Ion Mobility Mass Spectrometry Studies of Alpha$_1$-Antitrypsin

Institute of Structural
Molecular Biology

MPhil

Natacha Océane Lee

Declaration

I Natacha Lee, declare that the work presented in this thesis is my own. Any information obtained from other sources has been indicated in the thesis.

Abstract

Mass Spectrometry (MS) has significantly contributed to the field of structural biology in the last decades. In this thesis, native MS and ion mobility-MS (IM-MS) is used to study alpha$_1$-antitrypsin (α_1-AT) in order to better understand its pathophysiology, especially in terms of aggregation, which can cause early onset emphysema and liver cirrhosis in patients with certain mutations of this protein.

Initial work involved the study of recombinant α_1-AT by comparing the collision cross sections (CCS) and unfolding patterns of different recombinant mutations. Two stabilising variants (T114F and G117F) and a slowly polymerising variant (K154N) were studied alongside the recombinant protein. Indeed, the proteins did show differences in their CCSs and unfolding patterns as the stabilising variants retained their native structure more than the K154N variant at high collision energies.

Plasma α_1-AT variants (B, Z, and S) were then studied using native and IM-MS. Native MS was able to identify certain glycan residues attached to the plasma protein whilst IM-MS confirmed the stabilising effect of glycosylation on the protein upon increases in internal energy. IM-MS also revealed that the disease variants had larger CCS values than the plasma control, which suggests that in disease α_1-AT is more unfolded and primed to entering the aggregation pathway.

Finally, α_1-AT polymers were analysed by native MS and IM-MS. Although further work is needed, the spectra show differences in intensities of certain oligomeric species, which may lead to determining the pathway towards α_1-AT aggregation.

Acknowledgements

I would like to thank Dr. Thalassinos and Dr. Gooptu for their support, enthusiasm and knowledge throughout my research. Their input has been invaluable. I would also like to thank the chair of my thesis committee, John Christodoulou for his pastoral support in the last year.

In particular I would like to thank the group who have made the experience even more fun and to whom I can call friends. I feel extremely lucky to work among such a lively, friendly group of individuals. Adam, Juliette, Tanya and Zoja – you have been incredibly easy-to-work-with colleagues, I appreciate how rare it is to be able to get on with every one of you so thank you for putting up with my terrible jokes and comments. I would also like thank the office in general, Kieran, Claudie, Ana, Leo, Shaan and Amir, what wonderful people you are and I wish you all the best in whatever career paths you choose.

Lastly I would like to thank the Lomas group, especially Dr. Haq who dedicated much of his time to help with my sample preparation and the biochemistry of α_1-AT to ensure that my project ran smoothly and efficiently.

Table of Contents

List of Figures

List of tables

List of equations

List of abbreviations

Å	Ångström
α_1-AT	Alpha$_1$-antitrypsin
ATD	Arrival Time Distribution
BSA	Bovine Serum Albumin
C-terminus	Carboxyl-terminus
CCS	Collision Cross Section
CID	Collision Induced Dissociation
CIU	Collision Induced Unfolding
CRM	Charge Residue Model
CSD	Charge State Distribution
CX	Cross-linking
Da	Dalton
DC	Direct Current
DDA	Data Dependent Acquisition
DTIMS	Drift Time Ion Mobility Mass Spectrometry
E.coli	*Escherichia coli*
EM	Electron Microscopy
ESI	Electrospray Ionisation
et al.	*et alii* (and others)
ETD	Electron Transfer Dissociation
EHSS	Exact Hard Sphere Scattering
FWHM	Full Width at Half Maximum
GluC	Endoproteinase cleaving from C-terminus to glutamic acid residues
H/DX	Hydrogen/Deuterium Exchange
IEM	Ion Evaporation Model

IM	Ion Mobility
IMMS	Ion Mobility Mass Spectrometry
k	Kilo
L	Litre
LC	Liquid Chromatography
min	Minute
M	Molar (mol/dm^{-3})
m	Milli
MALDI	Matrix-Assisted Laser Desorption Ionisation
MDa	Mega Dalton
MPD	Microchannel Plate Detector
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
m/z	Mass-to-Charge Ratio
N-terminus	Amino-terminus
NMR	Nuclear Magnetic Resonance
OD	Optical Density
PA	Projection Approximation
PMF	Peptide Mass Fingerprinting
PLGS	ProteinLynx Global Server
ppm	Parts Per Million
PTM	Post-Translational Modification
PULSAR	Protein Unfolding for Ligand Stabilisation and Ranking
Q-ToF	Quadrupole Time-of-Flight
RF	Radio Frequency
RCL	Reactive Centre Loop
s	Second

SCX	Strong Cation Exchange
SDS-PAGE	Sodiumdodecyl Sulfate Polyacrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
SRIG	Stacked Ring Ion Guide
T-Wave	Travelling Wave
TWIMS	Travelling Wave Ion Mobility Mass Spectrometry
TJM	Trajectory Model
ToF	Time of Flight
Tris	Tris(2-carboxyethyl)phosphine
V	Voltage
WT	Wild-type
µ	micro

Chapter 1 Introduction to Mass Spectrometry

Mass Spectrometry (MS) is a powerful technique that principally measures the mass-to-charge ratio (*m/z)* of ionised molecules in the gas phase. These charged particles can be manipulated by electric and magnetic fields. Traditionally used to identify compounds smaller than 1000 Da, Mass Spectrometry has now become an indispensable analytical technique applied across many different scientific fields as a biological technique used to probe molecular weight, composition, and abundance of large macromolecular structures. Such examples include the 690 kDa 20S proteasome from *Methanosarcina thermophile*[1–3] and highly ordered virus capsids[4,5].

1.1 Mass Spectrometry - A Brief History

While investigating the transmission of electricity through gases in 1887, Cambridge physicist J.J.Thompson invented the mass spectrometer[6]. He was studying the properties of cathode rays by deflecting them using magnetic and electric fields in a low-pressure tube. The first apparatus was only able to measure the *e/m* (charge-to-mass ratio) of electrons. It took another two years to build an instrument able to indirectly determine the mass of the electron by simultaneously measuring the *e/m* and *e*. For this work, J.J. Thompson was awarded the Nobel Prize in Physics in 1906. Further development of the mass spectrometer was achieved with the help of Thompson's protégé Francis Aston. With this instrument, they were able to measure the masses of ions. Ions were generated by the instrument and passed through parallel magnetic and electric fields. They were subsequently deflected into parabolic trajectories before being detected on a photographic plate. From 1910-1940, further instruments were developed to improve resolving power, which allowed scientists to separate and prove the existence of elemental isotopes.

Since the initial mass spectrometer built by Thompson, the field of mass spectrometry was reserved for the study of elemental physics. Enter Alfred Nier, an electrical engineer turned physicist who saw the potential of mass spectrometry in solving problems across many different scientific fields. He worked with biologists to prepare carbon-13 [7] used as an isotopic label to track specific reactions or metabolic pathways within cells. In addition, he had a profound impact on geology when he helped determine the age of the earth by measuring $^{207}Pb/^{206}Pb$ in the planet's crust[8]. More notably however, was his discovery of ^{235}U, the isotope responsible for slow nuclear fission, which sparked the beginning of the nuclear age.

Despite the commercial availability of mass spectrometers, many remained unsure of what went on inside the instrument. Three chemists at the time, Fred McLafferty, Klaus Biemann and Carl Djerassi, as well as Keith Jennings, set out to understand the fragmentation mechanisms of different classes of organic molecules. Their methodical experiments ultimately allowed chemists to determine the structures of unknown molecules by MS. This laid the foundations for modern biological MS research.

Later, in 1970, Alan Marshall and Melvin Comisarow combined their knowledge of nuclear magnetic resonance (NMR) and ion cyclotron resonance (ICR) respectively to develop Fourier Transform ICR (FTICR MS)[9].

Despite all the improvements made to instrumentation, analysing proteins and large biomolecular structures was challenging. The problem lay with the harsh ionisation methods at the time, namely Chemical ionisation and Electron Impact ionisation, which caused extensive fragmentation and decomposition of larger molecules. Many had tried to ionise proteins with new techniques including fast atom bombardment (FAB), thermospray ionisation, and plasma desorption although none were successful[6]. In 1988, both electrospray ionisation (ESI)[10] and matrix assisted laser desorption ionisation (MALDI)[11] emerged and revolutionised biological MS, both are further explained below. Nonetheless, these techniques are still used to this day and have brought mass spectrometry to the forefront of analytical and structural biology.

1.2 Ionization Methods

Ionization is a process that yields ions from neutral molecules or atoms. Many different types of ionization techniques have been developed but only those commonly used in biological contexts will be discussed here. It is worth noting that these are soft ionization techniques designed to produce gas-phase protein ions. With regards to small molecular structures, other, harsher ionization techniques are employed. These include electron ionization (EI) whereby a beam of electrons is used to ionize the sample or chemical ionization (CI), which involves the reaction of an atom or neutral molecule with an ion.

1.2.1 Electrospray Ionisation

Biological studies using mass spectrometry are almost exclusively carried out using ESI or nanoelectrospray ionization (nESI). This gentle ionization technique has enabled us to analyse large intact macromolecular structures in the gas–phase[12]. ESI generates charged gas-phase ions in the following way: a strong electric field is added to a gold-coated borosilicate capillary; the pressure increases at the tip due to an increase in positive charge; the surface tension of the solvent breaks and forms a Taylor Cone. The droplets become smaller as the solvent evaporates leaving the surface charge density of the droplet to increase until the electric field strength reaches a critical point (Rayleigh Limit) within the charged droplet. Coulombic Expulsion then breaks apart the droplet, leading to smaller droplets. The process repeats until multiply charged analyte molecules are formed[13]. These events are depicted in Figure 1.1.

There are however, certain problems involved with conventional electrospray ionization[14]. From a biological perspective, perhaps the most important reason for the introduction of nESI is the ability to preserve precious biological sample. Given the sensitivity of the instrument, ESI can be extremely wasteful. As a result of μL/min flow rates, the diameters of the droplets formed in conventional ESI measure μm in diameter[15,16]. In nESI the initial droplets exiting the capillary are around 1000 times smaller in diameter[17]. This allows the droplets to fission more quickly allowing them to tolerate higher salt loads[18]. In addition, the ion signal is greater

because the resulting analytes are more highly charged, increasing their detection[14,19–21]. Hence, nanoelectrospray ionization is favoured for mass spectrometric studies of proteins.

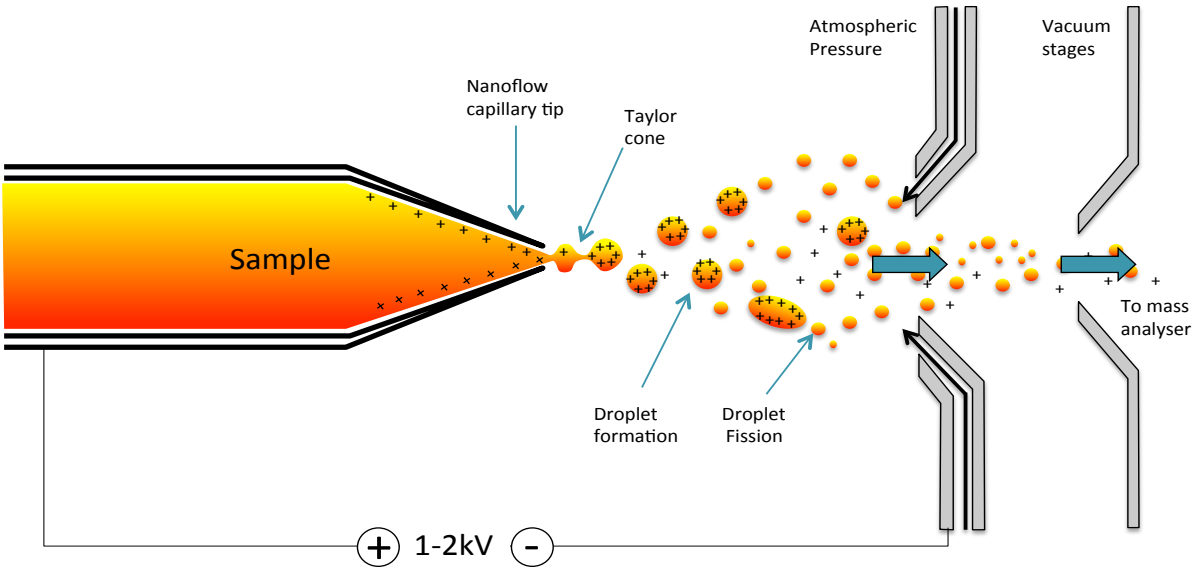


Figure 1.1. Electrospray Ionisation Process. Analyte molecules are added to a charged solvent (orange/red). As a strong electric field is applied to the capillary, the charge accumulates at the tip resulting in a Taylor cone. Eventually, a fine mist of charged droplets form and following several sequences of evaporation and droplet fission, the analyte becomes multiply charged and enters the mass spectrometer.

1.2.1.1 The formation of Gas-Phase Ions from Charged Droplets

The two main mechanisms proposed for the formation of gas-phase ions from highly charged ESI-formed droplets are the charge residue model (CRM) and the ion evaporation model (IEM), (Fig.1.2).

The charge residual model occurs as a result of a series of solvent evaporation and Coulomb fission. Eventually, only one analyte molecule will remain in a charged droplet measuring ~1nm. As the solvent from this last droplet evaporates, the residual charge on the surface of the droplet lands on the analyte molecule, hence giving the name CRM. The original hypothesis for CRM was proposed by Dole *et al.* [22,23] and was further supported by Winger and his group who showed that macromolecules most likely follow this mechanism[24]. From the mass spectra, they determined an empirical correlation between the analyte mass and the observed average charge state.

Fernandez De La Mora[25] probed further and used this equation to calculate the radius of the protein as:

$$\left(\frac{4}{3}\pi R^3\right)\rho N_A = M \qquad \text{Equation 1.1}$$

Where R is the radius of the protein, ρ is the density of water, N_A is Avogardro's constant and M, the molecular mass of the analyte. This assumes that the ultimate charged water droplet would be slightly larger in size than the protein and that the globular protein had the same density as that of water (ρ = 1 g/c.c).

Rather than depositing the charges from the droplet onto the surface of the molecule as in CRM, IEM shows that the radius of the droplet does not go below 10nm before ions are emitted from the droplet, taking charges with them. In other words, past a droplet radius of 10nm, ion emission dominates over Rayleigh fission. Iribarne and Thomson[15] who proposed the mechanism state that the rate constant k_1 for the emission of an ion from a highly charged

droplet surface can be given by the following equation:

$$k_1 = \frac{k_b T}{h} \exp\left(-\frac{\Delta G*}{RT}\right) \qquad \text{Equation 1.2}$$

Where k_b is the Boltzmann constant, T is the temperature, h is the plank constant, R is the gas constant, and $\Delta G*$ is the activation energy. Their experimental evidence suggests that this is often the route of ion production for small inorganic and organic ions.

Another, more recent model has emerged called the Chain Ejection Model (CEM) (Fig.1.2.c). This is a proposed mechanism for the formation of ions in non-polar polymer chains and was determined by molecular dynamic simulations. This occurs when the protein is highly disordered or unfolded, resulting in non-polar residues being solvent accessible. Proteins of this nature are unstable inside the Rayleigh-charged nano-droplet and favourably migrate to the surface of the droplet where part of the chain will be ejected first, followed by the remaining polypeptide chain. Although similar in part to IEM, this mechanism only applies to polymer chains that are partially hydrophobic, disordered and capable of binding excess charge carriers[26,27].

A fourth mechanism called the Fenn's model[10], proposes that the analyte remains neutral within the charge droplet as the charges are spread across the surface. When an analyte such as a protein is intrinsically charged, the neutrality is maintained by a counter ion. Meanwhile, as successive solvent evaporation and coloumbic fission occurs, the surface charge density of the droplet increases. At some point, part of the analyte will encounter the proton-rich surface. These protons then anchor the analyte molecule closer to the surface. Eventually, thermal activation is thought to provide the energy required for the analyte to partially move outside the droplet. The protonated surface and analyte residues repel each other, which ejects the remaining portion of the molecule from the droplet. Analytes with more charged residues will escape the droplet more readily due to an increase in Coulomb repulsion. The degree of charging relies on the shape, size, and orientation of the analyte as well as the variable charge

spacing and electric field at the droplet surface due to successive Coulombic fission events. As the droplet size decreases, the analyte molecule would be more charged. In practise, as more and more charged species are ejected from the droplet, both the number of species and number of charges will decrease in the droplet resulting in the bell-shaped distribution of the ion abundances in the charge state distribution.

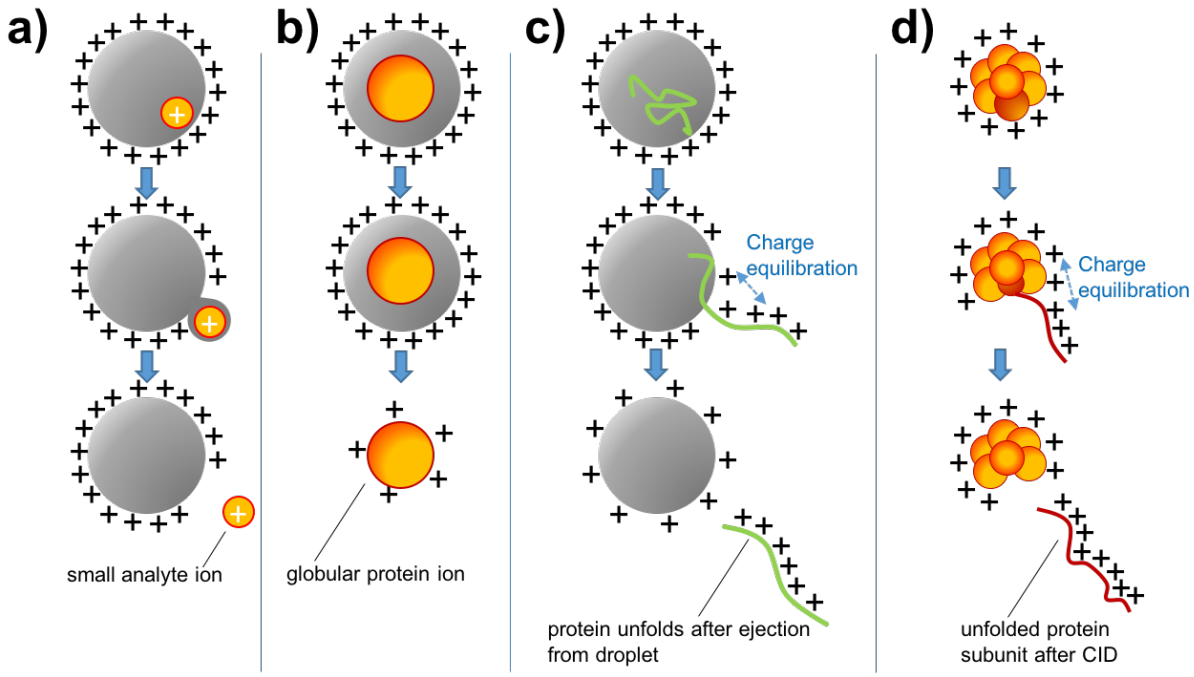


Figure 1.2. Proposed Mechanisms for Ion Formation in ESI. **a)** IEM, the ion is ejected from a charged nanodroplet. **b)** CRM, the charge deposits onto the analyte molecule as the solvent dissolves. **c)** CEM, the unfolded protein is ejected. **d)** CID of a gaseous multiprotein complex. Equlibration of charge is indicated by red dotted arrows in panels **c** and **d**. Figure adapted from Analytical Chemistry[26].

1.2.1.2 Multiply Charged Ions

Regardless of how gas-phase ions are produced in ESI, the spectra always shows multiple peaks for the same protein ion corresponding to its different charge states (Fig.1.3). For each adjacent peak, the charge difference will always be one. Knowing this, the charge states and the molecular masses of any ion can be calculated by solving both these simultaneous equations:

$$m_1 = \frac{(M+n)}{n} \qquad \text{Equation 1.3}$$

$$m_2 = \frac{(M+n+1)}{(n+1)} \qquad \text{Equation 1.4}$$

Where m_1 and m_2 are the m/z ratios of the two adjacent peaks, n is the number of protons associated with the ion, and M the molecular mass. Rearranging for n gives:

$$n = \frac{(m_2-1)}{(m_1-m_2)} \qquad \text{Equation 1.5}$$

With multiply charged ions the m/z ratio of large molecules is lowered, enabling low-range mass analysers to detect them.

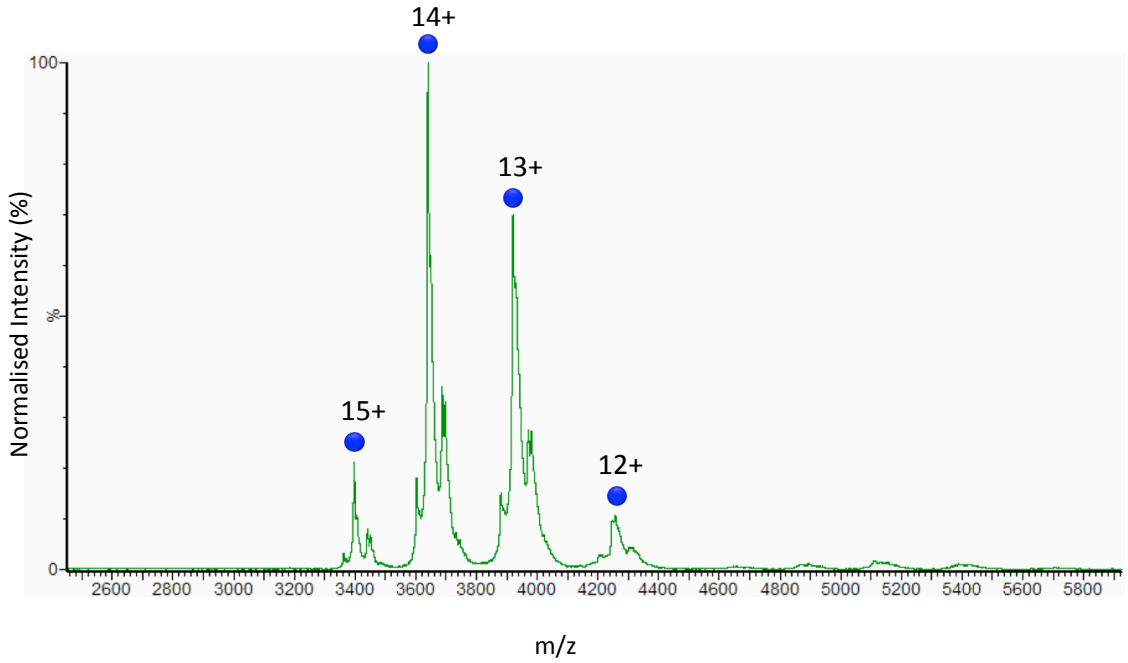


Figure 1.3. ESI Mass Spectrum of a Multiply Charged Protein Ion Under Native Conditions. The native mass spectrum of healthy plasma alpha$_1$-antitrypsin with an experimentally determined mass of 50960 +- 7 Da. The charge states of the monomer are shown for the highest peaks, other peaks refer to the same charge state but different states of glycosylation.

1.3 Ion separation – mass analysers

Mass Analysers separate gas-phase ions according to their mass-to-charge ratios[28]. All mass analysers fall into two categories: scanning or non-scanning analysers. Scanning analysers transmit ions of different masses along a time scale whereas other analysers transmit all ions simultaneously. Analysers can be subdivided based on other properties, for example: ion beam versus ion traps; low versus high kinetic energies; continuous versus pulsed analysis. Over time, mass analysers have improved dramatically and instrument versatility has increased due to the possibility of combining mass analysers in sequence[29].

The performance of a mass analyser is ranked on its mass range limit, scan speed, ion transmission, mass accuracy and resolving power. The mass range is defined as the upper *m/z* limit of the analyser, past this threshold the analyser can no longer measure the ion. The units of mass range are given either as Th or u for an ion carrying one elementary charge.

The scan speed is expressed in mass units per second (u s^{-1}) or mass units per millisecond (u ms^{-1}). This refers to the speed a mass analyser measures the ions over a particular mass range.

Transmission refers to the ratio of the ions reaching the detector to those entering the mass analyser. Often, ions will be lost through sections of the mass analyser called electric lenses, found before and after the analyser. Note that this is not the same as a duty cycle, which is a characteristic of the whole instrument.

Mass accuracy refers to the difference between the experimental *m/z* compared to the theoretical *m/z*, the closer these values, the more accurate the mass analyser. It is often expressed in millimass units (mmu) or parts per million (ppm). Analysers with better mass

accuracy tend to have a higher resolution and tend to be more stable[30].

Lastly, the resolving power is the ability of an analyser to distinguish two ions with very small *m/z* differences. Peaks are considered resolved if the trough between them is at 10% the height of the weaker peak intensity for ICR or magnetic instruments and 50% the height for instruments with quadrupoles and ion traps. Mass analysers with the greatest resolving power are able to distinguish ions with the smallest differences in mass. Resolving power can also be determined for single peaks where the difference in mass is taken across the peak width at 50%, this is called the full width at half maximum (FWHM) (Fig.1.4). High-resolution analysers are considered to have a resolving power above 10000.

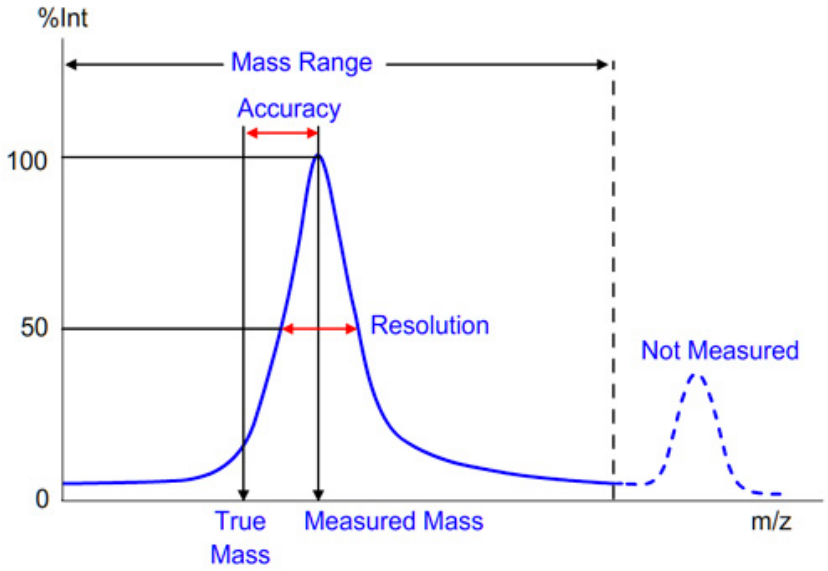


Figure 1.4. Resolution in Mass Spectrometry. The width at half the height of the peak is considered the resolution. The smaller the peak width the better resolved the peak. Accuracy is determined by how close the experimental value is from the true value. Figure taken from[149]

1.3.1 Quadrupole Mass Analyser

The quadrupole mass analyser was developed by W.Paul and H.S. Steinwedel in the 1950s to overcome the difficulties associated with magnetic fields that led to the degradation of resolution[31,32]. As the name suggests, this mass analyser uses alternating quadrupolar electric fields. The quadrupole acts as a mass filter comprising four parallel rods as depicted in Fig.1.5. Each opposite pair of rods has the same applied potential -(U+Vcos(ωt)) and (U+Vcos(ωt)), U is a DC voltage, V is an AC voltage, ω is the frequency and t is time. By changing these applied voltages, one can also change the ion's trajectory down the centre of the four rods.

Positive ions entering the quadrupole will be drawn to a negatively charge rod, if the potential of the rod changes before the ion comes into contact with it, then the ion will change direction and will now be attracted toward the adjacent rod which had also changed potential at the same time and is now negatively charged. If the ion does not discharge onto the rod at any point then it will have a stable trajectory through the instrument and will be 'filtered' from other ions that enter the instrument at the same time.

DC and AC voltages can be tuned for a specific *m/z* by varying the amplitude of the RF voltage or by adjusting the DC/RF ratio[33,34]. This will favour a stable trajectory through the mass analyser of certain mass-to-charge ratios. All other ions will have an unstable trajectory which will result in them being thrown out of their original paths. Alternatively, the entire mass spectrum can be scanned by simultaneously varying the amplitude of the DC and RF voltages.

