Evaluation of Sequential Chemoselective Peptide Ligation and Molecular Dynamics Simulations as Tools for the Total Synthesis of Proteins; an example using Bovine Pancreatic Ribonuclease A.

by

Claire Elizabeth Council

Submitted for the degree of Doctor of Philosophy

Department of Chemistry
Faculty of Engineering and Physical Sciences
University of Surrey

2013

Abstract

Chemoselective ligation between two peptides can be used to produce synthetic peptides and proteins that cannot easily be synthesised as single peptides by solid phase peptide synthesis. Chemoselective reaction between an aldehyde and hydrazine or hydroxylamine, masking the aldehyde as a 1, 2-amino alcohol, has been investigated as a method of sequential ligation that will allow synthesis of individual peptides in high yield and purity, enabling more than two peptides to be ligated.

Protected amino acids were used as precursors for the synthesis of 1, 2-amino alcohol derivatives that could be used in solid phase peptide synthesis for the production of peptides bearing a C-terminal 1, 2-amino alcohol, as a masked aldehyde for use in ligation reactions. Limitations of this method led to alternative investigations, using peptides bearing an N-terminal serine as a masked aldehyde, and a C-terminal hydrazide. Bovine pancreatic ribonuclease A was chosen as an example protein for sequential ligation and using this method, peptides were synthesised in high yield and purity for use in ligation reactions. Trial ligation reactions with short test peptides were performed successfully, however problems were experienced during ligations using peptide fragments of ribunuclease due to involvement of the cysteine side chains. Methods to overcome these unwanted reactions resulted in insoluble peptide fragments.

Computer modelling using molecular dynamics simulations has been used to investigate the effect of replacing native peptide bonds in ribonuclease on the structure of the protein. The method of molecular dynamics simulation was validated through comparison of the structure of a mutant of ribonuclease from experimental NMR data to the structure produced after a molecular dynamics simulation. Results of the modelling simulations suggest that replacement of native peptide bonds with the chemoselective bond formed through reaction of an aldehyde and hydrazine will have only minor implications for the structure of ribonuclease, and therefore should only have a small impact on enzyme activity.

Contents

Chapter 1	**Introduction**	
1.0	Introduction	1
1.1	Peptides and Proteins	3
1.2	Peptide Synthesis	4
1.2.1	Solid Phase Peptide Synthesis	5
1.3	Synthesis of Proteins	8
1.3.1	Fragment Condensation	8
1.3.2	Chemoselective Ligation	9
1.3.2.1	Non-amide Ligation	10
1.3.2.2	Native Chemical Ligation	13
1.3.3	Orthogonal and Sequential Ligation for Joining more than Two Peptide Fragments	19
1.4	Bovine Pancreatic Ribonuclease A	21
1.4.1	Catalytic Mechanism of Ribonuclease	22
1.4.2	Other Important Active Site Residues	24
1.4.3	Anfinsen's Folding Experiments	25
1.5	Molecular Modelling of Proteins	25
1.5.1	Molecular Dynamics Simulations of RNase A	26
1.6	Aim and Objectives	28

Chapter 2	**Synthesis of Amino Acid Derivatives for use in the Preparation of C-terminal Peptide Aldehydes**	
2.0	Introduction	29
2.1	Synthesis of C-terminal Peptide Aldehydes	30
2.1.1	The 1,2-diol as a Masked Aldehyde	31
2.2	History of Diols and Improvements	32
2.2.1	Use of C-terminal Diols as Masked Aldehydes in Sequential Chemoselective Ligation	34
2.3	Amino Alcohols – New Synthetic Route and Method	

	Development	37
2.3.1	1, 3-amino-2-hydroxypropane (Glycine Amino Alcohol)	38
2.3.2	A New Synthetic Route to 1-amino, 2-alcohol Derivatives of Amino Acids	38
2.3.2.1	Epoxide Derivatives of Fmoc Amino Acids	39
2.3.2.2	Amino Alcohol Derivatives Directly from Fmoc-protected Amino Acids	43
2.3.2.3	Stability of Amino Alcohol Derivatives of Fmoc Amino Acids	49
2.4	Protection of the Alcohol during Peptide Synthesis	50
2.5	Synthesis of a Small C-terminal Peptide Aldehyde	52
2.6	Synthesis of Longer C-terminal Peptide Aldehydes	56
2.7	Synthesis of Ribonuclease Peptide Fragments Bearing a C-terminal Amino Alcohol	58
2.8	Conclusion	61

Chapter 3 Chemoselective Ligation

3.0	Introduction	63
3.1	Comparison of Structure of Non-amide Chemoselective Bonds	63
3.2	N-terminal Serine as a Masked Aldehyde	66
3.3	Peptides Containing a C-terminal Hydrazide	68
3.4	Synthesis of Peptides for use in Trial Ligations	69
3.4.1	Test peptide with N-terminal Serine	69
3.4.2	Test peptide with C-terminal Hydrazide	70
3.5	Periodate Oxidation of Peptides bearing N-terminal Serine	72
3.6	Chemoselective ligation between an N-terminal Glyoxal and a C-terminal Hydrazide	74
3.6.1	Use of Aniline as a Nucleophilic Catalyst	75
3.6.1.1	Trial Ligations using Aniline	76
3.6.1.2	Large Scale Ligation	78
3.7	Conclusion	78

Chapter 4 Synthesis of Bovine Pancreatic Ribonuclease A

4.0	Introduction	80

4.1	Ligation Sites	80
4.1.1	Chemoselective Ligation between an N-terminal Hydrazide and C-terminal Aldehyde	81
4.2	Synthesis of Peptide Fragments Bearing C-terminal Amino Alcohols	84
4.3	Synthesis of Peptide Fragments Bearing C-terminal Hydrazides	85
4.4	Chemoselective ligation of RNase A	90
4.4.1	Oxidation of Peptide Fragments of RNase A Containing N-terminal Serine	92
4.4.1.1	Protection of Cysteine During Periodate Oxidation	94
4.6	Peptide Fragments of RNase A with Protected Cysteines	96
4.6.1	Oxidation of Peptide 4 of RNase A with S-tBu Protected Cysteines	97
4.6.2	Chemoselective Ligation with S-tBu Protected Cysteines	98
4.6.3	Acetamydomethyl as Alternative Cysteine Protection	99
4.6.4	Oxidation of Peptide 4 of RNase A with ACM Protected Cysteines	99
4.6.5	Chemoselective Ligation with ACM Protected Cysteines	100
4.7	Conclusion	101

Chapter 5	**Molecular Dynamics Simulations of Bovine Pancreatic Ribonuclease A**	
5.0	Introduction	103
5.1	Theory of Molecular Dynamics Simulations	103
5.1.1	Forcefield	104
5.1.2	Energy Minimisation	106
5.1.3	Molecular Dynamics Calculations	107
5.1.3.1	Initial Velocities	109
5.1.4	Constraints	110
5.1.4.1	Non-bonded Cut-off	111
5.1.4.2	Solvent	111
5.1.4.3	Periodic boundary conditions	112
5.1.4.4	NTP Ensemble	112

5.2	Root Mean Square Deviation as a tool for interpreting structural similarities	112
5.3	Molecular Dynamics Simulations	113
5.3.1	Molecular Dynamics Simulation of Native Bovine Pancreatic Ribonuclease A	114
5.3.2	Replacement of Ligation Sites with Chemoselective Bonds	117
5.3.2.1	Replacement of Ala20 with an Oxime Bond	117
5.3.2.2	Replacement of Ala20, Ala52 and Thr82 with a Hydrazide Chemoselective Bond	118
5.3.2.3	Molecular Dynamics Simulation of RNase A containing Three Hydrazide Bonds	120
5.3.2.4	Replacement of Cys 65 with Ala	123
5.4	Conclusions	125
Chapter 6	**Conclusions and Future Work**	
6.0	Conclusions and Future Work	127
Chapter 7	**Experimental Methods**	
7.0	Introduction	132
7.1	Materials	132
7.2	Experimental Methods	134
7.2.1	Solid Phase Peptide Synthesis	134
7.2.1.1	Fmoc Amino Acid Side Chain Protecting Groups	134
7.2.1.2	Fmoc Deprotection	135
7.2.1.3	Coupling Reaction	136
7.2.1.4	Acetylation	136
7.2.1.5	Cleavage	136
7.2.1.6	Qualitative Ninhydrin Assay	137
7.2.1.6.1	Preparation of Ninhydrin Assay Solutions	137
7.2.1.6.2	Ninhydrin Assay	137
7.2.1.7	Automated Peptide Synthesis	137
7.2.1.8	Calculation of Resin Substitution Level	138
7.2.2	Synthesis of N-nitrosomethyl Urea	139

7.2.3	Generation of Diazomethane	139
7.2.4	Synthesis of Fmoc-Phe-epoxide	139
7.2.4.1	Preparation of Fmoc-Phe-diazomethylketone [1]	140
7.2.4.2	Preparation of Fmoc-Phe-bromomethylketone [2]	140
7.2.4.3	Preparation of Fmoc-Phe-epoxide [3]	141
7.2.5	Synthesis of 1-amino, 2-alcohol Derivatives	141
7.2.5.1	Preparation of Fmoc-Phe-azidomethylketone [4]	142
7.2.5.2	Preparation of Fmoc-Phe-azidomethylalcohol [5]	142
7.2.5.3	Preparation of Fmoc-Phe-amino alcohol [6]	143
7.2.5.4	Synthesis of Amino Alcohol Derivatives of Other Amino Acids	143
7.2.5.4.1	Preparation of Fmoc-Leu-diazomethylketone [7]	144
7.2.5.4.2	Preparation of Fmoc-Leu-bromomethylketone [8]	144
7.2.5.4.3	Preparation of Fmoc-Leu-azidomethylketone [9]	144
7.2.5.4.4	Preparation of Fmoc-Leu-azidomethylalcohol [10]	144
7.2.5.4.5	Preparation of Fmoc-Leu-amino alcohol [11]	144
7.2.5.4.6	Preparation of Fmoc-Ile-diazomethylketone [12]	145
7.2.5.4.7	Preparation of Fmoc-Ile-bromomethylketone [13]	145
7.2.5.4.8	Preparation of Fmoc-Ile-azidomethylketone [14]	145
7.2.5.4.9	Preparation of Fmoc-Ile-azidomethylalcohol [15]	145
7.2.5.4.10	Preparation of Fmoc-Ile-amino alcohol [16]	145
7.2.5.4.11	Preparation of Fmoc-Ala-diazomethylketone [17]	146
7.2.5.4.12	Preparation of Fmoc-Ala-bromomethylketone [18]	146
7.2.5.4.13	Preparation of Fmoc-Ala-azidomethylketone [19]	146
7.2.5.4.14	Preparation of Fmoc-Ala-azidomethylalcohol [20]	146
7.2.5.4.15	Preparation of Fmoc-Ala-amino alcohol [21]	147
7.2.6	Stability of 1, 2 Amino Alcohol derivatives of Fmoc Protected Amino Acids	147
7.2.7	Synthesis of Peptides with a C-terminal 1, 2 Amino Alcohol	147
7.2.7.1	Substitution of 2-Chlorotrityl Resin with 1, 3-amino-2-hydroxypropane	147
7.2.7.2	Synthesis of Peptides on 1, 3-amino-2-hydroxypropane Substituted 2-chlorotrityl Resin	148
7.2.7.3	Substitution of 2-Chlorotrityl Resin with Fmoc-Phe amino	

	alcohol[6]	148
7.2.7.3.1	Synthesis of Z-RLF-amino alcohol	149
7.2.7.3.2	Synthesis of Z-RLF-aldehyde	149
7.2.8	Synthesis of Peptides with a C-terminal Hydrazide	149
7.2.8.1	Hydrazine Substituted 2-chlorotrityl Chloride Resin	150
7.2.8.2	Synthesis of Peptides on Hydrazine Substituted 2-chlorotrityl Resin	150
7.2.8.3	Synthesis of 8Aoa-fQWAVGHL-NH-NH2	150
7.2.8.4	Synthesis of KETAAAKFERQHMDSSTSA-NH-NH2 (Rib 1)	151
7.2.8.5	Synthesis of SSSSNYCNQMMKSRNLTKDRCKPVNTFVHE SL-NH-NH2 (Rib 2)	151
7.2.8.6	Synthesis of SDVQAVCSQKNVACKNGQTNCYQSYSTMSI -NH-NH2 (Rib 3)	151
7.2.8.7	Synthesis of SSSSNYC(S-tBu)NQMMKSRNLTKDRC(S-tBu)KPVNTFVHESL-NH-NH2 (Rib 2 StBu)	152
7.2.8.8	Synthesis of SDVQAVC(S-tBu)SQKNVAC(StBu)KNGQT NC(S-tBu)YQSYSTMSI-NH-NH2 (Rib 3 StBu)	152
7.2.8.9	Synthesis of SSSSNYC(ACM)NQMMKSRNLTKDRC(ACM) KPVNTFVHESL-NH-NH2 (Rib 2 ACM)	153
7.2.8.10	Synthesis of SDVQAVC(ACM)SQKNVAC(ACM)KNGQT NC(ACM)YQSYSTMSI-NH-NH2 (Rib 3 ACM)	153
7.2.9	Synthesis of N-terminally modified peptide fragments for use in ligation	153
7.2.9.1	Synthesis of H2N-HN-DCRETGSSKYPNCAYKTTQANKHIIV ACEGNPYVPVHFDASV (Rib 4)	154
7.2.9.2	Synthesis of SDCRETGSSKYPNCAYKTTQANKHIIVACEGN PYVPVHFDASV (Rib 4)	154
7.2.9.3	Synthesis of SDC(S-tBu)RETGSSKYPNC(StBu)AYKTTQA NKHIIVAC(S-tBu)EGNPYVPVHFDASV (Rib 4 StBu)	154
7.2.9.4	Synthesis of SDC(ACM)RETGSSKYPNC(ACM)AYKTTQAN KHIIVAC(ACM)EGNPYVPVHFDASV (Rib 4 ACM)	155
7.2.9.5	Synthesis of SDCRETGSSK	155
7.2.9.6	Synthesis of SDC(S-tBu)RETGSSK	156

7.2.10	Oxidation of Peptide Fragments Bearing an N-terminal Serine or C-terminal Amino Alcohol	156
7.2.10.1	Polymer Supported Oxidation	156
7.2.10.1.1	Polymer Supported Oxidation of KPVNTFVHESL-amino alcohol	157
7.2.10.1.2	Polymer Supported Oxidation of YQSYSTMSI-amino alcohol	157
7.2.10.2	Solution Phase Oxidation	157
7.2.10.2.1	Oxidation of SLIGRL-NH2	157
7.2.10.2.1	Oxidation of SDCRETGSSK	158
7.2.10.2.2	Oxidation of SDC(S-tBu)RETGSSK	158
7.2.10.2.3	Oxidation of SDCRETGSSKYPNCAYKTTQANKHIIVACE GNPYVPVHFDASV, and S-tBu and ACM protected versions	158
7.2.11	Chemoselective Ligation of Peptides	159
7.2.11.1	Initial Ligations of 8Aoc-fQWAVGHL-NH2-NH2 to glyoxal-LIGRL-NH2	159
7.2.11.1	Optimised Ligation of 8Aoc-fQWAVGHL-NH2-NH2 to glyoxal-LIGRL-NH2	159
7.3	Equipment and Instrumentation	160
7.3.1	Reverse Phase High Performance Liquid Chromatography	160
7.3.1.1	HPLC Analysis	160
7.3.1.2	HPLC Purification	161
7.3.2	Mass Spectrometry	161
7.3.2.1	Electrospray Mass Spectrometry of Peptides	161
7.3.3	Nuclear Magnetic Resonance Spectroscopy	162
7.3.4	Optical Spectroscopy	162
7.4	Molecular Dynamics Simulations	162
7.4.1	Running a Molecular Dynamics Simulation	162
7.4.2	Molecular Dynamics Simulation of Bovine Ribonuclease A	163
7.4.2.1	Replacement of Ala20 with an Oxime Bond	163
7.4.2.2	Replacement of Ala20, Ala52 and Thr82 with a hydrazide chemoselective bond	164
7.4.2.3	Molecular Dynamics Simulation of RNase A containing ligated	

Bonds 164
7.4.2.4 Replacement of Cys65 with Ala 164
7.4.2.5 Molecular Dynamics Simulation of Cys65Ala RNase A 165

Bibliography 166

Appendix I Conferences
Appendix IIa Analytical Data -Derivatives of Fmoc-Phenylalanine
Appendix IIb Analytical Data – Z-RLF aldehyde

Figures and Tables

Figure 1.1	Amino Acid, Peptide and Protein	4
Figure 1.2	Scheme for Solid Phase Peptide Synthesis	6
Figure 1.3	Non-native chemoselective ligation of a C-terminal aldehyde to an N-terminal hydrazine, amino-oxy or cysteine functionality	11
Figure 1.4	Thiazolidine ligation of a membrane permeable sequence to a functional domain	12
Figure 1.5	Scheme showing thiaproline ligation via a thiazolidine intermediate	14
Figure 1.6	Native chemical ligation	16
Figure 1.7	Staudinger ligation	18
Figure 1.8	Structure of Bovine Pancreatic Ribonuclease A	22
Figure 1.9	Mechanism of Ribunuclease A Catalysis	23
Figure 2.1	Reduction of Weinreb Amide	30
Figure 2.2	Oxidation of a diol or amino alcohol to aldehyde	32
Figure 2.3	Syntheiss of amino acids from phthalyl protected amino acids	33
Figure 2.4	Synthesis of diols directly from Fmoc protected amino acids	34
Figure 2.5	Sequential chemoselective ligation of four peptides by oxime bond	36
Figure 2.6	ESMS of peptide formed by oxime sequential chemoselective ligation	37
Figure 2.7	Putative synthetic strategy for attachment of epoxide to amino resin	40
Figure 2.8	^{1}H NMR of Fmoc-phenylalanine diazomethylketone	41
Figure 2.9	^{1}H NMR of Fmoc-phenylalanine bromomethylketone	42
Figure 2.10	Synthesis of 1, 2 amino alcohol derivative of Fmoc phenylalanine	44
Figure 2.11	^{1}H NMR of Fmoc-phenylalanine azidomethylketone	45
Figure 2.12	^{1}H NMR of Fmoc-phenylalanine azidomethylalcohol	46

Figure 2.13 HPLC and MS data for intermediates and products in the synthesis of the 1, 2 amino alcohol derivative of Fmoc phenylalanine 48
Figure 2.14 ESMS of Fmoc-Phe amino alcohol immediately after synthesis and after storage for 1 week at 4ºC 50
Figure 2.15 HPLC of Ac-AAA-amino alcohol 51
Figure 2.16 Synthesis of the peptide aldehyde Z-RLF-al 52
Figure 2.17 HPLC of Z-RLF-amino alcohol 53
Figure 2.18 HPLC of Z-RLF-aldehyde 54
Figure 2.19 ESMSof Z-RLF-aldehyde 55
Figure 2.20 NMR spectrum of Z-RLF-aldehyde 55
Figure 2.21 ESMS data for C-terminal aldehyde peptides 57
Figure 2.22 Amberlyst ion exchange resin treated with sodium periodate 57
Figure 2.23 Primary sequence of bovine pancreatic ribonuclease A, split into four peptide fragments 59

Figure 3.1 Comparison of structures of natural tripeptide Xaa-Xaa-Xaa, with different chemoselective bond replacements 65
Figure 3.2 N-terminal serine oxidation to glyoxal 67
Figure 3.3: ESMS data for SLIGRL-NH2 70
Figure 3.4: ESMS of 8Aoc-fQWAVGHL-NH2-NH2 72
Figure 3.5: ESMS of glyoxal-LIGRL-NH2 74
Figure 3.6: ESMS of trial ligation of 8Aoc-fQWAVGHL-NH2-NH2 to glyoxal-LIGRL-NH2 after 8 hours 77
Figure 3.7: ESMS of 8Aoc-fQWAVGHL-hydrazone-LIGRL-NH2 78

Figure 4.1: Primary sequence of bovine pancreatic ribonuclease A 81
Figure 4.2: Hydrazide chemoselective ligation of C-terminal aldehyde to N-terminal hydrazine 82
Figure 4.3: Image of RNase A (3rn3), showing sites of ligation and residues critical to ribonuclease activity 83
Figure 4.4 Primary sequence of RNase A, split into four peptide fragments,

using strategy 1 84
Figure 4.5: Sequence of C-terminal peptide fragment of RNase A showing location of pseudoproline dipeptides 85
Figure 4.6: Primary sequence of RNase A, split into four peptide fragments, using strategy 2 86
Figure 4.7: Sequence of peptides 2 and 3 showing location of pseudoproline dipeptides and difficult couplings 87
Figure 4.8: HPLC of RNase A Peptide 4 88
Figure 4.9: ESMS of RNase A Peptide 1 88
Figure 4.10: ESMS of RNase A Peptide 2 89
Figure 4.11: ESMS of RNase A Peptide 3 89
Figure 4.12: ESMS of Peptide 4 of RNase A 90
Figure 4.13: Proposed ligation of four peptide fragments by hydrazone ligation 91
Figure 4.14: Oxidation of RNase A peptide 4 92
Figure 4.15: ESMS showing periodate oxidation of RNase A peptide 4 after 10 minutes 93
Figure 4.16: ESMS showing periodate oxidation of RNase A peptide 4 after 5 minutes 94
Figure 4.17: ESMS of SDCRETGSSK with and without cysteine protection, before and after oxidation 96
Figure 4.18: Maximum Entropy plot for glyoxal of Peptide 4 protected with S-tBu 98
Figure 4.19: Maximum Entropy plot for glyoxal of Peptide 4 protected with ACM 100

Figure 5.1 Native bovine pancreatic ribonuclease A solvated with H_2O 114
Figure 5.2 Superposition of native RNase A before and after MD 116
Figure 5.3 Overlay of native ribunuclease and ribunuclease with Ala20 replaced with an oxime bond 118
Figure 5.4 Overlaid structures of Ile81-Thr82-Asp83 in native RNase A and RNase A containing a hydrazide chemoselective bond 119
Figure 5.5 Native ribonuclease superposed with ligated ribonuclease before molecular dynamics simulation 120

Figure 5.6 Energy plot (KJ/Mol) for MD simulation of ribonuclease containing hydrazide bonds 121
Figure 5.7 Native RNase A superposed with ligated RNase A after molecular dynamics simulation (t = 200 ps) 122
Figure 5.8 Superposition of RNaseA containing three hydrazide bonds from three different timepoints; t = 100 ps, t = 150 ps and t = 200 ps 123
Figure 5.9 Cys65Ala RNase A superposed with native RNase A 125

Table 7.1 Chemicals and Suppliers 132
Table 7.2 The standard side chain protecting groups used in Fmoc peptide synthesis 135
Table 7.3 Standard automated synthesis program 138

Abbreviations

ACM	acetamidomethyl
Ac_2O	Acetic anhydride
Boc	Tert-butoxy carbonyl
CTR	2-chlorotrityl chloride resin
DBU	Diaza(1,3)bicyclo[5.4.0]undecane
DCM	Dichloromethane
DDT	Dithiothreitol
DIPEA	Diisopropyl ethylamine
DMF	Dimethylformamide
DNA	Deoxyribonucleic acid
ESMS	Electrospray Mass Spectrometry
EtOH	Ethanol
Et_2O	Diethyl ether
Fmoc	9-Fluorenylmethoxycarbonyl
HOBt	N-Hydroxybenzotriazole
HBTU	2-(1 H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HIV	Human Immunodeficiency Virus
HOBt	N-Hydroxybenzotriazole
HPLC	High Pressure Liquid Chromatography
$hsPLA_2$	Human Secretory Phospholipase A_2
MD	Molecular Dynamics
MeOH	Methanol
MPS	Membrane-Permeable Sequence
Msc	2-(methylsulfonyl) ethyl carbonate
NCL	Native Chemical Ligation
NMP	N-methyl pyrolidone
NMR	Nuclear Magnetic Resonance
PAR2	Protease-activated Receptor-2
Pbf	2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl

PyBOP	Benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate
RMSD	Root Mean Square Deviation
RP-HPLC	Reverse Phase High Pressure Liquid Chromatography
RNA	Ribonucleic acid
RNase A	Ribonuclease A
SNB	S-(5-thio-2-nitrobenzoic acid)
SPPS	Solid Phase Peptide Synthesis
S-tBu	Thio-tertiary butyl
t-Bu	Tertiary butyl
TCEP	tris(2-carboxyethyl) phosphine
TEA	Triethylamine
TFA	Trifluroacetic acid
THF	Tetrahydrofuran
TIS	Triisopropyl silane
TLC	Thin Layer Chromatography
Trt	Trityl

Acknowledgements

This thesis is dedicated to Howard Moreton, who sadly passed away during the preparation of this thesis. I know you would have been proud of this achievement.

Many thanks go to Ann Moreton and Steve Council for their support and patience, particularly during the last few months. Thank you for always believing in me and encouraging me to finish what I had started.

I would like to thank everyone at Peptide Protein Research Ltd, particularly Robert Broadbridge for advising and guiding the project and to Michele Farris for proof-reading and suggestions.

Thanks to Neil Ward and Brendan Howlin for their supervision and support, and to everyone in the Chemistry division at the University of Surrey for their help over the years.

Chapter 1: Introduction

1.0 Introduction

Deoxyribonucleic acid (DNA) is often called the molecule of life, but it is the proteins encoded by DNA that are responsible for the cellular structure and biological processes that we call life. Proteins provide structural support in cells, they are catalysts for most cellular chemical reactions, and they are antibodies, transport molecules and cell receptors. Proteins are polymers of amino acids, the sequence being determined by the order of nucleotide bases in DNA, and it is this sequence of amino acids and the three-dimensional structure it adopts that determines the protein's biological function. The study of protein sequence, structure, function and interaction is essential for the more detailed understanding of the biological processes of life.

The study of a protein's structure and function requires a source of that protein. Some proteins, particularly the earliest examples studied, can be extracted from cells and purified. Early protein purification was achieved using crystallisation techniques such as precipitation from ammonium sulphate, used by Kunitz (1940) for isolation of crystalline Bovine Ribonuclease A. This technique works if the protein is present in the cell in large quantities, for example digestive enzymes, but very few proteins are that abundant. Ion exchange chromatography, which can separate proteins with different net charge, was first used in 1950 by Paléus *et al.* to purify cytochrome C. This technique allowed proteins to be purified to a much higher purity than simple precipitation, for example Hirs *et al.* (1953) separated ribonuclease A from ribonuclease B (which differs from the A form only by a single carbohydrate molecule) by ion exchange chromatography. However the problem still remained that only those proteins abundant in cells could be extracted and purified.

Recombinant technology, developed by Cohen *et al.* (1973) can be used to insert the gene for a protein of interest into a prokaryotic organism like *E. Coli*, which can then be induced to express that protein in large quantities. Recombinant proteins however

are costly and time consuming, particularly if many mutant versions are required. The process of producing recombinant protein first requires the synthesis of recombinant DNA. Recombinant DNA means DNA that has been created through molecular cloning techniques, usually combining DNA from the protein of interest with other DNA sequences that are required for the host cell to replicate the DNA and express it to produce protein. Using enzymes that break and join DNA, the recombinant DNA is inserted into a cyclic DNA ring called a plasmid, and this plasmid is transplanted into the host bacterial cell in a process called transfection. The transfected bacteria are fermented and the plasmid DNA is replicated along with the host cell DNA, and the protein of interest is expressed.

Recombinant proteins are limited to the twenty natural amino acids. There are a large number of proteins that undergo enzymatic post-translational modification of the natural amino acids, for example glycosylation, and these modifications cannot be copied in recombinant cells, as the host cells do not have the enzymes required. There are also a number of proteins that cannot be produced by prokaryotic expression, either because the proteins are toxic to the host cell, or because they are folded incorrectly in prokaryotic cells. This insoluble protein precipitates in the cell to form an inclusion body that cannot be easily purified. Eukaryotic expression systems such as yeast cells can be used (Botstein *et al.* (1979)), but are significantly more expensive and lower yielding than prokaryotic expression.

Total chemical synthesis can provide an alternative source of protein, with the advantage that amino acid substitutions can easily be made with either natural or unnatural amino acids. Total synthesis of proteins is in its infancy compared to recombinant protein production, but a number of small proteins have been synthesised, for example the chemical synthesis of a HIV-1 protease analogue by thioether ligation (Englebretsen *et al.* (1995)) and the review by Kent and Dawson (2000) gives the details of several proteins synthesised by chemical ligation. A number of other synthetic proteins are discussed in more detail in **Section 1.3**.

1.1 Peptides and Proteins

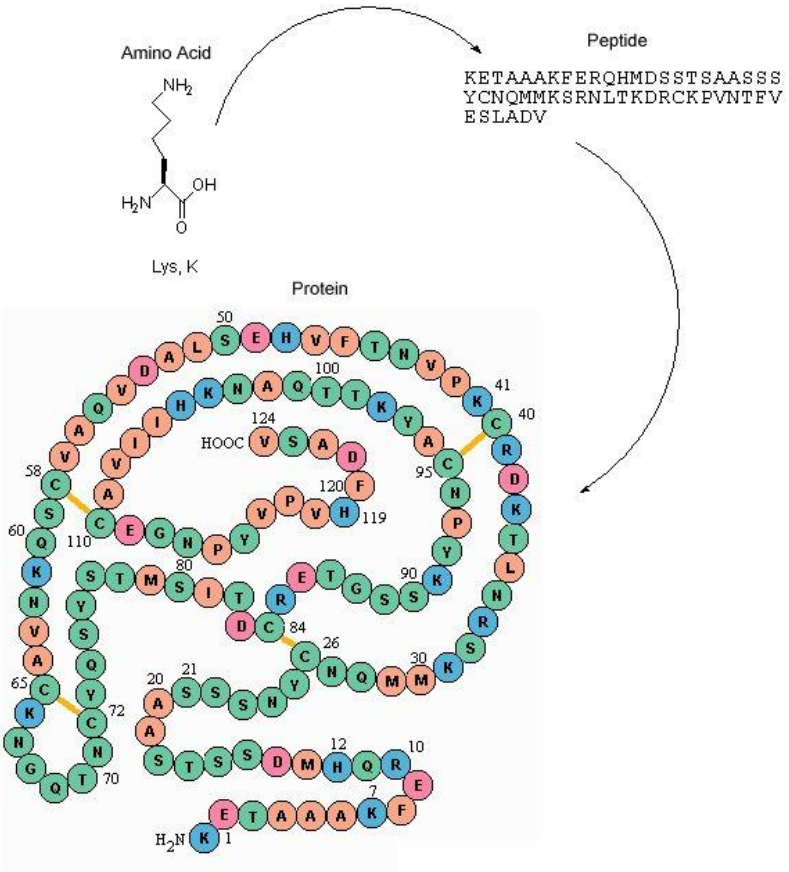
Amino acids when joined together by an amide bond are called peptides; from two amino acids, which form a dipeptide, up to many amino acids, known as polypeptides or proteins. There are twenty natural amino acids, which have a general structure consisting of an amino and a carboxylic acid group, with each amino acid bearing a different chemical group on the side chain. The characteristics of the side chain groups can be split into either hydrophobic or polar residues, with some of the polar side chains also having acidic or basic properties. An amino acid chain folds to form a three-dimensional structure due to interactions between these amino acid side chains. It is this structure together with specific side chain functionalities that give a protein its biological activity.

It has become common practice to define any chain of less than 100 amino acids as a peptide, and longer chains as either polypeptides or proteins. A more accurate definition for a protein is an amino acid sequence determined by the order of nucleotide base pairs in the DNA that encodes that protein [1]. In practice there are very few proteins that contain fewer than 100 amino acids, because short peptides do not have enough side chain groups to interact and form a specific three-dimensional structure. Chemical synthesis has a size limit on the number of amino acids that can easily be joined together of about 100, which is why this figure is often incorrectly used to define the difference between peptides and proteins.

Peptide and protein sequences are, by convention, written from the N to C direction (**Section 1.2.1**). This is called the primary sequence and it gives just the order of amino acids within the peptide or protein, with no structural information. The folding of a peptide or protein sequence is called secondary structure. Secondary structure includes covalently bonded disulfide bridges between cysteine residues and regions of the protein that form loops or structural motifs called alpha helixes and beta sheets, which are held by non-covalent hydrogen bonds. **Figure 1.1** shows the relationship between an individual amino acid, a peptide primary sequence and a full protein sequence (bovine ribonuclease A) showing simple secondary structure loops and disulfides, but not areas of alpha helix or beta sheet. The full three dimensional

structure of all the amino acids within a protein is called tertiary structure, and the tertiary structure of ribonuclease A, represented in a ribbon structure is shown in **Figure 1.9** [2].

Figure 1.1: Amino Acid, Peptide and Protein (Bovine Ribonuclease A used as the example protein, adapted from [3])



1.2 Peptide Synthesis

The synthesis of even a small peptide requires many individual chemical steps. Traditional solution peptide synthesis is an extremely time consuming process which often produces low yields of peptides. The invention of Solid Phase Peptide Synthesis (SPPS) by Merrifield (1963) revolutionised the synthesis of peptides, and now allows the routine synthesis of high yielding peptides in up to kilogram quantities. Solution phase synthesis is often still the technique of choice for multi-

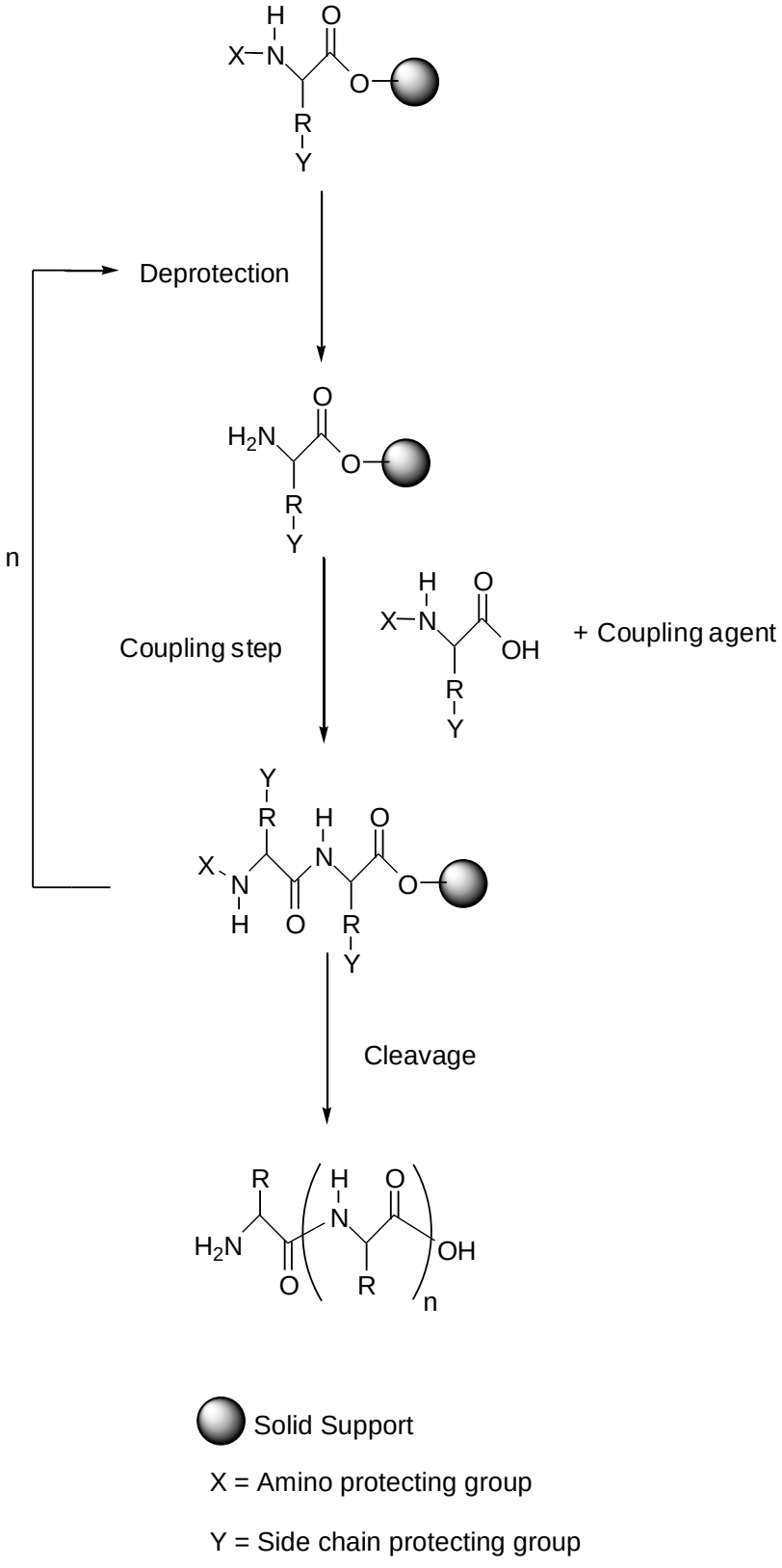
gram scale peptide synthesis, mainly due to the much higher cost of the reagents for solid phase synthesis; however, a huge amount of process development is required to determine a solution phase strategy for each individual peptide. Recently solid phase synthesis has started to be used for large scale manufacture of peptide drugs. The HIV fusion inhibitor T20 (Fuzeon® enfuvirtide) manufactured by Hoffmann-La Roche, Inc was the first peptide drug synthesised by solid phase peptide synthesis in multi-ton quantities. In collaboration between Trimeris and Roche, a synthesis strategy was developed to synthesise the peptide in three fragments by solid phase synthesis and then use solution phase synthesis to join the fragments together, producing the peptide in higher yield and purity than synthesis by solid phase synthesis alone (review by Bruckdorfer *et al.* (1994)). The increase in demand for Fmoc protected amino acids and the polymeric solid support required for this synthesis is reputed to be responsible for the dramatic worldwide drop in price for these reagents in the early 2000s [4].

1.2.1 Solid Phase Peptide Synthesis

In solid phase peptide chemistry an amino acid is attached to a solid polymeric support through the carboxylic acid, and amino acids are sequentially added by forming a peptide bond between the two amino acids. The amino acid to be added is activated with a coupling agent, which forms an active ester of the carboxylic acid. This reacts with the amino acid attached to the solid support to form the amide bond. The technique is outlined in the diagram in **Figure 1.2** and described below, but more information can be found in the book Solid Phase Synthesis: A Practical Approach, by Atherton *et al.* (1989). Synthesis proceeds from the C to the N direction while peptide sequences are written from N to C, which can be confusing to those outside the field. The convention of N to C which is the same direction as the term “amino acid” was defined before the invention of solid phase synthesis, which was synthetically less challenging in the C to N direction.

