Mass Spectrometric Analysis of Pilus Assembly, Amyloid Fibril Formation, and Membrane Proteins in their Native State

UNIVERSITY OF LEEDS

Aneika Corrine Leney

Submitted in accordance with the requirements for the degree of PhD

The University of Leeds

Astbury Centre for Structural Molecular Biology

October 2012

Publications

Chapter 2 includes work from the following publications:

(i) Leney A. C., Phan, G., Allen, W., Verger, D., Waksman G., Radford, S. E., Ashcroft, A. E. Second Order Rate Constants of Donor-Strand Exchange Reveal Individual Amino Acid Residues Important in Determining the Subunit Specificity of Pilus Biogenesis. *J. Am. Soc. Mass Spectrom.* 2011 (22) 1214-1223.

In this publication, all the mass spectrometry data were acquired and analysed by myself. Protein purification was equally performed by a combination of G. Phan, W. Allen, D. Verger and myself. S.E. Radford, A.E. Ashcroft and G. Waksman helped with the scientific input in the paper and provided help writing the manuscript.

(ii) Morrissey B., Leney A. C., Toste Rego, A., Phan, G., Allen, W. J., Verger, D., Waksman, G., Ashcroft, A. E., Radford, S. E. The Role of Chaperone-subunit Usher Domain Interactions in the Mechanism of Bacterial Pilus Biogenesis Revealed by ESI-MS. *Mol. Cell. Prot.* 2012 (11) M111.015289 1-10.

In this publication, the protein purification was performed by myself in combination with A. Toste Rego, G. Phan, W. J. Allen and D. Verger. All the mass spectrometry analysis was carried out by B. Morrissey. The manuscript was written by B. Morrissey with help from G. Waksman, S. E. Radford, A. E. Ashcroft and myself.

Chapter 4 includes work from the following publication:

Leney A. C., McMorran L. M., Radford S. E., Ashcroft A. E. Amphipathic Polymers Enable the Study of Functional Membrane Proteins in the Gas Phase. *Anal. Chem.* 2012 (22) 9841-9847.

In this publication, all the experimental work was performed by myself. L. M. McMorran provided the plasmids for the OmpT expression and purification and helped interpret the data. S. E. Radford and A. E. Ashcroft helped with the scientific input in the paper and provided help writing the manuscript.

Acknowledgements

Firstly, I would like to thank my supervisors Alison Ashcroft and Sheena Radford for their constant support and advice throughout my PhD studies. I would also like to thank the BBSRC for funding and Dr. R Bateman, Dr. J. Hoyes, Dr. M. Morris and Dr. K. Gilles from Waters for their advice and funding throughout my PhD studies. The pilus chapter was carried out with help from the Waksman group (Birkbeck College, London). I would also like to thank team Amyloid for useful discussions and Lindsay McMorran, Bethny Morrissey, Alice Bartlett and Gerard Huysmans for help with the membrane protein project.

I would like to thank all current and past members of the Radford lab. It is a brilliant group to be a part of and none of this work could have been done without them. In particular, I would like to thanks my office friends: Bethny for being my pilus best friend (pbf), Claire for among other things providing skittles and chocolate at crucial stages of writing, and Maya for her entertainment with jelly, fire at Astbury retreats, honest comments and elastic bands wars.

I would also like to thank the mass spectrometry lab for help and entertainment. Charlotte Scarff, in particular, has been a great help both scientifically and as a close friend. I would like to thank Lucy Woods for putting up with me as a housemate

Finally I would like to thank my family for all their support; my mum for trying to understand what I do and my dad for learning the word Biochemistry although I am still trying to persuade him that Academia is not an island somewhere near Greece. I would like to thank Richard for support and encouragement throughout my PhD especially throughout the toughest parts during my shoulder injury.

Abstract

Structural analysis of proteins and their complexes is crucial to understanding protein function. This thesis demonstrates the application of mass spectrometry (MS) to the study of pilus assembly on Gram-negative bacteria and amyloid formation in dialysis-related amyloidosis. In addition, it also tackles the challenging area of membrane protein analysis by MS.

Pili are hair-like appendages located on the outer membrane of bacteria that are involved in the transmission of infection. A periplasmic chaperone and an outer membrane usher protein coordinate pilin subunit assembly. Electrospray ionisation (ESI)-MS was used to elucidate the mechanism of subunit assembly. Experiments revealed the specific amino acids on the N-terminal extension of the pilus subunits that are important in catalysing the subunit assembly process. Further MS/MS analysis indicated differences in the stability of the chaperone-subunit-usher ternary complexes formed, providing new insights into the role of the usher in orchestrating pilus biogenesis.

Ion mobility spectrometry (IMS)-MS was next used to characterise oligomeric intermediates formed during beta-2 microglobulin (β_2m) assembly into amyloid fibrils. Analysis of the oligomers formed by a range of β_2m point mutants that affect the kinetics of amyloid fibril formation highlighted the complexity of this fibril-forming process. Further detailed characterisation of the β_2m mutant H51A revealed subtle differences in the subunit exchange dynamics of the oligomers involved in fibril formation.

Finally, this thesis shows a novel method for solubilising membrane proteins for MS analysis. Amphipathic polymers, termed amphipols, were used to fold and enhance the stability of two bacterial outer membrane proteins, OmpT and PagP. The utility of amphipols to study the structural and functional properties of membrane proteins by ESI-IMS-MS was then developed. Together these data show the power of ESI-IMS-MS in separating conformationally-distinct populations of amphipathic polymers from the amphipol-membrane complex whilst maintaining a ‘native-like’ membrane protein structure in the gas phase.

Table of Contents

Amyloid

3.1. Introduction

Membrane Proteins

4.1. Introduction

4.2. Results and Discussion

List of Figures

Introduction

Pilus Assembly

Amyloid

Membrane Proteins

List of Tables

Abbreviations

General Abbreviations

APCI	atmospheric pressure chemical ionisation
APol	amphipol
ATP	adenosine triphosphate
AUC	analytical ultra-centrifugation
BD	2, 3-butanedione
BR	bacteriorhodopsin
β_2m	beta-2 microglobulin
CCS	collision cross-section
CD	circular dichroism
CI	chemical ionisation
CID	collision induced dissociation
CMC	critical micelle concentration
CRM	charge residue model
CsI	caesium iodide
DDM	dodecylmaltoside
DEPC	diethylpyrocarbonate
DLPC	1,2-dilauroyl-sn-glycero-3-phosphocholine
DRA	dialysis related amyloidosis
DSC	donor strand complementation
DSE	donor strand exchange
ECD	electron capture dissociation
EDTA	ethylenediaminetetraacetic acid disodium salt
EI	electron impact
eIF3	eukaryotic initiation factor 3
EM	electron microscopy
EmrE	*E.coli* multidrug resistant transporter protein (EmrE)

ESI	electrospray ionisation
ETD	electron transfer dissociation
FWHM	full width at half-maximum
GSP	general secretory pathway
HDX	hydrogen deuterium exchange
HLA	human leukocyte antigen complex 1
IAPP	islet amyloid polypeptide
IEM	ion evaporation model
IMS	ion mobility spectrometry
IPTG	isopropyl β-D-1-thiogalactopyranoside
ITC	isothermal titration calorimetry
LB	Luria-Bertani
LDAO	lauryldimethylamine-oxide
LPS	lipopolysaccharide
MALDI	matrix assisted laser desorption ionisation
MCP	multi-channel plate
MHC	major histocompatibility complex
MOMP	major outer membrane protein
MS	mass spectrometry
MW	molecular weight
MWCO	molecular weight cut off
m/z	mass to charge ratio
NAPol	non-ionic amphipol
NCE	normalised collision energy
NHSA	sulfo*N*-hydroxysuccinimide acetate
NMR	nuclear magnetic resonance
Ntd	N-terminal deleted
Nte	N-terminal extension
Omp	outer membrane protein
Pap	pylenophritis associated pilus
PDB	protein data bank

PMSF	phenyl methyl sulponyl fluoride
p-NPP	*p*-nitrophenol palmitate
p-NP	*p*-nitrophenol
PSA	projection superposition approximation
QToF	quadrupole time of flight
RFU	relative fluorescence units
rpm	revolutions per minute
SEC	size exclusion chromatography
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SID	surface induced dissociation
SPR	surface plasmon resonanc
TB	terrific broth
ThT	thioflavin T
TIC	total ion chromatogram
Tof	time-of-flight
TPP^+	tetraphenylphosphonium cation
TTR	transthyretin
UsherC	C-terminal domain of the usher
UsherN	N-terminal domain of the usher
UsherP	plug domain of the usher
WT	wild-type

Symbols

Ar_m	mass of Argon
C_m	mass of complex
E	electric field strength
E_K	kinetic energy
Θ	ellipticity
D	dimer
h	hours

I	intensity
l	path length
k	rate constant
K	mobility
k_B	Boltzmann constant
K_D	dissociation constant
K_0	reduced mobility
M	monomer
M_w	molecular weight
n	charge state
Ω	collision cross-section
P	pentamer
Q	tetramer
t	time
T	trimer
t_D	drift time
V_D	velocity through the drift tube
V_0	void volume
V_t	total column volume
v/v	volume/volume
λ	wavelength
ω	frequency
w/w	weight/weight
z	charge

1. Introduction

Mass Spectrometry

1.1 Biophysical Analysis of Proteins

Proteins are essential in maintaining both the structure and function of living cells. They rarely function alone and form partners with other molecular species to fulfil their specific roles within the cell; there are a variety of such species, including small molecules, metal ions, oligonucleotides and peptides, in addition to other proteins. Biophysical techniques for the analysis of proteins have developed significantly over recent years. Spectroscopic approaches (such as circular dichroism (CD), light scattering, fluorescence), isothermal titration calorimetry (ITC), surface plasmon resonance (SPR), analytical ultracentrifugation (AUC), electron microscopy (EM), X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) are all commonly used for protein analysis; each method having its own strengths and weaknesses.

Traditional chromatography and electrophoresis based methods, such as size exclusion chromatography (SEC) and gel electrophoresis, can be used to determine the purity of the protein sample of interest in addition to the molecular weight and subunit stoichiometry of protein complexes. Careful consideration is needed due to the sensitivity of the elution and mobility times to the shape and physical properties of the protein of interest. By use of a calibration obtained with protein standards most relevant to the protein of interest low resolution molecular weight determination and stoichiometry information can be determined. AUC is another low resolution technique commonly used to determine the molecular mass, shape and equilibrium constants of proteins and their complexes, the development of which was awarded the Nobel prize in 1926 (Svedberg *et al.* 1926). AUC does not give as accurate molecular weights as other biochemical techniques such as MS, but it is still very powerful at providing sub-stoichiometric information on protein complexes.

X-ray crystallography and NMR spectroscopy are commonly used to provide high resolution structural information on proteins and their interactions with various

binding partners. X-ray crystallography provides a three-dimensional image, in some cases, at sub-atomic resolution (<1 Å) (Petrova *et al.* 2004). NMR spectroscopy, in addition to structural information, can provide dynamic information from a range of experiments including multi-dimensional and heteronuclear experiments, relaxation measurements, and hydrogen-deuterium exchange (Shortle 1996; Dempsey 2001; Dominguez *et al.* 2011; Zhao 2012). Due to the high complexity in NMR spectral analysis, the structure of proteins with molecular mass above ~50 kDa is extremely difficult to determine (Clore *et al.* 1998). A minor disadvantage of both X-ray crystallography and NMR spectroscopy is the use of high sample concentrations that they require which can lead to non-specific aggregation.

Another common biophysical technique, SPR, can be used to study non-covalent interactions in solution. SPR assumes that immobilising a molecule does not affect its interactions with other molecules in solution, and that the interaction measured reflects that in bulk solution, even though measurements are taken very close to a surface (Li *et al.* 2009).

Mass spectrometry (MS) is the technique used throughout this thesis. MS analysis has several advantages over other techniques including low sample consumption, no upper mass limit, and the ability to distinguish multiple species in solution by their unique molecular weight. MS does have several disadvantages in that samples in volatile buffers are required, the ions analysed are in the gas phase, and samples cannot be re-used (since it is a destructive technique). A brief history of MS and current MS techniques will be discussed further in this chapter followed by its application to protein structure analysis.

1.2. Mass Spectrometry

The first mass spectrometry experiment was performed by Joseph John Thomson in 1897. In this work cathode rays were deflected by an electromagnetic field in an evacuated tube which led to the discovery of the electron (Thomson 1899). Subsequent to this, Thomson and his student at the time, Francis W. Aston, refined these magnetic sector instruments to enable separation of isotopes of neon (Thomson 1911) and many other elements (Aston 1919). These experiments awarded them with the Nobel prize for physics in 1922.

Since the initial discovery of electrons by mass spectrometry, the field of mass spectrometry has been transformed more than could have been predicted. By the 1940s, industrial chemists were using commercially available mass spectrometers to quantify levels of known substances and by the 1950s these were used further for routine identification of unknown compounds (Griffiths 2008). Over the last 60 years, advances in ionisation methods and separation techniques, such as chromatography and ion mobility, that have been coupled to mass spectrometers have led to a great expansion in the application range of MS. MS has been used for applications as wide as the analysis of the chemical composition on planet Mars (Fenselau *et al.* 2003) in addition to the drug testing of athletes at the Olympic games (Hemmersbach 2008).

1.3. Overview of Mass Spectrometry Instrumentation

MS is an analytical technique for determining the molecular weight of gaseous ions based on their unique mass to charge ratios *(m/z*). There are three fundamental parts to the mass spectrometer, the ionisation source, the analyser and the detector (Figure 1.1).

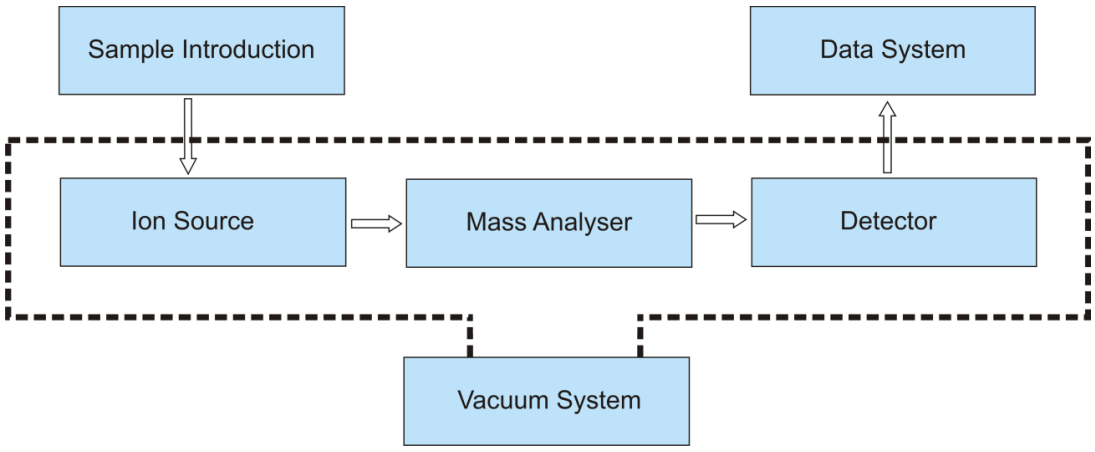


Figure 1.1 – Schematic representation of a modern mass spectrometer.

1.3.1 Ionisation

For mass spectrometry analysis, the analyte of interest must first be ionised and transmitted into the gas phase in order for it to be detected. There are many different methods of sample ionisation. Electron impact (EI) ionisation (Dempster 1918; Beynon 1956) and chemical ionisation (CI) (Munson *et al.* 1966) are suitable for small, vaporised organic compounds, whereas atmospheric pressure chemical ionisation (APCI) (Carroll *et al.* 1975) and electrospray ionisation (ESI) (Fenn *et al.* 1989) are useful for analysing samples from solution. This advance in MS ionisation methods led to a dramatic increase in the number of applications of MS to study biological systems, in particular protein analysis.

1.3.1.1 Matrix Assisted Laser Desorption/Ionisation

It was not until the 1980s that 'soft' ionisation techniques such as matrix-assisted laser desorption ionisation (MALDI) and ESI were developed (Karas *et al.* 1987; Fenn *et al.* 1989). In MALDI, the sample in solution and the matrix are often added sequentially to the MALDI plate. In the majority of cases, the matrix consists of 3,5-dimethoxy-4-hydroxycinnamic acid or α-cyano-4-hydroxycinnamic acid for protein and peptide analysis, respectively (Cohen *et al.* 1996). Typically the sample is spotted onto a MALDI plate, air-dried, and a laser fired at the analyte-matrix mixture to transfer the sample into the mass spectrometer. Upon laser ablation, protons are transferred to the analyte of interest either during or after the desorption process generating singly-charged ions (Figure 1.2A). This is disadvantageous since the organic matrix does not provide native conditions for proteins. ESI is the preferable ionisation method for protein structural analysis since, using this method of ionisation, it is possible to ionise samples from purely aqueous solution (Loo 1997; Ashcroft 2005; Benesch *et al.* 2006; Heck 2008; Bich *et al.* 2009). Electrospray ionisation is favoured over MALDI for non-covalent protein analysis since it can report on protein conformation due to multiply charged ions being produced in the ionisation process (Section 1.6, Figure 1.8).

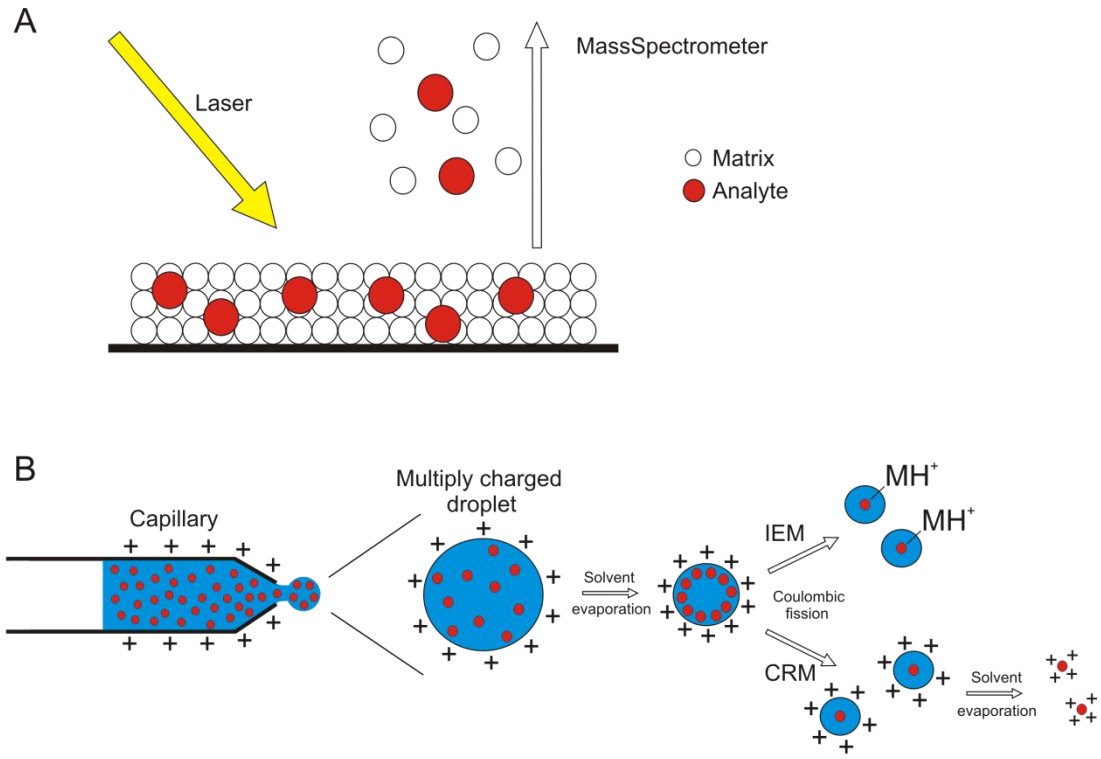


Figure 1.2 - A) Matrix assisted laser desorption/ionisation (MALDI) showing the analyte and matrix in red and white respectively. B) Electrospray ionisation (ESI) showing the solution containing the analyte of interest (red) being sprayed from an electrically charged capillary. The solvent (blue) evaporates until coulombic fission occurs resulting in positively charged ions. The differences between the ion evaporation model (IEM) (Dole *et al.* 1968) and the charge residue model (CRM) (Iribarne 1976) are highlighted.

1.3.1.2 Electrospray Ionisation

Electrospray ionisation is a gentle ionisation technique which takes place under atmospheric conditions (Fenn *et al.* 1989). During ESI, sample solution is passed through a charged capillary and drawn into a filament before it forms a mist of charged droplets around the end of the capillary (Figure 1.2B). The droplets reduce in size as they travel down a potential and pressure gradient through the inlet of the mass spectrometer towards the analyser. As the solvent evaporates from the solution, the droplets increase in charge density until the Rayleigh limit (the point at which the charge repulsion between the ions exceeds the surface tension) is reached. At this point either the solvent evaporates and Coulombic fission occurs until the droplets are of a certain radius to enter the mass spectrometer (ion evaporation model) (Dole *et al.* 1968) or the evaporation/fission cycles continue until all the solvent has evaporated and only the charged analyte molecule remains (charge residue model) (Figure 1.2b) (Iribarne 1976).

ESI can be used under aqueous conditions; however, the extra energy needed for ion desolvation from aqueous solvent (compared with volatile organic solvents) can damage native interactions. Nano-ESI, a low flow-rate modification of ESI, produces smaller initial droplets than ESI (diameter of 0.15 μm compared with 1.5 μm) (Wilm *et al.* 1996), reducing the energy needed for desolvation and fission events (Juraschek *et al.* 1999). Nano-ESI is, therefore, widely used in analysis of non-covalent native protein complexes (Sharon *et al.* 2007), and is the ionisation method used in this thesis.

There are two commonly used nanoESI methods for introducing samples into the mass spectrometer; coated capillaries and the automated NanoMate system. The NanoMate Triversa developed by Advion, uses a microfluidic chip consisting of an array of nanoESI nozzles. A voltage is applied to the surface of the chip while the sample of interest is sprayed through the nanoESI nozzle. The NanoMate offers a very efficient and stable spray due to the precise field strength created by the nanoESI nozzles. The NanoMate Triversa has been used to analyse a variety of non-covalent biological complexes including the interactions of peptide ligands with the vascular endothelial growth factor protein (Dyachenko *et al.* 2010), the interaction of tRNA with a tRNA-modifying enzyme, tRNA-guanine transglycosylase (Ritschel *et al.* 2009), and the self-assembly of the protein, beta-2 microglobulin during amyloid fibril formation (Smith *et al.* 2010; Smith *et al.* 2011).

Borosilicate glass capillaries are also commonly used to analyse large non-covalent complexes (Kirshenbaum *et al.* 2010). The capillaries used to spray the biological complexes used in this thesis were pulled in-house and coated in a gold/platinum mixture. Due to the ability to optimise the capillary diameter at the end of the electrospray tip, the coated capillaries produce mass spectra with increased signal to noise compared with the nanomate source. Poor reproducibility makes the nanomate system advantageous when performing multiple repetitive experiments.

1.3.2 Analyser

Once the sample has been ionised, the analytes must be separated according to their unique *m/z* ratios. A variety of different mass analysers are available. The first mass

spectrometers used magnets to deflect ions according to their mass to charge ratio. Since then, other analysers such as quadrupole, ion trap, time-of-flight (TOF) and ion cyclotron resonance instruments have been developed to achieve ion separation. Mass analysers aim to maximise resolution, mass range and mass accuracy.

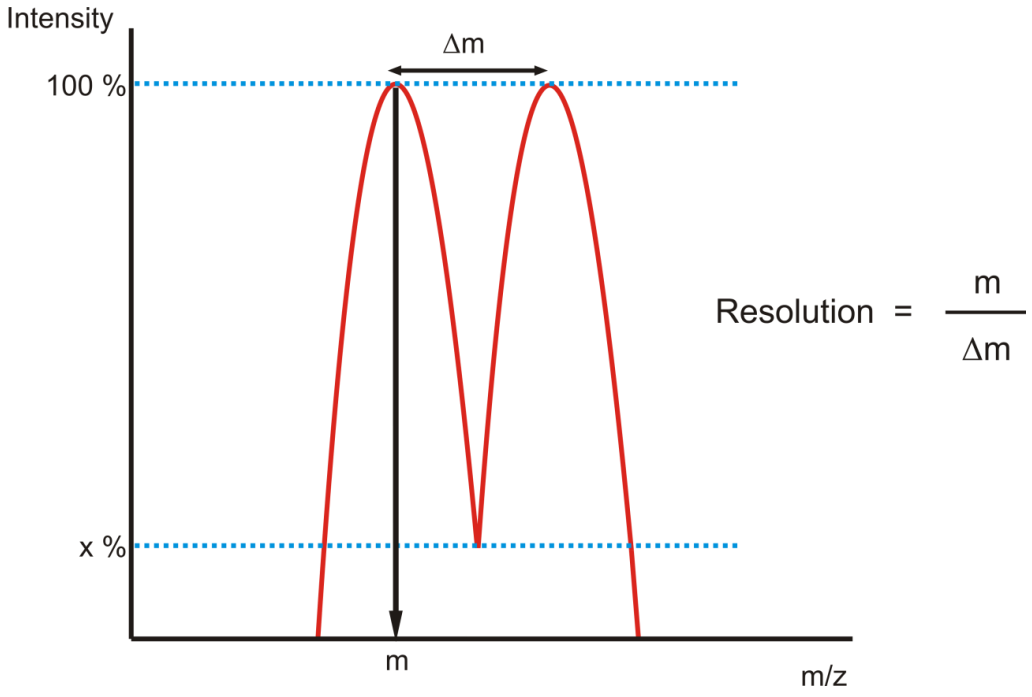


Figure 1.3 – Definition of resolution between two peaks of mass (m) and (m+Δm) in a mass spectrum. Peaks are resolved if the valley between them is at x % of the smallest peak intensity.

In mass spectrometry, resolution is defined as the smallest mass distance between two fully resolved peaks. Peak are considered resolved if the valley between them is at x %, where x is usually 10 or 50 %. The equation for resolution is shown in Figure 1.3. Resolution can also be defined by the peak width at half height for a single peak, termed the full width at half-maximum (FWHM). Resolution in mass spectrometry is highly dependent on the type of mass analyser used. The mass spectrometers used for Aston's early work had a resolution of 130 Da (Aston 1919), whereas the more recent Fourier transform instruments described by Marshall and co-workers had a resolution of 8, 000, 000 Da (Shi *et al.* 1998). Quadrupole and time-of-flight analysers were used in this thesis and are explained in more detail below.

Quadrupole Mass Analysers

Quadrupole mass analysers detect ions of each *m/z* sequentially. They consist of four rods arranged in pairs, parallel to one another (Figure 1.4). A direct current potential, U, and a radio-frequency oscillating potential, V, are applied to the rods. The potential on the rods (Φ_0) is described in Equations 1.1 and 1.2 where U is the magnitude of the direct potential, V is the amplitude of the oscillating potential, ω is the frequency, and t is time.

$$\Phi_0 = + (U - V \cos \omega t) \qquad \text{Equation 1.1}$$

$$- \Phi_0 = - (U - V \cos \omega t) \qquad \text{Equation 1.2}$$

The ions traversing the quadrupole will have a particular trajectory depending on their *m/z*. Due to the potentials applied to the rods, the ions will oscillate in the x and y direction until they reach the detector. If the trajectory of an ion through the quadrupole is stable, then the ion will reach the detector. Alternatively, if the potentials applied cause an unstable ion trajectory, then the ion will hit the walls of the quadrupoles, thus not being detected (de Hoffmann 2002).

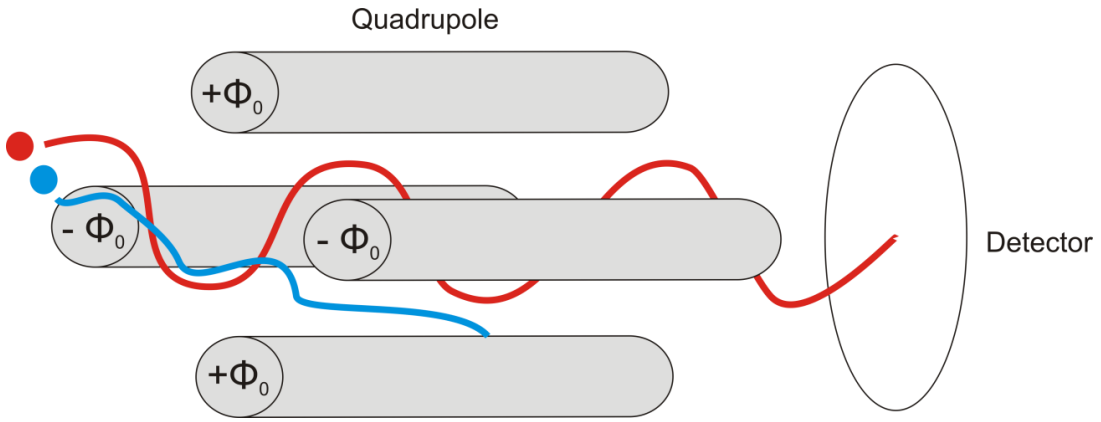


Figure 1.4 - Schematic of a quadrupole analyser showing an ion with a stable trajectory path (red) being transmitted to the detector whilst another ion with a different *m/z* ratio (blue) hits the walls of the quadrupole and is thus not detected.

Time-of-Flight (Tof) Mass Analysers

Time-of-flight (Tof) analysers were also used for the work in this thesis. Tof analysers have the advantage of having no theoretical upper mass limit. In Tof analysers, ions exiting the ionisation source are accelerated initially by a potential (V), giving each analyte ion a fixed amount of kinetic energy (E_K). Ions with the same kinetic energy then traverse a field-free region of fixed distance (d). The velocity (v) an ion has in this field-free region is inversely proportional to the square root of its mass (m) (Equation 1.3) (de Hoffmann 2002). For a fixed kinetic energy, the analyte ions velocity is determined by its mass and total charge (ze), where z is the number of charges on the analyte ion and e the charge on an electron (Equation 1.3).

$$v = \sqrt{\frac{2E_K}{m}} = \frac{\sqrt{2zeV}}{m} \qquad \text{Equation 1.3}$$

Since the time (t) an ion takes to traverse the field-free region is proportional to the length of the flight tube (d) and inversely proportional to its velocity (v), the time for ions to reach the detector can be related to the ion's *m/z* according to the following equation (Equation 1.4).

$$t = \sqrt{\frac{m}{2zeV}} \times d \qquad \text{Equation 1.4}$$

Unfortunately ions of the same *m/z* can have different initial kinetic energies causing them to reach the detector at different times. This can lead to peak broadening and decreased resolution (de Hoffmann 2002). Delayed pulse extraction was a technique developed to reduce the spread of kinetic energies between ions of the same *m/z* and is now used in all commercial Tof instruments to improve resolution (Vestal *et al.*

1995). This extraction pulse provides ions that have moved a small distance into the flight tube (i.e. whose kinetic energy is too low) with more kinetic energy thus allowing both ions to hit the detector simultaneously. Another method developed to increase the mass resolution in Tof instruments was the addition of a reflectron (Mamyrin *et al.* 1973) (Figure 1.5). Reflectrons consist of a ring of stacked electrodes, creating a retarding field, which acts as an ion mirror. Reflectrons compensate for differences in kinetic energies of analyte ions since ions with more kinetic energy will traverse further into the reflectron and as a result take longer to be reflected back than ions with less kinetic energy. The mass resolution is also increased by addition of the reflectron by effectively doubling the length of the flight tube. Another major advantage of the reflectron is that the instrument dimensions and the vacuum system can remain constant while the advantages of doubling the flight path in terms of resolution are realised.

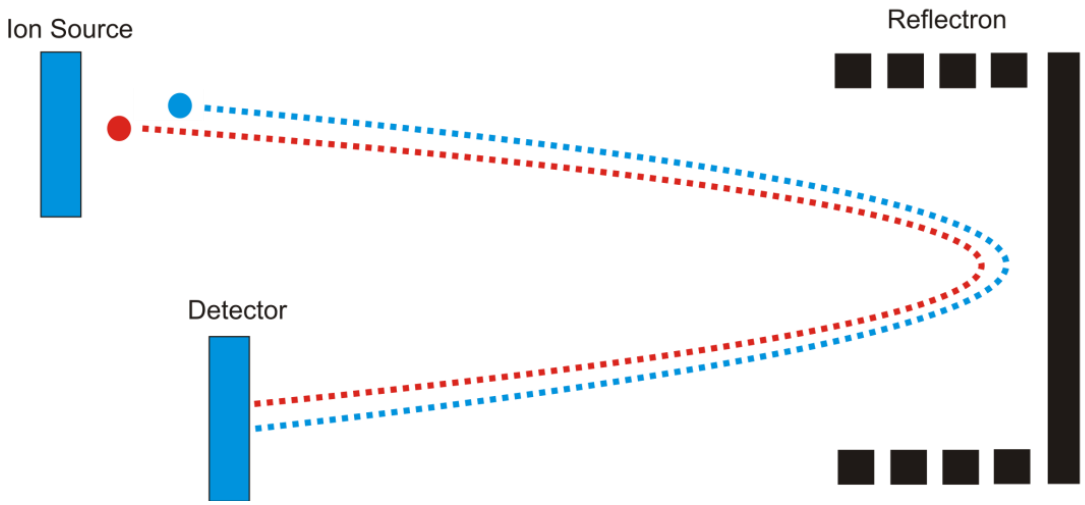


Figure 1.5 - Schematic of a Tof analyser showing two ions of same *m/z* but different kinetic energies (blue and red) being deflected to different extents by the reflectron and thus reaching the detector simultaneously.

1.3.3 Tandem Mass Spectrometry

Tandem mass spectrometry (MS/MS) involves the combined use of two or more mass analysers. MS/MS can be carried out on tandem instruments whereby two of the same analyser (for example, tandem quadrupole) or of two different analysers (hybrid) (for example, quadrupole-time-of-flight (Q-Tof)) are separated by a collision cell in which ions can be fragmented. Q-Tof instruments have been used throughout this thesis not only to characterise non-covalent complexes but also to perform structural characterisation of proteins through tandem mass spectrometry experiments. During MS/MS, a single precursor ion is selected in the first mass analyser (i.e. the quadrupole in the case of a Q-Tof), subjected to fragmentation, and the subsequent product ions detected in the second (Tof) mass analyser.

Collision-induced dissociation (CID) is the most commonly used fragmentation method. During CID, an ion of interest is subjected to collisions with an inert buffer gas (for example, argon) resulting in its fragmentation. In peptides, this results in predominantly but not exclusively b and y fragment ions (Roepstorff *et al.* 1984; Biemann 1992) which can be used to sequence the peptide of interest (Figure 1.6). Although CID is a relatively simple, straight forward fragmentation technique, the energy required to fragment a peptide bond between different amino acid residues varies and acidic residues are cleaved preferentially (Summerfield *et al.* 1997). This can limit full sequence coverage by CID and has led to development of the more novel fragmentation methods, electron capture dissociation (ECD) and electron transfer dissociation (ETD). ECD was first described by the McLafferty group and is based on ion-electron reactions in which the positive analyte ion captures electrons resulting in its fragmentation (Zubarev *et al.* 1998). In peptides this leads to the cleavage between the α-carbon and nitrogen atoms yielding predominantly c and z fragment ions (Figure 1.6). ECD generally allows full sequencing of peptides; however the production of high populations of electrons needed to induce fragmentation in Q-Tof instruments is challenging. ETD is comparable to ECD but instead of electrons, uses a small singly charged molecule, for example anthracene, to transfer electrons to the peptide precursor ion (Syka *et al.* 2004). Thus ETD avoids problems associated with high density electron production.

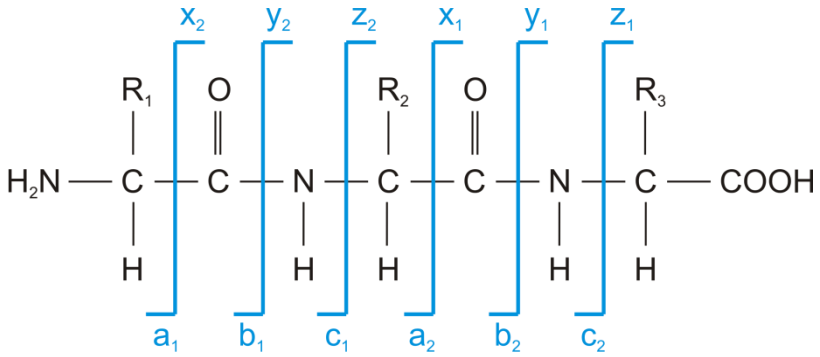


Figure 1.6 - Peptide ion nomenclature where a and x, b and y, and c and z represent fragments ions (Roepstorff *et al.* 1984; Biemann 1992). CID and ECD/ETD produce b and y, and c and z ions respectively.

CID, ECD and ETD have been used to reveal structural details of proteins and their complexes (Griffiths *et al.* 2001; Sobott *et al.* 2002; McCammon *et al.* 2005; Bakhtiar *et al.* 2006; Wu *et al.* 2007; Esteban *et al.* 2008; Gordiyenko *et al.* 2010; Morrissey *et al.* 2012) (see Section 1.9, Figure 1.10).

1.3.4 Detection

Ions exiting mass analysers are typically detected by electron multipliers which amplify the charge induced or the current produced when an ion hits its surface (Koppenaal *et al.* 2005). Tof instruments typically use multi-channel plates (MCP) of high resistive material that allow simultaneous detection of multiple ions (Koppenaal *et al.* 2005). By correlating the detection of ions with time in a Tof instrument, information in terms of a mass-to-charge (*m/z*) *versus* intensity plot can be obtained. Tof instruments are very sensitive as they have the ability to monitor all ions produced continuously (magnetic and quadrupole instruments can only detect one *m/z* at any one time i.e. they 'scan' the *m/z* range).

Over the last five years, a commercial instrument in which an additional separation technique called ion mobility spectrometry (IMS) has been inserted has aided the analysis of complex mixtures by MS and the structural analysis of proteins. The

development of IMS will be discussed in more detail below and examples of its application in protein structure analysis discussed further in Section 1.6.3.

1.4 Ion Mobility Spectrometry

Ion mobility spectrometry (IMS) can be used as an additional tool together with MS to separate ions based on their shape and/or charge (McDaniel 1973; Karasek 1974). During IMS, gas-phase ions are separated based on their mobility through a drift tube containing a buffer gas (Karasek 1974). Larger ions will take longer to traverse through the drift tube compared to smaller ions as they will experience more collisions with the buffer gas than smaller ions. Due to the electric field applied across the drift tube, differently charged ions of the same shape can also be separated with more highly-charged ions taking less time to traverse the drift tube than ions with less charge. Rotationally-averaged collision cross-section measurements can be calculated based on drift time measurements from IMS data. Using conventional IMS coupled with MS, collision cross-section measurements can be calculated directly from an ion's mobility (K) (Revercomb *et al.* 1975). Under these conditions, the mobility of an ion is equal to its velocity through the drift tube (V_D) divided by the electric field strength (E) (Equation 1.5) (Creaser *et al.* 2004). This can also be shown in terms of drift time (t_D), whereby the velocity through the drift tube (V_D) is related to the time it takes for a gas phase ion to travel the length (L) of the drift tube (Equation 1.6).

$$K = \frac{V_D}{E} \qquad \text{Equation 1.5}$$

$$K = \frac{L}{t_D E} \qquad \text{Equation 1.6}$$

As the ion's mobility through the drift cell depends on the buffer gas density, the mobility of an ion is normally reported in terms of its reduced mobility (K_0) (Equation 1.7), where T is the temperature in Kelvin and P the pressure in Torr.

$$K_0 = \frac{L}{t_D E} \times \frac{273}{T} \times \frac{P}{760} \qquad \text{Equation 1.7}$$

With conventional IMS, the mobility of an ion can be related to its collision cross-section (Ω) by the following equation (Equation 1.8), where z is the number of charges on the ion, e is the charge on an electron, k_B is the Boltzmann constant, and N the buffer gas number density.

$$\mathrm{K} = \frac{(18\pi)^{\frac{1}{2}}}{16} \frac{\mathrm{ze}}{(\mathrm{k_B T})^{\frac{1}{2}}} \left[\frac{1}{\mathrm{m_I}} + \frac{1}{\mathrm{m_N}}\right]^{\frac{1}{2}} \frac{1}{\mathrm{N}} \frac{1}{\Omega} \qquad \text{Equation 1.8}$$

The term $[1/m_I + 1/m_N]$ refers to the reduced mass of the buffer gas (m_N) and the analyte ion (m_I). Rearranging equation 1.8 for collision cross-section and substituting in reduced mobility (K_0) for mobility (K) from Equation 1.7 gives

$$\Omega = \frac{(18\pi)^{\frac{1}{2}}}{16} \frac{\mathrm{ze}}{(\mathrm{k_B T})^{\frac{1}{2}}} \left[\frac{1}{\mathrm{m_I}} + \frac{1}{\mathrm{m_N}}\right]^{\frac{1}{2}} \frac{760}{\mathrm{P}} \frac{\mathrm{T}}{273} \frac{1}{\mathrm{N}} \frac{\mathrm{t_D E}}{L} \qquad \text{Equation 1.9}$$

where the collision cross-section is proportional to the drift time (t_D) of an analyte ion of molecular mass (m_I) and charge z.

These home built, IMS instruments can suffer from poor sensitivity due to radial losses during IMS separation and duty cycle limitations. As such, detecting low-level, low-energy analytes such as macromolecular non-covalent protein complexes is challenging. The launch of the first commercially available travelling wave IMS-MS instrument, the Synapt HDMS, in 2007 (Pringle *et al.* 2007) was accompanied with a dramatic increase in the number of structural studies on proteins and macromolecular complexes by IMS-MS.

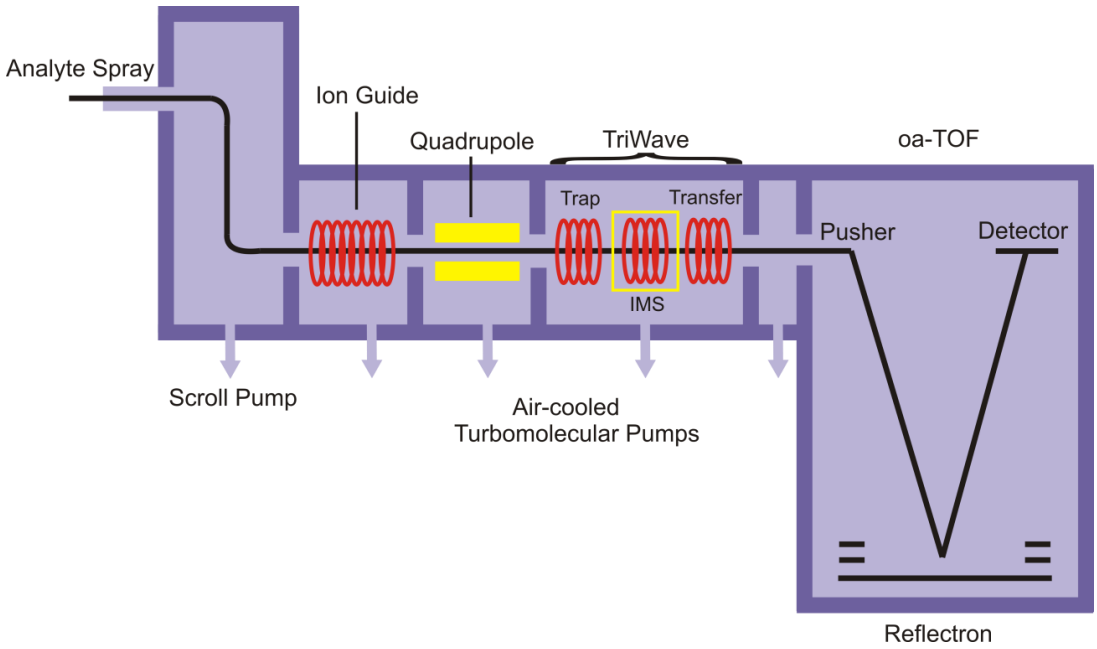


Figure 1.7 – Schematic representation of the Synapt HDMS instrument highlighting the quadrupole (used for isolating ions in CID-MS experiments) and the ion mobility cell (IMS). The Figure was adapted from Pringle *et al.* (Pringle *et al.* 2007).

In contrast to drift-cell ion mobility in which a constant low electric field is applied to the mobility cell (described above), travelling wave ion mobility uses a travelling wave of electric potential created by a series of stacked ring ion guides throughout the drift tube (Giles *et al.* 2004). The travelling wave ion guide can be used to separate ions of different mobility since the ability of the analyte ion to surf along the wave of electric potential through an inert gas depends on its mobility. Higher mobility ions are carried by the wave and reach the detector faster than lower mobility ions that roll over the top of the wave, thus taking longer to reach the detector (Figure 1.8) (Giles *et al.* 2004).

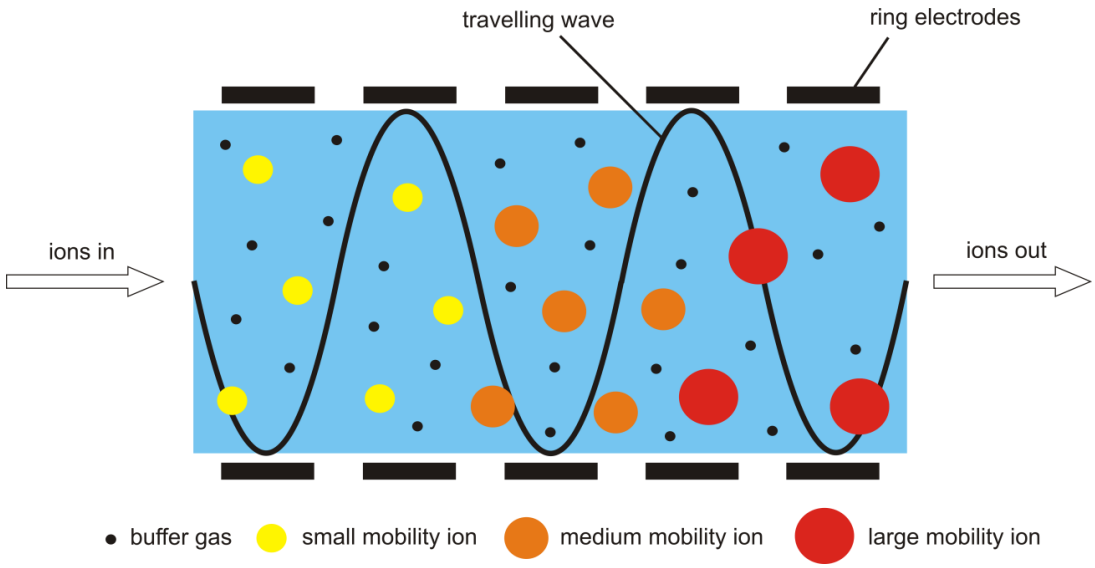


Figure 1.8 – Schematic of travelling wave ion mobility separation. Ions with a higher mobility (red) exit the drift cell before ions of lower mobility (yellow).

The travelling wave device in commercial instruments complicates the relationship between the collision cross-section and the drift time of the analyte through the drift cell. The cross-sections in travelling wave ion mobility, therefore, are calculated relative to the drift time of ions of known collision cross-section (i.e. that have been measured previously by use of conventional IMS-MS). The equation for the collision cross-section (Equation 1.9) can be adapted to contain constants, A and B, to account for the travelling wave characteristics.

$$\Omega = \frac{(18\pi)^{\frac{1}{2}}}{16}\frac{ze}{(k_B T)^{\frac{1}{2}}}\left[\frac{1}{m_I}+\frac{1}{m_N}\right]^{\frac{1}{2}}\frac{760}{P}\frac{T}{273}\frac{1}{N}At_D{}^{B} \qquad \text{Equation 1.10}$$

where A corrects for the electric field parameters (E and L in conventional IMS, Equation 1.9) and B for the non-linear effect of the travelling wave device (Wildgoose 2006).

The collision cross-section is often quoted as a charge and mass independent quantity in $Å^2Da^{-1}C^{-1}$ termed Ω' (Equation 1.11). Equation 1.11 can be further simplified to equation 1.12 by combining most of the parameters into a single constant A', where A' incorporates the correction for temperature, pressure and the electric field parameters.

$$\Omega' = \frac{(18\pi)^{\frac{1}{2}}}{16} \frac{1}{(k_B T)^{\frac{1}{2}}} \frac{760}{P} \frac{T}{273} \frac{1}{N} A t_D{}^B \qquad \text{Equation 1.11}$$

$$\Omega' = A' t_D{}^B \qquad \text{Equation 1.12}$$

Altogether, based on Equations 1.10, 1.11 and 1.12, the collision cross-section can be expressed as follows (Equation 1.13):

$$\Omega = ze \left[\frac{1}{m_I} + \frac{1}{m_N}\right]^{\frac{1}{2}} A' t_D{}^B \qquad \text{Equation 1.13}$$

$$= ze \left[\frac{1}{m_I} + \frac{1}{m_N}\right]^{\frac{1}{2}} \Omega'$$

A universal cross-section calibration method for travelling wave IMS is still in the process of being developed and different research groups tend to use different approaches for calibration. The calibrations used in this thesis are based on the use of denatured protein calibrants (Smith *et al.* 2009), that have had their absolute cross-sections measured by Clemmer *et al.* (Liu *et al.* 1997; Shelimov *et al.* 1997; Valentine *et al.* 1997a; Valentine *et al.* 1997b) (Amyloid Materials and Methods Section 3.4.10.2). The T-wave height, wave velocity and buffer gas pressure can all be optimised to separate the analyte of ions of interest in the ion mobility cell. For

the cross-section calibration to be valid, the ion mobility calibrants must be measured under the same experimental conditions as the analyte of interest (Shvartsburg *et al.* 2008; Leary *et al.* 2009). Where possible calibrants should be chosen to reflect mass, the mobility space covered and compound class.

1.5 Protein Analysis by MS

1.5.1 Protein Structure by ESI-MS

Mass spectrometry has often been applied within the field of proteomics, aiming to identify the primary structure of proteins along with their expression levels, post-translational modifications, protein-protein interactions and their function within the cell (Aebersold *et al.* 2003). MS over the last 20 years has expanded to the study of native quaternary structure of proteins and their complexes (Ganem *et al.* 2003; Chait 2011; Hilton *et al.* 2012), and it is this aspect of MS that will be covered throughout the rest of this thesis.

One major assumption throughout this thesis is that protein structure in solution is preserved on transition into the gas phase. There is significant evidence that the structure of non-covalent complexes is preserved on transition into the gas phase (Benesch *et al.* 2007; Sharon *et al.* 2007; Heck 2008; Wang *et al.* 2010). Initially there was concern as to whether all solution structures could be preserved into the gas phase since it is known that electrostatic interactions are increased and hydrophobic interactions weakened in the gas phase (Robinson *et al.* 1996; Loo 2000). As such, if protein complexes are maintained by predominantly hydrophobic interactions then preserving these structures in the gas phase was thought not to be possible (Hernandez *et al.* 2007). Recent studies have now shown that hydrophobic interactions can be maintained in the gas phase if the correct conditions are used (Liu *et al.* 2009; Liu *et al.* 2011). Indeed the binding affinities of these hydrophobic complexes as measured by ESI-MS closely match those observed in solution (Liu *et al.* 2011). With the launch of IMS-MS, evidence has increased for the maintenance of solution-phase structure into the gas phase (Ruotolo *et al.* 2005; Benesch *et al.* 2010). As such, it is now widely accepted that under controlled experimental conditions and short time-scales MS can be used to report accurately on solution-phase structures.

1.5.2 Protein Conformation by ESI-MS

Providing native protein structure is preserved on transition into the gas phase, the charge state distribution of proteins in the *m/z* spectrum can be related to the conformation of proteins in solution. During positive ESI, protons attach to the exposed, basic sites of proteins producing multiply charged ions. As such, a folded protein in solution will pick up less charge than its unfolded counterpart resulting in a narrower charge state distribution (Figure 1.9). This technique of determining protein conformation in solution has been widely accepted by the mass spectrometry community. Thus, ESI-MS is frequently used to probe protein conformational changes during pH titrations and temperature ramping and to observe conformational changes on protein/ligand binding.

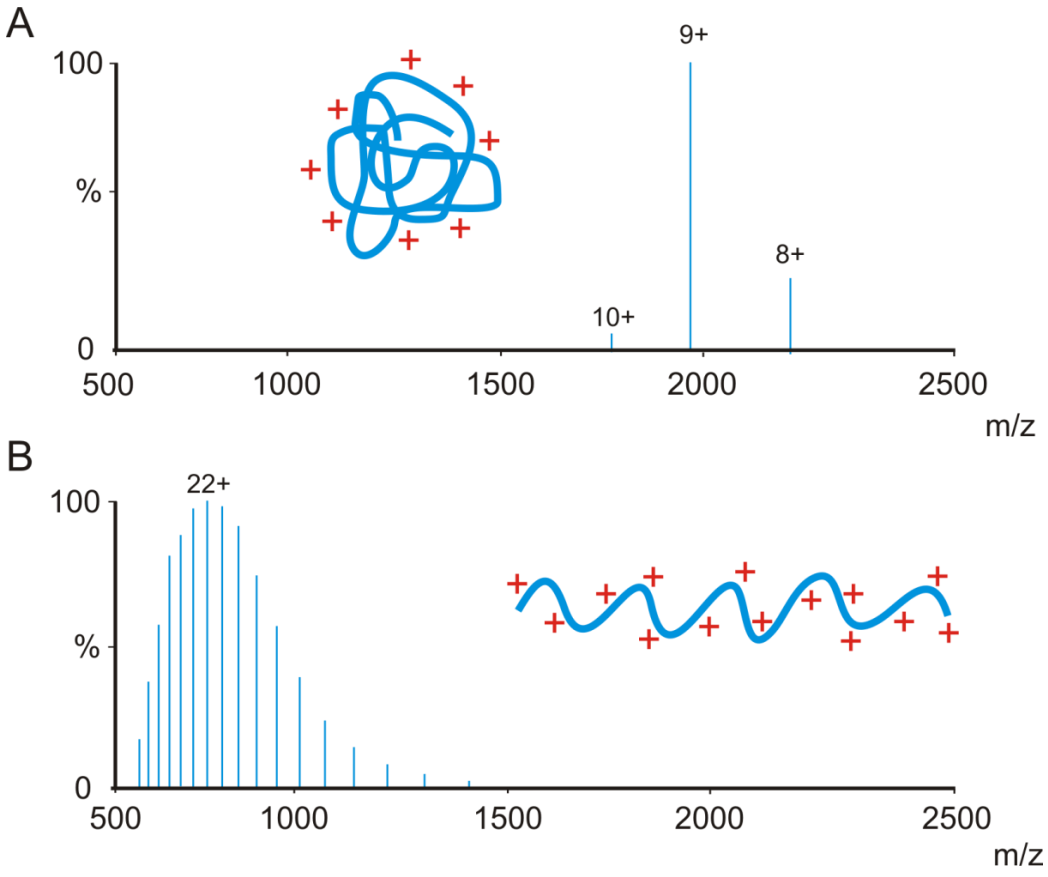


Figure 1.9 – Schematic of a mass spectrum of A) a folded and B) an unfolded protein in solution. The inserts represent the folded and unfolded protein picking up a different number of charges during the electrospray ionisation process.

More recently, the combination of IMS with MS has revealed more detailed information on protein conformations in solution (Kanu *et al.* 2008). IMS-MS can reveal conformational changes within proteins on folding/unfolding. Albert Heck and colleagues have used IMS-MS to look at substrate folding by the chaperonin complex GroEL-GroES. In these studies, the collision cross-sections were identical between the GroEL-GroES complex and the GroEL-GroES-substrate complex confirming that substrate folding occurs within the GroEL-GroES cavity (Van Duijn *et al.* 2009).

In addition to protein conformation, MS can also be used to determine the subunit stoichiometry (Section 1.5.3) and subunit dynamics of protein complexes in solution (Section 1.5.4). Additional advances in MS techniques have allowed quantification of protein complexes in solution (Section 1.5.5) leading to the determination of binding affinities by MS, for example with protein-ligand complexes. This, in combination with IMS-MS can reveal additional possible structural changes that can occur on ligand binding (Sharon *et al.* 2007). For example, a subtle conformational change was observed by IMS-MS on binding of ABT-737 (a small molecule inhibitor) to the Bcl-xL protein involved in cell survival (Atmanene *et al.* 2012). IMS-MS in combination with molecular modelling is very powerful in refining structural details of proteins and their complexes in solution. This is discussed in more detail in Section 1.5.7.

1.5.3 Subunit Stoichiometry by MS

In addition to the molecular weight of macromolecular protein complexes, MS can be used to determine the subunit stoichiometry of protein complexes in solution (McCammon *et al.* 2005). Tandem MS (in particular CID-MS) can be used to increase the internal energy of the complex once transmitted into the gas phase. This increase in energy causes the complex to dissociate into its individual components which can then be identified in the mass spectrum. For example, the dissociation pathway of part of the yeast nuclear-cytoplasmic exosome protein complex (a hexameric ring bound to proteins Rrp4, Rrp40 and a large exoribonuclease, Dis3) was analysed by increasing the collision energy applied to the exosome complex (Figure 1.10). On increasing collision energy, the Rrp40 dissociated first followed by

the Rrp4 protein with the hexameric ring remaining intact and bound to the Dis exoribonuclease protein throughout. These data indicate that the core hexameric ring is stable in the gas phase and is likely to be simply decorated in solution with the weakly bound proteins, Rrp4 and Rrp40.

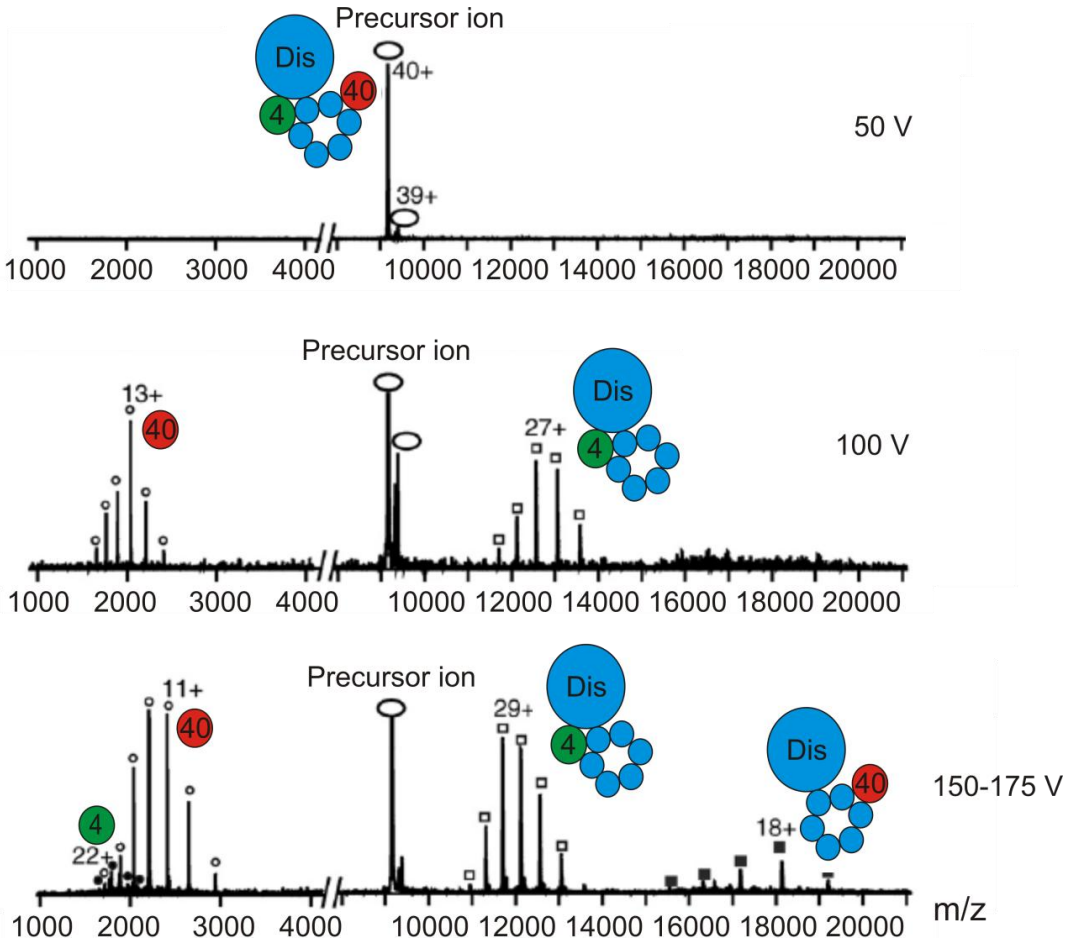


Figure 1.10 - MS/MS spectra of the exosome subcomplex. The exosome subcomplex, consisting of the ring proteins Rrp4 (4), Rrp40 (40) and Dis3 (Dis), with 40 + charges (○) was selected as the precursor ion and a collision voltage of 50 V (top), 100 V (middle) and 150-175 V (bottom) applied to it. On increasing collision voltage from 50 V to 100 V, two charge distributions were observed corresponding to the Rrp40 being released from the complex. On further increasing collision voltage to 150-175 V, the Rrp4 subunit was further released from the complex. This Figure was adapted from Synowsky et al. (Synowsky *et al.* 2009).

Tandem MS has also been used to probe the subunit interactions within the human eukaryotic initiation factor 3 (eIF3) (Zhou *et al.* 2008). Human eIF3 is an 800 kDa protein complex made up of 13 different subunits, the interactions between which were previously unknown. To create an interaction map of the protein subunits that

make up the eIF3 complex, ESI-MS was used initially to transfer eIF3 in its native state into the gas phase. Subsequent CID-MS experiments and in solution dissociation experiments (involving increasing ionic strength to dissociate the complex) allowed 27 subcomplexes of eIF3 to be identified. These data along with previously obtained data on the yeast eIF3 complex (Hinnebusch 2006), allowed a comprehensive interaction map to be proposed for the human eIF3 complex (Zhou *et al.* 2008).

IMS in combination with tandem MS has also been used to determine structural information on intermediates involved in the formation of large macromolecular complexes, for example virus capsids (Knapman *et al.* 2010). Two transient on-pathway intermediates were identified by ESI-MS during MS2 virus capid formation. On further analysis, by ESI-IMS-MS and CID-MS/MS, the mass, stoichiometry and cross-section of these intermediates were identified revealing information on the building blocks and thus the assembly mechanism of the MS2 virus capsid (Knapman *et al.* 2010).

Tandem MS has also been proposed as a tool to determine the nature of protein-protein interfaces based on the charge state distributions of the gas phase dissociated product ions in the MS/MS spectrum (Dodds *et al.* 2011) (see Amyloid Chapter). Additionally, it has only been with the high collision energies provided by CID-MS that membrane complexes have been able to be released from their complexes with detergent in solution and thus be detected by ESI-MS (Barrera *et al.* 2011) (see Membrane Protein Chapter).

1.5.4 Subunit Exchange by MS

Complex biological molecules rarely remain as static entities in solution and are constantly undergoing conformational changes whether it is through ligand binding, enzymatic function or movement throughout the cell (Kaltashov *et al.* 2002). Studying the folding and dynamics of protein complexes and their subunit exchange events is of utmost importance in understanding protein function. This subunit exchange within protein complexes can be detected by MS simply by a change in

mass if hetero-oligomers (subunits with different masses) or homo-oligomers (identical subunits with ^{13}C or ^{15}N labelling) are used (Keetch *et al.* 2005).

The oligomerisation of transthyretin (TTR), a tetrameric protein associated with amyloid disease, has previously been studied by subunit exchange MS analysis (Keetch *et al.* 2005). In this study, the differences in the subunit exchange kinetics were monitored between WT TTR and a more amyloidogenic TTR mutant, L55P. Interestingly, subunit exchange between the L55P mutant reached equilibrium in less than 20 minutes as opposed to WT TTR that took over 9 hours to reach equilibrium. These experiments highlight how subunit exchange can reveal differences in the subunit dynamics between structurally similar oligomeric complexes, revealing novel information, in this case, about the mechanism of amyloid fibril formation. Subunit exchange by MS has more recently been applied to look at the structural dynamics of the protein, β_2-microglobulin, whose oligomers also result in amyloid fibril formation (see Amyloid Chapter) (Smith *et al.* 2010) and virus capsid assembly intermediates (Stockley *et al.* 2007).

1.5.5 Quantification by MS

MS can be used as a quantitative tool to analyse protein complexes in solution. To determine kinetic and thermodynamic parameters of protein complexes by MS, the mass spectrum needs to reflect accurately the distribution of all molecules in solution. The intensity of ions in the *m/z* spectrum has been shown to be comparable for similar protein species that are present in equal concentrations in solution (Benesch *et al.* 2011); however there is an *m/z* dependence in current mass analysers (Liu *et al.* 2011) and detectors (Fraser 2002) that needs to be resolved to obtain absolute quantification of different species in solution based on signal intensity alone. Efforts have been made to enable quantification of proteins in complex with lipids in solution. Indeed, work by the Klassen group has developed methods to account for non-specific binding (reporter molecule method) (Sun *et al.* 2009), in-source dissociation (reference ligand method) (El-Hawiet *et al.* 2010) and interactions difficult to detect by ESI-MS alone (proxy protein method) (El-Hawiet *et al.* 2012).

1.5.6 Kinetic Analysis by MS

Dynamic processes can span a wide range of timescales from femtoseconds to days (Kandt *et al.* 2010). Although MS is limited by speed of sample introduction and the mass analyser, MS can measure structural kinetics on the millisecond timescale (Simmons *et al.* 2002; Pan *et al.* 2010; Stocks *et al.* 2010). The simplest method of kinetic analysis by MS is to monitor the reaction of interest continuously in real time. Prolonged spraying in ESI can cause electrochemical changes in the solutions which may affect the reaction kinetics (Van Berkel 2000). Thus, more commonly manual sampling methods are used at different time points to monitor the reaction over the required time course. Indeed MS has been used for kinetic analysis of proteins in pilus assembly, by monitoring changes in mass spectra as a function of time (Rose *et al.* 2008) (see Pilus Chapter 2.1). MS has also been used to monitor the assembly of the heterohexameric MtGimC chaperone complex (Fandrich *et al.* 2000), revealing information about its assembly mechanism in addition to the reaction kinetics.

1.5.7 Molecular Modelling of Protein Complexes

The combination of collision cross-section measurements from IMS with coarse-grain modelling is very powerful in elucidating structures of proteins and protein-protein complexes. Several methods have been proposed to calculate the cross-section of an ion based on a three-dimensional model of its structure. The projection approximation model is the first model, described in a preliminary, pre-computer state, by Mack in 1925 (Mack 1925), and further developed by Bowers *et al.* (Von Helden *et al.* 1991; Von Helden *et al.* 1993). This model assumes each atom is a hard sphere and calculates the cross-section by averaging the projected shadow of the analyte ion over all possible orientations. The projection approximation model underestimates the collision cross-section of proteins as interactions with the buffer gas are not taken into consideration (Jarrold 1999). More recent extensions to this model have been made to take these problems into account (Mesleh *et al.* 1996; Shvartsburg *et al.* 1996).

A modification of the projection approximation algorithm has also been developed by Dr. T. W. Knapman and Dr. J. T. Berryman *et al.* (formerly University of Leeds), termed the Leeds algorithm and is used in Chapter 2 and 3 of this thesis (Smith *et al.* 2009; Knapman *et al.* 2010). The Leeds algorithm assigns atomic radii based on a PDB structure of the protein or protein complex of interest and a snapshot of the protein taken based on its projection in the x-y plane. An appropriately sized box is then drawn around the projected structure and the area of the structure that would theoretically collide with the buffer gas in the drift cell of the mass spectrometer calculated based on the ratio of hits to misses, where a hit represents an area of protein structure that can collide with the buffer gas and a miss an area where the projected structure is absent. The structure is then rotated and the procedure repeated until the cross-section over all possible orientations is calculated. This method has been used successfully for collision cross-sectional area calculations of the amyloid forming proteins, β_2-microglobulin (Smith *et al.* 2010; Smith *et al.* 2011; Woods *et al.* 2011), in addition to virus capsids (Knapman *et al.* 2010; Shepherd *et al.* 2011).

Molecular dynamics simulations in combination with collision cross-sections can provide additional structural information on the protein-protein interaction of interest. Indeed molecular dynamics simulations on the human amyloid forming protein, islet amyloid polypeptide (IAPP) showed that it adopts an expanded β-hairpin conformation in contrast to the non-amyloid forming rat IAPP whose cross-section matches more of an α-helical conformation (Dupuis *et al.* 2009). Molecular dynamics on large protein complexes is challenging and too computationally expensive for an atomistic approach. Therefore, molecular dynamics on large complexes is typically carried out by dividing each complex into its individual subunit components.

MS is a good technique to study proteins and their complexes in solution. In particular, recent advances in IMS-MS have allowed structural information to be obtained on a large number of protein complexes in addition to information on subunit stoichiometry, binding affinity and subunit dynamics which were not possible by other biochemical techniques.

1.6 Thesis Overview

There are three main objectives of the work presented in this thesis. Firstly, to use mass spectrometry to determine the mechanism of pilus assembly in Gram-negative bacteria. Secondly, to determine details of oligomeric intermediates of amyloid fibril formation for the protein beta-2 microglobulin. Thirdly, to combine a novel method for membrane protein solubilisation for structural analysis of membrane proteins by mass spectrometry.

An introduction to the proteins of interest and current relevant mass spectrometry methods for analysis of these proteins is given at the start of each chapter. A description of all methods employed in this study and the material used, along with the references, are given at the end of each chapter. Chapter 2 includes the development of a mass spectrometry method to determine second order rate constants and its application to revealing insight into pilus assembly. In contrast to the carefully orchestrated pilus assembly process, Chapter 3 shows the application of IMS-MS to the study of the protein, beta-2 microglobulin, which assembles spontaneously into amyloid fibrils resulting in the disease, dialysis related amyloidosis. Chapter 4 includes the analysis of membrane proteins by mass spectrometry. All results are discussed together along with the future applications of mass spectrometry to protein analysis in the final remarks section.

References

Aebersold, R. and Mann, M. (2003). Mass spectrometry-based proteomics. *Nature* **422**(6928): 198-207.

Ashcroft, A. E. (2005). Recent developments in electrospray ionisation mass spectrometry: noncovalently bound protein complexes. *Natural Product Reports* **22**(4): 452-464.

Aston, F. W. (1919). A positive ray spectrograph. *Philosophical Magazine* **38**(228): 707-714.

Atmanene, C., Petiot-Becard, S., Zeyer, D., Van Dorsselaer, A., Hannah, V. V. and Sanglier-Cianferani, S. (2012). Exploring key parameters to detect subtle ligand-induced protein conformational changes using traveling wave ion mobility mass spectrometry. *Anal. Chem.* **84**(11): 4703-4710.

Bakhtiar, R. and Guan, Z. Q. (2006). Electron capture dissociation mass spectrometry in characterization of peptides and proteins. *Biotechnology Lett.* **28**(14): 1047-1059.

Barrera, N. P. and Robinson, C. V. (2011). Advances in the mass spectrometry of membrane proteins: From individual proteins to intact complexes. *Annu. Rev. Biochem.* **80**: 247-271.

Benesch, J. L. and Robinson, C. V. (2006). Mass spectrometry of macromolecular assemblies: Preservation and dissociation. *Curr. Opin. Struct. Biol.* **16**(2): 245-251.

Benesch, J. L. P., Ruotolo, B. T., Simmons, D. A. and Robinson, C. V. (2007). Protein complexes in the gas phase: Technology for structural genomics and proteomics. *Chemical Rev.* **107**(8): 3544-3567.

Benesch, J. L. P., Ruotolo, B. T., Simmons, D. A., Barrera, N. P., Morgner, N., Wang, L. C., Helen, R. and Robinson, C. V. (2010). Separating and visualising protein assemblies by means of preparative mass spectrometry and microscopy. *J. Struct. Biol.* **172**(2): 161-168.

Benesch, J. L. P. and Ruotolo, B. T. (2011). Mass spectrometry: Come of age for structural and dynamical biology. *Curr. Opin. Struct. Biol.* **21**(5): 641-649.

Beynon, J. H. (1956). The use of the mass spectrometer for the identification of organic compounds. *Microchimica Acta* **44**(1): 437-453.

Bich, C. and Zenobi, R. (2009). Mass spectrometry of large complexes. *Curr. Opin. Struct. Biol.* **19**(5): 632-639.

Biemann, K. (1992). Mass-spectrometry of peptides and proteins. *Annu. Rev. Biochem.* **61**: 977-1010.

Bleiholder, C., Wyttenbach, T. and Bowers, M. T. (2011). A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (i). Method. *Int. J. Mass Spectrom.* **308**(1): 1-10.

Carroll, D. I., Dzidic, I., Stillwell, R. N., Haegele, K. D. and Horning, E. C. (1975). Atmospheric-pressure ionization mass-spectrometry - corona discharge ion-source for use in liquid chromatograph mass spectrometer-computer analytical system. *Anal. Chem.* **47**(14): 2369-2373.

Chait, B. T. (2011). Mass spectrometry in the postgenomic era. *Annu. Rev. Biochem.* **80**: 239-246.

Clore, G. M. and Gronenborn, A. M. (1998). NMR structure determination of proteins and protein complexes larger than 20 kDa. *Curr. Opin. Chem. Biol.* **2**(5): 564-570.

Cohen, S. L. and Chait, B. T. (1996). Influence of matrix solution conditions on the MALDI-MS analysis of peptides and proteins. *Anal. Chem.* **68**(1): 31-37.

Creaser, C. S., Griffiths, J. R., Bramwell, C. J., Noreen, S., Hill, C. A. and Thomas, C. L. P. (2004). Ion mobility spectrometry: A review. Part 1. Structural analysis by mobility measurement. *Analyst* **129**(11): 984-994.

de Hoffmann, E., Stroobant, V. (2002). Mass spectrometry: Principles and applications. Chichester, John Wiley and Sons, Ltd.

Dempsey, C. E. (2001). Hydrogen exchange in peptides and proteins using nmr spectroscopy. *Progress in N.M.R. Spectroscopy* **39**(2): 135.

Dempster, A. J. (1918). A new method of positive ray analysis. *Phys. Rev.* **11**(4): 316-325.

Dodds, E. D., Blackwell, A. E., Jones, C. M., Holso, K. L., O'Brien, D. J., Cordes, M. H. J. and Wysocki, V. H. (2011). Determinants of gas-phase disassembly behavior in homodimeric protein complexes with related yet divergent structures. *Anal. Chem.* **83**(10): 3881-3889.

Dole, M., Mack, L. L. and Hines, R. L. (1968). Molecular beams of macroions. *J. Chem. Phys.* **49**(5): 2240.

Dominguez, C., Schubert, M., Duss, O., Ravindranathan, S. and Allain, F. H. T. (2011). Structure determination and dynamics of protein-RNA complexes by NMR spectroscopy. *Progress in N.M.R. Spectroscopy* **58**(1-2): 1-61.

Dupuis, N. F., Wu, C., Shea, J. E. and Bowers, M. T. (2009). Human islet amyloid polypeptide monomers form ordered beta-hairpins: A possible direct amyloidogenic precursor. *J. Am. Chem. Soc.* **131**(51): 18283-18292.

Dyachenko, A., Goldflam, M., Vilaseca, M. and Giralt, E. (2010). Molecular recognition at protein surface in solution and gas phase: Five VegF peptidic ligands show inverse affinity when studied by NMR and CID-MS. *Biopolymers* **94**(6): 689-700.

El-Hawiet, A., Kitova, E. N., Liu, L. and Klassen, J. S. (2010). Quantifying labile protein-ligand interactions using electrospray ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.* **21**(11): 1893-1899.

El-Hawiet, A., Kitova, E. N., Arutyunov, D., Simpson, D. J., Szymanski, C. M. and Klassen, J. S. (2012). Quantifying ligand binding to large protein complexes using electrospray ionization mass spectrometry. *Anal. Chem.* **84**(9): 3867-3870.

Esteban, O., Bernal, R. A., Donohoe, M., Videler, H., Sharon, M., Robinson, C. V. and Stock, D. (2008). Stoichiometry and localization of the stator subunits E and G in *thermus thermophilus* H+-ATPase/synthase. *J. Biol. Chem.* **283**(5): 2595-2603.

Fandrich, M., Tito, M. A., Leroux, M. R., Rostom, A. A., Hartl, F. U., Dobson, C. M. and Robinson, C. V. (2000). Observation of the noncovalent assembly and disassembly pathways of the chaperone complex MtGimC by mass spectrometry. *Proc. Nat. Acad. Sci. U.S.A.* **97**(26): 14151-14155.

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. and Whitehouse, C. M. (1989). Electrospray ionization for mass-spectrometry of large biomolecules. *Science* **246**(4926): 64-71.

Fenselau, C. and Caprioli, R. (2003). Mass spectrometry in the exploration of Mars - introduction. *J. Mass Spectrom.* **38**(1): 1-2.

Fraser, G. W. (2002). The ion detection efficiency of microchannel plates (MCPs). *Int. J. Mass Spectrom.* **215**(1-3): 13-30.

Ganem, B. and Henion, J. D. (2003). Going gently into flight: Analyzing noncovalent interactions by mass spectrometry. *Bioorganic & Med. Chem.* **11**(3): 311-314.