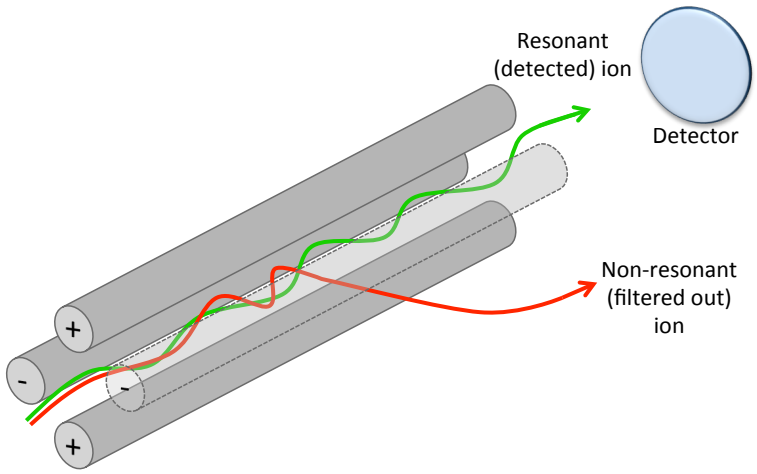


Figure 1.5. Schematic of a Quadrupole Mass Analyser.

1.3.2 Time-of-Flight Mass Analyser

Stephens first described the concept of time-of-flight (ToF) in 1946[35,36] but the instrument was designed and developed into a commercial instrument by W.L.Wiley and I.H.McLaren in 1955.[31] In short, a ToF mass analyser determines the *m/z* ratio of an ion by measuring the time taken for the ion to travel through a field-free region between the source and the detector. Smaller ions reach the detector faster than larger ions.

Linear ToF mass spectrometers (Fig.1.6) accelerate ions upon entering the source by applying a potential difference between an electrode and an extraction grid. All ions acquire the same kinetic energy, therefore they can be distinguished by their varying degrees of velocity. The following process occurs as follows, as an ion with mass m and total charge $q=ze$ is accelerated by a potential V_s its electrical potential energy E_{el} is converted into kinetic energy E_k via the following equation:

$$E_k = \frac{mv^2}{2} = qV_s = zeV_s = E_{el} \qquad \text{Equation 1.6}$$

Therefore, the velocity of the ion leaving the source is given as:

$$v = \left(\frac{2zeV_s}{m}\right)^{1/2} \qquad \text{Equation 1.7}$$

Once accelerated, the ion travels at a constant velocity in a straight line toward the detector. The time t for the ion to travel the distance L before reaching the detector is given as:

$$t = \frac{L}{\left(\frac{2zeV_S}{m}\right)^{1/2}} \qquad \text{Equation 1.8}$$

And rearranged to give:

$$t^2 = \frac{m}{z}\left(\frac{L^2}{2ezV_s}\right) \qquad \text{Equation 1.9}$$

Hence, m/z can be calculated by measuring t^2 if $\left(\frac{L^2}{2ezV_s}\right)$ remains constant. With all other factors being equal, this equation proves that ions of lower mass will reach the detector in a faster time due to differences in velocities.

Linear ToFs (Fig.1.6) however have low resolution due to differences in starting times, initial kinetic energies, starting locations, and initial directions of motions between the ions. These ions are therefore not detected at the same time[37]. Reflectrons (Fig.1.7) aim to improve the resolving power of the instrument by compensating for the initial energy spread. Developed by Mamyrin,[38] the reflectron is an ion mirror that focuses ions of different kinetic energies in time. As the ions hit the reflector they are exposed to a retarding electric field, which causes the ions to reach zero kinetic energy before being expelled from the reflector in the opposing direction. Ions with greater kinetic energies will penetrate the reflectron further, thus allowing the reflector to correct the time-of-flight in order for ions of different kinetic energies to hit the detector at the same time, hence improving resolution.

Due to its simple design, low cost and unlimited mass range the ToF is heavily incorporated into instruments analyzing large macromolecules.

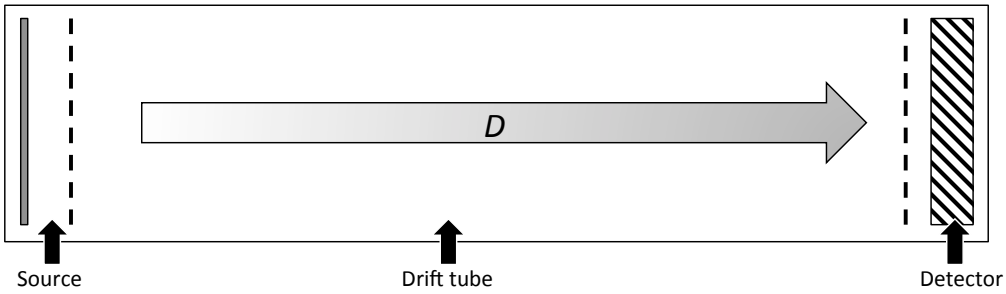


Figure 1.6. Schematic of a Linear Time-of-Flight Mass Analyser. Ions enter at the source and travel through a field free region in the flight tube. The direction of the ions is shown by the grey arrow, ***D*** before hitting the detector.

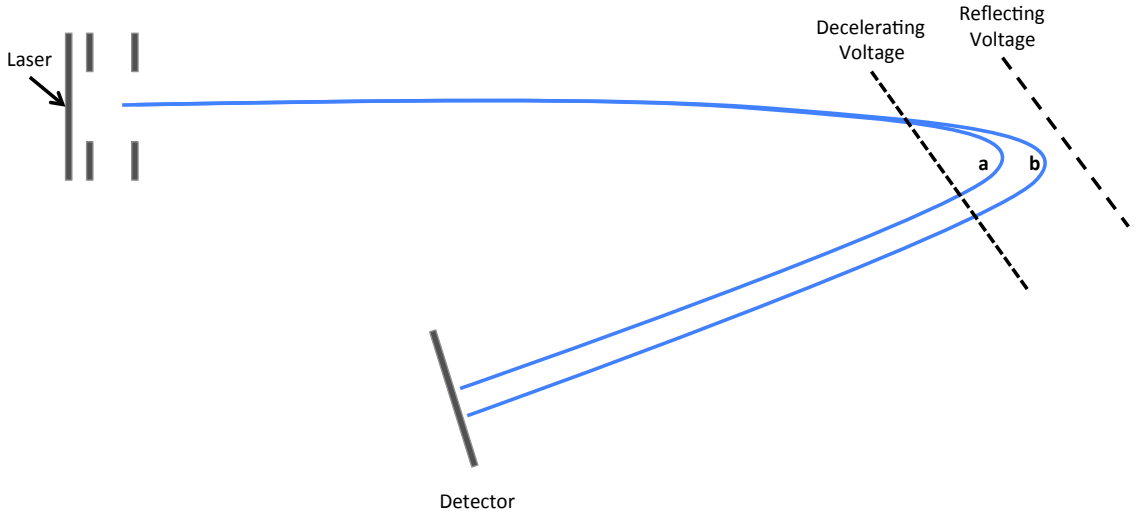


Figure 1.7. Schematic of a Single-Stage Reflectron. Two ions of the same *m/z* but with different kinetic energies (**a** and **b**) will arrive at the same time at the detector as both ions have different trajectories (blue lines). Image adapted from http://what-when-how.com/proteomics/time-of-flight-mass-spectrometry-proteomics/.

1.4 Detectors

Ions that have been separated by *m/z* are then transferred to the detector to generate a mass spectrum. Types of detectors include Faraday cups, photon multipliers, Electro-Optical Ion Detectors, array detectors, photographic plates and electron multipliers. Electron multipliers are the most commonly used ion detector in MS[39], an array of these electron multiplier channels forms the basis of the microchannel plate (MCP) detector (Fig.1.8). In this detector, ions that have travelled through the mass analyser are further accelerated by a conversion dynode (electrode) held at a high potential of 3-30kV opposite to the charge of the ion. Positive or negative ions then strike the conversion dynode and cause the emission of several secondary particles that can be positive ions, negative ions, neutrals or electrons. Secondary particles are converted to electrons at the first dynode and these are then amplified by a cascade effect in the MCP to produce a current. MCP detectors are able to detect many different ions at the same time because for every ion, only a few channels are occupied. If the ion beam is too large however, saturation can occur and a time period is needed for the channels to recover before new signals can be detected.

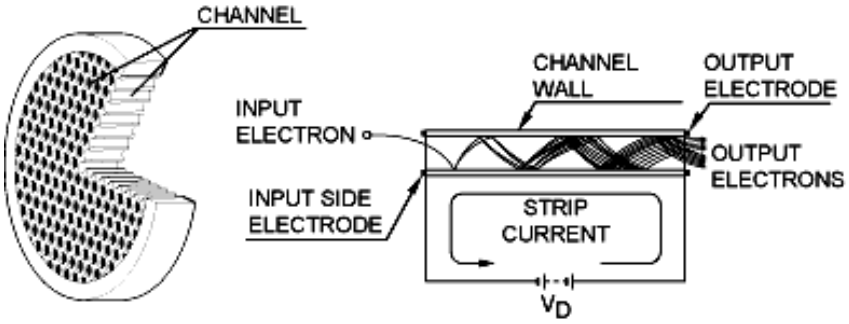


Figure 1.8. Schematic of a Multichannel Plate Detector. Image taken from http://www.dmphotonics.com/MCP_MCPImageIntensifiers/microchannel_plates.htm

1.5 Ion mobility mass spectrometry

Ion mobility (IM) separates ions according to their molecular shape, size, and charge[40–42]. The ions are driven through the drift cell by an electric field and the time taken for the ion to travel through an inert carrier buffer gas is directly proportional to the ion's collision cross section (CCS). We can use this information to determine the differences of CCSs between different proteins and complexes or to investigate the rate of unfolding. Smaller and more compact ions will collide less frequently with the buffer gas than larger more extended ions, resulting in faster arrival times (Fig.1.9). Coupled with MS, IM-MS can differentiate ions in two-dimensions separating them based on their *m/z* ratio and also on their overall structure. When the two techniques were first used in tandem, the ions in question were only simple molecules or isomers[43]. Nowadays, with the invention of ESI, larger more complex samples can be studied[44–49].

IM-MS can probe the structural properties of larger ions by reporting their rotationally averaged collision cross sections (CCS). Jarrold and Clemmer[46] pioneered the use of IMMS to study biological molecules using drift cell technology with a uniform static electric field. The CCS (Ω) can be calculated from the recorded drift time according to the Mason-Schamp equation[50]:

$$\Omega = \frac{3ze}{16N}\left(\frac{2\pi}{\mu k_b T}\right)^{\frac{1}{2}}\frac{1}{K_0} \qquad \text{Equation 1.10}$$

Where e is the elementary charge, z is the charge state of the ion, N the number density of the drift gas, Ω the CCS, μ the reduced mass of the ion/gas pair, K_0 the reduced mobility (under standard conditions), k_b the Boltzmann constant, and T the temperature in kelvin. Note that the relationship between Ω and K_0 only holds true at or below the 'low-field limit' where the ratio between electric field strength and buffer gas density is below $2\text{x}10^{-17}$ V cm^2. CCS calculations

can be compared to other techniques used to solve structures such as NMR or X-ray crystallography[42,51,52]. Initially, the quality of structural information obtained from Drift Time Ion Mobility Mass Spectrometry (DTIMS) suffered from low duty cycle as many ions were being lost in the instrument before striking the detector. Newer generation DTIMS instruments are able to accumulate ions before the drift cell in ion-trapping funnels[50].

Travelling-wave ion mobility spectrometry (TWIMS) comprises a series of ring electrodes called a stacked ring ion guide (SRIG)[53,54]. A travelling voltage wave is applied to the SRIG. Oppositely phased radio-frequency voltages are applied to adjacent electrodes and radially confine the ions. Application of a transient direct current (DC) voltage to each successive electrode in the SRIG propels the ions axially. It is this DC voltage that allows ions to surf and traverse the drift cell axially (Fig.1.10). Optimising ion separation can be achieved by altering the speed and magnitude of the travelling voltage wave. Ions with higher mobility are carried in the troughs of the waves whereas ions with lower mobility will roll over the wave and take longer to traverse the mobility cell. The changing electric field no longer satisfies the Mason-Schamp equation, therefore accurate CCS calculations require a calibration step using analytes with known CCS values[55,56]. Many protocols[57,58] and CCS databases[56,59–61] are available and an increase in software designed to process ATDs directly to CCS means that analysis time is significantly reduced[62].

Theoretical CCS values can be determined by rotationally averaging cross sections from input coordinate files obtained from X-ray crystallography, NMR, or molecular dynamics (MD). The most commonly used models include Exact Hard Sphere Scattering (EHSS), Trajectory Model (TJM), and Projection Approximation (PA). The latter was used for the analysis of the recombinant α_1-AT as the other software was not available for use. PA rotates the three-dimensional structure in all possible orientations in space giving a projected area of the molecule. The algorithm was developed by Mike Bowers at UCSB whilst investigating fullerene

with IMMS[155] and employs Monte Carlo integration to describe the interaction between the buffer gas Helium and an individual atom in the molecular ion. Equation 1.11 explains this interaction:

$$V(R) = \frac{n\varepsilon}{n(3+\gamma)-12(1+\gamma)} \times \left[\frac{12}{n}(1+\gamma)\left(\frac{r}{R}\right)^{n} - 4\gamma\left(\frac{r}{R}\right)^{6} - 3(1-\gamma)\left(\frac{r}{R}\right)^{4}\right] \quad \text{Equation 1.11}$$

Where V is the interaction between an atom in the ion compound and the Helium atom separated by distance R, ε is the depth and r the position of the potential well. The term γ defines the relative contributions of R^{-4} and R^{-6} and is dimensionless. The term R^{-4} describes the induced dipole in the Helium atom as a result of its interaction with the charge of the specific atom in the molecular ion.

Studies have shown that there is a <2% difference between theoretically calculated CCS values and experimentally derived CCS for lysozyme, myoglobin and cytochrome C[53]. The PA method is computationally inexpensive but comparisons with experimental data consistently show that the PA method underestimates the CCS due to its inability to take into account the long range scattering interactions[50,155]. EHSS will take the long range scattering into account but ignores long range interactions between the ion and the buffer gas. Hence, the most accurate but computationally expensive method was TJM. Projection Approximation was still used however in this project as an initial estimate of recombinant α_1-AT's CCS to ensure that we were observing its native conformation. Following these experiments, there is no atomic resolution data for any of the plasma variants and therefore, theoretical calculations are unnecessary.

Scientists often question the validity of studying proteins in the gas-phase and whether they retain their native structures. Several studies support IMMS as a means of studying protein conformation. Studies have shown that experimentally calculated CCS values matched those theoretically calculated from known crystal structures[53]. Clemmer's group published a time-dependent IMMS study which proved that natively folded proteins were preserved in the gas

phase on a sub-50 ms timescale[63], longer than the flight time of an ion through a linear mass spectrometer. Furthermore, the Cooks group[64,65] were able to collect lysozyme from the tobacco mosaic virus following mass spectrometric analysis and found that the lysozyme still exhibited normal activity despite having previously 'flown' through the mass spectrometer. They concluded that lysozyme remained intact throughout its time in the gas-phase. Furthermore, technical skill and a basic understanding of the instrument parameters ensures that the conditions employed are non-denaturing.

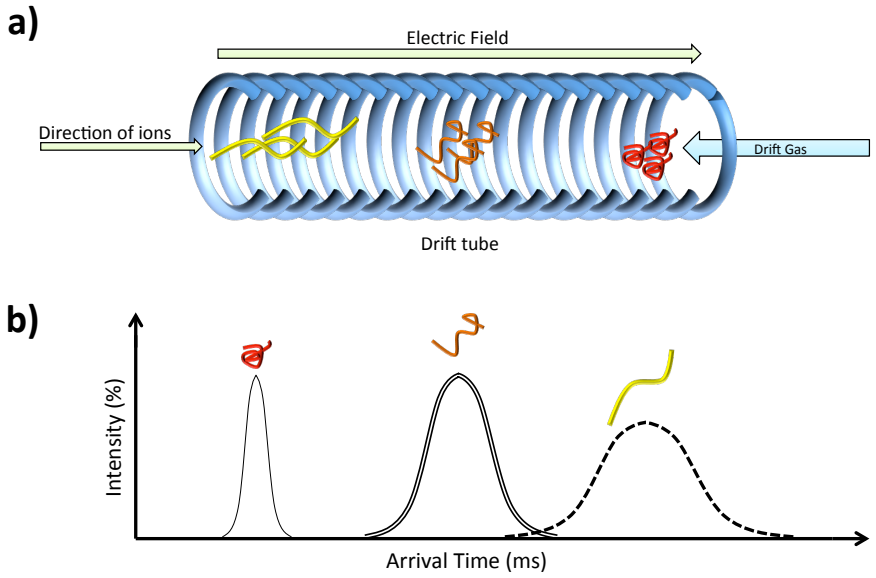


Figure 1.9. Ion Mobility Mass Spectrometry. Ions of the same *m/z* but different sizes or conformations (yellow, orange, or red) can be separated through a drift tube (**a**). The resulting arrival time distributions are shown for each ion in (**b**).

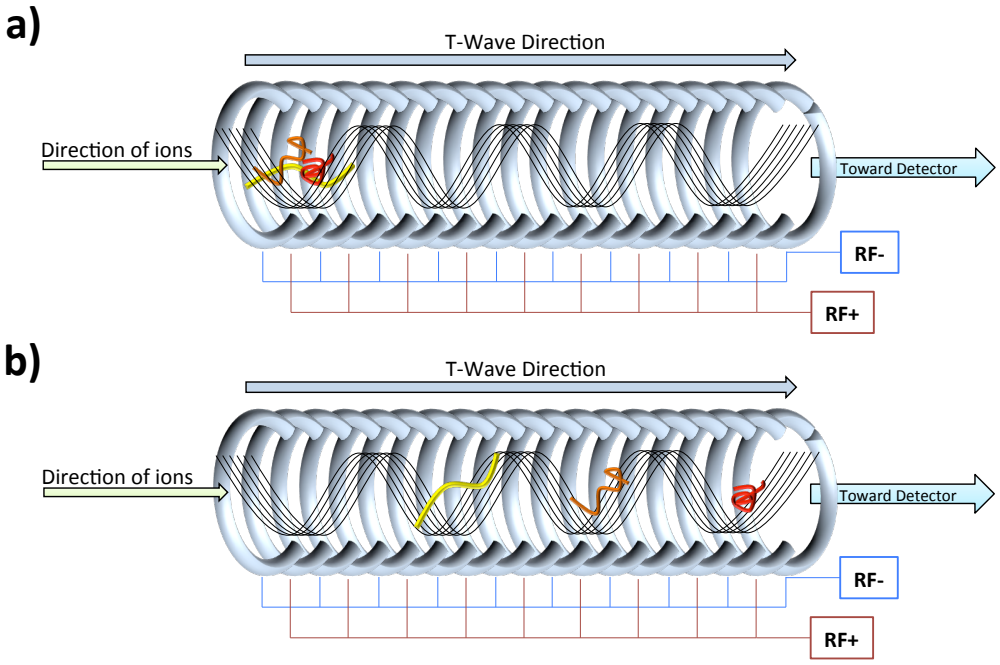


Figure 1.10. Travelling Wave Ion Mobility Mass Spectrometry. Ions are separated by a travelling wave caused by the alternating RF voltages on the SRIGs. The larger ions are not as easily moved by the travelling wave and thus take longer to traverse the drift cell.

1.5.1 Mass Spectrometry & Structural Biology

Mass spectrometry has been shown to retain native macromolecular structures in the gas-phase, thus it has become increasingly important in protein structural analyses of dynamic proteins unable to be crystallised or studied by nuclear magnetic resonance (NMR).

1.5.2 Tandem Mass Spectrometry & Collision Induced Dissociation

Tandem mass spectrometry (MS/MS) involves at least two stages of mass analysis, often combined with a dissociation step, to determine the amino acid profile of a peptide (Fig.1.11). The technique of fragmenting ions in this way was pioneered in the late 1960s by McLafferty and Jennings[66,67] and was termed Collision Induced Dissociation. Most commonly, the first analyser will isolate a precursor ion, which can then be activated to yield product ions that are analysed by the second mass analyser. This requires the collision cell to be placed between both mass analysers in the instrument.

MS/MS can occur either in time, by performing a sequence of events in an ion storage device, or in space, by coupling two physically distinct instruments. I will focus on the latter as this is the instrument available to us in the lab, the Waters SynaptG_2Si (Fig.1.12). Here, the instrument involves a Quadrupole and a ToF. The quadrupole will filter the ion according to its particular *m/z*, the precursor ion is then accelerated into the trap region and allowed to collide with an inert gas (typically Argon). The resulting fragment ions, known as product ions, are then analysed by ToF. Interpretation of the product ion spectrum provides information about the amino acid sequence of the un-fragmented precursor peptide ion that in turn can help identify the specific protein that peptides belonged to.

Collision Induced Dissociation is also used to probe the inter-connectivity of protein subunits within a large non-covalent complex[68,69]. The complex is filtered through the quadrupole and transferred into the collision cell where the internal energy as a result of collisions with the inert gas is dissipated through the complex ion. The increase in energy is just enough to break weak intermolecular bonds resulting in the dissociation of subunits from the

complex ion. The mechanism of dissociation is poorly understood but resulting mass spectra give evidence of asymmetric charge partitioning[70]. A resulting mass spectrum would show the charge-depleted complex at higher *m/z* than its intact state alongside the highly charged dissociated subunit.

Cross-linking has emerged as a new method of determining protein structure using mass spectrometry. By cross linking amino acid residues within the structure, it is possible to then identify which amino acids are close in space when the protein is folded. The proteins are cross-linked in an intact state before being subjected to tryptic digestion and MS/MS[71–75]. The presence of the cross-linker is detected post-data acquisition and the link can be detected. Many different cross-linkers have been developed to improve detection of cross-linked peptides and probe more information regarding protein-protein interactions within the cell. *In vivo* cross linkers can traverse the cell's lipid bilayer, whilst cross-linkers of different lengths enable us to begin assigning potential distances of interaction between closely interacting partners.

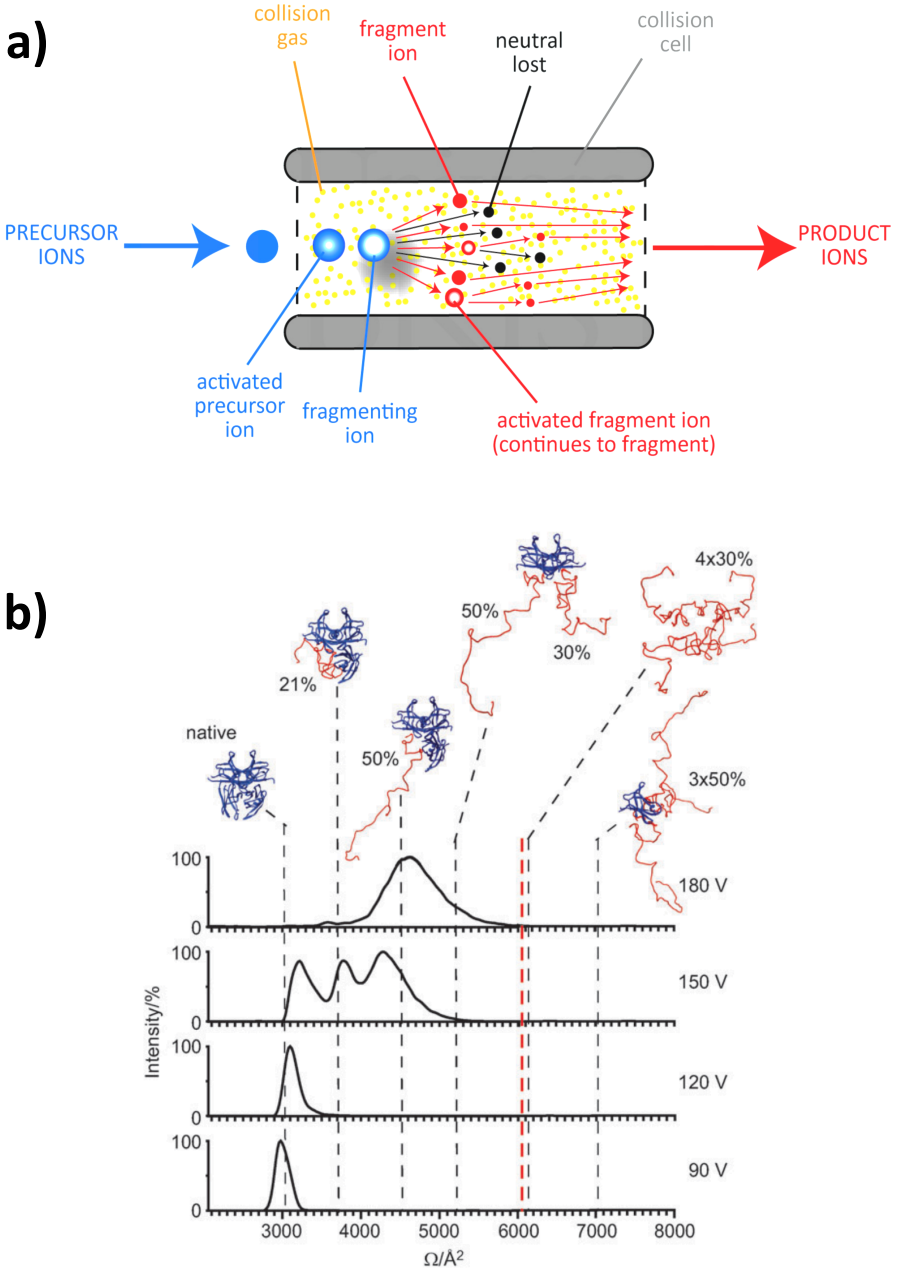


Figure 1.11. a) Collision Induced Dissociation. The precursor ions are shown in blue and are subsequently fragmented into product ions (red) in the collision cell. **b)** A protein is subject to varying degrees of collisional energy, causing it to unfold. The resulting arrival time distribution becomes more extended and contains more than one conformational species. Figure taken from[40].

1.5.3 Collision-Induced Unfolding

Due to the nature of the Travelling Wave ion mobility set-up, it is possible to unfold proteins in the gas phase[76–78]. Proteins ionised by electrospray are activated in the 'trap' region (the T-wave cell located immediately before the ion mobility cell) as they collide with argon gas. These collisions increase their internal energy[79], which induces their unfolding and reveals information about their stability. The stronger the intramolecular forces within the molecule, the higher the energy required to induce unfolding. The resulting ion conformations are then separated by the mobility cell and analysed by ToF-MS (Fig.1.11).

Changing the 'Trap' collision energy voltage (the potential offset between the source ion guide and the trap region) can control the extent of unfolding[77]. This is due to the resulting kinetic energies of the ions entering the trap and making collisions with argon. Voltage limits are protein-dependent however and start at a voltage where no unfolding is observed - the experimental CCS lies within those theoretically calculated for that protein - and ends where no further unfolding occurs.

Ruotolo's group have studied monoclonal antibodies (mAb) by CIU and IM to distinguish isoforms from each other. Their data generated fingerprint plots that track the extent of unfolding as a function of the energy provided during the collisions. [80] They predicted that with this method, it would be easier to analyse biosimilars in a high throughput fashion to develop new mAb biotherapeutics for drug discovery.

CIU can also be useful in determining the dissociation patterns of certain complexes, whereby increasing the voltage in the trap ion guide dissociates large complexes and detects subunits within them[81,82].

Given that CIU is being increasingly employed as a tool in structural biology, the demand for rapid data analysis and interpretation has grown. In return, several publications introduce software modules that are able to process, analyse, compare and classify CIU data (CIUSuite[83],

PULSAR[84], Amphitrite[62]). This should provide a more systematic way of analysing CIU data, by creating CIU fingerprint libraries.