The peptide synthesis and cleavage from the resin occurs in a reaction vessel containing a frit which retains the resin, as the peptide is bound to the solid support

Figure 1.2: Scheme for Solid Phase Peptide Synthesis



during synthesis. Large excesses of the amino acids with respect to the growing peptide chain are used to force the reactions to completion, and the excess is simply washed away from the resin after the reaction. The amino group has a temporary protection, to prevent more than one amino acid reacting per peptide chain, which is removed after addition of that amino acid to the growing peptide chain, ready for addition of the next amino acid. The two most common amine protecting groups used are tertiary-butoxycarbonyl (t-Boc or Boc) and 9 fluorenyl methoxycarbonyl (Fmoc). The amino acid side chain reactive groups are protected with chemical groups to prevent them participating in the coupling reaction, that are stable to the conditions used for removing the amine protecting group, but can be removed at the end of the synthesis. The side chain protecting groups are usually removed using the same chemical conditions as cleavage of the peptide from the solid support. The use of a set of protecting groups on the amino group and the side chain of an amino acid, which have different conditions of removal, is known as orthogonal protection (Merrifield *et al.* (1977)). Orthogonal protection means that under the deprotection conditions of one protecting group the other is stable and vice versa. An example of this is the Fmoc and Boc amine protecting groups which can be used for example on the alpha and epsilon amines of lysine during peptide synthesis. The Fmoc group is removed under basic conditions during chain extension, and the Boc group is removed under acidic conditions to deprotect the side chain. At the end of synthesis the peptide is cleaved from the resin which releases the peptide from the solid support, allowing the peptide to be filtered and collected while the resin is retained on the frit in the reaction vessel. The side chain protecting groups are usually removed using the same chemical conditions as cleavage of the peptide from the solid support, but if protected peptides are required then a solid support can be used that allows cleavage without deprotection of the side chains.

Synthesis by Fmoc chemistry has become the most common strategy used, due to the requirements in Boc chemistry for hydrogen fluoride (HF) to cleave the peptide from the solid support. Synthesis by Boc chemistry often produces higher quality crude peptides, attributed to use of trifluoroacetic acid for deprotection of the Boc group breaking up chain aggregation, but the safety considerations with using HF have meant that most solid phase synthesis is now performed using Fmoc chemistry.

There is a limit to the size of peptides that can be synthesised on the solid support, due to formation of secondary structure, and aggregation of the peptide, making it difficult for the incoming amino acid to access the end of the growing peptide chain. The number of amino acids that can be joined is dependent on the peptide sequence, but it is difficult to synthesise any peptide sequence longer than approximately 100 amino acids. Even if each individual reaction is very high yielding, after 100 reaction cycles the cumulative effect of small impurities makes the final protein difficult to purify.

1.3 Synthesis of Proteins

Although solid phase peptide synthesis has physical limits on the size of peptides that can be made, this has not prevented the total chemical synthesis of much larger peptides and proteins. The solution has been to join together peptide fragments synthesised on the solid phase, a process that is called peptide ligation. Many different ligation techniques have been successfully employed in the production of peptides that cannot be synthesised as a single peptide chain by SPPS. The ligation methods can be split into two broad classes, fragment condensation and chemoselective ligation.

1.3.1 Fragment Condensation

Fragment condensation is used to join a side chain protected peptide fragment to another side chain protected fragment, either with both fragments in solution or with one fragment still attached to the solid phase. Peptide fragments are usually synthesised on an acid labile solid support, such as 2-chlorotriryl chloride resin, allowing cleavage of the peptide from the solid support, but leaving the side chain protections in place. A normal coupling reaction can be employed for the condensation of the two fragments as the side chain functionalities are protected from participation. Typically the reaction requires long coupling times, in part due to the low solubility of protected peptide fragments, which means large excesses of the

carboxyl-containing peptide are required. This often leads to racemisation at the coupling site. Racemisation can be avoided by ensuring that the site for fragment condensation contains a glycine, proline, or a pseudoproline (a protected dimer containing serine or threonine with a structure similar to proline developed by Mutter *et al.* (1995)) as the carboxyl-containing residue. Glycine is an achiral residue, and proline or pseudoproline does not undergo epimerisation during the coupling reaction, allowing racemisation-free fragment condensation.

Protected peptide fragments cannot easily be purified due to difficulty in solubilising the peptides, so impurities in both fragments will be present in the final product, which is therefore often difficult to purify to high purity. For this reason fragment condensation is most useful for the synthesis of short peptides of 20-40 amino acids that have proven difficult to synthesise as a single peptide, using fragments of about 10-20 amino acids. Short peptides can be made more easily to high purity so it is not so important that the fragments are not easily purifiable. A thorough review of peptide synthesis via fragment condensation can be found in Methods in Molecular Biology, Vol 35 Peptide Synthesis Protocols (Nyfeler, R. (1994)).

1.3.2 Chemoselective Ligation

The solubility and purification problems associated with protected fragments as seen with fragment condensation have been avoided by using unprotected peptides, which are more readily soluble in aqueous solution. Unprotected peptide fragments can be purified by high performance liquid chromatography (HPLC) and analysed by mass spectrometry in the same way as any other synthetic peptide, so purification of the final protein simply requires separation of the product from any unreacted peptide fragments, to yield a high purity protein. The use of unprotected fragments requires a specific reaction between the N-terminus of one fragment and the C-terminus of the other, without unwanted side reactions from the participation of the unprotected side chain functional groups. This reaction that produces the ligation of the peptides is called chemoselective ligation and it can be any chemical reaction that achieves the specific ligation of the two peptides. Chemoselective ligation reactions can be split

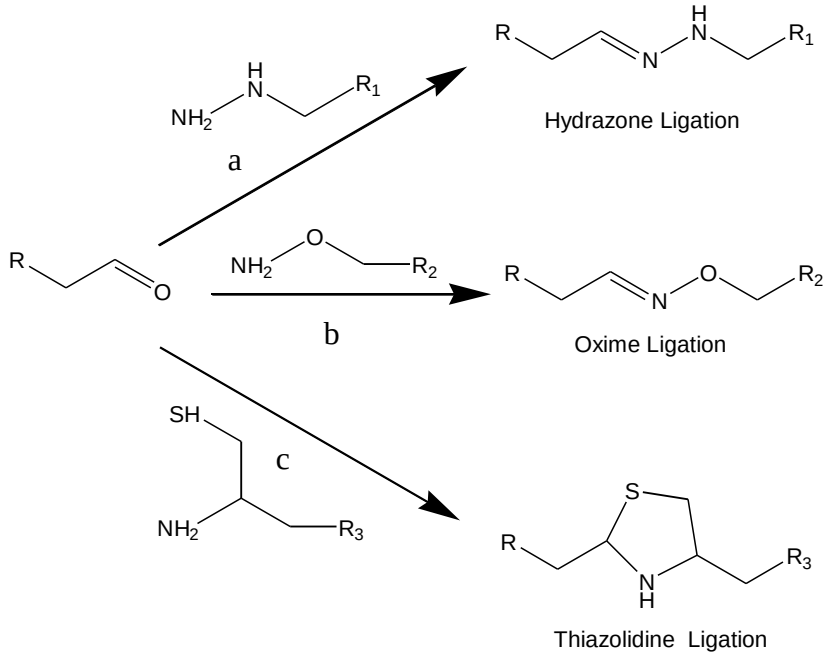
into two types, non-amide ligation and native chemical ligation. Non-amide ligation uses mutually reactive groups on the peptide fragments, which once ligated form an unnatural or non-amide linkage (**Section 1.3.2.1**). Native chemical ligation forms a natural peptide bond at the ligation site, and is a two-stage reaction where first a non-amide ligation occurs, followed by an intramolecular transfer reaction to form the natural peptide bond (**Section 1.3.2.2**).

An important feature of any ligation technique is the ease in which the groups required for ligation can be incorporated into the synthesis of the peptide fragments. Ideally the C-terminal group is added by reacting a linker or small molecule to a commercially available resin at the start of synthesis, and then normal SPPS can be used to construct the peptide fragment. For the fragment containing the N-terminal functionality, the specific functional group is coupled to the peptide on the solid phase at the end of synthesis, before cleavage from the resin.

1.3.2.1 Non-Amide Ligation

There have been many different chemical reactions used for non-amide ligation, and they have been used for a number of different applications. These applications include the joining of two peptide segments (Liu *et al.* (1994)), the formation of peptide dendrimers, where a peptide is ligated multiple times onto a template molecule as demonstrated by Tuchscherer (1993) and Rose (1994), and as a disulfide bridge replacement (Wahl *et al.* (1996)). One common approach uses an aldehyde, which at acidic pH is unreactive towards any natural amino acid side chain, reacting with a nucleophile. Nucleophiles of choice have been either an N-terminal cysteine residue, which forms a thiazolidine ring on reaction with an aldehyde, or a hydrazine or amino-oxy group, which produces hydrazone or oxime bonds (**Figure 1.3**). The aldehyde can be generated from periodate oxidation of a 2 amino alcohol or 1, 2-diol. The ligation reaction is carried out in acidic aqueous buffer (usually about pH 5.5) at room temperature (Rose (1994)).

Figure 1.3: Non-native chemoselective ligation of a C-terminal aldehyde to an N-terminal hydrazine, amino-oxy or cysteine functionality (adapted from Tam *et al.* (1995))

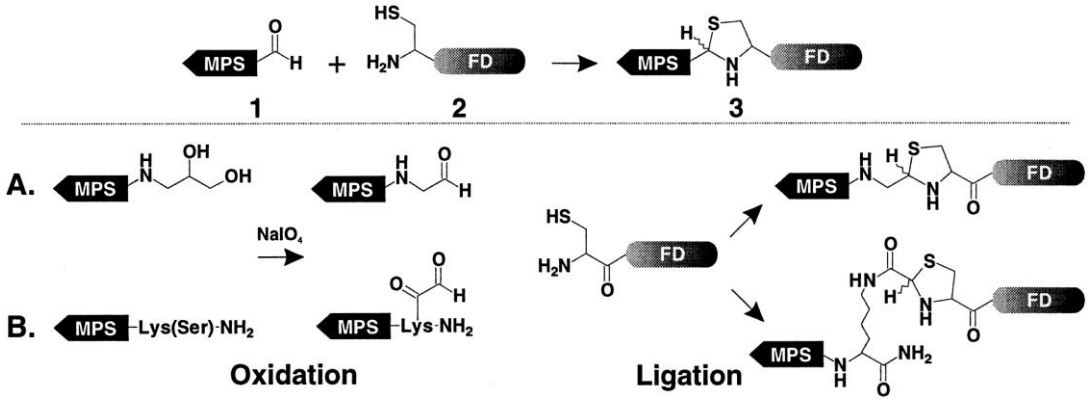


a) Hydrazine ligation
b) Amino-oxy ligation
c) Cysteine ligation

An example of thiazolidine ligation was used by Zhang *et al.* (1998) to ligate an active peptide fragment to a membrane permeable sequence (MPS) so that the functional peptide could be imported into cells (**Figure 1.4**). A C-terminal aldehyde was created in one of two ways. Either a serine was added to the side chain of a C-terminal lysine residue, or 3-amino 1-2-propanediol was attached to chlorotrityl resin, and then the MPS was synthesised on this functionalised resin. Periodate oxidation of the cleaved peptide generated an aldehyde at the C-terminus. Ligation in aqueous buffer to the functional peptide fragment which contained an N-terminal cysteine produced a cell-permeable bioactive peptide, which could be used in biological assays immediately after ligation, without product isolation or purification.

The presence of a non-amide bond did not affect the peptide's cell uptake or function when compared to peptide synthesised as a single chain.

Figure 1.4: Thiazolidine ligation of a membrane permeable sequence to a functional domain (Zhang *et al.* (1998))



MPS: Membrane permeable sequence

FD: Functional domain

Oxime ligation has been a popular choice in the assembly of peptide dendrimers. Tuchscherer (1993) used a template with four aldehyde groups to ligate four identical peptide fragments each containing an N-terminal hydroxylamine functionality to form an oxime bond. Similarly Rose (1994) used the oxime bond to synthesise a 20kD peptide dendrimer, by ligating six identical peptides with aminooxyacetic acid at the C-terminus to a template molecule containing six aldehyde groups. An oxime bond has also been used to replace the disulfide in Oxytocin, a pituitary hormone, by Wahl *et al.* (1996) by replacing the two cysteine residues with a serine, as an aldehyde precursor, and a side chain hydroxylamine group.

Most of the reagents for these non-native ligation techniques are now commercially available with the appropriate protecting groups for use in solid phase peptide synthesis. The hydrazine group can be introduced as tri-Boc-hydrazinoacetic acid, and the hydroxylamine group as bis-Boc-amino-oxyacetic acid, and both groups can be coupled onto the N-terminus of a peptide at the end of synthesis. Fmoc-3-

amino,1-2-propanediol functionalised chlorotrityl resin is also available, enabling the facile incorporation of the groups required for ligation into solid phase synthesis. Other reagents are available which contain the groups for non-native ligation on the side chains of modified amino acids, allowing the formation of branched or cyclic ligated products (several commercially available reagents are detailed in **Chapter 3**).

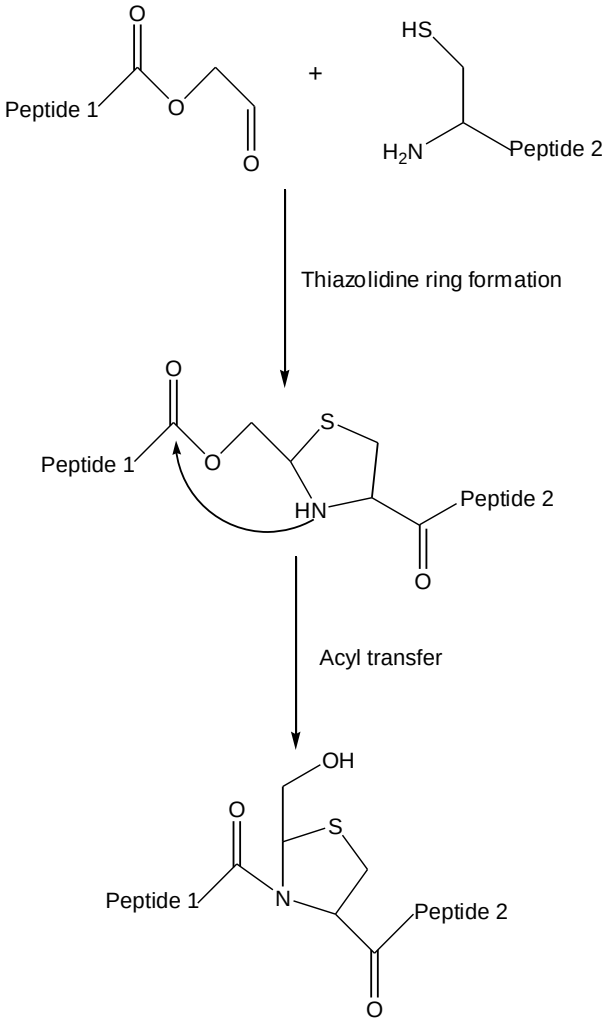
The inclusion of a non-natural bond in a protein would be expected to have some effect on the structure of that protein. The three-dimensional structure adopted by a protein is formed through interactions between amino acid side chain groups and also through interactions with the backbone amide groups. When a non-amide linkage is inserted, this not only replaces the amide bond with an alternative structure, but often one or both of the amino acid side chains at the ligation site are either structurally altered or absent (**Chapter 3, Figure 3.1**). In order to minimise the effect that this non-amide bond may have, the ligation site must be chosen carefully. An obvious site for non-native ligation is at a glycine-glycine bond, due to the absence of side chain groups. A thioether bond was used by Englebretsen *et al.* (1995) as a glycine-glycine bond replacement to synthesise enzymatically active HIV-1 protease. The bond was formed between an N-terminal bromoacetylated fragment and a C-terminal thiol. The protein had very similar activity and substrate specificity to the native protein, which suggests that a non-native bond has a minimal effect on protein structure, if the ligation sites are carefully chosen. However, there have been no investigations into the precise conformations of these non-amide linkages or the structures of the ligated proteins, which is one of the objectives of this study.

1.3.2.2 Native Chemical Ligation

The one major disadvantage to non-native ligation is the presence of a non-amide bond within the protein product. These non-amide bonds not only have different structures to the natural amide bond but they are also less stable at high or low pH. A more desirable chemoselective ligation is one that forms an amide bond at the site of ligation.

In an extension to thiazolidine ligation, Liu and Tam (1994) demonstrated that if a glyoxylaldehyde is used instead of a simple aldehyde, then ligation to form a thiazolidine ring is followed by an acyl transfer reaction to give a thiaproline bond (**Figure 1.5)**. This thiaproline linkage is an amide bond so has similar properties to a proline peptide bond, unlike the thiazolidine ring formed in non-amide ligation, which is unstable at high and low pH. Replacement of Pro39 in HIV-1 protease with a thiaproline bond did not affect the folding, conformation or activity of the protein Liu and Tam (1996).

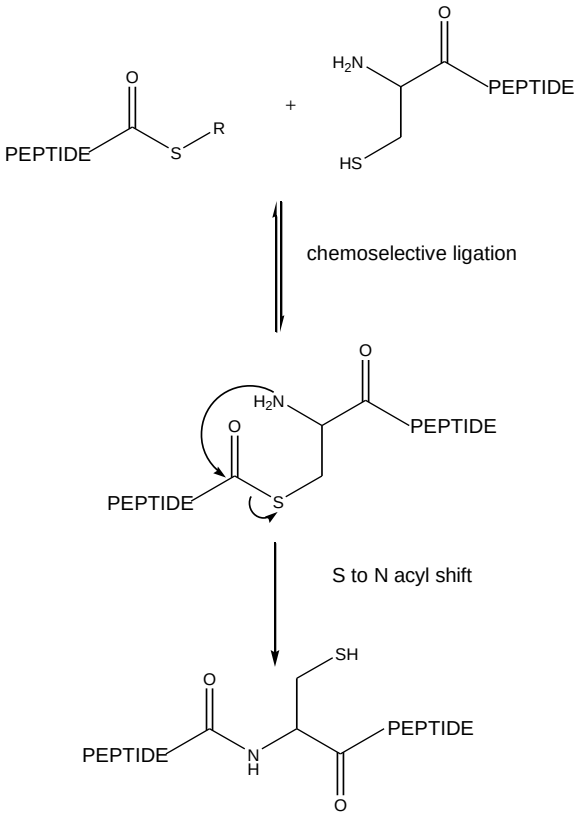
Figure 1.5: Scheme showing thiaproline ligation via a thiazolidine intermediate (adapted from Liu and Tam (1994))



Although the thiaproline bond is an amide bond it does not have the same structure as the natural peptide bond. A chemoselective ligation methodology has been developed that does create a native peptide bond at the ligation site, given the term native chemical ligation (NCL).

NCL, like the thiaproline ligation, is a two-stage reaction, first forming a chemoselective linkage, which then undergoes a rearrangement reaction to form the native peptide bond. It is based on the reaction and acyl rearrangement of a thioester with an amino thiol, first reported by Wieland (1953). The use of this reaction in the ligation of peptides was developed by the group directed by Kent (Dawson *et al.* (1994)) and it was Kent who gave the reaction the name native chemical ligation. A peptide containing a C-terminal thioester and another peptide containing an N-terminal cysteine are reacted in aqueous buffer at pH7. Firstly a thio-transesterification reaction occurs by nucleophilic substitution of the thiol in the N-terminal cysteine for the thiol in the thioester. This forms an intermediate that spontaneously rearranges by S to N acyl shift to form a natural peptide bond at the site of ligation (**Figure 1.6**).

Figure 1.6: Native chemical ligation (adapted from Kent and Dawson ((2000))



If the peptides for ligation contain other cysteines in addition to the N-terminal cysteine, the thio-transesterification reaction can occur with any of the cysteine thiols as a reversible reaction, but it is only when this occurs with the N-terminal cysteine that the amine is positioned for the spontaneous S to N acyl shift to form the amide bond. The stability of the amide bond is such that the acyl shift reaction can be considered irreversible.

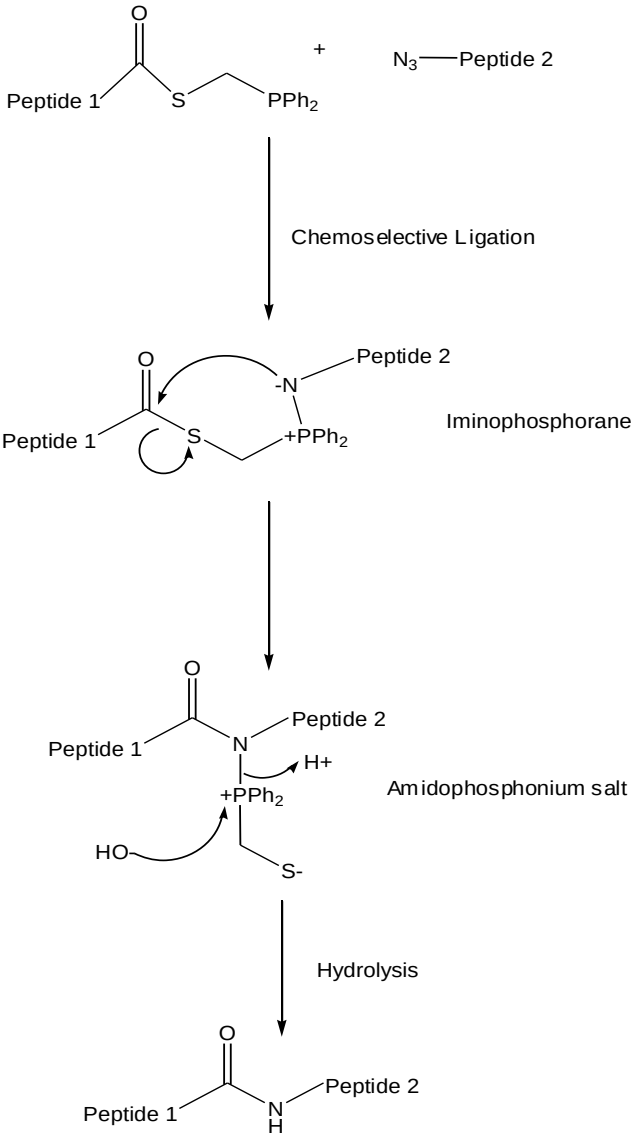
Peptides containing a C-terminal thioester can be synthesised directly using Boc chemistry, as the thioester is stable to the reaction conditionsused in Boc chemistry. Initial synthesis of C-terminal thioester peptides by Boc chemistry (Hacking *et al.* (1999)) utilised a mercapto-carboxylic acid at the C-terminus, attached to the solid support through the carboxyl. Cleavage of the peptide from the solid support would yield the crude thioester peptide which after purification can be used in the NCL reaction. However, the use of Boc chemistry has become disfavoured due to the requirement for HF cleavage. The synthesis of C-terminal thioesters by Fmoc

chemistry is more challenging. This is mainly because the thioester bond is not stable to the conditions of amino acid deprotection used in each cycle of Fmoc peptide synthesis. Therefore, in Fmoc chemistry, the thioester must be formed after synthesis of the peptide sequence. One method is to react the C-terminus of a fully side chain and N-terminal protected peptide with a thiol compound, followed by removal of protecting groups. This causes racemisation with any amino acid other than glycine so is not an ideal method. Alternatively the Kenner safety catch linker (Kenner *et al.* (1971)) has been used by Ingenito *et al.* (1999) to release the C-terminal thioester using a thiol as the nucleophile for peptide cleavage. None of the current methods of synthesis of thioesters by Fmoc chemistry are high yielding and so synthesis of the thioester containing fragment continues to be a challenge, and a serious limitation of the technique.

Another major limitation of NCL is the requirement for a cysteine residue at the site of ligation. More recent research has focused on adaptations of the NCL reaction to remove this requirement. A simple modification has been the use of homocysteine in place of cysteine (Tam (1998)), which can be methylated after ligation to form methionine. In a similar way cysteine has been de-sulfurised to alanine by Yan and Dawson (2001). The most promising development was the use of a removable auxiliary 1-phenyl-2-mercaptoethyl by Low *et al.* (2001) which is removed from the peptide bond after ligation. These adaptations have allowed synthesis of ligated peptides and proteins without a cysteine at the ligation site. However, the yields of these reactions are considerably lower than for standard NCL, and so these methods have not been adopted as routine methods of synthesis.

An alternative native chemical ligation technique that does not require a cysteine residue at the ligation site is known as Staudinger ligation, developed by Nilsson *et al*. (2000). It utilises a reaction by Staudinger (1919) where an azide and a trialkyl phosphine react to form an iminophosphorane intermediate, which then rearranges to form an amidophosphonium salt. Hydrolysis yields an amine and a phosphine oxide (**Figure 1.7**).

Figure 1.7: Putative scheme for Staudinger ligation (adapted from Nilsson *et al.* (2000))



Staudinger ligation has been used for ligation of a number of dipeptides with C-terminal glycine phosphinethioester to longer peptide fragments without racemisation (Nilsson *et al.* (2000), Soellner *et al.* (2002)), but the Staudinger reaction has not yet been successfully extended to the ligation of longer peptide fragments. In the recent review by Schilling *et al.* (2011), there are numerous examples where Staudinger ligation has been adopted as a successful method of bioconjugation, used for labelling of biomolecules with biotin and fluorescent dyes as an alternative to other bioconjugation methods such as Michael addition of a

maleimide to a thiol. Research regarding Staudinger ligation for peptide to peptide ligation has focused on extending the method to other amino acids than glycine as the C-terminal phosphinethioester, and Lue *et al.* (2004) concluded that although ligation using glycine and alanine C-terminal phosphinethioesters is possible, that valine C-terminal phosphinethioester is not possible. This limitation of the technique together with presumed problems in extending the ligation to longer peptides mean that this method of ligation has not been a successful replacement for NCL.

The limitations of NCL requiring a cysteine at the ligation site and difficulties in synthesising thioester peptides in high yields by Fmoc chemistry have meant that although NCL is a more desirable ligation, from the perspective that the ligated product has a fully stable natural peptide bond, non-amide ligation is a serious alternative, with the major advantage that there are no restrictions on the positions of ligation sites. Of course consideration needs to be made of the effect of a non amide bond, which is one of the aims of this investigation.

1.3.3 Orthogonal and Sequential Ligation for Joining More Than Two Peptide Fragments

Chemoselective ligation can be used to join two peptide segments together, enabling the synthesis of longer peptides and proteins than can be synthesised by stepwise SPPS in a single chain. If the chemical synthesis of much larger proteins is to be realised, then several fragments must be joined together, as the size of individual fragments is limited to the length of a peptide that can be synthesised by SPPS.

Orthogonal ligation describes where two mutually exclusive ligation techniques are used, allowing three peptide fragments to be joined together, for example a non-native ligation followed by native chemical ligation. An advantage of orthogonal ligation is that because the two ligation techniques use different chemistries, one ligation technique can be used followed by the other with no alterations to the chemistries of the reactive groups. In a proof of principle experiment Nilsson *et al.* (2003) ligated the dipeptide ribonuclease A 110-111 to RNase A 112-124 via

Staudinger ligation. This was then ligated to ribonuclease A 1-109 by native chemical ligation, to produce enzymatically active ribonuclease, with natural amide bonds at the sites of ligation, as both methods of ligation generate natural peptide bonds.

An alternative to orthogonal ligation techniques is sequential ligation, where the same ligation method is used to join several peptide fragments. This is more complicated than orthogonal ligation as a temporary protection is required to mask one of the chemoselective groups during the first ligation, which is then removed to allow ligation of a third peptide fragment. Depending on the ligation strategy used, different temporary protecting groups have been developed.

Sequential NCL was used by Hackeng *et al.* (1999) to join four peptide fragments together to create the 124 amino acid human secretory phospholipase A_2 (hsPLA_2). The N-terminal cysteines of peptide fragments 2 and 3 were temporarily protected with 2-(methylsulfonyl) ethyl carbonate (Msc) (Balvert-Geers, *et al.* (1975)). After ligation of peptide fragments 4 and 3, the Msc group was removed by a five-minute treatment at pH 13, unmasking the cysteine residue ready for ligation to peptide 2. The cycle of deprotection and ligation was repeated to ligate peptide fragment 1, yielding synthetic hsPLA_2 that was indistinguishable from recombinant protein.

A combination of both non-amide ligation, and native chemical ligation was used by Muir *et al.* (1997) to assemble four peptide fragments. Two peptides, one bearing a C-terminal aldehyde and the other a C-terminal hydroxylamine group, were ligated to form an oxime bond. Each peptide contained an N-terminal cysteine, which were orthogonally protected with either Msc or S-5-thio-2-nitrobenzoic acid (SNB). After oxime ligation the SNB group was removed with tris(2-carboxyethyl) phosphine (TCEP), and a peptide thioester was ligated to this by native chemical ligation. The Msc group was then removed in high pH from the other N-terminal cysteine, which was then ligated to a second thioester-bearing peptide, to yield the final ligated product.

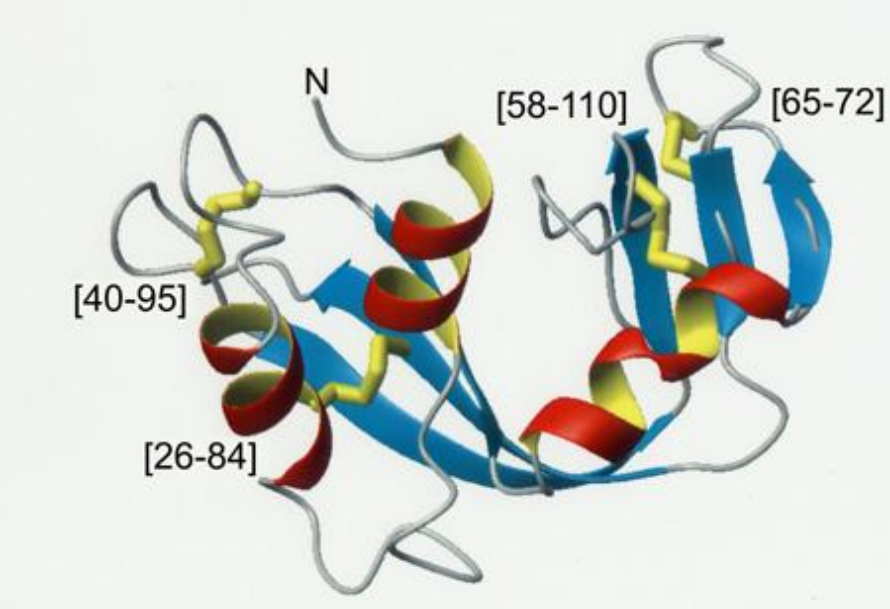
Sequential non-amide ligation has been used to ligate three peptide fragments, to create a peptide based on the C-terminus of a cytokine that had proven problematic to synthesise as a single chain or by fragment condensation (Broadbridge *et al.* (2001)). The Fmoc-3-amino-1-2-propanediol (glycine diol) method of creating C-terminal aldehydes was adapted by synthesising diols of other amino acids. This allows a chemoselective ligated peptide to more closely mimic the structure of a native peptide, by including the side chain of the amino acid close to the ligation site. The diol, which is a masked aldehyde, was used as the temporary protection during the ligation reactions. This work is reviewed in more detail in **Chapter 2 Section 2.2**. A significant limitation of the diol is, with the exception of Fmoc glycine diol, the low level of substitution achieved with chlorotrityl resin. A 1, 2-amino alcohol is an alternative masked aldehyde to the 1, 2-diol, and the amino functionality is significantly more reactive towards chlorotrityl resin than the alcohol. An objective of this study was the development of a synthetic route for the preparation of 1, 2-amino alcohol derivatives of Fmoc protected amino acids, and the use of these derivatives in the preparation of peptides containing C-terminal aldehydes for use in non-amide chemoselective ligation.

1.4 Bovine Pancreatic Ribonuclease A

Bovine pancreatic ribonuclease A (RNase A) was chosen as an example for the total synthesis of a protein using sequential chemoselective (non-amide) ligation. It is a pancreatic enzyme that catalyses the hydrolysis of ribonucleic acid (RNA), and consists of a single peptide chain of 124 amino acids with four disulfide bridges (**Figure 1.8**). RNase A is the unmodified form of ribonuclease produced in the pancreas, where the “A” is used to distinguish the protein from RNase B, C and D, which are different glycosylated forms of ribonuclease, also produced by the pancreas. Ribonuclease A, mainly due to its small size and ease of availability, is probably the most studied and the best characterised protein. Between 1960 and 1963 the work of Hirs, Spackman and Smyth (Hirs (1960) Hirs *et al.* (1960) Spackman *et al.* (1960) Smyth *et al.* (1962) and Smyth *et al.* (1963)) lead to the determination of the primary sequence of ribonuclease, the first protein to have been sequenced. It was

also the first whole protein to be synthesised by stepwise solid phase peptide synthesis by Gutte and Merrifield (1971) using Boc chemistry.

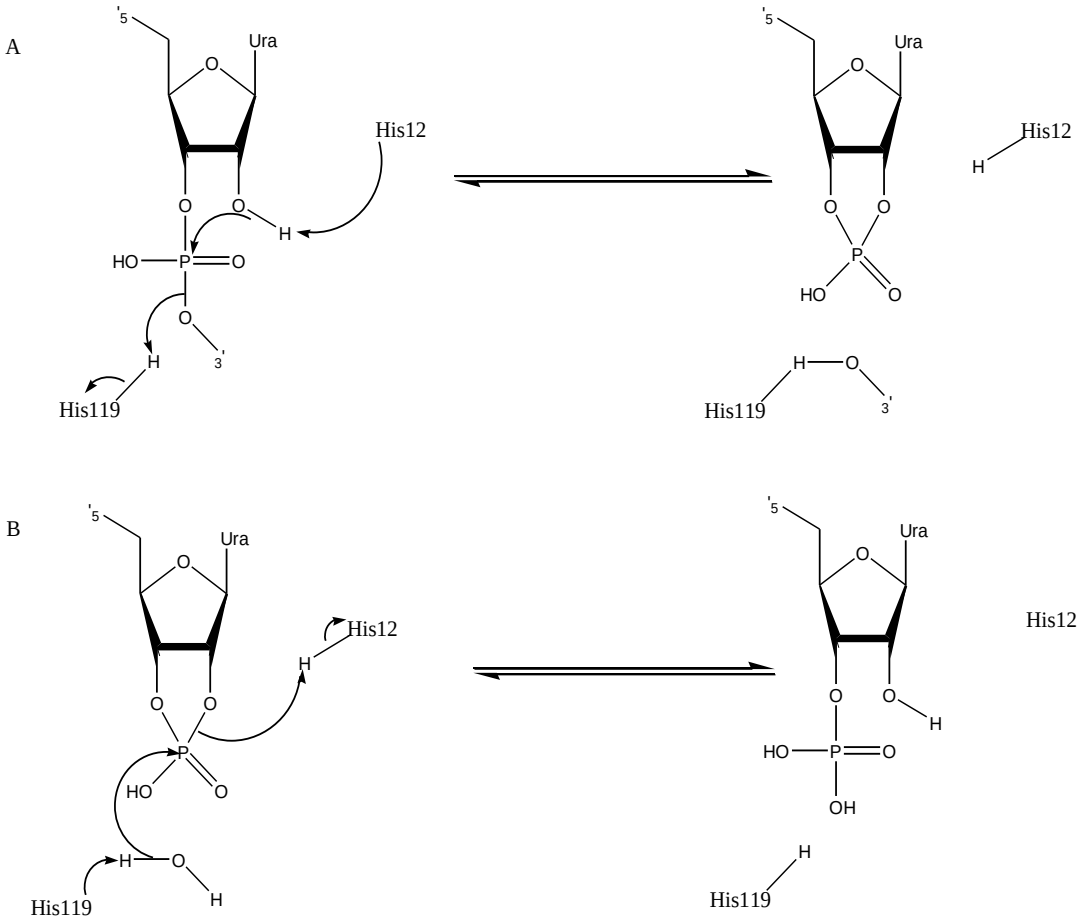
Figure 1.8: Structure of Bovine Pancreatic Ribonuclease A [2]



1.4.1 Catalytic Mechanism of Ribonuclease

RNase A catalyses the hydrolysis of RNA by cleaving the phosphodiester backbone on the 3' side of uracil and cytosine. The catalysis involves two histidine residues, His12 and His119, as proton donors and acceptors (Findlay *et al.* (1961).

Figure 1.19: Mechanism of Ribunuclease A Catalysis adapted from Schultz *et al.* (1998)



A: Transphosphorylation

B: Hydrolysis

The catalysis is a two step reaction, firstly a fast transphosphorylation reaction forming a pentavalent phorphorus intermediate, then a separate slower hydrolysis reaction (**Figure 1.19**). The use of various single site mutants have been used to elucidate the mechanism and active site residues involved in the catalytic mechanism. For example, replacement of either His12 or His119 with Ala by Thompson *et al.* (1994) produced little change in the structure of ribonuclease, but had a dramatic decrease in the catalytic activity of the enzyme.

1.4.2 Other Important Active Site Residues

Asp121 lies in close proximity to His119 and was shown by Schultz *et al.* (1998) to form a catalytic dyad with His119 in a similar manner to the catalytic triad of serine proteases, aligning His119 with the protonated nitrogen in the correct position for catalysis. Replacement of Asp121 with Asn or Ala by site directed mutagenesis did not affect the crystal structure of ribonuclease but did have a 10-100 fold decrease in the catalysis of cleavage of poly-(cytidylic acid), a synthetic ribonuclease substrate.

Phe120 has an interesting role in catalysis. Mutagenesis of this residue to a number of other amino acids such as Gly, Ala and Trp had very little effect on rates of catalysis, despite its close proximity to the active site (Tanimizu *et al. (*1998) and Chatani *et al.* (2001)). However, a mutant containing an ester bond as a non-amide bond replacement was reported by Raines (2004) to cause a 1 x 10^4 decrease in catalytic activity suggesting that the backbone amide from Phe120 has an important role in ribonuclease catalysis.

Lys41 is another residue known to be important at the active site of ribunuclease. Work by Murdock *et al.* (1966) using chemical modification of lysine had suggested that this residue was important, and mutagenesis experiments performed by Trautwein *et al.* 1991, replacing Lys with Arg confirmed this, showing with hydrolysis of cytidine 2´,3´-cyclic phosphate that the mutant had only 2% of the native ribonuclease activity.

A number of other residues, although not at the active site of the enzyme, have been found through a series of mutagenesis experiments to be essential for catalytic activity. Tarragona-Fiol *et al.* (1993) confirmed the importance of residues Asn71 and Glul11, through site-directed mutagenesis experiments replacing with Ala resulting in a 40-fold reduction in the catalytic activity for the hydrolysis of dinucleotides. The review by Nogues *et al.* (1995) details these and other important non-active site residues.