Giles, K., Pringle, S. D., Worthington, K. R., Little, D., Wildgoose, J. L. and Bateman, R. H. (2004). Applications of a travelling wave-based radio-frequency only stacked ring ion guide. *Rapid Comm. Mass Spectrom.* **18**(20): 2401-2414.

Gordiyenko, Y., Videler, H., Zhou, M., McKay, A. R., Fucini, P., Biegel, E., Muller, V. and Robinson, C. V. (2010). Mass spectrometry defines the stoichiometry of ribosomal stalk complexes across the phylogenetic tree. *Mol. Cell. Prot.* **9**(8): 1774-1783.

Griffiths, J. (2008). A brief history of mass spectrometry. *Anal. Chem.* **80**(15): 5678-5683.

Griffiths, W. J., Jonsson, A. P., Liu, S. Y., Rai, D. K. and Wang, Y. Q. (2001). Electrospray and tandem mass spectrometry in biochemistry. *Biochem. J.* **355**: 545-561.

Heck, A. J. R. (2008). Native mass spectrometry: A bridge between interactomics and structural biology. *Nature Meth.* **5**(11): 927-933.

Hemmersbach, P. (2008). History of mass spectrometry at the olympic games. *J. Mass Spectrom.* **43**(7): 839-853.

Hernandez, H. and Robinson, C. V. (2007). Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nature Protocols* **2**(3): 715-726.

Hilton, G. R. and Benesch, J. L. P. (2012). Two decades of studying non-covalent biomolecular assemblies by means of electrospray ionization mass spectrometry. *J. Royal Soc. Interface* **9**(70): 801-816.

Hinnebusch, A. G. (2006). EIF3: A versatile scaffold for translation initiation complexes. *Trends in Biochem. Sci.* **31**(10): 553-562.

Iribarne, J. V., Thomson, B.A. (1976). On the evaporation of small ions from charged droplets. *J. Chem. Phys.* **64**(6): 2287-2294.

Jarrold, M. F. (1999). Unfolding, refolding, and hydration of proteins in the gas phase. *Accounts of Chem. Res.* **32**(4): 360-367.

Juraschek, R., Dulcks, T. and Karas, M. (1999). Nanoelectrospray - more than just a minimized-flow electrospray ionization source. *J. Am. Soc. Mass Spectrom.* **10**(4): 300-308.

Kaltashov, I. A. and Eyles, S. J. (2002). Studies of biomolecular conformations and conformational dynamics by mass spectrometry. *Mass Spectrom. Rev.* **21**(1): 37-71.

Kandt, C. and Monticelli, L. (2010). Membrane protein dynamics from femtoseconds to seconds. *Meth. Mol. Biol.* **654**: 423-440.

Kanu, A. B., Dwivedi, P., Tam, M., Matz, L. and Hill, H. H. (2008). Ion mobility-mass spectrometry. *J. Mass Spectrom.* **43**(1): 1-22.

Karas, M., Bachmann, D., Bahr, U. and Hillenkamp, F. (1987). Matrix-assisted ultraviolet-laser desorption of nonvolatile compounds. *Int. J. Mass Spectrom. and Ion Processes* **78**: 53-68.

Karasek, F. W. (1974). Plasma chromatography. *Anal. Chem.* **46**(8): A710.

Keetch, C. A., Bromley, E. H. C., McCammon, M. G., Wang, N., Christodoulou, J. and Robinson, C. V. (2005). L55P transthyretin accelerates subunit exchange and leads to rapid formation of hybrid tetramers. *J. Biol. Chem.* **280**(50): 41667-41674.

Kirshenbaum, N., Michaelevski, I. and Sharon, M. (2010). Analyzing large protein complexes by structural mass spectrometry. *J. visualized experiments* (40).

Knapman, T. W., Berryman, J. T., Campuzano, I., Harris, S. A. and Ashcroft, A. E. (2010a). Considerations in experimental and theoretical collision cross-section measurements of small molecules using travelling wave ion mobility spectrometry-mass spectrometry. *Int. J. Mass Spectrom.* **298**(1-3): 17-23.

Knapman, T. W., Morton, V. L., Stonehouse, N. J., Stockley, P. G. and Ashcroft, A. E. (2010b). Determining the topology of virus assembly intermediates using ion mobility spectrometry-mass spectrometry. *Rapid Comm. Mass Spectrom.* **24**(20): 3033-3042.

Koppenaal, D. W., Barinaga, C. J., Denton, M. B., Sperline, R. P., Hieftje, G. M., Schilling, G. D., Andrade, F. J. and Barnes, J. H. (2005). MS detectors. *Anal. Chem.* **77**(21): 418-427.

Leary, J. A., Schenauer, M. R., Stefanescu, R., Andaya, A., Ruotolo, B. T., Robinson, C. V., Thalassinos, K., Scrivens, J. H., Sokabe, M. and Hershey, J. W. B. (2009). Methodology for measuring conformation of solvent-disrupted protein

subunits using T-wave ion mobility MS: An investigation into eukaryotic initiation factors. *J. Am. Soc. Mass Spectrom.* **20**(9): 1699-1706.

Li, R., Dai, B. C., Zhao, Y. D. and Lu, K. (2009). Application and development of spectroscopy methodologies in the study on non-covalent interactions. *Spectroscopy and Spectral Analysis* **29**(1): 240-243.

Liu, J. J. and Konermann, L. (2011). Protein-protein binding affinities in solution determined by electrospray mass spectrometry. *J. Am. Soc. Mass Spectrom.* **22**(3): 408-417.

Liu, L., Bagal, D., Kitova, E. N., Schnier, P. D. and Klassen, J. S. (2009). Hydrophobic protein-ligand interactions preserved in the gas phase. *J. Am. Chem. Soc.* **131**(44): 15980.

Liu, L., Kitova, E. N. and Klassen, J. S. (2011). Quantifying protein-fatty acid interactions using electrospray ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.* **22**(2): 310-318.

Liu, Y. S., Valentine, S. J., Counterman, A. E., Hoaglund, C. S. and Clemmer, D. E. (1997). Injected-ion mobility analysis of biomolecules. *Anal. Chem.* **69**(23): 728-735.

Loo, J. A. (1997). Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrom. Rev.* **16**(1): 1-23.

Loo, J. A. (2000). Electrospray ionization mass spectrometry: A technology for studying noncovalent macromolecular complexes. *Int. J. Mass Spectrom.* **200**(1-3): 175-186.

Mack, E. (1925). Average cross-sectional areas of molecules by gaseous diffusion methods. *J. Am. Chem. Soc.* **47**: 2468-2482.

Mamyrin, B. A., Karataev, V. I., Shmikk, D. V. and Zagulin, V. A. (1973). Mass-reflectron a new nonmagnetic time-of-flight high-resolution mass-spectrometer. *Zhurnal Eksperimentalnoi I Teoreticheskoi Fiziki* **64**(1): 82-89.

McCammon, M. G. and Robinson, C. V. (2005). Me, my cell, and I: The role of the collision cell in the tandem mass spectrometry of macromolecules. *Biotechniques* **39**(4): 447.

McDaniel, E. W., Mason, E.A. (1973). The mobility and diffusion of ions in gases *New York, John Wiley and Sons, Inc.*

Mesleh, M. F., Hunter, J. M., Shvartsburg, A. A., Schatz, G. C. and Jarrold, M. F. (1996). Structural information from ion mobility measurements: Effects of the long-range potential. *J. Phys. Chem.* **100**(40): 16082-16086.

Morrissey, B., Leney, A. C., Toste Rego, A., Phan, G., Allen, W. J., Verger, D., Waksman, G., Ashcroft, A. E. and Radford, S. E. (2012). The role of chaperone-subunit usher domain interactions in the mechanism of bacterial pilus biogenesis revealed by ESI-MS. *Mol. Cell. Prot.* **11**(7): M111.015289.

Munson, M. S. B. and Field, F. H. (1966). Chemical ionisation mass spectrometry. 1. General introduction. *J. Am. Chem. Soc.* **88**(12): 2621.

Pan, J. X., Han, J., Borchers, C. H. and Konermann, L. (2010). Characterizing short-lived protein folding intermediates by top-down hydrogen exchange mass spectrometry. *Anal. Chem.* **82**(20): 8591-8597.

Petrova, T. and Podjarny, A. (2004). Protein crystallography at subatomic resolution. *Rep. Progress Phys*. **67**(9): 1565-1605.

Pringle, S. D., Giles, K., Wildgoose, J. L., Williams, J. P., Slade, S. E., Thalassinos, K., Bateman, R. H., Bowers, M. T. and Scrivens, J. H. (2007). An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *Int. J. Mass Spectrom.* **261**(1): 1-12.

Revercomb, H. E. and Mason, E. A. (1975). Theory of plasma chromatography gaseous electrophoresis - review. *Anal. Chem.* **47**(7): 970-983.

Ritschel, T., Atmanene, C., Reuter, K., Van Dorsselaer, A., Sanglier-Cianferani, S. and Klebe, G. (2009). An integrative approach combining noncovalent mass spectrometry, enzyme kinetics and X-ray crystallography to decipher Tgt protein-protein and protein-RNA interaction. *J. Mol. Biol.* **393**(4): 833-847.

Robinson, C. V., Chung, E. W., Kragelund, B. B., Knudsen, J., Aplin, R. T., Poulsen, F. M. and Dobson, C. M. (1996). Probing the nature of noncovalent interactions by mass spectrometry. A study of protein-CoA ligand binding and assembly. *J. Am. Chem. Soc.* **118**(36): 8646-8653.

Roepstorff, P. and Fohlman, J. (1984). Proposal for a common nomenclature for sequence ions in mass-spectra of peptides. *Biomed. Mass Spectrom.* **11**(11): 601-601.

Rose, R. J., Verger, D., Daviter, T., Remaut, H., Paci, E., Waksman, G., Ashcroft, A. E. and Radford, S. E. (2008). Unraveling the molecular basis of subunit specificity in

P pilus assembly by mass spectrometry. *Proc. Nat. Acad. Sci. U.S.A.* **105**(35): 12873-12878.

Ruotolo, B. T., Giles, K., Campuzano, I., Sandercock, A. M., Bateman, R. H. and Robinson, C. V. (2005). Evidence for macromolecular protein rings in the absence of bulk water. *Science* **310**(5754): 1658-1661.

Sharon, M. and Robinson, C. V. (2007). The role of mass spectrometry in structure elucidation of dynamic protein complexes. *Annu. Rev. Biochem.* **76**: 167-193.

Shelimov, K. B., Clemmer, D. E., Hudgins, R. R. and Jarrold, M. F. (1997). Protein structure *in vacuo*: Gas-phase confirmations of BPTI and cytochrome C. *J. Am. Chem. Soc.* **119**(9): 2240-2248.

Shepherd, D. A., Veesler, D., Lichiere, J., Ashcroft, A. E. and Cambillau, C. (2011). Unraveling lactococcal phage baseplate assembly by mass spectrometry. *Mol. Cell. Prot.* **10**(9).

Shi, S. D. H., Hendrickson, C. L. and Marshall, A. G. (1998). Counting individual sulfur atoms in a protein by ultrahigh-resolution fourier transform ion cyclotron resonance mass spectrometry: experimental resolution of isotopic fine structure in proteins. *Proc. Nat. Acad. Sci. U.S.A.* **95**(20): 11532-11537.

Shortle, D. R. (1996). Structural analysis of non-native states of proteins by NMR methods. *Curr.Opin.Struct. Biol.* **6**(1): 24.
Shvartsburg, A. A. and Jarrold, M. F. (1996). An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chem. Phys. Lett.* **261**(1-2): 86-91.

Shvartsburg, A. A. and Smith, R. D. (2008). Fundamentals of traveling wave ion mobility spectrometry. *Anal. Chem.* **80**(24): 9689-9699.

Simmons, D. A. and Konermann, L. (2002). Characterization of transient protein folding intermediates during myoglobin reconstitution by time-resolved electrospray mass spectrometry with on-line isotopic pulse labeling. *Biochemistry* **41**(6): 1906-1914.

Smith, D. P., Knapman, T. W., Campuzano, I., Malham, R. W., Berryman, J. T., Radford, S. E. and Ashcroft, A. E. (2009). Deciphering drift time measurements from travelling wave ion mobility spectrometry-mass spectrometry studies. *Euro. J. Mass Spectrom.* **15**(2): 113-130.

Smith, D. P., Radford, S. E. and Ashcroft, A. E. (2010). Elongated oligomers in beta(2)-microglobulin amyloid assembly revealed by ion mobility spectrometry-mass spectrometry. *Proc. Nat. Acad. Sci. U.S.A.* **107**(15): 6794-6798.

Smith, D. P., Woods, L. A., Radford, S. E. and Ashcroft, A. E. (2011). Structure and dynamics of oligomeric intermediates in beta(2)-microglobulin self-assembly. *Biophys. J.* **101**(5): 1238-1247.

Sobott, F., Hernandez, H., McCammon, M. G., Tito, M. A. and Robinson, C. V. (2002). A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal. Chem.* **74**(6): 1402-1407.

Stockley, P. G., Rolfsson, O., Thompson, G. S., Basnak, G., Francese, S., Stonehouse, N. J., Homans, S. W. and Ashcroft, A. E. (2007). A simple, RNA-mediated allosteric switch controls the pathway to formation of a T=3 viral capsid. *J. Mol. Biol.* **369**(2): 541-552.

Stocks, B. B. and Konermann, L. (2010). Time-dependent changes in side-chain solvent accessibility during cytochrome C folding probed by pulsed oxidative labeling and mass spectrometry. *J. Mol. Biol.* **398**(2): 362-373.

Summerfield, S. G., Whiting, A. and Gaskell, S. J. (1997). Intra-ionic interactions in electrosprayed peptide ions. *Int. J. Mass Spectrom. and Ion Processes* **162**(1-3): 149-161.

Sun, N., Sun, J. X., Kitova, E. N. and Klassen, J. S. (2009). Identifying nonspecific ligand binding in electrospray ionization mass spectrometry using the reporter molecule method. *J. Am. Soc. Mass Spectrom.* **20**(7): 1242-1250.

Svedberg, T. and Fahraeus, R. (1926). A new method for the determination of the molecular weight of the proteins. *J. Am. Chem. Soc.* **48**(1): 430-438.

Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J. and Hunt, D. F. (2004). Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Nat. Acad. Sci. U.S.A.* **101**(26): 9528-9533.

Synowsky, S. A., van Wijk, M., Raijmakers, R. and Heck, A. J. R. (2009). Comparative multiplexed mass spectrometric analyses of endogenously expressed yeast nuclear and cytoplasmic exosomes. *J. Mol. Biol.* **385**(4): 1300-1313.

Thomson, J. J. (1899). On the masses of the ions in gases at low pressures. *Philosophical Magazine* **48**: 547-567.

Thomson, J. J. (1911). Rays of positive electricity. *Philosophical Magazine* **21**(122): 225-249.

Valentine, S. J., Anderson, J. G., Ellington, A. D. and Clemmer, D. E. (1997a). Disulfide-intact and -reduced lysozyme in the gas phase: Conformations and pathways of folding and unfolding. *J. Phys. Chem.* **101**(19): 3891-3900.

Valentine, S. J., Counterman, A. E. and Clemmer, D. E. (1997b). Conformer-dependent proton-transfer reactions of ubiquitin ions. *J. Am. Soc. Mass Spectrom.* **8**(9): 954-961.

Van Berkel, G. J. (2000). Electrolytic deposition of metals on to the high-voltage contact in an electrospray emitter: Implications for gas-phase ion formation. *J. Mass Spectrom.* **35**(7): 773-783.

Van Duijn, E., Barendregt, A., Synowsky, S., Versluis, C. and Heck, A. J. R. (2009). Chaperonin complexes monitored by ion mobility mass spectrometry. *J. Am. Chem. Soc.* **131**(4): 1452-1459.

Vestal, M. L., Juhasz, P. and Martin, S. A. (1995). Delayed extraction matrix-assisted laser-desorption time-of-flight mass-spectrometry. *Rapid Comm. Mass Spectrom.* **9**(11): 1044-1050.

Von Helden, G., Hsu, M. T., Kemper, P. R. and Bowers, M. T. (1991). Structures of carbon cluster ions from 3 to 60 atoms - linears to rings to fullerenes. *J. Chem. Phys.* **95**(5): 3835-3837.

Von Helden, G., Hsu, M. T., Gotts, N. and Bowers, M. T. (1993). Carbon cluster cations with up to 84 atoms - structures, formation mechanism, and reactivity. *J. Phys. Chem.* **97**(31): 8182-8192.

Wang, S. C., Politis, A., Di Bartolo, N., Bavro, V. N., Tucker, S. J., Booth, P. J., Barrera, N. P. and Robinson, C. V. (2010). Ion mobility mass spectrometry of two tetrameric membrane protein complexes reveals compact structures and differences in stability and packing. *J. Am. Chem. Soc.* **132**(44): 15468-15470.

Wildgoose, J. L., Giles, K., Pringle, S.D., Koeniger, S., Valentine, S.J., Bateman, R.H., Clemmer, D.E. (2006). A comparison of travelling wave and drift tube ion mobility separations. *Proc. 54th ASMS Conference on Mass Spectrometry and Allied Topics, Seattle, WA, USA.*

Wilm, M. and Mann, M. (1996). Analytical properties of the nanoelectrospray ion source. *Anal. Chem.* **68**(1): 1-8.

Woods, L. A., Platt, G. W., Hellewell, A. L., Hewitt, E. W., Homans, S. W., Ashcroft, A. E. and Radford, S. E. (2011). Ligand binding to distinct states diverts aggregation of an amyloid-forming protein. *Nat. Chem. Biol.* **7**(10): 730-739.

Wu, S. L., Huehmer, A. F. R., Hao, Z. Q. and Karger, B. L. (2007). On-line LC-MS approach combining collision-induced dissociation (CID), electron-transfer dissociation (ETD), and CID of an isolated charge-reduced species for the trace-level characterization of proteins with post-translational modifications. *J. Prot. Res.* **6**(11): 4230-4244.

Zhao, X. (2012). Protein structure determination by solid-state NMR. *Topics in current chemistry* **326**: 187-213.

Zhou, M., Sandercock, A. M., Fraser, C. S., Ridlova, G., Stephens, E., Schenauer, M. R., Yokoi-Fong, T., Barsky, D., Leary, J. A., Hershey, J. W., Doudna, J. A. and Robinson, C. V. (2008). Mass spectrometry reveals modularity and a complete subunit interaction map of the eukaryotic translation factor EIF3. *Proc. Nat. Acad. Sci. U.S.A.* **105**(47): 18139-18144.

Zubarev, R. A., Kelleher, N. L. and McLafferty, F. W. (1998). Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J. Am. Chem. Soc.* **120**(13): 3265-3266.

2.1. Introduction

Pilus Assembly

2.1 Bacterial Secretion Systems

Secretion systems are used by a variety of bacteria to transport proteins across their outer membranes. The secretion of virulence factors, such as toxins, and the assembly of adhesive organelles on the bacterial cell surface are crucial steps in the transmission of infection and hence are possible targets for anti-microbial reagents. Current successful antimicrobial agents target different microbial modes of action such as DNA replication or protein synthesis (Cegelski *et al.* 2008). Targeting these factors stimulates bacterial evolution due to the effect antimicrobial agents have on microbial survival. An alternative method of pathogenesis prevention is to target only the virulence properties of bacteria, for example the attachment of bacteria to host cells via hair-like appendages termed pili. In this way, microbes can still survive but no longer infect the host, and hence problems with antibiotic resistance will be kept to a minimum.

The general secretory pathway (GSP) involves the translocation of proteins across the cytoplasmic membrane *via* the sec machinery, and the assembly of these proteins into pili on the bacterial surface using a specific secretion apparatus (Stathopoulos *et al.* 2000). This thesis focuses on one type of GSP, the chaperone-usher pathway, that assembles over 30 adhesins each associated with a specific disease (Hung *et al.* 1998; Sauer *et al.* 2000).

2.1.1 Chaperone-usher pathway

The two widest researched chaperone-usher systems are the Type 1 pilus and the P pilus from uropathogenic *Escherichia coli*, encoded for by the fim and pap gene clusters, respectively (Hull *et al.* 1981; Sauer *et al.* 2004). P pili and Type 1 pili are involved in pyelonephritis and cystitis respectively, and their transmission is highly dependent on the conserved subunits' order and number in the final pilus structure (Wiles *et al.* 2008).

P pili are assembled by the PapD chaperone and the PapC usher, and consist of six subunits of 2 nm in diameter (PapG, PapF, PapE, PapK, PapA, PapH) arranged in a specific order (Figure 2.1.1) (Dhakal *et al.* 2009). PapG has an adhesion domain ($PapG_{adhesin}$) and a pilin domain ($PapG_{pilin}$) is the first subunit translocated across the outer membrane. This adhesin binds to host tissues, *via* a di-galactose binding domain, and initiates bacterial infection. In between the adapter subunits, PapF and PapK, is the major component of the tip fibrillum, PapE, which is present in five to ten copies (Kuehn *et al.* 1992). After the assembly of the flexible tip, thousands of PapA subunits translocate across the outer membrane and assemble into a rigid right-handed helix (Mu *et al.* 2006). Finally, pilus assembly terminates by incorporating a single copy of PapH.

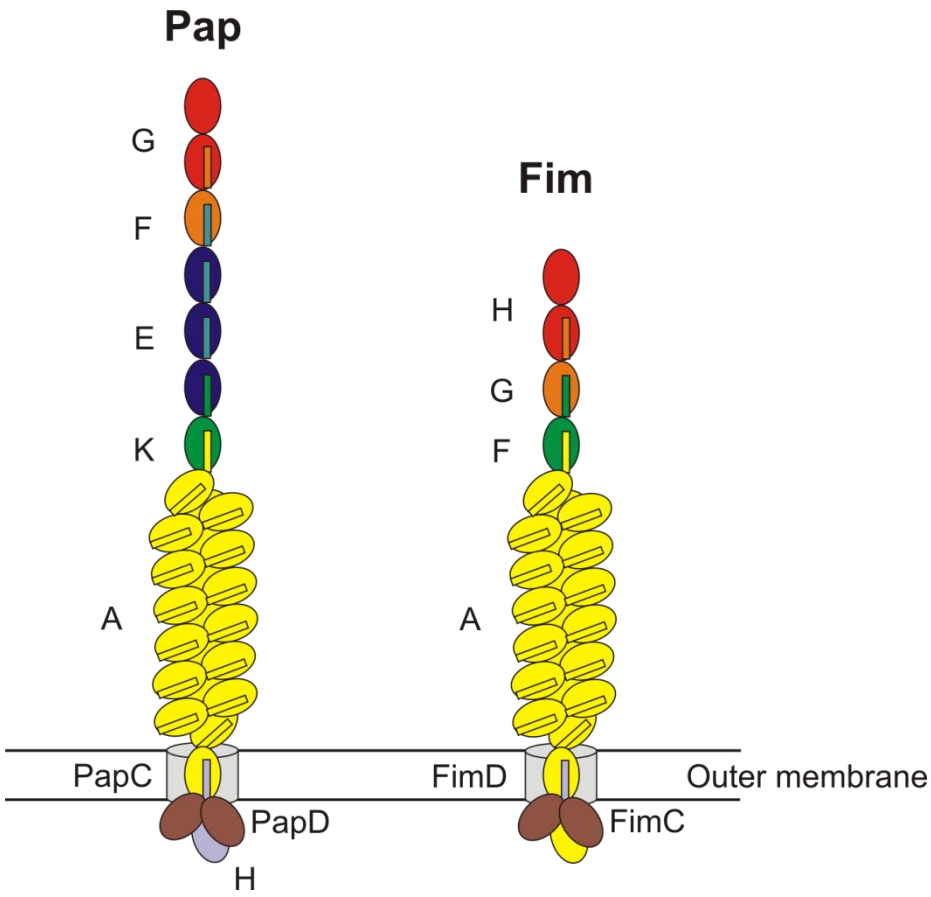


Figure 2.1.1 – Schematic structure of P pili (left) and Type 1 pilus (right) assembled on the outer membrane. In P pilus (Pap) subunits are labelled A (yellow), K (green), E (blue), F (orange), G (red) and H (light blue), the chaperone D (brown) and the usher C (grey). In Type 1 pilus (Fim) the subunits are labelled A (yellow), F (green), G (orange) and H (red), the chaperone C (brown) and the usher D (grey) (Figure adapted from (Leney *et al.* 2011). Both PapG and FimH contain both an adhesion and pilin domain.

Type 1 pili have only four subunits (FimA, FimF, FimG, FimH) and are assembled by a chaperone, FimC and the usher, FimD. FimA subunits fold into a similar helical rod to those of PapA in P pili; however the tip fibrillum is shorter (Jones *et al.* 1995), consisting of an adhesion domain, FimH, and adaptor subunits FimF and FimG (Figure 2.1.1).

2.1.2 Role of the Periplasmic Chaperone

Periplasmic chaperones play an important role in the chaperone-usher pathway. The chaperone was first thought to have a role in the chaperone-usher pathway when its high sequence identity among Gram-negative pathogens was observed (Hung *et al.* 1996). In 1989, the first X-ray structure of the PapD chaperone was solved, showing PapD consisted of two immunoglobulin domains positioned at right angles to each other (Figure 2.1.2) (Holmgren *et al.* 1989). The chaperones all have a conserved, characteristic hydrophobic core and a series of hydrophobic residues on the surface-exposed G_1 strand.

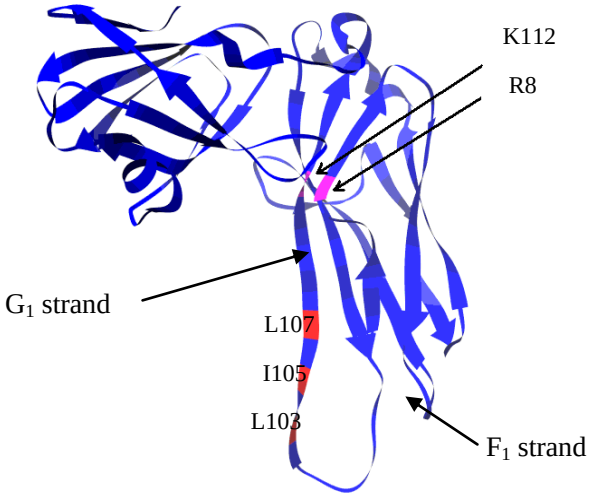


Figure 2.1.2 - Crystal structure of PapD chaperone (PDB file 3DPA) (Holmgren *et al.* 1989). The hydrophobic residues (L103, I105, L107) on the G_1 strand (red) and the basic residues (R8 and K112) (magenta) are conserved between all pilus related chaperones. Structure drawn using Pymol (DeLano 2002).

2.1.3 Chaperone-subunit complexes

All pilus subunits alone fold into a structure with an Ig-like fold, but the C-terminal β-strand of the typical Ig domain is lacking. These pilus subunits, therefore, are unstable and prone to aggregation. During the chaperone-usher pathway pilus subunits are stabilised by binding to a periplasmic chaperone, preventing them from degradation by proteases such as DegP (Bakker *et al.* 1991). This mechanism, known as donor strand complementation (DSC), results in a stable chaperone-subunit interaction whereby the chaperone's hydrophobic G_1 strand is inserted into the subunit's hydrophobic pocket (Figure 2.1.3a). The first evidence of DSC was found in the crystal structure of FimCFimH (Knight *et al.* 1997) and was further supported by the following PapDPapK, PapDPapH, PapDPapE, PapDPapF and PapDPapA chaperone-subunit crystal structures from P pilus (Sauer *et al.* 1999; Sauer *et al.* 2002; Verger *et al.* 2006; Verger *et al.* 2007; Verger *et al.* 2008). In each case specific chaperone residues, named P1-P4, insert into the subunits' P1-P4 hydrophobic pockets (Figure 2.1.3b).

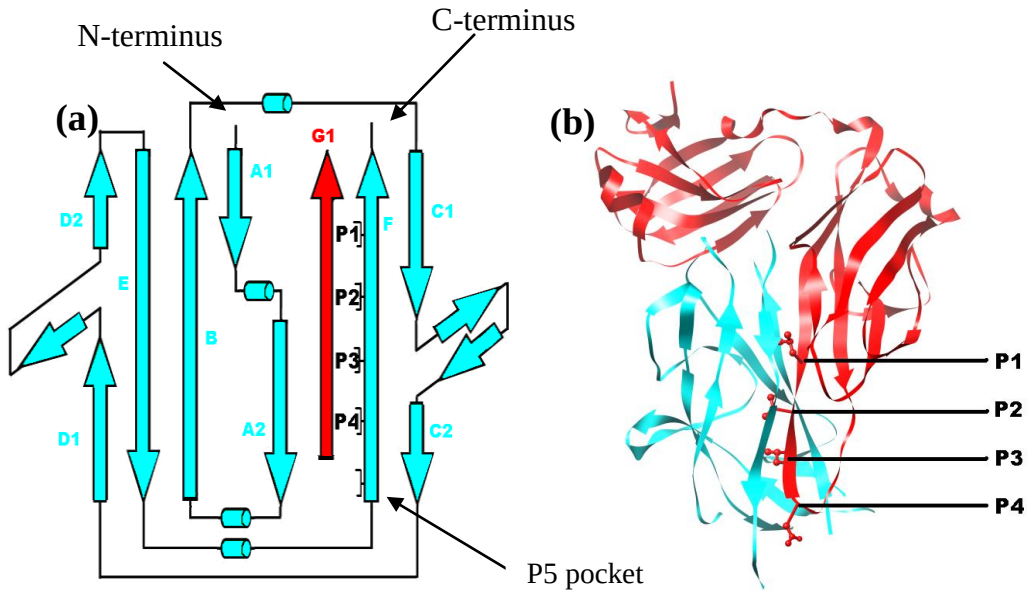


Figure 2.1.3 - a) Topology diagram of chaperone G_1 strand (red) bound to a subunit (blue) showing residues P1-P4 fitting into pockets P1-P4 (the empty P5 pocket is also shown). The subunit also has an N-terminal extension (not shown) that extends from the A1 strand. b) Crystal structure of the chaperone PapD (red) bound to the PapE subunit (blue) with P1-P4 residues of the chaperone indicated as ball-and-stick structures (PDB file 1N01) (Sauer *et al.* 2002). Structure drawn using Pymol (DeLano 2002).

2.1.4 Subunit-subunit complexes

All pilus subunits have an N-terminal extension (Nte) made up of alternating hydrophobic residues. This is absent in the X-ray crystal structure due to its disordered nature, but plays a vital role in subunit-subunit interactions. Figure 2.1.4 shows the crystal structure of PapE bound to K_{Nte} (a peptide representing the Nte of PapK subunit). As expected, due to the similarity in alternating hydrophobic residues, the Nte binds into the same hydrophobic groove (this time, using the P1-P5 pockets) on the subunit as the chaperone's G_1 strand does during DSC. In this case the Nte aligns antiparallel to the F strand (unlike the chaperone G_1 strand) in an interaction known as donor strand exchange (DSE).

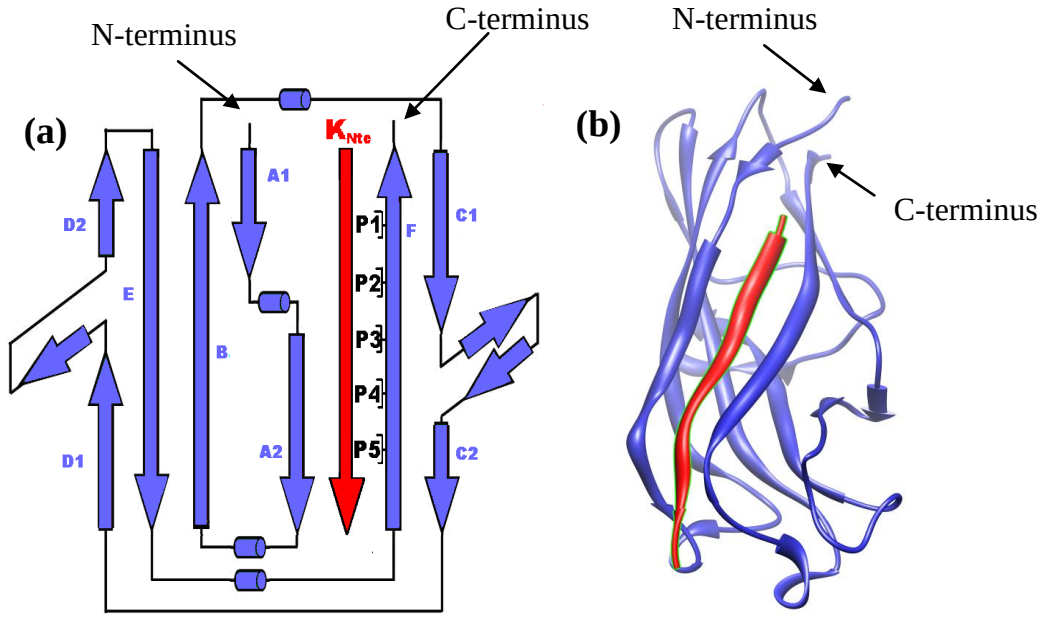


Figure 2.1.4 - (a) Topology diagram of $PapK_{Nte}$ (red) bound to a $PapE_{Ntd}$ subunit (blue) showing residues P1-P5 fitting into pockets P1-P5; (b) Crystal structure of a DSE complex with $PapE_{Ntd}$ (blue) bound to K_{Nte} (red). (PDB file 1N12) (Sauer *et al.* 2002). $PapE_{Ntd}$ is a version of PapE in which the Nte is deleted to prevent self-polymerisation of PapE subunits in solution. The structure was drawn using Pymol (DeLano 2002).

2.1.5 The donor strand exchange (DSE) mechanism

At the usher, chaperone-subunit complexes undergo a mechanism termed donor strand exchange (DSE) resulting in the formation of the final subunit-subunit complex and its translocation across the outer membrane. During DSE, the chaperone's G_1 strand in the chaperone-subunit complex is gradually replaced,

residue by residue with the N-terminal extension of an incoming subunit in a zip-in-zip-out mechanism.

Mass spectrometry, in combination with other biochemical techniques has been used to follow this DSE mechanism in detail (Remaut *et al.* 2006; Rose *et al.* 2008; Rose *et al.* 2008; Verger *et al.* 2008). Mass spectrometry experiments by Rose *et al*. set out to test how thousands of structurally similar subunits could assemble in the precise order and number essential for pilus function (Rose *et al.* 2008). Since the subunit specificity was thought to be determined by the pilus subunits themselves (Lee *et al.* 2007), all 25 possible DSE reactions between chaperone-subunit complexes (PapD bound to PapG, PapF, PapE, PapK, PapA) and subunit Ntes were carried out, and their observed reaction rates determined (Rose *et al.* 2008). On comparison of the DSE rates of reaction, it was seen that cognate reactions (i.e. those observed *in vivo*) occur at higher rates than their non-cognate counterparts. It was concluded, therefore, that each subunit groove interacts differently with different Ntes, hence the Nte must impart some specificity to the DSE reaction.

By comparing the different subunit Nte sequences, shown in Figure 2.1.5, it is not obvious how this subunit specificity is conferred. Further detailed MS experiments using chimeric peptides showed that the residues on the Nte at, and around, the P5 pocket (i.e. residues P5-1, P5, P5+1 and P5+2) are the most important in determining the DSE reaction rate (Rose *et al.* 2008). Further kinetic analysis of peptides with mutations in these residues were needed to gain an insight into the precise role of individual residues on the subunit Nte in conferring specificity in the DSE mechanism; the results of these experiments will be discussed further in this Chapter.

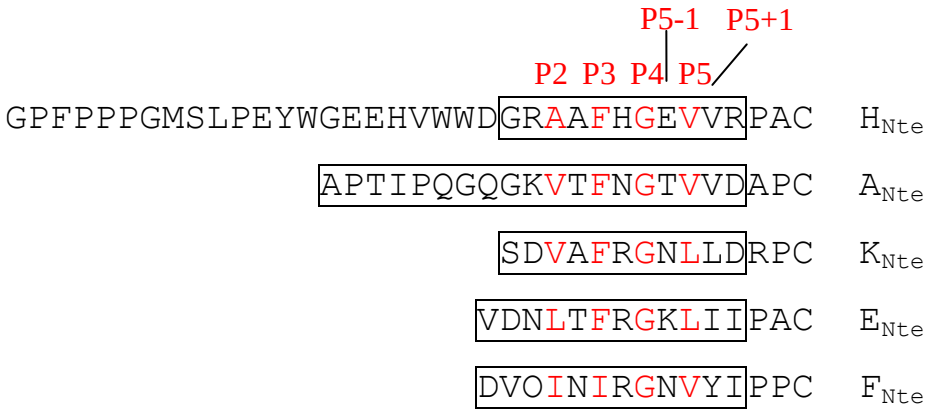


Figure 2.1.5 - Subunit Nte sequences of PapX subunits with P2-P5 residues highlighted in red. The black rectangles indicate the section of the Nte used in DSE experiments by Rose *et al.* 2008.

2.1.6 Additional factors that may contribute to subunit specificity

Although the N-terminal extension on the incoming subunit alone contributes to the subunit ordering in pilus biogenesis, this is not the only factor thought to be involved as DSE occurs in minutes *in vivo* at the usher without an external energy source (Jacob-Dubuisson *et al.* 1994) as opposed to the hour timescale observed with *in vitro* experiments (Rose *et al.* 2008). Indeed simulations using simple mathematical models showed that using the pseudo first order rates of DSE obtained using peptides equivalent to the Nte, a correct assembled pilus would be unlikely to occur (E. Paci, R. J. Rose, S. E. Radford, A. E. Ashcroft, unpublished data).

One major component absent in the *in vitro* studies described above is the outer membrane usher. More recent studies have, therefore, concentrated on the usher and the role it may have in pilus subunit specificity and DSE catalysis. Studies of the usher are described in the second part of this chapter.

2.1.7 Role of the Outer Membrane Usher in DSE

The usher protein is crucial for the translocation of pilus subunits across the bacterial outer membrane, since on deletion of the membrane usher, chaperone-subunits accumulate in the periplasm, and no pili form (Klemm *et al.* 1990). Cryo-electron

microscopy and other biophysical studies have shown that the usher outer membrane protein is a dimer, consisting of two monomers connected by detergent molecules (Li *et al.* 2004; Ng *et al.* 2004; So *et al.* 2006; Remaut *et al.* 2008); however, it has been suggested that only one usher is needed for function (Huang *et al.* 2009).

How pilus assembly at the usher occurs without an external energy source, e.g., ATP, is fascinating (Jones *et al.* 1993). The energy needed to drive the self-assembly process is thought to come from the difference in stabilities between the chaperone-subunit and the subunit-subunit interactions (Zavialov *et al.* 2003). The subunit-subunit complex adopts a more compact conformation compared with that of the chaperone-subunit complex. It is thought that the chaperone traps the subunit in a higher energy, partially folded state, which then collapses on binding an Nte to form a lower energy, more compact structural arrangement (Zavialov *et al.* 2003). The increased stability obtained from this conformational change releases the free energy needed to drive pilus assembly (Zavialvo *et al.* 2005).

The usher, both from P pilus and Type 1 pili, consists of four domains: an N-terminal domain (~125 residues) termed UsherN, a middle plug domain (residues ~220 to ~325), termed UsherP, surrounded by a central β-barrel domain (residues ~135 to ~640), and a C-terminal domain (residues ~638 to ~809) termed UsherC (Figure 2.1.6) (Brandenburg *et al.* 2005; Capitani *et al.* 2006; Bleiholder *et al.* 2011). Together the usher domains act to catalyse DSE and to thread the pilus subunits one by one through the usher β-barrel until the final pilus is formed.

Figure 2.1.6 - Usher domains: N-terminal domain (blue, UsherN); plug domain (pink, UsherP); β-barrel domain (black); C-terminal domain (green, Usher C_1 and Usher C_2).

The periplasmic N-terminal domain of the usher is thought to be the initial binding site for chaperone-subunit complexes prior to DSE and pilus biogenesis (Ng *et al.* 2004). The differential affinity of chaperone-subunit complexes for the usher's N-

terminal domain matches the order in which the subunits are translocated (Dodson *et al.* 1993). It has been suggested that this difference in affinity partly explains how the tip subunit, PapG, binds preferentially to the usher over PapA, which is present at a significantly higher concentration than all other pilus subunits combined in the periplasm (Ahn *et al.* 2004). Additional evidence from X-ray crystallography shows UsherN to bind predominantly to the N-terminal domain of the chaperone and partly to the chaperone-bound subunit in the $FimD_{N}FimCFimH_{pilin}$ (UsherN-chaperone-subunit) complex (Nishiyama *et al.* 2005) (Figure 2.1.7). This binding of the chaperone-subunit complexes to UsherN is thought to catalyse DSE by positioning the incoming chaperone-subunit complex in close proximity to the preceding chaperone-subunit complex (Remaut *et al.* 2008).

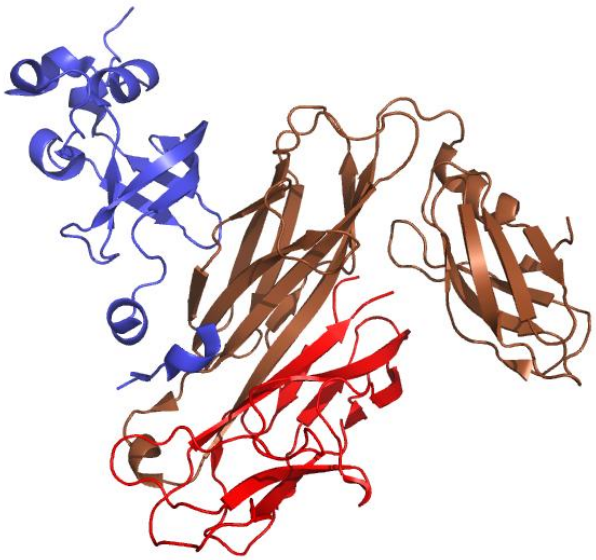


Figure 2.1.7 – Crystal structure of the chaperone-subunit complex, $FimCFimH_{pilin}$ (brown-red) bound to the N-terminal domain of the Usher, FimD (blue) (PDB file 1ZE3) (Nishiyama *et al.* 2005). $FimH_{pilin}$ corresponds to FimH where the adhesion domain has been removed. The structures were drawn using Pymol (DeLano 2002).

Prior to pilus biogenesis, the usher's plug domain resides inside the usher's β-barrel preventing the passage of solutes or periplasmic proteins across the channel (Remaut *et al.* 2008; Phan *et al.* 2011) (Figure 2.1.8A). Upon binding of the first chaperone-subunit complexes to the N-terminal domain, the usher β-barrel undergoes a conformational change resulting in the plug being released from the β-barrel (Saulino *et al.* 1998; Bayburt *et al.* 2010; Aminlashgari *et al.* 2012; Bechara *et al.*

2012) (Figure 2.1.8b). Once the usher is activated, the plug remains in the periplasm and here it is thought to play an additional role in donor strand exchange catalysis (Huang *et al.* 2009).

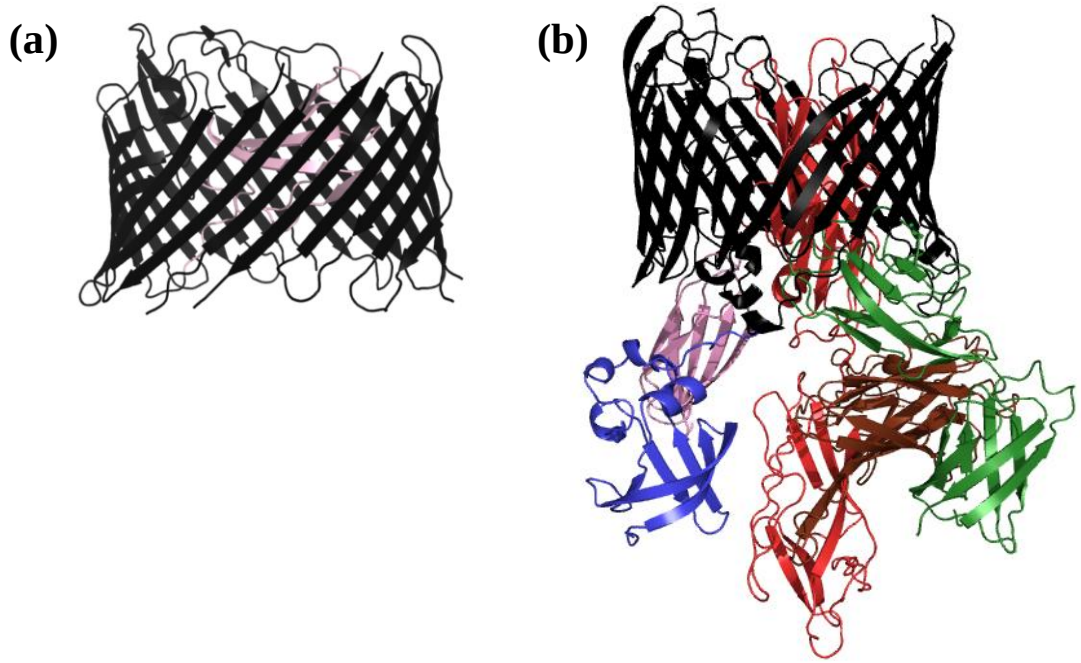


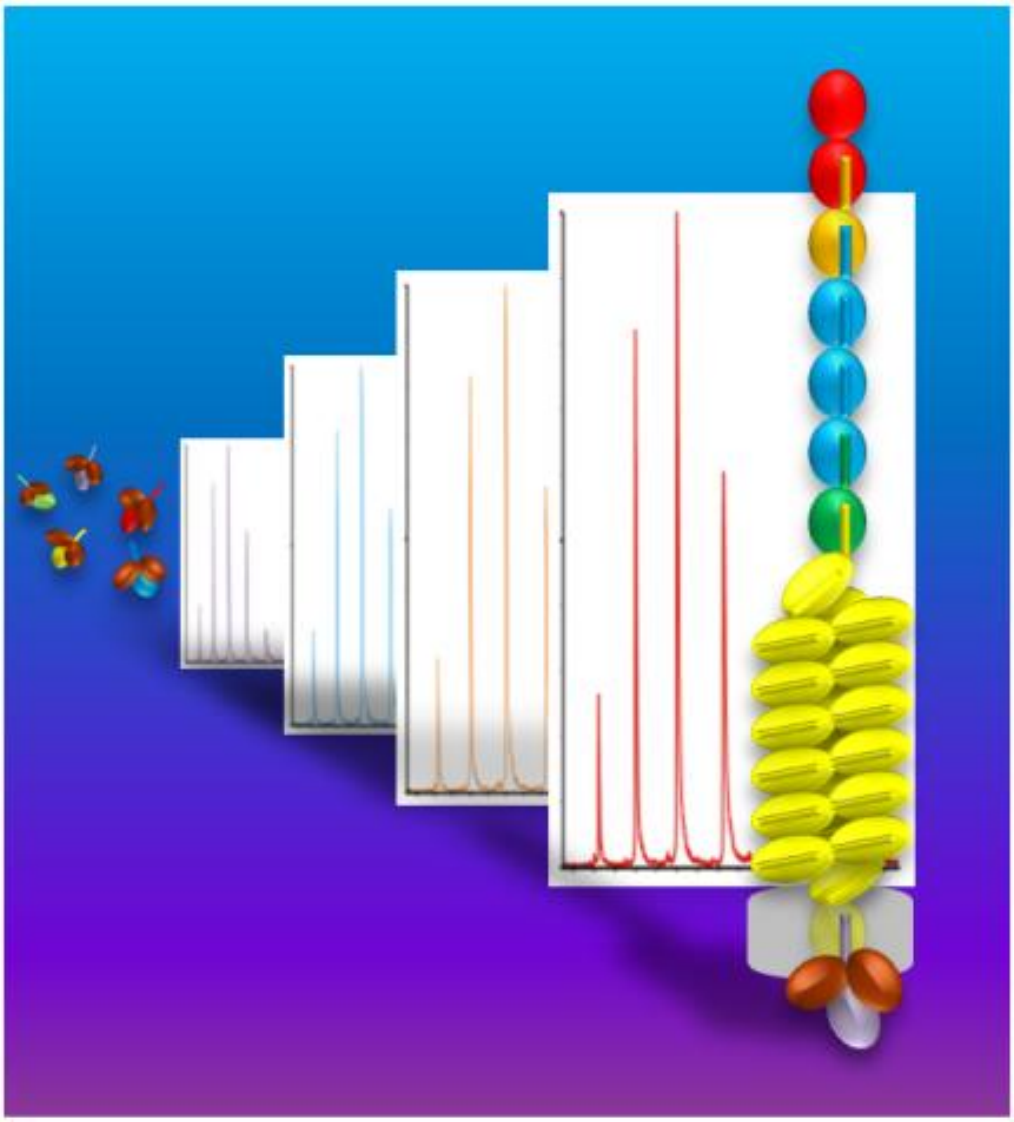
Figure 2.1.8 (a) Crystal structure of the inactivated outer membrane usher, FimD (PDB file 3OHN) (Phan *et al.* 2011) showing the plug domain (pink) located inside the barrel (black). The N-terminal and C-terminal domains were deleted in this construct. (b) Crystal Structure of the chaperone-subunit complex FimCFimH (brown-red) translocating the FimD usher (PDB file 3RFZ) (Phan *et al.* 2011). The N-terminal domain, plug, β-barrel and C-terminal domain of the usher are highlighted in blue, pink, black and green respectively. The structures were drawn using Pymol (DeLano 2002).

After the chaperone-subunit complex has bound to UsherN and the plug is released from the β-barrel, the chaperone-subunit complex is passed (directly or indirectly) to the C-terminal domain, either during or after DSE (Aminlashgari *et al.* 2012) (Figure 2.1.7b). The exact function of the usher's C-terminal domain in pilus biogenesis is unknown; however, it is crucial in pilus formation since on deletion of the C-terminal domain no pilus biogenesis occurs (Huang *et al.* 2009). Recent studies have, however, suggested a role in subunit discrimination for UsherC (Ahn *et al.* 2004) in addition to a potential role in usher activation (Saulino *et al.* 1998).

Altogether, there are still key questions remaining about how the usher catalyses pilus biogenesis and how it discriminates between different pilus subunits to result in the correct ordering of subunits in the final macromolecular pilus structure. This thesis aims to characterise further the binding of chaperone-subunits to different usher domains from P pilus. Mass spectrometry has been used to detect the different chaperone-subunit-usher interactions and measure any potential differences in their gas phase stabilities. In addition, ion mobility spectrometry-mass spectrometry, in combination with molecular modelling, has been used to try and predict the binding interface of chaperone-subunit-usher interactions that have not been characterised previously by other biophysical methods.

2.2. Results and Discussion

Pilus Assembly



2.2.1 Monitoring Donor-Strand Exchange by ESI-MS

Pilus donor strand exchange has been studied by a variety of biochemical techniques including cation exchange chromatography (Vetsch *et al.* 2006), SDS-PAGE (Sauer *et al.* 2002; Barnhart *et al.* 2003; Remaut *et al.* 2008) and fluorescence based assays (Nishiyama *et al.* 2010). Analysis of the precise mechanistic details of pilus biogenesis by these methods is challenging due to difficulties analyzing these complex multi-component systems. Mass spectrometry is an ideal technique to study protein complexes within a multi-complex system since each component can be separated by its unique *m/z*. Throughout this chapter, an ESI-MS method based on previous ESI-MS experiments (Remaut *et al.* 2006; Rose *et al.* 2008; Rose *et al.* 2008; Verger *et al.* 2008) was used to measure second order rate constants of DSE and to determine factors affecting the rates of DSE during pilus biogenesis.

The $PapD_{his}PapE_{Ntd}$ and $PapD_{his}PapH_{Ntd1}$ complexes used in the first part of this chapter were expressed and purified by a combination of Denis Verger, Gilles Phan, William Allen (Birkeck College, London) and myself (Materials and Methods Section 2.4.8). A typical example of the protein purification for the chaperone-subunit complexes is shown for $PapD_{his}PapE_{Ntd}$ in Figure 2.2.1. All chaperone-subunit complexes were expressed periplasmically (Materials and Methods Section 2.4.8). A single peak eluted off the nickel affinity column in all cases corresponding to a mixture of dissociated $PapD_{his}$ and the non-covalent $PapD_{his}PapE_{Ntd}$ complex (Figure 2.2.1A). Any dissociated $PapD_{his}$ was further removed by cation exchange chromatography (Figure 2.2.1B). The final purity of the chaperone-subunit complex was confirmed by SDS-PAGE (Figure 2.2.1C) and the molecular weight determined using ESI-MS (Materials and Methods Section 2.4.3).

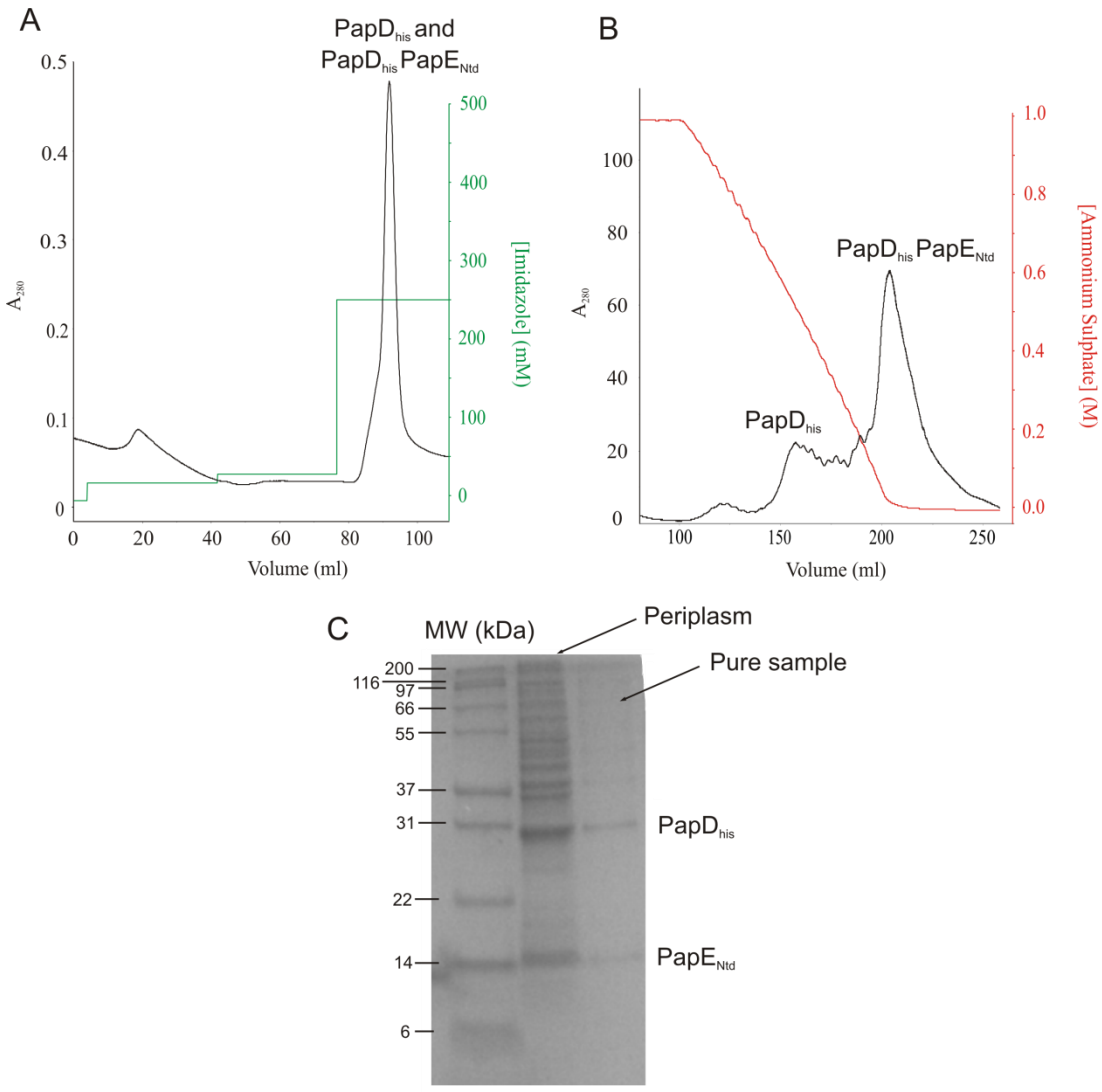


Figure 2.2.1 – Protein purification of the chaperone-subunit complex, $PapD_{his}PapE_{Ntd}$. A) Metal affinity chromatography showing $PapD_{his}PapE_{Ntd}$ and $PapD_{his}$ eluting off the column on increasing the concentration of imidazole; B) Cation exchange chromatography showing on decreasing ammonium sulphate concentration the free chaperone, $PapD_{his}$, is separated from the $PapD_{his}PapE_{Ntd}$ complex; C) SDS-PAGE showing the periplasmic fraction before purification and the final pure $PapD_{his}PapE_{Ntd}$ complex.

ESI-MS conditions were initially optimised using a LCT Premier (nano-ESI Tof MS) for detection of all chaperone-subunit complexes. The LCT Premier offers increased sensitivity over the Q-Tof (previously used in (Remaut *et al.* 2006), (Rose *et al.* 2008) and (Verger *et al.* 2008)) and hence is a more sensitive instrument for analysis of DSE reactions. Optimum pH, buffer concentration, chaperone-subunit concentration and MS ionisation conditions were chosen to minimise salt adduct formation and increase the signal to noise ratio. Additional electron microscopy (EM) experiments were carried out to determine if pilus donor strand exchange is

possible in the volatile buffer used. Pili self-assembled from the intact chaperone-subunit complex PapDPapE$_{V56E}$ were observed on incubation alone overnight at room temperature and a concentration of 20 μM in ammonium acetate buffer at pH 6.0 (the mutation on PapE was designed for other experiments to prevent interaction with the usher and is not thought to affect the DSE properties of PapDPapE) (Figure 2.2.2).

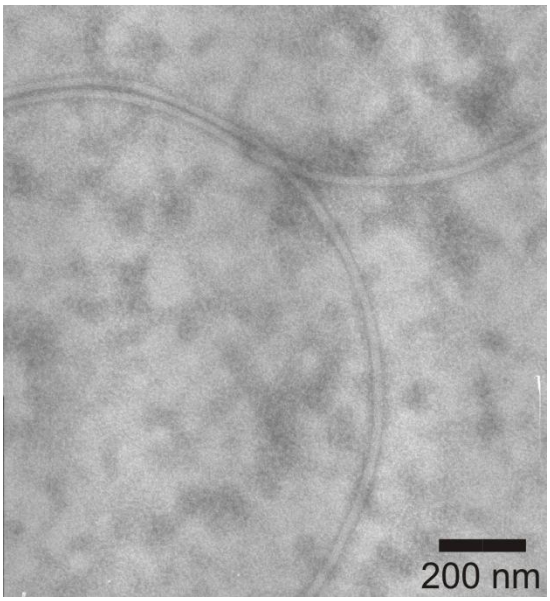


Figure 2.2.2 – Negative stain electron microscopy image of PapDPapE$_{V56E}$ fimbriae formed in 100 mM ammonium acetate pH 6.0, 22 ºC. The scale bar represents 200 nm.

The non-covalent ESI (*m/z*) mass spectrum of the chaperone-subunit complex, PapD$_{his}$PapE$_{Ntd}$, is shown in Figure 2.2.3. Since the PapDPapE chaperone-subunit complex can undergo DSE with itself, a truncated PapE construct was used in which the entire Nte was removed, creating the construct PapE$_{Ntd}$ (Sauer *et al.* 2002) (Section 2.4.8.1). This construct has been shown previously to be a good model for a donor-strand acceptor (Sauer *et al.* 2002; Verger *et al.* 2007). ESI-MS analysis of the PapD$_{his}$PapE$_{Ntd}$ complex showed a narrow charge state distribution suggesting the non-covalent chaperone-subunit complex has been preserved on transition into the gas phase. A small amount of dissociated chaperone, PapD$_{his}$ is also observed which is present in solution (Rebecca Rose – PhD Thesis, University of Leeds, 2008). Very

little $PapE_{Ntd}$ is observed in the *m/z* spectrum suggesting the chaperone, $PapD_{his}$ may ionise preferentially to the subunit $PapE_{Ntd}$ during electrospray ionization.

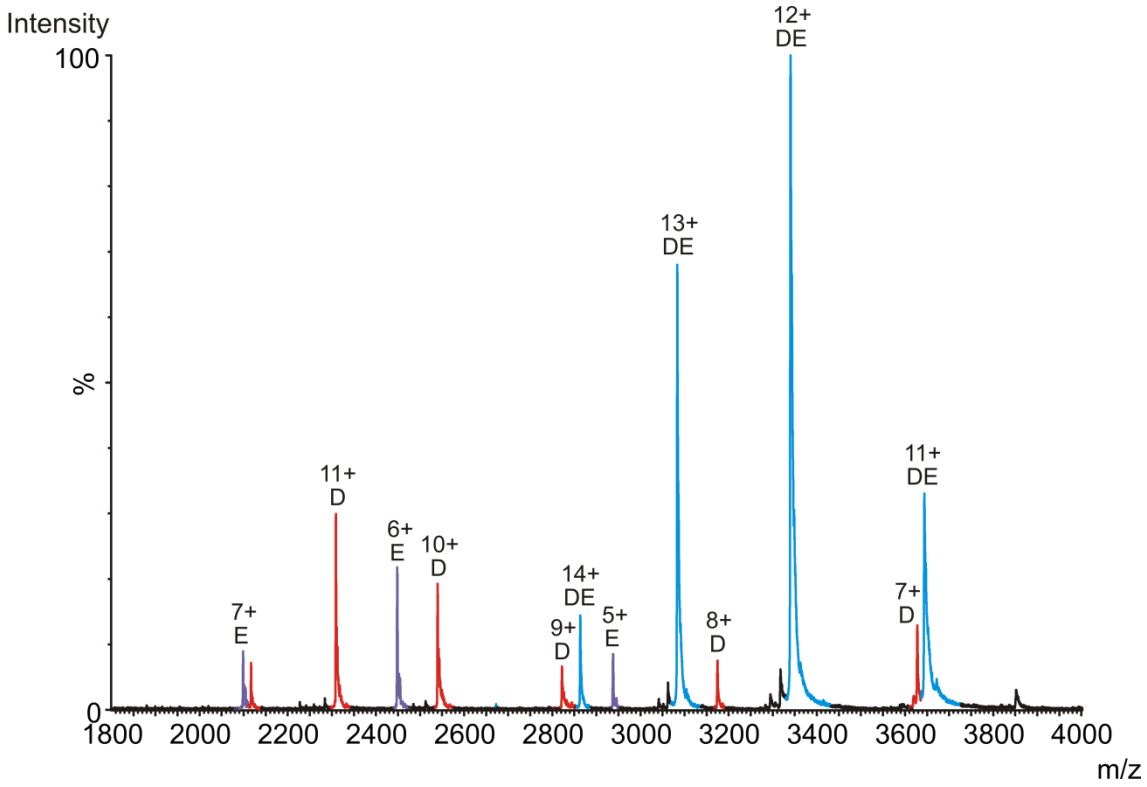


Figure 2.2.3 – ESI-MS spectrum of chaperone-subunit complex, $PapD_{his}PapE_{Ntd}$ in 5 mM ammonium acetate pH 6.0. Peaks corresponding to $PapD_{his}PapE_{Ntd}$ (DE), $PapD_{his}$ (D) and $PapE_{Ntd}$ (E) are shown in blue, red and purple respectively.

To test the new parameters and the different instrument conditions for reproducibility with previous work published on the specificity and pseudo first order rates of donor strand exchange (Rose *et al.* 2008), a donor strand exchange (DSE) reaction between the chaperone-subunit complex, $PapD_{his}PapE_{Ntd}$ and the N-terminal extension of PapE, termed E_{Nte}, was monitored using nanoESI-MS. This chaperone-subunit complex, $PapD_{his}PapE_{Ntd}$, was chosen for the study since it is the fastest chaperone-subunit complex to undergo DSE *in vitro* (Rose *et al.* 2008). The Nte peptide, E_{Nte}, was used to mimic the Nte of the full length subunit, PapE. This assumption is thought to be valid since the subunit's Nte is thought to act independently from the subunit to which it is bound (Lee *et al.* 2004; Lee *et al.*

2007). The sequence used for the E_{Nte} is shown below in Table 2.2.1. Note that two lysine residues were added C-terminally to increase solubility.

Peptide	Peptide Sequence
E_{Nte}	VDNLTFRGKLIIKKamide

Table 2.2.1 – Peptide sequence used to model DSE reaction between chaperone-subunit complex $PapD_{his}PapE_{Ntd}$ and pilin subunit $PapE_{Ntd1}$. KKamide was also added C-terminally to the peptide to increase its solubility.

The ESI-MS spectra aquired 1, 5 and 24 h after initiating DSE at room temperature, in 5 mM ammonium acetate pH 6.0, by addition of peptide E_{Nte} (25 µM) to the chaperone-subunit complex, $PapD_{his}PapE_{Ntd}$ (250 µM) are shown in Figure 2.2.4. As the reaction goes to completion, the peaks corresponding to $PapD_{his}PapE_{Ntd}$ decrease in intensity as the products $PapE_{Ntd}E_{Nte}$ and $PapD_{his}$ are formed. Small amounts of the reaction intermediate, $PapD_{his}PapE_{Ntd}E_{Nte}$, is also observed. A control experiment whereby $PapD_{his}PapE_{Ntd}$ alone was incubated under identical conditions remained the same after 24 h of incubation at room temperature, indicating the changes in the mass spectra in Figure 2.2.4 are due to the DSE reaction between $PapD_{his}PapE_{Ntd}$ and E_{Nte} (data not shown).

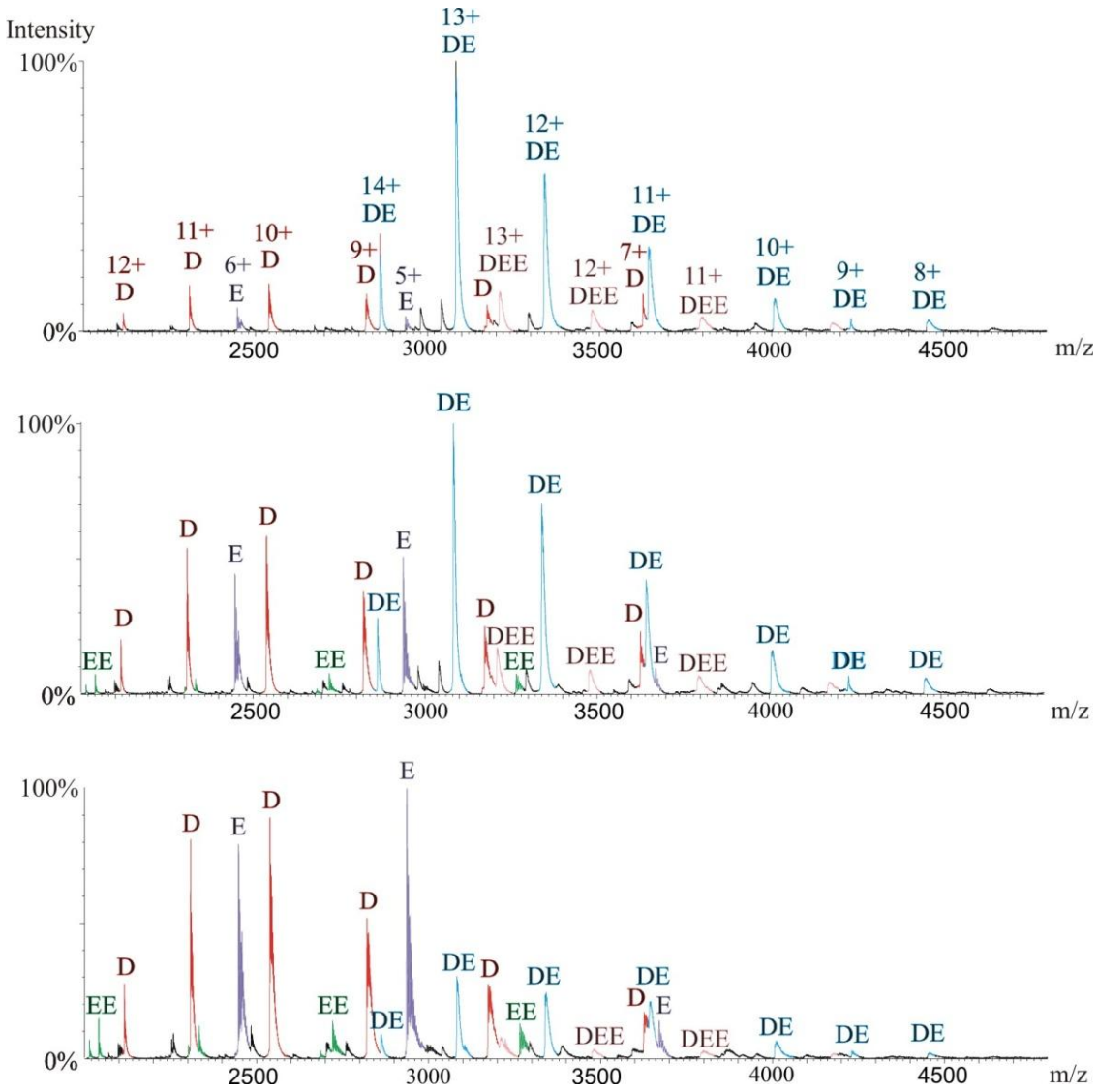


Figure 2.2.4 – Mass spectra of DSE reaction between $PapD_{his}PapE_{Ntd}$ (DE) and E_{Nte} after 1 hour (top), 5 hours (middle) and 24 hours (bottom). $PapD_{his}PapE_{Ntd}$ (DE - blue) decreases with time and $PapD_{his}$ (D - red), $PapE_{Ntd}E_{Nte}$ (EE - green), $PapE_{Ntd}$ (E - purple) increase. Small amounts of $PapD_{his}PapE_{Ntd}E_{Nte}$ intermediate (DEE - pink) also can be observed. The charge states observed for each species are shown in the top spectrum.

Free $PapE_{Ntd}$ is seen to increase in intensity in the mass spectra throughout the reaction time course with E_{Nte} suggesting it is a product of DSE. One possible reason for this is the dissociation of the $PapE_{Ntd}E_{Nte}$ reaction product in the gas phase. Indeed if MS source ionisation conditions are lowered (less energy available for sample ionisation), the $PapD_{his}$ and $PapE_{Ntd}$ dissociation products decrease in intensity and the intensity of the $PapE_{Ntd}E_{Nte}$ product increases (Figure 2.2.5). Even at the lowest ionisation conditions used, cone voltage 30 V, source temperature 30 ºC, ion energy 80 V and capillary voltage 1.6 kV, $PapE_{Ntd}$ is still observed in the

spectrum, hence it is possible that dissociation is occurring in solution. Further experimental evidence is needed for a firm conclusion but experiments were continued based on the supposition that product dissociation occurs in the MS ionisation source, and hence that the dissociated products formed do not affect the rate of DSE. Indeed, in support of this assumption, collision induced dissociation (CID) experiments have shown subunit-subunit complexes to be less stable in the gas phase than all chaperone-subunit complexes (Rebecca Rose – PhD Thesis, University of Leeds, 2008).

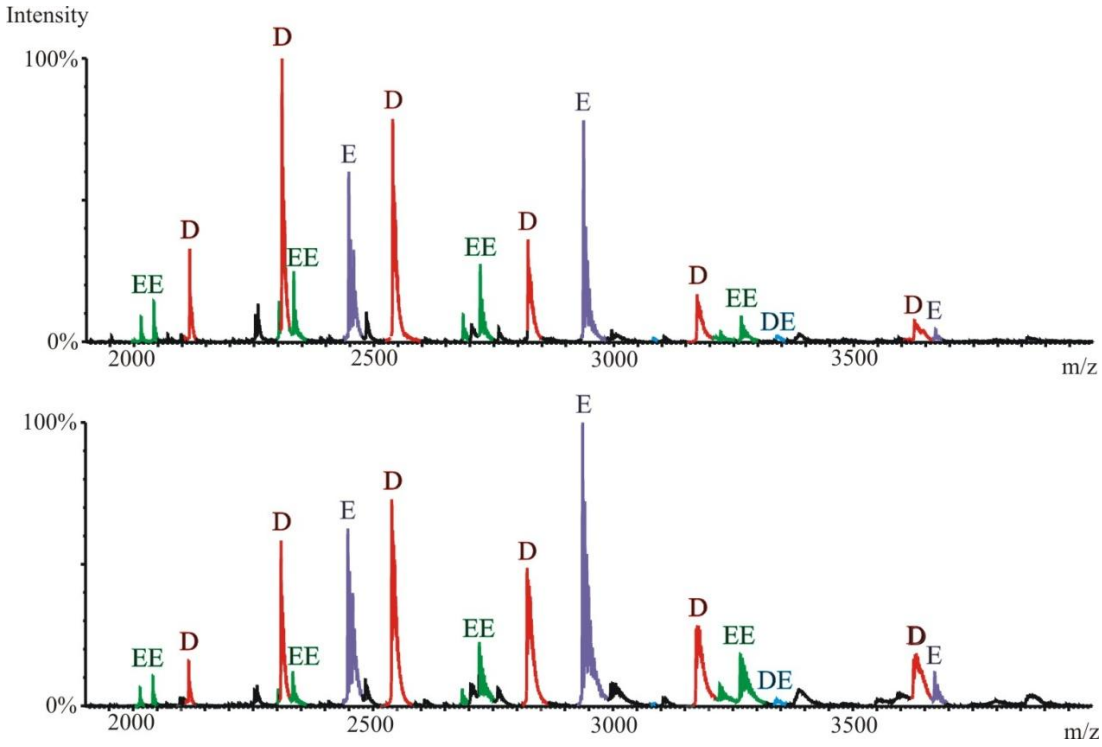


Figure 2.2.5 – Mass spectra of the DSE reaction between $PapD_{his}PapE_{Ntd}$ after 112 hours. Low energy ionisation conditions (cone voltage 30 V, source temperature 30 ºC, ion energy 80 V, capillary voltage 1.6 kV) were used in the top spectrum, higher energy ionisation conditions (cone voltage 70 V, source temperature 60 ºC, ion energy 100 V, capillary voltage 1.75 kV) in the bottom spectrum. The intensity of peaks corresponding to $PapE_{Ntd}E_{Nte}$ (EE, green) increases on lowering the energy of the ESI source conditions, while the intensity of peaks corresponding to $PapD_{his}$ (D, red) and $PapE_{Ntd}$ (E, purple) decrease (compared with the higher energy ESI source conditions). Small amounts of $PapD_{his}PapE_{Ntd}$ (DE, blue) are also shown. Since dissociation of $PapE_{Ntd}E_{Nte}$ occurs in both spectra, the higher energy ionisation conditions were chosen for use in all subsequent MS experiments.

The mass spectral data shown in this thesis are comparable to the ESI-MS results published by Rose *et al*. 2008 indicating the newly optimized MS conditions can be used for comparison with previous pilus DSE MS data (Rose *et al.* 2008).

2.2.2 Determining Second Order Rate Constants by ESI-MS

Previous mass spectrometry methods used to monitor pilus assembly have provided observed rates of the donor strand exchange reaction, only comparable within an experiment (Remaut *et al.* 2006; Rose *et al.* 2008). A method was developed to determine quantitatively second order rate constants for DSE. This method allows rate constants independent of initial reactant concentration to be determined, allowing a more accurate and quantitative comparison of the reactivity of chaperone-subunit complexes with different N-terminal extension peptides.

2.2.3 Quantification of ESI-MS data

The initial concentration of chaperone-subunit and peptide can be measured experimentally (see Sections 2.4.12 and 2.4.3 respectively). Exact quantification by ESI-MS during the DSE reaction is challenging due to the difference in ionisation properties of proteins of comparable molecular weights. Mass spectrometry is an ideal technique for the detection of multiple components in solution allowing the relative change in populations within a complex mixture to be determined. Providing the ion intensity measured directly correlates with concentration, quantitative kinetic analysis by ESI-MS is possible.

To ensure that the relative ion intensity of the chaperone-subunit complexes correlates with chaperone-subunit concentration in solution, and that the concentration used is in the linear range of the detector, an internal standard was added to the reaction mixture. Internal standards are commonly used in mass spectrometry to increase the reproducibility and precision of data between consecutive data acquisitions, and to highlight any instrumental problems occurring. The chaperone-subunit complex, $PapD_{his}PapH_{Ntd1}$, was chosen as an internal standard due to its inability to undergo DSE (Verger *et al.* 2006) (Section 2.2.6). A stable PapH construct, referred to as $PapH_{Ntd1}$, was used whereby the proline-rich N-terminal region of the Nte, known not to be involved in DSE and potentially responsible for insolubility and non-specific aggregation of the subunit, was removed (Verger *et al.* 2006) (Section 2.4.8.1). $PapD_{his}PapH_{Ntd1}$ is an ideal internal

standard since it is structurally similar to all other chaperone-subunit complexes, and hence it can be assumed that its ability to undergo ionisation is similar to that of other chaperone-subunit complexes.

