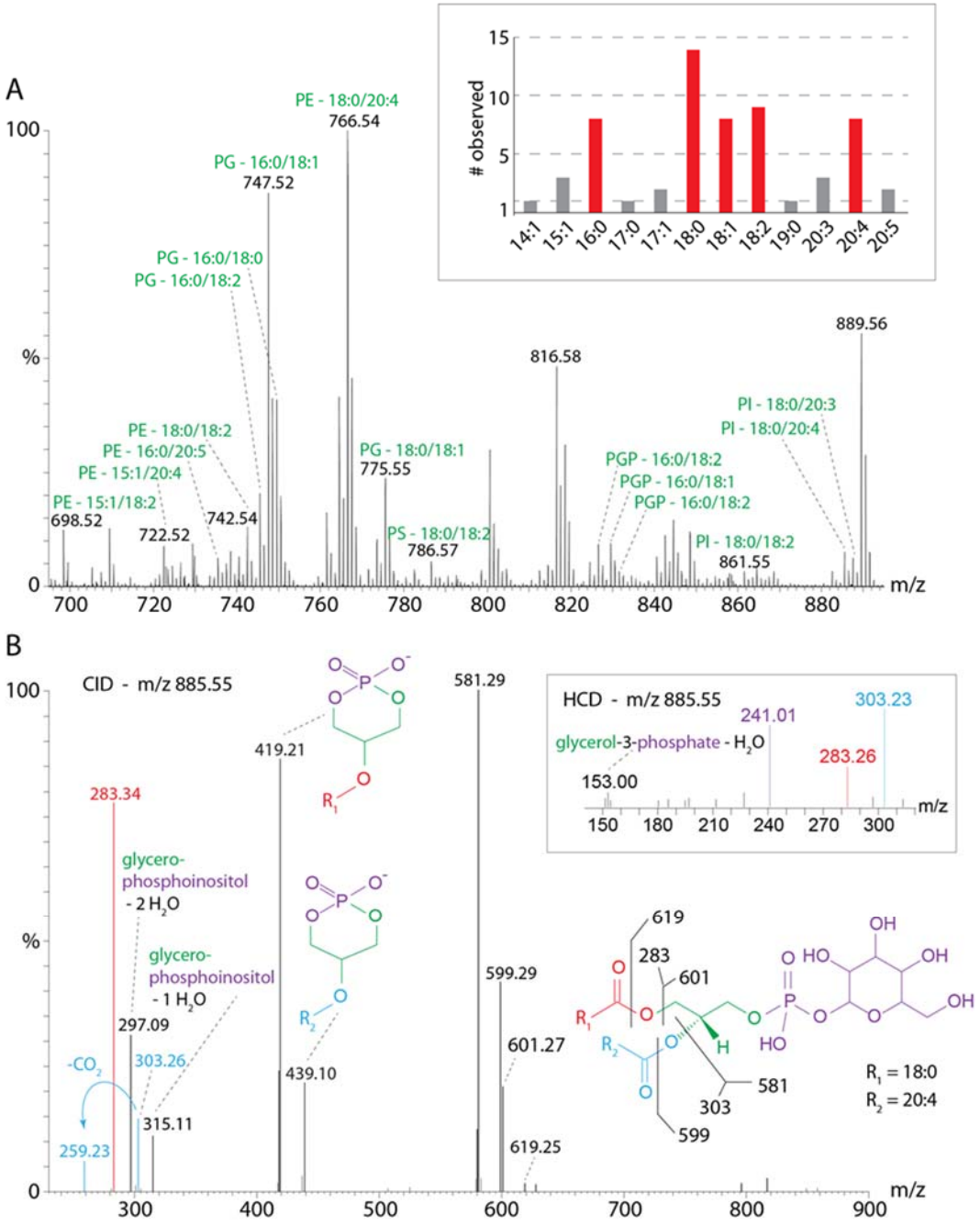


Figure 5.5: Lipid identification. (a) Combined mass spectrum over LC elution time. Most abundant precursors are labelled. PE is highly abundant in the CcO complex. The favoured fatty acid side chains are shown (inset). (b) Fragment spectrum of PI (*m/z* 885.55). Specific fragments are labelled. HCD further confirmed the presence of PI by additional low mass fragments (inset).

These novel lipid associations could be verified by their MS/MS spectra; an example for PI-18:0/20:4 is shown in Figure 5.5b. This high level of lipids implies considerable association of lipids with this complex, sufficient to form a protective shield but that they may reside within the subunit interfaces. Indeed, the mass difference between the anticipated mass of the dimer and the measured mass equates to 25 kDa, consistent with over 30 phospholipids. It is therefore likely that the CcO complex is stabilised by these multiple lipids in the natural membrane and these lipids may also play a role in mediating dimerization.

5.4 Post-translational modification of cytochrome *c* oxidase

Having identified a large set of lipid bound to the protein complex that potentially stabilises the dimer interface, the attention was shifted to the identification and function of PTMs on the protein stability. Initially the presence of all protein subunits was confirmed by liquid chromatography-coupled tandem-mass spectrometry (LC-MS/MS). The protein complex was denatured in low pH solution and was separated from the detergent micelle and bound lipids prior to enzymatic cleavage. Due to the fact that CcO contains many integral membrane protein subunits, tryptic and chymotryptic digestions were performed to allow generation of sufficient peptides for accurate protein identification. Subunits I-III, VIIb and VIII are mostly embedded in the membrane and have only few tryptic peptides; their sequence coverage after tryptic digestion was therefore low (6–40 percent). However, after chymotryptic digestion, these subunits could be identified with high sequence coverage (20–72 percent). Using this approach the presence of all 13 subunits in the complex was confirmed.

The exact mass of each peptide was determined and compared with the theoretical mass, allowing for potential modification to be identified. The exact peptide masses were determined from the mass spectra of ten out of the thirteen subunits. The missing subunits, I, II and VIIb are very hydrophobic and presumably their separation by LC is therefore impeded. Interestingly, in several cases the proteins eluted over several minutes and in different chromatographic peaks, which indicate multiple forms of the protein. In addition, higher masses than the theoretical values were obtained for most of the protein subunits suggesting that they are post-translationally modified.

Standard proteomics methods were performed to uncover the PTMs and the modification sites of CcO. These methods included potential phosphorylation of serine, threonine and tyrosine residues as well as acetylation of lysine residues in database searches. In total, there were 28 modified residues (14 phosphorylated and 14 acetylated) identified in seven of the 13 CcO subunits.

Locating the identified PTMs in the crystal structure of the CcO dimer (PDB ID 2OCC) shows that phosphorylation occurs primarily in the soluble subunits on the matrix side of the inner membrane. Acetylated lysine residues are more widely distributed and often in close proximity to phosphorylation sites. Many acetylated residues are located at the protein interfaces and, importantly, at least three of the sites (K71, K77, K171) are located at the dimer interface suggesting a stabilising role in the CcO complex.

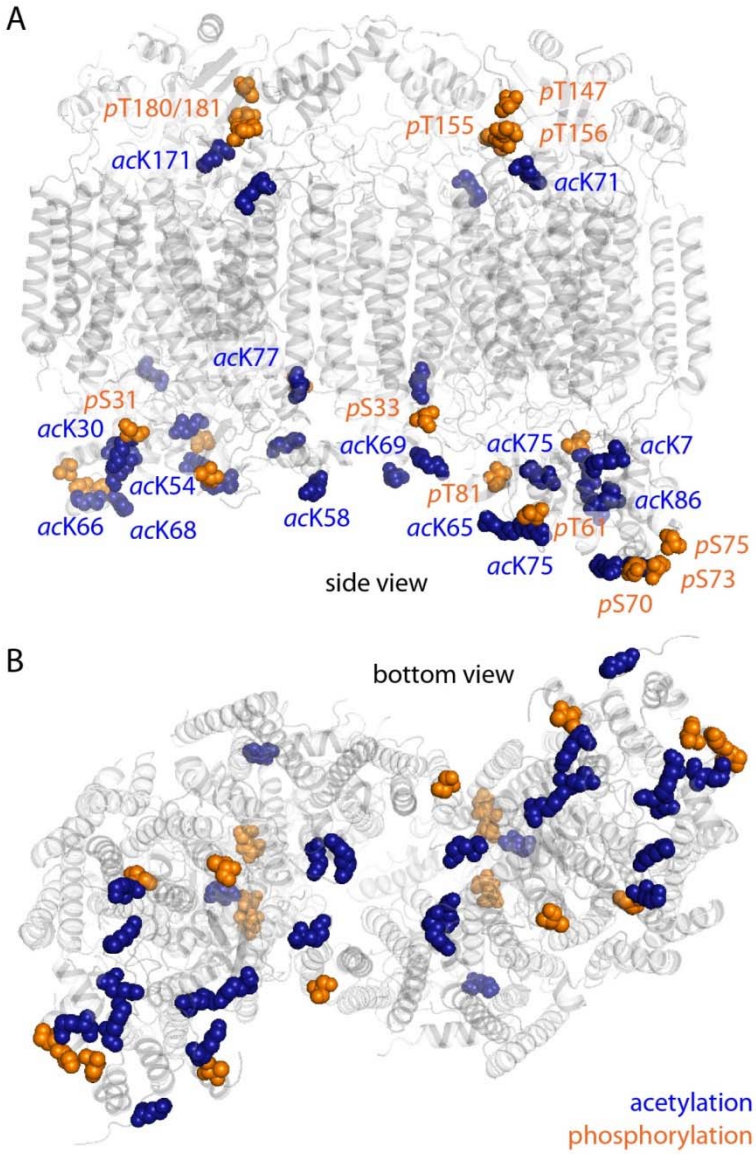


Figure 5.6: PTMs are located in the peripheral subunits and in the dimer interface of CcO. Phosphorylated and acetylated residues are highlighted (orange and blue space fillings, respectively). Residue numbers are given. Phospho-sites are mainly located in peripheral subunits. Acetylated residues are located at protein interfaces and in the CcO dimer interface. (a) Side view of CcO, (b) bottom view of CcO (PDB ID 2OCC).

Table 5.3: Most of the modified residues are located in subunits II, IV, Va and Vb with at least four PTMs in each of these subunits. Interestingly none of these subunits shows a preference for phosphorylation or acetylation, all four subunits are equally modified by both PTMs. However, it appears that all phosphorylated residues are serines or threonines with no phosphorylated tyrosine detected.

Sub-unit	PTM	Modified Residue	Peptide sequence	max. Mascot score
II	Acetylation	*ac*K171	AVPSLGLKTDAIPGRL	68
III	Acetylation	*ac*K77	ESTFQGHHTPAVQKGLR	64
IV	Acetylation	*ac*K29	AHGSVVKSEDYALPSYVDR	58
IV	Phosphorylation	*p*S30	AHGSVVKSEDYALPSYVDR	86
IV	Acetylation	*ac*K53	DYPLPDVAHVKNLSASQK	56
IV	Acetylation	*ac*K67	EKEKASWSSLSIDEK	88
IV	Acetylation	*ac*K69	EKASWSSLSIDEK	79
IV	Phosphorylation	*p*S72	EKEKASWSSLSIDEK	67
IV	Phosphorylation	*p*S74	ASWSSLSIDEK	66
IV	Acetylation	*ac*K86	RLKFKESFAEMNR.S	53
Va	Phosphorylation	*p*S47	SHGSHETDEEFDARWVTY	75
Va	Phosphorylation	*p*T60	WVTYFNKPDIDAWELR	89
Va	Acetylation	*ac*K64	WVTYFNKPDIDAWELR	42
Va	Acetylation	*ac*K74	KGMNTLVGYDLVPEPK	112
Vb	Acetylation	*ac*K57	KGQDPYNILAPK	78
Vb	Acetylation	*ac*K68	KGQDPYNILAPKATSGTK	71
Vb	Acetylation	*ac*K74	ATSGTKEDPNLVPSITNK	66
Vb	Phosphorylation	*p*T84	AASGTKEDPNLVPSITNK	71
VIa	Acetylation	*ac*K70	TKPFSWGDGNHTFFHNPR	100
VIc	Acetylation	*ac*K7	STALAKPQMR	61

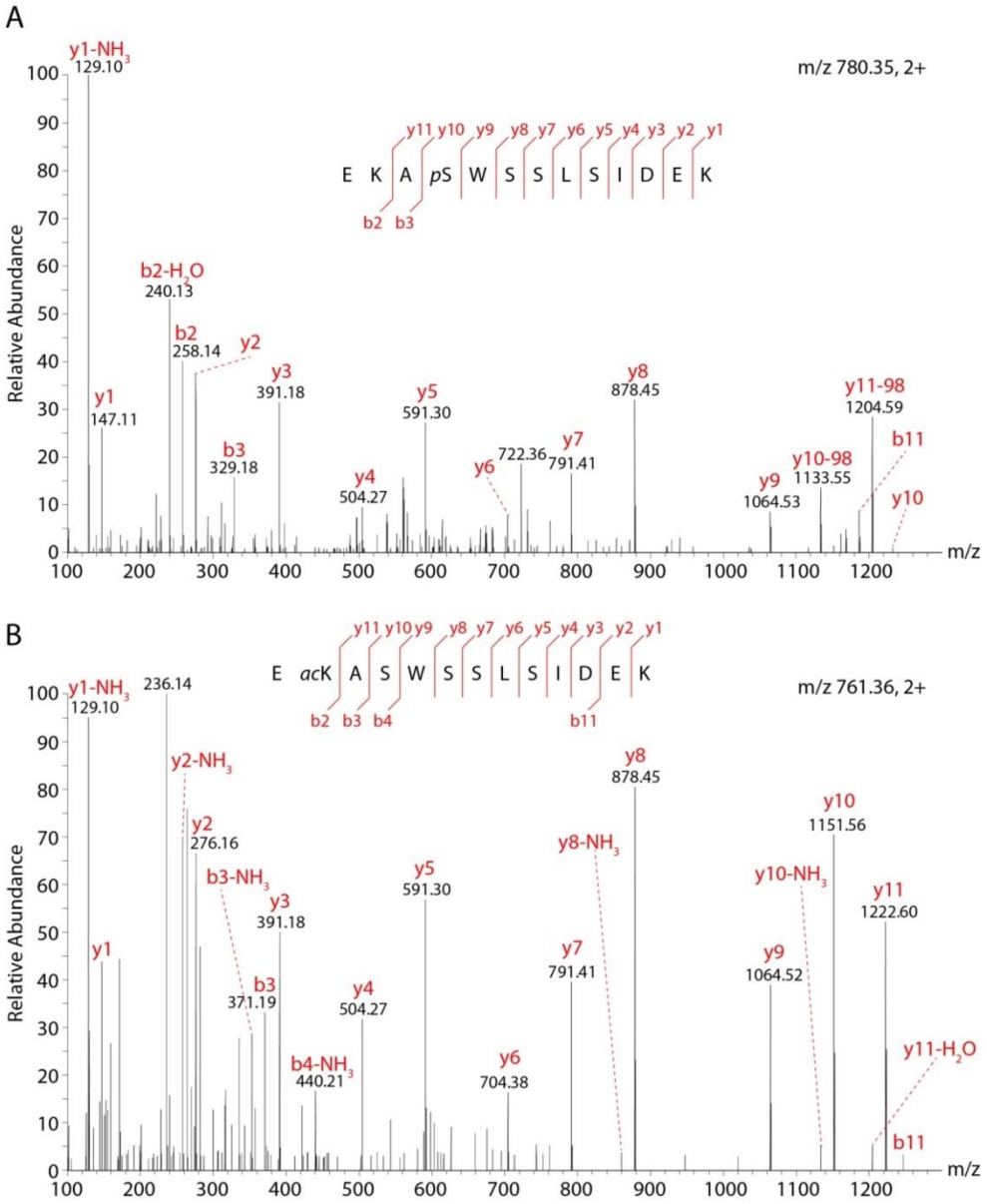


Figure 5.7: Example MS/MS spectra of phosphorylated (a) and acetylated (b) peptides.