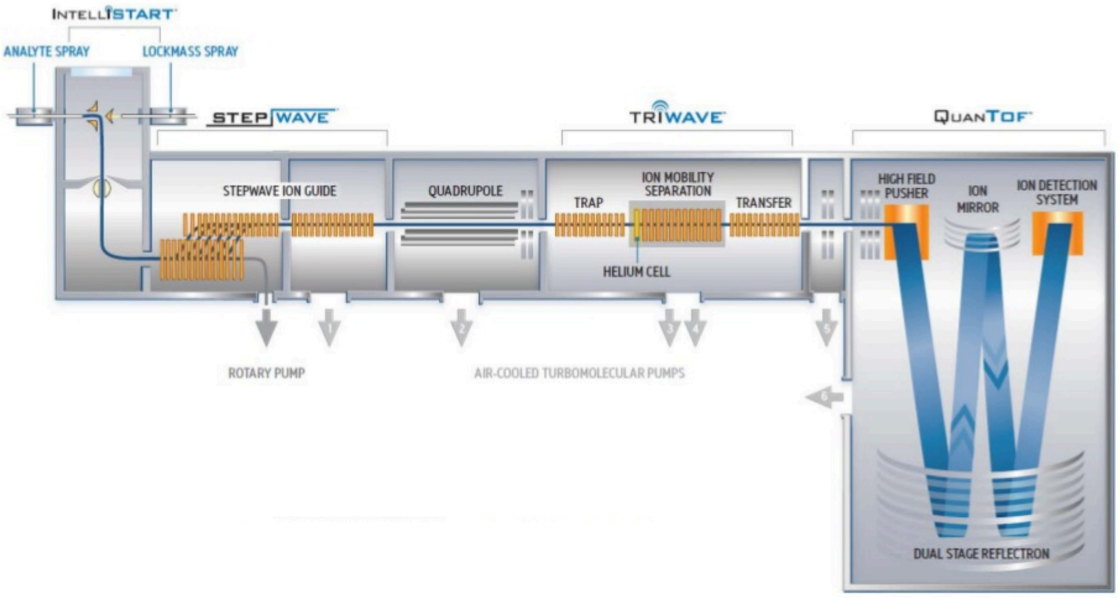


Figure 1.12. Schematic of the SYNAPT G2-Si Q-ToF manufactured by Waters. Image adapted from[150]

1.6 General Introduction to Alpha1-antitrypsin

Alpha1-antitrypsin (α_1-AT) is among the most abundant proteins circulating in human plasma. Belonging to the **ser**ine **p**roteinase **in**hibitor (serpin) superfamily it predominantly targets neutrophil elastase - an enzyme with broad substrate specificity that is secreted by neutrophils at inflammatory sites[85–87]. Mutations in the serpin lead to unstable, misfolded protein intermediates that aggregate as polymers within the endoplasmic reticulum of hepatocyte cells. Errors in folding trigger quality control pathways to prevent secretion of potentially harmful protein conformers. Patients homozygous for the most severe variants have only 10% active α1-AT[88,89]. This hereditary disorder can lead to early onset emphysema, liver hepatitis, and in some severe cases, neo-hepatocellular cancer.

Since its discovery as a genetic disease around half a century ago[90], much has been unveiled about its physiological and pathophysiological nature. This involves polymerisation - the central feature of the disease - which can be related to other conformational diseases such as Parkinson's, Huntington's, Prion diseases and Alzheimer's[91]. Despite the positive future implications following a greater understanding of α_1-AT, there is still uncertainty with regards to the polymerisation pathway and the pathological intermediate. This introduction aims to give a brief overview of the research thus far and offers ideas that may help resolve questions left unanswered.

1.6.1 Physiological properties

1.6.1.1 Native Structure

The initial structure of α_1-AT, solved by Loebermann in 1984[92], comprised three β-sheets (A-C) and nine α-helices (A-I). Over a decade later, Ryu *et al.*[93] crystallized α_1-AT again (Fig.1.13.*a*) which revealed an exposed reactive centre loop (RCL). Studies went on to show that in its active form, α_1-AT exists in a metastable state with the twenty residues in the exposed RCL acting as a pseudo-substrate for the target serine protease (Fig.1.13).

1.6.1.2 Mechanism

At sites of inflammation, the serine protease elastase is released by neutrophils to breakdown connective tissue, enabling blood cells to repair any damage. Alpha$_1$-antitrypsin is purpose-built to prevent the spread of serine protease destructive machinery throughout the body.

The pseudo-substrate binds to the protease active site forming a Michaelis-Menten complex via ester bond formation (Fig.1.13.*b*). This comes as a result of P1-P1' (M358-S359) cleavage which releases the P1' residue. Once this step is complete, the serpin undergoes a significant conformational change of 70Å that enables it to flip the protease from its upper to lower pole. The RCL then inserts as an extra sixth strand in the fourth position of β-sheet A (Fig.1.13.*c*). This transition dramatically stabilises the serpin from its metastable to stable/relaxed conformation, thus increasing its thermal stability from ~50C to >120C[94].

The protease becomes inactive due to disruption of the catalytic triad and loss of the oxyanion hole. α_1-AT is also inactivated and does not undergo further inhibition cycles. As a complex, this suicide molecule binds to the lipoprotein receptor family or to the proteasome and is degraded[95].

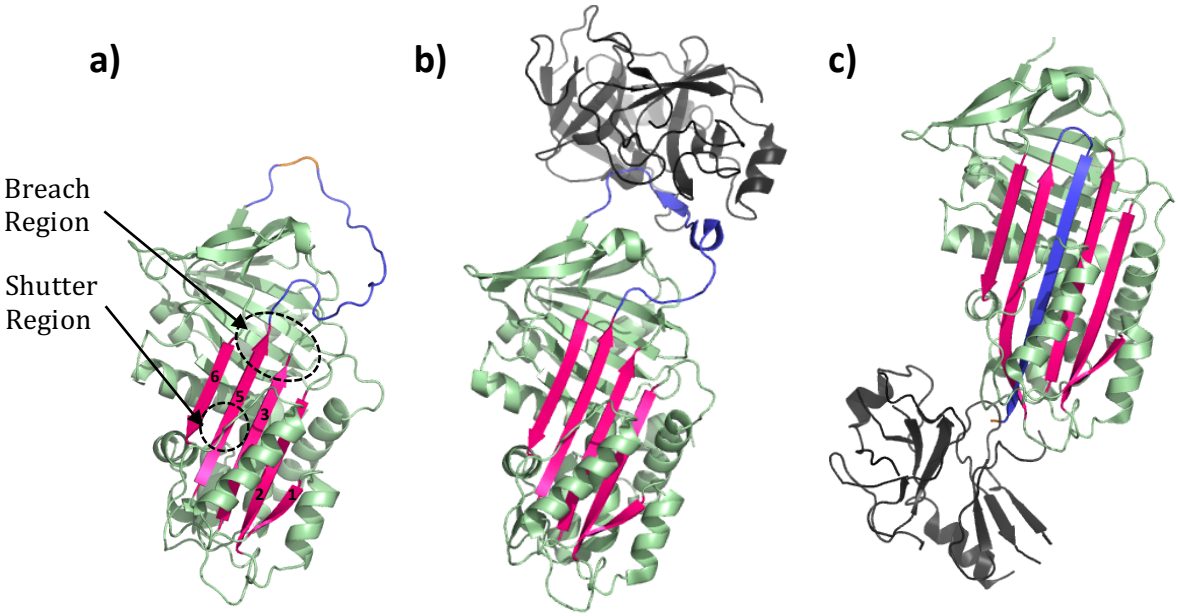


Figure 1.13. Physiological Alpha$_1$-antitrypsin. **a)** The native metastable three-dimensional structure of α_1-Antitrypsin. The colour scheme shows β-sheet A (pink) with numbered strands (1-6), the reactive centre loop/s4a (blue) and the P1-P1'/M358-S359 (orange). The remainder of the molecule (3NE4) is represented in green. **b**) The serpin-enzyme intermediate, the RCL is recognized by the protease (dark grey) and is cleaved at the P1-P1' bond. **c**) The final serpin-enzyme complex.

1.6.1.3 Other Physiological Roles

Hepatocytes produce 70-80% of the total α_1-AT in the body. The remainder is synthesised by macrophages, monocytes, pulmonary alveolar cells, corneal epithelium, and intestinal cells[96]. Unpublished transcriptomic data by Hebe Chen and Bibek Gooptu (Kings College London) suggests that there is expression of α_1-AT in CD4+ and CD8+ T cells. Higashiyama[97] also discovered *de novo* synthesis in adenocarcinomas, highlighting the ubiquitous nature and significance of this protein in health and disease.

In view of these findings, it is clear that α_1-AT does not act solely as an inhibitor. The transcriptomic data explores the possible relationship between vitamin D signalling and α_1-AT expression to enhance production of Interleukin-10 (IL-10) cytokine, an anti-inflammatory often involved in respiratory health. Discovering α_1-AT synthesis in other cells highlights its additional immunomodulatory, antimicrobial and anti-inflammatory properties[96].

1.6.1.4 Physiological Folding

The process by which α_1-AT nascent chains fold into metastable native protein (M) is not fully understood. It might be the case that, given the protein is a large monomer, its native state is a predominant structure within an ensemble of similar conformations. Chaperones are not needed for α_1-AT folding, leaving scientists unsure how the nascent polypeptide chain forms a quaternary structure. Several hypothesised mechanisms for the native structure folding include initial secondary structure formation followed by diffusion-collision; nucleation followed by growth; hydrophobic collapse into a molten globule intermediate; or, a novel folding pathway. Dolmer and Gettins[98] have tried to answer this question using Circular Dichroism (CD), Nuclear Magnetic Resonance (NMR) and native gel (PAGE) using modified cysteine α_1-AT fragments. They focused around the idea that C-terminal extreme residues are implicated in the native to latent transition. Thus, these same residues were responsible for initial folding of the protein

into a metastable state. From their experiments, they concluded that the s5A inserts first, followed by s1C and s4B, s5B and s4A respectively.

1.6.2 Pathophysiology

1.6.2.1 Pathological Alpha1-antitrypsin

It comes as no surprise that a metastable native protein is vulnerable to kinetic traps or disease states due to mutations. Figure 1.14 gives an overview of the possible routes to aggregation following mutations in the shutter region and/or at the base of the RCL.

To date, over 90 genetic variants of human α_1-AT have been determined. Of those, perhaps the most common and characterised of the deficiency variants are Z (Glu342Lys) and S (Glu264Val). The mean gene frequency of Z homozygosity is 0.026 in northwest Europe whilst S has a mean gene frequency of 0.0564 and is more highly distributed in southern Europe[99]. Severity of disease positively correlates with the variant's ability to polymerise. ZZ Homozygotes have 10-15% α_1-AT in their plasma compared to MM.

Z α1-AT polymerises spontaneously at physiological conditions to yield long polymers that aggregate within the ER of hepatocytes as inclusion bodies that are visible by electron microscopy[100]. These polymeric forms are considered responsible for the early onset arrival of liver cirrhosis, panniculitis, neonatal hepatitis and hepatocellular cancer[101–104]. At the same time, low α_1-AT plasma concentrations lead to unregulated neutrophil elastase activity that results in the degradation of alveolar tissue causing panlobular emphysema. Parmar[105] discovered that interstitial polymers acted as neutrophil chemoattractants that stimulated myeloperoxidase release, neutrophil adhesion and degranulation, further exacerbating inflammation (Fig.1.15).

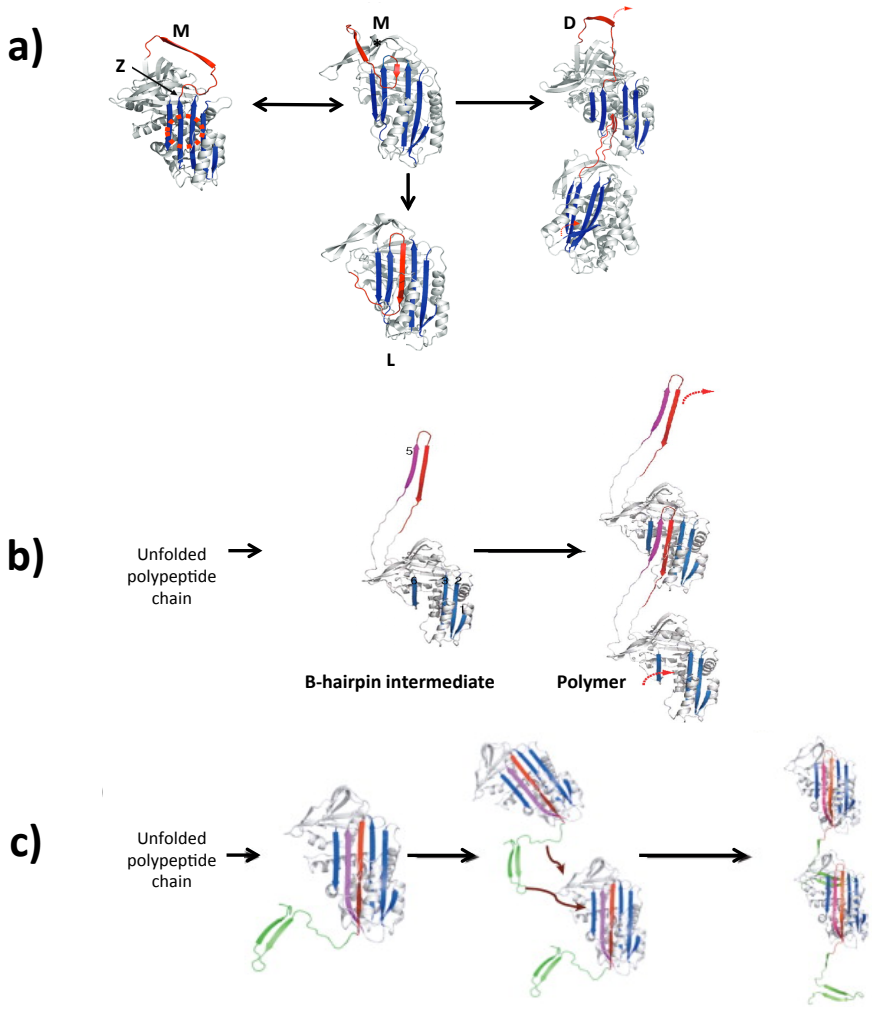


Figure 1.14. Current a_1-Antitrypsin Polymerisation Pathways. **a**) β-sheet A loop insertion. Native α_1-AT (M) is shown on the far left with functional elements β-sheet A (blue) and RCL (red). Mutations may arise in the RCL (Glu342Lys, arrowed) or in the shutter domain region (circled red), which yields an unstable intermediate (M*) by opening the β-sheet A, favouring partial loop insertion. Polymerisation proceeds as β-sheet A accepts a loop from another α_1-AT molecule to initially form a dimer (D). Alternatively, the RCL self-inserts to form an inactive latent (L) conformer. **b**) β-hairpin linkage. Model extrapolated from dimers of antithrombin[109], the two β-strands are formed from s5A (purple) and RCL/s4A (ref) and insert together in the extended β-sheet A of another β-hairpin intermediate. **c**) Triple strand linkage model. Yamasaki *et al.*[151] later solved an α_1-AT crystal structure and proposed this mechanism comprimsing three C-terminal β-strands (sC1, sB4, and sB5; green) in the intermolecular linkage. Figure adapted from[152].

1.6.2.2 Z Polymer Formation

Specifically, the Z variant is retained in the hepatocytes as ordered polymers in the Endoplasmic Reticulum (PAS positive and diastase resistant experiments). It is thought that 70% of Z α_1-AT is degraded by ERAD, 15% folds and is secreted and 10% forms polymers. Some polymers are degraded by autophagy but some remain in inclusion bodies. Z α_1-AT also activates NK KB by a calcium-mediated pathway independent of the Unfolded Protein Response (UPR), termed the ER overload response. ER overload results in the release of IL-6 and IL-8, both mediators of the inflammatory response. Studies have shown that cells expressing Z display a more prominent UPR when stressed with a 2^{nd} hit such as glucose depletion or the accumulation of other misfolded proteins[106]. This confirms autophagy as a specific response to the accumulation of Z.

Polymerisation is a key feature of α_1-AT deficiency though an accurate depiction of the polymerisation pathway is still lacking. Current models, though still widely disputed, are outlined and depicted in Figure 1.14.

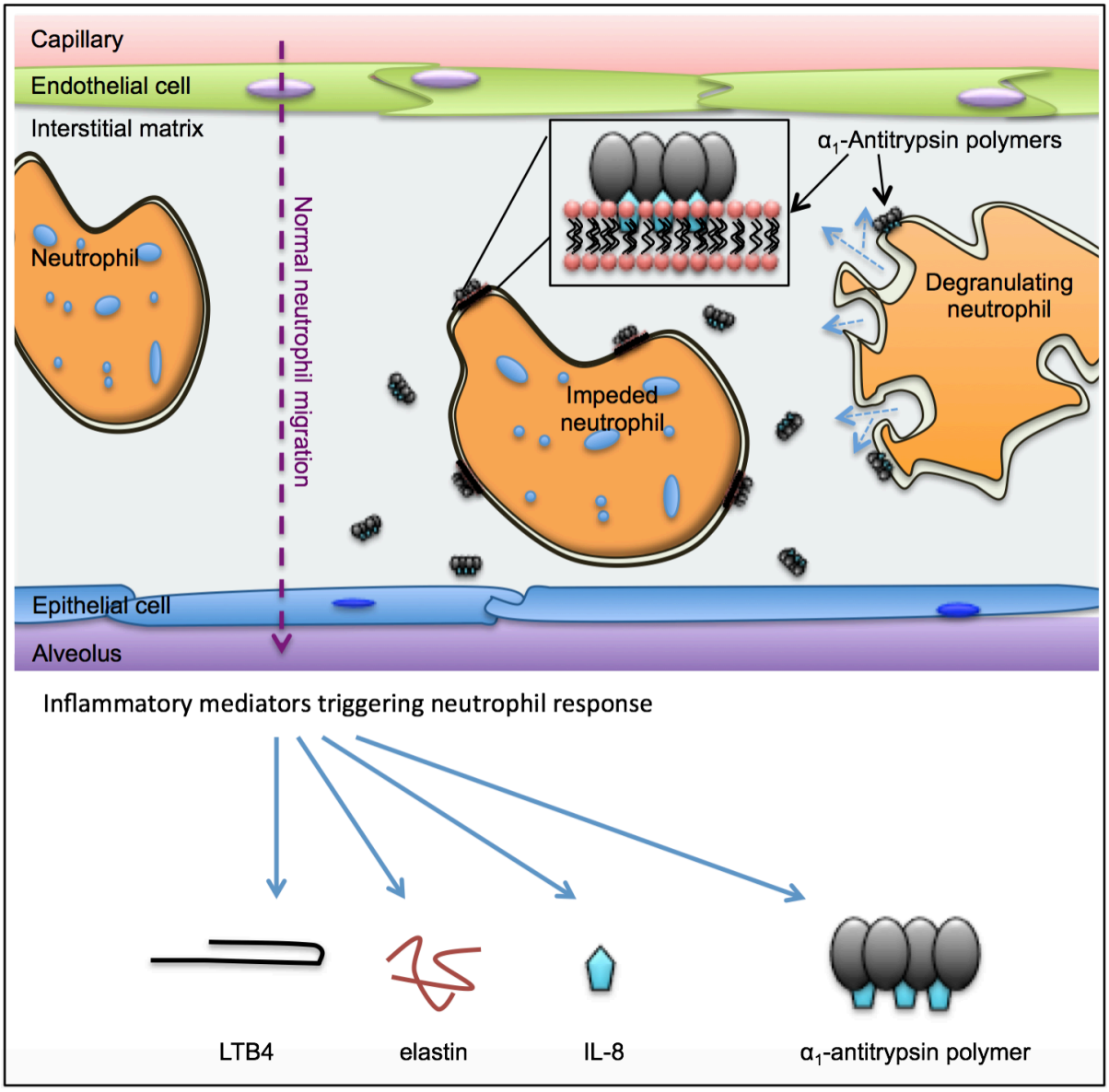


Figure 1.15. The Events Underlying Neutrophil Degranulation in the Gas-Exchanging Membrane of the Lung. Polymerised plasma α_1-AT migrates through the endothelial cells (green) and enters the interstitial matrix, binding to interleukin-8 (cyan). Figure adapted from (Gooptu *et al.*, 2009).

1.6.2.3 β-sheet A Reaction Loop Insertion

Initially proposed by[91,94,100,107,108] this model was widely accepted for nearly two decades. Via this mechanism, the Glu342Lys mutation located at the base of the RCL directs folding to an unstable intermediate (M*). M* then undergoes β-sheet A expansion that opens the s4a position in the domain. M* is primed for intermolecular insertion of an exogenous RCL loop from a nearby Z molecule, thus forming ordered polymers (Fig.1.14.*a*).

1.6.2.4 Polymerisation Proceeds Via a β-hairpin Intermediate

A closed dimer crystal structure of antithrombin and a trimer of α_1-AT disulphide mutant proposed by[109], later contradicted the above model. Given that serpin polymers can readily propagate under native conditions, they hypothesise that the M* state should resemble the monomeric component of the polymer. When they subjected α_1-AT to limited proteolysis their results suggested that s1C and the loop connecting strand 5 and helix I are the first two elements to fold. The authors also conclude from the data that the free-energy from β-sheet A completion is used to unfold helix I and the following coiled region. They believed that this unfolding event led to the exposure of a 30-residue linker, which might be responsible for the formation of linear polymers via lateral association of the monomeric proteins (Fig.1.14.*b*).

Huntington and Whisstock[110] investigated the physical limits of serpin 'loop-sheet' polymers by creating models of alpha$_1$-antitrypsin pentamers (Fig.1.16) but found that these had a morphology entirely distinct from that found in human tissue[111] (Fig.1.17.*a*). Huntington's models were more constrained, lacking the physical properties of serpin polymers formed either *in vitro* or *in vivo*. Thus, they concluded that large scale domain swapping underlies serpin polymerisation but acknowledged that the precise mechanism will depend on the serpin, its mutation, and the conditions used for induction. Naturally, many groups[112] aimed to test the β-hairpin model claiming that under their conditions, the postulated intermediate does not withhold and so the debate continues.

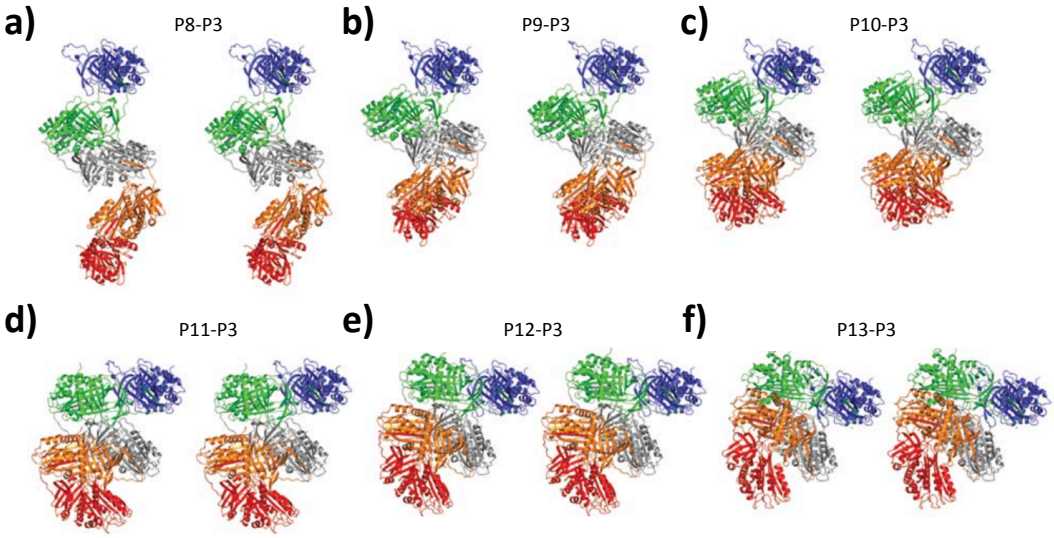


Figure 1.16. Loop-sheet Pentamer Stereo Representations. **a**) shows flexibility of the P8-P3 pentamer formed by loop-sheet A insertion but as the molecules continue to insert toward the RCL in the N-terminal direction (**b**-**f**) flexibility decreases as the polymer becomes more compact. Each monomer is shown in different colours. Figure adapted from[110].

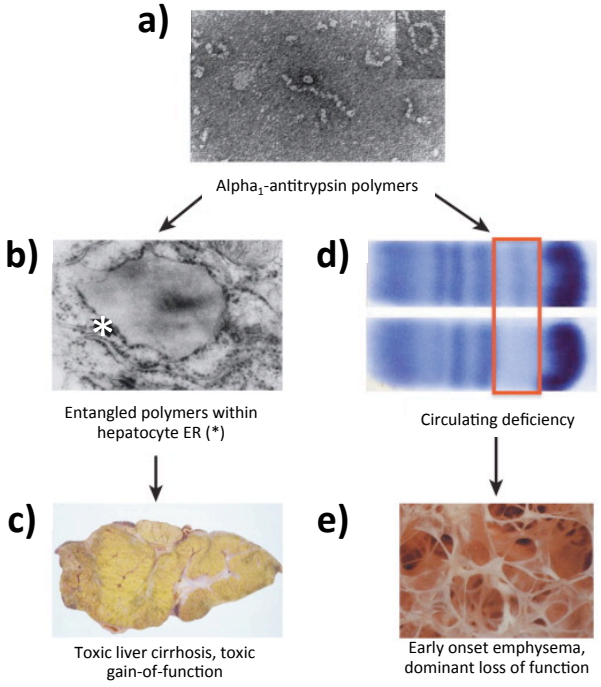


Figure 1.17. **a**) Electron Microscopy Images of Polymers Purified from Plasma. **b**) Polymers accumulating in the endoplasmic reticulum of hepatoctyes. The inability to secrete accumulated **α_1**-AT leads to hepatic cirrhosis in the ZZ homozygote (**c**). Serum tests reveal plasma deficiency (**d**). **α_1**-AT migration is shown in the red box. Top lane; normal control, lower lane; patient with **α_1**-AT deficiency, which leads to early onset emphysema (**e**) from the uncontrolled activity of proteases on the tissue.

1.6.2.5 β-sheet C Insertion

C-sheet insertion was proposed before the high-resolution crystal structure of α_1-AT was solved. Bottomley *et al.*[113] determined that polymerisation might occur by loop A- and C-sheet mechanisms depending on the *in vitro* conditions. Lomas also proposes C-sheet mechanism of polymerisation in the M_{malton} alpha1 antitrypsin variant[114]. From their data, Lomas concluded that the polymers result from the RCL of one molecule to the β-pleated sheet of another. Loop-C sheet-linked polymers are thought to be formed from over-insertion of the RCL in the protease inhibitor, which in turn spontaneously releases the s1C through conformational changes.

Given that the protein most likely exists as an ensemble of particular conformations in equilibrium, there is sufficient data to believe that treatment of these polymers *in vitro* might significantly alter conformation. Shifts in the natural equilibrium potentially favours different conformations/intermediate states, which in turn favour a host of different polymerisation folding pathways. Thus, all the above models might be possible, if only under their respective *in vitro* conditions. More recently, the analysis of conformational disease is seen as more representative under physiological conditions using minimally perturbing techniques and mild disease mutants[112]. The researchers suggest using a slowly polymerising variant such as K154N α_1-AT to allow enough time to study the solution behaviour of proteins. Their data, acquired under more physiological conditions, suggests that there is little unfolding in the intermediate, and disproves the domain-swapped models.

1.6.2.6 Polymer Characterisation

Polymer length varies between alleles and their surrounding environment eg. S polymers comprise of 15-17 α_1-AT monomers whilst M_{malton} only has 3-5. Other than hypotheses, data fail to provide insight into this phenomenon.

1.6.2.7 Pathological Intermediate

Establishing the structure of the pathological intermediate (M*) would further our understanding of polymerisation and would enable more targeted therapeutic strategies to prevent it. Given the inherent flexible nature of the protein, obtaining the intermediate structure using crystallography is challenging. Many groups have attempted this but none have convincingly done so. Ion mobility mass spectrometry (IMMS) was employed to study α_1-AT monomer, dimer and polymer[115]. The area of the CCS between that of the monomer and the dimer was thought to be the intermediate, that complemented the β-sheet A loop insertion model. Arguably, this could simply be a high charge state monomer in a more extended conformation. With the development of charge reducing agents[116], it seems more appropriate to conduct these experiments where the proteins are in more native conformations in order to validate conclusions.