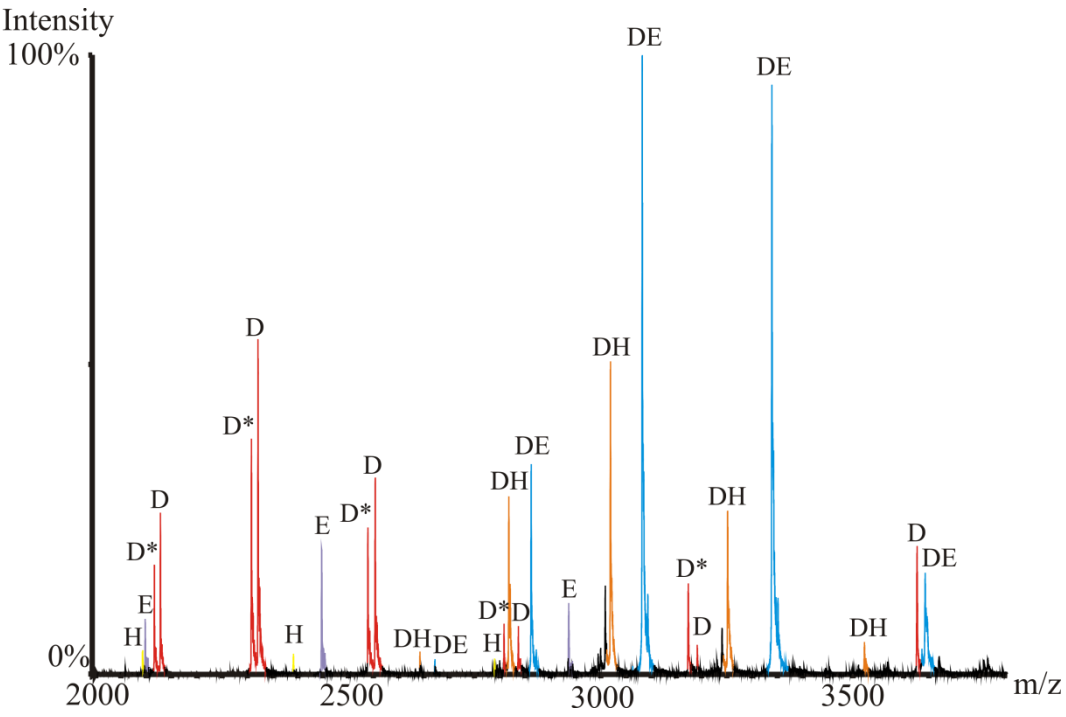


Figure 2.2.6 – Mass spectrum of 4 μM $PapD_{his}PapE_{Ntd}$ (DE, blue) and 2.5 μM $PapD_{his}PapH_{Ntd1}$ (DH, orange) in 5 mM ammonium acetate pH 6.0, room temperature. Dissociated products $PapD_{his}$ (D from $PapD_{his}PapE_{Ntd}$ and D* from $PapD_{his}PapH_{Ntd1}$, red), $PapE_{Ntd}$ (E, purple) and $PapH_{Ntd1}$ (H, yellow) are shown. The difference in molecular weight between D and D* is the addition of a GS linker between the C-terminus of PapD and the his-tag present in the $PapD_{his}PapH_{Ntd1}$ construct.

Mass spectra were acquired for the chaperone-subunit complex, $PapD_{his}PapE_{Ntd}$, at concentrations (0.4 μM-3.6 μM), containing a fixed $PapD_{his}PapH_{Ntd1}$ concentration (4 μM). Figure 2.2.6 shows an example of one of the resulting spectra with charge state distributions corresponding to $PapD_{his}PapE_{Ntd}$ and $PapD_{his}PapH_{Ntd1}$ labelled in blue and orange, respectively. Additional peaks correspond to $PapD_{his}$, $PapE_{Ntd}$ and $PapH_{Ntd1}$ and are a result of $PapD_{his}PapE_{Ntd}$ and $PapD_{his}PapH_{Ntd1}$ dissociation.

The relative ion intensities for the charge states corresponding to the chaperone-subunit complexes $PapD_{his}PapE_{Ntd}$ and $PapD_{his}PapH_{Ntd1}$, respectively, were summed and plotted as a ratio (relative protein response) against $PapD_{his}PapE_{Ntd}$

concentration (Figure 2.2.7). The data are linear over the concentration range used. Data for $PapD_{his}PapE_{Ntd}$ concentrations below 0.75 μM deviate slightly from the line due to inaccuracies in measuring peak areas when the signal reaches the noise level in the instrument. Data obtained for chaperone-subunit concentrations below 0.75 μM, therefore, were discarded in future DSE experiments.

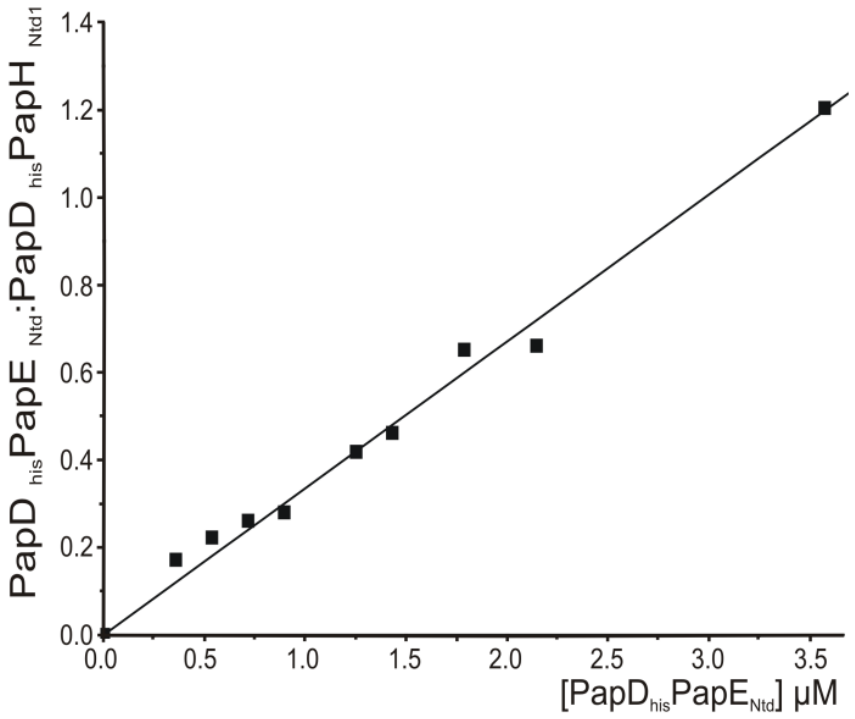


Figure 2.2.7 – Linearity of relative protein response as a function of concentration. The intensity of chaperone-subunit ($PapD_{his}PapE_{Ntd}$) peaks in the mass spectrum relative to the intensity of the internal standard ($PapD_{his}PapH_{Ntd1}$) as a function of the chaperone-subunit ($PapD_{his}PapE_{Ntd}$) concentration is plotted. The conditions of the experiment are as shown in Figure 2.2.5.

One common problem in ESI-MS is ion suppression. Ion suppression is caused when less volatile species change the efficiency of droplet formation or evaporation, which in turn prevents less charged ions entering the gas phase and reaching the detector (Annesley 2003). Concern was raised as to whether the large excess of peptide added in DSE experiments would affect the relative ion intensity of $PapD_{his}PapH_{Ntd1}$ and $PapD_{his}PapE_{Ntd}$, and whether their ionisation levels would be significantly reduced to levels close to the noise in the instrument.

To test this possibility, the intensity ratio of $PapD_{his}PapE_{Ntd}$ to the $PapD_{his}PapH_{Ntd1}$ was monitored on addition of various concentrations of raffinose. Raffinose, a trisaccharide composed of glucose, galactose and fructose (Figure 2.2.8), is commonly used as an internal standard in MS. Raffinose was used since it ionises easily by ESI, it does not interact with any chaperone-subunit complex, and most importantly it is small enough in molecular weight to avoid analysis problems such as overlapping peaks in the mass spectra.

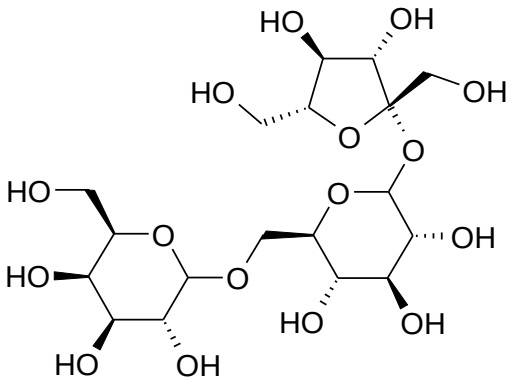


Figure 2.2.8 – Raffinose structure (average molecular weight = 504.42 Da) used to monitor ion suppression during ESI-MS.

Mass spectra of $PapD_{his}PapE_{Ntd}$ and $PapD_{his}PapH_{Ntd1}$ on addition of raffinose at concentrations 15 µM and 30 µM are shown in Figure 2.2.9. Peaks corresponding to $PapD_{his}PapE_{Ntd}$ (DE, blue) and $PapD_{his}PapH_{Ntd1}$ (DH, orange) remain at a constant ion intensity ratio of 1:1 independent of raffinose concentration. The total ion count (TIC) for each spectral acquisition also remained approximately constant (Figure 2.2.9), proving the addition of raffinose does not suppress the ionisation of $PapD_{his}PapE_{Ntd}$ or $PapD_{hishis}PapH_{Ntd1}$.

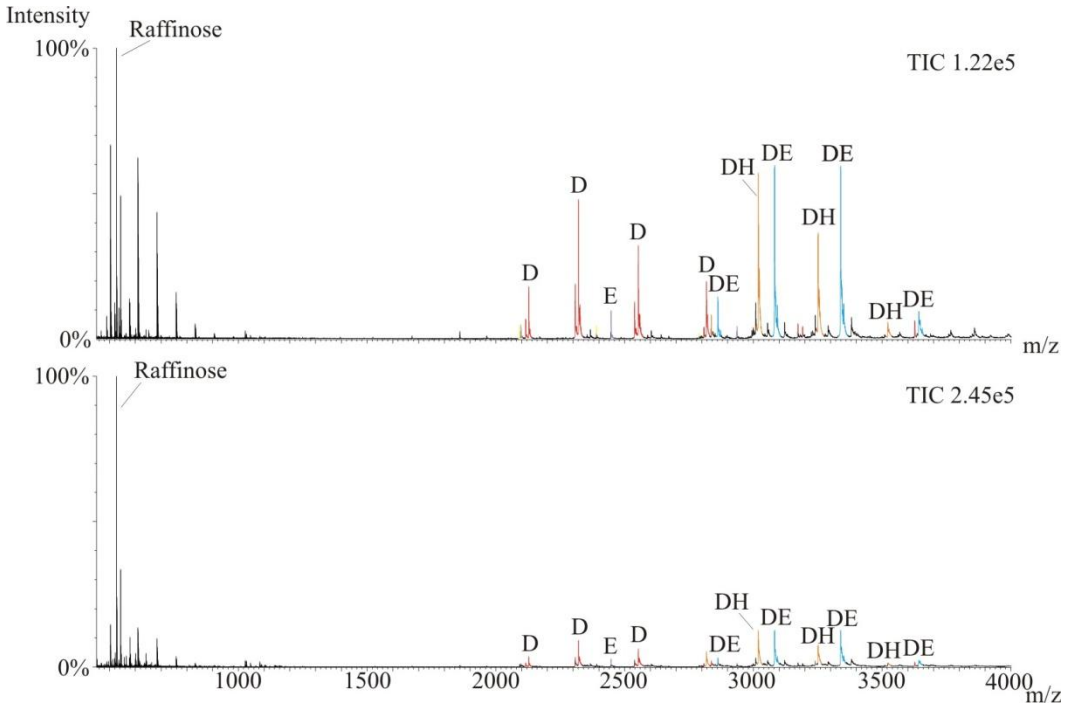


Figure 2.2.9 – Mass spectra for 2.5 μM $PapD_{his}PapE_{Ntd}$ (DE – blue) and 2.5 μM $PapD_{his}PapH_{Ntd1}$ (DH – orange) on addition of raffinose at concentrations 15 μM (top) and 30 μM (bottom). Raffinose peak labelled corresponds to its sodium adduct (*m/z* 527). All peaks are colour coded as in Figure 2.2.6. TIC = total ion count (aquired over one minute in each case).

Addition of raffinose concentrations up to 50 μM did not affect the $PapD_{his}PapE_{Ntd}$ to $PapD_{his}PapH_{Ntd1}$ ratio (Table 2.2.2). Altogether, these ESI-MS experiments show that chaperone-subunit concentration can be measured quantitatively by ESI-MS by comparison of the chaperone-subunit intensity of interest with the internal standard ($PapD_{his}PapH_{Ntd1}$), if the correct experimental conditions are used.

Raffinose Concentration (μM)	$PapD_{his}PapE_{Ntd}$ intensity / $PapD_{his}PapH_{Ntd1}$ intensity
3.18	0.96
3.75	0.92
6.25	0.92
7.50	0.89
12.5	0.91
15.0	1.02
25.0	0.96
30.0	0.98
50.0	1.03

Table 2.2.2 - The effect of increasing raffinose concentration on the $PapD_{his}PapE_{Ntd}$ to $PapD_{his}PapH_{Ntd1}$ ratio. The intensity ratio is constant within 5% error, suggesting addition of a large excess of peptide with not effect the method for determination of chaperone-subunit concentrations. Note, a constant concentration of 2.5 μM was maintained for both chaperone-subunit complexes in this experiment (5 mM ammonium acetate pH 6.0, room temperature).

2.2.4 Derivation of second order rate constants

Since the chaperone-subunit concentration can now be determined quantitatively throughout the DSE reaction through addition of an internal standard, $PapD_{his}PapH_{Ntd1}$, the second order rate constant rather than observed rate (determined from the relative loss of chaperone-subunit ions compared to all other ions in the mass spectrum) can be determined.

The derived equation for the DSE rate constant was based on the assumption that one chaperone-subunit complex, PapDPapX, reacts with one Nte of an incoming subunit, X_{Nte}, to form a subunit-Nte complex, PapX-X_{Nte}, via a ternary intermediate, PapDPapX-X_{Nte} (Equation 2.1).

$$\mathrm{PapDPapX} + \mathrm{X_{Nte}} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{PapDPapX - X_{Nte}} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \mathrm{PapD} + \mathrm{PapX - X_{Nte}} \qquad \text{Equation 2.1}$$

For simplicity, the backward reaction rate constant, k_{-2}, was assumed to be negligible. This assumption is valid since chaperone-subunit complexes undergo a conformational change to a more compact, thermodynamically stable structure on transition to the subunit-peptide complex, causing the release of energy needed to drive pilus assembly (Jones *et al.* 1993; Zavialov *et al.* 2003; Zavialvo *et al.* 2005; Vetsch *et al.* 2006). In previous studies a transient chaperone-subunit-Nte intermediate was detected in the Saf pilus system (Remaut *et al.* 2006), a system closely related to Pap. In P pilus, this intermediate does not accumulate during the DSE reaction of P pili and hence is rarely observed in the mass spectra of DSE reactions (Rose *et al.* 2008). Once this PapDPapX-X_{Nte} intermediate complex forms, DSE must therefore occur rapidly in the Pap system suggesting that the binding of the Nte to the chaperone-subunit complex (k_1) is the rate limiting step in the DSE reaction of P pili. Based on these assumptions, Equation 2.1 can be fitted to a second order rate equation from which the rate constants, k_1 and k_{-1}, can be determined (Equation 2.2). For this study, since no or very little DSE intermediate was observed (Rose *et al.* 2008) it can be assumed from Equation 2.1 that $k_2 >> k_{-1}$, and hence k_{-1}[PapDPapX-X_{Nte}] can be approximated to zero (Equation 2.3).

$$\frac{-\mathrm{d}[\mathrm{PapDPapX}]}{\mathrm{dt}} = \mathrm{k}_1[\mathrm{PapDPapX}][\mathrm{X_{Nte}}] + \mathrm{k}_{-1}[\mathrm{PapDPapX}{-}\mathrm{X_{Nte}}] \qquad \text{Equation 2.2}$$

$$\frac{-\mathrm{d}[\mathrm{PapDPapX}]}{\mathrm{dt}} = \mathrm{k}_1[\mathrm{PapDPapX}][\mathrm{X_{Nte}}] \qquad \text{Equation 2.3}$$

The above equation (Equation 2.3) can be integrated and rearranged in terms of chaperone-subunit concentration, [PapDPapX] (Equation 2.5) based on Equation 2.4 that states that the decrease in the concentration of the chaperone-subunit complex is equal to the loss of Nte peptide over time, since one chaperone-subunit reacts with one peptide during DSE (See Appendix 1 for full derivation).

$$[\mathrm{X_{Nte}}]_t = [\mathrm{X_{Nte}}]_0 - ([\mathrm{PapDPapX}]_0 - [\mathrm{PapDPapX}]_\mathrm{t}) \qquad \text{Equation 2.4}$$

where $[\mathrm{X_{Nte}}]_0$ is the initial concentration and $[\mathrm{X_{Nte}}]_t$ is the concentration at time t.

$$[\mathrm{PapDPapX}]_\mathrm{t} = \mathrm{y}\left\{\frac{[\mathrm{PapDPapX}]_0 \mathrm{e}^{-10^{\log \mathrm{k}_1([\mathrm{X_{Nte}}]_0-[\mathrm{PapDPapX}]_0)\mathrm{t}}} - \left(\frac{[\mathrm{PapDPapX}]_0{}^2 \mathrm{e}^{-10^{\log \mathrm{k}_1([\mathrm{X_{Nte}}]_0-[\mathrm{PapDPapX}]_0)\mathrm{t}}}}{[\mathrm{X_{Nte}}]_0}\right)}{1 - \left(\frac{[\mathrm{PapDPapX}]_0 \mathrm{e}^{-10^{\log \mathrm{k}_1([\mathrm{X_{Nte}}]_0-[\mathrm{PapDPapX}]_0)\mathrm{t}}}}{[\mathrm{X_{Nte}}]_0}\right)}\right\} \qquad \text{Equation 2.5}$$

where y is the dilution factor with the internal standard, $\mathrm{PapD_{his}PapH_{Ntd1}}$, prior to mass spectral acquisition.

All subsequent DSE experiments where the loss of chaperone-subunit complex with time is quantified can be fitted, therefore, to Equation 2.5 to determine an accurate second order rate constant, k_1, for the DSE reaction studied.

2.2.5 Experimental determination of second order rate constants using MS

To determine the rate constant of the DSE reaction between $PapD_{his}PapE_{Ntd}$ and E_{Nte}, ESI mass spectra were acquired over a time course of 7 h on addition of E_{Nte} to $PapD_{his}PapE_{Ntd}$. A constant concentration of $PapD_{his}PapH_{Ntd1}$ was added 30 s prior to each spectral acquisition, to ensure all non-specific interactions between $PapD_{his}PapH_{Ntd1}$ and species in DSE reaction solution were kept to a minimum.

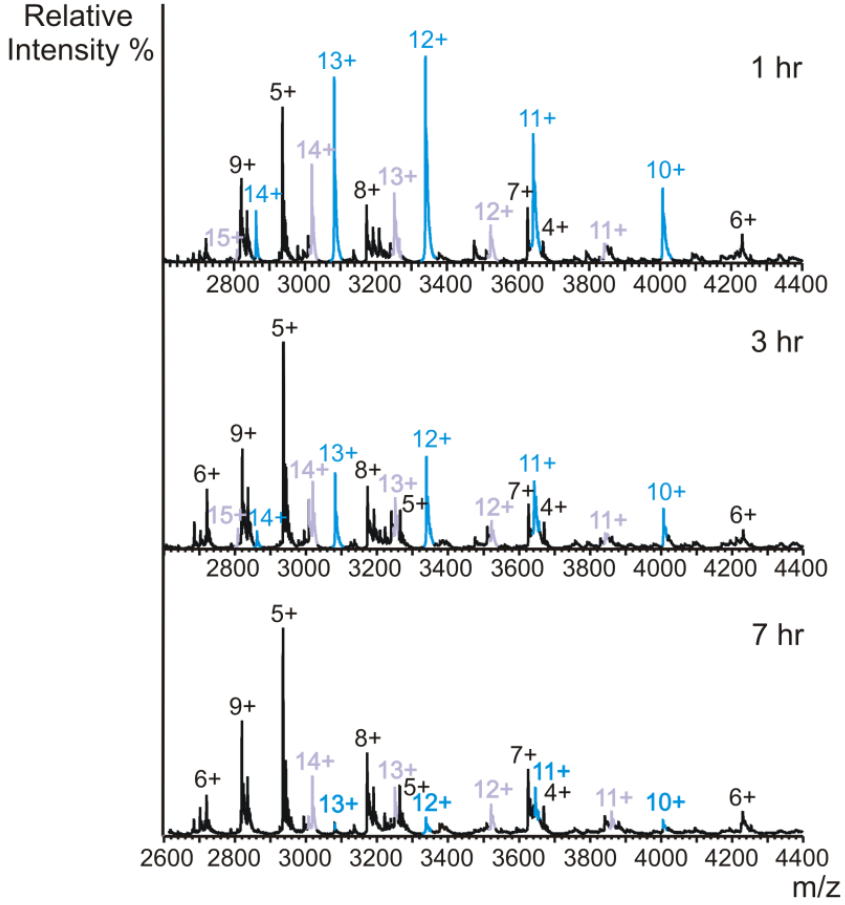


Figure 2.2.10 – ESI mass spectra of the DSE reaction between $PapD_{his}PapE_{Ntd}$ and E_{Nte} after 1 h (top), 3 h (middle) and 7 h (bottom). The charge state series of ions corresponding to $PapD_{his}PapE_{Ntd}$, the internal standard $PapD_{his}PapH_{Ntd1}$, and the DSE products ($PapE_{Ntd}$-E_{Nte}, $PapD_{his}$ and $PapE_{Ntd}$) are shown in blue, lilac and black respectively. The rate constant was measured from the decrease in intensity of the 11+, 12+, and 13+ ions originating from $PapD_{his}PapE_{Ntd}$ compared with the 12+, 13+ and 14+ ions of the internal standard.

Initially, the predominant species are the chaperone-subunit complex, $PapD_{his}PapE_{Ntd}$, and the internal standard, $PapD_{his}PapH_{Ntd1}$ (Figure 2.2.10). Small amounts of dissociated chaperone-subunit complex are also observed, equivalent to

less than 5% of the total protein concentration (determined by native PAGE, Rebecca Rose - PhD Thesis, University of Leeds, 2008). On addition of peptide Nte, the chaperone-subunit peaks decrease in intensity and the peaks relating to the chaperone and subunit-Nte DSE products increase relative to the internal standard, $PapD_{his}PapH_{Ntd1}$. The theoretical and experimental masses for all species observed are accurate to ± 0.05 % and are listed in the Materials and Methods Section 2.4.14.3, Table 2.4.3. Since the internal standard intensity is assumed to remain constant throughout the DSE reaction, the DSE reaction rate was determined in real time as the DSE reaction proceeds, by fitting the decrease in $PapD_{his}PapE_{Ntd}$:$PapD_{his}PapH_{Ntd1}$ ratio over time to the DSE rate equation (Equation 2.5) (Figure 2.2.11).

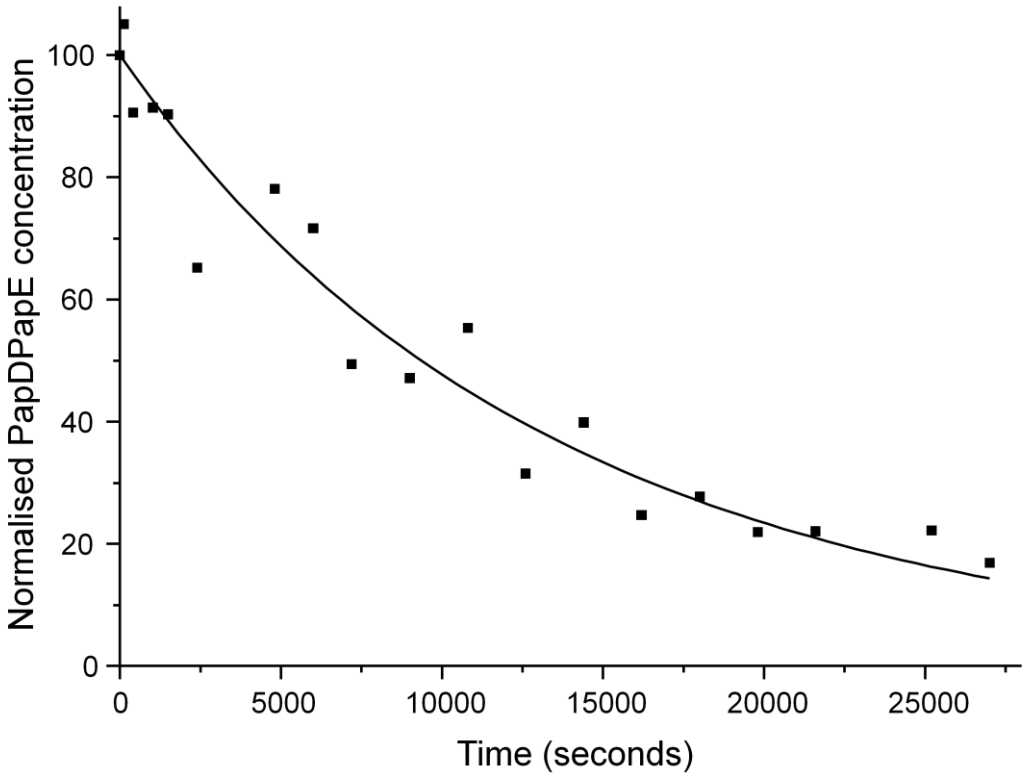


Figure 2.2.11 – The decrease of the normalized $PapD_{his}PapE_{Ntd}$ concentration (determined from the relative ion intensity of $PapD_{his}PapE_{Ntd}$:$PapD_{his}PapH_{Ntd1}$) over time during a DSE reaction with E_{Nte} peptide. Only one repeat of the data is shown for simplicity, although to demonstrate reproducibility three repeats were performed and the errors between replicates reported in Table 2.2.3.

The second order rate constant for the DSE reaction between $PapD_{his}PapE_{Ntd}$ and the peptide E_{Nte} was determined to be 0.32 $M^{-1}s^{-1}$ (Table 2.2.3). This is in reasonable agreement with the rate of 0.03 $M^{-1}s^{-1}$ reported by others for Fim pilus DSE using

cation exchange chromatography methods (Vetsch *et al.* 2006), taking into account for the difference in pilus system and that a ten-fold rate difference can be observed between different chaperone-subunit complexes and their corresponding cognate Ntes (Rose *et al.* 2008). The quantitative nanoESI-MS method described thus allows determination of second order rate constants for complex reactions with ease and an accuracy which can enable detailed analysis of the role of individual residues in controlling the rate and specificity of pilus assembly. On comparison to the previous method of DSE analysis by ESI-MS (Rose *et al.* 2008) a shorter time scale is also used. This reduces problems with chaperone-subunit dissociation, solvent evaporation and peak broadening which inevitably increase with time.

2.2.6 The P5 residue in DSE is important in determining subunit specificity

Previously, by using E_{Nte} and H_{Nte}-derived peptides in which the C-terminal five residues were swapped, it was shown that the C-terminal half of the Nte plays a crucial role in determining DSE rates (Rose *et al.* 2008). Based on studies of Saf DSE, it is thought that the P5 residue on the Nte fits tightly into the P5 pocket to initiate the unzipping of the chaperone's G_1 strand of the chaperone from the chaperone-subunit complex (Verger *et al.* 2008). On analysis of the crystal structure of the PapDPapH chaperone-subunit complex, the P5 pocket was found to be closed, explaining why PapDPapH does not undergo DSE and hence its role in terminating pilus biogenesis by anchoring the final pilus to the outer membrane (Verger *et al.* 2008). It was predicted, therefore, that on substituting the P5 residue of E_{Nte} with a smaller residue (i.e. one that does not fit as well into the P5 pocket) the DSE rate would decrease. To investigate this possibility, the synthetic peptide equivalent to the N-terminal 12 residues of PapE (E_{Nte}) was resynthesised where the P5 residue was altered from leucine to valine or alanine (sequences shown in Figure 2.2.12), and the rate constant of DSE determined using the ESI-MS approach described above.

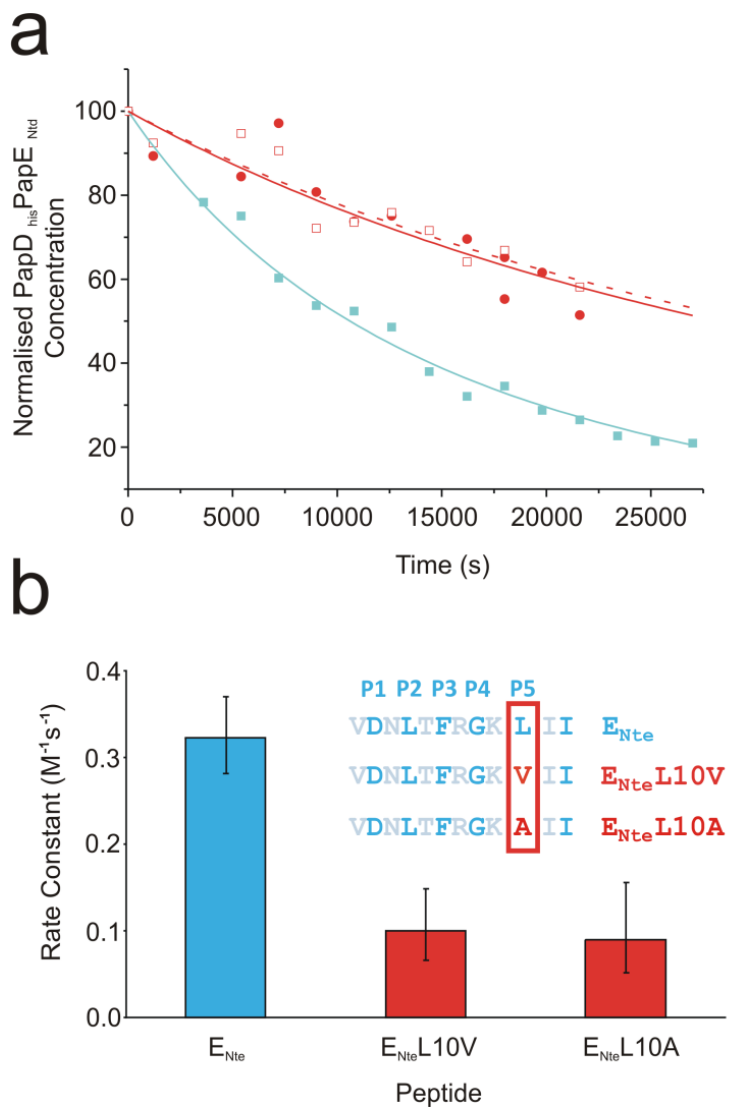


Figure 2.2.12 a) The decrease of the normalized $PapD_{his}PapE_{Ntd}$ concentration (determined from the relative ion intensity of $PapD_{his}PapE_{Ntd}$:$PapD_{his}PapH_{Ntd1}$) over time during a DSE reaction with peptides E_{Nte} (blue), E_{Nte}L10V (red solid line) and E_{Nte}L10A (red dashed line). b) DSE rate constants for the same DSE reactions showing the standard error between triplicate measurements. The insert shows the peptide sequences used with the alternate hydrophobic residues in bold font and the residue substituted highlighted in red.

The results show that substituting the Nte P5 residue to smaller residues decreases the rate of DSE significantly from 0.32 $M^{-1}s^{-1}$ for the wild-type E_{Nte} sequence to 0.10 $M^{-1}s^{-1}$ and 0.09 $M^{-1}s^{-1}$ for the valine and alanine substituted peptides, respectively (Figure 2.2.12, Table 2.2.3). As predicted, the P5 residue fitted less tightly into the subunit's P5 pocket, therefore, making DSE less likely to occur and thus the DSE rate constant was reduced for the DSE reactions between $PapD_{his}PapE_{Ntd}$ and the P5 substiuted peptides. The DSE rate constants for $PapD_{his}PapE_{Ntd}$ reaction with E_{Nte}L10V and E_{Nte}L10A are comparable suggesting that once the P5 residue is small

enough not to fit tightly into the P5 pocket, making it even smaller does not affect the DSE rate further. This was surprising since a clearer effect on DSE rate was observed for the SafA-SafB complex on the altering the size of the P5 residue (Remaut *et al.* 2006). Altogether, the results support the view that the P5 residue of the Nte in the incoming subunit is a critical determinant of the rate of DSE.

Peptide	Residue Substituted	Rate Constant, logk ($\mu M^{-1}s^{-1}$)	Rate Constant, k ($M^{-1}s^{-1}$)
E_{Nte}	None	-6.49 ± 0.06	0.32 (+ 0.05, - 0.04)
E_{Nte}L10V	P5	-7.00 ± 0.24	0.10 (+ 0.05, - 0.03)
E_{Nte}L10A	P5	-7.05 ± 0.17	0.09 (+ 0.07, - 0.04)

Table 2.2.3 - Second order rate constants (pH 6.0, 22 ºC) between $PapD_{his}PapE_{Ntd}$ and the substituted peptides E_{Nte}L10V and E_{Nte}L10A. Rate constants are displayed in terms of logk and the derived rate constant, k. The standard error between triplicate DSE reactions are shown.

2.2.7 The P5+1 and P5-1 residues modulate the rate of DSE

The complementarity between subunit grooves and Ntes that imparts specificity to DSE, is thought to contribute to the correct incorporation of subunits into the growing pilus fibre (Rose *et al.* 2008). The C-terminal half of the Nte (P5-1, P5, P5+1, P5+2 residues) has been shown to be the more important in determining the rate of DSE, with docking of the P5 residue into the P5 pocket initiating DSE (Rose *et al.* 2008; Verger *et al.* 2008). The role of the P5-1 and P5+1 residues in DSE, however, has not been determined. Although the side-chains of the P5+1 and P5-1 residues are directed away from the subunit's hydrophobic groove in the crystal structure of the DSE product (Figure 2.2.13) (Sauer *et al.* 2002), these residues, nonetheless, could play a role in the progress of the DSE reaction and thus be important determinants of subunit specificity.

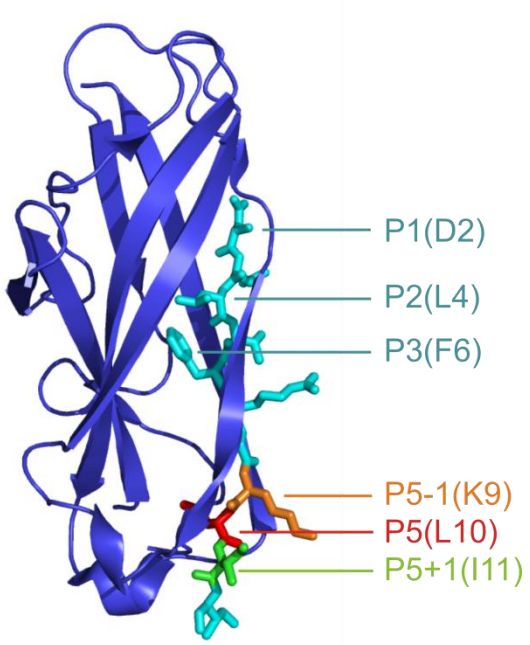


Figure 2.2.13 - Structure of the subunit-Nte DSE product, $PapE_{Ntd}E_{Nte}$ (blue-pale blue). The structure was modelled based on the $PapE_{Ntd}K_{Nte}$ crystal structure (PDB 1N12) (Sauer *et al.* 2002) with residues from K_{Nte} substituted with the corresponding residues in E_{Nte}. The P5-1, P5 and P5+1 residues are highlighted in orange, red and green, respectively.

To investigate the potential role of the P5-1 and P5+1 residues in determining the rate of DSE, E_{Nte} peptides were designed in which these residues were substituted individually with alanine in the variants E_{Nte}K9A and E_{Nte}I11A, respectively (Figure 2.2.13) and the rate of DSE with $PapD_{his}PapE_{Ntd}$ was determined using these peptides by the ESI-MS approach described above. Substituting at the P5+1 position to alanine had a surprising effect on the rate constant of DSE with $PapD_{his}PapE_{Ntd}$ with a 3-fold decrease in second order rate constant observed from 0.32 $M^{-1}s^{-1}$ for the wild type E_{Nte} to 0.12 $M^{-1}s^{-1}$ on substitution from isoleucine to alanine at the P5+1 position (Figure 2.2.14, Table 2.2.4). Remarkably, this difference in rate constant is similar in magnitude to that of the valine or alanine substitution at the P5 residue, suggesting that the P5+1 and the P5 residues are equally important in determining the rate of DSE for this subunit:Nte pair (Table 2.2.4). In comparison, substitution at the P5-1 position decreased the rate of DSE only slightly (from 0.32 $M^{-1}s^{-1}$ for wild-type to 0.22 $M^{-1}s^{-1}$ for E_{Nte}K9A) (Figure 2.2.14, Table 2.2.4) implying that the P5-1 residue has a relatively minor effect in determining the rate of DSE.

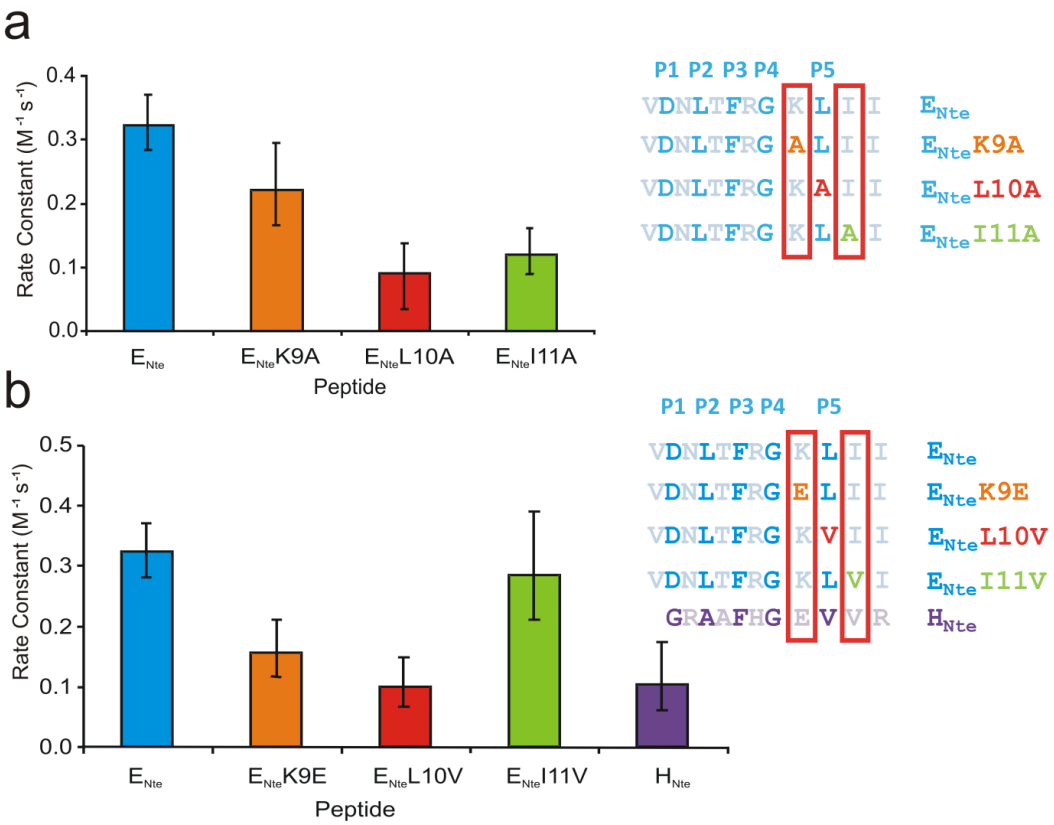


Figure 2.2.14 a) DSE rate constants for the DSE reaction between $PapD_{his}PapE_{Ntd}$ and the peptides E_{Nte} (blue), E_{Nte}K9A (orange), E_{Nte}L10A (red) and E_{Nte}I11A. b) DSE rate constants for the DSE reaction between $PapD_{his}PapE_{Ntd}$ and the peptides E_{Nte} (blue), E_{Nte}K9E (orange), E_{Nte}L10V (red), E_{Nte}I11V and H_{Nte} (purple). The standard error between triplicate measurements is shown in all cases. The inserts show the peptide sequences used with the alternate hydrophobic residues in bold font and the residues substituted highlighted in orange, red and green for the P5-1, P5 and P5+1 residues, respectively.

Peptide	Residue Substituted	Rate Constant, logk ($\mu M^{-1}s^{-1}$)	Rate Constant, k ($M^{-1}s^{-1}$)
E_{Nte}	None	-6.49 ± 0.06	0.32 (+ 0.05, - 0.04)
E_{Nte}K9A	P5-1	-6.66 ± 0.13	0.22 (+ 0.07, - 0.06)
E_{Nte}L10A	P5	-7.05 ± 0.17	0.09 (+ 0.07, - 0.04)
E_{Nte}I11A	P5+1	-6.92 ± 0.13	0.12 (+ 0.04, - 0.03)

Table 2.2.4 - Second order rate constants (pH 6.0, 22 ºC) between $PapD_{his}PapE_{Ntd}$ and the substituted peptides E_{Nte}K9A, E_{Nte}L10A and E_{Nte}I11A. Rate constants are displayed in terms of logk and the derived rate constant, k. The standard error between triplicate DSE reactions are shown.

As the residues immediately C-terminal to the P5 residue (P5+1) in E_{Nte} were found to influence the DSE rate, we next investigated the effect of these residues on DSE in $PapD_{his}PapE_{Ntd}$ by substituting residues from the non-cognate H_{Nte} sequence into the sequence of E_{Nte} at the P5-1, P5 and P5+1 positions (Figure 2.2.14b). This was

performed in order to aid our understanding of the role of the residues within the Nte in determining DSE specificity. Indeed if residues at the P5-1, P5 and P5+1 sites are important in modulating the rate of DSE, the DSE rate constant with these substituted peptides would be expected to decrease substantially, possibly to that of slowest $PapD_{his}PapE_{Ntd}$/Nte DSE reaction i.e., that with its non-cognate peptide (H_{Nte}) (Rose *et al.* 2008). Second order rate constants for the DSE reaction between $PapD_{his}PapE_{Ntd}$ and the cognate (E_{Nte}) and non-cognate (H_{Nte}) peptides showed a decrease in rate constant from 0.32 $M^{-1}s^{-1}$ for the cognate reaction to 0.10 $M^{-1}s^{-1}$ for the non-cognate reaction (Table 2.2.5). This is comparable to previous data that showed non-cognate DSE reactions, i.e. DSE reaction between chaperone-subunit:Nte pairs that are not adjacent to each other in the final pilus structure, are much slower than the equivalent DSE reaction that occurs *in vivo* (the cognate reaction) (Rose *et al.* 2008). Substitution of the P5 residue from leucine (in E_{Nte}) to valine (in H_{Nte}) in the variant E_{Nte}L10V demonstrates that replacement of a critical residue at the DSE initiation site reduces the rate constant of DSE substantially, to a rate equivalent to that observed for the reaction with H_{Nte} (Figure 2.2.14 and Table 2.2.5). Strikingly, substitution from lysine (the P5-1 residue in E_{Nte}) to glutamic acid (the equivalent residue in H_{Nte}) in the variant E_{Nte}K9E also decreased the second order rate constant substantially, to a rate comparable with that with H_{Nte} (Figure 2.2.13 and Table 2.2.5). In comparison, substitution of isoleucine to valine at the P5+1 position E_{Nte}I11V had very little effect on the DSE rate constant, with only a minor decrease from 0.32 $M^{-1}s^{-1}$ for E_{Nte} to 0.29 $M^{-1}s^{-1}$ with E_{Nte}I11V (Figure 2.2.14 and Table 2.2.5).

Peptide	**Residue Substituted**	**Rate Constant, logk ($\mu M^{-1}s^{-1}$)**	**Rate Constant, k ($M^{-1}s^{-1}$)**
E_{Nte}	None	-6.49 ± 0.06	0.32 (+ 0.05, - 0.04)
E_{Nte}K9E	P5-1	-6.80 ± 0.13	0.16 (+ 0.06, - 0.04)
E_{Nte}L10V	P5	-7.05 ± 0.17	0.09 (+ 0.07, - 0.04)
E_{Nte}I11V	P5+1	-6.54 ± 0.13	0.29 (+ 0.10, - 0.08)
H_{Nte}	None	-6.98 ± 0.22	0.10 (+ 0.07, - 0.04)

Table 2.2.5 - Second order rate constants (pH 6.0, 22 ºC) between $PapD_{his}PapE_{Ntd}$ and the substituted peptides E_{Nte}K9E, E_{Nte}L10V, E_{Nte}I11V and H_{Nte}. Rate constants are displayed in terms of logk and the derived rate constant, k. The standard error between triplicate DSE reactions are shown.

The H_{Nte} substituted peptides show differences in the relative importance of the P5-1 and P5+1 sites compared with those of the alanine scan (Figure 2.2.14A, B), where the P5+1 substitution had a greater effect on DSE than the P5-1 substitution. The different substitutions at the P5-1 and P5+1 position in E_{Nte} show the roles of the P5+1 and P5-1 residues are dependent on the precise nature of the side-chain introduced. Electrostatic interactions at the P5-1 position are thought to play a role in DSE initiation since a larger effect on the DSE rate constant of the E_{Nte}K9E substitution was observed compared with the E_{Nte}K9A substitution at the P5-1 position (Table 2.2.4 and 2.2.5). A trend comparable to that observed at the P5 position was observed on decreasing the size of the amino acid side chain at the P5+1 site from isoleucine (in E_{Nte}) to valine (the equivalent residue in H_{Nte}), or further to alanine, showing smaller amino acids substituted at this P5+1 position decrease the DSE rate constant with $PapD_{his}PapE_{Ntd}$ (Tables 2.2.4, 2.2.5), (Figure 2.2.12, Table 2.2.3). This decrease in side-chain volume may prevent hydrophobic surface interactions taking place which are needed to dock the Nte onto the chaperone-subunit complex before the chaperone's G_1 strand can unzip from the subunit's hydrophobic groove in the chaperone-subunit complex.

2.2.8 P5+1 and P5-1 residues alone cannot restore the fast kinetics of E_{Nte}

Previous experiments using chimeric peptides in which the N-terminal seven residues of one peptide Nte is fused to the C-terminal five residues of another peptide's Nte demonstrated that the C-terminal region of the Nte alone determines the rate of DSE (Rose *et al.* 2008). To investigate further the extent of the P5+1 and P5-1 residues' roles in determining the rate of DSE with $PapD_{his}PapE_{Ntd}$, E-H_{Nte} chimeric peptides were used which undergo very slow rates of DSE (Figure 2.2.15) (Rose *et al.* 2008). E_{Nte} residues were substituted individually into the E-H_{Nte} peptide at the P5-1, P5 or P5+1 positions with the aim of increasing the rate of DSE to mirror that obtained between the $PapD_{his}PapE_{Ntd}$ and E_{Nte} reaction. The control DSE reaction with $PapD_{his}PapE_{Ntd}$ and the peptide E-H_{Nte} showed a decrease in rate constant from 0.32 $M^{-1}s^{-1}$ for the cognate DSE reaction ($PapD_{his}PapE_{Ntd}$ + E_{Nte}) to a value comparable to the non-cognate reaction ($PapD_{his}PapE_{Ntd}$ + H_{Nte}) of 0.05 $M^{-1}s^{-1}$ (Figure 2.2.15, Table 2.2.6). This supports previous ESI-MS data confirming that the

residues C-terminal to the glycine residue at the P4 site are the most important in determining the rate of DSE (Rose *et al.* 2008).

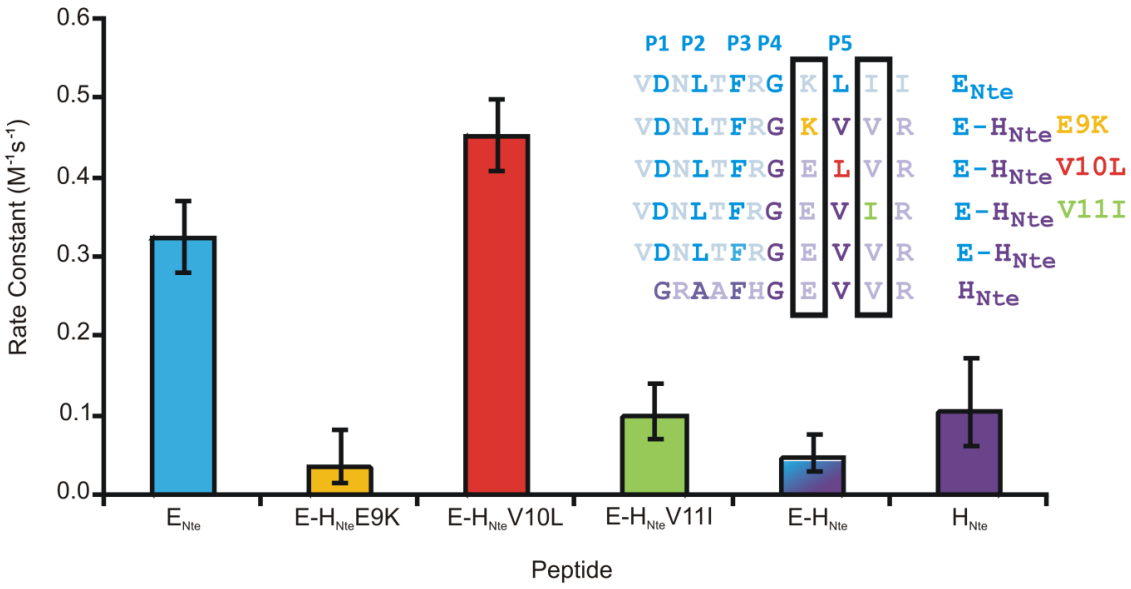


Figure 2.2.15 - DSE rate constants for the DSE reaction between $PapD_{his}PapE_{Ntd}$ and the peptides E_{Nte} (blue), E-H_{Nte}E9K (orange), E-H_{Nte}V10L (red), E-H_{Nte}V11I (green), E-H_{Nte}, (blue-purple) and H_{Nte} (purple). The standard error between duplicate measurements is shown in all cases. The inserts show the peptide sequences used with the alternate hydrophobic residues in bold font, the E_{Nte} and H_{Nte} sequences in blue and purple respectively, and the residues substituted highlighted in orange, red and green for the P5-1, P5 and P5+1 residues respectively.

Substitution of E_{Nte} residues back into E-H_{Nte} individually at the P5-1 and P5+1 positions in E-H_{Nte} (i.e., E-H_{Nte}E9K and E-H_{Nte}V11I, respectively) had very little effect on the second order rate constants of DSE (Table 2.2.6, Figure 2.2.15). These results suggest that in the absence of the P5 residue, the P5+1 and P5-1 residues cannot restore the DSE rate to that of the cognate reaction. The E-H_{Nte}V10L substitution, which introduces the cognate P5 leucine residue of E_{Nte} into E-H_{Nte}, increases the DSE rate constant to 0.45 $M^{-1}s^{-1}$, a value similar to that of the cognate reaction rate constant with E_{Nte} (Table 2.2.6, Figure 2.2.15). The cognate P5 residue alone can act to modulate DSE rates (Figure 2.2.15), by aiding the docking of the Nte into the empty P5 pocket of the accepting subunit for DSE.

Peptide	Residue Substituted	Rate Constant, logk ($\mu M^{-1}s^{-1}$)	Rate Constant, k ($M^{-1}s^{-1}$)
E_{Nte}	None	-6.49 ± 0.06	0.32 (+ 0.05, - 0.04)
E-H_{Nte}E9K	P5-1	-7.44 ± 0.36	0.04 (+ 0.05, - 0.02)
E-H_{Nte}V10L	P5	-6.35 ± 0.04	0.45 (+ 0.05, - 0.04)
E-H_{Nte}V11I	P5+1	-7.00 ± 0.15	0.10 (+ 0.04, - 0.03)
E-H_{Nte}	None	-7.31 ± 0.19	0.05 (+ 0.03, - 0.02)

Table 2.2.6 - Second order rate constants (pH 6.0, 22 ºC) between $PapD_{his}PapE_{Ntd}$ and the substituted peptides E-H_{Nte}E9K, E-H_{Nte}V10L, E-H_{Nte}V11I and E-H_{Nte}. Rate constants are displayed in terms of logk and the derived rate constant, k. The standard error between duplicate DSE reactions are shown.

The DSE rate constants in this thesis show the crucial role of the P5 residue on the incoming subunits' N-terminal extension in DSE initiation. Indeed, substitution of this single amino acid can decrease the rate constant of DSE by up to 3-fold (Figures 2.2.12 and 2.2.14). This supports and extends previous data that has showed the role of the P5 pocket in the chaperone-subunit complex in subunit specificity during DSE (Verger *et al.* 2006; Verger *et al.* 2008). Furthermore, the closed P5 pocket of the chaperone-subunit complex, $PapD_{his}$PapH, prevents DSE and hence *in vivo* this chaperone-subunit complex is used to terminate pilus biogenesis (Verger *et al.* 2006). The residues either side of the P5 residue on the N-terminal extension of PapE were shown here to play a role also in determining subunit specificity during pilus biogenesis (Figures 2.2.14 and 2.2.15), possibly by aiding the initial docking of the subunit N-terminal extension onto the accepting chaperone-subunit complex prior to DSE.

The DSE rates between chaperone-subunit complexes and their cognate N-terminal extensions shown here are too slow to explain how pili assemble on the minute timescale *in vivo*. Additional factors are needed for pilus biogenesis such as the interaction with the large β-barrel transmembrane usher protein, PapC, through which subunits translocate after DSE. *In vitro* experiments involving the fim usher from Type 1 pili have shown that the usher can catalyse pilus assembly compared to when the usher is absent (Vetsch *et al.* 2006). The precise mechanism of the catalysis and subunit ordering of the usher in DSE is still largely unknown. The role of the PapC usher in pilus biogenesis, in particular in determining subunit specificity was, therefore, investigated further using ESI-MS.

2.2.9 Role of the Usher in Pilus Biogenesis

During pilus biogenesis, the chaperone-subunits are thought to be targeted initially to the ushers' N-terminal domain (UsherN) (Ng *et al.* 2004). The binding of the chaperone-subunits to UsherN is thought to aid DSE catalysis through correctly positioning the donor N-terminal extension of the incoming chaperone-subunit close to the P5 pocket on the accepting subunit (Remaut *et al.* 2008). The binding of the chaperone-subunit complexes to UsherN was, therefore, further investigated using ESI-MS.

For all subsequent experiments, the chaperone-subunit complexes $PapD_{his}PapE_{Ntd}$ and $PapD_{his}PapH_{Ntd1}$ constructs were used as previously described (Section 2.2.1 and 2.2.3 respectively). Since PapA is known to self-associate, a truncated construct was used in which the region N-terminal to the conserved Nte binding sequence of PapA

was removed, and the conserved glycine at the P4 site was substituted with Asn, creating $PapA_{Ntd1G15N}$ (Verger *et al.* 2007). Wild-type PapK, PapF and PapG subunits were used.

The following work in Sections 2.2.10, 2.2.11 and 2.2.12 was carried out as a team with Dr. Bethny Morrissey (University of Leeds) and Gabriel Waksmans' group (Birkbeck College, London). Gilles Phan, William Allen (Birkbeck College, London) and myself equally contributed to the expression and purification of the chaperone-subunit complexes $PapD_{his}PapA_{Ntd1G15N}$, $PapD_{his}PapK$, PapDPapF, $PapD_{his}PapE_{Ntd}$ and $PapD_{his}PapH_{Ntd1}$). I expressed and purified all the $PapD_{his}PapG$ and $PapD_{his}PapG_{pilin}$ chaperone-subunit complexes used in addition to the $PapG_{adhesion}$ protein used in ternary complex formation. All usher domains were expressed and purified by Ana Toste-Rego, Gilles Phan and William Allen (Birkbeck College, London). The mass spectra in all cases for the usher project were acquired by Dr. Bethny Morrissey (University of Leeds). All modeling of the chaperone-subunit-usher ternary complexes was performed by myself.

The protein purification for the $PapD_{his}PapG_{pilin}$ chaperone-subunit complex is shown (for example in Figure 2.2.16) and was expressed periplasmically (Materials and Methods Section 2.4.8). A single peak eluted off the nickel affinity column corresponding to a mixture of dissociated $PapD_{his}$ and the non-covalent $PapD_{his}PapG_{pilin}$ complex (Figure 2.2.16A). Any dissociated $PapD_{his}$ was further removed by cation exchange chromatography (Figure 2.2.16B). The final purity of the chaperone-subunit complexes in all cases were confirmed by SDS-PAGE (Figure 2.2.16C) and the molecular weight determined using ESI-MS (Materials and Methods Section 2.4.3).

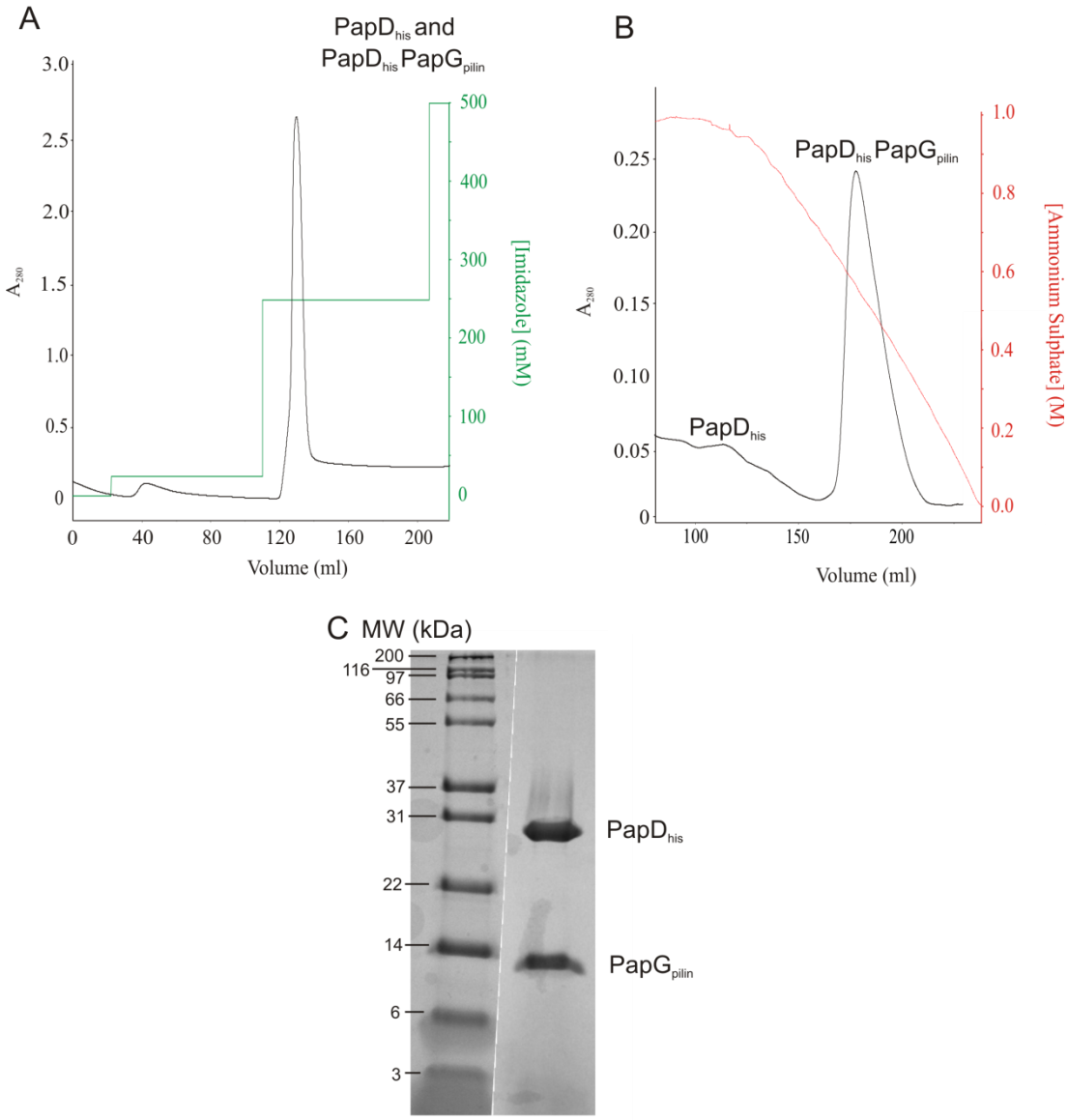


Figure 2.2.16 – Protein purification of the chaperone-subunit complex, $PapD_{his}PapG_{pilin}$. A) Metal affinity chromatography showing $PapD_{his}PapE_{Ntd}$ and $PapD_{his}$ eluting off the column on increasing the concentration of imidazole; B) Cation exchange chromatography showing on decreasing ammonium sulphate concentration the free chaperone, $PapD_{his}$, is separated from the $PapD_{his}PapG_{pilin}$ complex; C) SDS-PAGE showing the periplasmic fraction before purification and the final pure $PapD_{his}PapG_{pilin}$ complex.

2.2.10 Chaperone-Subunit-Usher Ternary Complexes identified by ESI-MS

The interaction of chaperone-subunit complexes with UsherN had not been studied previously by ESI-MS. Chaperone-subunit complexes $PapD_{his}PapG$, PapDPapF, $PapD_{his}PapE_{Ntd}$, $PapD_{his}PapK$, $PapD_{his}PapA_{Ntd1G15N}$ and $PapD_{his}PapH_{Ntd1}$ were incubated separately with the N-terminal domain of the PapC usher (UsherN) in 5 mM ammonium acetate pH 6.8 (Morrissey *et al.* 2012). After 5 minutes, the ESI-MS

mass spectra (acquired by Dr. Bethny Morrissey, University of Leeds) showed the presence of ternary complexes between all chaperone-subunits and the UsherN. All complexes were confirmed by their charge state distribution and corresponding molecular weight (Morrissey *et al.* 2012). One usher domain was bound to one chaperone-subunit complex in each case, suggesting the binding is specific. The mass spectrum of the ternary complex, $PapD_{his}$PapGUsherN, is shown for example in Figure 2.2.17.

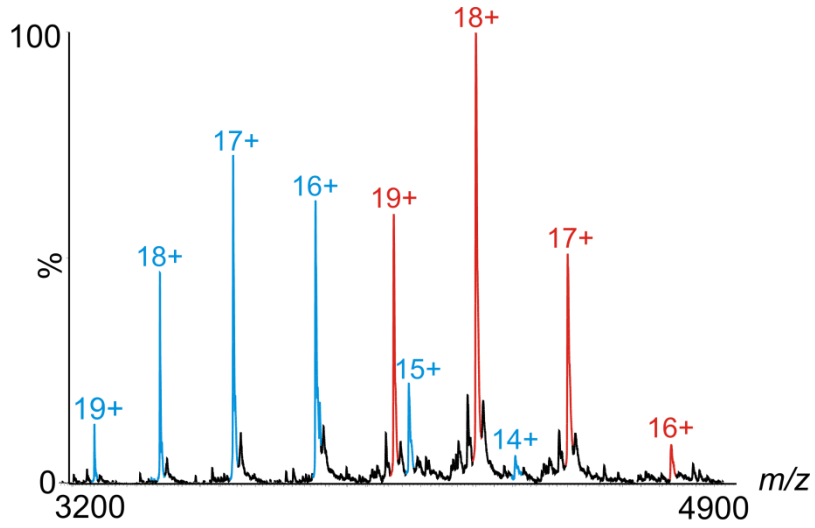


Figure 2.2.17 - ESI-MS of the chaperone-subunit complex, $PapD_{his}$PapG, bound to UsherN. The charge series corresponding to $PapD_{his}$PapG and PapDPapG$_{his}$UsherN are highlighted in blue and red respectively. Unbound UsherN is not shown on this *m/z* scale. $PapD_{his}$PapG was expressed and purified by myself. The mass spectrum was acquired by Dr. Bethny Morrissey (formerly University of Leeds).

2.2.11 Stability of Chaperone-Subunit-Usher Ternary Complexes

The differential binding of subunits to the N-terminal domain of the usher is thought to contribute to the subunit specificity in pilus biogenesis (Dodson *et al.* 1993; Ng *et al.* 2004). Unfortunately the K_D of the chaperone-subunit-UsherN interactions could not be determined accurately from the ESI mass spectra due to differences in the ionisation properties of the chaperone-subunits, chaperone-subunit-usher complexes, and their individual components, in addition to problems associated with potential complex dissociation in the gas phase. Since all chaperone-subunit complexes are

capable of forming stable interactions with UsherN in the gas phase, the difference in the gas phase stabilities of these ternary complexes was investigated by Dr. Bethny Morrissey (University of Leeds) to probe further the origins of subunit specificity during pilus biogenesis (Morrissey *et al.* 2012).

Collision induced dissociation (CID) MS/MS has been widely used to compare the gas phase stability of protein complexes (Benesch *et al.* 2003; Canon *et al.* 2009; Dodds *et al.* 2011), including the stability effect of ligand binding (Van Duijn *et al.* 2006; Hopper *et al.* 2009) and mutations (Hyung *et al.* 2009) on protein complexes. The chaperone-subunit-UsherN ternary complexes were subjected to CID-MS/MS (by Dr. Bethny Morrissey, formerly University of Leeds) in the trap region of the Synapt HDMS. The collision energy applied to the ternary complex precursor ions was increased until no complex remained and the chaperone-subunit-usher complex had completely dissociated. The fraction of the chaperone-subunit-UsherN ion intensity remaining compared with the total ion intensity of all ions in the mass spectrum was plotted as a function of normalized collision energy (Figure 2.2.18), and the energy required to fragment 50 % each of the ternary complexes compared (Figure 2.2.18). PapDPapGUsherN was found to be the most stable ternary complex *in vacuo* (Figure 2.2.18). This is consistent with the known high affinity of PapDPapG for UsherN determined using SPR experiments (Dodson *et al.* 1993). The other five chaperone-subunit-usher complexes showed substantially weaker, but nonetheless significant, gas phase stabilities relative to PapDPapGUsherN (Figure 2.2.18) (Morrissey *et al.* 2012).

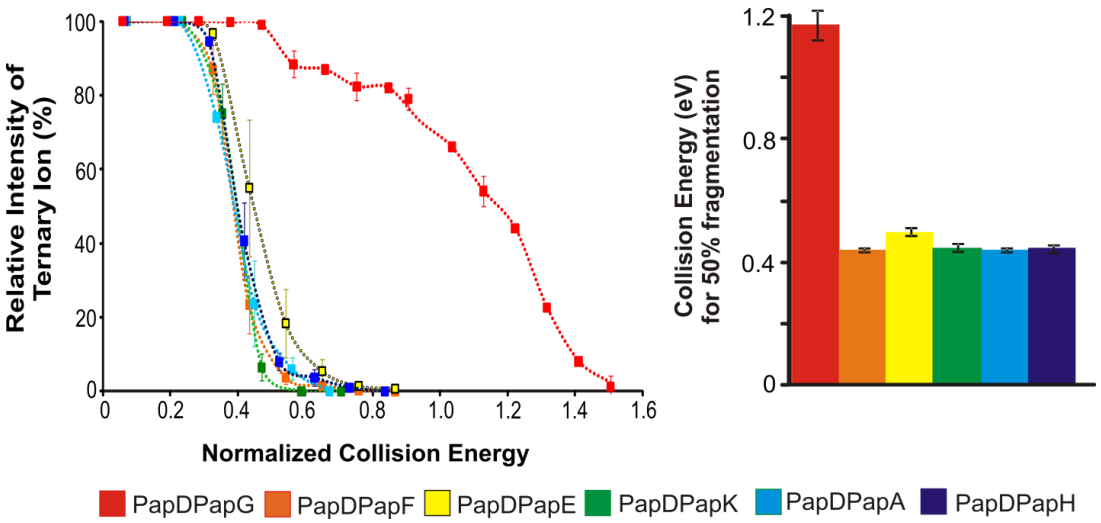


Figure 2.2.18 - Stability of different $PapD_{his}$PapXUsherN ternary complexes determined using CID-MS/MS. The CID-MS/MS fragmentation profiles of the ternary complexes are shown (left) along with the normalised collision energy required to fragment 50 % of the ternary complex into its individual components (right). $PapD_{his}$PapG, $PapD_{his}PapE_{Ntd}$, $PapD_{his}$PapK, $PapD_{his}PapH_{Ntd1}$ were expressed and purified by myself. The MS data were acquired by Dr. Bethny Morrissey (formerly University of Leeds).

It is interesting to note that PapG is the only subunit amongst the pilin subunits that contains two domains, a pilin domain, $PapG_{pilin}$, and an adhesion domain, $PapG_{adhesion}$ (Figure 2.2.19). This additional adhesion domain could be enhancing the gas phase stability of the $PapD_{his}$PapGUsherN compared with all other $PapD_{his}$PapXUsherN complexes. Indeed, the presence of the adhesion domain in FimH, an equivalent subunit to PapG in the Type 1 pilus system, binds strongly to the Fim usher and plays a crucial role in removing the plug domain form the β-barrel of the usher, ready for subunit translocation through the usher during pilus biogenesis (Phan *et al.* 2011).

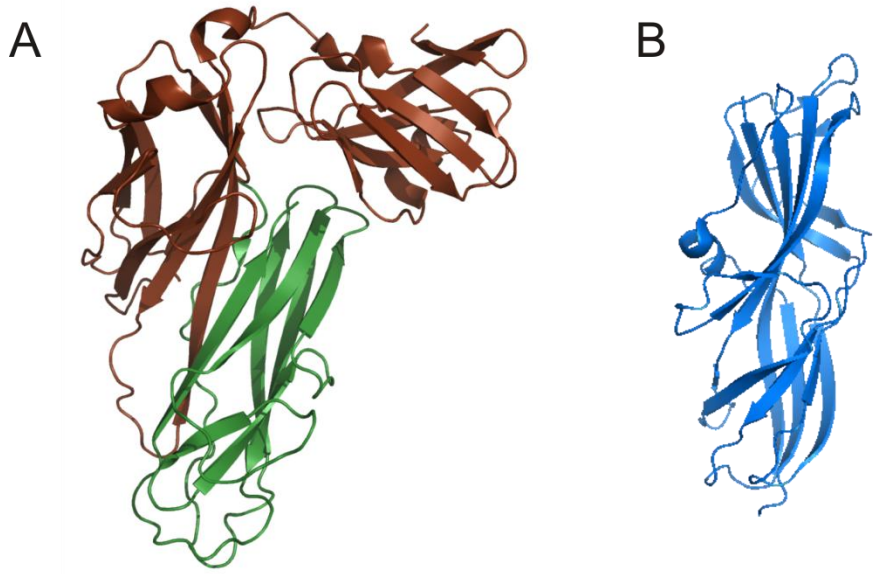


Figure 2.2.19 – Crystal structures of A) PapD (brown) bound to $PapG_{pilin}$ (green) in the $PapDPapG_{pilin}$ complex (PDB file: 2WMP) and B) $PapG_{adhesion}$ (blue) (PDB file:1J8S).

To determine whether the PapG adhesion domain enhances the gas phase stability observed for the $PapD_{his}$PapGUsherN complex, constructs were designed whereby the adhesion domain of PapG in the chaperone-subunit complex, PapDPapG, was deleted (creating $PapDPapG_{pilin}$) and the $PapG_{adhesin}$ domain (residues 1-196 of PapG) was expressed alone. Both $PapDPapG_{pilin}$ and $PapG_{adhesion}$ bound to the UsherN, forming complexes in a 1:1 stoichiometry as detected by ESI-MS. On comparison of the CID/MS/MS of the ternary complexes formed (by Dr. Bethny Morrissey, formerly University of Leeds), the $PapDPapG_{pilin}$UsherN complex stability is dramatically reduced compared with that of the full PapDPapGUsherN complex (Figure 2.2.20). This suggests a role of the $PapG_{adhesin}$ in PapDPapXUsherN complex stability, enhancing the gas phase binding affinity of PapG for UsherN over all other pilin subunits (Figure 2.2.20). Additional CID-MS/MS binding experiments (by Dr. Bethny Morrissey, formerly University of Leeds) between $PapG_{adhesion}$ and the UsherN revealed gas phase stabilities comparable to that of $PapDPapG_{pilin}$ with UsherN bound (Figure 2.2.20). Together, these data suggest that a potential binding interface exists between the UsherN domain of PapC and the adhesion domain of PapG. This additional interaction would explain the increased stability of the PapDPapGUsherN ternary complex over all other PapDPapXUsherN complexes and

may act in facilitating the activation of the usher, required to initiate pilus biogenesis.

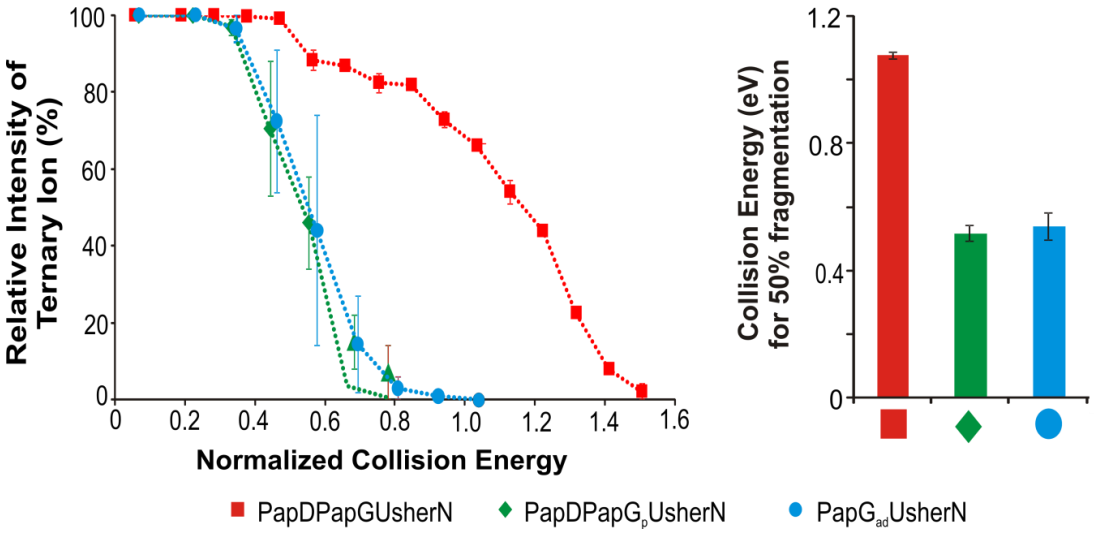


Figure 2.2.20 - Stability of different PapDPapXUsherN ternary complexes determined using CID-MS/MS. The CID-MS/MS fragmentation profiles of the ternary complexes are shown (left) along with the normalised collision energy required to fragment 50 % of the ternary complex into its individual components (right). PapD$_{his}$PapG, PapD$_{his}$PapG$_{pilin}$ (PapDPapG$_p$) and PapG$_{adhesion}$ (PapG$_{ad}$) were expressed and purified by myself. The MS data were acquired by Dr. Bethny Morrissey (formerly University of Leeds).

2.2.12 Predicting the Binding Interface of Chaperone-Subunit-Usher Ternary Complexes

Since a previously unexpected binding interface between the chaperone-subunit complex, PapDPapG, and the N-terminal domain of PapC has been suggested (Figure 2.2.20), ion mobility spectrometry mass spectrometry (IMS-MS) was used by Dr. Bethny Morrissey (formerly University of Leeds) to analyse the complexes between each domain of PapG with UsherN. In parallel, other PapDPapX complexes were analysed by ESI-IMS-MS for comparison. The IMS-MS mass spectrum of the PapDPapFUsherN ternary complex is shown in Figure 2.2.21. A single charge state distribution is observed corresponding to the ternary PapDPapFUsherN complex with a calculated molecular weight of 55,805 ± 2 Da which is within 0.01 % of the molecular weight predicted from the amino acid sequence (55,801 Da). A small amount of free PapDPapF is also observed in the ESI-IMS-MS driftscope which is

thought to be in equilibrium with the PapDPapFUsherN ternary complex in solution. The cross-sectional areas of the PapDPapF and PapDPapFUsherN complexes were determined by Dr. Bethny Morrissey (formerly University of Leeds) from the arrival time distributions of the lowest charge state ions corresponding to each complex. The collision cross-section of PapDPapF was 2520 $Å^2$ for the 12+ charge state ion which corresponds well with the theoretical cross section, 2441 $Å^2$, calculated (by myself) using the Leeds Algorithm (Smith *et al.* 2009; Knapman *et al.* 2010). This suggests that PapDPapF adopts a 'native-like' structure on transition into the gas phase. The collision cross-sectional area of the lowest charge state of the PapDPapFUsherN complex was calculated to be 2878 $Å^2$. This is 10 % smaller than the expected cross section (3218 $Å^2$) based on the theoretically determined cross section calculated (by myself) using the PapDPapFUsherN crystal structure. It is possible that structural collapse is occurring on transition into the gas phase resulting in a smaller measured cross-section. The difference observed here is more likely due to the experimental error in particular in measuring collision cross-sectional areas when using a denatured calibration (Bush *et al.* 2010). Indeed, additional work using a different calibration with native proteins will be needed to take this work further.

