

*Chemical labelling strategies for mass
spectrometric peptide analysis*

Helen Kathryn Robinson

Doctor of Philosophy

Chemistry

University of York

September 2015

Abstract

The work in this thesis describes the use of self-assembled monolayers (SAMs) to capture peptides on a gold-coated MALDI chip for mass spectrometric analysis. A SAM was formed on the gold-coated MALDI chip, onto which 4-bromophenylalanine was coupled (giving peaks due to 4-bromophenylalanine-SAM species a characteristic isotopic distribution). Simple single peptides and tryptic peptides from single proteins have been analysed using the developed technology. The capture of peptides on SAMs for mass spectrometric analysis has many advantages; all steps (chemistry and analysis) take place *in situ* on the gold-coated MALDI chip. This means that there are no sample transfer steps, reducing sample loss. Additionally, unreacted reagents and buffers can be removed from the surface of the gold-coated MALDI chip through washing. Buffers which are not compatible with mass spectrometric analysis (e.g. phosphate-based buffers) can be used for peptide capture as the buffer is washed away prior to analysis.

This thesis describes an alternative cleaning strategy for the removal of organic material from the surface of the gold-coated MALDI chips. Two plasma instruments were investigated for their abilities to remove organic material from the gold-coated MALDI chips, and both were shown to be a suitable safe alternative cleaning strategy to the caustic, potentially explosive piranha solution currently used.

Finally, in-solution dimethyl labelling was investigated, as a labelling strategy which could be adapted for used with the SAM technology. Quantification was performed using standard solutions of light and heavy labelled simple protein digests mixed in a range of ratios. Two different commercially-available software packages were investigated to assess their ability to analyse the generated data in order to determine whether one is more suitable for dimethyl labelling quantification. A set of complex standard samples was dimethyl labelled and analysed, to define parameters for most effectively determining quantification ratios, before application to a set of 'real' samples. Muscle protein samples were analysed and a set of potentially differentially abundant proteins identified for further validation.

Contents

Abstract	3
Contents	5
List of tables	9
List of figures	12
List of equations	19
Acknowledgements	21
Author's declaration	23
Chapter 1: Introduction.....	25
1.1. Protein analysis	27
1.1.1. Gel electrophoresis	28
1.1.2. Mass spectrometry.....	30
1.2. Mass spectrometry.....	31
1.2.1. Ionisation	32
1.2.1.1. Electrospray ionisation (ESI).....	32
1.2.1.2. Matrix-assisted laser desorption/ionisation (MALDI).....	34
1.2.2. Mass analysers	38
1.2.2.1. Time-of-flight (ToF).....	38
1.2.2.2. Fourier transform ion cyclotron resonance (FT-ICR)	41
1.2.3. Tandem mass spectrometry.....	45
1.3. Protein labelling strategies	47
1.3.1. Stable isotope labelling by amino acids	50
1.3.2. Isotope-coded affinity tag	51
1.3.3. Tandem mass tags.....	54
1.3.4. Isobaric tag for relative and absolute quantification	57
1.3.5. Acid-labile isotope-coded extractants	61
1.3.6. Dimethyl labelling	63
1.3.7. <i>N,N</i> -Dimethyl leucine	65
1.3.8. ¹⁸ O oxygen incorporation	66
1.3.9. Bromine incorporation	67
1.3.10. Summary	69
1.4. Self-assembled monolayers.....	71
1.5. Aims of this thesis	76
Chapter 2: Thiol-derived SAM.....	79

2.1.	Introduction.....	83
2.1.1.	Aims and overview	86
2.2.	Formation of a thiol-derived SAM.....	87
2.3.	Investigation into the most suitable matrix	90
2.4.	Peptide capture on a 4-bromophenylalanine tagged SAM	93
2.4.1.	Activation of SAM carboxylic acid groups	94
2.4.2.	Coupling of 4-bromophenylalanine to activated SAM	95
2.4.3.	Activation of 4-BrPhe-tagged SAM carboxylic acid groups	98
2.4.4.	Capture of single peptides on activated 4-BrPhe-tagged SAM	100
2.5.	Peptides captured on untagged SAM	107
2.5.1.	Activation of SAM carboxylic acid groups	107
2.5.2.	Peptide capture on activated SAM carboxylic acid groups	108
2.5.3.	Use of alternative activating reagents and conditions to investigate peptide capture on untagged SAM	114
2.6.	Investigation of inefficient activation of carboxylic acid group activation for peptide capture	116
2.6.1.	Natural exchange	118
2.6.2.	Investigating potential hydrolysis during 4-bromophenylalanine coupling to activated SAM	120
2.6.3.	Investigating potential hydrolysis during peptide capture on activated SAM ...	122
2.7.	Conclusions and future work.....	125
Chapter 3: Disulfide-derived SAM.....		129
3.1.	Introduction.....	133
3.1.1.	Aims and overview	134
3.2.	Formation of a disulfide-derived SAM	134
3.3.	Peptide capture on SAM	138
3.3.1.	MS/MS analysis of peptides captured on the disulfide-derived SAM	148
3.4.	4-Bromophenylalanine and phenylalanine coupling to SAM	151
3.5.	Conclusions and future work.....	157
Chapter 4: Using plasma effluent as a novel cleaning technique		161
4.1.	Introduction.....	163
4.1.1.	Aims	169
4.2.	Formation of a thiol-derived SAM.....	170
4.3.	Testing the kHz plasma instrument for cleaning SAM-coated gold	171

4.3.1.	Optimising conditions for the kHz plasma instrument	171
4.3.2.	Incremental treatment time course study	172
4.4.	Testing the radiofrequency (RF) plasma instrument for cleaning SAM-coated gold	175
4.4.1.	Optimising conditions for the RF plasma instrument	175
4.4.2.	Time course study	176
4.5.	Investigation into unassigned mass spectrometric peaks observed after plasma irradiation using either kHz or RF plasma instruments	179
4.6.	Respotting of thiol-derived SAM on wells cleaned with a plasma effluent	181
4.7.	Conclusions and future work	182
Chapter 5: Dimethyl labelling		185
5.1.	Introduction	187
5.1.1.	Aims and overview	188
5.2.	Validation of dimethyl labelling for quantification using standard solutions of light and heavy dimethyl-labelled protein digests	191
5.2.1.	Analysing standard ratios using aliquots of the same BSA digest	191
5.2.1.1.	MALDI-MS analysis	192
5.2.1.2.	LC-ESI-MS analysis	197
5.2.1.3.	Pseudo LC-MALDI-MS analysis	204
5.2.2.	Reproducibility of protein digestion and labelling	204
5.3.	Spiking labelled protein digests into labelled <i>E. coli</i> lysate digest	206
5.4.	Applying dimethyl labelling to relative quantification of muscle protein samples	220
5.5.	Conclusions and future work	231
Chapter 6: Integration, Conclusions and Future Work		235
6.1.	Integration	237
6.2.	Conclusions	241
6.3.	Future Work	243
Chapter 7: Experimental		249
7.1.	General methods	251
7.1.1.	Mass spectrometry	251
7.1.1.1.	solariX	251
7.1.1.2.	ultraflex	252
7.1.1.3.	maXis HD	253
7.1.2.	Cleaning of gold-coated MALDI chips	254
7.2.	Chapter 2	254

7.2.1.	Thiol-derived SAM formation (Figure 2.4).....	254
7.2.2.	NHS/EDC activation (Figure 2.9 and Figure 2.12).....	254
7.2.3.	PFP/EDC activation.....	255
7.2.4.	NHSS/EDC activation	255
7.2.5.	4-Bromophenylalanine coupling to the thiol-derived SAM (Figure 2.10)	255
7.2.6.	Peptide capture on the thiol-derived SAM (Figure 2.13 and Figure 2.16)	255
7.3.	Chapter 3.....	256
7.3.1.	Disulfide-derived SAM formation (Figure 3.3).....	256
7.3.2.	Peptide coupling (Figure 3.7)	256
7.3.3.	4-Bromophenylalanine/phenylalanine coupling	257
7.4.	Chapter 4.....	258
7.4.1.	SAM formation.....	258
7.4.2.	Operation of kHz plasma instrument.....	258
7.4.3.	Operation of RF plasma instrument	258
7.5.	Chapter 5.....	259
7.5.1.	Tryptic digestion	259
7.5.2.	Dimethyl labelling	259
7.5.3.	ZipTip treatment	260
7.5.4.	ProteinScape processing	260
7.5.4.1.	MALDI-MS.....	260
7.5.4.2.	LC-ESI-MS	260
7.5.5.	Mascot Distiller processing	261
7.5.6.	FDR determination	261
7.6.	Chapter 6.....	263
7.6.1.	Tryptic digestion	263
7.6.2.	Labelling of tryptic digestion.....	263
	Abbreviations	265
	References.....	271

List of tables

Chapter 1

Table 1.1: The combination of different isotopic reagents used for the five-plex dimethyl labelling system.....	64
Table 1.2: Summary of labelling methods, developed from the table by Boersema <i>et al.</i> ¹²³ .	70

Chapter 2

Table 2.1: Peak assignment for mass spectrum of thiol-derived SAM formation on a gold-coated MALDI chip	90
Table 2.2: Matrices and solvents used for the investigation (number of replicates = 8).....	91
Table 2.3: THAP concentrations and volumes spotted on wells (number of replicates = 3) .	92
Table 2.4: Peak assignment for mass spectrum of 4-bromophenylalanine tagged SAM.....	98
Table 2.5: Peptides captured on SAM construct and expected peaks (the product is defined as observed if the full isotopic distribution was observed in the mass spectra)	102
Table 2.6: Peak assignments for mass spectrum of Leu-Gly-Gly captured on activated 4-BrPhe tagged SAM	104
Table 2.7: Peptides and amounts used for capture, activating reagent used, expected m/z values of product captured on the activated SAM and whether the product was observed (the product is defined as observed if the full isotopic distribution of either of the species $[M+Na]^+$ and $[M+2Na-H]^+$ was observed in the mass spectra)	110

Chapter 3

Table 3.1: Peak assignments for mass spectrum of disulfide-derived SAM formation on a gold-coated MALDI chip.....	137
Table 3.2: Peptides spotted on the SAM and expected m/z values of product captured on the SAM.....	140
Table 3.3: Peak assignments for MALDI mass spectrum of leucine enkephalin captured on the disulfide-derived SAM	141
Table 3.4: Table showing peptides spotted at 5 and 0.05 nmol levels on the disulfide-derived SAM, and percentage of spectra with species for peptide capture and unreacted SAM (number of replicates = 12)	145
Table 3.5: Peaks associated with disulfide-derived SAM and 4-BrPhe coupled to the disulfide-derived SAM	153
Table 3.6: Peaks associated with disulfide-derived SAM and Phe captured on the SAM...	154

Table 3.7: Table for 4-BrPhe and Phe spotted at 5 and 0.05 nmol levels on the disulfide-derived SAM, and percentage of spectra with species for 4-BrPhe/Phe capture and unreacted SAM (number of replicates = 8).	155
---	-----

Chapter 4

Table 4.1: Treatment combinations for kHz plasma instrument, and for each treatment the percentage of spectra in which peaks for the SAM were detected (number of wells per condition = 4).....	172
---	-----

Table 4.2: Treatment combinations for the RF plasma instrument, and for each treatment the percentage of spectra in which peaks for the SAM were detected (number of wells analysed for each condition = 2)	176
---	-----

Chapter 5

Table 5.1: Ratios generated using PS for MALDI-MS analysis of prepared standard solutions of BSA tryptic digest (geometric standard deviations (SD _{geo}) calculated for the three replicate solutions prepared for each ratio) (number of peptides and sequence coverage for the three replicate solutions indicated) (see 7.5.6 for FDR values)	194
---	-----

Table 5.2: Ratios generated for each BSA digest solution (geometric standard deviations (SD _{geo}) for the average of three replicate solutions per ratio) (number of peptides and sequence coverage for individual solution replicates for each ratio) (see 7.5.6 for FDR values).....	200
---	-----

Table 5.3: Ratios obtained using MD for LC-ESI-MS analysis of different BSA digest solutions (see Section 7.5.6 for FDR values).....	205
--	-----

Table 5.4: Proteins used for spiking into <i>E. coli</i> lysate digest, and the ratios (L:H and H:L) they were prepared at	207
--	-----

Table 5.5: Number of proteins identified and quantified for each of the L:H and H:L solutions prepared with proteins spiked in at 1% and 5% of the total <i>E. coli</i> lysate digest amount (see Section 7.5.6 for FDR values)	208
---	-----

Table 5.6: Table showing the proteins with a q-value of less than 0.05 for proteins spiked into the <i>E. coli</i> lysate digest at 1% of the total lysate digest amount	216
--	-----

Table 5.7: Table showing the proteins with a q-value of less than 0.05 for proteins spiked into the <i>E. coli</i> lysate digest at 1% of the total lysate digest amount	217
--	-----

Table 5.8: Number of peptides identified and quantified across the muscle protein samples (see 7.5.6 for FDR values)	221
--	-----

Table 5.9: Table showing proteins with a q-value < 0.05 for muscle protein samples	226
--	-----

Chapter 6

Table 6.1: Proposed peptide assignments for peaks observed for acetone/ acetone-d ₆ labelled peptides	240
--	-----

Chapter 7

Table 7.1: Peptides used for capture reactions and their corresponding product peaks observed by MALDI-MS analysis	256
Table 7.2: Peptides used for capture reaction and their corresponding product peaks observed by MALDI-MS analysis	257
Table 7.3: FDR values calculated for analysis in Section 5.2.1.1 (- = no results from peptide search)	261
Table 7.4: FDR values calculated for analyses in Section 5.2.1.2.....	262
Table 7.5: FDR values calculated for analyses in Section 5.2.2.....	262
Table 7.6: FDR values calculated for analyses in Section 5.3.....	263
Table 7.7: FDR values calculated for analyses in Section 5.4.....	263

List of figures

Chapter 1

Figure 1.1: Edman degradation process.....	28
Figure 1.2: ESI source schematic.....	32
Figure 1.3: Formation of a Taylor cone from the spherical droplet emerging from the capillary (left) to a Taylor cone (right).....	32
Figure 1.4: Charged residue model	33
Figure 1.5: Ion-evaporation model.....	33
Figure 1.6: Matrix molecules used for MALDI.....	34
Figure 1.7: MALDI ionisation process	35
Figure 1.8: The unified MALDI analyte protonation mechanism restricted to positively precharged analyte ions such as peptides and proteins ⁵⁸ (where A = analyte, ma = matrix, X = counterion, H = proton)	37
Figure 1.9: Linear ToF mass analyser in linear orientation.....	38
Figure 1.10: Reflectron ToF mass analyser (both ions have the same m/z , but the blue circles represent ions with less kinetic energy than the red circles)	40
Figure 1.11: Orthogonal acceleration reflectron ToF mass analyser (both ions have the same m/z , but the blue circles represent ions with less kinetic energy than the red circles)	41
Figure 1.12: An example of a cubic ICR cell, where B = magnetic field (trapping plates are coloured blue)	42
Figure 1.13: The three motions that ions undergo in the cell.....	43
Figure 1.14: The motion of an ion within the ICR cell as a result of trapping, cyclotron and magnetron motions	43
Figure 1.15: Typical arrangement of an in-space tandem mass spectrometer	45
Figure 1.16: The fragmentation which can occur in a peptide. The different colours indicate the different types of bonds which can break.....	46
Figure 1.17: Strategies for protein quantification, detailing the possible different stages involved and the potential stages at which experimental variation can occur. ⁸⁹	48
Figure 1.18: Schematic of SILAC labelling strategy (where blue and red colours represent the two different cell cultures)	51
Figure 1.19: ICAT reagent, consisting of three regions: a biotin region to isolate ICAT-labelled peptides, a linker region which incorporates the stable isotopes and a reactive group which reacts with thiols (X = H (light) or X=D (heavy))	52
Figure 1.20: The ICAT strategy for quantifying differential protein expression.....	52
Figure 1.21: ICAT analysis of a protein from a standard protein mixture. Full scan mass spectrum at time 19.76 minutes of the LC-MS analysis (expanded m/z region to illustrate m/z difference between light and heavy labelled doubly charged peptide ions). At least four different peptide doublets were identified. ¹⁰⁵	53
Figure 1.22: First generation of the TMT tag (R = protein reactive functionality, X = H/D) ...	54

Figure 1.23: Pierce™ TMT reagent (TMT ⁰ shown) (* indicates isotope position).....	55
Figure 1.24: (a) Tandem mass spectrum of the peptide LSFNPTQLEEQCHI of β -lactoglobulin (m/z 1944.902), labelled with TMT and iodoacetamide. (b) Expanded depiction of the reporter ion region providing peptide relative quantification in a model four-protein mixture at six different relative concentrations (1:2:3:3:5:10) through the abundance of the reporter ions at $m/z = 126.1, 127.1, 128.1, 129.1, 130.1$ and 131.1 . ¹¹⁰	56
Figure 1.25: iTRAQ label consisting of three regions: a reporter group, a balancer group and a reactive group which targets primary amines. Table shows combinations of isotopes used to create each reporter group with the relevant balancer group, to make isobaric tags ^{102,113}	57
Figure 1.26: iTRAQ strategy for quantitative analysis	58
Figure 1.27: Example tandem mass spectrum of the peptide TPHPALTEAK from a protein digest mixture prepared by labelling four separate digests with each of the four iTRAQ reagents and combining the reaction mixtures in a 1:1:1:1 ratio. (i) isotopic distribution of the precursor ($[M+H]^+$, m/z 1352.84), (ii) low mass region showing signature iTRAQ ions used for quantification, (iii) isotopic distribution of the b_6 -ion, and (iv) isotopic distribution of the y_7 -ion. The peptide is labelled at both the N-terminus and C-terminus (lysine side-chain). The precursor ion and all internal fragment ions (e.g. b- and y-ions) contain all four members of the tag set, but remain isobaric. ¹¹²	59
Figure 1.28: 8-plex iTRAQ label consisting of three regions: a reporter group, a balancer group and a reactive group which targets primary amines	60
Figure 1.29: ALICE chemical structure (X=H (light), X=D (heavy)) with three regions: reactive group, linker chain and acid-labile functionality	62
Figure 1.30: Example tandem mass spectrum of a BSA peptide with three cysteine residues (CCAADDKEACFAVEGPK) labelled with the light ALICE label ¹²¹	62
Figure 1.31: Dimethyl labelling strategy developed by Hsu <i>et al.</i> (X = H (light) or D (heavy)) ¹²²	63
Figure 1.32: Labelling schemes for triplex isotope dimethyl labelling (R = remainder of peptide)	64
Figure 1.33: General structure of the DiLeu chemical label (* indicate labels in the 115 (smallest) label).....	66
Figure 1.34: Reaction scheme for the incorporation of ¹⁸ O during proteolytic digestion using labelled water	67
Figure 1.35: Bromine-containing solid-support N-terminal tags.....	68
Figure 1.36: Tandem MALDI mass spectra from the peptide WHWLQLKPGQPMY (a) before bromoacetylation and (b) after bromoacetylation. The Br-tag was only observed in the b-ion series. ¹³⁰	69
Figure 1.37: Generic structure of a SAM	71
Figure 1.38: Gas-phase reagent reacting with a SAM ¹³⁵	72

Chapter 2

Figure 2.1: 64 well gold-coated MALDI-MS chip.....	83
Figure 2.2: SPOT synthesis of a peptide library using Fmoc-protected amino acids.	84
Figure 2.3: Schematic representation of the capture of peptides on 4-bromophenylalanine tagged thiol-derived SAM.....	86
Figure 2.4: Formation of the thiol-derived SAM	88
Figure 2.5: MALDI mass spectrum of the thiol-derived SAM formed on a gold-coated MALDI chip.....	89
Figure 2.6: Structures of matrix molecules used for the investigation	90
Figure 2.7: L-4-bromophenylalanine	93
Figure 2.8: Simulated isotopic distribution of peaks associated with $[M+Na]^+$ mixed disulfide having one alcohol terminus and one carboxylic acid terminus on which the 4-BrPhe chemical tag is incorporated	94
Figure 2.9: NHS/EDC activation of the thiol-derived SAM.....	95
Figure 2.10: 4-Bromophenylalanine coupling to the NHS-activated carboxylic acid group ...	96
Figure 2.11: MALDI mass spectrum of 4-BrPhe coupled to the NHS-activated SAM	97
Figure 2.12: NHS/EDC activation of the 4-BrPhe-tagged SAM.....	99
Figure 2.13: Generic reaction of a peptide with the activated 4-BrPhe tagged SAM	101
Figure 2.14: MALDI mass spectrum of 50 nmol Leu-Gly-Gly spotted on the PFP-activated 4-BrPhe-tagged SAM (* indicates a PFP-derived impurity (unidentified because product ion analysis failed))	103
Figure 2.15: Plot showing ratio between peak intensity of the product and the starting material MALDI signals for peptides Gly-Gly-Val and Leu-Gly-Gly on both the SAM and the 4-BrPhe-tagged SAM (error bars show one standard deviation either side of the mean value) (number of replicates = 4).	105
Figure 2.16: Generic reaction of a peptide with activated SAM.....	109
Figure 2.17: Plot showing the mean ratio of peak intensity for product:starting material MALDI signals for each of the peptides with NHS as the activating reagent (Error bars show one standard deviation either side of the mean) (number of replicates = 4)	111
Figure 2.18: Plot showing the mean ratio of peak intensity for product:starting material MALDI signals for each of the peptides with PFP as the activating reagent (Error bars show one standard deviation either side of the mean) (number of replicates = 4)	112
Figure 2.19: (a) <i>N</i> -hydroxysuccinimide and (b) <i>N</i> -hydroxysodiumsulfosuccinimide	114
Figure 2.20: Proposed hydrolysis of NHS-activated carboxyl SAM.....	116
Figure 2.21: Natural exchange of the oxygen in the buffer (red) with the oxygen in the carboxylic acid (blue)	118
Figure 2.22: Plot showing the mean peak intensity ratio for $[M+2]:M$ for peaks associated with the SAM after ^{18}O exchange overnight (Error bars show one standard deviation either side of the mean) (number of replicates = 8)	119

Figure 2.23: Plot showing the mean peak intensity ratio for [M+2]:M for peaks associated with the SAM after coupling of the 4-BrPhe to the activated SAM when using two different activating reagents (Error bars show one standard deviation either side of the mean) (number of replicates = 8)	121
Figure 2.24: Plot showing the mean peak intensity ratio [M+2]:M for peaks associated with the SAM after capture of the peptide Gly-Gly-Gly on the activated SAM when using two different activating reagents (Error bars show one standard deviation either side of the mean) (number of replicates = 8).....	123
Figure 2.25: Plot showing the mean peak intensity ratio for [M+2]:M for peaks associated with the SAM after capture of the peptide Phe-Gly-Gly on the activated SAM when using two different activating reagents (Error bars show one standard deviation either side of the mean) (number of replicates = 8).....	124

Chapter 3

Figure 3.1: NHS ester disulfide	133
Figure 3.2: mPEG disulfide	133
Figure 3.3: Formation of the disulfide-derived SAM.	135
Figure 3.4: MALDI mass spectrum of the disulfide-derived SAM formed on a gold-coated MALDI chip.....	136
Figure 3.5: NHS ester by-product (containing a stable <i>N</i> -acyl urea)	137
Figure 3.6: Carboxylic acid reacting with DCC to form <i>O</i> -acyl isourea intermediate and subsequent rearrangement or substitution (R = rest of the disulfide, Cy = cyclohexyl)	138
Figure 3.7: Generic reaction of a peptide with the disulfide-derived pre-activated SAM.....	139
Figure 3.8: MALDI mass spectrum obtained following reaction for one hour of leucine enkephalin at 5 nmol with the disulfide-derived SAM	141
Figure 3.9: MALDI-product ion spectrum of [M+2Na-H] ⁺ for capture of the peptide Gly-Gly-Val on the SAM (precursor at <i>m/z</i> 1070).....	149
Figure 3.10: MALDI-product ion spectrum of [M+2Na-H] ⁺ for capture of the peptide leucine enkephalin on the SAM (precursor at <i>m/z</i> 1394)	150
Figure 3.11: Proposed fragments for product generated on CID of leucine enkephalin captured on disulfide-derived SAM	151
Figure 3.12: MALDI mass spectra obtained following reaction for one hour of 4-BrPhe at 5 nmol and 0.05 nmol with the disulfide-derived SAM. Top = 5 nmol of 4-BrPhe spotted on the disulfide-derived SAM, bottom = 0.05 nmol of 4-BrPhe spotted on the disulfide-derived SAM	152
Figure 3.13: MALDI mass spectra obtained following reaction for one hour of Phe at 5 nmol and 0.05 nmol with the disulfide-derived SAM. Top = 5 nmol of Phe spotted on disulfide-derived SAM, bottom = 0.05 nmol of Phe spotted on disulfide-derived SAM.....	154

Figure 3.14: Structure of bis(sulfosuccinimidyl) suberate	159
---	-----

Chapter 4

Figure 4.1: Potential structure in a plasma sheath (ϕ = potential, x = distance from the cold surface) ²¹¹	164
Figure 4.2: Simplified model for an electron travelling through stationary neutral species in a 3D model (left) and 2D model (right) (e^- = electron)	165
Figure 4.3: RF plasma electrode set up (left) and kHz plasma electrode set up (right). The electrode gap on the RF plasma electrode is 30 x 1 x 1 mm. On the kHz plasma electrode, the lower electrode is powered, whilst the upper electrode is grounded.	168
Figure 4.4: Damaged gold chip (arrow indicates where the metal layer under the gold can be seen after repeated use of piranha solution)	169
Figure 4.5: MALDI mass spectrum of the alcohol thiol-derived SAM	170
Figure 4.6: The electrode region of the kHz plasma instrument	171
Figure 4.7: Plot showing the S/N of m/z 861 after SAM-bearing, matrix-spotted wells were treated with the plasma effluent of the kHz plasma instrument	173
Figure 4.8: MALDI mass spectrum obtained from a well with SAM but no matrix treated with the effluent of the kHz plasma instrument for 10 seconds	174
Figure 4.9: The electrode region of the RF plasma instrument	175
Figure 4.10: Plot showing S/N of m/z 861 after SAM-bearing, matrix spotted wells were treated with the plasma effluent of the RF plasma instrument. Results are shown for the addition of 0.25% nitrogen (blue) or 0.5% oxygen (red)	177
Figure 4.11: MALDI mass spectrum obtained from a well with SAM but no matrix treated with the effluent with nitrogen incorporation of the RF plasma instrument for ten seconds.....	178
Figure 4.12: MALDI mass spectra obtained after treatment with the kHz plasma instrument (top) and RF plasma instrument (bottom) (* indicates peaks which are associated with the SAM)	179
Figure 4.13: MALDI mass spectrum obtained for a new gold chip before (top) and after (bottom) cleaning with piranha solution	180
Figure 4.14: MALDI mass spectrum showing alcohol thiol-derived SAM respotted on a well cleaned with plasma effluent (m/z 861 ($[M+Na]^+$, disulfide having two alcohol termini))	181

Chapter 5

Figure 5.1: Dimethyl labelling strategy developed by Hsu <i>et al.</i> (X = H (light) or D (heavy)) ¹²²	188
Figure 5.2: Workflow showing the steps in the acquisition of quantification data from MALDI-MS analysis using WARP-LC and PS.....	193

Figure 5.3: Plot showing the geometric mean generated from three replicate solutions analysed by MALDI-MS against the prepared ratio (error bars show the geometric standard deviation).....	195
Figure 5.4: Data workflow for data processed through PS, including the data format at each stage.....	198
Figure 5.5: Data workflow through MD, including the data format at each stage	198
Figure 5.6: Plot showing the geometric mean generated from three replicate solutions analysed by LC-ESI-MS against the prepared ratio (error bars show the geometric standard deviation) (PS = blue diamonds, MD = red squares)	201
Figure 5.7: Venn diagram showing the number of proteins quantified across all three replicates for L:H solutions with proteins spiked in at 1% of the total <i>E. coli</i> lysate digest amount (Venn diagram produced using Venny ²³⁷).	209
Figure 5.8: Venn diagram showing the number of proteins quantified across all three replicates for L:H solutions with proteins spiked in at 5% of the total <i>E. coli</i> lysate digest amount (Venn diagram produced using Venny ²³⁷).	209
Figure 5.9: Venn diagram showing the number of proteins quantified across all three replicates for H:L solutions with proteins spiked in at 1% of the total <i>E. coli</i> lysate digest amount (Venn diagram produced using Venny ²³⁷).	210
Figure 5.10: Venn diagram showing the number of proteins quantified across all three replicates for H:L solutions with proteins spiked in at 5% of the total <i>E. coli</i> lysate digest amount (Venn diagram produced using Venny ²³⁷).	210
Figure 5.11: Volcano plot showing $-\log_{10}(\text{q-value})$ against $\log_2(\text{protein ratio})$ for proteins spiked in at 1% of the total <i>E. coli</i> lysate digest amount (L:H) (Blue = solution 1, red = solution 2, green = solution 3).....	212
Figure 5.12: Volcano plot showing $-\log_{10}(\text{q-value})$ against $\log_2(\text{protein ratio})$ for proteins spiked in at 1% of the total <i>E. coli</i> lysate digest amount (H:L) (Blue = solution 1, red = solution 2, green = solution 3).....	213
Figure 5.13: Volcano plot showing $-\log_{10}(\text{q-value})$ against $\log_2(\text{protein ratio})$ for proteins spiked in at 5% of the total <i>E. coli</i> lysate digest amount (L:H) (Blue = solution 1, red = solution 2, green = solution 3).....	214
Figure 5.14: Volcano plot showing $-\log_{10}(\text{q-value})$ against $\log_2(\text{protein ratio})$ for proteins spiked in at 5% of the total <i>E. coli</i> lysate digest amount (H:L) (Blue = solution 1, red = solution 2, green = solution 3).....	215
Figure 5.15: Venn diagram showing the number of proteins identified in the unstretched:stretched 1 L:H and H:L experiments (Venn diagram produced using Venny ²³⁷)	222
Figure 5.16: Venn diagram showing the number of proteins quantified in the unstretched:stretched 1 L:H and H:L experiments (Venn diagram produced using Venny ²³⁷)	222

Figure 5.17: Venn diagram showing the number of proteins identified in the unstretched:stretched 2 L:H and H:L experiments (Venn diagram produced using Venny²³⁷) 222

Figure 5.18: Venn diagram showing the number of proteins quantified in the unstretched:stretched 2 L:H and H:L experiments (Venn diagram produced using Venny²³⁷) 223

Figure 5.19: Volcano plot showing $-\log_{10}(\text{q-value})$ against $\log_2(\text{protein ratio})$ for unstretched:stretched 1 (Blue = L:H, red = H:L) 224

Figure 5.20: Volcano plot showing $-\log_{10}(\text{q-value})$ against $\log_2(\text{protein ratio})$ for unstretched:stretched 2 (Blue = L:H, red = H:L) 224

Chapter 6

Figure 6.1: Proposed workflow for capturing protein digests on a SAM, with incorporation of the dimethyl-derived label 237

Figure 6.2: Labelling of a generic peptide with acetone (top) or acetone-d₆ 238

Figure 6.3: MALDI mass spectrum of a BSA digest labelled with acetone and acetone-d₆ (1:1) (blue = acetone labelled peptides, red = acetone-d₆ labelled peptides) (not all peptide pairs are labelled on the mass spectrum, but all identified peptides are detailed in Table 6.1) 239

Figure 6.4: Schematic of the chemical label, including a carbonyl group and an azide group 243

Figure 6.5: A potential chemical label using hydrogen and deuterium to generate the light and heavy labels (X indicate positions of either hydrogen or deuterium) 243

Figure 6.6: Potential synthetic scheme for synthesis of hydrogen (light) label²⁸⁹ 244

Figure 6.7: A potential chemical label using ¹²C and ¹³C to generate the light and heavy labels (* indicate positions of either ¹²C or ¹³C)..... 244

Figure 6.8: Proposed synthesis of the ¹³C chemical label (* indicates ¹³C position) 245

Figure 6.9: Potential 'reactive' component of the SAM 245

Figure 6.10: Potential spacer component of the SAM (where n = 4 or 5)..... 246

List of equations

Chapter 1

Equation 1.1: Equation relating the kinetic energy of an accelerated ion to its velocity mass	38
Equation 1.2: Rearrangement of Equation 1.1, to give velocity as the subject	39
Equation 1.3: Equation relating the time taken for an ion to travel the length of the flight tube to its velocity.....	39
Equation 1.4: Equation showing that m/z can be calculated from a measurement of the time taken to travel the length of the flight tube	39
Equation 1.5: Cyclotron frequency, and its dependence on mass, charge and magnetic field strength	44
Equation 1.6: Maximum resolution achievable in FT-ICR-MS	44

Chapter 2

Equation 2.1: Equation used to generate relative intensity ratio of product:starting materials	105
Equation 2.2: Mann-Whitney U-test calculation	106

Chapter 4

Equation 4.1: Equation showing the degree of ionisation of a plasma (n_e = electron density, n_o = number density of neutral atoms)	163
Equation 4.2: Equation for cross section of an atom (r_g = radius of an atom of the gas used in the plasma).....	165
Equation 4.3: Equation (upper) showing that virtually the entire face xy is obscured when the cuboid extends as far as one mean free path (and simplified (lower))	165
Equation 4.4: Equation for the frequency of encounters for electrons (ν = frequency of encounters, \bar{u} = mean speed)	165

Chapter 5

Equation 5.1: Formula used by Bruker PS for the determination of the protein quantification “median”	193
Equation 5.2: Formula for the determination of the protein quantification “mean”	193
Equation 5.3: Formula for the determination of the coefficient of variation of data set X	194
Equation 5.4: Bonferroni correction (S = significance level, N = number of tests)	211
Equation 5.5: Benjamini and Hochberg correction (N = number of tests)	212

Acknowledgements

Firstly, I would like to thank my supervisors Jane Thomas-Oates and Anne Routledge for giving me the opportunity to undertake this project. Your support and belief in me has been invaluable. My apologies go to them both for sometimes forgetting that I wasn't still in industry (and my 'Helen signature strange formulations') when writing my TAP reports and thesis.

I would like to thank members of the JTO and AR groups, both past and present, who helped me throughout my PhD, not least by providing me with a great working atmosphere. In particular, I would like to thank Siân who suggested I apply for this project (essentially, it's all her fault!). Thanks also to Salina, Andy, James, Rachel S, Jo, Kirsty, Rachel B, Emily, Chris and Tom. Rachel B, thank you for being such a brilliant MChem student!

Thanks also to the boys and girl over in the TF: Jerry, Dave, Adam and Rachel. Thank you for all your help on the proteomics side of things and for letting me take up valuable space and time within the TF and your office.

A big thank you must go to Karl for putting up with me as the student technician on the service for four years. You have taught me so much, and I will forever be grateful to you for sharing your knowledge with me, trusting me with the instruments and also for your friendship.

Thank you to the friends both inside and outside of the mad world of PhD-ing. Thank you for providing me with distractions from the madness. Thank you to Kate, David and Joe for the entertainment provided during demonstrating and also for their friendship outside of work. Thank you to Kirsty for being my gym buddy and a brilliant listener and logic-tester, but most of all for being an amazing friend over the last three years.

I would like to thank my family; Mum, Dad, Louise, Henry and Alfie. Thank you for attempting to understand what I've been doing for the last four (possibly eight) years, and nodding along to what I've said! Without your love, support and encouragement throughout my PhD, I would have had a much harder time of it.

Finally, thank you to Peter for letting me bend his ear when things were not going well at work, telling me that there was a light at the end of the tunnel and pushing me to get this thing finished! Thank you for your unconditional love (tested thoroughly at times) throughout this long process.

Author's declaration

I hereby declare that the work presented in this thesis is my own, except where otherwise acknowledged, and has not previously been submitted for a degree at this or any other university.

Chapter 1: Introduction

1.1. Protein analysis

Proteomics is the study of the proteome; which is all the proteins being expressed in a particular organism, tissue or cell, at a particular time by a genome. It is a complementary technique to genomics. The field of proteomics aims to use protein-level measurements of gene expression to characterise biological processes.¹ Therefore, the ability to identify, and potentially quantify (see Section 1.3), proteins in organelles, cells, tissues or fluids in order to understand biological processes is fundamental to the field of proteomics. The term proteome was first used by Wasinger *et al.* in 1995.^{2,3} Proteomics is complicated by the molecular complexity and dynamic nature of proteomes;³ it can change with stimuli, sometimes on a very short timescale. Proteins can be expressed at different levels in response to changes within the cell and external conditions. In addition to protein expression levels changing, proteins can be modified in the process of post-translational modification (PTM). PTMs are physiological changes which happen to the protein either during or after protein synthesis, and increase the diversity of the proteome.⁴ PTMs can occur on amino acid side-chains, or at the N- or C-terminal, and are usually the addition, removal or modification of a functionality. The modifications that a protein can undergo include phosphorylation, glycosylation, methylation, acetylation, ubiquitination and N- and C-truncation.⁴ Whilst the human genome comprises between 20,000 and 25,000 genes⁵, the number of proteins generated from these genes is estimated to be nearly 2 million.⁶

A protein cannot be subjected to multiplication techniques such as polymerase chain reactions which are used in genomics; therefore, however much of a protein there is, that is the amount available for analysis, as it cannot be amplified.⁷

The Edman degradation technique has been used to sequence peptides and proteins for many years (Figure 1.1).⁸ The alpha amino group on the amino acid at the N-terminus is reacted with phenylisothiocyanate under basic conditions. The resulting derivative cyclises and the derivatised terminal amino acid is cleaved and then identified using techniques such as electrophoresis or liquid chromatography. The process can be repeated multiple times, each time removing the terminal amino acid group of the remaining chain to uncover another alpha amino group. However, these consecutive cleavages result in a lower yield at each step, because the cyclisation and cleavage may not reach completion at each step, and so the cumulative effect of many consecutive steps results in less and less sample for each subsequent step. The Edman degradation process is also a very lengthy process, but can be fully automated. One of the main limitations of the process is that, if there is a modification to the N-terminus which prevents the N-terminal alpha amino group from being nucleophilic, the amino acid will not react with the phenylisothiocyanate.

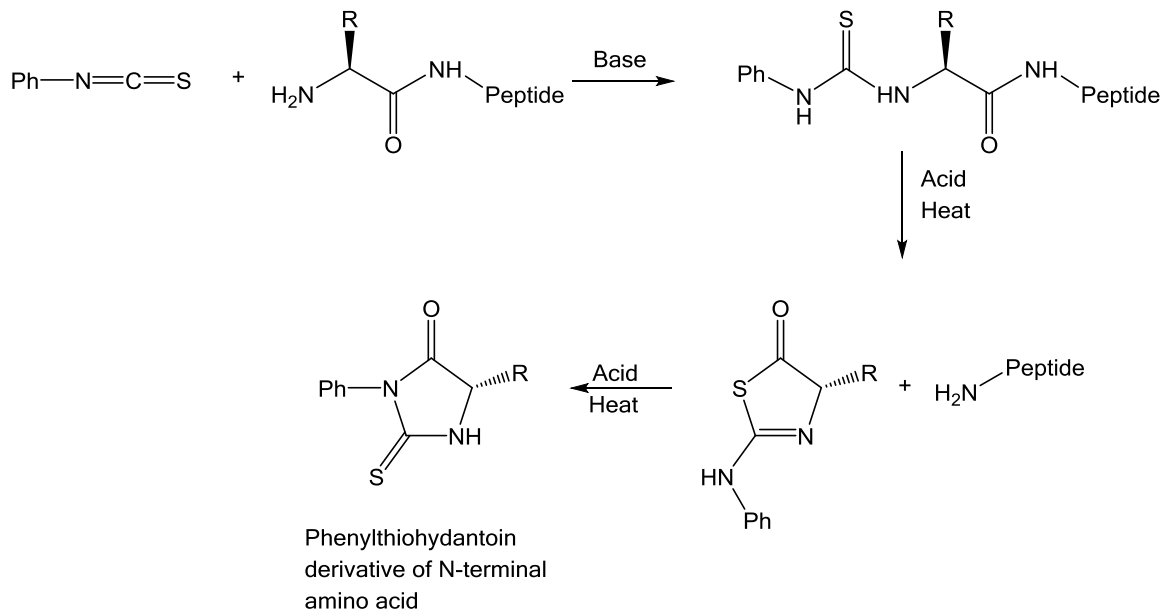


Figure 1.1: Edman degradation process

1.1.1. Gel electrophoresis

Gel electrophoresis is a technique used to separate charged molecules based on properties such as charge or size, as they pass through the gel. In gel electrophoresis, samples are loaded into wells at one end of a slab of gel. This gel is then placed in an electrophoresis chamber with an electrophoresis buffer solution. A voltage is applied across the ends of the gel, which creates a potential and causes the proteins to migrate in the electric field. The rate that a protein moves through the gel is related to its size and charge (i.e. a larger protein moves more slowly than a smaller protein).⁹ Gel electrophoresis is usually performed in either one or two dimensions.

One-dimensional gel electrophoresis (1-DGE) is a fast way to determine protein molecular weight and assess protein purity. There are different forms of 1-DGE, with the two main examples relevant to proteomics being sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)¹⁰ (used for determining molecular weight) and isoelectric focusing (used to determine the isoelectric point of a protein, and useful to detect small changes in proteins due to PTMs)¹¹. Laemmli incorporated the use of SDS (an anionic detergent which denatures proteins by non-covalently binding to the protein chain) into the native PAGE systems which had previously been used, to create the first SDS-PAGE systems.¹⁰ Once the proteins have been denatured and given a negative charge due to binding the SDS, they are then subjected to gel electrophoresis. The proteins are essentially separated by molecular weight, since the longer the protein chain, the more SDS

molecules they bind. Proteins can be stained, using stains such as Coomassie brilliant blue stain¹² or silver staining¹³ in order to visualise them. This generates an image of the gel, and allows proteins from the sample(s) to be visualised. Their migration can then be compared to those of standard proteins of known molecular mass, which are run alongside the unknown samples, and provide a 'molecular weight ladder'.¹⁴ The whole gel lane of the sample (or a small section of it) can be excised and the proteins in it subjected to in-gel proteolytic digestion, with the resulting peptides being subjected to mass spectrometric analysis.¹⁵

Two-dimensional gel electrophoresis (2-DGE) generally separates proteins based on their isoelectric point (pI) and size. The technique was described independently in 1975 by both O'Farrell¹⁴ and Klose¹⁶. In the first dimension, the protein mixture is loaded into a polyacrylamide gel strip, which has a pH gradient immobilised in it.¹⁷ A voltage is applied across the gel which causes the proteins to move through the gel until they reach the pH value which corresponds to their isoelectric point (the pH at which the protein has no net charge), at which point they stop moving. The proteins are then separated in an orthogonal second dimension based on their molecular weight, using SDS-PAGE. The proteins can then be stained, and following image analysis, protein spots can be excised and in-gel digested before mass spectrometric analysis.

A western blot can be used to detect specific proteins in a developed gel.^{18,19} The proteins are transferred to a nitrocellulose blotting membrane from the gel, which creates a 'copy' of the gel, which can be further analysed. A voltage is used which creates a potential and allows the negatively charged proteins from the gel to be attracted to the positively charged nitrocellulose membrane.¹⁸ Once the proteins have been transferred, the membrane is treated with a generic protein solution (e.g. bovine serum albumin (BSA)) to bind to 'sticky' places on the nitrocellulose (i.e. anywhere on the surface of the nitrocellulose where protein binding could occur). The binding of a generic protein to the remaining sites on the membrane prevents non-specific binding of a subsequent antibody, with the aim to maximise specific binding.²⁰ The antibody used is chosen to have specificity for a target protein, and binding, and thus staining, is obtained where a cross-reacting protein is present on the surface of the membrane. A common detection strategy is to use a two-step process in which primary and secondary antibodies are used.²¹ The primary antibodies are generated when a host species or immune cell culture is exposed to the target protein (or a part thereof). The membrane is incubated in a solution of the primary antibodies which recognise and bind to the target (and other cross-reacting) protein.²⁰ After washing the membrane to remove unbound primary antibodies, secondary antibodies are introduced, which are directed towards a species-specific portion of the primary antibodies. The secondary antibodies are usually linked to a conjugate, such as a reporter enzyme (e.g. horseradish peroxidase), which reacts with a development reagent or is used to cleave a

chemiluminescent reagent, resulting in a degree of luminescence which is proportional to the amount of protein present.

1.1.2. Mass spectrometry

Protein analysis has seen rapid development since the early 1990s saw an increased interest in the field.⁸ In particular, mass spectrometry has seen significant development to make protein analysis a viable analytical tool. Proteins and peptides can be separated using a method such as gel electrophoresis or HPLC, and then analysed by mass spectrometry. Mass spectrometry is used to determine accurate masses, amino acid sequences, and the nature and site of modifications of proteins and peptides.²²

Proteomics experiments can be classified as top-down or bottom-up experiments. In a bottom-up experiment, one or more proteins are broken down into constituent peptides before mass spectrometric analysis.²³ A proteolytic enzyme, for example, trypsin, is used. Trypsin cleaves the protein on the carboxyl (C)-terminal side of arginine and lysine residues (unless followed by proline), creating peptides that generally therefore have a basic residue (arginine or lysine) at the C-terminus of the peptides.²⁴ Bottom-up proteomics is a high throughput analysis method, and many proteins can be identified in a complex mixture.²⁵ Quantitative proteomics strategies are well established for application in the bottom-up proteomics approach, which can yield a lot of information. However, post-translational modification and isoform information may be lost as only one or two peptides from a protein may be detected, leading to protein identification based only on those one or two peptides, which may not contain the modifications.²⁶ Information about low abundance peptides/proteins can also be lost using this approach, as ions due to these peptides/proteins may be lost amongst information from more abundant peptides/proteins.

In a top-down proteomics experiment, a protein is analysed in its intact form, and usually subjected to gas-phase fragmentation in the mass spectrometer to allow protein sequencing.²⁷ There are advantages to the top-down approach, for example post-translational modifications can be detected (which may be lost in a bottom-up experiment),²⁷ and localisation of non-covalently bound ligands may be possible.^{26,28,29} Protein isoform identification is also possible with the top-down approach.²⁹ However, pure samples or simple mixtures are required, and there can be limited sensitivity and throughput.²⁸

There are three mass spectrometry strategies which are commonly used in proteomic workflows: shotgun, directed and targeted.³⁰ The term 'shotgun' proteomics is used when an entire proteome is digested and analysed; usually through a combination of liquid chromatography-tandem mass spectrometry (LC-MS/MS) and protein sequence database searching.^{7,31,32} The LC component of the analysis gives separation of the tryptic peptides

generated upon digestion. A typical approach is to record an MS survey scan, and allow the software to select a pre-defined number of precursor ions for MS/MS analysis, to enable identification of the peptides.^{30,31} There are disadvantages to the shotgun proteomic technique, with the main one being that the connection between intact proteins and their peptides is lost, and has to be recreated computationally after analysis.³³ Another disadvantage is the fact that lots of peptides are generated during proteome digestion, but not all of these peptides may be observed during mass spectrometric analysis.³⁴⁻³⁶ If a sample is analysed multiple times, then it is likely that different peptides are identified in the replicate analyses. This is largely influenced by the sampling speed of the mass spectrometer, the complexity of the sample and the separation capability of the LC system.³⁶ Although there have been vast improvements in the capabilities of both mass spectrometers and LC systems, there is a drive to analyse more and more complex samples. With complex proteomes the mass spectrometer cannot sample the LC output quickly enough to perform MS/MS analysis of every peptide as it elutes from the column, but coverage is improved by the use of replicates.³⁶

Directed and targeted proteomic strategies are used when there is prior information which defines specific protein or peptide sets to be analysed.³⁰ A directed proteomics strategy differs from shotgun proteomics in that quantification and identification of tryptic peptides occur in separate experiments.³⁰ At least two LC-MS or LC-MS/MS analyses are undertaken; one collects survey scans for quantification, and the second records product-ion spectra based on an inclusion list generated following analysis of the survey scans to determine which peptides need to be identified.³⁷

Targeted proteomic studies differ from directed studies in that targeted experiments are hypothesis driven. This means that a specific subset of peptides is targeted in the analysis.³⁰ Targeted analyses are usually performed on a triple quadrupole mass spectrometer, using selected reaction monitoring^{38,39} of a set of target peptides.

1.2. Mass spectrometry

Mass spectrometry is an instrument-based analytical technique, first introduced by J. J. Thomson in the early 1900s, whereby the mass-to-charge ratio (m/z) of ions from the sample under investigation is determined. A mass spectrometer is made up of three main components; an ionisation source, a mass analyser and a detector.

1.2.1. Ionisation

1.2.1.1. Electrospray ionisation (ESI)

Electrospray ionisation (ESI) was developed by Fenn *et al.* in the 1980s, and is termed a soft ionisation technique due to the low energy input and consequently very limited fragmentation of the analyte.⁴⁰ A schematic of an ESI source is shown in Figure 1.2.

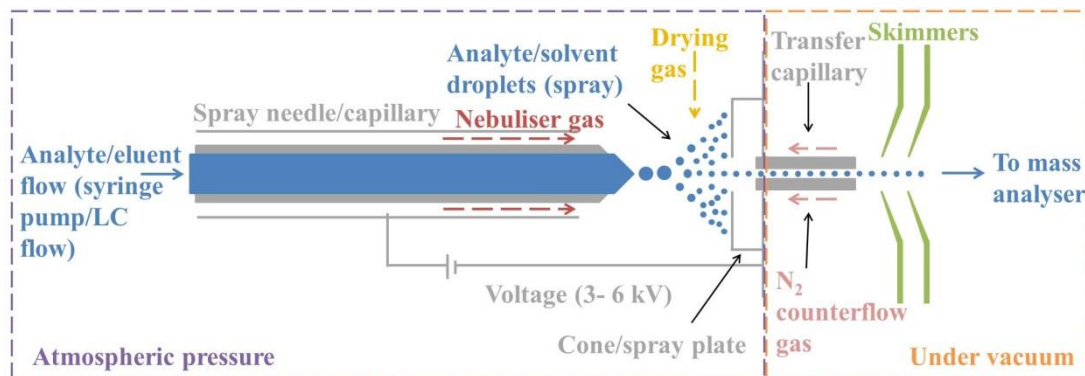


Figure 1.2: ESI source schematic

An analyte is introduced in solution through a capillary at atmospheric pressure.⁴⁰ The capillary has a voltage applied to it, which generates analyte ions (either positive or negative depending on the required mode) within the solution.^{40,41} The charges in the liquid repel each other and lead to the liquid at the surface 'stretching out' and eventually to droplets pinching off to 'escape' from the repelling charges. This is known as the formation of a Taylor cone (Figure 1.3).^{42,43}

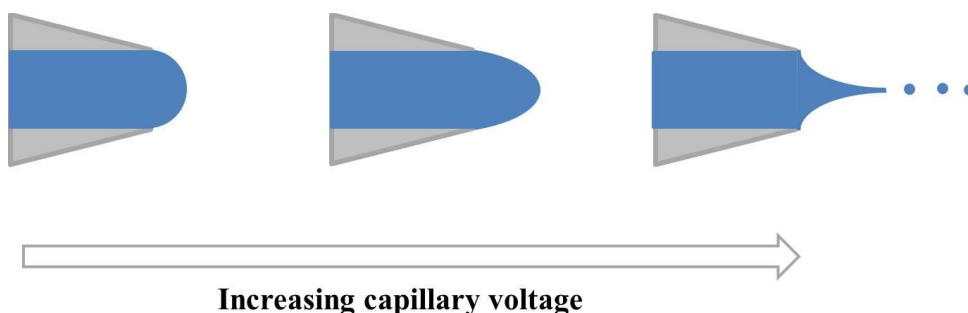


Figure 1.3: Formation of a Taylor cone from the spherical droplet emerging from the capillary (left) to a Taylor cone (right)

The droplet at the end of the capillary is initially spherical but it elongates due to charge accumulation in the capillary tip, forming the Taylor cone. Eventually the Coulombic repulsion forces within the droplet are higher than the surface tension and a stream of small droplets is released from the surface.⁴⁴ The stream of droplets is released into the ionisation chamber where they are attracted to the cone/spray plate because the cone/spray plate is

held at a lower potential than the capillary.⁴¹ The expelled droplets undergo solvent evaporation as they travel towards the analyser; solvent evaporation is aided by a stream of drying gas through which the droplets pass. The analyte ions are subjected to a counterflow of nitrogen gas as they pass through the transfer capillary in order to remove remaining solvent vapour, before subsequently being accelerated by the skimmers, which are set at increasing potential, towards the mass analyser.

There are two proposed methods for ion formation from the charged droplets: the charged residue model (IUPAC nomenclature)^{45,46} and the ion-evaporation model.⁴⁷ The charged residue model was proposed by Dole *et al.* in the 1960s.⁴⁶ In the charged residue model, the droplet size decreases through solvent evaporation, resulting in an increase in the charge density at the droplet surface. Eventually the Rayleigh limit is reached, where the Coulombic force of ion repulsion overcomes the solvent surface tension. This causes the droplet to undergo Coulombic explosion, creating smaller droplets. This process of solvent evaporation, charge repulsion and Coulombic explosion is repeated until all the solvent has been removed, leaving a series of unsolvated charged analyte ions (Figure 1.4).

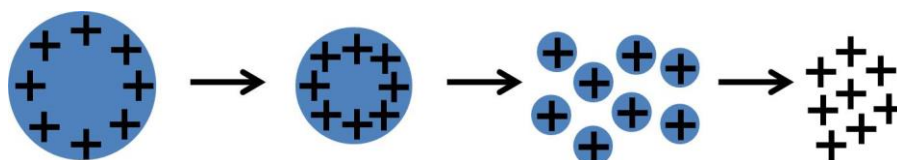


Figure 1.4: Charged residue model

The ion-evaporation model was proposed by Iribarne and Thomson in the 1970s.⁴⁷ The model assumes that when a droplet reaches a critical radius, through evaporation of solvent, the droplet surface field strength becomes high enough to desorb a solvated ion directly from the droplet and into the gas phase (Figure 1.5). This temporarily reduces the surface tension within the remaining droplet, until further solvent evaporation occurs and the charge density within the droplet becomes too high again causing further desolvation. The process of evaporation and ion desorption continues until a series of charged analyte molecules are left.

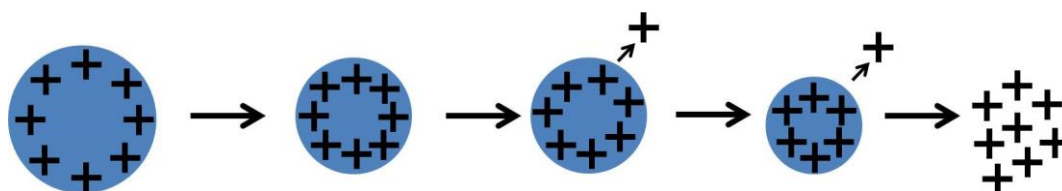


Figure 1.5: Ion-evaporation model

Limited fragmentation is observed in ESI mass spectra.⁴⁸ Usually, the molecular species observed are either protonated ($[M+H]^+$) or deprotonated ($[M-H]^-$) molecules, depending on

whether the voltage applied to the capillary is positive or negative.⁴⁴ Other molecular species, such as $[M+Na]^+$ or $[M+Cl]^-$ can arise, depending on contaminants present in the sample, or additional elements in the mobile phase (e.g. buffers or modifiers). It is also likely that high molecular mass analytes, such as proteins, with multiple protonation sites can acquire multiple charges.^{41,44,46}

1.2.1.2. Matrix-assisted laser desorption/ionisation (MALDI)

Matrix-assisted laser desorption/ionisation (MALDI) was introduced as a mass spectrometric ionisation technique by Karas and Hillenkamp in the 1980s.^{49,50} It uses a matrix and laser to ionise non-volatile molecules, including proteins and peptides. As with ESI, little fragmentation is observed in MALDI mass spectra.⁴⁹

The matrix - a small organic molecule with a strong UV absorption at the wavelength of the laser - is combined with the analyte on a MALDI plate to insulate the analyte molecules, absorb laser light and transfer the consequent excitation to the analyte.⁵¹ Molecules, such as 2,5-dihydroxybenzoic acid (DHB), α -cyano-4-hydroxycinnamic acid (CHCA) and 2',4',6'-trihydroxyacetophenone monohydrate (THAP) are all typical MALDI matrices (Figure 1.6). CHCA and DHB are often used in the analysis of proteins and peptides, whereas THAP is often used for oligonucleotides.⁵² The sample and the matrix are each dissolved in a solvent. Aliquots of the sample and the matrix solutions can be combined and spotted onto the MALDI plate. Alternatively, the two solutions can be spotted separately onto the same well on the plate. The well is dried (usually by leaving it to air-dry), causing a deposit on the surface of the plate of matrix and analyte crystals, which can then be analysed.

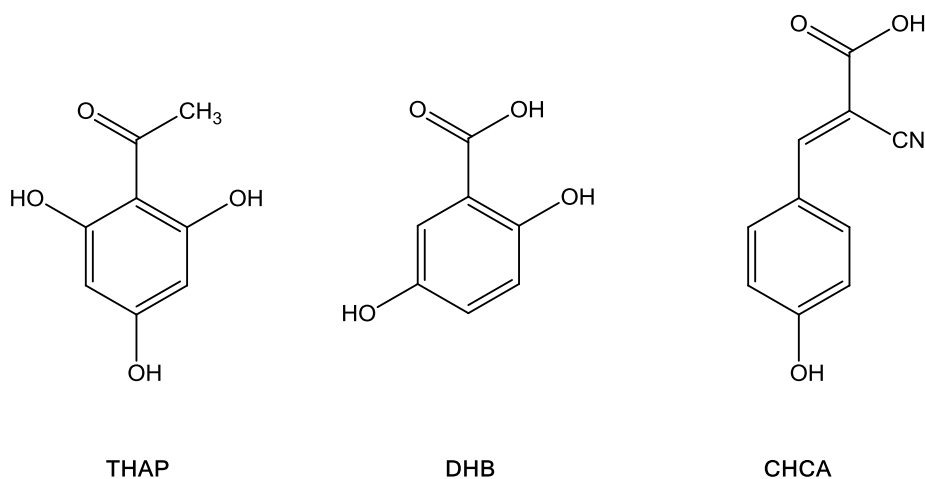


Figure 1.6: Matrix molecules used for MALDI

When the MALDI plate is in the mass spectrometer the crystals are subjected to a brief pulse of irradiation from a laser beam (Figure 1.7).

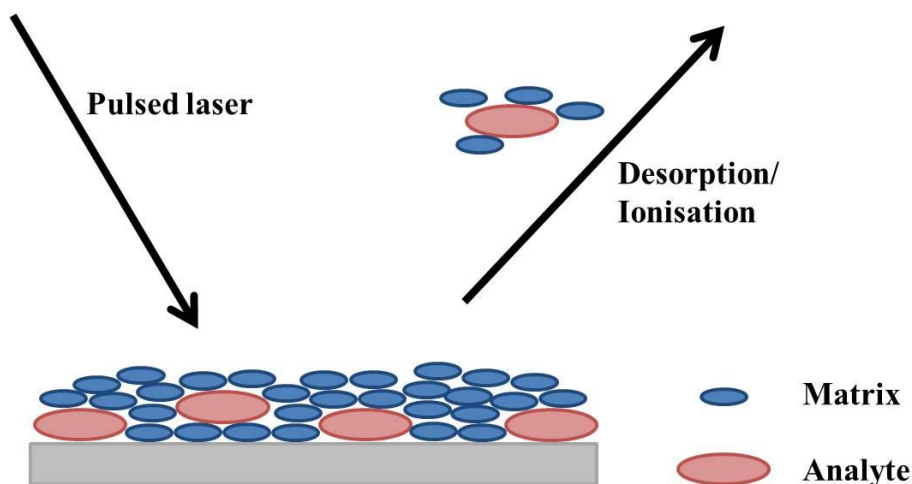


Figure 1.7: MALDI ionisation process

The laser beam ablates portions of the solid deposit where it hits the surface of the sample spot. There are many theories as to how the solid deposits on the surface of the MALDI plate generate the ions seen in the resulting mass spectrum. In 1998, Zenobi and Knochenmuss stated that it is very probable that there is no one mechanism that can explain all of the ions observed in a MALDI mass spectrum.⁵³

In early publications, gas-phase photoionisation, ion-molecule reactions, disproportionation, excited-state proton transfer, energy pooling, thermal ionisation and desorption of preformed ions were all proposed as pathways for MALDI ion formation.⁵³ It was proposed that there were two major processes which occur at the ionisation stage: generation of primary ions and subsequent conversion to more favourable secondary products.⁵⁴ The primary ions are generated either during or shortly after the laser pulse, with the secondary ions (e.g. radical cations, protonated molecules and cationised molecules) being the commonly observed ions in positive ion mode.⁵⁴ There are many proposed mechanisms for the formation of primary ions. The most favoured by Zenobi and Knochenmuss for positive mode UV-MALDI were multiphoton ionisation, energy pooling and multicentre models and desorption of preformed ions.⁵³ Primary ions go on to undergo further reactions to form the secondary products, which are observed in the mass spectrum.⁵⁴ Conditions within the plume must be considered for secondary ionisation mechanisms. Molecular dynamics simulations show that the plume can be described as being formed on very rapid, even explosive, solid-to-gas phase transition.⁵³ The secondary ionisation mechanisms most favoured by Zenobi and Knochenmuss are gas-phase cationisation and electron transfer.⁵³ Alkali cations, such as sodium and potassium, do not need to be added to the sample in order for $[M+Na]^+$ and $[M+K]^+$ ions to be detected. This is due to the fact that sodium and potassium are known to be ubiquitous contaminants observed in mass spectrometry.⁵³

Two of the favoured ionisation mechanisms in MALDI-MS are the lucky survivors hypothesis⁵⁵ and gas phase protonation⁵³ model.

The lucky survivors hypothesis was proposed by Karas *et al.* in 2000.⁵⁵ It is based on the premise that the ions detected are pre-formed (i.e. they exist in the sample spot before ablation). During ablation, most ions recombine with counterions, but a few (the lucky survivors) survive and go on to be detected. The lucky survivors mechanism provides an alternative ionisation model to those proposed previously, but retains elements such as photoionisation. As a result of laser desorption, small clusters of material are released from the surface of the sample spot, which consists of matrix, analyte and ionic species embedded in the matrix crystal. The clusters have enough excess energy to evaporate/desolvate to molecular and ionic species. Upon evaporation of the matrix molecules ion-ion and neutral-ion chemical reactions occur, driven by reaction enthalpy. Highly pre-charged analytes are neutralised, which is a strongly favoured reaction. However, a percentage of these analytes retain a single charge, and these analytes are known as the lucky survivors, which are observed in MALDI mass spectra.

The gas phase protonation model was proposed in a paper by Ehring *et al.* in 1992.⁵⁶ It is based on the hypothesis that neutral analytes are released into the gas phase after desorption/ablation, either through incorporation in the matrix crystals as uncharged species, or from charge recombination of a precharged analyte and its counterion. It is also possible for neutral matrix molecules to be protonated in a similar manner. The neutral analytes can undergo gas phase collisions with protonated or radical cation matrix species, to form charged, protonated analytes.

Karas and Krüger published a paper in 2003 in which the cluster ionisation mechanism was discussed as a refinement to the lucky survivors mechanism.⁵⁷ There is evidence to show that a mixture of matrix clusters and neutral evaporation are the fundamental ablation products. These clusters then break down and react en route to the detector, and matrix and analyte ions are observed in the mass spectrum. The presence of these clusters alters the MALDI ionisation process, with ionisation being linked to the occurrence of the clusters.⁵⁷

Jaskolla and Karas published a paper in 2011, in which they present evidence for both the lucky survivor model and the gas phase protonation model.⁵⁸ They proposed that both of the models were valid pathways to protonation in MALDI-MS. They also showed that one ionisation pathway can be more favourable than the other, depending on the conditions used. For example, increasing the laser fluence favours the gas phase protonation model. Ultimately, a unified MALDI analyte protonation mechanism was proposed, which combines and connects elements of both ionisation models (Figure 1.8).

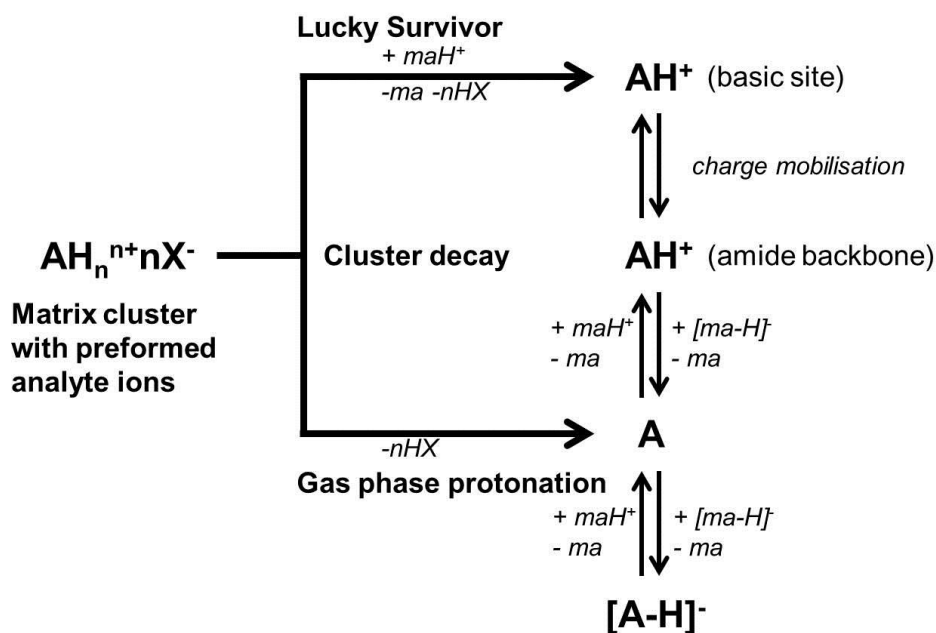


Figure 1.8: The unified MALDI analyte protonation mechanism restricted to positively precharged analyte ions such as peptides and proteins⁵⁸ (where A = analyte, ma = matrix, X = counterion, H = proton)

In 2012, a paper by Bae *et al.* discussing the degree of ionisation of peptides in MALDI-MS was published.⁵⁹ The authors suggested that the current lucky survivor and gas phase protonation models were both valid to a degree for peptide ion formation, but both could be modified to describe gas-phase peptide ion formation. The authors proposed, from experimental data, that laser ablation releases a peptide into the gas phase either as a free protonated peptide (Hypothesis 1), as a protonated peptide in a neutral ion pair (Hypothesis 2) or as a neutral peptide (Hypothesis 3). In Hypotheses 2 and 3, peptide ions were assumed to form via gas-phase reactions of peptide neutrals with matrix-derived cations. In Hypothesis 1, the ion emission is assumed to occur only for preformed analyte ions which are dielectrically screened from counter anions by neutral matrix molecules in a solid sample, and hence are easily released into the gas phase.

There is still uncertainty as to the exact MALDI mechanism, but the lucky survivor and gas phase protonation models seem to be accepted as the models which best explain the MALDI process.⁵⁸ However, there is still no definitive explanation of the ionisation process, as can be seen from the fact that authors of papers in recent years have adopted aspects of both models, depending on what their experimental data showed.

1.2.2. Mass analysers

The mass analyser separates the ions formed according to their m/z values. The way in which ions are separated depends on the type of mass analyser used. The time-of-flight (ToF) mass analyser and the Fourier transform ion cyclotron resonance (FT-ICR) mass analysers will be discussed, as these have been used in the work presented in this thesis.

1.2.2.1. Time-of-flight (ToF)

The ToF mass analyser was first described by Stephens in 1946,⁶⁰ and the first commercial instrument was developed by Wiley and McLaren in the 1950s.⁶¹ The ToF mass analyser became a prominent tool in the field of mass spectrometry in the 1960s.⁶²

A ToF mass analyser works by measuring the time it takes the ions produced in the ion source to travel through a field-free region, known as the flight tube, and reach the detector. There are two types of ToF mass analyser: the linear⁶¹ and reflectron⁶³ forms.

As described by Wiley and McLaren, in the linear ToF (Figure 1.9), the ion packets travel from the ion source, through a flight tube, and hit a detector at the other end.⁶¹

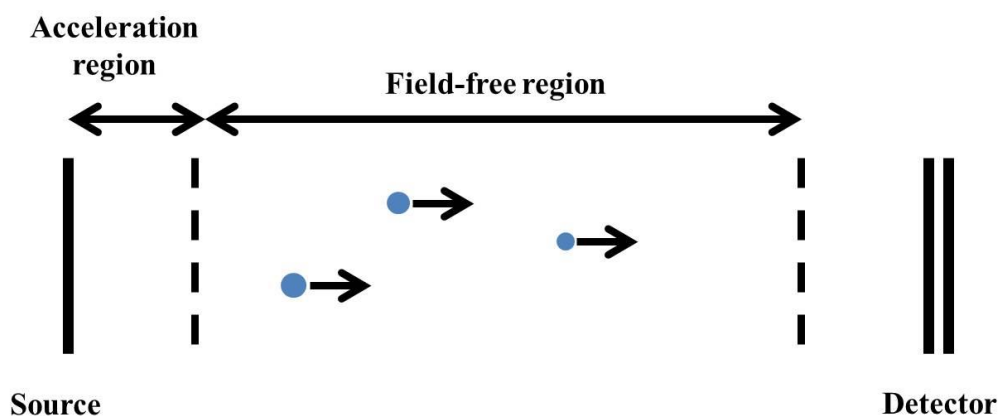


Figure 1.9: Linear ToF mass analyser in linear orientation

Bundles of ions are expelled from the source and accelerated into a flight tube by a potential difference. The ions separate in the flight tube based on their velocity (v), which is proportional to their m/z value. The kinetic energy (E_k) of an accelerated ion depends on its mass (m) and charge ($q = ze$), but also on the accelerating potential (V_s) (Equation 1.1).⁶²

$$E_k = \frac{mv^2}{2} = zeV_s$$

Equation 1.1: Equation relating the kinetic energy of an accelerated ion to its velocity mass

The velocity of an ion can be found by rearrangement of Equation 1.1:

$$v = \left(\frac{2zeV_s}{m} \right)^{\frac{1}{2}}$$

Equation 1.2: Rearrangement of Equation 1.1, to give velocity as the subject

The time (t) taken for an ion to travel the length (L) of the flight tube is given by Equation 1.3.

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

Equation 1.3: Equation relating the time taken for an ion to travel the length of the flight tube to its velocity

By replacing velocity in Equation 1.3 with the result of Equation 1.2, m/z values can be calculated from the time taken for different ions to travel the length of the flight tube:

$$t^2 = \frac{m}{z} \left(\frac{L^2}{2eV_s} \right)$$

Equation 1.4: Equation showing that m/z can be calculated from a measurement of the time taken to travel the length of the flight tube

As the ions are accelerated into the flight tube, a spread of kinetic energies is possible for each m/z value.⁶³ This is due to the fact that when the accelerating potential is applied, the distribution of the potential is not even. This can mean that ions with the same m/z do not experience exactly the same potential, and subsequently do not have exactly the same kinetic energy. The distribution of kinetic energies can be made worse by collisions within the plume, which can transfer energies between ions, and result in an uneven distribution of kinetic energies.⁶²

In a linear ToF, ions with the same kinetic energy can arrive at the detector at different times. This spread in detection times can be caused by differences (prior to acceleration) in the position of the ions within the source, in the direction of travel of the ions and in the time at which ions of the same m/z are created.⁶² This means that ions of the same m/z can hit the detector at different times, which can result in peak broadening and reduced resolution in the resulting mass spectrum.

The effects of kinetic energy spreads can be reduced using a reflectron, an idea which was introduced by Mamyryn *et al.* in 1973.⁶³ In the reflectron ToF (Figure 1.10), the ions enter the flight tube and reach an ion mirror or reflectron - a series of rings of increasing potential.⁶³

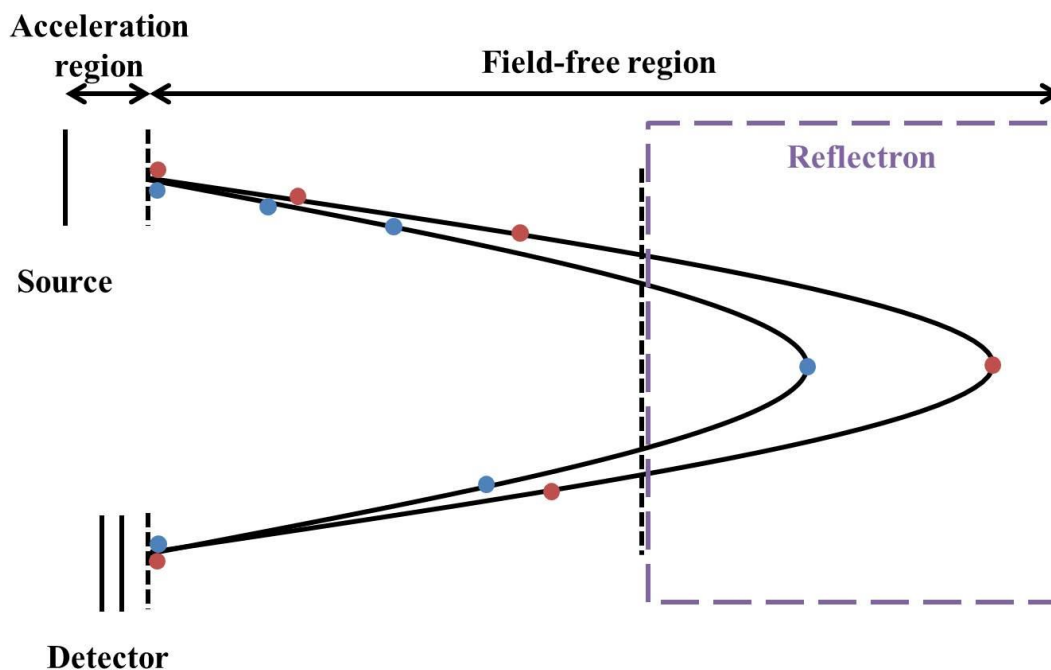


Figure 1.10: Reflectron ToF mass analyser (both ions have the same m/z , but the blue circles represent ions with less kinetic energy than the red circles)

The ions which have the most kinetic energy will travel further into the ion mirror, whilst ions with less kinetic energy will travel less far into the ion mirror. When the ions leave the reflectron, ions of the same m/z value should now arrive at the detector at the same time, creating a more uniform response at the detector, which should result in better resolution in the mass spectrum. One of the additional benefits of the reflectron is that the ions essentially travel twice the length of the flight tube, which increases the resolution of the mass analyser when compared to a linear ToF.⁶³

ToF mass analysers are clearly very appropriate for use with pulsed ion sources (such as MALDI) because the pulsed ion source generates packets of ions. These packets of ions have a short, precisely defined ionisation time, which generates defined packets of ions to be accelerated into the flight tube. Therefore, a packet of ions can be accelerated into the flight tube and detected, before the next packet of ions is generated. In the 1980s, Guilhaus and Dodonov 'rediscovered' orthogonal acceleration, which readily allows continuous ion sources (such as ESI) to be used with ToF mass analysers.^{64,65} The original work using an orthogonal geometry was attempted in the 1960s by O'Halloran, but the work was not published in the general scientific literature.⁶⁵

In orthogonal acceleration instruments (Figure 1.11), the continuous stream of ions generated in the ion source is focussed into a parallel beam which travels through an orthogonal accelerator.⁶⁴ The orthogonal accelerator produces an injection pulse at regular time intervals, which is orthogonal to the ion beam.⁶⁵ This raises the kinetic energy of the

ions in the direction orthogonal to the beam, and sends them towards the orthogonally placed flight tube.⁶⁴ The accelerator has two modes: a fill-up mode and a push-out mode.^{64–66} In the fill-up mode, the first stage of the accelerator must be field-free, to avoid ions in the ion beam being deflected from their trajectory, and causing a spatial spread of the ions. The ion beam enters the orthogonal accelerator at a right angle to the entrance to the flight tube. If there is no field, the ions are able to continue in their parallel ion beam, which fills the orthogonal accelerator. When the push-out pulse is applied, the ions in the orthogonal accelerator are pushed out of the accelerator, in an orthogonal direction to the ion beam. The ions then travel through the flight tube, towards the detector.

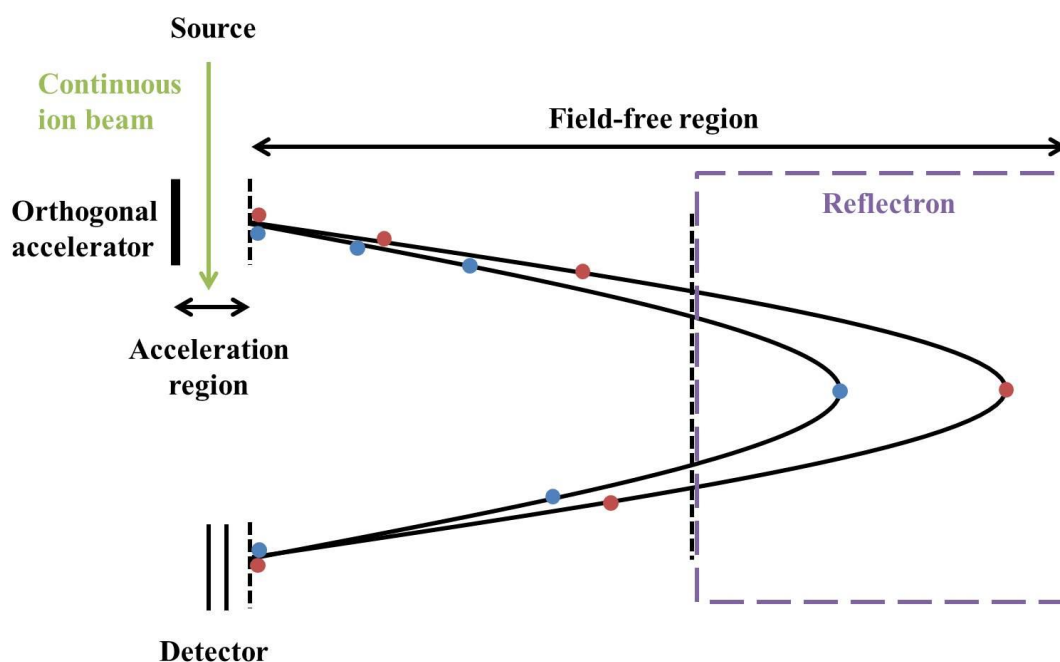


Figure 1.11: Orthogonal acceleration reflectron ToF mass analyser (both ions have the same m/z , but the blue circles represent ions with less kinetic energy than the red circles)

1.2.2.2. Fourier transform ion cyclotron resonance (FT-ICR)

FT-ICR mass spectrometry was first described by Comisarow and Marshall in 1974.^{67,68} The FT-ICR mass analyser works on the basic principle that the trajectory of an ion can be controlled by a magnetic field; if an ion has a low velocity, but is in an intense magnetic field, it can be made to move in a circular motion, thus trapping it in a space.⁵²

The main components of an FT-ICR-MS are a magnet, an ICR cell, an ultra-high vacuum and a sophisticated data system.⁶⁹ The magnet surrounds the ICR cell and is usually superconducting. An ultra-high vacuum is needed to produce the high resolution associated with the FT-ICR mass analyser by minimising collisions within the cell, which subsequently

increases the mean free path of the ions.⁶⁹ There are several stages (or sections) within the mass spectrometer, before the ions reach the ICR cell. Each stage has a higher vacuum than the previous stage.⁷⁰ The ICR cell is where the ions are stored, mass analysed and detected and is typically either cubic (Figure 1.12) or cylindrical in shape.

The basic experimental sequence consists of four stages: quench, ionise, excite and detect.⁶⁹ The quench step removes from the ICR cell ions remaining from a previous experiment, by applying a high potential difference across the trapping plates, which forces ions out of the cell axially. Once the ions are in the cell, they are excited into a coherent motion, which allows the ions to be detected.

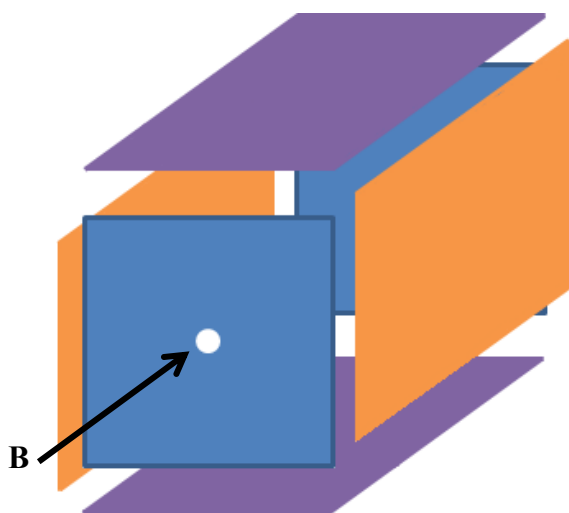


Figure 1.12: An example of a cubic ICR cell, where B = magnetic field (trapping plates are coloured blue)

The ICR cell consists of six plates, which are divided into three pairs.⁷¹ There are two plates (the plates with holes in them) which are known as the trapping plates, or end caps (Figure 1.12), where the ions can enter and exit the cell. Of the two remaining pairs of plates, one pair is known as the excitation plates, and the other pair is known as the detection plates.

Ions enter the cell, with an incoherent motion, meaning that they simply spiral inside the cell, and are not detected.⁷² Once inside the cell, they undergo three types of ion motion: trapping motion, cyclotron motion and magnetron motion (Figure 1.13).⁷⁰

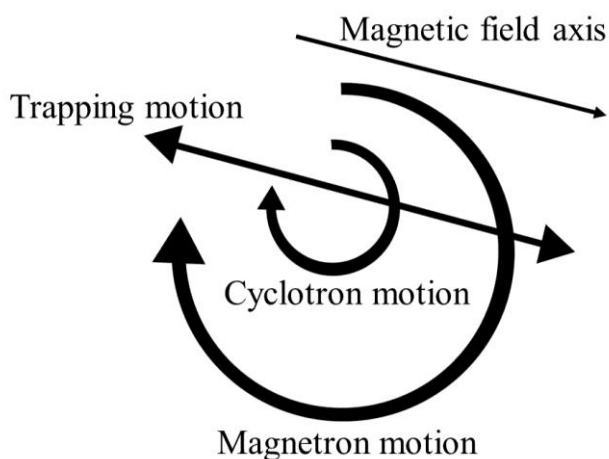


Figure 1.13: The three motions that ions undergo in the cell

The trapping motion traps the ions in the cell in a path parallel to the magnetic field. A voltage applied to the end caps, which are perpendicular to the magnetic field, traps the ions and causes them to undergo a harmonic oscillation. The ions undergo cyclotron motion simultaneously to the trapping motion, causing them to perform a circular motion within the cell. The magnetron motion experienced by ions comes from the combination of the magnetic field and electric field experienced by an ion.⁷⁰ The electric field acts to drive ions away from the centre of the ICR cell, whilst the magnetic field stops the ions from being accelerated into the walls of the cell. The passage of the ions within the ICR cell which is produced due to the combination of the three motions is summarised in Figure 1.14.

