

University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

**A Thesis Submitted for the Degree of PhD at the University of Warwick**

<http://go.warwick.ac.uk/wrap/34610>

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.

# Investigating Protein Structure by Means of Mass Spectrometry

Charlotte A. Scarff BSc (Hons)

A thesis submitted for the degree of Doctor of Philosophy

University of Warwick  
Department of Biological Sciences

July 2010

*'For the things we have to learn before we can do, we learn by doing.'*

*Aristotle*

*'Read not to contradict and confute, not to believe and take for granted, not to find  
talk and discourse, but to weigh and consider.'*

*Sir Francis Bacon*

# Table of Contents

<b>List of Figures</b> .....	<b>vi</b>
<b>List of Tables</b> .....	<b>xiii</b>
<b>Acknowledgements</b> .....	<b>xiv</b>
<b>Declaration</b> .....	<b>xv</b>
<b>Summary</b> .....	<b>xvi</b>
<b>Abbreviations</b> .....	<b>xvii</b>
<b>Chapter One: Introduction</b> .....	<b>1</b>
<b>1.1 Mass spectrometry</b> .....	<b>2</b>
1.1.1 What is a mass spectrometer? .....	3
1.1.2 Ionisation methods .....	3
1.1.3 Mass analysers .....	8
1.1.4 Detectors .....	12
1.1.5 Tandem mass spectrometry.....	14
1.1.6 Ion mobility mass spectrometry .....	17
1.1.7 Travelling-wave ion mobility mass spectrometry.....	19
1.1.8 Travelling-wave calibration .....	24
<b>1.2 Investigating Protein Structure</b> .....	<b>27</b>
1.2.1 What is protein structure? .....	27
1.2.2 Protein structure vs. protein conformation.....	28
1.2.3 Classic tools of structural biology.....	28
<b>1.3 Analysing protein structure by mass spectrometry</b> .....	<b>30</b>
1.3.1 Probing three-dimensional protein conformation with ESI-MS .....	30
1.3.2 Studying protein interactions, dynamics and complexes .....	31
1.3.3 The role of IM-MS .....	33
1.3.4 Solution-phase vs. gas-phase structure .....	34

1.3.5 Preserving protein structure in the gas phase .....	36
1.3.6 Native MS as a structural biology tool.....	37
<b>1.4 Aims and Objectives .....</b>	<b>39</b>
<b>1.5 Research Papers .....</b>	<b>40</b>
<b>1.6 Conference papers (Peer-reviewed).....</b>	<b>40</b>
<b>1.7 Oral Presentations.....</b>	<b>42</b>
<b>1.8 References .....</b>	<b>43</b>
<b>Chapter 2: Travelling-Wave Ion Mobility Mass Spectrometry Studies of Protein Structure .....</b>	<b>52</b>
<b>2.1 Introduction .....</b>	<b>53</b>
<b>2.2 Materials and Methods .....</b>	<b>55</b>
2.2.1 Material suppliers.....	55
2.2.2 Sample preparation.....	55
2.2.3 Travelling-wave ion mobility mass spectrometry experiments .....	56
2.2.4 ATDs, calibration and estimation of cross-sections.....	56
2.2.5 Theoretical cross-section calculations.....	59
<b>2.3 Results and Discussion.....</b>	<b>61</b>
2.3.1 Theoretical cross-section calculations.....	61
2.3.2 TWIM calibration.....	62
2.3.3 Estimated cross-sections .....	65
2.3.4 Unfolding restraints.....	69
2.3.5 Biological significance.....	81
<b>2.4 Conclusions .....</b>	<b>83</b>
<b>2.5 References .....</b>	<b>85</b>
<b>Chapter 3: Probing Hemoglobin Structure .....</b>	<b>88</b>
<b>3.1 Introduction .....</b>	<b>89</b>
<b>3.2 Materials and Methods.....</b>	<b>91</b>

3.2.1 Samples and sample preparation .....	91
3.2.2 Ion mobility mass spectrometry .....	91
<b>3.3 Results and Discussion .....</b>	<b>93</b>
3.3.1 Instrument acquisition parameters .....	93
3.3.2 Calibration .....	93
3.3.3 ESI-MS spectra .....	93
3.3.4 Alpha and beta monomers .....	95
3.3.5 Apo- or holo-? .....	96
3.3.6 Heme-deficient dimer .....	97
3.3.7 Hemoglobin tetramer assembly .....	98
3.3.8 Hb A vs Hb SS .....	99
<b>3.4 Conclusions .....</b>	<b>103</b>
<b>3.5 References .....</b>	<b>104</b>
<b>Chapter 4: A Quantitative and Conformational Study of the Autophosphorylation of the Histidine Kinase VanS .....</b>	<b>106</b>
<b>4.1 Introduction .....</b>	<b>107</b>
4.1.1 Histidine kinases .....	108
4.1.2 Vancomycin and vancomycin resistance .....	108
4.1.3 The mechanism of vancomycin resistance .....	110
4.1.4 The VanS-VanR two-component system .....	112
<b>4.2 Materials and Methods .....</b>	<b>115</b>
4.2.1 Expression and purification of VanS <sub>A</sub> truncates .....	115
4.2.2 Sample preparation for MS .....	116
4.2.3 MS analysis .....	116
4.2.4 Phosphorylation analysis .....	117
4.2.5 Modelling the mechanism of autokinase activity .....	118
<b>4.3 Results and Discussion .....</b>	<b>119</b>

4.3.1 ESI-MS spectra .....	119
4.3.2 Autokinase activity.....	121
4.3.3 Conformational studies .....	127
4.3.4 Model of VanS <sub>A</sub> autokinase activity .....	131
<b>4.4 Conclusions .....</b>	<b>134</b>
<b>4.5 References .....</b>	<b>135</b>
<b>Chapter Five: Diagnosis of Hemoglobinopathies .....</b>	<b>138</b>
<b>5.1 Introduction .....</b>	<b>139</b>
5.1.1 Hemoglobin.....	139
5.1.2 Hemoglobin disorders .....	139
5.1.3 Incidence of hemoglobinopathies .....	142
5.1.4 Hemoglobinopathy screening and diagnosis.....	143
5.1.5 Identification of Hb variants using mass spectrometry.....	146
5.1.6 Clinical screening by MS? .....	149
<b>5.2 Materials and Methods .....</b>	<b>152</b>
5.2.1 Sample preparation for Q-TOF MS analysis .....	152
5.2.2 Q-TOF mass spectrometric analysis .....	152
5.2.3 Testing the Q-TOF MS approach to hemoglobinopathy diagnosis .....	153
5.2.4 Ion mobility mass spectrometry .....	154
5.2.5 High-throughput screening .....	155
<b>5.3 Results and Discussion.....</b>	<b>158</b>
5.3.1 Case Study One: Patient A.....	161
5.3.2 Case Study Two: Patient B .....	166
5.3.3 Case Study Three: Patient C .....	172
5.3.4 High-throughput screening .....	177
<b>5.4 Conclusions .....</b>	<b>183</b>
<b>5.5 References .....</b>	<b>185</b>

<b>Chapter Six: Conclusions and Future Work.....</b>	<b>188</b>
<b>6.1 Investigating three-dimensional protein structure by means of mass spectrometry .....</b>	<b>189</b>
6.1.1 Biological significance.....	189
6.1.2 Protein unfolding dynamics .....	189
6.1.3 Hemoglobin analysis.....	190
6.1.4 VanS studies.....	190
6.1.5 Future directions.....	190
<b>6.2 Diagnosis of hemoglobinopathies by means of mass spectrometry .....</b>	<b>191</b>
6.2.1 Identification of rare variants .....	191
6.2.2 High-throughput screening .....	191
6.2.3 Future directions.....	192
<b>6.3 Concluding remarks .....</b>	<b>192</b>
<b>6.4 References .....</b>	<b>193</b>

# List of Figures

<b>Figure 1.1:</b> Schematic representation of a modern mass spectrometer.....	3
<b>Figure 1.2:</b> Schematic representation of the proposed MALDI ionisation process...5	
<b>Figure 1.3:</b> Schematic representation of the two proposed ESI ionisation mechanisms. Adapted from Kerbale (2000). .....	7
<b>Figure 1.4:</b> Schematic representation of a quadrupole mass analyser. ....	9
<b>Figure 1.5:</b> Schematic illustration of a reflectron TOF mass analyser.....	11
<b>Figure 1.6:</b> Schematic representations of a.) a discrete-dynode EM b.) a continuous-dynode EM and c.) a MCP detector. Adapted from Koppenaal <i>et al.</i> (2005).....	13
<b>Figure 1.7:</b> Peptide ion nomenclature, as proposed by Roepstorff, Fohlmann and Biemann. N-terminal fragment ions are termed a, b or c ions whilst C-terminal fragments are termed x, y or z ions. A subscript denotes the number of residues within a fragment ion. ....	15
<b>Figure 1.8:</b> The main scan modes in tandem mass spectrometry.....	16
<b>Figure 1.9:</b> A schematic representation of a stacked ring ion guide. ....	20
<b>Figure 1.10:</b> Schematic illustration of mobility separation in a TWIG. ....	21
<b>Figure 1.11:</b> Schematic representation of the Synapt instrument with an enlargement of the TriWave ion mobility device inset. Adapted from Pringle <i>et al.</i> (2007). ....	22
<b>Figure 1.12:</b> Stepwise evolution of globular protein structure after ESI. ....	35
<b>Figure 2.1:</b> a.) A typical spectrum obtained for sperm whale myoglobin, for calibration purposes, with b.) corresponding extracted arrival time distributions for charge states $[M + 11H]^{11+}$ to $[M + 16H]^{16+}$ . ....	58
<b>Figure 2.2:</b> Calibration curve of corrected arrival times for sperm whale myoglobin against normalised published cross-sections.....	59
<b>Figure 2.3:</b> Comparison of estimated cross-sections for sperm whale myoglobin and equine myoglobin obtained under denaturing conditions. ....	63
<b>Figure 2.4:</b> Comparison of published with experimentally estimated cross-sections for different charge states of equine cytochrome c (under denaturing conditions)....	64
<b>Figure 2.5:</b> Experimentally estimated cross-sections for different charge states of a.) chicken and b.) human lysozyme c under denaturing and near-physiological	

conditions with average theoretical cross-sections for PA and TM models shown. Error bars are shown to illustrate measurements that would be within $\pm 2$ %.....	66
<b>Figure 2.6:</b> Experimentally estimated cross-sections for different charge states of equine cytochrome c under denaturing and near-physiological conditions with average theoretical cross-sections for PA and TM models shown. Error bars are shown to illustrate measurements that would be within $\pm 2$ %.....	67
<b>Figure 2.7:</b> Experimentally estimated cross-sections for different charge states of equine myoglobin under denaturing and near-physiological conditions with average theoretical cross-sections for PA and TM models shown. Error bars are shown to illustrate measurements that would be within $\pm 2$ %.....	68
<b>Figure 2.8:</b> Experimentally estimated cross-sections for different charge states of sperm whale myoglobin under denaturing and near-physiological conditions with average theoretical cross-sections for PA and TM models shown. Error bars are shown to illustrate measurements that would be within $\pm 2$ %.....	69
<b>Figure 2.9:</b> Experimentally estimated cross-sections for different charge states of chicken lysozyme c, equine cytochrome c and equine myoglobin.....	70
<b>Figure 2.10:</b> Chicken lysozyme c oxidised in 10 mM ammonium acetate (top spectrum), oxidised in 50 % ACN 0.2 % HCOOH (middle spectrum) and reduced in 50 % ACN 0.2 % HCOOH (bottom spectrum) with deconvoluted spectra inset. ....	72
<b>Figure 2.11:</b> Arrival time distributions for charges states $[M + 6H]^{6+}$ to $[M + 10H]^{10+}$ for chicken lysozyme c oxidised in 10 mM ammonium acetate (top), oxidised in 50 % ACN 0.2 % HCOOH (middle) and reduced in 50 % ACN 0.2 % HCOOH (bottom). ....	74
<b>Figure 2.12:</b> Experimentally estimated cross-sections for different charge states of chicken lysozyme c under near-physiological conditions, denaturing conditions and under disulphide-reduced denaturing conditions. ....	75
<b>Figure 2.13:</b> Human lysozyme c oxidised in 10 mM ammonium acetate (top spectrum), oxidised in 50 % ACN 0.2 % HCOOH (middle spectrum) and reduced in 50 % ACN 0.2 % HCOOH (bottom spectrum) with deconvoluted spectra inset. ....	76
<b>Figure 2.14:</b> Arrival time distributions for charges states $[M + 6H]^{6+}$ to $[M + 10H]^{10+}$ for human lysozyme c oxidised in 10 mM ammonium acetate (top), oxidised in 50 % ACN 0.2 % HCOOH (middle) and reduced in 50 % ACN 0.2 % HCOOH (bottom).....	77

<b>Figure 2.15:</b> Experimentally estimated cross-sections for different charge states of human lysozyme c under near-physiological conditions, denaturing conditions and under disulphide-reduced denaturing conditions. ....	78
<b>Figure 2.16:</b> Experimentally estimated cross-sections for different charge states of disulphide-reduced chicken lysozyme c, equine cytochrome c and equine myoglobin. ....	80
<b>Figure 2.17:</b> Estimations of native cross-sections for proteins studied calculated theoretically and experimentally and plotted against molecular weight. ....	81
<b>Figure 3.1:</b> Representation of hemoglobin tetramer structure. $\alpha$ -chains are displayed in red and $\beta$ -chains in yellow. Heme groups are shown in green. ....	89
<b>Figure 3.2:</b> Mass spectra of normal (Hb A) and sickle (Hb SS) hemoglobin analyzed by ESI-TOF-MS under near-physiological conditions. ....	94
<b>Figure 3.3:</b> Average estimated cross-sections for charge states of a.) apo- $\alpha$ and holo- $\alpha$ monomers and b.) apo- $\beta$ and holo- $\beta$ monomers observed within three datasets. ...	96
<b>Figure 3.4:</b> Mass spectra of a.) commercially-sourced Hb A and b.) fresh Hb A analysed by ESI-TOF-MS under near-physiological conditions. Spectra are labelled with charge states of heterodimer (D), apo- and holo-monomers (subscripts “a” and “h” refer to apo- and holo-forms, respectively) and heme-deficient dimer. ....	97
<b>Figure 3.5:</b> Average estimated cross sections for holo- $\alpha$ , holo- $\beta$ , heterodimer, and Hb A tetramer, from three datasets. ....	98
<b>Figure 3.6:</b> Arrival time distributions for charge states states $[M + 15H]^{15+}$ to $[M + 18H]^{18+}$ of Hb SS (top) and Hb A (bottom). ....	99
<b>Figure 3.7:</b> Estimated cross sections for Hb A and Hb SS tetramers for three different charge states, showing averaged values from three datasets with corresponding errors. ....	100
<b>Figure 3.8:</b> Average estimated cross-sections for $\beta^a$ , $\beta^h$ , $\beta_s^a$ and $\beta_s^h$ monomers, from three datasets. ....	102
<b>Figure 3.9:</b> Average estimated cross sections for heterodimers $\alpha^h\beta^h$ and $\alpha^h\beta_s^h$ , from three datasets. ....	102
<b>Figure 4.1:</b> The structure of vancomycin. Adapted from Walsh et al. (1996). ....	109
<b>Figure 4.2:</b> The mode of action of vancomycin. ....	111
<b>Figure 4.3:</b> Mechanism of induction of vancomycin resistance by the VanS <sub>A</sub> -VanR <sub>A</sub> two-component system. A signal is received by VanS <sub>A</sub> stimulating autophosphorylation on a conserved histidine residue. VanS <sub>A</sub> transfers this	

phosphate to a conserved aspartic acid residue of VanR<sub>A</sub>. Once phosphorylated, VanR<sub>A</sub> activates transcription of the genes which confer vancomycin resistance. Adapted from Healy et al. (2000). ..... 112

**Figure 4.4:** Schematic of VanS<sub>A</sub> domain organisation based upon protein sequence information. Characteristic H, N, G1, F and G2 motifs, common to sensor histidine kinases are highlighted. Adapted from (Depardieu *et al.* 2003). ..... 114

**Figure 4.5:** Representation of the crystal structure of the cytoplasmic domain of a sensor histidine kinase (HK853) from *T. maritima* (PDB; 2C2A). The two HK853 monomers (blue and green) exhibit a small dimerisation region within which the conserved histidine residue is shown in red. A non-hydrolysable ATP analog is located in each ATP binding site within each CA domain and is shown in orange. 118

**Figure 4.6:** Comparison of ESI-MS spectra obtained for VanS<sub>A</sub> truncates a.) Δ110 b.) Δ140 c.) Δ150 and d.) Δ155 with, inset, corresponding deconvoluted spectra respectively. *m/z* species representative of dimer are illustrated by an asterisk (\*). 119

**Figure 4.7:** Mass spectra obtained for VanS<sub>A</sub>Δ110 samples allowed to phosphorylate for a.) 0, b.) 20, and c.) 60 minutes respectively. Inset spectra show a magnified region of each spectra containing the dominant charge state [M + 12H]<sup>12+</sup>. A decrease in the intensity of peaks relating to non-phosphorylated protein is observed and an increase in the intensity of peaks relating to phosphorylated protein is observed over time. .... 121

**Figure 4.8:** Deconvoluted mass spectra obtained for VanS<sub>A</sub>Δ110 samples allowed to phosphorylate for a.) 0, b.) 20, and c.) 60 minutes respectively. VanS<sub>A</sub>Δ110-P is observed at a mass of 32208 Da (± 1 Da), 80 Da higher than that of VanS<sub>A</sub>Δ110 (32128 ± 1 Da). Additional peaks observed within spectra are consistent with those expected to be observed for sodiated protein species. .... 122

**Figure 4.9:** The percentage of total protein phosphorylated after different time intervals for protein batch one (■) and protein batch two (▲) respectively. The averages of phosphorylation levels calculated, from three datasets for protein batch two, are shown. Error bars indicate the maximum and minimum values calculated for time points with three replicates. .... 123

**Figure 4.10:** Deconvoluted mass spectra obtained for VanS<sub>A</sub>Δ110 samples a.) batch one and b.) batch two allowed to autophosphorylate overnight. VanS<sub>A</sub>Δ110-P is observed at a mass of 32207 Da (± 1 Da), approximately 80 Da higher than that of

VanS <sub>A</sub> Δ110 (32128 ± 1 Da ). Batch one is virtually 100 % phosphorylated whilst batch two exhibits only 69 % phosphorylation.....	124
<b>Figure 4.11:</b> The percentage of total protein phosphorylated after different time intervals for protein batch one, measured by MS (■) and by means of a radioactive labelling assay (▲) respectively.....	125
<b>Figure 4.12:</b> Deconvoluted mass spectrum obtained for VanS <sub>A</sub> Δ155 allowed to autophosphorylate overnight. VanS <sub>A</sub> Δ155-P is observed at a mass of 26997 Da (± 1 Da), approximately 80 Da higher than that of VanS <sub>A</sub> Δ155 (26917 ± 1 Da ). Approximately 90 % of the protein is phosphorylated. ....	126
<b>Figure 4.13:</b> Comparison of ESI-MS spectra obtained for VanS <sub>A</sub> truncates a.) Δ140 b.) Δ150 c.) Δ155 at 10 μM in 10 mM ammonium bicarbonate pH 7.8 following incubation with ATP and MgCl <sub>2</sub> for 1 minute at room temperature with, inset, corresponding deconvoluted spectra. Additional peaks (•) are observed at a mass 529 Da (± 1 Da) higher than those representing each protein species. ....	128
<b>Figure 4.14:</b> Extracted arrival time distributions for the most abundant charge state observed within spectra for a.) VanS <sub>A</sub> Δ140, b.) VanS <sub>A</sub> Δ150, c.) VanS <sub>A</sub> Δ155, d.) VanS <sub>A</sub> Δ140-ATP, e.) VanS <sub>A</sub> Δ150-ATP and f.) VanS <sub>A</sub> Δ155-ATP. ....	129
<b>Figure 4.15:</b> Extracted arrival time distributions for the most abundant charge state ([M+12H] <sup>12+</sup> ) for VanS <sub>A</sub> Δ110 (red) and VanS <sub>A</sub> Δ110-P (blue) observed within spectra obtained from TWIM-MS experiments for three different datasets.....	131
<b>Figure 4.16:</b> Pictorial representation of VanS <sub>A</sub> autokinase activity.....	132
<b>Figure 5.1:</b> HPLC trace from TOSOH HLC-723 HbG7 analyser for blood sample from normal adult patient. Peaks shown represent the elution of heterodimers (αβ dimers).....	144
<b>Figure 5.2:</b> Box plot illustrating the distribution of 30 normal β-chain mass measurements. ....	158
<b>Figure 5.3:</b> The shift in mass measurement for the β-chain obtained upon MS analysis of samples containing 0 %, 20 %, 40 %, 60 %, 80 % and 100 % of a -1Da variant.....	159
<b>Figure 5.4:</b> HPLC trace from TOSOH HLC-723 HbG7 analyser for blood sample from patient A. Retention time window for Hb S is shown in blue.....	161
<b>Figure 5.5:</b> Deconvoluted spectrum obtained after ESI-MS analysis of blood sample from patient A with the α-chain used as an internal calibrant. ....	162

<b>Figure 5.6:</b> Deconvoluted spectrum obtained after ESI-MS analysis of blood sample from patient A with the $\beta$ -chain used as an internal calibrant. ....	163
<b>Figure 5.7:</b> ESI-spectra of tryptic digest of blood sample from patient A (red) and control (green). Inset spectra show the 999-1002 $m/z$ region in greater detail. A peak is observed at 999.17 $m/z$ in the sample which is not observed in the control.....	164
<b>Figure 5.8:</b> Schematic of $\alpha$ T9, amino acids 60-92 in the $\alpha$ -chain (shown in single-letter code). The potential single amino acid mutations which would result in a -1 Da mass change are illustrated.....	164
<b>Figure 5.9:</b> a.) MS/MS spectra of 999.5 $m/z$ from variant sample (red) and control (green). ....	165
<b>Figure 5.10:</b> HPLC trace from TOSOH HLC-723 HbG7 analyser for blood sample from patient B. Retention time windows for Hb D, Hb S and Hb C are shown in green, blue and red respectively.....	166
<b>Figure 5.11:</b> Deconvoluted spectrum obtained after ESI-MS analysis of blood sample from patient B. The $\alpha$ -chain was used as an internal calibrant. The main peaks within the spectrum are identified as $\alpha$ -chain, $\alpha$ -chain variant (+14 Da), $\beta$ -chain variant (-30 Da) and $\beta$ -chain. ....	167
<b>Figure 5.12:</b> ESI-MS spectra of tryptic digest of blood sample from patient B (red) and a control (green). An additional ion is observed within the spectrum for the patient's sample at 922.5 $m/z$ . This corresponds to $(\beta$ T1 -30 Da) <sup>1+</sup> .....	168
<b>Figure 5.13:</b> Schematic of $\beta$ T1, amino acids 1-8 in the $\beta$ -chain (shown in single-letter code). The potential single amino acid mutations which would result in a -30 Da mass change are illustrated.....	169
<b>Figure 5.14:</b> MS/MS spectra for 922.5 $m/z$ from variant sample (red) and 952.5 $m/z$ from control (green) with expected fragment ions for 952.5 $m/z$ shown. ....	169
<b>Figure 5.15:</b> ESI-MS spectra of tryptic digest of blood sample from patient B (red) and a control (green). An additional singly-charged ion is observed at 717.4 $m/z$ and an additional triply-charged ion at 771.4 $m/z$ within the spectrum for the patient's sample. ....	170
<b>Figure 5.16:</b> Schematic of $\alpha$ T9, amino acids 60-92 in the $\alpha$ -chain. The tryptic peptides produced for the mutant $\alpha$ 68(N→K) are shown.....	171
<b>Figure 5.17:</b> Deconvoluted spectrum obtained after ESI-MS analysis of blood sample from patient C with the $\alpha$ -chain used as an internal calibrant. An additional	

peak at 15156.27 Da is observed within the spectrum. This represents the presence of a +30 Da  $\alpha$ -chain variant. .... 172

**Figure 5.18:** ESI-MS spectra of tryptic digest of blood sample from patient C (red) and a control (green). The spectrum for the patient's sample indicates that a doubly-charged ion may be present at 932.4  $m/z$  which is masked by  $\beta T2^{1+}$  at 932.5  $m/z$ . . 173

**Figure 5.19:** Schematic of  $\alpha T6$ , amino acids 41-56 in the  $\alpha$ -chain. The potential single amino acid mutations which would result in a +30 Da mass change are illustrated. .... 173

**Figure 5.20:** MS/MS spectrum of 933  $m/z$  ( $\pm 2 m/z$  units) from blood sample of patient C. Inset spectrum shows the precursor ions selected. .... 174

**Figure 5.21:** Illustration of the separation of overlapping doubly- and singly-charged ions by travelling wave ion mobility mass spectrometry. A.) The 933  $m/z$  region was selected for ion mobility analysis. B.) The arrival time distribution showed the presence of two separate peaks corresponding to the arrival times for the C.) doubly-charged and D.) singly-charged species. .... 175

**Figure 5.22:** MS/MS spectrum for mobility-separated 932.4  $m/z$  (doubly-charged) with the corresponding peptide sequence and fragment ions. Interpretation of the spectrum confirms the mutation site as  $\alpha 51(G \rightarrow S)$ . .... 176

**Figure 5.23:** Mass error in  $\alpha$ -chain measurement obtained for 174 blood samples (two replicates). .... 177

**Figure 5.24:** Mass error in  $\beta$ -chain measurements obtained for 174 blood samples (average of two replicates). .... 179

**Figure 5.25:** Estimated glycosylated hemoglobin levels within each of 174 blood samples obtained by MS analysis (average of two replicates). .... 180

**Figure 5.26:** Estimated percentage of  $\delta$ -chain present within each of the 174 blood samples relative to the  $\beta$ -chain (average of two replicates). .... 181

# List of Tables

<b>Table 2.1:</b> Theoretical cross-sections for standard proteins calculated from published NMR and X-ray structures using the MOBCAL program and the PA, EHSS and TM models. ....	61
<b>Table 2.2:</b> Estimated cross-sections (smallest and largest) for proteins studied, and differences between these cross-sections observed.....	79
<b>Table 5.1:</b> The common clinically significant structural variants, the mutations which cause them and their clinical manifestations (data obtained from HbVar)...	140
<b>Table 5.2:</b> Summary of information acquired by MS intact analysis and MS and MS/MS digest analysis of 10 control samples. The disease states of each hemoglobin sample were correctly identified. ....	160

# Acknowledgements

First and foremost, I would like to thank my supervisor, Prof. Jim Scrivens, for providing me with the opportunity to undertake this PhD, for his support and guidance and for his continuous encouragement. Thank you to the BBSRC and Waters Corporation who provided the financial support for this work.

I would like to thank all of the members of the Scrivens Group, past and present, for providing a wonderful working environment. Many of the group are not only colleagues but dear friends. Special thanks go to Gill and Kostas who have provided me with great support during some difficult times. I will also be forever grateful to Prof. Keith Jennings for his guidance and advice and help with the finer grammatical details within this thesis. I now, at least, understand the concept of a 'hanging participle' and apologise if any have crept in to this work.

I would like to thank Prof. Michael T. Bowers for providing cross-section measurements for equine myoglobin for calibration purposes.

For the VanS work, I would like to thank my collaborators, Dr. Andrew Quigley and Dr. David Roper, for their expertise.

I would like to thank Dr Nicholas Jackson and Yvonne Elliot, University Hospitals Coventry and Warwickshire NHS Trust, for providing blood samples. For the investigation of hemoglobin abnormalities in South Asians, thanks go to my collaborators, Dr. Jeetesh Patel and Julie Chackathayil.

Finally my biggest thank you goes to my family who have always supported me and encouraged me throughout my career. An extra special thank you goes to my husband, Scarffy, who has provided relentless support, has always believed in my abilities and has spent hours listening to me talk about my work even though he hasn't always understood it!

# Declaration

I hereby declare that this thesis, submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy and entitled “ Investigating Protein Structure by Means of Mass Spectrometry”, represents my own work and has not been previously submitted to this or any other institution for a degree, diploma or other qualification. Work undertaken by my collaborators is explicitly stated where appropriate.

Charlotte A. Scarff

July 2010

# Summary

The three-dimensional conformation of a protein is central to its biological function. Mass spectrometry (MS) has become an important tool for the study of various aspects of protein structure. This project investigates the use of MS for diagnosis of hemoglobinopathies, through primary structure identification, and for three-dimensional protein structure analysis, through comparison to established methods and application to protein systems.

Travelling-wave ion mobility mass spectrometry (TWIM-MS) was used to investigate the biological significance of gas-phase protein structure. Protein standards were analysed by TWIM-MS. Cross-sections were estimated for proteins studied, for charge states most indicative of native structure, and were found to be in good agreement with those calculated from published X-ray crystallography and nuclear magnetic resonance structures. These results illustrated that the TWIM-MS approach can provide biologically-relevant data on three-dimensional protein structure.

TWIM-MS was then used to study the structural properties of the hemoglobin tetramer and its components. Results showed that globin monomers exist in similar conformations whether in apo- or holo- forms and that a heme-deficient dimer is unlikely to be a prerequisite for hemoglobin tetramer assembly. TWIM-MS was used to successfully differentiate between normal and sickle hemoglobin tetramers.

The conformational changes occurring in VanS, a histidine kinase, upon autophosphorylation were investigated by TWIM-MS. Results provided insights into the mechanism of autophosphorylation. MS was used to follow the rate of the autophosphorylation and results obtained compared well with those from an established method. This demonstrated that MS offers a simple, reproducible alternative to conventional methods for the study of phosphorylation rates.

MS was used to provide positive identification of a range of hemoglobinopathies caused by single point mutations. A high-throughput method was used to screen for hemoglobinopathies in South Asians with and without cardiovascular disease. Results showed a positive correlation between patients with hemoglobinopathies and those with cardiovascular disease.

# Abbreviations

## A

$\alpha$	Alpha globin
$\alpha^a$	Apo-alpha globin
$\alpha^G$	Hb G-Philadelphia
$\alpha^h$	Holo-alpha globin
$\alpha T9$	9th tryptic peptide of the alpha globin chain
ACN	acetonitrile
ADC	Analog-to-digital converter
ADP	Adenosine-5'-diphosphate
AGE	Agarose gel electrophoresis
ATD	Arrival time distribution
ATP	Adenosine-5'-triphosphate

## B

$\beta$	Beta globin
$\beta^a$	Apo-beta globin
$\beta^h$	Holo-beta globin
$\beta^S$	Hb S
$\beta T1$	1st tryptic peptide of the beta globin chain

## C

CA	Catalytic ATP-binding
CAE	Cellulose acetate electrophoresis
CHCA	$\alpha$ -cyano-4-hydroxycinnamic acid
CI	Chemical ionisation
CID	Collision-induced dissociation
CIMT	Carotid-artery intima media thickness
CRM	Charge residual model
CVD	Cardiovascular disease

## D

$\delta$	Delta globin
Da	Dalton
DC	Direct current
DCIM	Drift-cell ion mobility
DDA	Data directed analysis
DHp	Dimerisation phosphotransfer
DTT	Dithiothreitol

## **E**

ECD	Electron capture dissociation
EHSS	Exact hard sphere scattering
EI	Electron impact
EM	Electron multiplier
ESI	Electrospray ionisation
ETD	Electron transfer dissociation

## **F**

FAB	Fast atom bombardment
FT-ICR	Fourier transform ion cyclotron resonance
FWHM	Full-width half-maximum

## **G**

$\gamma$	Gamma globin
GlcNAc	N-acetylglucosamine

## **H**

Hb	Hemoglobin
Hb A	Normal adult hemoglobin (major component)
Hb A <sub>1c</sub>	Glycated hemoglobin
Hb A <sub>2</sub>	Normal adult hemoglobin (minor component)
Hb S	Sickle cell hemoglobin
HDL	High-density lipoprotein
HKs	Histidine kinases
HPLC	High performance liquid chromatography

## **I**

IEF	Isoelectric focussing
IEM	Ion evaporation model
IM	Ion mobility
IM-MS	Ion mobility mass spectrometry
IMS	Ion mobility spectrometry

## **M**

MALDI	Matrix-assisted laser desorption ionisation
MCA	Multi-channel acquisition
MCP	Microchannel plate

MD	Molecular dynamics
MgCl <sub>2</sub>	Magnesium chloride
MRM	Multiple reaction monitoring
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MurNAc	N-acetylmuramic acid
<i>m/z</i>	Mass-to-charge ratio
<b>N</b>	
NHS	National health Service
NMR	Nuclear magnetic resonance
<b>O</b>	
oa-TOF	Orthogonal acceleration time-of-flight
<b>P</b>	
PA	Projection approximation
PDB	Protein Databank
PTMs	Post translational modifications
<b>Q</b>	
Q	Quadrupole
<b>R</b>	
RF	Radio frequency
RR	Response regulator
rt	Retention time
<b>S</b>	
SIM	Selected ion monitoring
SRIG	Stacked ring ion guide
<b>T</b>	
TDC	Time-to-digital converter
TM	Trajectory method
TOF	Time-of-flight
T-Wave	Travelling-wave
TWIG	Travelling-wave ion guide
TWIM	Travelling-wave ion mobility

## U

UHCW University Hospitals Coventry and Warwickshire

## V

VanR <sub>A</sub>	Response regulator VanR phenotype A
VanS <sub>A</sub>	Sensor protein VanS phenotype A
VanS <sub>A</sub> -ATP	ATP-associated VanS <sub>A</sub>
VanS <sub>A</sub> Δ110	Sensor protein VanS phenotype A lacking 110 amino acids from the N-terminus
VanS <sub>A</sub> Δ110-P	Phosphorylated VanS <sub>A</sub> Δ110
VanS <sub>A</sub> Δ140	Sensor protein VanS phenotype A lacking 140 amino acids from the N-terminus
VanS <sub>A</sub> Δ150	Sensor protein VanS phenotype A lacking 150 amino acids from the N-terminus
VanS <sub>A</sub> Δ155	Sensor protein VanS phenotype A lacking 155 amino acids from the N-terminus
VanS <sub>A</sub> Δ155-P	Phosphorylated VanS <sub>A</sub> Δ155

Other abbreviations are explained, where appropriate, in the text.

Standard 3 and 1 letter amino acid abbreviations are used throughout this thesis.

<b>Amino Acid</b>	<b>3 Letter Code</b>	<b>1 letter code</b>
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic Acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic Acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

# Chapter One: Introduction

---

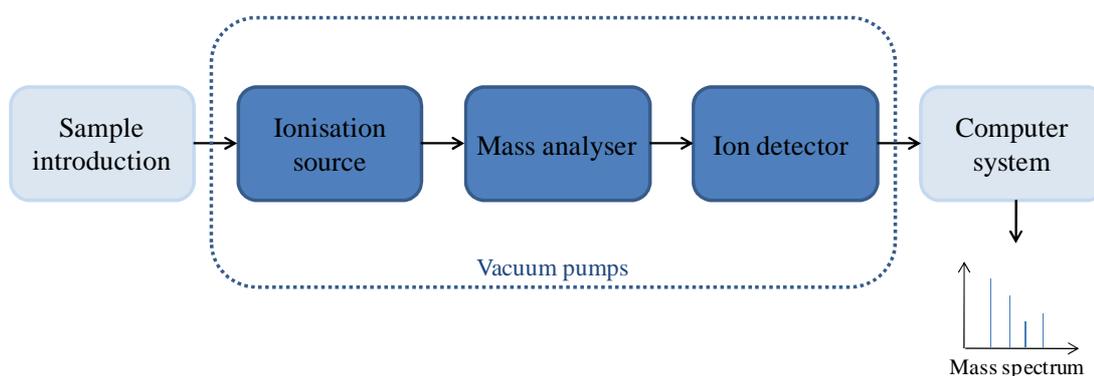
## 1.1 Mass spectrometry

Mass is a highly specific characteristic of any molecule and the ability to measure it precisely is fundamental to many scientific investigations. The technique of mass spectrometry was developed in the quest to understand the nature of cathode rays. Some researchers believed they were electromagnetic in nature whilst others believed they consisted of particles. For those who believed in the particle theory, the solution was to measure the mass of the particles (Griffiths 2008). This was achieved in 1899 by Joseph John Thomson, who successfully measured these particles, which he named electrons (Thomson 1899). This pioneering work, for which Thomson was awarded the 1906 Nobel Prize in Physics, led to the development of the first mass spectrometer. This instrument, termed the parabola spectrograph, was used to separate the isotopes of neon (Thomson 1911). Thomson's work was continued by his student Francis W. Aston, who designed and built a new mass spectrometer with improved resolving power. This allowed him to study the isotopes of many other elements (Aston 1919) and he was subsequently awarded the Nobel Prize for his work in 1922.

Since the pioneering work of Thomson the field of mass spectrometry has been transformed. Mass spectrometry started as a physicists' tool for studying the nature of the atom but as instrument design improved the potential applications of MS began to be envisaged. By the 1940s, mass spectrometers were commercially available and industrial chemists used them to quantify levels of known substances. In the 1950s, the relationship between the mass spectra produced and the molecular structure of the compound being analysed was investigated and the use of MS to identify unknown compounds was established (Griffiths 2008). Pivotal developments in ionisation methods have followed and mass spectrometry has been coupled with other separation techniques, such as chromatography and ion mobility. These developments, along with improvements in instrument design and the commercial availability of instruments, have greatly expanded the application range of mass spectrometry such that it is now an important tool for scientists from multiple disciplines.

### 1.1.1 What is a mass spectrometer?

A mass spectrometer is an instrument that can measure the mass-to-charge ratio ( $m/z$ ) of an ion in the gas phase. To analyse a sample of interest ions are first generated from the sample. These ions are then separated according to their  $m/z$ , detected by a detector and their relative abundances recorded. Figure 1.1 shows a schematic of the major components of a modern mass spectrometer.



**Figure 1.1:** Schematic representation of a modern mass spectrometer

An analyte of interest is introduced into the mass spectrometer and ionised within the ion source. The ions produced are then separated within a mass analyser, according to their  $m/z$ . A detector registers the ions detected at each  $m/z$  and a computer system records this information and converts it into a comprehensible format. Variations in each of the major components of the mass spectrometer have resulted in a broad array of instrument types for multiple applications.

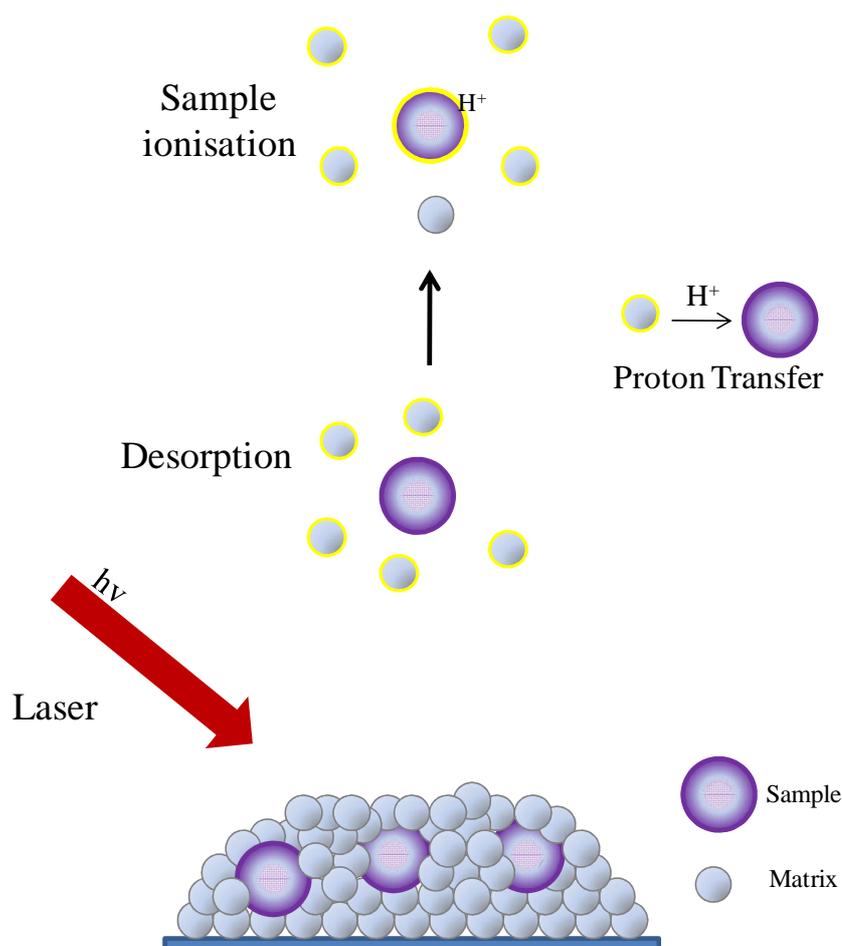
### 1.1.2 Ionisation methods

A number of ionisation methods can be used to generate ions for subsequent mass analysis by mass spectrometry. Up until the 1980s, the predominant ionisation techniques applied were electron impact (EI) and chemical ionisation (CI). EI and CI can only be used to ionise molecules within an already vaporised sample. This limits their application to the analysis of volatile thermally-stable compounds and so they have been used extensively to characterise small organic molecules. Biological

molecules are usually polar and thermally-labile and so cannot be analysed by EI- or CI-MS without prior derivatisation (Griffiths *et al.* 2001).

The routine MS analysis of polar thermally-labile molecules, of a few thousand Dalton (Da) in mass, was made possible by the introduction of fast atom bombardment (FAB) (Barber *et al.* 1981). Other ionisation techniques, such as field desorption and thermospray ionisation, were also being applied to the analysis of small biomolecules, at a similar time, but with limited success (Griffiths 2008). The development of both matrix-assisted laser desorption/ionisation (MALDI) and electrospray ionisation (ESI), at the end of the 1980s, transformed the field of biological mass spectrometry. MALDI and ESI are still the dominant ionisation methods used for biomolecule analysis to date and their application areas are still growing.

MALDI is a pulsed soft ionisation technique whilst ESI is a continuous ionisation method. MALDI involves the co-crystallisation of a sample, in low concentration, with a matrix. A sample is first dissolved in a suitable solvent and mixed with an appropriate matrix. The mixture is spotted onto a MALDI plate (commonly made of stainless steel) and air-dried. A laser is used to ablate the analyte:matrix mixture. The matrix absorbs this laser energy and this leads to desorption and ionisation of the analyte. A nitrogen laser at a wavelength of 337 nm is most commonly used (de Hoffmann and Stroobant 2002). Matrices are chosen depending on the properties of the analyte but must be able to absorb radiation at the wavelength of the laser.  $\alpha$ -cyano-4-hydroxycinnamic acid (CHCA) is often the matrix of choice for peptide analysis whilst sinapinic acid (3,5-dimethoxy-4-hydroxycinnamic acid) is used for the analysis of intact proteins. The exact mechanism of the MALDI ionisation process is not fully understood but there are several hypotheses. The generally accepted theory is that protons are transferred from the matrix molecules to the analyte molecules, either during the desorption process or just subsequently to it (Zenobi and Knochenmuss 1998) (see Figure 1.2). The majority of ions formed during MALDI are singly-charged species resulting from the attachment of a single proton.

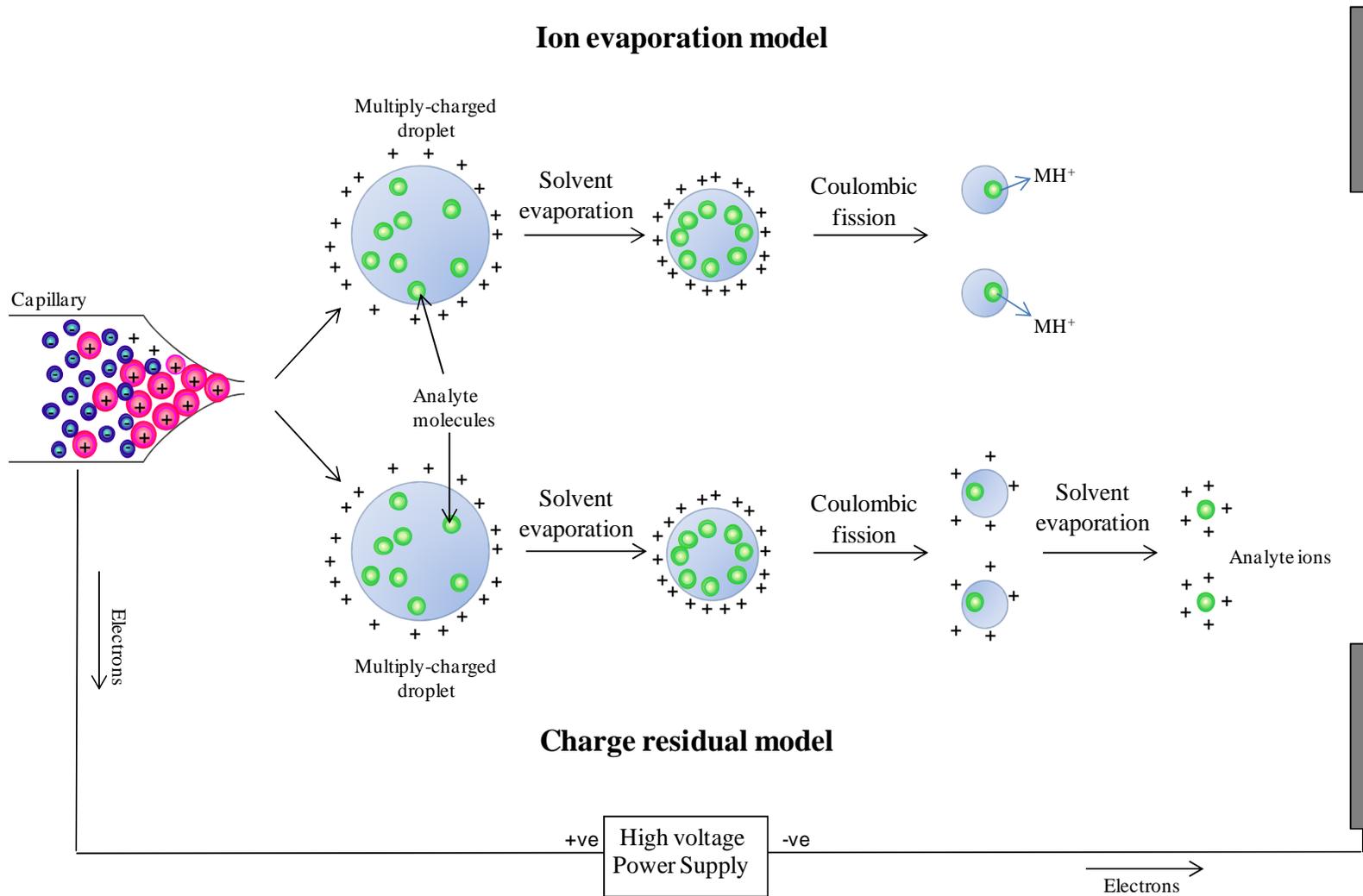


**Figure 1.2:** Schematic representation of the proposed MALDI ionisation process.

ESI takes place at atmospheric pressure and is a very gentle ionisation technique. For ESI, the sample to be analysed must first be dissolved in an aqueous/organic solution. The most appropriate solvent system to use depends on the properties of the analyte and whether ionisation in positive or negative mode is required. For ionisation in positive mode, a sample is routinely prepared in an aqueous solution that contains appropriate concentrations of an organic solvent (such as acetonitrile (ACN) or methanol) and an acid, such as formic or acetic. The sample solution is sprayed through a capillary needle, which is held at a high electrical potential with respect to the entrance of the mass spectrometer. Charge accumulation at the meniscus of the solution, at the end of the capillary, is induced by the electrical potential. A Taylor cone is formed at the end of the capillary that minimises the

charge-to-surface ratio. At a critical limit, the tip of the cone extends into a filament and highly-charged droplets are released from it (Kearle 2000). The production of droplets is aided by a nebulising gas (usually nitrogen) that flows around the outside of the capillary, towards the inlet of the mass spectrometer. The formation of gas-phase ions from these highly-charged droplets is thought to proceed through two mechanisms, the ion evaporation model (IEM) and the charge residual model (CRM). Solvent evaporates from the droplets, leading to a decrease in droplet radius but no change in charge. At a particular radius, for a given charge, the charge repulsion (or Coulombic repulsion) within a droplet exceeds the surface tension, which results in Coulombic fission. The point at which this occurs is known as the Rayleigh Limit. The IEM proposes that these solvent evaporation and Coulombic fission cycles continue until droplets are of a certain radius at which direct emission of ions into the gas phase occurs. The CRM model suggests that these evaporation/fission cycles continue until no more solvent remains and only the charged analyte remains (see Figure 1.3). There is a general consensus that ions with a small number of charges, at low  $m/z$  values, are preferentially formed by the IEM whilst multi-charged ions, at high  $m/z$  values, are formed by the CRM (Kearle 2000).

MALDI was developed by Michael Karas and Franz Hillenkamp in 1988 (Karas and Hillenkamp 1988). The original work on soft laser desorption as an ionisation method, however, was carried out by Koichi Tanaka in 1985. ESI was invented by John Fenn (Fenn *et al.* 1989). In 2002, Tanaka and Fenn were awarded the Nobel Prize in Chemistry for their development of soft desorption ionisation methods for mass spectrometric analyses of biological macromolecules.

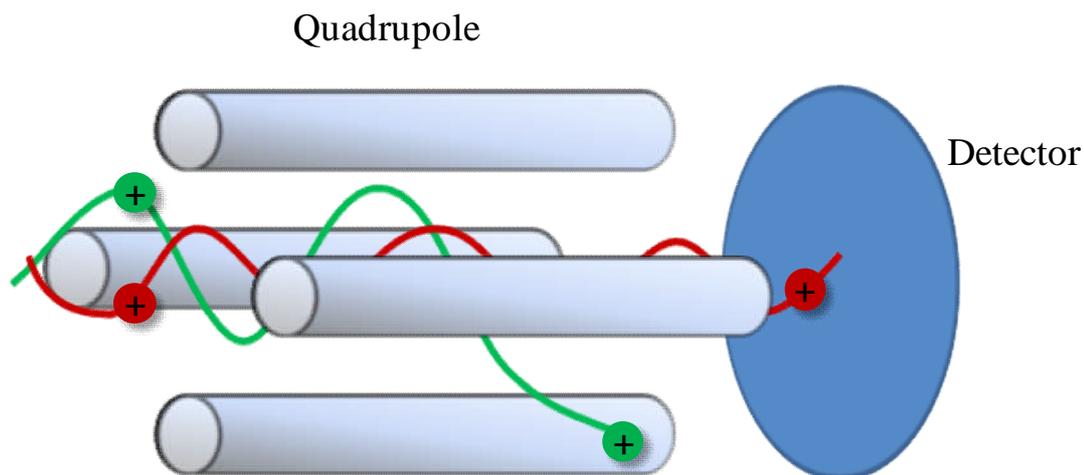


**Figure 1.3:** Schematic representation of the two proposed ESI ionisation mechanisms. Adapted from Kerbale (2000).

### 1.1.3 Mass analysers

The separation of ions, according to their  $m/z$ , is performed by a mass analyser. Scanning mass analysers, such as quadrupoles and magnetic sectors, detect ions of each  $m/z$  sequentially whilst others, such as the time-of-flight mass analyser and ion traps, allow the near-simultaneous detection of all ions (de Hoffmann and Stroobant 2002). Mass analysers, dependent on their type and design, can transmit ions with a given efficiency and have an upper mass detection limit and resolving power. The transmission efficiency of an analyser is the ratio of the number of ions it detects to the number of ions produced in the ion source. The upper mass detection limit of an analyser is the highest  $m/z$  that it can measure. The resolving power of a mass analyser is a measure of its ability to distinguish between ions of similar  $m/z$ . It is usually defined as  $(m/\Delta m)$  where  $\Delta m$  represents the smallest mass difference for which two peaks with masses  $m$  and  $\Delta m$  are resolved. For an isolated peak, resolution can be estimated by  $m/\Delta m$  where  $\Delta m$  is the full width of the peak at half-maximum (FWHM).

Quadrupole mass analysers are comprised of four round or hyperbolic rods, arranged in parallel as two pairs of electrically connected rods (see Figure 1.4). A combination of radio frequency (RF) and direct current (DC) voltages are applied to each pair of rods. As ions traverse the quadrupole mass analyser, within a particular electric field, they undergo a complex oscillating motion. The RF and DC voltages applied allow ions of particular  $m/z$  to have a stable trajectory within the quadrupole and to be transmitted through to the detector. The remaining ions collide with the sides of the rods, or are lost at the walls of the analyser. A broad packet of ions can be transmitted through the quadrupole if it is operated in RF-only mode or ions of a specific  $m/z$  can be selected for transmission by application of a specific electric field. Multipoles, with different numbers of parallel rods, to which only RF voltages are applied, are often used as ion beam focussing devices.



**Figure 1.4:** Schematic representation of a quadrupole mass analyser. An ion with a stable trajectory through the quadrupole to the detector is shown in red whilst an ion with an unstable trajectory is shown in green.

The time-of-flight mass analyser was originally described by Wiley and McLaren in 1955 (Wiley and McLaren 1955). A time-of-flight mass analyser measures the time it takes an ion to traverse a field-free region. This time can then be related to the  $m/z$  of the ion. If ions of different mass ( $m$ ) possess a fixed kinetic energy ( $E_k$ ) then they will have different velocities ( $v$ ) (Equation 1.1). This means that they will traverse a field-free region of fixed distance in different times (Equation 1.2). This time can be related to their mass and charge (Equation 1.3). Within the time-of-flight mass analyser, ions are first accelerated by a potential ( $V$ ) which gives each ion a fixed amount of kinetic energy. The ions then enter the field-free region, known as the flight tube, where they travel at a velocity which is inversely proportional to the square root of their mass (de Hoffmann and Stroobant 2002). For a fixed  $E_k$ , an ion with a mass ( $m$ ) and total charge ( $ze$ ), where  $z$  is the number of charges and  $e$  is the electronic charge, will have a velocity ( $v$ ) determined by:

$$v = \sqrt{\frac{2E_k}{m}} = \frac{\sqrt{2zeV}}{m}$$

Equation 1.1

The time ( $t$ ) an ion takes to traverse a flight tube of fixed length ( $d$ ) is inversely proportional to  $v$ :

$$t = \frac{d}{v}$$

Equation 1.2

Given that  $v$  is constant and  $V$  and  $d$  are known:

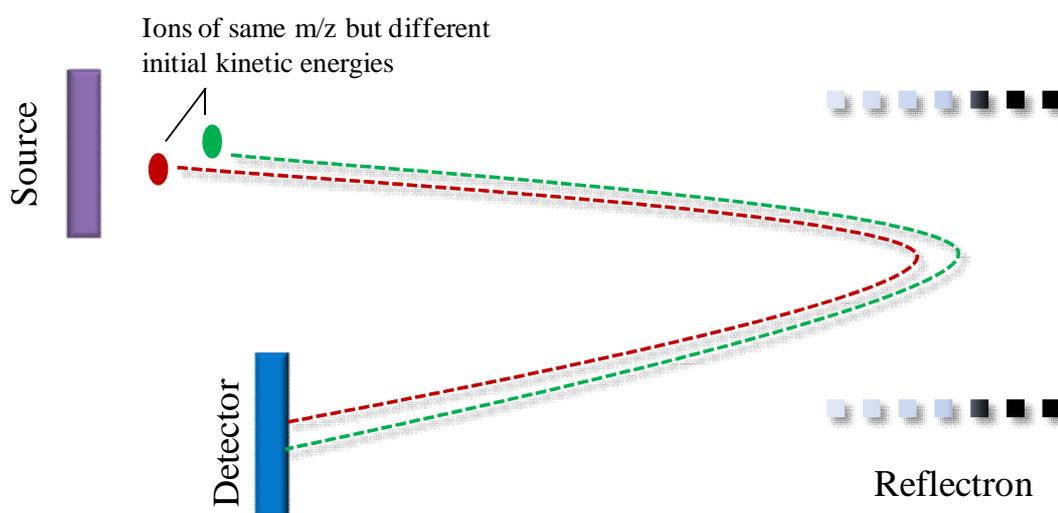
$$t = \sqrt{\frac{m}{z}} \times \text{constant}$$

Equation 1.3

A linear TOF mass analyser is highly sensitive and does not have, in theory, an upper mass detection limit. The resolving power of a linear TOF mass analyser is affected by multiple factors. Variations in the initial kinetic energy of an ion, the length of the ion formation pulse and the length of the flight tube all change the resolving power of a TOF mass analyser (de Hoffmann and Stroobant 2002). If ions of the same  $m/z$  have different initial kinetic energies they will traverse the flight tube with slightly different velocities. They will reach the detector at slightly different times, a broad ion peak will be recorded and resolution will be reduced. Conversely, as mass resolution is proportional to ion flight time, the longer the flight tube the higher the resolving power of the analyser.

Peak broadening at the detector can be reduced by compensating for differences in the initial kinetic energies of ions by use of delayed pulsed extraction (Vestal *et al.* 1995). When this technique is used ions are initially allowed to move into the flight-tube, with their velocities determined by their initial kinetic energy. An extraction pulse is then applied, which provides ions that have moved a smaller distance into the flight tube, whose kinetic energy was initially too low, with more kinetic energy. This reduces the kinetic energy spread among ions with the same  $m/z$ .

Higher mass resolution has also been achieved in modern TOF instruments by the addition of a reflectron. The use of a reflectron effectively doubles the flight path of an ion and additionally compensates for differences in initial kinetic energies of ions. A reflectron consists of a series of ring electrodes and creates a retarding field that acts as an ion mirror (Mamyrin *et al.* 1973), see Figure 1.5. Ions are accelerated into the flight tube and reflected back, when they reach the reflectron, down the flight tube to a detector. The distance an ion travels before it is reflected is related to its initial kinetic energy. Instruments with a reflectron can often be operated in V- or W-mode where ions are reflected once or twice by the reflectron respectively. A compromise is made when a reflectron is used as, although resolution is improved, there is an accompanying loss in sensitivity and a reduction in the upper mass limit of detection.



**Figure 1.5:** Schematic illustration of a reflectron TOF mass analyser.

The path of two ions, of the same  $m/z$ , through the analyser is represented. The ion with more initial kinetic energy (green) travels further into the reflectron before it is deflected than the ion with less initial kinetic energy (red). Both ions reach the detector simultaneously.

### 1.1.4 Detectors

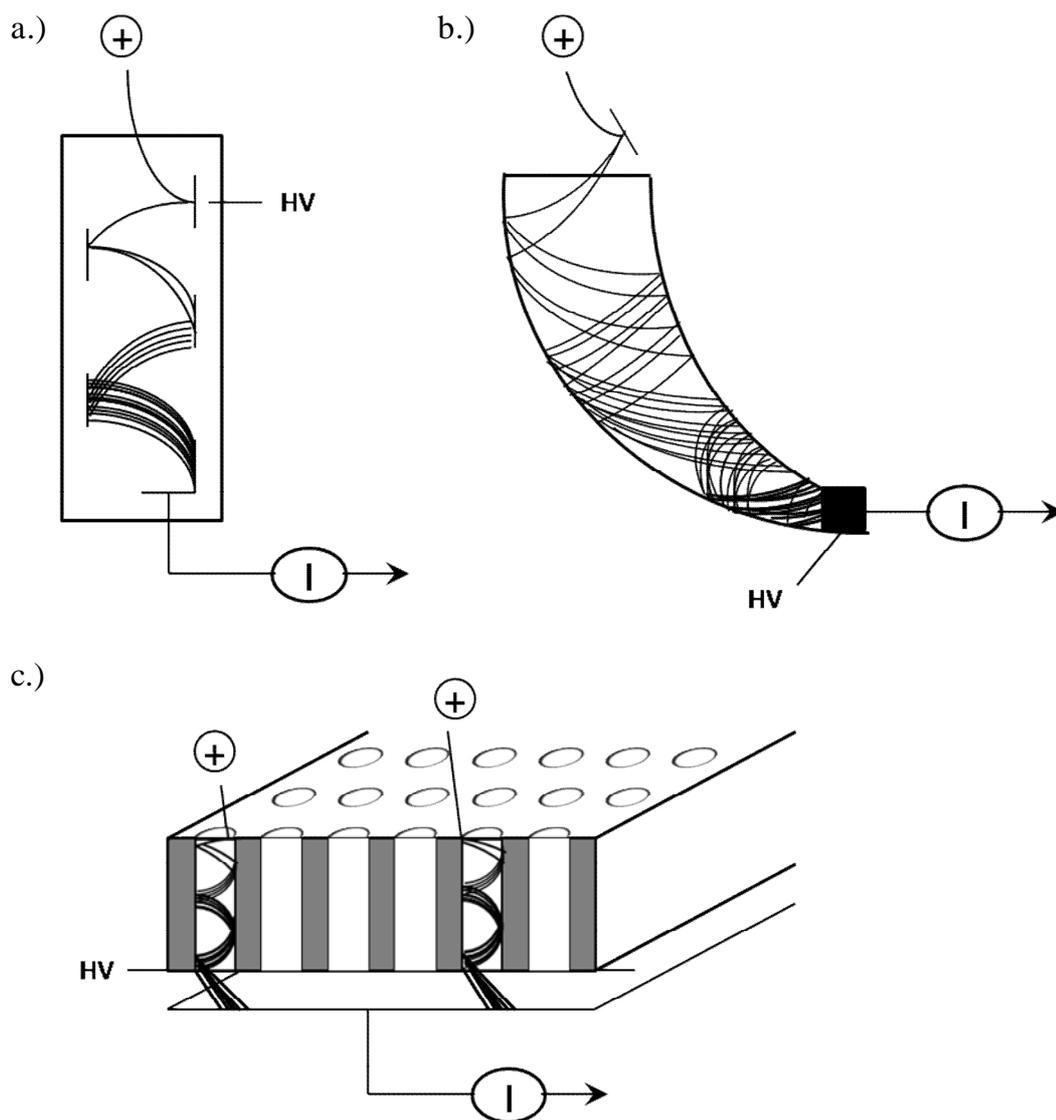
Once ions have been separated within the mass analyser they are detected by a detector. Numerous types of detector have been developed which either directly measure ions arriving at a detector, such as photographic plates, or detect an amplification signal produced as a consequence of an ion signal, such as electron multipliers.

Photographic plates were the first detectors to be used. An ion of a particular  $m/z$  hits a plate in a specific place and a calibration scale can be used to derive  $m/z$  value. The intensity of any spot observed on the photographic plate gives an indication of the relative intensity of ions present within the beam at the corresponding  $m/z$  value (de Hoffmann and Stroobant 2002). The disadvantages of this type of detector include poor sensitivity, short dynamic range and the requirement for off-line image development and calibration (Koppenaar *et al.* 2005).

Electron multiplier (EM) detectors are the most commonly used and measure ion currents produced following amplification of an ion signal. Ions exiting the mass analyser strike a conversion dynode and emit secondary electrons. These electrons are accelerated and collide with subsequent dynodes (in discrete-dynode EMs) or repeatedly with a continuous dynode (in continuous-dynode EMs) creating an electron cascade, which results in a measureable current (see Figure 1.6). The signal from a single ion can thus be amplified by  $10^6$  or more (Koppenaar *et al.* 2005).

A widely used detector in modern mass spectrometry, which is a variant of the continuous-dynode EM, is the microchannel plate (MCP) detector (Wiza 1979) (see Figure 1.6). Within a MCP, an array of microchannels, each a few tens of micrometers in diameter and a few millimetres in length, forms the amplifying volume (Koppenaar *et al.* 2005). An ion striking the MCP will emit secondary electrons and create electron cascades in nearby microchannels. Only a few channels will be affected by a single ion and so other channels can be used to detect the presence of different ions simultaneously. The electron current exiting a microchannel can be directed onto an anode for electronic detection of the current or onto a fluorescent surface for ion-beam imaging (Koppenaar *et al.* 2005). MCP

detectors use very short electron paths and very short electron pulse widths which means they can record the precise arrival time of an ion very accurately. This makes them the ideal detectors to couple with TOF mass analysers. The performance of an MCP detector can be affected by dead time effects, when the MCP becomes saturated with ions; the microchannels need time to recover before they can detect new signals effectively.



**Figure 1.6:** Schematic representations of a.) a discrete-dynode EM b.) a continuous-dynode EM and c.) a MCP detector. Adapted from Koppenaal *et al.* (2005).

Modern EM detectors use either an analog-to-digital converter (ADC) or time-to-digital converter (TDC) to process ion currents into  $m/z$  values. ADCs register the

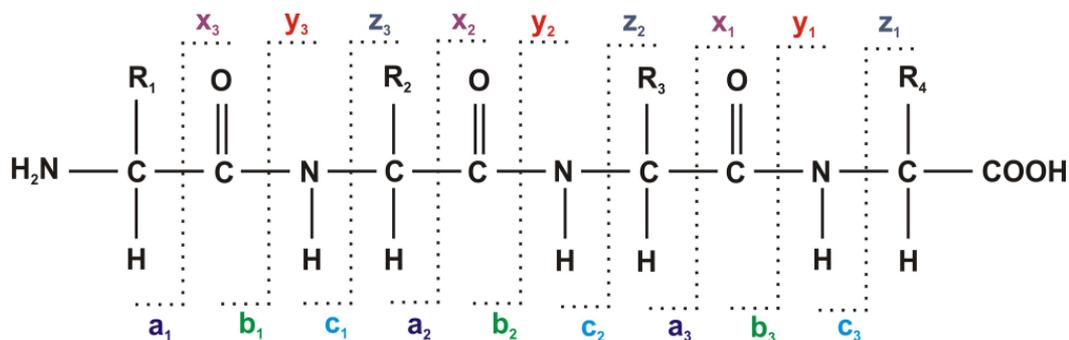
ion current created, amplify its signal and filter it to remove high frequency noise. The voltage derived from the ion current is then plotted on a  $m/z$  scale with respect to previously acquired calibrated data. TDCs register the time at which an ion strikes the detector and process this into an  $m/z$  value, again with respect to previously acquired calibrated data. Spectra are then produced by summing together individual ion recordings. TDC detectors have the advantage of greater speed of detection and sensitivity, whereas ADC detectors have an improved dynamic range. A dual ADC/TDC detector can be used to achieve fast acquisitions of data with both high sensitivity and dynamic range (Koppenaal *et al.* 2005).

### 1.1.5 Tandem mass spectrometry

Tandem mass spectrometry (MS/MS) refers to the use of at least two mass analysis steps within an experiment. MS/MS is often used in conjunction with a dissociation process to determine the product ions produced by fragmentation of a precursor ion. In this experiment, the first mass analyser is used to isolate the precursor ion, which then undergoes some fragmentation process. The product ions produced are then analysed by the second mass analyser. Interpretation of the resulting product ion spectrum can provide structural information about the precursor ion.

The MS/MS analysis of a peptide, for example, can provide peptide sequence information. The nomenclature used to describe peptide product ions was proposed by Roepstorff, Fohlmann and extended by Biemann, and is represented in Figure 1.7 (Roepstorff and Fohlman 1984; Biemann 1992).

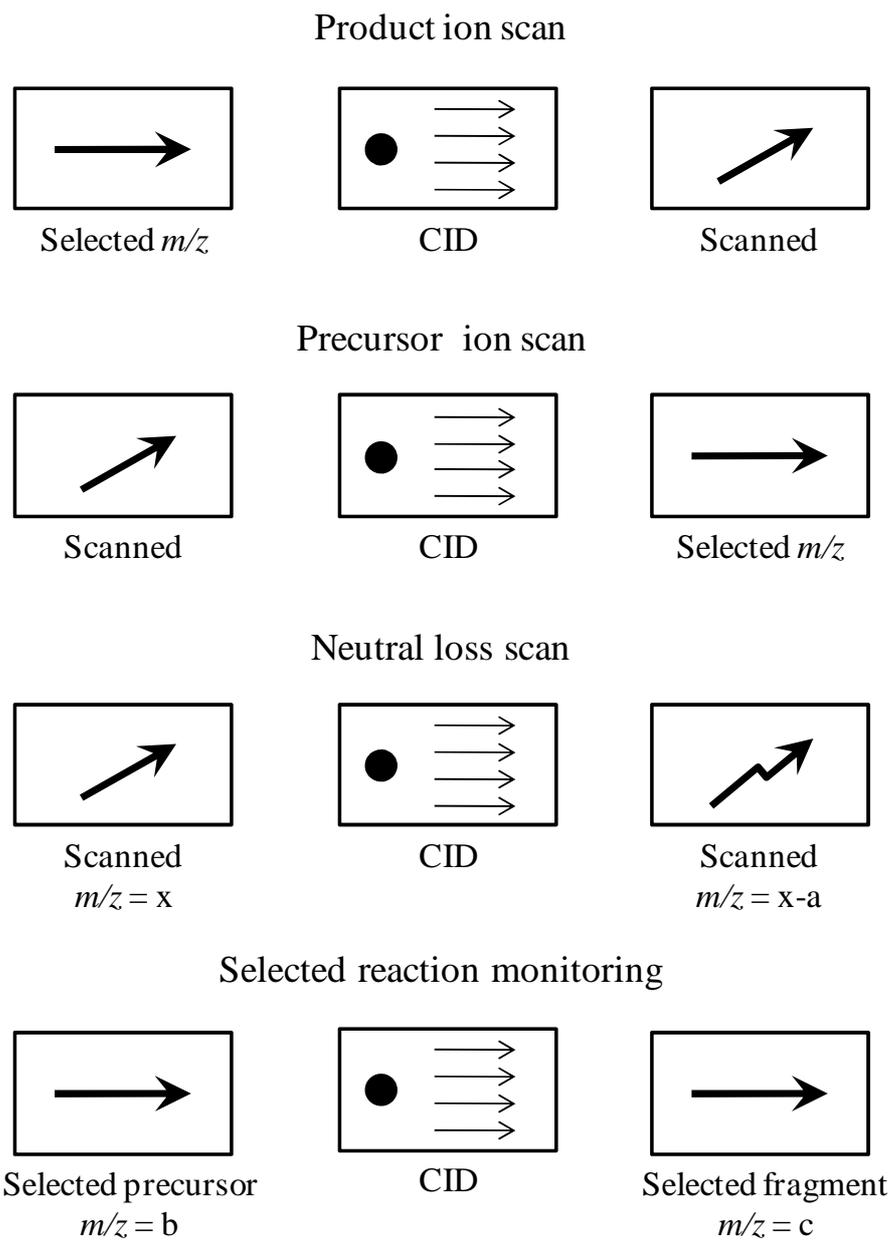
One of the most commonly used fragmentation methods employed in MS/MS experiments is collision-induced dissociation (CID), pioneered by Jennings and McLafferty in the 1960s (McLafferty and Bryce 1967; Jennings 1968). In CID, an ion is subjected to collisions with inert gas molecules. These collisions transfer energy to the ion and the subsequent increase in the internal energy of the ion induces fragmentation. In peptides, this fragmentation occurs mostly at amide bonds to give *b* and *y* fragment ions.