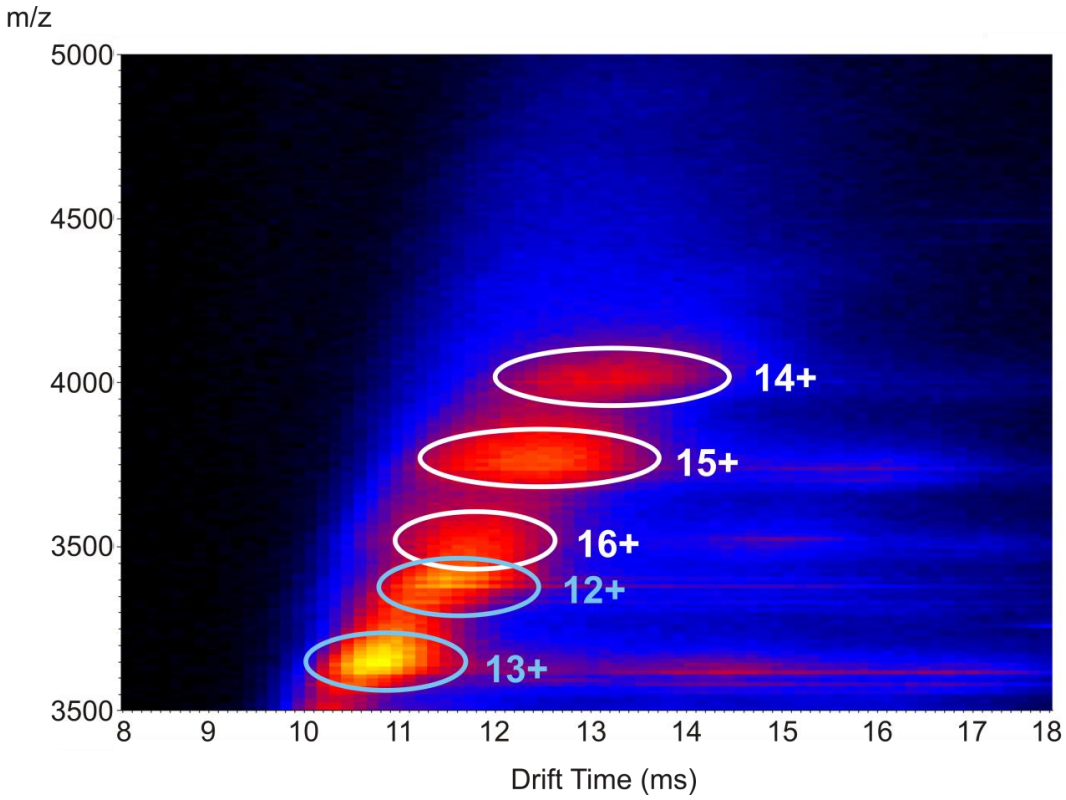


Figure 2.2.21 – ESI-IMS-MS driftscope plot of PapDPapFUsherN. The peaks corresponding to the 12+ and 13+ charge states of the free PapDPapF complex are highlighted in blue. The 14+, 15+, and 16+ charge states corresponding to the ternary PapDPapFUsherN complex are highlighted in white. The MS data were acquired by Dr. Bethny Morrissey (formerly University of Leeds).

Since the crystal structure of PapDPapFUsherN has been determined, molecular modeling was used (by myself) to determine if the binding interface between UsherN and the chaperone-subunit complex PapDPapF could have been predicted based on the information obtained from the IMS-MS data alone. Rigid body docking was used whereby UsherN was manually docked using Pymol (DeLano 2002) onto all possible areas of the chaperone-subunit complex PapDPapF and the theoretical collision cross section calculated using the Leeds algorithm (Smith *et al.* 2009; Knapman *et al.* 2010). The most structurally compact and expanded structures along with the PapDPapFUsherN crystal structures are highlighted in Figure 2.2.22 along with their theoretically calculated collision cross-sectional areas. The difference between collision cross-sections between all structures shown in Figure 2.2.22 is comparable to the experimental error associated with measuring collision cross-sections from the IMS data alone. These results thus show that the binding interface between PapDPapF and UsherN cannot be predicted based on the IMS data alone.

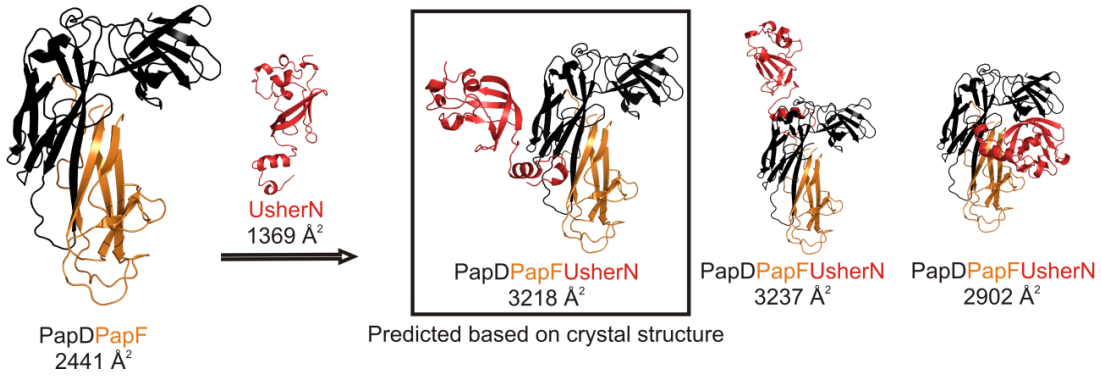


Figure 2.2.22 - Crystal Structure of PapDPapF (black-orange) PDB file:2W07 (Verger *et al.* 2008), UsherN (red) PDB file:1ZE3 (Nishiyama *et al.* 2005) and PapDPapFUsherN PDB files:2W07 (Verger *et al.* 2008) and 1ZE3 (Nishiyama *et al.* 2005) (black-orange-red) along with modeled PapDPapFUsherN structures. The theoretical collision cross sectional areas, calculated using the Leeds algorithm (Smith *et al.* 2009; Knapman *et al.* 2010) are shown for all structures. Note the collision cross-sectional areas observed experimentally are for the lowest charge state.

The ESI-IMS-MS data show that ternary PapDPapXUsherN complexes can be separated from free PapDPapF and UsherN based on their collision cross-sectional areas. Unfortunately increased IMS resolution is needed to further interpret cross section measurements in regards to determining binding interfaces of the PapDPapFUsherN complexes. Steps are being made to increase the resolution of the IMS-MS approach with the more modern instruments (Synapt G2 and Synapt G2S) compared with the Synapt HDMS used in these experiments (Giles *et al.* 2011). Although determining the exact binding interface of chaperone-subunit-usher ternary complex remains challenging, in the future, IMS could be utilised as a rapid, efficient tool for analyzing protein-protein complexes by ruling out unlikely binding sites based on cross-sectional area measurements alone.

2.3. Conclusions

Pilus Assembly

Despite making several inroads into the subject, the complex mechanism by which multiple pilin subunits assemble in such a precise order and stoichiometry to create a functional pilus on the outer membrane of *E.coli* still presents intriguing questions. In this thesis, an *in vitro* method has been developed to determine second order rate constants of DSE using ESI-MS. The use of a chaperone-subunit complex, $PapD_{his}PapH_{Ntd1}$, as an internal standard enabled quantitative analysis of the concentration of the chaperone-subunit complex, $PapD_{his}PapE_{Ntd}$, on DSE carried out with multiple substituted N-terminal extension peptides. The method developed allowed second order DSE rate constants to be determined in real time over reduced reaction times compared with previous studies, minimising problems associated with solvent evaporation and complex dissociation which inevitably increase with reaction time. This method produces rapid, reliable results, consistent with other ion-exchange methods for determining second-order DSE rate constants *in vitro* (Vetsch *et al.* 2006).

The second order rate constants of DSE measured in this thesis reveal new information about the role of individual residues in the Nte in controlling binding of the incoming subunit's Nte to the chaperone-subunit complex, and hence the rate of DSE initiation. The size and availability of the P5 pocket in the chaperone-subunit complex has been shown to modulate rates of DSE (Remaut *et al.* 2006; Rose *et al.* 2008; Verger *et al.* 2008). Therefore, the role of the P5 residue on the N-terminal extension of the incoming subunit that fits tightly into the P5 pocket on the accepting chaperone-subunit complex in DSE was initially investigated. Decreasing the size of the P5 residue on the N-terminal extension resulted in the decrease of the rate of DSE (Figure 2.2.12) showing the amino acid residue at the P5 position is crucial for efficient DSE. The role of the residues either side of the P5 residue (the P5+1 and P5-1 residues) on the N-terminal extension were also investigated. The data show that although the side-chains of the P5+1 and P5-1 residues are orientated away from the binding groove towards the solvent in the final subunit-Nte complex (Figure

2.2.13) (Sauer *et al.* 2002), they play a role in determining the rate of DSE possibly by stabilising the Nte-chaperone-subunit complex. Indeed, crystallography data on Fim subunits from Type 1 pili show that residues surrounding the subunits' hydrophobic groove can flip in and out of the hydrophobic groove on transition from the loose, chaperone-bound state to the more compact subunit-subunit interaction in the pilus tip, thus stabilising the native pilus structure (Le Trong *et al.* 2010). The results presented here reveal that every residue in the C-terminal half of the Nte is involved in DSE, irrespective of its location in the final DSE product.

The effect of the P5, P5+1 and P5-1 residues on the rate of DSE demonstrated here, whilst significant, is not sufficient to rationalise the ordering of subunits in the fully assembled pilus, especially once the concentration of different subunits in the periplasm is taken into account (as PapA is found in vast excess to the other subunits). The usher's differential affinities for subunits, along with its catalytic ability, therefore, must also play a key role in controlling the order of subunit polymerisation (Ng *et al.* 2004; Nishiyama *et al.* 2008).

Work in conjunction with Dr. Bethny Morrissey (formerly University of Leeds) showed that ESI-MS could be used to examine the binding of the chaperone-subunit complexes to the N-terminal domain of the usher, UsherN. CID-MS/MS experiments by Dr. Bethny Morrissey showed that the chaperone-subunit complex PapDPapG had an increased gas phase stability compared with all other chaperone-subunit complexes, consistent with the surface plasmon resonance data that showed PapDPapG to have a higher affinity for the usher compared to all other chaperone-subunit complexes (Dodson *et al.* 1993; Saulino *et al.* 1998; Sauer *et al.* 2004). Further resolution of PapDPapG into its components $PapDPapG_{pilin}$ and $PapG_{adhesion}$ revealed a previously unexpected role of the adhesion domain of the PapG subunit in targeting the chaperone-subunit complex, PapDPapG to the usher, ensuring it binds preferentially to all other chaperone-subunit complexes for pilus biogenesis initiation.

A model was proposed for pilus biogenesis involving the usher (Figure 2.3.1) (Morrissey *et al.* 2012). Pilus assembly initiates through the binding of PapDPapG to the N-terminal domain of the usher. This results in the removal of the plug domain

from the β-barrel into the periplasm. The PapDPapG complex is then passed to the C-terminal domain of the usher either directly or indirectly via the plug domain freeing up the N-terminal domain for binding to the second chaperone-subunit complex, PapDPapF. Once bound and the P5 residue on the PapF N-terminal extension is correctly positioned next to the P5 pocket on PapG, DSE can occur resulting in the release of the chaperone, PapD, from the PapDPapG complex. Finally PapG is translocated across the lumen of the usher simultaneously with PapDPapF moving across to the C-terminal domain, ready for the next incoming chaperone-subunit, $PapD_{his}$PapE, to bind UsherN. More recently, binding studies using biolayer interferometry have shown the interaction of the plug domain with the N-terminal of the usher and its contribution to the final subunit stoichiometry in the pilus (Volkan *et al.* 2012) consistent with the results presented here and in Morrissey *et al.* (Morrissey *et al.* 2012). A refined model in which the plug domain binds to UsherN upon exiting the lumen was proposed (Volkan *et al.* 2012) (Figure 2.3.1).

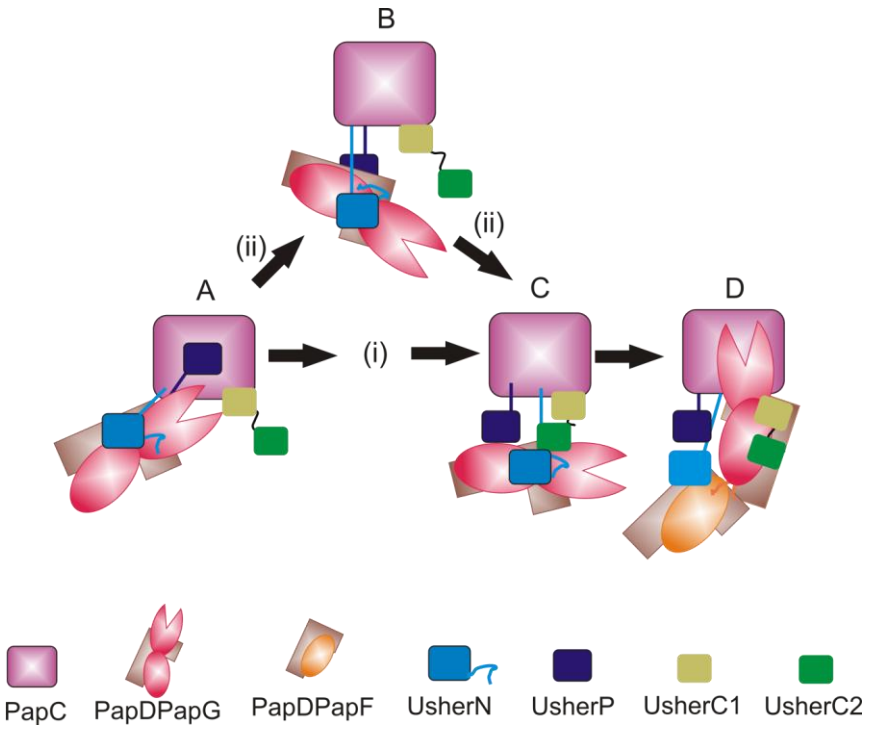


Figure 2.3.1 - Mechanism of pilus biogenesis. Pilus biogenesis commences with PapDPapG (brown-red) binding to UsherN (blue) (A). Once bound to the usher, the plug domain (dark blue) then moves out of the lumen of the barrel (B, C, D). PapDPapG is then transported either directly (i) or indirectly (ii, iii) via the plug domain to the ushers C-terminal domain (Usher C comprised of UsherC1 (light green) and UsherC2 (green)). UsherN is then released ready for binding to the second chaperone-subunit PapDPapF (brown-orange). Once bound to the UsherN, PapDPapF is positioned close to PapDPapG ready for DSE and PapG subunit translocation through the ushers β-barrel. Figure taken from Morrissey *et. al.* 2012 (Morrissey *et al.* 2012).

In summary, the data on DSE rate constants on the role of the P5, P5-1 and P5+1 residues in the N-terminal extension of the incoming chaperone-subunit complex combined with the differential affinity of different subunits to the usher together help explain how multiple subunits can arrange in such a specific order on the outer membrane during pilus biogenesis. Additional DSE experiments (building on previous data shown in Sections 2.2.5 - 2.2.8) involving the intact PapC usher will be needed to develop further our understanding of the DSE mechanism, allowing the role of the individual usher domains in defining the rate controlling the order of subunit assembly to be discerned.

2.4. Materials and Methods

Pilus Assembly

2.4.1 Reagents

Mass Spectrometry

CH_3OH, CH_3CH_2OH, HCl and NH_4OH were purchased from Fisher Scientific (Loughborough, UK). CH_3COONH_4 and CsI were purchased from Sigma-Aldrich Corporation (Poole, Dorset, UK) and raffinose was from ULTRA Scientific (Kent, UK).

Protein Expression and Purification

Agar, yeast extract, tryptone, ampicillin, isopropyl β-D-1-thiogalactopyranoside (IPTG) and Tris base were purchased from Melford Laboratories Ltd. (Suffolk, UK). Glycerol, SDS, imidazole and NaCl were purchased from Fisher Scientific (Loughborough, UK). 30 % Acrylamide, instant blue stain, DTT, Tris-tricine SDS buffer (10x) and the protein standard marker for SDS-PAGE were purchased from Severn Biotech (Kidderminster, UK), Expedeon (Cambridgeshire, UK), Formedium (Norfolk, UK), Invitrogen (Paisley, UK) and Fermentas (Germany), respectively. All other chemicals used in protein expression and purification were purchased from Sigma-Aldrich Corporation (Poole, Dorset, UK).

2.4.2 Buffer Preparation

Ammonium acetate buffers were made at 5 mM in deionised (18 MΩ) water. The pH was adjusted by addition of HCl or NH_4OH. The ammonium acetate buffer was stored at 4 ^{o}C for one week maximum and filtered through a 200 nm-pore filter (minisart high-flow single-use sterile filter, Sartorius, Surrey, UK) before use.

2.4.3 Peptide Preparation

The Nte peptides from PapE along with the substituted Nte peptides were purchased from CSS Albachem, Gladsmuir, UK. Each peptide was C-terminally amidated, and two lysine residues were added to the C-terminus of each sequence to aid solubility. Peptides were dissolved in 5 mM CH_3COONH_4, pH 6.0 and stored at -20 °C. The concentration of peptide was determined accurately by quantitative amino acid analysis (performed by Peter Sharratt, University of Cambridge, UK).

2.4.4 Growth Media

Luria-Bertani (LB) medium was purchased from Merck (Germany) and sterilised by autoclaving at 120 °C at 15 p.s.i. for 20 min. Terrific Broth (TB) was made by dissolving 12 g tryptone, 24 g yeast extract and 4 mL glycerol in 1 L deionised (18 MΩ) water. The TB was autoclaving at 120 °C at 15 p.s.i. for 20 min and once cool, filter sterilised KH_2PO_4 (2.31 g) and K_2HPO_4 (12.54 g) added to a final concentration of 17 mM and 72 mM, respectively. Filter-sterilised antibiotics were added to the LB/TB media when appropriate.

2.4.5 LB agar plates

1.5 g of bacto-agar was added per 100 mL of LB medium and sterilised by autoclaving at 120 °C at 15 p.s.i. for 20 min. Antibiotics were added to the agar solution once the temperature had reached 55 °C. 20 mL of the molten agar mixture was poured into a 100 mm diameter petri dish and allowed to set at room temperature before storage at 4 °C.

2.4.6 Competent Cell Preparation

Single C600 cell colonies from freshly streaked agar plates were grown in 10 mL LB medium at 37 ºC, shaking at 200 rpm for 16 h. Cells were then sub-cultured (1:20) into 100 mL fresh LB medium and grown at 37 ºC, shaking at 200 rpm until an $O.D._{600}$ of 0.45 was reached. Cell were then centrifuged at 4000 rpm (JS 5.3,

Beckman Coulters, UK) at 4 ºC for 10 min and the pellet resuspended in 10 mL of sterile, pre-chilled 100 mM $CaCl_2$ and stored on ice for 2 h. Cells were then centrifuged again, as before, and resuspended in 2 mL of 100 mM $CaCl_2$, 30 % (*v/v*) glycerol at 4 ºC. The competent cells were divided into 100 µL aliquots in sterile reaction tubes, snap frozen and stored at -80 ºC (Sambrook 1989).

2.4.7 Transformation of plasmid DNA

The plasmids encoding the chaperone-subunit complexes, $PapD_{his}PapE_{Ntd}$, $PapD_{his}PapA_{Ntd1G15N}$, $PapD_{his}PapG(II)$, $PapD_{his}PapG_{pilin}$, PapDPapK, $PapD_{his}PapF$, $PapD_{his}PapH_{Ntd1}$ and $PapG_{adhesion}$ were supplied by Professor Gabriel Waksman's group (Birkbeck College, London). *Escherichia coli* (*E.coli*) C600 cells were purchased from Stratagene (Cambridge, UK) and made competent using the protocol described in Section 2.4.6.

1 µL of purified plasmid DNA was incubated on ice with 50 µL of competent C600 *E.coli* cells. After 30 min, the cell/plasmid mixture was heat-shocked at 42 ºC for 45 s before returning the cells to ice for a further 2 min. 500 µL of sterile LB media was then added and the cells incubated, shaking at 200 rpm, for 1 h at 37 ºC. 20-100 µL of transformed cells were plated onto agar plates containing the appropriate antibiotic selection and incubated overnight at 37 ºC.

2.4.8 Protein Expression

2.4.8.1 Chaperone-subunit complexes

All Pap chaperone-subunit complexes were expressed and purified by a combination of Dr Denis Verger, William Allen and Gilles Phan at Birkbeck College, University of London and myself here at the University of Leeds. The same protocol was used in each case. All chaperone-subunit complexes expressed and purified are described below.

$PapD_{his}$-$PapH_{Ntd1}$ complex - The $PapH_{Ntd1}$ used is a truncated PapH construct with the proline-rich region of the Nte (residues Pro2-Gln22) removed (Verger *et al.* 2006). The truncated version of PapH was used to aid solubility (Verger *et al.* 2006).

$PapD_{his}$-$PapE_{Ntd}$ complex - The $PapE_{Ntd}$ construct has the residues Asp2-Ile12 on its Nte deleted to prevent PapE self-polymerisating (Sauer *et al.* 2002; Verger *et al.* 2006).

PapD$PapF_{G8N}$ complex - The P4 residue on the PapF subunit was mutated from Gly to Asn making the construct $PapF_{G8N}$ to prevent PapF self-polymerisation (Verger *et al.* 2008).

$PapD_{his}$-$PapA_{G15NNtd1}$ complex - The $PapA_{G15NNtd1}$ construct has the P4 residue substituted with Asn, in addition to the N-terminal extension being truncated to prevent self-polymerisation (Verger *et al.* 2007).

$PapD_{his}$-$PapG_{pilin}$ complex - Residues 214-336 from PapG(II) were used to create the $PapG_{pilin}$ construct (Morrissey *et al.* 2012).

$PapG_{adhesion}$ complex - The N-terminal 196 amino acids of PapG were used to create the $PapG_{adhesion}$ complex (Dodson *et al.* 2001).

Slightly different histidine-tags and linkers were used for the different Pap chaperone constructs (Table 2.4.1), all will be referred to throughout simply as $PapD_{his}$. The presence of his-tags has previously been shown not to affect the function of the PapD chaperone (Sauer *et al.* 1999; Sauer *et al.* 2002; Verger *et al.* 2006; Verger *et al.* 2007).

Pilin Complex	Vector	Antibiotic Resistance	Inducer	Tag on PapD
$PapD_{his}PapE_{Ntd1}$	pMMB91 ($PapD_{his}$), pTrc99A ($PapE_{Ntd1}$)	Kan ($PapD_{his}$), Amp ($PapE_{Ntd1}$)	IPTG	6his on PapD (no linker)
$PapD_{his}PapA_{NtdG15N}$	pTrc99A	Amp	IPTG	6his (GS link)
$PapD_{his}PapG(II)$	pTrc99A	Amp	IPTG	6his (GS link)
PapDPapK	pMMB91 (PapD), pTrc99A (PapK)	Kan (PapD), Amp (PapK)	IPTG	no tag
$PapD_{his}PapF$	pTrc99A	Amp	IPTG	6his on PapD (GA link)
$PapD_{his}PapH_{Ntd1}$	pTrc99A	Amp	IPTG	6his (GS link)
$PapD_{his}PapG(II)_{pillin}$	pTrc99A	Amp	IPTG	6his (GS link)
$PapG_{adhesin}$	pTrc99A	Amp	IPTG	-

Table 2.4.1 – PapD constructs used in the chaperone-subunit complexes along with the antibiotic resistance of the plasmid to which the gene is inserted.

2.4.8.2 Pre-culture

Initially bacterial glycerol stocks were prepared by inoculating 10 mL of sterile LB media (containing the appropriate antibiotic) with one colony from an agar plate and incubating at 37 ºC overnight, shaking at 200 rpm. 700 µL of the overnight starter culture was then mixed with 300 µL sterile glycerol solution in a 1.5 mL Eppendorf tube to make the final bacterial glycerol stock. Glycerol stocks were stored at -80 ºC.

Pre-cultures were mediated by addition of 20 µL bacterial glycerol stock to 500 mL sterile LB media. The media were incubated overnight at 37 ºC overnight, shaking at 200 rpm and used immediately for large-scale protein over-expression.

2.4.8.3 Periplasmic protein expression

Eleven 2 L flask each containing 1 L of TB media were pre-warmed overnight at 37 ºC. 30 mL of pre-culture was added to each flask along with 1 mL of 100 mg/mL ampicillin and/or 30 mg/mL kanomycin (stored at -20 ºC). The culture was incubated at 37 ºC, shaking at 200 rpm in an orbital shaking incubator, and the optical density (O.D) at 600 nm measured at regular intervals. At an OD of 0.6, the

C600 cells were induced by addition of 1 mL of 1 M filter sterilised IPTG. After 3.5 hours, the cultures were harvested at 5000 rpm (JLA 8.1, Beckman Coulters, UK), 4 ºC.

2.4.8.4 Periplasmic Extraction

The wet cell pellets were resuspended gently in 20 % (*w/v*) sucrose, 20 mM Tris-HCl, pH 8.0 (4 mL sucrose solution for every 1 g cells). 40 µL of 0.5 M EDTA and 40 µL of lysozyme (10 mg/mL) were added and the cells mixed on ice for 20 min to extract the periplasm. Finally 80 µL of 1M $MgCl_2$ was added to stop the reaction and the resulting mixture centrifuged for 20 mins at 12, 000 rpm (JA 25.50, Beckman Coulters, UK), 4 ºC. The pellet was discarded and the supernatant containing the periplasmic proteins dialysed overnight into 20 mM Tris-HCl, 150 mM NaCl, pH 8.0, 5 % (*v/v*) glycerol, 4 ºC for the his-tagged PapDPapX complexes or 20 mM MES pH 6.0, 10 mM NaCl, 1 mM EDTA, 5 % (*v/v*) glycerol, 4 ºC for the non his-tagged PapDPapX complexes. Following dialysis, the periplasmic extract was centrifuged for a further 10 min at 12, 000 rpm (JA 25.50, Beckman Coulters, UK), 4 ºC to remove any residual cell debris before protein purification.

2.4.9 Protein purification

2.4.9.1 $PapD_{his}PapX$ complexes

Metal affinity chromatography

The periplasmic extract was initially purified by nickel affinity chromatography using a HisTrap HP 5 mL column (GE Healthcare, Buckinghamshire, UK). All solutions were filtered using a 0.22 µm Steritop filter (Millipore, Massachusetts, USA) and degassed using a Buchi Vac V-500 vacuum pump (Buchhi Labortechnik AG, Switzerland) prior to passing through the column. The column was equilibrated with 10 column volumes of 20 mM Tris-HCl, 150 mM NaCl, pH 8.0 at 4 ºC and the periplasmic extract was then added at a flow rate of 2 mL/min using a peristaltic pump. The column containing the his-tagged proteins was transferred to an ÄKTA prime purification system (GE Healthcare, Buckinghamshire, UK) and washed with 5 column volumes of 20 mM Tris-HCl, 150 mM NaCl, pH 8.0 at 4 ºC at a flow rate

of 2 mL/min. The column was further washed with 20 mM, 40 mM and 50 mM imidazole in 20 mM Tris-HCl, 150 mM NaCl, pH 8.0 to remove any non-specifically bound proteins. Finally the his-tagged proteins were eluted by addition of 20 mM Tris-HCl, 150 mM NaCl, 500 mM imidazole, pH 8.0. The fractions containing his-tagged protein as detected by spectrophotometry at A_{280} and SDS-PAGE were combined and stored on ice before further purification by hydrophobic interaction chromatography.

Hydrophobic interaction chromatography

The his-tagged proteins were further purified by hydrophobic interaction chromatography using a HiTrap Phenyl HP 5 mL column (GE Healthcare, Buckinghamshire, UK) for the chaperone-subunit complexes $PapD_{his}PapE_{Ntd}$, $PapD_{his}PapA_{Ntd1G15N}$, $PapD_{his}PapG_{pilin}$, PapDPapK, $PapD_{his}PapF$, and $PapD_{his}PapH_{Ntd1}$. A HiTrap Butyl FF 5 mL column (GE Healthcare, Buckinghamshire, UK) was alternatively used for purification of the $PapD_{his}$PapG(II) complex. For all chaperone-subunit complexes, the column was initially washed with 0.5 M NaCl to remove any proteases present from previous purification procedures and then equilibrated with ten column volumes of 20 mM Tris-HCl, 1 M ammonium sulphate, pH 8.0, 4 ºC. All solutions were filtered using a 0.22 μm Steritop filter and degassed using a Buchi Vac V-500 vacuum pump prior to passing through the column. 20 mM Tris-HCl, 3 M ammonium sulphate, pH 8.0 was added drop by drop to the his-tagged protein mixture until a final concentration of 1 M ammonium sulphate was reached. The protein solution was then loaded onto the column using a peristaltic pump at a flow rate of 2 mL/min. The column was transferred to an ÄKTA prime purification system and was washed with 5 column volumes of 20 mM Tris-HCl, 1 M ammonium sulphate, pH 8.0. Bound proteins were then eluted over a linear gradient of 1-0 M ammonium sulphate in 20 mM Tris-HCl, pH 8.0 over 100 mL using a flow rate of 2 mL/min. $PapD_{his}$ typically eluted at 30 mL followed by the $PapD_{his}$PapX complex. The pooled $PapD_{his}$PapX fractions, conformed by SDS-PAGE, were concentrated to 500 μL in 20 mM TrisHCl, 150 mM NaCl, pH 8.0 using a Vivaspin 20 concentrator with a 30,000 MWCO (Sartorius Stedim Biotech., Germany) and stored in 30 % (*v/v*) glycerol at - 20 ºC until further use.

2.4.9.2 PapDPapX complexes

Cation exchange chromatography

For protein complexes containing PapD lacking a his-tag, protein complexes were expressed as described above. The periplasmic extract was initially purified by cation exchange chromatography using a SP HiTrap FF 5 mL column (GE Healthcare, Buckinghamshire, UK). All solutions were filtered using a 0.22 μm Steritop filter and degassed using a Buchi Vac V-500 vacuum pump prior to passing through the column. The column was equilibrated with 10 column volumes of 20 mM MES, 150 mM NaCl, pH 6.0 at 4 ºC and the periplasmic extract added at a flow rate of 2 mL/min using a peristaltic pump. The column containing the PapDPapX proteins was transferred to an ÄKTA prime purification system (GE Healthcare, Buckinghamshire, UK) and washed with 5 column volumes of 20 mM MES, 150 mM NaCl, pH 6.0 at 4 ºC at a flow rate of 2 mL/min. The proteins were eluted by applying a linear gradient of 0-150 mM NaCl in 20 mM MES pH 6.0, over 100 mL. The fractions containing PapDPapX protein as detected by spectrophotometry at A_{280} and SDS-PAGE were combined and stored on ice before further purification by hydrophobic interaction chromatography.

Hydrophobic interaction chromatography

The non his-tagged PapDPapX proteins were purified by hydrophobic interaction chromatography in a similar way to that of his-tagged proteins (Section 2.4.9.1) however 20 mM MES, pH 6.0 was used instead of 20 mM TrisHCl, pH 8.0. The fractions corresponding to the PapDPapX complexes were confirmed by SDS-PAGE and concentrated using a Vivaspin 20 concentrator with a 30,000 MWCO (Sartorius Stedim Biotech., Germany) to 500 μL in 20 mM TrisHCl, pH 8.0 and 30 % (*v/v*) glycerol added for storage at -20 ºC.

2.4.9.3 $PapG_{adhesion}$

The $PapG_{adhesion}$ purification procedure was based on that reported previously by the Waksman group (Dodson *et al.* 2001). $PapG_{adhesion}$ was expressed periplasmically and the periplasm, containing the $PapG_{adhesion}$, extracted according to the procedure mentioned previously (Section 2.4.8.4). $PapG_{adhesion}$ was selectively precipitated in

30 % (*w/v*) ammonium sulphate and the solution centrifuged at 4000 rpm (JS 5.3, Beckman Coulters, UK) for 10 min. The pellet containing the $PapG_{adhesion}$ was resuspended in 20 mM MES pH 5.8 and further dialysed into 20 mM MES pH 5.8 at 4 ºC, overnight.

Cation exchange chromatography

Using a peristaltic pump, the solution containing $PapG_{adhesion}$ was loaded onto a 5 mL S Sepharose column (GE Healthcare, Buckinghamshire, UK) that had been previously equilibrated with 10 column volumes of 20 mM MES pH 5.8 at 4 ºC. The column was then transferred to an ÄKTA prime purification system (GE Healthcare, Buckinghamshire, UK) and washed with 5 column volumes of 20 mM MES pH 5.8 at 4 ºC at a flow rate of 2 mL/min. The proteins were eluted by applying a linear gradient of 0-500 mM NaCl in 20 mM MES pH 6.0, over 100 mL. The fractions containing $PapG_{adhesion}$ protein as detected by spectrophotometry at A_{280} and SDS-PAGE were combined and stored on ice before further purification by hydrophobic interaction chromatography.

Hydrophobic interaction chromatography

$PapG_{adhesion}$ was further purified by hydrophobic interaction chromatography using a HiTrap Butyl FF 5 mL column (GE Healthcare, Buckinghamshire, UK). The column was initially equilibrated with ten column volumes of 20 mM MES, 1 M NaCl pH 5.8, 4 ºC. The protein solution was loaded onto the column in 20 mM MES, 1 M NaCl pH 5.8 using a peristaltic pump at a flow rate of 2 mL/min. The column was transferred to an ÄKTA prime purification system and $PapG_{adhesin}$ eluted over a linear gradient of 1-0 M NaCl in 20 mM MES, pH 5.8 over 100 mL using a flow rate of 2 mL/min. The pooled $PapG_{adhesin}$ fractions, confirmed by SDS-PAGE, were concentrated to 500 µL in 20 mM MES, 150 mM NaCl, pH 5.8 using a Vivaspin 20 concentrator with a 30,000 MWCO (Sartorius Stedim Biotech., Germany) and stored in 30 % (*v/v*) glycerol at - 20 ºC until further use.

2.4.10 Usher Domains

PapC usher domain UsherN (residues 4-131) was expressed periplasmically and purified as reported (Morrissey *et al.* 2012). UsherN was additionally Strep-tagged to aid purification. All usher domains were kindly provided by Ana Toste-Rego, Gilles Phan and William Allen (Birkbeck College, London) and stored at –20 ºC in 20 mM TrisHCl pH 8.0, 150 mM NaCl, 30 % (*v/v*) glycerol.

2.4.11 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE gels were used to monitor the purification of all chaperone-subunit complexes. Tris-tricine buffered gels were used in all cases (Table 2.4.2).

Solution Component	Resolving Gel volume (mL)	Stacking Gel volume (mL)
30 % (*w/v*) Acrylamide: 0.8 % (*w/v*) bis-acrylamide	7.5	0.83
3 M Tris-HCl, 0.3 % (*w/v*) SDS pH 8.45	5.0	1.55
H_2O	0.44	3.72
Glycerol	2.0	-
10 % (*w/v*) Ammonium persulphate	0.1	0.20
TEMED	0.01	0.01

Table 2.4.2 – Tri-tricine buffered SDS-PAGE gel. The volumes to add are sufficient for two 8 cm by 10 cm gels using a 1.5 mm spacer.

Protein samples were diluted two-fold in 2 x loading buffer (50 mM Tris-HCl pH 6.8, 100 mM DTT, 2 % (*w/v*) SDS, 0.1 % (*w/v*) bromophenol blue, 10 % (*v/v*) glycerol) and boiled for a minimum of five min before loading onto the gel. Invitrogen Mark 12 standard protein markers were loaded into one lane to aid identification of protein bands. A cathode buffer containing 200 mM TrisHCl, 200 mM Tricine, 0.2 % (*w/v*) SDS pH 8.25 was used along with the anode buffer containing 400 mM TrisHCl pH 8.8. Gels were electrophoresed at a constant current of ~ 30 mA until the samples entered the resolving gel, at which point the current was adjusted to 55-60 mA. Gels were stained using Instant Blue stain and imaged directly. SDS-PAGE gels were photographed using the Syngene InGenius gel documentation system.

2.4.12 Dialysis Procedure

The chaperone-subunit complexes and usher domains were dialysed into mass spectrometry-compatible buffers (typically 5 mM ammonium acetate pH 6.0) before analysis. 25 μL of protein was inserted into the well of a dialysis button (Hampton Research, Laguna Niguel, USA) and covered with a cellulose dialysis membrane (molecular weight cut-off at 3500 Da, Spectrum Laboratories, Inc.) and held in place by a rubber O-ring. The button was left in 100 mL of ammonium acetate buffer at 4 °C for 2-3 days. The buffer was changed every 4-8 hours to ensure complete removal of salt. Once dialysis was complete, the sample was removed from the microdialysis button using a 5 mL Hamilton syringe.

The final concentration of each chaperone-subunit complex was calculated using the Beer-Lambert law by measuring the absorbance at 280 nm and using the theoretical extinction coefficients determined from the amino acid sequence (using Protparam online software) (Edelhoch 1967; Gill *et al.* 1989; Pace *et al.* 1995).

2.4.13 Electron Microscopy

Carbon/formvar-coated 300 mesh electron microscopy grids were purchased from the Electron Microscopy suite in the Faculty of Biological Sciences, University of Leeds. 2 % (*w/v*) solution of uranyl acetate was obtained from the Electron Microscopy suite in the Faculty of Biological Sciences, University of Leeds.

10 μM PapDPapE_{V56E} (kindly provided by Gilles Phan, Birkbeck College, London) was incubated at room temperature overnight in 5 mM ammonium acetate pH 6.0. 10 μl of the pili formed were then placed on freshly ionised formvar- and carbon-coated EM grids for 30 sec. The grids were then blotted with filter paper to remove excess solvent and the samples stained with 2 % (*w/v*) uranyl acetate for 30 sec. The grids were then blotted again and air-dried before analysis. All images were taken using a Philips CM10 electron microscope operating at 80 keV.

2.4.14 Acquisition of Mass Spectrometry Data

2.4.14.1 Calibration

Before all experiments were carried out, the mass spectrometers were calibrated using a 0.1 mg/mL aqueous solution of CsI. Ion clusters of range *m/z* 1691.8 Da to 4809.5 Da were used to calibrate the required *m/z* range and ensure a mass accuracy of 0.02 %.

2.4.14.2 MS Configuration and Settings

A LCT Premier mass spectrometer (Waters UK Ltd., Manchester, UK) equipped with a chip-based nano-electrospray ionisation source (Triversa NanoMate, Advion Biosciences Corpn., Ithaca, NY, USA) was used to monitor the DSE reactions. A cone voltage of 70 V, capillary voltage of 1.75 kV, nebulising gas pressure of 0.5 psi, and a source temperature of 50 ºC was used in all MS experiments. Spectra were acquired using a 1 second scan, over a *m/z* range of 1000-5000 for all complexes except those containing the PapDPapG chaperone-subunit, in which case ions were acquired over the *m/z* range of 1000-7000. Spectra from a steady total ion chromatogram (TIC) over 1 minute were combined to produce the final summed mass spectra.

2.4.14.3 Processing of Mass Spectrometry Data

MassLynx software (Waters UK Ltd., Manchester, UK) was used to process all mass spectrometry data. A CsI calibration was applied initially to the combined mass spectra to correct the mass to within 0.01 % mass accuracy. All data were smoothed (twice over ten channels) using a Savitsky-Golay algorithm (Savitzky *et al.* 1964). The baseline was subtracted where necessary, and peaks centred according to area (95 % peak height with a minimum peak width of 5 channels at half peak height). Different species were identified by their molecular weight (M_w in Equations 2.4.1 and 2.4.2).

$$m/z = \frac{M_w + nH^+}{n} \qquad \text{Equation 2.4.1}$$

$$n = \frac{(m/z)_{n+1} - H^+}{(m/z)_n - (m/z)_{n+1}}$$ Equation 2.4.2

where n = charge state, (*m/z*) = mass to charge ratio with n charges

The experimental and expected molecular weights are shown in Table 2.4.3. All molecular weights match those predicted within 0.02 %.

Protein/Peptide involved in DSE Reaction	Measured mass (Da)	Calculated mass MW (Da)
N-terminal Extension Peptides		
E_{Nte}	1644	1644
E_{Nte} L10V	1630	1630
E_{Nte} L10A	1602	1602
E_{Nte} K9A	1587	1587
E_{Nte} I11A	1602	1602
E_{Nte} K9E	1645	1645
E_{Nte} I11V	1630	1630
$E\text{-}H_{Nte}$	1660	1660
$E\text{-}H_{Nte}$ E9K	1660	1659
$E\text{-}H_{Nte}$V10L	1674	1674
$E\text{-}H_{Nte}$V11I	1674	1674
H_{Nte}^{+}	1496	1496
DSE Products		
$PapD_{his}^{\#}$	25,382	25,385
$PapD_{his}$*	25,527	25,529
$PapE_{Ntd}$ - E_{Nte}	16,322	16,327
$PapE_{Ntd}$ - E_{Nte} L10V	16,309	16,313
$PapE_{Ntd}$ - E_{Nte} L10A	16,280	16,284
$PapE_{Ntd}$ - E_{Nte} K9A	16,263	16,269
$PapE_{Ntd}$ - E_{Nte} I11A	16,281	16,284
$PapE_{Ntd}$ - E_{Nte} K9E	16,327	16,328
$PapE_{Ntd}$ - E_{Nte} I11V	16,305	16,313
$PapE_{Ntd}$ - $E\text{-}H_{Nte}$	16,348	16,342
$PapE_{Ntd}$ - $E\text{-}H_{Nte}$ E9K	16,337	16,342
$PapE_{Ntd}$ - $E\text{-}H_{Nte}$V10L	16,353	16,357
$PapE_{Ntd}$ - $E\text{-}H_{Nte}$V11I	16,354	16,357
$PapE_{Ntd}$ - H_{Nte}^{+}	16,171	16,178
Chaperone-Subunit Complexes		
$PapD_{his}PapE_{Ntd}$	40,063	40,064
$PapD_{his}PapH_{Ntd1}$	42,262	42,263
$PapD_{his}$PapK	42,222	42,222
$PapD_{his}PapA_{Ntd1G15N}$	41,285	41,288
$PapD_{his}PapF_{G8N}$	40,311	40,314
$PapD_{his}$PapG	60,994	61,013

Table 2.4.3 - A comparison between the calculated molecular masses based on amino acid sequences and the measured molecular masses from the nanoESI-MS analyses. $^{+}$ represents an N-terminally acetylated peptide. # and * correspond to $PapD_{his}$ from $PapD_{his}PapE_{Ntd}$ and $PapD_{his}PapH_{Ntd1}$, $PapD_{his}$PapG, $PapD_{his}PapA_{Ntd1G15N}$ respectively.

2.4.15 Analysis of Mass Spectrometry Data

2.4.15.1 Determination of Second-Order Rate Constants

25 μM of chaperone-subunit complex, $PapD_{his}PapE_{Ntd}$, was incubated at 25 ºC individually with each peptide at a concentration of 250 μM in 5 mM ammonium acetate pH 6.0. Time points were taken every 30 min until a maximum of 8 h. At each time-point, the chaperone-subunit and peptide mixture was diluted 1 in 10 with the internal standard $PapD_{his}PapH_{Ntd1}$ (2.5 μM final concentration in 5 mM ammonium acetate pH 6.0) before spraying immediately into the mass spectrometer.

The concentration of $PapD_{his}PapE_{Ntd}$ *versus* time was determined by comparison of the $PapD_{his}PapE_{Ntd}$ ion intensity with that of the internal standard $PapD_{his}PapH_{Ntd1}$. Using these data, the protein concentration *versus* time was derived from the assumption that one chaperone-subunit complex reacts with only one Nte peptide, and hence the loss of chaperone-subunit is equal to the loss of peptide with time. The $PapD_{his}PapE_{Ntd}$ to $PapD_{his}PapH_{Ntd1}$ intensity ratio for the DSE reactions was then plotted *versus* time and fitted to equation 2.5 in each case using OriginPro7.5 software (OriginLab Corpn., Northampton, MA, US) to determine the second order rate constant, k_1.

2.4.15.2 IMS-MS of Chaperone-Subunit-Usher Complexes

All the ESI-IMS-MS data on the chaperone-subunit-usher ternary complexes were acquired by Bethny Morrissey (formerly University of Leeds). For these IMS-MS experiments, a ramped wave height of 3-11 V at a speed of 200 m/s was used. Drift times were corrected for both mass dependent and mass independent times, (Knapman *et al.* 2010) and drift time cross-section functions were calibrated as previously described (Smith *et al.* 2009). The cross section corresponding to the lowest charge state was reported in each case.

2.4.15.3 Computational Modelling of Chaperone-Subunit-Usher Complexes

Computational cross-sections were calculated from Protein Data Bank (PDB) structures using the Leeds Method Algorithm (Smith *et al.* 2009; Knapman *et al.*

2010). PDB codes for the chaperone-subunit complex PapDPapF$_{G8N}$ (pdb 2W07) was used. Since no crystal structure is available for UsherN, the equivalent residues from the crystal structure in the Fim system were used from Type 1 pili (FimD$_{Ntd}$ - res. 3-124) (pdb 1ZE3) (Phan *et al.* 2011) to estimate the collision cross section area. Tags used to aid purification of usher domains were not included in the modelled collision cross sectional area calculations.

No solved crystal structures for Pap ternary chaperone-subunit-usher were available. As such, the chaperone-subunit:usher binding interface was predicted based on an equivalent chaperone-subunit:usher binding interface found in Type 1 pili, for which crystal structures are available (Eidam *et al.* 2008). For the ternary complexes involving UsherN, the chaperone-subunit complex, PapDPapX, was sequenced aligned with the equivalent FimCFimF chaperone-subunit complex in Type 1 pili from the FimCFimFUsherN crystal structure (Eidam *et al.* 2008) to create the modelled PapDPapXUsherN$_{(Fim)}$ complex. Rigid body docking was used in which structures were manually docked onto each other using Pymol (DeLano 2002) to predict the most expanded and compact ternary complexes.

References

Annesley, T. M. (2003). Ion suppression in mass spectrometry. *Clinical Chemistry* **49**(7): 1041-1044.

Bakker, D., Vader, C. E. M., Roosendaal, B., Mooi, F. R., Oudega, B. and Degraaf, F. K. (1991). Structure and function of periplasmic chaperone-like proteins involved in the biosynthesis of K88 and K99 fimbriae in *Enterotoxigenic Escherichia-coli*. *Mol. Microbiol*. **5**(4): 875-886.

Barnhart, M. M., Sauer, F. G., Pinkner, J. S. and Hultgren, S. J. (2003). Chaperone-subunit-usher interactions required for donor strand exchange during bacterial pilus assembly. *J. Bacteriol.* **185**(9): 2723-2730.

Benesch, J. L. P., Sobott, F. and Robinson, C. V. (2003). Thermal dissociation of multimeric protein complexes by using nanoelectrospray mass spectrometry. *Anal. Chem.* **75**(10): 2208-2214.

Bush, M. F., Hall, Z., Giles, K., Hoyes, J., Robinson, C. V. and Ruotolo, B. T. (2010). Collision cross sections of proteins and their complexes: A calibration framework and database for gas-phase structural biology. *Anal. Chem*. **82**(22): 9557-9565.

Canon, F., Pate, F., Meudec, E., Marlin, T., Cheynier, V., Giuliani, A. and Sarni-Manchado, P. (2009). Characterization, stoichiometry, and stability of salivary protein-tannin complexes by ESI-MS and ESI-MS/MS. *Anal. and Bioanal. Chem*. **395**(8): 2535-2545.

Capitani, G., Eidam, O. and Grutter, M. G. (2006). Evidence for a novel domain of bacterial outer membrane ushers. *Proteins-Structure Function and Bioinformatics* **65**(4): 816-823.

Cegelski, L., Marshall, G. R., Eldridge, G. R. and Hultgren, S. J. (2008). The biology and future prospects of antivirulence therapies. *Nat. Rev. Microbiol*. **6**(1): 17-27.

DeLano, W. L. (2002). The pymol molecular graphics system, *DeLano Scientific*, Palo Alto, CA, USA.

Dhakal, B. K., Bower, J. M., Mulvey, M. A. and Moselio, S. (2009). Pili, fimbriae. *Encyclopedia of microbiology*. Oxford, Academic Press: 470.

Dodds, E. D., Blackwell, A. E., Jones, C. M., Holso, K. L., O'Brien, D. J., Cordes, M. H. J. and Wysocki, V. H. (2011). Determinants of gas-phase disassembly behavior in homodimeric protein complexes with related yet divergent structures. *Anal. Chem.* **83**(10): 3881-3889.

Dodson, K. W., Jacobdubuisson, F., Striker, R. T. and Hultgren, S. J. (1993). Outer-membrane papC molecular usher discriminately recognizes periplasmic chaperone pilus subunit complexes. *Proc. Nat. Acad. Sci. U.S.A.* **90**(8): 3670-3674.

Dodson, K. W., Pinkner, J. S., Rose, T., Magnusson, G., Hultgren, S. J. and Waksman, G. (2001). Structural basis of the interaction of the pyelonephritic *E. coli* adhesin to its human kidney receptor. *Cell* **105**(6): 733-743.

Edelhoch, H. (1967). Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry* **6**(7): 1948-1954.

Eidam, O., Dworkowski, F. S., Glockshuber, R., Grutter, M. G. and Capitani, G. (2008). Crystal structure of the ternary FimC-FimF(t)-FimD(n) complex indicates conserved pilus chaperone-subunit complex recognition by the usher FimD. *FEBS Lett* **582**(5): 651-655.

Giles, K., Williams, J. P. and Campuzano, I. (2011). Enhancements in travelling wave ion mobility resolution. *Rapid Comm. in Mass Spectrom.* **25**(11): 1559-1566.
Gill, S. C. and Vonhippel, P. H. (1989). Calculation of protein extinction coefficients from amino-acid sequence data. *Anal. Biochem.* **182**(2): 319-326.

Holmgren, A. and Branden, C. I. (1989). Crystal-structure of chaperone protein papD reveals an immunoglobulin fold. *Nature* **342**(6247): 248-251.

Hopper, J. T. S. and Oldham, N. J. (2009). Collision induced unfolding of protein ions in the gas phase studied by ion mobility-mass spectrometry: The effect of ligand binding on conformational stability. *J. Am. Soc. Mass Spectrom.* **20**(10): 1851-1858.

Huang, Y. H., Smith, B. S., Chen, L. X., Baxter, R. H. G. and Deisenhofer, J. (2009). Insights into pilus assembly and secretion from the structure and functional characterization of usher PapC. *Proc. Nat. Acad. Sci. U.S.A.* **106**(18): 7403-7407.

Hull, R. A., Gill, R. E., Hsu, P., Minshew, B. H. and Falkow, S. (1981). Construction and expression of recombinant plasmids encoding type-1 or D-mannose-resistant pili from a urinary-tract infection *Escherichia-coli* isolate. *Infection and Immunity* **33**(3): 933-938.

Hung, D. L., Knight, S. D., Woods, R. M., Pinkner, J. S. and Hultgren, S. J. (1996). Molecular basis of two subfamilies of immunoglobulin-like chaperones. *EMBO J.* **15**(15): 3792-3805.

Hung, D. L. and Hultgren, S. J. (1998). Pilus biogenesis via the chaperone/usher pathway: An integration of structure and function. *J. Struct. Biol.* **124**(2-3): 201-220.

Hyung, S. J., Robinson, C. V. and Ruotolo, B. T. (2009). Gas-phase unfolding and disassembly reveals stability differences in ligand-bound multiprotein complexes. *Chemistry & Biology* **16**(4): 382-390.

Jacob-Dubuisson, F., Striker, R. and Hultgren, S. J. (1994). Chaperone-assisted self-assembly of pili independent of cellular-energy. *J. Biol. Chem.* **269**(17): 12447-12455.

Jones, C. H., Pinkner, J. S., Nicholes, A. V., Slonim, L. N., Abraham, S. N. and Hultgren, S. J. (1993). FimC is a periplasmic PapD-like chaperone that directs assembly of type-1 pili in bacteria. *Proc. Nat. Acad. Sci. U.S.A.* **90**(18): 8397-8401.

Jones, C. H., Pinkner, J. S., Roth, R., Heuser, J., Nicholes, A. V., Abraham, S. N. and Hultgren, S. J. (1995). FimH adhesin of type-1 pili is assembled into a fibrillar tip structure in the *Enterobacteriaceae*. *Proc. Nat. Acad. Sci. U.S.A.* **92**(6): 2081-2085.

Klemm, P. and Christiansen, G. (1990). The FimD gene required for cell-surface localization of *Escherichia-coli* type-1 fimbriae. *Molecular & General Genetics* **220**(2): 334-338.

Knapman, T. W., Berryman, J. T., Campuzano, I., Harris, S. A. and Ashcroft, A. E. (2010). Considerations in experimental and theoretical collision cross-section measurements of small molecules using travelling wave ion mobility spectrometry-mass spectrometry. *Int. J. Mass Spectrom.* **298**(1-3): 17-23.

Knight, S., Mulvey, M. and Pinkner, J. (1997). Crystallization and preliminary X-ray diffraction studies of the FimC-FimH chaperone-adhesin complex from escherichia coli. *Acta Crystallographica Section D-Biological Crystallography* **53**: 207-210.

Kuehn, M. J., Heuser, J., Normark, S. and Hultgren, S. J. (1992). P pili in uropathogenic *Escherichia-coli* are composite fibers with distinct fibrillar adhesive tips. *Nature* **356**(6366): 252-255.

Lee, Y. M., DiGiuseppe, P. A., Silhavy, T. J. and Hultgren, S. J. (2004). P pilus assembly motif necessary for activation of the Cpxra pathway by papE in *Escherichia coli*. *J. Bacteriol.* **186**(13): 4326-4337.

Lee, Y. M., Dodson, K. W. and Hultgren, S. J. (2007). Adaptor function of papF depends on donor strand exchange in P-pilus biogenesis of *Escherichia coli*. *J. Bacteriol.* **189**(14): 5276-5283.

Le Trong, I., Aprikian, P., Kidd, B. A., Thomas, W. E., Sokurenko, E. V. and Stenkamp, R. E. (2010). Donor strand exchange and conformational changes during *E. coli* fimbrial formation. *J. Struct. Biol.* **172**(3): 380-388.

Leney, A. C., Phan, G., Allen, W., Verger, D., Waksman, G., Radford, S. E. and Ashcroft, A. E. (2011). Second order rate constants of donor-strand exchange reveal individual amino acid residues important in determining the subunit specificity of pilus biogenesis. *J. Am. Soc. Mass Spectrom.* **22**(7): 1214-1223.

Li, H. L., Qian, L. P., Chen, Z. Q., Thibault, D., Liu, G., Liu, T. B. and Thanassi, D. G. (2004). The outer membrane usher forms a twin-pore secretion complex. *J. Mol. Biol.* **344**(5): 1397-1407.

Li, Q. Q., Ng, T. W., Dodson, K. W., So, S. S. K., Bayle, K. M., Pinkner, J. S., Scarlata, S., Hultgren, S. J. and Thanassi, D. G. (2010). The differential affinity of

the usher for chaperone-subunit complexes is required for assembly of complete pili. *Mol. Microbiol.* **76**(1): 159-172.

Morrissey, B., Leney, A. C., Toste Rego, A., Phan, G., Allen, W. J., Verger, D., Waksman, G., Ashcroft, A. E. and Radford, S. E. (2012). The role of chaperone-subunit usher domain interactions in the mechanism of bacterial pilus biogenesis revealed by ESI-MS. *Mol. Cell. Prot.* **11**(7): M111.015289.

Mu, X. Q. and Bullitt, E. (2006). Structure and assembly of P-pili: A protruding hinge region used for assembly of a bacterial adhesion filament. *Proc. Nat. Acad. Sci. U.S.A.* **103**(26): 9861-9866.

Munera, D., Hultgren, S. and Fernandez, L. A. (2007). Recognition of the N-terminal lectin domain of FimH adhesin by the usher FimD is required for type 1 pilus biogenesis. *Mol. Microbiol.* **64**(2): 333-346.

Munera, D., Palomino, C. and Fernandez, L. A. (2008). Specific residues in the N-terminal domain of FimH stimulate type I fimbriae assembly in Escherichia coli following the initial binding of the adhesin to FimD usher. *Mol. Microbiol*. **69**(4): 911-925.

Ng, T. W., Akman, L., Osisami, M. and Thanassi, D. G. (2004). The usher n terminus is the initial targeting site for chaperone-subunit complexes and participates in subsequent pilus biogenesis events. *J. Bacteriol.* **186**(16): 5321-5331.

Nishlyama, M., Vetsch, M., Puorger, C., Jelesarov, I. and Glockshuber, R. (2003). Identification and characterization of the chaperone-subunit complex-binding domain from the type 1 pilus assembly platform FimD. *J. Mol. Biol.* **330**(3): 513-525.

Nishiyama, M., Horst, R., Eidam, O., Herrmann, T., Ignatov, O., Vetsch, M., Bettendorff, P., Jelesarov, I., Grutter, M. G., Wuthrich, K., Glockshuber, R. and Capitani, G. (2005). Structural basis of chaperone-subunit complex recognition by the type 1 pilus assembly platform FimD. *EMBO J.* **24**(12): 2075-2086.

Nishiyama, M., Ishikawa, T., Rechsteiner, H. and Glockshuber, R. (2008). Reconstitution of pilus assembly reveals a bacterial outer membrane catalyst. *Science* **320**(5874): 376-379.

Nishiyama, M. and Glockshuber, R. (2010). The outer membrane usher guarantees the formation of functional pili by selectively catalyzing donor-strand exchange between subunits that are adjacent in the mature pilus. *J. Mol. Biol.* **396**(1): 1-8.

Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. and Gray, T. (1995). How to measure and predict the molar absorption-coefficient of a protein. *Prot. Sci.* **4**(11): 2411-2423.

Phan, G., Remaut, H., Wang, T., Allen, W. J., Pirker, K. F., Lebedev, A., Henderson, N. S., Geibel, S., Volkan, E., Yan, J., Kunze, M. B. A., Pinkner, J. S., Ford, B., Kay, C. W. M., Li, H. L., Hultgren, S. J., Thanassi, D. G. and Waksman, G. (2011).

Crystal structure of the FimD usher bound to its cognate FimC-FimH substrate. *Nature* **474**(7349): 49-U71.

Remaut, H., Rose, R. J., Hannan, T. J., Hultgren, S. J., Radford, S. E., Ashcroft, A. E. and Waksman, G. (2006). Donor-strand exchange in chaperone-assisted pilus assembly proceeds through a concerted beta strand displacement mechanism. *Mol. Cell* **22**(6): 831-842.

Remaut, H., Tang, C., Henderson, N. S., Pinkner, J. S., Wang, T., Hultgren, S. J., Thanassi, D. G., Waksman, G. and Li, H. L. (2008). Fiber formation across the bacterial outer membrane by the chaperone/usher pathway. *Cell* **133**(4): 640-652.

Rose, R. J., Verger, D., Daviter, T., Remaut, H., Paci, E., Waksman, G., Ashcroft, A. E. and Radford, S. E. (2008). Unraveling the molecular basis of subunit specificity in P pilus assembly by mass spectrometry. *Proc. Nat. Acad. Sci. U.S.A.* **105**(35): 12873-12878.

Rose, R. J., Welsh, T. S., Waksman, G., Ashcroft, A. E., Radford, S. E. and Paci, E. (2008). Donor-strand exchange in chaperone-assisted pilus assembly revealed in atomic detail by molecular dynamics. *J. Mol. Biol.* **375**(4): 908-919.

Sambrook, J., Fritsch, E.F., Maniatis T. (1989). Molecular cloning - a laboratory manual.

Sauer, F. G., Futterer, K., Pinkner, J. S., Dodson, K. W., Hultgren, S. J. and Waksman, G. (1999). Structural basis of chaperone function and pilus biogenesis. *Science* **285**(5430): 1058-1061.

Sauer, F. G., Barnhart, M., Choudhury, D., Knights, S. D., Waksman, G. and Hultgren, S. J. (2000). Chaperone-assisted pilus assembly and bacterial attachment. *Curr. Opin. Struct. Biol.* **10**(5): 548-556.

Sauer, F. G., Pinkner, J. S., Waksman, G. and Hultgren, S. J. (2002). Chaperone priming of pilus subunits facilitates a topological transition that drives fiber formation. *Cell* **111**(4): 543-551.

Sauer, F. G., Remaut, H., Hultgren, S. J. and Waksman, G. (2004). Fiber assembly by the chaperone-usher pathway. *Biochimica Et Biophysica Acta-Mol. Cell Res.* **1694**(1-3): 259-267.

Saulino, E. T., Thanassi, D. G., Pinkner, J. S. and Hultgren, S. J. (1998). Ramifications of kinetic partitioning on usher-mediated pilus biogenesis. *EMBO J.* **17**(8): 2177-2185.

Savitzky, A. and Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **36**(8): 1627-1639.

Smith, D. P., Knapman, T. W., Campuzano, I., Malham, R. W., Berryman, J. T., Radford, S. E. and Ashcroft, A. E. (2009). Deciphering drift time measurements

from travelling wave ion mobility spectrometry-mass spectrometry studies. *Eur. J. Mass Spectrom*. **15**(2): 113-130.

So, S. S. K. and Thanassi, D. G. (2006). Analysis of the requirements for pilus biogenesis at the outer membrane usher and the function of the usher C-terminus. *Mol. Microbiol*. **60**(2): 364-375.

Stathopoulos, C., Hendrixson, D. R., Thanassi, D. G., Hultgren, S. J., St Geme, J. W. and Curtiss, R. (2000). Secretion of virulence determinants by the general secretory pathway in Gram-negative pathogens: An evolving story. *Microbes and Infection* **2**(9): 1061-1072.

van Duijn, E., Simmons, D. A., van den Heuvel, R. H. H., Bakkes, P. J., van Heerikhuizen, H., Heeren, R. M. A., Robinson, C. V., van der Vies, S. M. and Heck, A. J. R. (2006). Tandem mass spectrometry of intact GroEL-substrate complexes reveals substrate-specific conformational changes in the trans ring. *J. Am. Chem. Soc.* **128**(14): 4694-4702.

Thanassi, D. G., Stathopoulos, C., Dodson, K., Geiger, D. and Hultgren, S. J. (2002). Bacterial outer membrane ushers contain distinct targeting and assembly domains for pilus biogenesis. *J. Bacteriol*. **184**(22): 6260-6269.

Verger, D., Miller, E., Remaut, H., Waksman, G. and Hultgren, S. (2006). Molecular mechanism of P pilus termination in uropathogenic *Escherichia coli*. *EMBO Reports* **7**(12): 1228-1232.

Verger, D., Bullitt, E., Hultgren, S. J. and Waksman, G. (2007). Crystal structure of the P pilus rod subunit PapA. *PLOS Pathogens* **3**(5): 674-682.

Verger, D., Rose, R. J., Paci, E., Costakes, G., Daviter, T., Hultgren, S., Remaut, H., Ashcroft, A. E., Radford, S. E. and Waksman, G. (2008). Structural determinants of polymerization reactivity of the P pilus adaptor subunit PapF. *Structure* **16**(11): 1724-1731.

Vetsch, M., Erilov, D., Moliere, N., Nishiyama, M., Ignatov, O. and Glockshuber, R. (2006). Mechanism of fibre assembly through the chaperone-usher pathway. *EMBO Reports* **7**(7): 734-738.

Volkan, E., Ford, B. A., Pinkner, J. S., Dodson, K. W., Henderson, N. S., Thanassi, D. G., Waksman, G. and Hultgren, S. J. (2012). Domain activities of PapC usher reveal the mechanism of action of an *Escherichia coli* molecular machine. *Proc. Nat. Acad. Sci. U.S.A.* **109**(24): 9563-9568.

Wiles, T. J., Kulesus, R. R. and Mulvey, M. A. (2008). Origins and virulence mechanisms of uropathogenic *Escherichia coli*. *Experimental and Molecular Pathology* **85**(1): 11-19.

Zavialov, A. V., Berglund, J., Pudney, A. F., Fooks, L. J., Ibrahim, T. M., MacIntyre, S. and Knight, S. D. (2003). Structure and biogenesis of the capsular F1

antigen from *Yersinia Pestis*: Preserved folding energy drives fiber formation. *Cell* **113**(5): 587-596.

Zavialov, A. V., Tischenko, V. M., Fooks, L. J., Brandsdal, B. O., Aqvist, J., Zav'Yalov, V. P., MacIntyre, S. and Knight, S. D. (2005). Resolving the energy paradox of chaperone/usher-mediated fibre assembly. *Biochem. J.* **389**: 685-694.

Amyloid

3.1. Introduction

3.1.1 Amyloidosis

The correct folding of proteins into their native, functional state is highly dependent on protein sequence, the environment, and the presence of chaperones (Anfinsen 1973; Hartl 1996). If any of the above conditions is altered, protein misfolding can occur resulting in disease. Amyloidosis is the term given to protein misfolding/aggregation that results in long straight, unbranched polymers with a cross-β structure termed amyloid fibrils. These amyloid fibril structures are associated with over thirty different diseases, the most common of which in humans are Alzheimer's, Parkinson's and Type II diabetes mellitus (Chiti *et al.* 2006).

Amyloid fibrils are typically characterised by their ordered array of β-strands running perpendicular to the fibril long axis, termed a cross-β structure (Figure 3.1.1A). This cross-β structure is shown clearly in X-ray fibre diffraction data whereby the signals present correspond to the distance between β-strands within the β-sheets (4.7 Å) and the spacing between β-sheets (10 Å) (Figure 3.1.1A, B) (Luhrs *et al.* 2005). Imaging of these fibrils by electron microscopy is also used to confirm amyloid fibril formation with long, straight, twisted fibrils being the typical characteristics observed (Figure 3.1.1C) (Gras *et al.* 2011). The presence of amyloid fibrils is also demonstrated by their ability to bind dyes such as Congo red and Thioflavin T (ThT). Congo red is used to stain amyloid fibrils in tissues and exhibits apple-green birefringence upon visualisation with polarised light (Ladewig 1945; Westermark *et al.* 1999; Hatters *et al.* 2011). Thioflavin T binds amyloid fibrils and causes an increase in fluorescence at 480 nm on excitation at 440 nm (Naiki *et al.* 1989; Biancalana *et al.* 2010).

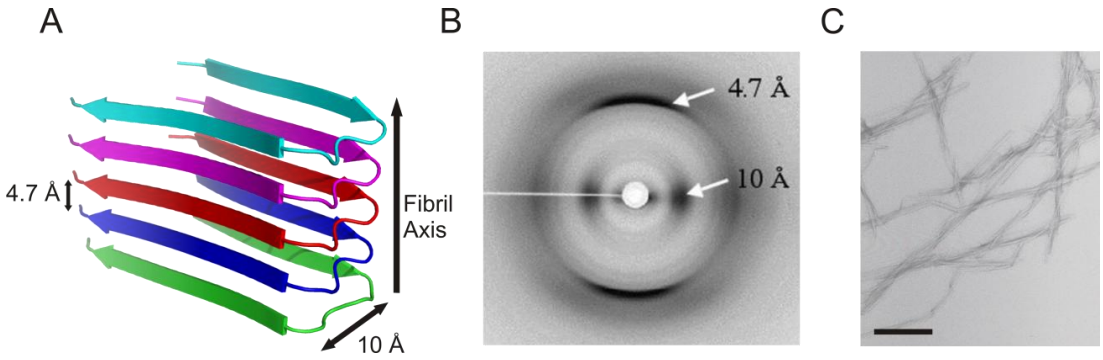


Figure 3.1.1 - Characterisation of Amyloid Fibrils A) Molecular model (PDB file: 2BEG) of the cross-β structure suggested for the $A\beta_{1-42}$ amyloid fibrils associated with Alzheimer's disease (Luhrs *et al.* 2005); B) X-ray fibre diffraction pattern of amyloid fibrils showing scattering at 4.7 Å and 10 Å corresponding to the distance between the β-strands within the β-sheets and the distance between β-sheets, respectively; C) Electron microscopy image of long, straight beta-2 microglobulin amyloid fibrils (taken myself). Scale bar represents 200 nm. A and B were adapted from Luhrs *et al.* (Luhrs *et al.* 2005) and Stromer *et al.* (Stromer *et al.* 2005), respectively.

3.1.2 Mechanism of Amyloid Formation

Amyloid formation is thought to follow a typical nucleated growth mechanism comprising a lag phase (or nucleation phase) and an elongation phase (Figure 3.1.2). The lag phase is the rate-limiting step whereby monomers unfold and rearrange to form the critical nucleus that is rapidly extended to form mature fibrils during the elongation phase (Harper *et al.* 1997; Chiti *et al.* 2003; Xue *et al.* 2008; Kashchiev *et al.* 2010). Interestingly, this lag phase can be eliminated by the addition of short pre-formed amyloid fibrils, called seeds, at the start of the reaction (Figure 3.1.2).

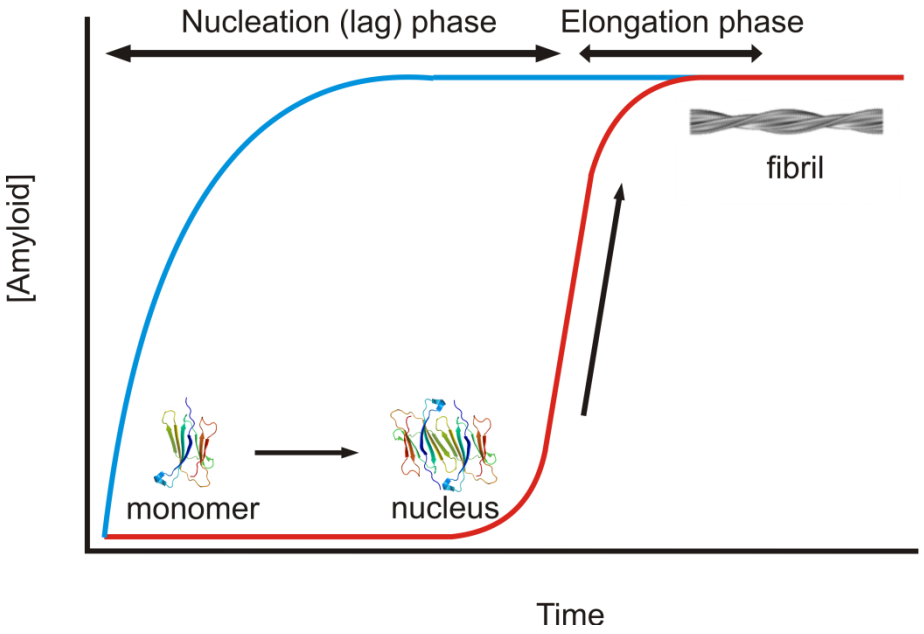


Figure 3.1.2 - Kinetics of amyloid fibril formation by a *de novo* nucleation-dependent mechanism (red) or after addition of amyloid seeds (blue). *De novo* fibril formation comprises a nucleation (lag) phase whereby monomers rearrange to oligomers before reaching the critical nucleus, and an elongation phase whereby the nucleus rapidly extends resulting in amyloid fibril formation. If amyloid seeds are added to monomeric protein (blue), the lag phase is reduced and the elongation phase is evident immediately.

During the lag phase of fibril formation, monomers are thought to assemble into oligomeric species. The presence of these heterogeneous oligomeric intermediates has been shown to correlate with amyloid disease progression (Lue *et al.* 1999; Wang *et al.* 1999; Tomic *et al.* 2009). It has been suggested that the oligomers formed during fibrillogenesis are the toxic species rather than the fibril end products themselves (Bucciantini *et al.* 2002; Reixach *et al.* 2004; Winner *et al.* 2011). Attention has turned to the structure of these amyloidogenic intermediates and their role in toxicity. The structure of the amyloidogenic intermediates of an amyloid forming protein beta-2 microglobulin (β_2m) will be the main focus of this thesis chapter.

3.1.3 Beta-2 microglobulin

β_2m is a 99-residue protein that is the light chain of the non-covalently bound major histocompatibility complex 1 (MHC 1). In its native state, β_2m consists of seven β-strands (A-G) arranged as two β-sheets (ABED and CFG) in a β sandwich fold connected by a disulphide bond between residues Cys 25 and Cys 80 (Figure 3.1.3 and 3.1.4). In healthy individuals, β_2m dissociates from the MHC I complex and circulates in the blood before being degraded by the kidneys (Figure 3.1.3) (Floege *et al.* 1996); however, in patients with kidney failure who undergo long-term dialysis, the concentration of β_2m increases up to 60-fold due to its lack of removal by the kidneys (Floege *et al.* 1996). Eventually, in these cases, β_2m is found deposited as amyloid fibrils in osteoarticular tissues, resulting in a condition termed dialysis-related amyloidosis (DRA) (Floege *et al.* 1996).

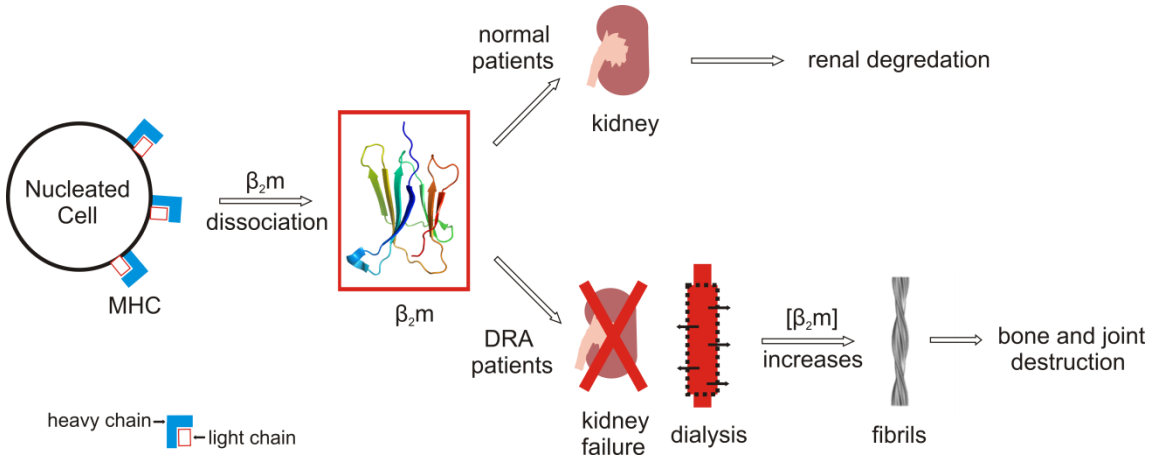


Figure 3.1.3 - Dialysis related amyloidosis (DRA) schematic showing β_2m (PDB file: 2YXF) (Iwata *et al.* 2007) dissociating from the major histocompatibility complex 1 (MHC 1). β_2m circulating in the blood is then either degraded by the kidneys and removed from the body or in the case of renal failure, accumulates in the body and forms fibrils that lead to bone and joint destruction causing the condition dialysis related amyloidosis (DRA).

Despite the wealth of information in the literature on DRA, the precise mechanism of β_2m amyloid assembly remains unknown. *In vitro* studies have shown β_2m remains soluble at pH 7.5 with no evidence of amyloid formation even after 100 days of incubation at 37 ºC (Eichner *et al.* 2009). Additional biological factors such as the patients' age (Van Ypersele de Strihou *et al.* 1991), the duration of kidney failure

(Davison 1995) and post-translational modifications of β_2m (Capeillereblandin *et al.* 1991; Miyata *et al.* 1996; Niwa *et al.* 1997; Esposito *et al.* 2000; Bellotti *et al.* 2001; Heegaard *et al.* 2002) are needed to facilitate amyloid formation *in vivo*.

3.1.4 β_2m Fibril Formation *in vitro*

Amyloid fibrils can be formed *in vitro* from wild-type human β_2m at pH 2.5 in a short time-scale using low ionic strength (< 100 mM) buffer (Kad *et al.* 2001). The highly dynamic acid-unfolded β_2m forms long straight, unbranched, twisted fibrils (as observed by electron microscopy) via a nucleation-dependent mechanism (Xue *et al.* 2008) (Figure 3.1.2). Fibrils formed at acidic pH display all the hallmarks of amyloid including a cross-β fibre diffraction pattern, Congo red birefringence, nucleation-dependent polymerisation kinetics, in addition to binding thioflavin T, apolipoprotein E, glycosaminoglycans, the serum amyloid P component and the anti-amyloid antibody WO1 (Smith *et al.* 2003; Gosal *et al.* 2005; Myers *et al.* 2006) (Figure 3.1.1).

Approximately 60 % of the β_2m sequence, as determined using small synthetic peptides and theoretical prediction algorithms (Kozhukh *et al.* 2002; Hasegawa *et al.* 2003; Ivanova *et al.* 2003; Jones *et al.* 2003; Ivanova *et al.* 2004; Pawar *et al.* 2005; Ivanova *et al.* 2006), has the potential to assemble into amyloid fibrils *in vitro*. In full length β_2m only a small region (residues 60-70) has been shown to affect fibril nucleation and elongation kinetics at pH 2.5 and low ionic strength (Platt *et al.* 2008; Routledge *et al.* 2009). This region contains six aromatic residues and is predicted to be highly aggregation prone (Pawar *et al.* 2005); however, contrasting proline scanning studies have revealed additional residues (Leu23, His51 and Val82) crucial in elongating fibril seeds (Chiba *et al.* 2003). Both the thermodynamics involved in fibril assembly and the stability of the fibrils affect the assembly of β_2m amyloid fibrils (Plattt *et al.* 2005; Routledge *et al.* 2009). The presence of the disulphide bond between the β-strands B and F which introduces residual structure into the acid-unfolded state is crucial for long-straight amyloid fibril formation (Smith *et al.* 2001; Hasegawa *et al.* 2003).