1.4.3 Anfinsen's Folding Experiments

Ribonuclease A was the enzyme used in Anfinsen's famous protein refolding experiments (Anfinsen *et al.* (1960) and Anfinsen *et al.* (1961)). He showed that after protein denaturing and reduction of the disulfides with urea and β-mercaptoethanol, the denatured protein, which lacked any enzymatic activity, would slowly gain full catalytic activity after dialysis to remove the urea and β-mercaptoethanol. Denatured protein, reoxidized in the presence of urea, would only gain 1% of its full catalytic activity once the urea was removed. This showed that under denaturing conditions the four disulfides form randomly, with only one combination having enzymatic activity. Addition of a small quantity of β-mercaptoethanol catalysed the rearrangement of the disulfides, with full catalytic activity restored after about 10 hours.

One reason RNase A was chosen for this study is its spontaneous oxidation to form the active conformation. If the unnatural bonds incorporated into ribonuclease do not significantly affect its ability to fold, then the correct disulfides should form under oxidizing conditions to produce active protein.

1.5 Molecular Modelling of Proteins

Molecular dynamics (MD) is the term used to describe the modelling of molecules by computer simulation over time. These simulations complement the structural experimental data obtained from NMR and X-ray crystal structures, and have aided the development of a greater understanding of molecular structure, particularly of large biomolecules.

The first molecular dynamics simulation of a protein was performed by McCammon *et al.* (1977) on the bovine pancreatic trypsin inhibitor (BPTI), which is a small globular protein whose crystal structure was first solved by Deisenhofer (1975). The results of this simulation and other similar studies revealed that proteins do not have

a rigid structure, but instead are a dynamic system of interconnected atoms that can exhibit complex movement.

MD simulations of proteins have been used in a number of different applications. Simulations can mimic laboratory experimental conditions, generating data for the movement of a protein in solution over a period of time. This can be used to study a protein's structure, or used to predict experimental properties, for example thermodynamic properties of a system. The study of processes such as ligand docking and binding to a protein have been studied using a process called simulated annealing, which simulates a decrease in temperature from a high temperature to a low temperature, reducing the kinetic energy of the system and resulting in a static structure. This can be used to compare to known crystal structures of bound ligands and to allow predictions to be made for other ligand binding. MD can be used for the limited study of dynamic processes such as protein folding, however with current algorithms and computing capabilities it is only just possible to simulate compete folding of some simple proteins. Recently, increases in the length of simulation time due to increased processing speeds have allowed dynamic studies towards ms timescales. Piana *et al.* (2011) designed a variant of a fast-folding β-sheet containing protein and have modelled reversible folding and unfolding of the protein over a 600μs simulation, starting from an extended confirmation. Their results closely matched experimental folding times, and allowed them to make accurate predictions about the folding times of a triple amino acid mutant of their protein.

1.5.1 Molecular Dynamics Simulations of RNase A

MD simulations of RNase A have been performed by a number of different investigators, in studies of RNase structure, disulfide formation, protein folding and enzyme catalysis. Santoro *et al.*, *(*1993) used MD simulations to generate eight different structures of RNase A. They compared the root mean square deviation (RMSD) of these eight structures to an X-ray crystal structure of the protein and found close structural similarity.

A dynamics study by Merkley *et al.* (2008) looked at the change in conformation of RNase A at different temperatures by performing six simulations of RNase A over a range of temperatures from 278 K up to 340 K. They used RMSD of the Cα atoms to allow comparison of the structures as a function of temperature. There is thermal transition known to occur in RNase A at approximately 320 K, and the results of their dynamics simulations showed the presence of a conformational transition between 310 and 320 K.

An investigation by Formoso *et al.* (2010) used molecular dynamics simulations to study the dynamics, relaxation and solvation of RNase A at different stages of the transesterification and cleavage reaction that occurs during enzyme catalysis. The MD simulations were run on nanosecond timescales, and their results have provided interesting insight into the catalysis mechanism. Their results suggest that Lys41 rather than His12 could be the acid/base catalyst of the transphosphorylation and hydrolysis reactions, which presents further scope for research into the mechanism of RNase A catalysis. This recent research demonstrates a role that computer simulations are taking in developing our understanding of protein structure, and enzyme mechanisms.

1.6 Aim and Objectives

The aim of this study was to identify and develop a methodology for synthesis and sequential chemoselective ligation of peptide fragments, facilitating the total chemical synthesis of peptides and proteins that cannot easily be synthesised as single peptides by solid phase peptide chemistry.

Ribonuclease A was chosen as the example protein for the following reasons.

1. It is a small protein which formed the focus of early research into the structure/function relationship of proteins, and therefore has been well characterised.
2. The native protein has full catalytic activity after denaturing and refolding.
3. At 124 amino acids, it is beyond the length that a peptide can reliably be synthesised by solid phase peptide synthesis in a single chain.

Specific objectives were to:

1. Identify and develop a ligation strategy that enables the synthesis of peptide fragments by solid phase peptide synthesis, in high yield for ligation
2. optimise the conditions for resin loading, fragment synthesis and ligation reactions,
3. synthesise biologically active Ribonuclease A by sequential chemoselective ligation, and
4. use molecular modelling to model the chemoselective bonds and predict the implications of these non-amide bonds to protein structure and function.

If sequential non-amide ligation is going to become a viable technique for the assembly of large synthetic proteins, development of a methodology for sequential ligation and detailed investigations into the conformation of these ligation bonds and their effect on protein structure and activity are necessary.

Chapter 2: Synthesis of Amino Acid Derivatives for use in the Preparation of C-terminal Peptide Aldehydes

2.0 Introduction

Chemoselective ligation by an oxime or hydrazide bond has a requirement for an aldehyde at the N or C terminus of one of the peptides for ligation, and initial studies for this thesis were directed towards reactions between peptides containing a C-terminal aldehyde and an N-terminal amino-oxy or hydrazide. This work was an extension of the studies initiated by Sharma *et al.* (1979) and improved by Broadbridge (1998) which used C-terminal 1, 2-diols as precursors for the synthesis of C-terminal peptide aldehydes. These C-terminal diols were used as masked aldehydes in sequential chemoselective ligation (Broadbridge *et al.* (2001)), but limitations with the method lead to the proposal for this thesis of using a 1, 2-amino alcohol in place of a 1, 2-diol.

An amino acid derivative bearing either a 1,2-diol or 1, 2-amino alcohol functionality in place of the carboxylic acid is a desirable building block for the synthesis of C-terminal peptide aldehydes, as oxidation with sodium metaperiodate produces the aldehyde in high yield (Clamp *et al.* (1965)). However the only commercially available amino acid derivatives bearing either the 1, 2-diol or 1, 2-amino alcohol functionality are derivatives of the simple amino acid glycine. Therefore derivatives of any other amino acid requires a method of synthesis.

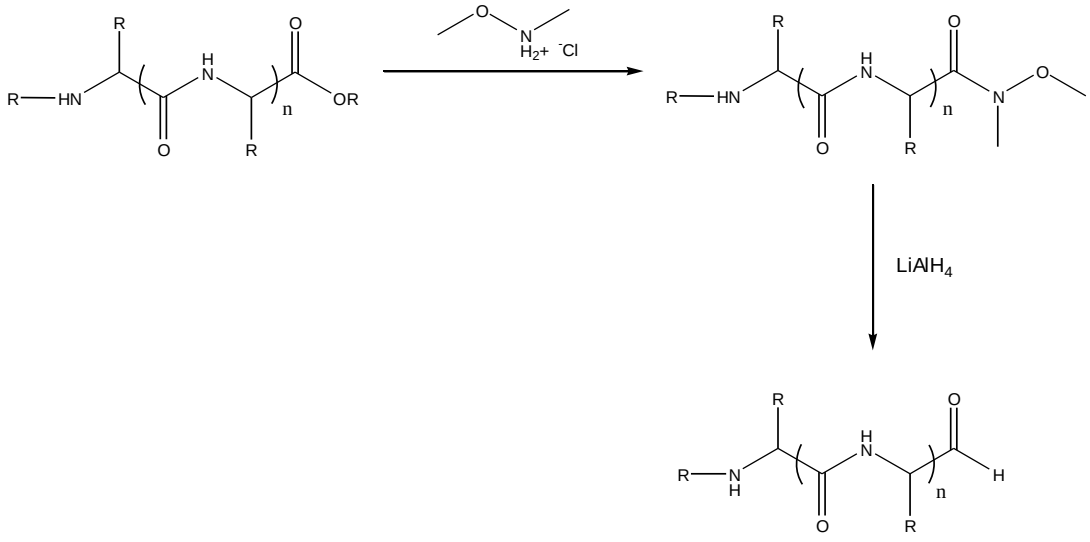
This chapter summarises the historical synthesis of 1, 2-diol derivatives of amino acids other than glycine by Sharma *et al.* (1979) and improvements to the synthesis of these diols that were made by Broadbridge (1998) and Broadbridge *et al.* (2001). The chapter then focuses on studies into the suitability of 1, 2-amino alcohols as replacement for 1, 2-diols, and development of a synthetic route to 1, 2-amino alcohol derivatives of amino acids. Finally this chapter details the use of these derivatives in the synthesis of peptides containing C-terminal 1, 2-amino alcohols,

and oxidation of the amino alcohol to produce C-terminal peptide aldehydes, with the aim of preparing several peptide fragments each containing a C-terminal 1, 2-amino alcohol as a masked aldehyde for use in sequential chemoselective ligation.

2.1 Synthesis of C-terminal Peptide Aldehydes

The generation of long peptide fragments bearing a C-terminal aldehyde can be problematic. Methods currently used include the reduction of a peptide-Weinreb amide (Fehrentz *et al.* (1995), **Figure 2.1**), (the Weinreb amide was first described by Nahm and Weinreb (1981)) or the attachment of an Fmoc amino aldehyde to a threonine resin via an oxazolidine ring (Ede *et al.* (2000)), followed by standard Fmoc peptide synthesis, and a two stage deprotection of amino acid side chains and cleavage from the solid support. Both of these methods have only been applicable to the formation of short peptide aldehydes.

Figure 2.1 Reduction of Weinreb Amide



With the Weinreb amide resin the number of equivalents of reductant required to cleave the peptide aldehyde from the solid support increases with every amide bond, which makes this method unsuitable for the synthesis of long peptide aldehydes due to increasing amounts of side reactions when large excesses of reductant are used [5].

Using the threonine oxazolidine linker, acylation of the oxazolidine ring resin during peptide synthesis often resulted in very low yields for this method, even for short peptide sequences, leading to protection of the oxazolidine nitrogen with Boc to prevent acylation (Al-Gharabli *et al.* (2006)). Resins preloaded with Fmoc amino aldehydes and Boc protected on the oxazolidine ring are now available for some amino acids (amino aldehydes attached to H-Thr-Gly-NovaSyn® TG resin, Novabiochem) but significant side reactions during the two stage deprotection and cleavage is reportedly responsible for low yields of peptides (Tanaka *et al.* (2007)).

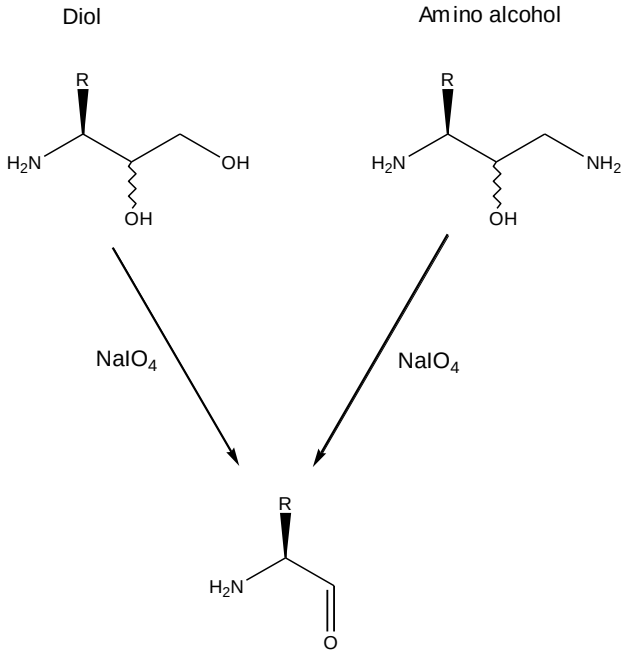
Recent improvements to the oxazolidine linker by Tanaka *et al.* (2007) are reported to have improved yields of a model 10mer peptide from 17% to 45% and to allow synthesis of a 33mer with a C-terminal leucine aldehyde at a yield of 19% which could not be synthesised on other oxazolidine linkers. This is the first example of solid phase synthesis of a long C-terminal peptide aldehyde with a residue other than glycine at the C-terminus. However this method produces aldehyde peptides directly by cleavage from the resin making it inappropriate for sequential chemoselective ligation, which requires masked aldehydes on all but one of the peptide fragments.

2.1.1 The 1,2-diol as a Masked Aldehyde

A 1, 2-diol or 1, 2-amino alcohol is a convenient aldehyde precursor (**Figure 2.2**) , as oxidation with sodium periodate proceeds cleanly to the aldehyde under mild conditions (Clamp *et al.* (1965)). Historically this reaction has been used for the generation of either N or C-terminal peptide aldehydes by the periodate oxidation of an N-terminal serine (Geoghegan et al. (1992)) or C-terminal 1-amino-2, 3-

propanediol (Zhang et al. (1998)). 1-amino-2, 3-propanediol as an aldehyde precursor can conveniently be incorporated into the peptide sequence at the start of solid phase peptide synthesis, but inserts a spacer between the peptide and the aldehyde. The synthesis of diol derivatives of other amino acids allows the synthesis of C-terminal peptide aldehydes bearing the amino acid side chain at the C-terminus.

Figure 2.2 Oxidation of a diol or amino alcohol to aldehyde

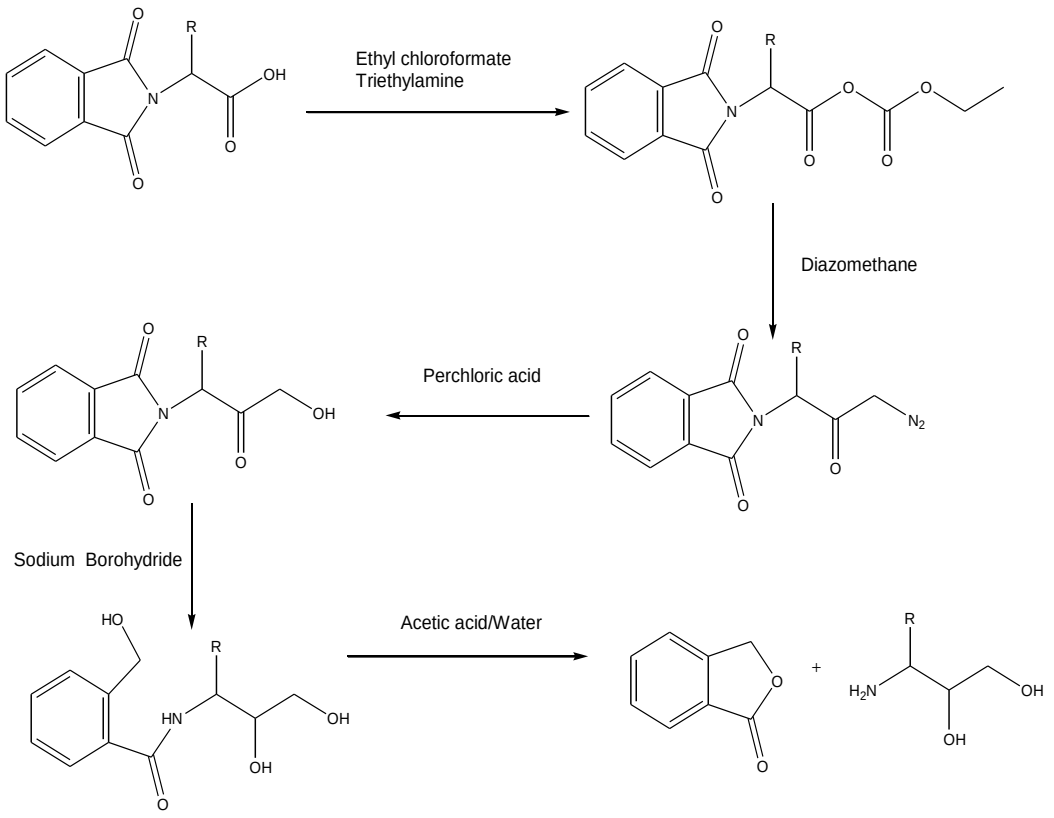


2.2 History of Diols and Improvements

The original synthesis of amino acid diols from protected amino acids was published by Sharma *et al.* (1979). This method converted a Boc protected amino acid to the diol via acid decomposition of a diazomethyl ketone to a trifluoroacetoxy ketone and reduction with borohydride to the diol. The main limitation with this synthesis was production of a cyclic impurity on treatment with trifluoroacetic acid, which accounted for a loss of over half the yield of product. It was suggested that this may

have been due to involvement of the Boc protected amine, and so this method was improved upon by Broadbridge (1998), by changing the amino protection to phthalyl which fully protected the amine, and using aqueous perchloric acid rather than trifluoroacetic acid for the acid decomposition to form an α-hydroxyketone which on reduction with borohydride resulted in a much improved yield of the amino 1, 2-diol ((Broadbridge (1998) (**Figure 2.3**)).

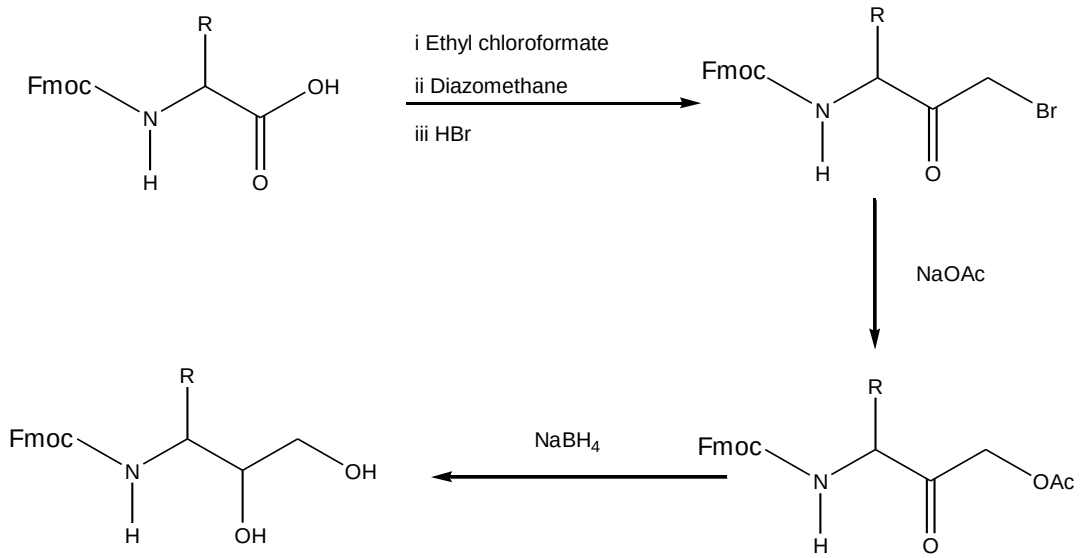
Figure 2.3 Synthesis of amino acid diols from phthalyl protected amino acids (adapted from Broadbridge (1998))



Further improvements were made by Broadbridge and Winsor *et al.* (2001) to the diol synthesis. Instead of phthalyl protection and acid decomposition of the diazoketone with perchloric acid, the Fmoc protected amino acid was converted to the bromomethyl ketone via HBr treatment of the diazoketone. This was reacted with

sodium acetate, and then sodium borohydride to produce the Fmoc protected diol (**Figure 2.4**). This route allowed synthesis of diols directly from Fmoc protected amino acids, with their corresponding side chain protections for Fmoc chemistry (Broadbridge *et al.* (2001).

Figure 2.4: Synthesis of diols directly from Fmoc protected amino acids, Broadbridge *et al.* (2001)



2.2.1 Use of C-terminal Diols as Masked Aldehydes in Sequential Chemoselective Ligation

This improved method of synthesis for diols was used for the synthesis of a peptide based on the C-terminus of a cytokine, that had proven difficult to synthesise as a single peptide (Broadbridge *et al.* 2001). The peptide sequence was split into four peptide fragments, which were sequentially ligated using chemoselective ligation by oxime bonds.

In sequential ligation a temporary protection is used to mask one of the chemoselective groups during the first ligation, which is then removed to allow ligation of a third peptide fragment. The diol was used as a masked aldehyde for this purpose. Two of the four peptide fragments required just a single reactive group at either their N or C-terminus, a C-terminal aldehyde on one and an N-terminal hydroxylamine on the other. The two central fragments required both an N-terminal hydroxylamine and a C-terminal masked aldehyde.

Leucine diol and Valine diol were synthesised via the method outlined in **Figure 2.4** and attached to chlorotrityl resin, then each of the three peptides requiring C-terminal aldehydes were synthesised on the chlorotrityl resin and purified. The fourth peptide required an amide at the C-terminus and so was synthesised on Rink amide resin. The hydroxylamine was incorporated onto the N-terminus of the three appropriate peptides using bis-Boc protected amino-oxy acetic acid. Bis-Boc protection prevents double acylation of the hydroxylamine that can occur when using the mono-Boc protected reagent. The scheme detailing the sequential ligation of the peptide fragments is shown in **Figure 2.5** and MS data for the ligated peptide is shown in **Figure 2.6**.

Figure 2.5: Sequential chemoselective ligation of four peptides by oxime bond (Broadbridge *et al.* 2001)

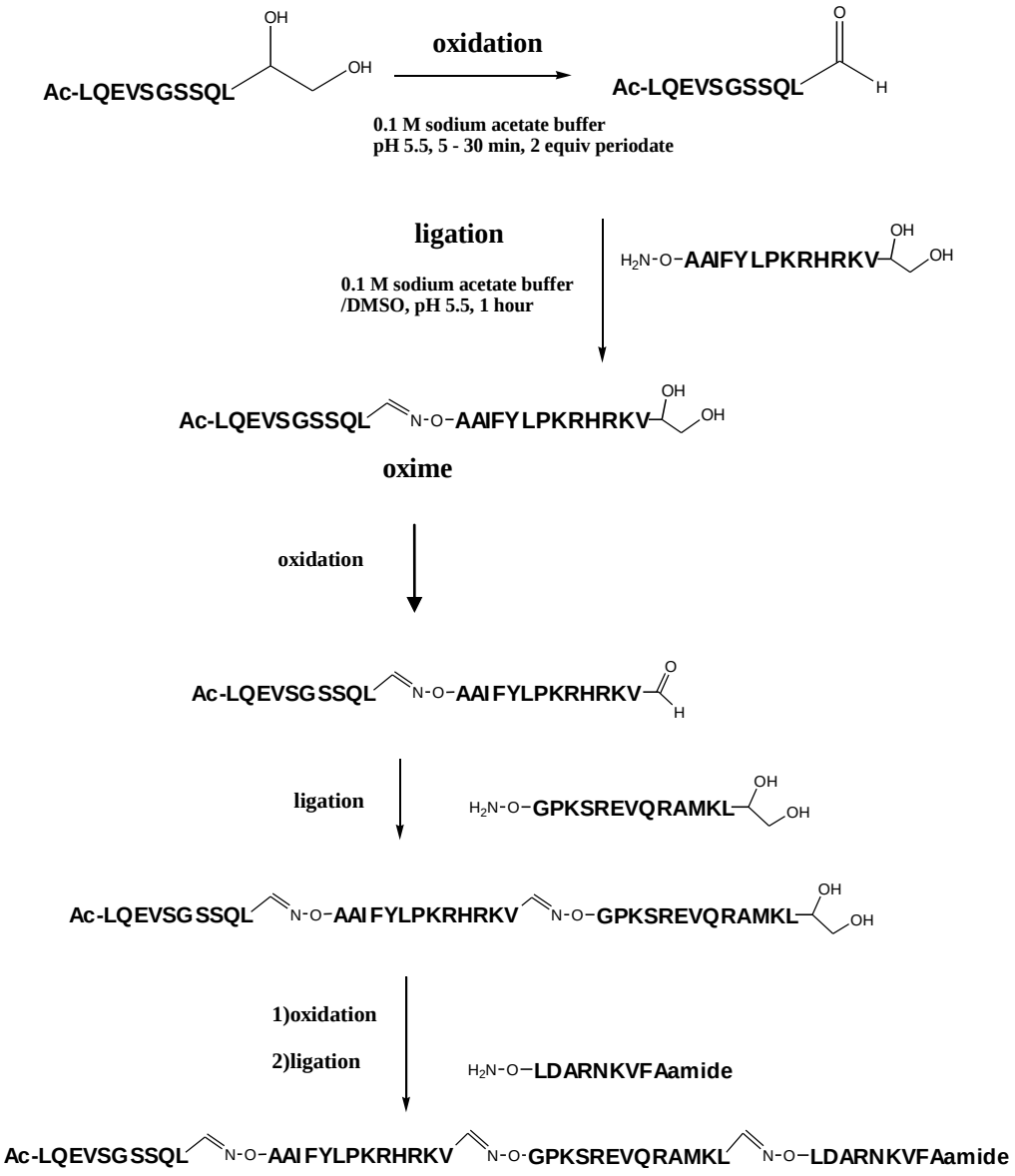
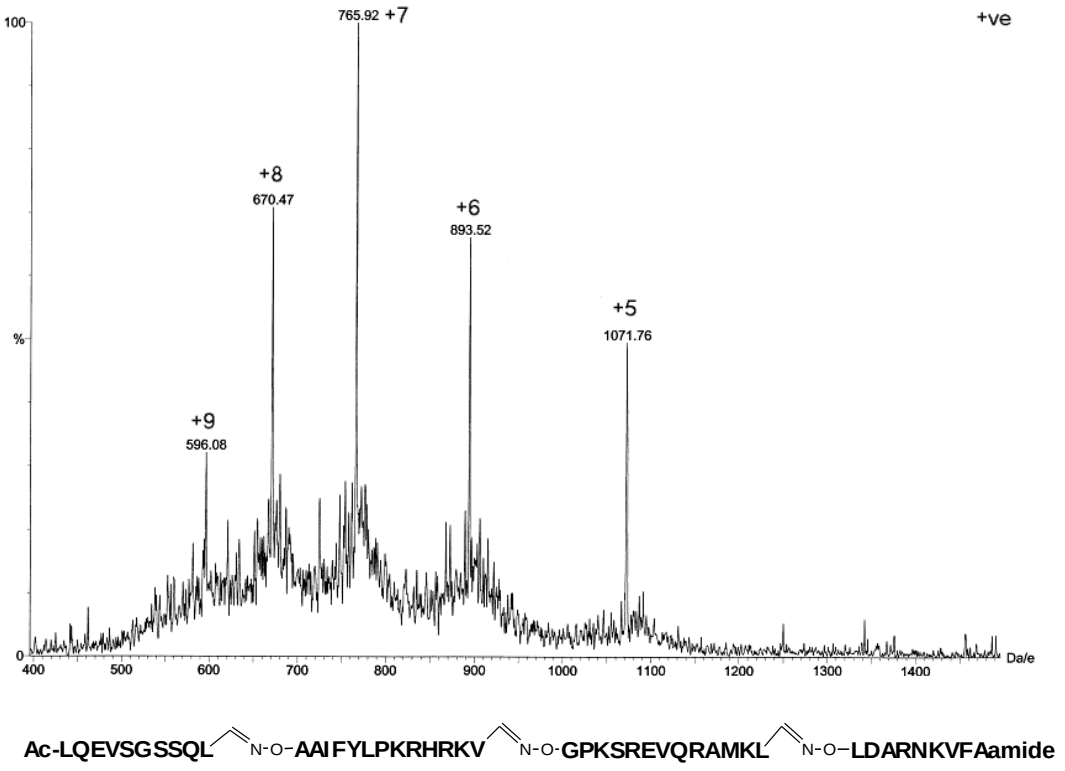


Figure 2.6: ESMS of Peptide formed by oxime sequential chemoselective ligation



[M +H] 5363

2.3 Amino Alcohols – New Synthetic Route and Method Development

A major restraint with the diol system was the low level of substitution achieved with chlorotrityl resin with any derivative other than glycine-diol (Broadbridge *et al.* 2001). This contributed to very low yields of product. An amine is significantly more reactive to trityl chloride than an alcohol, and the 1, 2-amino alcohol functionality has often been used at the N-terminus of a peptide as an aldehyde precursor (incorporated as an N-terminal serine) leading to this investigation of the 1, 2-amino alcohol as a replacement for the diol at the C-terminus of the peptide.

2.3.1 1, 3-amino-2-hydroxypropane (Glycine Amino Alcohol)

The potential of amino alcohols as replacements for diols for the synthesis of C-terminal peptide aldehydes, and their application in sequential chemoselective ligation was investigated using the commercially available 1, 2- amino alcohol derivative of glycine, 1, 3-amino-2-hydroxypropane. 1, 3-amino-2-hydroxypropane was attached to 2 chlorotrityl chloride resin. The amino functionality is significantly more reactive than the hydroxyl, so it was assumed that attachment to the resin would be via either of the amino groups. The optimum conditions found were 2 equivalents of 1, 3-amino-2-hydroxypropane, using DMF/DCM 1:1 and one equivalent of DIPEA, for 2 hours (**Chapter 7, Section 7.2.7.1**).

The 1, 3-amino-2-hydroxypropane substituted resin was then used in the synthesis of peptides containing a C-terminal glycine-amino alcohol. After HPLC purification, peptides were oxidised using $NaIO_4$ (**Chapter 7, Section 7.2.10**) to produce C-terminal peptide aldehydes. Following the successful development of this method it was adopted as a standard procedure for the custom synthesis of peptides containing C-terminal glycine aldehydes at Peptide Protein Research Ltd. Although detailed studies comparing yields of peptides synthesised using 1, 3-amino-2-hydroxypropane (glycine amino alcohol) and 1-amino-2, 3-propanediol (glycine diol) were not performed, yields of long peptide glycine aldehydes were improved with use of the amino alcohol, attributed to the increased stability of the amine attached to trityl linker compared to the alcohol, during peptide synthesis.

2.3.2 A New Synthetic Route to 1-amino, 2-alcohol Derivatives of Amino Acids

Once it had been demonstrated that the amino alcohol was a suitable replacement for the diol, a synthetic route for the synthesis of amino alcohol derivatives of amino acids had to be devised, as the glycine derivative was the only amino acid

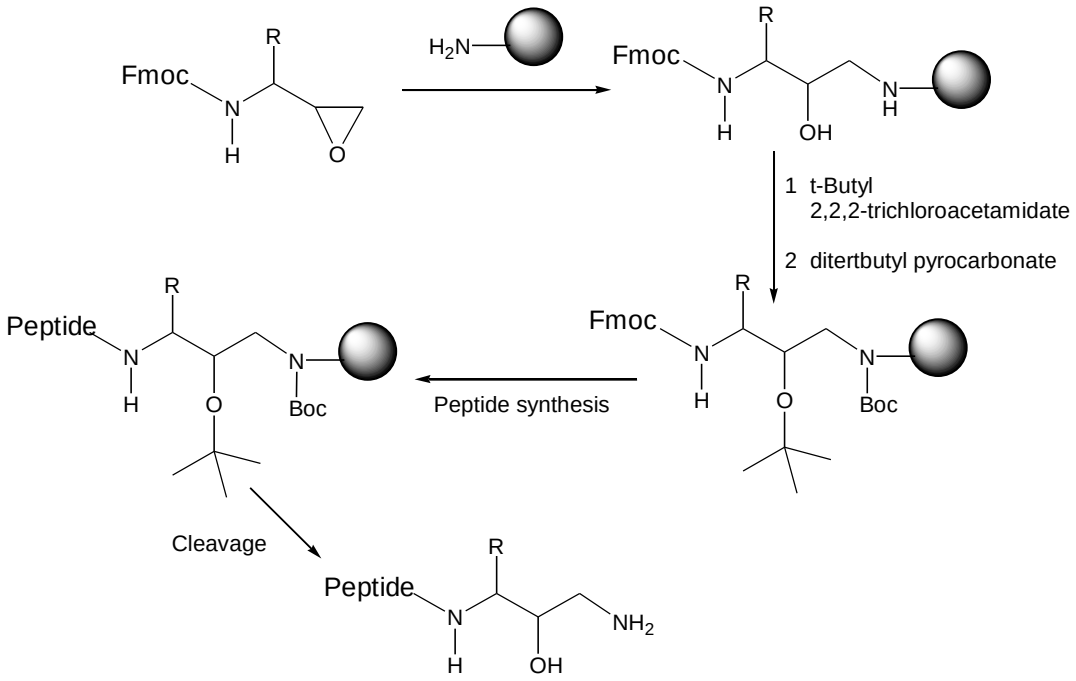
commercially available. A route was prefered that would convert an Fmoc protected amino acid with appropriate side chain protection directly to the amino alcohol, without the need to change the amine protecting group. It was important that any method developed for use in peptide fragment preparation for chemoselective ligation was cost effective, which would allow them to be used in commercial applications at Peptide Protein Research Ltd. Fmoc protected amino acids (**Chapter 7, Table 7.2**) are inexpensive due to their use in standard Fmoc peptide synthesis, whereas amino acids with Fmoc compatible side chain protecting groups but alternative amine protection are vastly more expensive.

2.3.2.1 Epoxide Derivatives of Fmoc Amino Acids

It was known that an epoxide derivative of an N-α–Boc protected amino acid could be reacted with an amine to give the amino alcohol functionality. Opening of the epoxide with either ammonia, or a primary amine is not compatible with an N-α–Fmoc protected epoxide, as the amine is sufficiently basic to remove the Fmoc protecting group under the refluxing conditions of the reaction. However, several papers reported the opening of epoxides with amines under mild conditions (room temperature reactions in solvents such as DCM), using a metal salt as catalyst. Originally lithium perchlorate was used (Chili *et al.* (1990)), but this was required in stoichiometric (100% mol) quantities, and perchlorates can undergo dangerous dcomposition reactions in the presence of organic molecules, so safer alternatives were investigated. Lithium or lanthanide III trifluoromethanesulfonates (triflates), which were safer, more efficient catalysts, were shown to be required in 10-50% mol quantities, and required only short reaction times at room temperature (Chini *et al.* (1994)). Therefore for this study it was attempted to react Fmoc protected epoxides with amines using the catalyst lithium triflate, which was the cheapest of these catalysts. Particularly of interest was the reaction of the epoxide with an amine functionalised solid support, for example Rink amide resin, which is employed as a standard linker in SPPS for the synthesis of C-terminal amides, and therefore could

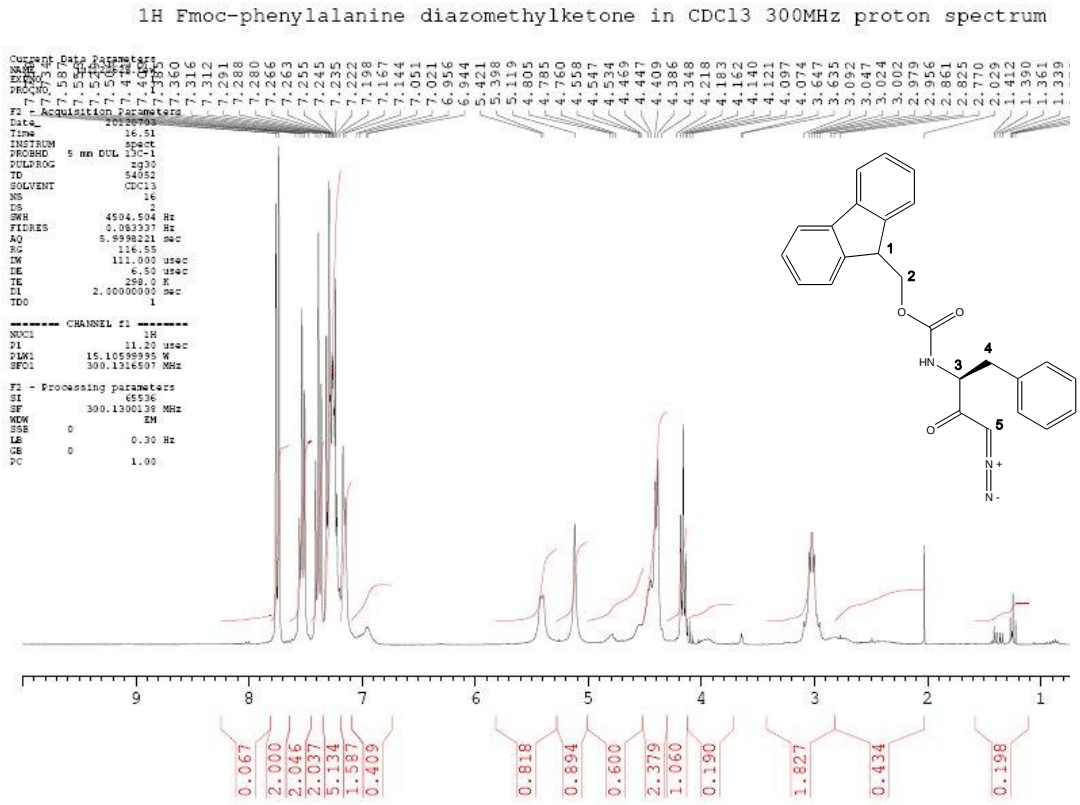
be used to yield the amino alcohol directly upon cleavage of the peptide from the resin (**Figure 2.7**).

Figure 2.7: Putative synthetic strategy for attachment of epoxide to Rink amide resin



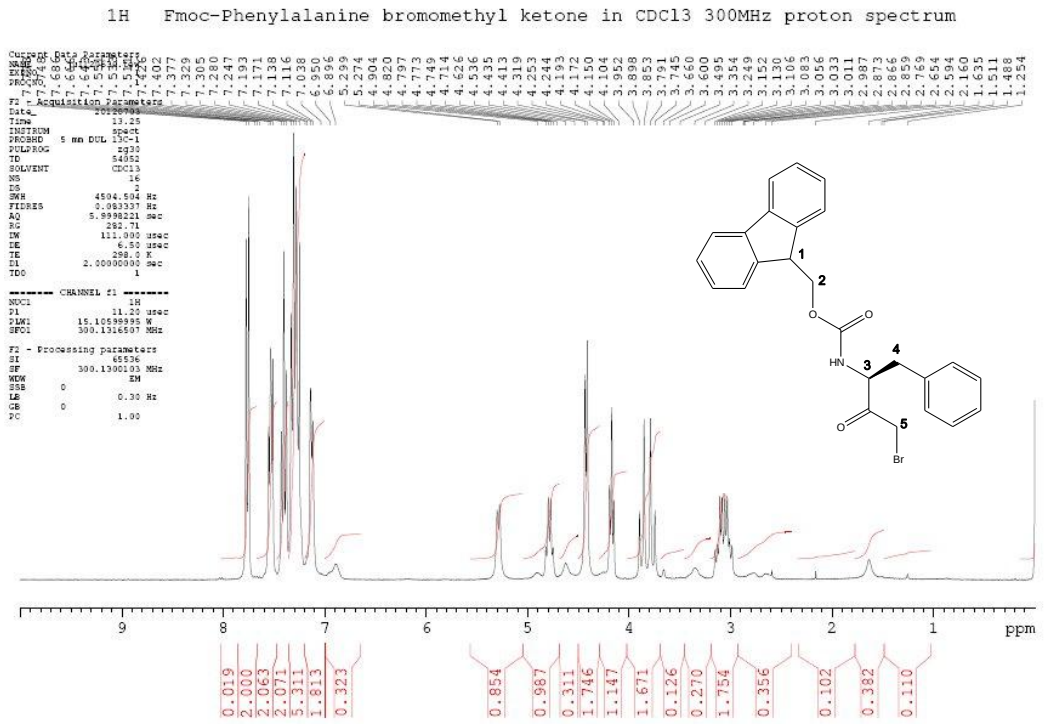
The epoxide of Fmoc-phenylalanine was synthesised as described in **Chapter 7, Section 7.2.4**. Fmoc-phenylalanine was activated using ethyl chloroformate and reacted with diazomethane to generate the diazomethylketone (**Chapter 7, Section 7.2.4.1**). This was purified by column chromatography and analysed by IR, ^{1}H NMR (**Figure 2.8**) and ^{13}C NMR and CHN analysis.