**Figure 1.7:** Peptide ion nomenclature, as proposed by Roepstorff, Fohlmann and Biemann. N-terminal fragment ions are termed a, b or c ions whilst C-terminal fragments are termed x, y or z ions. A subscript denotes the number of residues within a fragment ion.

Complementary fragmentation methods include electron capture dissociation (ECD) and electron transfer dissociation (ETD). ECD is based on ion-electron reactions in which the capture of electrons by a gaseous positive ion leads to fragmentation and neutralisation of that positive ion (Zubarev *et al.* 1998). In peptides, ECD leads to extensive cleavage of the peptide backbone between nitrogen atoms and  $\alpha$ -carbon atoms to yield *c* and *z* ions. The production of fragment ions by ECD is an available option on commercial Fourier transform ion cyclotron resonance (FT-ICR) instruments. ETD is a similar fragmentation approach to ECD, which is applicable to other instrumentation.

MS/MS experiments can either be performed in-space, by use of two or more spatially separated mass analysers, or in-time. In-space MS/MS experiments are commonly performed by use of a Q-TOF (a quadrupole mass analyser coupled to a TOF instrument), TOF-TOF or triple quadrupole (QQQ) instrument. Four main scan modes are used; product ion scan, precursor ion scan, neutral loss scan and selected reaction monitoring (see Figure 1.8). The choice of the most appropriate experiment depends on the information required. QQQ instruments are the most flexible and are able to operate in all four scan modes (but with limitations on sensitivity). Q-TOF and TOF-TOF instruments have largely been used for product ion scanning experiments.



**Figure 1.8:** The main scan modes in tandem mass spectrometry. CID stands for collision-induced dissociation. Adapted from de Hoffmann and Stroobant (2002).

### 1.1.6 Ion mobility mass spectrometry

Ion mobility spectrometry (IMS) is a well-established analytical technique that is used throughout the industrial world to characterise drugs, explosives and chemical warfare agents (Hill *et al.* 1990). IMS, invented in the 1970s and also known as plasma chromatography or ion chromatography, is a technique that separates ions based on their mobility (Cohen and Karasek 1970; Karasek 1974). The mobility of an ion is a measure of how rapidly it moves through a buffer gas under the influence of a weak electric field and is related to its rotationally averaged collision cross-section, mass and charge (Karasek 1974).

A constant electric field ( $E$ ), which may be produced by a series of ring electrodes, is used to propel ions through a drift cell whilst collisions with the buffer gas decelerate the motion of these ions. This process gives each ion a quasi-constant velocity ( $v_d$ ,  $\text{cm}^2 \text{s}^{-1}$ ). The velocity of an ion, under these conditions, is directly proportional to the strength of the electric field ( $E$ ,  $\text{V cm}^{-1}$ ) applied (Creaser *et al.* 2004). The mobility of an ion ( $K$ ,  $\text{cm}^2 \text{s}^{-1} \text{V}^{-1}$ ) is given by:

$$K = \frac{v_d}{E}$$

Equation 1.4

If the time taken to traverse a drift cell of length  $d$  (cm) is  $t_d$  (s), then:

$$K = \frac{d}{t_d E}$$

Equation 1.5

The mobility of an ion is routinely expressed as the reduced mobility ( $K_0$ ) under standard temperature  $T_0$  and pressure  $P_0$  ( $T_0$  273 Kelvin and  $P_0$  760 Torr):

$$K_0 = K \frac{PT_0}{TP_0} = K \frac{P273}{T760}$$

Equation 1.6

The reduced mobility  $K_0$  of an ion can be related to its collision cross-section by means of the following equation, deduced from kinetic theory (Mason and McDaniel 1988).

$$K_0 = \frac{3ze}{16N_0} \times \frac{1}{\sigma} \sqrt{\left(\frac{2\pi}{\mu k_B T}\right)}$$

Equation 1.7

where  $z$  is the number of charges,  $e$  is the electronic charge,  $N_0$  the buffer gas number density at standard temperature and pressure,  $\sigma$  is the average collision cross-section,  $\mu$  is the reduced mass of the buffer gas and ion,  $k_B$  is the Boltzmann constant and  $T$  is the effective temperature.

The mobility of an ion, at a given drift gas pressure and temperature, is therefore determined by the reduced mass, charge and collision cross-section of that ion (Creaser *et al.* 2004). The average collision cross-section of an ion can therefore be obtained from the average of all possible collision geometries of that ion (Clemmer and Jarrold 1997). Models may be used to calculate collision cross-section, estimate ion mobility and establish structure/mobility correlations (Creaser *et al.* 2004). Several methods have been proposed to calculate the average collision cross-section of an ion from a three-dimensional model of its structure. The simplest model is the projection approximation (PA), first described by Mack in 1925 (Mack 1925) and since developed by Bowers and co-workers (von Helden *et al.* 1991; Jarrold and Bower 1993; von Helden *et al.* 1993). This approach treats each atom within an ion as a hard sphere. Each geometric cross-section of an ion is measured by its projection (shadow) and all the projections created by every possible orientation of the ion are averaged to provide a rotationally-averaged collision cross-section. The PA is an adequate approximation for small molecules but can underestimate the cross-section of proteins when interactions with the buffer gas become important (Jarrold 1999). Extensions of this model have thus been described which take into account scattering effects of collisions between ions and buffer gas molecules and long-range interactions between ions and the buffer gas, which affect ion mobilities (Mesleh *et al.* 1996; Shvartsburg and Jarrold 1996).

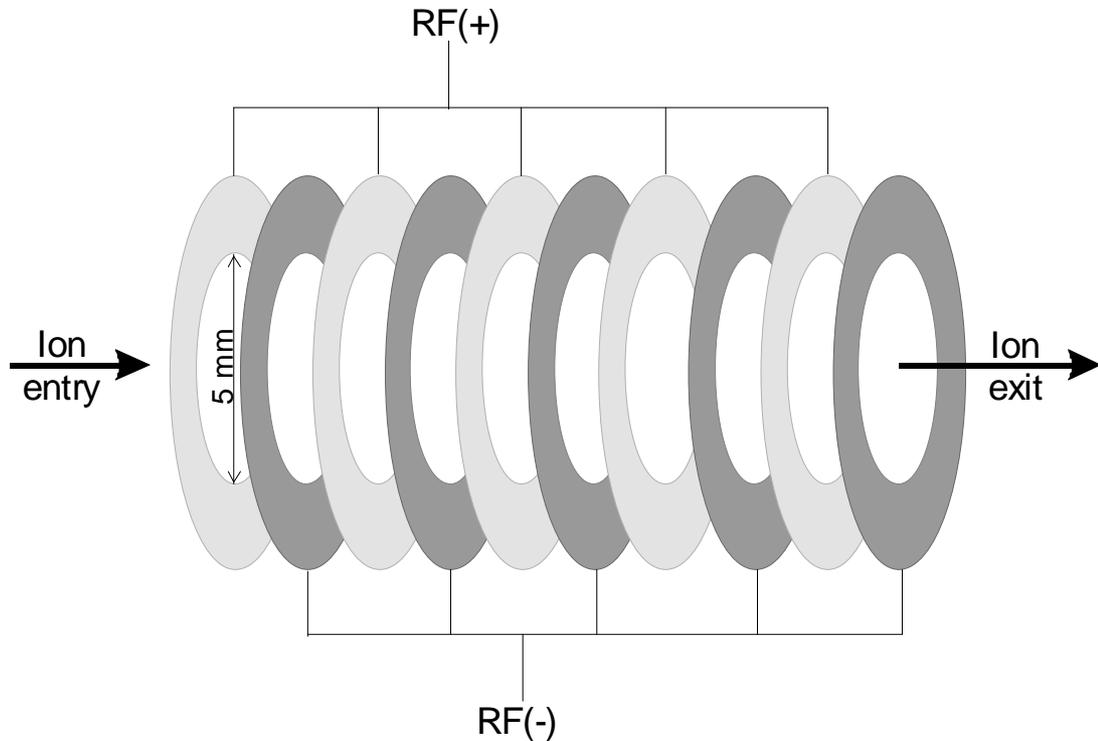
The technique of ion mobility spectrometry has been coupled with MS to produce several different types of ion mobility mass spectrometry (IM-MS) instrumentation. ESI and MALDI ionisation sources have been used to generate ions prior to IM analysis and quadrupole mass analysers (Shelimov *et al.* 1997), TOF mass analysers (Srebalus *et al.* 1999) and FT-ICR spectrometers (Bluhm *et al.* 2000) have all been used after IM separation to perform mass analysis. Recently an instrument which uses two and three drift cells in tandem has also been described (Koeniger *et al.* 2006; Merenbloom *et al.* 2006). Two main forms of ion mobility have been developed for use with MS, drift-cell ion mobility (DCIM) and travelling-wave (T-Wave) ion mobility (TWIM).

DCIM-MS instruments contain a drift cell (IMS cell) to which a constant electric field is applied, as described above. The majority of DCIM-MS instruments have been designed, developed and built in-house by specialist labs that possess the necessary engineering, electrical, mechanical and software expertise required. A DCIM device has been incorporated into a commercial instrument (McCullough *et al.* 2008) and a drift cell has been added into a commercial Q-ToF hybrid instrument (Thalassinos *et al.* 2004). The coupling of IM separation with a TOF mass analyser offers the advantage that mass spectra can be acquired on the microsecond time scale whilst mobility separation occurs on the millisecond time scale. A large number of mass spectra can thus be obtained by a TOF mass analyser during a single mobility experiment. DCIM-MS instruments have been used to conduct experiments to investigate the structural properties of an extensive range of molecules such as carbon clusters (von Helden *et al.* 1993), polymers (Gidden *et al.* 2002), peptides (Kaleta and Jarrold 2003), proteins (Wytttenbach *et al.* 1996; Shelimov *et al.* 1997) and nucleic acids (Gidden and Bowers 2003).

### **1.1.7 Travelling-wave ion mobility mass spectrometry**

In contrast to DCIM, in which a constant low electric field is applied to the mobility cell, TWIM uses a travelling-wave, comprised of a series of transient DC voltages, to propel ions through an inert gas in a stacked ring ion guide (SRIG) (Giles *et al.* 2004). The SRIG is composed of a series of ring electrodes that are arranged orthogonally to the ion transmission axis. Opposite phases of RF voltage are applied

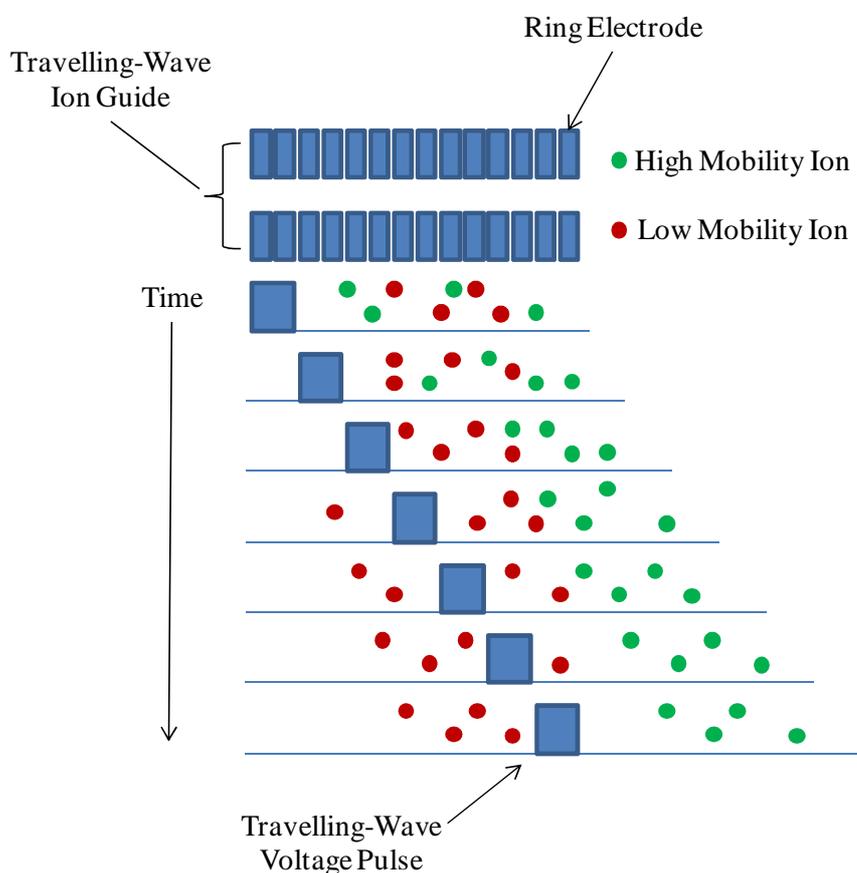
to adjacent rings; this creates a radially-confining effective potential barrier (Giles *et al.* 2004) (Figure 1.9). This stabilises the trajectory of ions as they traverse the SRIG and prevents ion loss via radial diffusion.



**Figure 1.9:** A schematic representation of a stacked ring ion guide. Adapted from Giles *et al.* (2004).

Ions are propelled through the SRIG by superimposing a DC potential on the confining RF of each pair of adjacent electrodes in succession. The movement of this potential along the ion guide provides a travelling voltage wave on which the ions can surf. Ions move along the SRIG in waves and it is thus referred to as a travelling-wave ion guide (TWIG). The TWIG can be used as a mobility separation device as the ability of an ion to surf along a travelling voltage wave through an inert gas depends on its mobility. Mobility separation of ions is thus achieved within the TWIG, under a given gas pressure, travelling-wave height and velocity conditions, because ions with high mobility are carried with the wave and exit the device faster

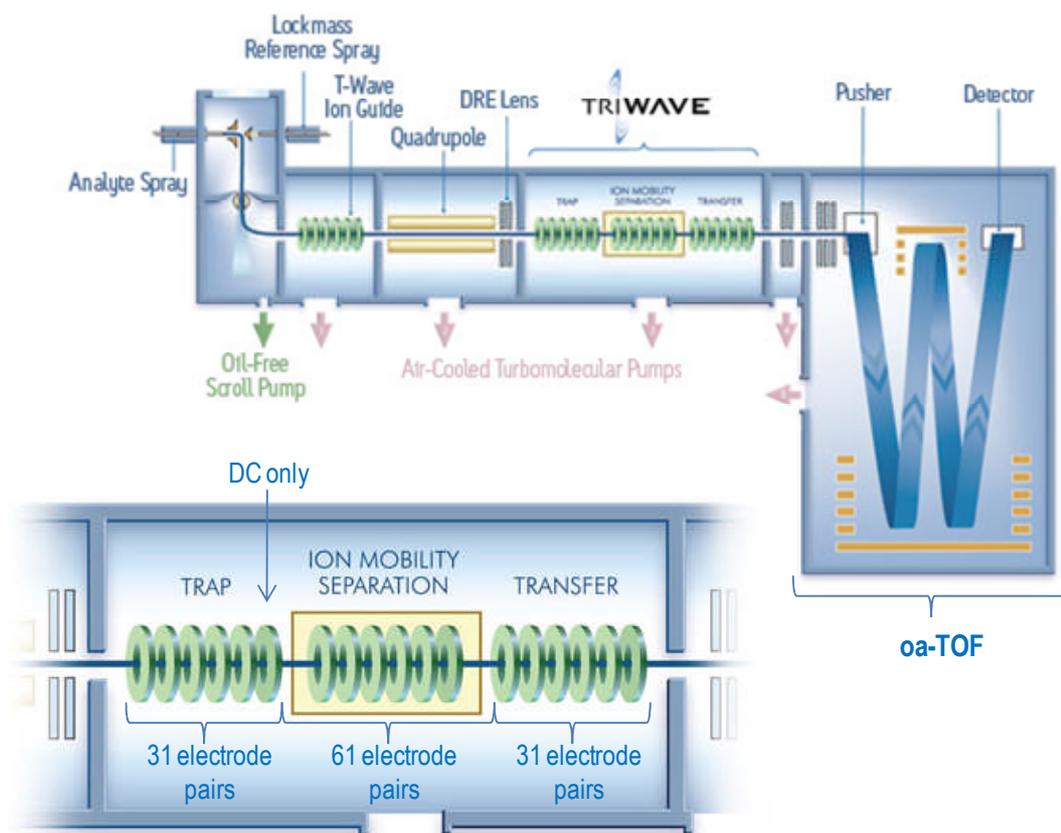
than ions of lower mobility that roll over the wave top and, as a consequence, exit the device later (Giles *et al.* 2004) (see Figure 1.10).



**Figure 1.10:** Schematic illustration of mobility separation in a TWIG. Here high mobility ions surf upon the wave whilst low mobility ions roll back over the top of the wave. Adapted from Giles *et al.* (2004).

TWIGs have been incorporated into a Q-ToF instrument to create the commercially-available Synapt HDMS system (Waters, Manchester, UK) (Pringle *et al.* 2007). A schematic of this instrument is shown in Figure 1.11. The instrument is a hybrid quadrupole ion mobility orthogonal acceleration time-of-flight (oa-ToF) mass spectrometer. Samples are ionised by means of ESI, and more recently MALDI, and enter the instrument via a z-spray source. Three TWIGs form the TWIM separator, named the TriWave or T-Wave (see Figure 1.11), which is situated between the

quadrupole and oa-TOF with the first TWIG referred to as the trap, the second as the mobility cell, and the third as the transfer.



**Figure 1.11:** Schematic representation of the Synapt instrument with an enlargement of the TriWave ion mobility device inset. Adapted from Pringle et al. (2007).

The trap and transfer regions are 100 mm long and consist of 31 electrode pairs. A travelling-wave voltage pulse is not applied to the trap. The final electrode of the trap is DC-only and its voltage is modulated so that it periodically gates ion packets into the mobility cell for separation to occur. The mobility cell is 185 mm long and consists of 61 electrode pairs. The flow of buffer gas within the mobility cell can be optimised for each sample and a maximum pressure of 1 mbar can be sustained. The travelling-wave velocity is given in m/s which is derived from the distance between pairs of electrodes divided by the time the pulse remains in each pair. The travelling-wave voltage is applied to the 1<sup>st</sup> and 7<sup>th</sup> pair of electrodes then the 2<sup>nd</sup> and 8<sup>th</sup> pair of

electrodes and so on. A 10 ms pulse time will result in an average velocity of 300 m/s since the distance between each pair of electrodes is ~ 3 mm. The transfer region has a travelling-wave voltage pulse applied to it continually in order to maintain the mobility separation of the ions achieved within the mobility cell as they are transferred to the oa-TOF. The initial push of the oa-TOF pusher is synchronised with the gated release of ion packets from the trap into the mobility cell. The gate is typically opened for 100  $\mu$ s. For each gate pulse, 200 orthogonal acceleration pushes of the TOF analyser are recorded to form one ion mobility experiment. The overall mobility recording time is  $200 \times t_p$ , where  $t_p$  is the pusher period, which is determined by the mass acquisition range (Pringle *et al.* 2007). The synchronisation of gated release of ions into the ion mobility separator with TOF acquisition allows both arrival time distribution (ATD) information and mass spectral data to be obtained simultaneously. The process is repeated until a mass spectrum with the desired signal-to-noise is obtained. This mode of mobility acquisition is one of the major advantages of the Synapt instrument.

Within the Synapt, CID experiments can be performed in the trap or transfer regions which means that a variety of different experiments can be performed. To add to this flexibility, the quadrupole present before the T-Wave can be operated in RF-only mode, to allow passage of all ions, or in resolving mode to selectively isolate ions of particular  $m/z$ . Another advantage of the Synapt instrument is that the TWIM cell does not compromise the intrinsic sensitivity of the hybrid Q-TOF instrument, unlike the majority of DCIM devices. Experiments conducted within the Synapt instrument show excellent reproducibility and the instrument is easy to use. The Synapt is available with a MALDI or ESI interface and with quadrupoles of different upper mass detection limits, it can therefore be tailored to a range of applications and subsequently its commercial availability has significantly increased the number of ion mobility users. The Synapt does currently have a number of limitations relating to resolving power, mobility cell design and the method of mobility separation. The resolving power of the T-Wave is not as high as that obtained within specialized in-house developed DCIM-MS instruments. The mobility cell has an upper pressure limit of 1 mbar and its temperature cannot be controlled. The physical principles behind DCIM are well characterised and data obtained can be used to measure the absolute collision cross-section of an ion. The mobility of an ion through the TWIM

cell cannot be directly related to its absolute collision cross-section. The electric field applied within the cell is non-uniform and time-dependent and the motion of an ion through the TWIM cell is complex and not fully understood. Several calibration approaches have been proposed, however, to obtain estimates for the collision cross-sections of ions from their mobilities, as measured by the TWIM device (Ruotolo *et al.* 2005; Scrivens *et al.* 2006; Wildgoose *et al.* 2006; Ruotolo *et al.* 2008; Thalassinos *et al.* 2009).

### 1.1.8 Travelling-wave calibration

It has been shown that reliable estimates of collision cross-sections for unknown compounds can be obtained from their mobilities through the T-Wave by reference to the mobilities of standards with known cross-sections (Ruotolo *et al.* 2005; Wildgoose *et al.* 2006; Thalassinos *et al.* 2009). Standards that have had their absolute collision cross-sections determined by means of DCIM-MS experiments are generally used (Clemmer ; Shelimov *et al.* 1997; Valentine *et al.* 1997a; Valentine *et al.* 1997b). A universal calibration method has not been developed and slightly different approaches are used by different research groups. The calibration method which is used in this work has been developed by the Scrivens research group (Scrivens *et al.* 2006; Williams and Scrivens 2008; Thalassinos *et al.* 2009) and is adapted from the method described by Wildgoose *et al.* (2006). To perform the calibration, standards are analysed by TWIM-MS, under the experimental conditions at which the analyte is analysed. The following steps are undertaken to create a calibration:

1. The orthogonal acceleration push of the TOF analyser (scan number ( $n$ )) in which each calibrant ion arrives at the detector is recorded
2. This is converted into an arrival time ( $t_d$ ) by:

$$t_d = n \times \text{pusher period (ms)}$$

Equation 1.8

3. The arrival time is corrected for  $m/z$  independent flight time (T-Wave offset)

The T-Wave offset is the time it would take an ion to traverse the mobility region and the transfer region at a particular wave velocity in the absence of a buffer gas. At a wave velocity of 300 m/s, it would take an ion 10  $\mu\text{s}$  to traverse one pair of ring electrodes. The mobility cell contains 61 electrode pairs and the transfer region contains 31 electrode pairs.  $[(61 \times 10) + (31 \times 10)] \mu\text{s} = 920 \mu\text{s}$ . The  $m/z$  independent corrected drift time ( $t'_d$ ), corrected for the T-Wave offset at a wave velocity  $WV$ , can thus be calculated from the following equation:

$$t'_d = t_d - \left( \frac{920 \times 300}{WV} \right)$$

Equation 1.9

4. The arrival time is further corrected for  $m/z$  dependent flight time

The  $m/z$  dependent flight time is proportional to the square root of the  $m/z$  value. It must be subtracted to obtain the corrected effective drift time ( $t''_d$ ), i.e. the time taken to traverse the mobility cell. At 1000  $m/z$ , the time taken for an ion to traverse the TOF mass analyser is 44  $\mu\text{s}$  and the transit time from being expelled from the transfer region to the TOF mass analyser is 41  $\mu\text{s}$ . The addition of these two times is equal to 85  $\mu\text{s}$ . The corrected effective drift time ( $t''_d$ ) is thus given by:

$$t''_d = t'_d - \sqrt{\frac{m/z}{1000}} \times 0.085$$

Equation 1.10

5. Calibration coefficients are obtained from published absolute cross-section data ( $\sigma$ ). Published cross-sections are corrected to take into account the effects of reduced mass and charge state. Where  $e$  is the charge on the ion,  $m_i$  is the mass of the ion and  $m_n$  is the mass of the mobility gas, normalised cross-sections are given by:

$$\sigma' = \frac{\sigma}{e \times \sqrt{\left( \frac{1}{m_i} + \frac{1}{m_n} \right)}}$$

Equation 1.11

6.  $\sigma'$  is plotted against  $t''_d$ .
7. A power series fit ( $y = Ax^B$ ) or a linear series fit ( $y = Ax + B$ ) to the data points is applied. A power series fit has been shown to provide a more reliable calibration for large compounds, such as proteins, whereas a linear relationship has been found to be more appropriate for smaller molecules, such as peptides (Thalassinos *et al.* 2009).
8. Experimental T-Wave mobility measurements obtained for an analyte are converted into estimated cross-sections by correction for reduced mass and charge and application of the power series fit (Equation 1.12) or the linear series fit (Equation 1.13) as appropriate.

$$\sigma = A \times t''_d{}^B \times e \times \sqrt{\left(\frac{1}{m_i} \times \frac{1}{m_n}\right)}$$

Equation 1.12

$$\sigma = [(A \times t''_d) + B] \times e \times \sqrt{\left(\frac{1}{m_i} \times \frac{1}{m_n}\right)}$$

Equation 1.13

The T-Wave height, wave velocity and pressure of the buffer gas within the mobility cell can all have an effect on the arrival time of an ion but these parameters do not affect the calibration provided that mobility measurements for the calibration standard are obtained under the same experimental conditions as the analyte (Leary *et al.* 2009). The corrected arrival times of the calibration standards must also bracket those of the analyte for the calibration to be valid (Shvartsburg and Smith 2008).

## **1.2 Investigating Protein Structure**

Determination of protein structure, function and interactions is vital to our understanding of biological processes. Many biophysical methods have been developed to investigate protein structure, protein dynamics and protein interactions, in the quest to relate protein structure to function.

### **1.2.1 What is protein structure?**

Protein structure may be divided into four main categories, primary, secondary, tertiary and quaternary structure. The primary structure of a protein relates to its amino acid sequence. The secondary structure of a protein relates to the local spatial arrangement of atoms within the backbone of a polypeptide chain. The tertiary structure of a protein is its three-dimensional arrangement in space. Many proteins consist of multiple polypeptide chains, loosely termed subunits, and the spatial arrangement of these forms the quaternary structure of a protein (Voet and Voet 1995). Complete functional protein assembly may also require the association of ligands, for example metal ions or heme.

The primary structure of a protein is ultimately responsible for its three-dimensional conformation and function. Knowledge of a protein's primary structure can be used to indicate its functionality, evolutionary relationships and cellular localisation (Nelson and Cox 2005). Many inherited diseases are caused by proteins with modifications in their primary sequences, which give rise to proteins with altered three-dimensional structures that do not function correctly. Some aspects of primary protein structure are not uniquely determined by gene sequence. Post-translational modifications (PTMs) are often applied to amino-acid sequence such as glycosylation to asparagine residues and phosphorylation to threonine, serine or tyrosine residues. PTMs are of significant importance and greatly influence the functionality of proteins. PTMs can determine the cellular localisation of a protein, target a protein for degradation, activate a particular function of a protein or conversely deactivate it (Mann and Jensen 2003).

The secondary structure of a protein is determined by local intra-chain interactions, typically hydrogen bonds between backbone carboxyl- and amino-groups. These interactions produce secondary structure elements such as alpha-helices and beta-sheets. The complete three-dimensional structure of a folded polypeptide chain is held together by non-covalent interactions (van der Waals forces, hydrophobic interactions, electrostatic interactions and hydrogen bonding) between secondary structural elements and disulphide bonds between cysteine residues (Voet and Voet 1995). These non-covalent interactions also preserve quaternary protein structure.

### **1.2.2 Protein structure vs. protein conformation**

Proteins within their native-states are not static entities; they are flexible. The degree of this flexibility ranges widely, from local structural fluctuations to global conformational changes, but is often vital for protein function (Kaltashov and Eyles 2002). Correct understanding of protein function therefore not only requires a good knowledge of a protein's structure but also how this structure may change its conformation.

### **1.2.3 Classic tools of structural biology**

The field of structural biology deals with the determination of molecular structure of biological macromolecules and how these structures affect their functions. The secondary structure of a protein is often analysed by spectroscopic techniques such as circular dichroism, infrared and Raman spectroscopy (Pelton and McLean 2000). The best tools a structural biologist currently has to characterise the three-dimensional structure of a protein are solid-phase X-ray crystallography, solution-phase nuclear magnetic resonance (NMR) spectroscopy and cryo-electron microscopy.

X-ray crystallography is a spectroscopy technique based upon the diffraction pattern of X-rays striking a crystal. A computer-controlled diffractometer is used to record diffraction data and high-speed computers are used to analyse the data collected in order to determine the position of atoms within a crystal (Deschamps and George

2003). The application of X-ray crystallography is therefore limited to the study of proteins which form crystals.

NMR spectroscopy exploits the fact that certain nuclei, such as  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$ , have nuclear spin. Nuclear spin generates a magnetic dipole so that when a strong magnetic field of uniform strength is applied to a protein solution, the magnetic dipoles are aligned either parallel or anti-parallel to the field (Nelson and Cox 2005). Short pulses of electromagnetic energy are then applied at right angles to the nuclei aligned in the magnetic field. The nuclei absorb some of this energy and this provides information about the identity of the nuclei and their chemical environment (Nelson and Cox 2005). NMR spectroscopy is largely used to study the three-dimensional structure of proteins that have a molecular weight less than 30,000 Da and do not form crystals (Delepierre and Lecroisey 2001). Solution-phase NMR can be used to investigate protein dynamics, providing information pertaining to transient conformational states (Delepierre and Lecroisey 2001).

Cryo-electron microscopy involves imaging a specimen at a temperature near to that of liquid nitrogen ( $-196\text{ }^\circ\text{C}$ ) with an electron microscope. It can be used to study macromolecular assemblies in different forms, symmetries, sizes and shapes but only at low-medium resolutions (Chiu *et al.* 2006).

NMR spectroscopy and X-ray crystallography provide atomic resolution of protein structure and are assumed to reflect biological form but do have limitations. Whilst X-ray crystallography provides detailed structural information, this relates to only one static structure sampled by the protein and is reliant on the production of protein crystals that will diffract. Production of protein crystals is often considered an art rather than a science and is a laborious process (Nelson and Cox 2005). NMR experiments have the advantage of being conducted in solution but typically require that the sample be isotopically labelled and in a high concentration, which can be difficult to obtain. Interpretation of NMR data also becomes more difficult when analysing higher molecular weight species due to spectral complexity (Nelson and Cox 2005).

### **1.3 Analysing protein structure by mass spectrometry**

The main application of mass spectrometry to protein analysis has been primary structure analysis within the field of proteomics. Proteomics involves the large-scale analysis of the complement of proteins expressed by a cell at a particular point in time. Proteomic studies aim to identify the proteins within a proteome and measure their expression levels, modifications, localisations, interactions and functions (Heuvel and Heck 2004). Mass spectrometry has been at the forefront of these studies, initially being used to identify proteins, then in quantitative studies and in studies of post-translational modifications (Aebersold and Mann 2003). The application of mass spectrometry to secondary structure analysis has been limited. It is the application of mass spectrometry to higher order protein structure analysis, however, which is of most interest with regard to this thesis.

Over the last ten to twenty years, with the advancements in instrumentation and ionisation methods as discussed above and extensive research efforts, mass spectrometry has become a valuable structural biology tool for the study of three-dimensional protein structure and the quaternary structure of non-covalent protein complexes.

#### **1.3.1 Probing three-dimensional protein conformation with ESI-MS**

The first use of mass spectrometry to obtain information regarding protein conformation was reported by Chowdhury *et al.* in 1990. They observed changes which occurred in the charge state distribution of cytochrome c mass spectra as a function of solution pH. They stated that the multiply-charged ions produced in the ESI process are primarily a result of proton attachment to exposed basic sites of the protein and that the availability of these sites is dependent on the conformation of the protein under the conditions of the study. A tightly folded conformation will have fewer exposed basic sites, and would accommodate fewer charges on the protein surface, than an unfolded conformation of the same protein (Chowdhury *et al.* 1990). If the charge states of gas-phase ions observed in the ESI mass spectrum of a protein reflect the charge states of that protein in solution, then the spectrum would provide

conformational information relating to the protein's solution state. At low pH, they observed that cytochrome c unfolded and accepted a large number of protons. At higher pH, cytochrome c accepted far fewer protons and was in a more tightly-folded conformation. These observations demonstrated the viability of ESI-MS as a new physical method to probe conformational changes in proteins under varying conditions, such as changes in pH, temperature and solvent composition.

A wealth of research into the use of mass spectrometry to provide information relating to three-dimensional protein structure has since followed. The following points are thus generally accepted. A typical ESI mass spectrum of a monomeric protein, obtained under controlled, non-denaturing, experimental conditions, will contain a narrow charge state distribution with a single maximum. The observation of a wide charge state distribution, with more than one maximum, within the spectrum, is indicative of the presence of more than one conformational state (Chowdhury *et al.* 1990). The lowest charge states detected in the spectrum for a protein are thought to be most representative of the native structure of that protein (Hoaglund-Hyzer *et al.* 1999). Conversely, partially-folded protein conformers have larger solvent-exposed surface areas and therefore can accept more charge upon transition from the solution phase to the gas phase (Mohimen *et al.* 2003).

### **1.3.2 Studying protein interactions, dynamics and complexes**

Soon after the initial work of Chowdhury *et al.* (1990) non-covalent interactions of proteins began to be studied by ESI-MS (Ganem *et al.* 1991; Katta and Chait 1991; Smith *et al.* 1992). The gentleness of the ESI process allowed non-covalent interactions, present in solution, to be preserved in the gas phase. This meant that ESI-MS could be used to study quaternary protein structure and protein interactions with nucleic acids, ligands and cofactors. The introduction of nanoflow-ESI (Wilm and Mann 1994) increased the applicability of ESI-MS to the analysis of protein complexes and interactions as it made the study of minute amounts of sample possible without a compromise in sensitivity (Wilm and Mann 1996). MALDI has also been used to study protein assemblies (Kiselar and Downard 2000; Strupat *et al.* 2000) but, due to the sample preparation required and the relatively higher internal energy of the ions formed, is not as applicable as ESI to the analysis of large

proteins, or protein assemblies, under near-physiological conditions (Heck and van den Heuvel 2004). As such it is not discussed further here.

Approaches based upon mass spectrometry have been used to investigate subunit stoichiometry and interactions within protein complexes, protein-ligand interactions, protein-DNA/RNA interactions, acid-denaturation and non-covalent complex assembly and disassembly. Examples of these investigations by mass spectrometry have been described in several extensive reviews (Loo 1997; Loo 2000; Kaltashov and Eyles 2002; Heck and van den Heuvel 2004; Ashcroft 2005) and will not be repeated in detail here. Some of the methods used are discussed below.

MS and MS/MS experiments have been used to determine the stoichiometry of subunits within a protein complex. Given an intact mass of a complex its composition can be determined by the sum of the masses of individual subunits (Hernandez and Robinson 2007). The arrangement of subunits within a complex can be investigated by MS/MS experiments as subunits on the outside of a complex are generally the most readily dissociated (Sharon *et al.* 2006). As the number of charges accepted by a protein complex, upon ESI, is related to the number of exposed basic sites, which can accept charge, this ionisation method can be used with MS to determine the oligomeric state of a complex. If the number of charges accepted by a tetrameric complex were on average double the number accepted by a dimeric complex then they would not be so easily distinguished (Heck and van den Heuvel 2004). As subunits are dissociated from a protein complex, by CID, they tend to retain a large portion of the charge which was on the intact complex. This asymmetric charge partitioning, with respect to mass, has been shown to be symmetric with respect to surface area (Benesch *et al.* 2006). Monomers, which are dissociated from a complex, are proposed to be largely unfolded (Jurchen and Williams 2003). The surface area of a stripped oligomer (an oligomer from which a subunit has been stripped by CID), can thus be estimated from the charge partitioning observed within mass spectra, between stripped oligomer and subunit, and the surface area of the unfolded subunit (Benesch *et al.* 2006).

The analysis of large proteins and protein complexes has largely been conducted within Q-TOF mass spectrometers. TOF mass analysers have extensive  $m/z$  ranges,

high sensitivity and fast acquisition rates. The combination of a quadrupole mass analyser with a TOF mass analyser allows MS/MS experiments to be conducted. The  $m/z$  range of the quadrupole mass analyser obviously limits the size of complex which can be investigated by mass spectrometry. Within the Robinson group, a tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies, up to 2-3 mDa in size, has been developed (Sobott *et al.* 2002b). Recently, Q-TOF instruments with 32 kDa quadrupoles have become commercially-available.

MS has also been applied to the analysis of protein dynamics and folding (Kaltashov and Eyles 2002; Eyles and Kaltashov 2004) and the structural elucidation of dynamic protein complexes (Sharon and Robinson 2007). Subunit exchange in solution has been observed by MS, by use of real-time monitoring (Sobott *et al.* 2002a; Painter *et al.* 2008). The folding and self-assembly of non-covalent protein complexes (Boys and Konermann 2007) and kinetic studies of protein-refolding/unfolding reactions (Konermann *et al.* 1997a; Konermann *et al.* 1997b) have also been measured by means of similar approaches.

### **1.3.3 The role of IM-MS**

The development of higher resolution ion mobility spectrometry techniques coupled with MS has provided a powerful tool for the determination of molecular structure (Kanu *et al.* 2008), providing information on shape as well as mass. IM-MS can be used to investigate the conformational states exhibited by a protein or protein complex within the gas phase. It has been applied to most of the applications areas discussed above. IM-MS can be used to investigate conformational changes that occur within proteins as they unfold/refold. It can be used to study global changes in protein structure occurring upon ligand-binding (Sharon and Robinson 2007). In the analysis of protein complexes, IM-MS can simultaneously provide information regarding subunit composition and overall topology (Ruotolo *et al.* 2005; Ruotolo *et al.* 2007; Ruotolo *et al.* 2008). An important outcome of the application of IM-MS to protein structure analysis is that it provides additional evidence that solution-phase structure is reflected in the gas phase.

### 1.3.4 Solution-phase vs. gas-phase structure

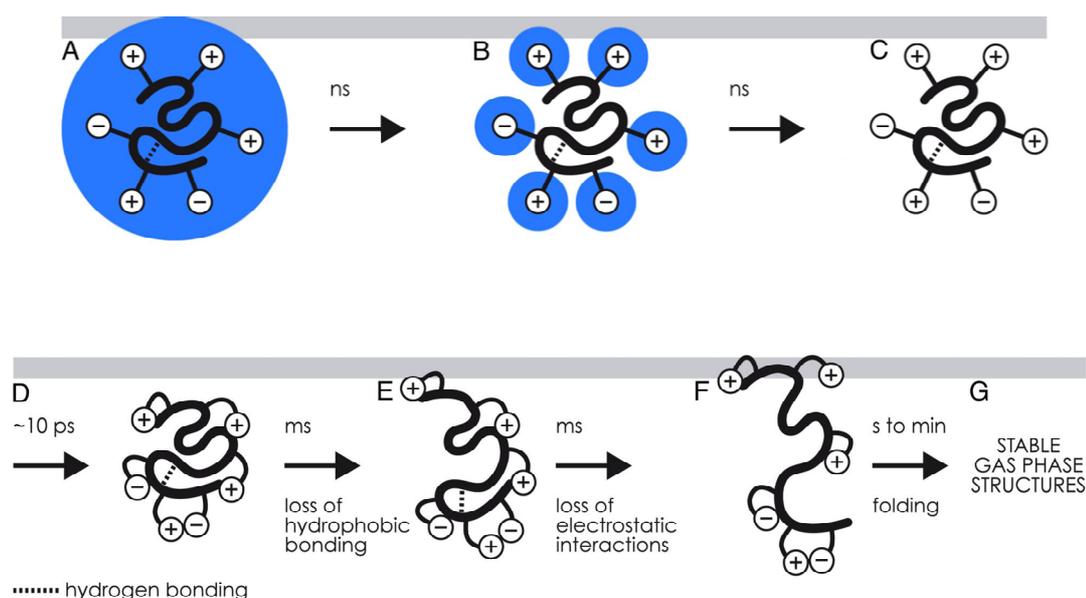
Key to the applicability of ESI-MS to native protein structure analysis is whether the solution-phase structure is conserved within the gas phase. There is now increasing evidence to suggest that protein conformations in the gas phase can mirror those in the solution phase, particularly over short time periods (Badman *et al.* 2005; Ruotolo *et al.* 2005; Ruotolo and Robinson 2006; Breuker and McLafferty 2008). Studies have shown that under controlled ESI conditions protein conformations can be preserved upon ionisation and gas-phase ions can retain conformational properties of structures present in the solution phase over the time they exist within the mass spectrometer (Hoaglund-Hyzer *et al.* 1999).

Evidence that proteins retain a memory of their solution-phase conformations in the gas phase has come from ion mobility studies where cross-sections, for the lowest charge states observed, for proteins studied have compared well with those calculated from X-ray and NMR structures (Shelimov *et al.* 1997; Shelimov and Jarrold 1997; Myung *et al.* 2002; Ruotolo *et al.* 2005; Scarff *et al.* 2008). Evidence has also been gained from the study of binding interactions and subunit stoichiometry (Loo 1997; Heck and van den Heuvel 2004) and charge partitioning upon dissociation of a complex (Jurchen and Williams 2003).

Solution-phase structures may not be preserved in the gas phase in all cases, however. Native protein structure is stabilised by a range of non-covalent interactions in solution but in the gas phase electrostatic interactions are heightened and hydrophobic interactions are weakened (Robinson *et al.* 1996; Loo 2000). It therefore may not be possible to observe/maintain an intact protein complex in the gas phase if its interactions are predominantly hydrophobic (Hernandez and Robinson 2007). Nevertheless, examples of the maintenance of hydrophobic interactions within the gas phase have been shown (Liu *et al.* 2009). In general, it is now widely accepted that under controlled experimental conditions and over short timescales gas-phase protein structure reflects that in the solution phase.

Breuker and McLafferty have proposed that globular proteins undergo a temporal evolution of structures after ESI, summarised in Figure 1.12 (Breuker and

McLafferty 2008). Molecular dynamics (MD) simulations have shown that following desolvation, on the picosecond timescale, structural changes occur on the exterior of the native protein (Steinberg *et al.* 2008). Charged amino acid side chains form a network of electrostatic interactions on the protein surface, which stabilise the native fold. MD results have shown that this structure can remain essentially unchanged for milliseconds and this has been supported by experimental evidence. Ion mobility studies conducted on cytochrome c ions, where ions were trapped prior to mobility analysis for different time periods, showed that a native-like structure persisted for 30 ms within the gas phase (Badman *et al.* 2005). After this time, hydrophobic interactions and subsequently electrostatic interactions dissociate and protein ions refold into stable gas phase structures over a timescale of seconds to minutes (Breuker and McLafferty 2008).



**Figure 1.12:** Stepwise evolution of globular protein structure after ESI.

(A) Native protein is covered with a monolayer of water, water is lost and concomitant cooling occurs to give the native protein with exterior ionic functionalities still hydrated (B), further water loss and cooling results in a dry protein (C). The dry protein undergoes collapse of its exterior ionic functionalities, millisecond loss of hydrophobic bonding (D) and millisecond loss of electrostatic interactions (E). Transiently unfolded ions form new non-covalent bonds in seconds (F) and refold into stable gaseous ion structures in minutes (G). Adapted from Breuker and McLafferty (2008).

### 1.3.5 Preserving protein structure in the gas phase

To maintain protein structures in the gas phase that reflect solution-phase structures MS experiments (termed native MS experiments) need to be conducted under controlled, non-denaturing conditions. Controlled non-denaturing conditions are obtained by careful optimisation of solvent and instrument settings. Hernandez and Robinson have produced a general protocol for this optimisation, detailing the main requirements for maintaining intact protein complexes within the gas phase (Hernandez and Robinson 2007). The key requirements are given below and are also relevant for the intact analysis of monomeric proteins.

Native MS measurements are usually performed in a buffered spray solution at the physiological pH of the analyte. The buffers commonly used for protein purification and storage, such as Tris and HEPES, are often non-volatile and are therefore incompatible with native MS (van Duijn 2010). Non-volatile substances are not amenable to the electrospray ionisation process and cause suppression of ionisation and/or extensive adduct formation (Hernandez and Robinson 2007). Protein samples are therefore buffer-exchanged into an ESI-compatible buffer, such as aqueous ammonium acetate at neutral pH. Ammonium acetate salt concentrations used typically range from 5 mM to 1M (Heck 2008) with 3 M concentrations found to be beneficial in some cases (Hernandez and Robinson 2007). High concentrations of ammonium acetate within the buffer aid in reducing the effects of non-volatile buffer components (Iavarone *et al.* 2004) but may cause dissociation of the protein complex under investigation.

The correct instrument conditions are vital to maintaining protein complexes intact in the gas phase. The major considerations are electrospray flow rate, mass range and voltage and pressure settings in the ESI source (Hernandez and Robinson 2007). Most native MS experiments are now conducted by nanoflow-ESI instead of standard ESI as the lower flow rate improves sensitivity and increases tolerance to buffer salts whilst requiring less sample volume (Hernandez and Robinson 2007). Non-covalent complexes and large proteins exhibit ions of high  $m/z$ . The transmission of high  $m/z$  ions can be improved by collisional-cooling or damping (Krutchinsky *et al.* 1998; Chernushevich and Thomson 2004). Collisional-cooling

involves cooling the ions by reducing their energy so they are more effectively transmitted into the mass spectrometer. One way to achieve this is to increase the pressure in the transfer region of the instrument between the source and the analyser (Tahallah *et al.* 2001). This increase in pressure reduces the energy of high  $m/z$  ions and helps to focus the ion beam, improving transmission (Heuvel and Heck 2004). The  $m/z$  range of TOF mass analysers, typically used for native MS experiments, is only limited by ion detection and the use of collisional-cooling can significantly improve this limit of detection.

### **1.3.6 Native MS as a structural biology tool**

Mass spectrometry-based analysis of global protein structure can be especially useful when structural characterisation by more conventional techniques, such as NMR spectroscopy or X-ray crystallography, is difficult or not possible. Recently, however, MS has been established as a structural biology tool in its own right, providing useful complementary information to other methods and possessing some distinct advantages.

MS has the unique advantage of being able to analyse heterogeneous populations and can be applied to the analysis of large protein complexes over a mDa in size, well beyond the scope of NMR. MS is a very sensitive and selective technique and thus allows the analysis of endogenous protein complexes (Heck and van den Heuvel 2004; Benesch *et al.* 2007). MS can be used to determine unambiguously the stoichiometry of protein complexes (Hernandez and Robinson 2007; Sharon and Robinson 2007). MS has also successfully been applied to the analysis of membrane proteins (Barrera *et al.* 2008), one of the most important yet difficult protein families to study. An initial analysis of a protein sample by MS under various conditions can provide X-ray crystallographers with information to expedite X-ray structural determination (Cohen and Chait 2001).

The time frame of a typical native MS experiment means that the dynamics of proteins, protein complexes and protein interactions can all be studied in real-time. X-ray crystallography structures provide a static snapshot of a protein or protein complex and thus do not provide information on dynamic processes. NMR has

widely been used to study dynamic processes but it cannot be used to provide information about individual protein complexes within a mixture. MS can be used to study the dynamics of multiple protein subpopulations in real-time experiments (Sharon and Robinson 2007). Native MS can be coupled with ion mobility to conduct experiments which provide information on shape as well as mass. Whilst IM-MS only provides a rather crude measurement of protein global structure it can be performed using small concentrations of protein and within a short time scale in comparison to X-ray crystallography or NMR spectroscopy.

The applicability of native MS is however limited by a number of factors. To perform a native MS experiment on a protein or protein complex it must be within a buffer compatible with MS and be of a sufficient purity. Some proteins are not stable within these buffers and so cannot be analysed, others require the presence of cofactors or salts, which interfere with sample ionisation. The quality of the sample is thus the main factor that affects the outcome of a native MS experiment (van Duijn 2010). Although heterogeneous populations can be studied by MS, their spectra are highly complicated and it can be extremely difficult to correctly assign peaks; protein complexes of similar mass give rise to overlapping charge states within mass spectra. Differences in the ionisation, charging and transmission efficiencies of different species means that spectra provide only qualitative and not quantitative information (Hernandez and Robinson 2007). Some protein complexes may not maintain conformations that reflect their native solution-phase conformations upon transmission into the gas phase if they contain a large proportion of hydrophobic interactions. Ultimately, MS is limited by mass range. As the mass of an analyte increases, desolvation, ion transmission and detection become less efficient (Hernandez and Robinson 2007).

## 1.4 Aims and Objectives

The work presented in this thesis involved the investigation of the use of mass spectrometry-based approaches to study aspects of protein structure. The three-dimensional structure of various protein systems is studied by means of travelling-wave ion mobility mass spectrometry and the use of mass spectrometry as a diagnostic tool for hemoglobinopathies is investigated.

The use of travelling-wave ion mobility mass spectrometry to study three-dimensional structure is a relatively new concept. The Synapt HDMS system was only made commercially-available in January 2007. Assessment of the use of this approach was conducted in initial work and then applied to two different systems; hemoglobin and VanS from the vancomycin resistance pathway in *Enterococcus faecium*.

The application of mass spectrometry to the diagnosis of hemoglobinopathies is not a new concept. Mass spectrometry is used to identify disorders in rare cases, which cannot be identified by classic approaches. It is not, however, considered an alternative approach for population screening due to cost and data interpretation skills required. Here an established mass spectrometry-based method for hemoglobinopathy diagnosis is adapted for application on a Q-TOF instrument and used to provide diagnosis of hemoglobin disorders. A potential method for high-throughput screening by mass spectrometry is introduced. The prospect of establishing a method based upon mass spectrometry as the routine approach for population screening is discussed.

The aims of this project were to:

1. Utilise travelling-wave ion mobility mass spectrometry to study protein structure and assess the biological significance of measurements made by comparison to those from X-ray crystallography and nuclear magnetic resonance studies (Chapter Two).

2. Investigate the structure of hemoglobin tetramers and their constituents by means of travelling-wave ion mobility mass spectrometry (Chapter Three).
3. Analyse conformational changes occurring in VanS, a histidine kinase, upon autophosphorylation. Investigate the properties of autophosphorylation by mass spectrometry-based approaches (Chapter Four).
4. Investigate the practical use of mass spectrometry for hemoglobinopathy diagnosis (Chapter Five).

## 1.5 Research Papers

The research presented in this thesis has already resulted in two peer-reviewed and published papers.

**Scarff, C. A., Patel, V. J., Thalassinos, K. and Scrivens, J. H.** (2009). Probing hemoglobin structure by means of travelling-wave ion mobility mass spectrometry. *Journal of the American Society for Mass Spectrometry*. **20**, 625-631.

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

## 1.6 Conference papers (Peer-reviewed)

Patel, J. V., Chackathayil, J., Tracey, I., Lovick, A., Gill, P. S., **Scarff, C. A.**, Thalassinos, K., Scrivens, J.H., Lip, G.Y.H. and Hughes, E.A. Haemoglobin disorders and low HDL cholesterol in South Asians – Are they connected? *Proc. 24<sup>th</sup> HEART UK Annual Conference, 2010, Edinburgh, UK.*

**Scarff, C. A.**, Kondrat, F. D. L., Booyjzsen, C., Mukherjee, A., Sadler, P. J. and Scrivens, J. H., Probing transferrin-metal interactions by means of travelling-wave ion mobility mass spectrometry, *Proc. 58<sup>th</sup> ASMS Conf. on Mass Spectrometry and Allied Topics, 2010, Salt Lake City, USA.*

Kondrat, F. D. L., **Scarff, C. A.**, Leszczyszyn, O. I., Kowald, G. R., Blindauer, C. A. and Scrivens, J. H., Conformational changes associated with the removal of zinc ions from proteins in a cyanobacterial Zn<sup>2+</sup> homeostatic system, *Proc. 58<sup>th</sup> ASMS Conf. on Mass Spectrometry and Allied Topics, 2010, Salt Lake City, USA.*

Scrivens, J. H., **Scarff, C. A.** and Snelling J., Thermally-assisted extractive electrospray ion mobility mass spectrometry characterisation of medicinal spray formulations, *Proc. 58<sup>th</sup> ASMS Conf. on Mass Spectrometry and Allied Topics, 2010, Salt Lake City, USA.*

Harvey, D. J., Crispin, M., Scalan, C., Bonomelli, C., Sobott, F., **Scarff, C. A.**, Thalassinou, K., Scrivens, J. H., Use of Ion Mobility and Negative Ion CID Mass Spectrometry for the Identification of N-glycans from Nanomolar Amounts of Glycoprotein, *Proc. 58<sup>th</sup> ASMS Conf. on Mass Spectrometry and Allied Topics, 2010, Salt Lake City, USA.*

Kondrat, F. D. L., **Scarff, C. A.**, Leszczyszyn, O. I., Blindauer, C. A. and Scrivens, J. H., An ion mobility mass spectrometry-based study of the metalloprotein SmtB, *Proc. 18th International Mass Spectrometry Conference, 2009, Bremen, Germany.*

**Scarff, C. A.**, Quigley, A. M., Lloyd, A. J., Roper, D. I. and Scrivens, J. H., Mass spectrometry-based studies of the Vancomycin resistance pathway in *Enterococcus faecalis*, *Proc. 57th ASMS Conf. on Mass Spectrometry and Allied Topics, 2009, Philadelphia, USA.*

Scrivens, J. H., Kondrat, F. D. L., **Scarff, C. A.**, Blindauer, C. A., Sanghera, N., Hilton, G. R., Gill, A. C., Pinheiro, T. and Thalassinou, K., A shape selective study of conformational changes in metal containing proteins, *Proc. 57th ASMS Conf. on Mass Spectrometry and Allied Topics, 2009, Philadelphia, USA.*

Scrivens, J. H., Thalassinou, K., **Scarff, C. A.**, Hilton, G. R., and Patel, V. J., Gas phase protein structure: Biological significance and comparison with X-ray and NMR measurements, *Proc. 56th ASMS Conf. on Mass Spectrometry and Allied Topics, 2008, Denver, USA.*

Jennings, K. R., Patel, N., **Scarff, C. A.**, Slade, S. E., Thalassinou, K. and Scrivens, J. H., A comparative analysis of electron transfer dissociation and collision-induced dissociation of non-phosphorylated and phosphorylated peptides, *Proc. 30th Annual Meeting of the British Mass Spectrometry Society, 2008, University of York, York.*

**Scarff, C. A.**, Thalassinou, K., Efstathiou, G., Williams, J. P., Green, B. N. and Scrivens, J. H., An expert-based system approach for the identification of single point human hemoglobin variants from mass spectrometric data, *Proc. 55th ASMS Conf. on Mass Spectrometry and Allied Topics, 2007, Indiana, USA.*

Patel, V. J., **Scarff, C. A.**, Thalassinou, K., Williams, J. P. and Scrivens, J. H. Structure-property relationships of hemoglobin tetramers studied by travelling-wave-based ion mobility-mass spectrometry, *Proc. 29th Annual Meeting of the British Mass Spectrometry Society, 2007, Herriot-Watt University, Edinburgh.*

## **1.7 Oral Presentations**

**Scarff, C. A.**, Thalassinou, K., Hilton, G. R., Patel, V. J. and Scrivens, J. H., Gas phase protein structure: Biological significance and comparison with X-ray and NMR measurements, *Proc. 30<sup>th</sup> Annual Meeting of the British Mass Spectrometry Society, 2008, University of York, York.*

## 1.8 References

- Aebersold, R. and Mann, M.** (2003). Mass spectrometry-based proteomics. *Nature*. **422**, 198-207.
- Ashcroft, A. E.** (2005). Recent developments in electrospray ionisation mass spectrometry: noncovalently bound protein complexes. *Natural Product Reports*. **22**, 452-464.
- Aston, F. W.** (1919). A positive ray spectrograph. *Philosophical Magazine*. **38**, 707 - 714.
- Badman, E. R., Myung, S. and Clemmer, D. E.** (2005). Evidence for Unfolding and Refolding of Gas-Phase Cytochrome c Ions in a Paul Trap. *Journal of the American Society for Mass Spectrometry*. **16**, 1493-1497.
- Barber, M., Bordoli, R. S., Sedgwick, R. D. and Tyler, A. N.** (1981). Fast atom bombardment of solids (FAB): a new ion source for mass spectrometry. *Journal of the Chemical Society, Chemical Communications*. **7**, 325-327.
- Barrera, N. P., Di Bartolo, N., Booth, P. J. and Robinson, C. V.** (2008). Micelles Protect Membrane Complexes from Solution to Vacuum. *Science*. **321**, 243-246.
- Benesch, J. L. P., Aquilina, J. A., Ruotolo, B. T., Sobott, F. and Robinson, C. V.** (2006). Tandem Mass Spectrometry Reveals the Quaternary Organization of Macromolecular Assemblies. *Chemistry & Biology*. **13**, 597-605.
- Benesch, J. L. P., Ruotolo, B. T., Simmons, D. A. and Robinson, C. V.** (2007). Protein Complexes in the Gas Phase: Technology for Structural Genomics and Proteomics. *Chemical Reviews*. **107**, 3544-3567.
- Biemann, K.** (1992). Mass spectrometry of peptides and proteins. *Annual Review of Biochemistry*. **61**, 977-1010.
- Bluhm, B. K., Gillig, K. J. and Russell, D. H.** (2000). Development of a Fourier-transform ion cyclotron resonance mass spectrometer-ion mobility spectrometer. *Review of Scientific Instruments*. **71**, 4078-4086.
- Boys, B. L. and Konermann, L.** (2007). Folding and Assembly of Hemoglobin Monitored by Electrospray Mass Spectrometry Using an On-line Dialysis System. *Journal of the American Society for Mass Spectrometry*. **18**, 8-16.
- Breuker, K. and McLafferty, F. W.** (2008). Stepwise evolution of protein native structure with electrospray into the gas phase,  $10^{-12}$  to  $10^2$  s. *Proceedings of the National Academy of Sciences*. **105**, 18145-18152.
- Chernushevich, I. V. and Thomson, B. A.** (2004). Collisional Cooling of Large Ions in Electrospray Mass Spectrometry. *Analytical Chemistry*. **76**, 1754-1760.

**Chiu, W., Baker, M. L. and Almo, S. C.** (2006). Structural biology of cellular machines. *Trends in Cell Biology*. **16**, 144-150.

**Chowdhury, S. K., Katta, V. and Chait, B. T.** (1990). Probing conformational changes in proteins by mass spectrometry. *Journal of the American Chemical Society*. **112**, 9012-9013.

**Clemmer, D. E.**

[http://www.indiana.edu/~clemmer/Research/cross%20section%20database/Proteins/protein\\_cs.htm](http://www.indiana.edu/~clemmer/Research/cross%20section%20database/Proteins/protein_cs.htm). Accessed on: 30th April 2008.

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

**Cohen, M. J. and Karasek, F. W.** (1970). Plasma chromatography TM-new dimension for gas chromatography and mass spectrometry. *Journal of Chromatographic Science*. **8**, 330-337.

**Cohen, S. L. and Chait, B. T.** (2001). Mass Spectrometry as a Tool for Protein Crystallography. *Annual Review of Biophysics and Biomolecular Structure*. **30**, 67-85.

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

**de Hoffmann, E. and Stroobant, V.** (2002). Mass Spectrometry Principles and Applications. Chichester, *John Wiley and Sons, Ltd*.

**Delepierre, M. and Lecroisey, A.** (2001). The interface between microbiology and structural biology as viewed by nuclear magnetic resonance. *Research in Microbiology*. **152**, 697-705.

**Deschamps, J. R. and George, C.** (2003). Advances in X-ray crystallography. *Trends in Analytical Chemistry*. **22**, 561-564.

**Eyles, S. J. and Kaltashov, I. A.** (2004). Methods to study protein dynamics and folding by mass spectrometry. *Methods*. **34**, 88-99.

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

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

**Gidden, J. and Bowers, M. T.** (2003). Gas-phase conformations of deprotonated trinucleotides (dGTT-, dTGT-, and dTTG-): the question of zwitterion formation. *Journal of the American Society for Mass Spectrometry*. **14**, 161-170.

**Gidden, J., Bowers, M. T., Jackson, A. T. and Scrivens, J. H.** (2002). Gas-phase conformations of cationized poly(styrene) oligomers. *Journal of the American Society for Mass Spectrometry*. **13**, 499-505.

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

**Griffiths, J.** (2008). A Brief History of Mass Spectrometry. *Analytical Chemistry*. **80**, 5678-5683.

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

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

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

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

**Heuvel, R. H. H. v. d. and Heck, A. J. R.** (2004). Native protein mass spectrometry: from intact oligomers to functional machineries. *Current Opinion in Chemical Biology*. **8**, 519-526.

**Hill, H. H., Siems, W. F. and St. Louis, R. H.** (1990). Ion mobility spectrometry. *Analytical Chemistry*. **62**, 1201A-1209A.

**Hoaglund-Hyzer, C. S., Counterman, A. E. and Clemmer, D. E.** (1999). Anhydrous Protein Ions. *Chemical Reviews*. **99**, 3037-3080.

**Iavarone, A. T., Udekwu, O. A. and Williams, E. R.** (2004). Buffer Loading for Counteracting Metal Salt-Induced Signal Suppression in Electrospray Ionization. *Analytical Chemistry*. **76**, 3944-3950.

**Jarrold, M. F.** (1999). Unfolding, Refolding, and Hydration of Proteins in the Gas Phase. *Accounts of Chemical Research*. **32**, 360-367.

**Jarrold, M. F. and Bower, J. E.** (1993). Mobilities of metal cluster ions: Aluminum and the electronic shell model. *The Journal of Chemical Physics*. **98**, 2399-2407.

**Jennings, K. R.** (1968). Collision-induced decompositions of aromatic molecular ions. *International Journal of Mass Spectrometry and Ion Physics*. **1**, 227-235.

**Jurchen, J. C. and Williams, E. R.** (2003). Origin of Asymmetric Charge Partitioning in the Dissociation of Gas-Phase Protein Homodimers. *Journal of the American Chemical Society*. **125**, 2817-2826.

**Kaleta, D. T. and Jarrold, M. F.** (2003). Noncovalent Interactions between Unsolvated Peptides: Dissociation of Helical and Globular Peptide Complexes. *Journal of Physical Chemistry B*. **107**, 14529-14536.

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

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

**Karas, M. and Hillenkamp, F.** (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical Chemistry*. **60**, 2299-2301.

**Karasek, F. W.** (1974). Plasma chromatography. *Analytical Chemistry*. **46**, 710A-720a.

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

**Kebarle, P.** (2000). A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *Journal of Mass Spectrometry*. **35**, 804-817.

**Kiselar, J. G. and Downard, K. M.** (2000). Preservation and detection of specific antibody-peptide complexes by matrix-assisted laser desorption ionization mass spectrometry. *Journal of the American Society for Mass Spectrometry*. **11**, 746-750.

**Koeniger, S. L., Merenbloom, S. I., Valentine, S. J., Jarrold, M. F., Udseth, H. R., Smith, R. D. and Clemmer, D. E.** (2006). An IMS-IMS Analogue of MS-MS. *Analytical Chemistry*. **78**, 4161-4174.

**Konermann, L., Collings, B. A. and Douglas, D. J.** (1997a). Cytochrome c Folding Kinetics Studied by Time-Resolved Electrospray Ionization Mass Spectrometry†. *Biochemistry*. **36**, 5554-5559.

**Konermann, L., Rosell, F. I., Mauk, A. G. and Douglas, D. J.** (1997b). Acid-Induced Denaturation of Myoglobin Studied by Time-Resolved Electrospray Ionization Mass Spectrometry†. *Biochemistry*. **36**, 6448-6454.

**Koppelaar, D. W., Barinaga, C. J., Denton, M. B., Sperline, R. P., Hieftje, G. M., Schilling, G. D., Andrade, F. J., Barnes, J. H. and Iv, I. V.** (2005). MS Detectors. *Analytical Chemistry*. **77**, 418 A-427 A.

**Krutchinsky, A. N., Chernushevich, I. V., Spicer, V. L., Ens, W. and Standing, K. G.** (1998). Collisional Damping Interface for an Electrospray Ionization Time-of-Flight Mass Spectrometer. *Journal of the American Society for Mass Spectrometry*. **9**, 569-579.

**Leary, J. A., Schenauer, M. R., Stefanescu, R., Andaya, A., Ruotolo, B. T., Robinson, C. V., Thalassinou, K., Scrivens, J. H., Sokabe, M. and Hershey, J. W. B.** (2009). Methodology for Measuring Conformation of Solvent-Disrupted Protein Subunits using T-WAVE Ion Mobility MS: An Investigation into Eukaryotic Initiation Factors. *Journal of the American Society for Mass Spectrometry*. **20**, 1699-1706.

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

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

**Loo, J. A.** (2000). Electrospray ionization mass spectrometry: a technology for studying noncovalent macromolecular complexes. *International Journal of Mass Spectrometry*. **200**, 175-186.

**Mack, E.** (1925). Average cross-sectional areas of molecules by gaseous diffusion methods. *Journal of the American Chemical Society*. **47**, 2468-2482.

**Mamyrin, B. A., Karataev, V. L., Shmikk, D. V. and V.A., Z.** (1973). The mass reflectron, a new non-magnetic time-of-flight spectrometer with high resolution. *Soviet Physics JETP*. **37**, 45-48.

**Mann, M. and Jensen, O. N.** (2003). Proteomic analysis of post-translational modifications. *Nature Biotechnology*. **21**, 255-261.

**Mason, M. A. and McDaniel, E. W.** (1988). Transport Properties of Ions in Gases, Wiley, New York.

**McCullough, B. J., Kalapothakis, J., Eastwood, H., Kemper, P., MacMillan, D., Taylor, K., Dorin, J. and Barran, P. E.** (2008). Development of an Ion Mobility Quadrupole Time of Flight Mass Spectrometer. *Analytical Chemistry*. **80**, 6336-6344.

**McLafferty, F. W. and Bryce, T. A.** (1967). Metastable-Ion characteristics: Characterization of Isomeric molecules. *Chemical Communications* 1215.

**Merenbloom, S. I., Koeniger, S. L., Valentine, S. J., Plasencia, M. D. and Clemmer, D. E.** (2006). IMS–IMS and IMS–IMS–IMS/MS for Separating Peptide and Protein Fragment Ions. *Analytical Chemistry*. **78**, 2802-2809.

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

**Mohimen, A., Dobo, A., Hoerner, J. K. and Kaltashov, I. A.** (2003). A Chemometric Approach to Detection and Characterization of Multiple Protein Conformers in Solution Using Electrospray Ionization Mass Spectrometry. *Analytical Chemistry*. **75**, 4139-4147.

**Myung, S., Badman, E. R., Lee, Y. J. and Clemmer, D. E.** (2002). Structural Transitions of Electrosprayed Ubiquitin Ions Stored in an Ion Trap over ~10 ms to 30 s. *Journal of Physical Chemistry A*. **106**, 9976-9982.

**Nelson, D. L. and Cox, M. M.** (2005). *Lehninger Principles of Biochemistry*, W. H. Freeman and Company.

**Painter, A. J., Jaya, N., Basha, E., Vierling, E., Robinson, C. V. and Benesch, J. L. P.** (2008). Real-Time Monitoring of Protein Complexes Reveals their Quaternary Organization and Dynamics. *Chemistry & Biology*. **15**, 246-253.

**Pelton, J. T. and McLean, L. R.** (2000). Spectroscopic methods for analysis of protein secondary structure. *Analytical Biochemistry*. **277**, 167-176.

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

**Robinson, C. V., Chung, E. W., Kragelund, B. B., Knudsen, J., Aplin, R. T., Poulsen, F. M. and Dobson, C. M.** (1996). Probing the Nature of Noncovalent Interactions by Mass Spectrometry. A Study of Protein CoA Ligand Binding and Assembly. *Journal of the American Chemical Society*. **118**, 8646-8653.

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

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

**Ruotolo, B. T., Giles, K., Campuzano, I., Sandercock, A. M., Bateman, R. H. and Robinson, C. V.** (2005). Evidence for Macromolecular Protein Rings in the Absence of Bulk Water. *Science*. **310**, 1658-1661.

**Ruotolo, B. T., Hyung, S.-J., Robinson, P. M., Giles, K., Bateman, R. H. and Robinson, C. V.** (2007). Ion Mobility-Mass Spectrometry Reveals Long-Lived, Unfolded Intermediates in the Dissociation of Protein Complexes. *Angewandte Chemie International Edition*. **46**, 8001-8004.

**Ruotolo, B. T. and Robinson, C. V.** (2006). Aspects of native proteins are retained in vacuum. *Current Opinion in Chemical Biology*. **10**, 402-408.

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

**Scrivens, J. H., Thalassinos, K., Hilton, G., Slade, S. E., Pinheiro, T. J. T., Bateman, R. H. and Bowers, M. T.** (2006). Use of a Travelling Wave-Based Ion Mobility Approach to Resolve Proteins of Varying Conformation. *Proc. 55th ASMS Conf. Mass Spectrometry and Allied Topics*, Indianapolis.

**Sharon, M. and Robinson, C. V.** (2007). The Role of Mass Spectrometry in Structure Elucidation of Dynamic Protein Complexes. *Annual Review of Biochemistry*. **76**, 167-193.

**Sharon, M., Taverner, T., Ambroggio, X. I., Deshaies, R. J. and Robinson, C. V.** (2006). Structural Organization of the 19S Proteasome Lid: Insights from MS of Intact Complexes. *PLoS Biology*. **4**, e267.

**Shelimov, K. B., Clemmer, D. E., Hudgins, R. R. and Jarrold, M. F.** (1997). Protein Structure in Vacuo: Gas-Phase Conformations of BPTI and Cytochrome c. *Journal of the American Chemical Society*. **119**, 2240-2248.

**Shelimov, K. B. and Jarrold, M. F.** (1997). Conformations, Unfolding, and Refolding of Apomyoglobin in Vacuum: An Activation Barrier for Gas-Phase Protein Folding. *Journal of the American Chemical Society*. **119**, 2987-2994.

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

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

**Smith, R. D., Lightwahl, K. J., Winger, B. E. and Loo, J. A.** (1992). Preservation of noncovalent associations in electrospray ionisation-mass spectrometry: Multiply charged polypeptide and protein dimers. *Organic Mass Spectrometry*. **27**, 811-821.

**Sobott, F., Benesch, J. L. P., Vierling, E. and Robinson, C. V.** (2002a). Subunit Exchange of Multimeric Protein Complexes. *Journal of Biological Chemistry*. **277**, 38921-38929.

**Sobott, F., Hernández, H., McCammon, M. G., Tito, M. A. and Robinson, C. V.** (2002b). A Tandem Mass Spectrometer for Improved Transmission and Analysis of Large Macromolecular Assemblies. *Analytical Chemistry*. **74**, 1402-1407.

