



Development and Application of Methods for Qualitative and Quantitative Analysis of Amino Acid Post-Translational Modifications Using Liquid Chromatography Coupled to Mass Spectrometry

Nikita D. Loik

Chemistry Research Laboratory
Department of Organic Chemistry
Corpus Christi College

Michaelmas Term 2012

A thesis submitted to the board of the Faculty of Physical Sciences
at the University of Oxford in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

Development and Application of Methods for Qualitative and Quantitative Analysis of Amino Acid Post-Translational Modifications Using Liquid Chromatography Coupled to Mass Spectrometry

The 2-oxoglutarate and ferrous ion dependent oxygenases are a super family of enzymes that are involved in a wide range of biological processes regulated through the mechanism of post-translational modification (PTM). Such biological processes include hypoxia sensing (through regulating HIF transcription), fatty acid metabolism (through carnitine production), transcriptional regulation (through demethylation of histones), and collagen structure formation (through proline and lysine hydroxylation).

To understand the underlying mechanisms of such regulatory processes, and to develop clinically useful inhibitors, and thereby regulate these processes in living organisms, requires sensitive methods for monitoring enzyme activity. The use of indirect methods such as quantification of reaction products ($^{14}\text{CO}_2$ or succinate) can be problematic, as both products can result from competitive reactions. Alternative direct measurement of substrate modifications using mass spectrometry-based proteomics can be applied; however, (1) for this technique the limit of detection is often prohibitive, (2) the method is best suited for the confirmation of known modifications, rather than for the discovery of new modifications, and (3) sequence coverage may often be only 60%, and therefore many modifications can be missed. The aim of the research presented in this thesis was to develop amino acid analysis and to apply these methods to the identification and quantification of PTMs catalysed by 2-oxoglutarate and ferrous ion dependent oxygenases.

A range of LC-MS approaches were investigated including: (1) C18 reversed phase chromatography of quinoline derivatised amino acids, (2) ion pairing chromatography, and (3) mixed mode chromatography with either UV, or conventional molecular MS, or isotope ratio mass spectrometry detection. Analysis of the elution patterns for those separation techniques enabled estimation of the retention parameters of modified amino acids and the identification of the modifications, where no standards were available.

The most sensitive approach developed employed mixed mode chromatography coupled to isotope ratio mass spectrometry which was optimised for the analysis of modified amino acids. This was shown to have a limit of detection two orders of magnitude lower ($0.01\mu\text{M}$) than other conventional mass spectrometry techniques. Using amino acid labelling in cell culture, a quantification protocol was developed which employed a non-labelled internal standard and selectively labelled cell culture. The method was shown to be suitable for both very accurate quantification at low concentration levels and metabolic studies, allowing us to track back the modifications to their precursors.

The analytical methods developed for amino acid analysis were successfully applied to the analysis of modifications resulting from 2-oxoglutarate and Fe dependent oxygenase activity. Stereochemistry of lysyl hydroxylation in the splicing regulatory protein Luk7L2 by JmjD6 as well as of the self hydroxylation of the JmjD6 was identified. The stereochemistry was shown to be different from that of previously reported for the collagen hydroxyline, hydroxylated by the another member of this enzyme family. mbP4H enzyme was shown to catalyse prolyl hydroxylation of taODD resulting in 4R-hydroxyprolyl. Amino acid analysis was used in order to verify the mechanism of the hBBOX catalysed rearrangement of Mildronate. Using the method developed for the analysis of non-derivatised amino acids the screening of potential substrates of hBBOX enzyme was carried out and two new substrates were identified. The isotope ratio mass spectrometry protocol was applied to the study of histones from cell culture; low levels of hydroxylated and methylated amino acids were quantified.

The analytical methods described were developed to complement to the well established proteomics techniques. The methods developed enable investigation into the region- and stereo- chemistry of the modified groups within the modified AA residue and has proved to be a powerful tool of exploratory PTM analysis.

Acknowledgements

I would like to express my sincere gratitude to both my supervisors, Dr James McCullagh and Professor Chris Schofield, for engaging me in this exciting project. I would like to thank Chris for letting me into the realm of biochemistry research and providing a word of advice and encouragement when this, for me, novel discipline was showing its unkind side. I would like to thank James for sharing his technical and scientific expertise with me and for demonstrating a truly critical analytical approach to the chemistry problems.

I owe a enormous debt of thanks to Lingzhi Gong, Robin Procter, Colin Sparrow, and Dr James Wickens for sharing their profound knowledge and providing endless support with the very nuts and bolts of the chromatography and mass spectrometry. Without them my scientific journey would have stopped at the first red light an instrument flashed.

I am most grateful to Monica Mantri whose cheerful attitude buoyed me when we became bogged down with the project. I would also like to thank her for mentoring my enzymatic assays and sharing her expertise in biochemistry lab work.

I would like to thank Dr Emily Flashman for introducing me to the work horse of chemical biology – the *E. coli* cell – thus making it less of a Trojan horse for me.

I would like to thank Dr Mukram Mackeen for showing me the ropes while I was sailing through kidney cells in the Petri dish. I am especially grateful to Louise Walport and Wei Shen Aik who kept watch over me tirelessly, thus preventing the stress-induced death of numerous kidney cells in the Petri dish and also neurons in my body.

I would like to thank James McCullagh for his constructive criticism on the thesis structure. My greatest thanks go to Monica Mantri, Anna Rydzik, Ivanhoe Leung, Areej Abuhammad for the innumerable cycles of proof reading. I am greatly indebted to Louise Walport, Adam Hardy and Inga Pfeffer for dotting all the i's and crossing all the t's when proof reading. I would also like to thank Dr Maggie Charles for explaining my grammar mistakes to me.

It was a special experience for me to work with Wei Shen Aik, Anna Boleininger, Luc Henry, Monica Mantri, Anna Rydzik, Rok Sekirnik, Hannah Tarhonskaya, Louise Walport – who, apart from being great researchers, were always good friends.

Special thanks to Inga Shpilevaya, Emma Mckinley, Karim Sutton and Amber Thompson, who made my office work a much more joyful and pleasant experience, and who created a Wonderland, where 'it's always tea-time'.

My research was made possible by Scatcherd European Scholarship. Additional financial support came from the British Mass Spectrometry Society and Corpus Christi College.

Finally, I would like to thank my family for their support during the last four years and for being understanding even when I showed the gloomiest side of myself.

Table of Contents

1	CHAPTER ONE – INTRODUCTION	1
1.1	2OG- AND Fe(II)- DEPENDENT OXYGENASES: OVERVIEW.....	4
1.1.1	<i>Oxygen Sensing: Maintaining Homeostasis</i>	<i>5</i>
1.1.2	<i>Collagen Post-Translational Modifications: Maintaining Structural Integrity</i>	<i>7</i>
1.1.3	<i>Antibiotic Biosynthesis</i>	<i>9</i>
1.1.4	<i>Fatty Acid Metabolism</i>	<i>10</i>
1.1.5	<i>Histone Modifications</i>	<i>11</i>
1.2	ASSAYS FOR THE 2OG OXYGENASES.....	12
1.3	ANALYSIS OF POST-TRANSLATIONAL MODIFICATIONS BY PROTEOMICS.....	14
1.3.1	<i>Stable Isotope Labelling in Cell Culture</i>	<i>15</i>
1.3.2	<i>Isobaric Tags for Relative and Absolute Quantification.....</i>	<i>16</i>
1.3.3	<i>Absolute Quantification Using an Internal Standard</i>	<i>17</i>
1.3.4	<i>Identification and Quantification of Post-Translational Modifications Using Proteomic Techniques</i>	<i>17</i>
1.4	AIMS OF THE WORK DESCRIBED IN THIS THESIS	20
1.5	REFERENCES	21
2	CHAPTER TWO – APPLICATION OF LC TO THE ANALYSIS OF PTMS	24
2.1	CHROMATOGRAPHIC SEPARATION OF AAS.....	24
2.1.1	<i>Derivatised AA analysis.....</i>	<i>30</i>
2.2	CHROMATOGRAPHY TESTING AND DEVELOPMENT	33
2.2.1	<i>RP chromatography of AQC-derivatised AAs</i>	<i>36</i>
2.2.1.1	Method background	36
2.2.1.2	Method development	36
2.2.2	<i>RP-IP chromatography.....</i>	<i>42</i>
2.2.2.1	Method background	42
2.2.2.2	Method development	45

2.2.3	Mixed mode chromatography of AAs	48
2.2.3.1	Method background	48
2.2.3.2	Method development	49
2.3	CONCLUSIONS	60
2.4	REFERENCES	61
3	APPLICATION OF RI TO THE ANALYSIS OF PTMS	64
3.1	INTRODUCTION OF THE RI SCALE	64
3.1.1	RI Prediction	66
3.1.2	Revision of the RI Concept.....	66
3.1.3	RI Scale for AA Analysis.....	66
3.2	RETENTION PATTERN RECOGNITION.....	68
3.2.1	Comparison of Different Chromatography Modes.....	68
3.2.1.1	Gradient invariability	70
3.2.1.2	Comparison of IP reagents.....	71
3.2.1.3	Comparison of different chromatographic modes.....	72
3.2.2	Analysis of Retention Patterns of Modified AAs	74
3.2.2.1	Homologous Series Analysis	74
3.2.2.2	Hydroxyproline Isomers.....	78
3.2.2.3	Hydroxylated AAs.....	79
3.2.2.4	N-Methylated AAs.....	80
3.3	CONCLUSIONS	82
3.4	REFERENCES	83
4	CHAPTER FOUR – QUANTITATIVE ANALYSIS OF AAS: METHOD VALIDATION AND DESCRIPTION	84
4.1	DETECTION TECHNIQUES	84
4.2	QUANTIFICATION METHOD DESCRIPTION.....	87
4.3	REPORTING THE RESULTS OF THE CALIBRATION TESTS.	91
4.3.1	Linearity Testing.....	92
4.3.2	LoD (LoD).....	94
4.3.3	Limit of Quantification (LoQ)	97

4.3.4	<i>Homoscedasticity</i>	98
4.3.5	<i>Quantitative Method Description Summary</i>	99
4.4	QUANTITATIVE METHOD VALIDATION	100
4.4.1	<i>Calibration Design</i>	100
4.4.2	<i>Results and Discussion</i>	102
4.4.2.1	Linearity	106
4.4.2.2	Homoscedasticity of the Experimental Data.....	110
4.4.2.3	LoDs and LoQs for Modified AAs.....	113
4.4.2.4	Comparison of Modified and Non-Modified AAs.....	114
4.4.3	<i>Conclusions</i>	115
4.5	REFERENCES.....	116
5	CHAPTER FIVE – APPLICATION OF AA ANALYSIS TO THE STUDY OF PTMS	118
5.1	AA RESIDUE HYDROXYLATION STEREOCHEMISTRY	118
5.1.1	<i>Hydroxylation of Lysyl Residues Catalysed by JmjD6</i>	118
5.1.2	<i>Hydroxylation of Proline Catalysed by mbP4H</i>	123
5.2	VERIFICATION OF ARYLAMINE N-ACETYLTRANSFERASE INHIBITION MECHANISM.....	126
5.3	STUDY OF AN HBBOX CATALYSED REARRANGEMENT	130
5.4	IDENTIFICATION OF POTENTIAL SUBSTRATES OF HBBOX	135
5.5	CONCLUSIONS	139
5.6	REFERENCES	141
6	CHAPTER SIX – DEVELOPMENT OF LC-IRMS FOR QUANTITATIVE ANALYSIS OF PTMS	143
6.1	METHOD BACKGROUND	143
6.1.1	<i>Isotope Ratio MS</i>	143
6.1.2	<i>Chromatographic Methods for LC-IRMS</i>	148
6.2	DETECTION AND QUANTIFICATION OF PTMS USING LC-IRMS	149
6.2.1	<i>Defining the Quantity of the Standard to be Added</i>	152
6.2.2	<i>Calibration Methods</i>	155
6.2.2.1	Theoretical Considerations	155

6.2.3	<i>Simulation Calibration Graphs</i>	158
6.2.3.1	Calibration Experiment Design.....	162
6.2.3.2	Comparison of Linear Fit and Model Fit	163
6.2.3.3	Limit of Detection	166
6.3	THE EFFECT OF THE ISOTOPE RATIO CORRECTIONS ON THE QUANTIFICATION METHOD	168
6.4	APPLICATION OF THE METHOD TO THE ANALYSIS OF HISTONE PTMS	172
6.4.1	<i>Histone PTMs Identification</i>	176
6.4.2	<i>Histone PTMs Quantification</i>	180
6.5	CONCLUSIONS AND FURTHER WORK	188
6.6	REFERENCES	189
7	CHAPTER SEVEN – SUMMARY	192
8	CHAPTER EIGHT – MATERIALS AND METHODS	196
8.1	CHEMICALS AND STANDARDS.....	196
8.2	CHROMATOGRAPHY METHODS	197
8.2.1	<i>Chromatographic Separation of AQC Derivatives of AAs</i>	197
8.2.2	<i>Chromatographic Separation of the Non-Derivatised AA Using Ion-Pairing Chromatography</i> ..	198
8.2.3	<i>Chromatographic Separation of Non-Derivatised AA Using Mixed Mode Chromatography</i>	199
8.2.4	<i>Chromatographic Separation of the Non-Derivatised AA by MM Chromatography Using Carbon Free Mobile Phase</i>	200
8.2.5	<i>List of Columns Used</i>	201
8.2.6	<i>Derivatisation with 6-Aminoquinolyl-N-Hydroxysuccinimidyl Carbamate</i>	201
8.3	DETECTION METHODS.....	204
8.3.1	<i>UV Detection</i>	204
8.3.2	<i>SIM and SRM Mode MS Detection</i>	204
8.3.3	<i>Scanning Mode MS Detection</i>	206
8.3.4	<i>IRMS Detection</i>	206
8.4	CHROMATOGRAPHY MS DATA PROCESSING	207
8.5	SIGNAL PROCESSING.....	207

8.5.1	Mass Chromatogram Data Presentation	207
8.5.2	Data Smoothing	209
8.6	HISTONE EXTRACTION AND PURIFICATION	215
8.6.1	Cell Growth	215
8.6.2	Acid Extraction of Histones	216
8.6.3	Total Protein Content Determination (Bicinchoninic Acid Assay)	217
8.6.4	Histone Separation Using Gel Electrophoresis	217
8.6.5	In Gel Histone Digestion.....	218
8.7	PEPTIDE AND PROTEIN DIGESTION	220
8.7.1	Acid Hydrolysis	220
8.7.2	Enzymatic Digestion.....	221
9	APPENDICES	A
9.1	APPENDIX A – CORRECTION COMPARISON	A
9.1.1	The Craig Correction	a
9.1.2	The Santrock, Studley, and Hayes correction	c
9.1.3	The Allison correction.....	e
9.2	APPENDIX B – RETENTION DATA	H
9.2.1	Reversed Phase Mixed Mode Chromatography Retention Data	h
9.2.2	Reversed Phase Chromatography of AQC Derivatised Amino Acids Retention Data	j
9.2.3	Reversed Phase Ion Pairing Chromatography Retention Data	k
9.3	REFERENCES	M

2OG	2-oxoglutarate	M	molar
2S,5R-HyL	(5R)-5-hydroxy- <i>L</i> -lysine	MeArg	<i>N</i> -methylarginine
3LHyArg	3 <i>L</i> -hydroxy- <i>L</i> -arginine	MeGly	sarcosine
3RHyArg	3 <i>R</i> -hydroxy- <i>L</i> -arginine	MeHis (1)	1-methyl- <i>L</i> -histidine
4D-HyP	cis-4-hydroxy- <i>D</i> -proline	MeHis (3)	3-methyl- <i>L</i> -histidine
4L-HyP	<i>L</i> -4-hydroxyproline	MeLys	<i>Nε</i> -methyl- <i>L</i> -lysine
5HyTrp	5-hydroxy- <i>L</i> -tryptophan	Met	methionine
A	abundance	min	minutes
AAd	α -aminoadipic acid	MMLC	mixed mode liquid chromatography
AcGlu	<i>N</i> -acetyl- <i>DL</i> -glutamic acid	MS	mass spectrometry
AcN	acetonitrile	NLe	norleucine
Ala	alanine	NVa	norvaline
Ala	alanine	Orn	ornithine
AQC	6-aminoquinolyl- <i>n</i> -hydroxysuccinimidyl carbamate	Phe	phenylalanine
Arg	arginine	PhSer	<i>O</i> -phospho- <i>L</i> -serine
Asn	asparagine	Pip	pipecolic acid
Asp	aspartic acid	Pro	proline
Car	carnitine	PTM	post-translational modification
cis 3L-HyP	cis- <i>L</i> -3-hydroxyproline	RPLC	reversed phase liquid chromatography
Cys	cysteine	Ser	serine
D,L-HyL	<i>DL</i> -5-hydroxylysine	SIM	selected ion monitoring
Da	Dalton	SRM	selected reaction monitoring
DiHyPhe	3,4-dihydroxyphenylalanine	Thr	threonine
DiMeArg	<i>N,N</i> -dimethylarginine	TIC	total ion count
DiMeGly	<i>N,N</i> -dimethylglycine	ToF	tim-of-flight
diMeLys	<i>Nε,Nε</i> -dimethyllysine	t_R	retention time
DNA	deoxyribonucleic acid	trans 3L-HyP	<i>trans</i> -3-hydroxy- <i>L</i> -proline
GG	gas chromatography	triMeLys	<i>Nε,Nε,Nε</i> -trimethyllysine
Gln	glutamine	Trp	tryptophan
Glu	glutamic acid	Tyr	tyrosine
Gly	glycine	UV	ultraviolet
Gly	glycine	V	volt
His	histidine	Val	valine
HSe	homoserine	XIC	extracted ion chromatogram
Ile	isoleucine	α,γ -DiAbu	α,γ -diaminobutyric acid
IP	ion pairing	α -ABu	α -aminobutyric acid
IPLC	ion pairing liquid chromatography	α -AHe	α -aminoheptanoic acid
IR	isotope ratio	α AlBu	α -aminoisobutyric acid
IRMS	isotope ratio mass spectrometry	α -AOc	α -aminooctanoic acid
LC	liquid chromatography	α -IBu	α -aminoisobutyric acid
Leu	leucine	β AlBu	β -aminoisobutyric acid
LoD	limit of detection	β -IBu	β -aminoisobutanoic acid
LoQ	limit of quantification		
Lys	lysine		

1 Chapter One – Introduction

The proteome of all known living organism contains 22 proteinogenic AAs. Of these 20 AAs are normally directly coded for by DNA sequences which can be transformed into proteins through the processes of transcription and translation. Interestingly there are over 500 compounds occurring in nature which can be ascribed to the chemical class of AAs^{1,2}, *i.e.* those compounds have both an amino group and a carboxylic group. About 140 of those compounds can be formed as a result of post-translational modifications (PTMs) to proteins³, implying that they are chemically derived by modification of the 20 primary DNA encoded AA residues. PTMs occur as a result of covalent modification following protein assembly. The huge discrepancy between the numbers of genetically encoded AAs and those occurring in the proteome raises question of the origins of PTMs.

It was not until about 50 years ago that deciphering the genetic code began⁴; only several decades later did PTMs come under detailed scientific scrutiny^{5,6}. The major reason for this delay was the generally low abundance of most PTMs, coupled with a lack of standards which until recently made it impossible to detect many of them. The most abundant PTM AAs – 4-hydroxyproline and 5-hydroxylysine (comprising up to 20% of collagen proline and lysine residues⁷) – were discovered about a century ago, but to discover other species it was necessary to significantly lower the limits of detection (LoD) for PTMs.

Of the various PTMs, most attention has probably been paid to phosphorylation because of the discovery of its involvement in numerous regulatory processes including transcriptional regulation⁸, trans-membrane signalling⁹, and cell apoptosis^{10,11}. Phosphorylation alone is proposed to be responsible for over 30% of all PTMs¹². As a result of the phosphorylation process a phosphate group of adenosine triphosphate is transferred to threonine, tyrosine, serine or cysteine residue of the protein. The process is catalysed by a group of enzymes

referred to as protein kinases. The reverse process of dephosphorylation is catalysed by protein phosphatases¹³.

A relatively recent discovery in the field of PTMs has been the modification of histones. Histones are the group of proteins which are responsible for the compact packaging of DNA in the nuclei of eukaryotic cells. There are five histone eukaryotic types: histone 1 (H1), histone 2a (H2a), histone 2b (H2b), histone 3 (H3) and histone 4 (H4). The last four (H2a, H2b, H3, and H4) are termed the inner histones or core histones, forming the core of the nucleosome. They make a complex, which works as a spool around which a stretch of DNA (about 145 bases) is wrapped, thus, allowing compact packaging of the DNA¹⁴ and control of the availability of certain genes for transcription. Histone 1 is associated with the linker part of the DNA, which is the part between the wrapped stretches. Schematically the nucleosome structure is presented in Figure 1-1.

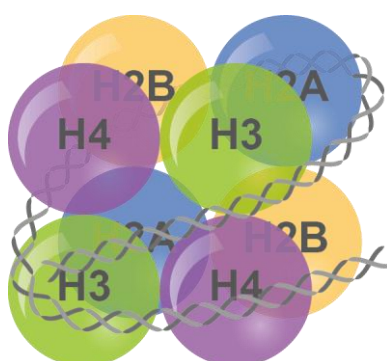


Figure 1-1. Nucleosome structure

H1, H2A, H2B and H4 represent different core histone types.

The accessibility of the DNA fragment stored in such way can be regulated by the modification of the core histone tails. Core histones are known for possessing a wide variety of PTMs. Such modifications can either enable or disable certain parts of the DNA sequence that are transcribed. Table 1-1 provides a summary of the human core histone PTMs that are known to date¹⁵. The predominant histone modifications comprise phosphorylation,

methylation and acetylation. Lysine acetylation is regulated by histone acetyl transferases and histone deacetylases (HDACs). It has been found that an increase in acetylation is associated with an increase in transcription, whereas hypoacetylation results in suppression of gene transcription¹⁶. Phosphorylation of histone residues is hypothesised to assist mitosis and meiosis and along with acetylation it promotes gene transcription¹⁷, as a result of a combination of negative charge on the phosphate groups, and neutralisation of positive charge through acetylation, which together lead to the de-condensation of the chromatin (the complex of double stranded DNA fibre with histone and non-histone proteins)¹⁸. Lysine and arginine methylation are hypothesised to regulate transcriptional processes, though the molecular mechanisms of action are still under investigation¹⁹.

Table 1-1. Summary of human histone modifications

(numbers indicate the modified residue in the histone)

Histone	Length (# residues)	Modificaitons
H2a	130	sym-diMeArg 4 AcLys 6 AcSer 2 PhSer 2 PhThr 121
H2b	126	AcLys 6, 12, 13, 16, 17, 21, 24, 109 MeLys 47, 109 diMeLys 58 PhSer 7, 15, 37
H3	136	asym-diMeArg 3, 18 sym-diMeArg 9 AcLys 5, 10, 15, 19, 24, 28, 37, 57, 80 MeLys 5, 10, 19, 24, 28, 37, 57, 65, 80 diMeLys 10, 28, 37, 57, 80 triMeLys 10, 28, 37, 57, 80, 123 PhThr 4, 7, 12 PhSer 11, 29, 32, 58, 81 PhTyr 108
H4	103	MeArg 4 asym-diMeArg 4 sym-diMeArg 4 MeLys 21 diMeLys 21 triMeLys 21 AcLys 6, 9, 13, 17, 32, 60, 92 AcSer 2 PhSer 2, 48 PhTyr 52, 89

Another important PTM is ubiquitination. Ubiquitin is a small protein comprising 76 AA residues. Ubiquitin targets either proteins which are programmed for degradation or proteins which degradation has been induced through phosphorylation²⁰. Ubiquitin also targets proteins which have been damaged through oxidation or mutation²¹. Ubiquitination occurs in a three step process. First, the C-terminus of ubiquitin forms a thioester with the ubiquitin activating enzyme (E1), then, the ubiquitin migrates to the thiol group of the ubiquitin conjugating enzyme (E2), and, finally, the ubiquitin complex with E2 binds to the ubiquitin ligase complex with the target protein. The target protein which has been modified by ubiquitin in such a manner is then forwarded to the protease for degradation²² (Figure 1-2).

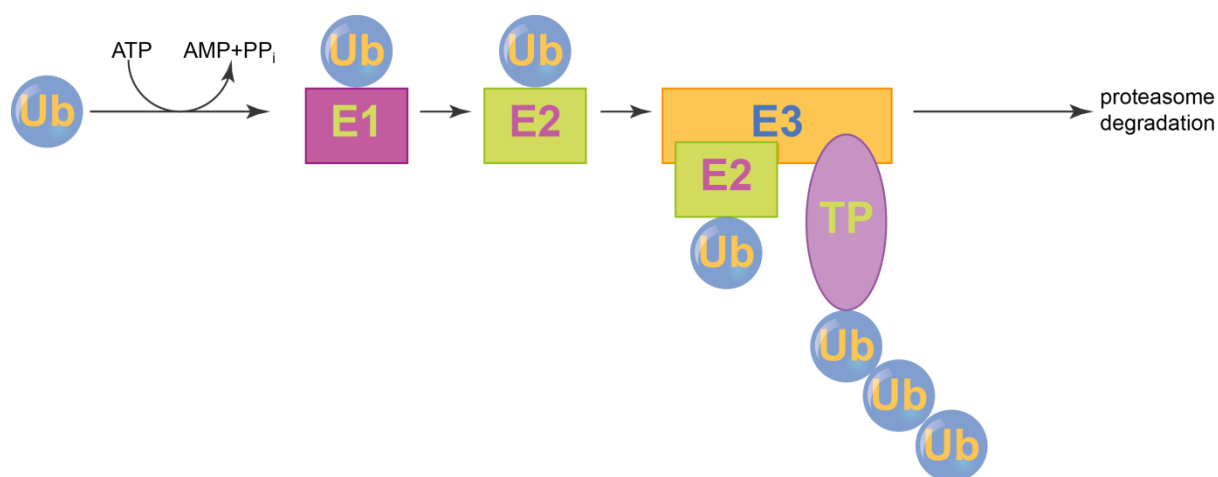


Figure 1-2. Schematic presentation of ubiquitin-proteasome system

Ub – ubiquitin; E1 – activating enzyme; E2 – conjugating enzyme; E3 – ubiquitin ligase; TP – target protein

ATP – adenosine triphosphate; AMP – adenosine monophosphate; PP_i – pyrophosphate.

1.1 2OG- and Fe(II)- dependent Oxygenases: Overview

The 2-oxoglutarate (2OG) and Fe(II) dependent oxygenases enzyme super-family is responsible for a broad spectrum of biochemical processes²³. 2OG oxygenases employ non-heme Fe(II) as a co-factor and 2OG as a co-substrate. The general scheme of the oxidation reaction catalysed by the 2OG- and Fe(II)- dependent oxygenases is presented in Figure 1-3. The Fe(II) ion is coordinated by a highly conserved triad of histidine (taking the n position in

the enzyme), aspartate or glutamate (taking the position n+2 in the enzyme), and another histidine (taking the position n+x in the enzyme). The remaining three positions of the hexacoordinated are occupied by the 2OG through the 1-carboxy group in the bi-dentate manner and dioxygen. In the catalytic cycle the 2OG is oxidised to CO₂ and succinate. The other atom of dioxygen is predominantly incorporated into the hydroxyl-group of the product²⁴. Ascorbic acid has been reported to stimulate the activity of some 2OG oxygenases both *in vitro* and *in vivo* tests. However, its precise molecular role has not been revealed to date²⁵.

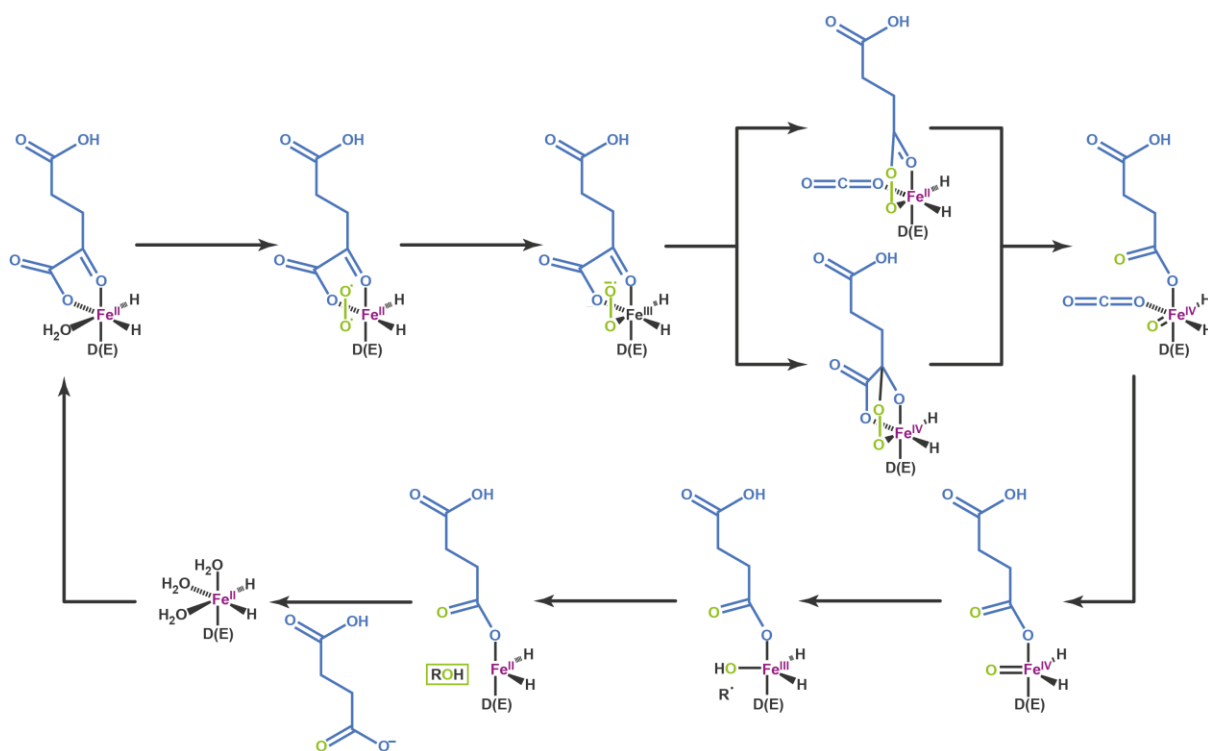


Figure 1-3. General scheme of the oxidation catalysed by 2OG- and Fe(II)- dependent oxygenases

The Fe(II) ion is coordinated by the highly conserved triad of histidine (H; taking the n position in the enzyme), aspartate or glutamate (D(E); taking the position n+2 in the enzyme), and another histidine (H; taking the position n+x in the enzyme).

1.1.1 Oxygen Sensing: Maintaining Homeostasis

The heterodimeric hypoxia inducible transcription factor (HIF) plays a central role in the chronic regulation of oxygen homeostasis in animals. In humans the levels and activity of the

HIF-1 α isoforms are controlled by prolyl- (P402 and P564) and asparaginyl- (N803) hydroxylation of the HIF- α subunit as catalysed by three prolyl-4-hydroxylase domain enzymes (PHDs) and FIH (factor inhibiting hypoxia inducible factor), respectively (Figure 1-4) – both of which belong to the superfamily of 2OG dependent oxygenases²⁶⁻³⁰.

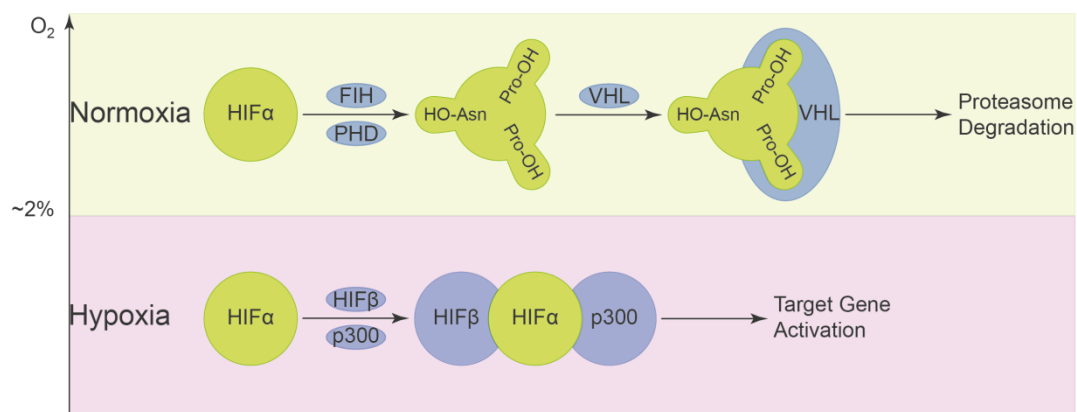


Figure 1-4. HIF-1 α normoxia degradation and hypoxic response

HIF α = hypoxia inducible transcription factor; PHD = prolyl-4-hydroxylase domain enzyme; VHL = von Hippel Lindau protein; FIH = factor inhibiting hypoxia inducible factor.

Under normoxic conditions two processes lead to the reduction of HIF α activity: prolyl hydroxylation promotes binding of HIF α subunit to the von Hippel Lindau (VHL) protein, which results in HIF α ubiquitination and subsequent degradation; HIF α asparaginyl hydroxylation blocks binding of HIF α to transcription coactivators and so suppresses HIF-mediated transcription. Under hypoxic conditions PHD and FIH activity is reduced and levels of active HIF α are correspondingly increased, leading to a co-ordinated response to hypoxia, including upregulation of angiogenesis and erythropoiesis and a shift from oxidative metabolism to glycolysis²⁶⁻³⁰ (Figure 1-5).

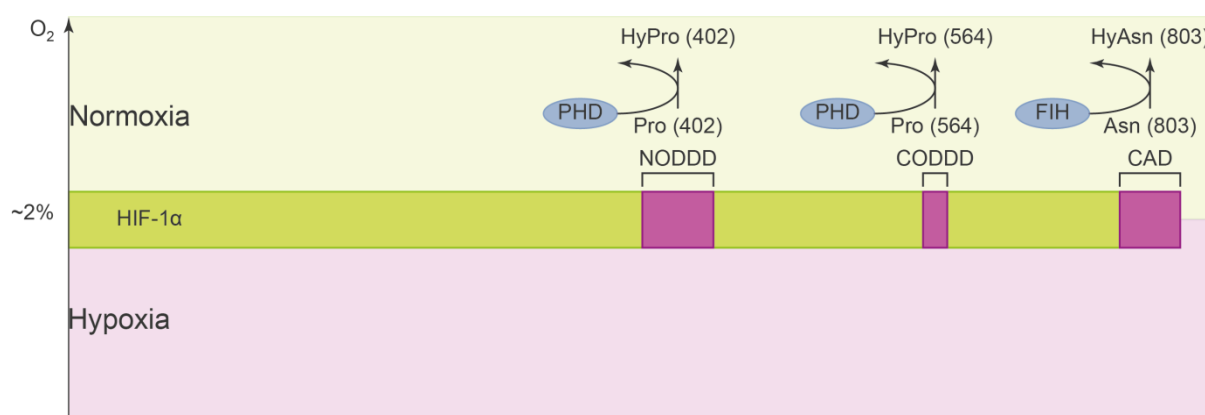


Figure 1-5. HIF-1 α hydroxylation under normoxic conditions

HIF α = hypoxia inducible transcription factor; PHD = prolyl-4-hydroxylase domain enzyme; FIH = factor inhibiting hypoxia inducible factor; NODDD = amino-terminal oxygen-dependent degradation domain; CODDD = carboxy-terminal oxygen-dependent degradation domain; CAD = carboxy-terminal activation domain.

Cellular hypoxia is observed in disease states where oxygen delivery systems are compromised including cancer^{31,32}, metabolic disorders³³ and myocardial infarction³⁴. As part of the long term regulation of the body's oxygen homeostasis the up-regulation of HIF α in these cases enables the activation of the genes which compensate for hypoxic effects. The inhibition of PHDs and FIH is of interest to enable artificially triggering of the HIF-regulated hypoxic response³⁵ for therapeutic benefits for anaemia and ischemia related disorders including diabetes³⁶, stroke³⁷ and myocardial infarction³⁸.

1.1.2 Collagen Post-Translational Modifications: Maintaining Structural Integrity

Collagens are the family of extracellular matrix proteins that maintain the integrity of the tissues of an organism³⁹. Collagens account for about one third of the organism's total protein by weight. The structure of protein fibrils is enabled through hydrogen bonds between the hydroxyl groups of 3- and 4- hydroxyprolines and 5-hydroxylysine residues of collagen. Prolyl- and lysyl-hydroxylases which catalyse modification of collagen residues belong to the family of 2OG oxygenases, found in the endoplasmic reticulum⁴⁰ (Figure 1-6).

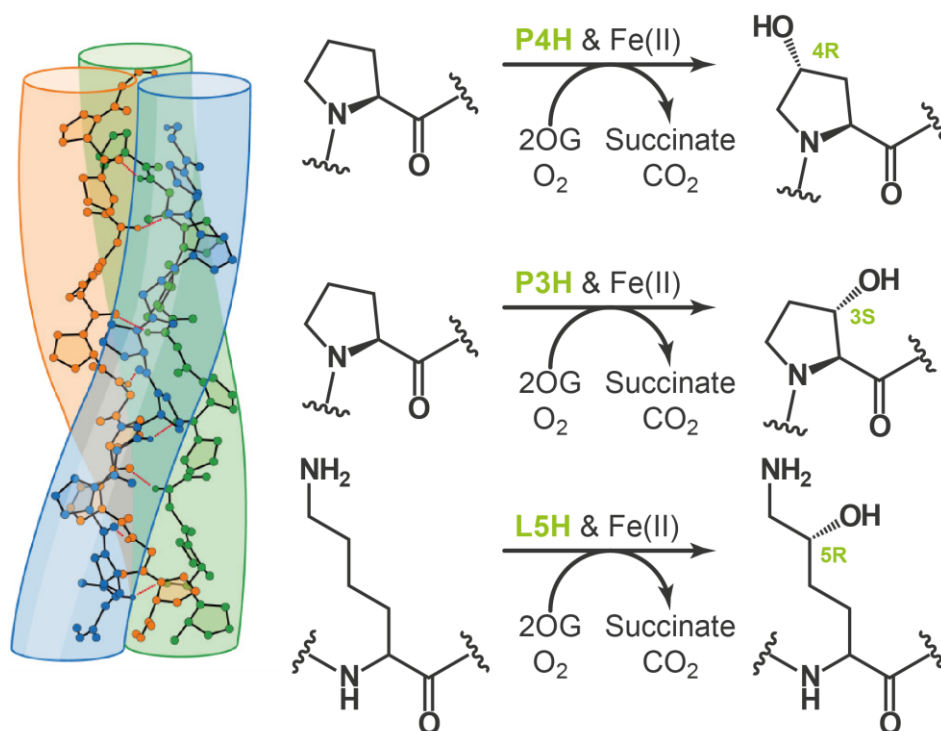


Figure 1-6. Collagen structure and reaction schemes of collagen prolyl- and lysyl hydroxylation

Collagen structure is adapted from Protein Structure and Function by Gregory A Petsko and Dagmar Ringe.

The most abundant of the collagen PTMs is 4-prolyl hydroxylation as occurs in the –X-Pro-Gly- fragment of the collagen polypeptide chain. This modification is also referred to as co-translational because the modification occurs in the nascent protein chain. This modification is responsible for the conformation stability of the triple helix of collagen, which is enabled through the formation of intra-molecular hydrogen bonds⁴¹.

3-Hydroxyproline is a less common modification occurring in collagen. It normally appears in the fragment -3-HyPro-4-HyPro-Gly-. It was shown that the presence of this residue destabilises the triple helix structure of collagen in fibrils, but it was hypothesised that this enables the stabilisation of the collagen structure in basement membranes, where much more interaction occurs with other biological molecules⁴².

3-Hydroxylysine enables the cross linking of the collagen fibrils. Analogously to the prolyl hydroxylase, the collagen lysyl hydroxylases target a triad of –X-Lys-Gly- of a nascent

collagen chain. Hydroxylated lysyl residues can undergo further modification – glycosylation, which was assumed to be responsible for the regulation of the collagen fibril diameter⁴³.

1.1.3 Antibiotic Biosynthesis

2OG oxygenases have also been shown to be responsible for very sophisticated biosynthetic processes such as biosynthesis of penicillins and cephalosporins. Isopenicillin N synthase (IPNS) enables the formation of the bicyclic core of the penicillins in one step. Deacetoxycephalosporin C synthase (DAOCS) catalyses the expansion of the penicillin nucleus to create the cephalosporin nucleus. Both IPNS and DAOCS belong to the family of 2OG oxygenases^{44,45} (Figure 1-7). Clavaminc acid synthase is another example of 2OG oxygenase responsible sophisticated biosynthetic pathway carried out through oxidative cyclisation and desaturation⁴⁶.

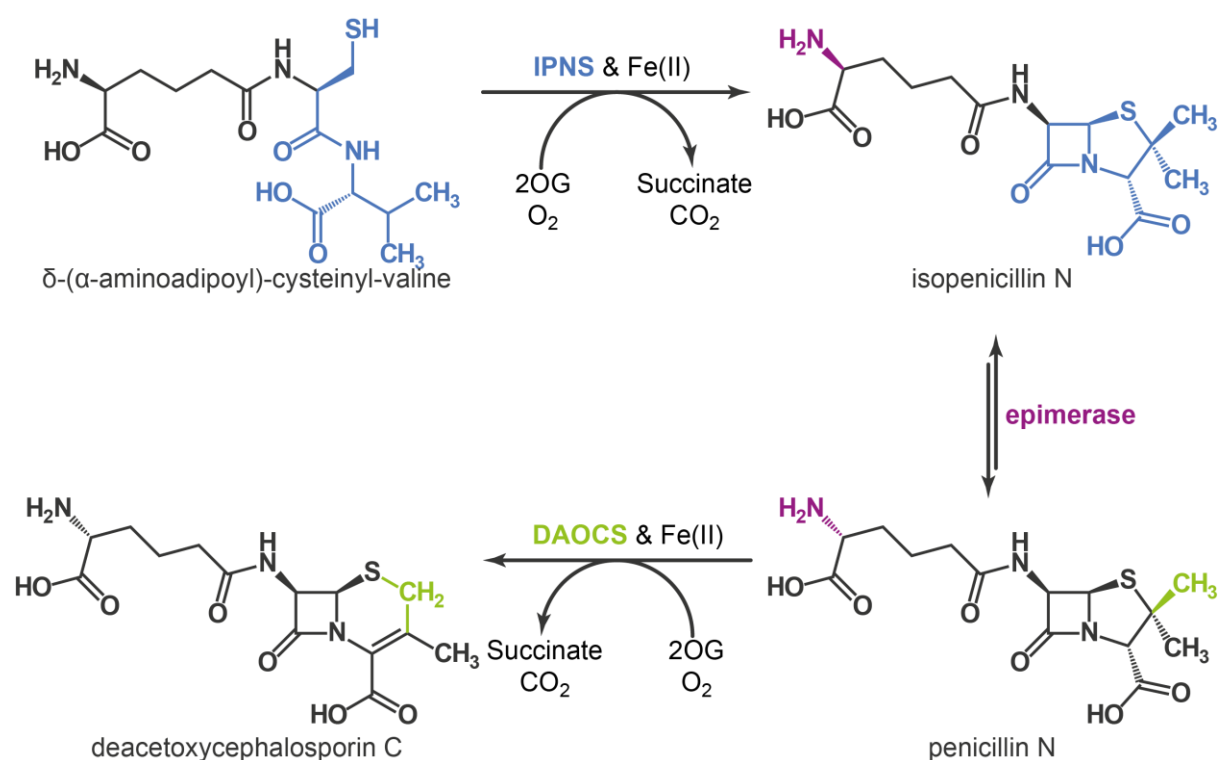


Figure 1-7. Scheme biosynthesis of penicillin and cephalosporin⁴⁷

IPNS = isopenicillin N synthase; DAOCS = deacetoxycephalosporin C synthase.

1.1.4 Fatty Acid Metabolism

Carnitine (see Figure 1-8) has long been established as playing a central role in the metabolism of fatty acids through promoting the oxidation of fatty acids in mitochondria. Fatty acids are delivered to mitochondria in the form of their O-acyl carnitine esters⁴⁸.

Carnitine is obtained from the diet of animals, but can also be produced *de novo* in the organism from N^ε-trimethyllysine (Figure 1-8). In the first stage of carnitine production a lysine residue is trimethylated using methionine as the source of methyl groups. It is proposed that lysine methylated within proteins such as histones, myosin and actin. N^ε-trimethyllysine is liberated after protein degradation and so acts as a carnitine precursor^{49,50}.

In the next step trimethylated lysine is hydroxylated to 3-hydroxy-6-N-trimethyllysine by N^ε-trimethyllysine hydroxylase (TMLH). The product is then cleaved to butyrobetaine aldehyde and glycine⁵¹. The butyrobetaine aldehyde is then oxidised to butyrobetaine, which is hydroxylated by the γ -butyrobetaine hydroxylase (BBOX) to carnitine⁵². Both BBOX and TMLH are members of the 2OG oxygenases family.

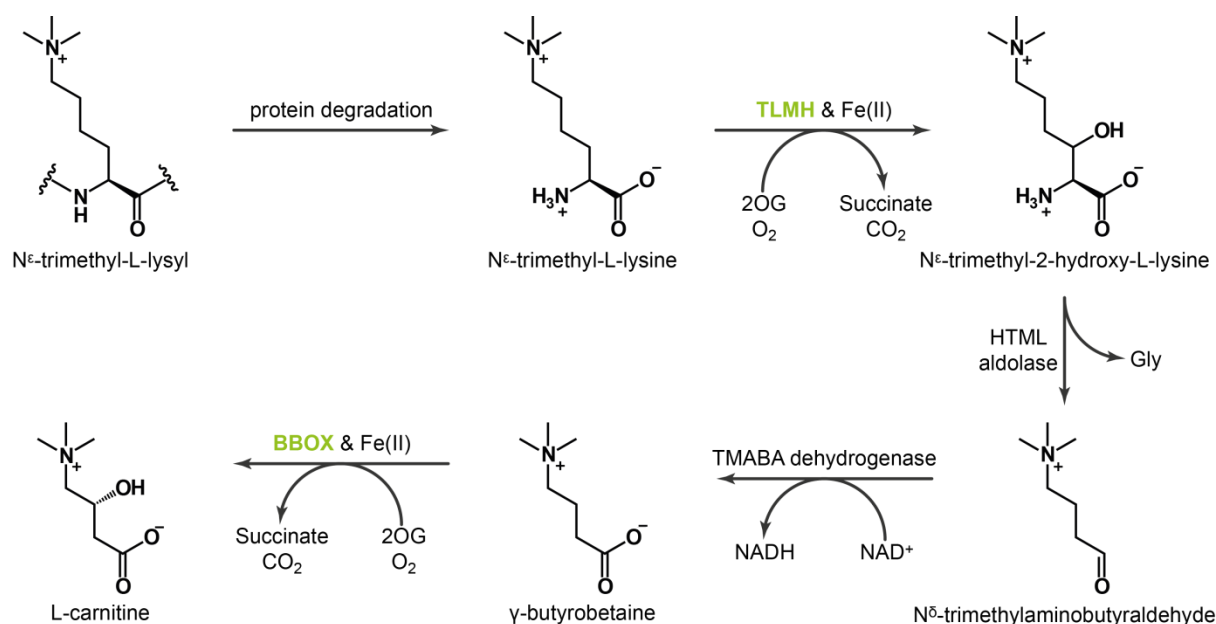


Figure 1-8. Carnitine biosynthesis pathway

TMLH = N ϵ -trimethyllysine hydroxylase; BBOX = γ -butyrobetaine hydroxylase.

1.1.5 Histone Modifications

The modifications of proteinogenic AA residues in proteins can enable extra functionality of proteins and regulate their biochemical activity. Regulation of protein function can be performed in a reversible manner through modifications such as phosphorylation, acetylation and methylation. In such cases the modification acts as a chemical switch and enables or disables protein function.

Methylation plays a pivotal role in gene transcription regulation. Both arginyl and N ϵ -lysyl methylation play an important role in transcription. Methylated lysyl residues both regulate transcription and can facilitate the repair of DNA¹⁷. Lysyl methylation at specific histone residues can either activate or suppress transcription^{53,54}.

The process of lysyl- and arginyl- methylation is regulated by the broad spectrum of methyltransferases. Demethylation of lysyl residues can be regulated by the family of 2OG oxygenases – JmjC histone demethylases^{55,56}. It still remains unclear what are the precise

roles of different methylated forms of both lysine and arginine. Currently research into the function of the 2OG oxygenases in DNA transcription regulation is ongoing.

Given the importance of regulating oxygen homeostasis, enabling structural stability and the regulation of a broad variety of chemical signalling processes, the role of post-translational hydroxylation as catalysed by members of by 2OG- and Fe(II)- dependent oxygenases, could potentially be comparable to that of post-translational phosphorylation. The work in this thesis focuses on modifications catalysed by 2OG oxygenases including lysine and proline hydroxylation, γ -butyrobetaine hydroxylase (hBBOX) catalysed hydroxylation and histone post-translational methylation-demethylation.

1.2 Assays for the 2OG Oxygenases

The medicinal importance of the 2OG oxygenases has prompted further research into their biological function and mechanisms. The possibilities to control – enhance or inhibit – biological processes, regulated by the 2OG oxygenases enzyme family, could enable significant new therapies. Therefore, it is desirable to develop robust and sensitive methods, to monitor the enzymatic activity of 2OG catalysed processes in a broad spectrum of biological matrices. The methods generally used for the 2OG oxygenase quantitative assays have been reviewed³⁵ as summarised in Table 1-2.

Table 1-2. Summary of the assays of 2OG oxygenase activity³⁵

Analyte	Analytical Technique
2OG	spectrophotometry
succinate	spectrophotometry, LC-MS, radioisotope measurement
CO ₂	¹⁴ C radioisotope measurement
proteins and peptides	LC-MS; MALDI-MS, NMR
³ H	³ H radioisotope measurement
¹⁸ O	¹⁸ O radioisotope measurement

2OG, succinate, CO₂ (Table 1-2) are the co-substrate and the products of oxidative reaction (Figure 1-3), respectively, their concentrations can be, but not always, stoichiometrically

related to those of the hydroxylated peptides. Thus, monitoring the amount of the succinate or CO₂ produced, or the decrease in the initial concentration of the 2OG, can enable indirect quantification of the protein hydroxylation.

These methods can be useful for *in vitro* experiments, in which the amount of all the analytes can be easily controlled. On the other hand, these methods cannot be used in cell and tissue based experiments or *in vivo* experiments. 2OG, succinate and CO₂ are not specific for 2OG oxygenases catalysed transformations, these compounds are also involved in a wide range of metabolic processes such as the citric acid cycle⁵⁷ – the pathway which allows all aerobic organisms to generate energy. The often high background consumption and production of those metabolites makes it impossible to identify the sources of non-specific analytes in *in vivo* experiments, thus making it difficult to use these methods to assay for 2OG oxygenase activity.

Measurement of CO₂ presents certain difficulties even for *in vitro* experiments associated with the necessity to scale up experiments to achieve sufficient CO₂ quantities for analysis and the safety regulations imposed on radioactive experiments, which often make this approach prohibitive⁵⁸. A method for ¹⁴CO₂ analysis for cells has been reported⁵⁹; CO₂, however, is not a specific product of a particular enzyme but can result from a wide range of metabolic processes, making it less specific and hence reliable as an indicator of a particular enzyme activity.

Therefore, in cell based and *in vivo* experiments, a preferable approach would be direct monitoring changes in substrate and product concentrations. Few approaches offer this capability at the LoD required, with quantitative MS based proteomics being one of the few possible exceptions.

1.3 Analysis of Post-Translational Modifications by Proteomics

Proteomics uses a wide spectrum of analytical methods and techniques, which enable identification of protein sequence, identification and localisation of protein PTMs, and study of protein-protein interactions. Generic proteomic experiments include (1) the fractionation of the sample in order to extract target proteins, (2) enzymatic digestion (often using trypsin) of the target protein to create a mixture of peptides, (3) LC separation of the digest mixture, (4) MS or MS-MS analysis of the peptides and (5) protein identification (often using protein libraries)⁶⁰.

Fractionation of the crude proteomic sample can be carried out using gel separation, chromatography, filtration or centrifugation. Enzymatic digestion involves cleavage of the protein into smaller peptides using enzymes with different selectivity such as trypsin (cleaves at the C-terminal site of Lys and Arg residues) or chymotrypsin (cleaves after Phe, Trp, and Tyr residues). For LC separation reversed phase chromatography is normally used.

Two strategies are often used in proteomics: ‘top-down’ and ‘bottom-up’ – the difference being whether enzymatic digestion of target proteins is involved or not (Figure 1-9). In the ‘top-down’ approach, intact proteins are fragmented within the mass spectrometer and the MS spectrum is analysed^{61,62}. Alternatively, in the ‘bottom-up’ approach, the protein is enzymatically cleaved into smaller peptides, and then the mixture of peptides is analysed. The ‘top-down’ approach is more rarely used.

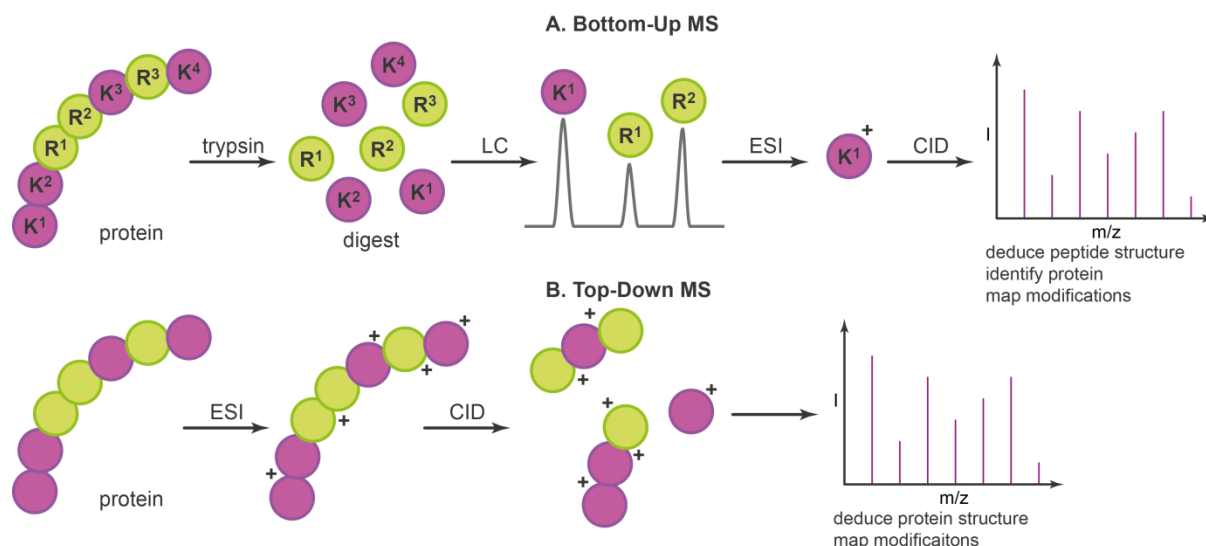


Figure 1-9. Comparison of ‘top-down’ and ‘bottom-up’ MS approach

LC = LC; ESI = electrospray ionisation; CID = collision induced dissociation.

K stands for the fragments with carboxy-terminal lysine; R stands for the fragments with carboxy-terminal arginine.

Proteomics provides various methods which enable identification and quantification of PTMs. Most of such methods allow relative quantification, *i.e.* the quantities produced are related to a reference sample. Relative quantification allows comparison of samples exposed to different conditions. In order to minimise error the samples exposed to different experimental conditions are mixed together and analysed as one mixture. In order to differentiate the sample from the reference, mass tags can be introduced. Several approaches have been developed in order to introduce mass tags including chemical tagging and metabolic tagging⁶³⁻⁶⁵. Chemical tags are normally introduced after the crude sample purification process, whereas the metabolic tags are introduced in the process of the cell or tissue growth^{64,65}.

1.3.1 Stable Isotope Labelling in Cell Culture

Stable isotope labelling in cell culture (SILAC) is a metabolic labelling method used for relative quantification in proteomic analysis. For SILAC experiments an isotope labelled AA, or number of AAs, is introduced via the growth media to create cells containing specific

isotopically labelled proteins. As a result the whole proteome is labelled with a selected AA. Initially the method was developed for introduction of enriched leucine but was later expanded to the use of other labelled AAs as well⁶⁶.

The analysis of a particular protein is preceded by extraction and purification procedures. The clone of the enzyme for the digest defines the choice of labels introduced. In ‘bottom-up’ analysis the protein of interest is often digested with trypsin. Since trypsin cleaves the protein at the C-terminal arginyl and lysyl residues, arginine and lysine are the AAs often used for the SILAC labelling experiment. Use of these two AAs usefully provides an isotopic marker in each of the peptides resulting from the tryptic digest⁶⁷.

Stable isotope labelling is used for comparative analysis, in which pools of cells exposed to different experimental conditions carry different isotope labels. When such pools are combined it is possible to get distinct signals for identical species derived from different pool within a single sample. Relating the signal intensities for species derived from different pools allows comparison of the conditions which these pools were exposed to. Normally up to 5 samples depending on the types of the labels introduced can be compared.

1.3.2 Isobaric Tags for Relative and Absolute Quantification

Isobaric tags for relative and absolute quantification (iTRAQ) is an elegant method relying on the chemical modification of peptides⁶⁸. The reagent, which as a result of fragmentation in the mass spectrometer, first generates the isobaric fragment which in turn produces specific fragment ions, which are different depending on the isotopes used in the tag (Figure 1-10). As in SILAC it becomes possible to combine the samples with different labels and quantify those within one experiment.

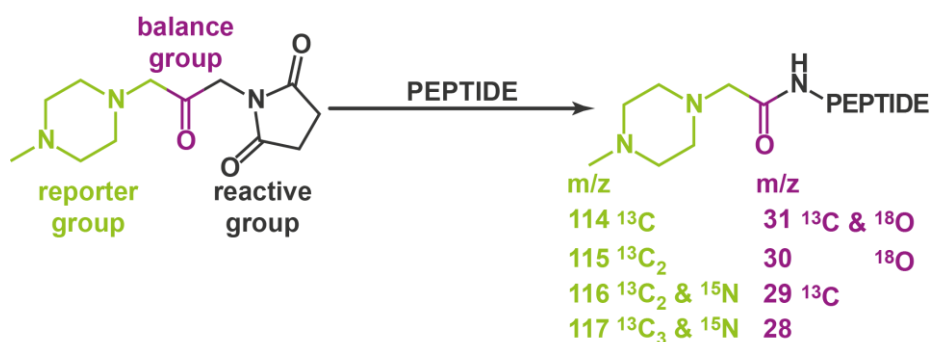


Figure 1-10. Diagram of isobaric tagging⁶⁸

Even though the iTRAQ method relies on the ions derived from the tag, rather than from the peptide, it is important to avoid co-elution of the isobaric peptides. The criticism of the method includes high cost of patented isobaric tags and poor sensitivity⁶⁹.

1.3.3 Absolute Quantification Using an Internal Standard

Absolute quantification using an internal standard (termed as AQUA) relies on addition of an isotopically labelled peptide marker to the digest of the target protein. This method requires *a priori* information about the target peptide AA sequence and availability of synthetic labelled peptide⁶⁶. This makes the analysis of the PTMs extremely complicated, due to the fact that it is more expensive to incorporate a modified AA into the peptide, and for some modifications the standards are simply not available.

1.3.4 Identification and Quantification of Post-Translational Modifications Using Proteomic Techniques

The identification of PTMs still presents a major analytical challenge with the quantity required for the characterisation of the modified protein presenting a particular issue. A particular PTM is not necessarily present in every instance of the protein, in fact it can often be abundant at only a few per cent of the protein and variable depending on cellular conditions. For example, an *in vitro* quantitative study of the ankyrin repeat domain hydroxylation by FIH showed the variation in hydroxylation level from 20% (under

normoxia) to below 1% (under hypoxic conditions)⁷⁰. In addition one particular protein can have different modifications, and various combinations of such modifications. Therefore, the target modification is often present at levels below the LoD⁷¹.

The number of modifications associated with one particular protein can also complicate the identification of the chromatographic peak corresponding to the modified species. Protein modification changes the hydrophobicity of the protein, and as a result the retention time of the chromatographic peak corresponding to a particular peptide.

In order to simplify the identification of the proteins and peptides, library matching can be used. Library matching can be carried out using annotated peptide libraries, *i.e.* collections of MS-MS fragmentation spectra with assigned peptide sequences. The identification is carried out on the basis of the matching algorithm; the level of confidence is assigned on the basis of the similarities of the measured and database spectra. To test the similarities a mathematical algorithm such as vector comparison in multidimensional space or comparison of the odds ratio is applied^{72,73}. The major disadvantage of matching the sample with the data from the library is that *a priori* information about expected PTMs is required. This means that the PTMs that have not been explicitly searched for will be overlooked and providing an exhaustive list of modifications for every protein is not necessarily practical. In addition to that, in the tandem MS technique, not only one peak is affected because of introduction of modification, but all the fragment-peaks as well⁷¹. As a result, a system like this cannot be resolved, because there are more unknown parameters, than given data. This situation benefits from prior exploratory analysis which provides information on the modifications present, thus, narrowing down the search.

Another important problem associated with the ‘bottom-up’ approach is that MS analysis is normally preceded by enzymatic digestion of analyte proteins creating smaller peptides. Unfortunately, enzymatic digests yield different quantities for different peptides. Some

peptides might be less abundant and happen to be below the LoD. As a result in the mixture of peptides only a small fraction can be identified and an even smaller number of peptides can be quantified. In a recent review⁶⁵ it was suggested that for single cellular organisms about 50% of proteins can typically be identified and for higher organisms only about 10% can be identified, even less than that can be quantified; for quantification much higher concentrations are required. The sensitivity of the MS techniques is another important issue. It is recommended to use at least 1 µg or several picomoles of purified protein⁷¹. Sensitivity at subattomolar level has been claimed in a recent review¹². Another review does not report a suggested LoD⁶³, but does cite a paper reporting the detection of α -casein fragments after tryptic digest of total quantity at 500 fmol⁷⁴.

A publication on histone quantification provides data from which the LoD could be estimated at the low micromolar (µmol/L) level⁷⁵. The authors use LC coupled to a MS ESI-ToF MS detector operated in single reaction mode to quantify ubiquitin in cells and tissues.

Both experimental data are in agreement with my laboratory experience, which shows that LoDs can be extremely variable and as such it is not realistic to quote numbers without specifying the instruments and matrix system used. In general, although, sub-micromolar levels of detection can be realised, these are usually for high purity standards and are rarely achievable for ‘real’ samples extracted from biological media.

As an example of the quantitative analysis problem associated with PTMs is exemplified when considering hypoxia inducible factor α -domain (HIF-1 α) undergoing PTM as a result of 2OG oxygenases. HIF-1 α is made up of 826 AA residues. HIF-1 α can potentially have two prolyl residues and one asparaginyl residue which can be hydroxylated. Hydroxylation can occur in as few as 1% of residues *in vivo*. The amount of total protein available from the *in vitro* assays for the analysis is about 1 mg. This means that less than 0.1 nmol of protein may be available for analysis. The spectrum of potential sample matrices encompasses *in vitro*

assays, cell based experiments, and *in vivo* experiments. It should also be noted that significant losses can result from multiple sample purification steps prior to LC-MS analysis.

1.4 Aims of the Work Described in this Thesis

The work described in this thesis concerns the development and application of analytical methods for the identification and quantification of PTMs. The aim of the research was to develop a range of methods which can provide information complimentary to that of proteomics to enable (1) the identification of stereo- and regio- isomeric PTMs, (2) the identification of the novel PTMs, thus, providing starting information for proteomics, and (3) the absolute quantification of low abundance PTMs.

In order to address the qualitative analysis of PTMs AA analysis was investigated. AA analysis allows direct detection of the PTMs and does not require *a priori* information about the modifications present. AA analysis presents two major challenges. First, the similarity of the physical and chemical properties of AAs and their modifications makes the separation and distinguishing of AAs from one another challenging. Second, modified AAs are present in quantities hundreds of times lower than those of non-modified proteinogenic AAs.

It was therefore, necessary to develop analytical methods to enable (1) the separation of compounds with close physical and chemical properties (selectivity), (2) the detection of the compounds at the lowest concentrations possible (sensitivity).

- a. In order to ensure the selectivity, LC coupled to MS techniques were developed and optimised. The compatibility of LC-MS methods with the biological matrices was tested.
- b. In order to evaluate the sensitivity of the methods developed they were rigorously tested using appropriate standards and relevant biological samples. A novel

quantification protocol using isotope ratio MS (IRMS) was developed and its application was tested for the quantification of histone PTMs.

- c. Application of the methods developed was carried out in collaboration with Christopher Schofield's Group (Department of Chemistry, University of Oxford), who focus their research on the functions of human 2OG oxygenases. A number of studies utilising the analytical methods developed were carried out as part of the work for this thesis and have subsequently been published or are in the preparation for publication; these studies demonstrated the applicability of the methods for the analysis of PTMs as catalysed by 2OG oxygenases.

1.5 References

- (1) Akram, M.; Asif, H. M.; Uzair, M.; Akhtar, N.; Madni, A.; Shah, S. M. A.; ul Hasan, Z.; Ullah, A. *Journal of Medicinal Plants Research* **2011**, 5, 3997.
- (2) Vranova, V.; Rejsek, K.; Skene, K. R.; Formanek, P. *Plant and Soil* **2011**, 342, 31.
- (3) Uy, R.; Wold, F. *Science* **1977**, 198, 890.
- (4) Chapeville, F.; Ehrenstein, G. V.; Benzer, S.; Weisblum, B.; Ray, W. J.; Lipmann, F. *Proceedings of the National Academy of Sciences of the United States of America* **1962**, 48, 1086.
- (5) Soffer, R. L. *Molecular and Cellular Biochemistry* **1973**, 2, 3.
- (6) Udenfrie, S. *Science* **1966**, 152, 1335.
- (7) Gustavson, K. H. *Nature* **1955**, 175, 70.
- (8) Hunter, T.; Karin, M. *Cell* **1992**, 70, 375.
- (9) Sibley, D. R.; Benovic, J. L.; Caron, M. G.; Lefkowitz, R. J. *Cell* **1987**, 48, 913.
- (10) Cheung, W. L.; Ajiro, K.; Samejima, K.; Kloc, M.; Cheung, P.; Mizzen, C. A.; Beeser, A.; Etkin, L. D.; Chernoff, J.; Earnshaw, W. C.; Allis, C. D. *Cell* **2003**, 113, 507.
- (11) Yousefi, S.; Green, D. R.; Blaser, K.; Simon, H. U. *Proceedings of the National Academy of Sciences of the United States of America* **1994**, 91, 10868.
- (12) Kamath, K. S.; Vasavada, M. S.; Srivastava, S. *Journal of Proteomics* **2011**, 75, 127.
- (13) Krebs, E. G.; Beavo, J. A. *Annual Review of Biochemistry* **1979**, 48, 923.
- (14) Shechter, D.; Dormann, H. L.; Allis, C. D.; Hake, S. B. *Nature Protocols* **2007**, 2, 1445.
- (15) 2002-2012.
- (16) De Ruijter, A. J. M.; Van Gennip, A. H.; Caron, H. N.; Kemp, S.; Van Kuilenburg, A. B. P. *Biochemical Journal* **2003**, 370, 737.
- (17) Strahl, B. D.; Allis, C. D. *Nature* **2000**, 403, 41.
- (18) Shogren-Knaak, M.; Ishii, H.; Sun, J. M.; Pazin, M. J.; Davie, J. R.; Peterson, C. L. *Science* **2006**, 311, 844.