Figure 3.1.4 - MHC 1 complex (PDB file: 2X70) (Celie *et al.* 2009) showing the heavy α-chain (black ribbon) and the light chain (β_2m coloured). The non-covalently bound β_2m molecules that are released aggregate into fibrils associated with the human disease DRA. The β-strands of the β_2m molecule are labelled A-G.

It is only once β_2m is released from MHC 1 that amyloid formation occurs; hence, it is important to note the structural contacts β_2m makes with MHC 1 since these regions are exposed on release from the MHC 1 complex. The D β-strand and the connecting DE loop of β_2m are involved in strong binding to the α-chain of MHC 1. Mutations in this loop (e.g. W60G) have been shown to prevent MHC 1 assembly (Esposito *et al.* 2008). On release from the MHC 1 complex this D β-strand becomes highly dynamic (Hodkinson *et al.* 2009) and aggregation-prone (Esposito *et al.* 2008; Ricagno *et al.* 2008; Santambrogio *et al.* 2010).

3.1.5 Fibril Formation as monitored by ESI-MS

Mass spectrometry has enabled significant structural advances in the understanding of the mechanism of amyloid formation, in particular concerning the oligomeric intermediates on- or off-pathway to fibril formation (Ashcroft 2010). From ESI-MS alone, structural details about monomeric proteins are immediately evident based on their charge state distribution in the *m/z* spectrum (Introduction Section 1.6, Figure 1.8). Studies on the intrinsically disordered protein, α-synuclein monomer whose aggregation is associated with Parkinson's disease, has shown multiple distinct populations present in the mass spectrum (Frimpong *et al.* 2010; Natalello *et al.* 2011). Monomeric β_2m has been studied extensively by ESI-MS, revealing predominantly one native-like conformation at neutral pH. β_2m is thought to populate three distinct conformations at low pH (a compact, partially compact and expanded structure) (Figure 3.1.5) (Borysik *et al.* 2004; Smith *et al.* 2007).

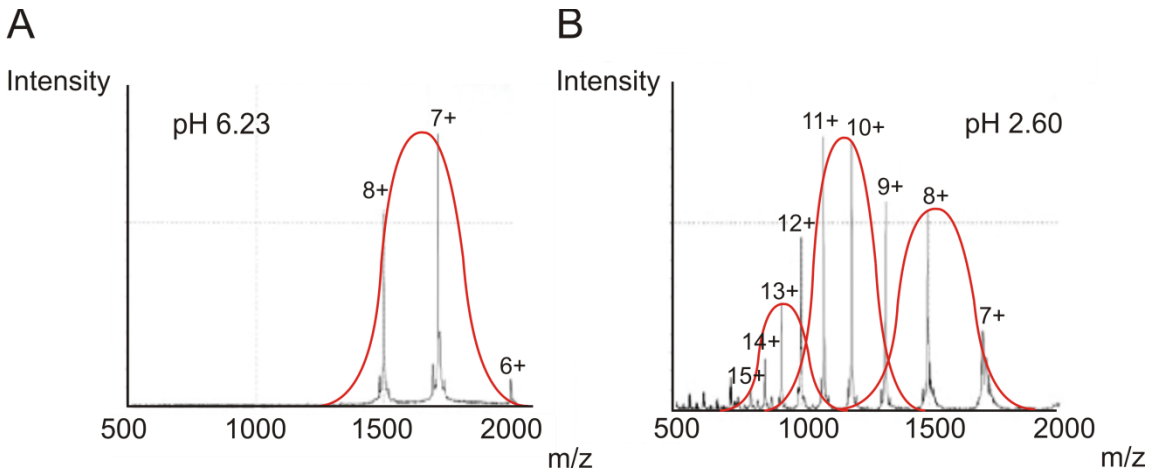


Figure 3.1.5 - ESI mass spectra of β_2m at A) pH 6.23 and B) pH 2.6. The charge state distributions corresponding to the different β_2m conformations are highlighted as red Gaussian curve. This Figure was adapted from Smith *et al.* (2007).

Recent advances in IMS-MS have enabled detailed structural characterisation of co-populated amyloidogenic intermediates that is not possible by ESI-MS alone. The peptide Aβ, whose assembly into amyloid fibrils is associated with Alzheimer's disease, is just one of many examples that have been studied extensively by ESI-IMS-MS (Bernstein *et al.* 2005; Berstein *et al.* 2009; Kloniecki *et al.* 2011; Gessel *et al.* 2012a; Gessel *et al.* 2012b). Recent work in this area by Bowers and co-workers has suggested two alternative pathways of fibril formation for the peptides $A\beta_{40}$ and

$A\beta_{42}$. $A\beta_{40}$ forms compact tetramers that slowly convert into amyloid fibrils, whereas $A\beta_{42}$ forms hexamers and dodecamers on the same time-scale, consistent with its more aggregation-prone tendency (Bernstein *et al.* 2009).

ESI-IMS-MS has also been used to separate structurally distinct β_2m monomeric conformations based on their shape and charge (Smith *et al.* 2007). The compact conformer present at low pH is comparable by collision cross-section calculations to the native β_2m monomer structure (Smith *et al.* 2009) as characterised by NMR (Verdone *et al.* 2002; Eichner *et al.* 2011) and X-ray diffraction (Trinh *et al.* 2002) showing that significant structure is still present to some extent under denaturing conditions.

ESI-IMS-MS can also be used to separate individual non-covalently bound oligomeric species from the highly heterogeneous samples present during amyloid formation. This technique has advantages over other biochemical techniques such as fluorescence or light-scattering which only report an average of all the oligomeric species present in solution. ESI-IMS-MS has been used previously to observe small β_2m oligomers on-pathway to long, straight fibril formation at pH 2.5 *in vitro*; however, no large oligomers (> hexamer) were observed under these conditions (Smith *et al.* 2010). This lack of higher ordered species observed combined with numerical modelling studies suggests six non-covalently bound β_2m monomers come together to form a nucleus which rapidly extends by monomer addition into mature amyloid fibrils (Xue *et al.* 2008; Smith *et al.* 2010). Collision cross-section measurements of the β_2m oligomers at low pH and ionic strength have allowed further structural insight into β_2m amyloid fibril formation. The collision cross-sections of the β_2m oligomers observed are consistent with a stacked conformation of end-to-end β_2m monomers rather than an alternative more compact, globular arrangement (Smith *et al.* 2010) (Figure 3.1.6).

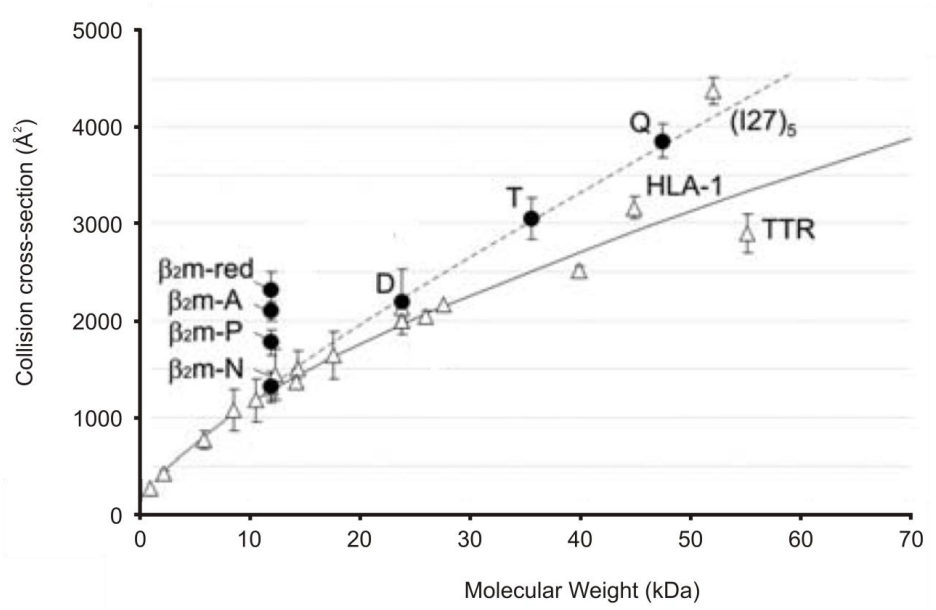


Figure 3.1.6 - Collision cross-sections of β_2m monomer and oligomers at pH 2.5 (black circles) along with a range of protein standards (open triangles). The cross-section of the disulphide bond-reduced β_2m monomer (β_2m-red), elongated/acid unfolded (β_2m-A), partially compact (β_2m-P), compact/native (β_2m-N) are shown with the average cross-sections of all the dimer (D), trimer (T) and tetramer (Q) charge state ions. The linear I27 concatanated pentamer, $(I27)_5$ and more compact protein standards human leukocyte antigen complex 1 (HLA-1) and transthyretin (TTR) as analysed at pH 6.5 are highlighted for comparison with the β_2m oligomer cross-sections. The solid line represents the cross-sections expected for a range of molecular weights assuming a sphere of density 0.44 Da/A^3. The dotted line shows the assembly of β_2m. This Figure was adapted from Smith *et al.* (2010).

Here, mass spectrometry in parallel with other biochemical techniques was used to unravel structural details at the molecular level of the β_2m oligomeric intermediates involved in amyloid fibril formation at pH 2.5. By creating single point mutations throughout the β_2m sequence and subsequent analysis of the fibril formation kinetics, new residues important in β_2m fibrilogenesis were revealed building on previous experiments by Routledge *et al.* and Platt *et al.* (Platt *et al.* 2008; Routledge *et al.* 2009). One variant in particular, H51A, on further analysis by ESI-IMS-MS revealed striking differences not only in the oligomer number compared to wild-type β_2m, but also differences in the structural dynamics of the oligomers formed consistent with a role in impeding amyloid fibril formation.

3.2. Amyloid

Results and Discussion

3.2.1 β_2m mutagenesis

Although several hypotheses have been put forward, the nature of the interfaces between β_2m monomers within different oligomeric species during fibril formation is yet to be determined. To unravel details of these structural intermediates, mutagenesis was performed on a variety of residues throughout the β_2m sequence (Figure 3.2.1). We hypothesised that a mutated residue present in an oligomeric interface could affect the stability of that oligomer, and may result in differences in its fibril formation kinetics compared with the wild-type protein. Alternatively, if the mutated residue was not involved in oligomer formation, the oligomer stability should remain comparable with that of the wild-type protein. In all cases, mutagenesis to alanine was performed due to its small side chain and its inability to undergo hydrogen bonding and partake in electrostatic interactions.

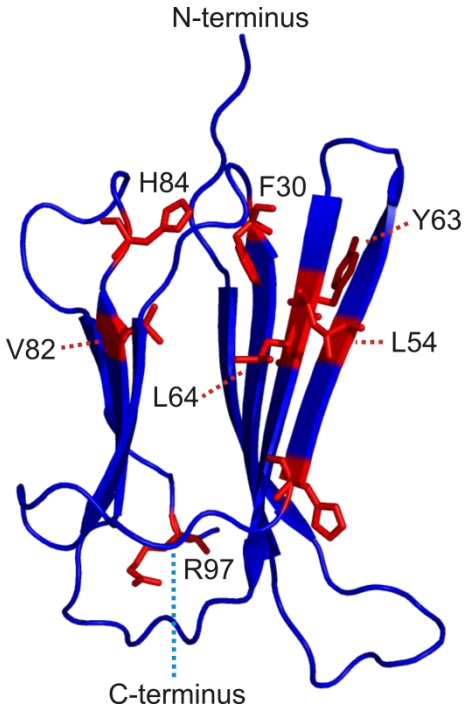


Figure 3.2.1 – Native β_2m structure (blue ribbon) (PDB file:2YXF) (Iwata *et al.* 2007) showing the position of the variants (red sticks) used to probe the structural interface of the β_2m oligomeric intermediates.

All β_2m variants were expressed and purified by ion exchange chromatography followed by size exclusion chromatography (SEC) (Materials and Methods Section 3.4.3). Wild-type β_2m, L64A, H84A and ^{15}N H51A were expressed and purified by myself. The F30A, Y63A, L54A and R97A proteins were kindly provided by Thomas Jahn (formerly University of Leeds). The H51A and V82A proteins were kindly provided by Katy Routledge (formerly University of Leeds). All variants were confirmed to be pure by SDS-PAGE, monomeric by SEC (data not shown) and each was identified by their correct molecular weight using ESI-MS by myself (Materials and Methods Section 3.4.5, Table 3.3.3). One example of a typical purification is shown in Figure 3.2.2 below.

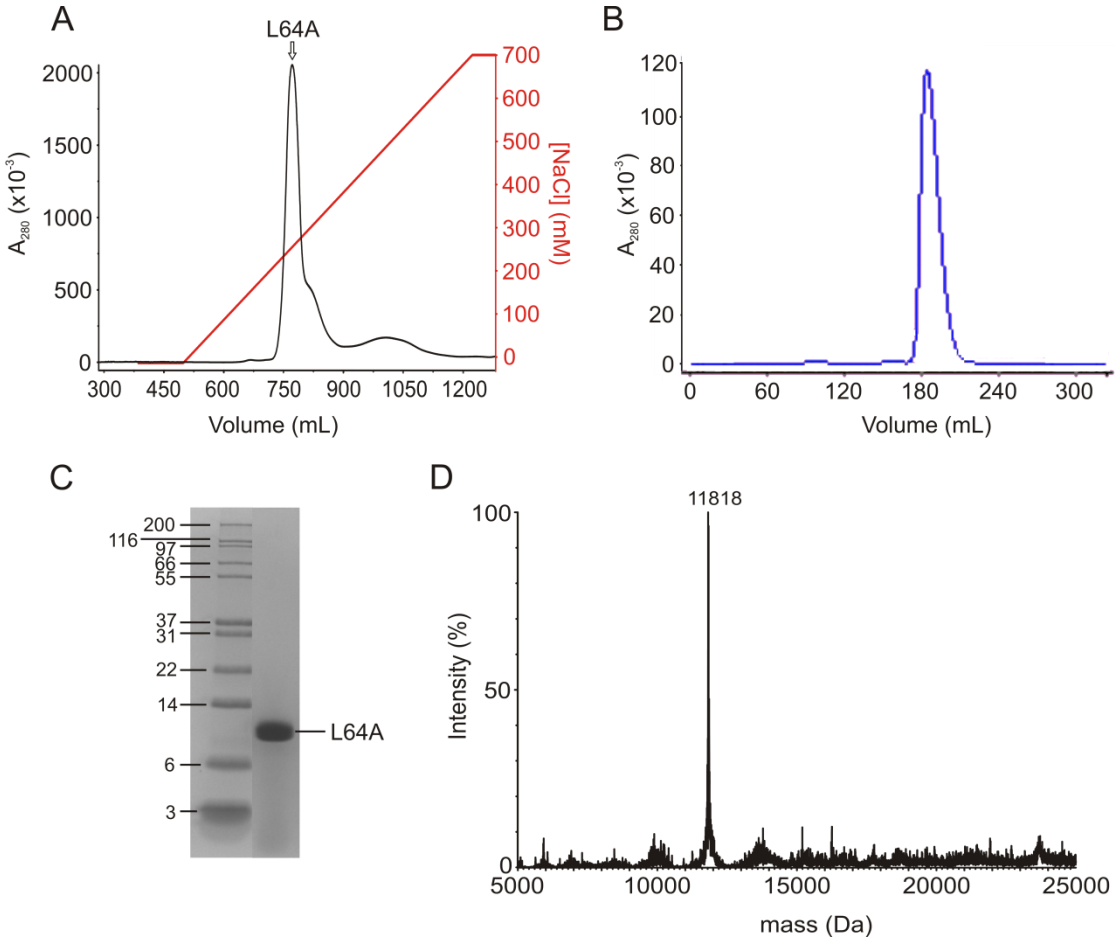


Figure 3.2.2 - Expression and purification of the β_2m variant, L64A. A) Anion exchange chromatography where the first peak corresponds to L64A; B) SEC showing a single monomeric peak eluting off the Superdex 75 column; C) SDS PAGE showing pure β_2m L64A and the molecular weight markers (Da); D) ESI-MS of purified L64A β_2m of correct molecular weight (expected mass = 11818 Da).

An example of the data obtained from protein purification is shown for the β_2m variant L64A (Figure 3.2.2). Two peaks typically eluted off the column during anion exchange chromatography (Figure 3.2.2A), the first larger peak corresponded to the protein of interest and the second peak to other protein impurities present in the sample (data not shown). SEC data confirmed the L64A β_2m variant was monomeric, with only a single peak eluting off the Superdex 75 column (Figure 3.2.3B). Other peaks were occasionally observed during SEC with the other β_2m variant corresponding to dimer and other higher order β_2m oligomers (data not shown). In these cases, only the peak corresponding to the monomeric protein was collected for further experiments. The final purity of the β_2m L64A variant was confirmed by SDS-PAGE (Figure 3.2.2C) and ESI-MS confirmed the molecular weight of the β_2m variant protein to within a mass accuracy of 0.01 % compared with that predicted based on the protein's amino acid sequence (Figure 3.2.2D) (Materials and Methods Section 3.4.5, Table 3.4.3).

3.2.2 Fibril formation of β_2m variants at pH 2.5

All β_2m variants were selected since they have been shown previously to form fibrils at pH 2.5 in 25 mM sodium phosphate, 25 mM sodium acetate (Routledge *et al.* 2009). Since mass spectrometry will be the primary tool in this thesis to study the oligomers formed during fibril formation, a volatile buffer compatible with ESI-MS was needed in which amyloid fibrils form. Fibril formation, therefore, was performed with the wild-type protein and all the other β_2m variants in 100 mM ammonium formate pH 2.5, incubating at 37 ºC with agitation at 600 rpm (Smith *et al.* 2010). On analysis by negative stain EM, all β_2m variants formed fibrils with a long straight morphology comparable to that of wild-type β_2m in 100 mM ammonium formate pH 2.5 (Figure 3.2.3). The volatile mass spectrometry buffer used here does not affect the fibril morphology and, therefore, was the buffer used throughout this thesis chapter.

Using low resolution techniques such as EM, very few differences are observed in the final fibril morphology between the β_2m variants, showing that these single amino acid substitutions do not grossly alter the course of amyloid fibril formation.

If any amino acid side chains altered are present in the oligomer interfaces or are involved in oligomer formation, differences in the fibril formation kinetics might be expected. Experiments by Routledge *et al.* have shown the lag times of fibril formation to increase by up to 3-fold in the case of the H51A variant compared with that of the wild-type β_2m protein (Routledge *et al.* 2009).

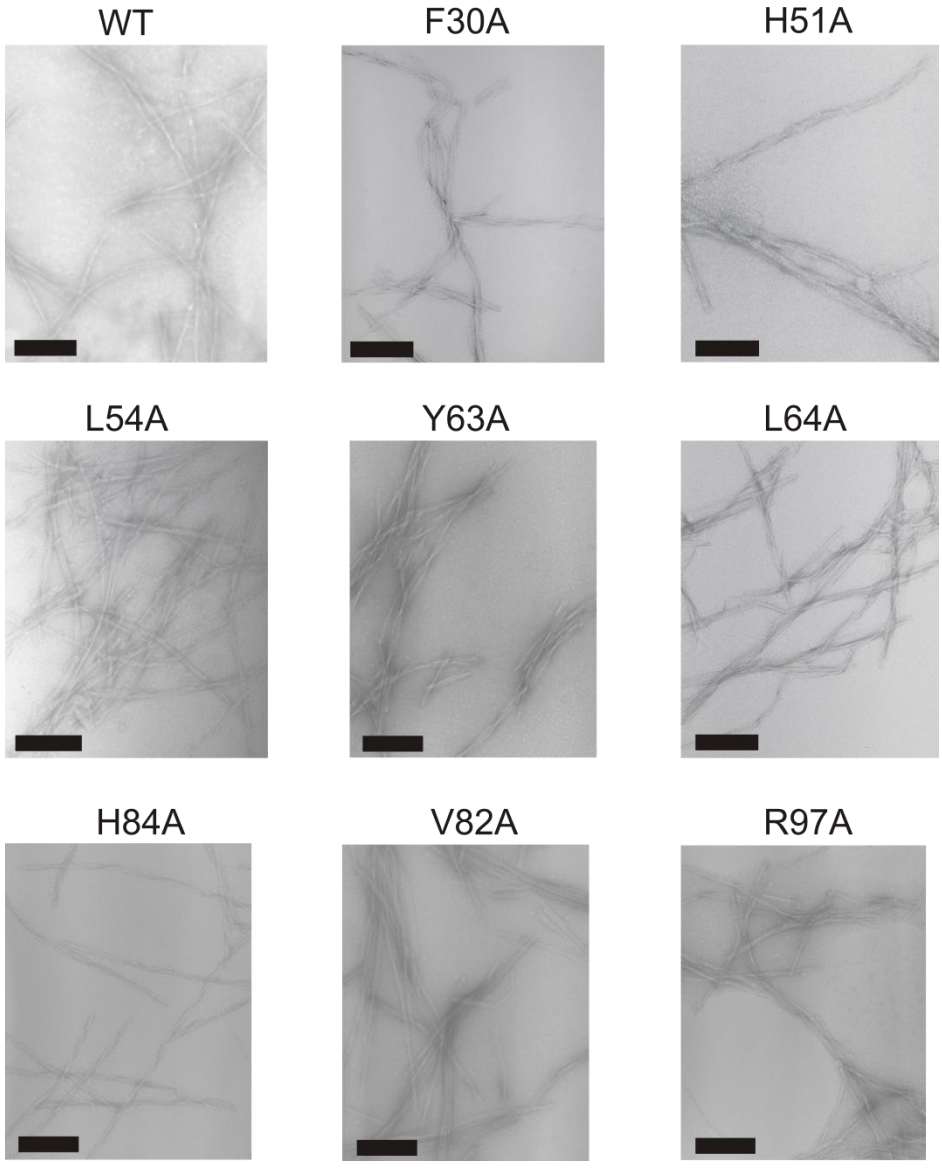


Figure 3.2.3 – Negative stain electron microscopy of fibrils formed in 100 mM ammonium formate pH 2.5, 37 ºC, 600 rpm, by wild-type β_2m and the β_2m variants, F30A, H51A, L54A, Y63A, L64A, H84A, V82A and R97A. The scale bar in each case represents 200 nm.

3.2.3 Fibril formation kinetics of β_2m variants

Thioflavin T (ThT), an amyloid fibril binding dye (see Section 3.1.2), was added to fibril growth assays of each variant to monitor the rate of fibril formation (Materials and Methods Section 3.4.8). Due to difficulties in reproducing lag times of fibril formation accurately between experiments, wild-type β_2m was analysed simultaneously with each of the β_2m variants in parallel on each day of analysis and the fold change in lag time (Materials and Methods Section 3.4.9) for each variant was compared with that of the wild-type β_2m (Figure 3.2.4). Of the β_2m variants studied, F30A, Y63A, L64A, H84A and R97A all have a lag time comparable to that of the wild-type β_2m (Figure 3.2.4) suggesting these mutations do not affect the rate of amyloid fibril formation at pH 2.5.

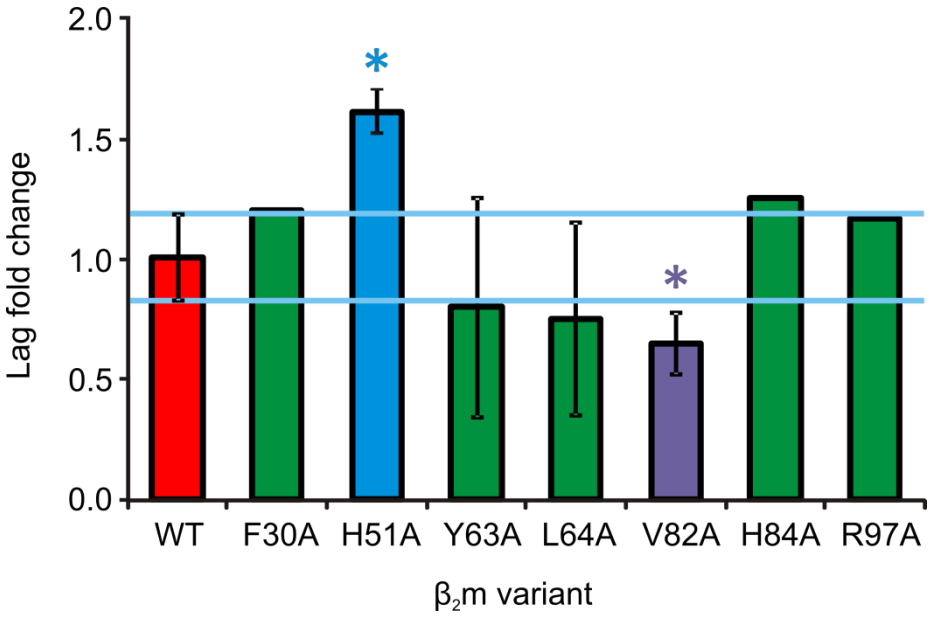


Figure 3.2.4 - Comparison of the kinetics of fibril formation of the variants F30A, H51A, Y63A, L64A, V82A, H84A and R97A with respect to wild-type (red). The fold change in lag time compared to wild-type β_2m is shown for comparison. The horizontal blue lines represent ± 1 standard deviation from 5 repeats of the wild-type β_2m proteins lag time. The variants, H51A and V82A, which show repeatedly different lag times compared with the wild-type β_2m, are highlighted in blue and purple, respectively, and marked with an asterisk. The error bars (where shown) represent ± 1 standard deviation. All experiments were performed in 100 mM ammonium formate pH 2.5, shaking at 600 rpm (Materials and Methods 3.4.8).

The removal of a charged residue towards the C-terminus of β_2m in the variant, R97A, has previously been shown to enhance the ability of β_2m to form fibrils at pH 7.0 (Jones *et al.* 2003). At pH 2.5 the lag time of fibril formation is comparable to that of the wild-type protein (Figure 3.2.4). This is most likely due to the lack of native structure in the C-terminus of the monomeric β_2m at the low pH used in this study. Val82 was the only residue that decreased the lag time of fibril formation on substitution to alanine. Val82 is located close to the disulphide bond (Cys80) and has been shown to slow fibril formation on mutation to proline (Chiba *et al.* 2003). The nature of the backbone and side-chains around the disulphide bond in β_2m appears to be important in determining the rate of amyloid fibril formation, even at low pH.

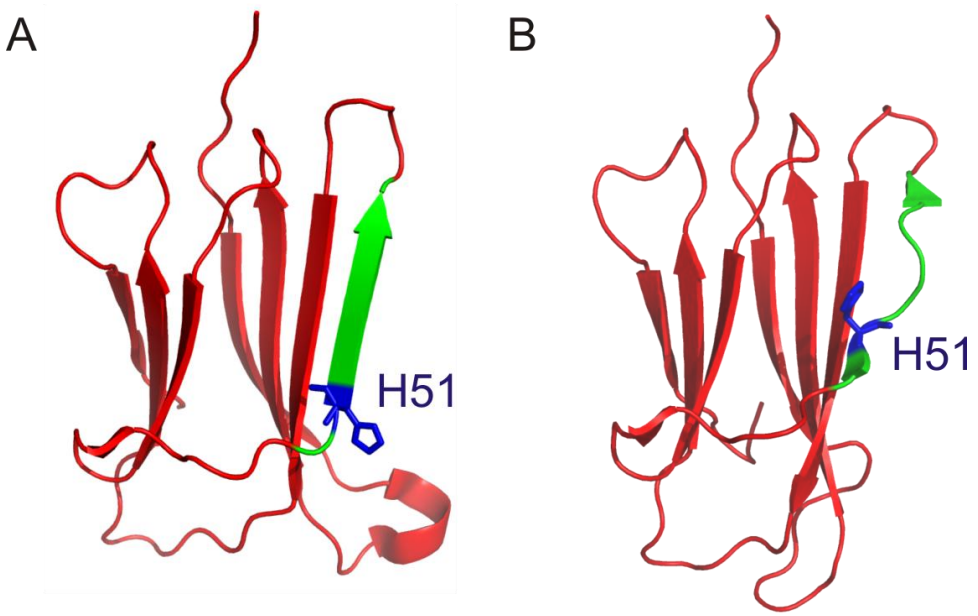


Figure 3.2.5 - Structures of β_2m highlighting the different conformations of the D stand (green) between free β_2m (PDB file: 1LDS) (Trinh *et al.* 2002) and the MHC 1 bound form (PDB file: 3AM8). The location of the His51 is highlighted in blue on each structure.

The following Sections (3.2.3 to 3.2.6) focus in more detail on the β_2m variant, H51A, since this is the only variant shown here to increase the lag time of fibril formation significantly (Figure 3.2.4). His51 is located on the D strand of native β_2m and is involved in the non-covalent binding of β_2m to MHC 1 (Figure 3.2.5). All crystal structures to date show a bulged D-strand in complex with MHC 1 with the residue Asp53 (the β-bulge residue at the centre of the D strand) hydrogen bonded to Arg35 on the MHC 1 (Shields *et al.* 1998). On release from the MHC 1 complex, this D-strand becomes solvent exposed and thus may be involved in initiating β_2m

aggregation. Crystal structures of free β_2m have shown that the D-strand can adopt a straight conformation (Figure 3.2.5A), with this regular edge strand thought to be involved in promoting aggregation (Trinh *et al.* 2002). More recent observations have suggested the D-strand adopts a bent conformation in the β_2m oligomeric state (Calabrese *et al.* 2008). These studies were performed at neutral pH upon addition of Cu^{2+}, and not under the acidic conditions used in this thesis. NMR studies additionally have indicated that the D-strand is very flexible in solution (Okon *et al.* 1992; Verdone *et al.* 2002) with molecular dynamics simulations showing the straight D-strand to be populated increasingly on decreasing pH (Park *et al.* 2006). On transition of the D-strand from the bent to the straight conformation (Figure 3.2.5), His51 rotates 180º. Substitution of this residue to alanine is thought to cause slight structural perturbations in the D-strand conformation, which results in increased lag time of amyloid formation observed in this thesis (Figure 3.2.6).

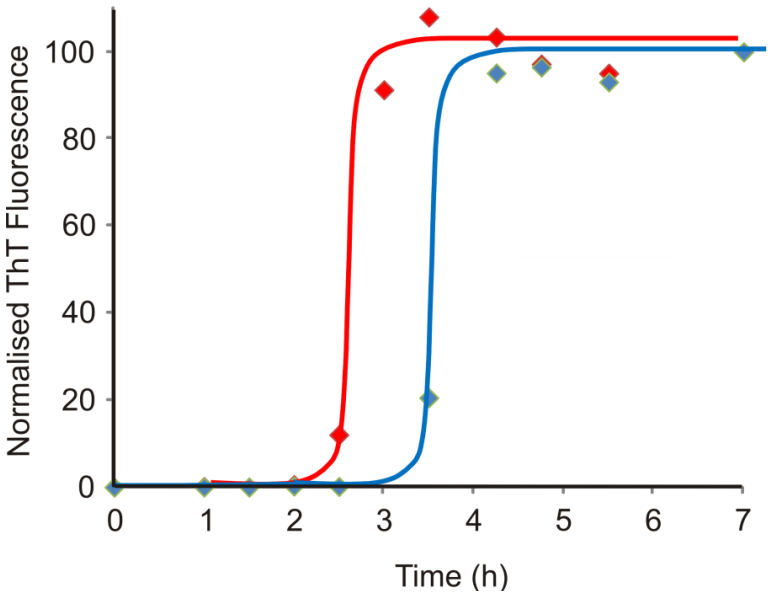


Figure 3.2.6 - Plot of the normalised ThT kinetics of fibril formation for β_2m wild-type (red) and H51A (blue). One replicate is shown for simplicity and a sigmoidal curve fitted to guide the eye through the data points. Data was acquired using the PTI spectrofluorimeter (Materials and Methods Section 3.4.8.2).

At neutral pH, native-like features are present in the β_2m amyoidogenic intermediate (Jahn 2006). Under acidic conditions, β_2m is unfolded with very little secondary structure persists (Plattt *et al.* 2005; Smith *et al.* 2010). The extent as to which the conformation of the D strand in β_2m plays a role at pH 2.5 in amyloid fibril formation is unknown and remains to be determined.

ThT fluorescence is a powerful technique for analysing fibrillation kinetics, but it does not report on the oligomeric species formed during the lag phase of fibrillogenesis (Figure 3.2.6). Mass spectrometry is an ideal technique for the analysis of these intermediates since it can detect individual oligomeric species from within a heterogeneous mixture, but also give an idea of the shape of these oligomers through the use of IMS-MS. ESI-MS/ESI-IMS-MS was used, therefore, to unravel any structural differences in the oligomeric intermediates between H51A and wild-type β_2m formed at pH 2.5 which may explain the differences in ThT kinetics observed (Figures 3.3.4 and 3.2.6).

3.2.4 Comparison of ESI Mass Spectra from wild-type β_2m and H51A

Previous ESI-MS studies have analysed the conformers formed within monomeric β_2m at different pH values together with their role in initiating fibril formation (Borysik *et al.* 2004; Borysik *et al.* 2004; Smith *et al.* 2007). Both β_2m (Figure 3.1.6) and the β_2m variant H51A (data not shown) populate three distinct monomeric populations at pH 2.5, an unfolded, partially folded and compact state. Here, we chose to investigate β_2m oligomers on-pathway to amyloid fibril formation and their role in fibrillogenesis. On initial dilution of lyophilised wild-type β_2m into 100 mM ammonium formate pH 2.5 oligomers up to trimer were instantly observed in the ESI mass spectrum (Figure 3.2.7). In contrast, with the H51A variant, many more higher ordered oligomers were present in the H51A *m/z* spectrum compared with those detected for wild-type β_2m (Figure 3.2.7). For H51A, charge state distributions corresponding to oligomers up to a pentamer in size are observed clearly suggesting the H51A variant is more highly aggregation prone compared with wild-type β_2m at low pH.

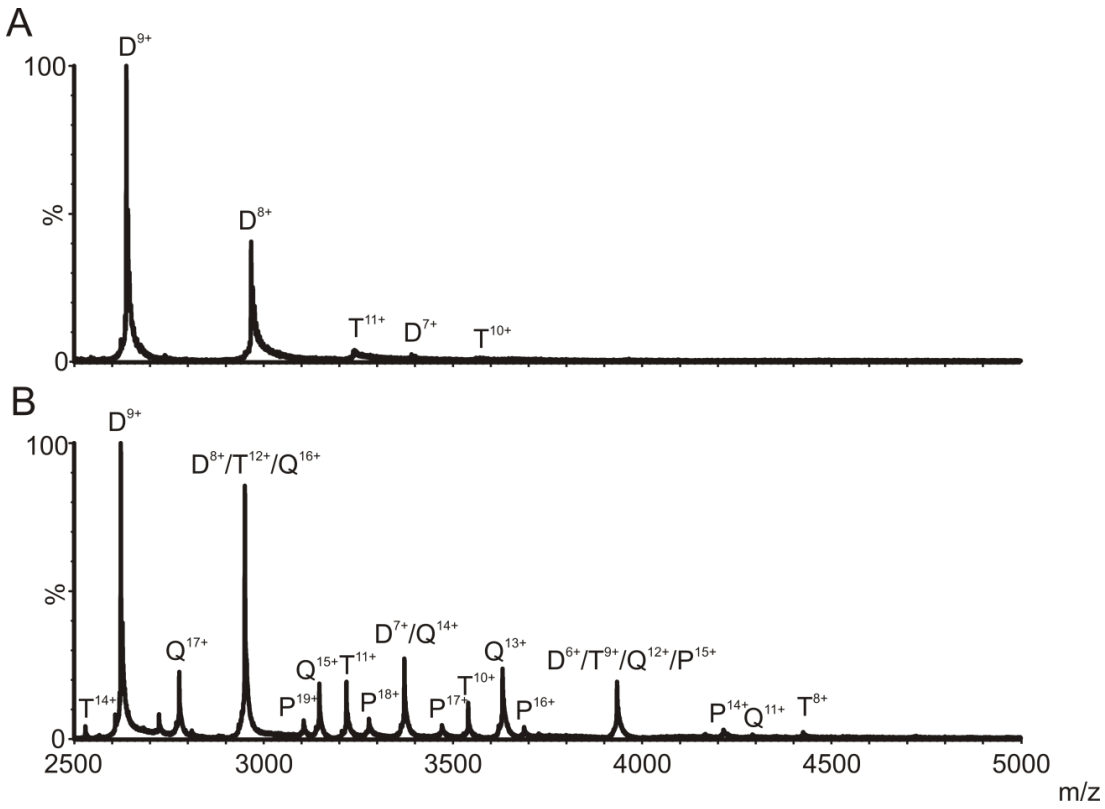


Figure 3.2.7 - ESI-MS mass spectra of wild-type β_2m (A) and H51A (B) over 2500-5000 *m/z* range in 100 mM ammonium formate pH 2.5. The charge states corresponding to the dimer (D), trimer (T), tetramer (Q) and pentamer (P) are highlighted. The spectra were obtained 1 min after dissolution of each protein into the MS buffer at pH 2.5.

The ESI-MS mass spectra of wild-type β_2m and H51A show several oligomeric species contributing to the same *m/z* ratio. For example, at *m/z* 2949.5 in the H51A mass spectrum, co-populated species are present corresponding to the 8+ dimer, 12+ trimer and 16+ tetramer ions, and at *m/z* 3932.2 the 6+ dimer, 9+ trimer, 12+ tetramer and 15+ pentamer ions (Figure 3.2.7B). The added dimension of ion mobility spectrometry was used to separate these co-populated species. The IMS-MS driftscope plots of wild-type β_2m and the β_2m variant H51A in 100 mM ammonium formate pH 2.5 are shown in Figure 3.2.8. The data show the power of IMS in separating multiple populations within a single sample, and as a result, each oligomer was assigned uniquely based on its *m/z* value and isotopic distribution. On comparison of the arrival time distributions for the oligomeric complexes, the higher order species were found to have a shorter drift time compared with lower order species of the same *m/z*. Charge also has an effect on the arrival time distribution,

with lower charged ions taking longer to reach the detector than more highly charged species.

For wild-type β_2m and H51A, oligomeric species up to hexamer are observed one minute after dilution into pH 2.5 buffer (Figure 3.2.8), albeit these larger species being in very low abundance (< 5% of the total ion count) (Figure 3.2.7). The additional IMS separation can not only distinguish between oligomers of same *m/z*, but can also enable identification of higher order oligomers by separating them from ions coming from the buffer.

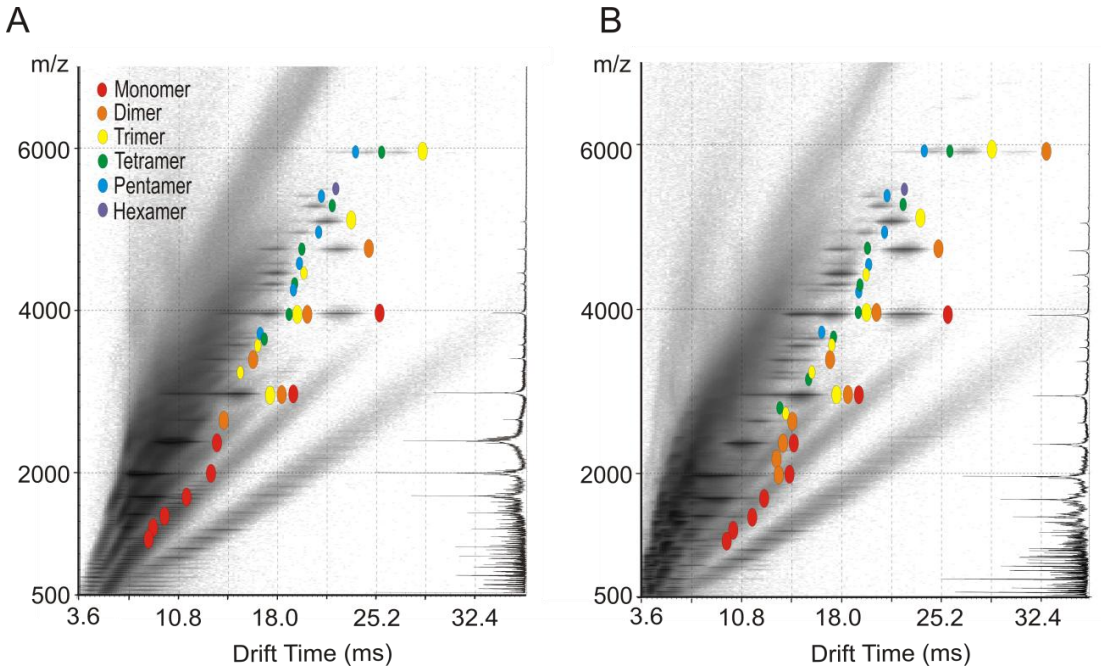


Figure 3.2.8 - ESI-IMS-MS of A) wild-type β_2m and B) the H51A β_2m variant in 100 mM ammonium formate pH 2.5. The mass spectrum in each case is shown on the right and the oligomers highlighted (monomer-red, dimer-orange, trimer-yellow, tetramer-green, pentamer-blue, hexamer-purple). The spectra were obtained 1 min after dissolution of each protein into the MS buffer at pH 2.5.

The collision cross-sectional areas for all charge states arising from all of the wild-type and H51A oligomers were estimated from the experimental data (Materials and Methods Section 3.4.10.2), and the lowest charge states shown in Figure 3.2.9 for comparison. The cross-section of the wild-type and H51A oligomers were identical within the experimental error (5 %). Consistent with previous characterisation of β_2m oligomers by ESI-IMS-MS (Smith *et al.* 2010), the oligomers adopt a more

extended structure compared with that of globular proteins. The cross sections of an end-to-end linked concatamer of five titin I27 Ig domains $(I27)_5$ (52 kDa; 4,297 $Å^2$) and the globular transthyretin tetramer (TTR) (47 kDa; 2,900 $Å^2$) were determined by Dr. David Smith experimentally (formerly University of Leeds) (Smith *et al.* 2010) and are shown as a comparison. The wild-type β_2m tetramer has a collision cross-section of 4172 $Å^2$ and the collision cross-section of the H51A tetramer is 4067 $Å^2$, which is consistent with a more expanded than globular structure, comparable with previously published data (Smith *et al.* 2010).

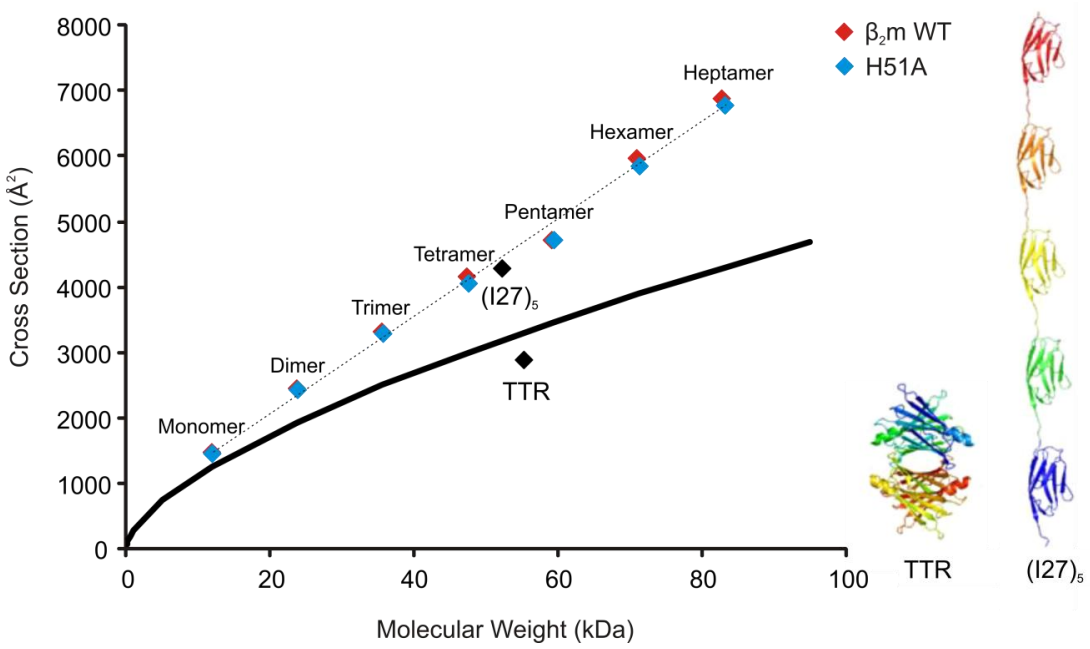


Figure 3.2.9 - Collision cross-sectional areas for the lowest charge state ions measured by ESI-IMS-MS for β_2m wild-type and H51A monomer. and their oligomers. The cross-sections for the transthyretin tetramer (TTR) and $(I27)_5$ (a linear pentamer of I27 domains) as determined by Dr. David Smith (formerly University of Leeds) are shown for comparison (Smith *et al.* 2010) along with their structures (TTR PDB file:1ICT, $(I27)_5$ PDB file:1TIT). The solid line represents the expected cross-sections calculated for a range of molecular weights assuming a perfect sphere of density 0.44 Da/$Å^3$. The dotted line is simply to guide the eye through the data points.

Although more oligomers of higher order are populated initially at low pH for the H51A β_2m variant compared with wild-type β_2m, it is interesting to compare how the oligomer distribution changes throughout the lag time of fibril formation. An experiment was set up whereby wild-type β_2m or the H51A variant were incubated separately at 37 ºC, shaking at 600 rpm, in 100 mM ammonium formate pH 2.5. At

various time points, samples were removed for analysis by ThT fluorescence analysis (to determine how far the sample was through the lag-phase of fibril formation) and were simultaneously analysed by ESI-IMS-MS (Materials and Methods Sections 3.4.8.2 and 3.4.10.2). The ESI-IMS-MS mass spectra for H51A and wild-type β_2m at 50 % through the lag phase of fibril formation are shown in Figure 3.2.10.

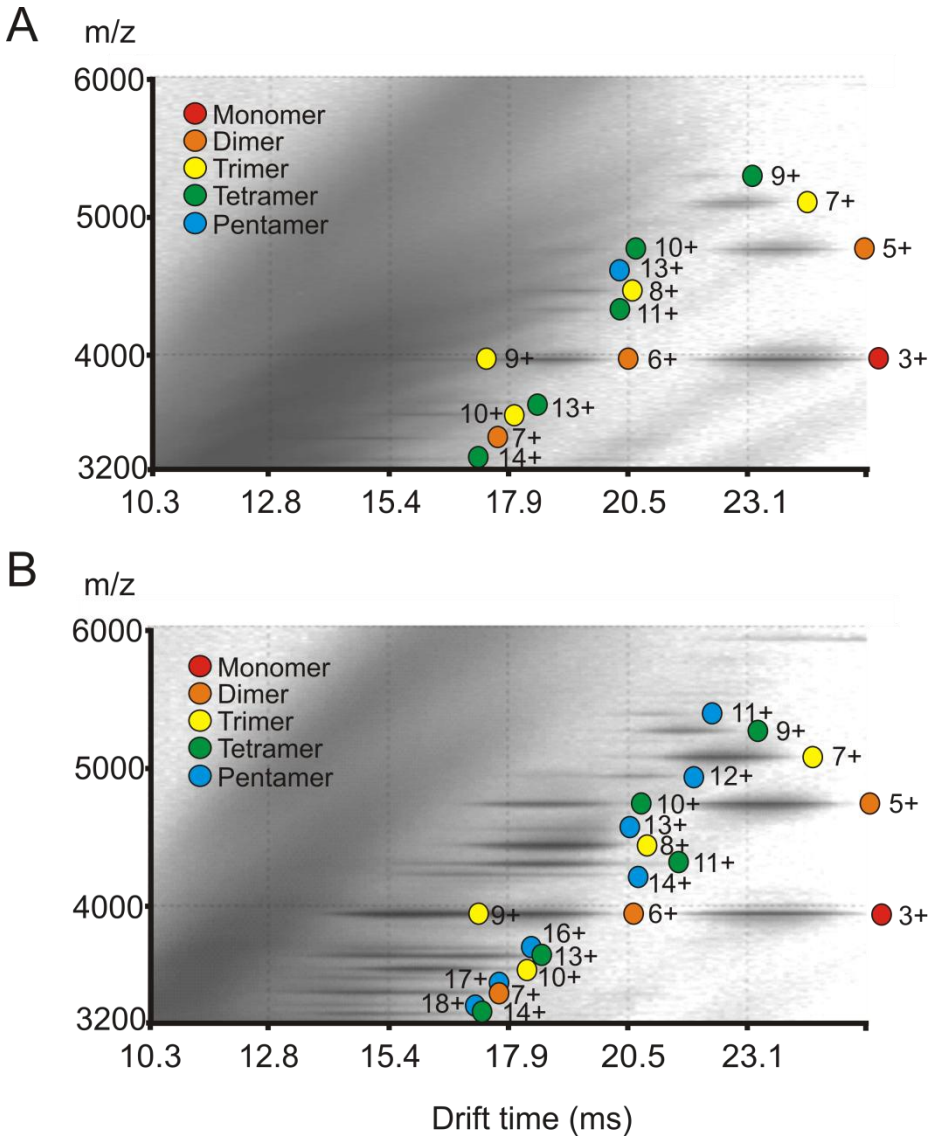


Figure 3.2.10 - ESI-IMS-MS of A) wild-type β_2m and B) the H51A β_2m variant in 100 mM ammonium formate pH 2.5 at 50 % of the lag phase of amyloid fibril formation (1 h and 1.5 h, respectively). The monomer and oligomers are highlighted in both Driftscope plots (monomer-red, dimer-orange, trimer-yellow, tetramer-green, pentamer-blue).

On analysis of the ESI-IMS-MS driftscope plot for wild-type β_2m, oligomers up to tetramer are readily observed (Figure 3.2.10A). One charge state is also visible corresponding to the 13+ pentamer; however, this is very low in abundance. In comparison with the same ESI-IMS-MS driftscope plot taken at 0 % through the lag phase of fibril formation, the oligomer population has decreased (for example, charge state distributions corresponding to the pentamer and hexamer are no longer observed 50 % into the lag phase) suggesting these oligomers have rapidly extended into pre-fibrillar species or into amyloid fibrils (Figures 3.2.8A and 3.2.10A). At the end of the lag phase, all oligomer peaks disappear and the sample becomes challenging to spray into the gas phase, presumably due to the high population of insoluble amyloid fibrils present. This suggests the oligomers observed are on-pathway or transient off-pathway species formed during amyloid fibril formation since very little remain detectable at the end of the lag phase post-fibril formation.

The higher order oligomers are still populated for the H51A variant at 50 % of the lag phase (Figure 3.2.10B). It is possible that these oligomers are slower to convert into amyloid fibrils and hence remain visible by ESI-IMS-MS for longer time periods compared with wild-type β_2m. These subtle differences in the oligomer structure may not be evident from low resolution cross-section measurements alone. Therefore, the oligomers formed from wild-type β_2m and H51A were investigated further by CID-MS and subunit exchange dynamics to determine whether any differences in the stability of the oligomers could explain the differences in the lag time of fibre assembly of the two proteins.

3.2.5 Gas Phase Oligomer Stability using ESI-CID-MS

Since no structural differences were apparent in the oligomer collision cross-sections between wild-type β_2m and the H51A variant by ESI-IMS-MS, the gas phase stability of the oligomers formed was further investigated by the use of CID-MS. If His51 is involved in the dimer interface, then the gas phase dimer stability may be affected. Ions corresponding to the 9+ dimer of wild-type β_2m and H51A were separately selected in the quadrupole of the Synapt HDMS mass spectrometer and subjected to CID fragmentation in the trap region. On increasing the collision energy

applied to the dimer 9+ ions, product ions were detected corresponding to the wild-type β_2m or H51A monomers (Figures 3.2.11, 3.2.12). The fragmentation profiles for wild-type and the H51A dimer are identical suggesting both dimers have the same gas phase stability under the ESI-MS conditions used here.

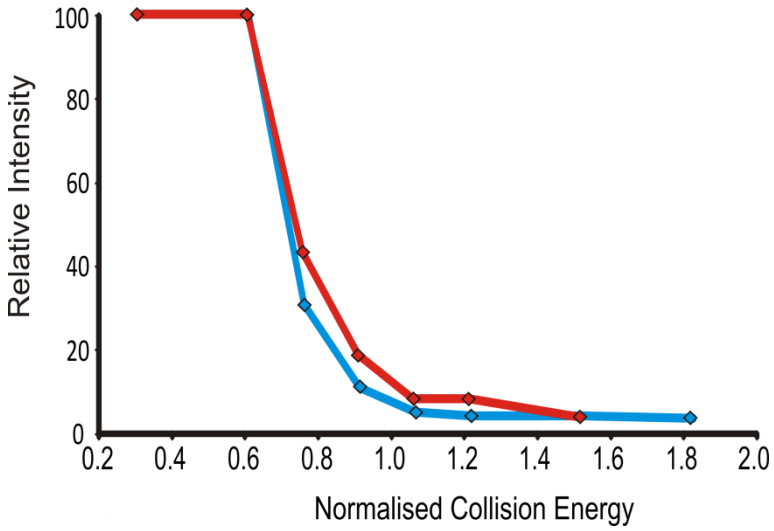


Figure 3.2.11 - ESI-CID-MS fragmentation profile for the wild-type dimer 9+ (*m/z* 2637) (red) and the H51A dimer 9+ (*m/z* 2622) (blue). One example is shown for simplicity. In each case the dimer 9+ was selected in the transfer region of the Synapt HDMS for fragmentation by CID.

The products of CID-MS have been used to unravel structural details including subunit stoichiometries, in addition to detecting ligand binding to multi-complex protein assemblies (McCammon *et al.* 2005). Typically during CID, the internal energy provided to dissociate the protein complex causes one subunit to unfold, taking the majority of the charge with it, before it is finally released and the product ions detected (Sciuto *et al.* 2011). This is termed asymmetric division since the charge is not equally distributed between the product ions in the *m/z* spectrum (Figure 3.2.12B). Alternatively, although relatively uncommon, symmetrical division can occur. In this case, the energy needed to dissociate the complex is much less than the energy needed to unfold one of the subunits, resulting in product ions of equal charge in the *m/z* spectrum (Figure 3.2.12A) (Dodds *et al.* 2011). On comparison of surface induced dissociation (SID) and CID experiments on protein complexes, Wysocki and co-workers suggested that analysis of the charge distribution in the product ion spectrum can reveal information about the nature of the protein-protein

interaction present (Dodds *et al.* 2011). For example, if electrostatic interactions are present at the protein-protein interface which are particularly stable in the gas phase, subunit unfolding is likely to occur resulting in asymmetric division. Alternatively, if only hydrophobic interactions are present, then symmetric division will be more favourable (Dodds *et al.* 2011).

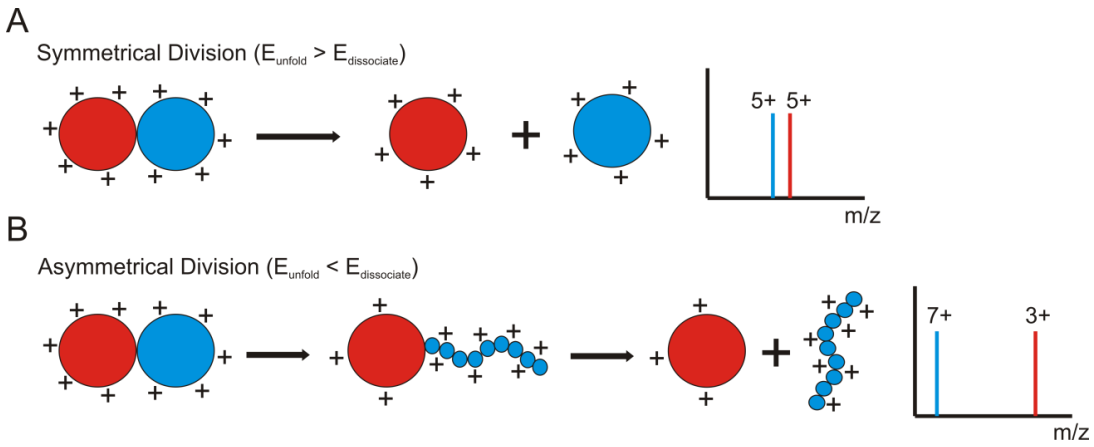


Figure 3.2.12 - Schematic of dimer CID dissociation behaviours; A) symmetrical division where the energy for subunit unfolding is greater than that of subunit dissociation; B) asymmetrical division where the energy for subunit dissociation is greater than subunit unfolding resulting in one subunit unfolding prior to subunit dissociation, taking majority of the charge with it. The two different monomers are shown as blue and red balls.

The product ions resulting from CID-MS of the wild-type β_2m and H51A 9+ and 11+ dimers are shown in Figure 3.2.13. In all cases, the most abundant product ions are the 4+ and 5+ monomer, and the 5+ and 6+ monomer for the dimers 9+ and 11+ respectively. This symmetric division suggests that the wild-type and H51A dimers are held weakly together in the gas phase by predominantly hydrophobic, rather than electrostatic interactions. Further experimental evidence is needed to confirm the nature of this interaction. Since the dimers here are formed at pH 2.5 where little residual structure is present in the β_2m monomer, it is possible the lack of subunit unfolding upon CID activation is simply due to the β_2m monomer already being slightly unfolded in the dimer and thus further unfolding cannot be observed. The β_2m trimer was also observed to undergo symmetrical division (data not shown), suggesting the nature of the monomer-monomer interface and dimer-monomer interface could be conserved.

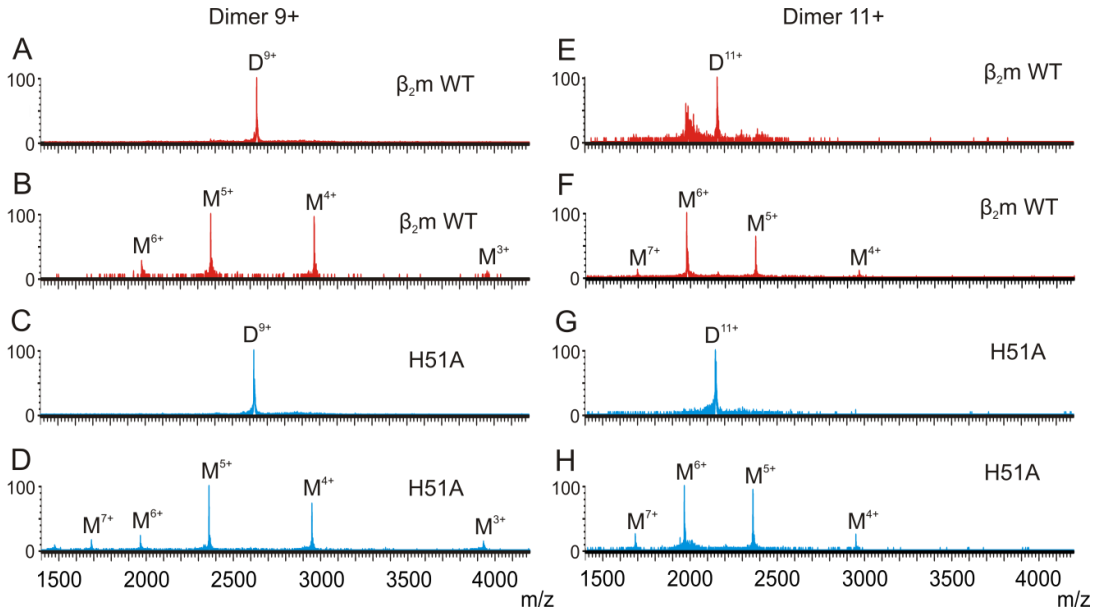


Figure 3.2.13 – ESI-CID/MS Product ion spectrum of wild-type β_2m (red) and H51A (blue) of dimer 9+ (A, B, C, D) and dimer 11+ (E, F, G, H) at a collision energy of 0 eV (A, C, E, G) and 80 eV (B, D, F, H). The ions relating to the dimer and monomer ions are labelled D and M respectively.

Unfortunately, no differences were observed by CID-MS between wild-type β_2m and the H51A β_2m variant (Figures 3.2.11 and 3.2.13). The gas phase stability cannot be used to explain differences in fibrillation kinetics; therefore, the subunit dynamics of the oligomers formed during amyloid fibril formation were next investigated.

3.2.6 Measuring Subunit Dynamics by ESI-MS

The dynamics of the β_2m oligomers formed during fibrillogenesis can be determined using subunit exchange experiments whereby ^{14}N and ^{15}N β_2m oligomers are mixed and the rate of exchange between the two species monitored by ESI-IMS-MS (Smith *et al.* 2010). If the oligomers are dynamic entities and undergo rapid exchange with one another, then mixed oligomers containing both ^{14}N and ^{15}N protein will be present in the mass spectrum (Figure 3.2.14). For example, three peaks would be observed for a rapidly exchanging dimer in a 1:2:1 ratio (^{14}N/^{14}N, ^{14}N/^{15}N and ^{15}N/^{15}N), four peaks for a rapidly exchanging trimer in a 1:3:3:1 ratio (^{14}N/^{14}N/^{14}N, ^{14}N/^{14}N/^{15}N, ^{14}N/^{15}N/^{15}N and ^{15}N/^{15}N/^{15}N) and five peaks for a rapidly exchanging tetramer in a 1:4:6:4:1 ratio (^{14}N/^{14}N/^{14}N/^{14}N, ^{14}N/^{14}N/^{14}N/^{15}N, ^{14}N/^{14}N/^{15}N/^{15}N, ^{14}N/^{15}N/^{15}N/^{15}N and ^{15}N/^{15}N/^{15}N/^{15}N). Alternatively, if the oligomers are static

entities, no exchange will occur and only ions corresponding to all ^{14}N and all ^{15}N proteins will be observed at a 1:1 ratio in the mass spectrum (Figure 3.2.14).

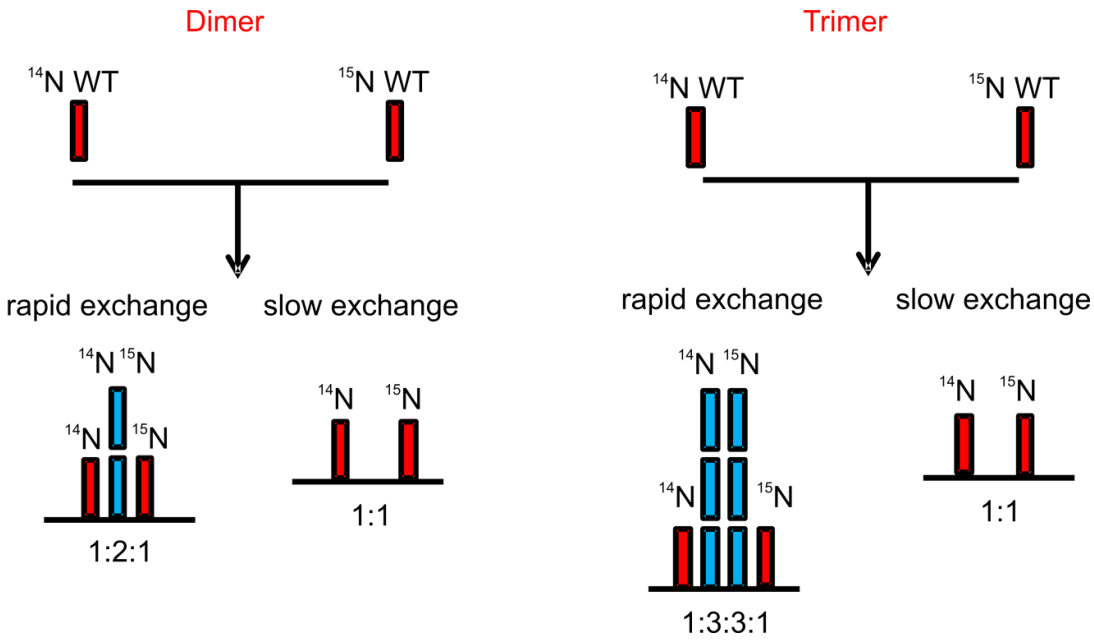


Figure 3.2.14 - Subunit exchange schematic showing the expected $^{14}N/^{15}N$ ratios for fully exchanged dimers and trimers and dimers and trimers that do not undergo exchange. ^{14}N and ^{15}N dimers and trimers are shown in red, mixed $^{14}N/^{15}N$ dimers and trimers are shown in blue. Figure adapted from Smith *et al.* (Smith *et al.* 2010).

^{14}N and ^{15}N wild-type protein were incubated separately in 100 mM ammonium formate pH 2.5 at room temperature for a few minutes to allow an equilibrium to establish, then mixed gently and a mass spectrum acquired after 1 minute (Figure 3.2.15). The peaks corresponding to the dimer 6+, trimer 9+ and tetramer 12+ ions overlap in the *m/z* spectrum, hence, the data for the exchange of these species was discarded. After one minute of incubating ^{14}N and ^{15}N β_2m, the dimer 5+ and 7+ both undergo rapid exchange with one another forming mixed $^{14}N/^{15}N$ dimers as evident from the ESI mass spectrum. This shows the dimers are dynamic entites, rapidly dissociating and associating with other monomers in solution. It is possible that the mixed oligomers observed here do not form in solution and instead form during the ESI process. Previous experiments whereby ^{14}N and ^{15}N wild-type β_2m monomers were sprayed separately but simultaneously into the ESI source suggest this is not the case, since no mixed oligomers were observed using this sample

introduction and ionisation process (Smith *et al.* 2010). The rapidly exchanging dimer is consistent with the CID-MS data (Section 3.2.5) which suggests the monomers are held together by weak hydrophobic interactions in the gas phase. A rapidly exchanging dimer would dissociate very readily as analysed by CID-MS.

The ^{14}N trimer 8+ undergoes a small amount of exchange in solution in comparison to the trimer 11+ which is fully exchanged (as shown by a 1:3:3:1 ratio of $^{14}N^{14}N^{14}N$: $^{14}N^{14}N^{15}N$: $^{14}N^{15}N^{15}N$: $^{15}N^{15}N^{15}N$ in the mass spectrum) after only a few minutes of incubation with ^{15}N β_2m (Figure 3.2.15). This is consistent with previous experimental data that have shown two β_2m trimers present in solution; one lower charged, stable trimer and another highly charged, dynamic trimeric species (Smith *et al.* 2010). In comparison, the wild-type β_2m tetramer undergoes very little exchange in solution suggesting it is more stable and less dynamic than the expanded trimer and all the dimeric species observed. On comparison of all charge states of the dimer, trimer and tetramers, the higher charged species undergo exchange more rapidly than the lower charged species. This suggests that the more compact, lower charged species are more stable and less dynamic than their more expanded, higher charged counterparts.

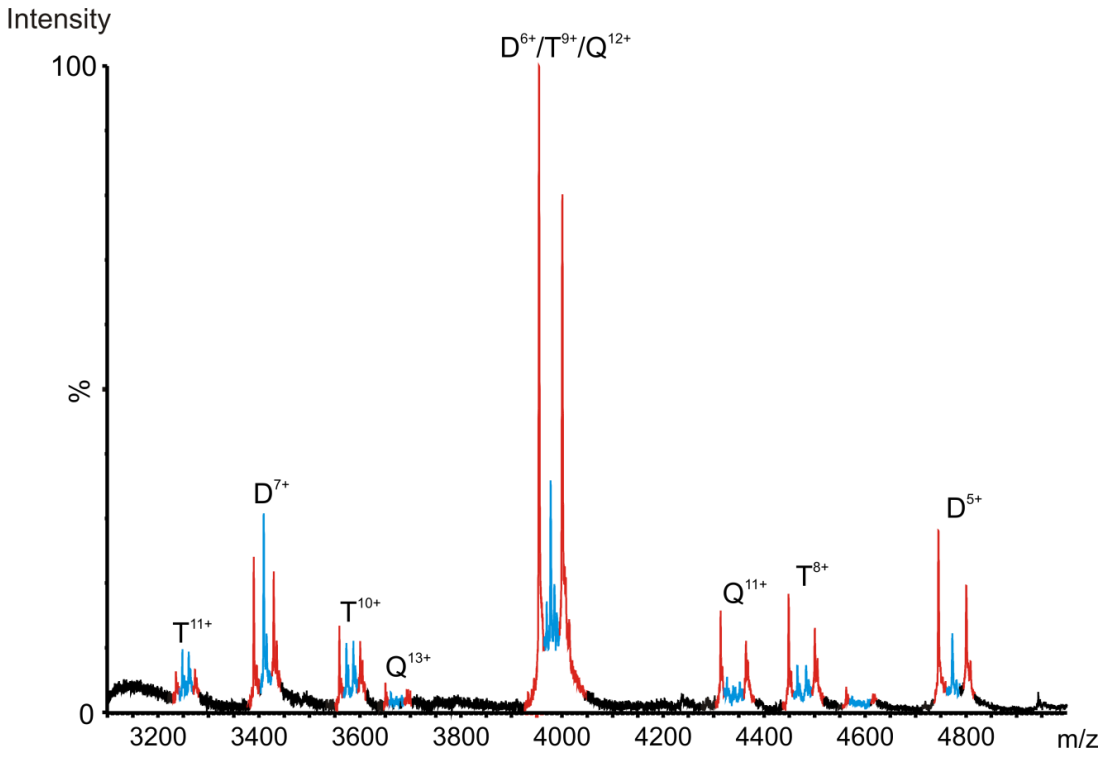


Figure 3.2.15 - ^{14}N and ^{15}N subunit exchange of the dimer, trimer and tetramer ions of wild-type β_2m. Samples of ^{14}N and ^{15}N wild-type β_2m were mixed in a 1:1 ratio for 1 min in 100 mM ammonium formate pH 2.5 before mass spectral acquisition. Peaks corresponding to all ^{14}N and all ^{15}N oligomers are shown in red, and peaks corresponded to mixed ^{14}N/^{15}N oligomers in blue.

Although the subunit exchange data reveal differences in the stabilities of the different oligomeric intermediates formed instantaneously at pH 2.5, it is interesting to determine whether the dynamics of these oligomers changes over time during fibril formation and whether any differences can be observed in the wild-type oligomers compared with those of other β_2m variants. Thus, the subunit dynamics of β_2m fibril formation were investigated at various time points throughout the lag phase of fibril formation. In these experiments, ^{14}N wild-type β_2m was incubated at 37 ºC, shaking at 600 rpm, in 100 mM ammonium formate pH 2.5 and at various time points during the lag phase ^{15}N wild-type β_2m (kept on ice throughout) was added and the subunit exchange monitored by ESI-MS over a 50 minute time-course at room temperature (Materials and Methods Section 3.4.10.4). In all cases, the lag phase was additionally monitored using Thioflavin T kinetics and fibril formation was at the end of the reaction confirmed by negative stain EM (data not shown). Note, due to the lack of agitation and incubation of the ^{15}N labelled sample, fibrils do not form. The mass spectrum of ^{15}N wild-type β_2m protein therefore remained

constant throughout the subunit exchange experiments (data not shown). Peaks corresponding to the ^{15}N wild-type β_2m protein can be used effectively as an internal standard to compare mass spectra between different time points assuming no subunit exchange occurs between the β_2m oligomers.

The subunit exchange kinetics for the dimer (6+, 7+ and 8+), trimer (10+, 11+) and tetramer (13+) of wild-type β_2m are shown in Figure 3.2.15. After 1 min of incubation with ^{15}N wild-type β_2m, subunit exchange had occurred between all the ^{14}N and ^{15}N wild-type β_2m oligomers present (Figures 3.2.15 and 3.2.15A top spectrum). After 50 min of mixing with ^{15}N wild-type β_2m, more mixed oligomers were observed suggesting that the dimer, trimer and tetramer all undergo rapid exchange with one another until they reach an equilibrium in which equal amounts of ^{14}N and ^{15}N monomers are present in the oligomeric state (Figure 3.2.16A bottom spectrum). The amount of subunit exchange for ^{14}N wild-type after 50 min of mixing with ^{15}N wild-type is comparable whether the ^{14}N wild-type was taken at 0 % or 50 % of the lag time of fibril formation (Figure 3.2.16A and B top spectra). This suggests the oligomers observed in the mass spectrum under these experimental conditions are the same throughout the lag time of fibrillogenesis.

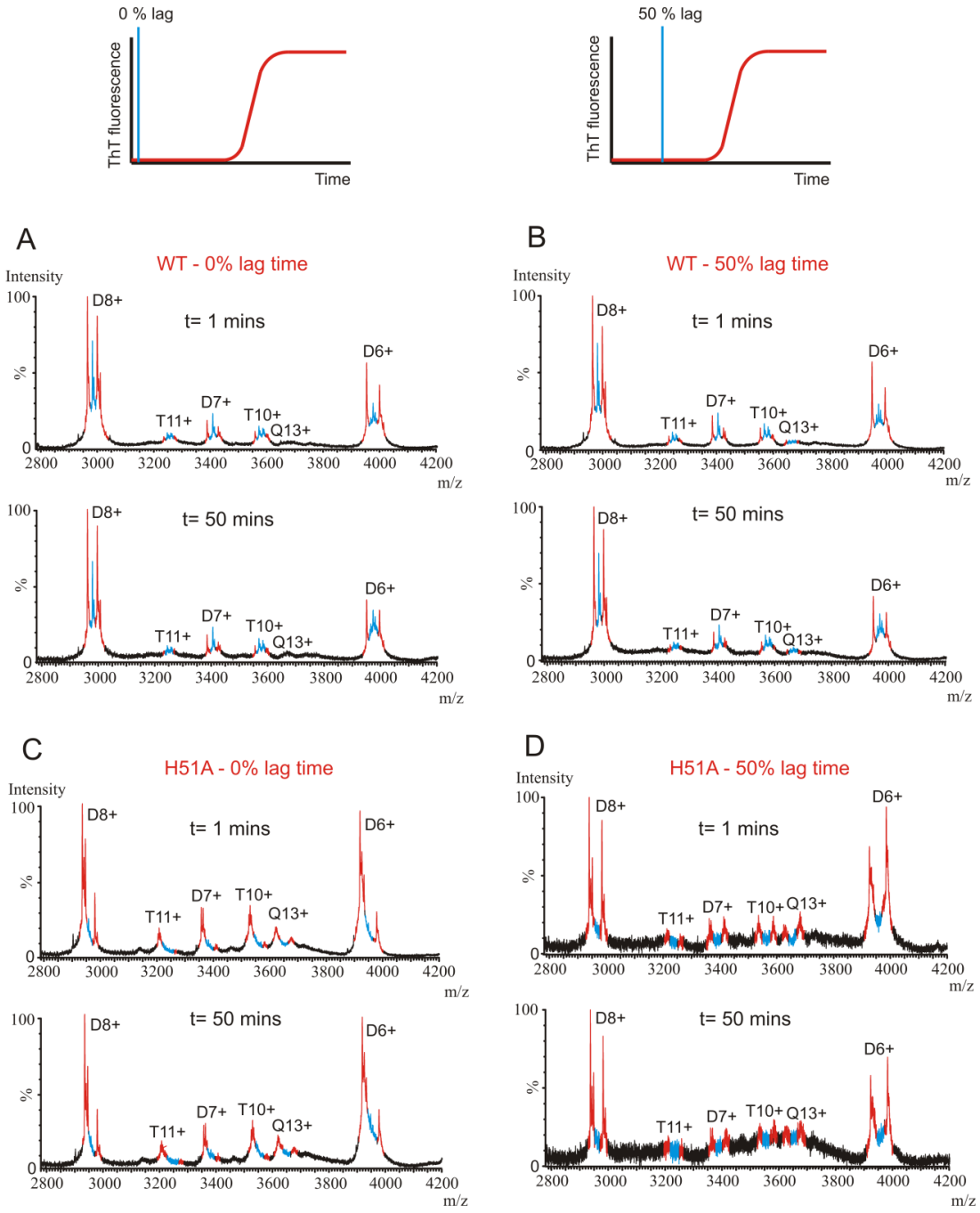


Figure 3.2.16 - ^{14}N and ^{15}N subunit exchange of the dimer 6+, 7+, 8+, trimer 10+, 11+ and tetramer 13+ ions. Samples of ^{14}N wild-type (A and B) or H51A (C and D) were mixed with ^{15}N wild-type and ^{15}N H51A, respectively in a 1:1 ratio for 1 min (top spectra in each case) and 50 min (bottom spectra in each case) at 0 % (A and C) and 50 % (B and D) of the lag phase. Peaks corresponding to all ^{14}N and all ^{15}N oligomers are shown in red, and peaks corresponded to mixed $^{14}N/^{15}N$ oligomers in blue.

In contrast to wild-type β_2m, on mixing ^{14}N H51A with ^{15}N H51A very little subunit exchange occurred after 1 min (Figure 3.2.16C top spectrum). After 50 min of incubation, no exchange had occurred with no peaks observed corresponding to the

mixed ^{14}N H51A:^{15}N H51A dimers, trimers and tetramers (Figure 3.2.16C bottom spectrum). The ratio of the unexchanged ^{14}N H51A to ^{15}N H51A oligomer peaks are not of equal abundance in the *m/z* spectrum at the start of the reaction even though ^{14}N H51A and ^{15}N H51A were mixed in a 1:1 ratio (Figure 3.2.16C). This is most likely due to the experimental error associated with measuring the ^{14}N H51A and ^{15}N H51A protein concentrations initially in combination with the manual mixing of the ^{14}N H51A and ^{15}N H51A samples in a 1:1 (*v/v*) ratio. For this reason, the expected 1:2:1, 1:3:3:1 and 1:4:6:4:1 ratios for the dimer, trimer and tetramer, respectively will not be observed. Peaks would still be expected to be visible corresponding to mixed oligomers if subunit exchange was occurring in solution.

