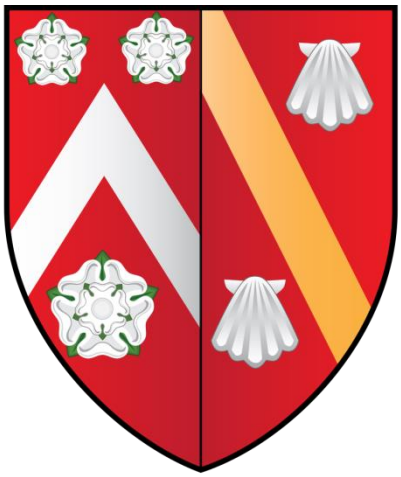
Probing Environmental Effects on Gas-Phase Protein Structure

Dominic James Hewitt

University of Oxford

Wadham College



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

2017

Probing environmental effects on gas-phase protein structure

Dominic James Hewitt, Wadham College

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

Hilary term 2017

Abstract

Growth in the use of mass spectrometry as an analytical tool for studying protein structure and functionality has posed questions regarding correlation of native structures and observed gas-phase conformations for proteins. This thesis explores the relationship between the solution-phase and gas-phase structures of soluble proteins, and investigates the effect of the lipid environment on membrane proteins.

To assess gas-phase structure, ion-mobility mass spectrometry was used to measure directly the collision cross-sections of soluble proteins across a broad mass range. These collision cross-sections values were then compared with those derived from solution-phase data, calculated from dynamic light scattering experiments. Using computational methods collision cross-sections were calculated from structures reported in the protein data bank using a variety of techniques. Differing physical environments in solution and gas phases are not shown to affect protein structure by the dynamic light scattering and ion-mobility measurements. Contrastingly the computational calculations demonstrate that there is less correspondence between x-ray structures and their gas-phase counterparts.

Despite the effect that lipid environment has on membrane protein structure, capturing this native environment has been challenging. Lipodisqs look to overcome this problem by removing the portion of the membrane surrounding the embedded protein. Lipodisqs were

investigated as potential vehicles for membrane protein mass spectrometry and compared with conventional detergent based approaches. Lower charge states are observed for proteins released from lipodisqs, compared to those released from detergent micelles, increasing the probability of native features being retained including native-like protein-lipid interactions. Lipodisqs therefore demonstrate significant potential as a tool for membrane protein mass spectrometry.

In summary, through this body of experimental work relationships have been established between gas-phase and solution-phase structures as well as computational methods. The use of Lipodisqs to eject membrane proteins from regions of the native membrane has been investigated.

Table of Contents

Declaration of Authorship

This thesis is a summary of research performed at the Department of Chemistry, University of Oxford between September 2012 and April 2017. The work detailed here is wholly my own and not the outcome of a collaboration except where 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 the thesis does not exceed the word limit.

Dominic James Hewitt

Oxford, April 2017

Abbreviations

BME – β-mercaptoethanol

BR – Bacteriorhodopsin

CID – Collision induced dissociation

DDM – n-Dodecyl-β-D-Maltoside

DLS – Dynamic light scattering

DMPC – 1,2-Dimyristoyl-sn-glycero-3-phosphocholine

EHSS – Exact hard sphere scattering

GDH – Glutamate dehydrogenase

GFP – Green fluorescent protein

HCD – High-energy collision induced dissociation

IM-MS – Ion mobility mass spectrometry

IPTG - Isopropyl β-D-1-thiogalactopyranoside

MATE – Multidrug and toxic compound extrusion protein

MS – Mass Spectrometry

MS-MS – Tandem mass spectrometry

MscL – Mechanosensitive channel of large conductance

OG – Octly Glucoside

PA – Projection Approximation

PBD – Protein data bank

PSA - Projected superposition approximation

RF – Radiofrequency

SEC – Size exclusion chromatography

SMA – Styrene Maleic acid

TJM – Trajectory method

ToF – Time-of-flight

Acknowledgements

Firstly, and most importantly, I'd like to thank Carol for her support, and for giving me the opportunity to undertake my PhD in her group. Despite a few setbacks during the last four and a half year she's always been hugely supportive and encouraging regardless. Without a doubt her support has been vital in ensuring the completion of this project and I cannot thank her enough for that.

Secondly, my two supervisors; Toni and Jon both of whom offered help and guidance at a number of points. Toni for his support at the start of my time studying in the group, guiding me during my part 2, and then during the start of my PhD. His help getting me in touch with Dave Scott allowing me to move on with the DLS project was vital. Jon, who encouraged me to take on the Lipodisq project and continued to support me despite the many setbacks during it. In addition his positive attitude throughout helped to keep me motivated.

To Erik and Matteo thank you for helping me delve into the horrific world of programming. Without their help none of the computational work would have been even remotely possible. Erik's tireless patience explaining how to run calculations using MOBCAL was hugely appreciated. Matteo for his help teaching me how to use python and also for his patience and the level of happiness he always brings to the office.

A huge thank you to my fellow DPhils in particular Idi and Heidi without whom I cannot imagine the last few years. Idi for his constant willingness to help with virtually any problem I had and introducing me to Albanian alcohol. Heidi who showed there is a positive way to look at everything. Possibly most importantly for helping campaign for Christmas music in the lab as soon as Christmas was even vaguely near.

To Dave Scott and Stephen Carr who helped to make the DLS project happen allowing me to use their space at Harwell briefly and their useful discussions. Particularly to Dave, who also set up a visit to the National Centre for Macromolecular Hydrodynamics allowing for some extra measurements to take place.

From the Watts group I'd like to thank Garrick Taylor, Juan Juarez, Peter Judge and of course Anthony Watts himself. Firstly to Garrick for helping get the project started and not giving up after the first few mishaps. Then to Juan who was instrumental in the project and many hours sat together purifying and acquiring data. Peter for his advice and oversight of the project and to Tony for his encouragement in pursing it.

Finally to all the other members of the Robinson group and the Benesch group for providing a wonderful atmosphere to work in. To all of those who braved the cold and the occasional sun to play football it was a much needed stress relief. Joe G for all his work getting the QExactive up and running and then patiently explaining how not to break it. To Tim for answering the questions about protein expression and purification and not letting all the columns die. In a similar vein a thank you to Todd for his work constantly fixing instruments and Helena and Lauren for their work keeping the lab running efficiently.

On a personal level outside of the lab I'd like to thank a number of people in particular for their tolerance during the less successful periods. Mikey, Chris and Max for enduring rants about experiments not working and constantly encouraging me that it would end and get submitted. In addition to Kev, Chris and Max for taking catching up to completely unnecessary and ridiculous extremes that are the highlight of any month. Lastly to my family and Amy whose support and belief in me was unwavering.

1. Introduction:

1.1. Motivation and Aims:

The developments of matrix-assisted laser desorption/ionisation (MALDI) and electrospray ionisation (ESI) as ionisation methods for mass spectrometry (MS) has led to an explosion in the field of MS. Both these techniques allowed for the use of MS as an analytical technique for studying proteins in 1989[1, 2, 3, 4]. Whilst this was a significant breakthrough in the field, the initial studies were somewhat limited by the necessity of using denaturing solvents. This means the data obtainable is limited to simple mass measurements as the organic solvents required for ease of ionisation also result in denaturation of the proteins.

Development of the ESI process allowed for the use of aqueous solvents enabling the possibility of mass spectrometry on non-denatured proteins[5, 6, 7]. This in turn opened up the avenue of research into native protein structures and properties using MS. However due to the nature and conditions in which MS takes place there is understandable scepticism about how native-like structures can remain whilst experiencing these conditions.

The development of an orthogonal ion-mobility time-of-flight spectrometer in 1999[8] showed the ability to couple ESI with an ion-mobility cell and the possibility for using this set-up for proteins. This laid the foundation for the development of a commercial ion-mobility mass spectrometry by Waters in 2006 allowing for a relatively straightforward method to obtain information about the shape and structure of proteins. However whilst the shape and structure of soluble proteins is well understood in solution there is some

debate as to the validity of shape and structural information determined using IM-MS due to the proteins being in the gas-phase and therefore in an anhydrous environment.

The first section of this thesis deals with the comparison of ion-mobility mass spectrometry (IM-MS) data with low-resolution structural information obtained in solution using dynamic light scattering (DLS) coupled with size exclusion chromatography (SEC).

One of the other consequences of the development of IM-MS was an advance in computational techniques that allow for the calculation of a collision cross-section for a protein in the gas-phase based upon a protein structure solved either using X-ray crystallography[9, 10, 11, 12, 13] or NMR spectroscopy[14, 15, 16]. These methods for calculation include but are not limited to the projection approximation[17], the exact hard sphere scattering method[18], the trajectory method[19] and the projected superposition approximation[20]. These methods are designed to validate the collision cross-sections obtained through IM-MS experiments either through estimation of a collisional cross-section or mimicking the ion-mobility experiment. These calculated cross-sections can be utilised to predict the cross-sections for proteins in the gas-phase.

Traditionally the most common method used for solving these structures has been X-ray crystallography, with over one hundred thousand structures solved utilising this method as of 2015[21] compared with just over eleven thousand by nuclear magnetic resonance (NMR) and just over eight hundred and fifty by electron microscopy (EM) as shown in **Figure 1-1**.

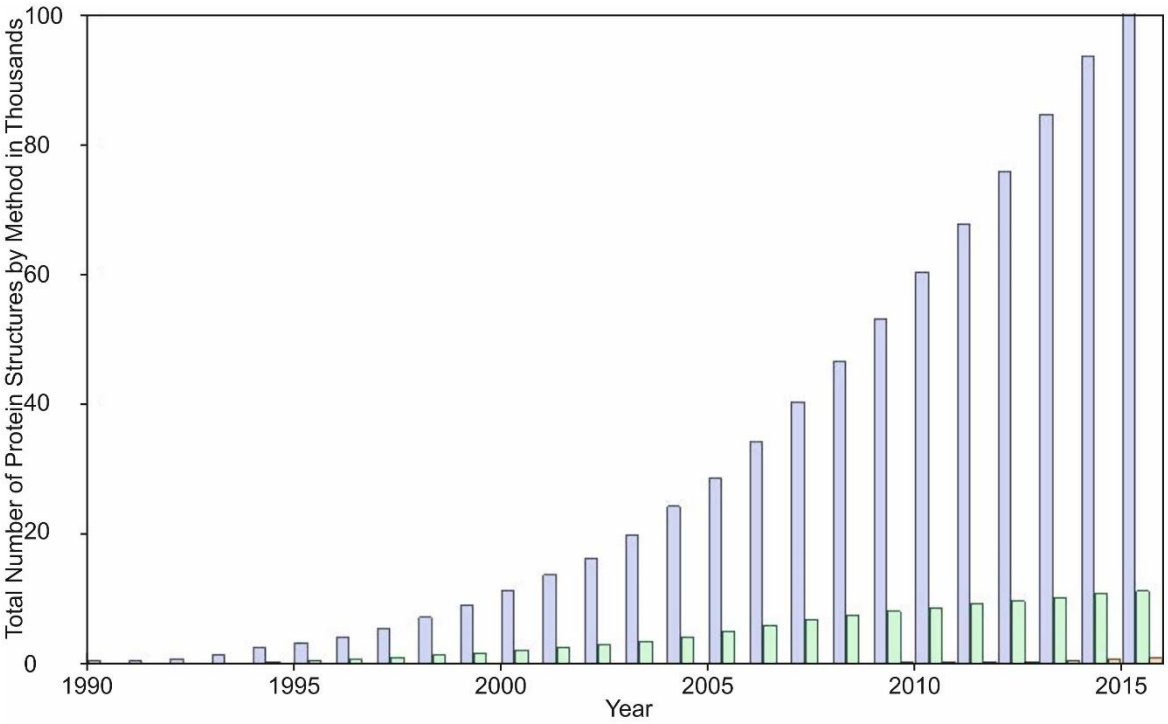


Figure 1-1: Comparison of number of protein structures solved by different experimental methods from 1990 to 2015[21]. X-ray crystallography (blue) vs NMR (green) vs EM (orange)

The work detailed in chapter two demonstrating consistency between low resolution shape and structural information between the solution-phase and gas-phase. However comparison between those measurements and the computationally derived cross-sections based on known x-ray crystal structures do not show the same consistency[22]. The second section of this thesis investigates if structures obtained utilising alternative methods to x-ray crystallography for the starting structure for the computational methods affects the collision cross-section calculated. Comparing results from DLS and IM-MS to the structures solved using either cryo-EM or NMR in addition to x-ray crystallography where possible.

The final part of this thesis investigates one of the most recent areas of application of mass spectrometry that of membrane proteins. Whilst the first spectra for a native-like soluble protein were recorded in the early 1990s[5, 6] due to the inherent difficulties

associated with working on membrane proteins this was only possible some years later in the early 2000s[23, 24]. In order to study membrane proteins outside of the cell membrane it is necessary to use an agent in order to mimic its native environment after extraction of the protein from the membrane. Failure to do so results in the protein precipitating out.

Traditionally detergents have been used to solubilise membrane proteins to allow for their analysis. However typically the use of these detergents in order to solubilise the protein can lead to reduced stability of the protein causing it to precipitate or degrade[3, 25]. Even when protein stability is not seriously compromised, detergents will still cause delipidation of the protein which can be problematic as it has been shown that lipids can have an effect on protein oligomerisation[26, 27, 28].

In an attempt to overcome this problem with regard to the study of membrane proteins multiple methods have been developed to complement or act as an alternative to the use of detergents. These include but are not limited to Bicelles[29], nanodiscs[30] and amphipols[31]. These three techniques have all been used as alternative vehicles for MS[32], however a major drawback of these methods is the requirement to use detergent to solubilise the protein initially before incorporating it into any of these. Utilising detergent in the process opens up the possibility that during the initial solubilisation the detergent could cause the protein to unfold or disassociate its oligomeric state. There is also a secondary risk that the protein could become delipidated losing protein specific lipids which are required for it to retain its native oligomeric state and structure.

The final chapter of this thesis investigates the possibility of using native nanodiscs to solubilise the protein instead of any of the previously described methods. The primary advantage theoretically in adopting this approach is the ability to solubilise membrane

proteins directly from cell membranes[33, 34]. This approach not only circumvents the need for any detergent, avoiding any potentially unfavourable interactions with the detergent, it also allows for the protein to be purified with its native lipids potentially allowing for the investigation of crucial protein lipid interactions.

1.2. Non-Covalent Mass Spectrometry:

Mass spectrometry is a technique in which the m/z of an ion, or a series of ions, is determined by measuring the change in a property of the ion, such as its flight time or trajectory in response to magnetic or electric fields. The ion goes through four stages during the experiment; ionisation, separation, detection and then a computer generates an output in the form of a spectrum as shown in a simplified schematic in **Figure 1-2**.

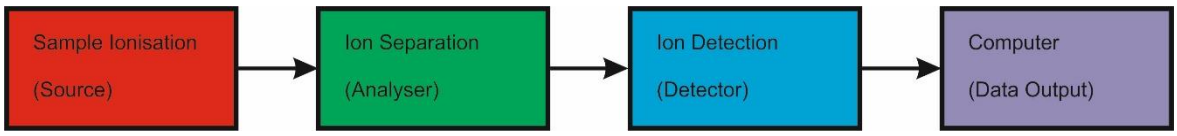


Figure 1-2: Simple schematic of the process a sample goes through during mass spectrometry.

Initially mass spectrometry ionisation techniques such as electron impact ionisation and chemical ionisation are classed as hard ionisation techniques to generate the ions. This means that whilst they are effective at ion generation they also impart so much energy it leads to the breaking of chemical bonds in addition to non-covalent interactions[35]. While this is structurally informative for small molecules it is not possible to analyse large biomolecules, such as peptides or proteins, without extensive derivatization using such ionisation methods.

Over time numerous softer ionisation techniques have been developed including electrospray ionisation (ESI)[36, 37] and matrix-assisted laser desorption ionisation (MALDI)[38]. One of the first examples of softer using ionisation for mass spectrometry was

demonstrated in an investigation of protein folding reactions in 1993[39], which followed on from the work of Katta and Chait[5]. Here they showed that with a softer ionisation method the myoglobin remained associated with its heme group. Under these soft energy regimes the non-covalent interactions in a molecule are preserved following electrospray ionisation.

Further developments of these ionisation techniques allowed for the preservation of vital non-covalent interactions which was not previously possible. This was a huge breakthrough for the field of mass spectrometry due to the vast majority of proteins existing as oligomers held together by these non-covalent interactions. Without maintaining these interactions in the mass spectrometer it would be impossible to make any conclusions on the biological implications of any findings as the protein would not be in anything close to a native state.

1.2.1. Electrospray Ionisation (ESI):

ESI is a soft ionisation technique developed in 1989 by Fenn[36]. It is described as a "soft" ionisation technique as the ionisation process does not cause extensive fragmentation of ions. ESI works by using electric potentials to create an electric field which draws charged ions in the sample into the aperture at the front of the mass spectrometer. This potential difference initially leads to the formation of a cone from the end of the needle containing the sample towards the aperture of the instrument, resulting from combination of the electrostatic attraction to the aperture and the surface tension of the sample. When the electrostatic attraction overcomes the surface tension then electrospray ionisation occurs. The development of ESI as a technique for ionisation in MS opened the door to

the possibility of using the technique to investigate proteins interactions using this methodology[7, 40].

1.2.2. ESI Mechanism:

ESI in its first implementation took place in a stainless steel hypodermic needle which is attached to the positive side of a high voltage power supply whilst the cone at the entrance of the mass spectrometer is attached to the negative side. The potential difference between these two leads to the formation of a Taylor cone at the end of the needle, as the positively charged ions in the needle are attracted to the negative charge that is building up on the cone as shown below **(Figure 1-3)**.

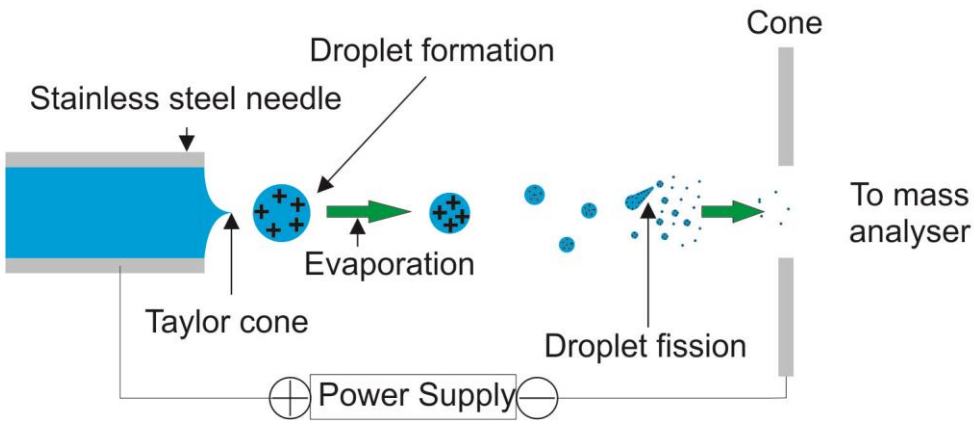


Figure 1-3: Formation of a Taylor cone due to the electric potential between the capillary and the cone. Adapted from [41]

The formation of a stable electrospray is conditional on the electric field, solvent surface tension and the radius of the capillary as given in *Equation 1.1*[42] where E_0 is the electric field, r_c is the radius of the capillary, γ is the surface tension of the solvent and ε_0 is the permittivity of free space.

(Equation 1.1)

$$E_0 = \left(\frac{2\gamma cos49°}{\varepsilon_0 r_c}\right)^{\frac{1}{2}}$$

When a stable electrospray has been established the charged droplet will be pulled towards the cone of the mass spectrometer. During this process the droplet starts to undergo Coulombic fission which allows it to reduce in size so that ideally at the point when it enters the instrument each charged droplet will only contain one analyte molecule. This process takes places due to solvent evaporating which in turn leads to an increase in the charge density of the droplet **(Figure 1-4)**.

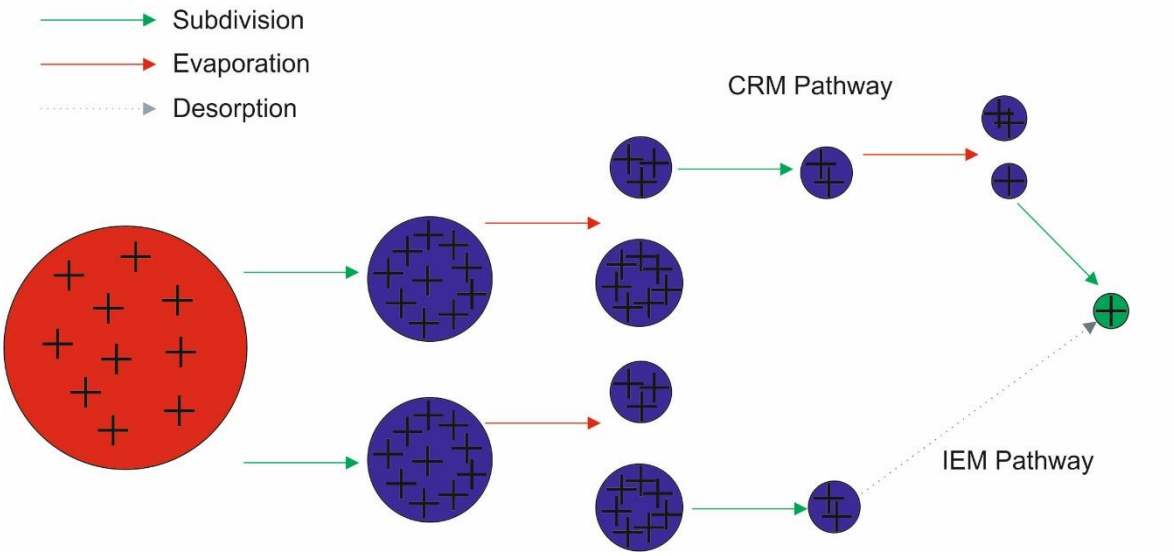


Figure 1-4: Demonstration of the two alternative pathways for generation of ions from an ESI created droplet.

With the charge density increasing it will reach the Rayleigh limit, given in *Equation 1.2*[42], which is the point at which the surface tension of the solvent is insufficient to maintain that level of charge density leading to the droplet undergoing fission.

(Equation 1.2)

$$q = 8\pi\sqrt{\varepsilon_0 \gamma R^3}$$

Here q is the charge on the droplet, R is the radius of the droplet, γ is the surface tension of the solvent and ε_0 is the permittivity of free space. This process continues with each

fission event as the droplet moves on towards the source of the mass spectrometer. These smaller droplets will continue undergoing these fission events until each droplet only contains a single ion, which is the ideal situation for MS. From this point onwards there are two models described for the release of desolvated ions from the remaining droplet.

There are two main models proposed for the fission pathways of a droplet containing a single ion to the point where a desolvated ion enters the spectrometer. These models are known as the charge residue model[37, 43] (CRM) and the ion evaporation model[44, 45] (IEM). They differ in terms of the final separation of the charged analyte ion from the bulk droplet.

In the CRM the droplet undergoes evaporation losing more and more of the solvent shell until only the analyte ion remains. By contrast the IEM model proposes that a highly charged ion can be ejected directly from the droplet. In reality larger protein molecules and their complexes will be ionised mainly through the CRM whereas smaller proteins may also be ionised partly through the IEM[46].

1.3. Methods for Mass Analysis:

There are three main types of analysers for MS which are widely in use at present. These three are: Time-of-flight (ToF), quadrupole and orbitrap analysers, each can be used individually or can be used in tandem. Whilst only one is required in order to have a functioning mass spectrometer the use of multiple analysers allows for extra levels of detail to be obtained during MS for example in tandem mass spectrometry which allows for the selection of an ion to later fragment for further analysis.

1.3.1. Time-of-Flight (ToF) Analysers:

ToF analysers work, as the name suggests, by measuring the time it takes for ions to travel a known distance after being accelerated through an electric field. From this measurement it is possible to determine the mass to charge ratio (m/z) for the ion. When a continuous method for ion generation is used, such as electrospray, it is necessary to perform this measurement perpendicular to the direction of travel through the instrument up until this point, as shown below in **Figure 1-5**[47]. However if a pulsed ionisation technique such as MALDI is used there is no need for this measurement to be done perpendicularly. The primary advantage in the use of ToF analysers is that they have no theoretical maximum m/z range which they can work up to making them ideal for use in studying large protein complexes. ToF analysers often make use of a reflectron which doubles the effective distance the ions travel before detection. The main reason for using a reflectron is that it allows for a reduction in the spread of arrival times for any series of ions with the same m/z but varying kinetic energies upon arrival into the analyser cell which in turn decreases the error in the measurement for the m/z for the ions.

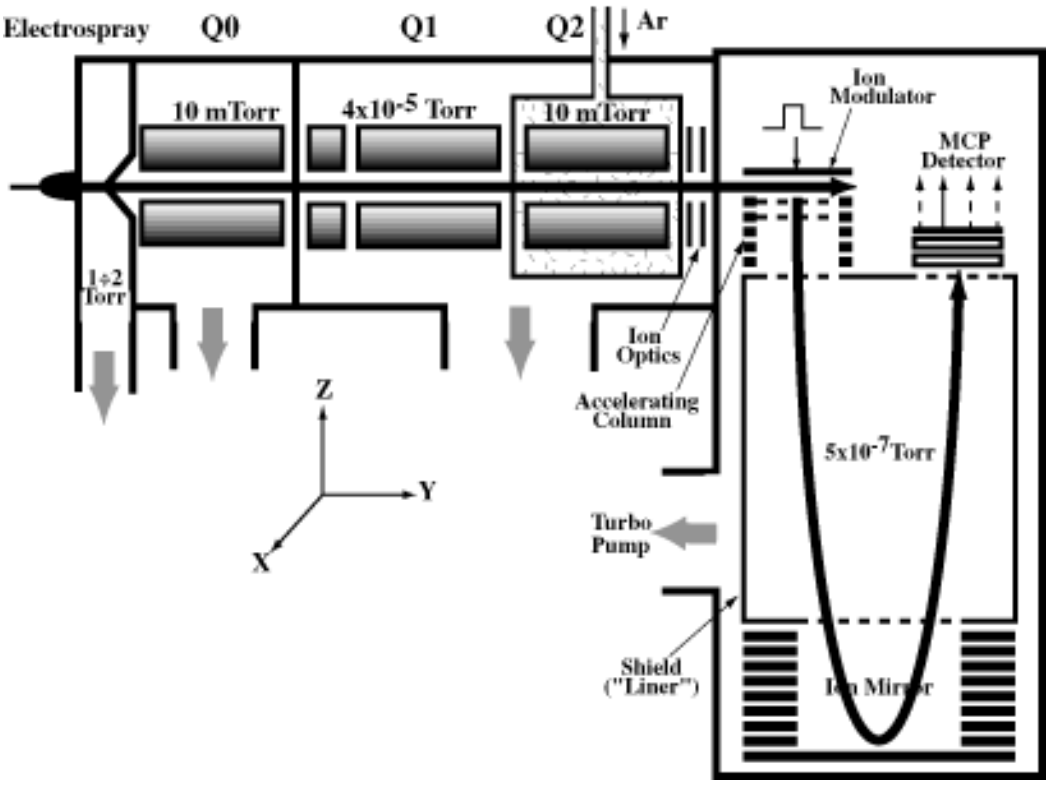


Figure 1-5[47]: A quadrupole-time-of-flight mass spectrometer as the ions enter the time of flight analyser at the end they are propelled perpendicular to the direction of travel throughout the instrument until that point.

1.3.2. Quadrupole Analysers:

Quadrupole analysers work by utilising four rods where the rods opposite one another are connected together electrically in pairs. The two pairs of rods alternate their polarity based upon the radiofrequency that is applied to them. The quadrupole can be used in one of two modes at any one time; scanning mode or transmission mode. In scanning mode the direct current is superimposed upon the radio frequency to only allow ions within a specified m/z range to have a stable trajectory throughout the quadrupole as illustrated in **Figure 1-6**. Scanning mode takes advantage of the fact that ions with too high an m/z will not be accelerated enough, and their initial momentum will cause them to fly into the quadrupole and not be transmitted. In contrast ions with too low an m/z will experience too great an acceleration, and their trajectories will rapidly become unstable and pass out of the central channel of the quadrupole. This ability to selectively

transmit only ions in a certain m/z range is particularly useful as it allows for the possibility of multiple levels of analysis within a single experiment. This is a fact which is taken advantage of in the use of tandem mass spectrometry which has developed into a widespread technique in the field.

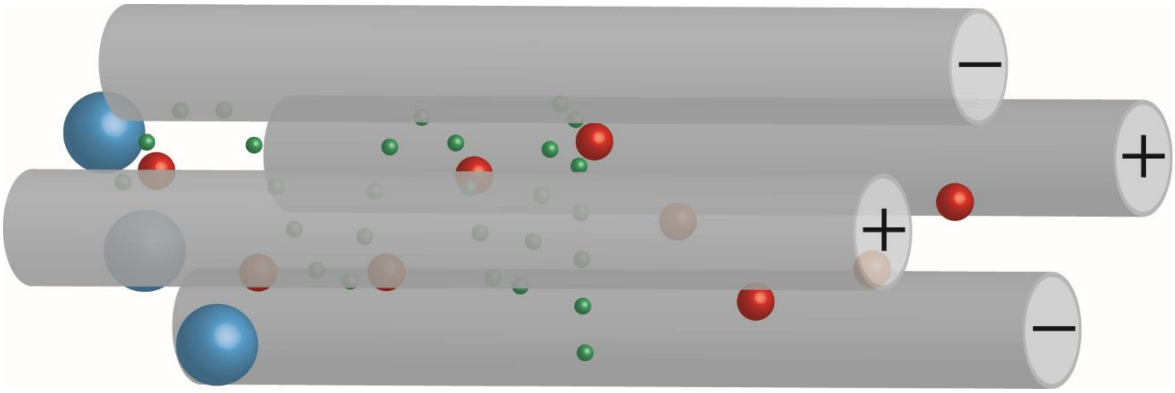


Figure 1-6: A schematic of a quadrupole analyser showing selection of the red ion. The ions in blue have too high a m/z to be guided by the quadrupole whereas the ions in green have too low a m/z and so are ejected out of the quadrupole.

By comparison in transmission mode the quadrupole acts as a funnel pushing the ions into the central region of the quadrupole. This is done by using the quadrupole in rf only mode which only takes advantage of its ability to concentrate the ion-beam as opposed to trying to selectively transmit certain ions.

Whilst using quadrupoles for native protein MS it is necessary to make sure they have appropriate resolving power for the proteins of interest. The maximum m/z resolvable for a given quadrupole is determined by the maximum radiofrequency voltage that can be applied between opposite pairs of rods (V), the radius of the rods in meters (r_0) and the frequency of the machine in Hertz (ν) as shown in *Equation 1.3*[48].

(Equation 1.3[48]*)*

$$m/z_{Max} = \frac{7 \times 10^6 V}{\nu^2 r_0^2}$$

1.3.3. Orbitrap Analysers:

The Orbitrap analyser was initially developed by Makarov in 2000[49] building upon the idea of trapping ions around a central spindle as initially formulated by Kingdon[50]. This type of mass spectrometer was then developed due to the potential for increased sensitivity with this type of analyser. This culminated in the development in a commercially available Orbitrap as detailed by Cooks in 2005[51]. The analyser measures the m/z from their oscillating frequencies allowing for much higher resolution than traditional ToF analysers making them vital for low mass measurements. This can be seen in particular with the initial use of Orbitrap mass spectrometers in proteomics work[52, 53, 54]. In 2012 it was first shown that it was possible to adapt an orbitrap mass spectrometer for use for high mass protein complexes[55].

Orbitrap analysers work by using a wire stretched along the axis of the outer cylinder with flanges sealing the trapping volume. With a voltage applied between the wire and the cylinder ions are attracted to the wire and only those with sufficient tangential velocity miss the wire and form a stable orbit around the wire[49]. In addition to the wire providing an axis for rotation of ions, the flanges of the outer cylinder add a curvature to the field keeping the ions restricted to the length of the cylinder.

Whilst theoretically the m/z for an ion could be determined from its frequency of rotation about the central axis this method would lead to poor resolution. This lack of resolution results from the strong dependency of this rotational frequency on the initial velocity of the ion upon entering the cylinder, and the initial radius of rotation. However developments in the field lead to the possibility of determining the m/z of any ion by using the frequency of oscillation along the axis of the field instead. Through the addition

of a modified outer electrode it is possible to introduce an axial quadrupole term which allows for controlled oscillation of ions along the axial axis of the analyser. The benefit of this modification is that it creates a measurable oscillation whose frequency is governed solely by the m/z ratio of the ion in question and is independent of any of the initial properties of the ion apart from its m/z.

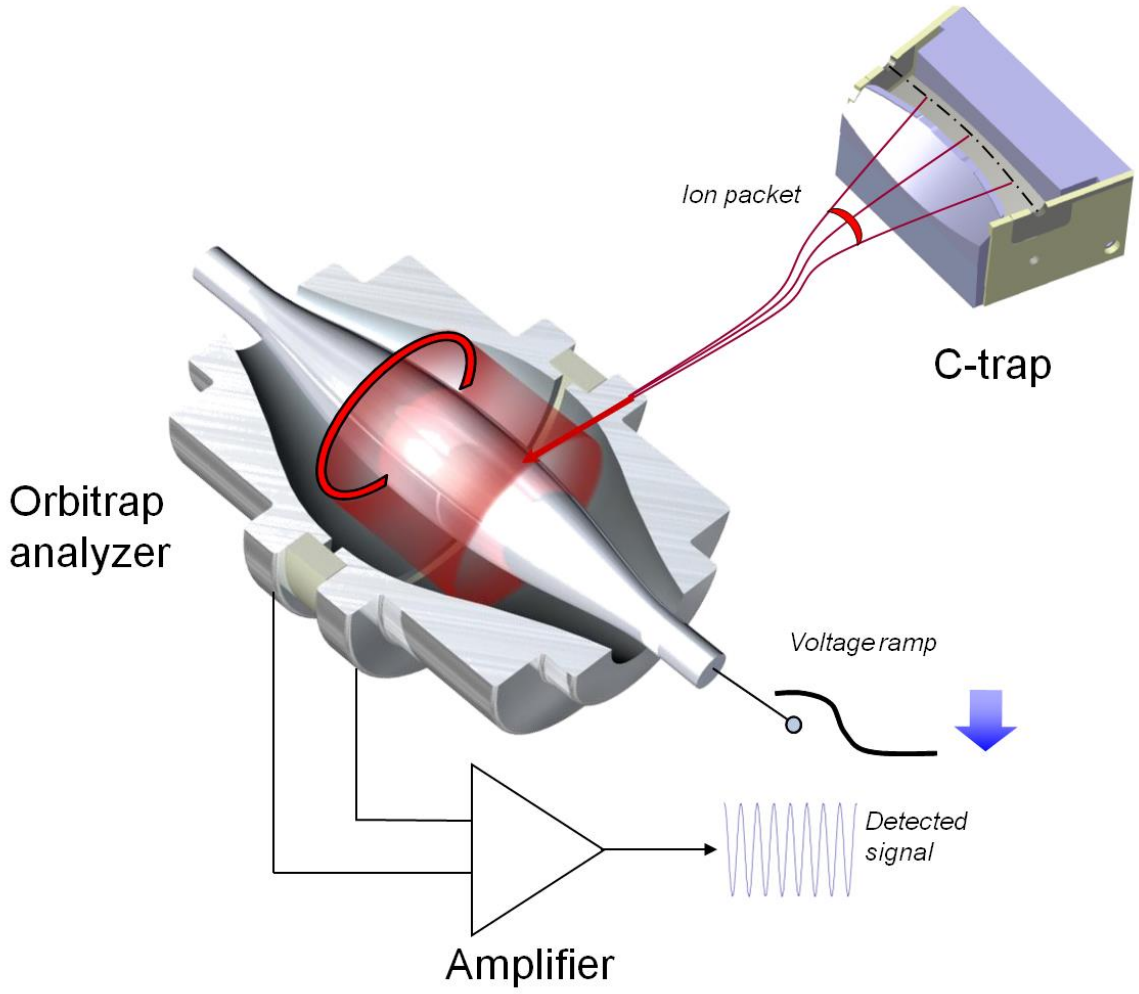


Figure 1-7: Schematic of an Orbitrap analyser with a C-trap which allows for the collection of the ions before injecting them into the Orbitrap for analysis, taken from Thermo Scientific.

The primary benefit of the orbitrap analyser over more traditional mass analysers such as ToF and quadrupoles is the increase in resolution. This is possible due to the ability to

accurately determine the axial frequency of oscillation which is independent of the energy and positions of the ions in the analyser.

1.3.4. Tandem Mass Spectrometry (MS-MS):

Tandem mass spectrometry is MS where there are multiple stages of selection of m/z for an ion during a single experiment. One of the main techniques utilised in MS-MS is collision induced dissociation (CID) which uses collision between a collision gas and the sample ion to fragment the ion. The technique is particularly useful when samples contain multiple complexes with only the use of CID there is no way of determining which ions have come from which initial species or if an ion existed in the sample before ionisation, or is a genuine dissociation product. In order to enhance the information obtained from CID experiments tandem mass spectrometry (MS-MS) was developed. This technique takes advantage of the use of multiple analysers in a single instrument leading to multiple levels of analysis. Typically with the use of a quadrupole analyser to initially select the ion of interest, then a collision cell to perform the dissociation experiment and then a second analyser, often a ToF analyser to record the resulting fragments.

The obvious benefit here lies in being able to trace the fragmented ions back to which parent ions they came from. This allows for the certainty that some ions are dissociation products as opposed to ions that exist in solution. For this reason MS-MS is an increasingly popular technique in the field and is becoming more readily available in commercial mass spectrometers.

1.3.5. Collision Induced Dissociation (CID):

Collision induced dissociation (CID) is a development in MS which takes advantage of a collision cell built into the mass spectrometer designed to allow for additional energy to

be applied to the analyte ions as they pass through. The cell used is filled with a collision gas, typically Argon, for the ions to collide with and then a voltage is applied across the cell to accelerate the ions into the gas molecules to cause collisions. Due to the nature of this process the addition of energy to the analyte ion is a slow cumulative one. This means that the kinetic energy from these collisions can then be converted into internal energy leading to unfolding of the ion in question and then the breaking of non-covalent interactions causing the loss of part of the analyte. Typically this means that the products from CID result from the loss of one highly charged subunit leaving a charge reduced complex behind[56]. This usually means that with increasing energy being added into a CID experiment you will see the loss of a first then a second subunit as opposed to seeing the loss of a dimer in a single shot. The combination of CID and multiple analysers in a mass spectrometer allows for the possibility of MS-MS as discussed later.

1.4. Detectors:

Two different detectors were utilised in obtaining the data detailed in this thesis. In the quadrupole-ToF hybrid MS instruments the detector is a microchannel plate (MCP) detector, in the Orbitrap MS instrument ions are detected by measuring their image current on the outer electrode.

1.4.1. Microchannel Plate (MCP) Detectors:

MCP detectors consist of an array of 10^4-10^7 miniature electron multipliers parallel to each other on a glass plate[57]. These channels allow for a large electron multiplication factor on a short timescales, due to the large number of collisions of the ion with the semiconducting wall of the channel. They work because a single collision between an ion and the wall of a single electron multiplier causes the release of multiple electrons, these

in turn then hit the wall and release more electrons causing a large multiplication of the signal, allowing for easier detection of the impact of the ion **(Figure 1-8)**.

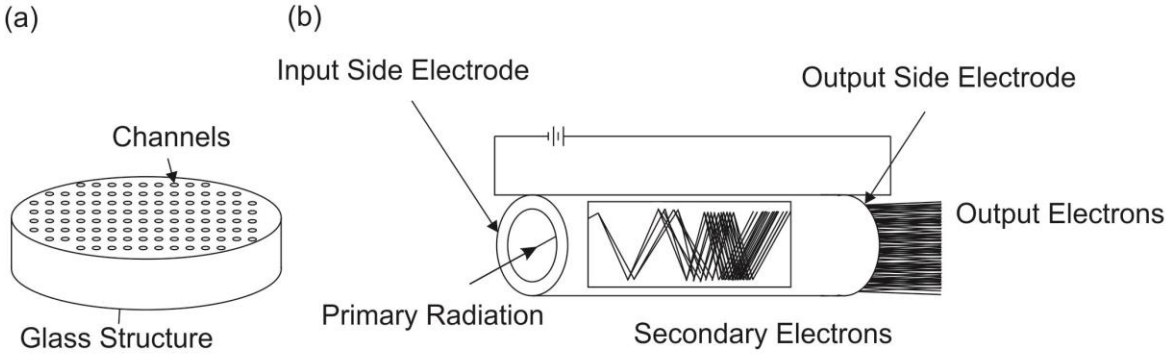


Figure 1-8: a) shows a simple view of a MCP detector while b) shows a zoomed in view of a single electron multiplier showing the process of electron release after initial impact with the wall of the channel. Figure adapted from Hoffmann et al[57].

1.4.2. Orbitrap Detectors:

The Orbitrap detects ions through measuring the image current on the outer electrode. This process of detection works by utilising the fact that as the ions orbit between two electrodes induces a radio frequency image charge in the electrodes which is then detected as an alternating image current in the detection unit which is connected to the electrodes themselves. This is then run through a fast Fourier transform algorithm which converts the recorded time-domain signal to a m/z spectrum[51].

1.5. Advances in Mass Spectrometry for Non-Covalent Mass Spectrometry:

With the growth of the field of non-covalent MS, additional techniques were developed to allow for more information to be obtained while analysing a sample. These developments include but are not limited to; nano-electrospray ionisation (nESI), collisional cooling and ion-mobility. Each of these developments on their own lead to a

significant improvement in the quality of information obtainable by mass spectrometry, when combined they constitute a significant advance in the scope of the field.