5.5 Conclusion

The five enzymes of the electron transport chain in mitochondria establish a proton gradient across the inner membrane and use ATP synthase for energy production. Cytochrome *c* oxidase is the last of these enzymes and represents a large transmembrane protein complex. Although dimerisation of cytochrome *c* oxidase has

been proposed, the dimer has not been confirmed in solution. Using non-denaturing mass spectrometry the oligomeric state of cytochrome *c* oxidase was elucidated providing the first evidence that both the monomer and the dimer are present in solution and potentially are stabilised by many associated lipids.

CcO purified from bovine heart was characterised and accurate protein masses were determined and identified PTMs and lipids associated with the intact complex. Transferring intact CcO into the gas phase of a mass spectrometer showed that high energy is required for releasing the protein-lipid assembly from the detergent micelle causing dissociation of the protein subunits. However, exchanging the detergent allowed release of intact monomeric and dimeric CcO, containing 13 and 26 subunits, respectively. The masses observed for these complexes revealed a large lipid entity associated with both oligomers.

A stabilising role for lipids was previously described,[225] but the role of lipids in stabilising the dimeric CcO was debated due to the low number of known bound lipids. However, in this study a large number of bound lipids were found. Using non-denaturing mass spectrometry it was found that the monomer binds an additional 10 kDa corresponding to approximately 13 phospholipids, assuming an average molecular weight of 775 Da. The lipids associated with dimeric CcO constitute a larger additional mass of 25 kDa; assuming that both individual monomers are stabilised by lipids yields about six phospholipids stabilising the dimeric interface. The subunit interface in the crystal structure forms a cavity, which may be where these lipids are located.

Recent molecular dynamics simulations identified seven binding sites for cardiolipin in CcO,[237] most of which are contradictory to previously described sites identified from

crystal structures.[225,238] The question therefore remains which of the associated lipids stabilise dimeric and monomeric CcO and where their binding sites are located in the CcO structure. Ion mobility measurements to guide molecular dynamics simulations may provide the means to investigate the binding of lipids in greater detail.

Employing conventional lipidomics methods, five lipid classes, each with several isomers were identified. Peak series corresponding to cardiolipin were obtained, however, the present experimental set-up did not allow their verification by tandem-MS. Considering the number of isomers identified and the peak intensities recorded, phosphoethanolamine appears to be the most prevalent lipid. Phosphoethanolamine and phosphatidylethanolamine are most abundant in bovine heart tissue[239] suggesting that they represent the annular lipids of the CcO complex. The lipid analysis also revealed three new lipid classes bound to CcO: phosphatidylinositol, phosphatidylserine and phosphatidylglycerolphosphate. These lipids are minor constituents of the mitochondrial inner membrane[239] suggesting that they bind to the CcO complex specifically. The role of these lipids in stabilisation could be further investigated by collision induced activation and measuring how they may change resistance to unfolding of the protein complex.

Acetylation and phosphorylation sites were also identified to gain insights into CcO regulation. Interestingly, subunits II, IV, Va and Vb were modified to the greatest extent. Locating PTMs in the crystal structure of CcO shows that they are essential for both stabilising and regulating the enzyme. Phosphorylation sites were present predominantly in soluble domains on the matrix side of the inner membrane accessible for kinases and phosphatases. Strikingly, phosphorylation was mainly detected on serine

and to a lower extent on threonine residues, but not on tyrosine residues. In agreement with other studies, [240] this indicates that these sites have a regulatory role rather than a stabilising role, which is often ascribed to phosphorylation on tyrosine residues.[240] A regulatory role for acetylation of lysine residues has not yet been shown. Interestingly, acetylation sites are in close proximity to phosphorylation sites and also in the dimeric interface, suggesting a stabilising role than a regulatory one.

Aside from the biological insights obtained, this protein-lipid assembly is the largest membrane embedded intact complex analysed by MS so far. Combining proteomics, lipidomics and MS of the intact complex meant insights could be gained into the different aspects of CcO regulation and stability. The data present here demonstrate for the first time an intact CcO dimer in solution and at the same time suggest equilibrium between the monomeric and dimeric forms of the enzyme.

Chapter Six: Conclusions

Over the last twenty years by providing structural information on samples with ever increasing complexity, mass spectrometry has become an important technique in the investigation of proteins. Recent advances have pushed further the boundaries of the technique, enabling the analysis of large biological membrane protein machineries. Key to these has been an ever increasing understanding of the underling mass spectrometry processes and mechanisms, from protein ion formation, activation to detection.

Due to the limitations of many analytical methods used for the study of membrane proteins a challenge is identifying conditions that can maintain membrane proteins in folded and active conformations in the absence of the natural lipid bilayer. Finding that detergents can directly affect charge state distributions of membrane proteins, and importantly revealing their charge carrier properties, has facilitated the identification of detergents compatible with native mass spectrometry, and those that are more likely to preserve the native-like conformations of membrane proteins in the gas phase. In addition, the amount of activation energy and the resistance to Coulombic repulsion are two key parameters that influence protein conformation in a mass spectrometer. The ability to access lower charge states and release membrane proteins at low activation energy makes polyethylene glycol detergents and the detergent LDAO eminently suitable. With improvement in sample preparation techniques and screening a variety of detergents in a timely fashion, finding detergents that maintain membrane proteins in solution whilst simultaneously being mass spectrometry compatible is possible.

In tandem with ion mobility, mass spectrometry based methods can characterise the structure of membrane proteins and provide important information regarding their folded structure, oligomerisation and the connectivity of subunits. An important application will be the utilisation of mass spectrometry to guide other analytical methods that provide high resolution structural insight. For instance, different detergents can be used while screening parameters such as incubation time and temperature for membrane proteins extraction from lipid bilayers. The quality of the samples, as analysed by mass spectrometry, can then be used to guide the production of highly homogenous samples that may be useful for protein crystallisation. At the very least, it may assist in narrowing the crystallisation screening process to those detergents that maintain the native oligomeric state of a protein. The ability to characterise the stability of proteins in the gas phase can be utilised to guide identification of mutations that enhance structural stability.

Membrane proteins are different to soluble proteins in their interactions with lipids. Non-denaturing mass spectrometry is a useful method for identification of ligand binding to proteins, with minimal sample pre-treatments, and the ability to directly measure binding stoichiometry. Instrumental developments applied in this research such as in-source high-voltage activation, reduction of inlet pressure or IRMPD activation allows the identification lipids bound to membrane proteins. Detergent removal in the source region is proposed as a method capable of isolating the bound state for further characterisation. The ability to perform tandem mass spectrometry on membrane proteins and interrogate the lipid bound peaks by collisional activation allows for high accuracy identification of bound lipids.

Structurally important lipids typically bind to membrane proteins with high affinity and these interactions are preserved in the gas phase. By applying a novel method to quantify the stabilising effects of lipid binding, structurally important lipids can be identified. Following previous results found by this method characterisation of *Pf*MATE stability upon lipid binding showed that cardiolipin, unlike other phospholipids, enhances *Pf*MATE resistance to unfolding, suggesting that this lipid is structurally important. The role of lipopolysaccharides in stabilising outer membrane proteins has also been suggested; measurements performed on OmpF show that lipopolysaccharide stabilises this protein. Current models of the arrangement of lipids in bilayers suggest the possibility of islands, stabilised by protein-protein interactions, in the outer membrane.[241] It may be possible to quantify the stabilisation provided by lipids in protein-lipid-protein assemblies. Here the lipids may act by stabilising the interactions of weakly associating regions of the protein interfaces, and this may provide a better understanding of compartmentation of cellular membranes.

The investigation of membrane protein stabilisation by lipids using mass spectrometry is at the moment performed on membrane proteins with synthetic lipids or a mixture of endogenous lipids bound. A limiting factor in these analyses is the inability to unambiguously identify bound lipids. Advances in high resolution mass analysers, such as Orbitrap technology[12], mean that it is now possible to resolve differences in chain length of bound lipids. If these measurements can be combined with ion mobility, it may be possible to both identify and quantify the stabilising effects of endogenous lipid mixtures bound to membrane proteins.

Lipid-protein interactions were studied in the context of pH, which can modulate the ionic interactions at the lipid-protein interface. Assuming that the most dominant ionic interactions will be established between chargeable residues in the vicinity of the lipid head group, a global analysis of the percentage area of chargeable residues in close proximity to lipid head groups was performed on all membrane proteins with known structure. Membrane proteins with a high percentage area of chargeable residues where the lipid head groups bind were identified, suggesting that for these proteins ionic interactions at the membrane interface could be important for stability and selectively of lipid binding. Interestingly, the majority of membrane proteins with a high percentage area of chargeable residues were found to be outer membrane proteins. To probe the effect of chargeable residues on lipid binding by mass spectrometry, positive and negative polarity were used to protonate or deprotonate the proteins and directly measure the effect of lipid binding from the degree of protonation. Based on these studies POPG was found to preferentially bind to OmpF in protonated conditions, suggesting that this interaction may be also favoured at low pH. It has been observed that anionic lipids bind predominately to basic residues at physiological pH by pairing of opposing charges. Surprisingly, in this research protonation of acidic residues modulates the binding of anionic lipids to basic residues at low pH. Electrophysiology experiments found that the open conformation of OmpF is stabilised by POPG at low pH. Furthermore, a correlation between structure stabilisation and function was observed from the cooperativity in binding of a peptide to OmpF in the presence of lipid. These findings suggest that membrane proteins and the surrounding lipids can be thought to function as one unit, and in studying membrane proteins it is important to consider the dynamics of membrane protein-lipid interactions.

Mammalian membrane proteins provide additional challenges for structure characterisation by non-denaturing mass spectrometry. In addition to the presence of a diverse lipid environment they are typically post-translationally modified, resulting in highly heterogeneous samples. Nevertheless, the methods developed in this dissertation were successfully utilised to characterise the oligomeric state and the stability of cytochrome *c* oxidase, a particularly complex membrane protein. A drawback of high resolution structures elucidated by crystallography is that it is sometimes unclear whether the oligomeric state observed in a crystal is present in solution, or is a crystallographic artefact. By preserving the non-covalent interactions and identifying conditions to charge resolve this large membrane protein complex, both monomer and dimer were found to be present, providing the first clear evidence of the dimeric species in solution. Furthermore, these results were able to suggest that a large lipid entity may stabilise the weak dimeric interface. Using conventional mass spectrometry methods new post-translational modifications and bound lipids were identified. Importantly, acetylation sites were found to be positioned in close proximity to the dimeric interface and the membrane region. This suggests a stabilisation role of both post-translational modifications and lipids in the oligomeric state of cytochrome *c* oxidase.

This dissertation exemplifies that from an increased understanding of the mechanistic aspects of mass spectrometry, advances can be made in its application to studying difficult samples such as membrane proteins. It is hoped that the methods developed and applied here will at least partly form an important foundation upon which mass spectrometry can be further extended and fruitfully used to characterise the incredibly complex biological world.

Chapter Seven: Experimental Procedures*

7.1 General methods

Materials

All detergents (Sol-Grade) were purchase from Affymetrix. Lipids were purchased in powder form at 99 percent purity from Avanti Polar Lipids and stored at stock concentrations of 10 mg/mL in water. All other reagents were purchased from Sigma with exception of imidazole that was purchased from Acros Organics and protease inhibitor tablets from Roche. Water used was from a Millipore Milli-Q system.

Devices

Samples were buffer exchange prior to MS analysis using Micro Bio-Spin 6 columns (Bio-Rad). For protein concentration and buffer exchange Vivaspin 500 100 kDa or 50 kDa MWCO devices and Ultrafree-0.5 MWCO devices (Amicon) were used. Needles were prepared in-house by gold coating borosilicate glass capillaries (Harvard Apparatus) of 1.0 mm OD × 0.5 mm ID for measurements on Waters instruments, and filaments of 1.2 mm OD × 0.58 mm ID for measurements on the Q-Exactive instrument.[11] Superdex 200 10/300 GL and HisTrap HP 5 mL columns (GE Healthcare) were used for protein purification. Benchtop detergent screening was performed using Ni-NTA agarose (Sepharose CL-6B support, 50 percent suspension in 30 percent ethanol, pre-charged with Ni^{2+}, Qiagen).

*The experiment section was taken from publication as given in pages: xv-xvi

Membrane protein plasmids

Plasmids containing the genes of the membrane proteins expressed were provided by Arthur Laganowsky, Timothy Allison and Eamonn Reading with exception of OmpF that was provided by Nick Housden.[64,184,242] Prior to transformation DNA sequencing was performed using the Sanger sequencing service of Source Biosciences to confirm the integrity of the genes.

Transformation

Chemically competent cell lines BL21-Gold(DE3) (Bio Labs New England) and C43(DE3) (Lucigen) were used for protein over-expression. Cells were thawed from −80 °C on ice. In a pre-chilled 1.5 mL test tube 50 µL of cells were combined with 25 ng of plasmid DNA and incubated on ice for 30 min. Cells were heat shocked for 45 s at 42 °C and incubated on ice for an additional 2 min. Pre-warmed SOC medium was added and cells were allowed to grow for 1 h with shaking at 37 °C. 50 µL of cells were plated on LB-agar plates supplemented with either 100 µg/mL ampicillin or 50 µg/mL kanamycin depending on the antibiotic resistance harboured on the plasmid. Plates were incubated overnight at 37 °C and stored at 4 °C.