**Srebalus, C. A., Li, J., Marshall, W. S. and Clemmer, D. E.** (1999). Gas-Phase Separations of Electrosprayed Peptide Libraries. *Analytical Chemistry*. **71**, 3918-3927.

**Steinberg, M. Z., Elber, R., McLafferty, F. W., Gerber, R. B. and Breuker, K.** (2008). Early Structural Evolution of Native Cytochrome c after Solvent Removal. *ChemBioChem*. **9**, 2417-2423.

**Strupat, K., Sagi, D., Bonisch, H., Schafer, G. and Peter-Katalinic, J.** (2000). Oligomerization and substrate binding studies of the adenylate kinase from *Sulfolobus acidocaldarius* by matrix-assisted laser desorption/ionization mass spectrometry. *Analyst*. **125**, 563-567.

**Tahallah, N., Pinkse, M. M., C. S. and Heck, A. J. R.** (2001). The effect of the source pressure on the abundance of ions of noncovalent protein assemblies in an electrospray ionization orthogonal time-of-flight instrument. *Rapid Communications in Mass Spectrometry*. **15**, 596-601.

**Thalassinos, K., Grabenauer, M., Slade, S. E., Hilton, G. R., Bowers, M. T. and Scrivens, J. H.** (2009). Characterization of Phosphorylated Peptides Using Traveling Wave-Based and Drift Cell Ion Mobility Mass Spectrometry. *Analytical Chemistry*. **81**, 248-254.

**Thalassinos, K., Slade, S. E., Jennings, K. R., Scrivens, J. H., Giles, K., Wildgoose, J., Hoyes, J., Bateman, R. H. and Bowers, M. T.** (2004). Ion mobility mass spectrometry of proteins in a modified commercial mass spectrometer. *International Journal of Mass Spectrometry*. **236**, 55-63.

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

**Thomson, J. J.** (1911). Rays of positive electricity. *Philosophical Magazine*. **6**, 752-767.

**Valentine, S. J., Anderson, J. G., Ellington, A. D. and Clemmer, D. E.** (1997a). Disulfide-Intact and -Reduced Lysozyme in the Gas Phase: Conformations and Pathways of Folding and Unfolding. *Journal of Physical Chemistry B*. **101**, 3891-3900.

**Valentine, S. J., Counterman, A. E. and Clemmer, D. E.** (1997b). Conformer-Dependent Proton-Transfer Reactions of Ubiquitin Ions. *Journal of the American Society for Mass Spectrometry*. **8**, 954-961.

**van Duijn, E.** (2010). Current Limitations in Native Mass Spectrometry Based Structural Biology. *Journal of the American Society for Mass Spectrometry*. **21**, 971-978.

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

**Voet, D. and Voet, J. G.** (1995). *Biochemistry*, John Wiley & Sons.

**von Helden, G., Hsu, M.-T., Kemper, P. R. and Bowers, M. T.** (1991). Structures of carbon cluster ions from 3 to 60 atoms: Linears to rings to fullerenes. *The Journal of Chemical Physics*. **95**, 3835-3837.

**von Helden, G., Hsu, M., Gotts, N. and Bowers, M.** (1993). Carbon cluster cations with up to 84 atoms: Structures, formation mechanism, and reactivity. *Journal of Physical Chemistry*. **97**, 8182-8192.

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

**Wiley, W. C. and McLaren, I. H.** (1955). Time-of-Flight Mass Spectrometer with Improved resolution. *Review of Scientific Instruments*. **26**, 1150-1157.

**Williams, J. P. and Scrivens, J. H.** (2008). Coupling desorption electrospray ionisation and neutral desorption/extractive electrospray ionisation with a travelling-wave based ion mobility mass spectrometer for the analysis of drugs. *Rapid Communications in Mass Spectrometry*. **22**, 187-196.

**Wilm, M. and Mann, M.** (1996). Analytical Properties of the Nanoelectrospray Ion Source. *Analytical Chemistry*. **68**, 1-8.

**Wilm, M. S. and Mann, M.** (1994). Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last? *International Journal of Mass Spectrometry and Ion Processes*. **136**, 167-180.

**Wiza, J. L.** (1979). Microchannel plate detectors. *Nuclear Instruments and Methods*. **162**, 587-601.

**Wytttenbach, T., von Helden, G. and Bowers, M. T.** (1996). Gas-Phase Conformation of Biological Molecules: Bradykinin. *Journal of the American Chemical Society*. **118**, 8355-8364.

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

**Zubarev, R. A., Kelleher, N. L. and McLafferty, F. W.** (1998). Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *Journal of the American Chemical Society*. **120**, 3265-3266.

# Chapter 2: Travelling-Wave Ion Mobility Mass Spectrometry Studies of Protein Structure

---

## 2.1 Introduction

The tertiary conformation of a protein is vital in determining its biological activity and as such is a key structural feature. The most established methods for studying protein conformation are X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. These techniques do not allow for the analysis of all proteins; many proteins do not form crystals and NMR studies require high concentrations of labelled protein. This excludes the use of proteins obtained directly from biological samples. Techniques that allow the measurement of protein conformation at biological concentration and within biological timescales are of great value in understanding biological processes.

The characterisation of aspects of three-dimensional protein structure by mass spectrometry is an area of much interest as the gas phase conformation, in many instances, can be related to that of the solution phase. Previous studies by drift-cell ion mobility mass spectrometry (DCIM-MS) have shown that cross-sections estimated for the lower ESI-generated charge states of several proteins, including bovine cytochrome c (Shelimov *et al.* 1997), bovine ubiquitin (Myung *et al.* 2002), and equine apomyoglobin (Shelimov and Jarrold 1997), compare well with those calculated from X-ray crystal structures. As the charge state increases there is an increase in overall cross-section of the molecule. This is thought to be attributable to the effects of Coulomb repulsion (Shelimov and Jarrold 1997; Badman *et al.* 2001).

In this work, the biological significance of gas-phase protein structure measurements made by the use of travelling-wave ion mobility mass spectrometry (TWIM-MS) was investigated.

As described in Chapter 1 Section 1.1.7, the travelling-wave ion mobility separator (T-Wave) (Giles *et al.* 2004) has considerably increased the sensitivity and speed of analysis in comparison with many earlier drift-cell ion mobility devices, so enabling one to analyse samples at biologically-relevant concentrations. The T-Wave has been integrated into a commercial quadrupole time-of-flight instrument, the Synapt HDMS System (Waters Corporation, Milford, MA, USA) (Pringle *et al.* 2007). The

Synapt operates with excellent reproducibility and mass accuracy, and is capable of providing a wealth of information since it is able to achieve ion mobility tandem mass spectrometry with the choice of performing collision induced dissociation before or after the mobility cell. For each ion mobility experiment, 200 complete mass spectra are acquired in a time of 15-30 ms. The nature of the T-Wave device means that the ion drift time is not inversely proportional to ion mobility as is the case when using DCIM-MS, and thus absolute collision cross-sections cannot be obtained directly from drift time measurements. It has been shown, however, that estimates of collision cross-sections can be obtained by reference to samples with known cross-sections (Ruotolo *et al.* 2005; Wildgoose *et al.* 2006; Thalassinos *et al.* 2008), provided that the data are obtained under the same experimental conditions; i.e. mobility gas, gas pressure, wave velocity, wave height, pusher frequency and injection energy.

Protein standards were analysed by means of TWIM-MS under denaturing and near-physiological solvent conditions. For each protein studied, mobility drift times for each of the charge states detected within spectra were recorded. Estimates of rotationally-averaged collision cross-sections for the charge states observed were obtained with reference to published cross-sections of known standards. These were compared with those calculated by the well-established MOBCAL program (Mesleh *et al.* 1996; Shvartsburg and Jarrold 1996). Cross-sections were calculated, for the various proteins studied, from published X-ray crystallography structures and NMR structures held in the protein databank (Berman *et al.* 2000).

The cross-section of a protein measured by ion mobility mass spectrometry varies with charge state. This allows the unfolding of proteins in the gas phase to be studied. The unfolding transition for disulphide-reduced and disulphide-intact lysozyme c was investigated by TWIM-MS to examine the restraints on protein unfolding attributable to the presence of disulphide bonds.

## **2.2 Materials and Methods**

### **2.2.1 Material suppliers**

All chemical reagents (analytical grade) and protein standards were obtained from Sigma-Aldrich Ltd (Gillingham, UK), Fisher Scientific (Loughborough, UK) or GE Healthcare (Chalfont St Giles, UK) unless otherwise stated. Mass spectrometry solvents were supplied by Mallinckrodt Baker Inc. (Phillipsburg, NJ, USA).

### **2.2.2 Sample preparation**

Protein standards (chicken lysozyme c, human lysozyme c, sperm whale myoglobin, equine myoglobin and equine cytochrome c) were diluted to a concentration of 10  $\mu$ M in 50 % acetonitrile 0.1 % formic acid for denaturing conditions and in 10 mM ammonium acetate pH 6.8 for near-native conditions.

Prior to mass spectrometric analysis, samples were desalted by means of centrifugation. 500  $\mu$ L of each sample was added to a pre-washed Ultrafree-0.5 5 kDa cut-off (Millipore, Billerica, MA, USA) centrifugal filter which was then centrifuged at 11,500 g for 10 minutes. Sample retained on the filter was reconstituted in the appropriate solvent (either 50 % acetonitrile 0.1 % formic acid or 10 mM ammonium acetate pH 6.8) and the centrifugation step repeated. The resulting samples were recovered from the filter devices into Eppendorf tubes and reconstituted to a volume of 500  $\mu$ l in the appropriate solvent.

The disulphide-bonds within lysozyme samples were reduced by dithiothreitol (DTT). Human and chicken lysozyme c protein stocks were prepared at 1 mg/mL in a 10 mM ammonium bicarbonate solution. A solution of DTT was prepared at 60 mg/mL in 100 mM ammonium bicarbonate. 10  $\mu$ L of the DTT solution was added to 90  $\mu$ L of each lysozyme protein stock. The resulting solutions were incubated for five minutes at room temperature. Lysozyme c samples which had been treated with DTT (disulphide-reduced) and untreated (disulphide-intact) protein stocks were

diluted to a concentration of 10  $\mu\text{M}$  in 50 % acetonitrile 0.1 % formic acid before undergoing mass spectrometric analysis.

### **2.2.3 Travelling-wave ion mobility mass spectrometry experiments**

Travelling-wave ion mobility mass spectrometry (TWIM-MS) experiments were performed on a Synapt HDMS System (Waters Corporation, Milford, MA, USA), described in detail in Chapter 1 Section 1.1.7 and elsewhere (Pringle *et al.* 2007). Samples were introduced into the source region of the Synapt HDMS system by direct infusion nano-ESI by means of fused silica nanospray needles (Waters Corporation, Milford, MA, USA). The instrument was operated in ESI positive mode with a capillary voltage of 1.2 kV, cone voltage of 70 V and source temperature of 90 °C for all experiments. The TOF mass analyser was tuned in V-mode to give an operating resolution of 8,000 (FWHM) and was calibrated using 2 mg/ml cesium iodide in 50 % aqueous propan-2-ol. A mass acquisition range of 800-3000  $m/z$  was used. The pressure within the ion mobility cell, containing nitrogen, was optimised by changing the nitrogen flow rate. Optimised separation was achieved at an indicated pressure of 0.55 mbar, corresponding to a flow rate of 30 mL/min. The travelling wave velocity and wave height were altered in increments from 100-600 m/s and 8-20V respectively. The conditions that provided the optimal mobility separation, a travelling wave height and velocity of 9 V and 300 m/s respectively, were used for all following experiments.

MassLynx™ (v4.1) software (Waters Corporation, Milford, MA, USA) was used to acquire data and perform subsequent processing required. Mass spectra were recorded at an acquisition rate of two spectra/s with an interscan delay of 100 ms. Spectra obtained were deconvoluted onto a true mass scale with MaxEnt, a maximum entropy modelling program available within MassLynx™ (v4.1) software.

### **2.2.4 ATDs, calibration and estimation of cross-sections**

The arrival time distributions for each ion of interest were extracted and interpreted. Each peak observed within an ATD for a particular ion was identified as relating to a particular conformation of that ion. The scan number (TOF push number) at which

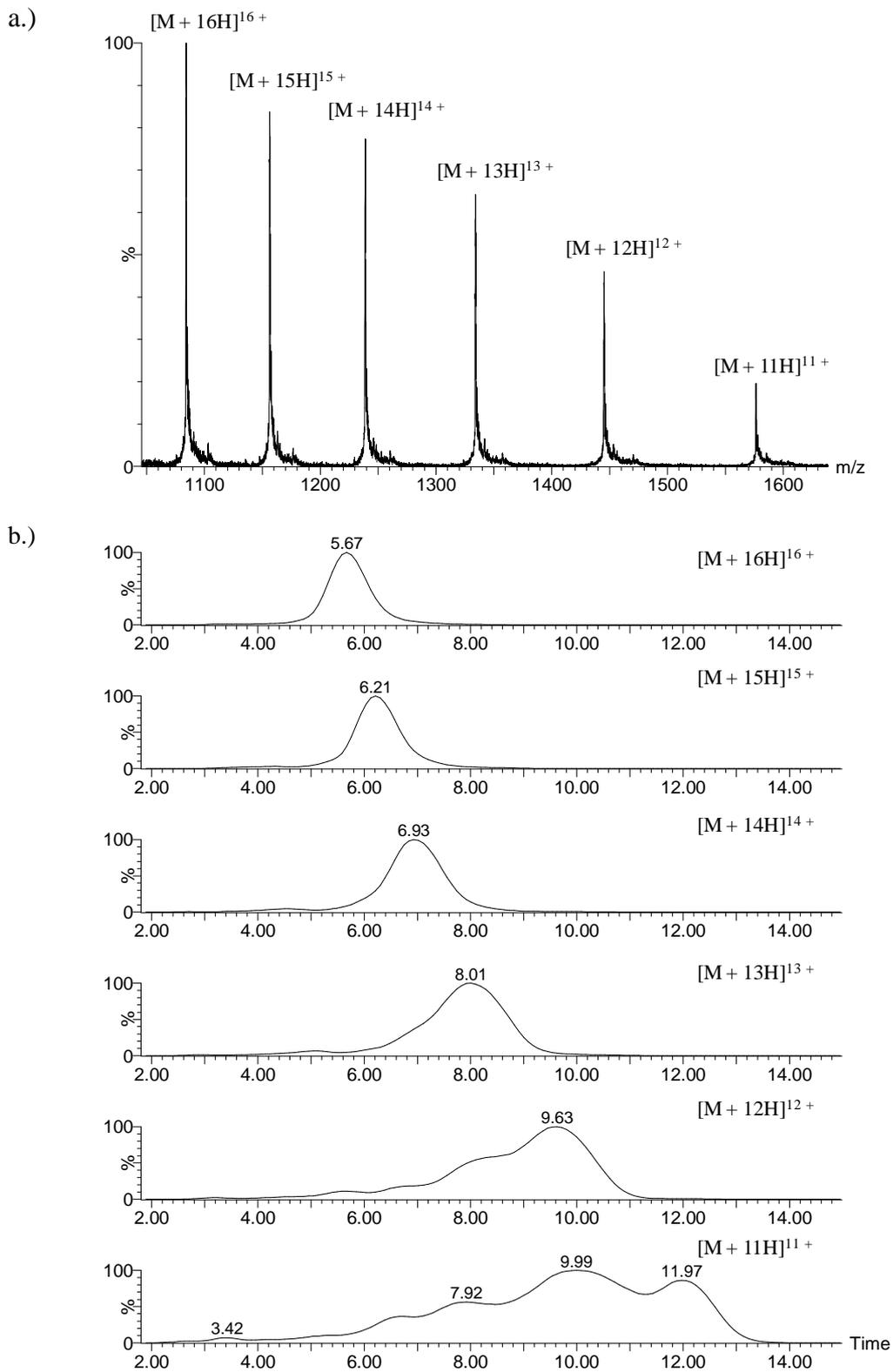
the apex of each peak occurred was taken and converted into an arrival time (the arrival scan number multiplied by the pusher frequency).

Calibration methods for the T-Wave device, to allow estimations of rotationally-averaged collision cross-sections from ion mobilities obtained, have been previously described (Ruotolo *et al.* 2005; Scrivens *et al.* 2006; Wildgoose *et al.* 2006; Ruotolo *et al.* 2008; Thalassinos *et al.* 2008). Here calibrations were performed following an in-house procedure based on earlier work, described in detail in Chapter 1 Section 1.1.8 and briefly below (Ruotolo *et al.* 2005; Scrivens *et al.* 2006; Wildgoose *et al.* 2006; Williams and Scrivens 2008).

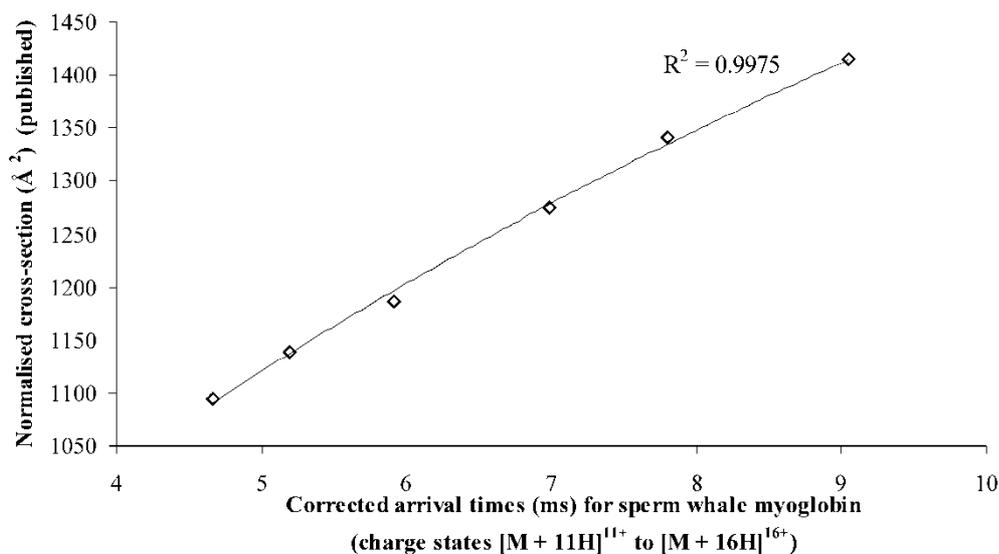
Equine myoglobin or sperm whale myoglobin at a concentration of 10 $\mu$ M in 50 % aqueous acetonitrile containing 0.2 % formic acid was used as the T-Wave calibrant. The myoglobin was analysed under the same instrument conditions that were used to analyse the analyte of interest. Sperm whale myoglobin was used here.

Recorded arrival times were corrected to exclude time spent outside of the ion mobility cell. Mass-independent time, spent in the transfer region, and mass-dependent time, spent between the transfer region and time-of-flight mass analyser, were subtracted. These corrected arrival times were used to create a calibration curve for cross-section measurements against values obtained from drift cell ion mobility mass spectrometry (DCIM-MS) studies. Cross-section measurements from DCIM-MS studies for sperm whale myoglobin were obtained from Clemmer's on-line database whilst those for equine myoglobin were obtained from Prof. Michael T. Bowers (personal communication). Normalised cross-sections (corrected for charge and reduced mass) were plotted against corrected arrival times to create a calibration with a power series fit. The calibration produced was then used to estimate the rotationally-averaged collision cross-sections for various charge states of the proteins studied under the same instrument operating conditions.

A typical spectrum obtained for sperm whale myoglobin, for calibration purposes, is shown in Figure 2.1, along with corresponding extracted arrival time distributions for charge states  $[M + 11H]^{11+}$  to  $[M + 16H]^{16+}$ . An example of a calibration curve is illustrated in Figure 2.2.



**Figure 2.1:** a.) A typical spectrum obtained for sperm whale myoglobin, for calibration purposes, with b.) corresponding extracted arrival time distributions for charge states  $[M + 11H]^{11+}$  to  $[M + 16H]^{16+}$ .



**Figure 2.2:** Calibration curve of corrected arrival times for sperm whale myoglobin against normalised published cross-sections.

The calibration allows one to estimate the cross-section of a molecule of interest, when analysed under the same instrument conditions as the calibrant, provided that the mobilities (corrected arrival times) for that molecule lie within the mobilities observed for the calibrant. This is irrespective of the size range of cross-sections for the calibrant (Shvartsburg and Smith 2008; Thalassinou *et al.* 2008). The calibration produced was only used to estimate rotationally-averaged collision cross-sections for ions of interest if their corrected arrival times fell along the calibration curve.

### 2.2.5 Theoretical cross-section calculations

To evaluate the measurements for rotationally-averaged collision cross-section obtained experimentally, comparisons were made with theoretically calculated approximations of the measurement, where possible. MOBCAL, an open source program to calculate mobilities (Mesleh *et al.* 1996; Shvartsburg and Jarrold 1996), was used to calculate cross-sections from published NMR and X-ray crystallography structures held at the RCSB Protein Data Bank (Berman *et al.* 2000).

MOBCAL facilitates the use of three approximations to calculate cross-sections. The simplest method is the projection approximation (PA) which replaces the cross-section of an ion with its projection (shadow) and averages the projections created by every orientation of the ion. The PA is an adequate approximation for small molecules but underestimates the cross-section of proteins with highly convex structures where interactions with the buffer gas become important (Jarrold 1999). The trajectory method (TM) gives the most reliable estimate, incorporating all interactions but is computationally intense. A compromise is to use a third model, the exact hard sphere scattering (EHSS) method. This ignores electrostatic interactions so requires substantially less computational time, and can calculate cross-sections to within a few percent of values obtained by the trajectory method (Jarrold 1999).

To use MOBCAL to calculate cross-sections from NMR and X-ray structures an input file for MOBCAL needs to be generated and the MOBCAL code needs to be modified. Force 2.0 (free distribution software) was used to compile and edit the FORTRAN script. The input file needs to be in a specific format and contain the Cartesian coordinates from the PDB file. A Perl script was used to extract the coordinates from each appropriate PDB file and to create an input file. The MOBCAL code was originally written for the modelling of small molecules and so the pre-defined atom capacity is too small and needs to be increased. The 'len' variable was increased so that it was greater than the number of atoms within the input file. The number of iterations used by MOBCAL to calculate cross-section from an input file was also increased as suggested within the MOBCAL instructions file. These required modifications have previously been discussed by Ruotolo *et al.* (2008).

PA and EHSS approximations were performed on a Dell workstation with a Dual Core Intel 3.0 GHz processor running Windows XP Professional. Repeats of these computations and TM approximations were performed on a Mac Pro with 2 Dual Core Intel Xeon 3.0 GHz processors running Mac OS X 10.4. Each TM approximation took 7 to 10 days to run on this system.

## 2.3 Results and Discussion

### 2.3.1 Theoretical cross-section calculations

Theoretical cross-sections, calculated by the PA, EHSS and TM models in MOBCAL from X-ray and NMR structures for the proteins standards studied, are shown in Table 2.1. Cross-sections calculated from X-ray crystallographic data and NMR structures, for the same protein, show remarkable agreement (within 3 % of each other here), as has been observed previously (Jarrold 1999). Nominal differences (approximately 2 %) are observed in the cross-sections calculated for the same protein from different organisms, which is not unexpected given their structural similarities and common functions. Cross-sections calculated, for the proteins studied, by use of the EHSS and TM approximations are in good agreement (within a few percent).

**Table 2.1:** Theoretical cross-sections for standard proteins calculated from published NMR and X-ray structures using the MOBCAL program and the PA, EHSS and TM models.

Protein				Theoretical cross-section ( $\text{\AA}^2$ )		
Description	Swiss-Prot ID	Experimental Method	PDB file	PA	EHSS	TM
Chicken lysozyme c	LYSC_CHICK	X-ray	1DPX	1180	1475	1499
		NMR	1GXX	1172	1451	1445
Human lysozyme c	LYSC_HUMAN	X-ray	2NWD	1218	1523	1551
		NMR	1IY3	1244	1568	1563
Sperm whale myoglobin	MYG_PHYCA	X-ray	1VXG	1375	1734	1730
		NMR	1MYF	1394	1754	1753
Equine myoglobin	MYG_HORSE	X-ray	1WLA	1366	1719	1716
Equine cytochrome c	CYC_HORSE	X-ray	1HRC	1055	1318	1313
		NMR	1LC1	1065	1324	1314

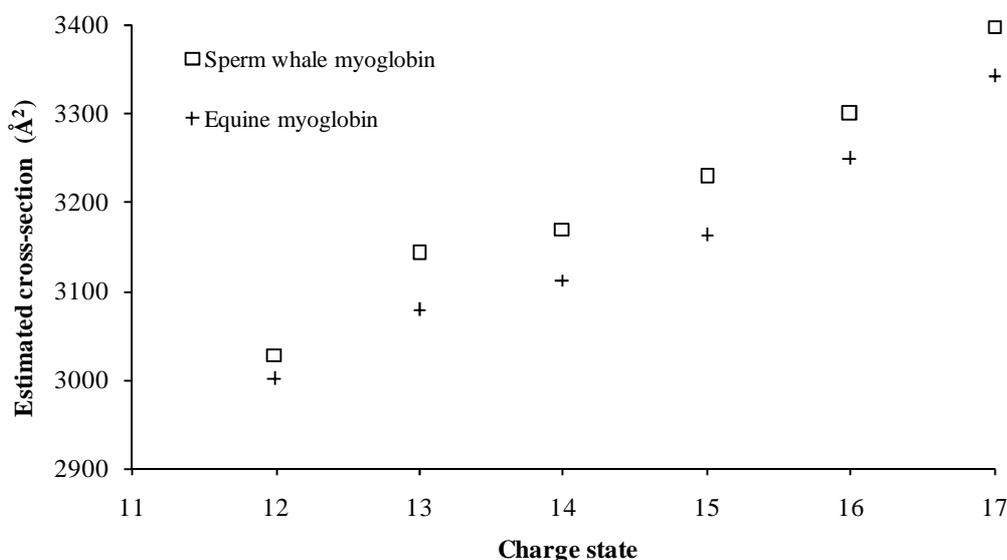
Jarrold has previously noted that cross-sections calculated for proteins by the EHSS method are within a few percent of those calculated by the trajectory method. This indicates that performing the computationally intensive TM approximation is not necessary in these cases. The PA model, however, has been shown to underestimate the cross-section of larger molecules. For larger molecules, interactions with the buffer gas have been observed to have a increased effect on mobility (Jarrold 1999).

Care needs to be taken when comparing estimated cross-sections generated by use of MOBCAL from different input structures and values obtained by experiment. X-ray crystallography structures can contain ligands, which may have bound due to the crystallisation conditions and which may not be natural ligands. If MOBCAL is applied to a PDB structure with missing hydrogen atoms, or one containing a different ligand, the cross-section calculated may be quite different from the cross-section calculated for the complete structure. Hopper and Oldham investigated collision induced unfolding of apo- and holo- protein ions in the gas phase by IM-MS (Hopper and Oldham 2009). They used MOBCAL to give theoretical approximations for the collision cross-sections of these proteins and typically reported a difference in calculated cross-section between apo- and holo- forms of 5 %. All PDB files used here were modified appropriately, where possible, to make sure that all hydrogen atoms and ligands were accounted for.

### **2.3.2 TWIM calibration**

Calibration of the TWIM device is made possible by comparing mobilities obtained for an analyte with those obtained for a standard, for which absolute cross-sections are known. The only publicly available database containing protein cross-sections, for multiple charge states, come from Clemmer's group (Clemmer). The protein with the largest cross-sections for which data is available is apomyoglobin. The stated source of this myoglobin was sperm whale yet equine myoglobin is the more widely available standard. It was, therefore, of interest to compare the mobilities of these two species. The correlation between estimated cross-sections for sperm whale myoglobin and equine myoglobin obtained for various charge states, under denaturing conditions, is illustrated in Figure 2.3.

For simplicity only charge states for which a single conformation was observed are shown. Equine myoglobin has a molecular weight of 16951.5 Da and sperm whale myoglobin, in its naturally occurring form, a molecular weight of 17199.9 Da. The sperm whale material used here was sourced from Sigma-Aldrich, UK as a recombinant product, expressed in *E. coli*, which has an N-terminus methionine present giving a molecular weight of 17331.1 Da. The equine and sperm whale myoglobins have an 87 % sequence identity (Altschul *et al.* 1997).

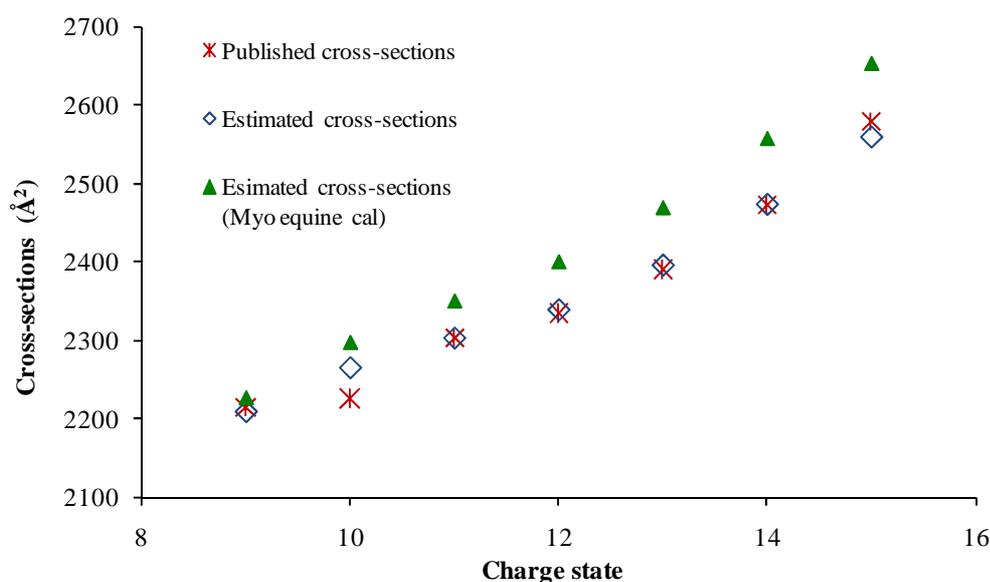


**Figure 2.3:** Comparison of estimated cross-sections for sperm whale myoglobin and equine myoglobin obtained under denaturing conditions.

The cross-sections estimated for the various charge states of the two proteins are similar but distinct; consistently, equine myoglobin is estimated to have a smaller cross-section than sperm whale myoglobin. The difference in cross-section observed is approximately 2 % and reflects the powerful resolving ability of this technique. This result indicates that the same protein from a different organism cannot be assumed to have cross-sections of sufficient similarity for them to be interchangeable for calibration purposes.

Experimentally estimated cross-sections for equine cytochrome c under denaturing conditions are shown in Figure 2.4 and are compared to published values obtained from DCIM-MS experiments (Clemmer). By use of the calibration created from

sperm whale myoglobin data, equine cytochrome c cross-sections may be estimated to within 2 % of absolute values. This experiment was repeated regularly over several months, under multiple experimental conditions and results were extremely reproducible with estimated cross-sections consistently within 2 % of published values. This, along with previous studies (Ruotolo *et al.* 2005; Wildgoose *et al.* 2006; Thalassinos *et al.* 2008), provides substantial evidence to support that the Synapt TWIM cell can be calibrated successfully by use of standards with known cross-sections. Measurements for equine cytochrome c are within 1 % of published values, with the exception of the  $[M + 10H]^{10+}$  charge state. This anomaly has been observed on various occasions and given the agreement observed between the rest of the charge states it is plausible that this may be because the conformation observed in the TWIM-MS experiment is different to that observed and published in the DCIM-MS instrument.



**Figure 2.4:** Comparison of published with experimentally estimated cross-sections for different charge states of equine cytochrome c (under denaturing conditions).

Cross-sections were estimated utilising a calibration of data obtained for sperm whale myoglobin against published values for sperm whale myoglobin and utilising a calibration of data obtained for equine myoglobin against published values for sperm whale myoglobin.

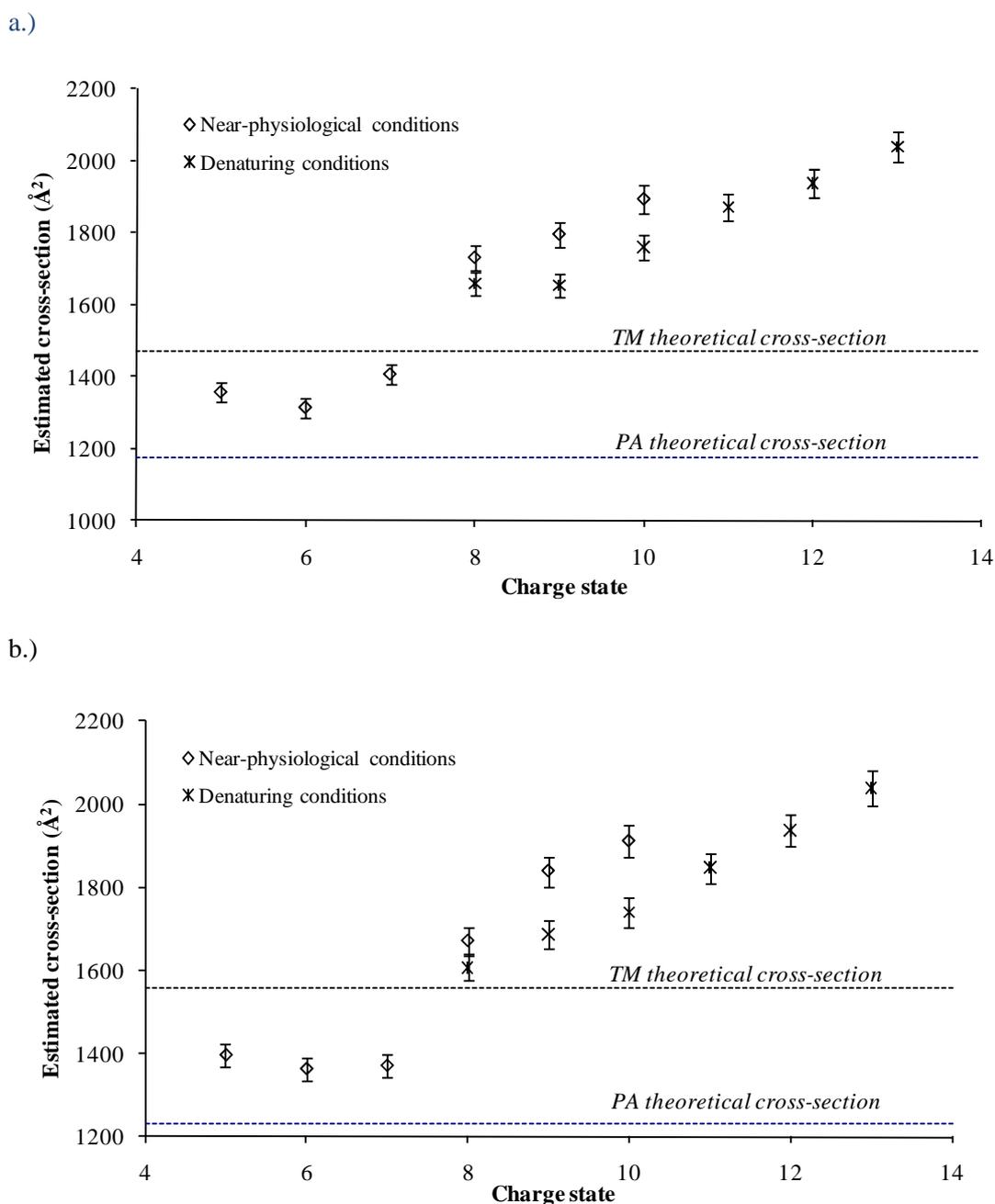
The agreement is not so impressive when the calibration curve is produced by using corrected arrival times for equine myoglobin but the absolute cross-sections for sperm whale myoglobin. Two of the seven cytochrome c charge states evaluated have estimated cross-sections within 2 % of published values when the calibration is performed as described above. All measurements are still, however, within 4 % of published values. These findings show that the use of equine myoglobin in place of sperm whale myoglobin can still produce reproducible results and yet highlights the resolving capabilities of the TWIM-MS approach.

It is the potential of this technique to probe biologically relevant conformations which is of the most prospective interest. The timescale of the TWIM-MS approach means that biologically relevant conformations can be preserved within the gas phase and not have time to rearrange into stable gas-phase structures (Badman *et al.* 2005; Breuker and McLafferty 2008). In order to evaluate this aspect of the technique, experiments were performed to obtain estimates of cross-sections for protein standards under denaturing and near-native conditions for different charge states.

### **2.3.3 Estimated cross-sections**

Figure 2.5 shows estimated cross-sections for chicken and human lysozyme c, respectively, obtained under both denaturing and near-physiological conditions.

The estimates of cross-section obtained for the protein standards, based on measurements conducted over a number of months, are reproducible between datasets to within  $\pm 2$  %. This estimated error is indicated as error bars within the relevant Figures.

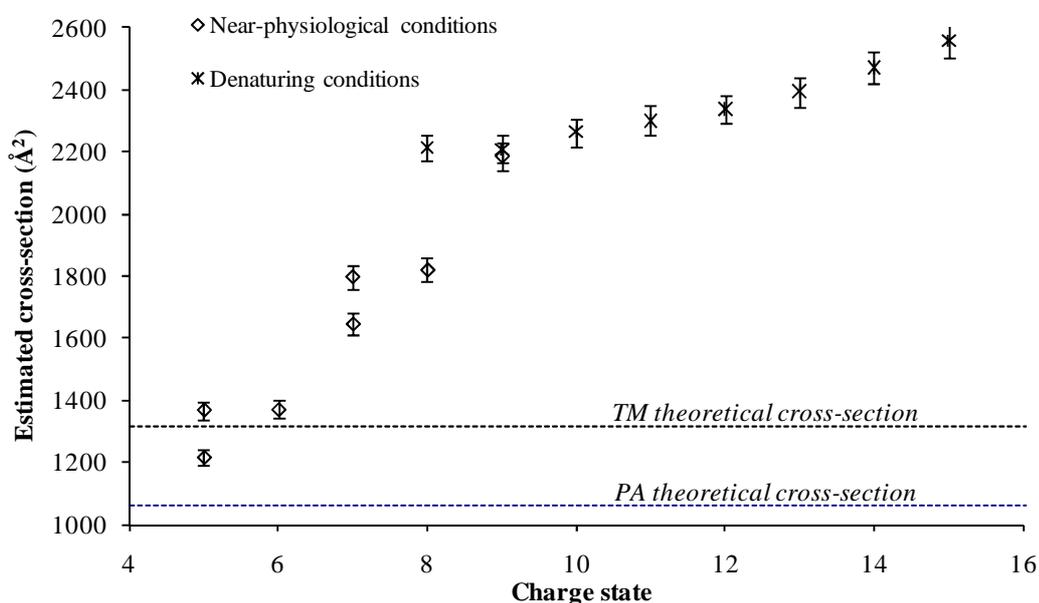


**Figure 2.5:** Experimentally estimated cross-sections for different charge states of a.) chicken and b.) human lysozyme c under denaturing and near-physiological conditions with average theoretical cross-sections for PA and TM models shown. Error bars are shown to illustrate measurements that would be within  $\pm 2\%$ .

Under physiological conditions conformations with significantly smaller cross-sections are observed for charge states  $[M + 5H]^{5+}$  to  $[M + 7H]^{7+}$ , in comparison to charge states  $[M + 8H]^{8+}$  to  $[M + 10H]^{10+}$ . These values fall between those calculated by means of the PA and TM approximations. These conformations have similar

cross-sections suggesting that these three charge states exist within similar conformations within the gas phase. Conformations determined under both denaturing and near-physiological conditions for charge states  $[M + 8H]^{8+}$  to  $[M + 10H]^{10+}$  are significantly different. This may be attributable to the different solvent conditions used.

Estimated cross-sections obtained for equine cytochrome c under both denaturing and near-physiological conditions are shown in Figure 2.6 below.

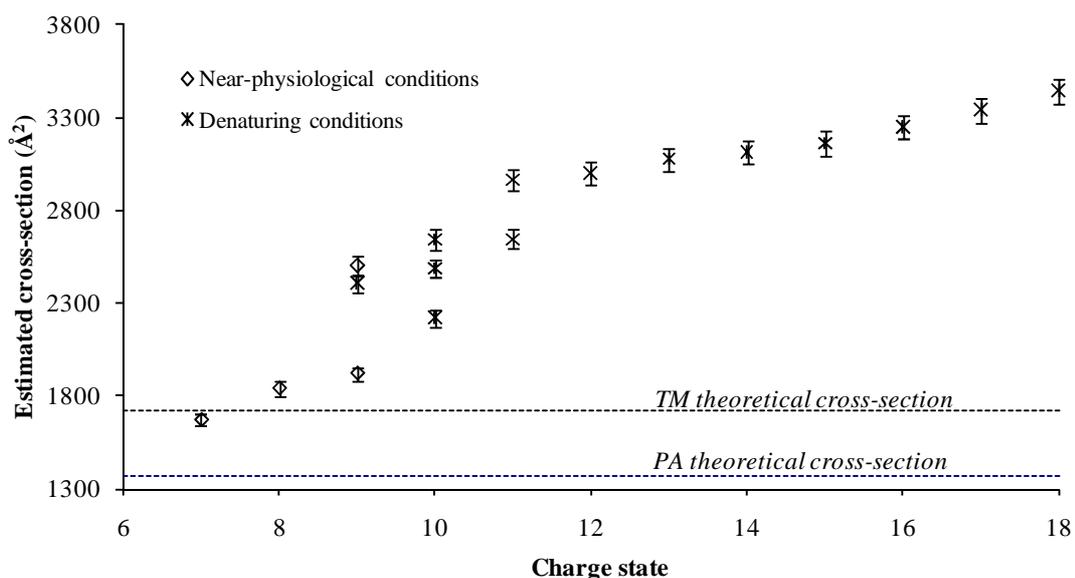


**Figure 2.6:** Experimentally estimated cross-sections for different charge states of equine cytochrome c under denaturing and near-physiological conditions with average theoretical cross-sections for PA and TM models shown. Error bars are shown to illustrate measurements that would be within  $\pm 2\%$ .

The change in cross-section with increase in charge state is quite different for cytochrome c from that of the two lysozymes. Lysozyme c possesses four disulphide bonds, which are maintained in the gas phase under denaturing as well as near-physiological conditions. Despite lysozyme c having a larger molecular weight than cytochrome c, cross-sections calculated for lysozyme c (Figure 2.5) under denaturing conditions are smaller than those observed for cytochrome c (Figure 2.6). Fewer charge states are observed for lysozyme c than for cytochrome c. The difference in

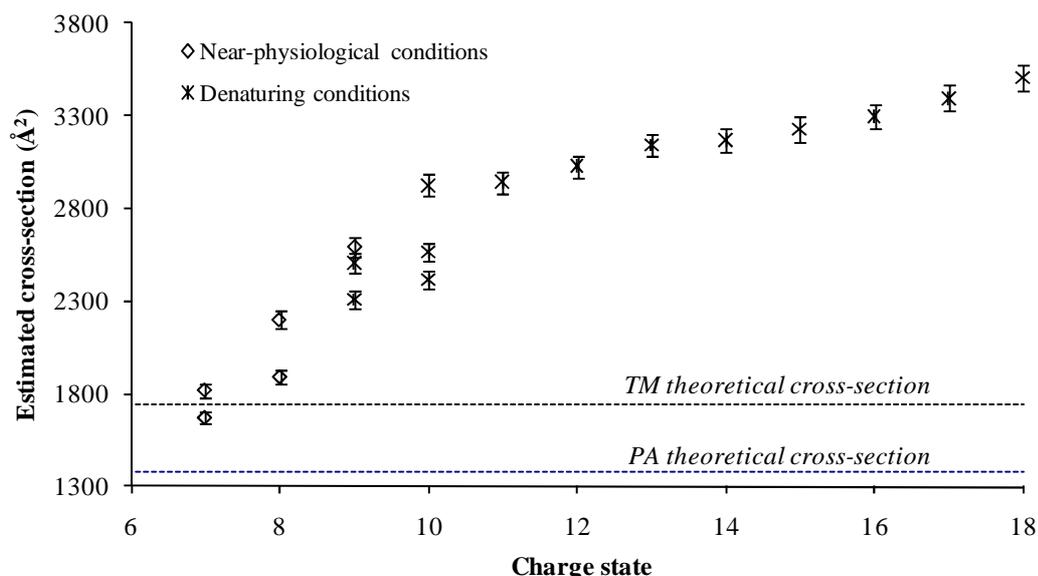
cross-section between the most compact and least compact structures observed for cytochrome c is much greater than that observed for the lysozymes.

Estimated cross-sections obtained for equine myoglobin and sperm whale myoglobin under both denaturing and near-physiological conditions are shown in Figures 2.7 and 2.8 respectively.



**Figure 2.7:** Experimentally estimated cross-sections for different charge states of equine myoglobin under denaturing and near-physiological conditions with average theoretical cross-sections for PA and TM models shown. Error bars are shown to illustrate measurements that would be within  $\pm 2\%$ .

Multiple conformations, for some charge states are observed, under the same solvent conditions, for cytochrome c and both myoglobins but not for lysozyme c. Conformations observed under near-physiological conditions for cytochrome c and myoglobin vary more significantly in cross-section than those observed for lysozyme c.



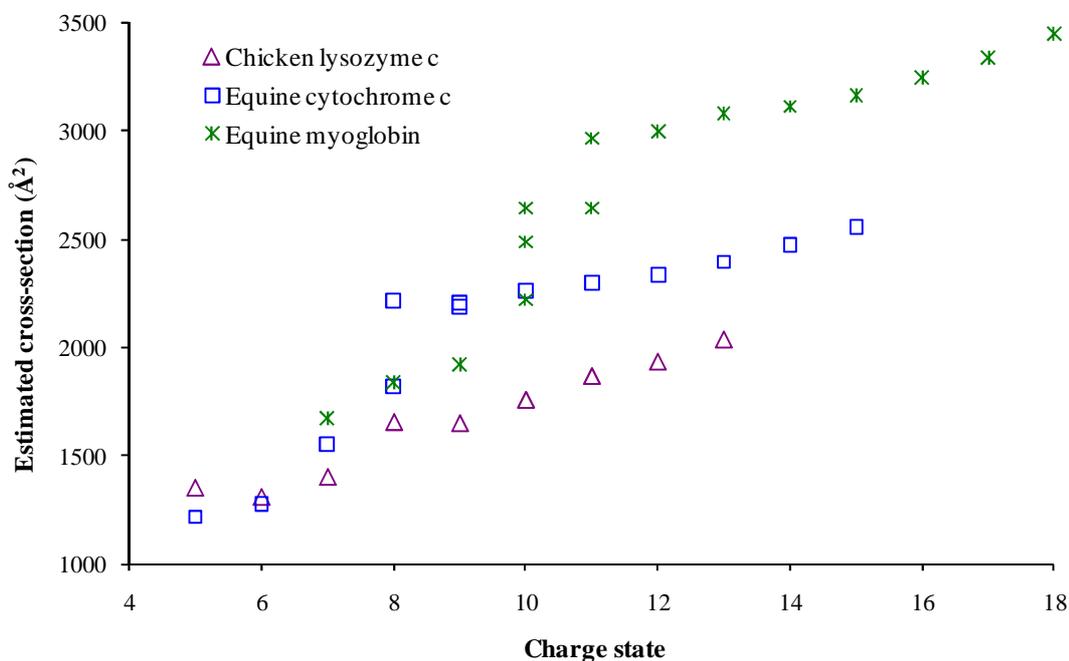
**Figure 2.8:** Experimentally estimated cross-sections for different charge states of sperm whale myoglobin under denaturing and near-physiological conditions with average theoretical cross-sections for PA and TM models shown. Error bars are shown to illustrate measurements that would be within  $\pm 2\%$ .

Ion mobility mass spectrometry is able to resolve differences in protein conformation under denaturing and near-native solvent conditions. Different conformations for the same charge state of lysozyme c are witnessed under the two solvent conditions investigated. The differences observed may be related to differences in solvent interactions. Conformations observed under near-physiological conditions change in cross-section less significantly with increase in charge than those observed under denaturing conditions.

### 2.3.4 Unfolding restraints

The degree of protein unfolding in the gas phase can be indicated by the increase in estimated cross-section with increase in charge state. The rise in estimated cross-section observed as charge state increases is shown for some of the proteins investigated, under denaturing conditions, in Figure 2.9. The unfolding transition is driven by Coulomb repulsion and is thought to be similar to acid denaturation in solution (Jarrold 1999). The change in cross-section can be affected, however, by the

presence of intact disulphide bonds (Hoaglund-Hyzer *et al.* 1999) and, possibly, by the association of ligands.



**Figure 2.9:** Experimentally estimated cross-sections for different charge states of chicken lysozyme c, equine cytochrome c and equine myoglobin.

The change in cross-section with increase in charge is greater for myoglobin than for cytochrome c. The heme group contained within the native structure of these proteins is covalently bound within cytochrome c but not within myoglobin. The heme group is lost when myoglobin is analysed under denaturing conditions but retained within cytochrome c. The presence of the heme group potentially stabilises more compact gas-phase conformations.

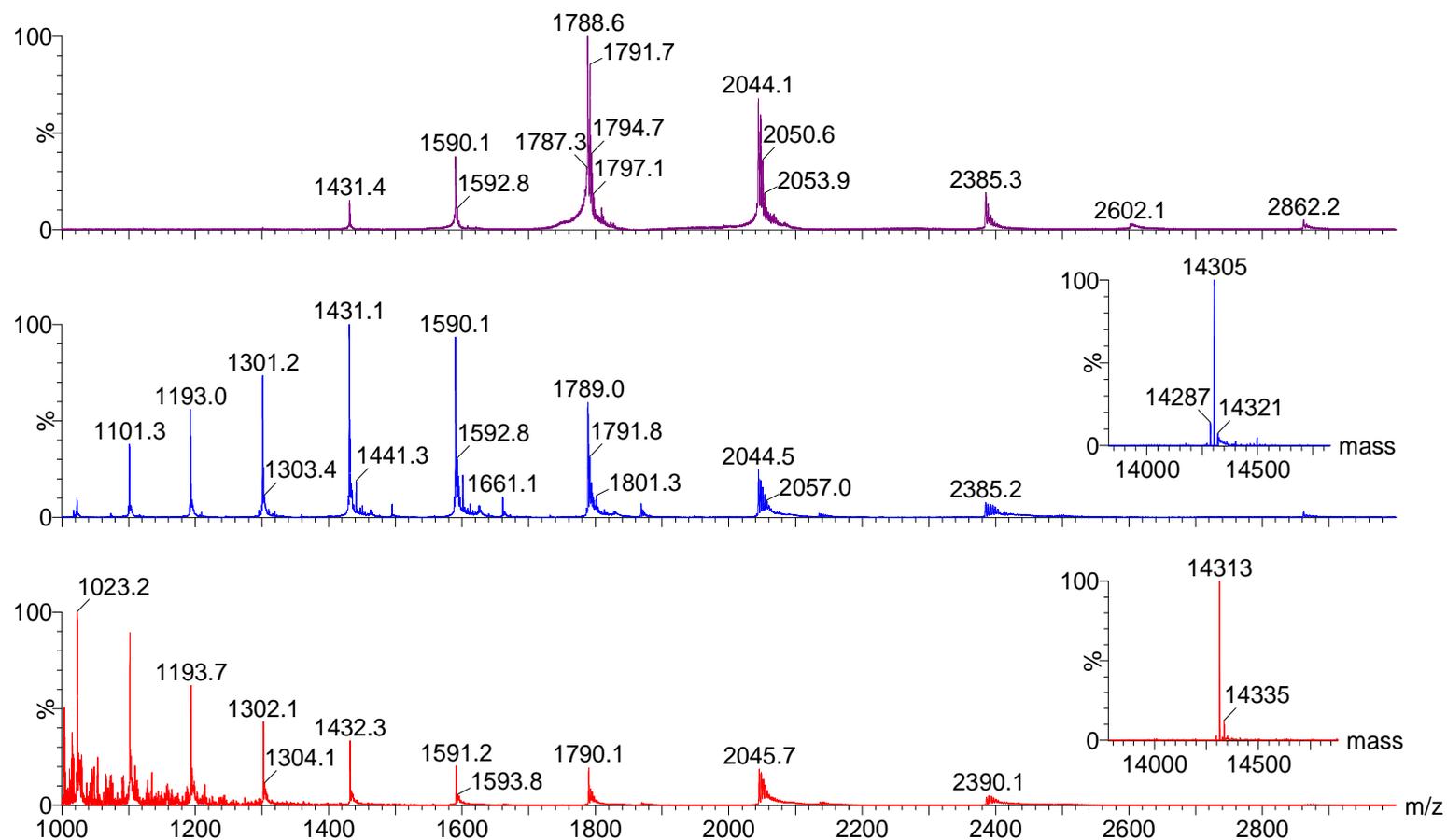
Differences are also seen between the unfolding patterns of myoglobin and cytochrome c. The unfolding of the three-dimensional structure of myoglobin with increase in charge state appears to progress more quickly than that of cytochrome c providing support for the conjecture that the heme group stabilises gas-phase protein conformation.

This is in agreement with early work by Elieser and Wright which suggested that the removal of heme from myoglobin highly destabilises the structure of the globin fold (Eliezer and Wright 1996) and supports previous DCIM-MS studies conducted by Shelimov and Jarrold (Shelimov and Jarrold 1997).

Under near physiological conditions, multiple conformations are observed for a number of charge states for myoglobin and cytochrome c but not for lysozyme c. This could suggest that the myoglobin and cytochrome c structures are less constrained in the gas phase and are able to adopt multiple stable conformations whereas the disulphide bond-constrained lysozyme c has less freedom of motion. The multiple conformations observed here are consistent with those that have been observed previously, in DCIM-MS studies, for cytochrome c equine (Jarrold 1999; Badman *et al.* 2001) and bovine (Clemmer *et al.* 1995; Shelimov *et al.* 1997) and equine apomyoglobin (Shelimov and Jarrold 1997).

The relative change in cross-section with increase in charge state for lysozyme c is smaller than for the other proteins studied, presumably because the presence of the disulphide bonds restricts how much the protein can unfold. To explore this concept lysozyme c was investigated further, under denaturing conditions, in which the disulphide-bonds were reduced by the addition of the reducing agent DTT.

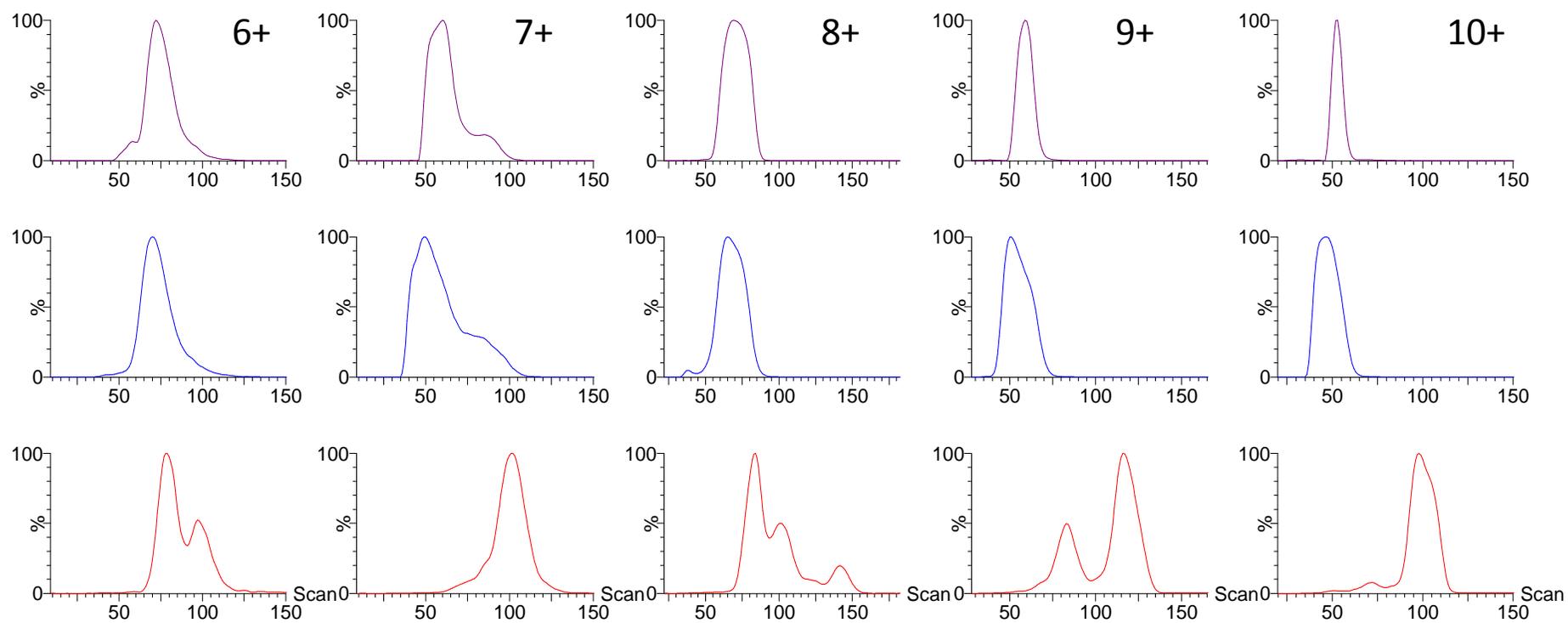
Figure 2.10 illustrates mass spectra obtained for chicken lysozyme c when analysed under oxidised near-physiological and denaturing conditions and under reduced denaturing conditions. Deconvoluted spectra obtained, by use of the MaxEnt processing algorithm, confirm that all disulphide bonds present have been reduced as an increase in mass of 8 Da is observed.



**Figure 2.10:** Chicken lysozyme c oxidised in 10 mM ammonium acetate (top spectrum), oxidised in 50 % ACN 0.2 % HCOOH (middle spectrum) and reduced in 50 % ACN 0.2 % HCOOH (bottom spectrum) with deconvoluted spectra inset.

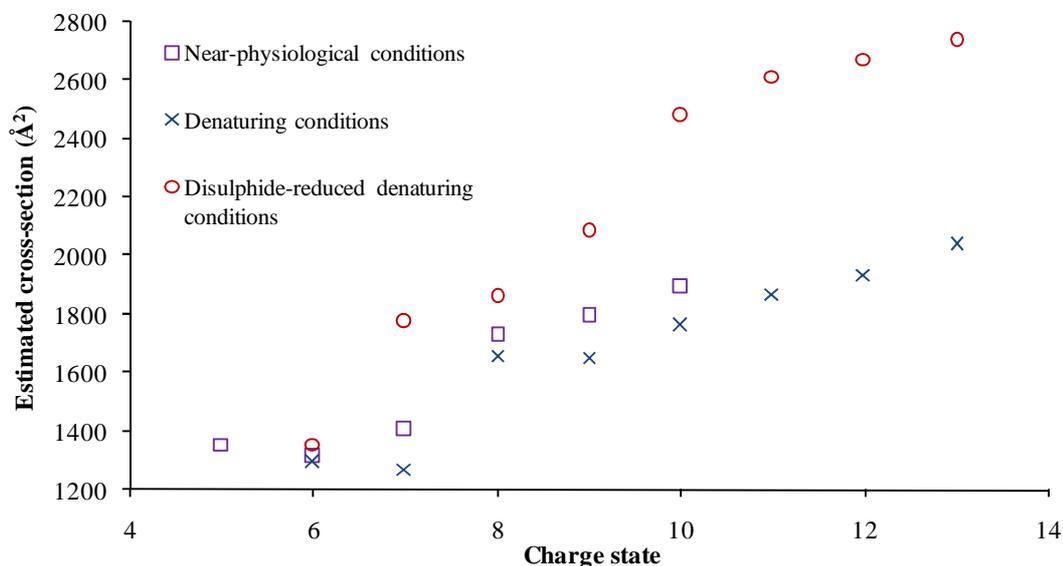
The shape of the charge state distributions observed in spectra obtained under oxidised conditions suggest that the protein predominantly occupies one conformation under these conditions. Under denaturing oxidised conditions the protein predominantly accepts two more charges than under near-physiological conditions. Under reduced conditions, however, the charge state distribution observed changes significantly suggesting that the protein has largely unfolded and now exists in more than one conformation.

This observation is supported by the data produced from TWIM-MS experiments (see Figure 2.11). Arrival time distributions for various charge states suggest that under oxidised conditions the protein is able to maintain a similar conformation whether analysed under non-denaturing or denaturing solvent conditions. Arrival time distributions are generally broader for all charge states under denaturing solvent conditions suggesting conformations are more flexible under these solvent conditions. When disulphide-bonds are reduced, however, multiple peaks in arrival time distributions are observed at higher arrival times; the protein adopts conformations which are less folded.



**Figure 2.11:** Arrival time distributions for charges states  $[M + 6H]^{6+}$  to  $[M + 10H]^{10+}$  for chicken lysozyme c oxidised in 10 mM ammonium acetate (top), oxidised in 50 % ACN 0.2 % HCOOH (middle) and reduced in 50 % ACN 0.2 % HCOOH (bottom).

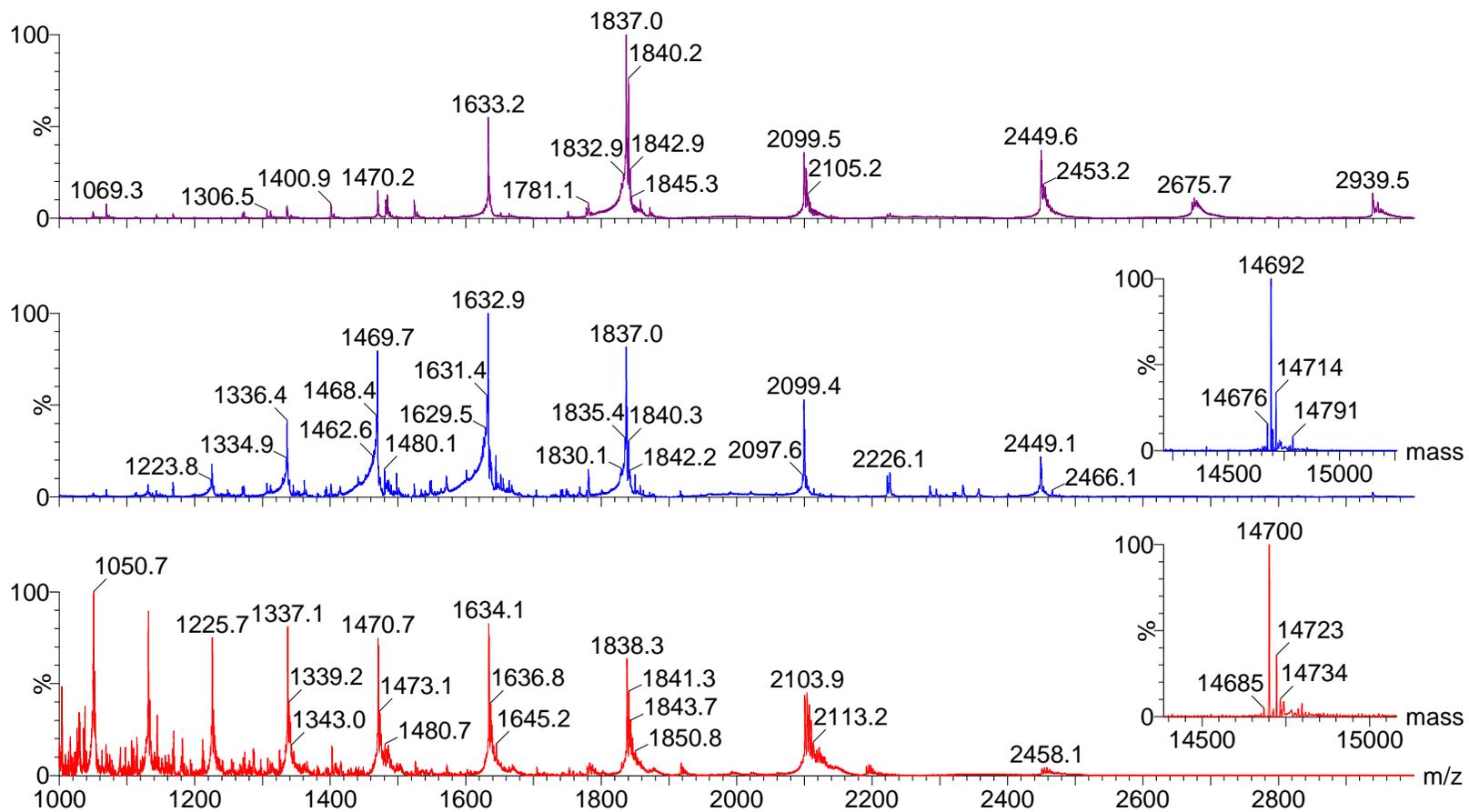
The estimated cross-sections for the protein conformations producing the dominant peaks in the arrival time distributions are illustrated below in Figure 2.12.



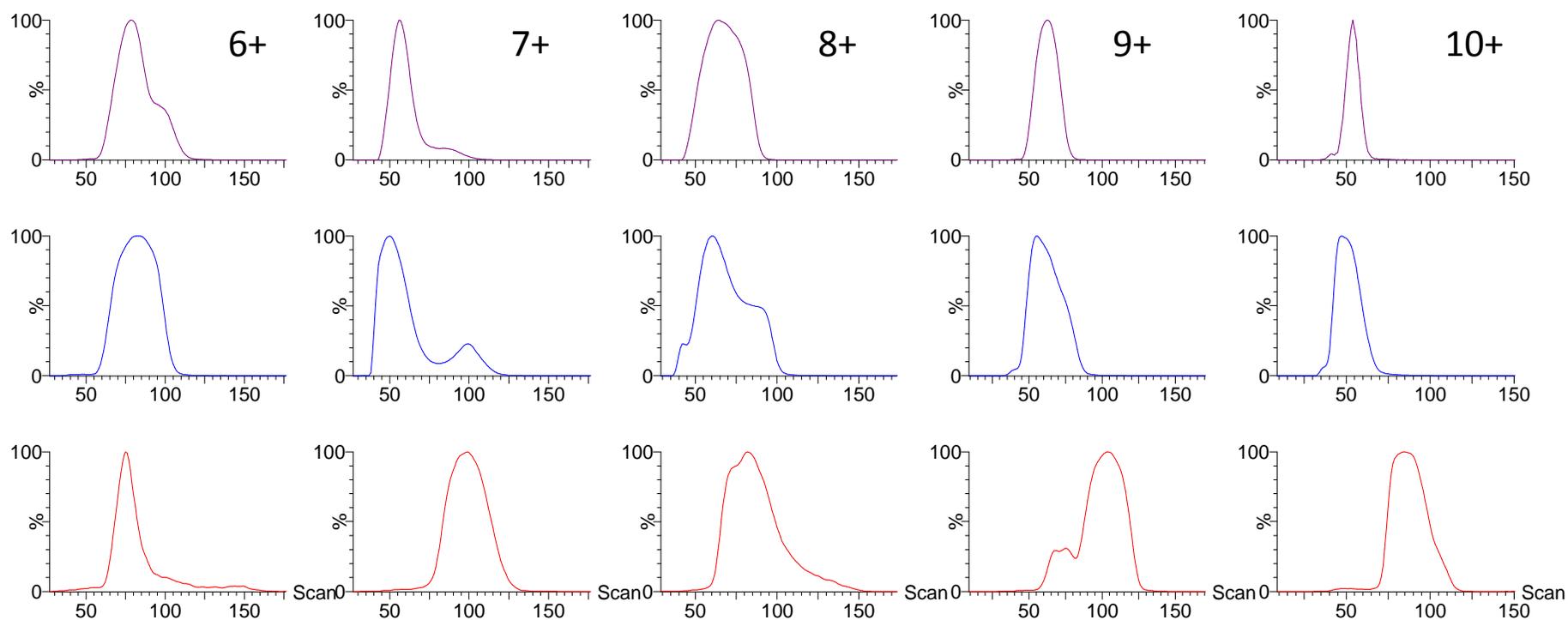
**Figure 2.12:** Experimentally estimated cross-sections for different charge states of chicken lysozyme c under near-physiological conditions, denaturing conditions and under disulphide-reduced denaturing conditions.

The cross-section estimations obtained for lysozyme c analysed under denaturing conditions are very similar to those obtained in earlier experiments, as would be expected (see Figure 2.5). The cross-sections estimated for charge state  $[M + 6H]^{6+}$  of the protein under oxidised and reduced conditions are also comparable. This is a surprising result which suggests that a very similar structure to the oxidised structure can be maintained when disulphide bonds are reduced.

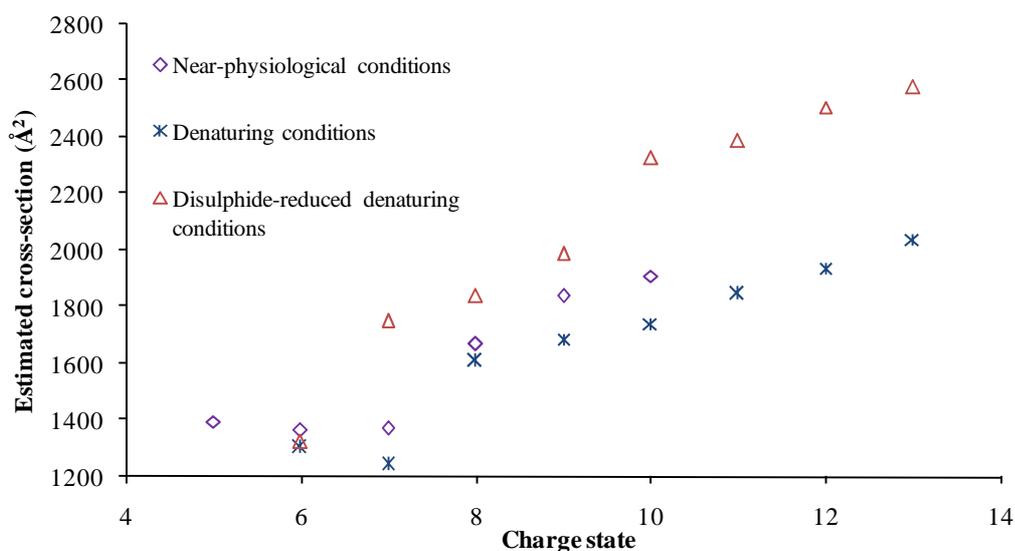
These results, observed for chicken lysozyme c under reduced and oxidised conditions, are also generally reflected for human lysozyme c (see Figures 2.13-2.15).