Figure 2.8: ^{1}H NMR of Fmoc-phenylalanine diazomethylketone



The ^{1}H NMR spectrum was assigned with the aid of the COSY spectrum. The peaks at 7.1-7.8 were assigned as 13 aromatic protons. The peak at 5.4 was identified as the NH proton, and showed correlation with the peak at approximately 4.5, part of a multiplet signal. This was therefore assigned as the CH proton at position **3**. This peak in turn correlated to the peak at 3.0, assigned as the two CH2 protons at position **4**. The singlet peak at 5.1 showed no correlation to other peaks and was assigned as the CH proton in position **5**. The peak at 4.2 showed correlation to the double integral peak at 4.4, and so were assigned as the CH in position **1** and the CH2 in position **2** respectively. The ^{1}H NMR spectrum was therefore considered to be consistent with the diazomethylketone.

Figure 2.9: ^{1}H NMR of Fmoc-phenylalanine bromomethylketone



The ^{1}H NMR spectrum was assigned with the aid of the COSY spectrum. The peaks at 7.1-7.8 were assigned as 13 aromatic protons. The peak at 5.3 was identified as the NH proton, and showed correlation with the peak at 4.8, which was therefore assigned as the CH proton at position **3**. This peak in turn correlated to the double integral peak at 3.1, assigned as the two CH2 protons at position **4**. The CH2 peak at 3.9 showed no correlation to other peaks and gave a coupling constant of J ~12Hz consistent with only geminal coupling, and so was assigned to the two CH2 protons at position **5**. The peak at 4.2 showed correlation to the double integral peak at 4.4, and so were assigned as the CH in position **1** and the CH2 in position **2** respectively. The ^{1}H NMR spectrum was therefore considered to be consistent with the bromomethylketone.

Fmoc-phenylalanine diazomethylketone was reacted with hydrogen bromide to form the bromomethylketone (**Chapter 7, Section 7.2.4.2**). The bromomethylketone was not further purified, and was analysed by IR, ^{1}H NMR (**Figure 2.9**) and ^{13}C NMR and CHN analysis.

Fmoc-phenylalanine epoxide was reacted with a number of different primary amines, and Rink amide resin, using lithium triflate as a catalyst, however even the mild reaction conditions of room temperature reactions in DCM resulted in the removal of the Fmoc protecting group by the primary amine, making this an unsuitable route of synthesis for the amino alcohol directly from Fmoc protected amino acids. It was decided that rather than using Boc chemistry and the epoxide, that an alternative route to the 1-amino, 2-alcohol derivative directly from the Fmoc protected amino acid would be attempted.

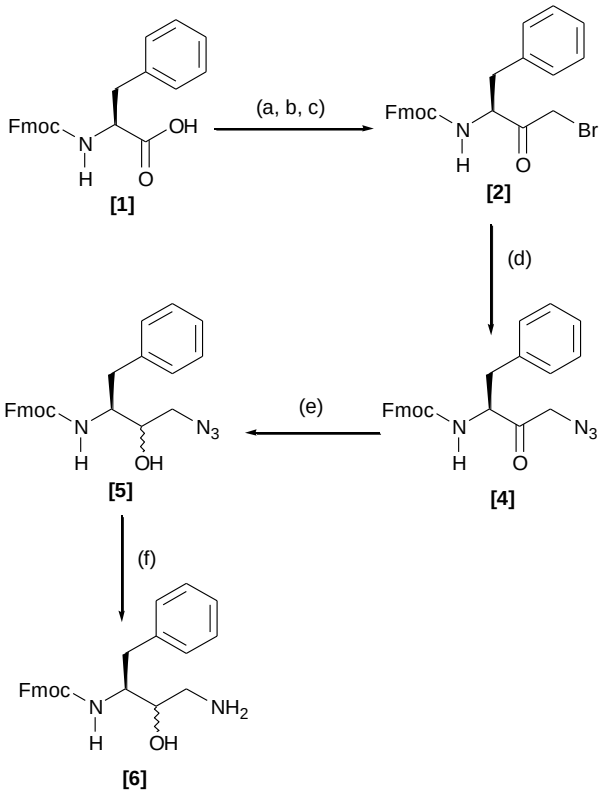
2.3.2.2 Amino Alcohol Derivatives Directly from Fmoc-protected Amino Acids

Returning to the synthetic route for the synthesis of diols directly from Fmoc protected amino acids, a strategy was developed that converted the bromomethyl ketone to an azido methyl ketone, which on reduction of the ketone and hydrogenation of the azide would yield an amino alcohol (**Figure 2.10**). Fmoc phenylalanine was the first amino acid chosen to test the synthetic strategy (**Chapter 7 Section 7.2.5**).

The bromomethylketone derivative of phenylalanine (**Figure 2.10, [2])** was synthesised as before, and this was converted to a azidomethylketone (**Figure 2.10, [4])** by reaction with sodium azide (**Chapter 7, Section 7.2.5.1**). The bromomethyl ketone was very hydrophobic so would not dissolve in polar solvents, and sodium azide was only soluble in aqueous or aqueous miscible solvents, therefore a number

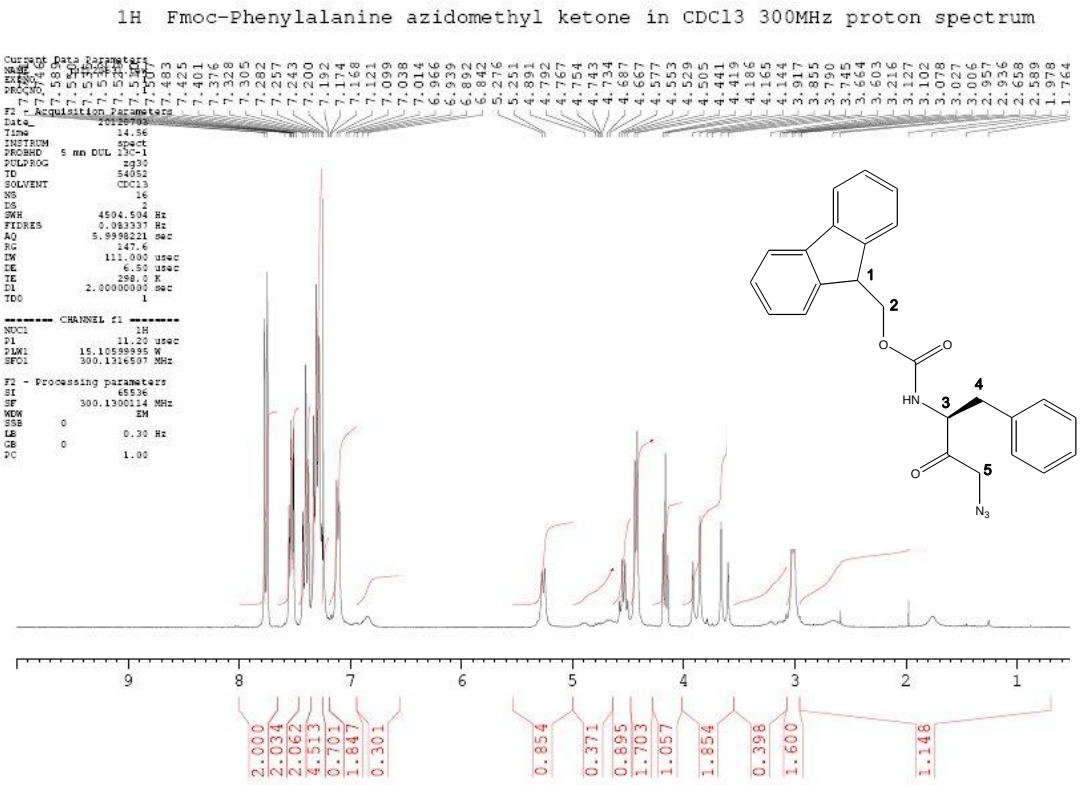
of different solvent systems were tried. The most successful solvent system found was a biphasic system, having the bromomethyl ketone dissolved in ethyl acetate and vigorously stirring with sodium azide dissolved in water, which gave a yield of the azidomethylketone of greater than 90% for all amino acids investigated (**Chapter 7, Section 7.2.5.1 [4], 7.2.5.4.3 [9], 7.2.5.4.8 [14] and 7.2.5.4.13 [19]**). Fmoc-phenylalanine azidomethylketone was analysed by IR, ^{1}H NMR (**Figure 2.11**) and ^{13}C NMR and CHN analysis.

Figure 2.10: Synthesis of 1, 2 amino alcohol derivative of Fmoc phenylalanine



Reagents and conditions: (a) EtOCOCl, TEA (2:2 equiv.), THF, 4h; (b) CH_2N_2 Et_2O 3 h; (c) HBr (48% aq, 1 equiv.) acetone 5 min; (d) NaN_3 (10 equiv), EtOAc/H_2O 24h; (e) $NaBH_4$ (1.4 equiv.) MeOH 10 min; (f) Ammonium formate (4 equiv.) Pd-C 10%, MeOH, 60°C, 30 min.

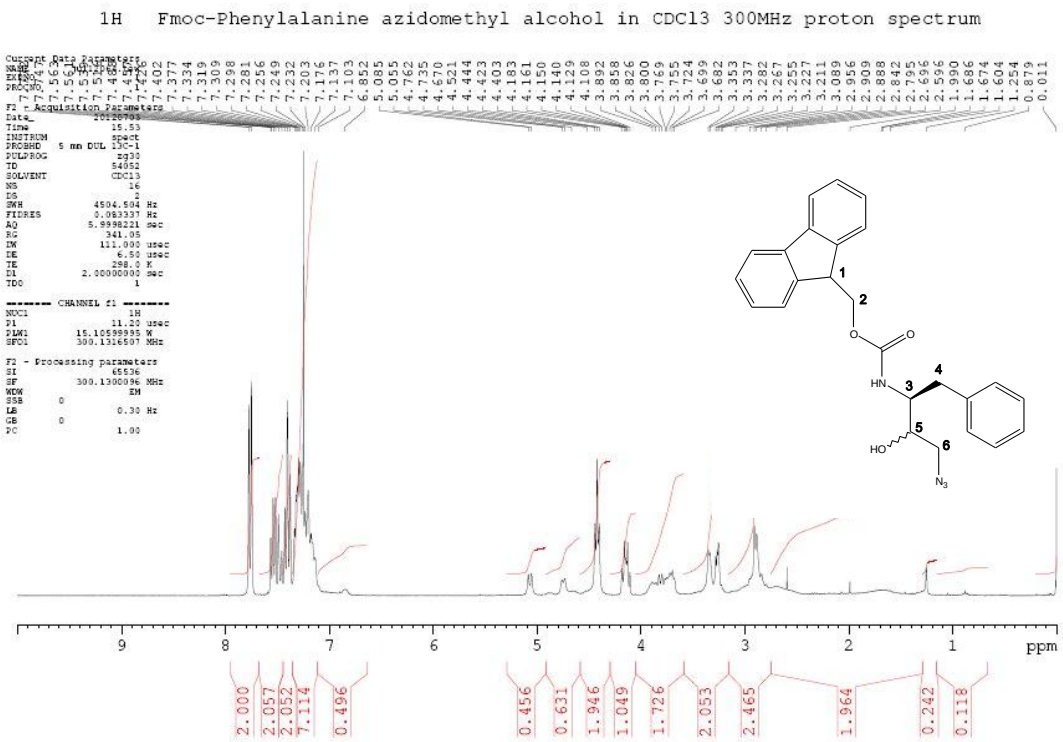
Figure 2.11: ^{1}H NMR of Fmoc-phenylalanine azidomethylketone



The ^{1}H NMR spectrum was assigned with the aid of the COSY spectrum. The peaks at 7.1-7.8 were assigned as 13 aromatic protons. The peak at 5.4 was identified as the NH proton, and showed correlation with the peak at 4.5, which was therefore assigned as the CH proton at position **3**. This peak in turn correlated to the double integral peak at 3.0, assigned as the two CH2 protons at position **4**. The CH2 peak at 3.75 showed no correlation to other peaks and gave a coupling constant of J ~12Hz consistent with geminal coupling, and so was assigned to the two CH2 protons at position **5**. The peak at 4.2 showed correlation to the double integral peak at 4.4, and so were assigned as the CH in position **1** and the CH2 in position **2** respectively. The ^{1}H NMR spectrum was therefore considered to be consistent with the azidomethylketone.

Reduction of the ketone to the alcohol (**Figure 2.10, [5])** was perfomed using sodium borohydride and the same conditions that were used by Broadbridge (1998) in the final step of production of Fmoc protected diols (**Figure 2.4**) (**Chapter 7, Section 7.2.5.2**). Fmoc-phenylalanine azidomethyl alcohol was analysed by LCMS, IR, ^{1}H (**Figure 2.12**) and ^{13}C NMR and CHN analysis. Reduction of the ketone produced mixed isomers at the hydroxyl. The presence of the chiral centre of the L amino acid means that two diastereomers were formed which are separable by HPLC, estimated to be approximately 60:40 (**Figure 2.13 [5]**). The stereochemistry of the hydroxyl was not important as oxidation of the amino alcohol to produce the aldehyde removes this chiral centre.

Figure 2.12: ^{1}H NMR of Fmoc-phenylalanine azidomethylalcohol

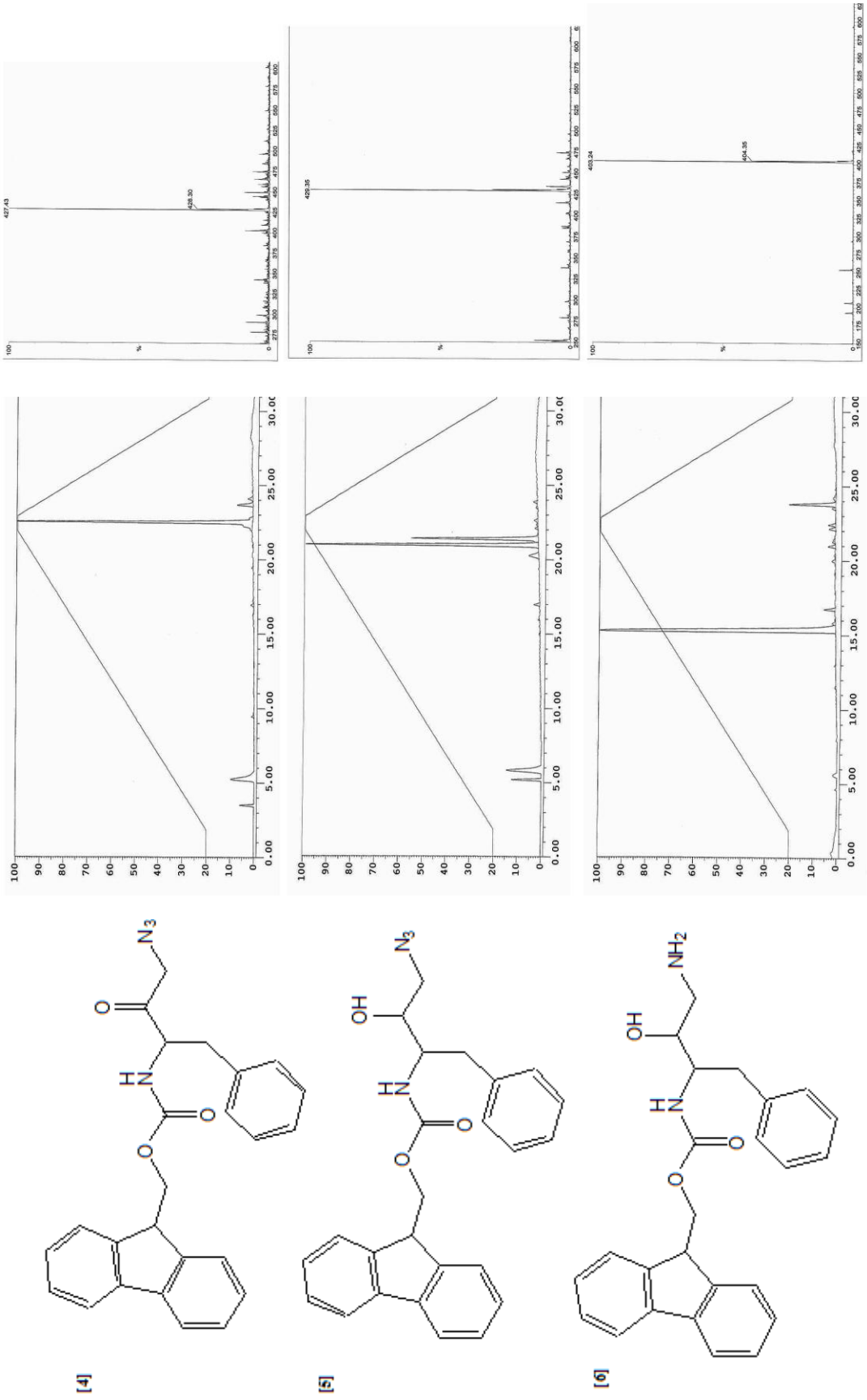


The ^{1}H NMR spectrum was assigned with the aid of the COSY spectrum. The spectra were not as clear as for the previous samples, attributed to the fact that the amino alcohol was a pair of diastereomers. The peaks at 7.1-7.8 were assigned as 13

aromatic protons. The peak at 5.1 was identified as the NH proton, and showed correlation with the peak at 3.7, which was therefore assigned as the CH proton at position **3**. This in turn correlated to the peak at 3.3, which was assigned as the two CH2 protons in position **4**. The peak at 4.2 showed correlation to the double integral peak at 4.4, and so were assigned as the CH in position **1** and the CH2 in position **2** respectively. The peak at 4.8 was assigned as the OH proton. This showed correlation to the peak at 3.9, which was assigned as the CH proton in position 5. This showed correlation to the double integral peak at 2.9, which was assigned as the CH2 peak in position 6. The ^{1}H NMR spectrum was therefore considered to be consistent with the azidomethylalcohol.

The final step of the synthesis required hydrogenation of an azide to a primary amine. Hydrogenation using hydrogen gas requires special equipment, and there is a safety consideration associated with the use of hydrogen gas. For this reason, the use of catalytic hydrogen transfer was investgated. This methodology has been thoroughly reviewed by Ram and Ehrenkaufer (1988 and references within). Of particular interest to this study was the findings of Gatriser *et al.* (1983) who employed ammonium formate as a source of hydrogen for hydrogenation of azides to amines. This method of catalytic hydrogen transfer has been used in a number of publications at room temperature. In this study, no reaction was observed for hydrogenation of the azidomethylalcohol (**Figure 2.10**, **[5])** using ammonium formate and 10% Pd-C at room temperature (approx 20 ºC). Hydrogenation of a diazide using ammonium formate was reported at 60 ºC by Glaçon *et al.* (1996) and heating the reaction to 60 ºC resulted in rapid conversion to the amine with no observable side reactions (**Chapter 7, Section 7.2.5.3**). Analysis by HPLC of the 1,2 amino alcohol (**Figure 2.13, [6])** gave the expected shift in retention time due to the increased hydrophilicity from the amine. There was no visible separation of the two diastereomers, but the peak width was broadened consistent with coelution of the diastereomers.

Figure 2.13: HPLC and MS data for intermediates and products in the synthesis of the 1, 2 amino alcohol derivative of Fmoc phenylalanine



2.3.2.3 Stability of Amino Alcohol Derivatives of Fmoc Amino Acids

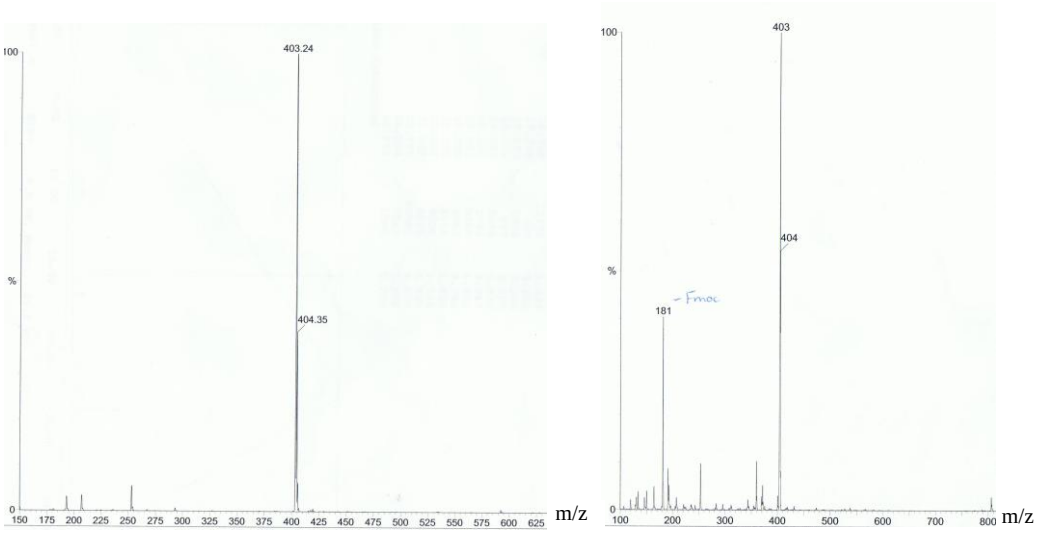
It was noticed that upon storage of the 1, 2-amino alcohol derivatives of Fmoc amino acids it became more difficult to solubilise the compound in organic solvent. Analysis by mass spectrometry revealed that the Fmoc group was not stable and was slowly being removed from the alpha amine. Mass spectrometry is a very sensitive method of detecting small quantities of Fmoc removal. The diamine degradant gives a stronger MS signal than the Fmoc protected compound as it more easily ionised, so even a very small amount of Fmoc removal can be detected. For this reason MS cannot be used to quantify the level of degradation as the peak size is not proportional to the quantity, but as the compound is a diamine with one amine protected, any significant level of Fmoc deprotection would render the derivative unsuitable for subsequent reactions, as resin loading could be to either amine.

The likely explanation is that the amine of the 1-amino 2-alcohol was sufficiently basic to remove the N-α-Fmoc group. A stability study was performed on both the amino alcohol and the azido alcohol, stored at 4º C, and analysed after 24 h, 1 week and 1 month (**Chapter 7, Section 7.2.6**). The azido alcohol showed good stability after 1 month, and had no sign of Fmoc removal by MS. The amino alcohol had detectable quantities of Fmoc removal after 24 hours, and after 1 week the quantity of Fmoc removal was considered significant (**Figure 2.14**). After 1 month there was a further increase in the proportion missing the Fmoc group.

Amino acid derivatives were therefore stored as the azido alcohol until required, then hydrogenated to the amino alcohol and immediately attached to chlorotrityl resin to avoid the problems of Fmoc removal.

Figure 2.14: ESMS of Fmoc-Phe amino alcohol immediately after synthesis (a) and after storage for 1 week at 4ºC (b)

a b



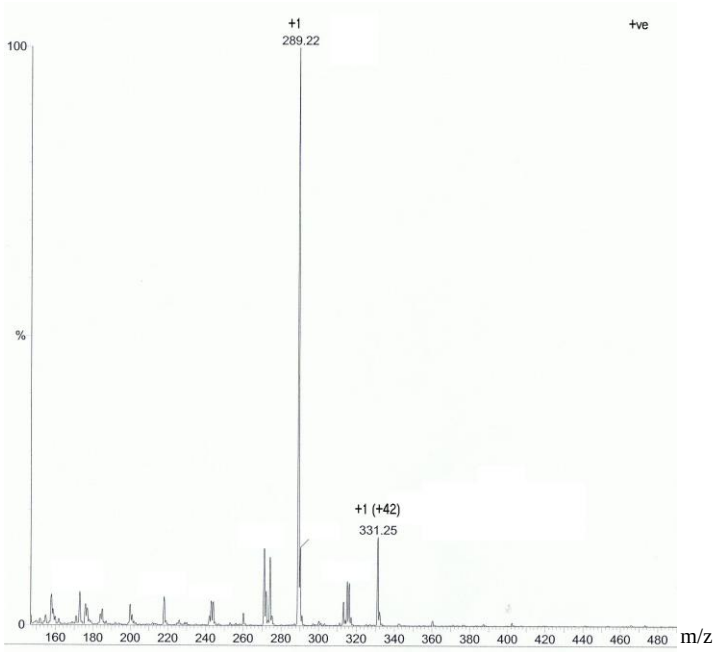
Fmoc-Phe-amino alcohol [M + H] 403 Phe amino alcohol [M + H] 181

2.4 Protection of the Alcohol during Peptide Synthesis

Following attachment of the N-α-Fmoc protected 1, 2-amino alcohol to chlorotrityl resin, protection of the alcohol as a tertiary butyl ether is required to prevent acylation during peptide synthesis. The method of protection used with glycine diol by *Zhang et al.* (1998) and in this study was *t*-butyl protection using *t*-butyl-2,2,2-trichloroacetamidate in DCM. If the hydroxyl is not protected then the steric bulk of the chlorotrityl linker prevents reaction of the hydroxyl with most reagents. However acetic anhydride, which is used to cap any unreacted amino groups in difficult coupling reactions and therefore prevents production of deletion peptides, (**Chapter 7, Section 7.2.1.4**) is a small molecule which is able to acetylate the hydroxyl. If this occurs then cleavage of the peptide does not yield a C-terminal amino alcohol, but instead the acetylated derivative, which cannot be oxidised to an aldehyde through treatment with sodium metaperiodate.

Protection of the hydroxyl of 1, 3-amino-2-hydroxypropane (glycine amino alcohol) was considered to have proceeded to completion, as use of capping steps with acetic anhydride during peptide synthesis did not produce any acetylation of the hydroxyl. However synthesis of the test peptide Ac-AAA-amino alcohol, where the N-terminus was acetylated with acetic anhydride, showed by ESMS in addition to the expected product peak, an additional peak +42 mass units which would correspond to the acetylated hydroxyl (**Figure 2.15**). Peptides were therefore synthesised without use of acetic acid capping to prevent undesirable acetylation.

Figure 2.15: HPLC of Ac-AAA-amino alcohol

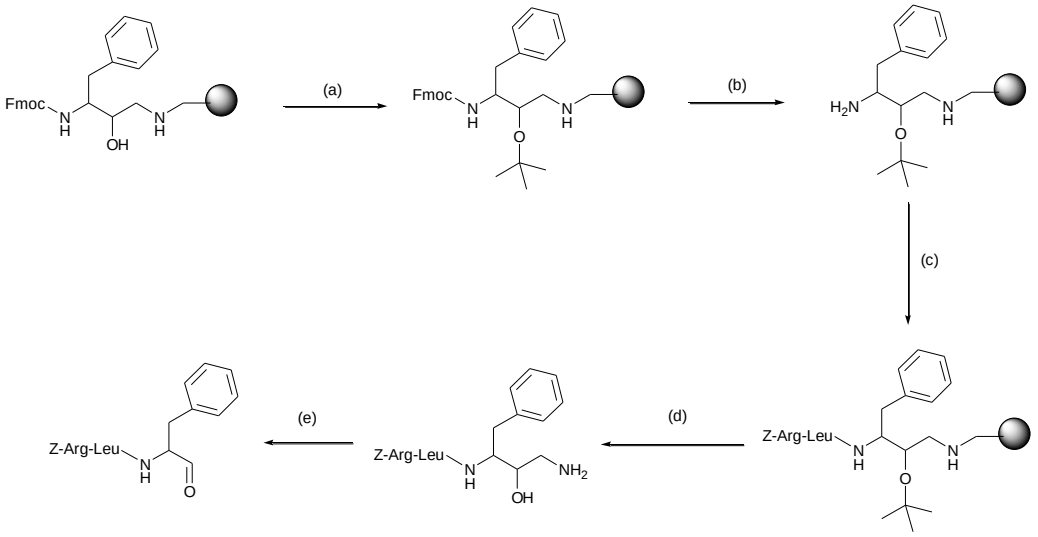


Ac-AAA-amino alcohol [M+H] 289, acetylated hydroxyl [M+H] 331

2.5 Synthesis of a Small C-terminal Peptide Aldehyde

The peptide Z-RLF-aldehyde, which is an analogue of the proteinase inhibitor chymostatin, was synthesized to demonstrate the use of the amino alcohol derivative in the synthesis of a C-terminal peptide aldehyde (**Figure 2.16**). This small peptide allowed characterisation of the aldehyde by HPLC (**Figure 2.17**), ESMS (**Figure 2.18**) and NMR (**Figure 2.19**).

Figure 2.16: Synthesis of the peptide aldehyde Z-RLF-H



Reagents and conditions: (a) *t*-butyl-2,2,2-trichloro acetamidate, DCM, 1h; (b) 20% piperidine/DMF, 20min; (c) Fmoc-Leu-OH, HBTU, HOBt (3 equiv.), DIPEA (9 equiv.) 1h; 20% piperidine/DMF, 20min; Fmoc-Arg(Pbf)-OH, HBTU, HOBt (3 equiv.), DIPEA (9 equiv.) 1h; 20% piperidine/DMF, 20min; Z- chloroformate (3 equiv.), DIPEA (9 equiv.), DMF; (d) TFA/TIS/H_2O 95:2.5:2.5 2h; (e) NaIO4, 0.1M NaOAc buffer pH5.5, 30min.

Fmoc-Phe amino alcohol was synthesized (**Chapter 7 Section 7.2.5**) and attached to 2-chlorotrityl chloride resin using similar conditions to that used for attachment of 1, 3-amino-2-hydroxypropane. The resin substitution was calculated by measuring the absorbance at 304 nm for the deprotection of the Fmoc group (**Chapter 7, Section 7.2.1.8**). Resin loading for Fmoc-Phe amino alcohol was very low compared to 1, 3-

amino-2-hydroxypropane, and after 1-2 hours no detectable loading had occurred. Reaction times of up to 24 hours were required to achieve a resin substitution of 0.1 mmol/g, which was sufficient resin substitution to allow for subsequent peptide synthesis. Extending reaction times further was not performed due to the risk of Fmoc removal from the alpha amine, which could lead to attachment of the derivative through this amine, which would insert the derivative in the wrong direction.

The peptide Z-RLF-amino alcohol was then synthesised by manual solid phase peptide chemistry **(Chapter 7, Section 7.2.7.3.1),** cleaved from the resin and purified by RP-HPLC (**Figure 2.13**). After purification, the peptide Z-RLF-amino alcohol was oxidised with sodium metaperiodate (**Chapter 7, Section 7.2.9.2**) to yield the final peptide Z-RLF-H **(Chapter 7, Section 7.2.7.3.2, Figure 2.16)**.

Figure 2.17: HPLC of Z-RLF-amino alcohol

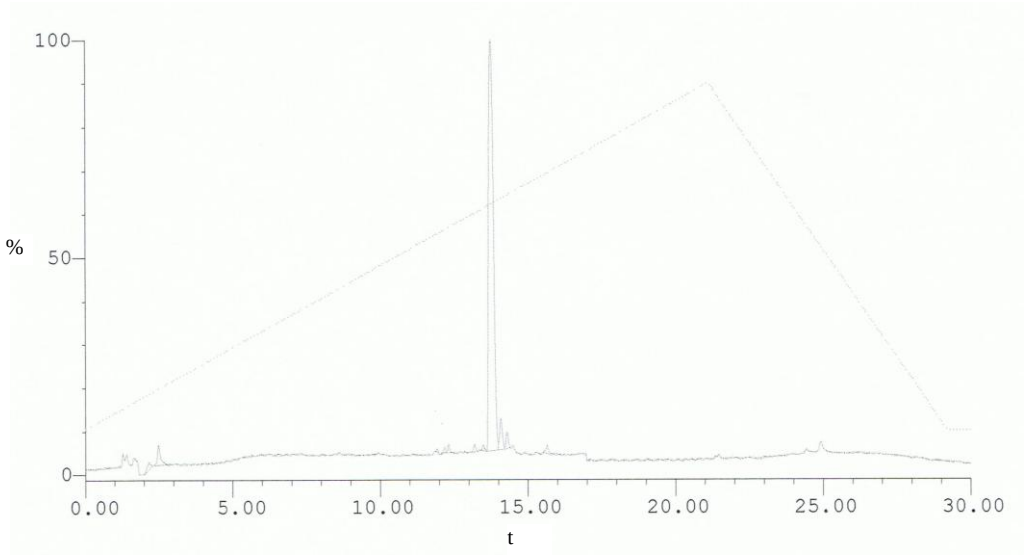
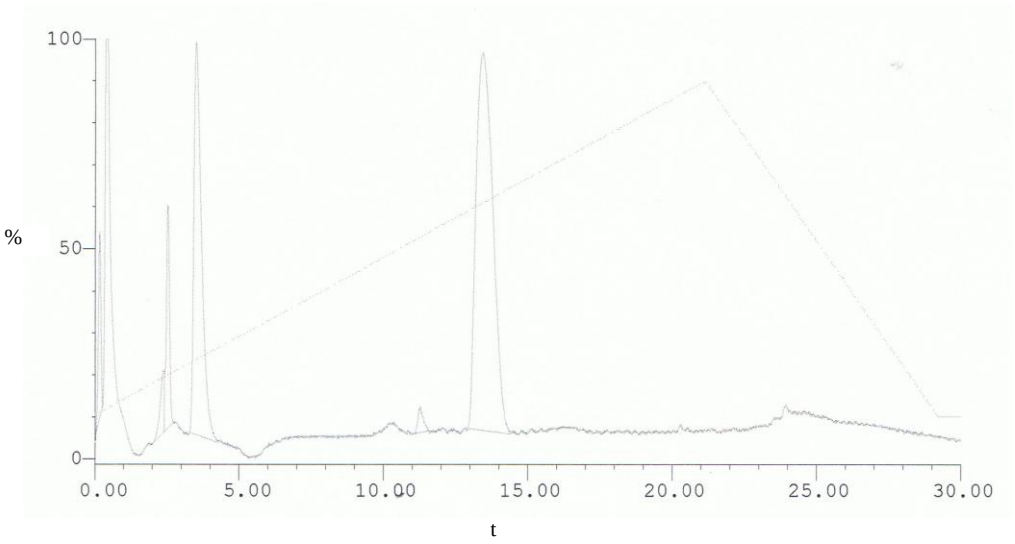
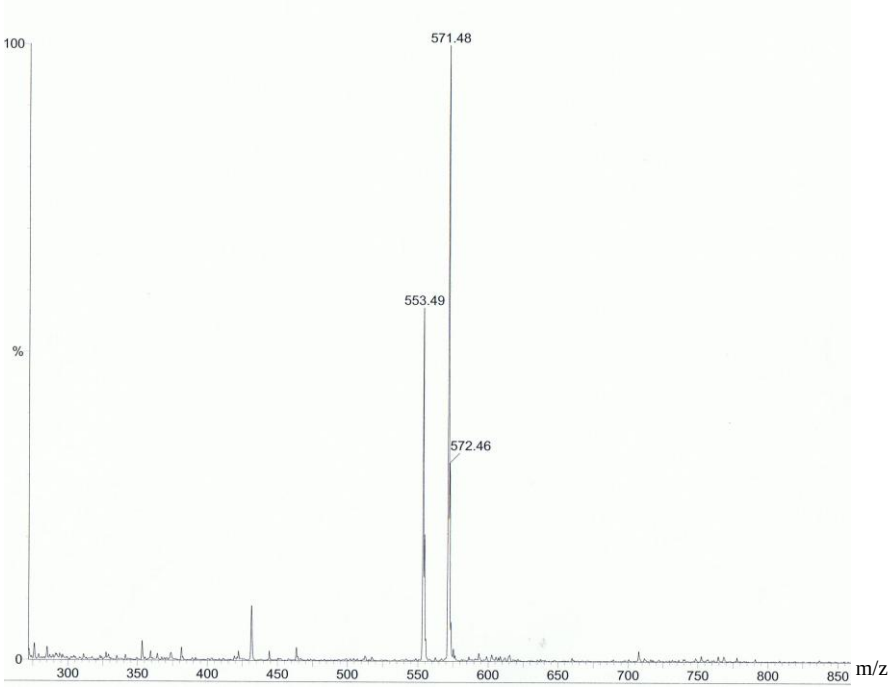


Figure 2:18 HPLC of Z-RFL-aldehyde



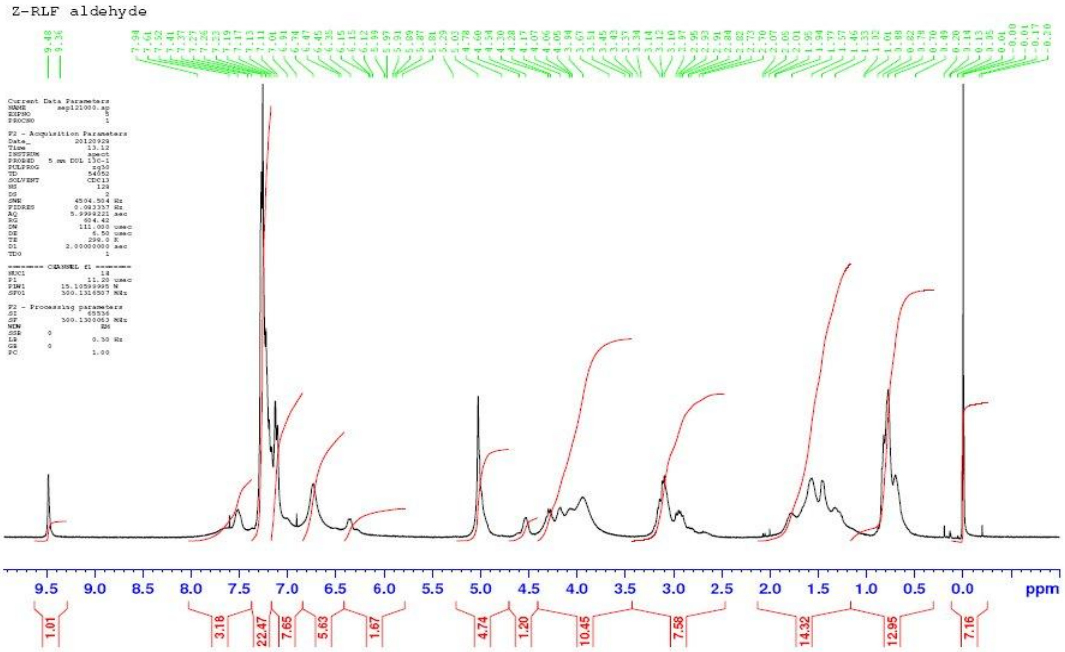
A broad HPLC peak was seen for the aldehyde (**Figure 2.18**), which is normal for C-terminal peptide aldehydes as in aqueous solution they predominantly exist in the hydrated form, as seen by ESMS **(Figure 2.19)**. The broad peak can be explained by the aldehyde and its hydrated form coeluting by HPLC. The low yield of peptide Z-RLF aldehyde meant that only a low resolution ^{1}H NMR spectrum was obtained (**Figure 2.20)**. The NMR spectrum was not fully assigned, but showed a weak aldehyde signal at δ 9.5 ppm due to exchange of the aldehyde proton with water.