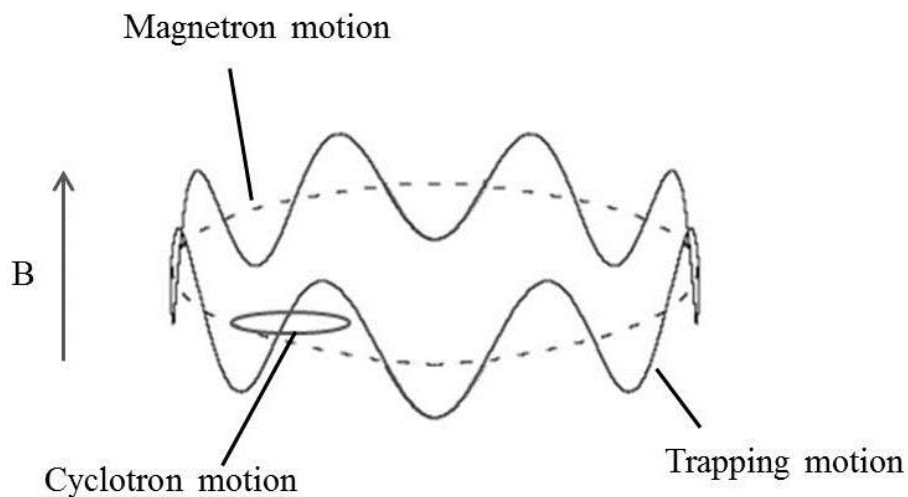


Figure 1.14: The motion of an ion within the ICR cell as a result of trapping, cyclotron and magnetron motions

The frequency with which an ion repeats its orbit within the ICR cell is called its cyclotron frequency.⁷¹ FT-ICR mass analysers determine the m/z value of ions based on their cyclotron frequency in a constant magnetic field.⁷¹ The cyclotron frequency (Equation 1.5) of

an ion is measured experimentally and subsequently converted to its m/z value.⁷² The frequency (f_c) is inversely proportional to the m/z of the ion, and directly proportional to the strength of the applied magnetic field (B). As the magnetic field remains constant, the only unknown quantity is the m/z value. As can be seen in Equation 1.5, the lower the m/z value, the higher the cyclotron frequency.

$$f_c = \frac{zB}{2\pi m}$$

Equation 1.5: Cyclotron frequency, and its dependence on mass, charge and magnetic field strength

If all the ions enter the cell with the same kinetic energy, they are in a packet. This packet then travels backwards and forwards within the cell (due to the trapping motion). When the ions enter the cell, their cyclotron radius is too small to be detected.⁷¹ Ions at each m/z are excited by applying a sweeping radiofrequency (RF) pulse across the excitation plates, which excites all m/z values.⁷³ This is a sinusoidal voltage, which increases the kinetic energy of the ions, and increases their radius when the RF is the same as the cyclotron frequency.⁷⁰ When the two values are the same, the ions absorb energy from the RF and their cyclotron radius is increased. The increase in radius allows the ions to be detected by the detector plates.⁶⁹ As the packet of orbiting ions passes near each of the opposing pairs of detector plates, electrons are attracted first to one plate and then the other. For example, if the instrument is operated in positive mode, the ion packet travelling towards the plate attracts electrons in the plate. This creates a differential current between the two detection plates, forming a sinusoidal image current, which is then digitised by the computer software.⁷⁰ The time domain sinusoidal curve undergoes Fourier transformation to produce a frequency spectrum, which is then converted to a mass spectrum.

Over time, or if the ions do not all enter the cell with the same kinetic energy, a packet of ions with the same m/z may spread out along the axis of the cell.⁷⁴ However, this does not necessarily affect the detection of the ions as they will still pass each of the detector plates at the same time, just at different points along the length of the cell.

The high resolution achievable using FT-ICR-MS is a consequence of the length of time that the ions are detected (transient duration, T)⁷⁰ and the cyclotron frequency: resolution increases as transient duration and cyclotron frequency increase (Equation 1.6).

$$R = \frac{f_c T}{2}$$

Equation 1.6: Maximum resolution achievable in FT-ICR-MS

1.2.3. Tandem mass spectrometry

Soft ionisation (such as ESI or MALDI) gives very few fragment ions, but for structure analysis fragments need to be generated. Fragments can be generated by collisions with a neutral gas following ion acceleration.

There can either be two mass analysers in line, with a collision cell in the middle to cause fragmentation (MS/MS 'in-space'), or there can be one mass analyser which performs all the functions (MS/MS 'in-time').⁷⁵ The typical set-up for an in-space tandem mass spectrometry experiment is shown in Figure 1.15, with the two mass analysers used for recording data referred to as MS1 and MS2. The collision cell is where the ions from MS1 may be fragmented, before being passed through to the second mass analyser.



Figure 1.15: Typical arrangement of an in-space tandem mass spectrometer

There are four major MS/MS experiments: product ion, precursor ion, constant neutral loss and selected reaction monitoring experiments.⁷⁶ In a product ion experiment, which is often used for sequencing peptides and oligosaccharides, and for analyte identification, the first mass analyser selects a specific precursor ion.⁷⁷ This ion then undergoes fragmentation, and the product ions are analysed in the second mass analyser. The resulting mass spectrum shows the ions which have been produced from the selected precursor.

Fragmentation of ions can be achieved using different techniques, with the most commonly used technique being collision induced dissociation (CID).⁷⁸⁻⁸¹ In CID, the ions are accelerated and then collided with a neutral gas (such as helium or nitrogen) in the collision cell. The collisions increase the internal energy of the ions by converting some of their kinetic energy to internal energy, which causes subsequent fragmentation of the ions through bond breakage.⁷⁹ The mass of the collision gas can influence the amount of energy that the ion receives, and therefore the maximum amount of kinetic energy which is converted to internal energy.⁸²

When a protein or peptide fragments (in most modern spectrometers, this is by low energy CID), it fragments in a well-defined way.⁸³ All three bonds in the peptide backbone are susceptible to breaking to different extents. When the bond in a singly-charged precursor breaks, the resulting charge can be left on either of the resulting fragments (Figure 1.16). If the charge is on the peptide's N-terminal portion, the fragment produced is termed an a-, b- or c-ion.⁸³ If the charge is on the peptide's C-terminal portion, the fragment is labelled either an x-, y- or z-ion.⁸³ In addition to being labelled as an a-, b-, c-, x-, y- or z-ion, the fragment

is given a number. The number is dependent on the number of amino acid residues in the fragment produced. For example, if it is the second amino acid from the N-terminus, and the charge is on the N-terminal portion, either an a_2^- , b_2^- or c_2^- ion will be formed.

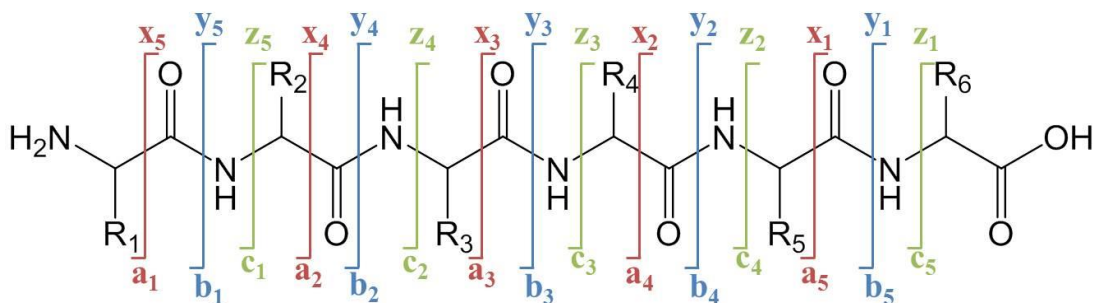


Figure 1.16: The fragmentation which can occur in a peptide. The different colours indicate the different types of bonds which can break.

The amide (peptide) bond between the amino acids is the most likely to break as it has the lowest energy pathway.⁸⁴⁻⁸⁶ Identification and interpretation of sequence ions, predominantly the b- and y- ions, but also, less frequently with low energy CID, a-, c-, x- and z-ions, allows peptide amino acid sequences to be determined. Immonium ions can be formed if double y/a-type cleavage occurs, which creates an ion derived from a single amino acid.⁸⁴ These cleavages can identify amino acids present in a peptide, but they do not give information relating to where in the sequence these amino acids are.

Computational methods are used to compare experimentally acquired mass spectra with known protein and gene sequences. Search engines (for example Mascot⁸⁷ and SEQUEST⁷⁷) allow protein sequence databases to be queried using experimental data. Experimental mass values are compared with calculated peptide mass or fragment ion mass values, which are generated from predicted proteolytic digestion of proteins in the database.⁸⁸ A database is often specified, to reduce the number of proteins being searched. The peptides from the theoretical fragments which have the 'right' mass are then pulled out, and theoretical spectra are generated for them. These are then matched against the experimental spectra. Mascot uses a probability-based matching algorithm to match the theoretically predicted fragments for all proteins in the database and the experimental fragments.⁸⁷ The software produces a list of possible matches between the experimental results and the theoretical results, and scores them based on closeness of the match between the theoretical and experimental results.

1.3. Protein labelling strategies

When discussing a protein labelling strategy, there are multiple ways the term can be understood. The term protein labelling can refer to isotopic labelling (introduced with or without the use of a chemical label) and chemical labelling (which may or may not introduce an additional chemical moiety sometimes referred to as a tag) of proteins. In the context of this thesis, protein labelling strategies have been used for the chemically-mediated isotopic labelling of proteins at the level of their tryptic peptides.

The use of protein labelling strategies plays an important part in proteomics, in particular in quantitative proteomics as a means of being able to quantify differences in the levels of particular proteins in two (or more) different biological samples.⁸⁹ However, protein labelling strategies have also been used for non-quantitative proteomics experiments (for example, proteins and peptides have been labelled to aid mass spectrometric identification).⁹⁰

Quantitative proteomics has been used to study subjects as diverse as biomarker discovery in pancreatic cancer⁹¹ and cold stress responses in petunia seedlings.⁹² In order to perform a quantitative proteomics experiment, there are a number of different routes which can be taken (Figure 1.17). The different options can be broken down into four different generic strategies: metabolic labelling, chemical labelling, spiked peptides and label free. Within each of these generic strategies, there are then a number of different experiments which can be performed (e.g. chemical labelling can be done using one of a range of different chemical labels). With each of the general strategies, the proteins must be extracted from the cells or tissue, prior to digestion to generate peptides. The resulting peptides are then analysed by mass spectrometry, and the data interpreted.

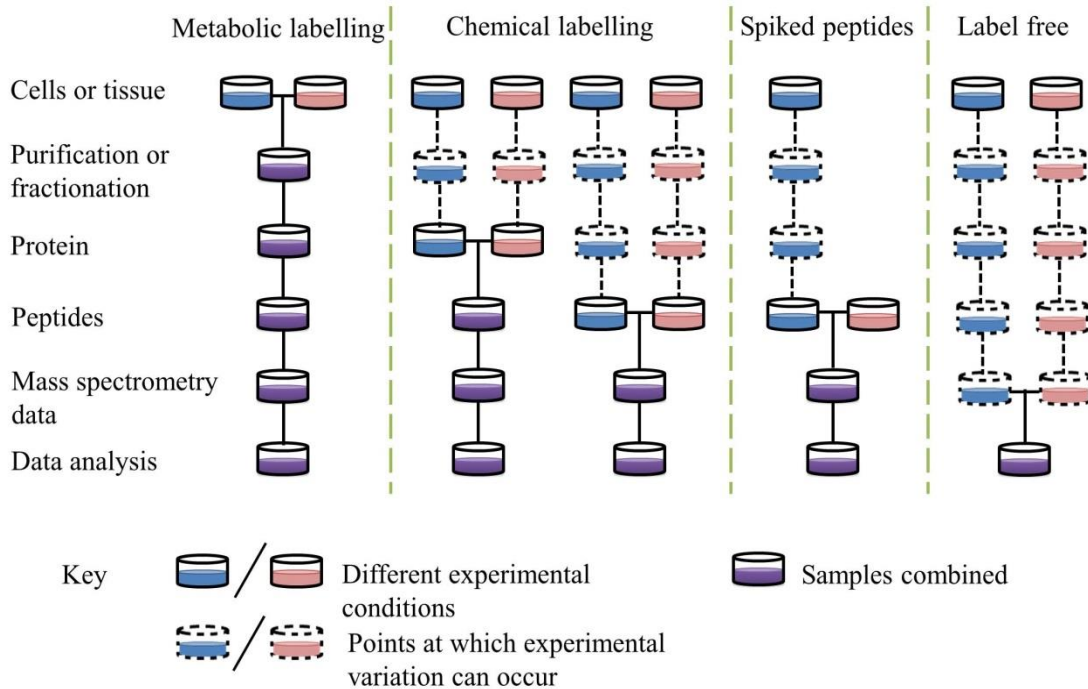


Figure 1.17: Strategies for protein quantification, detailing the possible different stages involved and the potential stages at which experimental variation can occur.⁸⁹

Metabolic labelling strategies involve the incorporation of a label into proteins through the use of an isotopically labelled amino acid (e.g. deuterated leucine) during cell growth and division. Two sets of cells are grown in cell medium. One is grown using non-labelled amino acids. The isotopically labelled amino acid is included in the cell culture medium for the other set of cells, and is incorporated into the proteins synthesised whilst in the cell culture medium.⁹³ Labelling is at the protein level, with the protein extracts resulting from the different experimental conditions being combined prior to any subsequent purification, digestion and analysis.

With a chemical labelling strategy, a label is usually introduced chemically or enzymatically into proteins or peptides.⁸⁹ One example of an enzymatically-mediated labelling strategy is the incorporation of ^{18}O during protein digestion.^{94,95} Chemical labelling strategies generally label peptides generated on protein digestion and the samples are not usually combined until a later stage. The samples from the different experimental conditions are invariably extracted and/or fractionated as individual samples, prior to labelling. Once the peptides have been generated and labelled, the samples from the different experimental conditions are combined. A protein or peptide can be chemically labelled at the N-terminus, the C-terminus or at a particular amino acid side chain within the sequence. Chemical labelling can either introduce an isotopic label into a native component of the protein (such as a carbon, oxygen or nitrogen atom already present in the peptide structure) or introduce a new chemical moiety (such as an additional organic group).

An alternative strategy is to spike synthetic peptides into a protein digest.⁹⁶ After protein digestion, a known quantity of a stable isotopically-labelled peptide is added. These peptides are synthetic versions of peptides produced during the digest, but incorporate stable isotopes to generate peptides with the same amino acid sequence as a peptide from the digest. The spiked digest is then analysed mass spectrometrically, where precise and absolute quantitative determination of the non-labelled peptides can be carried out. Such labelled peptides for spiking can be synthesised, but QconCAT is an alternative strategy for producing labelled peptides, and uses genes to express artificial proteins.^{97,98} A synthetic gene (which encodes for all the required peptides) is designed, inserted into a vector and expressed in *E. coli*. The cell medium is isotopically enriched to ensure that the expressed peptides are isotopically labelled. The artificial protein may be spiked into the samples before digestion, releasing its isotopically labelled peptides when the samples are digested.

Label-free strategies involve the two experimental conditions being kept separate throughout the analysis, until after the data have been acquired. In a shotgun proteomics experiment, label-free peptide quantification is achieved by either measuring mass spectral peak intensities or by spectral counting.⁹⁹ Measuring mass spectral peak intensities of peptide ions has been shown to provide relative quantification data.¹⁰⁰ Spectral counting involves comparing the number of tandem mass spectra assigned to each protein, and using these to compare the relative abundances of these proteins in different samples. It has been demonstrated that the number of times a peptide is selected for fragmentation directly reflects its relative abundance (and thus the relative abundance of its parent protein(s)) in the sample.¹⁰¹

Within all quantification strategies there are stages where experimental variation can occur (Figure 1.17), and so lead to quantification errors. This variation includes the fact that the amount of protein/peptide recovered from one experimental condition may not be exactly the same as that from another. Therefore, two experiments which may have started with similar amounts of protein may end up having differing amounts by the time the samples are combined, leading to errors in the quantification.

Metabolic labelling strategies incorporate the label within the proteins from the very earliest stage (during cell growth and division), which is a huge advantage of the metabolic labelling strategy compared to the other three strategies.⁹³ This is because pooling of samples can occur immediately after labelling, and so the possible experimental variation is dramatically reduced. As the labelled proteins are combined at the level of intact cells, this reduces sources of quantification error. With chemical labelling strategies, experimental variation can occur at any of the stages in the analysis that take place before samples are combined, as the individual samples undergo separate procedures (e.g. sample clean-up as well as labelling) before they are combined. Sample could be lost during these stages, or labelling efficiencies may vary, which would not be accounted for as the samples are handled

separately. The spiked peptides strategy can be subject to experimental variation, particularly if the spiking does not occur until protein digestion (as depicted in Figure 1.17). Like with the chemical labelling strategies, sample could be lost from, for example, sample clean-up or fractionation. With label-free strategies, there is the chance for experimental variation to occur at all of the different stages between generating the samples, obtaining the proteins and acquiring the data.

There are a number of criteria which should be met when considering the use of a label. These include the fact that the labelling strategy should label all proteins or peptides present in a sample in very high yield, and there should be minimal treatment of the sample to add the label.¹⁰² Mass spectrometric analysis is generally used to analyse labelled peptides and proteins. There are many tags which are commonly used in the world of bio-molecule analysis.

1.3.1. Stable isotope labelling by amino acids

Stable isotope labelling by amino acids in cell culture (SILAC) is an *in vivo* metabolic labelling strategy first introduced in 2002 to monitor the relative abundance of proteins when cells are grown in different culture media.⁹³ One culture medium contains standard amino acids ('light'), whilst the other(s) contains isotopically-modified amino acids ('heavy').⁹³ When the technique was first introduced by Ong *et al.* in 2002,⁹³ the group used deuterated leucine as the isotopically-modified amino acid to investigate the relative quantification changes in protein expression during muscle cell differentiation. Other isotopically-modified amino acids commonly used are lysine and arginine where (usually) ^{13}C and ^{15}N are incorporated into the amino acids.¹⁰³ These amino acids are incorporated into the proteins synthesised by the cells or organism grown in the 'light' or 'heavy' media. Enough cycles of cell growth must be completed to ensure that the isotopically-modified amino acids are fully incorporated into the proteins growing in the modified medium.¹⁰³ The different experimental conditions are combined and processed as a single sample. The resulting peptides are subjected to mass spectrometric analysis (Figure 1.18), with quantification carried out at the peptide level.

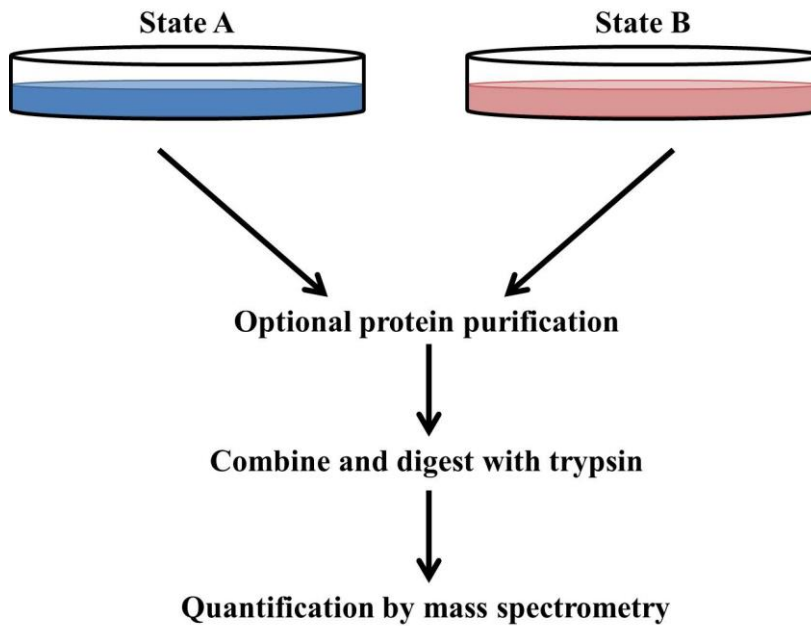


Figure 1.18: Schematic of SILAC labelling strategy (where blue and red colours represent the two different cell cultures)

When SILAC was first introduced as a labelling strategy, it was as a duplex experiment.⁹³ However, it is possible to perform a triplex experiment by use of 'light', 'medium' and 'heavy' labels.¹⁰⁴ In a standard duplex experiment, pairs of peptides which differ by a set mass difference, as determined by the number of heavier isotopes incorporated into the modified amino acids and the number of those amino acid in the peptides, can then be identified.

1.3.2. Isotope-coded affinity tag

The isotope-coded affinity tag (ICAT) was introduced as a chemical label for quantitative proteomics in 1999.¹⁰⁵ It contains a biotin affinity tag, linker region and cysteine-reactive group (Figure 1.19). ICAT is an example of a residue-specific chemical label, as it only reacts with cysteine side-chains (through the cysteine-reactive group). The biotin affinity tag is used as a way of isolating the ICAT-tagged peptides through avidin affinity chromatography.¹⁰⁵ The linker region contains the stable isotopes (either hydrogen or deuterium).

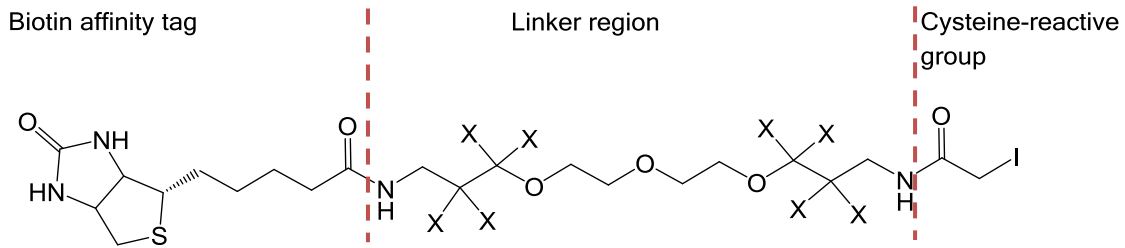


Figure 1.19: ICAT reagent, consisting of three regions: a biotin region to isolate ICAT-labelled peptides, a linker region which incorporates the stable isotopes and a reactive group which reacts with thiols (X = H (light) or X=D (heavy))

One sample is treated with the heavy ICAT reagent, whilst the other is treated with the light ICAT reagent (Figure 1.20).¹⁰⁵ The two samples are then combined, and enzymatically digested. The ICAT-labelled peptides (which must contain a cysteine residue) are isolated using biotin/avidin affinity chromatography, and analysed by mass spectrometry. Peptide pairs are identified by the presence of two peaks in the mass spectrum which differ by one eight mass unit increment for each ICAT moiety incorporated, and which appear in the same peak in the chromatogram.¹⁰⁵ They are then isolated for further analysis, in which their sequence is determined. Quantification of peptides occurs at the MS level by comparing the relative signal intensities of the peaks for the peptide ion pairs.

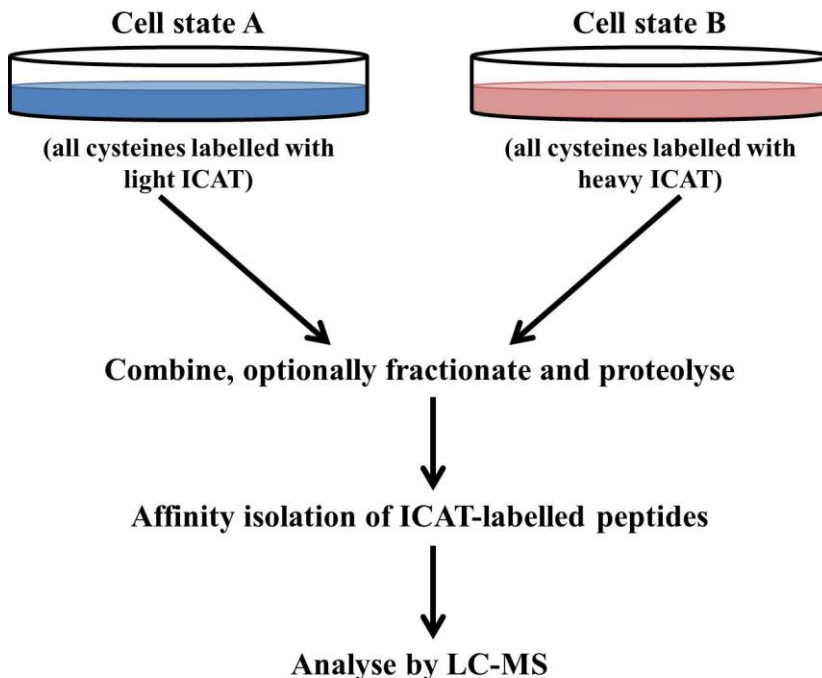


Figure 1.20: The ICAT strategy for quantifying differential protein expression

Gygi and co-workers initially looked at the applicability of the ICAT labelling strategy through the use of a standard protein mixture.¹⁰⁵ Two mixtures consisting of the same proteins, but present at different concentrations were labelled (one with the light ICAT label, and the other with the heavy ICAT label) and treated as schematically illustrated in Figure 1.20. All six proteins were identified and quantified (an example mass spectrum, which was obtained at 19.76 minutes is shown in Figure 1.21). The group then went on to compare protein expression in the yeast *Saccharomyces cerevisiae* using either ethanol or galactose as a carbon source.¹⁰⁵

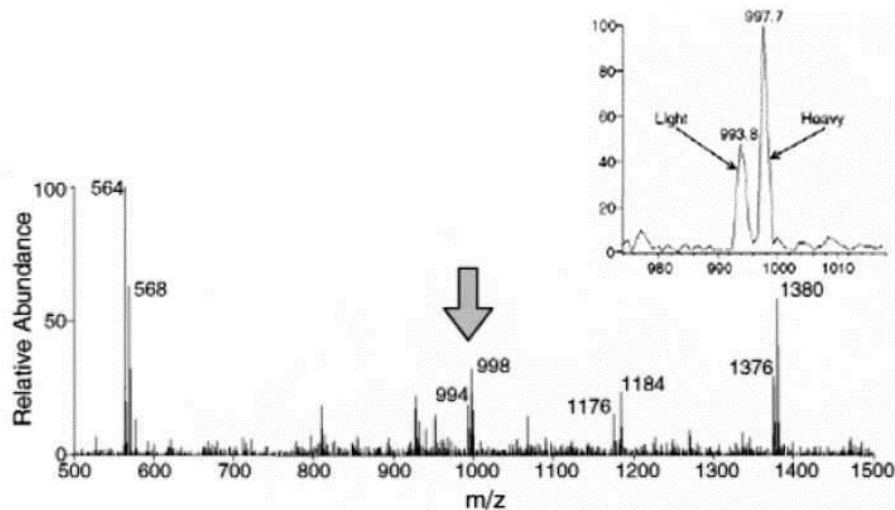


Figure 1.21: ICAT analysis of a protein from a standard protein mixture. Full scan mass spectrum at time 19.76 minutes of the LC-MS analysis (expanded m/z region to illustrate m/z difference between light and heavy labelled doubly charged peptide ions). At least four different peptide doublets were identified.¹⁰⁵

There are limitations to the ICAT technology. For example, the protein must have at least one cysteine residue in order to be represented in the peptides analysed. Another concern is that although the heavy and light versions of the peptides should co-elute in the HPLC, there is a chance that they may have slightly different retention times due to the presence of deuterium, which can affect the chromatography and thus the quantification.¹⁰⁶ Post-translational modifications may not be seen, as they would need to occur on the peptides containing cysteine.¹⁰⁷ There is also the possibility that the avidin affinity stage could suffer from non-specific binding.¹⁰⁷

There have been variations on the original ICAT molecule, such as the acid-cleavable ICAT.¹⁰⁶ This uses ^{12}C and ^{13}C isotopes instead of H and D, which avoids D/H chromatographic separation. It is called 'cleavable' because the linker group can be cleaved under acidic conditions, to remove the biotin.¹⁰⁷

1.3.3. Tandem mass tags

Tandem mass tags (TMTs) were proposed as an N-terminal labelling strategy by Thompson *et al.* in 2003.¹⁰⁸ The tag consists of an amine-reactive NHS group, a mass normaliser group and a mass reporter group. The tandem mass tags are a set of isobaric tags, meaning that each of the tags has the same overall mass. However, when the tagged peptides are fragmented in the mass spectrometer, ions of different m/z values are produced for each of the different isobaric tags. TMT labelling is a peptide quantification method which uses an MS/MS based analysis method. This technique allows for both the simultaneous determination of the identity and relative abundances of peptide pairs using CID. There have been different generations of the tandem mass tag. The first generation (duplex) TMT is shown in Figure 1.22, whilst the current PierceTM Amine-Reactive TMT⁰ is shown in Figure 1.23.

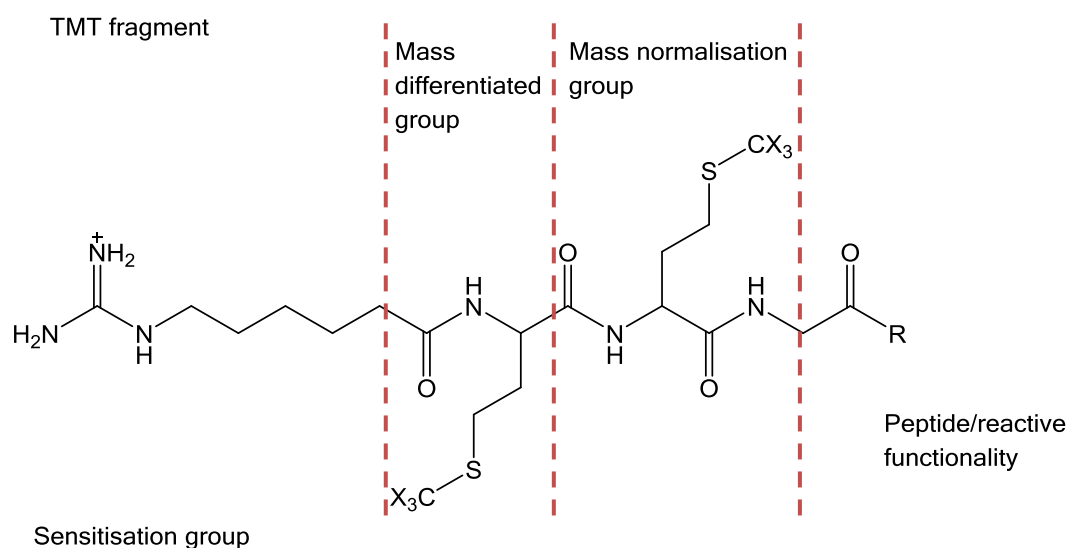


Figure 1.22: First generation of the TMT tag (R = protein reactive functionality, X = H/D)

The first generation of the TMT tag, as shown in Figure 1.22, comprises several different regions: sensitisation group, mass differentiated group, mass normalisation group and a reactive functionality.¹⁰⁸ The combination of the sensitisation group and mass differentiated group make up the TMT fragment that is generated upon tandem mass spectrometric analysis. The tag consists of a series of amide bonds in the backbone, connected to the sensitisation group, with the mass differentiated group and mass normalisation group being methionine residues. The sensitisation group is a guanidino functionality. Guanidination of lysine residues from a tryptic digest has been shown to improve the signal response of C-terminal lysine fragments upon tandem mass spectrometric analysis.¹⁰⁹

In 2008, Dayon *et al.* published a 6-plex TMT strategy which was used for proof of principle studies using standard protein mixtures and also to investigate proteins in human

cerebrospinal fluids.¹¹⁰ This has essentially gone on to become the Pierce™ TMT reagent (Figure 1.23), where a combination of carbon and nitrogen isotopes are used instead of the hydrogen and deuterium used in the first generation TMT, to ensure chromatographic co-elution.

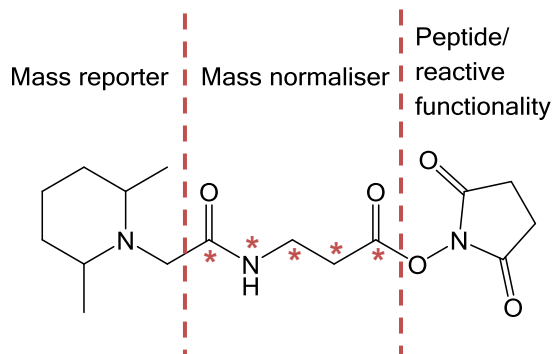


Figure 1.23: Pierce™ TMT reagent (TMT⁰ shown) (* indicates isotope position)

The current Pierce™ TMT reagent (Figure 1.23) bears little resemblance to the first generation TMT reagent devised by Thompson *et al.*. The element of the peptide/reactive functionality denoted 'R' in Figure 1.22, has been retained in the form of the NHS ester of the Pierce™ TMT reagent. However, the mass normaliser region no longer contains deuterium/hydrogen combinations, whilst the mass reporter part of the reagent is also significantly different from that in the first generation TMT.

The Dayon *et al.* paper first discussed the relative quantification ability of the 6-plex TMTs with a model four-protein mixture with protein abundances at six different concentrations.¹¹⁰ Six mixtures were prepared of a model protein mixture (with each mixture containing the same standard proteins), with varying ratios of tryptic peptides from the standard proteins. Each of the model protein mixtures was labelled with a different TMT label, and then aliquots of the six mixtures were combined. An example product ion spectrum obtained for a single peptide from the model protein mixture is shown in Figure 1.24, showing the detail of the reporter ion region used for relative quantification. Quantification is based on the ratios between the intensities of the reporter group ions in the tandem mass spectra.

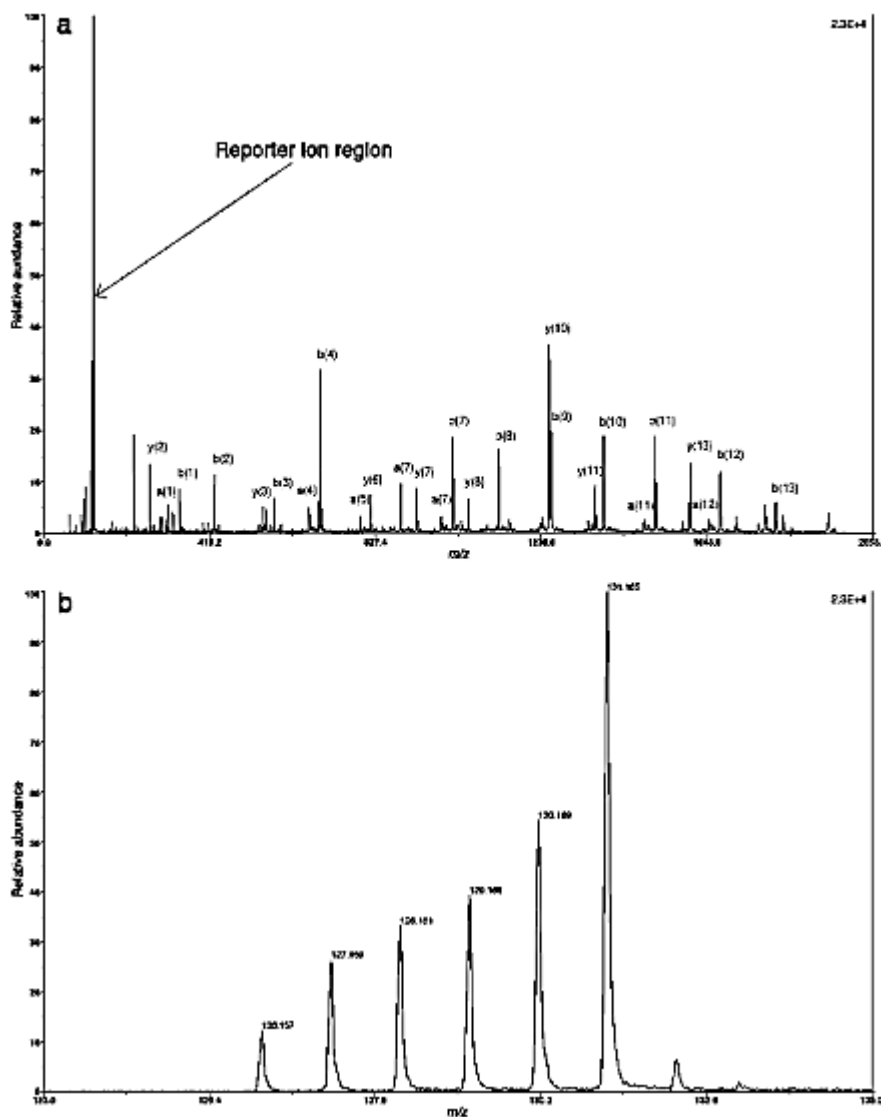


Figure 1.24: (a) Tandem mass spectrum of the peptide LSFNPTQLEECHI of β -lactoglobulin (m/z 1944.902), labelled with TMT and iodoacetamide. (b) Expanded depiction of the reporter ion region providing peptide relative quantification in a model four-protein mixture at six different relative concentrations (1:2:3:3:5:10) through the abundance of the reporter ions at $m/z = 126.1, 127.1, 128.1, 129.1, 130.1$ and 131.1 .¹¹⁰

Cerebrospinal fluid proteins from ante-mortem and post-mortem samples were then interrogated using TMT labelling. Through the TMT labelling, the authors were able to identify a number of proteins with expression differing between the ante-mortem and post-mortem cerebrospinal fluid samples. The results for several of these proteins (glial fibrillary acidic protein, protein DJ-1 and protein S100B) were validated by additional methods.¹¹⁰

The advantages of the TMT labelling strategy, reported by Thompson *et al.* include the fact that the isobaric tagged peptides co-migrate in chromatographic separation.¹⁰⁸ Another

advantage is the use of CID for quantification, which allows untagged material to be ignored.¹⁰⁸ However, the precursor peptide must be isolated exclusively, but this is not always achieved.¹¹¹ If multiple precursor peptides are isolated at the same time, then this can skew the ratios produced.

1.3.4. Isobaric tag for relative and absolute quantification

The isobaric tag for relative and absolute quantification (iTRAQ) (Figure 1.25) was proposed in 2004 as a multiplexing quantification technique.¹¹² As with TMT, the iTRAQ label contains a reporter group and a balancer group and reacts with primary amines.

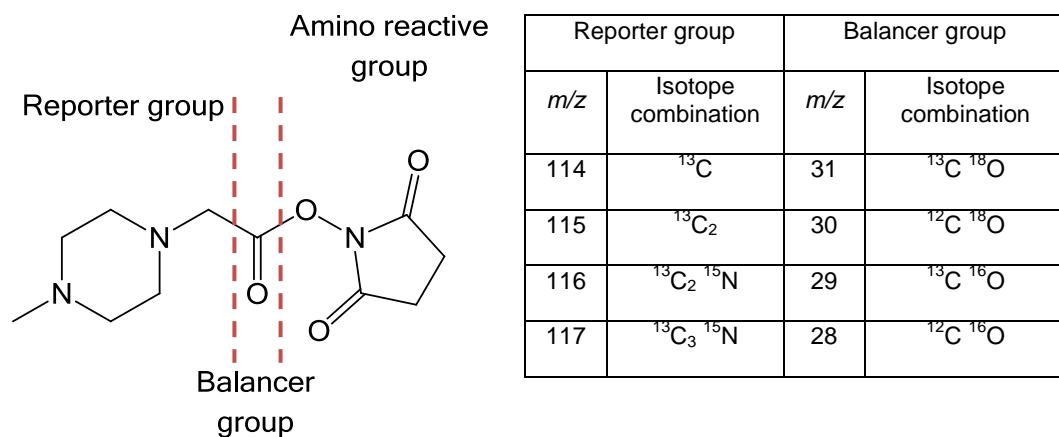


Figure 1.25: iTRAQ label consisting of three regions: a reporter group, a balancer group and a reactive group which targets primary amines. Table shows combinations of isotopes used to create each reporter group with the relevant balancer group, to make isobaric tags^{102,113}

iTRAQ is used as a label after protein digestion, at the peptide level. Proteins are enzymatically digested, and the peptides produced are labelled at the N-terminus (meaning all peptides are labelled), with each digest being labelled with a different version of the iTRAQ reagent (Figure 1.26).

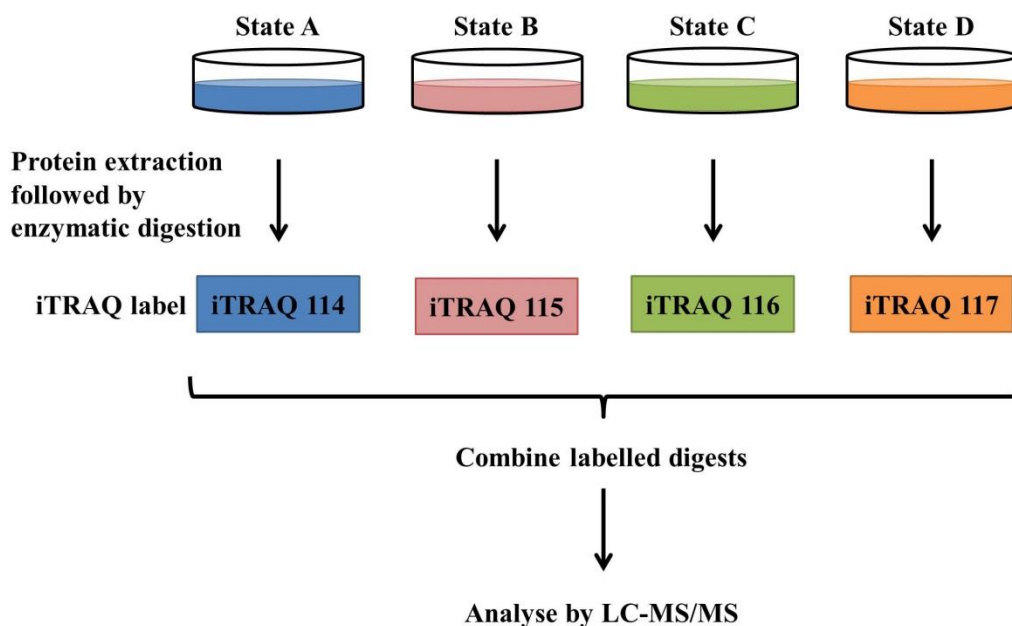


Figure 1.26: iTRAQ strategy for quantitative analysis

There were four versions of the original iTRAQ reagent allowing the comparison of four samples. As with TMT tags, the iTRAQ labels are isobaric, meaning that they have the same overall mass but different combinations of the reporter group and balancer group. Each sample is labelled with a different iTRAQ label. After labelling of the peptides with iTRAQ, the solutions are then combined and analysed by LC-MS. When subjected to MS/MS analysis, the reporter group fragments from the peptide to give rise to an ion at m/z 114-117, depending on the combinations of carbon and nitrogen isotopes used in the reporter group. Quantification is based on the reporter group ions produced on MS/MS (i.e. the intensities of the peaks at m/z 114-117), and the peptide is identified through the fragment ions also observed in the tandem mass spectrum.¹¹²

In 2004, Ross *et al.* published a paper in which they used iTRAQ labelling to examine the global protein expression of a wild-type yeast strain and mutant strains (*upf1Δ* and *xrn1Δ*). An example tandem mass spectrum is shown in Figure 1.27.

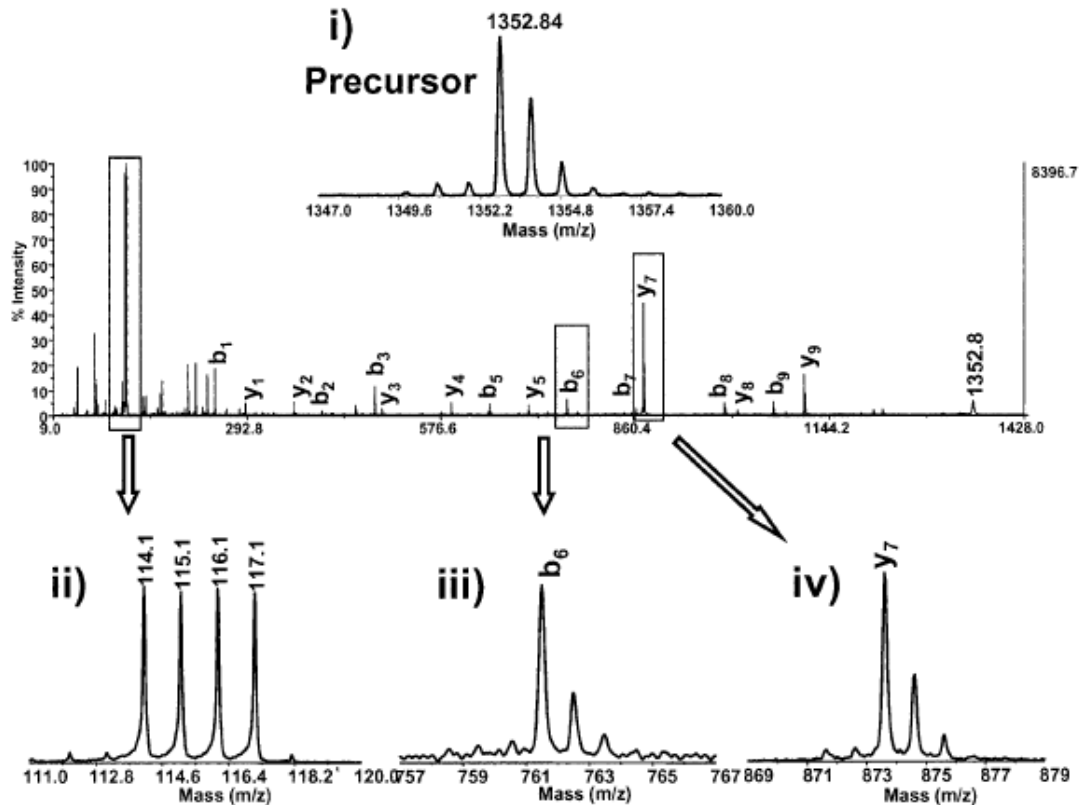


Figure 1.27: Example tandem mass spectrum of the peptide TPHPALTEAK from a protein digest mixture prepared by labelling four separate digests with each of the four iTRAQ reagents and combining the reaction mixtures in a 1:1:1:1 ratio. (i) isotopic distribution of the precursor ($[M+H]^+$, m/z 1352.84), (ii) low mass region showing signature iTRAQ ions used for quantification, (iii) isotopic distribution of the b_6 -ion, and (iv) isotopic distribution of the y_7 -ion. The peptide is labelled at both the N-terminus and C-terminus (lysine side-chain). The precursor ion and all internal fragment ions (e.g. b- and y-ions) contain all four members of the tag set, but remain isobaric.¹¹²

The group found that with both of the mutant strains, there was an increase in expression of a common set of proteins which are involved in amino acid biosynthesis and general nitrogen metabolism. However, they found that there were differences between the two mutant strains in terms of protein down-regulation: the *upf1* Δ strain showed down-regulation of proteins involved in DNA replication and RNA transcription, whereas the *xrn1* Δ strain showed down-regulation of proteins involved in the translation process (e.g. ribosomal proteins).¹¹²

An 8-plex version of the iTRAQ reagent (Figure 1.28) has also been developed (and is commercially available) where eight samples can be analysed in a single experiment.¹¹⁴

The reporter group masses are 113-119 Da and 121 Da (there is no reporter group mass at 120 Da due to potential contamination from phenylalanine immonium ions in the MS/MS

spectrum).¹¹⁴ Pierce *et al.* published a paper in 2008 detailing the 8-plex iTRAQ reagents, and demonstrated their use by comparing the activity of six oncogenes - all leukemogenic tyrosine kinases.¹¹⁴ Initially, the group used a defined mixture to show that the 8-plex reagent was capable of detecting differences in peptide levels between two samples, in line with the 4-plex strategy. They then went on to analyse the six oncogenes by several methods: iTRAQ, exon array, QTPCR and western blotting. There was good agreement between the results of the different analytical methods.

8-plex iTRAQ was also used by Ow *et al.* in 2008 for a shotgun proteomics experiment investigating enriched heterocysts from *Nostoc* sp. PCC 7120.¹¹⁵ They used the 8-plex iTRAQ strategy to simultaneously analyse four replicate phenotypical families in one experiment, rather than having to use multiple 4-plex reagent kits, which is highlighted in the paper as allowing an increase in throughput and reduction in experimental and mass spectrometry analysis time.¹¹⁵

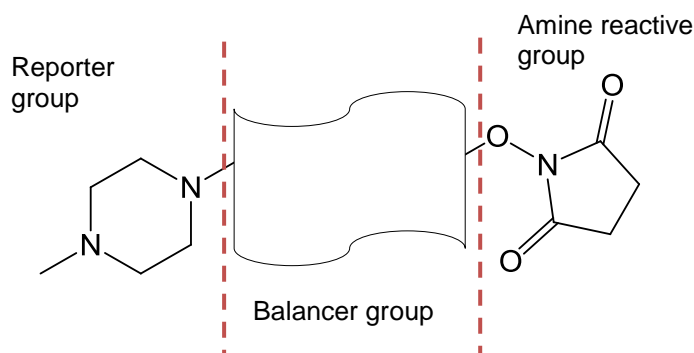


Figure 1.28: 8-plex iTRAQ label consisting of three regions: a reporter group, a balancer group and a reactive group which targets primary amines

The fact that multiple samples can be analysed at the same time is a great advantage of iTRAQ and TMT. There is also the advantage that all peptides are labelled at the N-terminus, which means that iTRAQ and TMT are less of a specific labelling strategy than ICAT, which only labels cysteine residues. This means that greater coverage of the peptide can be observed, and can also lead to several peptides being used to generate a quantification ratio for one protein, which can improve quantification accuracy.

The iTRAQ and TMT reagents are quite expensive, and it is difficult to detect PTMs and protein degradation, due to analysis occurring at the peptide level.¹⁰² An additional concern is the fact that the tags are not suitable for analysis on all mass spectrometers. For example, the signature masses for iTRAQ at 114 to 117 Da can lie out of the range of the instrument, as can happen with an ion trap instrument when operated in MS/MS mode.¹⁰² There has been much information published on ratio under-representation of fold change observed when using iTRAQ.^{116,117} Isotopic impurities, interference of ammonium ions derived from peptide fragmentation and peptide co-fragmentation due to mixed MS/MS on

fold change were evaluated by Ow *et al.* as possible reasons for underrepresentation of fold change.¹¹⁶ The problem of co-isolation with the isobaric-reagent labelled peptide ion was discussed by Christoforou and Lilley in 2011.¹¹⁸ Here they presented the strategies developed by two groups to counter this issue. Ting *et al.* almost completely eliminated ratio distortion due to co-isolation through the use of MS³.¹¹⁹ The group used 6-plex TMT reagents to model a two-proteome peptide mixture sample. Yeast peptides were labelled with all six TMT reagents, and combined to give a range of ratios between 2.5:1 and 10:1, whilst HeLa peptides were labelled with three of the TMT reagents and combined to give a 1:1:1 ratio. A variety of methods to try and reduce interference were investigated: estimation of the extent of interference based on data from full scan spectra; sample fractionation; reduction of the precursor ion isolation width; and MS³ data acquisition.¹¹⁹ They found that the best method to reduce interference was to perform an MS³ analysis, and that even though fewer peptide and protein identifications were made compared with MS² analysis, the quality of the quantification data justified this loss in analytical depth. A target ion was selected for CID MS² analysis, with the resulting data used for peptide-ion identification. The most intense MS² fragment ion was then selected for MS³ analysis to yield the quantification data in the resulting spectrum. This MS³ strategy was found to reduce interference with the TMT reporter ions when performing quantification.

Wenger *et al.* published an alternative approach to address the issue of co-isolation: 'QuantMode'.¹²⁰ Here, a variety of methods was also investigated to reduce the problem of co-isolation interference: narrowing the precursor ion selection width; data filtering after acquisition; and a real-time filtering approach to ensure suitably uncontaminated ions were analysed.¹²⁰ Ultimately, the group decided upon a method whereupon once a precursor ion is selected, it undergoes charge reduction through proton-transfer ion-ion reactions. If there were two precursor ions at m/z 500 (with different charge states (e.g. one with +2 and the other with +3)), then through proton-transfer ion-ion reactions these would become m/z 999 (now +1 charge state) and m/z 749.5 (now +2 charge state).¹²⁰ The two ions now have different m/z values, and therefore isolation of m/z 999 would result in isolation of only one of the original two precursors. The group devised the 'QuantMode' scan function in order to implement this process. Of the two approaches suggested by Ting *et al.* and Wenger *et al.*, Christoforou and Lilley summarise that the approach presented by Ting *et al.* is perhaps the most intuitive and (currently) easier of the two to implement.¹¹⁸

1.3.5. Acid-labile isotope-coded extractants

The acid-labile isotope-coded extractants (ALICE) tag was described by Qui *et al.* in 2002 (Figure 1.29) and is a tag (which reacts with cysteine-containing peptides) attached to a non-biological polymer bead.¹²¹ Capture of the peptides on the resin takes place at a neutral pH (pH 7.0-7.5), before cleaving the peptide and ALICE tag from the resin using a mild

acidolysis reaction.¹⁰⁷ The tag itself consists of a thiol-reactive group which reacts with the cysteine side chain, a linker chain which contains the chemical label element of the tag, and an acid-labile group which connects the tag to the cross-linked polymer bead.

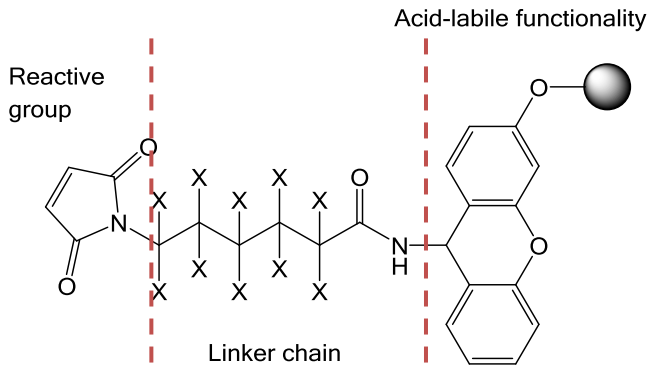


Figure 1.29: ALICE chemical structure (X=H (light), X=D (heavy)) with three regions: reactive group, linker chain and acid-labile functionality

Two aliquots of a protein mixture are digested; one is reacted with the light ALICE tag and the other with the heavy ALICE tag. The two ALICE-tagged digests are then mixed. The tagged peptides can be extensively washed, in order to remove untagged peptides, and then cleaved from the resin by mild acid-catalysed elution and subjected to LC-MS. Relative quantification is possible with the ALICE tag by calculating the relative peak heights of peptides labelled with the heavy and light tags.¹²¹ An example CID spectrum of a peptide containing three cysteine residues labelled by ALICE is shown in Figure 1.30.

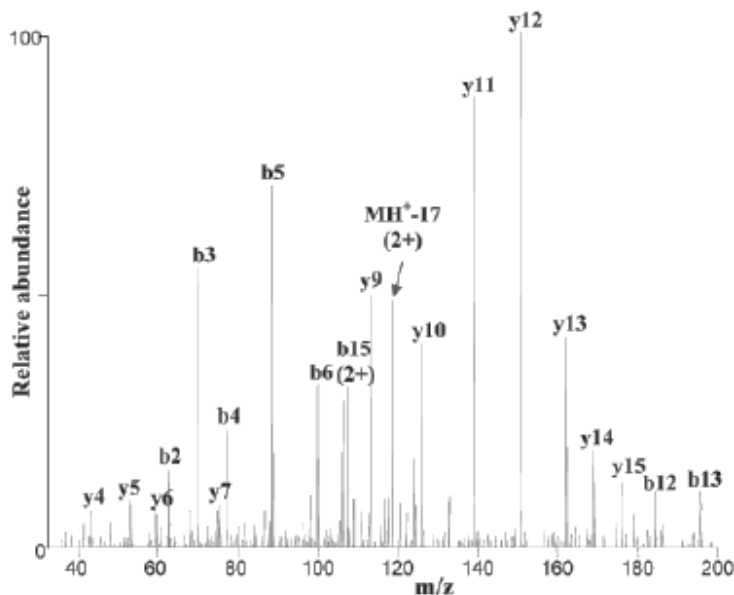


Figure 1.30: Example tandem mass spectrum of a BSA peptide with three cysteine residues (CCAADDKEACFAVEGPK) labelled with the light ALICE label¹²¹

One of the main advantages of the ALICE tag is the fact that only derivatised peptides are present in the mass spectrometric analysis. Peptides that have not been labelled are washed away prior to cleaving of the tagged peptides from the resin, which should simplify the mass spectrum obtained.

However, the cysteine-specific nature of the tag means that a small subset of peptides from the protein are analysed, which can hinder the identification of the protein. This may be the reason that there is little evidence to suggest that the ALICE labelling strategy has been adopted by the proteomics community for routine use.

1.3.6. Dimethyl labelling

Dimethyl labelling uses reductive amination to label the primary amino group at the N-terminus of the peptides and in the side chain of lysine residues. It was first proposed by Hsu *et al.* in 2003 as a global labelling strategy and is described as a reliable, cost-effective and undemanding labelling strategy.^{122,123} The labelling process increases the peptide mass by 28 Da from the unlabelled peptide when formaldehyde is used (light), and by 32 Da when deuterated formaldehyde is used (heavy) (Figure 1.31). The method is described as a relative quantification method, as the ratio between the intensities of the peaks for the light- and heavy-labelled peptide pairs can be compared. Hsu *et al.* performed dimethyl labelling on a standard protein mixture containing ovalbumin, bovine serum albumin and myoglobin to demonstrate the method was applicable for quantitative proteomics. Solutions of the three peptides were labelled with either the light or heavy formaldehyde labels and combined to give heavy/light ratios of 2.5:1 (ovalbumin), 1:1 (bovine serum albumin) and 0.25:1 (myoglobin) for LC-MS analysis. The experimental results were in good correlation with the prepared ratios.

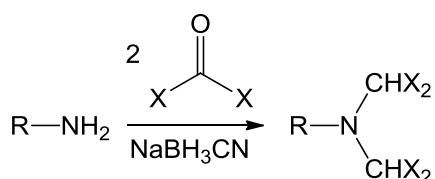


Figure 1.31: Dimethyl labelling strategy developed by Hsu *et al.* (X = H (light) or D (heavy))¹²²

In 2009, a paper by Boersema *et al.* described a multiplex dimethyl labelling strategy, which introduced another version of the dimethyl label including ¹³C and D in the formaldehyde, and D in the sodium cyanoborohydride (Figure 1.32).¹²³ They described a number of strategies which could be used for labelling: in-solution, online and on-column. In each of the strategies, the labelling was complete within 60-90 minutes.

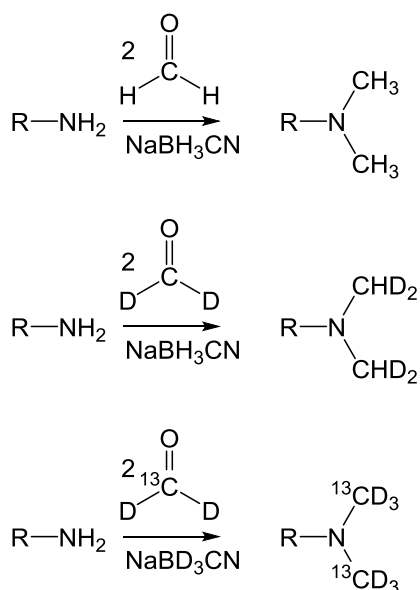


Figure 1.32: Labelling schemes for triplex isotope dimethyl labelling (R = remainder of peptide)

In 2014, Wu *et al.* published a paper which proposes a five-plex isotope dimethyl labelling system, with a mass difference of 2 Da between the different isotopic labels (Table 1.1).¹²⁴

Table 1.1: The combination of different isotopic reagents used for the five-plex dimethyl labelling system

Label	1	2	3	4	5
Formaldehyde isotope	H ₂ CO	H ₂ CO	D ₂ CO	D ₂ CO	D ₂ ¹³ CO
Cyanoborohydride isotope	NaBH ₃ CN	NaBD ₃ CN	NaBH ₃ CN	NaBD ₃ CN	NaBD ₃ CN
ΔMass (Da, one active site)	28.0313	30.0439	32.0564	34.0690	36.0757

The group demonstrated the feasibility of the five-plex strategy in proteome and phosphoproteome quantification of HeLa cell samples. They showed that high quantification accuracy could be obtained, although the quantification rates (quantified number/identified number) were lower for the five-plex labelling strategy (~78%) than the conventional dimethyl labelling strategies (>90%).¹²⁴ However, the authors contend that it is still better to run one five-plex experiment than two or three conventional experiments.

Dimethyl labelling has the advantages of being relatively quick and of using cheap reagents. The N-terminus of every peptide will theoretically be labelled, which can increase the sequence coverage¹²⁵ when compared with a labelling strategy such as ICAT which only targets cysteine-containing peptides. There is also the fact that multiple labels can be added if the peptide contains a lysine residue, which creates a larger mass difference between the

light and heavy labelled peptides, which in turn reduces the chance of isotopic envelopes overlapping in the mass spectrum.

A disadvantage to the dimethyl labelling strategy is the fact that separation of deuterated and non-deuterated forms of isotopically substituted compounds on chromatography is well-known to occur. This could be problematic when the samples are analysed by LC-MS, as a slight difference in retention times for the light formaldehyde and heavy formaldehyde labelled peptides is likely to be seen, which could affect the ability to perform quantification. Hsu *et al.* noted a negligible retention time shift when the authors first proposed the idea of dimethyl labelling; however they only looked at one labelled tryptic peptide from a BSA digest, which may not have been representative of all the peptides analysed in their experiment.¹²² Boersema *et al.* proposed that the reason for chromatographic separation of hydrogen and deuterium containing peptides may be due to the higher hydrophilicity of deuterium over hydrogen, meaning that the deuterium-labelled peptides travel through the column faster than the hydrogen-labelled peptides (if reversed phase LC is used).¹²³ Boutilier *et al.* have also studied the chromatographic behaviour of peptides following dimethylation with light and heavy formaldehyde.¹²⁶ The authors observed a statistically significant separation of dimethylated peptides, with the deuterium-labelled peptides eluting before the hydrogen-labelled peptides. However, they concluded that the separation was inconsequential in relation to peptide quantification in terms of point measurements, and that the dimethyl labelling strategy is suitable for quantitative proteomics.

1.3.7. *N,N*-Dimethyl leucine

N,N-Dimethyl leucine (DiLeu) labelling involves the labelling of peptide amine groups (i.e. the N-terminus and lysine residues). It was first proposed as a 4-plex chemical labelling strategy by Xiang *et al.* in 2010.¹²⁷ The tag consists of an amine reactive group, a balance group and a reporter group (Figure 1.33).

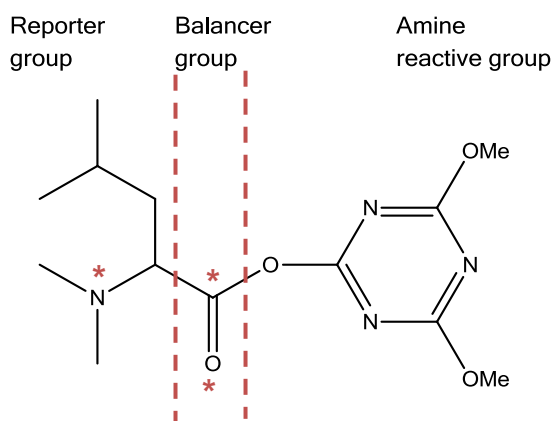


Figure 1.33: General structure of the DiLeu chemical label (* indicate labels in the 115 (smallest) label)

Quantification is based on the relative intensities of the reporter ions (m/z 115 – 118) in the tandem mass spectra. Xiang *et al.* compared the labelling efficiency and quantification capabilities of DiLeu labelling with those of the commercially available isobaric tag iTRAQ.¹²⁷ Their work showed that the DiLeu labelling strategy produces a comparable performance to iTRAQ.

The advantages of the labelling strategy are the fact that the labels are synthetically simple, there is high labelling efficiency and improved fragmentation efficiency.¹²⁷ Another big advantage of the DiLeu label over the commercially available isobaric tags iTRAQ and TMT is the fact that it is much cheaper to perform a quantification experiment with DiLeu than with iTRAQ or TMT.¹²⁷ The group have also developed 8-plex¹²⁸ and 12-plex¹²⁹ versions of the DiLeu labels.

1.3.8. ¹⁸O oxygen incorporation

¹⁸O can be incorporated into peptides at the carboxyl terminus by enzymatic digestion of the protein in the presence of H₂¹⁸O (Figure 1.34). One protein mixture is digested in the presence of normal ¹⁶O-water, and another is digested in the presence of ¹⁸O-water. The two samples are then combined for mass spectrometric analysis. The addition of the ¹⁸O leads to an increase of either 2 or 4 Da (one or two ¹⁸O).¹⁰²

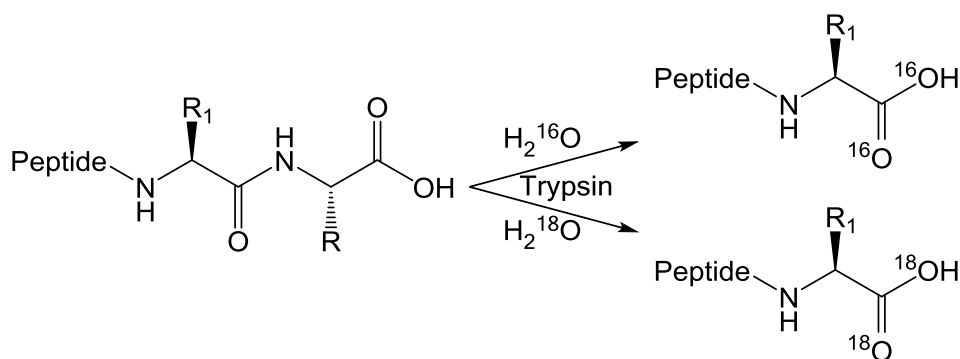


Figure 1.34: Reaction scheme for the incorporation of ^{18}O during proteolytic digestion using labelled water

The small mass difference can lead to a quantification issue as the mass of the peptides increases: the larger the mass of the peptide, the more complicated the isotopic distribution becomes, and so an overlap can occur between the isotopic envelopes of the light- and heavy-labelled peptides. An algorithm can be used to identify the proportions of the peaks which belong to the light- and heavy-labelled peptides when overlapping isotopic distributions are observed in the mass spectrum. The main advantage of this technique is that the two versions of the peptide do not suffer any of the effects previously mentioned upon chromatographic separation, such as retention time shifts. The labelling procedure itself is fairly straightforward, which also promotes its use as a labelling strategy.

However, there is the possibility of incomplete labelling, or back-exchange of ^{16}O and ^{18}O , which affect the quantification accuracy.^{102,107} Additionally, H_2^{18}O is an expensive reagent.

1.3.9. Bromine incorporation

A protein and peptide labelling strategy which is a qualitative, rather than a quantitative one, is the incorporation of bromine into proteins and peptides. The two isotopes of bromine (^{79}Br and ^{81}Br) can be simultaneously incorporated into tags for protein analysis, and proves useful due to its characteristic isotopic signature when analysed by mass spectrometry. The incorporation of bromine allows easy identification of bromine-containing species in a mass spectrum that can have a very high background from other components. The incorporation of a chemical group or tag can also help improve mass spectrometric response of peptides, especially in the presence of impurities that may be present in some 'real' samples. Finally, a tag can also help direct fragmentation on CID, or facilitate interpretation of fragmentation data from tagged peptides. Consequently, it has been included in this overview of labelling strategies, and also because this work is a very relevant precursor to the new work presented in this thesis.⁹⁰

Hudson *et al.* published a paper in 2012 detailing the preparation and use of two bromine-containing solid-support N-terminal tagging strategies (Figure 1.35).⁹⁰ Unlike most solid support labelling strategies, the authors looked to perform labelling and cleavage of the tagged peptide in a single step. This simplifies the workflow for peptide tagging, which is an advantage of the technique, as it reduces the number of stages required in order to obtain tagged peptide, and therefore possible sample loss etc.. The tagging construct consists of a divinylbenzene cross-linked polystyrene solid-support, a linker and the analytical tag. The main advantage of this technique is the fact that unreacted tag and by-products bound to the resin are removed by filtration, resulting in simple preparation for analysis by mass spectrometry.⁹⁰ However, there was an issue with the sensitivity of the approach, as the lower limits of detection (approximately 4 μg of single peptide, in this instance leucine enkephalin)⁹⁰ are well above the amount of each peptide which would typically be available in a proteomic digest, which could be expected to contain nanograms to milligrams total protein.³

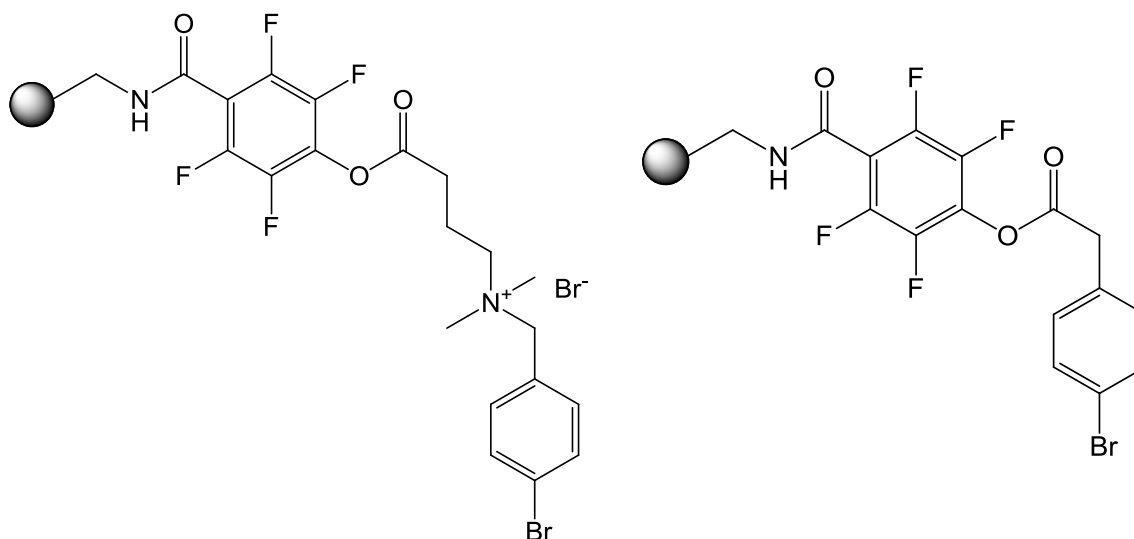


Figure 1.35: Bromine-containing solid-support N-terminal tags

In 2012, Song and Kim described the N-terminal bromoacetylation of peptides, and the subsequent sequencing by MALDI-MS, by reacting the N-terminus with bromoacetic anhydride.¹³⁰ The authors were able to selectively label the N-terminus over the lysine side chains by controlling the pH of the reaction. They found that the b-ions contained the bromine tag, but not the y-ions, which facilitated de novo sequencing of peptides (Figure 1.36).

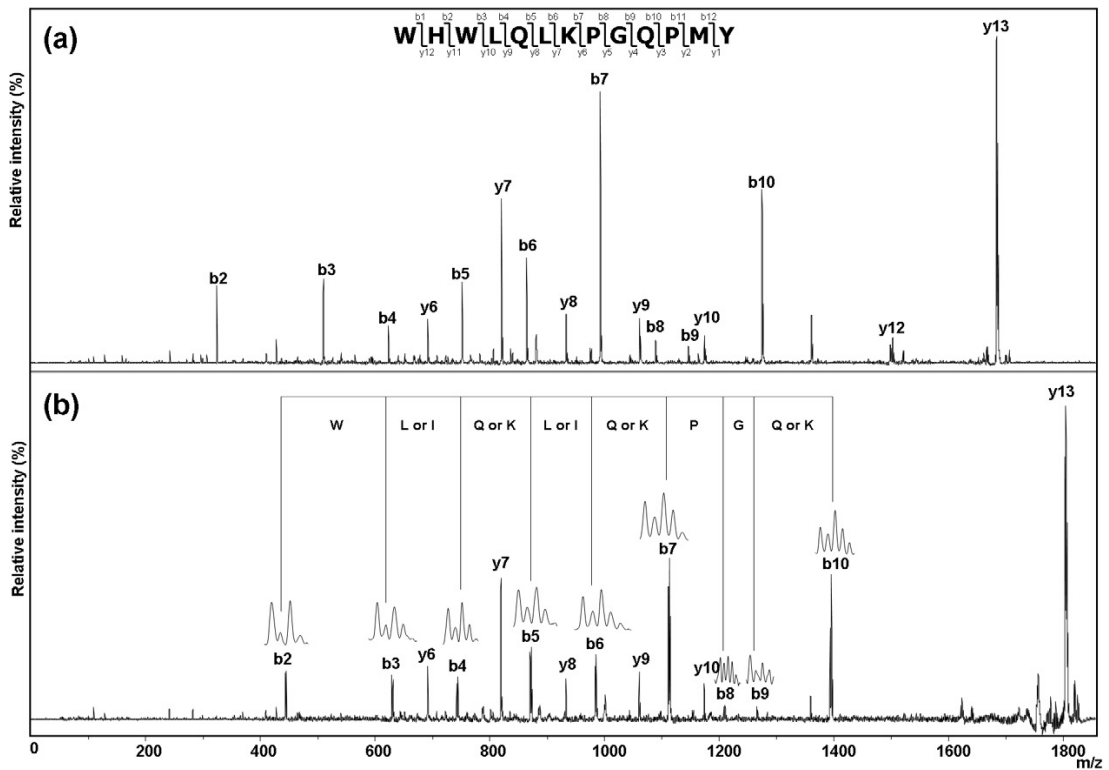


Figure 1.36: Tandem MALDI mass spectra from the peptide WHWLQLKPGQPMY (a) before bromoacetylation and (b) after bromoacetylation. The Br-tag was only observed in the b-ion series.¹³⁰

1.3.10. Summary

There are many different labelling strategies possible for proteomics work, most, but not all, of which have been designed to enable relative or absolute quantification. Each has its own advantages and disadvantages, which have been summarised in Table 1.2. The commercially available chemical labels, such as iTRAQ and TMT, are well established for quantitative proteomics analyses. Labelling strategies which are not commercialised, such as dimethyl labelling, are discussed as part of the development of the work in this thesis (see Chapter 5 for further discussion).

Table 1.2: Summary of labelling methods, developed from the table by Boersema *et al.*¹²³

Labelling strategy	Method	Strengths	Weaknesses
SILAC	Incorporation of chemically labelled amino acids	<ul style="list-style-type: none"> • Introduction of isotope label at organism level • No limits to amount of sample labelled • Quantify at MS level 	<ul style="list-style-type: none"> • Not applicable to human samples • Expensive to culture and reach full incorporation of isotopic labels
Dimethyl labelling	N-terminus and lysine residues	<ul style="list-style-type: none"> • Cheap reagents • Quick labelling 	<ul style="list-style-type: none"> • Introduction of isotopes at peptide level • Small isotope effect in LC separation
TMT	N-terminus	<ul style="list-style-type: none"> • Able to perform multiplex experiments • No chromatographic separation 	<ul style="list-style-type: none"> • Introduction of isotope label at peptide level • Quantification at MS/MS level • Potential co-isolation of precursor ions
iTRAQ	N-terminus	<ul style="list-style-type: none"> • Labelling up to eight samples at a time • No chromatographic separation • No increased complexity at MS level • Applicable to any sample 	<ul style="list-style-type: none"> • Introduction of isotope label at peptide level • Not very chemically stable, expensive reagents • Quantification occurs at MS/MS level • Reporter ions at low <i>m/z</i> values
DiLeu	N-terminus and lysine residues	<ul style="list-style-type: none"> • Cheap reagents • Enhanced fragmentation • Quick labelling 	<ul style="list-style-type: none"> • Introduction of isotope label at peptide level • Quantification occurs at MS/MS level • Small isotope effect in LC separation
ICAT	Cysteine groups	<ul style="list-style-type: none"> • Determines relative abundances using isotope pairs • Peptides isolated and purified before analysis due to biotin region 	<ul style="list-style-type: none"> • Introduction of isotope label at peptide level • Isotope effect in LC separation due to eight deuterium atoms • Limited to cysteine-containing peptides
ALICE	Cysteine groups	<ul style="list-style-type: none"> • Peptides isolated and purified before analysis 	<ul style="list-style-type: none"> • Limited to cysteine-containing peptides
¹⁸ O incorporation through protein digestion	C-terminal	<ul style="list-style-type: none"> • No chromatographic separation • Relatively cheap reagents as need small amounts of H₂¹⁸O 	<ul style="list-style-type: none"> • Mass difference of +4 Da can lead to overlapping signals

1.4. Self-assembled monolayers

In 1946, a paper published by Zisman, Pickett and Bigelow described the formation of a film on a clean metal surface by adsorption.¹³¹ However, the importance of this discovery was not recognised for many years. In the 1980s, research into self-assembled monolayers (SAMs) accelerated, and discoveries in the next 15 years resulted in the use of many SAMs in varying applications.¹³² The following section is thus not a comprehensive review of SAM chemistry, but instead focuses on the biological and analytical applications of SAMs

In its simplest embodiment, a self-assembled monolayer is an organised layer of molecules, and arises by the spontaneous arrangement of molecules on a surface by adsorption.¹³³ The composition of the molecules used means that SAMs have a unique combination of physical and chemical properties. Each molecule in a SAM contains a head group, a spacer group and a terminal group (Figure 1.37).

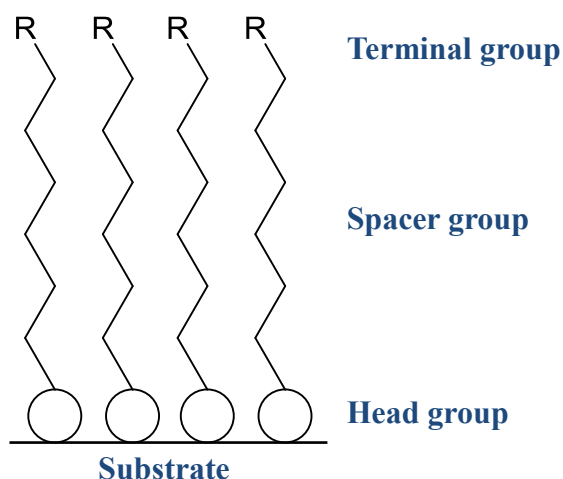


Figure 1.37: Generic structure of a SAM

The head group has a high affinity for the solid surface substrate (e.g. sulfur has a strong affinity for gold¹³⁴). The initial part of the self-assembly is the rapid reaction between the head group and the substrate, forming a bond between the two. The organisation of the terminal groups then follows over a slower period of time. This self-ordering of the terminal functional groups creates a high degree of organisation in SAMs. The degree of order within the SAM is due to factors such as electrostatic and dipole-dipole interactions, and affinity of the head group for the substrate.¹³³ The terminal group extends beyond the spacer groups, and therefore plays a large part in how the SAM reacts with the external environment. The terminal groups do not need to be solvated in order to react.¹³³ This is unusual, and means that the terminal groups can be reacted either with a solution or a gas-phase reagent.¹³³ The first reported example of a gas-phase reagent reacting with a SAM was published by Sun *et al.* in 1991, where they reported the reaction between surface-confined 4-

hydroxythiophenol and 4-aminothiophenol with gaseous dimethyloctylchlorosilane (Figure 1.38).¹³⁵

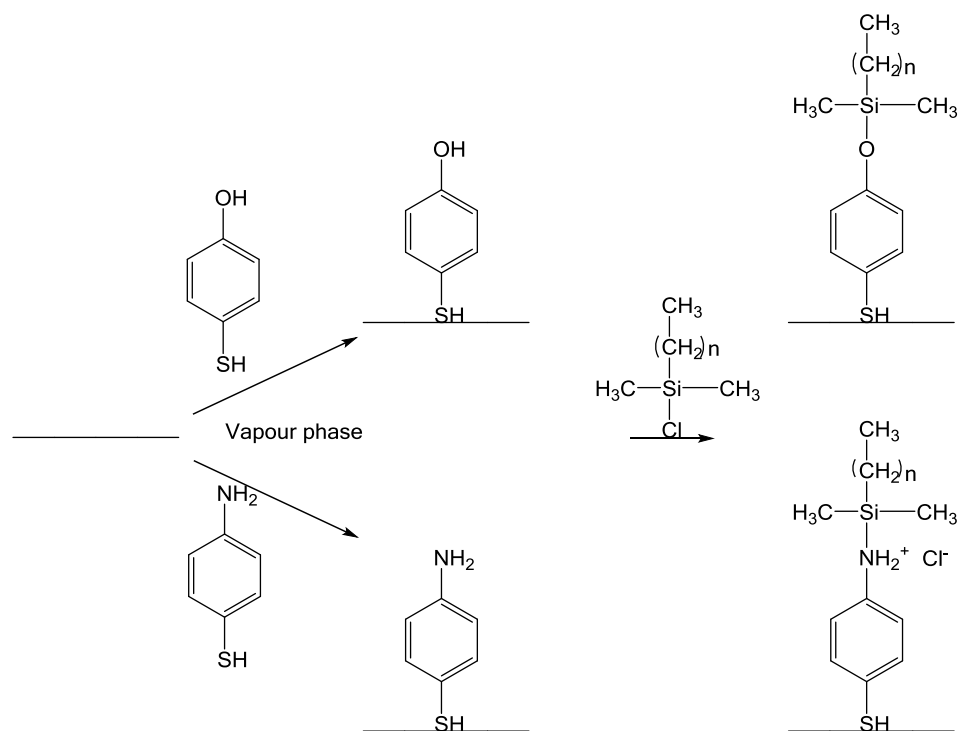


Figure 1.38: Gas-phase reagent reacting with a SAM¹³⁵

There are many advantages to using SAMs; they are easy to prepare; they can be used as building blocks for a more complex structure (usually confined to the solid surface); and the surface properties can be changed by modification of the molecular structure.¹³⁶ The most commonly used terminal groups for surface reactions are carboxyl, amino or hydroxyl groups.¹³³

Defects and pinholes can arise within a SAM structure, which have a detrimental effect on the SAM. These defects can arise from SAM formation on a surface with defects, the length of the spacer group and the conditions used to prepare the SAM.¹³⁷ It is also possible for reagents or by-products to adhere to the substrate, causing defects.

The field of self-assembled monolayers is vast, with many different applications. These include electrochemistry¹³⁸, DNA binding^{138,139} and chemical sensing^{139,140}. Due to the nature of the work discussed in this thesis, the applications focus for this introduction to self-assembled monolayers is on SAMs in relation to biological applications.

SAMs have been used for many years to provide a model system for studying protein interaction with organic surfaces. One such example was published in 1991 by Prime and Whitesides, in which they used combinations of five alkanethiols, and monitored the adsorption of five proteins (ribonuclease A, pyruvate kinase, fibrinogen, lysozyme and

chymotrypsinogen) on prepared SAMs.¹⁴¹ The five structures used were a mixture of hydrophilic and hydrophobic alkanethiols, thus allowing the formation of mixed SAMs. They found that protein adsorption was lowest for SAMs which contained high concentrations of the hydrophilic alkanethiols.