**Figure 2.13:** Human lysozyme c oxidised in 10 mM ammonium acetate (top spectrum), oxidised in 50 % ACN 0.2 % HCOOH (middle spectrum) and reduced in 50 % ACN 0.2 % HCOOH (bottom spectrum) with deconvoluted spectra inset.



**Figure 2.14:** Arrival time distributions for charges states  $[M + 6H]^{6+}$  to  $[M + 10H]^{10+}$  for human lysozyme c oxidised in 10 mM ammonium acetate (top), oxidised in 50 % ACN 0.2 % HCOOH (middle) and reduced in 50 % ACN 0.2 % HCOOH (bottom).



**Figure 2.15:** Experimentally estimated cross-sections for different charge states of human lysozyme c under near-physiological conditions, denaturing conditions and under disulphide-reduced denaturing conditions.

Valentine *et. al.* (1997) have previously investigated the conformations of disulphide-intact and reduced turkey lysozyme in the gas phase by means of DCIM-MS. The results obtained in this study are in agreement with their findings. They found that reduced turkey lysozyme produced gas-phase ions with distinctly different conformations from those of oxidised lysozyme. They reduced turkey lysozyme by boiling the sample in DTT. They did not see the lower charge states observed here for the reduced lysozyme without adding gas-phase charge stripping reagents. It may be suggested that boiling the sample resulted in a greater unfolding of the lysozyme and thus more charge acceptance than would have been observed by simply reducing the sample. It is interesting to note that the  $[M + 6H]^{6+}$  charge state formed when the reduced sample was subjected to gas-phase charge stripping reagents had a similar conformation to that observed for the  $[M + 6H]^{6+}$  charge state from the oxidised sample.

The unfolding pattern observed for chicken and human lysozyme c of increase in cross-section with increase in charge state, under disulphide-reduced conditions, now reflects much more closely that observed for the other proteins studied. Table 2.2

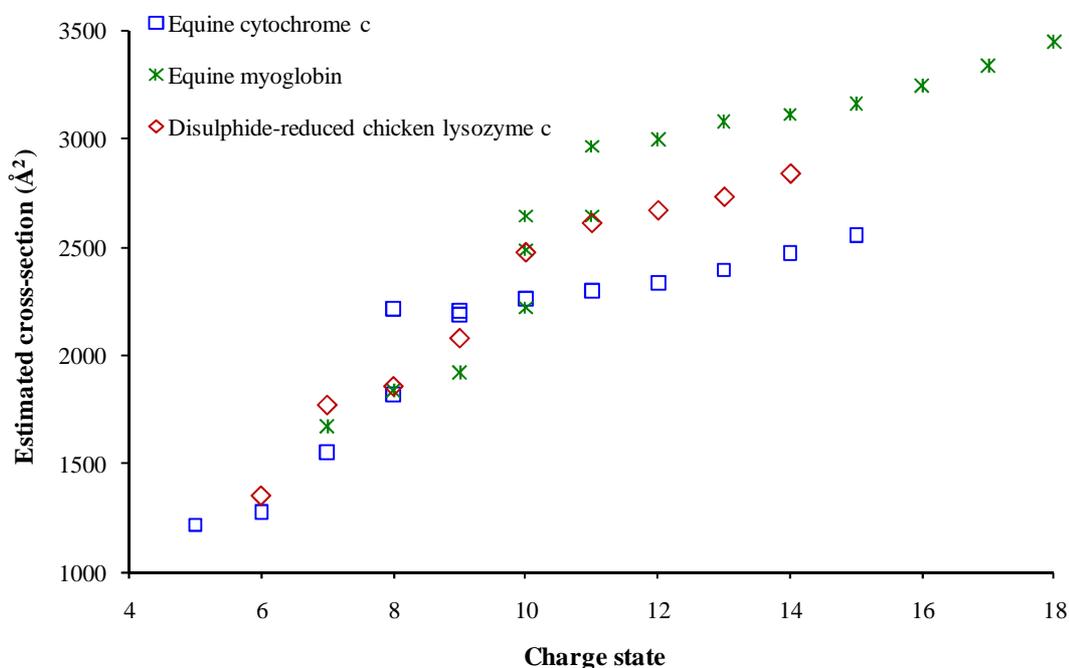
illustrates the smallest and largest estimated cross-sections obtained for each of the proteins studied reflecting the lowest charge state observed under non-denaturing conditions and the highest charge state observed under denaturing conditions respectively.

**Table 2.2:** Estimated cross-sections (smallest and largest) for proteins studied, and differences between these cross-sections observed.

Protein	MW (Da)	Estimated cross-section ( $\text{\AA}^2$ )		Change in cross-section	Percentage increase in cross-section (smallest to largest)
		Smallest	Largest		
Sperm whale myoglobin	17331	1671	3508	1836	109.9
Equine myoglobin	16952	1671	3446	1774	106.2
Equine cytochrome c	12359	1219	2559	1340	109.9
Chicken lysozyme c oxidised	14305	1314	2039	726	55.2
Chicken lysozyme c reduced	14313	1357	2736	1379	101.6
Human lysozyme c oxidised	14692	1363	2039	676	49.6
Human lysozyme c reduced	14700	1325	2578	1253	94.6

The percentage increase in cross-section for myoglobin and cytochrome c, from the smallest to the largest estimated cross-section, is approximately 105-110 % whilst this is only approximately 50-55 % for the oxidised lysozymes. Under reduced conditions, however, the lysozymes also show a percentage increase in cross-section close to 100 %. This demonstrates the restraints on protein unfolding attributable to the presence of the disulphide bonds. Once the disulphide bonds are reduced the change in cross-section observed with change in charge state is similar for all proteins studied.

If Figure 2.9 is reproduced replacing experimentally estimated cross-sections for different charge states of disulphide-intact chicken lysozyme c with disulphide-reduced chicken lysozyme c, a clear pattern emerges (see Figure 2.16).

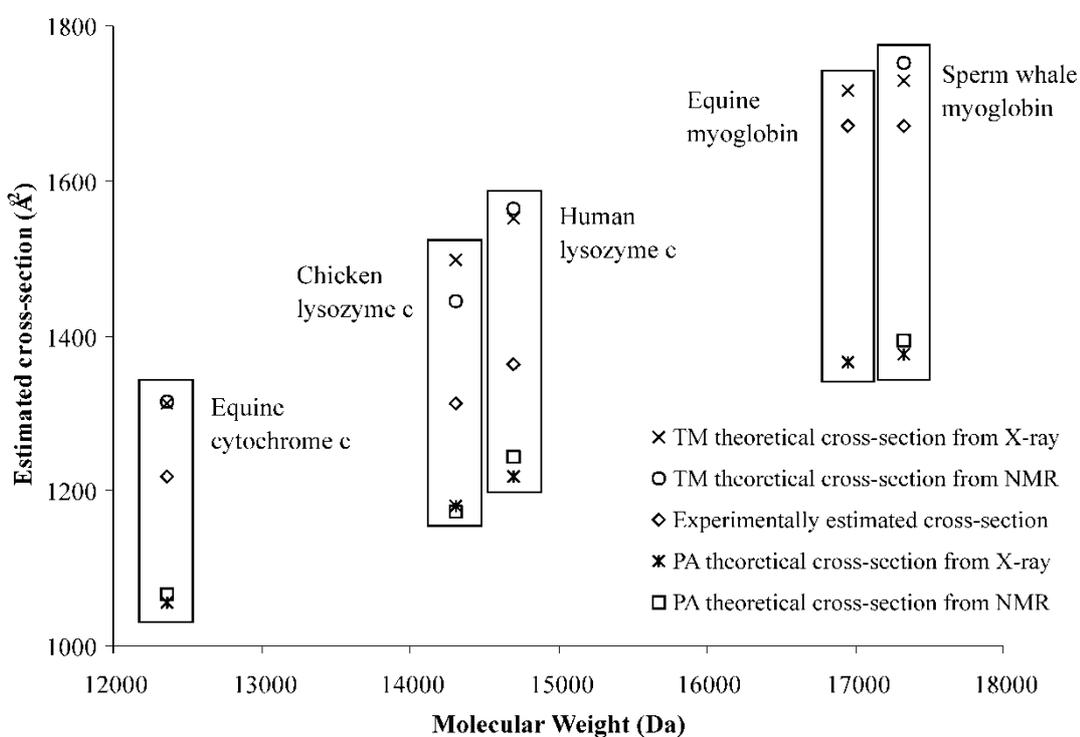


**Figure 2.16:** Experimentally estimated cross-sections for different charge states of disulphide-reduced chicken lysozyme c, equine cytochrome c and equine myoglobin.

The cross-sections estimated for the higher charge states (11 and above), reflecting largely unfolded conformations, of each of the proteins, increase with increase in charge state at a similar rate. The smallest estimated cross-sections now represent the lowest molecular weight protein and the largest the highest. In general, a protein with a larger molecular weight would be expected to have a larger collision cross-section (Ruotolo *et al.* 2005).

### 2.3.5 Biological significance

For each protein studied in this work the lowest charge state has been observed to have an estimated cross-section between the theoretical approximations (PA and TM) (Figure 2.17). The smallest estimated cross-sections for the two myoglobins lie closer to the TM approximations whilst the smallest estimated cross-sections for the lysozyme c proteins sit firmly between the PA and TM calculations.



**Figure 2.17:** Estimations of native cross-sections for proteins studied calculated theoretically and experimentally and plotted against molecular weight.

Cross-sections estimated for the lowest charge states observed under physiological conditions are in good agreement with theoretical measurements calculated from published X-ray and NMR structures. Multiply charged ions produced by ESI are primarily thought to be the result of proton attachment to exposed basic sites of the protein. These results (Figure 2.17) support the hypothesis that the lowest charge states observed (under near-native conditions) are most representative of the native protein structure due to the fact that a tightly folded conformation has fewer exposed

basic sites than an unfolded conformation of the same protein (Chowdhury *et al.* 1990).

Estimated cross-sections for the lowest charge states are smaller than the TM approximation values. This is in agreement with previous work which has concluded that gas phase conformations can be more compact than the crystal structure (Hoaglund-Hyzer *et al.* 1999). In the gas phase, it is proposed that intramolecular interactions become more dominant and make the polar side chains collapse on to the protein surface. This results in the protein adopting a more compact conformation (Shelimov *et al.* 1997).

The experimental time period is crucial to the maintenance of solution-phase structure as after an extended period it is known that gas-phase structures can rearrange to account for their change in environment. Badman *et al.* have shown evidence for unfolding and refolding of cytochrome c ions in the gas phase with transition from compact to unfolded structures occurring within 30-60ms, and unfolded structures refolding into an array of folded structures after ions are trapped for times exceeding 250 ms (Badman *et al.* 2005). The TWIM-MS experiment operates within a 15-30 ms time frame. This limits the opportunity for the solution-phase structures to rearrange within the gas phase.

## 2.4 Conclusions

This work illustrates that TWIM-MS can be used to investigate the three-dimensional structure of a protein and the stability of that structure within the gas phase.

Cross-sections estimated experimentally for proteins studied, for charge states most indicative of native structure, are in good agreement with measurements calculated from published X-ray and NMR structures. These results illustrate that the TWIM-MS approach can provide data on three-dimensional protein structure of biological relevance.

The results presented here are also in general agreement with data presented from DCIM-MS studies (Shelimov *et al.* 1997; Shelimov and Jarrold 1997; Valentine *et al.* 1997; Jarrold 1999; Badman *et al.* 2001). In general, it has been observed that the estimated cross-sections for protein charge states representative of denatured conformations favour a narrow distribution; a single highly-folded conformation is observed at low charge states and intermediate charge-states exist in multiple conformations with different cross-sections (Valentine *et al.* 1997). This is reflected within this work. At low charge states, the protein remains highly-folded, strengthened by increased electrostatic intramolecular interactions (Breuker and McLafferty 2008) whereas at high charge states increased Coulomb repulsion between charges forces the protein to adopt more open conformations (Li *et al.* 1999). Various intermediate conformations are formed due to the interplay between attractive-folding and repulsive-Coulombic interactions (Valentine *et al.* 1997).

Recently, studies were conducted by Smith *et al.* on deciphering drift time measurements from TWIM-MS studies (Smith *et al.* 2009). They investigated chicken lysozyme c and equine cytochrome c by TWIM-MS and the estimated cross-sections, for the lowest charge state observed, that they reported, were within 2 % of our estimated values. This highlights the robustness and reproducibility of the TWIM-MS approach.

The cross-section of a protein calculated from IM-MS measurements varies with charge state. This allows the unfolding of proteins in the gas phase, with increase in charge, to be studied. This work, along with that previously conducted by Valentine *et al.* (1997) by DCIM-MS studies, illustrates that the unfolding transition of lysozyme c is restrained by the presence of disulphide-bonds. The unfolding transition is similar for each protein when the restraints of disulphide-bonds are removed. At high charge states, in particular, there is a similar relative increase in cross-section with increase in charge state for each protein studied. As described above, at these higher charge states increase in cross-section with increase in charge is largely due to the corresponding increase in Coulomb repulsion.

The T-Wave device has advantages over many homemade drift-cell devices in that it is able to operate at biologically-relevant sample concentrations and its ease of use facilitates its operation in an automated fashion. The mobility resolution of the Synapt HDMS system, however, is lower than in many drift-cell devices. When this work was conducted the TWIM-MS approach was in the early stages of its development. Since then further improvements in resolution of the technique have been made and a new commercial instrument the Synapt HDMS G2 (Waters, Manchester, UK) has been launched. This will likely extend the application range of TWIM-MS.

## 2.5 References

**Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J.** (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*. **25**, 3389-3402.

**Badman, E. R., Hoaglund-Hyzer, C. S. and Clemmer, D. E.** (2001). Monitoring Structural Changes of Proteins in an Ion Trap over ~10-200 ms: Unfolding Transitions in Cytochrome c Ions. *Analytical Chemistry*. **73**, 6000-6007.

**Badman, E. R., Myung, S. and Clemmer, D. E.** (2005). Evidence for Unfolding and Refolding of Gas-Phase Cytochrome c Ions in a Paul Trap. *Journal of the American Society for Mass Spectrometry*. **16**, 1493-1497.

**Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E.** (2000). The Protein Data Bank. *Nucleic Acids Research*. **28**, 235-242.

**Breuker, K. and McLafferty, F. W.** (2008). Stepwise evolution of protein native structure with electrospray into the gas phase,  $10^{-12}$  to  $10^2$  s. *Proceedings of the National Academy of Sciences*. **105**, 18145-18152.

**Chowdhury, S. K., Katta, V. and Chait, B. T.** (1990). Probing conformational changes in proteins by mass spectrometry. *Journal of the American Chemical Society*. **112**, 9012-9013.

**Clemmer, D. E.**

[http://www.indiana.edu/~clemmer/Research/cross%20section%20database/Proteins/protein\\_cs.htm](http://www.indiana.edu/~clemmer/Research/cross%20section%20database/Proteins/protein_cs.htm). Accessed on: 30th April 2008.

**Clemmer, D. E., Hudgins, R. R. and Jarrold, M. F.** (1995). Naked Protein Conformations: Cytochrome c in the Gas Phase. *Journal of the American Chemical Society*. **117**, 10141-10142.

**Eliezer, D. and Wright, P. E.** (1996). Is Apomyoglobin a Molten Globule? Structural Characterization by NMR. *Journal of Molecular Biology*. **263**, 531-538.

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

**Hoaglund-Hyzer, C. S., Counterman, A. E. and Clemmer, D. E.** (1999). Anhydrous Protein Ions. *Chemical Reviews*. **99**, 3037-3080.

**Hopper, J. T. S. and Oldham, N. J.** (2009). Collision Induced Unfolding of Protein Ions in the Gas Phase Studied by Ion Mobility-Mass Spectrometry: The Effect of Ligand Binding on Conformational Stability. *Journal of the American Society for Mass Spectrometry*. **20**, 1851-1858.

**Jarrold, M. F.** (1999). Unfolding, Refolding, and Hydration of Proteins in the Gas Phase. *Accounts of Chemical Research*. **32**, 360-367.

**Li, J., Taraszka, J. A., Counterman, A. E. and Clemmer, D. E.** (1999). Influence of solvent composition and capillary temperature on the conformations of electrosprayed ions: unfolding of compact ubiquitin conformers from pseudonative and denatured solutions. *International Journal of Mass Spectrometry*. **185-187**, 37-47.

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

**Myung, S., Badman, E. R., Lee, Y. J. and Clemmer, D. E.** (2002). Structural Transitions of Electrosprayed Ubiquitin Ions Stored in an Ion Trap over ~10 ms to 30 s. *Journal of Physical Chemistry A*. **106**, 9976-9982.

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

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

**Ruotolo, B. T., Giles, K., Campuzano, I., Sandercock, A. M., Bateman, R. H. and Robinson, C. V.** (2005). Evidence for Macromolecular Protein Rings in the Absence of Bulk Water. *Science*. **310**, 1658-1661.

**Scrivens, J. H., Thalassinos, K., Hilton, G., Slade, S. E., Pinheiro, T. J. T., Bateman, R. H. and Bowers, M. T.** (2006). Use of a Travelling Wave-Based Ion Mobility Approach to Resolve Proteins of Varying Conformation. *Proc. 55th ASMS Conf. Mass Spectrometry and Allied Topics*, Indianapolis.

**Shelimov, K. B., Clemmer, D. E., Hudgins, R. R. and Jarrold, M. F.** (1997). Protein Structure in Vacuo: Gas-Phase Conformations of BPTI and Cytochrome c. *Journal of the American Chemical Society*. **119**, 2240-2248.

**Shelimov, K. B. and Jarrold, M. F.** (1997). Conformations, Unfolding, and Refolding of Apomyoglobin in Vacuum: An Activation Barrier for Gas-Phase Protein Folding. *Journal of the American Chemical Society*. **119**, 2987-2994.

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

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

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

**Thalassinos, K., Grabenauer, M., Slade, S. E., Hilton, G. R., Bowers, M. T. and Scrivens, J. H.** (2008). Characterization of Phosphorylated Peptides Using Traveling Wave-Based and Drift Cell Ion Mobility Mass Spectrometry. *Analytical Chemistry*. **81**, 248-254.

**Valentine, S. J., Anderson, J. G., Ellington, A. D. and Clemmer, D. E.** (1997). Disulfide-Intact and -Reduced Lysozyme in the Gas Phase: Conformations and Pathways of Folding and Unfolding. *Journal of Physical Chemistry B*. **101**, 3891-3900.

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

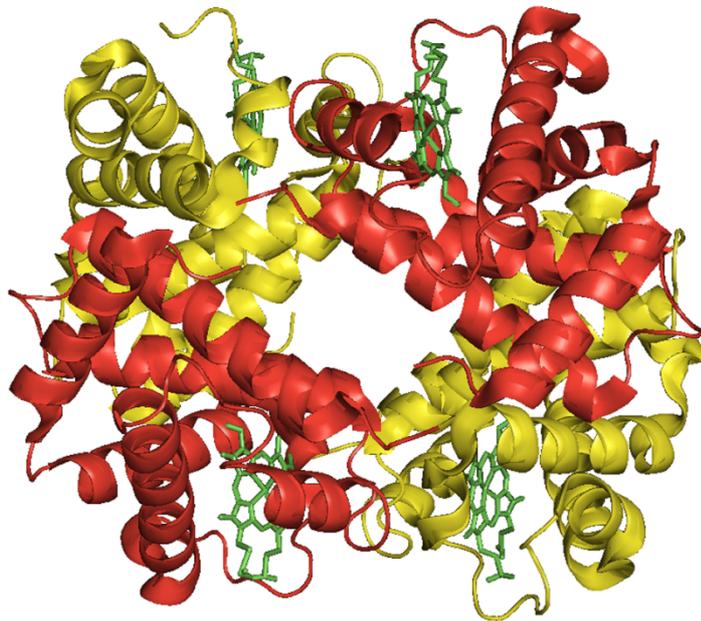
**Williams, J. P. and Scrivens, J. H.** (2008). Coupling desorption electrospray ionisation and neutral desorption/extractive electrospray ionisation with a travelling-wave based ion mobility mass spectrometer for the analysis of drugs. *Rapid Communications in Mass Spectrometry*. **22**, 187-196.

## Chapter 3: Probing Hemoglobin Structure

---

### 3.1 Introduction

Hemoglobin (Hb) is a tetramer consisting of four globin chains, two  $\alpha$ - and two  $\beta$ -, each associated with a heme group (see Figure 3.1). It is the major oxygen-transport protein found in the red blood cells of all vertebrates. Disorders of hemoglobin are the most common of all inherited disorders and consequently the molecule has been extensively studied.



**Figure 3.1:** Representation of hemoglobin tetramer structure.  $\alpha$ -chains are displayed in red and  $\beta$ -chains in yellow. Heme groups are shown in green.

The most debilitating Hb variant is that which causes sickle-cell anaemia. This disease occurs when a person inherits two particular mutated copies of the  $\beta$ -globin gene. The sickle-cell mutation results in the production of a  $\beta$ -chain with a single amino acid substitution ( $\beta 6$  Glu  $\rightarrow$  Val) and changes the conformation of the assembled tetramer to allow molecular stacking. Polymerisation of this sickle-cell hemoglobin molecule (Hb SS), in deoxygenated blood, causes the characteristic alteration in shape of red-blood cells from biconcave discs to crescentic (Murayama 1967).

ESI-MS has been widely used to detect Hb variants in hemoglobin (Shackleton *et al.* 1991; Wild *et al.* 2001; Thalassinou *et al.* 2004; Daniel *et al.* 2005; Shimizu *et al.* 2006). These approaches are discussed in Chapter 5. ESI-MS has also been used to investigate the structural assembly of hemoglobin into a non-covalent complex (Ofori-Acquah *et al.* 2001; Griffith and Kaltashov 2003; Boys and Konermann 2007; Griffith and Kaltashov 2007) and its corresponding disassembly (Versluis and Heck 2001).

Whilst there is extensive knowledge, on the atomic level, of hemoglobin quaternary structure, the exact assembly pathway for the hemoglobin tetramer is still a matter of debate. It is known that one  $\alpha$ - and one  $\beta$ -monomer come together to form a heterodimer and that two of these dimers associate to form the tetramer.  $\alpha$ - and  $\beta$ -monomers can exist in heme-free (apo,  $\alpha^a$  and  $\beta^a$ ) and heme-bound (holo,  $\alpha^h$  and  $\beta^h$ ) forms (Boys and Konermann 2007) and it is unclear as to whether the heme groups are attached to both  $\alpha$ - and  $\beta$ -monomers prior to dimer formation, or if association leads to heme recruitment.

Griffith and Kaltashov have suggested that the formation of a heme-deficient dimer intermediate ( $\alpha^h\beta^a$ ) occurs, consisting of a natively folded holo- $\alpha$ -globin ( $\alpha^h$ ) and a partially unfolded apo- $\beta$ -globin ( $\beta^a$ ), prior to complete dimer formation, to ensure correct tetramer structure arrangement (Griffith and Kaltashov 2003; Griffith and Kaltashov 2007). The Konermann group have reported, however, that the heme-deficient dimer is not observed when using freshly-prepared samples, in contrast to it being seen consistently when commercially-available samples, in the form of lyophilised powder, are used (Boys *et al.* 2007). They studied the acid-induced denaturation of bovine Hb and concluded that it followed a highly symmetric mechanism:  $(\alpha^h\beta^h)_2 \rightarrow 2 \alpha^h\beta^h \rightarrow 2\alpha^h_{\text{folded}} + 2\beta^h_{\text{folded}} \rightarrow 2\alpha^a_{\text{unfolded}} + 2\beta^a_{\text{unfolded}} + 4$  heme.

This work attempts to elucidate further the structural properties of the hemoglobin tetramer and its components and to determine whether conformational differences between the Hb A and Hb SS molecules can be observed by TWIM-MS.

## 3.2 Materials and Methods

### 3.2.1 Samples and sample preparation

Samples of fresh whole blood were supplied by University Hospitals Coventry and Warwickshire (UHCW) NHS Trust. Commercially-prepared hemoglobin was purchased from Sigma Aldrich Ltd (Gillingham, UK) and prepared at a concentration of 20 mg/mL in 10 mM ammonium acetate pH 6.8. Sample preparation for mass spectral analysis was adapted from that detailed by Ofori-Acquah *et al.* (Ofori-Acquah *et al.* 2001). Samples (20  $\mu$ l) were diluted 10-fold in 10 mM ammonium acetate pH 6.8 and spun at 3000 *g* for 15 minutes in centrifugal filter units with a 10 kDa cut-off (Microcon® YM-10, Millipore Corporation, Billerica, MA, USA). Sample retained on the filter was diluted a further 20-fold with 10 mM ammonium acetate and desalted by agitating for two ten-minute periods with approximately 5 mg of ion-exchange mixed bed resin (AG 501-X8, Bio-Rad Laboratories, Hercules, CA, USA) that had been prepared for use by rinsing twice in liquid chromatography MS grade water.

### 3.2.2 Ion mobility mass spectrometry

TWIM-MS experiments were conducted on a Synapt HDMS system, as described in Chapter 1 Section 1.1.7 and Chapter 2 Section 2.2.3. Instrument acquisition parameters were optimised, as described in Chapter 2, for hemoglobin tetramer analysis. Samples were introduced into the source region of the Synapt HDMS system by direct infusion nano-ESI by means of fused silica nanospray needles (Waters Corporation, Milford, MA, USA). The instrument was operated in ESI positive mode with a capillary voltage of 1.2 kV, cone voltage of 60 V and source temperature of 110 °C for all experiments. The TOF mass analyser was tuned in V-mode to give an operating resolution of 8,000 (FWHM) and was calibrated using 2 mg/mL cesium iodide in 50 % aqueous propan-2-ol. A mass acquisition range of 1000-4500 *m/z* was used. The ion mobility cell was operated at an indicated pressure of 0.68 mbar, corresponding to a flow rate of 38 mL/min.

Calibration of data, protein cross-section estimation and theoretical cross-section calculations were performed as detailed in Chapter 2 Sections 2.2.4 and 2.2.5. Equine myoglobin was used as the T-Wave calibrant. Data obtained for each hemoglobin tetramer over the  $m/z$  range 3000-4500 were deconvoluted on to a true mass scale using the MaxEnt processing algorithm, within MassLynx™ (v4.1) software, to provide an estimate of molecular mass. Experiments were carried out in triplicate.

## **3.3 Results and Discussion**

### **3.3.1 Instrument acquisition parameters**

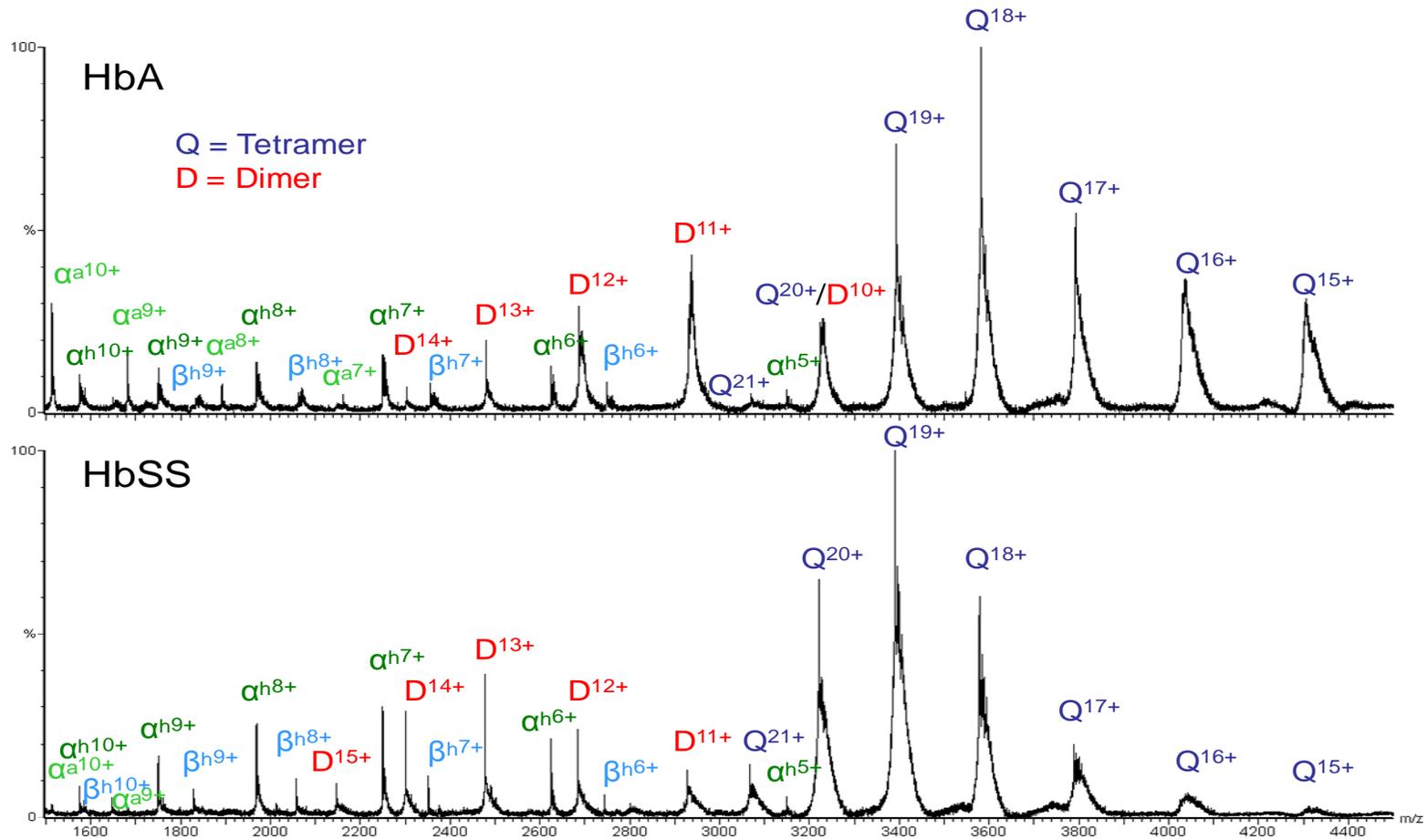
Considerable optimisation of instrument acquisition parameters is required for each individual application of ion mobility separation. These must be tailored to the sample of interest as optimal conditions are dependent on ionic species and mass-to-charge ratio (Tahallah *et al.* 2001). Controlled optimisation of instrument acquisition parameters indicated that a backing pressure of between 6.6 and 6.8 mbar was ideal for intact hemoglobin tetramer analysis. The optimal ion mobility separation of the tetramer was achieved using a travelling wave velocity and wave height of 400 m/s and 18 V respectively.

### **3.3.2 Calibration**

A calibration curve was used to allow the estimation of cross-sections for different constituents of hemoglobin in different charge states. Cross-sections calculated for equine myoglobin were within two percent of absolute values obtained by DCIM-MS experiments. These results were reproducible across the three datasets acquired.

### **3.3.3 ESI-MS spectra**

Representative spectra for normal (Hb A) and sickle (Hb SS) hemoglobin analysed by means of ESI-TOF-MS under non-denaturing conditions are shown in Figure 3.2. The data were deconvoluted to give masses of 64,454.7 Da for Hb A and 64,395.8 Da for Hb SS which were very close to the theoretical masses of 64,453.2 Da and 64,393.4 Da respectively (Ofori-Acquah *et al.* 2001).



**Figure 3.2:** Mass spectra of normal (Hb A) and sickle (Hb SS) hemoglobin analyzed by ESI-TOF-MS under near-physiological conditions. Spectra are labelled with charge states of tetramer (Q), heterodimer (D), and apo- and holo-monomers (superscripts “a” and “h” refer to apo- and holo-forms, respectively).

The hemoglobin spectra obtained show the presence of the tetramer ( $((\alpha^h\beta^h)_2)$ ), heterodimer ( $\alpha^h\beta^h$ ), and apo- and holo-monomer species. The trimer is not seen, as would be expected, as the accepted mechanism of formation of the hemoglobin tetramer involves the non-covalent association of two  $\alpha^h\beta^h$ -dimers. Carefully controlled near-physiological conditions were used in preparing the sample and the absence of any trimer supports the assertion that the species observed within the spectra represent those that exist naturally in solution. This is consistent with results from isotope labelling studies which showed that non-tetrameric ions in the spectrum corresponded to species present in solution (Hossain and Konermann 2006) rather than products of fragmentation formed during the ESI process (Kuprowski *et al.* 2007).

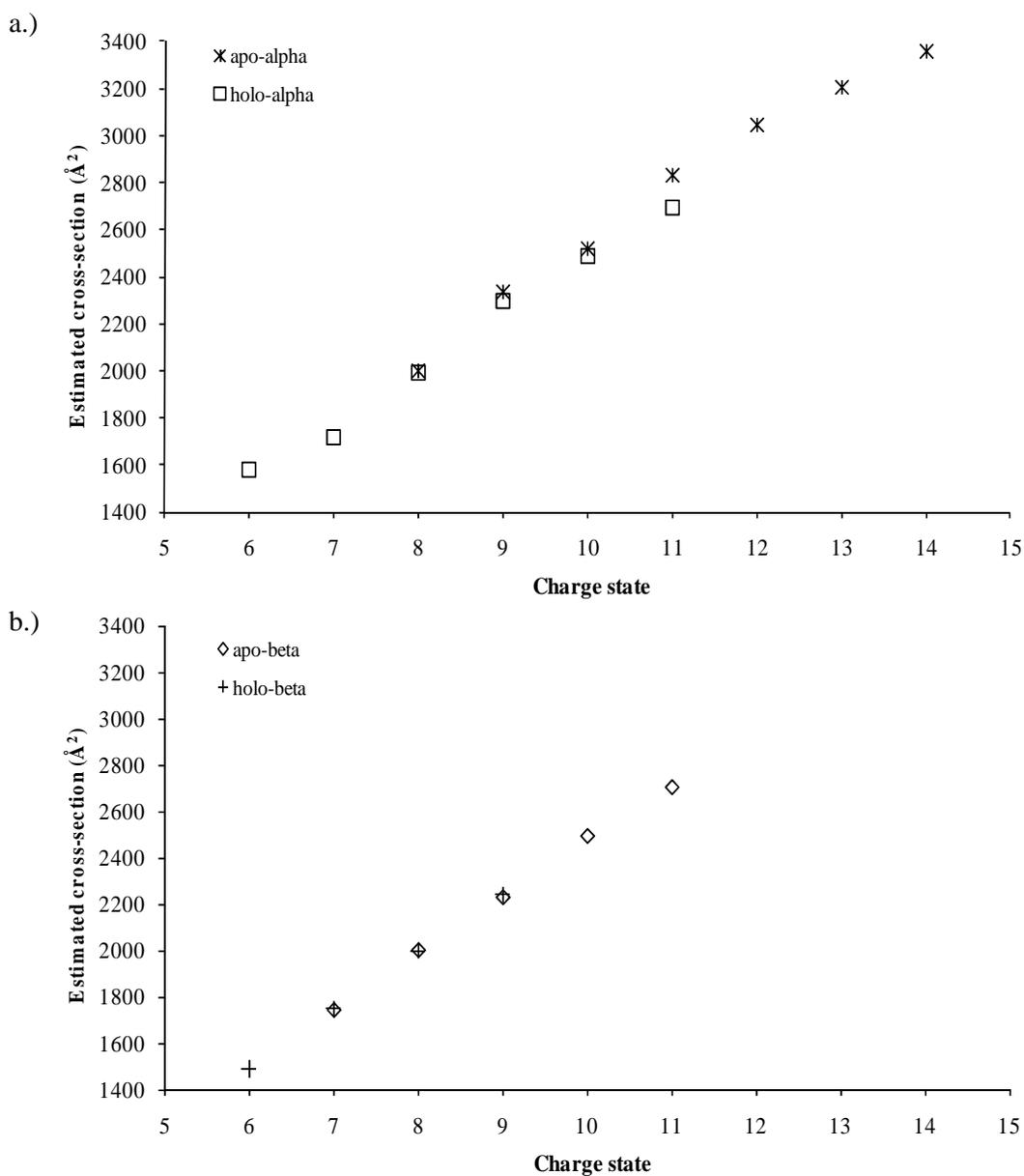
The most intense charge state for the Hb A tetramer observed within spectra was  $[M + 18H]^{18+}$  whilst that for the Hb SS tetramer was  $[M + 19H]^{19+}$ . The substitution of a glutamic residue for a valine residue results in the  $\beta$ -chain carrying one less negative charge. As two  $\beta$ -chains are present within each tetramer the Hb A molecule may have been expected to routinely accept two fewer charges than the Hb SS tetramer at the pH at which this study was carried out.

### 3.3.4 Alpha and beta monomers

Alpha and beta monomers are observed within the Hb A spectrum in both apo- and holo- forms. In a previous study, Griffith and Kaltashov suggested that an  $\alpha^h$  monomer first becomes associated with an  $\beta^a$  monomer to enable the beta-chain to incorporate the heme group (Griffith and Kaltashov 2003). This observation was based upon the absence of  $\beta^h$  in the spectrum. A subsequent study by Boys and Konermann detected very small quantities of heme-deficient dimer and found that both  $\alpha$ - and  $\beta$ -monomers were capable of binding heme. The discrepancies observed are thought to be attributable to differences between the commercially prepared and freshly obtained samples used (Boys *et al.* 2007). In the work reported here, in which fresh blood samples were used,  $\beta^h$  was observed in multiply-charged states, which is consistent with the above suggestion.

### 3.3.5 Apo- or holo-?

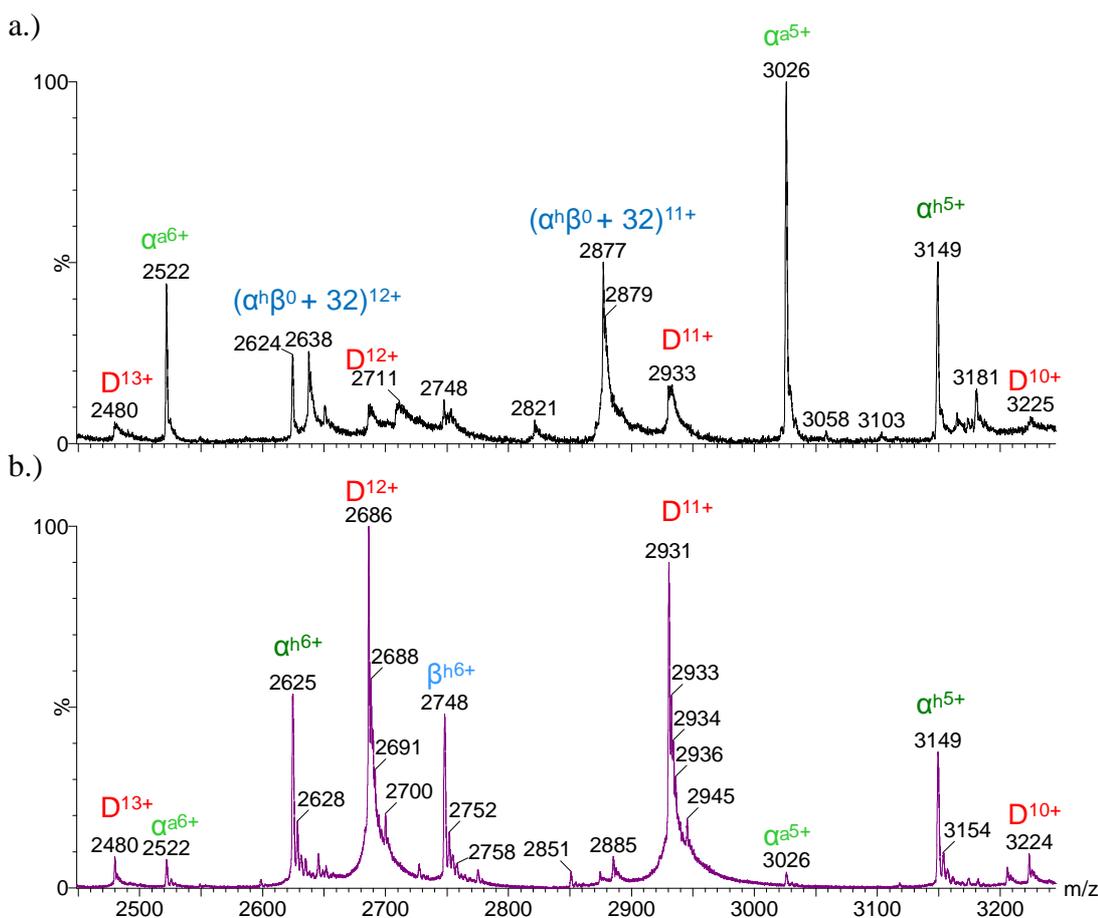
It has been reported that without the attachment of the heme group,  $\alpha$ - and  $\beta$ -monomers adopt extensively unfolded conformations (Leutzinger and Beychok 1981). Cross-sections for various charge states of  $\alpha$ - and  $\beta$ -monomers in both the apo- and holo- forms have been estimated and our observations suggest that the predominant conformations of  $\alpha$ - and  $\beta$ -monomers in the gas phase are similar to each other and show little change in the absence or presence of heme (Figure 3.3).



**Figure 3.3:** Average estimated cross-sections for charge states of a.) apo- $\alpha$  and holo- $\alpha$  monomers and b.) apo- $\beta$  and holo- $\beta$  monomers observed within three datasets.

### 3.3.6 Heme-deficient dimer

The heme-deficient dimer observed in previous studies is not observed here. The existence of both apo- and holo- forms of  $\alpha$ - and  $\beta$ -monomers, all of similar cross-sections, does not support the need for a  $\beta^a$  to associate with  $\alpha^h$  in order for the  $\beta$ -monomer to recruit heme. Analysis of commercially sourced human hemoglobin produced spectra containing additional peaks to those observed in spectra for fresh samples (see Figure 3.4 below). These peaks had mass-to-charge ratios consistent with those that would be produced by the presence of heme-deficient dimer with the addition of approximately 32 Da. This is in agreement with previous work conducted by the Konermann group on bovine hemoglobin. They attributed the mass shift to the occurrence of oxidative modifications in the commercial protein (Boys *et al.* 2007).

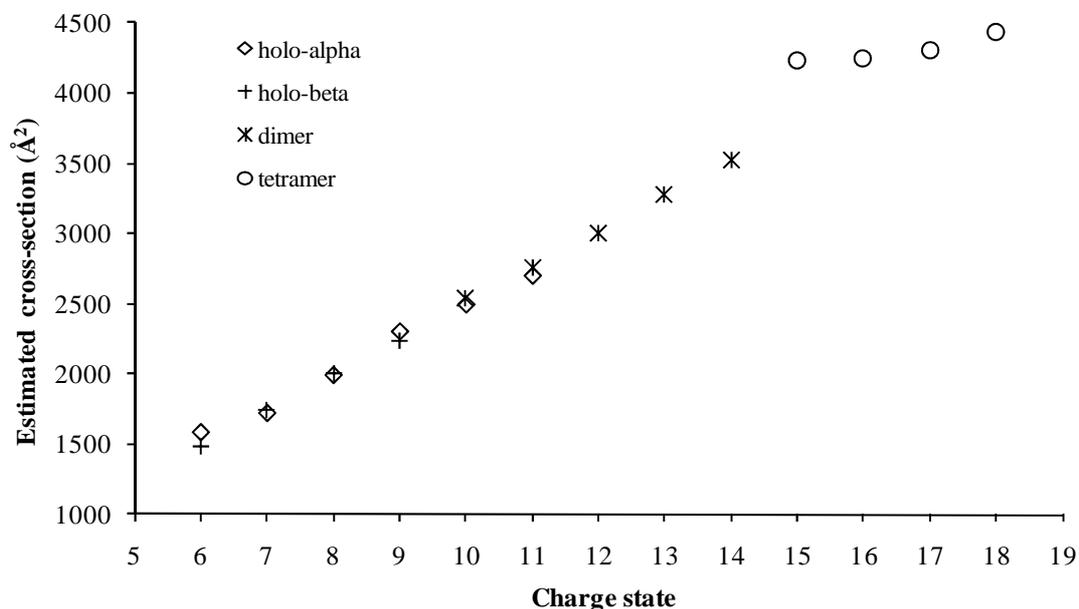


**Figure 3.4:** Mass spectra of a.) commercially-sourced Hb A and b.) fresh Hb A analysed by ESI-TOF-MS under near-physiological conditions. Spectra are labelled with charge states of heterodimer (D), apo- and holo-monomers (subscripts “a” and “h” refer to apo- and holo-forms, respectively) and heme-deficient dimer.

In this work, many more charge states of  $\alpha^a$  than  $\beta^a$  are observed. The number of charges accepted on a protein is related to the number of exposed basic sites on the protein's surface. A more folded protein has fewer of its basic sites exposed than an unfolded conformation and thus cannot accept as many charges. This may suggest that the  $\alpha$ -chain adopts more unfolded conformations in the gas phase than is possible for the  $\beta$ -chain but, alternatively, the absence of higher charge states of  $\beta^a$  in the spectra may be due to differences in the desolvation behaviours of  $\alpha$ - and  $\beta$ -monomers. The  $\alpha$ -chain ionises preferentially over the  $\beta$ -chain due to its greater non-polar character, thereby competing more effectively for charge (Kuprowski *et al.* 2007).

### 3.3.7 Hemoglobin tetramer assembly

By estimating the cross-sections of Hb A monomer, dimer and tetramer, one can obtain a picture of the assembly process (Figure 3.5).

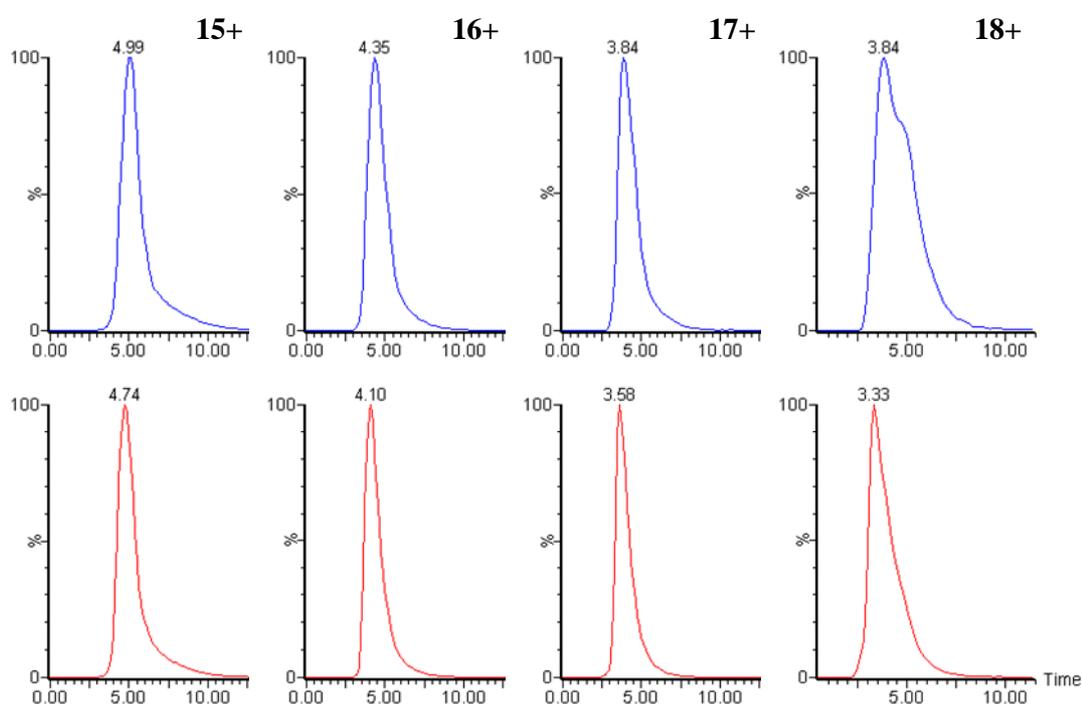


**Figure 3.5:** Average estimated cross sections for holo- $\alpha$ , holo- $\beta$ , heterodimer, and Hb A tetramer, from three datasets.

The  $[M + 12H]^{12+}$  charge state of dimer has an estimated cross-section of  $3001 \text{ \AA}^2$ . The  $[M + 6H]^{6+}$  charge states of  $\alpha^h$  and  $\beta^h$  have estimated average cross-sections of  $1583$  and  $1488 \text{ \AA}^2$  respectively. If these two globin monomers came together to form a dimer and if one assumes no major structural changes occurred, the cross-section of that dimer would be approximately the sum of the cross-sections of the two constituent parts. One would further expect that the cross-section observed would be slightly smaller than the sum of the monomer subunits as the contact area on both of the monomers would be compacted and contribute less to the overall cross-section. The observed data supports this argument.

### 3.3.8 Hb A vs Hb SS

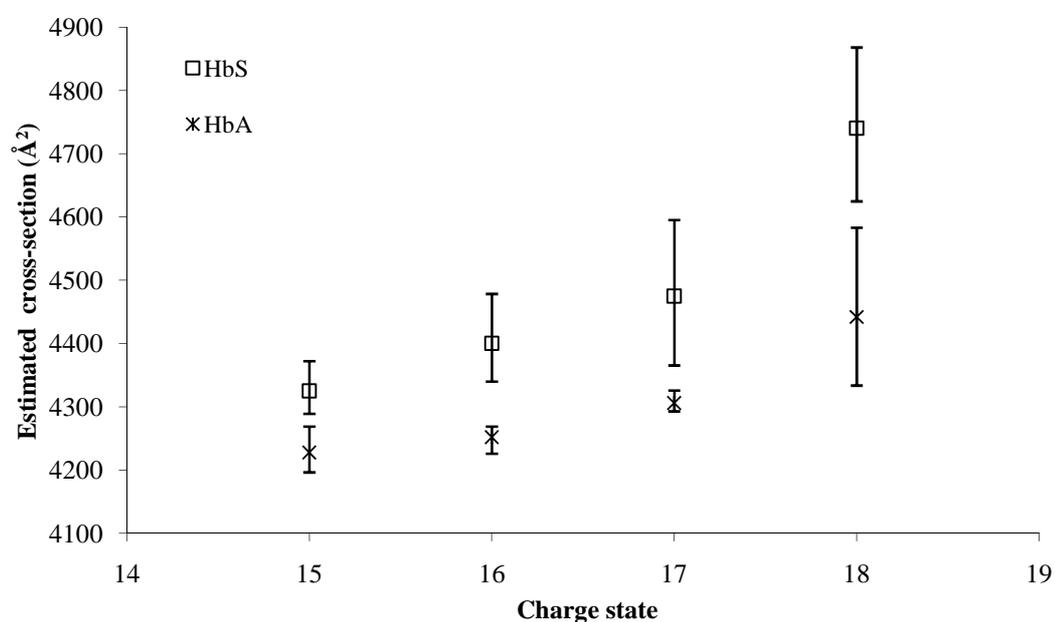
The arrival time distributions obtained for four different charge states of Hb A and Hb SS, for a single dataset, are illustrated below in Figure 3.6.



**Figure 3.6:** Arrival time distributions for charge states states  $[M + 15H]^{15+}$  to  $[M + 18H]^{18+}$  of Hb SS (top) and Hb A (bottom).

For each charge state, the arrival time distribution for Hb A is similar to that of Hb SS and they are irresolvable. The average arrival time for Hb A is consistently shorter than that for Hb SS. This indicates that Hb A exists within a more compact conformation than Hb SS.

The average estimated collision cross-sections for Hb A and Hb SS, for four different charge states, are illustrated in Figure 3.7. The data indicate a difference in cross-section between normal and sickle-cell hemoglobin, and a variation in cross-section with charge state.



**Figure 3.7:** Estimated cross sections for Hb A and Hb SS tetramers for three different charge states, showing averaged values from three datasets with corresponding errors.

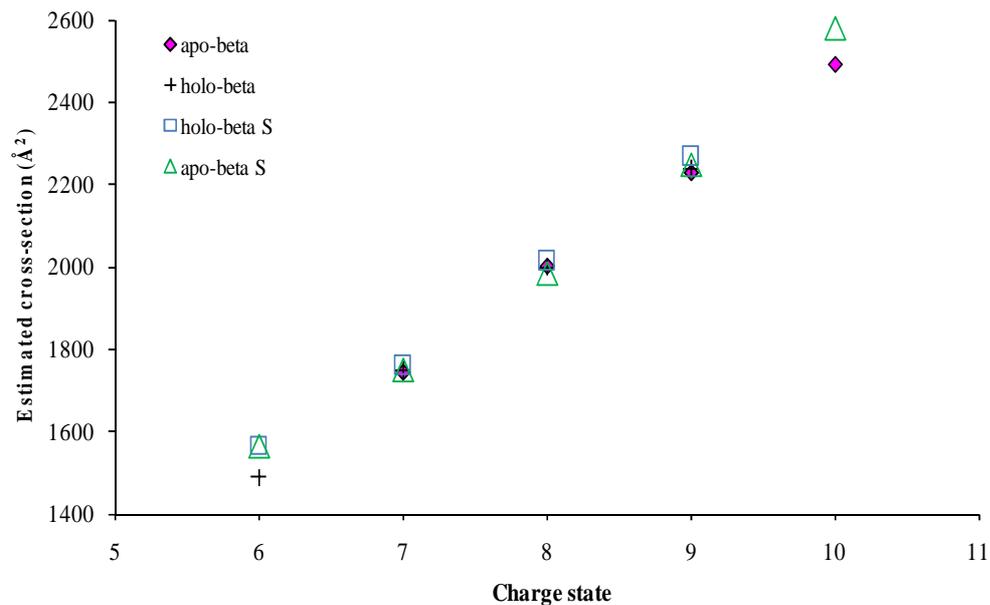
For the charge states studied, the cross-sections observed for Hb SS are somewhat larger than those of Hb A. Secondary, tertiary and quaternary structural considerations make it difficult to determine what the charge state of a molecule should be, theoretically, within a particular solvent at a particular pH. It has been shown that the lowest charge states observed under near physiological conditions are, in many cases, most representative of the native protein structure (Scarff *et al.* 2008). The  $[M + 18H]^{18+}$  charge state, for Hb A and Hb SS, may represent a tetrameric structure which is beginning to denature.

The reproducibility of the cross-sections estimated for the  $[M + 15H]^{15+}$ ,  $[M + 16H]^{16+}$  and  $[M + 17H]^{17+}$  charge states of Hb A, between the three replicate datasets, is  $\pm 1\%$  and this is believed to be representative of the reproducibility capabilities of the experiment. The larger deviation in estimation of cross-section for the same charge states for Hb SS, of  $\pm 3\%$ , may reflect the presence of a broader population of conformations of the Hb SS molecule, of similar cross-section. The arrival time distributions for the different charge states of Hb SS are also broader than those for Hb A. This provides further evidence to suggest that the Hb SS molecule exists in a broader population of conformations in comparison to the Hb A molecule.

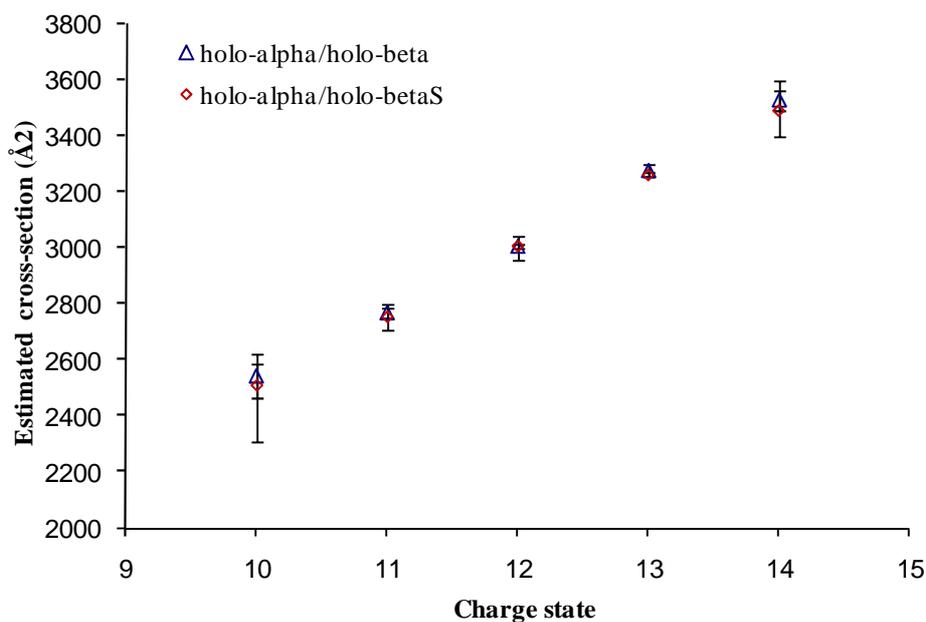
The rotationally averaged cross-sections for Hb A and Hb SS calculated from X-ray crystallographic structures were  $3313 \text{ \AA}^2$  and  $3733 \text{ \AA}^2$  for the PA and  $4343 \text{ \AA}^2$  and  $4775 \text{ \AA}^2$  for the EHSS respectively.

Values estimated experimentally for the  $[M + 15H]^{15+}$ ,  $[M + 16H]^{16+}$  and  $[M + 17H]^{17+}$  charge states of Hb A and Hb SS fall between these two theoretical approximations and agree with the X-ray observation that Hb SS has a larger cross-section than Hb A. Gas-phase conformations, although illustrative of solution-phase structures under controlled conditions, have been shown previously to be smaller than those predicted by EHSS approximations (Hoaglund-Hyzer *et al.* 1999); a more compact conformation is thought to be adopted in the gas phase as increased intramolecular interactions cause polar side chains to collapse onto the protein's surface (Shelimov *et al.* 1997).

Whilst differences in the overall quaternary structure of Hb A and Hb SS were detected no significant differences in mobility were observed between either normal and sickle beta monomers (Figure 3.8), or  $\alpha\beta$  and  $\alpha\beta_S$  dimers (Figure 3.9).



**Figure 3.8:** Average estimated cross-sections for  $\beta^a$ ,  $\beta^h$ ,  $\beta_s^a$  and  $\beta_s^h$  monomers, from three datasets.



**Figure 3.9:** Average estimated cross sections for heterodimers  $\alpha^h\beta^h$  and  $\alpha^h\beta_s^h$ , from three datasets.

This suggests that any differences present in the monomer structures are not of sufficient magnitude to enable them to be resolved by these experiments. The differences in the cross-sections of the intact tetramers are resolvable by this method.

### 3.4 Conclusions

This work has demonstrated the use of TWIM-MS to probe gas-phase conformations of three-dimensional protein structure and non-covalent complexes.

TWIM-MS has been successfully used to study hemoglobin tetramers. Cross-sections calculated for intact hemoglobin tetramers are comparable to those estimated from published X-ray crystallography data and conformational differences are observed between the Hb A and Hb SS molecules. Non-tetrameric species observed, including apo- and holo- forms of  $\alpha$ - and  $\beta$ -monomers and  $\alpha^h\beta^h$ -dimers, are thought to be naturally present in equilibrium in solution and not products of fragmentation during the ESI process.

$\alpha$ - and  $\beta$ -monomers have similar cross-sections to each other suggesting that they maintain a similar fold in the gas phase. Apo- and holo- forms of the monomers also have similar cross-sections suggesting that  $\alpha$ - and  $\beta$ -monomers can retain a folded structure in the absence and presence of the heme group. Extensively disordered monomer structures are not observed.

A heme-deficient dimer is not observed in the analysis of fresh blood samples. The results do not suggest that the association of  $\beta^a$  with  $\alpha^h$  is required in order for the  $\beta$ -monomer to recruit heme. The results, obtained on fresh blood samples rather than commercially prepared samples, do not support the hypothesis that a heme-deficient dimer is an essential intermediate in the tetramer assembly process.

### 3.5 References

**Boys, B. L. and Konermann, L.** (2007). Folding and Assembly of Hemoglobin Monitored by Electrospray Mass Spectrometry Using an On-line Dialysis System. *Journal of the American Society for Mass Spectrometry*. **18**, 8-16.

**Boys, B. L., Kuprowski, M. C. and Konermann, L.** (2007). Symmetric Behavior of Hemoglobin  $\alpha$ - and  $\beta$ - Subunits during Acid-Induced Denaturation Observed by Electrospray Mass Spectrometry. *Biochemistry*. **46**, 10675-10684.

**Daniel, Y. A., Turner, C., Haynes, R. M., Hunt, B. J. and Dalton, R. N.** (2005). Rapid and specific detection of clinically significant haemoglobinopathies using electrospray mass spectrometry-mass spectrometry. *British Journal of Haematology*. **130**, 635-643.

**Griffith, W. P. and Kaltashov, I. A.** (2003). Highly asymmetric interactions between globin chains during hemoglobin assembly revealed by electrospray ionization mass spectrometry. *Biochemistry*. **42**, 10024-10033.

**Griffith, W. P. and Kaltashov, I. A.** (2007). Protein Conformational Heterogeneity as a Binding Catalyst: ESI-MS Study of Hemoglobin H Formation. *Biochemistry*. **46**, 2020-2026.

**Hoaglund-Hyzer, C. S., Counterman, A. E. and Clemmer, D. E.** (1999). Anhydrous Protein Ions. *Chemical Reviews*. **99**, 3037-3080.

**Hossain, B. M. and Konermann, L.** (2006). Pulsed Hydrogen/Deuterium Exchange MS/MS for Studying the Relationship between Noncovalent Protein Complexes in Solution and in the Gas Phase after Electrospray Ionization. *Anal. Chem.* **78**, 1613-1619.

**Kuprowski, M. C., Boys, B. L. and Konermann, L.** (2007). Analysis of Protein Mixtures by Electrospray Mass Spectrometry: Effects of Conformation and Desolvation Behavior on the Signal Intensities of Hemoglobin Subunits. *Journal of the American Society for Mass Spectrometry*. **18**, 1279-1285.

**Leutzinger, Y. and Beychok, S.** (1981). Kinetics and mechanism of heme-induced refolding of human  $\alpha$ -globin. *Proceedings of the National Academy of Sciences*. **78**, 780-784.

**Murayama, M.** (1967). Structure of Sick Cell Hemoglobin and Molecular Mechanism of the Sickling Phenomenon. *Clinical Chemistry*. **13**, 578-588.

**Ofori-Acquah, S. F., Green, B. N., Davies, S. C., Nicolaides, K. H., Serjeant, G. R. and Layton, D. M.** (2001). Mass spectral analysis of asymmetric hemoglobin hybrids: Demonstration of Hb FS ( $\alpha_2\gamma\beta^S$ ) in sickle cell disease. *Analytical Biochemistry*. **298**, 76-82.

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

**Shackleton, C. H. L., Falick, A. M., Green, B. N. and Witkowska, H. E.** (1991). Electrospray mass spectrometry in the clinical diagnosis of variant hemoglobins. *Journal of Chromatography: Biomedical Applications*. **562**, 175-190.

**Shelimov, K. B., Clemmer, D. E., Hudgins, R. R. and Jarrold, M. F.** (1997). Protein Structure in Vacuo: Gas-Phase Conformations of BPTI and Cytochrome c. *Journal of the American Chemical Society*. **119**, 2240-2248.

**Shimizu, A., Nakanishi, T. and Miyazaki, A.** (2006). Detection and characterization of variant and modified structures of proteins in blood and tissues by mass spectrometry. *Mass Spectrometry Reviews*. **25**, 686-712.

**Tahallah, N., Pinkse, M. M., C. S. and Heck, A. J. R.** (2001). The effect of the source pressure on the abundance of ions of noncovalent protein assemblies in an electrospray ionization orthogonal time-of-flight instrument. *Rapid Communications in Mass Spectrometry*. **15**, 596-601.

**Thalassinos, K., Slade, S. E., Jennings, K. R., Scrivens, J. H., Giles, K., Wildgoose, J., Hoyes, J., Bateman, R. H. and Bowers, M. T.** (2004). Ion mobility mass spectrometry of proteins in a modified commercial mass spectrometer. *International Journal of Mass Spectrometry*. **236**, 55-63.

**Versluis, C. and Heck, A. J. R.** (2001). Gas-phase dissociation of hemoglobin. *International Journal of Mass Spectrometry*. **210-211**, 637-649.

**Wild, B. J., Green, B. N., Cooper, E. K., Lalloz, M. R., Erten, S., Stephens, A. D. and Layton, D. M.** (2001). Rapid identification of hemoglobin variants by electrospray ionization mass spectrometry. *Blood Cells, Molecules, and Diseases*. **27**, 691-704.

Chapter 4: A Quantitative and  
Conformational Study of the  
Autophosphorylation of the Histidine  
Kinase VanS

---

## 4.1 Introduction

The investigation of protein structure by approaches based upon mass spectrometry can yield information relating to protein stoichiometry, non-covalent interactions, ligand interactions and three-dimensional conformation (Heck and van den Heuvel 2004). When coupled with ion mobility separation, mass spectrometry can be exploited to provide information pertaining to global protein structure. IM-MS is particularly useful as a biophysical tool to study protein dynamics, in real time, when conformational changes occur upon interaction with a ligand. This is especially useful when structural characterisation by more conventional approaches, such as NMR spectroscopy or X-ray crystallography, is difficult if not impossible.

Mass spectrometry can also be used to examine enzymatic protein function by monitoring the conversion of a substrate to a product (Liesener and Karst 2005). Multiple MS-based-assays and detection methods have been developed to follow enzyme kinetics and have been validated by comparison with results obtained by other methods such as UV-absorbance, fluorescence and radioactive labelling assays (Ge *et al.* 2001; Norris *et al.* 2001; Steinkamp *et al.* 2004). Methods based on MS have focussed on the analysis of small substrates and have used a multiple reaction monitoring (MRM) or selected ion monitoring (SIM) scan mode to quantify accurately the amount of product produced. Assays based on MS are an attractive alternative to established approaches as they do not require labels, are non-radioactive, can simultaneously monitor the fate of multiple analytes and unambiguously identify reaction products (Liesener and Karst 2005).

In this study TWIM-MS was used to probe conformational changes which are proposed to occur during autokinase activity of the sensor histidine kinase protein, VanS<sub>A</sub>, from the human pathogen *Enterococcus faecium*. The rate of autophosphorylation of this protein was also monitored by the use of a mass spectrometry-based method.

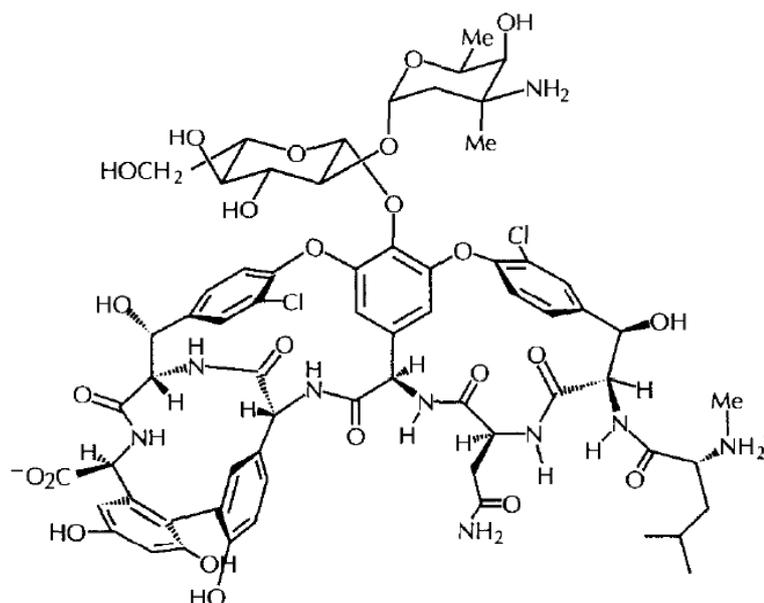
### 4.1.1 Histidine kinases

Histidine kinases (HKs), of which VanS<sub>A</sub> is an example, are the predominant mediators of prokaryotic signal-transduction and often possess multiple functions such as autokinase, phospho-transfer and phosphatase activities (Stock *et al.* 2000; Marina *et al.* 2005). The majority are homodimeric membrane proteins with a predicted extracellular sensor domain and a cytoplasmic core comprising a minimum of two distinct functional domains: a catalytic ATP-binding (CA) domain and a dimerisation phosphotransfer (DHp) domain (Gao and Stock 2009). The different enzymatic activities HKs carry out require the involvement of one or both of these domains and are thought to relate to different conformational states. A crystal structure representing the entire cytosolic portion of a sensor histidine kinase has recently been published (Marina *et al.* 2005). An additional structure containing the CA and DHp domains of a histidine kinase has also been obtained (Bick *et al.* 2009). These structures have both suggested that conformational changes are required for histidine kinase function. Several molecular structures have been reported for the CA and DHp domains of HKs in isolation but there is still limited structural characterisation of different functional states (Marina *et al.* 2005). HKs are therefore suitable candidates for study by IM-MS.

VanS<sub>A</sub> is of particular interest because it is the critical sensor kinase controlling the induction of high-level glycopeptide resistance in enterococci.

### 4.1.2 Vancomycin and vancomycin resistance

Vancomycin (Figure 4.1) is a naturally occurring glycopeptide antibiotic which is clinically important as a front-line treatment for Gram-positive bacterial infections including methicillin-resistant *Staphylococcus aureus* (Hong *et al.* 2008).



**Figure 4.1:** The structure of vancomycin. Adapted from Walsh *et al.* (1996).

Bacterial resistance to vancomycin was first reported in 1988 (Leclercq *et al.* 1988) and since then the spread of vancomycin resistance through Enterococcal bacterial populations has become a severe public health concern.

Vancomycin targets the synthesis of the bacterial cell wall peptidoglycan. Mature peptidoglycan consists of polymerised glycan chains cross-linked by short peptides (Healy *et al.* 2000). The glycan chains consist of alternating  $\beta$ -1,4 linked N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc) residues. A pentapeptide stem replaces the D-lactoyl group of each MurNAc residue and usually comprises of L-Ala- $\gamma$ -D-Glu-L-Lys-D-Ala-D-Ala in gram-positive bacteria (Vollmer *et al.* 2008). Cross-linking of the pentapeptide stems generally occurs between the carboxyl group of D-Ala at position 4 and the amino group of the diamino acid at position 3, through transpeptidase action, and results in the loss of a terminal D-Ala (Hong *et al.* 2008; Vollmer *et al.* 2008).