Culture media

LB media were prepared either from LB (Miller) powder (Fisher Scientific) or prepared by mixing 10 g NaCl, 10 g tryptone and 5 g yeast extract per 1 L of water. LB-agar plates were prepared by addition of 20 g agar to LB media. LB media were autoclaved prior to use. Antibiotic(s) were added immediately prior to pouring.

Preparation of lipids for membrane protein-lipid binding

Lipid powders were dissolved in a chloroform:methanol (50:50 (v/v)) solution to a final concentration of 10 mg/mL under a nitrogen atmosphere. Lipids were acidified to remove sodium adducts: the dissolved lipid mixture was transferred into a glass separator funnel and water added to reach 20 percent of the final volume (e.g. 1 mL water to 4 mL dissolved lipid mixture). Hydrochloric acid was added to the upper aqueous phase, dropwise, until the pH reached ~3.0 (at high pH the phosphate groups of the lipid will become protonated). The solutions were shaken and allowed to separate into phases. This process was repeated twice more. The lower organic phase containing lipid was then transferred to a glass vial.

To prepare the lipids for hydration, either non-acidified or acidified, the dissolved lipid samples were lyophilized by first drying under a nitrogen atmosphere until a small volume of solution remained and then drying overnight by placing under vacuum so that a lipid film forms. Saline (154mmol NaCl) was added in a minimum volume to suspend the lipid. This suspension was then frozen on dry ice. The frozen lipid 'cake' was placed under vacuum to lyophilise until dry. The freeze dried lipid was sealed and stored at −20 °C ready to be hydrated.

To hydrate the lipid an aqueous medium was added to the container of dry lipid and then agitated to form liposomes. The temperature of the hydrating medium was always above the gel-liquid crystal transition temperature (*Tc*) of the lipid before adding. Lipids were hydrated for approximately 1 h with vigorous shaking, mixing and stirring or until a clear suspension was obtained. Due to susceptibility to hydrolysis, causing lipids to

form their lyso lipid and free fatty acid derivatives, lipids were not stored in aqueous solution for more than 5–7 days.

Membrane proteins were incubated with lipids in a ratio of 1:2 protein-to-lipid.

7.2 Methods for Chapter Two

Small scale detergent screening of membrane proteins was performed following protocols that have been previously described.[64] The buffers used for the detergent screening were:

Extraction Buffer:	200 mM NaCl, 20 mM imidazole, 50 mM Tris pH 7.4, detergent solubilisation concentration (see Table 7.1)
Wash Buffer:	200 mM NaCl, 20 mM imidazole, 50 mM Tris pH 8.0, 2×CMC detergent.
Elution Buffer:	100 mM NaCl, 250 mM imidazole, 50 mM Tris pH 8.0, 2×CMC detergent.
MS Buffer:	200 mM ammonium acetate pH 7.4, 2×CMC detergent.

The detergent screening protocol used was as follows:

1. *Column preparation:* 400 µL of Ni-NTA agarose beads are added to a 1 mL Bio-Rad micro-centrifuge column. The columns were gravity packed or centrifugal force at 0.1 *g* for ~3 s. The waste was then collected into a Bio-Spin waste collector and disposed.

2. *Ni-bead pre-wash:* 4×500 µL aliquots of miliQ H_2O were run over the column either by gravity. Then 500 µL of elution buffer followed by 3×500 µL of wash buffer were run over and eluted.
3. *Equilibration of protein with Ni-NTA beads and desired detergent:* 300 µL of extraction buffer and 30 µL of solubilised His-tagged protein were added to the beads and mixed. The column was then capped and rolled. The components were left to incubate for 30 min. The solution was eluted and recycled over the column twice more for increased protein binding.
4. *Detergent wash:* 5×500 µL aliquots of wash buffer were washed over the column.
5. *Protein elution:* 2×250 µL of elution buffer was then added dropwise to the centre of the beads and the eluate collected.
6. *Protein concentration:* The eluted protein solution was concentrated using an Amicon centrifugal concentrator to 60–100 µL.
7. *MS preparation:* The concentrated protein was then buffer exchanged into the MS buffer using a Micro Bio-Spin 6 column. The sample can now be used for MS analysis.
8. *Column wash and storage:* 1×500 µL of elution buffer, 2×500 µL of H_2O and 2×500 µL of 20 percent ethanol. The columns were stored in 20 percent ethanol for further use as desired.

Protein expression, purification and preparation for mass spectrometry

Proteins were either cloned into a pET15b vector containing a TEV protease cleavable C-terminal fusion to superfolder green fluorescent protein (GFP) followed by

a 6× His-tag (vector 1) or into a pET15b vector containing a TEV protease cleavable N-terminal fusion to maltose binding protein (MBP) preceded by a secretion signal peptide (pelB) and a 10× His-tag (vector 2), as previously described.[184] With the exception of ELIC where the containing plasmid (vector 3) was a gift from I. Zimmermann and R. Dutzler[243] and contains a thrombin cleavable protease site.

Plasmids were transformed into either *E. coli* BL21 (DE3) Gold (Agilent) or *E. coli* OverExpress™ C43 (DE3) (Lucigen) cells. Proteins were expressed and purified as previously described,[64,184] in detergent and buffer containing either 10 percent or zero glycerol (Table 7.1). For MS analysis purified membrane proteins were buffer exchanged into 2×CMC of detergent and 200 mM ammonium acetate pH 8.0 (pH adjusted with ammonium hydroxide) using a centrifugal buffer exchange device (Micro Bio-Spin 6, Bio-Rad).

Table 7.1: Summary of membrane proteins used, abbreviations used within the dissertation, buffers and detergents used for both solubilisation and throughout purification of membrane proteins. The membrane proteins were exchanged into detergents of interest prior to MS analysis.

Membrane protein	**Abbr.**	**Vector**	**Glycerol / Detergent used for solubilisation**	**Buffer/ Detergent used for purification**
Ammonia transporter from *Escherichia coli*	AmtB	2	10 % Glycerol OG (240 mM)	10 % Glycerol DDM
Water channel from *Escherichia coli*	AqpZ	1	10 % Glycerol OG (240 mM)	10 % Glycerol DDM or OG
Formate transporter from *Escherichia coli*	FocA	1	10 % Glycerol NG (65 mM)	10 % Glycerol DDM
Arciflavine resistance protein from *Escherichia coli*	AcrB	1	0 % Glycerol DDM (20 mM)	0 % Glycerol DDM or TX-100

Membrane protein	Abbr.	Vector	Glycerol / Detergent used for solubilisation	Buffer/ Detergent used for purification
Mechanosensitive ion channel of small conductance from *Escherichia coli*	MscS	1	10 % Glycerol DDM (20 mM)	10 % Glycerol DDM
Multidrug resistance protein from *Escherichia coli*	MdtF	1	10 % Glycerol DDM (20 mM)	10 % Glycerol DDM
Glycerol conducting channel from *Escherichia coli*	GlpF	1	10 % Glycerol DDM (20 mM) or OG (240 mM)	10 % Glycerol DDM
Pentameric ligand gated ion channel from *Erwinia Chrysanthemi*	ELIC	3	10 % Glycerol DDM (20 mM)	10 % Glycerol DDM
Multidrug and toxic compound extrusion transporter from *Pyrococcus furiosus*	MATE	1	0 % Glycerol OGNG (17 mM)	0 % Glycerol DDM
Mechanosensitive ion channel of large conductance from *Mycobacterium tuberculosis*	MscL	1	0 % Glycerol TX-100 (40 mM) or OGNG (17 mM)	0 % Glycerol DDM or OGNG

Recombinant TTR containing a C-terminal His-tag was a gift from Mark Pepys. The proteins were buffer exchanged into 200 mM ammonium acetate pH 7.4 and either sprayed directly or supplemented with 2×CMC of DDM, LDAO C8E4 or E4 using a Micro Bio-Spin 6 column (Bio-Rad) at 4 °C. TTR concentration was 5.3 μM prior to buffer exchange.

Mass spectrometry

Mass spectrometer settings for a modified Q-TOF 2 mass spectrometer (Micromass, Manchester, UK) with a Z-spray source[74] were set to values for membrane proteins as previously described.[64] Typical instrument values were 5–7 μbar for source pressure, 1.5–1.8 kV for capillary voltage, 150–190 V for cone voltage, 1–10 V for extraction voltage, 180–200 V for collision voltage, argon for collision gas, and 0.2–0.3 MPa for collision gas pressure. Mass spectra were analysed using MassLynx v4.1 software (Waters) or PULSAR.[152]

Ion mobility mass spectrometry

Ion mobility measurements were performed on a modified Synapt G1 HDMS instrument with the travelling-wave ion mobility cell replaced by an 18 cm drift cell with radial RF ion confinement and a linear voltage gradient to direct ions along the axis of transmission to the time-of-flight (TOF) mass analyser.[54] The ion mobility mass spectrometer was typically set to a source pressure of 5–7 mBar, capillary voltage of 1.4–1.7 kV, capillary nanoflow of 0.03–0.2 mBar, argon as collision gas with flow rate set to 5–8 mL/min (~6.6–6.7 e^{-2} mBar).

For gas-phase unfolding experiments the collision voltage in the trap was typically varied from 50–240 V ramped at 5 V steps. The sample and extraction cone and trap bias voltages, quadrupole profile, and collision gas pressure were optimized for maximal ion intensity of the target membrane protein complex. Helium was the drift cell gas of choice set to a pressure of ~1.6–1.8 Torr (40 mL/min flow rate). Pusher time, pulse width and TDC inhibit were set to 180 μs, 7 μs and 7 μs, respectively.

Theoretical collision cross section calculations

The Projection Approximation (PA) calculation within MOBCAL[244,245] was used to calculate the CCS (Ω_{PA}) from respective crystal structure coordinates. Collision cross section values were corrected ($\Omega_{Corr'}$) by an empirical correction factor of 1.14 and corrected for missing residues between the protein construct (M_{Exp}) and crystal structure (M_{PDB}); assuming that residue addition will cause spherical grow in ion size[3,246]:

$$\Omega_{Corr'} = 1.14\ \Omega_{PA}\left(\frac{M_{Exp}}{M_{PDB}}\right)^{\frac{2}{3}}$$

Drop shape analysis

Surface tension measurements were performed using an IT Concepts tracker drop shape analysis instrument. Detergents were dissolved at 2×CMC in 200 mM ammonium acetate pH 7.4 and then underwent three freeze-thaw cycles before analysis. A static interfacial tension was determined by the pendent bubble method where a vapour drop is formed within a bulk liquid (water) creating a vapour-liquid interface.

By comparing the experimental drop shape to a theoretical Laplacian drop curvature,[247] *via* digital processing, and applying Tates law the surface tension γ is calculated;[248,249] with the drop shape being determined by the opposition between the surface tension and gravity:

$$\Delta P = \gamma\left(\frac{1}{R} + \frac{1}{R'}\right)$$

ΔP is the pressure difference across the vapour-liquid interface resultant of the interface curvature, R and R' are the radii of curvature of the interface.

$$2\pi r\gamma\, sin\theta = V(\rho_{water} - \rho_{air})g + \pi r^2 p$$

$$sin\theta = \frac{dz}{ds}$$

V the volume of the bubble under the plane of altitude, r is the capillary radius, g is the acceleration due to gravity, s is the bubble arc length, z is the bubble Cartesian coordinate, θ is the turning angle, p is the pressure per drop, ρ_w the density of water and ρ_l the density of the air (the density of water was used as the detergent concentrations used gave small changes in bulk liquid density).

Theoretical Rayleigh limit

The Rayleigh limit (Z_R) describes the maximum charge-per-droplet of a given size, where γ is the surface tension and R the radius of the droplet (assuming a density (ρ) of 1 g/cm^3)[24]:

$$Z_R e = 8\pi(\epsilon_0 \gamma R^3)^{\frac{1}{2}}$$

$$Z_R = 4\left(\frac{\pi\gamma\varepsilon_0}{\rho e^2 N_A}\right)^{0.5} . M^{0.5} = 0.078M^{0.5}$$

Charge state calculation

Charge state determination was performed using custom Python software developed in-house[184] using a mass spectra deconvolution approach.[250] All power law fits (z = a M^b) and statistics were carried out in Origin Pro 8.5.1., where z is the ion charge; either the maximum charge (z_R) or the average charge state (z_{ave}). M is the protein mass, b is the exponent scaling factor and a is an empirical offset constant. To investigate the

effect of collisional activation on detergent stripped membrane protein ions the relative difference in charge states were calculated ($z_{\text{Difference}}$):

$$z_{Difference} = \frac{z_{R/ave} - z_{R/ave}^{Initial}}{z_{R/ave}^{Initial}}$$

Detergent release

To quantify the degree of detergent removal as a function of collision energy two values were considered: the mass resolving power (R) and the average number (N) of detergent bound. The peak resolution was calculating by fitting a Gaussian function to each peak using the least squares algorithm. All fitting was performed within PULSAR. The resulted mass-to-charge ratio ($m/z = M$) full width at half maximum (ΔM) was then used to calculate the mass resolving power.

$$R = M_{center}/\Delta M$$

The average number of detergent bound was calculated by:

$$N_{detergent\ bound} = \frac{M_{exp} - M_{theor}}{M_{detergent}}$$

where the experimental mass-to-charge ratio (M_{exp}) was determined from the centre of the Gaussian function fitted to the peak.

Gas-phase unfolding

Gas-phase unfolding analysis and modelling was performed using the custom software PULSAR.[152]

7.3 Methods for Chapter Three

Protein purification and preparation for mass spectrometry

AmtB, AqpZ, *Pf*MATE were expressed and purified as described in the methods for chapter two. NapA D157N, OmpF, NTSR1 and ATPase proteins were kind donations from D. Drew, N. Housden, A. Pleuckthun and D. Stock, respectively.

Instrument settings were: Capillary voltage 1.5 V, cone voltage 130 V, extraction voltage 4 V, collision voltages in the trap ranging between 20 V and 200 V, and transfer collision voltage 50 V. Measurements were also performed in negative mode by switching the polarity from the tune page and applying 1.2 V on the capillary and 100 V on the cone. For high voltage measurements the instruction was modified to extend the cone voltage to 400 V by upgrading the electronic board and using a high voltage power supply.

The pressure was adjusted between 1.8 mbar to 7.5 mbar as measured by the built-in sensor on the rough pumping foreline and displayed in the MassLynx tune page. Adjustments were made by gradually opening or closing the Speedivalve (BOC Edwards, Crawley/Sussex, UK) that connects the first pumping stage between the sample cone and the extractor cone with the rotary pump, and in this manner controls the pumping efficiency in this region.