-
- (19) Greer, E. L.; Shi, Y. *Nat Rev Genet* **2012**, 13, 343.
 - (20) Brooks, C. L.; Gu, W. *Current Opinion in Cell Biology* **2003**, 15, 164.
 - (21) Wilkinson, K. D. *Seminars in Cell & Developmental Biology* **2000**, 11, 141.
 - (22) Welchman, R. L.; Gordon, C.; Mayer, R. J. *Nature Reviews Molecular Cell Biology* **2005**, 6, 599.
 - (23) Loenarz, C.; Schofield, C. J. *Nature Chemical Biology* **2008**, 4, 152.
 - (24) Clifton, I. J.; McDonough, M. A.; Ehrismann, D.; Kershaw, N. J.; Granatino, N.; Schofield, C. J. *Journal of Inorganic Biochemistry* **2006**, 100, 644.
 - (25) Flashman, E.; Davies, S. L.; Yeoh, K. K.; Schofield, C. J. *Biochemical Journal* **2010**, 427, 135.
 - (26) Schofield, C. J.; Ratcliffe, P. J. *Biochemical and Biophysical Research Communications* **2005**, 338, 617.
 - (27) Kaelin, W. G.; Ratcliffe, P. J. *Molecular Cell* **2008**, 30, 393.
 - (28) Semenza, G. L. *Cell* **2001**, 107, 1.
 - (29) Hirota, K.; Semenza, G. L. *Biochemical and Biophysical Research Communications* **2005**, 338, 610.
 - (30) Schofield, C. J.; Ratcliffe, P. J. *Nature Reviews Molecular Cell Biology* **2004**, 5, 343.
 - (31) Brown, J. M.; William, W. R. *Nature Reviews Cancer* **2004**, 4, 437.
 - (32) Pouyssegur, J.; Dayan, F.; Mazure, N. M. *Nature* **2006**, 441, 437.
 - (33) Bordoni, A.; Angeloni, C.; Leoncini, E.; Danesi, F.; Maranesi, M.; Biagi, P. L.; Hrelia, S. *Nutrition Metabolism and Cardiovascular Diseases* **2005**, 15, 166.
 - (34) Kido, M.; Du, L. L.; Sullivan, C. C.; Li, X. D.; Deutsch, R.; Jamieson, S. W.; Thistlethwaite, P. A. *Journal of the American College of Cardiology* **2005**, 46, 2116.
 - (35) Rose, N. R.; McDonough, M. A.; King, O. N. F.; Kawamura, A.; Schofield, C. J. *Chemical Society Reviews* **2011**, 40, 4364.
 - (36) Yamada, N.; Horikawa, Y.; Oda, N.; Iizuka, K.; Shihara, N.; Kishi, S.; Takeda, J. *Journal of Clinical Endocrinology & Metabolism* **2005**, 90, 5841.
 - (37) Ogle, M. E.; Gu, X. H.; Espinera, A. R.; Wei, L. *Neurobiology of Disease* **2012**, 45, 733.
 - (38) Parisi, Q.; Biondi-Zoccai, G. G. L.; Abbate, A.; Santini, D.; Vasaturo, F.; Scarpa, S.; Bussani, R.; Leone, A. M.; Petrolini, A.; Silvestri, F.; Biasucci, L. M.; Baldi, A. *International Journal of Cardiology* **2005**, 99, 337.
 - (39) Vanderrest, M.; Garrone, R. *Faseb Journal* **1991**, 5, 2814.
 - (40) Krane, S. M. *AAAs* **2008**, 35, 703.
 - (41) Myllyharju, J. *Matrix Biology* **2003**, 22, 15.
 - (42) Jenkins, C. L.; Bretscher, L. E.; Guzei, I. A.; Raines, R. T. *Journal of the American Chemical Society* **2003**, 125, 6422.
 - (43) Yamauchi, M.; Uzawa, K.; Deterding, L. J.; Tomer, K.; Verdelis, K.; Chandler, G. S.; Katz, E. P. *Journal of Dental Research* **1998**, 77, 643.
 - (44) Baldwin, J. E.; Abraham, E. *Natural Product Reports* **1988**, 5, 129.
 - (45) Schofield, C. J.; Baldwin, J. E.; Byford, M. F.; Clifton, I.; Hajdu, J.; Hensgens, C.; Roach, P. *Current Opinion in Structural Biology* **1997**, 7, 857.
 - (46) Baggaley, K. H.; Brown, A. G.; Schofield, C. J. *Natural Product Reports* **1997**, 14, 309.
 - (47) Valegard, K.; van Scheltinga, A. C. T.; Lloyd, M. D.; Hara, T.; Ramaswamy, S.; Perrakis, A.; Thompson, A.; Lee, H.-J.; Baldwin, J. E.; Schofield, C. J.; Hajdu, J.; Andersson, I. *Nature* **1998**, 394, 805.
 - (48) Bremer, J. *Journal of Biological Chemistry* **1962**, 237, 3628.
 - (49) Bremer, J. *Physiological Reviews* **1983**, 63, 1420.
-

- (50) Rebouche, C. J.; Seim, H. *Annual Review of Nutrition* **1998**, 18, 39.
- (51) Vaz, F. M.; Ofman, R.; Westinga, K.; Back, J. W.; Wanders, R. J. A. *Journal of Biological Chemistry* **2001**, 276, 33512.
- (52) Leung, I. K. H.; Krojer, T. J.; Kochan, G. T.; Henry, L.; von Delft, F.; Claridge, T. D. W.; Oppermann, U.; McDonough, M. A.; Schofield, C. J. *Chemistry & Biology* **2010**, 17, 1316.
- (53) Zhang, Y.; Reinberg, D. *Genes & Development* **2001**, 15, 2343.
- (54) Kouzarides, T. *Current Opinion in Genetics & Development* **2002**, 12, 198.
- (55) Hong, S. H.; Cho, Y. W.; Yu, L. R.; Yu, H.; Veenstra, T. D.; Ge, K. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, 104, 18439.
- (56) Kwon, D. W.; Ahn, S. H. *Biochemical and Biophysical Research Communications* **2011**, 410, 614.
- (57) Raimundo, N.; Baysal, B. E.; Shadel, G. S. *Trends in Molecular Medicine* **2011**, 17, 641.
- (58) Sabourin, P. J.; Bieber, L. L. *Journal of Biological Chemistry* **1982**, 257, 7460.
- (59) Krieg, R. C.; Liotta, L. A.; Petricoin Iii, E. F.; Herrmann, P. C. *Journal of Biochemical and Biophysical Methods* **2004**, 58, 119.
- (60) Aebersold, R.; Mann, M. *Nature* **2003**, 422, 198.
- (61) Chait, B. T. *Science* **2006**, 314, 65.
- (62) Stephenson, J. L.; McLuckey, S. A.; Reid, G. E.; Wells, J. M.; Bundy, J. L. *Current Opinion in Biotechnology* **2002**, 13, 57.
- (63) Witze, E. S.; Old, W. M.; Resing, K. A.; Ahn, N. G. *Nature Methods* **2007**, 4, 798.
- (64) Ong, S. E.; Mann, M. *Nature Chemical Biology* **2005**, 1, 252.
- (65) Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B. *Analytical and Bioanalytical Chemistry* **2007**, 389, 1017.
- (66) Richards, D. P.; Sojo, L. E.; Keller, B. O. *Journal of Labelled Compounds & Radiopharmaceuticals* **2007**, 50, 1124.
- (67) Mann, M. *Nature Reviews Molecular Cell Biology* **2006**, 7, 952.
- (68) Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlett-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J. *Molecular & Cellular Proteomics* **2004**, 3, 1154.
- (69) Johnson, D. W. *Journal of Chromatography B* **2011**, 879, 1345.
- (70) Singleton, R. S.; Trudgian, D. C.; Fischer, R.; Kessler, B. M.; Ratcliffe, P. J.; Cockman, M. E. *Journal of Biological Chemistry* **2011**, 286, 33784.
- (71) Mann, M.; Jensen, O. N. *Nature Biotechnology* **2003**, 21, 255.
- (72) Craig, R.; Cortens, J. C.; Fenyo, D.; Beavis, R. C. *Journal of Proteome Research* **2006**, 5, 1843.
- (73) Elias, J. E.; Gibbons, F. D.; King, O. D.; Roth, F. P.; Gygi, S. P. *Nature Biotechnology* **2004**, 22, 214.
- (74) Carr, S. A.; Huddleston, M. J.; Annan, R. S. *Analytical Biochemistry* **1996**, 239, 180.
- (75) Kaiser, S. E.; Riley, B. E.; Shaler, T. A.; Trevino, R. S.; Becker, C. H.; Schulman, H.; Kopito, R. R. *Nature Methods* **2011**, 8, 691.

2 Chapter Two – Application of LC to the analysis of PTMs

This chapter discusses the development of three established chromatographic techniques for the separation of AAs and their modifications: reversed phase (RP) chromatography of 6-aminoquinolyl-n-hydroxysuccinimidyl carbamate (AQC) derivatised AAs, RP chromatography assisted by ion pairing (IP) reagent, and RP mixed mode (MM) chromatography. The aim was to investigate the applicability of these techniques to the analysis of PTMs including the compatibility of the LC methods with the MS detection, the possibility to separate the diastereomers of hydroxylated AAs, and potential limitations of these methods.

2.1 Chromatographic separation of AAs

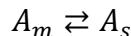
The first observation of an AA (cystine) dates back to 1810; and by 1922 the list of the twenty commonly observed proteinogenic AAs was completed with the discovery of methionine¹. However, it was not until the middle of the 20th century, when the quantitative analysis of the AAs in a protein, haemoglobin, was carried out². Since then AAs have been routinely analysed.

Separation of AAs began with paper chromatography³ (1948), followed by starch chromatography⁴ (1949), and ion exchange chromatography⁵ (1952). The next stage was marked by the introduction of gas chromatography techniques⁶ (1959), electrophoresis⁷ (1964), and thin layer chromatography⁸ (1969). Column chromatography (1900) was replaced first by high performance LC (1970), and subsequently by ultra performance LC (2004). A key change at each stage was the decrease in the particle size of the packed columns. This allowed significant improvement in resolution at higher flow rate⁹, thus, reducing analysis time.

The similarity of physical and chemical properties of AAs made selectivity one of the major challenges of analytical method development. The introduction of modified AAs into the analytical mixture increases the number of compounds with similar properties, which may require revision and development of the existing analytical methods. This has been addressed in two ways – through the use of selective MS detection and through the development of separation techniques.

AA analysis of PTMs is generally limited to HyP, and HyL and to a lesser extent the analysis of phosphorylated AAs. Hydroxylation and methylation regulated by 2OG- and Fe(II)-dependent oxygenases only subtly affect the physical and chemical properties of AAs such as solubility, acidity, and hydrophobicity.

In general the chromatographic process can be described using the distribution equilibrium between stationary phase and mobile phase:



A denotes analyte, and the subscripts **m** and **s** refer to the mobile and stationary phase respectively. The distribution constant is referred to as the capacity factor (*k*):

$$k = \frac{[A_s]}{[A_m]}$$

The equilibrium concentrations of the analyte in stationary phase ($[A_s]$) and in the mobile phase ($[A_m]$) are proportional to the time spent by the analyte in either stationary (the corrected retention time ($t_R - t_0$)) or mobile phase (retention time of the void volume (t_0))¹⁰ respectively. Therefore, this equation can be rewritten using conventional for chromatography units of time:

$$k = \frac{t_R - t_0}{t_0}$$

The distribution constant is linked to the thermodynamics of the process by Gibbs free energy¹⁰:

$$\Delta G = -RT \ln k$$

In order to describe RP chromatography assisted by IP reagents, which is addressed in this chapter, four more equations have to be introduced¹¹(Table 2-1).

Table 2-1. Equations describing processes in IP chromatography

A = analyte, H = hetaeron, L = the ligand of the stationary phase. The subscripts m and s attribute to the mobile phase and stationary phase respectively.

(a) describes the direct adsorption of the analyte by the stationary phase. (b) and (c) describe the process, in which first the ion pair is formed in a solution (b) and after that the ion pair is adsorbed on the surface (c). The last two equations describe the alternative process, in which the stationary phase is first modified by the hetaerons (d) and after that the analyte is being absorbed (e).

$A_m + L_s \rightleftharpoons AL_s$	$K_{d.a.a.} = \frac{[AL_s]}{[A_m] \cdot [L_s]}$	<u>d</u> irect <u>a</u> nalyste <u>a</u> bsorption (a)
$A_m + H_m \rightleftharpoons AH_m$	$K_{i.p.f.} = \frac{[AH_m]}{[A_m] \cdot [H_m]}$	<u>i</u> on <u>p</u> air <u>f</u> ormation (b)
$AH_m + L_s \rightleftharpoons AHL_s$	$K_{i.p.a.} = \frac{[AHL_s]}{[AH_m] \cdot [L_s]}$	<u>i</u> on <u>p</u> air <u>a</u> bsorption (c)
$H_m + L_s \rightleftharpoons HL_s$	$K_{h.a.} = \frac{[HL_s]}{[H_m] \cdot [L_s]}$	<u>h</u> etaeron <u>a</u> bsorption (d)
$A_m + HL_s \rightleftharpoons AHL_s$	$K_{m.e.a.} = \frac{[AHL_s]}{[E_m] \cdot [HL_s]}$	<u>m</u> odified surface <u>e</u> luate <u>a</u> bsorption (e)

In summary, it is important to point out two processes which result in the retention of the analyte – either direct adsorption of the analyte (for all chromatography types), or the adsorption of the ion pair (for IP chromatography only).

Chromatographic separation depends on the different types of interaction between analyte and stationary phase. Weaker interactions result in shorter retention times of the analytes. Energies of molecular interactions collected by Buszewski *et al.*¹² are shown in Table 2-2. The hydrophobic interaction energy for the AA side chains was assessed by transfer from water into a non-polar medium and was found to be 1-10 kJ/mol¹³. RP chromatography involves hydrophobic interactions between the molecules of the analyte and the stationary phase, and the non-specific hydrogen bond interaction between the molecules of analyte and

the non-protected hydroxyl groups of the stationary phase. Therefore, using RP chromatography the retention time of the AAs will be shorter, than for MM or ion exchange chromatography, where ionic interactions are involved.

Table 2-2. Energy of molecular interactions

Interaction Type	Energy (kJ/mol)
Ion-Ion	250-1050
Ion-Dipole	15
Dipole-Dipole	25-40
Hydrogen Bond	20-30
Donor-Acceptor	8-188

In order to develop a method for the separation of AAs, targeting in particular modified AAs, it is worth considering the properties of AAs and the chromatographic conditions, which affect their separation. According to Subirats¹⁴, the retention of AAs using normal phase and RP chromatography is substantially affected by AA polarity and hydrophobicity (Table 2-3).

pK_a values or the isoelectric point[#] can be used as a measure of AA polarity. Three non-correlating scales have been used to describe the hydrophobicity of AAs¹⁵. One only considers the hydrophobicity caused by the carboxyl and amino groups; another considers AA side-chains, as they are in the proteins; and the last one takes into account both the carboxyl and amino groups and side-chains. It is the last one which is applicable to the chromatographic analysis of free AAs.

A summary of data on the polarity and hydrophobicity is presented in Table 2-3. The pK_a values are taken from a recent publication by Henchoz *et al.*¹⁶, and values for isoelectric points were calculated from this data. The values of polarity of AA (AA) and the side chain (SC) have been adapted from Vallat *et al.*¹⁷. The authors have derived the polarity by comparison of partition coefficients in the octanol-water system for AAs with the distribution of the *n*-alkane having identical molecular volume. It can be noted, that the SC polarity, in

[#] Isoelectric point is the pH value at which the compound has no charge, or the charges are compensating each other in case of zwitterions.

the way it is introduced by Vallat *et al.*¹⁷, eliminates the contribution of the carboxyl and amino groups by subtracting the value for Gly. Undoubtedly, such an approach does not take into account the interaction between the carboxyl and amino groups and the side chains. Finally, the solubility is given as molality values for two different temperatures¹⁸, which are close to those commonly used in chromatographic analysis.

Table 2-3. Selected properties of the 20 standard AAs

AA	pKa			pI	Polarity		Solubility	
	Carboxyl Group	Amino Group	Side Chain		AA	SC	25°C	50°C
Ala	2.39 ± 0.08	9.82 ± 0.10		6.11	5.7	0.3	1.867	2.443
Arg	1.95 ± 0.22	9.04 ± 0.08	>12	>10.5	9.1	3.7	1.125	2.494
Asn	2.18 ± 0.11	8.8 ± 0.04		5.49	7.3	1.9	0.190	0.402
Asp	1.99 ± 0.08	10.01 ± 0.09	3.73 ± 0.07	2.86	7.2	1.8	0.038	0.086
Cys	2.04 ± 0.08	10.19 ± 0.45	8.42 ± 0.11	9.31	6.0	0.6	0.211	0.302
Glu	2.29 ± 0.11	9.87 ± 0.08	4.17 ± 0.06	3.23	7.6	2.2	0.291	0.596
Gln	2.24 ± 0.11	9.16 ± 0.06		5.70	7.4	2.0	0.060	0.142
Gly	2.43 ± 0.08	9.65 ± 0.08		6.04	5.4	0.0	3.370	5.295
His	1.64 ± 0.45	9.14 ± 0.08	6.09 ± 0.06	7.62	7.4	2.0	0.281	0.436
Ile	2.39 ± 0.07	9.67 ± 0.08		6.03	6.3	0.9	0.242	0.292
Leu	2.43 ± 0.05	9.71 ± 0.07		6.07	6.2	0.8	0.167	0.204
Lys	2.28 ± 0.22	9.23 ± 0.09	10.7 ± 0.13	9.97	8.7	3.3	1.687	2.512
Met	2.26 ± 0.10	9.17 ± 0.04		5.72	6.6	1.2	0.375	0.551
Phe	2.21 ± 0.15	9.17 ± 0.06		5.69	6.7	1.3	0.170	0.259
Pro	2.03 ± 0.16	10.77 ± 0.10		6.40	6.2	0.8	11.298	14.622
Ser	2.24 ± 0.11	9.15 ± 0.05		5.70	6.1	0.7	3.480	4.952
Thr	2.28 ± 0.13	9.03 ± 0.04		5.66	6.4	1.0	0.822	1.154
Trp	2.46 ± 0.12	9.43 ± 0.10		5.95	7.3	1.9	0.065	0.086
Tyr	2.36 ± 0.10	9.15 ± 0.12	10.56 ± 0.15	9.86	7.5	2.1	0.003	0.006
Val	2.42 ± 0.15	9.63 ± 0.08		6.03	6.2	0.8	0.501	0.621

The use of MS detection limits the choice of the solvents which can be used as components of the mobile phase. A water-acetonitrile mixture is commonly applied as the stationary phase in LC-MS analyses. Extensive research into the behaviour of the ionisable analytes in water-acetonitrile mobile phases has been carried out by Bosch *et al.*¹⁹. The authors have studied the pH changes in buffer-acetonitrile solutions and the variation of the pK_a value dependent on the content of acetonitrile (MeCN) in the solution.

Bosch *et al.* have shown that the pH of the buffer rises linearly with increase of the MeCN fraction. The increase of MeCN concentration from 0 to 60% results in the buffer pH increase by 0.8 to 1.4 units, depending on the initial concentration (pH) of the buffer. On the other hand, the pK_a of acids increases by 1 to 1.5 units of pK_a for the same change in the MeCN concentration, depending on the nature of the acid. For the bases the pK_a decreases by 0.4 units for the 60% change in MeCN concentration.

In order to enable MS detection in positive ion mode the AAs have to be positively charged, *i.e.* protonated. In order to enable protonation of AAs during electrospray ionisation, the chromatographic separation, preceding the ionisation, is carried out under a pH value below the isoelectric point of AAs. Therefore, for zwitterionic compounds such as AAs, the increase of the MeCN concentration should not affect the charge state of the carboxyl groups because both pH and pK_a values have a symbatic shift of comparable value. Even though the change of the pK_a value in the presence of aprotic solvent is more pronounced in the case of the amino groups, the pH of the mobile phase is still much lower than the pK_a of the amino group and the amino groups can be considered protonated over the pH and MeCN concentration range.

The way in which the pH of buffers changes with an increase of the MeCN fraction in the mobile phase has been previously studied^{20,21}. For acetic acid and FA (formic acid) – most commonly used acids when chromatography is combined with MS detection – the increase of MeCN concentration from 0 to 60% results in the increase of pH by about 1 unit.

The variation in the pK_a value of the amino group can be neglected because 27% of variation in the MeCN content will lead to only 0.2 units of pK_a variation of the amino function. In addition, during RP-IP chromatography, the majority of the AA amino groups will be blocked by the IP reagent. It is therefore, only worth considering the pK_a increase of the AA carboxylic groups.

It has also been pointed out that the retention of acidic and basic analytes on RP columns depends on two factors: hydrophobicity and degree of the ionisation¹⁴. It has been suggested that the retention time of acidic and basic ionic compounds can be presented as a linear combination of retention time of its protonated and non-protonated forms. It was shown that for the acidic compounds the retention time of the protonated form is higher than that of the non protonated; for the basic compounds the opposite was shown to be true.

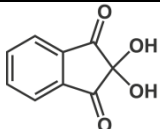
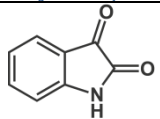
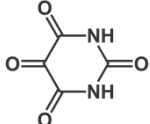
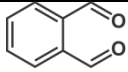
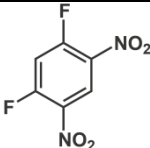
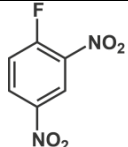
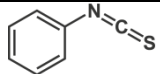
Using molecular modelling the pK_a of the TFA (trifluoroacetic acid) has been estimated, to be -0.5²². This value can be applied to other perfluorinated acids, which are used as the IP reagents. Therefore, such acids with very low pK_a values can be considered fully ionised despite the increase in MeCN concentration, and included from further consideration.

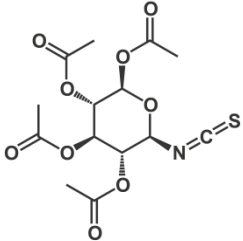
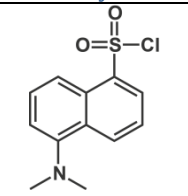
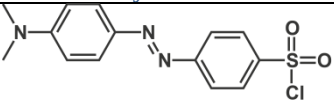
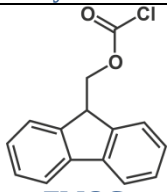
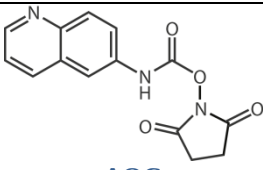
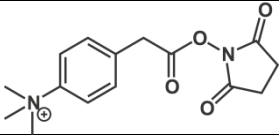
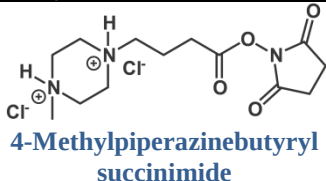
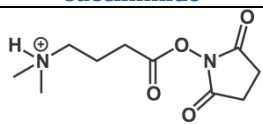
2.1.1 Derivatised AA analysis

UV absorbance and fluorescence detectors are commonly coupled to LC systems for analytical purposes. UV detectors measure the absorbance at a particular wavelength, whereas fluorescent detectors measure emission of a particular wavelength after excitation. Only three AAs – phenylalanine, tyrosine and tryptophan – have strong chromophoric or fluorophoric groups²³ (as a result of their conjugated aromatic systems) which allow their direct detection using UV or fluorescent detection. For most other AAs, derivatisation protocols (introduction of chromophoric or fluorophoric groups in order to improve absorbance or emission, respectively) have been previously developed (Table 2-4). The derivatisation of AAs has a significant impact on the chromatography as well. Derivatisation of the polar groups minimises interactions with the polar sites of the stationary phase, thus, reducing the retention time and improving the chromatographic peak shape. Derivatisation, however, has significant drawbacks such as longer preparation time, reagent interference, inability to derivatise secondary AAs and instability of the derivatives.

Table 2-4 shows a selection of common derivatising reagents that have been applied in LC for the AA analysis. Even though AAs have two polar groups – an amino group and a carboxylic group – derivatisation mostly targets the amino group. The reason for this is that amino groups readily take part in nucleophilic substitutions (Figure 2-1, Figure 2-2), which can be carried out under mild conditions. The carboxylic group, however, can be esterified, which normally requires longer reaction times and higher temperatures; alternatively, the substitution of a hydroxyl group can be carried out, however, the hydroxyl group is not a good leaving group.

Table 2-4. Derivatising reagents applied for the LC AA analysis

Derivatising Reagent	Comments	Ref.
 Ninhydrin (1948)	Post-column derivatisation. Used in paper chromatography.	24
 Isatin (1956)	Post-column derivatisation. Used in paper chromatography. Provides higher limit of detection than ninhydrin.	25
 Alloxan (1956)	Post-column derivatisation. Used in paper chromatography. Provides higher limit of detection than ninhydrin.	25
 o-Phthalaldehyde (1976)	Both pre- and post-column derivatisation. Not suitable for compounds with secondary amino group.	26,27
 Marfey's Reagent (1984)	L-Ala analogue is suitable for the separation of enantiomers via diastereomeres.	28
 Sanger's Reagent (1945)	Initially developed for determining of terminal AAs.	29
 Phenylisothiocyanate	Edman degradation developed initially for analysis of the peptide sequence starting N-terminal AA.	30,31

 2,3,4,6-Tetra-O-acetyl-β-D-glucopyranosyl isothiocyanate	Developed for separation of enantiomers of AAs.	32
 Dansyl chloride	Before introduction of AQC; this was the most common derivatising reagent for analysis of AAs.	33
 Dabsyl chloride	60 times more sensitive than Sanger's Reagent.	34
 FMOC	Suitable for secondary amino groups.	35
 AQC	Chemical analogues of this compound were developed for analysis of AAs.	36
	Developed for LC-MS analysis; stable fragment with mass 177 Da.	37
 4-Methylpiperazinebutyryl succinimide	Developed for MS-MS detection as an alternative to iTRAQ.	38
	Developed for MS-MS detection as an alternative to iTRAQ.	38

In the course of development of derivatising reagents for AA analysis, one of the major aims was to introduce a strong chromophore for UV and fluorescent detection; this is why most derivatising reagents have aromatic moieties as part of their structure. Another important

aspect was development of reagents which were suitable for both primary and secondary amino groups. Introduction of Sanger's Reagent²⁹ and, later, Marfey's Reagent²⁸ enabled derivatisation of AAs with secondary amino groups. Another problem, which was addressed in the development of derivatising reagents, was to enable the separation of AA enantiomers. Finally, when MS detection became common derivatising reagents specific for MS were developed.

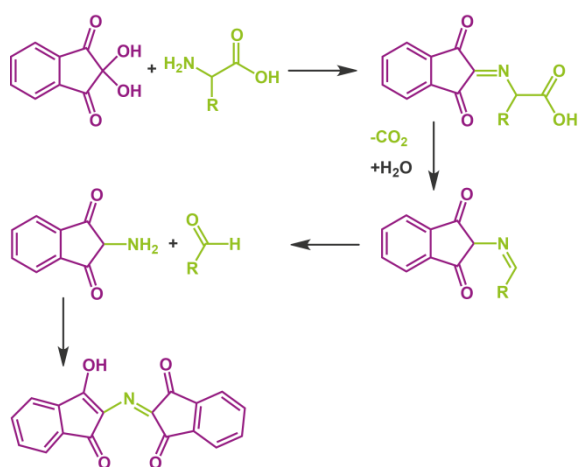


Figure 2-1. Scheme of the ninhydrin derivatisation

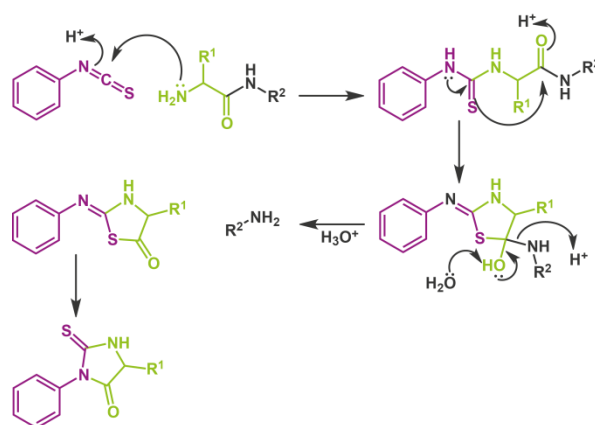


Figure 2-2. Scheme of Edman degradation

2.2 Chromatography testing and development

One of the major objectives of the work described in this thesis was the development of methods suitable for the analysis of PTMs, particularly of methylation and hydroxylation catalysed by 2OG oxygenases. Introduction of modified AAs increases the number of analytes, thus, increasing the possibility of co-elution of the AAs. The use of MS detection enables selective analysis of co-eluted compounds, providing they have distinct masses. The requirements to the analytical methods will therefore, differ from those applied to the analysis of AAs in general.

1. In order to enable MS detection chromatographic mobile phases have to be compatible with MS detection. To enable accurate quantitative analysis, it is necessary to ensure base-line separation of isobaric compounds (isomers and diastereomers).
2. 2OG- and Fe(II)- dependent oxygenase catalysed hydroxylation introduces a hydroxyl group into the carbon backbone of the AA, thus, creating a new chiral centre. In order to establish the stereochemistry of the hydroxylation it is necessary to separate the diastereomers of hydroxylated AAs.
3. Analysis of PTMs implies work with biological samples as well as samples from biochemical assays. It is necessary to ensure that the analytical techniques are compatible with the specific components of the matrices such as salts and buffers.

In order to increase the resolution of AAs similar in hydrophobicity and to minimise the analysis time RP chromatographic methods were investigated. Three types of chromatography have been tested and optimised in this project: (1) RP chromatography of AQC-derivatised AAs (Figure 2-3 A), (2) RP chromatography assisted by the IP reagent (Figure 2-3 B), and (3) RPMM chromatography (Figure 2-3 C). In cases (A) and (B) the polar amino groups are ‘shielded’ by the derivative group and by the IP reagent respectively. In case (C) the amino group of the AA can directly interact with the acidic groups of the stationary phase, thus, increasing retention.

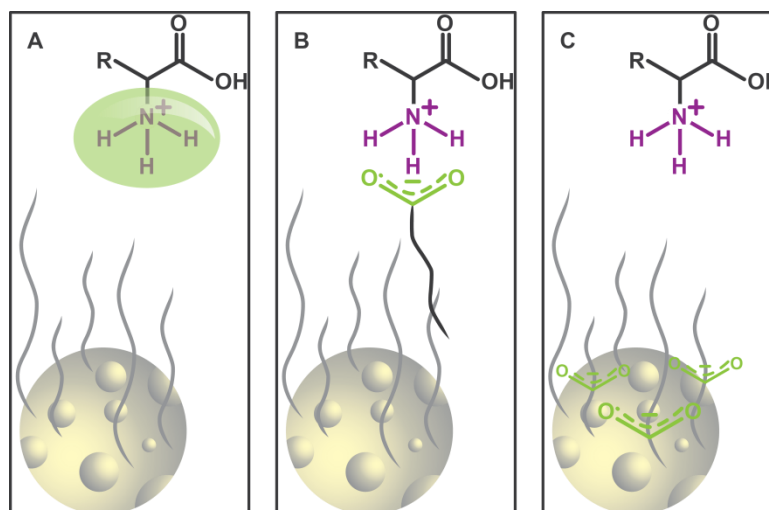


Figure 2-3. Schematic representation of stationary phase with analyte interaction for the three different chromatographic modes

- A. RP chromatography of derivatised AAs
- B. RP-IP chromatography
- C. RP MM chromatography

In this thesis it was attempted to assess the applicability of the chromatographic methods for the separation and identification of a wide range of PTMs using a limited range of commercially available AA standards. In order to test chromatography total run time (the elution time of the 20 proteinogenic amino acids) was identified. The average peak width was estimated for each chromatography type. To assess the chromatographic performance, chromatographic resolution was also measured. Chromatographic resolution was defined as the ratio of total run time to the peak width and shows the peak capacity or the maximum number of compounds, which can be fully resolved. To test the resolution of isobaric compounds structural isomers and diastereomers of AAs were tested. Resolution of Leu and Ile was used as a selectivity test for structural isomers.

First, proteinogenic AAs were selected to test the chromatography methods. Second, modified AAs catalysed by 2OG- and Fe(II)- dependent oxygenases were analysed including HyP, HyL, HyArg, HyHis, HyLeu, MeLys, diMeLys, triMeLys, MeArg, diMeArg. Third, in order to calibrate the chromatographic system, a homologous series of α -AAs was used.

Finally, commercially available positional isomers, and homologues of proteinogenic AAs were used in order to establish retention pattern. Calibration of the chromatographic system and retention pattern recognition are described in detail in Chapter Four. The exhaustive list of AAs tested for each type of chromatography is provided in Appendix B.

2.2.1 RP chromatography of AQC-derivatised AAs

2.2.1.1 Method background

RP chromatography of AQC (see Table 2-4) is a method that emerged in the 1990s. The derivatising reagent (Figure 2-4) was synthesised by S. Cohen *et al.*³⁶, and in the same paper a basic chromatographic method was suggested for the analysis of AAs. Different detection methods have been reported for AQC AA derivatives including fluorimetry³⁹ and electrochemical detection⁴⁰, though the quantitative methods were comparable to the one using UV detection, suggested initially⁴¹.

For this method, submicromolar limits of detection have been reported for UV detection⁴²⁻⁴⁴. Initial chromatography has been performed on the RP C18 and applied by many authors^{45,46}. It has been shown by many authors that the derivatisation enables both UV and fluorescent detection⁴³.

2.2.1.2 Method development

In this work, UV detection was compared with MS detection in scanning mode and MS detection in selective ion monitoring (SIM) mode. UV detection was carried out at 254 nm, at which wavelength the absorbance is maximal for all the AQC derivatised AAs³⁶. For MS analysis, two MS systems were used: LCT Premier (Waters), a time-of-flight (ToF) MS analyser, and Quattro Micro (Waters), a triple-quadrupole analyser. MS analysis in the scanning mode (scanning range 160 to 510 Da) was carried out in positive ion mode. The MS peak width was found to be approximately 0.1 Da. For quantitative analysis, extracted ion

chromatograms were created for the window centred on the maximum of the peak ± 0.05 Da. MS analysis in the SIM mode was carried out in positive ion mode. For standard AAs, the mass of the peak was identified by directly injecting the mixture of derivatised AAs. The MS peak width of the spectrum was approximately 0.5 Da; the maximum of the mass peaks for different AAs was shown to correspond to the molecular mass of the derivatised AA expressed in Da $+0.2$ Da. This mass was chosen for the analysis in SIM mode.

As a starting point in this work the separation of proteinogenic AAs using AQC derivatisation was explored⁴⁷. Analysis involves derivatisation of AAs with AQC under basic conditions of borate buffer (pH = 9.5). AQC reacts with both primary and secondary amino groups (Figure 2-4, AQC derivatisation scheme), which makes the method suitable for analysis of proline and its modifications.

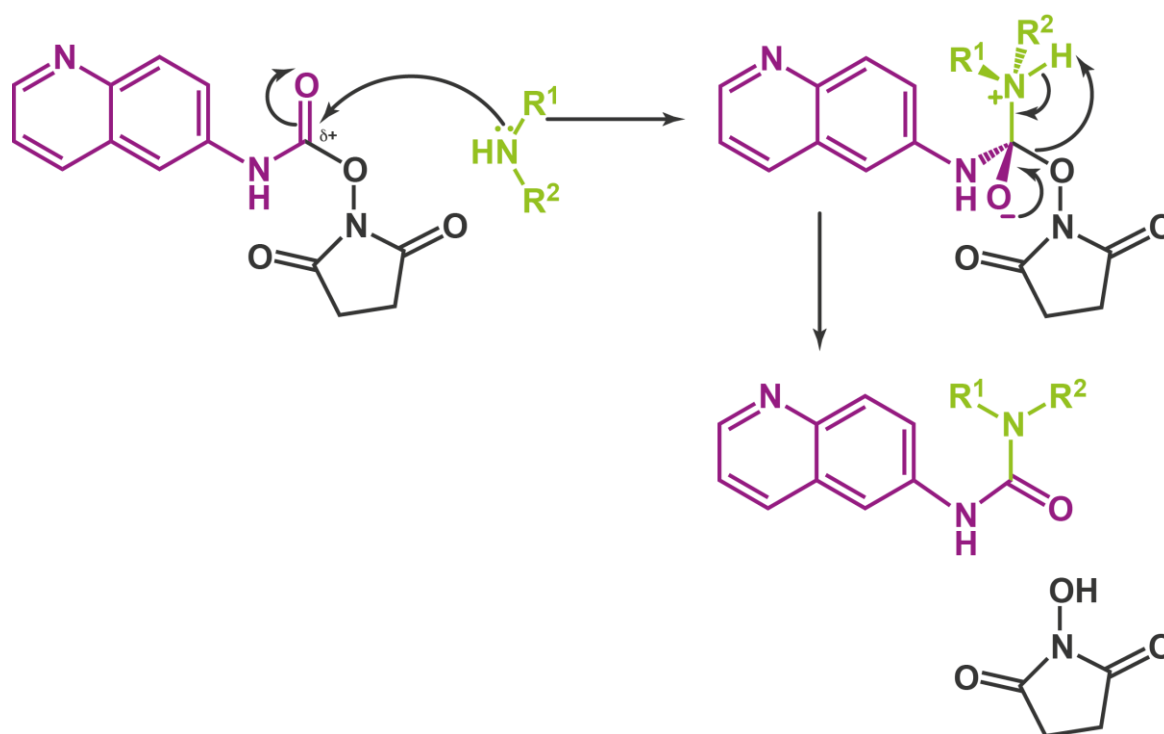


Figure 2-4. Scheme of derivatisation by AQC

The application of the RP chromatography to the analysis of the AQC-derivatised AA derived from biochemical samples was tested. The introduction of a C18 UPLC column (AccQTag Ultra C18, 2.1×100mm, 1.7µm, Waters) allowed reduction of the total run time in comparison with the original publication ($t_{R, \text{Phe}} < 8 \text{ min}^{48}$ (Figure 2-5 B) vs. $t_{R, \text{Phe}} = 38 \text{ min}^{36}$). In order to simplify the chromatography and make the elution pattern predictable multi-step gradient was substituted with a single-step gradient; the chromatography was carried out in RP mode from 100 to 75% water fraction in 25 min (Chapter Eight), resulting in an increase of the total run time ($t_{R, \text{Phe}} < 13 \text{ min}$, Figure 2-5 A).

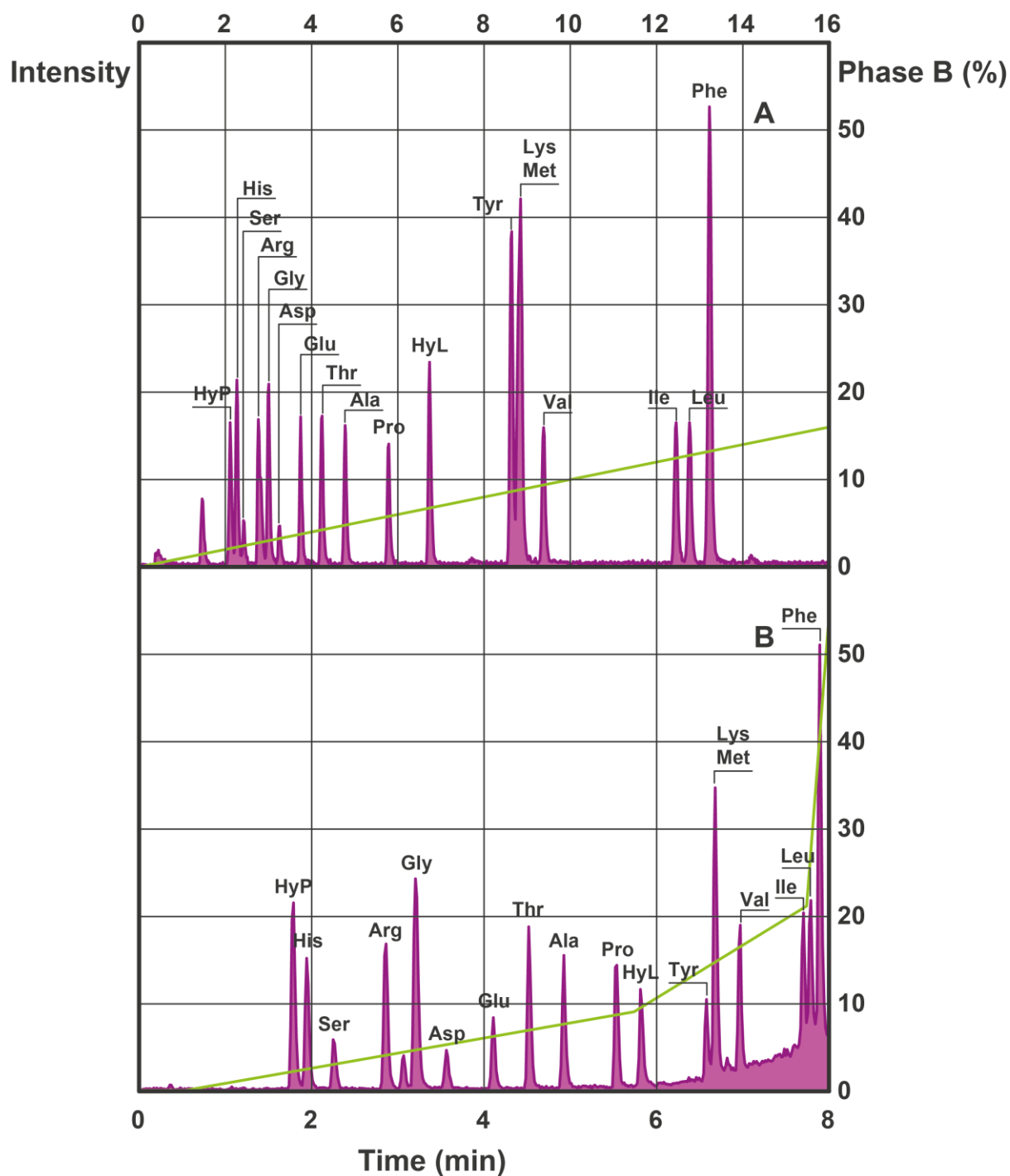


Figure 2-5. Single-step and multi-step gradient RP chromatograms of AQC-derivatised AAs

- A. Single-step gradient
- B. Multi-step gradient

Injection of a solution containing only borate buffer pH = 9.5 (blank) gave a single peak corresponding to the void volume (Figure 2-6 B, t_R = 0.35 min); whereas analysis of the buffer solution containing derivatising agent (reagent blank) reveals a product of self-

derivatisation (Figure 2-8, $t_R = 8.10$ min), ammonia adduct (Figure 2-7, $t_R = 1.65$ min), and a compound which was not assigned a structure (Figure 2-9, $t_R = 1.75$ min). The TIC of the reagent blank is presented in Figure 2-6 A.

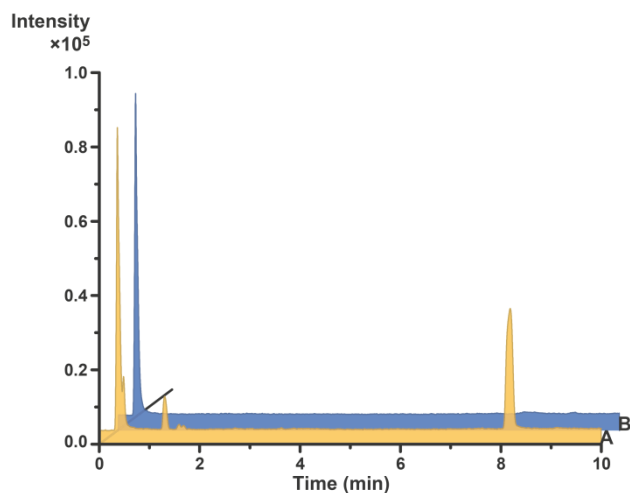


Figure 2-6. Chromatogram of blank runs (TIC)

1. Blank with added derivatising reagent
2. Blank without added derivatising reagent

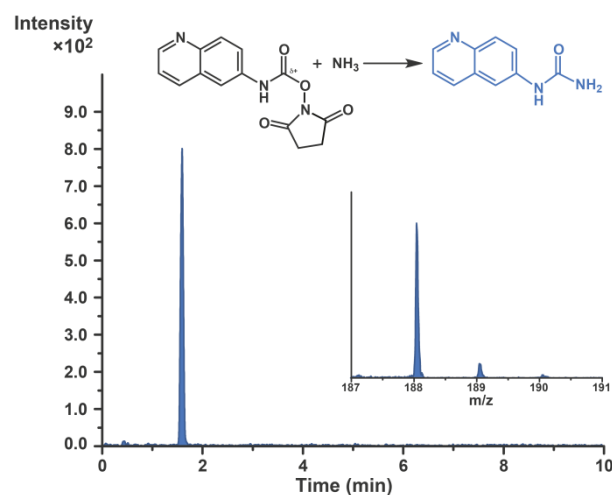


Figure 2-7. Chromatogram (XIC, $m/z = 188.1 \pm 0.05$) and the spectrum of derivatised ammonia ($t_R = 1.65$ min)

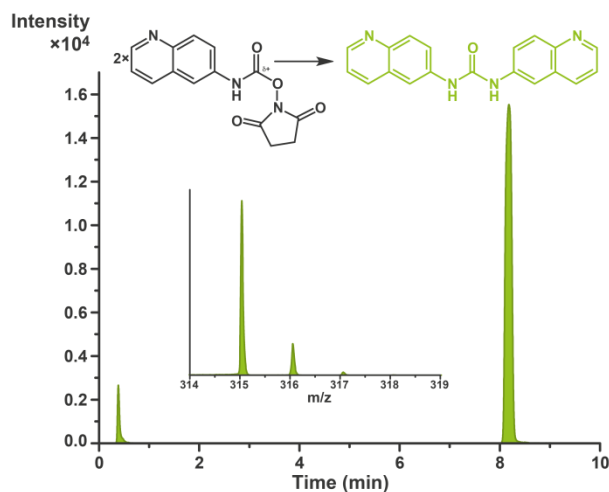


Figure 2-8. Chromatogram (XIC, $m/z = 315.1 \pm 0.05$) and the spectrum of the AQC self-derivatisation product ($t_R = 8.10$ min)

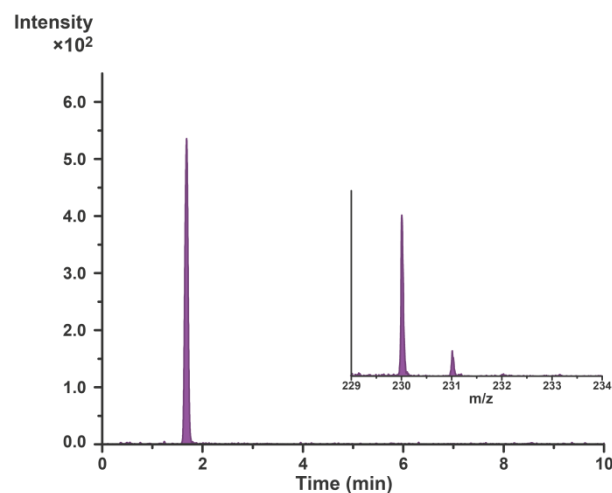


Figure 2-9. Chromatogram (XIC, $m/z = 230.1 \pm 0.05$) and the spectrum of the unknown compound ($t_R = 1.75$ min)

In order to test the applicability of RP chromatography of AQC-derivatised AAs to the separation of the isobaric AA diastereomers and positional isomers, commercially available

standards were analysed (Table 2-5). The chromatographic peak width for the standard AAs was 0.1-0.2 min. It can be seen from Table 2-5 that base-line separation is achieved for pairs of diastereomers and positional isomers tested using the single-step linear gradient described above, thus, enabling accurate quantification of the isobaric isomers and diastereomers.

Table 2-5. Separation of AQC-derivatised isobaric AAs

m/z (Da)	Name	t _R	Name	t _R
89	N-MeGly	3.57	β -Ala	3.96
103	β -AlBu	5.75	α -AlBu	6.52
115	D-AllylGly	7.75	Pro	5.77
119	HSer	3.03	Thr	4.21
131	4- <i>trans</i> -HyP	2.07	3- <i>trans</i> -HyP	2.24
	4- <i>cis</i> -HyP	3.58	3- <i>cis</i> -HyP	3.66
	Ile	12.43	Leu	12.75
162	2S,5S-HyL	6.53	2S,5R-HyL	6.81
169	3-MeHis	2.39	1-MeHis	2.53
171	2S,3S-HyHis	0.83	2S,3R-HyHis	1.19
174	Arg	2.79	N,N-diMeLys	2.89
190	2S,3R-HyArg	1.9	2S,3S-HyArg	1.97

Tris buffer is commonly used in biochemical assays to maintain an appropriate pH for enzyme action (often pH 7-9). The primary amino group of Tris buffer is readily derivatised by AQC. AQC-derivatised Tris gives a heavily tailed (~ 0.5 min) chromatographic peak (t_R ~ 3 min, m/z = 292.1) with a peak width of 0.25 min, as compared to AAs with a peak width of *ca.* 0.10 min. Heavy tailing can be explained by the presence of 3 hydroxyl groups, which may take part in non-specific interactions with hydroxyl groups of the stationary phase through hydrogen bond formation. In total, Tris affects over 5% of the chromatogram (0.75 min out of under 15 min of total run time) and can potentially compromise the analysis of the co-eluted compounds due to ionisation suppression of target analytes caused by the excess of Tris ions. The interference of the Tris in the analysis is described in Chapter Five. It was shown before⁴⁹ that alcohols react in exactly the same way as amines³⁹.

Tris buffer at 50 mM is normally used in biochemical assays. The concentration of substrates does not exceed 1 mM, with 10% of modification. Thus, the Tris concentration may exceed

the concentration of the target analyte by 500 times. For derivatisation, normally a 4 fold excess of AQC is used, as compared to the estimated AA concentration (Chapter Eight). Therefore, the large excess of Tris hinders the derivatisation of target AA with AQC and makes the method unsuitable for AA analysis. As discussed previously, the majority of derivatising agents are targeting amino groups of AAs. Therefore, for the samples in which Tris buffer cannot be avoided it was necessary to develop an alternative AA separation method which does not require derivatisation. In order to overcome the limitations imposed by the derivatisation procedure chromatography of non-derivatised AAs was investigated.

2.2.2 RP-IP chromatography

2.2.2.1 Method background

Separation of non-derivatised AAs can be performed in several modes – RP, ion exchange, or MM. RP-IP chromatography is a relatively new technique that emerged at the beginning of the 1970s. It was introduced for achieving a balance between lack of retention in the case of the analysis of polar compounds on RP columns, and long analysis time in the case of ionic chromatograms⁵⁰. In IP chromatography, retention is affected by both electrostatic forces and hydrophobic interaction⁵¹. Schematically this process is presented in Figure 2-3 B.

A wide range of compounds including alkylsulphonates⁵² and alkylsulphates⁵³ have been tested as IP agents in AAs analysis. Perfluorinated carboxylic acids proved most useful for analytical purposes due to their higher volatility^{54,55}. The first member of this homologous family, TFA, was reported to form ‘very strong ion pairs’ with eluate, which do not dissipate under the conditions of the electrospray ionisation source, and thus, prevent ionisation of analytes⁵⁶. In addition, TFA affects both spray stability and signal intensity as a result of high conductivity and high surface tension. These can be circumvented by addition of organic sheath liquid, highly concentrated weak acid, which competes for pairing with analysed AAs,

thus, reducing their pairing with TFA. Even though the techniques mentioned above eliminate the problems connected with TFA this compound remains an undesirable IP agent because all the precautions associated with its use drastically increase the time of analysis, while its corrosive properties can affect the equipment⁵⁶.

It was shown by Petritis *et al.* that within a homologous series of perfluorinated carboxylic acids that the increase of the chain length results in the increase of adsorbed quantities on the non-polar stationary phase (commonly C18 or C8). Thus, for Purospher RP-18e and the 1 μ M concentration of IP reagent in the mobile phase the absorbed quantities were shown to be as follows: 173 μ M pentadecafluorooctanoic acid (PDFOA), 65 μ M tridecafluoroheptanoic acid (TDFHA), 6 μ M nonafluoropentanoic acid (NFPA), 3 μ M heptafluorobutyric acid (HFBA), *ca.* 0 μ M (TFA)⁵⁴. The use of longer chain IP reagent should result in longer retention and better separation of the AAs.

On one hand, an increase in surfactant concentration leads to an increase in relative retention time of the analyte and potentially higher resolution. On the other hand, the higher the surfactant concentration, the lower the detector response⁵⁷ and the higher the risk of analyser contamination⁵⁸. It was shown, however, that the use of perfluorinated acids as IP agents does not affect the sensitivity of MS detection of AAs; on the contrary, enhancement in sensitivity was observed in comparison with the analysis carried out without those additives⁵⁹.

Another aspect affecting detection that should be taken into consideration is co-elution with the interfering agent. Response of the detector to Egacid Yellow M was shown to decay exponentially with the increase in concentration of co-eluted Egacid Blue A2G, and vice versa⁵⁷. Therefore, it is extremely important to achieve resolution when analysing minor components, since a higher concentration of the interfering substance may reduce the response of the target compound to zero.

It was found that an increase in MeCN concentration in the water-organic mobile phase leads to an increase in the MS response of the detector to the following types of AAs: neutral, basic, acidic and aromatic. The response on MeCN dependency reaches its maximum in the 50-70 percent MeCN concentration range⁵⁸. This was attributed to the weak surface tension ($\gamma = 0.03 \text{ N}\cdot\text{m}^{-1}$, as compared to water $\gamma = 0.07 \text{ N}\cdot\text{m}^{-1}$) and low polarity of MeCN ($P = 5.8$, as compared to water $P = 9$), which allows better ion desolvation from the charged droplets. The authors also showed that the maximum of the MS response was observed in the range between 5 and 50 mM concentration of nonafluoropentanoic acid.

Table 2-6. Examples of IP chromatography for separation of AAs

Ref.	Column	Phase A	Phase B	Programme	Flow Rate
60	Phenomenex Luna 150×2.0 mm; 5 μm 40°C	30 mM NH_4OAc pH 6.0 with AcOH	5% AcOH	Isocratic 12.5% A 42 min	0.2 ml/min
61	Alltima C18 250×4.6 mm; 5 μm N/A°C	TFA in MeCN 0.1% (v/v)	–	–	N/A
62	QS UptiSphere 50×2.0 mm; 3 μm 40°C	H_2O ; 5 mM TDFHA sample dissolved in 5 mM TDFHA	MeCN	Gradient 100→75% A 15 min	N/A
63	Hypercarb S 150×2.0 mm; 5 μm 40°C	H_2O ; 20 mM NFPA	MeCN	N/A	0.2 mL/min, split 1/10
64	Discovery C18 50×2.1 mm; 5 μm 40°C	H_2O ; 0.1% TDFHA	MeCN; 0.1% TDFHA	Gradient 90→2% A	0.3 mL/min
65	Purospher RP-18e 50×2.1 mm Ambient	H_2O ; 0.5 mM PDFOA	MeCN	N/A	1 mL/min
65	Supelcosil ABZ+ Plus 150×4.6 mm Ambient	H_2O ; 1 mM TDFHA	MeCN	N/A	1 mL/min
66	50×2.1 mm 1.8 μm ambient	H_2O ; 0.5 mM TDFHA	MeCN	Gradient 100→75% A 18 min	0.2 mL/min

As shown in Table 2-6, the solutions of perfluorinated carboxylic acids in water-MeCN are most commonly used for the separation of non-derivatised AAs. In some cases, MeCN was adjusted to the same concentration of IP reagent as the water solution, which probably allows for better equilibration. The same effect is expected when AA solutions are diluted with

perfluorinated carboxylic acid. The most common flow rate for the purposes of non-derivatised AA separation is $0.2 \text{ mL} \cdot \text{min}^{-1}$. The concentration of surfactant is within the range 0.5 to 1mM, but in some cases may exceed that concentration and reach 20 mM.

2.2.2.2 Method development

UV detection is not practical for non-derivatised AA analysis, because only Phe, Tyr, and Trp have strong chromophore groups. Therefore, MS detection was used. RP-IP chromatography was coupled to a Quattro Micro triple quadrupole MS. Positive mode SIM and SRM detection techniques were developed. In order to identify cone voltage and collision energy at which the signal is maximal, 26 AAs were injected directly ($100 \text{ } \mu\text{M}$, $10 \text{ } \mu\text{L}/\text{min}$) (Appendix B). The average cone voltage was found to be $13.4 \pm 0.8 \text{ V}$ (d.f. = 24); the collision energy was found to be $15.4 \pm 2.2 \text{ V}$ (d.f. = 24). Cys was shown to require a higher cone voltage (31 V) and collision energy (33 V). Introduction of a hydroxyl group results in approximately 2 V decrease in both cone voltage and collision energy and introduction of a methyl group does not significantly affect detection. Ionisation suppression of target analytes caused by an excess of Tris ions was observed. For SIM mode, reduction of the base-line was observed where Tris was eluted.

In my work, as a starting point 0.1% PDFHA was used as the IP reagent. Generally, the IP reagent is added to phase A only (Table 2-6). In order to facilitate the reconditioning of the column, IP reagent was added at a concentration of 0.1% by weight (*ca.* 2.4 mM) to both mobile phases A and B. In order to simplify the mobile phase preparation procedure, the concentration of the IP reagent was made 2 mM. As suggested by Petritis⁶³, an RP C18 column was used. Use of a UPLC chromatographic column (BEH C18, $2.1 \times 50 \text{ mm}$, $1.7 \mu\text{m}$) allowed us to reduce the total runtime ($> 30 \text{ min}$ in original publication⁶³) and elute 20

proteinogenic AAs under 6 min (Figure 2-10) using a gradient of water fraction 100 to 70% (Chapter Eight).

In order to select optimal separation conditions three IP reagents (tridecafluoroheptanoic acid – TDFHA, pentadecafluorooctanoic acid – PDFOA, and heptadecafluorononanoic acid – HDFNA) were compared (Appendix B). 46 AA standards were run in 5 replicates (t_R confidence interval <0.1 min). It can be seen (Appendix B) that from TDFHA to PDFOA and then to HDFNA the retention time decreases by approximately 0.5 min for the compounds eluted before NLe ($t_R = 4.57$ min using TDFHA) and increases by approximately 1.5 min for the compounds eluted after NLe. It can also be seen that the resolution of compounds eluted after NLe decreases for the longer chain IP reagent. The resolution of the compounds eluted before Ala ($t_R = 0.78$ min using TDFHA) is not affected by the type of IP reagent. Total run time increases from TDFHA ($t_R = 5.41$ min) to PDFOA ($t_R = 6.79$ min) and then to HDFNA ($t_R = 8.06$ min). Thus, TDFHA was chosen for the analysis of AAs, because it provides better resolution and allows reduction of the total run time.

The width of the chromatographic peak was found to be 0.1-0.2 min. Cis- and trans-diastereomers of HyP were base-line resolved (Table 2-7); however, the positional isomers 3-trans HyP and 4-trans HyP were closely eluted. Diastereomers of HyL were eluted as one peak. Positional isomers of MeHis were not fully resolved using this method. Ile and Leu were base-line separated. For the quantitative analysis of isobaric compounds a shallower gradient or isocratic mode should be used.

Table 2-7. Separation of non-derivatised isobaric AAs using RP-IP chromatography (TDFHA)

m/z (Da)	Name	t_R	Name	t_R
119	HSer	0.59	Thr	0.61
131	3-trans-HyP	0.35	4-trans-HyP	0.39
	4-cis-HyP	0.48	–	–
	Ile	4.26	Leu	4.45
162	5-HyL	5.18	Carnitine	2.26
169	3-MeHis	4.94	1-MeHis	5.03
174	Arg	5.41	N,N-diMeLys	5.46

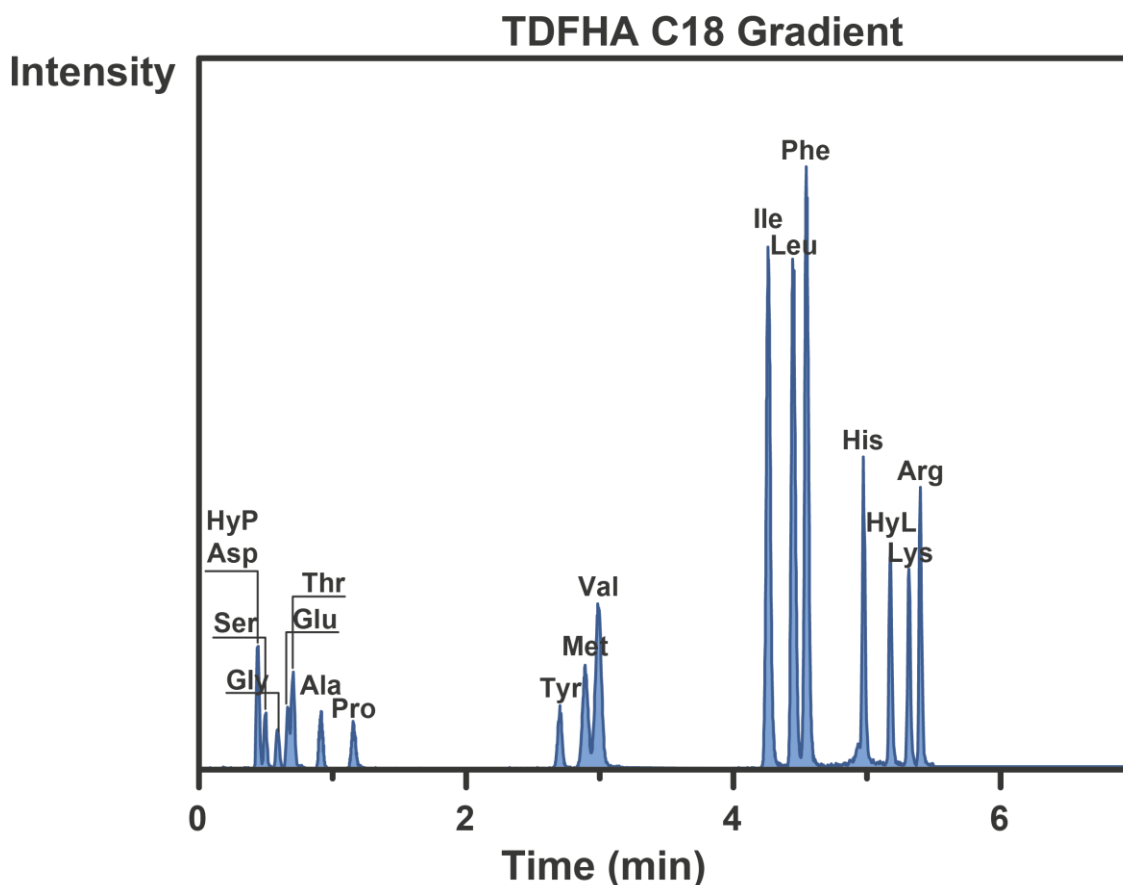


Figure 2-10. RP-IP chromatogram of the non-derivatised 20 standard AAs

20 standard AAs separated on C18 using TDFHA as an IP reagent

Thus, RP-IP chromatography proves to be a very fast chromatographic method, providing high resolution of the structural isomers. Unfortunately, the IP reagent was shown to cause contamination of the mass spectrometer with the IP reagent ions (m/z series $94+n \times 50$ Da) compromising detection in negative ion mode for subsequent analytes. It was shown that ionisation suppression resulted from the heterodimer of analyte and IP reagent formation⁶⁷. It is possible that such mechanism might take place in the gas phase as well, when the contaminating IP reagent forms a heterodimer with the analyte. Within the shared instrument

facility, it was unacceptable to pursue these experiments further. Therefore, alternative separation methods for non-derivatised AAs separation were investigated.

2.2.3 Mixed mode chromatography of AAs

2.2.3.1 Method background

MM chromatography has already been used for the separation of proteinogenic AAs as well as HyLys and HyPro from collagen and other proteins⁶⁸⁻⁷⁰. These publications suggested using a gradient of sulphuric acid in water⁶⁸ (total run time > 2 h) using IRMS detection. Sulphuric acid as a mobile phase additive is not compatible with molecular MS due to its highly corrosive properties and suppression of analyte ionisation. It was therefore, necessary to develop an alternative method suitable for electrospray ionisation (ESI) MS. It was preferential to use organic solvents which reduce the surface tension of the droplets in the ESI source and lead to faster evaporation, thus, providing higher ion intensity.

The SIELC (Prospect Heights, IL, USA) MM cation exchange columns (PrimeSep A series) have an acidic group attached to the surface by a linker of four carbon atoms and an aliphatic C12 carbon chain. Both the aliphatic chain and the acidic group are assembled in one molecule and then attached to the surface of silica gel⁷¹. The structure of the surface ligand is schematically presented in Figure 2-3 C. Depending on the type of column the pK_a of the acidic group of the stationary phase varies in broad range (Table 2-8), thus, allowing selection of the optimal properties for the separation.

Table 2-8. pK_a of the acidic groups of SIELC chromatographic columns of PrimeSep A series

Column type	pK_a
PrimeSep A	0
PrimeSep 100	1
PrimeSep 200	2
PrimeSep 500	5

2.2.3.2 Method development

For MM chromatography, MS detection in the SIM mode was used. The parameters of detection of non-modified AAs were the same as previously developed for MS detection coupled to RP-IP chromatography.

The retention of AAs using MM chromatography is determined predominantly by ionic interactions. Therefore, the separation should be defined by the pK_a of the AAs and that of the acidic groups of the stationary phase. PrimeSep 100 (2.1×250mm, 5 μ m) and PrimeSep 200 (2.1×250mm, 5 μ m) were selected for the initial screening because of the similar values of stationary phase acidic group pK_a to those of carboxyl groups of AAs (Table 2-3). In order to compare the results, PrimeSep A (2.1×250mm, 5 μ m) and PrimeSep 500 (2.1×250mm, 5 μ m) were also tested using optimal gradient conditions.

Two factors were expected to affect the retention of AAs using MM columns – hydrophobicity (lypophilicity) and the pK_a of the AAs (or ionisability). Therefore, two factors affecting the elution time were considered. First, the presence of the organic solvent increases the solubility of hydrophobic AAs such as phenylalanine, thus, reducing the retention time. Second, the counter ion (in this case cation) in the mobile phase, which can compete with AAs for the anionic groups of the stationary phase, was considered.

The choice of organic modifiers of the mobile phases compatible with MS is normally limited to methanol and MeCN⁷². For PrimeSep columns, the former was not an option because of the possible methylation of acidic groups of the stationary phase. Therefore, MeCN was used in all the experiments.

The choice of competing counter ions is much broader. The rationale behind the choice was stronger interaction between the counterion and the anionic group of the stationary phase. The interaction of the charged species depends on the charge density; the bigger the cation

the lower the charge density, given the same charge. Proton has the highest possible charge density therefore, its interaction with the anionic group should be stronger than that of the amino group of the AA. Ammonium formate (AF) was suggested as a source of ammonia ions, which should be close in size to that of the amino group of AAs.

In order to develop optimal chromatographic conditions for AAs separation by RP MM chromatography, separation by PrimeSep 100 and PrimeSep 200 were compared using different mobile phase concentrations of MeCN (0, 10, 20 and 30%; Figure 2-11 and Figure 2-12), AF (1, 5, 10, 25, 50 mM; Figure 2-13 and Figure 2-14), and FA (0, 0.1, 0.3, 0.5 and 0.7%; Figure 2-15 and Figure 2-16).

PrimeSep 100 was shown to be very retentive for standard AAs, and it was not possible to elute all the AAs within 30 min even at high concentrations of FA or AF. The advantage of PrimeSep 100 is its capability of separating the very closely eluted pair of isobaric Leu and Ile. The best separation ($R = 0.75$) was achieved at 10% MeCN and 0.05% FA (Figure 2-12 C). This level of separation was not achieved with either PrimeSep A or PrimeSep 200, which does not show a clear pattern in pK_a of acidic groups of stationary phase.

For PrimeSep 200 FA at 0.1% (*ca.* 26mM) concentration (Figure 2-15 D) has the same effect as AF at 25 mM concentration (Figure 2-13 B) on elution of the most strongly retained arginine (t_R just over 15 min). Thus, for PrimeSep 200 FA rather than AF was used in order to avoid introduction of new ions.