Vancomycin binds to the D-Ala-D-Ala termini of these pentapeptide chains and sterically prevents formation of cross-linking interactions between them, which are required for proper maturation of the cell wall (see Figure 4.2). Without this type of

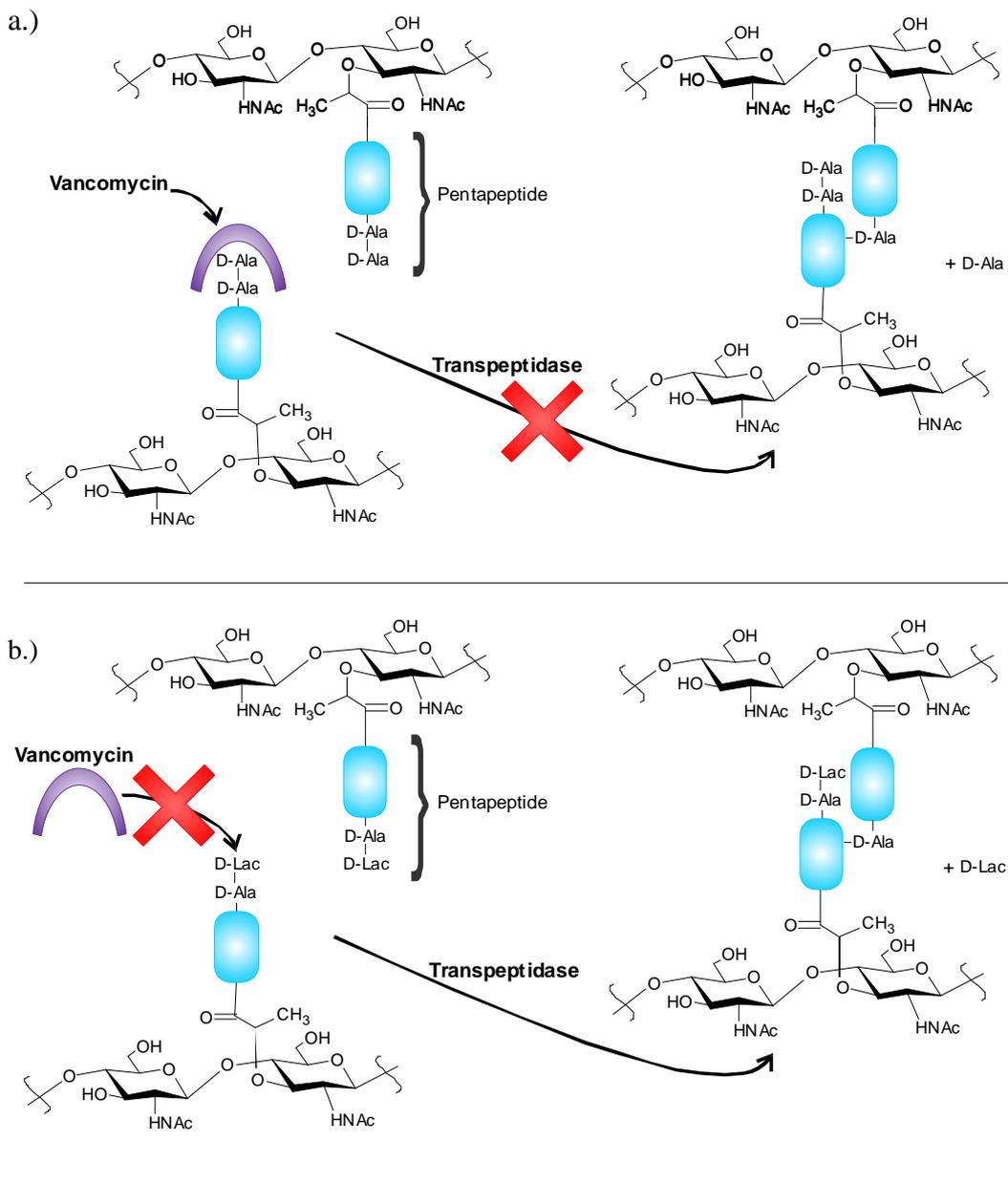
interchain cross-linking, the peptidoglycan layer is weakened, such that it cannot withstand the internal osmotic pressure of the cell. This leads to cell lysis (Walsh *et al.* 1996).

#### **4.1.3 The mechanism of vancomycin resistance**

Six vancomycin resistant phenotypic strains of Enterococci have been identified, designated VanA, B, C, D, E and VanG (Courvalin 2006). These phenotypes exhibit different mechanisms for conferring resistance, involving multiple genes. Resistance is caused by one of two different pathways which lead to the replacement of the terminal D-Ala on pentapeptide precursors with D-Lac or D-Ser (Pootoolal *et al.* 2002). Vancomycin no longer effectively binds to these precursors and so can not interrupt mature peptidoglycan synthesis.

The mechanism of vancomycin resistance in enterococci is controlled by a two-component system (Stock *et al.* 2000) consisting of VanS, a sensor histidine kinase (HK), and VanR, a response regulator (RR) (Arthur *et al.* 1992a). The genes involved within the vancomycin resistance pathway depend on the phenotypic strain. Here VanS and VanR from the VanA phenotypic strain are investigated and are thus described as VanS<sub>A</sub> and VanR<sub>A</sub> to denote their resistance type. The VanA phenotype is carried on a transposon and confers high-level inducible vancomycin resistance. It is the phenotype of most clinical importance for which Vancomycin-resistant *Staphylococcus aureus* strains have been reported (Courvalin 2006).

The pentapeptide stems of VanA phenotypic strains are terminated by D-Ala-D-Lac (Arthur *et al.* 1992b). The replacement of the terminal D-Ala with D-Lac, reduces the affinity of vancomycin for this terminal stem by over 1000-fold (Bugg *et al.* 1991) (see Figure 4.2).

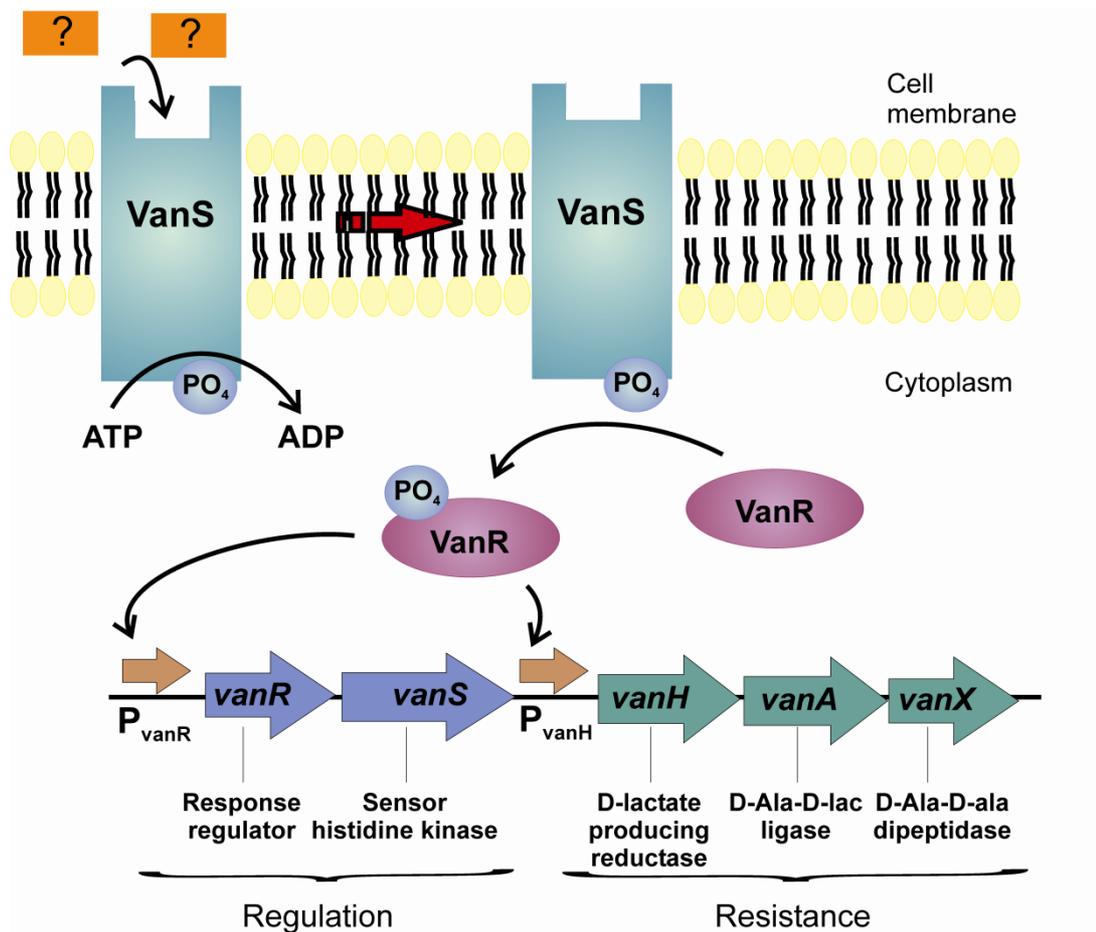


**Figure 4.2:** The mode of action of vancomycin.

a.) Vancomycin binds to the D-Ala-D-Ala termini of extracellular peptidoglycan precursors. Transpeptidase is unable to access its substrate, cross-linking of pentapeptide chains cannot occur and formation of mature peptidoglycan is blocked. b.) Vancomycin resistant bacteria (VanA phenotype) produce peptidoglycan precursors terminating in D-Ala-D-Lac. The affinity of vancomycin for D-Ala-D-Lac is much lower than for D-Ala-D-Ala so transpeptidation is able to occur and mature peptidoglycan is synthesised. (Adapted from Healy et al., 2000; Hong et al., 2008 and Walsh et al., 1996).

#### 4.1.4 The VanS-VanR two-component system

The VanS-VanR two-component system regulates the induction of vancomycin resistance. In response to a signal the sensor histidine kinase (VanS<sub>A</sub>) autophosphorylates on a conserved histidine residue (His164) and transfers this phosphate to an aspartic acid residue (Asp53) on the response regulator (VanR<sub>A</sub>) (Wright *et al.* 1993). Once phosphorylated, the response regulator then associates with the promoter region of the *van* operon (Holman *et al.* 1994). This leads to expression of the genes which induce vancomycin resistance (see Figure 4.3).



**Figure 4.3:** Mechanism of induction of vancomycin resistance by the VanS<sub>A</sub>-VanR<sub>A</sub> two-component system. A signal is received by VanS<sub>A</sub> stimulating autophosphorylation on a conserved histidine residue. VanS<sub>A</sub> transfers this phosphate to a conserved aspartic acid residue of VanR<sub>A</sub>. Once phosphorylated, VanR<sub>A</sub> activates transcription of the genes which confer vancomycin resistance. Adapted from Healy *et al.* (2000).

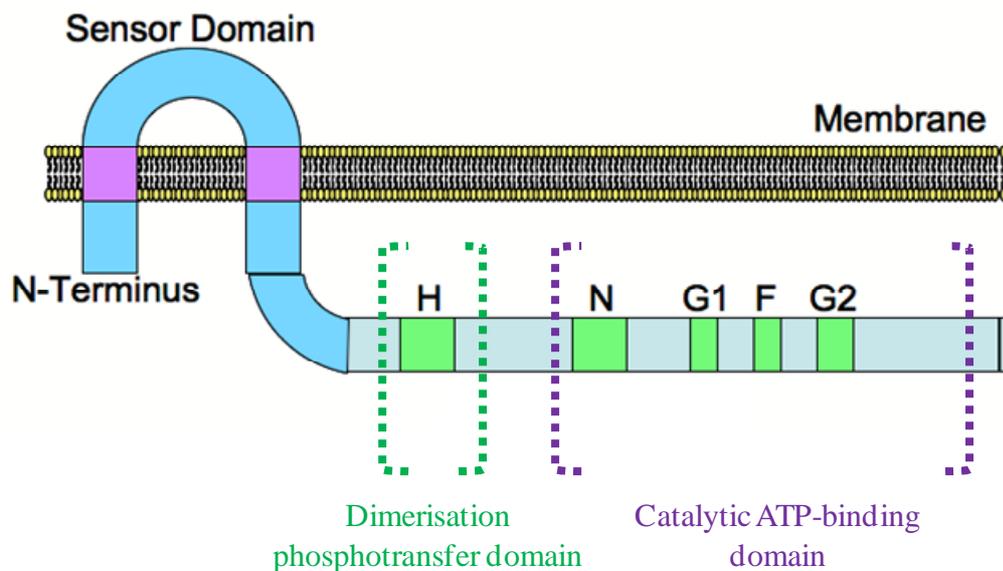
In the absence of vancomycin, VanS is thought to work as a phosphatase, ensuring that VanR remains unphosphorylated so that the *van* genes are not expressed (Wright *et al.* 1993).

The autophosphorylation, phosphotransferase and dephosphorylation activities of VanS were demonstrated in a study by Wright *et al.* (1993) by use of a fusion protein construct (Wright *et al.* 1993). The following reaction schemes were characterised:

- a) Autophosphorylation:  $\text{VanS-His} + \text{ATP} \rightarrow \text{HK-His}\sim\text{P} + \text{ADP}$
- b) Phosphotransfer:  $\text{HK-His}\sim\text{P} + \text{VanR-Asp} \rightarrow \text{HK-His} + \text{VanR-Asp}\sim\text{P}$
- c) Dephosphorylation:  $\text{VanR-Asp}\sim\text{P} + \text{H}_2\text{O} \rightarrow \text{VanR-Asp} + \text{P}_i$

The nature of the direct molecular ligand that interacts with the VanS sensor domain has not been determined for any VanS-VanR signal transduction system (Hong *et al.* 2008). Signal transduction is initiated in the presence of vancomycin. It is therefore thought that VanS is activated either directly by vancomycin itself or by an intermediate in cell wall biosynthesis which accumulates due to vancomycin action (Hong *et al.* 2008).

VanS<sub>A</sub> has a molecular weight of 44 kDa and contains 384 amino acid residues. It has an extracellular sensor domain, two transmembrane domains and a histidine kinase domain (amino acids 161-384). Within the histidine kinase domain amino acid motifs (the H, N, G1, F and G2 boxes) are found, which are conserved within the majority of HKs and are involved in autophosphorylation and ATP-binding (Gao and Stock 2009). The H box contains the histidine residue which is the site of autophosphorylation and is situated in the DHp domain whilst the N, G1, F and G2 boxes are found in the CA domain and form the ATP binding site. Figure 4.4 below illustrates domain organisation within VanS<sub>A</sub>.



**Figure 4.4:** Schematic of VanS<sub>A</sub> domain organisation based upon protein sequence information. Characteristic H, N, G1, F and G2 motifs, common to sensor histidine kinases are highlighted. Adapted from (Depardieu *et al.* 2003).

There are currently no examples of a VanS sensor histidine kinase structure in the Protein Data Bank despite extensive crystallisation trials (Quigley and Roper, unpublished results). How the VanS-VanR two-component system is controlled and the exact mechanisms of VanS autophosphorylation, phosphotransfer and phosphatase activity are not understood. *In vivo*, VanS<sub>A</sub> functions as a dimer and becomes phosphorylated in response to the activation of its sensor domain. *In vitro*, without the presence of its sensor domain, VanS<sub>A</sub> possesses autokinase activity (Wright *et al.* 1993).

In the present study these observations are investigated further by the use of TWIM-MS. VanS<sub>A</sub> truncates, lacking the membrane-spanning domains, are examined as they undergo autophosphorylation. TWIM-MS is used to explore whether conformational changes in VanS<sub>A</sub> are associated with autokinase activity. An MS-based method is also used to monitor autophosphorylation of VanS<sub>A</sub> over time and the results of this approach are compared to those obtained from radioactive labelling studies.

## 4.2 Materials and Methods

### 4.2.1 Expression and purification of VanS<sub>A</sub> truncates

Expression and purification of VanS<sub>A</sub> truncates was conducted by Andrew Quigley, Department of Biological Sciences, University of Warwick, as detailed below.

Truncated forms of the *vanS<sub>A</sub>* gene, were amplified, by PCR, from the full-length *vanS<sub>A</sub>* gene sequence, from *Enterococcus faecium* BM4147 (Leclercq *et al.* 1988), which had previously been cloned into the pTTQ18 vector (Saidijam *et al.* 2005). VanS<sub>A</sub> truncated forms were  $\Delta$ 110,  $\Delta$ 140,  $\Delta$ 150 and  $\Delta$ 155, lacking 110, 140, 150 and 155 amino acids from the N-terminus of the full-length protein respectively. PCR primers incorporated 5'-*Bsa*I and 3'-*Xho*I restriction sites allowing for ligation of each PCR construct into a pET28a expression vector, in frame with a C-terminal His<sub>6</sub> tag. The His<sub>6</sub>-fused VanS<sub>A</sub> truncates were expressed in *Escherichia coli* BL21(DE3) harbouring pRosetta in LB medium containing 35  $\mu$ g/mL kanamycin at 25°C. Expression was carried out using the Studier auto-induction system (Studier 2005) at 25 °C with shaking at 180 rpm. Cells were harvested and stored at -20°C in Buffer A (20 mM HEPES (adjusted to pH 7.8), containing 500 mM NaCl, protease inhibitor cocktail and 10 % (v/v) glycerol) supplemented by 20 mM imidazole. Cells were thawed and disrupted by three 1 minute cycles of sonication on ice and centrifuged at 10,000g in a Beckman JA-25.50 rotor to pellet cell debris. The resulting supernatant was clarified by centrifugation at 50,000g for 30 minutes and applied to a pre-packed, 5 mL, Chelating High performance Nickel-Sepharose column, previously equilibrated and washed with Buffer A and 20 mM imidazole. VanS<sub>A</sub> truncates were eluted with Buffer A supplemented by 500 mM imidazole. Protein eluate was concentrated before application to a Superdex 200 HR 10/300 size exclusion column pre-equilibrated in 20 mM HEPES (adjusted to pH 7.8) containing 100 mM sodium chloride and 10 % (v/v) glycerol. A PD10 desalting column (GE Healthcare) in 20 mM HEPES (adjust to pH 7.8) was used to further desalt the protein. The protein was concentrated and reconstituted in fresh buffer to 5 mg/mL, flash frozen with liquid nitrogen, and stored at -80 °C.

#### 4.2.2 Sample preparation for MS

Protein construct stocks were thawed and diluted to 10  $\mu$ M in 10 mM ammonium bicarbonate (aq) adjusted to pH 7.8 (JT Baker, NJ, USA). Histidine phosphorylation is acid-labile so care was taken to keep the solution conditions above pH 7 to maintain the integrity of any phospho-histidine residues. Samples were desalted by use of centrifugal filter devices (Biomax-10 Membrane, Millipore, MA, USA). Samples (100  $\mu$ L) were diluted 5-fold in 10 mM ammonium bicarbonate (aq) pH 7.8 and spun at 10,000 *g* for 3 minutes. Sample retained on the filter was reconstituted in fresh solvent and the centrifugation step repeated three times. For each truncate, several samples were prepared as detailed above. The autophosphorylation reaction was induced by the incubation of a protein construct sample with 1 mM ATP solution and 1 mM MgCl<sub>2</sub> solution. The reaction was quenched, if required, by the addition of 5 mM EDTA solution.

#### 4.2.3 MS analysis

A Synapt HDMS System was used to perform all MS and TWIM-MS experiments described, under the conditions detailed in Chapter 2 section 2.2.3 unless otherwise stated. A mass acquisition range of 1000-4500 *m/z* and a source temperature of 110 °C were used for all experiments.

Each of the VanS<sub>A</sub> truncates were analysed by mass spectrometry at 10  $\mu$ M in 10 mM ammonium bicarbonate solution (pH 7.8) before and after incubation with 1 mM ATP solution and 1 mM MgCl<sub>2</sub> solution for one minute. Differences observed with the spectra obtained before and after incubation with ATP and magnesium chloride were recorded. These observations were used to decide upon following experiments. Corresponding TWIM-MS experiments were then conducted on VanS<sub>A</sub> truncates  $\Delta$ 140,  $\Delta$ 150 and  $\Delta$ 155.

VanS<sub>A</sub> truncates  $\Delta$ 110 and  $\Delta$ 155 were incubated with 1 mM ATP solution and 1 mM MgCl<sub>2</sub> solution for different time periods (0, 2, 5, 10, 20, 30, 45, 60 minutes), before quenching the reaction by the addition of 5 mM EDTA solution. These samples were desalted again by use of centrifugal filter devices, in the manner described above,

prior to MS analysis. The results of the MS analysis were interpreted to determine the amount and rate of autophosphorylation of the VanS<sub>A</sub> truncates  $\Delta 110$  and  $\Delta 155$ . The phosphorylation assay was repeated with two different batches of VanS<sub>A</sub> $\Delta 110$  to assess relative activity in each. The phosphorylation assay was repeated in triplicate for batch two of VanS<sub>A</sub> $\Delta 110$  to illustrate reproducibility with incubation time periods of 0, 2, 10, 15, 25, 30, 45 and 60 minutes. Incubation time periods were changed to allow examination of phosphorylation levels between previously used time points. The number of time points used was limited to reduce experimental time and sample use. The phosphorylation time course was only conducted on the smallest and largest VanS<sub>A</sub> truncates investigated due to availability of sample and time constraints.

To allow for the study of phosphorylated and ATP-bound species the autophosphorylation reaction was allowed to proceed without quenching with EDTA. To study non-phosphorylated protein the reaction mix was prepared without the addition of ATP.

#### **4.2.4 Phosphorylation analysis**

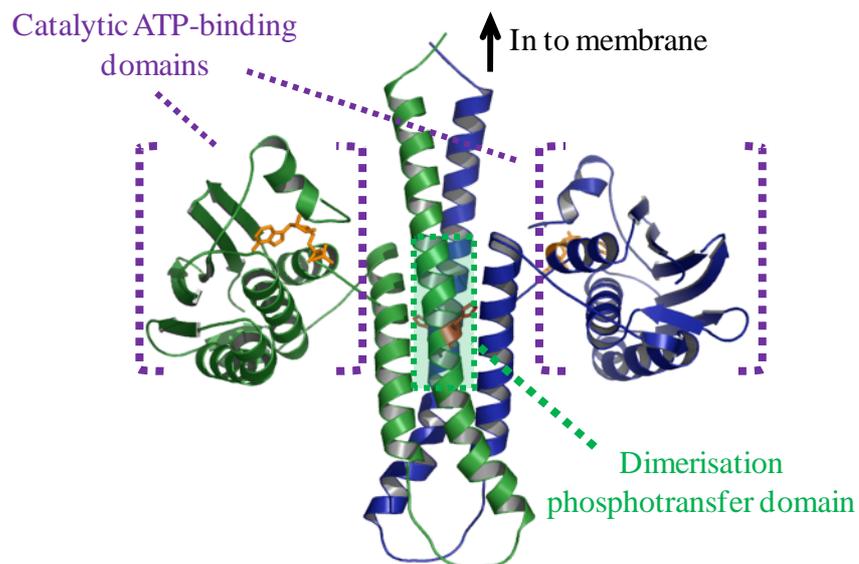
Spectra acquired for VanS<sub>A</sub> samples that were allowed to undergo phosphorylation for different time periods were deconvoluted onto a true mass scale, to provide a molecular mass measurement, with MaxEnt, a maximum entropy algorithm (Ferrige *et al.* 1992). Mass measurements obtained for the different truncates were compared to theoretical masses expected for each construct. The relative amount of phosphorylation observed was calculated from the ratio of non-phosphorylated to phosphorylated protein present, based on data derived from the deconvoluted spectra. This was only possible if spectra contained sufficient resolution such that the peak in the deconvoluted spectrum corresponding to phosphorylated protein was separated from those corresponding to alternative adduct peaks, i.e. tri-sodiated or tetra-sodiated peaks.

Data obtained on the amount of phosphorylated product produced over time were used to create enzyme progress curves, fitted to a non-linear regression model, in Prism (GraphPad Software Inc., CA, USA).

Results obtained for this phosphorylation assay for VanS<sub>A</sub>Δ110 batch one, were compared with results obtained by use of radioactive labelling.

#### 4.2.5 Modelling the mechanism of autokinase activity

In order to relate experimental observations to the phosphorylation mechanism of VanS<sub>A</sub> a pictorial representation of the protein was used. This was based upon the X-ray crystallographic structure of the cytosolic domain of a sensor histidine kinase (HK853) from *Thermatoga maritima* (Marina *et al.* 2005) illustrated in Figure 4.5 below. This protein shares only 26 % sequence identity with VanS<sub>A</sub> but, as a histidine kinase, it possesses CA and DHp domains which are likely to be highly conserved.

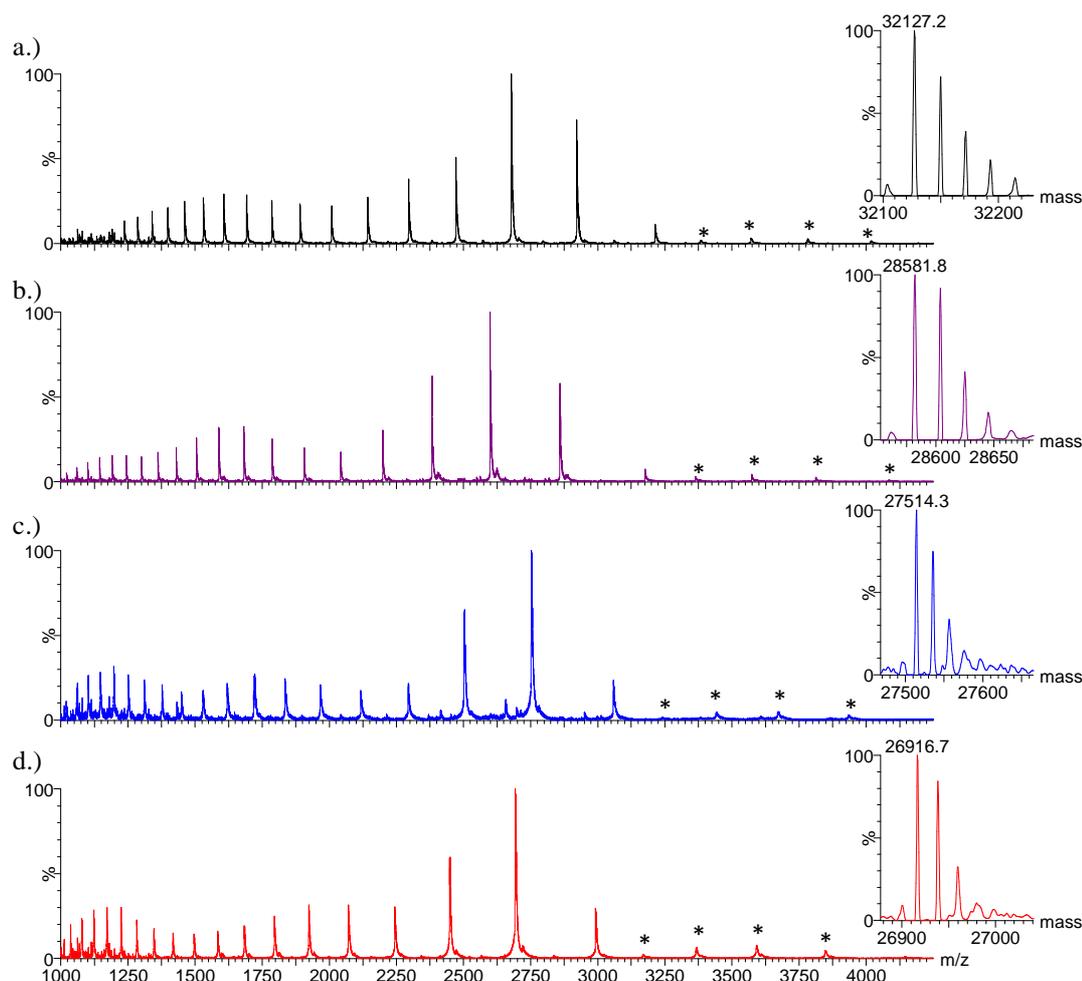


**Figure 4.5:** Representation of the crystal structure of the cytoplasmic domain of a sensor histidine kinase (HK853) from *T. maritima* (PDB; 2C2A). The two HK853 monomers (blue and green) exhibit a small dimerisation region within which the conserved histidine residue is shown in red. A non-hydrolyzable ATP analog is located in each ATP binding site within each CA domain and is shown in orange.

## 4.3 Results and Discussion

### 4.3.1 ESI-MS spectra

Analysis of VanS<sub>A</sub> truncates, under non-denaturing conditions, by use of ESI-MS showed the presence of primarily monomeric species and suggested the presence of multiple conformational states (see Figure 4.6).



**Figure 4.6:** Comparison of ESI-MS spectra obtained for VanS<sub>A</sub> truncates a.)  $\Delta 110$  b.)  $\Delta 140$  c.)  $\Delta 150$  and d.)  $\Delta 155$  with, inset, corresponding deconvoluted spectra respectively.  $m/z$  species representative of dimer are illustrated by an asterisk (\*).

Many more charge states were observed within the spectra than would be expected to be obtained from a protein having a single conformation, analysed under non-denaturing conditions. A small amount of homodimer was also observed within the

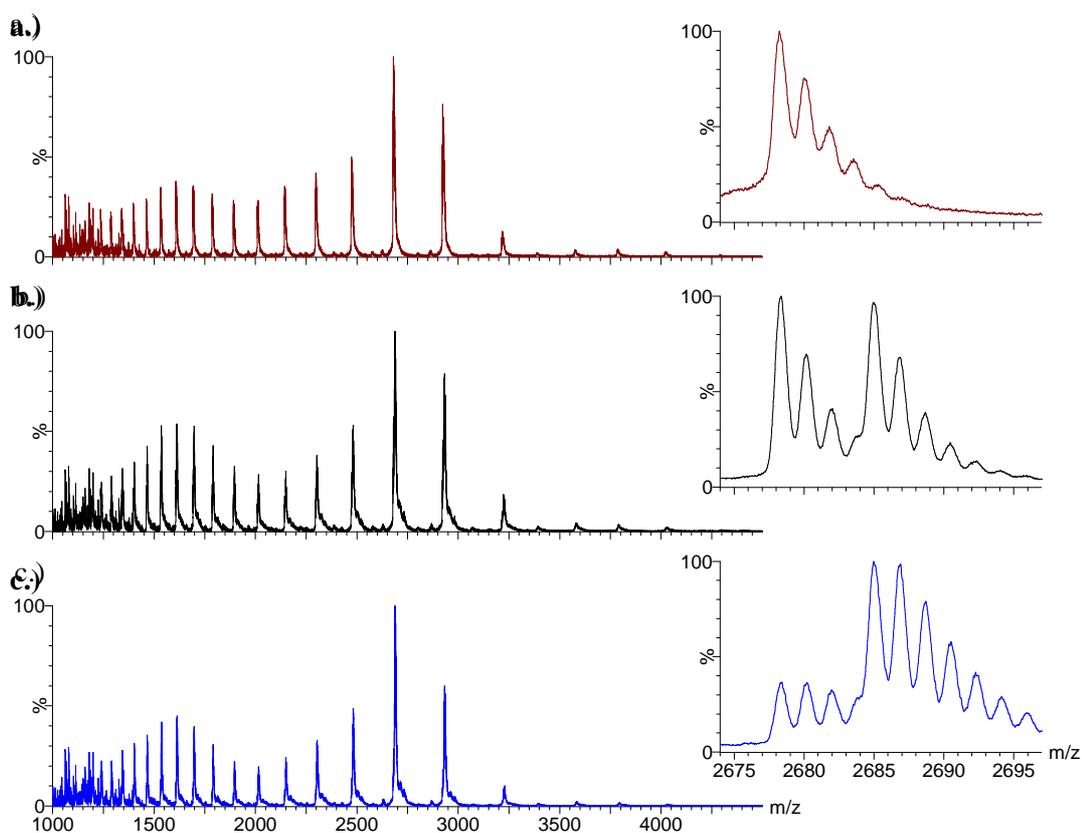
spectra. Data were deconvoluted to give masses of 32,127.2 Da, 28581.8 Da, 27514.3 Da and 26916.7 Da, for VanS<sub>A</sub> truncates  $\Delta$ 110,  $\Delta$ 140,  $\Delta$ 150 and  $\Delta$ 155 respectively, in good agreement with the theoretical masses of these constructs (32,127.5 Da, 28582.4 Da, 27514.3 Da and 26916.6 Da respectively).

Analytical gel filtration results suggest VanS<sub>A</sub> exists in solution as a dimer (Dr Quigley, personal communication). It may dissociate readily to monomer upon entry in to the gas phase because the interactions at the dimer interface are mainly hydrophobic. Hydrophobic interactions are not conserved well in the gas phase since they rely on the presence of surrounding solvent molecules. ESI-spectra obtained for VanS<sub>A</sub> indicated the presence of approximately 10 % dimer. The observation of an indication of a small amount of complexed species (10-20 %) within spectra is consistent with that observed in previous studies performed on non-covalent complexes where hydrophobic interactions were known to play a major role in their solution states (Robinson *et al.* 1996; Wu *et al.* 1997). Increasing the backing pressure did not lead to increased intensity of dimeric species within the mass spectra nor a change in the charge-state distribution observed. It is therefore likely that multiple dimeric conformations of the protein exist in solution with some relating to partially unfolded structures. Upon ionisation, interactions at the dimeric interface are lost but the structure of the monomer is likely preserved. Leary *et al.* have shown evidence to support the suggestion that a protein subunit can be gently dissociated from a macromolecular complex and maintain the same structure as it possesses when examined individually (Leary *et al.* 2009).

The charge-state distribution observed in the spectrum for VanS<sub>A</sub> $\Delta$ 110 contains two maxima. This indicates that there are at least two monomeric gas-phase conformations of this protein construct. The spectra for the other truncates have charge-state distributions exhibiting at least three maxima, signifying that these truncates exist in a minimum of three monomeric gas-phase conformations. The additional distribution of multiply-charged ions is observed at low  $m/z$  within these spectra, which suggests that it relates to a more unfolded conformation.

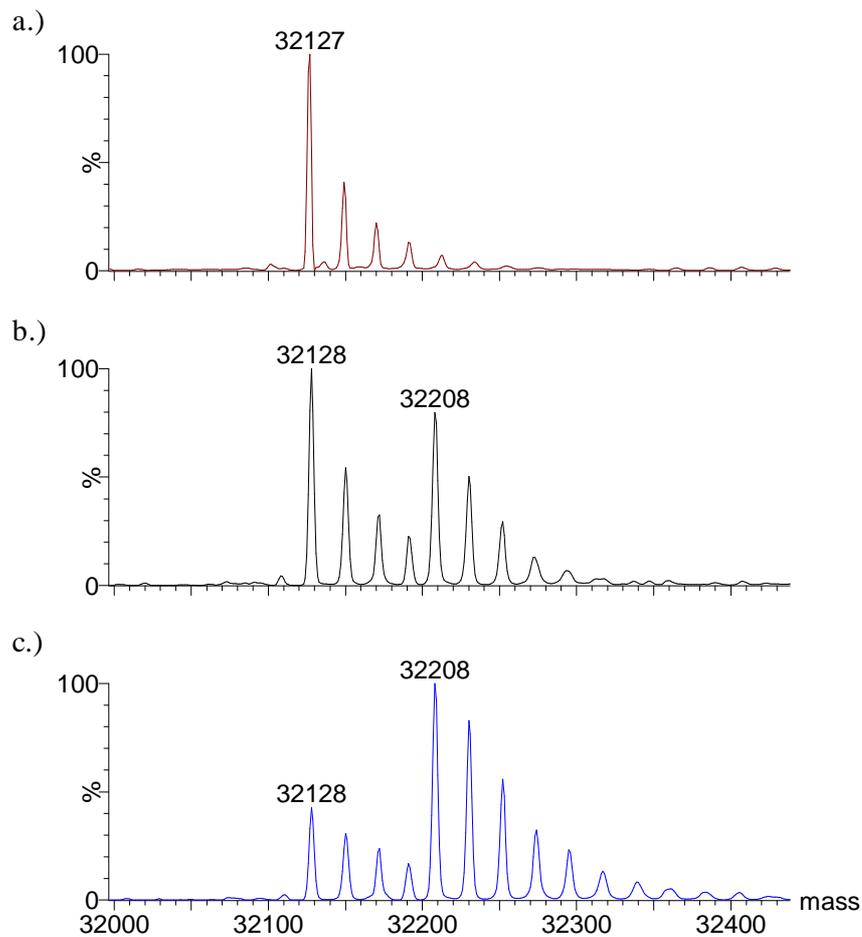
### 4.3.2 Autokinase activity

The ESI-MS spectra acquired, after VanS<sub>A</sub>Δ110 had been allowed to phosphorylate for different time periods, illustrated a relative decrease in non-phosphorylated protein and a relative increase in phosphorylated protein over time (Figure 4.7). The profile of the charge state distribution observed within the spectrum for non-phosphorylated protein was reflected in mass spectra obtained for all partially phosphorylated samples.



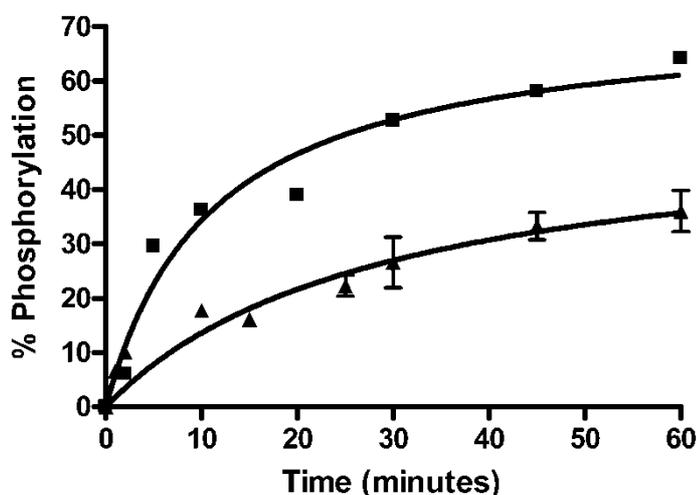
**Figure 4.7:** Mass spectra obtained for VanS<sub>A</sub>Δ110 samples allowed to phosphorylate for a.) 0, b.) 20, and c.) 60 minutes respectively. Inset spectra show a magnified region of each spectra containing the dominant charge state [M + 12H]<sup>12+</sup>. A decrease in the intensity of peaks relating to non-phosphorylated protein is observed and an increase in the intensity of peaks relating to phosphorylated protein is observed over time.

Figure 4.8 shows deconvoluted spectra obtained from spectra illustrated in Figure 4.7 above. Phosphorylated VanS<sub>A</sub>Δ110 (VanS<sub>A</sub>Δ110-P) was observed at a mass 80 Da higher than that of VanS<sub>A</sub>Δ110. This suggested that, following transfer of the  $\gamma$ -phosphoryl group from ATP to the conserved histidine residue (H164), that ADP was released. This is in agreement with what is known about the mode of action of this protein.



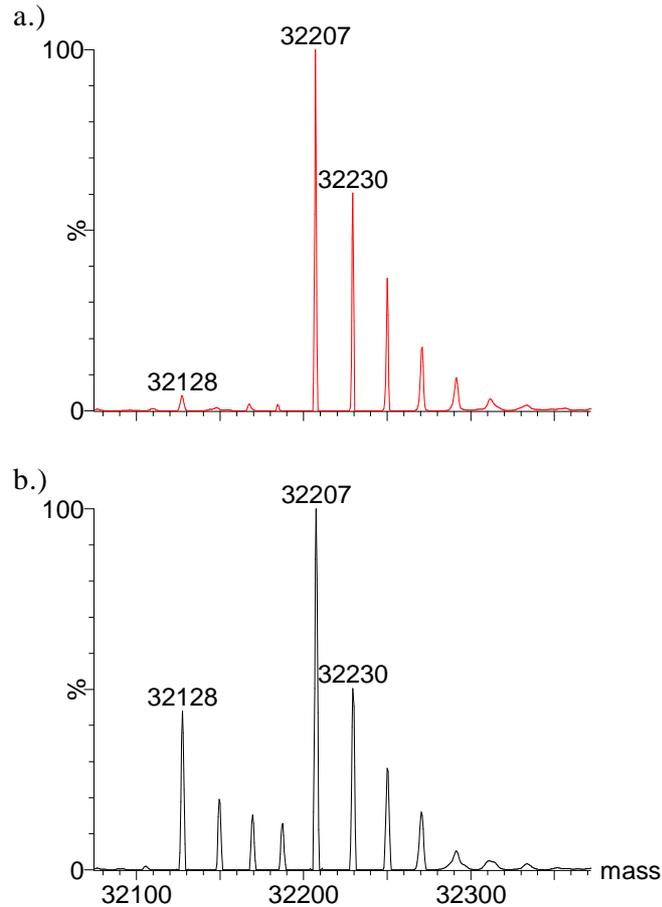
**Figure 4.8:** Deconvoluted mass spectra obtained for VanS<sub>A</sub>Δ110 samples allowed to phosphorylate for a.) 0, b.) 20, and c.) 60 minutes respectively. VanS<sub>A</sub>Δ110-P is observed at a mass of 32208 Da ( $\pm 1$  Da), 80 Da higher than that of VanS<sub>A</sub>Δ110 ( $32128 \pm 1$  Da). Additional peaks observed within spectra are consistent with those expected to be observed for sodiated protein species.

The ratios of VanS<sub>A</sub>Δ110 to VanS<sub>A</sub>Δ110-P, observed within spectra, were used to calculate the rate of autophosphorylation of VanS<sub>A</sub>Δ110. This is illustrated, for two different expression batches of this protein, in Figure 4.9. The phosphorylation assay was conducted in triplicate for the second batch of protein and phosphorylation levels calculated as an average. All data points were within  $\pm 5\%$  of average values obtained, illustrating the reproducibility of this approach.



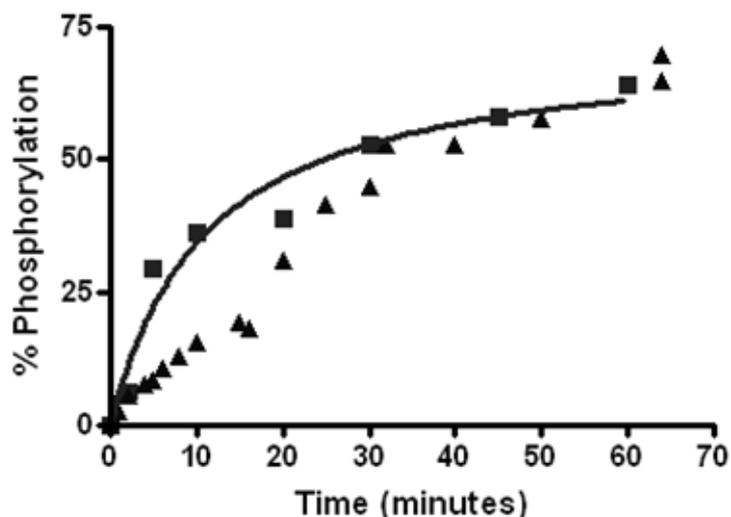
**Figure 4.9:** The percentage of total protein phosphorylated after different time intervals for protein batch one (■) and protein batch two (▲) respectively. The averages of phosphorylation levels calculated, from three datasets for protein batch two, are shown. Error bars indicate the maximum and minimum values calculated for time points with three replicates.

Approximately 67 % of VanS<sub>A</sub>Δ110 (batch one) was phosphorylated after phosphorylation had been allowed to proceed for 60 minutes but the second batch was somewhat less phosphorylated after the same time, suggesting that some of the protein was not functionally active. This was confirmed by allowing phosphorylation to take place overnight after which the batch one sample was virtually 100 % phosphorylated whereas the second batch was only 69 % phosphorylated (see Figure 4.10). This presumably accounts for the difference observed in phosphorylation levels after 60 minutes.



**Figure 4.10:** Deconvoluted mass spectra obtained for VanS<sub>A</sub>Δ110 samples a.) batch one and b.) batch two allowed to autophosphorylate overnight. VanS<sub>A</sub>Δ110-P is observed at a mass of 32207 Da ( $\pm 1$  Da), approximately 80 Da higher than that of VanS<sub>A</sub>Δ110 (32128  $\pm 1$  Da). Batch one is virtually 100 % phosphorylated whilst batch two exhibits only 69 % phosphorylation.

The autophosphorylation of VanS<sub>A</sub>Δ110 batch one over time was also measured by means of a radioactive labelling assay (Quigley and Roper, unpublished data), performed as detailed in Wright *et al.* (Pootoolal *et al.* 2002). A comparison of the results obtained by this assay with those obtained by the MS-based-assay can be seen in Figure 4.11 below.



**Figure 4.11:** The percentage of total protein phosphorylated after different time intervals for protein batch one, measured by MS (■) and by means of a radioactive labelling assay (▲) respectively.

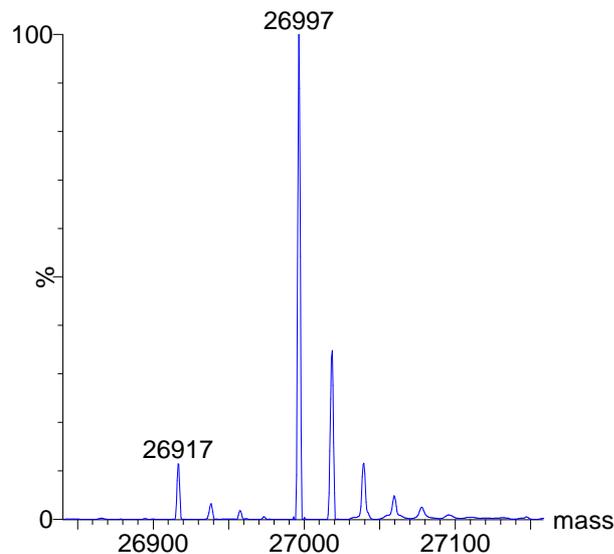
Good correlation between the levels of autophosphorylation measured by the two different methods at later time points is seen. Data collected by the two assays for the earlier time points show significant discrepancies. The conditions used in the radioactive labelling assay may have meant that this assay underestimated the phosphorylation levels at lower time points.

The radioactive labelling assay involves the incubation of VanS with magnesium chloride and  $[\gamma\text{-}^{32}\text{P}]\text{ATP}$  for different time periods.  $^{32}\text{P}$ -labelled  $\gamma$ -phosphate is transferred onto the VanS protein from  $[\gamma\text{-}^{32}\text{P}]\text{ATP}$  as autophosphorylation occurs. The reaction is stopped by the addition of SDS-PAGE loading buffer containing 25 mM EDTA solution. The EDTA solution stops further phosphorylation by chelating the  $\text{Mg}^{2+}$  ions bound to ATP. Once all the time points are complete, the samples are resolved by SDS-PAGE to separate protein from unbound  $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ . The amount of protein phosphorylation is determined by phosphoimaging the SDS-PAGE gel and calculating the levels of radioactively-labelled phosphorous incorporated. The loading buffer used to stop the reaction, which is at pH 6.8, may destabilise the phospho-histidine bond and cause the protein to dephosphorylate. If this is the case, phosphate loss would be expected to be more pronounced in samples with smaller reaction times as these samples would be exposed to the loading buffer for longer

time periods. This is what is observed. This may explain the differences in the experimental values obtained by the two assays at shorter reaction times.

A phosphorylation time course was also followed for VanS<sub>A</sub>Δ155 but it phosphorylated much more slowly than VanS<sub>A</sub>Δ110, and to a much lesser extent, with only 25 % of the protein becoming phosphorylated after 60 minutes. The amount of phosphorylation observed at small incubation times meant it was difficult to quantify accurately phosphorylation amounts and so only the phosphorylation level reached after 60 minutes is reported.

VanS<sub>A</sub>Δ155 reached 90 % phosphorylation after overnight incubation with ATP and magnesium chloride suggesting that the lower rate of phosphorylation observed was not due to the presence of significant amounts of functionally inactive protein (see Figure 4.12 below).



**Figure 4.12:** Deconvoluted mass spectrum obtained for VanS<sub>A</sub>Δ155 allowed to autophosphorylate overnight. VanS<sub>A</sub>Δ155-P is observed at a mass of 26997 Da ( $\pm 1$  Da), approximately 80 Da higher than that of VanS<sub>A</sub>Δ155 ( $26917 \pm 1$  Da ). Approximately 90 % of the protein is phosphorylated.

The difference in reaction rate observed between VanS<sub>A</sub>Δ110 and VanS<sub>A</sub>Δ155 truncates therefore indicates that the additional residues deleted in the smaller

truncate either play a role in phosphorylation or provide conformational stability. The additional residues may aid in forming an interaction between the CA and DHP domains, which is required for phosphorylation to occur.

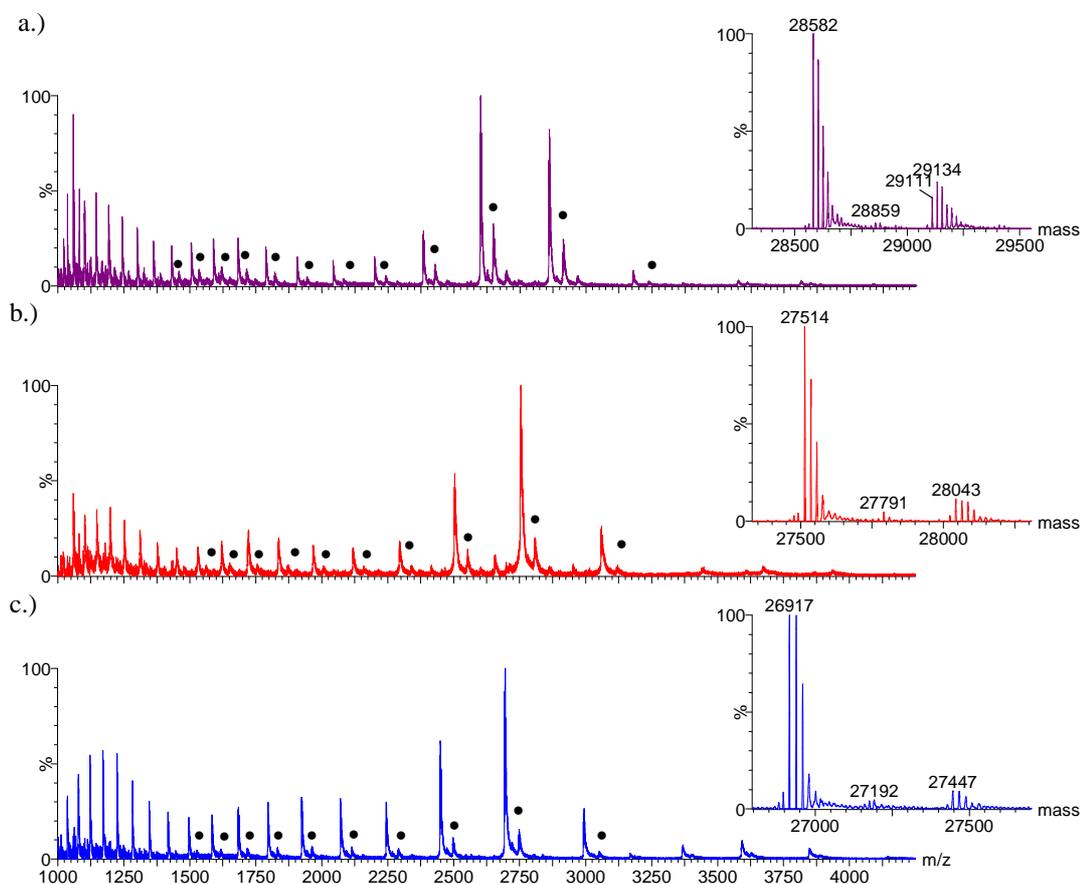
This work illustrates the use of an alternative approach for the measurement of autokinase activity, which could be used in the study of other systems. Although multiple methods based on mass spectrometry have been developed to monitor enzymatic conversions they have concentrated on the quantitation of low molecular weight products. Wind *et al.* (Wind *et al.* 2005) analysed histidine phosphorylation of another histidine kinase, CheA, by means of ESI-MS and element mass spectrometry. They quantified the degree of phosphorylation observed after *in vitro* phosphorylation for 60 minutes but did not conduct a phosphorylation time course. They did, however, elegantly illustrate the instability of histidine phosphorylation by following the rate of dephosphorylation occurring over time upon exposure of the sample to acidic conditions.

These results indicate that the use of this mass spectrometry-based approach to follow autokinase activity forms a valid alternative to more conventional biophysical methods. It provides reproducible results without requiring the use of radioactive material and can be completed within a shorter time period. The MS-based assay can be completed within one day whilst a radioactive labelling study takes two days. This work also suggests that the MS-based-assay can provide more accurate quantitation of phosphorylation at lower time points when compared to the radioactive labelling assay.

### **4.3.3 Conformational studies**

ESI-MS spectra acquired for VanS<sub>A</sub>Δ140, VanS<sub>A</sub>Δ150 and VanS<sub>A</sub>Δ155, following incubation with ATP and MgCl<sub>2</sub> for 1 minute at room temperature, exhibited additional peaks to those observed for protein alone (see Figure 4.13 in comparison to Figure 4.6). Deconvoluted spectra showed that these additional peaks related to the presence of a species 529 Da ( $\pm 1$  Da) higher in mass than each protein species. This mass difference indicated that ATP and magnesium were non-covalently associating with the protein (as magnesium-bound ATP has a molecular weight of

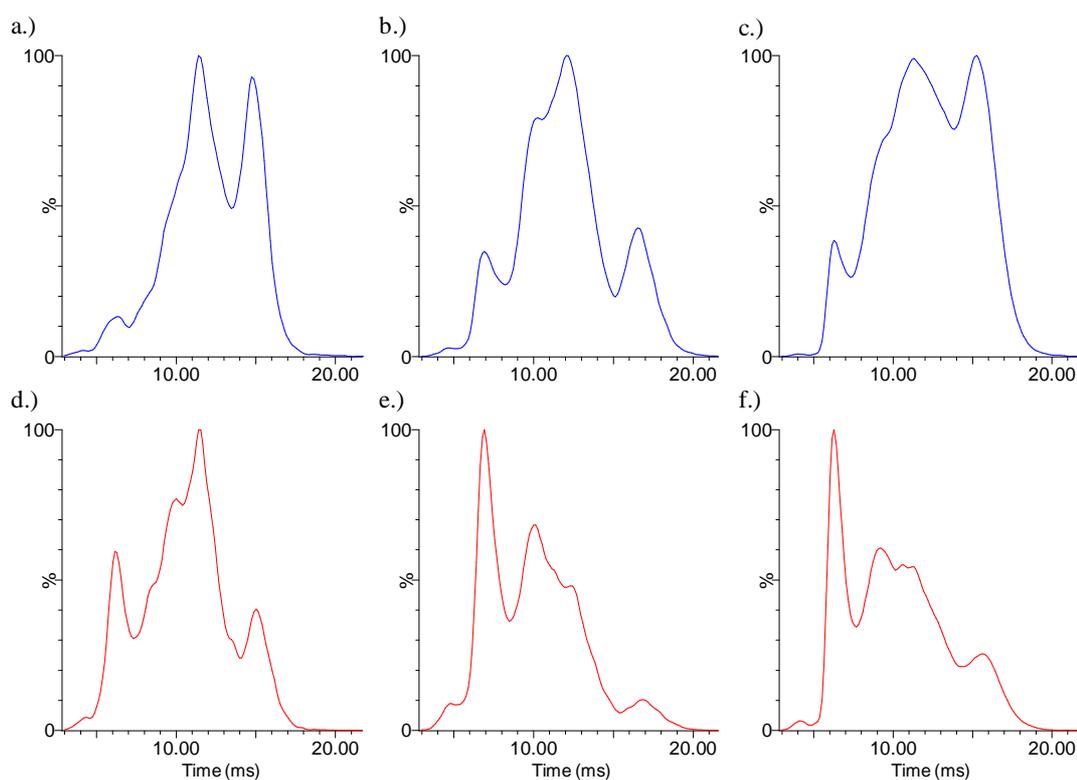
529.5 Da) and this association was being conserved in the gas phase. The spectra provided no evidence of the presence of phosphorylated species.



**Figure 4.13:** Comparison of ESI-MS spectra obtained for VanS<sub>A</sub> truncates a.) Δ140 b.) Δ150 c.) Δ155 at 10 μM in 10 mM ammonium bicarbonate pH 7.8 following incubation with ATP and MgCl<sub>2</sub> for 1 minute at room temperature with, inset, corresponding deconvoluted spectra. Additional peaks (•) are observed at a mass 529 Da ( $\pm 1$  Da) higher than those representing each protein species.

The observation of ATP-associated VanS<sub>A</sub> species in the gas phase provided an opportunity for an investigation of the three-dimensional structure of ATP-associated VanS<sub>A</sub> (VanS<sub>A</sub>-ATP) and VanS<sub>A</sub> to be carried out by means of TWIM-MS. ESI-spectra acquired for VanS<sub>A</sub>Δ110, following incubation with ATP and MgCl<sub>2</sub> solution did not show the presence of ATP-associated species at sufficiently high intensity for them to be investigated. It is likely that, VanS<sub>A</sub> Δ110-ATP did not persist long enough to be observed because, once ATP-association had occurred, phosphorylation and ADP release happened too quickly.

Extracted arrival time distributions (ATDs) for the most abundant charge states of VanS<sub>A</sub>Δ140, VanS<sub>A</sub>Δ150 and VanS<sub>A</sub>Δ155 and corresponding ATP-associated VanS<sub>A</sub> species observed in TWIM-MS experiments are shown in Figure 4.14.



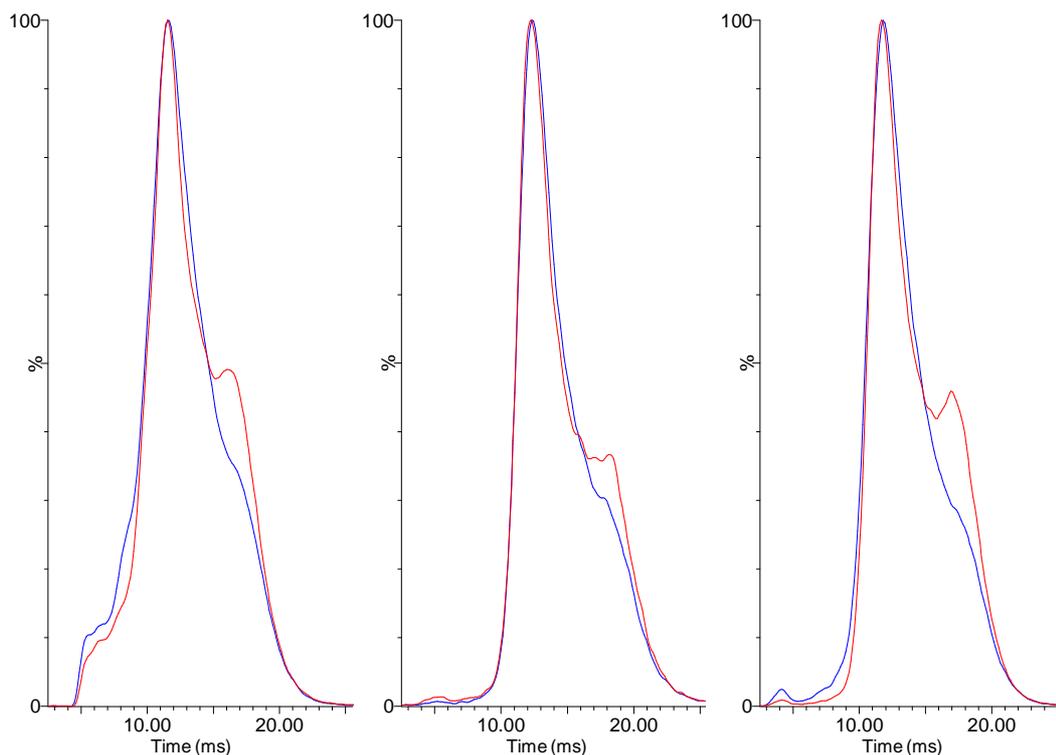
**Figure 4.14:** Extracted arrival time distributions for the most abundant charge state observed within spectra for a.) VanS<sub>A</sub>Δ140, b.) VanS<sub>A</sub>Δ150, c.) VanS<sub>A</sub>Δ155, d.) VanS<sub>A</sub>Δ140-ATP, e.) VanS<sub>A</sub>Δ150-ATP and f.) VanS<sub>A</sub>Δ155-ATP.

Each peak within an ATD for a particular ion is taken to represent at least one gas-phase conformation of that ion. Examination of the ATDs for the most abundant charge states of VanS<sub>A</sub>Δ140, VanS<sub>A</sub>Δ150 and VanS<sub>A</sub>Δ155 (Figure 4.14 above) suggest that these constructs exist in multiple gas-phase conformations. Peaks observed within the ATDs are not resolved and whilst at least three maxima are observed within each ATD it appears that these constructs exhibit many more gas-phase conformations that cannot be distinguished by this technique.

The ATDs obtained for VanS<sub>A</sub>-ATP species and for VanS<sub>A</sub> species show a number of interesting features. The apexes of peaks observed in the ATDs for VanS<sub>A</sub> species and corresponding ATP-associated VanS<sub>A</sub> species occur at similar times but there are stark differences in the relative intensities of the peaks observed. This suggests that whilst both VanS<sub>A</sub>-ATP species and corresponding VanS<sub>A</sub> species exist in analogous gas-phase conformations, they adopt these conformations to different extents. When ATP is bound VanS<sub>A</sub> species appear to favour conformations of higher mobility, i.e. more compact conformations.

Ion mobility experiments were conducted on VanS<sub>A</sub>Δ110 and VanS<sub>A</sub>Δ110-P. As stated above, VanS<sub>A</sub>Δ110 was not observed in complex with ATP and magnesium in sufficient intensity in spectra for conclusions to be drawn about its conformational distribution. Extracted ATDs for the most abundant charge states of VanS<sub>A</sub>Δ110 and VanS<sub>A</sub>Δ110-P observed in spectra obtained from TWIM-MS experiments are shown in Figure 4.15. ATDs from three datasets, which were obtained at intervals of several months, are illustrated. Extracted ATDs for VanS<sub>A</sub>Δ110 charge state [M+12H]<sup>12+</sup> contain two maxima indicating that this protein construct exhibits at least two gas-phase conformations. ATDs for VanS<sub>A</sub>Δ110-P charge state [M+12H]<sup>12+</sup> show primarily one peak. On comparison of ATDs for VanS<sub>A</sub>Δ110-P with those for VanS<sub>A</sub>Δ110 it can be seen that they exist mainly as conformations having similar mobilities.

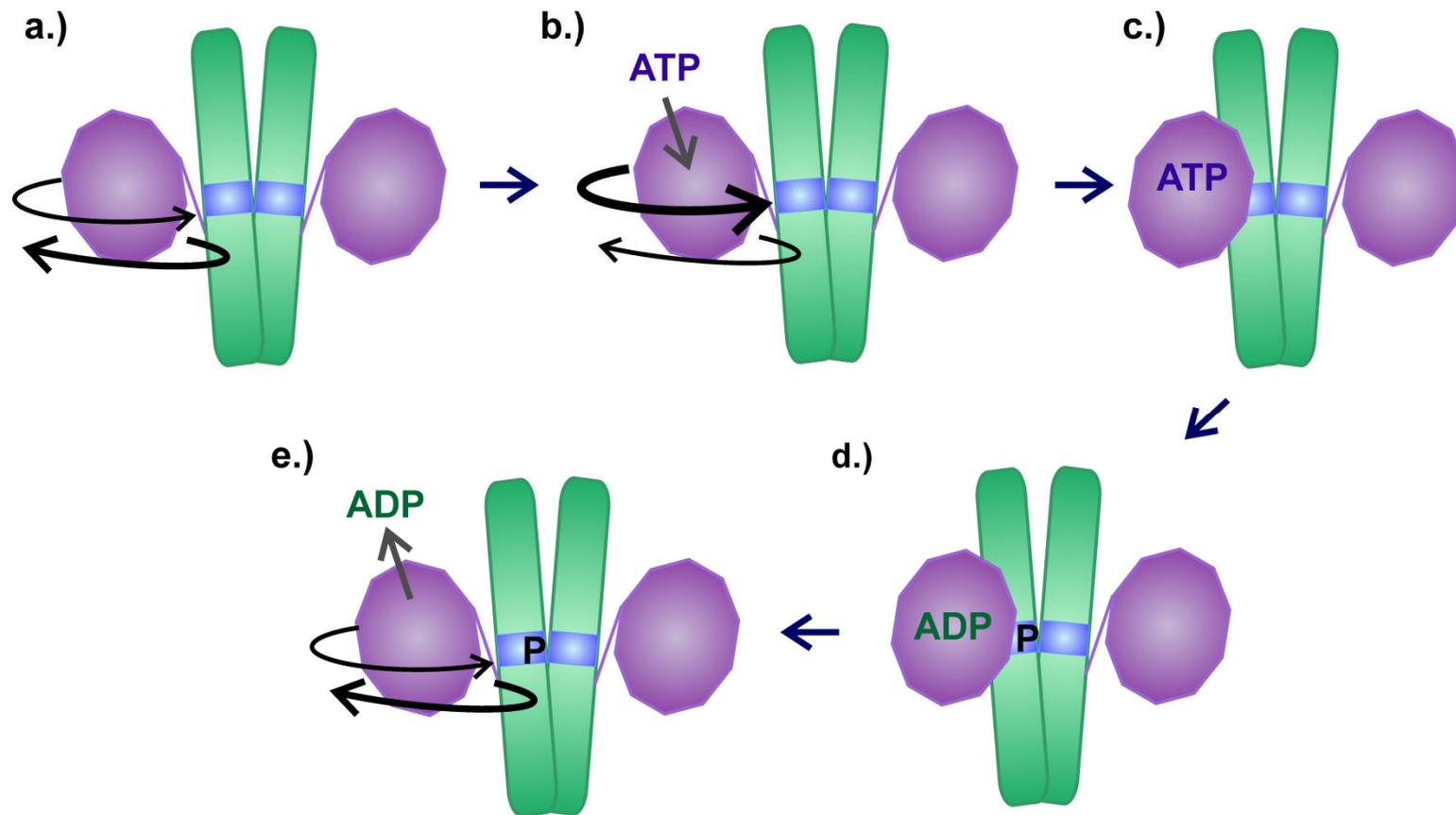
The maxima of each of the main peaks observed in the ATDs extracted from the three data sets are within ±0.36 ms of each other. The three datasets were obtained at intervals of several months on the same batch of protein. Caution has to be taken when comparing the ATDs across the three datasets as some protein degradation would be expected to occur over this time frame. This being said, the reproducibility of the data, over this period, gives confidence in the reliability and robustness of the method.



**Figure 4.15:** Extracted arrival time distributions for the most abundant charge state ( $[M+12H]^{12+}$ ) for VanS<sub>A</sub>Δ110 (red) and VanS<sub>A</sub>Δ110-P (blue) observed within spectra obtained from TWIM-MS experiments for three different datasets.

#### 4.3.4 Model of VanS<sub>A</sub> autokinase activity

The results from the ion mobility experiment conducted on non-phosphorylated, phosphorylated and ATP-bound forms of VanS<sub>A</sub> have been used, along with prior knowledge, to suggest a mechanism for VanS<sub>A</sub> autokinase activity. This is depicted in Figure 4.16 and is similar to that suggested for HK853 by Marina *et al.* (Marina *et al.* 2005).



**Figure 4.16:** Pictorial representation of VanS<sub>A</sub> autokinase activity. The catalytic ATP-binding domain is represented in purple whilst the dimerisation phosphotransfer domain is represented in blue. a.) The kinase domain is free to move around the phosphotransfer domain, hinging around the linker region between the two domains. b.) Upon ATP-binding, VanS<sub>A</sub> favours a more compact conformation. c.) The kinase domain comes into close proximity to the phosphotransfer domain. d.) This allows for transfer of phosphate from the associated ATP molecule to VanS<sub>A</sub> (H164). e.) Once phosphorylated, VanS<sub>A</sub> no longer favours a more compact conformation. The kinase domain is free to move. This allows for ADP release.

The CA domain is proposed to be free to rotate around the DHp domain, hinging around the linker region between the two domains. This results in an array of conformational states being exhibited by the protein. Upon ATP binding, the most compact of the multiple conformations observed is favoured i.e. more protein preferentially occupies this state than any other. This more closed conformation likely permits the interaction between the ATP molecule and the site of phosphorylation, so phosphorylation can occur. No significant differences between the conformations occupied between phosphorylated and non-phosphorylated protein are observed and, therefore, after phosphorylation, the most compact conformation is no longer favoured. It is likely that the ATP-binding domain moves out to allow ADP release. In the cell, this would then permit the interaction between the phosphotransfer domain of VanS<sub>A</sub> and the response regulator, VanR<sub>A</sub>.

This model is in agreement with others which have been suggested for different HKs (Marina *et al.* 2005; Bick *et al.* 2009) but these results give a further indication of protein dynamics. VanS<sub>A</sub> exists in similar conformational states whether ATP is associated or not; ATP-binding is not likely to cause a conformational transition. It is proposed that the bound ATP molecule interacts with the DHp domain when they are sufficiently close in a conformational state.

## 4.4 Conclusions

Ion mobility mass spectrometry has been used to provide information relating to the conformational states of the VanS<sub>A</sub> protein core under non-phosphorylated, ATP-associated and phosphorylated conditions. MS has been used to confirm that VanS<sub>A</sub> is able to autophosphorylate in the presence of ATP and MgCl<sub>2</sub>. The rate at which phosphorylation proceeds and the amount of phosphorylation observed has been studied.

This work confirms that ATP association leads to protein phosphorylation and ADP release. Ion mobility experiments provide data enabling conformational changes occurring during VanS<sub>A</sub> autophosphorylation to be observed. When in complex with magnesium and ATP, which are required for phosphorylation to occur, VanS<sub>A</sub> appeared to favour a more compact conformation. This conformational state is proposed to facilitate autokinase activity by bringing the ATP molecule and the site of phosphorylation into sufficient proximity for phospho-transfer to occur.

For all VanS<sub>A</sub> truncates studied, ion mobility experiments illustrated the presence of multiple protein conformations. They indicate that, in the gas phase, VanS<sub>A</sub>Δ155 exists in a larger number of stable conformations than VanS<sub>A</sub>Δ110. This is supported by ESI-MS spectra obtained for these two species. The charge-state distribution observed in the spectrum for VanS<sub>A</sub>Δ110 contains fewer maxima than the charge-state distribution observed in the spectrum for VanS<sub>A</sub>Δ155. The presence of multiple conformations, in solution, provides an explanation as to why histidine kinases have proved so difficult to crystallise. Mass spectrometry could be used as a tool to help determine protein constructs which would crystallise in this case. Mass spectrometry has previously been shown to be a valuable tool for the protein crystallographer as it can be used to verify correct protein expression (Chait 1994), elucidate protein domains (Cohen 1996) and analyse protein crystals (Cohen and Chait 2001).

The use of mass spectrometry to study phosphorylation rates is a valuable technique, which offers a simple, reproducible alternative to conventional biophysical methods.