Ion mobility settings were: wave velocity 300 m/s with a wave height of 13 V (IMS cell), wave velocity of 248 m/s with a wave height of 13 V (transfer). Nitrogen was used as drift cell gas with a pressure of 1.6 Torr. Collision cross-section (CCS) calibrations were performed using pyruvate kinase, yeast alcohol dehydrogenase and

concanavalin A (Sigma) as calibrants. The 12+ charge state of NapA was found to have a CCS of 4,600 $Å^2$, suggesting a slightly compacted native-like conformation compared to the calculated value of 4,900 $Å^2$.

Spectra were analysed using MassLynx v4.1 software (Waters, Milford, MA). Peak intensities were extracted using mMass 3.9.5 and unfolding trajectory plots were generated by PULSAR.

Infra-red multiphoton dissociation. IRMPD had been implemented in an Ultima tandem quadrupole/time-of-flight (Q-ToF) mass spectrometer (Waters/Micromass, Manchester, UK). Initially the mass spectrometer was modified for high, up to 30,000 *m/z* values, as described elsewhere.[74] Additionally, the collision cell in the mass spectrometer had been transformed to a linear ion trap by applying time-dependent controllable trapping voltages to its entrance and exit electrodes.

An IR window (ZnSe for 10.6 μm or CaF_2 for 2.94 μm, Crystran, UK) and a gold-coated copper mirror (Laser Beam Products, UK) were introduced in the TOF part of the instrument. The laser beam was steered into the mass spectrometer by gold-coated copper mirrors and focused at the exit from the collision cell by a 50 cm ZnSe or CaF_2 lens (Crystran, UK).

Three lasers were employed for IRMPD in alternation: a far-IR (10.6 μm wavelength, 25 W) continuous CO_2 laser (Synrad, Mukilteo, WA), a far-IR pulsed (50–100 ns, 0.6 J/pulse) CO_2 laser (Lightmachinery, Canada), and a mid-IR (2.94 μm) pulsed Er:YAG laser (Spectrum, Germany). The latter laser has two regimes: a ‘spiky’ mode producing an array of 500 ns laser pulses within 60 μs envelope and 40 mJ of

total energy, or a short pulse mode (50 ns, 15 mJ). Laser power/pulse energy was measured before the turning mirror on the top of the TOF part of the mass spectrometer. During the alignment the laser power was also measured in the ion guide region, and losses of the laser radiation due to the small apertures of the ion track were estimated. The information on the maximum laser power/pulse energy and the radiation losses on the apertures of the mass spectrometer for all three lasers are given in Table 7.2. Operation of the trap was synchronised with operation of the lasers using a waveform generator (Agilent, Santa Clara, CA).

For activation by the continuous laser the trap was open for the ions to enter during all time of irradiation (*ca*. 100–250 ms). At the end of trapping/irradiation period the voltages at the entrance and exit electrodes of the collision cell were changed for 7 ms to the values that allowed ejection of the ions to the TOF detector for mass analysis. For pulsed laser activation the ions were stored in the trap for 93 ms (far-IR) or 493 ms (mid-IR), and by the end of this period a laser pulse was fired followed by the ejection of the ions to the TOF detector. An IR photodiode (VIGO, Poland) was used to monitor the laser pulse quality and the alignment of the beam. IRMPD mass spectra presented here are produced by averaging over a period of 25–100 scans, scan time 1–5 s.

CID experiments were carried out in the collision cell of the same mass spectrometer, or in a Q-ToF2 mass spectrometer (Waters/Micromass, Manchester, UK) modified to allow for the collision voltage of up to 400 V (200 V standard) as described elsewhere. Protein solutions were electrosprayed in 200 mM ammonium acetate, pH 7.5, at concentrations of 1–5 μM with capillary voltage 1.4–1.7 kV and sampling cone

voltage 100–200 V. Mass spectra were calibrated using CsI cluster ions (20 mg/mL solution).

Table 7.2: Parameters of IR laser radiation used for activation of protein complexes.

IR pulse duration /wavelength	**CW / 10.6 µm**	**100 ns / 10.6 µm**	**60 µs / 2.94 µm**	**50 ns / 2.94 µm**
Maximum IR energy at entrance to MS, E_{in}	5.6 J	0.55 J	40 mJ	15 mJ
Trapping Time, t	243 ms	93 ms	493 ms	493 ms
Maximum IR Power, P	23.2 W	5.5 MW	670 W	0.3 MW
IR energy at exit, E_{out} as % of E_{in}	50 %	20 %	*ca.* 100 %	*ca.* 100 %

7.4 Methods for Chapter Four:

Membrane protein and lipid incubation

Membrane proteins were purified as described in methods for chapter two, with the exception of OmpF and FvpA that were a kind donation from N. Housden and NavMS from B. Wallace. Lipids were prepared as described in the general methods.

Mass spectrometry

Mass spectrometry parameters were identical to parameters described in the methods for chapter three, in both positive and negative polarity.

High resolution non-denaturing MS

Experiments were performed using a Thermo Scientific™ Q-Exactive™ hybrid quadrupole-Orbitrap™ mass spectrometer modified for the transmission and detection of high *m/z* ions.[12] Hardware alterations include a lower RF frequency applied to the selection quadrupole and a gas line allowing higher pressures to be achieved in the HCD cell. The instrument was operated in "Native Mode" where both the RF frequencies applied to transfer optics and Orbitrap voltages were optimised for high mass species. Ions were generated in the positive ion mode from a static nanospray source using gold-coated capillaries prepared in-house, then passed through a temperature controlled transfer tube (40–60 °C), RF-lens, injection flatapole and bent flatapole. After traversing the selection quadrupole, which was operated with a wide selection window (2,000–20,000 *m/z*), ions were trapped in the HCD cell before being transferred into the C-trap and Orbitrap mass analyser for detection. Transient times were 64 ms unless otherwise stated and AGC target was 1×10^6. Removal of the micelle was achieved through increased voltages applied in the HCD. Argon was used as the collision gas and the pressure in the HCD cell was maintained at around 1×10^{-9} mbar unless otherwise stated. Spectra were acquired with 10 microscans, and averaged with a noise level parameter of 3. Data from 15–50 scans were averaged using Thermo Scientific™ Xcalibur™ 2.1. Calibration was performed prior to mass measurement up to 11,300 *m/z* using clusters of CsI.

Measurement of OBS1-OmpF dissociation constant

Measurement of the dissociation constant of OBS1-OmpF was performed as described previously.[251] Briefly OBS1 peptide (NH_2-^{2}SGGDGRGHNTGAHSTSG^{18}-

$CONH_2$) was diluted from a single stock to two-fold the desired concentration and mixed with equal volume of OmpF in 25 mM ammonium acetate of 1 percent (w/v) β-OG immediately prior to mass measurement. Conditions for nESI-MS were verified as to generate spectra of sufficient quality to obtain sharp peaks without incurring ligand dissociation. 200 V was applied in the HCD cell and 0 V to the S-lens. Spectra were deconvoluted using UniDec software.[252] The Data Collector module was used to extract and normalize intensities summed over all charge states. Extracted intensities for the unbound and bound states were fitted to K_D binding models using either a single fixed K_D for all binding sites or free K_Ds for each individual site. After fitting to both models, an F-test was used to determine the best model.

Lipid head group binding region analysis of membrane protein structures

The PDBTM[253] was exploited to obtain information about membrane protein amino acids exposed to lipid head groups. This database lists 2,387 transmembrane proteins (June 2015) featured in the PDB databank, providing information about their alignment in membrane. These include chain selections, transformation matrices and biomatrix operations needed to obtain a biological assembly aligned along the z axis (bilayer on the xy plane, z=0 being its centre). The database also provides information about half bilayer thickness, hereon zPDBTM.

The analysis used only proteins having more than one strand crossing the bilayer, which reduced the database to 2,187 entries. After having produced correctly aligned biological assemblies according to PDBTM, their amount of basic (LYS, ARG, HIS) and acidic (GLU, ASP) residues in contact with lipid heads were quantified. The following method was applied, for outer and inner leaflet independently:

1. Build a dense point mesh (1 Å spacing) around the protein's head group region, representing lipid head groups location. On x and y axis, the mesh is large enough to encompass all the protein atoms. For inner leaflet, z=[-zPDBTM-t, -zPDBTM], for outer leaflet z=[zPDBTM, zPDBTM +t], where t = 3 Å mesh thickness along z axis.
2. Remove from the mesh all points at less than c = 3 Å from the protein, where c is a clashing distance threshold.
3. Launch a DBSCAN[254] clustering algorithm on the remaining points. This algorithm considers contiguous points as part of the same cluster. Select only the points sharing the cluster with a mesh corner point: as a result, the selected points completely encircle the protein, without clashing with it. Thus, if the protein is a channel, points located in its central cavity are now removed.
4. Detect all protein atoms within d = 5 Å from the mesh, and compute their solvent accessible area (SASAall) using Shrake-Rupley ("rolling ball") algorithm. Between these atoms, identify the subset part of basic or acidic amino acids. Compute the SASA of both these subsets, and the resulting presence ratios ρacidic = SASAacidic / SASAall and ρbasic = SASAbasic / SASAall.

As a result from this protocol, every protein is characterized by four numbers, i.e. the presence ratio of basic and acidic amino acids on bottom and top bilayer. This protocol is controlled by three parameters, mesh thickness t, and mesh clash distance c and protein-mesh contact distance d. In order to assess the sensitivity of the protocol to these parameters, several runs with different parameters set were performed, namely [t,c,d]=[3,3,5], [5,3,5], [3,2.5,5], [3,3.5,5], [6,3.5,5], [6,3,5]. By computing mean and standard deviation of all resulting ρ scores, 6.8 percent of performed measures had a

standard deviation above 5, (0.5 percent above 10). These cases were inspected manually: the difference was always caused by a misdetection of internal cavity by one of the different methods. Parameters used in this work have been chosen selecting the set of parameters displaying the smallest deviation from mean.

Molecular dynamic simulation

The MemProtMD pipeline[255] was used to set up MARTINI coarse-grained simulations[256] of AmtB (PDB:1U7G) and OmpF (PDB:2ZFG) in membranes composed of equal quantities of POPC and POPG. In order to simulate protonation of the lipids, the charge on the phosphate beads of POPC and POPG were changed from −1 C to 0 C, resulting in a neutrally charged POPG and a positively charged POPC. Protonation of the protein was simulated by changing the charge of the sidechain beads of aspartic acid and glutamic acid from −1 C to 0 C, and by changing the charge of a single sidechain bead of histidine from 0 C to 1 C. For each molecule, 20 independent coarse-grained self-assembly simulations were set up, five simulating a deprotonated system, five with only lipids protonated, five with only proteins protonated and five with the entire system protonated. Simulations were run for a total of 1 μs each and the first 200 ns discarded to allow the bilayer to self-assemble.4,000 frames were taken from the last 800 nanoseconds of the simulation and lipid headgroups (NC3 bead of POPC and GLH bead of POPG) within 6 Angstroms of a protein sidechain bead were identified using MDAnalysis.[257] Percentage POPG residence time and POPC residence time were assigned to each residue. POPG preference for each residue was calculated by subtracting POPC residence time from POPG residence time.

Electrophysiology

Planar lipid bilayer recordings were carried out in 20 mM sodium acetate, 1 M KCl, pH 4, at 22.0 ± 1.5 °C. Briefly, a bilayer of 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine alone or with addition of 25 percent 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (Avanti Polar Lipids, Alabaster, AL, USA) was positioned across an aperture, 100 μm in diameter, in a Teflon film (Goodfellow) that separated the cis and trans compartments (1 mL each) of the apparatus. Current-voltage relation was measured at pH 4 from -100 mV to + 100 mV after the cis-addition of OmpF (1 μL, 4 μM) (*cis* at ground). The OmpF pore closure was measured at +100 mV and at –100 mV after the *cis* chamber had been perfused with 20 mM sodium acetate, 1 M KCl, pH 4, to remove excess protein. Ionic currents were measured by using Ag/AgCl electrodes with a patch-clamp amplifier (Axopatch 200B, Axon Instruments). The amplified signal was low-pass filtered at 2 kHz and sampled at 10 kHz with a Digidata 1440A digitizer (Axon Instruments). Data analysis was performed with pClamp software (Molecular Devices).

7.5 Methods for Chapter Five

CcO purification

The CcO complex was purified from bovine heart as described[258] and was a kind donation from S. Yoshikawa.

Protein and PTM identification

Proteins were separated by SDS-PAGE and digested as described.[259] Peptides were separated by nano-LC using the EASY nLC 1000 system (Thermo Scientific; mobile phase A, 0.1 percent (v/v) formic acid (FA)/5 percent (v/v) DMSO; mobile phase B, 100 percent (v/v) ACN/0.1 percent (v/v) FA/5 percent (v/v) DMSO). The peptides were loaded onto a trap column (5 mm, PepMap RSLC, C18, 300 µm I.D. particle size 3 µm; Thermo Scientific), separated with a flow rate of 200 nL/min and a gradient of 7–30 percent (v/v) mobile phase B over 30 min on an analytical C18 column (50 cm, PepMap RSLC, EASY-spray column, C18, 75 µm ID particle size 3 µm; Thermo Scientific) and directly eluted into a Q-Exactive Orbitrap mass spectrometer (Thermo Scientific). MS conditions were: spray voltage of 2.1 kV; capillary temperature of 320 °C; operation in data-dependent mode. MS spectra were acquired in the Orbitrap (*m/z* 350−1,500) with a resolution of 70,000 and AGC target at 3×10^6. The ten most intense ions were selected for HCD at an AGC target of 50,000.

Raw data were searched against Uniprot-Swissprot database (541,762 sequences) using Mascot v2.4.1 search engine (Matrix Science). Search parameters: Mass accuracy filter: 20 ppm for precursor ions; 1.0 Da for MS/MS fragment ions; Enzyme: Trypsin or Chymotrypsin with maximal two missed cleavage sites; Carbamidomethylation of cysteine and oxidation of methionine as well as phosphorylation of serine, threonine and tyrosine and acetylation of lysine as variable modifications.

Denaturing LC-MS

The proteins were separated by nano-LC using a DionexUltiMate 3000 RSLC nano System (Thermo Scientific); mobile phase A: 0.05 percent (v/v) trifluoroacetic acid (TFA), mobile phase B: 50 percent (v/v) acetonitrile/50 percent (v/v) isopropanol/0.04 percent (v/v) TFA). The proteins were dissolved in 0.05 percent (v/v) TFA and loaded onto a monolithic PS-DVB column (100 µm × 25 cm, Thermo Scientific). Protein subunits were separated with a flowrate of 600 nL/min and a gradient from 8 to 98 percent solvent B over 30 min. The proteins were directly eluted into a QSTAR XL mass spectrometer (ABSciex). MS conditions were: ion spray voltage 2,980 V, declustering potential 75 V and collision energy 17 V.