1.6.3 Therapeutic Strategies

There are currently no clinically effective methods of treatment or cure for α_1-AT deficiency. Following decades of studies aiming to understand the interplay between protein folding and QC mechanisms. However, we have now reached a point where we can start to develop new therapeutic means of targeting it and its symptoms.

Patients deficient in alpha$_1$-antitrypsin currently undergo augmentation therapy intravenously on a weekly basis. This aims to restore circulating α_1-AT above the 11μM threshold and aims to slow down the progression of lung disease. As well as showing little-to-no clinical effectiveness, this choice of treatment is expensive and impractical. An improvement on this type of treatment involves the inhalation of industrially produced recombinant α_1-AT. In addition, patients suffering from emphysema can be treated with bronchodilators and inhaling steroids and oxygen when needed.

Gene therapy is being considered using non-viral gene transfer, recombinant adenovirus (rAd), recombinant adeno-associated virus (rAAV), and gammaretrovirus vectors[117]. As such, one phase II trial (using rAAV1 as a vector) is underway. Despite showing positive results, it is off its target range by up to 97%, highlighting the need to increase the dose of the vector.

More recently, induced pluripotent stem cells (iPSCs) have been employed to generate hepatocyte-like cells by exposing them to growth factors in order to recapitulate features of α1-AT deficiency with polymers forming inclusion bodies[118]. Consequently, one is able to acquire an infinite supply of human-derived cells expressing Z. This is a promising alternative given that scientists are still unable to generate primary human hepatocytes expressing this variant. With these model cells, scientists aim to create genetic lesions in order to stimulate gene editing, essentially correcting them into normal functioning cells. The next step would involve a safe method of introducing these corrected cells into patients to treat their deficiency.

Non-augmentation therapy is also another possible treatment route. This involves the use of multiple targets to prevent polymerisation. The strategy aims to manipulate autophagy, boost ERAD and use chemical chaperones. This aims to increase polymer clearance and encourage correct protein folding. Studies have already shown that autophagy-enhancing drugs (carbamazepine, rapamycin, transcription factor EB) reduced α_1-AT burden and hepatic fibrosis in mice[119]. It might also be useful to boost ERAD activity to increase misfolded α_1-AT variant degradation.

Small molecules that block/fill the s4A reactive centre loop insertion site with RCL analogous peptides are another suggestion. Tetrapeptide (TTAI) binds successfully to α_1-AT to prevent polymerisation of the protein[120] but cannot be used therapeutically as it is quickly degraded by proteases before having a chance to act on α_1-AT. Further studies, scanning large-scale libraries are being undertaken to determine other small molecules, which may prevent

polymerisation. The ideal binding molecule would encourage protease inhibitor activity but would prevent polymerisation of the molecule or formation of the pathological intermediate state, thus acting as a chemical chaperone.

Silencing the mutation using small interfering RNA therapy (siRNA) is also being considered[121]. Here, the siRNA constructs were placed in rAAV vectors, which were then packed into AAV8 capsids before being added to mice livers expressing human Z variant. Remarkably, a significant decrease in Z monomer and polymer was observed within weeks of injection. As a result, the AAV8-3X-siRNA vector may fall into the pipeline of treatments for α_1-AT deficiency in the near future.

A more recent study[122] involving the use of antisense oligonucleotide treatment on mice ceased liver disease progression after short-term treatment, reversed liver disease after long-term treatment and prevented liver disease in young animals. Such positive results deserve further investigation and may prove successful at treating deficiency.

1.6.4 Implications and Future Considerations

Although the suggestions above seem theoretically logical there are several hurdles remaining. Gene therapy has huge ethical implications. Severe side effects from other clinical trials[123] have hampered its reputation as a treatment method. Additionally, off-target results thus far might not simply be solved by an increase in dose. iPSCs are promising but, ideally, researchers would be able to express α1-AT variants in human derived hepatocytes rather than using an artificial system.

The use of multiple targets to prevent polymerisation sounds more like a shotgun approach to treating α_1-AT deficiency. Widespread attempts to prevent polymerisation might work but could be costly, severely disrupting the system's homeostatic state. RNAi silencing is still very

much in its infancy but could prove effective. Methods of delivery however would still prove problematic due to the species barrier. If small molecules successfully inhibit polymerisation they must be specific and selective to avoid systemic problems. It is worth noting that other polymerogenic variants (Siiyama, S, M_{malton}) have not been characterised as extensively as Z. For s4A inhibitors, there is a balance between inhibiting the greatest number of polymerogenic variants and molecular interaction with other proteins in the body, causing potential side-effects. It might be better therefore to focus on other sites specific only to polymerogenic serpin molecules that stabilise when misfolded and dissociate when correctly folded to yield active α_1-AT.

1.7 Aims and objectives

The aims of this project are to investigate α_1-AT with mass spectrometry and ion mobility to determine whether its conformational behaviour and its unfolding allude to its behaviour in health and disease.

Recombinant α_1-AT will first be studied to investigate whether ion mobility is capable of determining any differences between the proteins' CCS and their unfolding behaviour. Two stabilizing variants (G117F and T114F) and one slowly polymerizing variant (K154N) will be compared to wild-type recombinant α_1-AT and to previous biochemical observations made by other groups in the field.

Once a good understanding of the recombinant proteins is achieved, plasma α_1-AT variants (M, B, Z, and S) from healthy and α_1-AT deficient patients will be studied to investigate the effect glycosylation has on the conformational ensemble of the protein and its ability to unfold in the gas-phase. Any differences between the variants or the healthy variant might help us understand the potential pathway leading to aggregation.

Following the study of the plasma monomers, ion mobility mass spectrometry will experimentally determine the CCS of the dimer which can then be compared to three potential models of polymerization derived from the crystal structures. Intact polymers will also be analysed to determine their complexity and perhaps give us an idea of the type of polymers being formed (tight aggregates vs. long polymer chains). In addition, CID of the dimer would shed light on the dissociation of the monomers from each other, providing an insight into the type of non-covalent linkage within the polymer.

Chapter 2 Materials and Methods

2.1 Reagents

All mass spectrometric measurements were performed with 200mM Ammonium acetate, freshly prepared on the day of acquisition. All proteins were stored in PBS 5% glycerol unless stated otherwise.

2.2 Protein Purification

2.2.1 Purification of recombinant α_1-antitrypsin

Recombinant α_1-AT was purified using standard protocol as described below.

2.2.1.1 Protein expression

Escherichia coli (strain XL-1 blue) cells were transformed with vector pQE31 cDNA coding for a hexahistidine-tagged alpha$_1$-antitrypsin. The bacterial cells were then grown, induced and harvested by incubating the inoculated media at 37°C at 200-250rpm for 2.5hrs before the temperature was lowered to 30°C for a further 30mins or until the optical density of the culture relative to 1 mL of 2xTY control was 0.6-0.9. Protein expression was induced by adding 5mM Isopropyl β-D-thiogalactopyranoside (IPTG) and left to incubate for a further 4 hours at 30°C. The cells were then centrifuged into a pellet at 6000G for 15mins and subsequently snap-frozen in liquid nitrogen for storage at -80°C.

2.2.1.2 Protein purification

Thawed cells were re-suspended in an equal volume of Ni-seph Buffer A (20mM sodium phosphate pH8, 20mM imidazole, 0.5M NaCl) and sonicated for 2-4 minutes at pulse ~40 before passing through a C3 homogeniser (Avestin) 2-3 times. The lysate was spun at 18000 rpm, 4°C, 30mins in a JA 25.5 rotor. The Ni-seph column was recharged with 5 column volumes Buffer A , 5 column volumes of Buffer B (20mM sodium phosphate pH8, 200mM imidazole, 0.5M NaCl)

and 5-10 column volumes Buffer A. AKTA (GE Healthcare LifeSciences) prime lines were washed with 10-20mL Buffer B followed by an equal amount of Buffer A. Supernatant was added to the column using a peristaltic pump and transferred to the AKTA. The column was then washed with Buffer A to baseline and eluted with 40mL (gradient 0-100%) Buffer B. Fractions were collected and run on SDS gels (ThermoFisher Scientific). The column was further cleaned with 5-10 volumes of Buffer B and 5-10 volumes Buffer A and stored at 4°C for further use. Fractions with α_1-AT were pooled and dialysed overnight in 5L Q-seph Buffer A (10mM Tris/HCl pH7.4, 5mM EDTA, 1mM βME) followed by a further 2 hour dialysis in fresh buffer.

Next, the Q-sepharose column was recharged with 5-10 column volumes high salt wash (10mM Tris/HCl pH 7.4, 2.5M NaCl, 5mM EDTA, 1mM βME) and 5-10 volumes Q-seph buffer A. The sample was loaded and washed to baseline with 15-20 volumes Q-seph Buffer A. A 0-100% gradient of 100mL Q-seph Buffer B (10mM Tris/HCl pH7.4, 500mM NaCl, 5mM EDTA, 1mM βME) was added to the column to elute the proteins and fractions were collected. SDS PAGE gels were run to determine which fractions contained alpha$_1$-antitrypsin.

2.2.2 Purification of plasma α_1-antitrypsin

Patient plasma (M, Z, S and B) was obtained from the London Antitrypsin service (John Hurst, Bibek Gooptu, David Lomas) and prepared in the UCL Respiratory Department with the help of Imran Haq.

300mL plasma was centrifuged (eppendorf 5810) for 10 minutes at 4000 rpm, the supernatant was removed and re-spun prior to filtering with cellulose acetate filters. An Alpha$_1$ select column (GE Healthcare) was initially washed with Buffer 1 (20mM PBS, pH 2.0) before adding Buffer 2 (20mM Tris, 150mM NaCl, pH 7.4). The filtered supernatant was then loaded onto the column at a slower rate and the flowthrough fractions were collected. The column was then washed with Buffer 2 for a further 5 column volumes before eluting with Buffer 3 (20mM Tris, 2M $MgCl_2$, pH 7.4). Column fractions were run on an SDS-PAGE gel (Novex by life

technologies) (Fig.2.2). Fractions containing alpha 1 antitrypsin were pooled and dialysed overnight at 4°C in 5 L 20mM Tris, pH 8.0. Dialysis was repeated in fresh buffer for a further 4 hours. The Hi TRAP Q column was connected to the AKTA Prime plus (GE Healthcare) and cleaned with 50 mL Buffer 4 (PBS, 5% sucrose, pH 7.4) followed by 50 mL Buffer 3. Dialysed sample was loaded at 4 mL/min followed by a further 30 mL buffer 3. Column was eluted with Buffer 4 at 4mL/min and 4mL fractions were collected. Appropriate fractions were run on SDS- and Native gels (Fig.2.3). The concentration of pooled fractions was measured at 1.88mg/mL determined by NanoDrop (ThermoFisher Scientific) and the sample further buffer changed into PBS 5% Glycerol before storing at -80°C.

2.3 Preparation of α1-antitrypsin conformers

2.3.1 Alpha1-antitrypsin polymers

Polymers were induced by incubating plasma monomer from patients expressing wild-type, B, S, or Z. Once purified using the protocol described above, the proteins were incubated at 60 °C for 18 hours before they were gel filtrated. The fractions used in our analysis are shown in Fig.2.1 alongside the resultant native PAGE gel Fig.2.3. The samples were then dialysed in 2 x 5L 200 mM ammonium acetate for 48 hours at 5°C. The samples were directly added to the borosilicate capillary (Harvard Apparatus) and analysed by mass spectrometry.

2.3.2 Proteolytically cleaved α_1-antitrypsin

Plasma M as prepared above was buffer exchanged into 50mM ammonium bicarbonate pH 8 for 24 hours. We used *Staphylococcus aureus* GluC (Peprotech) to cleave α_1-AT at the P5-P6 bond (Met358-Ser359) on the RCL and reconstituted it in water to make a stock solution. GluC was added in a 1:20 ratio GluC: α_1-AT and incubated for 4 hours at 37°C. The cleaved protein was run on an SDS-PAGE gel and GluC was inactivated by incubating the solution at 90°C for 4 hours. A 0.22um filter was used to remove the aggregates and the purified cleaved M was stored at -80°C.

2.3.3 Induction of the latent conformer of α_1-antitrypsin

Latent M was prepared according to[114]. Purified α_1-AT (0.05mg/mL) was incubated at 68°C for 48 hours in 20mM Tris, 0.7M sodium citrate pH 7.4. This was then dialysed (3x 5L) with 20mM Tris pH 8.6 to remove the citrate and α_1-AT was concentrated to 1mg/mL solution. The resulting solution was then heated to 60°C for 3 hours to convert remaining active α_1-AT to polymer. Anion-exchange chromatography was used to separate monomer from polymer. Latent was separated from polymer over 60min with a 0-1M NaCl gradient in 20mM Tris pH 8.6. Fractions were collected, run through SDS-PAGE and Native gels and pooled before measuring the concentration and activity of latent α_1-AT.

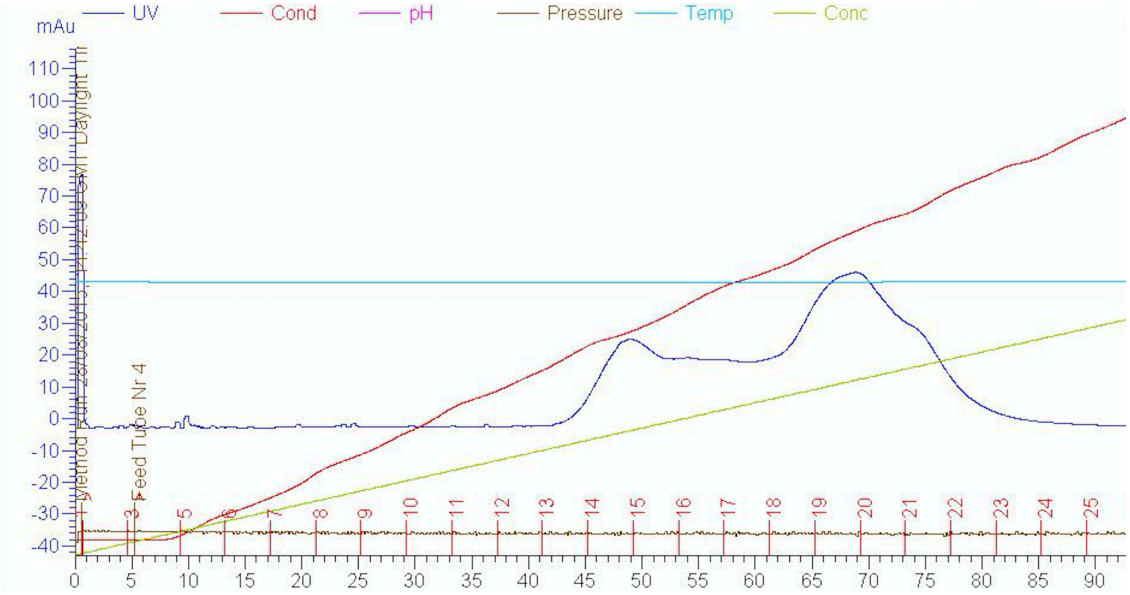


Figure 2.1. Elution of polymers by Q-sepharose. Polymers started eluting at fraction 14-22 (navy blue) as the salt gradient increased (red). Image taken by the AKTA software.

2.4 Protein Characterisation

2.4.1 Sodium dodecyl sulphate (SDS) PAGE

In addition to native-PAGE gels, SDS-PAGE was used to confirm the purified variants against a protein ladder and known plasma M and recombinant proteins (Fig.2.2 and Fig.2.3).

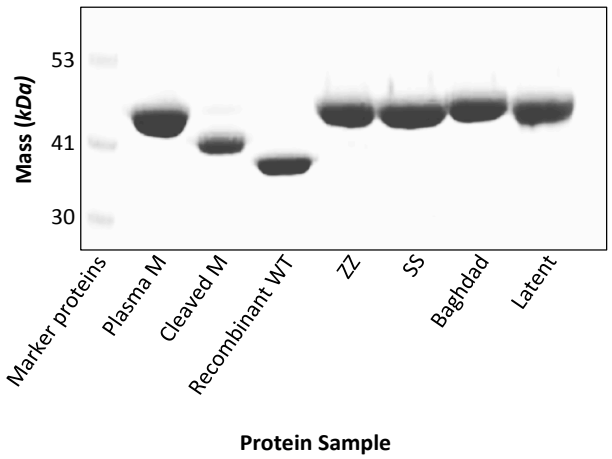


Figure 2.2. SDS-PAGE analysis of the plasma proteins. The approximate mass of each protein (labelled) can be determined by comparing with the marker proteins.

2.4.2 Non-denaturing PAGE

Biochemical analyses were performed for all proteins purified to confirm a successful purification. Fig.2.3 shows the native-PAGE gel of all the variants.

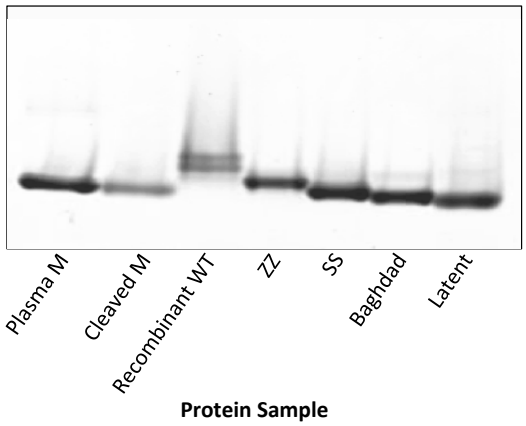


Figure 2.3. Native PAGE gel of the plasma variants. The plasma variants used for analysis are shown on the gel. The cleaved protein runs the same as the uncleaved plasma M because the RCL is still inserted into the β-sheet A. The recombinant protein is smaller due to no glycosylation and hence migrates to a lesser extent through the gel. All other proteins migrate similarly, hence the need for a high-resolution technique.

2.5 Preparation of α_1-antitrypsin for Native Mass & Ion Mobility Mass Spectrometry

All samples as prepared above and used for calibrating the instrument were buffer changed using Amicon Ultra 0.5 mL centrifugal filters (Millipore UK Ltd. Watford) into 200mM Ammonium Acetate and diluted to a concentration of 20μM. All samples were run on a Synapt G1 HDMS Q-TOF instrument (Waters, UK).

2.5.1 Calibrating the Instrument

The Q-TOF HDMS instrument was mass calibrated up to 12000*m/z* using 33μM Cesium iodide in 200mM Ammonium acetate. The TWIMS was calibrated using proteins Bovine Serum Albumin (BSA), Concanavalin A (Con A), and β-lactoglobulin (β-lac) purchased from Sigma Aldrich (St. Louis, MO) which have known CCS values[156]. The samples were then run under the ion mobility conditions shown in table 2.1. Amphitrite software developed in house[62] was then used to create a calibration curve by selecting charge states which match published CCS values. ATDs from selected peaks of the calibrants are then plotted and a power fit to the data is then calculated and plotted. From this, the experimental CCS values for the proteins under investigation can be determined from their ATDs. The calibration calculation and software is further explained in[62].

2.5.2 Parameters for data acquisition

The settings employed in the mass spectrometer for both native and Ion Mobility are given in Table.2.1.

Table 2.1. The settings employed for native ion mobility mass spectrometry analyses including those used for collision induced unfolding.

Settings	**Native**	**CIU**
Capillary Voltage (kV)	1.1	1.1
Sampling Cone (V)	30	30
Extraction Cone (V)	1	1
Source Temperature (°C)	40	40
Trap Collision Energy (eV)	10	10-50
Transfer Collision Energy	10	10
Mass Range (*m/z)*	1000-8000	1000-8000
Bias Voltage (V)	10	20
Backing Pressure (*m*bar)	0.48	0.48
IMS Wave Height (V)	10	8
IMS Wave Velocity (*ms^{-1})*	350	350
Transfer Wave Height (V)	8	8
Transfer Wave Velocity (*ms^{-1}*)	100	100
LM/HM Resolution	5.0/15.0	5.0/15.0

2.5.3 Theoretical collision cross section calculations

Theoretical CCS values of α_1-AT were calculated using the Projection Approximation method as described in section 1.5 for the pdb files 3NE4 (native α_1-AT), 3DRM (T114F), 3DRU (G117F), 1QMB (cleaved), 7API (S). Other variants have not been crystallised. The calculated CCS using PA are shown in Table 2.2.

Table 2.2. Theoretical calculations of recombinant and plasma alpha$_1$-antitrypsin variants. The PDBIDs are given alongside the theoretically calculated masses of the proteins based on the sequence data provided. The theoretical calculations were calculated by the Projection Approximation method. Theoretical values for unsolved structures was not possible and hence are left blank.

Protein Variant	PDBID	Theoretical Mass (Da)	Theoretical CCS (Å^2)
recWT	3NE4	44324	2769
G117F	3DRU	44386	3085
T114F	3DRM	44342	2692
K154N	-	44310	-
plasma M	-	44324	-
cleaved	1D5S	44324	2823
plasma S	7API	44294	2729
plasma Z	-	44323	-
plasma B	-	44350	-

Chapter 3- Biochemical and biophysical studies of recombinant alpha1-antitrypsin

3.1 Introduction

α_1-AT has only recently been studied with mass spectrometry to investigate recombinant wild type α_1-AT, one *forme fruste* recombinant variant, and artificially glycosylated α_1-AT[62,112,124,125]. Here, we aimed to probe differences, if any, in CCS and unfolding between stabilising and intermediate-like variants as a proof-of-principle before studying *ex vivo* samples. The proteins used for this were G117F, T114F, K154N, and recombinant WT. Thr114Phe and Gly117Phe mutations border the surface cavity of a drug design target. The T114F mutation causes partial cavity blockage and its structure has proved useful for novel pharmacophores[126] as it resists polymer formation whilst preserving inhibitory function[94]. The G117F mutation changes the local packing and the position of the F-helix down by half a turn. Studies show how these mutations can rescue defective folding of Z α_1-AT in COS-7 cell secretion model[94].

K154N however is a slowly polymerising α_1-AT variant and occupies a polymerogenic intermediate state along the polymerisation pathway[112]. Under the relatively native conditions employed, K154N exhibited 82% native-like structural and dynamic behaviour. The authors concluded that this would reject the domain swapping models of polymerisation in favour of the β-sheet A insertion model.

The aims and objectives of this project are to analyse these recombinant proteins with mass spectrometry to solidify previous results obtained by crystallography, NMR, or PAGE gels. In addition, collision induced unfolding will be used to assess the protein's stability over a range of different collision energies. Both studies will be used as a proof-of-principle experiment for further studies on plasma variants.

3.2 Results

3.2.1 Comparing the CCS values of recombinant monomers

Figure 3.1 shows the native mass spectra of each variant (Fig.3.1.*a*) alongside the arrival time distribution and collision cross section (Fig.3.1.b, Fig.3.1.c) for the full spectra. The spectra clearly show two distinct charge state distributions where the higher charge state distribution for each variants are shown. In these examples, there seems to be a large difference in the ratio between the high charge state distributions and the low charge state distributions where G117F variant is a much greater proportion of high charge state distribution. As indicated, (Fig.3.1.*a*) the 13+ charge state was isolated for collision induced dissociation (CID) across trap voltages ranging from 10-50V and are described in section 3.2.2.

Recombinant WT 13+ charge state has a CCS value of 3122 $Å^2$, G117F 3172 $Å^2$, T114F 3104 $Å^2$, and K154N 3153 $Å^2$. The peak widths of the CCS distributions indicate the dynamics of the protein's conformation ensemble. Recombinant WT has a CCS width of 142 $Å^2$, G117F and T114F have a width of 81 $Å^2$ and 121 $Å^2$ respectively, and the K154N variant has a peak width of 109 $Å^2$.

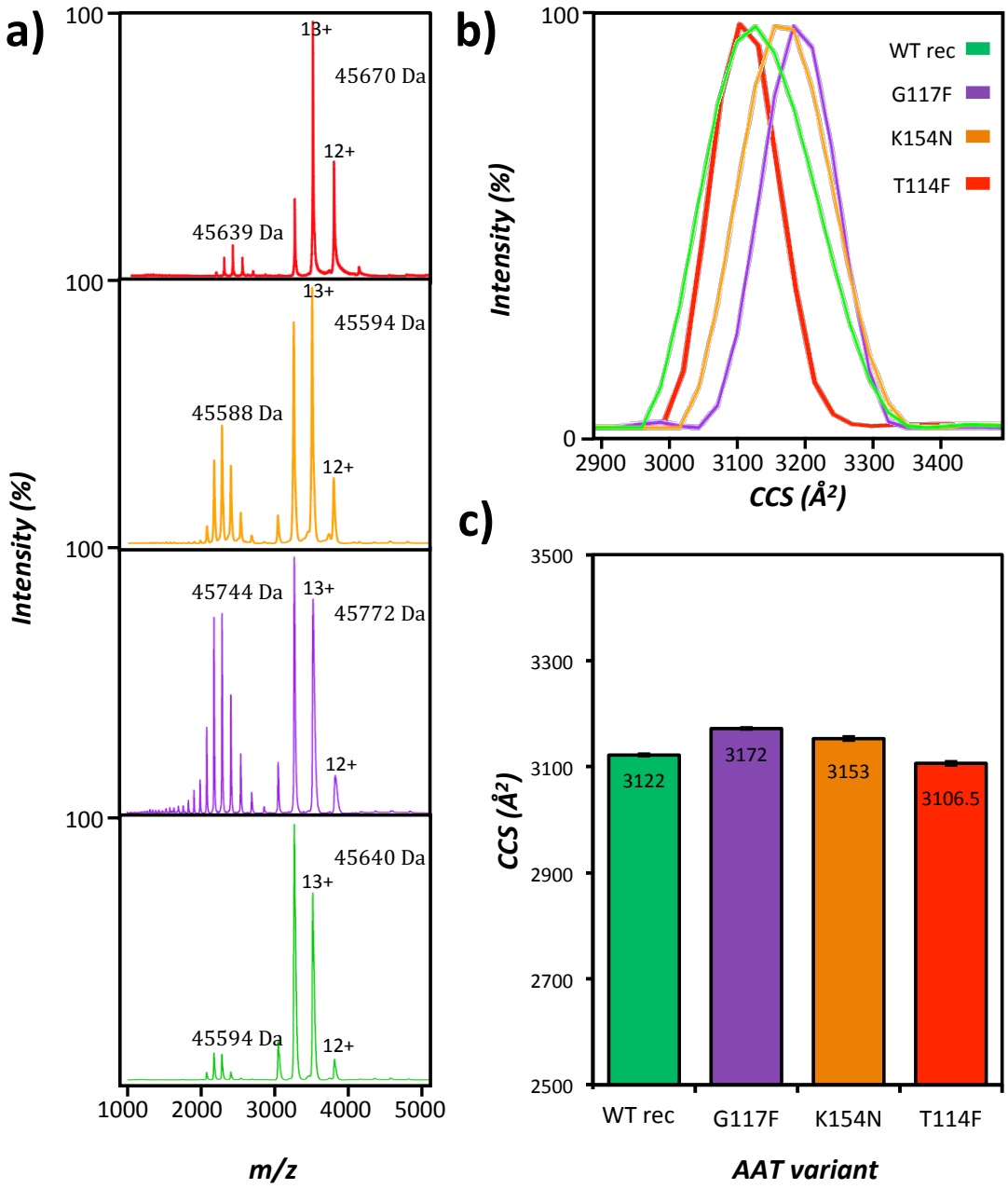


Figure 3.1. Native and Ion Mobility Analyses of Recombinant Alpha$_1$-antitrypsin. **A)** The native mass spectrum for each variant. The variants are WT(green), G117F (purple), and K154N(orange), and T114F (red). The 13+ charge state is that used for charge state isolation. B) The Arrival Time Distribution of the recombinant variants. These samples were incubated at 40°C before analyses. C) Barchart showing the collision cross section and Standard Deviation as a result of ion mobility shown in b). The data is plotted as an average of three repeats.