SAMs have been used to capture peptides and can be engineered to prevent the non-specific adsorption of peptides or proteins on surfaces.¹⁴² This is through the inclusion of polyethylene glycol (or oligoethylene glycol) in the spacer groups. There are several theories which have been proposed to explain why polyethylene glycol groups prevent protein adsorption. Jeon *et al.* theorised that as a protein approaches the polyethylene glycol groups by diffusion, van der Waals attraction between the polyethylene glycol groups and protein dominates.^{143,144} As the protein further approaches the polyethylene glycol layer, the polyethylene glycol chains compress, which increases the surface density of the polyethylene glycol layer leading to repulsion of the protein.¹⁴⁴ The more polyethylene glycol groups there are in the polymer (at least three), the greater the ability to resist protein adsorption.¹⁴⁵ In 1991, Jeon *et al.* studied the *in silico* theoretical experimental interaction between a model polyethylene glycol with a hypothetical 'protein' of infinite size (with the SAM substrate being in water).^{143,144} The group theorised that as a protein approaches the terminal polyethylene glycol chains, the protein is prevented from reaching the substrate surface due to a combination of steric repulsion, van der Waals forces and hydrophobic interaction. The polyethylene glycol chains compress (proposed to be due to the approaching protein causing the chains to start to flatten) causing a loss of entropy due to the limited configurations accessible to polyethylene glycol. This loss of entropy means that compression is disfavoured and steric repulsion occurs, so hindering protein adsorption. Ultimately, the theory summarises that surface-bound polyethylene glycol groups prevent protein adsorption due to the high conformational freedom of the chains and the entropy of the solvated PEG chain at the surface. The group summarised that the higher the surface density and the longer the chain length of the polyethylene glycol groups, the better resistance to protein adsorption was observed.¹⁴⁴ These findings have been corroborated by others, such as Michel *et al.* who published a paper in 2005 detailing the influence of polyethylene glycol architecture on protein adsorption and conformation.¹⁴⁶

Whitesides *et al.* surmised that there is a tightly bound layer of water at the interface between the SAM and the surrounding solution.^{147,148} Functionalities which can act as hydrogen bond acceptors (such as polyethylene glycol) are efficient at resisting protein adsorption, which is thought to be due to their ability to bind water molecules.¹⁴⁸ In contact with water, ethylene oxide chains are self-avoiding.¹⁴⁹ Prime and Whitesides concluded that any chain length for the polyethylene glycol groups (even as small as one unit) is enough to prevent protein adsorption, but that the longer the chain, the more effectively the surface can be covered, and therefore the less dense the polyethylene glycol chains need to be.¹⁴⁷

The biological applications of SAMs include the use of SAMs for enzyme-substrate reactions^{150,151}, carbohydrate chemistry^{152,153} and biomolecule immobilisation¹⁵⁴.

There have many been publications which have looked at enzyme-substrate reactions, with the incorporation of SAMs. Halling *et al.* reviewed the understanding of enzyme action on immobilised substrates (although there are also examples of enzymes being immobilised on a solid phase, and the substrate being in solution).¹⁵¹ Substrates are attached to a solid phase, meaning that the enzyme must migrate to the substrate sites. The paper discusses an increasing interest in automated synthesis and screening protocols, solid supported chemistry (with enzymes potentially being used for selected steps in a solid-phase synthesis) and biochemistry as potential reasons for the increased interest in studying these types of systems.

In the most cited example of carbohydrate-based SAM technology, Houseman and Mrksich detailed a chemical strategy for preparing carbohydrate arrays and the use of these in subsequent substrate-protein interaction studies.¹⁵² The authors prepared chips containing carbohydrate arrays, which were then used to study the interaction of the carbohydrates with proteins and enzymes. The arrays were incubated with bovine β -1,4-galactosyltransferase. The substrates were then incubated with rhodamine-labelled lectins from either *E. cristagalli* or *B. simplicifolia* l. Fluorescence imaging of the arrays revealed substrate specificity of the enzyme, and demonstrated the use of carbohydrate arrays for characterising enzymatic activities towards carbohydrate substrates. The advantages of this technology were described as the ability to immobilise carbohydrates on a chip in order to participate in binding interactions with proteins and enzymes, the prevention of non-specific adsorption of proteins on the SAM substrate, and the high yields and excellent selectivity of the arrays towards immobilisation.¹⁵² Houseman *et al.* published a paper in 2003 detailing the use of maleimide-functionalised SAMs for the preparation of peptide and carbohydrate biochips.¹⁵³ In 2012, Tantakitti *et al.* published a paper in which the authors created a variety of SAMs featuring carbohydrate thiols and oligoethylene glycol thiols.¹⁵⁵ Pure and mixed SAMs were formed, with the resulting SAMs monitored over a period of time. The authors concluded that the components of the mixed SAM segregated into increasingly pure and ordered domains over time.

Wagner *et al.* published a paper which looked at the covalent immobilisation of native biomolecules on SAMs, with the purpose of capturing amino group-containing biomolecules on surfaces via functionalised SAMs.¹⁵⁴ A variety of molecules including amino acids and model proteins were captured on the SAM. The group studied the resulting products by scanning tunnelling microscopy (STM) and atomic force microscopy (AFM), and found that SAMs on gold surfaces could provide tools in a number of nanotechnological applications (e.g. biosensing).

Houseman *et al.* published a paper detailing the use of peptide chips for the quantitative evaluation of protein kinase activity.¹⁵⁶ A kinase substrate (AcIYGEFKKKC-NH₂) was immobilised on an alkanethiolate-derived SAM, to reduce the need for blocking procedures used for peptide and protein chip assays. Phosphorylation of the peptide by *c*-Src kinase was monitored using phosphorimaging. SAMs have also been used as a means of immobilising antimicrobial peptides for study.^{157,158} Humblot *et al.* published a paper detailing the immobilisation of antimicrobial peptide amines on activated carboxylic acid groups on the SAM, with the aim to demonstrate antimicrobial peptides as an alternative antibacterial coating to conventional antibiotics.¹⁵⁸ The authors found that the inclusion of the antibacterial peptides on the SAM reduced the adhesion of three Gram-positive bacteria at the surface, whilst the peptides were also still able to kill over 50% of the bacterial cells.

Gold is often used as a surface for SAM formation due to its limited reactivity, but high affinity for sulfur. Molecules which have a sulfur head group, such as alkanethiols, dialkyl sulfides or alkyl xanthates, have been used for the formation of SAMs on gold surfaces.¹³² One of the first such uses of gold was reported in 1983 by Nuzzo and Allara.¹⁵⁹ The authors experimented with preparing, and structurally characterising, orientated monolayers of disulfide-terminated polyfunctional organic molecules with a variety of molecular structures on to a zerovalent gold substrate.¹⁵⁹

SAMs formed from sulfur head groups can be formed using either thiols or disulfides. In 1989, Bain *et al.* published a paper detailing their study of the formation of SAMs on gold using thiols and disulfides.¹⁶⁰ X-ray photoelectron spectroscopy (XPS) was used to determine the composition of the SAM formed. The XPS spectra of the sulfur bound to the surface suggested that both thiol and disulfide species yielded the same species on the surface, most probably in the form of a gold thiolate.¹⁶⁰ The authors found that, although there was a preference for the thiol to adsorb to the surface of the gold over the disulfide when thiols and disulfides are used simultaneously to form a SAM, there was little difference between the SAMs formed from pure thiols or disulfides. The preference for thiol adsorption over disulfide adsorption was summarised to be due to a kinetic phenomenon.¹⁶⁰ The conclusion that there is little difference in the SAMs formed from pure thiols and disulfides was due to the fact that the same species were formed on the substrate surface (observed by XPS), and the fact the contact angles recorded for the SAMs was largely independent of the precursor species used (i.e. a thiol or a disulfide). In 1994, Biebuyck *et al.* released a further paper which looked at formation of a SAM using solutions containing dialkyl disulfides and alkanethiols.¹⁶¹ It was found that the SAMs formed were indistinguishable in many ways from the two starting materials. However, they found that a SAM made only from disulfides had a lower contact angle than a SAM formed from alkanethiols. This meant that, due to the hydrophobic nature of the SAM, there may have been incomplete SAM formation when the disulfide was used, which would increase the hydrophilicity of the surface, and therefore

reduce the contact angle. When the surface with the dialkyl disulfide SAM formed on it was submerged in a solution of only alkanethiol, and the contact angle retested, the contact angle was indistinguishable from that for the SAM formed from the alkanethiol only. This suggests that a SAM formed from disulfides is not as tightly packed as a SAM formed from thiols.

In 2009, a paper was published investigating the surface processes involved in the generation of ions from MALDI-MS analysis of alkanethiolates on a gold surface.¹⁶² It has long been observed that intact, characteristic ions (usually disulfides) are seen in the resulting mass spectra.¹⁶² In this paper, it was proposed that the matrix solution dissolves alkanethiols from the SAM, which form disulfide species in solution. These then co-crystallise with matrix molecules, and are detected when analysed by MALDI-MS. Previous to the paper, it had been found that direct desorption of alkanethiolate SAM molecules from the gold surface led to formation of disulfides in the mass spectrometer vacuum.¹⁶² The topic of gold-immobilised SAMs is discussed in more detail in Chapters 2 and 3.

1.5. Aims of this thesis

Most proteomic chemical labelling strategies, such as iTRAQ¹¹² and TMT¹⁰⁸, are carried out in solution. It is also possible to label peptides using a labelled solid-support,^{121,163,164} although this approach has been less well explored. This is likely to be due to the numerous steps required to obtain labelled peptides: immobilisation of a label on the solid phase, capture and subsequent cleavage of the derivatised peptides from the solid phase. Work by Hudson *et al.* showed a labelling strategy using a solid phase, whereby the peptide cleaved the label from the solid phase during the labelling step.⁹⁰ The overall aim of the work in this thesis was to work towards developing a relative quantification proteomics technique for application to single protein digests; through the miniaturisation and localisation of a bottom-up proteomics workflow, so that everything except protein digestion takes place on a gold-coated MALDI chip. The gold-coated MALDI chips can be reused for a number of experiments, after cleaning. The strategy aims to provide simultaneous peptide capture and labelling with minimal sample manipulation and *in situ* mass spectrometric analysis.

This project aimed to investigate a new chemical labelling strategy, initially based on work by Hudson *et al.* using a bromine-containing chemical label,⁹⁰ and work by Flitsch and co-workers.^{150,165,166} The Flitsch group work involved the synthesis of carbohydrates and peptides on self-assembled monolayers and monitoring by MALDI-MS analysis.

The work may be divided into two strands which are brought together at the very end of the thesis: work involving the use of SAMs and work using in-solution dimethyl labelling.

The SAM-based work can be broken down further into peptide capture on SAMs on gold-coated MALDI chips and the subsequent removal of SAMs from the gold-coated MALDI chips. The aim of the peptide capture work was to develop a generic, flexible labelling strategy to capture, chemically derivatise and analyse protein digests directly on gold-coated MALDI chips.

In-solution dimethyl labelling was investigated to establish whether quantitative results could be obtained, with the ultimate aim being to attempt to adapt the dimethyl label for incorporation into the SAM technology.

Chapter 2 describes the capture of a variety of peptides on a thiol-derived SAM, with and without the use of a chemical tag, 4-bromophenylalanine.

Chapter 3 describes the capture of a variety of peptides, 4-bromophenylalanine and phenylalanine on a disulfide-derived SAM. The aim of the work in Chapter 3 was to improve the peptide capture efficiency from that seen in Chapter 2. Additionally, an investigation was undertaken to determine whether the presence of Phe at the N-terminus of the peptide reacting with the SAM reduced capture efficiency.

Chapter 4 describes the use of an alternative technique (plasma treatment) for cleaning gold-coated MALDI chips. The aim of this work was to determine whether treating the wells with plasma effluent was an alternative, less corrosive technique which could be used for cleaning the gold-coated MALDI chips.

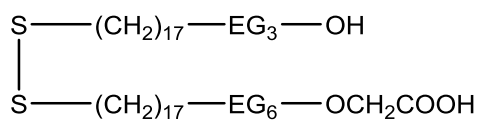
Chapter 5 describes the application of Boersema's in-solution dimethyl labelling strategy¹²³ to a set of standard solutions for validation and then it was used to analyse a set of muscle protein samples.

Chapter 2: Thiol- derived SAM

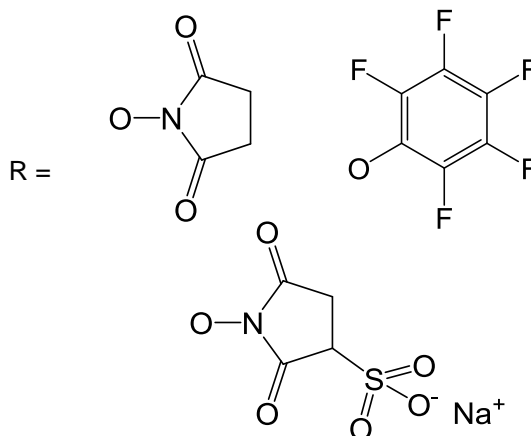
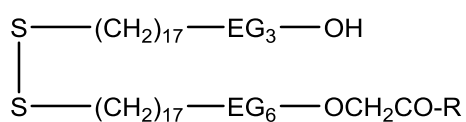
Throughout Chapter 2, the following terminology will be used to describe the SAM at the different stages of the method.

Terminology**Structure**

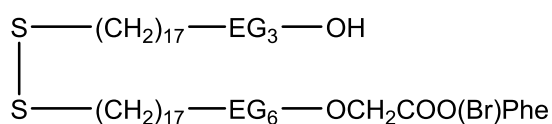
SAM



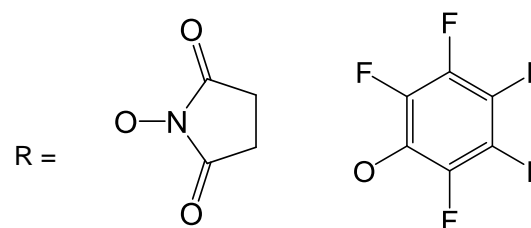
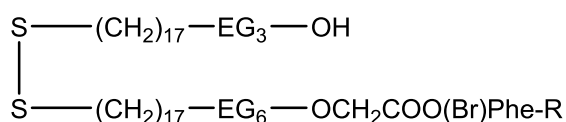
Activated SAM



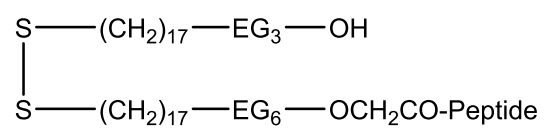
4-BrPhe-tagged SAM



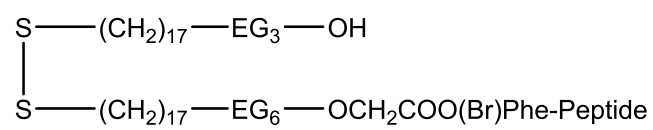
Activated 4-BrPhe-tagged SAM



Peptide captured on the
activated SAM



Peptide captured on the
activated 4-BrPhe-tagged SAM



2.1. Introduction

The capture of peptides on a gold surface for matrix-assisted laser desorption/ionisation mass spectrometry (MALDI-MS) analysis is being investigated with the potential specific intention of development for application to analysis of single protein digests, for example collagen in species identification studies.¹⁶⁷ The process involved creating a SAM on the surface of a gold-coated MALDI chip (Figure 2.1). This gold-immobilised SAM was modified to introduce a chemical tag, on which peptides were captured, and then desorbed and analysed when the gold-coated MALDI chip was placed in the MALDI source.

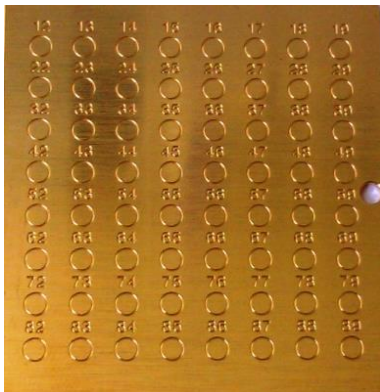


Figure 2.1: 64 well gold-coated MALDI-MS chip

The Flitsch group has investigated the use of SAMs on gold-coated MALDI chips for the SPOT synthesis (synthesis of large numbers of peptides on membrane sheets) of peptide libraries¹⁶⁶ and their evaluation as enzyme substrates.^{165,168} In that work, the thiol-derived SAM consisted of two components which terminated in either an alcohol group or a carboxylic acid group ($\text{HS}(\text{CH}_2)_{17}(\text{OCH}_2\text{CH}_2)_3\text{OH}$ or $\text{HS}(\text{CH}_2)_{17}(\text{OCH}_2\text{CH}_2)_6\text{OCH}_2\text{COOH}$). Their SPOT synthesis of peptides on the surface of a gold-coated MALDI chip is summarised in Figure 2.2. Having formed the SAM on the gold-coated MALDI chip, the carboxylic acid groups were activated using *N*-hydroxysuccinimide (NHS) prior to reaction with *N*-Fmoc diaminobutane. The subsequent removal of the Fmoc protecting group revealed the amine functionalised surface, which could be further reacted with Fmoc-protected amino acids to build a peptide on the surface of the SAM.

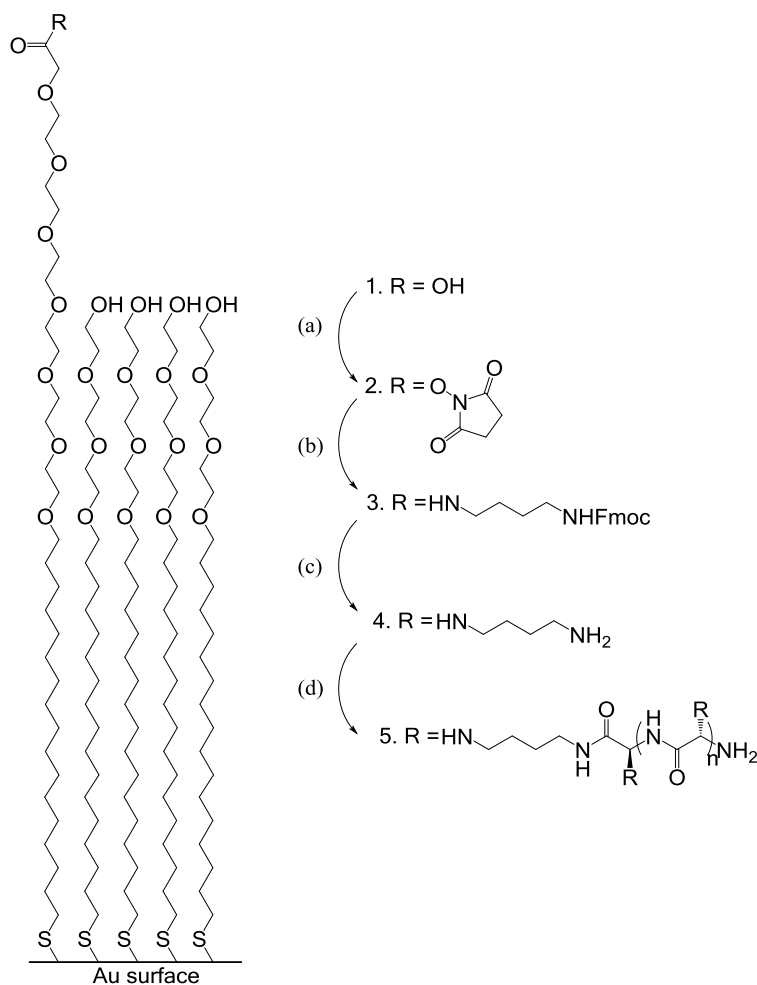


Figure 2.2: SPOT synthesis of a peptide library using Fmoc-protected amino acids.

Reagents and conditions: (a) NHS/EDC in DMF (1 h, ambient temperature); (b) H₂N(CH₂)₄NHFmoc in DMF (16 h, ambient temperature); (c) 20% piperidine in DMF (10 min, ambient temperature); (d) (i) SPOT synthesis (Fmoc-amino acid, PyBOP, DIPEA in DMF then 20% piperidine in DMF) (ii) 50% TFA, DCM (side-chain deprotection if required)

The success of the SAM formation, and each subsequent reaction, was monitored by MALDI-MS. The advantage of the on-chip synthesis is that the product does not need to be removed from the surface of the gold-coated MALDI chip prior to analysis, as the MALDI laser cleaves the immobilised material from the gold surface. The group synthesised a peptide library containing 56 tripeptides (all built on the surface of the SAM), which they used to probe the specificity of the protease thermolysin (a protease with a preference for cleaving at the N-terminus of hydrophobic residues). The gold-coated MALDI chip was incubated overnight in a solution of the enzyme, to assess whether there was a preference for a particular type of peptide, and then analysed by MALDI-MS. For some of the tripeptides, the terminal amino acid residue was cleaved by thermolysin (e.g. GFS, which was cleaved to FS). For other tripeptides, a dipeptide was cleaved by the thermolysin (e.g.

SAL was cleaved to L). Other tripeptides were unaffected by the presence of the thermolysin (e.g. GPS). The results showed the enzyme has a preference for hydrophobic and/or aromatic residues. The major benefits of the work were described as the *in situ* mass spectrometric analysis, the simple workflow with reduced sample manipulation, the possibility of multiplexing due to the 64 wells of the gold-coated MALDI chip and the fact that the gold-coated MALDI chip can be reused until the gold coating deteriorates.

This SAM-based technology could have many applications, as has been initially proven by the Flitsch group who have demonstrated applications in addition to SPOT synthesis.^{165,168} One such alternative application would be to adapt the SAM, and incorporate a chemical tag for quantitative or qualitative proteomics work. The use of chemical tags has long been established in proteomics work, and there are many variations which can be used, including labels for both quantitative and qualitative work. Collaborative work in the Flitsch and Routledge/Thomas-Oates groups identified bromine as a useful chemical label for incorporation into a tag for this application (theses^{169,170} and unpublished observations^{i,ii}).

ⁱ S. Hudson, University of York

ⁱⁱ R. Castangia, Manchester Interdisciplinary Biocentre, The University of Manchester

2.1.1. Aims and overview

The aim of the work presented in this chapter was to capture a variety of peptides on the Flitsch group's thiol-derived SAM, using an immobilised chemical tag to enable easy mass spectrometric identification of peaks associated with peptide captured on the SAM. A SAM was formed on the gold surface, with the carboxylic acid groups subsequently being activated for reaction with the chemical tag, 4-bromophenylalanine (4-BrPhe). The carboxylic acid groups then present on the tagged SAM were activated prior to reaction of the tagged SAM with a peptide (Figure 2.3). The activating agents NHS, 2,3,4,5,6-pentafluorophenol (PFP) and *N*-hydroxysulfosuccinimide (NHSS) have been tested for activating the carboxylic acid group.

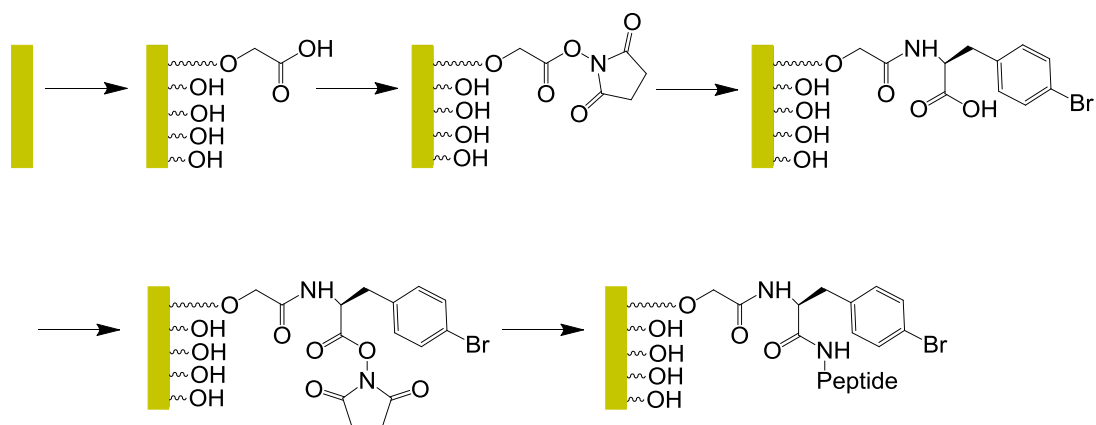


Figure 2.3: Schematic representation of the capture of peptides on 4-bromophenylalanine tagged thiol-derived SAM

An investigation into SAM activation inefficiency (and subsequent poor coupling) was undertaken using light and heavy labelled buffers.

2.2. Formation of a thiol-derived SAM

A gold-coated MALDI chip was cleaned with piranha solution (3:1 v/v conc. H₂SO₄:35% H₂O₂) prior to a SAM being formed on its surface. The two SAM components are alkane thiols containing PEG spacers and either a carboxylic acid (HS(CH₂)₁₇(OCH₂CH₂)₆OCH₂COOH) or alcohol (HS(CH₂)₁₇(OCH₂CH₂)₃OH) as their head group. The SAM components were each dissolved in DMSO at a concentration of 0.1 mg/mL, and a 1:4 molar ratio of carboxyl:alcohol terminated groups was created from these stock solutions.

Collaborators investigated the optimum ratio for the carboxylic acid groups to alcohol groups for capture of 4-BrPhe and Leu-Enk on the surface of the SAM.¹⁷⁰ A range of ratios was investigated ranging from a SAM being comprised solely of the carboxylic acid component, to a SAM being comprised of the carboxylic acid groups diluted 1:10 with the alcohol groups. For the 4-BrPhe coupling to the surface of the SAM, it was found that for a SAM composed solely of the carboxylic acid component, 4-BrPhe coupling was less effective than when the SAM was composed of both the carboxylic acid and alcohol groups. The optimum ratio was found to be between 1:2 and 1:4 (carboxylic acid groups:alcohol groups). Leu-Enk capture on SAM and 4-BrPhe-tagged SAM was also investigated, using the same SAM component ratios as for the 4-BrPhe coupling. In agreement with the results obtained for 4-BrPhe coupling to the SAM, the optimum ratio for the carboxylic acid groups and alcohol groups was found to be between 1:2 and 1:4. Capture of Leu-Enk on the 4-BrPhe-tagged SAM was found to be improved over the capture of Leu-Enk on the SAM. This was thought to be due to the hydrophobicity of the 4-BrPhe and the Leu-Enk possibly leading to an association of the 4-BrPhe and Leu-Enk in solution, increasing the likelihood of a reaction.

Based on the SAM ratio data, there is an indication that an element of steric hindrance occurs for both the 4-BrPhe and the Leu-Enk. The results show the importance of having a mixed SAM, containing both the carboxylic acid groups and the alcohol groups. The presence of the alcohol groups, creating the mixed SAM, increased the percentage of 4-BrPhe coupled to the SAM compared with starting material when analysed mass spectrometrically. Using this information, it is less likely that with the 1:4 ratio of carboxylic acid groups:alcohol groups, the coupling of one 4-BrPhe to the SAM would inhibit the coupling of another 4-BrPhe to the SAM.

Aliquots (0.4 µL) of the 1:4 molar ratio of carboxyl:alcohol terminated groups solution were spotted onto the wells on the gold-coated MALDI chip and the SAM was left to assemble overnight at ambient temperature (Figure 2.4). The gold-coated MALDI chip was washed with acetonitrile and dried under a stream of nitrogen.

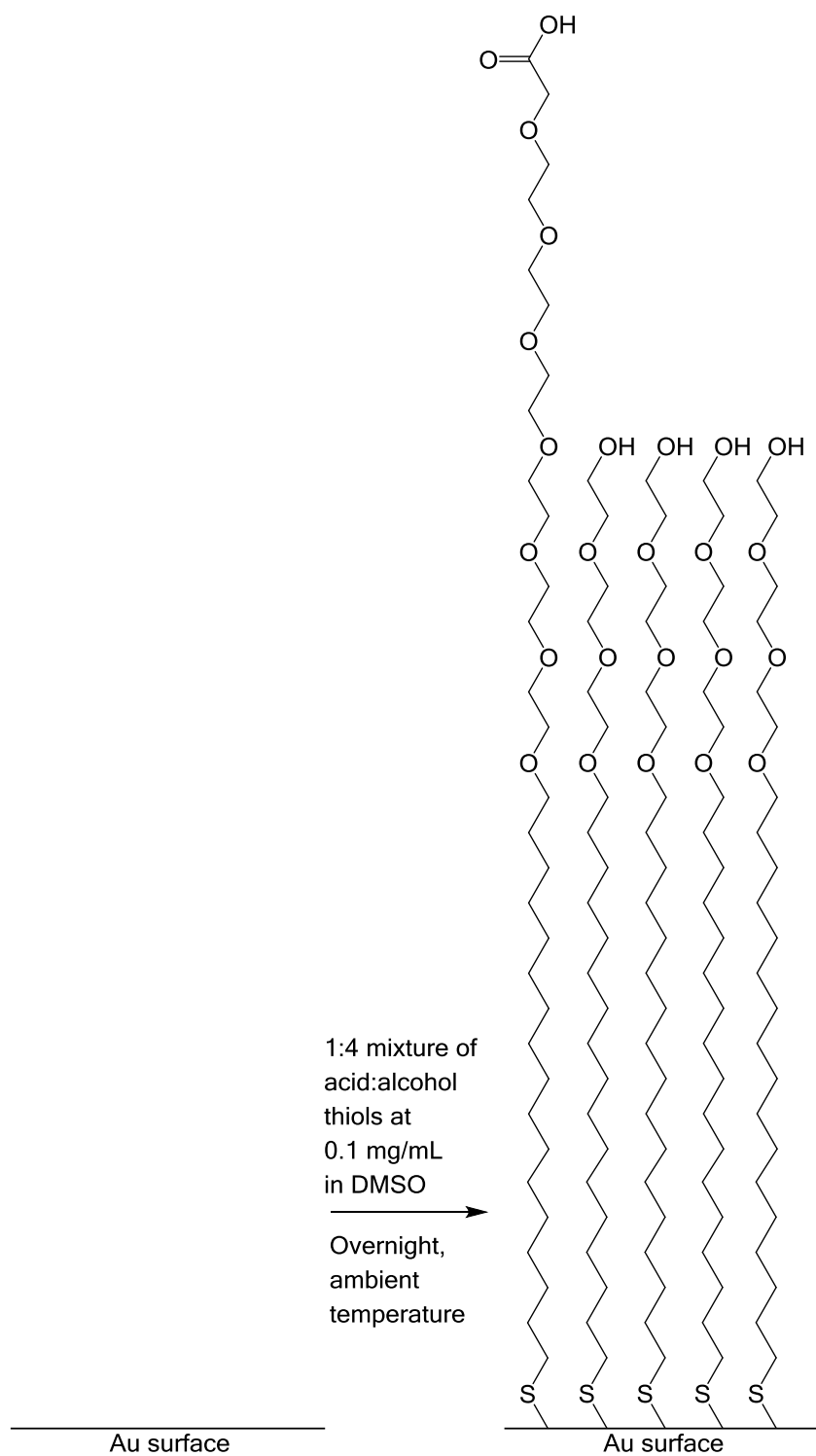


Figure 2.4: Formation of the thiol-derived SAM

The success of the formation of the SAM was assessed by MALDI-MS. Aliquots (0.3 μL) of THAP matrix solution were spotted on wells prior to mass spectrometric analysis. A typical MALDI mass spectrum is shown in Figure 2.5 (peak assignments shown in Table 2.1).

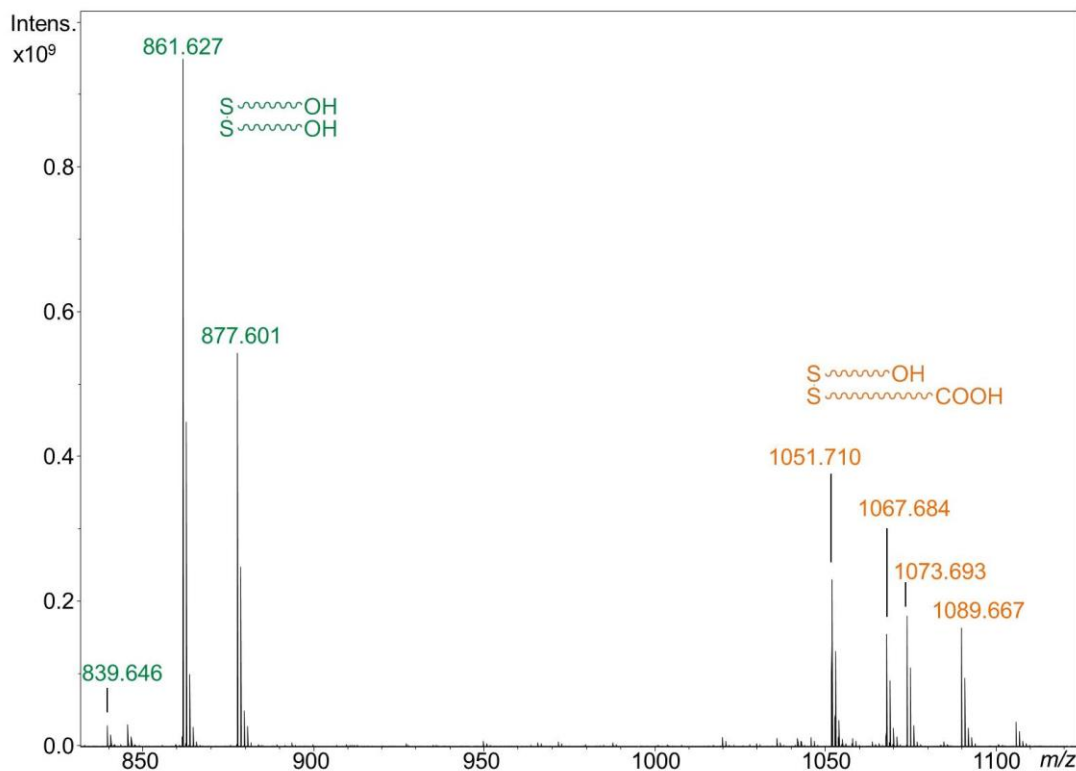


Figure 2.5: MALDI mass spectrum of the thiol-derived SAM formed on a gold-coated MALDI chip

As can be seen in Figure 2.5, peaks are observed for a disulfide having two alcohol termini, and also for a mixed disulfide having one alcohol terminus and one carboxylic acid terminus (Table 2.1). No peaks were observed for free thiols (data not shown), although there are many papers showing that upon MALDI-MS analysis, a combination of thiols and disulfides are formed before subsequent detection.^{171–175} In 1998, Trevor *et al.* published a paper detailing an investigation using two-laser mass spectrometry on thiolate, disulfide and sulfide SAMs.¹⁷¹ SAMs were formed on gold, and studied by nitrogen laser desorption followed by vacuum ultraviolet photoionisation of secondary neutrals on a ToF-MS. The authors found that dimers ($RSSR^+$) and disulfide ions were the predominant species observed in the mass spectra for all alkanethiolates and disulfides studied. The group found that no monomer ions (RS^+) were formed for linear thiolates. The group surmised from the data that dimerization occurred as a result of the recombination of nearest-neighbour thiolates on the surface, followed by intact dimer desorption, rather than recombination in the gas phase. In 2009, Ha *et al.* published a paper further investigating the MALDI process and SAMs.¹⁶² SAMs decorated with peptides and THAP matrix were subjected to two-laser mass spectrometry. The authors proposed that the matrix solution dissolves alkanethiolate molecules from SAMs, which results in formation of disulfide species in solution. The disulfides are then co-crystallised with matrix molecules and subsequently detected by MALDI-MS. The group found that when methanol was dropped onto a SAM surface, and re-collected after a few

seconds, the resulting mass spectrum was almost the same as one recorded directly from the SAM, when both areas were spotted with aliquots of matrix solution. Peaks associated with disulfide ions were observed from spectra obtained in both instances, indicating that the formation of these characteristic disulfide ions does not occur after laser irradiation.¹⁶²

Table 2.1: Peak assignment for mass spectrum of thiol-derived SAM formation on a gold-coated MALDI chip

Nominal m/z value	Assignment
839	$[M+H]^+$ - 2 alcohol termini
861	$[M+Na]^+$ - 2 alcohol termini
877	$[M+K]^+$ - 2 alcohol termini
1051	$[M+Na]^+$ - 1 alcohol terminus, 1 carboxyl terminus
1067	$[M+K]^+$ - 1 alcohol terminus, 1 carboxyl terminus
1073	$[M+2Na-H]^+$ - 1 alcohol terminus, 1 carboxyl terminus
1089	$[M+Na+K-H]^+$ - 1 alcohol terminus, 1 carboxyl terminus

2.3. Investigation into the most suitable matrix

THAP was used as a MALDI matrix for the experiments described above on the basis of work done by others using the same SAM components (S. Hudsonⁱⁱⁱ and R. Castangia^{iv}). An independent study was undertaken with non-nucleophilic solvents, and a range of matrices (Figure 2.6) to assess whether the optimum matrix was being used for the SAM work.

A SAM was formed on a gold-coated MALDI chip. Two separate solutions for the matrix molecules, THAP, DHB, CHCA and dithranol (DIT) were prepared, one in acetonitrile and one in acetone, with both prepared at concentrations of 15 mg/mL.

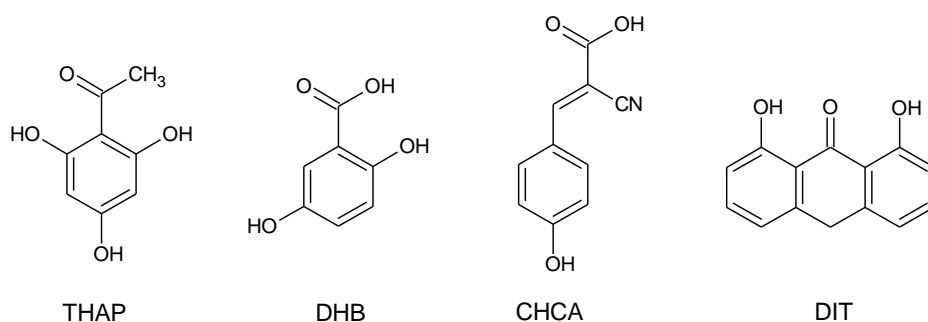


Figure 2.6: Structures of matrix molecules used for the investigation

ⁱⁱⁱ The University of York

^{iv} Manchester Interdisciplinary Biocentre, The University of Manchester

Aliquots of each matrix solution were spotted on wells of the gold-coated MALDI chip with the SAM formed on them, and the gold-coated MALDI chip was analysed by MALDI-MS. The S/N of the peak due to $[M+Na]^+$ for a disulfide having two alcohol termini (m/z 861) was extracted from each mass spectrum.

S/N is often used as a general indicator of the quality of a mass spectrum and enables comparison between samples prepared in different manners in which the absolute response may be affected by the chemical background: a high S/N is indicative of a 'high quality' mass spectrum (usually in combination with a large number of signals and narrow peaks).¹⁰¹ In 2012, Toh-Boyo *et al.* published a paper comparing sample preparation methods and evaluating intra- and intersample variability in MALDI-MS profiling.¹⁰¹ Multiple mass spectra were obtained from a MALDI spot containing an *E. coli* suspension. Data for absolute signal intensity (peak area), S/N and peak resolution from these spectra were extracted. It was found that the overall trends seen when using peak areas and S/N were similar.

The aim of comparing the different matrices and solvents analysed was to determine the appropriateness of each matrix and each solvent. In order to determine which matrix (and which solvent) was most appropriate for the SAM work, a measure which could be compared across the differently acquired spectra was needed, as different matrices have a particularly dramatic effect on spectral appearance which is why a means of comparing data was required. The S/N of the peak due to $[M+Na]^+$ ions was used to give an indication of the quality of data generated using the different matrix and solvent combinations for the SAM work.

The mean S/N of the resulting peak in the spectra from eight wells analysed with the same matrix and solvent were calculated, and the results summarised in Table 2.2.

Table 2.2: Matrices and solvents used for the investigation (number of replicates = 8)

Matrix	Solvent	Mean S/N of peak at m/z 861	Standard deviation	Coefficient of Variation / %	Matrix clusters observed
THAP	Acetonitrile	3904.0	1304.42	33.4	No
	Acetone	5311.5	2309.9	43.5	No
DHB	Acetonitrile	286.8	140.9	49.1	No
	Acetone	427.3	273.5	64.0	No
CHCA	Acetonitrile	39.0	25.4	65.2	Yes
	Acetone	41.2	20.8	50.5	Yes
DIT	Acetonitrile	21.7	10.5	48.1	Yes
	Acetone	11.1	4.4	39.8	Yes

The term 'matrix clusters observed' has been used to refer to when a mass spectrum is dominated by a series of peaks due to clusters of matrix molecules, which usually include signals from a variety of ionised species (e.g. $[M+H]^+$ and $[M+Na]^+$). A repeating pattern is observed in the mass spectrum. These dominant matrix cluster peaks can interfere with the signals from peaks associated with the analyte.

It can be seen that THAP gave a higher average S/N value than the other matrices tested. DHB gave a lower average S/N value than THAP. CHCA and DIT both showed matrix clusters, making them unsuitable for use with this work. This indicates that the most suitable matrix to use is THAP. Either of the solvents could be used with THAP, although acetonitrile gave S/N values with a smaller standard deviation and coefficient of variation than acetone. Therefore THAP prepared in acetonitrile was used for all further work. Although the standard deviation and coefficient of variation for THAP are large (which may, in part, be due to the innate variability observed with MALDI¹⁷⁶) THAP seems to be the 'least worst' option in terms of matrix to be used.

Having determined that THAP was the matrix to be used in future work, different THAP concentrations (with acetonitrile as the solvent) and volumes of matrix solution spotted on the wells were investigated to determine the impact on the quality of the data obtained by MALDI-MS. A mixed SAM containing the carboxylic:alcohol terminated groups at a 1:4 molar ratio was formed on the wells of a gold-coated MALDI chip. The S/N of the peak due to $[M+Na]^+$ ions (m/z 861) was used to give an indication of the quality of data generated using the different concentrations and volumes of matrix solutions. The mean S/N of the resulting peak in the spectra from three replicate wells was calculated, and the results summarised in Table 2.3.

Table 2.3: THAP concentrations and volumes spotted on wells (number of replicates = 3)

THAP concentration / mg/mL	Volume / μ L	Mean S/N of peak at m/z 861	Standard deviation	Coefficient of variation / %
5	1	461.5	76.4	16.6
	0.5	676.0	352.7	52.2
	0.3	283.6	55.5	19.6
10	1	285.4	127.7	44.7
	0.5	387.2	149.0	38.5
	0.3	726.2	111.8	15.4

As can be seen in Table 2.3, there is a difference observed, not only between the different THAP concentrations, but also between the volumes spotted on the wells. For 5 mg/mL THAP in acetonitrile, the mean S/N was lowest when 0.5 μ L of THAP solution was used. However, this produced the largest standard deviation and coefficient of variation. For both

1 μL and 0.3 μL volumes of THAP solution, the standard deviations and coefficients of variation are much smaller than with 0.5 μL volumes of matrix solution. For 10 mg/mL THAP in acetonitrile, the mean S/N increases as the volume of matrix solution decreases, and the standard deviation and coefficient of variation decrease. Therefore, THAP in acetonitrile at 10 mg/mL was used for all further work. A 0.3 μL volume of matrix solution was also used, as this gave the best results in terms of mean S/N, standard deviation and coefficient of variation compared to 1 μL and 0.5 μL aliquots of matrix solution.

2.4. Peptide capture on a 4-bromophenylalanine tagged SAM

There are many advantages to the incorporation of a chemical tag. In the case of 4-BrPhe (Figure 2.7), the two isotopes of bromine (^{79}Br and ^{81}Br) occur in nature in a ratio that is roughly 50:50. The incorporation of bromine (e.g. 4-BrPhe) gives a characteristic isotopic distribution when brominated species are analysed by mass spectrometry.

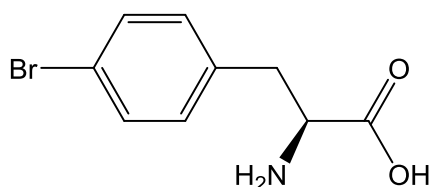


Figure 2.7: L-4-bromophenylalanine

The isotopic distribution of peaks associated with the SAM species containing one 4-BrPhe (i.e. one alcohol terminus and one carboxyl terminus carrying the 4-BrPhe tag) can be seen in Figure 2.8.

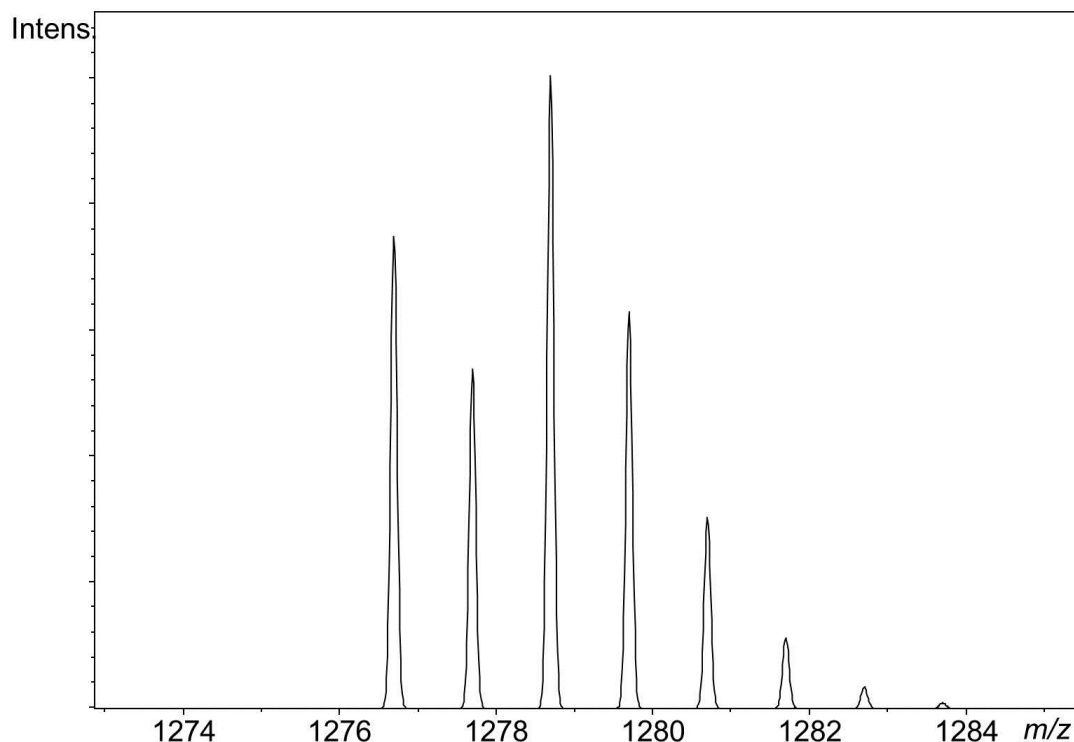


Figure 2.8: Simulated isotopic distribution of peaks associated with $[M+Na]^+$ mixed disulfide having one alcohol terminus and one carboxylic acid terminus on which the 4-BrPhe chemical tag is incorporated

As can be seen in Figure 2.8, a very distinctive isotopic distribution is observed following incorporation of 4-BrPhe on the SAM. Therefore, it should be easy to identify signals due to bromine-containing tagged peptides produced following capture on the tagged SAM.

2.4.1. Activation of SAM carboxylic acid groups

Having determined the SAM had formed successfully on the surface of the gold-coated MALDI chip, the carboxylic acid groups were activated toward nucleophilic attack by an amine. This was achieved using either PFP and 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) (each at 180 mM), or NHS and EDC (at 25 mM and 100 mM respectively), with both sets of reagents being prepared in anhydrous dimethylformamide (DMF) (NHS/EDC activation shown in Figure 2.9). Aliquots (1 μ L) (for PFP/EDC: 180 nmol of each spotted on a well, for NHS/EDC: 25 nmol of NHS and 100 nmol of EDC spotted on a well) of these solutions were left on the wells for an hour before washing the gold-coated MALDI chip with acetonitrile and drying it under a stream of nitrogen.

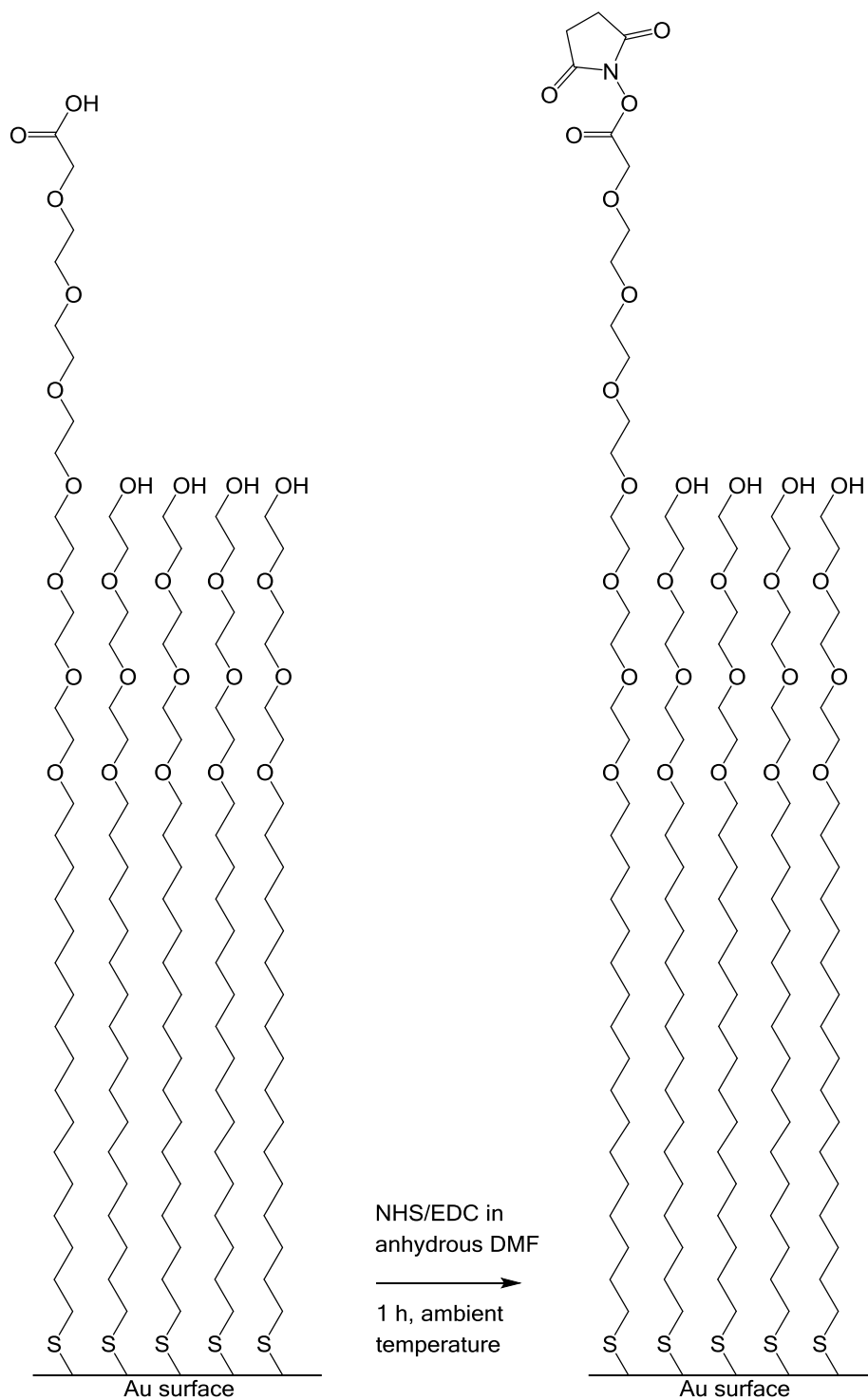


Figure 2.9: NHS/EDC activation of the thiol-derived SAM

2.4.2. Coupling of 4-bromophenylalanine to activated SAM

A 50 mM solution of 4-BrPhe was prepared in DMF:phosphate buffered saline (PBS) (50:50 v/v), and aliquots (1 μ L, 50 nmol of 4-BrPhe) were spotted onto the activated SAM. The

solution was left to react overnight at ambient temperature (Figure 2.10). The gold-coated MALDI chip was washed with acetonitrile and dried under a stream of nitrogen.

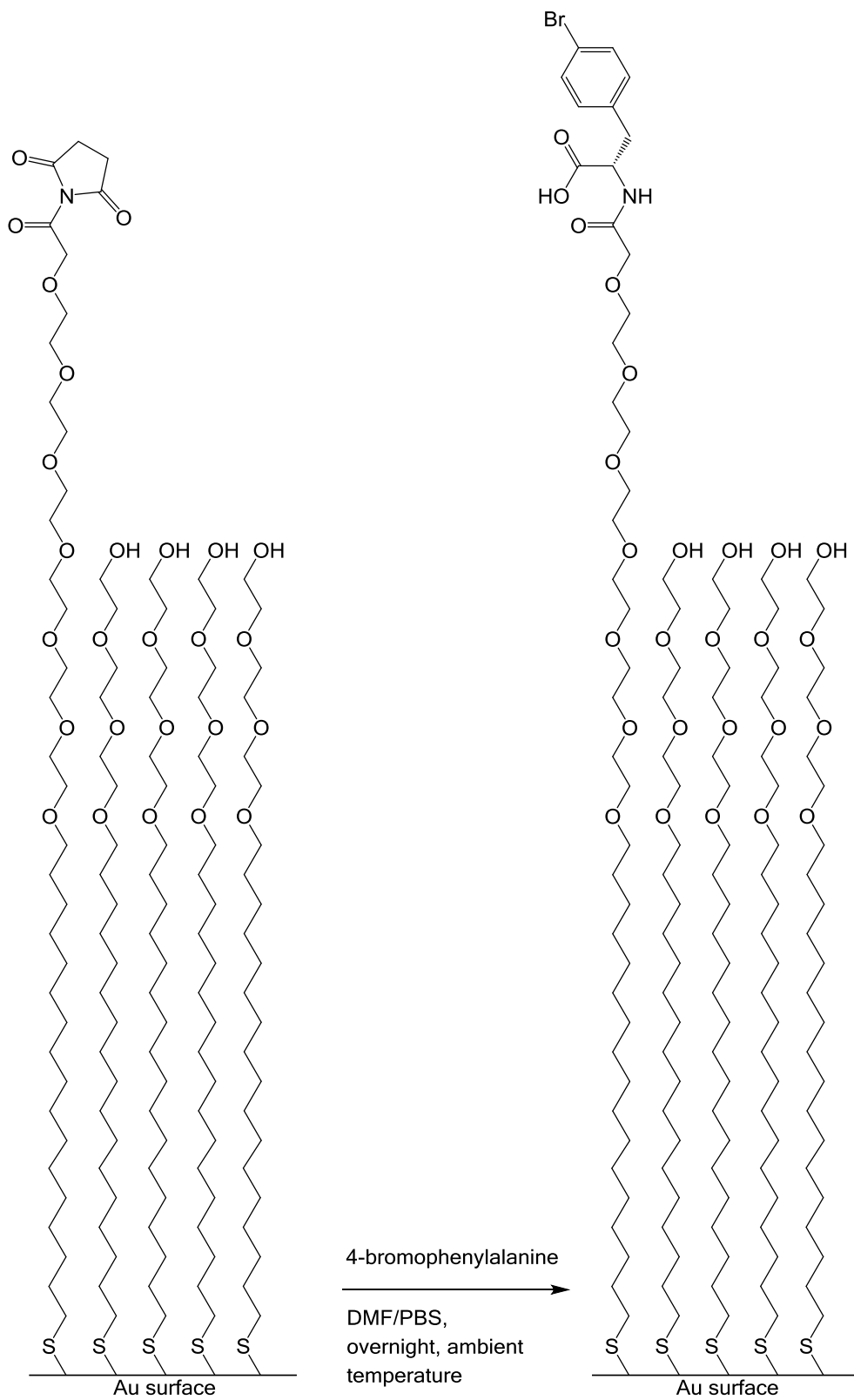


Figure 2.10: 4-Bromophenylalanine coupling to the NHS-activated carboxylic acid group

The gold-coated MALDI chip was subjected to mass spectrometric analysis to determine whether 4-BrPhe had been coupled to the activated SAM. Aliquots (0.3 μL) of THAP matrix solution were spotted on wells prior to mass spectrometric analysis. A representative mass spectrum showing the results of spotting 4-BrPhe on NHS-activated SAM is shown in Figure 2.11 (peak assignments shown in Table 2.4).

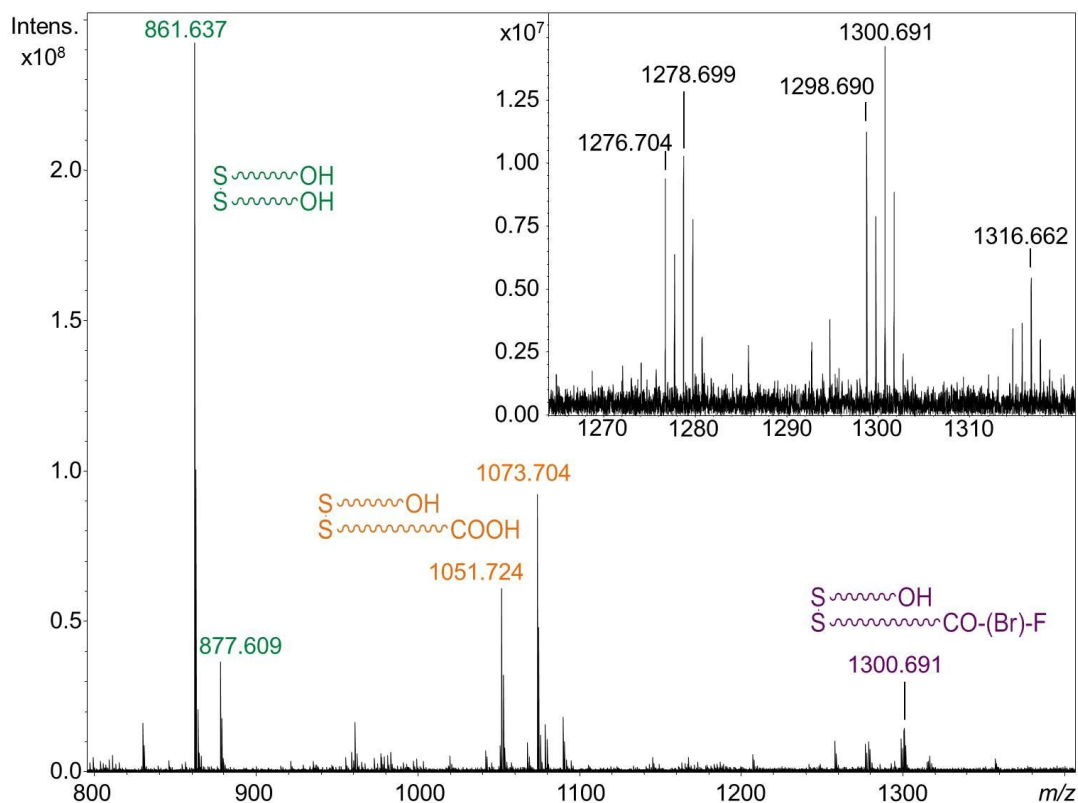


Figure 2.11: MALDI mass spectrum of 4-BrPhe coupled to the NHS-activated SAM

Peaks were observed at m/z 1276/1278 and 1298/1300, which correspond to a mixed disulfide having one alcohol terminus and one carboxylic acid terminus with 4-BrPhe coupled to it (Table 2.4). However, there is incomplete reaction of the 4-BrPhe with the SAM, as can be seen by the presence of peaks at m/z 1051 and 1073, which correspond to a mixed disulfide having one alcohol terminus and one carboxylic acid terminus (unreacted SAM).

Table 2.4: Peak assignment for mass spectrum of 4-bromophenylalanine tagged SAM

Nominal m/z value	Assignment
861	$[M+Na]^+$ - 2 alcohol termini
877	$[M+K]^+$ - 2 alcohol termini
1051	$[M+Na]^+$ - 1 alcohol terminus, 1 carboxyl terminus
1073	$[M+2Na-H]^+$ - 1 alcohol terminus, 1 carboxyl terminus
1276/1278	$[M+Na]^+$ - 1 alcohol terminus, 1 CO-4-BrPhe terminus
1298/1300	$[M+2Na-H]^+$ - 1 alcohol terminus, 1 CO-4-BrPhe terminus
1314/1316	$[M+Na+K-H]^+$ - 1 alcohol terminus, 1 CO-4-BrPhe terminus

2.4.3. Activation of 4-BrPhe-tagged SAM carboxylic acid groups

The carboxylic acid groups on the 4-BrPhe-tagged SAM were activated using PFP/EDC (180 mM/180 mM) or NHS/EDC (25 mM/100 mM) in anhydrous DMF for one hour (NHS/EDC activation shown in Figure 2.12). Aliquots (1 μ L) of these solutions (for PFP/EDC: 180 nmol of each spotted on a well, for NHS/EDC: 25 nmol of NHS and 100 nmol of EDC spotted on a well) were left on the wells for an hour at ambient temperature before washing the gold-coated MALDI chip with acetonitrile and drying it under a stream of nitrogen.

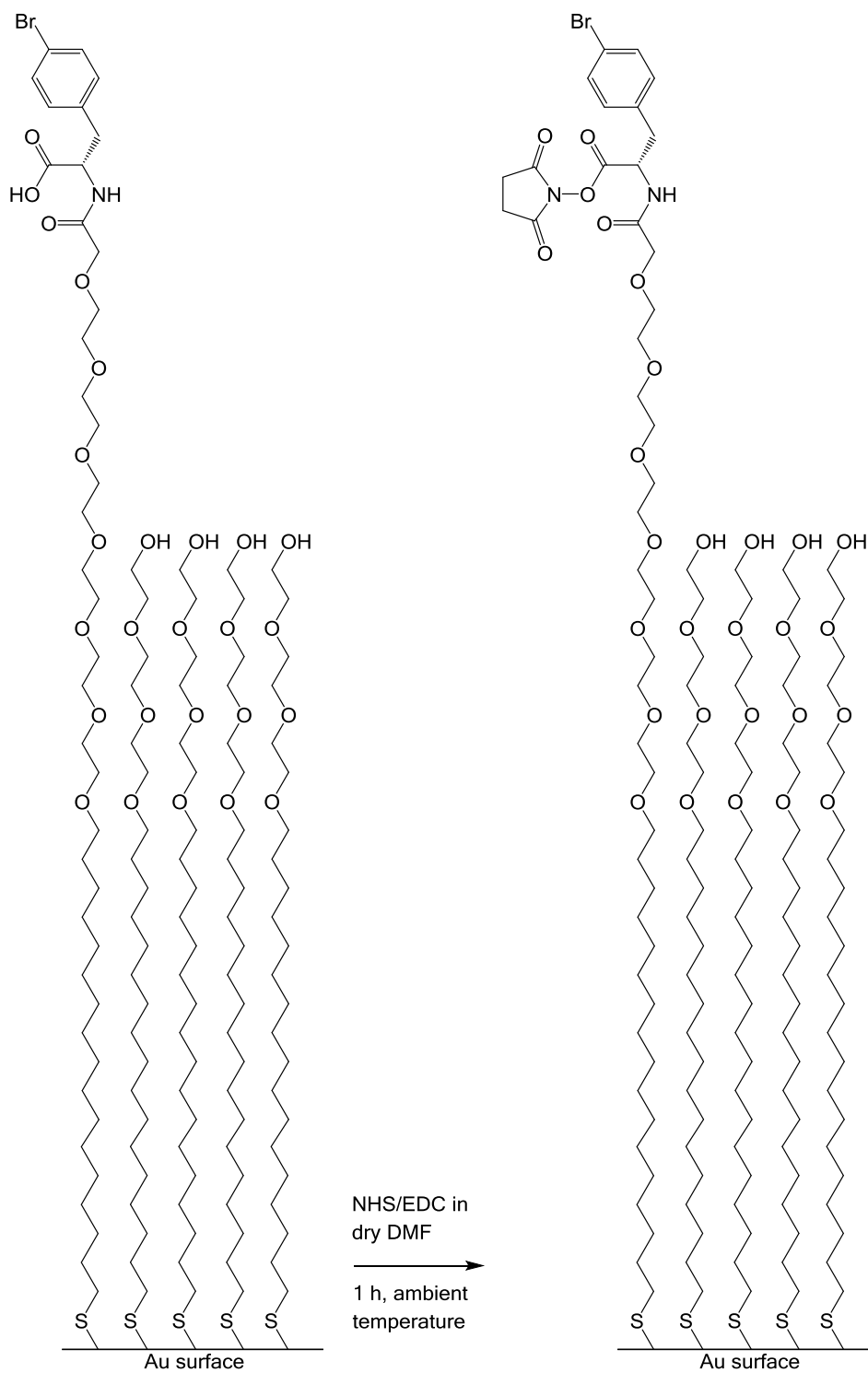


Figure 2.12: NHS/EDC activation of the 4-BrPhe-tagged SAM

2.4.4. Capture of single peptides on activated 4-BrPhe-tagged SAM

Capture of the peptides Leu-Gly-Gly and Gly-Gly-Val was attempted on the activated 4-BrPhe-tagged SAM (generic reaction shown in Figure 2.13). Stock solutions of the peptides Leu-Gly-Gly and Gly-Gly-Val were prepared at 50 mM in PBS. From these stock solutions, solutions were prepared to enable peptide capture with the peptides at four different amounts in 1 μ L aliquots: 50 nmol, 10 nmol, 5 nmol and 0.50 nmol in PBS. These amounts are all in excess of the estimated nanomole levels of carboxylic acid groups available for reaction: the surface area of a single well on a gold-coated MALDI chip is approximately 3.2 mm².¹⁶⁹ If the total density of a SAM on a clean gold surface is taken as approximately 585 pmol cm⁻²,¹⁵⁴ the amount of carboxylic acid functionality can be approximated (given that the ratio between carboxyl and alcohol terminated thiols is 1:4); this gives an estimate of a carboxylic acid level of 3.7 pmol on the surface of each well.

Aliquots (1 μ L) of solution containing peptide were spotted on the activated 4-BrPhe-tagged SAM and left overnight at ambient temperature. The gold-coated MALDI chip was washed with acetonitrile and dried under a stream of nitrogen before analysis by MALDI-MS. Aliquots (0.3 μ L) of THAP matrix solution were spotted on wells prior to mass spectrometric analysis.

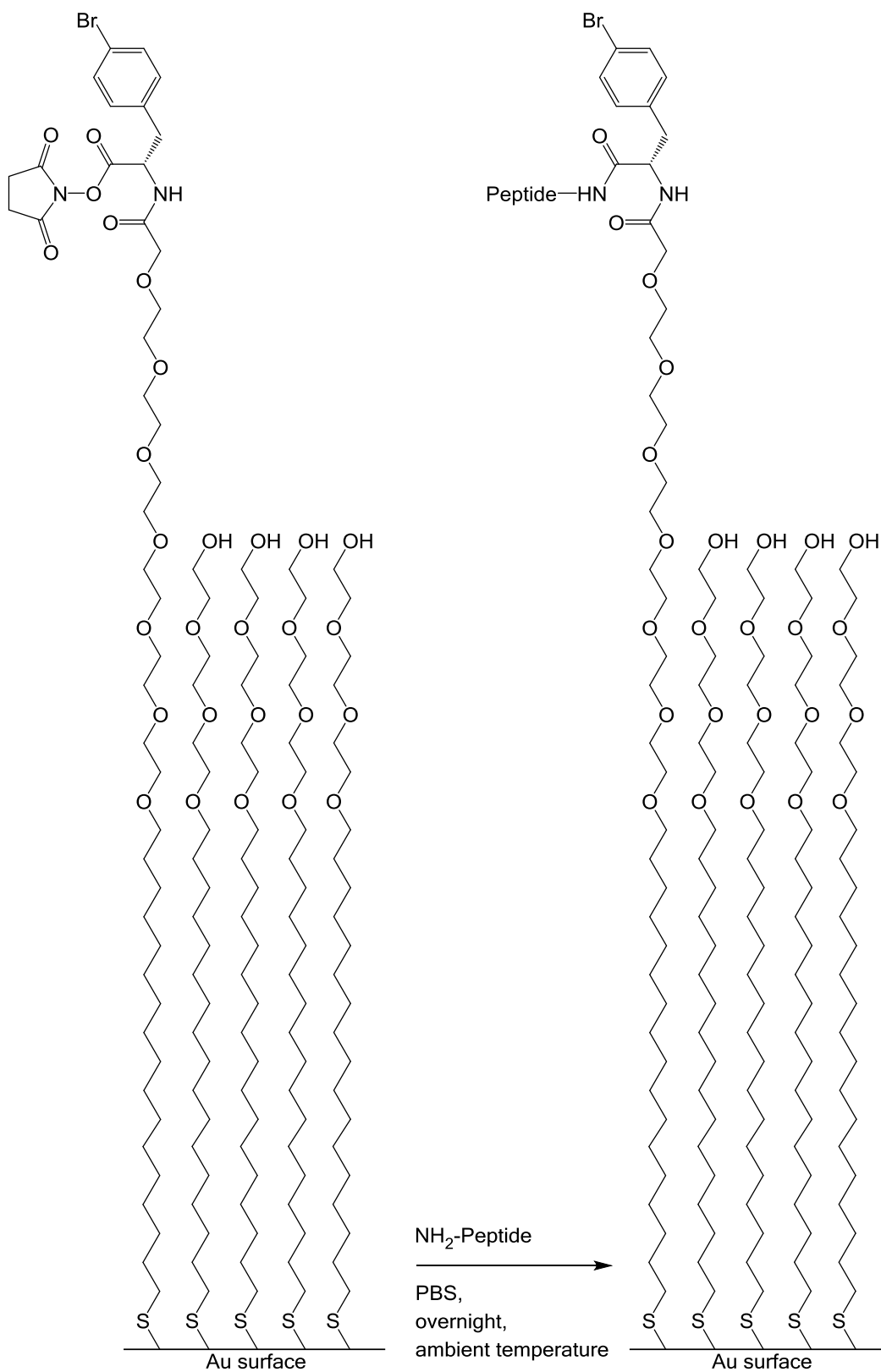


Figure 2.13: Generic reaction of a peptide with the activated 4-BrPhe tagged SAM

Table 2.5 shows the expected m/z values for Leu-Gly-Gly and Gly-Gly-Val captured on the activated 4-BrPhe-tagged SAM, the amount spotted on the wells, and whether peaks were observed in the mass spectrum for peptide captured on the activated 4-BrPhe-tagged SAM.

Table 2.5: Peptides captured on SAM construct and expected peaks (the product is defined as observed if the full isotopic distribution was observed in the mass spectra)

Peptide	Molecular mass of peptide / g mol^{-1}	Expected values of $[\text{M}+\text{Na}]^+$ and $[\text{M}+2\text{Na}-\text{H}]^+$ for peptide on 4-BrPhe tagged SAM / m/z	Amount of peptide spotted on well / nmol	Product observed
Leu-Gly-Gly (LGG)	245	1503/1505 1525/1527	50	Yes
			10	Yes
			5	No
			0.5	No
Gly-Gly-Val (GGV)	231	1489/1491 1511/1513	50	Yes
			10	Yes
			5	Yes
			0.5	No

A representative mass spectrum showing the results of spotting 50 nmol of Leu-Gly-Gly on the activated 4-BrPhe-tagged SAM spot is shown in Figure 2.14 (peak assignments are shown in Table 2.6).

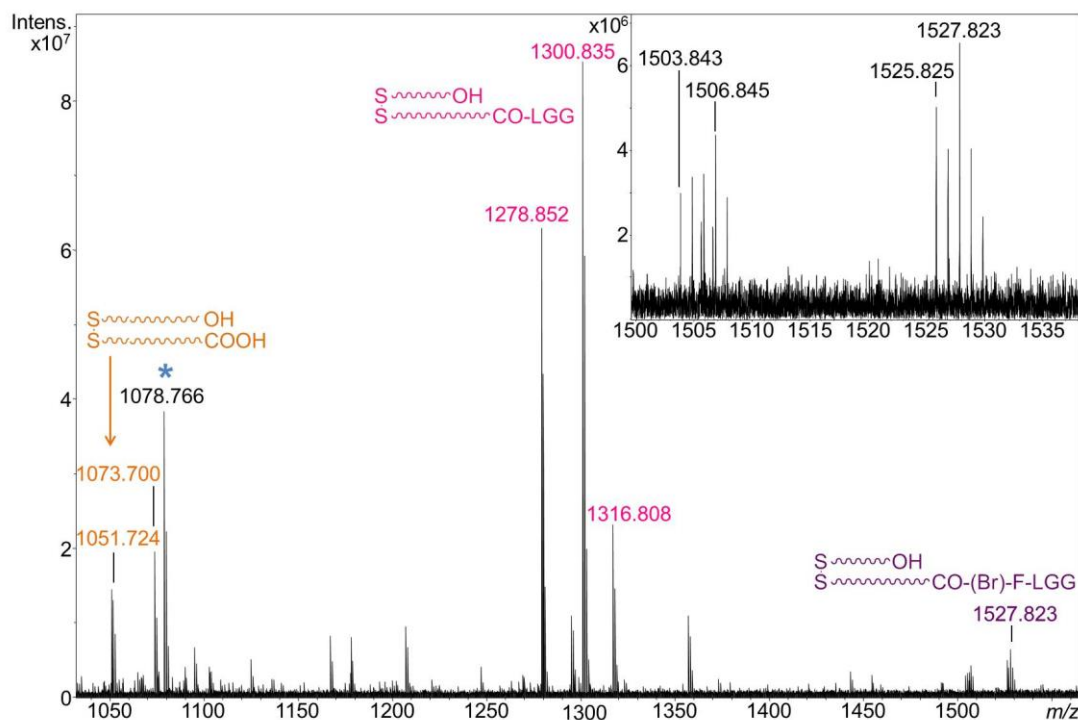


Figure 2.14: MALDI mass spectrum of 50 nmol Leu-Gly-Gly spotted on the PFP-activated 4-BrPhe-tagged SAM (* indicates a PFP-derived impurity (unidentified because product ion analysis failed))

Peaks were observed at m/z 1503/1505 and 1525/1527, which correspond to Leu-Gly-Gly captured by the 4-BrPhe tag attached to the carboxylic acid group terminating one chain of a mixed disulfide (the other chain contains an alcohol terminus) (Table 2.6). There were also peaks observed for a mixed disulfide having one alcohol terminus and one carboxylic acid terminus with Leu-Gly-Gly captured on it (in the absence of the 4-BrPhe tag) (m/z 1278, 1300 and 1316) and for unreacted SAM (m/z 1051 and 1073). Peaks were not observed for the activated SAM. The peak at m/z 1078 is only observed when PFP is used as the activating reagent, but has not been successfully identified.

Table 2.6: Peak assignments for mass spectrum of Leu-Gly-Gly captured on activated 4-BrPhe tagged SAM

Nominal m/z value	Assignment
1051	$[M+Na]^+$ - 1 alcohol terminus, 1 carboxyl terminus
1073	$[M+2Na-H]^+$ - 1 alcohol terminus, 1 carboxyl terminus
1278	$[M+Na]^+$ - 1 alcohol terminus, 1 CO-LGG terminus
1300	$[M+2Na-H]^+$ - 1 alcohol terminus, 1 CO-LGG terminus
1316	$[M+Na+K-H]^+$ - 1 alcohol terminus, 1 CO-LGG terminus
1503/1505	$[M+Na]^+$ - 1 alcohol terminus, 1 CO-4-BrPhe-LGG terminus
1525/1527	$[M+2Na-H]^+$ - 1 alcohol terminus, 1 CO-4-BrPhe-LGG terminus

When the peptide Gly-Gly-Val was captured on the 4-BrPhe-tagged SAM, peaks were observed for a mixed disulfide having one alcohol terminus and one carboxylic acid terminus with Gly-Gly-Val captured on it, with (m/z 1264 ($[M+Na]^+$) and 1286 ($[M+2Na-H]^+$)) and without (m/z 1489/1491 ($[M+Na]^+$) and 1511/1513 ($[M+2Na-H]^+$)) the 4-BrPhe tag, and also for unreacted SAM (m/z 1051 ($[M+Na]^+$) and 1073 ($[M+2Na-H]^+$)).

For both Leu-Gly-Gly and Gly-Gly-Val, peaks were observed which corresponded to the peptide captured on the activated SAM with and without the 4-BrPhe tag. This would suggest that during the activation of the 4-BrPhe-tagged SAM, untagged SAM was being activated, which subsequently reacted with the peptide (peaks were observed for unreacted SAM in spectra obtained after 4-BrPhe coupling to activated SAM). This resulted in peaks being observed which are associated with peptide captured on activated untagged SAM. In order to provide a comparison of the two peptide-containing products observed (peptide captured on the SAM with and without the 4-BrPhe tag), and therefore get a possible indication of the relative amounts of peptide captured on the tagged and untagged SAM, the peak intensity for each of the peaks in the isotopic cluster of each cationised species (e.g. $[M+Na]^+$) was extracted from the data. The total peak intensity of each of the peaks in the isotopic envelope was used due to the fact that peaks due to species including 4-BrPhe are split over the two Br isotope-containing species (Figure 2.8). Peak intensity has previously been used as a means of comparing differing SAM-containing species from mass spectrometric data by Castangia and Flitsch.¹⁷⁰

The idea of using peak intensity for comparing the relative amounts of different ions is well accepted with proteomic relative quantification strategies often using peak intensity to generate quantification ratios (examples include SILAC^{93,103} and TMT¹⁰⁸).

For each spectrum, the extracted peak intensities were entered into Equation 2.1. A mean was then taken for the peak intensity ratio of product:starting material peaks, and plotted

against the amounts of each of the peptides spotted on the SAM, with and without the 4-BrPhe tag (Figure 2.15). This produced a comparison of relative intensities of the peaks from the two products formed.

Mean peak intensity ratio of product/starting material

$$= \frac{(\text{peak intensity of product} + \text{Na}^+) + (\text{peak intensity of product} + \text{K}^+) + (\text{peak intensity of product} + (2\text{Na} - \text{H})^+)}{(\text{peak intensity of SAM} + \text{Na}^+) + (\text{peak intensity of SAM} + \text{K}^+) + (\text{peak intensity of SAM} + (2\text{Na} - \text{H})^+)}$$

Equation 2.1: Equation used to generate relative intensity ratio of product:starting materials

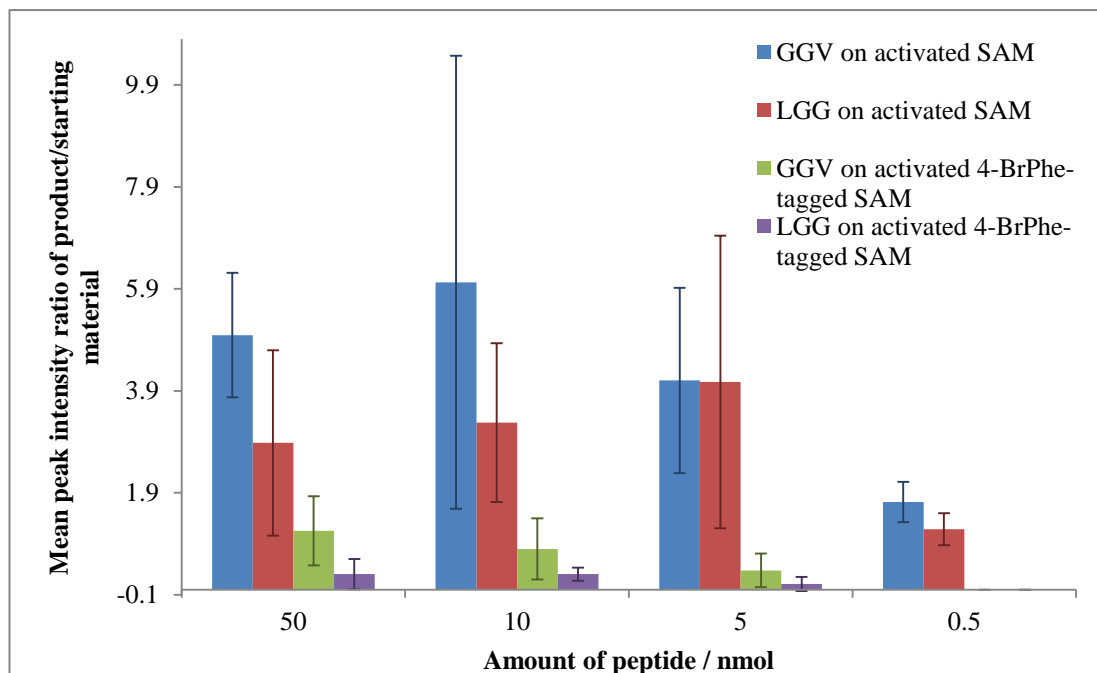


Figure 2.15: Plot showing ratio between peak intensity of the product and the starting material MALDI signals for peptides Gly-Gly-Val and Leu-Gly-Gly on both the SAM and the 4-BrPhe-tagged SAM (error bars show one standard deviation either side of the mean value) (number of replicates = 4).

As can be seen in Figure 2.15, the relative peak intensity ratios for the signals for the peptide captured on the activated SAM that had not incorporated the 4-BrPhe tag are much larger than those for the peptide captured on activated 4-BrPhe-tagged SAM. Of the two peptides, Gly-Gly-Val shows the higher peak intensity ratio of product:starting material.

For the two peptides, the lowest amount of peptide spotted on a well for which peptide captured on the activated 4-BrPhe-tagged SAM was observed to be different. Capture of Leu-Gly-Gly was observed at the 10 nmol level on the activated 4-Br-Phe-tagged SAM. However, capture of the peptide Gly-Gly-Val was observed at the 5 nmol level on the activated 4-BrPhe-tagged SAM. Both of these are well above the estimated amount of

carboxylic acid groups available on the surface of a well (3.7 pmol), suggesting that the amount of SAM should not have been limiting.

In order to determine whether any differences in mean values for peptide captured on activated SAM and peptide captured on activated 4-BrPhe-tagged SAM were statistically significant, the Mann-Whitney U-test was used. The Mann-Whitney U-test is a non-parametric test used to test the median of two populations without the need for the data to be normally distributed.^{177,178} A two-tailed U-test is applied when the analyst assumes that differences between the data sets can operate in both directions (i.e. either data set may be larger than the other).

In order to perform the U-test, the data from the two groups are ranked from smallest to largest, with the smallest being given the rank '1', the next lowest given the rank '2', through to the highest number. The ranks within each group are then summed. The larger of the two rank totals is given the name of T_x . The total number of data entries in each of the groups is calculated (N_1 and N_2), with N_x being denoted as the number associated with the group with the larger rank total (T_x). These values are then used to calculate the U-value for the data set (Equation 2.2).¹⁷⁹

$$U = N_1 \times N_2 + N_x \frac{(N_x + 1)}{2} - T_x$$

Equation 2.2: Mann-Whitney U-test calculation

The minimum number of data/scores which are generally accepted for use with the Mann-Whitney U-test is five within each data set.¹⁷⁹ Therefore, for small data sets with groups of less than five, an indication of the statistical significance can be gleaned, although the values should be taken as more of a guide due to the small sample set.