Mass spectrometry of intact CcO

Aliquots of CcO in *n*-decyl-β-D-maltopyranoside-containing buffer were exchanged against 200 mM ammonium acetate using Micro Bio-Spin 6 columns (Bio Rad). Spectra were acquired on a Q-ToF II or Synapt1 mass spectrometer (Waters) modified for high masses.[74] Optimized instrument parameters were as follows: capillary voltage 1.7 kV, cone voltage 190 V, extractor 5 V, source backing pressure 7–10 mbar and a collision cell pressure of 10 psi. Collision cell energy was 100–200 V. Spectra were processed and assigned using MassLynx software.

Detergent exchange

The detergent was exchanged as described previously in the methods for chapter two. Briefly, DM detergent was exchanged against C8E4, LDAO, TX-100, β-OG or foscholine-16 in 130 mM NaCl, 50 mM Tris pH 7.4 by size exclusion chromatography

using a Superdex 200 column (5/150; GE Healthcare) with a flow rate of 0.5 mL/min at 4 °C. Prior to MS analysis the buffer was exchanged against 200 mM ammonium acetate pH 7.4 supplemented with the respective detergents using the same chromatographic conditions.

Lipid identification

CcO proteins were digested with Trypsin at 37 °C. The peptide/lipid mixture was lyophilized and re-dissolved in 90 percent (v/v) methanol/0.05 percent (v/v) ammonia solution. For LC-MS/MS analysis, the peptide/lipid mixture was separated by nano-LC (DionexUltiMate 3000 RSLC nano System, Thermo Scientific; mobile phase A, 70 percent (v/v) methanol/25 mM ammonium acetate; mobile phase B, 100 percent (v/v) methanol/25 mM ammonium acetate). Peptides and lipids were loaded onto an analytical column (HPLC column Acclaim® PepMap 100, C8, 100 μm ID particle size 5 μm; Thermo Scientific) and separated with a flow rate of 300 nL/min and a gradient of 35–100 percent (v/v) mobile phase B over 21 min. Peptides and lipids directly eluted into an LTQ-Orbitrap XL hybrid mass spectrometer (Thermo Scientific).

MS conditions were: spray voltage of 1.5 kV and capillary temperature of 160 °C. The LTQ-Orbitrap XL was operated in negative ion mode and in data-dependent mode. Survey full scan MS spectra were acquired in the orbitrap (*m/z* 350−2000) with a resolution of 60,000 at *m/z* 400 and an automatic gain control (AGC) target at 5×10^5. The three most intense ions were selected for CID and HCD MS/MS fragmentation in the linear ion trap or the Orbitrap, respectively. Detection of previously selected precursors was excluded for 120 s. Ions with unrecognized charge state were excluded. CID fragmentation in the linear ion trap was performed at an AGC target of 3×10^4 and a

normalized collision energy of 38 percent at an activation of q = 0.25 and an activation time of 30 ms. HCD fragmentation was performed in the Orbitrap with a resolution of 7,500 at *m/z* 400 and a normalized collision energy of 50 percent at an activation time of 100 ms.

References

1 Thomson, J. J. Rays of positive electricity and thier application to chemical analysis. *Longmans Green, London* (1913).

2 Zhou, M., Morgner, N., Barrera, N. P., Politis, A., Isaacson, S. C., Matak-Vinkovic, D., Murata, T., Bernal, R. A., Stock, D. & Robinson, C. V. Mass spectrometry of intact V-type ATPases reveals bound lipids and the effects of nucleotide binding. *Science* **334**, 380-385 (2011).

3 Benesch, J. L. P. & Ruotolo, B. T. Mass spectrometry: come of age for structural and dynamical biology. *Current Opinion in Structural Biology* **21**, 641-649 (2011).

4 Katta, V. & Chait, B. T. Observation of the heme-globin complex in native myoglobin by electrospray-ionization mass spectrometry. *Journal of the American Chemical Society* **113**, 8534-8535 (1991).

5 Ganem, B., Li, Y. T. & Henion, J. D. Detection of noncovalent receptor-ligand complexes by mass spectrometry. *Journal of the American Chemical Society* **113**, 6294-6296 (1991).

6 McDaniel, E. W., Martin, D. W. & Barnes, W. S. Drift Tube-mass spectrometer for studies of low-energy ion-molecule reactions. *Review of Scientific Instruments* **33**, 2-7 (1962).

7 Lanucara, F., Holman, S. W., Gray, C. J. & Eyers, C. E. The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nature Chemistry* **6**, 281-294 (2014).

8 Stoeckli, M., Chaurand, P., Hallahan, D. E. & Caprioli, R. M. Imaging mass spectrometry: A new technology for the analysis of protein expression in mammalian tissues. *Nature Medicine* **7**, 493-496 (2001).

9 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).

10 Marcoux, J. & Robinson, C. V. Twenty years of gas phase structural biology. *Structure* **21**, 1541-1550 (2013).

11 Hernandez, H. & Robinson, C. V. Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nature Protocols* **2**, 715-726 (2007).

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

13 Munson, M. S. & Field, F.-H. Chemical ionization mass spectrometry. I. General introduction. *Journal of the American Chemical Society* **88**, 2621-2630 (1966).

14 Barber, M., Bordoli, R. S., Sedgwick, R. D. & Tyler, A. N. Fast atom bombardment of solids (FAB): A new ion source for mass spectrometry. *Journal of the Chemical Society, Chemical Communications*, 325-327 (1981).

15 Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64-71 (1989).

16 Dole, M., Mack, L., Hines, R., Mobley, R., Ferguson, L. & Alice, M. d. Molecular beams of macroions. *The Journal of Chemical Physics* **49**, 2240-2249 (1968).

17 Spengler, B., Kirsch, D., Kaufmann, R., Karas, M., Hillenkamp, F. & Giessmann, U. The detection of large molecules in matrix-assisted UV-laser desorption. *Rapid Communications in Mass Spectrometry* **4**, 301-305 (1990).

18 Madler, S., Boeri Erba, E. & Zenobi, R. MALDI-ToF mass spectrometry for studying noncovalent complexes of biomolecules. *Topics in Current Chemistry* **331**, 1-36 (2013).

19 Mager, F., Sokolova, L., Lintzel, J., Brutschy, B. & Nussberger, S. LILBID-mass spectrometry of the mitochondrial preprotein translocase TOM. *Journal of Physics: Condens Matter* **22**, 454132 (2010).

20 Takáts, Z., Wiseman, J. M., Gologan, B. & Cooks, R. G. Mass spectrometry sampling under ambient conditions with desorption electrospray ionization. *Science* **306**, 471-473 (2004).

21 Ferguson, C. N., Benchaar, S. A., Miao, Z., Loo, J. A. & Chen, H. Direct ionization of large proteins and protein complexes by desorption electrospray ionization-mass spectrometry. *Analytical Chemistry* **83**, 6468-6473 (2011).

22 Zenobi, R. & Knochenmuss, R. Ion formation in MALDI mass spectrometry. *Mass Spectrometry Reviews* **17**, 337-366 (1998).

23 Wong, S., Meng, C. & Fenn, J. Multiple charging in electrospray ionization of poly (ethylene glycols). *The Journal of Physical Chemistry* **92**, 546-550 (1988).

24 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).

25 Kebarle, P. & Tang, L. From ions in solution to ions in the gas phase-the mechanism of electrospray mass spectrometry. *Analytical Chemistry* **65**, 972A-986A (1993).

26 Fernández de La Mora, J. The fluid dynamics of Taylor cones. *Annual Review of Fluid Mechanics* **39**, 217-243 (2007).

27 Tang, K. & Smith, R. D. Physical/chemical separations in the break-up of highly charged droplets from electrosprays. *Journal of the American Society for Mass Spectrometry* **12**, 343-347 (2001).

28 Hoffmann, E. d. & Stroobant, V. *Mass Spectrometry: Principles and Applications*. 3rd edn, (John Wiley and Sons Ltd, 2007).

29 Yost, R. & Enke, C. Triple quadrupole mass spectrometry for direct mixture analysis and structure elucidation. *Analytical Chemistry* **51**, 1251-1264 (1979).

30 Dawson, P. H. Quadrupole mass analyzers: Performance, design and some recent applications. *Mass Spectrometry Reviews* **5**, 1-37 (1986).

31 Henchman, M. & Steel, C. Understanding the quadrupole mass filter through computer simulation. *Journal of Chemical Education* **75**, 1049 (1998).

32 Benesch, J. L., Ruotolo, B. T., Simmons, D. A. & Robinson, C. V. Protein complexes in the gas phase: technology for structural genomics and proteomics. *Chemical Reviews* **107**, 3544-3567 (2007).

33 Boyle, J. G., Whitehouse, C. M., Fenn, J. B. & Cotter, R. J. An ion-storage time-of-flight mass spectrometer for analysis of electrospray ions. *Rapid Communications in Mass Spectrometry* **5**, 400-405 (1991).

34 Doroshenko, V. M. & Cotter, R. J. Ideal velocity focusing in a reflectron time-of-flight mass spectrometer. *Journal of the American Society for Mass Spectrometry* **10**, 992-999 (1999).

35 Benesch, J. L., Ruotolo, B. T., Sobott, F., Wildgoose, J., Gilbert, A., Bateman, R. & Robinson, C. V. Quadrupole-time-of-flight mass spectrometer modified for higher-energy dissociation reduces protein assemblies to peptide fragments. *Analytical Chemistry* **81**, 1270-1274 (2009).

36 Mamyrin, B., Karataev, V., Shmikk, D. & Zagulin, V. The massreflect ron, a new non-magnetic time-of-flight mass spectrometer with high resolution. *Journal of Experimental and Theoretical Physics* **64**, 82-89 (1973).

37 Dawson, J. & Guilhaus, M. Orthogonal-acceleration time-of-flight mass spectrometer. *Rapid Communications in Mass Spectrometry* **3**, 155-159 (1989).

38 Ferguson, P. L. & Smith, R. D. Proteome Analysis by Mass Spectrometry. *Annual Review of Biophysics and Biomolecular Structure* **32**, 399-424 (2003).

39 Baykut, G. & Eyler, J. Fourier transform ion cyclotron resonance mass spectrometry. *TRAC Trends in Analytical Chemistry* **5**, 44-49 (1986).

40 Hu, Q., Noll, R. J., Li, H., Makarov, A., Hardman, M. & Graham Cooks, R. The Orbitrap: a new mass spectrometer. *Journal of Mass Spectrometry* **40**, 430-443 (2005).

41 Makarov, A. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Analytical Chemistry* **72**, 1156-1162 (2000).

42 Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198-207 (2003).

43 Fraser, G. W. The ion detection efficiency of microchannel plates (MCPs). *International Journal of Mass Spectrometry* **215**, 13-30 (2002).

44 Westmacott, G., Frank, M., Labov, S. E. & Benner, W. H. Using a superconducting tunnel junction detector to measure the secondary electron emission efficiency for a microchannel plate detector bombarded by large molecular ions. *Rapid Communications in Mass Spectrometry* **14**, 1854-1861 (2000).

45 De Hoffmann, E. & Stroobant, V. Mass spectrometry: principles and applications. 2007. *Iohn Wiley* **8**.

46 Ruotolo, B. T., Benesch, J. L., Sandercock, A. M., Hyung, S. J. & Robinson, C. V. Ion mobility-mass spectrometry analysis of large protein complexes. *Nature Protocols* **3**, 1139-1152 (2008).

47 Pagel, K. & Harvey, D. J. Ion mobility–mass spectrometry of complex carbohydrates: collision cross sections of sodiated N-linked glycans. *Analytical Chemistry* **85**, 5138-5145 (2013).

48 Ruotolo, B. T., Hyung, S. J., Robinson, P. M., Giles, K., Bateman, R. H. & Robinson, C. V. Ion mobility-mass spectrometry reveals long-lived, unfolded intermediates in the dissociation of protein complexes. *Angewandte Chemie International Edition in English* **46**, 8001-8004 (2007).

49 McDaniel, E., Martin, D. & Barnes, W. Drift tube-mass spectrometer for studies of low-energy ion-molecule reactions. *Review of Scientific Instruments* **33**, 2-7 (1962).

50 Mehmood, S., Marcoux, J., Hopper, J. T. S., Allison, T. M., Liko, I., Borysik, A. J. & Robinson, C. V. Charge Reduction Stabilizes Intact Membrane Protein Complexes for Mass Spectrometry. *Journal of the American Chemical Society* **136**, 17010-17012 (2014).

51 Wyttenbach, T., Kemper, P. R. & Bowers, M. T. Design of a new electrospray ion mobility mass spectrometer. *International Journal of Mass Spectrometry* **212**, 13-23 (2001).

52 Mason, E. A. & Chapman, S. Motion of small suspended particles in nonuniform gases. *The Journal of Chemical Physics* **36**, 627-632 (1962).

53 Creaser, C. S., Griffiths, J. R., Bramwell, C. J., Noreen, S., Hill, C. A. & Thomas, C. P. Ion mobility spectrometry: A review. Part 1. Structural analysis by mobility measurement. *Analyst* **129**, 984-994 (2004).

54 Bush, M. F., Hall, Z., Giles, K., Hoyes, J., Robinson, C. V. & Ruotolo, B. T. Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology. *Analytical Chemistry* **82**, 9557-9565 (2010).

55 Giles, K., Pringle, S. D., Worthington, K. R., Little, D., Wildgoose, J. L. & Bateman, R. H. Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid Communications in Mass Spectrometry* **18**, 2401-2414 (2004).

56 Esquenazi, E., Daly, M., Bahrainwala, T., Gerwick, W. H. & Dorrestein, P. C. Ion mobility mass spectrometry enables the efficient detection and identification of halogenated natural products from cyanobacteria with minimal sample preparation. *Bioorganic & Medicinal Chemistry* **19**, 6639-6644 (2011).

57 Shliaha, P. V., Bond, N. J., Gatto, L. & Lilley, K. S. Effects of traveling wave ion mobility separation on data independent acquisition in proteomics studies. *Journal of proteome research* **12**, 2323-2339 (2013).

58 Saba, J., Bonneil, E., Pomies, C., Eng, K. & Thibault, P. Enhanced sensitivity in proteomics experiments using FAIMS coupled with a hybrid linear ion trap/orbitrap mass spectrometer. *Journal of proteome research* **8**, 3355-3366 (2009).

59 Wyttenbach, T., von Helden, G., Batka, J. J., Carlat, D. & Bowers, M. T. Effect of the long-range potential on ion mobility measurements. *Journal of the American Society for Mass Spectrometry* **8**, 275-282 (1997).

60 Shvartsburg, A. A. & Jarrold, M. F. An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chemical Physics Letters* **261**, 86-91 (1996).