A difference between the H51A and wild-type oligomers was also evident on comparison of the subunit dynamics at 50 % of lag time (Figure 3.2.16B and D). For the H51A oligomers formed during the lag phase, very little subunit exchange occurs, suggesting these oligomers are more stable entities compared with the more rapidly exchanging wild-type oligomers. Very little exchange occurs even after 50 min of incubation with ^{15}N H51A (Figure 3.2.16D bottom spectrum). This difference in subunit exchange for H51A compared with wild-type β_2m may explain the difference in lag time observed (Figure 3.2.6). If the H51A oligomers formed are more stable and less dynamic that those of the wild-type protein then they could be slower at converting into the amyloidogenic state necessary for fibril formation. Alternatively, the H51A oligomers observed here may be off-pathway oligomers thus slowing the rate of fibril formation by diverting H51A monomers to dead-end products.

Together the H51A variant data suggests a role for the D-strand in amyloid fibril formation at pH 2.5. This supports previous data whereby the mutation of His51 to proline also decreased the rate of fibril formation (Chiba *et al.* 2003). It is possible, that the lack of charge at this residue is causing the differences in fibril formation kinetics observed; therefore, the role of charge residues throughout the β_2m sequence in fibril formation was investigated further.

3.2.7 Effect of Lysine Mutations on β_2m Fibril Formation

Since lack of a charged residue in H51A resulted in a decrease in lag time and also decreased the dynamics of the oligomers present, the effect of charged residues, in general, on β_2m fibril formation at pH 2.5 was investigated further. Six lysine residues were selected throughout the β_2m sequence and substituted one by one to alanine (Figure 3.2.17). These lysine residues selected throughout the β_2m sequence are all present in the loop regions of β_2m and thus may be acting to prevent β_2m aggregation (Richardson *et al.* 2002).

The work on the lysine variants was carried out with Dr. Charlotte Scarff (University of Leeds). The choice of mutation and primer design for these lysine variants was carried out by myself (Materials and Methods Section 3.4.2.4). All the expression, purification, ThT kinetics, EM, and mass spectrometry on the lysine variants K58A, K75A, K91A were carried out by myself. The expression, purification, ThT kinetics, EM and mass spectrometry of the K6A, K19A and K41A variants were carried out by Dr. Charlotte Scarff (University of Leeds).

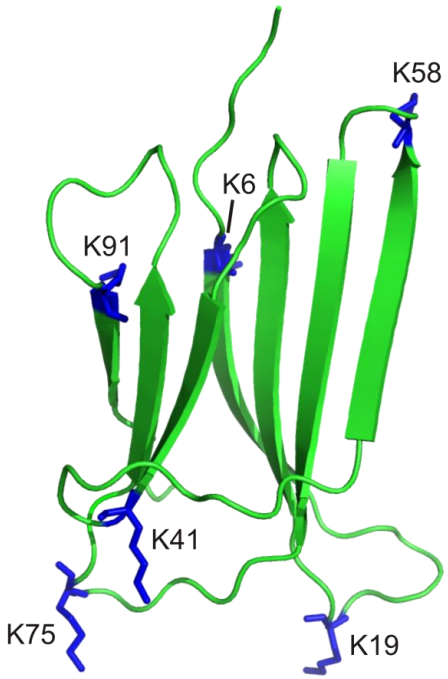


Figure 3.2.17 - Native β_2m structure (PDB file:2YXF) (Iwata *et al.* 2007) showing the position of the lysine variants used in this thesis. The mutated lysine residues are modelled as balls and sticks (blue).

All the lysine variants were expressed and purified as described in Section 3.2.1 and Materials and Methods 3.4.3. To determine whether these lysine variants are able to form amyloid fibrils, the variants were incubated individually in 100 mM ammonium formate pH 2.5 at 37 ºC, shaking at 600 rpm, and after approximately 48 h each sample was imaged by negative stain EM. The EM images show all variants formed long-straight fibrils (Figure 3.2.18).

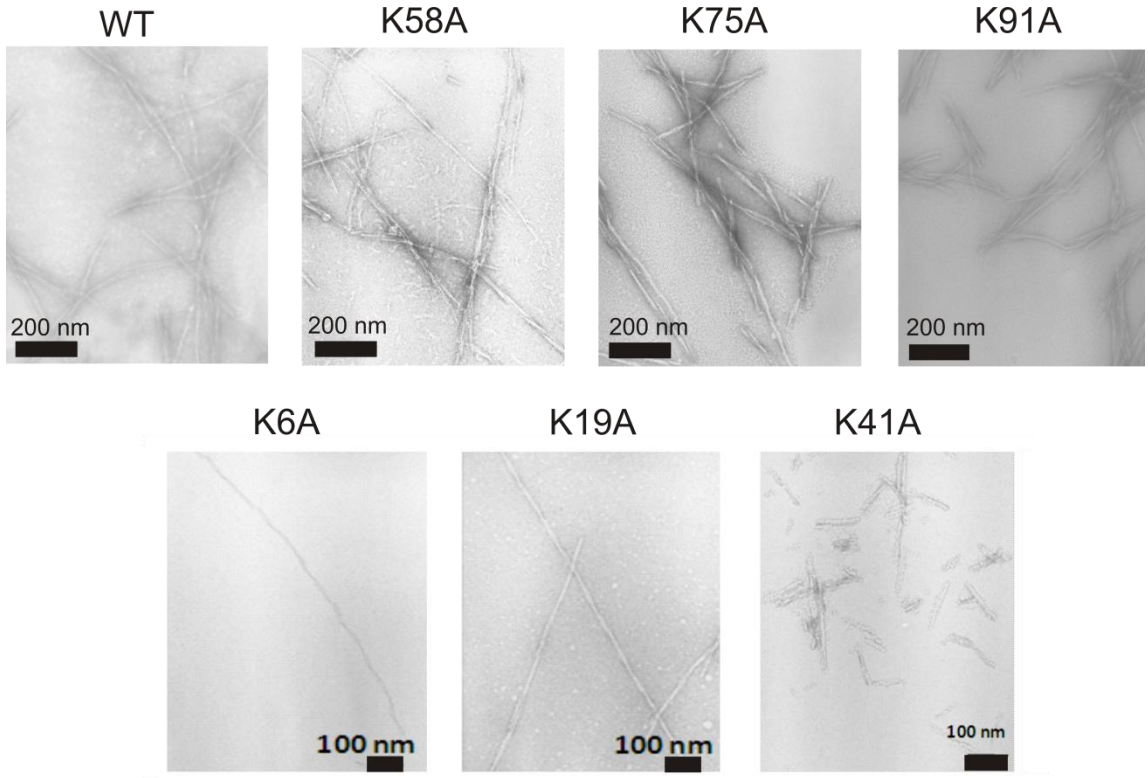


Figure 3.2.18 - Electron microscopy of fibrils formed in 100 mM ammonium formate pH 2.5, 37 ºC, 600 rpm, by wild-type β_2m and the β_2m variants, K6A, K19A, K41A, K58A, K75A and K91A. The scale bar for the K58A, K75A and K91A variants represents 200 nm. EM of the variants K6A, K19A and K41A was performed by Dr. Charlotte Scarff (University of Leeds), the scale bar for these variants represents 100 nm.

The kinetics of fibril formation for the lysine variants were also investigated using ThT fluorescence (Figure 3.2.19). If charged residues are important in amyloid fibril formation at pH 2.5, then differences in the fibril formation kinetics would be expected for these variants. An example of the ThT kinetic data for the lysine variants K58A, K75A and K91A is shown in Figure 3.2.19A. Strikingly, all the lysine variants analysed showed dramatic differences in their lag times suggesting

that charge is indeed important in amyloid fibril formation of β_2m in 100 mM ammonium formate pH 2.5.

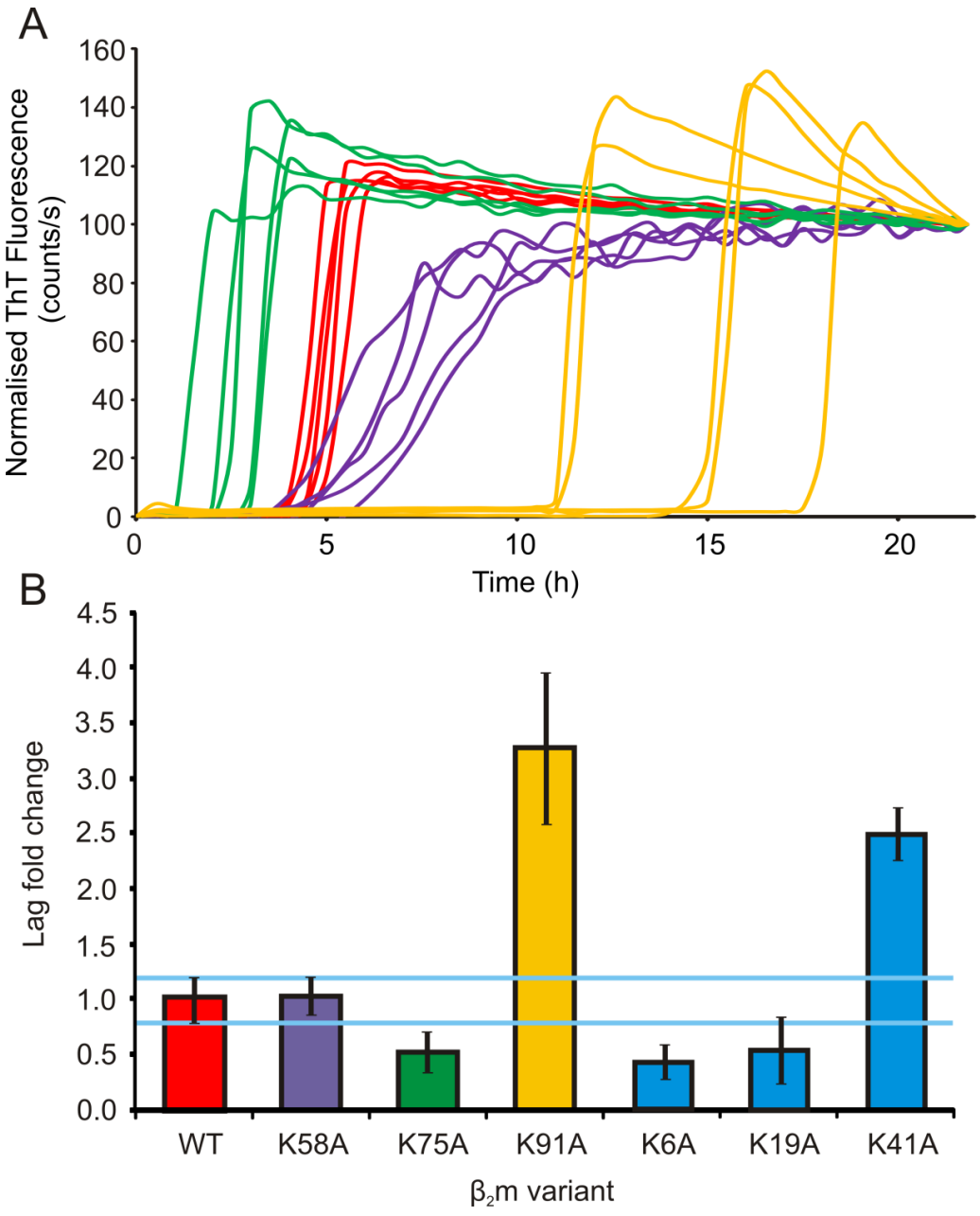


Figure 3.2.19 - A) Fibril formation kinetics of lysine variants, K75A (green), K58A (purple), K91A (yellow) along with wild-type β_2m (red). Five replicates are shown for each variant. Experiments were carried out using a plate reader, in 100 mM ammonium formate pH 2.5, 37 ºC, shaking at 600 rpm. All data were normalised to the end point of the reactions. B) Comparison of the kinetics of fibril formation of the lysine variants K58A (purple), K75A (green), K91A (yellow), K6A (blue), K19A (blue), K41A (blue) with respect to wild-type (red). The fold change in lag time compared to wild-type β_2m is shown for comparison. The horizontal blue lines represent ± 1 standard deviation from 5 repeats of the wild-type β_2m proteins lag time. The experimental data performed by Dr. Charlotte Scarff (University of Leeds) are shown in blue. The error bars represent ± 1 standard deviation.

On further analysis of the EM image of K58A, small oligomers are present in addition to long-straight fibrils which are absent in all other EM images (Figure 3.2.3 and 3.2.18). These could either be off-pathway oligomers or on-pathway oligomers that are simply slower at converting into mature amyloid fibrils consistent with the increased elongation rate observed. In addition, on further analysis of the EM images for K41A and K91A, the fibrils observed were shorter in length compared to all other fibrils formed from either the β_2m variants or wild-type β_2m (Figure 3.2.18). The increase in lag time observed in this case for the K41A and K91A variants may be due to the decrease in stability of the final amyloid fibrils causing the equilibrium between oligomers and amyloid fibrils to shift back towards the monomer/oligomer population. Additional experiments in which the stability of fibrils are measured, for example by guanidine denaturation experiments, are needed to confirm this hypothesis.

All the lysine variants displayed here showed comparable oligomer populations and collision cross-sections to wild-type β_2m as analysed by ESI-MS and ESI-IMS-MS in 100 mM ammonium formate pH 2.5 (data not shown). This suggests that these charged lysine residues are involved post-hexamer formation since no differences are observed up to this oligomer size in the ESI mass spectra, and also the initial oligomeric contacts have been suggested to be hydrophobic by ESI-CID-MS (Figure 3.2.13).

Overall, β_2m amyloid formation is a complex mechanism in which oligomers join together to form higher-order fibril structures. In this thesis, important residues involved in β_2m amyloid fibril formation have been highlighted. Although, mass spectrometry alone cannot be used to reveal a unique model to fit all the experimental data shown here, with additional experimental work the mechanism of amyloid fibril formation could be revealed.

3.3. Conclusions

In this thesis, residues have been highlighted that upon mutation affect the lag time of fibril formation as monitored by ThT fluorescence (Figures 3.2.4 and 3.2.19). One mutation in particular on the D-strand of β_2m, H51A, caused a significant two-fold increase in the lag time of fibril formation (Figures 3.2.4 and 3.2.6). Further analysis of the oligomeric intermediates of this variant by ESI-MS, revealed differences in the structural dynamics of the oligomers present throughout the lag phase (Figure 3.2.16). This suggests that either there is a balance between on- and off-pathway intermediates slowing the rate of fibril formation, or that the more stable, less dynamic H51A oligomers are slower at converting to a conformation that is required for fibril formation. These structural differences were not evident from the ESI-IMS-MS and ESI-CID-MS data alone with the β_2m wild-type and H51A oligomeric species appearing structurally and thermodynamically identical using these methods of analysis (Figures 3.2.9, 3.2.11 and 3.2.13).

Mutation of lysine residues throughout the β_2m sequence had dramatic effects on the lag time of fibril formation (Figure 3.2.19). The K41A and K91A variants both increased the lag phase by up to 3.5-fold compared with the wild-type β_2m (Figure 3.2.19). This increase in lag time has been shown previously for another variant of β_2m termed the triple variant whereby three residues on the E-strand were substituted to alanine (F62A, Y63A, Y67A) suggesting this aromatic region is also important in β_2m fibril formation (Plattt *et al.* 2005; Platt *et al.* 2008; Routledge *et al.* 2009). Looking together at the residues important in fibrillogenesis, no trend is obvious between the location of the important residues and the lag time of fibrillogenesis (with residues Phe62, Tyr63, Tyr67 in strand E, His51 in strand D, Lys41 in strand C and Lys91 in the F-G loop showing the most dramatic effects). The picture is more complicated using these types of experiments, since the mutated residue itself may not be causing the effects observed; local or long range effects of the mutation could be changing the overall protein conformation altering its ability to form oligomeric interactions important in fibrillogenesis. Additional experiments are needed to

determine the precise role of these important residues in amyloid fibril formation and to unravel further details on the oligomeric intermediates. Further insight into the mechanism of amyloid fibril formation will aid the design of novel inhibitor to prevent this complex process.

An alternative technique for studying protein conformation and dynamics is hydrogen deuterium exchange (HDX) (Percy *et al.* 2012). HDX on native proteins in solution can report on the whole structure and large scale as well as local conformational changes. Indeed, HDX experiments on β_2m have revealed an increase in the conformational dynamics of β_2m on its release from MHC 1 (Hodkinson *et al.* 2009; Hodkinson *et al.* 2012). Protein digestion, with (for example) the protease pepsin, can be used prior to MS analysis to provide additional site-specific information on the location of exposed peptide bonds in the structures being probed. For example, the regions involved in the partially unfolding of the SH3 domain (a small, modular domain found in many proteins) have been identified by proteolytic digestion combined with HDX MS (Wales *et al.* 2006). These experiments are not ideal since problems with back exchange can occur, complicating experimental analysis. MS/MS can be used instead of protein digestion with proteases to reveal site-specific information on proteins and their complexes. CID fragmentation methods have been used to report on HDX within protein structure; however, problems can arise whereby deuterium atoms exchange sites during the fragmentation process, termed scrambling (Jorgensen *et al.* 2005). This problem has recently been overcome by the use of ETD/ECD instead of CID fragmentation (Pan *et al.* 2008; Rand *et al.* 2009). Indeed a very recent publication by Rand *et al.* (2012) has shown great promise for studying protein conformation by MS, whereby deuterated ammonia was introduced into the ionisation source with subsequent fragmentation in the trap and transfer region of the Synapt mass spectrometer (Introduction Section 1.4, Figure 1.6) revealing residue-specific information on the gas phase conformations of ubiquitin and cytochrome c (Rand *et al.* 2012).

Oxidative labelling is another recently introduced technique used to label protein complexes for detection by ESI-MS (Konermann *et al.* 2010). Here, free hydroxy radicals are generated using a high energy laser. These $OH^{\cdot}$ radicals subsequently

react with exposed amino acids resulting in oxidation and an increase by 18 Da is observed on MS analysis. Oxidative labelling can be combined with proteolysis or CID/ECD/ETD fragmentation to determine site-specific labelling sites. Experiments by the Konermann group have shown the application of this technique to monitor the location of individual amino acids upon folding of membrane proteins from their denatured state (Pan *et al.* 2011; Pan *et al.* 2012). These labelling techniques are advantageous over HDX MS experiments since the labelling is irreversible and thus problems with back-exchange and scrambling are avoided. This technique offers great promise for determining oligomeric interfaces during ESI-IMS-MS analysis, since residues involved in the binding interface between the oligomers will be protected from labelling and thus could be identified.

In a similar way, covalent labelling has shown great promise for the analysis of oligomeric intermediates during β_2m fibril formation at neutral pH in the presence of Cu^{2+} (Mendoza *et al.* 2010; Mendoza *et al.* 2011). In these experiments, sulfo*N*-hydroxysuccinimide acetate (NHSA), diethylpyrocarbonate (DEPC) and 2, 3-butanedione (BD) were used to label lysines, histidine, amine and OH groups, and arginine side chains, respectively, and the protection/solvent exposure of these residues to labelling throughout the lag phase of fibril formation was used to determine which residues are buried/exposed in which oligomeric state of interest. Although from these data alone the precise structure of the amyloidogenic intermediates could not be predicted, the results obtained together with previous X-ray crystallography data allowed a more accurate prediction of the dimer and tetramer oligomeric sturctures, thus allowing a model of the mechanism of formation of these oligomeric species to be established (Mendoza *et al.* 2010; Mendoza *et al.* 2011).

Together, this thesis chapter has shown how ESI-IMS-MS can be used to analyse oligomeric intermediates in amyloid fibril formation. Although the data presented do not predict a mechanism of amyloid formation conclusively, with the recent advances in the mass spectrometry field (as discussed here), more detailed structural information on the β_2m oligomeric intermediates will be possible in the near future.

Amyloid

3.4. Materials and Methods

3.4.1 Reagents

Molecular Biology Reagents

The β_2m pINK$_{wil\text{-}type}$ plasmid was constructed by Dr Neil Kad (formerly of University of Leeds) by inserting the wild-type human β_2m gene sequence from plasmid pBJ192 (Reid *et al.* 1996) into a pET23a expression vector (Novagen). The plasmids encoding the β_2m variants K19A, F22A, F30A, H51A, L54A, Y63A, L64A, L65A, V82A, H84A and R97A were supplied by Thomas Jahn, Katy Routledge and Timo Eichner (formerly of University of Leeds). *Escherichia coli* (*E.coli*) BL21 (DE3) pLysS cells were purchased from Promega Corporation (Southampton, UK).

***E.coli* bacterial strains**

Strain	Characteristics
BL21(DE3)pLysS	*Hsd*S *gal* (λIts 857 *ind*1 Sam7 *nin*5 *lac*UV5-T7*gene*1)
XL1-Blue	*rec*A1 *end*A1 *gyr*A96 *thi*-1 *hsd*R17 *sup*E44 *rel*A1 *lac* (F' *pro*AB *lac*I^{q}ZΔM15 Tn*10*(Tet'))

Table 3.4.1 - *E.coli* bacterial strains used for extracting DNA for storage (XL1-Blue) and protein expression (BL21 (DE3) pLysS).

Reagents for Bacterial Growth

NZ Amine was purchased from Duchefia, Netherlands. Details of where all other molecular biology, protein expression and protein purification reagents were purchased are described in Chapter 2 (Pilus Materials and Methods 2.4).

Reagents used for fibril formation assays

Black 96 well plates were purchased from Corning Ltd, USA and clear sealing films from Thistle Scientifics, UK. Thioflavin T (ThT) was purchased from Sigma-Aldrich Corporation (Poole, Dorset, UK).

Electron Microscopy Materials

Carbon/formvar-coated 300 mesh electron microscopy grids were purchased from the Electron Microscopy suite in the Faculty of Biological Sciences, University of Leeds. 2 % (*w/v*) solution of uranyl acetate was obtained from the Electron Microscopy suite in the Faculty of Biological Sciences, University of Leeds.

Mass Spectrometry Reagents

Ammonium formate was purchased from Sigma-Aldrich Corporation (Poole, Dorset, UK), and HCl from Fisher Scientific (Loughborough, UK).

3.4.2 Molecular Biology

3.4.2.1 Plasmid DNA preparation

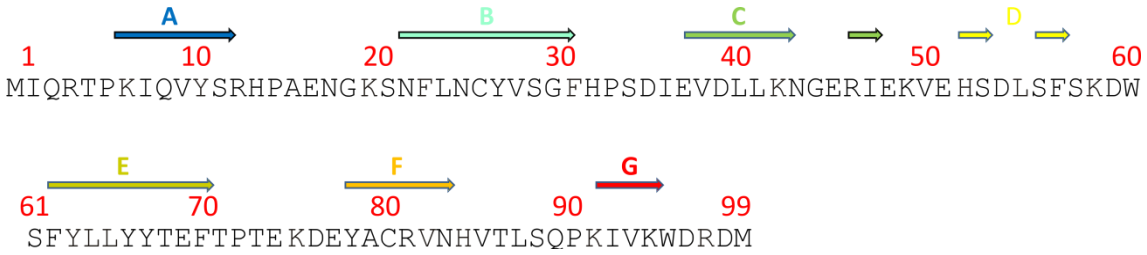
Plasmid DNA was isolated from 10 mL of fresh overnight liquid cultures of XL1-Blue cells using a Qiagen plasmid Mini-prep kit according to the manufacturer's instructions. To determine the concentration of DNA plasmid, an A_{260} of 1.0 was assumed equivalent to 50 µg/mL of double stranded DNA (Sambrook 1989).

3.4.2.2 Transformation of plasmid DNA

Plasmid DNA was transformed into XL1-Blue cells or BL21 (DE3) pLysS cells for DNA extraction or protein expression, respectively, according to the protocol mentioned previously (see Pilus Materials and Methods, Section 2.3). When the plasmid DNA was initially transformed into *E.coli* XL1-Blue cells after mutagenesis, NZY^+ medium was used to grow the cells after heat shocking instead of LB medium. To make NZY^+ medium, 10 g NZ Amine, 5 g yeast extract and 5 g NaCl were dissolved in a final volume of 1 L with >18 MΩ water. The medium was then autoclaved at 120 °C at 15 p.s.i for 20 min and before use, the following filter-sterilised supplements (initial stock concentrations) were added (*per* litre): 12.5 ml 1M $MgSO_4$, 12.5 ml 1M $MgCl_2$ and 20 ml 20 % (*w/v*) glucose.

3.4.2.3 β_2m amino acid sequence

The amino acid sequence of β_2m is shown below along with the location of the β-strands throughout the sequence.



3.4.2.4 Stratagene QuikChange® Mutagenesis

Site-directed mutagenesis (QuikChange) was carried out to create the alanine variants K6A, K41A, K58A, K75A and K91A. QuikChange mutagenesis was carried out according to the manufacturer's instructions, except that following mutagenesis, the DNA template was digested with 1 µL *DpnI* enzyme (specific for methylated DNA) for 16 h at 37 °C, rather than the recommended 1 h, to improve mutation frequency. All variants were sequenced from the T7 promoter to ensure that the gene contained the desired mutation and no other. The primers used to create the β_2m variants are shown in Table 3.4.2.

Variant	Primer Sequence (5′ to 3′)
K6A	tatacatatgattcaaagaactccagcaattcaggtttactcacgtcatc
K41A	ccgacattgaagttgacttactggcgaatggagagagaattgaaaaag
K91A	tgtgactttgtcacagcccgcgatagttaagtgggatcga
K75A	ctgaattcacccccactgaagcagatgagtatgcctgcc
K58A	cattcagacttgtctttcagcgcggactggtctttctatctctt

Table 3.4.2. Forward primers used for QuikChange® site-directed mutagenesis.

3.4.3 Expression and Purification of β_2m

3.4.3.1 Growth Media

Luria-Bertani (LB) medium was purchased from Merck (Germany) and sterilised by autoclaving at 120 °C at 15 p.s.i. for 20 mins. ^{15}N labelled minimal media were made

by dissolving 10 g K_2HPO_4, 10 g KH_2PO_4, 7.5 g Na_2HPO_4, 9 g K_2SO_4, 1 g $^{15}NH_4Cl$ (or 1 g $^{14}NH_4Cl$ for unlabelled minimal media) in 1 L deionised (18 MΩ) water and autoclaving at 120 °C at 15 p.s.i. for 20 mins. Once cool, filter sterilised 1M MgCl (2 mL), 1 M $CaCl_2$ (100 µL) and 20 % (*w/v*) glucose (20 mL) were added. Filter-sterilised antibiotics were added to the LB/minimal media when appropriate.

3.4.3.2 Pre-culture

Pre-cultures were made by addition of 20 µL of a bacterial glycerol stock (containing Bl21 (DE3) pLysS cells transformed with the relevant plasmid) and ampicillin (100 µg/mL) to 500 mL sterile LB media (Pilus Materials and Methods, Section 2.4). The media was incubated overnight at 37 ºC overnight, shaking at 200 rpm and used immediately for large-scale protein over-expression. For ^{15}N labelled protein expression, a pre-culture was made by addition of 20 µL bacterial glycerol stock to 500 mL of sterile unlabelled minimal medium and the mixture incubated for 24 hours at 37 ºC overnight, shaking at 200 rpm. This culture was then centrifuged at 6,000 rpm, 4 ºC and then resuspended in 5 mL ^{15}N labelled minimal media, ampicillin added to a final concentration of 100 µg/mL, and the culture used immediately for large-scale protein over-expression.

3.4.3.3 Cytoplasmic protein expression

Twelve 2 L flask each containing 1 L of LB media and 100 µg/mL ampicillin were pre-warmed overnight at 37 ºC. 30 mL of each pre-culture containing *E.coli* BL21 (DE3) pLysS cells with the relevant plasmid transformed into it were added to each flask along with 1 mL of 100 mg/mL carbenicillin (stored at -20 ºC). For ^{15}N labelled protein, 500 µL of pre-culture was added to each flask (typically five flasks containing 1 L of minimal media) containing 1 mL of 100 mg/mL carbenicillin. All cultures were incubated at 37 ºC, shaking at 200 rpm in an orbital shaking incubator, and the optical density (O.D) at 600 nm was measured at regular intervals. At an OD of 0.6, protein expression was induced by the addition of 1 mL of 1 M filter sterilised IPTG. After 16 hours, the cultures were harvested at 6,000 rpm (JLA 8.1 rotor, Beckman Coulter, High Wycombe, UK), 4 ºC.

3.4.3.4 Inclusion Body Isolation

The cell pellet was resuspended in 100 mL lysis buffer containing 20 mM Tris-HCl pH 8.0, lysozyme (100 µg/mL), phenyl methyl sulponyl fluoride (PMSF) (50 µg/mL) DNase (20 µg/mL), ethylenediaminetetraacetic acid disodium salt (EDTA) (1 mM) and incubated at room temperature for 30 mins. Cells were further lysed using a TS Series cell disruptor (Constant Systems Ltd, UK) at 20 kpsi. The cell lysate was then centrifuged at 12,000 rpm (JA 25.50 rotor, Beckman Coulter, High Wycombe, UK) for 1 h at 4 °C. The supernatant was discarded and the cell pellet was resuspended in 100 mL 20 mM Tris-HCl pH 8.0. The cell pellet was again centrifuged (12,000 rpm, for 1 h at 4 °C) and the pellet resuspended a further two times in 20 mM Tris-HCl pH 8.0, discarding the supernatant after each centrifugation step. The final cell pellet was solubilised in 150 mL 20 mM Tris-HCl pH 8.0 containing 8 M urea and incubated overnight on an orbital shaker at 70 rpm. Finally any insoluble material was discarded by centrifugation at 12,000 rpm (JA 25.50 rotor, Beckman Coulter, High Wycombe, UK) for 1 h at 4 °C.

3.4.3.5 Refolding By Dialysis

Protein in urea was dialysed against 5 L of 20 mM Tris-HCl pH 8.0 at 4 °C using a 3,500 Da molecular weight cut-off dialysis membrane. Any insoluble material was pelleted by centrifugation at 12,000 rpm (JA 25.50 rotor, Beckman Coulter, High Wycombe, UK) for 1h at 4 °C and discarded before protein purification.

3.4.3.6 Protein Purification

Anion Exchange Chromatography

The refolded β_2m protein solution was initially purified by anion exchange chromatography using a XK26/100 column containing 100 mL HiTrap Q-Sepharose fast flow resin (GE Healthcare, Buckinghamshire, UK). All solutions were filtered using a 0.22 µm Steritop filter (Millipore, Massachusetts, USA) and degassed using a Buchi Vac V-500 vacuum pump (Buchhi Labortechnik AG, Switzerland) prior to passing through the column. The column was equilibrated with 10 column volumes

of 20 mM Tris-HCl pH 8.0 and the impure protein solution loaded onto the column at a flow rate of 1.5 mL/min using an ÄKTA prime purification system (GE Healthcare, Buckinghamshire, UK). Unbound protein was washed through the column with 5 column volumes of 20 mM Tris-HCl pH 8.0 at a flow rate of 3 ml/min. Finally the protein was eluted with a linear gradient of 0-500 mM NaCl in 20 mM Tris-HCl pH 8.0, over 700 mL, flow rate 3 mL/min. All steps were performed at room temperature.

The pooled β_2m fractions, as detected by spectrophotometry at A_{280} and confirmed by SDS-PAGE (Pilus Materials and Methods Section 2.4.11), were dialysed into H_2O (18 MΩ) at 4 °C, freeze-dried and stored at -20 ºC until further use.

Size Exclusion Chromatography

The freeze dried β_2m was dissolved into 20 mM Tris-HCl pH 8.0 at a concentration of ~5mg/mL and filtered through a 0.22 μm Steritop filter (Millipore, Massachusetts, USA). 5 mL was then injected onto a HiLoad 26/60 Superdex 75 column (GE Healthcare, UK) equilibrated in 20 mM Tris-HCl pH 8.0 and controlled by an ÄKTA prime purification system. The protein was eluted at a flow rate of 3 mL/min. Typically monomeric β_2m eluted after 200 mL with oligomeric forms eluting prior to this. Several (usually six) gel filtration runs were necessary to purify the whole batch of protein. Finally all fractions containing pure β_2m protein, as confirmed by SDS-PAGE, were combined and dialysed into H_2O (18 MΩ) before freeze drying and storage at -20 ºC. The average yield of pure wild-type β_2m for this procedure was 50 mg/L of bacterial culture. β_2m variant yields for this procedure vary considerably, from 5 mg/L to 50 mg/L of bacterial culture.

3.4.4 Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Gel electrophoresis was routinely used to monitor over-expression and purification of recombinant proteins and to determine fibril yield (for example a yield of 50 mg/L

was obtained for wild type β_2m). Tris-tricine buffered gels were used in all cases (Pilus Materials and Methods, Section 2.4.11).

3.4.5 Characterisation of β_2m

3.4.5.1 Purity and Mass Determination

Purified protein was determined to be monomeric after gel filtration and was shown to be pure by SDS-PAGE (Pilus Materials and Methods, Section 2.4). The exact molecular weight of purified protein was determined by electrospray ionisation mass spectrometry (ESI-MS) (Pilus Materials and Methods, Section 2.4.15.3) using CsI to calibrate the mass spectrometer (Pilus Materials and Methods, Section 2.4.15.1). The expected and experimentally determined molecular weights are shown in Table 3.4.3.

β_2m variant	**Expected Molecular Weight (Da)**	**Determined Molecular Weight (Da)**	**Extinction Coefficient ($M^{-1}cm^{-1}$)**
wild-type	11860	11860	19940
K6A	11803	11800	19940
K19A	11803	11800	19940
K41A	11803	11801	19940
F30A	11784	11782	19940
H51A	11794	11793	19940
K58A	11803	11803	19940
Y63A	11768	11771	18450
L64A	11818	11818	19940
K75A	11803	11804	19940
V82A	11832	11833	19940
H84A	11794	11795	19940
K91A	11803	11801	19940
R97A	11775	11775	19940

Table 3.4.3 – A comparison between calculated molecular masses based on amino acid sequences and the measured molecular masses from nanoESI-MS analysis. The extinction coefficients used to calculate the protein concentration for each variant are also shown.

3.4.5.2 Extinction Co-efficient Determination

The concentration of each β_2m variant was calculated using the Beer-Lambert law by measuring the absorbance at 280 nm and using the theoretical extinction coefficients determined from the amino acid sequence (using Protparam online software) (Edelhoch 1967; Gill *et al.* 1989; Pace *et al.* 1995). The extinction coefficients used for the β_2m variants are shown in Table 3.3.3.

3.4.6 Electron Microscopy

10 µl of fibrils in 100 mM ammonium formate pH 2.5 (i.e. reaction end products) were placed on freshly ionised formvar- and carbon-coated EM grids for 30 sec. The grids were then blotted with filter paper to remove excess solvent and the samples stained with 2 % (*w*/*v*) uranyl acetate for 30 sec. The grids were then blotted again and air-dried before analysis. All images were taken using a Philips CM10 electron microscope operating at 80 keV.

3.4.7 Analytical Gel Filtration

Freeze dried β_2m was dissolved into 20 mM Tris-HCl pH 8.0 at a concentration of 100 µM. 200 µL was then injected onto a Superdex™ 75 10/300 GL column (GE Healthcare, UK) equilibrated in 20 mM Tris-HCl pH 8.0 and controlled by an ÄKTA prime purification system. The protein was eluted at a flow rate of 0.5 mL/min and the elution volume of the monomeric wild-type β_2m used for comparison with all other β_2m variants.

3.4.8 Fibril Growth Kinetics

3.4.8.1 Fibril Kinetics using the Plate Reader

β_2m or one of its variants was dissolved in 100 mM ammonium formate pH 2.5, 10 µM ThT (on ice) to a final protein concentration of 0.4 mg/mL (1 mL total volume). The protein concentration was determined by A_{280} using the specific extinction co-efficient determined for each variant (Table 3.4.3) and 100 µL aliquots placed into a

black 96 well plate. The plate was sealed with a clear heat-activated sealing film (Thistle Scientifics) allowing measurements *in situ* without evaporation of the sample. The plate was incubated at 37 °C with orbital shaking at 600 rpm for 48 h in a Fluostar Optima plate reader (BMG Labtech GmbH, Offenburg, Germany). ThT fluorescence was measured by the top down mode using a 440 nm (±10 nm) filter for excitation and a 480 nm (±12 nm) filter for emission. The data were normalised in each case to the final ThT fluorescence reading. At the end-point of the reaction, samples were taken for negative stain electron microscopy (Section 3.4.6).

3.4.8.2 Simultaneous Mass Spectrometry and Fibril Kinetics

β_2m or one of its variants was dissolved in 100 mM ammonium formate pH 2.5 (on ice) to a final protein concentration of 0.4 mg/mL. The protein concentration was determined by A_{280} using the specific extinction co-efficient determined for each variant (Table 3.4.3) and fibrillogenesis initiated by incubation at 37 °C, shaking at 600 rpm (Thriller Thermoshaker Incubator, PEQLAB Ltd. UK) for up to 48 h. 50 μL samples were taken at regular intervals and diluted with 950 μL Thioflavin T (ThT) (20 μM) in 100 mM ammonium formate pH 2.5. ThT fluorescence was measured in a PTI Quantamaster C-61 spectrofluorimeter using an excitation wavelength of 444 nm and emission at 480 nm with slit-widths of 4.5 nm. Simultaneous to the ThT reading, a 10 μL sample of the β_2m variant in 100 mM ammonium formate pH 2.5 was taken for analysis by mass spectrometry (Section 3.4.12). At the end-point of the reaction, samples were taken for negative stain electron microscopy (Section 3.4.6).

3.4.9 Calculation for the Lag Time of Fibril Growth

The observed growth traces showed different final ThT fluorescent signals for different β_2m variants. For simplicity all data were normalised assuming a final ThT signal of 1.0 and the resulting curves used to determine the lag times of fibril formation. The lag time was obtained by fitting a tangent to the steepest part of the normalised ThT growth phase and the time at which the line intersected the baseline was taken as the lag time (Hortschansky *et al.* 2005).

Fold changes in lag time were calculated using Equation 3.4.1 where $lag_{variant}$ is the average lag time of the β_2m of interest and $lag_{wiild\text{-}type}$ is the lag time of wild-type β_2m.

$$\text{Fold Change} = lag_{variant}/lag_{wild\text{-}type} \qquad \text{Equation 3.4.1}$$

3.4.10 Mass Spectrometry

3.4.10.1 Buffer Preparation

Ammonium formate buffers were made at 100 mM in deionised (18 MΩ) water. The pH was adjusted to pH 2.5 by addition of HCl. The ammonium formate buffer was stored at 4 ºC for a maximum of one week and filtered through a 200 nm-pore filter (minisart high-flow single-use sterile filter, Sartorius, Surrey, UK) before use.

3.4.10.2 ESI-(IMS)-MS

Ion mobility spectrometry-mass spectrometry (IMS-MS) experiments were performed on a Synapt HDMS, hybrid quadrupole-IMS-oaTOF (Waters Ltd., Manchester, UK) mass spectrometer interfaced with a NanoMate Triversa source (Advion Biosystems Inc. NY, USA). Positive nanoESI was used with a capillary voltage of 1.75 kV and a nitrogen nebulising gas pressure of 0.7 p.s.i. A cone voltage of 170 V and a backing pressure of 3.6 mbar were used to observe the β_2m oligomers. The ion accelerating voltages into the trap and transfer T-wave were set at 4 V and 40 V, respectively. The wave height was typically ramped from 2-25 V at 300 ms^{-1}.

Drift times were corrected for both mass dependent and mass independent times as reported previously (Knapman *et al.* 2010) and the drift time cross-section function was calibrated as described (Smith *et al.* 2009). Briefly, equine cytochrome C, sperm whale myoglobin and bovine ubiquitin were dissolved at a concentration of 10 µM in a mixture of acetonitrile/H_2O/acetic acid (50:40:10 *v/v/v*). The charge state ions for each protein were identified by their unique *m/z* and their measured drift times plotted against Ω' (see Introduction Section 1.4, Equation 1.11) calculated from David Clemmer's library of protein cross-sections determined by conventional IMS

(Liu *et al.* 1997; Shelimov *et al.* 1997; Valentine *et al.* 1997) (Figure 3.4.1, Table 3.4.4). The data were then fitted to the $y = Ax^B$ relationship to obtain the calibration curve (Figure 3.4.1). The values for A and B from the calibration curve were used to determine the collision cross-section of the protein charge state ions of interest from their drift time measurements. In all cases, the same IMS conditions were used for the protein of interest and the IMS calibrants.

Charge State	Ω (Å^2)			Ω' (x10^{21}Å^2 C^{-1} Da^{-1})		
	Ubiquitin	Cytochrome C	Myoglobin	Ubiquitin	Cytochrome C	Myoglobin
4	1004			8.277		
	1059			8.731		
5	1027			6.773		
	1137			7.499		
	1239			8.172		
6	1041			5.721		
	1220			6.705		
	1525			8.382		
7	1317			6.204		
	1580			7.443		
8	1442			5.944		
	1622			6.686		
9	1649	2215		6.042	8.120	
10	1732	2226	2796	5.712	7.344	9.228
11	1802	2303	2942	5.402	6.908	8.827
12		2335	3044		6.420	8.372
13		2391	3136		6.068	7.961
14		2473	3143		5.828	7.409
15		2579	3230		5.673	7.107
16		2679	3313		5.524	6.834
17		2723	3384		5.285	6.570
18		2766	3489		5.070	6.397
19		2800	3570		4.862	6.201
20			3682			6.076
21			3792			5.959
22			3815			5.723

Table 3.4.4 - Collision cross-sections (Ω) and charge and mass independent collision cross-sections (Ω') for the proteins bovine ubiquitin, horse cytochrome c and sperm whale apomyoglobin. Cross-sections are for protonated ions. Ω values were taken from David Clemmer's library (Liu *et al.* 1997; Shelimov *et al.* 1997; Valentine *et al.* 1997).

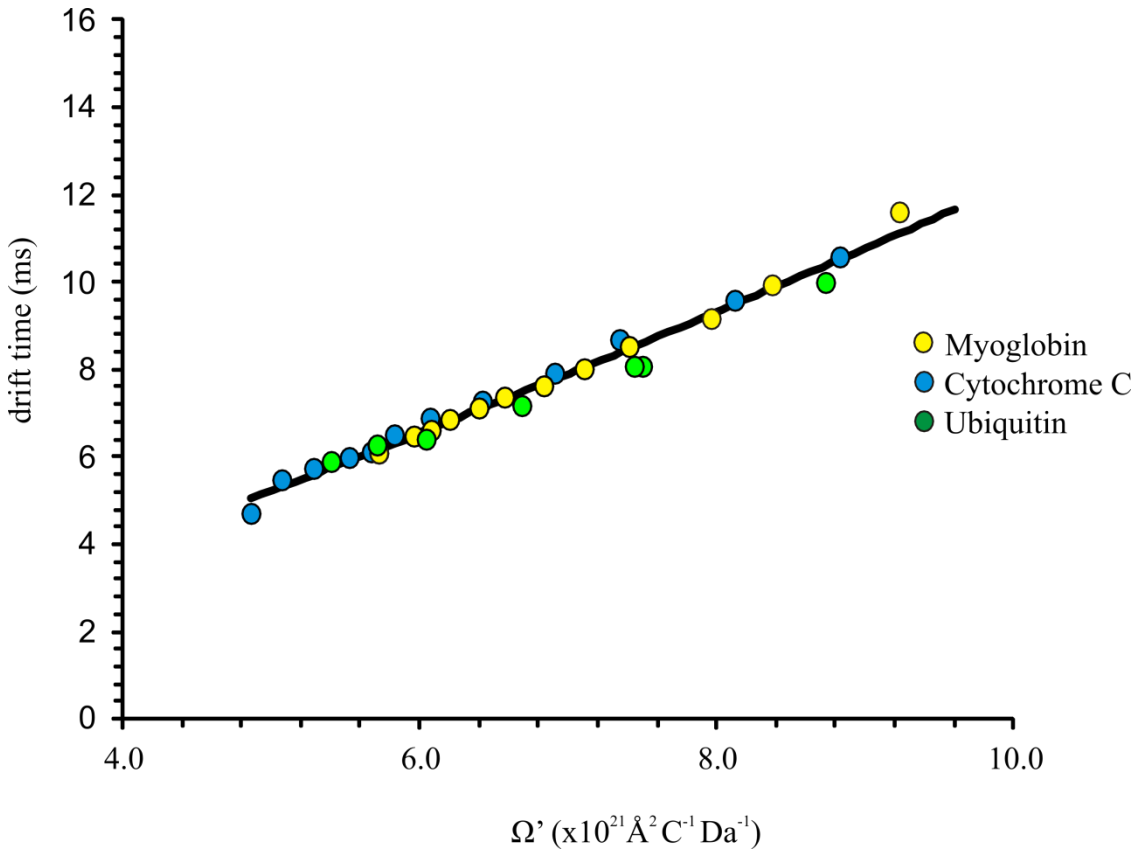


Figure 3.4.1 – Typical example of measured drift times for the proteins, myoglobin (yellow), cytochrome C (blue) and ubiquitin (green) versus Ω' (charge and mass independent collision cross-section) used for IMS calibration. All calibrants were acquired under denaturing conditions. The black line shows the calibration curve, fitted to the relationship $y = Ax^B$.

For all mass spectrometry experiments, the raw data were processed using MassLynx *v*.4.1 software (Waters Ltd., Manchester, UK) in combination with Driftscope *v*.3.0.

3.4.10.3 CID Methods

For CID-MS/MS experiments the *m/z* corresponding to the 9+ β_2m dimer ions was selected in the quadrupole region of the Synapt HDMS mass spectrometer for fragmentation in the transfer collision cell. Increasing collision energy was applied to the transfer in 20 eV increments until the parent ion was completely dissociated into fragment ions. The dimer 9+ was separated by IMS-MS from the tetramer 18+ ions prior to the transfer region of the mass spectrometer (Section 3.4.10.2), and Driftscope *v*.3.0 used to extract all fragment ions corresponding only to the drift time of the dimer 9+. Fragmentation profiles describing the fragmentation pathway of the

dimer 9+ were generated. These display the fraction of ternary ion intensity (as a percentage of total ion intensity) versus normalised collision energy (NCE).

$$NCE = \frac{E \,.\, Ar_m \,.\, z}{Ar_m + C_m} \qquad \text{Equation 3.4.2}$$

where E is the collision energy, Ar_m is the mass of Argon, C_m is the mass of the complex and z the charge of the parent ion.

The collision energy required to fragment 50 % of the dimer 9+ ions was reported for all β_2m variants. CID-MS data were repeated twice and the error reported in terms of ±1 standard deviation.

3.4.10.4 Subunit Exchange Experiments

For subunit exchange experiments, ^{15}N labelled wild-type β_2m was mixed in a 1:1 (*w/w*) ratio with ^{14}N wild-type β_2m or one of the β_2m variants (40 μM ^{15}N β_2m:40 μM β_2m variant) and ESI-MS acquired at regular time intervals using the NanoMate Triversa source. For subunit exchange measurements carried out at 50 % of the lag time of fibril growth, the lag time was estimated (using previously acquired data) and at the correct time point during fibril growth (Section 3.4.8.2 for fibril forming conditions) 100 μL of the β_2m variant was mixed in a 1:1 ratio with 100 μL of ^{15}N labelled wild-type β_2m or ^{15}N labelled β_2m variant (incubated on ice throughout the time course). Mass spectra were acquired immediately and at regular time intervals over the subsequent 4 h. All samples were incubated in the 96 well plate in the NanoMate Triversa source at 20 ºC throughout the subunit exchange time courses. The extent of subunit exchange was determined by comparing the area under the peaks corresponding to the ^{14}N β_2m variant and the ^{15}N wild-type β_2m/β_2m variant with the mixed ^{14}N β_2m variant:^{15}N wild-type β_2m/β_2m variant subunit exchanged oligomers.

References

Anfinsen, C. B. (1973). Principles that govern folding of protein chains. *Science* **181**(4096): 223-230.

Ashcroft, A. E. (2010). Mass spectrometry and the amyloid problem-how far can we go in the gas phase? *J. Am. Soc. Mass Spectrom.* **21**(7): 1087-1096.

Bellotti, V., Gallieni, M., Giorgetti, S. and Brancaccio, D. (2001). Dynamic of beta(2)-microglobulin fibril formation and reabsorption: The role of proteolysis. *Seminars in Dialysis* **14**(2): 117-122.

Bernstein, S. L., Wyttenbach, T., Baumketner, A., Shea, J. E., Bitan, G., Teplow, D. B. and Bowers, M. T. (2005). Amyloid beta-protein: Monomer structure and early aggregation states of a beta 42 and its Pro(19) alloform. *J. Am. Chem. Soc.* **127**(7): 2075-2084.

Bernstein, S. L., Dupuis, N. F., Lazo, N. D., Wyttenbach, T., Condron, M. M., Bitan, G., Teplow, D. B., Shea, J. E., Ruotolo, B. T., Robinson, C. V. and Bowers, M. T. (2009). Amyloid-beta protein oligomerization and the importance of tetramers and dodecamers in the aetiology of alzheimer's disease. *Nat. Chem.* **1**(4): 326-331.

Biancalana, M. and Koide, S. (2010). Molecular mechanism of thioflavin-T binding to amyloid fibrils. *Biochimica Et Biophysica Acta-Proteins and Proteomics* **1804**(7): 1405-1412.

Borysik, A. J. H., Radford, S. E. and Ashcroft, A. E. (2004). Co-populated conformational ensembles of beta(2)-microglobulin uncovered quantitatively by electrospray ionization mass spectrometry. *J. Biol. Chem.* **279**(26): 27069-27077.

Borysik, A. J. H., Read, P., Little, D. R., Bateman, R. H., Radford, S. E. and Ashcroft, A. E. (2004). Separation of beta(2)-microglobulin conformers by high-field asymmetric waveform ion mobility spectrometry (FAIMS) coupled to electrospray ionisation mass spectrometry. *Rapid Comm. Mass Spectrom.* **18**(19): 2229-2234.

Bucciantini, M., Giannoni, E., Chiti, F., Baroni, F., Formigli, L., Zurdo, J. S., Taddei, N., Ramponi, G., Dobson, C. M. and Stefani, M. (2002). Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* **416**(6880): 507-511.

Calabrese, M. F., Eakin, C. M., Wang, J. M. and Miranker, A. D. (2008). A regulatable switch mediates self-association in an immunoglobulin fold. *Nat. Struct. Mol. Biol.* **15**(9): 965-971.

Capeillereblandin, C., Delaveau, T. and Descampslatscha, B. (1991). Structural modifications of human beta-2 microglobulin treated with oxygen-derived radicals. *Biochemical J.* **277**: 175-182.

Celie, P. H. N., Toebes, M., Rodenko, B., Ovaa, H., Perrakis, A. and Schumacher, T. N. M. (2009). UV-induced ligand exchange in MHC class I protein crystals. *J. Am. Chem. Soc.* **131**(34): 12298-12304.

Chiba, T., Hagihara, Y., Higurashi, T., Hasegawa, K., Naiki, H. and Goto, Y. (2003). Amyloid fibril formation in the context of full-length protein - effects of proline mutations on the amyloid fibril formation of beta(2)-microglobulin. *J. Biol. Chem.* **278**(47): 47016-47024.

Chiti, F., Stefani, M., Taddei, N., Ramponi, G. and Dobson, C. M. (2003). Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* **424**(6950): 805-808.

Chiti, F. and Dobson, C. M. (2006). Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**: 333-366.

Davison, A. M. (1995). Beta(2)-microglobulin and amyloidosis: Who is at risk? *Nephrology Dialysis Transplantation* **10**: 48-51.

Dodds, E. D., Blackwell, A. E., Jones, C. M., Holso, K. L., O'Brien, D. J., Cordes, M. H. J. and Wysocki, V. H. (2011). Determinants of gas-phase disassembly behavior in homodimeric protein complexes with related yet divergent structures. *Anal. Chem.* **83**(10): 3881-3889.

Edelhoch, H. (1967). Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry* **6**(7): 1948-1954.

Eichner, T. and Radford, S. E. (2009). A generic mechanism of beta(2)-microglobulin amyloid assembly at neutral pH involving a specific proline switch. *J. Mol. Biol.* **386**(5): 1312-1326.

Eichner, T., Kalverda, A. P., Thompson, G. S., Homans, S. W. and Radford, S. E. (2011). Conformational conversion during amyloid formation at atomic resolution. *Mol. Cell* **41**(2): 161-172.

Esposito, G., Michelutti, R., Verdone, G., Viglino, P., Hernandez, H., Robinson, C. V., Amoresano, A., Dal Piaz, F., Monti, M., Pucci, P., Mangione, P., Stoppini, M., Merlini, G., Ferri, G. and Bellotti, V. (2000). Removal of the N-terminal hexapeptide from human beta 2-microglobulin facilitates protein aggregation and fibril formation. *Protein Sci.* **9**(5): 831-845.

Esposito, G., Ricagno, S., Corazza, A., Rennella, E., Gumral, D., Mimmi, M. C., Betto, E., Pucillo, C. E. M., Fogolari, F., Viglino, P., Raimondi, S., Giorgetti, S., Bolognesi, B., Merlini, G., Stoppini, M., Bolognesi, M. and Bellotti, V. (2008). Aggregation properties the controlling roles of Trp60 and Trp95 in beta(2)-microglobulin function, folding and amyloid aggregation properties. *J. Mol. Biol.* **378**(4): 887-897.

Floege, J. and Ehlerding, G. (1996). Beta-2-microglobulin-associated amyloidosis - discussion. *Nephron* **72**(1): 9-26.

Frimpong, A. K., Abzatimov, R. R., Uversky, V. N. and Kaltashov, I. A. (2010). Characterization of intrinsically disordered proteins with electrospray ionization mass spectrometry: Conformational heterogeneity of alpha-synuclein. Proteins-Structure Function and Bioinformatics **78**(3): 714-722.

Gessel, M. M., Bernstein, S., Kemper, M., Teplow, D. B. and Bowers, M. T. (2012a). Familial alzheimer's disease mutations differentially alter amyloid β-protein oligomerization. *Chemical Neuroscience*. DOI: 10.1021/cn300050d.

Gessel, M. M., Wu, C., Li, H. Y., Bitan, G., Shea, J. E. and Bowers, M. T. (2012b). A beta(39-42) modulates a beta oligomerization but not fibril formation. *Biochemistry* **51**(1): 108-117.

Gill, S. C. and Vonhippel, P. H. (1989). Calculation of protein extinction coefficients from amino-acid sequence data. *Anal. Biochem.* **182**(2): 319-326.

Gosal, W. S., Morten, I. J., Hewitt, E. W., Smith, D. A., Thomson, N. H. and Radford, S. E. (2005). Competing pathways determine fibril morphology in the self-assembly of beta(2)-microglobulin into amyloid. *J. Mol. Biol.* **351**(4): 850-864.

Gras, S. L., Waddington, L. J. and Goldie, K. N. (2011). Transmission electron microscopy of amyloid fibrils. Protein folding, misfolding, and disease: *Methods and protocols* **752**: 197-214.

Harper, J. D. and Lansbury, P. T. (1997). Models of amyloid seeding in alzheimier's disease and scrapie: Mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins. *Annu. Rev. Biochem.* **66**: 385-407.

Hartl, F. U. (1996). Molecular chaperones in cellular protein folding. *Nature* **381**(6583): 571-580.

Hasegawa, K., Ohhashi, Y., Yamaguchi, I., Takahashi, N., Tsutsumi, S., Goto, Y., Gejyo, F. and Naiki, H. (2003). Amyloidogenic synthetic peptides of beta 2-microglobulin - a role of the disulfide bond. *Biochem. Biophys. Res. Comm.* **304**(1): 101-106.

Hatters, D. M. and Griffin, M. D. W. (2011). Diagnostics for amyloid fibril formation: Where to begin? Protein folding, misfolding, and disease: *Methods and protocols* **752**: 121-136.

Heegaard, N. H. H., Roepstorff, P., Melberg, S. G. and Nissen, M. H. (2002). Cleaved beta(2)-microglobulin partially attains a conformation that has amyloidogenic features. *J. Biol. Chem.* **277**(13): 11184-11189.

Hodkinson, J. P., Jahn, T. R., Radford, S. E. and Ashcroft, A. E. (2009). HDX-ESI-MS reveals enhanced conformational dynamics of the amyloidogenic protein beta(2)-microglobulin upon release from the MHC-1. *J. Am. Soc. Mass Spectrom.* **20**(2): 278-286.

Hodkinson, J. P., Radford, S. E. and Ashcroft, A. E. (2012). The role of conformational flexibility in beta(2)-microglobulin amyloid fibril formation at neutral pH. *Rapid Comm. Mass Spectrom.* **26**(16): 1783-1792.

Hortschansky, P., Schroeckh, V., Christopeit, T., Zandomeneghi, G. and Fandrich, M. (2005). The aggregation kinetics of Alzheimer's beta-amyloid peptide is controlled by stochastic nucleation. *Protein Sci*. **14**(7): 1753-1759.

Ivanova, M. I., Gingery, M., Whitson, L. J. and Eisenberg, D. (2003). Role of the C-terminal 28 residues of beta 2-microglobulin in amyloid fibril formation. *Biochemistry* **42**(46): 13536-13540.

Ivanova, M. I., Sawaya, M. R., Gingery, M., Attinger, A. and Eisenberg, D. (2004). An amyloid-forming segment of beta 2-microglobulin suggests a molecular model for the fibril. *Proc. Nat. Acad. Sci. U.S.A.* **101**(29): 10584-10589.

Ivanova, M. I., Thompson, M. J. and Eisenberg, D. (2006). A systematic screen of beta(2)-microglobulin and insulin for amyloid-like segments. *Proc. Nat. Acad. Sci. U.S.A.* **103**(11): 4079-4082.

Iwata, K., Matsuura, T., Sakurai, K., Nakagawa, A. and Goto, Y. (2007). High-resolution crystal structure of beta(2)-microglobulin formed at pH 7.0. Journal of *Biochemistry* **142**(3): 413-419.

Jones, S., Manning, J., Kad, N. M. and Radford, S. E. (2003). Amyloid-forming peptides from beta(2)-microglobulin - insights into the mechanism of fibril formation in vitro. *J. Mol. Biol.* **325**(2): 249-257.

Jorgensen, T. J. D., Gardsvoll, H., Ploug, M. and Roepstorff, P. (2005). Intramolecular migration of amide hydrogens in protonated peptides upon collisional activation. *J. Am. Chem. Soc.* **127**(8): 2785-2793.

Kad, N. M., Thomson, N. H., Smith, D. P., Smith, D. A. and Radford, S. E. (2001). Beta(2)-microglobulin and its deamidated variant, N17D form amyloid fibrils with a range of morphologies *in vitro*. *J. Mol. Biol.* **313**(3): 559-571.

Kashchiev, D. and Auer, S. (2010). Nucleation of amyloid fibrils. *J. Chem. Phys*.**132**(21).

Kloniecki, M., Jablonowska, A., Poznanski, J., Langridge, J., Hughes, C., Campuzano, I., Giles, K. and Dadlez, M. (2011). Ion mobility separation coupled with ms detects two structural states of alzheimer's disease a beta 1-40 peptide oligomers. *J. Mol. Biol.* **407**(1): 110-124.

Knapman, T. W., Berryman, J. T., Campuzano, I., Harris, S. A. and Ashcroft, A. E. (2010). Considerations in experimental and theoretical collision cross-section measurements of small molecules using travelling wave ion mobility spectrometry-mass spectrometry. *Int. J. Mass Spectrom.* **298**(1-3): 17-23.

Konermann, L., Stocks, B. B., Pan, Y. and Tong, X. (2010). Mass spectrometry combined with oxidative labeling for exploring protein structure and folding. *Mass Spectrom. Rev.* **29**(4): 651-667.

Kozhukh, G. V., Hagihara, Y., Kawakami, T., Hasegawa, K., Naiki, H. and Goto, Y. (2002). Investigation of a peptide responsible for amyloid fibril formation of beta(2)-microglobulin by Achromobacter protease I. *J. Biol. Chem.* **277**(2): 1310-1315.

Ladewig, P. (1945). Double-refringence of the amyloid-congo-red-complex in histological sections. *Nature* **156**(3951): 81-82.

Lue, L. F., Kuo, Y. M., Roher, A. E., Brachova, L., Shen, Y., Sue, L., Beach, T., Kurth, J. H., Rydel, R. E. and Rogers, J. (1999). Soluble amyloid beta peptide concentration as a predictor of synaptic change in Alzheimer's disease. *Am. J. Pathology* **155**(3): 853-862.

Liu, Y. S., Valentine, S. J., Counterman, A. E., Hoaglund, C. S. and Clemmer, D. E. (1997). Injected-ion mobility analysis of biomolecules. *Anal. Chem.* **69**(23): 728-735.

Luhrs, T., Ritter, C., Adrian, M., Riek-Loher, D., Bohrmann, B., Doeli, H., Schubert, D. and Riek, R. (2005). 3D structure of Alzheimer's amyloid-beta(1-42) fibrils. *Proc. Nat. Acad. Sci. U.S.A.* **102**(48): 17342-17347.

McCammon, M. G. and Robinson, C. V. (2005). Me, my cell, and I: The role of the collision cell in the tandem mass spectrometry of macromolecules. *Biotechniques* **39**(4): 447.

Mendoza, V. L., Antwi, K., Baron-Rodriguez, M. A., Blanco, C. and Vachet, R. W. (2010). Structure of the preamyloid dimer of beta-2-microglobulin from covalent labeling and mass spectrometry. *Biochemistry* **49**(7): 1522-1532.

Mendoza, V. L., Baron-Rodriguez, M. A., Blanco, C. and Vachet, R. W. (2011). Structural insights into the pre-amyloid tetramer of beta-2-microglobulin from covalent labeling and mass spectrometry. *Biochemistry* **50**(31): 6711-6722.

Miyata, T., Taneda, S., Kawai, R., Ueda, Y., Horiuchi, S., Hara, M., Maeda, K. and Monnier, V. M. (1996). Identification of pentosidine as a native structure for advanced glycation end products in beta(2)-microglobulin-containing amyloid fibrils in patients with dialysis-related amyloidosis. *Proc. Nat. Acad. Sci. U.S.A.* **93**(6): 2353-2358.

Myers, S. L., Jones, S., Jahn, T. R., Morten, I. J., Tennent, G. A., Hewitt, E. W. and Radford, S. E. (2006). A systematic study of the effect of physiological factors on beta 2-microglobulin amyloid formation at neutral pH. *Biochemistry* **45**(7): 2311-2321.

Naiki, H., Higuchi, K., Hosokawa, M. and Takeda, T. (1989). Fluorometric-determination of amyloid fibrils *in vitro* using the fluorescent dye, thioflavine-T. *Anal. Biochem.* **177**(2): 244-249.

Natalello, A., Benetti, F., Doglia, S. M., Legname, G. and Grandori, R. (2011). Compact conformations of alpha-synuclein induced by alcohols and copper. *Proteins-Structure Function and Bioinformatics* **79**(2): 611-621.

Niwa, T., Katsuzaki, T., Miyazaki, S., Momoi, T., Akiba, T., Miyazaki, T., Nokura, K., Hayase, F., Tatemichi, N. and Takei, Y. (1997). Amyloid beta(2)-microglobulin is modified with imidazolone, a novel advanced glycation end product, in dialysis-related amyloidosis. *Kidney Int.* **51**(1): 187-194.

Okon, M., Bray, P. and Vucelic, D. (1992). H-1-nmr assignments and secondary structure of human beta-2-microglobulin in solution. *Biochemistry* **31**(37): 8906-8915.

Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. and Gray, T. (1995). How to measure and predict the molar absorption-coefficient of a protein. *Protein Sci.* **4**(11): 2411-2423.

Pan, J., Han, J., Borchers, C. H. and Konermann, L. (2008). Electron capture dissociation of electrosprayed protein ions for spatially resolved hydrogen exchange measurements. *J. Am. Chem. Soc.* **130**(35): 11574.

Pan, Y., Brown, L. and Konermann, L. (2011). Kinetic folding mechanism of an integral membrane protein examined by pulsed oxidative labeling and mass spectrometry. *J. Mol. Biol.* **410**(1): 146-158.

Pan, Y., Ruan, X., Valvano, M. A. and Konermann, L. (2012). Validation of membrane protein topology models by oxidative labeling and mass spectrometry. *J. Am. Soc. Mass Spectrom.* **23**(5): 889-898.

Park, S. and Saven, J. G. (2006). Simulation of pH-dependent edge strand rearrangement in human beta-2 microglobulin. *Protein Sci.* **15**(1): 200-207.

Pawar, A. P., DuBay, K. F., Zurdo, J., Chiti, F., Vendruscolo, M. and Dobson, C. M. (2005). Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases. *J. Mol. Biol.* **350**(2): 379-392.

Percy, A. J., Rey, M., Burns, K. M. and Schriemer, D. C. (2012). Probing protein interactions with hydrogen/deuterium exchange and mass spectrometry-a review. *Analytica Chimica Acta* **721**: 7-21.

Platt, G. W., McParland, V. J., Kalverda, A. P., Homans, S. W. and Radford, S. E. (2005). Dynamics in the unfolded state of beta(2)-microglobulin studied by NMR. *J. Mol. Biol.* **346**(1): 279-294.

Platt, G. W., Routledge, K. E., Homans, S. W. and Radford, S. E. (2008). Fibril growth kinetics reveal a region of beta(2)-microglobulin important for nucleation and elongation of aggregation. *J. Mol. Biol.* **378**(1): 251-263.

Rand, K. D., Zehl, M., Jensen, O. N. and Jorgensen, T. J. D. (2009). Protein hydrogen exchange measured at single-residue resolution by electron transfer dissociation mass spectrometry. *Anal. Chem.* **81**(14): 5577-5584.

Rand, K. D., Pringle, S. D., Morris, M. and Brown, J. M. (2012). Site-specific analysis of gas-phase hydrogen/deuterium exchange of peptides and proteins by electron transfer dissociation. *Anal. Chem.* **84**(4): 1931-1940.

Reid, S. W., Smith, K. J., Jakobsen, B. K., Ocallaghan, C. A., Reyburn, H., Harlos, K., Stuart, D. I., McMichael, A. J., Bell, J. I. and Jones, E. Y. (1996). Production and crystallization of MHC class I b allele single peptide complexes. *FEBS Lett.* **383**(1-2): 119-123.

Reixach, N., Deechongkit, S., Jiang, X., Kelly, J. W. and Buxbaum, J. N. (2004). Tissue damage in the amyloidoses: Transthyretin monomers and nonnative oligomers are the major cytotoxic species in tissue culture. *Proc. Nat. Acad. Sci. U.S.A.* **101**(9): 2817-2822.

Ricagno, S., Colombo, M., de Rosa, M., Sangiovanni, E., Giorgetti, S., Raimondi, S., Bellotti, V. and Bolognesi, M. (2008). DE loop mutations affect beta 2-microglobulin stability and amyloid aggregation. *Biochem. Biophys. Res. Comm.* **377**(1): 146-150.

Richardson, J. S. and Richardson, D. C. (2002). Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc. Nat. Acad. Sci. U.S.A.* **99**(5): 2754-2759.

Routledge, K. E., Tartaglia, G. G., Platt, G. W., Vendrusco, M. and Radford, S. E. (2009). Competition between intramolecular and intermolecular interactions in an amyloid-forming protein. *J. Mol. Biol.* **389**(4): 776-786.

Sambrook, J., Fritsch, E.F., Maniatis T. (1989). Molecular cloning - a laboratory manual.

Santambrogio, C., Ricagno, S., Colombo, M., Barbiroli, A., Bonomi, F., Bellotti, V., Bolognesi, M. and Grandori, R. (2010). DE-loop mutations affect beta 2 microglobulin stability, oligomerization, and the low-pH unfolded form. *Protein Sci.* **19**(7): 1386-1394.

Sciuto, S. V., Liu, J. J. and Konermann, L. (2011). An electrostatic charge partitioning model for the dissociation of protein complexes in the gas phase. *J. Am. Soc. Mass Spectrom.* **22**(10): 1679-1689.

Shelimov, K. B., Clemmer, D. E., Hudgins, R. R. and Jarrold, M. F. (1997). Protein structure *in vacuo*: Gas-phase confirmations of BPTI and cytochrome C. *J. Am. Chem. Soc.* **119**(9): 2240-2248.

Shields, M. J., Assefi, N., Hodgson, W., Kim, E. J. and Ribaudo, R. K. (1998). Characterization of the interactions between MHC class I subunits: A systematic

approach for the engineering of higher affinity variants of beta(2)-microglobulin. *J. Immunology* **160**(5): 2297-2307.

Smith, D. P. and Radford, S. E. (2001). Role of the single disulphide bond of beta(2)-microglobulin in amyloidosis *in vitro*. *Protein Sci.* **10**(9): 1775-1784.

Smith, D. P., Jones, S., Serpell, L. C., Sunde, M. and Radford, S. E. (2003). A systematic investigation into the effect of protein destabilisation on beta 2-microglobulin amyloid formation. *J. Mol. Biol.* **330**(5): 943-954.

Smith, D. P., Giles, K., Bateman, R. H., Radford, S. E. and Ashcroft, A. E. (2007). Monitoring copopulated conformational states during protein folding events using electrospray ionization-ion mobility spectrometry-mass spectrometry. *J. Am. Soc. Mass Spectrom.* **18**(12): 2180-2190.

Smith, D. P., Knapman, T. W., Campuzano, I., Malham, R. W., Berryman, J. T., Radford, S. E. and Ashcroft, A. E. (2009). Deciphering drift time measurements from travelling wave ion mobility spectrometry-mass spectrometry studies. *Eur. J. Mass Spectrom.* **15**(2): 113-130.

Smith, D. P., Radford, S. E. and Ashcroft, A. E. (2010). Elongated oligomers in beta(2)-microglobulin amyloid assembly revealed by ion mobility spectrometry-mass spectrometry. *Proc. Nat. Acad. Sci. U.S.A.* **107**(15): 6794-6798.

Stromer, T. and Serpell, L. C. (2005). Structure and morphology of the Alzheimer's amyloid fibril. *Microscopy Research and Technique* **67**(3-4): 210-217.

Tomic, J. L., Pensalfini, A., Head, E. and Glabe, C. G. (2009). Soluble fibrillar oligomer levels are elevated in Alzheimer's disease brain and correlate with cognitive dysfunction. Neurobiology of Disease **35**(3): 352-358.

Trinh, C. H., Smith, D. P., Kalverda, A. P., Phillips, S. E. V. and Radford, S. E. (2002). Crystal structure of monomeric human beta-2-microglobulin reveals clues to its amyloidogenic properties. *Proc. Nat. Acad. Sci. U.S.A.* **99**(15): 9771-9776.

Valentine, S. J., Counterman, A. E. and Clemmer, D. E. (1997). Conformer-dependent proton-transfer reactions of ubiquitin ions. *J. Am. Soc. Mass Spectrom.* **8**(9): 954-961.

van Ypersele de Strihou, C., Jadoul, M., Malghem, J., Maldague, B. and Jamart, J. (1991). Effect of dialysis membrane and patient's age on signs of dialysis-related amyloidosis. The working party on dialysis amyloidosis. *Kidney Int.* **39**(5): 1012-1019.

Verdone, G., Corazza, A., Viglino, P., Pettirossi, F., Giorgetti, S., Mangione, P., Andreola, A., Stoppini, M., Bellotti, V. and Esposito, G. (2002). The solution structure of human beta 2-microglobulin reveals the prodromes of its amyloid transition. *Protein Sci.* **11**(3): 487-499.

Wales, T. E. and Engen, J. R. (2006). Partial unfolding of diverse SH3 domains on a wide timescale. *J. Mol. Biol.* **357**(5): 1592-1604.

Wang, J., Dickson, D. W., Trojanowski, J. Q. and Lee, V. M. Y. (1999). The levels of soluble versus insoluble brain A beta distinguish Alzheimer's disease from normal and pathologic aging. *Experimental Neurology* **158**(2): 328-337.

Westermark, G. T., Johnson, K. H. and Westermark, P. (1999). Staining methods for identification of amyloid in tissue. *Amyloid, Prions, and Other Protein Aggregates* **309**: 3-25.

Winner, B., Jappelli, R., Maji, S. K., Desplats, P. A., Boyer, L., Aigner, S., Hetzer, C., Loher, T., Vilar, M., Campionic, S., Tzitzilonis, C., Soragni, A., Jessberger, S., Mira, H., Consiglio, A., Pham, E., Masliah, E., Gage, F. H. and Riek, R. (2011). *In vivo* demonstration that alpha-synuclein oligomers are toxic. *Proc. Nat. Acad. Sci. U.S.A.* **108**(10): 4194-4199.

Xue, W. F., Homans, S. W. and Radford, S. E. (2008). Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly. *Proc. Nat. Acad. Sci. U.S.A.* **105**(26): 8926-8931.

Membrane Proteins

4.1. Introduction

Integral membrane proteins play essential roles in many biological processes such as transport of solutes, signalling and energy transduction. Despite this, our fundamental knowledge of the structure of membrane proteins is limited. Membrane proteins are notoriously difficult to study; their expression and purification is challenging and problems occur when trying to preserve the structural activity and the stability of these proteins *in vitro* (Seddon *et al.* 2004).

Increasing numbers of biochemical techniques are being used to study membrane proteins and their complexes in solution (Marassi *et al.* 1998; Garcia-Saez *et al.* 2007; Carpenter *et al.* 2008; Raunser and Walz 2009). Mass spectrometry will be the primary analytical tool used for membrane protein analysis in this thesis. Although non-covalent ESI-MS of intact membrane proteins has only been established relatively recently (Barrera *et al.* 2011), mass spectrometry is a good technique to study membrane proteins since it can determine protein conformation (Wang *et al.* 2010) and the subunit stoichiometry of biomolecular complexes (Barrera *et al.* 2009), in addition to identifying any post-translational modifications present (Bechara *et al.* 2012).

4.1.1 Mass Spectrometry of Membrane Proteins

Membrane proteins are insoluble in volatile aqueous/organic solvents used typically in mass spectrometry analysis of proteins and their complexes. Membrane protein analysis by mass spectrometry has been challenging due to their hydrophobicity and thus to the large excess of detergent molecules needed to solubilise the membrane protein complexes in solution. Under soft ionisation conditions, membrane protein peaks may be swamped by multiple peaks arising from the detergent clusters in the *m/z* spectrum, making membrane protein analysis almost impossible. Initial ESI-MS analysis of the *E.coli* multidrug resistant transporter protein (EmrE) bound to the tetraphenylphosphonium cation (TPP^+) in dodecylmaltoside (DDM) detergent,

showed broad peaks in the *m/z* spectrum corresponding to a heterogeneous distribution of DDM molecules binding to the protein-ligand complex (Figure 4.1.1A) (Ilag *et al.* 2004; Tate 2006). These broad peaks made the exact molecular weight of the protein difficult to assign. The data showed great promise that membrane proteins complexed with detergent micelles could be preserved during the electrospray ionisation process and into the gas phase. Further analysis showed that on subjecting these broad peaks to collision induced dissociation, EmrE and TPP^+ were released from DDM and peaks detected corresponding to the TPP^+ and the 6+ charge state ions of EmrE (Ilag *et al.* 2004) (Figure 4.1.1B).

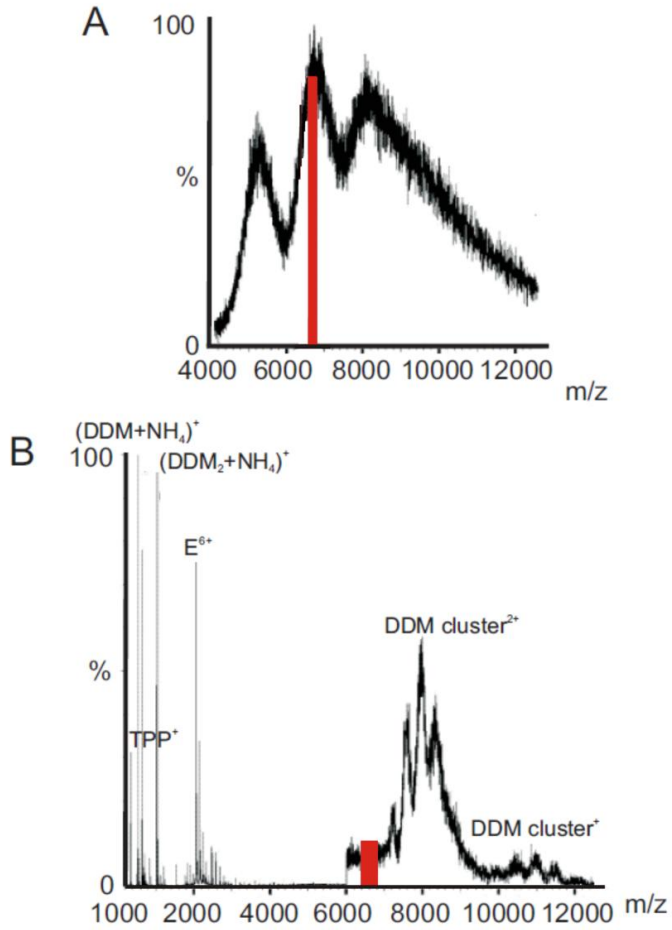


Figure 4.1.1 A) ESI-MS of EmrE in dodecylmaltoside (DDM) with tetraphenyl phosphonium (TPP^+); B) CID-MS of ion isolated at 6500 ± 100 *m/z* (red). Peaks corresponding to the TPP^+, EmrE (E) and DDM clusters products, released from the EmrE-TPP^+-DDM complex, are highlighted in the *m/z* spectrum. This Figure was adapted from Ilag *et al.* (Ilag *et al.* 2004).