1.5.1. Nano-Electrospray Ionisation:

With ESI-MS becoming more common place as the field grew some of the principle disadvantages of this ionisation technique became more obvious. For example the requirement to have the source sealed in a nitrogen atmosphere and also the issue of picking up non-specific interactions. These interactions arise from the large initial droplets formed in ESI being so big that it is difficult to ensure that only a single analyte molecule is present per droplet upon entering the instrument. This can be seen in the example spectra below where GroEL, which is usually a dual-ringed tetradecamer, shows mainly as one distribution corresponding to the tetradecamer in the nESI spectrum **(Figure 1-9 upper)** but as sets corresponding to tetradecamer, 28-mer and 42-mer in the ESI spectrum **(Figure 1-9 lower)**[56].

Nano-electrospray ionisation (nESI) was developed as a technique to overcome these problems. With the use of a smaller initial capillary the initial droplet size is significantly smaller and so far fewer fission events are required before the desired one analyte molecule per droplet criterion is satisfied. In addition to the improvements gained through nESI in terms of the reduction of non-specific interactions it also requires a smaller aperture for the source allowing the instrument to be left open to normal atmosphere. The final benefit of nESI over ESI can be seen in the spectrum below with sharper peaks seen as opposed to the broader ones observed for standard ESI[56].

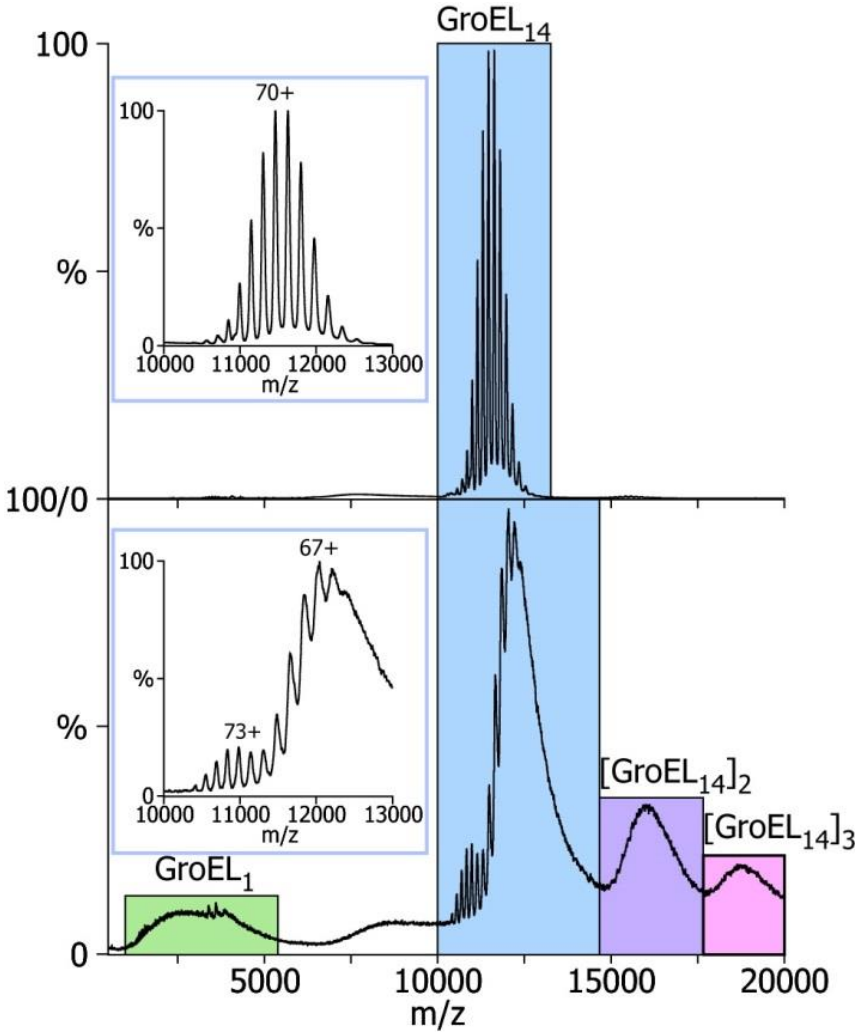


Figure 1-9: Comparison of mass spectrum for GroEL obtained using nESI (upper) and using ESI (lower).[56]

One of the practical benefits in using nESI as opposed to ESI as the ionisation method for a mass spectrometry experiments lies in the significant reduction in the aperture into the instrument. This in turn removes the need for the system to be contained under nitrogen instead it is possible to leave the instrument open to a laboratory environment. Whilst this does facilitate an easier experimental setup it also has an unintended consequence with regard to the spectra. With many nESI mass spectrometers being open to the environment compared to many ESI being under nitrogen there is a slight charge state effect due to the change in atmosphere as seen in **Figure 1-9**.

1.5.2. Collisional Cooling:

In the early days of native protein mass spectrometry it was reported that, despite the theoretical infinite limit of a ToF mass spectrometer, at normal pressure settings no ions could be detected for complexes over 100 kDa[58]. However it was noted that when the pressure was increased it became possible to observe signals corresponding to these higher mass complexes[59, 60, 61].

The reason for the emergence of these signals transpired to be due to collisional cooling, which is the process by which an ion undergoes collisions with gas molecules, causing its kinetic energy to be dissipated. This allows for better transmission of the ions through the mass spectrometer due to the high amount of energy the ions have upon entering the instrument due to gas expansion. For large molecules this increase in energy would cause them to have excessive axial and radial kinetic energy in the initial part of the instrument drastically reducing transmission as they will not be focussed enough to either exit the aperture at the end of the quadrupole or may cause them to miss the detector[62]. **Figure 1-10** shows the effect of collisional cooling on the transmission of ions depending on the size of the ion in question.

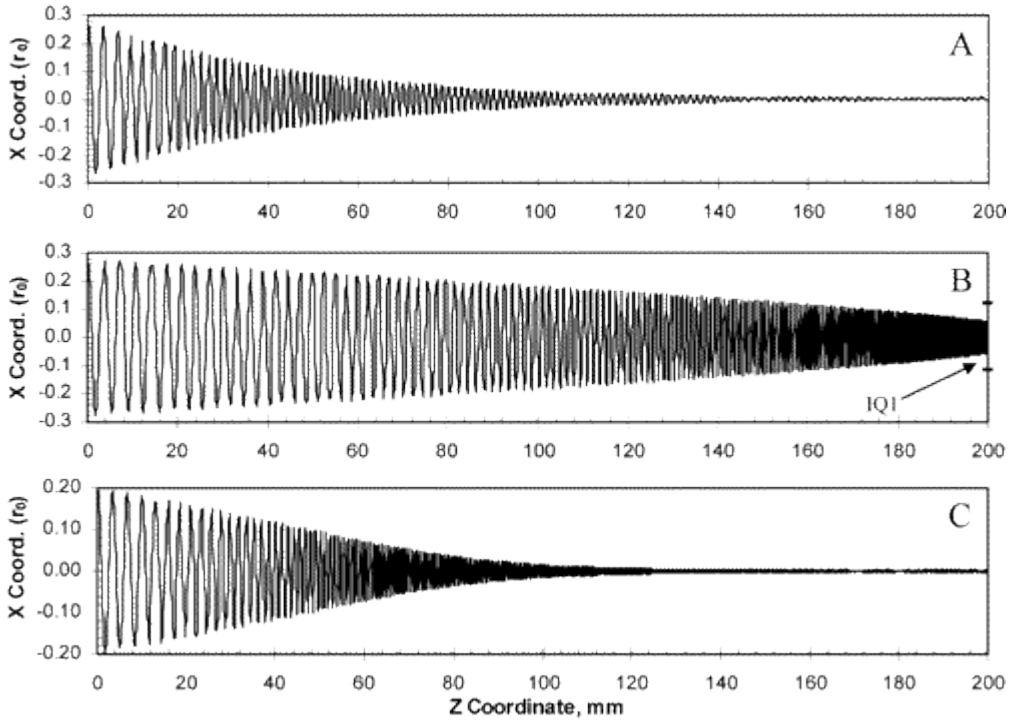


Figure 1-10: Simulations of ion trajectories in Q0 (XZ plane). (A) Myoglobin ion (M = 16 951 Da, σ = 2,520 Å2) at 8 mTorr; (B) proteasome 20S ion (M = 692 kDa, σ = 19 400 Å2) at 8 mTorr; (C) proteasome 20S ion at 30 mTorr.[62]

The higher the pressure used in the initial source cell of the instrument the greater the effect of collisional cooling will be, this however comes at a price as utilising too high a pressure in any one compartment of the instrument may cause a severe reduction in transmission as the collisions hinder the ions progress through the mass spectrometer. This leads to a balancing act which is dependent on the species of interest to try and optimise transmission dependent mainly on the size of the complex. Pressure settings optimised for a protein such as myoglobin will lead to very low transmission for a much larger protein for example proteasome 20S and vice versa as shown by Chernushevich[62], however with an understanding of collisional cooling it opens up the possibility of being able to using mass spectrometry for larger complexes.

1.5.3. Ion Mobility Mass Spectrometry (IM-MS):

Ion mobility mass spectrometry (IM-MS) is a technique which allows for information about the protein's shape and size to be obtained. The addition of mobility cell, filled with a known pressure of a suitable buffer gas, into a mass spectrometer is necessary for a mobility measurement to be undertaken. The time taken for ions to traverse the drift cell allows for a measure of the mobility due to the addition of a buffer gas in the cell. The ions passing through the cell will undergo collisions with this buffer gas hampering their progress, with smaller molecules tending to undergo fewer collisions, and therefore traversing the cell faster than their larger counterparts, which undergo more collisions and therefore tend to traverse the cell slower. The drift time for an ion traversing the cell is defined as t_d as below in *Equation 1.4*[22] where t_{d0} is the drift time for an ion outside the drift tube, which is dependent on the duty cycle of the instrument and the m/z of the ion, L is the length of the mobility cell, V is the voltage applied across the cell and K is the mobility constant of the ion[22].

(Equation 1.4)

$$t_d = \frac{L}{VK} + t_{d0}$$

There is a known relationship between the mobility constant (*K*) and the collision cross-section of an ion as shown in *Equation 1.5*. Through the use of this fact and the measured drift time for an ion it becomes relatively simple using IM-MS to obtain information about an ions collision cross-section[63].

(Equation 1.5)

$$K = \left(\frac{3q}{16N}\right)\left(\frac{2\pi}{kT}\right)^{1/2}\left(\frac{m+M}{mM}\right)^{1/2}\left(\frac{1}{\Omega}\right)$$

Initially ion-mobility used a linear field in the drift cell where a constant voltage gradient was across the cell as the method for drawing ions through. However more recently newer instruments make use of travelling wave technology in the drift cell[64]. Whilst this technology does lead to better transmission of the ions in the mass spectrometer it can cause a number of issues when it comes to measuring the exact mobility for the ion in question. The downside to the use of travelling wave ion mobility mass spectrometry comes in the requirement for use of calibration for accurate collision cross-section measurements. The calibrants used also need to be chosen carefully for any measurements due to the nonlinear relationship between travelling wave ion mobility and drift tube ion mobility. This means there are dangers we need to be aware of when performing a calibration particularly if it is necessary to extrapolate outside the range the chosen calibrants span.

1.6. Fast Protein Liquid Chromatography:

One of the principle methods used for purification of proteins before analysis is fast protein liquid chromatography (FPLC) and is used in the work detailed in chapters 2 and 4. This technique allows for the injection of a sample onto a column, which can then be used to separate out the protein of interest from any other contaminants present. The three main types of chromatography used in this body of work are; size exclusion chromatography, affinity chromatography and ion exchange chromatography. Typically this type of purification takes place in either a cold room or cold cabinet in order to allow for the protein to be kept stable throughout the process. A typical set up for a FPLC is shown in **Figure 1-11**. The two buffers can be mixed and then pumped over the column allowing for an increase in a chemical utilised to elute the protein of interest, common

when performing affinity chromatography. When doing size exclusion chromatography typically only one buffer is used and the column alone performs the job of separating out the components in the solution. The UV/Vis and conductance detectors allows for the detection of the fractions which contain protein and can also record useful information on concentrations of salts in the solutions.

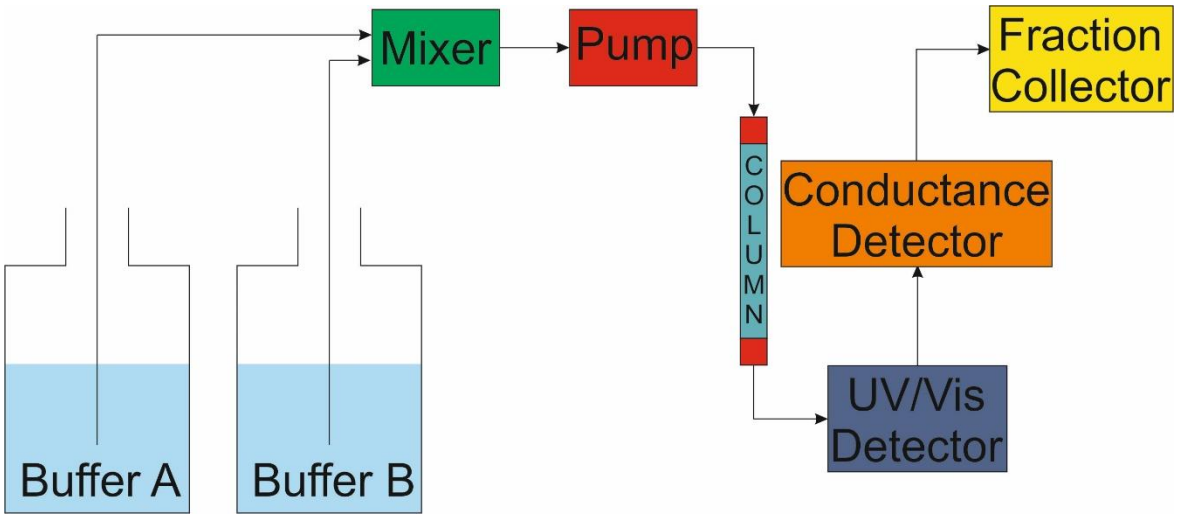


Figure 1-11: Cartoon demonstrating the basic setup of a FPLC experiment to allow for purification of proteins.

1.6.1. Size Exclusion Chromatography (SEC):

Size exclusion chromatography (SEC) is a chromatographic technique which allows for the separation of species by size. This is done using a column packed with resin that contains pores of different sizes. When the sample is injected onto the column the speed at which it travels down the column is dependent on its ability to be trapped by the various pores. The smallest species will be trapped by all the pores in the resin leading to it eluting later on during a run, in contrast the largest species will either not be trapped by any pores and therefore be eluted in the void volume or only by very few causing it to be eluted early on in a run. This leads to the possibility of separating a protein which exists as monomer and dimer into fractions containing solely monomer and those containing

solely dimer, provided that the timescale of exchange between oligomeric states is sufficiently slower than the timescale of the experiment.

Typically SEC is often used as a final stage in a purification process once the majority of contaminants have been removed using alternative purification methodologies. In addition SEC can also be used for buffer or detergent exchange, running a purified protein over a column pre-equilibrated in the desired buffer this allows for easy manipulation of the sample conditions without the risks involved in a simple Bio-Spin which can see proteins precipitate out.

1.6.2. Affinity Chromatography:

Affinity chromatography separates out proteins based on their binding affinity to the resin used. During the course of this work affinity chromatography was carried out using HisTrap columns (GE Healthcare) which are packed with Ni^{2+} ions which bind strongly with histidine residues in a protein. The basis of this technique lies in initially binding most of the proteins in a sample to the column initially and then adding in a competitor for the binding sites. With low concentrations of the competitor (Imidazole for HisTrap columns in this work) the weaker binding proteins are displaced from the binding sites and therefore eluted from the column. Increase concentration of the competitor will lead to the proteins with stronger binding affinities being displaced and therefore eluted.

Typically affinity chromatography is used as an early step in a purification protocol due to its ability to selectively remove a high percentage of unwanted products of any purification protocol. However it is worth nothing that this is rarely used as the only stage in a purification due to the fact there are always some proteins co-expressed which will have high numbers of histidines present causing them to also be eluted along with the

protein of interest meaning that a secondary purification step is usually optimal in order to guarantee the purity of the sample.

1.6.3. Ion Exchange Chromatography:

Ion exchange chromatography takes advantage of the different coulombic interactions of proteins and other molecules to allow for separation. Both cation and anion exchange are possible where the addition of salt to the buffer can lead to elution of the protein of interest due to either the Na^+ or Cl^- causing the displacement of the protein from the column.

This works by using a column loading the protein sample onto a column packed with strong ion-binding resins. This causes all proteins in the sample to bind to the column initially. This is typically followed by a gentle gradient of salt addition allowing for the slow elution of proteins with differing strengths of ion-interactions. Similarly to affinity chromatography this technique is usually used as an early step in a purification process before a second or third process to guarantee purity.

1.7. Dynamic Light Scattering (DLS):

Dynamic light scattering (DLS) is a technique that takes advantage of the way in which light can interact with particles in solution. The measurement is made using a polarised laser beam which is then passed through the sample, then polarised again before the resulting diffraction pattern is recorded using a photomultiplier. This pattern is known as a speckle pattern, and multiple versions of this pattern are collected over time and the resemblance of each pattern to its predecessors can be interpreted to give valuable information about the sample being analysed. The resemblance of a pattern to its predecessors is quantified by an auto-correlation function which has been shown to be

related to the diffusion co-efficient of a sample as shown in *Equation 1.6* with a Bragg wave vector as defined in *Equation 1.7*[65]:

(Equation 1.6)

$$g^2(\tau) - 1 = e^{(-D^0 q^2 \tau)}$$

(Equation 1.7)

$$q = (4\pi n/\lambda)\sin(\theta/2)$$

This shows the relationship between the auto-correlation function ($g^2(\tau)$) and the diffusion co-efficient D_0 at a time τ after the initial speckle pattern was recorded as well as the Bragg wave vector (q). This diffusion co-efficient quantifies the speed of movement of the particles inside the solution, the larger the diffusion co-efficient the more motion that will be occurring in solution and so the quicker the auto-correlation function will decay towards zero[66].

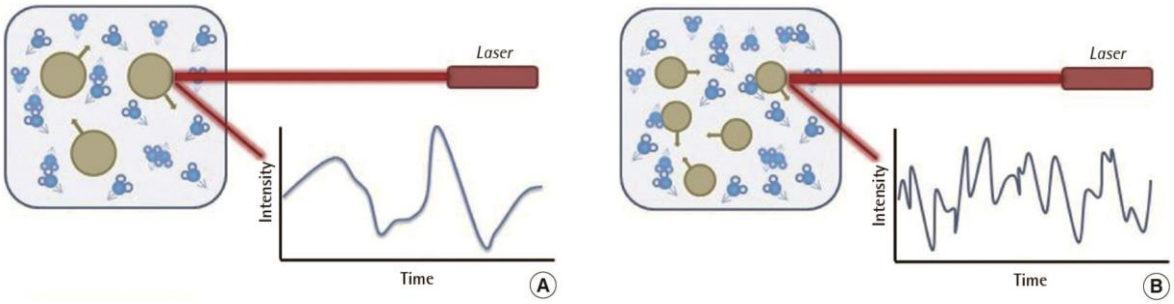


Figure 1-12[66]: Cartoon showing the differences in the rate in change of the intensities of speckle patterns depending on the ease of diffusion for the particle being studied. In a) a large slow diffusing particle is shown whereas in b) a smaller particle is shown with a much faster changing pattern.

Using *Equation 1.6* it is possible to fit the auto-correlation function data and determine the diffusion co-efficient of the particle in solution as shown in **Figure 1-13**. This technique can be particularly powerful when multiple measurements are taken in quick succession allowing for a high degree of confidence in the values calculated.

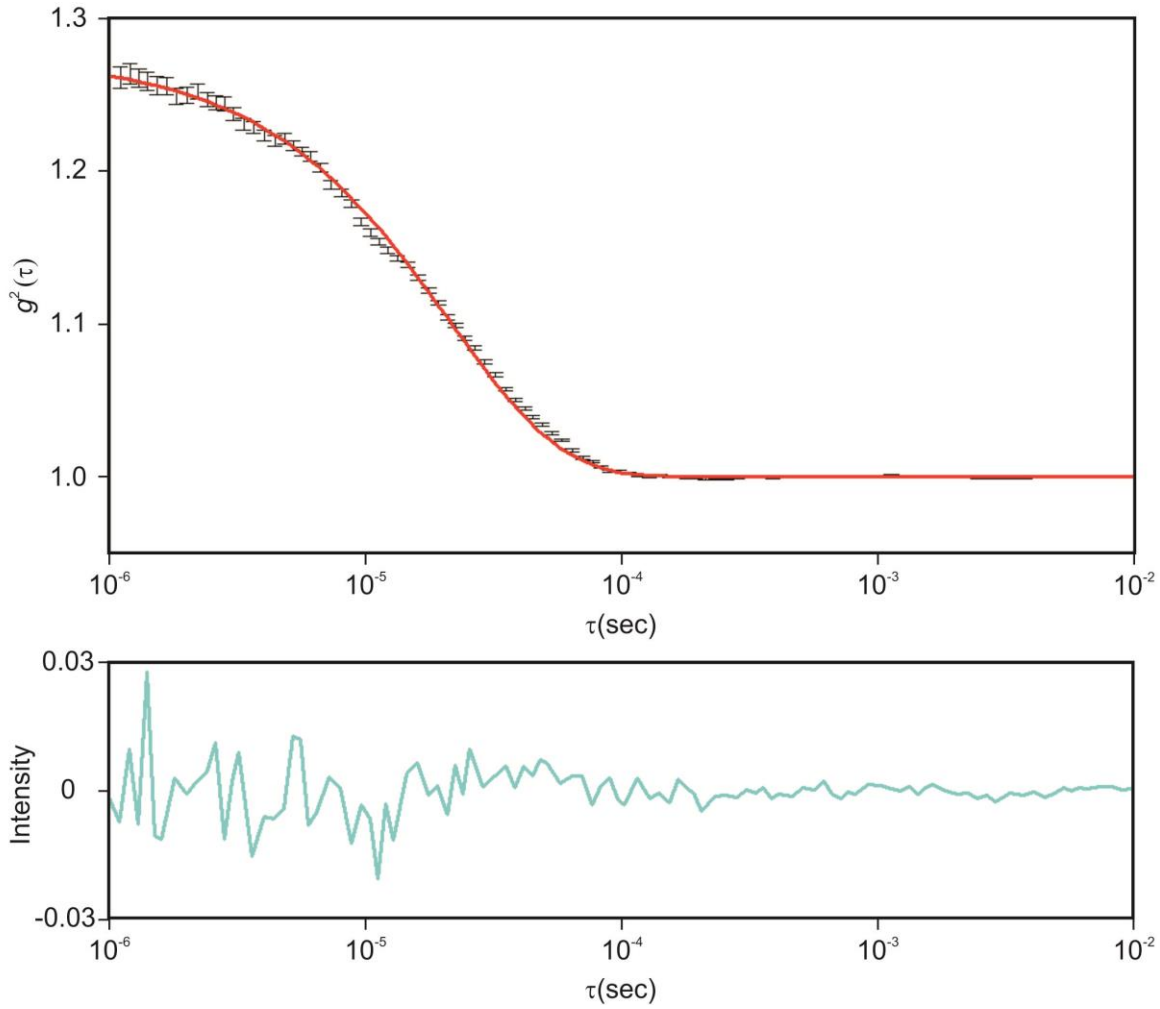


Figure 1-13: Auto-correlation function for one point for the alcohol dehydrogenase tetramer is shown in the upper panel with the fit (red) as well as individual points plotted with error bars (black dots). Lower panel shows the residuals at each point (green).

From the diffusion co-efficients calculated from the auto-correlation functions it is possible to determine the Stokes radius for the protein[67]. This is done using the Stokes-Einstein equation, where R_S is the Stokes radius, k_B is the boltzmann constant, T is the temperature in Kelvin, D is the diffusion co-efficient in Ficks units and η is the viscosity. The temperature of the experiment was also recorded to allow for the determination of the viscosity which in turn allows for the eventual calculation of the Stokes Radius:

(Equation 1.8)

$$R_s = \frac{k_B T}{6\pi\eta D^{\circ}}$$

1.8. Computational Techniques for Calculation of Collision Cross-Sections (CCS)

Since the original development of IM-MS there has been a growing interest in the ability to determine collision cross-sections for proteins for which the structure is already known. The initial interest in this was to confirm that proteins were maintaining native structure despite being subjected to the vacuum inside the mass spectrometer. Initially this was done by using methods which were not initially designed for use with proteins, for example the projection approximation was originally for use with clusters. However as the field grew new approaches were developed specifically aimed at modelling the interactions which take place during IM-MS. These methods which are more targeted at modelling the IM-MS process have led to more rigorous calculations which more accurately encompass the entirety of the process.

1.8.1. Projection Approximation (PA):

The projection approximation was first developed as a method by Bowers in 1993[17] to model the mobilities of carbon clusters during an ion-mobility experiment. This technique was later on adapted for using with proteins and their complexes but it is important to remember that the technique was initially developed for use with smaller less structurally complex carbon atom clusters.

The projection approximation works by taking the known structure of a protein from x-ray, NMR or EM, orientating it randomly around its centre of mass using three random Euler angles. The structure of the protein is then projected onto the xy plane and the cross-section determined using a Monte Carlo integration[17]. In the Monte Carlo method a two-dimensional box is placed around the projection of the protein, with edges tangent

to the extreme points of the projection in the x and y planes, and points inside the box are chosen randomly. If the point chosen is within the combined van-der-Waals radius of helium and the closest atom, to the nearest atom's centre of mass then a collision is deemed to have taken place. The cross-section is determined by the ratio of collisions to misses multiplied by the area of the box. This is then normally averaged over a number of different orientations to give an average collision cross-section.

1.8.2. Exact Hard Sphere Scattering (EHSS) Method:

The exact hard sphere scattering method was first developed in 1996 by Jarrold[18]. This was developed with the idea of providing a more rigorous methodology for modelling the interactions between ions in a drift tube and a buffer gas during the ion-mobility experiment with a reduced level of computational complexity when compared to the trajectory method.

The value in using this method over a simple projection approximation is it attempts to evaluate the collision integral for the subject in question rather than determining the cross-section based on a simple hit or miss philosophy. By considering the result of an impact of a buffer gas molecule with the ion traversing the mobility cell it automatically brings into consideration the possibility of multiple collisions between a single buffer gas molecule and the ion being studied. This effect is particularly important where the ion has a concave surface where the difference between measured cross-sections for PA and EHSS can be more than 20%.

The basis of all methods that advance upon the PA method work on trying to evaluate the orientationally averaged collision integral ($\Omega_{avg}^{1,1}$) which has a defined relationship with the zero-field mobility of an ion (K) as shown in *Equation 1.9* [18]:

(Equation 1.9)

$$K = \frac{(18\pi)^{1/2}}{16}\left[\frac{1}{m}+\frac{1}{m_b}\right]^{1/2}\frac{ze}{(k_BT)^{1/2}}\frac{1}{\Omega_{avg}^{1,1}}\frac{1}{N}$$

With the zero field mobility of an ion easily calculable from the drift time recorded, as shown earlier in *Equation 1.4,* and the rest of the constants in the equation above known or are easily obtained from the IM-MS experiment. To calculate the orientationally averaged collision integral it is important to consider the collision geometry determined by θ, γ and φ, the scattering angle χ(θ,γ,φ,g,b). $\Omega_{avg}^{1,1}$ is then calculated by evaluating the integrals shown in *Equation 1.10* where g is the relative velocity and b is the impact parameter, μ is the reduced mass of the protein and the buffer gas:

(Equation 1.10)

$$\Omega_{avg}^{1,1} = \left(\frac{1}{4\pi^2}\int_0^{2\pi}d\theta\int_0^{\pi}d\varphi\,sin\varphi\int_0^{2\pi}d\gamma\frac{\pi}{8}\right)\times\left(\left(\frac{\mu}{k_BT}\right)^3\int_0^{\infty}dg e^{-\mu g^2/2k_BT}\,g^5\right)$$
$$\times\int_0^{\infty}db\;2b(1-cos\chi(\theta,\gamma,\phi,g,b))$$

When we consider both the buffer gas and the atoms making up the ion being analysed as hard spheres it allows for the reduction of the terms relating to the scattering angle in *Equation 1.9* to πb^2_{min} with b^2_{min} being equal to the hard sphere contact distance. In doing this you can get *Equation 1.11* as an approximation of the collision integral. However in this approximation you are ignoring both long range potential interactions and the details of the scattering process.

(Equation 1.11)

$$\Omega_{avg}^{1,1} \approx \frac{1}{4\pi^2}\int_0^{2\pi}d\theta\int_0^{\pi}d\varphi\,sin\varphi\int_0^{2\pi}d\gamma\,\pi b_{min}^2$$

In order to try and take into account the scattering that will take place during ion-mobility measurements this method solves *Equation 1.11*. This allows for a more accurate evaluation of the collision integral as opposed to the previous approximation which effectively calculates a projection of the molecule. Whilst this has been shown to be equivalent to the collision integral for a perfectly convex molecule it does not hold true for molecules with concave surfaces. Whilst the deviation of the cross section integral from the projection is small when there are small concave surfaces when the overall molecular geometry is concave it can lead to much larger deviations. Overall the EHSS method provides a more rigorous methodology for determining the collision integral than the projection approximation without the computational demands of the trajectory method.

1.8.3. Trajectory Method (TJM):

The trajectory method for determining a collision cross-section for an ion was first developed in 1996 by Mesleh[19]. The method was developed as an alternative to the available hard sphere projection approximation. The main thrust behind this new method was to attempt to account for the potential electrostatic interactions between the buffer gas and the ion being modelled.

The trajectory method is often considered the “gold standard” of computational methods, for determining the theoretical collision cross-sections for proteins via ion-mobility. Like all of the computational methods it attempts to try and estimate a collision cross-section for a protein structure. This methodology is based around taking a known protein structure and trying to simulate the environment they would experience during an ion-mobility measurement.

The largest difference between this method and other computational methods is in attempting to account for long range potential interactions between the protein and the buffer gas. However it does, like many of the other methods available, also consider the effect of ion structure and surface roughness. This initially works from the zero field mobility and then incorporates elements to model the electrostatic interactions involved. The key difference with the trajectory method here comes in the treatment of collision integral where the scattering angle is treated differently here.

To take into account the effect of the electrostatic interactions between the buffer gas and the ion a Lennard-Jones potential is modelled. This is modelled using *Equation 1.12* shown below which is used to determine the potential work ($\phi(\theta, \varphi, \gamma, b, r)$), where ϵ and σ are the Lennard-Jones parameters of well depth and distance at which the potential becomes positive respectively and α is the polarizability of the buffer gas:

(Equation 1.12)

$$\phi(\theta,\varphi,\gamma,b\gamma,r) = 4\epsilon\sum_{i}^{n}\left[\left(\frac{\sigma}{r_i}\right)^{12} - \left(\frac{\sigma}{r_i}\right)^{6}\right] - \frac{\alpha}{2}\left(\frac{ze}{n}\right)^{2}\left[\left(\sum_{i}^{n}\frac{x_i}{r_i^3}\right)^{2} + \left(\sum_{i}^{n}\frac{y_i}{r_i^3}\right)^{2} + \left(\sum_{i}^{n}\frac{z_i}{r_i^3}\right)^{2}\right]$$

This equation allows for the modelling of both the two body 6-12 interactions and the ion-induced dipole interactions. The two body 6-12 interactions are modelled in the first term whilst the second term models the ion-induced dipole interactions. Experiments done where the long range potential interactions between the buffer gas molecules, as well as their scattering, are taken into account compared to a hard sphere projection approximation show significant discrepancies. These experiments were done using carbon fullerene clusters but it is thought that the discrepancies between the TJM and

hard sphere are likely to be exacerbated with increasing size of molecules being studied and with increasing polarizability of the molecule in question.

1.8.4. Projected Superposition Approximation (PSA):

In 2011 Bleiholder proposed the projected superposition approximation as a method for determining collision cross-section for an ion-mobility experiment[20]. The idea being to maintain the rigorous nature of the trajectory methods whilst reducing the computational power required to perform the calculations. This methodology was designed to produce accurate results for the cross-sections of larger ions such as proteins. In order to do this, this method makes use of two effects it coins the "size effect" and the "shape effect".

The size effect is used to describe the way in which the effective atomic radius of an atom can be increased due to its environment. The shape effect is the effect caused by the geometry of the molecule in question. This is a particularly crucial effect when it comes to molecules with concave surfaces which leads to the possibility of multiple collisions between a single buffer gas molecule and the ion undergoing IM-MS.

In order to achieve the goal of a highly accurate method of determining the collision cross-section without the computational complexity that is associated with the TJM this method builds upon the relatively simple PA method. This is done by first incorporating the size effect into the orientation averaged projection area ($\langle\Omega_{PSA}\rangle$) and then by multiplying this by a shape factor (ρ) as shown in *Equation 1.13*:

(Equation 1.13)

$$\Omega_{PSA}^{(1,1)}(T) = \langle\Omega_{PSA}\rangle \times \rho$$

The averaged projection area here differs from the projection given by the PA method by replacing the atomic hard-sphere radius for an atom with a probability of an atomic collision. The probability of a collision for an atom *a* with a buffer gas molecule is defined as p_a(T,x) as shown in *Equation 1.14* where r=|x-r_a| and $q_{a,coll}(T;r_{LJ}\varepsilon_{LJ}) = \sqrt{\omega_a^{(1,1)}}/\pi$ is the atomic collision radius of atom *a* defined by the momentum transfer integral $\omega_a^{(1,1)}$ for an isolated atom colliding with a buffer gas particle *X*. The interaction between the atom *a*, and the buffer gas particle *X* can be modelled using a (12,6,4) potential using the Lennard-Jones parameters r_{LJ} and ε_{LJ} the well position and depth respectively.

(Equation 1.14)

$$p_{a(T,x)=}\begin{cases} 1, & r \le q_{a,coll}(T;r_{LJ}\varepsilon_{LJ}) \\ \alpha e^{-(r-q_{a,coll})k_1 t^{l_1}} + (\alpha-1)e^{-(r-q_{a,coll})k_2 t^{l_2}} & r > q_{a,coll}(T;r_{LJ}\varepsilon_{LJ}) \end{cases}$$

This in turn leads to the cumulative collision probability *P(T,x)* for a system of *m* atoms can be defined as shown in *Equation 1.15*:

(Equation 1.15)

$$P(T,x) = \sum_{a=1}^{m} p_a(T,x)$$

This is turn allows for the orientationally averaged projection area to be define as shown in *Equation 1.16*:

(Equation 1.16)

$$\langle \Omega_{PSA} \rangle = \frac{1}{n} \sum_{i=1}^{n} \Omega_i \left[\sum_{a=1}^{m} p_a, (T,x) \right]$$

Whilst the above allows for the consideration of the size factor it does not account for the shape factor. In order to do this a simple ratio is used comparing the effective surface

area of the molecule (A_{mol}) to the area of convex envelope of the molecule (A_{CE}). This ratio gives the shape factor (*p)* as shown below in *Equation 1.17*:

(Equation 1.17)

$$p = \frac{A_{mol}}{A_{CE}}$$

The combination of these two gives the overall projection superposition approximation as define earlier in *Equation 1.13*.

1.9. Protein MS:

One of the primary challenges in protein MS is the transfer of the protein to the gas-phase. However this challenge has been addressed, mainly through the use of volatile solvents such as ammonium acetate to assist the transfer of the protein of interest to the gas-phase as well as other advances in the field mentioned earlier like nESI as an ionisation technique. This selection of a MS compatible solvent allows for the use of MS for the study of many proteins with relative ease. The only other issue it is important to be wary of during protein MS is the concentration of the sample. High concentrations can lead to non-specific binding events been observed and lead to difficulties in achieving a stable electrospray of the protein as it may crystallise in the tip of the capillary.

1.9.1. Soluble Protein MS:

Initial work using native MS for analysis of proteins focussed on soluble proteins due to the relative ease of handling them compared to their insoluble counterparts. However there are many questions with regard to the validity of the technique due to the interaction between hydrophilic regions of the soluble proteins becoming exposed to a vacuum inside the mass spectrometry. This is why many mass spectrometry experiments

are done in conjunction with other experimental techniques such as cross-linking[68, 69] and x-ray crystallography[70, 71] to corroborate the data.

1.9.2. Membrane Protein MS:

The primary challenge one is faced with when performing MS of membrane proteins is similar to the challenges faced for most other techniques used when working with membrane proteins. This issue is their lack of solubility in typical aqueous solutions due to the high degree of hydrophobicity of the membrane embedded part of the protein. In order to overcome this problem there is a necessity for the use of a solubilising agent in order to keep the membrane protein stable while in solution. This leads to a compromise between choosing the best solubilising agent for stabilising the protein and choosing the optimal solubilising agent for MS. Traditionally detergents have been used to solubilise membrane proteins and were also used when the first mass spectra for a membrane protein was recorded in 2008[25]. However the problem that arose is the difficulty in maintaining the correct protein stoichiometry whilst still removing the detergent bound to the protein.

In an effort to find a way around this problem many alternative vehicles for membrane protein MS have been used including bicelles, nanodiscs and amphipols[32]. However, whilst these methods have some advantages over detergent with regard to creating a more native like environment for the protein, they still require the characteristic high energies in order to strip them away and reveal the naked membrane protein. There is a brief discussion of current techniques that have been used for membrane protein MS below including a small discussion on their relative advantages and disadvantages.

1.9.2.1. Detergents:

Initial studies showed that it was possible to get good quality mass spectra from some detergent solubilised membrane proteins despite the high energies required to strip away the micelle[72]. The detergents make use of their hydrophobic tails to stabilise the transmembrane region of the protein shielding it from unfavourable interactions with water (**Figure 1-14)**. The hydrophilic heads of the detergent point outwards into the solvent providing a favourable interaction allowing for the creation of stable protein-detergent micelle.

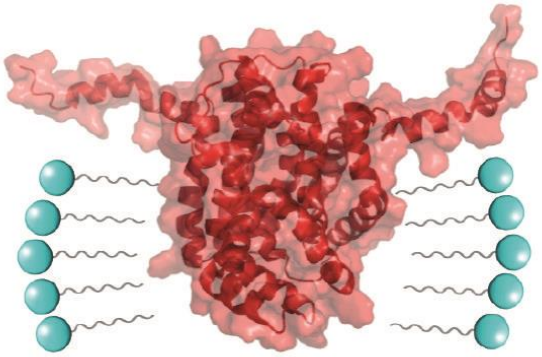


Figure 1-14: Cartoon illustrating a detergent solubilised membrane protein[32]. The teal spheres represent the hydrophilic head groups of the detergent whilst the grey spirals represent the hydrophobic tails. The membrane protein itself is shown in red with the transmembrane region solubilised by the detergent.

The advent of new charge reducing detergents have made it easier to remove the detergent micelle at much lower energies enabling the preservation of some of the non-covalent interactions present in membrane protein complexes[73]. For example it has been shown that with the use of C8E4 instead of a more traditional detergent, such as DDM, it is possible to see the apo protein for the multidrug and toxic compound extrusion protein, from *Pyrococcus furiosus*, with as little as 10V applied in the collision cell. This is far lower than the 100V that is normally required as a minimum for the same protein using the more traditional DDM[73].

In spite of these advances enabling better quality data to be collected there remains question marks over the suitability of detergents as a mimic for a lipid bilayer that a membrane protein usually exists in. Detergents solubilise membrane proteins by creating a micelle around them with the hydrophobic tails pointed in towards the transmembrane region of the protein to allow for minimisation of unfavourable interactions of the hydrophobic regions with water. This is very different to their native environment where the lipids in the bilayer would be lined up parallel to the protein inserted inside it.

1.9.2.2. Bicelles:

Bicelles are another of the vehicles that has been used to solubilise membrane proteins and then utilised for MS. Bicelles are a form of model lipid bilayer where two different lipids are used one which forms a bilayer around the membrane protein of interest and the other which form amphiphatic caps on the ends of the bilayer more similar to a detergent micelle **(Figure 1-15)**. Here the lipids align alongside the transmembrane region as opposed to having their hydrophobic tails pointing in forming a sphere around it like detergents. Whilst the lipids line up differently the detergent caps point in to shield the hydrophobic part of the bilayer in the same way as they shield the transmembrane region of a membrane protein in a micelle.

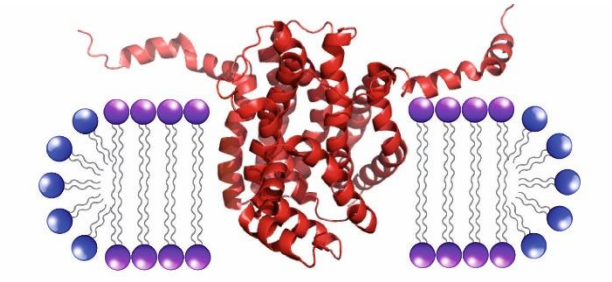


Figure 1-15: Cartoon illustrating a membrane protein solubilised inside a bicelle[32]. The lipids forming the bilayer are shown with purple head groups while the lipids forming the amphiphatic caps have blue head groups, again the membrane protein is shown in red with the transmembrane region solubilised by being buried inside the lipid bilayer.

The obvious theoretical advantages of bicelles, over simple detergent solubilisation, is that the environment the protein experiences is far more similar to the native environment of a cell membrane, due to the fact that the protein has lipids aligned along the transmembrane axis. However it has been demonstrated that these bicelles require a large amount of energy in order to release the membrane protein for analysis in the gas-phase[32].

1.9.2.3. Amphipols:

Amphipols are a class of surfactants that consist of a hydrophilic backbone with hydrophobic sidechains grafted on giving them an amphiphilic quality[74]. Amphipols have the benefit of stabilising membrane proteins in solution without the need for free amphipols in solution **(Figure 1-16)**[75, 76]. They have been shown to be capable of sustaining the activity of membrane enzymes in solution. Studies involving *E. coli* diacylglycerol kinase have shown that the activity for the enzyme solubilised with just amphipol have increased activity compared with the same membrane enzyme only solubilised using detergent[75]. However despite their ability to successfully solubilise and stabilise membrane proteins in solution sadly they are not ideal for MS. This is due to the fact that very high collision energies are required in order to remove the amphipol from the membrane protein making analysis difficult[32].