61 Shvartsburg, A. A., Schatz, G. C. & Jarrold, M. F. Mobilities of carbon cluster ions: critical importance of the molecular attractive potential. *The Journal of chemical physics* **108**, 2416-2423 (1998).

62 Bush, M. F., Campuzano, I. D. & Robinson, C. V. Ion mobility mass spectrometry of peptide ions: effects of drift gas and calibration strategies. *Analytical Chemistry* **84**, 7124-7130 (2012).

63 Bleiholder, C., Wyttenbach, T. & Bowers, M. T. A novel projection approximation algorithm for the fast and accurate computation of molecular

collision cross sections (I). Method. *International Journal of Mass Spectrometry* **308**, 1-10 (2011).

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

65 Shepherd, D. A., Marty, M. T., Giles, K., Baldwin, A. J. & Benesch, J. L. Combining tandem mass spectrometry with ion mobility separation to determine the architecture of polydisperse proteins. *International Journal of Mass Spectrometry* **377**, 663-671 (2015).

66 Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. & Kuster, B. Quantitative mass spectrometry in proteomics: a critical review. *Analytical and Bioanalytical Chemistry* **389**, 1017-1031 (2007).

67 Hilton, G. R. & Benesch, J. L. 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).

68 Loo, J. A. Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrometry Reviews* **16**, 1-23 (1997).

69 Ganem, B., Li, Y. T. & Henion, J. D. Observation of noncovalent enzyme-substrate and enzyme-product complexes by ion-spray mass spectrometry. *Journal of the American Chemical Society* **113**, 7818-7819 (1991).

70 Robinson, C. V., Chung, E. W., Kragelund, B. B., Knudsen, J., Aplin, R. T., Poulsen, F. M. & Dobson, C. M. Probing the nature of noncovalent interactions by mass spectrometry. A study of protein-CoA ligand binding and assembly. *Journal of the American Chemical Society* **118**, 8646-8653 (1996).

71 Sobott, F. & Robinson, C. V. Protein complexes gain momentum. *Current Opinion in Structural Biology* **12**, 729-734 (2002).

72 Sharon, M., & Robinson, C.V. The role of mass spectrometry in structure elucidation of dynamic protein complexes. *Annual Reviews* **76**,167-193 (2007)

73 Wilm, M. & Mann, M. Analytical properties of the nanoelectrospray ion source. *Analytical Chemistry* **68**, 1-8 (1996).

74 Sobott, F., Hernandez, H., McCammon, M. G., Tito, M. A. & Robinson, C. V. A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Analytical Chemistry* **74**, 1402-1407 (2002).

75 Rostom, A. A. & Robinson, C. V. Detection of the intact GroEL chaperonin assembly by mass spectrometry. *Journal of the American Chemical Society* **121**, 4718-4719 (1999).

76 Chernushevich, I. V. & Thomson, B. A. Collisional cooling of large ions in electrospray mass spectrometry. *Analytical Chemistry* **76**, 1754-1760 (2004).

77 Labastie, P. & Doy, M. A high mass range quadrupole spectrometer for cluster studies. *International Journal of Mass Spectrometry and Ion Processes* **91**, 105-112 (1989).

78 Heck, A. J. R. & van den Heuvel, R. H. H. Investigation of intact protein complexes by mass spectrometry. *Mass Spectrometry Reviews* **23**, 368-389 (2004).

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

80 Baca, M. & Kent, S. B. Direct observation of a ternary complex between the dimeric enzyme HIV-1 protease and a substrate-based inhibitor. *Journal of the American Chemical Society* **114**, 3992-3993 (1992).

81 Jones, J. L., Dongre, A. R., Somogyi, A. & Wysocki, V. H. Sequence dependence of peptide fragmentation efficiency curves determined by electrospray ionization/surface-induced dissociation mass spectrometry. *Journal of the American Chemical Society* **116**, 8368-8369 (1994).

82 Zubarev, R. A., Kelleher, N. L. & McLafferty, F. W. Electron capture dissociation of multiply charged protein cations. A nonergodic process. *Journal of the American Chemical Society* **120**, 3265-3266 (1998).

83 Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J. & Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy* **101**, 9528-9533 (2004).

84 Price, W. D., Schnier, P. D. & Williams, E. R. Tandem mass spectrometry of large biomolecule ions by blackbody infrared radiative dissociation. *Analytical Chemistry* **68**, 859-866 (1996).

85 Woodin, R., Bomse, D. & Beauchamp, J. Multiphoton dissociation of molecules with low power continuous wave infrared laser radiation. *Journal of the American Chemical Society* **100**, 3248-3250 (1978).

86 Bowers, M. T., Marshall, A. G. & McLafferty, F. W. Mass spectrometry: recent advances and future directions. *The Journal of Physical Chemistry* **100**, 12897-12910 (1996).

87 Brodbelt, J. S. Analytical applications of ion-molecule reactions. *Mass Spectrometry Reviews* **16**, 91-110 (1997).

88 Winger, B. E., Hofstadler, S. A., Bruce, J. E., Udseth, H. R. & Smith, R. D. High-resolution accurate mass measurements of biomolecules using a new

electrospray ionization ion cyclotron resonance mass spectrometer. *Journal of the American Society for Mass Spectrometry* **4**, 566-577 (1993).

89 Hall, Z., Politis, A., Bush, M. F., Smith, L. J. & Robinson, C. V. Charge-state dependent compaction and dissociation of protein complexes: insights from ion mobility and molecular dynamics. *Journal of the American Chemical Society* **134**, 3429-3438 (2012).

90 Zhou, M., Sandercock, A. M., Fraser, C. S., Ridlova, G., Stephens, E., Schenauer, M. R., Yokoi-Fong, T., Barsky, D., Leary, J. A. & Hershey, J. W. Mass spectrometry reveals modularity and a complete subunit interaction map of the eukaryotic translation factor eIF3. *Proceedings of the National Academy of Sciences* **105**, 18139-18144 (2008).

91 Baldwin, A. J., Lioe, H., Hilton, G. R., Baker, L. A., Rubinstein, J. L., Kay, L. E. & Benesch, J. L. The polydispersity of alphaB-crystallin is rationalized by an interconverting polyhedral architecture. *Structure* **19**, 1855-1863 (2011).

92 Benesch, J. L., Aquilina, J. A., Baldwin, A. J., Rekas, A., Stengel, F., Lindner, R. A., Basha, E., Devlin, G. L., Horwitz, J., Vierling, E., Carver, J. A. & Robinson, C. V. The quaternary organization and dynamics of the molecular chaperone HSP26 are thermally regulated. *Chemistry & Biology* **17**, 1008-1017 (2010).

93 Liu, L., Bagal, D., Kitova, E. N., Schnier, P. D. & Klassen, J. S. Hydrophobic Protein−Ligand Interactions Preserved in the Gas Phase. *Journal of the American Chemical Society* **131**, 15980-15981 (2009).

94 Sharon, M. & Robinson, C. V. The role of mass spectrometry in structure elucidation of dynamic protein complexes. *Annual Review of Biochemistry* **76**, 167-193 (2007).

95 Aquilina, J. A., Benesch, J. L., Ding, L. L., Yaron, O., Horwitz, J. & Robinson, C. V. Subunit exchange of polydisperse proteins: mass spectrometry reveals consequences of alphaA-crystallin truncation. *Journal of Biological Chemistry* **280**, 14485-14491 (2005).

96 Sims, J. J., Haririnia, A., Dickinson, B. C., Fushman, D. & Cohen, R. E. Avid interactions underlie the Lys63-linked polyubiquitin binding specificities observed for UBA domains. *Nature Structural & Molecular Biology* **16**, 883-889 (2009).

97 Ruotolo, B. T., Giles, K., Campuzano, I., Sandercock, A. M., Bateman, R. H. & Robinson, C. V. Evidence for macromolecular protein rings in the absence of bulk water. *Science* **310**, 1658-1661 (2005).

98 Dyachenko, A., Gruber, R., Shimon, L., Horovitz, A. & Sharon, M. Allosteric mechanisms can be distinguished using structural mass spectrometry. *Proceedings of the National Academy of Sciences* **110**, 7235-7239 (2013).

99 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).

100 Ebong, I. O., Morgner, N., Zhou, M., Saraiva, M. A., Daturpalli, S., Jackson, S. E. & Robinson, C. V. Heterogeneity and dynamics in the assembly of the heat shock protein 90 chaperone complexes. *Proceedings of the National Academy of Sciences* **108**, 17939-17944 (2011).

101 Siuzdak, G., Bothner, B., Yeager, M., Brugidou, C., Fauquet, C. M., Hoey, K. & Change, C.-M. Mass spectrometry and viral analysis. *Chemistry & Biology* **3**, 45-48 (1996).

102 Ouyang, Z., Takáts, Z., Blake, T. A., Gologan, B., Guymon, A. J., Wiseman, J. M., Oliver, J. C., Davisson, V. J. & Cooks, R. G. Preparing protein microarrays by soft-landing of mass-selected ions. *Science* **301**, 1351-1354 (2003).

103 Hopper, J. T. & 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. *Journal of the American Society for Mass Spectrometry* **20**, 1851-1858 (2009).

104 Zhou, M., Politis, A., Davies, R. B., Liko, I., Wu, K. J., Stewart, A. G., Stock, D. & Robinson, C. V. Ion mobility-mass spectrometry of a rotary ATPase reveals ATP-induced reduction in conformational flexibility. *Nature Chemistry* **6**, 208-215 (2014).

105 Zhong, Y., Han, L. & Ruotolo, B. T. Collisional and coulombic unfolding of gas-phase proteins: high correlation to their domain structures in solution. *Angewandte Chemie International Edition in English* **53**, 9209-9212 (2014).

106 Lingwood, D. & Simons, K. Lipid rafts as a membrane-organizing principle. *Science* **327**, 46-50 (2010).

107 Wallin, E. & Von Heijne, G. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Science* **7**, 1029 (1998).

108 von Heijne, G. Membrane-protein topology. *Nature Reviews: Molecular Cell Biology* **7**, 909-918 (2006).

109 Wagner, S., Bader, M. L., Drew, D. & de Gier, J.-W. Rationalizing membrane protein overexpression. *Trends in Biotechnology* **24**, 364-371 (2006).

110 Zilch, L. W., Kaleta, D. T., Kohtani, M., Krishnan, R. & Jarrold, M. F. Folding and unfolding of helix-turn-helix motifs in the gas phase. *Journal of the American Society for Mass Spectrometry* **18**, 1239-1248 (2007).

111 Loo, R. R., Dales, N. & Andrews, P. C. Surfactant effects on protein structure examined by electrospray ionization mass spectrometry. *Protein Science* **3**, 1975-1983 (1994).

112 Demmers, J. A., Haverkamp, J., Heck, A. J., Koeppe, R. E. & Killian, J. A. Electrospray ionization mass spectrometry as a tool to analyze hydrogen/deuterium exchange kinetics of transmembrane peptides in lipid bilayers. *Proceedings of the National Academy of Sciences* **97**, 3189-3194 (2000).

113 Demmers, J. A., van Dalen, A., de Kruijff, B., Heck, A. J. & Killian, J. A. Interaction of the K+ channel KcsA with membrane phospholipids as studied by ESI mass spectrometry. *FEBS Letters* **541**, 28-32 (2003).

114 Ilag, L. L., Ubarretxena-Belandia, I., Tate, C. G. & Robinson, C. V. Drug binding revealed by tandem mass spectrometry of a protein-micelle complex. *Journal of the American Chemical Society* **126**, 14362-14363 (2004).

115 Ilag, L. L., Ubarretxena-Belandia, I., Tate, C. G. & Robinson, C. V. Drug binding revealed by tandem mass spectrometry of a protein-micelle complex. *Journal of the American Chemical Society* **126**, 14362-14363 (2004).

116 Borysik, A. J., Hewitt, D. J. & Robinson, C. V. Detergent release prolongs the lifetime of native-like membrane protein conformations in the gas-phase. *Journal of the American Chemical Society* **135**, 6078-6083 (2013).

117 Marcoux, J. & Robinson, C. V. Twenty years of gas phase structural biology. *Structure* **21**, 1541-1550 (2013).

118 Drew, D., Lerch, M., Kunji, E., Slotboom, D. J. & de Gier, J. W. Optimization of membrane protein overexpression and purification using GFP fusions. *Nature Methods* **3**, 303-313 (2006).

119 Drew, D., Newstead, S., Sonoda, Y., Kim, H., von Heijne, G. & Iwata, S. GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in Saccharomyces cerevisiae. *Nature Protocols* **3**, 784-798 (2008).

120 Hopper, J. T., Yu, Y. T., Li, D., Raymond, A., Bostock, M., Liko, I., Mikhailov, V., Laganowsky, A., Benesch, J. L., Caffrey, M., Nietlispach, D. & Robinson, C. V. Detergent-free mass spectrometry of membrane protein complexes. *Nature Methods* **10**, 1206-1208 (2013).

121 Marcoux, J., Wang, S. C., Politis, A., Reading, E., Ma, J., Biggin, P. C., Zhou, M., Tao, H., Zhang, Q., Chang, G., Morgner, N. & Robinson, C. V. Mass spectrometry reveals synergistic effects of nucleotides, lipids, and drugs binding to a multidrug resistance efflux pump. *Proceedings of the National Academy of Sciences* **110**, 9704-9709 (2013).

122 Schmidt, C., Zhou, M., Marriott, H., Morgner, N., Politis, A. & Robinson, C. V. Comparative cross-linking and mass spectrometry of an intact F-type ATPase suggest a role for phosphorylation. *Nature Communications* **4**, 1985 (2013).

123 Boeri Erba, E., Ruotolo, B. T., Barsky, D. & Robinson, C. V. Ion mobility-mass spectrometry reveals the influence of subunit packing and charge on the dissociation of multiprotein complexes. *Analytical Chemistry* **82**, 9702-9710 (2010).

124 Iribarne, J. V. & Thomson, B. A. On the evaporation of small ions from charged droplets. *The Journal of Chemical Physics* **64**, 2287-2294 (1976).

125 Konermann, L., Ahadi, E., Rodriguez, A. D. & Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Analytical Chemistry* **85**, 2-9 (2012).

126 Gamero-Castano, M. & de la Mora, J. F. Mechanisms of electrospray ionization of singly and multiply charged salt clusters. *Analytica Chimica Acta* **406**, 67-91 (2000).

127 Fenn, J. B. Ion formation from charged droplets- roles of geometry, energy, and time. *Journal of the American Society for Mass Spectrometry* **4**, 524-535 (1993).

128 Iavarone, A. T. & Williams, E. R. Mechanism of charging and supercharging molecules in electrospray ionization. *Journal of the American Chemical Society* **125**, 2319-2327 (2003).