The Mann-Whitney U-test was used to determine whether any of the differences between the mean value for peptide captured on activated SAM and peptide captured on activated 4-BrPhe-tagged SAM were statistically significant for the different peptide amounts.

For Gly-Gly-Val: 50 pmol (Mann-Whitney, $U = 0$, $n_1 = n_2 = 4$, $P < 0.05$, two-tailed); 10 pmol (Mann-Whitney, $U = 0$, $n_1 = n_2 = 4$, $P < 0.05$, two-tailed); 5 pmol (Mann-Whitney, $U = 0$, $n_1 = n_2 = 4$, $P < 0.05$, two-tailed); and 0.5 pmol (Mann-Whitney, $U = 0$, $n_1 = n_2 = 4$, $P < 0.05$, two-tailed).

For Leu-Gly-Gly: 50 pmol (Mann-Whitney, $U = 0$, $n_1 = n_2 = 4$, $P < 0.05$, two-tailed); 10 pmol (Mann-Whitney, $U = 0$, $n_1 = n_2 = 4$, $P < 0.05$, two-tailed); 5 pmol (Mann-Whitney, $U = 3$, $n_1 = n_2 = 4$, $P < 0.05$, two-tailed); and 0.5 pmol (Mann-Whitney, $U = 0$, $n_1 = n_2 = 4$, $P < 0.05$, two-tailed).

A Mann-Whitney U-value of '0' is the critical value for data sets containing four values (two-tailed, $P < 0.05$). For seven of the eight data sets the U-value obtained meets the critical value for a two-tailed Mann-Whitney U-test where $P < 0.05$. Therefore, the differences between the mean values obtained for peptide captured on activated SAM and peptide captured on activated 4-BrPhe-tagged SAM could be deemed as statistically significant. However, in practice, a U-value of less than '0' is not possible. For Leu-Gly-Gly at 5 pmol, $U = 3$ indicates that there is no statistical significance between the mean values obtained for peptide captured on activated SAM and peptide captured on activated 4-BrPhe-tagged SAM. However, it is worth repeating that these values should be seen as a guide, and the results should be treated with caution due to the small data sets.

2.5. Peptides captured on untagged SAM

When peptides had been left to react with the activated 4-BrPhe-tagged SAM, peaks associated with unreacted SAM and peptide captured on the activated SAM were observed in the spectra obtained. Due to the inefficient coupling occurring between the activated 4-BrPhe-tagged SAM and the peptide, the efficiency of peptide capture on an activated SAM without the addition of the 4-BrPhe tag was investigated. The smallest amount of peptide which could be spotted on the SAM, and peaks observed in the resulting mass spectrum for peptide captured on the SAM, was also investigated. The peptides were captured at levels of 50 pmol and 5 pmol. These levels were chosen because they are closer to the realistic amounts which would be available in a typical proteomics experiment. It would also show if peptide capture was successful at lower levels than on the activated 4-BrPhe-tagged SAM.

2.5.1. Activation of SAM carboxylic acid groups

Having determined the SAM had formed on the surface of the gold-coated MALDI chip, the carboxylic acid groups were activated toward nucleophilic attack by an amine using either PFP (180 mM) and EDC (180 mM), or NHS (25 mM) and EDC (100 mM), with both sets of reagents being prepared in dry DMF (NHS/EDC activation shown in Figure 2.9). Aliquots (1 μ L) of these solutions (for PFP/EDC: 180 nmol of each spotted on a well, for NHS/EDC: 25 nmol of NHS and 100 nmol of EDC spotted on a well) were left on the wells for an hour before washing the gold-coated MALDI chip with acetonitrile and drying under a stream of nitrogen.

2.5.2. Peptide capture on activated SAM carboxylic acid groups

Capture of peptides on activated SAM was performed using the peptides Leu-Gly-Gly, Gly-Gly-Val, Gly-Pro-Gly-Gly and Thr-Tyr-Ser (generic reaction showing peptide capture shown in Figure 2.16). Stock solutions of the peptides detailed in Table 2.7 were prepared at 50 mM in PBS. Subsequent dilutions of the solutions were made to give solutions which would result in 50 pmol (0.05 mM) and 5 pmol (5 μ M) amounts of peptide being spotted on each well in a 1 μ L aliquot. The peptide solutions were spotted on wells and left overnight at ambient temperature. The gold-coated MALDI chip was washed with acetonitrile and dried under a stream of nitrogen prior to MALDI-MS analysis. Aliquots (0.3 μ L) of THAP matrix solution were spotted on wells prior to mass spectrometric analysis.

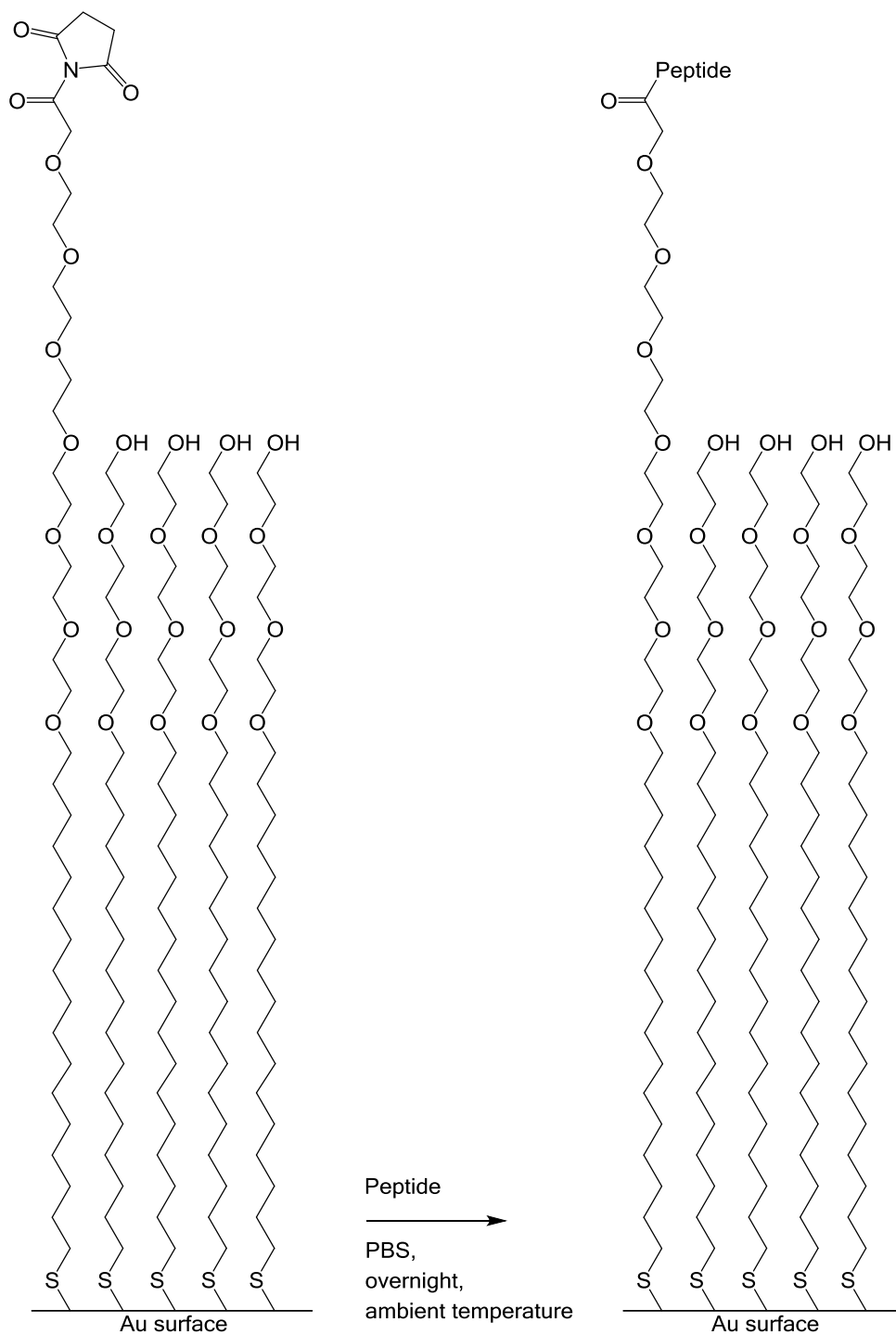


Figure 2.16: Generic reaction of a peptide with activated SAM

Table 2.7: Peptides and amounts used for capture, activating reagent used, expected m/z values of product captured on the activated SAM and whether the product was observed (the product is defined as observed if the full isotopic distribution of either of the species $[M+Na]^+$ and $[M+2Na-H]^+$ was observed in the mass spectra)

Peptide	Molecular mass of peptide / $g\ mol^{-1}$	Activated SAM	Amount of peptide spotted on well / pmol	Expected values of $[M+Na]^+$ and $[M+2Na-H]^+$ / m/z	Product observed
Leu-Gly-Gly (LGG)	245	NHS	50	1278	Yes
		NHS	5	1300	Yes
		PFP	50		Yes
		PFP	5		Yes
Gly-Gly-Val (GGV)	231	NHS	50	1264	Yes
		NHS	5	1286	No
		PFP	50		Yes
		PFP	5		Yes
Gly-Pro-Gly-Gly (GPGG)	286	NHS	50	1319	Yes
		NHS	5	1341	Yes
		PFP	50		Yes
		PFP	5		Yes
Thr-Tyr-Ser (TYS)	369	NHS	50	1402	Yes
		NHS	5	1424	No
		PFP	50		Yes
		PFP	5		No

Table 2.7 shows that for Leu-Gly-Gly, Gly-Gly-Val and Gly-Pro-Gly-Gly, peaks associated with the relevant peptide captured on the SAM were observed by MALDI-MS when wells were spotted with 5 pmol of peptide. For Thr-Tyr-Ser, no peaks were observed for peptide captured on the SAM at an amount of 5 pmol for either NHS/EDC or PFP/EDC activation.

In order to provide a comparison between the different peptides captured on the activated SAM, the amount of peptide spotted on the activated SAM and the activating reagent used (NHS or PFP), the peak intensity for each of the peaks in the isotopic cluster of each cationised species (e.g. $[M+Na]^+$) was extracted from the data. For each spectrum, these were entered into Equation 2.1. A mean was then taken for the peak intensity of each of the peptides spotted on the SAM (at each amount) and with both of the activating reagents (NHS and PFP). This produced a comparison of the relative intensities of the peaks from

the different peptides captured on the activated SAM and the different activating reagents used. Figure 2.17 and Figure 2.18 show the results for NHS and PFP activation respectively.

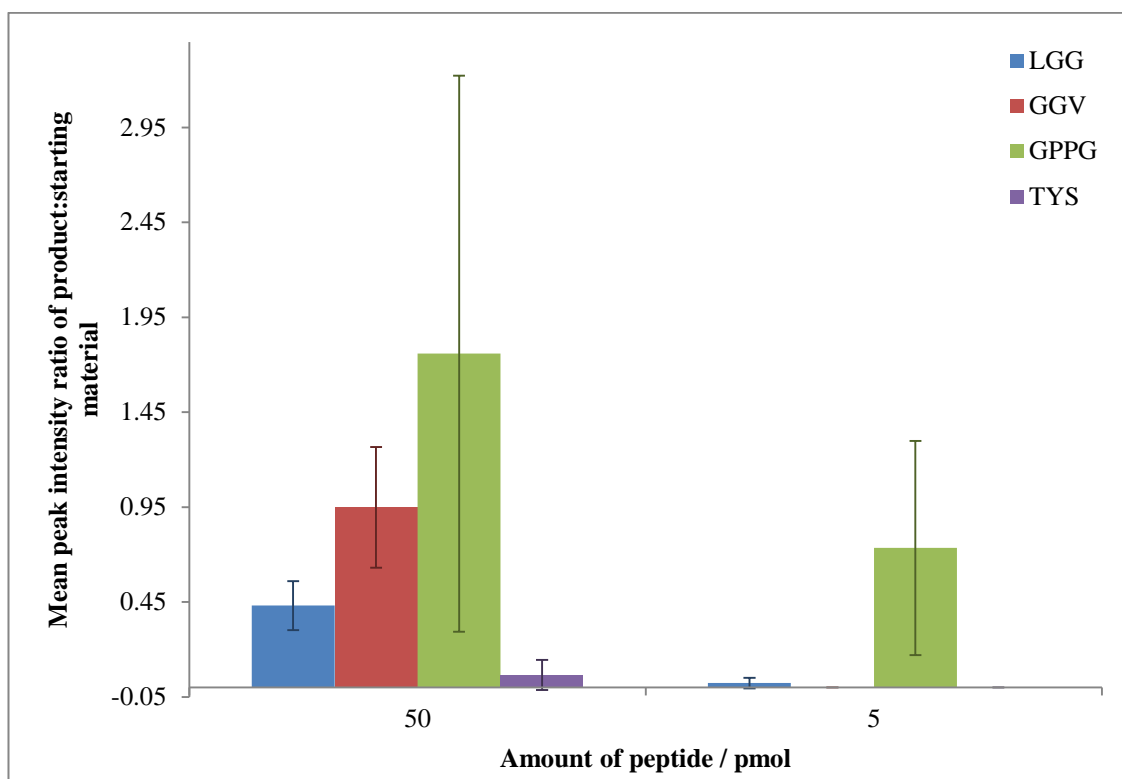


Figure 2.17: Plot showing the mean ratio of peak intensity for product:starting material MALDI signals for each of the peptides with NHS as the activating reagent (Error bars show one standard deviation either side of the mean) (number of replicates = 4)

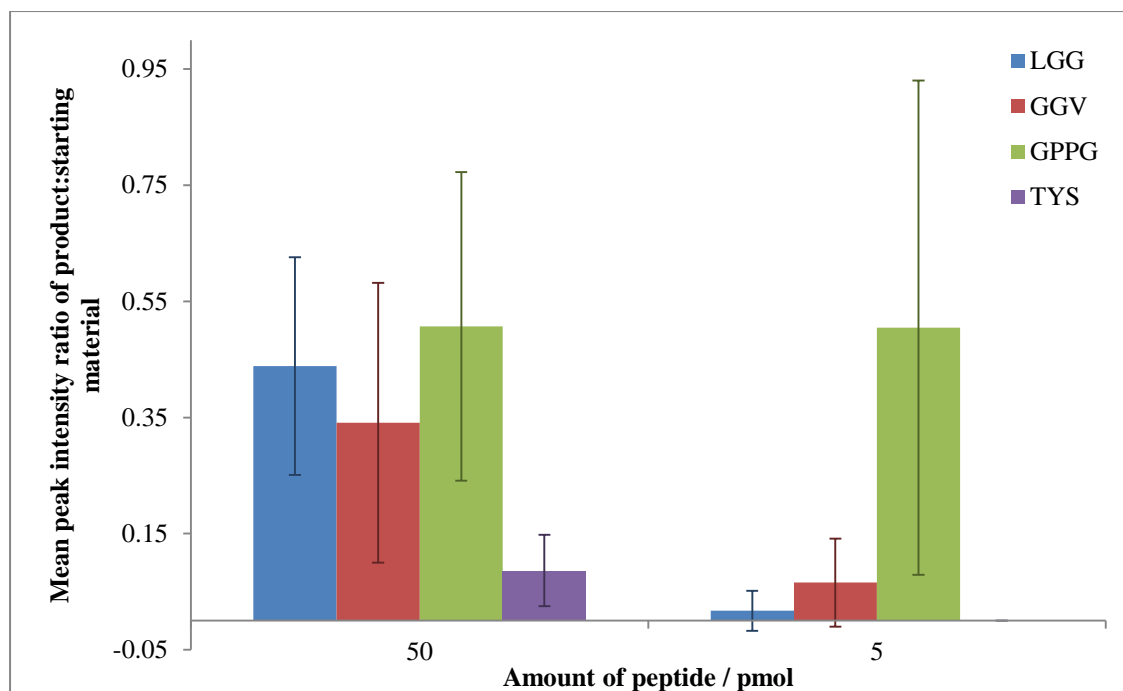


Figure 2.18: Plot showing the mean ratio of peak intensity for product:starting material MALDI signals for each of the peptides with PFP as the activating reagent (Error bars show one standard deviation either side of the mean) (number of replicates = 4)

With NHS used as the activating reagent, the peptides Leu-Gly-Gly and Gly-Pro-Gly-Gly were successfully captured on the activated SAM when 5 pmol of peptide was spotted on each well containing activated SAM. Capture of the peptides Gly-Gly-Val and Thr-Tyr-Ser was not observed when 5 pmol of peptide was spotted on each well containing activated SAM.

With PFP used as the activating reagent, the peptides Leu-Gly-Gly, Gly-Gly-Val and Gly-Pro-Gly-Gly were all successfully captured on the activated SAM when 5 pmol of peptide was spotted on each well containing activated SAM. Capture of the peptide Thr-Tyr-Ser was not observed when 5 pmol of peptide was spotted on each well containing activated SAM.

For both NHS and PFP activation, the mean ratio of peak intensity for product:starting material for capture of the peptide Thr-Tyr-Ser is much lower than for the other peptides when 50 pmol of peptide was spotted on each well containing activated SAM. This could be indicative of a problem with one of the amino acids in the peptide. Based on the results seen for 4-BrPhe coupling to the activated SAM (Section 2.4), it could be assumed that the Tyr amino acid, with its aromatic side chain, is the reason for the low ratio of peak intensity for product:starting material.

Figure 2.17 and Figure 2.18 both show that the peptide with the highest ratio of peak intensity of product:starting material is Gly-Pro-Gly-Gly. For the peptides Gly-Pro-Gly-Gly

and Gly-Gly-Val (at 50 pmol), the mean ratio is higher where NHS, rather than PFP, is the activation reagent (Figure 2.17). There have been numerous investigations into the susceptibility of activated esters to hydrolysis, where it has been surmised that NHS activated esters are more susceptible to hydrolysis than PFP activated esters.¹⁸⁰⁻¹⁸²

However, due to the higher product:starting material ratio obtained for NHS as the activation reagent, NHS was used rather than PFP as the activation reagent for subsequent reactions.

Due to the non-normal distribution of the data sets, the Mann-Whitney U-test was applied to the data, to compare the mean ratio of peak intensity for peptide captured at 50 pmol and 5 pmol on activated SAM for both NHS and PFP as the activation reagent. As was the case with the data in Section 2.4.4, only four replicates were performed, and therefore the U-values obtained from the small data sets should be treated as a guide.

For NHS: Leu-Gly-Gly (Mann-Whitney, $U = 0$, $n_1 = n_2 = 4$, $P < 0.05$, two-tailed); Gly-Gly-Val (Mann-Whitney, $U = 0$, $n_1 = n_2 = 4$, $P < 0.05$, two-tailed); Gly-Pro-Pro-Gly (Mann-Whitney, $U = 4$, $n_1 = n_2 = 4$, $P < 0.05$, two-tailed); and Thr-Tyr-Ser (Mann-Whitney, $U = 4$, $n_1 = n_2 = 4$, $P < 0.05$, two-tailed).

For PFP: Leu-Gly-Gly (Mann-Whitney, $U = 4$, $n_1 = n_2 = 4$, $P < 0.05$, two-tailed); Gly-Gly-Val (Mann-Whitney, $U = 0$, $n_1 = n_2 = 4$, $P < 0.05$, two-tailed); Gly-Pro-Pro-Gly (Mann-Whitney, $U = 2$, $n_1 = n_2 = 4$, $P < 0.05$, two-tailed); and Thr-Tyr-Ser (Mann-Whitney, $U = 9$, $n_1 = n_2 = 4$, $P < 0.05$, two-tailed).

There is a difference between the U-values obtained when NHS was used as the activation reagent compared to PFP as the activation reagent. More non-zero results were obtained for data obtained using PFP than NHS as the activation reagent. A Mann-Whitney U-value of '0' is the critical value for data sets containing four values ($P < 0.05$, two-tailed).

Therefore, Leu-Gly-Gly and Gly-Gly-Val captured on activated SAM using NHS as the activation reagent, and Gly-Gly-Val captured on activated SAM using PFP as the activation reagent meet the critical value for a two-tailed Mann-Whitney U-test where $P < 0.05$. This implies that, in these three instances, the difference between the mean peak intensity values calculated when 50 pmol and 5 pmol of peptide are spotted on activated SAM could be deemed as statistically significant. For Gly-Pro-Pro-Gly and Thr-Tyr-Ser captured on activated SAM using NHS as the activated reagent, and Leu-Gly-Gly, Gly-Pro-Pro-Gly and Thr-Tyr-Ser captured on activated SAM using PFP as the activation reagent, the difference between the mean peak intensity values calculated for the two peptide amounts is greater than '0', and should not be deemed statistically significant.

Levels of peptide capture were successful for a lower amount of peptide being spotted on wells containing activated SAM than for wells containing activated 4-BrPhe-tagged SAM (Section 2.4). This could be due to low levels of incorporation of the 4-BrPhe prior to peptide capture on the activated 4-BrPhe-tagged SAM, inefficient activation of the SAM (and

subsequent activated 4-BrPhe-tagged SAM) or simply an accumulation of inefficient chemistry at the different stages of the reaction.

2.5.3. Use of alternative activating reagents and conditions to investigate peptide capture on untagged SAM

Peaks have been observed for species associated with unreacted SAM when peptide capture has been attempted on the 4-BrPhe-tagged SAM and the untagged SAM using both NHS/EDC and PFP/EDC activation. Therefore, alternative activating reagents and conditions were explored in the hope of improving activation extent, and subsequent reactivity of the activated SAM species. Reviews of coupling reactions have shown a variety of coupling reagents have been used for peptide coupling.^{183,184}

NHSS (Figure 2.19) has been used in conjunction with EDC for carboxyl activation due to the fact that both are water soluble.^{133,185} The conditions used (Sections 2.4.1, 2.4.3 and 2.5.1) for activation using NHS/EDC or PFP/EDC in DMF are more suited to the solubilisation of the NHS and PFP than the EDC. This mismatch of solubilities could have been a reason for activation inefficiency (and therefore spectra recorded after reaction of activated SAM with peptides showing peaks for unreacted SAM). Therefore, the possibility that NHSS is more suited to this work was investigated as a potential way to increase reaction efficiency between the carboxylic acid group of the SAM and the activating moiety.

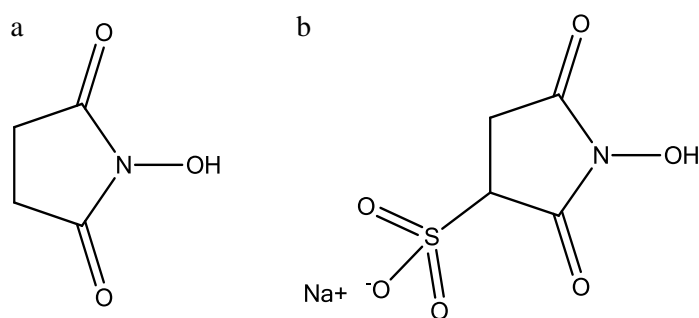


Figure 2.19: (a) *N*-hydroxysuccinimide and (b) *N*-hydroxysodiumsulfosuccinimide

Having prepared a SAM on a gold-coated MALDI chip, a comparison of the efficiency of activation using NHSS/EDC (25 mM/100 mM) chemistry or NHS/EDC (25 mM/100 mM) chemistry was performed. For both sets of activation reagents, solutions were prepared using two solvent conditions: DMF and acetonitrile/water (50:50 v/v). This was to provide a comparison with previous experiments with DMF, and secondly to use a solvent system more compatible with EDC and NHSS (i.e. acetonitrile/water). Aliquots (1 μ L) of the solutions (for NHS/EDC: 25 nmol of NHS and 100 nmol of EDC on a well, for NHSS/EDC: 25 nmol of NHSS and 100 nmol of EDC spotted on a well) were spotted onto individual

wells, and left for an hour at ambient temperature. The gold-coated MALDI chip was then washed with acetonitrile and dried under a stream of nitrogen.

The peptides Leu-Gly-Gly and Phe-Gly-Gly were prepared in PBS. Aliquots (1 μ L) (5 nmol of peptide) were spotted on wells containing SAM and left overnight at ambient temperature. The gold-coated MALDI chip was then washed with acetonitrile and dried under a stream of nitrogen prior to MALDI-MS analysis. Aliquots (0.3 μ L) of THAP matrix solution were spotted on the wells prior to mass spectrometric analysis.

When analysed, peaks were observed for each of the peptides captured on the SAM. For Leu-Gly-Gly, a peak at m/z 1278 indicated the presence of a disulfide having one alcohol terminus and one carboxylic acid terminus with the peptide coupled to it. For Phe-Gly-Gly, the presence of a peak at m/z 1312 indicated a disulfide having one alcohol terminus and one carboxylic acid terminus with the peptide coupled to it. Peaks were observed at m/z 1051 and 1073, which indicated the presence of a disulfide having one alcohol terminus and one carboxylic acid terminus (unreacted SAM).

For spectra which were obtained using acetonitrile/water as the activation solvent, unassignable peaks were observed. In spectra which were recorded for wells where NHS/EDC activation chemistry was used prior to peptide coupling, peaks were observed at m/z 1203 and m/z 1225. In spectra which were recorded for wells where NHSS/EDC activation chemistry was used prior to peptide coupling, peaks were observed at m/z 1237 and m/z 1259. Upon MS/MS analysis the same product ions were observed at m/z 932 and 958 for the two different unassigned precursors. However, no reasonable assignment has been made. The peaks were not observed when DMF was used as the solvent system. Therefore, it can be assumed that the formation of these ions was due to the use of the acetonitrile/water solvent system.

One of the reasons for trying NHSS was to increase the solubility of EDC by using a solvent more suited to solubilising EDC (acetonitrile/water)¹⁸³ than the DMF used for NHS/EDC and PFP/EDC activation. However, there were unassignable peaks observed in the mass spectra recorded after peptide capture when the acetonitrile/water solvent system was used. This means that in spite of the theoretical advantage of solubility in acetonitrile and water, in practice this solvent system was less practical to use than DMF; the data suggest that a side-reaction of some kind is taking place when acetonitrile/water is used. Therefore, the original methodology of NHS/EDC in DMF was used in further work.

2.6. Investigation of inefficient activation of carboxylic acid group activation for peptide capture

In spectra obtained after activation of the carboxylic acid groups followed by subsequent tagging with 4-BrPhe or peptide capture, peaks were observed in the spectra which corresponded to unreacted SAM (m/z 1051 and 1073). The coupling of the 4-BrPhe and of a peptide to the activated SAM is carried out in an aqueous medium, using PBS to prepare the 4-BrPhe tag/peptide solution and to control the pH. There are several potential scenarios which could be causing the presence of the peaks associated with unreacted SAM. The activated ester could have been hydrolysed before the tag or peptide could react with it which, due to the conditions used for the reaction, would mean trying to react a carboxylic acid and primary amine together at room temperature. The peptide could have been captured on the SAM, but then be hydrolysed again during the reaction (the least likely option due to the fact that this would involve hydrolysis of an amide). Finally, the 4-BrPhe tag may have not been coupled to the activated ester SAM efficiently in the first place.

An investigation was undertaken to determine whether, in these aqueous steps, hydrolysis of the activated carboxylic acid group could occur, which would convert the activated SAM back to the carboxylic acid group, so that it would then not be available to react with either the 4-BrPhe tag (Figure 2.20) or the peptide. This would reduce the efficiency of the chemistry by reducing the amount of SAM-based activated ester available for reaction.

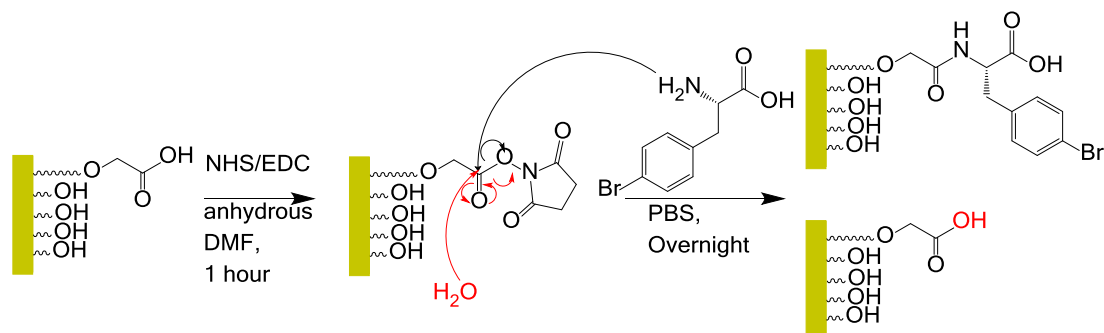


Figure 2.20: Proposed hydrolysis of NHS-activated carboxyl SAM

In order to investigate the potential hydrolysis of the activated ester, an investigation was undertaken using H₂¹⁸O-containing PBS. Parallel control analyses were undertaken using H₂¹⁶O-containing PBS for each H₂¹⁸O-containing PBS experiment. This meant that two gold-coated MALDI chips were used in parallel: a 'light/¹⁶O chip' and a 'heavy/¹⁸O chip'. For each investigation, new gold-coated MALDI chips were used with freshly prepared SAMs.

For each capture stage, two equal aliquots of PBS were dried down. An equivalent volume of water, either H₂¹⁶O or H₂¹⁸O was then added, to re-dissolve the solid salt. This created

two solutions, prepared in exactly the same way but one prepared in ^{18}O -labelled water and one prepared in unlabelled water.

After SAMs were formed on each of the two gold-coated MALDI chips, the chips were washed with acetonitrile and dried under a stream of nitrogen. The carboxylic acid groups, onto which tag would be captured, were activated using either NHS or PFP in anhydrous DMF, with EDC as the coupling reagent. Although PFP had previously been found to be less suitable for the activation of the SAM (due to the lower peak intensity ratio obtained for peptide captured on the untagged SAM), it was included in this investigation, to determine whether there was a preference for the theorised hydrolysis of one activated ester over the other. The activation solution was left on the wells for an hour, before being washed off and dried under a stream of nitrogen. The activated esters were then further reacted with either 4-BrPhe or a simple tripeptide.

Following mass spectrometric analysis, the peak intensity values for each of the peaks associated with the unreacted SAM were extracted from the spectra. The ratio of peak intensity for $[\text{M}+2]:\text{M}$ peaks was calculated for each of the isotopic peaks for each of the cationised unreacted SAM species (m/z 1051, 1073 and 1089) observed in the mass spectra for each well. The mean of the ratios for all of the cationised species was then taken for each well. Finally, the mean for all the wells was taken. A comparison was then made between the ratios obtained in unlabelled buffer and ^{18}O -labelled buffer.

Incorporation and detection of ^{18}O into the carboxylic acid group of the SAM was investigated by Roberto Castangia during the course of his PhD.¹⁷⁰ The carboxylic acid group was activated using NHS, before H_2^{18}O was spotted on wells containing the activated SAM, and left overnight. Upon MALDI-MS analysis, incorporation of the ^{18}O into the carboxylic acid groups was clearly visible in the resulting mass spectra. Since this positive control experiment had already been carried out and ^{18}O incorporation successfully demonstrated in this system, this experiment was not repeated as part of this work.

2.6.1. Natural exchange

It was possible that when a solution of PBS was spotted on the SAM, exchange of the oxygen between the buffer and the carboxylic acid group could occur naturally over time (Figure 2.21).

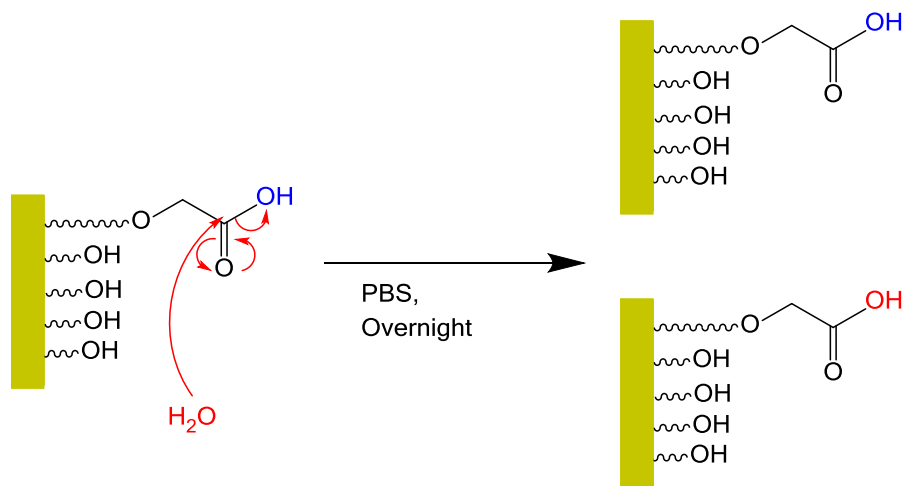


Figure 2.21: Natural exchange of the oxygen in the buffer (red) with the oxygen in the carboxylic acid (blue)

Wells with the SAM formed on them were spotted with aliquots (1 μL) of either the light or heavy PBS and left overnight at ambient temperature. The wells were then washed with acetonitrile and dried under a stream of nitrogen. Aliquots (0.3 μL) of THAP matrix solution were spotted on wells prior to mass spectrometric analysis. The wells were analysed by MALDI-MS, in order to determine the extent of background exchange between H_2^{18}O in the buffer and ^{16}O in the carboxylic acid group on the 'heavy chip'. If exchange between the oxygen in the buffer and carboxylic acid group took place then the isotopic distribution would be altered by an increase in the intensity of the $[\text{M}+2]$ peak, which would be detectable mass spectrometrically. The extracted data are shown in Figure 2.22.

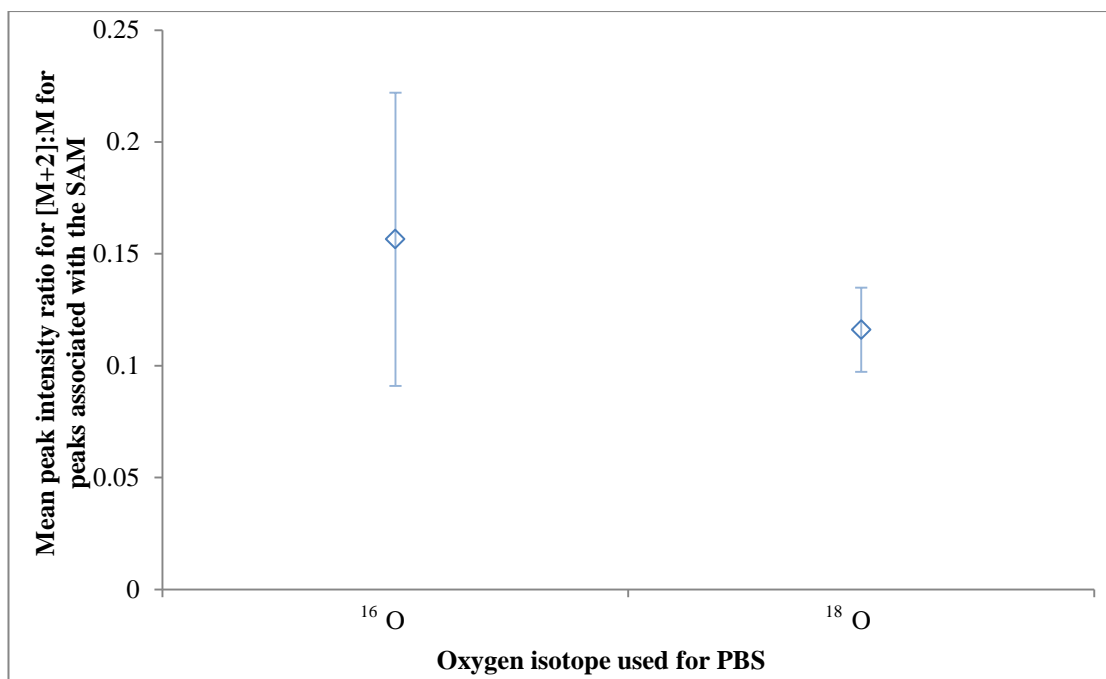


Figure 2.22: Plot showing the mean peak intensity ratio for [M+2]:M for peaks associated with the SAM after ^{18}O exchange overnight (Error bars show one standard deviation either side of the mean) (number of replicates = 8)

As can be seen in Figure 2.22, there is a small difference between the values obtained for the mean peak intensity ratio for [M+2]:M when unlabelled and H_2^{18}O -containing PBS were spotted on wells with SAM formed on them and left overnight. The mean ratio value obtained for wells spotted with H_2^{18}O -containing PBS is slightly lower than the ratio value obtained for unlabelled PBS. The error bars for the two sets of data overlap. Since the data are not normally distributed, it is not possible to determine the statistical significance of the difference between the mean ratio values using a T-test. Therefore a Mann-Whitney U-test was performed, to assess whether there was any statistical significance to the differences observed between the two values (Mann-Whitney, $U = 16.5$, $n_1 = n_2 = 8$, $P < 0.05$, two-tailed). Based on the U-value obtained, the result is not statistically significant.

2.6.2. Investigating potential hydrolysis during 4-bromophenylalanine coupling to activated SAM

Having determined that there was no statistically significant natural exchange of the ^{18}O in the PBS with the ^{16}O of the carboxylic acid group, the next step was to determine whether hydrolysis of the activated ester was occurring during the 4-BrPhe tag coupling step. The carboxylic acid groups were activated using either NHS/EDC or PFP/EDC in DMF for one hour prior to 4-BrPhe coupling.

Activation was achieved using either PFP (180 mM) and EDC (180 mM), or NHS (25 mM) and EDC (100 mM), with both sets of reagents being prepared in anhydrous DMF. Aliquots (1 μL) of these solutions (for PFP/EDC: 180 nmol of each spotted on a well, for NHS/EDC: 25 nmol of NHS and 100 nmol of EDC spotted on a well) were left on the wells for an hour before washing the gold-coated MALDI chip with acetonitrile and drying it under a stream of nitrogen.

A 50 mM solution of 4-BrPhe in PBS (light or heavy) was prepared, and aliquots (1 μL , 50 nmol of 4-BrPhe) were spotted onto each well. The gold-coated MALDI chips were left overnight at ambient temperature. The gold-coated MALDI chips were then washed with acetonitrile and dried under a stream of nitrogen. Aliquots (0.3 μL) of THAP matrix solution were spotted on wells prior to mass spectrometric analysis.

If hydrolysis of the activated SAM occurred during 4-BrPhe coupling, then the isotopic distribution would be altered by an increase in the intensity of the $[\text{M}+2]$ peak, which would be detectable mass spectrometrically. The extracted data are shown in Figure 2.23.

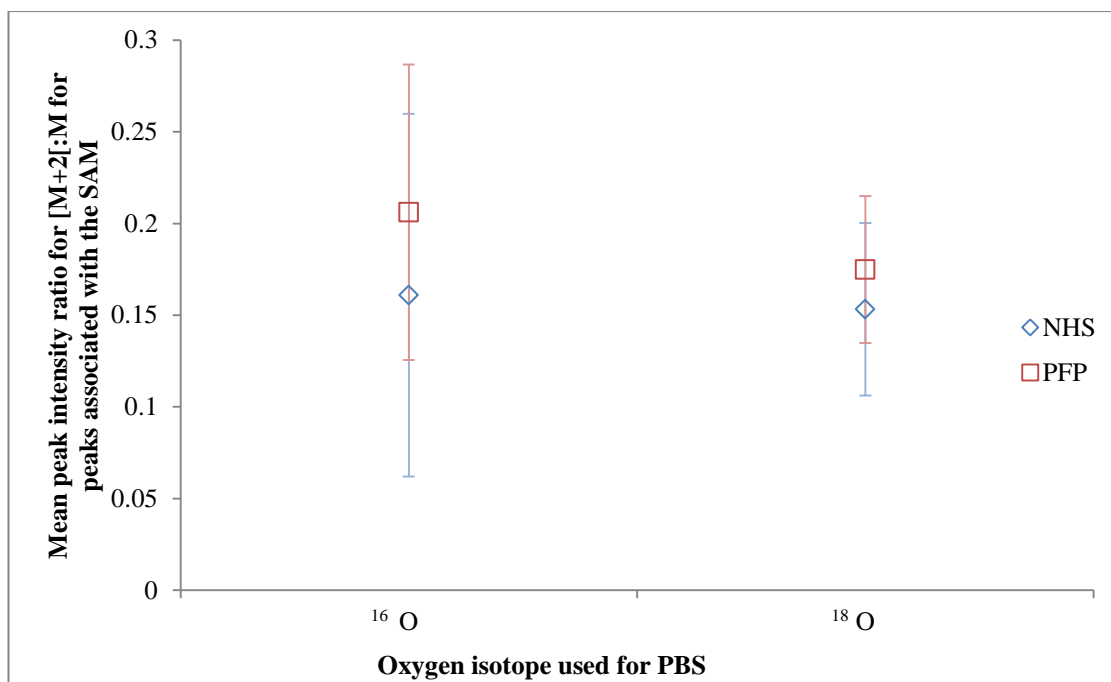


Figure 2.23: Plot showing the mean peak intensity ratio for [M+2]:M for peaks associated with the SAM after coupling of the 4-BrPhe to the activated SAM when using two different activating reagents (Error bars show one standard deviation either side of the mean) (number of replicates = 8)

As can be seen in Figure 2.23, the mean peak intensity ratio for [M+2]:M after coupling of the 4-BrPhe to the activated SAM is almost the same for wells spotted with 4-BrPhe in unlabelled PBS and H_2^{18}O -containing PBS when NHS was used for activation.

There is a slight difference in the mean peak intensity ratio for [M+2]:M after coupling of the 4-BrPhe to the activated SAM between wells spotted with 4-BrPhe in unlabelled PBS and H_2^{18}O -containing PBS when PFP was used for activation. The mean peak intensity ratio for [M+2]:M is lower for spectra recorded for wells spotted with 4-BrPhe in H_2^{18}O -containing PBS. The error bars for the two sets of data overlap, and the values are very comparable. Having examined the data, it is clear that the data are not normally distributed, and therefore a Mann-Whitney two-tailed U-test was performed on the data, to assess whether there was any statistical significance to the differences observed between the mean ratio values obtained for the ^{16}O and ^{18}O experiments using NHS and PFP. For NHS (Mann-Whitney, $U = 24$, $n_1 = n_2 = 8$, $P < 0.05$, two-tailed) and PFP (Mann-Whitney, $U = 32$, $n_1 = n_2 = 8$, $P < 0.05$, two-tailed) the results are not statistically significant.

Hydrolysis of the activated carboxylic acid groups during 4-BrPhe coupling was thus not detectable mass spectrometrically. Therefore, hydrolysis of the activated SAM during the 4-BrPhe capture is not the reason for peaks being observed for unreacted SAM in spectra obtained after 4-BrPhe had been spotted on wells containing activated SAM. This suggests

that the reason for peaks being observed for unreacted SAM is due to inefficient coupling of the 4-BrPhe to the activated SAM.

Peaks were not observed for an activated SAM when either NHS or PFP were used as the activating reagents. This, coupled with the fact that hydrolysis of the activated SAM during 4-BrPhe coupling was not detectable suggests that the activation of the SAM is inefficient.

2.6.3. Investigating potential hydrolysis during peptide capture on activated SAM

Having determined that hydrolysis of the activated esters was not occurring when coupling the 4-BrPhe to the activated SAM, capture of the simple tripeptides Gly-Gly-Gly and Phe-Gly-Gly was attempted on the activated SAM. This was to determine whether the same result that was observed when coupling the 4-BrPhe to the activated SAM (i.e. no detectable hydrolysis) was obtained when capturing peptides on the activated SAM. As a secondary element to the experiment, peptides were chosen which varied in their N-terminal amino acid, with one terminating in Phe. This allowed a comparison between a peptide with an N-terminal Phe and a similar tripeptide without an N-terminal Phe. It was possible that when attempting to capture 4-BrPhe on the activated carboxylic acid groups, the capture was inefficient, as low intensity peaks were observed for 4-BrPhe coupling to the SAM (see Sections 2.4.2 and 2.6.2). One theory was that this inefficient capture could have been due to steric hindrance due to the bulky aromatic side chain of the Phe amino acid. The presence of the bulky side chain in Phe could have been reducing the ability of the N-terminal of the 4-BrPhe to approach the activated SAM. The comparison of the peptides Gly-Gly-Gly and Phe-Gly-Gly allowed an investigation into whether the Phe at the amino terminus hinders a reaction with the SAM activated ester.

After SAM formation, the carboxylic acid groups were activated using either PFP/EDC, or NHS/EDC chemistry, with both sets of reagents being prepared in anhydrous DMF. Aliquots (1 μ L) of these solutions (for PFP/EDC: 180 nmol of each spotted on a well, for NHS/EDC: 25 nmol of NHS and 100 nmol of EDC spotted on a well) were left on the wells for an hour before washing the gold-coated MALDI chip with acetonitrile and drying it under a stream of nitrogen.

Stock solutions of the two peptides Gly-Gly-Gly and Phe-Gly-Gly were each prepared at 50 mM in PBS (light or heavy). Subsequent dilutions were prepared, in order to spot 50 pmol of peptide onto each well, and the gold-coated MALDI chips were left overnight at ambient temperature. The gold-coated MALDI chips were washed with acetonitrile and dried under a stream of nitrogen. Aliquots (0.3 μ L) of THAP matrix solution were spotted on wells prior to mass spectrometric analysis.

If hydrolysis of the activated SAM occurred during peptide capture, then the isotopic distribution would be altered by an increase in the intensity of the [M+2] peak, which would be detectable mass spectrometrically. The extracted data are shown in Figure 2.24 and Figure 2.25.

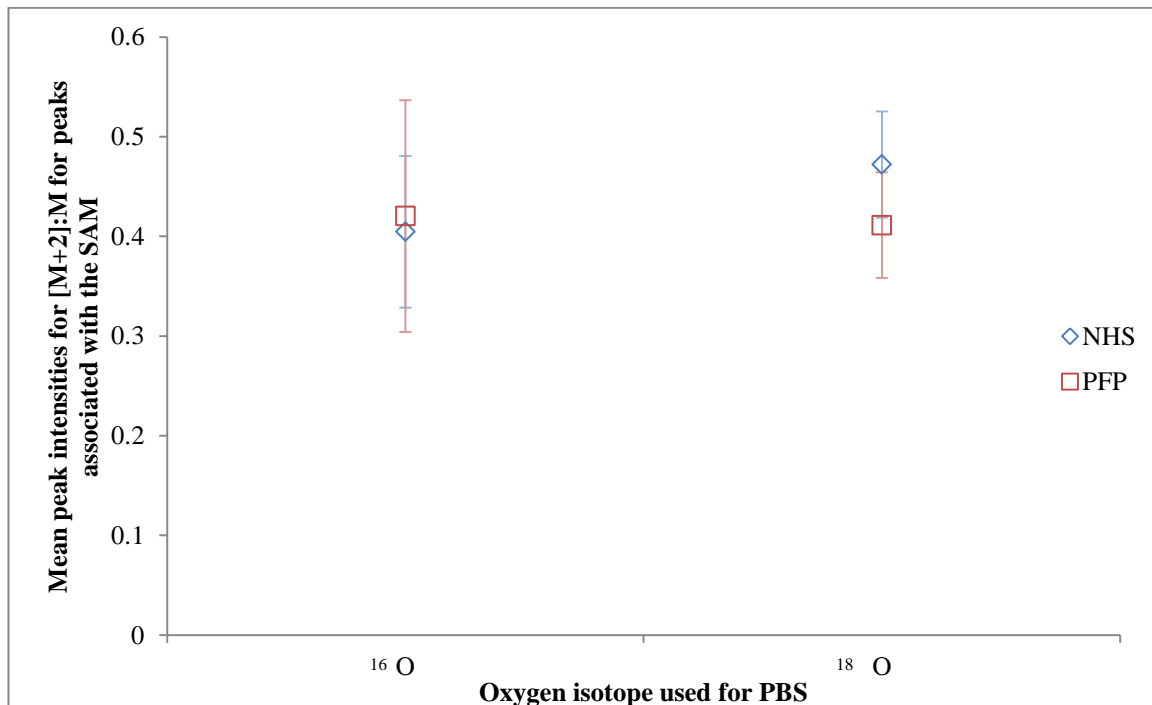


Figure 2.24: Plot showing the mean peak intensity ratio [M+2]:M for peaks associated with the SAM after capture of the peptide Gly-Gly-Gly on the activated SAM when using two different activating reagents (Error bars show one standard deviation either side of the mean) (number of replicates = 8)

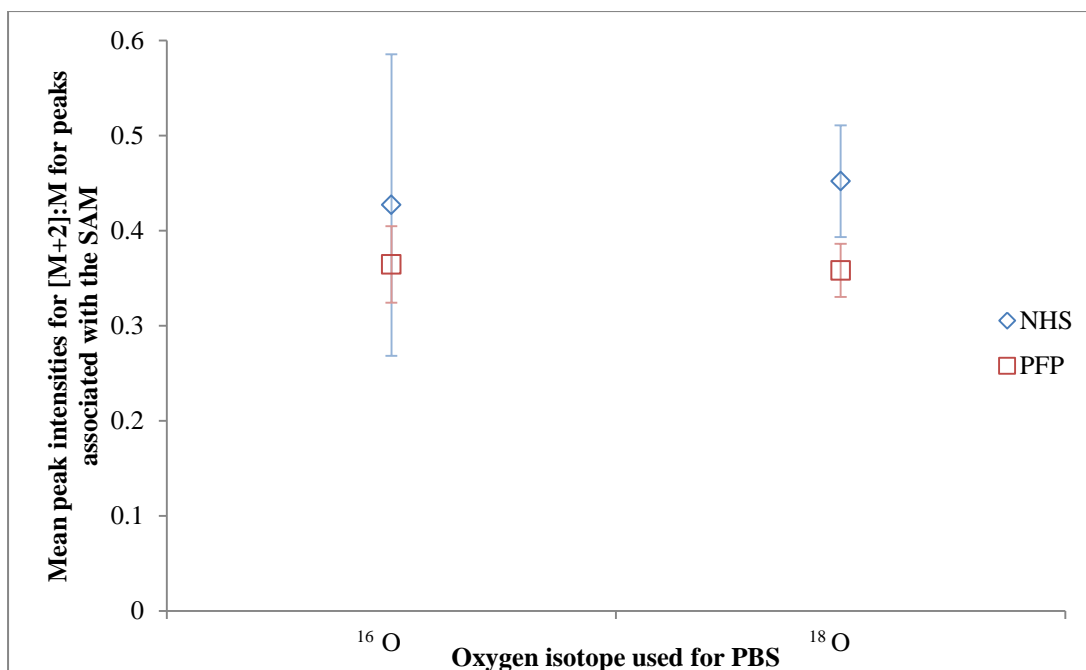


Figure 2.25: Plot showing the mean peak intensity ratio for [M+2]:M for peaks associated with the SAM after capture of the peptide Phe-Gly-Gly on the activated SAM when using two different activating reagents (Error bars show one standard deviation either side of the mean) (number of replicates = 8)

As can be seen in Figure 2.24, there is little difference between the mean ratio of peak intensities for [M+2]:M after capture of Gly-Gly-Gly on the activated SAM for wells spotted with Gly-Gly-Gly in unlabelled PBS and H_2^{18}O -containing PBS when PFP was used for activation. There is more of a difference in the mean ratio of peak intensities for [M+2]:M after capture of Gly-Gly-Gly on the activated SAM for wells spotted with Gly-Gly-Gly in unlabelled PBS and H_2^{18}O -containing PBS when NHS was used for activation. For NHS, the mean ratio of peak intensities for [M+2]:M is higher after capture of Gly-Gly-Gly on the activated SAM for wells spotted with Gly-Gly-Gly in H_2^{18}O -containing PBS than in unlabelled buffer.

As can be seen in Figure 2.25, there is little difference between the mean ratio of peak intensities for [M+2]:M after capture of Phe-Gly-Gly on the activated SAM for wells spotted with Phe-Gly-Gly in unlabelled PBS and H_2^{18}O -containing PBS when NHS and PFP were used for activation. As was observed in Figure 2.24 with NHS, the mean ratio of peak intensities for [M+2]:M is higher after capture of Phe-Gly-Gly on the activated SAM for wells spotted with Phe-Gly-Gly in H_2^{18}O -containing PBS than in unlabelled buffer.

The error bars for the sets of data overlap, and the values are comparable. Having examined the data, it is clear they are not normally distributed, and therefore a Mann-Whitney two-tailed U-test was performed on the data, to assess whether there was any

statistical significance to the differences observed between the ^{16}O and ^{18}O values obtained for NHS and PFP for the two peptides.

For Gly-Gly-Gly: For NHS (Mann-Whitney, $U = 14$, $n_1 = n_2 = 8$, $P < 0.05$, two-tailed) and PFP (Mann-Whitney, $U = 31$, $n_1 = n_2 = 8$, $P < 0.05$, two-tailed), the results are not statistically significant.

For Phe-Gly-Gly: For NHS (Mann-Whitney, $U = 16.5$, $n_1 = n_2 = 8$, $P < 0.05$, two-tailed) and PFP (Mann-Whitney, $U = 16.5$, $n_1 = n_2 = 8$, $P < 0.05$, two-tailed) the results are not statistically significant.

Hydrolysis of the activated carboxylic acid groups during Gly-Gly-Gly and Phe-Gly-Gly capture was thus not detectable mass spectrometrically. Therefore, as was seen with 4-BrPhe coupling (Section 2.6.2), hydrolysis of the activated SAM during Gly-Gly-Gly and Phe-Gly-Gly capture is not the reason for peaks being observed for unreacted SAM in spectra obtained after Gly-Gly-Gly and Phe-Gly-Gly had been spotted on wells containing activated SAM. This suggests that the reason for peaks being observed for unreacted SAM is due to inefficient capture of the peptides on the activated SAM.

2.7. Conclusions and future work

A variety of peptides have been captured on the thiol-derived SAM, with and without the presence of the chemical tag 4-BrPhe. Levels of peptide capture were lower for peptides captured on the activated SAM than for peptides captured on the activated 4-BrPhe-tagged SAM. This could be due to low levels of incorporation of the 4-BrPhe prior to peptide capture, inefficient activation of the SAM (and subsequent activated 4-BrPhe-tagged SAM) or simply an accumulation of inefficient chemistry at the different stages of the reaction.

Peaks for unreacted SAM were observed in mass spectra obtained after 4-BrPhe had been coupled to the activated SAM and also when peptide had been captured on the activated SAM. An investigation into the origin of these peaks was undertaken. This involved the use of H_2^{18}O -containing PBS, to determine whether hydrolysis of the activated carboxylic acid groups was occurring. However, it was found that the reason for the presence of peaks associated with unreacted SAM was not due to hydrolysis of the activated carboxylic acid groups. Therefore, the assumption was that the chemistry is not efficient. This may be due to the activation of the carboxylic acid groups not being efficient, or the capture of the 4-BrPhe or peptide being inefficient.

Peptides with aromatic N-terminal amino acids could be used to test the hypothesis that π -stacking is occurring. A variety of peptides with differing numbers of amino acids, with an

aromatic N-terminal amino acid, and their equivalent with a non-aromatic N-terminal amino acid should be used for peptide capture. This would allow a comparison between the capture efficiency of two peptides equivalent except for their N-terminal amino acids. If the capture efficiency was lower in all instances for peptides containing the aromatic N-terminal amino acid, then the theory that π -stacking was occurring could be supported. If the issue is one of steric hindrance, it would be sensible to use peptides containing a non-aromatic, but bulky, N-terminal side chain. This would allow a true conclusion to be made as to whether steric hindrance was the over-riding reason for inefficient peptide capture, or whether the presence of the aromatic side chain is the reason for inefficient peptide capture.

It is highly likely that the carboxylic acid head groups are not exposed as depicted in Figure 2.4, and are more likely to be buried within the main structure of the SAM. This would restrict their accessibility to the 4-BrPhe/peptide. The surface of the SAM could be analysed by a technique such as near-edge X-ray absorption fine structure (NEXAFS).¹⁸⁶ Nelson *et al.* published a paper in which they used NEXAFS to determine the long-range chain alignment within alkylthiolate monolayers.¹⁸⁶ NEXAFS provides structural information on the electronic structure and orientation of molecules or molecular fragments, by probing the absorption of electromagnetic radiation by excitation of core electrons into unoccupied bound or continuum states.¹⁸⁷ NEXAFS could be used to determine the structure of the SAM on the surface of the gold-coated MALDI chips, to determine whether the carboxylic acid groups are buried within the main structure of the SAM, which would limit the ability of the carboxylic acid groups to react with 4-BrPhe or a peptide. However, it is likely that it is not a single effect which is causing the presence of peaks associated with unreacted SAM after peptide capture, but a combination of a range of different effects.

Alternative activation and coupling chemistry may improve the activation of the SAM prior to capture of a chemical tag or peptide coupling. The alternative activation reagent NHSS was tested, to see whether a better solubility match between activation reagent and solvent would improve capture efficiency, as EDC (which is used in conjunction with the NHS and PFP) is more soluble in solvents such as acetonitrile, as opposed to DMF (which is used to solubilise the NHS and PFP). However, there were unassignable peaks observed in the mass spectra recorded after peptide capture when an acetonitrile/water solvent system was used. This negated the theoretical advantage of using a solvent which was more compatible with the EDC. There are other alternative activation and coupling chemistries which could be tested. One such example would be the conversion of the carboxylic acid groups to acid chloride groups. An acid chloride would be more susceptible to reaction with the primary amine of the 4-BrPhe/peptide than the carboxylic acid. The conversion of the carboxylic acid groups to acid chlorides was tested (using oxalyl chloride to convert the carboxylic acid groups to acid chloride groups), but mass spectra were obtained with a high S/N, and no peaks were observed for peptide captured on the acid chloride-derived SAM.

Another alternative which was briefly explored was the conversion of the carboxylic acid group to a ketene. This was based on the work by Chan *et al.* in 2012, where they modified the N-terminal of peptides and proteins using an alkyne-functionalised ketene.¹⁸⁸ However, when the SAM was derivatised to form the ketene, the resultant mass spectra were very noisy, and no peaks were observed for peptide captured on the ketene-derived SAM.

The spectra obtained after conversion of the carboxylic acid groups on the gold-coated MALDI chip were very noisy. This could have been due to unexpected chemistry occurring when the gold-coated MALDI chip with the SAM formed on the wells was placed in the reagents and solvents. The possibility of converting the carboxylic acid groups of the thiols to a more reactive group (such as the acid chloride or ketene) in solution, prior to SAM formation on the gold-coated MALDI chip could be explored. If the conversion of the carboxylic acid groups was to be performed prior to SAM formation, the possibility that some of the acid chloride or ketene could be converted back to the carboxylic acid during SAM formation would also need to be explored.

Another avenue would be to investigate alternative components to form a SAM (such as the pre-activated SAM discussed in Chapter 3). Another example would be to use the disulfide component used by Peelen *et al.* in 2006 for the capture of mammalian cells by cell surface receptor binding to a ligand immobilised on gold thin films.¹⁸⁹ The group used an aldehyde-terminated disulfide ((OHC-(CH₂)₉-S-)₂), to form a SAM on a gold-coated glass substrate. They then immobilised proteins on the aldehyde-terminated gold surfaces. This involved reacting a primary amine with the aldehyde, and then subsequent reduction of the imine using sodium cyanoborohydride. Aldehyde-terminated disulfides with PEG groups within the chains are also available (e.g. (-S-(CH₂)₁₁-(OCH₂CH₂)_nOCH₂CHO)₂). Aldehyde-terminated disulfides could potentially be used for the capture of peptides on a SAM for subsequent MALDI-MS analysis. A suitable spacer molecule would be needed, for example, (-S-(CH₂)₁₁-(OCH₂CH₂)_n-OH)₂ for combination with the aldehyde-terminated disulfide containing PEG groups.

Chapter 3: Disulfide- derived SAM

Throughout Chapter 3, the following terminology will be used to describe the SAM at the different stages of the peptide capture methodology.

Terminology	Structure
Disulfide-derived SAM	$\begin{array}{l} \text{S} - (\text{CH}_2)_2 - \text{EG}_6 - \text{OCH}_3 \\ \\ \text{S} - (\text{CH}_2)_2 - \text{EG}_8 - \text{C(O)O} - \text{NHS} \end{array}$
Disulfide-derived SAM containing NHS by-product	$\begin{array}{l} \text{S} - (\text{CH}_2)_2 - \text{EG}_6 - \text{OCH}_3 \\ \\ \text{S} - (\text{CH}_2)_2 - \text{EG}_8 - \text{C(O)O} - \text{N(Cy)C(O)NHCy} \end{array}$
Peptide captured on the disulfide-derived SAM	$\begin{array}{l} \text{S} - (\text{CH}_2)_2 - \text{EG}_6 - \text{OCH}_3 \\ \\ \text{S} - (\text{CH}_2)_2 - \text{EG}_8 - \text{C(O)O} - \text{Peptide} \end{array}$
4-BrPhe captured on the disulfide-derived SAM	$\begin{array}{l} \text{S} - (\text{CH}_2)_2 - \text{EG}_6 - \text{OCH}_3 \\ \\ \text{S} - (\text{CH}_2)_2 - \text{EG}_8 - \text{C(O)O} - (\text{Br})\text{Phe} \end{array}$
Phe captured on the disulfide-derived SAM	$\begin{array}{l} \text{S} - (\text{CH}_2)_2 - \text{EG}_6 - \text{OCH}_3 \\ \\ \text{S} - (\text{CH}_2)_2 - \text{EG}_8 - \text{C(O)O} - \text{Phe} \end{array}$

3.1. Introduction

Peptide capture on the thiol-derived SAM (Chapter 2) had limited success, as peaks associated with unreacted SAM were observed in spectra obtained after peptide capture. Therefore, alternative SAM components were investigated, in an attempt to improve SAM activation and peptide capture, and minimise the number of synthetic steps in construction on the SAM-peptide structure.

An alternative SAM component was a pre-activated disulfide (Figure 3.1). If it was the activation step (and not the peptide capture step) that was the problem when attempting to capture peptides on the thiol-derived SAM, the pre-activated disulfide approach removes the need to activate the SAM prior to capture of the tag or peptide (which also removes one on-chip synthesis step).

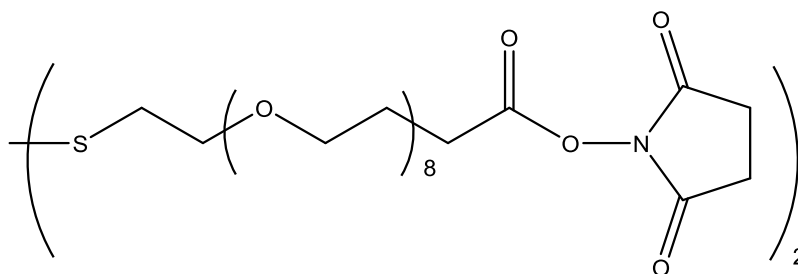


Figure 3.1: NHS ester disulfide

There have been several studies which have compared the reactivity of thiols and disulfides with a solid substrate.^{160,161,190} These studies concluded that thiols form a SAM on the substrate at a faster rate than disulfides. Bain *et al.* found that in a solution containing mixtures of a thiol and a disulfide, adsorption of the thiol was strongly preferred, by a factor of approximately 75:1.¹⁶⁰ Therefore, an mPEG disulfide spacer molecule (Figure 3.2) was used in combination with the NHS ester disulfide for construction of the SAM (to avoid preferential adsorption of spacer molecules when used with a disulfide NHS ester).

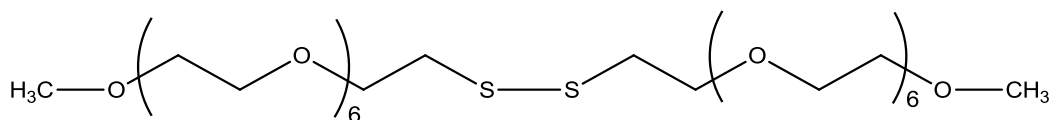


Figure 3.2: mPEG disulfide

3.1.1. Aims and overview

The aim of the work in this chapter was to investigate the suitability of a disulfide-derived SAM for peptide capture, prior to mass spectrometric analysis. A pre-activated SAM was formed on the gold surface and a variety of peptides captured at two levels (5 nmol and 0.05 nmol of peptide per well).

In Chapter 2, when the chemical tag 4-BrPhe was coupled to the thiol-derived SAM, peaks were observed in the mass spectra which corresponded to unreacted SAM. Coupling of the same 4-BrPhe chemical tag to the disulfide-derived SAM was thus attempted at the 5 nmol and 0.05 nmol level in order to determine whether more efficient coupling of the tag could be achieved using the disulfide-derived SAM.

An investigation into whether there was a problem with Phe being located at the N-terminus of the species (i.e. 4-BrPhe or peptide) to be captured on the disulfide-derived SAM was also undertaken.

3.2. Formation of a disulfide-derived SAM

A gold-coated MALDI chip was cleaned with piranha solution prior to SAM formation. The two components used for forming the SAM were both disulfides containing PEG spacers and either a methyl group ($((\text{SCH}_2\text{CH}_2(\text{OCH}_2\text{CH}_2)_6\text{OCH}_3)_2)$) or NHS ester ($((\text{S}(\text{CH}_2\text{CH}_2\text{O})_8\text{CH}_2\text{CH}_2\text{COONC}_4\text{O}_2\text{H}_4)_2)$) as their terminal group. The SAM components were each dissolved in DMSO at a concentration of 0.1 mg/mL, and a 1:4 molar ratio of NHS ester:mPEG terminated groups was created from these stock solutions. Aliquots (0.4 μL) of this mixture were then spotted onto wells of the gold-coated MALDI chip and the SAM was left to assemble overnight at ambient temperature (Figure 3.3). The gold-coated MALDI chip was washed with acetonitrile and dried under a stream of nitrogen.

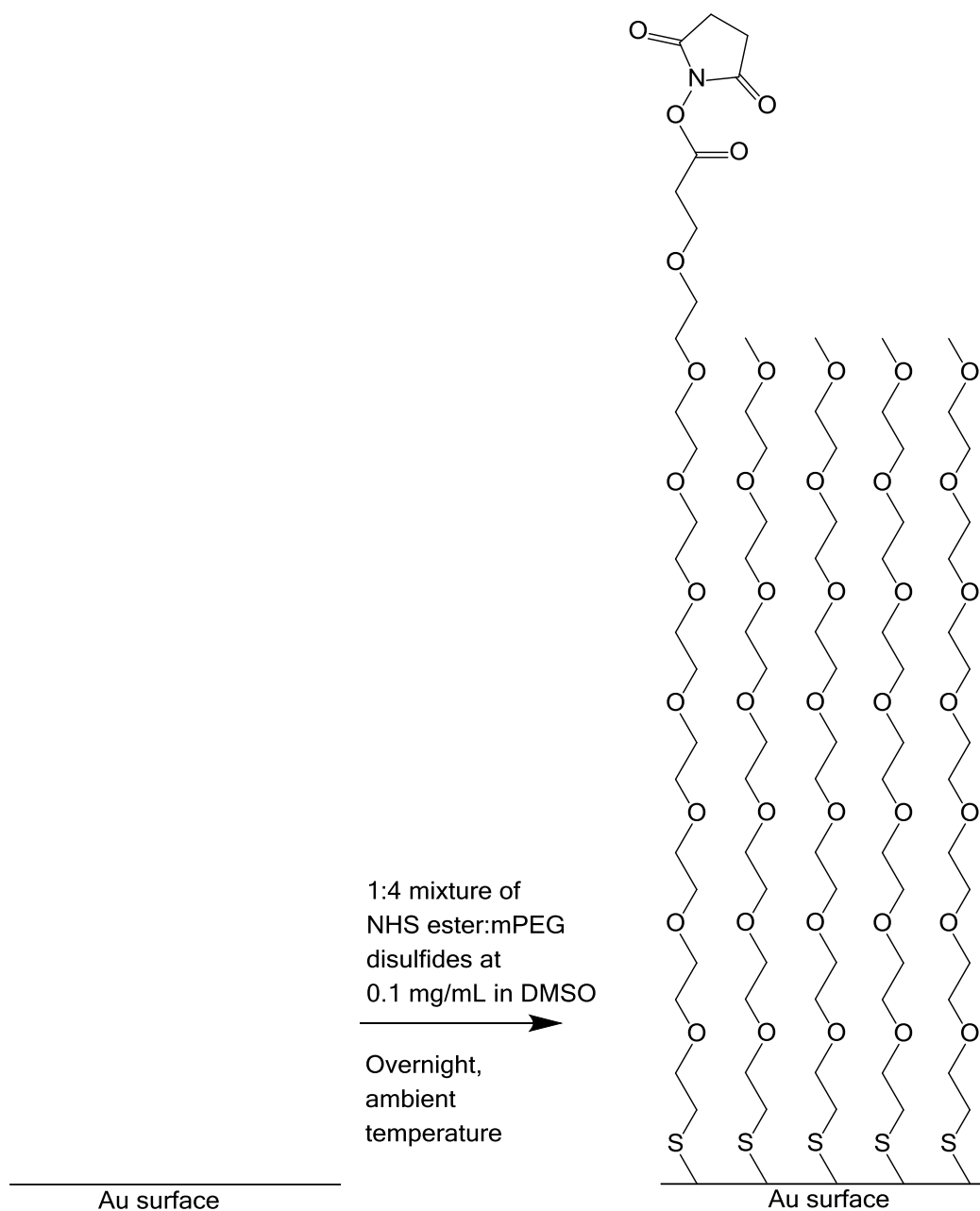


Figure 3.3: Formation of the disulfide-derived SAM.

Aliquots (0.3 μL) of THAP solution were spotted on wells, and the formation of the SAM was assessed by MALDI-MS. A typical MALDI mass spectrum is shown in Figure 3.4 (peak assignments shown in Table 3.1).

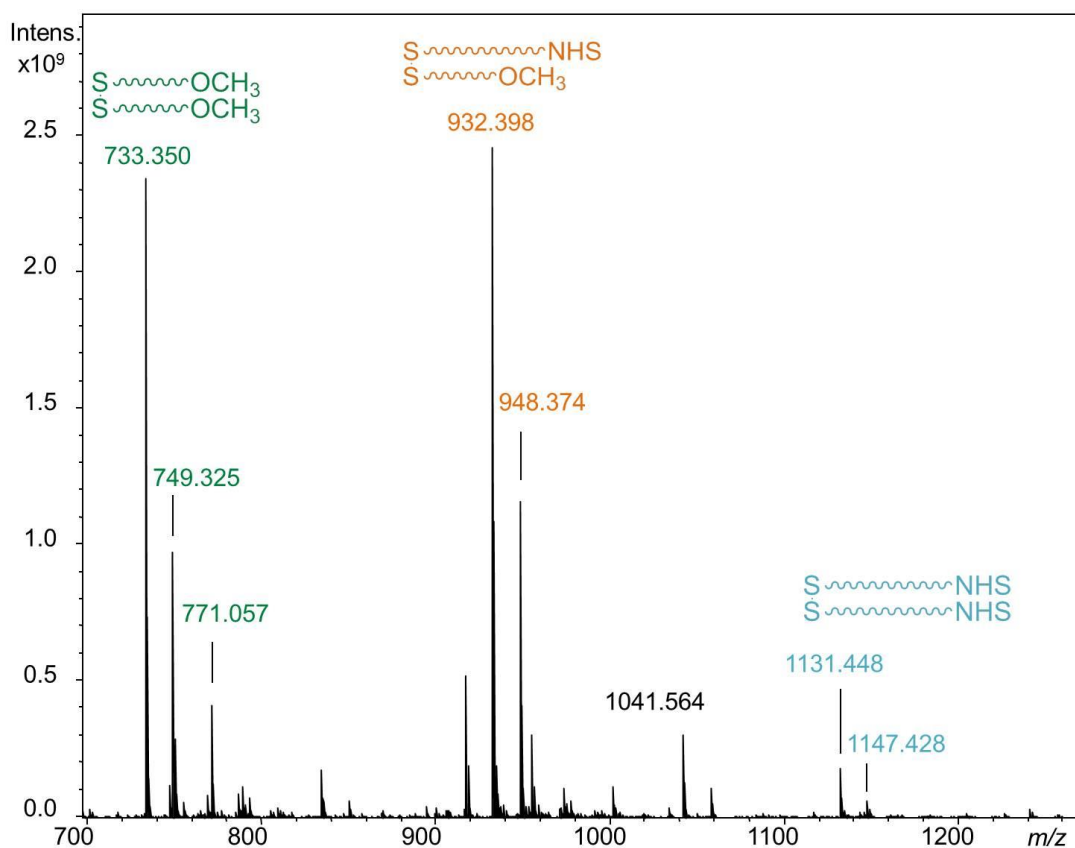


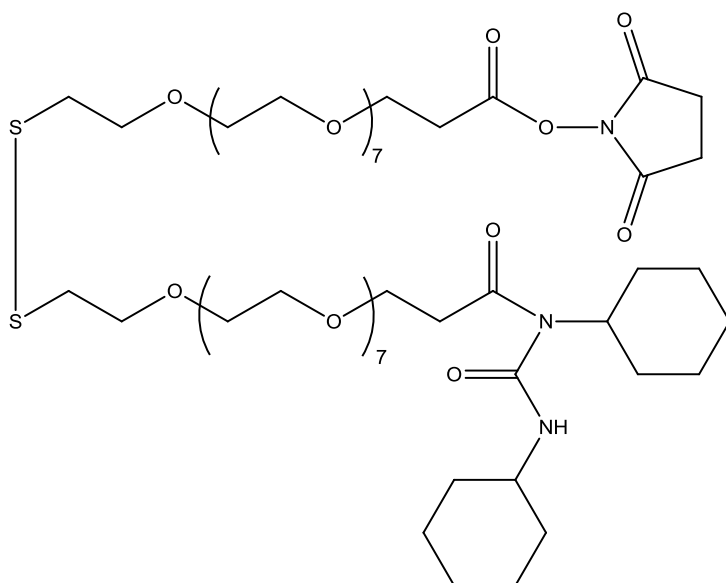
Figure 3.4: MALDI mass spectrum of the disulfide-derived SAM formed on a gold-coated MALDI chip

As can be seen in Figure 3.4, a SAM was formed on the gold-coated MALDI chip from the mixture of the two disulfides. Peaks can be seen which are associated with a disulfide having two mPEG termini, a disulfide having two NHS ester termini and a mixed disulfide having one mPEG terminus and one NHS ester terminus (Table 3.1). There is also a peak at m/z 1041 ($[M+Na]^+$) which was discovered to be due to the presence of a by-product in the commercial NHS ester disulfide.

Table 3.1: Peak assignments for mass spectrum of disulfide-derived SAM formation on a gold-coated MALDI chip

Nominal m/z value	Assignment
733	$[M+Na]^+$ - 2 mPEG termini
749	$[M+K]^+$ - 2 mPEG termini
771	$[M+Na+K-H]^+$ - 2 mPEG termini
932	$[M+Na]^+$ - 1 NHS ester terminus, 1 mPEG terminus
948	$[M+K]^+$ - 1 NHS ester terminus, 1 mPEG terminus
1041	$[M+Na]^+$ - 1 NHS ester by-product terminus, 1 mPEG terminus
1131	$[M+Na]^+$ - 2 NHS ester termini
1147	$[M+K]^+$ - 2 NHS ester termini

Polypure (the supplier) was contacted over the presence of the peak observed at m/z 1041. The information provided by Polypure highlighted the presence of a by-product in the commercial NHS ester (Figure 3.5).

**Figure 3.5: NHS ester by-product (containing a stable *N*-acyl urea)**

In order to confirm whether the by-product in the commercial NHS ester was contributing to the presence of the peak at m/z 1041, product ion analysis was performed. The ion generating the peak at m/z 1041 was found to be due to a mixed disulfide having one mPEG terminus and the stable *N*-acyl urea half of one by-product NHS ester terminus.

Based on information provided by Polypure, it is likely that the by-product is formed during the conversion of the carboxylic acid starting material to the NHS ester (Figure 3.6). The

carboxylic acid groups are activated using *N,N'*-dicyclohexylcarbodiimide (DCC). This produces an *O*-acyl isourea intermediate. The NHS ester then reacts with the activated carboxylic acid to form the NHS ester and dicyclohexylurea (Figure 3.6 – Activation route). The contaminant is formed by rearrangement of the *O*-acyl isourea intermediate to form the more stable *N*-acyl urea product (Figure 3.6 – Rearrangement route).

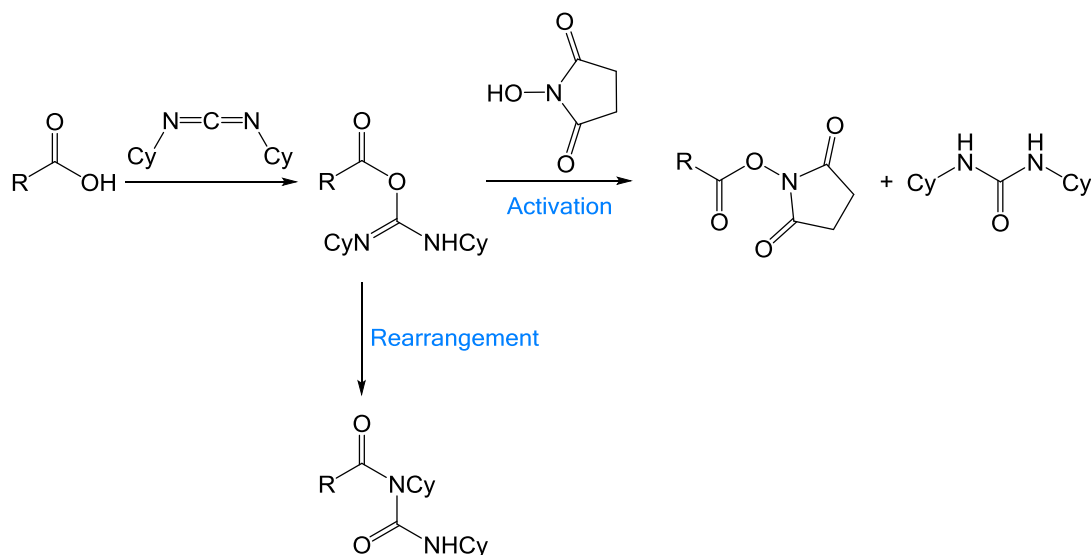


Figure 3.6: Carboxylic acid reacting with DCC to form *O*-acyl isourea intermediate and subsequent rearrangement or substitution (R = rest of the disulfide, Cy = cyclohexyl)

The *N*-acyl urea by-product is stable, and will not go on to react with NHS to form NHS ester groups at the termini of the disulfide (unlike the *O*-acyl isourea). The presence of the by-product reduces the amount of NHS groups available for reaction with a peptide, and the *N*-acyl urea will also not go on to react with a peptide. Therefore, the amount of reactive groups available on the surface of the SAM is less than expected. This could be accounted for when preparing the SAM, by including a higher amount of the NHS ester in the disulfide mixture prior to spotting the SAM components on the gold-coated MALDI chip.

3.3. Peptide capture on SAM

This work was done in collaboration with an MChem student (Rachel Bates). In previous work (Chapter 2), the peptide capture step on the activated SAM was carried out over 18 h. This was based on the method developed by the Flitsch group in Manchester.¹⁷⁰ However, there was no direct evidence that an overnight coupling reaction was necessary. In fact, the reaction between an NHS ester and a primary amine is expected to occur rapidly. The reaction of an NHS ester with an amine is typically performed between pH 7 and 9, and is completed between 30 minutes and 2 hours at ambient temperature.¹⁹¹ This suggests that

the reaction between the peptide and the SAM may not require overnight incubation. A peptide capture reaction of one hour on the disulfide-derived SAM was tested (generic reaction showing peptide capture shown in Figure 3.7). One hour was originally chosen as a compromise due to the fact that it lies between 30 minutes and 2 hours, to determine whether peptide capture was observed at all after one hour.

Stock solutions of the peptides shown in Table 3.2 were prepared at 50 mM in PBS. From these stock solutions, solutions were prepared to enable peptide capture on the disulfide-derived SAM with the peptides at two different amounts in 1 μL aliquots: 5 nmol and 50 pmol. Aliquots (1 μL) of the peptide solution were spotted on wells and left for one hour at ambient temperature. The gold-coated MALDI chip was washed with acetonitrile and dried under a stream of nitrogen.

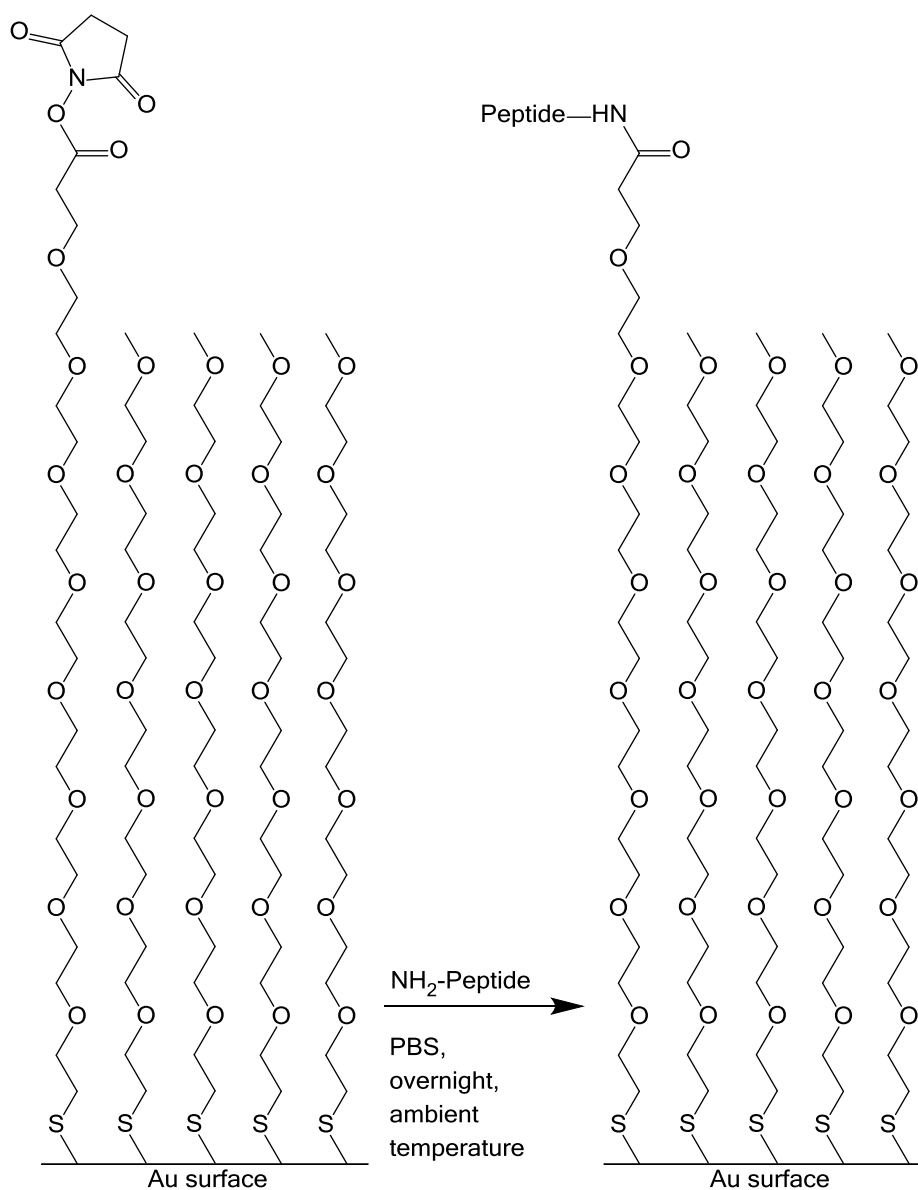


Figure 3.7: Generic reaction of a peptide with the disulfide-derived pre-activated SAM

Aliquots (0.3 μL) of THAP solution were spotted on wells, and the success of the capture of the peptides detailed in Table 3.2 was assessed by MALDI-MS. Table 3.2 shows the expected m/z values for the peptides captured on the SAM.