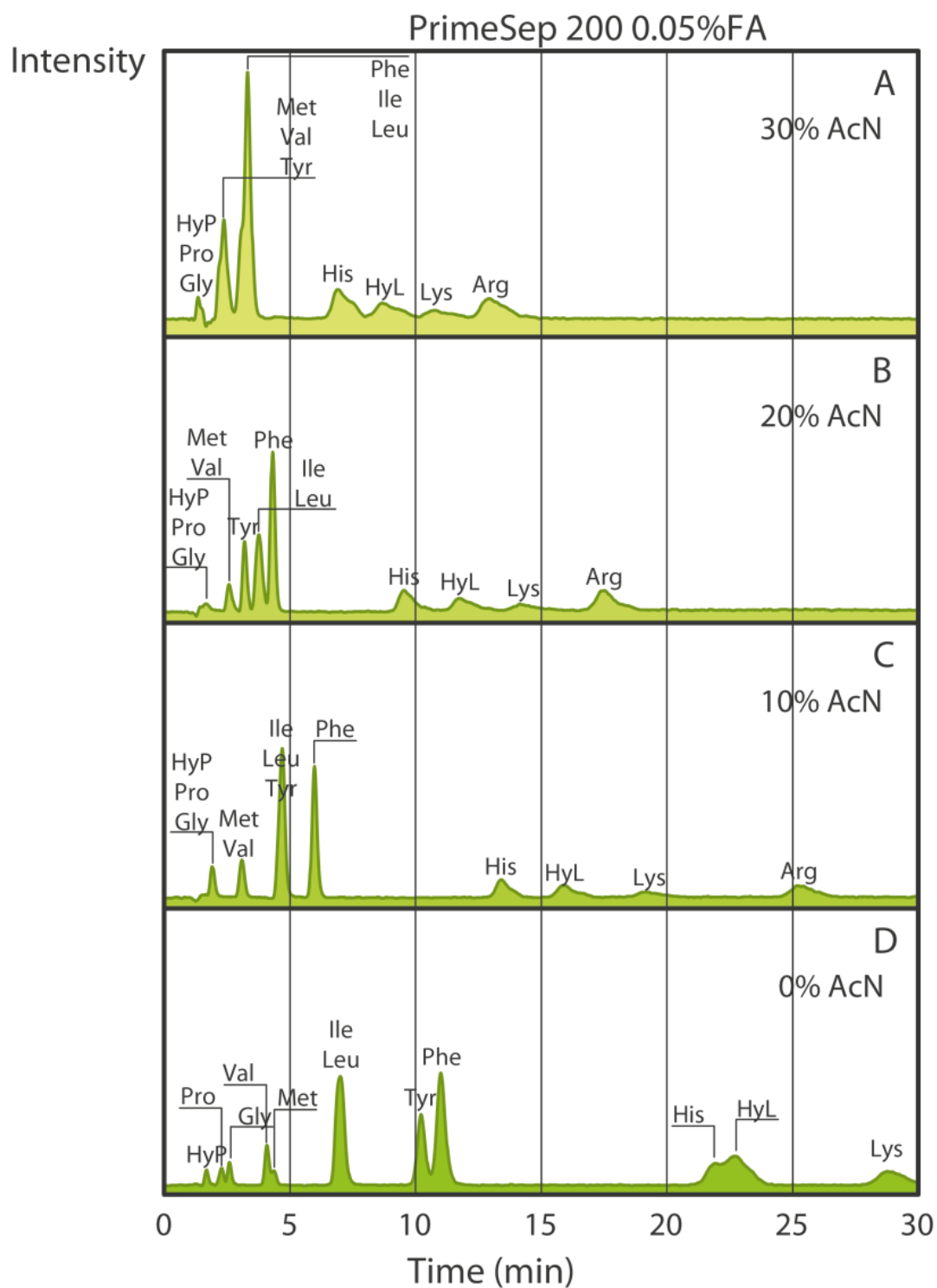


Figure 2-11. Chromatograms for PrimeSep 200 at different MeCN concentrations

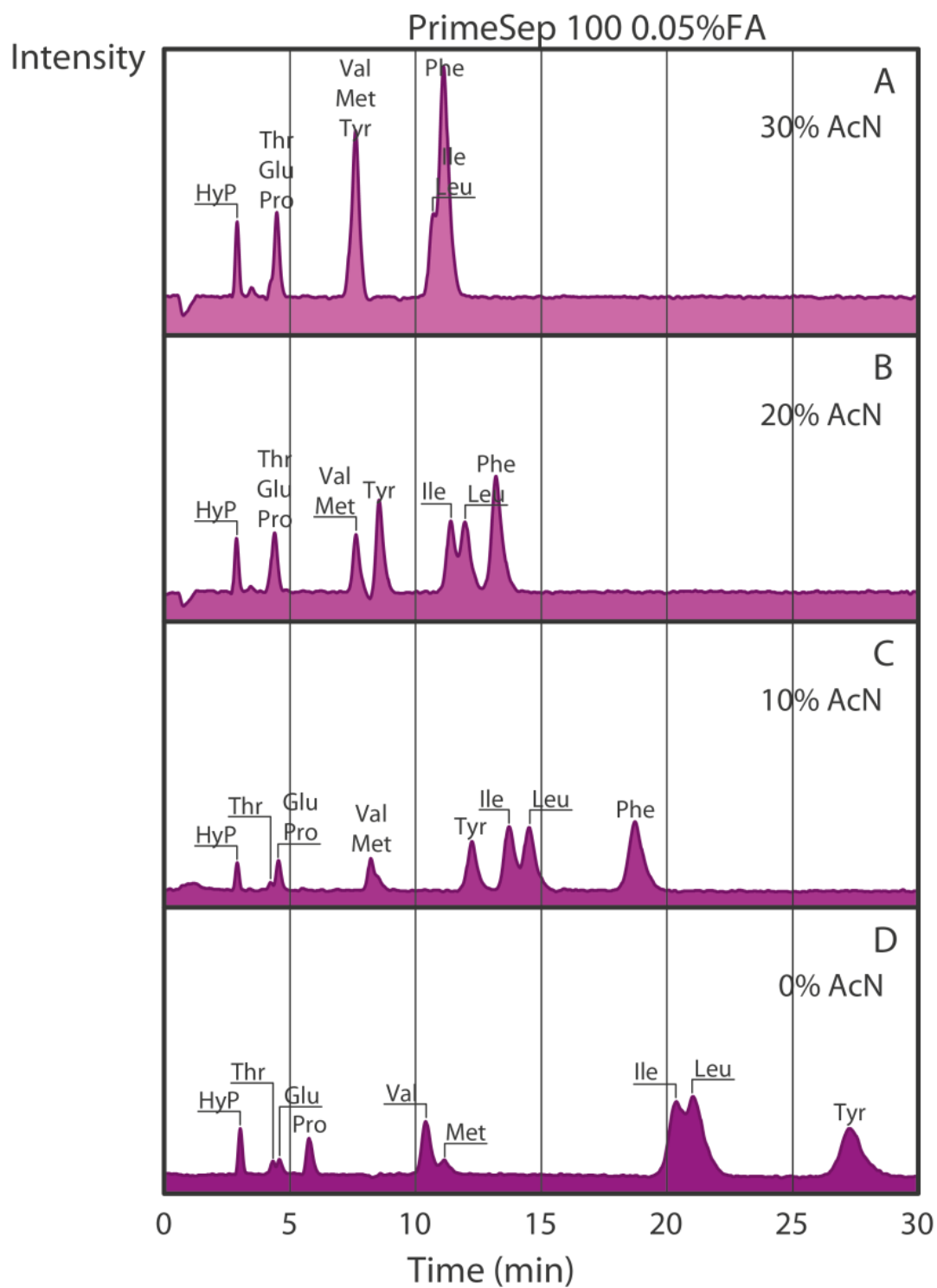


Figure 2-12. Chromatograms for PrimeSep 100 at different MeCN concentrations

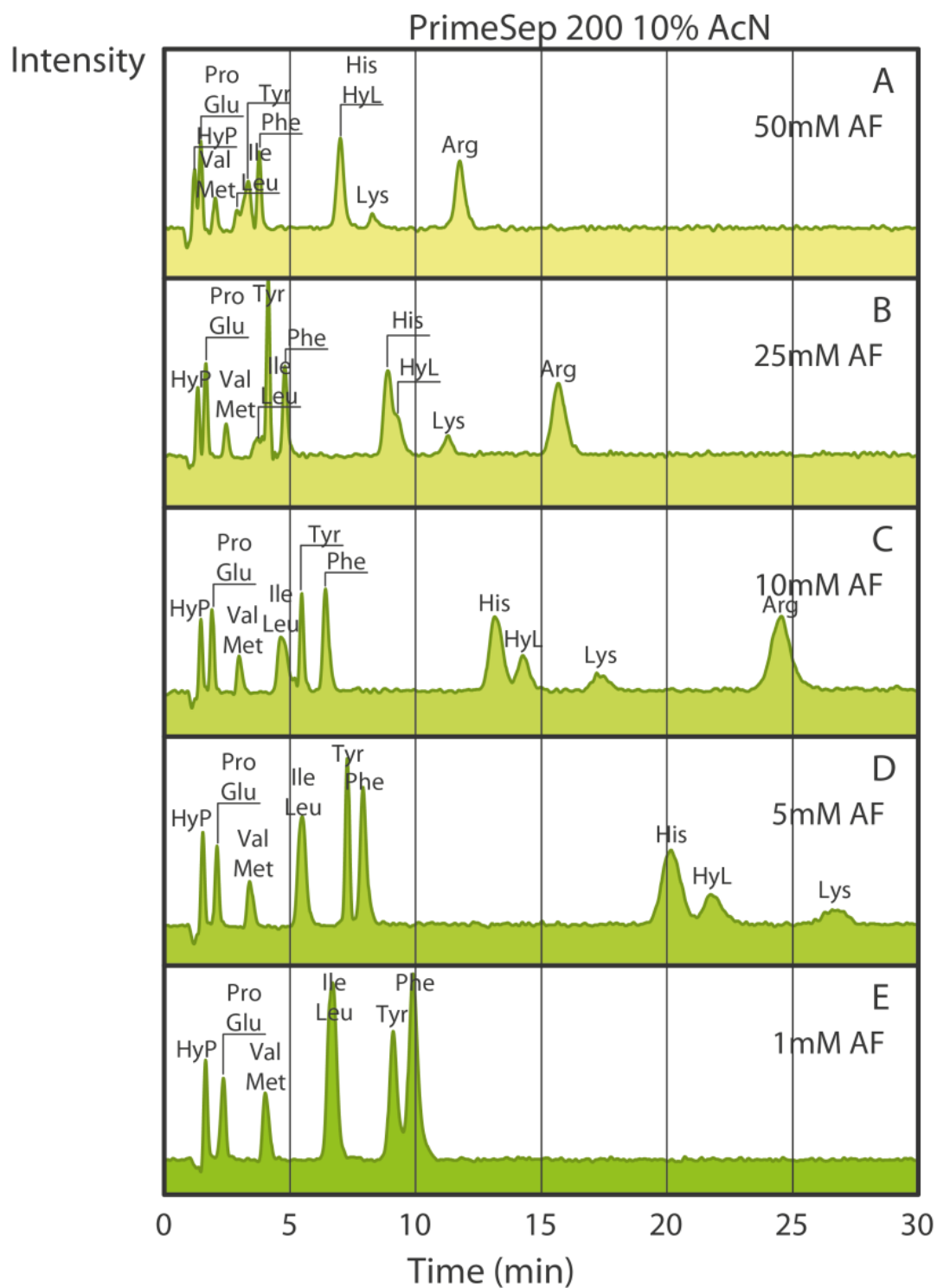


Figure 2-13. Chromatograms for PrimeSep 200 at different AF concentrations

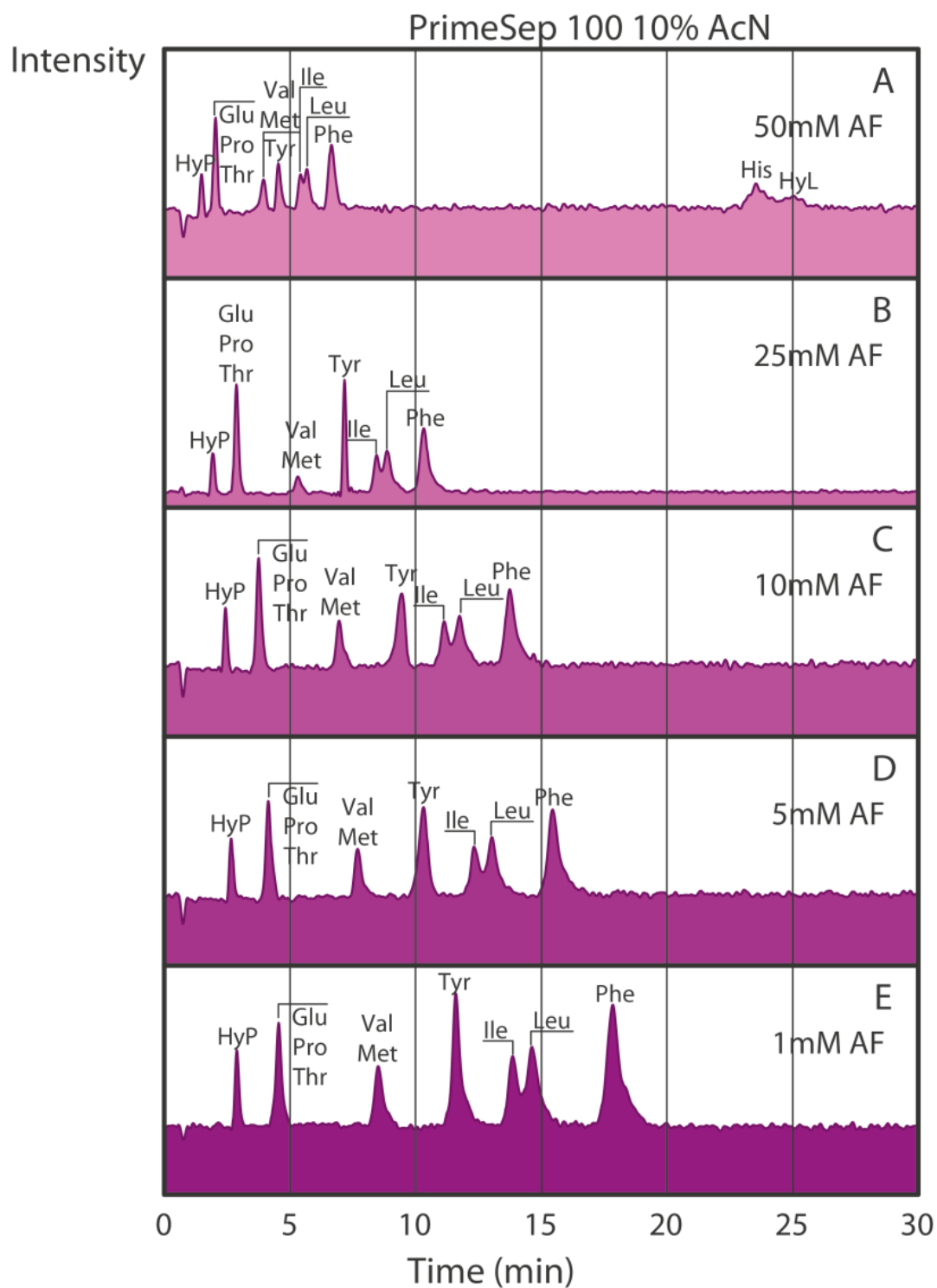


Figure 2-14. Chromatograms for PrimeSep 100 at different AF concentrations

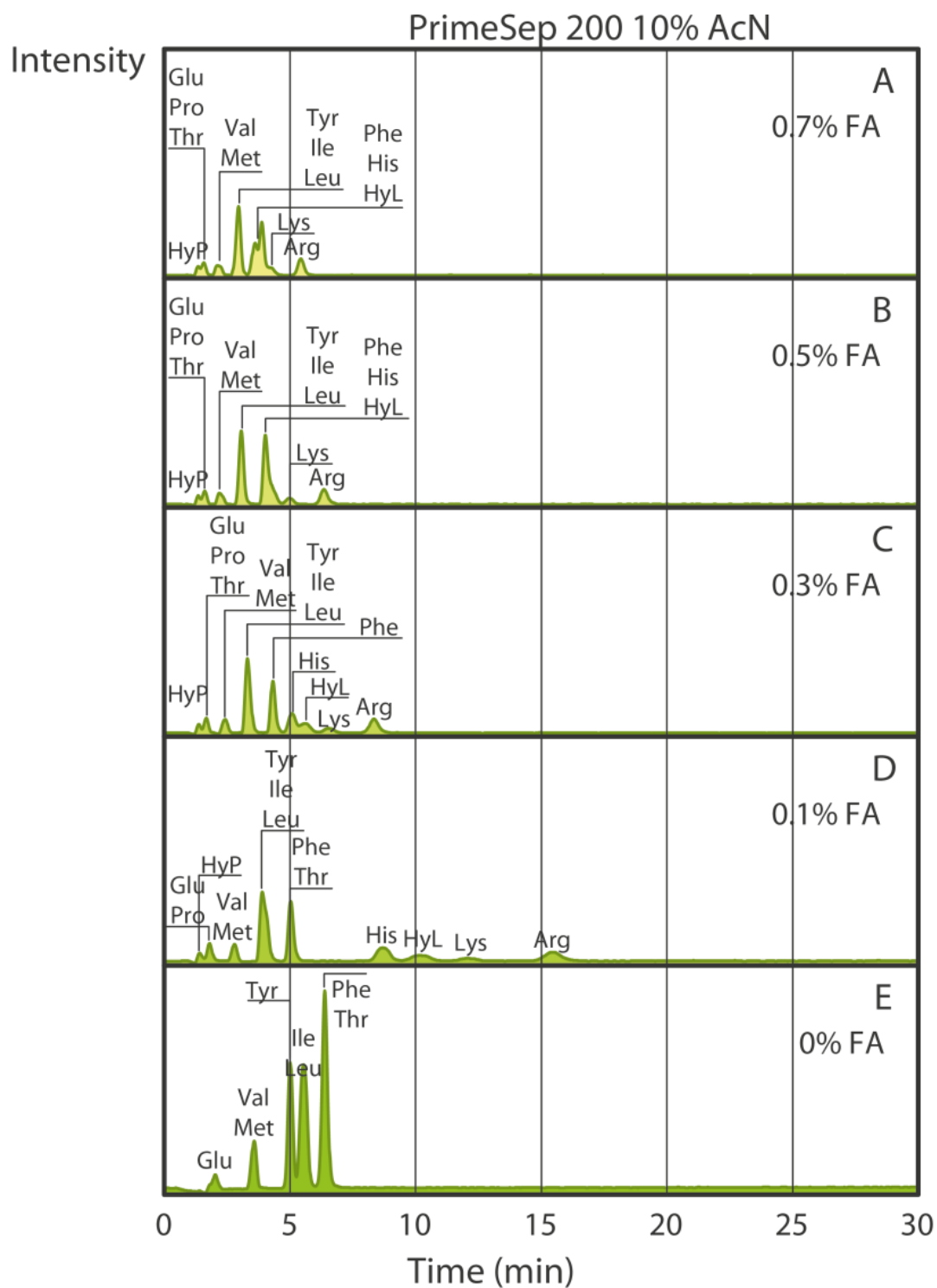


Figure 2-15. Chromatograms for PrimeSep 200 at different FA concentrations

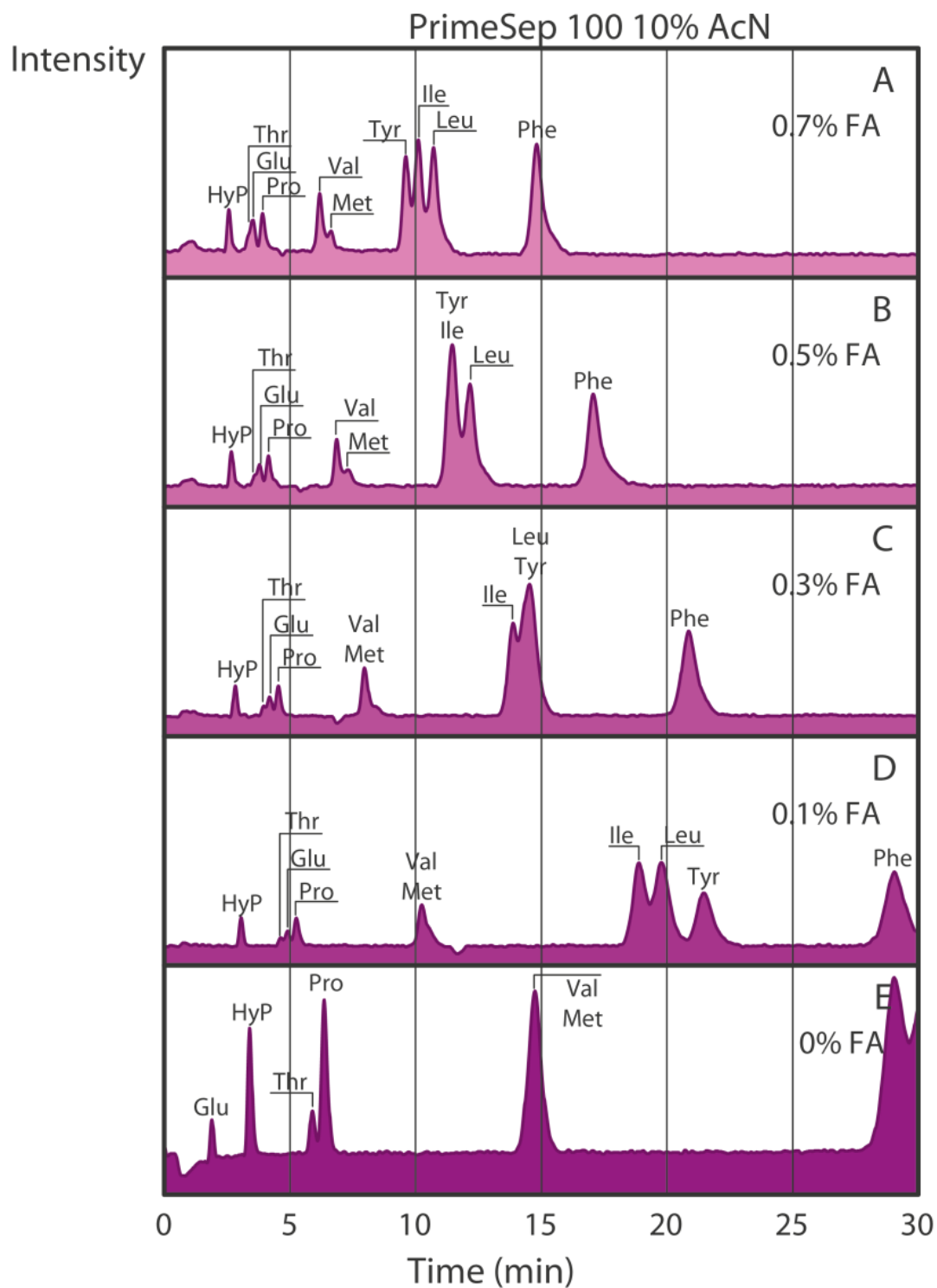


Figure 2-16. Chromatograms for PrimeSep 100 at different FA concentrations

As a result of the column testing for routine analysis of AAs, PrimeSep 200 was chosen for the routine analysis of AAs. The initial conditions were set at 10% MeCN, which allows fast elution of His, Lys and Arg. As a modifier FA rather than AF was chosen.

For MM chromatography, the peak width was found to be 0.3-0.6 min. The wider peaks are observed for AAs with several amino groups such as Lys, HyLys, Arg, His. PrimeSep 200 was shown to be suitable for separation of HyP diastereomers (Table 2-9). Leu and Ile were only partially resolved on PrimeSep 200.

Table 2-9. Separation of non-derivatised isobaric AAs using MM chromatography (PrimeSep 200)

m/z (Da)	Name	t_R	Name	t_R
89	<i>N</i> -MeGly	4.43	3Ala	9.45
103	diMeGly	4.37	2AIBu	6.18
	3AIBu	10.72	–	–
117	triMeGly	4.83	Val	7.13
119	Thr	4.5	HSer	4.7
131	3- <i>trans</i> -HyP	3.23	4- <i>trans</i> -HyP	3.54
	4- <i>cis</i> -HyP	4.07	3- <i>cis</i> -HyP	4.27
	Ile	9.89	Leu	10.09
162	5-HyL	16.91	Carnitine	12.13
169	3-MeHis	15.75	1-MeHis	17.41
174	Arg	20.39	<i>N,N</i> -diMeLys	21.96

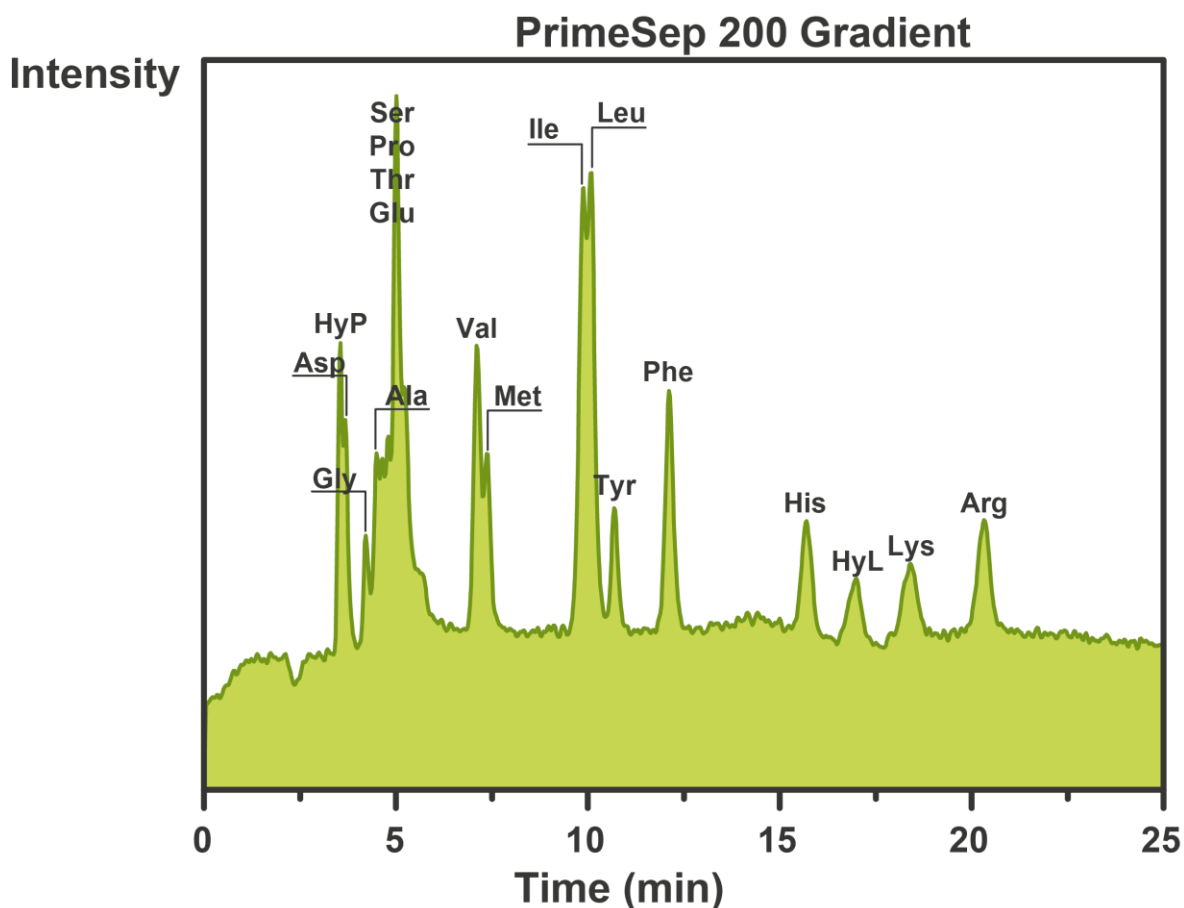


Figure 2-17. MM chromatogram of the non-derivatised 20 standard AAs

20 standard AAs separated on the PrimeSep 200

Injection of a mixture of proteinogenic AAs onto PrimeSep A (Figure 2-18 B) and PrimeSep 500 (Figure 2-18 A) columns showed that all the analytes do not elute in less than 30 min on PrimeSep A, and the standard AAs were shown to be weakly retained on PrimeSep 500 with a total run time at less than 5 min. For PrimeSep A TFA was used as the mobile phase modifying reagent; when FA was used all proteinogenic AAs were not eluted within 2 h.

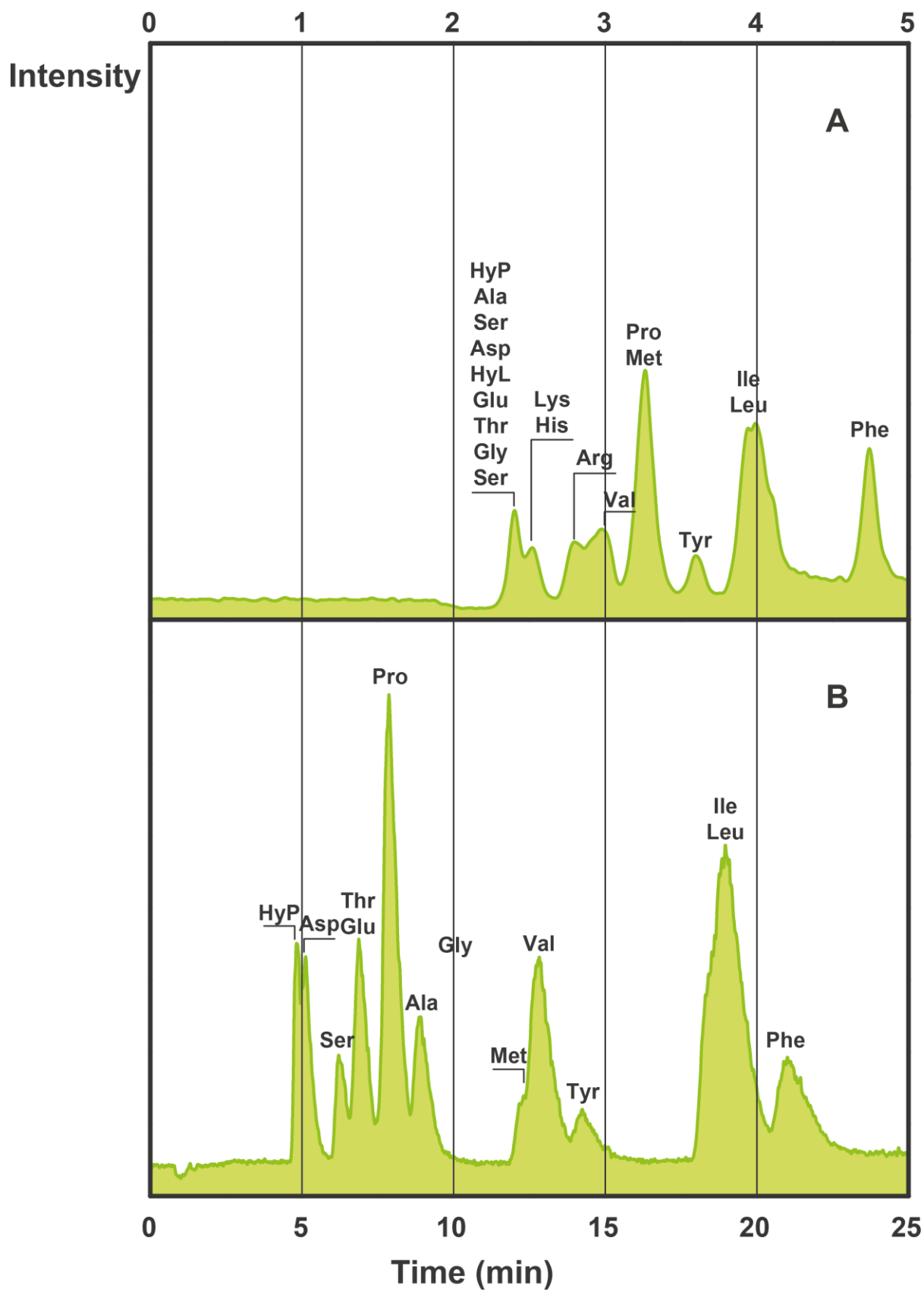


Figure 2-18. Chromatograms of proteinogenic AAs separated by MM chromatography

- A. Gradient 100 to 0% phase A in 25min, A (10% MeCN in H₂O, 0.05% FA), B (20% MeCN in H₂O, 0.2% FA)
- B. Gradient 80 to 0% phase A in 25min, A (10% MeCN in H₂O), B (20% MeCN in H₂O, 0.1% TFA)

2.3 Conclusions

Table 2-10 provides an overview of the performance differences between the three investigated chromatography methods. The total run time for the 20 standard AAs and resolution of the closely eluted pair Ile-Leu were used to assess the performance of the chromatographic system. Chromatographic peak width and chromatographic resolution were also determined.

Table 2-10. Comparison of the three investigated chromatography modes

RP = reversed phase chromatography of AQC-derivatised AAs; RP-IP = reversed phase ion pairing chromatography of non-derivatised AAs using TDFHA; MM = mixed mode chromatography of non-derivatised AAs.

Comparison Parameter	RP	IP	MM
MeCN Gradient (%)	0-75	0-27	10-20
Total Run Time (min)	15	6	20
Peak Width (min)	0.2	0.2	0.5
Chromatographic Resolution	75	30	40
$R_{\text{Ile/Leu}}$	1.2	2.6	0.6
Flow Rate (mL/min)	0.7	0.8	0.3
Injection Volume (μL)	1	1	2

RP chromatography with AQC-derivatised AAs proved to be a powerful tool for separation and analysis of structural isomers of AAs and of modified AA diastereomers (Table 2-5). The use of Tris buffer was shown to hinder the derivatisation of AAs present in minor concentrations. For AQC derivatisation it was determined that it is better to avoid buffers which contain compounds with free primary and secondary amino groups, or alcohols.

RP-IP chromatography was shown to separate the 20 standard AAs within 6 min. It was shown to provide high resolution for the pair Leu-Ile (Table 2-10) thus, demonstrating sensitivity towards structure of the AA sidechain. Separation of diastereomers of HyP (Table 2-7) demonstrates the sensitivity of this method to the changes in the polarity of AAs. Resolution of RP-IP chromatography can be adjusted using different concentrations of IP reagent, as well as by varying the length of the alkyl chain, and thus changing the mechanism

of the retention, which allows control of separation. It was not possible to separate the diastereomers of the modified AAs using RP-IP chromatography, except for the hydroxylated prolines. The method was also shown to compromise detection in negative ion mode.

MM chromatography of non-derivatised AAs was shown to be compatible with MS detection. MM chromatography requires longer separation time because of the ionic interaction between the stationary phase and analyte. MM chromatography using PrimeSep 200 does not provide base-line resolution for the pair of Leu-Ile and some other structural isomers (Table 2-9) and does not separate the diastereomers of modified AAs.

2.4 References

- (1) Vickery, H. B.; Schmidt, C. L. A. *Chemical Reviews* **1931**, 9, 169.
- (2) Foster, G. L. *Journal of Biological Chemistry* **1945**, 159, 431.
- (3) Martin, A. J. P.; Mittelman, R. *Biochemical Journal* **1948**, 43, 353.
- (4) Moore, S.; Stein, W. H. *Journal of Biological Chemistry* **1949**, 178, 53.
- (5) Hirs, C. H. W.; Moore, S.; Stein, W. H. *Journal of Biological Chemistry* **1952**, 195, 669.
- (6) Youngs, C. G. *Analytical Chemistry* **1959**, 31, 1019.
- (7) Nybom, N. *Physiologia Plantarum* **1964**, 17, 434.
- (8) Heathcot, J. G.; Haworth, C. *Journal of Chromatography* **1969**, 43, 84.
- (9) van Deemter, J. J.; Zuiderweg, F. J.; Klinkenberg, A. *Chemical Engineering Science* **1956**, 5, 271.
- (10) Smith, R. M. *Trac-Trends in Analytical Chemistry* **1984**, 3, 186.
- (11) Knox, J. H.; Hartwick, R. A. *Journal of Chromatography* **1981**, 204, 3.
- (12) Buszewski, B.; Gdzala-Kopciuch, R.; Michel, M. *Accreditation and Quality Assurance* **2011**, 16, 237.
- (13) Némethy, G. *Angewandte Chemie International Edition in English* **1967**, 6, 195.
- (14) Subirats, X.; Bosch, E.; Roses, M. *Journal of Chromatography A* **2004**, 1059, 33.
- (15) Kidera, A.; Konishi, Y.; Oka, M.; Ooi, T.; Scheraga, H. A. *Journal of Protein Chemistry* **1985**, 4, 23.
- (16) Henchoz, Y.; Schappler, J.; Geiser, L.; Prat, J.; Carrupt, P. A.; Veuthey, J. L. *Analytical and Bioanalytical Chemistry* **2007**, 389, 1869.
- (17) Vallat, P.; Gaillard, P.; Carrupt, P. A.; Tsai, R. S.; Testa, B. *Helvetica Chimica Acta* **1995**, 78, 471.
- (18) Amend, J. P.; Helgeson, H. C. *Pure and Applied Chemistry* **1997**, 69, 935.
- (19) Bosch, E.; Espinosa, S.; Roses, M. *Journal of Chromatography A* **1998**, 824, 137.
- (20) Subirats, X.; Bosch, E.; Roses, M. *Journal of Chromatography A* **2009**, 1216, 2491.
- (21) Jordan, F. *Journal of Physical Chemistry* **1973**, 77, 2681.

- (22) Goss, K. U. *Environmental Science & Technology* **2008**, 42, 456.
- (23) Teale, F. W. J.; Weber, G. *Biochemical Journal* **1957**, 65, 476.
- (24) Moore, S.; Stein, W. H. *Journal of Biological Chemistry* **1948**, 176, 367.
- (25) Saifer, A.; Oreskes, I. *Analytical Chemistry* **1956**, 28, 501.
- (26) Lindeberg, E. G. G. *Journal of Chromatography* **1976**, 117, 439.
- (27) Yokoyama, T.; Sakai, A.; Zenki, M. *Chromatographia* **2008**, 67, 535.
- (28) Marfey, P. *Carlsberg Research Communications* **1984**, 49, 591.
- (29) Sanger, F. *Biochemical Journal* **1945**, 39, 507.
- (30) Edman, P. *Acta Chemica Scandinavica* **1950**, 4, 283.
- (31) Heinrikson, R. L.; Meredith, S. C. *Analytical Biochemistry* **1984**, 136, 65.
- (32) Kinoshita, T.; Kasahara, Y.; Nimura, N. *Journal of Chromatography A* **1981**, 210, 77.
- (33) Negro, A.; Garbisa, S.; Gotte, L.; Spina, M. *Analytical Biochemistry* **1987**, 160, 39.
- (34) Lin, J. K.; Wang, C. H. *Clinical Chemistry* **1980**, 26, 579.
- (35) Butikofer, U.; Fuchs, D.; Bosset, J. O.; Gmur, W. *Chromatographia* **1991**, 31, 441.
- (36) Cohen, S. A.; Michaud, D. P. *Analytical Biochemistry* **1993**, 211, 279.
- (37) Shimbo, K.; Yahashi, A.; Hirayama, K.; Nakazawa, M.; Miyano, H. *Analytical Chemistry* **2009**, 81, 5172.
- (38) Johnson, D. W. *Journal of Chromatography B* **2011**, 879, 1345.
- (39) Cohen, S. A.; Deantonis, K. M. *Journal of Chromatography A* **1994**, 661, 25.
- (40) Li, G. D.; Krull, I. S.; Cohen, S. A. *Journal of Chromatography A* **1996**, 724, 147.
- (41) Pappa-Louisi, A.; Nikitas, P.; Agrafiotou, P.; Papageorgiou, A. *Analytica Chimica Acta* **2007**, 593, 92.
- (42) Liu, H. J.; Sanuda-Pena, M. C.; Harvey-White, J. D.; Kalra, S.; Cohen, S. A. *Journal of Chromatography A* **1998**, 828, 383.
- (43) Oreiro-Garcia, M. T.; Vazquez-Illanes, M. D.; Sierra-Paredes, G.; Sierra-Marcuno, G. *Biomedical Chromatography* **2005**, 19, 720.
- (44) Hernandez-Orte, P.; Ibarz, M. J.; Cacho, J.; Ferreira, V. *Chromatographia* **2003**, 58, 29.
- (45) Jorgensen, N. O. G.; Jensen, R. E. *Marine Chemistry* **1997**, 57, 287.
- (46) Reverter, M.; Lundh, T.; Lindberg, J. E. *Journal of Chromatography B* **1997**, 696, 1.
- (47) Ward, M.; Walker, J. M., Ed.; Humana Press: 1996, p 461.
- (48) Fiechter, G.; Mayer, H. K. *Journal of Chromatography B* **2011**, 879, 1353.
- (49) Motte, J. C.; Windey, R.; Delafortrie, A. *Journal of Chromatography A* **1996**, 728, 333.
- (50) Wittmer, D. P.; Nuessle, N. O.; Haney, W. G. *Analytical Chemistry* **1975**, 47, 1422.
- (51) Bidlingmeyer, B. A.; Deming, S. N.; Price, W. P.; Sachok, B.; Petrusek, M. *Journal of Chromatography* **1979**, 186, 419.
- (52) Deelder, R. S.; Linssen, H. A. J.; Konunenduk, A. P.; van de Venne, J. L. M. *Journal of Chromatography A* **1979**, 185, 241.
- (53) Seiler, N.; Knodgen, B. *Journal of Chromatography* **1985**, 341, 11.
- (54) Petritis, K. N.; Chaimbault, P.; Elfakir, C.; Dreux, M. *Journal of Chromatography A* **1999**, 833, 147.
- (55) van Leuken, R. G. J.; Kwakkenbos, G. T. C.; Duchateau, A. L. L. *Journal of Chromatography A* **1993**, 647, 131.

- (56) Garcia, M. C. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **2005**, 825, 111.
- (57) Holcapek, M.; Volna, K.; Jandera, P.; Kolarova, L.; Lemr, K.; Exner, M.; Cirkva, A. *Journal of Mass Spectrometry* **2004**, 39, 43.
- (58) de Person, M.; Chairnbault, P.; Elfakir, C. *Journal of Mass Spectrometry* **2008**, 43, 204.
- (59) Gustavsson, S. A.; Samskog, J.; Markides, K. E.; Langstrom, B. *Journal of Chromatography A* **2001**, 937, 41.
- (60) Thiele, B.; Fullner, K.; Stein, N.; Oldiges, M.; Kuhn, A. J.; Hofmann, D. *Analytical and Bioanalytical Chemistry* **2008**, 391, 2663.
- (61) Concha-Herrera, V.; Torres-Lapasio, J. R.; Vivo-Truyols, G.; Garcia-Alvarez-Coque, M. C. *Journal of Liquid Chromatography & Related Technologies* **2006**, 29, 2521.
- (62) Piraud, M.; Vianey-Saban, C.; Petritis, K.; Elfakir, C.; Steghens, J. P.; Bouchu, D. *Rapid Communications in Mass Spectrometry* **2005**, 19, 1587.
- (63) Petritis, K.; Chaimbault, P.; Elfakir, C.; Dreux, M. *Journal of Chromatography A* **2000**, 896, 253.
- (64) Zoppa, M.; Gallo, L.; Zacchello, F.; Giordano, G. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **2006**, 831, 267.
- (65) Chaimbault, P.; Petritis, K.; Elfakir, C.; Dreux, M. *Journal of Chromatography A* **1999**, 855, 191.
- (66) Armstrong, M.; Jonscher, K.; Reisdorph, N. A. *Rapid Communications in Mass Spectrometry* **2007**, 21, 2717.
- (67) Clifford, M. N.; Lopez, V.; Poquet, L.; Williamson, G.; Kuhnert, N. *Rapid Communications in Mass Spectrometry* **2007**, 21, 2014.
- (68) McCullagh, J. S. O.; Juchelka, D.; Hedges, R. E. M. *Rapid Communications in Mass Spectrometry* **2006**, 20, 2761.
- (69) McCullagh, J.; Gaye-Siessegger, J.; Focken, U. *Rapid Communications in Mass Spectrometry* **2008**, 22, 1817.
- (70) McCullagh, J. S. O. *Rapid Communications in Mass Spectrometry* **2010**, 24, 483.
- (71) Orlovsky, V., Information Request.
- (72) Jemal, M.; Hawthorne, D. J. *Rapid Communications in Mass Spectrometry* **1999**, 13, 61.

3 Application of RI to the analysis of PTMs

This chapter explores the application of the retention indices (RIs) to the identification of the modified AAs. The chapter presents the systematic analysis of the retention data of structural analogues (homologues and isomers) of AAs and establishes the structure-retention relationship. It describes how the retention time (t_R) of a modified AA can be estimated based on its chemical structure and how the estimated t_R can be used for the identification of the modified AAs in the absence of commercially available standards.

3.1 Introduction of the RI Scale

The concept of RIs was first introduced by Kovats in 1958¹. Later (1965), it was shown by Kovats that there is a strong relationship between RIs and the chemical structure of analytes².

Another application to the newly introduced scale was suggested by W.O. McReynolds³, who used the RIs of a set of compounds different in their chemical nature in order to compare different columns. For a compound tested on one column, he managed to estimate the t_R on other columns for which only the RIs of standards were known.

It was not until 1979 when the concept of RIs was adapted for LC⁴. It was primarily introduced for the analysis of drugs and narcotics. It can be speculated that, the compounds which are qualified as such cannot be found easily as standards, which led to the necessity to develop an analytical method which enables identification without standards.

Following the trend set by Kovats in gas chromatography, an homologous series of compounds were selected as reference compounds such as α -ketones⁴. Later, RI scales based on different homologous series such as alkyl aryl ketones⁵ and nitroalkanes⁶ were proposed. The introduction of alternative reference compounds was caused by the necessity to ensure that a strong chromophore group is present in order to enable UV detection, most commonly applied at that time. For the analysis of a wide range of a substituted and non-substituted

aromatic hydrocarbons, aromatic compounds – benzene (RI = 10), naphthalene (RI = 100) and phenanthrene (RI = 1000) – were suggested as reference compounds⁷. Elliott *et al.* (1996) based their reference scale on the drugs – paracetamol (RI = 100), aprobarbitone (RI = 200), demoxipam (RI = 300), triazolam (RI = 400), and lorazepam benzophenone (RI = 500)⁸.

The RIs are arbitrarily assigned to a set of reference compounds and often the RI values are taken to be proportional to the number of carbon atoms in the backbone of reference compounds such as alkanes (GC, Kovats)¹, alkan-2-ones (LC, Baker)⁴ or alkyl aryl ketones (LC, Smith)⁹. As can be seen from the reference scales suggested by Popl and Elliott⁸, this is not always the case.

The obvious advantage of introducing a scale, in which any two consequent members of the series differ by the same structural fragment – be that a methylene group or the fragment of benzene ring – is the possibility to trace correlation between structure and RI.

The use of RIs in LC has had less application compared to gas chromatography. One reason for this may be the absence of an agreement between the authors on a unified scale. A single scale makes no allowance for different mechanisms of retention. Therefore, the broad spectrum of types of LC – as compared to gas chromatography – cannot be described using a single unified RI scale.

The criteria for the RI scale have been clearly defined⁹. As a major requirement, a suitable set of reference compounds needs to be detectable at 254 nm¹⁰. Due to the presence of strong chromophore, the homologous series of alkyl aryl ketones meets this requirement. This was the only improvement, or facilitation, in comparison to the previously suggested alkan-2-ones. These two systems have been compared on several occasions^{11,12}, but preference was not given to either of the systems, and both are still used in parallel.

Smith (1990) suggested that “the reason for selecting a homologous series of standard compounds for RI scale is that they are usually systematically spaced throughout the chromatogram”. Initially, Smith suggested that reference compounds should be generally available and cover a wide range of polarities¹³. A more elaborated system has been suggested by Popl, who used structural analogues of aromatic compounds⁷.

3.1.1 RIs Prediction

The prediction of the RI of a compound can either be based on correlation between its retention and its physical properties, such as octanol-water partitioning coefficients, a method suggested by Baker¹⁴, or based on correlation between structure and retention, as suggested by Kovats for GC¹.

Popl attempted to predict the RIs for polycondensed aromatic compounds based on his scale⁷. Elliott introduced such scale “to produce a stable linear retention scale during the gradient elution”⁸.

3.1.2 Revision of the RI Concept

A plethora of LC modes (stationary phases and mobile phases) are available and the different processes involved in chromatographic separation including partitioning, adsorption, hydrophobic interaction, ionic interactions *etc.* make the introduction of single unified system of RIs challenging because each may have different effects on the retention of compounds.

The development of a RI scale in work for this thesis was based on structural analogies between the reference compounds and the analytes. As previously suggested, a set of reference compounds should be that of a homologous series, or a set of compounds with the same structural increment throughout the series.

3.1.3 RI Scale for AA Analysis

In the work for this thesis the RIs are defined as follows^{1,15}

$$RI = 100 \cdot z + 100 \cdot \frac{\log t_{R,x} - \log t_{R,z}}{\log t_{R,z+1} - \log t_{R,z}}$$

t_R corresponds to the retention time of the analyte or the reference compounds. The subscript x refers to the analyte; subscript z refers to the number of carbon atoms in the reference compound eluting immediately before the analyte, and $z+1$ is the number of carbon atoms in the reference compound eluting immediately after the analyte.

The t_R of an AA depends on its hydrophobicity and the degree of ionisation of its polar groups¹⁶. It is impossible to account for the degree of ionisation of AAs in the chromatographic system. However, using a homologous series of α -AAs with an increasing length of the alkyl-groups as reference compounds enables calibration of the chromatographic system both in terms of hydrophobicity and structure. Elongation of the aliphatic group is expected to result in an increase in hydrophobicity in the series, and growth of the aliphatic chain by one carbon atom is linked proportionally to the RI scale.

The homologous series of α -AAs (Table 3-1) was used as reference compounds to calibrate the RI scale. It can be seen (Table 3-1) that for aliphatic acids and amines, the elongation of the aliphatic group does not affect the pKa of the compound. Based on the comparison with aliphatic amines and carboxylic acids, the pKa values of the α -AAs can be assumed to remain constant with increasing length of the aliphatic group. Therefore, the introduction of such a scale allows us to eliminate the effect of the carboxyl-amino group and calibrate the system in terms of hydrophobicity.

Table 3-1. Introduction of the RIs scale based on α -AAspK_a data are adapted from Williams collection¹⁷

C – number of carbon atoms.

# C	Acids pK _a	Amines pK _a	Name	AAs		RI
				pK _a	pK _a	
2	4.76	10.70	Gly	2.34	9.60	200
3	4.86	10.60	Ala	2.35	9.69	300
4	4.83	10.77	α -ABu	2.55	9.60	400
5	4.84	10.63	NVa	–	–	500
6	4.85	10.56	NLe	–	–	600
7	4.89	10.67	α -AHe	–	–	700
8	4.89	10.65	α -AOc	–	–	800

For the three types of chromatography tested (RP (44 compounds), RP-IP (36 compounds) and MM (42 compounds)), the retention data acquired were converted to RIs. The results were averaged over 5 replicates (data are presented in Appendix B).

3.2 Retention Pattern Recognition

3.2.1 Comparison of Different Chromatography Modes

The retention times of the reference compounds were first measured for AQC-derivatised AAs using RP chromatography and for non-derivatised AAs using RP-IP chromatography and MM chromatography. In order to analyse and compare the three chromatography modes, the retention times of reference compounds were plotted against the RIs of reference compounds. Both RP chromatography of AQC-derivatised AAs (Figure 3-1) and MM chromatography (Figure 3-3) revealed a logarithmically escalating relationship between t_R and RI, which has been previously described by a number of authors^{4,12,18}. For IP chromatography, an s-shaped relationship between t_R and RI was found, which has not been reported before (Figure 3-2).

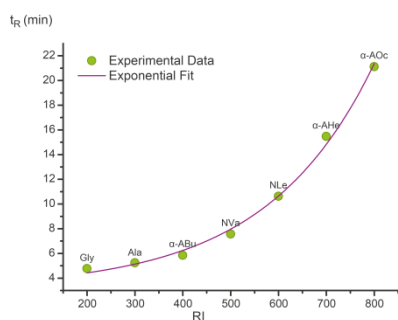


Figure 3-1. t_R vs. RIs dependence for RP-MM chromatography (PrimeSep 200)

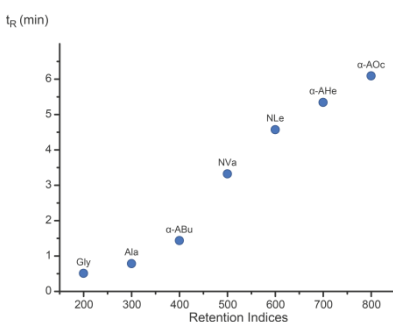


Figure 3-2. t_R vs. RIs dependence for RP-IP chromatography (TDFHA)

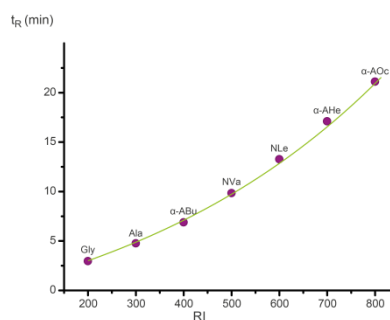


Figure 3-3. t_R vs. RIs dependence for RP chromatography of AQC-derivatised AAs

It is suggested that the s-shaped curve in the case of RP-IP chromatography (Figure 3-2) is the result of the change of the mechanism of absorption from ion pair formation to the dynamic ion exchange. It has been pointed out that retention controlled by ion pair formation can be enhanced by elongation of the alkyl group of the heteron¹⁹. However, the opposite should be true as well, and increase of the alkyl group of the analyte should increase the retention, because it increases the length, and hence hydrophobicity of the ion pair. It was observed that the t_R increases rapidly up to norvaline (NVa, Figure 3-2), and after this point, the trend changes to a slower escalation. This suggests that the second part of the curve may correspond to a dynamic ion exchange mechanism, which is less dependent on the chain length. The positive trend remaining in this part of the curve suggests that the competitive process of direct absorption of the analyte depends on hydrophobicity and increases with alkyl group length.

3.2.1.1 Gradient invariability

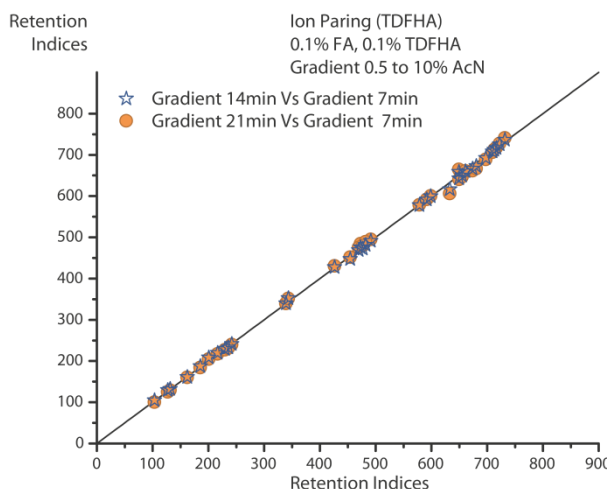


Figure 3-4. Comparison of different gradients for RP-IP chromatography; straight line is for $y = x$

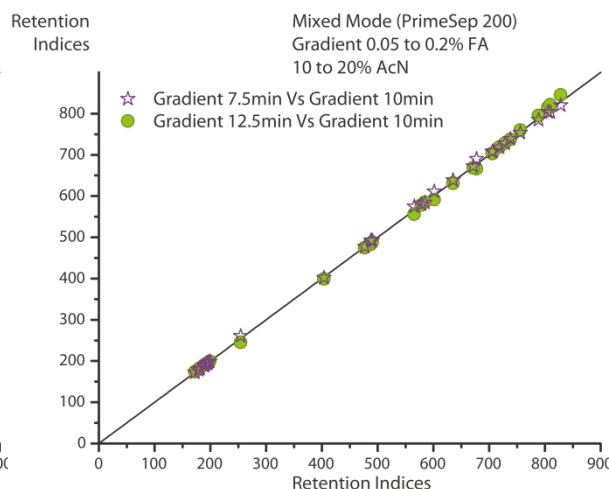


Figure 3-5. Comparison of different gradients for the RP-MM chromatography; straight line is for $y = x$

For RP-IP chromatography (TDFHA, Figure 3-4) and for MM chromatography (PrimeSep 200 column, Figure 3-5), three gradients were compared. The RIs of the test compounds for the shallower gradient and a steeper gradient have been plotted against the basic gradient. It can be seen that in both cases the data points slightly deviate from the straight line at 45° ($y = x$). Points belonging to the straight line at 45° ($y = x$) demonstrate the equality of the RIs for the different gradients. The variation of RI values for different gradients is comparable to the variation of the measurements for the same gradient at different runs.

Thus, RIs – as compared to t_R – are independent of the gradient for both IP and MM chromatography. The observed trend can be extrapolated for RP chromatography as well. Invariability of the RI scale allows for comparison of the retention data acquired under different elution gradients. Whereas it is impossible to compare the t_R under different gradient conditions, the RI are comparable, irrespective of the gradient.

3.2.1.2 Comparison of IP reagents

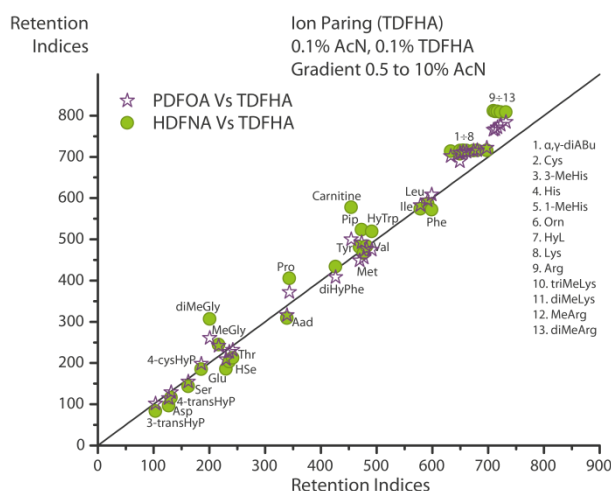


Figure 3-6. Comparison of different IP reagents for RP-IP chromatography; straight line is for $y = x$

The RIs of AAs have been measured for three different IP reagents varying in the number of carbon atoms in the alkyl chain: TDFHA, PDFOA, and HDFNA. The same molar concentration (2 mM solution) has been used in all experiments.

It has been shown previously²⁰ that the molar quantity of IP reagent absorbed by the stationary phase increases with an increase in the length of the alkyl chain. As a result, longer t_R is observed. It can be seen from the plot (Figure 3-6) that in general the green dots corresponding to the longer HDFNA are above the star-shaped dots, corresponding to the PDFOA, which means that not only absolute retention is increasing but also the relative retention in comparison to the reference compounds.

It can be seen that resolution of longer retained AAs (Figure 3-6, from α,γ -diABu to diMeArg) decreases with an increase in the alkyl chain length of the IP reagents. This might be caused by the change in the analyte absorption mechanism of the AAs eluted before NVa (Figure 3-2, NVa – RI = 500); and for the longer retained AAs the ionic interaction with the absorbed IP reagent dominates hydrophobic interaction. As a result, there is no separation of

the later eluted AAs using long chain perfluorinated surfactants, whereas with the shorter alkyl chain length the mechanism of AA separation may revert back to hydrophobicity interaction, thus enhancing the separation of AAs. Thus, even greater resolution of late eluted AAs can be expected when using UDFHA rather than for TDFHA. Unfortunately, this experiment could not be carried out because the use of these IP reagents became prohibitive in the lab due to their contamination of the MS system, particularly for the subsequent use in the analysis in the negative ion mode.

3.2.1.3 Comparison of different chromatographic modes

Three types of chromatography employed in this work were compared against each other by plotting the AA RIs measured for one mode of chromatography against the RIs measured for another mode (Figure 3-7, Figure 3-8, Figure 3-9, Figure 3-10).

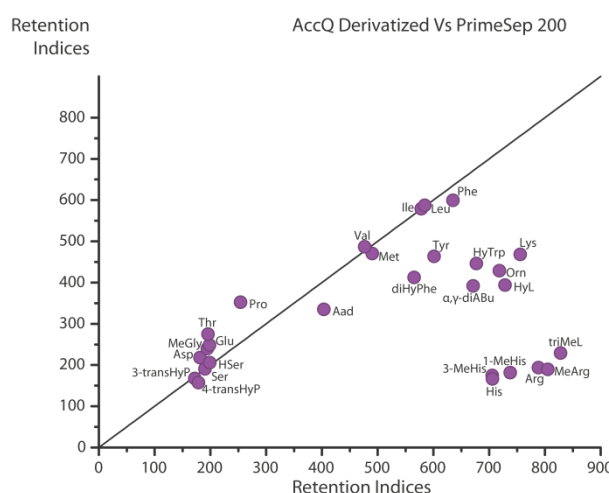


Figure 3-7. Comparison of RP chromatography of derivatised AAs and RP MM chromatography; straight line is for $y = x$

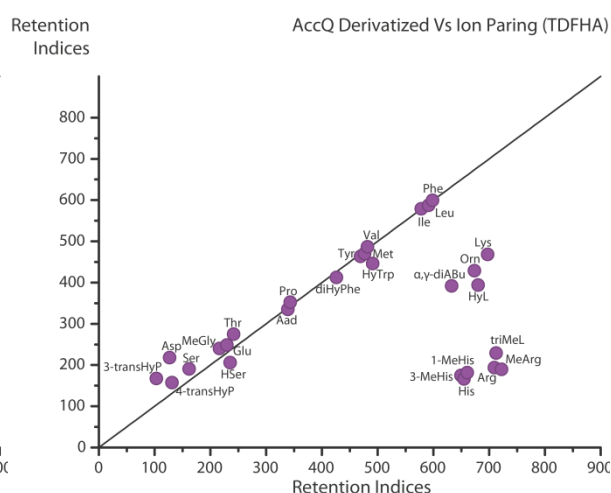


Figure 3-8. Comparison of RP chromatography of derivatised AAs and RP-IP chromatography; straight line is for $y = x$

The t_R using RP chromatography of AQC-derivatised AAs was compared with the t_R from MM chromatography of non-derivatised AAs. It can be seen that these two methods correspond to one another for the AAs with non-polar aliphatic or aromatic side chains (Figure 3-7). There is a significant difference in the retention of AAs which have a hydroxyl

group, and the trend is even stronger for aromatic AAs. Even greater difference in retention is observed for the AAs with additional amino groups such as Lys and Orn. Such differences can be explained by the fact that for RP chromatography the presence of the amino group reduces lypophilicity, thus reducing retention of the compound, whereas in the case of MM chromatography the presence of the additional amino group enables interaction of this group with the stationary phase, which increases the t_R .

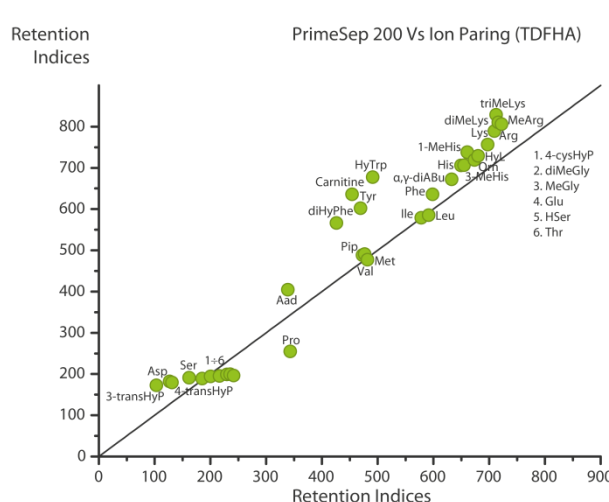


Figure 3-9. Comparison of RP-MM chromatography vs. RP-IP chromatography; straight line is for $y = x$

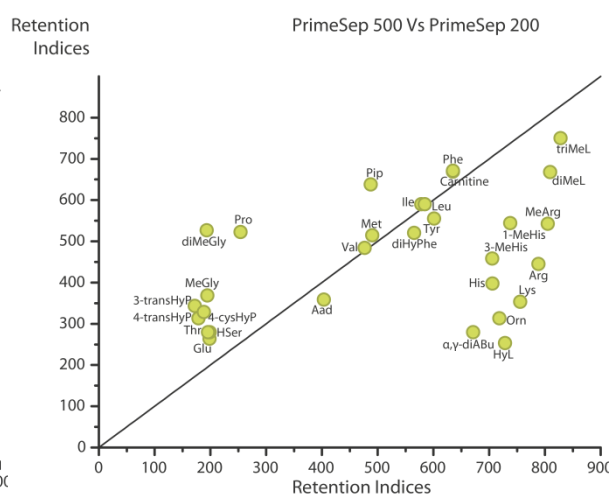


Figure 3-10. Comparison of two RP-MM chromatography columns PrimeSep 200 and PrimeSep 500; straight line is for $y = x$

A closer resemblance in RI is observed between RP chromatography and IP chromatography (Figure 3-8). Significant differences between the two types of chromatography are observed only when the second amino group is present.

It is worth pointing out that in both cases the RI of α,γ -diABu, Orn, Lys, and HyL form a separate cluster from RI of Arg, MeArg, His, 1-MeHis, 3-MeHis, MeLys, diMeLys and triMeLys. These differences result from the fact that the compounds of the first cluster are AQC-bis-derivatised, and therefore, they are, retained longer.

It can be seen (Figure 3-9) that the weakly retained AAs are hardly separated by MM chromatography and demonstrate much better resolution when separated in IP mode. The

homologues of α,γ -diABu, Orn, and Lys form a straight line nearly parallel to the straight line at 45°. This indicates that in both chromatography types there is a constant increment on the RIs scale for the second amino group added.

3.2.2 Analysis of Retention Patterns of Modified AAs

The increase in RI is associated with the increase in hydrophobicity. 100 units on the RI scale are proportional to an increase of the carbon backbone by one carbon atom (Table 3-1). In order to identify the structure-retention relationships, the RIs of AAs were compared to both the reference compound with the same number of the carbon atoms in the carbon backbone and the closest structural analogue amongst AAs available as standards. The AAs which form the homologous series were considered, and the effect of introduction of a particular functional group, such as hydroxyl, methyl or amino group was assessed.

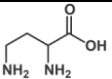
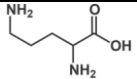
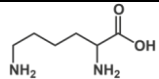
3.2.2.1 Homologous Series Analysis

The expected pattern for the homologous series is an increase by 100 RI units for every carbon atom added. Deviation from the expected pattern is observed for AAs with two amino groups (Table 3-2), AAs with two carboxyl groups (Table 3-3) and AAs with a hydroxyl group which can form an intra-molecular hydrogen bond (Table 3-5).

The first group of homologues investigated was the di-amino acids (di-AAAs), with one amino group at the α -carbon and a second one at the terminal carbon atom. As shown in Table 3-2, introduction of the second amino group does not result in a constant RI increment, as compared to the reference compounds. Di-AAAs do not follow the pattern expected for homologous series. However, di-AAAs form a different pattern (increase by 39 ± 9 RI units). The pattern observed is the result of the system being calibrated for a single amino group present in the compound. This pattern can also be clearly seen when different types of chromatography are compared (Figure 3-7 and Figure 3-8). It can be seen that α,γ -diABu,

Orn and Lys form a line parallel to the straight line at 45°. This trend can be expanded for the next members of the di-amino homologous series.