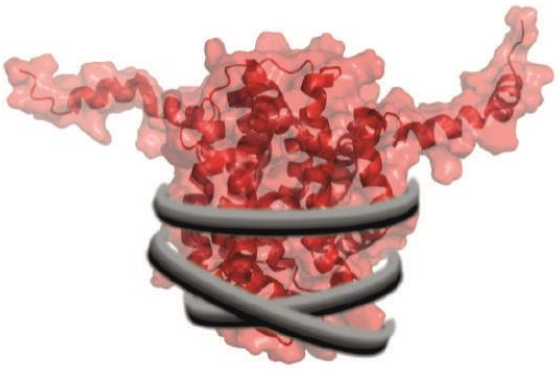


Figure 1-16: Cartoon illustration a membrane protein solubilised using amphipols[32]. The membrane protein is shown in red and is solubilised by rings of the amphipol shown in half grey half black rings. The grey half represent the hydrophobic side chains grafted onto the surfactant and are on the inside stabilising the hydrophobic transmembrane region whilst the black half represents the hydrophilic backbone which is free to interact favourably with the water in solution.

1.9.2.4. Nanodiscs:

Nanodiscs are made up of two components, a membrane scaffold protein (MSP), which has a ring structure, and lipids which fill into the centre of the MSP creating a disc[77]. Once again a principle theoretical advantage of using a nanodisc is to get away from the use of detergents and keep the protein solubilised with something more natural such as lipids. The presence of a MSP to constrain the lipids also can allow for some lateral pressure to be exerted in on the protein inserted replicating the normal environment in a lipid bilayer (**Figure 1-17)**.

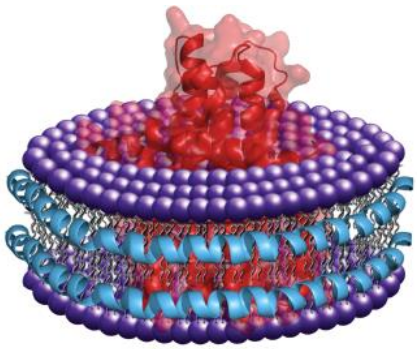


Figure 1-17: Cartoon representing a membrane protein solubilised using a nanodisc[32]. The blue alpha helix rings round the circumference of the disc represent the membrane scaffold protein, the purple denote the hydrophilic heads of the lipids with the hydrophobic tails hidden from view inside the disc and the red showing the membrane protein being solubilised.

Whilst the nanodisc seems a highly advantageous choice for MS, it requires high collision voltages in order to eject the membrane proteins inserted. Secondly the methods used for the preparation of the samples require the protein to be purified in detergent initially. The detergent is removed afterwards in conjunction with inserting the protein in to the nanodisc, meaning there are still concerns the interaction with the detergent could lead to some loss of native structure before it reaches the nanodisc.

1.9.2.5. Lipodisqs:

Lipodisqs closely resemble nanodiscs but with two major differences. Firstly instead of using MSP to constrain the disc a polymer is used. One of the more commonly used polymers is made of a mixture of styrene and maleic acid (SMA), and can stablilise a disc similarly to the way in which MSP can. Secondly the addition of the co-polymer to cell membranes will cause the co-polymer to act as a “cookie cutter” solubilising sections of the membrane[78].

The benefit of using lipodisqs over other techniques is clear as it has all the benefits of nanodiscs, coupled with the possibility of directly extracting the protein of interest from

its native environment including lipids. This benefit also means that the sample for MS can be made in a completely detergent free method. This means that there is no reason to believe that the protein should lose its native structure or stoichiometry. An example of a membrane protein solubilised inside a lipodisq is shown in **Figure 1-18**[33]**.**

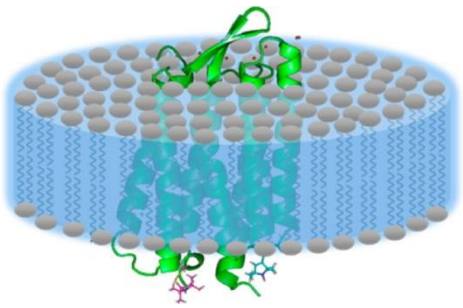


Figure 1-18: Cartoon showing a membrane protein solubilised inside a lipodisq[33]. The membrane protein is shown in green with the lipids in grey and the co-polymer is shown in light blue.

This technique has been shown to be generally effective at stably solubilising membrane proteins, with many membrane proteins already characterised inside these lipodisqs[33, 79]. It has been shown to be applicable to a number of different membrane proteins so far and it is wide applicability in combination with the ability to retain aspect of the native environment makes it an incredibly attractive tool. The work in this thesis details the use of this copolymer for membrane protein MS and addresses some of the challenges and opportunities this presents.

1.10. References:

1. Beavis, R.C. & Chait, B.T. Matrix-Assisted Laser-Desorption Mass Spectrometry Using 355 nm Radiation. *Rapid communications in mass spectrometry : RCM* **3**, 436-439 (1989).

2. Loo, J.A., Udseth, H.R. & Smith, R.D. Peptide and Protein-Analysis by Electrospray Ionization Mass-Spectrometry and Capillary Electrophoresis Mass-Spectrometry. *Analytical Biochemistry* **179**, 404-412 (1989).

3. Loo, J.A., Loo, R.R.O., Udseth, H.R., Edmonds, C.G. & Smith, R.D. Solvent-Induced Conformational-Changes of Polypeptides Probed by Electrospray-Ionization Mass-Spectrometry. *Rapid Communications in Mass Spectrometry* **5**, 101-105 (1991).

4. Gawinowicz, M.A., Merlini, G., Birken, S., Osserman, E.F. & Kabat, E.A. Amino-Acid Sequence of the FV Region of a Human Monoclonal IGM (NOV) with Specificity for the Capsular Polysaccharide of the Group B-Meningococcus and of Escherichia-Coli K1, Which Cross-Reacts with Polynucleotudes and with Denatured DNA. *Journal of Immunology* **147**, 915-920 (1991).

5. 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).

6. 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).

7. Hilton, G.R. & Benesch, J.L.P. Two Decades of Studying Non-Covalent Biomolecular Assemblies by Means of Electrospray Ionization Mass Spectrometry. *Journal of the Royal Society Interface* **9**, 801-816 (2012).

8. Henderson, S.C., Valentine, S.J., Counterman, A.E. & Clemmer, D.E. ESI/Ion Trap/Ion Mobility/Time-of-Flight Mass Spectrometry for Rapid and Sensitive Analysis of Biomolecular Mixtures. *Analytical Chemistry* **71**, 291-301 (1999).

9. Bushnell, G.W., Louie, G.V. & Brayer, G.D. High-Resolution 3-Dimensional Structure of Horse Heart Cytochrome-C. *Journal of Molecular Biology* **214**, 585-595 (1990).

10. Wilson, K.P., Malcolm, B.A. & Matthews, B.W. Structural and Thermodynamic Analysis of Compensating Mutations Within the Core of Chicken Egg-White Lysozyme. *Journal of Biological Chemistry* **267**, 10842-10849 (1992).

11. Maurus, R. *et al.* A Myoglobin Variant with a Polar Substitution in a Conserved Hydrophobic Cluster in the Heme Binding Pocket. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology* **1341**, 1-13 (1997).

12. Pjura, P.E., Lenhoff, A.M., Leonard, S.A. & Gittis, A.G. Protein Crystallization by Design: Chymotrypsinogen Without Precipitants. *Journal of Molecular Biology* **300**, 235-239 (2000).

13. Stein, P.E., Leslie, A.G.W., Finch, J.T. & Carrell, R.W. Crystal-Structure of Uncleaved Ovalbumin at 1.95 A Resolution. *Journal of Molecular Biology* **221**, 941-959 (1991).

14. Banci, L. *et al.* Solution Structure of Oxidized Horse Heart Cytochrome C. *Biochemistry* **36**, 9867-9877 (1997).

15. Schwalbe, H. *et al.* A Refined Solution Structure of Hen Lysozyme Determined Using Residual Dipolar Coupling Data. *Protein Science* **10**, 677-688 (2001).

16. Osapay, K., Theriault, Y., Wright, P.E. & Case, D.A. Solution Structure of Carbonmonoxy Myoglobin Determined from Nuclear-Magnetic-Resonance Distance and Chemical-Shift Constraints. *Journal of Molecular Biology* **244**, 183-197 (1994).

17. Vonhelden, G., Hsu, M.T., Gotts, N. & Bowers, M.T. Carbon Cluster Cations with up to 84 Atoms - Structures, Formation Mechanism, and Reactivity. *Journal of Physical Chemistry* **97**, 8182-8192 (1993).

18. 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).

19. 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. *Journal of Physical Chemistry* **100**, 16082-16086 (1996).

20. 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).

21. Berman, H.M. *et al.* The Protein Data Bank. *Nucleic Acids Research*, 235-242 (2000).

22. Bush, M.F. *et al.* Collision Cross Sections of Proteins and Their Complexes: a Calibration Framework and Database for Gas-Phase Structural Biology. *Anal Chem* **82**, 9557-9565 (2010).

23. Demmers, J.A.A., Haverkamp, J., Heck, A.J.R., 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 of the United States of America* **97**, 3189-3194 (2000).

24. Demmers, J.A.A., van Dalen, A., de Kruijff, B., Heck, A.J.R. & Killian, J.A. Interaction of the K+ Channel KcsA with Membrane Phospholipids as Studied by ESI Mass Spectrometry. *Febs Letters* **541**, 28-32 (2003).

25. 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).

26. Donlan, J.A.C., Gupta, K., Struwe, W.B., Gault, J. & Robinson, C.V. Probing the Effect of Lipid Binding on the Monomer-Dimer Equilibrium of a Prokaryotic Sugar Transporter by Native Mass Spectrometry. *Biophysical Journal* **110**, 423A-423A (2016).

27. Gupta, K. & Robinson, C. A Mass Spectrometric Platform to Reveal Lipid Mediated Oligomerisation of Membrane Proteins in Atomistic Detail. *Biophysical Journal* **110**, 424A-424A (2016).

28. Hansen, S.B., Tao, X. & MacKinnon, R. Structural Basis of PIP2 Activation of the Classical Inward Rectifier K+ Channel Kir2.2. *Nature* **477**, 495-U152 (2011).

29. Marcotte, I. & Auger, M. Bicelles as model membranes for solid- and solution-state NMR studies of membrane peptides and proteins. *Concepts in Magnetic Resonance Part A* **24A**, 17-37 (2005).

30. Denisov, I.G., Grinkova, Y.V., Lazarides, A.A. & Sligar, S.G. Directed self-assembly of monodisperse phospholipid bilayer nanodiscs with controlled size. *Journal of the American Chemical Society* **126**, 3477-3487 (2004).

31. Popot, J.L. *et al.* Amphipols: polymeric surfactants for membrane biology research. *Cellular and Molecular Life Sciences* **60**, 1559-1574 (2003).

32. Hopper, J.T.S. *et al.* Detergent-Free Mass Spectrometry of Membrane Protein Complexes. *Nature Methods* **10**, 1206-+ (2013).

33. Orwick-Rydmark, M. *et al.* Detergent-Free Incorporation of a Seven-Transmembrane Receptor Protein into Nanosized Bilayer Lipodisq Particles for Functional and Biophysical Studies. *Nano Letters* **12**, 4687-4692 (2012).

34. Rajesh, S., Knowles, T. & Overduin, M. Production of Membrane Proteins Without Cells or Detergents. *New Biotechnology* **28**, 250-254 (2011).

35. Munson, M.S.B. & Field, F.H. Chemical Ionization Mass Spectrometry .I. General Introduction. *Journal of the American Chemical Society* **88**, 2621-& (1966).

36. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. & Whitehouse, C.M. Electrospray Ionization for Mass-Spectrometry of Large Boimolecules. *Science* **246**, 64-71 (1989).

37. Dole, M., Mack, L.L. & Hines, R.L. Molecular Beams of Macroions. *Journal of Chemical Physics* **49**, 2240-& (1968).

38. Spengler, B. *et al.* The Detection of Large Molecules in Matrix-Assisted UV-Laser Desorption. *Rapid Communications in Mass Spectrometry* **4**, 301-305 (1990).

39. Li, Y.T., Hsieh, Y.L., Henion, J.D. & Ganem, B. Studies on Heme-Binding in Myoglobin, Hemoglobin, and Cytochrome-C by Ion-Spray Mass-Spectrometry. *Journal of the American Society for Mass Spectrometry* **4**, 631-637 (1993).

40. Loo, J.A. Studying Noncovalent Protein Complexes by Electrospray Ionization Mass Spectrometry. *Mass Spectrometry Reviews* **16**, 1-23 (1997).

41. Soares, R. *et al.* Mass spectrometry and animal science: Protein identification strategies and particularities of farm animal species. *Journal of Proteomics* **75**, 4190-4206 (2012).

42. Kebarle, P. A Brief Overview of the Present Status of the Mechanisms Involved in Electrospray Mass Spectrometry. *Journal of Mass Spectrometry* **35**, 804-817 (2000).

43. Mack, L.L., Kralik, P., Rheude, A. & Dole, M. Molecular Beams of Macroions .2. *Journal of Chemical Physics* **52**, 4977-& (1970).

44. Iribarne, J.V. & Thomson, B.A. On the Evaporation of Small Ions from Charged Droplets. *Journal of Chemical Physics* **64**, 2287-2294 (1976).

45. Thomson, B.A. & Iribarne, J.V. Field-Induced Ion Evaporation from Liquid Surfaces at Atmospheric-Pressure. *Journal of Chemical Physics* **71**, 4451-4463 (1979).

46. Konermann, L., Ahadi, E., Rodriguez, A.D. & Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Analytical Chemistry* **85**, 2-9 (2013).

47. Chernushevich, I.V., Loboda, A.V. & Thomson, B.A. An Introduction to Quadrupole-Time-of-Flight Mass Spectrometry. *Journal of Mass Spectrometry* **36**, 849-865 (2001).

48. Dawson, P.H. Quadruploe Mass Analyzers - Performance, Design and Some Recent Applications. *Mass Spectrometry Reviews* **5**, 1-37 (1986).

49. Makarov, A. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Analytical Chemistry* **72**, 1156-1162 (2000).

50. Kingdon, K.H. A Method for the Neutralization of Electron Space Charge by Positive Ionization at Very Low Gas Pressures. *Physical Review* **21**, 408-418 (1923).

51. Hu, Q.Z. *et al.* The Orbitrap: a New Mass Spectrometer. *Journal of Mass Spectrometry* **40**, 430-443 (2005).

52. Macek, B., Waanders, L.F., Olsen, J.V. & Mann, M. Top-Down Protein Sequencing and MS3 on a Hybrid Linear Quadrupole Ion Trap-Orbitrap Mass Spectrometer. *Molecular & Cellular Proteomics* **5**, 949-958 (2006).

53. Scigelova, M. & Makarov, A. Orbitrap Mass Analyzer - Overview and Applications in Proteomics. *Proteomics*, 16-21 (2006).

54. Lu, A. *et al.* Nanoelectrospray Peptide Mapping Revisited: Composite Survey Spectra Allow High Dynamic Range Protein Characterization Without LCMS on an Orbitrap Mass Spectrometer. *International Journal of Mass Spectrometry* **268**, 158-167 (2007).

55. Rose, R.J., Damoc, E., Denisov, E., Makarov, A. & Heck, A.J.R. High-Sensitivity Orbitrap Mass Analysis of Intact Macromolecular Assemblies. *Nature Methods* **9**, 1084-+ (2012).

56. Benesch, J.L.P., 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).

57. Wiza, J.L. Microchannel Plate Detectors. *Nuclear Instruments & Methods* **162**, 587-601 (1979).

58. Tahallah, N., Pinkse, M., Maier, C.S. & Heck, A.J.R. The Effect of the Source Pressure on the Abundance of Ions of Noncovalent Protein Assemblies in an Electrospray Ionization Orthogonal Time-of-Flight Instrument. *Rapid Communications in Mass Spectrometry* **15**, 596-601 (2001).

59. 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).

60. 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).

61. Sanglier, S., Ramstrom, H., Haiech, J., Leize, E. & Van Dorsselaer, A. Electrospray Ionization Mass Spectrometry Analysis Revealed a Similar to 310 kDa Noncovalent

Hexamer of HPr Kinase/Phosphatase from Bacillus Subtilis. *International Journal of Mass Spectrometry* **219**, 681-696 (2002).

62. Chernushevich, I.V. & Thomson, B.A. Collisional Cooling of Large Ions in Electrospray Mass Spectrometry. *Analytical Chemistry* **76**, 1754-1760 (2004).

63. Kanu, A.B., Dwivedi, P., Tam, M., Matz, L. & Hill, H.H. Ion Mobility-Mass Spectrometry. *Journal of Mass Spectrometry* **43**, 1-22 (2008).

64. Shvartsburg, A.A. & Smith, R.D. Fundamentals of Traveling Wave Ion Mobility Spectrometry. *Analytical Chemistry* **80**, 9689-9699 (2008).

65. Harding, S.E. & Jumel, K. Light Scattering. *Curr Protoc Protein Sci* **Chapter 7**, Unit 7 8 (2001).

66. Kim, H.-A., Seo, J.-K., Kim, T. & Lee, B.-T. Nanometrology and its Perspectives in Environmental Research. *Environ Health Toxicol* **29**, e2014016-2014010 (2014).

67. Edward, J.T. Molecular Volumes and Stokes-Einstein Equation. *Journal of Chemical Education* **47**, 261-& (1970).

68. Hall, Z., Schmidt, C. & Politis, A. Uncovering the Early Assembly Mechanism for Amyloidogenic beta(2)-Microglobulin Using Cross-linking and Native Mass Spectrometry. *Journal of Biological Chemistry* **291**, 4626-4637 (2016).

69. Schmidt, C. & Robinson, C.V. A Comparative Cross-Linking Strategy to Probe Conformational Changes in Protein Complexes. *Nature Protocols* **9**, 2224-2236 (2014).

70. Laganowsky, A. *et al.* Membrane Proteins Bind Lipids Selectively to Modulate their Structure and Function. *Nature* **510**, 172-+ (2014).

71. Claramunt, R.M. *et al.* The Structure of 1H-Perimidin-2(3H)-one adn its Derivatives in the Solid-State (X-Ray Crystallography and CP/MAS C-13-NMR), in Solution (C-13-NMR), and in the Gas-Phase (Mass-Spectrometry). *Helvetica Chimica Acta* **77**, 121-139 (1994).

72. Laganowsky, A., Reading, E., Hopper, J.T.S. & Robinson, C.V. Mass Spectrometry of Intact Membrane Protein Complexes. *Nature Protocols* **8**, 639-651 (2013).

73. Reading, E. *et al.* The Role of the Detergent Micelle in Preserving the Structure of Membrane Proteins in the Gas Phase. *Angewandte Chemie-International Edition* **54**, 4577-4581 (2015).

74. Tribet, C., Audebert, R. & Popot, J.L. Amphipols: Polymers that Keep Membrane Proteins Soluble in Aqueous Solutions. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 15047-15050 (1996).

75. Gorzelle, B.M. *et al.* Amphipols Can Support the Activity of a Membrane Enzyme. *Journal of the American Chemical Society* **124**, 11594-11595 (2002).

76. Prata, C. *et al.* Nonionic Amphiphilic Polymers Derived from Tris(hydroxymethyl)-Acrylamidomethane Keep Membrane Proteins Soluble and Native in the Absence of Detergent. *Biopolymers* **56**, 77-84 (2001).

77. Borch, J. & Hamann, T. The Nanodisc: a Novel Tool for Membrane Protein Studies. *Biological Chemistry* **390**, 805-814 (2009).

78. Orwick, M.C. *et al.* Detergent-Free Formation and Physicochemical Characterization of Nanosized Lipid-Polymer Complexes: Lipodisq. *Angewandte Chemie-International Edition* **51**, 4653-4657 (2012).

79. Knowles, T.J. *et al.* Membrane Proteins Solubilized Intact in Lipid Containing Nanoparticles Bounded by Styrene Maleic Acid Copolymer. *Journal of the American Chemical Society* **131**, 7484-+ (2009).

2. Hydrodynamic Comparison of Solution and Gas-Phase Protein Shape and Sizes:

2.1. Introduction:

This chapter details the work comparing the shape and size of proteins in solution, obtained using DLS, and in the gas-phase, using ion mobility. The significance of this work derives from the increasing popularity of IM-MS as a tool for interrogating protein structures, due to its ability to deduce information about a protein's shape and size[1, 2, 3]. IM-MS has significant advantages as a technique compared to traditional methods for obtaining low resolution structural information. Compared to small angle x-ray scattering (SAXS), analytical ultracentrifugation (AUC) and DLS, IM-MS has the benefit of requiring far less sample, only a few microlitres of sample at micromolar concentration, coupled with a significantly shorter timescale for the experiment.

The benefits in terms of speed and the reduction in the amount of sample required per experiment, has led to an explosion in the prevalence of mobility data for proteins. This in turn has caused more questions about its viability as a technique for interrogating proteins' native structure. These concerns are typically around the effect of ionisation and desolvation on the structure of a protein complex. However there have been studies to show that proteins can remain stable in the gas-phase for periods of time longer than the usual IM-MS experiment. It is important to note, however, that the gas-phase stability for any protein ion is dependent on both the charge and protein in question[4, 5].

Despite the evidence that proteins can remain folded and stable during an IM-MS experiment, there remains an issue in the different collision cross-sections obtained from experimental data and from those obtained using known crystal structures coupled with

computational methods. The best example of this can be seen in that despite the development of many advanced computational methods which simulate IM-MS[6, 7, 8, 9, 10] a simple scaled projection is still generally accepted as the best method to accurately predict the collision cross-section that will be measured experimentally[11]. A potential rationale for the lack of agreement between the more rigorous computationally calculated cross-sections and the experimentally obtained values would be a partial collapse of structure in the gas-phase. If there was to be a slight structural alteration of the protein when moving from the solution-phase to the gas-phase it would account for these discrepancies.

2.2. Experimental Outline:

In order to effectively probe whether protein structure is maintained after the proteins are transferred to the gas-phase, it is important to utilise two techniques which give comparable data. Given that the purpose of this work is to understand the relationship between IM-MS cross-sections and similar resolution structural information from another source, it is obvious to use IM-MS for the gas-phase technique. However, the choice of technique to use for comparable data is not as straightforward, with a range of options to choose from in analytical ultracentrifugation (AUC), small angle x-ray scattering (SAXS), and dynamic light scattering (DLS).

The choice between AUC, SAXS and DLS comes down to a number of different factors. These include ease of performing the experiments, as well as the levels of data it is possible to get out of the experiments. Both the AUC and DLS experiments take place in solution with the SAXS experiment being capable of running in solid or liquid form. All of these methods available allow for the freedom to choose our preferred buffer for the experiment. This means they can all be run in the same buffer that the IM-MS experiment is conducted in.

This leaves the choice of technique being dependent upon their ability to easily separate out the different oligomeric states of the proteins before the analysis takes place.

Using this rationale, it was decided to utilise DLS over AUC and SAXS, as it is possible to couple DLS with size exclusion chromatography (SEC). The combination of these two techniques has a double benefit: firstly, the use of SEC means that different oligomeric states are separated out during the course of a run, and secondly by introducing the sample eluting from the column directly into the DLS cell multiple data points can be collect from each experiment. This leads to much easier identification of outliers and a much higher level of confidence in the data acquired for each protein.

The DLS experiment allows us to gain multiple auto-correlation functions for each sample as discussed in the Introduction. These in turn allow for the calculation of a number of different Stokes radii for each sample in a single run as detailed earlier.

Experimental data was obtained during the course of a number of weeks. This lead to a natural variation in the experimental conditions during each run. The standardisation of all determined diffusion co-efficients eliminates any discrepancies which might arise from these factors. This can be done simply by standardising them using *Equation 2.1* where $D^{o}_{20,w}$ is the diffusion co-efficient at standard conditions of 20 degrees Celsius in water, D^{o} is the measured diffusion co-efficient, T is the temperature of the experiment in Kelvin, $\eta_{soln,T}$ is the viscosity of the solvent at the temperature of the experiment, and $\eta_{20,w}$ is the viscosity of water at 20 degrees Celsius.

(Equation 2.1)

$$D^{\circ}_{20,w} = D^{\circ}\left(\frac{293.15}{T}\right)\left(\frac{\eta_{soln,T}}{\eta_{20,w}}\right)$$

These standardised diffusion co-efficients can then be used with the Einstein-Stokes equation shown in *Equation 2.2*[12] to calculated standardised Stokes radii for the protein or protein complexes.

(Equation 2.2)

$$R_s = \frac{k_B T}{6\pi\eta D^{\circ}_{20,w}}$$

Whilst this gives us a Stokes radius for the proteins it is important to note that this is not the same as the radius the protein will have in the gas-phase. While proteins are in solution they will have an associated hydration shell of water molecules that will tend to move with the proteins[13]. This effect is important as it affects the initial measurement of the diffusion co-efficient underestimating it compared to the naked protein.

The radius of the anhydrous protein can then be calculated by making use of a hydration factor (δ), the partial volume of the protein ($\bar{v}$) and the density of the solvent (ρ). The hydration factor for a protein is defined as the grams of water association with a protein per gram of protein. Using this relationship (*Equation 2.4)* and the volume of a sphere (*Equation 2.3)* in combination it is possible to define the radius of the anhydrous protein as given in *Equation 2.5*:

(Equation 2.3)

$$V = \frac{4}{3}\pi r^3$$

(Equation 2.4)

$$\delta = \bar{v}\rho\left(\frac{V_{hy.}}{V_{anhy.}} - 1\right)$$

(Equation 2.5)

$$R_{anhyd.} = \frac{R_s \bar{v} \rho}{\sqrt[3]{\bar{v}^2 \rho^2 (\bar{v}\rho + \delta)}}$$

Having derived a radius for the protein of interest, stripped of its hydration shell, it is then possible to determine the expected collision cross-section (Ω) for this protein in the gas-phase. This can be done simply using *Equation 2.6*[11] below which accounts for the radius of the ion being analysed (R_i) as well as the buffer gas it interacts with (R_{He}) during the mobility measurement. However, one other term is necessary to account for the inelastic ion/neutral collisions - this is a drag enhancement term (ξ). Note this equation is only correct for when the ion-mobility experiment uses helium as its buffer gas.

(Equation 2.6)

$$\Omega = \pi (R_i + R_{He})^2 \times \xi$$

This drag factor has been determined empirically by Millikan to be 1.36, which according to Epstein relates to the percentage (91%) of neutrals that interact with the ion as shown in *Equation 2.7.*[14, 15, 16]

(Equation 2.7)

$$\xi = \frac{\pi 0.91}{8}$$

Through the use of all the experiments and equations described above it is possible to calculate a collision cross-section for proteins in the gas-phase from their measure diffusion co-efficients in solution:

(Equation 2.8)

$$\Omega = \pi \left(\frac{\frac{k_B T}{6\pi\eta D^{\circ}_{20,w}} \overline{v}\rho}{\sqrt[3]{\overline{v}^2\rho^2(\overline{v}\rho + \delta)}} + R_{He} \right)^2 \times \xi$$

2.3. Experimental Setup:

To perform the online SEC-DLS measurements all proteins detailed in **Table 2-1** were purchased as dry powders and prepared as detailed in Materials and Methods.

Protein	Mass (kDa)	*n*
Cytochrome C	11.8	1
Lysozyme	14.3	1
Myoglobin	17.7	1
Chymotrypsin	25.7	1
Ovalbumin	44.3	1
Transthyretin	55.1	4
Avidin	58.5	4
Bovine serum albumin	69.3	1
Alcohol dehydrogenase	147.4	4
Aldolase	157.5	4
Pyruvate kinase	232.6	4
Glutamate dehydrogenase	329.6	6
β-Galactosidase	475.6	4

Table 2-1: Summary of proteins used for the DLS experiments along with their masses and oligomeric states.

The prepared samples were then injected and then eluted using a SuperDex 75 10/300 as describe in Materials and Methods and then analysed using a Dawn Heleos II modified with an inboard quasielastic light scatterer. The UV absorbance, diffractive index and diffusion co-efficient were all obtained as a function of time over each run. A cartoon depicting the experimental setup is shown below **(Figure 2-2)**

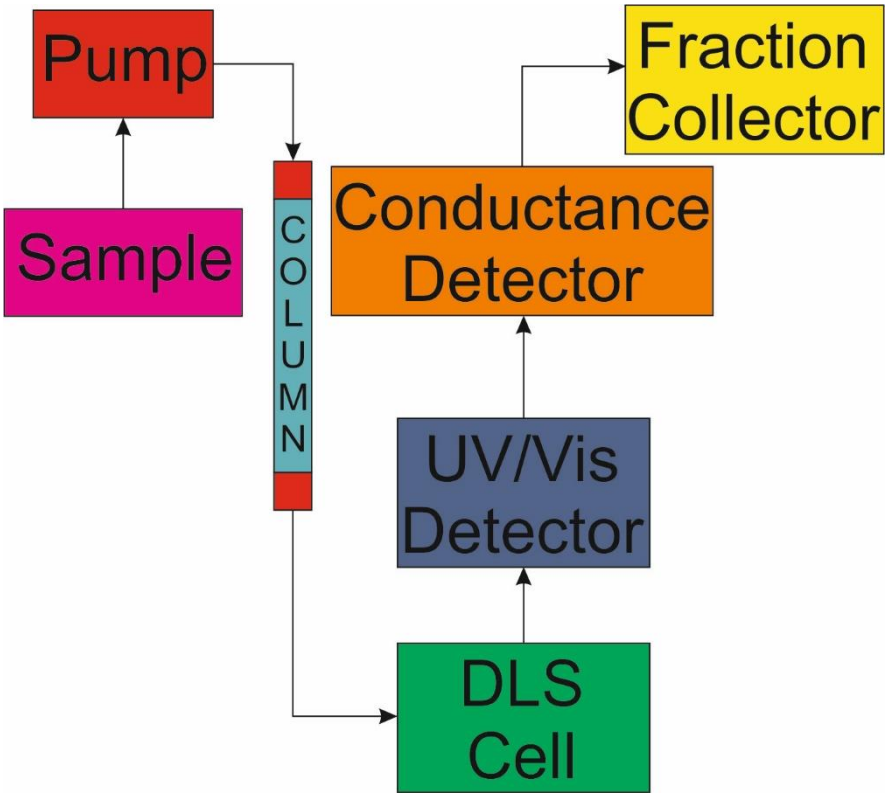


Figure 2-1: Cartoon schematic of the DLS experimental setup with the sample undergoing separation on the size exclusion column before analysis by DLS and then UV/Vis to identify the elution time for each protein.

2.4. Validation of Experimental Methodology:

A primary concern in performing this work was whether the different experimental methodologies would have similar shape dependencies. If there was a significant difference in the shape dependencies between the different methods, then despite no change in the protein structure between the solution and gas-phase, it would still be possible that the results would suggest that the structure was significantly different.

To validate this approach, formulae are needed to describe the relationship between shape factors and increasing deviation from perfectly symmetrical shapes. This is done by considering the change in shape factors for both oblate and prolate shapes as a result of deviation from a sphere. For DLS it is possible to describe the shape and size relationships using the Perrin translational friction factors, f(P) for two-component ellipsoids[17, 18]. The term f(P) is a radial shape factor (R_{sf}) defining the relationship between apparent radius and

ellipsoid axial ratio as shown in *Equations 2.9 & 2.10*,where q is the axial ratio of the semi major and semi minor ellipsoid axes (*a* and *b* respectively) given by *a/b* for prolate ellipsoids and *b/a* for oblate ellipsoids. The comparative effects of the deviation from sphericality for DLS and IM-MS are plotted in **Figures 2-3 and 2-4**.

(Equation 2.9)

$$f(P) = \frac{\sqrt{1-q^2}}{q^{2/3}\ln\frac{1+\sqrt{1-q^2}}{q}}$$

(Equation 2.10)

$$f(P) = \frac{\sqrt{q^2-1}}{q^{2/3}\tan^{-1}\sqrt{q^2-1}}$$

For the ion-mobility modelling the shape factors (Ω_{sf}) for elliptical size and shape relationships are estimated based on well-known relationships between the projected area (PA) and the surface area (SA) for convex bodies as shown in *Equation 2.11*.

(Equation 2.11)

$$SA = 4PA$$

Using *Equation 2.11* it is possible to derive formulae for Ω_{sf} for both oblate and prolate ellipsoids as shown in *Equations 2.12 and 2.13*,where α is equal to the angular eccentricity which for a prolate is arcos(*a/b*) and for an oblate is arcos (*b/a*):

(Equation 2.12)

$$\Omega_{sf}^{oblate} = \frac{2\pi\left(a^2 + \frac{ab\alpha}{\sin\alpha}\right)}{4\pi\left(\sqrt[3]{3V/4\pi}\right)^2}$$

(Equation 2.13)

$$\Omega_{sf}^{prolate} = \frac{2\pi\left[a^2 + \frac{b^2}{\sin\alpha}\ln\left(\frac{1+\sin\alpha}{\cos\alpha}\right)\right]}{4\pi\left(\sqrt[3]{3V/4\pi}\right)^2}$$

The equations for the volumes for an oblate and prolate respectively are given below in *Equation 2.14* and *2.15*:

(Equation 2.14)

$$V^{oblate} = \frac{4}{3}\pi a^2 b$$

(Equation 2.15)

$$V^{prolate} = \frac{4}{3}\pi b^2 a$$

Through substitution of the volumes given above into the relevant equations for the different shape factors it is possible to obtain *Equation 2.16* and *Equation 2.17*. This defines the shape factor for oblates and prolates solely in terms of *a, b* and *α*. This in turn allows for the definition of the shape factor in terms of the axial ration for the ellipsoids in question.

(Equation 2.16)

$$\Omega_{sf}^{oblate} = \frac{\left(a^2 + \frac{ab\alpha}{\sin\alpha}\right)}{2\left(\sqrt[3]{b^2 a}\right)^2}$$

(Equation 2.17)

$$\Omega_{sf}^{prolate} = \frac{\left[a^2 + \frac{b^2}{\sin\alpha}\ln\left(\frac{1+\sin\alpha}{\cos\alpha}\right)\right]}{2\left(\sqrt[3]{a^2 b}\right)^2}$$

In order to bring the shape factor for IM-MS in line with that for DLS it is rescaled to $\sqrt{\Omega_{sf}}$.

The plot for the shape factors against axial ratio is shown below in **Figure 2-2**:

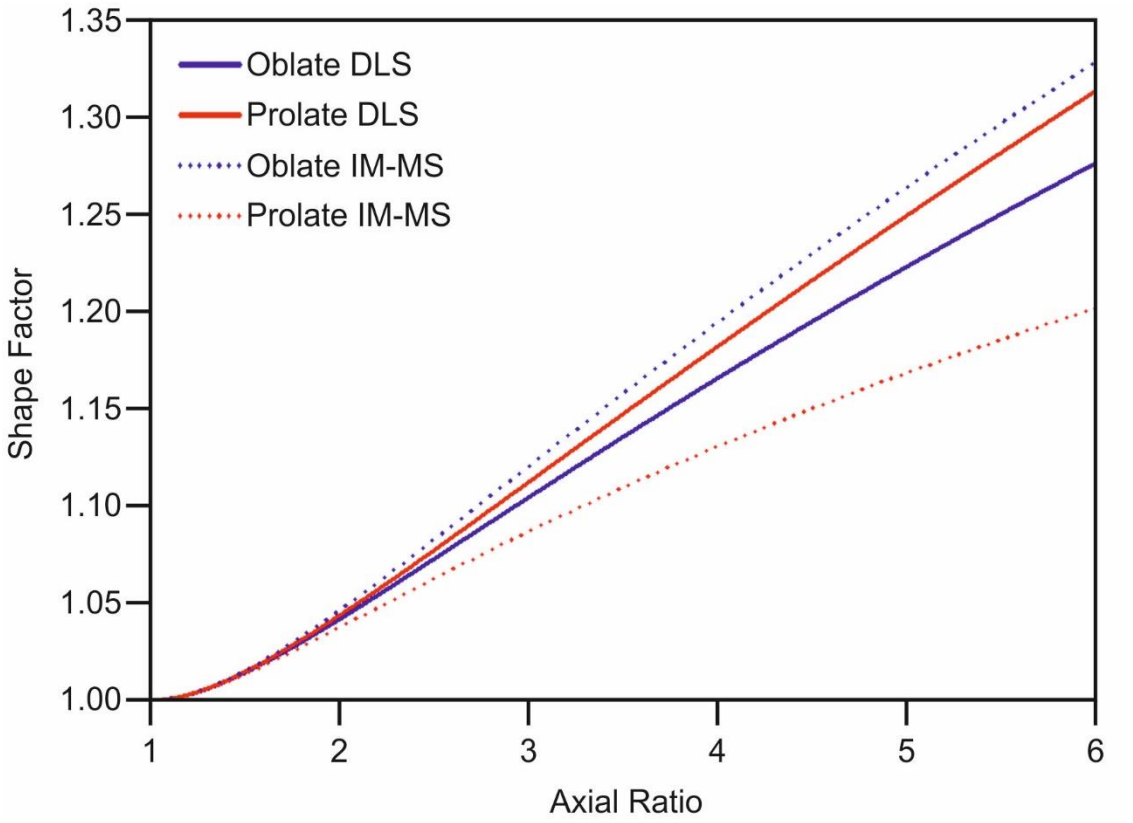


Figure 2-2: Shape factors for oblates are shown in blue whilst shape factors for prolates are shown in red. IM-MS shape factors are represented as dotted lines whilst those for DLS are represented by solid lines.

Figure 2-2 demonstrates that there is relatively little difference between the shape factors, provided the axial ratios are not too high. For this reason this study was undertaken using predominantly globular proteins where the distortion away from a perfect sphere is kept as minimal as possible. For simplicity the deviations for oblates and prolates between the different experimental techniques are plotted below. **Figure 2-3** demonstrates that even with an axial ratio as high as three, where there is a distinct deviation from sphericality, the difference between the two methods is less than 2.5%. This shows that for globular proteins there is little to no difference between the shape factors observed for DLS and IM-MS. From this it was concluded that making a direct comparison between IM-MS and DLS data is appropriate, provided the proteins used are all spherical or reasonably close to it.

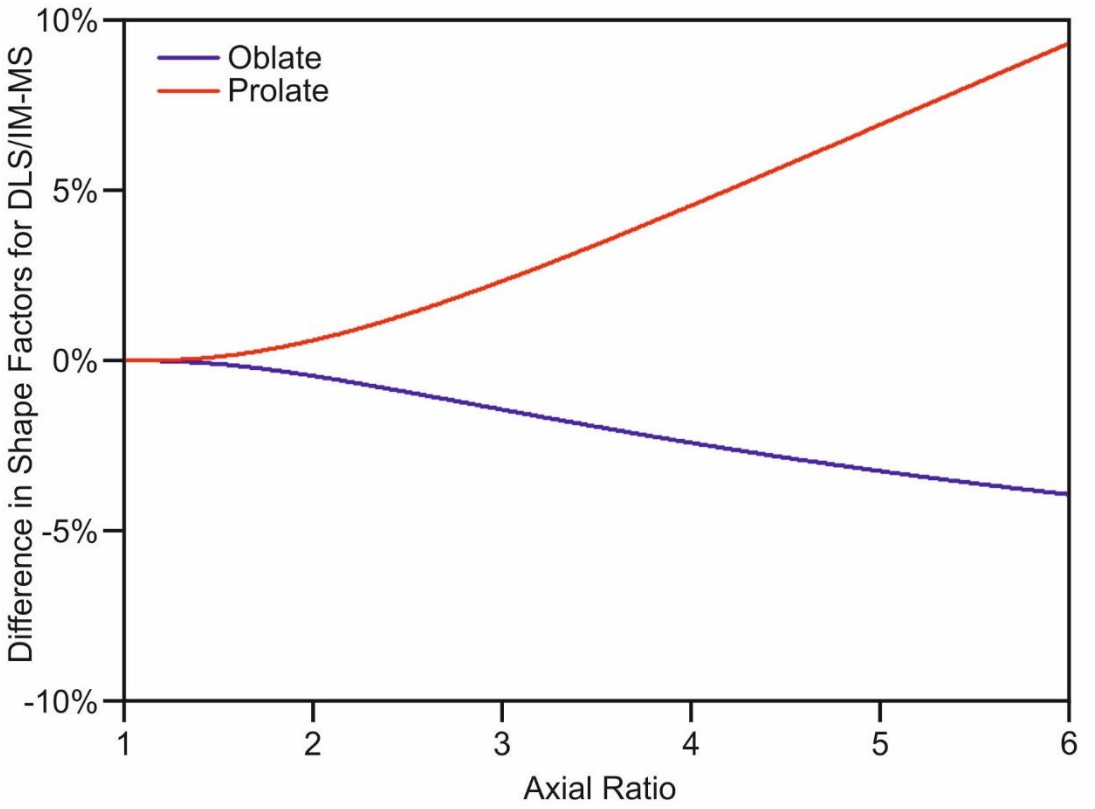


Figure 2-3: The percentage difference between the DLS and IM-MS shape factors defined as $R_{sf}/\sqrt{\Omega_{sf}}$. Deviations for oblates are shown in blue and those for prolates are shown in red.