Figure 2.19: ESMS of Z-RLF-aldehyde



Z-RLF-aldehyde [M + H] 553, hydrated form [M + H] 571

Figure 2.20: NMR spectrum of Z-RLF-aldehyde



2.6 Synthesis of Longer C-terminal Peptide Aldehydes

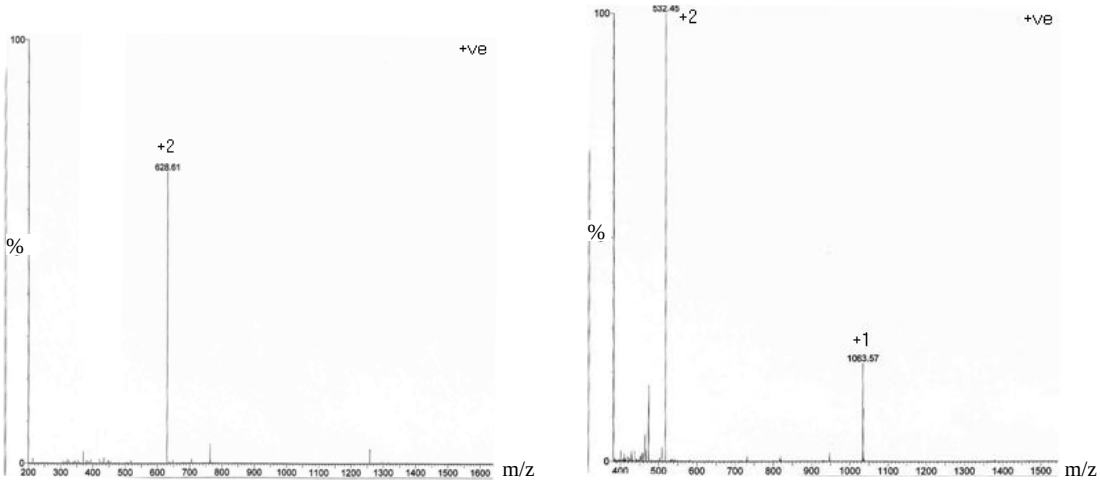
After synthesis of the small C-terminal aldehyde peptide Z-RLF-H, longer peptides were synthesised as a step towards the process of making long peptide fragments for use in chemoselective ligation. Bovine ribonuclease A had been chosen as the example protein for use in the ligation studies **(Chapter 1, Section, 1.4)**, and so the two peptides sequences shown below corresponding to short regions of bovine ribonuclease A were synthesised.

KPVNTFVHESL-aldehyde 11mer
YQSYSTMSI-aldehyde 9mer

Amino alcohol derivatives of leucine (**Chapter 7, Section 7.2.5.4.4, [11])** and isoleucine (**Chapter 7, Section 7.2.5.4.9, [16])** were synthesised following the same method as for synthesis of Fmoc-Phe amino alcohol (**Chapter 7 Section 7.2.5.4**). and attached to 2-chlorotrityl resin (**Chapter 7, Section 7.2.6.3**). The peptide amino alcohols KPVNTFVHESL-amino alcohol and YQSYSTMSI-amino alcohol were synthesised by standard Fmoc chemistry as detailed in **Chapter 7, Section 7.2.1.**

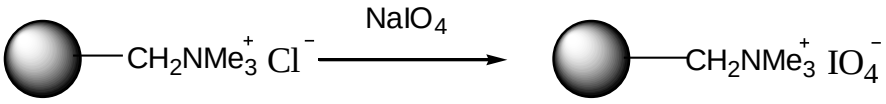
The peptide amino alcohols were cleaved from the solid support and purified by RP-HPLC, and the purified peptides oxidised to their corresponding C-terminal aldehydes (**Figure 2.21**) using a polymer supported periodate, synthesised from amberlyst ion exchange resin (**Chapter 7, Section 7.2.10.1, Figure 2.22**).

Figure 2.21: ESMS data for C-terminal aldehyde peptides



KPVNTFVHESL-aldehyde 1254.442 **YQSYSTMSI-aldehyde 1063.185**

Figure 2.22: Amberlyst ion exchange resin treated with sodium periodate



The advantage of using a polymer supported periodate is that the pure peptide aldehyde peptide can be filtered away from the resin, which after solvent removal requires no further purification. Reactions of polymer supported reagents such as periodate have been reported to be solvent specific (Chesney, (1999)), with several reactions only possible in methanol. For the oxidation of the peptide amino alcohols detailed above, methanol was found to be the only solvent suitable for the reaction to proceed. This limits use of this method to peptides that are readily soluble in methanol, which was the case for these two test peptides, but was predicted to be problematic for longer peptides.

Yields for the two test peptides were very low. After HPLC purification, KPVNTFVHESL-amino alcohol yielded 3.1 mg and YQSYSTMSI-amino alcohol yielded 2.4 mg, which corresponded to lower than 5% yield based on the starting resin substitution. Synthesis of peptides of this length would be expected to give a yield of approximately 50 mg (50% yield based on starting resin substitution. After oxidation to the aldehydes no further purification was performed and recovery from the polymer supported periodate was approximately 80%. A discussion of the low yields of the peptides with C-terminal amino alcohols follows in **Section 2.7**.

2.7 Synthesis of Ribonuclease Fragments Bearing a C-terminal Amino Alcohol

Synthesis of ribonuclease by sequential chemoselective ligation would require the synthesis of several long peptides with C-terminal amino alcohols as masked aldehydes. **Chapter 4, Section 4.1.1** describes how ligation sites were chosen using the available published data and crystal structures. Ligation sites in the positions planned required the synthesis of amino alcohol derivatives of alanine, leucine and isoleucine, and the replacement of two alanine residues and one threonine residue with either amino oxy acetic acid or hydrazino acetic acid (**Figure 2.23**).

Figure 2.23: Primary sequence of bovine pancreatic ribonuclease A, split into four peptide fragments

KETAAAKFERQHMDSSTSAASSSNYCNQMMKSRNLTKDRCKPVNTFVHESLADVQAVCSQK
NVACKNGQTNCYQSYSTMSITDCRETGSSKYPNCAYKTTQANKHIIVACEGNPYVPVHFDASV

Peptide 1) KETAAAKFERQHMDSSTSA-al
Peptide 2) X-SSSNYCNQMMKSRNLTKDRCKPVNTFVHESL-al
Peptide 3) X-DVQAVCSQKNVACKNGQTNCYQSYSTMSI-al
Peptide 4) X-DCRETGSSKYPNCAYKTTQANKHIIVACEGNPYVPVHFDASV

X = N-terminal hydroxylamine or hydrazide

In addition to amino alcohol derivatives of leucine and isoleucine synthesised for use in the preparation of the peptides detailed in **Section 2.6**, Fmoc-Ala amino alcohol (**Chapter 7, Section 7.2.5.4.15, [21])** was synthesised following the same method as for synthesis of Fmoc-Phe amino alcohol (**Chapter 7 Section 7.2.5.4).** The amino acid derivatives were attached to 2-chlorotrityl resin as described previously for phenylalanine derivatives (**Chapter 7 Section 7.2.6.3**). Peptides 2 and 3 (**Figure 2.23**) were longer versions of the short peptides detailed in **Section 2.6** above, and peptide 1 was a new peptide sequence. Peptide 4 had a C-terminal acid and so was synthesised on resin pre-loaded with Fmoc valine on the Wang linker. Peptide 4 was synthesised and purified and yielded 76 mg of purified peptide (49% yield from initial resin substitution). This peptide required a number of attempted syntheses to optimise the route and maximise the yield, and this is discussed in further detail in **Chapter 4, Section 4.1.2**. Several attempts at synthesis of the peptides with C-terminal amino alcohol were made, but none yielded full length peptide fragments in sufficient yield to use in further reactions. Extension of the peptide sequences beyond the 9mer and 11mer peptides detailed in **Section 2.6** resulted in such low yields, even for the shortest 19mer peptide fragment (**Figure 2.23, peptide 1**) of ribonuclease,

that although peptide of the correct mass could be detected by ESMS in the crude product (data was so poor that it is not presented here), no peptide was isolated after cleavage and HPLC purification. The lowering of final peptide yield during peptide elongation is likely to be due to cleavage of the peptide from the solid support during synthesis.

Previous ligation studies performed using C-terminal diols (Broadbridge *et al.* (2001), **Section 2.2.1**) had been performed using low milligram quantities of peptides, which could be monitored by LCMS and HPLC, but resulted in only just detectable quantities of peptides after a series of sequential ligations. The resin substitution for these amino acid diols was known to be low compared to glycine diol and although the yields of these peptides relative to starting resin substitution levels were not calculated, yields appear to be consistent with those seen for the ribonuclease peptides synthesised in this study (approx 5% yield, **Section 2.7**), suggesting that initial substitution levels are not solely responsible for low yields of peptides. Yields of peptides synthesised on commercially available glycine diol and glycine amino alcohol attached to chlorotrityl resin can be achieved at similar levels to those achieved with a carboxylic acid attached to chlorotrityl resin, ie with yields of approximately 50% based on the starting resin substitution level. Therefore it appears that for both diol and amino alcohol derivatives of other amino acids than glycine, not only are starting resin substitution levels lower, but also cleavage of the peptide from the resin occurs during synthesis leading to very low yields of peptides.

The following explanation for peptide cleavage during synthesis is proposed. One difference between glycine and the other amino acids was the ability to fully protect the hydroxyl of the amino alcohol or diol after attachment to the solid support with both glycine diol and glycine amino alcohol, but not with the other amino acids synthesised for this investigation (**Section 2.4**). Chlorotrityl resin is succeptible to cleavage by alcohols (for example trifluoroethanol (TFE) or hydroxybenotriazole (HOBt) can be used to cleave protected peptides from chlorotrityl resin), and an

unprotected hydroxyl could be sufficiently acidic to slowly cause cleavage of the peptide during peptide synthesis. More research would be required to confirm this hypothesis and to quantify the loss of peptide at each reaction cycle, however at this stage of the investigation it was decided that no further research would be continued into C-terminal masked aldehydes, and alternative strategies would be considered. Each peptide fragment would be required in at least 100 mg quantities, to allow for loss of product at each stage of ligation and oxidation, for any method of sequential ligation to become a viable method for peptide and protein synthesis.

2.8 Conclusion

A novel route for the synthesis of 1, 2 amino alcohol derivatives of Fmoc protected amino acids has been developed. The amino alcohol derivative of Fmoc amino acids can be used as a replacement for Fmoc-amino acid diols in the production of C-terminal peptide aldehydes. The route of synthesis of the Fmoc amino alcohol is high yielding and uses standard Fmoc protected amino acids as the building block. The route for synthesis was only tested using amino acids that did not require side chain protection, so further work would be required to determine the suitability of this method of synthesis for other amino acids.

Investigation showed the Fmoc amino alcohol to be unstable when stored at 4ºC , as the free amine is sufficiently basic to remove the Fmoc group from the alpha amine during storage. However, if the derivative is stored as the azide precursor at 4ºC, and the amine generated just prior to attachment to the polystyrene resin, this issue is resolved.

The aim of the amino alcohol was to achieve higher levels of substitution of the residue on chlorotrityl resin compared to diols, which would enable higher yields of peptides for use in chemoselective ligation reactions. Unfortunately, although an

amine is more reactive towards chlorotrityl resin than an alcohol, it appears that the amine in a 1, 2 amino alcohol is not significantly more reactive than the alcohol in a 1, 2 diol. However it was also discovered that it was not only the low level of substitution that was responsible for low yields of peptide. Spontaneous cleavage of peptide from the linker during peptide synthesis may have resulted in low yields of peptides. Therefore although development of a route of synthesis of C-terminal aldehydes via the synthesis of peptides containing C-terminal amino-alcohol has been achieved, it has not proven to be useful in the production of high yielding peptide fragments for use in sequential chemoselective ligation.

Returning to the literature for chemoselective ligation, and with the knowledge that an N-terminal serine is a convenient 1, 2 amino alcohol for use as an aldehyde precursor in peptides, the decision was made to investigate swapping of the groups for ligation so that the aldehyde was at the N-terminus with the hydrazide or amino-oxy group at the C-terminus. **Chapter 3** expands on this idea and details the synthesis of peptides for use in chemoselective ligation, and optimisation of the oxidation and ligation reactions. **Chapter 4** describes the synthesis of peptide fragments using these functional groups for synthesis of RNase A by chemoselective ligation.

Chapter 3: Chemoselective ligation

3.0 Introduction

Chemoselective ligation describes any chemical reaction used to join two unprotected peptides in solution, usually joining the the N-terminus of one peptide to the C-terminus of another peptide, replacing a natural peptide bond. The main aim of this work was to optimise a method of synthesis of peptide fragments for chemoselective ligation and to look at the effects of these unnatural linkages on the structure and function of a protein.

This chapter examines the structure of the linkages formed by a number of different chemoselective ligation reactions. The chapter then details the synthesis of peptides containing N-terminal serine as a precursor to an N-terminal glyoxal, and peptides containing a C-terminal hydrazide, for use in chemoselective ligation, and optimisation of the oxidation and ligation reactions. The aim of this section of study was to verify that ligation of an N-terminal glyoxal to a C-terminal hydazide is a viable method of ligation, allowing peptides to be synthesised and ligated in high yield.

3.1 Comparison of Structure of Non-amide Chemoselective Bonds

Non-amide chemoselective ligation has several aspects that must be considered when it is chosen as a method for peptide or protein synthesis. One is the ease of synthesis of the individual peptide fragments with their modified termini for the ligation, which is an important aspect of this thesis. Another is the rate and yield of reaction of the ligation reaction. However a third important feature is the structure of the bond formed at the ligation site, and how this differs from the natural peptide bond. **Chapter 5** details the molecular dymanics study that was performed to look at the effect of chemoselective bonds used in this thesis on the structure of an example

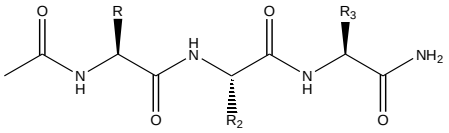
protein, bovine pancreatic ribonuclease A. In this chapter a more simple structural comparison of the bonds is made.

For the purpose of this thesis, only the structures of chemoselective bonds formed through reaction between an aldehyde and a hydrazine or hydroxylamine are compared. There are a number of commercial reagents available for use in peptide synthesis which can be used to generate peptides with terminal aldehydes, hydrazine and hydroxylamine functionalities. Hoewever, use of these reagents results in loss of one or more amino acids from the ligation site and generate a bond that does not closely resemble the natural peptide bond. These reagents are useful when using chemoselective ligation to join two molecules, such as linking a dye to a protein, or for template assembly of a number of peptides onto a scaffold. However for protein synthesis it is desirable that the ligated bond most closely mimics the natural peptides bond and that it includes as many amino acid side chains as possible close to the ligation site.

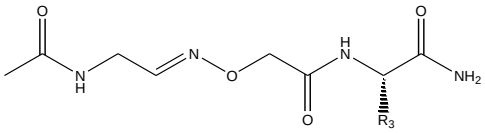
Figure 3.1 shows several different structures formed through chemoselective ligation of an aldehyde to a hydrazine or hydroxylamine, either using commercial reagents of peptide synthesis, or the methods used in this thesis.

Structure 1 shows the structure of a natural tripeptide, with the amino acids represented in three letter code with Xaa representing any of the 19 natural amino acids. **Structures 2** and **3** show the ligated bonds formed using commercially available reagents for chemoselective ligation. **Structure 2** is formed through the reaction of a peptide bearing a C-terminal glycine aldehyde (formed from a glycine diol precursor) ligated to a peptide bearing a N-terminal hydroxylamine (available from Novabiochem as Bis-Boc-amino-oxyacetic acid, catalogue number 851028). **Structure 3** is formed through the reaction of a peptide bearing a C-terminal side chain-amino-oxy-acetic acid (available from Novabiochem as Fmoc-Dpr(Boc-Aoa)-OH, catalogue number 852216) ligated to a peptide bearing an N-terminal glyoxal (fromed from an N-terminal serine precursor). Both of these bonds result in loss of two amino acid side chains, and show little structural similarity to the natural peptide sequence.

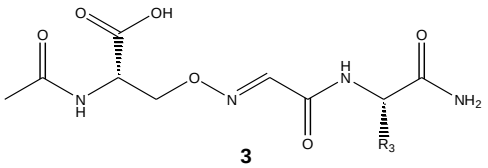
Figure 3.1: Comparison of structures of natural tripeptide Xaa-Xaa-Xaa, with different chemoselective bond replacements



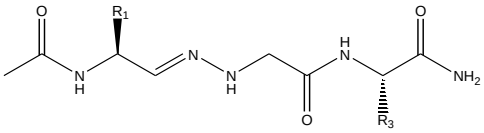
1



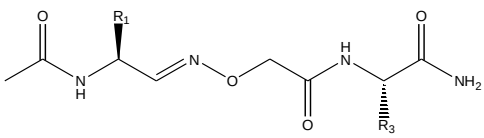
2



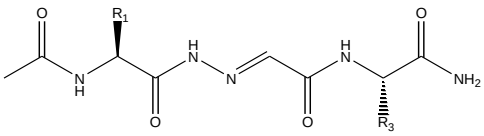
3



4



5



6

1: Natural peptide sequence

2: C-terminal glycine aldehyde ligated to N-terminal amino-oxy acetic acid

3: C-terminal side chain-amino-oxy-acetic acid ligated to N-terminal glyoxal

4: C-terminal aldehyde ligated to N-terminal hydrazine

5: C-terminal aldehyde ligated to N-terminal hydroxylamine

6: C-terminal hydrazide ligated to N-terminal glyoxal

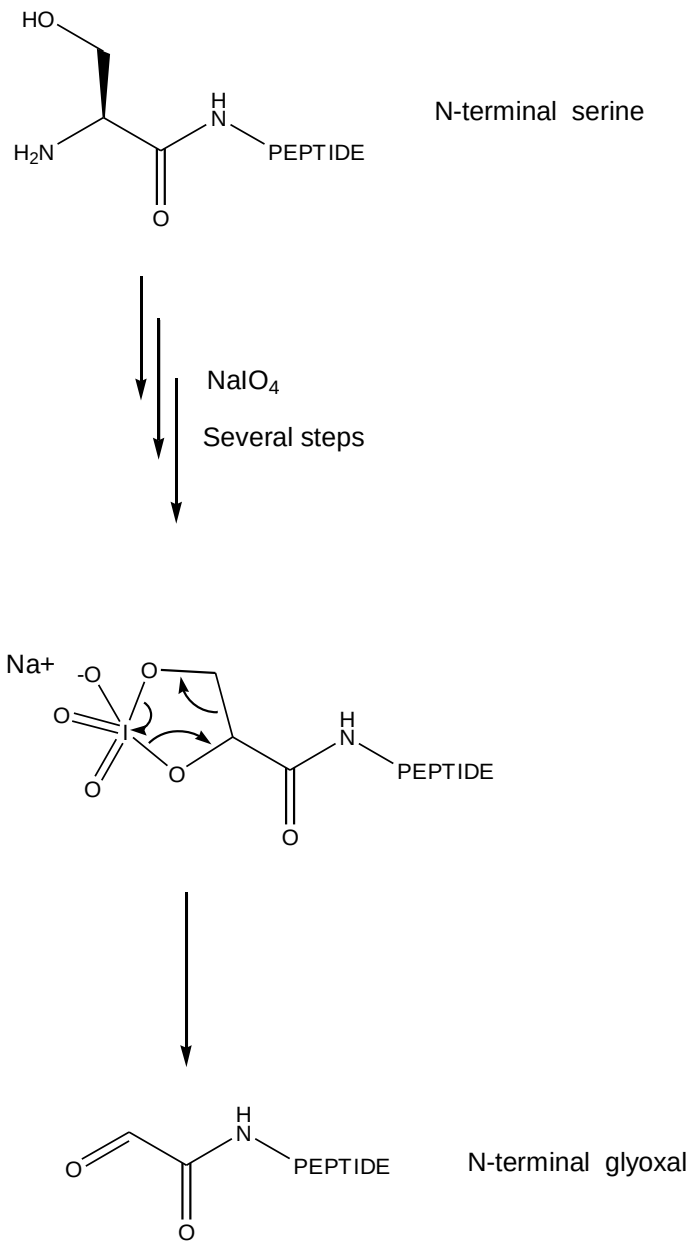
Structures 4 and **5** show the ligated bonds formed through reaction of a peptide bearing a C-terminal aldehyde (formed from the corresponding amino acid diol or 1, 2-amino alcohol, described in **Chapter 2**), reacting a peptide bearing either an N-terminal hydrazide (was available from Novabiochem as Tri-boc-hydrazinoacetic acid catalogue number 04-12-0245, now discontinued) or N-terminal hydroxylamine (Bis-Boc-amino-oxyacetic acid, catalogue number 851028 as used for **Structure 2**). Both of these bonds result in loss of just one amino acid side chain, and have a closer similarity to the natural peptide sequence. All of these examples have one additional bond length compared to the natural peptide sequence.

Chapter 2 concluded with the decision that the 1, 2-amino alcohol derivatives of Fmoc protected amino acids was not a suitable method for the synthesis of peptide fragments in high yield, which lead to the alternative investigation into reaction between an N-terminal aldehyde reacting with a C-terminal hydrazide. **Structure 6** shows the ligated bond formed between these two functionalities. Again there is loss of a single amino acid side chain and one extra bond length compared to the natural peptide bond, but the structure most closely resembles the natural peptide sequence,

3.2 N-terminal Serine as a Masked Aldehyde

An N-terminal serine is a 1, 2 amino alcohol and is therefore a convenient aldehyde precursor, as it can be incorporated into any peptide sequence. An N-terminal threonine can also be used. Oxidation with sodium metaperiodate produces a glyoxal.The actual mechanism is complicated and proceeds via a number of steps to form a cyclic intermediate, which then spontaneously rearranges to form the glyoxal (**Figure 3.2**).

Figure 3.2 N-terminal serine oxidation to glyoxal



At the site of chemoselective ligation, the N-terminal residue of the C-terminal peptide fragment is replaced with a serine, so the ligation results in loss of the side chain coresponding to the amino acid that is replaced by serine. An N-terminal serine as an aldehyde precursor was first reported by Geoghegan *et al.* (1992), and has since been used in numerous publications as a cheap and convenient route to an N-terminal aldehyde.

3.3 Peptides Containing a C-terminal Hydrazide

A C-terminal peptide hydrazide is a simple modification to the C-terminus of a peptide, formed through the reaction of the C-terminal carboxylic acid to hydrazine through an amide bond. Synthesis of peptides containing a C-terminal hydrazide can be achieved in a number of different ways. One method is to synthesise a fully side chain and N-teminal protected peptide by SPPS using a hyperacid labile linker such as 2-chlorotrityl, and react this protected peptide in solution to hydrazine using a coupling agent such as DIC or HBTU. The main disadvantage to this technique is racemisation of the C-terminal residue. Activation of the carboxyl of a peptide is much more prone to racemisation than activation of the carboxyl of a single amino acid, mainly because an activated amino acid reacts more quickly than an activated peptide giving less time for racemisation to occur. A superior method of synthesis is one that forms the C-terminal hydrazide through reaction of a C-terminal amino acid directly onto a hydrazine modified solid support, followed by synthesis of the peptide.

Hydrazide peptides were originally synthesised using this method by Wang and Merrifield (1969) by Boc chemistry using *t*-alkyloxycarbonylhydrazide resin and later by Wang (1973) using *p*-alkoxybenzyloxycarbonylhydrazide resin and 2-(4-biphenyl)isopropyloxycarbonyl (Bpoc), a hyperacid sensitive amine protecting group, as an alternative to Boc, which allowed cleavage from the resin in TFA rather than HF. Since the development of Fmoc as an orthogonal amine protecting group, this resin has also been used with Fmoc chemistry for the synthesis of peptide hydrazides. However, reported low yields and side reactions during peptide cleavage lead to the use of hydrazine bound to 2-chlorotrityl resin by Stavropoulos *et al.* (1995), allowing cleavage in low TFA concentrations to avoid the side reactions associated with high concentrations of TFA.

This method of synthesis was used in this investigation for the synthesis of peptides bearing C-terminal peptide aldehdyes. 2-chlorotrityl chloride resin was loaded with hydrazine using a 2% solution of hydrazine monohydrate in DMF (**Chapter 7, Method 7.2.8.1**) . Peptides were then synthesised on this resin using the Symphony

or Prelude automated peptide synthesisers (Protein Technologies Inc) . Cleavage from the chlorotrityl linker in 95% TFA and appropriate scavengers depending on the peptide synthesis produced C-terminal peptide hydrazides in high yield and purity (**Section 3.4.2**).

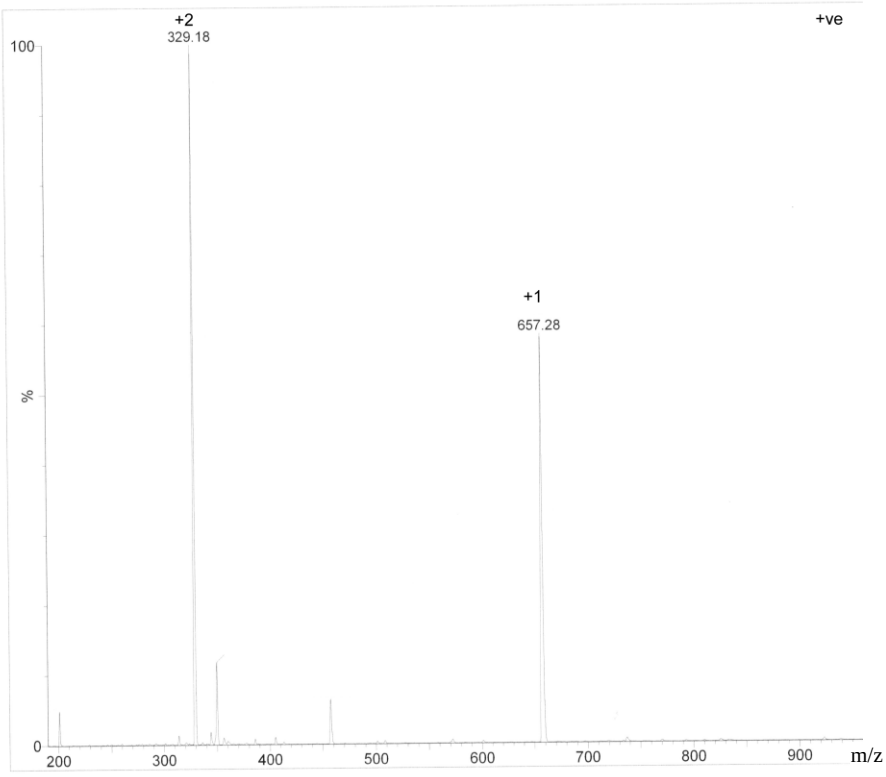
3.4 Synthesis of Peptides for use in Trial Ligations

Short peptide sequences were chosen and synthesised for use in trial scale oxidation and chemoselective ligation reactions. Two sequences were chosen that are peptides regularly ordered as custom peptides at PPR Ltd, which are straightforward to synthesise and purify in high yield.

3.4.1 Test peptide with N-terminal Serine

The peptide SLIGRL-NH2 which has an N-terminal serine and a C-terminal amide was chosen as a test sequence for the glyoxal peptide. This is a sequence derived from the N-terminus of the protease-activated receptor-2 (PAR2), and is a receptor agonist. The peptide was synthesised and purified by RP-HPLC as detailed in **Chapter 7, Sections 7.2.1 and 7.3.1** (ESMS data shown in **Figure 3.3**). The peptide was synthesised on 500 mg polystyrene resin preloaded with Fmoc-Rink amide MBHA linker, at a substitution level of 0.36 mmol/g. The final purified yield was 69 mg which corresponds to a 58% yield.

Figure 3.3: ESMS data for SLIGRL-NH2



SLIGRL-NH2 [M + H] 657

3.4.2 Test peptide with C-terminal Hydrazide

The peptide 8Aoc-fQWAVGHL-NH-NH2 which has an N-terminal 8-amino octanoyl group at the N-terminus, D-phenylalanine and a C-terminal hydrazide was chosen as a test sequence for the hydrazide peptide. This is a sequence derived from the C-terminus of bombesin and is a peptide antagonist.

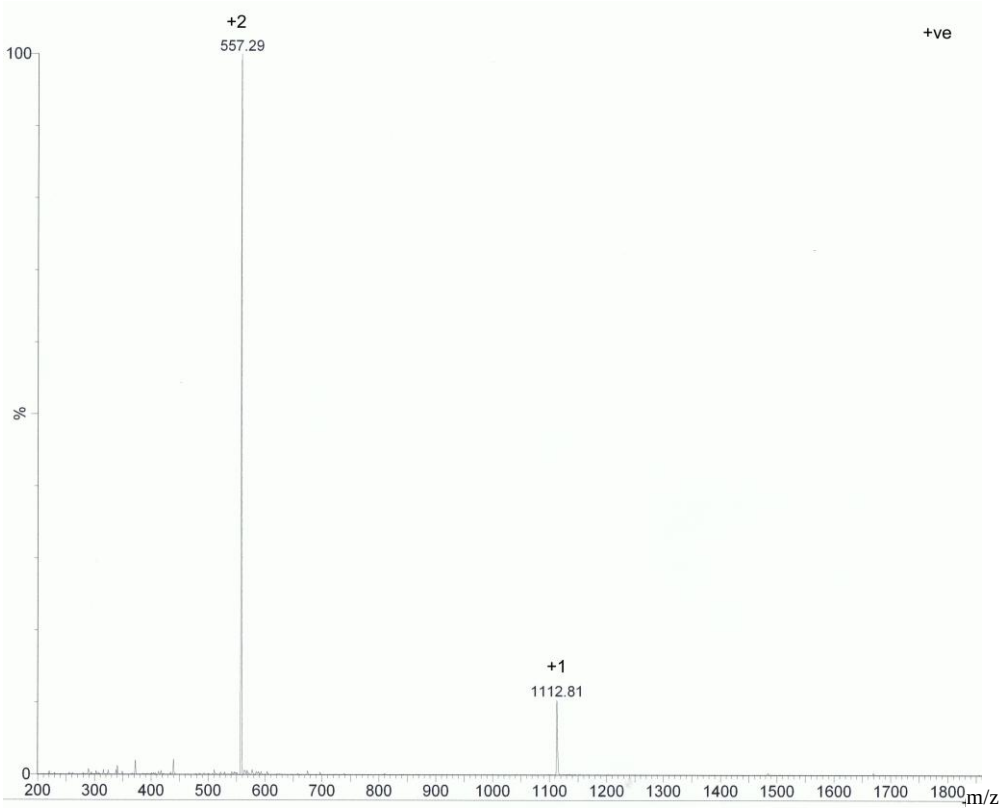
The peptide was synthesised on hydrazine substituted 2-chlorotrityl chloride resin (CTR) (**Chapter 7, Section 7.2.8.3**) . The substitution level of the hydrazine on CTR cannot be measured, but is assumed to be close to the level of substitution of the starting chloride resin (1.3 mmol/g). This is too high for peptide synthesis and so the substitution level was lowered during the first amino acid coupling by reacting Fmoc-Leu and Boc-Gly in a ratio of 4:1 to terminate a proportion of the resin with

Boc-Gly. The Gly-NH-NH2 product is easily separable from the correct peptide by HPLC, and so this method is a convenient way to reduce resin substitution. The proportion of resin terminated with Boc-Gly does not exactly correspond to the proportion of Boc-Gly used in the coupling reaction. Glycine reacts more quickly than other amino acids and Boc amino acids have better solubility than Fmoc amino acids, so a higher proportion of Boc-Gly reacts than the Fmoc amino acid.

This method of resin lowering is used regularly at PPR Ltd to lower the substitution of Wang preloaded resins for use on peptide synthesisers. A resin substitution level of 0.5 mmol/g or above produces high levels of terminated sequences during peptide synthesis, and so a substitution level of between 0.2 and 0.4 mmol/g is ideal for minimising peptide chain aggregation, while maximising the yield of peptide without requiring use of large amounts of resin. Wang resins at 0.5-0.8 mmol/g substitution that are lowered using a 4:1 ratio of Fmoc-amino acid to Boc-Gly produces a substitution level of between 0.2-0.4 g (investigations performed at PPR Ltd, results not shown here), so approximately a halving of resin substitution.

If a large excess of amino acid is reacted to CTR the highest substitution achieved is approx 0.8 mmol/g. It was assumed that the hydrazine-CTR would react similarly, and so use of a 4:1 ratio of Fmoc-Leu to Boc-Gly would result in a substitution level of approx 0.4 mmol/g. The actual substitution level of the Fmoc-Leu-hydrazide resin was not measured due to difficulty accessing a photospectrometer, but the yield of 8Aoc-fQWAVGHL-NH-NH2 produced from 300 mg of resin was 48 mg, which would correspond to a 40% yield at 0.4 mmol/g substitution or an 80% yield at 0.2 mmol/g substitution. As 50% yields are typical in peptide synthesis, and peptide synthesis proceeded without signs of terminated products it is likely that the substitution levels achieved were between 0.2-0.4 mmol/g. ESMS data for 8Aoc-fQWAVGHL-NH-NH2 is shown in **Figure 3.4**.

Figure 3.4: ESMS of 8Aoc-fQWAVGHL-NH-NH2



8Aoc-fQWAVGHL-NH-NH2 [M + H] 1113

3.5 Periodate Oxidation of Peptides bearing N-terminal Serine

The peptide SLIGRL-NH2 was oxidised using sodium metaperiodate to produce the N-terminal glyoxal. The method of Gaertner *et al.* (1992) was followed, which used a 2 fold excess of $NaIO_4$ in imidazole hydrochloride buffer (50 mM, pH 6.9) (**Chapter 7, Section 7.2.10.2**).

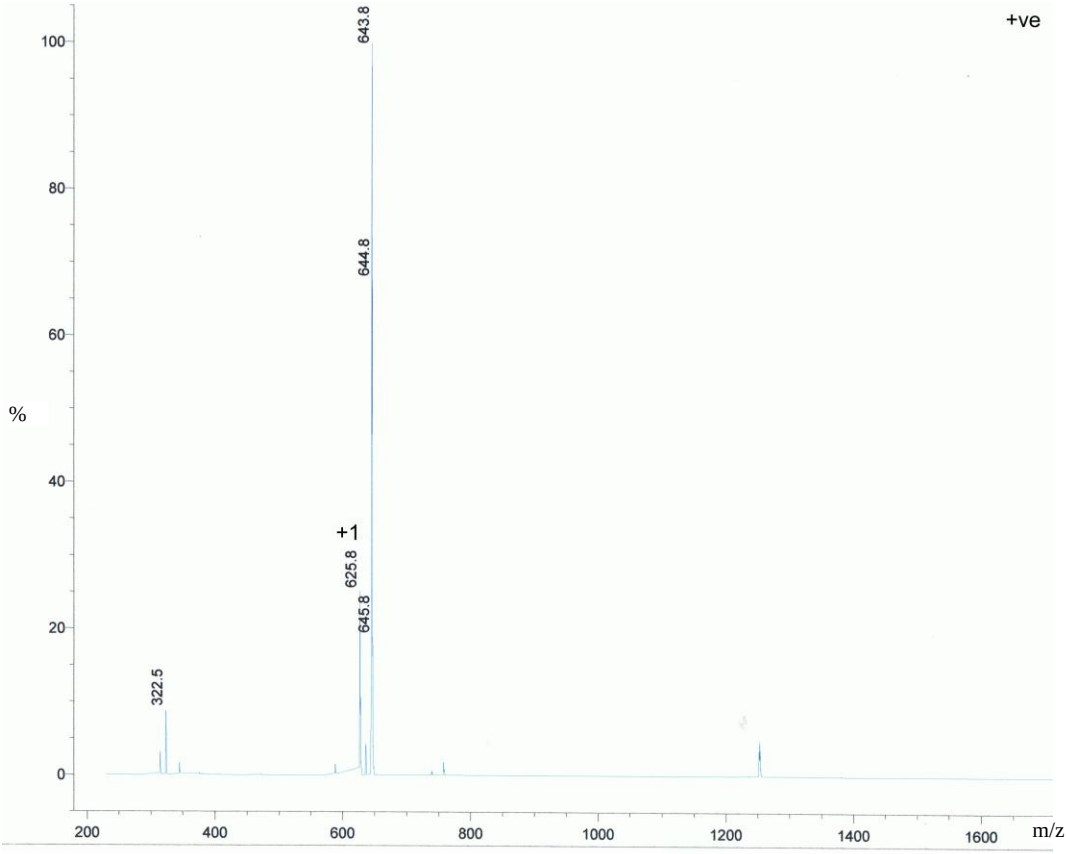
A trial scale oxidation was first performed using 1 mg of peptide at a concentration of 5 mg/ml. A small sample of the peptide was quenched after 10 minutes with ethylene glycol (2 equivalents over $NaIO_4$), and analysed by LCMS.

LCMS was used to separate the immidizole buffer from the peptide by HPLC, allowing the peptide to be analysed by MS. The peptide was loaded onto the LCMS in 5% MeCN and after elution of the buffer the MeCN was increased to 50%, causing co-elution of the starting peptide and glyoxal product.

The major peak seen by ESMS was [M + H] 645 which would correspond to the hydrated form of the glyoxal, and also visble was [M + H] 658 which is the starting peptide. Samples were also taken at 15 minutes and 20 minutes, and complete oxidation was seen at 15 minutes (no [M + H] 658 visible by MS) with no side reactions visible in the 20 minute sample.