## 4.5 References

- Arthur, M., Molinas, C. and Courvalin, P.** (1992a).The VanS-VanR two-component regulatory system controls synthesis of depsipeptide peptidoglycan precursors in *Enterococcus faecium* BM4147. *Journal of Bacteriology*. **174**, 2582-2591.
- Arthur, M., Molinas, C. and Courvalin, P.** (1992b).The VanS-VanR two-component regulatory system controls synthesis of depsipeptide peptidoglycan precursors in *Enterococcus faecium* BM4147. *J. Bacteriol.* **174**, 2582-2591.
- Bick, M. J., Lamour, V., Rajashankar, K. R., Gordiyenko, Y., Robinson, C. V. and Darst, S. A.** (2009).How to Switch Off a Histidine Kinase: Crystal Structure of *Geobacillus stearothermophilus* KinB with the inhibitor Sda. *Journal of Molecular Biology*. **386**, 163-177.
- Bugg, T. D. H., Wright, G. D., Dutka-Malen, S., Arthur, M., Courvalin, P. and Walsh, C. T.** (1991).Molecular basis for vancomycin resistance in *Enterococcus faecium* BM4147: biosynthesis of a depsipeptide peptidoglycan precursor by vancomycin resistance proteins VanH and VanA. *Biochemistry*. **30**, 10408-10415.
- Chait, B. T.** (1994).Mass spectrometry -- a useful tool for the protein X-ray crystallographer and NMR spectroscopist. *Structure*. **2**, 465-467.
- Cohen, S. L.** (1996).Domain elucidation by mass spectrometry. *Structure*. **4**, 1013-1016.
- Cohen, S. L. and Chait, B. T.** (2001).Mass Spectrometry as a Tool for Protein Crystallography. *Annual Review of Biophysics and Biomolecular Structure*. **30**, 67-85.
- Courvalin, P.** (2006).Vancomycin Resistance in Gram-Positive Cocci. *Clinical Infectious Diseases*. **42**, S25-S34.
- Depardieu, F., Courvalin, P. and Msadek, T.** (2003).A six amino acid deletion, partially overlapping the VanS<sub>B</sub> G2 ATP-binding motif, leads to constitutive glycopeptide resistance in VanB-type *Enterococcus faecium*. *Molecular Microbiology*. **50**, 1069-1083.
- Ferrige, A. G., Seddon, M. J., Green, B. N., Jarvis, S. A., Skilling, J. and Staunton, J.** (1992).Disentangling electrospray spectra with maximum entropy. *Rapid Communications in Mass Spectrometry*. **6**, 707-711.
- Gao, R. and Stock, A. M.** (2009).Biological Insights from Structures of Two-Component Proteins. *Annual Review of Microbiology*. **63**, 133-154.
- Ge, X., Sirich, T. L., Beyer, M. K., Desaire, H. and Leary, J. A.** (2001).A Strategy for the Determination of Enzyme Kinetics Using Electrospray Ionization with an Ion Trap Mass Spectrometer. *Analytical Chemistry*. **73**, 5078-5082.

**Healy, V. L., Lessard, I. A. D., Roper, D. I., Knox, J. R. and Walsh, C. T.** (2000). Vancomycin resistance in enterococci: reprogramming of the -Ala--Ala ligases in bacterial peptidoglycan biosynthesis. *Chemistry & Biology*. **7**, R109-R119.

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

**Holman, T. R., Wu, Z., Wanner, B. L. and Walsh, C. T.** (1994). Identification of the DNA-binding site for the phosphorylated VanR protein required for vancomycin resistance in *Enterococcus faecium*. *Biochemistry*. **33**, 4625-4631.

**Hong, H.-J., Hutchings, M. I. and Buttner, M. J.** (2008). Vancomycin Resistance VanS/VanR Two-Component Systems. *Bacterial Signal Transduction: Network and Drug Targets*, Landes Bioscience.

**Leary, J. A., Schenauer, M. R., Stefanescu, R., Andaya, A., Ruotolo, B. T., Robinson, C. V., Thalassinou, K., Scrivens, J. H., Sokabe, M. and Hershey, J. W. B.** (2009). Methodology for Measuring Conformation of Solvent-Disrupted Protein Subunits using T-WAVE Ion Mobility MS: An Investigation into Eukaryotic Initiation Factors. *Journal of the American Society for Mass Spectrometry*. **20**, 1699-1706.

**Leclercq, R., Derlot, E., Duval, J. and Courvalin, P.** (1988). Plasmid-mediated vancomycin and teicoplanin resistance in *Enterococcus faecium*. *N. Engl. J. Med.* **319**, 157-161.

**Liesener, A. and Karst, U.** (2005). Monitoring enzymatic conversions by mass spectrometry: a critical review. *Analytical and Bioanalytical Chemistry*. **382**, 1451-1464.

**Marina, A., Waldburger, C. D. and Hendrickson, W. A.** (2005). Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. *The European Molecular Biology Organization Journal*. **24**, 4247-4259.

**Norris, A. J., Whitelegge, J. P., Faull, K. F. and Toyokuni, T.** (2001). Analysis of Enzyme Kinetics Using Electrospray Ionization Mass Spectrometry and Multiple Reaction Monitoring: Fucosyltransferase V. *Biochemistry*. **40**, 3774-3779.

**Pootoolal, J., Neu, J. and Wright, G. D.** (2002). Glycopeptide Antibiotic Resistance. *Annual Review of Pharmacology and Toxicology*. **42**, 381-408.

**Robinson, C. V., Chung, E. W., Kragelund, B. B., Knudsen, J., Aplin, R. T., Poulsen, F. M. and Dobson, C. M.** (1996). Probing the Nature of Noncovalent Interactions by Mass Spectrometry. A Study of Protein CoA Ligand Binding and Assembly. *Journal of the American Chemical Society*. **118**, 8646-8653.

**Saidijam, M., Bettaney, K. E., Szakonyi, G., Psakis, G., Shibayama, K., Suzuki, S., Clough, J. L., Blessie, V., Abu-Bakr, A., Baumberg, S., Mueller, J., Hoyle, C.**

**K., Palmer, S. L., Butaye, P., Walravens, K., Patching, S. G., O'Reilly, J., Rutherford, N. G., Bill, R. M., Roper, D. I., Phillips-Jones, M. K. and Henderson, P. J. F.** (2005). Active membrane transport and receptor proteins from bacteria. *Biochem. Soc. Trans.* **33**, 867-872.

**Steinkamp, T., Liesener, A. and Karst, U.** (2004). Reaction monitoring of enzyme-catalyzed ester cleavage by time-resolved fluorescence and electrospray mass spectrometry: method development and comparison. *Analytical and Bioanalytical Chemistry.* **378**, 1124-1128.

**Stock, A. M., Robinson, V. L. and Goudreau, P. N.** (2000). Two-Component Signal Transduction. *Annual Review of Biochemistry.* **69**, 183-215.

**Vollmer, W., Blanot, D. and de Pedro, M., A.** (2008). Peptidoglycan structure and architecture. *FEMS Microbiology Reviews.* **32**, 149-167.

**Walsh, C. T., Fisher, S. L., Park, I. S., Prahalad, M. and Wu, Z.** (1996). Bacterial resistance to vancomycin: five genes and one missing hydrogen bond tell the story. *Chemistry & Biology.* **3**, 21-28.

**Wind, M., Wegener, A., Kellner, R. and Lehmann, W. D.** (2005). Analysis of CheA Histidine Phosphorylation and Its Influence on Protein Stability by High-Resolution Element and Electrospray Mass Spectrometry. *Analytical Chemistry.* **77**, 1957-1962.

**Wright, G. D., Holman, T. R. and Walsh, C. T.** (1993). Purification and characterization of VanR and the cytosolic domain of VanS: A two-component regulatory system required for vancomycin resistance in *Enterococcus faecium* BM4147. *Biochemistry.* **32**, 5057-5063.

**Wu, Q., Gao, J., Joseph-McCarthy, D., Sigal, G. B., Bruce, J. E., Whitesides, G. M. and Smith, R. D.** (1997). Carbonic Anhydrase-Inhibitor Binding: From Solution to the Gas Phase. *Journal of the American Chemical Society.* **119**, 1157-1158.

# Chapter Five: Diagnosis of Hemoglobinopathies

---

## 5.1 Introduction

### 5.1.1 Hemoglobin

Hemoglobin (Hb) is the oxygen-transport metalloprotein found in the red blood cells of all vertebrates; it binds oxygen in the lungs and supplies it to the organs and tissues. It is tetrameric in structure, consisting of four globular protein subunits: two alpha ( $\alpha$ ) chains and two non- $\alpha$  chains, each associated with a prosthetic heme group. The  $\alpha$ -chains are encoded on chromosome 16 by two closely related genes,  $\alpha 1$  and  $\alpha 2$ . The non- $\alpha$  chains, beta ( $\beta$ ), delta ( $\delta$ ) and gamma ( $\gamma$ ), are encoded by a cluster of genes on chromosome 11 (Hartwell *et al.* 2005). A diploid cell therefore has four  $\alpha$ -globin genes and two  $\beta$ -like genes (Clarke and Higgins 2000). The type of non- $\alpha$  chain within the tetramer is used to define the hemoglobin type. In the predominant form of adult hemoglobin, Hb A, the non- $\alpha$  chains are  $\beta$ -chains, in the minor form, Hb A<sub>2</sub> (2-3.5 %), these are  $\delta$ -chains. In a fetus, the majority of hemoglobin is HbF ( $\alpha_2\gamma_2$ ). Hb F has increased oxygen affinity relative to that of adult hemoglobin. This allows oxygen to be effectively transferred from maternal to fetal cells, across the placenta. The  $\alpha$ - and  $\beta$ -chains consist of 141 and 146 amino acid residues, respectively. There is some sequence homology between the two chains (64 individual amino acid residues in identical positions), and the  $\beta$ -chain differs from the  $\delta$ - and  $\gamma$ -chains by 10 and 39 residues, respectively (Clarke and Higgins 2000).

### 5.1.2 Hemoglobin disorders

Hemoglobin disorders may be characterised into structural variants and thalassemiias. Structural variants are produced when the amino acid composition of a protein chain is altered. A thalassaemia syndrome is caused when a mutation or deletion event leads to diminished or no production of one of the globin chains of the hemoglobin molecule. Mutations leading to thalassaemia syndromes mostly occur within the non-coding regions of the globin genes. Over 1000 structural variants and 300 mutations responsible for the thalassaemia syndromes have been detected so far. Details of these are housed in HbVar, a relational database of human hemoglobin variants and thalassaemia mutations, which may be found at the Globin Gene Server (Hardison *et*

al. 2002). The majority of the structural variants and thalassemia syndromes identified are clinically silent, creating no ill effects in the carrier, but a small number are of significant clinical importance (Clarke and Higgins 2000).

Hemoglobin variants are often named after their place of first discovery or family name of the primary case. A systematic nomenclature is used to characterise structural variants. Thus Hb S ( $\beta 6 \text{ Glu} \rightarrow \text{Val}$ ) describes that Hb S is produced by an amino acid mutation in the  $\beta$ -chain at position six from a glutamic acid to a valine. The common clinically significant structural variants are Hb S, C, D<sup>Punjab</sup>, E and O<sup>Arab</sup>, caused by single amino acid substitutions in the  $\beta$ -chain, and the Lepores, which are  $\delta:\beta$  chain hybrids. The mutations responsible for these variants and their clinical manifestations are summarised in Table 5.1 below.

**Table 5.1:** The common clinically significant structural variants, the mutations which cause them and their clinical manifestations (data obtained from HbVar). Mass change shown is the change in globin chain mass produced as a result of the mutation.

Hemoglobin	Mutation	Mass Change (Da)	Clinical manifestation (Heterozygote/ Homozygote)
S	$\beta 6 \text{ (Glu} \rightarrow \text{Val)}$	-30	Carrier, no symptoms/ Sickle cell disease
C	$\beta 6 \text{ (Glu} \rightarrow \text{Lys)}$	-1	Carrier, no symptoms/ Mild anaemia
D <sup>Punjab</sup>	$\beta 121 \text{ (Glu} \rightarrow \text{Gln)}$	-1	Carrier, no symptoms/ Mild anaemia
E	$\beta 26 \text{ (Glu} \rightarrow \text{Lys)}$	-1	Mild microcytosis/ Thalassemia minor
O <sup>Arab</sup>	$\beta 121 \text{ (Glu} \rightarrow \text{Lys)}$	-1	Carrier, no symptoms/ Mild anaemia
Lepore-Hollandia	Hybrid: $\delta 1:22, \delta/\beta 23:49, \beta 50:146$	-30	Thalassemia minor
Lepore-Baltimore	Hybrid: $\delta 1:50, \delta/\beta 51:85, \beta 86:146$	-45	Thalassemia minor
Lepore-Boston-Washington	Hybrid: $\delta 1:87, \delta/\beta 88:116, \beta 116:146$	-2	Thalassemia minor

The inheritance of homozygosity for Hb S (Hb SS, sickle cell anaemia) or Hb S with Hb C (Hb SC disease), Hb D<sup>Punjab</sup> (Hb SD disease), Hb E (Hb SE disease) or Hb O<sup>Arab</sup> (Hb SO disease) result in sickling disorders. The sickling disorders are so

named because the abnormal hemoglobins they produce precipitate and polymerise in red blood cells causing the cells to become sickle in shape (Clarke and Higgins 2000). The sickling disorders are associated with severe anaemia and life-threatening vaso-occlusive crises. Anaemia (meaning lack of blood) leads to poor oxygen supply to the organs and tissues. Sickle cell crises result when the sickled red blood cells obstruct blood vessels and restrict blood supply to the organs, causing severe pain (NHS Sickle Cell and Thalassaemia Screening Programme 2009). Numerous complications can result as a consequence of these afflictions and sickle cell sufferers require life-long treatment. Heterozygotes, possessing one normal and one sickle  $\beta$ -globin gene, have sickle cell trait but are symptom free.

The thalassemys produce a syndrome characterised by anaemia, the severity of which depends on the type of thalassaemia syndrome present. The most severe thalassemys,  $\alpha$ - and  $\beta$ -thalassaemia major, result from homozygous genetic defects in the  $\alpha$ -globin and  $\beta$ -globin synthesis, respectively (Hartwell *et al.* 2005).

There are four  $\alpha$ -globin genes and therefore a single  $\alpha$ -globin gene deletion ( $-\alpha/\alpha$ ) gives  $\alpha$ -thal-2 trait and has no effect on hemoglobin synthesis. A double deletion ( $--/\alpha\alpha$  or  $-\alpha/-\alpha$ ) results in mild microcytic hypochromic anaemia (production of small red blood cells with low levels of hemoglobin per cell) (Clarke and Higgins 2000). If three  $\alpha$ -globin genes are deleted Hb H disease is produced. The excess production of  $\beta$ -chains leads to the formation of  $\beta_4$  tetramers. These precipitate in red cell precursors leading to ineffective red cell production and enlargement of the spleen and liver (Clarke and Higgins 2000). If all four  $\alpha$ -globin genes are deleted (homozygous  $\alpha$ -thal-1) Hb Bart's hydrops fetalis is caused which is incompatible with postnatal life.

Single  $\beta$ -globin gene deletion or diminished synthesis of one  $\beta$ -globin chain results in  $\beta$ -thalassaemia minor. This disorder presents with mild microcytic hypochromic anaemia (Clarke and Higgins 2000). Sufferers of this disorder produce more Hb A<sub>2</sub> than in normal individuals. Quantification of Hb A<sub>2</sub> levels above 3.5 %, along with the recording of a low mean cell hemoglobin level (as part of a full blood count), is used to provide diagnosis of heterozygous  $\beta$ -thalassaemia (Ryan *et al.* 2010). Hb E and Hb Lepore have thalassaemic manifestations that lead to decreased  $\beta$ -globin

production and microcytic hypochromic anaemia in homozygotes. Homozygous  $\beta$ -thalassemia sufferers can present with  $\beta$ -thalassemia intermedia or major. This depends on whether the thalassaemia results in greatly diminished ( $\beta^+$ ) or no  $\beta$ -chain production ( $\beta^0$ ).  $\beta$ -thalassemia major sufferers present with severe microcytic hypochromic anaemia and rely on life-long regular blood transfusions. About fifty-percent of  $\beta$ -thalassemia major sufferers, in the UK, die before the age of 35 (Modell *et al.* 2000).

As well as these common clinically significant hemoglobinopathies many more of the structural variants reported so far are of clinical relevance. A simple search of the HbVar database shows that there are 32 structural  $\alpha$ -chain variants and 88 structural  $\beta$ -chain variants which lead to anaemia in heterozygotes (Hardison *et al.* 2002).

### **5.1.3 Incidence of hemoglobinopathies**

Worldwide, in 1998, there was estimated to be 269 million carriers of hemoglobin disorders (Angastiniotis and Modell 1998). Hemoglobinopathies are a significant healthcare problem and the most common inherited disorders. Over 300,000 babies with major hemoglobin disorders are born every year and the majority die undiagnosed, untreated or under-treated (Angastiniotis and Modell 1998). The incidence of hemoglobinopathies is also on the increase; hemoglobinopathies such as Hb S, C and E have spread through migration from their native areas in the Mediterranean, Africa and Asia and are now endemic throughout Europe, the Americas and Australia (Angastiniotis and Modell 1998). Some variants have maintained a significant prevalence in certain areas because they provide some evolutionary advantage. Hb S is maintained at approximately 10 % frequency in malaria-endemic regions (Flint *et al.* 1998). Whilst Hb S homozygotes have sickle disease, heterozygotes have a 10-fold reduced risk of severe malaria (Ackerman *et al.* 2005). A similar protection is thought to be provided by E trait and  $\beta$ -thalassemia.

In the UK, the highest disease burden relates to sickling disorders in black ethnic groups (Angastiniotis and Modell 1998). In England, approximately 3000 babies (0.47 %) with sickle cell trait, 140-175 with sickle cell disease, 2800 (0.44 %) with

$\beta$ -thalassemia trait and 10-25 with  $\beta$ -thalassemia major/intermedia are born annually (Hickman *et al.* 1999).

#### **5.1.4 Hemoglobinopathy screening and diagnosis**

Highly sensitive detection and quick and accurate identification of hemoglobinopathies is necessary to allow early initiation of treatment in affected individuals and to uncover the presence of any inheritance issues. The majority of countries in the developed world recognise this need and have their own hemoglobinopathy screening programmes. In the UK, this is the NHS Sickle Cell and Thalassaemia Screening Programme. This consists of two linked screening programmes, one for antenatal screening and one for newborn screening (NHS Sickle Cell and Thalassaemia Screening Programme 2009).

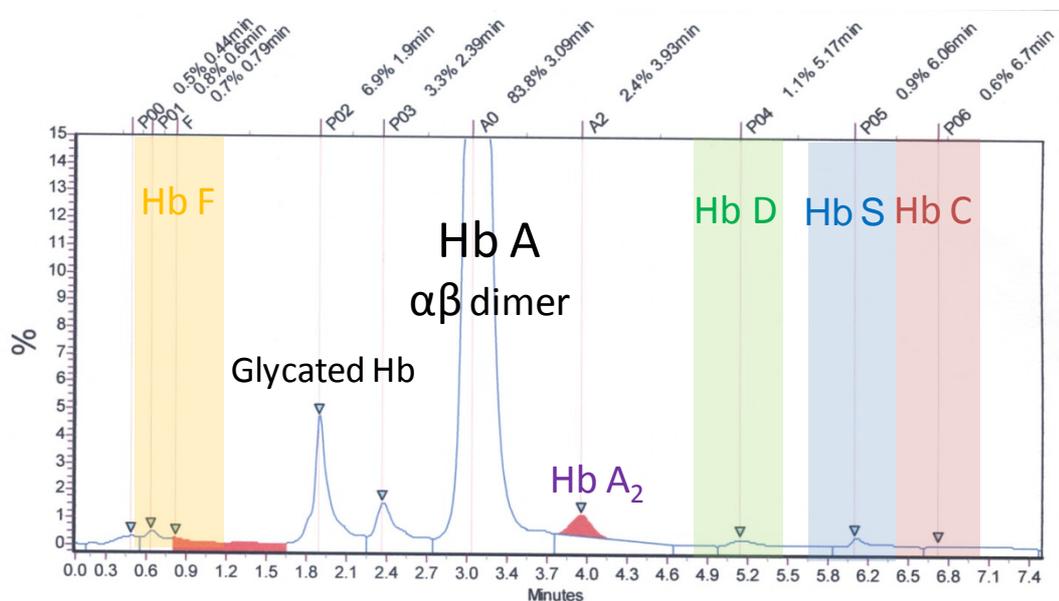
The aim of the antenatal screening is to offer sickle cell and thalassaemia screening to all pregnant women. In England, screening for sickle cell, thalassaemia and the other common clinically significant hemoglobin variants, C, D-Punjab, E, O-Arab, H and Lepore is offered to all women early in pregnancy in high prevalence areas. High prevalence areas are those which have sickle cell incidence in newborns of 1.5 per 10,000 pregnancies or greater. In low prevalence areas, pregnant women are first given a 'full blood count' blood test and the results of this, together with information regarding their families' origins, are used to decide whether a full screen needs to be performed (NHS Sickle Cell and Thalassaemia Screening Programme 2009).

Newborn babies are offered screening for sickle cell disease and  $\beta$ -thalassemia major as part of the routine 'heel prick' or 'blood spot' test. Men and women can ask for a test before they become parents, so they can know if they are carriers of a clinically significant hemoglobinopathy or thalassaemia and can evaluate the risk to any children they may have.

There are a number of techniques that can be used for screening of hemoglobinopathies but the most commonly used and recommended techniques are high-performance liquid chromatography (HPLC) and isoelectric focussing (IEF) (Ryan *et al.* 2010). The requirements of the antenatal screening programme within

England is that a screening method detects the presence of Hb A, S, C, D-Punjab, E, O-Arab, H and Lepore and can provide quantitation of Hb F and Hb A<sub>2</sub> for determination of  $\beta$ -thalassemia trait (Ryan *et al.* 2010). The neonatal screen must indicate the presence of sickle cell and/or  $\beta$ -thalassemia major.

HPLC separates variants based on their different retention times on a chromatography column. Several commercial automated HPLC systems have been developed which provide separation of the clinically important hemoglobin variants as well as Hb A<sub>2</sub> and Hb F quantitation. These systems require relatively low capital investment but have high running costs. An example of a HPLC trace produced from the analysis of a normal blood sample is shown below in Figure 5.1. Peaks observed within the chromatogram represent the elution of hemoglobin heterodimers ( $\alpha\beta$  dimers). Peaks showing the elution of Hb A and Hb A<sub>2</sub> are indicated. The retention windows within which Hb F, D, S and C elute on the HLC-723 HbG7 analyser (Tosoh Bioscience Ltd., Redditch, UK) (Chevenne *et al.* 1999) are shown. A single hemoglobinopathy screen by this approach takes less than eight minutes.



**Figure 5.1:** HPLC trace from TOSOH HLC-723 HbG7 analyser for blood sample from normal adult patient. Peaks shown represent the elution of heterodimers ( $\alpha\beta$  dimers). Retention time windows for Hb F, A, A<sub>2</sub>, D, S and C are shown.

HPLC provides sufficiently accurate quantification of Hb A<sub>2</sub> to be used for the diagnosis of  $\beta$ -thalassemia trait (Ryan *et al.* 2010). Some variants co-elute with Hb A, S or other clinically significant variants which means that the presence of a variant could be missed or misinterpreted. HPLC also provides separation of glycosylated and other derivative forms of hemoglobin, which can make interpretation more difficult. Careful examination of every chromatogram is therefore essential and any identification of variants made is provisional.

Glycated Hb (Hb A<sub>1c</sub>) levels have been used over the last 30 years to assess whether a sufferer of diabetes has their diabetes under good control. Hb A<sub>1c</sub> testing has also been shown to help predict the likelihood that patients will develop diabetes in the future (Edelman *et al.* 2004).

IEF separates variants based on their isoelectric point, the position along a pH gradient where they have no net charge (Hartwell *et al.* 2005). IEF can provide resolution of Hb F from Hb A and Hb S, C, D, O and G (Clarke and Higgins 2000). IEF is not a validated technique for Hb A<sub>2</sub> quantitation.

If the presence of a clinically important variant is suggested by HPLC or IEF analysis then another technique is applied to confirm this such as IEF or HPLC as appropriate or cellulose acetate electrophoresis (CAE) and/or agarose gel electrophoresis (AGE). CAE provides separation of Hb A and F but cannot separate Hb C from Hb E, A<sub>2</sub> and O or Hb S from D-Punjab. AGE provides separation of Hb F, S and C but cannot separate Hb A from Hb E, A<sub>2</sub> and D-Punjab (Bain 2006).

Results of these analyses are used in conjunction with those from a full blood count to provide presumptive diagnosis. To obtain definitive diagnosis, DNA or protein sequence analysis is required but is not currently a requirement under the screening programme as positive presumptive identification by two approaches is considered sufficient to provide diagnosis. Samples may be referred to DNA reference laboratories or to mass spectrometry laboratories in some cases, however, when a thalassemia syndrome or variant is clearly present but when its identification cannot be presumptively confirmed.

A DNA approach to screening would provide excellent sensitivity but is not widely available, has considerable reagent costs and is unable to detect the presence of post-translational modifications (Wild *et al.* 2001). The use of mass spectrometry to identify variants is becoming more widespread and may have the potential to be used as a screening method (Ryan *et al.* 2010).

### **5.1.5 Identification of Hb variants using mass spectrometry**

The use of mass spectrometry to identify hemoglobin variants was first reported by Wada *et al.* in 1981. Wada used field desorption mass spectrometry to analyse tryptic peptides of abnormal and normal hemoglobins (Wada *et al.* 1981). Since then a number of mass spectrometry-based approaches have been developed and used to identify hemoglobin variants.

A recent review by Zanella-Cleon *et al.* summarises the phenotypic determination of hemoglobinopathies by mass spectrometry to date (Zanella-Cleon *et al.* 2009) and references over 50 publications which report the use of mass spectrometry to identify hemoglobinopathies. Over 20 of the publications referenced report the detection of novel variants by mass spectrometry-based approaches. Mass spectrometry has been used to characterise variants with one or two amino acid substitutions, insertions, deletions and post-translational modifications (Zanella-Cleon *et al.* 2009).

The mass spectrometry-based approaches used for hemoglobinopathy detection can largely be divided into two camps, those undertaken to identify unknown variants when an abnormal Hb is detected by phenotypic methods and those developed to screen for specific variants.

The general method adopted for unknown variant identification by mass spectrometry is a multi-step approach, first illustrated by Wild *et al.* (2001). ESI-MS on the intact globin chains is performed under denaturing conditions so that the globin chains and heme group are analysed as separate entities. Two globin chains that differ in mass by 6 Da or more may be resolved by ESI-MS and their average masses can be determined to within  $\pm 0.1$  Da (Wild *et al.* 2001). Globin chains that differ by less than 6 Da are detected by a change in the mass of the composite peak

(Rai *et al.* 2003). A -1 Da  $\beta$ -chain variant present at 20 % intensity (with normal  $\beta$ -chain at 80 %) would be expected to reduce the mass measurement of the composite peak by 0.2 Da. This analysis allows the variant present to be associated with the  $\alpha$ - or  $\beta$ -chain and a list of possible amino acid substitutions that could be responsible for the variant to be formed, based on the mass change seen. ESI-MS analysis of the tryptic digest of the globin chains then allows identification of the tryptic peptide containing the mutation and further ESI-MS/MS analysis of the variant tryptic peptide can allow positional determination of the amino acid change present.

The majority of structural variants identified are produced by single amino acid substitutions. This approach was originally developed for the detection of these, in adults, but can easily be adapted for the identification of mutants with insertions, deletions or post-translation modifications. This approach has also been adapted and applied successfully to newborn screening for sickle cell disease and homozygous  $\beta^0$ -thalassemia (Wild *et al.* 2004).

An alternative to this multi-step approach is to use a more directed approach whereby only known variants of interest are screened for. The tryptic digest of a blood sample is analysed by mass spectrometry using a multiple reaction monitoring (MRM) scan mode, on a triple quadrupole, or in data directed analysis (DDA) mode. The occurrence of particular fragment ions upon CID of precursors is looked for. In MRM mode, unique transitions from precursor mass to fragment mass are pre-defined for particular variants. In DDA mode, when ions are detected at a particular  $m/z$  CID is performed upon them and the MS/MS spectrum produced is interpreted to identify any variant present.

Daniel *et al.* (2005) have described a rapid and specific MRM approach for the simultaneous detection of the clinically significant variants Hb S, C, D-Punjab, E and O-Arab. They analysed 200 blood samples by this approach, in parallel with existing screening methods, and provided correct identification of all samples. They have also developed this approach further for the measurement of  $\delta$ : $\beta$ -globin peptide ratios and have shown that this is a suitable surrogate marker for Hb A<sub>2</sub> quantitation (Daniel *et al.* 2007). Their work has shown the potential for mass spectrometry to be

used as the primary screening method for hemoglobinopathies in the clinical setting as it fulfils all the current method requirements for antenatal/neonatal screening.

Basilico *et al.* (2007) have developed another directed approach to hemoglobin variant detection based on the HPLC-ESI-MS/MS analysis of the peptide mixture of a hemoglobin tryptic digest. They developed an in-house database containing theoretical mass spectra for various known variants. Experimental data obtained could then be correlated with that held in the peptide database to provide successfully identification of known variants (Basilico *et al.* 2007).

A limitation of both of these approaches is that they can only be used to identify known variants, which the experiment is designed to detect. The approach described by Basilico *et al.* (2007) has the further limitation that it does not provide a method for Hb A<sub>2</sub> quantitation.

The use of a mass spectrometry-based approach for glycosylated hemoglobin quantitation has also been evaluated by several groups (Roberts *et al.* 2001; Nakanishi *et al.* 2002). These have shown that ESI-MS analysis provides a precise measurement of glycosylated Hb, Hb A<sub>1c</sub>, comparable with measurements obtained by other methods. Nakanishi *et al.* (2002) have also reported that the presence of abnormal Hb S can cause erroneous Hb A<sub>1c</sub> measurements by commercial methods (such as HPLC). Quantitation of Hb A<sub>1c</sub> by MS is not affected by the presence of variants and in these cases provides more reliable quantitation results.

A fundamental limitation of the application of mass spectrometry to hemoglobinopathy screening is that it can only detect variants which result in a change in globin chain mass. The common clinically significant variants, however, all result in a mass change (Table 5.1). No variants produced by a Ile→Leu (or vice-versa) mutation have been reported in HbVar (Hardison *et al.* 2002). Gln→Lys mutations produce only a small change in mass and three variants resulting from this mutation have been reported. These variants can be detected easily upon tryptic digest analysis, however, as a new trypsin cleavage site is formed by the mutation.

### 5.1.6 Clinical screening by MS?

Whilst it is clear that identification of hemoglobin variants by MS provides complementary results and can provide more information than HPLC or IEF, it has not been adopted within the clinical setting for population screening. This is despite an approach for screening by MS having been developed by Daniel *et al.* (2005 and 2007) and being of a similar time requirement and cost per test to the current HPLC systems. HPLC and IEF systems require relatively small capital investment but have high running costs, whereas the consumable costs for MS analysis are small but the initial capital investment required is significantly higher. Daniel *et al.* (2005) suggest that the economics are dependent on the size of the population being screened and that a tandem MS approach would be more cost-effective with larger populations. The limitation of a tandem MS, MRM based approach, is that it only provides diagnosis of known variants, for which MRM transitions are defined. The appeal of a mass spectrometry-based method to screening would be greater if it provided substantially more information than is obtained by HPLC or IEF, for a similar cost and time requirement.

A MS-based screening approach could be developed which would provide definitive diagnosis of unknown variants, as well as the common clinically significant ones and quantification of Hb F, Hb A<sub>2</sub> and Hb A<sub>1c</sub>. A multi-step approach to variant identification as described by Wild *et al.* (2001) would be favoured for this approach as an MRM-based approach could not identify unknowns. This MS identification process, however, currently requires a highly-skilled operator and extensive manual data interpretation. Any MS-based screening approach would therefore need to overcome this limitation.

A potential solution would be to provide a high-throughput screen of the intact globin chains. This would then be used to determine if a variant was present and would quantify  $\delta:\beta$  globin ratios and hemoglobin glycation. When the presence of a variant was indicated by this approach it could then be definitively identified by analysis of the peptide mixture produced by tryptic digestion. This solution would require that the analysis could be performed in a high-throughput manner and that

the requirement for manual data processing and interpretation be simplified or removed.

Within this chapter, the use of mass spectrometry to provide definitive identification of several variants unidentifiable by HPLC is described. This is achieved through use of a multi-step approach similar to that of Wild *et al.* (2001) utilising a mass spectrometer of Q-TOF geometry. The use of a potential high-throughput approach to hemoglobinopathy screening is then investigated.

The application of a triple quadrupole instrument for rapid hemoglobinopathy screening has been shown to produce reproducible results in previous studies. The methods, previously described by Wild *et al.* (2001) and Daniel *et al.* (2005 and 2007), both have their disadvantages. The multi-step approach described requires extensive manual data processing and interpretation whilst the directed tandem MS approach only looks for and identifies known variants (screened for by unique MRM transitions).

The use of an intact-globin screen in a high-throughput manner on a triple quadrupole is described. A Xevo-TQ (Waters Corporation, Milford, MA, USA) equipped with a TriVersa NanoMate (Advion Biosciences Ltd., Ithaca, NY, USA) automated sample introduction system was used to analyse blood samples for hemoglobin abnormalities (chain variants, high  $\delta$ -chain, high glycation). This approach was applied to the analysis of samples provided by Dr. Patel (University of Birmingham Centre for Cardiovascular Sciences). Dr. Patel was interested in screening samples from South Asians with and without cardiovascular disease (CVD) for the presence of hemoglobin anomalies.

In the UK, deaths from cardiovascular disease (CVD) are 40 % more common amongst people from the Indian Subcontinent (South Asian). It is unclear why South Asians are so susceptible to CVD. Hemoglobin disorders are also common in South Asians, but a connection to CVD has not been explored. It was therefore of interest to investigate whether hemoglobin disorders were more common to South Asians with CVD than those without. Hemoglobin abnormalities were detected by use of the high-throughput mass spectrometry-based approach and correlated with patient

information by Dr Patel. It was hypothesised that more hemoglobin abnormalities would be observed in CVD patients than in health controls.

This work does not attempt to provide a complete clinical study but rather focuses on the value of an automated mass spectrometry-based approach to hemoglobinopathy screening for selected demographic groups. It is hoped that with further development a mass spectrometry-based approach to screening could be used within the clinical setting.

## 5.2 Materials and Methods

### 5.2.1 Sample preparation for Q-TOF MS analysis

Anonymised patient blood samples were obtained from UHCW NHS Trust. Blood samples were prepared for analysis as follows. 10  $\mu\text{L}$  of blood from an EDTA-tube was diluted in 490  $\mu\text{L}$  of HPLC-grade water to create a stock solution. 100  $\mu\text{L}$  of stock solution was diluted 10-fold in 50 % acetonitrile 0.2 % formic acid for intact analysis and 100  $\mu\text{L}$  of stock was used for tryptic digestion. The tryptic digestion was performed as detailed in Wild *et al.* (2001). Prior to intact analysis, the sample was desalted by agitation for 30 seconds with approximately 5 mg of AG 50W-X8 cation-exchange beads (Bio-Rad Laboratories).

### 5.2.2 Q-TOF mass spectrometric analysis

A Q-Tof Ultima (Waters Corporation, Manchester, UK), equipped with a standard flow ESI source and controlled by MassLynx™ (v4.1) software (Waters Corporation, Milford, MA, USA) was used to analyse various blood samples. Samples were introduced into the ESI source at a flow rate of 5  $\mu\text{L}/\text{min}$  from a 250  $\mu\text{L}$  syringe using a syringe pump. The instrument was operated in ESI positive mode with a capillary voltage of 3 kV, 60 V cone voltage and source temperature of 110 °C for all experiments. The TOF mass analyser was tuned in V-mode to give an operating resolution of 1,500 (FWHM) for intact analysis and an operating resolution of 6,000 (FWHM) for digest analysis. The TOF mass analyser was calibrated using the  $\alpha$ -chain from a control blood sample. Data were acquired in continuum mode at a speed of 1 spectra/second and total ion counts were kept below 400 per scan to minimise dead time error.

The reduced instrument operating resolution of 1,500 (FWHM) for intact Hb analysis was achieved by reducing the gas pressure within the collision cell. This operating resolution allowed the data produced to reflect that which would be obtained on a triple quadrupole instrument. This operating resolution produced spectra that were more compatible with conventional MaxEnt processing. The

MaxEnt software, within MassLynx, which is used to deconvolute the data, was originally written for the deconvolution of triple quadrupole data. It assumes that peaks within the spectrum to be processed have a similar peak width and are Gaussian in shape. Spectra produced upon a Q-ToF instrument have peaks with asymmetric distributions and varying peak width with  $m/z$ . This does not create a significant problem when variant peaks are separable within the MaxEnt spectra. When a variant is present at -1 Da, however, its presence needs to be identified by a shift in mass in deconvoluted spectra. Globin chain mass, therefore, needs to be measured reproducibly to within  $\pm 0.1$  Da to detect a  $\pm 1$  Da variant at 10 % intensity. The increased resolution available on a Q-ToF instrument is not an advantage as it is not sufficient to separate the two species. Even on a FT-ICR operating at 200,000 resolution at 400  $m/z$  the two isotopic distributions would not be separated (Kleinert *et al.* 2008).

### 5.2.3 Testing the Q-TOF MS approach to hemoglobinopathy diagnosis

The ability to reproducibly measure the  $\beta$ -chain to within  $\pm 0.1$  Da was evaluated by analysing a control sample 30 times. Spectra acquired over one minute were combined and processed as described in Rai *et al.* (2003). Data were background subtracted with a 25-order polynomial such that 5 % of the data fell below the new baseline. Subtracted, combined, spectra were then deconvoluted onto a true mass scale by use of MaxEnt, a maximum entropy-based program, available within the MassLynx software. The MaxEnt mass output parameters used were a mass range of 14800-16800 Da and resolution of 0.2 Da/data point. The simulated isotope pattern damage model was used at a spectrometer blur width of 0.2-0.3 Da. This was chosen such that the  $\alpha$ -chain measurement fell within 0.2 Da of its sequence mass at the beginning of the MaxEnt iteration process. 40 % minimum intensity ratios were used and the process was allowed to iterate until convergence. The deconvoluted spectrum obtained was then smoothed and centred and the mass scale shifted linearly so that the  $\alpha$ -chain mass was equal to its sequence mass. The  $\beta$ -chain mass obtained was then recorded.

A -1 Da  $\beta$ -chain variant present at 10 % of the intensity of the normal  $\beta$ -chain would be expected to produce a shift in measured  $\beta$ -chain mass of 0.1 Da. To assess

whether this could actually be observed by our approach samples containing different concentrations of a -1 Da variant were produced and analysed. A normal blood sample was mixed in different ratios with a sample containing a homozygous -1 Da  $\beta$ -chain in order to create variant concentrations of 0 %, 25 %, 50 %, 75 % and 100 %.

The potential of the Q-TOF to identify variants was tested by analysing 10 blood samples for the presence of hemoglobin variants using the multi-step approach described by Wild *et al.* (2001). These blood samples had previously been characterised by Brian Green (Waters Corporation, Manchester, UK) by use of the Wild *et al.* approach on a triple quadrupole instrument. The samples were analysed without prior knowledge of the variants which they contained. Once they had been analysed the results acquired were compared with those previously obtained.

This MS-based approach to hemoglobin variant identification was then applied to a series of samples from UHCW NHS Trust for which HPLC results had indicated the presence of a variant. These samples had been submitted for analysis as part of the NHS Sickle Cell and Thalassaemia Antenatal Screening Programme.

HPLC experiments were performed by UHCW NHS Trust on a TOSOH HLC-723 HbG7 analyser. This system provides quantification of Hb F and Hb A2 and presumptive identification of hemoglobinopathies within a sample in less than 8 minutes.

#### **5.2.4 Ion mobility mass spectrometry**

Ion mobility mass spectrometry was used to aid in the identification of a particular variant. TWIM-MS was used to separate a precursor ion from an interfering ion at the same  $m/z$  value prior to MS/MS analysis. This allowed MS/MS data for the precursor of interest alone to be obtained. A Synapt HDMS system was used to perform TWIM-MS experiments and operated as described in Chapter 2 Section 2.2.3 unless otherwise stated. Optimised ion mobility conditions used were a travelling wave height of 11 V and a travelling wave velocity of 300 m/s. A mass acquisition range of 50-2000  $m/z$  was used.

### 5.2.5 High-throughput screening

A Xevo-TQ (Waters Corporation, Milford, MA, USA) equipped with a TriVersa NanoMate (Advion Biosciences Ltd., Ithaca, NY, USA) automated sample introduction system was used to acquire data on 174 samples provided by Dr. Patel (University of Birmingham Centre for Cardiovascular Sciences).

A preliminary case-control study was used to compare vascular damage, hemoglobin abnormalities and iron levels in South Asian CVD patients with those in controls. Vascular damage and iron levels were determined by Dr. Patel. Vascular damage was determined by ultrasound imaging: carotid-artery intima media thickness (CIMT). High-density lipoprotein (HDL) subfractions were separated using ultracentrifugation and analysed for iron. CIMT and HDL cholesterol levels provide a measure of the severity of heart disease present. Measurement of CIMT with ultrasound is a non-invasive, sensitive, and reproducible technique for identifying and quantifying subclinical vascular disease and for evaluating CVD risk (Stein *et al.* 2008). Higher CIMT and lower HDL cholesterol levels indicate higher disease risk.

Blood samples were assessed for the presence of single-point mutation hemoglobin variants, high  $\delta$ -chain and high glycation levels without knowledge of their disease status. The results of this analysis were then correlated with patient information by Dr. Patel. 10 hemoglobinopathy controls (as used above), with known hemoglobinopathy status, were also analysed by this approach.

Blood samples were prepared for analysis by diluting 0.5  $\mu$ L of blood in 250  $\mu$ L of 50 % acetonitrile 0.2 % formic acid within a 96-well plate, compatible with the TriVersa NanoMate. The samples were not desalted prior to analysis due to time constraints. Data were acquired through use of a sample list, overnight, in duplicate. For each sample, a three minute acquisition was performed in multi-channel acquisition (MCA) mode at a scan speed of two spectra/sec.

To allow the high-throughput analysis of results obtained the BioPharmaLynx software (Waters Corporation) was used for data processing. BioPharmaLynx allows

one to deconvolute multiple spectral outputs and compares the results of these analyses. The deconvolution parameters used reflected those used by Wild *et al.* (2001). A Gaussian damage model had to be used as an isotope damage model is not available within this software package.

The analysis identified the masses and intensities of all peaks present within the deconvoluted spectrum. The masses for the  $\alpha$ ,  $\beta$ ,  $\delta$  and glycosylated globin chains were pre-defined within the software and automatically detected within each sample. All data files obtained were analysed by the BioPharmaLynx method. The mass and intensity information for all chains of interest and any peaks observed in the spectrum at greater than 20 % intensity were imported into Excel. Excel was then used to compare the results for  $\alpha$ -chain measurement,  $\delta$ -chain intensity and glycosylated  $\alpha$ - and  $\beta$ -chain intensity between samples. The error in  $\alpha$ - and  $\beta$ -chain measurement was used to indicate whether an  $\alpha$ - or  $\beta$ -chain variant (with a mass shift 6 Da or less) was present. The presence of any additional peaks within the spectra at greater than 20 % intensity suggested the presence of a structural variant resulting from a mass shift of more than 6 Da.

The percentage of glycosylated hemoglobin present was calculated as described by Roberts *et al.* (2001) from intensity information obtained for the  $\alpha$ - and  $\beta$ -chain and singly glycosylated  $\alpha$ - and  $\beta$ -chain species from the deconvoluted spectrum for each sample.

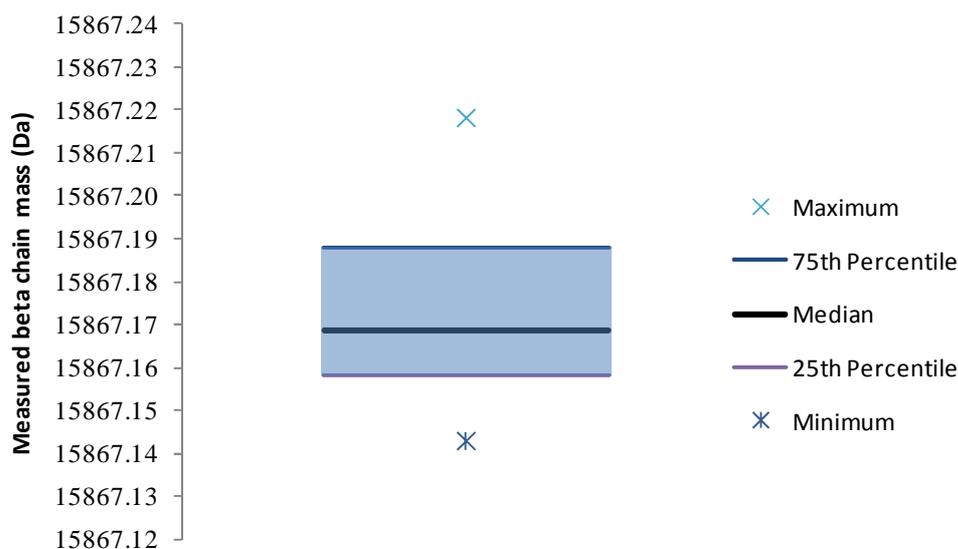
$\% \text{ GHb} = 50 * [\alpha_g / (\alpha + \alpha_g) + \beta_g / (\beta + \beta_g)]$  where GHb is glycosylated Hb,  $\alpha$  and  $\beta$  represent the intensities of the  $\alpha$ - and  $\beta$ -chains, and  $\alpha_g$  and  $\beta_g$  represent the intensities of the glycosylated  $\alpha$ - and  $\beta$ -chains respectively (Roberts *et al.* 2001).

This approximation of glycosylated hemoglobin level is based on several assumptions including that there is no significant contribution to glycosylated hemoglobin from other globin chains, that all  $\alpha$ -chain species have the same sensitivity and that all  $\beta$ -chain species have the same sensitivity. These assumptions are considered reasonable given that results reported by this approach have been shown to be highly correlated with those obtained by other techniques (Roberts *et al.* 2001).

In a similar way, estimation of  $\delta$ -chain levels was achieved by performing a ratio calculation of  $\delta$ -chain intensity to  $\beta$ -chain intensity. Here it is assumed that the  $\beta$ - and  $\delta$ -chains have similar ionisation efficiencies. This approach uses a  $\delta:\beta$  ratio of the intact chains, in contrast to the work of Daniel *et al.* (2007) who used a peptide-based method, to provide a surrogate marker of Hb A<sub>2</sub> levels.

### 5.3 Results and Discussion

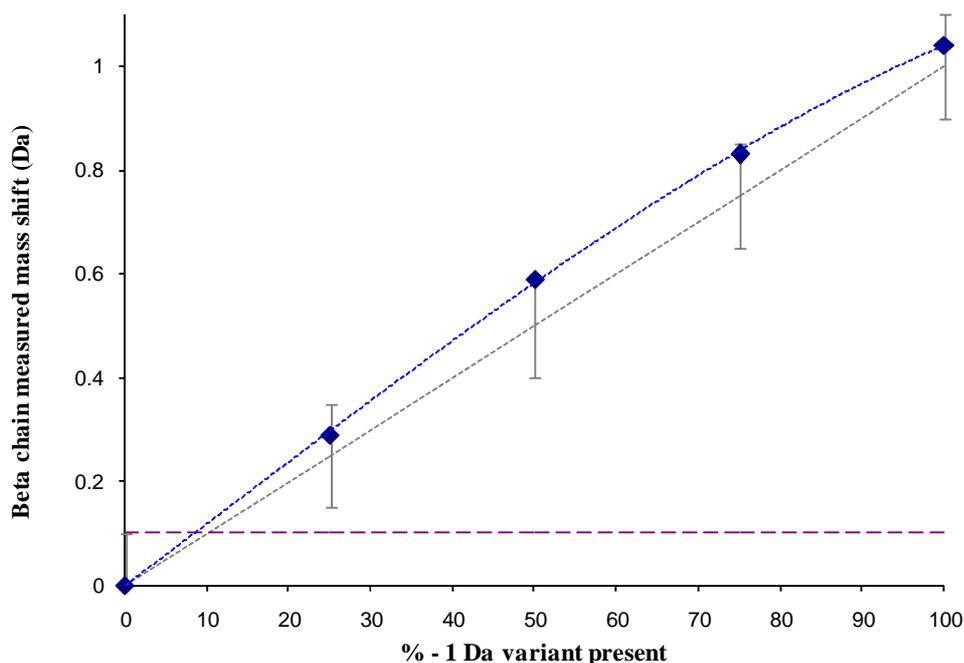
Normal  $\beta$ -chain mass measurements obtained by Q-TOF MS analysis for 30 replicates were all within 0.1 Da of the  $\beta$ -chain average mass (15867.2405 Da), following internal mass correction with the known  $\alpha$ -chain value (see Figure 5.2 below).



**Figure 5.2:** Box plot illustrating the distribution of 30 normal  $\beta$ -chain mass measurements.

The mean measured mass of the  $\beta$ -chain was 15867.1744 Da with a standard deviation of 0.0227 Da. The  $\beta$ -chain measurements obtained were all lower than the true  $\beta$ -chain mass. It is not clear why this is the case but may relate to systematic errors within the method of data acquisition or the method of data processing. The reproducibility obtained suggests, however, that this approach can be successfully used to obtain mass measurements to within  $\pm 0.1$  Da.

Figure 5.3 below illustrates the shift in  $\beta$ -chain mass measurement obtained upon analysis of MS data for samples containing 0 %, 20 %, 40 %, 60 %, 80 % and 100 % of a -1Da variant.



**Figure 5.3:** The shift in mass measurement for the  $\beta$ -chain obtained upon MS analysis of samples containing 0 %, 20 %, 40 %, 60 %, 80 % and 100 % of a -1Da variant. The dotted blue line indicates experimentally obtained data. The dotted grey line indicates the theoretical values which should be obtained, with error bars indicating values which are within a  $\pm 0.1$  Da tolerance. The dotted purple line is shown to indicate the limit of detection.

This analysis confirms that this approach can be used to successfully detect the presence of  $\pm 1$  Da variants at greater than 10 % intensity by observing a variation in  $\beta$ -chain mass measurement of greater than 0.1 Da.

All 10 control samples analysed for hemoglobin variants by the Q-TOF MS-based approach were correctly identified. The results of this analysis are summarised in Table 5.2 below.

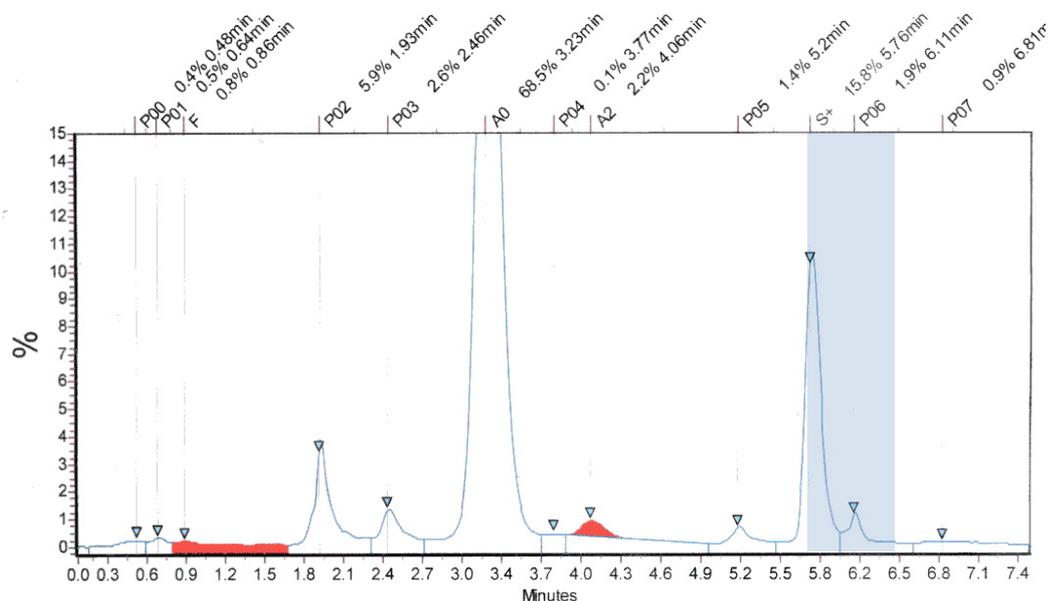
This approach has been used to provide definitive identification of variants present in more than 25 samples which were submitted to UHCW NHS Trust for antenatal screening and for which HPLC analysis did not provide variant identification. Here three case studies are presented which highlight the applicability of the MS-based approach to hemoglobinopathy diagnosis.

**Table 5.2:** Summary of information acquired by MS intact analysis and MS and MS/MS digest analysis of 10 control samples. The disease states of each hemoglobin sample were correctly identified.

Hemoglobin Sample	MS intact analysis	MS and MS/MS digest analysis
Sickle-trait (A/S)	-30 Da $\beta$ -chain detected	Additional peptide at 922.5 $m/z$ observed, ( $\beta$ T1 -30 Da) <sup>1+</sup> . MS/MS analysis of 922.5 $m/z$ confirmed sickle trait, $\beta$ 6(E→V).
D-Punjab/C heterozygote	$\beta$ -chain mass measurement (-0.89 Da) indicated -1 Da homozygote/heterozygote	Additional peptide at 694.5 $m/z$ observed, characteristic of Hb C, $\beta$ 6(E→K). Peptide at 689.4 $m/z$ , ( $\beta$ T13 -1Da) <sup>2+</sup> . MS/MS analysis of 689.4 $m/z$ confirmed $\beta$ 121(E→Q), D-Punjab.
E-trait (A/E)	$\beta$ -chain mass measurement (-0.34 Da) indicated -1 Da variant present	New peptides identified at 916.5 $m/z$ and 416.3 $m/z$ . Confirms $\beta$ 26(E→K), Hb E. Mutation provides a new trypsin cleavage site and two new peptides are formed with characteristic $m/z$ values.
O-Arab trait	$\beta$ -chain mass measurement (-0.43 Da) indicated -1 Da variant present	New doubly-charged peptide identified at 625.3 $m/z$ , presence of new peptide confirms $\beta$ 121(E→K), O-Arab trait.
D-Punjab homozygote	$\beta$ -chain mass measurement (-0.98 Da) indicated -1 Da homozygote/heterozygote	Peptide at 689.4 $m/z$ , (T13 -1Da) <sup>2+</sup> . MS/MS analysis of 689.4 $m/z$ confirmed $\beta$ 121(E→Q), D-Punjab.
Sickle/D-Punjab heterozygote	Two peaks observed for $\beta$ -chain. Mass measurements indicated the presence of a -30 Da variant and -1Da variant	Additional peptide at 922.5 $m/z$ observed, ( $\beta$ T1 -30 Da) <sup>1+</sup> . MS/MS analysis of 922.5 $m/z$ confirmed sickle trait, $\beta$ 6(E→V). Peptide at 689.4 $m/z$ , ( $\beta$ T13 -1Da) <sup>2+</sup> . MS/MS analysis of 689.4 $m/z$ confirmed $\beta$ 121(E→Q), D-Punjab.
Sickle homozygote	$\beta$ -chain mass measurement -30 Da from normal	Peptide at 922.5 $m/z$ observed, ( $\beta$ T1 -30 Da) <sup>1+</sup> . MS/MS analysis of 922.5 $m/z$ confirmed sickle, $\beta$ 6(E→V). No peak at 952.5 $m/z$ observed for normal $\beta$ T1 <sup>1+</sup> .
Lepore-Hollandia/ E heterozygote	Two peaks observed for $\beta$ -chain. Mass measurements indicated the presence of a -31 Da variant and -1Da variant	New peptides identified at 916.5 $m/z$ and 416.3 $m/z$ . Confirms $\beta$ 26(E→K), Hb E. Lepore-Hollandia suspected as no evidence for single E→P mutation. MS/MS analysis of 480.3 $m/z$ , ( $\beta$ T2 + 27 Da) <sup>2+</sup> , confirmed $\beta$ 9(S→T) and $\beta$ 12(T→N). MS/MS analysis of 628.8 $m/z$ confirmed $\beta$ 22(E→A).
Normal	$\beta$ -chain mass measurements in normal range (- 0.05 Da)	N/A
Normal	$\beta$ -chain mass measurements in normal range (- 0.03 Da)	N/A

### 5.3.1 Case Study One: Patient A

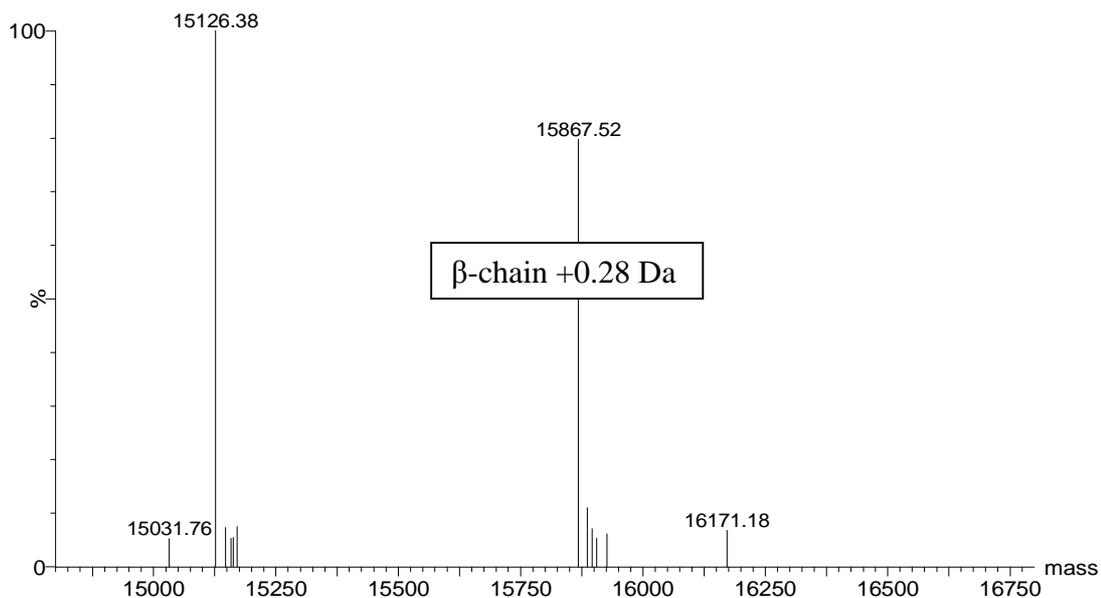
The HPLC trace from the analysis of a blood sample from patient A is shown below in Figure 5.4.



**Figure 5.4:** HPLC trace from TOSOH HLC-723 HbG7 analyser for blood sample from patient A. Retention time window for Hb S is shown in blue.

The HPLC trace shows the presence of an additional peak with a retention time (rt) of approximately 5.8 minutes. The variant responsible for this peak represents approximately 15.8 % of the total hemoglobin. This peak elutes just inside the Hb S retention time window but is not Hb S. The presence of a variant at such a low percentage indicates that it is probably an  $\alpha$ -chain variant. As there are four  $\alpha$ -globin genes, a mutation within one of these genes usually leads to the expression of an  $\alpha$ -chain variant 10-25 % of the time. A  $\beta$ -chain variant is usually expressed at 40-50 %. The identification of this variant cannot be determined by HPLC or IEF analysis.

ESI-MS analysis of this blood sample followed by spectral deconvolution was used to investigate whether the variant was in the  $\alpha$ - or  $\beta$ -chain. The  $\alpha$ -chain was first used as an internal data calibrant. The deconvoluted spectrum obtained after linear correction with the  $\alpha$ -chain is shown below in Figure 5.5. This shows an error in  $\beta$ -chain mass measurement of +0.28 Da. This alone would suggest the presence of a +1 Da  $\beta$ -chain variant.

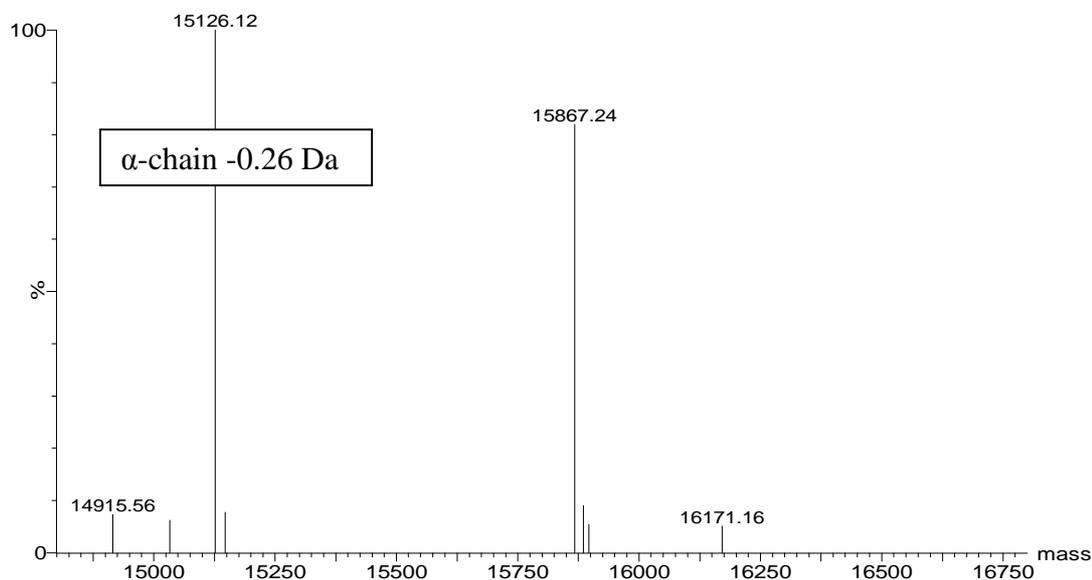


**Figure 5.5:** Deconvoluted spectrum obtained after ESI-MS analysis of blood sample from patient A with the  $\alpha$ -chain used as an internal calibrant.

As the calibration of the data was performed with the  $\alpha$ -chain, however, the  $\alpha$ -chain mass measurement was corrected to normal. It was therefore possible that the variant was a -1 Da  $\alpha$ -chain variant instead. The intact MS analysis, alone, cannot confirm whether the variant is a -1 Da  $\alpha$ -chain variant or a +1 Da  $\beta$ -chain variant. Electrophoretic charge change information or MS analysis of the tryptic digest is required to confirm this.

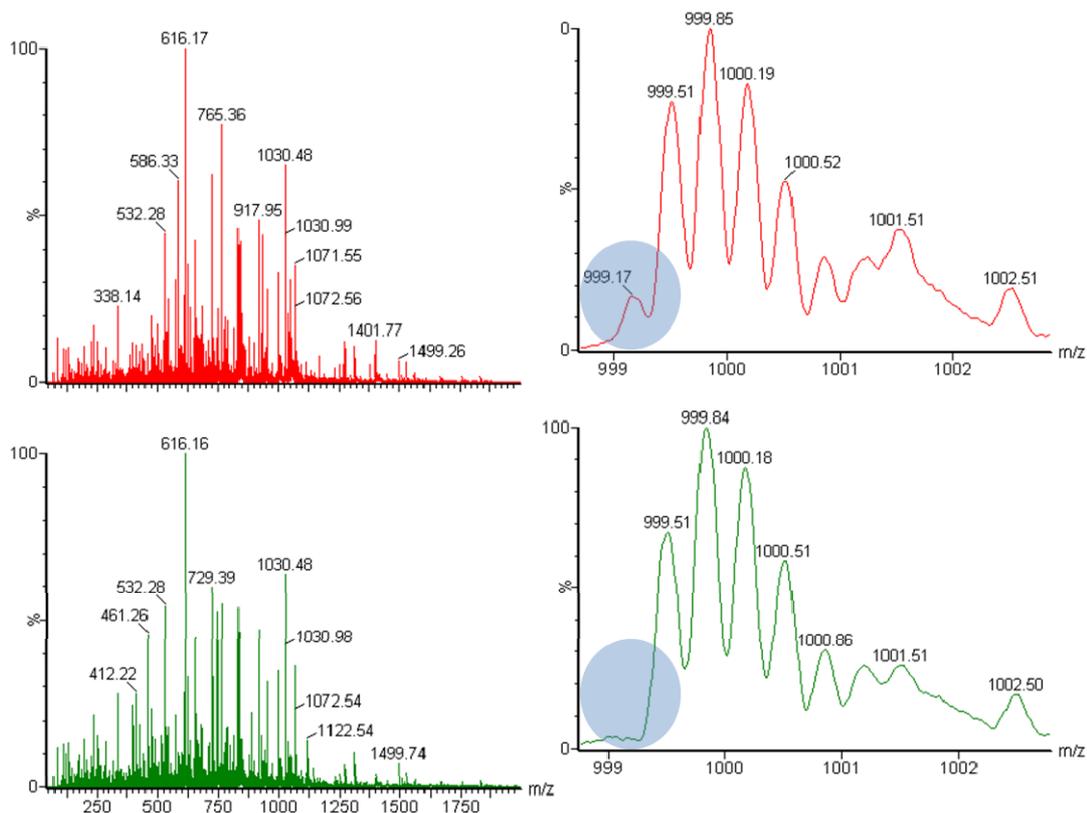
The HPLC trace showed that the variant had a positive polarity change and so it was inferred that the variant was an  $\alpha$ -chain variant. A single amino acid mutation which produces a zero or negative mass change gives a positive or zero electrophoretic charge change and the converse is also true (Rai *et al.* 2003). The positive  $\beta$ -chain mass change originally observed, together with the positive polarity change observed in the HPLC trace, indicated, therefore, that the variant was actually in the  $\alpha$ -chain. The  $\beta$ -chain was used to re-calibrate the data file prior to deconvolution. The resulting deconvoluted spectrum is shown in Figure 5.6. The  $\alpha$ -chain mass measurement (after linear correction with the  $\beta$ -chain) was -0.26 Da from the normal  $\alpha$ -chain mass which indicated that a -1 Da  $\alpha$ -chain was present. If HPLC information was not available, or this analysis was being conducted in an automated fashion,

intact MS analysis would indicate the presence of a -1 Da  $\alpha$ -chain variant or a +1 Da  $\beta$ -chain variant. In this case the sample would be subjected to further analysis as the presence of a variant was indicated and this further analysis would be used to determine the identification of the variant.



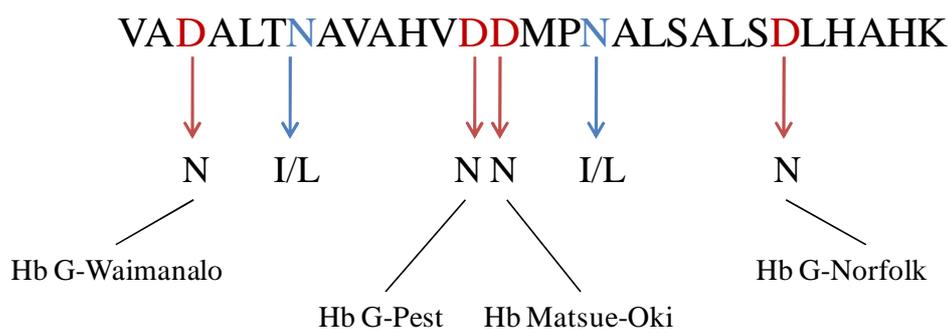
**Figure 5.6:** Deconvoluted spectrum obtained after ESI-MS analysis of blood sample from patient A with the  $\beta$ -chain used as an internal calibrant.

The blood sample from patient A was subsequently subjected to tryptic digestion and the digested sample analysed by means of ESI-MS. The spectrum obtained showed the presence of an additional peak at 999.17  $m/z$  (Figure 5.7). The additional peak at 999.17  $m/z$  was from a triply-charged ion. This corresponded to that which would be produced from a peptide 1 Da smaller than the 9<sup>th</sup> tryptic peptide of the  $\alpha$ -chain ( $\alpha$ T9). This suggested that the  $\alpha$ -chain variant present was the result of a single amino acid mutation within  $\alpha$ T9(62-90), which produced a -1 Da mass change. There are four possible amino acid changes which result in a -1 Da mass shift (D $\rightarrow$ N, E $\rightarrow$ Q, N $\rightarrow$ I/L, E $\rightarrow$ K) and six positions within  $\alpha$ T9 where one of these changes could have occurred (Figure 5.8). This information combined with the positive electrophoretic charge change observed in the HPLC trace would suggest that the mutation was D $\rightarrow$ N but does not give the exact mutation site.



**Figure 5.7:** ESI-spectra of tryptic digest of blood sample from patient A (red) and control (green). Inset spectra show the 999-1002  $m/z$  region in greater detail. A peak is observed at 999.17  $m/z$  in the sample which is not observed in the control.