With that in mind the following proteins were selected for use in this work: Cytochrome C, Myoglobin, Lysozyme, Chymotrypsin, Ovalbumin, Transthyretin, Avidin, Bovine Serum Albumin, Alcohol Dehydrogenase, Aldolase, Pyruvate Kinase, Glutamate Dehydrogenase and β-Galactosidase. These proteins have the benefit of all being commercially available in addition to fulfilling the requirement of being close to spherical as can be seen in **Figure 2-4**:

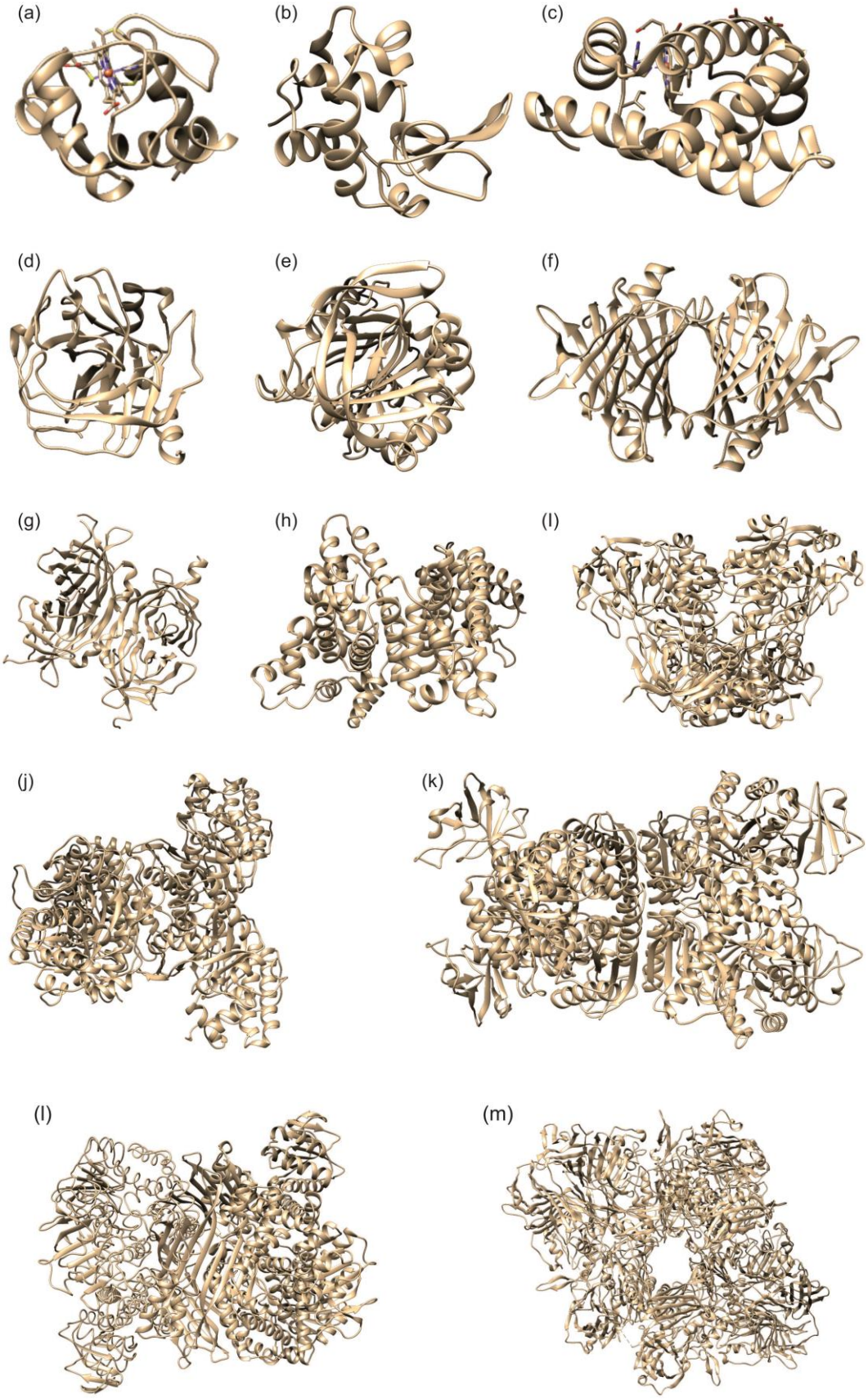


Figure 2-4: X-ray crystal structures of all proteins used (a) Cytochrome C[19], (b) Lysozyme[20], (c) Myoglobin[21], (d) Chymotrypsin[22], (e) Ovalbumin[23], (f) Transthyretin[24], (g) Avidin[25], (h) Bovine Serum Albumin[23], (i) Alcohol Dehydrogenase[26], (j) Aldolase[27], (k) Pyruvate Kinase[28], (l) Glutamate Dehydrogenase[29], (m) β-Galactosidase[30]

2.5. Results:

Results for this chapter are broken down into a number of constituent parts. The first details with the initial results from the DLS experiment. The following section discusses how to put these results into context by calibrating for the buffers used. The third subsection discusses the raw ion-mobility results, before a complete discussion of the insights we can get about ion-mobility data from the DLS data. Finally in the last subsection of the results I discuss the way our results compare with the Millikan drag factor depending on the analysis methodology used.

2.5.1. DLS Results:

The diffusion co-efficients ($D°$) for each protein were obtained using SEC-DLS in 200 mM ammonium acetate as described earlier. These diffusion co-efficients are determined from the auto-correlation function from each protein. The setup used of an online DLS system connected to the SEC column allows for values for $g^2(\tau)$ to be taken every two seconds. This, combined with all proteins being eluted over a volume of more than one ml, allows for over 100 independent values of $D°$ to be collected per elution of protein or protein complex. From the initial data sets acquired an initial filter was manually applied to remove the fringe 10% of values. This was done due to the lower concentration of protein present when the DLS experiment was performed leading to a large variance in the values determined for $D°$ **(Figure 2-5).** In addition to this initial filter for each data set the mean and standard deviation were then calculated and any data points that were more than three standard deviations away from the mean were identified as outliers and removed from the determination of the average diffusion co-efficient.

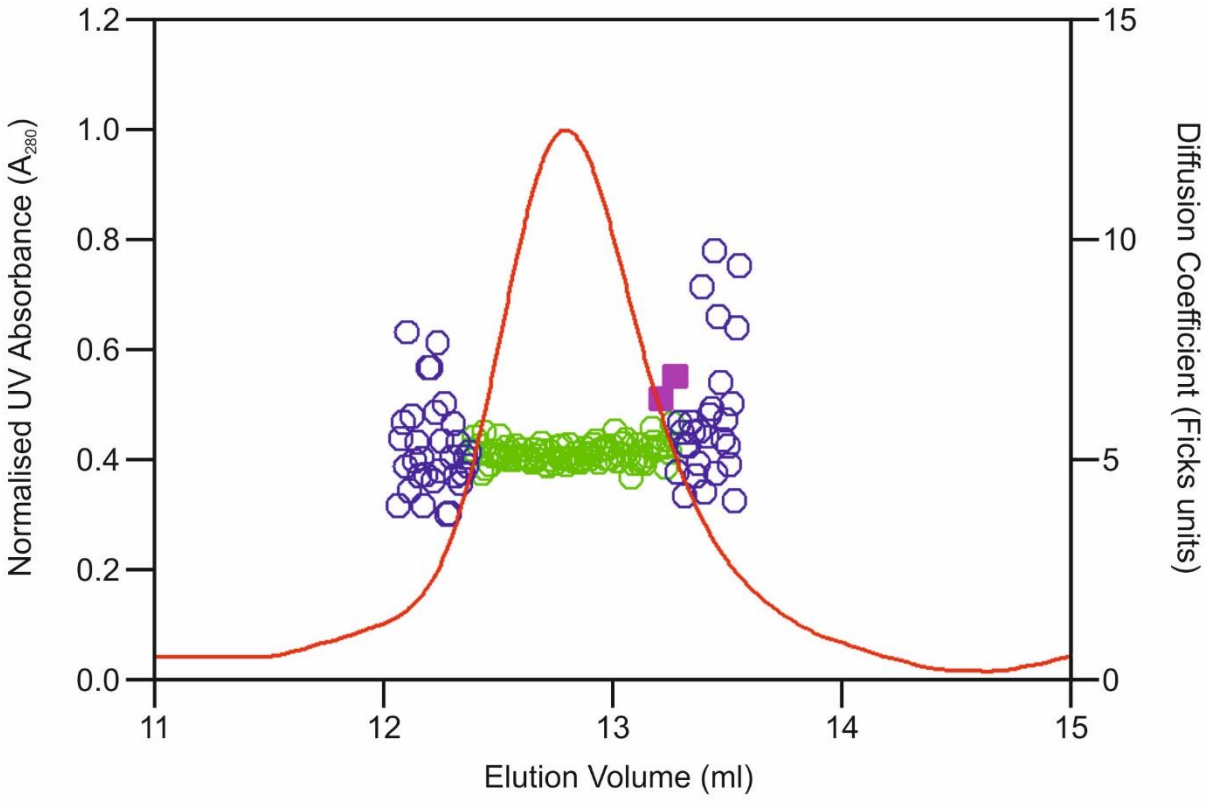


Figure 2-5: UV absorption for the elution peak for the alcohol dehydrogenase tetramer (red) with the diffusion coefficients for each point shown. The peripheral values are excluded (hollow blue circles) with the remaining values inside three standard deviations used (hollow green circles) whilst values outside three standard deviations (filled purple squares) are considered outliers and removed from the dataset.

For each protein all the measured *D°* values need to be standardised to account for the solvent viscosity and temperature. This can be done using *Equation 2.1* which allows for this conversion to take place, providing the temperature of the initial experiment and relevant viscosities are known. However there is a well-documented link between viscosity and temperature, so in order to effectively standardise the protein diffusion co-efficients a calibration curve for viscosity of 200 mM ammonium acetate is required.

2.5.2. Creating a Calibration Curve:

To create a calibration curve the viscosity of 200 mM ammonium acetate was measured at 5°C, 15°C and 30°C using a rolling ball viscometer. Measurements for the viscosity at each temperature were taken three times and then averaged. Then a calibration curve was generated by making use of the relationship between the viscosity of the solvent and the solution as shown in *Equation 2.18* where $\eta_{AA,T}$ is the viscosity of ammonium acetate at temperature T, η_{AA,T_c} is the viscosity of ammonium acetate at the calibration temperature, $\eta_{H_2O,T}$ is the viscosity of water at temperature T and η_{H_2O,T_c} is the viscosity of water at the calibration temperature.

(Equation 2.18)

$$\frac{\eta_{AA,T}}{\eta_{AA,T_c}} = \frac{\eta_{H_2O,T}}{\eta_{H_2O,T_c}}$$

The viscosity of water can be calculated from the temperature as shown using *Equation 2.19* and *Equation 2.20*. *Equation 2.19* is used when the temperature is above 20°C and *Equation 2.20* is used when the temperature is less than 20°C.

(Equation 2.19)

$$\log_{10}\eta_T = \frac{1301}{998.333 + 8.1855(T-20) + 0.00585(T-20)^2} - 1.30233$$

(Equation 2.20)

$$\log_{10}\frac{\eta_T}{\eta_{20}} = \frac{1.3272(20-T) - 0.001053(T-20)^2}{T+105}$$

The calibration curve is then generated by averaging the viscosity generated from each of the three different calibration points. The calibration curve generated for 200 mM

ammonium acetate is shown below in **Figure 2-6** with a comparison with pure water shown as well.

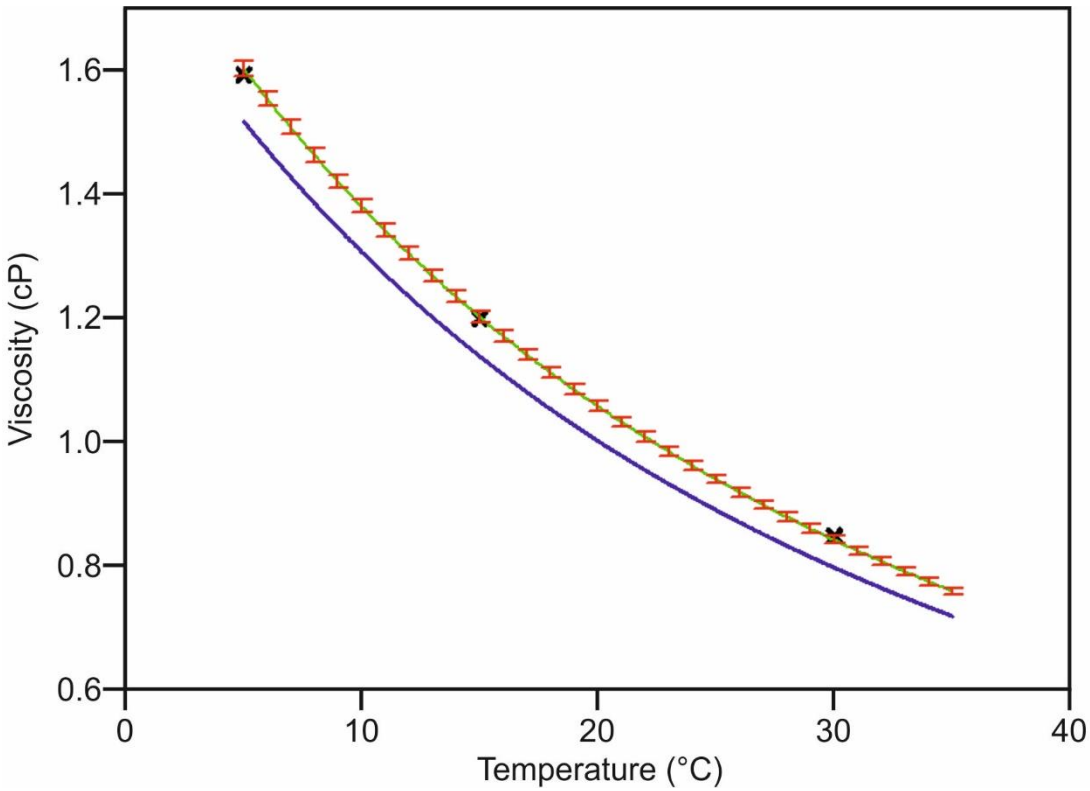


Figure 2-6: Calibration curve for viscosity of 200 mM ammonium acetate depending on temperature (green) is shown with the standard deviation from the three different calibration temperatures (red) shown. Calibration viscosities are also shown (black crosses) as well as the viscosity of water for comparison (blue)

Once the calibration curve was made, it was possible to combine the data obtained using the DLS experiment with the calibration curve to derive standardised diffusion co-efficients for each of the samples used. Combining the standardised diffusion co-efficients with *Equation 2.2* to get a Stokes radius for each hydrated protein or protein complex in this work is detailed in **Table 2-2:**

Protein	Mass (kDa)	*n*	$D^{\circ}_{20,w}$	R_S (nm)
Cytochrome C	11.8	1	13.0±1.6	1.62±0.2
Lysozyme	14.3	1	11.6±0.4	1.85±0.1
Myoglobin	17.7	1	11.7±1.1	1.83±0.2
Chymotrypsin	25.7	1	9.6±0.3	2.23±0.1
Ovalbumin	44.3	1	7.0±0.3	3.06±0.1
Transthyretin	55.1	4	6.3±0.1	3.38±0.1
Avidin	58.5	4	6.3±0.5	3.42±0.3
Bovine serum albumin	69.3	1	6.1±0.6	3.50±0.3
Alcohol dehydrogenase	147.4	4	4.8±0.2	4.48±0.2
Aldolase	157.5	4	4.9±0.2	4.34±0.3
Pyruvate kinase	232.6	4	4.1±0.2	5.17±0.2
Glutamate dehydrogenase	329.6	6	3.6±0.1	6.02±0.2
β-Galactosidase	475.6	4	3.4±0.1	6.35±0.1

Table 2-2: Summary of data obtained from the DLS experiment. Protein masses and stoichiometries (n) were obtained from previously reported values. $D^{\circ}_{20,w}$ were determined experimentally in 200 mM ammonium acetate as described earlier and given in Fick's units. Stokes radii were calculated from the diffusion co-efficients using *Equation 2.3*. Errors are ± one standard deviation of the mean.

2.5.3. IM-MS Results:

Collison cross-sections were obtained via direct measurement on an in-house linear field radiofrequency confined ion-mobility mass spectrometer, where they were unavailable from literature. Drift times were obtained directly from the experiment and all of the temperature, pressure and acceleration voltage were recorded. From the drift time measured and knowing t_{d0} for the instrument the mobility for each protein can be calculated using *Equation 1.4*. The mobilities were then all standardised to obtain a K_0 value as shown in *Equation 2.21*.

(Equation 2.21)

$$K_0 = K \times \frac{273.15}{T} \times \frac{P}{760}$$

From the standardised mobilities the collision cross-section can then be determined using *Equation 1.5*. For each protein the collision cross-sections were averaged, both across all

available charge states, and over all replicate experiments. The reported error was the standard deviation of all the Ω values obtained. IM-MS shows one of its primary benefits here as the presence of multiple charge states allows for multiple measurements of the collision cross-section during the course of a single experiment.

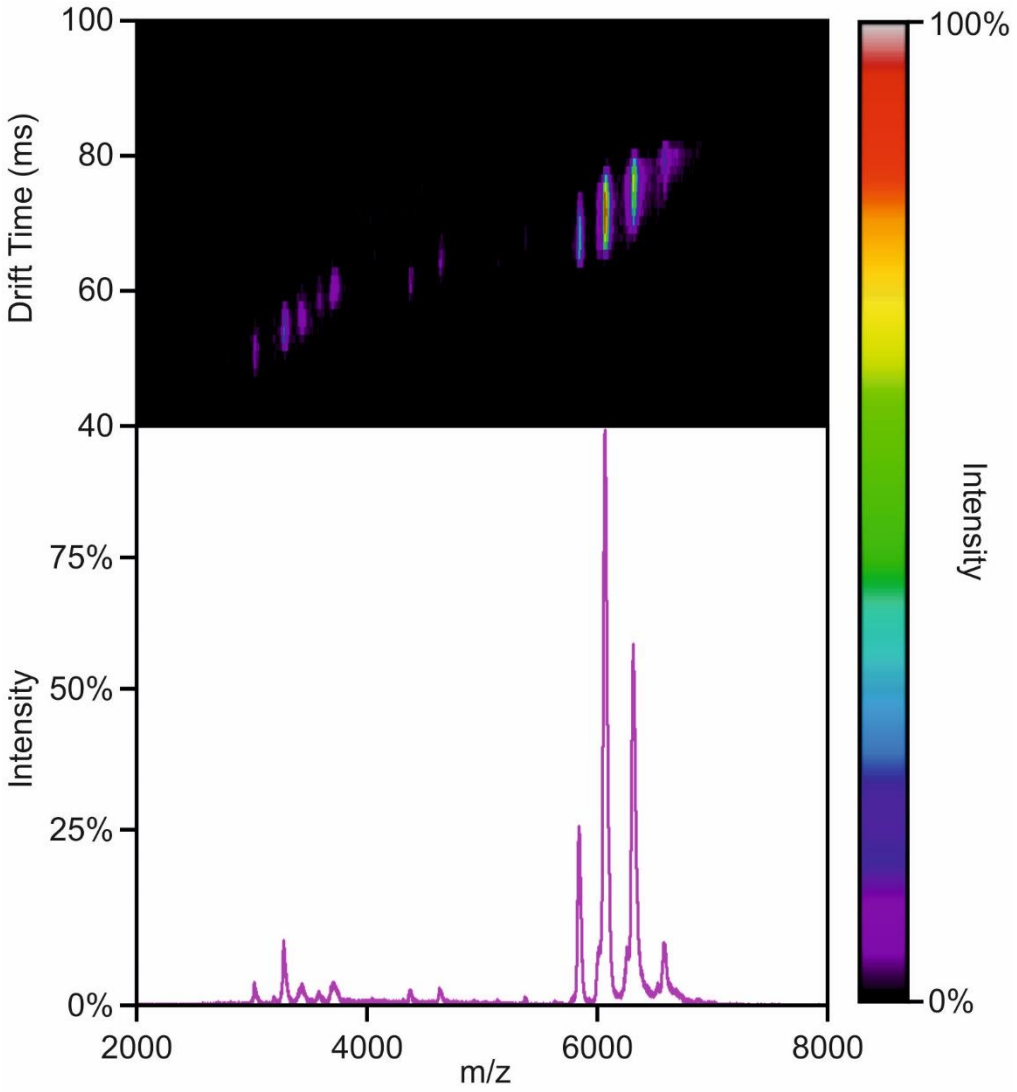


Figure 2-7: Ion-mobility spectrum (upper) and mass spectrum (lower) for aldolase as recorded on the in-house linear field synapt.

Each peak in the mobility spectrum was taken and fitted to a Gaussian distribution, as shown in **Figure 2-8**. The mean value for the peak is then used for the drift time measurement for each mobility calculation, and the standardised mobility constants are

then used in conjunction with *Equation 1.5* to calculate the collision cross-sections.

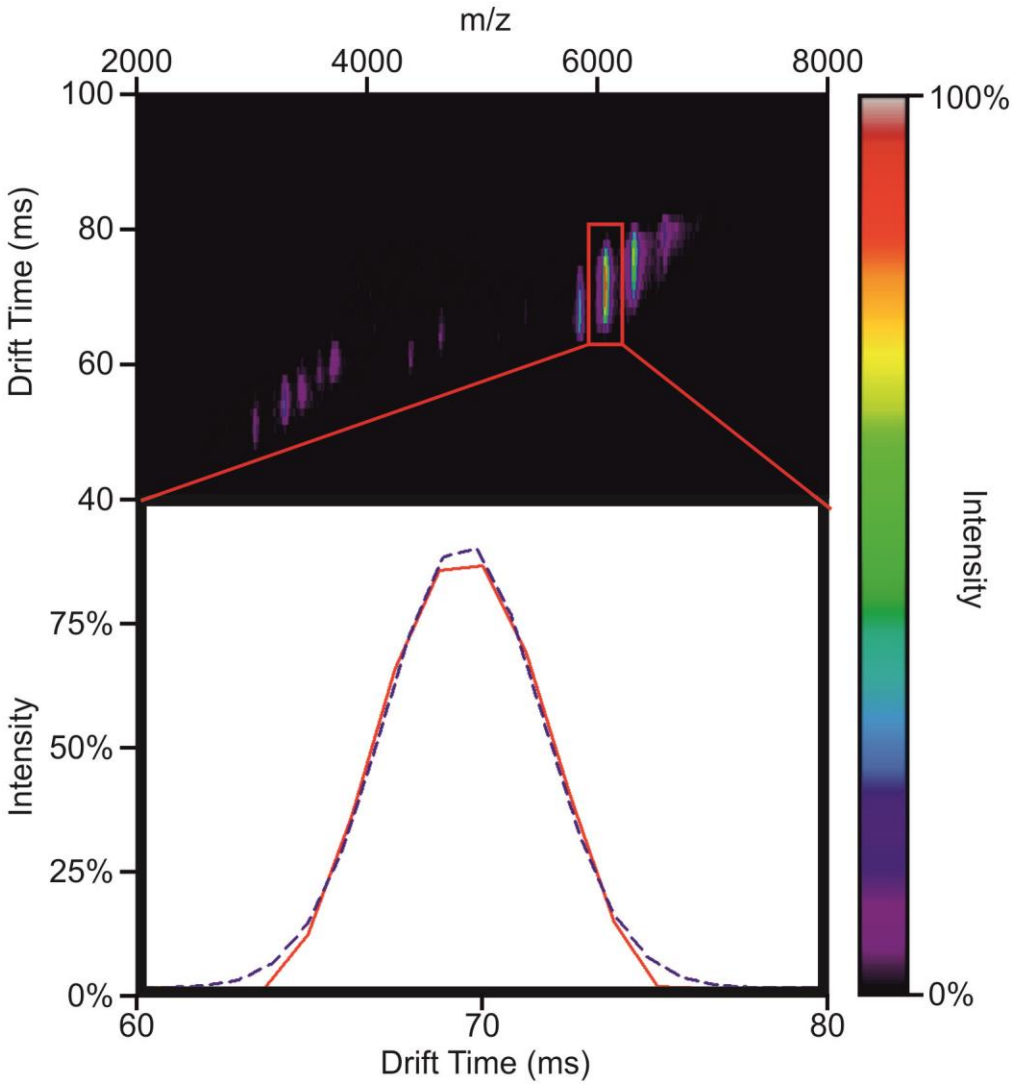


Figure 2-8: Example of the fitting of measured drift times to a Gaussian distribution for the 26+ charge state of aldolase. The experimental drift times are shown in red whilst the Gaussian fit is shown in blue. The drift time used for the collision cross-section is the centre of the Gaussian fit.

The collision cross-sections for the twelve different proteins used were obtained using the experimental setup described previously or taken from literature where available[11]. The results for these are shown in **Table 2-3**, along with their mass and the oligomeric state the protein is observed in:

Protein	Mass (kDa)	n observed	Ω (nm^2)
Cytochrome C	11.8	1	12.6±0.3[11]
Lysozyme	14.3	1	14.7±0.9
Myoglobin	17.7	1	16.7±0.1[11]
Chymotrypsin	25.7	1	18.7±0.4
Ovalbumin	44.3	1	31.8±0.1
Transthyretin	55.1	4	34.0±0.2[11]
Avidin	58.5	4	36.4±0.0[11]
Bovine serum albumin	69.3	1	40.7±0.3[11]
Alcohol dehydrogenase	147.4	4	68.6±1.0[11]
Aldolase	157.5	4	80.1±0.5
Pyruvate kinase	232.6	4	102.2±1.2[11]
Glutamate dehydrogenase	329.6	6	128.0±0.0[11]
β-Galactosidase	475.6	4	155.2±0.5

Table 2-3: Summary of data obtained from the IM-MS experiment. Protein masses and stoichiometries (n) were obtained from previously reported values. Ω is the mean experimental collision cross-section from all charge states. Errors are ± one standard deviation of the mean. Values taken from literature are

2.6. Comparison of DLS Data with IM-MS Data:

From the results above it is now possible to compare the predicted collision cross-section for each protein based on the radius measured in solution. This can be done by using *Equation 2.6* with a value of 1.36 used for ξ. For this methodology, a series of predicted cross-sections can be determined and compared directly with the experimentally determined cross-sections as shown in **Figure 2-9**:

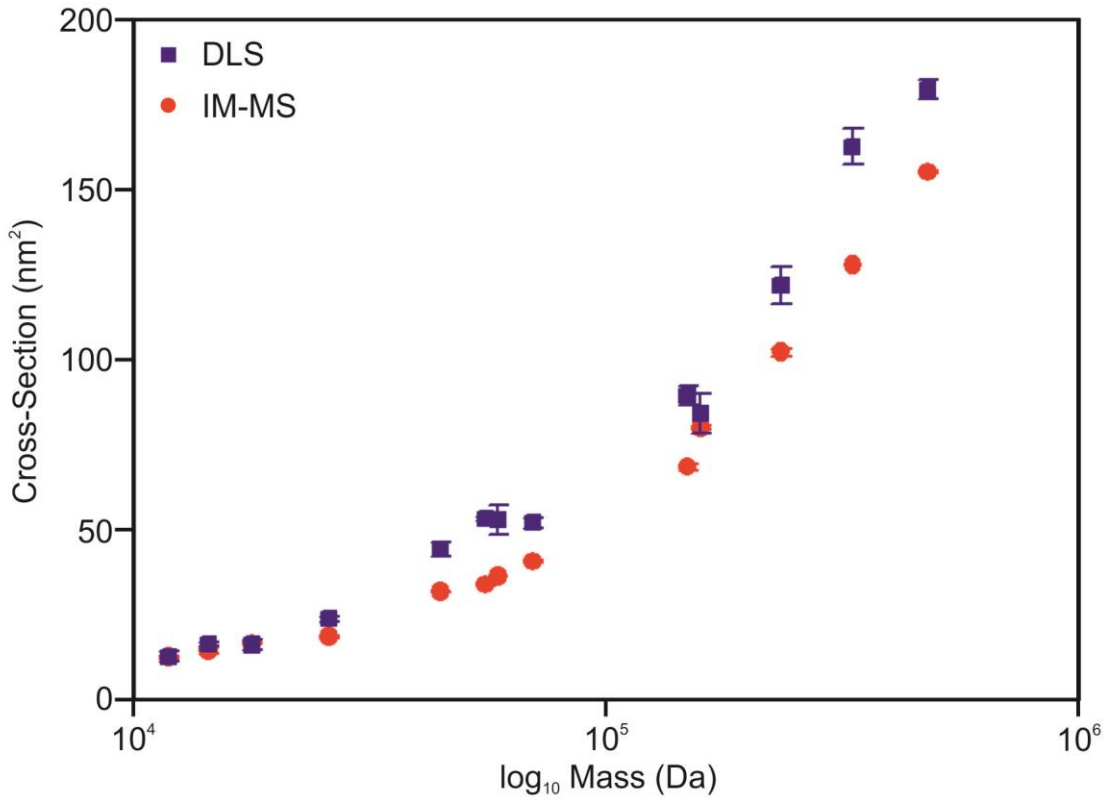


Figure 2-9: Experimental collision cross-sections (red) compared to those calculated based on the Stokes radii obtained via DLS (Blue).

This comparison above shows that the Ω values generated from the DLS experiment are incompatible with those directly measured using ion-mobility. The DLS derived collision cross-sections in blue are all significantly larger than those measured directly in the IM-MS experiment. This could be interpreted as evidence supporting the theory that proteins undergo collapse in the gas-phase leading to a reduction in the observed Ω values. However, an alternative theory is that, by using raw Stokes radii for the proteins, the water associated with them in solution leads to inflated radii compared to the gas-phase measurement.

In order to achieve a more realistic comparison between the DLS data and the IM-MS data the next step is to try to take into account the hydration of the protein. In order to do this, two possible approaches were considered in order to account for the hydration shell surrounding the protein. The hydration shell is defined as the water that differs from the bulk of the solution due to its interaction with the protein. The amount of water associated

with the protein in solution (δ) can be estimated using the Kuntz method[31]. This methodology works on the assumption that the hydration factor for any protein can be determined as a combination of the hydration factors for all the constituent amino acids. This is done as shown below using *Equation 2.22* to generate a hydration factor for any protein with a known amino acid sequence, where H_i is the hydration factor for amino acid "i," N_i is the number of amino acid "i" and M_r is the mass of protein.

(Equation 2.22)

$$\delta = \frac{18}{M_r}\sum H_i N_i$$

There are potential drawbacks to using this method, as there are concerns as to whether all of the assumptions made in this methodology hold true. In particular, there is reason to doubt the assumption that, regardless of the environment, all the amino acids are hydrated to the same extent. This is unlikely as with larger proteins some amino acids will be buried in an area less accessible to solvent, therefore those residues are very unlikely to be hydrated as fully as those which are fully solvent accessible. This can be seen in that, for the proteins with reported δ by Kuntz, the observed values are generally significantly lower than those calculated[31]. The computer program SEDNTERP (http://www.rasmb.bbri.org/) combined with the FASTA sequences available from the protein data bank for each protein used to calculate the hydration factors is detailed below in **Table 2-4**:

Protein	Mass (kDa)	PDB	δ ($g_{water}/g_{protein}$)
Cytochrome C	11.8	1HRC[19]	0.447
Lysozyme	14.3	1HEL[20]	0.338
Myoglobin	17.7	1WLA[21]	0.442
Chymotrypsin	25.7	1EX3[22]	0.328
Ovalbumin	44.3	1OVA[32]	0.328
Transthyretin	55.1	1F41[24]	0.436
Avidin	58.5	1VYO[25]	0.366
Bovine serum albumin	69.3	3V03[23]	0.450
Alcohol dehydrogenase	147.4	2HCY[26]	0.398
Aldolase	157.5	1ADO[27]	0.411
Pyruvate kinase	232.6	1F3W[28]	0.408
Glutamate	329.6	1NR7[29]	0.402
β-Galactosidase	475.6	1DP0[30]	0.385

Table 2-4: Data listing the hydration factors for each protein used in this work based upon the Kuntz method calculated using the PDB IDs listed above and the SEDNTERP software.

The second option is to assume that the δ is similar for all proteins and use one value for all. Values for δ have been reported as falling in between 0.2 and 0.5 across the literature, but the use of an average value for δ simplifies accounting for hydration, given most proteins with measured hydration factors fall inside the range of 0.2 to 0.4 for δ[33, 34]. There is a significant difference in the end results of the theoretical cross-sections depending on the method used for determining δ as can be seen in **Figure 2-10**.

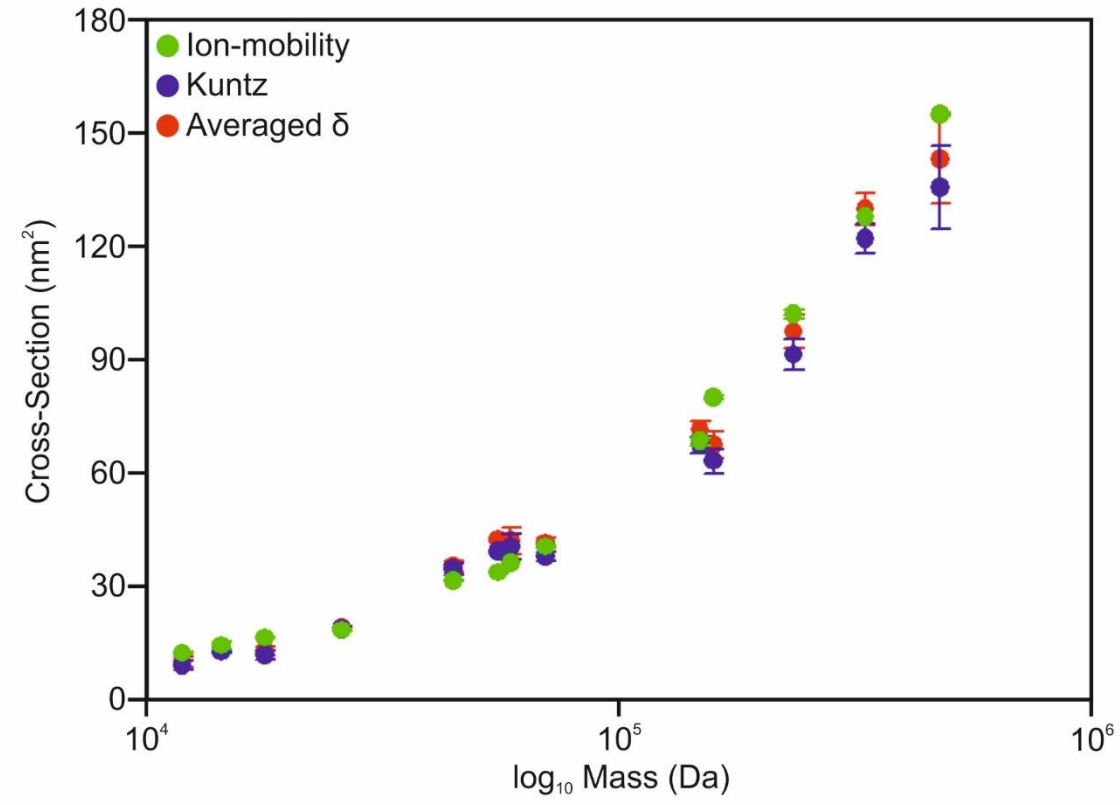


Figure 2-10: Figure showing the cross-sections calculated from the DLS data based on two different hydration factors either the Kuntz method (blue) or the averaged value of 0.3 (red) as well as the cross-sections obtained directly from ion-mobility measurements (green).

Figure 2-10 shows the difference between the ion mobility cross-sections and the ones derived from the two different options for δ. One of the most interesting results from this is that, independent of the methods used for determining the water associated with the protein, for the majority of proteins the largest collision cross-section is observed directly from the ion-mobility measurement. This result is particularly interesting since a main criticism of MS as a technique is the concern that proteins in the gas-phase might undergo collapse due to the hydrophobic environment in the vacuum, causing the protein to become unstable. If this was the case the results should show a larger cross-section for the proteins determined from their solution-phase properties. Whilst this was the case for the very initial results as shown by **Figure 2-9,** to not take into consideration the water associated with the protein in solution would be a considerable oversight.

2.7. Validation of the Millikan Drag Factor:

The final insight that this body of work can give with regard to comparisons of the solution phase and gas phase protein structures comes in the validation of the Millikan drag factor. This factor has been suggested in the past to have a value of 1.36, as discussed earlier in the Experimental Outline. With the data gathered from the DLS experiment it is also possible to experimentally determine a value for the drag factor by using *Equation 2.6*.

In order to do this, a comparison between the experimentally determined collision cross-sections from the IM-MS experiments and the de-wetted Stokes radii of the proteins can be made. Using *Equation 2.6* and substituting in the radius for the de-wetted protein ($R_{anyd.}$) for the radius of the ion (R_i) links together the experimentally determined collision cross-section and the solution-phase determined de-wetted Stokes radii, leading to *Equation 2.23* below:

(Equation 2.23)

$$\Omega = \pi(R_{anhyd.} + R_{He})^2 \times \xi$$

From *Equation 2.23,* if we substitute in using the equation for $R_{anhyd.}$ as given in *Equation 2.5,* we can directly link the calculated Stokes radii to the ion-mobility measurements as shown in *Equation 2.24*.

(Equation 2.24)

$$\Omega = \pi\left(\frac{R_s\bar{v}\rho}{\sqrt[3]{\bar{v}^2\rho^2(\bar{v}\rho+\delta)}} + R_{He}\right)^2 \times \xi$$

In order to use *Equation 2.24* it is important to note the specific partial volumes will be specific to each protein used. For this work the individual specific partial volumes were calculated from the protein sequences using the SEDNTERP computer program, allowing for

accurate calculation of expected cross-sections based on each measurement for the Stokes radius in solution.

In order to determine a value for the Millikan drag factor it is necessary to make a series of plots of the expected radii of the proteins in solution against their cross-section values available in the gas-phase. **Figures 2-11, 2-12 and 2-13** show a range of plots, each of which could be used to determine a value for the Millikan drag factor. Whilst **Figure 2-11** shows a simple plot of the experimentally determined Stokes radii against the collision cross-sections, **Figure 2-12 and 2-13** are for Stokes radii which have both attempted to compensate for the dewetting of the protein as discussed earlier, covering both the Kuntz method for accounting for protein hydration, as well as a simplified average of the hydration factors for all proteins used.

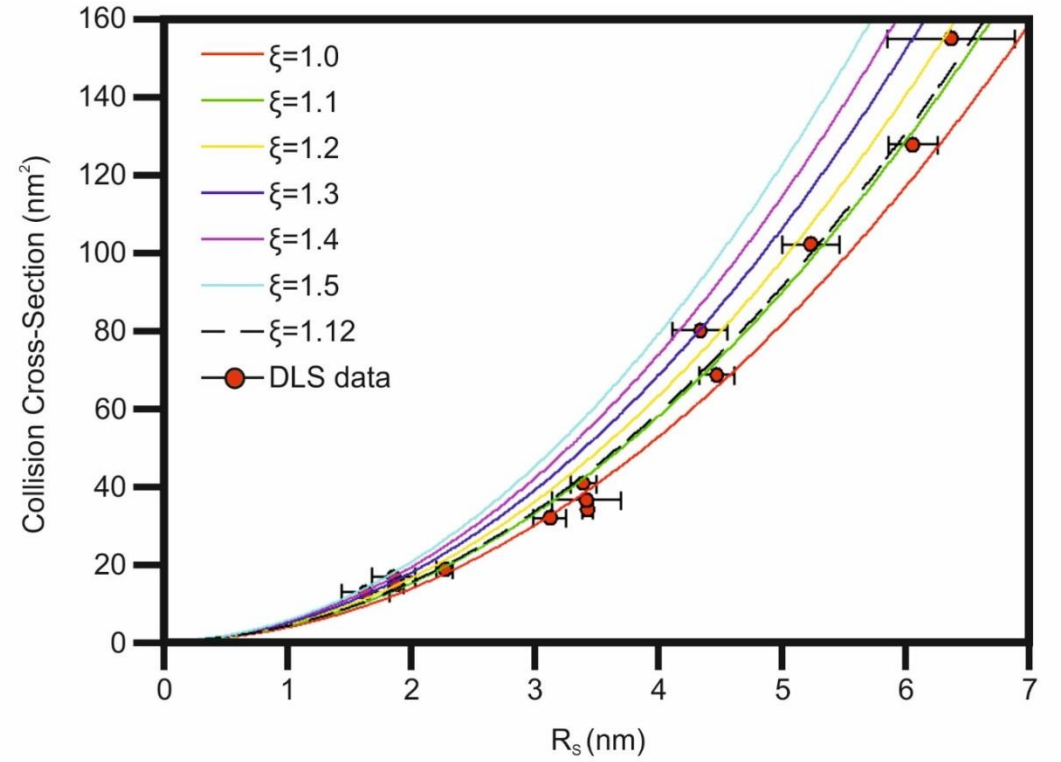


Figure 2-11: Figure comparing the radii measured directly from DLS data with the measured collision cross-sections from the ion-mobility experiment. In addition to the DLS data (red with black error bars), for reference a number of Millikan drag factors are plotted using Equation 2.23 (represented with solid lines of differing colours). Additionally, the best fit from minimising the R^2 error this is shown in a dashed black line.

Figure 2-11 shows that the calculated value for the Millikan drag factor is significantly at odds with the conventional expectation. However, as explained earlier in this chapter, this is not unexpected due to the significant effect of protein hydration. This merely reinforces the fact that not taking into account this affect when interpreting the DLS data could lead to a significant misinterpretation of the results.