129 Taflin, D. C., Ward, T. L. & Davis, E. J. Electrified droplet fission and the Rayleigh limit. *Langmuir* **5**, 376-384 (1989).

130 Felitsyn, N., Peschke, M. & Kebarle, P. Origin and number of charges observed on multiply-protonated native proteins produced by ESI. *International Journal of Mass Spectrometry* **219**, 39-62 (2002).

131 Kaltashov, I. A. & Mohimen, A. Estimates of protein surface areas in solution by electrospray ionization mass spectrometry. *Analytical Chemistry* **77**, 5370-5379 (2005).

132 Hogan, C. J., Jr., Carroll, J. A., Rohrs, H. W., Biswas, P. & Gross, M. L. Combined charged residue-field emission model of macromolecular electrospray ionization. *Analytical Chemistry* **81**, 369-377 (2009).

133 Allen, S. J., Schwartz, A. M. & Bush, M. F. Effects of polarity on the structures and charge states of native-like proteins and protein complexes in the gas phase. *Analytical Chemistry* **85**, 12055-12061 (2013).

134 Barrera, N. P., Isaacson, S. C., Zhou, M., Bavro, V. N., Welch, A., Schaedler, T. A., Seeger, M. A., Miguel, R. N., Korkhov, V. M., van Veen, H. W., Venter, H., Walmsley, A. R., Tate, C. G. & Robinson, C. V. Mass spectrometry of

membrane transporters reveals subunit stoichiometry and interactions. *Nature Methods* **6**, 585-587 (2009).

135 Privé, G. G. Detergents for the stabilization and crystallization of membrane proteins. *Methods* **41**, 388-397 (2007).

136 Helenius, A. & Simons, K. Solubilization of membranes by detergents. *Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes* **415**, 29-79 (1975).

137 Bechara, C., Noll, A., Morgner, N., Degiacomi, M. T., Tampe, R. & Robinson, C. V. A subset of annular lipids is linked to the flippase activity of an ABC transporter. *Nature Chemistry* **108**, 255-262 (2015).

138 Contreras, F.-X., Ernst, A. M., Wieland, F. & Brügger, B. Specificity of intramembrane protein–lipid interactions. *Cold Spring Harbor perspectives in biology* **3**, a004705 (2011).

139 Lichtenberg, D. Characterization of the solubilization of lipid bilayers by surfactants. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **821**, 470-478 (1985).

140 Michel, H. in *International Tables for Crystallography Volume F: Crystallography of biological macromolecules* 94-99 (Springer, 2001).

141 Morgner, N., Montenegro, F., Barrera, N. P. & Robinson, C. V. Mass spectrometry-from peripheral proteins to membrane motors. *Journal of Molecular Biology* **423**, 1-13 (2012).

142 Friemann, R., Larsson, D. S., Wang, Y. & van der Spoel, D. Molecular dynamics simulations of a membrane protein− micelle complex in vacuo. *Journal of the American Chemical Society* **131**, 16606-16607 (2009).

143 Loo, R. R. O., Lakshmanan, R. & Loo, J. A. What protein charging (and supercharging) reveal about the mechanism of electrospray ionization. *Journal of the American Society for Mass Spectrometry* **25**, 1675-1693 (2014).

144 McKay, A. R., Ruotolo, B. T., Ilag, L. L. & Robinson, C. V. Mass measurements of increased accuracy resolve heterogeneous populations of intact ribosomes. *Journal of the American Chemical Society* **128**, 11433-11442 (2006).

145 Korepanova, A. & Matayoshi, E. D. HPLC-SEC characterization of membrane protein-detergent complexes. *Current Protocols in Protein Science*, 29.25. 21-29.25. 12 (2012).

146 Borysik, A. J. & Robinson, C. V. Formation and dissociation processes of gas-phase detergent micelles. *Langmuir* **28**, 7160-7167 (2012).

147 Griffith, W. P. & Kaltashov, I. A. Highly asymmetric interactions between globin chains during hemoglobin assembly revealed by electrospray ionization mass spectrometry. *Biochemistry* **42**, 10024-10033 (2003).

148 Simmons, D. A., Wilson, D. J., Lajoie, G. A., Doherty-Kirby, A. & Konermann, L. Subunit disassembly and unfolding kinetics of hemoglobin studied by time-resolved electrospray mass spectrometry. *Biochemistry* **43**, 14792-14801 (2004).

149 Benesch, J. L. Collisional activation of protein complexes: picking up the pieces. *Journal of the American Society for Mass Spectrometry* **20**, 341-348 (2009).

150 Ruotolo, B. T., Hyung, S. J., Robinson, P. M., Giles, K., Bateman, R. H. & Robinson, C. V. Ion Mobility-mass spectrometry reveals long-lived, unfolded intermediates in the dissociation of protein complexes. *Angewandte Chemie International Edition* **46**, 8001-8004 (2007).

151 Beveridge, R., Covill, S., Pacholarz, K. J., Kalapothakis, J. M. D., MacPhee, C. E. & Barran, P. E. A Mass-spectrometry-based framework to define the extent of disorder in proteins. *Analytical Chemistry* **86**, 10979-10991 (2014).

152 Allison, T. M., Reading, E., Liko, I., Baldwin, A. J., Laganowsky, A. & Robinson, C. V. Quantifying the stabilizing effects of protein-ligand interactions in the gas phase. *Nature Communications* **6** (2015).

153 Morgner, N., Montenegro, F., Barrera, N. P. & Robinson, C. V. Mass spectrometry-from peripheral proteins to membrane motors. *Journal of Molecular Biology* **423**, 1-13 (2012).

154 Pattnaik, P. Surface plasmon resonance. *Applied Biochemistry and Biotechnology* **126**, 79-92 (2005).

155 Hyung, S. J., Robinson, C. V. & Ruotolo, B. T. Gas-phase unfolding and disassembly reveals stability differences in ligand-bound multiprotein complexes. *Chemistry & Biology* **16**, 382-390 (2009).

156 Niu, S., Rabuck, J. N. & Ruotolo, B. T. Ion mobility-mass spectrometry of intact protein–ligand complexes for pharmaceutical drug discovery and development. *Current Opinion in Chemical Biology* **17**, 809-817 (2013).

157 Hopper, J. T. & 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. *Journal of the American Society for Mass Spectrometry* **20**, 1851-1858 (2009).

158 Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W. & Gygi, S. P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proceedings of the National Academy of Sciences* **100**, 6940-6945 (2003).

159 Hoffmann, E. *Mass spectrometry*. (Wiley Online Library, 1996).

160 McLuckey, S. A. & Goeringer, D. E. Slow heating methods in tandem mass spectrometry. *Journal of Mass Spectrometry* **32**, 461-474 (1997).

161 Hart, K. J. & McLuckey, S. A. Relative dissociation energy measurements using ion trap collisional activation. *Journal of the American Society for Mass Spectrometry* **5**, 250-259 (1994).

162 Benesch, J. L., Ruotolo, B. T., Simmons, D. A. & Robinson, C. V. Protein complexes in the gas phase: technology for structural genomics and proteomics. *Chemical Reviews* **107**, 3544-3567 (2007).

163 Wysocki, V. H., Joyce, K. E., Jones, C. M. & Beardsley, R. L. Surface-induced dissociation of small molecules, peptides, and non-covalent protein complexes. *Journal of the American Society for Mass Spectrometry* **19**, 190-208 (2008).

164 Light-Wahl, K. J., Schwartz, B. L. & Smith, R. D. Observation of the noncovalent quaternary associations of proteins by electrospray ionization mass spectrometry. *Journal of the American Chemical Society* **116**, 5271-5278 (1994).

165 Winger, B. E., Light-Wahl, K. J., Ogorzalek Loo, R. R., Udseth, H. R. & Smith, R. D. Observation and implications of high mass-to-charge ratio ions from electrospray ionization mass spectrometry. *Journal of the American Society for Mass Spectrometry* **4**, 536-545 (1993).

166 Laskin, J., Denisov, E. & Futrell, J. A comparative study of collision-induced and surface-induced dissociation. 1. Fragmentation of protonated dialanine. *Journal of the American Chemical Society* **122**, 9703-9714 (2000).

167 Ruotolo, B. T. & Robinson, C. V. Aspects of native proteins are retained in vacuum. *Current Opinion in Chemical Biology* **10**, 402-408 (2006).

168 Shukla, A. K. & Futrell, J. H. Tandem mass spectrometry: dissociation of ions by collisional activation. *Journal of Mass Spectrometry* **35**, 1069-1090 (2000).

169 McLuckey, S. A. Principles of collisional activation in analytical mass spectrometry. *Journal of the American Society for Mass Spectrometry* **3**, 599-614 (1992).

170 Marzluff, E. M., Campbell, S., Rodgers, M. & Beauchamp, J. Low-energy dissociation pathways of small deprotonated peptides in the gas phase. *Journal of the American Chemical Society* **116**, 7787-7796 (1994).

171 Schmitz, K. R. & Sauer, R. T. Substrate delivery by the AAA plus ClpX and ClpC1 unfoldases activates the mycobacterial ClpP1P2 peptidase. *Molecular Microbiology* **93**, 617-628 (2014).

172 Thompson, M. W. & Maurizi, M. R. Activity and specificity of Escherichia-coli ClpAp protease in cleaving model peptide-substrates. *Journal of Biological Chemistry* **269**, 18201-18208 (1994).

173 Andersson, F. I., Tryggvesson, A., Sharon, M., Diemand, A. V., Classen, M., Best, C., Schmidt, R., Schelin, J., Stanne, T. M., Bukau, B., Robinson, C. V., Witt, S., Mogk, A. & Clarke, A. K. Structure and function of a novel type of ATP-dependent Clp protease. *Journal of Biological Chemistry* **284**, 13519-13532 (2009).

174 Tryggvesson, A., Stahlberg, F. M., Mogk, A., Zeth, K. & Clarke, A. K. Interaction specificity between the chaperone and proteolytic components of the cyanobacterial Clp protease. *Biochemical Journal* **446**, 311-320 (2012).

175 Stanne, T. M., Pojidaeva, E., Andersson, F. I. & Clarke, A. K. Distinctive types of ATP-dependent Clp proteases in cyanobacteria. *Journal of Biological Chemistry* **282**, 14394-14402 (2007).

176 Zeiler, E., List, A., Alte, F., Gersch, M., Wachtel, R., Poreba, M., Drag, M., Groll, M. & Sieber, S. A. Structural and functional insights into caseinolytic proteases reveal an unprecedented regulation principle of their catalytic triad. *Proceedings of the National Academy of Sciences* **110**, 11302-11307 (2013).

177 Bolanos-Garcia, V. M., Kiyomitsu, T., D'Arcy, S., Chirgadze, D. Y., Grossmann, J. G., Matak-Vinkovic, D., Venkitaraman, A. R., Yanagida, M., Robinson, C. V. & Blundell, T. L. The crystal structure of the N-terminal region of BUB1 provides insight into the mechanism of BUB1 recruitment to kinetochores. *Structure* **17**, 105-116 (2009).

178 Dunbar, R. C. Infrared radioative cooling of gas-phase ions. *Mass Spectrometry Reviews* **11**, 309-339 (1992).

179 Aiello, A. E., Larson, E. L. & Levy, S. B. Consumer Antibacterial Soaps: Effective or Just Risky? *Clinical Infectious Diseases* **45**, S137-S147 (2007).

180 Benesch, J. L., Aquilina, J. A., Ruotolo, B. T., Sobott, F. & Robinson, C. V. Tandem mass spectrometry reveals the quaternary organization of macromolecular assemblies. *Chemistry & Biology* **13**, 597-605 (2006).

181 Schmidt, A., Bahr, U. & Karas, M. Influence of pressure in the first pumping stage on analyte desolvation and fragmentation in nano-ESI MS. *Analytical Chemistry* **73**, 6040-6046 (2001).

182 Blanksby, S. J. & Mitchell, T. W. Advances in mass spectrometry for lipidomics. *Annual Review of Analytical Chemistry* **3**, 433-465 (2010).

183 Andréll, J. & Tate, C. G. Overexpression of membrane proteins in mammalian cells for structural studies. *Molecular Membrane Biology* **30**, 52-63 (2013).

184 Laganowsky, A., Reading, E., Allison, T. M., Ulmschneider, M. B., Degiacomi, M. T., Baldwin, A. J. & Robinson, C. V. Membrane proteins bind lipids selectively to modulate their structure and function. *Nature* **510**, 172-175 (2014).

185 Contreras, F. X., Ernst, A. M., Wieland, F. & Brugger, B. Specificity of intramembrane protein-lipid interactions. *Cold Spring Harb Perspect Biol* **3** (2011).

186 Schmidt, D., Cross, S. R. & MacKinnon, R. A gating model for the archeal voltage-dependent K+ channel KvAP in DPhPC and POPE:POPG decane lipid bilayers. *Journal of Molecular Biology* **390**, 902-912 (2009).

187 Lee, S.-Y., Lee, A., Chen, J. & MacKinnon, R. Structure of the KvAP voltage-dependent K+ channel and its dependence on the lipid membrane. *Proceedings of the National Academy of Sciences* **102**, 15441-15446 (2005).

188 Long, S. B., Tao, X., Campbell, E. B. & MacKinnon, R. Atomic structure of a voltage-dependent K+ channel in a lipid membrane-like environment. *Nature* **450**, 376-382 (2007).

189 Xu, Y., Ramu, Y. & Lu, Z. Removal of phospho-head groups of membrane lipids immobilizes voltage sensors of K+ channels. *Nature* **451**, 826-829 (2008).

190 D'Avanzo, N., McCusker, E. C., Powl, A. M., Miles, A. J., Nichols, C. G. & Wallace, B. A. Differential lipid dependence of the function of bacterial sodium channels. *PLoS ONE* **8**, e61216 (2013).

191 Sonntag, Y., Musgaard, M., Olesen, C., Schiøtt, B., Møller, J. V., Nissen, P. & Thøgersen, L. Mutual adaptation of a membrane protein and its lipid bilayer during conformational changes. *Nature Communications* **2**, 304 (2011).

192 Zheng, H., Liu, W., Anderson, L. Y. & Jiang, Q.-X. Lipid-dependent gating of a voltage-gated potassium channel. *Nature Communications* **2**, 250 (2011).

193 Suh, B.-C. & Hille, B. PIP2 is a necessary cofactor for ion channel function: how and why? *Annual review of biophysics* **37**, 175 (2008).

194 Li, L., Shi, X., Guo, X., Li, H. & Xu, C. Ionic protein–lipid interaction at the plasma membrane: what can the charge do? *Trends in Biochemical Sciences* **39**, 130-140 (2014).