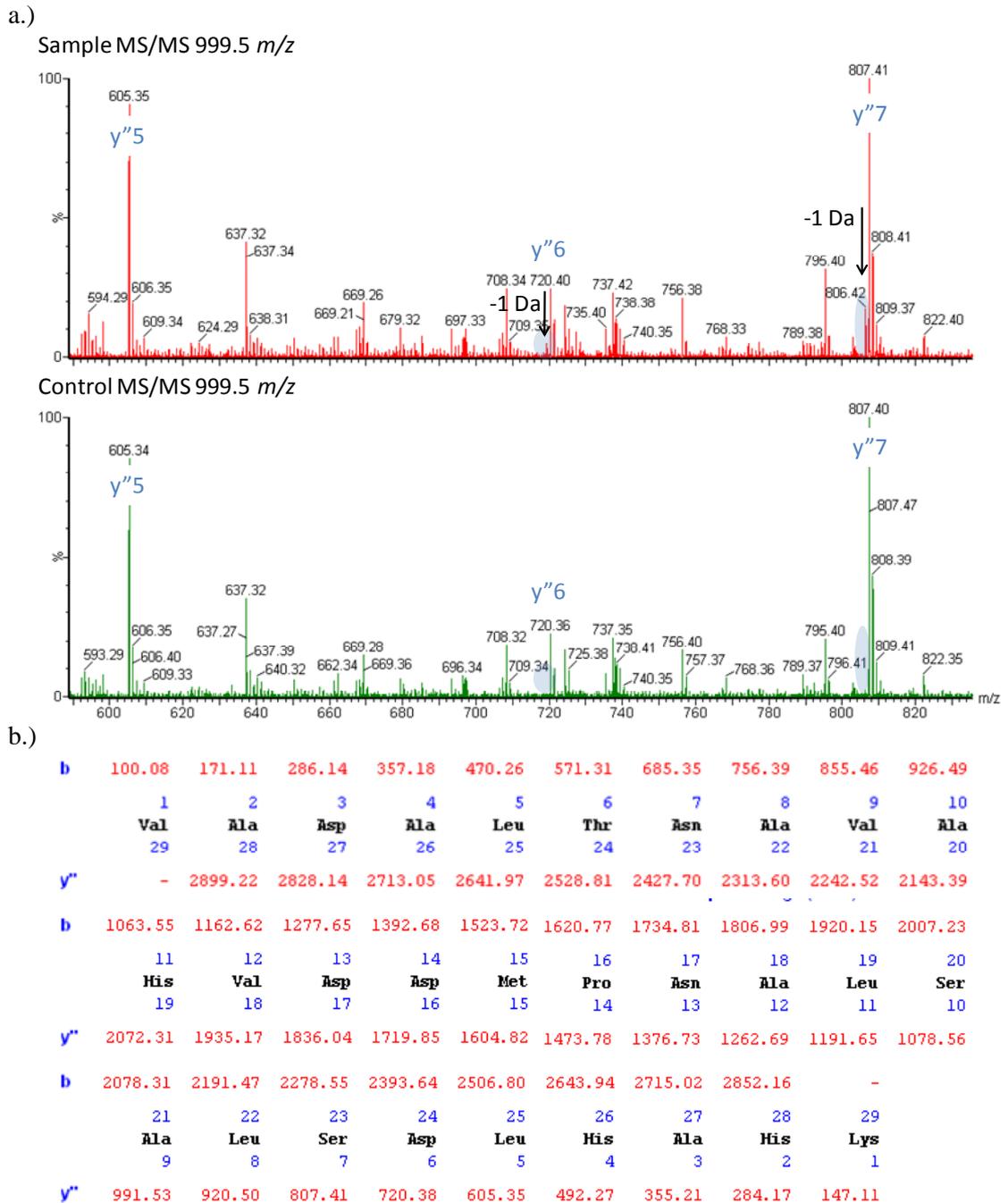
### $\alpha$ T9(62-90)



**Figure 5.8:** Schematic of  $\alpha$ T9, amino acids 60-92 in the  $\alpha$ -chain (shown in single-letter code). The potential single amino acid mutations which would result in a -1 Da mass change are illustrated.

The mutation site was confirmed by MS/MS analysis. Spectra obtained for MS/MS analysis on 999.5 ( $\pm 0.5$ )  $m/z$  ( $[\alpha$ T9 + 3H] $^{3H+}$ ) from a control and the sample are shown in Figure 5.9. Presence of fragment ions at -1 Da for  $y''6$  and larger  $y''$  ions

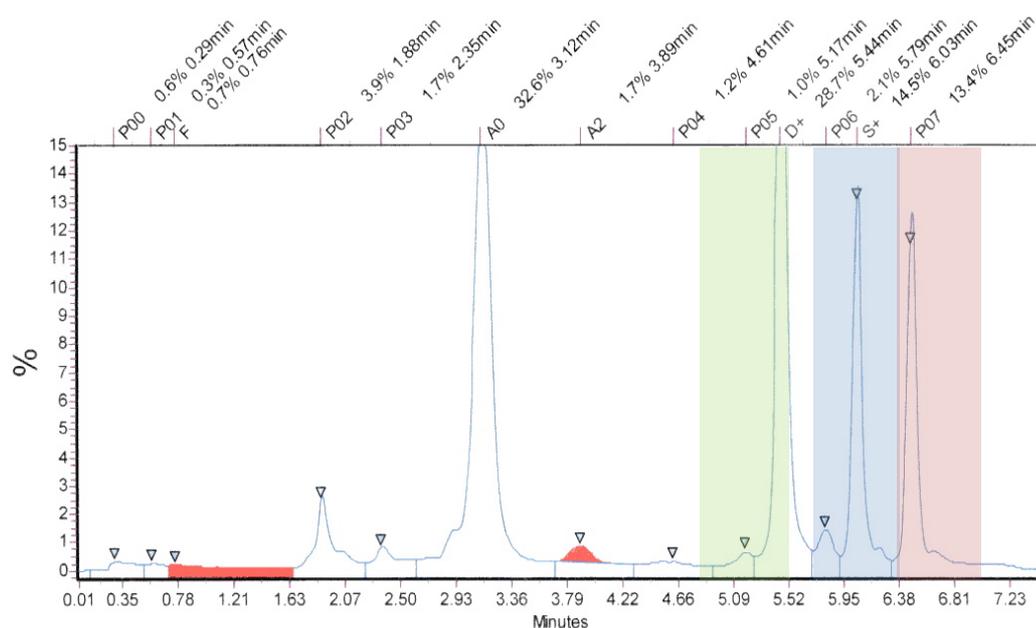
but not for  $y''5$  confirms that the mutated site is  $\alpha 85(\text{Asp} \rightarrow \text{Asn})$ . A search of HbVar showed that the variant present was Hb G-Norfolk. This variant is found in a few families in England and Canada and is associated with increased oxygen affinity.



**Figure 5.9:** a.) MS/MS spectra of 999.5  $m/z$  from variant sample (red) and control (green). Additional fragment ions are observed in the sample spectrum at -1 Da from the  $y''6$  ion and larger  $y''$  ions. These additional fragment ions confirm the identity of the variant present as  $\alpha 85(\text{Asp} \rightarrow \text{Asn})$ . b.) Expected fragment ions for 999.5  $m/z$  are shown.

### 5.3.2 Case Study Two: Patient B

The HPLC trace from the analysis of a blood sample from patient B is shown below in Figure 5.10.

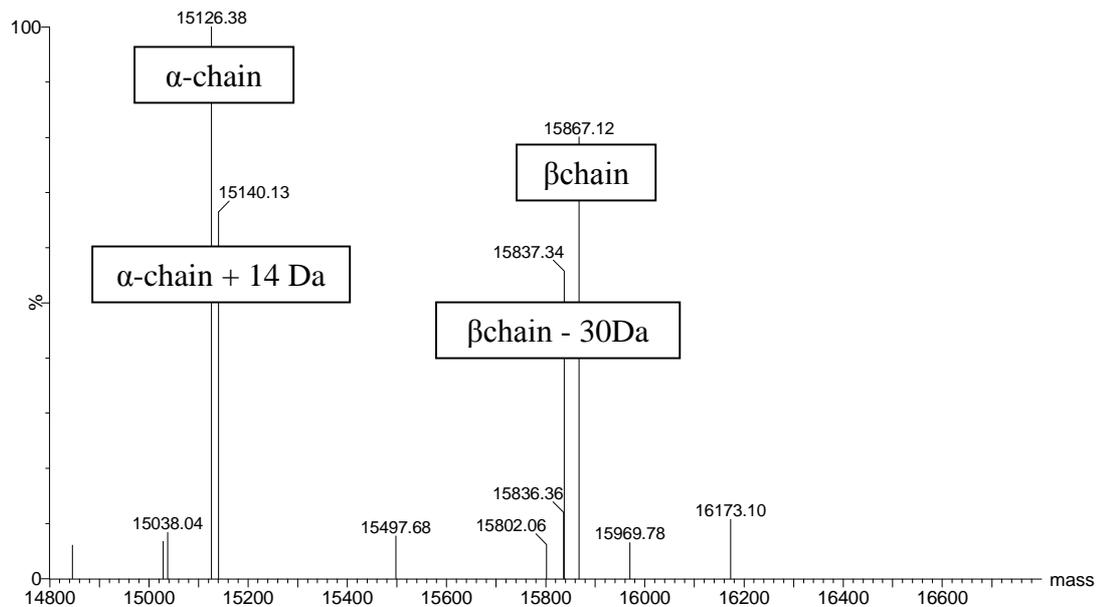


**Figure 5.10:** HPLC trace from TOSOH HLC-723 HbG7 analyser for blood sample from patient B. Retention time windows for Hb D, Hb S and Hb C are shown in green, blue and red respectively.

The HPLC trace showed the presence of three additional peaks with retention times in the Hb D, Hb S and Hb C retention windows respectively. The presence of four main peaks in the HPLC trace suggested that two variants were present within this sample.

The deconvoluted spectrum obtained following ESI-MS analysis of the blood sample from patient B is shown below in Figure 5.11. Four main peaks are seen in the spectrum at masses 15126.38, 15140.13, 15837.34 and 15867.12 Da. The peaks are identified as  $\alpha$ -chain,  $\alpha$ -chain variant (+14 Da,  $\alpha+14$ ),  $\beta$ -chain variant (-30 Da,  $\beta-30$ ) and  $\beta$ -chain respectively. The mass measurement of the  $\beta$ -chain in this case is not as accurate as would be expected. The mass error of -0.12 Da would ordinarily be expected to represent the presence of a  $\beta$ -chain variant at a low percentage. There are only two  $\beta$ -chain gene loci, however, and a peak is shown representing the presence

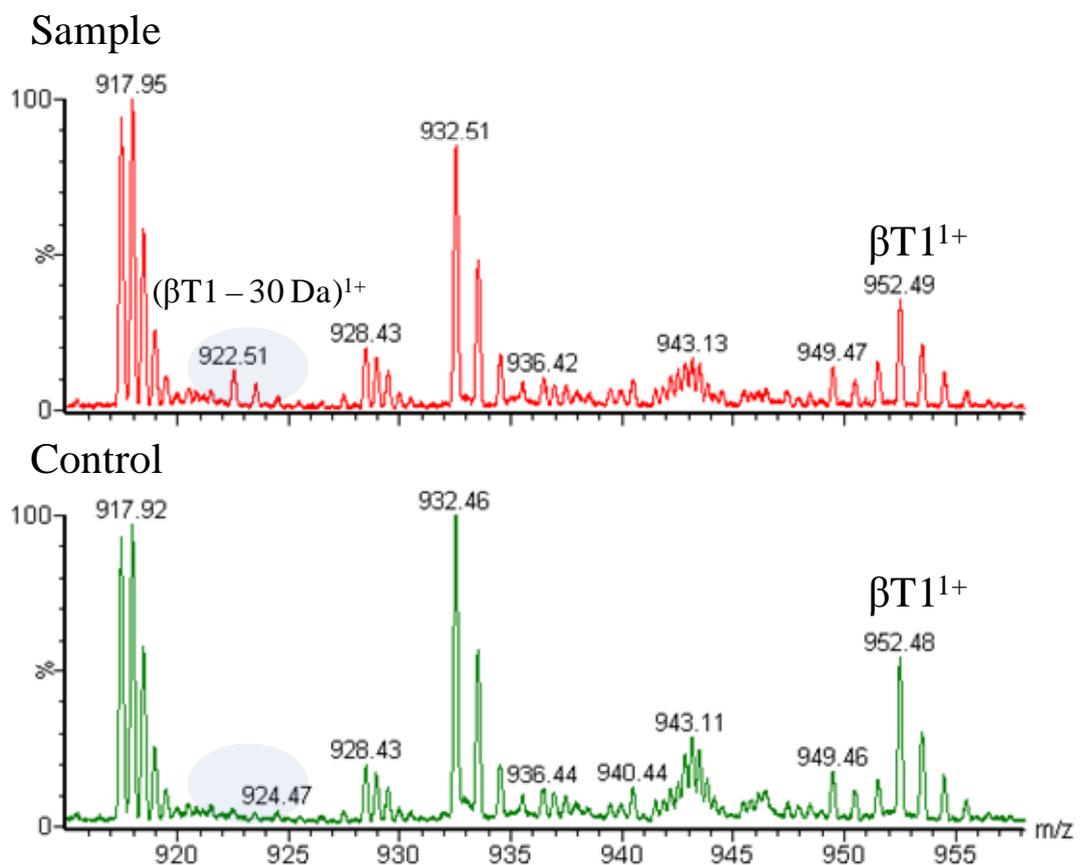
of a -30 Da  $\beta$ -chain variant. The other  $\beta$ -chain present must be normal. If an additional variant was present the HPLC trace for this sample would be different. There would be further additional peaks, representing variant dimers, and no peak corresponding to the elution of normal  $\alpha\beta$ -dimer.



**Figure 5.11:** Deconvoluted spectrum obtained after ESI-MS analysis of blood sample from patient B. The  $\alpha$ -chain was used as an internal calibrant. The main peaks within the spectrum are identified as  $\alpha$ -chain,  $\alpha$ -chain variant (+14 Da),  $\beta$ -chain variant (-30 Da) and  $\beta$ -chain.

When additional peaks are observable in the deconvoluted mass spectrum the mass measurement of individual peaks is often not within  $\pm 0.1$  Da. The error in mass measurement is increased to  $\pm 0.25$  Da here. This does not cause a problem as the observation of an error in mass measurement greater than 0.1 Da is always investigated further. The percentage of  $\alpha$ -chain variant present within the sample was estimated, by integration of the areas under the peaks corresponding to  $\alpha$ -chain and  $\alpha$ -chain variant within the deconvoluted spectrum, to be 51 %. The  $\beta$ -chain variant was estimated to represent 30 % of the total  $\beta$ -chain. The percentage of  $\alpha$ -chain variant present was higher than would normally be expected if only one  $\alpha$ -gene locus was mutated. This suggested the possibility that this patient had  $\alpha$ -thalassemia. The blood sample from patient B was subsequently subjected to tryptic digestion and the digested sample analysed by ESI-MS.

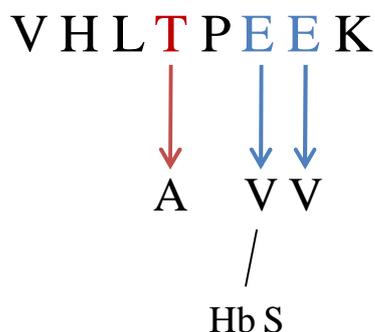
The spectra obtained showed the presence of some additional peaks in comparison to a control (Figures 5.12 and 5.15). Additional singly-charged ions were observed at 717.4 and 922.5  $m/z$  and an additional triply-charged ion was observed at 771.4  $m/z$ . The ion at 922.5  $m/z$  corresponded to that which would be produced by a peptide 30 Da smaller than the 1<sup>st</sup> tryptic peptide of the  $\beta$ -chain ( $\beta$ T1). There are five single amino acid substitutions which can result in a -30 Da mass change (T $\rightarrow$ A, W $\rightarrow$ R, S $\rightarrow$ G, M $\rightarrow$ T and E $\rightarrow$ V) and three positions in  $\beta$ T1 where one of these substitutions could have occurred (Figure 5.13).



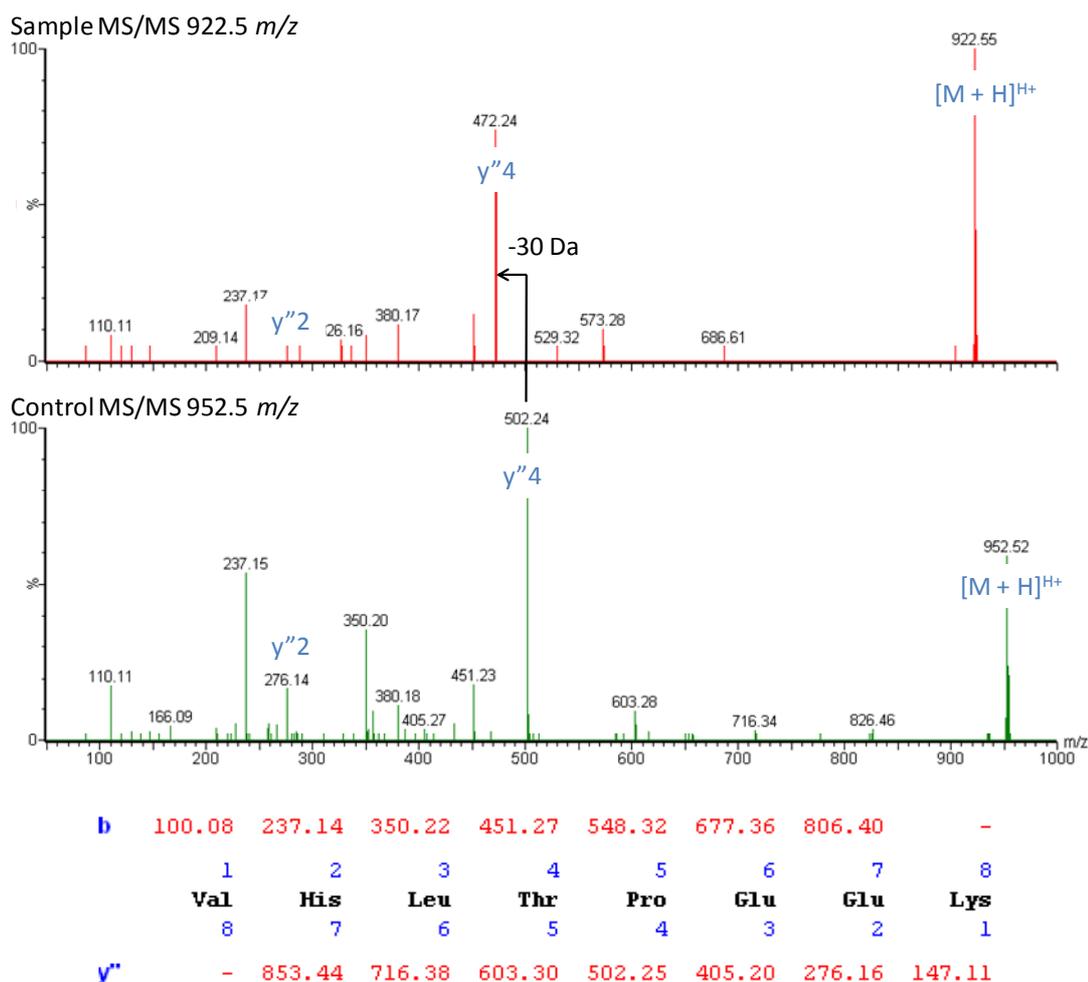
**Figure 5.12:** ESI-MS spectra of tryptic digest of blood sample from patient B (red) and a control (green). An additional ion is observed within the spectrum for the patient's sample at 922.5  $m/z$ . This corresponds to  $(\beta$ T1 - 30 Da)<sup>1+</sup>.

MS/MS analysis of 922.5  $m/z$  in comparison to 952.5  $m/z$  ( $\beta$ T1<sup>1+</sup>) was used to confirm the identification of the  $\beta$ -chain variant present (Figure 5.14). MS/MS analysis showed that the mutation was  $\beta$ 6(E $\rightarrow$ V) ( $\beta^S$ ), corresponding to Hb S.

# $\beta$ T1(1-8)

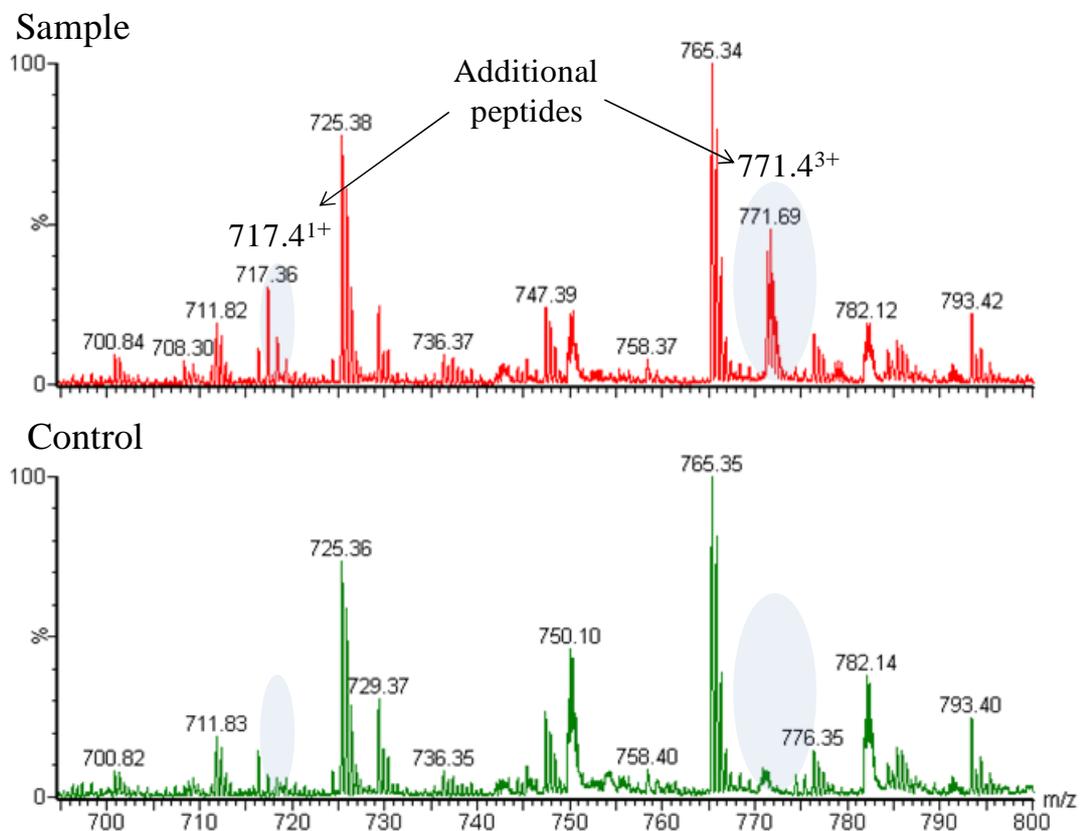


**Figure 5.13:** Schematic of  $\beta$ T1, amino acids 1-8 in the  $\beta$ -chain (shown in single-letter code). The potential single amino acid mutations which would result in a -30 Da mass change are illustrated.

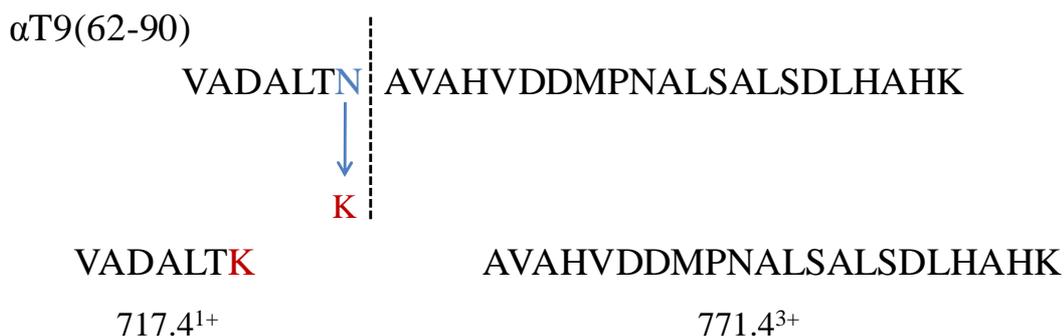


**Figure 5.14:** MS/MS spectra for 922.5 m/z from variant sample (red) and 952.5 m/z from control (green) with expected fragment ions for 952.5 m/z shown.

The  $\alpha$ -chain variant with a +14 Da mass change could have been produced by one of five possible single amino acid substitutions (N $\rightarrow$ K, D $\rightarrow$ E, G $\rightarrow$ A, S $\rightarrow$ T or V $\rightarrow$ I/L). As two other additional peptides need to be accounted for within the ESI-MS spectra for the tryptic digest (Figure 5.15 below) a N $\rightarrow$ K mutation is most likely. This mutation produces an additional trypsin cleavage site and so would produce two new peptides. The  $\alpha$ -chain sequence was then analysed to determine the new peptides which would be produced by a single N $\rightarrow$ K mutation. It was found that the additional peptides produced could be explained by an  $\alpha$ 68(N $\rightarrow$ K) (Figure 5.16). The  $\alpha$ -chain variant was thus identified as corresponding to Hb G-Philadelphia ( $\alpha^G$ ).



**Figure 5.15:** ESI-MS spectra of tryptic digest of blood sample from patient B (red) and a control (green). An additional singly-charged ion is observed at 717.4  $m/z$  and an additional triply-charged ion at 771.4  $m/z$  within the spectrum for the patient's sample.

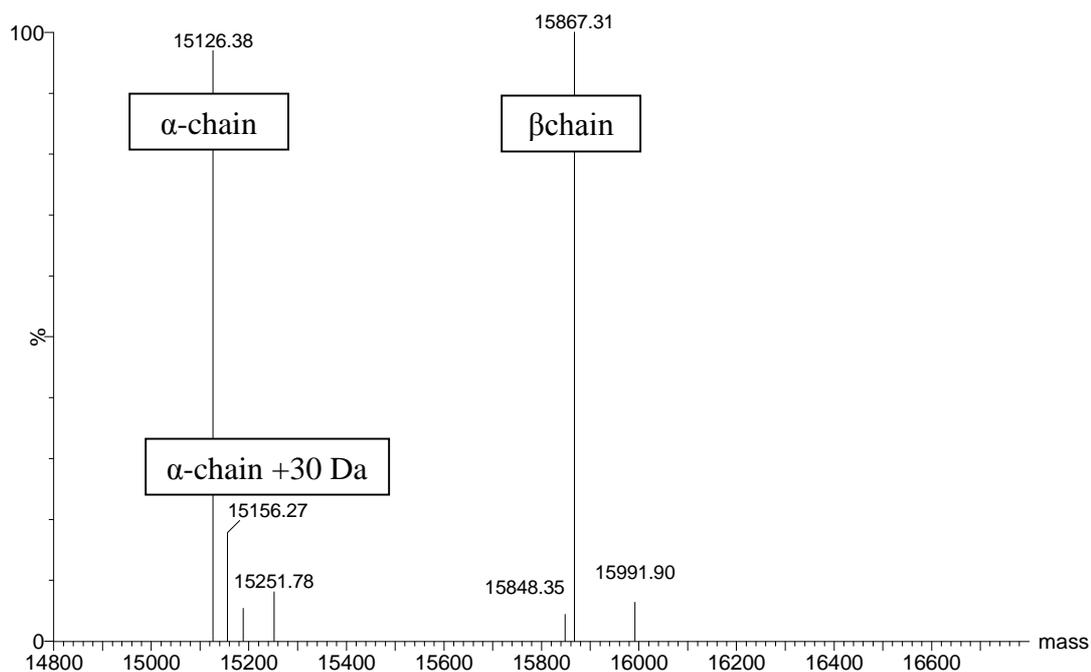


**Figure 5.16:** Schematic of  $\alpha$ T9, amino acids 60-92 in the  $\alpha$ -chain. The tryptic peptides produced for the mutant  $\alpha$ 68(N $\rightarrow$ K) are shown.

This patient has Hb S and Hb G-Philadelphia. Hb G-Philadelphia in Black populations is present with the 3.7 kilobase deletion alpha-thal-2 ( $-\alpha^G/\alpha\alpha$ ); the quantity of Hb G in such individuals varies between 30 and 35 %. When this deletion occurs in *trans* ( $-\alpha^G/-\alpha$ ) the quantity of Hb G is increased to approximately 45 %. These patients have a distinct microcytosis and hypochromia. The high percentage of Hb G-Philadelphia, estimated by the MS analysis, suggested that this individual may have this form of  $\alpha$ -thalassemia. With the variants present identified, the four main peaks present in the HPLC trace could also be assigned. They represented the elution of, from left to right,  $\alpha\beta$  dimer,  $\alpha^G\beta$  dimer,  $\alpha\beta^S$  dimer and  $\alpha^G\beta^S$  dimer respectively. The relative intensities that these peaks were observed at within the HPLC trace could now be used to calculate the relative intensities of the variant chains within the sample. Calculations based upon the HPLC analysis showed that approximately 31 % of the  $\beta$ -chain variant was present and 47 % of the  $\alpha$ -chain variant was present. The percentages estimated by the mass spectrometric analysis of 30 % and 51 % respectively are in good agreement with these measurements. The high percentage of Hb G-Philadelphia observed strongly suggests that this patient has *trans* alpha-thal-2. In this case, mass spectrometric analysis has been used successfully to provide identification of a clinically important variant.

### 5.3.3 Case Study Three: Patient C

The HPLC trace from the analysis of a blood sample from patient C showed no additional peaks but gave an abnormal presentation of the Hb A1c fraction. The deconvoluted spectrum obtained following ESI-MS analysis of the blood sample from patient C is shown below in Figure 5.17.

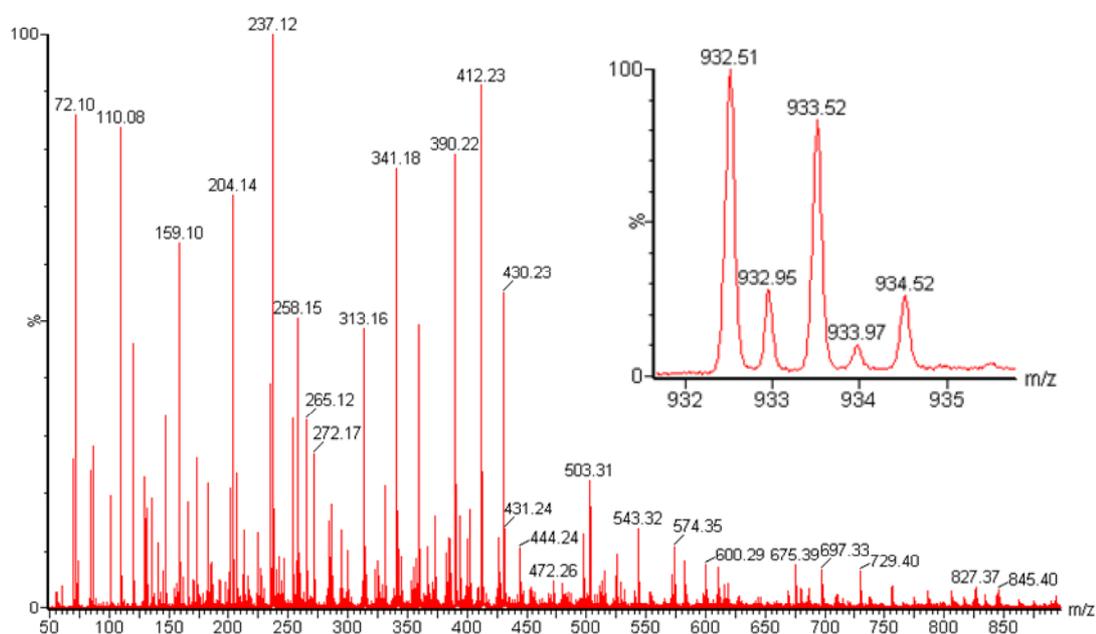


**Figure 5.17:** Deconvoluted spectrum obtained after ESI-MS analysis of blood sample from patient C with the  $\alpha$ -chain used as an internal calibrant. An additional peak at 15156.27 Da is observed within the spectrum. This represents the presence of a +30 Da  $\alpha$ -chain variant.

The  $\alpha$ -chain variant with a +30 Da mass change could have been produced by one of five possible single amino acid substitutions (A $\rightarrow$ T, R $\rightarrow$ W, G $\rightarrow$ S, T $\rightarrow$ M or V $\rightarrow$ E). Analysis of the tryptic digest of the blood sample indicated the potential presence of a doubly-charged ion at 932.4  $m/z$  which would correspond to the peptide  $\alpha$ T6 + 30 Da (Figure 5.18). This ion is masked by the presence of  $\beta$ T2<sup>1+</sup> at 932.5  $m/z$ . There are four possible positions in the 6<sup>th</sup> tryptic peptide of the  $\alpha$  chain where a single amino acid mutation could have resulted in a 30 Da mass increase (Figure 5.19).



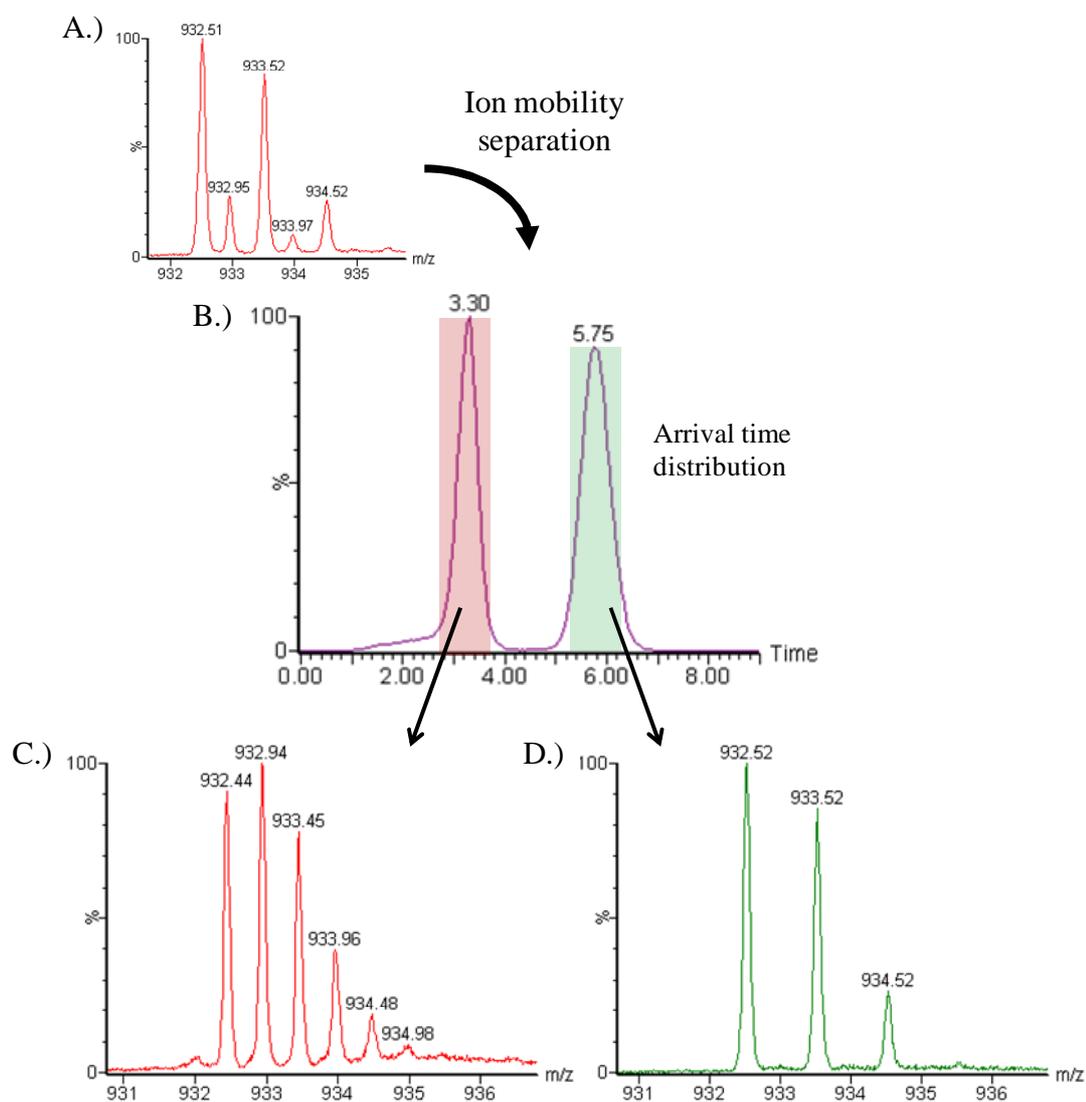
MS/MS analysis of 933  $m/z$  ( $\pm 2 m/z$  units) produced a complicated spectrum which was difficult to interpret (Figure 5.20).



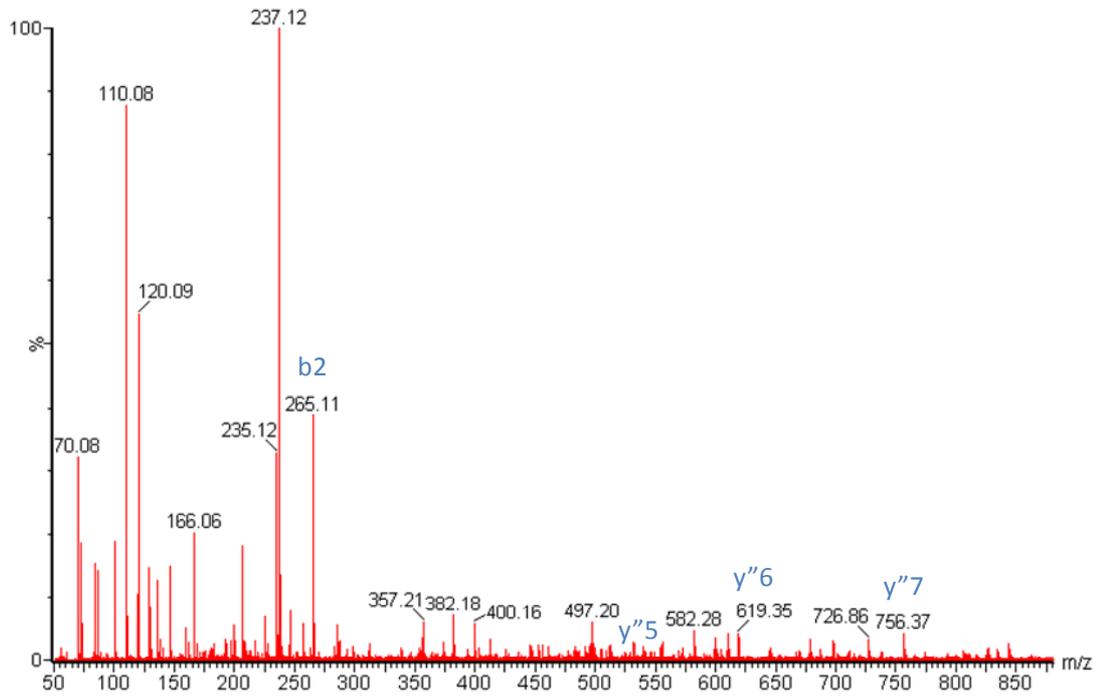
**Figure 5.20:** MS/MS spectrum of 933  $m/z$  ( $\pm 2 m/z$  units) from blood sample of patient C. Inset spectrum shows the precursor ions selected.

To aid in variant identification ion mobility separation was used to separate the doubly-charged precursor ion at 932.4  $m/z$  from the singly-charged precursor ion at 932.5  $m/z$  prior to MS/MS analysis (Figure 5.21). The quadrupole of the Synapt HDMS instrument was operated in resolving mode to allow the transmission of 933  $m/z$  ( $\pm 2 m/z$  units). The arrival time distribution of the mobility separated species showed two peaks, one corresponding to the ATD of the doubly-charged species present and one to the singly-charged species present. Collision-induced dissociation of these species was then performed within the transfer region of the instrument. This allowed product ion spectra for both the doubly-charged and singly-charged species to be obtained. The MS/MS spectrum obtained for the doubly-charged ion was interpreted to confirm the identification of the variant as Hb Riccarton,  $\alpha 51(G \rightarrow S)$  (Figure 5.22). Ion mobility was therefore successfully used to reduce spectral complexity and to facilitate the identification of this hemoglobin variant. Ion

mobility has also been used previously to facilitate the identification of the variants Hb Fort Worth and Hb J-Paris-I (Williams *et al.* 2008).



**Figure 5.21:** Illustration of the separation of overlapping doubly- and singly-charged ions by travelling wave ion mobility mass spectrometry. A.) The 933  $m/z$  region was selected for ion mobility analysis. B.) The arrival time distribution showed the presence of two separate peaks corresponding to the arrival times for the C.) doubly-charged and D.) singly-charged species.

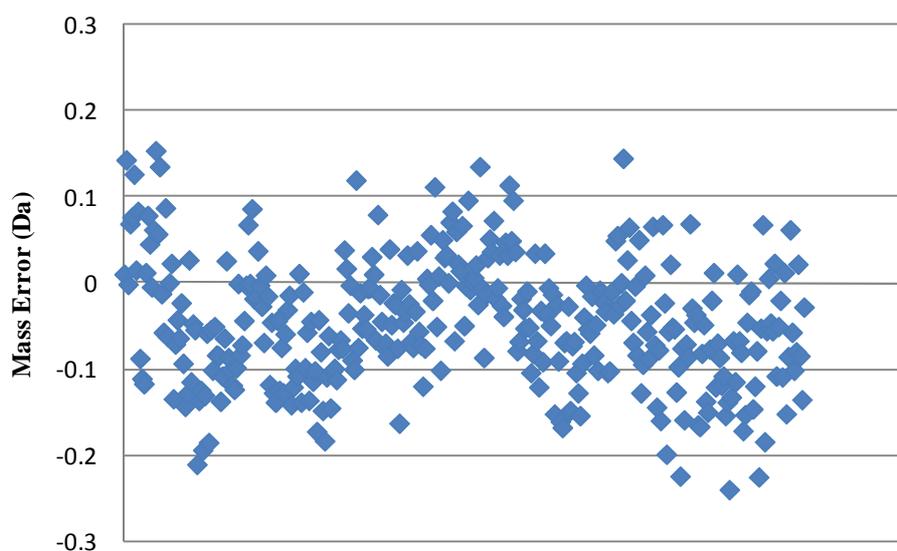


<b>b</b>	102.06	265.12	412.19	509.24	646.30	793.37	908.39	1021.48
	1	2	3	4	5	6	7	8
	Thr	Tyr	Phe	Pro	His	Phe	Asp	Leu
	16	15	14	13	12	11	10	9
<b>y''</b>	-	1763.95	1599.79	1452.72	1355.67	1218.61	1071.54	956.52
<b>b</b>	1108.51	1245.57	1332.60	1419.63	1490.67	1618.73	1717.80	-
	9	10	11	12	13	14	15	16
	Ser	His	Ser	Ser	Ala	Gln	Val	Lys
	8	7	6	5	4	3	2	1
<b>y''</b>	843.43	756.40	619.34	532.31	445.28	374.24	246.18	147.11

**Figure 5.22:** MS/MS spectrum for mobility-separated 932.4  $m/z$  (doubly-charged) with the corresponding peptide sequence and fragment ions. Interpretation of the spectrum confirms the mutation site as  $\alpha 51(G \rightarrow S)$ .

### 5.3.4 High-throughput screening

The error in  $\alpha$ -chain mass measurements acquired by the high-throughput, automated sample introduction, Xevo-TQ approach to screening intact hemoglobin chains coupled with BioPharmaLynx software processing are shown below in Figure 5.23.



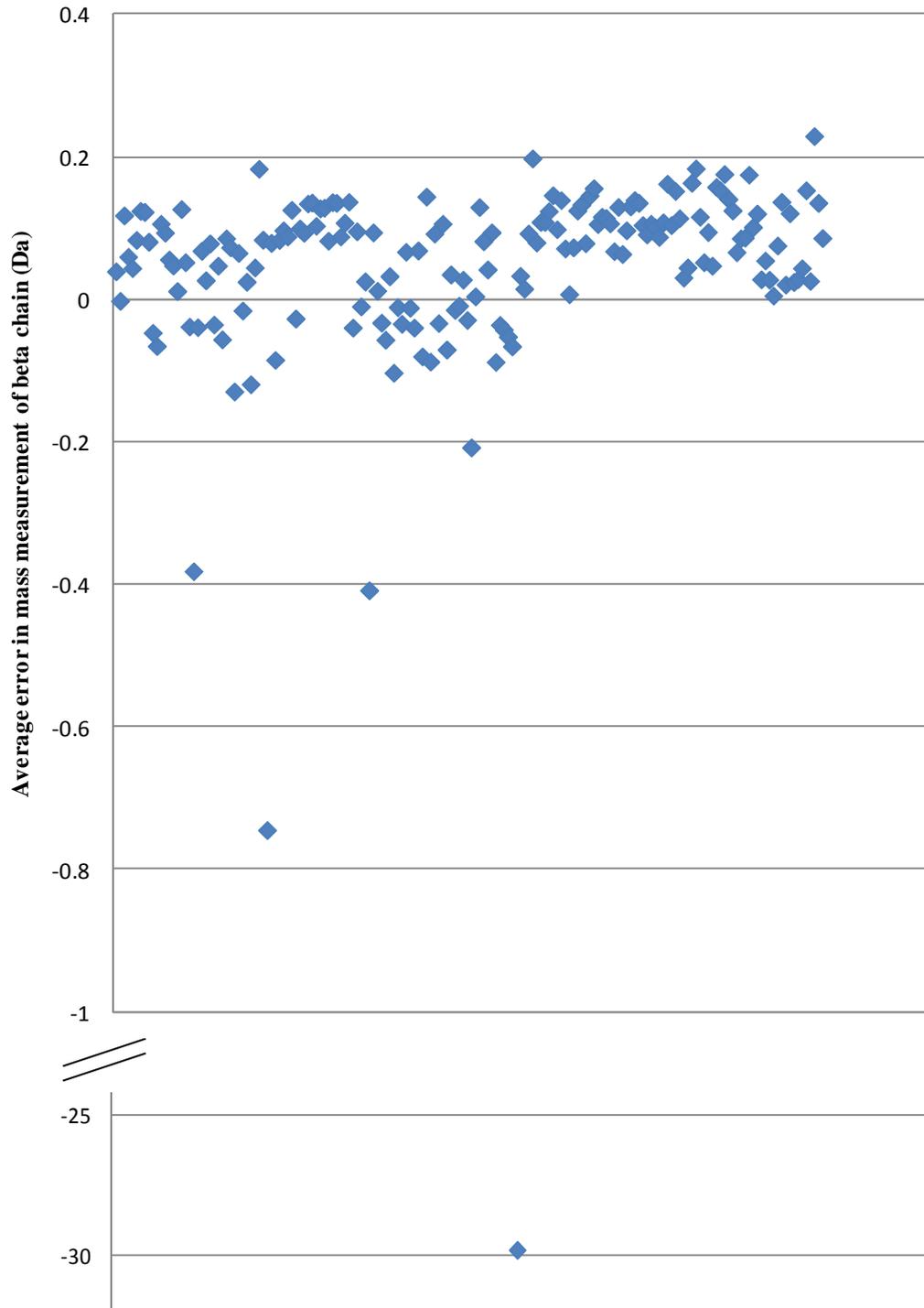
**Figure 5.23:** Mass error in  $\alpha$ -chain measurement obtained for 174 blood samples (two replicates).

The mean  $\alpha$ -chain measurement has an error of -0.04476 Da and a standard deviation of 0.073091 Da. This would suggest that this approach to measuring globin chains would indicate the presence of a variant at  $\pm 1$  Da occurring at greater than 20 % intensity (with 95 % confidence). 10 control samples, as analysed by the Q-TOF approach previously described, were also analysed by this approach. Measurements correctly indicated the presence of variants in these samples, where present, through the occurrence of new peaks within the deconvoluted spectrum or a shift in mass measurement of the globin chains.

The accuracy in mass measurement obtained here is lower than that reported by Wild *et al.* (2001) when this approach was used in a non-automated fashion for globin chain detection. This could have several causes. The samples analysed by this approach were up to one year old. Over this time period the globin chains can bind

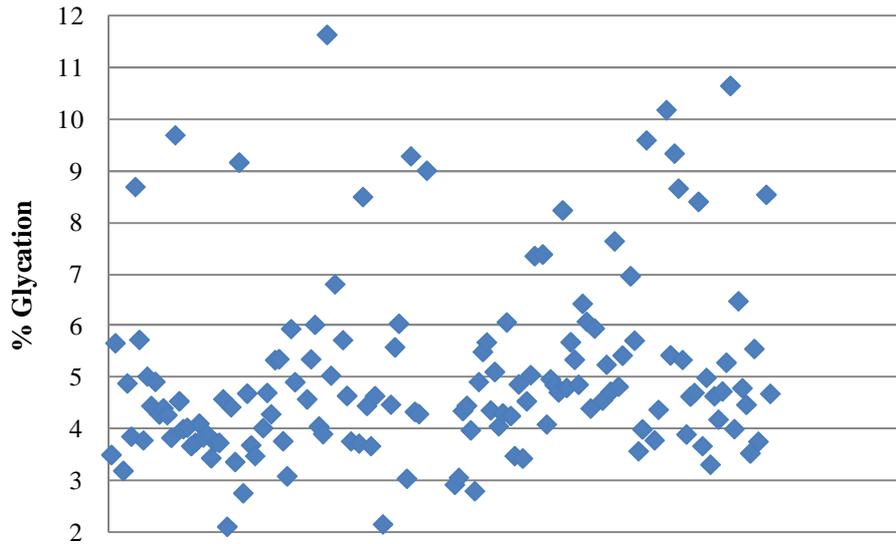
adducts, such as alkali metals. Peaks corresponding to these adducted species are observed within the spectra and can interfere with spectral deconvolution. Samples are usually desalted using cation-exchange beads prior to analysis to reduce this effect. This desalting was not performed here due to time constraints. The accuracy in mass measurement could have also been influenced by the parameters used within the BioPharmaLynx processing method. These may not have been optimal for the analysis of all the samples. The manual processing method described by Wild *et al.* (2001) is difficult to implement using the automated approach since it involves optimisation of the deconvolution parameters used for the processing on an individual sample basis. It is probable that both the sample preparation and data processing could be improved to provide an automated solution which measures the normal globin chains to within  $\pm 0.1$  Da (with 95 % confidence) as determined by the manual approach.

The average error in  $\beta$ -chain mass measurement obtained for each of the samples, from two replicates, (after linear correction with the  $\alpha$ -chain) is shown below in Figure 5.24. The majority of these were within  $\pm 0.2$  Da of the normal  $\beta$ -chain mass. When a  $\beta$ -chain measurement of more than 0.2 Da from normal was indicated manual data interpretation was conducted. In five samples an average change in  $\beta$ -chain mass measurement of greater than 0.2 Da was indicated. In one sample the presence of two  $\beta$ -chain variants was indicated, one at -30 Da and one at -1 Da from the normal  $\beta$ -chain mass. Manual data interpretation showed that four out of these five samples did contain variants. BioPharmaLynx automated processing, therefore, gave an erroneous result in one case.



**Figure 5.24:** Mass error in  $\beta$ -chain measurements obtained for 174 blood samples (average of two replicates).

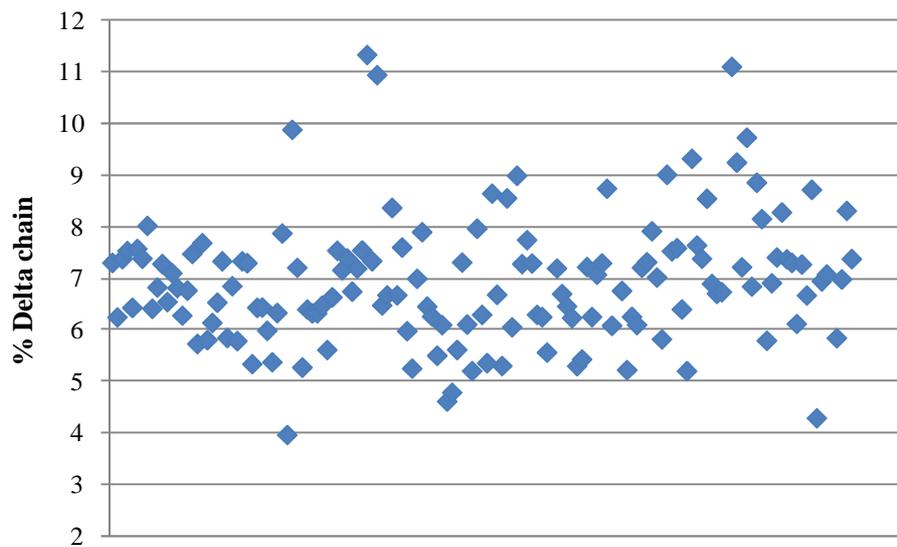
The estimated percentage of glycated hemoglobin calculated for each of the 174 samples is shown below in Figure 5.25.



**Figure 5.25:** Estimated glycated hemoglobin levels within each of 174 blood samples obtained by MS analysis (average of two replicates).

Normal adult hemoglobin glycation levels are between 4-6 % of total hemoglobin. As the actual glycation levels within these samples are unknown it is not possible to directly assess the reliability or accuracy of the approach used here to estimate glycation levels. This being said, approximately 50 % of the measurements estimated by the MS analysis fell within the expected normal range. The mean glycation level estimated by the MS analysis was 5.17 %. Glycation was recorded as high if greater than 6.5 %. This meant 21 blood samples were recorded as having a high hemoglobin glycation level.

The estimated percentage of  $\delta$ -chain present within each of the 174 samples relative to the  $\beta$ -chain is shown below in Figure 5.26. The percentage of  $\delta$ -chain within a blood sample, relative to the  $\beta$ -chain, would normally be expected to be between 2 and 3.5 %. In adults, Hb A<sub>2</sub> comprises about 2-3.5 % of total hemoglobin. The average percentage  $\delta$ -chain calculated here was much higher than this. This could be attributed to interference from sodium potassium-bound  $\beta$ -chain molecules. The isotopic pattern of the  $\beta$  chain with one potassium and one sodium atom bound overlaps with that of the  $\delta$ -chain. The  $\delta$ -chain measurements could therefore be higher than expected because a contribution to them is being made by interfering peaks.



**Figure 5.26:** Estimated percentage of  $\delta$ -chain present within each of the 174 blood samples relative to the  $\beta$ -chain (average of two replicates).

Examination of the deconvoluted spectra for each of the 174 blood samples show peaks corresponding to sodium-bound and potassium-bound  $\beta$ -chains. The intensity of these adduct peaks is greater than would normally be expected for fresh blood samples. The blood samples analysed here were up to one year old and desalting of the samples prior to analysis was not performed. Whilst the accurate quantitation of the  $\delta$ -chain relative to the  $\beta$ -chain was not possible by this approach the calculated percentage of  $\delta$ -chain could be used to provide an indication of a higher than normal measurement. The  $\delta$ -chain intensity measurement was recorded as high if greater than 8 %. Therefore 19 samples were recorded as having high  $\delta$ -chain levels.

The results of this analysis were presented to Dr. Patel. Dr. Patel correlated these results with the known disease status of these patients. All blood samples which were reported to have high hemoglobin glycation levels came from patients whom were diabetic. Hemoglobin anomalies detected by ESI-MS were more frequent for CVD patients than controls (35 % vs. 14 %,  $P=0.002$ ). Hemoglobin with elevated  $\delta$ -chain content was a common trait among CVD patients (19 %), who had more advanced atherosclerosis ((i) greater CIMT (0.75 vs. 0.65 mm,  $P=0.08$ ) and (ii) lower HDL cholesterol (0.78 vs. 1.03 mmol/L,  $P=0.03$ ) than normal patients independent

of age, gender and other confounding factors). These results are promising and, whilst preliminary, suggest that hemoglobin disorders may contribute to vascular damage in this population. Further research is needed to validate the approach used here and to investigate whether such proteomic techniques can be used in clinical practice to indicate CVD risk.

The high-throughput approach described above requires further optimisation and validation but shows promise. The presence of structural variants at 10 % intensity need to be identified so intact mass measurements of the globin chains need to be within  $\pm 0.1$  Da (with 95 % confidence). Future work will involve development of this approach and a small clinical trial, in collaboration with Coventry and Warwickshire NHS Trust, to assess its performance in comparison to the HPLC methodology. The continuation of this project will involve the development of objective-based software for automatically interpreting the processed data.

## 5.4 Conclusions

Mass spectrometry already provides a vital resource for the characterisation of hemoglobin variants not identifiable by phenotypic methods currently used within the clinical setting. A number of NHS trusts refer abnormal samples for mass spectrometric analysis. In this work, the definitive identification of a number of hemoglobinopathies, caused by single point mutations, by means of mass spectrometry-based approaches were shown.

$\beta$ -chain mass measurements obtained, for a number of standards, confirmed that a Q-TOF MS-based approach can be used to detect the presence of  $\pm 1$  Da hemoglobin variants at greater than 10 % intensity.

A high-throughput method, which used a TriVersa NanoMate, automated sample introduction system coupled with a QQQ instrument, was successfully used to identify the presence of hemoglobinopathies within 10 control samples. South Asians with and without CVD were subsequently screened by means of this approach. A positive correlation between patients with CVD and those with hemoglobinopathies was shown. These results were preliminary and further work to investigate whether hemoglobinopathies contribute to vascular damage is required.

The potential for the use of a mass spectrometry-based approach for antenatal and/or neonatal hemoglobinopathy screening within the clinical setting, in place of HPLC analysis, has been addressed by several groups (Wild *et al.* 2004; Daniel *et al.* 2005).

Any mass spectrometry-based approach to screening has the limitation that it can not detect variants which do not result in a mass change but only a small number of these exist. DNA analysis will always be required to confirm the presence of  $\alpha$ -thalassemias by detection of the mutations in non-coding regions of the globin gene responsible for them.

Mass spectrometry can be used to provide all the information which is currently supplied by the HPLC screen. In addition to this, an MS-based screen could provide

simultaneous detection of other variants. This could be important in the future as the increase in incidence of hemoglobinopathies could lead to the combination of hemoglobin variants, which, in the compound heterozygote state, produce severe disease. Reliable quantitation of Hb A<sub>2</sub> and Hb A<sub>1c</sub> can be provided by MS-based approaches (Roberts *et al.* 2001; Daniel *et al.* 2007). Erroneous results for quantification of Hb A<sub>2</sub> and Hb A<sub>1c</sub> can be provided by HPLC methods in the presence of interfering variants. In these cases, quantification by MS has been shown to be a superior approach.

The high-throughput approach to screening introduced here shows promise and with further development could lead to a population screening method suitable for implementation within the clinical setting.

## 5.5 References

- Ackerman, H., Usen, S., Jallow, M., Sisay-Joof, F., Pinder, M. and Kwiatkowski, D. P.** (2005). A Comparison of Case-Control and Family-Based Association Methods: The Example of Sickle-Cell and Malaria. *Annals of Human Genetics*. **69**, 559-565.
- Angastiniotis, M. and Modell, B.** (1998). Global Epidemiology of Hemoglobin Disorders. *Annals of the New York Academy of Sciences*. **850**, 251-269.
- Bain, B. J.** (2006). Laboratory techniques for the identification of abnormalities of globin chain synthesis. *Hemoglobinopathy diagnosis*. M. Khan. Oxford, Blackwell Publishing Ltd. **2nd Edition**, 26-62.
- Basilico, F., Di Silvestre, D., Sadini, S., Petretto, A., Levreri, I., Melioli, G., Farina, C., Mori, F. and Mauri, P. L.** (2007). New approach for rapid detection of known hemoglobin variants using LC-MS/MS combined with a peptide database. *Journal of Mass Spectrometry*. **42**, 288-292.
- Chevenne, D., Marle, N., Chauffert, M., Noel, M., Ducrocq, R. and Trivin, F.** (1999). Evaluation of the Tosoh HLC-723GHb V A1c 2.2 hemoglobin A1c analyzer. *Clinical Biochemistry*. **32**, 487-490.
- Clarke, G. M. and Higgins, T. N.** (2000). Laboratory Investigation of Hemoglobinopathies and Thalassemias: Review and Update. *Clinical Chemistry*. **46**, 1284-1290.
- Daniel, Y. A., Turner, C., Haynes, R. M., Hunt, B. J. and Dalton, N.** (2007). Quantification of hemoglobin A(2) by tandem mass spectrometry. *Clinical Chemistry*. **53**, 1448-1454.
- Daniel, Y. A., Turner, C., Haynes, R. M., Hunt, B. J. and Dalton, R. N.** (2005). Rapid and specific detection of clinically significant haemoglobinopathies using electrospray mass spectrometry-mass spectrometry. *British Journal of Haematology*. **130**, 635-643.
- Flint, J., Harding, R. M., Boyce, A. J. and Clegg, J. B.** (1998). 1 The population genetics of the haemoglobinopathies. *Baillière's Clinical Haematology*. **11**, 1-51.
- Hardison, R. C., Chui, D. H., Giardine, B., Riemer, C., Patrinos, G. P., Anagnou, N., Miller, W. and Wajcman, H.** (2002). HbVar: A relational database of human hemoglobin variants and thalassemia mutations at the globin gene server. *Human Mutation*. **19**, 225-233.
- Hartwell, S. K., Srisawang, B., Kongtawelert, P., Christian, G. D. and Grudpan, K.** (2005). Review on screening and analysis techniques for hemoglobin variants and thalassemia. *Talanta*. **65**, 1149-1161.

**Hickman, M., Modell, B., Greengross, P., Chapman, C., Layton, M., Falconer, S. and Davies, S. C.** (1999). Mapping the prevalence of sickle cell and beta thalassaemia in England: estimating and validating ethnic-specific rates. *British Journal of Haematology*. **104**, 860-867.

**Kleinert, P., Schmid, M., Zurbriggen, K., Speer, O., Schmutz, M., Roschitzki, B., Durka, S. S., Leopold, U., Kuster, T., Heizmann, C. W., Frischknecht, H. and Troxler, H.** (2008). Mass Spectrometry: A Tool for Enhanced Detection of Hemoglobin Variants. *Clin Chem*. **54**, 69-76.

**Modell, B., Khan, M. and Darlison, M.** (2000). Survival in [beta]-thalassaemia major in the UK: data from the UK Thalassaemia Register. *The Lancet*. **355**, 2051-2052.

**Nakanishi, T., Miyazaki, A., Shimizu, A., Yamaguchi, A. and Nishimura, S.** (2002). Assessment of the effect of hemoglobin variants on routine Hb A1c measurements by electrospray ionization mass spectrometry. *Clinica Chimica Acta*. **323**, 89-101.

**NHS Sickle Cell and Thalassaemia Screening Programme** (2009). Handbook for Laboratories. ISBN 13:978-0-9554319-2-0.

**Rai, D. K., Griffiths, W. J., Landin, B., Wild, B. J., Alvelius, G. and Green, B. N.** (2003). Accurate mass measurement by electrospray ionization quadrupole mass spectrometry: detection of variants differing by <6 Da from normal in human hemoglobin heterozygotes. *Analytical Chemistry*. **75**, 1978-1982.

**Roberts, N. B., Amara, A. B., Morris, M. and Green, B. N.** (2001). Long-term evaluation of electrospray ionization mass spectrometric analysis of glycosylated hemoglobin. *Clinical Chemistry*. **47**, 316-321.

**Ryan, K., Bain, B. J., Worthington, D., James, J., Plews, D., Mason, A., Roper, D., Rees, D. C., Salle, B. d. I. and Streetly, A.** (2010). Significant haemoglobinopathies: guidelines for screening and diagnosis. *British Journal of Haematology*. **149**, 35-49.

**Stein, J., H. , Korcarz, C., E. , Lonn, E., Kendall, C., B. , Mohler, E., R., Rembold, C., M. and Post, W., S.** (2008). Use of Carotid Ultrasound to Identify Subclinical Vascular Disease and Evaluate Cardiovascular Disease Risk: A Consensus Statement from the American Society of Echocardiography Carotid Intima-Media Thickness Task Force Endorsed by the Society for Vascular Medicine. *Journal of the American Society of Echocardiography : official publication of the American Society of Echocardiography*. **21**, 93-111.

**Wada, Y., Hayashi, A., Fujita, T., Matsuo, T., Katakuse, I. and Matsuda, H.** (1981). Structural analysis of human hemoglobin variants with field desorption mass spectrometry. *Biochimica et Biophysica Acta (BBA) - Protein Structure*. **667**, 233-241.

**Wild, B. J., Green, B. N., Cooper, E. K., Lalloz, M. R., Erten, S., Stephens, A. D. and Layton, D. M.** (2001). Rapid identification of hemoglobin variants by electrospray ionization mass spectrometry. *Blood Cells, Molecules, and Diseases*. **27**, 691-704.

**Wild, B. J., Green, B. N. and Stephens, A. D.** (2004). The potential of electrospray ionization mass spectrometry for the diagnosis of hemoglobin variants found in newborn screening. *Blood Cells, Molecules, and Diseases*. **33**, 308-317.

**Williams, J. P., Giles, K., Green, B. N., Scrivens, J. H. and Bateman, R. H.** (2008). Ion mobility augments the utility of mass spectrometry in the identification of human hemoglobin variants. *Rapid Communications in Mass Spectrometry*. **22**, 3179-3186.

**Zanella-Cleon, I., Joly, P., Becchi, M. and Francina, A.** (2009). Phenotype determination of hemoglobinopathies by mass spectrometry. *Clinical Biochemistry*. **42**, 1807-1817.

## Chapter Six: Conclusions and Future Work

---

The work presented in this thesis illustrates the use of TWIM-MS and MS to investigate properties relating to three-dimensional protein structure and to diagnose hemoglobinopathies.

## **6.1 Investigating three-dimensional protein structure by means of mass spectrometry**

### **6.1.1 Biological significance**

In recent years, mass spectrometry has established itself as a valuable structural biology tool for the investigation of tertiary and quaternary protein structure. It is now generally accepted that under controlled experimental conditions and over short timescales gas-phase protein structure reflects that in the solution phase. MS has thus been applied to the study of a number of protein systems, protein complexes and protein interactions.

In this work, the applicability of TWIM-MS to three-dimensional protein structure investigation was shown. Cross-sections estimated experimentally for protein standards and hemoglobin tetramers studied, for charge states most indicative of native structure, were in good agreement with those calculated from published X-ray and/or NMR structures. The results presented here show that TWIM-MS experiments can be used to obtain biologically-relevant data on protein structure.

### **6.1.2 Protein unfolding dynamics**

The cross-section of a protein calculated from IM-MS measurements varies with charge state. This allows the unfolding of proteins in the gas phase, with increase in charge, to be studied. Results obtained within this work support previous observations made within DCIM-MS studies (Shelimov *et al.* 1997; Shelimov and Jarrold 1997; Valentine *et al.* 1997; Jarrold 1999; Badman *et al.* 2001). The increase in estimated cross-section observed with increase in charge state for a particular protein depends on structural properties of the protein. The unfolding transition of a protein may be restrained by the presence of ligands or disulphide-bonds.

### **6.1.3 Hemoglobin analysis**

The global conformations of normal hemoglobin tetramers and sickle hemoglobin tetramers were distinguished by TWIM-MS measurements. Apo- and holo- forms of  $\alpha$ - and  $\beta$ -monomers were shown to have similar cross-sections. This suggests that  $\alpha$ - and  $\beta$ -monomers can retain a folded structure in the absence and presence of the heme group. Extensively disordered monomer structures were not observed. A heme-deficient dimer was not observed in the analysis of fresh blood samples and is unlikely to be an essential intermediate for hemoglobin tetramer assembly.

### **6.1.4 VanS studies**

TWIM-MS was used to provide information relating to the conformational states of the VanS<sub>A</sub> protein core under non-phosphorylated, ATP-associated and phosphorylated conditions. When in complex with magnesium and ATP, VanS<sub>A</sub> appeared to favour a more compact conformation. This conformational state is proposed to facilitate autokinase activity by bringing the ATP molecule and the site of phosphorylation into sufficient proximity for phospho-transfer to occur. The rate at which phosphorylation proceeded and the amount of phosphorylation observed was studied and results obtained were comparable with those from an established method. The use of mass spectrometry to study phosphorylation rates is a valuable technique, which offers a simple, reproducible alternative to conventional biophysical methods.

### **6.1.5 Future directions**

The application of mass spectrometry to three-dimensional protein structure is greatly expanding and is likely to continue to do so. MS has some unique advantages over conventional methods applied for protein structure analysis. It can be used to study heterogeneous populations, in real time, at biologically-relevant concentrations and within relative short time periods. When used in combination with other structural biology techniques, MS results can be used to provide detailed structural models for protein complexes (van Duijn 2010).

The launch of the commercially-available Synapt HDMS instrument significantly expanded the user-base of IM-MS for protein structure analysis. The recent introduction of the Synapt HDMS G2 instrument (Waters, Manchester, UK), which offers significant improvements over the first instrument, will surely increase the application of TWIM-MS to other research areas and larger and more complex systems. The Synapt G2 has improved mobility resolution (approximately four times that of the Synapt), an improved TOF resolution (over 40,000 FWHM) and a hybrid TDC/ADC detector which provides higher dynamic range.

## **6.2 Diagnosis of hemoglobinopathies by means of mass spectrometry**

### **6.2.1 Identification of rare variants**

MS is often used for identification of abnormal hemoglobin variants as illustrated by the vast number of publications (over 50 referenced within a recent review) within this area (Zanella-Cleon *et al.* 2009). The definitive identification of a number of hemoglobinopathies, caused by single point mutations, by means of a Q-TOF MS-based approach, was shown within this work. The identification of variants can be facilitated in some cases by the use of TWIM-MS.

### **6.2.2 High-throughput screening**

A high-throughput method, which used a TriVersa NanoMate, automated sample introduction system coupled with a QQQ instrument, was developed to provide a rapid screen for the presence of hemoglobin abnormalities. This method was used to screen blood samples from South Asians with and without CVD as part of a preliminary case-control study. A positive correlation between patients with CVD and those with hemoglobinopathies was shown.

### **6.2.3 Future directions**

The link between hemoglobinopathies and vascular damage will be explored further by expanding the preliminary work already conducted to larger sample sizes and other demographics.

There is potential for a mass spectrometry-based approach for antenatal and/or neonatal hemoglobinopathy screening to replace HPLC analysis within the clinical setting and this has been addressed by several groups (Wild *et al.* 2004; Daniel *et al.* 2005). The high-throughput approach to screening introduced here shows promise and with further development could lead to a population screening method suitable for implementation within the clinical setting.

### **6.3 Concluding remarks**

The application areas of mass spectrometry are constantly growing and many research areas, across the sciences, have benefited from the advancements in MS made over the last 10 to 20 years. Mass spectrometry is used extensively within a number of biological research fields including structural biology, proteomics, metabolomics and lipidomics.

MS is being used increasingly as a biomedical research tool for the diagnosis of disease and for biomarker discovery. Mass spectrometers are already used within hospitals for the diagnosis of metabolic disorders and their use for disease diagnosis is only likely to continue as new advancements are made and MS becomes more accepted as a viable diagnostic tool. Of particular interest is the development of mass spectrometric imaging, which can provide information about the spatial arrangement of biomolecules. This has the potential to be used routinely within the surgical setting, imaging tissues dissected during surgical operations (Chughtai and Heeren 2010). Over the next 10 to 20 years it can be envisaged that mass spectrometry will be used increasingly to bring disease diagnosis from the bench to the bedside.