Table 3.2: Peptides spotted on the SAM and expected m/z values of product captured on the SAM

Peptide	Molecular mass of peptide / g mol^{-1}	Expected values of SAM-peptide signal ($[\text{M}+\text{Na}]^+$ and $[\text{M}+2\text{Na}-\text{H}]^+$) / m/z
Leu-Gly-Gly (LGG)	245	1062 / 1084
Gly-Gly-Val (GGV)	231	1048 / 1070
Ala-Ala-Ala (AAA)	231	1048 / 1070
Gly-Gly-Gly (GGG)	189	1006 / 1028
Phe-Gly-Gly (FGG)	279	1096 / 1118
Leucine enkephalin (Leu-Enk)	555	1372 / 1394

Figure 3.8 shows a MALDI mass spectrum obtained following reaction for one hour of leucine enkephalin at 5 nmol with the disulfide-derived SAM (peak assignments shown in Table 3.3).

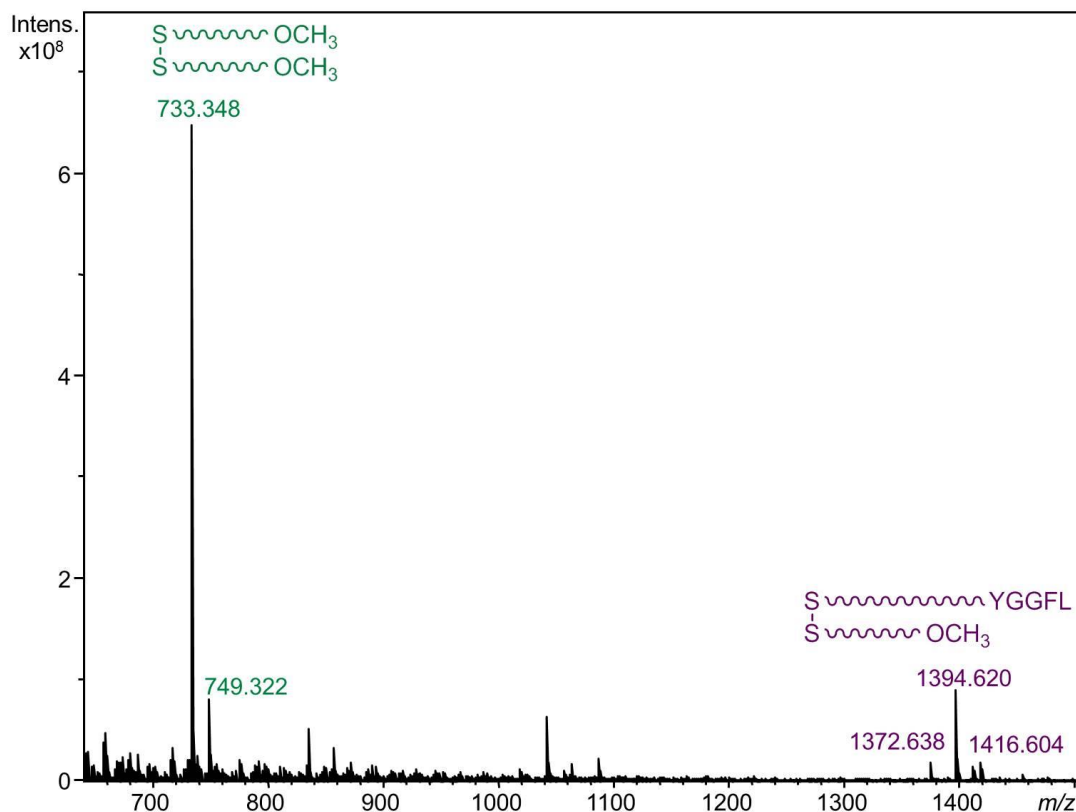


Figure 3.8: MALDI mass spectrum obtained following reaction for one hour of leucine enkephalin at 5 nmol with the disulfide-derived SAM

Peaks were observed for species containing leucine enkephalin captured on the SAM (m/z 1372, 1394 and 1416). No peaks were observed corresponding to species bearing unreacted NHS groups (m/z 932, 948, 1131 and 1147).

Table 3.3: Peak assignments for MALDI mass spectrum of leucine enkephalin captured on the disulfide-derived SAM

Nominal m/z value	Assignment
733	$[M+Na]^+$ - 2 mPEG termini
749	$[M+K]^+$ - 2 mPEG termini
1372	$[M+Na]^+$ - 1 mPEG terminus, 1 CO-Leu-Enk terminus
1394	$[M+2Na-H]^+$ - 1 mPEG terminus, 1 CO-Leu-Enk terminus
1416	$[M+Na+K-H]^+$ - 1 mPEG terminus, 1 CO-Leu-Enk terminus

The reduced incubation time of one hour for capture of a peptide on the NHS ester of the SAM was shown to be successful for all peptides tested. This means that one of the overnight reaction steps can be removed from our protocol, which is convenient in reducing

the overall sample handling time. This considerable time-saving for the workflow could potentially mean that the entire process (from SAM formation to peptide capture and analysis) could take place in two days. The formation of the SAM would remain an overnight step based on literature evidence that in order to ensure SAM formation, the disulfide solution should be left in contact with the gold-coated MALDI chip for at least 12 hours.^{160,190} Although one overnight step would still be needed to form the SAM, the protein digest protocol used in this work (Chapter 7) is also an overnight reaction. Therefore, the SAM could be left to form overnight simultaneously with the protein being digested.

Having established that peptide capture was observed after one hour, a time course study was performed, with five minute intervals used between 5 and 60 minutes, in order to determine whether one hour was still an excessive amount of time for peptide capture. The time course showed that a time of between 40 and 60 minutes was optimum for the peptide capture reaction (depending on the peptide). This was measured by analysis of the mass spectra obtained after peptide capture for the specified times, and comparing the relative intensities of peaks due to SAM species bearing unreacted NHS groups and peaks due to peptide captured on the SAM. After 40-60 minutes, there was a decrease observed in the relative amounts of SAM species bearing unreacted NHS groups and peptide captured on the SAM, possibly due to a deterioration of the SAM on the surface of the gold-coated MALDI chip over time. For most peptides, the mass spectra obtained following peptide capture showed a combination of peaks. In some instances, peaks were seen for peptide captured on the SAM and also for species bearing unreacted NHS groups. In other instances, peaks were observed for peptide captured on the SAM, but no peaks were observed for species bearing unreacted NHS groups. In both instances, no peaks were observed for hydrolysis of the NHS esters to carboxylic acid groups. This variability could just be a reflection of the innate variability of MALDI.¹⁹² It can be difficult to obtain reproducible results using MALDI, mainly due to the fact that ion signals produced are not reproducible from sample to sample or from spot to spot, and shot to shot analysis of the same spot can produce varying results.¹⁹²⁻¹⁹⁴ Part of the reason for this may be due to inhomogeneity of the sample, due to the way that the matrix crystallises on the surface of the gold-coated MALDI chip.

Power analysis can be used to determine the sample size required to detect an effect of a given size with a given degree of confidence.¹⁹⁵ The power of an experiment is the probability that it can detect a treatment effect, if it is present. Statistical power analysis exploits the mathematical relationship between power, sample size, significance criterion and the population effect size.¹⁹⁵ There are examples of power analysis being applied to mass spectrometric proteomics analysis. These include the use of power analysis to determine the smallest fold-change reliably detected for a complex peptide mixture.¹⁹⁶ In 2011, Levin published a paper detailing the role of statistical power analysis in quantitative

proteomics.¹⁹⁷ The author sought to establish the importance of power analysis to calculate the number of biological replicates required for a study. In 2007, Meng *et al.* published a paper detailing LC-MS analysis of complex peptide mixtures, where six standard peptides were spiked into a processed plasma background.¹⁹⁶ The authors used a power analysis to show that with a sample size of 12, a 1.5-fold difference was expected to be detected about 90% of the time. However, in the instance of the work in this thesis, a qualitative comparison has been made, and therefore, a power analysis is not appropriate.

There have been various publications in recent years which have looked into the variability observed during MALDI-MS analysis. In 2013, Hu *et al.* published a paper discussing the coffee-ring effect.¹⁷⁶ Here, the authors discussed the heterogeneous distribution of molecules within dry deposits of suspensions and solutions, which can compromise analytical performance and reproducibility.

One of the compromises of the SAM technology is being unable to mix the analyte (the SAM) and the matrix prior to spotting the two on the gold-coated MALDI chip. Therefore, a homogeneous solution cannot be formed prior to spotting. It is possible that the matrix may be crystallising heterogeneously over the surface of the SAM, which could affect the quality of the spectra obtained.

It has been shown over the years that sweet spots occur within a spot on a MALDI plate¹⁹⁸ (or chip), and that coffee-rings can also occur.¹⁷⁶ Both of these phenomena can compromise the quality of the mass spectra obtained, and lead to inter- and intra-spectral variability. These MALDI mass spectral effects could have affected the spectra obtained, resulting in differences obtained for the same peptide spotted on different wells of the SAM.

Alternatively, the variability may derive from the protocol itself. This could be a problem with the SAM formation. A 1:4 mixture of the mPEG:NHS disulfides was spotted on the wells, and so it is tempting to assume that the distribution of species across the wells is random, as depicted in Figure 3.3. There is no guarantee that SAMs formed on different wells of the same chip would be identical. Within a mixed SAM, separation of the components can occur.¹⁹⁹ Therefore, there could be regions which have a high density of NHS-bearing species, with very little of the spacer component, and vice versa. If areas exist where NHS groups are packed closely together with few spacer molecules, peptide capture could be compromised as it may not be possible for peptide molecules to access and react with all of the NHS groups due to steric hindrance (hence peaks would be observed for peptide captured on the SAM and also for species bearing unreacted NHS groups). Spectra in which no peaks were observed for species bearing unreacted NHS groups may arise if all the NHS groups were sufficiently separated by the spacer molecules to make them all accessible for peptide capture (hence peaks would only be observed for species bearing peptide captured on the SAM).

The presence of peaks due to different scenarios (peaks associated with peptide captured on the SAM, peaks associated with no peptide capture and no peaks associated with SAM or peptide captured on the SAM) for the same peptide, is likely to be a combination of the two possible sources of variability: innate MALDI variability and protocol variability. Surface analysis of the SAM on the gold-coated MALDI chip would provide information on the surface composition of the SAM. STM has been used previously to demonstrate phase segregation in mixed-component SAMs.¹⁹⁹ Stranick *et al.* showed clear differences between the STM images obtained from SAMs comprising a single component, and SAMs comprising two components.¹⁹⁹ The image produced from the mixed SAM (comprising a methyl terminated SAM and an ester terminated SAM) showed definitive areas for the two components, showing that the two components had segregated in areas to form single-component domains. It is possible that a similar phenomenon is occurring for the disulfide-derived SAM used in this work.

The mass spectra obtained for each peptide were examined for the presence of peaks due to species associated with peptide captured on the SAM and species bearing unreacted NHS groups. The percentage of spectra (twelve spectra from twelve wells) where peaks were observed for species associated with peptide captured on the SAM and species bearing unreacted NHS groups was determined (Table 3.4).

Table 3.4: Table showing peptides spotted at 5 and 0.05 nmol levels on the disulfide-derived SAM, and percentage of spectra with species for peptide capture and unreacted SAM (number of replicates = 12)

Peptide	Amount spotted on wells / nmol	Peaks seen for peptide captured on SAM?	Percentage ^v of spectra with signal for captured peptide	Peaks seen for unreacted SAM?	Percentage ^v of spectra with signal for unreacted SAM
AAA	5	Yes	20%	Yes	90%
	0.05	Yes	100%	Yes	10%
LGG	5	Yes	58%	Yes	50%
	0.05	Yes	100%	Yes	64%
GGV	5	Yes	75%	Yes	25%
	0.05	Yes	100%	Yes	45%
GGG	5	Yes	100%	No	0%
	0.05	Yes	75%	Yes	92%
FGG	5	Yes	92%	Yes	8%
	0.05	Yes	17%	Yes	100%
Leu-Enk	5	Yes	100%	Yes	8%
	0.05	Yes	50%	Yes	100%

As can be seen in Table 3.4, the different peptides showed differing values for both the percentage of spectra with signals for peptide captured on the SAM, and for the percentage of spectra with signals for species bearing unreacted NHS groups.

For each peptide, multiple experiments were performed, using a number of gold-coated MALDI chips. Due to the fact that multiple chips were used for the peptide capture experiments, multiple batches of the disulfide-derived SAM components were prepared in solution and spotted on the gold-coated MALDI chips. Therefore, if there were problems with one solution of the disulfide-derived components, this should have been noticeable when compared with the results of other experiments that used the same peptide.

^v One mass spectrum recorded from each well. Data extracted from spectra for relevant peaks (e.g. species bearing unreacted NHS groups). Percentage values reported as a percentage of the total number of spectra recorded for each peptide at 5 nmol and 0.05 nmol amounts spotted on the SAM.

When the peptide Phe-Gly-Gly was spotted on the SAM, peptide capture appeared to be more successful when 5 nmol of peptide was reacted with the SAM than when 0.05 nmol of peptide was reacted with the SAM. The percentage of spectra which had peaks for peptide captured on the SAM was higher from the wells spotted with 5 nmol of peptide (92%) than wells spotted with 0.05 nmol of peptide (17%). The peptide Gly-Gly-Gly, which differs from Phe-Gly-Gly only by the presence of Gly at the N-terminus in place of Phe, showed a much higher percentage of spectra with peaks associated with peptide capture on the disulfide-derived SAM at 0.05 nmol (75%) compared to Phe-Gly-Gly. This suggests that the presence of the phenylalanine residue may be reducing the success of peptide capture on the disulfide-derived SAM. The percentage of spectra with signals for unreacted SAM was also higher for Phe-Gly-Gly reacting with the disulfide-derived SAM when compared to Gly-Gly-Gly reacting with the disulfide-derived SAM at both the 5 nmol and 0.05 nmol levels. The lower percentage of spectra with peaks associated with Phe-Gly-Gly capture on the disulfide-derived SAM is in agreement with the results obtained when reacting 4-BrPhe and Phe-Gly-Gly with the thiol-derived SAM (Chapter 2), where peaks associated with unreacted SAM were observed in the mass spectra obtained after 4-BrPhe had been reacted with the SAM. The result obtained for Phe-Gly-Gly reacting with the disulfide-derived SAM suggests that the presence of Phe at the N-terminus may lead to inefficient capture of the peptide or amino acid residue.

When Phe-Gly-Gly was used for capture, the results showed better peptide capture at the higher level of 5 nmol of peptide when compared to reacting 0.05 nmol of peptide with the disulfide-derived SAM. The phenyl group makes the Phe residue hydrophobic. However, the peptide was dissolved in PBS, which is hydrophilic. Therefore, there is a mismatch in the properties of the solvent and the solute. There have been many studies which have investigated peptide aggregation in solvents.²⁰⁰⁻²⁰² It has been shown that peptides can form ordered aggregates in aqueous solution.²⁰⁰ Non-polar amino acid side chains are removed from water and packed together in the interior of the aggregate. This could be the case for Phe-Gly-Gly. The hydrophobic phenyls in the side chains of the Phe could be causing the peptides to aggregate, and pack the N-termini of the peptides inside the aggregate. This would mean that the N-termini would not be available to react with the SAM, which could have resulted in the lower capture efficiency observed with Phe-Gly-Gly.

Although the peptide Phe-Gly-Gly is small, Jayakumar *et al.* showed that ordered aggregates in aqueous solution could be formed by a tetrapeptide (Val-Leu-Pro-Phe).²⁰⁰ Therefore, the possibility that a tripeptide could form aggregates in aqueous solution is viable. The authors believed that the formation of aggregates is due to two processes: the positive entropy from water release due to intermolecular association of ionic moieties (between 5 °C and 40 °C) and intramolecular ionic interaction (above 40 °C). These processes are driven by the desire to remove non-polar amino acid side chains from water,

and pack them together in the interior of an aggregate. Jayakumar *et al.* monitored the formation of aggregates through fluorescence spectroscopy, where the critical micelle concentration for the peptide was determined. This technique could be applied to Phe-Gly-Gly due to the presence of the aromatic phenylalanine, which is naturally fluorescent. It would be possible to prepare solutions of the peptide Phe-Gly-Gly at varying concentrations, and subject these to fluorescence spectroscopic analysis, in order to determine whether aggregates were forming in solution (which would subsequently inhibit reaction with the SAM).

Alternatively, hydrophobic solvents could be used, to determine whether hydrophobic interactions occurred with the peptide in PBS. A hydrophobic solvent should interrupt/prevent aggregates forming, as there would be no need to protect the hydrophobic amino acid side chains from water.

The phenomenon of π -stacking in proteins is well established.^{203,204} π -stacking occurs when non-covalent interactions occur between aromatic rings, causing the aromatic rings to stack together. It is also possible that π -stacking of the aromatic side chains of the Phe amino acid could be occurring in solution. As with the theory that the hydrophobic Phe amino acids could be causing the formation of aggregates, the possibility of π -stacking of the aromatic side chain of the Phe amino acids would also result in aggregates being formed in solution. If aggregates were formed in solution, this would mean that the peptide N-terminus would not be available to react with the SAM, which would have resulted in the lower capture efficiency observed with Phe-Gly-Gly. This theory could be tested by using peptides with other aromatic terminal amino acids (e.g. those containing tyrosine or tryptophan). The same amounts of peptide should be spotted on wells (e.g. 5 nmol and 50 pmol), in order to provide a comparison with the behaviour of Phe-Gly-Gly. The extent of peptide capture on the disulfide-derived SAM could be calculated from the spectral data (in a similar manner to that used in Chapter 2) for the different peptides, in order to establish whether phenylalanine was causing the problem, or if it is a problem with all aromatic amino acid residues.

A study by Samal and Geckeler in 2001 showed unexpected solute aggregation in water on dilution.²⁰⁵ They looked at aggregation in aqueous solutions of fullerene-cyclodextrin conjugates, β -cyclodextrin, sodium chloride, sodium guanosine monophosphate and a DNA oligonucleotide. Their conclusion was that larger aggregates can exist in dilute aqueous solutions than in more concentrated solutions. For some peptides, peaks were observed for unreacted SAM after peptide capture. This theoretically, should not be the case, as 0.05 nmol is still above the amount of reactive groups on the surface of the SAM. Therefore, there is excess peptide, which should mean no unreacted SAM. For the peptide Gly-Gly-Gly, when 5 nmol of peptide was spotted on wells, 25% of spectra contained signals for unreacted SAM. When 50 pmol of Gly-Gly-Gly was spotted on wells, 45% of spectra contained signals for unreacted SAM. The fact that peaks are observed for unreacted SAM

in a higher percentage of spectra at the 50 pmol level than the 5 nmol level could imply a higher level of aggregation at the 50 pmol amount of peptide, which would mean that the peptides are less likely to react with the SAM.

The fact that peptide aggregates may be forming in the PBS suggests that PBS is not the most suitable solvent for the peptide capture step. PBS was used as a result of data obtained in previous work by the Flitsch group,¹⁷⁰ where there has been a history of use of PBS with the SAM components discussed in Chapter 2. With the possibility that this was not the most suitable solvent in mind, peptide capture on the SAM using 5% *N,N*-diisopropylethylamine (DIPEA) in DMF was attempted. However, no peaks were seen for peptide captured on the SAM and no further work was undertaken using DIPEA in DMF. A variety of other solvents were also investigated, including triethylammonium bicarbonate (TEAB), DMF, phosphate buffer, acetonitrile and water. However, none of these solvents proved to be more suitable than PBS for the reaction of a peptide with the disulfide-derived SAM. Therefore, its use was continued during this work.

Having established that peptide capture was possible at 50 pmol, peptide capture was attempted at 5 pmol for the peptides mentioned in Table 3.2. For each peptide, four wells were spotted with the relevant peptide solution. However, peptide capture was not observed for any of the peptides used. Therefore, the limit of detection for the disulfide-derived SAM is between 50 pmol and 5 pmol. This is quite a large amount of peptide, and is currently well above the amount which might be expected in a typical proteomics experiment.

3.3.1. MS/MS analysis of peptides captured on the disulfide-derived SAM

Product ion spectra were recorded in order to determine the structures of the species giving rise to the signals in the MALDI mass spectra obtained following peptide capture.

Product ion spectra obtained from the ions corresponding to $[M+2Na-H]^+$ for the peptides Gly-Gly-Val and leucine enkephalin captured on the SAM (m/z 1070 and 1394 respectively) are shown in Figure 3.9 and Figure 3.10 respectively.

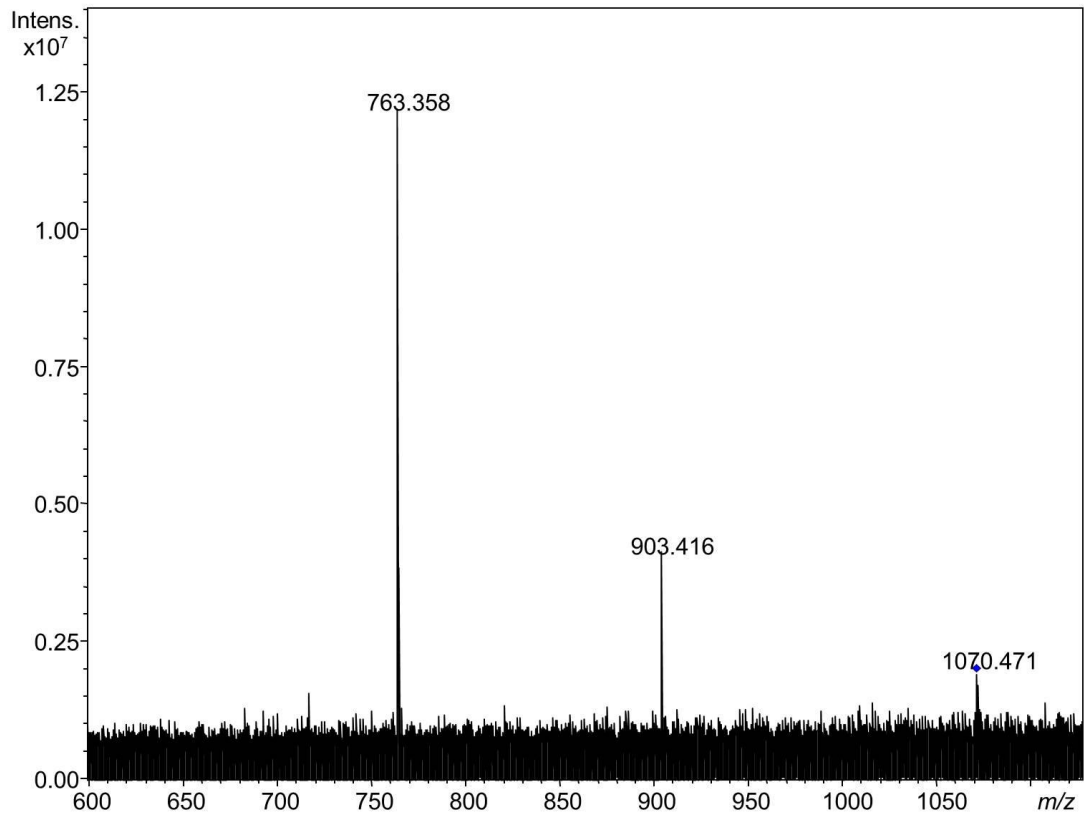


Figure 3.9: MALDI-product ion spectrum of $[M+2Na-H]^+$ for capture of the peptide Gly-Gly-Val on the SAM (precursor at m/z 1070)

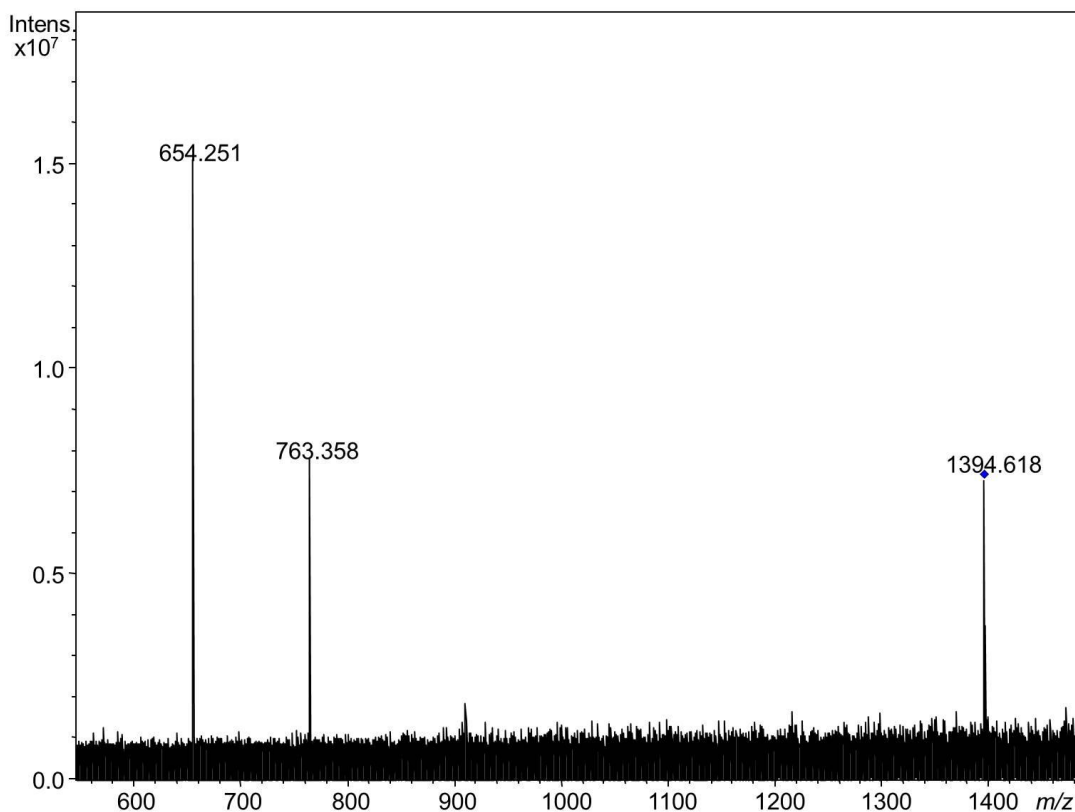


Figure 3.10: MALDI-product ion spectrum of $[M+2Na-H]^+$ for capture of the peptide leucine enkephalin on the SAM (precursor at m/z 1394)

In both Figure 3.9 and Figure 3.10, there is a peak at m/z 763. This suggests that the product ion originates from the SAM, and is not peptide specific. There is also a change from even to odd m/z from the parent ion to the product ion observed at m/z 763, indicating the number of nitrogen atoms has changed from odd to even. It was calculated that the ion at m/z 763 is due to the $[M+Na]^+$ species of the SAM fragment shown in Figure 3.11, with CID-induced cleavage occurring within the PEG chain of the disulfide-peptide species, and not at the amide bond as might have been expected. In Figure 3.10, the peak at m/z 654 corresponds to the $[M+2Na-H]^+$ species of the SAM-peptide structure shown in Figure 3.11, presumably produced via β -elimination of the peptide-containing fragment of the SAM-peptide complex.

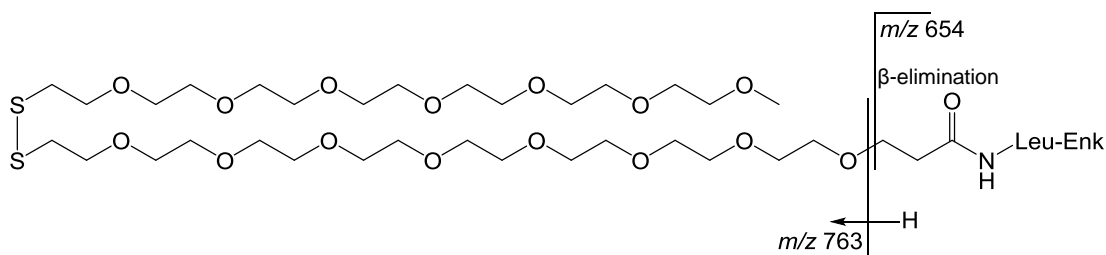


Figure 3.11: Proposed fragments for product generated on CID of leucine enkephalin captured on disulfide-derived SAM

In Figure 3.9, the peak at m/z 903 is unassigned.

The fact that ions at the same m/z value (m/z 763 ($[M+Na]^+$)) were observed in the CID spectra of both peptides captured on the SAM (the fragment shown in Figure 3.11) could be a convenient diagnostic ion for peptide capture products. For a more complicated peptide mixture in which the peptides may not have been identified (for example a protein digest captured on the SAM), CID MS/MS analysis could very conveniently be used to demonstrate successful peptide capture, using the ion at m/z 763 as a diagnostic fragment for the disulfide-derived SAM product (Figure 3.11).

3.4. 4-Bromophenylalanine and phenylalanine coupling to SAM

In previous results (Sections 2.4 and 2.6, and Section 3.3), problems have been noted when attempting to capture 4-BrPhe and Phe-Gly-Gly on both the thiol-derived SAM (Section 2.4 and 2.6) and the disulfide-derived SAM (Section 3.3). In Chapter 2, peaks associated with unreacted SAM were observed in spectra for peptide captured on the SAM. This is most likely to be due to a cumulative inefficiency in the different chemical steps in the protocol. With the disulfide-derived SAM, the SAM is already pre-activated and so further investigation of coupling Phe-containing species to the SAM was undertaken using the disulfide-derived SAM, and 4-BrPhe and Phe.

In the case of coupling 4-BrPhe to the SAM, it was not clear whether the presence of the Phe, or the combination of the Phe with the bromine in the chemical tag was contributing to the poor coupling of the tag to both the thiol-derived SAM and the disulfide-derived SAM. It is possible that the size of the bromine atom could be contributing to any steric hindrance when the 4-BrPhe was reacted with the disulfide-derived SAM. Therefore, coupling of 4-BrPhe and Phe was attempted at 5 nmol and 0.05 nmol levels on the disulfide-derived SAM.

4-BrPhe and Phe were prepared in PBS for capture on the disulfide-derived SAM at 5 nmol and 0.05 nmol levels. Aliquots (1 μ L) were spotted on wells and left for one hour at ambient temperature. The gold-coated MALDI chip was washed with acetonitrile and dried under a stream of nitrogen. Aliquots (0.3 μ L) of THAP solution were spotted on wells prior to MALDI-MS analysis.

Figure 3.12 shows the MALDI mass spectrum obtained following reaction for one hour of 4-BrPhe at 5 nmol and 0.05 nmol with the disulfide-derived SAM (peak assignments shown in Table 3.5).

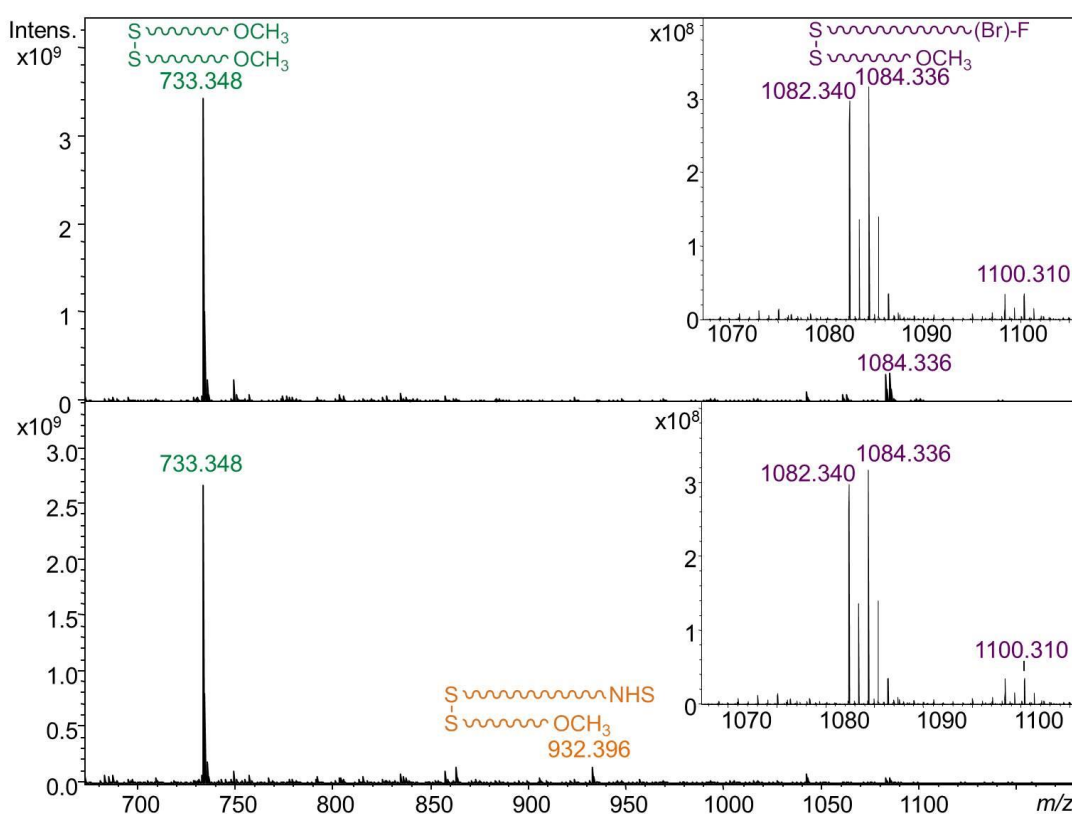


Figure 3.12: MALDI mass spectra obtained following reaction for one hour of 4-BrPhe at 5 nmol and 0.05 nmol with the disulfide-derived SAM. Top = 5 nmol of 4-BrPhe spotted on the disulfide-derived SAM, bottom = 0.05 nmol of 4-BrPhe spotted on the disulfide-derived SAM

As can be seen in Figure 3.12, 4-BrPhe coupling was achieved when 5 nmol or 0.05 nmol of 4-BrPhe were used. At the 0.05 nmol level, peaks can be seen in the mass spectra which correspond to species bearing unreacted NHS groups (Table 3.5), which are not observed in the spectrum obtained for the higher 4-BrPhe amount (5 nmol).

Table 3.5: Peaks associated with disulfide-derived SAM and 4-BrPhe coupled to the disulfide-derived SAM

Nominal m/z value	Assignment
733	$[M+Na]^+$ - 2 mPEG termini
932	$[M+Na]^+$ - 1 mPEG terminus, 1 NHS terminus
1082/1084	$[M+2Na-H]^+$ - 1 mPEG terminus, 1 CO-4-BrPhe terminus
1098/1100	$[M+Na+K-H]^+$ - 1 mPEG terminus, 1 CO-4-BrPhe terminus

For all mass spectra obtained following reaction for one hour of 4-BrPhe at 5 nmol and 0.05 nmol with the disulfide-derived SAM, peaks associated with coupling of the 4-BrPhe to the SAM consistently showed a higher S/N value at the 5 nmol level than the 0.05 nmol level. The data suggest that coupling of the 4-BrPhe to the disulfide-derived SAM was more efficient when 5 nmol of 4-BrPhe was reacted with the disulfide-derived SAM than 0.05 nmol of 4-BrPhe reacted with the disulfide-derived SAM. There is a 100-fold increase in the amount of 4-BrPhe spotted on the wells at 5 nmol compared to 0.05 nmol, which is likely to be the reason for the coupling being more efficient, as there is a larger excess of 4-BrPhe. This could drive the coupling of the 4-BrPhe and the SAM.

Figure 3.13 shows MALDI mass spectra obtained following reaction for one hour of Phe at 5 nmol and 0.05 nmol with the disulfide-derived SAM (peak assignments shown in Table 3.6).

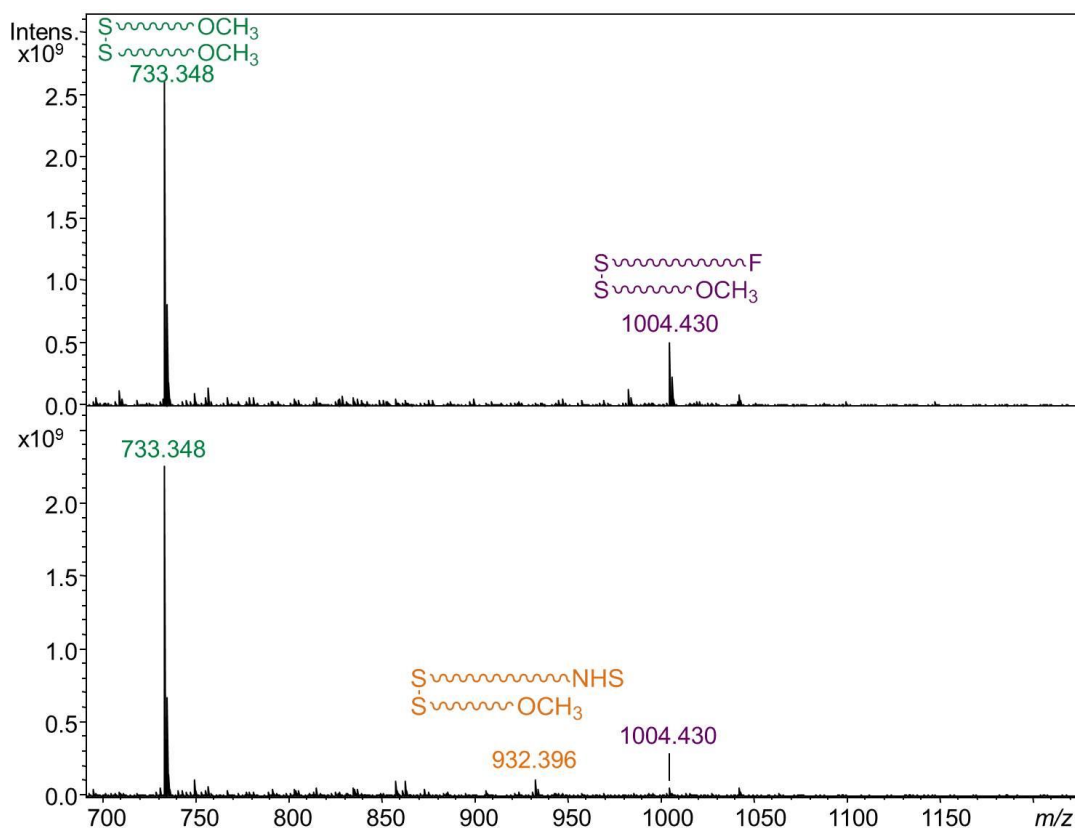


Figure 3.13: MALDI mass spectra obtained following reaction for one hour of Phe at 5 nmol and 0.05 nmol with the disulfide-derived SAM. Top = 5 nmol of Phe spotted on disulfide-derived SAM, bottom = 0.05 nmol of Phe spotted on disulfide-derived SAM

As can be seen in Figure 3.13, coupling of Phe was successful when both 5 nmol and 0.05 nmol were spotted on the SAM. As with 4-BrPhe, at the 0.05 nmol amount, peaks can be seen in the mass spectra which correspond to species bearing unreacted NHS groups.

Table 3.6: Peaks associated with disulfide-derived SAM and Phe captured on the SAM

Nominal m/z value	Assignment
733	$[\text{M}+\text{Na}]^+$ - 2 mPEG termini
932	$[\text{M}+\text{Na}]^+$ - 1 mPEG terminus, 1 NHS terminus
1004	$[\text{M}+2\text{Na}-\text{H}]^+$ - 1 mPEG terminus, 1 CO-Phe terminus

As with the coupling of 4-BrPhe to the disulfide-derived SAM, the mass spectra obtained following reaction for one hour of Phe at 5 nmol with the disulfide-derived SAM consistently showed a higher S/N compared with that from the spectra obtained following reaction for one hour of Phe at 0.05 nmol with the disulfide-derived SAM. The data suggest that coupling of the Phe to the disulfide-derived SAM was more efficient when 5 nmol of Phe was

reacted with the disulfide-derived SAM than 0.05 nmol of Phe reacted with the disulfide-derived SAM. There is a 100-fold increase in the amount of Phe spotted on the wells at 5 nmol compared to 0.05 nmol. This could be the reason for the coupling to be more efficient, as there is an excess of Phe, which could drive the coupling of the Phe and the SAM.

The mass spectra obtained for 4-BrPhe and Phe coupling to the SAM were examined for the presence of peaks due to species associated with 4-BrPhe/Phe on the SAM and species bearing unreacted NHS groups. The percentage of spectra where peaks were observed for species associated with 4-BrPhe/Phe captured on the SAM and species bearing unreacted NHS groups was calculated (Table 3.7).

Table 3.7: Table for 4-BrPhe and Phe spotted at 5 and 0.05 nmol levels on the disulfide-derived SAM, and percentage of spectra with species for 4-BrPhe/Phe capture and unreacted SAM (number of replicates = 8).

	Amount spotted on wells / nmol	Peaks seen for capture on SAM?	Percentage ^{vi} of spectra with signal for capture	Peaks seen for unreacted SAM?	Percentage ^{vi} of spectra with signal for unreacted SAM
4-BrPhe	5	Yes	100%	Yes	0%
	0.05	Yes	66%	Yes	6%
Phe	5	Yes	100%	Yes	0%
	0.05	Yes	50%	Yes	25%

As can be seen in Table 3.7, the percentage of spectra for which peaks associated with coupling of the 4-BrPhe/Phe to the SAM is higher at 5 nmol than at 0.05 nmol.

As was suggested and discussed in Section 3.3, it is possible that partitioning of the 4-BrPhe and Phe is occurring in solution due to the mismatch in the properties of the solvent and the solute.^{200–202} The hydrophobic phenyls in the side chains of the Phe could be causing aggregation, and pack the N-termini of the 4-BrPhe or Phe inside the aggregate. This would result in lower coupling efficiency observed for 4-BrPhe or Phe reacting with the disulfide-

^{vi} One mass spectrum recorded from each well. Data extracted from spectra for relevant peaks (e.g. species bearing unreacted NHS groups). Percentage values calculated as a percentage of the total number of spectra recorded for each peptide at each amount spotted on the SAM.

derived SAM than an amino acid with no phenyl side chain, as the N-termini of the 4-BrPhe and Phe would be hidden within the aggregate.

Alternatively, it is also possible that π -stacking of the aromatic side chains of the Phe amino acid could be occurring in solution. As with the theory that the hydrophobic Phe amino acids could be causing the formation of aggregates, the possibility of π -stacking of the aromatic side chain of the Phe amino acids would also result in aggregates being formed in solution. If aggregates were formed in solution, this would mean that the N-termini of the 4-BrPhe and Phe would not be available to react with the SAM, which would have resulted in the low coupling efficiency observed with 4-BrPhe and Phe.

Consistent with the results seen in Section 3.3, coupling of 4-BrPhe and Phe with the disulfide-derived SAM was more successful at the 5 nmol level than at the 0.05 nmol level. This could again be due to larger aggregates being present in the more dilute solution (where 0.05 nmol of 4-BrPhe or Phe was reacted with the disulfide-derived SAM), which would prevent reaction of the 4-BrPhe or Phe with the disulfide-derived SAM.²⁰⁵ There is excess 4-BrPhe or Phe, when compared to the theoretical amount of reactive-termini on the SAM, which should theoretically mean no peaks being observed for unreacted SAM.

When the results seen for 4-BrPhe and Phe reacting with the disulfide-derived SAM are combined with the results seen in Section 3.3 for the peptide Phe-Gly-Gly reacting with the disulfide-derived SAM, the data may suggest that the presence of the phenylalanine residue at the N-terminus of a peptide reduces capture efficiency between the peptide and the SAM. This is likely to be due to some form of aggregation of the peptide, 4-BrPhe or Phe when in solution in the PBS and spotted on wells containing the SAM. Aggregation could be due to the hydrophobic nature of the Phe amino acid or π -stacking of the aromatic side chains of the Phe amino acid. Both of these would cause the aggregation of the Phe-containing species in the PBS, which would reduce the efficiency of the capture on the disulfide-derived SAM.

There is also the possibility that the presence of the Phe amino acid increases the amount of peptide/amino acid needed for capture on the disulfide-derived SAM, due to the fact that a larger proportion of spectra contained peaks due to unreacted SAM at the 0.05 nmol level than the 5 nmol level. This could be due to the fact that larger aggregates could occur in the more dilute solution (0.05 nmol of peptide, 4-BrPhe or Phe reacted with the disulfide-derived SAM) than in the more concentrated solution (5 nmol of peptide, 4-BrPhe or Phe reacted with the disulfide-derived SAM).

It is possible that peptides which contain amino acids with aromatic side chains (Phe, Tyr, His and Trp) at the N-terminus may exhibit the same properties as Phe-Gly-Gly, 4-BrPhe and Phe when reacted with the disulfide-derived SAM. This could be tested by reacting a variety of peptides containing an amino acid with an aromatic side chain with the disulfide-

derived SAM. Ideally, an equivalent peptide containing the same sequence, but an amino acid without an aromatic side chain (such as Gly) in place of the Phe, Tyr, His or Trp should be used. This would build on the work undertaken using Phe-Gly-Gly and Gly-Gly-Gly (Chapter 2 and Section 3.3) by allowing a comparison between the two types of amino acids (those containing an aromatic side chain, and those without an aromatic side chain). It would also be interesting to see whether the position of the aromatic side chain had any effect on the efficiency of the peptide capture on the disulfide-derived SAM (i.e. whether the capture efficiency is improved if the aromatic side chain is moved further away from the N-terminus).

3.5. Conclusions and future work

A variety of peptides has been captured on the disulfide-derived SAM at 5 nmol and 0.05 nmol levels. The peptide capture step has been reduced from an overnight reaction to one hour. This is a considerable time saving for the workflow, and could potentially mean that the entire process (from SAM formation to peptide capture and analysis) could take place in two days with one overnight step for simultaneous SAM formation and protein digestion.

The data suggest that when reacting Phe-Gly-Gly, 4-BrPhe or Phe with the disulfide-derived SAM, the capture is inefficient. When 5 nmol and 0.05 nmol levels of the Phe-containing species were reacted with the disulfide-derived SAM, the capture of the Phe-containing species was more efficient at the 5 nmol level than the 0.05 nmol level. This is likely to be due to some form of aggregation occurring in the peptide solution. There is a mismatch in the properties of the solvent (PBS) and the solute, which could result in the hydrophobic Phe amino acid causing aggregation in the aqueous solution, and packing the N-termini of the Phe-containing species inside the aggregate. This would mean that the N-termini were not available for reaction with the SAM. Alternatively, π -stacking of the aromatic side chain of the Phe amino acid could also be causing the formation of aggregates within the PBS.

A possible explanation for the fact that for some peptides peaks were still observed for unreacted SAM after peptide capture at the 0.05 nmol level (even though this is above the estimated amount of reactive groups on the surface of the SAM) could be the fact that larger aggregates were formed in the more dilute peptide solution. Therefore, there would be less peptide available to react with the SAM when 0.05 nmol of peptide was used, as larger aggregates could be present than when 5 nmol of peptide was reacted with the SAM.

Further work into possible alternative solvents for the peptide solution (and therefore subsequent reaction of the peptide and SAM) should be undertaken.

Alternative chemical tags should also be investigated. Due to the fact that there is a problem with phenylalanine, 4-BrPhe is not a suitable chemical tag to use. A possible alternative would be to adapt a currently-used proteomics chemical labelling strategy (see Chapter 5 and Chapter 6).

The disulfide-derived SAM consists of the sulfur head group, the PEG groups and the terminal groups. This is in contrast to the SAM components used in Chapter 2, which contain an alkyl chain between the sulfur head group and the PEG groups. The PEG groups do not have the structural rigidity of alkyl chains. This is because these chains can be solubilised by water, with hydrogen bonds formed between the PEG groups and the water.²⁰⁶ The solubilisation of the PEG groups will disrupt the monolayer. Therefore, the cartoon image of the SAM with the PEG chains protruding from the solid surface, and ultimately the NHS groups projecting beyond the end of the SAM, may well be an inaccurate interpretation. Due to the incorporation of water between the PEG chains, it is possible that the PEG chains may be tangled, and the NHS groups may be buried within the SAM. Therefore, a peptide would not be able to react with the NHS group. The disulfide-derived SAM components may lack structural integrity when the SAM is formed on the gold-coated MALDI chip. In order to regain some of the rigidity that may be required for the SAM, an alternative SAM should be investigated. The SAM components should include an alkyl region to give some rigidity to the SAM, as well as a PEG region, to stop non-specific adsorption on the gold surface. Either thiol-derived or disulfide-derived components could be used to form a SAM for subsequent peptide capture.

Alternative SAM components which could be explored include those published by Murray *et al.* in 2014.²⁰⁷ The group (in collaboration with Steven Johnson's group at the University of York) synthesised a variety of alkanethiol-oligoethyleneglycols to produce a variety of functionalised SAMs. One SAM which was formed consisted of two thiols: one terminating in an alcohol group (the spacer molecule) and the other terminating in a primary amine. These two components both contain alkyl chains followed by a PEG region, which would meet the proposed criteria for potential SAM components. The group then reacted the amine termini of the SAM with bis(sulfosuccinimidyl) suberate (BS3) (Figure 3.14) to form an activated SAM to capture the antibody anti-hCG through amide bond formation. This could potentially be adapted for use with the peptide capture work.

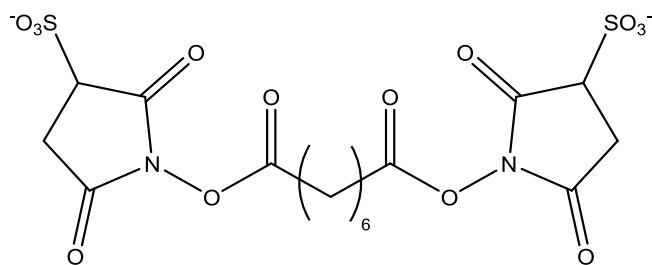


Figure 3.14: Structure of bis(sulfosuccinimidyl) suberate

The BS3 molecule could be used as the chemical tag (a potential alternative chemical tag to 4-BrPhe). A 'heavy' version of the BS3 would need to be synthesised, to include either deuterium or ^{13}C atoms (or a combination of the two), most likely within the alkyl region in the middle of the BS3 molecule. After the SAM had been formed on the gold-coated MALDI chip, a 50:50 mixture of the light and heavy BS3 molecules would be reacted with the SAM. The peptide(s) would then be reacted with the BS3-labelled SAM and the resulting SAM-peptide structure analysed by MALDI-MS.

Chapter 4: Using plasma effluent as a novel cleaning technique

4.1. Introduction

Plasma is often referred to as the fourth state of matter. It is an ionised gas which, on the macroscale, contains equal numbers of positive and negative charges. However, on the microscale, strong electric fields can exist between the particles, due to an imbalance of charges in a particular region. The charges arise from molecules of gas being split into ions and electrons. Although the electrons and ions are no longer bound together, they cannot be described as free, as there is still an interaction between the two.

There are some very different types of plasma (e.g. stars, lightning and nuclear fusion reactors), and plasmas can be used for many different applications (e.g. etching, X-ray lasers and semiconductor technology).^{208–210}

Plasma is generated either through heating, or by application of an electric field to a gas. When applying an electric field (e.g. to a laboratory-based plasma), the gas is turned into plasma by flowing it through the space between a pair of electrodes. The high electric charge tears electrons from the particles of gas, creating charges within the gas. This is a cascading process, creating more free electrons by collisions between the free electrons and ions already generated within the body of gas. The degree of ionisation of the plasma^{vii} is expressed as a ratio between the number density of ions and the number density of neutrals, and ranges from fully ionised to partially ionised (Equation 4.1).

$$I = \frac{n_e}{n_e + n_o}$$

Equation 4.1: Equation showing the degree of ionisation of a plasma (n_e = electron density, n_o = number density of neutral atoms)

Within plasma, Debye shielding occurs. This essentially means that electrons tend to gather near a positive charge within the plasma, creating an electrostatic shield around the charge and vice versa. The shielding length, known as the Debye length, is an important characteristic of plasma, as it must be much smaller than the physical size of the plasma. The Debye length is essentially a measure of how far electrostatic effects of a charge carrier persist.

A quasi-neutral state exists throughout plasma, except near the boundaries (i.e. the edges of the plasma). An example of a confined plasma is a laboratory-based plasma, where a plasma sheath exists on all cold walls (e.g. the electrode casing) with which the plasma makes contact. A plasma sheath is a layer of charge imbalance near the cold surface. It consists of a layer of negative charge, which repels electrons. Electrons have a higher thermal velocity than ions and so stand a greater chance of colliding with the wall, and

^{vii} The number of species per cm³

therefore the loss of electrons from the plasma is greater than the loss of ions. A net positive charge thus develops over time, and raises the potential of the bulk of the plasma above the potential of the walls (i.e. makes the potential of the walls negative with respect to the bulk of the plasma). This potential drop provides a barrier for electrons and attracts ions. Therefore, adjacent to the negative boundary layer is a layer of positive space charge, which has a rapidly increasing potential across it (as the distance from the wall increases), until a quasi-neutral state is reached (Figure 4.1).²¹¹

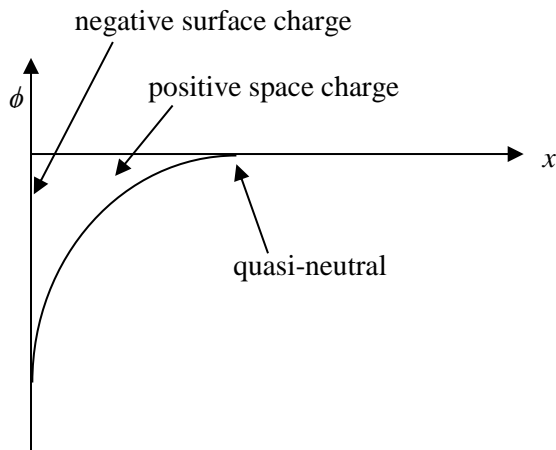


Figure 4.1: Potential structure in a plasma sheath (ϕ = potential, x = distance from the cold surface)²¹¹

In plasma, the different particles that can arise fall into two categories: charged (such as ions and electrons) and neutral (such as radicals and metastables). Metastables refer to molecules in a metastable state. These are excited states which have a finite lifetime and relax after a period of time into a more stable state (i.e. the ground state). A metastable state can be described as an energy trap, and is often visualised as a well in the energy profile for a molecule. Therefore, metastable molecules can be thought of as molecules which spend an extended period of time in a long lived configuration, which is not the ground state.

Particles within plasma are free to move about, and so can collide and interact with each other. In a simplified model, electrons can be thought of as the objects that move within the plasma, and the neutral species can be thought of as staying stationary. The electrons then make their way past these stationary neutral species (Figure 4.2). When an electron collides with a neutral species in a plasma it bounces off, as the neutral species is assumed to be a hard sphere, and the electron behaves like a point mass.²¹¹ Using a collision cross section (σ), it is possible to quantify the probability of a collision between a neutral species and an electron, or how far into the space filled by neutral species an electron can travel before having a collision. In the simplified model, the cross section of the species that the electron

sees depends on the energy of the electron encountering the gas (the total kinetic energy in the encounter).²¹¹

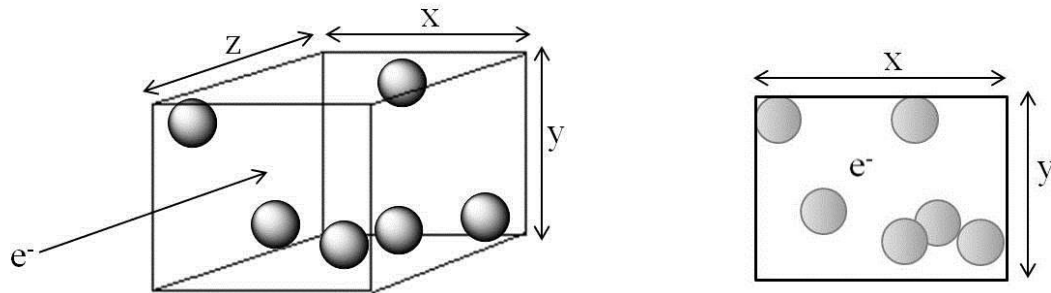


Figure 4.2: Simplified model for an electron travelling through stationary neutral species in a 3D model (left) and 2D model (right) (e⁻ = electron)

The number of target atoms in the cuboid xyz is n_gxyz (where n_g = number of gas units per unit volume). Each neutral species presents a cross section (σ) of:

$$\sigma = \pi r_g^2$$

Equation 4.2: Equation for cross section of an atom (r_g = radius of an atom of the gas used in the plasma)

When the path of the electron is viewed through the xy plane, the total area blocked by neutral species is $n_gxy\lambda\sigma$. If the cuboid extends as far as one mean free path (λ), virtually the entire face xy is obscured:

$$(n_gxy\lambda)\sigma = xy$$

$$\lambda = \frac{1}{n_g\sigma}$$

Equation 4.3: Equation (upper) showing that virtually the entire face xy is obscured when the cuboid extends as far as one mean free path (and simplified (lower))

These equations show that the electron is very likely to have a collision. The frequency of encounters can be estimated for the electrons:

$$\nu = \bar{v}n_g\sigma$$

Equation 4.4: Equation for the frequency of encounters for electrons (ν = frequency of encounters, \bar{v} = mean speed)

However, this is a simplified model, and only takes into consideration simple elastic encounters between neutral species and electrons. Cross sections are not independent of energy.²¹¹ High-energy electrons travel so fast that their interactions with the outer spheres of neutral species are reduced.

Elastic and inelastic collisions are possible within plasma. In an elastic collision, the total kinetic energy is conserved, but the momentum is redistributed between particles and there is a change of direction for the particles involved in the collision.²¹¹ Electrons cannot lose much energy to the ions or neutrals, but the change in momentum between the two can be large. In an inelastic collision, excited states can occur due to energy transfer within the particles involved in the collision, and the total number of particles resulting from the collision does not have to be maintained.²¹¹ A fraction of the initial kinetic energy is transferred to internal energy to form excited states, which can go on to fragment. Electrons can lose almost all of their kinetic energy when they collide with ions or neutrals.

Plasma temperature is usually referenced in Kelvin. At low temperatures, the ions and electrons within the plasma are likely to recombine to form atoms. If enough recombination occurs, then the plasma returns to a gaseous state. Temperatures are defined within the plasma for electrons (T_e), ions (T_i) and neutrals (T_{neutrals}). It is possible for a laboratory-based plasma to either be thermal or non-thermal. For thermal plasma, the bulk of the plasma is at thermal equilibrium, meaning that the three specified temperatures are the same. For non-thermal plasma, the electron temperature is much higher than the ion and neutral temperatures. Due to the relative differences in mass between the ions and the neutrals versus the electrons, energy transfer is not efficient between the high temperature electrons and the larger, cooler ions and neutrals, and so overall the plasma is considered to be at room temperature. This gives rise to the term 'atmospheric pressure plasma'. The two plasma instruments used in the work described in this chapter were both operated at atmospheric pressure, and are termed an 'atmospheric-pressure plasma jet' (APPJ). These plasma sources offer advantages over low-pressure plasmas, such as the fact that expensive vacuum systems are not required.²¹² However, it can be difficult to maintain a glow discharge in an atmospheric-pressure plasma due to the absence of a vacuum.²¹² Higher voltages are required, and arcing can occur between the electrodes.

An example of a high temperature plasma is the plasma in an inductively coupled plasma mass spectrometer (ICP-MS) which uses a high temperature, inductively coupled plasma source with a mass spectrometer.

Laboratory-generated plasmas can be thought of as steady, non-thermal equilibrium states.²¹¹ Continuous electrical discharges are often used to create and maintain a dynamic steady state, such as forcing an electric current through a gas in order to ionise it and create plasma. The electric current must be continually supplied to ensure the plasma is

maintained. A balance exists between production and loss of ions. For laboratory-generated plasmas, where production and loss processes are generally separated physically, energy which is invested in the production of ions is effectively transported to the place of loss (sink).

When plasma comes into contact with a solid surface, the surface acts as a sink, which 'drains away' the plasma. There is a recombination of electrons and ions at the surface due to the fact that the electrons and ions are suddenly at lower temperature. The environment surrounding the plasma (such as the physical boundaries of a laboratory-generated plasma, or the blanket of neutral gaseous molecules surrounding the plasma effluent) can also act as a sink.

In plasma generated using an admixture (e.g. helium and a small amount of another gas), the species generated can be altered from those produced using a single gas. The key chemistry occurring within the plasma is largely dependent on the ratio of the other gas (often oxygen or nitrogen) added to the helium. Oxygen and nitrogen allow different species to be produced within the plasma, which may not be observed with a pure helium gas plasma. With the addition of oxygen, reactive oxygen species (ROS) are produced (e.g. hydroxyl radicals and ozone), whilst the addition of nitrogen leads to reactive nitrogen species (RNS) being produced (e.g. nitric oxide and nitrogen dioxide). The effect of these different species on the properties of the plasma effluent is not yet understood.²¹³

Plasmas have previously been used for cleaning of metal surfaces, but have mainly been generated using argon.^{214,215} In 2005, Raiber *et al.* published a paper in which they used plasmas generated from hydrogen or oxygen to remove an alkanethiolate SAM from a gold surface.²⁰⁹ They compared the efficiency of the two plasmas using XPS, STM, ellipsometry (to determine the thickness of the film) and contact angle measurements. The authors found that although use of the oxygen plasma resulted in a level of oxidation on the gold surface, the changes in surface topology detected by STM were not significant. Therefore, they were satisfied that either of the plasmas could be used for the removal of alkanethiolate SAMs from a gold surface.

In 2012, Favrat *et al.* published a paper detailing the removal of stearic acid from a gold surface using atmospheric-pressure plasma.²¹⁶ Although this is not removal of SAMs from a gold surface using plasma, it is still removal of an organic layer from a metal surface. The group used a Kelvin probe to monitor residual stearic acid on the surface. The atmospheric plasma was generated using compressed air (giving a high composition of nitrogen and oxygen in the plasma effluent). The group found that after four exposures to the plasma effluent, residual contamination (remaining stearic acid) was between 5.7 and 8.1% compared to initial contamination.

For the work described in this chapter, two different plasma instruments were used: a kHz dielectric barrier discharge (DBD) plasma instrument and a radiofrequency (RF) plasma instrument. A DBD plasma is a non-equilibrium plasma, which exhibits similar charged-species density to a weakly ionised gas.²¹² The RF plasma instrument is operated with insulated or external electrodes and can be operated at a wide range of pressures, whilst the DBD plasma instrument operates at or near atmospheric pressure. Both the DBD and RF plasma instruments used in this work are classed as APPJs. A schematic of the area where the plasma is generated is shown in Figure 4.3 for each of the two plasma instruments.

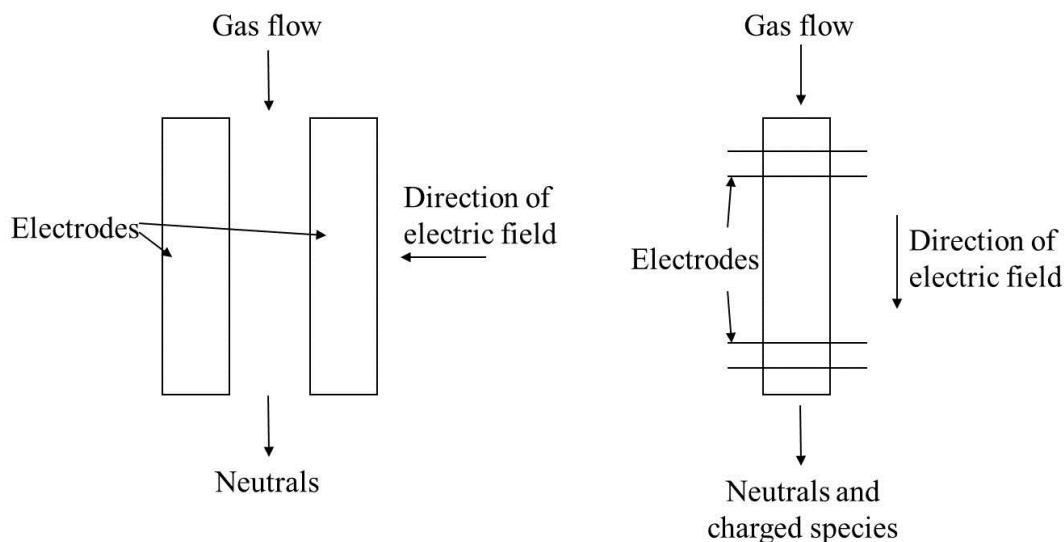


Figure 4.3: RF plasma electrode set up (left) and kHz plasma electrode set up (right). The electrode gap on the RF plasma electrode is $30 \times 1 \times 1$ mm. On the kHz plasma electrode, the lower electrode is powered, whilst the upper electrode is grounded.

With the RF plasma instrument, the electric field is perpendicular to the flow of gas, which results in charged species being confined within the electrode region. This means that they are not present in the plasma effluent, and therefore not interacting with the surface of the gold-coated MALDI chip. The kHz plasma instrument has an electric field which is parallel to the flow of the gas, resulting in no confinement of the charged species within the electrode region.

This work was undertaken in collaboration with Martin Blake and Deborah O'Connell of the York Plasma Institute and Victor Chechik of the Department of Chemistry.

4.1.1. Aims

The method currently used for cleaning the gold-coated MALDI chips is to use a piranha solution. However, this is a corrosive cleaning technique, and removes the outermost layer of gold from the gold-coated MALDI chip. After a number of cleans with piranha solution, the thin layer of gold which coats the gold-coated MALDI chips can become eroded enough for the metal underneath to be seen (Figure 4.4). A helium-based plasma effluent was suggested as an alternative to the piranha solution to remove the SAM from the gold-coated MALDI chip, whilst potentially causing minimal damage to the gold surface. Therefore, a helium-based plasma effluent was investigated as an alternative method for cleaning the gold-coated MALDI chips being used in the SAM-based workflow.

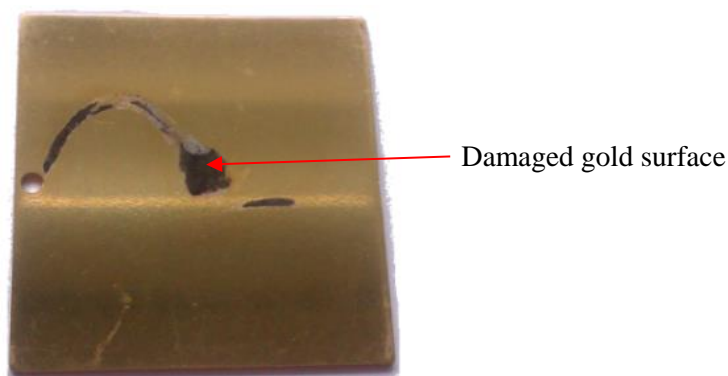


Figure 4.4: Damaged gold chip (arrow indicates where the metal layer under the gold can be seen after repeated use of piranha solution)

4.2. Formation of a thiol-derived SAM

A gold-coated MALDI chip was cleaned with piranha solution and spotted with aliquots of an alcohol thiol ($\text{HS}(\text{CH}_2)_{17}(\text{OCH}_2\text{CH}_2)_3\text{OH}$), which was left to assemble overnight at room temperature. The gold-coated MALDI chip was washed with acetonitrile and dried under a stream of nitrogen. Aliquots of matrix were spotted on wells, and the success of the formation of the SAM was assessed by MALDI mass spectrometry (Figure 4.5) before wells were subjected to the plasma effluent. Aliquots ($0.3 \mu\text{L}$) of THAP matrix solution were spotted on the wells prior to mass spectrometric analysis.

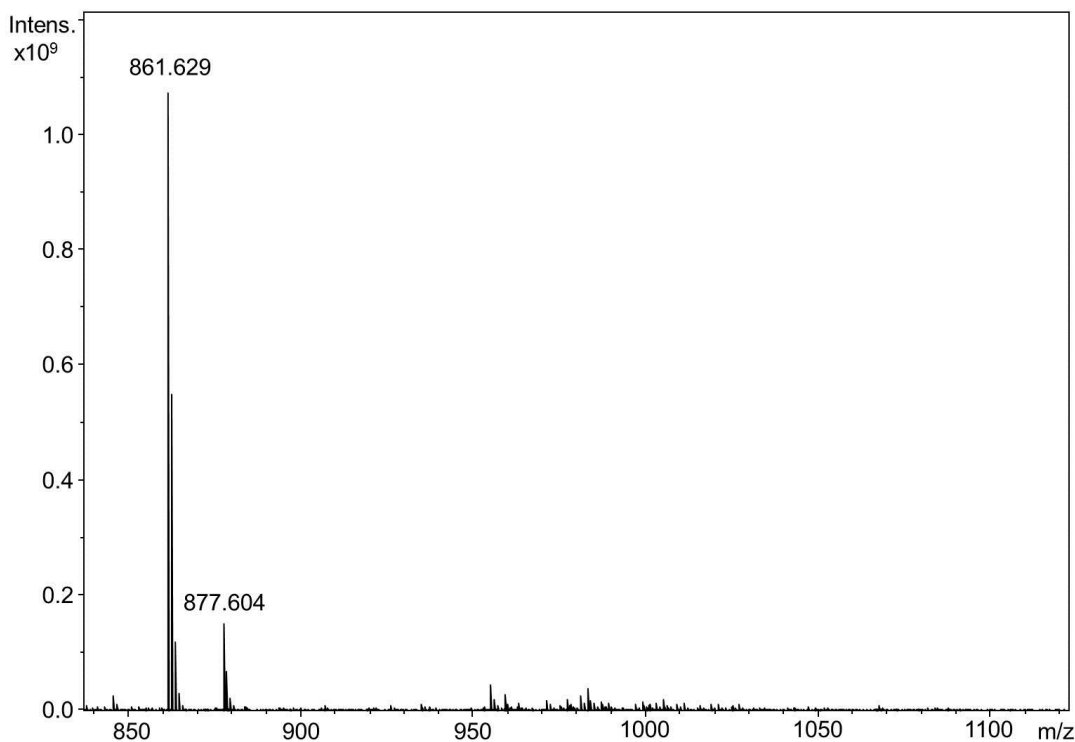


Figure 4.5: MALDI mass spectrum of the alcohol thiol-derived SAM

As can be seen in Figure 4.5, the alcohol thiol successfully formed a SAM. A peak at m/z 861 indicates the presence of an $[\text{M}+\text{Na}]^+$ disulfide having two alcohol termini. A peak at m/z 877 indicates the presence of an $[\text{M}+\text{K}]^+$ disulfide having two alcohol termini.

4.3. Testing the kHz plasma instrument for cleaning SAM-coated gold

The suitability of the kHz plasma instrument for SAM removal was assessed. It was observed that the plasma effluent extended beyond the end of the casing (Figure 4.6), and that the effluent didn't cover the surface area of a well. Consequently, the gold-coated MALDI chip needed to be moved whilst a well was being treated with the plasma effluent, in order to treat the whole surface of the well.

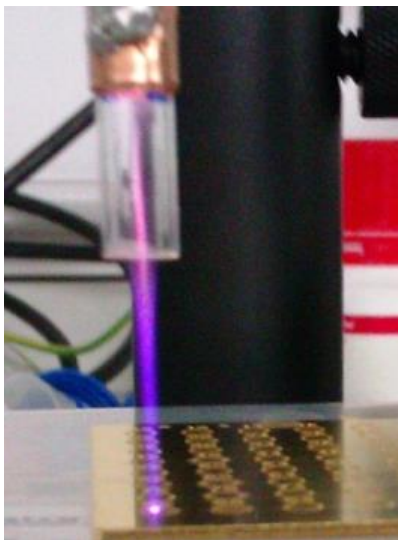


Figure 4.6: The electrode region of the kHz plasma instrument

4.3.1. Optimising conditions for the kHz plasma instrument

Wells which had SAM and matrix on them were treated with plasma effluent. The time that wells were subjected to the plasma effluent (30, 60, 90 and 120 seconds), and the distance of the plasma casing from the gold surface (2 mm, 10 mm, 17 mm, 25 mm and 40 mm), were varied to ascertain the potential cleaning ability of the plasma. Multiple wells were treated for each treatment combination (Table 4.1). The treated wells were then respotted with aliquots (0.3 μ L) of matrix solution and reanalysed by MALDI-MS. Spectra were obtained from the middle of each well and also from the outer edge of each well. This was due to the fact that the gold-coated MALDI chip had to be moved whilst the well was being treated with the plasma effluent, to see whether the movement had been successful.

Table 4.1: Treatment combinations for kHz plasma instrument, and for each treatment the percentage of spectra in which peaks for the SAM were detected (number of wells per condition = 4)

Treatment		<i>m/z</i> 861 detected	
		Middle of the well	Outer edge of the well
17 mm	30 s	75%	100%
	60 s	25%	100%
	90 s	25%	100%
	120 s	25%	75%
30 s	2 mm	100%	100%
	10 mm	100%	100%
	17 mm	100%	100%
	25 mm	100%	100%
	40 mm	100%	100%

As can be seen in Table 4.1, it was found that peaks associated with the SAM (m/z 861 ($[M+Na]^+$)) were observed in spectra obtained after plasma treatment for all treatment combinations tested. However, the S/N obtained after plasma treatment was approximately a fifth to a tenth of the S/N value from spectra obtained before plasma treatment. Peaks associated with the SAM were observed in mass spectra obtained after plasma treatment at all distances.

As a comparison, wells which had been cleaned with piranha solution were spotted with aliquots of matrix and analysed by MALDI-MS. No peaks were observed for species associated with the SAM (m/z 861 and 877).

Peaks associated with the SAM were observed in spectra obtained after plasma treatment of wells with SAM formed on them, but no peaks were observed in spectra obtained after piranha cleaning of wells with SAM formed on them. The data suggest that the plasma treatments used in this experiment were not as effective at removing the SAM from the gold-coated MALDI chips as the piranha solution.

4.3.2. Incremental treatment time course study

While it was evident that the relative intensities of SAM-derived mass spectrometric signals were decreased by kHz plasma treatment, it was not possible to determine from this experiment whether this was because cleavage within the SAM was occurring, or whether the SAM was being cleaved at the gold-sulfur bond. Therefore, a time course study was

undertaken, with sequential wells being subjected to increasing periods of treatment with plasma effluent; periods from 10 seconds to 160 seconds were tested, with periods increased by 10 second increments. A parallel experiment was carried out to determine whether removal of the SAM from the gold-coated MALDI chip was affected by matrix addition; a duplicate experiment was performed without adding matrix to the wells before treatment (in both cases wells were subjected to irradiation periods increasing in increments of 10 seconds). The plasma effluent was created using helium with 0.5% oxygen.

The S/N of the peak at m/z 861 ($[M+Na]^+$, disulfide containing two alcohol termini) was extracted from each mass spectrum recorded for wells which had been spotted with matrix prior to plasma irradiation, and plotted (Figure 4.7).

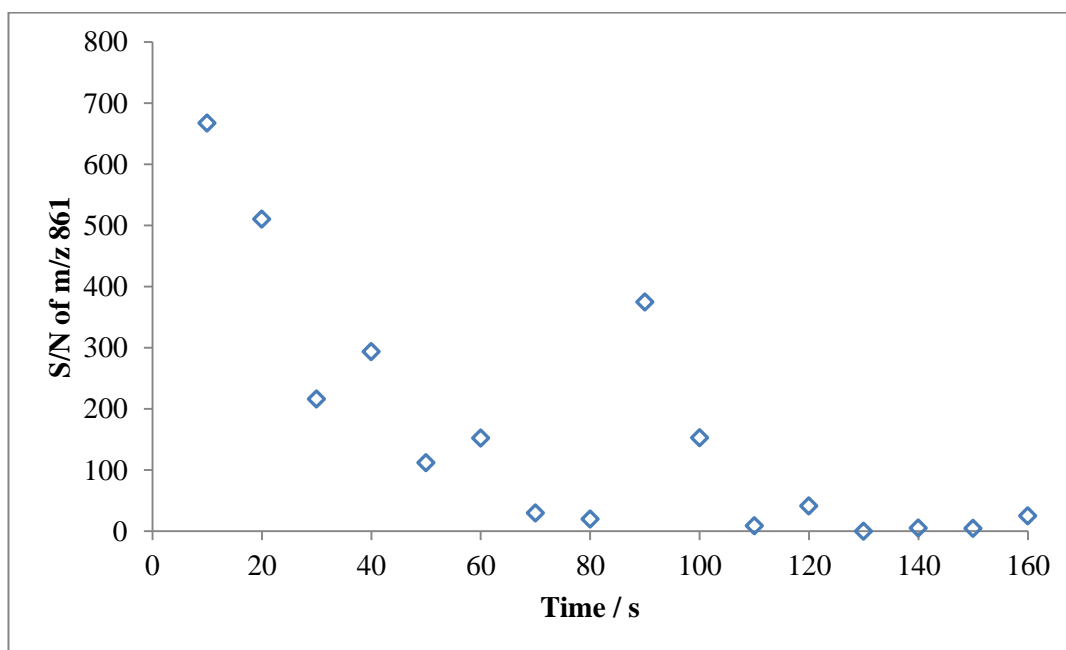


Figure 4.7: Plot showing the S/N of m/z 861 after SAM-bearing, matrix-spotted wells were treated with the plasma effluent of the kHz plasma instrument

Figure 4.7 shows an overall decrease in the relative S/N of m/z 861 with increasing period of plasma treatment. However, the downward trend is not smooth, with spikes seen at some time points (e.g. 90 seconds). This is likely to be due to MALDI's innate variability¹⁹² and the fact that only one well was analysed for each time point; there was no reanalysis of the wells (to allow an average value to be determined) and the experiment was not repeated. This was due to time-constraints, and subsequent loss of access to the plasma instruments. Had there been the opportunity to repeat the analysis, a minimum of four wells would have been analysed for each time point, allowing mean S/N values to be calculated for each time point. This would allow easier identification of anomalous results, which could then be excluded from the data. Although the data point at 90 seconds does not fit the downward trend

observed for the time-course study, this result was included in the overall analysis due to the fact that only single wells were analysed.

An example of the mass spectra obtained after plasma treatment of matrix-free SAM-containing wells is shown in Figure 4.8.

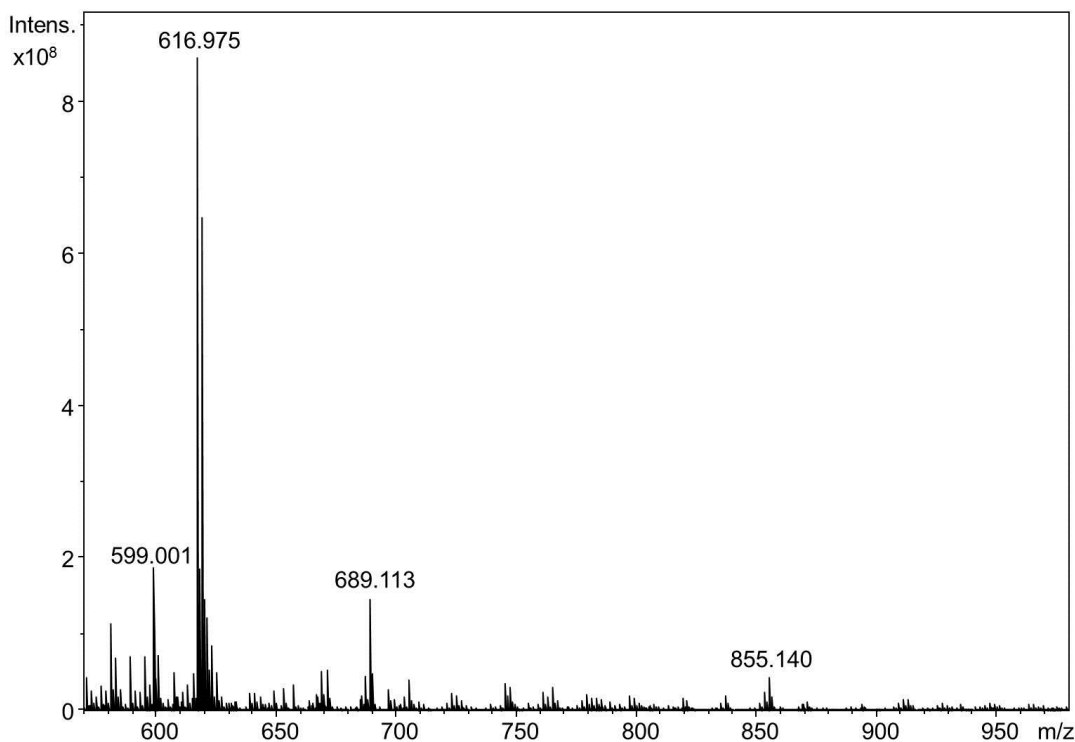


Figure 4.8: MALDI mass spectrum obtained from a well with SAM but no matrix treated with the effluent of the kHz plasma instrument for 10 seconds

As can be seen in Figure 4.8, there was no detectable signal at m/z 861 or 877, which would have indicated an $[M+Na]^+$ and $[M+K]^+$ for a disulfide having two alcohol termini. This indicates that the SAM had been reduced below the levels of detection from wells within the shortest treatment period of 10 seconds. The fact that the presence of the SAM had been removed from a well in less than 10 seconds indicates that this could be a suitable alternative method to the piranha solution for removing SAMs from gold-coated MALDI chips, provided the matrix solution is rinsed from the wells before treatment with the plasma effluent.

4.4. Testing the radiofrequency (RF) plasma instrument for cleaning SAM-coated gold

The plume from the RF instrument plasma (Figure 4.9) did not protrude from the end of the casing, in contrast to that of the kHz plasma instrument. Consequently, the end of the electrodes was placed 2 mm from the surface of the gold-coated MALDI chips. The distance between the two electrodes is 3 mm, which is the approximate diameter of a well on the gold-coated MALDI chip. Therefore, the gold-coated MALDI chip was not moved around whilst each well was being treated with the plasma (contrary to treating wells with the kHz plasma instrument).