Table 3-2. di-AAs RIs Comparison

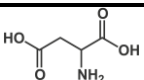
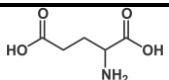
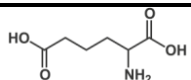
	α -ABut RI=400  α,γ -diABu		NVa RI=500  Orn		NLe RI=600  Lys	
	RI	Δ RI	RI	Δ RI	RI	
AQC-Derivatised	390	35	425	39	464	
IP	619	48	667	26	693	
MM	671	47	718	38	756	

Next, the group of dicarboxyl-AAs was studied. Analogous to the di-AAs, deviation from the pattern observed for the reference compounds might be expected. In the case of the AQC-RP chromatography $RI_{Glu} - RI_{Asp} = 30$ and $RI_{AAd} - RI_{Glu} = 87$. The observation can be explained by the formation of the intra-molecular hydrogen bonds in Asp. Such hydrogen bonding results in the formation of a six-membered ring system, which decreases the ionisation of the compound, hence increasing the t_R . Therefore, for this type of chromatography, the very high t_R of Asp is an outlier in the observed pattern; and for the following members of the homologous series it should be expected that the increment is about 100 RI units.

Such a scenario is not observed for IP chromatography, where the increment remains the same and corresponds to the retention pattern expected. For IP chromatography, it is the ion pair which is retained, rather than the individual AA. The pK_a of the carboxylic group of Asp (3.7) is higher than that of TDFHA (-0.5). In addition to this, the concentration of the IP reagent is about 100 times higher than that of the AA. Statistically, it is more likely to form the ion pair with the IP reagent, of which its ionised form is present in solution in a thousand times higher concentration. Therefore, the equilibrium of the system is shifted from intra-molecular hydrogen bond formation towards ion pair formation, thus decreasing the polarity of the carboxyl group less than expected.

It is more difficult to explain the observed deviation for MM chromatography. In this case, apparently two competitive processes have to be taken into consideration – intra-molecular hydrogen bond formation causing a decrease in polarity, and the repulsion of the anionic groups of the stationary phase and carboxylic groups. As ionic forces (250-1050 kJ/mol) are much stronger than hydrogen bonds (20-30 kJ/mol) and much stronger than hydrophobic forces. It can be suggested that, for Asp and Glu, retention is significantly affected by weak hydrophobic interaction between the analytes and the stationary phase. Ion interaction between the anionic group of the stationary phase is hindered because of the intra-molecular hydrogen bond formation and majorly because of repulsion between the two carboxylic groups of the AA – blocking the amino group – and the anionic group of the stationary phase. The increase in chain length may make the amino group more accessible, which enhances its interaction with the polar group of the stationary phase, thus increasing t_R .

Table 3-3. Dicarboxyl-AAs RIs Comparison

	α -ABut RI=400  Asp		NVa RI=500  Glu		NLe RI=600  AAd	
	RI	Δ RI	RI	Δ RI	RI	
AQC-Derivatised	218	30	248	87	335	
IP	127	102	229	110	339	
MM	107	86	193	209	402	

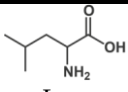
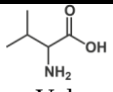
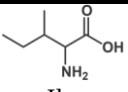
The next set of compounds investigated was the closely related branched alkyl chain AAs (Table 3-4). It is worth pointing out two trends for such compounds. Firstly, the RIs for branched alkyl chain compounds are very similar for all types of chromatography. The same RI scale has been used for all three types of chromatography. Therefore, the RI of the branched alkyl chain compounds will be invariant of the chromatography type.

Another important aspect is that Ile is always eluted before Leu. This can be explained by looking at their pK_a values: Ile (2.39) and Leu (2.43). It becomes less important for MM

chromatography, because that retention is here mostly affected by ionic interaction rather than hydrophobicity. As a result, the two compounds are poorly resolved on MM columns.

It is not possible to translate the pK_a values onto the RI scale. All three AAs have very similar pK_a values: Ile (2.39 ± 0.07), Leu (2.43 ± 0.05), and Val (2.42 ± 0.15) – with the variance taken into account, these values are identical.

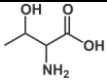
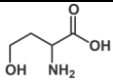
Table 3-4. Valine Homologues RIs Comparison

	NVa RI=500		α -ABut RI=400		NVa RI=500	
						
	Leu		Val		Ile	
	RI	Δ RI	RI	Δ RI	RI	
AQC Derivatized	587	-101	486	92	578	
IP	592	-108	484	94	578	
MM	585	-109	476	103	579	

The next group of compounds studied has a hydroxyl group separated from the carboxyl group by one or several carbon atoms. In the case of RP chromatography of derivatised AAs, for the Ser-Thr pair, an increment close to 100 is observed. One might expect the same difference for the pair of Ser-HSer, which, however, proves to be much lower. In case of the Ser-Thr pair, hydrophobicity of the molecule increases with elongation of the alkyl chain. In the case of the Ser-HSer pair, increase in hydrophobicity is nullified by intra-molecular interactions, and one extra carbon atom is not enough to completely remove those interactions. For the next member of this homologous series, a rapid increase in RI value is expected, since two carbon atoms are enough for isolation of the polar groups. For IP chromatography, in which case intramolecular interaction through hydrogen bonding is suppressed by the presence of the IP reagent, it can be seen that the increase in the chain length has a similar effect, irrespectively of the position of the extra carbon atom. In the case of MM chromatography, the increasing accessibility of the amino group plays a predominant

role in comparison to the elongation of the short alkyl group, and the retention increases by a greater extent when the amino group is spaced from the hydroxyl group.

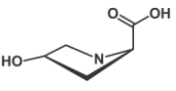
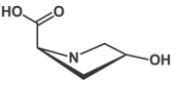
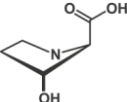
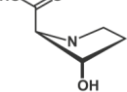
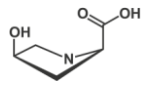
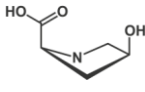
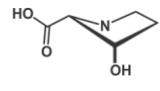
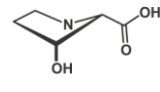
Table 3-5. Serine Homologues RIs Comparison

	α -ABut RI=400  Thr		Ala RI=300  Ser		α -ABut RI=400  HSer
	RI	Δ RI	RI	Δ RI	RI
AQC-Derivatised	275	-89	186	19	205
IP	241	-80	161	72	233
MM	179	-26	153	42	195

3.2.2.2 Hydroxyproline Isomers

For hydroxyprolines, there are two positional isomers that have so far been found in nature, 3- and 4-hydroxyprolines. Each of these can have two pairs of diastereomers. It is possible to differentiate between the diastereomers chromatographically (Chapter Five) because they have two polar groups – a hydroxyl group and a carboxyl group – either on the same side or on opposite sides relative to the plane of the ring. When both groups are positioned on the same side, formation of an intra-molecular hydrogen bond is possible, which reduces polarity of the compound. It is difficult to explain though why the effect is much stronger in case of RP and MM chromatography than in the case of IP chromatography.

Table 3-6. Hydroxyproline Isomers RIs Comparison

	 R,S-4-HyP  S,R-4-HyP	 R,R-3-HyP  S,S-3-HyP	 R,R-4-HyP  S,S-4-HyP	 R,S-3-HyP  S,R-3-HyP
AQC-Derivatised	138	151	247	245
IP	131	103	186	N/A
MM	92	59	143	160

3.2.2.3 Hydroxylated AAs

For RP chromatography, introduction of a hydroxyl group into the aliphatic chain, so that it is isolated from the other polar groups, results in a decrease of the RI by 73 ± 29 units. In the case MM chromatography, the decrease in RI was by 39 ± 37 . For the IP reagent, the expected decrease in the RI value should be 30 ± 10 units of RI, but this should be confirmed by a larger set of data.

Table 3-7. Influence of a Hydroxyl Group in AAs (Ala and Ser)

			
	Ala		Ser
	RI	Δ RI	RI
AQC-Derivatised	300	-114	186
IP	300	-139	161
MM	300	-147	153

Table 3-8. Influence of a Hydroxyl Group in AAs (Lys and 5-HyL)

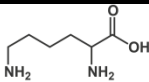
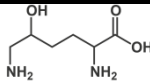
			
	Lys		5-HyL
	RI	Δ RI	RI
AQC-Derivatised	468	-78	390
IP	693	-20	673
MM	756	-28	728

Table 3-9. Influence of a Hydroxyl Group in AAs (His and 3-HyHis)

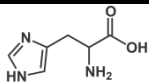
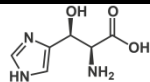
			
	His		3-HyHis
	RI	Δ RI	RI
AQC-Derivatised	150	-85	65

Table 3-10. Influence of a Hydroxyl Group in AAs (Arg and 3-HyArg)

	Arg		3-HyArg
	RI	Δ RI	RI
AQC-Derivatised	190	-55	135
MM	790	-49	741

Table 3-11. Influence of a Hydroxyl Group in AAs (Trp and HyTrp)

	Trp		HyTrp
	RI	Δ RI	RI
AQC-Derivatised	620	-174	446
MM	755	-77	678

Table 3-12. Influence of the Hydroxy Group Introduction (Phe, Tyr, diHyPhe)

	Phe		Tyr		diHyPhe
	RI	Δ RI	RI	Δ RI	RI
AQC-Derivatised	599	-136	463	-51	412
IP	600	-129	471	-43	428
MM	635	-34	601	-36	565

3.2.2.4 N-Methylated AAs

Another common modification of AAs is methylation. Unfortunately, the set of data of the methylated compounds tested is very small and varied in terms of structure, which makes it difficult to make precise predictions.

Generally, there is an increase of 20 ± 10 units of RI for each methyl group added.

It is worth pointing out that the analysis of AQC-derivatised AAs with two amino groups showed a drastic decrease in the RI value. This was observed when either of the amino groups was dimethylated. This occurs because the derivatisation could not happen at a tertiary amino group.

Table 3-13. Influence of the Methyl Group Introduction (His, 1-MeHis, 3-HyHis)

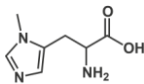
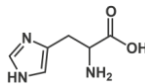
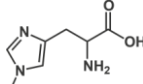
	 3-MeHis		 His		 1-MeHis
	RI	Δ RI	RI	Δ RI	RI
AQC-Derivatised	162	12	150	23	173
IP	644	6	650	11	661
MM	705	1	706	32	738

Table 3-14. Influence of the Methyl Group Introduction (Lys, MeLys, diMeLys, triMeLys)

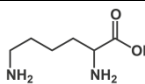
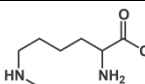
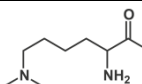
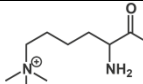
							
	Lys		MeLys		diMeLys		triMeLys
	RI	Δ RI		Δ RI	RI	Δ RI	RI
AQC-Derivatised	468	39	507	-338	196	33	229
IP	693	N/A	N/A	N/A	717	5	712
MM	756	27	783	29	812	19	831

Table 3-15. Influence of the Methyl Group Introduction (Arg, MeArg, sym-diMeArg)

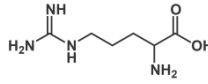
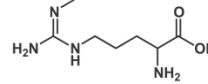
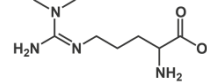
	 Arg		 MeArg		 sym-diMeArg
	RI	Δ RI	RI	Δ RI	RI
AQC-Derivatised	708	17	725	12	737
IP	190	-6	184	N/A	N/A
MM	790	18	808	N/A	N/A

Table 3-16. Influence of the Methyl Group Introduction (Gly, MeGly, diMeGly, triMeGly)

	 Gly		 MeGly		 diMeGly		 triMeGly
	RI	Δ RI	RI	Δ RI	RI	Δ RI	RI
AQC-Derivatised	200	40	240	N/A	N/A	N/A	N/A
IP	200	18	218	13	205	N/A	N/A
MM	200	-27	173	-5	168	80	248

All the predictions presented are associated with error resulting from the lack of statistical data – often the statistics is based on two or three species.

3.3 Conclusions

In this chapter, the concept of RI was revised in order to simplify the identification of modified AAs. Introduction of RIs allowed the use of retention data acquired using different gradients and allowed comparison of different chromatography types.

Analysis of retention data of standard compounds allowed for identification of a structure-retention relationship. RI increments were estimated for AAs modified with methyl and hydroxyl groups (Table 3-17).

Table 3-17. Introduction of methyl- and hydroxyl- group RI increments

Chromatography Type	Methylation		Hydroxylation	
	ΔRI	$\pm \Delta$	ΔRI	$\pm \Delta$
AQC Derivatized	25	11	107	43
IP	11	9	96	121
MM	21	9	67	56

The major challenge associated with the use of RI structure increments results from the limited compounds available and larger errors in RI associated with this. Whereas addition of a methyl group gives a consistent RI increment with a small error (Table 3-17), introduction of a hydroxyl group results in a greater error associated with the increment, because introduction of a hydroxyl group to an aromatic centre (Table 3-11, Table 3-12, Table 3-13) differs from that introduced to an sp^3 -carbon (Table 3-7, Table 3-8, Table 3-9). A larger dataset is expected to reduce the error significantly and make t_R prediction more accurate.

The RI system developed may assist the identification of unknown components and enable prediction of t_R for modified AAs of known structure. In this work the RI system was applied to the identification of the hBBOX catalysed rearrangement of 3-(2,2,2-trimethylhydraziniumyl)-propionate, and to the identification of potential hBBOX substrates and is described in more detail in Chapter Five.

3.4 References

- (1) Kovats, E. *Helvetica Chimica Acta* **1958**, 41, 1915.
- (2) Kovats, E.; Weisz, P. B. *Berichte Der Bunsen-Gesellschaft Fur Physikalische Chemie* **1965**, 69, 812.
- (3) McReynolds, W. O. *Journal of Chromatographic Science* **1970**, 8, 685.
- (4) Baker, J. K.; Ma, C. Y. *Journal of Chromatography* **1979**, 169, 107.
- (5) Smith, R. M. *Analytical Chemistry* **1984**, 56, 256.
- (6) Bogusz, M.; Aderjan, R. *Journal of Chromatography* **1988**, 435, 43.
- (7) Popl, M.; Dolansky, V.; Mostecky, J. *Journal of Chromatography* **1976**, 117, 117.
- (8) Elliott, S. P.; Hale, K. A. *Journal of Chromatography B* **1997**, 694, 99.
- (9) Smith, R. M. *Journal of Chromatography* **1982**, 236, 313.
- (10) Cohen, S. A.; Michaud, D. P. *Analytical Biochemistry* **1993**, 211, 279.
- (11) Larbi, H.; Didaoui, L.; Foudil-Cherif, Y.; Touabet, A. *Chromatographia* **2011**, 74, 459.
- (12) Smith, R. M.; Finn, N. *Journal of Chromatography* **1991**, 537, 51.
- (13) Smith, R. M. *Trac-Trends in Analytical Chemistry* **1984**, 3, 186.
- (14) Baker, J. K. *Analytical Chemistry* **1979**, 51, 1693.
- (15) van Den Dool, H.; Dec. Kratz, P. *Journal of Chromatography A* **1963**, 11, 463.
- (16) Subirats, X.; Bosch, E.; Roses, M. *Journal of Chromatography A* **2004**, 1059, 33.
- (17) Williams, R.
- (18) Morishita, F.; Kakihana, H.; Kojima, T. *Analytical Letters* **1984**, 17, 2385.
- (19) Horvath, C.; Melander, W.; Molnar, I.; Molnar, P. *Analytical Chemistry* **1977**, 49, 2295.
- (20) Petritis, K. N.; Chaimbault, P.; Elfakir, C.; Dreux, M. *Journal of Chromatography A* **1999**, 833, 147.

4 Chapter Four – Quantitative Analysis of AAs: Method Validation and Description

When developing a new quantitative method, it is normal to apply a set of tests in order to validate the method. Such tests provide information about limit of detection (LoD) and quantification (LoQ), linearity, sensitivity, and robustness of the analytical method¹. Reporting the results of quantitative test method can present a challenge. In many cases reported values cannot be used by other researchers because some of them, such as linear regression parameters, are valid only for particular experimental conditions whereas parameters such as LoQ are normally defined by the precision required for the particular experiment and are not universally applicable.

Analytical LC-MS methods comprise both separation and detection. Methods for the chromatographic separation of AAs were introduced in Chapter Two. The work described in this chapter focused on detection techniques and comparison of different methods of quantitative analysis of AAs, revealing the limitations of such methods, and assessing the applicability of such methods for the analysis of modified AAs. The presentation of chromatography MS data is also discussed, due to a lack of agreement and consistent reporting criteria in the field. Ways of describing and comparing the methods of quantitative analysis are considered as are the ways of presenting method validation parameters in such a way that these can be used by other researchers.

4.1 Detection Techniques

Ultra violet (UV) and mass spectrometry (MS) detection methods were used for the research described in this thesis. UV detection was carried out at 254 nm, at which wavelength the absorbance is maximal for all the AQC derivatised AAs². It was noted that at this wavelength, the absorbance maximum is similar for all the derivatised AAs, except for doubly derivatised

AAs such as lysine and cystine². Therefore, sensitivity is most affected by the signal produced by the derivative fragment, rather than that produced from the AA itself. For MS analysis, two spectrometers were used: LCT Premier (Waters) with a time-of-flight (ToF) MS analyser, and Quattro Micro (Waters) with a triple-quadrupole analyser.

MS analysis uses two basic functions, at least one of which is available in every mass-spectrometer: mass scanning (scanning masses over a given range within a period of time) and mass selection (monitoring one particular mass over a given period of time). Mass scanning is reduced to mass selection when the mass range is narrowed to the width of the mass peak. MS systems can be divided into two groups – ones which can perform multiple mass selection and those which can only monitor a mass range (however narrow it is) at a time. The Quattro Micro is a triple quadrupole instrument, which allows mass scanning as well as consecutive monitoring of several selected masses. The LCT Premier is a ToF mass spectrometer, which only allows scanning over a defined mass range. An important characteristic of mass spectrometer detectors is their resolving power, *i.e.* the ability of the mass spectrometer to distinguish between close masses and this can be expressed as follows:

$$R = \frac{M}{\Delta M}$$

Where M is the target mass and ΔM is the difference between target mass and the closest mass which can be distinguished from the target mass (base line resolution). In this respect a large difference in resolving power is observed between the ToF and Triple Quadrupole Systems. LCT Premier ($R = 1000$) has resolving power about five times higher, than that for Quattro Micro ($R = 200$). These numbers are slightly lower than those cited³ because the target compounds – AAs – have very low masses up to ~200 Da.

These two mass analysers present different possibilities for the analysis of AAs. The LCT Premier enables recording of a spectrum with high resolution for every time point of the

mass-chromatogram, which can be further processed and can be mathematically extracted from the raw data. Detailed discussion of the ion extraction process used in this work is provided in Chapter Eight. From the extracted ion chromatograms, given the high resolution of the mass chromatograms, it is possible to maximise signal to noise ratios, thus, minimising the LoD. The Quattro Micro provides a number of different modes of operation (Figure 4-1). For quantitative analysis, particularly useful modes are precursor ion scan and selected reaction monitoring. Precursor ion scanning, if carried out without the fragmentation of a selected ion, allows monitoring of the molecular ion. This technique is referred to as selected ion monitoring (SIM). In the case of selected reaction monitoring (SRM) the selected ion is fragmented and a particular fragment is monitored.

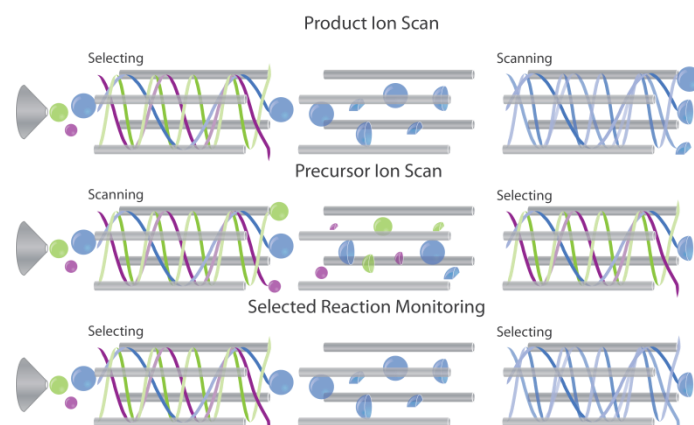


Figure 4-1. Triple Quadrupole MS analyser modes

Therefore, the LCT Premier and the Quattro Micro machines enable two different kinds of selectivity. The first provides high resolution with the possibility to mathematically ‘filter out’ the ‘irrelevant’ masses in post processing thus, enhancing the level of quantification. The second enables ‘filtering out’ the noise by physically removing the irrelevant ions in the mass spectrometer. With the Quattro Micro it is possible to employ both physical mass filtering and ‘chemical filtering’, in which case the precursor ions, which do not generate particular fragments, are removed.

Out of the four detection techniques available: UV detection, MS detection in scanning mode, MS detection in SIM mode, and MS detection in SRM mode – it was necessary to test and select the detection approach enabling optimal sensitivity and selectivity.

4.2 Quantification Method Description

Quantitative methods enable estimation of the amount of analyte based on fact that intensity of the analytical signal is normally proportional to the quantity of analyte in a sample. An important part of quantitative method development is to establish a relationship between the analytical signal and the quantity of an analyte. Usually, the most accurate way to test the quantitative method is to carry out a calibration, because calibration requires no assumptions about the analytical signal to concentration relationship on the calibration concentration range and therefore, is unbiased in its approach. Calibrations provide a large amount of data and statistical description of quantitative methods allows this data to be reduced. It is crucial, however, to present the data in such a way that any parameters produced as a result of the data reduction, can be used by other researchers for the assessment of the method's efficacy and sensitivity for a particular type of analysis and can also be compared with other methods.

Despite a number of very well established ways to describe quantitative methods, there appears to exist a discrepancy in the literature between recommended methods and those used in practice. In order to assess how methods are described and characterised in the literature 25 publications were selected at random for the period 2001 to 2011. To select publications, the Web of Science search engine was used to search through the Expanded Science Citation Index by topic “quanti* method validation”. Only original publications were considered. The method validation routine from each publication is summarised in Table 4-1.

Table 4-1. Method validation parameters utilised for the quantitative analysis assessment

Ref.	Summary
4	Calibration (0, 2, 6, 20, 60, 200 ng per injection); LoD = 3S/N; LoQ = 10S/N; LoD (12 to 380 pg per injection)
5	Calibration (0 to 50 ng/mL, 0.12 to 1 ng/mL); LoD = 3s _y ; LoQ = 10s _y ; R ² as a proof of linearity
6	Calibration (0.1 to 10 µg/mL); LoQ = 0.1 µg/mL; linearity stated
7	Calibration (different range for different compounds); LoD is below the calibration range; R ² as a proof of linearity
8	Calibration (1, 5, 10, 20, 40 ng/mL); LoD = 3σ of blank, LoQ = 10σ of blank; LoD = 1.5 ng/mL; LoQ = 5 ng/mL; R ² > 0.99 as a proof of linearity; regression line $y = 39.8x + 3.36$
9	Calibration (calibration range not clearly stated); R ² as a proof of linearity; provides the list of linear regression equations
10	Calibration (calibration range not clearly stated); LoD and LoQ are derived from the calibration curve
11	Calibration of isotopologues ratio on bilogarithmic axes; LoD not clearly stated (about 30 fmoles based on authors recommendation)
12	Calibration (calibration range not clearly stated); LoD = 3S/N, LoQ = 10S/N; R ² as a proof of linearity
13	Calibration (LoQ, 0.5, 1, 5, 20, 30 µg/L); LoQ = 5S/N, Mandel's fitting test as a proof for linearity
14	LoD lowest detectable concentration, LoQ = 10 LoD/3; reported linearity range 1-2 orders of magnitude, r ² as a proof of linearity
15	Calibration (0.25–10 µg/mL); LoD = 3S/N, LoQ = 10S/N; LoD = 0.05 µg/mL, LoQ = 0.15 µg/mL
16	Calibration (1 ng/mg to 100 ng/mg); LoD = 3S/N, LoQ = 5LoD; LoD up to 3 orders of magnitude below the calibration range
17	Calibration; LoD close to the calibration range; R ² as a proof of linearity
18	Calibration (0.01–3.00 µg/mL); LoD close to the calibration range; linearity reported but not validated
27	LoD or LoQ (the difference is not clear from the paper); linear range reported, no justification is provided
19	Calibration; R ² > 0.99 as an evidence of linearity, linear regression equation provided
20	Calibration (1 ng/mL to 1 µg/mL); no clear distinction between LoD and LoQ; LoQ < 5 ng/mL; R ² > 0.99 as a proof of linearity
21	Multiple injections of blank and of a sample, LoD = 2SD of noise, LoQ: RSD < 20%
22	LoD = 3SD of noise, LoQ is defined through the required precision and accuracy
23	Calibration; LoD below the calibration range; LoQ: RSD < 20%
24	Calibration (0.01 to 100 ng per sheet, all the data points are given); LoD = 3S/N, LoQ = 2LoD; RSD is calculated as well, and the results are compared to the LoQ
25	Calibration (0.03 to 10 µg/mL; 6 data points); LoD = 3S/N, LoD below the lowest concentration point, LoQ not clearly defined
26	Calibration (5 to 500 nM); LoD = 3S/N, LoQ = 10S/N; R ² as a proof of linearity

From the summary in Table 4-1, it can be seen that there is no clear trend in which quantitative data is reported, *i.e.* some data are missing or not calculated in the same way

between publications. However, the majority of authors commonly cite several parameters including LoD (LoD), limit of quantification (LoQ), linearity range, and a linear regression equation. In most cases LoD and LoQ are treated as directly related parameters different only by some factor (normally 3.3, to make it outstand 10 standard deviations of the blank signal from the blank signal mean value; some authors use different factors usually without providing a rationale behind their choice). The most common way to test a quantitative analysis method is to carry out a calibration. Multiple injections of a blank or a sample can also be used, when a calibration appears to be impractical, *e.g.* when a standard is not available. Most commonly the coefficient of determination (R^2) is cited as sufficient proof of linearity, and more rarely lack of fit tests²⁷ such as Mandel's fitting test are applied. Often the linearity range is cited, but no explanation of how it was determined is provided. Problems are also encountered with the using of LoD and LoQ values. In three papers the values of either LoD or LoQ were used, but what was meant by these values was not stated^{6,11,19}. In one of the papers the LoD and LoQ values were used interchangeably and no difference was drawn between them²⁰. Only 5 papers stated the values clearly and distinguishably from one another, with explanation of the use of these values.

The lack of comparable information makes it difficult to identify an appropriate analytical method without further testing. This means that in most cases, the quantitative parameters cannot be used by others without verification and in some cases the information provided in such a concise way can be misleading, *e.g.* reporting the LoD without reporting the lowest concentration actually analysed may result in an underestimation of the LoD.

Thus, on the basis of analysis of the calibration (three replicates for every data point, from 2.5 μM to 100 μM concentration range) data collected²⁸ one set of authors report that the correlation coefficient "to be greater than 0.998 or better". Apparently, based on the

assumption of linearity, extrapolation allows deduction of the LoD to be 5 nM; no validation of the linearity assumption was apparently carried out.

These sort of conclusions do not provide any valuable or reliable information for the next researcher. Firstly, even though the authors report the number of replicates, they do not provide information on how the calibration was designed (*i.e.* what concentration points were used; depending on the concentrations used mathematical artefacts such as the leverage effect can be created) and – even though the calibration covers almost two orders of magnitude – what concentration points have been used within the cited concentration range. Should the calibration be performed using only two extreme concentration points, it is hardly surprising, that the correlation coefficient is close to unity. Another point made in the same paper²⁸ is that the LoD reported is more than two orders of magnitude beyond the concentration covered by the reported calibration range. For this to be valid it requires an assumption that the response follows a straight line for two more orders of magnitude, evidence of which is not provided.

In another publication²⁹ the estimated LoD and LoQ are reported. The estimation was based on the triplicate measurements of a single concentration point at 10 μM and the signal of the blank solution. This estimation suggested the LoD to be well below 0.1 μM , which is more than two orders of magnitude lower than the concentration actually measured. The authors report the LoD with three significant figures, which implies that using this method one can distinguish between concentrations with a difference of 0.001 μM , which is in contradiction with the reported LoD, because the difference between concentrations, which can be distinguished from each other, corresponds to the LoD, providing that data is distributed homoscedastically.

Therefore, summarising and presenting method validation data presents the scientist with a significant challenge to convey the results to potential users of the method in such a way that

biased interpretation or repetition of costly method validation is not required. This section revises theoretical approaches for presenting quantitative methods and addresses the practical use of such methods. It addresses several examples of quantitative methods described in the literature and analyses the reliability of the conclusions based on such data. The calibration curve and the information which can be derived from the calibration curve are considered. The section concludes with a calibration experiment design which is subsequently used for quantitative AA analysis validation and provides a set of the descriptive methods, which are used when reporting the calibration experimental work in this thesis and which can be recommended for the method description in general.

4.3 Reporting the results of the calibration tests.

Describing data with limited a set of parameters can be misleading, which can be illustrated using the example of Anscombe's Quartet³⁰. Four different datasets are presented in Figure 4-2. Even though they are obviously different, statistically they present the same result and can be summarised using the same set of parameters (Figure 4-2). Importantly, the R^2 values, normally presented as evidence of linearity, do not allow differentiation between a linear trend and quadratic curve (Figure 4-2, A & B). Additionally, not presenting the data as a scatter plot prohibits the inspection of outliers; a straight line with a single outlier could be mistaken for a noisy dataset (Figure 4-2, A & C). Finally, the leverage effect, illustrated in graph D (Figure 4-2) is particularly important when a calibration curve is constructed. In cases when the concentration points outstand from one another beyond an order of magnitude, the fitted curve will be dominated predominantly by data points related to higher concentrations. Thus, the conclusions drawn on the basis of the regression forced in such a way will be biased as a result of this effect.

When analysing calibration datasets, the two cases presented in graphs B and D (Figure 4-2) are of pivotal importance. It is often the case that misuse of the R^2 value (using it as a proof

of linearity) results in a false conclusion about the linearity of the dataset. When this is combined with the leverage effect, caused by inadequate calibration design, the conclusion drawn from such analysis affects not only the calculated values at lower concentrations but also the LoD.

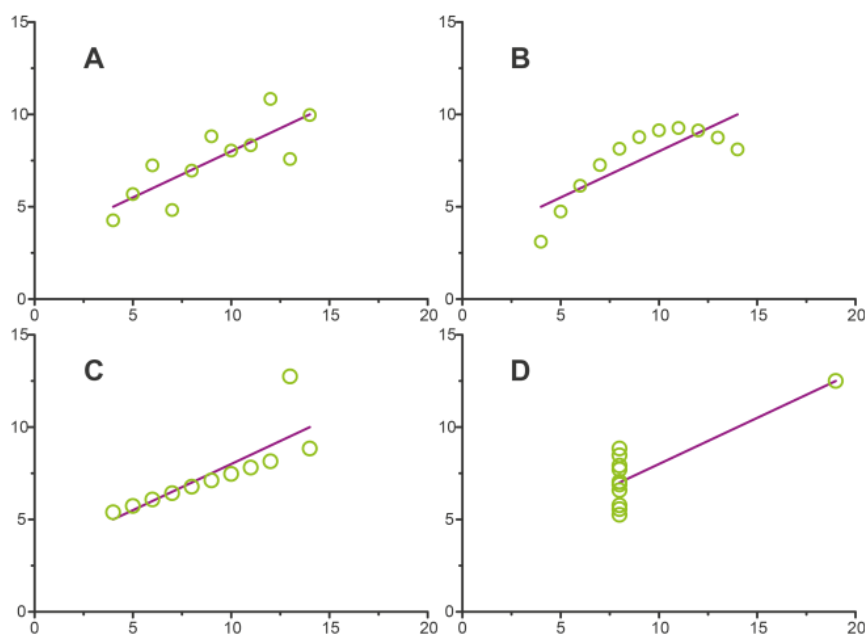


Figure 4-2. Anscombe's quartet³⁰

Presented data sets are statistically identical. Number of observations (n) = 11; mean of the x 's ($\bar{x}$) = 9.0; mean of the y 's ($\bar{y}$) = 7.5; equation of regression line: $y = 3 + 0.5x$; sum of squares of x - 110.0; regression sum of squares = 27.50 (1 d.f.); residual sum of squares of y = 13.75 (9 d.f.); estimated standard error of slope = 0.118; $R^2 = 0.667$

4.3.1 Linearity Testing

It appears widely accepted in the published literature that the results of quantitative concentration calibrations are appropriately represented by use of the correlation coefficient R ³¹.

$$R = \frac{\sum_i ((x_i - \bar{x}) \cdot (y_i - \bar{y}))}{\sqrt{(\sum_i (x_i - \bar{x})^2) \cdot (\sum_i (y_i - \bar{y})^2)}}$$

Where x refers to the concentration and y refers to the signal.

By analogy, the coefficient of determination is R^2 . This parameter shows the fraction of the variance in the one variable (in my case – y – signal) that is explained by the linear model in the other variable (in my case – x – AA concentration).

Neither of these parameters can be applied to advocate the linearity of the dataset. As shown in the previous section, application of this parameter is not characteristic of the curve shape^{32,33}, and the correlation coefficient is strongly affected by value of the slope, and for the datasets, with constant variation and different only in the value of the slope, the correlation coefficient increases with the increase of the slope. It has also been reported that the straight line can be forced through an *a priori* non-linear dataset and provide correlation coefficient and coefficient of determination close to unity. An example of a non-linear dataset fitted with a straight line ($R^2 = 0.9998$) is provided in Chapter Six. It is worth noting the types of problems that can be caused by assigning linearity to a non-linear dataset. Normally, the calibration dataset presents either a concave or a convex curve. The leverage effect will inevitably result in adjusting the position of the fitted line closer to the higher concentrations. This will in due course result in the underestimation of the intercept and hence the LoD in case of concave calibration dataset, and overestimation of the LoD in the case of a convex dataset.

Numerous tests exist which allow us to identify whether a dataset best fit is a straight line or not³⁴⁻³⁶. However, it is likely that the most robust way of calibration data set assessment is to analyse the residual plot or the Q-Q plot either visually, or by applying one of the normality tests to the data, which in this case should be normally distributed relative to zero³⁷.

Almost any dataset can be fitted with the linear equation. Choice as to whether this kind of fit is adequate or not should be based on the requirements to the errors associated with the fit chosen; error assignment based on the linear fit will be presented later in this section.

4.3.2 LoD (LoD)

Prior to discussing the ways of finding the LoD it is necessary to define this parameter. According to IUPAC:

The LoD, expressed as the concentration, C_L , or the quantity, q_L , is derived from the smallest measure, x_{LoD} , that can be detected with reasonable certainty for a given analytical procedure. The value of x_{LoD} is given by the equation

$$x_{LoD} = \bar{x}_b + k \cdot s_b$$

Where $\bar{x}_b$ is the mean of the blank measures, s_b is the standard deviation of the blank measures, and k is a numerical factor chosen according to the confidence level desired³⁸.

In other words, LoD is the lowest concentration, which can be distinguished from the noise with a given certainty. This certainty reflects the probability of mis-finding the analyte (making α error), *i.e.* finding the analyte in a sample, which does not have it. With k equal to 3, and assumption that the noise follows a normal distribution, this probability is 0.15%.

Such a definition gives a researcher certain freedom. First, the parameter k can be chosen depending on the required level of certainty. Even though, conventionally, the value for k is set at 3, this value can be adjusted according to the needs of the researcher and heightened to provide greater confidence in the finding or lowered to introduce more uncertainty into the results. The parameter, $\bar{x}_b$, is strongly dependent on the method used to estimate the mean of the blank measures. It can be estimated by measuring directly the signal of the blank sample, without analyte. This can only be efficiently applied to the analytical methods in which the signal has only one dimension, *e.g.* absorbance, and is not suitable for the chromatographic methods in which the signal (chromatographic peak area) has two dimensions. In this case, the calibration curve can be applied in order to identify the signal value at zero concentration.

The signal is calculated from the calibration curve equation at zero concentration. Linear calibration is most often applied for the sake of its simplicity, *i.e.*

$$y = a + b \cdot x$$

Where y is the signal value, x is the concentration of the analyte, a is the signal value at zero concentration of the analyte, and b is the slope of the calibration curve.

Therefore, in signal terms the LoD will be defined as $y_{LoD}=b+3s_{y/x}$, and in concentration or quantity terms it will be $x_{LoD}=3s_{y/x}/a$, where y_{LoD} is the minimum signal distinguishable from the signal of the blank sample, with 99.7% confidence, and x_{LoD} is the minimum concentration, corresponding to this signal.

In practice, another approach is often applied, termed signal to noise ratio (SNR); the LoD is estimated by relating the signal value of the analyte of known concentration to the signal value of the background, then finding the concentration at which the SNR will be equal to 3. All these methods, even though they are well established within analytical chemistry, when applied to chromatography, have the strong disadvantage of treating the two dimensional signal as one dimensional, thus, losing information.

For the calibration dataset fitted with a straight line slope and intercept of such line have to be calculated first. The slope is calculated so that the sum of the least squares of residuals is minimised

$$b = \frac{\sum_i ((x_i - \bar{x}) \cdot (y_i - \bar{y}))}{\sum_i (x_i - \bar{x})^2}$$

And the intercept is calculated as follows:

$$a = \bar{y} - b \cdot \bar{x}$$

The random errors y (signal) have to be calculated according to the formula:

$$s_{y/x} = \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{n - 2}}$$

Where $\hat{y}$ stands for the estimate derived from the linear fit equation.

It is then necessary to calculate the deviation associated with the slope, the deviation associated with the intercept and the deviation which will be assigned to the concentration value calculated using such a linear fit.

The standard deviation of the slope is calculated according to the formula:

$$s_b = \frac{s_{y/x}}{\sqrt{\sum_i (x_i - \bar{x})^2}}$$

The standard deviation of the intercept is calculated:

$$s_a = s_{y/x} \sqrt{\frac{\sum_i x_i^2}{n \cdot \sum_i (x_i - \bar{x})^2}}$$

This allows us to calculate the LoD. In signal terms the LoD can be calculated as follows

$$LoD_y = a + 3 \cdot s_{y/x}$$

And the concentration LoD is given by:

$$LoD_x = \frac{3 \cdot s_{y/x}}{b}$$

Finally the standard deviation of the concentration value calculated from such fit can be calculated as follows

$$s_{x_0} = \frac{s_{y/x}}{b} \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(y_0 - \bar{y})^2}{b^2 \cdot \sum_i (x_i - \bar{x})^2}}$$

With ***n*** stands for the number of points used for the calibration curve, and ***m*** is the number of replicate measurements of the sample. In practice often only one measurement is performed for the sample.

4.3.3 Limit of Quantification (LoQ)

Another parameter frequently reported is the limit of quantification (LoQ). This can be an ambiguous parameter because of the various ways it has been defined (Table 4-1). First, it is necessary to explain the use of factor 3 (or 3.3) together with the LoD when defining the LoQ. Generally, the LoQ is defined as the lowest concentration at which the quantification can be performed with a defined precision^{39,40}. The measure of precision is the standard deviation of either signal or noise. The standard deviation itself depends only on the analytical method (instrument settings, buffers, nature of the analytes etc.) used and cannot be varied without changing the method. Yet, it is the researcher who defines the relative standard deviation (RSD) suitable for the purposes of their analysis⁴¹. This value is by no means standard and depends on the precision required; therefore, if the LoQ is reported, the value of the RSD used should also be provided. Providing the LoQ is equal to 3 (or 3.3) times the LoD the concentration at which the β error (the error of reporting the absence of analyte, whereas it is present) is 5%, and RSD corresponds to 10%⁴². In fact, reporting LoQ this way is not only misleading, and does not give others an idea of what the level of precision is, but is also wasteful, because it conveys information, which can be easily deduced from the LoD value.

Instead of quoting the LoQ as $10 \times s_{y/x}$, the definition of the LoQ should be used. According to IUPAC the measure of precision is the standard deviation, which can be calculated as follows:

$$s = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{(n - 1)}}$$

In order to identify the LoQ, the cut off point on the RSD scale should first be clearly stated⁴³.

4.3.4 Homoscedasticity

An important test which might be useful when describing the calibration dataset is the homoscedasticity test; this shows whether the data variance along the y axes remains constant for all values of x. For a calibration curve it means that the signal variance is constant for any concentration point within the calibration range³⁷. Often, the signal variance increases with the concentration, for example, the calibration for HyP using UV detection and MS detection in SIM mode (Figure 4-3 and Figure 4-4). The scatter plot for HyP calibration using the UV detection (Figure 4-3) shows distribution close to homoscedastic, whereas the scatter plot presenting the results of HyP calibration using the MS detection in SIM mode (Figure 4-4) is an example of a heteroscedastic dataset.

This observation may seem insignificant, however, if the heteroscedasticity is revealed, the standard error associated with the concentration will increase proportionally to the signal variance, which means that small concentrations will have a smaller standard error associated with them.

Although rarely accounted for, homoscedasticity drastically affects the quantification experiment design. Often the variance increases with concentration; in a such situation certain analytical methods (such as standard addition) should be treated with caution because by adding a standard amount of analyte, the concentration of analyte is increased, and so is the variance. Depending on the variance increase, this can make quantification of small quantities of analyte impossible. Likewise, if the decrease of component in a reaction is measured and its concentration is in the area of high variance, it might be necessary to measure the increasing concentration of the conjugated product, which concentration might be lower, but will have lower variance.

4.3.5 Quantitative Method Description Summary

When presenting a new quantitative method it is vital to provide calibration data in full. Neither the correlation coefficient nor the coefficient of determination should be used as a proof of linearity of the dataset. Sufficient proof of linearity can be provided if the residuals follow a normal distribution over the full range of the calibration; alternatively goodness of fit tests can be applied. Calibration failing the linearity test can be fit with a straight line if the error resulting from such a fit meets the requirements imposed by the researcher.

LoD can only be considered valid if measurement of the signal has been performed at the concentration corresponding to the LoD. Any other way of identifying LoD should be considered as an appreciation only. When appreciation of the LoD is performed, the method used for the assessment should however be clearly stated.

The importance of LoQ is dependent on the requirements of the research and, therefore, such requirements should be stated; the level of RSD suffices to indicate the level of required precision and as RSD equal to 10% might not always be at the required level of precision. RSD plotted against concentration throughout the complete calibration range can be used in order to show whether the required level of precision is achievable or not.

The homoscedasticity of the data should be considered throughout the calibration range. This information is vital for quantification experiment design and cannot be conveyed by using alternative parameters.

It is desirable that the concentrations within a calibration range should be equidistant in order to apply a linear fit throughout the calibration range. When several orders of magnitude are covered by the calibration, the calibration design, in which concentrations are equidistant within one order of concentration magnitude, can be applied. If such a design is applied in

general it would be preferable to perform the linear fit for every order of magnitude, unless a larger data set has been shown to follow a linear trend.

4.4 Quantitative Method Validation

4.4.1 Calibration Design

For the analysis of AAs, six analytical methods were tested: (1) analysis of the AQC-derivatised AAs with UV detection (chosen as a reference standard); methods using AQC-derivatised AA analysis using MS detection in (2) scanning mode and in (3) selection ion monitoring mode; to overcome the limitations imposed by the AQC derivatisation, two chromatographic methods were developed for non-derivatised AA analysis – reversed phase ion pairing chromatography of non-derivatised AAs was combined with MS detection (4) in selected ion monitoring mode and (5) in selected reaction monitoring mode; mixed mode chromatography was followed by MS detection in selected ion monitoring mode (6).

Colour coding is applied to the calibration data presented in this chapter. Purple is used for analytical methods using derivatised AAs and green is used for the analytical methods using non-derivatised AAs.

Analytical standards were not commercially available for most of the PTMs of AAs. It was, therefore, necessary to be able to estimate the suitability of the method without the use of the standards, relying entirely on the data available for species of similar physical and chemical properties. Commercially available hydroxyproline and hydroxylysine, together with the non-modified AAs were used as test compounds for the analytical methods developed. It was assumed in this work that sensitivities were similar for the modified and non-modified AAs and the conclusions drawn for the non-modified AAs and hydroxylysine or hydroxyproline were valid for modification of other AAs as well. In cases, when the standards are not

available, it is suggested that the closest structural analogue can be used as a standard if the calibration is needed.

PTMs of AAs occur in a wide variety of samples such as synthetic mixtures, *in vitro* enzyme assays, cell based experiments, and *in vivo* experiments. Depending on the sample type, the concentration of the analyte may vary significantly. Therefore, it was necessary to test the applicability of the analytical methods developed in a wide concentration range. A concentration calibration was carried out over the range 0.1 μM to 100 μM , thus, covering three orders of magnitude.

Within an order of magnitude, concentrations of the analytical standards were equally distanced from each other.

- 0.1, 0.2, ... 1 μM
- 2, 3 ... 10 μM
- 20, 30 ... 100 μM

This calibration design eliminates leverage effect within a particular order of magnitude and allows unbiased identification of the LoD and limit of quantification. Using this calibration design, it was possible to appreciate the variance of the data at each concentration point in a range.

For each order of magnitude, the calibration was prepared by a serial dilution of the highest concentration in the range. The highest concentration in the range was prepared by serial dilution of a stock solution. 5 replicate peak area measurements were made for the calibrant at every concentration, thus, resulting in 140 calibration points per curve. For the acquisition, concentrations were arranged in descending order. After five injections a blank (sample without analyte) was analysed. Details of the mass chromatogram peak integration are discussed and described in Chapter Eight.

4.4.2 Results and Discussion

The results of the calibration are summarised in Table 4-2, Figure 4-3 and Figure 4-4. The UV detection coupled with the chromatography of the derivatised AAs appears to have lowest LoD; the SIM mode MS detection has a slightly higher LoD. The UV detection method shows a very narrow distribution in comparison to the MS methods. This can be explained by the fact that UV detection is determined primarily by the introduced AQC group with high absorbance, and individual properties of AAs are less important. However, with the MS methods, individual properties of AAs are responsible for the charge state and the abundance of the charged species; as a result the variance in LoD is greater.

Table 4-2. Concentration calibration statistics (0.1 to 100 µM)

$$y = a + b \cdot x$$

y is the area of the chromatographic peak, and x is the concentration, a is an intercept and b is the slope; LoD_{av} is the average LoD for 20 standard AAs and HyL and HyP given together with the standard deviation.

	Derivatised AAs						Non-Derivatised AAs					
	UV		Scanning		SIM		(IP) SIM		(IP) SRM		(MM) SIM	
	HyP	HyL	HyP	HyL	HyP	HyL	HyP	HyL	HyP	HyL	HyP	HyL
R2	0.995	0.996	0.918	0.983	0.986	0.986	0.955	0.984	0.878	0.935	0.916	0.920
b	51.4	23.3	1.07	0.80	141	96	6713	245	0.30	0.38	793	484
a	-122	-93	1.3	-0.5	14	-281	32156	-714	4.95	0.23	-7063	-4657
b SD	0.3	0.1	0.03	0.01	1.4	1.0	124	3	0.010	0.009	22	13
a SD	12	5	1.0	0.3	53	39	4625	100	0.38	0.34	823	488
Sy/x	116	51	9.8	3.2	518	374	44760	966	3.70	3.33	7968	4718
LoD (µM)	7	7	28	12	11	12	20	12	37	26	30	29
LoD_{av} (µM)	6.7±0.4		23.0±12.9		10.5±2.1		14.0±4.2		24.1±9.1		17.5±3.6	

The LoD for UV detection is affected by the complexity of the sample; however, the data presented here was obtained using LC-grade standards. In the case of more complex mixtures, where other UV absorbing species are present, the LoD will be higher than that for the mixture of standards, so identification of the AAs present at small quantities is less reliable.

For MS methods the LoD appears to be proportional to the abundance of the charge state. Charge state abundance is crucial for AAs with two amino groups, which can carry two protons, such as lysine, hydroxylysine, ornithine, cystine. Cysteine presents another analytical problem as it tends to form cystine under alkaline conditions and at high

concentrations; this should be accounted for when analysing both cysteine and its modifications.

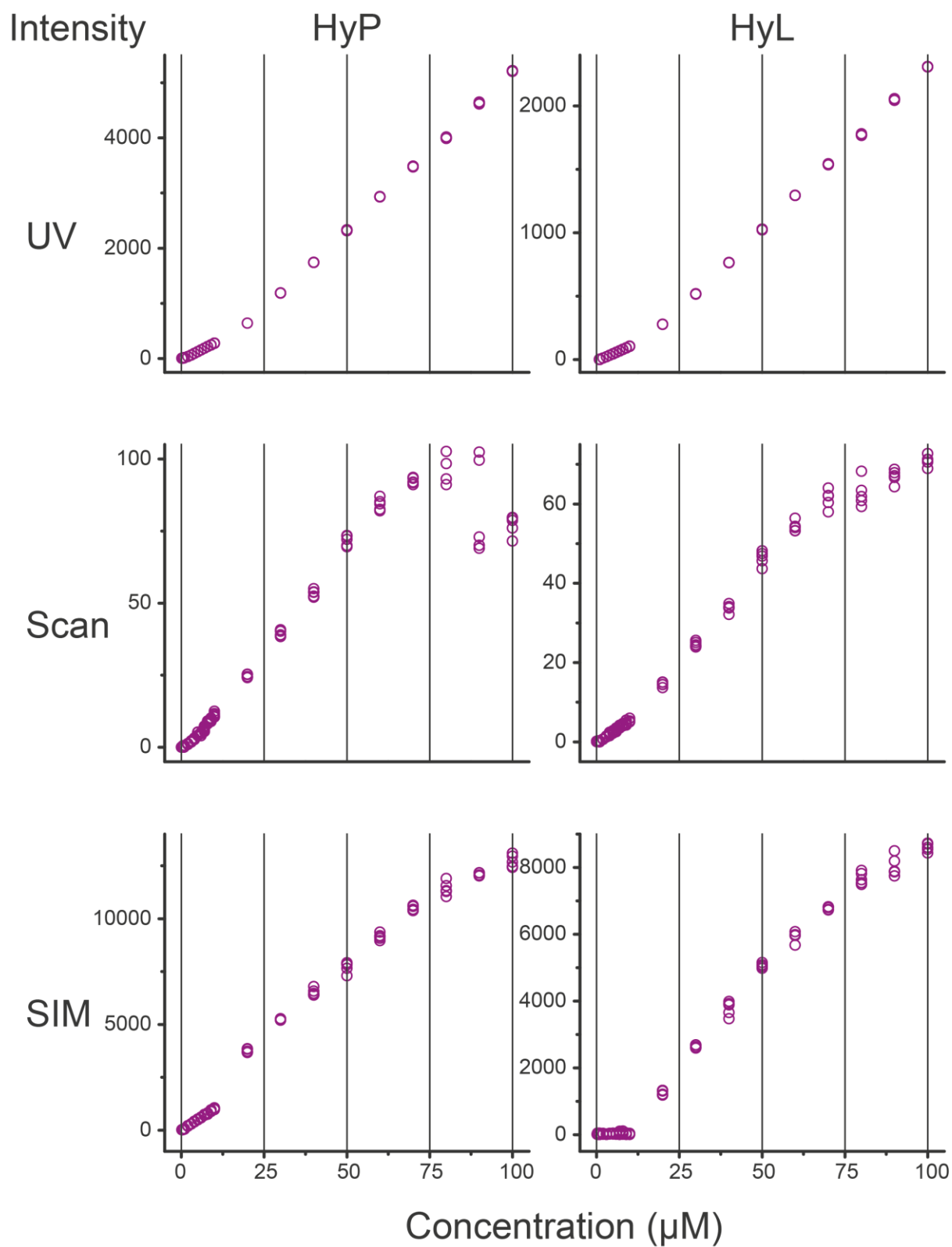


Figure 4-3. Calibration results for derivatised AAs (0.1 to 100 μM)

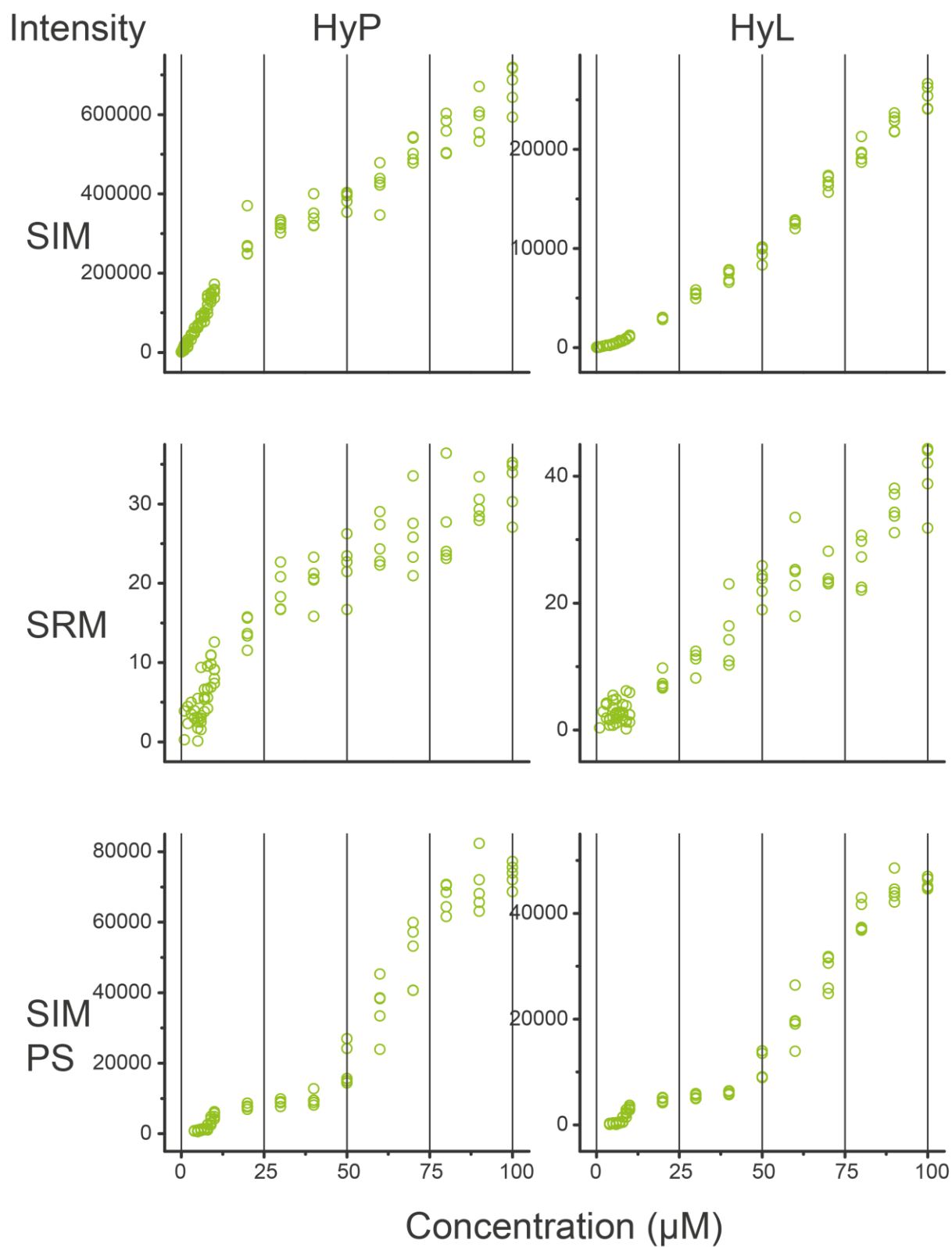


Figure 4-4. Calibration results for non-derivatised AAs (0.1 to 100 μM)

4.4.2.1 Linearity

All the quantitative methods tested demonstrate a high coefficient of determination (R^2) (Table 4-2). However, none of the calibrations can be considered to be linear. To demonstrate this point, the calibration datasets were divided into three subsets based on concentration of the standards as follows (0.1 to 1 μM ; 1 to 10 μM ; and 10 to 100 μM) in order to avoid leverage effects as discussed previously. Statistical parameters of linear regression were applied to the data subsets and are presented in Table 4-3, Table 4-4, and Table 4-5. Comparison of the slopes of the fitted lines for different calibration ranges show that none correspond to a straight line; on the contrary the calibration datasets follow either concave or convex curve.

The non-linear calibrations can be approximated with linear intervals. In this work, the linear intervals were taken to be equal to one order of magnitude. For the purposes of quantitative analysis, two methods of validating the performance of the analytical system were suggested: single point calibration or simplified calibration. For the single point calibration, one concentration point was used on the border of linear intervals. For example 10 μM is sufficient for the quantification of the samples in the concentration range 1 to 10 μM as well as 10 to 100 μM . Simplified calibration includes the limiting points of the linear intervals and LoD, if different from the lowest limiting point.

Table 4-3. Concentration calibration statistics (10 to 100 μM)

$$y = a + b \cdot x$$

y is the area of the chromatographic peak, and x is the concentration, a is an intercept and b is the slope of the fitted line

	UV		Derivatised AAs				Non-Derivatised AAs					
			Scanning		SIM		(IP) SIM		(IP) SRM		(MM) SIM	
	HyP	HyL	HyP	HyL	HyP	HyL	HyP	HyL	HyP	HyL	HyP	HyL
R²	0.998	0.999	0.744	0.952	0.969	0.973	0.928	0.984	0.796	0.885	0.898	0.907
b	55.9	24.9	0.85	0.75	126	96	5154	277	0.23	0.39	897	550
a	-431	-205	16.3	2.8	1074	-290	140216	-2928	9.86	-0.25	-14533	-9427
b SD	0.3	0.1	0.07	0.02	3.2	2.3	207	5	0.017	0.020	44	25
a SD	20	8	4.5	1.5	201	142	12873	317	1.04	1.23	2713	1574
Sy/x	64	26	14.7	4.9	659	465	42138	1036	3.41	4.01	8879	5152
LoD (μM)	3	3	52	20	16	15	25	11	44	31	30	28

Table 4-4. Concentration calibration statistics (1 to 10 µM)

$$y = a + b \cdot x$$

y is the area of the chromatographic peak, and x is the concentration, a is an intercept and b is the slope of the fitted line

	Derivatised AAs						Non-Derivatised AAs					
	UV		Scanning		SIM		(IP) SIM		(IP) SRM		(MM) SIM	
	HyP	HyL	HyP	HyL	HyP	HyL	HyP	HyL	HyP	HyL	HyP	HyL
R2	0.999	0.999	0.964	0.974	0.995	0.055	0.953	0.933	0.542	0.010	0.729	0.744
b	29.7	11.5	1.21	0.58	104	2	15882	113	0.89	0.06	772	511
a	-22	-11	-1.5	-0.4	-21	16	-8737	-145	-0.12	2.25	-3342	-2551
b SD	0.2	0.0	0.03	0.01	1.1	1.1	511	4	0.108	0.079	46	30
a SD	1	0	0.2	0.1	7	7	3173	27	0.67	0.49	288	184
Sy/x	3	1	0.7	0.3	22	22	10386	88	2.18	1.60	944	601
LoD (µM)	0.3	0.2	2	1	1	36	2	2	7	77	4	4

Table 4-5. Concentration calibration statistics (0.1 to 1 µM)

$$y = a + b \cdot x$$

y is the area of the chromatographic peak, and x is the concentration, a is an intercept and b is the slope of the fitted line

	Derivatised AAs						Non-Derivatised AAs					
	UV		Scanning		SIM		(IP) SIM		(IP) SRM		(MM) SIM	
	HyP	HyL	HyP	HyL	HyP	HyL	HyP	HyL	HyP	HyL	HyP	HyL
R2	0.988	0.255	0.478	0.007	0.937	0.001	0.758	0.687	–	–	–	–
b	16.1	1.5	0.46	0.02	55	-1	12609	50	–	–	–	–
a	-1	-0.04	-0.1	0.1	4	18	326	7	–	–	–	–
b SD	0.2	0.1	0.07	0.03	1.8	5.5	1029	5	–	–	–	–
a SD	0.1	0.04	0.0	0.0	1	3	639	3	–	–	–	–
Sy/x	0.4	0.1	0.1	0.1	4	11	2090	10	–	–	–	–
LoD (µM)	0.1	0.3	1	10	0	-25	0.5	0.6	–	–	–	–

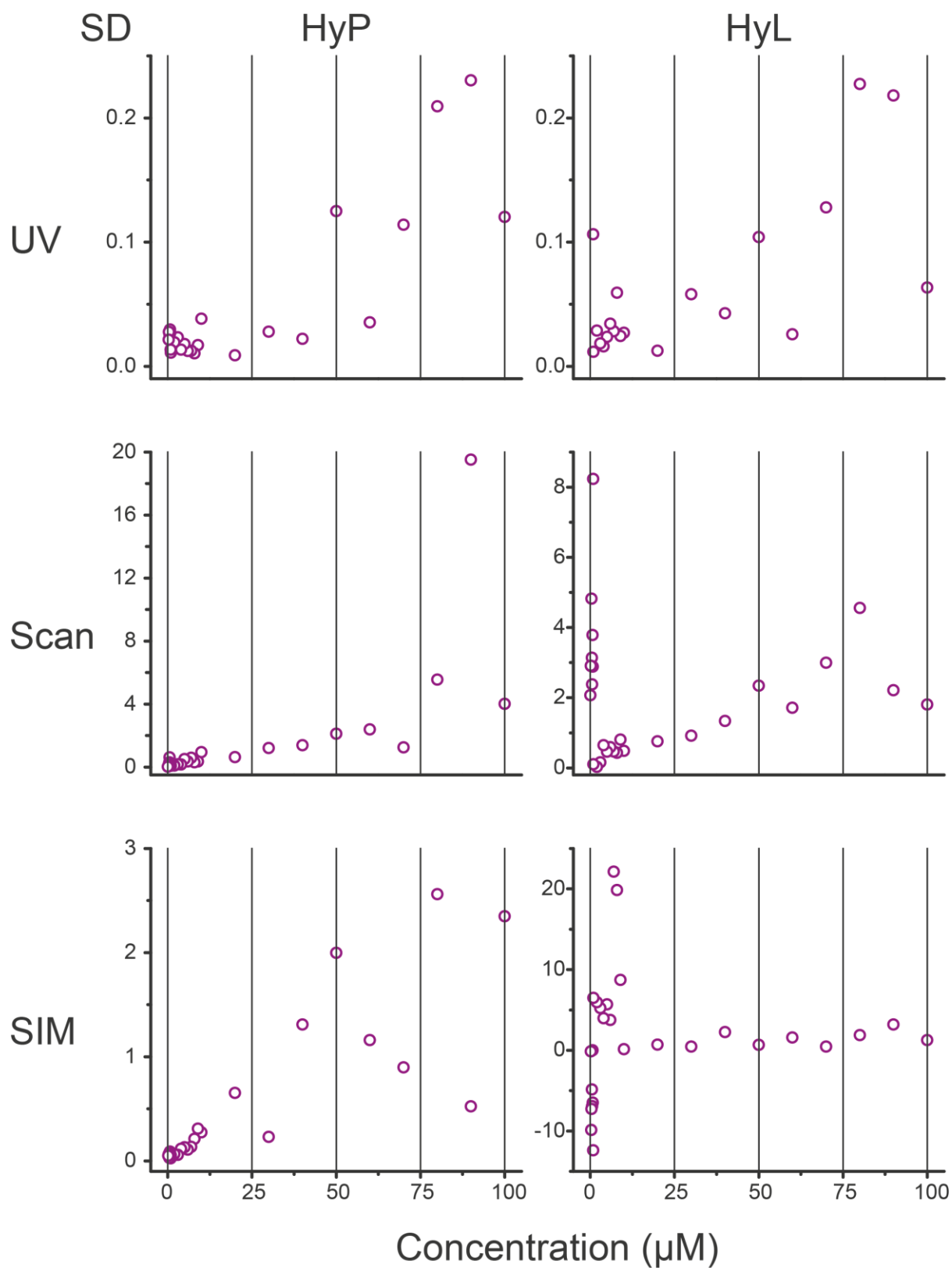


Figure 4-5. SD on concentration dependency for derivatised AAs

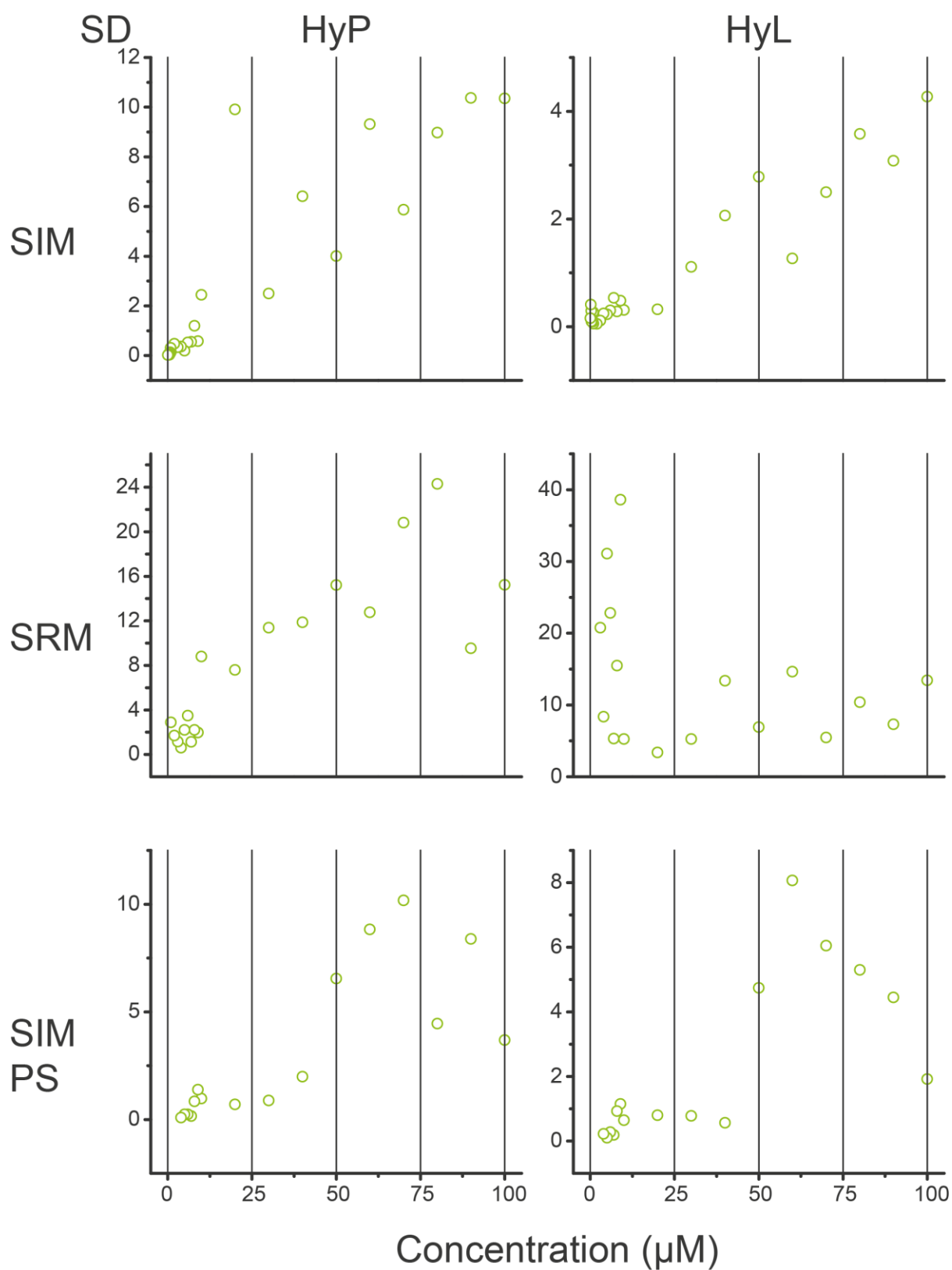


Figure 4-6. SD on concentration dependency for non-derivatised AAs

4.4.2.2 Homoscedasticity of the Experimental Data

Figure 4-5 and Figure 4-6 show that for all the methods, the standard deviation increases with concentration. This should be taken into account when standard addition methods are designed. It should also be taken into account when fitting the calibration data with a straight line; to compensate for the increasing standard deviation weighted regression can be used³⁷, or the calibration range can be divided into several intervals and fitted separately.