3.2.2 Collision Induced Unfolding of α1-antitrypsin recombinant variants

Data for wild type α_1-AT under a CIU experiment are shown in Fig.3.2. As the collision energy increases from 10eV through 50eV there is a clear transition between the native, folded conformation (3080 Å^2) and one which is more unfolded (3570 Å^2). Three distinct species are observed under these conditions. Under native conditions the CCS is 3080 Å^2, close to the theoretically calculated CCSs of 2601 Å (Projection Approximation) and 3379 Å (Exact Hard Sphere Scattering), indicating that the protein is folded and intact in the gas-phase. At 20eV the peak width gets wider but the peak top remains the same. At 30eV the protein occupies three distinct conformers, their top peaks are 3094 Å^2, 3371 Å^2, and 3542 Å^2. An increase in 10eV shows no sign of the native structure but the 3371 Å^2 species increases in abundance as well as the unfolded species. At 50eV, the protein has reached its upper limit of unfolded which is measured as 3567 Å^2 and a peak width of 181 Å^2 compared to the natively folded species with a peak width of 100 Å^2.

T114F and G117F have slightly different folding patterns to each other and also compared to wild-type. The T114F and G117F have a native fold of 3050 Å^2 (peak width 110 Å^2) and 3066 Å^2 (peak width 123 Å^2) respectively. Both stabilising variants have negligible differences at 10 and 20eV although the peak widths widen due to their gained internal energies. A protein's internal energy refers to the increased vibration of atoms in the structure upon an increase in energy. At 30eV, most of the T114F is in the native state with smaller peaks at 3210 Å^2, and 3330 Å^2. For the G117F mutant, the percentage of the intermediate at 3340 Å^2 is much greater than for T114F. At 40eV the predominant peak top of T114F is 3360 Å^2 followed by the unfolded species at 3590 Å^2. For the G117F at 40eV, there are two intermediates, highest peak being more natively folded with a peak top 3366 Å^2 and the second intermediate with a peak top 3500 Å^2. There is only slight indication of the unfolded state at 40eV for the G117F mutant. At 50eV, both proteins have reached their upper limits and are unfolded; the peak tops are 3580 Å^2 (peak width 196 Å^2), and 3590 Å^2 (peak width 150 Å^2) for T114F and G117F respectively.

K154N has a CCS of 3094 Å^2 (peak width 143 Å^2) starts to unfold as the collision energy is ramped to 20eV. The predominant species is still natively folded but the wider peak width of 241 Å^2 suggests that an intermediate is present. At 30eV protein unfolding is well underway, with the folded (3094 Å^2), intermediate (3379 Å^2), and unfolded state (3597 Å^2) in a 2:3:2 ratio. At 40eV the protein occupies in part the intermediate and unfolded state. At 50eV, the protein is unfolded at 3597 Å^2 (peak width 278 Å^2).

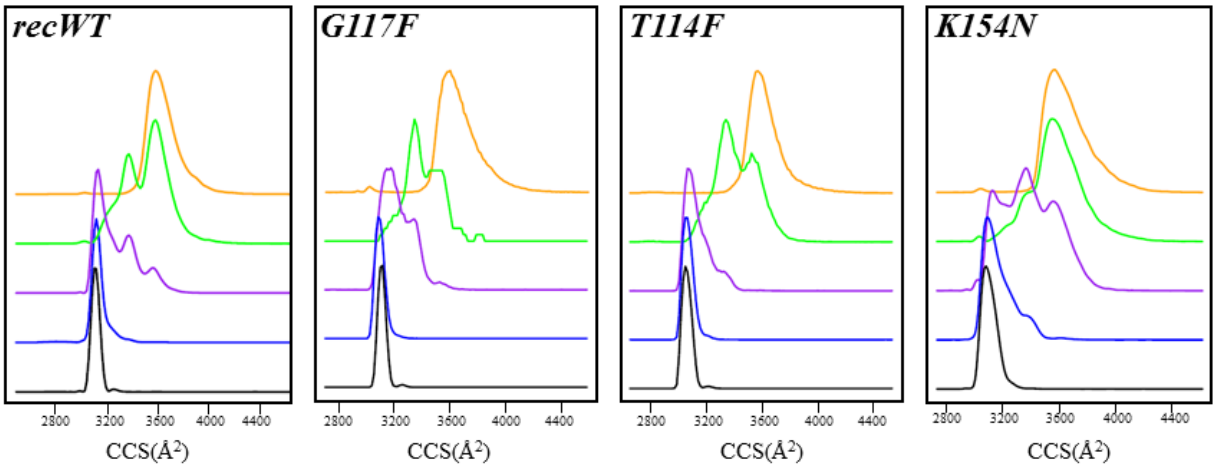


Figure 3.2. Collision induced dissociation of recombinant variants. From left to right: recombinant WT, G117F, T114F, K154N. The collision energy profiles show the protein unfolding as the internal energy is increased. The collision energies are shown: 10eV (black), 20eV (blue), 30eV (purple), 40eV (green), and 50eV (orange).

3.3 Discussion and Conclusions

3.3.1 Comparing the CCS values of recombinant monomers

The data presented agrees with past studies showing that K154N unfolds at 40°C[112] to give a larger CCS value than recombinant wild-type. The G117F has a larger CCS value than both of the proteins but this could be a result of the downwards shift of the F-helix as the CCS value obtained at lower temperatures is greater than K154N and WT. In addition, the peak width is smaller than the other variants indicating reduced flexibility of the protein. This would confirm data already known about the G117F variant which stabilises the Z mutation. Furthermore, the CIU data confirms that the G117F retains its native species for longer despite having an increased CCS value. T114F has a smaller CCS change than all the variants suggesting that it unfolds to a lesser extent than the other species.

There is sufficient evidence to suggest that the high charge state distribution is a result of the histidine tag used for protein purification of the recombinant proteins. Firstly, the high charge states have a CCS similar to that of the low charge state species, therefore there is little unfolding occurring in the structure (Fig.A.3). Secondly, the tag carries six histidine residues resulting in a charge state distribution 6 charges greater than the low charge state species. Finally, the masses of the high charge state and low charge state distributions are not greater than 155Da (the mass of one Histidine residue) therefore concluding that the Histidine tag is not cleaved but somehow unfolds from the protein, hence revealing the six additional charges. Therefore, the high charge state species is not indicative of the protein unfolding but rather the histidine-tag remaining on the protein.

3.3.2 Collision Induced Dissociation of recombinant alpha$_1$-antitrypsin

It is currently thought that α_1-AT unfolds via an intermediate in the form M→ I→ U , where I is the intermediate, and U is the unfolded structure. Recombinant WT follows the same pattern as the stabilising variants for the first two collision energy intervals whereby at 20eV, the

protein's internal energy gain has resulted in a more flexible and dynamic nature (indicated by the increased peak width) allowing it to occupy a much wider range of conformers, some more extended, some less extended. Following a further increase in eV, the protein unfolds into two further species. The peak top in the middle with a CCS of 3371 $Å^2$ can be considered the intermediate whereas the third peak at 3542 $Å^2$ seen at 40 and 50eV could be the most unfolded species under these conditions. Higher energies were attempted but signal losses meant that the 13+ charge state was no longer visible past these conditions. Hence, at 30eV, the protein already reaches the unfolded state. Any increases in intensity of the native conformation would suggest that point mutations had an effect on the stability of the protein. Experiments like these enable us to postulate the effects of single point mutations on the dynamic behaviour of an intermediate.

Analysis of G117F α_1-AT mutant also revealed interesting findings. The crystal structure of this variant shows a downward shift of the F-helix by half a turn to re-introduce packing of the helix's residues between β-sheet $2A^{94}$. This would explain why G117F is larger than the other variants despite being a stabilising mutant. The data also supports G117F as a stabilising variant as it retains natively folded conformers more than either recWT or K154N. This mechanism is thought to be responsible for the stabilisation of WT and Z α_1-AT in hepatocyte cell models.

In contrast, T114F functions by partial cavity blockage. The native mass spectrum presented here shows T114F to have a reduced intensity of the high charge state distribution with respect to the lower, more native charge state distribution. In addition, the lower charge state distribution seems to have shifted to the right. This is perhaps the most stabilising variant under these conditions as the peak width only starts to widen at 30eV and the intensity of the 3210 $Å^2$ intermediate is lower than the other variants. At 40eV the CCS distribution is very similar to the G117F but both exhibit and intermediate as the greatest intensity with lower levels of the unfolded species than both WT and K154N. Despite having an unfolded CCS slightly

larger than the WT, the peak width is smaller, suggesting that the protein even at high collision energies is less flexible and hence more stable than WT.

K154N is a slowly polymerising variant, occupying the polymerogenic intermediate in native conditions. Indeed it has a larger CCS and wider peak width indicating that the polymerogenic intermediate is occupying a larger conformational ensemble. The CCS value of the protein incubated at 40°C was 3153 $Å^2$, similar to that obtained by Nyon *et al.*[112] who observed a CCS value of 3220 $Å^2$ when heated at 40°C. The CCS of the K154N at native conditions taken for the CIU data is 3094 $Å^2$. This was collected from sample incubated at 4°C and agrees with findings that the K154N does not have a larger collision cross section below 40°C[112]. However, as this experiment collisionally activates the protein, the data still agrees with findings that K154N is a slowly polymerising variant that more readily occupies the intermediate state because it unfolds much more readily at 30eV than any of the other variants (T114F, G117F, and recWT).

In conclusion, the biophysical data presented here compliment known biochemical and computational data in that the stabilising variants show a stabilised unfolding pathway, the slowly polymerising variant adopts an unfolded intermediate earlier than the recombinant wild type and the experimental CCSs are close to those computationally derived from crystal structures.

Chapter 4- Ion mobility mass spectrometry and collision induced unfolding studies of ex-vivo *Alpha$_1$-antitrypsin plasma variants*

4.1 Introduction

As explained in section 1.6, α_1-AT deficiency causes both early onset emphysema and predisposes patients to liver cirrhosis. In this chapter, disease variants extracted from affected patients were studied with ion mobility mass spectrometry and CIU in the hope of finding variations between the disease monomers in conformational flexibility, structure and/or stability. The hypothesis being that depending on their conformational behaviours, the proteins might follow a different path to polymerisation. By understanding this pathway to aggregation, one might be able to interfere at specific points to reduce the pathological effects.

4.1.1 Glycosylation

Here, the carbohydrate composition of different plasma variants was explored using MS and IM-MS to determine whether there were any differences in glycan structure that explain their behaviour in disease.

Glycosylation is one of the most complex post-translational modifications that proteins can undergo[127]. It is important in protein folding, flexibility, stability, and signalling. Carbohydrate residues are attached to the proteins at certain residues such as Asparagine, Threonine and Serine and serve to increase protein stability by protecting them from proteolysis and degradation[128,129] Glycans play a role in protein folding and polymerisation because they bring about differences in relaxation and mobility[125]. Specific oligosaccharides are able to modulate protein steric interactions and affect specific protein functions due to their ability to bind to different receptors. In addition, glycans also play important roles as signalling molecules in innate immunity[128].

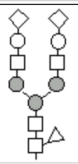
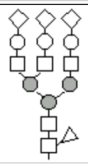
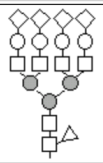
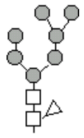
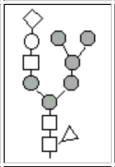
Traditionally, glycosylation patterns of proteins are identified by Isoelectric Focusing (IEF), which separates proteins on their molecular masses but also on their isoelectric points.

The different glycosylation residues are shown here in Fig.4.1 and represent the most common residues capable of binding to proteins. Biochemical studies of α_1-AT glycosylation have enabled McCarthy *et al.*[127] to computationally add the glycans to crystallised α_1-AT, the image is shown in Fig.4.2. This enabled us to visualise the plasma variant at atomic resolution.

Human α_1-AT has three *N*-glycosylation sites, N70, N107, and N271. N70 is often bi-antennary di-sialylated with few tri-antennary trisialylated branches and is not often fucosylated[125,127,130]. Glycans on site N107 have shown the most variation in structure with bi-, tri-, and tetra-antennary, bi-, tri-, and tetra-sialylation, core glycans and fucosylation. N271 has some core fucosylation and is often bi-antennary di-sialylated. Changes in α_1-AT glycosylation are often linked to an increase in fucosylation.

In healthy patients, five isoforms of α_1-AT are found due to *N*-glycan heterogeneity[125,131]. Glycans on site N107 have shown the most variation in structure. Differences in the IEF pattern are detected if patients have different genetic variants or congenital disorders of glycosylation (CDG). Abnormalities in IEF patterns can also be due to changes in both the peptide sequence and the *N*-glycans therefore other analytical techniques are required to understand the molecular basis of these differences.

a)

	complex sialylated fucosylated **diantennary**	**complex** sialylated fucosylated **triantennary**	**complex** sialylated fucosylated **tetraantennary**	**high mannose** (man 7)	**hybrid**	common core str. subunit	sialylated branch subunit	branch subunit
N-Linked Structures								
Composition	$C_{90}O_{65}N_6H_{146}$	$C_{115}O_{83}N_8H_{186}$	$C_{140}O_{101}N_{10}H_{226}$	$C_{64}O_{49}N_2H_{106}$	$C_{83}O_{62}N_4H_{136}$	$C_{34}O_{25}N_2H_{56}$	$C_{25}O_{18}N_2H_{40}$	$C_{14}O_{10}N_1H_{23}$
Mono. Mass	2350.8303	3007.0579	3663.2856	1686.5864	2180.7612	892.3172	656.2276	365.1322
Av. Mass	2352.13	3008.72	3665.30	1687.51	2181.96	892.81	656.59	365.33

b)

Residue	Symbol	Residue Composition	Mono Isotopic Mass	Average Mass
Sialic Acid	◇	$C_{11}O_8NH_{17}$	291.0954	291.26
Galactose	○	$C_6O_5H_{10}$	162.0528	162.14
N-Acetylglucosamine	□	$C_8O_5NH_{13}$	203.0794	203.19
Mannose	●	$C_6O_5H_{10}$	162.0528	162.14
Fucose	△	$C_6O_4H_{10}$	146.0579	146.14

Figure 4.1. Common Protein Glycans. **a**) The masses of N-linked structures are shown as well as their simplified structural formulas. **b**) The masses and molecular formula of common glycan residues. Figure taken from http://www.ionsource.com/Card/carbo/carbstr.htm

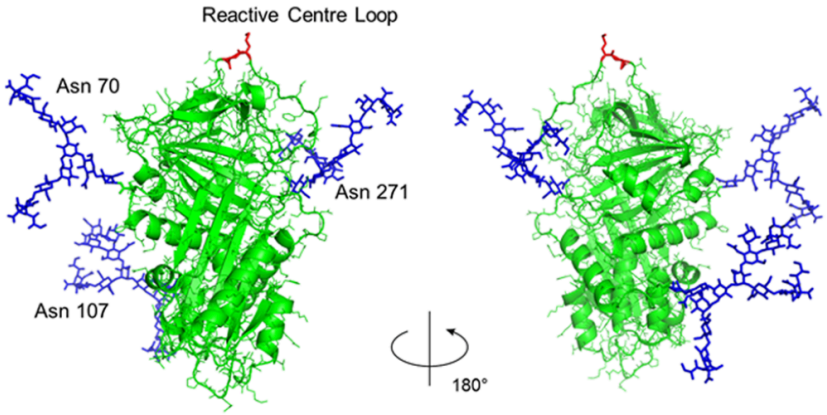


Figure 4.2. Computationally modelled glycosylated α_1-antitrypsin. The glycans were added computationally following extensive glycan sequencing by mass spectrometry. The protein (3NE4.pdb) is shown in green, the reactive centre loop in red and the glycans in blue. Figure reproduced from[127] McCarthy C. *et al.*

4.1.2 Current Structural Understanding of Plasma Antitrypsin

Thus far, atomic resolution of plasma α_1-AT has not been achieved, leaving a limited understanding of the structural nature of these proteins with attached glycans. Plasma variants are representative of the α_1-AT that circulates in human bodies. Plasma Z and S have long been characterised by several groups, but their crystal structures are still unknown[132–136]. Therefore, investigating the overall structure of these macromolecules might reveal differences in the phenotypes of α_1-AT deficient patients.

In 2015, a novel deficiency variant, B (A336P, Baghdad) mutant was characterised as being more polymerogenic than Z[137]. In addition to the reduced activity in B compared to wild-type protein, the patient presented with a total of 85-95% functional deficiency. Researchers also found that despite having less impaired folding, B α_1-AT could populate the polymerogenic intermediate more readily than Z[88]. Given the differences in mutations on the resulting phenotype, this further supports the need for more individualised care.

4.2 Results

4.2.1 Ion Mobility in the Study of Glycosylated Proteins

The native mass spectra of the plasma variants are shown in Fig.4.3. The full spectrum is shown to the left in Fig.4.3.*a*. Interestingly, the recombinant protein has two charge state distributions from 12-15+ and from 18-23+ charges whereas the plasma variants have only a single charge state distribution for the monomer. Naturally, the monomers from plasma are glycosylated and this explains the ~5kDa difference in mass between the plasma variants and the recombinant WT. For each spectrum in Fig.4.3.*a* the 13+ charge state is highlighted in blue and further focused upon in the remaining panels. The 13+ charge state was chosen because its CCS is within those theoretically calculated for the recombinant protein and thus is believed to be the more natively-folded in structure. In addition, this peak has a high signal allowing more data to be collected during charge state isolation.

Fig.4.3.*b* shows the different 13+ peaks with varying levels of glycosylation. For each peak highlighted in a given colour, the mass is shown in the same colour. For different masses in the same colour the peaks go from left to right *m/z* and represent the different masses as a result of the glycans attached to the protein.

As well as separating the differently glycosylated α_1-AT by *m/z,* they could be separated using ion mobility due to their differences in conformation. The arrival times are shown in Fig.4.3.*c* where each ATD corresponds to the same coloured peak and mass in Fig.4.3.*b*. In order to determine which glycans are attached to the plasma variants further tests are necessary but we can go some way into predicting those that might be attached to the protein. Figure 4.1 shows the N-linked structures, their composition, and mass. Fig.4.1 shows in more detail the residues making up the N-linked structure their composition and monoisotopic masses. Using these known values and the mass differences between the different peaks in Fig.4.3.*a* it is possible to predict the possible glycans that exist on each plasma variant. This is shown in Table

4.1 where the difference in masses could correspond to the loss/gain of specific carbohydrate residues.

Plasma M has a difference of 531 between the green and purple peak in Fig.4.3.*b*, this could be due to the additional mass of 527 Da from a branch subunit and a galactose or mannose residue. Given that the difference in mass is +- 17.4 Da, the proposed glycans above fit well within these error values. Secondly, the difference in mass between the green and orange peaks is 654 +- 20 Da. The difference in mass here is likely to be a sialylated branch subunit with a mass of 656 Da.

The S plasma variant has a difference of 658+- 4 Da between the purple and green peaks. Again, this mass will likely correspond to a sialylated branch subunit. Between the green and orange peaks, the mass difference is 139 +- 5 Da. From Fig.4.1 there is no residue smaller than fucose with a mass of 146 Da so although only 2Da outside the range of error calculated for the difference in mass, it is most likely this as no other residue has a mass closer to 146 Da.

The relatively novel plasma variant, B had a mass difference of 657 +- 11Da between the purple and green highlighted peaks and between the green and orange peak. As above, it is most likely that these mass differences are due to the addition of a sialylated branch subunit with a molecular mass of 656 Da.

Finally, the Z variant has several smaller peaks within the two main that are highlighted in purple and green. The mass differences between the smaller peaks are 64 Da which could perhaps be due to the loss of an oxidated Methionine residue. The difference in mass between the purple and green peaks here are 658 +- 5 Da, which again would be considered to be the loss of a sialylated branch subunit. Bi-, tri-, and tetra-sialylation are found on the N70, N107, and N271 glycosylation sites of alpha$_1$-antitrypsin as shown in McCarthy's paper[127].

Fig.4.3.*c* also shows the separation of these differently glycosylated monomers via ion mobility. The drift time is recorded and shows that the more glycosylated the species - hence a greater mass - the greater the arrival time and the wider the arrival time distribution.

Table 4.1. Predicted glycans responsible for the differences in masses shown in Fig.4.3. The peak numbers correspond to the peaks in Fig.4.3 going left-right across the mass spectrum. Numbers in bold represent the theoretical masses of the proteins without glycosylation, the italic numbers are the mass differences between the peak number and the previous peak. For peak 1 this is the difference between the 1st peak and the theoretical mass. Numbers in brackets are the masses calculated from the Fig.4.1 and are those thought to be causing the mass difference.

Mass Difference Between Glycosylated Species & Possible Glycans					
	Theoretical mass (Da)	**Peak 1 (Da)**	**Peak 2 (Da)**	**Peak 3 (Da)**	**Peak 4 (Da)**
cleavedM	**44324**	***49496***	***50099***	***50599***	***51189***
			sialic acid, galactose/mannose, fucose (599) these could have come off and N-Ac remains-all peaks remain glycosylated	branch subunit + fucose(511) or 3mannose/galactose (486)	2x sialic acid (582)
M	**44324**	***50429***	***50960***	***51614***	
			branch subunit + galactose/mannose (527)	sialylated branch subunit (656)	
S	**44294**	***51008***	***51666***	***51805***	
		complex tetra, 2x common core subunit (5449)	sialylated branch subunit (656)	fucose (146)	
B	**44350**	***51059***	***51716***	***52370***	
			sialylated branch subunit (656)	sialylated branch subunit(656)	
		complex tetra, 2x common core subunit (5449)			
Z	**44323**	***50934***	***50998***	***51592***	***51656***
		due to complex tetra and high mannose (5352)		2x sialic acid (582)	loss of Met-ox

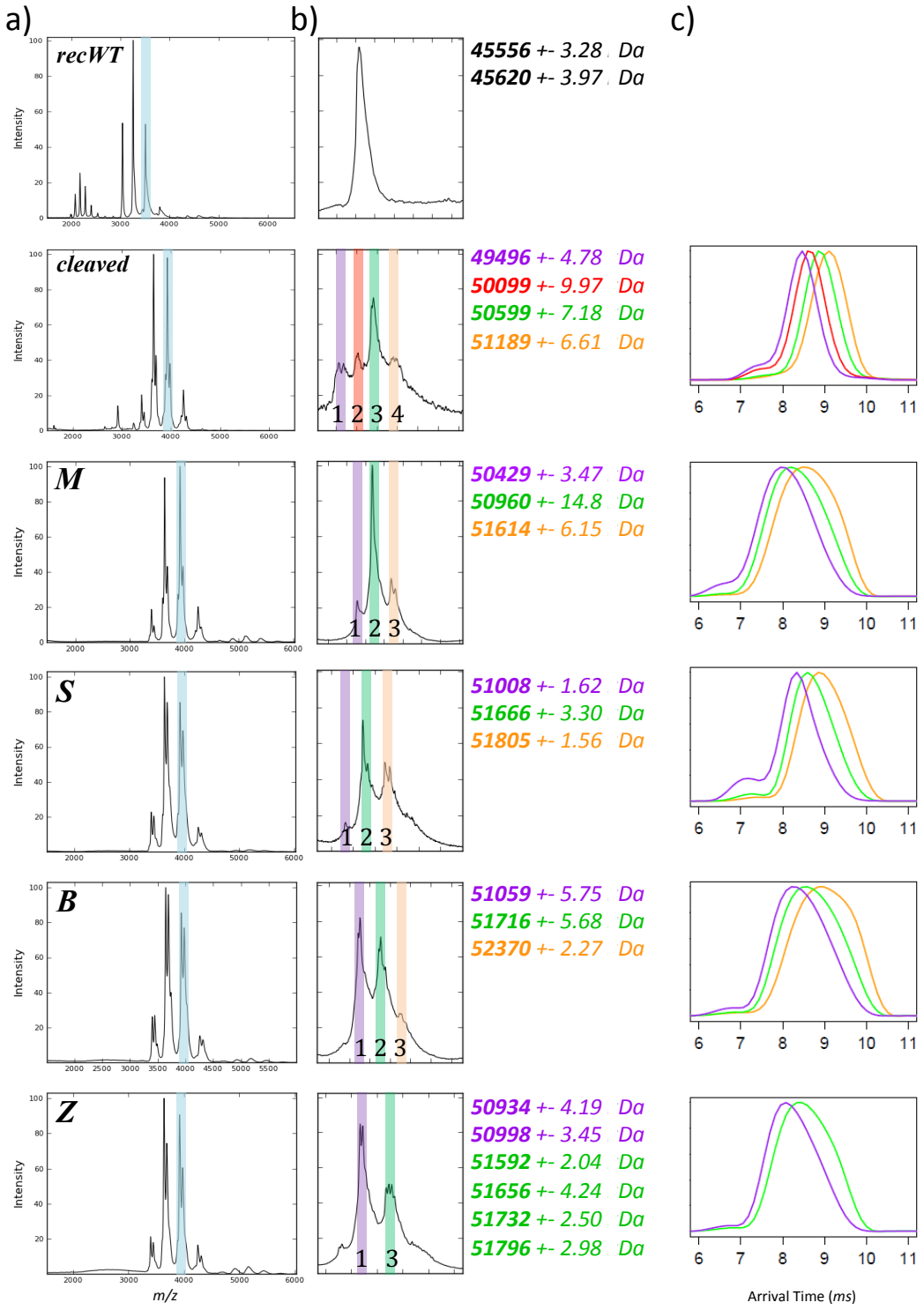


Figure 4.3. a. Native mass spectra of plasma variants. The charge state distribution is shown with the 13+ charge state highlighted in blue. **b**) The 13+ charge state mass spectrum. The glycosylation peaks are coloured purple, red, green, and orange corresponding to their respective masses as indicated by their colours. **c**) The arrival time distribution of the different glycosylated peaks. The coloured peaks represent those coloured in **b**.

4.2.2 Comparing the CCS values of plasma variants

The arrival time distributions in Fig.4.3.*c* and the CCS values in Fig.4.4 show variations between the species. Fig.4.4 shows the difference in CCS between the different variants against the recombinant and plasma M controls. All plasma variants are approximately 5kDa larger in mass than the recombinant control due to glycosylation.

Figure 4.4 shows the mean CCS values (from 2 repeats) obtained in a single day's acquisition for all samples to ensure that all conditions were the same. The recombinant wild type has a mean CCS of 3089 Å^2 and a +- 9.9 Å^2 range. Plasma M has a CCS of 3165 Å^2 and a range of +- 3.5 Å^2 whilst the compact controls; cleaved M and latent M, have larger CCS of 3192 Å^2 and 3188 Å^2 and a range +-3.5 Å^2 and +- 1.4 Å^2 respectively. The disease variants S, B and Z have CCS values of 3195 Å^2, 3223 Å^2, and 3200 Å^2 with a range of +-10.6 Å^2, +-0.7 Å^2, and +-0.7 Å^2 respectively.

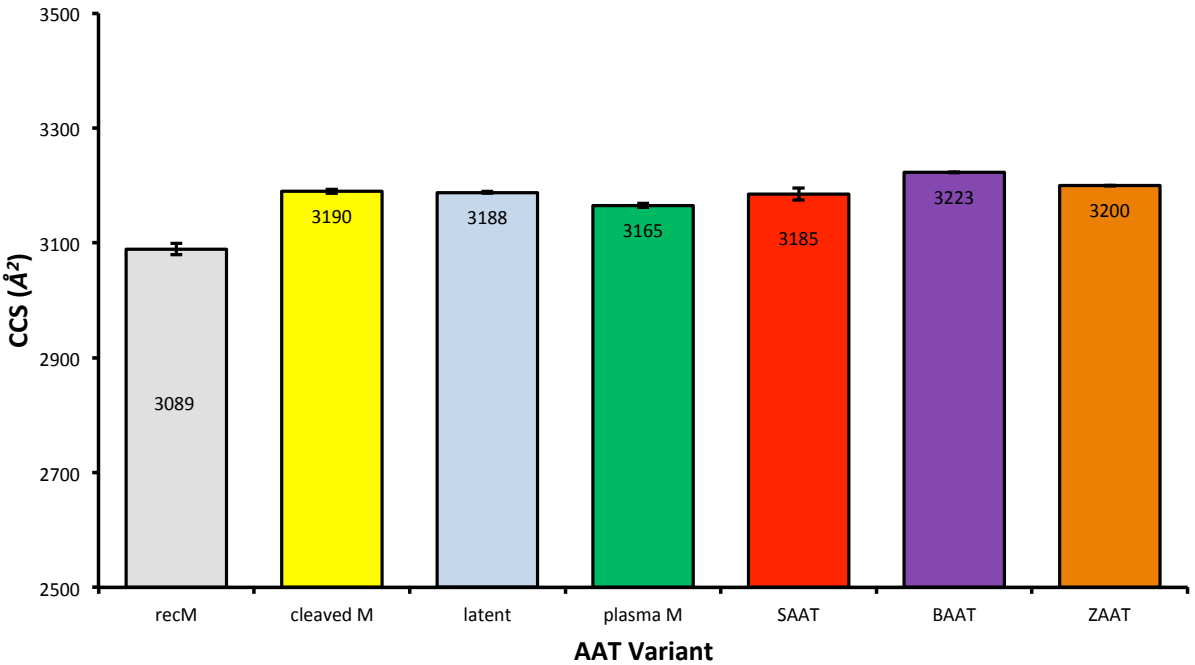


Figure 4.4. Mean collisional cross section of the plasma variants. Each plasma variant is labelled; the mean CCS value is given inside the column and the error bars for each sample show the standard deviation.