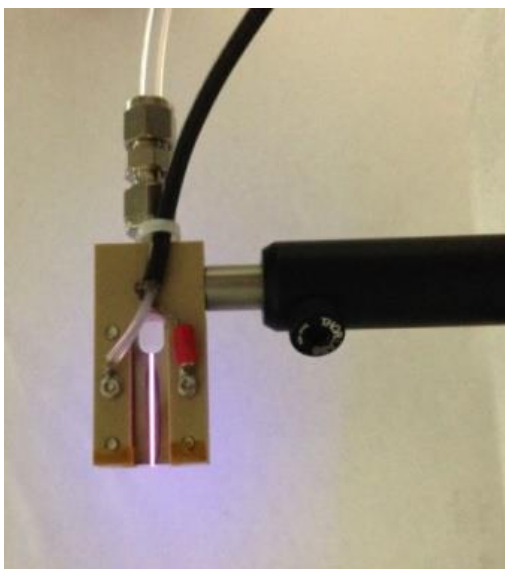


Figure 4.9: The electrode region of the RF plasma instrument

4.4.1. Optimising conditions for the RF plasma instrument

The RF plasma instrument is usually run with addition of small amounts (e.g. 0.3%) of nitrogen or oxygen to the helium. The addition of oxygen or nitrogen alters the properties of the plasma. The possible benefits of addition of one or other of the gases to the helium were investigated, to determine whether addition of such a modifier gas produced a plasma that exhibited improved cleaning properties compared to only helium being used. The times that the wells were treated with the RF plasma instrument were longer than those used for the kHz plasma instrument. These times were suggested by the collaborators due to the differences in the orientation of the electric fields of the two plasma instruments.

After SAM formation, aliquots (0.3 μL) of matrix solution were spotted on wells, and the wells were treated with the RF plasma instrument (Table 4.2). Multiple wells were subjected to

the same treatment conditions. The wells were then respotted with aliquots (0.3 μL) of matrix solution and analysed by MALDI-MS.

Table 4.2: Treatment combinations for the RF plasma instrument, and for each treatment the percentage of spectra in which peaks for the SAM were detected (number of wells analysed for each condition = 2)

Gas added	Percentage added	Time / mins	m/z 861 detected
Nitrogen	0.1%	3	100%
		5	100%
	0.25%	3	100%
		5	100%
	0.5%	3	100%
		5	100%
	0.75%	3	100%
		5	100%
Oxygen	0.1%	3	100%
		5	50%
	0.25%	3	50%
		5	50%
	0.5%	3	50%
		5	50%
	0.75%	3	50%
		5	50%

As can be seen in Table 4.2, the addition of oxygen to the helium had more of an impact than the addition of nitrogen, as the number of wells from which the mass spectra obtained after plasma treatment contained the peak associated with the SAM (m/z 861) was lower. Where the ion at m/z 861 was observed in the mass spectrum, the S/N value was lower than in spectra recorded prior to plasma treatment (an approximate tenth to twentieth decrease in the S/N value). For both nitrogen and oxygen incorporation, spectra obtained for wells which were treated for five minutes showed lower S/N values for the peak associated with the SAM (m/z 861) than spectra obtained for wells which were treated for three minutes.

4.4.2. Time course study

A similar experiment to the time course study with the kHz plasma instrument was performed. It was evident that the relative intensities of SAM-derived mass spectrometric

signals were decreasing with increasing time of plasma treatment, but it was not possible to determine whether this was because cleavage was occurring within the SAM, or whether the SAM was being cleaved at the gold-sulfur bond. Therefore, a time course study was undertaken, with sequential wells being subjected to increasing periods of treatment with the RF plasma effluent; periods from 10 to 160 seconds were tested with periods increased by ten second increments. A parallel experiment was carried out to determine whether removal of the SAM from the gold-coated MALDI chips was affected by matrix addition, by carrying out the same experiment without adding matrix to the wells before cleaning. The plasma effluent was created using helium with either oxygen or nitrogen incorporation. Two gas mixtures were tested: 0.25% nitrogen incorporation in helium and 0.5% oxygen incorporation in helium. The wells were irradiated with the plasma effluent at a distance of 2 mm.

The S/N of the peak at m/z 861 ($[M+Na]^+$, disulfide containing two alcohol termini) was extracted from each mass spectrum recorded for wells which had been spotted with matrix prior to plasma treatment for both oxygen and nitrogen incorporation, and plotted (Figure 4.10).

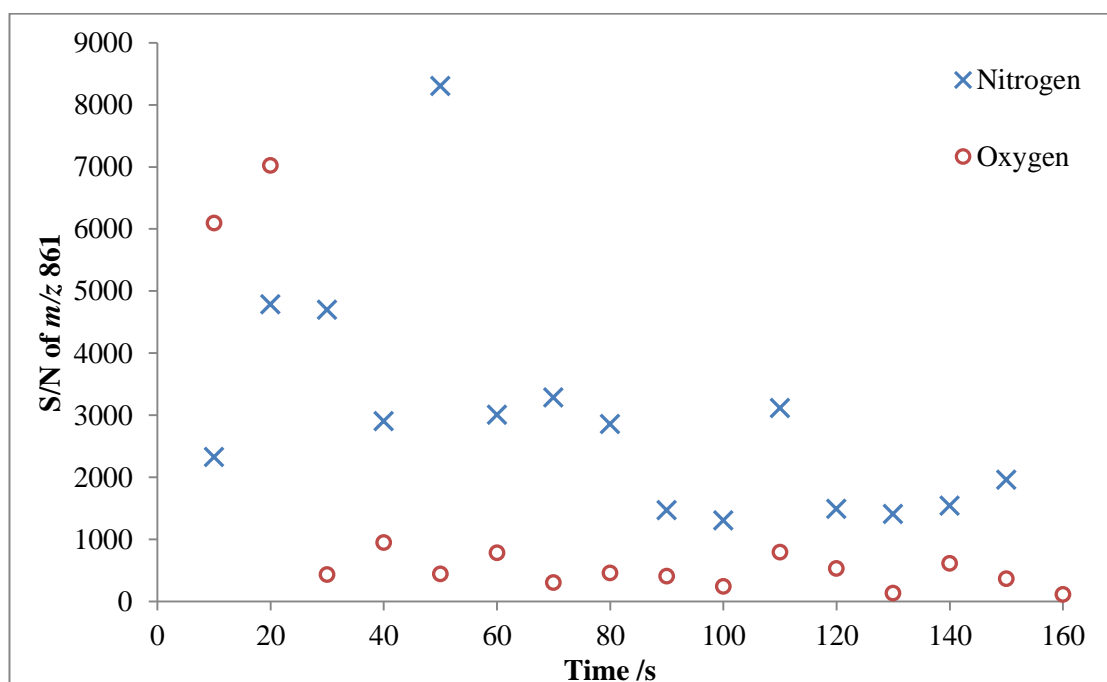


Figure 4.10: Plot showing S/N of m/z 861 after SAM-bearing, matrix spotted wells were treated with the plasma effluent of the RF plasma instrument. Results are shown for the addition of 0.25% nitrogen (blue) or 0.5% oxygen (red)

Figure 4.10 shows an overall decrease in the relative S/N of m/z 861 with increasing period of plasma treatment. However, as was seen with the kHz plasma instrument (Figure 4.7), spikes were seen at some time points (e.g. oxygen incorporation at 50 seconds). This is likely to be for the same reasons as with the kHz plasma instrument.

There is a sudden decrease in the relative S/N of the peak at m/z 861 after 20 seconds treatment with the plasma formed with oxygen incorporation. Based on the graph produced, it can be implied that incorporation of oxygen into the helium is the better option, rather than nitrogen incorporation. This is in agreement with the data in Table 4.2, which suggest that oxygen incorporation provides better SAM removal than nitrogen incorporation. However, as with the earlier analysis with the kHz plasma instrument, these conclusions are based on a single well for each time point, and so replicate analyses should be performed to allow mean values to be obtained.

An example of the spectra obtained after plasma treatment of matrix-free SAM-containing wells is shown in Figure 4.11.

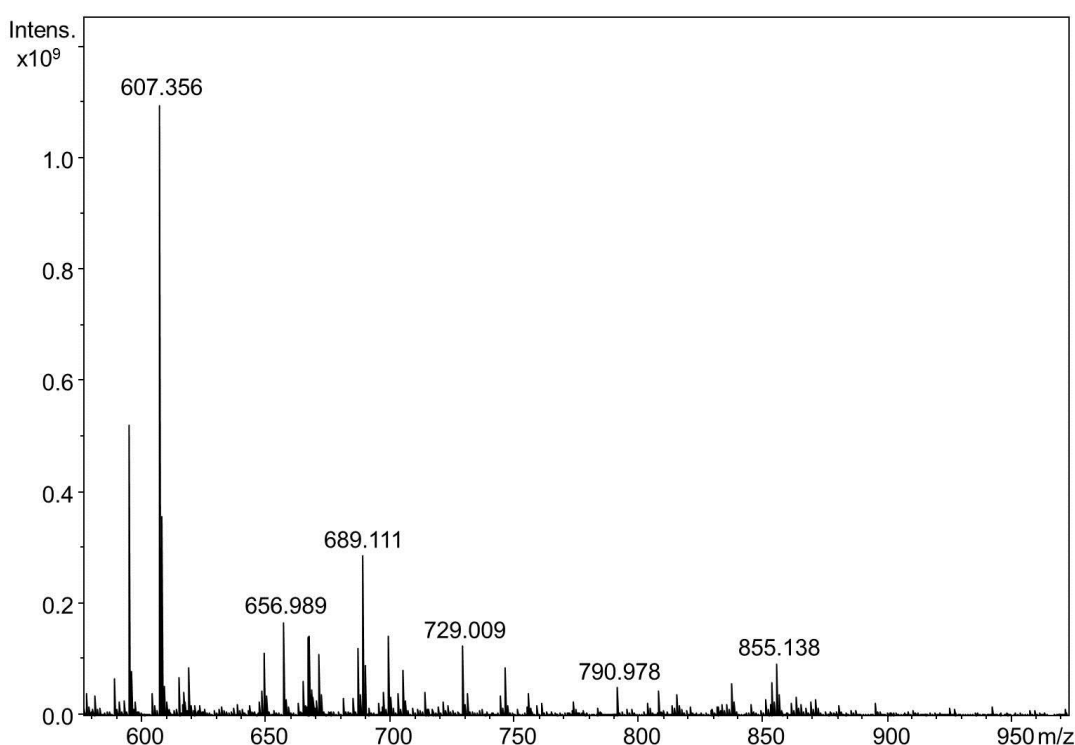


Figure 4.11: MALDI mass spectrum obtained from a well with SAM but no matrix treated with the effluent with nitrogen incorporation of the RF plasma instrument for ten seconds

When mass spectra were obtained from wells which only had SAM on them, there was no detectable signal at m/z 861 ($[M+Na]^+$, disulfide having two alcohol termini). This indicates that the presence of the SAM had been reduced below the levels of detection from wells within the shortest treatment period of 10 seconds. The fact that the SAM could be removed from a well in less than 10 seconds indicates that this could be a suitable alternative method to the piranha solution for removing SAMs from the gold-coated MALDI chips, provided the matrix was washed off the surface of the gold-coated MALDI chip prior to plasma treatment.

4.5. Investigation into unassigned mass spectrometric peaks observed after plasma irradiation using either kHz or RF plasma instruments

Peaks (such as those at m/z 607, 638, 668, 689), which were not associated with the SAM, were observed in mass spectra recorded after wells were subjected to plasma treatment using either one of the different instruments (Figure 4.12).

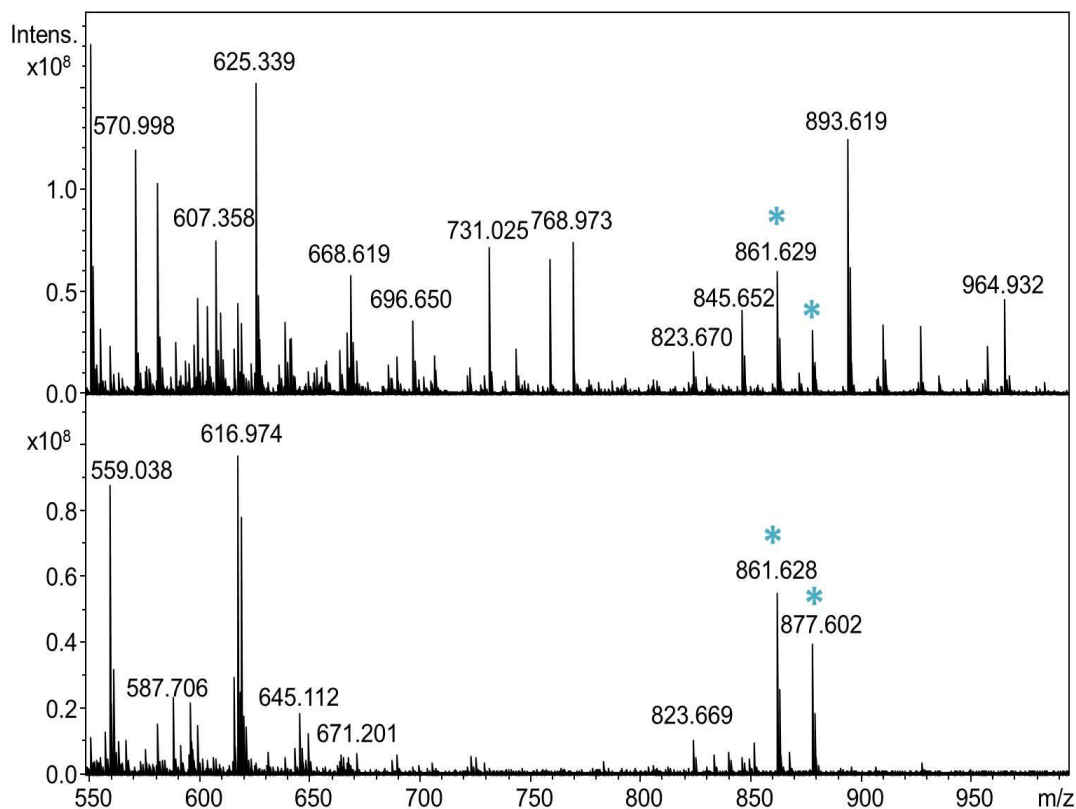


Figure 4.12: MALDI mass spectra obtained after treatment with the kHz plasma instrument (top) and RF plasma instrument (bottom) (* indicates peaks which are associated with the SAM)

The source of these peaks was investigated using the kHz plasma instrument (30 second irradiation), to determine whether they were related to cleavage of the SAM, the matrix or the gold surface.

Wells on which no SAM was formed were spotted with matrix and irradiated with the plasma effluent prior to mass spectrometric analysis. The unassigned peaks were observed in the resulting mass spectra. This suggested that the peaks observed were derived from either the matrix or the gold as they had been observed in mass spectra in the absence of the SAM on the gold-coated MALDI chip, and so do not derive from cleavage of the SAM.

Wells with no SAM and no matrix were irradiated with the plasma effluent prior to mass spectrometric analysis. The unknown peaks were observed in the resulting mass spectra, suggesting that the peaks were due to the gold, and not to an interaction between the plasma and the matrix.

In order to determine if the species giving rise to the unknown peaks were present on the gold surface before plasma treatment, a new gold chip (which had never been cleaned) was analysed. Wells on the gold-coated MALDI chip were spotted with aliquots (0.3 μL) of matrix solution and analysed by MALDI-MS. The gold-coated MALDI chip was then treated with piranha solution before being spotted with matrix and reanalysed. Figure 4.13 shows example mass spectra obtained before and after cleaning a new gold-coated MALDI chip with piranha solution.

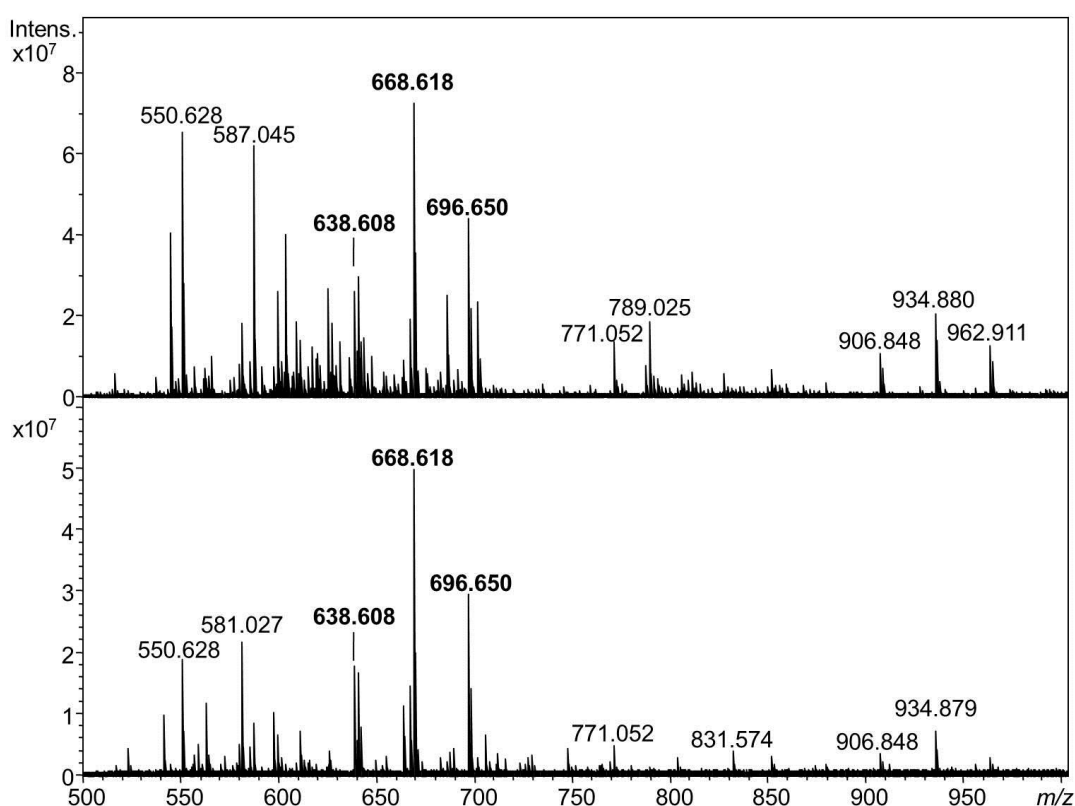


Figure 4.13: MALDI mass spectrum obtained for a new gold chip before (top) and after (bottom) cleaning with piranha solution

As can be seen in Figure 4.13, peaks at m/z 638, 668 and 696, observed in all spectra recorded from the plasma experiments, are also present in spectra recorded before the gold-coated MALDI chip had been subjected to any of the SAM-forming or cleaning treatments. It can thus be concluded that cleaning the gold-coated MALDI chips with plasma effluent does not appear to alter the gold-coated MALDI chip, as far as is observable through MALDI mass spectrometry.

4.6. Respotting of thiol-derived SAM on wells cleaned with a plasma effluent

Having determined that the SAM could be removed from the gold-coated MALDI chip by treatment with the kHz and RF plasma instruments, the ability to reform a SAM on wells which had been subjected to plasma effluent was investigated.

Wells on a gold-coated MALDI chip were treated using the RF plasma instrument. A solution of the alcohol thiol was then spotted on wells of the gold-coated MALDI chip and left to assemble overnight. The success of the formation of the SAM was determined by MALDI mass spectrometry (Figure 4.14).

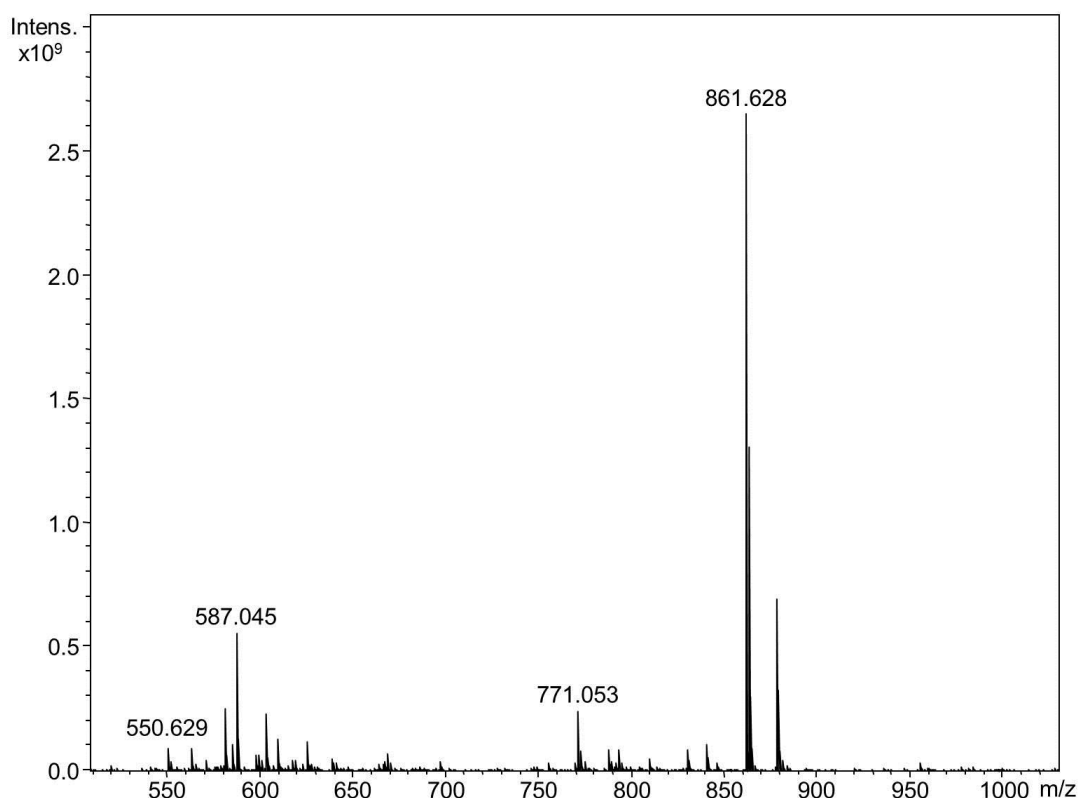


Figure 4.14: MALDI mass spectrum showing alcohol thiol-derived SAM respotting on a well cleaned with plasma effluent (m/z 861 ($[M+Na]^+$, disulfide having two alcohol termini))

As can be seen in Figure 4.14, the formation of the SAM on wells which had previously been irradiated with plasma effluent was successful, as indicated by the presence of the peak at m/z 861 (c.f. Figure 2.4 which shows formation of a mixed SAM, with one component being the alcohol thiol on a well cleaned with piranha solution). Therefore, a SAM can be successfully formed on the surface of a well which has been treated with plasma effluent.

4.7. Conclusions and future work

An alcohol thiol-derived SAM was removed from gold-coated MALDI chips using kHz and RF plasma instruments. Removal of the SAM was much easier without the presence of matrix on the surface of the wells. Therefore, it is suggested that the matrix deposited on wells after MALDI-MS analysis be removed by washing (e.g. with acetonitrile), before cleaning the gold-coated MALDI chip using either of the plasma instruments. This has the potential to be an effective alternative to the piranha treatment currently used. Removal of the matrix deposit prior to plasma treatment would allow each well to be cleaned in less than 10 seconds. This represents a considerable decrease in the time taken to clean a gold-coated MALDI chip over that required using piranha solution as a full chip could be cleaned in less than 11 minutes, as opposed to approximately 30 minutes when cleaning with the highly caustic piranha solution. A SAM has also been reformed on wells which had been treated with plasma effluent. Plasma treatment provides a promising new method for cleaning the gold-coated MALDI chips, which may lead to the chips being usable for a longer period of time, as the caustic piranha solution damages the gold surface (Figure 4.4).

The theory that removing the matrix by washing the gold-coated MALDI chip and then treating the wells with either plasma instrument is an effective way of removing SAM from wells of the chip would need to be tested. This would involve a simple comparison of plasma treatment of wells that had only had SAM formed on them and wells that had SAM formed on them and were then spotted with matrix solution and washed. Provided that washing the gold-coated MALDI chip to remove the matrix, followed by plasma treatment of the wells proved an effective way to clean the wells, this would provide an alternative cleaning strategy to the caustic piranha treatment used currently.

Replicates should be analysed for the time course studies with both of the plasma instruments to allow a greater confidence in the results obtained. This would enable mean values of the S/N of m/z 861 to be derived for each time point, which could help reduce the effects of experimental variability.

While MALDI mass spectrometry allows the formation or removal of a SAM to be studied, it does not analyse the surface of the wells directly, to see if they are being damaged by the plasma effluent. Techniques such as XPS, AFM and STM could be used to give insight into the effects of the plasma on the gold surface. XPS and STM were used by Raiber *et al.* to monitor the removal of organic monolayers from a gold surface.²⁰⁹

XPS is a surface-sensitive quantitative technique, which measures elemental composition of a material, and gives an indication of the chemical state of the element. In the case of the gold-coated MALDI chips, it would give an indication of whether the gold is in an oxidised state on the surface of the chip after treatment with the different plasma effluents or piranha

solution.²¹⁷ The purpose of cleaning the gold-coated MALDI chips is to clean the surface, but also to ensure the gold is in an unoxidised state.

AFM and STM are used to investigate the structure and profile of a surface. AFM essentially involves the movement of a tip over a surface.²¹⁸ This tip is attached to a cantilever. The movement of the tip across the surface is monitored, and movement up or down is recorded. In this way, a model of the surface can be produced. STM is a technique used to image a surface at the atomic level.²¹⁹ A probe is held close to the sample surface. Electrons tunnel between the surface and the probe, causing an electrical signal. The probe scans across the surface of the sample, generating an image of the surface through the current produced. These techniques could be used to determine whether the surface of a well has been completely freed from SAM by comparison with an underivatized area of the gold-coated MALDI chip which had not been treated with plasma effluent, and also with piranha solution-cleaned wells. AFM and STM could also be used to determine whether damage to the surface had occurred by assessing whether tracks or other surface damage to the surface of the gold-coated MALDI chip had resulted from any of the cleaning treatments.

XPS could be applied to the analysis of the surface of the gold-coated MALDI chip due to the fact that the technique can be used to look at the chemical and electronic states of the elements. This would allow oxidation of the gold to be detected. Oxidation of the gold surface after cleaning will reduce the ability of the SAM to form a perfect monolayer as atomic gold is needed. AFM could be applied to determine whether damage had occurred to the surface of the gold-coated MALDI chip after the different cleaning treatments, but also to check that the plasma effluent is capable of removing all traces of the SAM from the surface of a well.

Chapter 5: Dimethyl labelling

5.1. Introduction

Having established (Chapters 2 and 3) that 4-BrPhe was not wholly suitable as the chemical tag for use with the SAM technology, alternative labelling strategies were sought. Non-commercial chemical labels (e.g. dimethyl labelling^{122,123}) and commercial labels (e.g. iTRAQ¹¹² or TMT¹⁰⁸) were researched, with the aim to adapt an already-validated and well adopted labelling strategy for incorporation into the SAM/gold-coated MALDI chip technology. iTRAQ and TMT are well-established commercial labels, which use expensive reagents, and were therefore discounted. The DiLeu strategy uses quantification at the MS/MS level. This may provide an additional complication when combining with the SAM. Dimethyl labelling and ¹⁸O incorporation through protein digestion were both considered. With ¹⁸O incorporation, there is the possibility of either one or two ¹⁸O (either 2 or 4 Da) being incorporated. If only one ¹⁸O is incorporated, a mass difference of 2 Da would mean a large overlap of the isotopic envelopes, which could lead to quantification errors. The chemical labelling strategy which appeared to be most promising for potential use with the SAM technology, and was thus chosen for further investigation, and possible adaptation, was dimethyl labelling. This was chosen for a number of reasons: dimethyl labelling occurs on the amino terminus of peptides which makes it readily compatible with the SAM technology thus far developed; the technology is not commercialised, making it amenable to modification and cheaper to work on; the dimethyl label has good potential to be modified for incorporation into the SAM technology (see discussion in Chapter 6); and there was an interest to develop dimethyl labelling as a mainstream labelling strategy within the group.

Dimethyl labelling uses reductive amination with formaldehyde to label the primary amino group at the N-terminus of peptides and in the side chain of lysine residues, replacing the two hydrogen atoms on the primary amine with methyl groups. One advantage of dimethyl labelling is that it does not rely on costly commercial reagents like some other labelling strategies (e.g. iTRAQ and TMT), since it uses readily available rather than custom reagents. In addition, it has the advantages of being reliable, cost-effective and procedurally undemanding.¹²² The dimethyl labelling approach allows relative quantification, in which ratios of the intensities of the peaks for light and heavy labelled peptide pairs are compared.

The addition of two methyl groups to a primary amine results in an increase of 28 Da for the light label and an increase of 32 Da for the heavy label (Figure 5.1) when compared with unlabelled peptide. This means that there is a mass difference of 4 Da between the light and heavy versions of a singly-derivatised peptide.

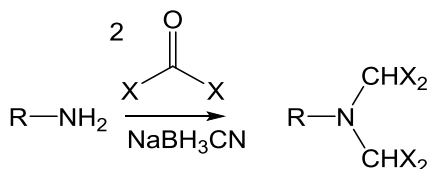


Figure 5.1: Dimethyl labelling strategy developed by Hsu *et al.* (X = H (light) or D (heavy))¹²²

In 2009, a paper by Boersema *et al.* described three strategies which could be used for dimethyl labelling: in-solution, online and on-column.¹²³ In each of the strategies, labelling was reported to be complete within 60-90 minutes. The in-solution dimethyl labelling strategy described¹²³ was the basis for the work reported in this chapter.

5.1.1. Aims and overview

It was important to determine that quantitative results could be obtained using in-solution dimethyl labelling, before attempting to adapt the dimethyl label for incorporation into the SAM technology. Therefore, the aim of the work described in this chapter was to investigate the viability of dimethyl labelling for quantitative analysis, using two ionisation methods for mass spectrometric analysis: MALDI and ESI. ESI is the more usual choice of ionisation technique for chemical labelling and quantitative proteomics work. However, MALDI is highly relevant for the SAM aspect of this thesis, and therefore the ability to combine dimethyl labelling and MALDI mass spectrometric analysis has an important bearing on the potential to combine dimethyl labelling with the SAM technology. Having determined the viability of dimethyl labelling for quantitative analysis, the method was applied to muscle protein samples (these samples were readily available for analysis, and there was interest in making a comparison with an emPAI analysis being undertaken on these samples at the same time).

The Boersema paper described the application of triplex dimethyl labelling to a BSA digest, but the authors themselves did not apply dimethyl labelling to 'real' samples.¹²³ There have, however, been several examples of dimethyl labelling being used for quantitative proteomics of 'real' samples. In 2009, Ji *et al.* published a paper using dimethyl labelling for absolute quantification of therapeutic monoclonal antibodies.²²⁰ The group used a labelling strategy based on previous publications, which involved the guanidination of lysine side-chains, to prevent multiple labels being incorporated into lysine-containing peptides.²²¹ The publication details an LC-MS method (and its subsequent validation) for absolute quantification of a therapeutic monoclonal antibody in biological fluids using dimethyl labelling.

In 2011, Munoz *et al.* published a paper detailing the use of dimethyl labelling for the quantitative proteomic analysis of human-induced pluripotent stem cells and embryonic stem cells.²²² The authors used one of the labelling strategies (presumed to be the in-solution

labelling strategy due to there being no mention of column details for online or on-column strategies) detailed by Boersema *et al.*¹²³ Human embryonic stem cells, two different human-induced pluripotent stem cell lines and their precursor fibroblast cell lines were investigated. The group found a high level of similarity between human-induced pluripotent stem cells and embryonic stem cells proteomes, with a small subset of proteins differentially expressed in the two human-induced pluripotent stem cell lines.

Wilson-Grady *et al.* published a paper in 2013, which appeared during the course of the work described in this thesis, detailing the use of dimethyl labelling for the quantitative comparison of the phosphoproteomes of fasted and re-fed mouse livers.²²³ They used an on-line desalting and reductive dimethylation strategy. The authors varied the pH at which the labelling occurred using 2-(N-morpholino)ethanesulfonic acid, and found that at increasing pH, there were fewer peptide and protein identifications following LC-MS analysis. An optimum pH of between pH 5 and pH 6 was identified, and therefore pH 5.5 was used for the quantitative phosphoproteomics work. The group used a labelling solution of either 0.2% formaldehyde or D₂-formaldehyde with 60 mM sodium cyanoborohydride in 0.25 mM 2-(N-morpholino)ethanesulfonic acid.

Of the three papers mentioned above, only one uses an un-adapted version of the dimethyl labelling strategies published by Boersema in 2009.²²⁴ There have been multiple variations on the dimethyl labelling of peptides. Due to the fact that the overall aim of this work was to develop a labelling strategy which could be modified for use with the SAM technology, it was important that the labelling strategy chosen was an in-solution approach for best compatibility with the SAM workflow. The online and on-column strategies proposed by Boersema *et al.* are clearly not ideal for integration with the SAM technology, as they would add extra complexity to the overall workflow.

The in-solution dimethyl labelling method (as used by Boersema *et al.*)²²⁴ was thus first validated for quantification using standard solutions of light and heavy labelled simple protein digests mixed in a range of ratios. This was because there are variants of the dimethyl labelling strategy, and therefore demonstration that the in-solution dimethyl labelling strategy was suitable for quantification (and subsequent adaptation for use with the SAM technology) was sought, as well as the need to test its application with ESI and MALDI. Two different software packages were also assessed for analysing the data to determine whether one or other was more suitable for dimethyl labelling quantification: Bruker ProteinScape (PS) and Mascot Distiller (MD).

Having demonstrated that for a single protein digest, standard solutions could be analysed and quantified, more complex samples were analysed. An *E. coli* lysate digest (labelled with light:heavy reagents at 1:1) was spiked with single protein digests labelled and mixed in ratios that differed from 1:1. These protein digests were spiked in at 1% or 5% (w/w) of the

total *E. coli* lysate digest amount. The resultant spiked solutions were analysed by LC-ESI-MS. The levels at which spiked proteins could be identified and reliably quantified needed to be ascertained, in order to determine how abundant a protein would need to be in a 'real' sample for a change in protein expression between two samples to be detectable.

After determining the limits of detection for analysing complex samples, the validated approach was applied for the analysis of muscle protein samples. Having established the in-solution dimethyl labelling, analysis of the muscle samples provided a convenient way to analyse samples and compare the quantification results obtained with those obtained via emPAI analysis. The detailed comparison of the data is beyond the scope of this thesis. The muscle protein samples were one unstretched muscle sample, and two stretched muscle samples, where the stretched muscle samples were stretched for three hours. The aim of the experiment was to determine whether differences in the abundance of proteins between the unstretched and stretched muscle samples could be identified.

Data from this chapter, except for the muscle protein data, are available in an electronic format: MSV000080344 (BSA digests – ESI), MSV000080345 (BSA digest – MALDI) and MSV000080346 (*E. coli* spiking data). The raw muscle protein data are embargoed at present, and so not available for viewing at this time.

5.2. Validation of dimethyl labelling for quantification using standard solutions of light and heavy dimethyl-labelled protein digests

For dimethyl labelling to be used for relative quantitation of unknowns, it must be validated (i.e. shown to generate the expected experimental results from standards containing defined levels of different components). This is in order to demonstrate that the approach gives the expected results, before modifying it to develop a SAM-compatible version of the method. In order to assess the viability of dimethyl labelling for relative quantitation work, a simple system was tested, using digests of a single protein (in this case BSA). A standard protein tryptic digestion protocol was used, as recommended for use with iTRAQ (Section 7.5).^{112,225} This digestion protocol was used because it is well adopted and generates peptides under conditions appropriate for subsequent N-terminal labelling. The digestion solvent system is free of primary amines, which reduces the chance of competition when dimethyl labelling of peptides is underway. There was also an interest in comparing dimethyl labelling with iTRAQ, and therefore using the same digestion protocol for both labelling strategies provided a level of comparability. Digests were labelled with heavy or light formaldehyde and mixed in a range of different relative amounts, and subjected to mass spectrometric analysis.

The peak areas of the $[M+H]^+$ signals for the heavy and light peptide pairs in the resulting spectra were used to determine relative amounts. In all instances, each standard solution was prepared in triplicate, and forward and reverse labelling was performed by comparing both light:heavy (L:H) (forward) and heavy:light (H:L) (reverse) ratios. It is standard practice in relative quantification proteomics to prepare solutions allowing comparison of intensities when the light, and also the heavy label is the more abundant. The function of forward and reverse labelling is to ensure there is no bias associated with which sample carries the heavy or the light label. There are several reasons to prepare triplicate solutions, and perform forward and reverse experiments. These include the ability to test the reproducibility of preparing the solutions (e.g. are there any pipetting errors?), to allow systematic errors due to labelling to be identified,²²⁶ and to identify errors arising from overlapping isotopic envelopes in comparing the light and heavy label intensities.

5.2.1. Analysing standard ratios using aliquots of the same BSA digest

It is important to establish the dynamic range of a labelling strategy, which is to say the dynamic range over which accurate and reliable relative quantification data can be generated. Hsu *et al.* demonstrated a dynamic range over two orders of magnitude for

dimethyl labelling.¹²² Ideally, a proteomics labelling strategy should be able to provide accurate quantification data when there is a large or small difference in the expression of a protein.

Tryptic digestion of BSA was performed (as detailed in Section 7.5.1), and two aliquots of the digest were labelled: one with the light formaldehyde and the other with the heavy formaldehyde. Aliquots of the light and heavy labelled digests were dissolved to the same concentrations and combined in known ratios (v/v) and the resulting solutions analysed by MALDI-MS and LC-ESI-MS. The MALDI-MS data were analysed using PS, and the LC-ESI-MS data were analysed using PS and MD.

5.2.1.1. MALDI-MS analysis

MALDI-MS analysis was coordinated through WARP-LC software and the results were reported in PS (both are software by Bruker Daltonics) (Figure 5.2). The ratios of the solutions prepared were 1:1, 2:1, 3:1, 5:1 and 10:1 (L:H and H:L). A mass spectrum was acquired from each solution (each sample was prepared in triplicate and analysed). From the resulting data, a list of all possible peptide pairs was generated, indicating m/z values for peptides with masses 4 Da or 8 Da apart (peaks for peptides with a difference of 4 Da contain one dimethyl group or 8 Da contain two dimethyl groups), and the ratio of peak areas for each peptide pair was calculated. The ratios were calculated by WARP-LC from the mass spectrum by summing the areas of the peaks associated with the isotopic envelope for a peak for a heavy labelled peptide and dividing it by the area of the peaks associated with the isotopic envelope for the peak of the corresponding light labelled peptide. The software then selected up to ten m/z values for product ion analysis (specified in the method as the peaks with the highest S/N). Once the MS/MS acquisition had been completed, the data were sent to PS. Subsequently, a protein search was performed, where the product ion spectra were submitted to the search engine Mascot, in order to identify the peptides, and ultimately the protein that had been labelled. Once the search had been completed, statistics for the protein quantification were generated in PS.

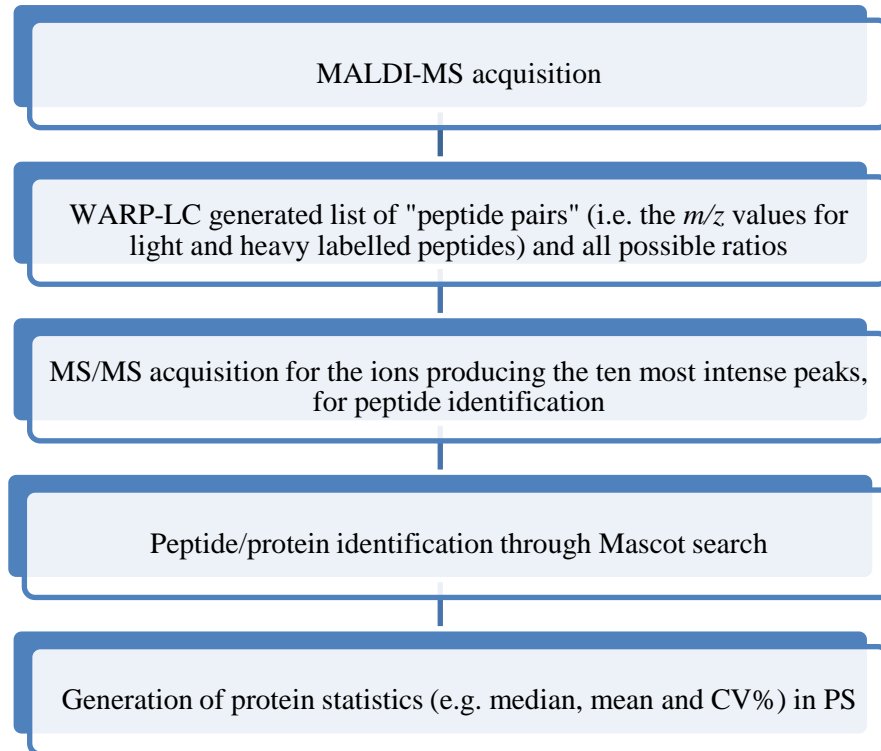


Figure 5.2: Workflow showing the steps in the acquisition of quantification data from MALDI-MS analysis using WARP-LC and PS

In PS, what Bruker (personal communication from Susanne Fütteer, in Proteineer Support/Software Development) refer to as a ‘median’ value is reported for the relative quantification of each quantified protein. However, in spite of its name, this is not a true median. In fact, a lognormal distribution is used by WARP-LC and PS to calculate a defined median (Equation 5.1), mean (Equation 5.2) and coefficient of variance (CV) (Equation 5.3) for each quantified protein.²²⁷ The peptide ratios calculated from the area of the heavy labelled peptide’s isotopic envelope divided by the area of the light labelled peptide’s isotopic envelope are defined as data set X. These values are then converted to their natural log values, creating data set Y. From the mean (μ) and variance (σ^2) of Y, the so-called ‘median’, mean and CV of X are calculated.

$$\mathbf{Median}_x = e^\mu$$

Equation 5.1: Formula used by Bruker PS for the determination of the protein quantification “median”

$$\mathbf{Mean}_x = \mu_x = e^{(\mu+0.5\sigma^2)}$$

Equation 5.2: Formula for the determination of the protein quantification “mean”

$$CV_x = \sqrt{e^{\sigma^2} - 1}$$

Equation 5.3: Formula for the determination of the coefficient of variation of data set X

A geometric mean and standard deviation was calculated using the ratios generated in PS for each of the triplicate solutions. Table 5.1 gives the ratios generated from the measurements carried out on the standard solutions (plotted in Figure 5.3). FDR values of 0.00% were obtained for all searches (Table 7.3).

Table 5.1: Ratios generated using PS for MALDI-MS analysis of prepared standard solutions of BSA tryptic digest (geometric standard deviations (SDgeo) calculated for the three replicate solutions prepared for each ratio) (number of peptides and sequence coverage for the three replicate solutions indicated) (see 7.5.6 for FDR values)

Ratio prepared (L:H)	Protein ratio	SDgeo	Number of peptides	Sequence coverage / %
0.10:1	0.06:1	1.03	5, 5, 5	19, 20, 16
0.20:1	0.11:1	1.10	4, 3, 4	7, 6, 10
0.33:1	0.20:1	1.14	1, 3, 4	1, 11, 9
0.50:1	0.32:1	1.04	5, 6, 7	12, 16, 18
0.67:1	0.51:1	1.06	2, 2, 7	5, 5, 16
1:1	1.02:1	1.03	1, 4, 1	2, 8, 6
1:1 (H:L)	1.02:1	1.06	2, 6, 5	4, 16, 12
1.5:1	1.62:1	1.03	2, 3, 5	4, 10, 13
2:1	2.57:1	1.06	3, 4, 3	9, 9, 12
3:1	5.00:1	1.04	3, 7, 5	6, 17, 11
5:1	10.54:1	1.05	0, 7, 6	0, 19, 16
10:1	12.02:1	1.33	4, 2, 7	14, 9, 19

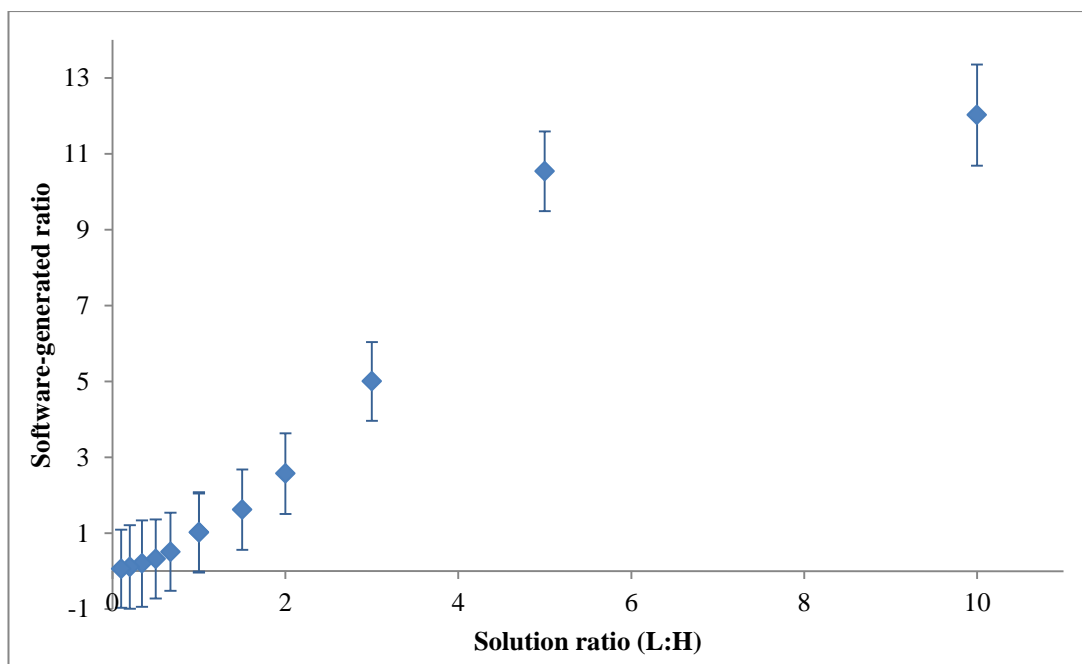


Figure 5.3: Plot showing the geometric mean generated from three replicate solutions analysed by MALDI-MS against the prepared ratio (error bars show the geometric standard deviation)

As can be seen from the results in Table 5.1 and Figure 5.3, the ratios obtained using PS deviated from the ratios of the standard solutions that had been prepared when the ratios are above 2:1, with the more abundant label being over-represented in each case. The geometric standard deviations are relatively large for each of the PS ratios. For 5:1 (L:H) solution 1, no useable data was obtained (the mass spectrum was very noisy, and no MS/MS data were acquired automatically, as the S/N ratios of the peptide signals were not high enough to trigger product ion spectra acquisition). Upon reanalysis of the well, the same result was obtained. This meant that either there was a problem with the formation of the dried spot on the well, or with the solution itself. However, the decision was taken not to respot the well, as an inspection of the data showed that the PS ratios were deviating considerably from the prepared solutions above 2:1, and therefore further investigation would be required anyway.

The largest geometric standard deviation generated is for the 10:1 (L:H) ratio. The geometric standard deviation of 1.33 was generated from the PS ratios 8.33:1, 12.50:1 and 16.67:1 of the three 10:1 L:H solutions. There is a considerable difference between the ratios obtained from the three solutions. The geometric standard deviation value obtained for the ratios for the 10:1 H:L solutions is smaller (generated from the PS ratios 16.22:1, 15.16:1 and 16.17:1). Based on the values obtained for the six solutions (the three forward (L:H) and three reverse (H:L)), the ratio 8.33:1 could be deemed an outlier, due to the fact that it differs so much from the other five ratios (SD (all six values) = 3.23, SD (five excluding

8.33) = 1.68). This outlier could have been a result of the innate variability which is a characteristic of MALDI spectra,^{192,193} which could have a profound impact on the quantification data obtained. For MALDI-MS analysis, a single mass spectrum was obtained per spot, with multiple MS/MS acquisitions for peptide identification. For dimethyl labelling, the quantification takes place at the MS level, and the quality of the quantification is dependent on the quality of the mass spectrum obtained. MALDI's variability derives from the inherent inhomogeneity of the matrix/analyte spot with sweet spots a characteristic of MALDI analyses.^{101,192,194} Depending on where on the spot the laser hits, spectral quality and signal intensities can be very variable. With traditional MALDI-MS, the analyte is not distributed evenly across the surface of the MALDI spot, which can lead to the variability in different mass spectra obtained from a single MALDI spot.¹⁷⁶ This can impact the peptide (and ultimately protein) ratios generated from the data. This would theoretically be reduced with the SAM technology, as the analytes would be immobilised on the well surface, presumably uniformly.

The 2003 paper by Hsu *et al.* detailed the analysis of dimethyl labelled peptides by MALDI-MS.¹²² The group published a plot detailing the linearity range of the dimethyl labelling method, with the ratio of components in solutions being analysed by MALDI-MS. The plot shows a linear range for a single protein containing four lysine residues for ratios over the range 0.05:1 and 20:1. This is in contrast to the data shown in Figure 5.3, where the relationship is clearly not linear. Hsu *et al.* used a Micromass MALDI-ToF mass spectrometer. Samples were analysed using CHCA in ethanol, acetonitrile and 0.1% trifluoroacetic acid as the matrix solution. A 0.5 M solution of hydrochloric acid was mixed with both the matrix and the sample (1:2:1 v/v/v) prior to deposition of the solution onto the MALDI plate. In the work presented in this thesis, CHCA in acetonitrile was used as the matrix solution. The labelled BSA digest was spotted on the MALDI plate wells 1:1 with the matrix solution. Hsu *et al.* state that as the ratio increases above 1:5 (L:H), the method is less precise relative to small variations. The authors discuss the fact that they perform three measurements, although it is unclear whether this is three MALDI-MS acquisitions of the same well, or three samples analysed once (as in the case of the work in this thesis). The different matrix solution used in the work in this thesis compared to that in the work by Hsu *et al.*, and the inclusion of hydrochloric acid, along with the different mass spectrometers used could all contribute to the different results obtained in the two studies.

As the results obtained by MALDI-MS deviated from the ratios of the standard solutions that had been prepared, particularly as the standard solution ratio moved further away from 1:1, LC-ESI-MS was investigated as a possible alternative to MALDI-MS.

5.2.1.2. LC-ESI-MS analysis

The ratios generated by PS from the MALDI-MS data deviated from the expected ratios when greater than 2:1. The variance between data produced for solutions with the same ratio was quite large in some cases, and the more abundant of the labelled digests was over-represented. In order to determine whether this was a problem with analysing the samples by MALDI-MS, and to determine whether this could have been due to pipetting errors, the same solutions (subjected to a 1 in 100 dilution) were also analysed using LC-ESI-MS.

LC-ESI-MS analysis was controlled using HyStar (Bruker Daltonics), and the data analysed using PS and MD. Two software packages were tested in the event that the problem with the ratios generated was due to the software package. Other software packages (MaxQuant²²⁸, MSQuant²²⁹ and Skyline²³⁰) were researched, to determine what software is available and being used by the proteomics community and which might be suitable for the application described in this thesis. Some software packages were discounted due to incompatibility between the data format and the software. MD was trialled due to the fact that it is compatible with the data files acquired, but also contains a 'built-in' dimethyl quantification method. It is commercially available software (as is PS) which can be used for protein searches and quantification and is an interface with quantification algorithms for use with the locally available Mascot search engine. The data workflows for the two software packages are shown in Figure 5.4 (PS) and Figure 5.5 (MD).

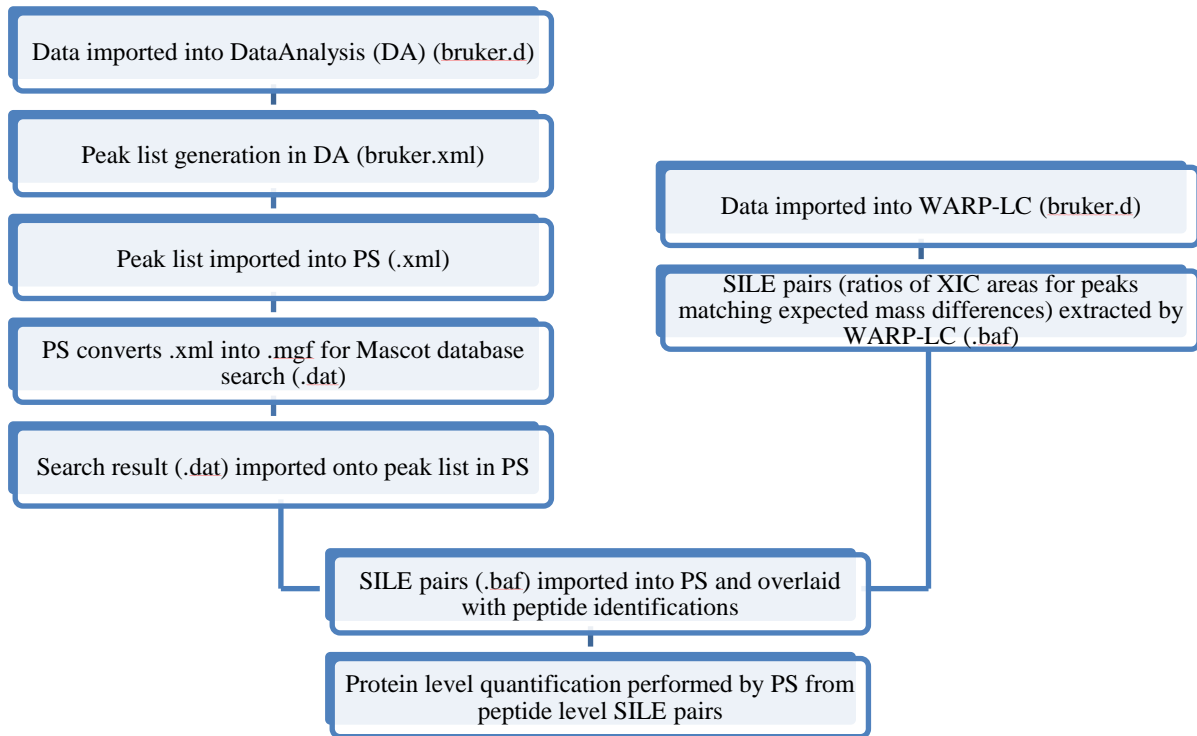


Figure 5.4: Data workflow for data processed through PS, including the data format at each stage

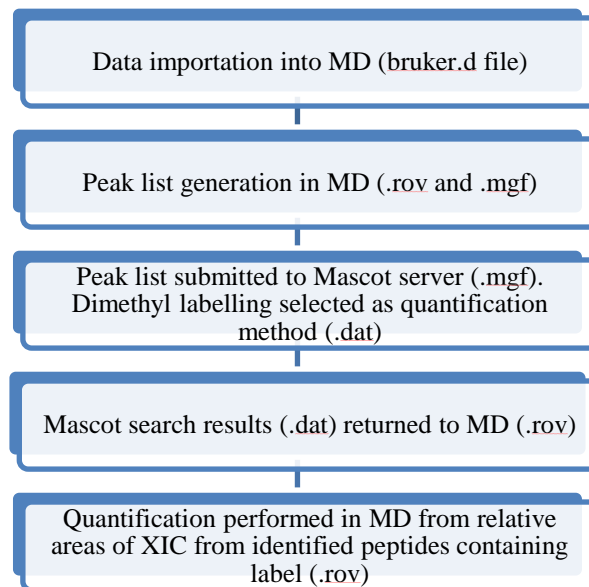


Figure 5.5: Data workflow through MD, including the data format at each stage

Through MD, the MS processing options used were: Un-centroiding (Peak half width: 0.2, Data points per Da: 10); Re-gridding (Data points per Da: 100); Multi-format spectra

(Preferred type: Profile); Peaks (Minimum number: 1, Maximum charge: 4); and Aggregation (Scan group aggregation method: Sum). The MS peak picking processing options used were: Filtering (Correlation threshold: 0.7, Minimum S/N: 10, Minimum peak m/z : 50, Maximum peak m/z : 100,000); Peak profile (Minimum peak width: 0.005 Da, Expected peak width: 0.025 Da, Maximum peak width: 0.125); and General (Baseline correction applied, Fit Method: Isotope distribution, Maximum peak iterations per scan: 500). The same options were used for the MS/MS peak picking processing options. The correlation threshold is a measure of the correlation between the predicted and observed precursor isotope distribution (the higher the correlation threshold, the greater the correlation between the predicted and observed precursor isotope distributions).

Peak picking is not performed using PS, but is performed in DA. The settings used by DA for peak picking were: Sum Peak settings were used in DA; Use the same width as in acquisition; S/N: 1; Relative intensity threshold: 0.001%; Absolute intensity threshold: 20.

The only parameters which are in common between the two peak picking processes are the S/N parameters (MD = 10, DA (PS) = 1). The threshold is set higher for data processed through MD than for data processed through PS. This should mean that less noise is picked up in the peak lists generated through MD than PS, which should lead to 'cleaner' data. Neither software package returns peak areas with the quantification results, meaning that it is not easily possible to compare peak areas obtained for individual peptides using the different software packages.

The data acquired for each solution were submitted to each of the two software packages for analysis. A geometric mean was calculated using the ratios generated by the software for the triplicate analyses, and a geometric standard deviation was generated for each ratio. Table 5.2 shows the ratios generated for the standard solutions using PS or MD software (plotted in Figure 5.6). For PS data, FDR values were in the range of 0.00-2.01%, and for MD data, FDR values were in the range of 0.00-2.13% (Table 7.4).

Table 5.2: Ratios generated for each BSA digest solution (geometric standard deviations (SDgeo) for the average of three replicate solutions per ratio) (number of peptides and sequence coverage for individual solution replicates for each ratio) (see 7.5.6 for FDR values)

Ratio prepared (L:H)	PS				MD			
	Protein ratio	SDgeo	Number of peptides	Sequence coverage / %	Protein ratio	SDgeo	Number of peptides	Sequence coverage / %
0.10:1	0.14:1	1.05	25, 30, 25	39, 52, 41	0.09:1	1.06	55, 54, 56	67, 63, 70
0.20:1	0.28:1	1.11	23, 22, 31	46, 42, 66	0.21:1	1.03	23, 29, 27	34, 43, 41
0.33:1	0.39:1	1.05	25, 32, 20	52, 64, 46	0.32:1	1.03	26, 27, 31	42, 37, 58
0.50:1	0.54:1	1.01	37, 29, 27	68, 60, 57	0.48:1	1.00	42, 31, 28	68, 46, 34
0.67:1	0.77:1	1.08	31, 31, 38	52, 59, 61	0.67:1	1.01	38, 29, 28	54, 47, 36
1:1	1.00:1	1.01	43, 36, 21	48, 57, 21	0.99:1	1.02	48, 42, 24	67, 64, 41
1:1 (H:L)	1.01:1	1.04	14, 22, 19	34, 37, 42	0.99:1	1.01	20, 26, 27	33, 36, 40
1.5:1	1.48:1	1.05	34, 32, 21	62, 54, 37	1.47:1	1.01	36, 37, 41	51, 55, 57
2:1	2.00:1	1.02	37, 25, 23	70, 48, 37	2.01:1	1.01	39, 31, 35	62, 56, 53
3:1	3.15:1	1.27	20, 20, 29	45, 38, 65	3.01:1	1.03	28, 39, 29	49, 59, 42
5:1	3.43:1	1.21	21, 26, 23	34, 43, 44	5.06:1	1.04	30, 29, 36	42, 42, 59
10:1	7.16:1	1.06	28, 28, 26	39, 41, 39	10.91:1	1.07	54, 71, 74	70, 72, 73

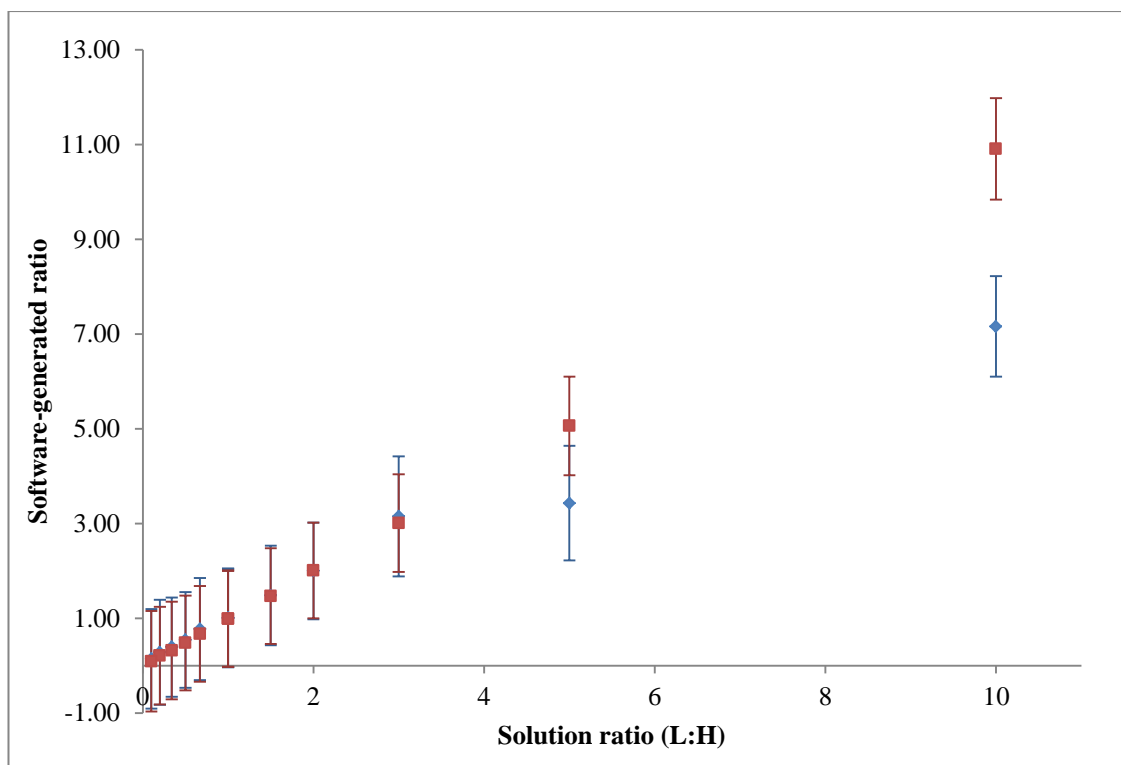


Figure 5.6: Plot showing the geometric mean generated from three replicate solutions analysed by LC-ESI-MS against the prepared ratio (error bars show the geometric standard deviation) (PS = blue diamonds, MD = red squares)

As can be seen in Table 5.2 and Figure 5.6, there is a difference between the ratios obtained using the two different software packages (PS and MD) applied to the same set of measurements. It can be seen in Figure 5.6 that as the ratio deviates from 1:1, the software-generated ratio obtained using PS deviates further from the expected values than with MD. As can be seen in Figure 5.6, with the MD software-generated ratios, the trend is much more linear. It should be noted that there is a difference between the number of peptides used for quantification and the sequence coverage for the two software packages (Table 5.2). For each of the software packages, different database searches were performed. For MD, at the time of the data being analysed it was not possible to import calibrated data into the software (personal communication, John Cottrell, Mascot Science). This meant that for database searches conducted through Mascot for MD, wider tolerances were needed for mass errors than for PS. PS is Bruker software and has the ability to use the lock-mass-corrected data generated from the Bruker maXis instrument. Better calibrated data should give more peptide identifications, meaning a greater number of peptides available for quantification. Therefore, it is unlikely that the search tolerances are the reason for the difference in the number of peptides used for quantification through the two software packages.

With the data analysed in PS, the greater the difference between the amount of the light and heavy labelled peptides, the more the ratio tends to be under-represented. It can be argued that identifying a difference is sufficient and that the exact extent of this is not the issue. As long as the difference in expression of a protein is identified, the numerical value obtained may not necessarily be important: the protein has been identified for further investigation and validation is necessary whatever the number determined from the mass spectrometric results (with the MALDI data presented in Section 5.2.1.1, it is important to note that the number of peptides used for quantification was low, especially bearing in mind that the experiment was performed using a simple protein digest, and raised the possibility of problems with peptide identification and quantification with a more complex mixture). What is important is to ensure that a protein which is differentially expressed between the two samples is called as different and not missed due to an under-representation of the ratio. With under-representation of the ratio, there is a danger of generating a false negative result (i.e. not recording a difference when there is one). With PS, the under-representation of the prepared ratio starts to become apparent at ratios greater than 2:1. Under-representation of ratios is commonly observed for isobaric labelling strategies.^{118,231,232} However, there appears to be little in the literature about labelling strategies, like dimethyl labelling, where quantification occurs using the mass spectral data rather than the tandem data. With under-representation of the ratios, a difference should still be observed (and could then be further investigated), although the real difference in expression could be greater than the difference returned by the software.

In the data generated by MD, the values obtained are much closer to the ratios of the standard solutions prepared. There is also no apparent systematic preference, as there is with the PS results, for the most abundant label to be over- or under-represented, as can be seen from the fact that in some cases the MD ratio is slightly higher or slightly lower than the solution ratio prepared (e.g. 5:1 L:H = 5.07:1 and H:L = 4.76:1). The ratios generated for 10:1 (L:H and H:L) are within 10% of the level in the standard solution prepared, which is much closer than the ratios for the same data from the same samples, determined using PS. The difference must be due to a difference in the software packages. The data generated by MD used, in general, a larger number of peptides to generate the protein ratios than did PS. MD has a built in dimethylation method. However, with PS a method for quantifying the dimethyl data had to be designed, so that the two methods certainly differ in some details. For example, in MD, the H/D elution time shift experienced with dimethyl labelling, although small, is automatically considered. With PS however, a small difference in elution times between the light and heavy labelled peptides had to be included in the created method. It is possible that a difference in the methods used by the two software packages could have resulted in the differing numbers of peptides used for quantification in MD and PS. The two software packages are very different; including in the way the peak lists are generated for the protein database search (as discussed earlier). Based on the fact that MD applies a

higher S/N threshold than PS, it could be assumed that a smaller number of peptides would be available for quantification through MD than through PS. However, this is clearly not the case. Therefore, the difference between the numbers of peptides available for quantification from the two software packages is likely not due to the peak picking parameters (certainly not due to the S/N parameters). However, S/N is only one of the parameters needed for peak identification; others include peak width and isotopic distribution settings. Within MD, there are several acceptance criteria which must be met before a peptide quantification result is accepted, which can be altered by the analyst. These include the correlation threshold (discussed above), the fraction threshold and the standard error threshold. The fraction threshold is the fraction of the peak area in the precursor region of the mass spectrum which is accounted for by the components (i.e. the light and heavy labelled peptides). This parameter is set as part of the quantification method. The standard error threshold is also set within the quantification method, and is the estimated standard error for the calculated ratio. Each peptide ratio is calculated by making a least squares fit to the component intensities from the scans in the extracted ion chromatogram peak.²³³

However, the ultimate difference between the two software packages, leading to the difference in the number of peptides available for quantification (and ultimately the superior quantification results obtained through MD over PS), has proved impossible to identify, and is also beyond the scope of this thesis. It would also be worth waiting until the calibration issue with Bruker data files in MD has been remedied, so that the same search parameters could be used for the database searches (i.e. peptide tolerance, 10 ppm; MS/MS tolerance, 0.1 Da), in order to determine whether this changes the quantification results in MD. It is possible that the tolerances used for database searching through MD and PS may be the reason for the differences in the number of peptides used for quantification. The vast majority of peptides identified through the two searches are the same. However, it is possible that peptides whose peptide and MS/MS tolerances fell just outside of the search tolerance limits would have been lost through PS (whereas these may have been included in MD, therefore accounting for the larger number of peptide quantification ratios available for protein quantification used through MD rather than PS), and therefore there would be fewer peptide quantification ratios available for protein quantification. In order to investigate this, several database searches were repeated with larger tolerances (peptide tolerance, 10 ppm; MS/MS tolerance, 0.1 Da), in order to determine whether the number of peptides identified increased. However, there was no increase in the number of peptides identified. This would suggest that it is not the search parameters which are creating the differences between the number of peptides identified through MD and PS (and subsequently available for peptide quantification), and may be due to the peak picking parameters. It would be interesting to compare the quantification results obtained from calibrated data using MD and PS, with those obtained from the uncalibrated data using MD.

The ratios produced from the data obtained using LC-ESI-MS analysis are much closer to the ratios of the standard solutions prepared than those produced using MALDI-MS analysis. The data suggest that the analysis method (MALDI-MS or LC-ESI-MS) plays a part in the accuracy of relative quantification results generated. The fact that multiple spectra were acquired over a chromatographic peak for LC-ESI-MS analysis, meaning that the quantification ratio can be averaged over multiple spectra, could have resulted in the generated ratios being more accurate for LC-ESI-MS than MALDI-MS. With MALDI-MS data, only one ratio value is generated for the peptide in each measurement. If the result for that peptide varies greatly from the expected ratio for the protein, it is possible that the software may not class the ratio for that peptide as an outlier as it has nothing else to compare it to.

Due to the fact that the ratios generated from the acquired data using MD were much closer to those of the standard solutions than those generated from the acquired data using PS, MD was used in future LC-ESI-MS analysis.

5.2.1.3. Pseudo LC-MALDI-MS analysis

MD cannot process MALDI data when the file consists of a single mass spectrum, and a few tandem mass spectra (a typical MALDI-MS acquisition format). The software is unable to read the file, as it is expecting an LC-type file. LC-MALDI-MS analysis of samples could be performed and the data analysed computationally using MD and PS to allow a comparison of the results. A 'pseudo LC-MALDI-MS' analysis was performed, in order to determine whether LC-MALDI-MS data could indeed be processed using MD. This was done by using a single ZipTip to load a labelled (1:1 L:H) BSA digest, and then fractionating eluting peptides using increasing percentages of acetonitrile in aqueous mixtures. The resulting fractions were then spotted on a pre-spotted LC-MALDI plate (this means that the sample wells are pre-spotted with matrix onto which aliquots of the fractions were spotted and allowed to air dry, and the calibrant and matrix are pre-spotted on the calibrant well). When the data from these fractions were imported into MD, the data could be processed as if the few different ZipTip fractions were LC fractions, and a quantification result was produced (1:0.98). When the same data were imported into PS, a quantification result was not obtained.

5.2.2. Reproducibility of protein digestion and labelling

In a quantitative proteomics experiment the samples to be compared are digested, and then differentially labelled, before being mixed in equal amounts for analysis. For example, a 'healthy' sample could be compared with a 'disease' sample in a forward and reverse

labelling experiment: the samples are distinct and so they are handled and digested separately before being labelled and mixed for analysis. This would enable differences not only in the levels of identical peptides to be revealed, but also differences in the peptides present in the two samples to be identified (e.g. if there were differences in a peptide sequence, only the light or heavy labelled version would be present for detection mass spectrometrically).

Since in the first validation experiment (Section 5.2.1) a single BSA digest was made and split for subsequent labelling and recombining, it was necessary to demonstrate that labelling and combining aliquots of different digests would not affect the measured ratios. For this purpose, two separate aliquots of BSA were digested. One was labelled with light formaldehyde and the other with heavy formaldehyde. The light and heavy labelled digests were then combined, in triplicate, in a 1:1 (v/v) ratio (L:H and H:L). The labelled solutions were analysed by LC-ESI-MS and analysed using MD (Table 5.3). FDR values ranged between 0.00-0.97% (Table 7.5).

Table 5.3: Ratios obtained using MD for LC-ESI-MS analysis of different BSA digest solutions (see Section 7.5.6 for FDR values)

Ratio	Replicate	Protein ratio	SD (geo)
L:H	1	0.94:1	1.101
	2	0.95:1	1.096
	3	0.95:1	1.084
H:L	1	1.02:1	1.119
	2	1.05:1	1.105
	3	1.03:1	1.091

As can be seen for the data in Table 5.3, the ratios obtained through MD consistently reflect the ratios of the prepared 1:1 standard solutions. The geometric standard deviations are all approximately 1.1. These values, combined with the consistency of the MD ratios for the six solutions give confidence in the relative quantification data produced and thus in the reproducibility of the sample handling procedures established.

5.3. Spiking labelled protein digests into labelled *E. coli* lysate digest

Having established that dimethyl labelling could be performed on a simple BSA digest, and acceptable ratios generated over the range 1:0.1 to 1:10, a more complex sample was generated for testing. In a relative quantitative proteomics experiment using dimethyl labelling, differences between peptides, and ultimately proteins, in two samples are identified and analysed. Following labelling, the two samples are combined in a 1:1 total protein weight ratio. On analysis, proteins whose relative levels have not changed (expected to be the majority) remain at 1:1, while proteins whose abundances have changed need to be validated using alternative experiments and methods. Validation can be achieved through using an alternative labelling or label free strategy to determine whether the change in level was also observed using an additional mass spectrometric approach. Alternatively, a non-mass spectrometric analysis technique such as western blotting could be used, to test for changes in abundance of a protein between the two samples. Alternative LC-MS strategies could also be investigated, using orthogonal techniques such as HILIC²³⁴, strong cation exchange²³⁵ or high pH reversed phase chromatography. In the work presented in this thesis, a strategy was employed which did not set a fold change threshold cut-off.

The obvious reason for a measured difference in levels of a particular protein could be due to under- or over-expression of that protein or differential breakdown of it in one of the two samples. However, perhaps less intuitively, it could also be due to a difference in the amino acid sequence of the protein in the two samples. For example, a 'disease' sample could have a mutation in the amino acid sequence of a protein. The mutation could be a change in the amino acid sequence, a PTM or a single nucleotide polymorphism.²³⁶ When analysed, such a situation could result in two separate peptide peaks for which there was no corresponding peak in the other labelled sample: one peak would correspond to the version of the peptide only present in the 'healthy' sample, and the other corresponding to the version of the peptide only present in the 'disease' sample. This would result in a measured difference in the abundance level of the protein between the two samples. Therefore, both a difference in expression and/or breakdown of a protein as well as the possibility of differences in the amino acid sequence of the protein need to be considered as reasons for a measured difference. In order to determine whether proteins whose expression differs from a 1:1 ratio could be detected not only in a single protein digest mixture but also in a complex background, a spiking experiment was performed. The proteins BSA, rabbit muscle glycogen phosphorylase (Phos B), human transferrin (Trans) and horse heart myoglobin (Myo) were digested with trypsin and aliquots labelled with light and heavy formaldehyde. Additionally, a tryptic digestion was performed on an aliquot of *E. coli* lysate (the background proteins), and aliquots labelled with light and heavy formaldehyde.

The light and heavy labelled *E. coli* lysate digests were combined to create a solution with a 1:1 (v/v) L:H ratio, with a total digest concentration of 100 ng/ μ L. Having labelled the additional separate protein digests, the protein digest solutions were combined in known ratios (v/v) (Table 5.4) which differed from the 1:1 ratio of the *E. coli* lysate digest. The labelled protein digest mixtures were then spiked into aliquots of the 1:1 labelled *E. coli* lysate digest, at either 1% (i.e. 1 ng/ μ L) or 5% (5 ng/ μ L) of the total *E. coli* lysate digest amount. As with all other experiments, labelling was carried out to allow both the forward and reverse experiments to be performed (L:H and H:L), and each spiked solution was prepared in triplicate.

Table 5.4: Proteins used for spiking into *E. coli* lysate digest, and the ratios (L:H and H:L) they were prepared at

Protein	Ratio
BSA	2:1
Trans	3:1
Phos B	4:1
Myo	5:1

Following LC-ESI-MS analysis, the data were processed using MD. The number of proteins identified and quantified in each of the solutions was extracted from the data (Table 5.5). FDR values ranged from 0.15-0.91% (Table 7.6).

Table 5.5: Number of proteins identified and quantified for each of the L:H and H:L solutions prepared with proteins spiked in at 1% and 5% of the total *E. coli* lysate digest amount (see Section 7.5.6 for FDR values)

Solution	Replicate	Number of proteins identified	Number of proteins quantified
L:H 1%	1	427	245
	2	436	234
	3	427	251
L:H 5%	1	423	237
	2	423	228
	3	401	239
H:L 1%	1	423	246
	2	436	235
	3	423	236
H:L 5%	1	423	220
	2	412	161
	3	436	236

As can be seen in Table 5.5, the number of proteins identified and the number of proteins quantified were similar for the 1% and 5% spiked solutions. The number of proteins quantified across the three replicate solutions for each condition (i.e. L:H with proteins spiked in at 1% of the total *E. coli* lysate digest amount) are shown in Venn diagrams in Figure 5.7 to Figure 5.10.

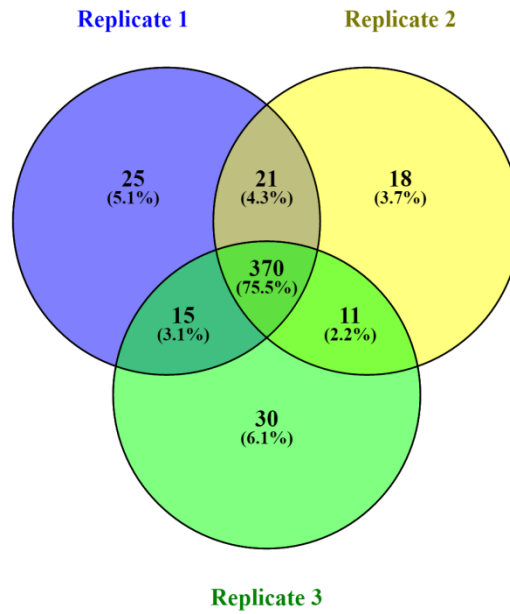


Figure 5.7: Venn diagram showing the number of proteins quantified across all three replicates for L:H solutions with proteins spiked in at 1% of the total *E. coli* lysate digest amount (Venn diagram produced using Venny²³⁷).

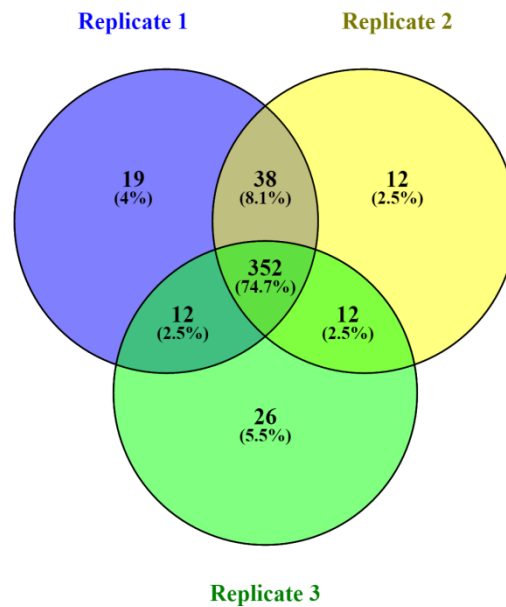


Figure 5.8: Venn diagram showing the number of proteins quantified across all three replicates for L:H solutions with proteins spiked in at 5% of the total *E. coli* lysate digest amount (Venn diagram produced using Venny²³⁷).

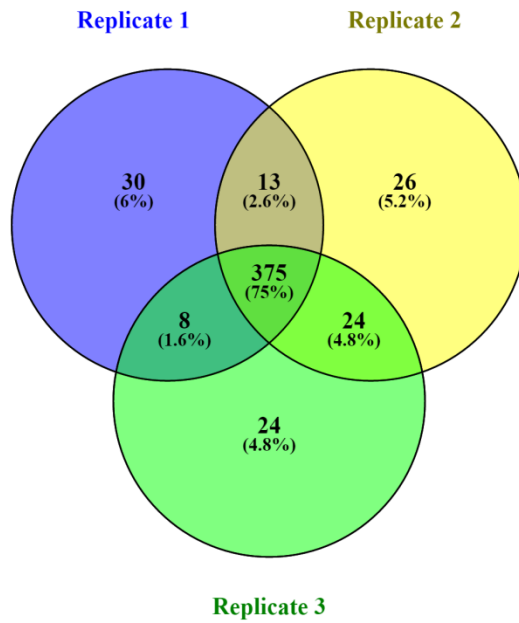


Figure 5.9: Venn diagram showing the number of proteins quantified across all three replicates for H:L solutions with proteins spiked in at 1% of the total *E. coli* lysate digest amount (Venn diagram produced using Venny²³⁷).

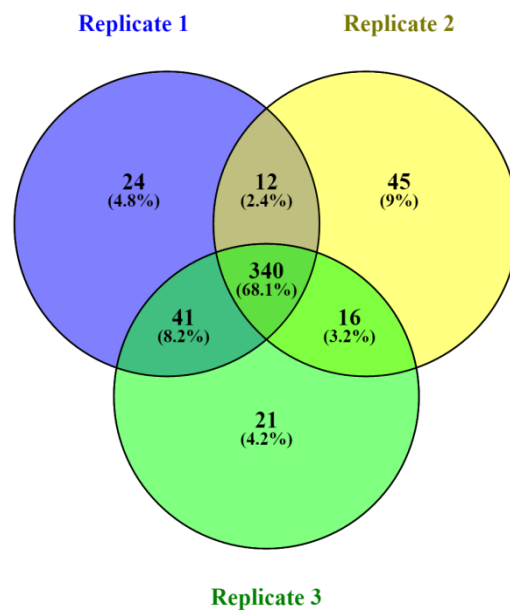


Figure 5.10: Venn diagram showing the number of proteins quantified across all three replicates for H:L solutions with proteins spiked in at 5% of the total *E. coli* lysate digest amount (Venn diagram produced using Venny²³⁷).

As can be seen in Figure 5.7 to Figure 5.10, the number of proteins quantified across all three replicate solutions is high for each condition. The high number of commonly quantified

proteins increases confidence in the results as it shows replicate solutions have a high degree of inter-sample reproducibility when considering MS/MS sampling.

In order to perform a statistical analysis of the spiked data, further interpretation of the data was performed. There are important considerations when performing statistical analysis of data sets. These include the multiple testing problem, which states that if a test is performed multiple times across a data set, the likelihood of random errors increases.²³⁸ Therefore, a number of results will falsely be declared positive results. In order to combat this, researchers have suggested a number of corrections, with the Bonferroni correction²³⁹ and Hochberg and Benjamini correction²⁴⁰ being two commonly used tests in proteomics.