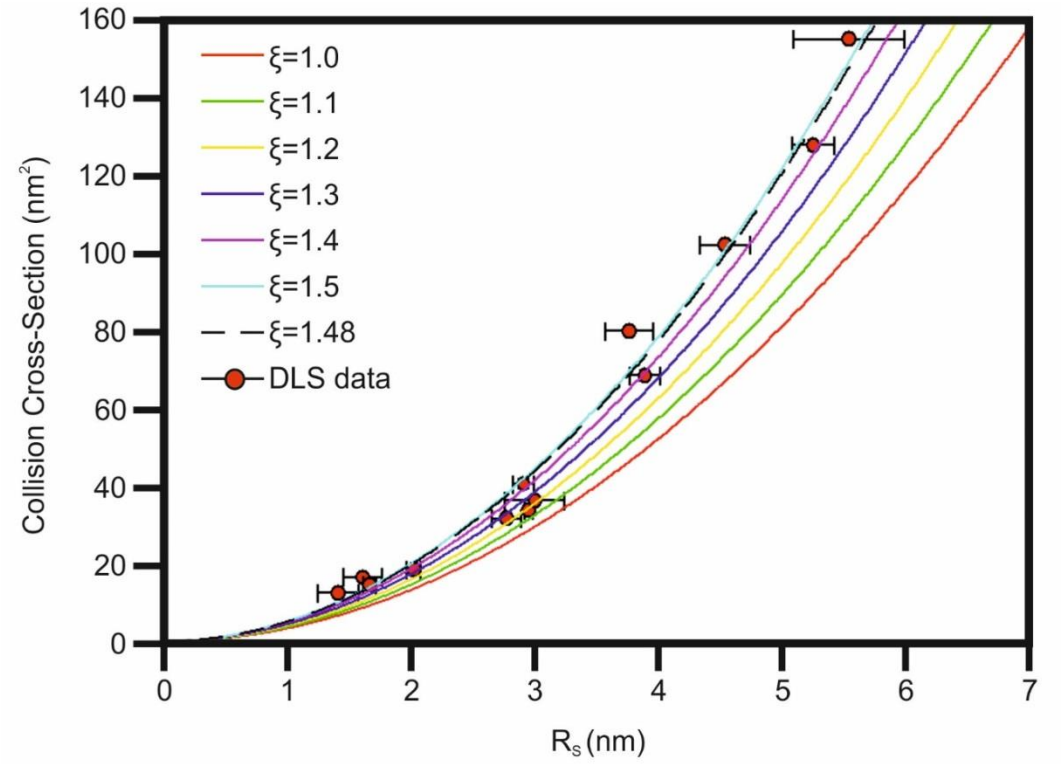


Figure 2-12: Figure comparing the radii measured directly from DLS data after dewetting the protein using the Kuntz methodology for determining the δ with the measured collision cross-sections from the ion-mobility experiment. In addition to the DLS data (red with black error bars), for reference a number of Millikan drag factors are plotted using Equation 2.23 (represented with solid lines of differing colours). Additionally, the best fit from minimising the R^2 error is shown in a dashed black line.

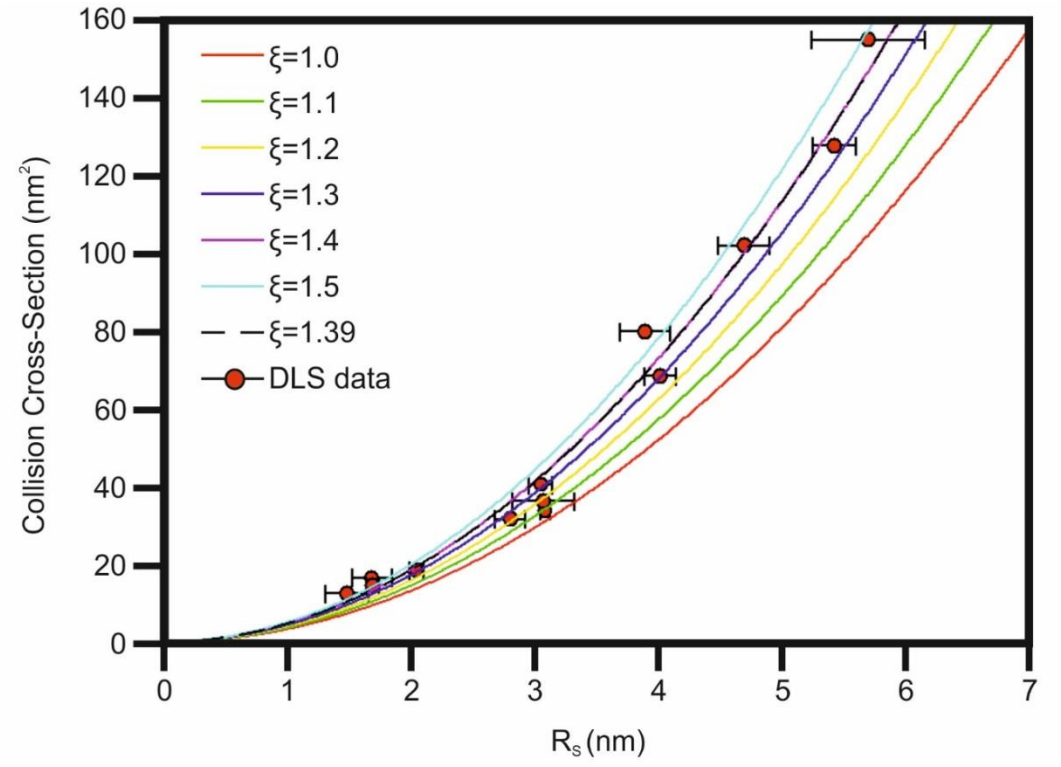


Figure 2-13: Figure comparing the radii measured directly from DLS data after dewetting the protein using a general δ value of 0.3 with the measured collision cross-sections from the ion-mobility experiment. In addition to the DLS data (red with black error bars), for reference a number of Millikan drag factors are plotted using Equation 2.23 (represented with solid lines of differing colours). Additionally the best fit from minimising the R^2 error is shown in a dashed black line.

Figures 2-12 and 2-13 demonstrate that the Millikan drag factor is in close agreement with the previously determined Millikan drag factor of 1.36. The value for the Millikan drag factor was determined by minimising the sum of mean squared errors as shown in *Equation 2.25,* where *E* is the difference between the expected value and the observed value, Ω is the collision cross-section, ξ is the Millikan drag factor, R_S, is the Stokes radii for the protein, and R_{He} is the radius of helium:

(Equation 2.25)

$$E = \sqrt{(\Omega - \xi * \pi(R_S + R_{He})^2)^2}$$

Depending on the methodology used to account for the hydration of the protein, the value calculated for the Millikan drag factor is calculated as either 1.48 using the Kuntz method for protein hydration factors or 1.39 using an average value of 0.3. Given the rigorous

investigations into the Millikan drag factor done previously, the much closer agreement with the known value of 1.36, using a δ of 0.3 instead of one calculated using the Kuntz method, suggests this is a better way to account for the hydration of the proteins in solution.

2.8. Conclusions:

There are a number of conclusions that can be drawn from the work detailed in this chapter. The first and most fundamental is that it is possible to obtain low resolution structural information about the proteins of interest using DLS and IM-MS. Secondly, by modelling the shape factors of the two different techniques we can see that it is possible to compare like for like between DLS and IM-MS as long as the proteins used are relatively globular in their structures.

When the DLS data is compared to the IM-MS data after appropriate steps have been taken to account for the hydration of the protein and the solvent viscosity, we can see a favourable comparison that suggests, within the resolution of the experiments, that the proteins maintain their solution-phase structure during the course of the IM-MS experiment. This work suggests that conclusions drawn from IM-MS data can be considered to be applicable to the proteins in the solution-phase as well.

In the final section of this work the comparison of the experimentally determined Millikan drag factors with the known values of 1.36 leads to one of two possible conclusions. The first of these is that with a δ of 0.3 $g_{water}/g_{protein}$ there is an excellent agreement with the expected value for the Millikan factor. The alternative, albeit less likely, conclusion could be that, in using a higher hydration factor (consistent with those calculated using the Kuntz method as detailed in **Table 2-4**), a higher Millikan drag factor is required in order to correlate the solution-phase and gas-phase data. The need for a higher Millikan drag factor

when it is known to be 1.36 could be interpreted as evidence for collapse of proteins in the gas-phase, as has been previously theorised. However, given that the Kuntz methodology is not well suited for use with larger protein complexes, as previously discussed, if there is an effect it is likely to small and not significant in terms of the overall protein structure.

2.9. References:

1. Kim, S.H. & Spangler, G.E. Ion Mobility Mass-Spectrometry of 2 Structurally Different Ions Having Identical Ion Mass. *Analytical Chemistry* **57**, 567-569 (1985).

2. Karpas, Z., Stimac, R.M. & Rappoport, Z. Differentiating Between Large Isomers and Derivation of Structural Information by Ion Mobility Spectrometry Mass-Spectrometry Techniques. *International Journal of Mass Spectrometry and Ion Processes* **83**, 163-175 (1988).

3. Clemmer, D.E. & Jarrold, M.F. Ion Mobility Measurements and Their Applications to Clusters and Biomolecules. *Journal of Mass Spectrometry* **32**, 577-592 (1997).

4. Schennach, M. & Breuker, K. Proteins with Highly Similar Native Folds Can Show Vastly Dissimilar Folding Behavior When Desolvated. *Angew Chem Int Ed Engl* **53**, 164-168 (2014).

5. Benesch, J.L. & Robinson, C.V. Biological Chemistry: Dehydrated but Unharmed. *Nature* **462**, 576-577 (2009).

6. 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).

7. Vonhelden, G., Hsu, M.T., Gotts, N. & Bowers, M.T. Carbon Cluster Cations with up to 84 Atoms - Structures, Formation Mechanism, and Reactivity. *Journal of Physical Chemistry* **97**, 8182-8192 (1993).

8. 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).

9. 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. *Journal of Physical Chemistry* **100**, 16082-16086 (1996).

10. Larriba, C. & Hogan, C.J. Free Molecular Collision Cross Section Calculation Methods for Nanoparticles and Complex Ions with Energy Accommodation. *Journal of Computational Physics* **251**, 344-363 (2013).

11. Bush, M.F. *et al.* Collision Cross Sections of Proteins and Their Complexes: a Calibration Framework and Database for Gas-Phase Structural Biology. *Anal Chem* **82**, 9557-9565 (2010).

12. Edward, J.T. Molecular Volumes and Stokes-Einstein Equation. *Journal of Chemical Education* **47**, 261-& (1970).

13. Chen, X.F., Weber, I. & Harrison, R.W. Hydration Water and Bulk Water in Proteins Have Distinct Properties in Radial Distributions Calculated from 105 Atomic Resolution Crystal Structures. *Journal of Physical Chemistry B* **112**, 12073-12080 (2008).

14. Millikan, R.A. The General Law of Fall of a Small Spherical Body Through a Gas, and its Bearing Upon the Nature of Molecular Reflection From Surfaces. *Physical Review* **22**, 1-23 (1923).

15. Epstein, P., S. On the Resistance Experienced by Spheres in their Motion Through Gases. *Phys. Rev.* **23**, 710-733 (1924).

16. Hogan, C.J., Jr. & de la Mora, J.F. Ion Mobility Measurements of Nondenatured 12-150 kDa Proteins and Protein Multimers by Tandem Differential Mobility Analysis-Mass Spectrometry (DMA-MS). *J Am Soc Mass Spectrom* **22**, 158-172 (2011).

17. Perrin, F. Mouvement Brownien d'un Ellipsoide (II). Rotation Libre et Dépolarisation des Fluorescences. Translation et Diffusion de Molécules Ellipsoidales. *Journal de Physique et le Radium* **7**, 1 - 11 (1936).

18. Vouk, V. Projected Area of Convex Bodies. *Nature* **162**, 330-331 (1948).

19. Bushnell, G.W., Louie, G.V. & Brayer, G.D. High-Resolution 3-Dimensional Structure of Horse Heart Cytochrome-C. *Journal of Molecular Biology* **214**, 585-595 (1990).

20. Wilson, K.P., Malcolm, B.A. & Matthews, B.W. Structural and Thermodynamic Analysis of Compensating Mutations Within the Core of Chicken Egg-White Lysozyme. *Journal of Biological Chemistry* **267**, 10842-10849 (1992).

21. Maurus, R. *et al.* A Myoglobin Variant with a Polar Substitution in a Conserved Hydrophobic Cluster in the Heme Binding Pocket. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology* **1341**, 1-13 (1997).

22. Pjura, P.E., Lenhoff, A.M., Leonard, S.A. & Gittis, A.G. Protein Crystallization by Design: Chymotrypsinogen Without Precipitants. *Journal of Molecular Biology* **300**, 235-239 (2000).

23. Majorek, K.A. *et al.* Structural and Immunologic Characterization of Bovine, Horse, and Rabbit Serum Albumins. *Molecular Immunology* **52**, 174-182 (2012).

24. Hornberg, A., Eneqvist, T., Olofsson, A., Lundgren, E. & Sauer-Eriksson, A.E. A Comparative Analysis of 23 Structures of the Amyloidogenic Protein Transthyretin. *Journal of Molecular Biology* **302**, 649-669 (2000).

25. Repo, S. *et al.* Binding Properties of HABA-Type Azo Derivatives to Avidin and Avidin-Related Protein 4. *Chemistry & Biology* **13**, 1029-1039 (2006).

26. Plapp, B.V. & Berst, K.B. *Human Alcohol Dehydrogenase 4: Mechanism, Specificity and Effects of Ethanol on Retinoid Metabolism*, vol. 12, 2006.

27. Blom, N. & Sygusch, J. Product Binding and Role of the C-Terminal Region in Class I D-Fructose 1,6-Bisphosphate Aldolase. *Nature Structural Biology* **4**, 36-39 (1997).

28. Wooll, J.O. *et al.* Structural and Functional Linkages Between Subunit Interfaces in Mammalian Pyruvate Kinase. *Journal of Molecular Biology* **312**, 525-540 (2001).

29. Banerjee, S., Schmidt, T., Fang, J., Stanley, C.A. & Smith, T.J. Structural Studies on ADP Activation of Mammalian Glutamate Dehydrogenase and the Evolution of Regulation. *Biochemistry* **42**, 3446-3456 (2003).

30. Juers, D.H. *et al.* High Resolution Refinement of beta-Galactosidase in a New Crystal Form Reveals Multiple Metal-Binding Sites and Provides a Structural basis for alpha-Complementation. *Protein Science* **9**, 1685-1699 (2000).

31. Kuntz Jr, I.D. Hydration of Macromolecules. III. Hydration of Polypeptides. *J Am Chem Soc*, 514-516 (1971).

32. Stein, P.E., Leslie, A.G.W., Finch, J.T. & Carrell, R.W. Crystal-Structure of Uncleaved Ovalbumin at 1.95 A Resolution. *Journal of Molecular Biology* **221**, 941-959 (1991).

33. Squire, P.G. & Himmel, M.E. Hydrodynamics and Protein Hydration. *Arch Biochem Biophys* **196**, 165-177 (1979).

34. Dam, J. & Schuck, P. Calculating Sedimentation Coefficient Distributions by Direct Modeling of Sedimentation Velocity Concentration Profiles. *Methods Enzymol* **384**, 185-212 (2004).

3. Comparison of Computationally Calculated Collision Cross-Sections with Experimentally Determined Collision Cross-Sections:

3.1. Introduction:

This chapter builds on the work done in the previous chapter by introducing a comparison of the experimental data with computationally derived data. With the growth in use of IM-MS data there have been multiple attempts to model the ion-mobility experiment computationally in order to have a comparison with known structures. These structures can be obtained using a variety of techniques such as; x-ray crystallography, solution phase NMR, solid state NMR and EM.

The comparison with computationally calculated data is particularly interesting given it might be possible in this data to see if a protein is undergoing large changes going into the gas-phase if it is possible to correctly predict the expected cross-section from structures in solution. In order to investigate the differences between the different methods all the proteins investigated in chapter two had their collision cross-sections calculated using a set of four different computational methods and then compared with data from direct gas-phase IM-MS collision cross-section data.

3.2. Experimental Outline:

To perform the experimental calculations a suitable atomic structure for each protein used needed to be found. In order to have consistency between the types of structures used for each calculation and the much greater availability of x-ray crystal structures it was decided to use structures that had been solved using that methodology. The PDB code for each structure used is listed below in **Table 3-1**:

Protein	Mass (kDa)	PDB code	n in PDB	n observed
Cytochrome C	11.8	1HRC[1]	1	1
Lysozyme	14.3	1HEL[2]	1	1
Myoglobin	17.7	1WLA[3]	1	1
Chymotrypsin	25.7	1EX3[4]	1	1
Ovalbumin	44.3	1OVA[5]	4	1
Transthyretin	55.1	1F41[6]	2	4
Avidin	58.5	1VYO[7]	2	4
Bovine serum albumin	69.3	3V03[8]	2	1
Alcohol dehydrogenase	147.4	2HCY[9]	4	4
Aldolase	157.5	1ADO[10]	4	4
Pyruvate kinase	232.6	1F3W[11]	8	4
Glutamate dehydrogenase	329.6	1NR7[12]	12	6
β-Galactosidase	475.6	1DP0[13]	4	4

Table 3-1: Details of proteins used during these experiments, the pdb codes used for the structures and the oligomeric states in the crystal structures and observed during IM-MS

Before any of these structures could be used for calculations there is a small amount of manipulation to do in order to make sure we making a fair comparison. This manipulation was necessary to obtain a structure that was reconcilable with the observed oligomeric states in the IM-MS and DLS experiments. There were three stages to this process, which was carried out by Erik Marklund: Firstly any atoms present in the protein sequence but not accounted for in the structure were added using the WHATIF server (http://swift.cmbi.ru.nl/servers/html/index.html). Secondly it was necessary to take into account some missing residues and build them on to the structures using ModLoop[14]. Finally in some cases it was necessary to actually alter the oligomeric state for some of the structures as can be seen in **Table 3-1**, the necessary adjustments were then made using UCSF Chimera[15]. Following this all structures were then protonated before running them through MOBCAL to determine the collision cross-sections for the PA[16], EHSS[17] and TJM[18] methods. For the PSA[19] method the structures were submitted to the PSA web server at 300K in Helium. The errors

reported for the calculations are one standard deviation from the mean for 16 different calculations for all the MOBCAL determined cross-sections, whilst for the PSA the error we record here is the same as taken directly from the report generated from the PSA server.

3.3. Computational Results:

All of the proteins listed in **Table 3-1** went through the manipulation process described in the experimental outline. The resulting computational calculations are tabulated below in **Table 3-2** with errors reported as described earlier. There are missing values for ΩPSA for β-Galactosidase due to an error when running the calculation on the PSA server:

Protein	Mass	Ω_{PA}	Ω_{EHSS}	Ω_{TJM}	Ω_{PSA}
Cytochrome C	11.8	10.9±0.1	13.5±0.0	13.5±0.1	13.2±0.1
Lysozyme	14.3	12.1±0.1	15.2±0.0	15.1±0.1	14.6±0.2
Myoglobin	17.7	14.0±0.1	17.6±0.0	17.7±0.1	17.3±0.2
Chymotrypsin	25.7	17.4±0.1	22.4±0.0	22.4±0.1	21.5±0.2
Ovalbumin	44.3	26.8±0.2	34.9±0.0	35.2±0.2	33.3±0.4
Transthyretin	55.1	32.1±0.3	41.6.±0.0	41.3±0.3	37.2±0.4
Avidin	58.5	32.0±0.2	41.7±0.0	41.7±0.3	38.8±0.3
Bovine serum albumin	69.3	38.7±0.3	51.0±0.0	50.8±0.3	47.2±0.5
Alcohol dehydrogenase	147.4	62.3±0.6	83.3±0.1	82.7±0.4	78.2±0.8
Aldolase	157.5	68.7±0.5	90.3±0.1	90.4±0.5	80.4±0.7
Pyruvate kinase	232.6	86.8±0.8	116.9±0.1	117.1±0.6	107.8±1.6
Glutamate dehydrogenase	329.6	112.0±0.8	149.7±0.1	148.8±0.7	169.7±2.3
β-Galactosidase	475.6	138.9±1.1	188.9±0.2	189.1±1.1	NA

Table 3-2: Summary of computationally calculated cross-sections for all the proteins used in this work.

The first and most immediate observation from these initial results is the difference between the calculated cross-sections from different methodologies. The most dramatic difference comes in the calculation based on the projection approximation. This is not necessarily surprising in and of itself due to the relatively simplistic nature of the

method. It has been shown previously that the best predictor for proteins collision cross-sections in the gas phase is a scaled projection approximation. This is done using a scaling factor of 1.14 which has been empirically derived from ion-mobility studies done previously[20].

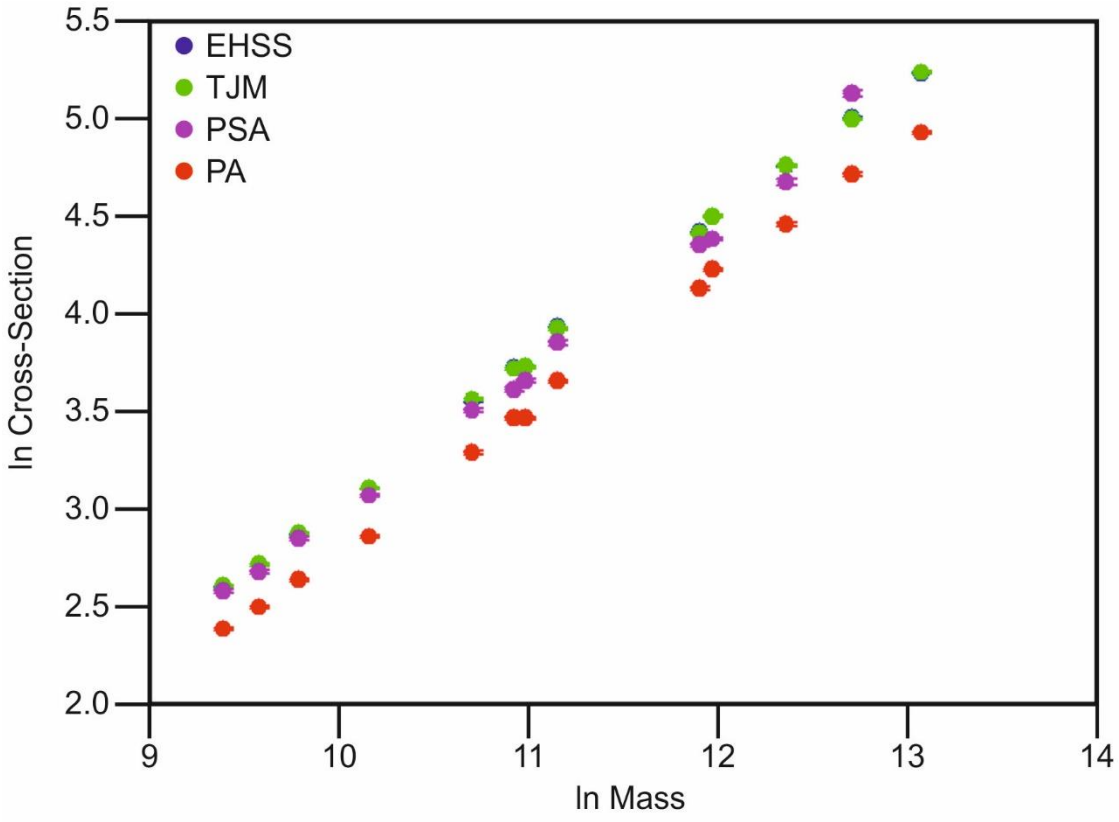


Figure 3-1: Plot of the ln of the mass of the proteins studied against the ln of the cross-sections determined using the computational methods described previously.

The second observation from this data is the relative consistency in terms of predicted cross-sections for the protein complexes for the EHSS and the TJM methodologies. Given the large differences in the computational power required to run the TJM compared to the EHSS it is remarkable that almost all the data points for the EHSS in **Figure 3-1**, are obscured by the corresponding value determined using the TJM. In addition to the very high level of correlation between EHSS and TJM calculations the PSA calculations also appear to be in close agreement with the PSA seeming to calculate a slightly lower value for the cross-sections across the board with the sole exception of Glutamate

Dehydrogenase where the PSA calculated is markedly larger than that determined with the other methods.

3.4. Comparison with DLS Experimental Data:

The comparison between the different computational methods and the experimentally determined collision cross-sections from DLS can be seen in **Figure 3-2**. All four computational methods are plotted alongside the experimental data, a scaled projection approximation with a 1.14 scaling factor and a linear line of best fit for all the experimental DLS data.

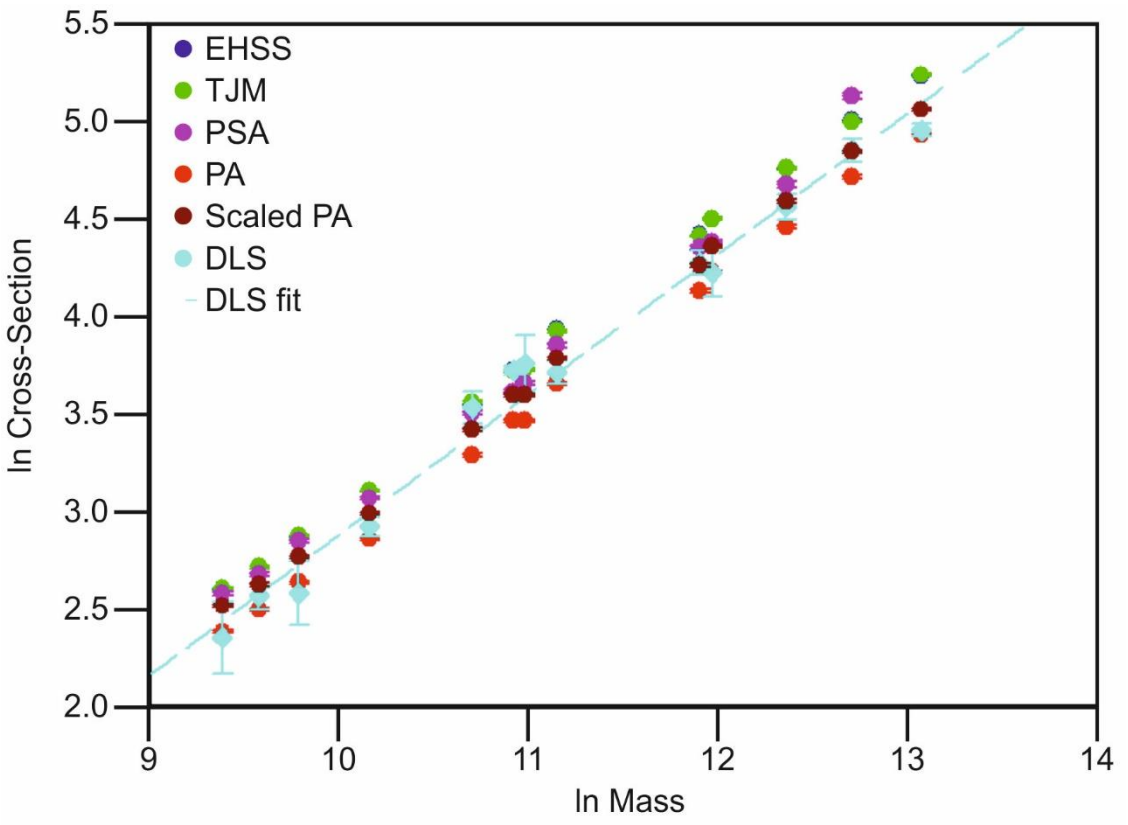


Figure 3-2: Graph showing the comparison between the various computational methods and the collision cross-sections derived from the DLS data.

From first observation from **Figure 3-2** is that the line of best fit for all the DLS data points suggests that a scaled projection approximation most accurately predicts the collision cross-section expected for any protein using the methodology described in the previous chapter.

However in spite of the close agreement in general for the line of best fit with the scaled projection approximation some of the data points do fall outside the margin of error for the experiment. These discrepancies between the computational methodology and the experimental data are explainable in one of three potential ways:

Firstly it is possible that there is a difference between the protein structures used for the computational calculations and the proteins structure in solution. The second explanation is a shortfall in the computational methodology leading to the discrepancy between the experimental data points and those determined computationally from known crystal structures. The final explanation for the deviations observed could be explained due to flaws in the methodology used to calculate the predicted gas-phase collision cross-section from the solution-phased DLS experiment. It is possible that the deviation from perfect correlation between the computationally calculated cross-sections and those calculated from DLS could be down to any one of these factors or a combination of all three.

3.5. Comparison with IM-MS Experimental Data:

In **Figure 3-3** you can see the comparison between the collision cross-sections measured directly using IM-MS and the computational methods described previously. Similarly to **Figure 3-2** all four computational methods are plotted alongside the experimental data, a line of best fit for the experimental date as well as the scaled projection approximation with a 1.14 scaling factor. It is worth noting that given the size of the scaling factor was determined from direct ion-mobility measurements we would expect to see a great deal of correlation between the IM-MS data and the scaled PA.

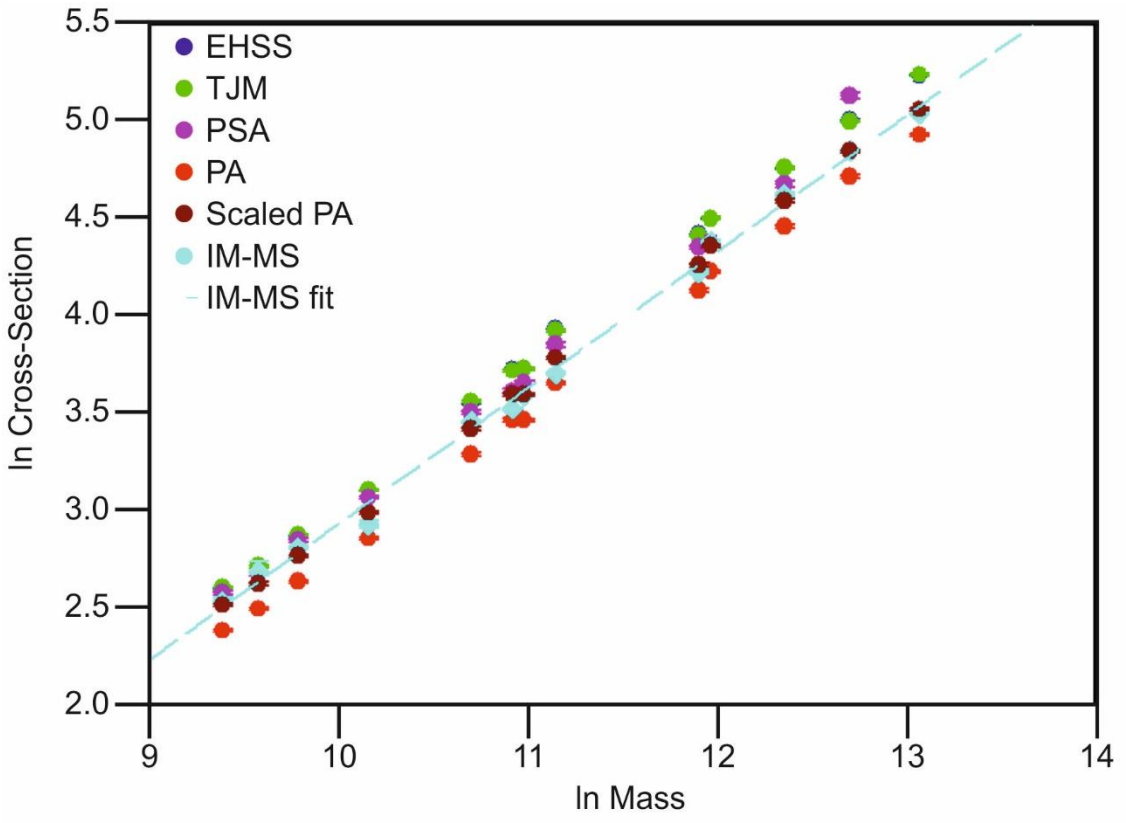


Figure 3-3: Graph showing the comparison between the various computational methods and the collision cross-sections from the IM-MS experiment.

The first observation to make is that similar to the DLS data the closest match between the experimental data and the computational methods comes with the scaled projection approximation. Despite the similarity with the DLS data, shown by the best fit still being the scaled projection approximation, there is a difference here illustrated by better agreement with the scaled projection approximation. This is evidenced by all bar one protein being inside the margin of error of the experiment from the cross-section calculated using the scaled projection approximation.

The second observation is that the next closest to fitting the ion-mobility data is the PSA where the difference between the experimental data and the PSA calculations only falls marginally outside the error of the experiment it does however still consistently

overestimate the collision cross-section when compared with the experimentally determined data.

There are a number of plausible explanations for the differences between the computational methods and the ion-mobility collision cross-sections and those directly determined from computational methodologies:

Firstly there is the possibility that during the ion-mobility experiment the structure of the protein differs from the structure used for the computational calculations. This could be down to either of two factors; that the protein might undergo collapse during the gas-phase leading to the experimental measurement being lower than the computational predictions, conversely it could be that the structure for the protein when crystallised differs slightly from the structure in solution allowing for a small discrepancy between the predicted and observed cross-sections.

Secondly there is the possibility that there is still a slight lack of understanding in all the factors at work during the ion-mobility experiment. This could lead to the calculations not including all relevant influences on the protein as it passes through the mobility cell and hence not completely replicating experimental conditions. This result would be interesting particularly in light of the most recently published PSA model providing the closest direct predictions for the collision cross-sections.

3.6. Comparisons of Structure Solving Methodologies:

The majority of proteins with known structures were solved using x-ray crystallography as shown in **Figure 1-1**. There is a possibility that the method used for solving the structure in the PDB might be a cause of the differences between the computationally calculated cross-sections and those from DLS and IM-MS. In order to investigate if the method used for solving the protein structure does have a significant effect on the calculated cross-sections it was decided to explore the results when the protein's structure has been solved using alternative methods.

X-ray crystallography is the primary method used for solving protein structures as demonstrated earlier. However as the name suggests this requires the proteins to be crystallised for the technique to be used, this in turn means the proteins are subjected to non-physiological conditions in order to facilitate their crystallisation. In order to investigate if these different conditions are the cause of the discrepancies noted earlier alternative structures for the proteins used, but solved with alternative methods were searched for. Following a search in the protein data bank[21] structures for a few of the proteins were found which had been solved using either solution-phase NMR or using cryo-EM.

Whilst proteins are often exposed to non-physiological conditions for crystallisation in theory this can be avoided with cryo-EM and solution NMR. The most obvious difference is for solution NMR structures the proteins can exist in their natural form as opposed to being crystallised.

The potential advantages of cryo-EM over x-ray crystallography are similar to the advantages of solution-phase NMR. Through the use of flash freezing the proteins and then analysing them at cryogenic temperatures is believed to allow the proteins to maintain their

native structure whilst undergoing analysis. There remains a significant drawback in using cryo-EM for the solving of protein structures. The resolution of the cryo-EM experiment compared to X-ray crystallography is much lower which, is a factor to be aware of when using structures solved using this methodology.

For the thirteen proteins used in this study, only four of them have structures solved by an alternate method for the same protein from the same organism. In addition, myoglobin has an NMR structure available from a different organism with 88% sequence similarity. The different proteins and their PDB codes are documented below in **Table 3-3:**

Protein	PDB code	Method	n PDB	n observed
Cytochrome C	1HRC	X-Ray	1	1
Cytochrome C	1AKK[22]	NMR	1	1
Lysozyme	1HEL	X-Ray	1	1
Lysozyme	1E8L[23]	NMR	1	1
Myoglobin	1WLA	X-Ray	1	1
Myoglobin*	1MYF[24]	NMR	1	1
Glutamate dehydrogenase	1NR7	X-Ray	12	6
Glutamate dehydrogenase	3JCZ[25]	Cryo-EM	6	6
β-Galactosidase	1DP0	X-Ray	4	4
β-Galactosidase	3J7H[26]	Cryo-EM	4	4

Table 3-3: For the proteins used which have structures solved by an alternate method to x-ray crystallography detailing what the alternate method is and the PDB reference for the structure.

The most immediate factor that stands out is the difference in oligomeric state for glutamate dehydrogenase. The structure determined during Cryo-EM has the same oligomeric state as observed during the IM-MS experiment. Whilst this is by no means definitive evidence that cryo-EM structures are a more appropriate base from which to work when attempting to predict gas-phase collision cross-sections it is an encouraging start point. The structures used for each comparison are shown below with the x-ray crystal structures on the left and the alternately solved structures on the right:

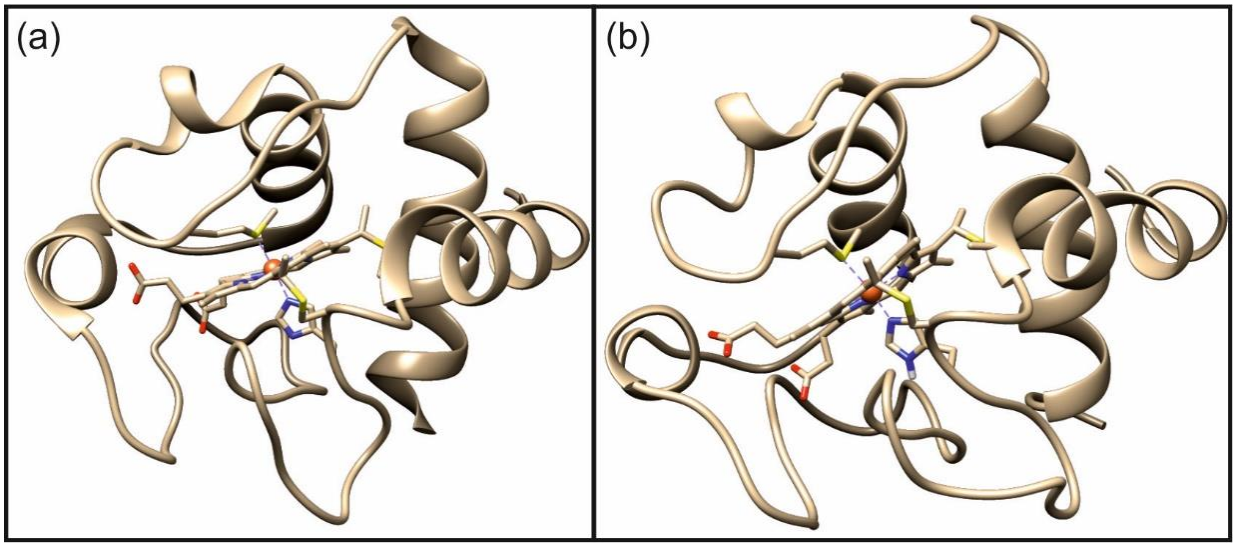


Figure 3-4: Cartoons of the X-ray crystal[1] (a) and solution NMR[22] (b) structures for cytchrome C using the PDB codes listed in Table 3-5 above and then processed before running the computational methods.

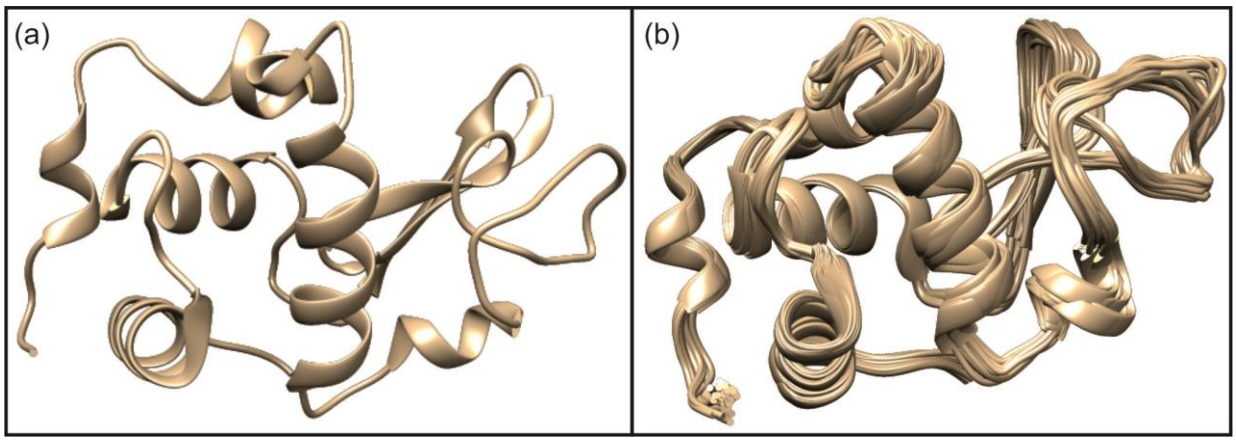


Figure 3-5: Cartoons of the X-ray crystal[2] (a) and the full ensemble of solution NMR[23] (b) structures for lysozyme using the PDB codes listed in Table 3-5 above and then processed before running the computational methods.

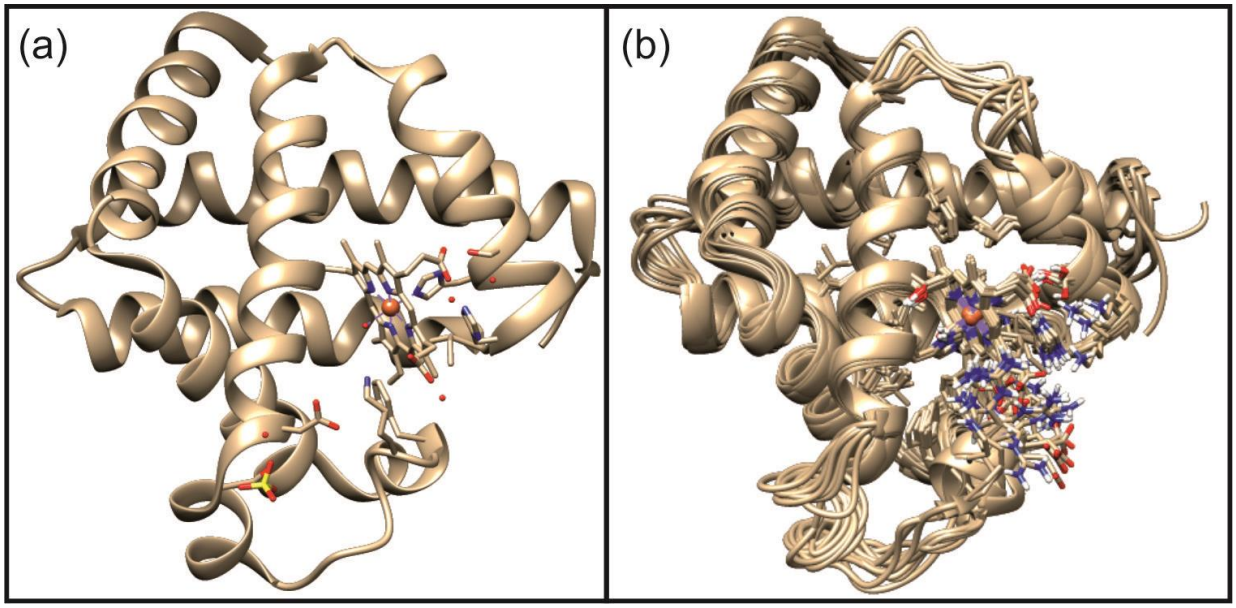


Figure 3-6: Cartoons of the X-ray crystal[3] (a) and the full ensemble of solution NMR[24] (b) structures for myoglobin using the PDB codes listed in Table 3-5 above and then processed before running the computational methods.