The reaction was repeated using 50 mg of peptide at a concentration of 5 mg/ml (**Chapter 7, Section 7.2.10.2.1**). The reaction was quenched after 20 minutes and the peptide was immediately purified by RP-HPLC and freeze dried. The ESMS of the purified glyoxal is shown in **Figure 3.5**. The purified glyoxal was stored at -20 º C until required. The glyoxal peptide was used over several months and showed no signs of degradation or reduction in reactivity over this time, although no formal stability studies were performed.

Figure 3.5: ESMS of glyoxal-LIGRL-NH2



glyoxal-LIGRL-NH2 [M + H] 626, [M + H_2O] 644

3.6 Chemoselective ligation between an N-terminal Glyoxal and a C-terminal Hydrazide

8Aoc-fQWAVGHL-NH-NH2 was ligated to glyoxal-LIGRL-NH2 in NaOAc buffer (0.1 M, pH 4.6). In an initial trial, the peptides were reacted in a 1:1 ratio at a concentration of 1 mg/ml, and the reaction was monitored by MS. After 24 h there was both starting peptides and ligated product visible by MS, and after 3 days the reaction had gone to completion (**Chapter 7, Section 7.2.11.1**). The ligation was repeated at 2 mg/ml, in an attempt to spead up the rate of recation, but there no noticeable difference in reaction rate. The ligation was attempted at 5 mg/ml, but the hydrazide peptide fragment was not soluble at this concentration.

3.6.1 Use of Aniline as a Nucleophilic Catalyst

During these investigations a communication published by Dirksen *et al.* (2006) demonstrated the use of aniline as a nucleophilic catalyst in oxime ligation to dramatically increase reaction rates.

A reason why ligation of hydroxylamines and hydrazides is so slow is that the reaction is in equilibrium between the imine (Schiff base) and the starting reagents, and the equilibrium constant (K_{eq}), (the ratio of the rate constant of formation to the rate constant of hydrolysis) for this reaction is low. This means that high concentrations of reactants are required to force the reaction to completion. Peptide ligation often requires quite low concentration of the peptides due to solubility problems and so long reaction times and incomplete reactions are often reported as a problem with this method of ligation.

The pK_a of aniline is 4.6, which is approximately the same as the pK_a of the hydrazide. Dirksen *et al.* (2006) proposed the use of aniline in a large excess would cause Schiff base formation with aniline at the pH of the ligation. A transimination with hydroxylamine or hydrazide would then occur and this reaction is rapid at the pH of the ligation. Therefore the rate determining step would be controlled by the concentration of aniline used in the reaction, and not the concentration of the peptides being ligated. They demonstrated this in their 2006 communication with oxime ligation, and achieved up to 400 fold increases in reaction rate.

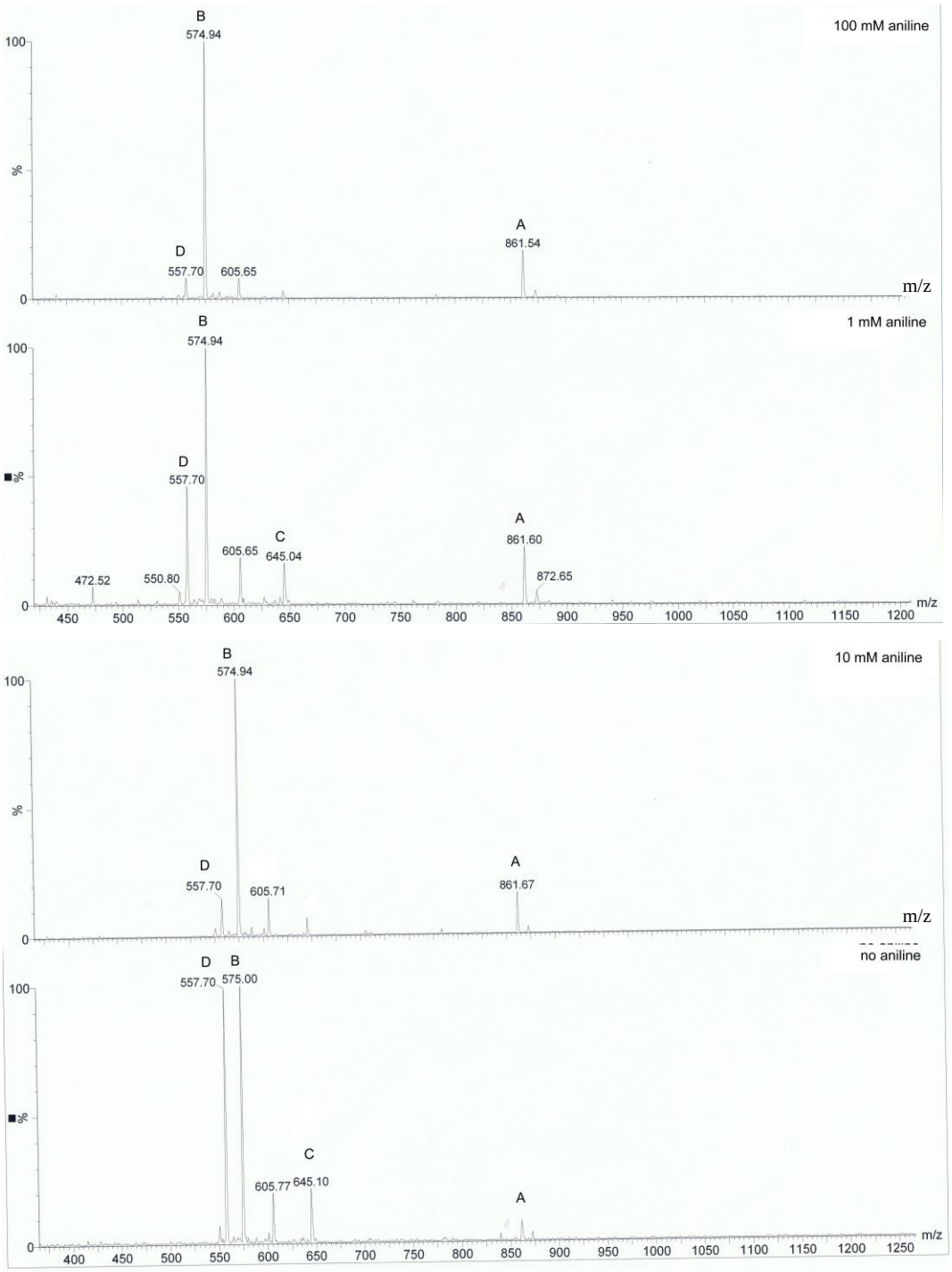
The use of aniline as a nucleophilic catalyst was therefore tested in this study for hydrazone ligation, as detailed below. A later publication by Dirksen *et al.* (2008) and so not available during this study used aniline to increase reaction rates in hydrazone ligation with similar effect.

3.6.1.1 Trial Ligations using Aniline

The trial ligation of 8Aoc-fQWAVGHL-NH-NH2 to glyoxal-LIGRL-NH2 in NaOAc buffer (0.1 M, pH 4.6) in a 1:1 ratio at a total peptide concentration of 1 mg/ml was performed in four separate reactions, using no aniline, 1 mM, 10 mM and 100 mM aniline. After 8 hours all the reactions were monitored by ESMS. **Figure 3.6** shows the ESMS for each reaction after 8 hours. All had starting peptide and ligated product visible. For the reaction that did not contain aniline, the two starting peptides (peak C and peak D) are clearly visible, along with the ligated product (peaks A and B). As the amount of aniline is increased, the quantity of starting peptides reduces. After 8 hours in the prescence of 10 mM and 100 mM aniline, the glyoxal peptide is bearly visible and the hydrazide peptide is greatly reduced compared to the reactions containing 1 mM or no aniline.

After 18 hours the reactions were monitored by MS and the reactions with 10 mM and 100 mM aniline had proceeded to completion. The reaction using 1 mM aniline was completed in 2 days. The reaction without aniline took 3 days to reach completion. It was therefore decided that 10 mM aniline would be most suitable for ligations, as a 10 fold increase to 100 mM did not increase the reaction rate sufficiently for the reaction to be completed in a day. Both 10 mM and 100 mM additions resulted in complete ligation in 18 hours, meaning ligation reactions could be performed overnight.

Figure 3.6: ESMS of trial ligation of 8Aoc-fQWAVGHL-NH-NH2 to glyoxal-LIGRL-NH2 after 8 hours



A = 8Aoc-fQWAVGHL-hydrazone- LIGRL-NH2 [M + 2H] 861 ([M + H] 1721)

B = 8Aoc-fQWAVGHL-hydrazone- LIGRL-NH2 [M + 3H] 575 ([M + H] 1721)

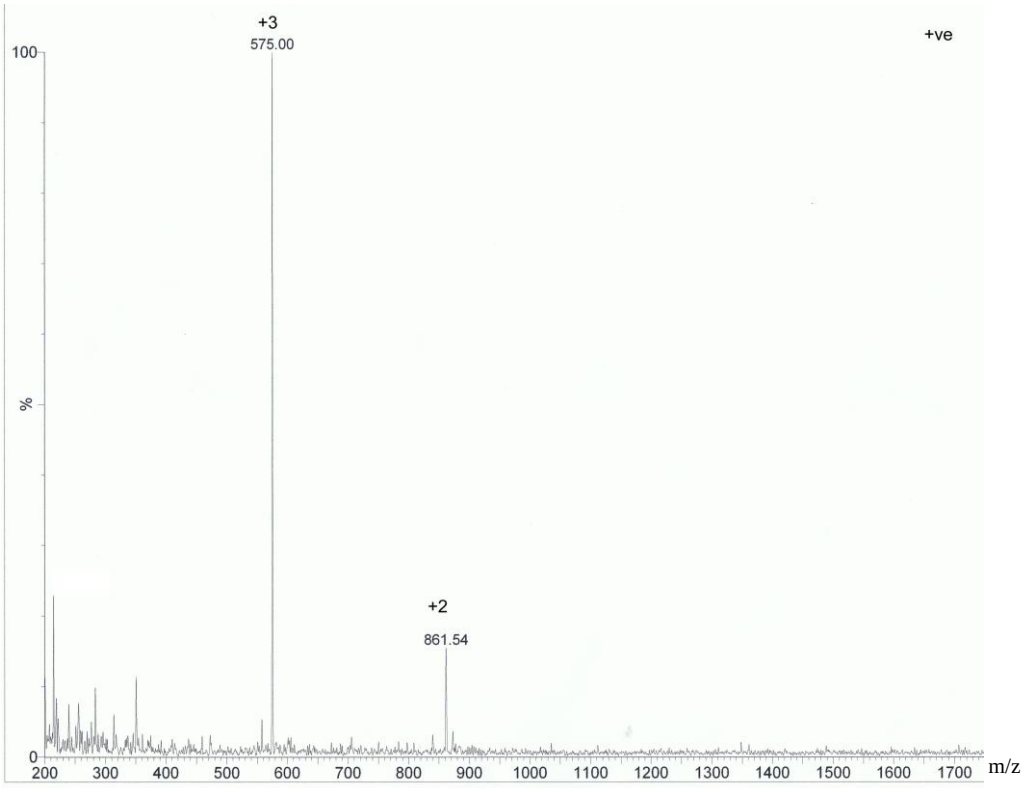
C = glyoxal-LIGRL-NH2 [M + H_3O] 645 ([M + H] 626)

D = 8Aoc-fQWAVGHL-NH-NH2 [M + 2H] 557 ([M + H] 1113)

3.6.1.2 Large Scale Ligation

The optimised conditions for ligation determined by the trial scale ligation was scaled up, so that the hydrazone product could be purified. 8Aoc-fQWAVGHL-NH-NH2 was ligated to glyoxal-LIGRL-NH2 in NaOAc buffer (0.1 M, pH 4.6) in a 1:1 ratio at a total peptide concentration of 1 mg/ml in the prescence of 10 mM aniline (**Chapter 7, Section 7.2.11.2**). The ligated product was purified by HPLC. Figure 4.7 shows the ESMS trace for the purified hydrazone peptide.

Figure 3.7: ESMS of 8Aoc-fQWAVGHL-hydrazone-LIGRL-NH2



8Aoc-fQWAVGHL-hydrazone-LIGRL-NH2 [M + H] 1721

3.7 Conclusion

Chemoselective ligation between two peptides, one bearing an N-terminal glyoxal and the other bearing a C-terminal hydrazide has been investigated as a potential method for use in the sequential ligation of several peptides. The studies detailed in

this chapter confirmed that this method is suitable for the ligation of two peptides in high yield, and therefore should be a suitable method to use in sequential ligation, using the serine as a masked glyoxal. During sequential ligation the N-terminal serine acts as a masked glyoxal, allowing ligation of two peptides, before unmasking of the glyoxal for use in the next ligation reaction.

The rate of hydrazone formation was increased by use of aniline as a nucleophilic catalyst, reducing reaction times from 3 days to less than 24 hours.

This method, using an N-terminal glyoxal ligated to a C-terminal hydrazide was therefore adopted for use in the sequential ligation of peptides. **Chapter 4** details the synthesis of peptide fragments of bovine pancreatic ribonuclease A and ligation of these peptides.

Chapter 4: Synthesis of Bovine Pancreatic Ribonuclease A

4.0 Introduction

Bovine pancreatic ribonuclease A (RNase A) was chosen as the example protein for total synthesis of a protein by sequential chemoselective ligation. It is one of the most studied small proteins, and so its structure and mechanism of catalysis are well understood (**Chapter 1, Section 1.4**).

At 124 amino acids, it is possible to synthesise as a single chain by SPPS, but this has only been achieved using Boc chemistry (Gutte *et al.* (1971), not Fmoc chemistry. Taking all this information into account it was decided that RNase A would be an ideal protein to synthesise by sequential chemoselective ligation, and to study the effects of the ligated non-natural bonds on the structure and function of the protein.

The sequence of ribonuclease was analysed for suitable ligation sites, and split into four sections. The N and C terminals of each fragment sequence needed the correct modifications added to allow sequential ligation. This chapter describes the design of peptide fragments using the original ligation method of an N-terminal hydrazide or hydroxylamine reacting with a C-terminal aldehyde as detailed in **Chapter 2**. The chapter then focuses on the design and synthesis of peptide fragments of RNase A containing an N-terminal glyoxal and C-terminal hydrazide following on from the studies described in **Chapter 3**.

4.1 Ligation Sites

The 124 amino acids that make up the primary sequence of bovine pancreatic ribonuclease were analysed for suitable ligation sites. It was decided to position the sites for ligation away from residues known to be involved in the active site, or

catalytic mechanism (**Figure 4.1**), to reduce the likelihood of the ligated bonds affecting the structure or activity of the enzyme.

Figure 4.1: Primary sequence of bovine pancreatic ribonuclease A

KETAAAKFERQHMDSSTSAASSSNYCNQMMKSRNLTKDRCKPVNTFVHESLADVQAVCSQK
NVACKNGQTNCYQSYSTMSITDCRETGSSKYPNCAYKTTQANKHIIVACEGNPYVPVHFDASV

Residues in **green** are involved in the active site and **pink** are essential for catalytic activity (**Chapter 1.4.1-1.4.2**).

4.1.1 Chemoselective Ligation between an N-terminal Hydrazide and C-terminal Aldehyde

At the ligation site the two amino acids either side of the ligated bond are important. In **Figure 4.2, (a)** shows a tetramer of four different amino acids with side chains R_1-R_4, **(b)** shows the two fragments that would be required for chemoselective ligation between a C-terminal aldehyde and N-terminal hydrazine (the original method of ligation detailed in **Chapter 2**) and **(c)** shows the product of the ligation. The amino acid with side chain R_2 is required as a C-terminal aldehyde and the amino acid side chain R_3 is not present in the ligated product.

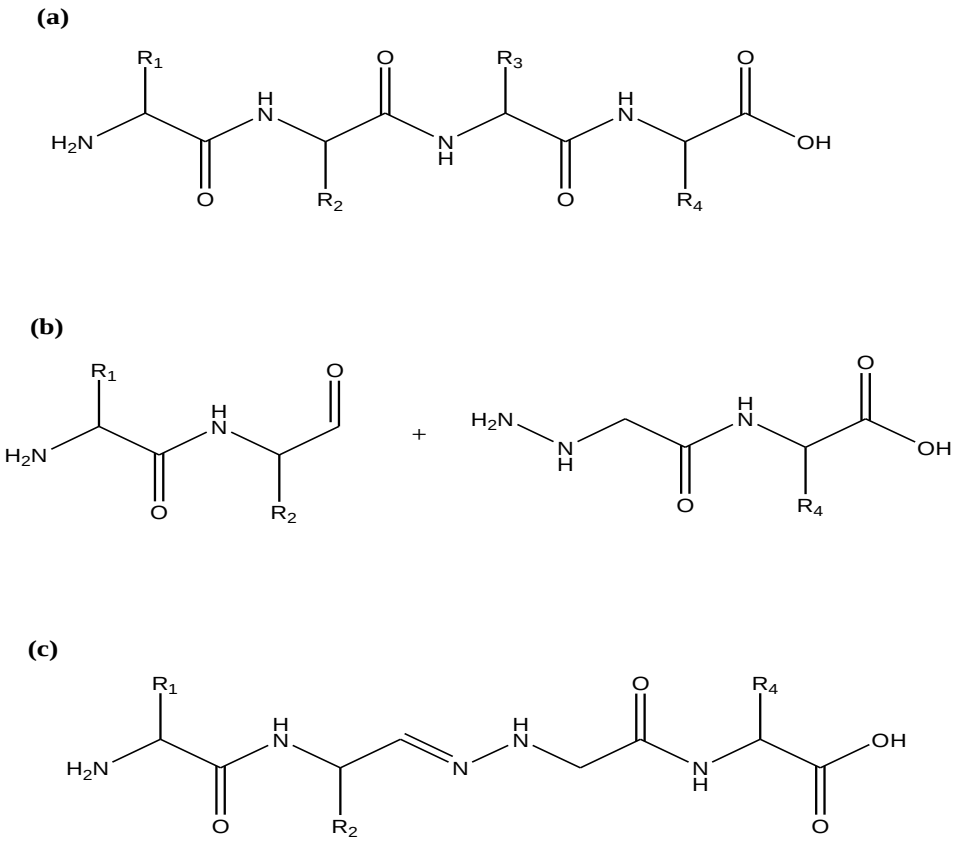
The ligation sites were chosen so that,

1) the C-terminal aldehyde of each peptide was an amino acid that did not require side chain protection, as the amino alcohol derivatives of these would be more simple synthetic targets, and

2) the amino acid side chain that would be lost would be a simple functional group so that removal of that side chain would be likely to have little effect on the structure of the enzyme. The best amino acid at this site would be a glycine as this amino does

not have a side chain functional group, but there were no glycines in suitable positions.

Figure 4.2: Hydrazide chemoselective ligation of C-terminal aldehyde to N-terminal hydrazine



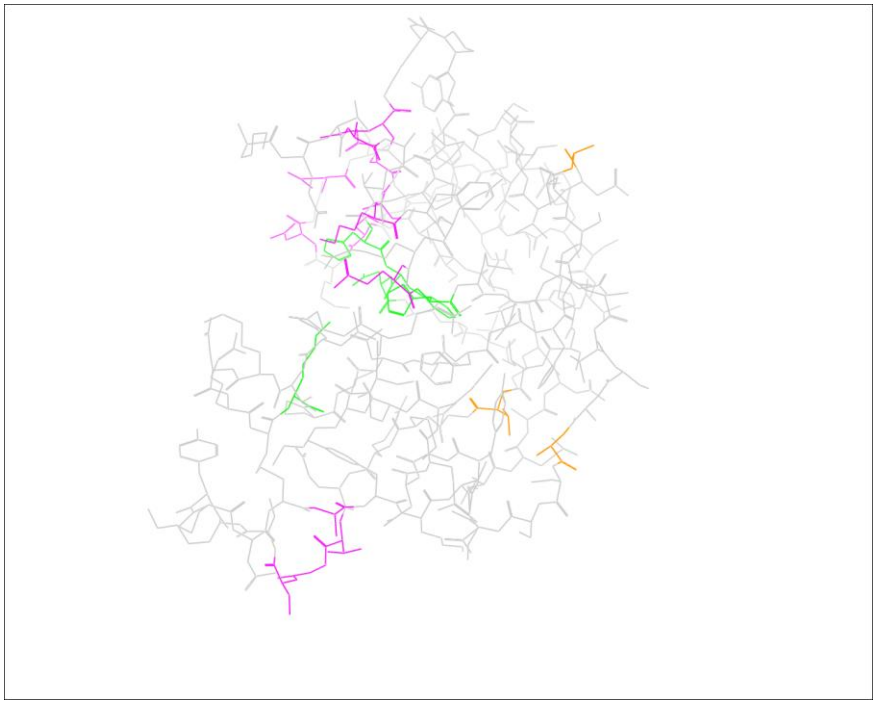
a) natural tetrapeptide
b) two peptide fragments required for chemoslective ligation
c) product of chemoselective ligation

Taking all these factors into consideration, three ligation sites were chosen splitting the sequence into four peptide fragments. Two of the ligation sites were able to be positioned at alanine residues, which has a simple methyl group side chain, and the third ligation site was positioned at a threonine residue which has a hydroxyl methyl side chain. **Chapter 2, Section 2.6** details the original plan for sequential ligation

using the C-terminal amino alcohol as a masked C-terminal aldehyde. The ligation sites and peptide fragments that would have been required are shown in **Chapter 2, Figure 2.12,** and repeated in **Figure 4.4** for clarity.

The crystal structure of bovine pancreatic ribonuclease A [3rn3 (Howlin *et al.* (1989))] was manipulated using MOE.2007.09 to show the residues essential for enzyme catalytic activity, and the three residues that would be lost in the ligation strategy shown above (**Figure 4.3**). This confirmed that these sites are all spatially arranged on the opposite face of RNase A from the residues known to be required for activity.

Figure 4.3: Image of RNase A (3rn3), showing sites of ligation and residues critical to ribonuclease activity



Green - active site residues, **Pink** - other important resides for catalytic activity, **Orange** – amino acids that would be replaced by proposed ligation sites.

4.2 Synthesis of Peptide Fragments Bearing C-terminal Amino Alcohols

Chapter 2, Section 2.6 describes the attempted synthesis of the three peptides bearing C-terminal amino alcohols (**Figure 4.4**, peptides 1, 2 and 3), synthesised on chlorotrityl resin substituted with amino alcohol derivatives of Fmoc amino acids. This method is referred to in this chapter as strategy 1.

Figure 4.4 Primary sequence of RNase A, split into four peptide fragments, using strategy 1

KETAAAKFERQHMDSSTSAASSSNYCNQMMKSRNLTKDRCKPVNTFVHESLADVQAVCSQK
NVACKNGQTNCYQSYSTMSITDCRETGSSKYPNCAYKTTQANKHIIVACEGNPYVPVHFDASV

Peptide 1) KETAAAKFERQHMDSSTSA-al
Peptide 2) X-SSSNYCNQMMKSRNLTKDRCKPVNTFVHESL-al
Peptide 3) X-DVQAVCSQKNVACKNGQTNCYQSYSTMSI-al
Peptide 4) X-DCRETGSSKYPNCAYKTTQANKHIIVACEGNPYVPVHFDASV

X = N-terminal hydroxylamine or hydrazide
al = C-terminal aldehyde
pink = residues involved at active site

The fourth peptide required for the synthesis of RNase A (**Figure 4.4**, peptide 4) has a C-terminal carboxylic acid, and therefore could be synthesised by standard Fmoc peptide synthesis using pre-loaded Fmoc-Val Wang resin. The synthesis of this peptide using the Symphony automated peptide synthesiser required a number of attempted syntheses to optimise the route and maximise the yield. Initial attempts at the synthesis using resin with a substitution level of 0.39 mmol/g and standard Fmoc peptide synthesis (**Chapter 7, Section 7.2.1**) resulted in several truncated peptides and deletion peptides visible in the crude synthesis. The peptide was resynthesised using a low loaded Wang resin with a substitution level of 0.27 mmol/g, which improved the crude peptide purity. The quality of the crude peptide was improved

further by the use of pseudoproline dipeptides (Mutter *et al.* (1995)), and the optimised location of these is shown in **Figure 4.5**.

Figure 4.5: Sequence of C-terminal peptide fragment of RNase A showing location of pseudoproline dipeptides (underlined)

Peptide 4) X-DCRETGSSKYPNCAYKTTQANKHIIVACEGNPYVPVHFDASV

X = hydrazinoacetyl or hydroxylamine

The optimised synthesis of this peptide with an N-terminal hydrazinoacetyl group yielded 76 mg of purified peptide (20% yield from initial resin substitution) (**Chapter 7, Section 7.2.9.1**).

This approach to the synthesis of RNase A (strategy 1) was not continued due to the problems encountered when using the 1, 2 amino alcohol at the C-terminus of the peptide fragments. The alternative method of synthesis (strategy 2) using C-terminal hydrazides ligated to N-terminal glyoxal was tested using short peptides (**Chapter 4**), and then extended to the synthesis of peptide fragments of RNase A. This chapter details the synthesis of these peptides and the oxidation and chemoselective ligation reactions required for the synthesis of RNase A.

4.3 Synthesis of Peptide Fragments Bearing C-terminal Hydrazides

Synthesis of RNase A by ligation between an N-terminal glyoxal and a C-terminal hydrazide required a redesign of the peptide fragments. The locations of the chemoselective bonds as detailed in **Section 4.5.1** were kept the same, but the C-terminal aldehyde was replaced with a C-terminal hydrazide and the N-terminal hydroxylamine or hydrazide was replaced with a serine (**Figure 4.6**).

Figure 4.6: Primary sequence of RNase A, split into four peptide fragments, using strategy 2

KETAAAKFERQHMDSSTSAASSSNYCNQMMKSRNLTKDRCKPVNTFVHESLADVQAVCSQK
NVACKNGQTNCYQSYSTMSITDCRETGSSKYPNCAYKTTQANKHIIVACEGNPYVPVHFDASV

Peptide 1) KETAAAKFERQHMDSSTSA-NH-NH2
Peptide 2) SSSSNYCNQMMKSRNLTKDRCKPVNTFVHESL-NH-NH2
Peptide 3) SDVQAVCSQKNVACKNGQTNCYQSYSTMSI-NH-NH2
Peptide 4) SDCRETGSSKYPNCAYKTTQANKHIIVACEGNPYVPVHFDASV

X = amino acid replaced with serine
NH-NH2 = C-terminal hydrazide
pink = residues involved at active site

The three peptides with C-terminal hydrazides (**Figure 4.6**, peptides 1, 2 and 3) were synthesised on 2-chlorotrityl resin substituted with hydrazine (**Chapter 7, Section 7.2.8.1**). The synthesis of peptide 1 produced high quality crude peptide that did not show signs of amino acid deletions or truncated peptides. After HPLC purification, 125 mg of purified peptide was obtained. The initial substitution level of the hydrazide resin was not measured so the yield could not be calculated, but the yield was consistent with starting substitution levels of between 0.2-0.4 mmol/g.

Peptides 2 and 3 required optimisation of the synthesis to include pseudoprolines, and also extended coupling times of difficult reactions (**Figure 4.7, Chapter 7, Section 7.2.9.1-2**). The peptides were synthesised using the Prelude peptide synthesiser which enabled the synthesis of the peptides to include automated addition of the pseudoproline dipeptides.

Figure 4.7: Sequence of peptides 2 and 3 showing location of pseudoproline dipeptides (underlined) and difficult couplings (bold)

Peptide 2) SSSSNYCNQMMKSRNLTKDRCKPVNTFVHESL-NH-NH2
Peptide 3) SDVQAVCSQK**N**VACK**N**GQT**N**CYQSYSTMSI-NH-NH2

After purification, peptide 2 yielded 65 mg and peptide 3 yielded 38 mg. If the substitution of the hydrazide resin was assumed to be the same as for peptide 1, which was consistent with starting substitution levels of 0.2-0.4 mmol/g, then the yields for these two peptides was significantly lower. These peptides are longer and were more difficult to synthesise, so the lower yield is likely to be due to a lower quality crude peptide and also possibly due to a small amount of cleavage of peptide from the chlorotrityl linker during peptide synthesis.

The fourth peptide (**Figure 4.6**, peptide 4) has a C-terminal acid and so was synthesised on Fmoc-Val Wang resin using the same resin substitution and pseudoprolines outlined in the method in **Section 4.1.2**. The yield of this peptide was 54 mg (15% yield based on starting resin substitution, experimental details given in **Section 7.2.9.2**). This yield was slightly lower than for the similar peptide with an N-terminal hydrazinoacetyl (**Section 4.2**), as the peptide was repurified by HPLC to achieve the highest purity possible. The peptide with N-terminal hydrazinoacetyl would be used in chemoselective ligation reactions directly, but the peptide with N-terminal serine required oxidation to the glyoxal so particular effort was made to achieve a very high purity. The repurified peptide was analysed by analytical HPLC and gave a single peak (**Figure 4.8**)

All four peptide fragments were analysed by HPLC and ESMS, and the ESMS data are shown in **Figure 4.9 – Figure 4.12 (**Full HPLC and ESMS data is in **appendix II).**

Figure 4.8: HPLC of RNase A Peptide 4

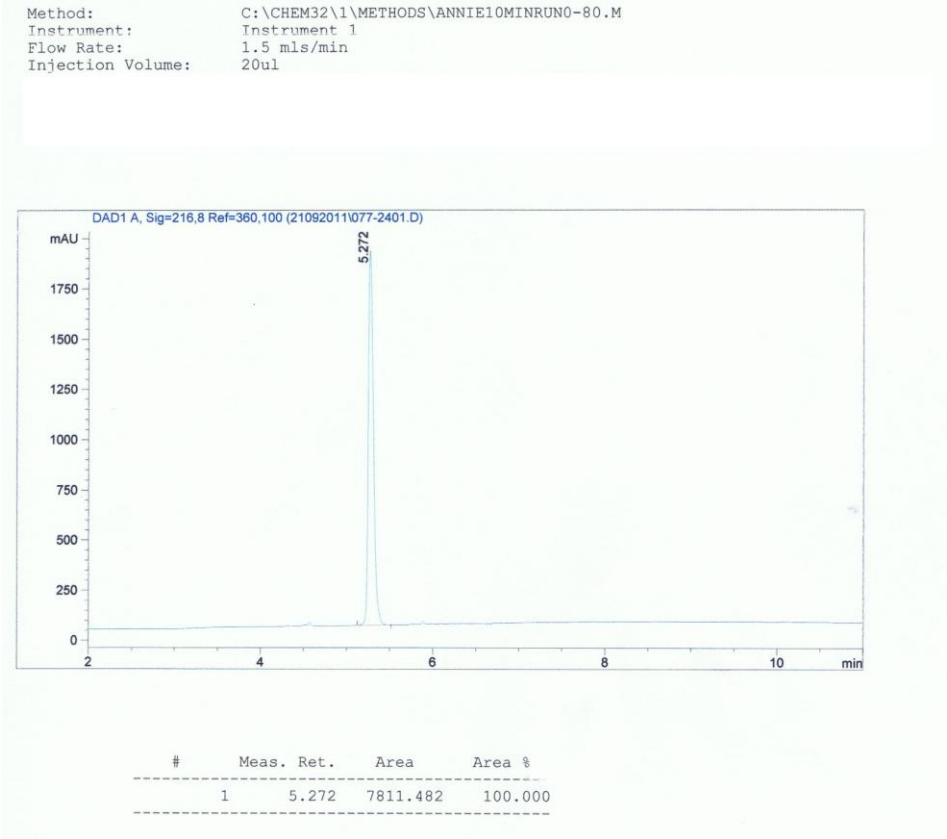
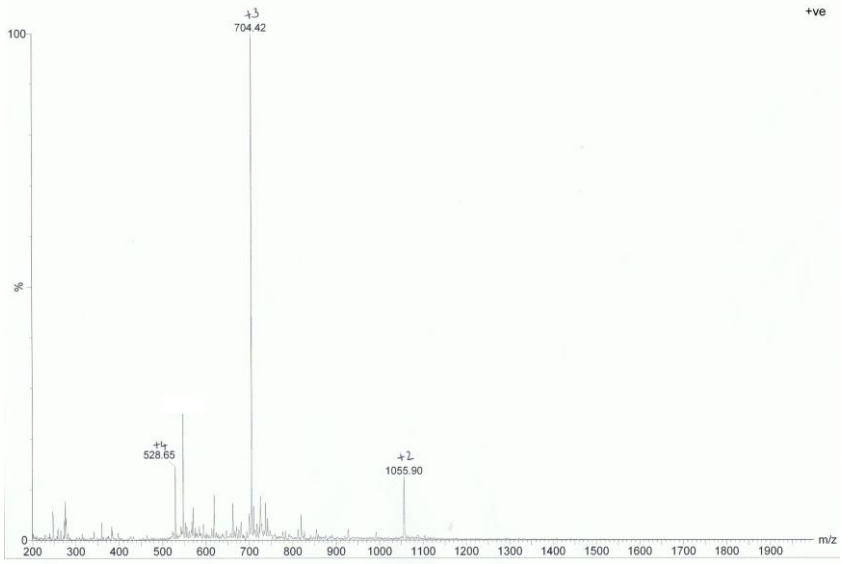
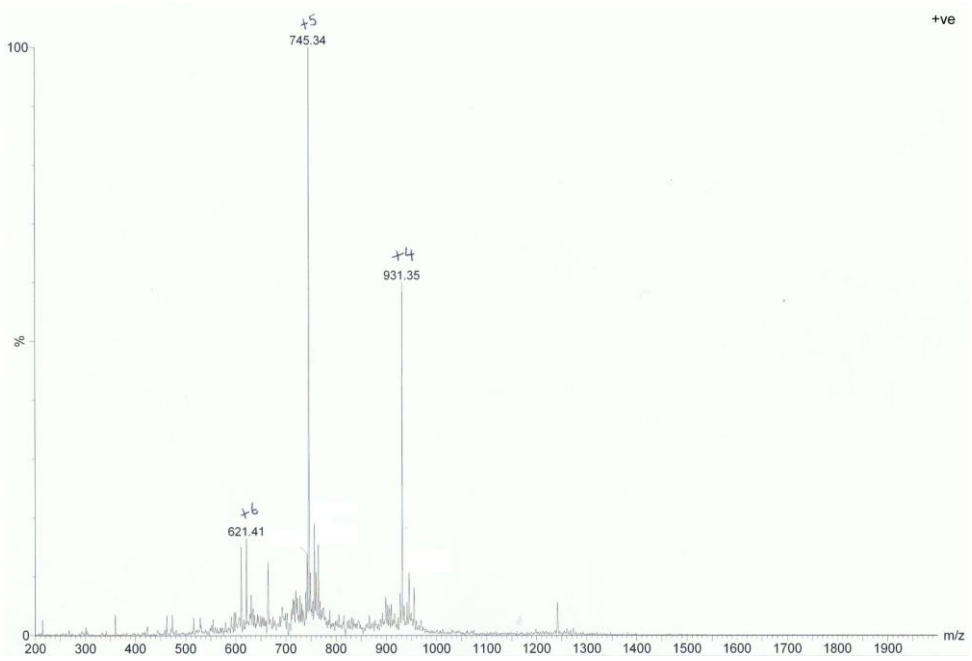


Figure 4.9: ESMS of RNase A Peptide 1



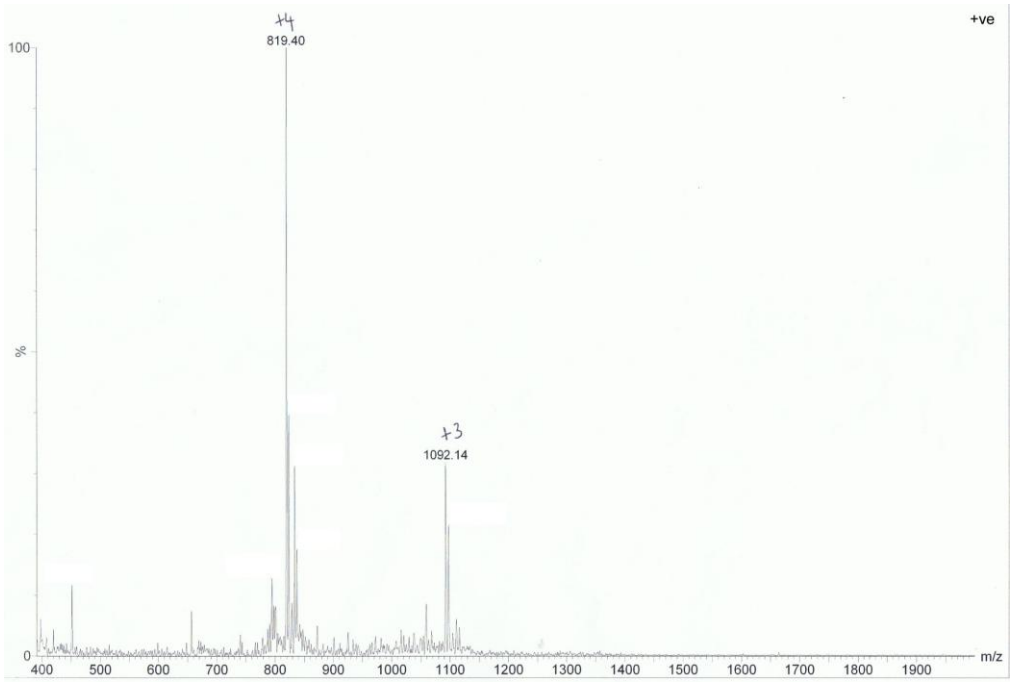
KETAAAKFERQHMDSSTSA-NH-NH2 [M + H] 2110

Figure 4.10: ESMS of RNase A Peptide 2



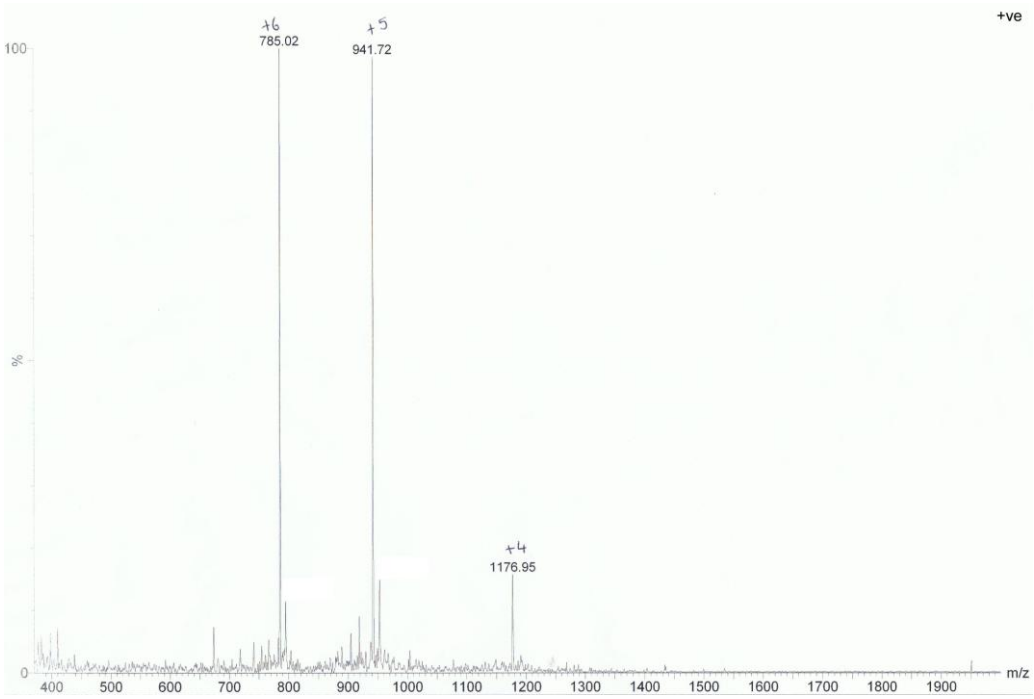
SSSSNYCNQMMKSRNLTKDRCKPVNTFVHESL-NH-NH2 [M + H] 3721