This breakthrough from the Robinson group in ESI-MS of membrane protein complexes led to the investigation of larger multi-subunit membrane protein

complexes. The first membrane protein complex to be analysed by ESI-MS was $BtuC_2D_2$, a vitamin B_{12} importer containing two BtuC transmembrane subunits and two BtuD cytoplasmic subunits (Locher *et al.* 2002; Barrera *et al.* 2008). Upon increasing the activation energy applied to the $BtuC_2D_2$-detergent micelle complex in the gas phase, the intact tetramer was released from the micelle and the corresponding ions detected in the *m/z* spectrum (Barrera *et al.* 2008). ESI-IMS-MS was used to determine the collision cross-section of the $BtuC_2D_2$ complex ejected from the detergent micelle (Wang *et al.* 2010). On comparison of this collision cross-section with that predicted based on the crystal structure, it was suggested that the high energy applied to release membrane proteins from detergent micelles in the gas phase did not affect membrane protein structure, with the gas phase cross-section measured mirroring that of the modelled cross-section for $BtuC_2D_2$ (Wang *et al.* 2010). More recently, the use of ESI-MS has expanded to include more complex, large macromolecular machines, for example the V-type ATPases from *Thermus thermophilus* and *Enterococcus hirae* (Zhou *et al.* 2011). In this study not only was the subunit stoichiometry of the complex determined, but also a potential regulatory role of nucleotide binding on ATP hydrolysis and proton translocation revealed.

Currently, the most common method to study intact membrane proteins by mass spectrometry involves initial solubilisation of the proteins in micelle-forming detergent molecules. Despite this breakthrough in membrane protein analysis, detergent micelles are not optimal for preserving membrane protein function due to their highly dynamic nature and thus protein unfolding and aggregation in detergent micelles are commonly observed (Garavito *et al.* 2001; Zhou *et al.* 2001). In addition, the spherical micelles do not provide a native-like environment, poorly mimicking the physical properties of the planar cellular membrane. Alternative membrane solubilisation techniques are therefore needed to be able to explore the dynamic nature and mode of functioning of membrane proteins (Seddon *et al.* 2004; Popot 2010).

4.1.2 Amphipols

One alternative to detergent micelles for solubilising membrane proteins are amphipols (APols): amphipathic polymers containing both hydrophobic and hydrophilic groups. Amphipols were designed to bind non-covalently to the transmembrane region of membrane proteins in a quasi-irreversible manner (Popot 2010; Popot *et al.* 2011). The amphipol A8-35, composed of octyl and carboxyl chains, was the first amphipol designed and synthesised to bind membrane proteins, and as a result is the most well studied amphipol to date (Figure 4.1.2A). At, and above, pH 7 most carboxylates in A8-35 are ionised in aqueous solution making the amphipol highly soluble in water. In aqueous solution, A8-35 self-assembles into a tetramer with an average particle size of 40 kDa as determined by small angle neutron scattering, NMR, analytical ultra-centrifugation, static and dynamic light scattering and size exclusion chromatography (Gohon *et al.* 2006).

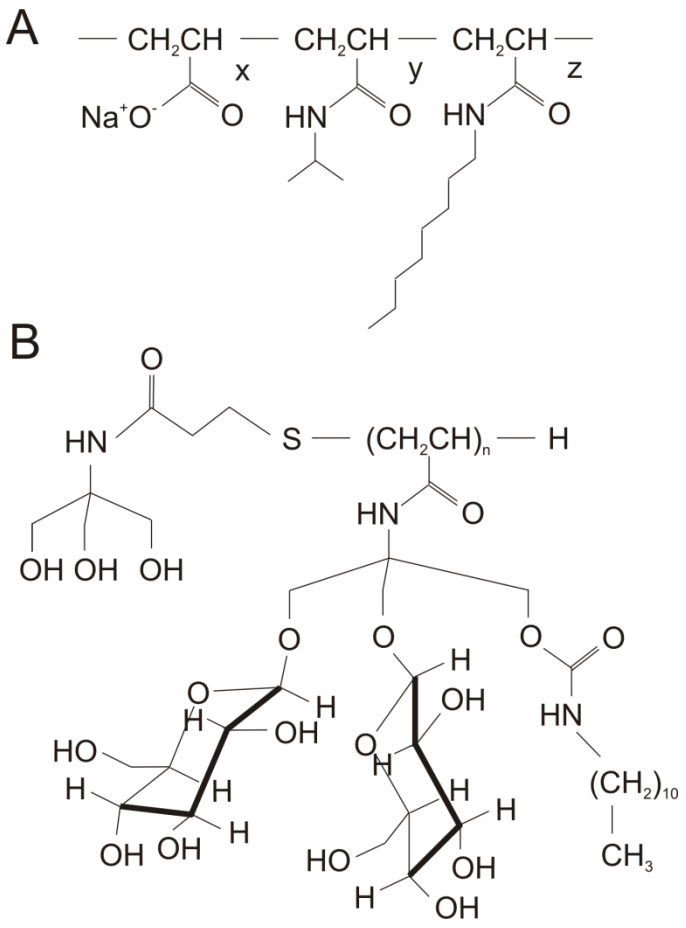


Figure 4.1.2 – Chemical structures of A) the amphipol, A8-35 where x = 29-34%, y = 25-28%, z =39-44% and B) the homopolymeric, glucosylated, non-ionic amphipol (NAPol). Figure adapted from Popot *et al.* (Popot *et al.* 2011).

Unfortunately, due to the high dependence on the ionisation status of the carboxylates for A8-35 solubility, membrane proteins in complex with the amphipol A8-35 can only be studied under basic conditions. These experimental constraints led to the design of novel non-ionic amphipols called NAPols (Figure 4.1.2B) (Sharma *et al.* 2008). As such, NAPols can now be used to solubilise membrane proteins over a wide pH range (Sharma *et al.* 2008; Bazzacco *et al.* 2012). Due to the lack of commercially available NAPols, the amphipol A8-35 has been used to trap the membrane proteins used in this thesis.

4.1.3 Folding Membrane Proteins into Amphipols

Apart from a few exceptions (Popot *et al.* 2003), amphipols cannot be used to extract membrane proteins directly from cell membranes. To transfer membrane proteins into amphipols, membrane proteins are first extracted and purified in detergent using standard methods (Arnold *et al.* 2008). The folded detergent-membrane protein mixture is then supplemented with amphipol and briefly incubated, before detergent removal from the detergent-amphipol-membrane protein complex (Figure 4.1.3A). Detergent can be removed from the solution in two ways. Firstly, the solution can be diluted to below the critical micelle concentration (CMC) of the detergent. Below the CMC, detergent disperses into its monomeric form which can be removed from the solution by dialysis. This method was used successfully to trap the nicotinic acetylcholine receptor (nAChR) in the amphipol, A8-35, restoring the native-like allosteric properties of nAChR (Martinez *et al.* 2002). Alternatively, detergent can be removed by the addition of polystyrene beads (Bio-Beads SM2). Detergent can absorb onto the Bio-Beads which can then be removed readily by centrifugation (Zoonens *et al.* 2005). This method is advantageous over the previous method for removing detergent as the sample remains concentrated and no dilution of the sample is required prior to analysis.

Membrane proteins can also be folded into amphipols straight from their denatured state (Figure 4.1.3B) (Pocanschi *et al.* 2006; Baneres *et al.* 2011). Two outer membrane β-barrel proteins, OmpA and FomA, were initially selected for analysis (Pocanschi *et al.* 2006). On dilution of the urea unfolded OmpA and FomA into urea-free buffer in the presence of the amphipol, A8-35, both proteins immediately

displayed β-sheet secondary structure (as analysed by circular dichroism), an electrophoretic mobility comparable to the native state (as analysed by cold SDS-PAGE) and resistance to trypsin proteolysis, suggesting both proteins had folded completely into A8-35 (Pocanschi *et al.* 2006). The folding of membrane proteins into amphipols from their denatured state has been extended more recently to a variety of G protein coupled receptors (GPCRs), including LTB_4 (Dahmane *et al.* 2009), 5-HT_{4a} a serotonin receptor (Dahmane *et al.* 2009), CB1 a cannabinoid receptor (Dahmane *et al.* 2009) and the ghrelin receptor GHSR-1a (Baneres *et al.* 2011).

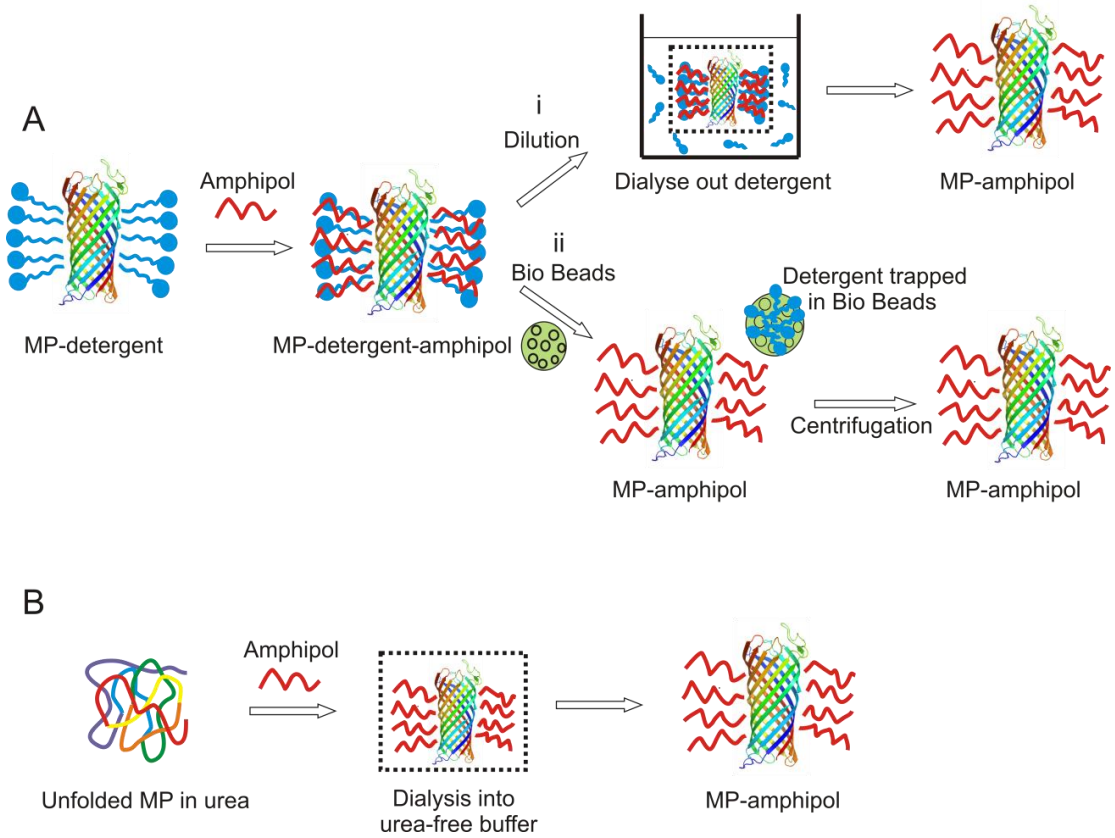


Figure 4.1.3 - Trapping membrane proteins (MP) in amphipols (red curved line). A) Upon addition of amphipol to the MP-detergent complex, detergent (blue) can be removed either (i) by dilution to below the CMC of the detergent followed by dialysis or (ii) by addition of Bio-Beads (green); B) Membrane protein folding from its urea-denatured state into amphipols is achieved by dialysis or rapid dilution into urea-free buffer.

Membrane protein folding into amphipols from their denatured state presents several advantages compared with transferring membrane proteins to amphipols from detergent. For well-studied membrane proteins, the detergent in which these

membrane proteins are stable is known. In most cases the membrane protein of interest has not been well characterised and the detergent type in which the membrane proteins is stable will take time to determine. The ability to purify membrane proteins in their denatured state and fold them directly into amphipols avoids these challenging steps of detergent optimisation and associated problems with membrane protein stability. The use of amphipols in this way, therefore, is thought to greatly simplify membrane protein sample preparation and is the method chosen in this thesis to trap membrane proteins for analysis by ESI-MS.

The membrane protein:amphipol ratio is a key parameter that needs to be optimised for preparation of membrane protein samples. If the membrane protein:amphipol ratio is too low, problems occur with protein solubility leading to aggregation. Alternatively, if the membrane protein:amphipol ratio is too high, the high concentration of amphipol can increase the viscosity of the solution to a level that compromises membrane protein function. The optimal membrane protein:amphipol ratio was investigated initially using the prototypical small α-helical inner membrane protein, bacteriorhodopsin, where a 1:3 (*w/w*) ratio of bacteriorhodopsin to amphipol, A8-35 produced the largest yield of natively folded bacteriorhodopsin (Gohon *et al.* 2008). Trapping experinents on larger protein complexes normally require a higher membrane protein:amphipol ratio of 1:5 (*w/w*) such as in the case for the multi-subunit cytochrome bc_1 membrane protein complex (Bechara *et al.* 2012).

4.1.4 Stability of Amphipol-trapped Membrane Proteins

One of the major advantages of solubilising membrane proteins in amphipols is the increased stability the amphipol provides to the protein (Popot *et al.* 2003; Popot 2010). Indeed, bacteriorhodopsin showed no evidence of denaturation over six months of storage at 4 ºC in amphipol compared with only a few days in the detergent, octylthioglucoside (Popot *et al.* 2011). Bacteriorhodopsin-A8-35 complexes can even survive freeze/thaw cycles without protein denaturation (Gohon *et al.* 2008).

The increase in membrane protein stability in amphipols has led to the use of amphipols as vectors for the delivery of membrane proteins and transmembrane peptides *in vivo*. Indeed, recent data have shown that vaccines formulated with the major outer membrane protein from *Chlamydia trachomatis* (MOMP):amphipol complex can protect mice against the *Chlamydia trachomatis* infection better than vaccines containing MOMP:detergent complexes, and to a similar level to vaccination with the live organism itself (Tifrea *et al.* 2011). Amphipols may greatly aid future vaccine developments in addition to helping solubilise membrane proteins/peptides for other therapeutic applications (Popot *et al.* 2011).

Over the last sixteen years, amphipols have been employed to study membrane protein complexes in conjunction with a variety of biochemical techniques including size exclusion chromatography (Gohon *et al.* 2008), electron microscopy (Cvetkov *et al.* 2011), analytical ultra-centrifugation (Gohon *et al.* 2008), fluorescence resonance energy transfer (Zoonens *et al.* 2007) and solution NMR (Zoonens *et al.* 2005; Catoire *et al.* 2009; Catoire *et al.* 2010). The application of amphipol-trapped membrane proteins to ESI-MS has not been developed. A recent publication (appearing after commencement of these studies) used MALDI-TOF MS determine the molecular mass of the amphipol-trapped membrane proteins, bacteriorhodopsin, OmpA, cytochrome b_6f and cytochrome bc_1 (Bechara *et al.* 2012). The use of MALDI rather than ESI meant that only molecular masses could be obtained, and no structural details of the membrane proteins in the gas phase could be reported. Additionally, the harsher conditions of MALDI compared with ESI caused the membrane protein complexes analysed to dissociate prior to MS analysis (Bechara *et al.* 2012).

This thesis chapter focuses on the ESI-MS analysis of two bacterial outer membrane proteins. PagP and OmpT. Both PagP and OmpT are synthesised in the cytoplasm and translocated across the inner membrane via the Sec translocon (Bos *et al.* 2007) before they fold correctly into the outer membrane with help from periplasmic chaperones and the Bam complex (Mogensen *et al.* 2005; Knowles *et al.* 2009).

4.1.5 Lipid A Palmitoyl transferase, PagP

PhoPQ-activated gene P (PagP) is an eight-stranded, 20 kDa, β-barrel protein with a small N-terminal α-helix (Hwang *et al.* 2002; Ahn *et al.* 2004). Once inserted into the outer membrane, PagP catalyses the transfer of a palmitate chain from phospholipids to lipid A, the hydrophobic anchor of the lipopolysaccharide, LPS. This helps reinforce the outer membrane protecting the bacterium from host immune defences (Bishop 2005). Lipid A acylation has shown resistance to cationic antimicrobial peptides *in vivo* (Guo *et al.* 1998).

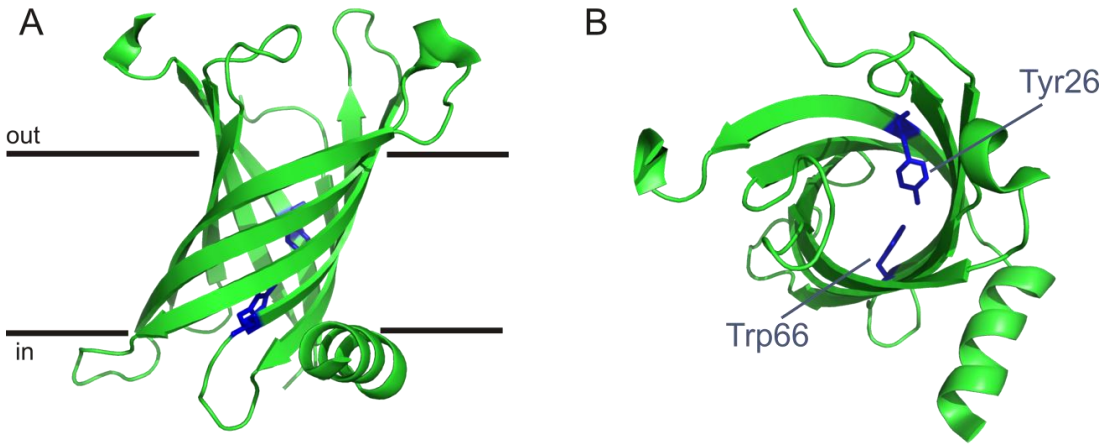


Figure 4.1.4 - Crystal Structure of PagP (PDB file: 1THQ) in lauryldimethylamine-oxide (Ahn *et al.* 2004) A), and B) rotated 90º out of the plane of the paper. The residues contributing to the Cotton Effect in the far-UV CD spectrum, Tyr26 and Trp66, are highlighted.

The structure of PagP in the detergent dodecylphosphocholine and lauryldimethylamine-oxide (Figure 4.1.5) has been determined by NMR (Hwang *et al.* 2002) and X-ray crystallography (Ahn *et al.* 2004), respectively. In both cases, PagP is tilted 25º compared to the membrane plane (Hwang *et al.* 2002; Ahn *et al.* 2004) (Figure 4.1.4) allowing the clustering of the aromatic residues at the membrane interface regions (Schulz 2002). The folding of PagP into liposomes has also been studied extensively by cold SDS-PAGE, Fourier transform infrared spectroscopy, fluorescence and circular dichroism (Huysmans *et al.* 2007). These studies have indicated that the N-terminal α-helix is not essential for PagP folding but acts as a post-assembly clamp, stabilising the folded PagP after insertion into the outer membrane (Huysmans *et al.* 2007).

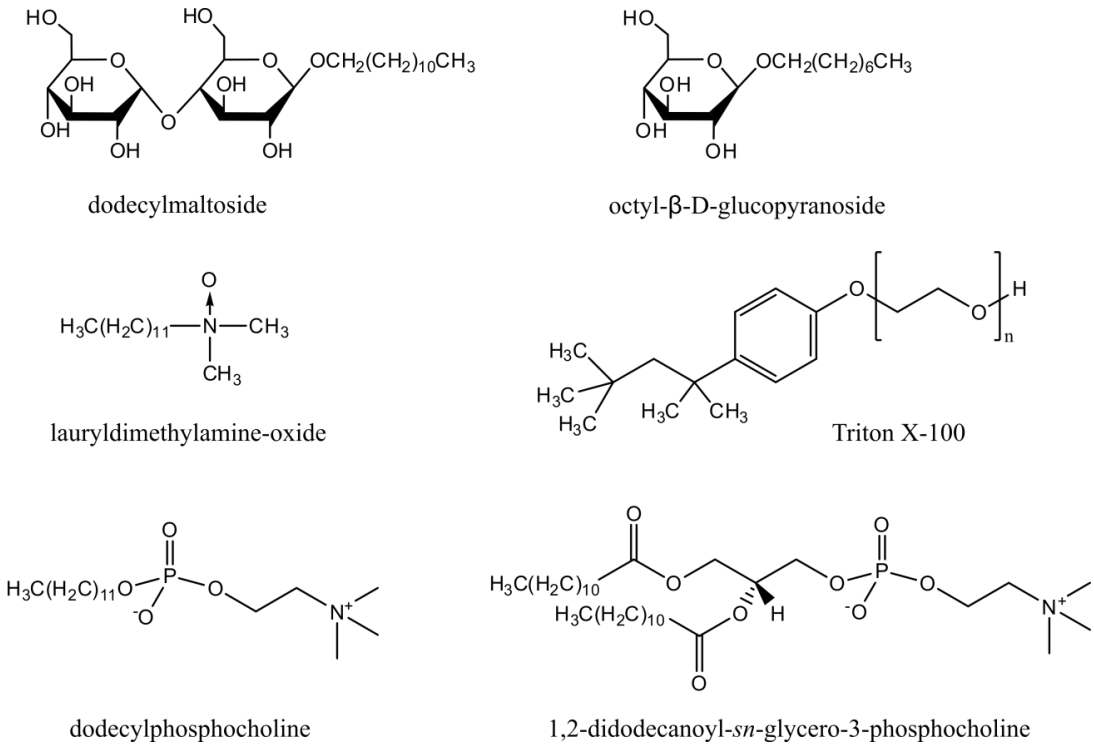


Figure 4.1.5 - Structure of the detergents dodecylmaltoside (DDM), octyl-β-D-glucopyranoside (OG), lauryldimethylamine-oxide (LDAO), Triton X-100, and lipids dodecylphosphocholine (DPC) and 1,2-didodecanoyl-*sn*-glycero-3-phosphocholine (DLPC).

In addition to performing functional studies on PagP (Bishop *et al.* 2000; Hwang *et al.* 2004; Huysmans *et al.* 2007), the correct natively folded PagP can also be characterised by a distinctive positive ellipticity band in the far-UV circular dichroism (CD) spectrum (Khan *et al.* 2007). This positive band at 232 nm, also known as a Cotton effect, arises from the Tyr26 and Trp66 residues (Figure 4.1.4) which pack closely together in the native PagP structure (Khan *et al.* 2007; Cuesta-Seijo *et al.* 2010). Although PagP folding into detergent micelles and liposomes has been studied extensively, the folding of PagP into amphipols has yet to be established.

4.1.6 Outer Membrane Protease, OmpT

The outer membrane protease, OmpT, is a ten-stranded, 34 kDa, β-barrel protein (Figure 4.1.5) that cleaves peptides between dibasic residues with arginine being preferential at the P1 and P1' positions (McCarter *et al.* 2004). OmpT can protect against a range of antimicrobial peptides and thus plays a role in bacterial pathogenicity (Hritonenko *et al.* 2007). The crystal structure of OmpT in octyl-β-D-glucopyranoside (Figure 4.1.5) shows that not all of OmpT is present in the transmembrane region, in fact long loops protruding out of the membrane are present which contain the enzyme's active site (Figure 4.1.6) (Vandeputte-Rutten *et al.* 2001). The folding of OmpT into detergent micelles has been studied by cold SDS-PAGE and far-UV CD (Kramer *et al.* 2000). However, in detergent alone OmpT was not found to be functional (Kramer *et al.* 2000). Indeed, only on addition of lipopolysaccharide (LPS), the major component of the outer membrane in Gram-negative bacteria, was OmpT functionally active as an outer membrane protease. The lack of activity in the absence of LPS is thought to be a safety mechanism to prevent the cleavage of cytosolic proteins during OmpT transport to the outer membrane (Kramer *et al.* 2002).

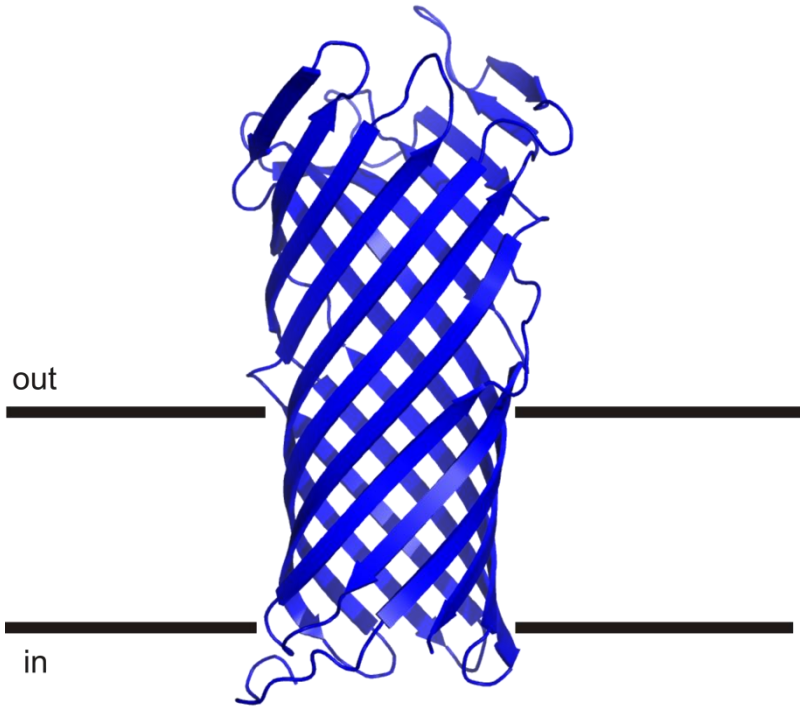


Figure 4.1.6 - Crystal Structure of OmpT (PDB file: 1I78) in octyl-β-D-glucopyranoside (Vandeputte-Rutten *et al.* 2001).

Although, a structurally similar outer membrane protease, OmpA, has been folded into the amphipol, A8-35 (Pocanschi *et al.* 2006), the folding of OmpT into amphipol has not yet been characterised.

In this thesis, ESI-MS has been applied to the study of the β-barrel outer membrane proteins PagP and OmpT in complex with amphipols. Both PagP and OmpT were successfully folded into the amphipol, A8-35, and their stability and functional activity confirmed by cold SDS-PAGE, size exclusion chromatography, far-UV CD and fluorescence. The membrane protein-amphipol complexes, once formed, were subjected to ESI-IMS-MS and the collision cross-sectional areas of the proteins determined. Based on these results, the future application of ESI-IMS-MS to study other membrane protein complexes will be evaluated.

Membrane Proteins

4.2. Results and Discussion

The compatibility of amphipol-membrane protein complexes with ESI-MS was investigated through the study of two outer-membrane proteins OmpT and PagP. Although OmpT and PagP are relatively simple β-sheet bacterial outer membrane proteins, their interaction with amphipols has not been reported previously.

4.2.1 Folding OmpT into A8-35

The OmpT plasmid was kindly provided by Lindsay McMorran (University of Leeds). OmpT plasmid was expressed in BL21 (DE3) cells and OmpT isolated as inclusion bodies (Materials and Methods Section 4.4.3). Purification by size exclusion chromatography in 25 mM TrisHCl pH 8.0, 250 mM NaCl, 6 M guanidine hydrochloride showed a single peak eluting off the column corresponding to > 95 % pure monomeric OmpT (Figure 4.2.1A) which was additionally confirmed by SDS-PAGE (Figure 4.2.1B). OmpT (yield ~15 mg/L bacterial culture) was dialysed into H_2O and stored at -80 ºC as a white precipitate.

One method commonly used to fold membrane proteins into amphipols involves initially solubilising the membrane protein of interest in detergent (Zoonens *et al.* 2005) (Figure 4.1.3). OmpT has been folded routinely into a variety of different detergent micelles for biochemical analysis (White *et al.* 1995; Kramer *et al.* 2000; Vandeputte-Rutten *et al.* 2001). OmpT was therefore dissolved into Triton X-100, amphipol added in a 1:5 (*w/w*) ratio of membrane protein:A8-35, and the detergent then removed by the addition of Bio-Beads. Unfortunately, OmpT did not completely solubilise in Triton X-100 so less than 10 % protein was incorporated as a result into the amphipol (data not shown).

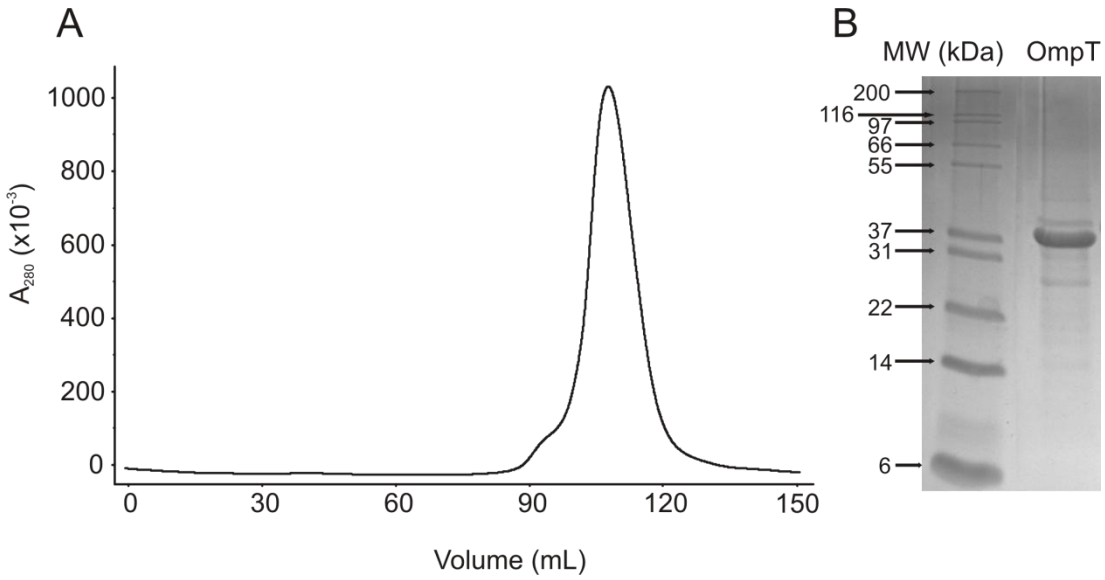


Figure 4.2.1 - A) Size exclusion chromatography trace (Superdex 75 column) of OmpT in 25 mM TrisHCl, 250 mM NaCl, 6 M guanidine hydrochloride pH 8.0; B) SDS-PAGE gel showing pure OmpT along with the molecular weight markers.

An alternative, more straightforward method was investigated to fold membrane proteins into amphipols. Amphipols have been used previously to fold the structurally similar β-barrel bacterial outer membrane protein OmpA from 8 M urea (Pocanschi *et al.* 2006). In this method, OmpA was solubilised initially in 8 M urea and then diluted rapidly 20-fold into urea-free buffer containing the amphipol, A8-35 (Figure 4.1.2A). In a similar way, to initiate OmpT folding, OmpT unfolded in 8 M urea, was mixed with the amphipol A8-35 in a 1:5 (*w/w*) ratio of OmpT:A8-35, and this mixture was dialysed immediately into 100 mM ammonium bicarbonate pH 8.0. Remarkably, using this method OmpT remained 100 % soluble in urea-free buffer, and no visible evidence of aggregation or precipitation was apparent.

Cold sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) can be used to distinguish between folded and fully denatured forms of membrane proteins (Schnaitm 1973). Thus, cold SDS-PAGE was used to determine the folding yield of OmpT into the amphipol, A8-35. On dilution of OmpT into 8 M urea, a single band was observed on the SDS-PAGE gel corresponding to unfolded OmpT (Figure 4.2.2). In comparison, the A8-35-solubilised OmpT migrated faster than its urea unfolded conformation indicating OmpT had folded completely into A8-35 (Figure 4.2.2). Additional heat denaturation of the OmpT:A8-35 complex again

unfolded OmpT causing it to migrate at a similar molecular weight to the urea unfolded OmpT sample. Altogether, the cold SDS-PAGE results indicate OmpT is more folded in amphipol compared with its denatured state; however, further analysis is required to determine whether the protein is correctly folded into its functionally active form in A8-35.

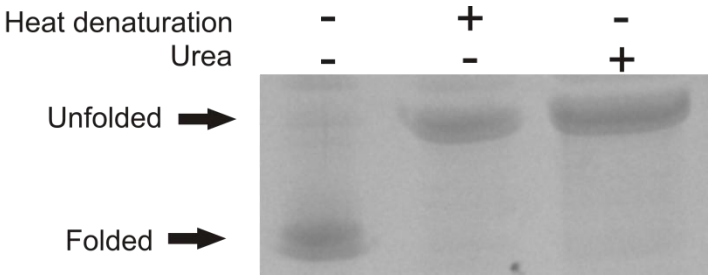


Figure 4.2.2 – SDS-PAGE gel of OmpT unfolded in urea and OmpT in A8-35 with and without heat denaturation prior to loading on the gel.

Tryptophan fluorescence can be used as an intrinsic probe of membrane protein folding (Moon *et al.* 2011). Membrane proteins unfolded in denaturant have a tryptophan emission spectrum whose λ_{max} is blue-shifted compared with their folded counterpart, for example, in detergent micelles or lipid. Since the amino acid sequence of OmpT contains eight tryptophan residues (Materials and Methods Section 4.4.2), the fluorescence emission spectrum of the tryptophans in OmpT can be used to determine the extent of OmpT folding into amphipol. The fluorescence emission spectrum of OmpT in A8-35 is shown in Figure 4.2.3. The λ_{max} for OmpT in A8-35 is 330 nm compared with 347 nm in the unfolded OmpT spectrum in urea. The average wavelength between 320 nm and 370 nm has been used to compare fluorescence emission spectra between OmpT folded species in liposomes (Lindsay McMorran, University of Leeds). This average wavelength value increases the sensitivity of the assay by avoiding variability in absolute intensities at a particular wavelength (due to pipetting errors associated with pippetting liposomes) through intensity normalisation (Material and Methods Section 4.4.6.5). To compare the folded OmpT:A8-35 spectrum to that of folded OmpT in liposomes, the average wavelength of OmpT in A8-35 and in 8 M urea was calculated. The average wavelength red-shifts from 345.2 nm to 341.6 nm on comparison of OmpT in A8-35

with OmpT in 8 M urea. This average wavelength of OmpT folded in A8-35 is comparable to that observed in 1,2-dilauroyl-*sn*-glycero-3-phosphocholine (DLPC) liposomes (Figure 4.1.5) at pH 8.0, in 3 M urea (342.6 nm) (Lindsay M. McMorran, University of Leeds, personal communication).

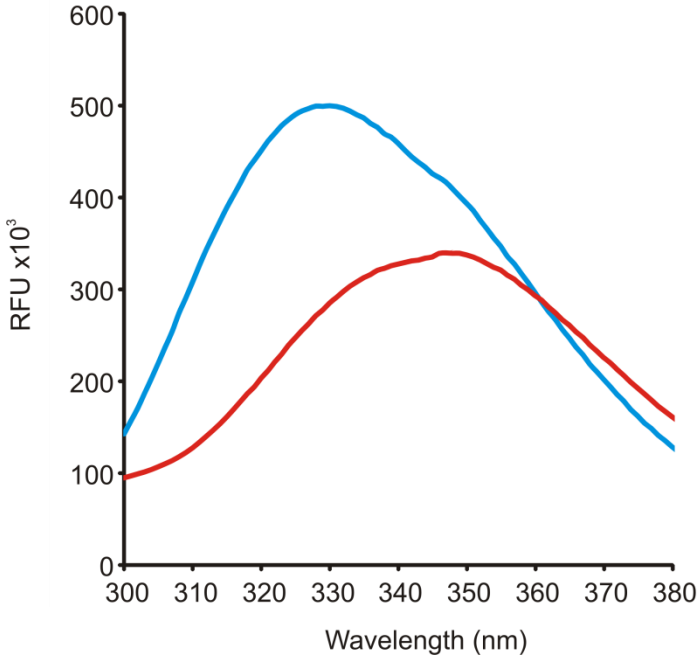


Figure 4.2.3 – Fluorescence emission spectrum of OmpT:A8-35 complex (blue) compared with spectrum in 8M urea (red).

Far-UV circular dichroism (CD) is a commonly used technique to determine protein secondary structure (Kelly *et al.* 2005). Far-UV CD was, therefore, used to determine whether β-sheet secondary structure is present in the OmpT:A8-35 complex. The far-UV CD spectrum of the OmpT:A8-35 complex showed a characteristic negative maximum at 218 nm (Figure 4.2.4A), which indicates that OmpT has β-sheet structure in complex with A8-35 and is therefore most likely to have folded into its native state in the amphipol.

The complementary technique of size exclusion chromatography (SEC) was used to study the stability and heterogeneity of the OmpT:A8-35 complex (Materials and Methods Section 4.4.6.6). A single, narrow peak was observed corresponding to the OmpT:A8-35 complex in solution (Figure 4.2.4B). The single OmpT:A8-35 species

elution volume corresponds to a higher molecular weight (~ 100 kDa) compared with that predicted based on the OmpT amino acid sequence (34 kDa). This may be due to the amphipol coating around the OmpT protein is increasing the Stokes radius of OmpT causing it to elute earlier off the gel filtration column. Nonetheless, the presence of a single OmpT:A8-35 peak shows only one folded OmpT:A8-35 species is present in solution.

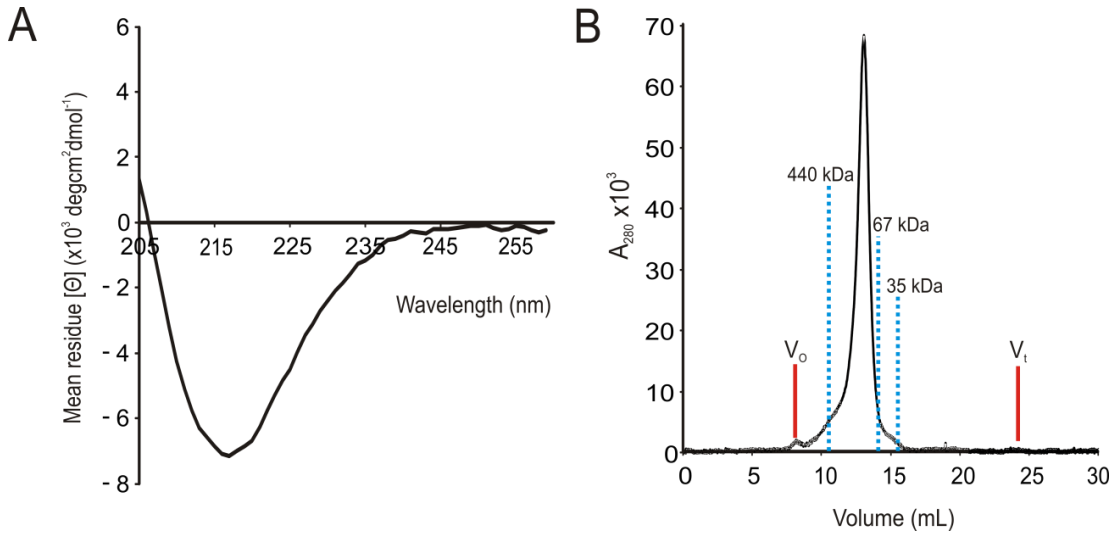


Figure 4.2.4 A) Far-UV CD spectrum of OmpT-A8-35 complex in 100 mM ammonium bicarbonate pH 8.0; B) SEC trace showing a single peak corresponding to the OmpT-A8-35 complex in 100 mM ammonium bicarbonate pH 8.0 is present in solution. The void volume (V_O), total column volume (V_t), and the elution times for the standard proteins ferritin (440 kDa), bovine serum albumin (67 kDa) and β-lactoglobulin (35 kDa) are shown.

The SDS-PAGE, SEC, CD and fluorescence emission spectra all indicate that OmpT is folded into the amphipol, A8-35; however, none of these techniques alone proves OmpT is folded into its native, functionally active state. The functionality of refolded OmpT was examined, therefore, by use of a fluorescence based assay. Native OmpT readily cleaves an internally quenched fluorogenic peptide (Abz-Ala-Arg-Arg-Ala-Tyr-(NO_2)-NH_2) in the presence of lipopolysaccharide (LPS) causing an increase in fluorescence at 430 nm. In the absence of LPS, no cleavage is observed (Kramer *et al.* 2002; Brandenburg *et al.* 2005). LPS is thought to activate OmpT allosterically via a subtle conformational change in the active site, and thus is thought to be a

safety mechanism to prevent cleavage of cytosolic proteins during trafficking to the LPS containing outer membrane (Kramer *et al.* 2002). OmpT:A8-35 complex was incubated at various concentrations with and without the presence of LPS and the specific OmpT enzyme activity measured based on the observed increase in fluorescence (Figure 4.2.5) (Materials and Methods Section 4.4.6.3). On increasing the concentration of the OmpT:A8-35 complex, the rate of cleavage of the fluorogenic peptide increased suggesting that OmpT is functional in the OmpT:A8-35 complex. The specific activity for OmpT was found to be 0.70 µmoles/mg enzyme/min, comparable to that observed in DLPC liposomes, pH 8.0 containing 3 M urea (0.75 µmoles/mg enzyme/min) (Lindsay M. McMorran, University of Leeds, personal communication). A small amount of activity is also observed in the absence of LPS (0.13 µmoles/mg enzyme/min) (Figure 4.2.5B); however, this is 6-fold reduced compared with the activity observed when LPS is present in solution showing that LPS is still required for efficient OmpT emzyme activity. The low activity of OmpT observed in the absence of LPS could be due to the presence of the amphipol causing a slight conformational change in OmpT resulting in the active site becoming more accessible for peptide cleavage. Together the SDS-PAGE, fluorescence emission, CD, SEC and the fluorescence activity data show OmpT readily folds into A8-35 to form a stable, native, functional OmpT:A8-35 complex.

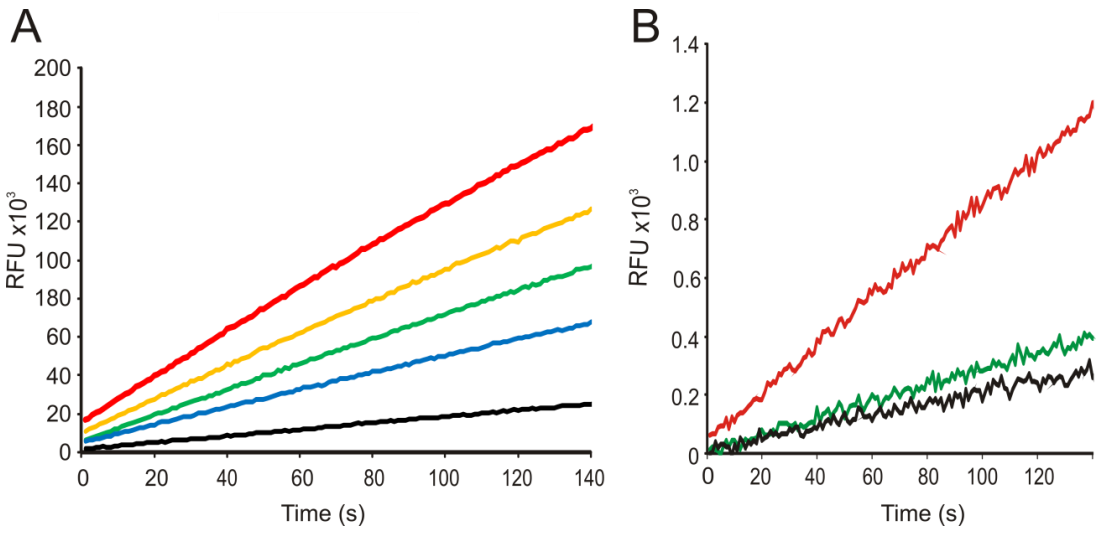


Figure 4.2.5 A) OmpT:A8-35 functional assay showing the fluorescence increase (relative fluorescence units) on cleavage of the peptide Abz-Ala-Arg-Arg-Ala-Tyr-(NO_2)-NH_2 on addition of 0.05 µM (black), 0.10 µM (blue), 0.15 µM (green), 0.20 µM (yellow), 0.30 µM (red) OmpT-A8-35 complex in the presence of LPS; B) OmpT:A8-35 functional assay without LPS present at an OmpT-A8-35 concentration of 0.05 µM (black), 0.15 µM (green) and 0.30 µM (red).

One significant advantage of amphipol-trapped membrane proteins over solubilising membrane proteins in detergent micelles is the increased stability that the amphipols provide. Bacteriorhodopsin/amphipol complexes have been shown to be stable for up to 6 months at 4 ºC (Gohon *et al.* 2008); hence, the stability of OmpT in A8-35 was investigated. After two months of storage at 4 ºC, the OmpT:A8-35 complex was analysed by cold SDS-PAGE, CD, SEC and its activity further measured by the fluorescence based assay described above. Strikingly, no differences in the SDS-PAGE, SEC and CD spectra were observed on comparison of freshly prepared OmpT:A8-35 samples with OmpT:A8-35 samples stored for 2 months in 100 mM ammonium bicarbonate buffer (Figure 4.2.6). Only a slight decrease in activity (less than 3-fold) was observed for the old sample compared with the newly folded OmpT:A8-35 complex.

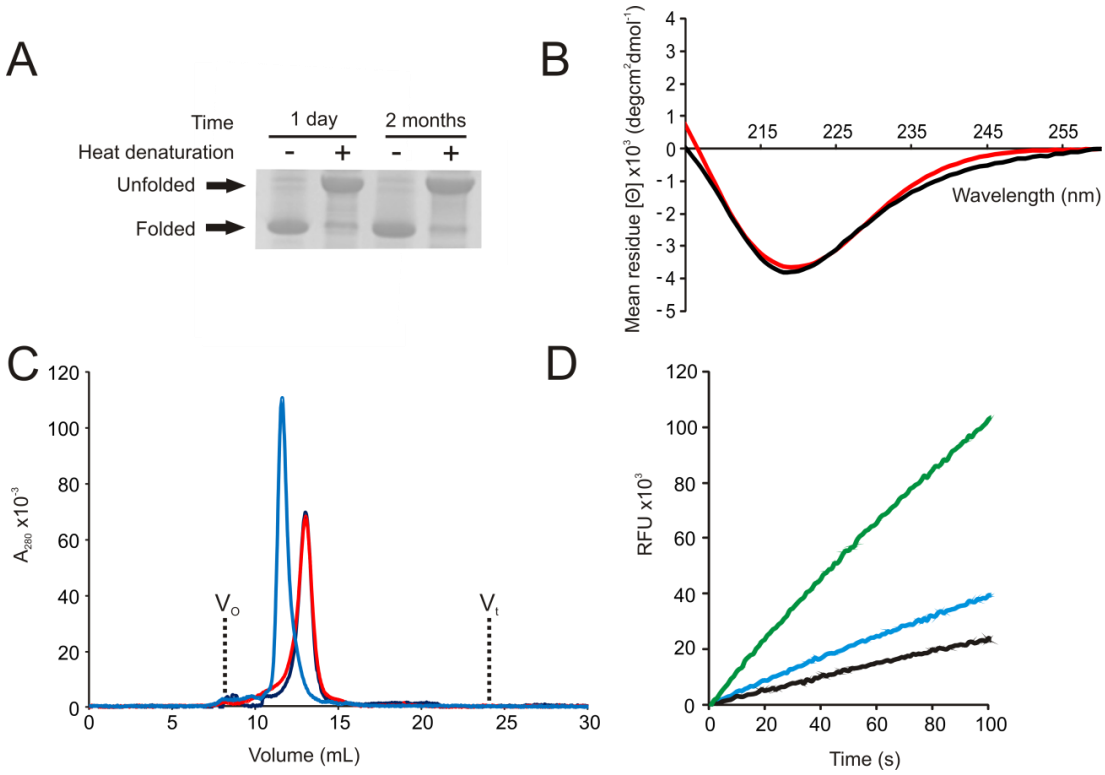


Figure 4.2.6 - OmpT:A8-35 complex stability upon storage at 4 °C for 2 months. A) Cold SDS-PAGE analysis showing folded OmpT remains folded in A8-35; B) Far-UV CD spectra showing OmpT:A8-35 before (red) and after (black) incubation at 4 °C for 2 months; C) SEC of OmpT in 8M urea (blue), OmpT:A8-35 (red) and OmpT:A8-35 after 2 months (black). The void volume (V_O) and total column volume (V_t) are shown; D) Activity assay for OmpT showing the fluorescence increase (relative fluoresecence units) of peptide on addition of 0.05 µM (black) and 0.15 µM (green) of freshly prepared OmpT:A8-35 compared with 0.15 µM of OmpT:A8-35 complex after 2 months of incubation at 4 °C (blue).

OmpT has been shown here for the first time to fold completely into the amphipol, A8-35 resulting in a stable OmpT:A8-35 complex. OmpT:A8-35 complex was further characterised by ESI-MS. IMS-MS was used to provide information on the structure of OmpT in the OmpT:A8-35 complex in addition to its accurate molecular weight. Before developing a method in which to study the OmpT:A8-35 complex by ESI-IMS-MS analysis, the mass spectrum of the amphipol, A8-35 alone was investigated.

4.2.2 ESI-IMS-MS Analysis of the Amphipol, A8-35

Due to their heterogeneous nature, the study of amphipols by mass spectrometry is challenging. The ESI mass spectrum of the amphipol, A8-35, is shown in Figure 4.2.6. Multiple peaks are observed that arise from the heterogeneous nature of the amphipol making the molecular weight of individual amphipol species alone very difficult to assign. Fragments of the amphipol with various combinations of the x, y and z species (highlighted in the amphipol structure) of the amphipol in Figure 4.2.7 are observed, indicating the correct amphipol of interest is present in solution.

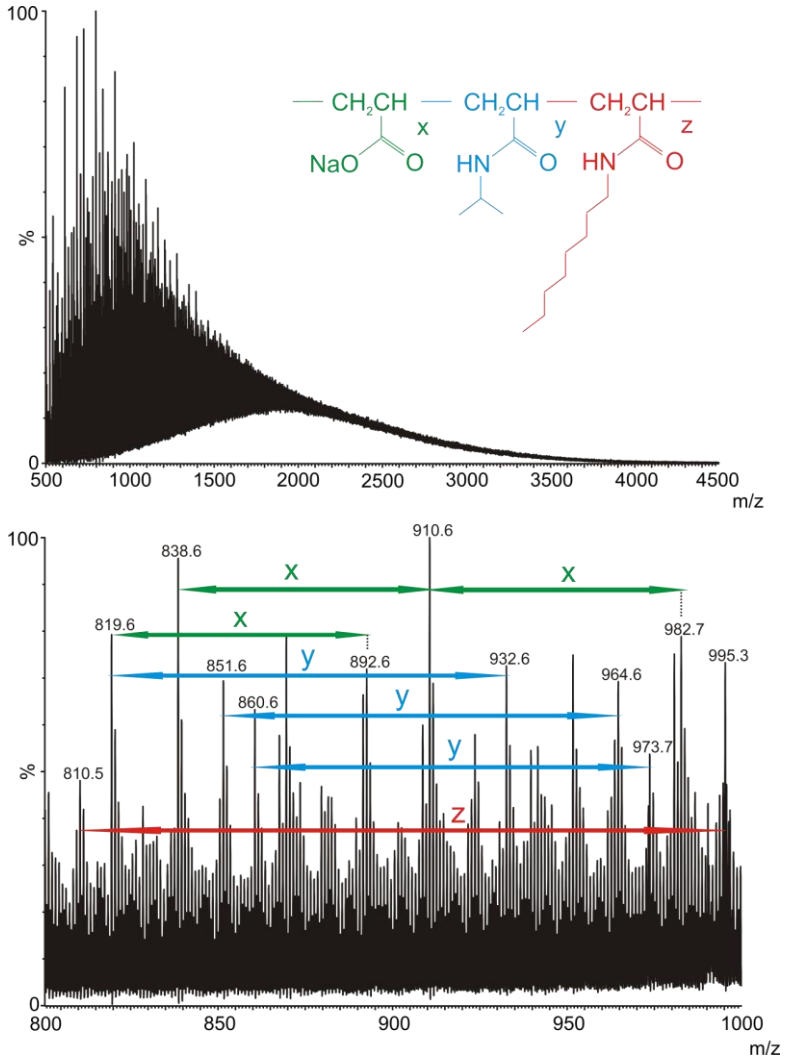


Figure 4.2.7 – ESI-MS analysis of amphipol, A8-35 alone over the entire *m/z* range (top) along with a zoomed in *m/z* spectrum between 800 and 1000 *m/z* where various A8-35 related peaks are highlighted (bottom). x, y and z correspond to the repeated regions of the amphipathic polymer, highlighted in the structure.

Over recent years, ion mobility spectrometry-mass spectrometry (IMS-MS) has been utilised increasingly to separate polymeric mixtures, allowing complex spectra to be assigned more accurately (Aminlashgari *et al.* 2012). IMS-MS was, therefore, used to separate the multiple charge states of A8-35 to determine whether a more accurate molecular weight could be assigned. The ESI-IMS-MS drifscope plot shows multiply charged ions (1+, 2+, 3+, 4+, 5+) present in the mass spectrum highlighting the power of ion mobility in separating out this complex amphipol mixture in the gas phase. The higher charge state ions (4+ and 5+) have a higher *m/z* (between 1,500 and 2,500) which correlate with the expected molecular weight for the amphipol (9-10 kDa). Unfortunately the mass spectrum does not reflect accurately the molecular

weight range of the amphipol due to smaller species ionising preferentially during the ESI process compared with the higher *m/z* species present.

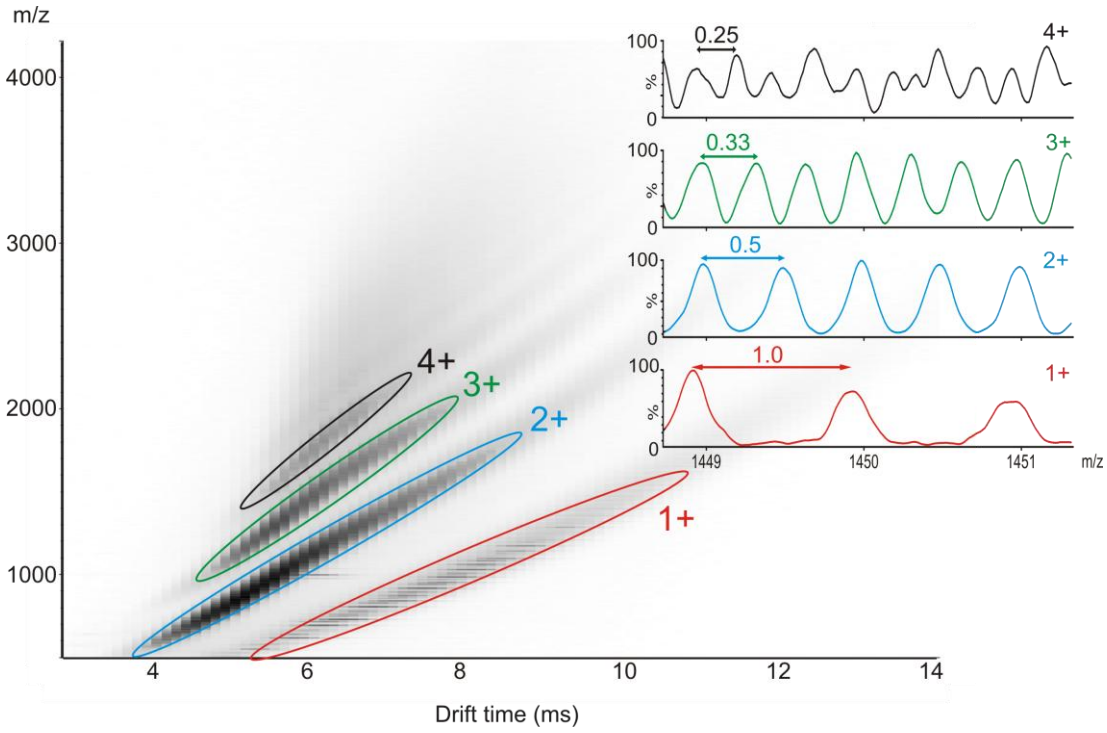


Figure 4.2.8 – ESI-IMS-MS drifscope plot of the amphipol, A8-35 alone highlighting the 1+. 2+, 3+ and 4+ charge states. The insert shows the isotope distribution in the mass spectrum (*m/z* range 1449-1451) for the extracted peaks highlighted from the driftscope plot (1+, 2+, 3+ and 4+ charge state ions).

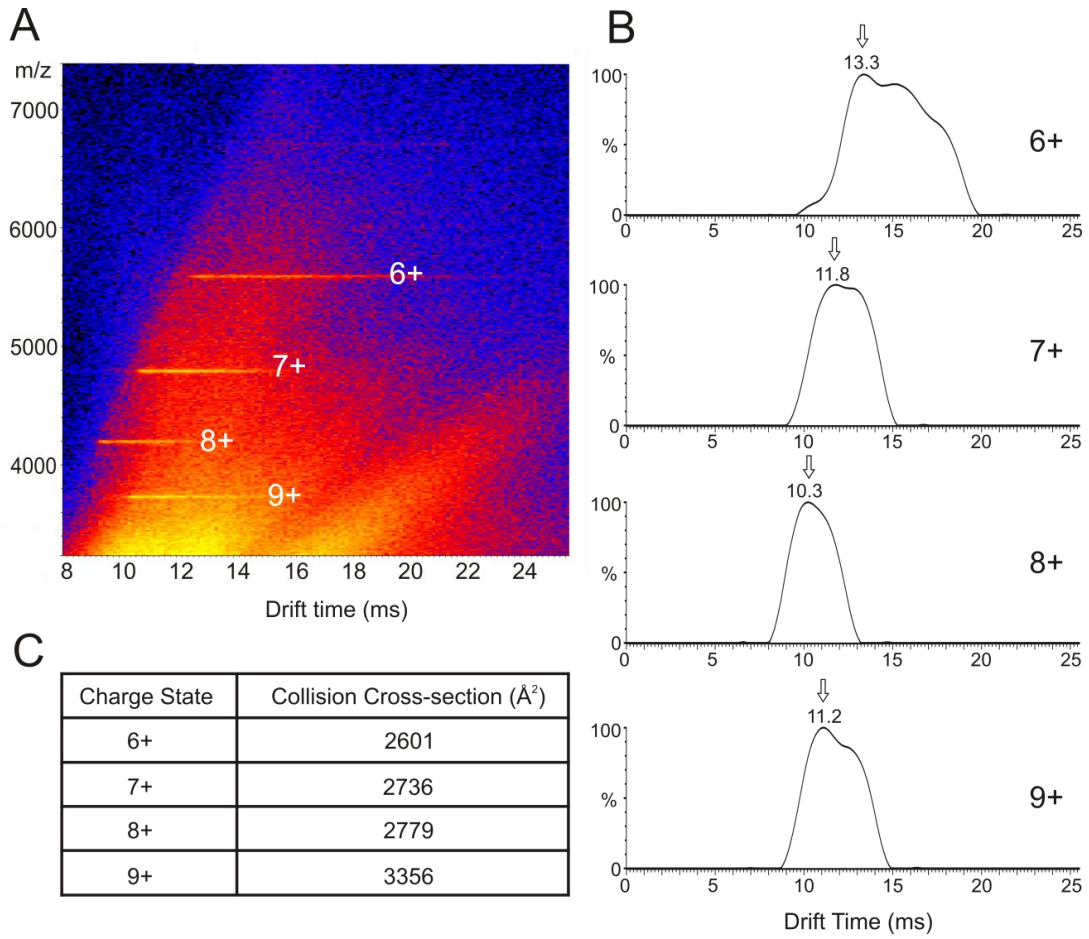
4.2.3 ESI-IMS-MS Analysis of OmpT:A8-35 Complex

To detect membrane proteins by ESI-MS, the membrane protein of interest must be released from the detergent micelle or lipid in the gas phase (Barrera *et al.* 2011). At low energy ionisation conditions, the OmpT in complex with amphipol was undetectable by ESI-MS (data not shown). When the energy in the trap and transfer T-wave devices of the instrument were increased from 6 V and 4 V to 150 V and 100 V, respectively (Materials and Methods Section 4.4.6.7), the membrane proteins could be released from the amphipol and ions corresponding to the multiply charged membrane proteins were detected (Figure 4.2.9A). The mass spectrum of OmpT shows a narrow charge state distribution corresponding to the 6+, 7+, 8+ and more

expanded 9+ ions with an experimentally determined MW of 33,462 ± 5 Da, within 0.01 % of calculated based on the amino acid sequence (33,460 Da).

IMS-MS was used additionally to determine if the OmpT is preserved in a native-like structure on release from amphipol in the gas phase. By calibrating the arrival time distributions of protein ions of known structure, the collision cross-sectional areas (CCS) of unknown proteins can be determined. An ideal ion mobility cross-section calibration involves using native protein calibrants of a similar molecular weight and size to the proteins of interest (Bush *et al.* 2010). The high energy conditions used in membrane protein analysis make the analysis of native calibrants impossible under the same ESI-IMS-MS conditions. In the future, it may be possible to calibrate membrane protein cross sections using previously analysed membrane proteins of known CCS areas in the gas phase. In the meantime, following previous analysis of membrane proteins' collision cross-sectional area measurements (Wang *et al.* 2010), a denatured calibration (using cytochrome c, myoglobin and ubiquitin) was used for membrane protein analysis (Amyloid Materials and Methods Section 3.4.10.2) as this is thought to equally reflect the state of the CCS areas of membrane proteins in the gas phase as the native calibration.

The experimentally determined CCSs for the 6+, 7+ and 8+ charge state ions of OmpT are 2601, 2736, 2779 $Å^2$, respectively (Figure 4.2.9C). The lowest charge state ion (6+) is consistent (within a 5 % error) with the theoretically determined cross-sectional area predicted from the crystal structures (2718 $Å^2$) using the projected superposition approximation (PSA) method (Bleiholder *et al.* 2011). Small shoulder peaks are present at increasing arrival time distributions for all of the OmpT charge states (Figure 4.2.9B). These are likely due to small amounts of gas-phase unfolding of the OmpT due to the high energies needed to eject OmpT from the OmpT:A8-35 complex. An additional increase in collision cross-section is observed with the 9+ charge state compared with the 8+ charge state ion (Figure 4.2.9C). This is thought to be due to Coulombic repulsion that is observed typically with highly charged ions. Together, the IMS-MS data shows the ability of amphipols to preserve native-like membrane protein structure on transition into the gas phase.



C

Charge State	Collision Cross-section (Å^2)
6+	2601
7+	2736
8+	2779
9+	3356

Figure 4.2.9 A) ESI-IMS-MS driftscope plot of OmpT:A8-35 showing the 6+, 7+, 8+ and 9+ charge states; B) Arrival time distributions of the charge states corresponding to OmpT. The arrows indicate the drift times used in the collision cross-section calculations; C) Table of the collision cross-sectional area measurements for each OmpT charge state using the ion mobility denatured calibration.

The combination of the biophysical assays on OmpT and mass spectral analysis show the amphipol, A8-35 is a useful tool for studying membrane proteins. The application of amphipols to study other membrane proteins of interest was investigated further to determine whether amphipols can be used as a generic tool for membrane protein analysis. Accordingly, another smaller bacterial outer membrane β-barrel protein, PagP, was chosen for analysis.

4.2.4 Folding PagP into A8-35

The smaller 20.2 kDa protein, PagP (Figure 4.1.4) transfers a palmitate chain from phospholipids to the lipid A moiety of LPS in the outer leaflet of the membrane, reinforcing the hydrocarbon core of the outer leaflet and protecting it from host

immune defences (Bishop 2005). The PagP plasmid was kindly provided by Lindsay McMorran (University of Leeds). PagP plasmid was expressed in BL21 (DE3) cells and PagP isolated as inclusion bodies (Materials and Methods Section 4.4.3). On PagP purification by metal affinity chromatography, a single peak eluted corresponding to PagP on increasing the concentration of imidazole to 250 mM in 25 mM TrisHCl pH 8.0, 250 mM NaCl, 6 M guanidine hydrochloride (Figure 4.2.10A). PagP was found to be pure as analysed by SDS-PAGE (Figure 4.2.10B) and subsequently dialysed into H_2O for storage as a white precipitate at -80 ºC (yield ~10 mg/L bacterial culture).

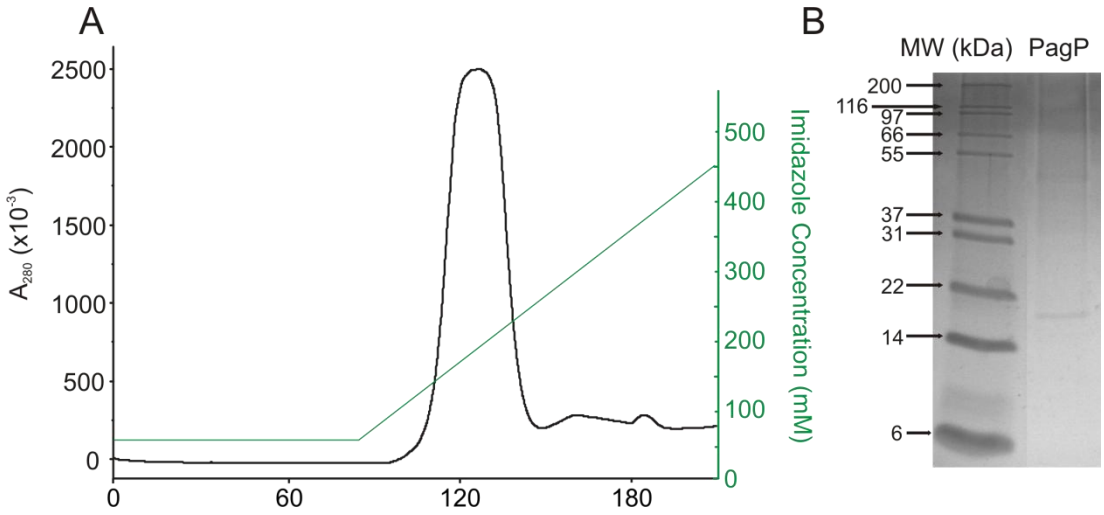


Figure 4.2.10 - A) Nickel affinity chromatography trace (HisTrap 5 mL column) of PagP in 25 mM TrisHCl, 250 mM NaCl, 6 M guanidine hydrochloride pH 8.0 on increasing imidazole concentration; B) SDS-PAGE gel showing pure PagP along with the molecular weight markers.

PagP was folded into A8-35 in the same way as OmpT. Briefly, A8-35 was added to PagP unfolded in urea and PagP folding initiated by dialysis in urea-free 100 mM ammonium bicarbonate buffer at pH 8.0. Cold SDS-PAGE analysis was used to determine the extent of folding of PagP into A8-35 (Figure 4.2.11). Despite much effort, it was not possible to fold PagP completely into A8-35, with two bands visible on the cold SDS-PAGE gel corresponding to a mixture of folded and unfolded PagP species in amphipol.

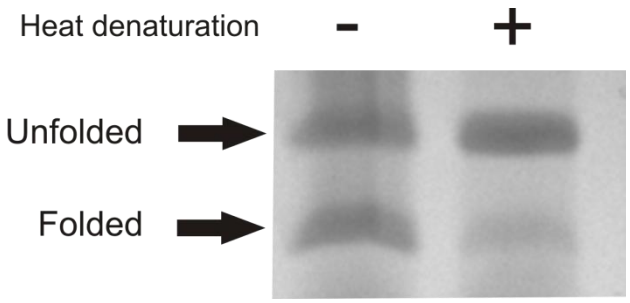


Figure 4.2.11 – SDS-PAGE gel of PagP in A8-35 (1:5 *w/w*) with and without heat denaturation prior to loading on the gel.

The ratio of amphipol to membrane protein can be optimised to increase the folding yield of membrane proteins in amphipols. For example, only partial refolding of bacteriorhodopsin (BR) into the amphipol, A8-35, was observed using a BR:A8-35 ratio of 1:2 (*w/w*); however complete refolding was observed by size exclusion chromatography on increasing the BR:A8-35 ratio to 1:5 (*w/w*) (Pocanschi *et al.* 2006). The *w/w* ratio of A8-35 was altered and its effect on PagP folding yield investigated (Figure 4.2.12). Increasing the PagP:A8-35 (*w/w*) ratio, increased the folding yield of PagP up to a maximum folding yield of ~50 % at a 1:5 (*w/w*) ratio of PagP to amphipol. A 1:5 (*w/w*) ratio was used for the data presented throughout the rest of this thesis. The rate of dialysis can also affect protein folding. Various step dialysis methods (for example from 8 M urea, to 7 M urea, 6 M urea and 5 M urea before the complete removal of urea) were carried out to slowly remove the urea from the amphipol-containing PagP sample, however no increase in PagP folding yield was observed (data not shown).

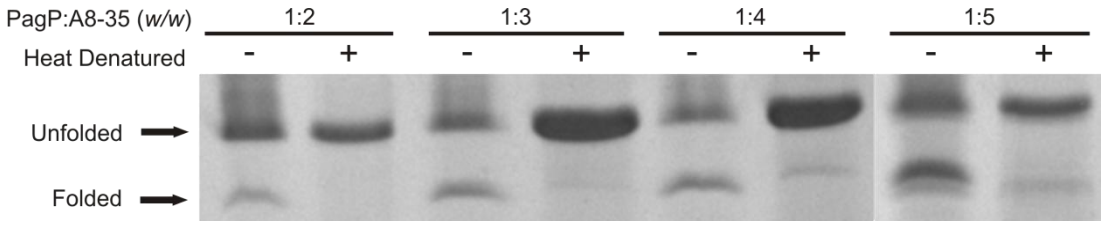


Figure 4.2.12 – SDS PAGE of PagP:A8-35 complex with and without heat denaturation prior to loading the samples on the gel.

Size exclusion chromatography was also used to determine the extent of PagP folding into A8-35 (Figure 4.2.13) (Materials and Methods Section 4.4.6.6). On solubilising PagP in 100 mM ammonium bicarbonate buffer containing 8 M urea, a single peak was observed after the void volume of the column corresponding to unfolded PagP. In comparison, the SEC data for the PagP:A8-35 complex showed a broad peak eluting off the column. This broad peak corresponds to the folded and unfolded PagP:A8-35 species that were also observed on the cold SDS-PAGE gel (Figures 4.2.11 and 4.2.12). The lower folding yield of PagP compared with OmpT is thought to be due to the highly aggregation-prone nature of PagP (Huysmans *et al.* 2007). Problems have been observed previously on folding PagP into liposomes with only complete folding into 1,2-dilauroyl-sn-glycero-3-phosphocholine ($diC_{12:0}PC$) observed in the presence of 7 M urea (Huysmans *et al.* 2007).

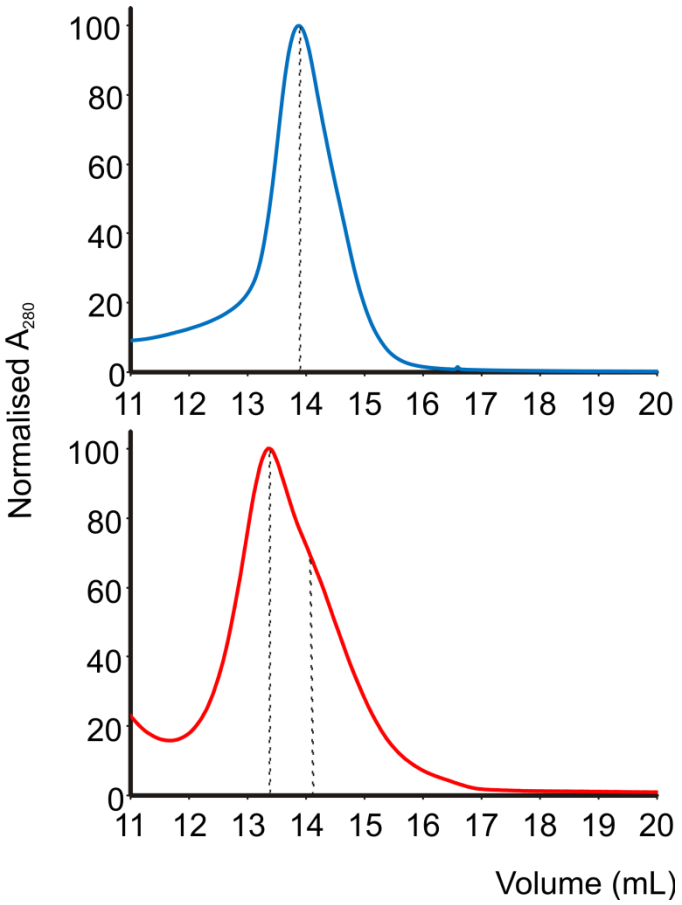


Figure 4.2.13 – Size exclusion chromatography of PagP unfolded in 8M urea (blue, top) and the PagP-A8-35 complex in 100 mM ammonium bicarbonate pH 8.0 (red, bottom).

Although, PagP was not completely folded into A8-35, far-UV CD was used to determine whether the folded PagP present contained β-sheet secondary structure. A characteristic negative maximum at 218 nm was observed, suggesting PagP had folded correctly into A8-35 (Figure 4.2.14). The positive molar ellipticity at 232 nm (shown in Figure 4.2.12A) commonly observed in the native PagP far-UV CD spectrum is absent (Khan *et al.* 2007). The band at 232 nm is due to a Cotton effect between residues Tyr26 and Trp66 which pack closely together in the native PagP structure (Figures 4.1.4, 4.2.14A). Slight structural perturbations in the PagP structure in the amphipol could be resulting in the loss of this Cotton band. Local structural modifications through mutation of residues in the active site of PagP has also shown the loss of the PagP Cotton band in the far-UV CD spectrum (Khan *et al.* 2010). The absence of the Cotton band could also be due to the mixture of folded and unfolded PagP species in A8-35 masking the natively folded PagP signal.

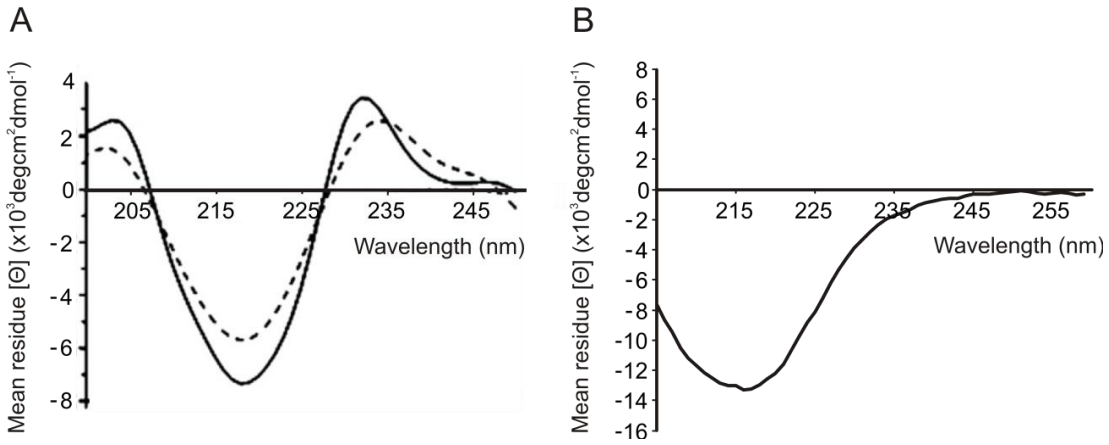


Figure 4.2.14 - Far-UV CD spectrum of A) PagP in the detergent cyclofos-7 (- - -) and 1,2-dilauroyl-sn-glycero-3-phosphocholine liposomes (-) showing a negative maximum at 218 nm and a Cotton Band at 232 nm; B) PagP:A8-35 complex showing a characteristic negative maximum at 218 nm corresponding to β-sheet secondary structure. Figure 4.2.14A was adapted from Huysmans *et. al.* (Huysmans *et al.* 2007).

Additional tryptophan fluorescence emission experiments were carried out to further investigate the folding of PagP into the amphipol, A8-35. PagP contains 12 tryptophan residues, two in the N-terminal α-helix, eight in the transmembrane region of the β-barrel, and two in the extracellular loops (Figure 4.4.2). The tryptophans and tyrosines in unfolded PagP in 8 M urea and the PagP:A8-35 complex in 100 mM ammonium bicarbonate pH 8.0 were excited at 280 nm and

their tryptophan emission spectra recorded between 300 and 380 nm (Figure 4.2.15). The data show that the λ_{max} shifts from 348 nm to 330 nm on folding the PagP from 8M urea into A8-35 (Figure 4.2.15) showing the 12 tryptophans in PagP are more buried from solvent in the PagP:A8-35 compared with the urea denatured state. The data are comparable with the λ_{max} for the fluorescence emission spectra of PagP refolded into cyclofos-7 micelles or DLPC (Figure 4.1.5) liposomes (335 nm) (Huysmans *et al.* 2007), suggesting PagP has correctly folded into the amphipol.