The false discovery rate (FDR) is an important metric when quoting the results of database searches. It is a measure of the proportion of errors committed by falsely rejecting the null hypothesis.²⁴⁰ The null hypothesis for a database search is that when assigning a significance measure to a match between a peptide sequence and a tandem mass spectrum, the peptide was not identified by the mass spectrometer.²⁴¹

The p-value is the smallest level of significance at which the null hypothesis would be rejected (i.e. the probability that the null hypothesis is not true).²⁴²⁻²⁴⁴ It is a statistical measure used to determine a level of significance for a result. The q-value is similar to the p-value, except it is a measure of significance in terms of the FDR.²⁴⁵

The Bonferroni correction sets the significance (q-value) at (significance level)/(number of tests) (Equation 5.4).^{238,242} For example, with 20 tests, and a significance level of 0.05, a null hypothesis would be rejected if the p-value was less than 0.0025. There are those who consider the Bonferroni correction method to be controversial.^{239,246} Perneger published an article in 1998 detailing what the author highlighted as problems with the Bonferroni correction. Firstly, that the general null hypothesis (that all null hypotheses are true simultaneously) is rarely of interest or use to researchers. Secondly, that the interpretation of a finding depends on the number of other tests performed; and finally, the likelihood of incorrect retention of a false null hypothesis (a false negative result) is increased.²⁴⁶

$$q = \frac{S}{N}$$

Equation 5.4: Bonferroni correction (S = significance level, N = number of tests)

Benjamini and Hochberg published a paper in 1995, detailing a way of controlling the FDR.²⁴⁰ The p-values are ordered from smallest to largest, and then given a rank from one to 'n'. The Benjamini and Hochberg correction is an estimation of the FDR, and is calculated by multiplying the p-value by the number of tests, and dividing this number by the p-value

rank (Equation 5.5). The calculated FDR is an estimated FDR in the list of 'significant' results above a threshold (a maximum desired FDR), rather than of the global data set.

$$q = \frac{p \times N}{rank}$$

Equation 5.5: Benjamini and Hochberg correction (N = number of tests)

For each solution data set, the peptide-level data were extracted. For each protein where the minimum number of peptides used for protein quantification was three, the p-value was calculated using the corresponding peptide data. Once p-values had been calculated for each of the protein results, a Hochberg and Benjamini multiple test FDR estimation was then applied to the p-values. The corrected p-values (q-values) were then used, in combination with the protein quantification ratios, to generate volcano plots^{247–249} (Figure 5.11 to Figure 5.14). A minimum number of peptides of three was required for the spiked data, as initial interrogation of the data showed a large number of false positive quantification results obtained when q-values were included for proteins quantified using two peptide ratios. In general, a minimum of two false positive quantification results were obtained when the minimum number of two peptide ratios were used to generate the protein quantification ratio. As this is the model data for which the answer is known, the decision was taken to exclude protein ratios generated using two peptide ratios.

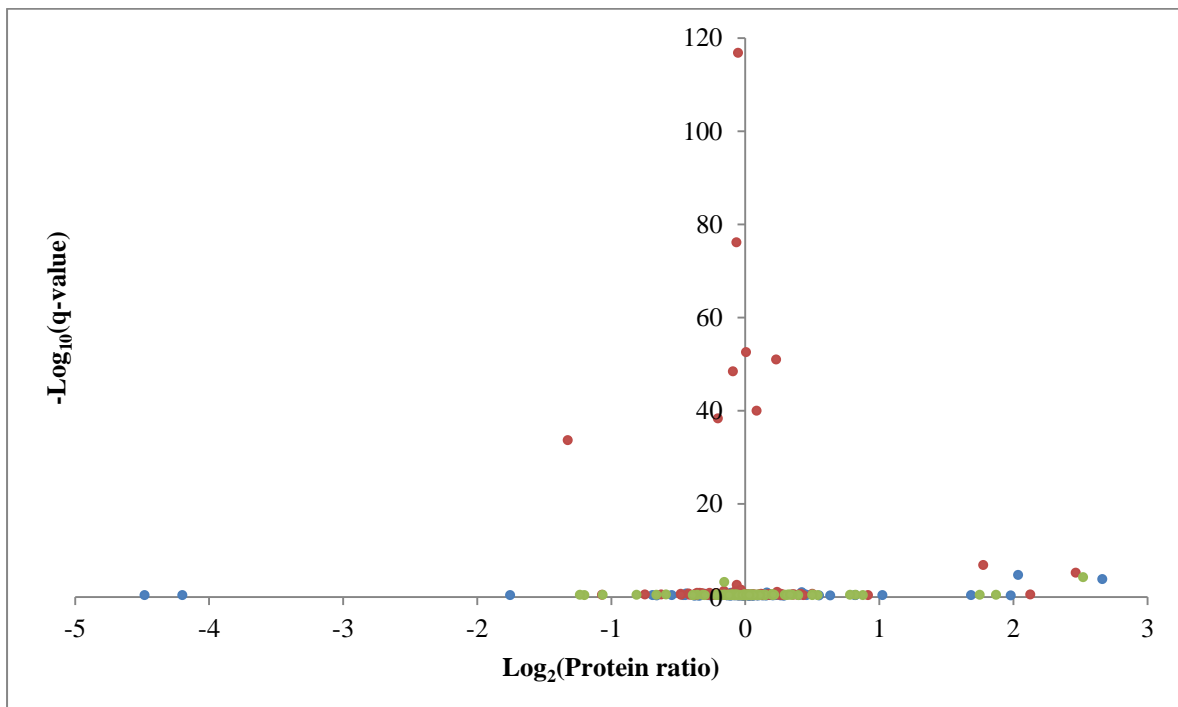


Figure 5.11: Volcano plot showing $-\log_{10}(\text{q-value})$ against $\log_2(\text{protein ratio})$ for proteins spiked in at 1% of the total *E. coli* lysate digest amount (L:H) (Blue = solution 1, red = solution 2, green = solution 3)

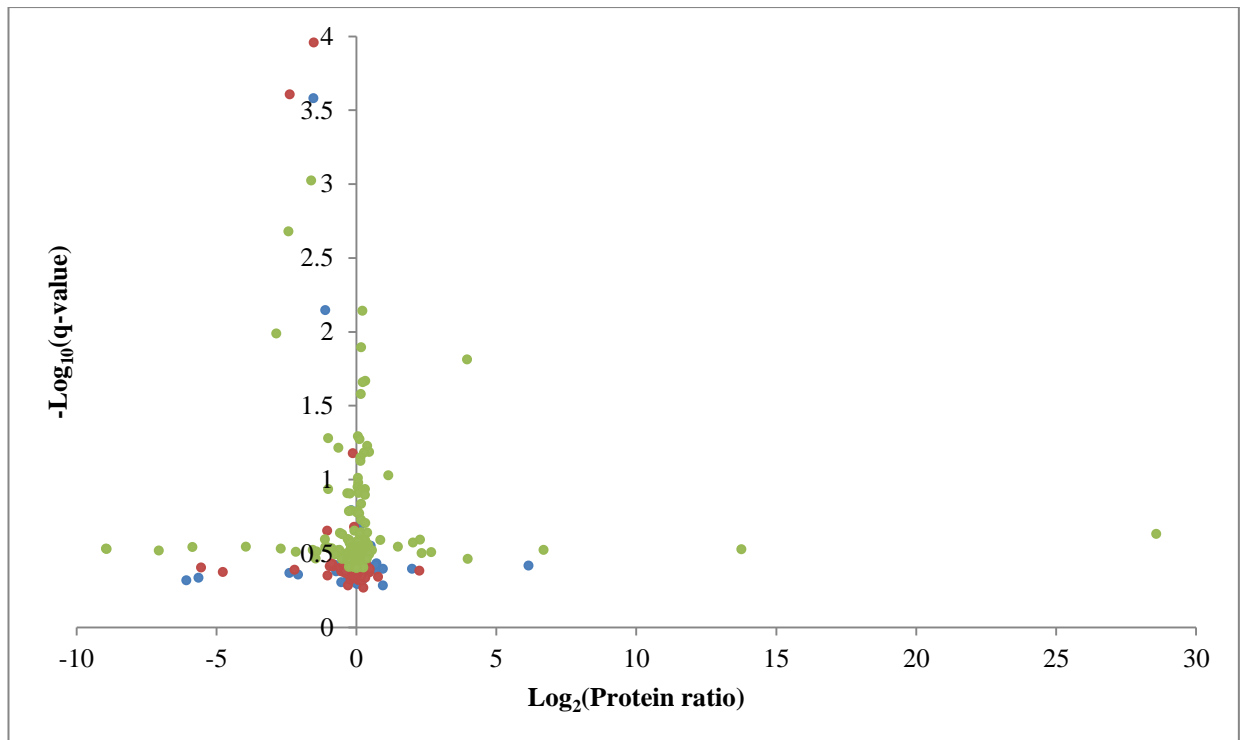


Figure 5.12: Volcano plot showing $-\log_{10}(\text{q-value})$ against $\log_2(\text{protein ratio})$ for proteins spiked in at 1% of the total *E. coli* lysate digest amount (H:L) (Blue = solution 1, red = solution 2, green = solution 3)

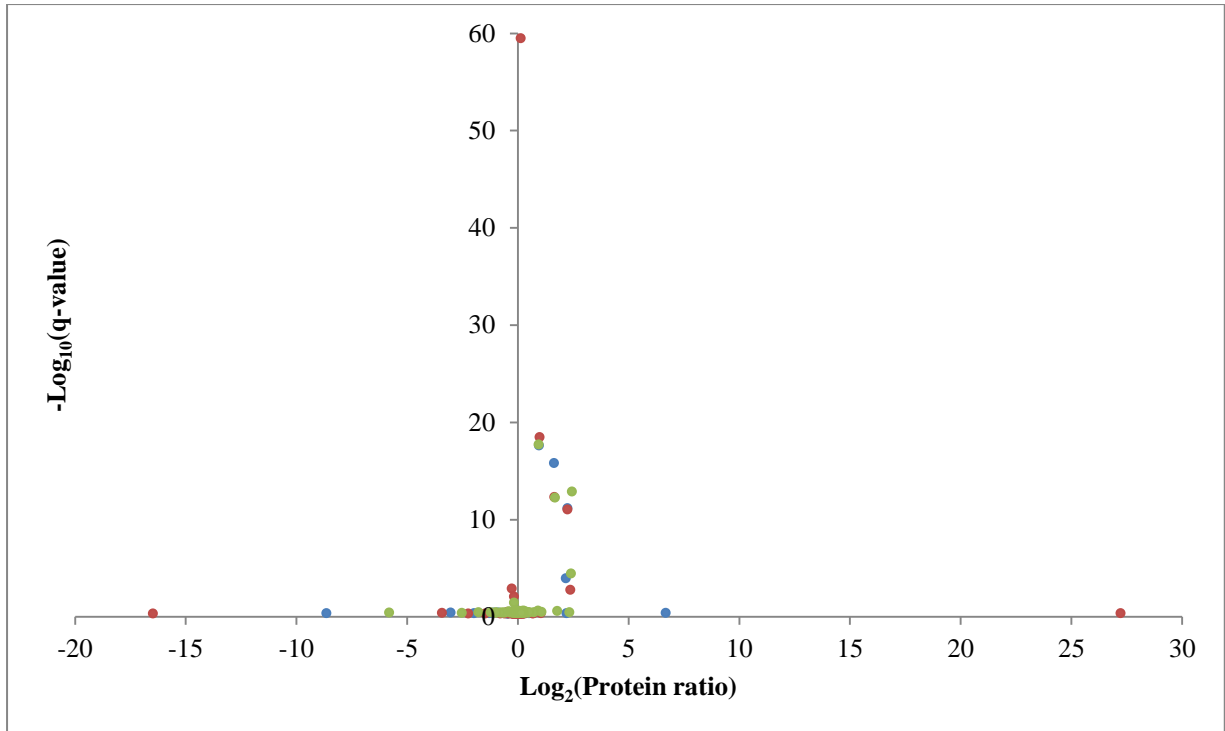


Figure 5.13: Volcano plot showing $-\log_{10}(\text{q-value})$ against $\log_2(\text{protein ratio})$ for proteins spiked in at 5% of the total *E. coli* lysate digest amount (L:H) (Blue = solution 1, red = solution 2, green = solution 3)

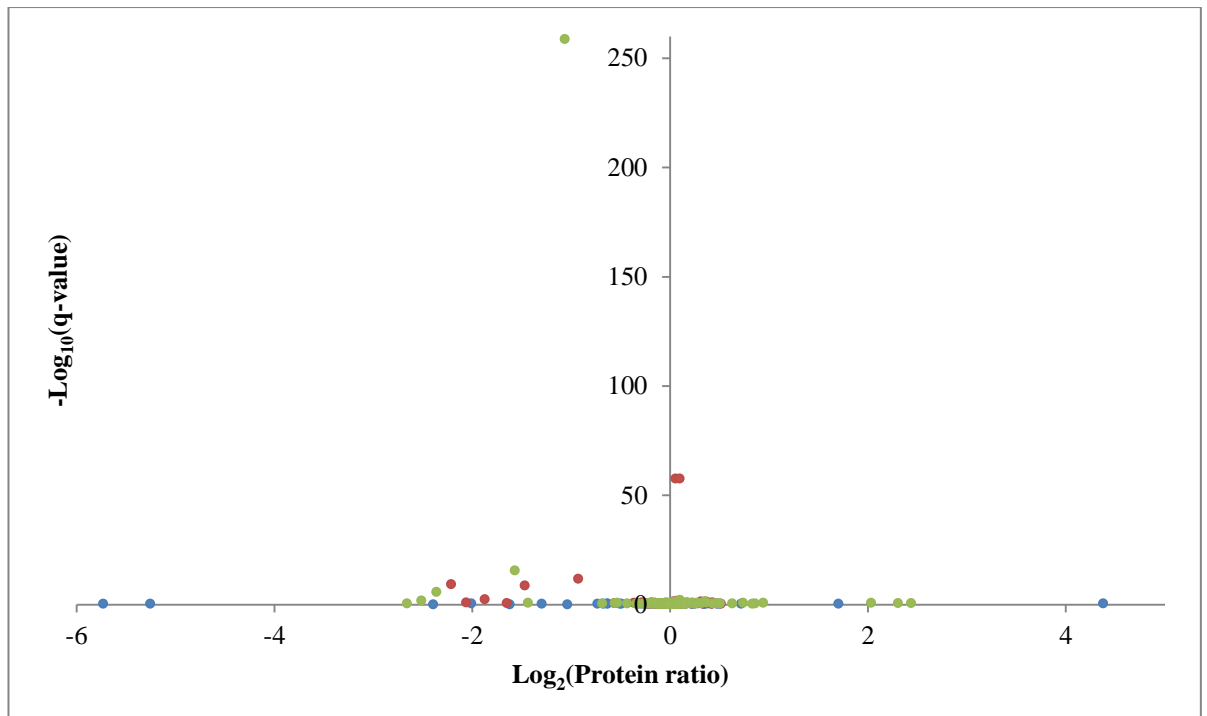


Figure 5.14: Volcano plot showing $-\log_{10}(\text{q-value})$ against $\log_2(\text{protein ratio})$ for proteins spiked in at 5% of the total *E. coli* lysate digest amount (H:L) (Blue = solution 1, red = solution 2, green = solution 3)

The data in Figure 5.11 to Figure 5.14 show that for each set of solutions (e.g. proteins spiked in at 5% of the total *E. coli* lysate digest amount (H:L)), the plots are very similar, as the majority of points overlap between the three replicate solutions. This emphasises the inter-sample reproducibility.

Proteins with a q-value of less than 0.05 were deemed to show a statistically significant difference between samples. These proteins are listed for solutions with proteins spiked in at 1% of total *E. coli* lysate digest amount (Table 5.6) and 5% of the total *E. coli* lysate digest amount (Table 5.7).

Table 5.6: Table showing the proteins with a q-value of less than 0.05 for proteins spiked into the *E. coli* lysate digest at 1% of the total lysate digest amount

L:H or H:L	Solution	Protein	q-value
L:H	1	Trans	1.99×10^{-7}
		Myo	1.57×10^{-4}
	2	EG10241	1.60×10^{-117}
		EG10805	7.72×10^{-77}
		EG20173	3.11×10^{-53}
		EG12316	1.16×10^{-51}
		EG10022	3.75×10^{-49}
		EG10906	1.03×10^{-40}
		EG10586	5.02×10^{-39}
		EG10770	2.44×10^{-34}
		Trans	1.43×10^{-7}
		Myo	5.92×10^{-6}
		EG12204	2.42×10^{-3}
		EG10576	2.86×10^{-2}
	EG10421	4.93×10^{-2}	
	3	Myo	5.41×10^{-5}
EG10586		6.39×10^{-4}	
H:L	1	Trans	2.64×10^{-4}
		BSA	7.15×10^{-3}
	2	Trans	1.10×10^{-4}
		Myo	2.48×10^{-4}
	3	Trans	9.47×10^{-4}
		Myo	2.09×10^{-3}
		EG10953	7.23×10^{-3}
		EG10699	1.03×10^{-2}
		EG11581	1.28×10^{-2}
		EG10158	1.54×10^{-2}
		EG10794	2.15×10^{-2}
EG10256		2.19×10^{-2}	
EG10497	2.65×10^{-2}		

Table 5.7: Table showing the proteins with a q-value of less than 0.05 for proteins spiked into the *E. coli* lysate digest at 1% of the total lysate digest amount

L:H or H:L	Solution	Protein	q-value
L:H	1	BSA	2.52×10^{-18}
		Trans	1.68×10^{-16}
		Myo	7.03×10^{-12}
		Phos B	1.10×10^{-4}
	2	EG10895	3.42×10^{-60}
		BSA	3.47×10^{-19}
		Trans	5.04×10^{-13}
		Myo	9.24×10^{-12}
		EG10550	1.25×10^{-3}
		Phos B	1.74×10^{-3}
		EG10576	8.47×10^{-3}
	3	BSA	1.93×10^{-18}
		Myo	1.31×10^{-13}
		Trans	5.65×10^{-13}
		Phos B	3.55×10^{-5}
EG10420		3.50×10^{-2}	
H:L	1	Trans	5.86×10^{-7}
		Myo	1.40×10^{-4}
		BSA	5.29×10^{-3}
	2	EG10789	2.68×10^{-58}
		EG11003	2.75×10^{-58}
		BSA	1.68×10^{-12}
		Myo	6.85×10^{-10}
		Trans	1.96×10^{-9}
		Phos B	4.30×10^{-3}
		EG10599	2.65×10^{-2}
	3	BSA	1.80×10^{-259}
		Trans	2.80×10^{-16}
		Myo	2.09×10^{-6}
		EG10870	9.88×10^{-3}
		Phos B	1.80×10^{-2}
EG10347		3.72×10^{-2}	

As can be seen in Table 5.6 and Table 5.7, the number of false positive quantification results (i.e. *E. coli* lysate proteins which should have been present at 1:1, and deviate from this) is higher when the proteins were spiked at 1% of the total *E. coli* lysate digest amount. A total of 19 false positive quantification results were obtained for solutions with the proteins spiked in at 1% of the total *E. coli* lysate digest amount. This is compared with the nine false positive quantification results that were obtained for solutions with the proteins spiked in at 5% of the total *E. coli* lysate digest amount. No *E. coli* lysate protein was identified as a false positive quantification result in both the forward and reverse labelled solutions. It would be expected that a true positive change in ratio would be evident in both the forward and reverse labelling data sets and in more than one replicate. That is not the case for any of the false positive quantification results generated from the *E. coli* lysate proteins.

For the solutions with proteins spiked in at 5% of total *E. coli* lysate digest amount, in the L:H labelling all four of the spiked proteins generated q-values of less than 0.05. In the reverse labelled solutions (H:L), for one of the solutions only three of the four spiked proteins generated q-values of less than 0.05. In this instance, the protein with the non-statistically significant q-value was Phos B (q-value = 0.27). For solution 1, a total of three peptides was used for generation of the p-value, and subsequent q-value. It is possible that the small number of peptides used for protein quantification, and subsequent q-value determination, is the reason for the non-statistically significant q-value. This is a false negative quantification result, meaning that a change in protein levels has not been detected in this instance.

The main difference between the solutions with proteins spiked in at 1% and 5% of the total *E. coli* lysate amount is the fact that a protein quantification ratio for Phos B was not obtained when Phos B was spiked in at 1% of the total *E. coli* lysate amount. Quantification of Phos B was not possible from the LC-ESI-MS data obtained either because only one peptide could be quantified (in which case MD does not include this), or no peptides could be quantified. Phos B was the not the smallest of the four proteins which were spiked into the *E. coli* lysate digest. Therefore, the reason for the lack of peptide quantification ratios for Phos B when the protein was spiked in at 1% of the total *E. coli* lysate digest amount is not due to a lack of peptides being produced in the digest. It is likely to be due to the fact that it has been observed within the group that Phos B does not ionise as well by ESI as BSA, Trans and Myo ionise (personal communication with Adam Dowle, Department of Biology). Phos B was identified during the Mascot search in all instances, but quantification was not possible. In the PRIDE database²⁵⁰, Project PXD000001 contains data from a TMT spike experiment. In that experiment, Phos B was spiked in to an equimolar *Erwinia carotovora* lysate with varying proportions in each channel of quantification.²⁵¹ A total of 44 peptides were identified for Phos B in the spiked experiment. This is compared with 24 peptides for BSA, 31 peptides for yeast enolase and two peptides for bovine cytochrome C. These

results are in contrast to the data obtained during the dimethyl labelling experiment presented in this thesis. However, whilst the dimethyl labelling experiments were analysed using a maXis q-TOF, the TMT spiking experiment was analysed using an LTQ Orbitrap velos. The different mass spectrometers have different mass analysers and different source designs, will have been tuned differently (not least because of their different designs) and will likely have different responses for some of the same peptides analysed on the two instruments, which may lead to the differing results. The 'background proteins' used in the dimethyl labelling experiment and TMT experiments are also different (*E. coli* lysate versus *Erwinia carotovora* lysate), which could have led to the differences observed for the number of Phos B peptides identified in the two sets of experiments.

Another PRIDE database entry (Project PXD000288) details the results of a deep skeletal muscle proteome investigation in mice.²⁵² Nearly complete sequence coverage was obtained for Phos B, which was found to be one of the 10 most abundant proteins, which in total make up 50% of the total protein mass in skeletal muscle. The analysis was performed using a Q Exactive mass spectrometer. The fact that Phos B was one of the most abundant proteins observed, and the fact that the analysis was performed on a Q Exactive could both be reasons for the high sequence coverage (i.e. a large amount of the protein is known to increase the chance of peptides from Phos B being selected for analysis, and the type of mass spectrometer could also affect the response of different peptides).

However, there are instances where Phos B has been recorded as having the lowest sequence coverage in a spiking experiment. In 2005, Silva *et al.* published results from an absolute quantification spiking experiment using six proteins spiked in to an *E. coli* protein digest.²⁵³ Of the six proteins spiked into the *E. coli* protein digest, Phos B generally had the lowest sequence coverage of the six spiked proteins. This analysis was performed on a q-ToF Ultima API. The fact that a q-ToF was used, along with the spiking-nature of the experiment means that this experiment, and its results, are more in line with the nature of the dimethyl experiments performed in this work.

The fact that Phos B was not quantified using the data obtained from the proteins spiked in at 1% of the total *E. coli* lysate digest amount, but was always quantified using the data obtained from the proteins spiked in at 5% of the total *E. coli* lysate digest amount is likely to be due to the fact that more of the protein was spiked in to the *E. coli* lysate digest for the 5% spiked solutions, and therefore more would be present in the solution.

In all cases for each of the proteins spiked in at 1% of the total *E. coli* lysate digest amount, the spiked protein which was quantified using the greatest number of peptides is BSA. The same is observed in the plots showing the proteins spiked in at 5% of the total *E. coli* lysate digest amounts apart from L:H replicate 3. This is not due to the fact that BSA is a larger protein than the other proteins used for spiking (in fact it is the second smallest). It is

therefore likely that the peptides produced from a BSA digest are particularly well ionised in this LC-ESI-MS system.

The data suggest that dimethyl labelling with LC-ESI-MS analysis and MD for quantification can be used for reliable quantification. The spiked proteins were identified as statistically significantly different from the data from all of the solutions where the proteins were spiked in at 5% of the total *E. coli* lysate digest amount. However, using the data obtained from the proteins spiked in at 1% of the total *E. coli* lysate digest amount, Phos B could not be quantified. Therefore, it is possible that in an experiment with two unknown samples, if the abundance of the protein which is changing is low (e.g. around 1% of the total amount of protein in the sample) or it ionises poorly, then the change may not be detected.

5.4. Applying dimethyl labelling to relative quantification of muscle protein samples

Having established that spiked proteins could be identified as statistically significantly different in a complex background (principally at 5% of the total background amount), and validated in-solution dimethyl labelling for relative quantification, the method was applied to a quantitative proteomics experiment with 'real' samples. A duplex experiment was performed using mouse-derived skeletal muscle cell samples (undifferentiated C2C12 myoblasts), prepared by Elliot Jokl (Department of Biology). After stretching, the cells were washed and scraped in PBS. Three protein extracts were prepared: one from unstretched muscle (U), and two from stretched muscle (stretched 1 (S1) and stretched 2 (S2)), where the unstretched muscle sample was to be compared with the two stretched muscle samples. Both stretched muscle samples were subjected to dynamic (continuous) stretching for three hours. The aim of the experiment was to determine whether there was a difference between levels of any of the proteins in extracts of stretched and unstretched muscle cells. At the time of the analysis, it was not known whether there are any differences in protein abundances between the unstretched and stretched samples. The protein extracts were subjected to an in-gel tryptic digestion, performed by Rachel Bates, prior to dimethyl labelling, as the samples were also analysed using label-free quantification (analysis performed by Adam Dowle; those results are not included in this work).

Aliquots of the unstretched muscle and stretched muscle proteins were labelled with both the light and heavy formaldehyde labels (in order to perform the forward and reverse experiments). The unstretched muscle sample was compared with each of the stretched muscle samples. Aliquots of these now-labelled digests were combined in a 1:1 (v/v) ratio (L:H and H:L) and analysed by LC-ESI-MS.

The number of proteins identified and quantified in each of the solutions was extracted from the data (Table 5.8). FDR values ranged from 0.00-0.30% (Table 7.7).

Table 5.8: Number of peptides identified and quantified across the muscle protein samples (see 7.5.6 for FDR values)

Solution	Replicate	Number of proteins identified	Number of proteins quantified
U:S1	L:H	455	142
	H:L	459	164
U:S2	L:H	492	152
	H:L	582	210

As can be seen in Table 5.8, the number of proteins identified and the number of proteins quantified were similar between the L:H and H:L unstretched:stretched 1 solutions. For the L:H and H:L unstretched:stretched 2 solutions, 90 more proteins were identified, and 58 more proteins were quantified from analysis of the H:L solution than the L:H solution. It is possible that for unstretched:stretched 2, the differences between the number of proteins identified and quantified for the L:H and H:L experiments could be due to systematic errors due to labelling, errors in the reproducibility of preparing the solutions, or errors arising from overlapping isotopic envelopes. The difference between the L:H and H:L experiments for unstretched:stretched 2 highlights the need for sample replicates and replicate analyses being carried out, to increase coverage and help to identify anomalous results.

The number of proteins identified and quantified across the L:H and H:L (unstretched:stretched 1 and unstretched:stretched 2) solutions are shown in Venn diagrams in Figure 5.15 to Figure 5.18.

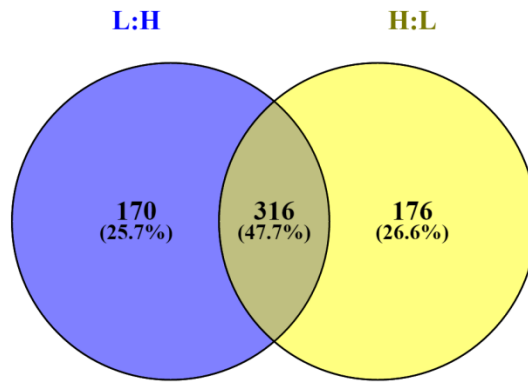


Figure 5.15: Venn diagram showing the number of proteins identified in the unstretched:stretched 1 L:H and H:L experiments (Venn diagram produced using Venny²³⁷)

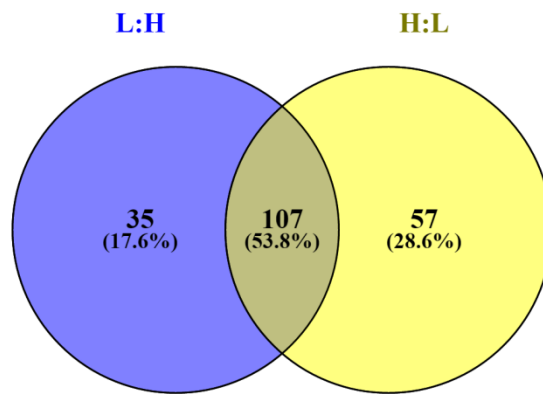


Figure 5.16: Venn diagram showing the number of proteins quantified in the unstretched:stretched 1 L:H and H:L experiments (Venn diagram produced using Venny²³⁷)

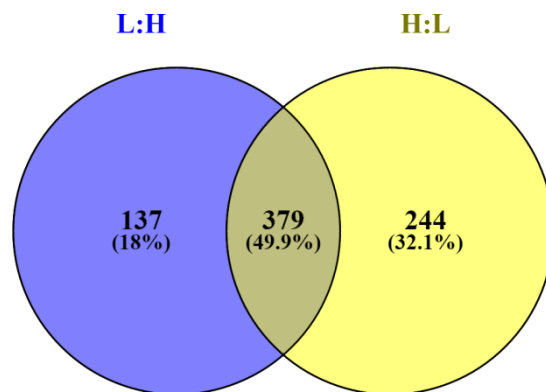


Figure 5.17: Venn diagram showing the number of proteins identified in the unstretched:stretched 2 L:H and H:L experiments (Venn diagram produced using Venny²³⁷)

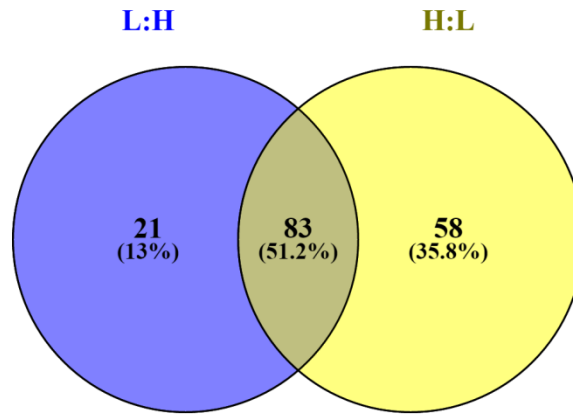


Figure 5.18: Venn diagram showing the number of proteins quantified in the unstretched:stretched 2 L:H and H:L experiments (Venn diagram produced using Venny²³⁷)

As can be seen in Figure 5.15 to Figure 5.18, the overlap between the proteins from unstretched and stretched samples is approximately 50% for proteins identified and proteins quantified.

With the *E. coli* lysate digest experiments, known amounts of protein were digested, so that it was easy to prepare a 1:1 (v/v) solution of the *E. coli* lysate digest and have confidence that it was indeed a 1:1 mixture. With the muscle samples, an unknown amount (approximately 100-200 ng) of protein was digested for each of the three samples. Therefore, although a 1:1 (v/v) solution was prepared, there was no guarantee that equal amounts of light and heavy labelled proteins were combined.

Due to the inability to ensure direct comparability of protein amounts being used for each digest (100-200 ng), a global normalisation was applied. It was expected that the vast majority of proteins would be unchanged in level between the two samples, and should therefore be present in equal amounts (i.e. give a ratio of 1:1). In MD, the 'median' normalisation option was used. MD's median option applies a correction factor to each peptide ratio such that the median of the ratios for all peptide matches that pass the quality test is 1:1.

As was the case with the *E. coli* lysate digest spiking experiments, the data were interrogated at the peptide level. For each protein where the protein ratio was generated using a minimum of three peptide ratios, a p-value was generated from the peptide ratio data. The p-values were then converted to q-values using the Hochberg and Benjamini multiple test FDR estimation. The resulting q-values were used along with the protein ratios to generate volcano plots (Figure 5.19 and Figure 5.20).

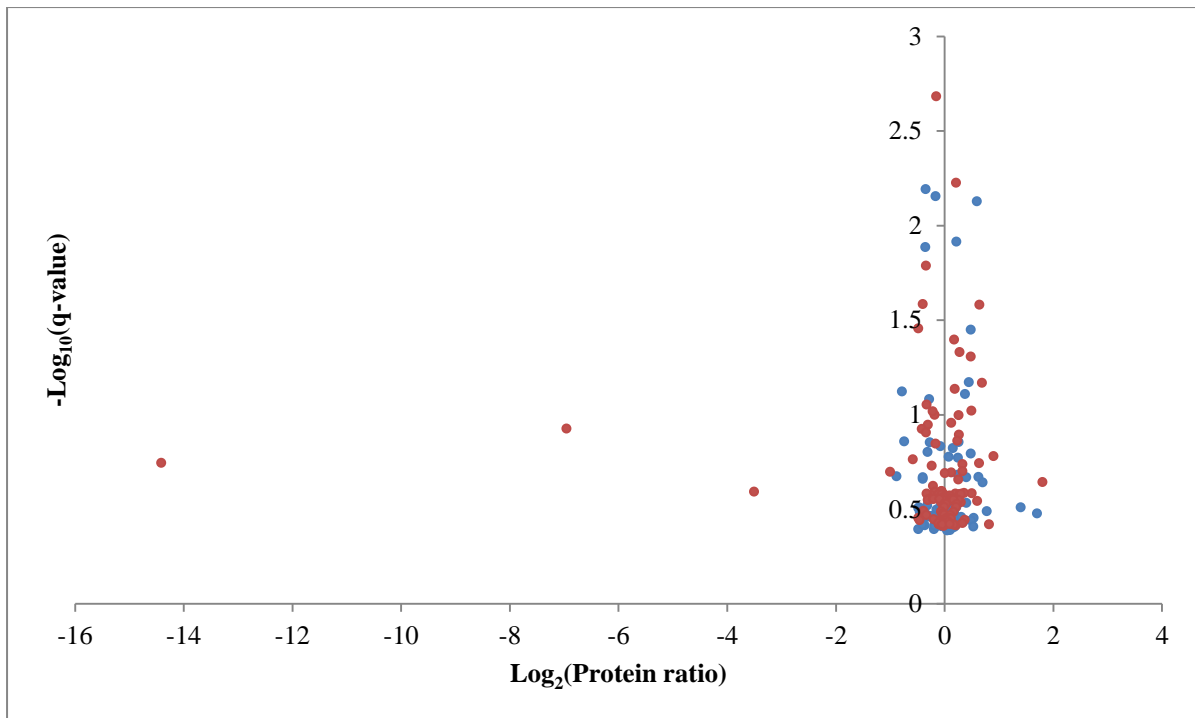


Figure 5.19: Volcano plot showing $-\log_{10}(\text{q-value})$ against $\log_2(\text{protein ratio})$ for unstretched:stretched 1 (Blue = L:H, red = H:L)

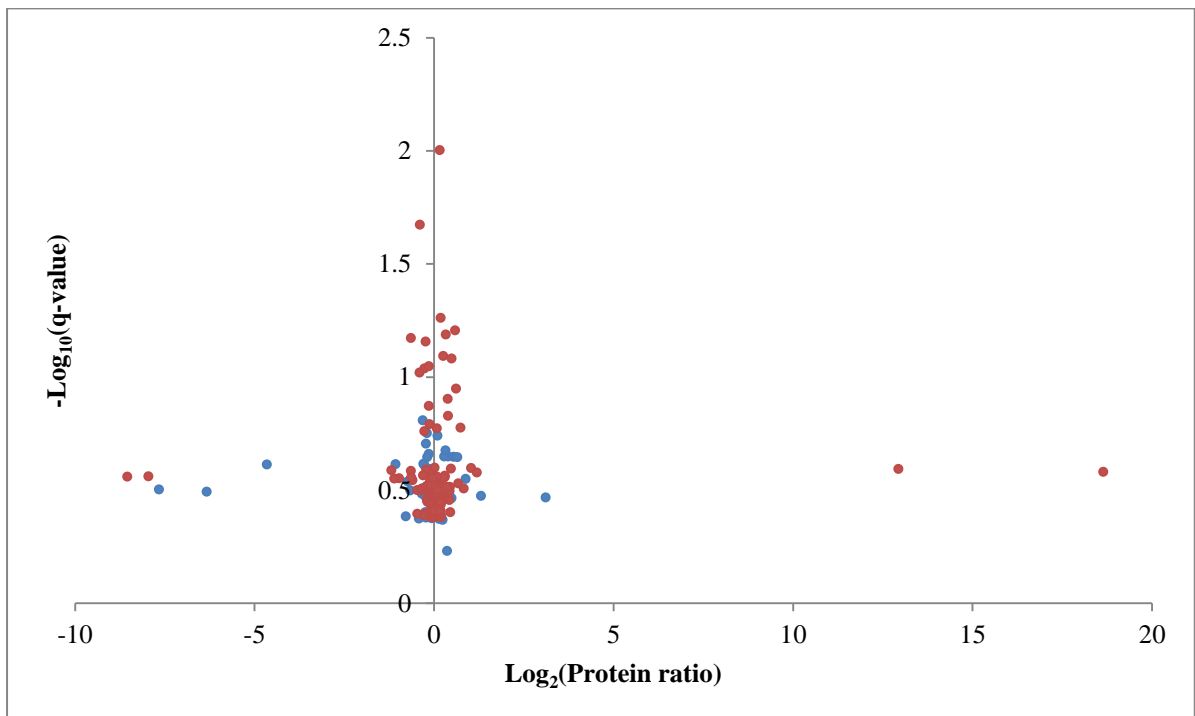


Figure 5.20: Volcano plot showing $-\log_{10}(\text{q-value})$ against $\log_2(\text{protein ratio})$ for unstretched:stretched 2 (Blue = L:H, red = H:L)

Figure 5.19 shows a wider spread of data points along the x-axis for the H:L solution than the L:H solution for unstretched:stretched 1. However, overall for both the unstretched:stretched 1 and unstretched:stretched 2 plots, the L:H and H:L plots look similar in the sense that the majority of points fall along the y-axis (as is expected).

The same parameters as were used in the *E. coli* lysate spiking experiments (a minimum of three peptides used for quantification, and a significance level of $q < 0.05$) were applied to the muscle protein data. The proteins identified as having a significant q-value are detailed in Table 5.9.

Table 5.9: Table showing proteins with a q-value < 0.05 for muscle protein samples

	L:H			H:L		
	Protein	q-value	Protein ratio	Protein	q-value	Protein ratio
Unstretched :stretched 1	6-phosphogluconate dehydrogenase, decarboxylating	7.46×10^{-3}	1.5120	Desmin	2.08×10^{-3}	0.9000
	Phosphoglycerate kinase 1	6.44×10^{-3}	0.7865	Aldehyde dehydrogenase, mitochondrial	5.95×10^{-3}	1.1600
	60S acidic ribosomal protein P0	7.01×10^{-3}	0.8928	Myosin-9	1.63×10^{-2}	0.7894
	60S ribosomal protein L13	1.30×10^{-2}	0.7834	Annexin A1	2.61×10^{-2}	0.7572
	Transitional endoplasmic reticulum ATPase	1.22×10^{-2}	1.1660	Phosphoserine aminotransferase	2.63×10^{-2}	1.5630
	Tubulin beta-4B chain	3.56×10^{-2}	1.4020	Fructose-bisphosphate aldolase A	4.01×10^{-2}	1.1300
				Chloride intracellular channel protein 1	3.50×10^{-2}	0.7173
				40S ribosomal protein SA	4.93×10^{-2}	1.3970
				ATP synthase subunit alpha, mitochondrial	4.69×10^{-2}	1.2140
				Heat shock protein HSP 90-beta	9.94×10^{-3}	1.1190
Unstretched :stretched 2			Desmin	2.13×10^{-2}	0.7657	

As can be seen in Table 5.9, more proteins were found to have a significant q-value for the unstretched:stretched 1 experiment than the unstretched:stretched 2 experiment. In fact, no proteins were found to have a significant q-value for the L:H solution of unstretched:stretched 2. Bearing in mind that more proteins were identified and quantified for the unstretched:stretched 2 experiment, this lack of significance serves as a reminder that small data sets (with no replicates or repeat sampling) can lead to inconclusive or misleading information, and of the need for replicate analyses. The significant q-values obtained for unstretched:stretched 1 experiment are all close to the significance threshold used (0.05). It is important to bear in mind that this was a quick investigation with no biological or technical replicates provided or possible due to the very limited sample amount and time. The results should thus be considered as a preliminary investigation aimed at generating candidates for consideration in further studies, and for comparison with the emPAI analysis carried out separately by Adam Dowle. Further work should be performed in the first instance to generate technical and biological replicates, to enhance the understanding of these samples and give greater confidence in these preliminary dimethyl labelling results obtained.

No protein was identified as changing levels in both the forward and reverse experiments for stretched 1 or stretched 2, which suggests these results may be false positive quantification results or that the level of change is low. Further investigation is needed to determine whether the changes reported in Table 5.9 are true positive results. However, an initial interpretation of the data was performed. There is some agreement between the dimethyl labelling experiment and the more detailed emPAI experiment (data not reported in this work):

The proteins mitochondrial ATP synthase subunit alpha and mitochondrial aldehyde dehydrogenase were both found to be up-regulated in the stretched muscle protein sample 1 compared with the unstretched muscle protein sample in both the dimethyl labelling experiment and the emPAI experiment. In the dimethyl labelling experiment, the protein 60S ribosomal protein L13 was found to be up-regulated, whilst in the emPAI experiment, the protein 60S ribosomal protein L5 was found to be up-regulated. For the unstretched:stretched 1 samples, the protein tubulin beta-4B chain was found to be down-regulated in both the dimethyl labelling and the emPAI experiments. For the unstretched:stretched 2 samples, heat shock protein HSP 90-beta was found to be up-regulated in both the dimethyl labelling and the emPAI experiments.

For the unstretched:stretched 1 samples, ribosomal proteins were found to show differential abundance. 40S ribosomal protein SA, 60S ribosomal protein L13 and 60S acidic ribosomal protein P0 were found to be up-regulated in stretched muscle protein sample 1 compared with the unstretched muscle sample. It has been found that early responses to damaged muscle include the early induction of transcription genes.²⁴⁸ The up-regulation of some

ribosomal proteins in the stretched muscle protein sample compared with the unstretched muscle protein sample would be consistent with this observation.

Transitional endoplasmic reticulum ATPase was found to be down-regulated in stretched muscle protein sample 1 compared with the unstretched muscle protein sample.

Transitional endoplasmic reticulum ATPase is necessary for the fragmentation of Golgi stacks during mitosis and for their reassembly after mitosis.²⁵⁴ The protein myosin-9 was found to be down-regulated in the unstretched:stretched 1 sample. Myosin-9 plays a role in cytoskeleton reorganisation, cytokinesis and cell shape.²⁵⁵ Decarboxylating 6-phosphogluconate dehydrogenase was found to be down-regulated in stretched muscle protein sample 1 compared with the unstretched muscle protein sample. The function of this decarboxylating enzyme is to catalyse the oxidative decarboxylation of 6-phosphogluconate to ribulose 5-phosphate and carbon dioxide, with concomitant reduction of NADP to NADPH.²⁵⁶

Heat shock protein HSP 90-beta was up-regulated in the stretched muscle protein 2 sample when compared with the unstretched muscle protein sample, which is not surprising. The muscle samples were stretched for three hours, and therefore were likely to have been showing shock responses. HSP 90 is a molecular chaperone, many of which are known to interact with cytoskeletal elements such as microfilaments and intermediate filaments, and regulate the folding of cytoskeletal or cytoskeleton-related proteins.^{257,258} HSPs are known to be produced in stressed cells (they are still an important part of the non-stressed cell), which would explain their up-regulation in the stretched muscle protein 2 sample.^{257,259}

Chloride intracellular channel protein 1 was found to be down-regulated in stretched muscle protein sample 1 when compared with the unstretched muscle protein sample. Chloride intracellular channel protein 1 is a protein which inserts into membranes to form chloride ion channels.²⁶⁰ In 2004, Baek *et al.* published findings which included a down-regulation of chloride intracellular channel protein 1 under oxidative stress in *Helicobacter pylori*-infected human gastric mucosa.²⁶¹ It is possible that a similar effect is observed during stretching of the muscle proteins, which would account for the down-regulation of chloride intracellular channel protein 1 in stretched muscle protein sample 1 compared with the unstretched control sample.

The annexin family of proteins are calcium and phospholipid binding proteins, and are involved in the regulation of calcium in cells. Annexin A1 was found to be down-regulated in stretched muscle protein sample 1 when compared with the unstretched muscle protein sample. This is perhaps contrary to the expected result as annexin A1 has been found to play an important role in cell repair.²⁶²⁻²⁶⁴ In 2015, Leikina *et al.* published a paper in which the authors proposed that a deficiency in annexin A1 did not have a detrimental effect on myofibre repair after lengthening contraction injury, although its absence did delay muscle

regeneration.²⁶² This suggests that its presence is not necessarily key to muscle regeneration over the seven day period investigated by Leikina *et al.*. However, the result observed during the dimethyl labelling investigation presented in this thesis could also be a false positive quantification result.

Desmin was found to be down-regulated in both the stretched muscle protein sample 1 and 2 compared with the unstretched muscle protein sample. Desmin is an intermediate filament protein which connects myofibrils to each other and to the plasma membrane in striated muscle. Due to the fact that the muscle samples were stretched, it is highly likely that damage occurred to the proteins. There have been several papers which have discussed the possibility that if a muscle is damaged, desmin can be broken down within the cell.^{248,265–267} Therefore, it is possible that the down-regulation of desmin in the stretched muscle protein samples compared with the unstretched control could be due to damage occurring to the stretched muscle proteins, and subsequent desmin break-down.

Two mitochondrial proteins were found to be up-regulated in stretched muscle protein sample 1 compared with the unstretched muscle protein sample. Mitochondrial aldehyde dehydrogenase and mitochondrial ATP synthase subunit alpha were both found to be up-regulated in unstretched:stretched 1. Mitochondrial aldehyde dehydrogenase is involved in the conversion of aldehydes to carboxylates.²⁶⁸ The up-regulation of aldehyde dehydrogenase proteins has been noted during organism stress, particularly oxidative stress.²⁶⁹ Mitochondrial ATP synthase subunit alpha is a protein which produces ATP from ADP, which in turn is used as energy by cells.²⁷⁰ This could explain its up-regulation in stretched muscle protein sample 1 compared with the unstretched muscle protein sample, due to a possible increased energy demand during stretching. Alternatively, muscle contraction is known to require energy,²⁷¹ and therefore it is possible that as a result of stretching (where subsequent contraction would be expected), the up-regulation of mitochondrial ATP synthase subunit alpha may occur.

Tubulin beta-4B chain was found to be down-regulated in the stretched muscle protein sample 1 compared with the unstretched control. It is a major constituent of microtubules.^{257,272} Tubulin beta-chain has previously been found to be up-regulated when cells are subjected to oxidative stress.²⁷³ Tubulin has also been found to be down-regulated during heat stress.^{274,275} In 2010, Weis *et al.* published a paper detailing the relationship between heat shock protein 90 and tubulin.²⁷⁵ Based on that work, it is possible that the up-regulation of heat shock protein 90 observed here is due to the down-regulation of tubulin beta-4B chain in the stretched muscle protein sample 1.

Phosphoglycerate kinase 1 and phosphoserine aminotransferase were both found to be up-regulated in the stretched muscle protein sample 1 compared with the unstretched muscle protein sample. Phosphoglycerate kinase 1 is a glycolytic protein.²⁷⁶ There have been

several publications reporting up-regulation of phosphoglycerate kinase 1 for different organisms under stress conditions.^{276,277} Therefore, the up-regulation of phosphoglycerate kinase 1 in stretched muscle protein sample 1 is perhaps not surprising given the fact that stretching the samples for three hours is likely to have induced stress. Phosphoserine aminotransferase catalyses the reversible conversion of 3-phosphohydroxypyruvate to phosphoserine.²⁷⁸ Phosphoserine aminotransferase has previously been found to be up-regulated in response to heat stress.²⁷⁹ It is possible that the effect could be the same for stretching stress as it is for heat stress.

Fructose-bisphosphate aldolase A was also found to be up-regulated in the stretched muscle protein sample 1 compared with the unstretched muscle protein sample. This protein plays an important role in glycolysis and gluconeogenesis.²⁸⁰ Fructose-bisphosphate aldolase A has previously been found to be up-regulated in mice after low-intensity exercise.²⁸¹ Therefore, the observation of fructose-bisphosphate aldolase A being up-regulated in the stretched muscle protein sample 1 is thus consistent with this previously-observed result.

Several of the proteins with significant q-values have been observed as being up-regulated when an organism is placed under stress (e.g. heat shock proteins and phosphoglycerate kinase 1). Wang *et al.* published a paper in 2009 where they analysed data from many different experiments to determine whether there were any generally detected proteins in comparative proteomics.²⁸² The authors concluded that the reason for commonly differentially abundant proteins could be due to cellular stress response. Some of the proteins identified as being frequently detected as having differential abundance include heat shock proteins, chloride intracellular protein 1, phosphoglycerate kinase 1 and fructose-bisphosphate aldolase A. Therefore, these proteins having differential abundance in the stressed muscle experiment may be due to a cellular stress response.

The fact that the most abundant proteins do not show a significant difference in abundance between the unstretched and stretched muscle samples is an interesting result. It is possible that it is some of the less abundant proteins within the muscle that show a difference in abundance between the unstretched and stretched samples.

Although the two stretched muscles samples were both subjected to dynamic stretching for three hours, the proteins with differential abundance are not the same in the two samples. This has also been noted through emPAI LC-ESI-MS analysis and western blotting analysis carried out by others on the same samples. For example, during the emPAI analysis, the protein 60S ribosomal protein L18 was found to be up-regulated in stretched muscle protein 1 sample compared with the unstretched muscle protein sample. However, it was not determined to be up-regulated in stretched muscle protein 2 sample compared with the unstretched control. It is not currently known why the two stretched muscle samples, which

came from the same flask of undifferentiated C2C12 myoblasts, and were subjected to the same stretching treatment, were found to have different proteins up-regulated and down-regulated when compared with the unstretched muscle protein sample. However, it is possible that this could be due to biological variation, or differing amounts of total protein in samples 1 and 2. It is possible that for unstretched:stretched 2, the differences between the number of proteins identified and quantified for the L:H and H:L experiments could be due to systematic errors due to labelling, errors in the reproducibility of preparing the solutions, or errors arising from overlapping isotopic envelopes.

5.5. Conclusions and future work

As detailed above, dimethyl labelling is now regarded as a routine labelling strategy in mainstream proteomics. However, due to time constraints affecting the work in this chapter, a full-scale investigation was not possible. The work in this chapter acts as a starting point in the broader aim to develop dimethyl labelling as a mainstream labelling strategy within the group, as well as to adapt and develop it for use with the SAM technology.

The in-solution dimethyl labelling strategy is cheap, quick and efficient. Labelling of standard samples was performed and samples mixed in a variety of ratios (both L:H and H:L), and the solutions were then analysed by MALDI-MS and LC-ESI-MS. The ratios generated from the LC-ESI-MS data are closer to those of the standard solutions than the MALDI-MS data, with a smaller spread in the data for LC-ESI-MS than MALDI-MS generated results, which suggests that LC-ESI-MS produced the more accurate and precise results.

Two software packages were tested for data analysis of the LC-ESI-MS data: PS and MD. Having compared the two software packages, MD provided better results than PS. In PS, the data generated for the standard solutions always resulted in an under-representation of the ratio. However, this problem was not observed with MD. With the *E. coli* lysate spiking experiments, more peptides were identified (and subsequently used for quantification) for proteins in the *E. coli* lysate spiking experiment data analysed with MD than with PS. Additionally, the spread in quantification results was much smaller for the data handled with MD than PS. With PS, the range of ratios obtained for *E. coli* lysate proteins (which should have been 1:1) was much larger than with MD.

There are differences between the two software packages used. For MD, at the time of the data being analysed, it was not possible to import calibrated data into the software. This meant that for database searches conducted through Mascot for MD, wider tolerances were needed for mass errors than for PS. Once MD can incorporate the calibrated Bruker data

files, it would be interesting to compare the quantification results obtained from calibrated data using MD and PS with uncalibrated data using MD. Although search tolerances wider than 10 ppm (peptide tolerance) and 0.1 Da (MS/MS tolerance) can be used in PS, uncalibrated data would need to be used in PS in order to gain a complete comparison. PS automatically uses the calibrated data, so that a way to use uncalibrated data would need to be investigated. The ultimate difference between the two software packages, leading to the difference in the number of peptides available for quantification (and ultimately the superior quantification results obtained through MD over PS), is unknown. This would be worth investigating as part of a larger comparison of different software packages for quantification.

An investigation using LC-MALDI-MS could be undertaken, in a similar manner to the LC-ESI-MS analysis. In this way, a set of standard labelled solutions of a single protein digest could be analysed. A set of spiked solutions could then be prepared and analysed. This would allow a direct comparison between LC-ESI-MS analysis and LC-MALDI-MS analysis. Ultimately, MALDI-MS would be the analysis technique used, if a dimethyl labelling derivative was used in combination with the SAM technology. Therefore, the suitability of MD for processing LC-MALDI-MS analysis should be investigated. It is possible that, with the analytes separated prior to MALDI-MS, the relative quantification may be improved. This would help mitigate the issue that when all the peptides are spotted in one well, peptides which ionise least efficiently may be less well represented in the mass spectrum, which may adversely affect the quantification ratio for that peptide. The LC-MALDI-MS technique may mean that these peptides are better represented in the resulting mass spectra if the spectra are not dominated by a few highly-ionisable peptides; this may improve the quantification results obtained.

Future work should include further investigation of the muscle protein samples. Replicate analyses of the samples should be performed (in order to maximise the protein coverage) (technical and biological replicates should be considered). An alternative quantification method, such as iTRAQ quantification, could be used to corroborate the results. Alternatively, an orthogonal chromatographic method could be used (e.g. HILIC or strong cation exchange chromatography). An alternative analysis would provide greater confidence in the data obtained. Potential changes in protein expression, observed through differing protein abundance between samples, should then be validated using alternative methods such as western blotting.

Potentially important information could be missed on the less abundant proteins (which may be up- or down-regulated), as it is possible that peptides associated with these proteins may not be detected due to their low abundance. The samples could be analysed using a longer

LC gradient (which may further separate peptides, and allow more peptides to be identified and used for quantification for the less abundant proteins). In Section 5.4, single injections of the samples were made. With more sample and replicate labelling, repeat injections of each sample would be possible, with the resultant data being combined for analysis. Additional peptides may well be identified and quantified in the different runs, as it is unlikely that exactly the same peptides would be identified in each run, and technical repeats may provide additional information.^{99,283,284} Having performed that analysis, a list of up- and down-regulated proteins could be generated, in order to target the analysis towards specific proteins. Alternatively, more sample could be loaded onto the instrument in order to try and increase the intensities of the less abundant peptide ions. However, this would be dependent on the intensities of the most abundant ions, as it is possible to saturate the mass spectrometric detector. Therefore, a balance must be struck between increasing the intensities of the less abundant ions and saturating the detector with the most abundant ions. It is also possible to deplete for some of the major proteins, in order to reduce the complexity of the sample and subsequently, potentially, increase the ability to detect and quantify the less abundant proteins in a sample. Depletion strategies usually involve immunoaffinity techniques to remove one or more specific proteins. An alternative chromatographic method, such as HILIC²³⁴, strong cation exchange²⁵ or a high pH reverse phase system could also be used.

An interesting approach would be to analyse a new set of *E. coli* lysate digest-spiked solutions using a wider variety of quantification strategies, such as label-free and iTRAQ, alongside dimethyl labelling. Due to the fact that these samples would be identical (other than the label), and exactly what was in them would be known, it would provide an interesting comparison of the different labelling strategies (e.g. is one strategy more suited to quantification of the spiked proteins, or is one strategy better at identifying spiked proteins at 1% over 5%). Having completed the process with a set of standard spiked solutions, a variety of real samples could be analysed in the same way and a direct comparison of the outcomes made.

Chapter 6: Integration, Conclusions and Future Work

6.1. Integration

The different strands of this thesis had been planned as part of a coherent whole, with the aim being to recombine these strands to form a new SAM technology for quantitative proteomics. The SAM technology as it stands shows potential as a novel analysis methodology for mass spectrometric peptide analysis and the dimethyl labelling strategy has the potential to be adapted for use in the SAM technology. The envisaged workflow can be seen in Figure 6.1.

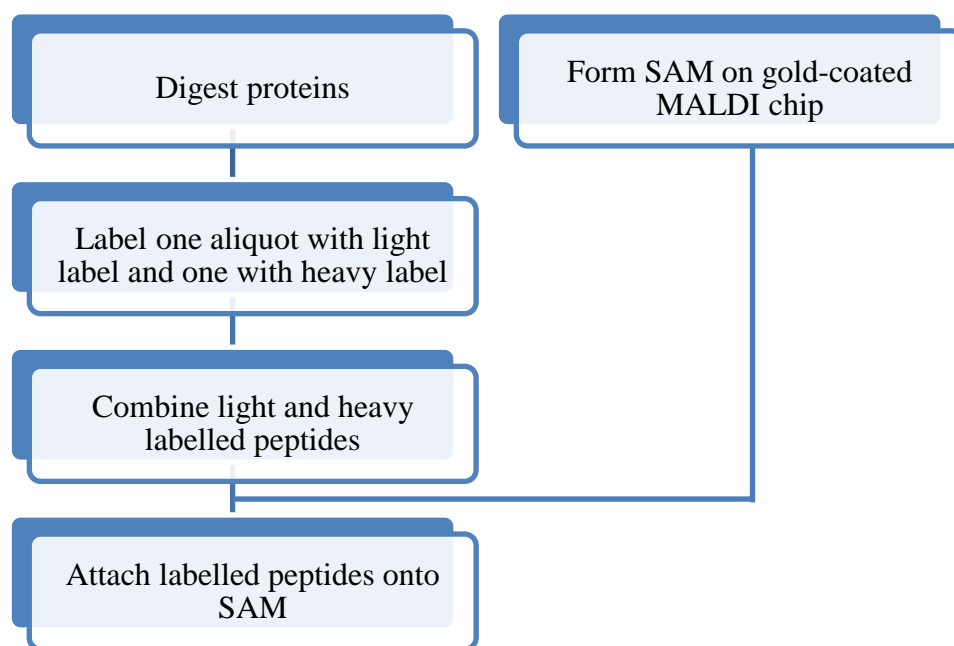


Figure 6.1: Proposed workflow for capturing protein digests on a SAM, with incorporation of the dimethyl-derived label

When the dimethyl labelling work was started, the ultimate aim had been to derivatise a ketone for use with the SAM technology. Formaldehyde has no other functional group, and therefore has no functionality to react with the SAM. Therefore, it is unsuitable for use with the SAM technology. However, a ketone can contain more than one functional group, meaning that a second group could be incorporated for reaction with the SAM. Having established dimethyl labelling (which uses an aldehyde) successfully within the proteomics facility, the next logical step was to label peptides from a tryptic digest with a simple ketone.

As a proof of principle that a molecule with a ketone carbonyl group could be used to label the peptides, aliquots of a simple BSA digest were labelled with acetone and acetone- d_6 . The labelling chemistry is similar for formaldehyde and acetone. Although there has been much research into dimethyl peptide labelling, there is limited literature on ketone labelling of peptides. One example is the labelling of peptides with acetone, which the authors termed ‘reductive alkylation by acetone (RABA)’ (Figure 6.2).²⁸⁵ In this paper, the group looked at

labelling standard peptides and tryptic digests of transgenic mouse tissues and human cell lysates.

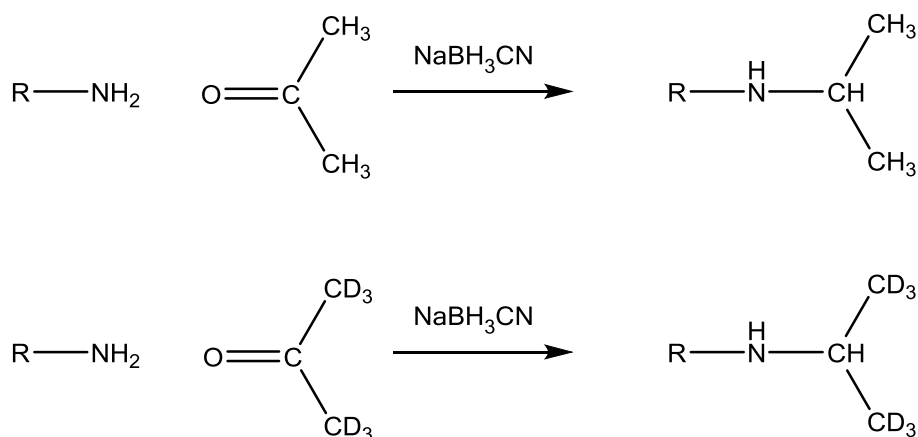


Figure 6.2: Labelling of a generic peptide with acetone (top) or acetone-d₆

The group labelled the peptides with acetone or acetone-d₆ using 90% acetonitrile (+0.1% formic acid) as the solvent system. This would add an additional step to the current methodology, as the protein digests are performed in TEAB at pH 8.0, and so would need to be dried and re-suspended in the 90% acetonitrile (+0.1% formic acid) solvent system. Therefore, the possibility of labelling the peptides using an adapted version (using TEAB instead of 90% acetonitrile +0.1% formic acid of the RABA labelling strategy) was investigated.

Aliquots of a BSA digest in TEAB were acidified to ~ pH 4.5 using 0.1% formic acid. One aliquot was then labelled with acetone, and the other with acetone-d₆. A 1:1 solution of the labelled digests was prepared and then analysed by MALDI-MS. The resulting mass spectrum is shown in Figure 6.3.

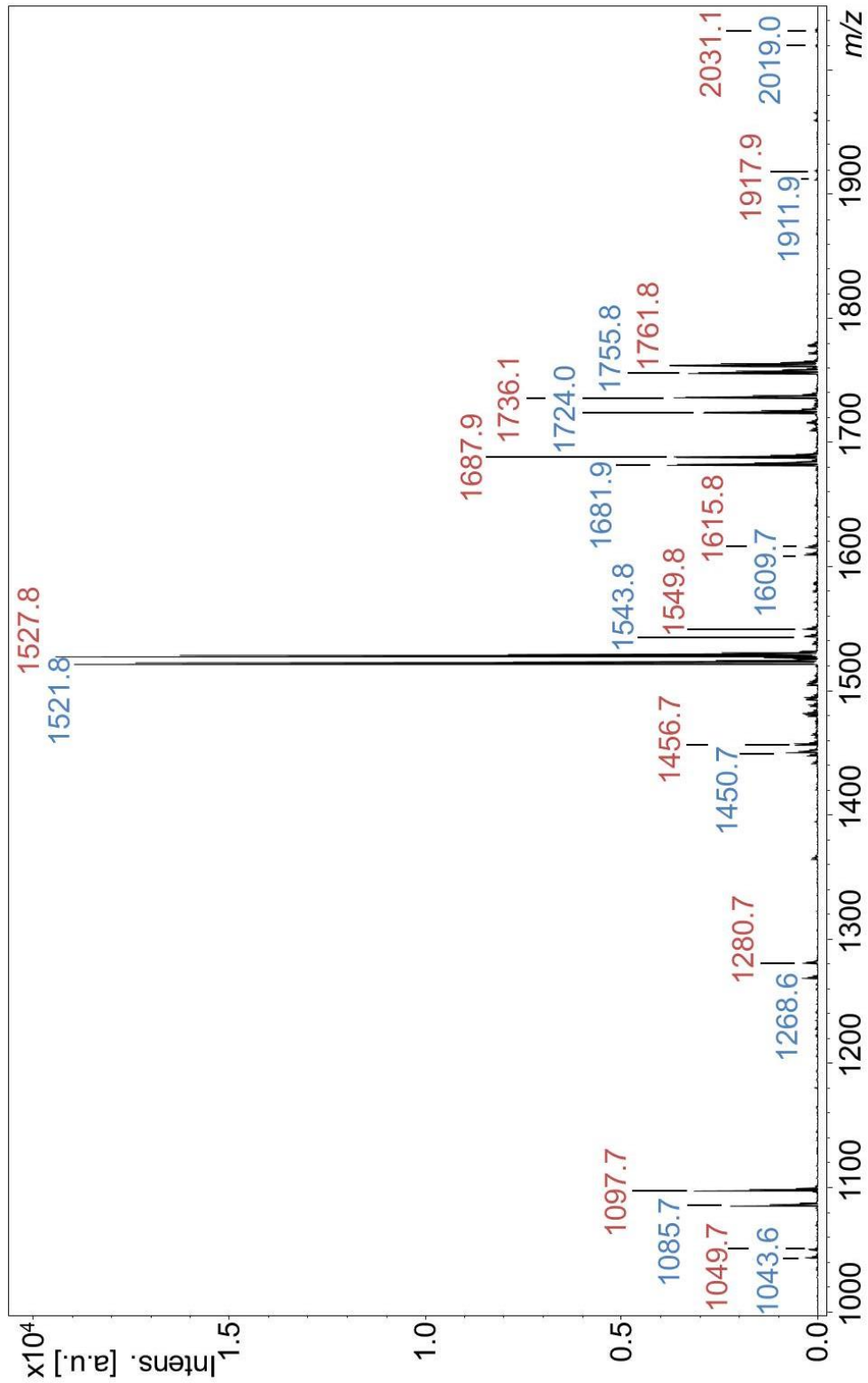


Figure 6.3: MALDI mass spectrum of a BSA digest labelled with acetone and acetone-d₆ (1:1) (blue = acetone labelled peptides, red = acetone-d₆ labelled peptides) (not all peptide pairs are labelled on the mass spectrum, but all identified peptides are detailed in Table 6.1)

As can be seen in Figure 6.3, BSA tryptic digest peptides have been successfully labelled with acetone and acetone- d_6 . There is a mass difference of 6 Da between the acetone and acetone- d_6 labelled peptides. In some instances, two labels have been incorporated into peptides which include a lysine residue. This suggests that the peptide has been labelled at the N-terminal and also at the lysine side-chain. No peaks were observed for unlabelled BSA digest peptides. Although no MS/MS data were acquired, the proposed peptide assignments for the spectrum shown in Figure 6.3 are given in Table 6.1. There remain two potential peptide pairs which are currently unassigned (m/z 1579.8/1585.8 and m/z 1960.8/1966.8).

Table 6.1: Proposed peptide assignments for peaks observed for acetone/ acetone- d_6 labelled peptides

Nominal m/z value	Number of labels	Peptide sequence
1043	1	ALKAWSVAR
1085	2	ALKAWSVAR
1268	2	CASIQKFGER
1441	1	TVMENFVAFVDK
1450	1	SLHTLFGDELCK
1481	1	RHPEYAVSVLLR
1492	2	SLHTLFGDELCK
1521	1	LGEYGFQNALIVR
1543	1	LRCASIQKFGER
1609	1	DAFLGSFLYEYSR
1681	1	KVPQVSTPTLVEVSR
1709	1	MPCTEDYLSLILNR (unmodified Cys)
1724	2	KVPQVSTPTLVEVSR
1755	1	MPCTEDYLSLILNR
1771	1	KFWGKLYEIIARR
1911	1	RPCFSALTPDETYVPK
2019	2	SLHTLFGDELCKVASLR

As can be seen from the data presented in Figure 6.3 and Table 6.1, peptides from aliquots of a BSA digest were successfully labelled with acetone and acetone- d_6 . This serves as a critical proof of principle that peptides can be labelled with a ketone (using a protocol based on a combination of the dimethyl labelling strategy used, and the RABA labelling strategy²⁸⁵), and is the crucial first stage in combining the different strands of the work in this thesis.

6.2. Conclusions

The work in this thesis has consisted of three different strands: creating SAMs on a gold-coated MALDI chip to capture peptides for mass spectrometric analysis, novel removal of immobilised material from the gold-coated MALDI chips and solution-based chemical labelling of tryptic protein digests for quantitative mass spectrometric analysis.

There are many potential benefits to the capture and analysis of peptides using gold-immobilised SAMs for mass spectrometric analysis. These include:

- (i) All chemistry and analysis occurs *in situ* on the gold-coated MALDI chip, minimising transfers and thus reducing sample losses, which are the main drawback of solution phase multistep and derivatisation-based protocols.
- (ii) Capture of peptide(s) on the SAM concentrates a sample, which is useful when dealing with small sample amounts.
- (iii) Unreacted reagents can be washed from the gold-coated MALDI chip before analysis, allowing the use of non-volatile buffers, such as PBS, where in solution-phase proteomic workflows typically demand a desalting step. Solid phase capture offers an alternative to these procedures, which is worth evaluating.
- (iv) All steps occurring on a multi-well plate which offers a convenient medium throughput sample handling and analysis process.

The approach, as it stands, still has significant drawbacks, the major one of which is the yield of the peptide capture chemistry. Our plans to exploit 'click' chemistry (see Section 6.3), which is noted for its good yields, should offer the potential for addressing this drawback and improving the overall levels of analyte at which the approach is applicable.

In Chapter 2, a thiol-derived SAM (with and without the incorporation of the chemical tag 4-BrPhe) was used to capture a variety of peptides for mass spectrometric analysis. Levels of peptide capture were lower for peptides captured on the untagged activated SAM than for peptides captured on the activated 4-BrPhe-tagged SAM. In experiments with and without the use of the 4-BrPhe, peaks were observed for unreacted SAM after peptide capture. An investigation using H₂¹⁸O-containing PBS for 4-BrPhe coupling and peptide capture on the activated SAM suggested that the presence of peaks associated with unreacted SAM was not due to hydrolysis of the activated SAM. This suggested that the reason for the peaks associated with the unreacted SAM could be due to inefficient chemistry occurring with the SAM. It is likely that a combination of inefficient activation of the SAM and inefficient capture of the peptide on the activated SAM led to the presence of peaks associated with unreacted SAM.

In Chapter 3, a disulfide-derived SAM was used (as an alternative to the thiol-derived SAM used in Chapter 2) to capture a variety of peptides for mass spectrometric analysis. The

disulfide-derived SAM contained a pre-activated component, in the form of an NHS ester disulfide (removing the need for an on-chip activation step, and meaning one less step in construction of the SAM-peptide structure). The peptide capture reaction with the SAM has been reduced from an overnight reaction to a one hour reaction. This considerable time saving could potentially mean that the entire process (from SAM formation to peptide capture and analysis) could take place over two days instead of the original three overnight steps. The data in Chapter 3 (Sections 3.3 and 3.4) suggest that the presence of Phe at the N-terminus reduces the efficiency of capture of the Phe-containing species on the disulfide-derived SAM. This could be due to aggregation occurring within the peptide solution, due to the hydrophobic nature of Phe, and/or π -stacking of the aromatic side chain. However, peptide capture was more successful when a lower amount (0.05 nmol) of peptide was spotted on the disulfide-derived SAM than on the thiol-derived SAM (5 nmol). This was an improvement from the thiol-derived SAM used in Chapter 2.

In Chapter 4, two plasma instruments were used to remove a SAM from the gold-coated MALDI chips. Both of these plasma instruments removed a thiol-derived SAM from the gold-coated MALDI chips. The proposed method would be to wash the matrix from the gold-coated MALDI chip, and then treat the gold-coated MALDI chip with plasma effluent. In this instance, the SAM could be removed from the wells of the gold-coated MALDI chip in less than 10 seconds. This is a considerable decrease in the time currently taken to clean the gold-coated MALDI chip with piranha solution (up to 30 minutes to clean a chip), and no corrosive solutions are needed.

In Chapter 5, dimethyl labelling of peptides from tryptic digestions was performed. Standard samples were labelled with either light or heavy formaldehyde and samples mixed in a variety of ratios (both L:H and H:L), and the solutions were analysed by MALDI-MS and LC-ESI-MS. The ratios generated from the LC-ESI-MS data were closer to the ratios of the standard solutions, resulting in LC-ESI-MS being carried forward as the analysis method for further work in the chapter. Two software packages (PS and MD) were tested for data analysis of the LC-ESI-MS data, with MD providing better results than PS. Dimethyl labelling and relative quantification was successfully performed on a complex standard sample, and also on a 'real' set of samples (muscle protein samples). The complex standard samples provided a set of criteria which could be applied to the muscle protein samples in order to determine the parameters outside of which a protein ratio could be deemed a potential positive quantification result. Dimethyl labelling of muscle protein samples showed that some proteins appear to be differentially abundant in unstretched and stretched muscle protein samples, some of which correspond with results obtained by emPAI analysis of the same cell extracts. Other potential positive quantification ratio results need further investigation.

6.3. Future Work

Peptides from a BSA tryptic digest have successfully been labelled with acetone and d_6 -acetone. This paves the way for combining the SAM technology and the in-solution dimethyl labelling strands of the work discussed in this thesis (Figure 6.1). The chemical label would be bifunctional, comprising a ketone for reductive alkylation of a peptide and an orthogonal group to react with the SAM (Figure 6.4). An alkyne could allow reaction with an azide-functionalised SAM using 'click' chemistry. One of the most common forms of 'click' chemistry is the reaction of an azide and alkyne in the presence of a copper catalyst (known as a copper-catalysed azide-alkyne cycloaddition).^{286–288}

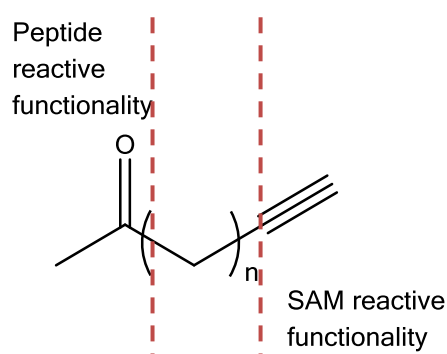


Figure 6.4: Schematic of the chemical label, including a carbonyl group and an alkyne group

For example, light and heavy versions of a suitable chemical tag could be generated using either hydrogen/deuterium (Figure 6.5) or $^{12}\text{C}/^{13}\text{C}$ (Figure 6.7).

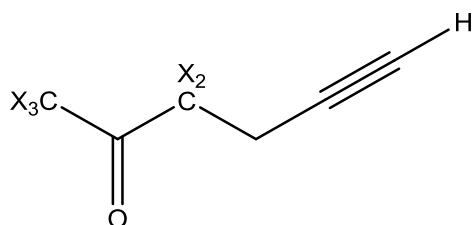


Figure 6.5: A potential chemical label using hydrogen and deuterium to generate the light and heavy labels (X indicate positions of either hydrogen or deuterium)

The molecule shown in Figure 6.5 contains the two key features needed for this strategy: a carbonyl group (for reaction with the peptide) and an alkyne (for reaction with the SAM). In the case of the potential hydrogen/deuterium chemical label shown in Figure 6.5, both light and heavy versions would need to be synthesised. For the chemical label in Figure 6.5, the mass difference between the light and heavy labels would be 5 Da.

In 1975, Sih *et al.* published a paper detailing the asymmetric total synthesis of (-)-prostaglandin E_1 and (-)-prostaglandin E_2 .²⁸⁹ As part of the synthesis of (-)-prostaglandin E_2 , 1-hexyn-5-one was synthesised (synthesis up to this point shown in Figure 6.6). Ethyl

acetoacetate was condensed with 3-bromo-1-propyne to generate 1-hexyn-5-one in a 70% yield.

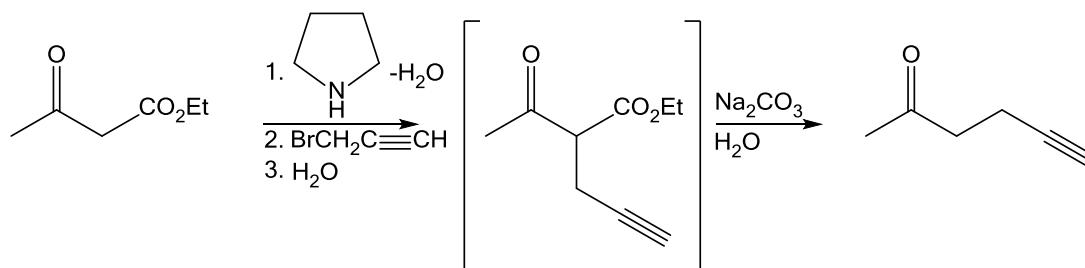


Figure 6.6: Potential synthetic scheme for synthesis of hydrogen (light) label²⁸⁹

For the light chemical label, the molecule could be synthesised by reacting ethyl acetoacetate with 3-bromo-1-propyne (Figure 6.6). Similarly, the heavy chemical label could be synthesised by reacting a deuterated ethyl acetoacetate (which would need to be synthesised) with 3-bromo-1-propyne.

In the case of the potential $^{12}\text{C}/^{13}\text{C}$ chemical label shown in Figure 6.7, the light version is commercially available. The heavy version would need to be synthesised. This could be synthesised by reacting ($^{13}\text{C}_2$)acetaldehyde with lithium ($^{13}\text{C}_2$)acetylide, followed by oxidation of the resulting secondary alcohol.

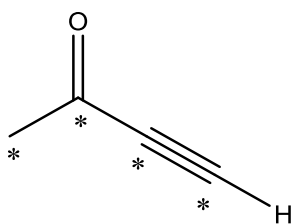


Figure 6.7: A potential chemical label using ^{12}C and ^{13}C to generate the light and heavy labels (* indicate positions of either ^{12}C or ^{13}C)

In 2007, Foot *et al.* published a paper detailing the cyclisation of O-propargylic hydroxylamine derivatives.²⁹⁰ As part of this synthesis, the authors reacted acetaldehyde with lithium acetylide to form an alcohol, which would be an ideal starting place for the synthesis of the ^{13}C chemical label. A proposed synthesis for the ^{13}C chemical label shown in Figure 6.7 is shown below (Figure 6.8).

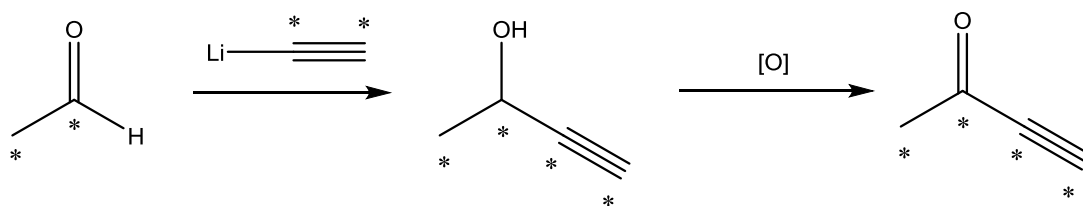


Figure 6.8: Proposed synthesis of the ^{13}C chemical label (* indicates ^{13}C position)

The oxidation of the alcohol to ketone shown in Figure 6.8, would need to be performed using conditions tolerant of the alkyne functionality. In 2003, Kang *et al.* published a paper which featured the oxidation of 3-butyn-2-ol to 3-butan-2-one, which used a Jones oxidation for the conversion.²⁹¹ This would therefore be worth exploring during the synthesis of the label shown in Figure 6.7.

For the chemical label in Figure 6.7, the mass difference between the light and heavy chemical labels would be 4 Da. This is the same mass difference as between the light and heavy formaldehyde labels used in Chapter 5. Using the $^{12}\text{C}/^{13}\text{C}$ chemical label would be structurally different to the hydrogen/deuterium chemical labels. The molecular weight of the labels would be lower than the hydrogen/deuterium chemical labels due to needing fewer CH_2 groups in order to incorporate the different labels. However, there is the possibility that due to the close proximity of the two functional groups required for this strategy, the $^{12}\text{C}/^{13}\text{C}$ chemical label may not be suitable.

Both the proposed hydrogen/deuterium and $^{12}\text{C}/^{13}\text{C}$ chemical labels contain a terminal alkyne group to react with an azide-terminated SAM. A potential azide-terminated thiol which could be used is shown in Figure 6.9, and has been synthesised by a collaborator with the group.²⁰⁷ This molecule contains an alkyl chain for rigidity, and also a PEG region to resist non-specific peptide adsorption. The presence of both of these regions has been shown to be important for future SAM components based on the work in Chapters 2 and 3.

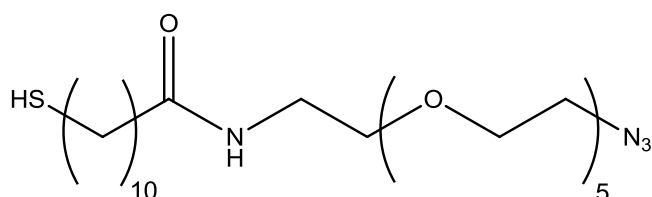


Figure 6.9: Potential 'reactive' component of the SAM

A potential spacer molecule which could be used with the azide-terminated thiol is shown in Figure 6.10.²⁰⁷ This alcohol-terminated thiol could be synthesised (within the group) with either four or five PEG groups.

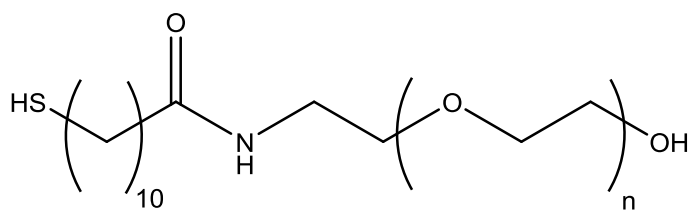


Figure 6.10: Potential spacer component of the SAM (where $n = 4$ or 5)

The chemical tags and SAM components would need to be tested in the proposed workflow. An important consideration would be whether the hydrogen/deuterium version or the $^{12}\text{C}/^{13}\text{C}$ alternative is the optimum chemical label. The ease and cost of synthesising the labels should be considered, as one of the benefits of dimethyl labelling is that it is relatively cheap compared with other quantitative proteomics strategies. Another important consideration when choosing the chemical label is the success of reaction with both peptides and the gold-immobilised azide-functionalised SAM. There must also be no preference for either the light or heavy labelled peptides to react with the immobilised SAM. Studies would need to be performed which compared the quantification data generated from the labelled peptides and the labelled peptides 'clicked' on the SAM, in order to ensure that data obtained from the peptides 'clicked' on the SAM are consistent with those obtained from the labelled peptides in solution.

There are examples in the literature of biological molecules being immobilised on solid surfaces. In 2006, Sun *et al.* published a paper in which they showed immobilisation of carbohydrates and proteins onto solid surfaces using Diels-Alder and copper-catalysed azide-alkyne cycloadditions.²⁹² The authors used a single-component SAM to immobilise a number of components onto the solid surface, including the protein thrombomodulin. The immobilisation of the protein was conducted in PBS (which would be compatible with the current methodology), using butanol, tris(triazolyl)amine solution and copper sulfate solution. The system was incubated at 4 °C for 72 hours without agitation (which again fits with the existing use of the gold-coated MALDI chips). Yang *et al.* detailed a carbohydrate microarray for the detection of glycan-protein interactions in 2015.²⁹³ A sodium ascorbate and copper sulfate solution was used in a phosphate buffer system. In 2014, Choi *et al.* published a paper detailing the regio- and chemoselective immobilisation of proteins on gold surfaces.²⁹⁴ Proteins were modified to include an alkyne moiety, which was subsequently used to capture the protein on the surface of an azide-terminated SAM. A phosphate buffer system was again used for the immobilisation of proteins on the gold surface.

Based on the work by Sun *et al.*, Yang *et al.* and Choi *et al.*, experiments using the proposed chemical tags for the future work of this project could use PBS as the solvent system. This would retain some of the conditions established during original work conducted during the course of this thesis.