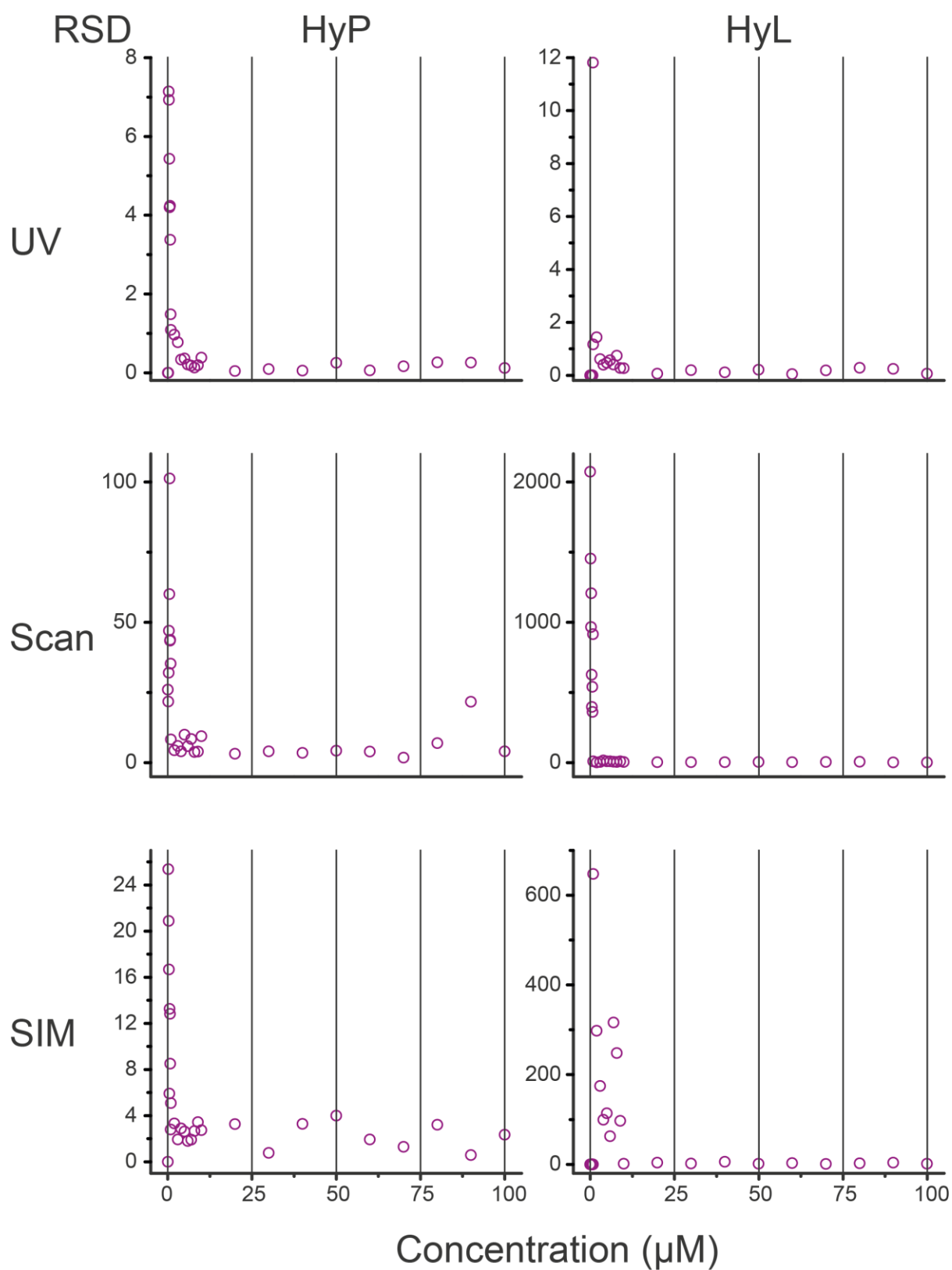


Figure 4-7. RSD on concentration dependency for derivatised AAs

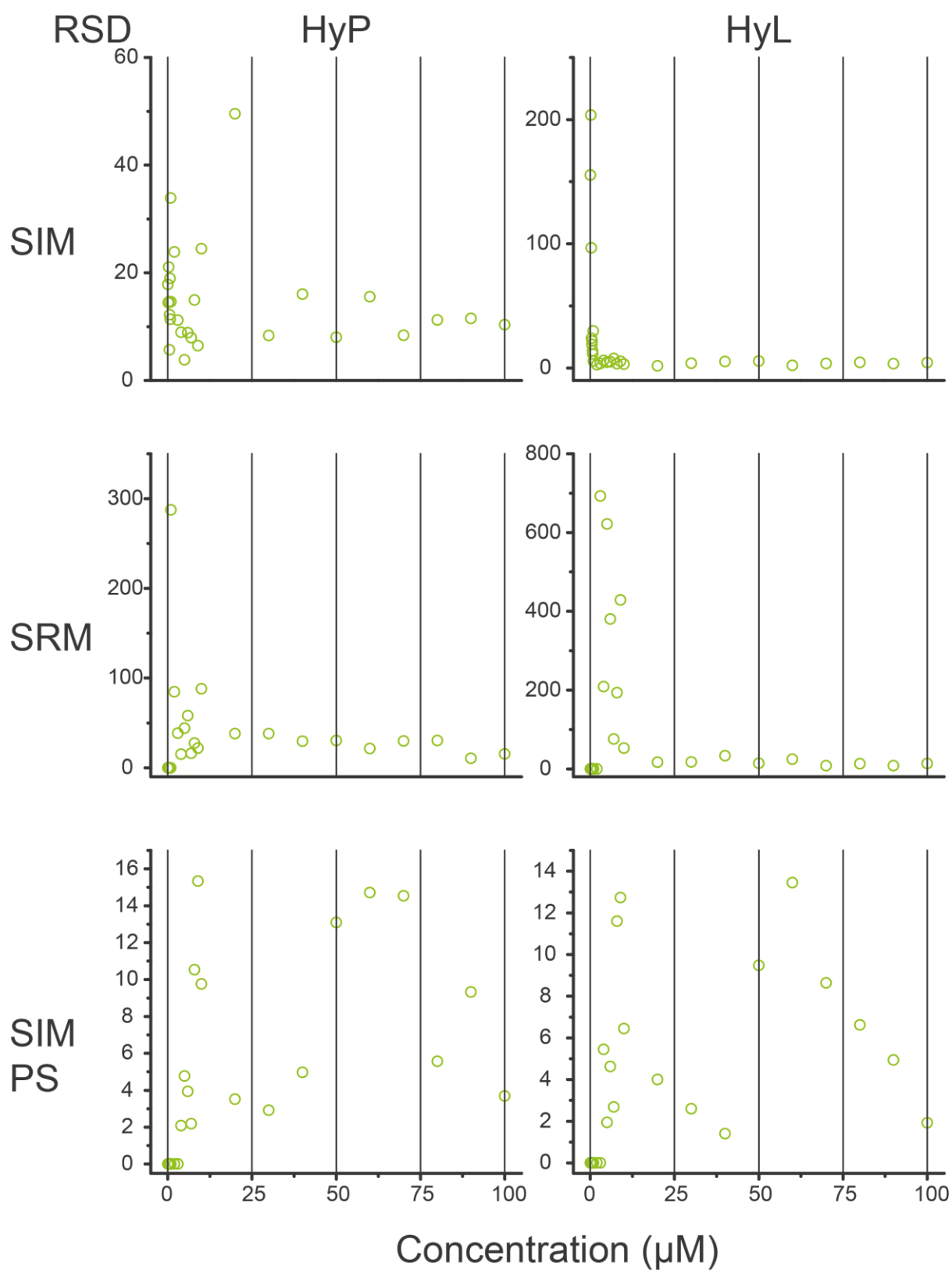


Figure 4-8. RSD on concentration dependency for non-derivatised AAs

4.4.2.3 LoDs and LoQs for Modified AAs

The LoD is dependent on the way linear regression is constructed. When a calibration curve is used to identify the LoD, two parameters affect the result: slope of the fitted curve (Figure 4-3 and Figure 4-4) and the variance of the signal (Figure 4-5 and Figure 4-6). With the calibration design used in this work, in order to avoid the leverage effect introduced by the data points at high concentration, it is necessary to narrow the calibration range to one order of magnitude. This generally also reduces the variance and allows for more accurate identification of the LoD.

From Table 4-5, it can be seen that the LoD for the UV detection method is close to the LoD for the SIM mode of detection for non-derivatised AAs separated by reversed phase chromatography using an ion pairing reagent. The UV detection of AAs is generally less practical, however, because of low selectivity of the method (any compound absorbing UV light at 254 nm interferes with the signal). Analysis of derivatised AAs using MS detection in scanning mode (Table 4-4) provides a reasonable alternative to UV detection combining high sensitivity and selectivity, which allows for unambiguous identification of the AAs.

Contrary to expectation, MS detection in SRM mode shows a higher LoD than for the SIM mode (Table 4-4). Fragmentation in SRM mode did not achieved 100% efficiency, resulting in a decrease of the intensity of the fragment peak in SRM mode compared with the intensity of the precursor peak in SIM mode. However, by using double mass filtering, SRM mode allows a reduction of the background noise. The former process, leads to a decrease in sensitivity, which is not compensated for by a reduction in the background noise.

Reversed phase mixed mode chromatography MS detection in SIM mode proved to have higher LoD than reversed phase chromatography coupled to the same detector. This can be explained by the fact that peaks are much broader in the case of reverse phase mixed mode

chromatography (Chapter Two), and therefore, contribution to the background noise is greater.

The methods used were characterised with relative standard deviation dependency on the concentration plots (Figure 4-7 and Figure 4-8), from which the limit of quantification can be easily deduced. It can be seen that for the derivatised AA analysis, the relative standard deviation reduces with an increase in concentration and remains at about the 4% level for higher concentrations. Lower precision is observed for the analysis of the non derivatised AAs; it approximately 10% for MS detection in SIM mode for the reversed phase ion pairing chromatography and about 15% for reversed phase mixed mode chromatography.

4.4.2.4 Comparison of Modified and Non-Modified AAs

LoD and sensitivity for hydroxylated lysine and proline were compared with lysine and proline. The results are presented in Table 4-6. It can be seen that LoDs are similar for both modified and non-modified species, whereas sensitivity is significantly different by all the methods except for MS detection in scanning mode. It can be concluded that the non-modified AA can be used for the presence or absence test of the related modified AA, however, for an accurate quantitative analysis, the modified AA itself should be used as a quantification standard.

Table 4-6. Comparison of the modified and non-modified AAs LoDs and sensitivities (0.1 to 10 µM)

Sensitivity of the method is taken as the slope of the calibration curve.

	UV		Derivatised AAs				Non-Derivatised AAs					
			Scanning		SIM		(IP) SIM		(IP) SRM		(MM) SIM	
	LoD (µM)	b	LoD (µM)	b	LoD (µM)	b	LoD (µM)	b	LoD (µM)	b	LoD (µM)	b
Pro	0.5	10	1	1	1	104	0.5	338	24	0.6	20	1577
HyP	0.6	28	2	1	1	68	0.4	458	37	0.3	19	855
Lys	0.2	6	2	0.5	16	0.2	1	102	16	1	14	616
HyL	0.3	11	1	0.5	30	0.2	0.5	51	26	0.4	15	526

4.4.3 Conclusions

Analysis of derivatised AAs using UV detection following separation with reversed phase chromatography provides a LoD at the sub μM level, combined with a high level of precision, *i.e.* the RSD goes down to 2% with an increase in concentration. There are two important limitations when using this method. First, a general disadvantage of the use of UV detection for the analysis of trace compounds is low selectivity, which results in an inability to distinguish an analyte present in minor concentrations from the major compound, especially in the case of closely eluted compounds. Second, any buffers with primary and secondary amino groups are not suitable for use with derivatising reagent (for example Tris is one of the most commonly used buffers in biochemical assays). For this type of chromatography, MS detection in scanning mode provides a reasonable alternative to UV detection. Even though the LoD appears to be higher, this method is suitable because of its high selectivity. Using MS detection in SIM mode produces a much higher LoD so it is not the best alternative for the UV detection with this type of chromatography.

Reversed phase ion pairing chromatography, combined with MS detection in SIM mode provides a LoD comparable to that of UV detection of derivatised AAs. The use of an ion pairing reagent may appear to be beneficial because it provides necessary separation of the AAs in far a shorter time, than the method using AQC-derivatised AAs. However, it should be noted that perfluorinated long chain carboxylic acids cause significant contamination of the MS system with negative ions, which compromises the detection in negative ion mode due to the ion suppression. The precision of ion pairing chromatography combined with SIM MS detection is lower, and it is not possible to go below RSD of 5% within this method. Combining this type of chromatography with MS detection in MS mode results in an even higher LoD and lower precision. From the observations made in this work, it can be concluded that for the analysis of AAs, MS detection in SRM mode would be applicable in

cases when very complicated matrices may be found, which cannot be eliminated during the sample preparation stage.

Reversed phase mixed mode chromatography is an alternative to reversed phase ion pairing chromatography and potentially more suitable because of the use of 'MS friendly' mobile phases. Due to the mechanism of separation involved in reversed phase mixed mode chromatography, peaks are considerably broader, and therefore, contribution to the noise is higher, which results in higher LoD, than for the other methods. The advantage of the method is that it requires no derivatisation and is compatible with any of the buffers used for the sample preparation.

4.5 References

- (1) Swartz, M. E.; Krull, I. S. *Lc Gc North America* **2006**, 24, 480.
- (2) Cohen, S. A.; Michaud, D. P. *Analytical Biochemistry* **1993**, 211, 279.
- (3) Edmond de Hoffman, V. S. MS. *Principles and Applications*; Third edition ed., 2007.
- (4) Castiglioni, S.; Zuccato, E.; Crisci, E.; Chiabrando, C.; Fanelli, R.; Bagnati, R. *Analytical Chemistry* **2006**, 78, 8421.
- (5) Cheng, C.; Tsai, H. R. *Analytica Chimica Acta* **2008**, 623, 168.
- (6) Dai, G. W.; Wei, X. H.; Liu, Z. F.; Liu, S. J.; Marcucci, G.; Chan, K. K. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **2005**, 825, 201.
- (7) Chen, Y. T.; Ling, Y. C. *Journal of MS* **2002**, 37, 716.
- (8) Fortin, T.; Salvador, A.; Charrier, J. P.; Lenz, C.; Lacoux, X.; Morla, A.; Choquet-Kastylevsky, G.; Lemoine, J. *Molecular & Cellular Proteomics* **2009**, 8, 1006.
- (9) Garcia-Reyes, J. F.; Hernando, M. D.; Ferrer, C.; Molina-Diaz, A.; Fernandez-Alba, A. R. *Analytical Chemistry* **2007**, 79, 7308.
- (10) Guzsvany, V.; Madzgalj, A.; Trebse, P.; Gaal, F.; Franko, M. *Environmental Chemistry Letters* **2007**, 5, 203.
- (11) Hanke, S.; Besir, H.; Oesterhelt, D.; Mann, M. *Journal of Proteome Research* **2008**, 7, 1118.
- (12) Lee, S. J.; Choi, J. Y.; Park, S.; Il Chung, J.; Jin, J. S.; Sung, N. J.; Bae, D. W.; Shin, S. C. *Journal of Food Composition and Analysis* **2010**, 23, 314.
- (13) Mattarozzi, M.; Musci, M.; Careri, M.; Mangia, A.; Fustinoni, S.; Campo, L.; Bianchi, F. *Journal of Chromatography A* **2009**, 1216, 5634.
- (14) Moore, S.; Smyth, W. F.; Gault, V. A.; O'Kane, E.; McClean, S. *Rapid Communications in MS* **2009**, 23, 1747.
- (15) Owens, J.; Hok, S.; Alcaraz, A.; Koester, C. *Journal of Agricultural and Food Chemistry* **2009**, 57, 4058.
- (16) Sealey-Voyksner, J. A.; Khosla, C.; Voyksner, R. D.; Jorgenson, J. W. *Journal of Chromatography A* **2010**, 1217, 4167.

- (17) Selvan, P. S.; Pal, T. K. *Journal of Pharmaceutical and Biomedical Analysis* **2009**, 49, 780.
- (18) Skrabakova, Z.; O'Halloran, J.; van Pelt, F.; James, K. J. *Rapid Communications in MS* **2010**, 24, 2966.
- (19) Bonizzi, I.; Buffoni, J. N.; Feligini, M. *Journal of Chromatography A* **2009**, 1216, 165.
- (20) Izrael-Tomasevic, A.; Phu, L.; Phung, Q. T.; Lill, J. R.; Arnott, D. *Journal of Proteome Research* **2009**, 8, 3132.
- (21) Breaud, A. R.; Harlan, R.; Di Bussolo, J. M.; McMillin, G. A.; Clarke, W. *Clinica Chimica Acta* **2010**, 411, 825.
- (22) Varanasi, K.; Sridhar, V.; Potharaju, S.; Shraddha, R.; Sivakumar, S. P. N.; Sabapathi, S. K.; Satheeshmanikandan, T. R. S.; Kumar, V. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **2008**, 865, 91.
- (23) Zhang, J. Y.; Yang, G. D.; Lin, R.; Hu, Z. *Biomedical Chromatography* **2011**, 25, 450.
- (24) Wille, K.; Claessens, M.; Rappé, K.; Monteyne, E.; Janssen, C. R.; De Brabander, H. F.; Vanhaecke, L. *Journal of Chromatography A* **2011**, 1218, 9162.
- (25) Rodríguez-Gonzalo, E.; García-Gómez, D.; Carabias-Martínez, R. *Journal of Chromatography A* **2011**, 1218, 9055.
- (26) Janzen, N.; Sander, S.; Terhardt, M.; Steuerwald, U.; Peter, M.; Das, A. M.; Sander, J. *Steroids* **2011**, 76, 1437.
- (27) Vial, J.; Jardy, A. *Hrc-Journal of High Resolution Chromatography* **1999**, 22, 217.
- (28) Oreiro-Garcia, M. T.; Vazquez-Illanes, M. D.; Sierra-Paredes, G.; Sierra-Marcuno, G. *Biomedical Chromatography* **2005**, 19, 720.
- (29) Bosch, L.; Alegria, A.; Farre, R. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **2006**, 831, 176.
- (30) Anscombe, F. J. *American Statistician* **1973**, 27, 17.
- (31) Miller J.N., M. J. C. *Statistics and Chemometrics for Analytical Chemistry*; Pearson Education Ltd, 2005.
- (32) Hibbert, D. B. *Accreditation and Quality Assurance* **2005**, 10, 300.
- (33) Huber, W. *Accreditation and Quality Assurance* **2004**, 9, 726.
- (34) Bruggemann, L.; Quapp, W.; Wennrich, R. *Accreditation and Quality Assurance* **2006**, 11, 625.
- (35) Mark, H. *Journal of Pharmaceutical and Biomedical Analysis* **2003**, 33, 7.
- (36) Hsieh, E.; Hsiao, C. F.; Liu, J. P. *Journal of Chemometrics* **2009**, 23, 56.
- (37) Castillo, M. A.; Castells, R. C. *Journal of Chromatography A* **2001**, 921, 121.
- (38) Mocak, J.; Bond, A. M.; Mitchell, S.; Scollary, G. *Pure and Applied Chemistry* **1997**, 69, 297.
- (39) Gibbons, R. D.; Grams, N. E.; Jarke, F. H.; Stoub, K. P. *Chemometrics and Intelligent Laboratory Systems* **1992**, 12, 225.
- (40) Araujo, P. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **2009**, 877, 2224.
- (41) Vial, J.; Jardy, A. *Analytical Chemistry* **1999**, 71, 2672.
- (42) Carre, M.; Excoffier, S.; Mermet, J. M. *Spectrochimica Acta Part B-Atomic Spectroscopy* **1997**, 52, 2043.
- (43) Cox, C. *Journal of the Royal Statistical Society Series C-Applied Statistics* **2005**, 54, 63.

5 Chapter Five – Application of AA Analysis to the Study of PTMs

This chapter describes work devoted to the application of AA analysis for the study of substrate modifications produced by 2OG- and Fe (II)- dependent oxygenase catalysis. AA analysis was used to determine the stereochemistry of JmjD6 catalysed lysyl hydroxylation, the regio- and stereo-selectivity of prolyl hydroxylation catalysed by a prolyl hydroxylase from *Monosiga brevicollis* (mbP4H), and for the study of the hBBOX catalysed rearrangement and identification of potential hBBOX substrates. Studies on cysteine modifications as a result of arylamine N-acetyltransferase inhibition are also described.

The projects described in this chapter were carried out in collaboration with the laboratory of Professor Chris Schofield (Department of Chemistry, University of Oxford), unless stated otherwise.

5.1 AA Residue Hydroxylation Stereochemistry

5.1.1 Hydroxylation of Lysyl Residues Catalysed by JmjD6

This project was carried out in collaboration with M. Mantri (Department of Chemistry, University of Oxford).

The regio- and stereo-specificity of product formation varies between 2OG- and Fe(II)- dependent oxygenases that catalyse post-translational hydroxylations. Aspartyl- and asparaginyl- residues of epidermal growth-factor-type domain proteins are hydroxylated to produce 3R-products¹. In contrast, in HIF and ankyrin repeat domains, hydroxylation of asparaginyl-residues results in the 3S products^{2,3}. Collagen prolyl hydroxylases catalyse the 3S- and 4R- stereochemistry, and collagen lysyl hydroxylases catalyse 5R-lysyl hydroxylation⁴ (see Chapter One).

JmjD6 has been reported to catalyse the hydroxylation of lysyl-residues at the C-5 position in U2AF65 and LUC7L2, a protein important in pre-mRNA splicing⁵. It was hypothesised that the stereochemistry of JmjD6 catalysed hydroxylation of the LUC7L2₂₆₇₋₂₇₈ peptide might result in 5*R*-hydroxylysyl as found in collagen. In order to test this hypothesis, a hydroxylated LUC7L2₂₆₇₋₂₇₈ peptide, produced by JmjD6 catalysis, was compared with synthetic LUC7L2₂₆₇₋₂₇₈ peptide with a 2*S*,5*R*-hydroxylysyl residue 269 (NPK(OH)RSRSREHRR) by AA analysis. LUC7L2₂₆₇₋₂₇₈ was synthesised and incubated with recombinant human JMJD6 produced in *Escherichia coli*.

Reversed phase chromatography of AQC-derivatised AAs, which was shown to provide separation of modified AA diastereomers (see Chapter Two), was chosen to establish the stereochemistry of JmjD6 catalysed lysyl-hydroxylation. Several adjustments had to be made to the AQC method. In order to obtain sufficient resolution between the two 5-hydroxylysine diastereomers, the gradient was modified from 6% Phase A per min to 1% Phase A per min (Figure 5-1). Thus, it was possible to improve the resolution of the two 5-hydroxylysine diastereomers to $R = 1.3$ from initial value of $R = 0.6$. A flow-splitter (split ratio 1:1) was also introduced before the MS detector, which enabled approximately a 2 fold increase in the signal intensity (Figure 5-2).

Due to the presence of two amino groups in lysine and its modified forms these compounds are doubly derivatised (except for the di- and tri-methylated lysine) by AQC. Indeed when using a fourfold excess of the derivatising reagent (Chapter Eight) singly derivatised species were not found. The same result was observed for hydroxylysine under these analytical conditions used.

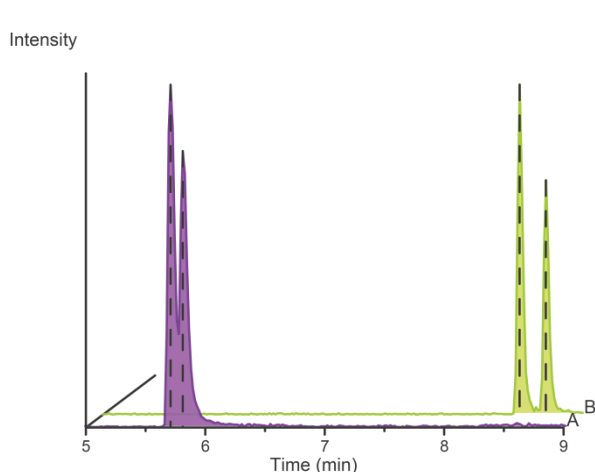


Figure 5-1. Gradient adjustment

(XIC; $m/z = 252.05 \pm 0.5$)

- A. Original gradient; decrease in Phase A 6% per min.
 B. Adjusted gradient; decrease in Phase A 1% per min.

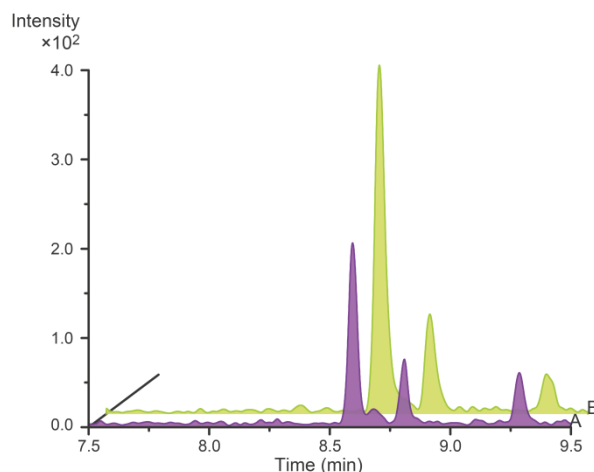


Figure 5-2. Flow splitter application

(XIC; $m/z = 252.05 \pm 0.5$)

- A. Without splitter
 B. Using the splitter (split ratio 1:1)

At the pH of the mobile phase used ($\text{pH} < 7$), both amino groups of lysine should be protonated (the pK_a value of the lysine amino groups are 9.2 and 10.7); however, it appears that approximately 30% of the species are singly charged (Figure 5-4). Because of the higher intensity of the double charged species $m/z = 252.05$, it was chosen for the analysis.

It has been previously shown that hydroxylysine undergoes epimerisation under the conditions of acid hydrolysis⁶ (excess 6 M HCl, 24 hours, 110°C). This observation was confirmed using collagen as a test material. Therefore, a double enzymatic digestion procedure, using trypsin ($2\ \mu\text{L}$ of $100\ \mu\text{g}\cdot\text{mL}^{-1}$, 2 hours, 37°C) followed by carboxypeptidase ($1\ \mu\text{L}$ of $5\ \text{mg}\cdot\text{mL}^{-1}$, overnight, 37°C), was developed using collagen to produce the hydroxyl-lysine. Both trypsin and carboxypeptidase were removed using a 10 kDa molecular weight cut-off filter (14000 rpm, 10 min); see Chapter Eight for the detailed protocol. It can be seen (Figure 5-5, B) that enzymatic hydrolysis of collagen provides only one peak corresponding to the 2S,5R-hydroxylysine standard (Figure 5-5, C). In contrast, acidic hydrolysis produces three peaks (Figure 5-5, A). Two of the peaks were identified as the two sets of 5-hydroxylysine diastereomers based on their m/z value and t_R . It was hypothesised

that the largest peak could correspond to a different hydroxylation site, which might be created as a result migration, possibly by the mechanism shown in Figure 5-3. Therefore, in order to establish the stereochemistry of 5-hydroxylysine acid hydrolysis procedure is unsuitable because of isomerisation.

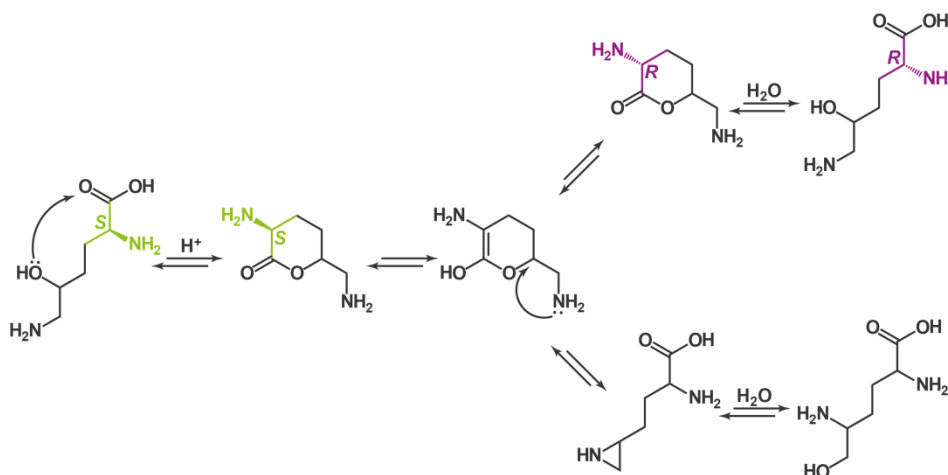


Figure 5-3. Possible mechanism of HyL isomerisation at C-2 and C-5 under the HCl hydrolysis conditions

In order to identify the product stereochemistry resulting from JmjD6 catalysed lysyl hydroxylation, LUC7L2₂₆₇₋₂₇₈ was incubated with JmjD6 in the presence of 2OG, Fe(II) and ascorbate⁷ (Chapter Eight). The results of the assay mixtures analysis are shown in Figure 5-6. It can be seen that JmjD6 catalysed lysyl-hydroxylation results in the product with the 2S,5S stereochemistry, which differs from the collagen lysyl hydroxylases, which give a products with the 2S,5R stereochemistry⁴.

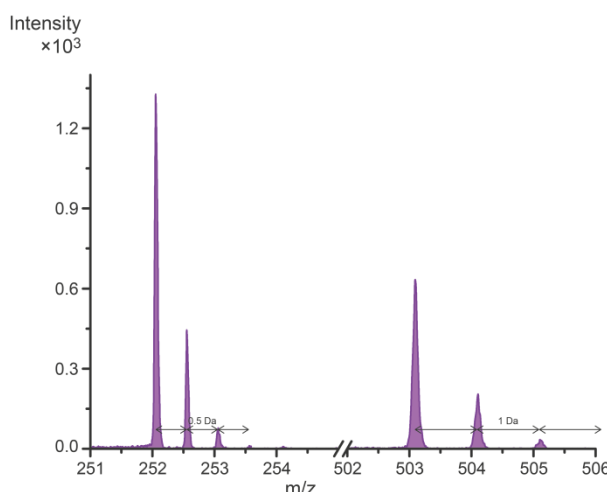


Figure 5-4. Singly charged and doubly charged doubly derivatised 5-hydroxylysine

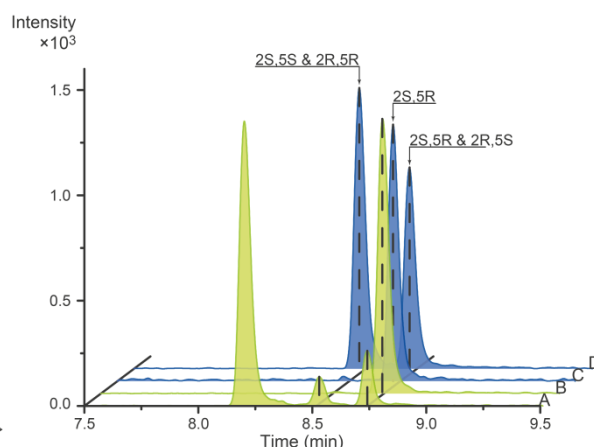


Figure 5-5. Effect of hydrolysis on the 5-hydroxylysine stereochemistry Test; (XIC; m/z = 252.05 ± 0.5)

- A. Collagen HCl hydrolysis
- B. Collagen enzymatic hydrolysis
- C. 2S,5R-HyL standard
- D. HyL racemate standard

It was also shown that JmjD6 catalyses self-hydroxylation (*i.e.* hydroxylates itself in the absence of substrate) of lysyl residues⁸. The 5-hydroxylysine produced by self-hydroxylation has the same stereochemistry as LUC7L2₂₆₇₋₂₇₈ (Figure 5-6). It can be seen in Figure 5-7, that in the absence of Fe(II) and 2OG, apparent hydroxylation still occurs, but at a lower level. This can be explained by the fact that, as with some other 2OG oxygenases, recombinant JmjD6 may co-purify with trace amounts of iron and, possibly, 2OG. Therefore, iron may be present in the assay mixture even when no iron is added, thus catalysing the hydroxylation reaction.

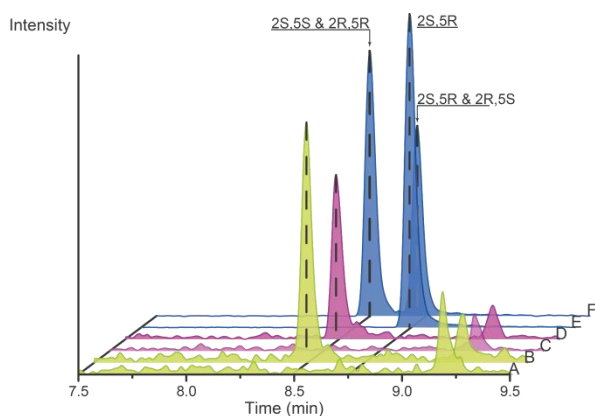


Figure 5-6. JmjD6 catalysed hydroxylation of LUC7L₂₆₇₋₂₇₈ and JmjD6 self-hydroxylation; (XIC; $m/z = 252.05 \pm 0.5$)

- A. JmjD6 reference
- B. JmjD6 assay
- C. LUC7L₂₆₇₋₂₇₈ incubation without JmjD6
- D. LUC7L₂₆₇₋₂₇₈ incubation with JmjD6
- E. 2S,2R-hydroxylysine standard
- F. Diastereomeric mixture of hydroxylysine standards

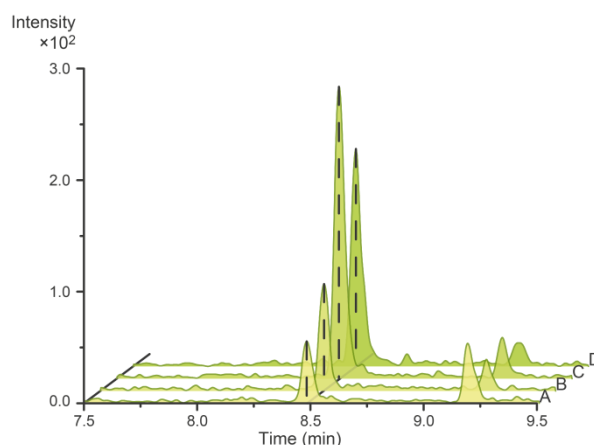


Figure 5-7. JmjD6 self hydroxylation assay; (XIC; $m/z = 252.05 \pm 0.5$)

- A. incubation without Fe(II) and 2OG
- B. incubation without 2OG
- C. incubation without Fe(II)
- D. incubation with Fe(II) and 2OG

5.1.2 Hydroxylation of Proline Catalysed by mbP4H

The project was carried out in collaboration with A. Boleininger (Department of Chemistry, University of Oxford).

Monosiga brevicollis is a unicellular eukaryotic organism which belongs to the class of choanoflagellates⁹. Recent research into the nuclear genome of *Monosiga brevicollis* suggests that the choanoflagellates are the closest unicellular relative of metazoans⁹. Analysis of the *Monosiga brevicollis* genome has shown that it may produce a homologue of the prolyl-4-hydroxylases (PHDs) responsible for oxygen sensing in metazoans. The putative protein found as a result of the genome sequence analysis was labelled *Monosiga brevicollis* prolyl-4-hydroxylase (mbP4H). Research into the function of mbP4H may not only contribute to our understanding of the biology of the choanoflagellates, but also may cast light on the origin of the 2OG oxygenases themselves as oxygen sensors.

Trichoplax adhaerens, which is the simplest animal, was chosen as a comparison organism due to the fact that hypoxia sensing in *Trichoplax adhaerens* has been previously studied; the

Trichoplax adhaerens prolyl hydroxylase (taPHD) has been identified as well as its natural HIF- α substrate, which contains oxygen dependent degradation domain (taODD)¹⁰.

mbP4H was incubated with taODD (a potential substrate) in the presence of Fe(II), 2OG and ascorbate overnight. A +16 Da peptide mass-shift was observed implying hydroxylation of taODD. However, it was also considered possible that the +16 Da mass shift may have been caused by aspartyl, asparaginyl or lysyl hydroxylation. To verify which of the residues was hydroxylated, MS-MS analysis using a Waters SYNAPT mass spectrometer was carried out (for data see A. Boleininger thesis) on the taODD peptide after incubation with mbP4H. As a positive control the baODD was also incubated with taP4H, 2OG and Fe(II) and the peptide product of this incubation was analysed by MS-MS.

It was found for the mbP4H peptide product that hydroxylation occurs in the same peptide fragment as for hydroxylation catalysed by taPHD. The results clearly indicated that the +16 Da shift was the result of either a prolyl- or phenylalanyl-residue hydroxylation. To clarify which residue the hydroxylation occurred on, AA analysis using AQC derivatisation, following acid hydrolysis was carried out. MS detection in scan mode was used. Two peaks corresponding to a hydroxylated proline were identified (t_R = 1.7 min and t_R = 3.6 min, Figure 5-8). The results were identical for incubation with the both mbP4H enzyme and taPHD reference enzyme. It appears that the level of peptide hydroxylation is 2 fold higher with the taPHD reference enzyme incubation. Such a difference could have resulted from experimental error or the kinetic properties of the enzymes. No peaks corresponding to hydroxylated phenylalanine were identified.

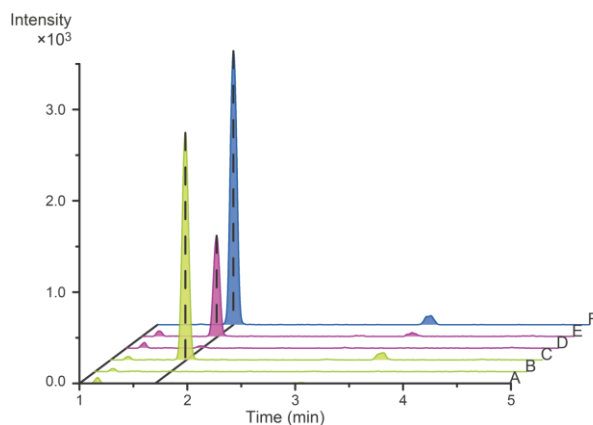


Figure 5-8. AA analysis of taODD peptide substrate by taPHD and mbP4H enzymes

- A. Incubation of taODD substrate in presence of ascorbate, 2OG, and Fe, without enzyme
- B. Incubation of taPHD enzyme in presence of ascorbate, 2OG, and Fe, without substrate
- C. Incubation of taODD substrate in presence of ascorbate, 2OG, and Fe, with taPHD enzyme
- D. Incubation of mbP4H enzyme in presence of ascorbate, 2OG, and Fe, without substrate
- E. Incubation of taODD substrate in presence of ascorbate, 2OG, and Fe, with mbP4H enzyme
- F. 4-HyP standard subjected to HCl hydrolysis

A peak corresponding to the retention time for hydroxyproline was also present in the incubation of mbP4H enzyme with Fe(II), 2OG and ascorbate, but without substrate (Figure 5-10). Therefore, the newly discovered mbP4H enzyme, unlike the taPHD reference enzyme, might undergo self hydroxylation (Figure 5-9 and Figure 5-10). However, this finding requires further verification analogous to that carried out for JmjD6.

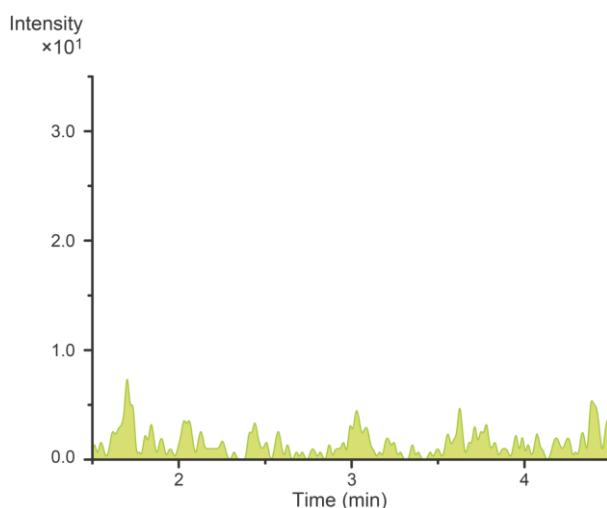


Figure 5-9. Incubation of taPHD enzyme in presence of ascorbate, 2OG, and Fe, without substrate

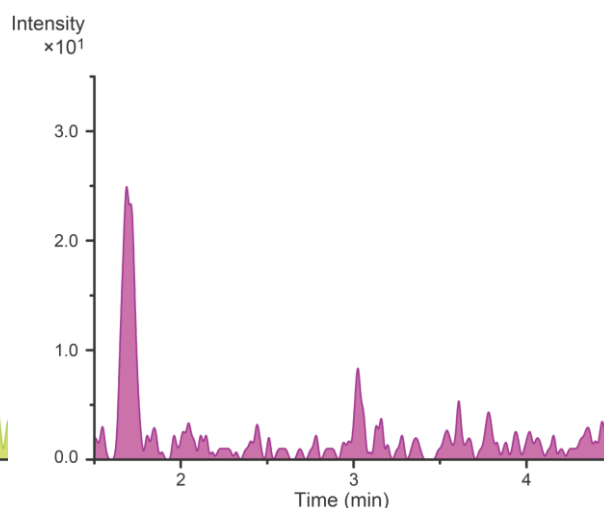


Figure 5-10. Incubation of mbP4H enzyme in presence of ascorbate, 2OG, and Fe, without substrate

AA analysis also allowed identification of the stereochemistry of mbP4H catalysed prolyl hydroxylation. Initially, 4*R*-hydroxyproline was suggested. In order to confirm this, AA analysis following the HCl hydrolysis was attempted. For the direct comparison between hydrolysed peptide and standards a 4*R*-hydroxyproline standard was also subjected to the HCl hydrolysis. It can be observed (see Figure 5-11) that as a result of acid hydrolysis some of the 2*S*,4*R*-hydroxyproline is converted to 2*S*,4*S*- or 2*R*,4*R*-hydroxyproline. The peak corresponding to the 2*S*,4*S*- or 2*R*,4*R*-hydroxyproline can also be seen after AA analysis of the assay mixture (Figure 5-8) and should be considered to be an artefact resulting from HCl hydrolysis. It is perhaps more likely, that epimerisation occurs at C-4 rather C-2; AA analysis, however, does not provide this information.

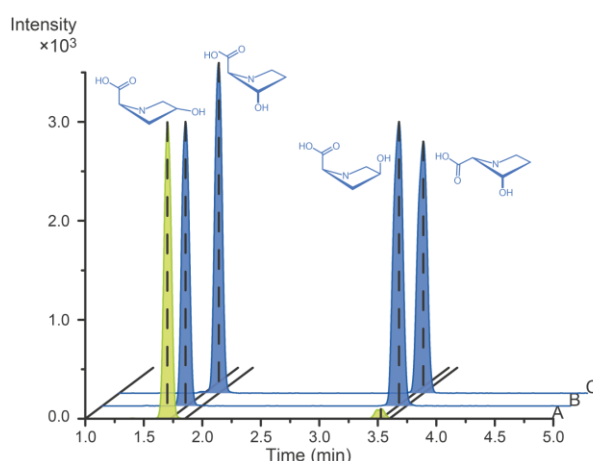


Figure 5-11. Analysis of AQC-derivatised hydroxy prolines

- A. 2*S*,4*R*-HyP HCl Hydrolysis test
- B. Mixture of 2*S*,4*R*-HyP (left peak) and 2*S*,4*S*-HyP (right peak)
- C. Mixture of 2*S*,3*S*-HyP (left peak) and 2*S*,3*R*-HyP (right peak)

5.2 Verification of Arylamine *N*-Acetyltransferase Inhibition Mechanism

This project was carried out in collaboration with A. Abuhammad (DPhil student, E. Sim Laboratory, Department of Pharmacology, University of Oxford).

The laboratory of E. Sim is studying inhibitors developed against arylamine *N*-acetyltransferase (NAT) of *Mycobacterium marinum*, a close analogue of the enzyme found

in *Mycobacterium tuberculosis*. Several analogues of an initial hit compound have been tested for NAT inhibition. A covalent mechanism of inhibition was proposed in which a modification of the active-site cysteine was proposed (Figure 5-12). In order to test this hypothesis, free cysteine was incubated overnight at ambient temperature with 3-benzoyl-1-methyl-4-phenyl-4-piperidinol (A1), which was provided by my collaborators. After incubation of cysteine with A1, derivatisation of the products with AQC was carried out followed by LC-MS analysis. The analysis confirmed a compound was formed with a mass corresponding to the mass of cysteine plus a fragment of the inhibitor compound (Figure 5-12). Reference compound A1 and cysteine were analysed separately (Figure 5-13).

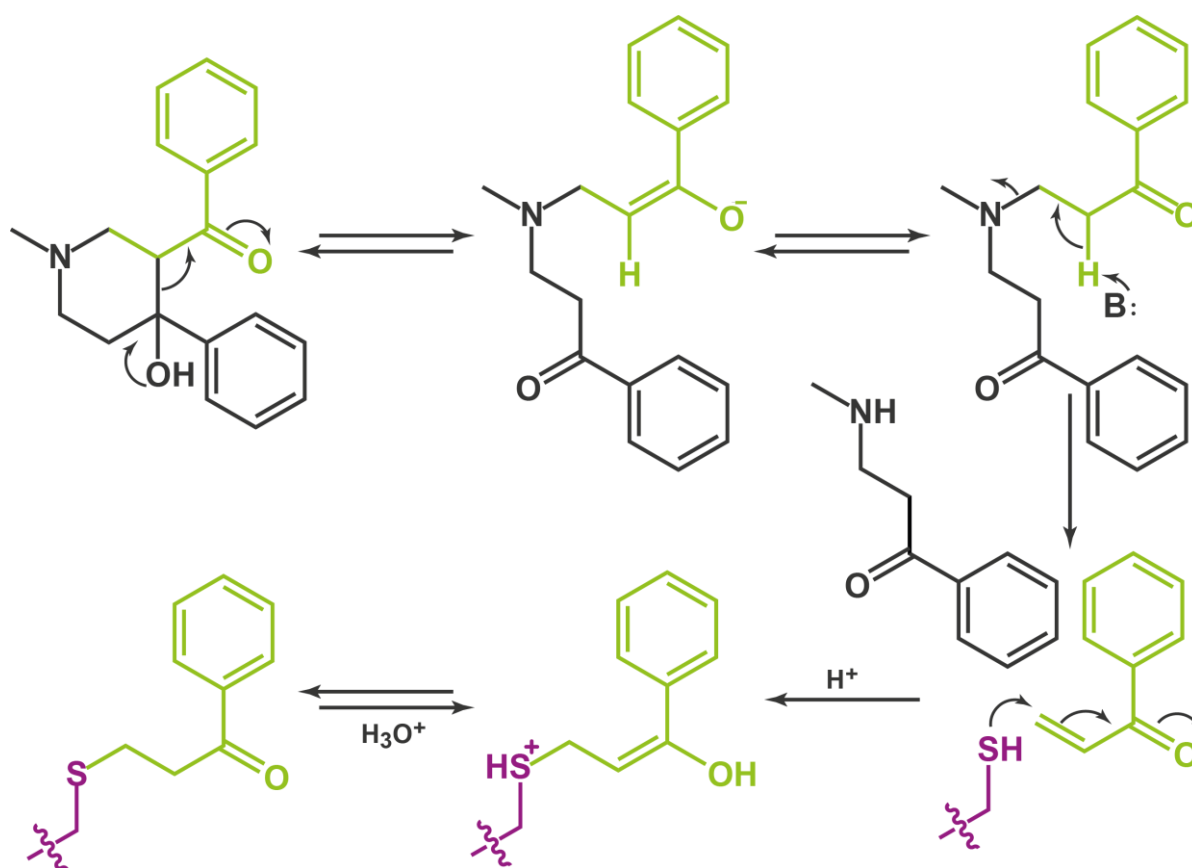


Figure 5-12. Proposed mechanism for the inhibition by 3-benzoyl-1-methyl-4-phenyl-4-piperidinol

Figure was adapted from D.Phil thesis of Dr Areej Abuhammad

The spectra of the A1-cysteine peak (Figure 5-13, A) supports the structure assigned to the adduct (Figure 5-15). This illustrates how AA analysis method can be extended from the identification of AAs to the identification of adducts of the AAs with organic ligands.

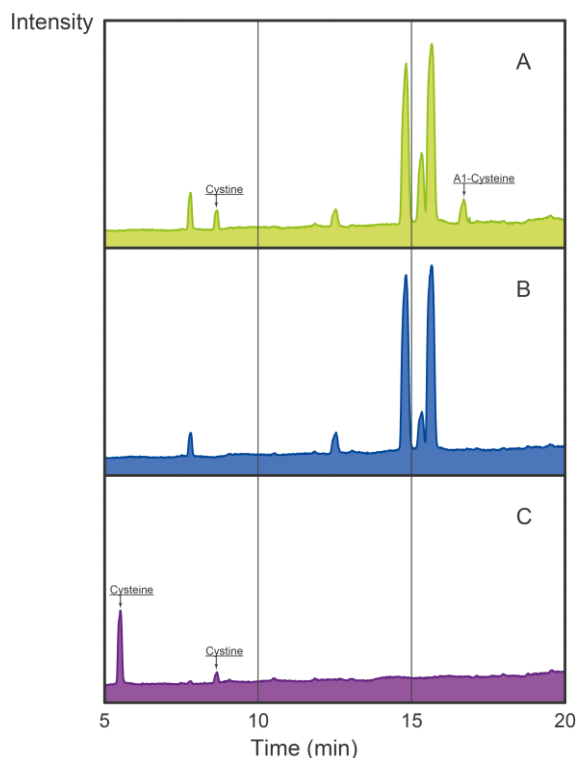


Figure 5-13. Chromatograms of the inhibition test set

- A. Assay of cysteine with A1
- B. Assay of A1 inhibitor without cysteine
- C. Assay of cysteine without A1 inhibitor



Figure 5-14. Mass spectra of cysteine and cystine

- A. Spectrum of AQC-derivatised cystine
 - B. Spectrum of AQC-derivatised cysteine
- Where **R** is the AQC fragment

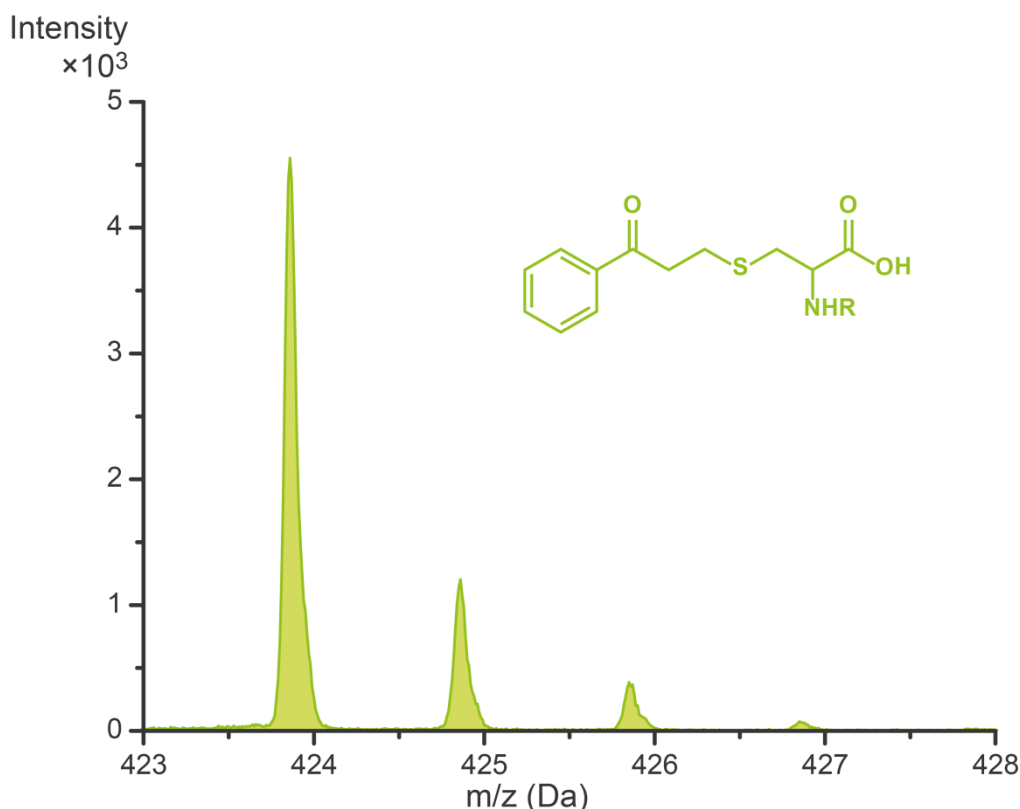


Figure 5-15. Spectrum of AQC-derivatised A1-cysteine adduct; where R is the AQC fragment

A structure related to the A1-cysteine adduct is phenylalanine. However, three extra carbon atoms in the carbon backbone of A1 should increase retention time significantly compared to phenylalanine; the presence of the extra polar ketone group is expected to reduce the retention time. On balance, the retention of such an adduct may indeed be close to that of phenylalanine (Figure 5-13).

Interestingly, a substantially different retention time for cysteine from that reported value was observed (Figure 5-13, C). Formation of a disulfide bridge under alkali conditions results in conversion of cysteine (Figure 5-14, B) to cystine (Figure 5-14, A), which gives a peak at 8.7 min. The retention time of the cysteine is expected to be closer to that of serine (2.7 min under the same experimental conditions), as a result of their similar structures. However, the lower polarity of the thiol group in comparison to the hydroxyl group. The retention time of cysteine was found to be 5.5 min under the conditions of analysis.

5.3 Study of an hBBOX Catalysed Rearrangement

The project was carried out in collaboration with L. Henry (Department of Chemistry, University of Oxford), who carried out incubations of 3-(2,2,2-trimethylhydraziniumyl)-propionate with hBBOX and I. Leung (Department of Chemistry, University of Oxford), who carried out NMR analyses.

L-Carnitine is responsible for transporting long-chain fatty acids into mitochondria. In animals, *L*-carnitine is obtained both from the diet and as a result of biosynthesis. The last step of *L*-carnitine production is catalysed by γ -butyrobetaine hydroxylase (BBOX), which belongs to the family of 2OG- and Fe(II)- dependent oxygenases¹¹ (see Chapter One).

The fatty acid degradation pathway is compromised under ischemic conditions. An inadequate supply of oxygen results in the inhibition of fatty acid oxidation and accumulation of their metabolites. Under certain conditions treatment with exogenous *L*-carnitine can facilitate the fatty acid degradation pathway; however, in advanced ischemia, *L*-carnitine-dependent oxidation of fatty acids is inhibited at the stage of acyl-carnitine formation^{12,13}. In this case, inhibition of *L*-carnitine production is proposed as being therapeutically beneficial. 3-(2,2,2-trimethylhydraziniumyl)propionate (THP, Figure 5-16 A) – also known as MET-88, meldonium or quaterine, and marketed under the brand name Mildronate – is used clinically for the treatment of angina and myocardial infarction.

In studies on the reaction of human γ -butyrobetaine hydroxylase (hBBOX) with THP (Figure 5-16, A) several unanticipated compounds were observed. These unanticipated products included hydrated malonate semialdehyde (Figure 5-16, B), formaldehyde, which forms a characteristic adducts with the amino group of Tris buffer (Figure 5-16, C), and 3-amino-4-(methyldamino)butanoic acid¹⁴ (Figure 5-16, D).

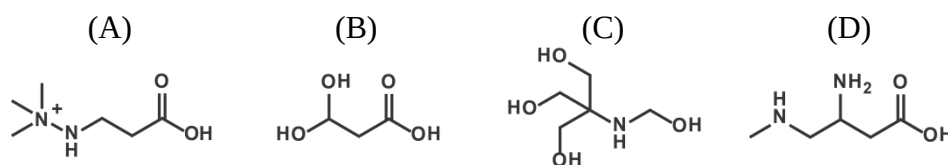


Figure 5-16. Products of hBBOX interaction with THP

3-Amino-4-(methylamino)butanoic acid (Figure 5-16, D) appears to result from a rearrangement product involving a C-C bond formation¹¹. This phenomenon has not been described before for 2OG oxygenases. It was therefore desirable to confirm the proposed structure of this compound. In order to identify the structure NMR studies were carried out. 1D TOCSY and 2D COSY experiments indicated that the structure has a 4 carbon atom chain (contrary to the 3 carbon atom chain of the THP), which could result either from rearrangement or, possibly from the reaction with the cosubstrate 2OG. ¹H-¹³C HMBC suggested the presence of a terminal carboxylic acid group. The ¹³C chemical shift of the putative C-3 (70 ppm) of the product indicated that it was likely to be connected to a heteroatom, probably nitrogen. The transfer of an additional carbon atom from 2OG was ruled out using ¹³C labelled 2OG in the hBBOX assay; the ¹³C labels were only found in succinate.

In order to investigate the origin of the ‘extra’ carbon atom in the backbone of the product 3-amino-4-(methylamino)butanoic acid as compared with the initial compound THP, THP with a single ¹³C labelled methyl group was used in the assay with hBBOX. The three N-methyl groups on THP are chemically equivalent. Upon rearrangement to 3-amino-4-(methylamino)butanoic acid one methyl group remains on nitrogen, one is removed as formaldehyde, and one contributes to the elongation of the carbon backbone. Therefore, if the rearrangement proceeds as predicted, the probability that the ¹³C label will remain in the final product is ⅓, and the probability that it leaves as formaldehyde is ⅓. Analysis by MS should

reveal a consistent change in the isotopic pattern associated with molecular ion of 3-amino-4-(methylamino)butanoic acid. Since the probability of the label remaining in the molecule is $\frac{2}{3}$ we should observe a ratio of 1:2 in the molecular ion peak corresponding to the unlabelled species and the ^{13}C labelled species (at +1 Da) respectively. Experimental analysis produced this result, thus confirming the idea that rearrangement takes place as suggested above (see Figure 5-17).

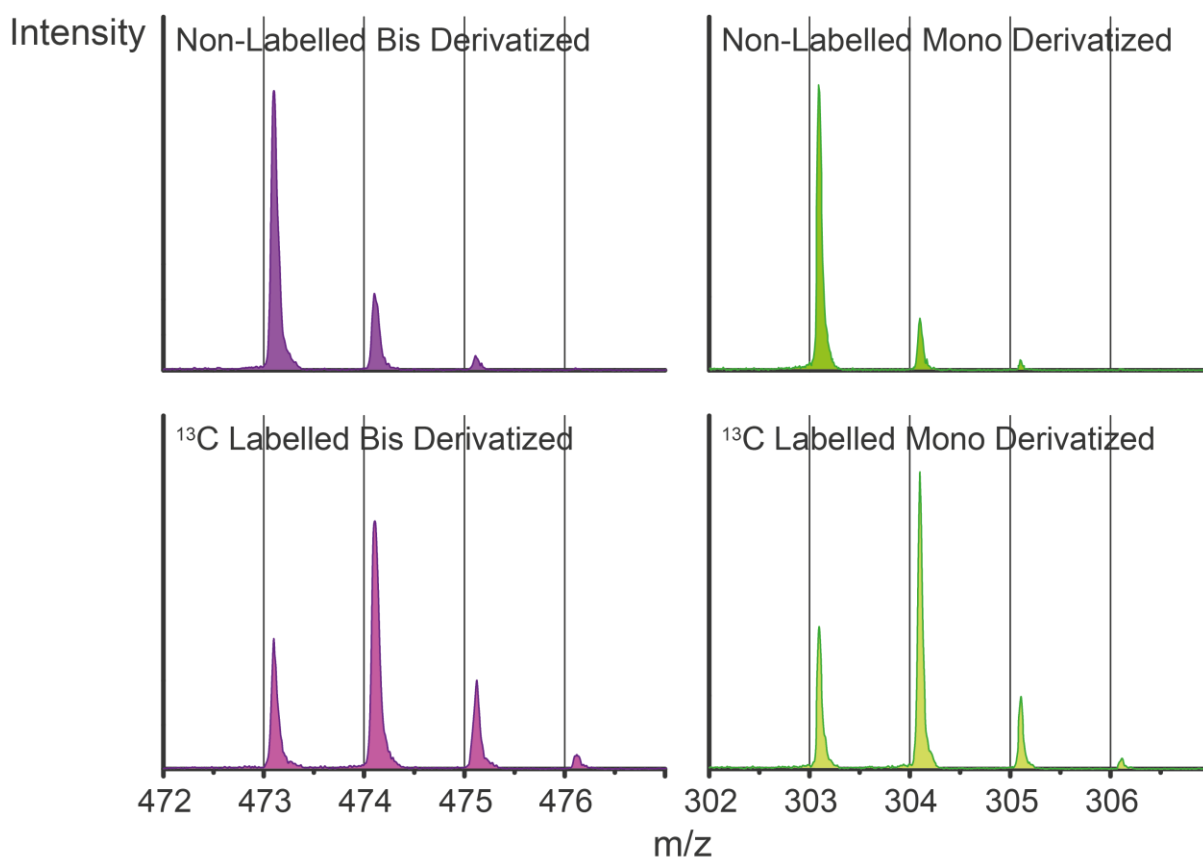


Figure 5-17. Spectra of the ^{13}C labelled and non-labelled mono- and bis- derivatised 3-amino-4-(methylamino)butanoic acid

In order to study THP and the rearrangement compounds LC-MS analyses of AQC derivatives were carried out. The reaction mixture contained high concentrations of Tris buffer, which causes problems for the analysis. Tris buffer reacts with the AQC derivatising agent due to the presence of the Tris primary amino group, which compromises the derivatisation of the target analyte. It is difficult to remove the Tris buffer from the analytical

mixture as it is present at several hundreds fold excess (as compared with the concentration of the hBBOX enzyme and THP substrate). One way to circumvent this problem is to use an excess of the derivatising reagent. However, this was found to affect the chromatographic analysis in several detrimental ways. First, a large tailing chromatographic peak was observed at the retention time of approximately three minutes, corresponding to the derivatised Tris buffer (Figure 5-18). The tailing might result from three hydroxyl groups present in the Tris molecule, interacting with the stationary phase hydroxyl groups due to hydrogen bonding. Second, the retention of compounds eluting immediately after Tris can be affected by the presence of a different component in the mixture. In this case the retention time of the analyte was not affected, because it eluted after the Tris peak tail. Finally, even though an excess of derivatising agent was added, some of the target compounds were not fully derivatised, *i.e.* compounds having two amino groups were present in both double and single derivatised form. This was the case with the target analyte (3-amino-4-(methylamino)butanoic acid). The chromatograms of the assay mixture of THP treated with hBBOX enzyme, and the enzyme only control mixture, are presented in Figure 5-18. Mono- and bis- derivatised products are labelled accordingly. Spectra of the mono- and bis- derivatised products are presented in Figure 5-17.

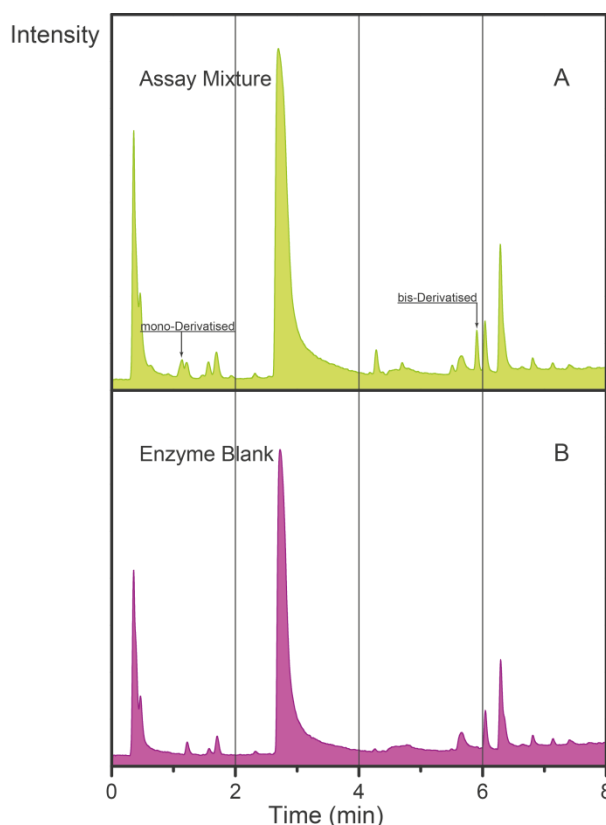


Figure 5-18. Chromatograms for THP Treated with hBBOX and hBBOX alone

- A. Chromatogram of the hBBOX assay mixture with THP substrate
- B. Chromatogram of the hBBOX assay mixture without THP substrate

It was necessary to estimate the retention time of 3-amino-4-(methylamino)butanoic acid. Since a standard for this compound was not initially available, it was considered that an estimated retention time might serve as an indirect tool for compound identification. Possibly, the closest available structural analogue of this compound is methylated α,γ -diaminobutyric acid. The retention index for the AQC-derivatised α,γ -diaminobutyric acid using RP chromatography was found to be 392. Given the methylation increment is 20 ± 10 RI units (Chapter Three), 3-amino-4-(methylamino)butanoic acid is predicted to have a RI 412 ± 10 . The experimental value of the index was found to be 405.

Another structural analogue of the 3-amino-4-(methylamino)butanoic acid was ornithine. The β -amino group still allows the formation of an inter-molecular hydrogen bond, which can have a substantial effect on the retention time. The methylation of the amino group might be

predicted to have a similar effect to that of the carbon backbone elongation. Therefore, we would expect very close elution of the two compounds. Ornithine was analysed separately as a reference, and the results are presented in Figure 5-19.

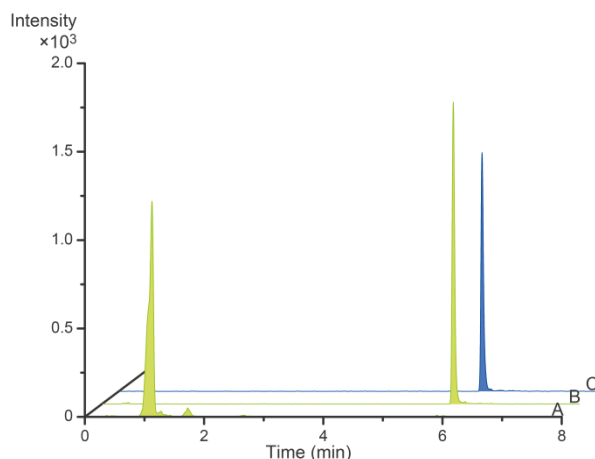


Figure 5-19. Extracted ion chromatogram analytes and ornithine as a reference compound.

- A. mono-derivatised 3-amino-4-(methylamino)butanoic acid (XIC, $m/z = 303$)
- B. bis-derivatised 3-amino-4-(methylamino)butanoic acid (XIC, $m/z = 473$)
- C. bis-derivatised ornithine (XIC, $m/z = 473$)

It was possible to support the mechanism of the 3-amino-4-(methylamino)butanoic acid as a result of rearrangement and new C-C bond formation. The major obstacle in the analysis was the presence of Tris buffer. The use of RI allowed prediction of the t_R of the 3-amino-4-(methylamino)butanoic acid, thus supporting its proposed structure.

5.4 Identification of Potential Substrates of hBBOX

Stereoselective hydroxylation of small molecules is a considerable synthetic challenge especially with non-activated C-H bonds¹⁵. Several recent reports have emerged on the metal catalysed hydroxylation of non-activated molecules^{16,17}. However to date enzyme catalysed reactions are more common^{18,19}. One aim of my work was to investigate the scope of enzymatic hydroxylation by hBBOX, which is known to stereoselectively catalyse hydroxylation of γ -butyrobetaine at C-3 (Table 5-1). The first stage of the project involved

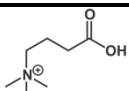
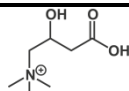
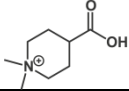
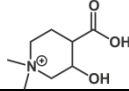
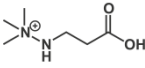
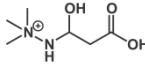
the design and synthesis of a library of potential substrates. Synthetic analogues of natural substrate were prepared by A. Rydzik (Department of Chemistry, University of Oxford).

The set of compounds tested comprised 32 compounds. The list of compounds together with their retention data is provided in Appendix B. The assay conditions for testing the enzyme activity with a particular substrate involved the use of Tris buffer which compromised the derivatisation procedure as explained above. In order to circumvent this potential problem MM chromatography was used (PrimeSep 200), which does not require derivatisation.