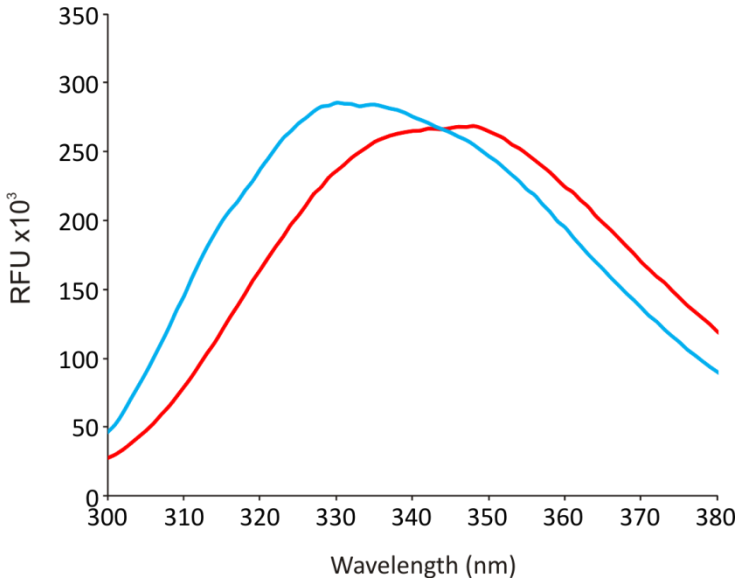


Figure 4.2.15 – Fluorescence emission spectrum of PagP:A8-35 complex (blue) compared with the urea unfolded PagP spectrum (red).

Despite the absence of the Cotton band in the CD spectrum, a functional assay was used additionally to determine if PagP is folded to a native functionally active state in A8-35. An activity assay in which native PagP hydrolyses *p*-nitrophenylpalmitate (*p*-NPP) to *p*-nitrophenol (*p*-NP) was monitored by the change in UV absorbance at 410 nm (Huysmans *et al.* 2007). As the concentration of PagP was increased, in the rate of *p*-NPP hydrolysis also increased and increasing amounts of the yellow *p*-NP were detected (Figure 4.2.16). Although not a strictly quantitative method, due to the poor solubility of the *p*-NPP substrate in the buffer used (Materials and Methods Section 4.4.6.4), the specific enzyme activity of the PagP:A8-35 complex was calculated to be 0.019 ± 0.006 nmol/min/µM, a value similar to that observed for PagP in detergent (0.068 ± 0.011 nmol/min/µM) (Huysmans *et al.* 2007) considering

the different buffers used and problems associated with *p*-NPP solubilisation in these buffers.

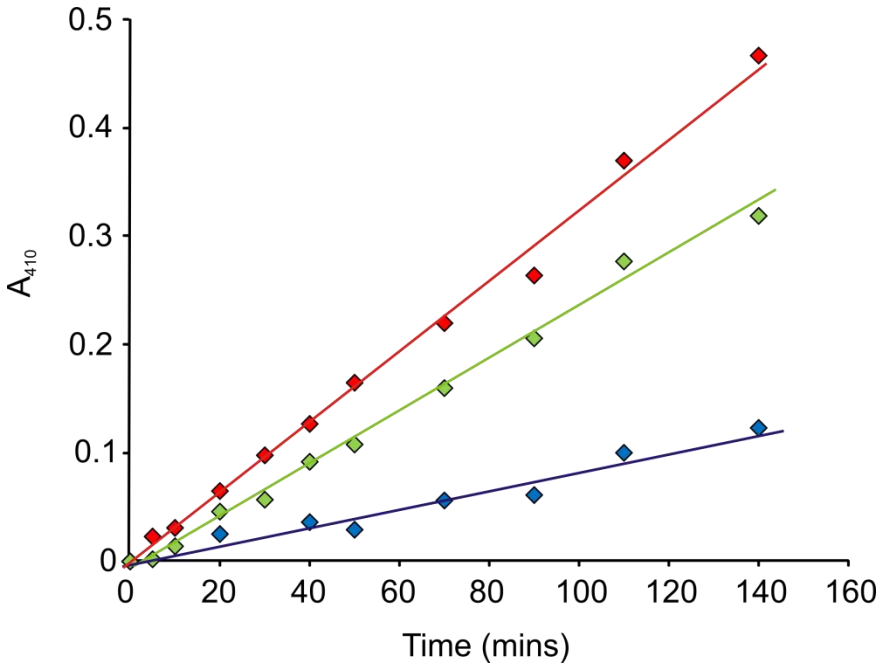


Figure 4.2.16 – PagP:A8-35 functional assay showing the absorbance increase at 410 nm on hydrolysis of *p*-NPP at a PagP:A8-35 concentration of 20 µM (blue), 30 µM (green), 40 µM (red).

Analysis of the OmpT:A8-35 complex showed that once formed the OmpT:A8-35 complex was stable for up to 2 months at 4 ºC (Figure 4.2.6). The stability of the PagP:A8-35 complex formed was investigated, therefore, by cold SDS-PAGE, far-UV CD and tryptophan fluorescence emission spectroscopy (Figure 4.2.17). Although PagP was not initially fully folded in the amphipol, A8-35, after two months of storage at 4 ºC, the PagP:A8-35 complex remained 50 % folded as analysed by cold SDS-PAGE (Figure 4.2.17A). The far-UV CD and tryptophan fluorescence emission spectra were also comparable between freshly prepared samples and samples stored at 4 ºC for 2 months Figure 4.2.17B,C). These data suggest that once folded into amphipol, PagP remains folded and stable in 100 mM ammonium bicarbonate buffer at pH 8.0.

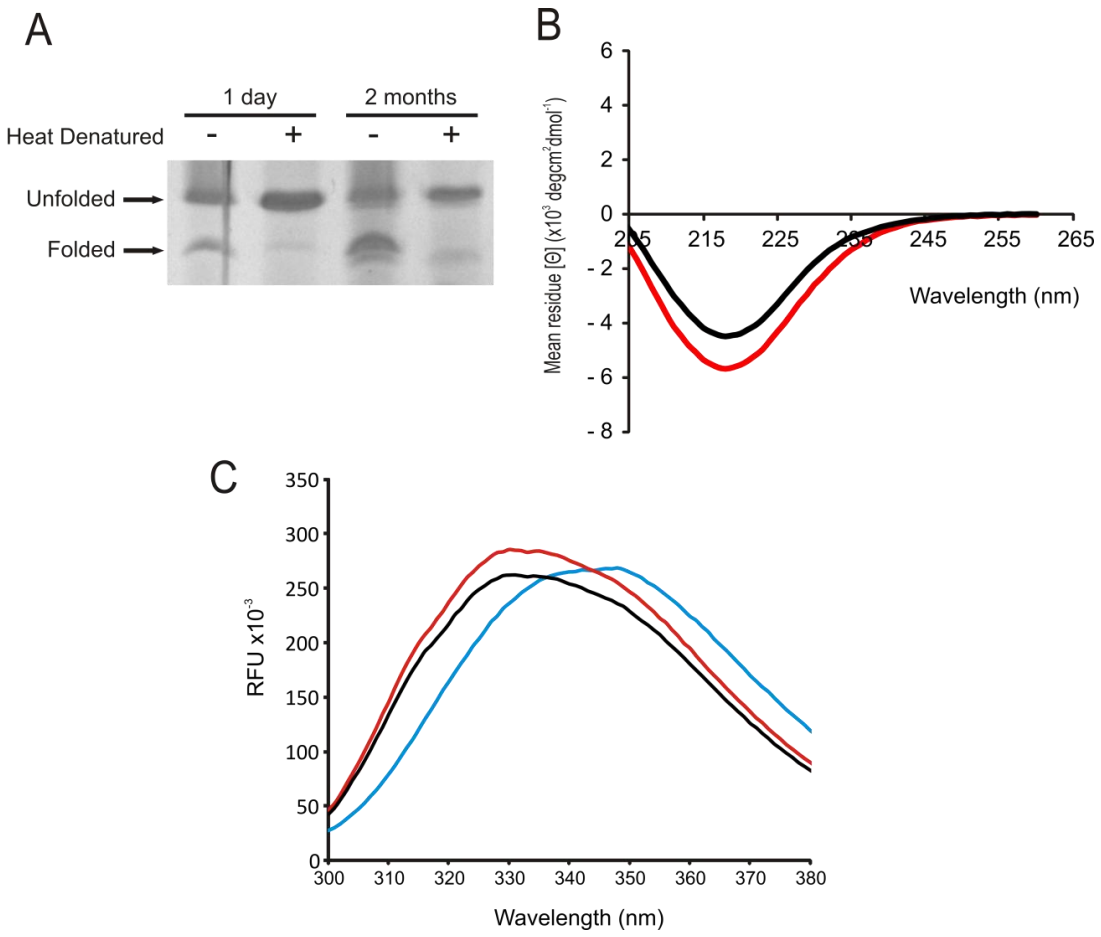


Figure 4.2.17 - PagP:A8-35 complex stability upon storage at 4 °C for 2 months. A) SDS-PAGE analysis showing PagP remains ~50 % folded in A8-35; B) Far-UV CD spectra showing PagP:A8-35 before (red) and after (black) incubation at 4 °C for 2 months; C) Fluorescence emission spectra of PagP in 8M urea (blue), PagP:A8-35 (red) and PagP:A8-35 after 2 months at 4 °C (black). The fluorescence intensity was reported in relative fluorescence units (RFU).

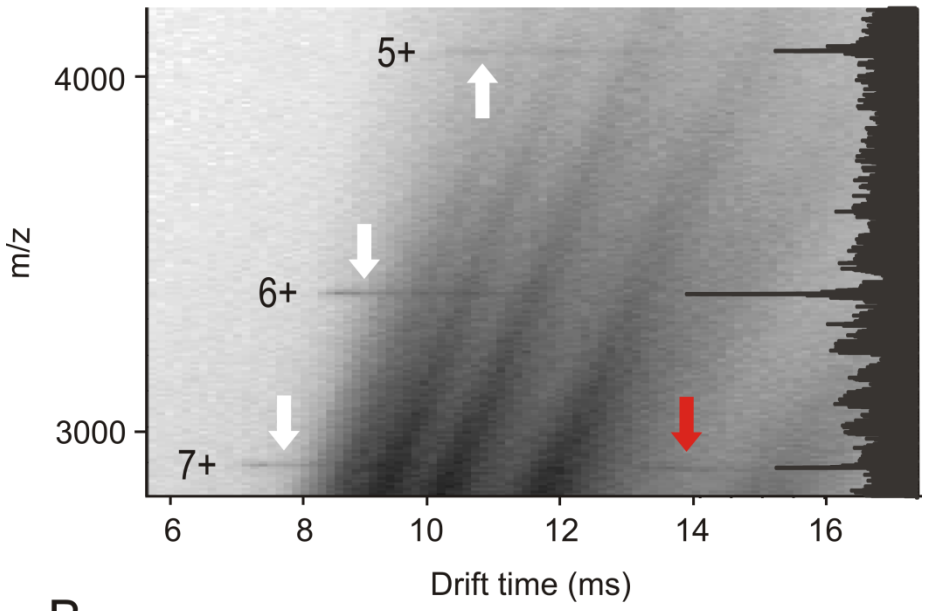
4.2.5 ESI-IMS-MS Analysis of PagP:A8-35 Complex

Together, the SDS-PAGE, far-UV CD, SEC and fluorescence emission spectra suggest that only ~ 50 % of PagP is folded into the amphipol, A8-35, in 100 mM ammonium bicarbonate pH 8.0. It is possible that ESI-IMS-MS can be used to analyse the conformation of the folded and unfolded PagP species in the gas phase, hence IMS-MS was carried out on the PagP:A8-35 complex mixture in 100 mM ammonium bicarbonate buffer. If the presence of amphipol can preserve membrane protein structure on transition into the gas phase (as shown with the OmpT:A8-35 complex mass spectral data (Figure 4.2.9)), at least two species should be observed

in the mass spectrum of PagP:A8-35 corresponding to a folded and unfolded conformation of PagP.

In comparison to ESI-MS experiments on the OmpT:A8-35 complex under low energy ionisation conditions (Materials and Methods Section 4.4.6.7), PagP in complex with A8-35 was undetectable by ESI-MS under low energy ionisation conditions (data not shown). When the trap and transfer T-wave devices were increased to 150 V and 100 V respectively, PagP was released from the amphipol and ions corresponding to the multiply charged membrane protein were detected (Figure 4.2.18). The ESI mass spectrum of PagP shows a narrow charge state distribution corresponding to the 5+, 6+, 7+ ions with an experimentally determined MW of 20,175 ± 1 Da, within 0.01 % of calculated based on the amino acid sequence (20,175 Da). Higher charge state ions (7+, 8+, 9+ and 10+) are also observed corresponding to a more unfolded PagP species, however, although present in the driftscope, these peaks are difficult to analyse due to their overlap with the high *m/z* amphipol peaks.

A



B

Charge State	Collision Cross-section (Å^2)
5+	1790
6+	1794
7+	1836
7+	3131

Figure 4.2.18 - A) ESI-IMS-MS driftscope plot of PagP:A8-35 complex. The charge states of the ions are highlighted. The white arrows represent the more compact PagP conformation and the red arrow the more expanded conformation. The mass spectrum of PagP is shown on the left-hand side; B) Table of the collision cross-sectional area measurements for each PagP charge state using the ion mobility denatured calibration. The collision cross-section for the more expanded PagP conformation is highlighted in red.

The experimentally determined CCS for the lowest charge state ions (5+) of PagP (1790 Å^2) are consistent with the values predicted from the crystal structure using the projected superposition approximation algorithm (Bleiholder *et al.* 2011) (1732 Å^2) within a 5 % error, highlighting the power of amphipols in preserving membrane protein structure on transition into the gas phase. IMS-MS also allowed an additional, more expanded, conformation of PagP to be separated from the amphipol, A8-35 in the gas phase (the 7+ charge state of which is highlighted in Figure 4.2.18). From the arrival time distribution, the CCS of this 7+ charge state ion (3131 Å^2) is

70 % larger than the more compact 7+ charge state ion (1836 $Å^2$). Since no additional expanded conformations of OmpT were observed in the IMS-MS driftscope plot for the OmpT:A8-35 complex analysed under identical mass spectrometry conditions (Figure 4.2.9), it is thought that the unfolded PagP species observed in the gas phase is also present in solution and is not an artefact of the ESI-IMS-MS process. In addition, since there are two conformations present for the 7+ charge state, the more expanded conformer is not likely to be due to Coulombic repulsion, in contrast to that observed for the 9+ charge state in the OmpT driftscope (Figure 4.2.9).

The observation of the unfolded PagP using the high energy ionisation conditions suggests the unfolded PagP protein is also surrounded by the amphipol and is being protected during the electrospray process into the gas-phase. The percentage of unfolded PagP remained constant during storage at 4 ºC for 2 months (Figure 4.2.17A) with no visible evidence of precipitation observed. This suggests that the amphipols can not only be used to study proteins in their native state but may also have implications in trapped structural intermediates in the membrane protein folding pathway.

The data presented show that PagP has been inserted into amphipol, A8-35. Solution biochemical data combined with ESI-IMS-MS indicate that PagP can fold into a stable PagP:A8-35 complex; however a 50 % folding yield is obtained. The data show the power of IMS-MS in separating folded and unfolded conformations of PagP in support of the in solution SDS-PAGE and SEC data indicating that amphipols can preserve membrane protein structure on transition into the gas phase.

4.3 Conclusions

Membrane protein analysis by biochemical techniques is challenging due to the difficulties associated with membrane protein purification and retaining native, functionally active complexes *in vitro*. Over the last sixteen years, amphipols have been increasingly used as an alternative to detergents to analyse membrane protein complexes in solution (Popot *et al.* 2011). In this thesis, two β-barrel outer membrane proteins, PagP and OmpT, were studied. Both membrane proteins were inserted for the first time into the amphipol, A8-35. Analysis by SDS-PAGE, SEC, CD and fluorescence emission spectroscopy revealed that OmpT had folded completely into the amphipol with a functional activity comparable to that observed in liposomes. PagP, however, did not completely fold into A8-35 with approximately 50 % remaining unfolded in solution. Further investigation is needed into alternative methods of folding membrane proteins into amphipols to optimise the folding yield of PagP into amphipol. Indeed, alternative amphipols, for example the non ionic amphipols (NAPols) (Bazzacco *et al.* 2012), have now been designed that may assist the folding of PagP into a more stable PagP:amphipol complex.

The analysis of membrane proteins by mass spectrometry has previously been established by solubilisation of membrane proteins in detergent micelles (Barrera *et al.* 2008; Barrera *et al.* 2009; Wang *et al.* 2010; Barrera *et al.* 2011; Zhou *et al.* 2011). This approach to membrane protein analysis has limitations due to problems associated prior to ESI-MS analysis with preserving protein structure, protein-protein and protein-ligand interactions in detergent solutions. Amphipols offer the ability to trap membrane proteins in detergent-free aqueous solutions, thus alleviating these limitations. Additionally, amphipols are particularly stable in aqueous solution containing low salt which are ideal conditions for mass spectral analysis. The ESI mass spectrum of the amphipol, A8-35 alone showed multiple peaks corresponding to the intact amphipol along with multiple fragment ions. The heterogeneity of this sample makes the exact molecular weight of amphipol impossible to determine accurately. Indeed MALDI-MS analysis of the same amphipol, A8-35 was not

detectable since it is thought the sample is too polydisperse for any individual species to be detected (Bechara *et al.* 2012).

Mass spectrometric analysis of amphipol-trapped membrane proteins has only been established relatively recently (Bechara *et al.* 2012). In this work, four different membrane proteins were studied, OmpA (an 8-stranded β-barrel outer membrane protein from *E.coli*), bacteriorhodopsin (a 7 α-helical retinal-containing proton pump from the archaebacterium *Halobacterium salinarum*), cytochrome b_6f (a superdimer from the green alga *Chlamydomonas reinhardtii* with 8 transmembrane subunits) and cyctochrome bc_1 (a multisubunit superdimer from beef heart mitochondria). MALDI-TOF MS analysis allowed the detection of all the membrane proteins studied, with the membrane proteins being released from the amphipols during the MALDI ionisation process. This mass spectrometry data allows the molecular weight, and subunit stoichiometries of membrane protein complexes to be determined. This MALDI-MS data alone does not provide any structural information on the membrane proteins in the gas phase and additional experiments are needed to determine if the membrane proteins are indeed functional in the MALDI matrix prior to mass spectrometry analysis. In this thesis an ESI-IMS-MS method was used to determine the molecular weight of the membrane protein of interest and also to investigate the structural properties of the membrane proteins in complex with amphipols.

Two different A8-35-trapped membrane proteins, OmpT and PagP were studied by ESI-IMS-MS. IMS-MS separated amphipol from the membrane proteins in the gas phase allowing ions to be detected corresponding to the membrane proteins alone. Additional collision cross-sectional area calculations showed that amphipols can preserve membrane protein structure on transition into the gas phase with experimental cross sections within 5 % of those obtained from theoretical values calculated based on the PDB structures. For example in the case with PagP, IMS-MS could separate multiple different proteins conformations in the complex with the amphipol, reflecting the in solution biochemical data.

The application of A8-35 to two different β-barrel membrane proteins suggests that this procedure could be applicable to other more complex membrane proteins. The

insertion of membrane proteins into amphipols is becoming increasingly popular for membrane protein structure analysis. The α-helical inner membrane protein, bacteriorhodopsin (Pocanschi *et al.* 2006; Gohon *et al.* 2008), β-barrel outer membrane proteins OmpX (Catoire *et al.* 2010), OmpA (Pocanschi *et al.* 2006), FomA (Pocanschi *et al.* 2006), and membrane protein complexes including ligand bound to G-protein coupled receptors (GPCR) (Catoire *et al.* 2010) and the NADH bound ubiquinone oxidoreductase (Flotenmeyer *et al.* 2007), are just some of the membrane proteins that have been inserted into amphipols over the last few years. Although in this thesis only small β-barrel outer membrane proteins were selected for MS analysis, the MS analysis of higher-order multi-subunit membrane protein complexes in amphipols is sure to become increasingly popular, since the higher the mass to charge ratio of the membrane protein complex studied, the easier their separation from amphipol during IMS-MS analysis.

Nanodiscs are an alternative way of solubilising membrane proteins in solution for study by mass spectrometry (Marty *et al.* 2012). Nanodiscs typically consist of a lipid bilayer surrounded by a stabilising scaffold protein to which a membrane protein can be inserted (Ritchie *et al.* 2009; Bayburt *et al.* 2010; Popot 2010). Although these membrane proteins inside the nanodiscs are thought to be in a more native-like environment, the preparation of nanodiscs containing membrane proteins is not straightforward. Several parameters need to be optimised, such as the lipid type, the lipid-protein ratio, in addition to the nanodisc forming conditions making it difficult to be a high throughput approach. Insertion of membrane proteins into amphipols, however, is relatively straightforward with only the protein:amphipol ratio needing optimisation. Although membrane proteins in nanodiscs are thought to have an increased stability compared with membrane proteins inserted into detergent or liposomes, the stability of membrane proteins in amphipols compared with nanodiscs is dramatically increased with the membrane protein:amphipol complexes. The PagP and OmpT that folded into the amphipol, A8-35 remained folded and functionally active during storage for two months at 4 ºC.

As amphipols become more commercially available, the study of membrane proteins in amphipols will become increasingly utilised within the membrane protein field. The additional stability that amphipols offer makes these membrane protein-

amphipol complexes highly favourable over methods that use detergent to solubilise the membrane proteins of interest. Since membrane proteins constitute approximately 30 % of the proteome and more than 50 % of all known drug targets, the study of membrane proteins by high throughput mass spectrometric techniques will become increasingly utilised amongst this challenging field of research.

Membrane Proteins

4.4. Materials and Methods

4.4.1 Reagents

All reagents were purchased as previously described in Materials and Methods Sections 2.4 and 3.4 unless otherwise stated below.

Protein Purification Reagents

Triton X-100 was purchased from Calbiochem, UK. Guanidine hydrochloride (GuHCl) was purchased from Melford Laboratories, Suffolk, UK.

Reagents for Enzyme Assays

Abz-Ala-Arg-Arg-Ala-Tyr-(NO_2)-NH_2 was purchased from Cambridge Peptides, UK. Lipopolysaccharide (LPS) was purchased from Calbiochem, UK, p-nitrophenyl palmitate (pNNP) from Sigma-Aldrich Corporation, Dorset, UK and isopropanol from Fisher Scientific, UK.

Amphipol

The amphipol, A8-35, was purchased from Affymetrix UK Ltd. (High Wycombe, UK).

Mass Spectrometry Reagents

Ammonium bicarbonate was purchased from Sigma-Aldrich Corporation (Poole, Dorset, UK).

4.4.2 Molecular Biology

Transformation of plasmid DNA

Plasmid DNA (kindly provided by Lindsay M. McMorran, University of Leeds) was transformed into BL21 (DE3) cells for protein expression according to the protocol mentioned previously (see Pilus Materials and Methods, Section 2.4.7).

OmpT Amino Acid Sequence

```
RAKLLGIVLT TPIAISSFAS TETLSFTPDN INADISLGTL SGKTKERVYL
AEEGGRKVSQ LDWKFNNAAI IKGAINWDLM PQISIGAAGW TTLGSRGGNM
VDQDWMDSSN PGTWTDESRH PDTQLNYANE FDLNIKGWLL NEPNYRLGLM
AGYQESRYSF TARGGSYIYS SEEGFRDDIG SFPNGERAIG YKQRFKMPYI
GLTGSYRYED FELGGTFKYS GWVESSDNDE HYDPGKRITY RSKVKDQNYY
SVAVNAGYYV TPNAKVYVEG AWNRVTNKKG NTSLYDHNNN TSDYSKNGAG
IENYNFITTA GLKYTF
```

Note: the N-terminal methionine is cleaved off during protein expression and hence not shown in this sequence.

PagP Amino Acid Sequence

```
MNADEWMTTF RENIAQTWQQ PEHYDLYIPA ITWHARFAYD KEKTDRYNER
PWGGGFGLSR WDEKGNWHGL YAMAFKDSWN KWEPIAGYGW ESTWRPLADE
NFHLGLGFTA GVTARDNWNY IPLPVLLPLA SVGYGPVTFQ MTYIPGTYNN
GNVYFAWMRF QFLEHHHHHH
```

Note: the C-terminus of PagP was His-tagged to aid protein purification

4.4.3 Expression and Purification of Membrane Proteins

4.4.3.1 Pre-culture

Pre-cultures were mediated by addition of 20 µL bacterial glycerol stock to 500 mL sterile LB media (Pilus Materials and Methods, Section 2.4.8). The medium, containing 100 µg/mL ampicillin, was incubated overnight at 37 °C overnight, shaking at 200 rpm and used immediately for large-scale protein over-expression.

4.4.3.2 Cytoplasmic Protein Expression

Two 2 L flasks each containing 1 L of LB media were pre-warmed overnight at 37 °C. 30 mL of pre-culture containing *E.coli* BL21 (DE3) cells with the relevant plasmid transformed into it were added to each flask along with 1 mL of 100 mg/mL carbenicillin (stored at -20 °C). The culture was incubated at 37 °C, shaking at 200 rpm in an orbital shaking incubator, and the optical density (O.D) at 600 nm was measured at regular intervals. At an OD of 0.6, protein expression was induced by addition of 1 mL of 1 M filter sterilised IPTG. After 3 hours, the cultures were harvested at 6,000 rpm, 4 °C.

4.4.3.3 Inclusion Body Isolation

The cell pellet was resuspended in 100 mL lysis buffer containing 50 mM Tris-HCl pH 8.0, phenyl methyl sulponyl fluoride (PMSF) (50 µg/mL), DNase (20 µg/mL), ethylenediaminetetraacetic acid disodium salt (EDTA) (5 mM) and incubated at room temperature for 30 min (lysozyme (100 µg/mL) was additionally added to the $PagP_{his}$ prep). Cells were further lysed using a TS Series cell disruptor (Constant Systems Ltd, UK) at 20 kpsi. The cell lysate was then centrifuged at 20,000 rpm for 30 min at 4 °C. The supernatant was discarded and the cell pellet resuspended in 100 mL 50 mM Tris-HCl pH 8.0. The cells were centrifuged (20,000 rpm, 30 min, 4 °C) and resuspended in 50 mM Tris-HCl pH 8.0, 1 % (*v/v*) Triton X-100. The cells were then centrifuged (20,000 rpm, 30 min, 4 °C) and resuspended in 50 mM Tris-HCl pH 8.0 before the final centrifugation step (20,000 rpm, 30 min, 4 °C). The final cell pellet was solubilised in 150 mL 25 mM Tris-HCl pH 8.0, 250 mM NaCl containing 6 M guanidine hydrochloride and incubated for 2 hours on an orbital shaker at 70 rpm, 4 °C. Any insoluble material remaining was discarded by centrifugation at 12,000 rpm for 1 h at 4 °C.

4.4.3.4 Metal Affinity Chromatography (PagP purification)

PagP was purified using a Histrap (5 mL) column (GE Healthcare, Buckinghamshire, UK). All solutions were filtered using a 0.22 µm Steritop filter (Millipore, Massachusetts, USA) and degassed using a Buchi Vac V-500 vacuum pump (Buchhi Labortechnik AG, Switzerland) prior to passing through the column. The column was equilibrated with 5 column volumes of 25 mM Tris-HCl pH 8.0, 250 mM NaCl, 6 M guanidine hydrochloride and the impure protein solution loaded onto the column at a flow rate of 2 mL/min using a peristaltic pump. The column containing the His-tagged protein was transferred to an ÄKTA prime (GE Healthcare, Buckinghamshire, UK) and washed with 5 column volumes of 25 mM Tris-HCl pH 8.0, 250 mM NaCl, 6 M guanidine hydrochloride, 25 mM immidazole at 4 ºC at a flow rate of 2 mL/min to remove any non-specifically bound proteins. Finally $PagP_{his}$ was eluted using 25 mM Tris-HCl pH 8.0, 250 mM NaCl, 6 M guanidine hydrochloride, 500 mM immidazole. The fractions containing $PagP_{his}$ (as detected by spectrophotometry at A_{280} and SDS-PAGE) were combined and dialysed

into H_2O (18 MΩ) at 4 °C. The PagP precipitate was pelleted by centrifugation and stored at -80 ºC until further use.

4.4.3.5 Size Exclusion Chromatography (OmpT purification)

OmpT inclusion bodies solubilised in 25 mM Tris-HCl pH 8.0, 250 mM NaCl, 6 M guanidine hydrochloride were filtered through a 0.22 μm Steritop filter (Millipore, Massachusetts, USA) and injected (5 mL per run) onto a HiLoad 26/60 Superdex 75 column (GE Healthcare, Buckinghamshire, UK) equilibrated in 20 mM Tris-HCl pH 8.0, 250 mM NaCl, 6 M guanidine hydrochloride and controlled by an ÄKTA prime. OmpT typically eluted after 100 mL using a flow rate of 3 mL/min. Several gel filtration runs were necessary to purify the whole batch of protein. Finally all fractions containing pure OmpT, as detected by spectrophotometry at A_{280} and SDS-PAGE, were combined and dialysed into H_2O (18 MΩ) before storage as a precipitate at -80 °C.

4.4.4 Characterisation of Membrane Proteins

Membrane Protein Purity

The purity of PagP and OmpT were determined by a single peak on the a HiLoad 26/60 Superdex 75 (GE Healthcare, Buckinghamshire, UK) gel filtration column in addition to a single band being observed at the correct molecular weight using SDS-PAGE analysis (Pilus Materials and Methods, Section 2.4.11).

Determination of Membrane Protein Concentration

The concentration of PagP and OmpT for all experiments was calculated using the Beer-Lambert law by measuring the absorbance at 280 nm and using the theoretical extinction coefficients determined from the amino acid sequence (using Protparam online software) (Edelhoch 1967; Gill *et al.* 1989; Pace *et al.* 1995). The extinction coefficients used for PagP and OmpT are 82390 and 78270 $M^{-1}cm^{-1}$ respectively.

4.4.5 Membrane Protein Folding into Ampipols

PagP and OmpT were initially dissolved in 100 mM ammonium bicarbonate pH 8.0 containing 8 M urea. A8-35 (purchased from Affymetrix Ltd., High Wycombe, UK) was then added at a protein to A8-35 ratio of 1:5 (*w/w*) and the solution dialysed into 100 mM ammonium bicarbonate pH 8.0 at 4 °C for 24 h. A final membrane protein concentration of 1 mg/mL was used for all experiments and an A8-35 concentration of 5 mg/mL unless otherwise stated.

4.4.6 Analysis of Membrane Protein-amphipol Complexes.

4.4.6.1 Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Gel electrophoresis was routinely used to monitor over-expression and purification of recombinant proteins. Additionally cold SDS-PAGE was used to measure the protein folding yield into amphipol (Schnaitm 1973). Tris-tricine buffered gels were used in all cases (Pilus Materials and Methods, Section 2.4.11). For the cold SDS-PAGE gels, the samples made up in loading buffer were not boiled prior to loading on the SDS-PAGE gel.

4.4.6.2 Circular Dichroism

CD spectra of all proteins were recorded on a Chirascan™ CD Spectrometer (Applied Photophysics, Surrey, UK) using a 0.1 mm cuvette. 8 scans from 200 to 260 nm with a bandwidth of 1.0 nm, and a scan speed of 20 nm/min, were accumulated and averaged. Background spectra containing the amphipol alone were subtracted for all samples. The recorded CD spectra were normalised to obtain the mean residue molar ellipicity [Θ] (λ), in degcm^2/dmol:

$$[\Theta](\lambda) = 100\frac{\Theta(\lambda)}{c \cdot n \cdot l}$$

where l is the path length of the cuvette (cm), Θ (λ) is the recorded ellipticity (deg), c is the concentration (moles/L) and n is the number of amino acid residues, 296 and 170 for OmpT and PagP, respectively.

4.4.6.3 Activity Assays OmpT

To measure OmpT protease activity, the change in fluorescence emission at 430 nm of the cleavable fluorogenic peptide (Abz-Ala-Arg-Arg-Ala-Tyr-(NO_2)-NH_2) was measured upon excitation at 325 nm (Kramer *et al.* 2000). The excitation and emission slit widths were set to 2 nm and the fluorescence measured over a 300 s reaction timescale using a Photon Technology International fluorimeter (Ford, West Sussex, UK). The temperature was regulated to 25 °C using a waterbath and a 1 cm path length cuvette was used throughout. Where necessary, lipopolysaccharide (LPS) was added to the OmpT:A8-35 solution at a concentration of 1 mg/mL, and in all cases the fluorogenic peptide was added immediately before analysis to a final concentration of 50 µM. Samples were mixed manually, resulting in a dead time of approximately 15 s. The average specific enzyme activity over a range of protein concentrations was reported in amount of substrate utilised/mg of enzyme/min. A control in which OmpT was unfolded in 8 M urea confirmed the activity was due to the natively folded OmpT:A8-35 complex. The fluorescence due to the buffer and A8-35 alone was subtracted from all spectra acquired.

4.4.6.4 Activity Assays PagP

The enzymatic assay for PagP was adapted from a previously described method (Huysmans *et al.* 2007). The hydrolysis of *p*-nitrophenyl palmitate (*p*-NNP) to *p*-nitrophenol by PagP was monitored by the increase in absorbance at 410 nm. *p*-NPP (10 mM in isopropanol) was diluted into 100 mM ammonium bicarbonate pH 8.0 containing various concentrations of PagP:A8-35 complex to a final substrate concentration of 1 mM and the rate of reaction monitored over 200 min. The increase in absorbance due to A8-35 addition alone was subtracted from all measurements. A control in which PagP was unfolded in 8 M urea was used to confirm the activity was due to the natively folded PagP:A8-35 complex. The average specific enzyme turnover over a range of protein concentrations was reported in nmol product/min/µM of PagP using an extinction coefficient of 3390 M^{-1} cm^{-1} for *p*-nitrophenol.

4.4.6.5 Fluorescence Emission

To measure the extent of membrane protein folding in amphipol, the tryptophan and tyrosine residues in PagP and OmpT (0.5 µM) were excited at 280 nm and the fluorescence emission spectrum measured between 300 – 400 nm using a Photon Technology International fluorimeter (Ford, West-Sussex, UK). The excitation and emission slit widths were set to 2 nm, the temperature was regulated to 25 °C using a waterbath, and a 1 cm path length cuvette was used throughout. PagP and OmpT spectra were recorded folded in A8-35 and unfolded in 8 M urea. The fluorescence emission due to the buffer and A8-35 alone was subtracted from all spectra acquired. The average wavelength (nm) was calculated as follows between 320 and 370 nm (Royer *et al.* 1993):

$$< \lambda >= \frac{\sum_i \lambda_i I_i}{\sum_i I_i}$$

where $<\lambda>$ is the average wavelength, λ_i the wavelength and I_i the fluorescence intensity at λ_i.

,

4.4.6.6 Size Exclusion Chromatography

Size exclusion chromatography was carried out using a Superdex 200 (10/300) column connected to an Äkta system. 200 µL of protein:A8-35 complex (or membrane protein alone) containing 0.2 mg protein was injected onto the column equilibrated in either 100 mM ammonium bicarbonate (folded) or 100 mM ammonium bicarbonate pH 8.0 containing 8 M urea (unfolded). The protein was eluted at a flow rate of 0.5 mL/min at room temperature. Elution profiles were followed by measuring the absorbance at 280 nm. The elution of PagP and OmpT was additionally confirmed by SDS-PAGE analysis.

4.4.6.7 Mass Spectrometry

Buffer Preparation

Ammonium bicarbonate buffers were made at 100 mM in deionised (18 MΩ) water. The ammonium bicarbonate buffer was stored at 4 ºC for one week maximum.

ESI-(IMS)-MS analysis

Experimental measurements were performed on a Synapt HDMS mass spectrometer (Waters Ltd., Manchester, UK) equipped with a NanoMate (Advion Biosystems Inc., Ithaca, NY, U.S.A) nanoESI autosampling device. Positive nano-ESI with a capillary voltage of 1.75 kV and a nitrogen nebulising gas pressure of 0.5 p.s.i. was used for all experiments. For the mass spectra acquired on the amphipol alone, a cone voltage of 70 V, source temperature 60 °C, bias 20 V, trap voltage 6 V and a transfer voltages of 4 V was applied.

For data acquired on the membrane protein-amphipol complexes, a cone voltage of 170 V was applied. The trap and transfer T-wave devices were additionally increased to 150 and 100 V respectively, the bias set to 100 V and the backing pressure to 4.7 mbar. Ion mobility separation was performed for all samples by ramping the wave height from 4.5-28.5 V at a speed of 300 m/s. Drift times were corrected for mass dependent and mass independent times (Knapman *et al.* 2010) and the drift time cross-section function calibrated as reported previously (Smith *et al.* 2009) (See Amyloid Materials and Methods, Section 3.4.10.2). Computer-based cross-sectional area calculations were made from Protein Data Bank structures using the projection superposition approximation (PSA) method described previously (Bleiholder *et al.* 2011). An aqueous solution of CsI was used for mass calibration. All data were acquired over the *m/z* range 500-8000 and the raw data processed by use of MassLynx *v*.4.1 and Driftscope *v*.3.0 (Waters Ltd., Manchester, UK).

References

Ahn, V. E., Lo, E. I., Engel, C. K., Chen, L., Hwang, P. M., Kay, L. E., Bishop, R. E. and Prive, G. G. (2004). A hydrocarbon ruler measures palmitate in the enzymatic acylation of endotoxin. *EMBO J.* **23**(15): 2931-2941.

Aminlashgari, N. and Hakkarainen, M. (2012). Emerging mass spectrometric tools for analysis of polymers and polymer additives. *Mass spectrometry of polymers - new techniques*. **248**: 1-37.

Arnold, T. and Linke, D. (2008). The use of detergents to purify membrane proteins. Current protocols in protein science/editorial board, John E. Coligan Chapter **4**: Unit 4.8.1-4.8.30.

Baneres, J. L., Popot, J. L. and Mouillac, B. (2011). New advances in production and functional folding of G-protein-coupled receptors. *Trends in Biotechnology* **29**(7): 314-322.

Barrera, N. P., Di Bartolo, N., Booth, P. J. and Robinson, C. V. (2008). Micelles protect membrane complexes from solution to vacuum. *Science* **321**(5886): 243-246.

Barrera, N. P., Isaacson, S. C., Zhou, M., Bavro, V. N., Welch, A., Schaedler, T. A., Seeger, M. A., Miguel, R. N., Korkhov, V. M., van Veen, H. W., Venter, H., Walmsley, A. R., Tate, C. G. and Robinson, C. V. (2009). Mass spectrometry of membrane transporters reveals subunit stoichiometry and interactions. *Nat. Methods.* **6**(8): 585-587.

Barrera, N. P. and Robinson, C. V. (2011). Advances in the mass spectrometry of membrane proteins: From individual proteins to intact complexes. *Annu. Rev. Biochem.* **80**: 247-271.

Bayburt, T. H. and Sligar, S. G. (2010). Membrane protein assembly into nanodiscs. *FEBS Lett.* **584**(9): 1721-1727.

Bazzacco, P., Billon-Denis, E., Sharma, K. S., Catoire, L. J., Mary, S., Le Bon, C., Point, E., Baneres, J. L., Durand, G., Zito, F., Pucci, B. and Popot, J. L. (2012). Nonionic homopolymeric amphipols: Application to membrane protein folding, cell-free synthesis, and solution nuclear magnetic resonance. *Biochemistry* **51**(7): 1416-1430.

Bechara, C., Bolbach, G., Bazzaco, P., Sharma, K. S., Durand, G., Popot, J.-L., Zito, F. and Sagan, S. (2012). MALDI-TOF mass spectrometry analysis of amphipol-trapped membrane proteins. *Anal. Chem.* **84**(14): 6128-6135.

Bishop, R. E., Gibbons, H. S., Guina, T., Trent, M. S., Miller, S. I. and Raetz, C. R. H. (2000). Transfer of palmitate from phospholipids to lipid a in outer membranes of gram-negative bacteria. *EMBO J.* **19**(19): 5071-5080.

Bishop, R. E. (2005). The lipid a palmitoyltransferase PagP: Molecular mechanisms and role in bacterial pathogenesis. *Mol. Microbiol.* **57**(4): 900-912.

Bleiholder, C., Wyttenbach, T. and Bowers, M. T. (2011). A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (i). Method. *Int. J. Mass Spectrom.* **308**(1): 1-10.

Bos, M. P., Robert, V. and Tommassen, J. (2007). Biogenesis of the Gram-negative bacterial outer membrane. *Annu. Rev. Microbiol.* **61**: 191-214.

Brandenburg, K., Garidel, P., Schromm, A. B., Andra, J., Kramer, A., Egmond, M. and Wiese, A. (2005). Investigation into the interaction of the bacterial protease OmpT with outer membrane lipids and biological activity of OmpT:Lipopolysaccharide complexes. *Euro. Biophys. J. with Biophys. Lett.* **34**(1): 28-41.

Bush, M. F., Hall, Z., Giles, K., Hoyes, J., Robinson, C. V. and Ruotolo, B. T. (2010). Collision cross sections of proteins and their complexes: A calibration framework and database for gas-phase structural biology. *Anal. Chem.* **82**(22): 9557-9565.

Carpenter, E. P., Beis, K., Cameron, A. D. and Iwata, S. (2008). Overcoming the challenges of membrane protein crystallography. *Curr. Opin. Struct. Biol.* **18**(5): 581-586.

Catoire, L. J., Zoonens, M., van Heijenoort, C., Giusti, F., Popot, J. L. and Guittet, E. (2009). Inter- and intramolecular contacts in a membrane protein/surfactant complex observed by heteronuclear dipole-to-dipole cross-relaxation. *J. Mag. Res.* **197**(1): 91-95.

Catoire, L. J., Damian, M., Giusti, F., Martin, A., van Heijenoort, C., Popot, J. L., Guittet, E. and Baneres, J. L. (2010a). Structure of a gpcr ligand in its receptor-bound state: Leukotriene B4 adopts a highly constrained conformation when associated to human BLT2. *J. Am. Chem. Soc.* **132**(26): 9049-9057.

Catoire, L. J., Zoonens, M., van Heijenoort, C., Giusti, F., Guittet, E. and Popot, J. L. (2010b). Solution NMR mapping of water-accessible residues in the transmembrane beta-barrel of OmpX. *Euro. Biophys. J. with Biophys. Lett.* **39**(4): 623-630.

Cuesta-Seijo, J. A., Neale, C., Khan, M. A., Moktar, J., Tran, C. D., Bishop, R. E., Pomes, R. and Prive, G. G. (2010). PagP crystallized from SDS/cosolvent reveals the route for phospholipid access to the hydrocarbon ruler. *Structure* **18**(9): 1210-1219.

Cvetkov, T. L., Huynh, K. W., Cohen, M. R. and Moiseenkova-Bell, V. Y. (2011). Molecular architecture and subunit organization of TrpA1 ion channel revealed by electron microscopy. *J. Biol. Chem.* **286**(44): 38168-38176.

Dahmane, T., Damian, M., Mary, S., Popot, J. L. and Baneres, J. L. (2009). Amphipol-assisted *in vitro* folding of G protein-coupled receptors. *Biochemistry* **48**(27): 6516-6521.

Edelhoch, H. (1967). Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry* **6**(7): 1948-1954.

Flotenmeyer, M., Weiss, H., Tribet, C., Popot, J. L. and Leonard, K. (2007). The use of amphipathic polymers for cryo electron microscopy of NADH:Ubiquinone oxidoreductase (complex i). *J. Microscopy-Ox.* **227**(3): 229-235.

Garavito, R. M. and Ferguson-Miller, S. (2001). Detergents as tools in membrane biochemistry. *J. Biol. Chem.* **276**(35): 32403-32406.

Garcia-Saez, A. J. and Schwille, P. (2007). Single molecule techniques for the study of membrane proteins. *App. Microbiol. Biotech.* **76**(2): 257-266.

Gill, S. C. and Vonhippel, P. H. (1989). Calculation of protein extinction coefficients from amino-acid sequence data. *Anal. Biochem.* **182**(2): 319-326.

Gohon, Y., Giusti, F., Prata, C., Charvolin, D., Timmins, P., Ebel, C., Tribet, C. and Popot, J. L. (2006). Well-defined nanoparticles formed by hydrophobic assembly of a short and polydisperse random terpolymer, amphipol A8-35. *Langmuir* **22**(3): 1281-1290.

Gohon, Y., Dahmane, T., Ruigrok, R. W. H., Schuck, P., Charvolin, D., Rappaport, F., Timmins, P., Engelman, D. M., Tribet, C., Popot, J. L. and Ebel, C. (2008). Bacteriorhodopsin/amphipol complexes: Structural and functional properties. *Biophysical J.* **94**(9): 3523-3537.

Guo, L., Lim, K. B., Poduje, C. M., Daniel, M., Gunn, J. S., Hackett, M. and Miller, S. I. (1998). Lipid A acylation and bacterial resistance against vertebrate antimicrobial peptides. *Cell* **95**(2): 189-198.

Hritonenko, V. and Stathopoulos, C. (2007). Omptin proteins: An expanding family of outer membrane proteases in Gram-negative *Enterobacteriaceae*. *Mol. Mem. Biol.* **24**(5-6): 395-406.

Huysmans, G. H. M., Radford, S. E., Brockwell, D. J. and Baldwin, S. A. (2007). The N-terminal helix is a post-assembly clamp in the bacterial outer membrane protein PagP. *J. Mol. Biol.* **373**(3): 529-540.

Hwang, P. M., Choy, W. Y., Lo, E. I., Chen, L., Forman-Kay, J. D., Raetz, C. R. H., Prive, G. G., Bishop, R. E. and Kay, L. E. (2002). Solution structure and dynamics of the outer membrane enzyme PagP by NMR. *Proc. Nat. Acad. Sci. USA* **99**(21): 13560-13565.

Hwang, P. M., Bishop, R. E. and Kay, L. E. (2004). The integral membrane enzyme PagP alternates between two dynamically distinct states. *Proc. Nat. Acad. Sci. USA* **101**(26): 9618-9623.

Ilag, L. L., Ubarretxena-Belandia, I., Tate, C. G. and Robinson, C. V. (2004). Drug binding revealed by tandem mass spectrometry of a protein-micelle complex. *J. Am. Chem. Soc.* **126**(44): 14362-14363.

Kelly, S. M., Jess, T. J. and Price, N. C. (2005). How to study proteins by circular dichroism. *Biochim. et Biophys. Acta-Proteins and Proteomics* **1751**(2): 119-139.

Khan, M. A., Neale, C., Michaux, C., Pomes, R., Prive, G. G., Woody, R. W. and Bishop, R. E. (2007). Gauging a hydrocarbon ruler by an intrinsic exciton probe. *Biochemistry* **46**(15): 4565-4579.

Khan, M. A., Moktar, J., Mott, P. J. and Bishop, R. E. (2010). A thiolate anion buried within the hydrocarbon ruler perturbs pagp lipid acyl chain selection. *Biochemistry* **49**(11): 2368-2379.

Knapman, T. W., Berryman, J. T., Campuzano, I., Harris, S. A. and Ashcroft, A. E. (2010). Considerations in experimental and theoretical collision cross-section measurements of small molecules using travelling wave ion mobility spectrometry-mass spectrometry. *Int. J. Mass Spectrom.* **298**(1-3): 17-23.

Knowles, T. J., Scott-Tucker, A., Overduin, M. and Henderson, I. R. (2009). Membrane protein architects: The role of the Bam complex in outer membrane protein assembly. *Nat. Rev. Microbiol.* **7**(3): 206-214.

Kramer, R. A., Zandwijken, D., Egmond, M. R. and Dekker, N. (2000). In vitro folding, purification and characterization of *Escherichia coli* outer membrane protease OmpT. *Euro. J. Biochem.* **267**(3): 885-893.

Kramer, R. A., Brandenburg, K., Vandeputte-Rutten, L., Werkhoven, M., Gros, P., Dekker, N. and Egmond, M. R. (2002). Lipopolysaccharide ride regions involved in the activation of *Escherichia coli* outer membrane protease OmpT. *Euro. J. Biochem.* **269**(6): 1746-1752.

Locher, K. P., Lee, A. T. and Rees, D. C. (2002). The *E-coli* BtuCD structure: A framework for ABC transporter architecture and mechanism. *Science* **296**(5570): 1091-1098.

Marassi, F. M. and Opella, S. J. (1998). NMR structural studies of membrane proteins. *Curr. Opin. Struct. Biol.* **8**(5): 640-648.

Martinez, K. L., Gohon, Y., Corringer, P. J., Tribet, C., Merola, F., Changeux, J. P. and Popot, J. L. (2002). Allosteric transitions of torpedo acetylcholine receptor in lipids, detergent and amphipols: Molecular interactions vs. Physical constraints. *FEBS Lett.* **528**(1-3): 251-256.

Marty, M. T., Das, A. and Sligar, S. G. (2012). Ultra-thin layer MALDI mass spectrometry of membrane proteins in nanodiscs. *Anal. and Bioanal. Chem.* **402**(2): 721-729.

McCarter, J. D., Stephens, D., Shoemaker, K., Rosenberg, S., Kirsch, J. F. and Georgiou, G. (2004). Substrate specificity of the *Escherichia Coli* outer membrane protease OmpT. *J. Bacteriol.* **186**(17): 5919-5925.

Mogensen, J. E. and Otzen, D. E. (2005). Interactions between folding factors and bacterial outer membrane proteins. *Mol. Microbiol.* **57**(2): 326-346.

Moon, C. P. and Fleming, K. G. (2011). Using tryptophan fluorescence to measure the stability of membrane proteins folded in liposomes. *Methods in Enzymology: Biothermodynamics*, **492**: 189-211.

Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. and Gray, T. (1995). How to measure and predict the molar absorption-coefficient of a protein. *Protein Sci.* **4**(11): 2411-2423.

Pocanschi, C. L., Dahmane, T., Gohon, Y., Rappaport, F., Apell, H. J., Kleinschmidt, J. H. and Popot, J. L. (2006). Amphipathic polymers: Tools to fold integral membrane proteins to their active form. *Biochemistry* **45**(47): 13954-13961.

Popot, J. L., Berry, E. A., Charvolin, D., Creuzenet, C., Ebel, C., Engelman, D. M., Flotenmeyer, M., Giusti, F., Gohon, Y., Herve, P., Hong, Q., Lakey, J. H., Leonard, K., Shuman, H. A., Timmins, P., Warschawski, D. E., Zito, F., Zoonens, M., Pucci, B. and Tribet, C. (2003). Amphipols: Polymeric surfactants for membrane biology research. *Cell. Mol. Life Sci.* **60**(8): 1559-1574.

Popot, J. L. (2010). Amphipols, nanodiscs, and fluorinated surfactants: Three nonconventional approaches to studying membrane proteins in aqueous solutions. *Annu. Rev. Biochem.* **79**: 737-775.

Popot, J. L., Althoff, T., Bagnard, D., Baneres, J. L., Bazzacco, P., Billon-Denis, E., Catoire, L. J., Champeil, P., Charvolin, D., Cocco, M. J., Cremel, G., Dahmane, T., de la Maza, L. M., Ebel, C., Gabel, F., Giusti, F., Gohon, Y., Goormaghtigh, E., Guittet, E., Kleinschmidt, J. H., Kuhlbrandt, W., Le Bon, C., Martinez, K. L., Picard, M., Pucci, B., Sachs, J. N., Tribet, C., van Heijenoort, C., Wien, F., Zito, F. and Zoonens, M. (2011). Amphipols from A to Z. *Annu. Rev. Biophys.* **40**: 379-408.

Raunser, S. and Walz, T. (2009). Electron crystallography as a technique to study the structure on membrane proteins in a lipidic environment. *Annu. Rev. Biophys.* **38**: 89-105.

Ritchie, T. K., Grinkova, Y. V., Bayburt, T. H., Denisov, I. G., Zolnerciks, J. K., Atkins, W. M. and Sligar, S. G. (2009). Reconstitution of membrane proteins in

phospholipid bilayer nanodiscs. Methods in enzymology; liposomes, pt f. Duzgunes, N. San Diego, *Elsevier Academic Press Inc.* **464**: 211-231.

Royer, C. A., Mann, C. J. and Matthews, C. R. (1993). Resolution of the fluorescence equilibrium unfolding profile of Trp aporepressor using single tryptophan mutants. *Protein Sci.* **2**(11): 1844-1852.

Schnaitm, C. (1973). Outer membrane proteins of *Escherichia-coli* .1. Effect of preparative conditions on migration of protein in polyacrylamide gels. *Arch. Biochem. Biophys.* **157**(2): 541-552.

Schulz, G. E. (2002). The structure of bacterial outer membrane proteins. *Biochim. et Biophys. Acta-Biomem.* **1565**(2): 308-317.

Seddon, A. M., Curnow, P. and Booth, P. J. (2004). Membrane proteins, lipids and detergents: Not just a soap opera. *Biochim. et Biophys. Acta-Biomem.* **1666**(1-2): 105-117.

Sharma, K. S., Durand, G., Giusti, F., Olivier, B., Fabiano, A. S., Bazzacco, P., Dahmane, T., Ebel, C., Popot, J. L. and Pucci, B. (2008). Glucose-based amphiphilic telomers designed to keep membrane proteins soluble in aqueous solutions: Synthesis and physicochemical characterization. *Langmuir* **24**(23): 13581-13590.

Smith, D. P., Knapman, T. W., Campuzano, I., Malham, R. W., Berryman, J. T., Radford, S. E. and Ashcroft, A. E. (2009). Deciphering drift time measurements from travelling wave ion mobility spectrometry-mass spectrometry studies. *Eur. J. Mass Spectrom.* **15**(2): 113-130.

Tate, C. G. (2006). Comparison of three structures of the multidrug transporter EmrE. *Curr. Opin. Struct. Biol.* **16**(4): 457-464.

Tifrea, D. F., Sun, G. F., Pal, S., Zardeneta, G., Cocco, M. J., Popot, J. L. and de la Maza, L. M. (2011). Amphipols stabilize the Chlamydia major outer membrane protein and enhance its protective ability as a vaccine. *Vaccine* **29**(28): 4623-4631.

Vandeputte-Rutten, L., Kramer, R. A., Kroon, J., Dekker, N., Egmond, M. R. and Gros, P. (2001). Crystal structure of the outer membrane protease OmpT from *Escherichia coli* suggests a novel catalytic site. *EMBO J.* **20**(18): 5033-5039.

Wang, S. C., Politis, A., Di Bartolo, N., Bavro, V. N., Tucker, S. J., Booth, P. J., Barrera, N. P. and Robinson, C. V. (2010). Ion mobility mass spectrometry of two tetrameric membrane protein complexes reveals compact structures and differences in stability and packing. *J. Am. Chem. Soc.* **132**(44): 15468-15470.

White, C. B., Chen, Q., Kenyon, G. L. and Babbitt, P. C. (1995). A novel activity of OmpT - proteolysis under extreme denaturing conditions. *J. Biol. Chem.* **270**(22): 12990-12994.

Zhou, Y. F., Lau, F. W., Nauli, S., Yang, D. and Bowie, J. U. (2001). Inactivation mechanism of the membrane protein diacylglycerol kinase in detergent solution. *Protein Sci.* **10**(2): 378-383.

Zhou, M., Morgner, N., Barrera, N. P., Politis, A., Isaacson, S. C., Matak-Vinkovic, D., Murata, T., Bernal, R. A., Stock, D. and Robinson, C. V. (2011). Mass spectrometry of intact V-type ATPases reveals bound lipids and the effects of nucleotide binding. *Science* **334**(6054): 380-385.

Zoonens, M., Catoire, L. J., Giusti, F. and Popot, J. L. (2005). NMR study of a membrane protein in detergent-free aqueous solution. *Proc. Nat. Acad. Sci. USA* **102**(25): 8893-8898.

Zoonens, M., Giusti, F., Zito, F. and Popot, J. L. (2007). Dynamics of membrane protein/amphipol association studied by Förster resonance energy transfer: Implications for *in vitro* studies of amphipol-stabilized membrane proteins. *Biochemistry* **46**(36): 10392-10404.

Final Remarks

The work contained in this thesis describes the application of mass spectrometry to the structural analysis of proteins and non-covalent protein complexes. Mass spectrometry was chosen due to its advantages over other, more established biochemical techniques. These features include low sample consumption, no theoretical upper mass limit, and the ability to distinguish between multiple species co-populated in solution on account of their molecular weight. Over the last 30 years, with the development of electrospray ionization mass spectrometry (ESI-MS), protein analysis by mass spectrometry has excelled. As a result, mass spectrometry is now routinely used not only to determine the molecular weight of non-covalently bound protein complexes in solution, but also to determine post-translational modifications (Witze *et al.* 2007; Bechara *et al.* 2012), subunit stoichiometries (Loo 1997; Barrera *et al.* 2009; Gordiyenko *et al.* 2010; Janganan *et al.* 2011), protein-protein interaction interfaces (Dodds *et al.* 2011; Liu *et al.* 2011), kinetic reaction mechanisms (Konermann *et al.* 1997; Fandrich *et al.* 2000; Rose *et al.* 2008), protein-ligand binding affinities (Wortmann *et al.* 2008; Jonker *et al.* 2011; Kitova *et al.* 2012) and in general to unravel complex protein assembly mechanisms (Knapman *et al.* 2010; Smith *et al.* 2010; Morrissey *et al.* 2012).

Mass spectrometry can be used as a tool to monitor biochemical reactions in real time. Mass spectrometry can also be used to follow protein folding on the millisecond timescale (Simmons *et al.* 2002; Pan *et al.* 2010; Stocks *et al.* 2010; Pan *et al.* 2011) and the complex assembly of protein subunits (Rose *et al.* 2008; Verger *et al.* 2008; Smith *et al.* 2010; Smith *et al.* 2011), for example in the assembly of the heterohexameric NtGimC chaperone complex (Fandrich *et al.* 2000). In this thesis mass spectrometry has been used to follow two contrasting assembly mechanisms; pilus biogenesis and amyloid fibril formation. During pilus assembly thousands of structurally similar protein subunits are assembled on the outer membrane of bacteria and mass spectrometry was used to study these protein subunits based on their molecular weights. In contrast, during amyloid assembly, thousands of identical subunits assemble with one another and it is only with the combination of ion

mobility spectrometry and mass spectrometry that this heterogeneous mixture of amyloidogenic oligomers was separated.

In particular, ESI-MS has been used in this thesis to determine accurately second order rate constants of pilus biogenesis in real time. The key to this success was the accurate quantification of the chaperone-subunit complex over time by the addition of a structurally similar internal standard. The application of this method led to the identification of specific amino acids in the N-terminal extension of the pilin subunits which are important in determining the specificity of pilus biogenesis (Leney *et al.* 2011).

ESI-IMS-MS was used to determine the relative populations of β_2m oligomers during amyloid fibril formation. Comparison of WT β_2m oligomers with oligomers derived from a series of mutants covering the entire β_2m protein sequence revealed specific residues important in amyloidosis, highlighting the complexity of this fibril-forming process. Further, detailed characterisation of the β_2m mutant H51A revealed subtle differences in the subunit exchange dynamics of the oligomeric species involved in fibril formation, even though their gas phase stabilities, cross sectional areas, and eventual fibril morphology remained comparable. Together these results, in combination with those on pilus assembly, demonstrate the power of mass spectrometry to discern protein reaction mechanisms in complex biological systems.

In addition to ESI-MS and ESI-IMS-MS analysis, tandem MS is a widely used technique to determine subunit stoichiometry (Barrera *et al.* 2009; Synowsky *et al.* 2009; Gordiyenko *et al.* 2010) in addition to revealing information about the nature of protein-protein interactions in solution (Chung *et al.* 1999; Erba *et al.* 2010; Dodds *et al.* 2011). CID-MS/MS can also reveal differences in binding affinities between protein-protein complexes (Daniel *et al.* 2002; Erba *et al.* 2010). This thesis shows how CID-MS/MS can be used to determine differences in the gas phase stabilities of the chaperone-subunit-usher ternary complexes formed during pilus assembly. By truncation of the PapG subunit into its adhesion and pilin domains, the reason for the increased affinity of the PapG subunit above all other subunits for the outer membrane usher was revealed (Morrissey *et al.* 2012).

CID MS/MS has also been suggested as a useful tool to understand more about the nature of protein-protein complexes, providing further insight as to whether their

binding interfaces are hydrophobic or electrostatic in nature. On subjecting dimer and trimer β_2m oligomers to CID-MS/MS, the oligomers underwent symmetrical division producing equally charged dissociation products. This is unusual for CID-MS/MS analysis of the dissociation products where asymmetric division is the proposed mechanism of dissociation (Sciuto *et al.* 2011). Although further data are needed, preliminary data suggest the lack of symmetrical division in the amyloidogenic intermediates is due to the interface being more hydrophobic rather than electrostatic in nature. Additional covalent labelling experiments for example oxidative labeling (Konermann *et al.* 2010) may reveal insights into the location of individual amino acid residues in the on-pathway oligomeric intermediates.

More recent attention in native protein mass spectrometry has turned to membrane proteins, which only in the last few years has been found possible. Membrane protein analysis by mass spectrometry is challenging due to the insolubility of these species in aqueous and organic buffers. Recent advances by Carol Robinson and collegues have shown it is possible to transfer membrane proteins in detergent micelles intact into the gas phase (Ilag *et al.* 2004; Barrera *et al.* 2008). Detergent micelles are not optimal for membrane protein analysis and alternative solubilisation techniques are needed to maintain membrane proteins in their native, functional state for structural and functional analysis (Seddon *et al.* 2004; Popot 2010). The use of amphipols for the analysis of membrane proteins by mass spectrometry has been investigated in this thesis. The two bacterial outer membrane proteins, OmpT and PagP, were inserted into amphipols and successfully analysed by ESI-MS. These data show great promise for the future of membrane protein analysis using mass spectrometry with membrane proteins remaining stable in amphipol for analysis for up to two months compared to a few days in detergent micelles and with minimal sample preparation required. The question now remains as to whether amphipols can protect membrane protein complexes on transition into the gas phase. This will greatly increase the applicability of amphipols for membrane proteins mass spectrometry analysis.

Altogether mass spectrometry holds great promise for becoming a routinely used analytical tool for analysis of protein structure. Advances in amino acid labeling techniques will allow single amino acid residues to be tracked throughout protein complex assembly, while developments in mass spectrometric imaging could

provide information simultaneously on the spatial arrangement of these biomolecules in the gas phase. In the meantime, the use of mass spectrometry in combination with other in solution biochemical techniques is the most powerful approach to elucidate complex biochemical mechanisms.

References

Barrera, N. P., Di Bartolo, N., Booth, P. J. and Robinson, C. V. (2008). Micelles protect membrane complexes from solution to vacuum. *Science* **321**(5886): 243-246.

Barrera, N. P., Isaacson, S. C., Zhou, M., Bavro, V. N., Welch, A., Schaedler, T. A., Seeger, M. A., Miguel, R. N., Korkhov, V. M., van Veen, H. W., Venter, H., Walmsley, A. R., Tate, C. G. and Robinson, C. V. (2009). Mass spectrometry of membrane transporters reveals subunit stoichiometry and interactions. *Nature Methods* **6**(8): 585-587.

Bechara, C., Bolbach, G., Bazzaco, P., Sharma, K. S., Durand, G., Popot, J.-L., Zito, F. and Sagan, S. (2012). MALDI-TOF mass spectrometry analysis of amphipol-trapped membrane proteins. *Anal. Chem*. **84**(14): 6128-6135.

Chung, E. W., Henriques, D. A., Renzoni, D., Morton, C. J., Mulhern, T. D., Pitkeathly, M. C., Ladbury, J. E. and Robinson, C. V. (1999). Probing the nature of interactions in SH2 binding interfaces - evidence from electrospray ionization mass spectrometry. *Protein Sci*. **8**(10): 1962-1970.

Daniel, J. M., Friess, S. D., Rajagopalan, S., Wendt, S. and Zenobi, R. (2002). Quantitative determination of noncovalent binding interactions using soft ionization mass spectrometry. *Int. J. Mass Spectrom.* **216**(1): 1-27.

Dodds, E. D., Blackwell, A. E., Jones, C. M., Holso, K. L., O'Brien, D. J., Cordes, M. H. J. and Wysocki, V. H. (2011). Determinants of gas-phase disassembly behavior in homodimeric protein complexes with related yet divergent structures. *Anal. Chem.* **83**(10): 3881-3889.

Erba, E. B., Ruotolo, B. T., Barsky, D. and Robinson, C. V. (2010). Ion mobility-mass spectrometry reveals the influence of subunit packing and charge on the dissociation of multiprotein complexes. *Anal. Chem*. **82**(23): 9702-9710.

Fandrich, M., Tito, M. A., Leroux, M. R., Rostom, A. A., Hartl, F. U., Dobson, C. M. and Robinson, C. V. (2000). Observation of the noncovalent assembly and disassembly pathways of the chaperone complex MtGimC by mass spectrometry. *Proc. Nat. Acad. Sci. U.S.A.* **97**(26): 14151-14155.

Gordiyenko, Y., Videler, H., Zhou, M., McKay, A. R., Fucini, P., Biegel, E., Muller, V. and Robinson, C. V. (2010). Mass spectrometry defines the stoichiometry of ribosomal stalk complexes across the phylogenetic tree. *Mol. Cell. Prot.* **9**(8): 1774-1783.

Ilag, L. L., Ubarretxena-Belandia, I., Tate, C. G. and Robinson, C. V. (2004). Drug binding revealed by tandem mass spectrometry of a protein-micelle complex. *J. Am. Chem. Soc.* **126**(44): 14362-14363.

Janganan, T. K., Bavro, V. N., Zhang, L., Matak-Vinkovic, D., Barrera, N. P., Venien-Bryan, C., Robinson, C. V., Borges-Walmsley, M. I. and Walmsley, A. R. (2011). Evidence for the assembly of a bacterial tripartite multidrug pump with a stoichiometry of 3:6:3. *J. Biol. Chem.* **286**(30): 26900-26912.

Jonker, N., Kool, J., Irth, H. and Niessen, W. M. A. (2011). Recent developments in protein-ligand affinity mass spectrometry. *Anal. Bioanal. Chem*. **399**(8): 2669-2681.

Kitova, E. N., El-Hawiet, A., Schnier, P. D. and Klassen, J. S. (2012). Reliable determinations of protein-ligand interactions by direct ESI-MS measurements. Are we there yet? *J. Am. Soc. Mass Spectrom.* **23**(3): 431-441.

Knapman, T. W., Morton, V. L., Stonehouse, N. J., Stockley, P. G. and Ashcroft, A. E. (2010). Determining the topology of virus assembly intermediates using ion mobility spectrometry-mass spectrometry. *Rapid Comm. Mass Spectrom*. **24**(20): 3033-3042.

Konermann, L., Collings, B. A. and Douglas, D. J. (1997). Cytochrome C folding kinetics studied by time-resolved electrospray ionization mass spectrometry. *Biochemistry* **36**(18): 5554-5559.

Konermann, L., Stocks, B. B., Pan, Y. and Tong, X. (2010). Mass spectrometry combined with oxidative labeling for exploring protein structure and folding. *Mass Spectrom. Rev.* **29**(4): 651-667.

Leney, A. C., Phan, G., Allen, W., Verger, D., Waksman, G., Radford, S. E. and Ashcroft, A. E. (2011). Second order rate constants of donor-strand exchange reveal individual amino acid residues important in determining the subunit specificity of pilus biogenesis. *J. Am. Soc. Mass Spectrom*. **22**(7): 1214-1223.

Liu, J. J. and Konermann, L. (2011). Protein-protein binding affinities in solution determined by electrospray mass spectrometry. *J. Am. Soc. Mass Spectrom*. **22**(3): 408-417.

Loo, J. A. (1997). Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrom. Rev.* **16**(1): 1-23.

Morrissey, B., Leney, A. C., Toste Rego, A., Phan, G., Allen, W. J., Verger, D., Waksman, G., Ashcroft, A. E. and Radford, S. E. (2012). The role of chaperone-

subunit usher domain interactions in the mechanism of bacterial pilus biogenesis revealed by ESI-MS. Mol. Cell. Prot. **11**(7): M111.015289: 1-10.

Pan, J. X., Han, J., Borchers, C. H. and Konermann, L. (2010). Characterizing short-lived protein folding intermediates by top-down hydrogen exchange mass spectrometry. *Anal. Chem*. **82**(20): 8591-8597.

Pan, Y., Brown, L. and Konermann, L. (2011). Kinetic folding mechanism of an integral m. *J. Mol. Biol.* **410**(1): 146-158.

Popot, J. L. (2010). Amphipols, nanodiscs, and fluorinated surfactants: Three nonconventional approaches to studying membrane proteins in aqueous solutions. *Annu. Rev. Biochem*. **79**: 737-775.

Rose, R. J., Verger, D., Daviter, T., Remaut, H., Paci, E., Waksman, G., Ashcroft, A. E. and Radford, S. E. (2008). Unraveling the molecular basis of subunit specificity in p pilus assembly by mass spectrometry. *Proc. Nat. Acad. Sci. U.S.A.* **105**(35): 12873-12878.

Sciuto, S. V., Liu, J. J. and Konermann, L. (2011). An electrostatic charge partitioning model for the dissociation of protein complexes in the gas phase. *J. Am. Soc. Mass Spectrom*. **22**(10): 1679-1689.

Seddon, A. M., Curnow, P. and Booth, P. J. (2004). Membrane proteins, lipids and detergents: Not just a soap opera. *Biochim. et Biophys. Acta-Biomembranes* **1666**(1-2): 105-117.

Simmons, D. A. and Konermann, L. (2002). Characterization of transient protein folding intermediates during myoglobin reconstitution by time-resolved electrospray mass spectrometry with on-line isotopic pulse labeling. *Biochemistry* **41**(6): 1906-1914.

Smith, D. P., Radford, S. E. and Ashcroft, A. E. (2010). Elongated oligomers in beta(2)-microglobulin amyloid assembly revealed by ion mobility spectrometry-mass spectrometry. *Proc. Nat. Acad. Sci. U.S.A.* **107**(15): 6794-6798.

Smith, D. P., Woods, L. A., Radford, S. E. and Ashcroft, A. E. (2011). Structure and dynamics of oligomeric intermediates in beta(2)-microglobulin self-assembly. *Biophysical J.* **101**(5): 1238-1247.

Stocks, B. B. and Konermann, L. (2010). Time-dependent changes in side-chain solvent accessibility during cytochrome c folding probed by pulsed oxidative labeling and mass spectrometry. *J. Mol. Biol*. **398**(2): 362-373.

Synowsky, S. A., van Wijk, M., Raijmakers, R. and Heck, A. J. R. (2009). Comparative multiplexed mass spectrometric analyses of endogenously expressed yeast nuclear and cytoplasmic exosomes. *J. Mol. Biol.* **385**(4): 1300-1313.

Verger, D., Rose, R. J., Paci, E., Costakes, G., Daviter, T., Hultgren, S., Remaut, H., Ashcroft, A. E., Radford, S. E. and Waksman, G. (2008). Structural determinants of polymerization reactivity of the P pilus adaptor subunit PapF. *Structure* **16**(11): 1724-1731.

Witze, E. S., Old, W. M., Resing, K. A. and Ahn, N. G. (2007). Mapping protein post-translational modifications with mass spectrometry. *Nature Methods* **4**(10): 798-806.

Wortmann, A., Jecklin, M. C., Touboul, D., Badertscher, M. and Zenobi, R. (2008). Binding constant determination of high-affinity protein-ligand complexes by electrospray ionization mass spectrometry and ligand competition. *J. Mass Spectrom.* **43**(5): 600-608.

Appendix 1

Derivation of equation for second-order rate constant

$$A + B \rightarrow X$$

(A = [peptide], B = [chaperone-subunit], X = product)

$$\frac{d[B]}{dt} = -k[A][B]$$

$[X]_0 = 0$ - since no product initially

$[A] \equiv [A]_t$ and $[B] \equiv [B]_t$

Expressing [A] and [B] in terms of [X]

$$[A] = [A]_0 - [X]$$
$$[B] = [B]_0 - [X]$$

Therefore the rate equation in terms of [X] is

$$\frac{d[B]}{dt} = -\frac{d[X]}{dt}$$

$$-\frac{d[X]}{dt} = -k([A]_0 - [X])([B]_0 - [X])$$

Rearranging this in order to integrate it gives

$$-\frac{d[X]}{dt} = -k([X] - [A]_0)([X] - [B]_0)$$

$$\frac{d[X]}{([X] - [A]_0)([X] - [B]_0)} = kdt$$

This can be integrated using partial fractions

Using z and y as constants, making this a sum of two simpler fractions

$$\frac{1}{([X] - [A]_0)([X] - [B]_0)} = \frac{z}{[X] - [A]_0} + \frac{y}{[X] - [B]_0}$$

making a common denominator and multiplying out

$$\frac{1}{([X]-[A]_0)([X]-[B]_0)} = \frac{z([X]-[B]_0)+y([X]-[A]_0)}{([X]-[A]_0)([X]-[B]_0)}$$

$$\frac{1}{([X]-[A]_0)([X]-[B]_0)} = \frac{z([X]-[B]_0)+y([X]-[A]_0)}{([X]-[A]_0)([X]-[B]_0)}$$

$$1 = z([X]-[B]_0)+y([X]-[A]_0)$$

or $1 = z[X]-z[B]_0+y[X]-y[A]_0$

$$z[X]+y[X]=0$$

$$z[X]=-y[X]$$

$$z=-y$$

$$1=-z[B]_0-y[A]_0$$

or $1=-z[B]_0+z[A]_0$

so $z=\frac{1}{[A]_0-[B]_0}$ $\qquad$ $y=-\frac{1}{[A]_0-[B]_0}$

substituting these back into the previous equation

$$\frac{1}{([X]-[A]_0)([X]-[B]_0)} = \frac{1}{([X]-[A]_0)([A]_0-[B]_0)} - \frac{1}{([X]-[B]_0)([A]_0-[B]_0)}$$

$$\frac{d[X]}{([X]-[A]_0)([A]_0-[B]_0)} - \frac{d[X]}{([X]-[B]_0)([A]_0-[B]_0)} = kdt$$

Integrating this gives

$$\frac{1}{([A]_0-[B]_0)}\ln\left\{\frac{([X]-[A]_0)}{-[A]_0}\right\} - \frac{1}{([A]_0-[B]_0)}\ln\left\{\frac{([X]-[B]_0)}{-[B]_0}\right\} = kt$$

Changing the sign inside the ln and combining

$$\frac{1}{([A]_0-[B]_0)}\ln\left\{\frac{([A]_0-[X])}{[A]_0}\right\} - \frac{1}{([A]_0-[B]_0)}\ln\left\{\frac{([B]_0-[X])}{[B]_0}\right\} = kt$$

$$\frac{1}{([A]_0 - [B]_0)} \ln\left\{\frac{([A]_0 - [X])[B]_0}{[A]_0([B]_0 - [X])}\right\} = kt$$

$$\ln\left\{\frac{[B]_0([A]_0 - [X])}{[A]_0([B]_0 - [X])}\right\} = k([A]_0 - [B]_0)t$$

$$\ln\left\{\frac{[B]_0[A]}{[A]_0[B]}\right\} = k([A]_0 - [B]_0)t$$

Rearranging for [B]

(remember B = chaperone-subunit measured experimentally)

$k([A]_0 - [B]_0) = k'$ for simplicity

$$\frac{[B]_0[A]}{[A]_0[B]} = e^{k't} \qquad \text{since } [A] = [A]_0 - ([B]_0 - [B])$$

$$[B] = [B]_0[A]e^{-k't} \times \frac{1}{[A]_0}$$

$$[B] = [B]_0([A]_0 - ([B]_0 - [B]))e^{-k't} \times \frac{1}{[A]_0}$$

$$[B] = \frac{[B]_0}{[A]_0}([A]_0 - ([B]_0 - [B]))e^{-k't}$$

$$[B] = \frac{[B]_0}{[A]_0}e^{-k't}([A]_0 - [B]_0 + [B])$$

$$[B] = [B]_0 e^{-k't} - \frac{[B]_0{}^2 e^{-k't}}{[A]_0} + \frac{[B][B]_0 e^{-k't}}{[A]_0}$$

$$[B]-\frac{[B][B]_0 e^{-k't}}{[A]_0}=[B]_0 e^{-k't}-\frac{[B]_0{}^2 e^{-k't}}{[A]_0}$$

$$[B]\left(1-\frac{[B]_0 e^{-k't}}{[A]_0}\right)=[B]_0 e^{-k't}-\left(\frac{[B]_0{}^2 e^{-k't}}{[A]_0}\right)$$

$$[B]=\frac{[B]_0 e^{-k't}-\left(\frac{[B]_0{}^2 e^{-k't}}{[A]_0}\right)}{1-\left(\frac{[B]_0 e^{-k't}}{[A]_0}\right)}$$

$$[B]=\frac{[B]_0 e^{-k([A]_0-[B]_0)t}-\left(\frac{[B]_0{}^2 e^{-k([A]_0-[B]_0)t}}{[A]_0}\right)}{1-\left(\frac{[B]_0 e^{-k([A]_0-[B]_0)t}}{[A]_0}\right)}$$

Since $10^{\log k}$ = k, the reaction is ten fold diluted before analysis, and accounting for the initial ratio of PapDPapE:PapDPapH intensity at the different initial PapDPapH and PapDPapE concentrations used. The final equation for use in determining second order rate constants is as follows

$$[B]=x\left\{\frac{[B]_0 e^{-10^{\log k}([A]_0-[B]_0)t}-\left(\frac{[B]_0{}^2 e^{-10^{\log k}([A]_0-[B]_0)t}}{[A]_0}\right)}{1-\left(\frac{[B]_0 e^{-10^{\log k}([A]_0-[B]_0)t}}{[A]_0}\right)}\right\}$$

where [B] = concentration of PapDPapX
[A] = concentration of peptide
$[B]_0$ = [B] before addition of peptide
$[A]_0$ = [A] before addition of PapDPapX
k = second order rate constant
x = dilution factor before acquisition, and the DX:DH intensity ratio at t=0