## 6.4 References

- Badman, E. R., Hoaglund-Hyzer, C. S. and Clemmer, D. E.** (2001). Monitoring Structural Changes of Proteins in an Ion Trap over ~10-200 ms: Unfolding Transitions in Cytochrome c Ions. *Analytical Chemistry*. **73**, 6000-6007.
- Chughtai, K. and Heeren, R. M. A.** (2010). Mass Spectrometric Imaging for Biomedical Tissue Analysis. *Chemical Reviews*. **110**, 3237-3277.
- Daniel, Y. A., Turner, C., Haynes, R. M., Hunt, B. J. and Dalton, R. N.** (2005). Rapid and specific detection of clinically significant haemoglobinopathies using electrospray mass spectrometry-mass spectrometry. *British Journal of Haematology*. **130**, 635-643.
- Jarrold, M. F.** (1999). Unfolding, Refolding, and Hydration of Proteins in the Gas Phase. *Accounts of Chemical Research*. **32**, 360-367.
- Shelimov, K. B., Clemmer, D. E., Hudgins, R. R. and Jarrold, M. F.** (1997). Protein Structure in Vacuo: Gas-Phase Conformations of BPTI and Cytochrome c. *Journal of the American Chemical Society*. **119**, 2240-2248.
- Shelimov, K. B. and Jarrold, M. F.** (1997). Conformations, Unfolding, and Refolding of Apomyoglobin in Vacuum: An Activation Barrier for Gas-Phase Protein Folding. *Journal of the American Chemical Society*. **119**, 2987-2994.
- Valentine, S. J., Anderson, J. G., Ellington, A. D. and Clemmer, D. E.** (1997). Disulfide-Intact and -Reduced Lysozyme in the Gas Phase: Conformations and Pathways of Folding and Unfolding. *Journal of Physical Chemistry B*. **101**, 3891-3900.
- van Duijn, E.** (2010). Current Limitations in Native Mass Spectrometry Based Structural Biology. *Journal of the American Society for Mass Spectrometry*. **21**, 971-978.
- Wild, B. J., Green, B. N. and Stephens, A. D.** (2004). The potential of electrospray ionization mass spectrometry for the diagnosis of hemoglobin variants found in newborn screening. *Blood Cells, Molecules, and Diseases*. **33**, 308-317.
- Zanella-Cleon, I., Joly, P., Becchi, M. and Francina, A.** (2009). Phenotype determination of hemoglobinopathies by mass spectrometry. *Clinical Biochemistry*. **42**, 1807-1817.

# Travelling wave ion mobility mass spectrometry studies of protein structure: biological significance and comparison with X-ray crystallography and nuclear magnetic resonance spectroscopy measurements

Charlotte A. Scarff, Konstantinos Thalassinos, Gillian R. Hilton and James H. Scrivens\*

Department of Biological Sciences, University of Warwick, Coventry CV47AL, UK

Received 29 July 2008; Revised 27 August 2008; Accepted 27 August 2008

The three-dimensional conformation of a protein is central to its biological function. The characterisation of aspects of three-dimensional protein structure by mass spectrometry is an area of much interest as the gas-phase conformation, in many instances, can be related to that of the solution phase. Travelling wave ion mobility mass spectrometry (TWIMS) was used to investigate the biological significance of gas-phase protein structure. Protein standards were analysed by TWIMS under denaturing and near-physiological solvent conditions and cross-sections estimated for the charge states observed. Estimates of collision cross-sections were obtained with reference to known standards with published cross-sections. Estimated cross-sections were compared with values from published X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy structures. The cross-section measured by ion mobility mass spectrometry varies with charge state, allowing the unfolding transition of proteins in the gas phase to be studied. Cross-sections estimated experimentally for proteins studied, for charge states most indicative of native structure, are in good agreement with measurements calculated from published X-ray and NMR structures. The relative stability of gas-phase structures has been investigated, for the proteins studied, based on their change in cross-section with increase in charge. These results illustrate that the TWIMS approach can provide data on three-dimensional protein structures of biological relevance. Copyright © 2008 John Wiley & Sons, Ltd.

The tertiary conformation of a protein is vital in determining its biological activity and as such is a key structural feature. The most established methods for studying protein conformation are X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. These techniques do not allow the analysis of all proteins; many proteins do not form crystals and NMR studies require high concentrations of labelled protein. This excludes the use of proteins obtained directly from biological samples. Techniques that would allow the measurement of protein conformation at biological concentration and within biological timescales would be of great value in understanding biological processes.

Advancements in mass spectrometric analysis techniques along with ionisation methods have allowed for the study of aspects of three-dimensional protein structure. These include subunit stoichiometry and interactions, ligand interactions and acid denaturation.<sup>1</sup>

Electrospray ionisation (ESI) enables protein molecules to be analysed intact by mass spectrometry, producing an array of multiply charged gas-phase ions from protein molecules in solution. The multiply charged ions produced are pri-

marily a result of proton attachment to exposed basic sites of the protein and have been shown in many cases to be related to the solution-phase conformation of the protein.<sup>2</sup> It is hypothesised that the lowest charge states seen (under near-native conditions) are indicative of the native protein structure. A tightly folded conformation will have fewer exposed basic sites, and would accommodate fewer charges on the protein surface, than an unfolded conformation of the same protein.<sup>3,4</sup> Conversely, partially folded protein conformers can accept more charge upon transition from solution to the gas phase due to their larger solvent-exposed surface area.<sup>5</sup>

There is now increasing evidence to suggest that protein conformations in the gas phase mirror those in the solution phase, particularly over short time periods. Studies have shown that under controlled ESI conditions protein conformations can be preserved upon ionisation and gas-phase ions retain the conformational properties of structures present in the solution phase over the time that they exist within the mass spectrometer.<sup>2</sup>

Previous studies by drift-time ion mobility mass spectrometry (DTIMS) have shown that cross-sections estimated for the lower ESI-generated charge states of several proteins, including bovine cytochrome c,<sup>6</sup> bovine ubiquitin,<sup>7</sup> and equine apomyoglobin,<sup>8</sup> compare well with those calculated

\*Correspondence to: J. H. Scrivens, Department of Biological Sciences, University of Warwick, Coventry CV47AL, UK.  
E-mail: jh.scrivens@warwick.ac.uk

from X-ray crystal structures. As the charge state increases there is an increase in overall cross-section of the molecule, and this is thought to be attributable to the effects of Coulomb repulsion.<sup>8,9</sup>

In this study we analyse multiple proteins by travelling wave ion mobility mass spectrometry (TWIMS) under denaturing and near-physiological conditions and compare experimentally estimated collision cross-sections with those calculated from both X-ray and NMR structures utilising the well-established MobCal program.<sup>10,11</sup>

The travelling wave ion mobility separator (T-Wave)<sup>12</sup> has significantly increased sensitivity and speed in comparison with the traditional drift-time ion mobility device, allowing for the analysis of samples at biologically relevant concentrations. The T-Wave has been integrated into a quadrupole time-of-flight instrument, the Synapt HDMS system (Waters Corp., Milford, MA, USA).<sup>13</sup> The Synapt operates with excellent reproducibility and mass accuracy, and is capable of providing a wealth of information, being able to achieve ion mobility tandem mass spectrometry with the choice of performing collisionally induced dissociation (CID) before or after the mobility cell. For each ion mobility experiment, 200 complete mass spectra are acquired in a time frame of 15–30 ms. The nature of the T-Wave device means that, unlike in DTIMS, the ion drift time is not inversely proportional to ion mobility and thus absolute collision cross-sections cannot be obtained directly from drift-time information. It has been shown, however, that estimates of collision cross-sections can be obtained by reference to samples with known cross-sections,<sup>14–15</sup> provided that the data is obtained under the same experimental conditions; i.e. mobility gas, gas pressure, wave velocity, wave height, pusher frequency and injection energy.

## EXPERIMENTAL

### Sample preparation

Protein standards chicken lysozyme c, human lysozyme c, sperm whale myoglobin, equine myoglobin and equine cytochrome c were obtained from Sigma-Aldrich Ltd.

(Gillingham, UK). Each protein sample was diluted to a concentration of 10  $\mu$ M in 50% acetonitrile/0.1% formic acid (Mallinckrodt Baker Inc., Phillipsburg, NJ, USA) for denaturing conditions and in 10 mM ammonium acetate (Sigma-Aldrich Ltd.), pH 6.8, for near-native conditions. Prior to mass spectrometric analysis the samples were desalted by being spun twice at 11500 g for 10 min in an Ultraclear – 0.5 kDa cut-off (Millipore, Billerica, MA, USA) centrifugal filter. The sample retained on the filter was reconstituted in the same solvent.

### Ion mobility mass spectrometry

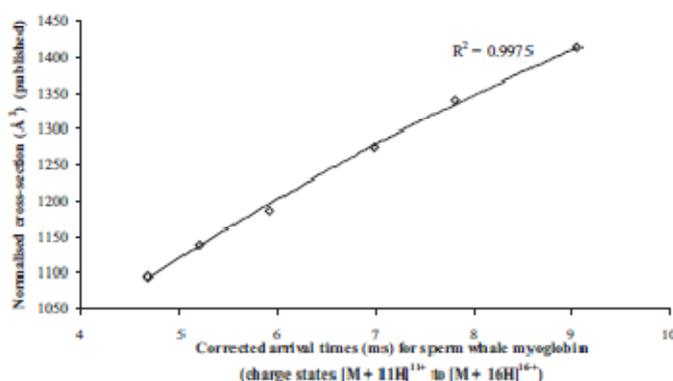
Experiments were performed on a Synapt HDMS system (Waters), described in detail elsewhere.<sup>13</sup> Samples were introduced into the source region by direct infusion nano-ESI.

The instrument was operated in ESI positive mode with a capillary voltage of 1.2 kV, cone voltage of 70 V and source temperature of 90°C. The ion mobility cell containing nitrogen was operated at an indicated pressure of 0.55 mbar corresponding to a flow rate of 30 mL/min. The travelling wave height and velocity were 9 V and 300 m/s, respectively.

Sperm whale myoglobin was analysed under denaturing conditions to create a calibration curve (Fig. 1) for cross-section measurements against values obtained from DTIMS studies.<sup>16</sup> The calibration was performed following an in-house procedure based on earlier work.<sup>14,15,17,18</sup> In brief, arrival times were corrected to exclude time spent outside of the ion mobility cell. Mass-independent time spent in the transfer region and mass-dependent time spent between the transfer region and time-of-flight mass analyzer were subtracted. Normalised cross-sections (corrected for charge and reduced mass) were then plotted against corrected arrival times to create a calibration with a power series fit. Rotationally averaged collisional cross-sections were estimated for the protein standards, at different charge states, under denaturing and near-native conditions.

### Theoretical cross-section calculations

To evaluate the measurements for rotationally averaged collisional cross-section obtained experimentally, compari-



**Figure 1.** Calibration curve of corrected arrival times for sperm whale myoglobin against normalised published cross-sections.

sons were made with theoretically calculated approximations of the measurement from published X-ray and NMR structures. MobCal, a program to calculate mobilities,<sup>10,11</sup> can be used to calculate cross-section via three methods.

The simplest method is the projection approximation (PA) which replaces the cross-section of an ion with its projection (shadow) and averages the projections created by every orientation of the ion. The PA is an adequate approximation for small molecules but underestimates the cross-section of proteins with highly convex structures where interactions with the buffer gas become important.<sup>19</sup> The trajectory method (TM) gives the most reliable estimate, incorporating all interactions, but is computationally intense. A compromise is to use a third model, the exact hard sphere scattering (EHSS) method. This ignores electrostatic interactions so requires substantially less computational time, and can calculate cross-sections to within a few percent of values obtained by the TM.<sup>19</sup>

Here PA, EHSS and TM approximations were calculated from available X-ray and NMR structures held at the RCSB Protein Data Bank<sup>20</sup> for all proteins studied. All computations were performed on a Mac Pro with two Dual Core Intel Xeon 3.0GHz processors running Mac OS X 10.4.

Each TM approximation took 7 to 10 days to run on this system.

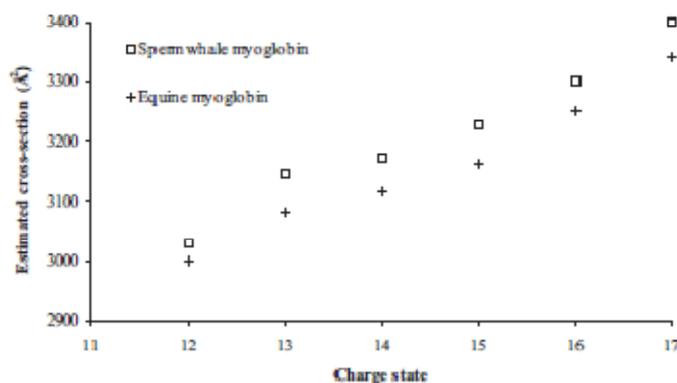
## RESULTS

Theoretical cross-sections calculated using the PA, EHSS and TM models in MobCal from X-ray and NMR structures for the proteins standards studied are shown in Table 1. Cross-sections estimated from X-ray and NMR structures for the same protein are within 3% of each other. Nominal differences (approximately 2%) are observed in the cross-sections calculated for the same protein from different organisms, which is not unexpected given their high sequence identity. Cross-sections calculated via the EHSS and TM approximations are in good agreement, to within a couple of percent, as expected.

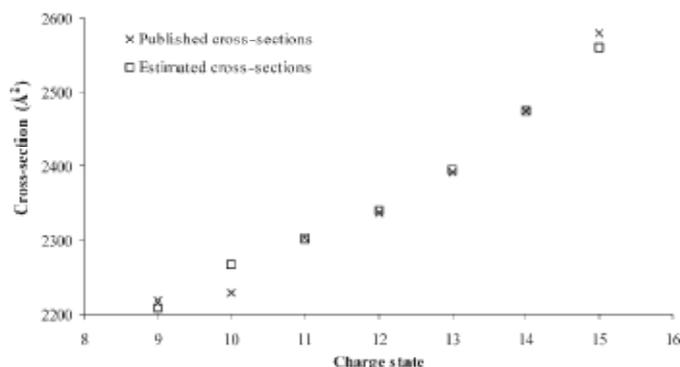
The correlation between estimated cross-sections for sperm whale myoglobin and equine myoglobin obtained for various charge states under denaturing conditions is illustrated in Fig. 2. For simplicity the charge states for which only one conformation was observed are shown. Equine myoglobin has a molecular weight of 16951.5 Da and sperm whale myoglobin, in its naturally occurring form, a

**Table 1.** Theoretical cross-sections for standard proteins calculated from published NMR and X-ray structures using the MobCal program and the PA, EHSS and TM models

Description	Protein			Theoretical cross-section ( $\text{\AA}^2$ )		
	Swiss-Prot ID	Experimental method	Pdb file	PA	EHSS	TM
Chicken lysozyme c	LYSC_CHICK	X-ray	1DPX	1180	1475	1499
		NMR	1GXX	1172	1451	1445
Human lysozyme c	LYSC_HUMAN	X-ray	2NWD	1218	1523	1551
		NMR	1IY3	1244	1568	1563
Sperm whale myoglobin	MYG_PHYCA	X-ray	1VXG	1375	1734	1730
		NMR	1MYF	1394	1754	1753
Equine myoglobin	MYG_HORSE	X-ray	1WLA	1366	1719	1716
Equine cytochrome c	CYC_HORSE	X-ray	1HRC	1055	1318	1313
		NMR	1LC1	1065	1324	1314



**Figure 2.** Comparison of estimated cross-sections for sperm whale myoglobin and equine myoglobin obtained under denaturing conditions.



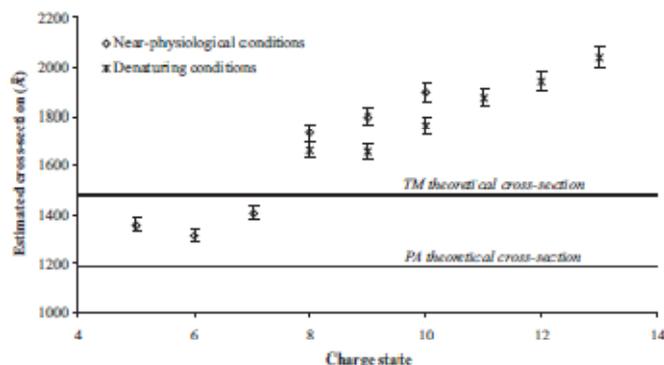
**Figure 3.** Comparison of published with experimentally estimated cross-sections for different charge states of equine cytochrome c (under denaturing conditions). Cross-sections were estimated utilising a calibration of data obtained for sperm whale myoglobin against published values.

molecular weight of 17199.9 Da. The sperm whale material used here was sourced from Sigma-Aldrich as a recombinant product, expressed in *E. coli*, which has an N-terminal methionine present giving a molecular weight of 17331.1 Da. The equine and sperm whale myoglobins have an 87% sequence identity.<sup>21</sup> The cross-sections estimated for the various charge states of the two proteins are similar but distinct; consistently we estimate equine myoglobin to have a smaller cross-section than sperm whale myoglobin, in agreement with theoretical calculations. This result indicates that these two proteins would not be interchangeable for calibration purposes since the differences in cross-section observed lie at the extremes of the experimental error of the calibration.

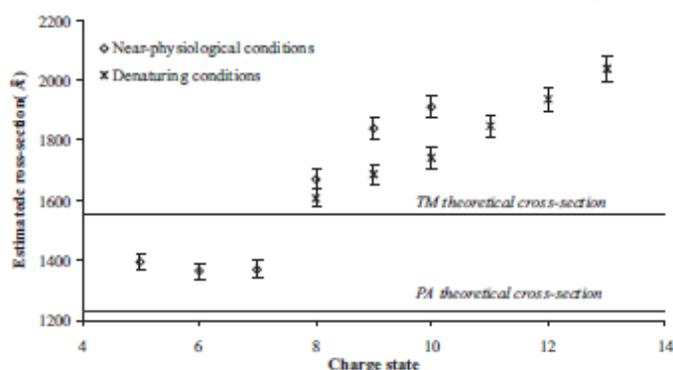
The experimentally estimated cross-sections for equine cytochrome c under denaturing conditions are shown in Fig. 3 and are compared to published values for DTIMS experiments.<sup>6</sup> The estimated cross-sections are within 1% of published values, except for the  $[M+10H]^{10+}$  charge state.

This result has been shown to be extremely reproducible having been repeated regularly over several months, under multiple experimental conditions. These results demonstrate the reproducibility and selectivity of the approach; it is the potential of the technique to probe biologically relevant conformations that is of the most prospective interest. In order to evaluate this aspect of the technique, estimated cross-sections for protein standards under denaturing and near-native conditions for different charge states have been obtained and compared.

Figures 4 and 5 show estimated cross-sections of lysozyme c obtained under both denaturing and near-physiological conditions. Under physiological conditions conformations with significantly smaller cross-sections are observed for charge states  $[M+5H]^{5+}$  to  $[M+7H]^{7+}$ , in comparison with charge states  $[M+8H]^{8+}$  to  $[M+10H]^{10+}$ , and these values fall between the PA and TM calculated values. These conformations have extremely similar cross-sections suggesting that these three charge states are stable within the gas phase.



**Figure 4.** Experimentally estimated cross-sections for different charge states of chicken lysozyme c under denaturing and near-physiological conditions with average theoretical cross-sections for PA and TM models shown.



**Figure 5.** Experimentally estimated cross-sections for different charge states of human lysozyme c under denaturing and near-physiological conditions with average theoretical cross-sections for PA and TM models shown.

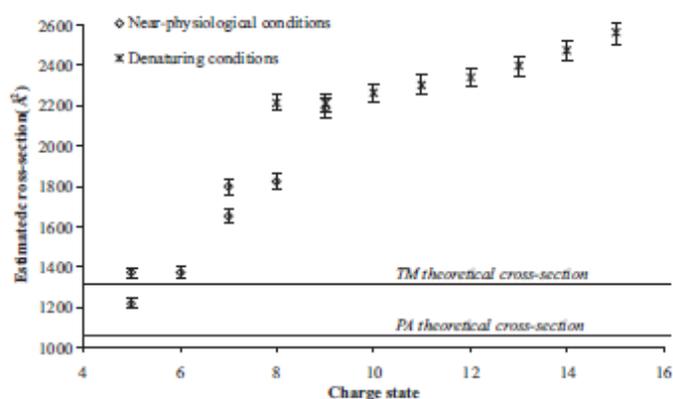
Conformations determined under both denaturing and near-physiological conditions for charge states  $[M+8H]^{8+}$  to  $[M+10H]^{10+}$  are significantly different. This may be attributable to the different solvent conditions used.

Lysozyme c possesses four disulphide bonds, which are maintained in the gas phase under denaturing as well as near-physiological conditions. Despite lysozyme c having a higher molecular weight than cytochrome c, cross-sections calculated for lysozyme c under denaturing conditions are smaller than those observed for cytochrome c (Fig. 6). Fewer charge states are also observed within the lysozyme c spectrum and the change in cross-section measured between the most compact structure and least compact structure is much greater for cytochrome c. Multiple conformations for some charge states are observed for cytochrome c and both myoglobins (Figs. 7 and 8) and conformations observed under near-physiological conditions vary more significantly in cross-section than those observed for lysozyme c.

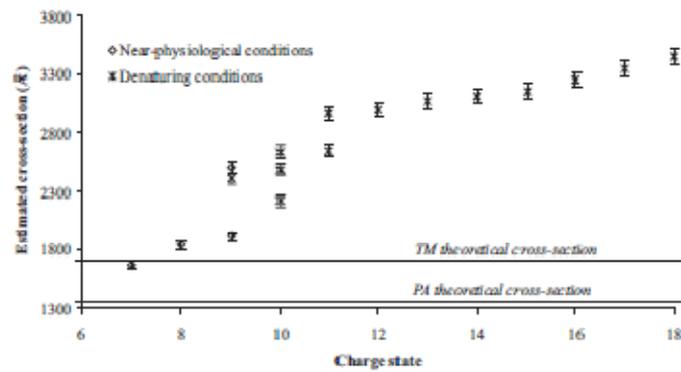
The change in cross-section with increase in charge is greater for myoglobin than for cytochrome c. The heme group contained within the native structure of these proteins is covalently bound within cytochrome c, but not within myoglobin. The heme group is lost when myoglobin is analysed under denaturing conditions but retained within cytochrome c. The presence of the heme group potentially stabilises more compact gas-phase conformations.

The estimates of cross-section obtained for the protein standards, based on measurements conducted over a number of months, are reproducible between datasets to within  $\pm 2\%$ . This estimated error is indicated in Figs. 4–8 as error bars.

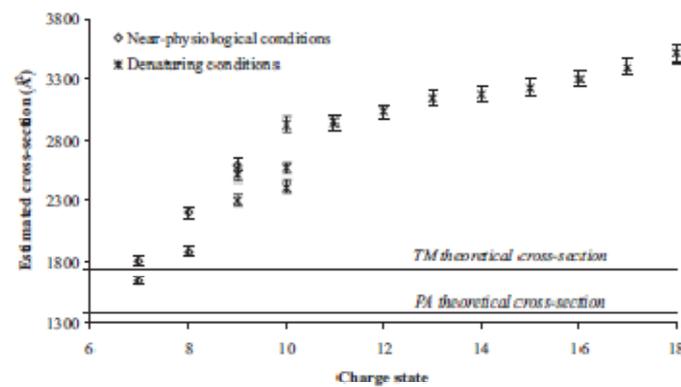
For each protein studied the lowest charge state observed has an estimated cross-section between the theoretical approximations PA and TM (Fig. 9). The estimated native cross-sections for the two myoglobins lie closer to the TM approximations while the cross-sections estimated for the lysozyme c proteins sit between the PA and TM calculations.



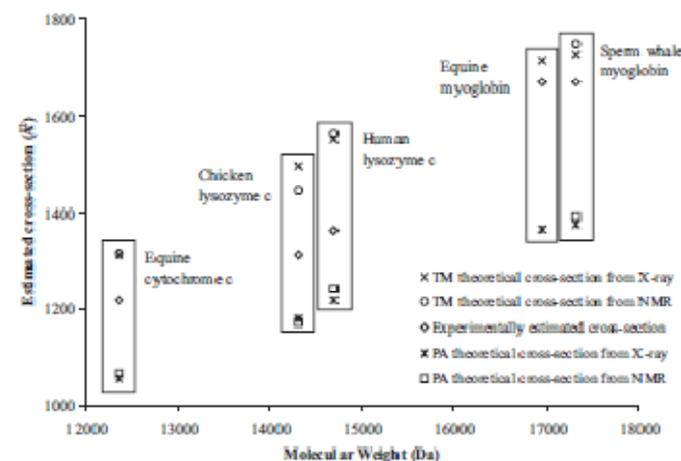
**Figure 6.** Experimentally estimated cross-sections for different charge states of equine cytochrome c under denaturing and near-physiological conditions with average theoretical cross-sections for PA and TM models shown.



**Figure 7.** Experimentally estimated cross-sections for different charge states of equine myoglobin under denaturing and near-physiological conditions with average theoretical cross-sections for PA and TM models shown.



**Figure 8.** Experimentally estimated cross-sections for different charge states of sperm whale myoglobin under denaturing and near-physiological conditions with average theoretical cross-sections for PA and TM models shown.



**Figure 9.** Estimations of native cross-sections for proteins studied calculated theoretically and experimentally and plotted against molecular weight.

Theoretical cross-sections calculated from X-ray crystallographic data and NMR structures show remarkable agreement, as has been observed previously.<sup>19</sup>

A comparison of cross-sections estimated for multiple charge states of equine myoglobin and sperm whale myoglobin under denaturing conditions shows curves of similar shape but with a clear difference in cross-sections of approximately 2% (Fig. 2). The observation of these small but notable differences between the same protein from different organisms, observed also for chicken and human lysozyme c, reflects the powerful resolving ability of this technique. It should also be noted that the same protein from different organisms cannot be assumed to have cross-sections of sufficient similarity for them to be interchangeable for calibration purposes.

Utilising the calibration created from sperm whale myoglobin data, equine cytochrome c cross-sections are estimated to within 2% of absolute values. This, along with previous studies,<sup>14</sup> provides substantial evidence that the Synapt TWIMS system can be calibrated successfully by this method. Measurements for equine cytochrome c are within 1% of published values, with the exception of the  $[M+10H]^{10+}$  charge state. This anomaly has been observed on various occasions and, given the agreement with the rest of the charge states, we believe that this may be caused by the conformation observed in the TWIMS experiment being different from that observed and reported in the DTIMS instrument.

Cross-sections estimated for the lowest charge states observed under physiological conditions are in good agreement with theoretical measurements calculated from published X-ray and NMR structures. Multiply charged ions produced by ESI are primarily a result of proton attachment to exposed basic sites of the protein. Our results (Fig. 9) support the hypothesis that the lowest charge states (under near-native conditions) are representative of the native protein structure due to a tightly folded conformation having fewer exposed basic sites than an unfolded conformation of the same protein.<sup>4</sup>

Estimated cross-sections for the lowest charge states are smaller than the TM approximation values. This is in agreement with previous work which has concluded that gas-phase conformations can be more compact than the crystal structure.<sup>2</sup> In the gas phase, it is proposed that intramolecular interactions become more dominant and make the polar side chains collapse onto the protein surface. This results in the protein adopting a more compact conformation.<sup>6</sup>

Ion mobility mass spectrometry is able to clearly resolve differences in protein conformation under denaturing and near-native solvent conditions. Different conformations for the same charge state of lysozyme c are witnessed under the two solvent conditions investigated. The differences observed are likely to be due to solvent interactions. Conformations observed under near-physiological conditions change in cross-section less significantly with increase in charge than those observed under denaturing conditions.

The degree of protein unfolding can be indicated by the increase in estimated cross-section with increase in charge

state. For the proteins investigated, under denaturing conditions, these curves show significant similarity. The unfolding transition is driven by Coulomb repulsion and is thought to be similar to acid denaturation in solution.<sup>19</sup> The change in cross-section is affected, however, by the presence of intact disulphide bonds.<sup>2</sup> The unfolding rate of lysozyme c is smaller than for the other proteins studied. Analysis of disulphide-reduced lysozyme c under the same experimental conditions will provide more information regarding the role of the disulphide bond in maintaining tertiary structure.

Differences are also seen between the unfolding patterns of myoglobin and cytochrome c. The unfolding of the three-dimensional structure of myoglobin with increase in charge state appears to progress more quickly than that of cytochrome c, providing evidence that the heme group stabilises gas-phase protein conformation. This is in agreement with early work by Eliezer and Wright which suggested that the removal of heme from myoglobin highly destabilises the structure of the globin fold,<sup>22</sup> and supports previous DTIMS studies conducted by Shelimov and Jarrold.<sup>8</sup>

Under near physiological conditions, multiple conformations are observed for the same charge state for myoglobin and cytochrome c but not for lysozyme c. This suggests that the structures of myoglobin and cytochrome c are less constrained in the gas phase and thus are able to adapt multiple stable conformations whereas the disulphide-bond-constrained lysozyme c has a less flexible existence. The multiple conformations observed here are consistent with those observed previously, and in greater number in DTIMS studies, for cytochrome c equine<sup>9</sup> and bovine,<sup>6,23</sup> and equine apomyoglobin.<sup>8</sup>

The experimental time period is crucial to the maintenance of solution-phase structure as after an extended period it is known that gas-phase structures can rearrange to account for their change in environment. Badman *et al.* have shown evidence for the unfolding and refolding of cytochrome c ions in the gas phase with transition from compact to unfolded structures occurring within 30–60 ms, and unfolded structures refolding into an array of folded structures after ions are trapped for times exceeding 250 ms.<sup>24</sup> The TWIMS experiment operates within a 15–30 ms time frame. This limits the opportunity for the solution-phase structures to rearrange within the gas phase.

## CONCLUSIONS

This work illustrates that TWIMS can be used to investigate the three-dimensional structure of proteins and the stability of that structure within the gas phase. The T-Wave device is able to operate at biologically relevant sample concentrations and its speed facilitates its use in an automated fashion. The considerable information provided by the graphs of estimated cross-sections against charge state can allow the unfolding of proteins in the gas phase to be studied. TWIMS can also be utilised to probe conformational changes in protein structure and could potentially be used as a fast screen to confirm that proteins synthesised via recombinant methods have the same cross-section as the targeted, biologically active, conformation. The TWIMS approach is still in the early stages of development and it is likely that

further improvements in resolution of the technique will extend the approach into a fast, sensitive, information-rich method for studying three-dimensional protein structure.

## REFERENCES

1. Heck AJR, van den Heuvel RHH. *Mass Spectrom. Rev.* 2004; **23**: 368.
2. Hoaglund-Hyzer CS, Counterman AE, Clemmer DE. *Chem. Rev.* 1999; **99**: 3037.
3. Kaltashov IA, Eyles SJ. *Mass Spectrom. Rev.* 2002; **21**: 37.
4. Chowdhury SK, Katta V, Chait BT. *J. Am. Chem. Soc.* 1990; **112**: 9012.
5. Mohimen A, Dobo A, Hoemer JK, Kaltashov IA. *Anal. Chem.* 2003; **75**: 4139.
6. Shelimov KB, Clemmer DE, Hudgins RR, Jarrold MF. *J. Am. Chem. Soc.* 1997; **119**: 2240.
7. Myung S, Badman ER, Lee YJ, Clemmer DE. *J. Phys. Chem. A* 2002; **106**: 9976.
8. Shelimov KB, Jarrold MF. *J. Am. Chem. Soc.* 1997; **119**: 2987.
9. Badman ER, Hoaglund-Hyzer CS, Clemmer DE. *Anal. Chem.* 2001; **73**: 6000.
10. Mesleh MF, Hunter JM, Shvartsburg AA, Schatz GC, Jarrold MF. *J. Phys. Chem.* 1996; **100**: 16082.
11. Shvartsburg AA, Jarrold MF. *Chem. Phys. Lett.* 1996; **261**: 86.
12. Giles K, Pringle SD, Worthington KR, Little D, Wildgoose JL, Bateman RH. *Rapid Commun. Mass Spectrom.* 2004; **18**: 2401.
13. Pringle SD, Giles K, Wildgoose JL, Williams JP, Slade SE, Thalassinou K, Bateman RH, Bowers MT, Scrivens JH. *Int. J. Mass Spectrom.* 2007; **261**: 1.
14. Wildgoose JL, Giles K, Pringle SD, Koeniger SJ, Valentine RH, Bateman RH, Clemmer DE. *Proc. 54th ASMS Conf. Mass Spectrometry and Allied Topics*, Seattle, 2006.
15. Ruotolo BT, Giles K, Campuzano I, Sandercocock AM, Bateman RH, Robinson CV. *Science* 2005; **310**: 1658.
16. Clemmer DE. Available: [http://www.indiana.edu/~clemmer/Research/cross%20section%20database/Proteins/protein\\_cs.htm](http://www.indiana.edu/~clemmer/Research/cross%20section%20database/Proteins/protein_cs.htm) (last accessed: 30 April 2008).
17. Scrivens JH, Thalassinou K, Hilton G, Slade SE, Pinheiro TTT, Bateman RH, Bowers MT. *Proc. 55th ASMS Conf. Mass Spectrometry and Allied Topics*, Indianapolis, 2006.
18. Williams JP, Scrivens JH. *Rapid Commun. Mass Spectrom.* 2008; **22**: 187.
19. Jarrold MF. *Acc. Chem. Res.* 1999; **32**: 360.
20. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. *Nucleic Acids Res.* 2000; **28**: 235.
21. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. *Nucleic Acids Res.* 1990; **18**: 3389.
22. Eliezer D, Wright PE. *J. Mol. Biol.* 1996; **263**: 531.
23. Clemmer DE, Hudgins RR, Jarrold MF. *J. Am. Chem. Soc.* 1995; **117**: 10141.
24. Badman ER, Myung S, Clemmer DE. *J. Am. Soc. Mass Spectrom.* 2005; **16**: 1493.

---

---

# Probing Hemoglobin Structure by Means of Traveling-Wave Ion Mobility Mass Spectrometry

Charlotte A. Scarff, Vibhuti J. Patel, Konstantinos Thalassinou, and James H. Scrivens

Department of Biological Sciences, University of Warwick, Coventry, United Kingdom

---

Hemoglobin (Hb) is a tetrameric noncovalent complex consisting of two  $\alpha$ - and two  $\beta$ -globin chains each associated with a heme group. Its exact assembly pathway is a matter of debate. Disorders of hemoglobin are the most common inherited disorders and subsequently the molecule has been extensively studied. This work attempts to further elucidate the structural properties of the hemoglobin tetramer and its components. Gas-phase conformations of hemoglobin tetramers and their constituents were investigated by means of traveling-wave ion mobility mass spectrometry. Sickle (HbS) and normal (HbA) hemoglobin molecules were analyzed to determine whether conformational differences in their quaternary structure could be observed. Rotationally averaged collision cross sections were estimated for tetramer, dimer, apo-, and holo-monomers with reference to a protein standard with known cross sections. Estimates of cross section obtained for the tetramers were compared to values calculated from X-ray crystallographic structures. HbS was consistently estimated to have a larger cross section than that of HbA, comparable with values obtained from X-ray crystallographic structures. Nontetrameric species observed included apo- and holo- forms of  $\alpha$ - and  $\beta$ -monomers and heterodimers;  $\alpha$ - and  $\beta$ -monomers in both apo- and holo- forms were found to have similar cross sections, suggesting they maintain a similar fold in the gas phase in both the presence and the absence of heme. Heme-deficient dimer, observed in the spectrum when analyzing commercially prepared Hb, was not observed when analyzing fresh blood. This implies that holo- $\alpha$ -apo- $\beta$  is not an essential intermediate within the Hb assembly pathway, as previously proposed. (J Am Soc Mass Spectrom 2009, 20, 625–631) © 2009 Published by Elsevier Inc. on behalf of American Society for Mass Spectrometry

---

---

Mass spectrometry (MS) has become an important tool for the study of various aspects of protein structure, including the assembly and disassembly of protein complexes, subunit interactions, and ligand interactions [1]. The study of noncovalent complexes has been facilitated by the use of collisional cooling or dampening of ions by elevating the pressure within the source region or by deceleration [2, 3].

The typical electrospray ionization (ESI) mass spectrum of a protein consists of an envelope of peaks attributed to a series of multiply charged gas-phase ions that can indicate the stability and compactness of its structure in the gas phase [4, 5]. Multiply charged ions are produced by proton attachment, predominantly to exposed basic sites on the protein [6], and those of lowest charge are thought to be most representative of native structure [7]. A highly compact protein would have fewer exposed basic sites than those of an unfolded conformation of the same protein and thus would accept fewer charges [4, 8].

Ion mobility spectrometry is a shape-selective technique, based on the time taken for an ion to traverse a mobility cell containing an inert gas under the influence of a weak electric field [9]. This time is related to the rotationally averaged collision cross section, mass, and charge of the ion. The coupling of ion mobility separation with mass spectrometry has provided a powerful method for the analysis of complex mixtures and for the determination of molecular structure [10].

Ion mobility mass spectrometry has emerged as a complementary technique to the well-established methods of X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy for three-dimensional protein structure analysis [11]. There is now substantial evidence to support the view that the gas-phase protein structure can reflect, under controlled conditions, the native solution-phase structure [1, 6, 12]. Multiple studies have shown good agreement between rotationally averaged cross-sectional measurements obtained from X-ray and NMR structures and those obtained by ion mobility experiments [7, 13–17].

In this study traveling-wave ion mobility mass spectrometry (TWIMS) was used to probe the gas-phase conformations of hemoglobin tetramers and their constituents.

---

Address reprint requests to Dr. James H. Scrivens, University of Warwick, Department of Biological Sciences, Coventry CV4 7AL, UK. E-mail: j.h.scrivens@warwick.ac.uk

A traveling-wave ion mobility separator (T-wave) [18] has been incorporated into a quadrupole time-of-flight (TOF) instrument, the Synapt HDMS system (Waters Corp., Milford, MA, USA) [9]. The T-wave allows for samples to be analyzed at biologically relevant concentrations, having speed and sensitivity advantages over the traditional drift-time ion mobility device. The T-wave device does not allow for absolute cross-sectional measurements to be obtained from drift-time information, although these may be estimated by using reference samples of known cross sections [7, 13, 19].

Hemoglobin (Hb) is a tetramer consisting of four globin chains, two  $\alpha$ - and two  $\beta$ -, each associated with a heme group. It is the major oxygen-transport protein found in the red blood cells of all vertebrates. Disorders of hemoglobin are the most common of all inherited disorders and, consequently, the molecule has been extensively studied.

The most debilitating Hb variant is that which causes sickle-cell anemia. This disease occurs when a person inherits two particular mutated copies of the  $\beta$ -globin gene. The sickle-cell mutation results in the production of a  $\beta$ -chain with a single amino acid substitution ( $\beta$  Glu  $\rightarrow$  Val) and changes the conformation of the assembled tetramer to allow molecular stacking. Polymerization of this sickle-cell hemoglobin molecule (HbS), in deoxygenated blood, causes the characteristic alteration in shape of red blood cells from biconcave discs to crescentic [20].

ESI-MS has been widely used to detect Hb variants in hemoglobin [21–25] and to investigate its structural assembly into a noncovalent complex [26–29] and its corresponding disassembly [30].

The exact assembly pathway for hemoglobin is still under debate. It is known that one  $\alpha$ - and one  $\beta$ -monomer come together to form a heterodimer and that two of these dimers associate to form the tetramer. Since  $\alpha$ - and  $\beta$ -monomers can exist in heme-free (apo,  $\alpha^a$  and  $\beta^a$ ) and heme-bound (holo,  $\alpha^h$  and  $\beta^h$ ) forms [28], it is unclear as to whether the heme groups are attached to both  $\alpha$ - and  $\beta$ -monomers before dimer formation or whether association leads to heme recruitment.

Griffith and Kaltashov have suggested that the formation of a heme-deficient dimer intermediate ( $\alpha^h\beta^a$ ) occurs, consisting of a natively folded holo- $\alpha$ -globin ( $\alpha^h$ ) and a partially unfolded apo- $\beta$ -globin ( $\beta^a$ ), before complete dimer formation, to ensure correct tetramer structure arrangement [26, 27]. The Konermann group, however, reported that the heme-deficient dimer seen consistently when using commercially available samples, in the form of lyophilized powder, was not observed when using freshly prepared samples [31]. They studied the acid-induced denaturation of bovine Hb and concluded that it followed a highly symmetric mechanism:  $(\alpha^h\beta^h)_2 \rightarrow 2\alpha^h\beta^h \rightarrow 2\alpha^h_{\text{folded}} + 2\beta^h_{\text{folded}} \rightarrow 2\alpha^a_{\text{unfolded}} + 2\beta^a_{\text{unfolded}} + 4\text{ heme}$ .

This work attempts to further elucidate the structural properties of the hemoglobin tetramer and its compo-

nents and to determine whether conformational differences between the HbA and HbS molecules can be observed by TWIMS.

## Experimental

### Samples and Sample Preparation

Samples of fresh whole blood were supplied by University Hospitals Coventry and Warwickshire NHS Trust. Sample preparation for mass spectral analysis was adapted from that detailed by Ofori-Acquah et al. [29]. Samples (20  $\mu$ L) were diluted 10-fold in 10 mM ammonium acetate (pH 6.8) and spun at 3000  $g$  for 15 min in centrifugal filter units with a 10-kDa cutoff (Microcon YM-10, Millipore Corp., Billerica, MA, USA). Sample retained on the filter was diluted a further 20-fold with 10 mM ammonium acetate and desalted by agitating for two 10-min periods with close to 5 mg of ion-exchange mixed bed resin (AG 501-X8, Bio-Rad Laboratories, Hercules, CA, USA) that had been prepared for use by rinsing twice in liquid chromatography MS grade water. The resulting solutions were introduced into the ESI source of a Synapt HDMS System (Waters) by means of fused silica nanospray needles. All solvents and calibration and protein standards were obtained from Sigma-Aldrich (St. Louis, MO, USA).

### Mass Spectral Analysis

Data were acquired by means of a Synapt HDMS system in ESI positive mode with a capillary voltage of 1.2 kV from 1000 to 4500  $m/z$ . The TOF mass analyzer was calibrated using 2 mg/mL cesium iodide in 50% aqueous propan-2-ol.

Instrument acquisition parameters were adjusted to provide the optimal ion mobility separation. The cone voltage was 60 V and the collision energy in the trap region was 10 eV. Source temperature and gas flow were 110 °C and 35 mL/min, respectively. Nitrogen was used as the gas in the ion mobility cell and the indicated pressure within the cell was 0.68 mbar, equivalent to a flow rate of 38 mL/min. The backing pressure was increased in increments from 2 to 8 mbar to identify the ideal pressure conditions for transmission of the relevant ionic species. The traveling-wave velocity and wave height were altered in increments from 100 to 600 m/s and 8 to 20 V, respectively, and the conditions that provided the optimal mobility separation were used for all following experiments.

The synchronization of gated release of ions into the ion mobility separator with TOF acquisition allows arrival time distributions of ions to be obtained. For each gate pulse, 200 orthogonal acceleration pushes of the TOF analyzer are recorded to form one ion mobility experiment. The overall mobility recording time is  $200 \times t_p$ , where  $t_p$  is the pusher period [9]. The pusher period depends on the mass acquisition range; for these

experiments, a pusher period of 120  $\mu$ s was used, giving a mobility recording time of 24 ms.

Equine myoglobin at a concentration of 10  $\mu$ M in 50% aqueous acetonitrile containing 0.2% formic acid was used to provide data that were used to create a calibration curve for cross-sectional measurements.

Data obtained for each hemoglobin tetramer over the  $m/z$  range 3000–4500 were deconvoluted onto a true mass scale using maximum entropy modeling (MaxEnt for short) software to provide an estimate of molecular mass. Experiments were carried out in triplicate.

#### Calibration, Modeling, and Estimation of Cross Section

The equine myoglobin data were used to create a calibration curve for each set of experiments. Absolute cross sections for equine myoglobin were obtained from drift-time ion mobility mass spectrometry (DTIMS) studies (Prof. Michael T. Bowers, personal communication). The calibration was performed using a procedure developed in-house based on previously published work [13, 19, 32, 33]. In brief, normalized cross sections (corrected for charge and reduced mass) were plotted against corrected arrival times (corrected to exclude time spent outside the ion mobility cell) to create a calibration with a power-series fit. The calibration allows one to estimate the cross section of a molecule of interest provided that the mobilities (corrected arrival times) for that molecule lie within the mobilities observed for the calibrant used, irrespective of the size range of cross sections for the calibrant [34, 35]. The calibration was used to estimate rotationally averaged collision cross sections of hemoglobin monomer, dimer, and tetramer for the different charge states observed, based on their arrival time distributions, provided that their corrected arrival times fell along the calibration curve.

To compare the experimental cross sections for the normal and sickle hemoglobin tetramers with accepted values, cross sections were calculated using MOBCAL, a program to calculate mobilities [36, 37], from published X-ray structures held at the RCSB Protein Data Bank [38].

MOBCAL facilitates the use of three approximations to calculate cross sections. The projection approximation (PA) typically results in an underestimation of the cross section of a large ion. It calculates the cross section by averaging the projections produced by every orientation of a molecule and so does not take into account interactions with the buffer gas. The trajectory method (TM) takes into account all interactions, but is computationally intense. The exact hard-sphere scattering model (EHSS) carries out trajectory calculations, while ignoring long-range interactions, but nevertheless gives values within a few percent of the TM approximations [7, 39]. For this work, cross sections were calculated using

the PA and EHSS methods to reduce computational time.

## Results and Discussion

### Instrument Acquisition Parameters

Considerable optimization of instrument acquisition parameters is required for each individual application of ion mobility separation. This must be tailored to the sample of interest because optimal conditions are dependent on ionic species and mass-to-charge ratio [3]. Controlled optimization of instrument acquisition parameters indicated that a backing pressure of between 6.6 and 6.8 mbar was ideal for intact hemoglobin tetramer analysis. The optimal ion mobility separation of the tetramer was achieved at a traveling-wave velocity and wave height of 400 m/s and 18 V, respectively.

### Calibration

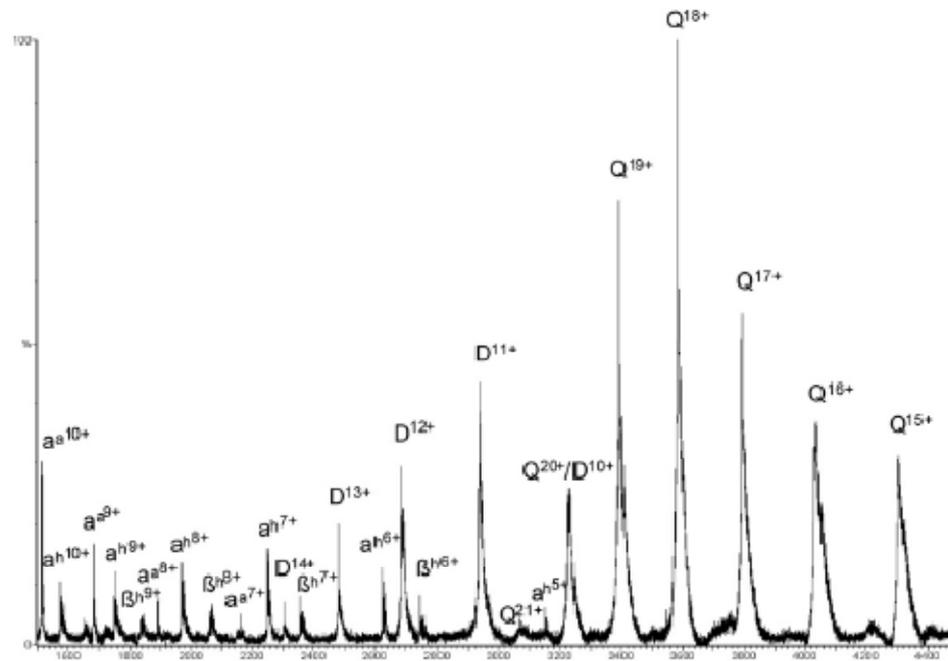
A calibration curve was used to allow the estimation of cross sections for different constituents of hemoglobin in different charge states. Cross sections calculated for equine myoglobin were within 2% of absolute values obtained by DTIMS experiments. These results were reproducible across the three datasets acquired.

### Hemoglobin Tetramer Analysis

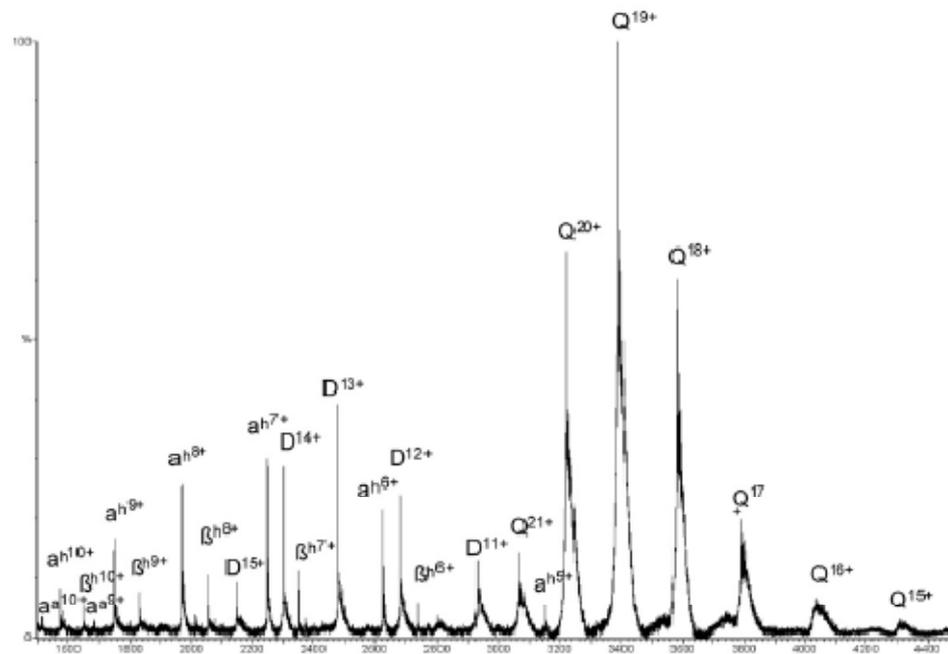
Representative spectra for normal (HbA) and sickle (HbS) hemoglobin analyzed by means of ESI-TOF-MS under nondenaturing conditions are shown in Figures 1a and b, respectively. The data were deconvoluted to give masses of 64,454.7 Da for HbA and 64,395.8 Da for HbS, which were very close to the theoretical masses of 64,453.2 and 64,393.4 Da, respectively [29].

The hemoglobin spectra obtained show the presence of the tetramer  $[(\alpha^h\beta^h)_2]$ , heterodimer  $(\alpha^h\beta^s)$ , and apo- and holo-monomer species. The trimer is not seen, as would be expected, because the formation of the hemoglobin tetramer involves the noncovalent association of two  $\alpha^h\beta^h$ -dimers. Carefully controlled near-physiological conditions were used in preparing the sample and the absence of any trimer implies that the species observed within the spectra exist naturally in solution. This is consistent with results from isotope labeling studies that showed that nontetrameric ions in the spectrum corresponded to species present in solution [40] rather than products of fragmentation formed during the ESI process [41].

Alpha and beta monomers are observed within the HbA spectrum in both apo- and holo- forms. In a previous study, Griffith and Kaltashov [26] suggested that an  $\alpha^h$  monomer first becomes associated with an  $\beta^s$  monomer, to enable the  $\beta$ -chain to incorporate the heme group. This observation was based on the absence of  $\beta^h$  in the spectrum. A subsequent study by Boys and Konermann [31] detected very small quantities of heme-



(a)



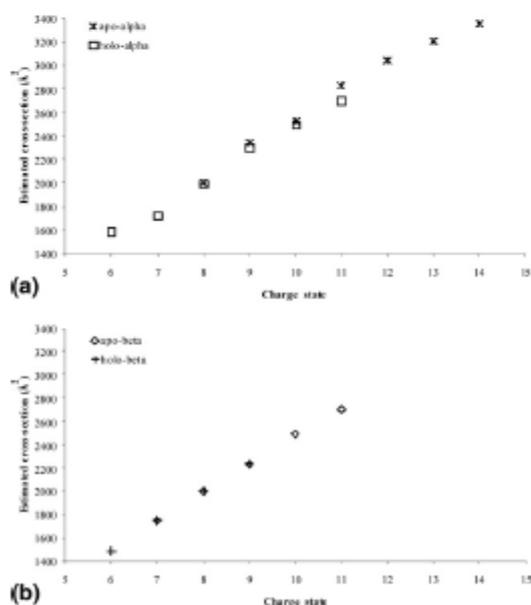
(b)

**Figure 1.** Mass spectra of (a) normal (HbA) and (b) sickle (HbS) hemoglobin analyzed by ESI-TOF-MS under near-physiological conditions, labeled with charge states of tetramer (Q), heterodimer (D), and apo- and holo-monomers (superscripts "a" and "h" refer to apo- and holo-forms, respectively).

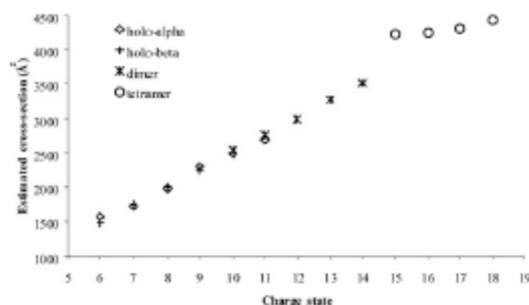
deficient dimer and found that both  $\alpha$ - and  $\beta$ -monomers were capable of binding heme. The discrepancies observed are thought to be attributable to differences between the commercially prepared and freshly obtained samples used. In the work reported here, in which fresh blood samples were used,  $\beta^h$  was observed in multiply charged states.

It has been reported [42] that without the attachment of the heme group,  $\alpha$ - and  $\beta$ -monomers adopt extensively unfolded conformations. Cross sections for various charge states of  $\alpha$ - and  $\beta$ -monomers in both the apo- and holo- forms have been estimated and our observations suggest that the predominant conformations of  $\alpha$ - and  $\beta$ -monomers in the gas phase are similar to each other and show little change in the absence or the presence of heme (Figure 2). The cross section of each of the molecules studied increases with an increase in charge, thought to be a result of the effects of Coulomb repulsion [16, 43].

The heme-deficient dimer observed in previous studies is not observed here. The existence of both apo- and holo- forms of  $\alpha$ - and  $\beta$ -monomers, all of similar cross sections, does not support the need for a  $\beta^a$  to associate with  $\alpha^h$  for the  $\beta$ -monomer to recruit heme. Analysis of commercially prepared human hemoglobin (data not shown) does show the presence of a heme-deficient dimer at a molecular mass 32 Da higher than expected; this is in agreement with previous work conducted by the Konermann group [31] on bovine hemoglobin, who attributed the additional mass to the occurrence of oxidative modifications in the commercial protein.



**Figure 2.** Average estimated cross sections for charge states of (a) apo- $\alpha$  and holo- $\alpha$  monomers and (b) apo- $\beta$  and holo- $\beta$  observed within three datasets.



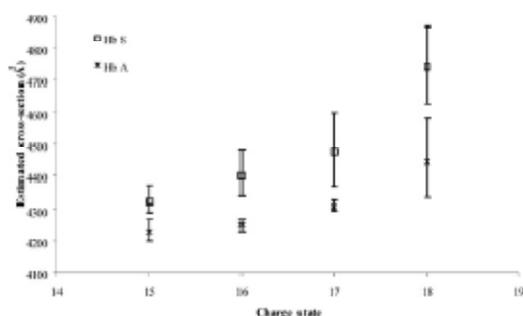
**Figure 3.** Average estimated cross sections for holo- $\alpha$ , holo- $\beta$ , heterodimer, and HbA tetramer, from three datasets.

We observe many more charge states of  $\alpha^a$  than of  $\beta^a$ . The number of charges accepted on a protein is related to the number of exposed basic sites on the protein's surface. A more folded protein has fewer of its basic sites exposed than an unfolded conformation and thus cannot accept as many charges. This may suggest that the  $\alpha$ -chain adopts more unfolded conformations in the gas phase than is possible for the  $\beta$ -chain but, alternatively, the absence of higher charge states of  $\beta^a$  in the spectra may be ascribed to the different desolvation behavior of  $\alpha$ - and  $\beta$ -monomers. The  $\alpha$ -chain ionizes preferentially over the  $\beta$ -chain because of its greater nonpolar character, thereby competing more effectively for charge [41].

By estimating the cross sections of monomer, dimer, and tetramer, a picture of the assembly process can be obtained (Figure 3). The  $[M + 12H]^{12+}$  charge state of dimer has an estimated cross section of  $3001 \text{ \AA}^2$ . The  $[M + 6H]^{6+}$  charge states of  $\alpha^h$  and  $\beta^h$  have estimated average cross sections of  $1583$  and  $1488 \text{ \AA}^2$ , respectively. If these two globin monomers come together to form a dimer, one would expect that the cross section of that dimer would be approximately the addition of the cross sections of the two constituent parts and, indeed, that is the case here. One would further expect that the cross section observed would be slightly smaller than the sum of the monomer subunits because the contact area on both of the monomers would be compacted and contribute less to the overall cross section. The data are in agreement with this.

The average estimated collision cross sections for HbA and HbS, for four different charge states, are illustrated in Figure 4. The data indicate a difference in cross section between normal and sickle-cell hemoglobin and a variation in cross section with charge state.

For HbA and HbS for the charge states studied, the cross sections observed for HbS are somewhat larger than those of HbA. Secondary, tertiary, and quaternary structural considerations make it difficult to determine what the charge state of a molecule should be, theoretically, within a particular solvent at a particular pH. It is clear from previous work that the lowest charge states observed under near-physiological conditions are most



**Figure 4.** Estimated cross sections for HbA and HbS tetramers for three different charge states, showing averaged values from three datasets with corresponding errors.

representative of the native protein structure [7]. The  $[M + 18H]^{18+}$  charge state, for HbA and HbS, is suspected to be representative of a tetrameric structure that is beginning to denature.

The reproducibility of the cross sections estimated for the  $[M + 15H]^{15+}$ ,  $[M + 16H]^{16+}$ , and  $[M + 17H]^{17+}$  charge states of HbA, between the three replicate datasets, is  $\pm 1\%$ , believed to be representative of the reproducibility capabilities of the experiment. The larger deviation in estimation of cross section for the same charge states for HbS, of  $\pm 3\%$ , may reflect the presence of a broader population of conformations of the HbS molecule of similar cross section.

The rotationally averaged cross sections for HbA and HbS calculated from X-ray crystallographic structures were 3313 and 3733 Å<sup>2</sup> for the PA and 4343 and 4775 Å<sup>2</sup> for the EHSS, respectively. Values estimated experimentally for the  $[M + 15H]^{15+}$ ,  $[M + 16H]^{16+}$  and  $[M + 17H]^{17+}$  charge states of HbA and HbS fall between these two theoretical approximations and agree with the X-ray observation that HbS has a larger cross section than that of HbA. Gas-phase conformations, although illustrative of solution-phase structures under controlled conditions, have been shown previously to be smaller than those predicted by EHSS approximations [6]; a more compact conformation is thought to be adopted in the gas phase because increased intramolecular interactions cause polar side chains to collapse onto the protein's surface [14].

## Conclusions

This work demonstrates the use of TWIMS to probe gas-phase conformations of three-dimensional protein structure and noncovalent complexes.

TWIMS has been successfully used to analyze hemoglobin tetramers. Cross sections calculated for intact hemoglobin tetramers are comparable to those estimated from published X-ray crystallography data and conformational differences are observed between the HbA and HbS molecules. Nontetrameric species observed, including apo- and holo- forms of  $\alpha$ - and

$\beta$ -monomers and  $\alpha\beta$ -dimers, are naturally present in equilibrium in solution and are not products of fragmentation during the ESI process.

Both  $\alpha$ - and  $\beta$ -monomers have cross sections similar to each other, suggesting that they maintain a similar fold in the gas phase. Apo- and holo- forms of the monomers also have similar cross sections, suggesting that  $\alpha$ - and  $\beta$ -monomers can retain a folded structure in the absence and the presence of the heme group. Extensively disordered monomer structures are not observed.

A heme-deficient dimer has not been observed and the results do not suggest the requirement for association of  $\beta^2$  with  $\alpha^2$  for the  $\beta$ -monomer to recruit heme. The results, obtained on fresh blood samples rather than commercially prepared samples, do not support the hypothesis that a heme-deficient dimer is an essential intermediate in the tetramer assembly process.

## Acknowledgments

The authors thank Dr. Nicholas Jackson and Yvonne Elliot, University Hospitals Coventry and Warwickshire NHS Trust, for providing samples and Prof. Michael T. Bowers for providing cross-sectional measurements for equine myoglobin.

## References

1. Heck, A. J. R.; van den Heuvel, R. H. H. Investigation of Intact Protein Complexes by Mass Spectrometry. *Mass Spectrom. Rev.* **2004**, *23*, 368–389.
2. Chernushevich, I. V.; Thomson, B. A. Collisional Cooling of Large Ions in Electrospray Mass Spectrometry. *Anal. Chem.* **2004**, *76*, 1754–1760.
3. Tahallah, N.; Pinkse, M.; Maier, C. S.; Heck, A. J. R. The Effect of the Source Pressure on the Abundance of Ions of Noncovalent Protein Assemblies in an Electrospray Ionization Orthogonal Time-of-Flight Instrument. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 596–601.
4. Chowdhury, S. K.; Katta, V.; Chait, B. T. Probing Conformational Changes in Proteins by Mass Spectrometry. *J. Am. Chem. Soc.* **1990**, *112*, 9012–9013.
5. Mohimen, A.; Dobo, A.; Hoerner, J. K.; Kaltashov, I. A. A Chemometric Approach to Detection and Characterization of Multiple Protein Conformers in Solution Using Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **2003**, *75*, 4139–4147.
6. Hoaglund-Hyzer, C. S.; Counterterman, A. E.; Clemmer, D. E. Anhydrous Protein Ions. *Chem. Rev.* **1999**, *99*, 3037–3080.
7. Scarff, C. A.; Thalassinou, 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.* **2008**, *22*, 3297–3304.
8. Kaltashov, I. A.; Eyles, S. J. Studies of Biomolecular Conformations and Conformational Dynamics by Mass Spectrometry. *Mass Spectrom. Rev.* **2002**, *21*, 37–71.
9. Pringle, S. D.; Giles, K.; Wildgoose, J. L.; Williams, J. P.; Slack, S. E.; Thalassinou, K.; Bateman, R. H.; Bowers, M. T.; Scrivens, J. H. An Investigation of the Mobility Separation of Some Peptide and Protein Ions Using a New Hybrid Quadrupole/Travelling Wave IMS/oa-ToF Instrument. *Int. J. Mass Spectrom.* **2007**, *261*, 1–12.
10. Kanu, A. B.; Dwivedi, P.; Tam, M.; Matz, L.; Hall, H. H. J. Ion Mobility-Mass Spectrometry. *J. Mass Spectrom.* **2008**, *43*, 1–22.
11. van der Heuvel, R. H.; Heck, A. J. Native Protein Mass Spectrometry: From Intact Oligomers to Functional Mechanisms. *Curr. Opin. Chem. Biol.* **2004**, *8*, 519–526.
12. Kaddis, C. S.; Lormel, S. H.; Yin, S.; Berhane, B.; Apostol, M. I.; Kickhoefer, V. A.; Rome, L. H.; Loo, J. A. Sizing Large Proteins and Protein Complexes by Electrospray Ionization Mass Spectrometry and Ion Mobility. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 1205–1216.
13. Ruotolo, B. T.; Giles, K.; Campuzano, I.; Sandercock, A. M.; Bateman, R. H.; Robinson, C. V. Evidence for Macromolecular Protein Rings in the Absence of Bulk Water. *Science* **2005**, *310*, 1658–1661.
14. Shellmov, K. B.; Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. Protein Structure In Vacuo: Gas-Phase Conformations of BPTI and Cytochrome c. *J. Am. Chem. Soc.* **1997**, *119*, 2240–2248.

15. Shelimov, K. B.; Jarrold, M. F. Conformations, Unfolding, and Refolding of Apomyoglobin in Vacuum: An Activation Barrier for Gas-Phase Protein Folding. *J. Am. Chem. Soc.* **1997**, *119*, 2987–2994.
16. Valentine, S. J.; Anderson, J. G.; Ellington, A. D.; Clemmer, D. E. Disulfide-Intact and -Reduced Lysozyme in the Gas Phase: Conformations and Pathways of Folding and Unfolding. *J. Phys. Chem. B* **1997**, *101*, 3891–3900.
17. Myung, S.; Badman, E. R.; Lee, Y. J.; Clemmer, D. E. Structural Transitions of Electrosprayed Ubiquitin Ions Stored in an Ion Trap over  $\sim 10$  ms to 30 s. *J. Phys. Chem. A* **2002**, *106*, 9976–9982.
18. Giles, K.; Pringle, S. D.; Worthington, K. R.; Little, D.; Wildgoose, J. L.; Bateman, R. H. Applications of a Travelling Wave-based Radio-Frequency-Only Stacked Ring Ion Guide. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 2401–2414.
19. Wildgoose, J. L.; Giles, K.; Pringle, S. D.; Kooniger, S. J.; Valentino, R. H.; Bateman, R. H.; Clemmer, D. E. A Comparison of Travelling Wave and Drift Tube Ion Mobility Separations. In *Proceedings of the 54th ASMS Conference on Mass Spectrometry and Allied Topics*, Seattle, WA, 2006.
20. Murayama, M. Structure of Sickle Cell Hemoglobin and Molecular Mechanism of the Sickling Phenomenon. *Clin. Chem.* **1967**, *13*, 578–588.
21. Daniel, Y. A.; Turner, C.; Haynes, R. M.; Hunt, B. J.; Dalton, R. N. Rapid and Specific Detection of Clinically Significant Haemoglobinopathies Using Electrospray Mass Spectrometry-Mass Spectrometry. *Br. J. Haematol.* **2005**, *130*, 635–643.
22. Shackleton, C. H. L.; Falick, A. M.; Green, B. N.; Wittkowska, H. E. Electrospray Mass Spectrometry in the Clinical Diagnosis of Variant Hemoglobins. *J. Chromatogr. B Biomed. Appl.* **1991**, *562*, 175–190.
23. Wild, B. J.; Green, B. N.; Stephens, A. D. The Potential of Electrospray Ionization Mass Spectrometry for the Diagnosis of Hemoglobin Variants Found in Newborn Screening. *Blood Cells Mol. Dis.* **2004**, *33*, 308–317.
24. Wild, B. J.; Green, B. N.; Cooper, E. K.; Laloz, M. R.; Erten, S.; Stephens, A. D.; Layton, D. M. Rapid Identification of Hemoglobin Variants by Electrospray Ionization Mass Spectrometry. *Blood Cells Mol. Dis.* **2001**, *27*, 691–704.
25. Shimizu, A.; Nakanishi, T.; Miyazaki, A. Detection and Characterization of Variant and Modified Structures of Proteins in Blood and Tissues by Mass Spectrometry. *Mass Spectrom. Rev.* **2006**, *25*, 686–712.
26. Griffith, W. P.; Kaltashov, I. A. Highly Asymmetric Interactions between Globin Chains during Hemoglobin Assembly Revealed by Electrospray Ionization Mass Spectrometry. *Biochemistry* **2003**, *42*, 10024–10033.
27. Griffith, W. P.; Kaltashov, I. A. Protein Conformational Heterogeneity as a Binding Catalyst: ESI-MS Study of Hemoglobin H Formation. *Biochemistry* **2007**, *46*, 2020–2026.
28. Boys, B. L.; Konermann, L. Folding and Assembly of Hemoglobin Monitored by Electrospray Mass Spectrometry Using an On-line Dialysis System. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 8–16.
29. Ofori-Aquah, S. F.; Green, B. N.; Davies, S. C.; Nicolaidis, K. H.; Serjooat, C. R.; Layton, D. M. Mass Spectral Analysis of Asymmetric Hemoglobin Hybrids: Demonstration of Hb PS ( $\alpha_2\gamma\beta^S$ ) in Sickle Cell Disease. *Anal. Biochem.* **2001**, *298*, 76–82.
30. Versluis, C.; Heck, A. J. R. Gas-Phase Dissociation of Hemoglobin. *Int. J. Mass Spectrom.* **2001**, *210–211*, 637–649.
31. Boys, B. L.; Kuprowski, M. C.; Konermann, L. Symmetric Behavior of Hemoglobin  $\alpha$ - and  $\beta$ -Subunits during Acid-Induced Denaturation Observed by Electrospray Mass Spectrometry. *Biochemistry* **2007**, *46*, 10675–10684.
32. Williams, J. P.; Scrivens, J. H. Coupling Desorption/Electrospray Ionization and Neutral Desorption/Extractive Electrospray Ionization with a Travelling-Wave Based Ion Mobility Mass Spectrometer for the Analysis of Drugs. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 187–196.
33. Scrivens, J. H.; Thalassinou, K.; Hilton, G.; Slade, S. E.; Pinheiro, T. I. T.; Bateman, R. H.; Bowers, M. T. Use of a Travelling Wave-based Ion Mobility Approach to Resolve Proteins of Varying Conformation. In *Proceedings of the 55th ASMS Conference on Mass Spectrometry and Allied Topics*, Indianapolis, IN, 2006.
34. Shvartsburg, A. A.; Smith, R. D. Fundamentals of Traveling Wave Ion Mobility Spectrometry. *Anal. Chem.* (November 6, 2008). doi:10.1021/ac8016295 [Epub ahead of print].
35. Thalassinou, K.; Grabenauer, M.; Slade, S. E.; Hilton, G. R.; Bowers, M. T.; Scrivens, J. H. Characterization of Phosphorylated Peptides Using Travelling Wave-based and Drift Cell Ion Mobility Mass Spectrometry. *Anal. Chem.* (December 4, 2008). [Epub ahead of print].
36. Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. Structural Information from Ion Mobility Measurements: Effects of the Long-range Potential. *J. Phys. Chem.* **1996**, *100*, 16082–16086.
37. Shvartsburg, A. A.; Jarrold, M. F. An Exact Hard-Sphere Scattering Model for the Mobilities of Polyatomic Ions. *Chem. Phys. Lett.* **1996**, *261*, 86–91.
38. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
39. Jarrold, M. F. Unfolding, Refolding, and Hydration of Proteins in the Gas Phase. *Acc. Chem. Res.* **1999**, *32*, 360–367.
40. Hossain, B. M.; Konermann, L. Pulsed Hydrogen/Deuterium Exchange MS/MS for Studying the Relationship between Noncovalent Protein Complexes in Solution and in the Gas Phase after Electrospray Ionization. *Anal. Chem.* **2006**, *78*, 1613–1619.
41. Kuprowski, M. C.; Boys, B. L.; Konermann, L. Analysis of Protein Mixtures by Electrospray Mass Spectrometry: Effects of Conformation and Desolvation Behavior on the Signal Intensities of Hemoglobin Subunits. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 1279–1285.
42. Leutzinger, Y.; Boychok, S. Kinetics and Mechanism of Hemo-Induced Refolding of Human  $\alpha$ -Globin. *Proc. Nat. Acad. Sci. U. S. A.* **1981**, *78*, 780–784.
43. Badman, E. R.; Hoaglund-Hyzer, C. S.; Clemmer, D. E. Monitoring Structural Changes of Proteins in an Ion Trap over  $\sim 10$ –200 ms: Unfolding Transitions in Cytochrome c Ions. *Anal. Chem.* **2001**, *73*, 6000–6007.