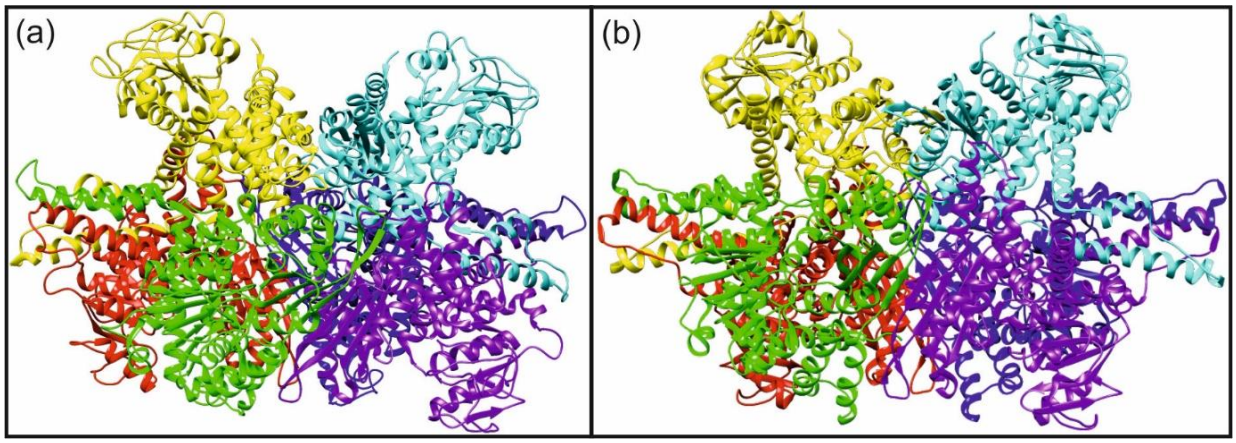


Figure 3-7: Cartoons of the X-ray crystal[12] (a) and the cryo-EM[25] (b) structures for glutamate dehydrogenase using the PDB codes listed in Table 3-5 above and then processed before running the computational methods. Each chain is coloured individually for ease of comparison between the two structures.

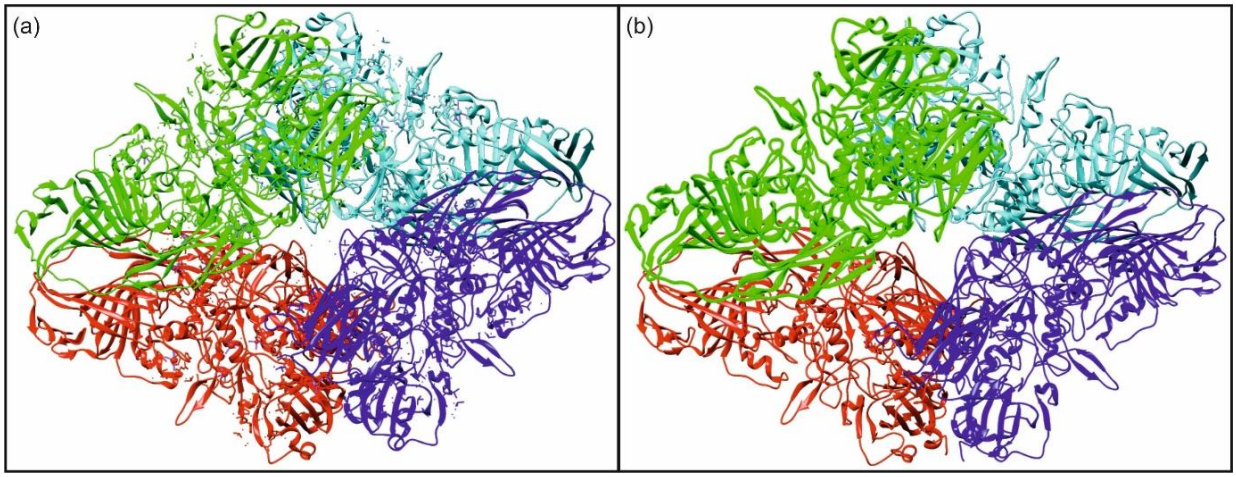


Figure 3-8: Cartoons of the X-ray crystal[13] (a) and the cryo-EM[26] (b) structures for β-galactosidase using the PDB codes listed in Table 3-5 above and then processed before running the computational methods. Each chain is coloured individually for ease of comparison between the two structures.

It can be seen in **Figures 3-4 – 3-8** that there is relatively little difference between the crystal structures and those solved using alternative methods. This is encouraging as we would anticipate the structures to be without any significant differences, any differences easily observable by eye would suggest too big a discrepancy to suggest both structures are of the protein in the same conformation.

Figures 3-5 (b) and **3-6 (b)** demonstrate that there are multiple possible conformers for both lysozyme (50) and myoglobin (12). There are two possible options for dealing with these structures with multiple conformers; firstly an average structure could be determined by averaging the positions for each atom, alternatively each computational method could be applied to each conformer and then an average taken at the end. Whilst opting to use and averaged structure would reduce the computational demands, by only requiring one calculation for each protein per method, averaging the structure in this way has potential drawbacks. This is because there is no guarantee that an average of all of the structures in the ensemble would provide and accurate representation of the structure of the protein.

Due to concerns about the validity of generating average structures for the ensemble the more computationally rigorous approach of determining cross-sections was chosen. This meant running all the methods for each conformer making up the ensemble and then averaging the results to calculate the collision cross-section.

Before any of these structures can be used to calculate collision cross-sections they were all manipulated as described earlier to account for missing atoms or residues. There was no need to adjust for the correct oligomeric state due to all of these structures having matching oligomeric states between their solved structures and the observed during IM-MS. The structures were then protonated as described previously to account for the full protein structure. Theoretical collision cross-sections were then determined as done previously for the x-ray crystal structures allowing for the generation of cross-sections based on the PA, EHSS and TJM methods as shown in **Table 3-4** below. The PSA was not calculated for this work due to technical difficulties with the PSA web server:

Protein	Method	Ω_{PA}	Ω_{EHSS}	Ω_{TJM}
Cytochrome C	X-Ray	10.9±0.1	13.5±0.0	13.5±0.1
Cytochrome C	NMR	11.7±0.1	14.6±0.1	14.3±0.1
Lysozyme	X-Ray	12.1±0.1	15.2±0.0	15.1±0.1
Lysozyme	NMR	14.0±0.2	17.4±0.2	16.9±0.2
Myoglobin	X-Ray	14.0±0.1	17.6±0.0	17.7±0.1
Myoglobin*	NMR	14.4±0.2	18.0±0.3	18.0±0.2
Glutamate dehydrogenase	X-Ray	112.0±0.8	149.7±0.1	148.8±0.7
Glutamate dehydrogenase	Cryo-EM	107.6±1.1	145.2±1.7	144.7±0.9
β-Galactosidase	X-Ray	138.9±1.1	188.9±0.2	189.1±1.1
β-Galactosidase	Cryo-EM	140.0±1.6	188.6±2.5	187.3±1.1

Table 3-4: Comparison between the calculated cross-sections derived from a variety of experimental methodologies.

The comparisons between the collision cross-sections determined from conventionally solved structures (using X-Ray crystallography) and structures solved with solution NMR

(Figure 3-9) and cryo-EM **(Figure 3-10)** are shown below. Values from direct IM-MS measurements and the DLS derived cross-sections are also included for reference.

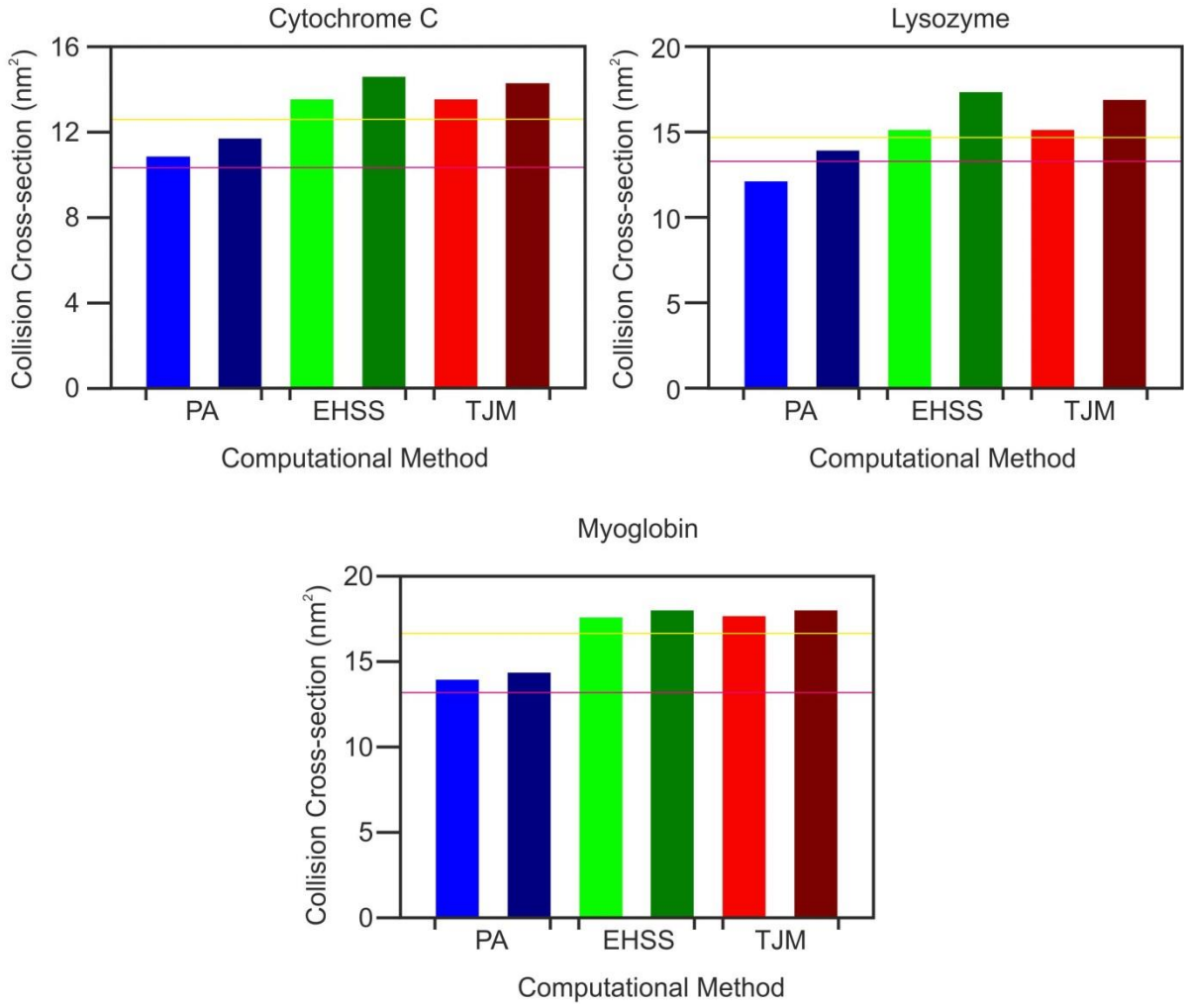


Figure 3-9: Summary of the computationally derived collision cross-sections for x-ray structures (light) and solution NMR structures (dark) for reference the directly measured ion-mobility cross-sections (yellow) and DLS derived cross-sections (pink) are shown.

Comparison between the NMR structures and x-ray crystallography structures all show a significant increase in the calculated cross-sections. With an average increase of 8% of the cross-section when determined from solution NMR compared to the x-ray. This is a significant increase and is difficult to rationalise given the stark contrast to the experimentally determined data. The computational data suggests that with either method for solving the protein structures the more rigorous TJM and EHSS methods significantly overestimate the cross-section when compared to the experimental data.

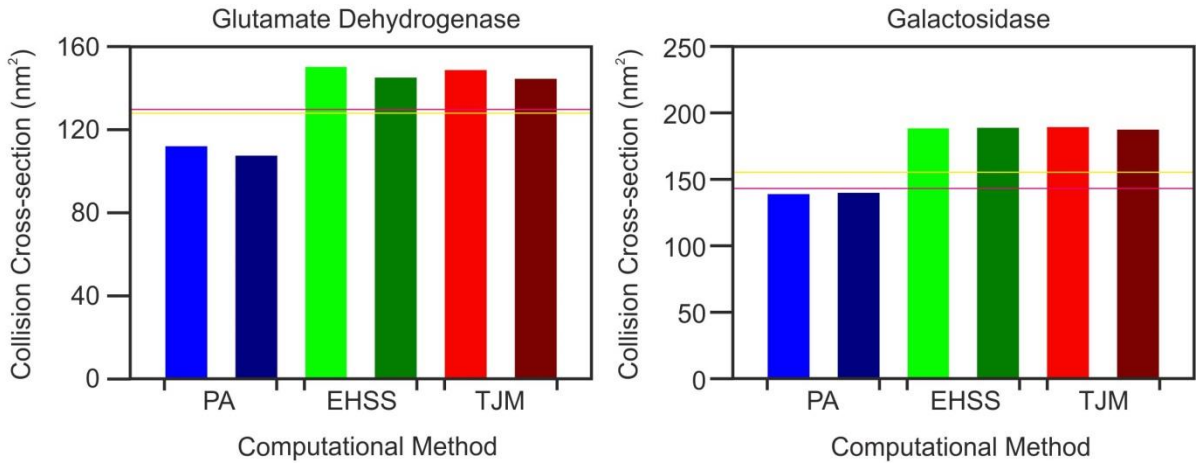


Figure 3-10: Summary of the computationally derived collision cross-sections for x-ray structures (light) and cryo-EM structures (dark) for reference the directly measured ion-mobility cross-sections (yellow) and DLS derived cross-sections (pink) are shown.

Comparisons for the two proteins with structures solved using cryo-EM with their x-ray analogues shows some interesting features. Overall there is a 2% drop in the calculated collision cross-section for the cryo-EM structures compared with those solved by the more traditional methodology. There is a difference for the two proteins however with GDH showing a marked drop in in the calculated cross-sections whereas the β-galactosidase returns virtually identical cross-sections. Whilst the data does show a slight decrease in the measured cross-sections it is only marginal and with the relative scarcity of proteins solved using multiple methods it is not practical to draw any significant conclusions from the data available.

3.7. Conclusions:

This work was done in an effort to rationalise the discrepancies between calculated collision cross-sections, based on solved protein structures, and the experimentally measured cross-sections, both from IM-MS and those derived from DLS measurements. Areas of interest in particular was investigating the relationship between the DLS derived cross-sections and the range of computational methods available.

Initial results seen previously where directly measured cross-sections from IM-MS are lower than all cross-sections determined using computational methods apart from the projection approximation. Furthermore, in comparison to the DLS measurements, all of the computational techniques, PA aside, calculate a significantly larger collision cross-section than is determined by a calculation based off the DLS determined protein radius. This leaves a significant disconnect between the computational methodologies currently available and the IM-MS experiment.

The hypothesis that a potential shortfall might not be in the computational methods, was then investigated, however, there is still a lack of readily available proteins whose structures have been solved by a wide range of methods. With the limited availability of proteins for investigation some interesting results were still yielded. NMR solved structures resulted in a significant increase in the collision cross-sections in contrast to the x-ray structures. Conversely the cryo-EM derived structures saw a mild decrease in collision cross-section when contrasted with the conventional x-ray structures.

Whilst the results are of interest, showing that one of the greatest differentiators for the final calculated cross-sections can be the method used in determining the structure, there is not currently a large enough library of structures for conclusive results to be obtained. At the time of writing just over 1,200 proteins have structures solved using at least two of cryo-EM, NMR and x-ray crystallography[27]. Whilst this would provide a significant number of proteins with which to perform a study, upon further investigation many of these in one method are solved under non-native conditions or have significant differences in sequence or host organism. In spite of that this work demonstrates there is potential for future investigation down this avenue to provide a more comprehensive overview.

3.8. References:

1. Bushnell, G.W., Louie, G.V. & Brayer, G.D. High-Resolution 3-Dimensional Structure of Horse Heart Cytochrome-C. *Journal of Molecular Biology* **214**, 585-595 (1990).

2. Wilson, K.P., Malcolm, B.A. & Matthews, B.W. Structural and Thermodynamic Analysis of Compensating Mutations Within the Core of Chicken Egg-White Lysozyme. *Journal of Biological Chemistry* **267**, 10842-10849 (1992).

3. Maurus, R. *et al.* A Myoglobin Variant with a Polar Substitution in a Conserved Hydrophobic Cluster in the Heme Binding Pocket. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology* **1341**, 1-13 (1997).

4. Pjura, P.E., Lenhoff, A.M., Leonard, S.A. & Gittis, A.G. Protein Crystallization by Design: Chymotrypsinogen Without Precipitants. *Journal of Molecular Biology* **300**, 235-239 (2000).

5. Stein, P.E., Leslie, A.G.W., Finch, J.T. & Carrell, R.W. Crystal-Structure of Uncleaved Ovalbumin at 1.95 A Resolution. *Journal of Molecular Biology* **221**, 941-959 (1991).

6. Hornberg, A., Eneqvist, T., Olofsson, A., Lundgren, E. & Sauer-Eriksson, A.E. A Comparative Analysis of 23 Structures of the Amyloidogenic Protein Transthyretin. *Journal of Molecular Biology* **302**, 649-669 (2000).

7. Repo, S. *et al.* Binding Properties of HABA-Type Azo Derivatives to Avidin and Avidin-Related Protein 4. *Chemistry & Biology* **13**, 1029-1039 (2006).

8. Majorek, K.A. *et al.* Structural and Immunologic Characterization of Bovine, Horse, and Rabbit Serum Albumins. *Molecular Immunology* **52**, 174-182 (2012).

9. Plapp, B.V. & Berst, K.B. *Human Alcohol Dehydrogenase 4: Mechanism, Specificity and Effects of Ethanol on Retinoid Metabolism*, vol. 12, 2006.

10. Blom, N. & Sygusch, J. Product Binding and Role of the C-Terminal Region in Class I D-Fructose 1,6-Bisphosphate Aldolase. *Nature Structural Biology* **4**, 36-39 (1997).

11. Wooll, J.O. *et al.* Structural and Functional Linkages Between Subunit Interfaces in Mammalian Pyruvate Kinase. *Journal of Molecular Biology* **312**, 525-540 (2001).

12. Banerjee, S., Schmidt, T., Fang, J., Stanley, C.A. & Smith, T.J. Structural Studies on ADP Activation of Mammalian Glutamate Dehydrogenase and the Evolution of Regulation. *Biochemistry* **42**, 3446-3456 (2003).

13. Juers, D.H. *et al.* High Resolution Refinement of beta-Galactosidase in a New Crystal Form Reveals Multiple Metal-Binding Sites and Provides a Structural basis for alpha-Complementation. *Protein Science* **9**, 1685-1699 (2000).

14. Fiser, A., Do, R.K. & Sali, A. Modeling of Loops in Protein Structures. *Protein Sci* **9**, 1753-1773 (2000).

15. Pettersen, E.F. *et al.* UCSF Chimera--a Visualization System for Exploratory Research and Analysis. *J Comput Chem* **25**, 1605-1612 (2004).

16. Vonhelden, G., Hsu, M.T., Gotts, N. & Bowers, M.T. Carbon Cluster Cations with up to 84 Atoms - Structures, Formation Mechanism, and Reactivity. *Journal of Physical Chemistry* **97**, 8182-8192 (1993).

17. 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).

18. 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. *Journal of Physical Chemistry* **100**, 16082-16086 (1996).

19. 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).

20. Bush, M.F. *et al.* Collision Cross Sections of Proteins and Their Complexes: a Calibration Framework and Database for Gas-Phase Structural Biology. *Anal Chem* **82**, 9557-9565 (2010).

21. Berman, H.M. *et al.* The Protein Data Bank. *Nucleic Acids Research* **28**, 235-242 (2000).

22. Banci, L. *et al.* Solution Structure of Oxidized Horse Heart Cytochrome C. *Biochemistry* **36**, 9867-9877 (1997).

23. Schwalbe, H. *et al.* A Refined Solution Structure of Hen Lysozyme Determined Using Residual Dipolar Coupling Data. *Protein Science* **10**, 677-688 (2001).

24. Osapay, K., Theriault, Y., Wright, P.E. & Case, D.A. Solution Structure of Carbonmonoxy Myoglobin Determined from Nuclear-Magnetic-Resonance Distance and Chemical-Shift Constraints. *Journal of Molecular Biology* **244**, 183-197 (1994).

25. Borgnia, M.J. *et al.* Using Cryo-EM to Map Small Ligands on Dynamic Metabolic Enzymes: Studies with Glutamate Dehydrogenase. *Molecular Pharmacology* **89**, 645-651 (2016).

26. Bartesaghi, A., Matthies, D., Banerjee, S., Merk, A. & Subramaniam, S. Structure of beta-Galactosidase at 3.2-Angstrom Resolution Obtained by Cryo-Electron Microscopy. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 11709-11714 (2014).

27. Berman, H.M. *et al.* The Protein Data Bank. *Nucleic Acids Research*, 235-242 (2000).

4. Novel Vehicles for Membrane Protein Mass Spectrometry:

4.1. Introduction:

The hydrophobic portion of membrane proteins that sits embedded inside the cell membrane bilayer (**Figure 4-1)** has always posed difficulties in the handling of this type of protein[1]. The same properties that are essential for a membrane protein to function in the membrane, also form significant challenges when trying to analyse them as this requires them to be handled in an aqueous environment.

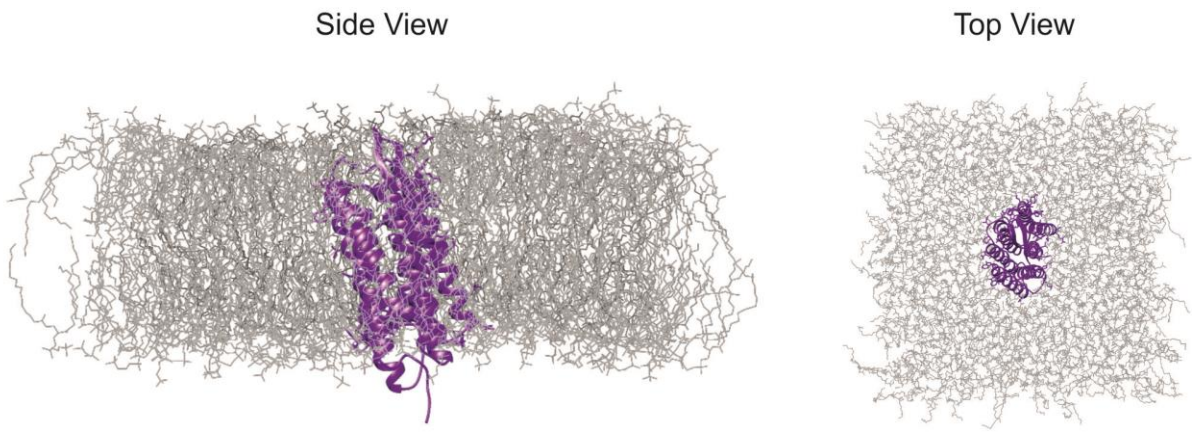


Figure 4-1: Bacteriorhodopsin (purple) inside a lipid bilayer (grey) as it would exist in the cell. Structure taken from PDB ID 4DM2 (currently unpublished but structure available from the PDB[2]) and inserted into a membrane using MemProt MD[3]

To study membrane proteins requires solubilisation of the protein from the cell membrane, using an excess of detergent under ill-defined conditions. This process causes the membranes to break up as they become solubilised into detergent micelles[4]. However, selecting conditions and a detergent that maintains the protein's stability and authentic conformation is challenging.

Whilst the process described above is an effective way to solubilise membrane proteins into a form where purification for other components can take place easily with chromatography, there are some drawbacks.

Firstly, with detergents not being naturally present in the cell, there are concerns that they can distort the structure or inhibit the function of the protein. Secondly, the detergents have a delipidating effect on proteins[5], which, whilst necessary for their solubilisation from the membrane, can cause changes in the proteins structure[6, 7, 8]. Finally another significant difference between a detergent solubilised membrane protein and one in the cell membrane is the lack of membrane lateral pressure.

As a result of these pitfalls associated with using detergents, a number of alternative methodologies have been developed, with the idea of keeping the protein in a more native-like environment[9, 10]. Generally speaking these split into one of two approaches; either using lipid vehicles to solubilise the membrane protein, or, the creation of an artificial vehicle designed to closely mimic the native lipid environment without lipids.

With the first approach the main methods used are liposomes, bicelles, nanodiscs and lipodisqs as shown in **Figure 4-2**. The proteins are initially solubilised using detergent and then exchanged into protein-lipid complexes. Nanodiscs are made up of three components, **(Figure 4-2 (a))** the membrane scaffold protein (blue), the lipids (purple) and the membrane protein (red). Similarly Lipodisqs **(Figure 4-2 (b))** are also made up of three components except the membrane scaffold protein is replaced with a non-protein scaffold, e.g. a polymer (blue). The membrane protein (green) is still surrounded by lipids (grey). Bicelle solubilisation consists of three components as well, as seen in **Figure 4-2 (c)**, the membrane protein (red) is embedded in a mini bilayer composed of lipids (purple), with the ends of the bilayer capped with detergent molecules (blue). In the case of liposomes, as seen in **Figure 4-2 (d)**, the entire vehicle consists of lipids (blue) with the membrane protein (green) embedded inside.

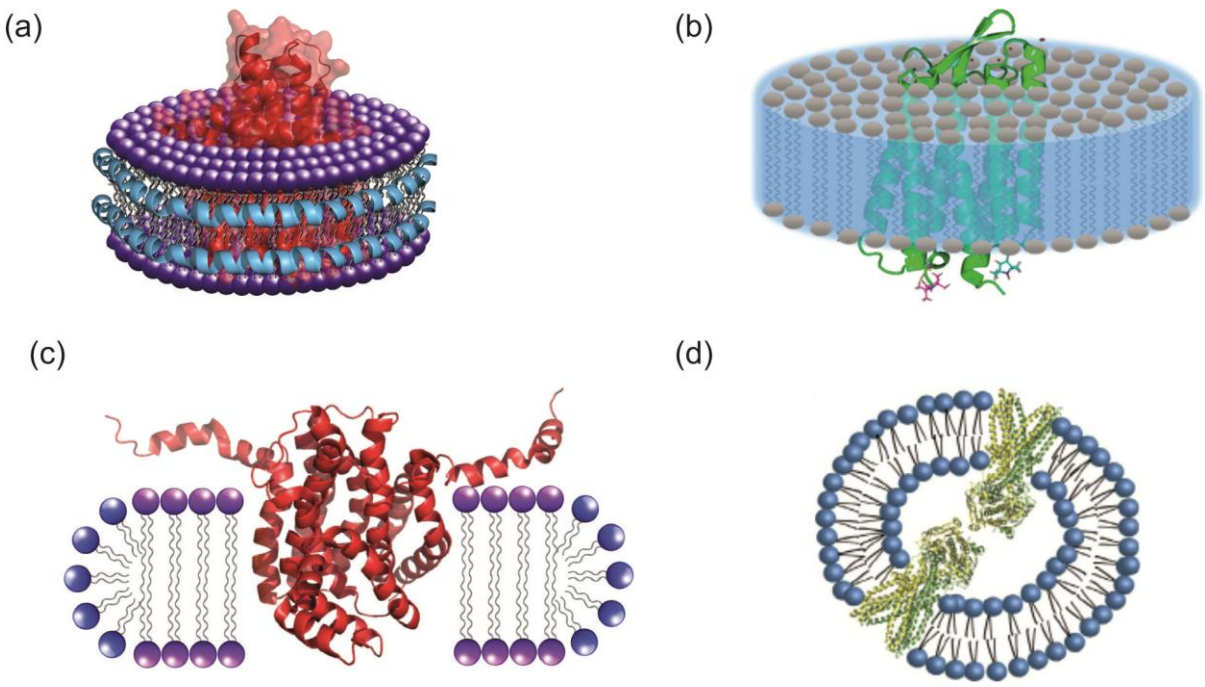


Figure 4-2: Examples of membrane proteins solubilised using lipids (a) membrane protein (red) solubilised in a nanodisc taken from Hopper et al.[11] (b) membrane protein (green) solubilised inside a lipodisq taken from Orwick-Rydmark et al.[12] (c) membrane protein (red) solubilised inside a bicelle taken from Hopper et al.[11] (d) membrane protein (green) solubilised inside a liposome taken from Claxton et al.[13]

The main benefit of these methods using lipids as the solubilisation agent for the membrane proteins is the prospect of keeping the protein in a native-like environment. Further to the initial benefits, three of these methods; nanodiscs, lipodisqs and liposomes, provide external pressure to the protein similar to a natural membrane. In the case of nanodiscs and lipodisqs, this is provided by the component that constitutes the exterior of the vehicles circumference, the membrane scaffold protein in nanodiscs and the polymer in lipodisqs[14]. The curvature of the liposomes provides the tension in for membrane proteins solubilised inside them. This is an important factor, as it has been suggested that some membrane proteins need this tension, in order to retain their functionality. In contrast, membrane proteins solubilised inside a bicelle will not experience any tension, which is a potential drawback to this solubilisation technique.

Unfortunately, apart from lipodisqs, in the aforementioned techniques, all the membrane protein comes into contact with detergent before being solubilised with lipid. This means there is potential for the detergent to cause delipidation of the membrane protein, potentially causing the loss of lipids which are structurally significant. One final pitfall is that for membrane proteins solubilised into liposomes, there is a significant chance of incorporating more than one protein in each liposome[15].

The main methodologies using artificial means to solubilise the proteins instead of lipids are; detergent solubilisation, using amphipols[16] or artificially created lipopeptides[17, 18, 19, 20] as shown in **Figure 4-3**. Detergent solubilisation works with the hydrophobic hydrocarbon tails pointing in to the hydrophobic region of the membrane protein, while the hydrophilic head groups point out into the solvent as seen in **Figure 4-3 (a)**. Amphipol solubilisation takes advantage of the fact that amphipols are designed with a hydrophobic side and a hydrophilic side[10]. This can be seen in **Figure 4-3 (b)** with the hydrophilic side, shown in black, exposed to the bulk solvent and with the hydrophobic side, shown in grey, coiling around the hydrophobic transmembrane region of the protein. Lipopeptide solubilisation combines a detergent with a small peptide to try and get close to replicating a bilayer. The hydrophobic hydrocarbon section of the lipopeptide lines up parallel to the hydrophobic region of the membrane protein, aligning similarly to way the lipid bilayer in the cell membrane does, whilst the hydrophilic alpha helical peptide is exposed to the solvent.

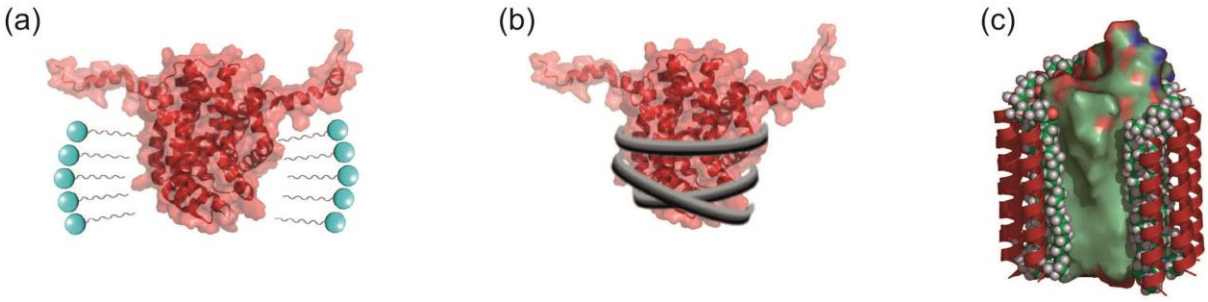


Figure 4-3: Examples of proteins solubilised without using lipids (a) membrane protein (red) solubilised inside a detergent micelle, (b) membrane protein (red) solubilised inside amphipols both taken from Hopper et al.[11], (c) membrane protein (green) solubilised inside lipopeptides taken from McGregor et al.[17]

There is a significant drawback in all of these methods, in that whilst they all attempt to mimic the native environment of the membrane protein, they are all artificial molecules. This leads to debate about whether the protein can be considered to be truly in a native-like state. The relative expense of amphipols and lipopeptides, compared with detergents, means that amphipol and lipopeptide solubilised membrane proteins are usually extracted with detergent before being exchanged into the preferred solubilisation vehicle afterwards[10, 21].

The second drawback for these methods, in common with the lipid based methods described earlier, is the issue of replicating the tension the protein experiences in the membrane. Neither detergent, nor lipopeptide solubilisation, can offer a comparable mimic for the tension experience in the cell membrane. In contrast, amphipols due to their nature coiling around the transmembrane region, they can provide tension, similar to that experienced in their native environment.

Many of these methods of solubilising membrane proteins have been used in conjunction with mass-spectrometry[11]. Detergents remain the most common method for solubilising membrane proteins however their drawbacks when it comes to mass spectrometry are well documented[22]. In addition to potential worries with regard to the preservation of crucial lipid interactions, the process of removing the micelle often requires large amounts of

energy to be added to the system, causing concern that the protein may undergo unfolding or other structural alterations before the micelle can be fully stripped away.

So far, alternative methods for solubilising membrane proteins have not managed to overcome this problem. The methods used by Hopper et al.[11] all require significantly more energy to release the protein from the vehicle they are solubilised in when compared to detergent. It is however promising that it is possible to release membrane proteins from a range of different vehicles during a mass spectrometry experiment:

4.2. Initial Experiment:

In the initial experiments conducted, I looked to use SMA-bound lipodisqs in the hope these "native nanodiscs" might allow for some insights to specific protein-lipid interactions. The logic behind this hypothesis is that with no need to use detergent at any point, no delipidation of the protein should occur.

For the use of lipodisqs with MS, there needs to be a significant advantage to their use compared to more traditional MS vehicles. For the purpose of this work, the initial aim was to achieve any of the following:

i. Protein-lipid interactions being maintained after the release of the protein from the lipodisq.
ii. Preservation of oligomeric states that are not captured during more traditional membrane mass spectrometry.
iii. A reduction in the energy required to release the protein from its vehicle.

The biggest potential benefit from using lipodisqs for MS is the ability to retain all of the lipids associated with the protein. Given the extraction process removes all the lipids in proximity to the protein at the time of extraction[23], where the lipids have a strong

interaction with the protein, it should be possible to observe if an effect of this type is present. Whilst this has the capacity for significant insights into the role lipids play, with regard to protein function, structure and oligomerisation there is an associated pitfall. Due to the heterogeneous make-up of the cell membrane there is a risk that using lipodisqs consisting of entirely native lipids will result in an overcomplicated mass spectrum, preventing analysis occurring.

In a bid to allow for easy "proof of principle" for lipodisqs as a vehicle for MS, lipodisqs were explored in which only DMPC was the lipid component. This decision was taken as making the lipid component of the lipodisq homogenous should allow for easy analysis, establishing if the protein can be released from the vehicle during the MS experiment. Additionally the low lipid-protein ratio in the purple membrane means additional lipid is required to facilitate the formation of lipodisqs.

For the initial work the membrane protein bacteriorhodopsin (BR), from *Halobacterium Salinarum*, was chosen as a test case. The choice of protein was governed again essentially on the ease of extraction and purification of the protein from the membrane. With BR being the only protein present in the purple membrane[24], this allows for simple purification, in addition to its purple appearance, which allows for simple visual tracking of whether the protein is present in the lipodisq after extraction.

4.3. Initial Results:

The initial work involved the study of BR solubilised into a DMPC lipodisq. The DMPC was doped into the lipodisq by addition of an excess of DMPC to the lipodisq formed from the initial addition of the SMA co-polymer, as described in materials and methods. Spectra were acquired in 200 mM ammonium acetate pH 8. The initial results from the first spectra were

acquired using BR in DMPC lipodisqs on a Waters Synapt HDMS, modified to use a linear field ion-mobility cell, are shown in **Figure 4-4**:

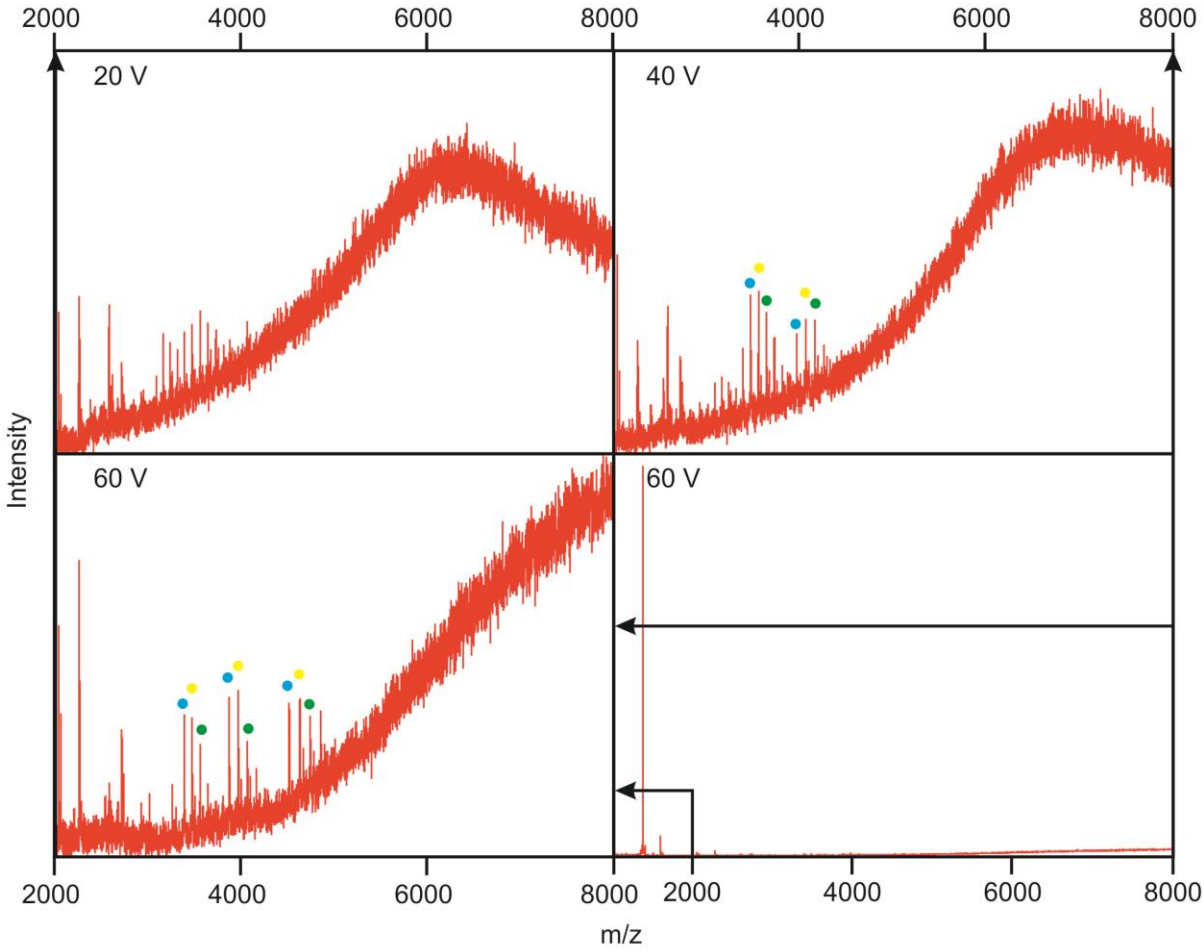


Figure 4-4: Initial spectra from BR out of DMPC lipodisqs. Spectra taken at increasing collision voltages applied in the collision cell, protein peaks can first be detected at 40V. The naked BR can be seen (blue) as well as BR with DMPC bound (yellow and green)

The initial results are promising for two reasons; the relatively low energies required to eject BR from the lipodisq and the observation of lipid bound peaks in the spectra. Whilst the initial data is promising there are also some potential problems identified simultaneously. In spite of the homogeneous make-up of the lipids inside the lipodisqs used, the initial peaks seen are not assignable, despite their obvious presence in the spectra acquired at a collision energy of 20V (**Figure 4-4 top left panel)**. Secondly the intensity of the lipids masks the signal for the protein ejected, the bottom right panel in **Figure 4-4** shows the full spectra

acquired where the signal for the protein being studied only exists in very low relative intensity peaks. Finally, given native BR is known to exist in the membrane as a trimer[25], there is a concern that this technique does not capture the trimeric form of BR.

The final issue with the oligomeric state not being correctly reflected in the MS results could be explained by the doping in of DMPC. If the endogenous lipids play a vital role in the stabilisation of the trimer, addition of an excess of DMPC could well lead to the loss of the trimeric state. Were this to be the cause then the loss of the native trimeric state, whilst still disappointing, is explicable.