Potential substrates were incubated with hBBOX in the presence of 2OG and Fe(II). Samples with enzyme added were compared against the reference without enzyme. For the substrates tested retention indices were calculated (Appendix B), and the hydroxylated species were compared with the predicted t_R values.

Apart from the natural GBB substrate (first compound in Table 5-1), chromatographic peaks with mass corresponding to the mass of the hydroxylated form were found for three other compounds (Table 5-1 and).

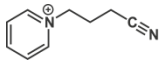
Table 5-1. Substrates of hBBOX and their hydroxylated forms retention data

Structure	t_R	RI	Structure	t_R	RI
	12.18	651		10.91	619
	14.26	676		11.99	631
	13.41	660		11.82	625

For the last compound (Table 5-2) a peak with the mass (164.1 Da) corresponding to the hydroxylated form was found both in the assay mixture and reference sample. The retention index calculated for the suspected hydroxylated form was 612 RI units. It was shown in Chapter Three that introduction of a hydroxyl group causes a decrease in retention by 36 ± 28

units of retention with 95% confidence ($n=4$). This statistic was amended by inclusion of three new compounds (Table 5-1) giving a 37 ± 18 retention index unit decrease with 95% confidence ($n=7$). Standard error decreased as a result of increase of the dataset. The observed retention shift was 117 RI units, which substantially exceeds the confidence limit. Therefore, the suspected compound was omitted from consideration. As a result of retention data analysis of the synthetic library further structure-retention relationship were revealed for reversed phase mixed mode chromatography.

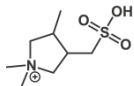
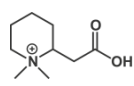
Table 5-2. hBBOX substrate test compound and its retention data.

Structure	t_R	RI
	17.13	729

Screening of the potential hBBOX substrates provided extensive retention data for the modified AAs and allowed us to identify the RI increments of the new structural fragments.

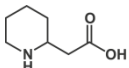
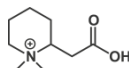
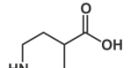
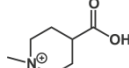
It was shown using a compound with a sulfonic group (Table 5-3) that introduction of a very polar acidic group results in a shift in retention time towards the void volume. It would be expected that an analogous compound with a carboxy group – instead of sulfonic group – would have a RI close to 700 units.

Table 5-3. hBBOX substrate test compound and its retention data.

Structure	t_R	RI	Structure	t_R	RI
	4.35	174		15.53	699

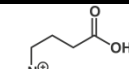
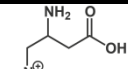
Methylation was shown to cause an increase in t_R (Table 5-4). The combined data allows an estimate of the increment introduced by a methyl group (24 ± 18).

Table 5-4. hBBOX substrates test compounds and its retention data.

Structure	t_R	RI	Structure	t_R	RI
	14.04	673		15.53	699
	11.86	629		14.26	676

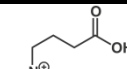
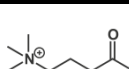
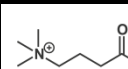
Introduction of a third amino group also results in an increase of retention (Table 5-5). There is no particular trend, however, and this data cannot be easily used for extrapolation purposes. Following on from the discussion in Chapter Three it can be suggested, that interaction of amino groups with acidic groups of stationary phase depends on specific intra-molecular hydrogen bonding, and should be considered individually for every compound.

Table 5-5. hBBOX substrates test compounds and its retention data.

n_C	Structure	t_R	RI	Structure	t_R	RI
12		12.18	651		15.46	697

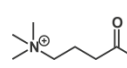
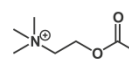
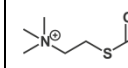
It should be mentioned, that substituting a carboxyl group for a neutral group results in a significant increase in t_R (Table 5-6).

Table 5-6. hBBOX substrates test compounds and its retention data.

Structure	t_R	RI	Structure	t_R	RI	Structure	t_R	RI
	12.18	651		17.30	732		17.55	735

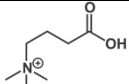
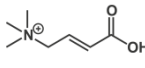
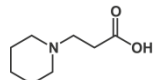
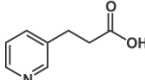
Carbon atom substitution with heteroatoms also appears to have some effect on retention (Table 5-7).

Table 5-7. hBBOX substrates test compounds and its retention data.

Structure	t_R	RI	Structure	t_R	RI	Structure	t_R	RI
	17.30	732		14.46	699		17.38	753

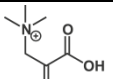
Interestingly the introduction of double bonds has little influence on retention time (Table 5-8).

Table 5-8. hBBOX substrates test compounds and its retention data.

Structure	t_R	RI	Structure	t_R	RI
	12.18	651		12.58	644
	15.25	694		14.81	692

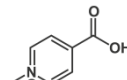
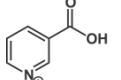
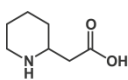
Notably, it appears that structural isomerism does not affect t_R (Table 5-9).

Table 5-9. hBBOX substrate test compound and its retention data.

Structure	t_R	RI
	12.27	647

Aromatic systems significantly reduce the retention (Table 5-10).

Table 5-10. hBBOX substrates test compounds and its retention data.

Structure	t_R	RI	Structure	t_R	RI	Structure	t_R	RI
	5.27	332		6.33	439		14.04	673

5.5 Conclusions

The chromatographic methods developed were shown to be suitable for *in vitro* enzymatic experiments. In order to reduce the interference of matrix components (buffers etc), and to reduce the background noise, additional purification steps specific to the sample type were found to be necessary.

Reversed phase chromatography of AQC-derivatised AAs proved to be a powerful tool for separating modified AA diastereomers. Due to the limited number of the pairs of diastereomers studied it was not possible to identify a general pattern for t_R values for

diastereomers elution. Therefore, it is necessary to use standards for the analysis of unknown diastereomers. The data for HyL diastereomers demonstrates that the 2*S*,5*S*- and 2*R*,5*R*-HyL pair of diastereomers is eluted before 2*S*,5*R*- and 2*R*,5*S*-HyL pair. Collagen HCl lysates were used as a test for resolution of the pairs of HyL diastereomers. HyP presents a special case of elution due to the conformational restrictions imposed by its ring structure. For both 4-HyP and 3-HyP the *cis*-diastereomers, as a result of intra molecular hydrogen bonding, appear to be less polar and eluted after the *trans*-diastereomers.

The choice of chromatographic methods is dependent on the concentration of the target analyte and potential interfering components of the matrix. Buffers, as well as any matrix components with primary or secondary amino groups compete for the AQC derivatising reagent, which has to be taken into account when performing these analyses.

The complexity of the sample also determines the choice of a detection method. UV detection is the least selective and cannot be recommended for samples with complicated matrices. MS detection in SRM mode is the most selective, but it was not suitable for concentrations below 10 μ M using instrumentation available for this research.

The methods developed are applicable for a broad variety of the concentrations ranging from a LoD under 1 μ M to the much higher concentrations over 100 μ M using synthetic mixtures.

It is worth pointing out the robustness of these chromatography methods. No evidence was found for the t_R of compounds studied being significantly affected by the matrix components such as buffers or salts. Buffers present in the sample do affect the degree of ionisation of the analytes; this can lead to signal suppression and a local decrease in the background ions present. Robustness of chromatography, combined with the retention pattern recognition routine, allows estimation of the elution time of unknown compounds which has been shown here to be a powerful and useful tool for PTM analysis.

A broad spectrum of modifications both natural and synthetic can be analysed by the methods developed. When analysing a new modification its polarity, molecular weight and structure have to be considered in order to estimate the appropriateness of a particular LC-MS method for its analysis. The chromatographic methods described (see Chapter Eight for further details) can be adjusted to the specific needs of the research by simple modification of the eluting gradients.

5.6 References

- (1) Przysiecki, C. T.; Staggers, J. E.; Ramjit, H. G.; Musson, D. G.; Stern, A. M.; Bennett, C. D.; Friedman, P. A. *Proceedings of the National Academy of Sciences* **1987**, *84*, 7856.
- (2) Coleman, M. L.; McDonough, M. A.; Hewitson, K. S.; Coles, C.; Mecinović, J.; Edelmann, M.; Cook, K. M.; Cockman, M. E.; Lancaster, D. E.; Kessler, B. M.; Oldham, N. J.; Ratcliffe, P. J.; Schofield, C. J. *Journal of Biological Chemistry* **2007**, *282*, 24027.
- (3) McNeill, L. A.; Hewitson, K. S.; Claridge, T. D.; Seibel, J. F.; Horsfall, L. E.; Schofield, C. J. *Biochem. J.* **2002**, *367*, 571.
- (4) Witkop, B. *Experientia* **1956**, *12*, 372.
- (5) Webby, C. J.; Wolf, A.; Gromak, N.; Dreger, M.; Kramer, H.; Kessler, B.; Nielsen, M. L.; Schmitz, C.; Butler, D. S.; Yates, J. R.; Delahunty, C. M.; Hahn, P.; Lengeling, A.; Mann, M.; Proudfoot, N. J.; Schofield, C. J.; Bottger, A. *Science* **2009**, *325*, 90.
- (6) Hamilton, P. B.; Anderson, R. A. *Journal of Biological Chemistry* **1955**, *213*, 249.
- (7) Mantri, M.; Loik, N. D.; Hamed, R. B.; Claridge, T. D. W.; McCullagh, J. S. O.; Schofield, C. J. *Chembiochem* **2011**, *12*, 531.
- (8) Mantri, M.; Webby, C. J.; Loik, N. D.; Hamed, R. B.; Nielsen, M. L.; McDonough, M. A.; McCullagh, J. S. O.; Bottger, A.; Schofield, C. J.; Wolf, A. *Medchemcomm* **2012**, *3*, 80.
- (9) King, N.; Westbrook, M. J.; Young, S. L.; Kuo, A.; Abedin, M.; Chapman, J.; Fairclough, S.; Hellsten, U.; Isogai, Y.; Letunic, I.; Marr, M.; Pincus, D.; Putnam, N.; Rokas, A.; Wright, K. J.; Zuzow, R.; Dirks, W.; Good, M.; Goodstein, D.; Lemons, D.; Li, W.; Lyons, J. B.; Morris, A.; Nichols, S.; Richter, D. J.; Salamov, A.; Sequencing, J. G. I.; Bork, P.; Lim, W. A.; Manning, G.; Miller, W. T.; McGinnis, W.; Shapiro, H.; Tjian, R.; Grigoriev, I. V.; Rokhsar, D. *Nature* **2008**, *451*, 783.
- (10) Loenarz, C.; Coleman, M. L.; Boleininger, A.; Schierwater, B.; Holland, P. W. H.; Ratcliffe, P. J.; Schofield, C. J. *EMBO Rep* **2011**, *12*, 63.
- (11) Henry, L.; Leung, I. K. H.; Claridge, T. D. W.; Schofield, C. J. *Bioorganic & Medicinal Chemistry Letters* **2012**, *22*, 4975.
- (12) Dambrova, M.; Liepinsh, E.; Kalvinsh, I. *Trends in Cardiovascular Medicine* **2002**, *12*, 275.
- (13) Simkhovich, B. Z.; Shutenko, Z. V.; Meirēna, D. V.; Khagi, K. B.; Mežapūke, R. J.; Molodchina, T. N.; Kalvīņš, I. J.; Lukevics, E. *Biochemical Pharmacology* **1988**, *37*, 195.

- (14) Leung, I. K. H.; Krojer, T. J.; Kochan, G. T.; Henry, L.; von Delft, F.; Claridge, T. D. W.; Oppermann, U.; McDonough, M. A.; Schofield, C. J. *Chemistry & Biology* **2010**, 17, 1316.
- (15) Crabtree, R. H. *Chemical Reviews* **2010**, 110, 575.
- (16) Gunay, A.; Theopold, K. H. *Chemical Reviews* **2010**, 110, 1060.
- (17) Chen, M. S.; White, M. C. *Science* **2010**, 327, 566.
- (18) de Montellano, P. R. O. *Chemical Reviews* **2010**, 110, 932.
- (19) Chen, M. S.; White, M. C. *Science* **2007**, 318, 783.

6 Chapter Six – Development of LC-IRMS for quantitative analysis of PTMs

This chapter describes work on the development and application of a novel method for modified AA identification and quantification using metabolic labelling in cell culture. It describes mixed mode chromatography coupled with isotope ratio MS utilised for the quantitative analysis of modified AAs present at low nanomolar concentrations. The experimental design enables *in vitro*, and potentially *in vivo*, analysis. This chapter initially provides a background to the analytical approach taken, then, discusses the development of the methodological strategy and, finally, the results of its application to the analysis of histone PTMs.

6.1 Method Background

Since the discovery of stable isotopes by Francis William Aston¹ in 1919, by means of a mass spectrograph, isotopes have been a vital part of MS. Isotopic analysis can provide information that is as important and useful for the identification of the molecular ion and pseudo-molecular ion peaks in mass spectra. However, it was not until 1947, that a magnetic sector mass spectrometer was constructed specifically for the purpose of precise determination of isotope ratios², which in 1984 was coupled to a gas chromatography inlet system³. A further important step in the development of this technique was made in the 1990s when the moving transport interface was combined with isotope ratio MS in order to enable an online LC inlet system⁴. A modified design was made commercially available in 2004 by the Thermo Electron corporation⁵.

6.1.1 Isotope Ratio MS

Isotope ratio MS is often used to measure the isotopic composition of light elements, which in practice are C (12, 13), N (14, 15), O (16, 18), S (32, 34) and H (1, 2) – which are

transformed into CO₂, N₂, CO, SO₂ and H₂O gases, respectively⁶ prior to introducing them into the ion source of the mass spectrometer. The work described in this chapter focused on the isotope ratio MS of carbon only. For the analysis of carbon, an isotope ratio mass spectrometer has three detectors; instead of scanning between masses it measures the ion count for three isotopes simultaneously. As a result the technique provides 3 traces for the analysis of the same sample, thus, allowing more counts to be acquired per trace.

The isotope ratio is defined as the ratio of a heavier (normally less abundant isotope) to the abundance of the lighter isotope. In the case of carbon we have the following:

$$R_i = \frac{v_{^{13}\text{C}}}{v_{^{12}\text{C}}}$$

Where R_i is the isotope ratio of a compound, and the v represents the number of atoms of isotope ^{13}C and ^{12}C , respectively, in one mole of a compound.

For example, let us consider the isotope ratio of proline (R_{Pro}) with a single carbon atom 99% labelled – *i.e.* one out of five Pro atoms has the ^{13}C abundance at the level of 99% and the other four have the ^{13}C and ^{12}C abundance at a natural level of about 1.1 and 98.9% respectively.

$$R_{\text{Pro}} = \frac{v_{^{13}\text{C}_{\text{Pro}}}}{v_{^{12}\text{C}_{\text{Pro}}}} = \frac{1 \cdot 0.99 + 4 \cdot 0.011}{1 \cdot 0.01 + 4 \cdot 0.989} = 0.2607161$$

The isotope ratio is a unitless fraction. It is normally quoted with 7 significant figures, and for calculations it is recommended to use double precision, *i.e.* 14 significant digits⁷. The isotopic precision for the contemporary isotope ratio mass spectrometers is reported to be generally at the level of 0.2 ‰ ($\Delta R=0.0002$)⁴.

The isotope ratio (the enrichment of a sample with one particular isotope) is affected by kinetic isotope effects^{8,9} and equilibrium isotope fractionation^{10,11} (differences in the reaction

rates for different isotopes). Thus, the enrichment in one particular isotope can provide indirect information about the processes involved in formation of a particular sample compared to another and enables differentiation between the two samples^{12,13}. This property is utilised in archaeological and environmental studies, in which the conclusions are drawn on the basis of the isotope ratio comparison of samples from different epochs and different geographical locations. In order to make the samples comparable, the use of a reference was suggested¹⁴: The isotope ratio of the sample is compared to that of the now extinct Pee Dee Belemnite (PDB) fossil standard ($R_{PDB}=0.0112372\pm0.00000009$)^{4,6}. The resulting value is termed the delta value⁴:

$$\delta^{13}C_i = \frac{R_i}{R_{PDB}} - 1$$

Subtraction of unity allows the judgement of whether the sample is enriched (sample delta value is greater than zero) or depleted (sample delta value is less than zero) in comparison with the standard. The value is normally expressed in parts per thousand (‰).

For the example of proline labelled with a single ¹³C atom as discussed above, delta value is calculated to be:

$$\delta^{13}C_{Pro} = \frac{R_{Pro}}{R_{PDB}} - 1 = \frac{0.2607161}{0.0112372} - 1 = 22201.16\text{‰}$$

Delta values encountered in nature normally range from -50 to 0 ‰⁴. Excess delta values, like the one mentioned above for the labelled proline, immediately stand out.

Although expression of the isotope ratio as delta values is conventional in many areas, in my work the isotope ratio itself is used. Firstly, this is because the aim of my work was not a comparison of samples across space and time, and, secondly, because the use of the simpler fraction allows one to simplify the calculations arising from experiments. It is worth pointing out, however, that for the method developed, both isotope ratio and delta values can be used

interchangeably, and all the statements made, where the isotope ratios are used, are valid for the cases when the delta values are used.

It is worth briefly discussing the way in which LC coupled to isotope ratio MS functions. Schematically this process is presented in Figure 6-1. First, the separation of a complex sample is carried out by conventional LC. In the next stage complete oxidation of the sample takes place. Thus, all the carbon present is converted to CO₂; other elements present in the sample are oxidised accordingly. Ionisation of CO₂ is carried out using an electron ionisation source, thus, producing CO₂⁺ ions. Three masses are recorded for the carbon dioxide via channels 44, 45 and 46 (Figure 6-1). The channels 44 and 45 are used to derive the raw information on the carbon isotope ratio. Trace 44 represents the variation of the signal derived from ¹²C¹⁶O¹⁶O and trace 45 represents the superposition of signals derived from ¹³C¹⁶O¹⁶O and ¹²C¹⁷O¹⁶O or ¹²C¹⁶O¹⁷O. Trace 46 is used to introduce correction for the presence of ¹⁷O and its contribution to the trace 45.

Three carbon isotopes are found in nature ¹²C, ¹³C and ¹⁴C³⁰ (Table 6-1) can be excluded from consideration. The abundance of ¹⁴C carbon isotope is several orders of magnitude lower than reported precision for analysis using the LC-IRMS technique (1 ppt *versus* 0.2 ‰); therefore, this isotope can be excluded from consideration.

Table 6-1. Carbon and oxygen average isotope abundance³⁰

Isotope	Abundance (%)	R
¹² C	98.90	1
¹³ C	1.10	0.0111223
¹⁴ C	≈1×10 ⁻¹⁰	≈1×10 ⁻¹²
¹⁶ O	99.76	1
¹⁷ O	0.038	0.0003809
¹⁸ O	0.20	0.0020048

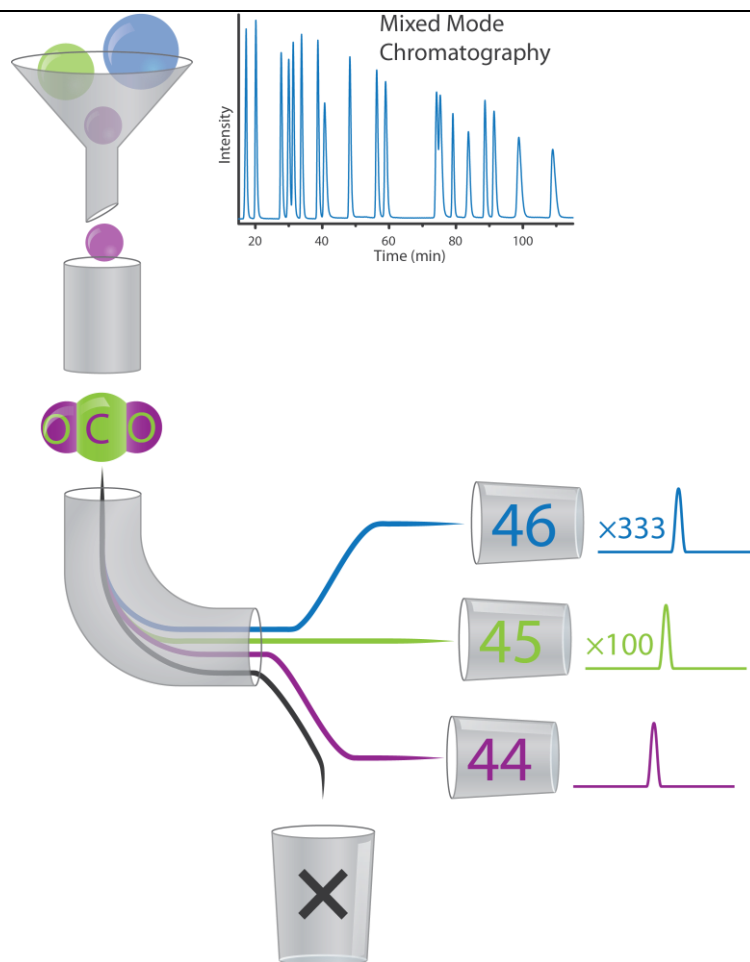


Figure 6-1. Schematic representation of an isotope ratio mass spectrometer coupled to a LC using chemical oxidation interface.

Many chemists are familiar with the isotope experiments carried out using molecular MS; in order to avoid confusion it is useful to compare some aspects of conventional MS and IRMS functionality. A comparison of the two techniques is presented in Figure 6-2. Consider a compound with six carbon atoms fully labelled with ^{13}C (that is all carbon atoms have been replaced with ^{13}C atoms). In the case of molecular MS a mass shift in the molecular ion, proportional to 6 Da (depending on the charge state), would be expected. Since all the carbon is converted to carbon dioxide, for isotope ratio MS, a 1 Da shift for the carbon dioxide (expected for molecular MS) will correspond to the flat line for the trace 44 and a peak for

the trace 45. Analogously, should only half of the carbons be labelled peaks for both trace 44 and trace 45 would be expected to provide the same intensity.

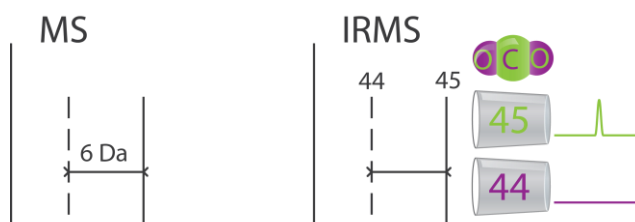


Figure 6-2. Comparison of molecular MS and IRMS for a 100% ^{13}C labelled molecule containing 6 carbon atoms.

In case of molecular MS 6 Da shift is observed for a 100% ^{13}C labelled molecule containing 6 carbon atoms will be observed; for the CO_2 1 Da shift is observed. In case of isotope ratio MS the trace 44 peak will disappear completely; peak of the trace 45 will be proportional to the quantity of the labelled compound.

Despite data for three mass channels being acquired I will present only traces 44 and 45 throughout. This is due to the fact that it is difficult to draw any conclusions from the analysis of trace 46 and its presence complicates the data presented, whereas traces 44 and 45 provide easily understood visual information. Throughout this chapter the chromatograms and illustrations are colour coded: purple used for trace 44 and green for trace 45.

6.1.2 Chromatographic Methods for LC-IRMS

The coupling of isotope ratio MS to LC via the chemical oxidation interface imposes several crucial requirements on the choice of chromatographic method⁵.

Firstly, the choice of the mobile phases is limited to carbon free mobile phases; a limitation which excludes the use of organic solvents and organic buffers. Secondly, it is not desirable to employ derivatisation for the analysis of AAs, as by doing so a new carbon source is added which will affect the isotopic composition of the sample. Chromatographic methods are restricted to separation using aqueous, inorganic elluents.

In the work presented in this thesis, a method modified from that developed by McCullagh and co-workers^{15,16} (Figure 6-3) for the separation of AAs was used. The method employs

mixed mode chromatography using a PrimeSep A (2.1×250mm, 5µm) column and gradient 0 to 0.03 M sulphuric acid in water solution as the mobile phase¹⁷ (Chapter Eight). For this type of detection it is important to achieve baseline separation for the target analytes. In order to ensure better resolution of the late eluting compounds ($t_R = 75$ min, Figure 6-3) shallower gradient was applied, as compared to the original work.

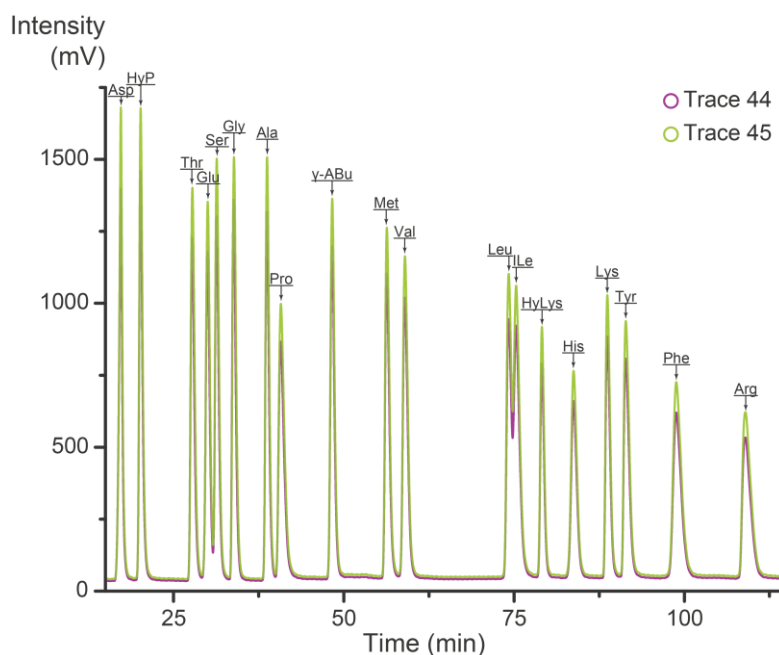


Figure 6-3. General view of the AAs separated on the mixed mode chromatographic column and detected by isotope ratio MS

6.2 Detection and Quantification of PTMs Using LC-IRMS

To the knowledge of the author, the use of LC-IRMS for the analysis of PTMs has not been previously reported. Indeed, IRMS generally sees most of its use in environmental and inorganic applications and has not been used as a tool in chemical biology. In order to apply LC-IRMS to chemical biology problems a novel experimental protocol was developed. The protocol developed employs metabolic isotope labelling followed by compound specific isotope ratio MS for the identification and quantification of PTMs. This method combines the very sensitive detection of isotope ratio changes and the selectivity of LC. The concept

behind this experimental method involves the use of metabolic labelling to incorporate a carbon isotope labelled AA (or selection of labelled AAs) into cell culture. The significance of this procedure is two-fold. It ensures that PTMs result from isotopically labelled AAs making them more easily identified. In addition, the added isotope label enables the quantification of the PTMs using IRMS. As LC-IRMS can be used to measure individual AA isotope ratios extremely accurately, down to changes in natural abundance at below 1‰ level⁴, the detection of modified AAs is straightforward if they contain artificially higher concentrations of ¹³C carbon. Indeed the spiking of analyte samples with natural ¹³C abundance standards is still possible whilst maintaining the ability to identify the presence of the PTM by elevated (but diluted) ¹³C carbon content. Thus, spiking of the standard enables both identification from the LC chromatogram and subsequent quantification using the increase in the ¹³C ratio compared to the standard. This is the general concept that lies behind the design and development of LC-IRMS protocol described below.

The quantification experiment design is based on the concept of isotope dilution[#] already developed for a wide range of compounds including AAs¹⁸. A known quantity of a standard is added directly to the sample, as opposed to single or multiple point calibration (in which case the set of standards is analysed separately). Two signals are acquired from the sample, one for the standard, and one for the sample. The ratio of signal intensities for the sample and for the standard allows quantification of the sample. This is based on the general analytical technique of standard addition.

Unlike the method using an internal standard, the method developed here does not allow measurement of two distinguishable signals from the sample and from the standard. Instead,

[#] The term ‘isotope dilution’ was initially introduced for substances composed of the same type of nuclides, and, according to the IUPAC, refers to the “mixing of given nuclide with one or more of its isotopes”. Development of MS enabled the application of this method to more complicated organic compounds. In this, case, however, the application of the term should be expanded and mixtures of given compounds with their isotopologues should also be included.

it provides two traces for different isotope species. As a result, it is possible to deduce the quantity of isotopically labelled molecules indirectly. This is performed by applying the isotope ratio concept. The increase in concentration of one of the isotopic forms, either labelled or non-labelled, leads to the proportional increase of the intensity of the corresponding mass peak in the spectrum (in the case of molecular MS) or increase of the chromatographic peak on the corresponding mass trace (in the case of isotope ratio MS), and, consequently, alters the isotope ratio. Thus, it is possible to track the alterations in isotope composition and, given knowledge of the isotope composition of the standard and sample, and the quantity of the standard added, it is possible to calculate the initial quantity of the sample (Figure 6-4).

The general formula describing relationship between sample and analyte concentration and isotopic composition of the analytical mixture is given as follows:

$$R_{a.m.} = \frac{C_{St} \cdot A_{St}^{13} + C_S \cdot A_S^{13}}{C_{St} \cdot A_{St}^{12} + C_S \cdot A_S^{12}}$$

Where subscript a.m. refers to the analytical mixture of sample and standard, and R is the isotope ratio of the mixture. C stands for the concentration and A for the abundance. The subscripts St and S refer to the standard and sample respectively. The superscripts 12 and 13 refer to the carbon isotopes.

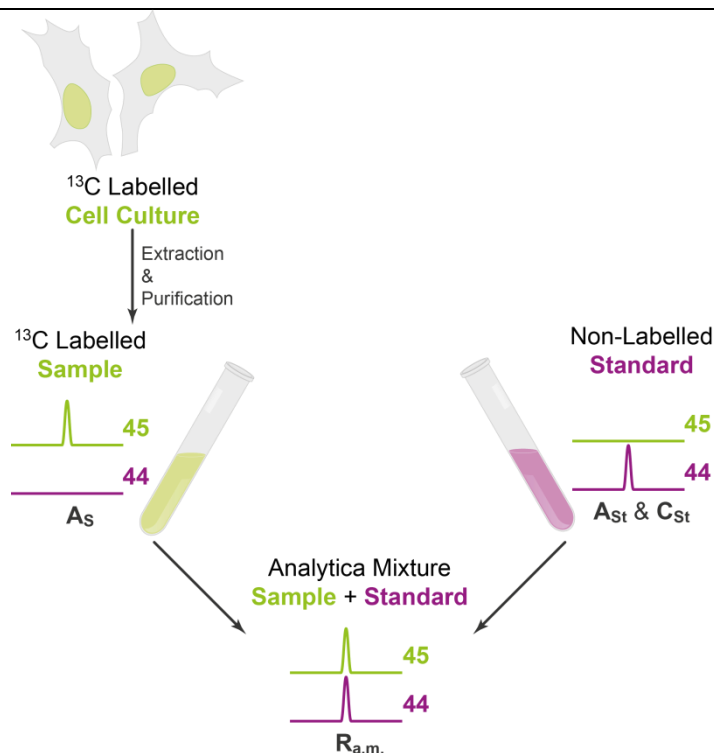


Figure 6-4. Isotope Labelling Experiment Design

$R_{a.m.}$ – isotope ratio of the analytical mixture (measured experimentally); The C_{St} – concentration of the standard (known *a priori*); (A_{St}) – isotopic abundance of the standard (either quoted by the supplier, or can be measured in a separate control experiment); (A_S) – isotopic abundances of the sample (deduced from *a priori* data on the sample, or measured in a separate control experiment).

The isotope ratio of the analytical mixture ($R_{a.m.}$) is measured experimentally. The concentration of the standard (C_{St}) is decided *a priori*. Isotopic abundances of the standard (A_{St}) are either quoted by the supplier, or can be measured in a separate control experiment, which does not contain added sample. Isotopic abundances of the sample (A_S) can either be deduced from *a priori* data on the sample, or can be measured in a separate control experiment, which does not contain added standard.

6.2.1 Defining the Quantity of the Standard to be Added

As pointed out above, it is important to ensure a minimum dilution of the sample when adding the standard modified AAs. On the other hand it is also necessary to ensure accurate and robust detection of the chromatographic peaks, and to minimise the variance of the

isotope ratio calculated for a given peak. The relationship between the peak size, and the isotope ratio associated with this peak becomes obvious when one considers the contribution of the background. In general, the smaller the peak area the greater the relative influence of the background, resulting from dissolved CO₂ (normally at about 10 mV level, Figure 6-3). When dealing with the isotope ratio the relative error becomes absolute, due to the fact that, by its nature, the ratio is a relative value. Previously it had been reported that for the analogous instrument limit of quantification was estimated to be 10 ng of carbon per injection¹⁹. It was necessary to verify whether isotope ratio measurements can be carried out at this level.

Minimum dilution and high accuracy are diametric conditions in their nature, but they can be reconciled through optimisation. In order to optimise the quantity of the added standard, three AAs differing in the number of the carbon atoms in their backbone – alanine ($n_C = 3$), proline ($n_C = 5$), and hydroxylysine ($n_C = 6$) – were selected. To test the influence of a modification chemistry hydroxyproline was selected ($n_C = 5$) as an example of an AA, which has the same number of carbon atoms as proline, but with an additional hydroxyl group. The AAs were injected as a mixture and separated using mixed mode chromatography column PrimeSep A (2.1×250mm, 5µm; Chapter Eight).

The isotope ratio measurement is directly related to quantity of carbon atoms present in the molecule, because the molecule is converted into carbon dioxide prior to MS analysis. The detector response is therefore, proportional to the amount of analyte present and records the amount of carbon contained in a sample. The concentration is presented as carbon content per injection. The quantification experiment was designed for a 10 µL sample volume to be injected on the column. It is therefore, worth considering the amount of added carbon in the units of nanograms per 10 µL.

For every AA the isotope ratio and corresponding peak areas were measured at four different concentrations with five replicates for every concentration point. The isotope ratio variance decreases with the increase in concentration (Figure 6-5).

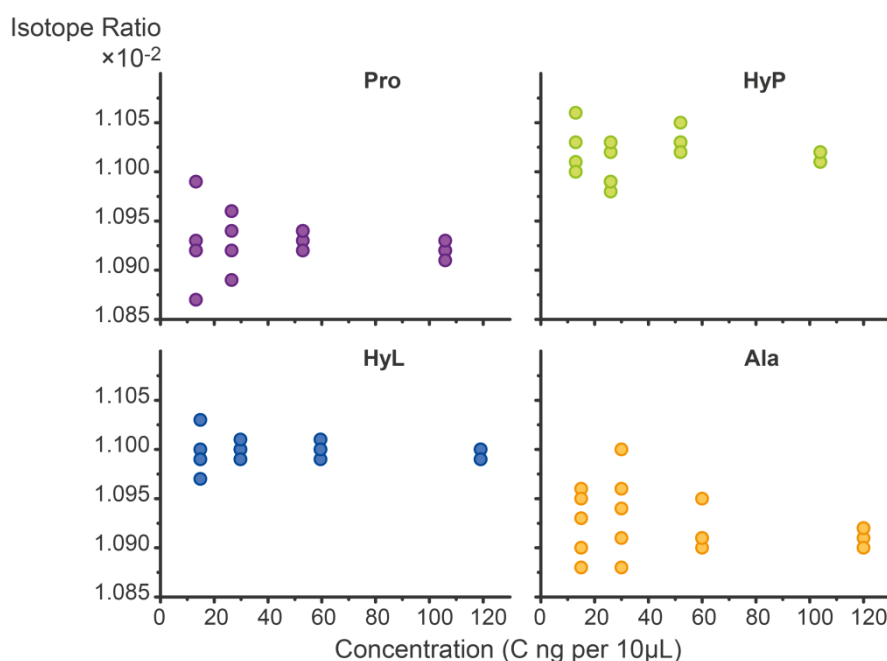


Figure 6-5. Isotope ratio variance dependence on concentration.

Table 6-2. Analysis of variance summary table

AA	Test F Values	
	Isotope Ratio	Area (mV·s)
Ala	0.636	725.22
Pro	0.402	13894
HyL	0.850	9308.0
HyP	2.555	6393.6

In order to identify the cut-off point at which the decrease in the R becomes insignificant with an increase of the concentration, analysis of variance was carried out (Table 6-2). The critical F value (for the F-distribution the value, which corresponds to the given level of confidence, when comparing the variances of two samples of given size) is independent of the actual values and is only dependant on degrees of freedom of the dataset. As equivalent datasets are used, the F critical will be constant in all the tests. This experiment was based on a dataset of

4 samples (4 different concentration levels), and 5 replicates at every concentration level. Therefore, the between-sample degrees of freedom can be defined as $4-1=3$, and within sample degrees of freedom can be defined as $4 \times (5-1)=16$. For the confidence level of 0.05 the critical F value is 3.007.

The test F values for the isotope ratio are well below the critical F value, which means that the between samples variation is insignificant at the 0.05 level and can be attributed to random error. The opposite situation is observed for the area. As anticipated, peak area was strongly dependent on the concentration.

Despite the fact that the analysis of variance proved to be unrevealing, probably due to the lack of the test power, it is worth noting the increase in confidence limits with the decrease of concentration, which is clearly seen in Figure 6-5. Based on the visual analysis of the data, it appears that, confidence limits remain constant above 60 ng of carbon per injection. This suggests that the addition of standards at a level of 60 ng of carbon per injection should be used. This is in agreement with previously published data for the analysis of carbohydrates²⁰. Stable isotope ratios were reported to be constant for the samples containing above 36 to 150 ng of carbon depending on the carbohydrate.

6.2.2 Calibration Methods

6.2.2.1 Theoretical Considerations

In order to understand, in which way sample and standard isotopic composition contribute to the value of the isotope ratio of the analytical mixture, it is worth considering the situation in a more general form, using abstract isotopic masses M_2 and M_1 ($M_2 > M_1$) and using the number of moles (v) instead of the concentrations (C). This general approach is suitable for any isotope dilution technique with any type of isotopes, and therefore, might be interesting for wider audience.

The equation for the isotope ratio of the analytical mixture ($R_{a.m.}$) can be rewritten as follows:

$$R_{a.m.} = \frac{v_{St} \cdot A_{St}^{M_2} + v_S \cdot A_S^{M_2}}{v_{St} \cdot A_{St}^{M_1} + v_S \cdot A_S^{M_1}} \quad (1)$$

where v stands for the number of moles of either standard or sample, A stands for the abundance of the chemically identical molecules, which differ in isotope composition and as a result, molecular weight. Molecular weight is shown as superscript (M_1 and M_2). Thus, the numerator gives the total quantity of the entities with mass M_2 , and the denominator gives the total quantity of the entities with mass M_1 . This is the first form of equation commonly used for the calibration.

Several authors, however, present the same equation in a different manner, and express the mole ratio of sample to standard in the mixture, rather than the isotope ratio in the mixture^{21,22}. Such presentation allows direct calculation of the quantity of the analyte. In this case the equation will have slightly different form:

$$\frac{v_S}{v_{St}} = \frac{R_m \cdot A_{St}^{M_1} - A_{St}^{M_2}}{A_S^{M_2} - R_m \cdot A_S^{M_1}} \quad (2)$$

Equation (2) is a rearranged form of equation (1).

Another approach is based on the idea that only the quantity of standard is known *a priori*, and isotope ratios of sample (R_S) and standard (R_{St}) and analytical mixture of the two ($R_{a.m.}$) can be determined *a posteriori*. Thus, it is necessary to express the mole ratio of sample to standard in the mixture via the isotope ratios of the mixture, standard and sample. For this it is necessary, to introduce two assumptions; firstly, all compounds are assumed to exist in two isotopic forms with weights M_1 and M_2 . The abundance (A) of the molecule of a particular

type corresponds to the probability of the molecule having either weight M_1 or weight M_2 .

This means that:

$$A_i^{M_1} + A_i^{M_2} = 1 \text{ and } R_i = \frac{A_i^{M_2}}{A_i^{M_1}}$$

$$A_i^{M_1} = \frac{1}{R_i + 1} \text{ and } A_i^{M_2} = \frac{R_i}{R_i + 1}$$

And therefore

$$\frac{v_S}{v_{St}} = \frac{(R_m - R_{St}) \cdot (R_S + 1)}{(R_S - R_m) \cdot (R_{St} + 1)} \quad (3)$$

This is the final form of the general equation commonly appearing in the literature.

According to many authors three limiting conditions should be considered²²⁻²⁴. First, if the abundance of molecules with molecular weight M_2 in the standard *only* is negligibly small ($A_{St}^{M_2} \rightarrow 0$) and, consequently, the isotope ratio of the standard tends to zero ($R_{St} \rightarrow 0$), then the equations (1-3) can be simplified:

$$\begin{array}{c|c|c} (1) & (2) & (3) \\ R_m = \frac{v_S \cdot A_S^{M_2}}{v_{St} \cdot A_{St}^{M_1} + v_S \cdot A_S^{M_1}} & \frac{v_S}{v_{St}} = \frac{R_m \cdot A_{St}^{M_1}}{A_S^{M_2} - R_m \cdot A_S^{M_1}} & \frac{v_S}{v_{St}} = \frac{R_m \cdot (R_S + 1)}{R_S - R_m} \end{array}$$

Second, if the abundance of molecules with molecular weight M_1 in the sample *only* is negligibly small ($A_S^{M_1} \rightarrow 0$) and, consequently the isotope ratio of the sample tends to infinity ($R_S \rightarrow \infty$), then the equation can be simplified:

$$\begin{array}{c|c|c} (1) & (2) & (3) \\ R_m = \frac{v_S \cdot A_S^{M_2}}{v_{St} \cdot A_{St}^{M_1}} + \frac{A_{St}^{M_2}}{A_{St}^{M_1}} & \frac{v_S}{v_{St}} = \frac{R_m \cdot A_{St}^{M_1} - A_{St}^{M_2}}{A_S^{M_2}} & \frac{v_S}{v_{St}} = \frac{R_m - R_{St}}{R_{St} + 1} \end{array}$$

Third, if *both* the abundance of molecules with molecular weight $\mathbf{M}_1$ in a sample is negligibly small ($A_S^{M_1} \rightarrow 0$, and, consequently, $R_S \rightarrow \infty$) *and* the abundance of molecules with molecular weight $\mathbf{M}_2$ in a standard is negligibly small ($A_{St}^{M_2} \rightarrow 0$, and, consequently, $R_{St} \rightarrow 0$), then the equation can be simplified as follows:

$$\begin{array}{c|c|c} \begin{array}{c} (1) \\ R_m = \frac{v_S \cdot A_S^{M_2}}{v_{St} \cdot A_{St}^{M_1}} \end{array} & & \begin{array}{c} (2) \\ \frac{v_S}{v_{St}} = \frac{R_m \cdot A_{St}^{M_1}}{A_S^{M_2}} \end{array} \\ \hline & & \begin{array}{c} (3) \\ \frac{v_S}{v_{St}} = R_m \end{array} \end{array}$$

The last equation has frequently been used without necessarily taking regard to the limitations discussed²³. Non-justified use of the fitting equation can result in significant error, when calculating the concentration of the analyte.

Vice versa, if *both* the abundance of molecules with molecular weight $\mathbf{M}_1$ in a standard is negligibly small ($A_{St}^{M_1} \rightarrow 0$, and, consequently, $R_{St} \rightarrow \infty$) *and* abundance of molecules with molecular weight $\mathbf{M}_2$ in a sample is negligibly small ($A_S^{M_2} \rightarrow 0$, and, consequently, $R_S \rightarrow 0$), then the equations can be rewritten in the following way:

$$\begin{array}{c|c|c} \begin{array}{c} (1) \\ R_m = \frac{v_{St} \cdot A_{St}^{M_2}}{v_S \cdot A_S^{M_1}} \end{array} & & \begin{array}{c} (2) \\ \frac{v_S}{v_{St}} = \frac{R_m \cdot A_S^{M_1}}{A_{St}^{M_2}} \end{array} \\ \hline & & \begin{array}{c} (3) \\ \frac{v_S}{v_{St}} = \frac{1}{R_m} \end{array} \end{array}$$

6.2.3 Simulation Calibration Graphs

In order to estimate the applicability of different analytical forms of the basic equation and the effect of these on the quantitative analysis calibration, simulations were carried out.

Calibration simulations for the isotope dilution were performed in the following manner: the quantity of standard used was fixed and the quantity of sample was varied. In practice, there

exist two possibilities – either standard or sample can be isotopically enriched. In this work the situation where the sample is enriched and the standard is relatively depleted will be discussed in detail.

Below four possible situations for the calibration will be described considering all three assumptions listed above as well as the situation, where no assumptions are valid. Every situation is illustrated with a scatter plot, where the calibration is simulated using the basic equation (no assumptions taken), as well as all of the simplified equations valid under the conditions discussed above. To the left of the graph the information on the isotope ratios of the sample and standard are given, and the fulfilled assumptions are highlighted, and the equation suitable for those conditions is provided.

Visual analysis of the calibration simulations allows one to compare the deviation from the basic data plot (no assumptions) caused by the application of the simplified equations. The plots are presented as the calculated values against the expected values of the mole ratio of sample to standard in the analytical mixture. Clearly, on such a scale, the ‘no assumption’ equation corresponds to the straight line at $y = x$.

In the basic case, when no assumptions are valid, it can be clearly seen, that only the basic equation, can be used for the calibration. The use of the different equation results in great deviation the original dataset (Figure 6-6).

$$R_S = 0.7/0.3$$

$$R_{St} = 0.3/0.7$$

$$\frac{v_S}{v_{St}} = \frac{(R_m - R_{St}) \cdot (R_S + 1)}{(R_S - R_m) \cdot (R_{St} + 1)}$$

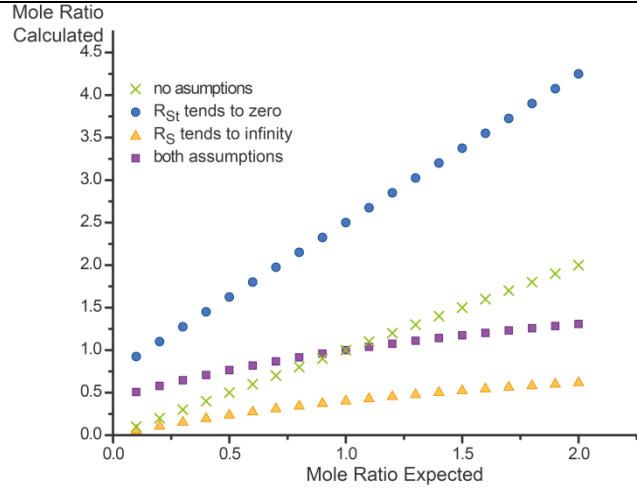


Figure 6-6. Comparison of Different Calibration Equations (conditions are listed to the left of the plot)

When a sample is relatively enriched in a heavy isotope and the standard is almost devoid of such an isotope, the calibration curve can be approximated by the equation of $(\frac{v_S}{v_{St}} = \frac{a \cdot R_m}{R_m + b})$ (Figure 6-7).

$$R_S = 0.7/0.3$$

$$A_{St}^{M_2} \rightarrow 0, R_{St} = 0.01/0.99$$

$$\frac{v_S}{v_{St}} = \frac{R_m \cdot (R_S + 1)}{R_S - R_m}$$

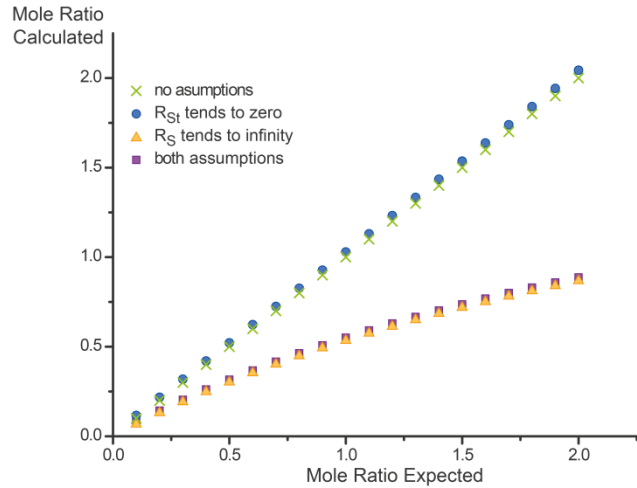


Figure 6-7. Comparison of Different Calibration Equations (conditions are listed to the left of the plot)

The third case occurs when the standard has both isotopes, and the sample is greatly enriched in the heavy isotope. In this case the calibration graph can be reduced to the straight line

$(\frac{v_S}{v_{St}} = R_m \cdot a - b)$ with the slope (a) equal to $\frac{1}{R_{St}+1}$ and an intercept (b) equal to $-\frac{R_{St}}{R_{St}+1}$. For

the approximation either this equation or the basic one can be used. It is crucial to point out that the deviation from the basic data set increases with the increase in the mole ratio, i.e. when the fraction of a sample in the analytical mixture is increasing (Figure 6-8).

$$A_S^{M_1} \rightarrow 0, R_S = 0.99/0.01$$

$$R_{St} = 0.3/0.7$$

$$\frac{v_S}{v_{St}} = R_m \cdot \frac{1}{R_{St} + 1} - \frac{R_{St}}{R_{St} + 1}$$

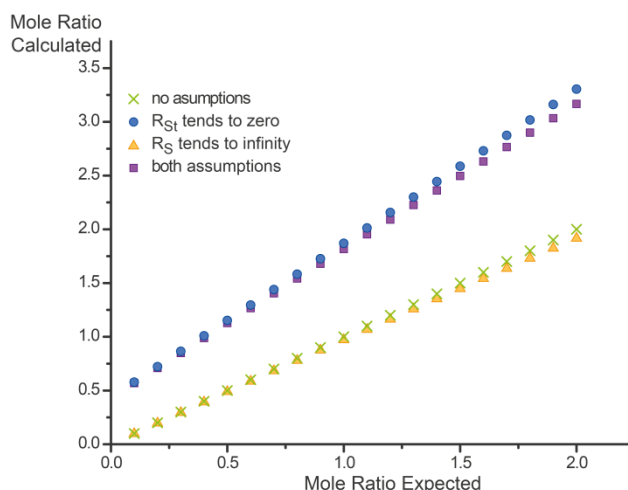


Figure 6-8. Comparison of Different Calibration Equations (conditions are listed to the left of the plot)

In the simplest case – the sample is highly enriched in one isotope, whereas the standard is extremely depleted in such an isotope – the calibration curve is a straight line with an intercept equal to zero and slope approaching unity. It was reported that the deviations of the slope from unity may occur as a result of (1) significant difference between masses used in calculations, (2) if an isotope effect takes place in the process of ionization, or if the standard is composed of differently labelled species²³. In this case either of the models can be used to approximate the data, as clearly seen from the plots (Figure 6-9).

$$A_S^{M_1} \rightarrow 0, R_S = 0.99/0.01$$

$$A_{St}^{M_2} \rightarrow 0, R_{St} = 0.01/0.99$$

$$\frac{v_S}{v_{St}} = R_m$$

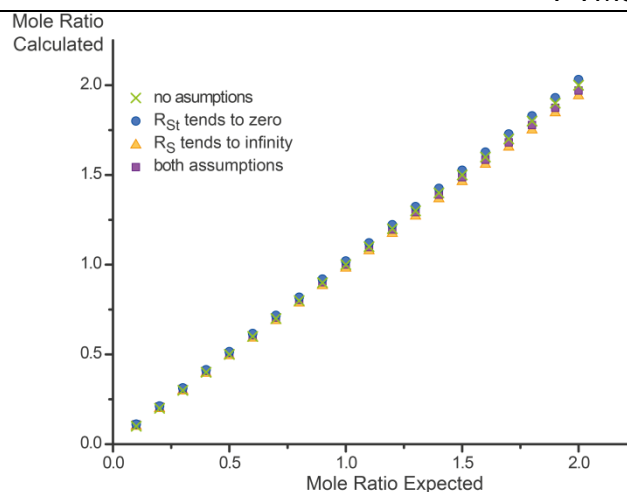


Figure 6-9. Comparison of Different Calibration Equations (conditions are listed to the left of the plot)

In the general case the calibration curve is given by the rational function which under certain conditions can be simplified to a linear one. It is clearly seen, that unless certain requirements of the isotope composition of the sample and standard are fulfilled no other approximations can be applied. The exemption is the narrow range of isotope ratios close to the point of an equimolar mixture of sample and standard, where a linear function can be applied, but this should be checked every time.

6.2.3.1 Calibration Experiment Design

In order to test the quantification protocol suggested a calibration was carried out. Single carbon 99% labelled proline was used as a test compound for the calibration. A standard solution of non-labelled proline at a concentration 60 ng of carbon per 10 μ L (100 μ M), was used to dilute all the labelled standards. The calibration was designed as follows: for the sample concentration range from 10 nM to 10 μ M (three orders of magnitude) with concentration points equally spaced within one order of magnitude: 10 nM, 20 nM, 30 nM..., 100 nM, 200 nM, 300 nM..., 1 μ M, 2 μ M, 3 μ M,..., 10 μ M, thus, giving 28 concentration points with 5 mechanical replicates each, totalling in 140 points per calibration curve.

Detailed discussion of the calibration design is provided in Chapter Four. The calibration curve was constructed as the isotope ratio of the analytical mixture dependency on the concentration of the labelled sample (Figure 6-10 and Figure 6-11).

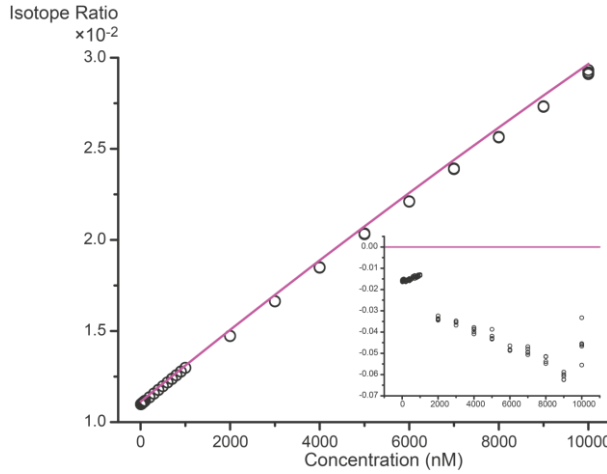


Figure 6-10. Calibration data fit with model equation

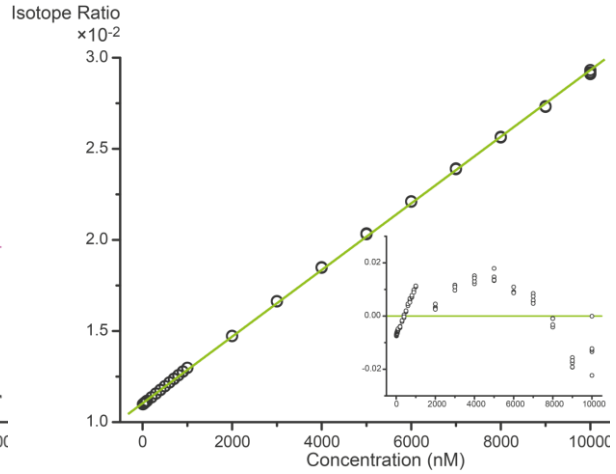


Figure 6-11. Calibration data fit with straight line

The secondary plot is showing the residual scatter plot (axes are as labelled in major plot)

$$R_{a.m.} = \frac{C_{St} \cdot A_{St}^{13} + C_S \cdot A_S^{13}}{C_{St} \cdot A_{St}^{12} + C_S \cdot A_S^{12}}$$

$$\sum (x_i - \hat{x})^2 = 4.9 \cdot 10^{-6}$$

$$R_{a.m.} = a + b \cdot C_S$$

$$a \sim \frac{A_{St}^{13}}{A_{St}^{12}} = 0.011$$

$$b \sim \frac{A_S^{13}}{C_{St}} = 1.8 \cdot 10^{-6}$$

$$R^2 = 0.9998$$

$$\sum (x_i - \hat{x})^2 = 9.5 \cdot 10^{-7}$$

6.2.3.2 Comparison of Linear Fit and Model Fit

In order to eliminate the necessity of recalibration of the system the model fit was suggested. In this section the problem of fitting the experimental data with either linear or model equations is discussed.

The calibration data is presented in Appendix B and graphically shown in Figure 6-10 and Figure 6-11. The low concentration range is presented in Figure 6-16. I have already discussed the conditions under which the calibration can be safely fitted with a straight line. The situations, where the concentration of the sample is close to that of the standard – *i.e.* the

concentration of the sample is 5 to 200% of that of the standard – were considered.

Undoubtedly, for the calibration constructed it is only the higher sample concentration range for which the assumption of closeness of the sample and standard concentrations is fulfilled, and, as a result none of the approximations discussed above can be valid.

However, the following assumptions remain valid for the calibration as well:

1. The abundance of the ^{12}C isotope in the standard tends towards unity ($A_{St}^{12} \rightarrow 1$, $A_{St}^{13} = 0.989$).
2. Throughout the concentration range it can be accepted that the concentration of sample is negligibly small in comparison to the concentration of standard (sample is 0.01 to 10% of the standard).

Using those limiting conditions the general formula

$$R_{a.m.} = \frac{C_{St} \cdot A_{St}^{13} + C_S \cdot A_S^{13}}{C_{St} \cdot A_{St}^{12} + C_S \cdot A_S^{12}}$$

can be simplified to:

$$R_{a.m.} = \frac{C_{St} \cdot A_{St}^{13} + C_S \cdot A_S^{13}}{C_{St} \cdot A_{St}^{12} + C_S \cdot A_S^{12}} = \frac{A_{St}^{13}}{A_{St}^{12}} + \frac{A_S^{13}}{C_{St} \cdot A_{St}^{12}} \cdot C_S$$

Such transformation means that the ^{12}C derived from the sample is neglected. In addition to that, the enrichment of the standard is taken to be 100% ^{12}C . Such conditions cannot be normally fulfilled. As a result we get a linear equation relating the isotope ratio of the analytical mixture to the concentration of the sample

$$R_{a.m.} = a + b \cdot C_S$$

where

$$a = \frac{A_{St}^{13}}{A_{St}^{12}}$$

$$b = \frac{A_S^{13}}{C_{St} \cdot A_{St}^{12}}$$

The results of both model fit and linear fit are given in Figure 6-10 and Figure 6-11.

In order to compare model fit and linear fit the following should be considered:

1. The residuals scatter plot
2. The sum of the residuals squared
3. The practical applicability of the fit.

The residuals scatter plots for both fits are given as the secondary plots in Figure 6-10 and Figure 6-11. It can be seen that neither of the fits is perfect, which can be concluded from the uneven distribution of the residuals relative to the zero level. The model fit overestimates the values of the isotope ratio (residuals scatter plot is below the zero level). This can be caused by the wrong estimation of the isotope ratio of the standard used in the calculation. The sum of the residuals squared is about five times greater in the case of the model fit, however, which implies that the choice should be made in favour of the linear fit.

Apart from linear fit higher order polynomials were suggested²⁵ in the literature to provide accurate description of isotope dilution MS experiments. However, from a practical point of view, a linear or polynomial fit is far from ideal, since it provides the results similar to the model fit but requires performance of a calibration, since the parameters *a* and *b* can be dependent on the performance of the instrument. In contrast, the model fit requires only the *a priori* information on the isotope ratios of sample and standard, and the quantity of the standard added. The suggestion to use the model equation for fitting the isotope dilution calibration data is also supported by the observation that linear bracketing design of the quantitative experiment will produce greater error than a single point calibration using model

fit²⁶. Based on the results of calibration experiment, in this work the model equation was used together with a ‘single point calibration’ in order to carry out the quantification.

6.2.3.3 Limit of Detection

The LoD was defined by the calibration data and was shown to be 10 nM for the single carbon atom labelled proline and will be proportionally lower for the compounds with higher ¹³C abundance. For the fully labelled proline for example one should expect the LoD to be at the level of 2 nM (*i.e.* 5 times lower). This result is in agreement with the minimum detectable difference in the isotope ratio cited in the literature ($\Delta R=0.0002$)⁴. This LoD is approximately two orders of magnitude lower than that for the molecular MS scanning mode for free AAs and about an order of magnitude lower than that for the single ion monitoring mode as reported in Chapter Four.

It is interesting to compare the results obtained for LC-IRMS with the data previously published for molecular MS following separation by LC of o-phthalaldehyde derivatised AAs²⁷. The authors report the minimum quantity to be injected for the stable isotope ratio measurement to be 25 pmol per injection, with the standard deviation of the isotope ratio measurement between 0.2 and 0.8%. Therefore, the LoD for the method reported is between 0.6 and 2.4% of the total amount injected, or transformed into concentration scale 150 to 600 fmol per injection of fully labelled isotopologue of AA. This result, however, is difficult to achieve due to the presence of o-phthalaldehyde fragment, which dilutes the isotopic label.

The extremely low LoD for the LC-IRMS is achieved mainly due to two contributing factors. One can be referred to as a ‘natural filter’. This makes use of measuring only two masses, thus, significantly reducing the background from other ions. It is similar in this way to the single ion monitoring mode for triple quadrupole MS. Another, can be called chemical

filtering achieved through oxidation of the analyte to carbon dioxide thus, filtering out all other components.

The influence of noise on the LoD can be illustrated using a simulation. The linear response to concentration was simulated according to the following equation

$$\text{Response} = 0.7 \cdot \text{Concentration} + 5 + \text{Random Error}$$

Two datasets were created with two different random errors (Figure 6-12). The green dots show the data with smaller variance ($\sigma=0.552$) and the purple crosses show the data with greater variance ($\sigma=1.494$). Based on this data concentration normal distribution (Figure 6-13) and the calibration curve (Figure 6-14) were constructed. It can be seen that the LoD for the dataset with lower noise is in the lower concentration range (green circles), whereas for the dataset with higher noise has a higher LoD (purple crosses). It is worth pointing out that the LoD in this case provides not only the information about the smallest amount which can be measured, but also about the smallest difference between concentrations which can be determined. This is illustrated in Figure 6-14 by the density of concentration points which can be distinguished from each other.

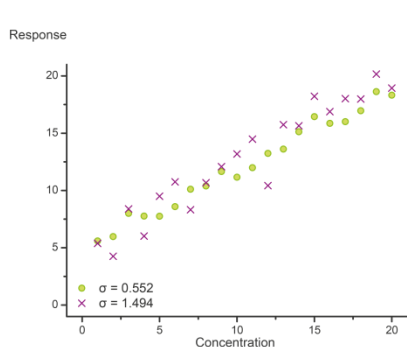


Figure 6-12. Simulated calibration dataset

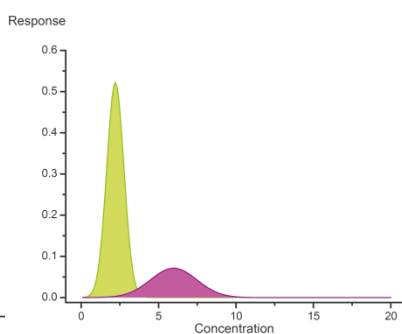


Figure 6-13. Normal distribution reconstructed for the simulated calibration dataset

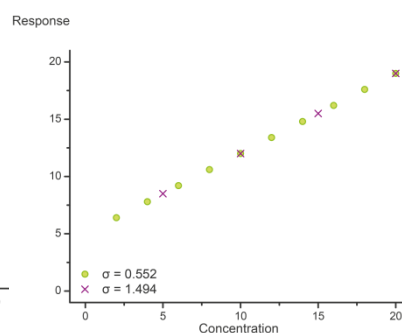


Figure 6-14. Linear fit of the simulated calibration dataset

The second ‘filtering’ involves the normalization by the reference gas, termed as ‘house gas correction’. This is a mathematical procedure, which apparently accounts for the variation in

the instrument performance from run to run. This type of correction was not found in the literature (often only isobaric interference corrections are considered⁶), but deduced from a spreadsheet provided by Dr. Oliver Kracht in a personal communication⁷. House gas correction allows us to reduce the variance in the isotope ratio value, thus, reducing the LoD. The result of such correction is shown in Figure 6-15 and Figure 6-16. Without this correction the LoD would be 4 times higher.

6.3 The Effect of the Isotope Ratio Corrections on the Quantification Method

This section provides the analysis of the influence of the corrections applied in the mathematical algorithm for converting the raw data into an isotope ratio.

Before a more detailed discussion of the mathematics behind the values reported here, it is worth mentioning what the raw data looks like. As discussed before, the raw data is comprised of three mass traces: trace 44, trace 45, and trace 46. The complete list of species contributing to the trace is given in Table 6-3. It can be seen from the table that carbon dioxide comprises six different isotopologues. The isotopologues with mass 47-49 are not considered, due to their extremely low abundance, which means they cannot be accurately detected.

Table 6-3. List of isotopologues of CO₂ and their relative abundance

Information on CO₂ isotopologue abundances is adapted from Stable Isotope Geochemistry (Jochen Hoefs, Springer, 2010)

44			$^{12}\text{C}^{16}\text{O}^{16}\text{O}$ 98.40%		
45		$^{13}\text{C}^{16}\text{O}^{16}\text{O}$ 1.11%	$^{12}\text{C}^{17}\text{O}^{16}\text{O}$ 374 ppm	$^{12}\text{C}^{16}\text{O}^{17}\text{O}$ 374 ppm	
46	$^{12}\text{C}^{17}\text{O}^{17}\text{O}$ 0.142 ppm	$^{12}\text{C}^{18}\text{O}^{16}\text{O}$ 0.20%	$^{12}\text{C}^{16}\text{O}^{18}\text{O}$ 0.20%	$^{13}\text{C}^{17}\text{O}^{16}\text{O}$ 4.2 ppm	$^{13}\text{C}^{16}\text{O}^{17}\text{O}$ 4.2 ppm
47	$^{13}\text{C}^{17}\text{O}^{17}\text{O}$ 1.6 ppb	$^{13}\text{C}^{16}\text{O}^{18}\text{O}$ 22.2 ppm	$^{13}\text{C}^{18}\text{O}^{16}\text{O}$ 22.2 ppm	$^{12}\text{C}^{17}\text{O}^{18}\text{O}$ 0.75 ppm	$^{12}\text{C}^{18}\text{O}^{17}\text{O}$ 0.75 ppm
48		$^{12}\text{C}^{18}\text{O}^{18}\text{O}$ 3.96 ppm	$^{13}\text{C}^{17}\text{O}^{18}\text{O}$ 8.4 ppb	$^{13}\text{C}^{18}\text{O}^{17}\text{O}$ 8.4 ppb	
49			$^{13}\text{C}^{18}\text{O}^{18}\text{O}$ 44.5 ppb		

The structure of raw data is schematically presented in Figure 6-1, and the sample chromatogram is given Figure 6-3. It can be seen, that each trace is a chromatogram, but to trace 44 and trace 45 intensities the factor of 100 and 333 correspondingly are applied. Apparently this has been introduced by Thermo Scientific for the sake of simplicity, so that all three traces can be viewed on the same intensity scale. This information was also deduced from the spreadsheet received in the personal communication⁷.

There are several correction algorithms (Appendix A) applied in calculating R. House gas correction was shown to have a significant effect on the applicability of the quantification method developed here. This correction is applied in order to compensate for the variations in the isotope ratio introduced by the ambient environment. As a reference point the VPDB (Vienna PDB) is used⁶.

In order to investigate the influence of the house gas correction calibration data for single carbon atom labelled proline for the range 10 to 100 nM was used. Raw data was extracted and the isotope ratio values were calculated both with and without applying the house gas correction. The calculations of the isotope ratio were carried out using the algorithm provided⁷, the results of the calculations correspond to the values calculated using Isodat Workspace IRMS data processing software.

First, consideration should be given to the isotope ratio value without any corrections. In order to do this, the area of the chromatographic peak for trace 45 has to be divided by the area of the chromatographic peak for the trace 44 from the isotope ratio system. The resulting value is the raw isotope ratio. These calculations have been carried out for the calibration dataset and the results are given in Figure 6-15.

In order to perform the house gas correction, three pulses of carbon dioxide are recorded at the beginning of every chromatogram. The isotope ratio of the house gas is calculated as

described above. This ratio should be equal to the isotope ratio of VPDB ($R_{VPDB}=0.01118028$). Due to the fluctuations caused by the instrument this is rarely the case.

The factor by which the two values differ is defined and applied to the raw isotope ratio.