4.2.3 Collision Induced Unfolding patterns of α1-antitrypsin plasma variants

Alpha$_1$-antitrypsin unfolding shows the effect of glycosylation in stabilising the protein compared to recombinant wild type (Fig.4.5). The unfolding pattern also shows slight differences in the stabilities of the plasma variants.

Initially, cleaved α_1-AT was included in the experiment as a compact control. Contrastingly, the collision cross section is significantly larger than plasma M. In addition, we see the reactive centre loop ejected from the protein as it begins to unfold at higher collision energies in the trap. This is indicated by the appearance of a narrow peak from 2700-3000Å^2. As the collision energy increases further from 40-50eV, the intensity of the RCL peak increases, further confirming the unfolding of the protein.

Recombinant WT is used as a control as well as plasma M (Fig.4.5). Recombinant WT is smaller in CCS than the other proteins (3089 Å^2), it is unaffected by 20eV collision energy but at 30eV, three conformations appear with the predominant conformation still being natively folded. At 40eV the natively folded state is no longer abundant and the majority of the protein is in the unfolded state. At 50eV the protein is completely unfolded and no other conformations are visible.

For M, Z, B, and S, the proteins remain unchanged from 10-20eV. At 30eV, they begin to occupy a more flexible configuration, shown in Fig.4.5 as a widening of the ATD. Further increases in collision energy to 40eV shows the introduction of another configuration, most likely the intermediate as shown in the second peak, the proteins now occupy a wider range of conformations including that of the native. A further 10eV results in the unfolded state and the native conformer is no longer present. Wider peak widths suggest that a greater number of larger conformations have formed as a result of the increased collisional energy.

S unfolds to a greater extent than B, Z or M. The CCS of the S variant is greater that both other mutants but it behaves in the same way for the three initial collision energies (10-30eV).

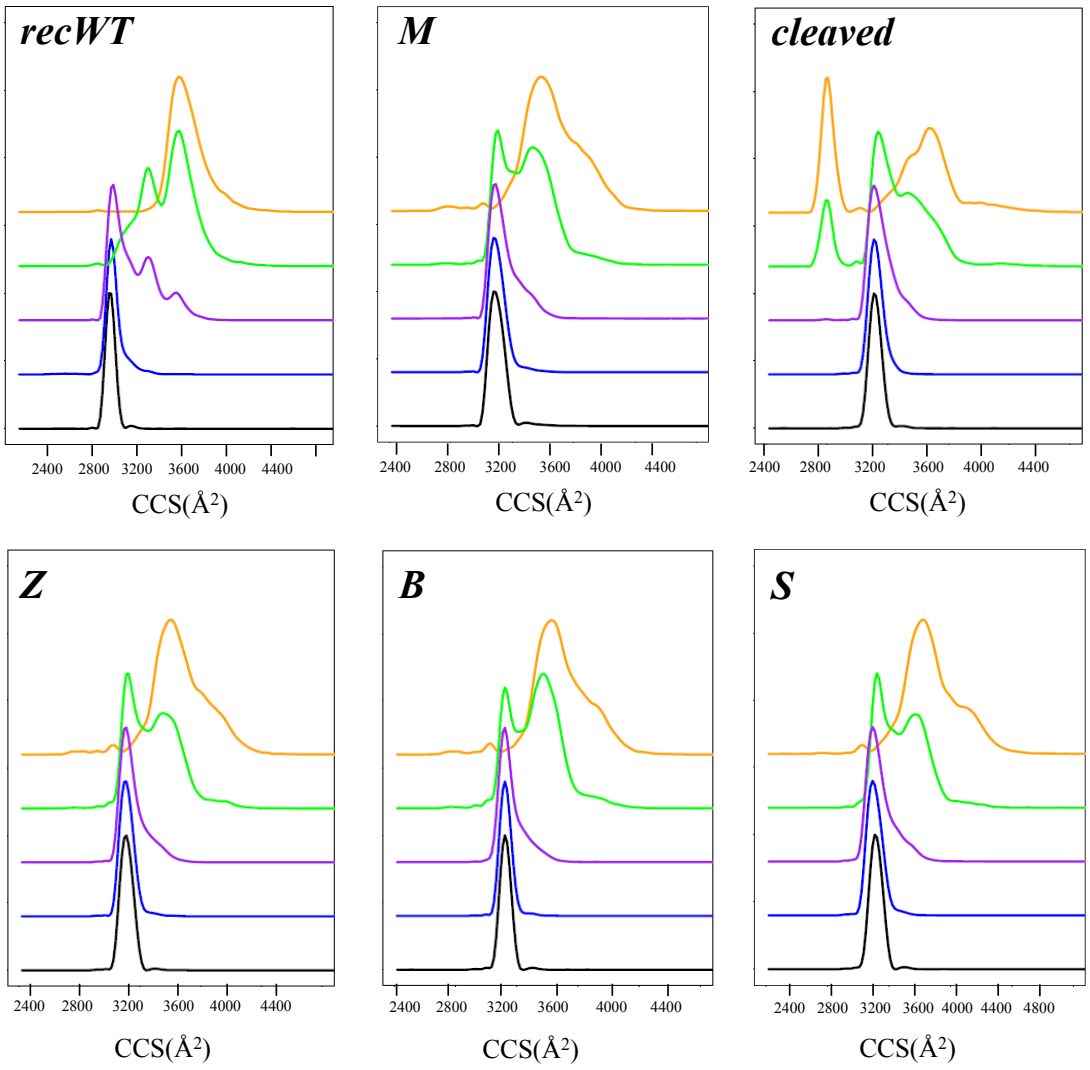


Figure 4.5. Collision Induced Unfolding of alpha$_1$-antitrypsin variants. Each panel corresponds to the collision induced unfolding of a specific variant. The extent of unfolding is indicated in Å^2. The different collision energies are shown; 10eV (black), 20eV (blue), 30eV (purple), 40eV (green), and 50eV (orange).

As the collision energy is ramped up to 40eV the variant occupies the intermediate and unfolded conformations with a 4:3 ratio of folded to intermediate unfolding present. At 50eV the protein is even more unfolded and exhibits a wider unfolding than the other variants with the range of conformations ranging from 3200 Å^2 to 4400 Å^2, 200 Å^2 wider than the other variants.

4.2.4 The Effect of Glycosylation on protein conformation and stability

From the data presented here, the effect of glycosylation was investigated by comparing the recombinant wild-type protein with all the plasma variants. Looking at Fig.4.5 the recombinant CIU fingerprint shows that at 30eV, there are three unfolded species. At 30eV for all other proteins, the predominant species is natively folded. At 40eV, there are two unfolded species with traces of the native fold. All other variants have only begun to unfold and exhibit a natively folded species and a more unfolded species. With 50eV of collision energy, the recombinant protein has unfolded 490 $Å^2$ larger than its native fold compared to plasma M which has only unfolded by 319 $Å^2$ despite its increase in mass.

4.3 Discussion and Conclusions

4.3.1 Glycosylation Analysis by Ion Mobility

Until recently, serpin glycosylation studies were scarce. The interest in glycosylation only re-emerged when studies proved that glycosylation of α_1-AT in augmentation therapy improved the stability and functional efficacy of treatment[127].

Glycosylation plays an important role in the processing and secretion of α_1-AT from the ER in hepatocytes, macrophages, neutrophils and lung epithelial cells. Alpha$_1$-antitrypsin is considered an Acute Phase Protein (APP) because it circulates the plasma and responds to inflammation either qualitatively or quantitatively by changing its glycosylation pattern. Such changes have been linked to inflammatory and malignant conditions and can therefore be used as potential biomarkers for diseases such as cancer.

The theoretical polypeptide masses of the α_1-AT variants are shown in Table 2.2. The mass spectrometry data therefore shows that the least glycosylated peaks, those furthest left, are still glycosylated. The theoretical mass difference between the non-glycosylated M, S, B, and Z were 4863, 5442, 5493, and 5368 *Da* respectively.

Some studies have investigated the Zα_1-AT glycosylation but there have been no previous glycosylation studies on S or B α_1-AT variants so the data cannot be cross validated with other biochemical or biophysical methods. We have however, identified potential glycans attached to these.

Zα_1-AT has a two-sialic acid residue difference between one peak and the other (green and purple). It is unknown whether Zα_1-AT glycosylation plays a role in polymer accumulation in the ER of hepatocytes but evidence shows the complete lack of sialic acid in hepatic polymers compared to those in serum[138]. Studies show that both the core and outer arm fucosylated glycans on Zα_1-AT are elevated compared to wild-type individuals[127]. The large 5368 Da

difference corresponds well to the attachment of a complex sialylated fucosylated tetra-antennary structure, which supports Jeppsson's and McCarthy's studies, and a high mannose structure[139]. The inclusion of high mannose structure has not been identified in previous studies but is the only structure that would satisfy the difference in mass. In 1978, researchers discovered an excess of mannose in Zα_1-AT isolated from the liver[139]. In 1986, Bolmer and Kleinerman isolated and characterised α_1-AT in PAS-positive hepatic granules from rats and discovered that granule α_1-AT had an increased amount of mannose, which is a signal for endoglucasaminidase H cleavage[140]. Both studies report an absence in sialyl residues, something we still observe in our data from the plasma Z variant. High mannose oligosaccharides are uncommon in plasma proteins[141]. Rather, they are often found on the cell surfaces of non-self entities and bind to receptors on macrophages to initiate pathogen phagocytosis[131,142]. Perhaps this contributes to the deficiency seen in patients homozygous for Z α_1-AT due to an increased rate of clearance, affecting the protein's half life. Further studies would be needed to decipher whether this is the case.

Sα_1-AT has three peaks corresponding to a sialylated branch subunit and fucose (Fig.4.3). The large mass difference of 5442 Da from the non-glycosylated Sα_1-AT and the least glycosylated peak is complex sialylated fucosylated tetra-antennary structure. Bα_1-AT has a mass difference corresponding to the sialylated branch subunit and a complex sialylated fucosylated tetra-antennary structure. Thus S, B and Z share the same complex sialylated fucosylated tetra-antennary structure where Z has a high mannose structure and S and B have a common subunit. B has two sialylated branch subunits, S has one plus a fucose residue and Z just has two sialic acids. Wild-type plasma M has mass differences that could be a result of the following carbohydrate residues sialylated branch subunit and a branch subunit with galactose/mannose.

Mass Spectrometry can be used further in combination with Collision Induced Dissociation

to sequence the glycans linked to the protein. To reduce the loss of sialylated glycans in the instrument, the mass spectrometer can be run in negative ion mode[143,144]. Alternatively, sialylated glycans can be stabilised by large amounts of methylation, which enables them to be analysed in positive ion mode alongside neutral atoms. In this way, full coverage of the oligosaccharides is possible.

4.3.2 Comparing the CCS Values of Plasma Variants

Only charge state isolation data was used for the comparison of CCS values and the analysis of CIU graphs. The data shown in Fig.4.3 has been acquired under full scan mode and although is necessary to separate out the differently glycosylated species, the mixture of charge-state species may interfere with the true CCS values, hence, more accurate results are obtained when the quadrupole filters through the desired mass-to-charge.

Recombinant wild-type is the smallest protein with a CCS of 3089 Å, this fits well within the theoretically calculated value of 2601 Å (Projection Approximation) and 3379 Å (EHSS). It also has the smallest range, +-9.9 Å. Given that the recombinant protein is not glycosylated and has a mass 4500 Da less than that of plasma M, it is no surprise that its CCS is smaller. In addition, it has a smaller CCS peak width than the other plasma variants, which might suggest that glycosylation has a large effect on the conformational flexibility of the native fold. Alternatively, peak broadening might be a result of the glycans adopting several conformations without any changes to the core α_1-AT structure. This does not however suggest that it is more stable as a protein. The effects of glycosylation and analysis of CIU are further discussed below.

The compact controls; cleaved M, and latent have larger CCS values than plasma M (Fig.4.4). This at first does not fit in with our hypothesis but it might be that under the mass spectrometer conditions employed here, the RCL is not fully inserted within the s4a region creating a dynamic arm, which increases the variant's rotationally averaged CCS. Another possibility may be due to the insertion of the RCL, β-sheet A expands hence resulting in a slightly larger protein than its

native, metastable fold. Peak width can be used as a measure of conformational flexibility. Despite the CCS of the compact controls being greater than M, their peak widths are smaller, suggesting that they occupy a smaller range of conformations than plasma M. Plasma M interestingly has the largest peak width among all the variants. Perhaps this is due to there being such heterogeneous glycosylation among α_1-AT proteins in healthy patients.

Disease variants also show differences in CCS. B α_1-AT is the largest of all the variants with a CCS of 3223 $Å^2$ followed by Z with a CCS of 3200 $Å^2$ and Sα_1-AT with a CCS of 3185 $Å^2$. Their peak widths are 129 $Å^2$, 145 $Å^2$, and 170 $Å^2$ for B, Z, and S respectively compared to that of plasma M (peak width 151 $Å^2$). The narrowest peak width, B, suggests that this mutation has a stabilising effect on the protein much like the stabilising recombinant variants explored in Chapter 3. Studies by *Haq.I et al.* show that folding is much less impaired in Ala336Pro than the Z variant but that it can populate the polymerogenic state and hence polymerise more readily than either wild-type or Z α_1-AT [88]. Given that Z has the same CCS as both latent and cleaved variants, it is possible that Z has an extended structure allowing it to form the polymerogenic intermediate and that this is even more pronounced in B where the structure is even more expanded and primed to enter a polymerogenic state. S on the contrary has a smaller CCS than both Z and B but is still larger than plasma M. Patients with the Siyama mutation are mildly deficient, have a low risk of lung disease and a medium risk of liver disease. The data presented here supports what is known about the S (E264V) variant already. Engh.R *et al.*[145] show that the mutation results in the loss of internal hydrogen bonds and disulphide bonds from E264 to Y38 and K487 respectively. They also note that although this is the case, no drastic amino acid rearrangement has taken place, which would explain why there is only a slight increase in molecular size compared to the plasma wild-type protein. As I have already alluded, the mutation removes certain intramolecular bonds, which thermally destabilise the protein and may also impair folding. Slower folding can lead to aggregation and perhaps the wider peak for

S α_1-AT is showing the range of conformations being tested before folding into its native structure.

4.3.3 Comparing the Collision Induced Unfolding of α_1-AT Variants

What remains very clear to see is the difference in unfolding pattern between the recombinant protein and the other glycosylated variants (Fig.4.5). Recombinant wild-type has the smallest CCS and is hence more compact than the other proteins. Following an increase in collision energy to 30eV, the protein begins to unfold into three dominant conformations whilst the plasma variants are mostly unaffected by the increase in internal energy. At 40eV the protein unfolds to an even greater extent and reaches the unfolded conformation at 50eV. Compared to its starting conformation, the percentage increase in CCS is 18%, nearly twice as much as all the other variants (CCS increase 8-10%). Given that the conditions employed were all constant, the only difference between the proteins is that recombinant is not glycosylated. Hence, one can hypothesise that the decrease in stability is due to the lack of glycosylation on the protein. Glycosylation studies[128,146] show how they were able to vary the degree of glycosylation to control the thermostability of proteins. Therefore, it is clear that recombinant wild-type is less thermally stable and unfolds differently to its glycosylated counterparts. Studies should therefore aim to use plasma variants more frequently as they most likely represent the situation *in vivo* more accurately.

Cleaved M has been cleaved at the F352-L353 residues allowing the RCL to insert into the s4a position of beta-sheet A. Crystallisation studies[91] of cleaved α_1-AT show that following self-insertion of P14-P7, an intermolecular linkage with a partially occupied β-sheet A of an adjacent molecule can occur by insertion of the P6-P3 residues.

The CIU fingerprint of Cleaved α_1-AT is very interesting to observe as one can see that at high collisional energies, the protein becomes unfolded to the point where the non-covalently

attached Reactive Centre Loop has escaped from the β-sheet s4a position. This occurs strictly after 40eV has been applied. At 40eV there is only a small amount of RCL that has been ejected from the molecule with the highest intensity peak still corresponding the native cleaved α_1-AT fold. At 50eV the natively folded species has further unfolded and most of the protein has ejected its RCL. The CCS of the RCL is 2800 Å despite being so much smaller in mass than the intact protein because it is completely unfolded and hence would have a large rotationally averaged CCS. This is a limitation of Ion Mobility in that it does not truly represent the structure if the analyte in question is not globular.

Plasma M has a relatively larger peak width than Z and B α_1-AT but this is most likely due to the varying degrees of glycosylation[86,127,130]. Much like cleaved, the protein remains stable in the first three increments of collision energy. At 40eV its peak width is much wider and there are likely 2 conformational species. At 50eV the natively folded M is no longer present and the protein is in a more extended structure, occupying a wider range of species due to the reduced stability of the protein.

The unfolding fingerprint of the disease variants agrees well with previous biochemical data and further supports the CCS analysis above. As the B starts to unfold at 40eV, it does so to a much greater extent than Z, S, or M whereby the unfolded species is much more prominent. However, S, with its greater flexibility unfolds to 3761 Å, over 200 Å greater than Z and B where the main peak tops are 3429 Å and 3509 Å respectively. This further supports the findings that S is less thermally stable than the other variants but that B enters the polymerogenic state more readily compared to wild-type α_1-AT and the well-characterised disease variants.

Carrying on this work should involve setting new upper-boundaries for the comparison of the plasma variants as it clear that even at 50eV the proteins are unfolding to an even greater extent. In addition, smaller collision energy intervals should be used especially in the 30-50eV range, this would provide more insight into the unfolding and may reveal further differences between the plasma variants that are not obvious here such as the specific collision energy

voltage at which the protein can no longer maintain its native state. The issue with the above will be acquisition times. Here, 7 samples - excluding the calibrants - were run. Prior to mass spectrometry analysis, the sample preparation takes an hour for every day of analysis. Each sample requires 5 minutes per collision energy setting therefore totalling 25 minutes acquisition time for 5 collision energy increments. For seven, efficiently-run samples repeated twice, the acquisition time would be 7 hours. In reality, there are always problems trying to spray the protein, adjusting the needle, loading the needle etc. so only two repeats were possible. In addition, instrument calibration takes on average 45 minutes. In total, a full day of sample acquisition with barely any problems along the way takes 9 hours without breaks and full supervision. Ideally, one would do this experiment across several days, taking drift time readings with smaller intervals of 2eV, and calibrating each day with proteins of known CCS including plasma M and recombinant wild-type, to get a more accurate CIU fingerprint.

It is worth mentioning that flexibility and protein stability are not entirely the same here. Flexibility of the species refers to the range of conformations the protein is able to acquire. Stability refers to how resilient the protein is when its internal energy increases which is based upon the number and type of intermolecular bonds within the structure. For example, the more glycosylated the species, the more the glycans are able to move about the protein at their points of attachment and hence the wider the CCS distribution. However, as shown, glycosylation increases stability of proteins because they unfold to a lesser extent as a result of increased internal energy.

In conclusion, comparisons between the plasma variants and the recombinant monomer unfolding patterns show the stabilising effect of glycosylation on α_1-AT. The sensitivity of the instrument can help assign mass differences to common glycan ligands, some of which are known to bind to α_1-AT in previous literature. Differences in CCS between the different variants also show slight differences in size and conformational ensemble. Furthermore, CIU of B α_1-AT

reveals that it unfolds more readily than any other disease variant agreeing with previous findings that it populates the polymerogenic state more readily than Z. Given the differences between the recombinant proteins and their plasma counterparts, this study shows that future experiments should focus more on the plasma variants that represent the disease *in vivo*.

Chapter 5- Ion mobility mass spectrometry of ex-vivo *pathogenic polymers*

5.1 Introduction

Following the study of plasma variants, our aim was to analyse α_1-AT polymers to probe their propensity to aggregate via a specific pathway. Monomers were extracted from patient's plasma and heated to induce polymerisation. Polymers formed by heating still retain the 2C1 epitope, proving that polymers formed in this way are relevant to disease.

Patients homozygous for Z have severe deficiency and are likely to develop early onset emphysema and liver cirrhosis. As polymers have not been observed to atomic resolution, two different models of polymerisation are thought to exist, see section 1.6.2.1.

The polymers have been analysed by ion mobility mass spectrometry and the dimer charge state CCSs have been measured and analysed. Differences in the polymer's CCS would go some way into determining whether different variants polymerise via alternative methods.

5.2 Results

5.2.1 Structural Analysis of the polymers by mass spectrometry

The polymers are shown here in Fig.5.1. Each variant sample displays the monomer, dimer, and polymer labelled. The lower end of the spectrum shows the typical charge state distribution for the monomer with the 13, 14, and 15+ charge states. Next, the dimer is seen with varying intensities. The polymer is shown from 6000-12000 *m/z*. The peak heights vary across the different polymers, where wild-type has a higher proportion of smaller oligomeric species compared to the B and Z polymers as shown from the higher intensity peaks at 6000-8000 *m/z*. Conversely, the Baghdad sample contains mostly monomer but the oligomeric states are more evenly distributed from 6000-12000*m/z*.

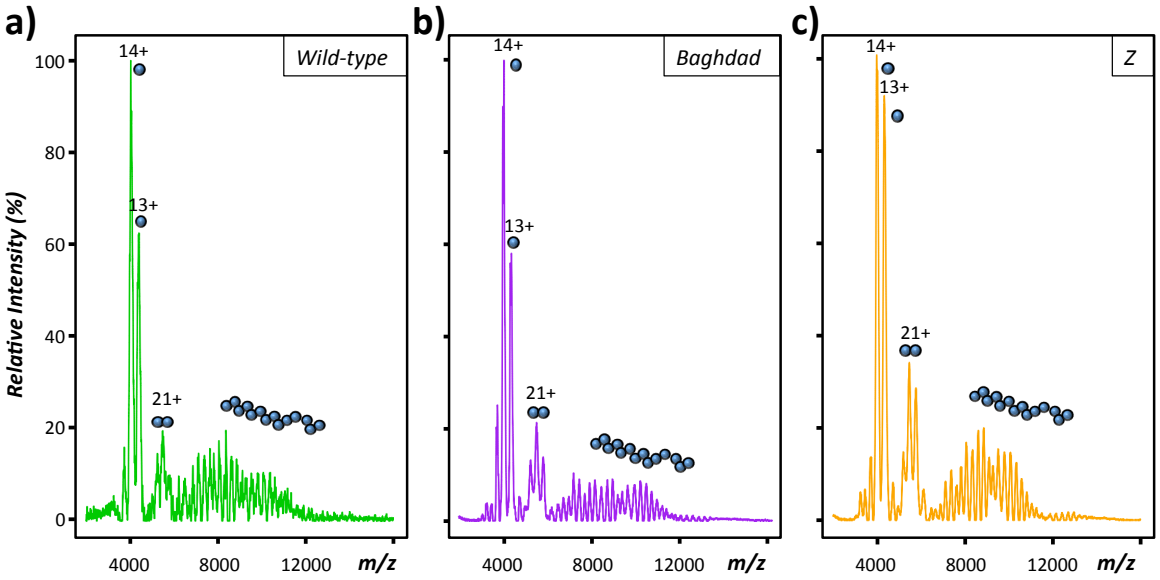


Figure 5.1. Native mass spectra of plasma-derived polymers. **a**) Wild-type polymer mass spectrum. **b**) Baghdad polymer mass spectrum. **c**) Z-polymer mass spectrum. The 13+ and 14+ charge states of the monomer, and the 21+ charge state of the dimer are shown alongside their respective cartoon representations. The peaks between 6000-14000*m/z* represent the different oligomeric states of the polymer.

5.2.2 Structural analysis of the polymers by ion mobility

Focusing on the dimer allows us to explore the structural intermolecular linkages between the monomeric intermediates as they proceed to polymerisation. Fitting the experimental CCSs (Fig.5.2) to current dimer models (Fig.5.3) will provide biophysical data supporting a model of polymerisation, which might differ depending on the variant. Interestingly, few have explored the importance of glycosylation on the structure and its involvement in the disease process. With ion mobility and CIU, one can assess how this particular post-translational modification affects the conformational ensemble and stability of α_1-AT.

Here, the 21+charge state of the dimer has been selected from the full ion mobility spectrum and the different variant dimers have been compared. CCS values are taken from the peak tops, peak widths are taken as the width across the peak at 50% intensity and refer to the conformational flexibility of the species and the standard deviation is taken from the differences in peak tops among the repeats. The mean CCSs of the dimers were 5807 Å^2 (range +-172 Å^2) for wild-type, 5935 Å^2 (range+-234 Å^2) for Baghdad, and 5674 Å^2 (range +-46 Å^2) for Z- α_1-AT. The

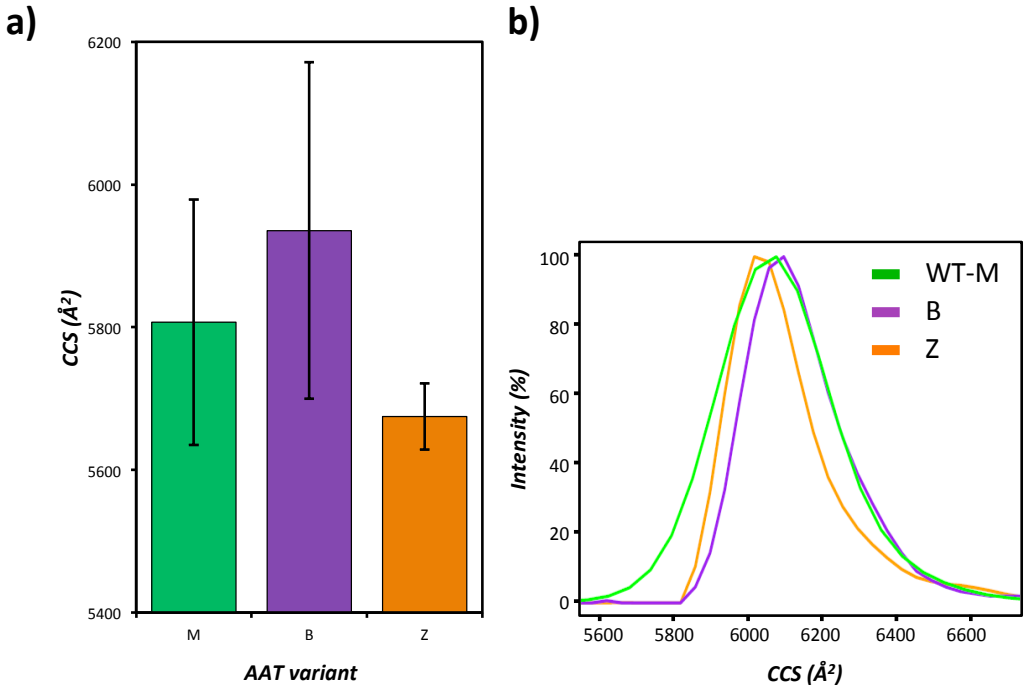


Figure 5.2. Collision Cross Sections of the α_1-AT disease dimers. **a**) Barchart showing the CCS of the dimer disease variants. The average CCS values as shown for the WT-M (green), Baghdad (purple), and Z (orange). Error bars represent thee standard error about the mean. **b**) The CCS distribution of α_1-AT disease dimers. The arrival time distribution was converted to Å^2 for the WT-M (green), Baghdad (purple), and Z (orange).

peak widths of the dimers were 355 Å^2 for wild-type, 280 Å^2 for Baghdad, and 240 Å^2 for Zα_1-AT.