195 Di Paolo, G. & De Camilli, P. Phosphoinositides in cell regulation and membrane dynamics. *Nature* **443**, 651-657 (2006).

196 Shin, J. J. & Loewen, C. J. Putting the pH into phosphatidic acid signaling. *BMC Biology* **9**, 85 (2011).

197 Gitlin, I., Carbeck, J. D. & Whitesides, G. M. Why are proteins charged? Networks of charge–charge interactions in proteins measured by charge ladders

and capillary electrophoresis. *Angewandte Chemie International Edition* **45**, 3022-3060 (2006).

198 Dobson, C. M. Protein folding and misfolding. *Nature* **426**, 884-890 (2003).

199 Garcia-Viloca, M., Gao, J., Karplus, M. & Truhlar, D. G. How enzymes work: analysis by modern rate theory and computer simulations. *Science* **303**, 186-195 (2004).

200 van den Bogaart, G., Meyenberg, K., Risselada, H. J., Amin, H., Willig, K. I., Hubrich, B. E., Dier, M., Hell, S. W., Grubmüller, H. & Diederichsen, U. Membrane protein sequestering by ionic protein-lipid interactions. *Nature* **479**, 552-555 (2011).

201 Young, B. P., Shin, J. J., Orij, R., Chao, J. T., Li, S. C., Guan, X. L., Khong, A., Jan, E., Wenk, M. R. & Prinz, W. A. Phosphatidic acid is a pH biosensor that links membrane biogenesis to metabolism. *Science* **329**, 1085-1088 (2010).

202 Lide, D. R. *CRC handbook of chemistry and physics*. (CRC press, 2004).

203 Khuong, T. M., Habets, R. L., Kuenen, S., Witkowska, A., Kasprowicz, J., Swerts, J., Jahn, R., van den Bogaart, G. & Verstreken, P. Synaptic PI (3, 4, 5) P 3 is required for syntaxin1A clustering and neurotransmitter release. *Neuron* **77**, 1097-1108 (2013).

204 Lam, A. D., Tryoen-Toth, P., Tsai, B., Vitale, N. & Stuenkel, E. L. SNARE-catalyzed fusion events are regulated by syntaxin1A–lipid interactions. *Molecular biology of the cell* **19**, 485-497 (2008).

205 Tusnády, G. E., Dosztányi, Z. & Simon, I. Transmembrane proteins in the Protein Data Bank: identification and classification. *Bioinformatics* **20**, 2964-2972 (2004).

206 O'Keeffe, A. H., East, J. M. & Lee, A. G. Selectivity in lipid binding to the bacterial outer membrane protein OmpF. *Biophysical Journal* **79**, 2066-2074 (2000).

207 Nestorovich, E. M., Rostovtseva, T. K. & Bezrukov, S. M. Residue ionization and ion transport through OmpF channels. *Biophysical Journal* **85**, 3718-3729 (2003).

208 Müller, D. J. & Engel, A. Voltage and pH-induced channel closure of porin OmpF visualized by atomic force microscopy1. *Journal of Molecular Biology* **285**, 1347-1351 (1999).

209 Deol, S. S., Bond, P. J., Domene, C. & Sansom, M. S. Lipid-protein interactions of integral membrane proteins: a comparative simulation study. *Biophysical Journal* **87**, 3737-3749 (2004).

210 Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & De Vries, A. H. The MARTINI force field: coarse grained model for biomolecular simulations. *The Journal of Physical Chemistry B* **111**, 7812-7824 (2007).

211 Gallego, M. & Virshup, D. M. Post-translational modifications regulate the ticking of the circadian clock. *Nature Reviews Molecular Cell Biology* **8**, 139-148 (2007).

212 Witze, E. S., Old, W. M., Resing, K. A. & Ahn, N. G. Mapping protein post-translational modifications with mass spectrometry. *Nature Methods* **4**, 798-806 (2007).

213 Capaldi, R. A. Structure and function of cytochrome c oxidase. *Annual Review of Biochemistry* **59**, 569-596 (1990).

214 Krause, F., Scheckhuber, C. Q., Werner, A., Rexroth, S., Reifschneider, N. H., Dencher, N. A. & Osiewacz, H. D. Supramolecular organization of cytochrome c oxidase-and alternative oxidase-dependent respiratory chains in the filamentous fungus Podospora anserina. *Journal of Biological Chemistry* **279**, 26453-26461 (2004).

215 Palmer, J. The organization and regulation of electron transport in plant mitochondria. *Annual Review of Plant Physiology* **27**, 133-157 (1976).

216 Babcock, G. T. & Wikstrom, M. Oxygen activation and the conservation of energy in cell respiration. *Nature* **356**, 301-309 (1992).

217 Kadenbach, B., Unigbauer, M., Jarausch, J., Buge, U. & Kuhn-Nentwig, L. The complexity of respiratory complexes. *Trends Biochem. Sci.* **8**, 398-400 (1983).

218 Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R. & Yoshikawa, S. The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 A. *Science* **272**, 1136-1144 (1996).

219 Yoshikawa, S., Muramoto, K. & Shinzawa-Itoh, K. Proton-pumping mechanism of cytochrome C oxidase. *Annu Rev Biophys* **40**, 205-223 (2011).

220 Yoshikawa, S., Muramoto, K., Shinzawa-Itoh, K. & Mochizuki, M. Structural studies on bovine heart cytochrome c oxidase. *Biochim Biophys Acta* **1817**, 579-589 (2012).

221 Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R. & Yoshikawa, S. Structures of metal sites of oxidized bovine heart cytochrome c oxidase at 2.8 A. *Science* **269**, 1069-1074 (1995).

222 Robinson, N. C. Specificity and binding affinity of phospholipids to the high-affinity cardiolipin sites of beef heart cytochrome c oxidase. *Biochemistry* **21**, 184-188 (1982).

223 Harrenga, A. & Michel, H. The cytochrome c oxidase from Paracoccus denitrificans does not change the metal center ligation upon reduction. *Journal of Biolology and Chemistry* **274**, 33296-33299 (1999).

224 Svensson-Ek, M., Abramson, J., Larsson, G., Tornroth, S., Brzezinski, P. & Iwata, S. The X-ray crystal structures of wild-type and EQ(I-286) mutant cytochrome c oxidases from Rhodobacter sphaeroides. *Journal of Molecular Biology* **321**, 329-339 (2002).

225 Shinzawa-Itoh, K., Aoyama, H., Muramoto, K., Terada, H., Kurauchi, T., Tadehara, Y., Yamasaki, A., Sugimura, T., Kurono, S., Tsujimoto, K., Mizushima, T., Yamashita, E., Tsukihara, T. & Yoshikawa, S. Structures and physiological roles of 13 integral lipids of bovine heart cytochrome c oxidase. *EMBO Journal* **26**, 1713-1725 (2007).

226 Muramoto, K., Ohta, K., Shinzawa-Itoh, K., Kanda, K., Taniguchi, M., Nabekura, H., Yamashita, E., Tsukihara, T. & Yoshikawa, S. Bovine cytochrome c oxidase structures enable O2 reduction with minimization of reactive oxygens and provide a proton-pumping gate. *Proceedings of the National Academy of Sciences* **107**, 7740-7745 (2010).

227 Kadenbach, B. & Huttemann, M. The subunit composition and function of mammalian cytochrome c oxidase. *Mitochondrion* **24**, 64-76 (2015).

228 Helling, S., Huttemann, M., Ramzan, R., Kim, S. H., Lee, I., Muller, T., Langenfeld, E., Meyer, H. E., Kadenbach, B., Vogt, S. & Marcus, K. Multiple phosphorylations of cytochrome c oxidase and their functions. *Proteomics* **12**, 950-959 (2012).

229 Saraste, M., Penttila, T. & Wikstrom, M. Quaternary structure of bovine cytochrome oxidase. *Europian Journal of Biochemistry* **115**, 261-268 (1981).

230 Thompson, D. A., Gregory, L. & Ferguson-Miller, S. Cytochrome c oxidase depleted of subunit III: proton-pumping, respiratory control, and pH dependence of the midpoint potential of cytochrome a. *Journal of Inorganic Biochemistry* **23**, 357-364 (1985).

231 Sedlak, E. & Robinson, N. C. Sequential dissociation of subunits from bovine heart cytochrome C oxidase by urea. *Biochemistry* **48**, 8143-8150 (2009).

232 Sedlak, E., Varhac, R., Musatov, A. & Robinson, N. C. The kinetic stability of cytochrome C oxidase: effect of bound phospholipid and dimerization. *Biophysic Journal* **107**, 2941-2949 (2014).

233 Wittig, I. & Schagger, H. Advantages and limitations of clear-native PAGE. *Proteomics* **5**, 4338-4346 (2005).

234 Yoshikawa, S., Shinzawa-Itoh, K., Nakashima, R., Yaono, R., Yamashita, E., Inoue, N., Yao, M., Fei, M. J., Libeu, C. P., Mizushima, T., Yamaguchi, H., Tomizaki, T. & Tsukihara, T. Redox-coupled crystal structural changes in bovine heart cytochrome c oxidase. *Science* **280**, 1723-1729 (1998).

235 Reading, E., Liko, I., Allison, T. M., Benesch, J. L., Laganowsky, A. & Robinson, C. V. 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).

236 Kadenbach, B. & Hüttemann, M. The subunit composition and function of mammalian cytochrome c oxidase. *Mitochondrion* **24**, 64-76 (2015).

237 Arnarez, C., Marrink, S. J. & Periole, X. Identification of cardiolipin binding sites on cytochrome c oxidase at the entrance of proton channels. *Scientific Reports* **3** (2013).

238 Tsukihara, T., Shimokata, K., Katayama, Y., Shimada, H., Muramoto, K., Aoyama, H., Mochizuki, M., Shinzawa-Itoh, K., Yamashita, E., Yao, M., Ishimura, Y. & Yoshikawa, S. The low-spin heme of cytochrome c oxidase as the driving element of the proton-pumping process. *Proceedings of the National Academy of Sciences* **100**, 15304-15309 (2003).

239 Daum, G. Lipids of mitochondria. *Biochimica et Biophysica Acta* **822**, 1-42 (1985).

240 Ly, T., Ahmad, Y., Shlien, A., Soroka, D., Mills, A., Emanuele, M. J., Stratton, M. R. & Lamond, A. I. A proteomic chronology of gene expression through the cell cycle in human myeloid leukemia cells. *Elife* **3**, e01630 (2014).

241 Rassam, P., Copeland, N. A., Birkholz, O., Toth, C., Chavent, M., Duncan, A. L., Cross, S. J., Housden, N. G., Kaminska, R., Seger, U., Quinn, D. M., Garrod, T. J., Sansom, M. S. P., Piehler, J., Baumann, C. G. & Kleanthous, C. Supramolecular assemblies underpin turnover of outer membrane proteins in bacteria. *Nature* **523**, 333-336 (2015).

242 Housden, N. G., Wojdyla, J. A., Korczynska, J., Grishkovskaya, I., Kirkpatrick, N., Brzozowski, A. M. & Kleanthous, C. Directed epitope delivery across the Escherichia coli outer membrane through the porin OmpF. *Proceedings of the National Academy of Sciences* **107**, 21412-21417 (2010).

243 Zimmermann, I. & Dutzler, R. Ligand Activation of the Prokaryotic Pentameric Ligand-Gated Ion Channel ELIC. *PLoS Biology* **9**, e1001101 (2011).

244 Shvartsburg, A. A. & Jarrold, M. F. An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chemical Physics Letters* **261**, 86-91 (1996).

245 Mesleh, M. F., Hunter, J. M., Shvartsburg, A. A., Schatz, G. C. & Jarrold, M. F. Structural information from ion mobility measurements: Effects of the long-range potential. *The Journal of Physical Chemistry* **100**, 16082-16086 (1996).

246 Hall, Z., Politis, A. & Robinson, C. V. Structural modeling of heteromeric protein complexes from disassembly pathways and ion mobility-mass spectrometry. *Structure* **20**, 1596-1609 (2012).

247 Young, T. An essay on the cohesion of fluids. *Philosophical Transactions of the Royal Society of London* **95**, 65-87 (1805).

248 Garandet, J. P., Vinet, B. & Gros, P. Considerations on the pendant drop method: A new look at Tate's law and Harkins' correction factor. *Journal of Colloid and Interface Science* **165**, 351-354 (1994).

249 Tate, T. On the magnitude of a drop of liquid formed under different circumstances. *Philosophical Magazine Series 4* **27**, 176-180 (1864).

250 Stengel, F., Baldwin, A. J., Bush, M. F., Hilton, G. R., Lioe, H., Basha, E., Jaya, N., Vierling, E. & Benesch, J. L. Dissecting heterogeneous molecular chaperone complexes using a mass spectrum deconvolution approach. *Chemistry & Biology* **19**, 599-607 (2012).

251 Housden, N. G., Hopper, J. T., Lukoyanova, N., Rodriguez-Larrea, D., Wojdyla, J. A., Klein, A., Kaminska, R., Bayley, H., Saibil, H. R., Robinson, C. V. & Kleanthous, C. Intrinsically disordered protein threads through the bacterial outer-membrane porin OmpF. *Science* **340**, 1570-1574 (2013).

252 Marty, M. T., Zhang, H., Cui, W., Gross, M. L. & Sligar, S. G. Interpretation and deconvolution of nanodisc native mass spectra. *Journal of the American Society for Mass Spectrometry* **25**, 269-277 (2014).

253 Kozma, D., Simon, I. & Tusnády, G. E. PDBTM: Protein data bank of transmembrane proteins after 8 years. *Nucleic Acids Research* **41**, D524-D529 (2013).

254 Birant, D. & Kut, A. ST-DBSCAN: An algorithm for clustering spatial–temporal data. *Data & Knowledge Engineering* **60**, 208-221 (2007).

255 Stansfeld, P. J., Goose, J. E., Caffrey, M., Carpenter, E. P., Parker, J. L., Newstead, S. & Sansom, M. S. MemProtMD: automated insertion of membrane protein structures into explicit lipid membranes. *Structure* **23**, 1350-1361 (2015).

256 Monticelli, L., Kandasamy, S. K., Periole, X., Larson, R. G., Tieleman, D. P. & Marrink, S.-J. The MARTINI coarse-grained force field: extension to proteins. *Journal of Chemical Theory and Computation* **4**, 819-834 (2008).

257 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).

258 Mochizuki, M., Aoyama, H., Shinzawa-Itoh, K., Usui, T., Tsukihara, T. & Yoshikawa, S. Quantitative reevaluation of the redox active sites of crystalline bovine heart cytochrome c oxidase. *The Journal of Biological Chemistry* **274**, 33403-33411 (1999).

259 Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Analytical Chemistry* **68**, 850-858 (1996).