4.4. Comparison with Detergent Data:

For a reasonable evaluation of the benefits of lipodisqs compared with more traditional solubilisation vehicles to be appropriate, it is important to also acquire data for the same protein using the traditional methods. Therefore BR was solubilised using octyl glucoside, a detergent commonly used for membrane protein MS, and elctrosprayed under the same conditions, and on the same modified Waters Synapt, as the lipodisq solubilised BR (**Figure 4-5)**. The main points of comparison to look for are; the ease of release of the BR from its solubilisation vehicle, protein oligomeric state, preservation of protein-lipid interactions, and, the observed charge states of the protein. Ideally for lipodisqs to be an advantageous technique they need to show at least equivalence in some of these categories to merit use over traditional detergent methods:

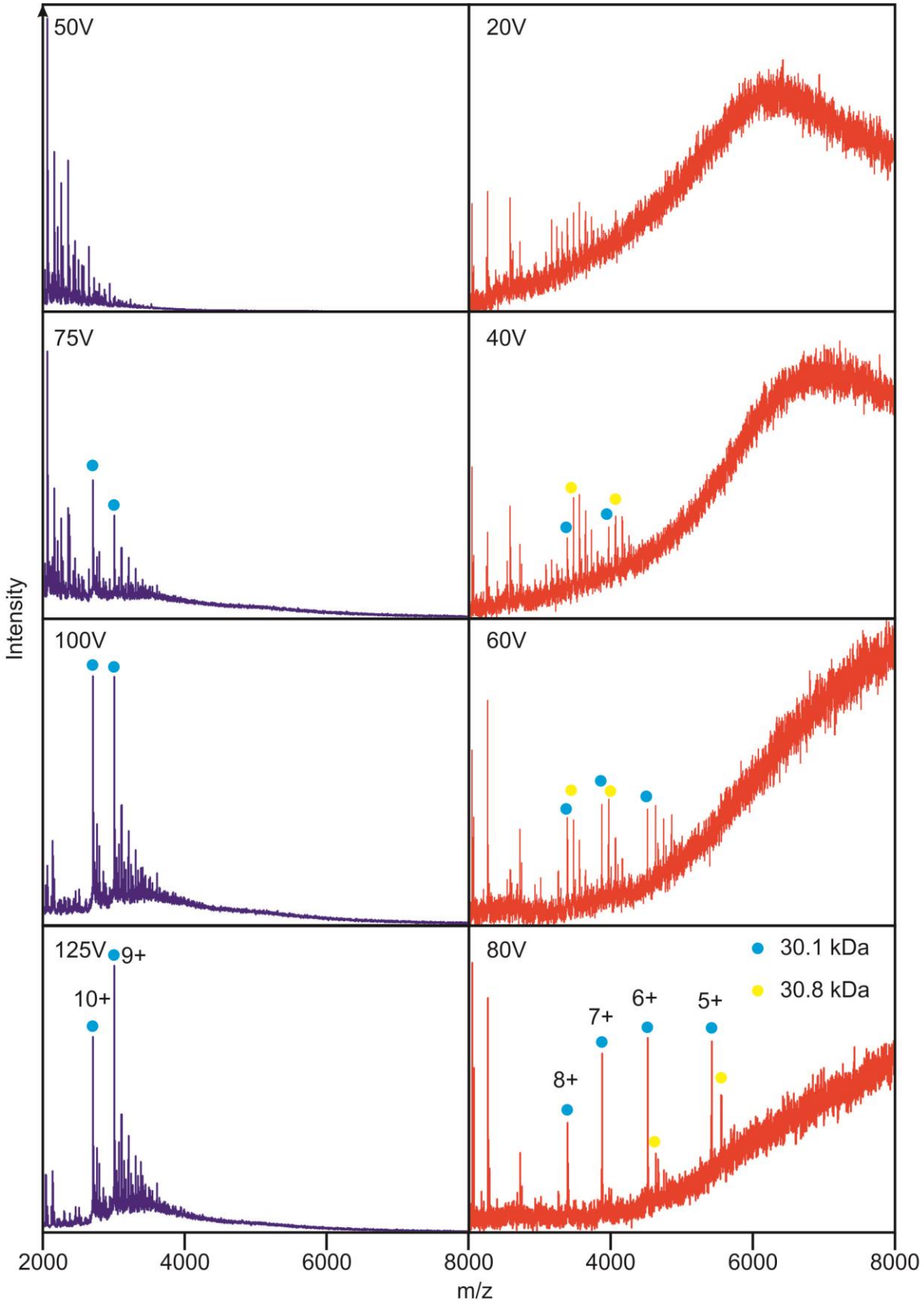


Figure 4-5: Comparison between MS of detergent (OG) solubilised BR (blue) and lipodisq solubilised BR (red). The blue dots highlight the naked BR peaks in both series of spectra, whilst the yellow peaks show the BR associated with a single DMPC lipid.

From the comparison between the techniques it can be seen that there is a twofold benefit here from using lipodisqs, in contrast to a more traditional detergent based solubilisation technique. Firstly, the collision energy required in order to release the protein from the solubilisation vehicle is much greater in the case of the detergent micelle. The first peaks associated with BR are seen in the spectra for the OG solubilised protein at a collision voltage of 75V, requiring an additional 35V more than for the lipodisq solubilised protein. The second noticeable difference comes in the observed charge states for the BR, after it has been ejected from the solubilisation vehicle. For the detergent solubilised spectra, the observed charge states are significantly higher than in the lipodisq solubilised spectra. The detergent solubilised protein has peaks from the nine plus and ten plus charge states, whereas the lipodisq solubilised protein has peaks from the five plus to eight plus charge states.

The final difference can be observed in the overall appearance of the spectra. The spectra from detergent solubilisation only have peaks corresponding to the naked protein, whilst the spectra from the lipodisqs solubilisation have peaks corresponding to binding of multiple DMPC lipids. Whilst it is unlikely to see DMPC bound to BR in its native environment, given it is only introduced to the protein in the preparation of the lipodisq sample, the lack of any peaks demonstrating lipids bound to the protein illustrate the delipidating effect of the detergent during membrane protein extraction.

These initial results show there is a lot of potential for the use of lipodisqs as a vehicle for membrane protein MS. In addition to the benefit of retaining the protein-lipid interactions, which, whilst possible in some cases during detergent solubilised membrane protein MS[6], the ability to release the protein at lower collision energies, and with lower charge states, is significant. The higher the collision energy used to release the protein using CID, the more

likely the protein may become unfolded, and less native like. Similarly, the lower the charge states for the membrane protein, the more likely it is to retain the folding, and other characteristics, associated with the protein in its native state[26]. Work has been done previously, highlighting the benefits of reducing the charge on a protein, leading to a greater ease of observation of non-covalent interactions[27]. It can be seen in **Figure 4-5** that, at a collision energy of 80V, it is only the much lower charge states of the 6+ and 5+ ions that have maintained the protein-lipid interaction, seen more extensively in the lower collision energy spectra. Given the promising nature of these initial results, it naturally follows that it is worth investigating further with additional proteins to see if the advantages shown here with BR are generally reproducible across a range of membrane proteins.

4.5. Secondary Experiments with Lipodisqs:

The Mechanosensitive channel of large conductance (MscL), from *Mycobacterium tuberculosis*, and the multidrug and toxic compound extrusion protein (MATE), from *Pyrococcus furiosus*, membrane proteins were chosen for use in confirming if membrane protein MS utilising lipodisqs is more generally applicable than simply working with BR. The selection of these proteins provided additional challenges, when compared to the somewhat simplified system from which BR is extracted. With these two proteins there is now a need to separate the protein of interest in the membrane from every other protein that is present inside the membrane.

Traditionally, this purification step is relatively simple, with the use of a recognisable tag on the protein that can be cleaved at a later point, if desired, to leave behind the protein of interest. This process has been used previously to isolate both of these proteins in detergent[28] and allowed for the acquisition of their spectra. However, there were initial problems during early attempts at the purification process for the lipodisq solubilised

proteins. After the addition of the SMA to the homogenised membranes, as detailed in chapter 5, the resulting mixture was run on a HisTrap column. The protein failed to bind to the column initially with no peak appearing in the UV trace even at high concentrations of imidazole. However, with an extended incubation period with the affinity media, the protein bound to the column, and a peak was observed upon elution. It is believed that the problem in the early attempts at purification arose due to an interaction between the co-polymer and the tag designed to bind to the column. This interaction seems to significantly slow the process of binding to the column, but does not prevent the addition of the imidazole causing elution.

Initially work was done with MscL in DMPC doped lipodisqs, and after early success, work was carried out on both MscL and MATE solubilised in native lipodisqs. Native lipodisqs are those lipodisqs which have not had their lipid composition artificially altered, after the initial extraction of the protein, from the membranes using SMA. With taking the step forward using native lipodisqs, there is a potential for great insights to be gained, about the specific interactions between lipids, and the membrane proteins embedded inside them. This could be hugely significant if it is possible to interrogate the data sufficiently to identify which lipids are involved in specific, as opposed to non-specific, non-covalent interactions.

4.5.1. DMPC Doped Lipodisqs:

The first spectrum for a protein, other than BR, was for GFP tagged MscL protein as shown in **Figure 4-6** and **Figure 4-7**. **Figure 4-6** shows the energy ramp from low energy (40V), where there are a series of features present, to higher energy (140V) where the MscL-GFP monomer can first be observed. **Figure 4-7** shows the best spectrum acquired for the MscL-GFP monomer at a slightly higher energy, where the spectrum is much cleaner and better resolved.

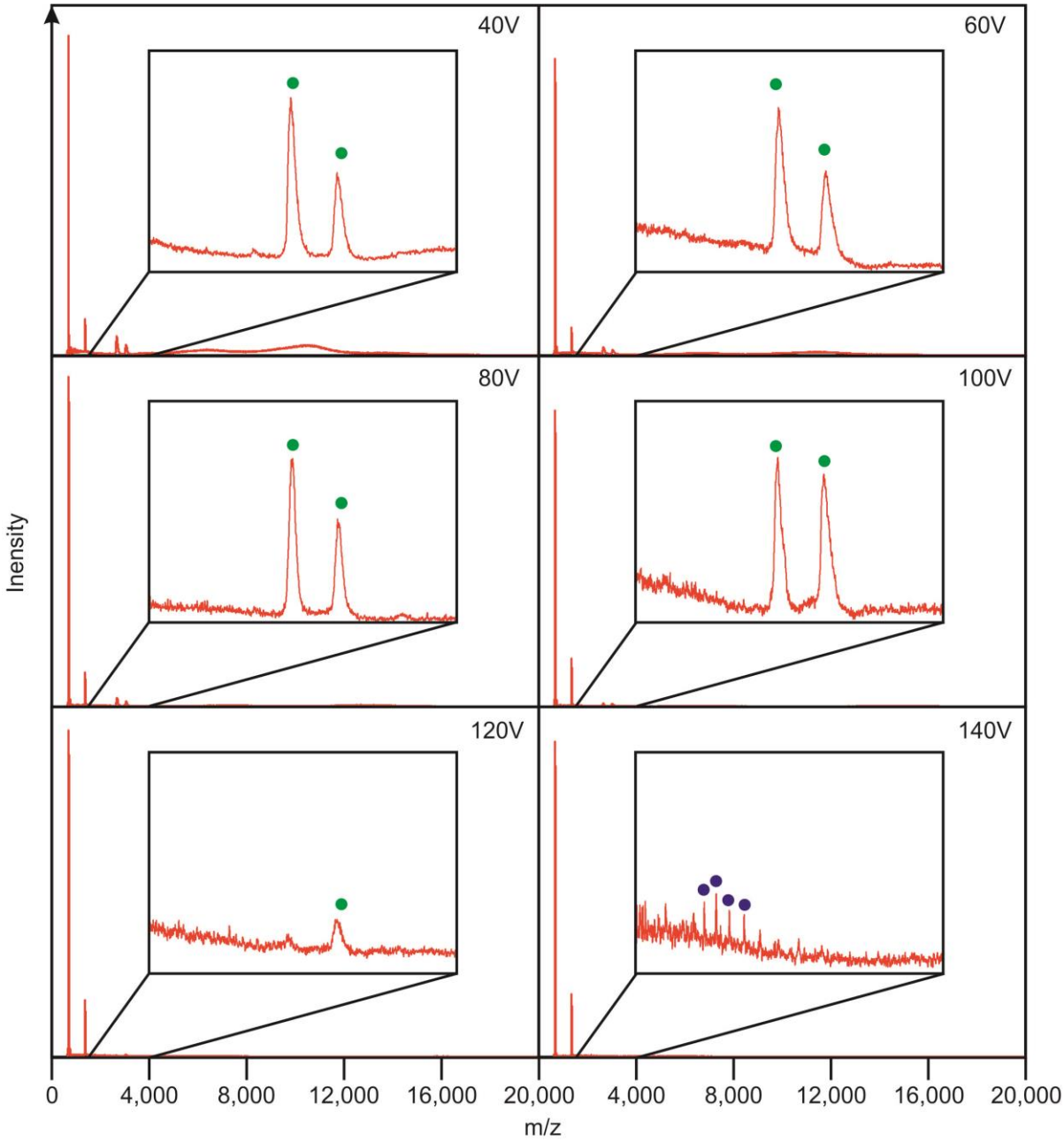


Figure 4-6: Spectra showing an energy ramp at 20V intervals from 40V to 140V for a Lipodisq of MscL-GFP with DMPC lipids doped in. The spectra inset are a zoomed in look of the 1,500 – 4,000 m/z region. With peaks corresponding to the MscL-GFP monomer marked in purple and an unassignable species thought to be a precursor marked in green.

There are a number of interesting features that appear consistently throughout most of the energy-ramp. Whilst the only peaks that were assignable to a protein were for the MscL-GFP monomer (purple dots in **Figure 4-6**), there is evidence of some form of precursor in the non-resolved peaks at 2,600 and 3,000 m/z (green dots in **Figure 4-6**). However these peaks are incredibly broad, with both peaks spanning over 100 m/z units, making the conclusive

assignment of them to any species impossible. The presence of the small humps in both the 40V and 60V spectra above is of interest. These could potentially represent the intact lipodisq before the collision energy increased to the point they are no longer observed.

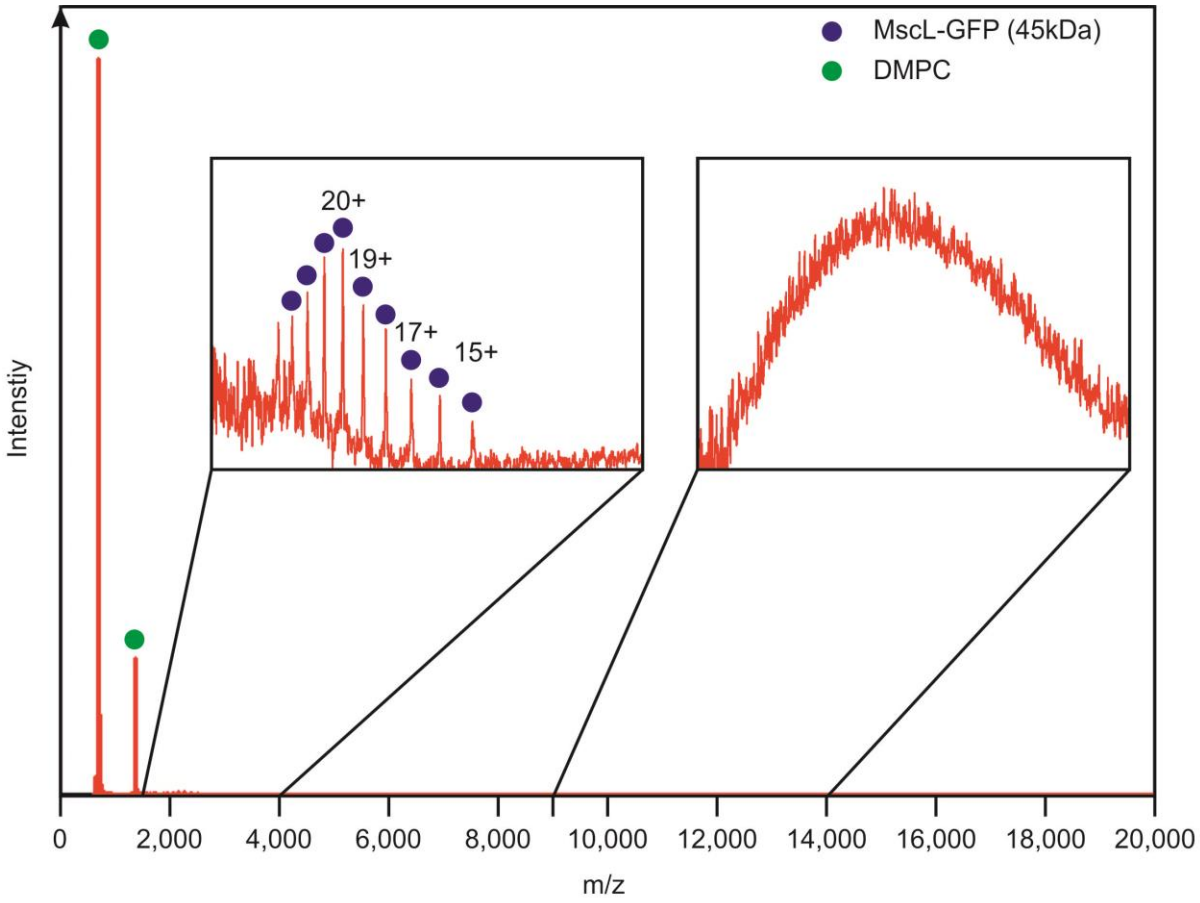


Figure 4-7: Mass Spectrum of MscL-GFP at 160V shows a monomer released from the lipodisq (purple) as well as large peaks for DMPC (green) and a large hump in the high mass region which is unassignable.

The spectrum above shows the best resolved spectrum obtained for the MscL-GFP, with nine charge states observable for the protein. It is clear that we can see the monomer here. However, there are a few drawbacks from this work already seen in three main areas; firstly, the oligomeric state observed does not correspond with its state in the membrane, secondly the lack of observable protein-lipid interactions, and finally, the intensity of the peaks corresponding to the protein.

The difference between the oligomeric state of MscL after extraction from the membrane, where it is observed as a pentamer[28], and the MscL-GFP mass spectra, where it is only observed as a monomer, could be caused by a couple of factors. Potentially the GFP-tag not being cleaved from the main body of the protein could hinder the ability of the protein to oligomerise. This however seems unlikely as work done by Reading et al.[28] has shown that in spite of the presence of a GFP-tag onto MscL it is possible to observe oligomerisation in a detergent environment suggesting the lack of observation of higher oligomers could be down to the lipodisq itself.

Alternatively, there is a reasonably high possibility that the observed peaks for MscL-GFP correspond to a highly charged monomer, which has been ejected from the main body of the protein complex. For this to be true, there has to be a parent ion, which is breaking apart into a highly charged monomer, and leaving behind a charge stripped partial parent ion.

In **Figure 4-8,** a magnified view of the lowest energy spectrum shown in **Figure 4-6**, there are two broad peaks in the spectrum, as well as the two non-resolved peaks shown in green. For the dissociation pathway suggested above, the parent ions would originate in the broad peak annotated in orange and dissociates to a charge stripped oligomer in the broad peak annotated in purple and the non-resolved peaks in green. The non-resolved peaks then release a highly charged monomer upon increased activation.

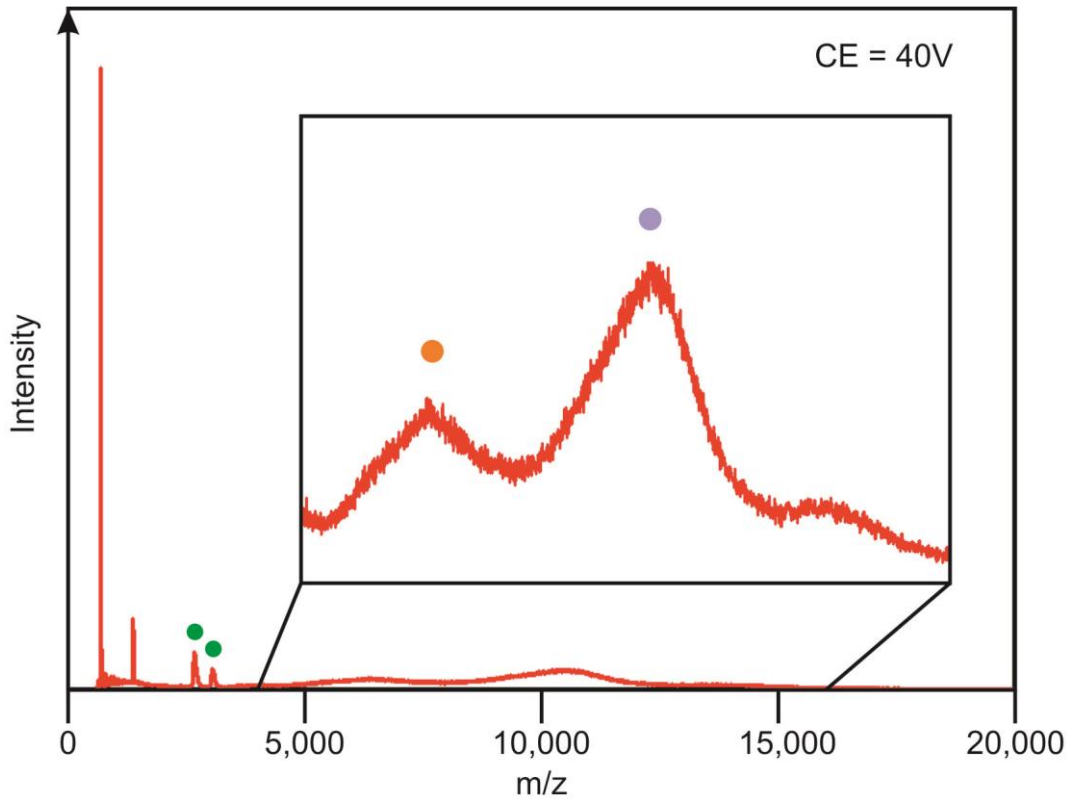


Figure 4-8: Magnified view of the spectrum of MscL-GFP in DMPC lipodisqs at 40V in the collision cell. The broad peak annotated in orange is a potential precursor, whilst the broad peak annotated in purple could correspond to a charge stripped product ion from CID.

The lack of observed peaks, corresponding to the protein bound to lipids, is a disappointing result; especially given the protein has been solubilised using lipids. However, given the high energy required to liberate the MscL-GFP monomer from the lipodisq the breaking of the protein-lipid interaction is understandable. The non-resolved peaks seen in the lower energy spectra could be caused by an ejected MscL-GFP complex with different lipids, or chain lengths of SMA bound to it.

Without further experiments it is impossible to conclude what the make-up of the observed non-resolved peaks actually is. MS-MS experiments were done in addition, selecting the non-resolved peaks marked by green dots in **Figure 4-6** and **Figure 4-8,** hoping to interrogate these peaks and determine their composition. Unfortunately the MS-MS experiments yielded no results on either of the two non-resolved peaks, with negligible total

ion counts recorded. It is believed that this is because these peaks are only observable once some energy is applied to the collision cell. For the lowest collision energy used (not shown) all that is observed is a broad peak in the higher m/z region. With the instrumentation available at the time of the work it was not possible to apply sufficient energy to the sample before selection by the quadrupole to allow for effective MS-MS of the desired peaks in this experiment.

The final issue noted during the work on the MscL-GFP construct in DMPC doped lipodisqs, was the difficulties in observing the protein peaks. In the optimised spectrum, shown in **Figure 4-7**, the peaks are still of low intensity making assignment difficult. This with low signal intensities could also be masking some of the problems in observing protein-lipid interactions. There is a possibility, that there are lipid bound peaks, which are of such low intensity they are lost amongst the noise in the baseline. In order to combat this issue many of the future experiments were also run on a modified Orbitrap instrument, where the increased sensitivity should allow for detection of any such peaks if they are present.

4.5.2. Native Lipodisqs:

Following the initial success seen in the release of the MscL-GFP protein from the DMPC doped lipodisqs the obvious next step forward was, to see if it was still possible to release a membrane protein from a native lipodisq. The native lipodisqs results proved to be significantly harder to interpret than the DMPC doped lipodisqs used initially. This does not come as a surprise, given that this will introduce a high level of heterogeneity to the composition of the lipids. Whilst software has been developed designed for the deconvolution of complex MS data[29], allowing for interpretation of the complex data, often obtained for membrane proteins still inside nanodiscs, this is not possible with lipodisqs.

This problem arises because, where nanodiscs have a protein of a known mass, acting as the belt of the disc, in lipodisqs the belt is formed by the co-polymer. This is problematic, as the co-polymer does not have a fixed mass, meaning there are now two sources of heterogeneity inside the sample, greatly increasing the difficulty of analysis.

The first experiments with native lipodisqs were done using the same MscL-GFP protein as for the DMPC doped lipodisqs. The results from the experiments with the MscL-GFP inside native lipodisqs are shown in **Figure 4-9:**

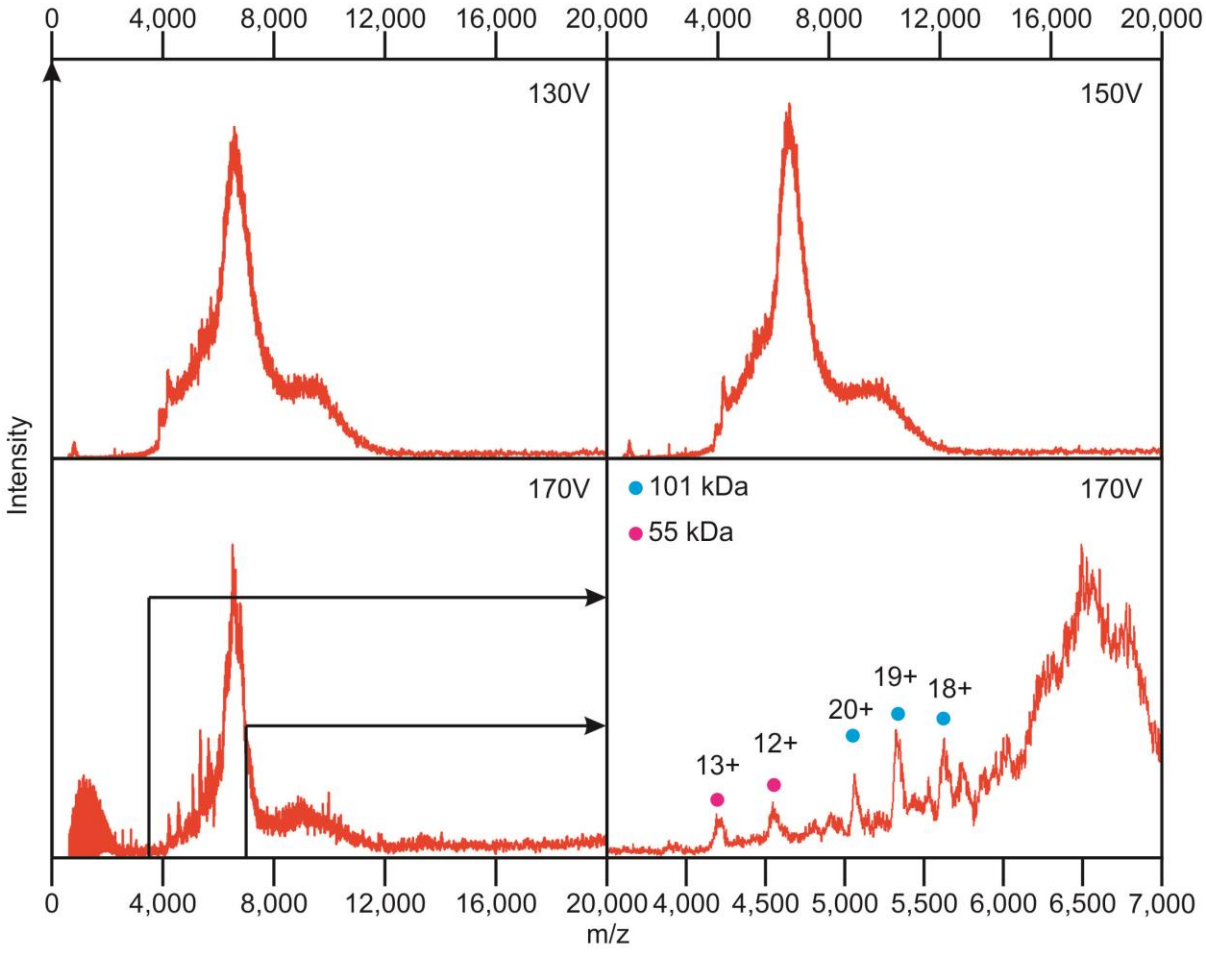


Figure 4-9: Energy ramp of MscL-GFP inside a native lipodisq. Poorly resolved peaks are observed in when the energy applied in the HCD cell of the Q-Exactive. These poorly resolved peaks can be used to determine an approximate mass for the complex corresponding to 55kDa for the peaks in pink and 101kDa for the peaks in blue.

The data acquired for MscL-GFP inside the native lipodisqs is not conclusive in determining that lipodisqs can be used universally for membrane protein MS. Whilst at higher collision energies poorly resolved peaks are eventually seen, due to their breadth, any mass assignment can only be an approximation. With these caveats in mind the two masses measured from the peaks still do not correspond to anything anticipated based on the mass of the monomeric unit of the protein.

However, one way to rationalise the data from this experiment, given the masses corresponding to the observed peaks, is the presence of some lipids or polymer to a block of protein released from the main body of the lipodisq. The width of the peaks in the spectra acquired suggests that; either the protein is particularly salty, or there is a range of lipids or some of the co-polymer with varying masses bound to the protein. The masses theoretically could correspond to monomer and dimer of the protein an additional 10 kDa of either lipids or co-polymer attached.

In addition to the work for MscL-GFP, the same experiment was performed using MATE-GFP. The results for this are shown in **Figure 4-10,** an energy ramp showing the evolution of the spectra acquired with increased energy being applied to the system, revealing a number of interesting features in the spectra:

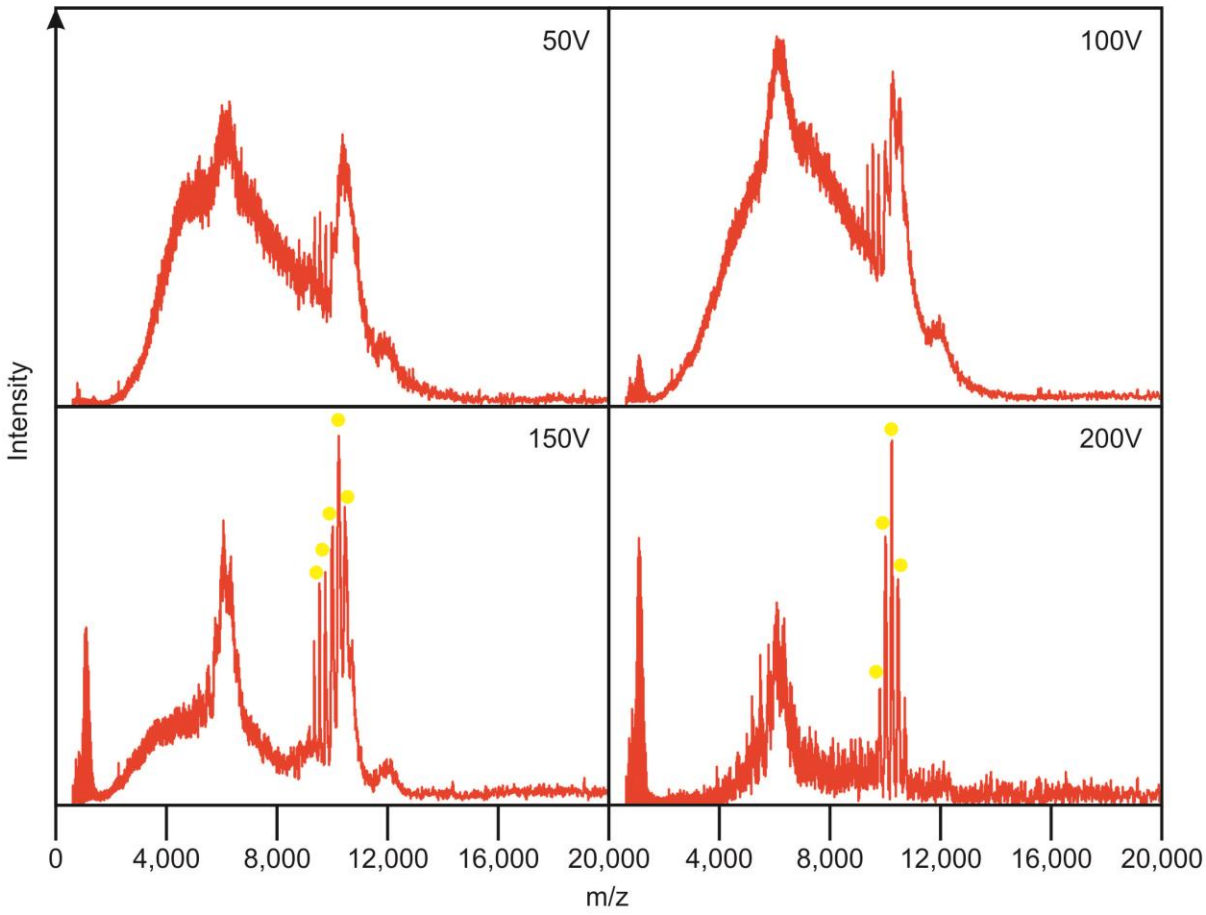


Figure 4-10: Spectra showing the energy ramp whilst trying to release the MATE-GFP protein from a native lipodisq. Eventually a clean set of peaks (yellow) is observed being released from the main body of the hump in the spectra.

There are a number of features of interest in the above series of spectra which are noteworthy. Firstly, the most obvious difference, when contrasted with spectra obtained for lipodisqs where DMPC lipids have been doped in, can be seen in the low m/z region. Typically for the DMPC doped lipodisqs, as shown in **Figure 4-7**, in the low m/z region only two peaks are observed, corresponding to one or two DMPC lipids with peaks at 678 and 1,356 respectively. Whereas, in the native lipodisqs a whole range of lipids are present, as shown by the level of population in the low mass region under 2000 m/z. Secondly, there is a series of peaks which arguably are present in all the spectra, but only become fully observed at higher collision energies. Finally, there is a low intensity series of non-resolved peaks that appear at the highest collision energy between 5,000 and 8,000 m/z. Zoomed in

views of both observed charge series at 200V are shown in **Figure 4-11** and **Figure 4-12** below:

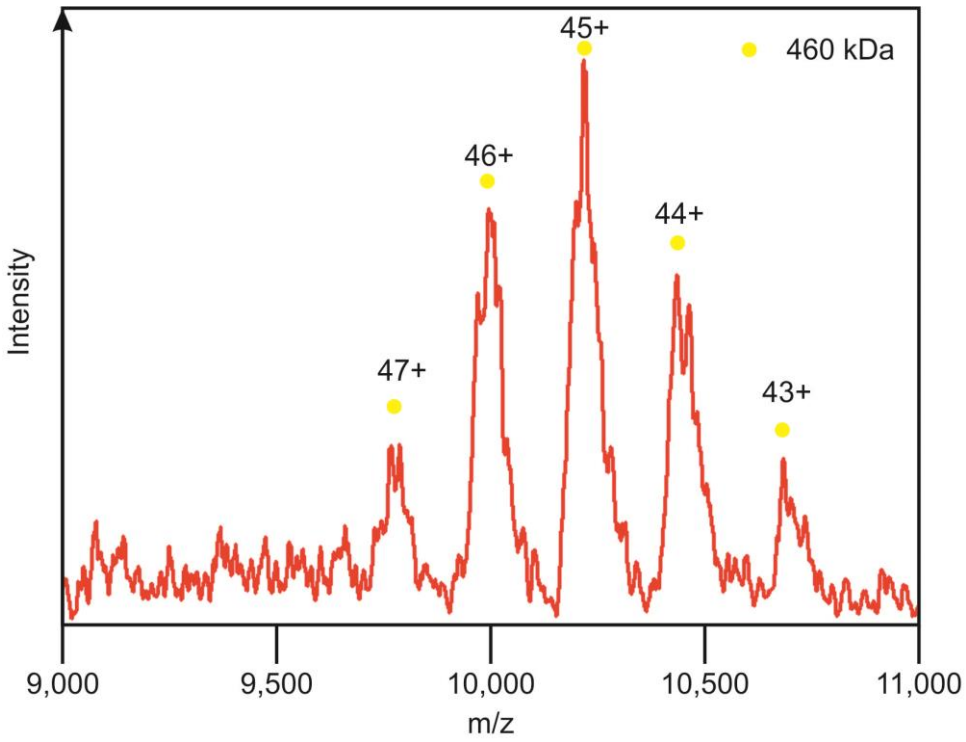


Figure 4-11: Spectrum of MATE-GFP acquired at 200V magnifying the 9,000 – 11,000 m/z region

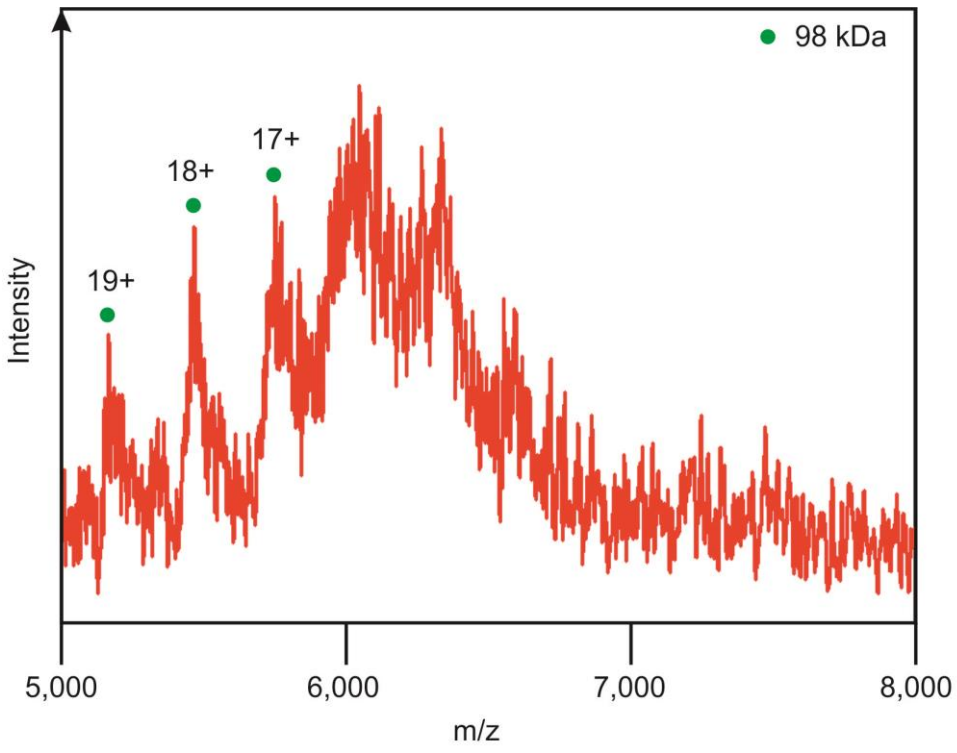


Figure 4-12: Spectrum of MATE-GFP acquired at 200V magnifying the 5,000 – 8,000 m/z region

There appears to be a common theme with the native lipodisq data acquired, in that it is very difficult to obtain clean spectra, seemingly due to the added complexity from the heterogeneous make-up of the lipid component. From the sequence of the MATE-GFP the anticipated mass is 78 kDa whilst the two masses observed in the spectra deviate significantly from this value.

Firstly, the lower mass species of around 98 kDa, annotated in green in **Figure 4-12**, could only reasonably be a monomeric unit of the protein, with some additional mass, either from; lipids or the SMA, contributing to the mass measurement. Whilst this is theoretically possible the data obtained from experiments so far prove to be inconclusive, given the low intensity and lack of resolution of the peaks seen here. The easiest way to confirm this would be through increasing the energy of the instrument until it was possible to see the naked protein once the additional species are stripped away. Unfortunately, the only spectra where these are visible are at the highest energy settings available on the instrument.

The second charge series corresponding to the higher mass species of approximately 460 kDa is also difficult to rationalise. In literature most of the evidence strongly suggests that MATE exists only as a monomer in the cell membrane[30]. The mass of a pentamer of MATE-GFP would only be equal to 390k Da, leaving it 70k Da short of the species observed in the spectra above. There is a possibility that this additional mass could be explained by a number of lipids or copolymer molecules being associated with the MATE-GFP. The lack of resolution for these peaks adds credibility to the hypothesis that they could correspond to proteins with lipids bound. However, without further experiments to corroborate, and given the lack of other evidence for MATE existing in higher oligomeric states, it is most likely this was an artificial artefact of the experiment.

4.6. Further Investigation of Bacteriorhodopsin in DMPC Lipodisqs:

The data obtained so far using lipodisqs with only native lipids, and no alteration of the lipid composition of the lipodisq, has proved to be too complex for conclusive analysis. Whilst it is often possible to speculate as to the identity of the species it is currently not been possible to conclusively prove these hypotheses. The challenges currently faced in performing MS on membrane proteins solubilised in native lipodisqs may well be solved in the future, either with developments in instrumentation, or with developments in the techniques used for analysis. However, at present the results obtained for native lipodisqs suggest that, the optimal way to use lipodisqs for membrane protein MS is by doping DMPC into the lipodisq, after the initial extraction of the protein using the co-polymer. For this reason the final data presented here is all from DMPC doped lipodisqs, where the removal of the heterogeneous lipids makes analysis possible.

Using this principle it is possible to observe similar spectra for BR when retinal has been removed after growing the protein in the presence of nicotine, and also when a deuterated sample is used.