Figure 4.11: ESMS of RNase A Peptide 3



SDVQAVCSQKNVACKNGQTNCYQSYSTMSI-NH-NH2 [M + H] 3274

Figure 4.12: ESMS of Peptide 4 of RNase A

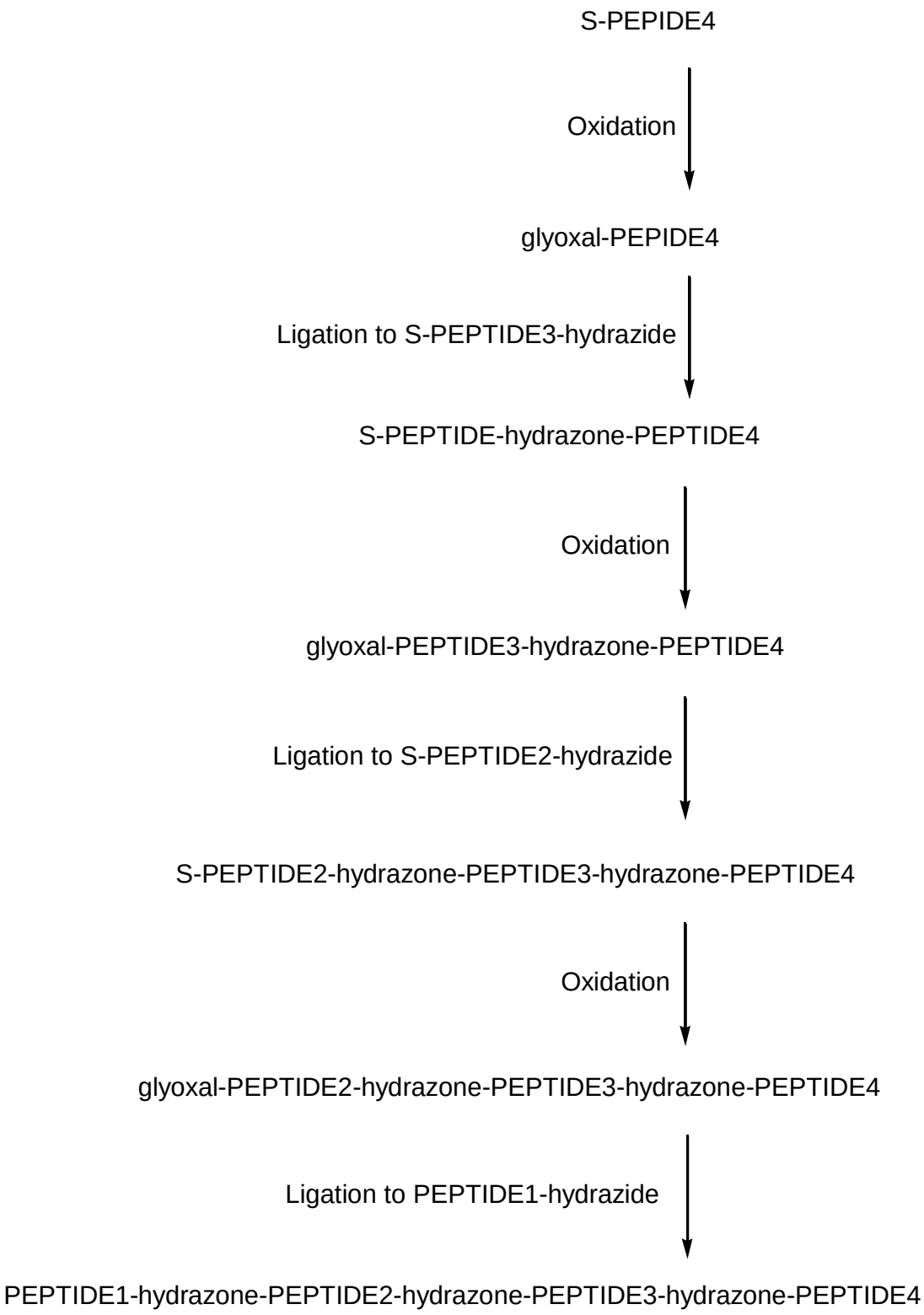


SDCRETGSSKYPNCAYKTTQANKHIIVACEGNPYVPVHFDASV [M + H] 4702

4.4 Chemoselective ligation of RNase A

Sequential chemoselective ligation of RNase A using the four peptides detailed in **Section 4.3, Figure 4.6** requires a cycle of oxidation, to produce an N-terminal glyoxal, followed by ligation of this glyoxal to a peptide bearing a C-terminal hydrazide and N-terminal serine, and then a repeat of the cycle of oxidation and ligation until all four peptides have been joined together. A schematic of this process is shown in **Figure 4.13.**

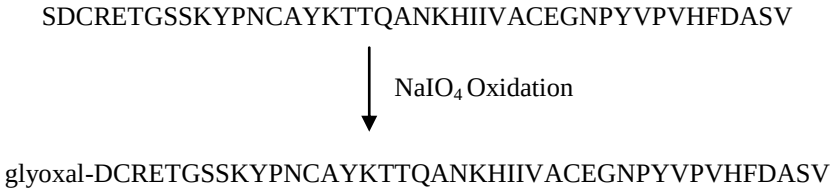
Figure 4.13: Proposed ligation of four peptide fragments by hydrazone ligation



4.4.1 Oxidation of Peptide Fragments of RNase A Containing N-terminal Serine

The first stage in the process of chemoselective ligation required the generation of an N-terminal glyoxal from the oxidation of peptide 4, which contains an N-terminal serine (**Figure 4.14**).

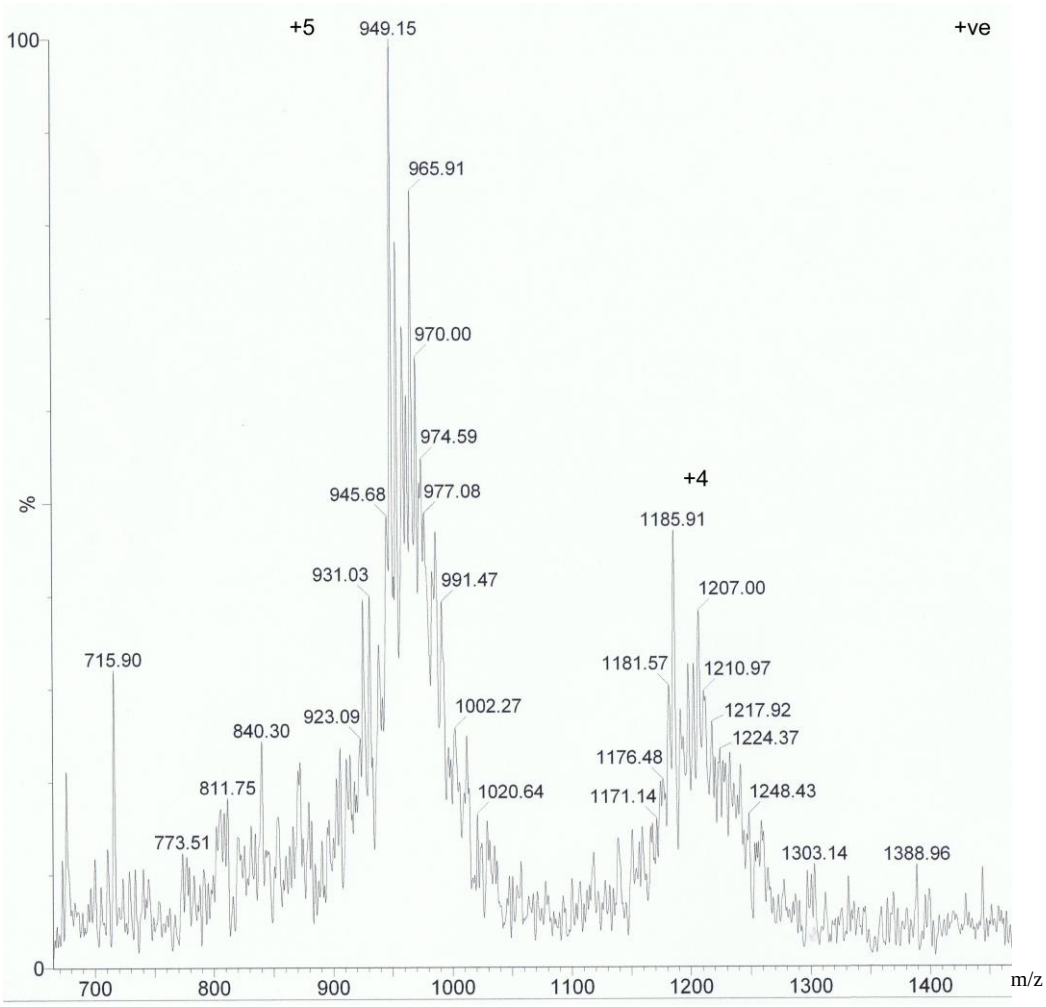
Figure 4.14: Oxidation of RNase A peptide 4



Peptide 4 was oxidised using the same conditions used for oxidation of the test peptide SLIGRL-NH2 (**Chapter 3, Section 3.5**). The reaction was quenched after 10 minutes with ethylene glycol (2 equivalents over $NaIO_4$), and a small sample was analysed by LCMS (**Figure 4.15**). The molecular mass of the glyoxal would be 4671, but ESMS of other glyoxal peptides shows the hydrated form which would be 4689, giving a +5 ion at 939 m/z. There was no peak visible at this m/z. The peptide showed multiple peaks in the ESMS of higher mass than expected, which corresponds to parent ions separated by 16 mass units. This suggested that as well as oxidising the serine to a glyoxal, other oxidation was occurring. It is difficult by ESMS to extrapolate the exact molecular mass of the parent ion, as ESMS only has an accuracy of +/- 1 mass unit, but the masses extrapolate approximately to parent ions with 3 x +16 to 8 x +16 additional mass units than the expected hydrated form of the glyoxal.

There are no methionines in the sequence (an amino acid prone to oxidation by sodium metaperiodate), so it was thought that perhaps the cysteines were being oxidised to cysteic acid.

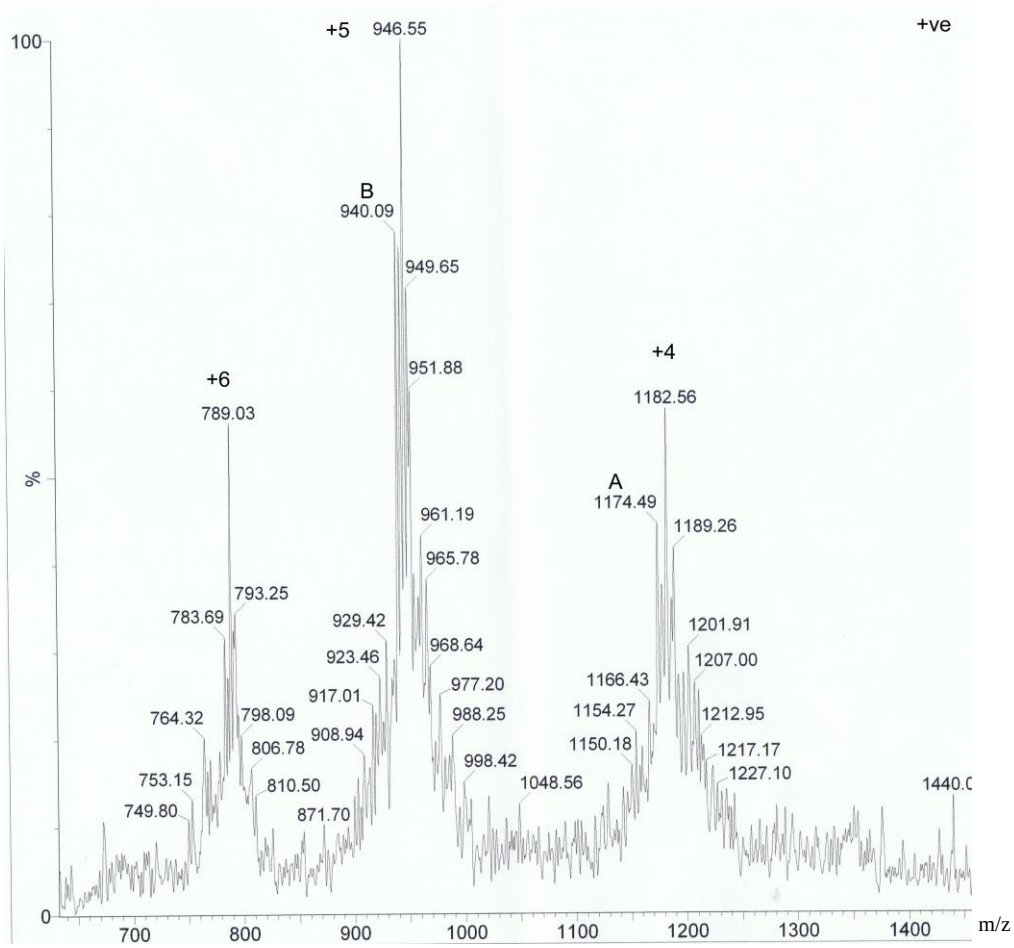
Figure 4.15 ESMS showing periodate oxidation of RNase A peptide 4 after 10 minutes



Peaks correspond to +4 and +5 ions of oxidised RNase A peptide 4 with multiple additional oxidations.

The oxidation reaction was repeated, but the reaction was quenched with ethylene glycol after 5 minutes. A sample was analysed by ESMS, and a similar pattern of peaks separated by 16 mass units was seen, except that peaks corresponding to the +5 and +4 ions of the hydrated glyoxal were visible along with the other oxidised products (**Figure 4.16**). Significant oxidation was therefore occurring even with short reaction times and so a solution to this oxidation needed to be found. As cysteine was the suspected source of the oxidation, protection of the cysteine during the oxidation and ligation reactions was proposed.

Figure 4.16: ESMS showing periodate oxidation of RNase A peptide 4 after 5 minutes



A = + 4 ion, glyoxal-peptide 4 [M + H_3O] 4689

B = +5 ion, glyoxal-peptide 4 [M + H_3O] 4689

4.4.1.1 Protection of Cysteine During Periodate Oxidation

It was decided that a trial would be performed, where just the N-terminal 10 residues of RNase A peptide 4 would be synthesised with and without a protecting group on the single cysteine present in this sequence. Periodate oxidation of these peptides would determine whether cysteine was responsible for the aditional oxidation seen, as a protecting group on cysteine would prevent its oxidation to cysteic acid. The

shorter peptide sequence would reduce the peptide complexity, by only containing one cysteine and also having a lower molecular mass making interpretation of MS data more straightforward.

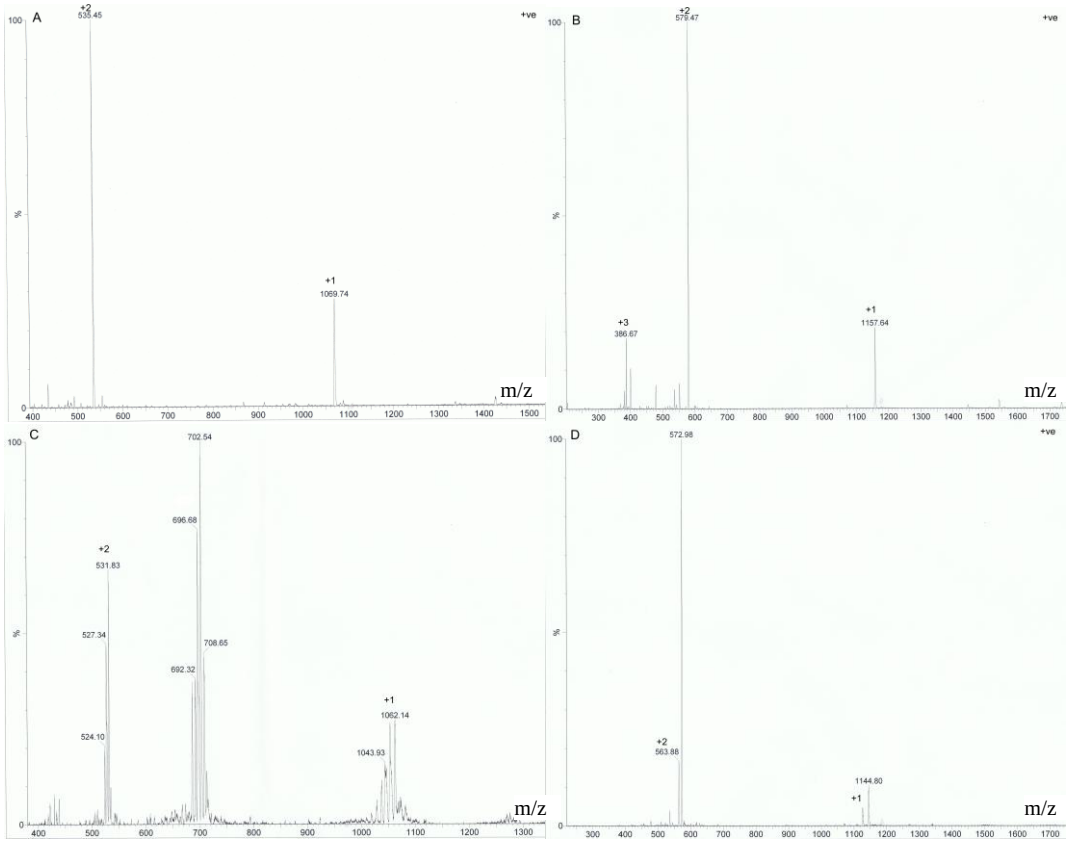
The protecting group chosen for protection of the cysteine was a thio-tertiary butyl group (S-tBu). This group forms a disulfide with the thiol side chain, and can be removed by a reducing agent such as dithiothreitol (DDT) or tris(2-carboxyethyl) phosphine (TCEP). The peptides SDCRETGSSK and SDC(S-tBu)RETGSSK were synthesised by standard Fmoc peptide chemistry using the Prelude peptide synthesiser. Both peptides were oxidised using $NaIO_4$ under the same conditions as used previously (**Chapter 3, Section 3.5**). The ESMS results of the oxidation are shown in **Figure 4.17**.

Oxidation of the S-tBu protected peptide proceeded to the glyoxal, as shown by ESMS (**Figure 4.17, D**), visible as the glyoxal and the hydrated form of the glyoxal. Oxidation of the peptide containing an unprotected cysteine produced peptide containing multiple products (**Figure 4.17 C**). There were no peaks visible corresponding to the glyoxal or hydrated form of the glyoxal, but there were multiple peaks which were +1 ions and +2 ions separated by 8 mass units, and also a cluster of peaks at around 700 mass units, separated by 6 mass units.

It was clear from these results that the unprotected cysteine residues were responsible for the multiple products seen during oxidation of the 4^{th} fragment of RNase A. As the focus of this investigation was on developing a route for sequential ligation, it was decided that the fragments of ribonuclease would be remade with S-tBu protection on the cysteines. No further investigation was made into the precise side reactions that were occurring with periodate oxidation in the presence of unprotected cysteine, but this is an area of study that can be investigated in the future.

The protection of cysteine with S-tBu allowed oxidation of the peptide to the glyoxal, and avoided the side reactions seen with unprotected cysteine.

Figure 4.17: ESMS of SDCRETGSSK with and without cysteine protection, before and after oxidation



A: SDCRETGSSK [M + H] 1069

B: SDC(S-tBu)RETGSSK [M + H] 1157

C: multiple unidentified products (glyoxal-DCRETGSSK [M + H] 1038 [M + H_3O] 1056 not visible)

D: glyoxal-DC(S-tBu)RETGSSK [M + H] 1026 [M + H_3O] 1144

4.6 Peptide Fragments of RNase A with Protected Cysteines

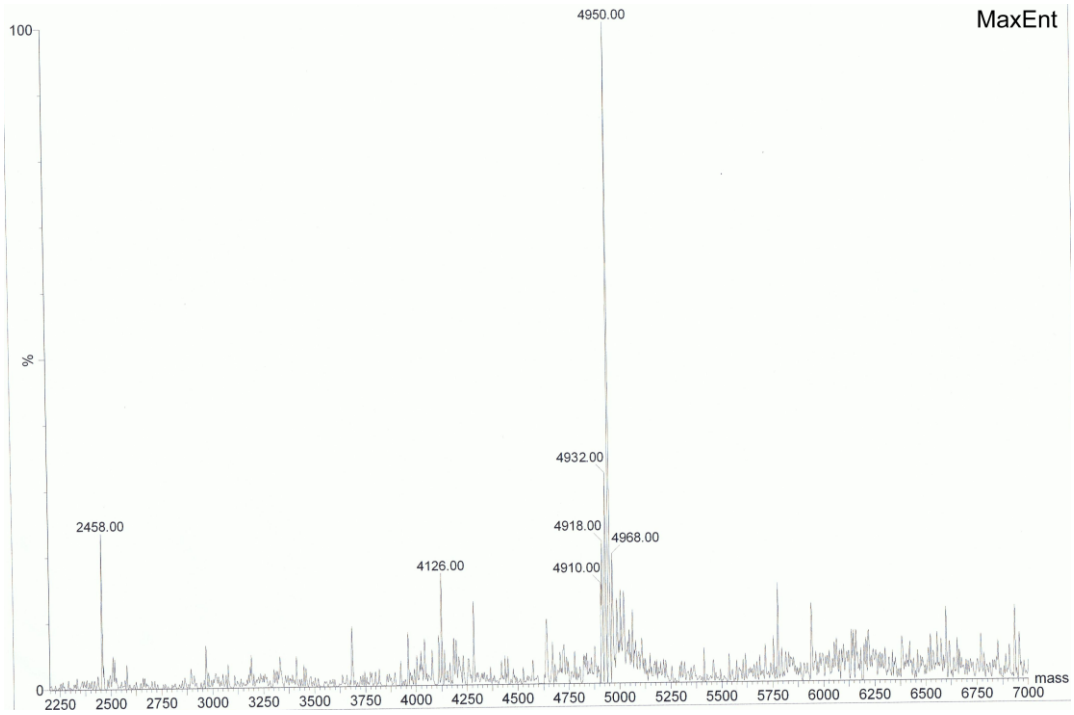
Three of the four peptide fragments of RNase A contained cysteine (**Figure 4.6**) and so these were resynthesised using Cys(S-tBu). The Prelude automated peptide synthesiser was used as this has capacity for up to 7 additional amino acids in addition to the 20 standard Fmoc amino acids. This meant that Cys(S-tBu) and the psuedoprolines required for the synthesis could be incorporated by the synthesiser,

and so no manual synthesis was required. The peptides were cleaved and purified as described previously (**Section 4.3**). As before the yields of the two peptides with C-terminal hydrazide (Peptides 2 and 3) could not be calculated as the initial resin substitution was not measured. Peptide 4 was produced in 27% yield from starting resin substitution (**Chapter 7, Section 7.2.9.4).** Peptides eluted at a later retention time by HPLC, which is consistent with more hydrophobic peptides due to the protecting groups on the side chains of the cysteine residues. Peptide 3 was particularly hydrophobic and difficult to purify due to its late retention time.

4.6.1 Oxidation of Peptide 4 of RNase A with S-tBu Protected Cysteines

Peptide 4 of RNase A protected with S-tBu on the three cysteines was oxidised to the N-terminal glyoxal. The peptide was not soluble in imidazole hydrochloride buffer, but was soluble in DMSO and so the peptide was first solubilised using a minimum quantity of DMSO and then diluted using imidazole buffer. The presence of DMSO did not affect the oxidation reaction and oxidation with $NaIO_4$ proceeded smoothly to yield the glyoxal (**Figure 4.18**).

Figure 4.18: Maximum Entropy plot for glyoxal of Peptide 4 protected with S-tBu



glyoxal-DC(S-tBu)RETGSSKYPNC(S-tBu)AYKTTQANKHIIVAC(S-tBu)EGNPYVPVHFDASV

[M + H] 4932

[M + H_3O] 4950

4.6.2 Chemoselective Ligation with S-tBu Protected Cysteines

Ligation was attempted between the glyoxal of peptide 4 and the hydrazide of peptide 3, with both peptides protected at the cysteines with S-tBu. Both peptides were very difficult to solubilise in NaOAc buffer and required use of DMSO to initially dissolve the peptides. The reaction was stirred overnight and the peptides had precipitated from solution. The reaction was resolubilised and a sample was analysed by ESMS, but only starting peptides were visible with no sign of ligated product.

Chemoselective ligation uses unprotected peptides as these are generally soluble in aqueous media. The use of S-tBu protection on the cysteines increased the hydrophobicity of the peptides such that they were no longer soluble in aqueous

buffer, and so an alternative protection for the cysteines was investigated that would be more aqueous soluble.

4.6.3 Acetamidomethyl as Alternative Cysteine Protection

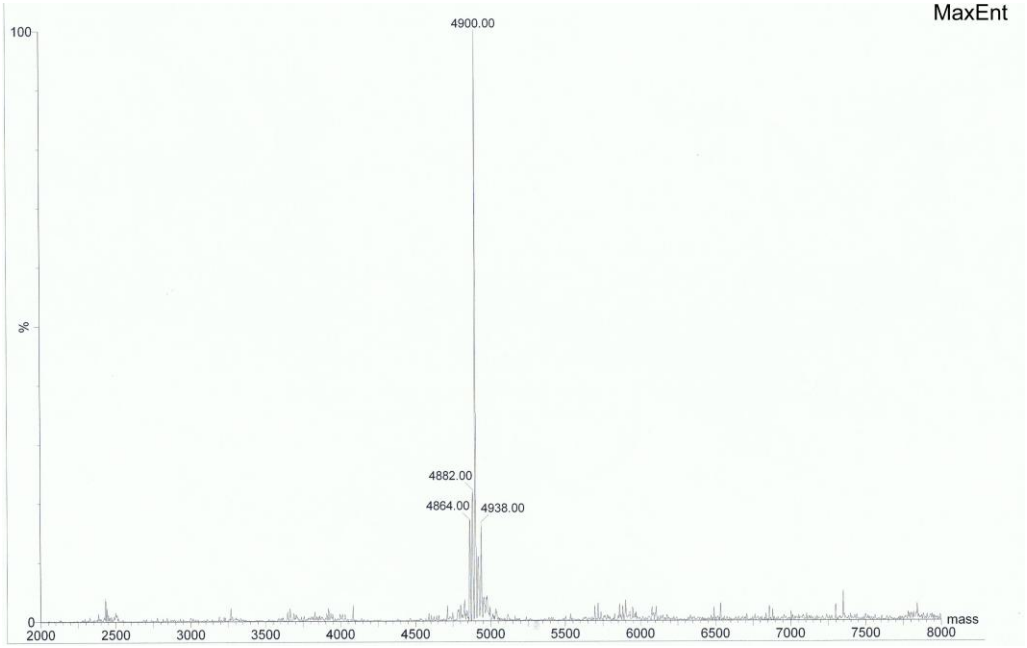
A common cysteine protection, usually used for forming disulfide bonds between cysteines is the acetamidomethyl group (ACM). ACM is less hydrophobic than S-tBu and so this was used as an alternative protection for the cysteines. The conditions for removal of this protection using I_2 forms a disulfide between two cysteines. Removal of the 8 ACM groups from the final ligated product would be likely to cause mixed disulfide formation, so these would need to be reduced using TCEP and the protein reoxidised to form the correct disulfides. For this reason ACM had initially been avoided. The other common cysteine protecting group is tBu but this is very difficult to remove from cysteine, and would be likely to have similar hydrophobicity problems as seen with S-tBu, therefore ACM was the only practical alternative.

The three peptide fragments were remade using Cys(ACM) and the same synthesis procedure as detailed in **Section 4.6** and **Chapter 7, Section 7.2.9.4**. Solubilisation of the peptides was easier that with S-tBu protection and the peptides eluted with an earlier retention time during purification by RP-HPLC.

4.6.4 Oxidation of Peptide 4 of RNase A with ACM Protected Cysteines

Peptide 4 of RNase A protected with ACM on the three cysteines was oxidised to the N-terminal glyoxal. The peptide was soluble in imidazole hydrochloride buffer and so no DMSO was required to solubilise the peptide. Oxidation with $NaIO_4$ proceeded smoothly to yield the glyoxal (**Figure 4.19**).

Figure 4.19: Maximum Entropy plot for glyoxal of Peptide 4 protected with ACM



glyoxal-DC(ACM)RETGSSKYPNC(ACM)AYKTTQANKHIIVAC(ACM)EGNPYVPVHFDASV

[M + H] 4882

[M + H_3O] 4900

4.6.5 Chemoselective Ligation with ACM Protected Cysteines

Ligation was attempted between the glyoxal of peptide 4 and the hydrazide of peptide 3, with both peptides protected at the cysteines with ACM. The glyoxal was easily solubilised in NaOAc buffer (pH 4.6), but peptide 3 required the use of DMSO to initially dissolve followed by dilution with NaOAc buffer. Ligation was tried at a number of different peptide concentrations and aniline concentrations but no ligation was achieved in any of these reactions. It was possible that the presence of DMSO was affecting the ligation, and so alternative solvents were also tried, including dimethylformamide (DMF) and hexafluoroisopropanol (HFIP). Use of DMF resulted in peptide precipitating from solution overnight. Use of HFIP maintained peptide solubility, but still did not result in any peptide ligation visible by ESMS.

4.7 Conclusion

Peptide fragments of bovine pancreatic ribonuclease A have been synthesised with N-terminal serine as a masked glyoxal and C-terminal hydrazide, for use in sequential chemoselective ligation. This method enabled the peptides to be synthesised in high yield and purity.

Periodate oxidation of the C-terminal peptide fragment gave multiple products separated by 16 mass units which suggested oxidation was occurring as a side reaction. It was thought that cysteine could be responsible for this oxidation and so a shorter peptide fragment containing just a single cysteine and N-terminal serine was synthesised to test this theory. Protection of the cysteine with a S-tBu group allowed successful oxidation of the N-terminal serine to the glyoxal, whereas the corresponding peptide with an unprotected cysteine produced multiple products with no glyoxal visible by ESMS.

The peptide fragments were resynthesised using S-tBu protection for the cysteines. Oxidation of the C-terminal peptide fragment proceeded to the glyoxal without the additional oxidation products seen with the unprotected version of the peptide. Ligation could not be performed using these peptides due to poor peptide solubility in aqueous buffer, and so an alternative cysteine protection using ACM was tried. Peptide solubility was improved, particularly for peptide 2, but peptide 3 still required the use of organic solvent to solubilise. No ligation was possible between these peptides, and so further work is required to improve peptide solubility, either by use of alternative cysteine protection, or perhaps through the use of a solubilising peptide tag. This is discussed further in **Chapter 6**.

During the time of this study, an alternative ligation approach using sequential native chemical ligation (NCL) has been used to synthesise the active protein by Fmoc synthesis in several fragments (Boerema (2007). The presence of the cysteines in that case provided the sites for ligation, and enabled synthesis of RNase A that was identical to the native protein. In the study presented in this thesis the presence of cysteines in the sequence caused problems with oxidation of peptide fragments. This

was overcome by the use of protecting groups, but produced the problem of peptide solubility which prevented ligation occurring. It may be that the problems caused by cysteine are not surmountable and that this method of chemoselective ligation can only be applied to proteins that do not contain cysteine. However, peptides containing cysteine now have a method of synthesis, sequential NCL, and so it is only in peptides not containing cysteine, or perhaps where the positions of the cysteines are not suitable sites for ligation that an alternative method of ligation is required. This is discussed in further detail in **Chapter 6.**

Chapter 5: Molecular Dynamics Simulations of Bovine Pancreatic Ribonuclease A

5.0 Introduction

Non-amide chemoselective ligation produces quite different structures to a natural peptide bond. In the chemoselective bonds studied for this thesis, the peptide backbone is one bond length longer than the natural peptide bond, and one or more amino acid side chains are lost or replaced with a different chemical group (**Chapter 3, Section 3.1**). Chemoselective ligation between an N-terminal glyoxal and C-terminal hydrazide produces a ligated product that closely resembles the natural peptide (**Chapter 3, Figure 3.1**). However there are still major structural differences, which when incorporated into a protein may have an effect on protein structure and function. The objective of this study was to model this chemoselective bond using a molecular dynamics (MD) simulation, and use these data to predict the implications of these non-amide bonds to protein structure and function, by studying the effect of these bonds on the structure of bovine pancreatic ribonuclease A (RNase A).

This chapter describes the basic theory behind MD calculations and the constraints that are required to allow calculations to be performed in hours, rather than days or weeks. The chapter then focuses on MD simulations of native ribounuclease and ribonuclease modified to contain chemoselective bonds. The structures were superposed and calculations performed to measure the similarity between the structures, allowing predictions to be made about the effect of chemoselective bonds on the structure of proteins.

5.1 Theory of Molecular Dynamic Simulations

Molecular dynamics, the computer simulation of interactions of atoms over time, using equations to approximate the physical interactions, has developed into an

exceptionally powerful tool for the simulation of complex molecular systems. The ever increasing computational power and availability of increasingly powerful software packages mean that molecular dynamic simulations have moved from the realm of the computational chemist into an accessible tool for all chemists. The software package available at the University of Surrey and used in this study is called Molecular Operating Environment (MOE) (MOE.2007.09, Chemical Computing Group Inc[7]). This section briefly describes the theory behind MD calculations and energy minimisation calculations required to perform an MD simulation.

5.1.1 Forcefield

A molecular dynamics simulation takes a system of atoms in defined positions and applies a forcefield (potential energy) between each atom, at a series of timesteps. The forcefield is an equation that approximates the total potential energy of the system. This is usually split into bonded atoms, with the forces approximated by a simple harmonic oscillator for bond length, with additional terms for changes in angles and torsions, and non bonded atoms (Van der Waals and electrostatic charges).

AMBER (Assisted Model Building and Energy Refinement) is a set of forcefields developed for the simulation of biomolecules such as proteins by Cornell *et al.* (1995). There are many versions of Amber and the version available within MOE at the time of this study was AMBER99. The general equation for the AMBER force field has the following form:

$$V = \sum k_b (l - l_0)^2 + \sum k_a (\theta - \theta_0)^2 + \sum \frac{1}{2} V_n [1 + \cos(n\phi - \phi_0)]$$
$$+ \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \{ [\frac{Aij}{R_{ij}^{12}} - \frac{Bij}{R_{ij}^{6}}] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \}$$

This equation calculates the total potential energy of a system V as the sum of four terms. The four terms in the equation describe the following four interactions:

1: $\sum k_b (l - l_0)^2$

k_b = force constant

l = bond length

l_0 = ideal bond length

the sum of all covalently bonded atoms in the system, modelled in AMBER as simple harmonic motion, so each covalent bond is modelled as a simple harmonic oscillator;

2: $\sum k_a (\theta - \theta_0)^2$

k_a = force constant

θ = bond angle

θ_0 = ideal bond angle

the sum of all angles, which calculates the energy due to the alignment of the orbitals in the covalent bonds also modelled as a simple harmonic oscillator;

3: $\sum \frac{1}{2} V_n [1 + \cos(n\phi - \phi_0)]$

V_n = barrier to rotation

n = periodicity

ϕ = torsion angle

ϕ_0 = phase factor

the sum of all torsions, which calculates the energy due to twisting of bonds, modelled as a periodic Fourier series, as bonds can rotate through 360°;

4: $$\sum_{j=1}^{N-1}\sum_{i=j+1}^{N}\{[\frac{Aij}{R_{ij}^{12}}-\frac{Bij}{R_{ij}^{6}}]+\frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}\}$$

for two atoms i and j

A and B = Van der Waal constants

R = non-bond distance (Van der Waals)

q = atomic charge

ε_0 = dielectric constant

r = non-bond distance (Electrostatics)

the sum of all non-bonded atoms. This is split into two parts, the first representing Van der Waals interactions using the Lennard-Jones potential model (Lennard-Jones, 1924)), and the second representing electrostatic charges. The AMBER forcefield models the electrostatic charges as excess or deficit of the sum of the total charge from the protons and electrons as a single partial point charge at the atom centre, using Coulomb's law.

The AMBER forcefield requires a set of parameters (standard bond lengths, angles, torsions and charges). A specific parameter set called ff99 (Wang *et al.* (2000)) for modelling of organic and biological molecules such as peptides and proteins was used in this study.

5.1.2 Energy Minimisation

Energy minimisation is required before a molecular dynamics simulation is run. This adjusts the positions of all the atoms in the system until a local energy minimum is found, which ensures that no atoms are positioned too closely together. It is essential that this is run prior to the molecular dynamics simulation; otherwise the system may be started in a very unstable energy state, which causes the whole system to fall apart once the simulation is run.

An energy minimisation calculation uses the principle that the forces on the system are equal to the negative gradient of the total potential energy. The calculation is performed on a static structure, i.e. the structure where temperature = 0 K. The potential energy of the system is calculated from the position of all the atoms, and then the confirmation of the atoms is changed in the direction of the forces on the atoms, which decreases the total potential energy of the system. The change in energy is plotted on an energy surface, and a minimum is found when the gradient of the potential energy is zero. The algorithms used for energy minimisation do not necessarily locate the global minimum of the system, as the calculations drive the energy of the system downwards (calculations would need to go upwards to move out of a local minimum), however a local minimum energy state is usually of sufficiently low energy that a MD simulation can be run.

5.1.3 Molecular Dynamics Calculations

Molecular dynamics most commonly uses classical mechanics as the basis of the calculations. Quantum mechanics can be used as a basis for calculations (defined as *ab initio* methods by Parr *et al.* (1950)). Quantum mechanical calculations model the electrons in a system as waves by approximating the Schrödinger equation, however the calculations are computationally complex and current computing power allows only small molecules containing tens of atoms to be studied in this way. For large biomolecules such as proteins, use of classical mechanics allows calculations to be performed on many hundreds of atoms.