Consider the following example. The sample peak areas for the trace 44 (A_S^{44}) and trace 45 (A_S^{45}) are equal to 7.17666559127536 mV·sec and 0.083615650073295 mV·sec respectively.

The reference gas peak areas for the trace 44 and trace 45 are equal to 75.0416284806179 mV·sec and 0.860146789031345 mV·sec respectively. Therefore, the sample raw isotope ratio (R_S^{raw}) can be calculated as follows:

$$R_S^{raw} = \frac{A_S^{45}}{A_S^{44}} = \frac{0.083615650073295}{7.17666559127536} = 0.0116510444871426$$

Analogously the raw reference isotope ratio (R_R^{raw}) can be calculated:

$$R_R^{raw} = \frac{A_S^{45}}{A_S^{44}} = \frac{0.860146789031345}{75.0416284806179} = 0.0114622617665275$$

The correction factor (f) can be calculated as follows:

$$f = \frac{R_{VPDB}}{R_R^{raw}} = \frac{0.01118028}{0.0114622617665275} = 0.975399116485813$$

In order to perform the house gas correction the sample raw isotope ratio has to be multiplied by the correction factor:

$$\begin{aligned} R_S^{corrected} &= R_S^{raw} \cdot f = 0.0116510444871426 \cdot 0.975399116485813 \\ &= 0.0113644184988958 \end{aligned}$$

The resulting values of the corrected isotope ratio for the calibration data are given in Figure 6-16. It can be seen that as a result of house gas correction the variance of the isotope ratio is reduced. The reduction of variance as a result of the house gas correction has not been

previously reported. This is an important mathematical artefact which allows one to lower the LoD 4 times.

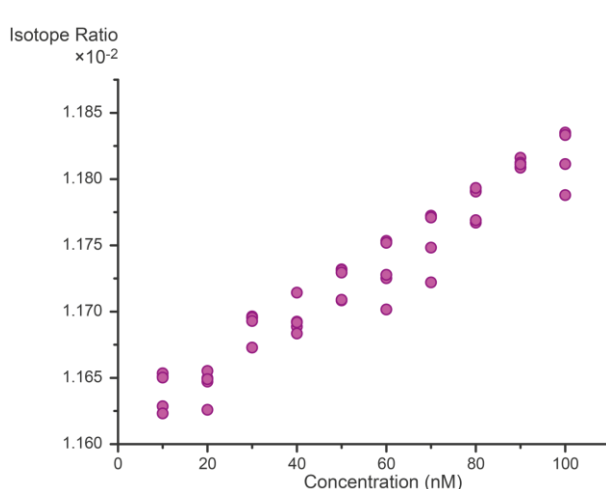


Figure 6-15. Isotope Ratio Calculated without Corrections

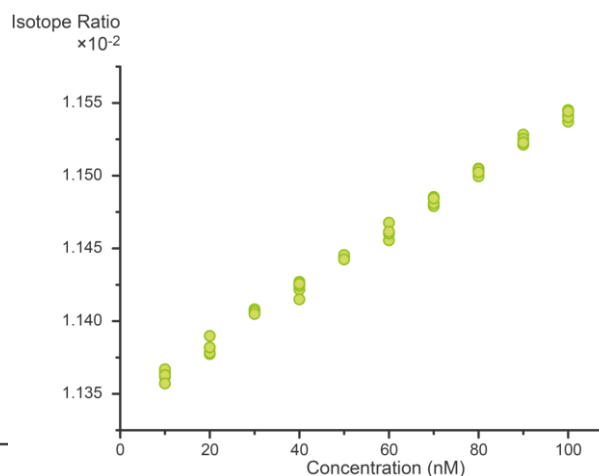


Figure 6-16. Isotope Ratio Calculated Using House Gas Correction

Another class of correction, termed ^{17}O corrections, suggested by Craig^{6,14}, by the group of authors Santrock, Studley, and Hayes²⁸ (SSH), and by Allison *et al.*^{29,30} is often automatically applied to calculate isotope ratios (discussed in detail in Appendix A). These three corrections are normally applied after the house gas correction has been performed. These corrections attempt to compensate for the contribution of ^{17}O to trace 45 by taking into account trace 46.

In order to investigate whether those corrections have an effect on the quantification results the calibration data acquired was subjected to those corrections (Figure 6-17).

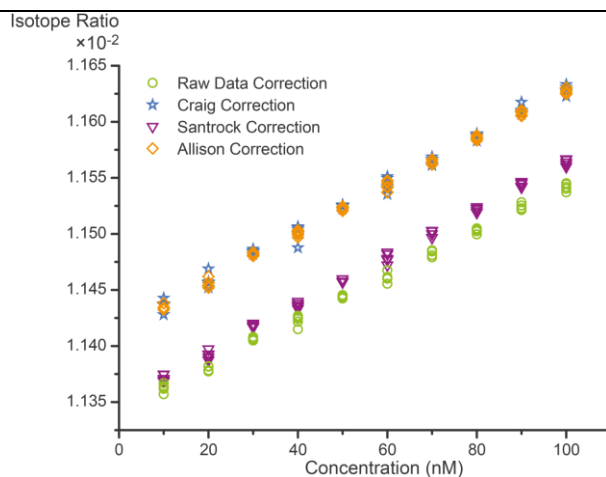


Figure 6-17. Isotope Ratio Correction Comparison

The results obtained using the Allison correction are almost identical to those from the Craig correction and only slightly different from those of the SSH correction and the data to which only house gas correction has been applied. However small the difference between these corrections, it is greater than the variance of the isotope ratio for the same type of correction, which means that one should always be careful when comparing the data against other data treated with different correction algorithm. In this work, house gas correction together with the SSH correction was applied.

It was possible to demonstrate, that unlike the house gas correction, the Craig, SSH or Allison corrections do not affect the isotope ratio variance of the sample. Therefore, quantification results are independent on the ¹⁷O correction used.

6.4 Application of the Method to the Analysis of Histone PTMs

The calibration data suggests that the method of quantification using IRMS (LoD 2 nM for the fully labelled Pro) is 2 to 3 orders of magnitude more sensitive than other LC-MS methods tested previously (Chapter Four). In order to test the applicability of the

quantification method developed to biologically relevant samples, core histones were targeted in human cell cultures grown with selected isotopically enriched AAs.

Histones were extracted from the labelled human embryonic kidney (HEK 293T) cells. Three biological repeats were carried out (Samples I, II and III; Figure 6-18), which came from 2 cell culture batches. The first batch of labelled HEK 293T cells was kindly provided by Dr Mukram Mackeen (Nuffield Department of Medicine, University of Oxford). It was fully labelled with both lysine ($^{13}\text{C}_6$, $^{15}\text{N}_2$) and arginine ($^{13}\text{C}_6$, $^{15}\text{N}_4$). The second batch was grown in the CRL, Oxford. For the second batch the cells were individually labelled with either arginine ($^{13}\text{C}_6$, $^{15}\text{N}_4$) or lysine ($^{13}\text{C}_6$, $^{15}\text{N}_2$).

A variety of protocols for cell labelling with AAs have been previously reported³¹. In this work Dulbecco's Modified Eagle Medium (DMEM) media without Lys or Arg was used. Labelled AAs were added separately: Lys (final concentration 0.0878 g/L, 0.398 mM), Arg (final concentration 0.1521 g/L, 0.798 mM). The cells were re-passaged 5 times with 1:4 split ratios (one quarter of the cells was transferred to the new media each time) before harvesting, thus, theoretically ensuring 99.9% labelling. Histones derived from cell culture without labelled AAs were used for comparison. In order to isolate histones from the cells acid extraction was applied (Chapter Eight).

Three separate histone samples were prepared. The first sample derived from the first batch of histones was analysed as a mixture of bulk histone proteins. The second sample was derived from the 6th passage of the second batch of cells individually labelled with either Lys or Arg; the histones were analysed as a mixture of bulk histones as well. The third sample was derived from the 7th passage of the second batch. The bulk histones extracted for the third sample were separated into individual histones (H2A, H2B, H3 and H4) using SDS

PAGE gel electrophoresis followed by in-gel trypsin digestion. H2A, H2B and H3 individual histones were subject to the analysis (Figure 6-18).

Two aliquots for each histone sample were digested into individual AAs using vapour phase acid hydrolysis. An excess of 6 M HCl was added to the vial containing several sample tubes, which was then incubated for 24 h at 110°C under vacuum (~10 mbar). Dry hydrolysate from the first aliquot was reconstituted in the mixture of quantification standards (HyL, Lys, MeLys, diMeLys, triMeLys, HyArg, Arg and MeArg) and was used for quantification. The dried hydrolysate from the second aliquot was reconstituted in 0.1 M HCl, and used to define the level of isotopic enrichment of the standards.

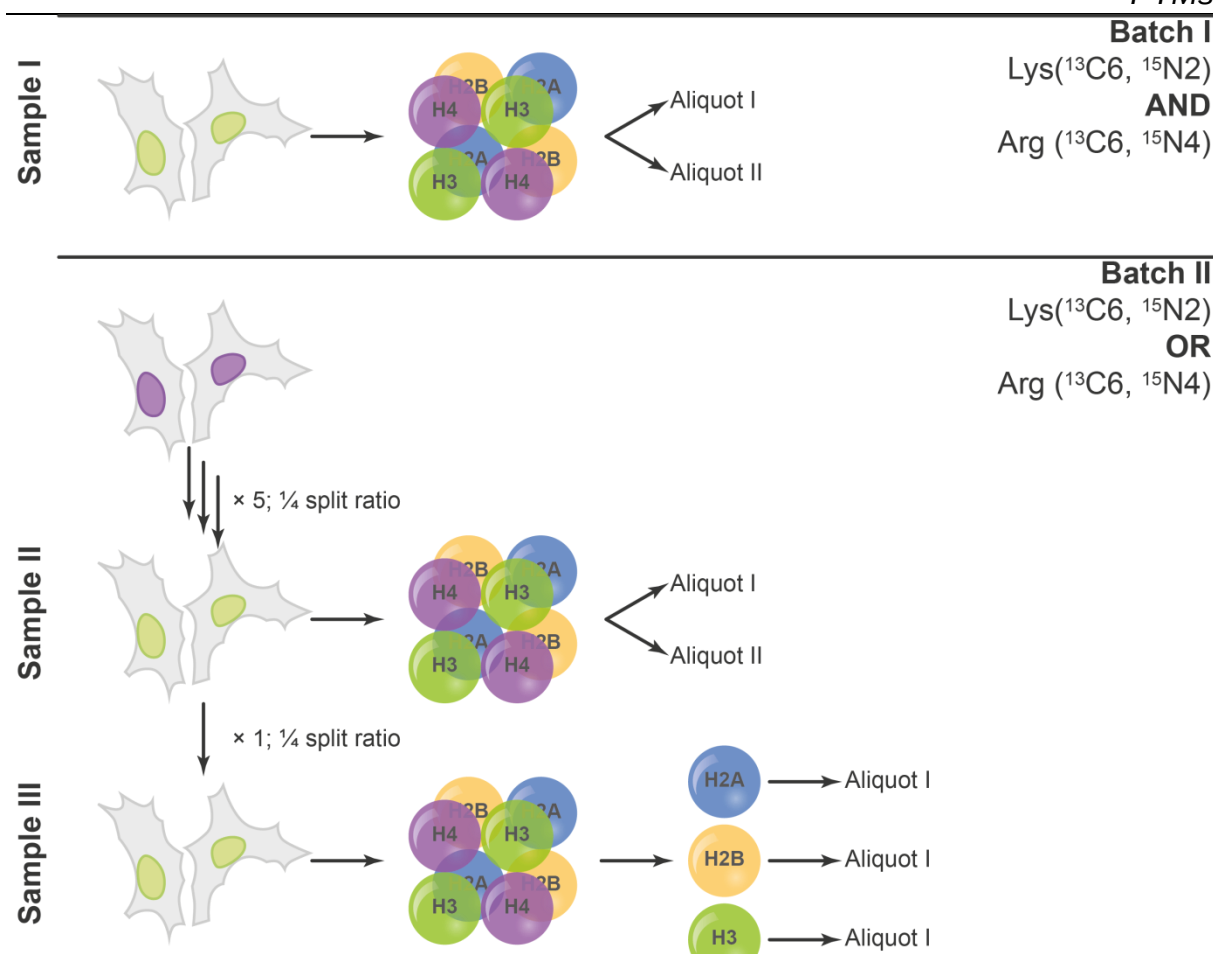


Figure 6-18. Histone analysis scheme

Aliquot I is reconstituted in the solution of quantification standards (60 ng of C per 10 μ L for each standard AA); used for quantification

Aliquot II is reconstituted in 0.1 M HCl; used for the identification of the isotope abundance in the sample

Cells are colour coded as follows: **green** for the labelled cells, **purple** for the non-labelled cells.

Histones are colour coded according to their type.

Under the conditions of acid hydrolysis phosphorylated and acetylated species are hydrolysed, due to the hydrolysis of the phosphate ester and amide bonds correspondingly. Protocols for cell culture labelling, histone extraction and acid digestion are described in Chapter Eight.

The quantity of histones extracted from each 15 cm plate was assessed using the bicinchoninic acid assay³² (see Chapter Eight for details). The total protein quantity for each 15 cm plate bulk histone extract was around 200×10^{-6} g. This amount of protein corresponds

to approximately 1.8×10^{-6} moles of total AAs; given that the average proteinogenic AA average molecular weight is 110 g/mole.

Of those at least one modified AA per protein is expected (*e.g.* either methylated Lys or Arg residues). The quantity of one particular modified AA is predicted to be roughly 1%³³ of the total AAs, resulting in 1×10^{-8} moles of particular AA per 200×10^{-6} g of histone mixture. According to the calibration carried out, the LoD for quantification using IRMS is around 10×10^{-15} mol/ μ L (10nM) using a 10 μ L injection for a single carbon labelled proline. Therefore, as much as 20×10^{-6} g is sufficient for a single analysis, which amount corresponds to 5×10^{-12} moles per 10 μ L of injection.

6.4.1 Histone PTMs Identification

In order to identify histone PTMs of Lys and Arg, bulk histones derived from the first batch were analysed using mixed mode chromatography coupled to IRMS (Figure 6-20). The following AA standards were used: HyLys, Lys, MeLys, diMeLys, triMeLys, HyArg, Arg and MeArg. Peak identification was based on ^{13}C abundance and retention times as compared to those of the standards (Table 6-4). A ^{13}C abundance higher than that of the standard shows that the species was derived from either lysine or arginine and the retention time was used to identify a particular modification. The following modifications of lysine and arginine were identified by spiking samples with the non-labelled standards (Figure 6-20 A): HyLys, MeLys, diMeLys, triMeLys, HyArg, and MeArg.

Figure 6-20 C shows the results of the analysis of the bulk histones derived from the non-labelled cell culture. Without artificial enrichment only abundant AAs such as lysine and arginine can be clearly seen. Figure 6-20 A and Figure 6-20 B show the results of the analysis of the bulk histones derived from the cell culture labelled with both Lys and Arg. A at peak $t_R = 147$ min was assigned the structure of diMeArg based on retention time and the expected

relatively high abundance of this particular modification³³. It was not possible to verify this finding and establish the particular positional isomer of diMeArg, because the standards were not available.

Three peaks $132 < t_R < 142$ min have a high abundance of ^{13}C and, therefore, are expected to be chemically related to lysine or arginine, but could not be identified using any of the reference standards available. Further investigation into the origin of these peaks is required.

A peak ($t_R = 92$ min), which was initially assigned the structure of HyArg was found in both Arg labelled histones ($R = 0.01662581$) and Lys labelled histones ($R = 0.03854693$) (Figure 6-22 A and B). When compared to the isotope ratio of the standard ($R = 0.01098025$) it appears that the concentration (0.5 nM) of the species in the Arg labelled sample is below the LoD (1.7 nM for the fully labelled HyArg). Whereas this species is much more abundant (2.3 nM) in lysine labelled histones. In order to investigate the source of the unknown peak labelled Lys and Arg used in the growth media was analysed separately using IRMS. The peaks ($t_R = 89.2$ min, 99.9 min and 134.5 min) were shown to be present in the labelled Arg standard. Trace 45 was used for the relative quantification of the abundance relative to Arg and was shown to be as follows: 0.15, 0.05 and 0.25% correspondingly. It is suggested that it was introduced into the cell sample as a result of co-purification during the histone extraction.

A peak ($t_R = 99$ min), which was initially assigned as potentially representative structure of HyLys (based on t_R of standard) was also found in both Arg labelled histones ($R = 0.01130928$) and Lys labelled histones ($R = 0.01148469$) (Figure 6-22 A and B). The isotope ratio suggests that the peak is not derived from the labelled source, and therefore, is not related to lysine or arginine.

In vivo the conversion of Arg to Pro takes place efficiently^{34,35}. Arg is first converted into ornithine by arginase in mitochondria. Subsequently, ornithine is converted into glutamic- γ -

semialdehyde by ornithine amino transferase. Glutamic- γ -semialdehyde cyclises to give 1-pyrroline-5-carboxylic acid, which is reduced to proline by pyrroline-5-carboxylate reductase. Conversion of arginine to proline was observed for the first batch of samples (Figure 6-20, $t_{R, HyPro}$ = 11 min, $t_{R, Pro}$ = 18 min). The amount of proline derived from the histones was assessed to be at the level of 13% of the total proline (IR = 0.14705938).

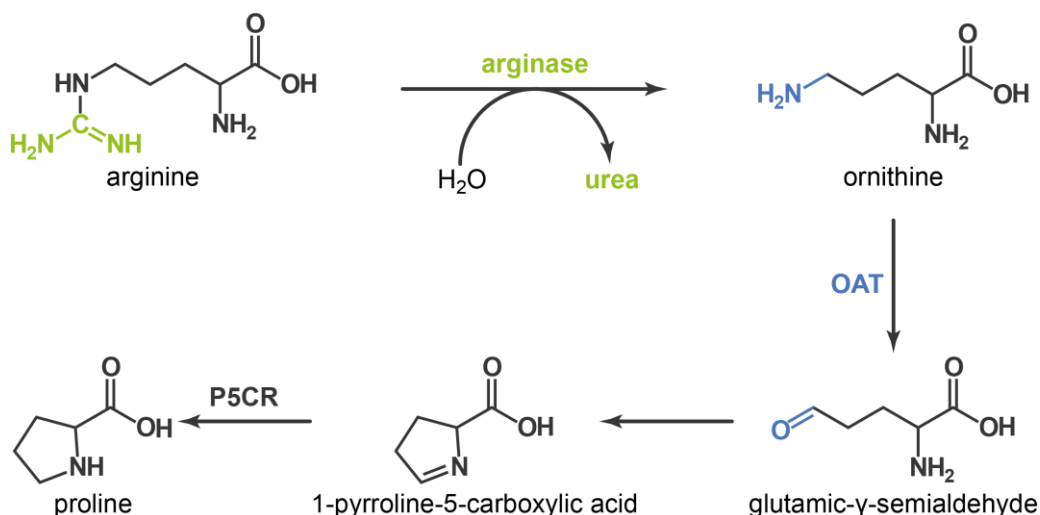


Figure 6-19. Arginine to proline conversion scheme

OAT – ornithine amino transferase; P5CR – pyrroline-5-carboxylate reductase.

In order to reduce the Arg-Pro conversion, non-labelled proline was added (0.2 g/L, 1.665 mM) to the second batch cell growth media containing labelled arginine. It has been proposed that if added proline is present in sufficient quantity in the growth media the production of proline from arginine will be minimised^{34,35}. Experiments showed this was the case; no labelled proline was found in the sample with non-labelled proline added to the cell-growth media (Figure 6-22). In order to make the cells labelled with individual AAs comparable non-labelled proline was added to the lysine labelled growth media as well.

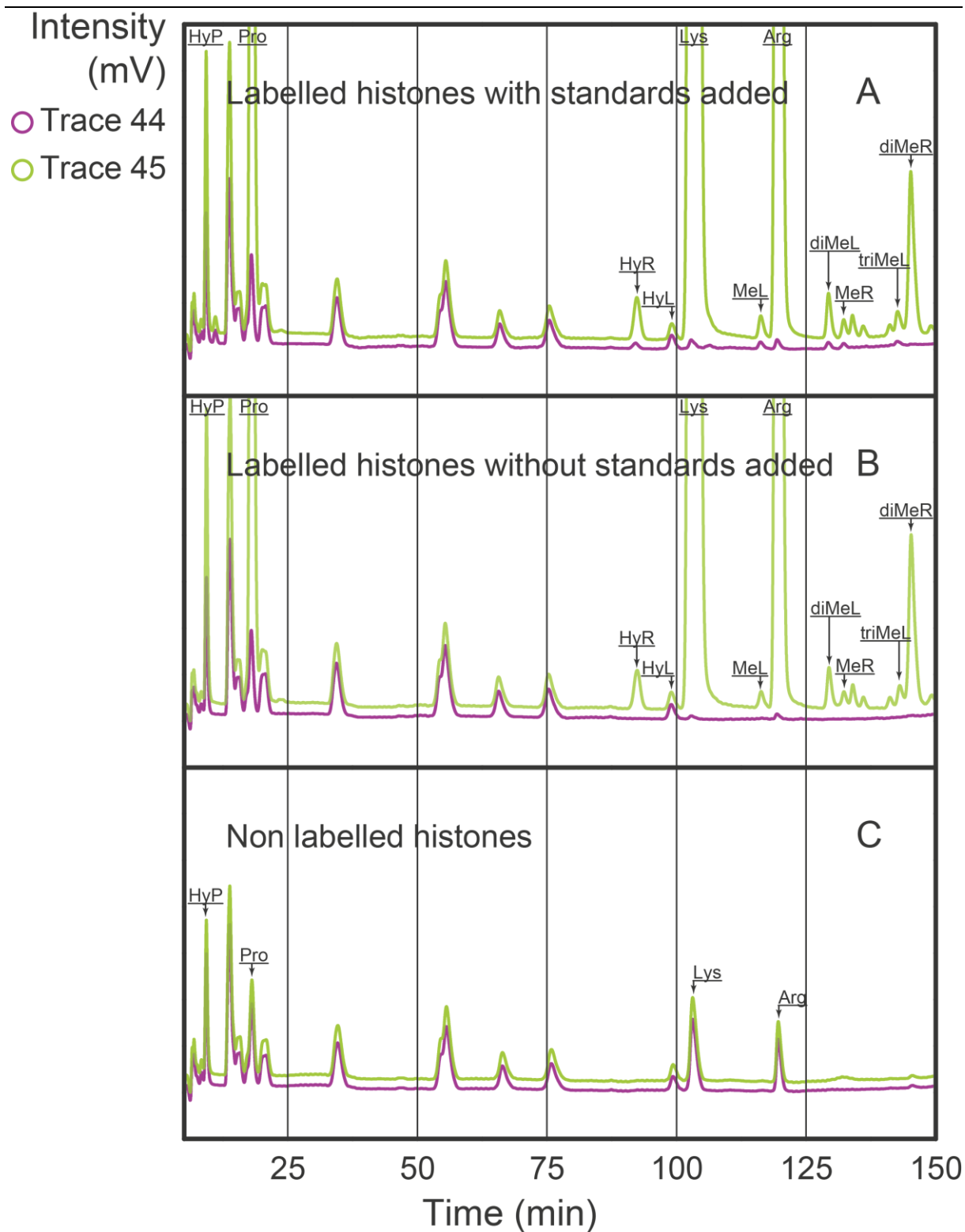


Figure 6-20. Bulk core histones labelled with Arg and Lys (Sample I).

6.4.2 Histone PTMs Quantification

Quantification was carried out as discussed previously in this chapter. HyLys and HyArg were ruled out at the identification stage and were not quantified. Quantification was carried out for the following AAs Lys, MeLys, diMeLys, triMeLys, Arg and MeArg. In order to carry out quantification for each sample, two aliquots of the same sample were analysed with and without addition of the quantification standard (Figure 6-21 A and B correspondingly). The standards used for the quantification were run as a mixture in triplicate (Figure 6-21 C). The retention time (t_R) and isotope ratios (R) for the standards are presented in Table 6-4. Abundance of ^{13}C isotope (^{13}C A) for the standards was calculated on the basis of the isotope ratio and is also presented in Table 6-4.

Table 6-4. Analysis of the standards

t_R = retention time; R = isotope ratio; ^{13}C A = ^{13}C abundance

Δ indicated the 95% confidence interval based on the analysis of 3 replicates.

AA	t_R (min)	$\pm$	Δt_R	R	$\pm$	ΔR	^{13}C A (%)	$\pm$	$\Delta^{13}\text{C}$ A (%)
HyPro	11.08	$\pm$	0.13	0.01098826	$\pm$	0.00002639	1.086883	$\pm$	0.002639
Pro	18.18	$\pm$	0.41	0.01111535	$\pm$	0.00006231	1.099316	$\pm$	0.006231
HyArg	92.10	$\pm$	1.32	0.01098025	$\pm$	0.00010086	1.086099	$\pm$	0.010085
HyLys	99.24	$\pm$	1.09	0.01110563	$\pm$	0.00005144	1.098365	$\pm$	0.005144
Lys	103.20	$\pm$	1.71	0.01098317	$\pm$	0.00004539	1.086385	$\pm$	0.004539
MeLys	116.05	$\pm$	1.49	0.01102493	$\pm$	0.00008011	1.090471	$\pm$	0.008010
Arg	119.49	$\pm$	1.35	0.01103560	$\pm$	0.00003242	1.091514	$\pm$	0.003242
diMeLys	129.19	$\pm$	1.03	0.01082500	$\pm$	0.00007665	1.070907	$\pm$	0.007664
MeArg	132.13	$\pm$	1.09	0.01098531	$\pm$	0.00011789	1.086594	$\pm$	0.011788
triMeLys	142.68	$\pm$	0.75	0.01076065	$\pm$	0.00012858	1.064609	$\pm$	0.012856

Three samples for histone cell culture experiments were analysed (Figure 6-18):

1. Both Arg and Lys labelled bulk histones (provided by Dr. M. Mackeen) (Table 6-6)
2. Individually labelled with either Lys or Arg bulk histones (Table 6-7)
3. Individually labelled with either Lys or Arg gel-separated histones (Table 6-8)

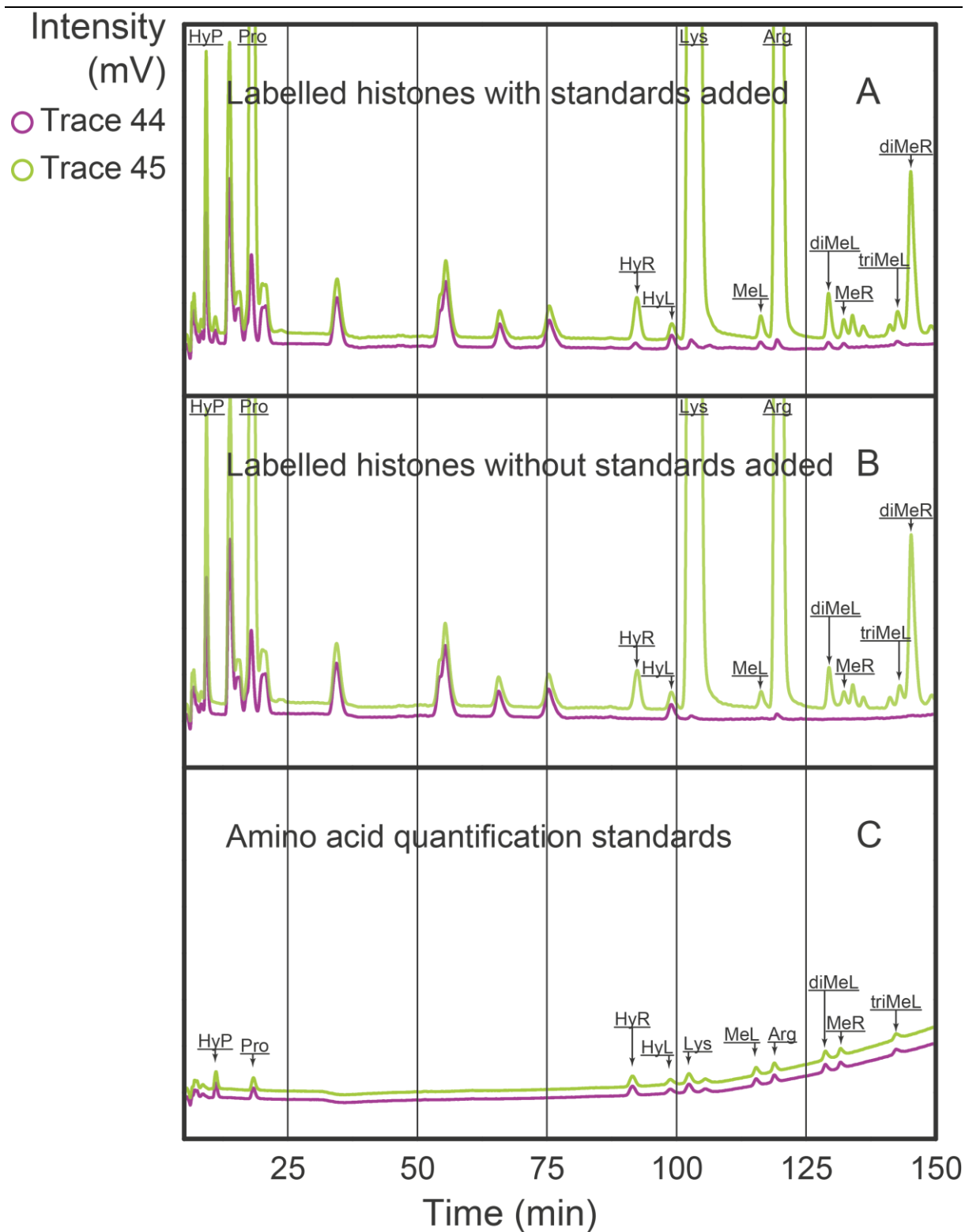


Figure 6-21. Bulk core histones labelled with Arg and Lys (Sample I).

Illustration of the quantification experiment.

To calculate the concentration of each AA present in the sample, first the sample without added standard was analysed (Figure 6-21, B). The bulk histones from batch one with both Lys and Arg labelled were analysed and the isotope ratios and ^{13}C abundances were identified (Table 6-5). The experimental ^{13}C abundance (A_e) is based on the isotope ratio value. The theoretical abundance (A_t) was calculated using the assumption that the methyl group introduced has a natural abundance of ^{13}C (~1%). It can be seen (Table 6-5) that the experimental value is close to the theoretical value in the case of lysine and arginine. Greater deviation is observed for the modified species. ^{13}C abundance, lower than theoretical, could potentially result from co-elution with a matrix component diluting the label. Such dilution will lead to an underestimation of the real concentration (Table 6-6).

Table 6-5. Analysis of the bulk histones (Sample I)

AA	tR (min)	IR	Ae. (%)	At. (%)
HyPro	N/A	N/A	N/A	100
Pro	17.95	0.14705938	12.8	100
HyArg	91.04	5.44737952	84.5	100
HyLys	98.96	0.01122925	1.1	100
Lys	102.85	56.02019360	98.2	100
MeLys	116.46	0.23201638	18.8	86
Arg	119.39	9.93147912	90.9	100
diMeLys	130.21	0.40084966	28.6	75
MeArg	132.13	0.75654862	43.1	86
triMeLys	141.56	0.11954007	10.7	67

The concentrations of the sample can be either calculated with the assumption that the ^{13}C is close to 100% and any deviation from this can be neglected or without this assumption (Table 6-6). In order to estimate the influence of the 100% labelling assumption, the concentrations were calculated using different formulas (1) and (2) below. If no assumption is made, then the data about the isotope abundance of standard and sample have to be used. The concentration in this case will be calculated according to the formula:

$$C_{n.a.} = \frac{C_{St} \cdot A_{St}^{13} - R_{a.m.} \cdot C_{St} \cdot A_{St}^{12}}{R_{a.m.} \cdot A_S^{12} - A_S^{13}} \quad (1)$$

On the other hand, if it is possible to neglect the fact that the enrichment is different from theoretical, the following simplified equation can be used:

$$C_a = R_{a.m.} \cdot C_{St} \cdot A_{St}^{12} - C_{St} \cdot A_{St}^{13} \quad (2)$$

It can be seen that greater deviation from the theoretical enrichment results in greater underestimation of the analyte concentration. It can also be seen that the use of theoretical enrichment can result in mis-finding of an analyte as in the case of hydroxylysine.

Table 6-6. Bulk histone analysis results (Sample I)

N/I=not identified (R is within the confidence interval of a standard); T/A=trace amount (R is below LoD)

AA	tR (min)	R	Cn.a. (nM)	Ca. (nM)
HyArg	92.06	0.08617115	8	6
HyLys	99.06	0.01123204	N/I	T/A
Lys	102.81	10.97701984	1144	904
MeLys	116.19	0.03316884	10	2
Arg	119.40	5.44686484	1092	448
diMeLys	129.29	0.06818636	15	4
MeArg	132.23	0.03084935	3	1
triMeLys	142.58	0.05100000	36	2

In order to confirm the findings quantification of the histone PTMs was carried out for the batch of cells labelled with either Arg or Lys (Table 6-7, Figure 6-22 A and B correspondingly). The quantification of the diMeLys might be less reliable because it was affected by the drift of the chromatographic baseline (Figure 6-22 B, $120 < t_R < 140$ min). It can be seen that the results are comparable with those obtained previously for the histones labelled with both lysine and arginine (Sample I).

Table 6-7. Individually labelled bulk histone analysis results (Sample II)

N/I=not identified (R is within the confidence interval of a standard); T/A=trace amount (R is below LoD)

Histone Label	AA	tR (min)	IR	C (nM)
L-Lys 13C6,15N2	HyLys	99.17	0.01149553	N/I
	Lys	103.14	8.11292414	668
	MeLys	116.233	0.06899527	5
	diMeLys	129.27	0.09759335	7
	triMeLys	142.63	0.05907491	4
L-Arg 13C6,15N4	HyArg	92.15	0.01662581	N/I
	Arg	119.48	3.59279692	295
	MeArg	132.12	0.01452200	N/I

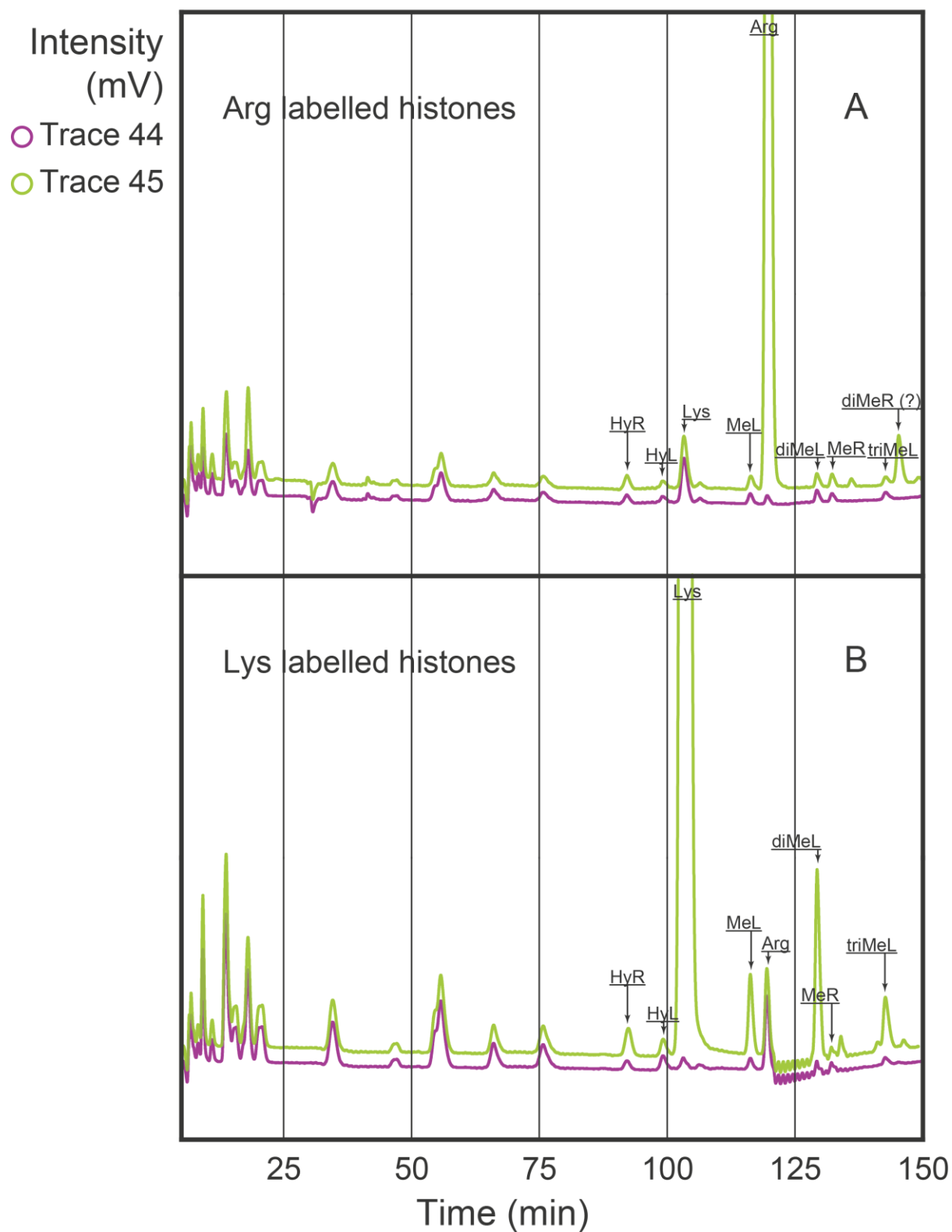


Figure 6-22. Bulk core histones labelled with Arg, Lys or Pro (Sample II)

The next stage of validation was to test whether the method could be used with samples separated by gel electrophoresis. SDS PAGE gel electrophoresis remains one of the most commonly used methods of protein separation, however, this method can lead to significant losses at the sample preparation stage. It has been shown that the losses can vary in the range 20 to 66%³⁶.

In order to test whether the quantification results are affected by preparation of the sample using electrophoresis, as compared against bulk histones, another batch of individually labelled histones was subject to 18% SDS PAGE gel separation. 10 µL of bulk histones were loaded on the gel, which corresponds to the amount used for the bulk histone analysis. As a result of gel separation histones H2A, H2B and H3 were extracted (Figure 6-23). The gel spots were cut out and in-gel tryptic digestion was carried out in order to release separated histones. Extracted individual histones were vacuum-dried and subjected to HCl hydrolysis.

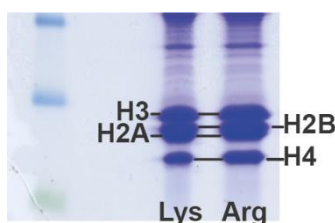


Figure 6-23. Individual AA labelled histones separated by gel

The quantification was carried out in the same manner as for the bulk histones (Table 6-8). Separation of individual histones and their subsequent analysis by excising the gel bands for acid hydrolysis and subsequent LC-IRMS analysis was successful and the losses, if any, were not significant (Table 6-7 vs. Table 6-8). It is important to notice, that variation in the level of lysine and arginine methylation of different histones was observed when the three samples were compared.

Table 6-8. Individually labelled gel separated histone analysis results (Sample III)

N/I=not identified (R is within the confidence interval of a standard); T/A=trace amount (R is below LoD)

Histone Label	AA	C (nM)		
		H2A	H2B	H3
L-Lys 13C6,15N2	Lys	131	162	259
	MeLys	6	N/I	1
	diMeLys	9	T/A	1
	triMeLys	5	N/I	T/A
L-Arg 13C6,15N4	Arg	70	24	38
	MeArg	N/I	N/I	N/I

Table 6-9 provides a summary of the results for methylation of the Samples I, II and III analysed. It is important to point out the difference in absolute Arg and Lys concentrations between batches 1 and 2. This could result from the inaccuracy associated with quantification of the quantity of histones using the total protein assay. The difference in Arg and Lys concentrations between batches 2 and 3 is more likely to be due to the fact that only H2A, H2B and H3 histones were subject to the analysis. In the second part of Table 6-9 the quantities of modifications are normalised by the quantity of the precursor either Lys or Arg. Importantly, of the Lys modifications, it can be seen that for all three batches diMeLys is the most abundant modification, and triMeLys is the least abundant. Based on the available data, the inter-laboratory variation is therefore greater than the variation between the batches of cells grown in one laboratory.

Table 6-9. Quantification results comparison

N/I=not identified (R is within the confidence interval of a standard)

Batch 1 – both Arg and Lys labelled histones (provided by Dr. M. Mackeen); analysed in bulk

Batch 2 – individually labelled with either Lys or Arg histones; analysed in bulk

Batch 3 – individually labelled with either Lys or Arg histones; sum of H2A, H2B and H3 analysed individually

AA	C (nM)			%		
	Sample I	Sample II	Sample III	Sample I	Sample II	Sample III
Lys	904	668	552	100	100	100
MeLys	2	5	7	0.22	0.75	1.27
diMeLys	4	7	10	0.44	1.05	1.81
triMeLys	2	4	5	0.22	0.60	0.91
Arg	448	295	132	100	100.00	100
MeArg	1	N/I	N/I	0.22	0.00	0.00

Several methods have been used for the quantitative analysis of histone PTMs³⁷⁻³⁹. These methods allow relative quantification of particular histone peptides and their modifications. It was possible to find only one publication providing relative abundance of the particular PTM at different sites³³. In order to derive the relative abundance of particular PTM of the whole histone the relative abundance of such PTM at different sites was summed up and divided by the number of relevant AA residues (Table 6-10).

Table 6-10. Summary of the histone Lys and Arg modifications

Information on the modification abundance is adapted from Tweedie-Cullen *et al.*³³

Information on the number of residues is adapted from UniProtKB.

Lys # Res. – number of Lys residues; Arg # Res. – number of Arg residues.

AA	H1	H2A	H2B	H3	H4	Bulk
Lys # Res.	56	14	20	13	11	114
MeLys %	0.1	0.2	0.1	4.5	1.2	0.7
diMeLys %	0.1	1.4	0.1	15.5	4.4	2.4
triMeLys %	0.0	0.0	0.0	2.0	1.2	0.3
Arg # Res.	6	12	8	18	14	58
MeArg %	0.0	0.2	0.0	0.3	1.1	0.4
diMeArg %	0.0	0.0	0.0	0.1	4.1	1.0

Figure 6-24 compares the quantitative data for bulk histone analysis reported in this thesis with the previously published data on relative histone PTMs quantification using proteomics³³. The analysis of individual histones following gel separation shows smaller quantities than those reported. It is possible this is the result of losses during the histone extraction.

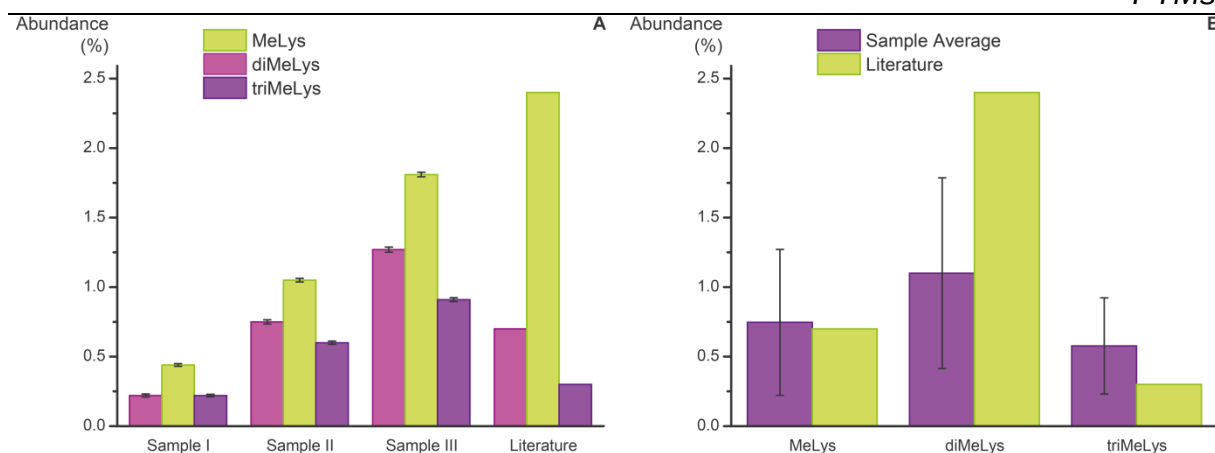


Figure 6-24. Methylated lysines relative abundance.

- Comparison of individual samples with the literature; error bars show the confidence intervals, based on the results of calibration and the errors of the standard measurements.
- Comparison of the sample (I, II and III) average with literature; error bars show SD for Samples I, II and III.

6.5 Conclusions and Further Work

This chapter has described the development and application of a novel method suitable for the identification and absolute quantification of AAs resulting from PTMs at low abundance in cell culture. Importantly, even though the method was developed and tested for the analysis of AAs, the conclusions made for the quantitative analysis may be valid for any class of organic compounds. The limiting factors are chromatography, which has to be developed for every class of compounds, and the availability of standards.

It was shown that the method developed provides a robust analytical platform for the exploratory analysis of histone PTMs. Selective isotopic labelling allows us to trace the modified species back to its metabolic precursor, which is also useful in establishing metabolic pathways. Mixed mode LC coupled to isotope ratio MS was shown to be a powerful tool for absolute quantitative analysis with the LoD at the lower nanomolar level. It was shown that the method is compatible with a gel electrophoresis protein extraction procedure.

Two batches of HEK 293T cells discussed in this chapter were grown in different laboratories and for the second batch two different passages were analysed (Figure 6-18). The Lys methylation pattern remains constant for all three samples with maximum abundance of diMeLys and least abundant MeLys. It is important to point out that the pattern observed for the three samples is in agreement with the previously published data (Table 6-10, Figure 6-24). It is also important to notice, that between-sample variation (variation between different batches of cells) appears to be greater than the error of the method developed.

One way to reduce the between-sample variation is to increase the precision of the quantification of the initial material, which is necessary to evaluate the biological levels of each modification. IRMS can provide very accurate quantification results. In order to make those applicable to the cell based work it is necessary to put in place quantitative methods for the particular intact protein in which the modifications are to be quantified using the LC-IRMS method. It would also be advantageous to explore alternatives to gel electrophoresis for preparative separation of peptides. The use of chromatographic techniques is currently being pursued. This should help to minimise sample preparation time by removing the gel extraction step, saving around two days (Chapter Eight). One alternative to gel electrophoresis is LC coupled to a UV or diode array detector which might allow both fast separation and accurate quantification of the proteins.

6.6 References

- (1) Aston, F. W. *Isotopes*; Edward Arnold & Co.: London, 1922.
- (2) Nier, A. O. *Review of Scientific Instruments* **1947**, *18*, 398.
- (3) Matthews, D. E.; Hayes, J. M. *Analytical Chemistry* **1978**, *50*, 1465.
- (4) Godin, J. P.; Fay, L. B.; Hopfgartner, G. *MS Reviews* **2007**, *26*, 751.
- (5) Krummen, M.; Hilkert, A. W.; Juchelka, D.; Duhr, A.; Schluter, H. J.; Pesch, R. *Rapid Communications in MS* **2004**, *18*, 2260.
- (6) Werner, R. A.; Brand, W. A. *Rapid Communications in MS* **2001**, *15*, 501.
- (7) Kracht, O., Isodate Workspace Software Request - result sheet.
- (8) Bigeleisen, J. *Journal of Physical Chemistry* **1952**, *56*, 823.

- (9) Scott, J. H.; O'Brien, D. M.; Emerson, D.; Sun, H.; McDonald, G. D.; Salgado, A.; Fogel, M. L. *Astrobiology* **2006**, 6, 867.
- (10) Bottinga, Y. *Journal of Physical Chemistry* **1968**, 72, 800.
- (11) Matsuhisa, Y.; Goldsmith, J. R.; Clayton, R. N. *Geochimica et Cosmochimica Acta* **1979**, 43, 1131.
- (12) Larsen, T.; Taylor, D. L.; Leigh, M. B.; O'Brien, D. M. *Ecology* **2009**, 90, 3526.
- (13) Tieszen, L. L.; Boutton, T. W.; Tesdahl, K. G.; Slade, N. A. *Oecologia* **1983**, 57, 32.
- (14) Craig, H. *Geochimica et Cosmochimica Acta* **1957**, 12, 133.
- (15) McCullagh, J. S. O.; Juchelka, D.; Hedges, R. E. M. *Rapid Communications in MS* **2006**, 20, 2761.
- (16) McCullagh, J.; Gaye-Siessegger, J.; Focken, U. *Rapid Communications in MS* **2008**, 22, 1817.
- (17) McCullagh, J. S. O. *Rapid Communications in MS* **2010**, 24, 483.
- (18) Shemin, D.; Foster, G. L. *Annals of the New York Academy of Sciences* **1946**, 47, 119.
- (19) Godin, J. P.; Hau, J.; Fay, L. B.; Hopfgartner, G. *Rapid Communications in MS* **2005**, 19, 2689.
- (20) Boschker, H. T. S.; Moerdijk-Poortvliet, T. C. W.; van Breugel, P.; Houtekamer, M.; Middelburg, J. J. *Rapid Communications in MS* **2008**, 22, 3902.
- (21) Pickup, J. F.; McPherson, K. *Analytical Chemistry* **1976**, 48, 1885.
- (22) Colby, B. N.; Rosecrance, A. E.; Colby, M. E. *Analytical Chemistry* **1981**, 53, 1907.
- (23) Colby, B. N.; McCaman, M. W. *Biomedical MS* **1979**, 6, 225.
- (24) Schoeller, D. A. *Journal of Clinical Pharmacology* **1986**, 26, 396.
- (25) Jonckheere, J. A.; Deleenheer, A. P. *Analytical Chemistry* **1983**, 55, 153.
- (26) Thienpont, L. M.; VanNieuwenhove, B.; Stockl, D.; DeLeenheer, A. P. *Journal of MS* **1996**, 31, 1119.
- (27) van Eijk, H. M. H.; Rooyakkers, D. R.; Soeters, P. B.; Deutz, N. E. P. *Analytical Biochemistry* **1999**, 271, 8.
- (28) Santrock, J.; Studley, S. A.; Hayes, J. M. *Analytical Chemistry* **1985**, 57, 1444.
- (29) Assonov, S. S.; Brenninkmeijer, C. A. M. *Rapid Communications in MS* **2003**, 17, 1017.
- (30) Allison C.E., F. R. J., Meijer H.A.J. In *Reference and intercomparison materials for stable isotopes of light elements* Vienna, 1993, p 155.
- (31) Samuel-Landtiser, M.; Zachariah, C.; Williams, C. R.; Edison, A. S.; Long, J. R. In *Current Protocols in Protein Science*; John Wiley & Sons, Inc.: 2001.
- (32) Bainor, A.; Chang, L.; McQuade, T. J.; Webb, B.; Gestwicki, J. E. *Analytical Biochemistry* **2011**, 410, 310.
- (33) Tweedie-Cullen, R. Y.; Brunner, A. M.; Grossmann, J.; Mohanna, S.; Sichau, D.; Nanni, P.; Panse, C.; Mansuy, I. M. *Plos One* **2012**, 7.
- (34) Lößner, C.; Warnken, U.; Pscherer, A.; Schnölzer, M. *Analytical Biochemistry* **2011**, 412, 123.
- (35) Park, S. K.; Liao, L.; Kim, J. Y.; Yates, J. R. *Nat Meth* **2009**, 6, 184.
- (36) Zhou, S.; Bailey, M. J.; Dunn, M. J.; Preedy, V. R.; Emery, P. W. *PROTEOMICS* **2005**, 5, 2739.
- (37) Knapp, A. R.; Ren, C.; Su, X.; Lucas, D. M.; Byrd, J. C.; Freitas, M. A.; Parthun, M. R. *Methods* **2007**, 41, 312.
- (38) van Leeuwen, F.; Gafken, P. R.; Gottschling, D. E. *Cell* **2002**, 109, 745.

(39) Beck, H. C.; Nielsen, E. C.; Matthiesen, R.; Jensen, L. H.; Sehested, M.; Finn, P.; Grauslund, M.; Hansen, A. M.; Jensen, O. N. *Molecular & Cellular Proteomics* **2006**, 5, 1314.

7 Chapter Seven – Summary

This chapter summarises the findings described in this thesis, the methods developed and tested for the analysis of PTMs, and makes some comment on the potential for future applications of AA analysis in the characterisation of PTMs. Table 7-1 provides a summary of the chromatographic approaches tested and the LoD analyte compatibility that were experimentally determined. It is hoped that the results will aid in facilitating the choice of method for the analysis of PTMs in the future.

It is important to characterise the niche which AA analysis may occupy amongst the many analytical methods dealing with PTMs. The approaches involving various forms of AA analysis described in this thesis are not suggested as an alternative to the well established proteomic techniques, rather as being complementary to them. Compared to proteomics, AA analysis lacks the power to establish the position of modified residues within proteins, AA analysis does, however, enable investigation into the region- and stereo- chemistry of the modified groups within the modified AA residue and has proved to be a powerful tool of exploratory PTM analysis.

The major aim of my work was to develop analytical methods to enable the identification and quantification of PTMs using AA analysis. The methods developed were shown to be useful for a broad range of samples including synthetic mixtures, *in vitro* enzymatic assays and cell based experiments, and potentially applicable to *in vivo* experiments. Apart from naturally occurring modifications the methods were also found to be useful for the analysis of synthetic adducts of AAs, as well as non-natural AAs.

Table 7-1. Summary of the analytical method developed and their limits of detection

RP = reversed phase; IP = ion pairing; MS = MS; scan = scanning mode; SIM = selected ion monitoring mode; SRM = selected reaction monitoring mode;

RSD = relative standard deviation; LoD = limit of detection

Chromatography Type	1nM	10nM	LoD 100nM	1µM	10µM
RP AQC derivatisation total run time ~ 15 min ☑ robust chromatography resolves diastereomers ☒ extra measures should be taken when working with buffers with primary and secondary amino groups RP-IP total run time ~ 6 min ☑ robust chromatography compatible with all buffers IP reagent enables tuneable resolution ☒ IP reagent affects the negative detection mode Mixed Mode total run time ~ 25 min ☑ robust chromatography compatible with all buffers ☒ requires equilibration after use of different mobile phases (up to 24 h) Mixed Mode Aqua total run time ~ 2 h ☑ NO organic solvents compatible with all buffers ☒ difficult to adjust resolution			UV RSD < 1% at C > 1 µM ☒ not selective, LoD for real samples is expected to be higher MS SIM RSD < 20% at C > 1 µM ☑ selective detection	MS Scan RSD < 10% at C > 1 µM ☑ high resolution ☒ charge state has to be verified for AAs with multiple amino groups MS SIM RSD < 5% at C > 1 µM ☑ selective detection MS SIM RSD < 15% at C > 5 µM ☑ selective detection	MS SRM RSD < 30% at C > 1 µM ☑ selective detection suitable for very complicated matrices ☒ charge state has to be verified for AAs with multiple amino groups
		IRMS for A(13C) = 20%, nC=5 RSD < 10% at C > 10 nM; RSD < 1% at C > 100 nM ☑ suitable for tracking the metabolic pathways compatible with all buffers ☒ LoD is dependent upon level of labelling and nC isotopic labelling is expensive and time consuming non specific detector, alternative methods of identification are required			

AA analyses of a number of proteins from the super-family of 2OG- and Fe(II)- dependent oxygenases were carried out. These enzymes catalyse hydroxylation of AAs as well as regulate demethylation of lysyl and arginyl residues, oxidative cyclisation and double bond formation. On-going research into the function of 2OG oxygenases requires robust and sensitive analytical methods that enable quantification and identification of the products in a wide range of samples, ranging from *in vitro* assays to *in vivo* experiments. Such a broad spectrum of samples requires methodologies which works over a broad range of concentrations as well as being capable of dealing with a range of different matrices.

One area that was a particular challenge was the need for modified AA standards to enable identification, however, many modified AAs are not commercially available. In order to find a way to solve this problem the use of retention indices was investigated and developed which allows identification of the products on the basis of chromatographic retention of the target compound and the retention of its structural analogues. Often the closest structural analogue is the non-modified AA, which is often available as a reference standard.

A separate problem is the identification of the stereochemistry of PTMs, which has only been addressed before for a few modifications such as hydroxylation of lysyl- and prolyl- residues of collagen. In this thesis AQC modified AA analysis was shown to be a powerful tool to separate the diastereomers of modified AAs. Using this method it was possible to establish the stereochemistry of lysyl hydroxylation as catalysed by JmjD6, and prolyl hydroxylation as catalysed by mbP4H.

The suit of methods developed for AA analysis in this thesis covers the concentration range from low nanomolar to high micromolar levels. Dealing with complex matrices in biochemical assays requires selective detection, so that the target component can be distinguished from interfering matrix components. In this work the selectivity and resolution of MS detectors was utilised to address this problem.

Isotopic labelling for the analysis of PTMs in the cell based experiments enables high selectivity. It allowed labelling of the whole cell with specific AAs and its metabolites, thus, it can highlight a set of metabolic pathways using a common precursor. Isotopic labelling was combined with IRMS to enable very sensitive detection suitable for cell based work. The disadvantage of this technique is that IRMS is not selective and does not allow analysis of individual traces produced by particular compounds, which makes the identification of multiple metabolites reliant on chromatographic resolution for identification.

The methods developed in this work were shown to be applicable both for in vitro and for cell based experiments. It is hoped that their continued application will be used to expand the current knowledge about the 2OG oxygenases. Currently further research into the stereochemistry of PTMs is being carried out using AQC-derivatised AA analysis. The methods developed are also used to study the level of enzymatic activity. The IRMS method developed is also applicable to the study of histone PTMs. The application of quantitative analysis should allow the insight of how hypoxia affects the PTMs of histones. It is expected that the expansion will be possible to apply to the study of PTMs of other proteins including ribosomal proteins.

8 Chapter Eight – Materials and Methods

8.1 Chemicals and Standards

The list of standards used for the work described in this thesis is provided in Table 8-1.

Table 8-1. AA standards used in this work

Name	Supplier
L-ornithine hydrochloride	Sigma
L-Tyrosine	Sigma
L-Phenylalanine	Sigma
Carnitine hydrochloride	Sigma
L-Methionine	Sigma
L-Isoleucine	Sigma
L-Leucine	Sigma
β-Alanine	Sigma
2-Aminoisobutyric acid	Sigma
3-Aminoisobutanoic acid	Sigma
L-Lysine monohydrochloride	Sigma
L-Homoserine	Sigma
Pipecolic acid	Sigma
L-Glutamic acid	Sigma
D,L-2,4-Diaminobutyric acid	Sigma
L-Threonine	Sigma
L-Asparagine	Sigma
L-2-Aminoadipic acid	Sigma
2-Aminoisobutyric acid	Sigma
DL-3-Aminoisobutyric acid	Sigma
DL-2-Aminocaprylic acid	Sigma
Methylated	
Nε-methyl-L-lysine hydrochloride	Sigma
Nε,Nε-dimethyl-L-lysine monohydrochloride	Sigma
Nε,Nε,Nε-trimethyllysine hydrochloride	Sigma
N-methyl-L-arginine acetate	Sigma
N,N-dimethylarginine hydrochloride	Sigma
sarcosine	Sigma
N,N-dimethylglycine	Sigma
1-methyl-L-histidine	Sigma
3-methyl-L-histidine	Sigma
Hydroxylated	
DL-5-hydroxylysine hydrochloride	Sigma
(5R)-5-hydroxy-L-lysine dihydrochloride monohydrate	Sigma
trans-3-hydroxy-L-proline	Sigma
cis-L-3-hydroxyproline	Chem-Impex International
L-4-hydroxyproline	Sigma
cis-4-hydroxy-D-proline	Sigma
3,4-dihydroxy-DL-phenylalanine	Sigma
5-hydroxy-L-tryptophan	Sigma
3L-hydroxy-L-arginine dihydrochloride	Sigma
3R-hydroxy-L-arginine dihydrochloride	Sigma
Acetylated	
N-acetyl-DL-glutamic acid	Sigma
Phosphorelated	
O-phospho-L-serine	Sigma

The derivatising agent for AQC derivatisation was purchased from Waters Ltd (Elstree, Hertfordshire, UK).

pH was verified by pH-metre Mettler Toledo with the Mettler Toledo InLab 413 pH probe, calibrated to pH 7 and 10 before use. The storage solution was 3 M potassium chloride.

Borate buffer of pH 11 was purchased from Sigma Aldrich.

6 M HCl for hydrolysis was purchased from Thermo Fisher Scientific (Northumberland, UK).

Water used for all the experiments was purified by Millipore Elix system with 0.22 μm filter at the outlet.

For the cell based work, SILAC cell-growth media and dialysed fetal bovine serum were purchased from Thermo Fisher Scientific (Northumberland, UK).

8.2 Chromatography Methods

Unless stated otherwise chromatographic separation was performed using Waters Acquity UPLC equipped with UV detector.

8.2.1 Chromatographic Separation of AQC Derivatives of AAs

AccQTag Ultra column (Waters, Milford, USA) was used (reversed phase C18, 2.1 $\times$ 100mm, particles 1.7 μm). The column was operated at flow rate of 0.7 mL $\cdot\text{min}^{-1}$.

Table 8-2. Reversed phase chromatography of AQC AAs mobile phase composition

Phase	Composition			
	H ₂ O	AccQTag Eluent A concentrate	MeCN	FA
A	95 (v)	5 (v)	–	–
B	–	–	100	1.3 (v)

A linear one-step gradient was applied: Phase A 100 to 75% over 25 min. The column was cleaned for 2 min at a higher acetonitrile concentration (40% Phase A) before reconditioning.

Table 8-3. Reversed phase chromatography of AQC AAs mobile phase gradient

Time (min)	% Phase A	Function
0	100	separation
25.0	75.0	
25.5	40.0	
27.5	40.0	cleaning
28.0	100.0	reconditioning
30.0	100.0	

The separation was further adjusted by applying shallower or steeper gradient, or by adding extra steps to the separation process.

The column temperature was kept at 55°C. The injection volume was 1 μ L. UV detection was performed at 260 nm wavelength. The flow was split at 1:1 ratio before entering into the mass spectrometer.

8.2.2 Chromatographic Separation of the Non-Derivatised AA Using Ion-Pairing Chromatography

BEH column (Waters, Milford, US) was used (reversed phase C18, 2.1 $\times$ 50mm, particles 1.7 μ m). The column was operated at flow rate of 0.8 mL $\cdot$ min⁻¹.

Table 8-4. Ion pairing chromatography of non-derivatised AAs mobile phase composition

Phase	Composition		
	H ₂ O : MeCN	FA	TDFHA
A	99.5 : 0.5 (v)	0.1 (v)	2 mM
B	10 : 90 (v)	0.1 (v)	2 mM

H₂O and AcN were mixed and 0.1% by volume of FA was added. TDFHA was added at 2 mM concentration. These mixtures were sonicated until the ion pairing reagent was completely dissolved (*ca.* 30 min). It was shown that concentration of the ion pairing reagent can be reduced to 0.5 mM without compromising separation efficiency¹.

A linear one-step gradient (Table 8-5) was applied: Phase A 100 to 75% in 25 min. The column was cleaned for 2 min at a higher acetonitrile concentration (40% Phase A) before reconditioning.

Table 8-5. Ion pairing chromatography of non-derivatised AAs mobile phase gradient

Time (min)	% Phase A	Function
0	99.9	separation
6.0	70	
10.0	99.9	reconditioning

The chromatography can be simplified by removing the FA from the mobile phase, thus reducing the number of the components of the mixture. The presence of perfluorinated acids enabled protonation of the analyte for the mass-spectrometric detection.

For the samples with complex biological matrices (cell extracts) the cleaning step at higher percentage of acetonitrile was introduced.

The column temperature was kept at 45°C. The injection volume was 1 µL and partial loop with needle overfill was used. SIM mode was used for detection.

8.2.3 Chromatographic Separation of Non-Derivatised AA Using Mixed Mode Chromatography

PrimeSep 200 column (SIELC, Prospect Heights, US) was used (mixed mode, 2.1×250mm, particles 5µm). The column was operated at flow a rate of 0.3 mL·min⁻¹ (Table 8-6).

Table 8-6. Mixed mode chromatography of non-derivatised AAs mobile phase composition

Phase	Composition		
	H ₂ O	MeCN	FA
A	90 (v)	10 (v)	0.05 (v)
B	80 (v)	20 (v)	0.2 (v)

Linear one-step gradient (Table 8-7) was applied: Phase A 100 to 0% in 25 min. The column was cleaned for 2 min at higher acetonitrile concentration (40% Phase A)

Table 8-7. Mixed mode chromatography of non-derivatised AAs mobile phase gradient

Time (min)	% Phase A	Function
0	100	separation
25.0	0	
26.0	100	
30.0	100	reconditioning

The chromatography can be adjusted by using 0% MeCN Phase A and a lower concentration of formic acid in Phase A, but not lower than 0.01%, this is necessary to enable positive ion detection. The cleaning step can be introduced with higher percentage of acetonitrile and formic acid.

The column temperature was kept at 45°C. The injection volume was 2 µL. SIM mode was used for detection. Signal acquisition was performed in one function and scanning mode was used in parallel as a control.

8.2.4 Chromatographic Separation of the Non-Derivatised AA by MM Chromatography Using Carbon Free Mobile Phase

The chromatographic method was adapted from the one reported previously². Chromatographic separation was performed using Surveyor HPLC system.

A PrimeSep 100 column (SIELC, Prospect Heights, US) was used (mixed mode, 2.1×250mm, particles 5µm).

Table 8-8. MM chromatography of non-derivatised AAs mobile phase composition

Phase	Composition	
	H ₂ O	H ₂ SO ₄
A	100	—
B	100	0.2 (v)

The flow rate was maintained at 0.7 mL/min. The column was operated at ambient temperature. The gradient used was as follows:

Table 8-9. MM chromatography of non-derivatised AAs mobile phase gradient

Time (min)	% Phase A
0	0
10	0
30	100
110	100

An injection volume of 10 µL was used.

8.2.5 List of Columns Used

Type	Dimensions	Particle Size	Brand Name	Manufacturer
Reversed Phase C18	2.1×100mm	1.7µm	AccQTag Ultra	Waters
Reversed Phase C18	2.1×100mm	1.7µm	BEH	Waters
Reversed Phase C18	2.1×50mm	1.7µm	BEH	Waters
Mixed Mode CE	2.1×250mm	5µm	PrimeSep A	SIELC
Mixed Mode CE	2.1×150mm	2µm	PrimeSep A	SIELC
Mixed Mode CE	2.1×250mm	5µm	PrimeSep 100	SIELC
Mixed Mode CE	2.1×250mm	5µm	PrimeSep 200	SIELC
Mixed Mode CE	2.1×250mm	5µm	PrimeSep 500	SIELC
Mixed Mode CE	2.1×250mm	5µm	PrimeSep D	SIELC

8.2.6 Derivatisation with 6-Aminoquinolyl-N-Hydroxysuccinimidyl Carbamate

AA hydrolysate standard contains 2500 pmoles/µL of each AA, cystine was the exception at 125pmol/µL.

Samples were stored in sealed septa at room temperature for up to 1 week.

AQC reagent was dissolved in 1mL of dry acetonitrile.

The mixture was vortexed briefly (*ca.* 10 sec) and heat (10mins) at 55°C

Dilute 100µL of AAA hydrolysate standard with 900µL of 0.1M HCl. Resulting concentration of each AA is 250pmol/µL, except for cystine – 125pmol/µL.

Derivatisation

The reagents were added in the following order: (1) borate buffer, (2) standard solution OR sample and (3) derivatizing agent. The mixture was vortexed briefly (*ca.* 10 sec), and heated for 10mins at 55°C to accomplish the derivatisation reaction.

Blank, Reagent Blank and AA standard or sample is made up as follows:

	Buffer (µL)	AA (µL)	Derivatizing Solution (µL)
Blank	100	–	–
Reagent Blank	80	–	20
Stock or Sample	70	10	20

Resulting concentration of each AA is 25pmol/μL, except for cystine which was 12.5pmol/μL.