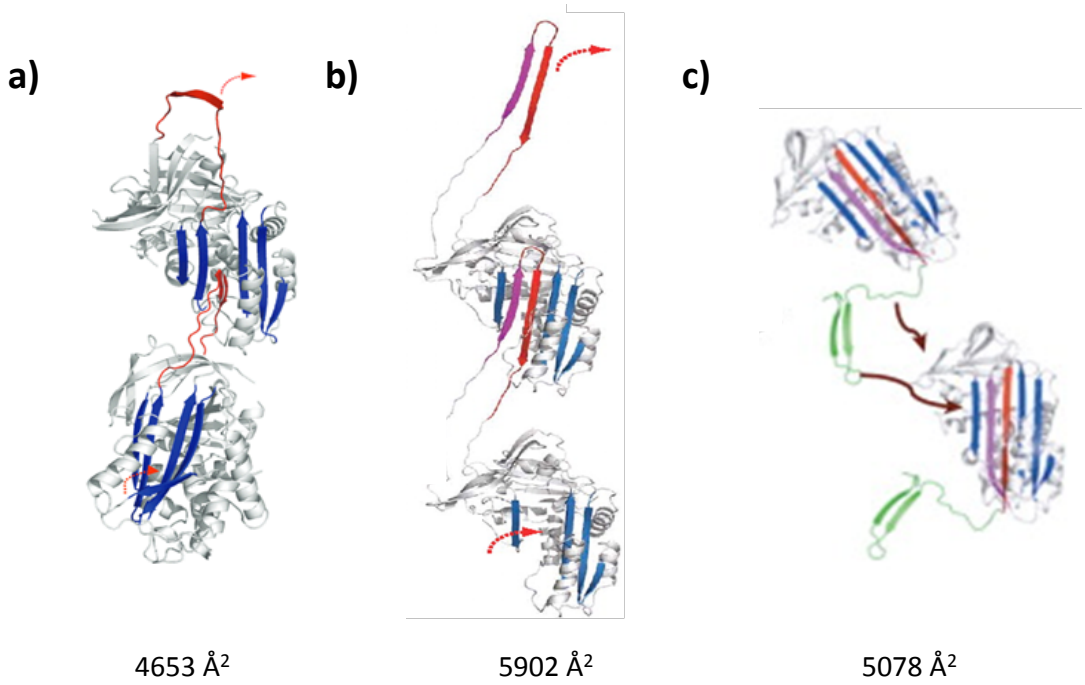


Figure 5.3. The theoretically calculated CCS values for the different polymerisation pathways. **a**) β-sheet A loop insertion with a theoretical CCS value of 4653 Å^2. **b**) The β-hairpin intermediate with insertion of the s4A and s5A strands gives a theoretical CCS value of 5902 Å^2. **c**) The C-sheet insertion model gives a CCS value of 5078 Å^2. All theoretical values were calculated by Projection Approximation. Figure adapted from [94].

5.2.3 Collision Induced Unfolding of Dimer

The 21+ charge state of the wild-type dimer was selected for Collision Induced Unfolding to assess whether this method would be useful in determining any differences between the polymer connections in the wild-type protein compared to the disease variants. The dimer was exposed to trap collision energies of 10-90eV and the arrival time was recorded for each setting, the experiment was repeated twice. The following CIU fingerprint (Fig 5.4) shows where each collision energy increment is shown in a separate colour.

The initial collision energy increase from 10-20eV does not alter the CCS distribution of the wild-type dimer. Exposure to 30eV collision energy starts the unfolding process but the peak

top is still 4765 Å^2. At 40eV there is a significant shift from 4796 Å^2 to 5017 Å^2 with a weak presence of the natively folded dimer. The dimer keeps unfolding from 50eV to 90eV from 5160 Å^2 to 5760 Å^2 resulting in an overall conformational expansion of 20%. 100eV was tested but there was a repeated loss of signal though the data suggests the dimer has not reached its upper unfolding limit. Throughout the increasing energy increments, there was no dissociation of the monomers from each other. Other variants were not tested further due to time restrictions and sample limitation.

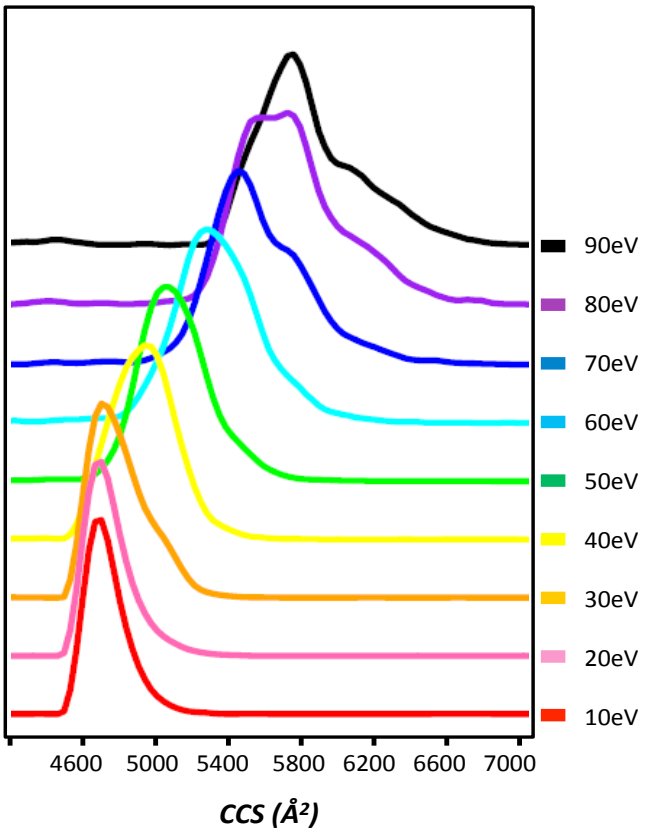


Figure 5.4. The collision induced unfolding of the 21+ charge state isolated dimer. The 21+ charge state of the Wild-type polymer collision cross section is measured against increasing collision energies 10-90eV. CCS values were converted using Amphitrite (Sivalingham G *et al.* 2013) and plotted in R.

5.3 Discussion and Conclusion

5.3.1 Structural Analysis of the polymers by mass spectrometry

Interestingly, the wild-type polymer contains dimer at a much lesser intensity than the other two disease variants. This would confirm that these disease variants form polymers more readily than wild-type α_1-AT. Deconvoluting the mass spectra for such heterogenous samples is difficult due primarily to the overlapping of charge states from different species and the glycosylation states which widen the peaks for those charge states. For example, the difference in mass between a non-glycosylated and a glycosylated 13+ charge state is on average 5000 Da. In a dimer, the non-glycosylated vs glycosylated difference will be twice as great and the peaks further apart, possibly overlapping with a trimer peak. In addition, as the polypeptide chain increases in length, the higher and the greater the number of charge states. If the peaks are too broad, it may comprise trimer, tetramer, or pentamer. A way to overcome this would be to isolate certain *m/z* regions using the quadrupole. The reason this did not occur in the data presented here is due to difficulties in spraying such large complexes, data acquisition times only lasted for a few minutes with low signal. In addition, making direct comparisons with different disease variants is difficult as they have different degrees of glycosylation, which alters the resulting CCS value.

Despite this, one can determine the ratios of the different –mers in the sample. For native wild-type M there is a greater intensity of polymer from 6000-9000 *m/z*, the higher chain polymers had a lower intensity. In the Baghdad mutant, the dimer is very clear and the intensity of the different polymer chains remains evenly distributed throughout the acquisition range. This would suggest that the stability of the polymer until 12000 *m/z* is rather stable and does not change with indirect proportion to the number of protein molecules in the chain. It also might suggest differences in the mutant's dynamics of elongation-nucleation.

The Z mutant has a much greater intensity of dimer and the larger chained polymers show a slight bell-shaped distribution where most of the polymers in the sample are between 4 and 9

protein molecules long. This might give an insight into the type of protein structures present *in vivo*.

In addition, the disease polymers do not show very high intensities of the trimer. This could elude to the chain forming via dimer-dimer association. The trimer is shown in the drift time spectrum at 4800-5600 *m/z* but the peak intensity is much lower in these regions compared to the dimer and the following oligomeric structures. Such polymerization phenomena have already been shown by mass spectrometry such as the formation of a dodecamer from 2 hexamers formed by a further 3 dimers[147].

Appendix A.2 shows the spreadsheet used to calculate the theoretical *m/z* of the different length polymers. The highlighted regions are those actually seen in the mass spectra by cross-validation of drift-time figures, Fig.A.3. From this, a theoretical mass spectrum was created and used to isolate peaks that corresponded to the particular species. This greatly facilitated the Amphitrite software, which had initial difficulties in simulating the spectrum. The final deconvolution of all the peaks is shown in Fig.A.4. However, this is not a finalized model due to the glycosylation of each polymer chain. In monomers, it is relatively easy to assign glycans to the different peaks but with an increase in chain length comes an increase in complexity and factorial increases in possible glycan adducts. Therefore, despite calculating a theoretical mass spectrum, shown in Fig.A.4, compared to the experimental data, the theoretical spectrum is too simple.

Improving the charge state assignment would include the use of the theoretical spectrum and use those *m/z* to isolate specific charge states for specific species. The issue here however is due to the glycosylation, which will change the *m/z.* Therefore, it may be best to acquire over the full range first and then isolate each peak top *m/z* value from the full spectrum to get many individual spectra. In this way, although laborious and time-consuming, for each peak one would be able to know exactly the CCS and whether there are other species with the same *m/z.* This can be technically achieved with a full spectrum ion mobility profile but the issue is the loss

of signal for the very high mass polymers, making it very difficult to assign specific peaks to their oligomer.

5.3.2 Structural analysis of the polymers by ion mobility

As expected, the CCS of all the dimers is less than the CCS values for twice the monomer, 5807 Å^2 (range +-172 Å^2) for wild-type rather than 6330 Å^2 (given that the monomer CCS for wild-type was 3165 Å^2). This is because a portion of the protein is inserted into its neighbour, hence reducing the surface area of the dimer complex. However, they do not agree with the computational values predicted above. Firstly, these models have been calculated without the glycans, which would underestimate the CCS. Secondly, the charge state of these dimer models is unknown and therefore a direct comparison cannot be used. Thirdly, the data has been taken from the full spectrum of the polymer and therefore, its ATD might be affected by the surrounding species and charge states travelling through the drift tube at the same time. The only sample with enough ion transmission to acquire data over a long period of time was the wild-type and therefore the 21+ charge state was isolated. Here, the wild-type dimer had a mean CCS value of 5411 Å^2 following two repeats which was similar to the CCS values extracted from the full spectrum. However, despite being able to isolate the charge state, we cannot yet do a direct comparison with other disease variants, nor can we conclude which polymerisation model best fits our data.

One of the difficulties encountered when attempting to measure the mass and collision cross sections of the polymers is the lack of signal as a result of high amounts of monomer in the sample. There are several explanations for this. Firstly, it is possible that the conditions in the mass spectrometer, despite being 'soft', still caused the dissociation of monomers from the polymers. If the polymer were only one length and only one monomer dissociated for every polymer chains then the ion signal would show a 1:1 ratio of monomer to polymer. This makes sense because for every oligomer chain length in the spectrum, an addition monomer is detected therefore the total monomer is much greater. A useful way of calculating the total

chain length of the polymer would be to find the ratio between the monomer signal and the polymer signal. A downside to this however is that all the spectra have different ratios despite showing polymers with *m/z* ratios of above 12,000. However, as one can see from every spectrum across all polymer samples, the monomer has a much greater intensity than the polymers so the instrument conditions may in fact be knocking off several monomers from the polymer.

Finally, it is likely that the instrument favours transmission of the monomer rather than the huge mass and size of the polymer. The complexity of the disease polymer samples can be significantly reduced if the samples were fractionated into sequential oligomer compositions. In this way, one would be able to see the full charge state distribution for each oligomer without the overlap from other oligomer charge states. In this way, one would easily be able to deconvolute the peaks we see in the full spectrum above (Fig.5.1).

5.3.3 Collision Induced Dissociation of Dimer

CIU of M polymer (Fig.5.4), correlates with known data that polymers are ultra stable under denaturing conditions such as 8M urea[124] and would explain why the dimer did not dissociate at very high collision energies. Despite this, we can go some way into predicting the unfolding of the dimer because, as evidenced, the CCS does become larger with increased internal energy and does occupy a wide range of CCS distributions. If the s4A strand is well inserted into the A-sheet then this part of the protein must be retaining its native fold because the monomers are not covalently bonded. The internal energy must be affecting other parts of the dimer other than the area of the dimer linkage.

To date, no papers provide information of the melting temperature of the species. Further experiments using mass spectrometry could investigate the effect of high temperatures on α_1-AT by increasing the source temperature. Performing these experiments on different dimer variants would go some way into explaining how the dimers unfold and whether there are any

differences between different disease mutants. This would reinforce the need for more individualised therapeutic approaches.

5.3.4 α_1-Antitrypsin Polymer Structure Determination

Determining the structure of α_1-Antitrypsin polymer from *ex vivo* patient samples and purified recombinant protein would enable us to determine a CCS of the polymer and may help us understand the pathway to polymerisation.

Cross-linking studies on native structures have been achieved using distance restraints (Chen and Herzog, 2014) but are challenging and computationally intense. Identifying the 2C1 antibody epitope by cross-linking would prove invaluable given that this antibody is specific only to polymerogenic variants. Consequently, we might be closer to understanding the changes α_1-Antitrypsin undergoes upon polymer formation and the differences in exposed regions of the disease variant on its way to polymerisation.

To summarise, the instrument was able to identify polymer chains up to 9 monomers long. The samples are complex due to glycosylated and potentially overlapping species at the higher *m/z* region but the intensities of the different polymer lengths provide a first insight into potential differences in nucleation-elongation. So far, no significant differences are observed between the variant dimer CCS values but CIU of the M α_1-AT dimer agrees with previously published data that the polymer linkage is exceptionally stable

Chapter 6- Final Remarks

The project here started with recombinant proteins and ended with the *ex vivo* samples. The results here confirm known crystallisation data [92,94]. It also confirms the importance of mass spectrometry in collecting low-resolution information about the structural conformations of proteins and provides important information about its stability by either exploiting the trap collision energy or by changing the temperature in the instrument to measure the protein's thermostability.

Continuing on with structural determination, further IM and CIU studies on the different charge states of the dimer across the different variants could have yielded valuable information on the stability of the polymer linkage. In addition, I would have been keen to study α_1-AT with cross-linking studies. Crosslinkers are capable of linking amino acid residues close in space and reveal a lot of information about the residues linked together in space. Given the number of cross-linkers available with specific spacer lengths it would have been interesting to explore and compare their efficacies of cross-linking and the information retrieved. Furthermore, *in vivo* cross-linkers are relatively new but have huge potential. The experiment would involve using a cross-linker specific for alpha$_1$-antitrypsin, allow the disease variant to be expressed and observe α_1-AT's interactions with other proteins. Depending on the mutation, different interactions and complexes may form; enhancing our knowledge on the effects disease mutations have in the cell. This principle can be applied to time-dependent studies, yielding even more information on the effects of disease variants in cells.

The information we are able to deduce on the protein's behaviour due to its structure is vital in understanding more about how the protein interacts in the system and how these interactions change with disease. Nowadays, the overall protein database contains just under 116,000 protein structures solved by X-ray crystallography and NMR[148]. The wealth of knowledge acquired about the structures of the proteins in our world gives us good ground to

understanding health and disease. As no protein acts alone, the real challenge remains in linking these individual pieces of information together to build a coherent picture. A downside with any research is the depth of knowledge; often individuals will focus specifically on their topic and choice of instrumentation that other techniques are dismissed. Thus, a more collaborative, multidisciplinary approach is necessary to achieve full understanding of a disease, with many biological, chemical, and physical techniques involved.

In the new age of science, with increased accessibility to newly published research and the ability to connect with many different scientific communities, mass spectrometry is well placed. It has earnt its place as a valued instrument across many different scientific fields but it has a huge range of diversity in the type of data it acquires and the experiments it is capable of. Its only limitation is the software designed to analyse the data, which will continue to improve with advancements in technology and the increasing number of computer scientists making a transition to the more traditional experimental fields of science.

This thesis presents the effects of mutations in the serpin, alpha$_1$-antitrypsin, which is highly important in maintaining the homeostasis of neutrophil elastase in the body to prevent excessive degradation of cellular tissue. Native MS was used to determine the mass of different variants and the glycan compositions as well as their oligomerisation states, stoichiometries and went some way into observing the subunit connectivity of the disease polymers. IM-MS was used to observe differences in collisional cross section between different α_1-AT variants and to start probing the dimer formed between two α_1-AT monomers in the hope of confirming the model of polymerisation.

The initial questions posed at the beginning of the thesis would not have been possible without mass spectrometry and its ability to cope with and analyse the conformational states of species in a heterogenous sample. Future experiments will involve mass spectrometry to quantify and qualify ligand binding, detect cross-linked residues to determine protein inter- and intra-molecular bonds and formulate an interactome for specific proteins as well as determine

insights of disease proteomes. As a biological technique, MS's ability to adapt to the growing complexities faced in understanding disease makes it an indispensable tool for solving them.

Finally, this body of work clearly shows that gas-phase stability correlates with published literature on the thermal stability of recombinant, plasma[153,154] and polymer samples[137]. The data also shows how post-translational modifications such as glycosylation increase the gas-phase stability of α_1-AT. Lastly, native MS revealed polymer chains up to nine monomers in length. Selecting the dimer from this heterogeneous sample for CIU further proved that the oligomeric linkages are exceptionally stable going some way into explaining the presence of oligomers *in vivo.* Hence, mass spectrometry is well suited to further study the role of α_1-AT mutations in disease.

Chapter 7- References

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

2. Mann, M. *et al.* Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trends Biotechnol.* **20,** 261–268 (2002).

3. Biemann, K. Four decades of structure determination by mass spectrometry: From alkaloids to heparin. *J. Am. Soc. Mass Spectrom.* **13,** 1254–1272 (2002).

4. Uetrecht, C. *et al.* High-resolution mass spectrometry of viral assemblies: molecular composition and stability of dimorphic hepatitis B virus capsids. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 9216–9220 (2008).

5. Uetrecht, C. *et al.* Stability and shape of hepatitis B virus capsids in vacuo. *Angew. Chemie - Int. Ed.* **47,** 6247–6251 (2008).

6. Griffiths, J. A brief history of mass spectrometry. *Analytical Chemistry* **80,** 5678–5683 (2008).

7. Nier, A. O. Some reminiscences of mass spectrometry and the Manhattan Project. *J. Chem. Educ.* **66,** 385 (1989).

8. De Laeter, J. & Kurz, M. D. Alfred Nier and the sector field mass spectrometer. *Journal of Mass Spectrometry* **41,** 847–854 (2006).

9. Hipple, J. A., Sommer, H. & Thomas, H. A. A precise method of determining the faraday by magnetic resonance. *Physical Review* **76,** 1877–1878 (1949).

10. Fenn, J. B. Ion formation from charged droplets: Roles of geometry, energy, and time. *J.*

Am. Soc. Mass Spectrom. **4,** 524–35 (1993).

11. Beavis, R. C. & Chait, B. T. Matrix-assisted laser-desorption mass spectrometry using 355 nm radiation. *Rapid Commun. Mass Spectrom.* **3,** 436–9 (1989).

12. Bich, C. & Zenobi, R. Mass spectrometry of large complexes. *Current Opinion in Structural Biology* **19,** 632–639 (2009).

13. Iavarone, A. T., Jurchen, J. C. & Williams, E. R. Effects of solvent on the maximum charge state and charge state distribution of protein ions produced by electrospray ionization. *J. Am. Soc. Mass Spectrom.* **11,** 976–85 (2000).

14. Karas, M., Bahr, U. & Dülcks, T. Nano-electrospray ionization mass spectrometry: addressing analytical problems beyond routine. *Fresenius. J. Anal. Chem.* **366,** 669–676 (2000).

15. Kebarle, P. & Tang, L. From Ions in Solution to Ions in the Gas Phase - The Mechanism of Electrospray Mass Spectrometry. *Anal. Chem.* **65,** 972A–986A (1993).

16. Tang, L. & Kebarle, P. Dependence of ion intensity in electrospray mass spectrometry on the concentration of the analytes in the electrosprayed solution. *Anal.Chem.* **65,** 3654–3668 (1993).

17. Wilm, M. S. & Mann, M. Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last? *Int. J. Mass Spectrom. Ion Process.* **136,** 167–180 (1994).

18. Bahr, U., Pfenninger, A., Karas, M. & Stahl, B. High-sensitivity analysis of neutral underivatized oligosaccharides by nanoelectrospray mass spectrometry. *Anal. Chem.* **69,** 4530–4535 (1997).

19. Wilm, M. & Mann, M. Analytical properties of the nanoelectrospray ion source. *Anal. Chem.* **68,** 1–8 (1996).

20. Mann, M. & Wilm, M. Electrospray Mass Spectrometry for Protein Characterisation. *Trends Biochem. Sci.* **20,** 219–224 (1995).

21. Juraschek, R., Dülcks, T. & Karas, M. Nanoelectrospray--more than just a minimized-flow electrospray ionization source. *J. Am. Soc. Mass Spectrom.* **10,** 300–8 (1999).

22. Dole, M. *et al.* Molecular Beams of Macroions. *J Chem.Phys.* **49,** 2240–2249 (1968).

23. Mack, L. L., Kralik, P., Rheude, A. & Dole, M. Molecular Beams of Macroions. II. *J. Chem. Phys.* **52,** 4977 (1970).

24. Banerjee, S. & Mazumdar, S. Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. *Int. J. Anal. Chem.* **2012,** 1–40 (2012).

25. Fernandez De La Mora, J. Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Anal. Chim. Acta* **406,** 93–104 (2000).

26. Konermann, L., Ahadi, E., Rodriguez, A. D. & Vahidi, S. Unraveling the mechanism of electrospray ionization. *Anal. Chem.* **85,** 2–9 (2013).

27. Konermann, L., Rodriguez, A. D. & Liu, J. On the formation of highly charged gaseous ions from unfolded proteins by electrospray ionization. *Anal. Chem.* **84,** 6798–6804 (2012).

28. Sysoev, A. A. Time-of-flight analysers with sector fields: advances and prospects. *Eur. J. Mass Spectrom.* **6,** 501–513 (2000).

29. Glish, G. L. & Burinsky, D. J. Hybrid Mass Spectrometers for Tandem Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **19,** 161–172 (2008).

30. Marshall, A. G. & Hendrickson, C. L. High-resolution mass spectrometers. *Annu. Rev. Anal. Chem. (Palo Alto. Calif).* **1,** 579–599 (2008).

31. Wiley, W. C. & McLaren, I. H. Time-of-Flight Mass Spectrometer with Improved Resolution. *Rev. Sci. Instrum.* **26,** 1150 (1955).

32. Paul, W. & Steinwedel, H. Ein neues Massenspektrometer ohne Magnetfeld. *Zeitschrift fur Naturforschung - Section A Journal of Physical Sciences* **8,** 448–450 (1953).

33. Chernushevich, I. V., Loboda, A. V. & Thomson, B. A. An introduction to quadrupole-time-of-flight mass spectrometry. *J. Mass Spectrom.* **36,** 849–865 (2001).

34. Steen, H., Küster, B. & Mann, M. Quadrupole time-of-flight versus triple-quadrupole mass spectrometry for the determination of phosphopeptides by precursor ion scanning. *J. Mass Spectrom.* **36,** 782–790 (2001).

35. Wolff, M. M. & Stephens, W. E. A pulsed mass spectrometer with time dispersion. *Rev. Sci. Instrum.* **24,** 616–617 (1953).

36. Cameron, A. E. & Eggers, D. F. An ion 'velocitron'. *Rev. Sci. Instrum.* **19,** 605–607 (1948).

37. Lehmann, W. D. Jürgen H. Gross: Mass spectrometry–A Textbook, 2nd ed. *Analytical and Bioanalytical Chemistry* **401,** 3033–3034 (2011).

38. Mamyrin, M. A. Laser assisted reflectron time-of-flight mass spectrometry. *Int. J. Mass Spectrom.* **131,** 1–19 (1994).

39. Ladislas Wiza, J. Microchannel plate detectors. *Nucl. Instruments Methods* **162,** 587–601 (1979).

40. Ruotolo, B. T. *et al.* Ion mobility-mass spectrometry reveals long-lived, unfolded intermediates in the dissociation of protein complexes. *Angew. Chem. Int. Ed. Engl.* **46,** 8001–4 (2007).

41. Kanu, A. B., Dwivedi, P., Tam, M., Matz, L. & Hill, H. H. Ion mobility-mass spectrometry.

Journal of Mass Spectrometry **43,** 1–22 (2008).

42. Thalassinos, K. *et al.* Ion mobility mass spectrometry of proteins in a modified commercial mass spectrometer. *Int. J. Mass Spectrom.* **236,** 55–63 (2004).

43. Karpas, Z. The structure and mobility in air of protonated ketones. *Int. J. Mass Spectrom. Ion Process.* **107,** 435–440 (1991).

44. Konijnenberg, A., Butterer, A. & Sobott, F. Native ion mobility-mass spectrometry and related methods in structural biology. *Biochim. Biophys. Acta* **1834,** 1239–56 (2013).

45. Sitkiewicz, E., Kłoniecki, M., Poznański, J., Bal, W. & Dadlez, M. Factors influencing compact-extended structure equilibrium in oligomers of aβ1-40 peptide--an ion mobility mass spectrometry study. *J. Mol. Biol.* **426,** 2871–85 (2014).

46. Clemmer, D. E. & Jarrold, M. F. Ion mobility measurements and their applications to clusters and biomolecules. *Journal of Mass Spectrometry* **32,** 577–592 (1997).

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

48. Kaddis, C. S. *et al.* Sizing Large Proteins and Protein Complexes by Electrospray Ionization Mass Spectrometry and Ion Mobility. *J. Am. Soc. Mass Spectrom.* **18,** 1206–1216 (2007).

49. Hall, Z., Politis, A. & Robinson, C. V. Structural modeling of heteromeric protein complexes from disassembly pathways and ion mobility-mass spectrometry. *Structure* **20,** 1596–1609 (2012).

50. Lanucara, F., Holman, S. W., Gray, C. J. & Eyers, C. E. The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics.

Nat. Chem. **6,** 281–294 (2014).

51. Thalassinos, K. & Scrivens, J. H. Applications of traveling wave ion mobility-mass spectrometry. in *Pract. Aspects Trapped Ion Mass Spectrom.* 205–236 (2009).

52. Scarff, C. A., Snelling, J. R., Knust, M. M., Wilkins, C. L. & Scrivens, J. H. New structural insights into mechanically interlocked polymers revealed by ion mobility mass spectrometry. *J. Am. Chem. Soc.* **134,** 9193–9198 (2012).

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

54. Smith, D. P. *et al.* Deciphering drift time measurements from travelling wave ion mobility spectrometry-mass spectrometry studies. *Eur. J. Mass Spectrom. (Chichester, Eng).* **15,** 113–130 (2009).

55. Salbo, R. *et al.* Traveling-wave ion mobility mass spectrometry of protein complexes: Accurate calibrated collision cross-sections of human insulin oligomers. *Rapid Commun. Mass Spectrom.* **26,** 1181–1193 (2012).

56. Forsythe, J. G. *et al.* Collision cross section calibrants for negative ion mode traveling wave ion mobility-mass spectrometry. *Analyst* **140,** 6853–6861 (2015).

57. Michaelevski, I., Kirshenbaum, N. & Sharon, M. T-wave ion mobility-mass spectrometry: basic experimental procedures for protein complex analysis. *J. Vis. Exp.* 1–7 (2010). doi:10.3791/1985

58. Sun, Y., Vahidi, S., Sowole, M. A. & Konermann, L. Protein Structural Studies by Traveling Wave Ion Mobility Spectrometry: A Critical Look at Electrospray Sources and Calibration

Issues. *J. Am. Soc. Mass Spectrom.* **27,** 31–40 (2016).

59. Bush, M. F. *et al.* Collision cross sections of proteins and their complexes: A calibration framework and database for gas-phase structural biology. *Anal. Chem.* **82,** 9557–9565 (2010).