Firstly, in the case of the BR experiment with nicotine, this is designed to show the effect of the addition of nicotine whilst growing cells. This has been shown to inhibit the production of retinal which, in turn, should allow for the observation of a reduced mass species of BR owing to the absence of retinal. When retinal is added back in to the sample this should allow for the observation of a charge series corresponding to BR as shown earlier. For comparison the spectra for BR grown in nicotine before the addition of nicotine and after the addition of nicotine are shown together:

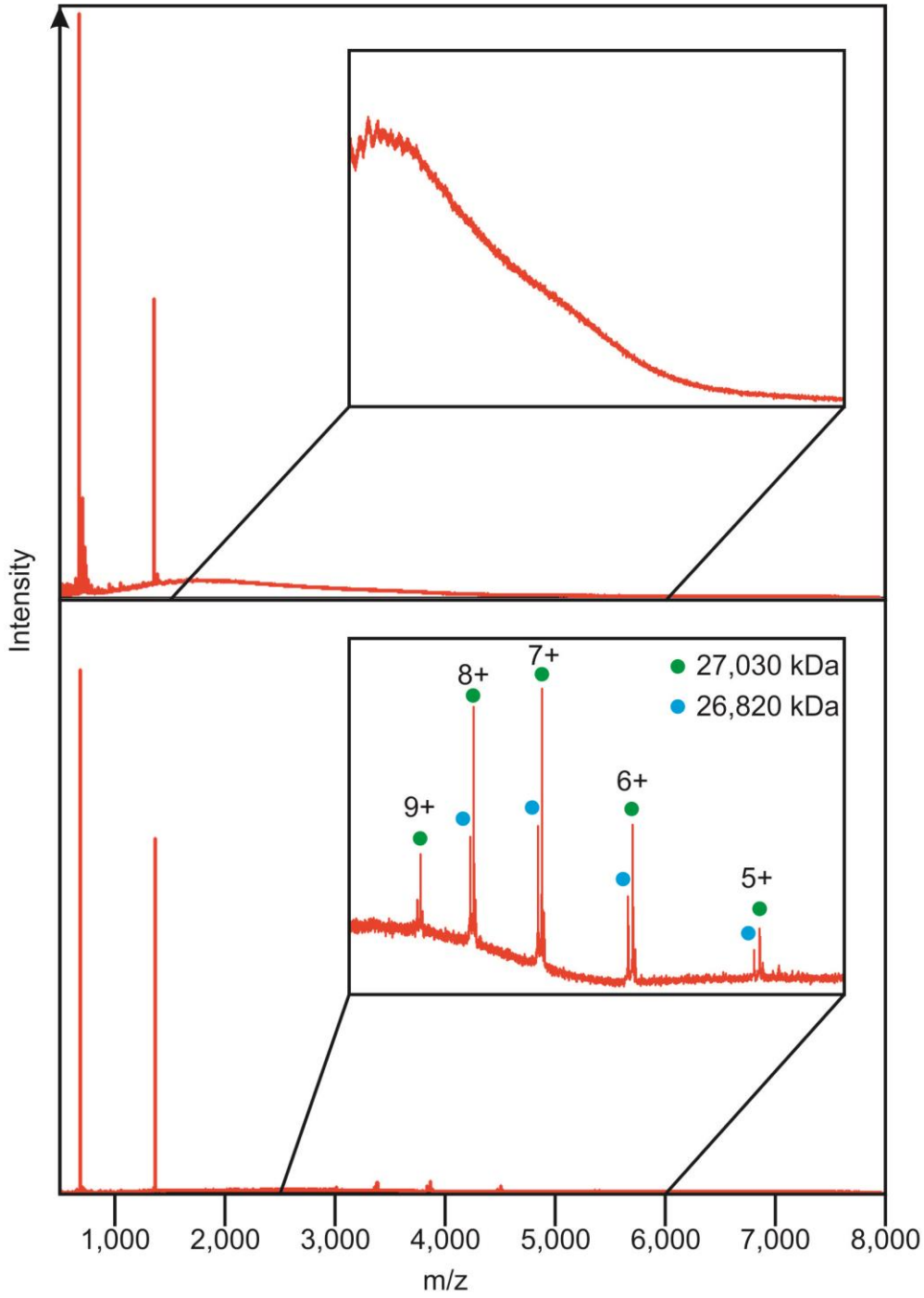


Figure 4-13: Comparison of the spectra for BR grown in nicotine (top) and then the same sample after retinal was re-added (bottom). Both spectra were acquired on a QToF Synapt at 100V collision energy.

The first thing immediately apparent in the spectra from the BR grown in the presence of nicotine is that no peaks are observed. This is in spite of the collision energy being well in excess of that usually required to liberate BR from a DMPC lipodisq, with an energy of 100V being applied in the CID cell. Spectra were acquired up to the maximum energy of the

instrument and still no peaks were observed for the BR, without the addition of retinal to the sample.

However, the most interesting data can be observed in the lower panel of **Figure 4-13.** Here, in the magnified section, not only can a charge series be observed for the conventional BR peaks, as seen previously (green dots), but a second series is also visible (blue dots). The second series of peaks correspond to the approximate mass of the BR without the retinal. The only plausible explanation for the lower mass series of peaks is the loss of the retinal with the reduction of observed mass being of the correct order to correspond to the retinal molecule.

This is particularly interesting, given that despite multiple attempts it was never possible to obtain a spectrum showing the lower mass species from the sample, without retinal added. Whilst not all the retinal has fully bound back to the protein, it would appear that it offers a stabilisation effect, which makes the release of the protein from the lipodisq possible, where it was not previously. This result suggests that the loss of the retinal has a significant structural effect on BR, particularly on the way it interacts with it lipids, causing them to become harder to remove in the sample with no retinal present.

The final experiment was conducted on a deuterated sample for BR. Whilst this sample has not undergone any alterations apart from the deuteration, the experiment was simply to determine if the protein can still be liberated from the lipodisq.

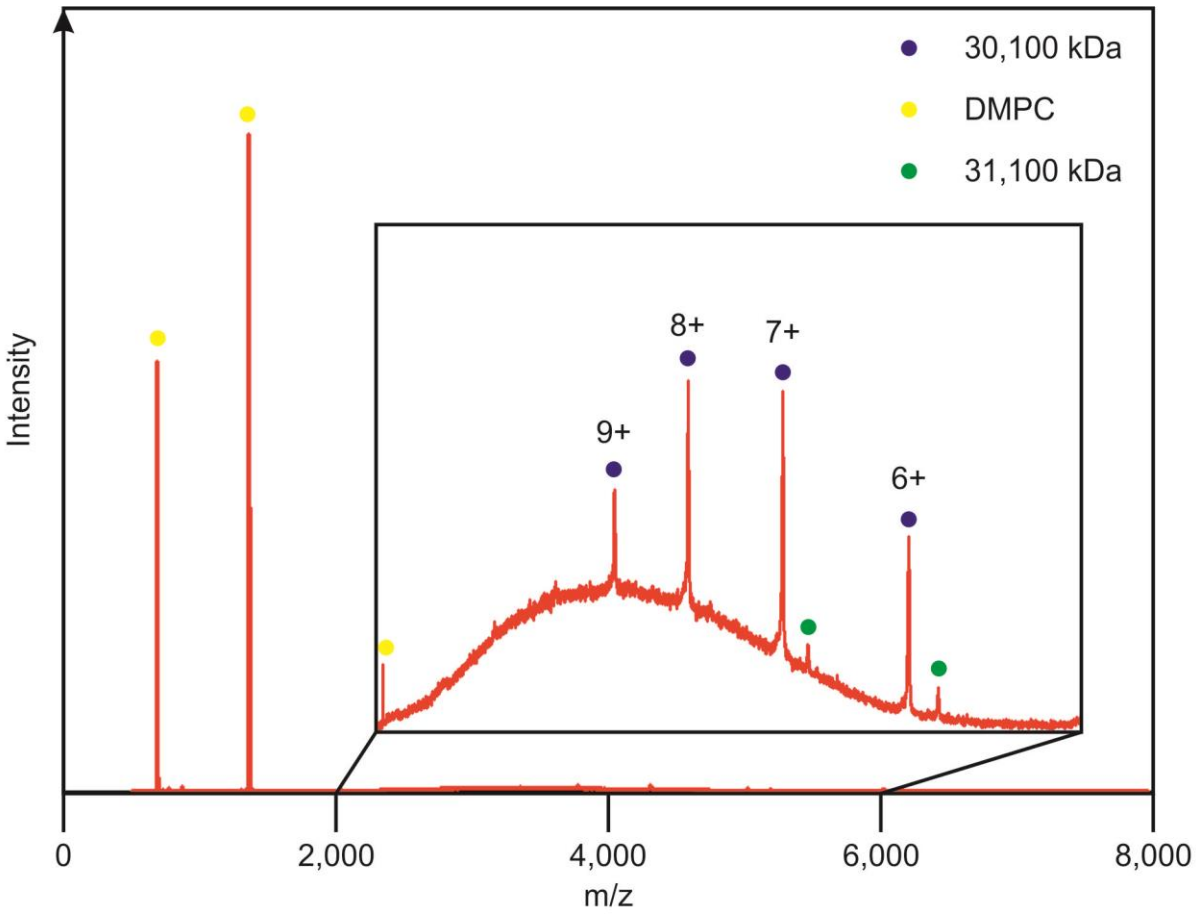


Figure 4-14: Spectrum of deuterated BR acquired at 100V on a QToF.

The spectra acquired shows, once again, the ease with which the BR can be released from the lipodisq. There are the same features in the spectra as seen with the non-deuterated samples. A charge series can be observed in purple showing the increased mass for the deuterated BR with the sample measuring in at 3,000 Daltons heavier than the non-deuterated sample. There is a significant amount of DMPC also present in the spectrum, annotated in yellow, consistent with the data collected for a normal BR sample showing small lipid clusters corresponding to 1-3 lipids.

Finally there is also an observable peak at a slightly higher m/z for both the 6+ and 7+ charge states (green dots in **Figure 4-14**). These small peaks correspond to the presence of lipid, yet to be stripped from the protein during the MS experiment.

In summary, it has been shown here that when coupled with BR, the lipodisq system is an effective vehicle for MS. In a range of conditions, and sample types, it remains possible to liberate the protein during the course of an MS experiment without the requirement of extreme collision voltages to do so. Whilst data set all relates to one protein the versatility of the technique with protein expressed in a range of conditions suggests that the technique could become more versatile, with further research and optimisation for membrane protein MS.

4.7. Conclusions:

The overall conclusion of this body of work is that the potential for MS using lipodisqs is significant; however there are still significant practical challenges to overcome in this area. Presently a gap exists between the complexities of native lipodidsq samples, and the instrumental and analytical techniques available in the field. There has been progress in the field in this area since the time this work was undertaken, with the development of techniques for the removal of the solubilisation vehicle in the source region on new instrumentation[31]. Developments such as this may allow for either, the acquisition of spectra with reduced complexity allowing for conclusive assignment of the observed peaks, or more multi-level experiments allowing for the interrogation of peaks seen during energy ramps, that were previously not selectable for MS-MS. Alternatively future breakthroughs in analytical techniques may allow for the deconvolution of spectra, allowing for some insights to be gleaned from the data currently available.

Away from the native lipodisqs, there is significant opportunity for the use of lipid-doped lipodisqs as a vehicle for membrane protein MS. Whilst in this chapter the main focus of the work was on trying to obtain data using native lipodisqs, preliminary experiments using

DMPC-doped lipodisqs showed a number of promising results. There is no doubt that the ideal result would be to analyse the proteins directly from the native lipodisqs. This would allow for easy determination of which lipids interact most strongly with the protein inside the membrane, and could provide invaluable information. Despite the setback that data acquired using native lipodisqs is currently too complex to interpret it does not mean there is still no valuable information to be obtained using lipodisqs where a lipid of choice has been doped in.

There is no reason to believe that with careful control of the doping in of the lipid to replace the bulk lipid that some information about the native lipid interactions cannot be gained. Similarly given the success already seen with BR and MscL it would appear that with a relatively homogeneous lipid composition of the lipodisq that liberation of the protein from the lipodisq is possible. Currently unpublished work from a collaboration between the Robinson group and the Watts group following on from this project has confirmed that this is possible. Data has been acquired for the membrane protein archaerhodopsin 4, and others, not only showing the release of the protein from a DMPC doped lipodisq, but also seeing both DMPC and an additional lipid bound thought to be from the cell membrane as shown below in **Figure 4-15,** acquired by Robinson group member Kin Hoi:

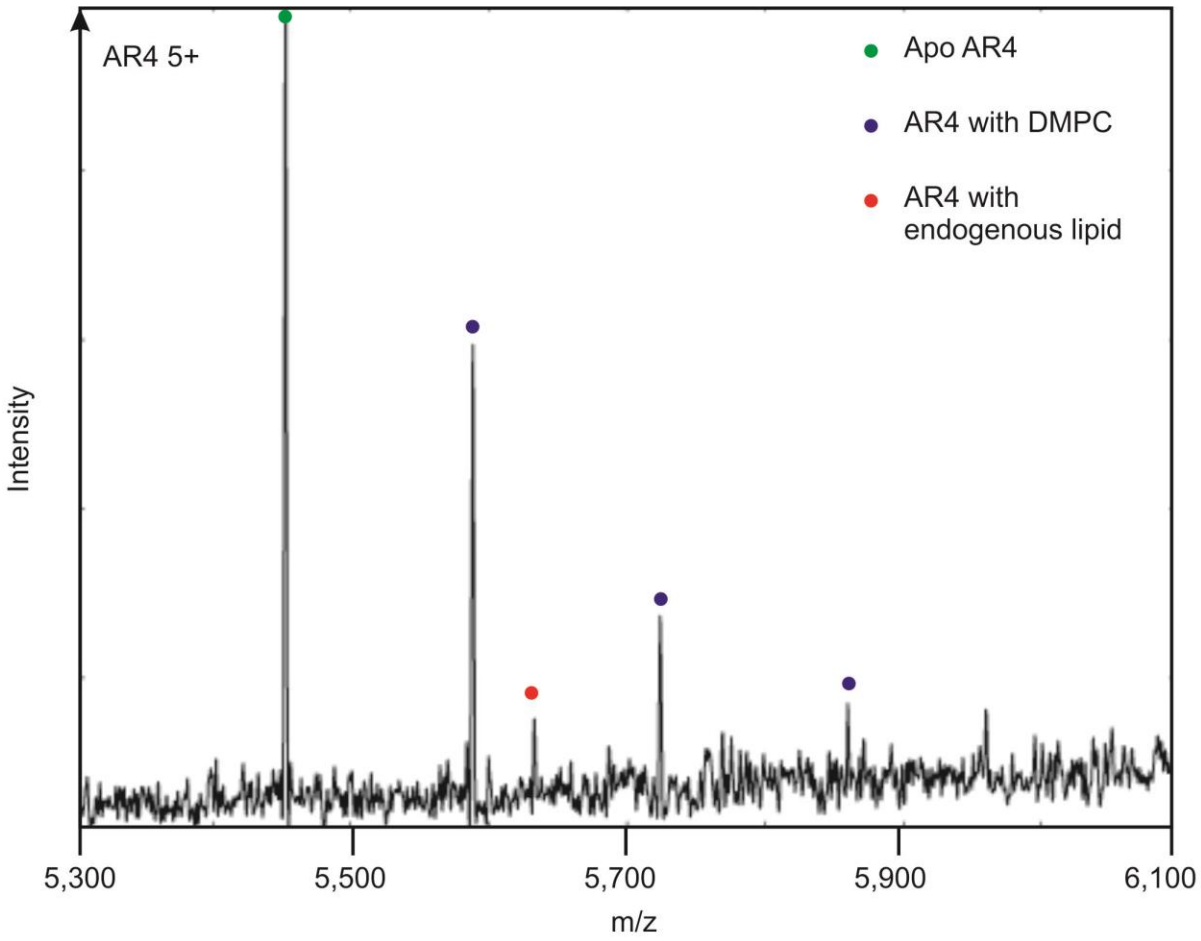


Figure 4-15: The first mass spectrum of the arachearhodopsin 4 protein 5+ charge state, developed and obtained from lipodisqs. A number of DMPC lipids bound (purple) in addition to a native lipid are observed (red). Spectrum acquired by Kin Hoi work currently unpublished.

In conclusion, lipodisqs have been demonstrated to offer a plausible alternative to more traditional solubilisation vehicles for MS. Their biggest advantage is the removal of the need for any detergent use, preventing either the delipidation of the protein, or the deformation of the proteins native structure. In addition to this the ability to observe some protein-lipid interactions makes it a very powerful technique for membrane protein analysis. Coupling this with the ability to dope out native lipids for lipids of choice might allow for a measure of the strength of different protein-lipid interactions to be assessed in the future. Overall the advantages of the technique outweigh some of the practical difficulties encountered when preparing the sample for MS inside the lipodisq and suggest an excellent way forward for studying membrane proteins in their native lipid environment.

4.8. References:

1. Bowie, J.U. Stabilizing Membrane Proteins. *Current Opinion in Structural Biology* **11**, 397-402 (2001).

2. Berman, H.M. *et al.* The Protein Data Bank. *Nucleic Acids Research*, 235-242 (2000).

3. Stansfeld, P.J. *et al.* MemProtMD: Automated Insertion of Membrane Protein Structures into Explicit Lipid Membranes. *Structure* **23**, 1350-1361 (2015).

4. Laganowsky, A., Reading, E., Hopper, J.T.S. & Robinson, C.V. Mass Spectrometry of Intact Membrane Protein Complexes. *Nature Protocols* **8**, 639-651 (2013).

5. Seddon, A.M., Curnow, P. & Booth, P.J. Membrane Proteins, Lipids and Detergents: not Just a Soap Opera. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1666**, 105-117 (2004).

6. Laganowsky, A. *et al.* Membrane Proteins Bind Lipids Selectively to Modulate their Structure and Function. *Nature* **510**, 172-+ (2014).

7. Donlan, J.A.C., Gupta, K., Struwe, W.B., Gault, J. & Robinson, C.V. Probing the Effect of Lipid Binding on the Monomer-Dimer Equilibrium of a Prokaryotic Sugar Transporter by Native Mass Spectrometry. *Biophysical Journal* **110**, 423A-423A (2016).

8. Gupta, K. & Robinson, C. A Mass Spectrometric Platform to Reveal Lipid Mediated Oligomerisation of Membrane Proteins in Atomistic Detail. *Biophysical Journal* **110**, 424A-424A (2016).

9. Borch, J. & Hamann, T. The Nanodisc: a Novel Tool for Membrane Protein Studies. *Biological Chemistry* **390**, 805-814 (2009).

10. Tribet, C., Audebert, R. & Popot, J.L. Amphipols: Polymers that Keep Membrane Proteins Soluble in Aqueous Solutions. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 15047-15050 (1996).

11. Hopper, J.T.S. *et al.* Detergent-Free Mass Spectrometry of Membrane Protein Complexes. *Nature Methods* **10**, 1206-+ (2013).

12. Orwick-Rydmark, M. *et al.* Detergent-Free Incorporation of a Seven-Transmembrane Receptor Protein into Nanosized Bilayer Lipodisq Particles for Functional and Biophysical Studies. *Nano Letters* **12**, 4687-4692 (2012).

13. Claxton, D.P., Kazmier, K., Mishra, S. & McHaourab, H.S. Chapter Twelve - Navigating Membrane Protein Structure, Dynamics, and Energy Landscapes Using Spin Labeling and EPR

Spectroscopy. In: Peter, Z.Q. & Kurt, W. (eds). *Methods in Enzymology*, vol. Volume 564. Academic Press, 2015, pp 349-387.

14. Orwick, M.C. *et al.* Detergent-Free Formation and Physicochemical Characterization of Nanosized Lipid-Polymer Complexes: Lipodisq. *Angewandte Chemie-International Edition* **51**, 4653-4657 (2012).

15. Rigaud, J.L. & Levy, D. Reconstitution of Membrane Proteins into Liposomes. *Liposomes, Pt B* **372**, 65-86 (2003).

16. Gorzelle, B.M. *et al.* Amphipols Can Support the Activity of a Membrane Enzyme. *Journal of the American Chemical Society* **124**, 11594-11595 (2002).

17. McGregor, C.L. *et al.* Lipopeptide Detergents Designed for the Structural Study of Membrane Proteins. *Nature Biotechnology* **21**, 171-176 (2003).

18. Wang, X.Q. *et al.* Solubilization and Stabilization of Isolated Photosystem I Complex with Lipopeptide Detergents. *Plos One* **8** (2013).

19. Tao, H.C. *et al.* Engineered Nanostructured beta-Sheet Peptides Protect Membrane Proteins. *Nature Methods* **10**, 759-+ (2013).

20. Prive, G.G. Detergents for the Stabilization and Crystallization of Membrane Proteins. *Methods* **41**, 388-397 (2007).

21. Prive, G.G. Lipopeptide Detergents for Membrane Protein Studies. *Current Opinion in Structural Biology* **19**, 379-385 (2009).

22. 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).

23. Dorr, J.M. *et al.* The Styrene-Maleic Acid Copolymer: a Versatile Tool in Membrane Research. *European Biophysics Journal with Biophysics Letters* **45**, 3-21 (2016).

24. Stoeckenius, W., Lozier, R.H. & Bogomolni, R.A. Bacteriorhodopsin and the Purple Membrane of Halobacteria. *Biochimica Et Biophysica Acta* **505**, 215-278 (1979).

25. Shiu, P.J., Chen, H.M. & Lee, C.K. One-Step Purification of Delipidated Bacteriorhodopsin by Aqueous-Three-Phase System from Purple Membrane of Halobacterium. *Food and Bioproducts Processing* **92**, 113-119 (2014).

26. Mehmood, S. *et al.* Charge Reduction Stabilizes Intact Membrane Protein Complexes for Mass Spectrometry. *Journal of the American Chemical Society* **136**, 17010-17012 (2014).

27. Reading, E. *et al.* The Role of the Detergent Micelle in Preserving the Structure of Membrane Proteins in the Gas Phase. *Angewandte Chemie-International Edition* **54**, 4577-4581 (2015).

28. Reading, E. *et al.* The Effect of Detergent, Temperature, and Lipid on the Oligomeric State of MscL Constructs: Insights from Mass Spectrometry. *Chemistry & Biology* **22**, 593-603 (2015).

29. Marty, M.T. *et al.* Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary Interactions to Polydisperse Ensembles. *Analytical Chemistry* **87**, 4370-4376 (2015).

30. Tanaka, Y. *et al.* Structural Basis for the Drug Extrusion Mechanism by a MATE Multidrug Transporter. *Nature* **496**, 247-+ (2013).

31. Landreh, M. *et al.* Controlling Release, Unfolding and Dissociation of Membrane Protein Complexes in the Gas Phase Through Collisional Cooling. *Chemical Communications* **51**, 15582-15584 (2015).

5. Materials and Methods:

5.1. Sample Preparation for DLS:

Proteins used in the DLS experiments, detailed in **Table 2-1** were purchased as lyophilized powders (Sigma Aldrich). Each protein was then dissolved in 200 mM ammonium acetate at concentrations between 2-10 mg/ml. Due to the relationship between intensity of light scattering and particle size the smaller proteins such as myoglobin and cytochrome C where dissolved at higher concentrations than the larger glutamate dehydrogenase and β-Galactosidase.

The protein samples were then filtered using a 4 mm 0.22 µm Millex filter (Millipore) in order to remove protein aggregates and any other undissolved impurities still present. 500µl of each protein was then injected onto a SuperDex 75 10/300 size exclusion column pre-equilibrated, using one and a half column volumes, in 200 mM ammonium acetate at 4 °C at a flow rate of 0.3 mL/min. Samples were passed into a Dawn Heleos II modified with an inboard quasi-elastic light scatterer (Wyatt Tech.) upon elution for the acquisition of the DLS data. Data was then analysed as described earlier in chapter two.

5.2. Calculating Collision Cross-Sections:

Ion-mobility data was acquired using an in house RF-confining linear field ion-mobility mass spectrometer[1] for aldolase, chymotrypsin, lysozyme, ovalbumin and β-galactosidase. Samples were prepared by dissolving the lyophilized proteins in 200 mM ammonium acetate to a final protein concentration of 10 µM. Proteins were then filtered using a 4 mm 0.22 µm Millex filter (Millipore) before being desalted using P6 Micro Bio-Spin columns (Bio-Rad), which were pre-equilibrated in 200mM ammonium acetate.

Collision cross-sections were then obtained in positive ion mode using helium as the gas in the drift cell. Drift times (t_d) were measured at six different acceleration voltages, between 75V and 200V for each protein, at a pressure of approximately 2.5 Torr in the drift cell. From these measurements the cross-section was then calculated using the gradient of a plot of t_d against 1/V as shown in *Equation 1.5*. For each protein a minimum of two replicates were obtained for the avoidance of error. All reported values area an average across each charge state for each protein.

5.3. Expression of Membrane Proteins:

The following membrane proteins were expressed during the course of this work; MscL-GFP and MATE-GFP in addition to the acquisition of BR which was expressed by a collaborator from the Watts group. The mechanosensitive channel of large conductance MscL from *Mycobacterium tuberculosis* and was fused to GFP in the C-terminus. In additionthe construct had a six histidine tag and a TEV cleave site for ease of purification as in Reading et al[2]. The MATE used was from *Pyrococcus furiosus*, and was also fused to GFP and a six histidine tag as in Reading et al[3] .

For both proteins the plasmids encoding them were then transformed into *Escherichia coli* BL21 (DE3) cells (Agilent). After the transformation the cells were grown on LB-Agar (20g Agar, 10g Sodium Chloride, 10g Peptone 140, 5g yeast extract per litre) plates in the presence of amphicilin (100mg/ml). Several colonies were then added to 50 ml of LB media containing 50mg/ml Amphicilin and grown overnight at 37°C under mild agitation. 1L of LB media containing 50mg/ml of Amphicilin was then inoculated with 7ml of the overnight culture and grown at 37°C under mild agitation until the OD_{600} reached 0.6. Cells were than induced with IPTG was then added to the media to a final concentration of 0.5mM then

grown for another 3 hours. Subsequently cells were harvested by centrifugation at 5,000g for 10 minutes at 4°C and resuspended in 20 ml of lysis buffer (300mM sodium chloride, 20mM 2-amino-2-hydroxymethyl-propane-1,3-diol (Tris), 5mM β-mercaptoethanol (BME), a complete protease inhibitor tablet (Roche), pH 7.4).

The resuspended cells were then lysed by passing them through an M-110 PS microfluidizer (Microfluidics) at 19,000 psi several times. After lysis cell debris was removed by pelleting it at 20,000g for 30 mins at 4°C. The supernatant was centrifuged at 100,000g for 2 hours at 4°C to collect the membranes. The membranes were then resusupended in ice-cold membrane resuspension buffer (100 mM sodium chloride, 20% glycerol, 5 mM BME and 20 mM Tris, pH 7.4). Membranes were homogenised using a Potter-Elvehjem Teflon pestle and glass tube ready for solubilisation.

BR was expressed by a collaborator in the Watts group as described by Oesterhelt et al[4].

5.4. Solubilisation of Membrane Proteins in Lipodisqs

Solubilisation of membrane proteins into lipodisqs was done by the addition of 3:1 SMA co-polymer (Malvern Cosmeceutics Ltd.) to the homogenised membranes[5]. For the formation of native lipodisqs 2.5% (w/v) SMA was added in a 1:1 ratio to the homogenised membranes and left to incubate for an hour. OD600 measurements were taken at 15 minute intervals to track the solubilisation.

For DMPC doped lipodisqs DMPC vesicles were formed by hydrating a lipid film in 50 mM Tris, 300 mM NaCl, at pH 8, to a final concentration of 30 mM followed by 10× freeze− thaw cycles. These were then extruded using a 400 nm filter multiple times to break up the larger vesicles. Pelleted membranes were then resuspended in the DMPC vesicles and sonicated

for 30 minutes with gentle heating. Then 2.5% (w/v) SMA was added in a 1:1 ratio to resuspended protein as for the native lipodisqs they were then left to incubate for an hour and OD600 measurements were taken at 15 minute. Then for both methods after solubilisation all insoluble particles were removed by pelleting using centrifugation at 40,000g for 30 minutes at room temperature.

5.5. Solubilisation of Membrane Proteins Using Detergent

Solubilisation of membrane proteins into detergent micelles took place by adding 1% (weight/volume) DDM overnight to the homogenised membranes while being gently agitated. Then all insoluble particles were removed by pelleting using centrifugation at 20,000g for 30 minutes at 4°C.

5.6. Purification of Membrane Proteins in Lipodisqs

All Lipodisqs containing membrane proteins were purified using SEC on a Superdex 200 10/300 column (GE Healthcare). In addition in the case of the MATE and MscL they were initially purified using a HisTrap column (GE Healthcare).

For the two Histidine-tagged proteins the lipodisqs samples were incubated overnight on a 5ml HisTrap HP column. The loaded column was then washed with 50 mM Tris, 300 mM NaCl 20 mM imidazole at 2 ml/min until a flat baseline was established allowing for the removal of all the initially unbound material. This was then followed by a stepwise addition of the same buffer with a 500mM imidazole concentration allowing the elution of the bound protein as shown in **Figure 5-1:**

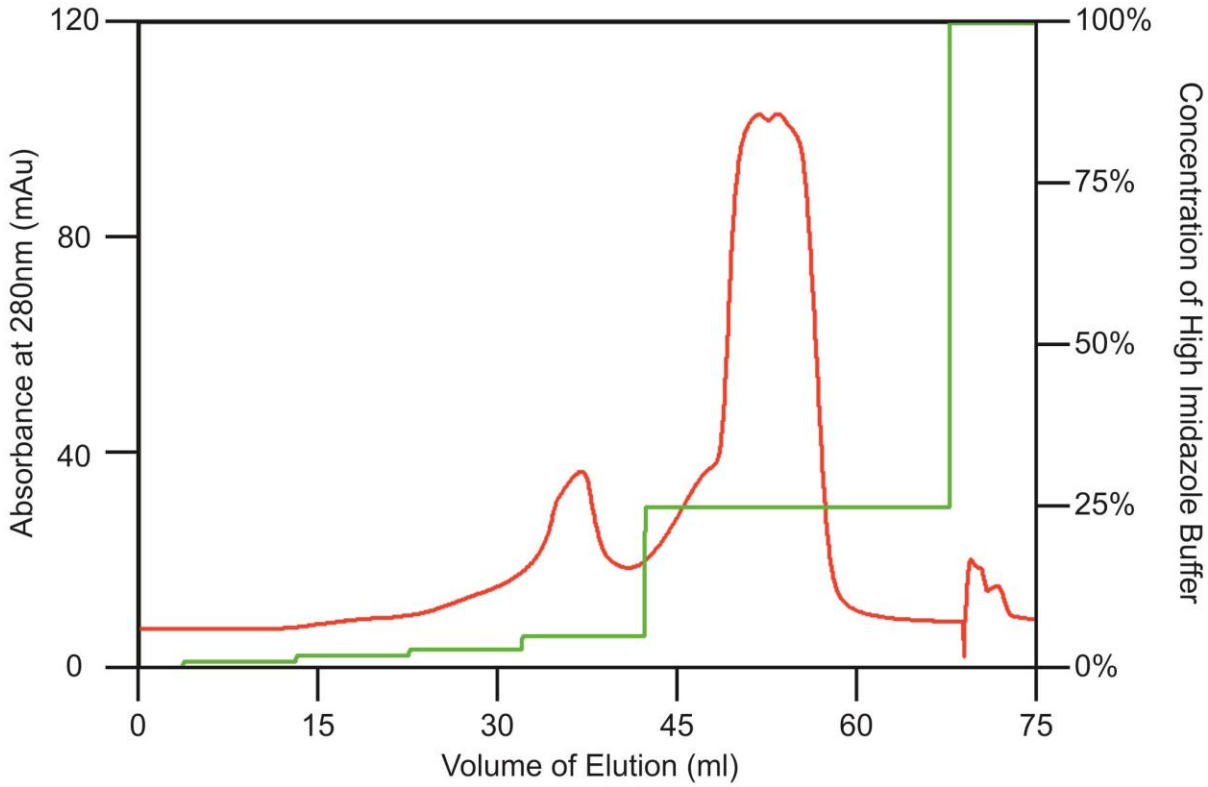


Figure 5-1: Example of a trace from a FPLC purification of a His-tagged protein (MscL) on a HisTrap column. The absorbance is shown in red whilst the concentration of the high imidazole buffer is shown in green.

For the SEC 500µl of the Lipodisq solubilised protein was injected onto the column prequilibrated in 200mM ammonium acetate pH 7.8. The sample was then eluted at 0.3 ml/min and the resulting trace collected as shown in **Figure 5-2** below:

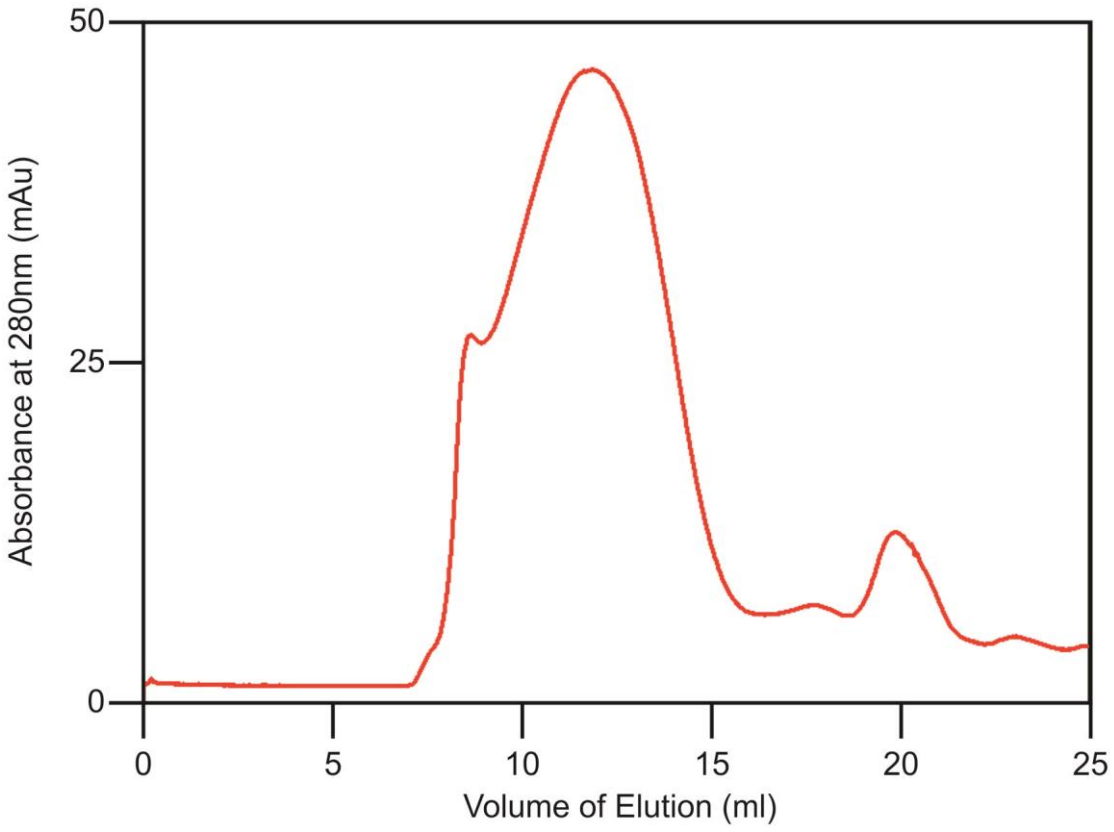


Figure 5-2: Example of a SEC trace for a lipodisq solubilised protein (BR). The absorbance is shown in red as a function of the elution volume.

The final step of the purification using SEC allows for the exchange of the sample into the ammonium acetate buffer to be used for MS. Following the SEC purification all lipodisq samples were also sized using DLS as shown in **Figure 5-3:**

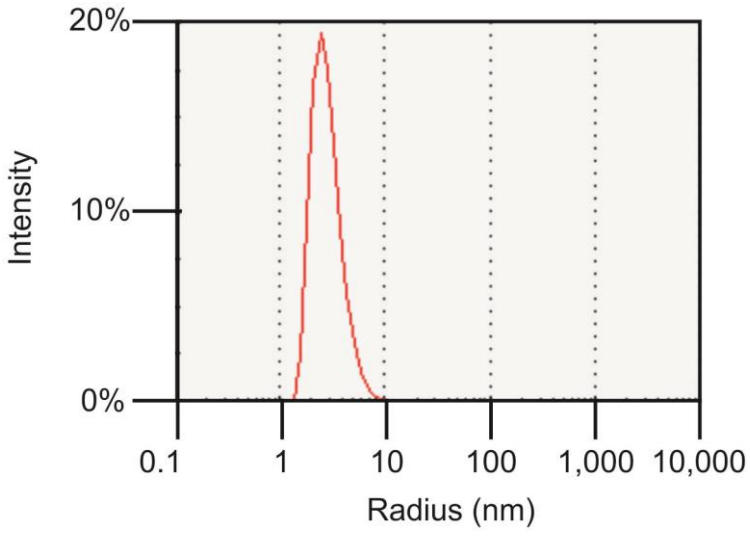


Figure 5-3: Example of DLS results for a lipodisq solubilised membrane protein (BR)

5.7. Purification of Membrane Proteins in Detergent:

Membrane proteins in detergent were purified as described by Laganowsky et al[6]. Briefly the solubilised proteins were initially purified using a 5 ml HisTrap column. The detergent solubilised protein was loaded onto the HisTrap column, pre-equilibrated in 200 mM sodium chloride, 20 mM imidazole, 0.025% DDM, and 50 mM Tris, pH 7.4, at 0.5 ml/min to allow the tag to bind. The protein was then washed with the same buffer to remove all non-bound material until a stable baseline for absorbance was obtained. Protein was then eluted over two column volumes using a linear gradient from 20mM to 500mM imidazole.

The main peak was then pooled, concentrated and then injected onto a Superdex 200 Increase 10/300, pre-equilibrated in 200 mM sodium chloride, 20 mM imidazole, 0.025% DDM, and 50 mM Tris, pH 7.4. Protein was then eluted at 0.3 ml/min as for the lipodisq solubilised proteins and the main peak collected.

5.8. References

1. Bush, M.F. *et al.* Collision Cross Sections of Proteins and Their Complexes: a Calibration Framework and Database for Gas-Phase Structural Biology. *Anal Chem* **82**, 9557-9565 (2010).

2. Reading, E. *et al.* The Effect of Detergent, Temperature, and Lipid on the Oligomeric State of MscL Constructs: Insights from Mass Spectrometry. *Chemistry & Biology* **22**, 593-603 (2015).

3. Reading, E. *et al.* The Role of the Detergent Micelle in Preserving the Structure of Membrane Proteins in the Gas Phase. *Angewandte Chemie-International Edition* **54**, 4577-4581 (2015).

4. Oesterhelt, D.S., W. Isolation of the Cell Membrane of Halobacterium Halobium and its Fractionation into Red and Purple Membrane. *Methods Enzymol*, 667-678 (1974).

5. Orwick-Rydmark, M. *et al.* Detergent-Free Incorporation of a Seven-Transmembrane Receptor Protein into Nanosized Bilayer Lipodisq Particles for Functional and Biophysical Studies. *Nano Letters* **12**, 4687-4692 (2012).

6. Laganowsky, A., Reading, E., Hopper, J.T.S. & Robinson, C.V. Mass Spectrometry of Intact Membrane Protein Complexes. *Nature Protocols* **8**, 639-651 (2013).

6. Conclusions:

The effect environment has on gas-phase protein structure is significant in a number of ways. Probing the contrast between the collision cross-sections determined experimentally in the gas-phase contrasted to those derived from solution-phase and solid-phase data yielded significant results. Results obtained using DLS in the solution-phase allowed for rationalisation of experimental collision cross-sections obtained from IM-MS.

Diffusion co-efficients from the DLS experiment compared well with ion-mobility cross-sections, after factoring in hydration affects. This demonstrates that the environment experienced by the protein in solution is the one that governs the gas-phase structure observed. This coupled with the agreement of the experimental data with the known value for the Millikan drag factor adds weight to the body of evidence showing IM-MS data provides insights into protein structure in solution.

Comparison of the computationally derived theoretical collision cross-sections from x-ray crystal structures to both DLS derived cross-section and those measured directly using IM-MS showed a surprising lack of correlation. For all the more rigorous methodologies (EHSS, TJM and PSA) the calculated values are significantly greater than those measured using IM-MS or those derived from DLS data whilst the simpler methodology (PA) calculated significantly lower values.

The possibility that the structures from x-ray do not correlate with those seen in solution was also investigated. Comparison of the collision-cross sections calculated from protein structures solved with cryo-EM and NMR as opposed to those solved by x-ray yield some results of interest. Whilst none of the three methods offer computational values that

correlate well with IM-MS data the difference in values obtained demonstrates the effect the environment has on the proteins structure. Whilst this hasn't allowed for a decisive insight into the optimal protein solution method for correlation with IM-MS data it has suggested that it is an area meriting future research.

Finally the investigation into the use of lipodisqs as method for membrane protein mass spectrometry yielded some promising results. The use of a vehicle that mimics the native environment of the protein has great potential for future insights. The holy grail for membrane protein MS is the observation of the protein in the gas-phase along with its native interaction. Whilst this work falls short of that ultimate aim it does represent a significant step forward in the journey towards it.

The use of lipodisqs for membrane protein MS allowed for the first spectra of membrane proteins which had never been exposed to a detergent environment. This is a ground-breaking development as the protein is not exposed to delipidating conditions at any point allowing for the conservation of the native environment. Whilst the acquisition of conclusive spectra was not possible for proteins inside the native lipodisqs the results demonstrate significant potential of the technique, especially in the light of new instrumental developments.

Lipodisqs with doped lipid components consistently gave promising data. BR can be liberated from the lipodisq environment repeatedly and protein-lipid interactions can be observed as well. Liberation of MscL-GFP from this environment as well shows the promise of the technique. Additionally projects in the Robinson group in collaboration with the Watts group following on from this work demonstrate the release of additional members of

the rhodopsin class of membrane proteins highlighting the potential for lipodisqs when lipid heterogeneity is reduced.

Lipodisqs represent a significant step forward in the range of techniques available for membrane protein MS. The ability for a detergent-free method from expression to analysis coupled with the option to control the make-up of the lipid component of the lipodisq adds a number of avenue for future research as this technique becomes more widely used.