The concentration of a particular AA in a peptide or protein is described by the following equation:

$$C_{AA} = \frac{A_{AA} \cdot Q_P}{MW_{AA} \cdot V_S}$$

C_{AA} – concentration of AA in sample (pmoles/μL)

A_{AA} – abundance of AA in protein(%), *i.e.* mass of AA (mg) in a protein divided by the weight of the protein (mg).

Q_P – quantity of protein in a sample (g)

MW_{AA} – molar weight of AA (g/moles)

V_S – volume of the sample (μL)

For example, for collagen, the least abundant AA is hydroxylysine (MW=162g/mole) at 0.62%. Therefore, a 1 mg/mL solution of collagen hydrolysate corresponds to the 38 pmole/μL concentration for hydroxylysine in the mixture.

$$C_{hydroxylysine} = \frac{A_{hydroxylysine} \cdot Q_P}{MW_{hydroxylysine} \cdot V_S} = \frac{0.0062 \cdot 0.001g}{162 \text{ g/mole} \cdot 1000\mu L} = 38 \text{ pmoles}/\mu L$$

Derivatising agent should be calculated for the total quantity of amino function in our solution. It is reasonable, therefore, to use an average molar weight of AA in protein, which is usually estimated to be 110g/mole.

$$C_{AF} = \frac{Q_P}{AMW_{AA} \cdot V_S}$$

C_{AF} – concentration of amino function in sample (pmoles/μL)

Q_P – quantity of protein in a sample (g)

AMW_{AA} – average molar weight of AA (g/moles)

V_S – volume of the sample (μL)

For example, the 1 mg/mL solution of collagen hydrolysate corresponds to the 38 pmoles/ μL concentration of amino function.

$$C_{AF} = \frac{Q_P}{AMW_{AA} \cdot V_S} = \frac{0.001g}{110 g/mole \cdot 1000\mu L} = 9.1 \text{ nmoles}/\mu L$$

Calculation of Derivatising Solution Volume

The concentration of the derivatizing agent in the Waters AAA kit is between 10 and 14 mole/ μL (average 12 mole/ μL); and it is usually used in 3 times excess. The volume of derivatizing solution can, therefore, be calculated as follows:

$$V_{DS} = F_E \cdot \frac{C_{AF} \cdot V_S}{C_{DA}}$$

V_{DS} – volume of derivatizing solution (μL)

C_{AF} – concentration of amino function in sample (nmoles/ μL)

Q_P – quantity of protein in a sample (g)

AMW_{AA} – average molar weight of AA (g/moles)

C_{DA} – concentration of derivatizing agent (nmoles/ μL)

F_E – excess factor

E.g.: 23 μL of derivatisation solution are required for derivatisation of 10 μL of sample solution at 1mg/mL concentration of collagen.

$$V_{DS} = F_E \cdot \frac{C_{AF} \cdot V_S}{C_{DA}} = 3 \cdot \frac{9.1 \text{ nmoles}/\mu L \cdot 10\mu L}{12 \text{ nmoles}/\mu L} = 23\mu L$$

For the sake of convenience this volume can be rounded down to 20 μL .

Storage and Stability

The derivatising reagent was stored at room temperature.

MeCN solution for the derivitisation reagent was stored in at 4°C.

Derivatised AAs samples were stored at 4°C while in the auto sampler and discarded immediately after the analysis.

8.3 Detection Methods

8.3.1 UV Detection

UV detection of AQC-derivatised AAs was carried out using a Waters Acquity UV detector operating at 254 nm.

8.3.2 SIM and SRM Mode MS Detection

MS detection in SIM and SRM modes was carried out using a Quattro Micro triple quadrupole mass spectrometer (Waters, Milford, US).

MS detection of the AAs was performed in positive V mode. The capillary voltage was at 3000 V; extractor voltage was at 6 V, and RF lens was at 0.8 V. The desolvation temperature was 200°C, and the source temperature was 120°C. The cone gas flow was at 30 L/min and the desolvation gas flow was 800 L/min.

For ion paring chromatography of non-derivatised AAs; detection of AAs was made using functions: (1) 0.0 – 0.8 min, (2) 0.8 – 1.0 min, (3) 1.0 – 1.7 min, (4) 2.3 – 3.2 min, (5) 4.0 – 4.7 min, (6) 4.7 – 5.5 min. The cone voltages and collision energies used for individual AAs are listed in Table 8-10. Cone voltage and collision energy were identified for the individual AAs by direct injection of 100 µM solution at 5 µL×min⁻¹. The conditions for every function are presented in Table 8-10.

Table 8-10. Mass spectrometer parameters for SIM and SRM modes of detection of the non-derivatised AAs

Name	SIM Mode		SRM Mode			tR	Acquisition Function
	Precursor	Cone	Fragment	Cone	Collision		
HyP	132.2	16	86	19	13	0.43	1
Asp	134.1	16	79	15	10	0.43	1
Ser	106	13	59.9	14	13	0.5	1
Gly	76	13	30	14	11	0.58	1
Glu	148.1	13	84	14	19	0.66	1
Thr	120.1	13	74	13	11	0.69	1
Ala	90.1	14	44	14	11	0.9	2
Pro	116.1	17	70	16	17	1.14	3
Tyr	182.2	12	136.2	13	15	2.66	4
Met	150.2	11	104.1	12	11	2.84	4
Val	118.1	13	72	13	11	2.94	4
Ile	132.2	13	86.1	13	11	4.25	5
Leu	132.2	13	86.1	13	11	4.44	5
Phe	166.2	14	120.1	13	16	4.54	5
Cys	122.1	31	76	33	17	4.94	6
His	156.2	15	110.1	17	16	4.99	6
HyL	163.2	10	128.1	11	15	5.19	6
Lys	147.2	13	84.1	14	18	5.34	6
Arg	175.2	17	70	17	32	5.43	6

For the reversed phase chromatography of the derivatised AAs the detection was arranged in seven acquisition functions: (1) 0.0 – 2.8 min, (2) 2.8 – 3.6 min, (3) 3.6 – 4.2 min, (4) 4.2 – 5.2 min, (5) 5.2 – 6.1 min, (6) 6.1 – 7.2 min, (7) 7.2 – 8.5 min. The conditions for every function are presented in Table 8-11.

Table 8-11. Mass spectrometer parameters for SIM mode of detection of the derivatised AAs

Name	Trace	tR	Function
HyP	302.2	2.07	1
His	326.2	2.28	1
Ser	276.2	3.17	2
Arg	345.2	3.35	2
Gly	246.2	3.48	2
Asp	304.2	3.83	3
Glu	318.2	4.33	4
Thr	290.2	4.7	4
Ala	260.2	5.08	4
Pro	286.2	5.65	5
HyL	503.2	5.9	5
Cys	291.2	6.56	6
Lys	487.2	6.62	6
Tyr	352.2	6.75	6
Met	320.2	6.89	6
Val	288.2	7.02	6
Ile	302.2	7.74	7
Leu	302.2	7.82	7
Phe	336.2	7.94	7

8.3.3 Scanning Mode MS Detection

MS detection in the scanning mode was carried out using a LCT Premier ToF mass spectrometer (Waters, Milford, US).

MS detection was performed in positive ion V mode. The capillary voltage was at 3000 V, and the cone voltage was at 100 V. The desolvation temperature was 250°C and source temperature was 120°C. The cone gas flow was at 30 L/min and desolvation gas flow was 550 L/min. The scanning range was 160 to 510 Da.

8.3.4 IRMS Detection

In this work a previously published method² was adapted. A DELTA V Advantage isotope ratio mass spectrometer was coupled to the HPLC system via Finnigan LC IsoLink interface (Thermo Electron Corp., Bremen, Germany). The Thermo LC IsoLink interface is composed of a liquid phase oxidation reactor, CO₂ separation unit, water separation system and an open split assembly.

Oxidation in the reactor is carried out using following reagents: sodium peroxodisulfate (0.8 M) entrained with phosphoric acid (1.7 M). The oxidation reagents are pumped separately at a flow rate 30 µL/min and were added to the eluate after separation. The oxidation reactor was maintained at 99.9°C.

Under these conditions all the organic compounds are converted to CO₂, which is removed from the liquid phase by a Gortex gas permeable membrane, dried and carried to the IRMS detector by the flow of helium. Helium flow rate is maintained at 2 mL/min².

8.4 Chromatography MS Data Processing

8.5 Signal processing

Signal processing was performed using the QuanLynx software (Waters, Milford, US). All the traces were smoothed prior to the analysis. Two types of smoothing were employed: moving average (1 iteration, 5 scans window) for the UV trace, and Savitzky-Golley (1 iteration, for the window corresponding to the number of scans of the half width of the peak).

8.5.1 Mass Chromatogram Data Presentation

I initially discuss the ways of presenting the chromatography MS data. Three types of data presentation are provided in Figure 8-1. Graph A corresponds to the traditional way of presenting MS chromatography peaks. This approach however can be misleading if one considers the lines joining the data point to be characteristic of the linear trend observed for the signal. Presenting the peak as it is shown in graph B can also be misleading, because the dot represents the signal at a particular time point, where as in reality the MS signal was acquired over certain time span. Apparently, in order to better reflect the reality of MS chromatographic data collection can be presented as bar charts, as given in graph C. It is worth pointing out that graph B corresponds to an adequate representation of the UV signal. It is worth bearing in mind therefore, that even though both UV and MS traced are normally presented as given in graph A, they have very different nature.

Choice of the mass window is important when creating an extracted ion chromatogram. As an example I will consider the peak of aminoquinoline carbamate derivatised hydroxylysine presented In Figure 8-2. As a criterion I will use signal to noise ratio, which is calculated as average (by scans) intensity of the signal related to the average intensity of the noise. Even though this criterion should not be used in general for the assessment of the analytical method, the signal to noise ratio can be a robust test for when the data processing is concerned.

The starting point is the total ion chromatogram, presented in graph D. It can be seen that by narrowing down the window we decrease significantly the overall intensity, but the signal to noise ratio is increasing. Graph B illustrates the situation in which the selected narrow mass window does not correspond to the maximum of the mass peak, as a result of which the signal to noise ratio is reduced in comparison to the broader window, which contains though the mass peak maximum.

Another important aspect of presenting mass chromatogram is to show the base line at zero of the intensity scale. It is important to point out, that unlike UV trace mass chromatograms have an absolute zero and showing the zero line clearly provides valuable information. For example in Figure 8-2 graph D showing the chromatogram from 750 of the intensity scale, might seem reasonable, since the peak becomes more pronounced and noise is treated as irrelevant. In fact such data representation – especially in cases when the intensity axes scale is not provided – can give a false impression of the higher signal to noise ratio.

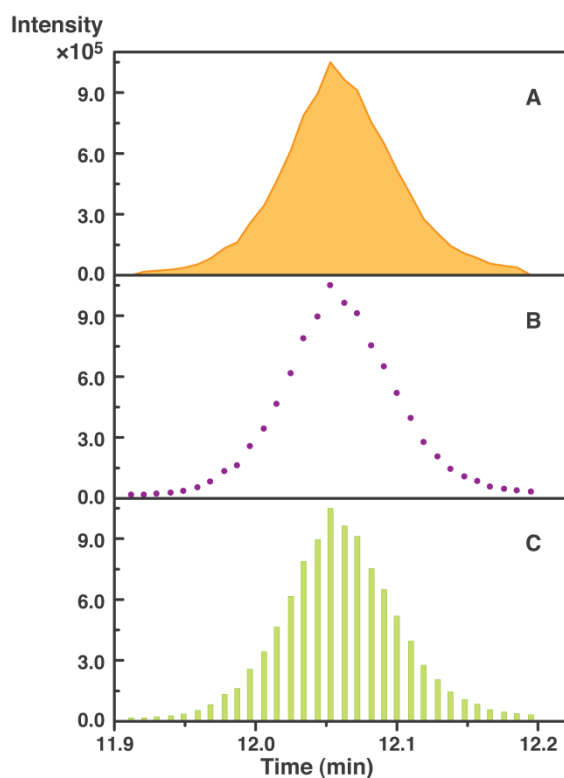


Figure 8-1. Ways of Presenting the Mass Spectrometry Chromatography Peak

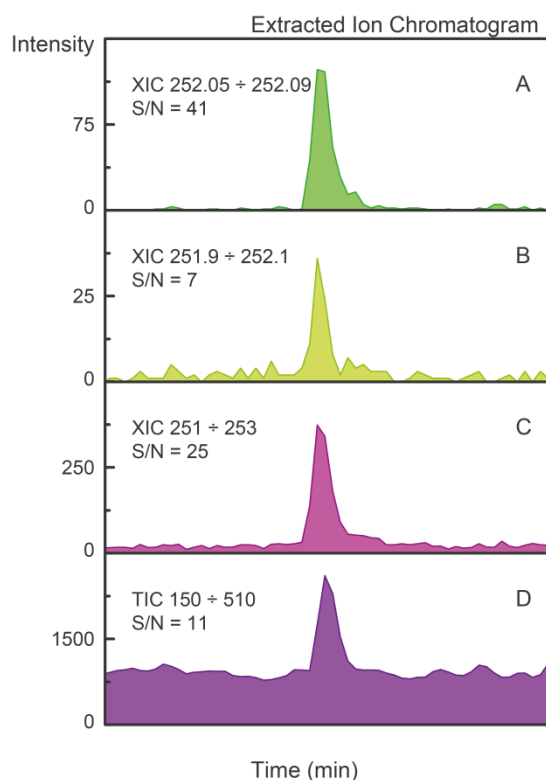


Figure 8-2. Extracted Ion Chromatogram Mass Window Selection

8.5.2 Data Smoothing

Either area of the chromatographic peak or its height can be used for the quantification of the compound. As an example raw MS chromatographic peak of aminoquinoline carbamate derivatised aspartate is presented in Figure 8-3. An attempt to integrate such peak and to identify its area will result in splitting the peak into multiple peaks, which in due course will result in poor reproducibility of the peak integration. The currently available software does not allow the work with raw data. In order to get the peak parameters it is necessary to further process such data.

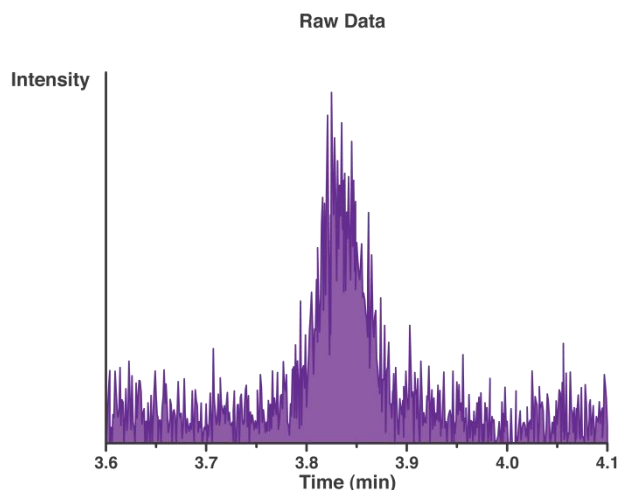


Figure 8-3. Raw LCMS Data.

Quatro Micro; AccQ-Asparagine; SIM ($m/z = 304.2$)

Normally, in order to make the integration possible peak smoothing is applied. Two peak smoothing algorithms are generally used: moving average or Savitzky-Golay³. Two algorithms are compared using the above raw data as an example. The width of the peak is about 130 data points. The results of smoothing the using different smoothing window width (from 10 to 70 points) are presented in Figure 8-4 and Figure 8-5. It can be seen that Savitzky-Golay algorithm has hardly any effect on the peak shape, whereas applying the moving average algorithm results in broadening of the peak and a decrease in height.

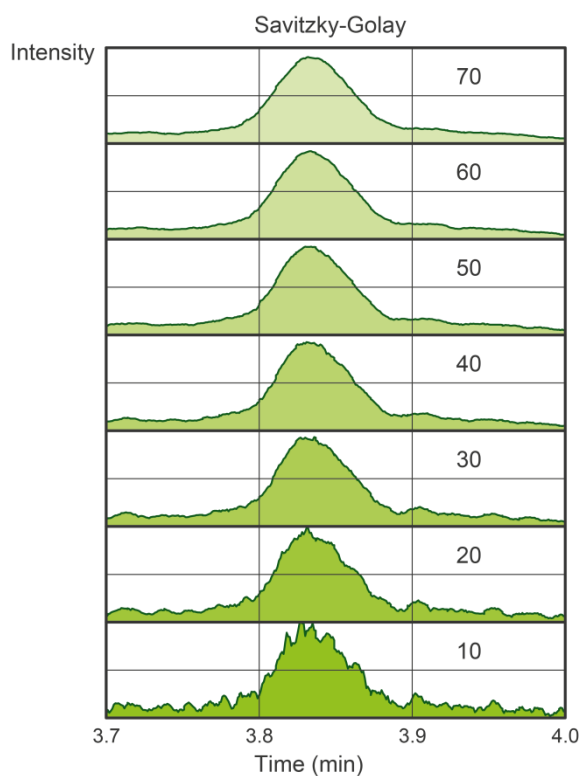


Figure 8-4. Savitzky-Golay Smoothing Using Different Smoothing Window Width

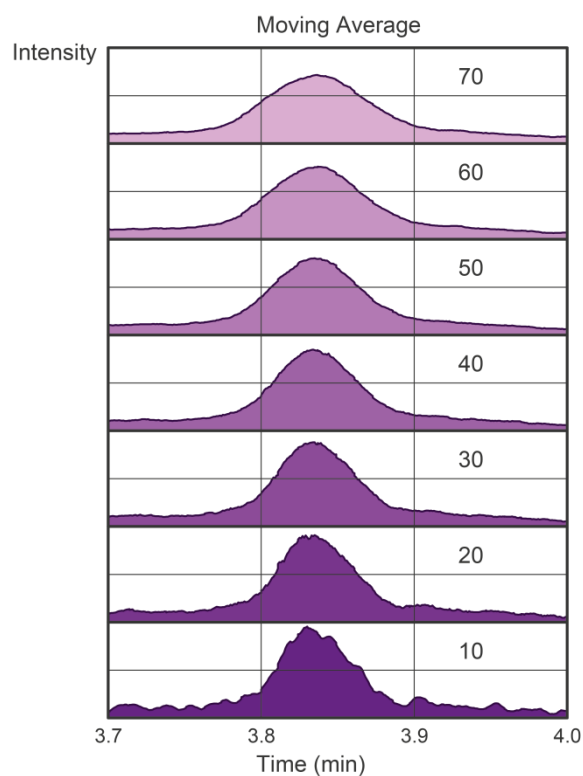


Figure 8-5. Moving Average Smoothing Using Different Smoothing Window Width

In order to demonstrate how the area and the height of the peak are affected I have used two different software packages OriginPro 8.5 and MassLynx. The results of the peak height comparison are presented in Figure 8-6 and Figure 8-7, and the results of the peak height comparison are given in Figure 8-8 and Figure 8-9.

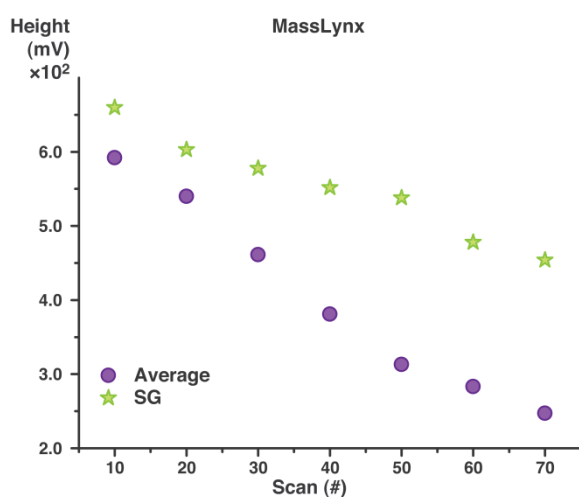


Figure 8-6. MassLynx Peak Height Result Using Different Smoothing Window and Different Algorithms

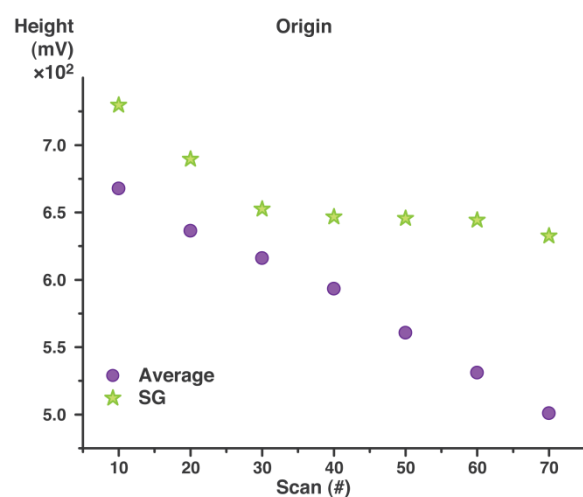


Figure 8-7. OriginPro 8.5 Peak Height Results Using Different Smoothing Window and Different Algorithms

Even though the increase of the smoothing window results in the decrease of height in case of both algorithms it is worth pointing out that Savitzky-Golay smoothing affects these parameters to a lesser extent. For the area there is a downward trend in case of the moving average being applied and almost no effect can be seen in case of the Savitzky-Golay. It is worth pointing out, that the variation of the area caused by the MassLynx package is about an order of magnitude greater than that caused by the OriginPro software. Apparently the reason for this is the different algorithms applied by the two software packages for identification of the beginning and ending of the mass chromatographic peak as well as different way of constructing the baseline of the peak. Whereas MassLynx draws a straight line between the starting and ending points of the peak the OriginPro software calculates the mathematical area of the curve, *i.e.* the area limited by the curve on one side and the x-axis on the other side.

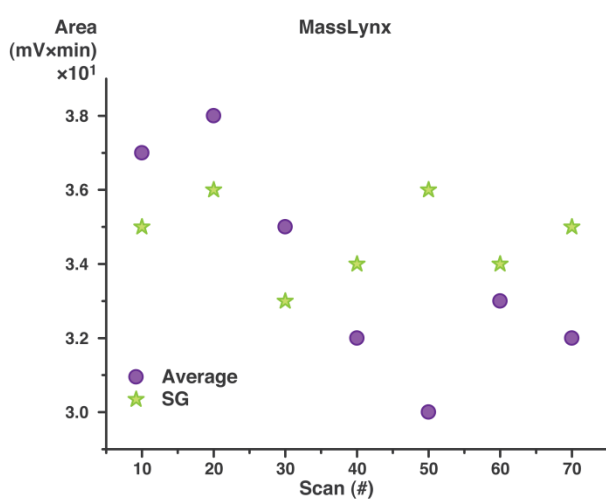


Figure 8-8. MassLynx Peak Integration Result Using Different Smoothing Window and Different Algorithms

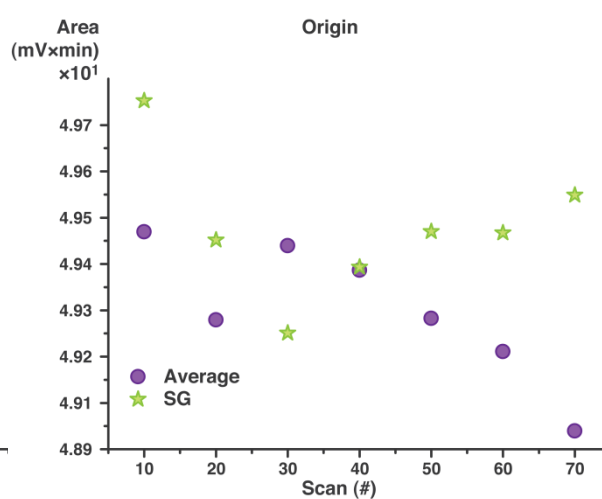


Figure 8-9. MassLynx Peak Integration Result Using Different Smoothing Window and Different Algorithms

Another important point concerns the data acquisition. It is generally recommended to have as many data points per peak as possible. This can be achieved by including fewer masses in the MS acquisition function operated in the SIM mode. The results are presented in Figure 8-10 and Figure 8-11. However, when the time scale is used, the increase in the number of scans does not affect the peak area, whereas when the area is calculated per scan,

proportionally greater response is achieved for greater number of scans. Unluckily, the MassLynx software, which is used for the batch processing cannot utilise this and operates only on the time scale.

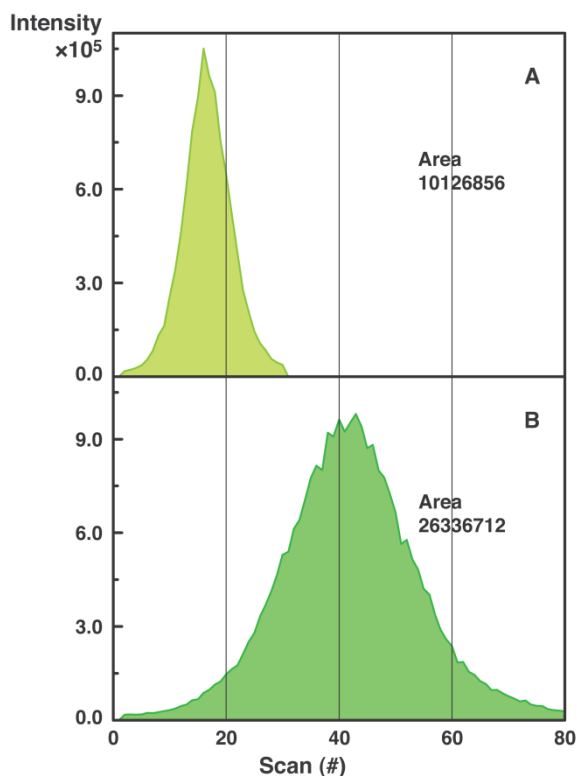


Figure 8-10. Phenylalanine Peak Presented on the Scan Scale

- A. 17 masses are monitored in parallel
- B. 1 mass is monitored

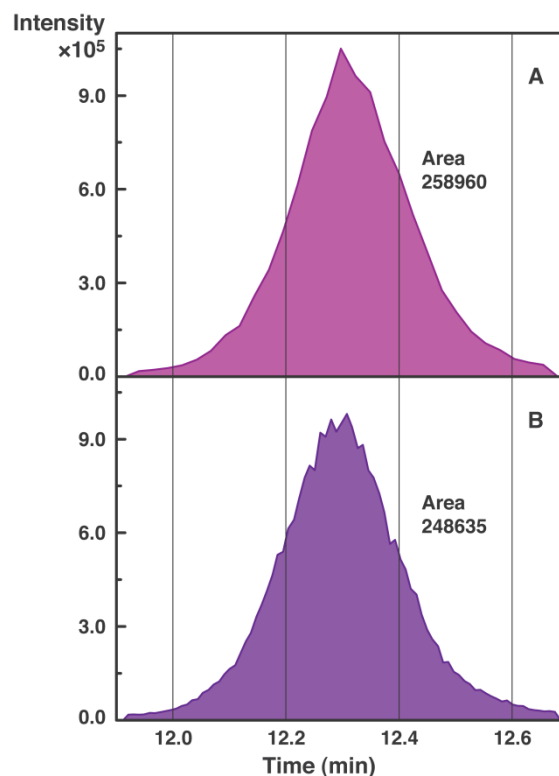


Figure 8-11. Phenylalanine Peak Presented on the Time Scale

- A. 17 masses are monitored in parallel
- B. 1 mass is monitored

Using the scan scale x-axis, we can show that neither of the smoothing methods is devoid of drawbacks. The results of two smoothing algorithms comparison are provided in Figure 8-12. It can be seen that both algorithms are create artefacts, which result in the increase of area in case of the Savitzky-Golay smoothing and in decrease of area when moving average is applied. The line parallel to the x-axes corresponds to the peak area, when no smoothing is applied.

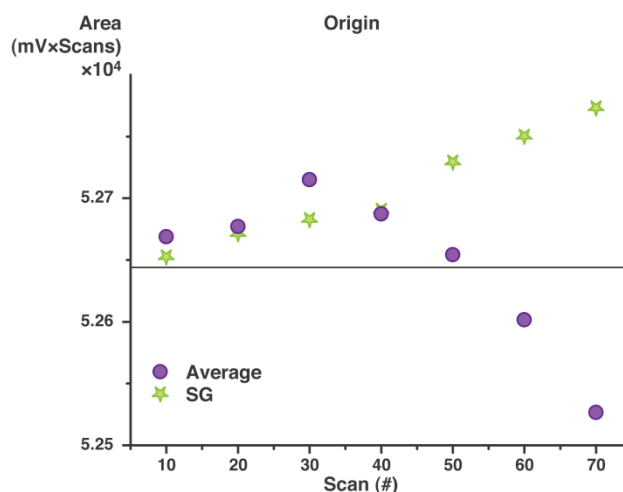


Figure 8-12. Comparison of the Smoothing Algorithms Using the Scan Scale

Absence of the software which allows the processing untreated mass chromatograms requires us to use the conventional ways of data processing, which enables batch processing of large data sets. Based on the discussion presented above I used Savitzky-Golay smoothing prior to the quantification. The window of smoothing will be set as half of the width peak – it was shown before, that this enables adequate smoothing and causes minimum data distortion.

Even though it is not conventional to present the mass chromatograms as bar charts it is necessary to understand the nature of the data behind the graph. The difference between UV chromatograms and mass chromatograms should be taken into consideration when processing and presenting the data. In case of mass chromatograms the zero of the intensity scale is absolute and should always be shown.

The absence of the software which allows ready batch processing of the raw mass chromatograms results in the necessity of data processing prior to its integration. The optimal processing algorithm was suggested by Savitzky and Goley and one iteration of the algorithm with window set at the number of data points corresponding to the half width of the peak⁴.

8.6 Histone Extraction and Purification

8.6.1 Cell Growth

In the cell based experiments HEK 293T cell line was used. The cells were grown in media containing labelled ^{13}C fully labelled arginine, lysine or proline (*i.e.* ^{13}C at all carbon atoms). DMEM Media for SILAC, purchased from Thermo Fisher Scientific (Northumberland, UK), was used to grow the cells. 500 mL of DMEM media 50 mL of Dialyzed Fetal Bovine Serum also purchased from Thermo Fisher Scientific (Northumberland, UK) was added to the DMEM media. The DMEM media for SILAC does not contain Arg, Lys or Pro. Therefore, those AAs were added to the media separately (Table 8-12). All three AAs were added to the media, so that every media has all the AAs and each media had one labelled AA and two non-labelled AA. The quantities of the AAs added to the media are listed in Table 8-12. For cell growth, 10 cm and 15 cm diameter plates were used. For every passage, cells were grown until they covered about 90% of the plate surface. The density of the cells was confirmed using a microscope. When the cells were grown to 90% density, roughly a $\frac{1}{4}$ of all the cells on the plate were transferred to a new plate. In order to maximise the percentage of labelling the cells were re-passaged 5 times, thus theoretically conferring 99.9% of isotope labelling to the cells. A final re-passage was carried out using a 15 cm diameter plate, after this the cells were harvested.

Table 8-12. AAs labelling of the media

	Name	MW (g/mol)	Supplier	g per 500 mL	mM
Labelled					
Arg	<i>L</i> -Arginine- $^{13}\text{C}_6$, $^{15}\text{N}_4$ hydrochloride	220.59	Aldrich	0.04390	0.398
Lys	<i>L</i> -Lysine- $^{13}\text{C}_6$, $^{15}\text{N}_2$ hydrochloride	190.65	Aldrich	0.07607	0.798
Pro	<i>L</i> -Proline- $^{13}\text{C}_5$	120.13	Aldrich	0.1	1.665
Non-Labelled					
Arg	<i>L</i> -Arginine hydrochloride	210.70	Sigma	0.041929	0.398
Lys	<i>L</i> -Lysine monohydrochloride	182.65	Fluka	0.072877	0.798
Pro	<i>L</i> -Proline	115.13	Sigma	0.1	1.737

8.6.2 Acid Extraction of Histones

For the histone extraction, cells were washed with PBS 3X buffer and scraped off the Petri dish into a falcon tube and spun at 1500 rpm for 5 min. The supernatant was discarded. Cell pellets were washed with 10 mL PBS and spun again at 1500 rpm for 5 min. The supernatant was discarded again. The cell pellets were resuspended in 0.5 mL ice-cold lysis buffer and transferred into a 1.5 mL micro-centrifuge tube kept on ice and spun at 3000 rpm (5 min, 4°C). The nuclear pellet was resuspended in 1 mL of the same lysis buffer and spun at 3000 rpm (5 min, 4°C). For nuclear extraction, 150 µL of 0.4 N HCl was added and left on ice for 1 hour. In order to remove nuclear debris the lysate was spun at 13000 rpm (20 min, 4°C). 75 µL of supernatant was transferred to fresh tube. 750 µL of ice cold acetone was added and left at -20°C overnight. The histones were spun at 13000 rpm (20 min, 4°C) and supernatant was removed. The pellet was dispersed in 1 mL of ice-cold acetone. The mixture was centrifuged again at 13000 rpm (20 min, 4°C) and acetone was discarded. Another 1 mL of acetone was added and spun again at 13000 rpm for 20 min at 4°C. the acetone was discarded. The histones were dried out of acetone using rotary pump and resuspended in 200 µL of MilliQ water.

Nuclei Buffer (1x):

- Triton X-100 0.2%
- HEPES, pH 7.6 10 mM
- MgCl₂ 1.5 mM
- KCl 1 mM

Lysis buffer:

- Nuclei buffer (10x stock) 1x
- Sodium butyrate (1 M stock) 1x (10 mM)
- Protease inhibitor w/o EDTA (Roche) 1x
- Phosphatase inhibitor (Roche) 1x

8.6.3 Total Protein Content Determination (Bicinchoninic Acid Assay)

- Add BSA standard for appropriate dilution range (2 mg/ml stock)
- Add mQ water to make up to 20 ul
- Assay cell lysate in triplicate/duplicate against a standard curve (eg. 10 ul lysate +10 ul mQ water = ½ dilution):
 - A few dilutions for each sample eg. 1/2, 1/5, 1/10
- Mix 0.4 ml BCA solution A (CuSO₄ solution) with 20 ml BCA solution B (Sigma kit)
- Add 200 ul of the mix to each well and incubate at 37°C 30 min
- Read A₅₆₂ (562 nm optimum but 540-590 nm ok)

8.6.4 Histone Separation Using Gel Electrophoresis

In order to separate histones, sodium dodecyl sulphate (SDS) polyacrylamide gel electrophoresis (PAGE) was used. PAGE allows separation of SDS modified proteins according to their size or molecular weight. The sample contained 10 µL of histone mixture extracted from cells plus 10×SDS gel loading buffer. The protein samples were heated at 95°C for 3-5 min, and loaded onto a gel, which was electrophoresed at 200 V for 1 hour. The gels were then stained with Coomassie Blue for 15 min and rinsed in the mixture of 10% methanol and 10% acetic acid.

Table 8-13. 10 x SDS-PAGE sample loading buffer

Reagent	Amount
Tris.HCl (0.5M pH 6.8)	10ml
Bromophenol blue (0.2% w/v)	0.2g
SDS (10% W/V)	20ml
Glycerol	12ml
β- Mercaptoethanol	5.0ml

Table 8-14. 10 x SDS-PAGE running buffer

Reagent	Amount (in 1 litre)
Tris.HCl	30g
Glycine	144g
SDS	10g

Table 8-15. Resolving gel: (for 1 gel)

Reagent	Amount (18%)
Tris.HCl (1.5M pH 8.8)	1.25ml
MilliQ water	0.65 mL
30% acrylamide (w/v)	3 mL
SDS (10%w/v)	50µl
10% APS (w/v)	50 µl
TEMED	4 µl

Table 8-16. Stacking gel: (for 1 gel)

Reagent	Amount (~3.0%)
Tris.HCl (0.5M pH 6.6)	0.63ml
MilliQ water	1.57ml
30% acrylamide (w/v)	0.25ml
SDS (10%w/v)	25µl
10% APS (w/v)	25 µl
TEMED	2.5 µl

8.6.5 In Gel Histone Digestion

Day one

Histone bands were cut from the gel using clean scalpel. Individual bands were cut into approximately 1 mm² pieces and placed in 1.5 mL tube. 200 µL of wash solution was added and kept overnight while shaking.

Day two

The wash solution was replaced with another 200 µL wash solution. The gel pieces were washed for 3 hours on rocking table. Tubes with gel pieces were centrifuged briefly and the wash solution was removed. In order to dehydrate the gel 200 µL of acetonitrile was added and left for 5 min on rocking table. After removing acetonitrile the procedure was repeated again. The gel pieces were dried in vacuum centrifuge for 3 min. 30 µL of 10 mM DTT buffer were added; the suspension was left on a rocking table for 30 min. The solution was removed. 30 µL of 50 mM iodoacetamide buffer and alkylate were added and the suspension was left on the rocking table for 30 min. the solution was removed. In order to dehydrate the

gel pieces another 200 μ L of acetonitrile was added for 5 min; afterwards acetonitrile was removed. Gel pieces were dried in a vacuum centrifuge for 3 min. Gel pieces were rehydrated in 200 μ L of 100 mM ammonium bicarbonate for 10 min on rocking table. Afterwards the rehydration solution was removed. Gel pieces were dehydrated again using 200 μ L of acetonitrile for 5 min. After acetonitrile had been removed the gel pieces were dried in a vacuum centrifuge for 3 min. 30 μ L of 20 ng/ μ L trypsin solution was added to the gel pieces and rehydrated on ice for 10 min exactly with occasional gentle mixing. Gel pieces were centrifuged and the excess of trypsin solution was removed. 5 μ L of 50 mM ammonium bicarbonate was added to the sample and the sample was incubated overnight at 37°C.

Day three

50 μ L of 50 mM ammonium bicarbonate buffer was added to the tube containing the gel pieces and incubated for 10 min with occasional gentle vortex mixing. supernatant was collected in a 1.5 ml tube. 50 μ L of the Extraction Buffer I was added to the tube containing the gel pieces and incubate for 10 minutes with occasional gentle vortex mixing. Supernatant was added to the collection tube. Another 50 μ L of the Extraction Buffer II was added to the tube containing the gel pieces and incubate for 10 minutes with occasional gentle vortex mixing. Supernatant was added to the collection tube. The sample was transferred to the hydrolysis vial and dried completely in a vacuum centrifuge.

All reagents were prepared freshly and sterile filtered using 0.2 μ m filter.

1. Wash solution

50 % methanol, 5 % acetic acid in MilliQ water

2. 100 mM ammonium bicarbonate (79 g/mol):

0.79 g NH_4HCO_3 was dissolved in 100 mL MilliQ water

3. 50 mM ammonium bicarbonate

100 mM stock was diluted 1:1 in MilliQ water

4. 1 M DTT (154 g/mol)

1.54 g dithiothreitol was dissolved in 10 mL MilliQ water

5. 10 mM DTT

10 μ L 1 M DTT was added to 1.0 mL of 100 mM NH_4HCO_3

6. 50 mM iodoacetamide (185 g/mol)

0.092 g iodoacetamide was dissolved in 1 mL MilliQ water

7. Extraction buffer I

- 50 % acetonitrile
- 5 % formic acid
- in MilliQ-H₂O

Extraction buffer II

- 85 % acetonitrile
- 5 % formic acid
- in MilliQ-H₂O

Buffer A

- 98 % MilliQ-H₂O
- 2% acetonitrile
- 0.1% formic acid

8.7 Peptide and Protein Digestion

8.7.1 Acid Hydrolysis

The peptide solution was placed in a hydrolysis tube and dried under vacuum using a rotary pump. Hydrolysis tubes with dried samples were placed in hydrolysis vials. Then 150 μ L of 6 M HCl was added to the hydrolysis vial. The hydrolysis vials were closed tightly and pumped out and flushed with nitrogen using the vacuum tight switch cap. The procedure was repeated three times. The vial is pumped out again and left under vacuum. Samples prepared in this way were then incubated in the oven for 24 hours at 110°C.

Hydrolysis tubes were purchased from Waters Ltd (Hertfordshire, UK). Constant boiling HCl for hydrolysis was purchased from Thermo Fisher Scientific (Northumberland, UK).

8.7.2 Enzymatic Digestion

Trypsin Digestion

Trypsin preferentially cleaves the peptide bond after the Lys and Arg residues.

5 μ L of Trypsin 100 μ g/mL solution in 15 mM NH_4OAc were added to the peptide mixture and incubate for 2 hours at 37°C. In order to confirm the efficiency of the tryptic digest 1 μ L of the digest mixture was analysed using MALDI mass spectrometry. In order to remove the trypsin the digest mixture was spun down two times for 10 min at 14 800 rpm. Supernatant was collected and incubated with carboxypeptidase.

Carboxypeptidase Digestion

1 μ L of carboxypeptidase added to the supernatant and incubated overnight at 37°C. the digest was filtered using 3 kDa molecular weight cut off filter at 14800 rpm for 10 min.

- (1) Armstrong, M.; Jonscher, K.; Reisdorph, N. A. *Rapid Communications in Mass Spectrometry* **2007**, 21, 2717.
- (2) McCullagh, J. S. O.; Juchelka, D.; Hedges, R. E. M. *Rapid Communications in Mass Spectrometry* **2006**, 20, 2761.
- (3) Savitzky, A.; Golay, M. J. E. *Analytical Chemistry* **1964**, 36, 1627.
- (4) Otto, M. *Chemometrics*, 2007.

9 Appendices

9.1 Appendix A – Correction comparison

9.1.1 The Craig Correction

Certain disambiguation is required, when addressing the correction procedure proposed by Harmon Craig¹. The major problem for a contemporary reader is how this author defines the isotope ratios. Definition of the R is straightforward for the R of a particular element element:

$$^{13}R = \frac{^{13}C}{^{12}C}$$

When dealing with the diatomic molecule of oxygen, the R is more complex:

$$^{17}R = \frac{^{17}O^{16}O}{^{16}O^{16}O}$$

$$^{18}R = \frac{^{18}O^{16}O}{^{16}O^{16}O}$$

And even more complex, when it comes to defining the isotopic ratio of carbon dioxide.

$$^{45}R = \frac{v^{45}CO_2}{v^{44}CO_2} = \frac{^{13}C^{16}O^{16}O + ^{12}C^{16}O^{17}O}{^{12}C^{16}O^{16}O}$$

$$^{46}R = \frac{v^{46}CO_2}{v^{44}CO_2 + v^{45}CO_2} = \frac{^{12}C^{16}O^{18}O + ^{13}C^{16}O^{17}O + ^{12}C^{17}O^{17}O}{^{12}C^{16}O^{16}O + ^{13}C^{16}O^{16}O + ^{12}C^{16}O^{17}O}$$

“¹³δ_{measured} – measured enrichment, calculated from the molecular masses collected”. With nowadays nomenclature in mind, it should be termed as ⁴⁵δ. Whereas, “¹³δ_{corrected} – the true enrichment” – should be understood as ¹³δ.

According to Craig, the carbon dioxide 45 isotope ratio comprises carbon 13 ratio and oxygen 17 ratio:

$$^{45}\text{R} = ^{13}\text{R} + ^{17}\text{R} \quad (1)$$

Another assumption by Craig is that:

$$\frac{^{17}\text{R}_S}{^{17}\text{R}_{St}} = \left(\frac{^{18}\text{R}_S}{^{18}\text{R}_{St}} \right)^{0.5} = (1 + ^{18}\delta)^{0.5} \cong 1 + \frac{^{18}\delta}{2} \quad (2)$$

A further $^{13}\delta$ correction is based largely on this assumption. Below his formulae for the correction is derived.

According to the definition of the delta value:

$$^{45}\delta = \frac{^{45}\text{R}_S}{^{45}\text{R}_{St}} - 1 \quad (3)$$

$$^{13}\delta = \frac{^{13}\text{R}_S}{^{13}\text{R}_{St}} - 1 \quad (4)$$

which alternatively could be rewritten as $^{13}\delta = \frac{^{13}\text{R}_S - ^{13}\text{R}_{St}}{^{13}\text{R}_{St}}$, if $^{13}\text{R}_S$ is substituted then in accordance with equation (1), we get $^{13}\delta = \frac{^{45}\text{R}_S - ^{17}\text{R}_S - ^{13}\text{R}_{St}}{^{13}\text{R}_{St}}$. Then $^{45}\text{R}_S$ can be substituted into the equation (2) to produce $^{13}\delta = \frac{^{45}\text{R}_{St} + ^{45}\delta^{45}\text{R}_{St} - ^{13}\text{R}_{St} - ^{17}\text{R}_S}{^{13}\text{R}_{St}}$; and finally, $^{45}\text{R}_{St}$ can be substituted according to the equation (1), which resulting in:

$$^{13}\delta = \frac{^{45}\delta^{45}\text{R}_{St} + ^{17}\text{R}_{St} - ^{17}\text{R}_S}{^{13}\text{R}_{St}}$$

The only unknown part of which is $^{17}\text{R}_S$ – which can be calculated from the second assumption, thus giving:

$$^{13}\delta = \frac{^{45}\delta \cdot ^{45}\text{R}_{St} + ^{17}\text{R}_{St} - ^{17}\text{R}_S}{^{13}\text{R}_{St}} = ^{45}\delta \cdot \left(\frac{^{45}\text{R}_{St}}{^{13}\text{R}_{St}} \right) - ^{18}\delta \cdot \left(\frac{^{17}\text{R}_{St}}{2 \cdot ^{13}\text{R}_{St}} \right)$$

The isotope ratios of the standard Craig adopts from the Nier's publication² after thorough revision.

$$^{13}\delta = 1.0676 \cdot ^{45}\delta - 0.0338 \cdot ^{18}\delta$$

The only remaining problem is the $^{18}\delta$ which Craig derives from the 46 carbon dioxide isotope ratio, which appears in his paper as follows:

$$^{46}R = \frac{v^{46}CO_2}{v^{44}CO_2 + v^{45}CO_2} = \frac{^{12}C^{16}O^{18}O + ^{13}C^{16}O^{17}O + ^{12}C^{17}O^{17}O}{^{12}C^{16}O^{16}O + ^{13}C^{16}O^{16}O + ^{12}C^{16}O^{17}O}$$

Then Craig makes the assumption that the last term in the numerator – $^{12}C^{17}O^{17}O$ – can be omitted because it is negligibly small in comparison to the abundance of other species, thus the starting point for the $^{18}\delta$ estimation is

$$^{46}R = \frac{v^{46}CO_2}{v^{44}CO_2 + v^{45}CO_2} = \frac{^{12}C^{16}O^{18}O + ^{13}C^{16}O^{17}O}{^{12}C^{16}O^{16}O + ^{13}C^{16}O^{16}O + ^{12}C^{16}O^{17}O}$$

$$1 + ^{46}\delta = \frac{\left\{ 1 + ^{18}\delta + \frac{(1 + ^{13}\delta) \cdot ^{13}R_{St} \cdot ^{17}R_{St} \cdot \left(1 + \frac{^{18}\delta}{2}\right)}{^{18}R_{St}} \right\} \cdot \{1 + ^{13}R_{St} + ^{17}R_{St}\}}{\left\{ 1 + \frac{^{13}R_{St} \cdot ^{17}R_{St}}{^{18}R_{St}} \right\} \cdot \left\{ 1 + (1 + ^{13}\delta) \cdot ^{13}R_{St} + ^{17}R_{St} \cdot \left(1 + \frac{^{18}\delta}{2}\right) \right\}}$$

When the equations 1, 2, and 4 are applied to the equation above, it can be transformed into

$$1 + ^{46}\delta = \frac{\left\{ 1 + ^{18}\delta + \frac{(1 + ^{13}\delta) \cdot ^{13}R_{St} \cdot ^{17}R_{St} \cdot \left(1 + \frac{^{18}\delta}{2}\right)}{^{18}R_{St}} \right\} \cdot \{1 + ^{13}R_{St} + ^{17}R_{St}\}}{^{46}R_{St}}$$

9.1.2 The Santrock, Studley, and Hayes correction

The major improvement to the correction algorithm proposed by the three authors, Santrock, Studley, and Hayes (often termed as the SSH correction)³, when compared to the procedure suggested by Craig, was their suggestion of clearly defined “isotopic abundances” – isotopic ratios (^nR) and fractional abundances (^nF) – which are defined by the authors as follows:

$$^nR = \frac{\text{isotopic species of mass } n}{\text{base species related to } n}$$

$${}^nF = \frac{\text{isotopic species of mass } n}{\text{sum of all species related to } n}$$

It can be easily derived, therefore, that:

$${}^{44}F = {}^{12}F \cdot {}^{16}F \cdot {}^{16}F$$

$${}^{45}F = {}^{13}F \cdot {}^{16}F \cdot {}^{16}F + 2 \cdot {}^{12}F \cdot {}^{16}F \cdot {}^{17}F$$

$${}^{46}F = {}^{12}F \cdot {}^{17}F \cdot {}^{17}F + 2 \cdot {}^{12}F \cdot {}^{16}F \cdot {}^{18}F + 2 \cdot {}^{13}F \cdot {}^{16}F \cdot {}^{17}F$$

$${}^{47}F = {}^{13}F \cdot {}^{17}F \cdot {}^{17}F + 2 \cdot {}^{13}F \cdot {}^{16}F \cdot {}^{18}F + 2 \cdot {}^{12}F \cdot {}^{17}F \cdot {}^{18}F$$

$${}^{48}F = {}^{12}F \cdot {}^{18}F \cdot {}^{18}F + 2 \cdot {}^{13}F \cdot {}^{17}F \cdot {}^{18}F$$

$${}^{49}F = {}^{13}F \cdot {}^{18}F \cdot {}^{18}F$$

And for the isotope ratios:

$${}^{45}R = {}^{13}R + 2 \cdot {}^{17}R$$

$${}^{46}R = 2 \cdot {}^{18}R + 2 \cdot {}^{13}R \cdot {}^{17}R + 2 \cdot {}^{17}R^2$$

$${}^{47}R = {}^{13}R \cdot {}^{17}R^2 + 2 \cdot {}^{13}R \cdot {}^{18}R + 2 \cdot {}^{17}R \cdot {}^{18}R$$

$${}^{48}R = {}^{18}R^2 + 2 \cdot {}^{13}R \cdot {}^{17}R \cdot {}^{18}R$$

$${}^{49}R = {}^{13}R \cdot {}^{18}R^2$$

The isotope ratio for the species of mass 47, 48 and 49 could not be measured and therefore the estimation of ${}^{17}\text{O}$ contribution to the 45 mass can only be carried out from the first two equations:

$${}^{45}R = {}^{13}R + 2 \cdot {}^{17}R$$

$${}^{46}R = 2 \cdot {}^{18}R + 2 \cdot {}^{13}R \cdot {}^{17}R + 2 \cdot {}^{17}R^2$$

We thus have 2 equations with 3 unknown parameters – ^{13}R , ^{17}R and ^{18}R . In order to resolve this system a third equation has to be introduced. This can be done through considering the correlation between ^{17}O and ^{18}O , which has previously been introduced by Harmon Craig. Santrock, Studley and Hayes, state that the power (a) should be between 0.50 and 0.53.

$$\frac{^{17}R_S}{^{17}R_{St}} = \left(\frac{^{18}R_S}{^{18}R_{St}} \right)^a$$

This equation can be rewritten as:

$$^{17}R = ^{18}R^a \cdot K$$

where $K = \frac{^{17}R_{St}}{^{18}R_{St}^a}$.

9.1.3 The Allison correction

Allison uses the definitions of “isotopic abundances” introduced by Santrock, Studley, and Hayes, but accepts the parameter as equal to 0.5 in the equation uniting the ^{17}O and ^{18}O of sample and standard⁴. Allison defines three equations and three unknown parameters.

$$^{45}R = ^{13}R + 2 \cdot ^{17}R$$

$$^{46}R = 2 \cdot ^{18}R + 2 \cdot ^{13}R \cdot ^{17}R + 2 \cdot ^{17}R^2$$

$$\frac{^{17}R_S}{^{17}R_{St}} = \left(\frac{^{18}R_S}{^{18}R_{St}} \right)^{0.5}$$

By substituting ^{18}R from $\frac{^{17}R_S}{^{17}R_{St}} = \left(\frac{^{18}R_S}{^{18}R_{St}} \right)^{0.5}$ and ^{13}R from $^{45}R = ^{13}R + 2 \cdot ^{17}R$ a quadratic equation in ^{17}R can be formed:

$$\left(2 \cdot \frac{^{18}R_{St}}{^{17}R_{St}^2} - 3 \right) \cdot ^{17}R_S^2 + 2 \cdot ^{45}R_S \cdot ^{17}R_S - ^{46}R_S = 0$$

This equation can be solved with respect to $^{17}R_S$ as follows:

$$^{17}R_S = \frac{-\beta + \sqrt{\beta^2 - 4\alpha\gamma}}{2\alpha}$$

where

$$\alpha = 2 \cdot \frac{^{18}R_{St}}{^{17}R_{St}^2} - 3$$

$$\beta = 2 \cdot ^{45}R_S$$

$$\gamma = -^{46}R_S$$

After that ^{13}R can be calculated from the equation

$$^{13}R = ^{45}R - 2 \cdot ^{17}R$$

And ^{18}R from the equation

$$^{18}R = \frac{^{46}R - 2 \cdot ^{13}R \cdot ^{17}R - 2 \cdot ^{17}R^2}{2}$$

Interestingly, there are differences between the way data is presented in this work and the way the data is collected by the NIST⁵. In my work the reference parameters are taken from the original papers as a primary source and only calculate other parameters where it is necessary (the calculated parameters are given in italic), using the algorithm provided by the author. NIST apparently provides the set of algorithms and set of the parameters, which could be used interchangeably, where as algorithm and parameter are defined by the authors and using different parameters for the same algorithm makes the data incomparable. NIST also provides opportunity to use Assonov correction, which is similar in principle to the corrections described above, but the author uses different values for the carbon and oxygen isotopic abundances based on the different sources^{6,7}.

Table 9-1. Parametres applied in different isotope ratio corrections

Authors	Craig	Santrock Studley Hayes	Allison Francey Meijer
Year	1957	1985	1995
C Standard	PDB	PDB	VPDB CO ₂
O Standard	PDB CO ₂	SMOW	
¹³ R	0.0112372		0.01123720
¹⁷ R	0.00075990	0.00037990	0.00038080
¹⁸ R	0.00415800	0.00200520	0.00208835
⁴⁵ R	0.01199710		0.01199881
⁴⁶ R	0.00411715		0.00418540
a	0.5	0.516	0.50000000
K		0.0099235	

9.2 Appendix B – Retention Data

9.2.1 Reversed Phase Mixed Mode Chromatography Retention Data

This data was produced using a PrimeSep 200 column.

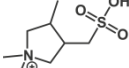
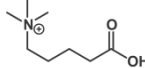
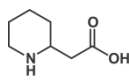
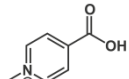
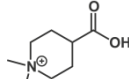
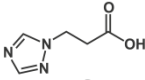
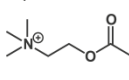
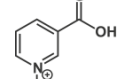
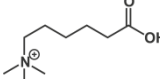
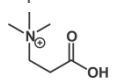
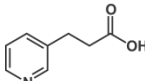
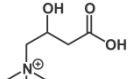
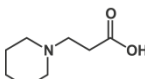
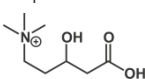
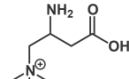
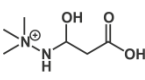
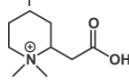
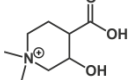
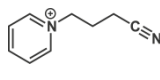
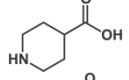
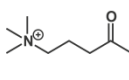
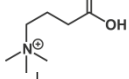
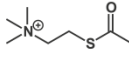
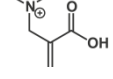
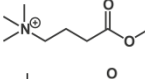
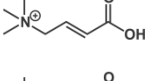
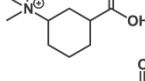
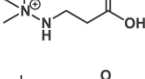
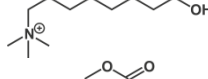
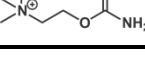
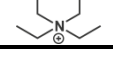
Table 9-2. Reference AAs retention data

Name	tR	RI
Gly	4.77	200
Ala	5.24	300
α-ABu	5.84	400
NLe	10.62	600
NVa	7.57	500
α-AHe	15.47	700
α-AOc	21.11	800

Table 9-3. Standard AAs retention data

Name	tR	RI	Name	tR	RI
PhSer	2.79	0	Ile	9.89	579
3-trans HyP	3.23	59	Leu	10.09	585
4-trans HyP	3.54	92	Tyr	10.69	601
Asp	3.68	107	3AIBu	10.72	602
4-cys HyP	4.07	143	Carnitine	12.13	635
Ser	4.18	153	Phe	12.14	635
3-cys HyP	4.27	160	α,γ -diABu	13.91	671
diMeGly	4.37	168	HyTrp	14.21	678
MeGly	4.43	173	3-MeHis	15.75	705
Thr	4.50	179	His	15.77	706
Glu	4.68	193	Orn	16.40	718
HSer	4.70	195	HyL	16.91	728
triMeGly	4.83	248	1-MeHis	17.41	738
Pro	5.02	253	HyArg	17.56	741
AAd	5.90	402	Trp	18.37	755
2AIBu	6.18	419	Lys	18.42	756
Val	7.13	476	MeLys	19.95	783
Pip	7.34	488	Arg	20.39	790
Met	7.39	491	MeArg	21.59	808
3Ala	9.45	564	diMeLys	21.96	812
diHyPhe	9.46	565	triMeLys	23.66	831

Table 9-4. Retention data for the potential hBBOX substrates

Structure	t _R	RI	Structure	t _R	RI
	4.35	174		13.56	680
	4.83	245		14.04	673
	5.27	332		14.26	676
	5.96	426		14.46	699
	6.33	439		14.65	706
	10.61	613		14.81	692
	10.91	619		15.25	694
	11.68	634		15.46	697
	11.82	625		15.53	699
	11.99	631		17.13	729
	11.86	629		17.30	732
	12.18	651		17.38	753
	12.27	647		17.55	735
	12.58	644		17.79	740
	13.41	660		18.08	774
	13.56	679		24.17	828

9.2.2 Reversed Phase Chromatography of AQC Derivatised Amino Acids Retention Data

Table 9-5. Reference AAs retention data

Name	tR	RI
Gly	2.95	200
Ala	4.75	300
α -ABu	6.89	400
NLe	13.26	600
NVa	9.83	500
α -AHe	17.09	700
α -AOc	21.11	800

Table 9-6. Standard AAs retention data

Name	tR	RI	Name	tR	RI
2S,3S-HyHis	0.83	54	AAd	5.40	335
2S,3R-HyHis	1.19	81	Cys	5.48	338
2S,3R-HyArg	1.90	134	3AIBu	5.75	351
2S,3S-HyArg	1.97	140	Pro	5.77	352
4-trans HyP	2.07	138	2AIBu	6.52	385
His	2.22	150	2S,5S-HyL	6.53	384
3-trans HyP	2.24	151	α,γ -diABu	6.68	392
3-MeHis	2.39	162	2S,5R-HyL	6.81	395
1-MeHis	2.53	173	diHyPhe	7.19	412
MeArg	2.70	184	Orn	7.63	429
Ser	2.73	186	D-AllylGly	7.75	431
Arg	2.79	190	HyTrp	8.11	446
diMeLys	2.89	196	Tyr	8.63	463
HSer	3.03	205	Cyst	8.64	463
Asp	3.22	218	Lys	8.77	468
triMeLys	3.39	229	Met	8.84	470
MeGly	3.57	240	Val	9.37	486
4-cys HyP	3.58	247	MeLys	10.04	507
3-cys HyP	3.66	245	Ile	12.43	578
Glu	3.71	248	Leu	12.75	587
3Ala	3.96	261	Phe	13.22	599
Thr	4.21	275	Trp	13.95	620

9.2.3 Reversed Phase Ion Pairing Chromatography Retention Data

Table 9-7. Reference AAs retention data

Name	tR	RI
Gly	0.51	200
Ala	0.78	300
α -ABu	1.43	400
NLe	4.57	600
NVa	3.32	500
α -AHe	5.34	700
α -AOc	6.09	800

Table 9-8. Standard AAs retention data

Name	tR	RI	Name	tR	RI
3-trans HyP	0.35	103	HyTrp	3.09	493
Asp	0.38	127	Ile	4.26	578
4-trans HyP	0.39	131	Leu	4.45	592
Ser	0.44	161	Phe	4.55	600
4-cys HyP	0.48	186	α,γ -diABu	4.81	619
diMeGly	0.51	205	Cys	4.93	658
MeGly	0.55	218	3-MeHis	4.94	644
Glu	0.58	229	His	4.98	650
HSer	0.59	233	1-MeHis	5.03	661
Thr	0.61	241	Orn	5.13	667
AAd	0.99	339	HyL	5.18	673
Pro	1.02	349	Lys	5.32	693
diHyPhe	1.79	428	Arg	5.41	708
Carnitine	2.26	451	triMeLys	5.43	712
Tyr	2.57	471	diMeLys	5.46	717
Pip	2.65	478	MeArg	5.50	725
Met	2.74	476	sym-diMeArg	5.57	737
Val	2.86	484			

Table 9-9. Gradients and IP reagents comparison

Gradient IP AA	7 min TDFHA		14 min TDFHA		21 min TDFHA		7 min PDFOA		9 min HDFNA	
	t _R	RI	t _R	RI	t _R	RI	t _R	RI	t _R	RI
3-trans HyP	0.35	104	0.33	105	0.38	103	0.34	102	0.26	83
Asp	0.38	127	0.36	129	0.42	127	0.36	115	0.27	96
4-trans HyP	0.39	131	0.36	132	0.43	131	0.38	129	0.29	116
Ser	0.44	162	0.40	161	0.49	161	0.42	155	0.31	143
4-cys HyP	0.48	186	0.44	188	0.54	186	0.50	198	0.35	185
Gly	0.51	200	0.46	200	0.58	200	0.50	200	0.37	200
diMeGly	0.51	201	0.48	209	0.59	205	0.65	261	0.51	306
MeGly	0.55	217	0.50	220	0.63	218	0.60	242	0.42	245
Glu	0.58	230	0.52	229	0.66	229	0.52	209	0.35	185
HSer	0.59	236	0.53	233	0.67	233	0.56	227	0.37	204
Thr	0.61	242	0.55	241	0.70	241	0.57	232	0.38	212
Ala	0.78	300	0.72	300	0.93	300	0.76	300	0.50	300
Aad	0.99	339	0.94	340	1.24	339	0.83	316	0.52	309
Pro	1.02	343	1.03	352	1.35	349	1.11	372	0.77	405
α -ABu	1.43	400	1.43	400	1.93	400	1.29	400	0.75	400
diHyPhe	1.79	426	2.00	428	2.67	428	1.38	409	0.95	434
Carnitine	2.26	454	2.52	447	3.35	451	2.88	500	2.63	578
Tyr	2.57	469	3.26	469	4.27	471	1.91	449	1.33	481
Pip	2.65	473	3.54	476	4.72	478	2.71	493	1.79	523
Met	2.74	477	3.41	473	4.49	476	2.02	456	1.20	467
Val	2.86	482	3.78	482	4.97	484	2.40	477	1.35	483
HyTrp	3.09	491	4.27	492	5.30	493	2.35	475	1.74	519
NVa	3.32	500	4.72	500	5.59	500	2.88	500	1.52	500
Ile	4.26	578	6.43	578	8.08	578	4.83	583	2.87	574
Leu	4.45	592	6.80	592	8.60	592	5.18	594	3.21	587
Phe	4.55	599	7.02	600	8.96	600	5.67	609	2.83	572
NLe	4.57	600	7.03	600	8.95	600	5.38	600	3.60	600
α,γ -diABu	4.81	633	7.30	616	9.11	619	6.50	702	7.95	713
Cys	4.93	649	8.08	659	10.80	658	6.35	689	7.61	707
3-MeHis	4.94	650	7.78	643	10.06	644	6.60	710	8.04	715
His	4.98	655	7.91	650	10.25	650	6.62	711	8.06	715
1-MeHis	5.03	661	8.11	660	10.69	661	6.63	712	8.05	715
Orn	5.13	674	8.24	667	10.70	667	6.66	715	8.03	715
HyL	5.18	681	8.36	673	10.87	673	6.68	716	8.06	715
Lys	5.32	698	8.76	693	11.60	693	6.75	722	8.06	715
α -AHe	5.34	700	8.91	700	11.98	700	6.48	700	7.24	700
Arg	5.41	710	9.04	708	12.13	708	6.79	766	8.06	811
triMeLys	5.43	713	9.10	712	12.28	712	6.80	767	8.05	810
diMeLys	5.46	717	9.17	717	12.40	717	6.82	772	8.05	810
MeArg	5.50	722	9.31	725	12.64	725	6.85	778	8.04	808
sym-diMeArg	5.57	732	9.50	737	12.99	737	6.88	785	8.04	808
α -AOc	6.09	800	10.58	800	14.61	800	6.96	800	7.97	800

Table 9-10. Cone voltage and collision energy of AA standards

AA	Formula	MW	Cone (V)	Precursor Ion	Collision (V)	Daughter
Ala	C ₃ H ₇ NO ₂	89.09318	90.1	14	90.1	11
Arg	C ₆ H ₁₄ N ₄ O ₂	174.201	175.2	17	175.2	32
Asn	C ₄ H ₈ N ₂ O ₃	132.1179	133.2	13	133.1	22
Asp	C ₄ H ₇ NO ₄	133.1027	134.1	16	134.1	10
Cys	C ₃ H ₇ NO ₂ S	121.1582	122.1	31	122.1	17
Glu	C ₅ H ₉ NO ₄	147.1293	148.1	13	148.1	19
Gln	C ₅ H ₁₀ N ₂ O ₃	146.1445	147.1	11	147.1	19
Gly	C ₂ H ₅ NO ₂	75.0666	76	13	76	11
His	C ₆ H ₉ N ₃ O ₂	155.1546	156.2	15	156.2	16
HyL	C ₆ H ₁₄ N ₂ O ₃	162.187	163.2	10	163.2	15
HyP	C ₅ H ₉ NO ₃	131.1299	132.2	16	132.2	13
Ile	C ₆ H ₁₃ NO ₂	131.1729	132.2	13	132.2	11
Leu	C ₆ H ₁₃ NO ₂	131.1729	132.2	13	132.2	11
Lys	C ₆ H ₁₄ N ₂ O ₂	146.1876	147.2	13	147.2	18
Met	C ₅ H ₁₁ NO ₂ S	149.2113	150.2	11	150.2	11
NLe	C ₆ H ₁₃ NO ₆	131.1729	132.2	13	132.2	11
Orn	C ₅ H ₁₂ N ₂ O ₂	132.161	133.2	12	133.2	22
Phe	C ₉ H ₁₁ NO ₂	165.1891	166.2	14	166.2	16
Pro	C ₅ H ₉ NO ₂	115.1305	116.1	17	116.1	17
Ser	C ₃ H ₇ NO ₃	105.0926	106	13	106	13
Thr	C ₄ H ₉ NO ₃	119.1192	120.1	13	120.1	11
triMeLys	C ₉ H ₂₁ N ₂ O ₂ ⁺	189.2747	189.3	16	189.3	13
Trp	C ₁₁ H ₁₂ N ₂ O ₂	204.2252	205.2	12	205.2	14
Tyr	C ₉ H ₁₁ NO ₃	181.1885	182.2	12	182.2	15
Val	C ₅ H ₁₁ NO ₂	117.1463	118.1	13	118.1	11
γ-ABu	C ₄ H ₉ NO ₄	103.1198	104.1	12	104.1	24

9.3 References

- (1) Craig, H. *Geochimica et Cosmochimica Acta* **1957**, 12, 133.
- (2) Nier, A. O. *Physical Review* **1950**, 77, 789.
- (3) Santrock, J.; Studley, S. A.; Hayes, J. M. *Analytical Chemistry* **1985**, 57, 1444.
- (4) Allison C.E., F. R. J., Meijer H.A.J. In *Reference and intercomparison materials for stable isotopes of light elements* Vienna, 1993, p 155.
- (5) NIST
<http://www.csl.nist.gov/div837/837.01/outputs/standards/algorithm/algorithm.html#Calc>.
- (6) Assonov, S. S.; Brenninkmeijer, C. A. M. *Rapid Communications in Mass Spectrometry* **2003**, 17, 1017.
- (7) Assonov, S. S.; Brenninkmeijer, C. A. M. *Rapid Communications in Mass Spectrometry* **2001**, 15, 2426.