An alternative strategy to consider, should the above ideas prove inefficient, could include selective reactivity of different functional groups on the peptide to attach the peptide to the SAM and also the label to the peptide. This is based on the isobaric peptide termini labelling (IPTL) strategy.²⁹⁵⁻²⁹⁷ Here, an alternative enzyme to trypsin (endoproteinase Lys-C) is used to digest the protein, which cleaves on the carboxyl side of lysine residues (essentially producing peptides with amino groups at both termini).²⁹⁸ Peptides are labelled on the C-terminal lysine residues with either 2-methoxy-4,5-dihydro-1*H*-imidazole or tetradeuterated 2-methoxy-4,5-dihydro-1*H*-imidazole. The labels selectively react with the C-terminal lysine residues over the N-terminal amino group. This leaves the N-terminal free for further reaction with tetradeuterated succinic anhydride or succinic anhydride. The mix of 2-methoxy-4,5-dihydro-1*H*-imidazole with tetradeuterated succinic anhydride and tetradeuterated 2-methoxy-4,5-dihydro-1*H*-imidazole with succinic anhydride creates isobaric masses.

Based on the IPTL strategy, a chemical label could thus be devised which reacts with the C-terminal lysine residues of the peptides. The N-terminal could then be reacted with the SAM, in order to anchor the labelled peptides to the surface of the gold-coated MALDI chip. However, this would be a non-isobaric strategy, as the SAM components would not be isotopically labelled.

Chapter 7: Experimental

7.1. General methods

PBS at pH 7.4, with a concentration defined as '10 ×' was obtained from Invitrogen. A working concentration of 1 × (137 mM NaCl, 2.7 mM KCl, 8 mM Na₂HPO₄ and 2 mM KH₂PO₄) was obtained by dilution with deionised water. The working concentration of 1 × was pH adjusted to pH 8.0 using 0.1 M NaOH. All other reagents were supplied by Sigma Aldrich, unless otherwise stated. Solvents were supplied by Fisher Scientific, except for TEAB (1M, pH 8.0) which was purchased from Sigma Aldrich. Peptides were supplied by Sigma Aldrich, with their purities taken as supplied. Peptides were examined mass spectrometrically, and no peptide contaminants were detected. Trypsin was supplied by Promega.

7.1.1. Mass spectrometry

7.1.1.1. solariX

Positive-ion MALDI mass spectra were obtained on a Bruker 9 Tesla solariX FT-ICR-MS instrument. It was operated with a Nd:YAG smartbeam™ laser (355 nm) with power setting of 40%. Mass spectra were acquired over the *m/z* range 500-3000 with a transient length of 2.5166 s to give a resolution of 440,000 at *m/z* 400. Each spectrum was acquired using 200 laser shots, a frequency of 200 Hz, and an 'average spectra' setting of 10.

The following ion transfer settings were used. Source optics: capillary exit, 100 V; deflector plate, 230 V; funnel 1, 100 V; skimmer 1, 15 V; funnel RF amplitude 180 Vpp. Octopole: frequency, 5 MHz; RF amplitude, 390 Vpp. Transfer optics: ToF, 2.000 ms; frequency, 4 MHz; RF amplitude 390 Vpp;.

Mass spectra were externally calibrated using a spot containing Bruker Daltonics Peptide Calibration Standard II. Peptide Calibration Standard II was dissolved in a mixture of acetonitrile and 0.1% TFA (1:2 v/v).

THAP was dissolved in acetonitrile (10 mg/mL). The solution was vortexed for 30 seconds. For sample wells, 0.3 µL of the THAP matrix solution was spotted on top of the already dried sample on each well, and the spot was left to air dry. For calibrant wells, 1 µL of the calibrant solution was spotted on the well, with 1 µL of THAP matrix solution, and the spot was left to air dry.

Spectra were obtained using Bruker solariXcontrol software and processed with Bruker DataAnalysis 4.0.

CID-MALDI mass spectra were obtained using argon as the collision gas, a collision energy of 45.0 V and an isolation window of m/z 4.0. A laser power setting of 45% was used. A transient length of 0.9088 s was used giving a resolution of 160,000 at m/z 400.

7.1.1.2. **ultraflex**

Positive-ion MALDI mass spectra were obtained on a Bruker ultraflex III instrument. It was operated in reflectron mode, with a Nd:YAG smartbeam™ laser with attenuation setting of 25%. Mass spectra were acquired over the m/z range 800-4000. Mass spectra were externally calibrated against an adjacent spot containing six peptides (des-Arg¹-bradykinin, 904.681; angiotensin I, 1296.685; Glu¹-fibrinopeptide B, 1750.677; ACTH (1-17 clip), 2093.086; ACTH (18-39 clip), 2465.198; ACTH (7-38 clip), 3657.929.). Monoisotopic masses were obtained using the peak-picking SNAP averagine algorithm (C 4.9384, N 1.3577, O 1.4773, S 0.0417, H 7.7583) and a S/N threshold of 2.

CHCA was dissolved in acetonitrile:water (50:50) + 0.1% TFA (10 mg/mL). The solution was vortexed for 30 s and then centrifuged for 1 minute. 100 μ L of the supernatant was pipetted into a new microcentrifuge tube, and 100 μ L of acetonitrile:water (50:50) + 0.1% TFA was added to create the matrix solution. 1 μ L of the sample solution was spotted on a steel MALDI target plate, with 1 μ L of CHCA matrix solution. The spots were left to air dry.

For each spot, the ten strongest ions, with S/N greater than 30, were selected for MS/MS. Fragmentation was performed in LIFT mode without the introduction of a collision gas. The default calibration was used for MS/MS spectra, which were baseline-subtracted and smoothed (Savitsky-Golay, width 0.15 m/z units, cycles 4); monoisotopic peak detection used a SNAP averagine algorithm (C 4.9384, N 1.3577, O 1.4773, S 0.0417, H 7.7583) with a minimum S/N of 6. Bruker flexAnalysis software (version 3.3) was used to perform spectral processing and peak list generation.

Tandem mass spectral data were submitted to database searching using a locally-running copy of the Mascot program (Matrix Science Ltd., version 2.4), through the Bruker PS interface (version 2.1). Search criteria specified: enzyme, trypsin; fixed modifications, carbamidomethyl (C); variable modifications, oxidation (M), dimethyl(Heavy_K) (K), dimethyl(Heavy_N-term) (N-term), dimethyl(Light_K) (K) and dimethyl(Light_N-term) (N-term); peptide tolerance, 250 ppm (initial searches performed using 100 ppm peptide tolerance, but known labelled peptide pairs were missed during the searches. Therefore, search parameters widened to allow these values to be included in the database search. Mascot Expect values were interrogated to monitor high mass error ppm values); MS/MS tolerance, 0.5 Da (standard setting used within the department); instrument, MALDI-TOF-TOF (the version and size of the database can be obtained from the Mascot result page).

Kings database (an in-house database) accessed between 18/06/2014 (694 sequences) and 31/07/2014 (707 sequences)). Results were filtered to accept only peptides with $p < 0.05$.

7.1.1.3. maXis HD

Samples were loaded onto a nanoAcquity UPLC system (Waters) equipped with a nanoAcquity Symmetry C₁₈ 5 μm trap (180 μm x 20 mm, Waters) and a nanoAcquity BEH130 1.7 μm C₁₈ capillary column (75 μm x 250 mm, Waters) (BSA digests)/ nanoAcquity HSS T3 1.8 μm C₁₈ capillary column (75 μm x 250 mm, Waters) (*E. coli* lysate spiking experiments/muscle protein sample experiments). The trap wash solvent was 0.1% (v/v) aqueous formic acid and the trapping flow rate was 10 $\mu\text{L}/\text{min}$. The trap was washed for 5 min before switching flow to the capillary column. The column temperature was 60 °C. The two mobile phases were: 0.1% (v/v) aqueous formic acid (solvent A) and acetonitrile containing 0.1% (v/v) formic acid (solvent B). The flow rate for the capillary column was 350 nL/min (BSA digests)/ 300 nL/min (*E. coli* lysate spiking experiments/muscle protein sample experiments). The gradient profile for the BSA digests was as follows: initial conditions 2% solvent B (2 min), followed by a linear gradient to 35% solvent B over 20 min. The gradient profile for the *E. coli* lysate spiking experiments/muscle protein sample experiments was as follows: linear 2-30% solvent B over 125 min then linear 30-50% solvent B over 5 min. In all instances, the gradient culminated in a wash with 95% solvent B for 2.5 min, prior to returning to initial conditions and re-equilibrating for 25 min before subsequent injections.

The nanoLC system was interfaced with a maXis HD LC-MS/MS system (Bruker Daltonics) with a CaptiveSpray ionisation source (Bruker Daltonics). Positive ESI- MS & MS/MS spectra were acquired using AutoMSMS mode. Instrument control, data acquisition and processing were performed using Compass 1.7 software (microTOF control, Hystar and DataAnalysis, Bruker Daltonics). Instrument settings were: ion spray voltage: 1,450 V, dry gas: 3 L/min, dry gas temperature 150 °C, ion acquisition range: m/z 150-2,000. For the BSA digests, AutoMSMS settings were: MS: 0.5 s (acquisition of survey spectrum), MS/MS (CID with N₂ as collision gas): ion acquisition range: m/z 350-2,000. For the *E. coli* lysate spiking experiments/muscle protein sample experiments, the additional settings quadrupole low mass: m/z 300, transfer time: 120 ms, collision RF: 1,400 Vpp, MS spectra rate: 5 Hz were recorded.

Precursors were fragmented using a fixed cycle time of 3 s (BSA digests/*E. coli* lysate spiking experiments)/1 s (muscle protein sample experiments) using a dynamic method adapting spectra rates between 1 and 10 Hz (BSA digests)/2 and 12 Hz (*E. coli* lysate spiking experiments)/5 and 20 Hz (muscle protein sample experiments) based on precursor intensities with an absolute threshold of 655 counts. The collision energy and isolation width settings were automatically calculated using the AutoMSMS fragmentation table, preferred

charge states were 2-4 and singly charged ions were excluded. A single MS/MS spectrum was acquired for each precursor and former target ions were dynamically excluded for 48 s with reconsideration of excluded precursors for fragmentation if their intensity rose by a factor of 4 within this time.

7.1.2. Cleaning of gold-coated MALDI chips

A gold-coated MALDI chip was placed in a Petri dish and cleaned with piranha solution (3:1 conc. H_2SO_4 : 35% H_2O_2) [caution hazardous]. A freshly prepared solution was poured over the gold-coated MALDI chip in the dish and left for approximately 30 minutes, with occasional agitation, until bubbles were no longer observed on the gold-coated MALDI chip surface. Piranha solution was neutralised using sodium hydroxide in deionised water, and the gold-coated MALDI chip was washed with deionised water and acetonitrile and dried under a stream of nitrogen.

7.2. Chapter 2

7.2.1. Thiol-derived SAM formation (Figure 2.4)

Stock solutions of the two SAM components, $(\text{HS}(\text{CH}_2)_{17}(\text{OCH}_2\text{CH}_2)_3\text{OH})$ and $\text{HS}(\text{CH}_2)_{17}(\text{OCH}_2\text{CH}_2)_6\text{OCH}_2\text{COOH}$, were prepared in DMSO (0.4 mg/mL) and stored at -18°C . For experiments, the stock solutions were diluted to 0.1 mg/mL and these were mixed in a 4:1 molar ratio (alcohol:acid thiols). Aliquots (0.4 μL) of this mixed solution were spotted onto the wells of the gold-coated MALDI chip. The gold-coated MALDI chip was sealed in a Petri dish and left overnight at ambient temperature then washed with acetonitrile and dried under a stream of nitrogen.

7.2.2. NHS/EDC activation (Figure 2.9 and Figure 2.12)

A solution containing NHS (2 mg, 25 mM) and EDC (12.8 mg, 100 mM) was prepared in anhydrous DMF (1 mL). Aliquots (1 μL) of this solution were spotted onto the wells of the gold-coated MALDI chip. The gold-coated MALDI chip was sealed in a Petri dish at ambient temperature for one hour then washed with acetonitrile and dried under a stream of nitrogen.

7.2.3. PFP/EDC activation

Solutions of PFP (32 mg, 180 mM) and EDC (35 mg, 180 mM) were prepared in anhydrous DMF (1 mL). Equal volumes of the two solutions were combined, and aliquots (1 μ L) of the resulting solution were spotted onto the wells of the gold-coated MALDI chip. The gold-coated MALDI chip was sealed in a Petri dish at ambient temperature for one hour then washed with acetonitrile and dried under a stream of nitrogen.

7.2.4. NHSS/EDC activation

A solution of NHSS (2 mg, 25 mM) and EDC (12.8 mg, 100 mM) was prepared in anhydrous DMF or acetonitrile (1 mL). Aliquots (1 μ L) of this solution were spotted onto the wells of the gold-coated MALDI chip. The gold-coated MALDI chip was sealed in a Petri dish at ambient temperature for one hour then washed with acetonitrile and dried under a stream of nitrogen.

7.2.5. 4-Bromophenylalanine coupling to the thiol-derived SAM

(Figure 2.10)

A solution of 4-BrPhe (6.3 mg, 50 mM) in 50:50 (v:v) DMF:PBS (1 \times , titrated to pH 8.0 with 0.1 M NaOH) (0.5 mL) was prepared. This solution was sonicated and centrifuged. Aliquots of the supernatant (1 μ L) were spotted onto the gold-coated MALDI chip.

A piece of filter paper was placed in the bottom of a Petri dish and wetted with deionised water (1 mL), to ensure the gold-coated MALDI chip did not dry out. The gold-coated MALDI chip was sealed in the Petri dish and left overnight at ambient temperature. The gold-coated MALDI chip was washed with acetonitrile and dried under a stream of nitrogen.

7.2.6. Peptide capture on the thiol-derived SAM (Figure 2.13 and

Figure 2.16)

Peptides were prepared at the specified concentration using PBS (1 \times , titrated to pH 8 with 0.1 M NaOH). Aliquots (1 μ L) were spotted onto the wells of the gold-coated MALDI chip. A piece of filter paper was placed in the bottom of a Petri dish, and wetted with deionised water (1 mL), to ensure the gold-coated MALDI chip did not dry out.

For investigations using H₂¹⁸O-containing PBS, a piece of filter paper was placed in the bottom of the dish with the gold-coated MALDI chip spotted with the H₂¹⁸O-containing PBS, and anhydrous DMF was pipetted onto the filter paper until it was covered.

In both methods, the Petri dish was sealed and the gold-coated MALDI chip left at ambient temperature overnight. The gold-coated MALDI chip was washed with acetonitrile and dried under a stream of nitrogen. The peptides were analysed by MALDI-MS and the expected products are shown in Table 7.1.

Table 7.1: Peptides used for capture reactions and their corresponding product peaks observed by MALDI-MS analysis

Peptide	Molecular mass / g mol^{-1}	Expected product ion (Tagged SAM) $[\text{M}+\text{Na}]^+ / m/z$ $\begin{array}{c} \text{S} \text{ ~~~~~ (Br)F-peptide} \\ \\ \text{S} \text{ ~~~~~ OH} \end{array}$	Expected product ion (SAM) $[\text{M}+\text{Na}]^+ / m/z$ $\begin{array}{c} \text{S} \text{ ~~~~~ peptide} \\ \\ \text{S} \text{ ~~~~~ OH} \end{array}$
Gly-Ala	146	1404/1406	-
Leu-Gly-Gly	245	1503/1505	1278
Gly-Gly-Val	231	1489/1491	1264
Gly-Pro-Gly-Gly	286	-	1319
Thr-Tyr-Ser	369	-	1402
Gly-Gly-Gly	189	-	1222
Phe-Gly-Gly	279	-	1312

7.3. Chapter 3

7.3.1. Disulfide-derived SAM formation (Figure 3.3)

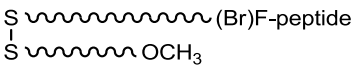
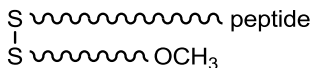
Stock solutions of the two disulfide-derived SAM components, $(\text{SCH}_2\text{CH}_2(\text{OCH}_2\text{CH}_2)_6\text{OCH}_3)_2$ and $(\text{SCH}_2\text{CH}_2\text{O}(\text{CH}_2\text{CH}_2\text{O})_7\text{CH}_2\text{CH}_2\text{COON}(\text{CO})_2(\text{CH}_2)_2)_2$, were prepared in DMSO (0.4 mg/mL) and stored at -18°C . For experiments, the stock solutions were diluted to 0.1 mg/mL and these were mixed in a 4:1 molar ratio (mPEG:NHS disulfides) by combining the two disulfide solutions in a 2.5 μL :1 μL (mPEG:NHS disulfides) ratio in a new microcentrifuge tube. Aliquots (0.4 μL) of this mixed solution were spotted onto the wells of the gold-coated MALDI chip. The gold-coated MALDI chip was sealed in a Petri dish and left overnight at ambient temperature then washed with acetonitrile and dried under a stream of nitrogen.

7.3.2. Peptide coupling (Figure 3.7)

Peptides were prepared at the specified concentration using PBS (1 \times , titrated to pH 8 with 0.1 M NaOH). Aliquots (1 μL) were spotted onto the wells of the gold-coated MALDI chip. A

piece of filter paper was placed in the bottom of a Petri dish, and wetted with deionised water (1 mL), to ensure the gold-coated MALDI chip did not dry out. The Petri dish was sealed and the gold-coated MALDI chip left at ambient temperature overnight. The gold-coated MALDI chip was washed with acetonitrile and dried under a stream of nitrogen. The peptides were analysed by MALDI-MS and the expected products are shown in Table 7.2.

Table 7.2: Peptides used for capture reaction and their corresponding product peaks observed by MALDI-MS analysis

Peptide	Molecular mass / g mol^{-1}	Expected product ion (Tagged SAM) $[\text{M}+\text{Na}]^+ / m/z$ 	Expected product ion (SAM) $[\text{M}+\text{Na}]^+ / m/z$ 
Gly-Ala	146	1404/1406	-
Leu-Gly-Gly	245	1503/1505	1278
Gly-Gly-Val	231	1489/1491	1264
Gly-Pro-Gly-Gly	286	-	1319
Thr-Tyr-Ser	369	-	1402
Gly-Gly-Gly	189	-	1222
Phe-Gly-Gly	279	-	1312

7.3.3. 4-Bromophenylalanine/phenylalanine coupling

A solution of 4-BrPhe (6.3 mg, 50 mM) or phenylalanine (8.6 mg, 50 mM) in 50:50 (v:v) DMF:PBS (1x, pH-adjusted to pH 8.0 with 0.1 M NaOH) (0.5 mL) was prepared. This solution was sonicated and centrifuged, and dilutions made with 50:50 (v:v) DMF:PBS (1x, pH-adjusted to pH 8.0 with 0.1 M NaOH). Aliquots of the supernatant (1 μL) were spotted onto wells on the gold-coated MALDI chip.

A piece of filter paper was placed in the bottom of a Petri dish and wetted with deionised water (1 mL), to ensure the gold-coated MALDI chip did not dry out. The gold-coated MALDI chip was sealed in a Petri dish and left overnight at ambient temperature. The gold-coated MALDI chip was washed with acetonitrile and dried under a stream of nitrogen.

7.4. Chapter 4

7.4.1. SAM formation

A stock solution of the SAM component, $(\text{HS}(\text{CH}_2)_{17}(\text{OCH}_2\text{CH}_2)_3\text{OH})$ was prepared in DMSO (0.4 mg/mL) and stored at $-18\text{ }^\circ\text{C}$. For experiments, the stock solution was diluted to 0.1 mg/mL. Aliquots (0.4 μL) were spotted onto the wells of the gold-coated MALDI chip. The gold-coated MALDI chip was sealed in a Petri dish and left overnight at ambient temperature then washed with acetonitrile and dried under a stream of nitrogen.

7.4.2. Operation of kHz plasma instrument

An atmospheric pressure kHz plasma instrument was used. The instrument consists of a simple electrode design with the powered electrode positioned 1 cm above the end of the quartz glass tube and the grounded electrode a further 2 cm above it. The quartz tube had an outer diameter of 6 mm and inner diameter of 4 mm, and a length of 100 mm. A gas flow of 2 standard litres per minute (SLM) of helium and various oxygen admixtures (ranging from 0 – 1.0%) were used. The kHz power supply was a high voltage pulsed nanosecond supply, with pulse voltage -3 - 18 kV at 75 Ω .

A gold-coated MALDI chip was placed under the quartz tube, on a plastic holder, with the end of the quartz tube a specified distance from the surface of the gold-coated MALDI chip. The gas flow was set to 2 SLM, with the specified percentage of oxygen added to the helium using a mixing control unit. The plasma was then generated by voltage application to the electrodes.

The gold-coated MALDI chip was moved so that the required well was under the end of the quartz tube, by using a pair of plastic tweezers. Timing of the cleaning was performed using a digital timer.

7.4.3. Operation of RF plasma instrument

A 13.56 MHz RF plasma instrument was operated at atmospheric pressure. The setup consists of one ground and one powered electrode driven by an RF generator at a frequency of 13.56 MHz applied through a matching unit. Typical operating powers are 15-25 W. The instrument was run with helium feed gas and varying oxygen or nitrogen admixtures.

A gold-coated MALDI chip was placed on a plastic holder, below the casing for the two electrodes, with the end of the electrodes 2 mm from the surface of the gold-coated MALDI chip. The gas flow was set to 1 SLM, with the specified percentage of oxygen or nitrogen added to the helium using a mixing control unit. The plasma was then generated by voltage application to the electrodes.

The gold-coated MALDI chip was moved so that the required well was under the end of the electrodes by using a pair of plastic tweezers. Timing of the cleaning was performed using a digital timer.

7.5. Chapter 5

7.5.1. Tryptic digestion

Up to 100 µg of protein was dissolved in TEAB (20 µL, 0.5 mM) in a 1.5 mL microcentrifuge tube. To this, tris-(2-carboxyethyl-phosphine) (TCEP) solution (2 µL, 50 mM) was added. The solution was vortexed and incubated at 60 °C for 1 h. Methyl methanethiosulfonate (MMTS) solution (1 µL, 200 mM) in isopropanol was added. The resulting sample solution was vortexed and incubated at ambient temperature for 10 minutes.

Trypsin solution (0.5 mM TEAB) was added to the protein solution to give a final trypsin:protein ratio between 1:100 and 1:20 (w/w). The solution was then incubated in a heating block at 37 °C overnight. An aliquot of the digest was then purified using a 10 µL C18 ZipTip, to clean the sample up for MALDI-MS analysis (see Section 7.5.3 for protocol).

7.5.2. Dimethyl labelling

For light or heavy dimethyl labelling, peptides (< 25 µg) were dissolved in TEAB (100 µL, 100 mM). For light dimethyl labelling, formaldehyde (4 µL, 4% (v/v)) was added. For heavy dimethyl labelling, D₂-formaldehyde (4 µL, 4% (v/v)) was added. The solutions were vortexed and centrifuged briefly. To each solution, sodium cyanoborohydride solution (4 µL, 0.6 M) was added. The solutions were vortexed and centrifuged briefly and then incubated at ambient temperature for 1 h in a fume hood. The labelling reactions were then quenched by adding ammonia solution (16 µL, 1% (v/v)), vortexed and centrifuged briefly. Formic acid (8 µL, 5% (v/v)) was then added to each. The differentially labelled solutions were then combined in defined ratios by volume to generate a series of ratio solutions.

For MALDI mass spectrometric analysis, aliquots of the solutions were purified using a 10 µL C18 ZipTip before mass spectrometric analysis (see Section 7.5.3 for protocol).

For LC-ESI mass spectrometric analysis, aliquots of the standard solutions were acidified to pH ~2 using a formic acid solution (10%) prior to mass spectrometric analysis.

7.5.3. ZipTip treatment

Conditioning (50:50 water:acetonitrile + 0.1% TFA) and washing (water + 0.1% TFA) solutions were prepared. A 10 μ L C18 ZipTip was attached to a pipette, which was set to dispense 10 μ L. The tip was conditioned twice by aspirating and expelling (to waste) aliquots (10 μ L) of the conditioning solution, and then washed twice by aspirating and expelling (to waste) with aliquots (10 μ L) of the washing solution. The sample solution was then loaded onto the ZipTip by repeated aspiration/expulsion of sample solution (10 μ L) through the ZipTip 10 times. The solid phase, to which peptides were bound was washed by aspirating and expelling (to waste) two aliquots (10 μ L) of the washing solution. The peptides were eluted from the ZipTip by aspirating an aliquot (10 μ L) of the conditioning solution and expelling it into a clean 0.5 mL microcentrifuge tube. This aliquot of conditioning solution was then aspirated and expelled back into the same centrifuge tube a further 9 times to increase sample recovery.

7.5.4. ProteinScape processing

7.5.4.1. MALDI-MS

Tandem mass spectral data were submitted to database searching using a locally-running copy of the Mascot program (Matrix Science Ltd., version 2.4), through the Bruker PS interface (version 2.1). Search criteria specified: enzyme, trypsin; fixed modifications, methylthio (C); variable modifications, oxidation (M), dimethyl(Heavy_K) (K), dimethyl(Heavy_N-term) (N-term), dimethyl(Light_K) (K) and dimethyl(Light_N-term) (N-term); peptide tolerance, 250 ppm; MS/MS tolerance, 0.5 Da; instrument, MALDI-TOF-TOF (the version and size of the database can be obtained from the Mascot result page. Kings database accessed between 18/06/2014 (694 sequences) and 31/07/2014 (707 sequences)). Results were filtered to accept only peptides with a p-value of 0.05 or lower.

7.5.4.2. LC-ESI-MS

Tandem mass spectral data were submitted to database searching using a locally-running copy of the Mascot program (Matrix Science Ltd., version 2.4), through the Bruker PS interface (version 2.1). Search criteria specified: enzyme, trypsin; fixed modifications, methylthio (C); variable modifications, oxidation (M), dimethyl(Heavy_K) (K), dimethyl(Heavy_N-term) (N-term), dimethyl(Light_K) (K) and dimethyl(Light_N-term) (N-term); peptide tolerance, 10 ppm; MS/MS tolerance, 0.1 Da; instrument, maXis (the version and size of the database can be obtained from the Mascot result page. Kings database accessed between 18/06/2014 (694 sequences) and 31/07/2014 (707 sequences)). Results were filtered to accept only peptides with a p-value of 0.05 or lower.

7.5.5. Mascot Distiller processing

A new project was created for each data file. A peak list was generated using the standard maXis method parameters. Mass spectra were submitted to a locally-running copy of the Mascot program (Matrix Science Ltd., version 2.4). Search criteria specified: enzyme, trypsin; fixed modifications, methylthio (C); variable modifications, oxidation (M), dimethyl (K), dimethyl (N-term), dimethyl:2H(4) (K), dimethyl:2H(4) (N-term), dimethyl:2H(6)13C(2) (K) and dimethyl:2H(6)13C(2) (N-term); quantification, dimethylation; peptide tolerance, 100 ppm; MS/MS tolerance, 0.1 Da; instrument, maXis (the version and size of the database can be obtained from the Mascot result page. Kings database accessed between 18/06/2014 (694 sequences) and 31/07/2014 (707 sequences). EcoProt database accessed between 27/02/2015 (4287 sequences) and 02/03/2015 (4287 sequences). Uniprot_mouse database accessed between 07/08/2015 (22123 sequences) and 11/08/2015 (22123 sequences)). Results were filtered to accept only peptides with an p-value of 0.05 or lower. Results reported as median (correlation threshold 0.7, std. err. threshold 0.3).

7.5.6. FDR determination

FDR values were determined by searching against a decoy database. The automatic decoy database search option was used on the Mascot search form. This option ensures that every time a protein sequence from the target database is tested, a decoy sequence of the same length is automatically generated and tested. FDR values were recorded as peptide matches above identity threshold

Table 7.3: FDR values calculated for analysis in Section 5.2.1.1 (- = no results from peptide search)

Ratio prepared (L:H)	Solution 1 FDR / %	Solution 2 FDR / %	Solution 3 FDR / %
0.1:1	0.00	0.00	0.00
0.2:1	0.00	0.00	0.00
0.33:1	0.00	0.00	0.00
0.5:1	0.00	0.00	0.00
0.67:1	0.00	0.00	0.00
1:1	0.00	0.00	0.00
1:1 (H:L)	0.00	0.00	0.00
1.5:1	0.00	0.00	0.00
2:1	0.00	0.00	0.00
3:1	0.00	0.00	0.00
5:1	-	0.00	0.00
10:1	0.00	0.00	0.00

Table 7.4: FDR values calculated for analyses in Section 5.2.1.2

Ratio prepared (L:H)	PS			MD		
	Solution 1 FDR / %	Solution 2 FDR / %	Solution 3 FDR / %	Solution 1 FDR / %	Solution 2 FDR / %	Solution 3 FDR / %
0.1:1	1.77	0.80	0.00	0.00	2.13	0.00
0.2:1	1.30	0.00	0.56	0.00	0.00	0.00
0.33:1	0.75	0.57	1.20	0.00	0.00	0.00
0.5:1	0.00	1.69	0.00	0.00	0.00	0.00
0.67:1	1.29	1.06	0.81	0.00	0.00	0.00
1:1	0.37	0.41	1.76	0.00	0.00	0.00
1:1 (H:L)	1.35	0.63	1.16	0.00	0.00	0.00
1.5:1	0.00	1.35	0.42	0.00	0.00	0.00
2:1	0.87	0.53	0.99	0.68	0.87	0.00
3:1	0.60	1.69	0.61	0.85	0.00	0.00
5:1	2.01	1.32	1.60	0.00	0.00	0.00
10:1	0.00	1.57	1.54	0.00	0.00	0.00

Table 7.5: FDR values calculated for analyses in Section 5.2.2

Ratio	Replicate	FDR / %
L:H	1	0.00
	2	0.00
	3	0.00
H:L	1	0.00
	2	0.00
	3	0.97

Table 7.6: FDR values calculated for analyses in Section 5.3

Solution	Solution 1 FDR / %	Solution 2 FDR / %	Solution 3 / FDR %
L:H 1%	0.27	0.29	0.23
L:H 5%	0.26	0.27	0.84
H:L 1%	0.24	0.26	0.15
H:L 5%	0.91	0.26	0.26

Table 7.7: FDR values calculated for analyses in Section 5.4

Solution	Replicate	FDR / %
U:S1	L:H	0.24
	H:L	0.00
U:S2	L:H	0.00
	H:L	0.30

7.6. Chapter 6

7.6.1. Tryptic digestion

Up to 100 µg of protein was dissolved in TEAB (20 µL, 0.5 mM) in a 1.5 mL microcentrifuge tube. To this, TCEP solution (2 µL, 50 mM) was added. This sample solution was vortexed and incubated at 60 °C for 1 h. MMTS solution (1 µL, 200 mM) in isopropanol was added. The resulting sample solution was vortexed and left at room temperature for 10 minutes.

Trypsin solution (in 0.5 mM TEAB) was added to the protein solution to give a final trypsin:protein ratio between 1:100 and 1:20 (w/w). The solution was then incubated in a heating block at 37 °C overnight. An aliquot of the digest was then purified using a 10 µL C18 ZipTip, to clean the sample up for MALDI-MS analysis (see Section 7.5.3 for protocol).

7.6.2. Labelling of tryptic digestion

For light or heavy dimethyl labelling, peptides (< 25 µg) were dissolved in TEAB (100 µL, 100 mM). Peptide solutions were acidified to ~ pH 4.5 with formic acid solution (0.1%). Sodium cyanoborohydride solutions were prepared in acetone and d₆-acetone (36 µg/µL). To a peptide solution (20 µL), the relevant sodium cyanoborohydride solution (5 µL) was added. The solutions were vortexed and centrifuged briefly and then incubated at ambient

temperature for 2 h in a fume hood. The differentially labelled solutions were then combined in a 1:1 (v/v) ratio.

For MALDI mass spectrometric analysis, an aliquot of the standard solution was purified using a 10 μ L C18 ZipTip before mass spectrometric analysis (see Section 7.5.3 for protocol).

Abbreviations

Abbreviations

Abbreviations

μ	Mean
μL	Microlitre
μM	Micromolar
ϕ	Potential
σ	Cross section
σ^2	Variance
λ	Mean free path
v	Velocity
ν	Frequency of encounters
\bar{v}	Mean speed
1-DGE	One-dimensional gel electrophoresis
2-DGE	Two-dimensional gel electrophoresis
4-BrPhe	4-Bromophenylalanine
AFM	Atomic force microscopy
Ala	Alanine
ALICE	Acid-labile isotope-coded extractants
APPJ	Atmospheric-pressure plasma jet
Arg	Arginine
B	Magnetic field
BS3	Bis(sulfosuccinimidyl) suberate
BSA	Bovine serum albumin
CHCA	α -Cyano-4-hydroxycinnamic acid
CID	Collision-induced dissociation
CV	Coefficient of variation
Cy	Cyclohexyl
DA	DataAnalysis
Da	Dalton
DBD	Dielectric barrier discharge
DCC	<i>N,N'</i> -Dicyclohexylcarbodiimide
DHB	2,5-Dihydroxybenzoic acid
DiLeu	<i>N,N</i> -Dimethyl leucine
DIPEA	<i>N,N</i> -Diisopropylethylamine
DIT	Dithranol
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
e^-	Electron
ECD	Electron capture dissociation
EDC	1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide
EG	Ethylene glycol

Abbreviations

E_k	Kinetic energy
emPAI	Exponentially Modified Protein Abundance Index
ESI	Electrospray ionisation
ETD	Electron transfer dissociation
f_c	Frequency
FDR	False discovery rate
FT-ICR	Fourier transform ion cyclotron resonance
Gly	Glycine
h	Hour
H:L	Heavy:light
His	Histidine
HPLC	High performance liquid chromatography
HSP	Heat shock protein
Hz	Hertz
ICAT	Isotope-coded affinity tag
ICP	Inductively coupled plasma
ICP-MS	Inductively coupled plasma mass spectrometry
ICR	Ion cyclotron resonance
IPTL	Isobaric peptide termini labelling
iTRAQ	Isobaric tag for relative and absolute quantification
kHz	Kilohertz
L	Litre
L	Length
L:H	Light:heavy
LC	Liquid chromatography
LC-ESI-MS	Liquid chromatography-electrospray ionisation-mass spectrometry
LC-MS	Liquid chromatography-mass spectrometry
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
Leu	Leucine
Leu-Enk	Leucine enkephalin
Lys	Lysine
m	Mass
m/z	Mass-to-charge ratio
MALDI	Matrix-assisted laser desorption/ionisation
MALDI-MS	Matrix-assisted laser desorption/ionisation-mass spectrometry
MD	Mascot Distiller
MHz	Megahertz
min	Minute
mL	Millilitre

Abbreviations

mM	Millimolar
mm	Millimetre
MMTS	Methanethiosulfonate
ms	Millisecond
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
Myo	Myoglobin
N	Number of tests
n_e	Electron density
NEXAFS	Near-edge X-ray absorption fine structure
n_g	Number of gas units per unit volume
ng	Nanogram
NHS	<i>N</i> -Hydroxysuccinimide
NHSS	<i>N</i> -Hydroxysulfosuccinimide
nL	Nanolitre
nmol	Nanomole
n_o	Number density of neutral atoms
PBS	Phosphate buffered saline
PEG	Polyethyleneglycol
PFP	2,3,4,5,6-Pentafluorophenol
Phe	Phenylalanine
Phos B	Phosphorylase B
pI	Isoelectric point
pmol	Picomole
Pro	Proline
PS	Bruker ProteinScape
PTM	Post-translational modification
q	Charge
RF	Radiofrequency
r_g	Radius of an atom of gas
RON	Reactive nitrogen species
ROS	Reactive oxygen species
s	Second
S	Significance level
S1	Stretched 1
S2	Stretched 2
S/N	Signal-to-noise ratio
SAM	Self-assembled monolayer
SD	Standard deviation

Abbreviations

SD _{geo}	Geometric standard deviation
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
Ser	Serine
SILAC	Stable isotope labelling in amino acid culture
SLM	Standard litre per minute
Std Dev.	Standard deviation
STM	Scanning tunnelling microscopy
t	Time
T	Transient duration
TCEP	Tris-(2-carboxyethyl-phosphine)
T _e	Electron temperature
TEAB	Triethylammonium bicarbonate
TFA	Trifluoroacetic acid
THAP	2',4',6'-Trihydroxyacetophenone
Thr	Threonine
T _i	Ion temperature
TMT	Tandem mass tag
T _{neutrals}	Neutrals temperature
ToF	Time-of-Flight
Trans	Transferrin
Trp	Tryptophan
Tyr	Tyrosine
U	Unstretched
UV-MALDI	Ultraviolet-matrix-assisted laser/desorption ionisation
V	Voltage
v/v	Volume/volume ratio
Val	Valine
VICAT	Visible isotope-coded affinity tag
V _s	Source potential
w/w	Weight/weight ratio
x	Distance from cold surface
XIC	Extracted ion chromatogram
XPS	X-ray photoelectron spectroscopy

References

References

References

1. N. L. Anderson and N. G. Anderson, *Electrophoresis*, 1998, **19**, 1853–1861.
2. V. C. Wasinger, S. J. Cordwell, A. Cerpa-Poljak, J. X. Yan, A. A. Gooley, M. R. Wilkins, M. W. Duncan, R. Harris, K. L. Williams, and I. Humphery-Smith, *Electrophoresis*, 1995, **16**, 1090–1094.
3. P. Mallick and B. Kuster, *Nat. Biotechnol.*, 2010, **28**, 695–709.
4. M. Mann and O. N. Jensen, *Nat. Biotechnol.*, 2003, **21**, 255–261.
5. M. Pertea and S. L. Salzberg, *Genome Biol.*, 2010, **11**, 206–212.
6. O. N. Jensen, *Curr. Opin. Chem. Biol.*, 2004, **8**, 33–41.
7. S. D. Patterson and R. H. Aebersold, *Nat. Genet.*, 2003, **33 Suppl**, 311–323.
8. H. Steen and M. Mann, *Nat. Rev. Mol. Cell Biol.*, 2004, **5**, 699–711.
9. O. Smithies, *Biochem. J.*, 1955, **61**, 629–641.
10. U. K. Laemmli, *Nature*, 1970, **227**, 680–685.
11. M. Gronow and G. Griffiths, *Fed. Eur. Biochem. Soc. Lett.*, 1971, **15**, 340–344.
12. S. Fazekas de St. Groth, R. G. Webster, and A. Datyner, *Biochim. Biophys. Acta*, 1963, **71**, 377–391.
13. C. R. Merril, D. Goldman, S. A. Sedman, and M. H. Ebert, *Science (80-.)*, 1981, **211**, 1437–1438.
14. P. H. O'Farrell, *J. Biol. Chem.*, 1975, **250**, 4007–4021.
15. A. Shevchenko, M. Wilm, O. Vorm, and M. Mann, *Anal. Chem.*, 1996, **68**, 850–858.
16. J. Klose, *Humangenetik*, 1975, **26**, 231–243.
17. B. Bjellqvist, K. Ek, P. G. Righetti, E. Gianazza, A. Görg, R. Westermeier, and W. Postel, *J. Biochem. Biophys. Methods*, 1982, **6**, 317–339.
18. H. Towbin, T. Staehelin, and J. Gordon, *Biochemistry*, 1979, **76**, 4350–4354.
19. W. N. Burnette, *Anal. Biochem.*, 1981, **112**, 195–203.
20. E. C. Jensen, *Anat. Rec.*, 2012, **295**, 369–371.
21. T. Mahmood and P. C. Yang, *N. Am. J. Med. Sci.*, 2012, **4**, 429–434.
22. R. Aebersold and M. Mann, *Nature*, 2003, **422**, 198–207.
23. K. Biemann and I. A. Papayannopoulos, *Acc. Chem. Res.*, 1994, **27**, 370–378.
24. W. E. Brown and F. Wold, *Biochemistry*, 1973, **12**, 828–834.
25. R. C. Dwivedi, V. Spicer, M. Harder, M. Antonovici, W. Ens, K. G. Standing, J. A. Wilkins, and O. V. Krokhin, *Anal. Chem.*, 2008, **80**, 7036–7042.
26. S. Wu, N. M. Lourette, N. Tolić, R. Zhao, E. W. Robinson, A. V. Tolmachev, R. D.

References

- Smith, and L. Paša-Tolić, *J. Proteome Res.*, 2009, **8**, 1347–1357.
27. N. L. Kelleher, H. Y. Lin, G. A. Valaskovic, D. J. Aaserud, E. K. Fridriksson, and F. W. McLafferty, *J. Am. Chem. Soc.*, 1999, **121**, 806–812.
28. N. Siuti and N. L. Kelleher, *Nat. Methods*, 2007, **4**, 817–821.
29. J. C. Tran, L. Zamdborg, D. R. Ahlf, J. E. Lee, A. D. Catherman, K. R. Durbin, J. D. Tipton, A. Vellaichamy, J. F. Kellie, M. Li, C. Wu, S. M. M. Sweet, B. P. Early, N. Siuti, R. D. LeDuc, P. D. Compton, P. M. Thomas, and N. L. Kelleher, *Nature*, 2011, **480**, 254–258.
30. B. Domon and R. Aebersold, *Nat. Biotechnol.*, 2010, **28**, 710–721.
31. A. L. McCormack, D. M. Schieltz, B. Goode, S. Yang, G. Barnes, D. Drubin, and J. R. Yates, *Anal. Chem.*, 1997, **69**, 767–776.
32. J. R. Yates, *J. Mass Spectrom.*, 1998, **33**, 1–19.
33. A. I. Nesvizhskii and R. Aebersold, *Mol. Cell. Proteomics*, 2005, **4**, 1419–1440.
34. A. J. Link, J. Eng, D. M. Schieltz, E. Carmack, G. J. Mize, D. R. Morris, B. M. Garvik, and J. R. Yates, *Nat. Biotechnol.*, 1999, **17**, 676–682.
35. M. P. Washburn, D. Wolters, and J. R. Yates, *Nat. Biotechnol.*, 2001, **19**, 242–247.
36. H. Liu, R. G. Sadygov, and J. R. Yates, *Anal. Chem.*, 2004, **76**, 4193–4201.
37. A. Schmidt, N. Gehlenborg, B. Bodenmiller, L. N. Mueller, D. Campbell, M. Mueller, R. Aebersold, and B. Domon, *Mol. Cell. Proteomics*, 2008, **7**, 2138–2150.
38. V. Lange, P. Picotti, B. Domon, and R. Aebersold, *Mol. Syst. Biol.*, 2008, **4**, 1–14.
39. R. Kiyonami, A. Schoen, A. Prakash, S. Peterman, V. Zabrouskov, P. Picotti, R. Aebersold, A. Huhmer, and B. Domon, *Mol. Cell. Proteomics*, 2011, **10**, DOI: 10.1074/mcp.M110.002931.
40. M. Yamashita and J. B. Fenn, *J. Phys. Chem.*, 1984, **88**, 4451–4459.
41. J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, C. M. Whitehouse, C. Meng, and M. Mann, *Science (80-.)*, 1989, **246**, 64–71.
42. G. Taylor, *Proc. R. Soc. A Math. Phys. Eng. Sci.*, 1964, **280**, 383–397.
43. M. S. Wilm and M. Mann, *Int. J. Mass Spectrom. Ion Process.*, 1994, **136**, 167–180.
44. M. Wilm, *Mol. Cell. Proteomics*, 2011, **10**, DOI: 10.1074/mcp.M111.009407.
45. K. K. Murray, R. K. Boyd, M. N. Eberlin, G. J. Langle, L. Li, and Y. Naito, *Pure Appl. Chem.*, 2013, **85**, 1515–1609.
46. M. Dole, L. L. Mack, R. L. Hines, R. C. Mobley, L. D. Ferguson, and M. B. Alice, *J. Chem. Phys.*, 1968, **49**, 2240–2249.
47. J. V. Iribarne and B. A. Thomson, *J. Chem. Phys.*, 1976, **64**, 2287–2294.

References

48. C. Whitehouse, R. Dreyer, M. Yamashita, and J. B. Fenn, *Anal. Chem.*, 1985, **57**, 675–679.
49. M. Karas, D. Bachmann, U. Bahr, and F. Hillenkamp, *Int. J. Mass Spectrom. Ion Process.*, 1987, **78**, 53–68.
50. M. Karas and F. Hillenkamp, *Anal. Chem.*, 1988, **60**, 2299–2301.
51. M. Karas, U. Bahr, A. Ingendoh, and E. Nordhoff, *Anal. Chim. Acta*, 1990, **241**, 175–185.
52. E. de Hoffmann and V. Sroobant, *Mass Spectrometry: Principles and Applications*, Wiley, 3rd edn., 2007.
53. R. Zenobi and R. Knochenmuss, *Mass Spectrom. Rev.*, 1998, **17**, 337–366.
54. R. Knochenmuss, *J. Mass Spectrom.*, 2002, **37**, 867–877.
55. M. Karas, M. Glückmann, and J. Schäfer, *J. Mass Spectrom.*, 2000, **35**, 1–12.
56. H. Ehring, M. Karas, and F. Hillenkamp, *Org. Mass Spectrom.*, 1992, **27**, 472–480.
57. M. Karas and R. Krüger, *Chem. Rev.*, 2003, **103**, 427–439.
58. T. W. Jaskolla and M. Karas, *J. Am. Soc. Mass Spectrom.*, 2011, **22**, 976–988.
59. Y. J. Bae, Y. S. Shin, J. H. Moon, and M. S. Kim, *J. Am. Soc. Mass Spectrom.*, 2012, **23**, 1326–1335.
60. W. E. Stephens, *Phys. Rev.*, 1946, **69**, 691.
61. W. C. Wiley and I. H. McLaren, *Rev. Sci. Instrum.*, 1955, **26**, 1150–1157.
62. M. Guilhaus, *J. Mass Spectrom.*, 1995, **30**, 1519–1532.
63. B. A. Mamyurin, V. I. Karataev, D. V. Shmikk, and V. A. Zagulin, *Sov. Phys. , JETP*, 1973, **26**, 45–48.
64. J. H. J. Dawson and M. Guilhaus, *Rapid Commun. Mass Spectrom.*, 1989, **3**, 155–159.
65. M. Guilhaus, D. Selby, and V. Mlynski, *Mass Spectrom. Rev.*, 2000, **19**, 65–107.
66. J. Coles and M. Guilhaus, *Trends Anal. Chem.*, 1993, **12**, 203–213.
67. M. Comisarow and A. Marshall, *Chem. Phys. Lett.*, 1974, **25**, 282–283.
68. M. Comisarow and A. Marshall, *Chem. Phys. Lett.*, 1974, **26**, 489–490.
69. J. Amster, *J. Mass Spectrom.*, 1996, **31**, 1325–1337.
70. D. G. Schmid, P. Grosche, H. Bandel, and G. Jung, *Biotechnol. Bioeng.*, 2001, **71**, 149–161.
71. A. G. Marshall, C. L. Hendrickson, and G. S. Jackson, *Mass Spectrom. Rev.*, 1998, **17**, 1–35.

References

72. A. G. Marshall, *Acc. Chem. Res.*, 1985, **18**, 316–322.
73. L. Schweikhard and A. G. Marshall, *J. Am. Soc. Mass Spectrom.*, 1993, **4**, 433–452.
74. M. Scigelova, M. Hornshaw, A. Giannakopoulos, and A. Makarov, *Mol. Cell. Proteomics*, 2011, **10**, M111.009431.
75. J. V. Johnson, R. A. Yost, P. E. Kelley, and D. C. Bradford, *Anal. Chem.*, 1990, **62**, 2162–2172.
76. B. Domon and R. Aebersold, *Science (80-.)*, 2006, **312**, 212–217.
77. J. K. Eng, A. L. McCormack, and J. R. Yates, *J. Am. Soc. Mass Spectrom.*, 1994, **5**, 976–989.
78. L. Sleno and D. a Volmer, *J. Mass Spectrom.*, 2004, **39**, 1091–1112.
79. F. W. McLafferty, P. F. Bente, R. Kornfeld, S.-C. Tsai, and I. Howe, *J. Am. Chem. Soc.*, 1995, **30**, 797–806.
80. S. A. McLuckey, *J. Am. Soc. Mass Spectrom.*, 1992, **3**, 599–614.
81. J. H. Beynon, R. G. Cooks, J. W. Amy, W. E. Baitinger, and T. Y. Ridley, *Anal. Chem.*, 1973, **45**, 1023A–1031A.
82. J. Gronowska, C. Paradisi, P. Traldi, and U. Vettori, *Rapid Commun. Mass Spectrom.*, 1990, **4**, 306–313.
83. P. Roepstorff and J. Fohlman, *Biomed. Mass Spectrom.*, 1984, **11**, 601.
84. K. Biemann, *Biomed. Environ. Mass Spectrom.*, 1988, **16**, 99–111.
85. V. H. Wysocki, G. Tsaprailis, L. L. Smith, and L. A. Breci, *J. Mass Spectrom.*, 2000, **35**, 1399–1406.
86. J. Yague, A. Paradela, M. Ramos, S. Ogueta, A. Marina, F. Barahona, J. A. López de Castro, and J. Vázquez, *Anal. Chem.*, 2003, **75**, 1524–1535.
87. D. N. Perkins, D. J. C. Pappin, D. M. Creasy, and J. S. Cottrell, *Electrophoresis*, 1999, **20**, 3551–3567.
88. J. R. Yates, *Electrophoresis*, 1998, **19**, 893–900.
89. M. Bantscheff, M. Schirle, G. Sweetman, J. Rick, and B. Kuster, *Anal. Bioanal. Chem.*, 2007, **389**, 1017–1031.
90. S. R. Hudson, F. L. Chadbourne, P. A. Helliwell, E. Pflimlin, J. E. Thomas-Oates, and A. Routledge, *ACS Comb. Sci.*, 2012, **14**, 97–100.
91. M. Gronborg, T. Z. Kristiansen, A. Iwahori, R. Chang, R. Reddy, N. Sato, H. Molina, O. N. Jensen, R. H. Hruban, M. G. Goggins, A. Maitra, and A. Pandey, *Mol. Cell. Proteomics*, 2006, **5**, 157–171.
92. W. Zhang, H. Zhang, L. Ning, B. Li, and M. Bao, *Front. Plant Sci.*, 2016, **7**, 1–13.

References

93. S.-E. Ong, B. Blagoev, I. Kratchmarova, D. B. Kristensen, H. Steen, A. Pandey, and M. Mann, *Mol. Cell. Proteomics*, 2002, **1**, 376–386.
94. X. Yao, A. Freas, J. Ramirez, P. A. Demirev, and C. Fenselau, *Anal. Chem.*, 2001, **73**, 2836–2842.
95. M. Miyagi and K. C. Sekhar Rao, *Mass Spectrom. Rev.*, 2007, **26**, 121–136.
96. S. a Gerber, J. Rush, O. Stemman, M. W. Kirschner, and S. P. Gygi, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 6940–6945.
97. R. J. Beynon, M. K. Doherty, J. M. Pratt, and S. J. Gaskell, *Nat. Methods*, 2005, **2**, 587–589.
98. J. Rivers, D. M. Simpson, D. H. L. Robertson, S. J. Gaskell, and R. J. Beynon, *Mol. Cell. Proteomics*, 2007, **6**, 1416–1427.
99. W. M. Old, K. Meyer-Arendt, L. Aveline-Wolf, K. G. Pierce, A. Mendoza, J. R. Sevinisky, K. A. Resing, and N. G. Ahn, *Mol. Cell. Proteomics*, 2005, **4**, 1487–1502.
100. D. Chelius and P. . Bondarenko, *J. Proteome Res.*, 2002, **1**, 317–323.
101. G. M. Toh-Boyo, S. S. Wulff, and F. Basile, *Anal. Chem.*, 2012, **84**, 9971–9980.
102. A. Prange and D. Pröfrock, *J. Anal. At. Spectrom.*, 2008, **23**, 432–459.
103. S.-E. Ong and M. Mann, *Nat. Protoc.*, 2006, **1**, 2650–2660.
104. J. S. Andersen, Y. W. Lam, A. K. L. Leung, S.-E. Ong, C. E. Lyon, A. I. Lamond, and M. Mann, *Nature*, 2005, **433**, 77–83.
105. S. P. Gygi, B. Rist, S. A. Gerber, F. Turecek, M. H. Gelb, and R. Aebersold, *Nat. Biotechnol.*, 1999, **17**, 994–999.
106. E. C. Yi, X.-J. Li, K. Cooke, H. Lee, B. Raught, A. Page, V. Aneliunas, P. Hieter, D. R. Goodlett, and R. Aebersold, *Proteomics*, 2005, **5**, 380–387.
107. A. Leitner and W. Lindner, *J. Chromatogr. B*, 2004, **813**, 1–26.
108. A. Thompson, J. Schäfer, K. Kuhn, S. Kienle, J. Schwarz, G. Schmidt, T. Neumann, and C. Hamon, *Anal. Chem.*, 2003, **75**, 1895–1904.
109. F. L. Brancia, A. Butt, R. J. Beynon, S. J. Hubbard, S. J. Gaskell, and S. G. Oliver, *Electrophoresis*, 2001, **22**, 552–559.
110. L. Dayon, A. Hainard, V. Licker, N. Turck, K. Kuhn, D. F. Hochstrasser, P. R. Burkhard, and J.-C. Sanchez, 2008, **80**, 2921–2931.
111. R. M. Sturm, C. B. Lietz, and L. Li, *Rapid Commun. Mass Spectrom.*, 2014, **28**, 1051–1060.
112. P. L. Ross, Y. N. Huang, J. N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M.

References

- Bartlet-Jones, F. He, A. Jacobson, and D. J. Pappin, *Mol. Cell. proteomics*, 2004, **3**, 1154–1169.
113. W. Yan and S. S. Chen, *Brief. Funct. Genomic. Proteomic.*, 2005, **4**, 27–38.
114. A. Pierce, R. D. Unwin, C. a Evans, S. Griffiths, L. Carney, L. Zhang, E. Jaworska, C.-F. Lee, D. Blinco, M. J. Okoniewski, C. J. Miller, D. a Bitton, E. Spooncer, and A. D. Whetton, *Mol. Cell. Proteomics*, 2008, **7**, 853–863.
115. S. Y. Ow, T. Cardona, A. Taton, A. Magnuson, P. Lindblad, K. Stensjö, and P. C. Wright, *J. Proteome Res.*, 2008, **7**, 1615–1628.
116. S. Y. Ow, M. Salim, J. Noirel, C. Evans, I. Rehman, and P. C. Wright, *J. Proteome Res.*, 2009, **8**, 5347–5355.
117. C. Evans, J. Noirel, S. Y. Ow, M. Salim, A. G. Pereira-Medrano, N. Couto, J. Pandhal, D. Smith, T. K. Pham, E. Karunakaran, X. Zou, C. a Biggs, and P. C. Wright, *Anal. Bioanal. Chem.*, 2012, **404**, 1011–1027.
118. A. Christoforou and K. S. Lilley, *Nat. Methods*, 2011, **8**, 911–913.
119. L. Ting, R. Rad, S. P. Gygi, and W. Haas, *Nat. Methods*, 2011, **8**, 937–940.
120. C. D. Wenger, M. V. Lee, A. S. Hebert, G. C. McAlister, D. H. Phanstiel, M. S. Westphall, and J. J. Coon, *Nat. Methods*, 2011, **8**, 933–935.
121. Y. Qiu, E. A. Sousa, R. M. Hewick, and J. H. Wang, *Anal. Chem.*, 2002, **74**, 4969–4979.
122. J.-L. Hsu, S.-Y. Huang, N.-H. Chow, and S.-H. Chen, *Anal. Chem.*, 2003, **75**, 6843–6852.
123. P. J. Boersema, R. Raijmakers, S. Lemeer, S. Mohammed, and A. J. Heck, *Nat. Protoc.*, 2009, **4**, 484–494.
124. Y. Wu, F. Wang, Z. Liu, H. Qin, C. Song, J. Huang, Y. Bian, X. Wei, J. Dong, and H. Zou, *Chem. Commun.*, 2014, **50**, 1708–1710.
125. M. L. Hennrich, S. Mohammed, a. F. M. Altelaar, and A. J. R. Heck, *J. Am. Soc. Mass Spectrom.*, 2010, **21**, 1957–1965.
126. J. M. Boutilier, H. Warden, A. A. Doucette, and P. D. Wentzell, *J. Chromatogr. B*, 2012, **908**, 59–66.
127. F. Xiang, H. Ye, R. Chen, Q. Fu, and L. Li, *Anal. Chem.*, 2010, **82**, 2817–2825.
128. D. C. Frost, T. Greer, F. Xiang, Z. Liang, and L. Li, *Rapid Commun. Mass Spectrom.*, 2015, **29**, 1115–1124.
129. D. C. Frost, T. Greer, and L. Li, *Anal. Chem.*, 2015, **87**, 1645–1654.
130. J. Song and H.-J. Kim, *Anal. Biochem.*, 2012, **423**, 269–276.

References

131. W. C. Bigelow, D. L. Pickett, and W. A. Zisman, *J. Colloid Sci.*, 1946, **1**, 513–538.
132. A. Ulman, *Chem. Rev.*, 1996, **96**, 1533–1554.
133. V. Chechik, R. M. Crooks, and C. J. Stirling, *Adv. Mater.*, 2000, **12**, 1161–1171.
134. G. M. Whitesides, J. K. Kriebel, and J. C. Love, *Sci. Prog.*, 2005, **1**, 17–48.
135. L. Sun, R. C. Thomas, and R. M. Crooks, *J. Am. Chem. Soc.*, 1991, **113**, 8550–8552.
136. F. Schreiber, *Prog. Surf. Sci.*, 2000, **65**, 151–256.
137. L. Yan, W. T. S. Huck, and G. M. Whitesides, *J. Macromol. Sci. Part C Polym. Rev.*, 2004, **44**, 175–206.
138. J. C. Love, L. A. Estroff, J. K. Kriebel, R. G. Nuzzo, and G. M. Whitesides, *Chem. Rev.*, 2005, **105**, 1103–1169.
139. N. K. Chaki and K. Vijayamohanan, *Biosens. Bioelectron.*, 2002, **17**, 1–12.
140. M. Frasconi, F. Mazzei, and T. Ferri, *Anal. Bioanal. Chem.*, 2010, **398**, 1545–1564.
141. K. L. Prime and G. M. Whitesides, *Science (80-.)*, 1991, **252**, 1164–1167.
142. B. Szeftczyk, R. Franco, J. Gomes, and M. N. Cordeiro, *J. Mol. Struct. THEOCHEM*, 2010, **946**, 83–87.
143. S. I. Jeon and J. D. Andrade, *J. Colloid Interface Sci.*, 1991, **142**, 159–166.
144. S. I. Jeon, J. H. Lee, J. D. Andrade, and P. G. de Gennes, *J. Colloid Interface Sci.*, 1991, **142**, 159–166.
145. E. Ostuni, R. G. Chapman, M. N. Liang, G. Meluleni, G. Pier, D. E. Ingber, and G. M. Whitesides, *Langmuir*, 2001, **17**, 6336–6343.
146. R. Michel, S. Pasche, M. Textor, and D. G. Castner, *Langmuir*, 2005, **21**, 12327–12332.
147. K. L. Prime and G. M. Whitesides, *J. Am. Chem. Soc.*, 1993, **115**, 10714–10721.
148. P. Harder, M. Grunze, R. Dahint, G. M. Whitesides, and P. E. Laibinis, *J. Phys. Chem. B*, 1998, **102**, 426–436.
149. H. Matsuura and K. Fukuhara, *J. Mol. Struct.*, 1985, **126**, 251–260.
150. R. V. Ulijn, B. Baragana, P. J. Halling, and S. L. Flitsch, *J. Am. Chem. Soc.*, 2002, **124**, 10988–10989.
151. P. J. Halling, R. V. Ulijn, and S. L. Flitsch, *Curr. Opin. Biotechnol.*, 2005, **16**, 385–392.
152. B. T. Houseman and M. Mrksich, *Chem. Biol.*, 2002, **9**, 443–454.
153. B. T. Houseman, E. S. Gawalt, and M. Mrksich, *Langmuir*, 2003, **19**, 1522–1531.
154. P. Wagner, M. Hegner, P. Kern, F. Zaugg, and G. Semenza, *Biophys. J.*, 1996, **70**,

References

- 2052–2066.
155. F. Tantakitti, J. Burk-Rafel, F. Cheng, R. Egnatchik, T. Owen, M. Hoffman, D. N. Weiss, and D. M. Ratner, *Langmuir*, 2012, **28**, 6950–6959.
 156. B. T. Houseman, J. H. Huh, S. J. Kron, and M. Mrksich, *Nat. Biotechnol.*, 2002, **20**, 270–274.
 157. F. Costa, I. F. Carvalho, R. C. Montelaro, P. Gomes, and M. C. L. Martins, *Acta Biomater.*, 2011, **7**, 1431–1440.
 158. V. Humblot, J.-F. Yala, P. Thebault, K. Boukerma, A. Héquet, J.-M. Berjeaud, and C.-M. Pradier, *Biomaterials*, 2009, **30**, 3503–3512.
 159. R. G. Nuzzo and D. L. Allara, *J. Am. Chem. Soc.*, 1983, **105**, 4481–4483.
 160. C. D. Bain, H. A. Biebuyck, and G. M. Whitesides, *Langmuir*, 1989, **5**, 723–727.
 161. H. A. Biebuyck, C. D. Bain, and G. M. Whitesides, *Langmuir*, 1994, **10**, 1825–1831.
 162. T. K. Ha, H. Bin Oh, J. Chung, T. G. Lee, and S. Y. Han, *Langmuir*, 2009, **25**, 3692–3697.
 163. H. Zhou, J. A. Ranish, J. D. Watts, and R. Aebbersold, *Nat. Biotechnol.*, 2002, **20**, 512–515.
 164. J.-S. Kim, Z. Dai, U. K. Aryal, R. J. Moore, D. G. Camp, S. E. Baker, R. D. Smith, and W.-J. Qian, *Anal. Chem.*, 2013, **85**, 6826–6832.
 165. N. Laurent, J. Voglmeir, A. Wright, J. Blackburn, N. T. Pham, S. C. C. Wong, S. J. Gaskell, and S. L. Flitsch, *Chembiochem*, 2008, **9**, 883–887.
 166. N. Laurent, R. Haddoub, J. Voglmeir, S. Wong, S. J. Gaskell, and S. L. Flitsch, *Chembiochem*, 2008, **9**, 2592–2596.
 167. M. Buckley, M. Collins, J. Thomas-Oates, and J. C. Wilson, *Rapid Commun. Mass Spectrom.*, 2009, **23**, 3843–3854.
 168. R. Haddoub, M. Dauner, F. a Stefanowicz, V. Barattini, N. Laurent, and S. L. Flitsch, *Org. Biomol. Chem.*, 2009, **7**, 665–670.
 169. S. Hudson, The University of York, 2010.
 170. R. Castangia, 2012.
 171. J. L. Trevor, K. R. Lykke, M. J. Pellin, and L. Hanley, *Langmuir*, 1998, **14**, 1664–1673.
 172. J. Su and M. Mrksich, *Angew. Chemie - Int. Ed.*, 2002, **41**, 4715–4718.
 173. M. Mrkisch, *ACS Nano*, 2008, **2**, 7–18.
 174. V. L. Marin, T. H. Bayburt, S. G. Sligar, and M. Mrksich, *Angew. Chemie - Int. Ed.*, 2007, **46**, 8796–8798.

References

175. Z. A. Gurard-Levin, M. D. Scholle, A. H. Eisenberg, and M. Mrksich, *Comb. Sci.*, 2011, **13**, 347–350.
176. J.-B. Hu, Y.-C. Chen, and P. L. Urban, *Anal. Chim. Acta*, 2013, **766**, 77–82.
177. F. Wilcoxon, *Biometrics Bull.*, 1945, **1**, 80–83.
178. H. B. Mann and D. R. Whitney, *Ann. Math. Stat.*, 1947, **18**, 50–60.
179. N. Nachar, *Tutor. Quant. Methods Psychol.*, 2008, **4**, 13–20.
180. M. R. Lockett, M. F. Phillips, J. L. Jarecki, D. Peelen, and L. M. Smith, *Langmuir*, 2007, **24**, 69–75.
181. P. Theato, *J. Polym. Sci. Part A Polym. Chem.*, 2008, **46**, 6677–6687.
182. R. M. Arnold, G. R. Sheppard, and J. Locklin, *Macromolecules*, 2012, **45**, 5444–5450.
183. E. Valeur and M. Bradley, *Chem. Soc. Rev.*, 2009, **38**, 606–631.
184. V. R. Pattabiraman and J. W. Bode, *Nature*, 2011, **480**, 471–479.
185. D. Bartczak and A. Kanaras, *Langmuir*, 2011, **27**, 10119–10123.
186. K. E. Nelson, L. Gamble, L. S. Jung, M. S. Boeckl, E. Naeemi, S. L. Golledge, T. Sasaki, D. G. Castner, C. T. Campbell, and P. S. Stayton, *Langmuir*, 2001, **17**, 2807–2816.
187. G. Hähner, *Chem. Soc. Rev.*, 2006, **35**, 1244–1255.
188. A. O.-Y. Chan, C.-M. Ho, H.-C. Chong, Y.-C. Leung, J.-S. Huang, M.-K. Wong, and C.-M. Che, *J. Am. Chem. Soc.*, 2012, **134**, 2589–2598.
189. D. Peelen, V. Kodoyianni, J. Lee, T. Zheng, M. R. Shortreed, and L. M. Smith, *J. Proteome Res.*, 2006, **5**, 1580–1585.
190. C. Jung, O. Dannenberger, Y. Xu, M. Buck, and M. Grunze, *Langmuir*, 1998, **14**, 1103–1107.
191. G. Mattson, E. Conklin, S. Desai, G. Nielander, M. D. Savage, and S. Morgensen, *Mol. Biol. Rep.*, 1993, **17**, 167–183.
192. S. H. Ahn, K. M. Park, Y. J. Bae, and M. S. Kim, *J. Am. Soc. Mass Spectrom.*, 2013, **24**, 868–876.
193. G. M. Toh-boyo, S. S. Wulff, and F. Basile, *Anal. Biochem.*, 2012, **84**, 9971–9980.
194. A. I. Gusev, W. R. Wilkinson, A. Proctor, and D. M. Hercules, *Anal. Chem.*, 1995, **67**, 1034–1041.
195. J. E. Cohen, *Curr. Dir. Psychol. Sci.*, 2004, **1**, 98–101.
196. F. Meng, M. C. Wiener, J. R. Sachs, C. Burns, P. Verma, C. P. Paweletz, M. T. Mazur, E. G. Deyanova, N. A. Yates, and R. C. Hendrickson, *J. Am. Soc. Mass*

References

- Spectrom.*, 2007, **18**, 226–233.
197. Y. Levin, *Proteomics*, 2011, **11**, 2565–2567.
198. O. Vorm, P. Roepstorff, and M. Mann, *Anal. Chem.*, 1994, **66**, 3281–3287.
199. S. J. Stranick, A. N. Parikh, Y.-T. Tao, D. L. Allara, and P. S. Weiss, *J. Phys. Chem.*, 1994, **98**, 7636–7646.
200. R. Jayakumar, C. Jayanthi, and L. Gomathy, *Int. J. Pept. Protein Res.*, 1995, **45**, 129–137.
201. M. N. Jones, C. P. Patrick, and M. C. Phillips, *J. Colloid Interface Sci.*, 1976, **55**, 116–125.
202. D. Thirumalai, D. K. Klimov, and R. I. Dima, *Curr. Opin. Struct. Biol.*, 2003, **13**, 146–159.
203. G. B. McGaughey, M. Gagné, and A. K. Rappé, *J. Biol. Chem.*, 1998, **273**, 15458–15463.
204. K. L. Copeland, J. A. Anderson, A. R. Farley, J. R. Cox, and G. S. Tschumper, *J. Phys. Chem. B*, 2008, **112**, 14291–14295.
205. S. Samal and K. E. Geckeler, *Chem. Commun.*, 2001, 2224–2225.
206. B. Heymann and H. Grubmüller, *Chem. Phys. Lett.*, 1999, **307**, 425–432.
207. J. Murray, D. Nowak, L. Pukenas, R. Azhar, M. Guillorit, C. Wälti, K. Critchley, S. Johnson, and R. S. Bon, *J. Mater. Chem. B*, 2014, **2**, 3741–3744.
208. H. Conrads and M. Schmidt, *Plasma Sources Sci. Technol.*, 2000, **9**, 441–454.
209. K. Raiber, A. Terfort, C. Benndorf, N. Krings, and H.-H. Strehblow, *Surf. Sci.*, 2005, **595**, 56–63.
210. S. Samukawa, M. Hori, S. Rauf, K. Tachibana, P. Bruggeman, G. Kroesen, J. C. Whitehead, A. B. Murphy, A. F. Gutsol, S. Starikovskaia, U. Kortshagen, J.-P. Boeuf, T. J. Sommerer, M. J. Kushner, U. Czarnetzki, and N. Mason, *J. Phys. D. Appl. Phys.*, 2012, **45**, 1–37.
211. N. S. J. Braithwaite, *Plasma Sources Sci. Technol.*, 2000, **9**, 517–527.
212. A. Schütze, J. Jeong, S. Babayan, J. Park, G. Selwyn, and R. Hicks, *Trans. Plasma Sci.*, 1998, **26**, 1685–1694.
213. D. B. Graves, *J. Phys. D. Appl. Phys.*, 2012, **45**, 1–42.
214. J. F. Kang, S. Liao, R. Jordan, and A. Ulman, *J. Am. Chem. Soc.*, 1998, **120**, 9662–9667.
215. J. I. Henderson, S. Feng, T. Bein, C. P. Kubiak, and W. Lafayette, *Langmuir*, 2000, **16**, 6183–6187.

References

216. O. Favrat, L. Aleya, and G. Monteil, *Surf. Coatings Technol.*, 2012, **206**, 3715–3720.
217. C. Yan, A. Götzhäuser, M. Grunze, and C. Wöll, *Langmuir*, 1999, **15**, 2414–2419.
218. G. Binnig and C. F. Quate, *Phys. Rev. Lett.*, 1986, **56**, 930–933.
219. G. Binnig and H. Rohrer, *Surf. Sci.*, 1983, **126**, 236–244.
220. C. Ji, N. Sadagopan, Y. Zhang, and C. Lepsy, *Anal. Chem.*, 2009, **81**, 9321–9328.
221. C. Ji, N. Guo, and L. Li, *J. Proteome Res.*, 2005, **4**, 2099–2108.
222. J. Munoz, T. Y. Low, Y. J. Kok, A. Chin, C. K. Frese, V. Ding, A. Choo, and A. J. R. Heck, *Mol. Syst. Biol.*, 2011, **7**, 1–13.
223. J. T. Wilson-Grady, W. Haas, and S. P. Gygi, *Methods*, 2013, **61**, 277–286.
224. P. J. Boersema, T. T. Aye, T. A. B. van Veen, A. J. R. Heck, and S. Mohammed, *Proteomics*, 2008, **8**, 4624–4632.
225. S. Ferret-Bernard, W. Castro-Borges, A. A. Dowle, D. E. Sanin, P. C. Cook, J. D. Turner, A. S. MacDonald, J. R. Thomas, and A. P. Mountford, *J. Proteomics*, 2012, **75**, 938–948.
226. H.-T. Lau, H. W. Suh, M. Golkowski, and S.-E. Ong, *J. Proteome Res.*, 2014, **13**, 4164–4174.
227. A. K. Singh, A. Singh, and M. Engelhardt, *The Lognormal Distribution in Environmental Applications*, 1997.
228. J. Cox and M. Mann, *Nat. Biotechnol.*, 2008, **26**, 1367–72.
229. P. Mortensen, J. W. Gouw, J. V. Olsen, S.-E. Ong, K. T. G. Rigbolt, J. Bunkenborg, J. Cox, L. J. Foster, A. J. R. Heck, B. Blagoev, J. S. Andersen, and M. Mann, *J. Proteome Res.*, 2010, **9**, 393–403.
230. B. MacLean, D. M. Tomazela, N. Shulman, M. Chambers, G. L. Finney, B. Frewen, R. Kern, D. L. Tabb, D. C. Liebler, and M. J. MacCoss, *Bioinformatics*, 2010, **26**, 966–968.
231. N. A. Karp, W. Huber, P. G. Sadowski, P. D. Charles, S. V. Hester, and K. S. Lilley, *Mol. Cell. Proteomics*, 2010, **9**, 1885–1897.
232. D. C. Frost, T. Greer, F. Xiang, Z. Liang, and L. Li, *Rapid Commun. Mass Spectrom.*, 2015, **29**, 1115–1124.
233. B. Ghesquière, V. Jonckheere, N. Colaert, J. Van Durme, E. Timmerman, M. Goethals, J. Schymkowitz, F. Rousseau, J. Vandekerckhove, and K. Gevaert, *Mol. Cell. Proteomics*, 2011, **10**, M110.006866-12.
234. P. J. Boersema, S. Mohammed, and A. J. R. Heck, *Anal. Bioanal. Chem.*, 2008, **391**, 151–159.

References

235. P. Horvatovich, B. Hoekman, N. Govorukhina, and R. Bischoff, *J. Sep. Sci.*, 2010, **33**, 1421–1437.
236. Y. Kim, C. Kang, B. Min, and G.-S. Yi, *BMC Med. Genomics*, 2015, **8**, S7.
237. J. C. Oliveros, <http://bioinfogp.cnb.csic.es/tools/venny/index.html>.
238. C. R. Genovese, N. A. Lazar, and T. Nichols, *Neuroimage*, 2002, **15**, 870–878.
239. R. A. Armstrong, *Ophthalmic Physiol. Opt.*, 2014, **34**, 502–508.
240. Y. Benjamini and Y. Hochberg, *J. R. Stat. Soc. Ser. B*, 1993, **46**, 289–300.
241. L. Kääll, J. D. Storey, M. J. MacCoss, and W. S. Noble, *J. Proteome Res.*, 2008, **7**, 29–34.
242. W. S. Noble, *Nat. Biotechnol.*, 2009, **27**, 1135–1137.
243. B. Bhattacharya and D. Habtzghi, *Am. Stat.*, 2002, **56**, 202–206.
244. J. Concato and J. A. Hartigan, *J. Investig. Med.*, 2016, **0**, 1–6.
245. J. D. Storey and J. D. Tibshirani, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 9440–9445.
246. T. V Perneger, *Br. Med. J.*, 1998, **316**, 1236–1238.
247. D. W. Mahoney, T. M. Therneau, C. J. Heppelmann, L. Higgins, L. M. Benson, R. M. Zenka, P. Jagtap, G. L. Nelsestuen, H. R. Bergen, and A. L. Oberg, *J. Proteome Res.*, 2011, **10**, 4325–4333.
248. G. L. Warren, M. Summan, X. Gao, R. Chapman, T. Hulderman, and P. P. Simeonova, *J. Physiol.*, 2007, **582**, 825–841.
249. I. Ntai, K. Kim, R. T. Fellers, O. S. Skinner, A. D. Smith, B. P. Early, J. P. Savaryn, R. D. LeDuc, P. M. Thomas, and N. L. Kelleher, *Anal. Chem.*, 2014, **86**, 4961–4968.
250. L. Martens, H. Hermjakob, P. Jones, M. Adamski, C. Taylor, D. States, K. Gevaert, J. Vandekerckhove, and R. Apweiler, *Proteomics*, 2005, **5**, 3537–3545.
251. L. Gatto and A. Christoforou, *Biochim. Biophys. Acta*, 2014, **1844**, 42–51.
252. A. S. Deshmukh, M. Murgia, N. Nagaraj, J. T. Treebak, J. Cox, and M. Mann, *Mol. Cell. Proteomics*, 2015, **14**, 841–53.
253. J. C. Silva, M. V Gorenstein, G.-Z. Li, J. P. C. Vissers, and S. J. Geromanos, *Mol. Cell. Proteomics*, 2005, **5**, 144–156.
254. H. H. Meyer, *Biochim. Biophys. Acta*, 2005, **1744**, 108–119.
255. S. S. Baxter, C. F. Dibble, W. C. Byrd, J. Carlson, C. R. Mack, I. Saldarriaga, and S. Bencharit, *Mol. Biosyst.*, 2014, **10**, 1881–1889.
256. M. J. Adams, G. H. Ellis, S. Gover, C. E. Naylor, and C. Phillips, *Structure*, 1994, **2**, 651–668.

References

257. P. Liang and T. H. MacRae, *J. Cell Sci.*, 1997, **110**, 1431–1440.
258. K. Kurogi, Y. Sakakibara, Y. Kamemoto, S. Takahashi, S. Yasuda, M.-C. Liu, and M. Suiko, *FEBS J.*, 2010, **277**, 3804–3811.
259. L. E. Hightower, *Cell*, 1991, **66**, 191–197.
260. H. Singh, M. A. Cousin, and R. H. Ashley, *FEBS J.*, 2007, **274**, 6306–6316.
261. H. Y. Baek, J. W. Lim, H. Kim, J. M. Kim, J. S. Kim, H. C. Jung, and K. H. Kim, *Biochem. J.*, 2004, **379**, 291–299.
262. E. Leikina, A. Defour, K. Melikov, J. H. Van der Meulen, K. Nagaraju, S. Bhuvanendran, C. Gebert, K. Pfeifer, L. V Chernomordik, and J. J. K, *Sci. Rep.*, 2015, **5**, 1–12.
263. V. Bizzarro, A. Petrella, and L. Parente, *J. Cell. Physiol.*, 2012, **227**, 3007–3015.
264. A. K. McNeil, U. Rescher, V. Gerke, and P. L. McNeil, *J. Biol. Chem.*, 2006, **281**, 35202–35207.
265. R. L. Lieber, L. E. Thornell, and J. Fridén, *J. Appl. Physiol.*, 1996, **80**, 278–284.
266. D. L. Morgan and D. G. Allen, *J. Appl. Physiol.*, 1999, **87**, 2007–2015.
267. J. Fridén and R. L. Lieber, *Med. Sci. Sports Exerc.*, 1992, **24**, 521–530.
268. V. Vasiliou, A. Pappa, and D. R. Petersen, *Chem. Biol. Interact.*, 2000, **129**, 1–19.
269. C.-H. Chen, L. Sun, and D. Mochly-Rosen, *Cardiovasc. Res.*, 2010, **88**, 51–57.
270. J. E. Walker, M. Saraste, M. Runswick, and N. J. Gay, *EMBO J.*, 1982, **1**, 945–951.
271. C. D. Moyes, B. J. Battersby, and S. C. Leary, *J. Exp. Biol.*, 1998, **201**, 299–307.
272. E. Nogales, M. Whittaker, R. A. Milligan, and K. H. Downing, *Cell*, 1999, **96**, 79–88.
273. S. Goswami, N. L. Sheets, J. Zavadil, B. K. Chauhan, E. P. Bottinger, V. N. Reddy, M. Kantarow, and A. Cvekl, *Invest. Ophthalmol. Vis. Sci.*, 2003, **44**, 2084–2093.
274. L. Parrotta, C. Faleri, M. Cresti, and G. Cai, *Planta*, 2016, **243**, 43–63.
275. F. Weis, L. Moullintraffort, C. Heichette, D. Chrétien, and C. Garnier, *J. Biol. Chem.*, 2010, **285**, 9525–9534.
276. S. Wu, J. M. Storey, and K. B. Storey, *J. Exp. Zool. Part A Ecol. Genet. Physiol.*, 2009, **311**, 57–67.
277. R. Parker, T. J. Flowers, A. L. Moore, and N. V. J. Harpham, *J. Exp. Bot.*, 2006, **57**, 1109–1118.
278. M.-J. Basurko, M. Marche, M. Darriet, and A. Cassaigne, *IUBMB Life*, 1999, **48**, 525–529.
279. C. Xu and B. Huang, *J. Exp. Bot.*, 2008, **59**, 4183–4194.

References

280. J. A. MacDonald and K. B. Storey, *Arch. Biochem. Biophys.*, 2002, **408**, 279–285.
281. J. Hyzewicz, J. Tanihata, M. Kuraoka, N. Ito, Y. Miyagoe-Suzuki, and S. Takeda, *Free Radic. Biol. Med.*, 2015, **82**, 122–136.
282. P. Wang, F. G. Bouwman, and E. C. M. Mariman, *Proteomics*, 2009, **9**, 2955–2966.
283. A.-C. Gingras, R. Aebersold, and B. Raught, *J. Physiol.*, 2005, **563**, 11–21.
284. W.-J. Qian, J. M. Jacobs, T. Liu, D. G. Camp, and R. D. Smith, *Mol. Cell. Proteomics*, 2006, **5**, 1727–44.
285. J. Zhai, X. Liu, Z. Huang, and H. Zhu, *J. Am. Soc. Mass Spectrom.*, 2009, **20**, 1366–1377.
286. P. Wu, A. K. Feldman, A. K. Nugent, C. J. Hawker, A. Scheel, B. Voit, J. Pyun, J. M. J. Fréchet, K. B. Sharpless, and V. V. Fokin, *Angew. Chemie - Int. Ed.*, 2004, **43**, 3928–3932.
287. J. E. Moses and A. D. Moorhouse, *Chem. Soc. Rev.*, 2007, **36**, 1249–1262.
288. V. Hong, S. I. Presolski, C. Ma, and M. G. Finn, *Angew. Chemie - Int. Ed.*, 2009, **48**, 9879–9883.
289. C. J. Sih, J. B. Heather, R. Sood, P. Price, G. Peruzzotti, L. F. Hsu Lee, and S. S. Lee, *J. Am. Chem. Soc.*, 1975, **97**, 865–874.
290. O. F. Foot, D. W. Knight, A. C. L. Low, and Y. Li, *Tetrahedron Lett.*, 2007, **48**, 647–650.
291. Y. K. Kang, K. S. Lee, K. H. Yoo, K. J. Shin, D. C. Kim, C.-S. Lee, J. Y. Kong, and D. J. Kim, *Bioorg. Med. Chem. Lett.*, 2003, **13**, 463–466.
292. X.-L. Sun, C. L. Stabler, C. S. Cazalis, and E. L. Chaikof, *Bioconjug. Chem.*, 2006, **17**, 52–57.
293. J. Yang, A. Moraillon, A. Siriwardena, R. Boukherroub, F. Ozanam, A. C. Gouget-Laemmel, and S. Szunerits, *Anal. Chem.*, 2015, **87**, 3721–3728.
294. S. Choi, J. S. Seo, R. F. H. Bohaty, and C. D. Poulter, *Bioconjug. Chem.*, 2014, **25**, 269–275.
295. C. J. Koehler, M. Strozynski, F. Kozielski, A. Treumann, and B. Thiede, *J. Proteome Res.*, 2009, **8**, 4333–4341.
296. C. J. Koehler, M. Arntzen, M. Strozynski, A. Treumann, and B. Thiede, *Anal. Chem.*, 2011, **83**, 4775–4781.
297. C. J. Koehler, M. Arntzen, G. A. de Souza, and B. Thiede, *Anal. Chem.*, 2013, **85**, 2478–2485.
298. P. A. Jekel, W. J. Weijer, and J. J. Beintema, *Anal. Biochem.*, 1983, **134**, 347–354.