60. Tao, L., McLean, J. R., McLean, J. A. & Russell, D. H. A Collision Cross-Section Database of Singly-Charged Peptide Ions. *J. Am. Soc. Mass Spectrom.* **18,** 1232–1238 (2007).

61. Qin, J. & Chait, B. T. Collision-induced dissociation of singly charged peptide ions in a matrix-assisted laser desorption ionization ion trap mass spectrometer. *Int. J. Mass Spectrom.* **190-191,** 313–320 (1999).

62. Sivalingam, G. N., Yan, J., Sahota, H. & Thalassinos, K. Amphitrite: A program for processing travelling wave ion mobility mass spectrometry data. *Int. J. Mass Spectrom.* **345-347,** 54–62 (2013).

63. Merenbloom, S. I., Koeniger, S. L., Bohrer, B. C., Valentine, S. J. & Clemmer, D. E. Improving the efficiency of IMS-IMS by a combing technique. *Anal. Chem.* **80,** 1918–1927 (2008).

64. Gologan, B. *et al.* Ion soft-landing into liquids: Protein identification, separation, and purification with retention of biological activity. *J. Am. Soc. Mass Spectrom.* **15,** 1874–1884 (2004).

65. Siuzdak, G. *et al.* Mass spectrometry and viral analysis. *Chem. Biol.* **3,** 45–48 (1996).

66. Nibbering, N. M. M. The McLafferty rearrangement: A personal recollection. *Journal of the American Society for Mass Spectrometry* **15,** 956–958 (2004).

67. Jennings, K. R. Collision-induced decompositions of aromatic molecular ions. *Int. J. Mass Spectrom. Ion Phys.* **1,** 227–235 (1968).

68. Sharon, M. How far can we go with structural mass spectrometry of protein complexes? *J. Am. Soc. Mass Spectrom.* **21,** 487–500 (2010).

69. Benesch, J. L. P. Collisional Activation of Protein Complexes: Picking Up the Pieces. *J. Am. Soc. Mass Spectrom.* **20,** 341–348 (2009).

70. Madsen, J. A. & Brodbelt, J. S. Asymmetric charge partitioning upon dissociation of DNA duplexes. *J. Am. Soc. Mass Spectrom.* **21,** 1144–1150 (2010).

71. Leitner, A. *et al.* Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol. Cell. Proteomics* **9,** 1634–1649 (2010).

72. Rappsilber, J. The beginning of a beautiful friendship: Cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J. Struct. Biol.* **173,** 530–540 (2011).

73. Sinz, A. Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes. *J. Mass Spectrom.* **38,** 1225–37 (2003).

74. Vasilescu, J., Guo, X. & Kast, J. Identification of protein-protein interactions using in vivo cross-linking and mass spectrometry. *Proteomics* **4,** 3845–3854 (2004).

75. Schmidt, C. & Robinson, C. V. A comparative cross-linking strategy to probe conformational changes in protein complexes. *Nat. Protoc.* **9,** 2224–36 (2014).

76. Shvartsburg, a a & Smith, R. D. Fundamentals of Traveling Wave Ion Mobility Spectrometry. *Anal. Chem.* **80,** 9689–9699 (2008).

77. Hopper, J. T. S. & Oldham, N. J. Collision Induced Unfolding of Protein Ions in the Gas Phase Studied by Ion Mobility-Mass Spectrometry: The Effect of Ligand Binding on Conformational Stability. *J. Am. Soc. Mass Spectrom.* **20,** 1851–1858 (2009).

78. Tian, Y., Han, L., Buckner, A. C. & Ruotolo, B. T. Collision Induced Unfolding of Intact Antibodies: Rapid Characterization of Disulfide Bonding Patterns, Glycosylation, and Structures. *Anal. Chem.* **87,** 11509–11515 (2015).

79. Covey, T. & Douglas, D. J. Collision cross sections for protein ions. *J. Am. Soc. Mass Spectrom.* **4,** 616–623 (1993).

80. Freeke, J., Bush, M. F., Robinson, C. V. & Ruotolo, B. T. Gas-phase protein assemblies: Unfolding landscapes and preserving native-like structures using noncovalent adducts. *Chem. Phys. Lett.* **524,** 1–9 (2012).

81. Hyung, S. J., Robinson, C. V. & Ruotolo, B. T. Gas-Phase Unfolding and Disassembly Reveals Stability Differences in Ligand-Bound Multiprotein Complexes. *Chem. Biol.* **16,** 382–390 (2009).

82. Landreh, M. *et al.* Controlling release, unfolding and dissociation of membrane protein complexes in the gas phase through collisional cooling. *Chem. Commun.* **51,** 15582–15584 (2015).

83. Eschweiler, J. D., Rabuck-Gibbons, J. N., Tian, Y. & Ruotolo, B. T. CIUSuite: A Quantitative Analysis Package for Collision Induced Unfolding Measurements of Gas-Phase Protein Ions. *Anal. Chem.* **87,** 11516–11522 (2015).

84. Allison, T. M. *et al.* Quantifying the stabilizing effects of protein-ligand interactions in the gas phase. *Nat. Commun.* **6,** 8551 (2015).

85. Stockley, R. A. Alpha1-antitrypsin review. *Clinics in Chest Medicine* **35,** 39–50 (2014).

86. Blanchard, V. *et al.* N-glycosylation and biological activity of recombinant human alpha1-antitrypsin expressed in a novel human neuronal cell line. *Biotechnol. Bioeng.* **108,** 2118–2128 (2011).

87. Griese, M. *et al.* alpha1-Antitrypsin inhalation reduces airway inflammation in cystic fibrosis patients. *Eur. Respir. J.* **29,** 240–50 (2007).

88. Haq, I. *et al.* Deficiency Mutations of α 1 -antitrypsin Differentially Affect Folding, Function and Polymerization. *Am. J. Respir. Cell Mol. Biol.* **54,** 150619071832005 (2015).

89. Stoller, J. K. & Aboussouan, L. S. Alpha1-antitrypsin deficiency. *Lancet (London, England)* **365,** 2225–36 (2005).

90. Laurell, C. B. & Erikkson, S. Electrophoretic Alpha1-globulin pattern of serum in Alpha1-antitrypsin Deficiency. *Scand. J. Clin. Lab. Investig.* **15,** 132–& (1963).

91. Huntington, J. A. *et al.* A 2.6 A structure of a serpin polymer and implications for conformational disease. *J. Mol. Biol.* **293,** 449–55 (1999).

92. Loebermann, H., Tokuoka, R., Deisenhofer, J. & Huber, R. Human α1-proteinase inhibitor. *J. Mol. Biol.* **177,** 531–557 (1984).

93. Ryu, S. E., Choi, H. J., Kwon, K. S., Lee, K. N. & Yu, M. H. The native strains in the hydrophobic core and flexible reactive loop of a serine protease inhibitor: crystal structure of an uncleaved alpha1-antitrypsin at 2.7 A. *Structure* **4,** 1181–1192 (1996).

94. Gooptu, B. *et al.* Crystallographic and cellular characterisation of two mechanisms stabilising the native fold of alpha1-antitrypsin: implications for disease and drug design. *J. Mol. Biol.* **387,** 857–68 (2009).

95. Qu, D., Teckman, J. H., Omura, S. & Perlmutter, D. H. Degradation of a mutant secretory protein, ?? 1-antitrypsin Z, in the endoplasmic reticulum requires proteasome activity. *J. Biol. Chem.* **271,** 22791–22795 (1996).

96. Janciauskiene, S. M. *et al.* The discovery of ??1-antitrypsin and its role in health and disease. *Respiratory Medicine* **105,** 1129–1139 (2011).

97. Higashiyama, M., Doi, O., Kodama, K., Yokouchi, H. & Tateishi, R. An evaluation of the prognostic significance of alpha-1-antitrypsin expression in adenocarcinomas of the lung: an immunohistochemical analysis. *Br. J. Cancer* **65,** 300–302 (1992).

98. Dolmer, K. & Gettins, P. G. W. How the serpin α 1-proteinase inhibitor folds. *J. Biol. Chem.* **287,** 12425–12432 (2012).

99. Luisetti, M. & Seersholm, N. Alpha1-antitrypsin deficiency. 1: epidemiology of alpha1-antitrypsin deficiency. *Thorax* **59,** 164–9 (2004).

100. Lomas, D. a, Evans, D. L., Finch, J. T. & Carrell, R. W. The mechanism of Z alpha 1-antitrypsin accumulation in the liver. *Nature* **357,** 605–607 (1992).

101. Propst, T. *et al.* Prevalence of hepatocellular carcinoma in alpha-1-antitrypsin deficiency. *J. Hepatol.* **21,** 1006–1011 (1994).

102. Zhou, H. & Fischer, H. P. Liver carcinoma in PiZ alpha-1-antitrypsin deficiency. *Am. J. Surg. Pathol.* **22,** 742–8 (1998).

103. Topic, A., Ljujic, M. & Radojkovic, D. Alpha-1-antitrypsin in pathogenesis of hepatocellular carcinoma. *Hepat. Mon.* **12,** (2012).

104. Perlmutter, D. H. Pathogenesis of chronic liver injury and hepatocellular carcinoma in alpha-1-antitrypsin deficiency. *Pediatr. Res.* **60,** 233–238 (2006).

105. Parmar, J. S. *et al.* Polymers of ??1-antitrypsin are chemotactic for human neutrophils: A new paradigm for the pathogenesis of emphysema. *Am. J. Respir. Cell Mol. Biol.* **26,** 723–730 (2002).

106. Ordóñez, A. *et al.* Endoplasmic reticulum polymers impair luminal protein mobility and sensitise to cellular stress in α (1) -antitrypsin deficiency. *Hepatology* (2012). doi:10.1002/hep.26173

107. Dunstone, M. a *et al.* Cleaved antitrypsin polymers at atomic resolution. *Protein Sci.* **9,** 417–420 (2000).

108. Schulze AJ1, Frohnert PW, Engh RA, H. R. Evidence for the extent of insertion of the active site loop of intact .alpha.1 proteinase inhibitor in .beta.-sheet A. *Biochemistry* **31,** 7560–5 (1992).

109. Yamasaki, M., Li, W., Johnson, D. J. D. & Huntington, J. a. Crystal structure of a stable dimer reveals the molecular basis of serpin polymerization. *Nature* **455,** 1255–1258 (2008).

110. Huntington, J. A. & Whisstock, J. C. Molecular contortionism - On the physical limits of serpin 'loop-sheet' polymers. in *Biological Chemistry* **391,** 973–982 (2010).

111. Lomas, D. a & Carrell, R. W. Serpinopathies and the conformational dementias. *Nat. Rev. Genet.* **3,** 759–768 (2002).

112. Nyon, M. P. *et al.* Structural dynamics associated with intermediate formation in an archetypal conformational disease. *Structure* **20,** 504–512 (2012).

113. James, E. L. & Bottomley, S. P. The mechanism of alpha 1-antitrypsin polymerization probed by fluorescence spectroscopy. *Arch. Biochem. Biophys.* **356,** 296–300 (1998).

114. Lomas, D. A. *et al.* ??1-antitrypsin Mmalton (Phe52-deleted) forms loop-sheet polymers in vivo: Evidence for the C sheet mechanism of polymerization. *J. Biol. Chem.* **270,** 16864–16870 (1995).

115. Ekeowa, U. I. *et al.* Defining the mechanism of polymerization in the serpinopathies. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 17146–17151 (2010).

116. Mehmood, S. *et al.* Charge reduction stabilizes intact membrane protein complexes for mass spectrometry. *J. Am. Chem. Soc.* **136,** 17010–17012 (2014).

117. Flotte, T. R., Mueller, C. & Flotte, T. R. Gene-based therapy for alpha-1 antitrypsin deficiency. *COPD* **10 Suppl 1,** 44–49 (2013).

118. Yusa, K. *et al.* Targeted gene correction of α1-antitrypsin deficiency in induced pluripotent stem cells. *Nature* **478,** 391–394 (2011).

119. Pastore, N. *et al.* Gene transfer of master autophagy regulator TFEB results in clearance of toxic protein and correction of hepatic disease in alpha-1-anti-trypsin deficiency. *EMBO Mol. Med.* **5,** 397–412 (2013).

120. Chang, Y. P., Mahadeva, R., Chang, W. S. W., Lin, S. C. & Chu, Y. H. Small-molecule peptides inhibit Z ??1-antitrypsin polymerization. *J. Cell. Mol. Med.* **13,** 2304–2316 (2009).

121. Cruz, P. E. *et al.* In vivo post-transcriptional gene silencing of alpha-1 antitrypsin by adeno-associated virus vectors expressing siRNA. *Lab. Invest.* **87,** 893–902 (2007).

122. Guo, S. *et al.* Antisense oligonucleotide treatment ameliorates alpha-1 antitrypsin-related liver disease in mice. *J. Clin. Invest.* **124,** 251–261 (2014).

123. Gore, M. G. Adverse effects of gene therapy: Gene therapy can cause leukaemia: no shock, mild horror but a probe. *Nature* **10,** (2003).

124. Nyon, M. P. *et al.* An integrative approach combining ion mobility mass spectrometry, X-ray crystallography, and nuclear magnetic resonance spectroscopy to study the conformational dynamics of α(1)-antitrypsin upon ligand binding. *Protein Sci.* **24,** 1301–1312 (2015).

125. Lusch, A. *et al.* Development and analysis of alpha 1-antitrypsin neoglycoproteins: The impact of additional N -glycosylation sites on serum half-life. *Mol. Pharm.* **10,** 2616–2629 (2013).

126. Berthelier, V., Harris, J. B., Estenson, K. N. & Baudry, J. Discovery of an inhibitor of Z-

alpha1 antitrypsin polymerization. *PLoS One* **10,** (2015).

127. McCarthy, C. *et al.* The role and importance of glycosylation of acute phase proteins with focus on alpha-1 antitrypsin in acute and chronic inflammatory conditions. *Journal of Proteome Research* **13,** 3131–3143 (2014).

128. Broersen, K., Voragen, A. G. J., Hamer, R. J. & De Jongh, H. H. J. Glycoforms of ??-Lactoglobulin with Improved Thermostability and Preserved Structural Packing. *Biotechnol. Bioeng.* **86,** 78–87 (2004).

129. Öberg, F. *et al.* Glycosylation Increases the thermostability of human aquaporin 10 protein. *J. Biol. Chem.* **286,** 31915–31923 (2011).

130. Kolarich, D., Weber, A., Turecek, P. L., Schwarz, H. P. & Altmann, F. Comprehensive glyco-proteomic analysis of human ??1- antitrypsin and its charge isoforms. *Proteomics* **6,** 3369–3380 (2006).

131. An, H. J., Peavy, T. R., Hedrick, J. L. & Lebrilla, C. B. Determination of N-Glycosylation Sites and Site Heterogeneity in Glycoproteins. *Anal. Chem.* **75,** 5628–5637 (2003).

132. Lomas, D. A. Molecular mousetraps, alpha1-antitrypsin deficiency and the serpinopathies. *Clinical medicine (London, England)* **5,** 249–257 (2005).

133. Stoller, J. K. & Aboussouan, L. S. Alpha1-antitrypsin deficiency. *Lancet* **365,** 2225–36 (2005).

134. Strange, C. *et al.* Genetic testing for Alpha1-antitrypsin deficiency. *Genet. Med.* **6,** 204–210 (2004).

135. Chapman, K. R., Stockley, R. a, Dawkins, C., Wilkes, M. M. & Navickis, R. J. Augmentation therapy for alpha1 antitrypsin deficiency: a meta-analysis. *COPD* **6,** 177–84 (2009).

136. Miranda, E. *et al.* A novel monoclonal antibody to characterize pathogenic polymers in liver disease associated with alpha1-antitrypsin deficiency. *Hepatology* **52,** 1078–88 (2010).

137. Haq, I. *et al.* Deficiency Mutations of Alpha-1 Antitrypsin. Effects on Folding, Function, and Polymerization. *Am. J. Respir. Cell Mol. Biol.* **54,** 71–80 (2015).

138. Jeppsson, J. O., Larsson, C., & Eriksson, S. Characterization of α1-antitrypsin in the inclusion bodies from the liver in α1-antitrypsin deficiency. *N. Engl. J. Med.* **293,** 567–579 (1975).

139. Hercz, A., Katona, E., Cutz, E., Wilson, J. R. & Barton, M. alpha1-Antitrypsin: the presence of excess mannose in the Z variant isolated from liver. *Science (80-.).* **201,** 1229–1232 (1978).

140. Bolmer, S. & Kleinerman, J. Isolation and characterization of alpha 1-antitrypsin in PAS-positive hepatic granules from rats with experimental alpha 1-antitrypsin deficiency. *Am J Pathol* **123,** 377–389 (1986).

141. Putman, F. *The Plasma Proteins V5: Structure, Function, and Genetic Control*. (1984).

142. An, H. J., Froehlich, J. W. & Lebrilla, C. B. Determination of glycosylation sites and site-specific heterogeneity in glycoproteins. *Current Opinion in Chemical Biology* **13,** 421–426 (2009).

143. Harvey, D. J. *et al.* Travelling-wave ion mobility and negative ion fragmentation of high-mannose N-glycans. *J. Mass Spectrom.* **51,** 219–235 (2016).

144. Harvey, D. J. *et al.* Travelling wave ion mobility and negative ion fragmentation for the structural determination of N-linked glycans. *Electrophoresis* **34,** 2368–2378 (2013).

145. Engh, R. *et al.* The S variant of human alpha 1-antitrypsin, structure and implications for

function and metabolism. *Protein Eng.* **2,** 407–15 (1989).

146. Chen, Y., Chen, X., Guo, T. L. & Zhou, P. Improving the thermostability of β-lactoglobulin via glycation: The effect of sugar structures. *Food Res. Int.* **69,** 106–113 (2015).

147. Bernstein, S. L. *et al.* Amyloid-β protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease. *Nat. Chem.* **1,** 326–331 (2009).

148. H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P. E. B. The Protein Data Bank. *Nucleic Acids Research* 28: 235–242 (2000). at <www.rcsb.org>

149. Taylor, T. & Langley, J. The CHROMacademy Essential Guide. *CHROMacademy* at <http://www.chromacademy.com/Mass_Analyzers_Traps_and_TOFs_The_Essential_Guide.html>

150. Mullin, L. & Ladak, A. Increasing Sensitivity for Tof-MS Detection of Polychlorinated Biphenyls (PCBs) Using Tof MRM. *Waters Corp.* (2015). at <http://www.waters.com/webassets/cms/library/docs/720005358en.pdf>

151. Yamasaki, M., Sendall, T. J., Pearce, M. C., Whisstock, J. C. & Huntington, J. a. Molecular basis of α1-antitrypsin deficiency revealed by the structure of a domain-swapped trimer. *EMBO Rep.* **12,** 1011–1017 (2011).

152. Gooptu, B., Dickens, J. A. & Lomas, D. A. The molecular and cellular pathology of α1-antitrypsin deficiency. *Trends in Molecular Medicine* **20,** 116–127 (2014).

153. Kwon, K. S., Yu, M. H. Effect of Glycosylation on the stability of α1-antitrypsin toward urea denaturation and thermal deactivation. *Biochim. Biophys. Acta.* **3**, 265-272 (1997)

154. Kwon, K. S., Kim, J., Shin, H.S., Yu, M.H. Single amino acid substitutions of alpha 1-antitrypsin that confer enhancement in thermal stability. *J Biol Chem*. **269,** 9627-9631(1994)

155. Wyttenbach, T., von Helden, G., Batka, J.J., Carlat, D., Bowers, M. T. Effect of the Long-Range Potential on Ion Mobility Measurements. *J Am Soc Mass Spectrom*. **8**, 275(1997)

156. Bush, M. F. Collision Cross Section Database. *Biophys. Bioanal. Mass. Spec.* (2013) *http://depts.washington.edu/bushlab/ccsdatabase/*

Chapter 8- Appendix

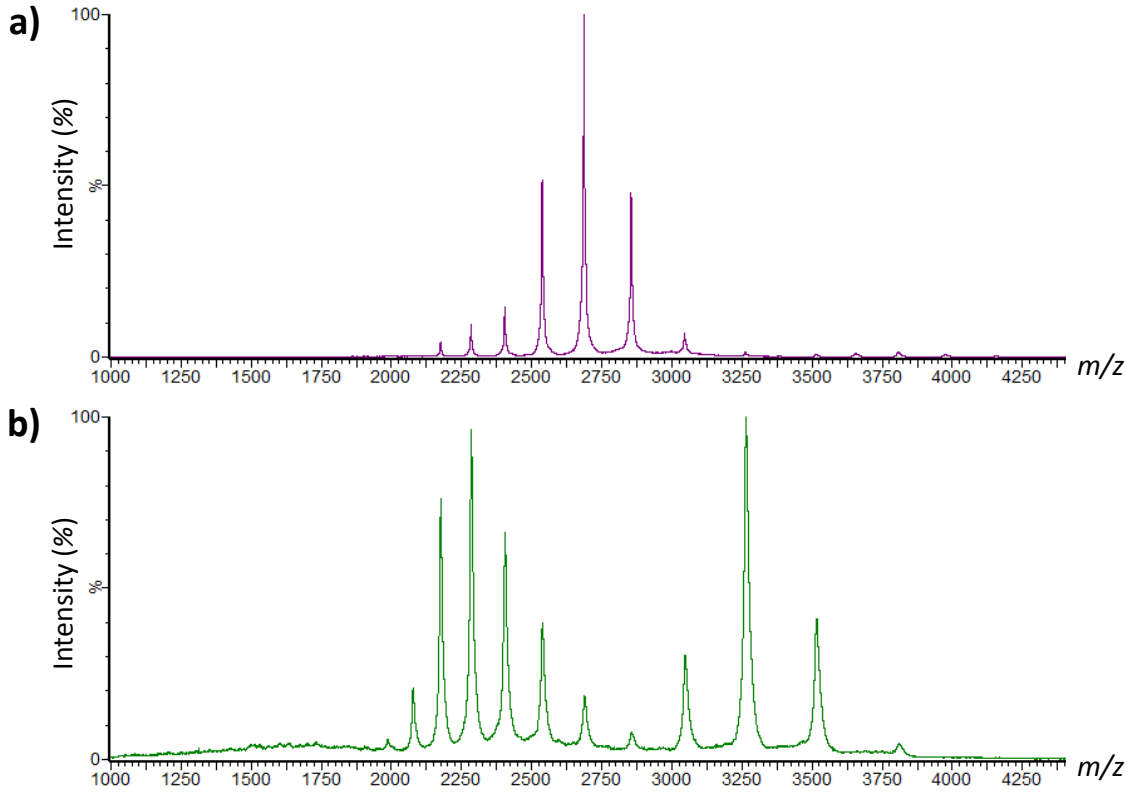


Figure A.1. Equilibration of G117F monomer. **a**) The initial spectrum following addition of the sample directly after the buffer exchange process. **b**) Following equilibration, the G117F variant now resembles the distinct 12-15+ charge states seen for all other variants.

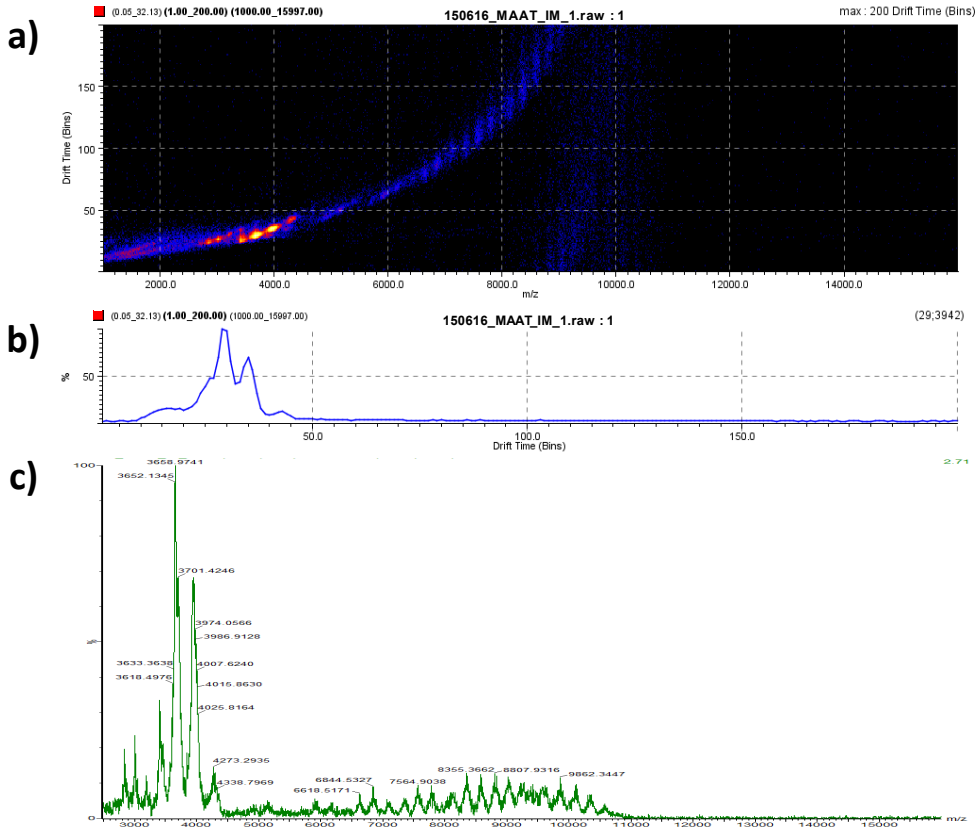


Figure A.2. Plasma M-derived polymer Ion Mobility Data in MassLynx. **a**) ATD vs *m/z* in driftscope. **b**) The arrival time distribution of M-derived polymer. **c**) The native mass spectrum of M-polymer.

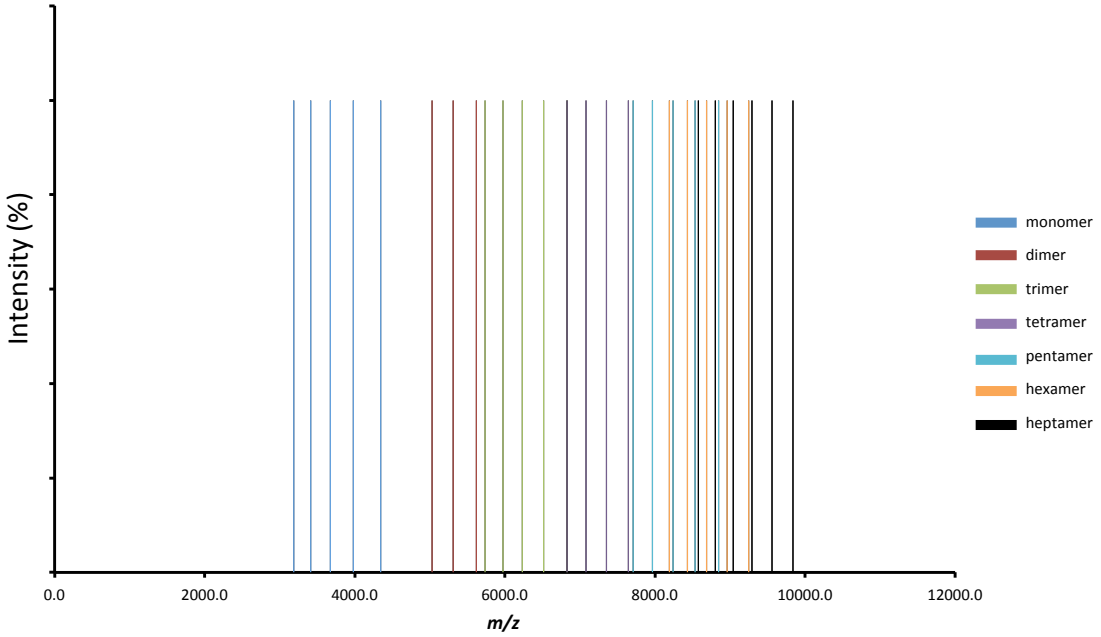


Figure A.3. The theoretical mass spectrum of the polymer. The oligomeric states for the monomer (blue), dimer (red), trimer (green), tetramer (purple), pentamer (cyan), hexamer (orange), and heptamer (black). The peaks are calculated from the excel spreadsheet (Fig.A.3).