The forcefields such as the AMBER forcefield described in **Section 5.1.1** models the atoms using geometric terms such as bond lengths and angles and the dynamics simulation uses Newtonian mechanics. Newton's second law of motion relates force to acceleration:

$$F = ma$$

The force on an atom can be related to the mass of each atom and its acceleration, expressed as the second derivative with respect to time.

$$\frac{F}{m} = \frac{d^2r}{dt^2}$$

Integrating this equation with respect to time gives the velocity of the atom:

$$\frac{dr}{dt} = \left(\frac{F}{m}\right)t + c = v$$

The force on each atom is calculated as the change in energy between its current position at time t and its position a small distance away at time t + δt. The position at time t + δt can be calculated from the position at time t using a truncated Taylor series:

$$r(t+\delta t) = r(t) + \frac{dr}{dt}\Delta t + \frac{d^2r}{dt^2}\Delta\frac{t^2}{2}$$

The truncation of this series is required for computational simplicity, but creates a source of error. This error can be partially addressed by using an algorithm known as the Verlet algorithm (Verlet (1967)). This involves two Taylor series expansions, one for r(t + δt) and the other for r(t - δt). Combining the two expansions gives an expression for position at time r(t + δt) that does not contain the velocity, so although the Verlet algorithm corrects some of the errors in the Taylor series expansion, velocities cannot be calculated directly.

An alternative, known as the leapfrog algorithm, developed by Hockney (1970), calculates the positions at integer time points and velocities at integer +1/2 time points such that they leapfrog over each other.

$$r(t+\delta t) = r(t) + \delta t v(t+\frac{1}{2}\delta t)$$
$$v(t+\frac{1}{2}\delta t) = v(t-\frac{1}{2}\delta t) + \delta t a(t)$$

The leapfrog algorithm therefore allows direct calculation of velocities and so was an improvement on the Verlet algorithm. The algorithm used within MOE is called velocity-Verlet (Swope *et al.* (1982)) and incorporates aspects of both the Verlet and leapfrog algorithms, calculating both position and velocity for the same timestep. The velocity-Verlet algorithm uses a half timestep to calculate the velocity but then uses this half timestep to calculate the velocity at the full timestep.

$$v(t+\frac{1}{2}\delta t) = v(t) + \frac{1}{2}at + \delta t$$
$$r(t+\frac{1}{2}\delta t) = r(t) + v(t+\frac{1}{2}\delta t)\delta t$$
$$v(t+\delta t) = v(t+\frac{1}{2}\delta t) + \frac{1}{2}a(t+\delta t)\delta t$$

The velocity-Verlet is the current best algorithm for compromise between computational efficiency and accuracy.

Another error in MD calculations is due to the assumption that acceleration and velocity remains constant between time t and time t + δt. This means that the longer the timestep the larger the error, as velocity and acceleration are not actually constant. For this reason the timesteps in MD must be short, in the order of 10^{-15} seconds, so that they are in the same timescale as the smallest scale vibrations in the system, i.e. bond vibrations.

5.1.3.1 Initial Velocities

In MD simulations the position of the atoms are usually taken from a crystal structure, and an energy minimisation is performed to adjust the position of the

atoms into a low energy state ready for the simulation to be performed. However an MD simulation also requires that each atom has an initial velocity, which is information that isn't known. The initial velocities are therefore assigned using the Maxwell-Boltzmann distribution.

$$f(V) = \left(\frac{m}{2\pi k_B T} \right)^{\frac{1}{2}} e^{-mv^2 / 2k_B T}$$

$f(V)$ = distribution of speeds V

m = mass

k_B = Boltzmann constant

T = temperature

A constant temperature is required throughout the simulation, however temperature tends to change during the simulation due to the non-bonded cut-off constraint (**Section 6.1.4.1**). The velocities are therefore periodically rescaled during the simulation to keep the temperature constant.

5.1.4 Constraints

A molecular dynamics simulation requires a number of constraints to be applied to the system. The time a simulation takes to run will depend on the number of atoms in the system, the size of the timestep and number of timesteps. For the simulation to be of relevance the total length of the simulation must be in the timescale of molecular motion, so ideally in the nanosecond range. However, even with modern computers simulations can take from days to weeks on multiple computers, so a series of constraints are required to reduce the computational complexity of the calculation. The constraints used in this study were a non-bonded cut-off, solvent model, periodic boundary conditions and the NTP ensemble, which are described below.

5.1.4.1 Non-bonded Cut-off

The most intensive part of the simulation is the calculation of the potential energy due to unbonded atoms, i.e. the calculation of Van der Waals and electrostatic interactions. These interactions are of importance throughout all space, but as distance increases the effect decreases as the reciprocal of distance, so short distance interactions have a greater importance than longer distance interactions. For this reason a non-bonded cut-off is applied which means that all non-bonded interactions at a distance greater than this cut-off are ignored. This compromise greatly reduces the size of the calculation without greatly affecting the outcome of the calculation.

5.1.4.2 Solvent

Simulations of biomolecules such as proteins are usually run in a solvent such as water, to most accurately model the system. There are two types of solvent that can be used, implicit or explicit. Implicit solvent models the solvent as a continuum rather than as discrete molecules. In explicit solvent models each solvent molecule individually interacts with the simulation giving a more realistic model than implicit solvent, but for computational efficiency the solvent is usually simplified to a rigid structure, therefore only interacting with the system via non-bonded interactions.

In this study water was used as an explicit solvent (to most closely simulate the laboratory environment), using a model called TIP3P (Jorgensen *et al.* (1983)), which models the water molecule as a rigid structure with the bond lengths and angles set to the average lengths and angles of a water molecule. Three point charges are applied to the three atoms, and the potential energy is calculated by a Lennard-Jones potential in a similar fashion to the non-bonded interactions in the AMBER forcefield (**Section 5.1.1**).

5.1.4.3 Periodic Boundary Conditions

Periodic boundary conditions (PBC) are applied, which means that although the simulation is run in a three dimensional box, or unit cell, molecules on one edge of the box can interact with those on the opposite side; indeed a molecule can disappear from one side of the cell and reappear on the other side with the same velocity. This avoids edge conditions, which can produce unusual results in the energy calculation. PBC usually uses the minimum-image convention, which means that each atom will only interact with the closest image of each other atom.

6.1.4.4 NTP Ensemble

An ensemble is a concept in statistical mechanics and thermodynamics, which is used to describe a system containing a large number of particles, such as an MD simulation. Four properties, pressure P, volume V, temperature T, and number of particles N are used. Three of these properties are held constant, allowing only one property to change.

The most relevant constraining system used in biomolecular molecular dynamics is known as the Isothermal-Isobaric ensemble or NTP. In this system N - number of atoms, T - temperature and P – pressure of the system are held constant, so only volume is allowed to change. This set of constraints most realistically represents the conditions of a laboratory experiment, and is the ensemble used in this study.

5.2 Root Mean Square Deviation as a Tool for Interpreting Structural Similarities

Root mean square deviation (RMSD) is a common measure of the similarity between two protein structures. Two protein structures of the same residue length are superposed and the coordinate position of each Cα atom that forms the backbone of the two structures is used to give the mean distance between each Cα atom. The root

of the square of the means gives a value in Å that can be used to compare structural similarities.

RMSD is calculated using the equation

$$RMSD = \left(\frac{1}{N} \sum_{i-1}^{N} \left(r_i - r_i' \right)^2 \right)^{\frac{1}{2}}$$

where r_i and r_i' correspond to the three dimensional coordinates of the Cα atoms for the two protein chains and N = number of residues. Within MOE, two protein sequences are superposed and aligned using a program called FITT (Finkelstein (1987)) which uses an algorithm to orientate the molecules so that the RMSD is minimised.

An RMSD value of zero would be produced for the superposition of two identical protein molecules. As the difference between the structures increases so does the RMSD value. A study comparing the RMSD of pairs of homologous proteins by Chothia *et al.* (1986) found that the RMSD was always <3 Å for homologous proteins, and so this figure has since been adopted as the figure for defining structural similarity between protein molecules.

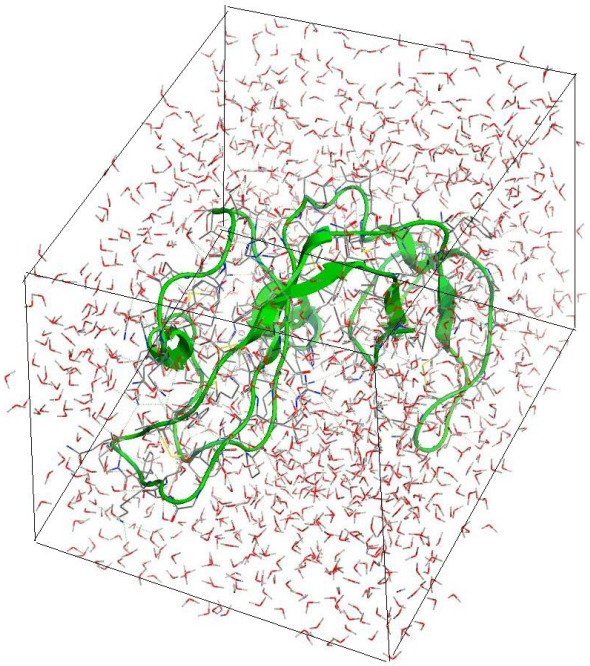
5.3 Molecular Dynamics Simulations

MD simulations were run on native RNase A and RNase A containing chemoselective bonds (**Chapter 7, Section 7.4.1**). MD was also performed on RNase A containing a single amino acid mutation of Cys to Ala, which allowed the method to be validated against published experimental results.

5.3.1 Molecular Dynamics Simulation of Native Bovine Pancreatic Ribonuclease A

A molecular dynamics simulation was first performed for native RNase A (**Chapter 7, Section 7.4.1.1**), which would allow subsequent simulations of ribonuclease containing chemoselective bonds to be compared to the native protein. **Figure 5.1** shows the structure of native RNase A placed inside a virtual box of water. The box size was made as small as possible to minimise the number of water molecules in the simulation, therefore minimising the computational complexity. After energy minimisation of this structure the MD simulation was run using 200 timesteps of 0.5 ps, to generate a database file with 200 entries representing 100 ps of simulation. This simulation took approximately 2 hours to run. 100 ps is a short simulation time, but as this simulation was of native ribonuclease with no modifications, it was only necessary to run a simulation until the energy of the system had stabilised, as any longer simulation would only produce a larger dataset giving the same information.

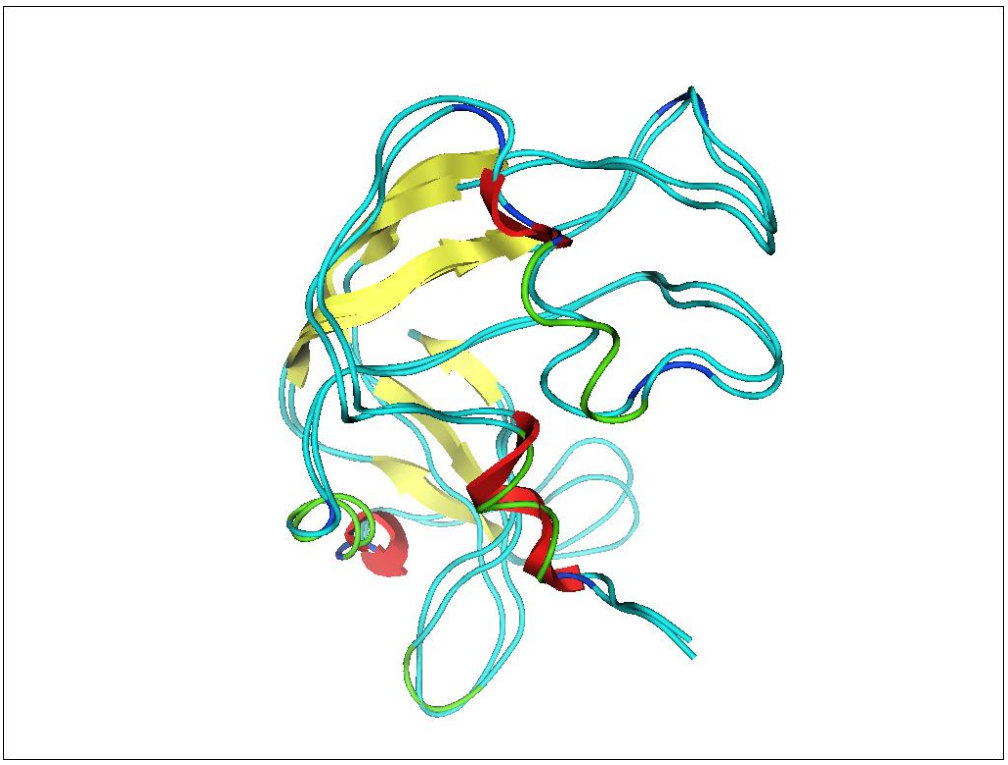
Figure 5.1: Native bovine pancreatic ribonuclease A solvated with H_2O



The final frame of the MD simulation (t = 100 ps) was loaded into MOE together with the energy minimised structure prior to running the simulation (**Figure 5.2**). Root mean square deviation (RMSD) calculations were made using the superpose function in MOE, which gave a figure of 1.51 Å. **Section 5.2** explains the theory and significance of RMSD as a tool for the comparison of superposed protein structures. The two structures contain identical atoms and bond connectivity. The only difference between the two structures is that a short MD simulation was run, allowing the atoms to shift from their energy minimised positions to final positions at the end of the simulation.

The two superposed structures are represented as ribbon cartoons, which show the backbone of each structure as a ribbon. The MOE software recognises regions of α-helix (red) and β-sheet (yellow) and loops (blue). Regions in green are recognised by the software as helical, but not of the exact α-helix conformation for the software to replace with the red ribbon. The process of MD has moved the atoms out of the rigid structures found in a crystal, and the final frame provides a snapshot of the molecule as it may exist at any one moment in solution.

Figure 5.2: Superposition of native RNase A before and after MD



RMSD = 1.51 Å

Similar RMSD calculations were perfomed by superposing the initial structure of ribonuclease before MD with two other timepoints through the simulation, at t = 50 ps and t = 75 ps. RMSD values of 1.45 Å and 1.10 Å were obtained.

A study by Santoro *et al.* (1993) compared the structures of 16 MD simulations of native RNase, and produced an average RMSD of 0.92 Å from the crystal structure, which is slightly smaller than the RMSD values calculated in this investigation.

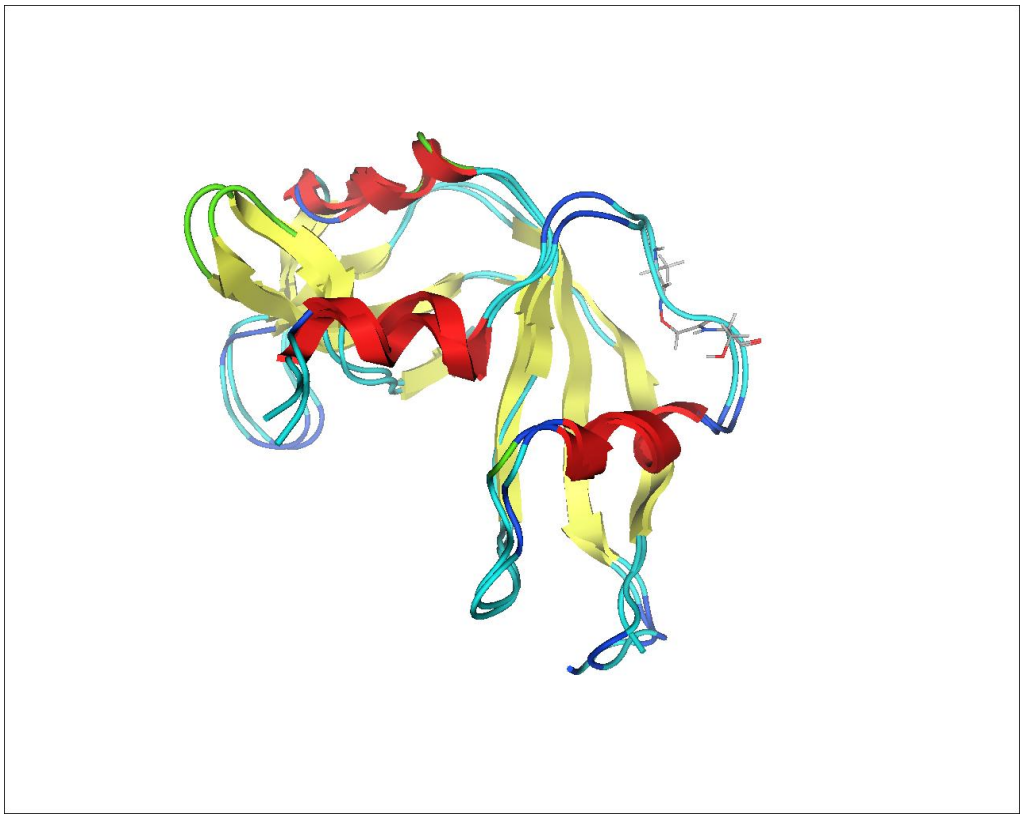
5.3.2 Replacement of Ligation Sites with Chemoselective Bonds

RNase A containing oxime and hydrazide chemoselective bonds were created within MOE and these were used in MD simulations as detailed below.

5.3.2.1 Replacement of Ala20 with an Oxime Bond

The structure of ribonuclease was created with Ala20 replaced with an oxime bond, formed through reaction of an N-terminal hydroxylamine with a C-terminal aldehyde (**Chapter 7, Section 7.4.1.2**). The energy minimised structure of native ribonuclease A was loaded into MOE and, and superposed with the modified version of ribonuclease containing the Ala20 oxime replacement. The molecules were selected in cartoon mode, which draws the backbone as a coloured ribbon showing regions of loops, alpha helices and beta sheets. As the oxime bond is not recognised as protein backbone by the software, this was represented with a stick structure. The overlaid structures after energy minimisation visually showed close structural similarity (**Figure 5.3**), although no RMSD calculations were performed on these structures, as this was a preliminary study to the main investigation.

Figure 5.3: Overlay of native ribunuclease and ribunuclease with Ala20 replaced with an oxime bond.



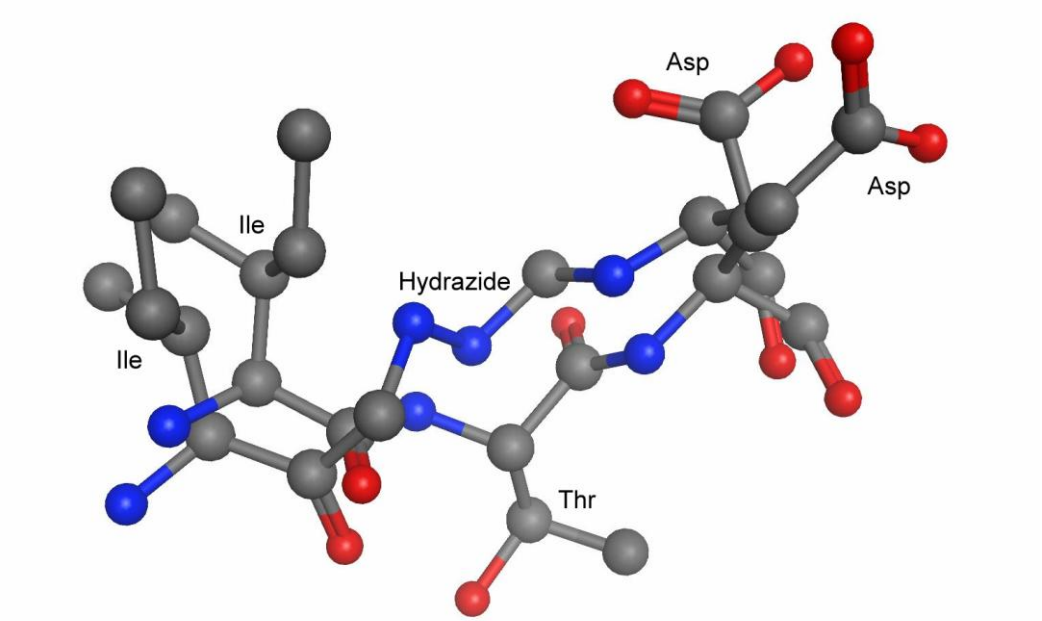
Whilst these molecular dynamic studies were being performed, the experimental synthesis of ribunuclease by chemoselective ligation was no longer continued using the oxime bond (**Chapter 3, Section 3.3)**, so no further modelling was perfomed using the oxime bond. The direction of the chemoselective bonds was reversed (**Chapter 3, Section 3.4)**, and all further modelling studies were performed using hydrazide bonds between an N-terminal glyoxal and a C-terminal hydrazide.

5.3.2.2 Replacement of Ala20, Ala52 and Thr82 with a Hydrazide Chemoselective Bond

Each site was in turn replaced with the structure of a hydrazide bond formed through reaction between an N-terminal glyoxal and a C-terminal hydrazide (**Chapter 7,**

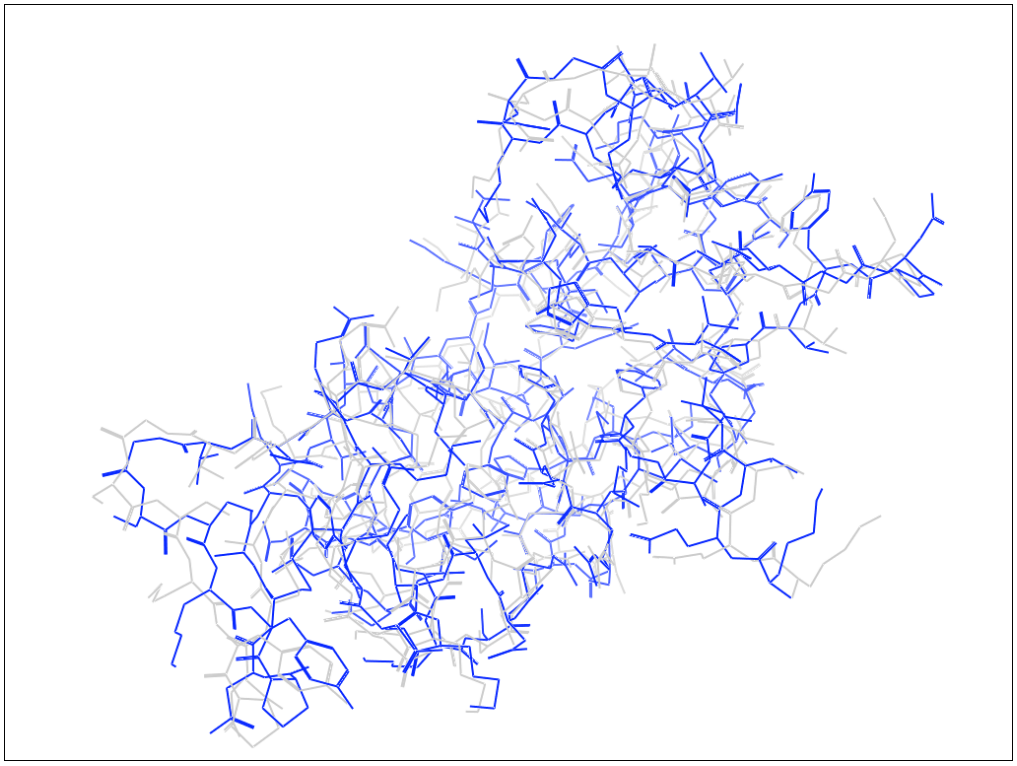
Section 7.4.1.3). **Figure 5.4** shows the structure of Ile81-Thr82-Asp83 in native RNAse A and in the ligated structure with a hydrazide bond.

Figure 5.4: Overlaid structures of Ile81-Thr82-Asp83 in native RNase A and RNase A containing a hydrazide chemoselective bond.



After each bond was replaced, the structure was run through an energy minimisation using the AMBER 99 forcefield, firstly just on the replaced atoms, and then using the proximity tool to identify all atoms close to the changed bonds. Once all three bonds had been replaced the whole system was then minimised once more. The structure was then superposed with the structure of energy minimised native ribonuclease (**Figure 5.5**). MOE was no longer able to recognise regions of α-helix or β-sheet using the ribbon mode of representing the structures. Therefore all structural comparisons of ribonuclease containing chemoselective bonds were converted to line drawings representing the carbon framework of the protein, as this was a clearer way to visualise the structures. RMSD calculations were made using the superpose function in MOE and gave a figure of 1.17 Å, which represents close structural corrrelation.

Figure 5.5: Native ribonuclease superposed with ligated ribonuclease before molecular dynamics simulation



Grey = native ribunuclease

Blue = ligated ribunuclease

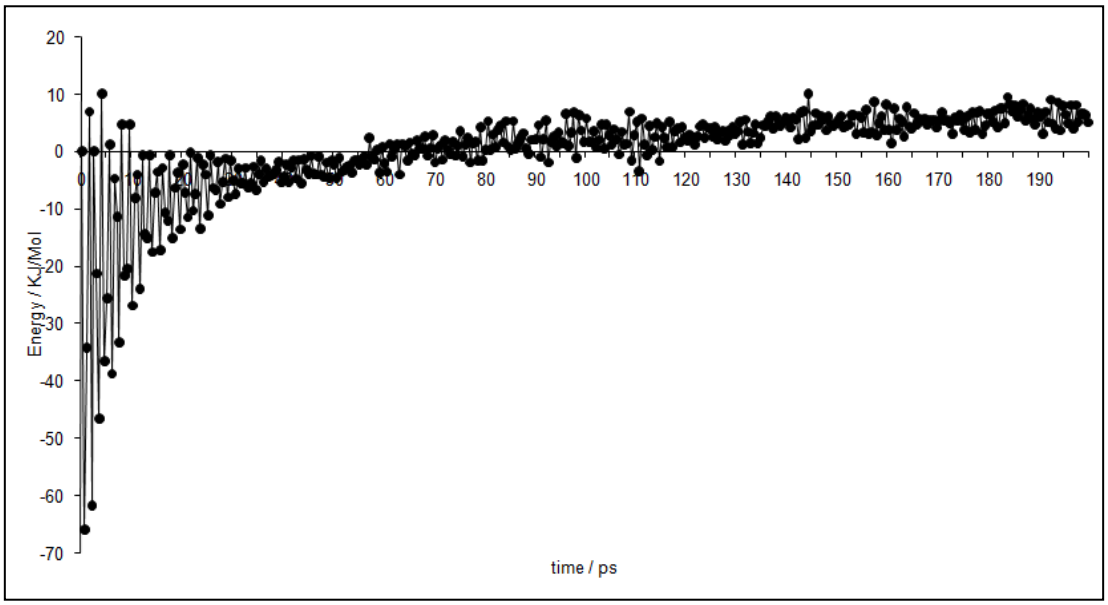
RMSD = 1.17 Å

5.3.2.3 Molecular Dynamics Simulation of Bovine Pancreatic Ribonuclease A containing Three Hydrazide Bonds

The molecular dynamics simulation was run on the structure of ribonuclease containing three hydrazide bonds using the same constraints used for native ribonuclease (**Chapter 7, Section 7.4.1.4**). **Figure 5.6** shows the total potential energy of the system plotted against time. For the first 30 ps of the simulation the energy of the system fluctuated quite dramatically which is due to the system being started in a high energy state, as energy minimisation calculations only find a local minimum energy state. This energy minimum allows the MD simulation to be run without the system falling apart due to calculation errors caused by unnaturally high

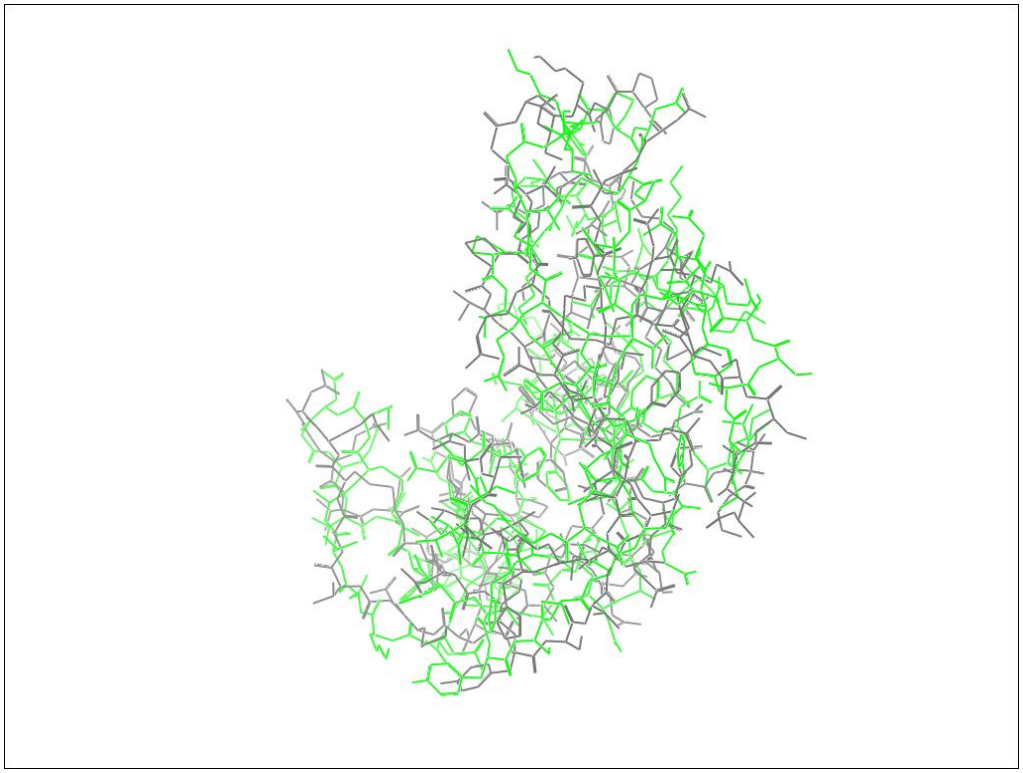
forces in the system. After approximately 50 ps the energy of the system stabilised, but showed a gradual overall increase in energy over the next approximately 100 ps. The final 50 ps of the simulation gave a stable energy trace, and so it was not necessary to run the simulation for longer timescales.

Figure 5.6: Energy plot (KJ/Mol) for MD simulation of ribonuclease containing hydrazide bonds



The structure was then superposed with the structure of native ribonuclease after molecular dynamics using the final frame of the MD simulation (t = 200 ps) (**Figure 5.7**). The calculated RMSD value was 2.20 Å. Any value under 3 Å is considered to be close structural similarity, and so a value of 2.20 Å represents a structure of comparable similarity to two homologous proteins, that is two proteins from different organisms sharing a common catalytic mechanism, but with differences in primary sequence. The results of the dynamic experiment for ribonuclease A containing three hydrazide bonds can be interpreted as producing protein with sufficient structural similarity to native ribonuclease that any effect on catalytic mechanism is likely to be negligible.

Figure 5.7: Native RNase A superposed with ligated RNase A after molecular dynamics simulation (t = 200 ps)



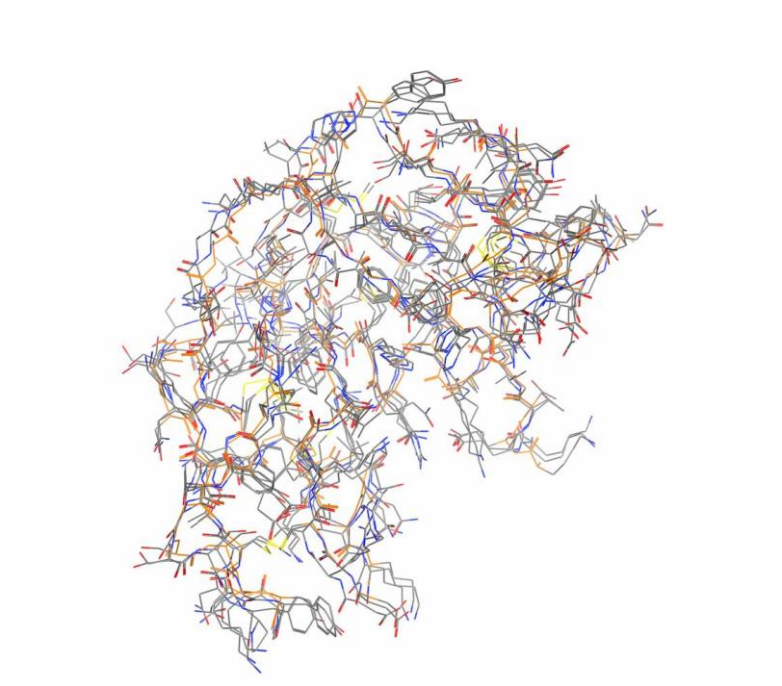
Grey = native ribunuclease

Green = ligated ribunuclease

RMSD 2.20 Å

Similar RMSD calculations were perfomed by superposing the initial structure of ribonuclease before MD with two other timepoints through the simulation, at t = 100 ps and t = 150 ps. RMSD values of 2.38 Å and 2.16 Å were obtained. **Figure 5.8** shows the superposition of the three structures produced at the timepoints; t = 100 ps, t = 150 ps and t = 200 ps.

Figure 5.8: Superposition of RNaseA containing three hydrazide bonds from three



different timepoints; t = 100 ps, t = 150 ps and t = 200 ps

5.3.2.4 Replacement of Cys 65 with Ala

Bovine pancreatic ribonuclease A has 4 disulfide bonds. Extensive studies have been performed to elucidate the folding and order of disulfide formation of ribonuclease, including a study by Talluri *et al.* (1994). In this investigation they removed a single disulfide bond between Cys65 and Cys72, either by reaction of these two thiols groups with iodoacetate, or replacement of cysteine with cysteamine to form des-[65-72] RNase A. They were following previous studies by Rothwarf *et al.* (1991) who used dithiothreitol as a reductant to partly denature RNase A. Their work had suggested the disulfide between Cys65 and Cys72 was the final of the four disulfides to form during the folding process. Talluri *et al.* (1994) compared NMR solution structures of des-[65-72] RNase A to the crystal structure of native RNase A, and obtained an RMSD of 1.53 Å. Their conclusion was that the close structural similarity between the two structures meant there was no high energy barrier to

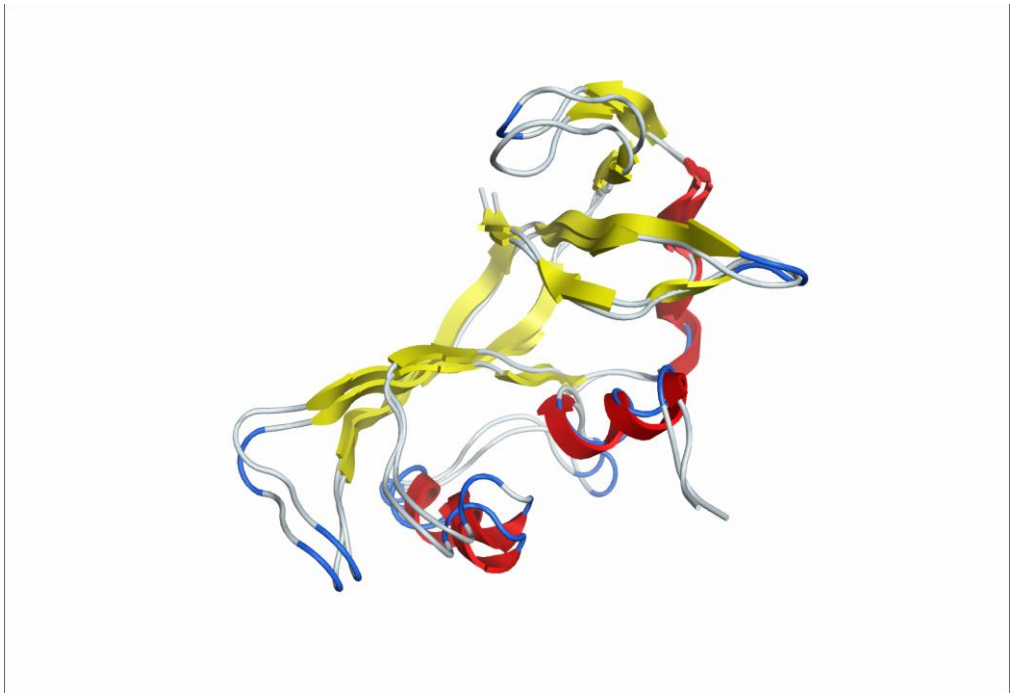
overcome during formation of the final disulfide, as only minor conformational changes would be required.

Des-[65-72] RNase A was chosen to test the validity of the MD simulations used in this study. Superposition of the structure of native RNase A with the structure of des-[65-72] RNase A generated through MD simulation should produce a value for the RMSD close to the value of 1.53 Å found by Talluri *et al.* (1994), from experimental NMR solution data.

The easiest way to produce a structure for Des-[65-72] RNase A within MOE was to create a single mutation of one cysteine that would prevent formation of the disulfide. Cys 65 was replaced with Ala (**Chapter 7, Section 7.4.1.5**), which removed the disulfide bond, as alanine has no thiol side chain for disulfide formation. The structure was energy minimised, and the MD simulation was run (**Chapter 7, Section 7.4.1.5**) using the same set of constraints as used for the previous dynamics experiments (**Sections 5.3.1 and 5.3.2.3, Chapter 7, Sections 7.4.1.1 and 7.4.1.4**).

The structures of native RNase A and Des-[65-72] RNase A after MD were superposed (**Figure 5.9**), using the final structure from the dynamic simulation at t = 200 ps. Ribbon structures were able to be used to represent the backbones of the structures as both molecules contain only natural amino acids and peptide bonds. The RMSD was calculated and gave a value of 1.29 Å.

Figure 5.9: Cys65Ala RNase A superposed with native RNase A



RMSD = 1.29 Å

The value of 1.29 Å calculated from the dynamic simulation was in close agreement with the value of 1.53 Å calculated by Talluri *et al.* (1994) from experimental data. It is therefore reasonable to assume that the MD simulations performed and RMSD values calculated in this study give a valid representation of the actual structure of RNase A containing chemoselective bonds.

5.4 Conclusions

Molecular dynamics simulations were run using bovine pancreatic ribonuclease A as an example protein, to compare the structures of native RNAse A to the structure of RNAse A containing chemoselective bonds. The study was limited to looking at the bonds used in the synthetic aspects of this thesis. The modelling studies were performed concurrently with the synthetic experimental studies, and so adaptations

were made to mirror the directions taken during progression of the experimental work.

Simulations of native RNase A and RNase A containing hydrazone chemoselective bonds were performed, and the structures superposed. RMSD calculations gave values lower than 3 Å which means the proteins had close structural similarity. This implies that the presence of these chemoselective bonds does not greatly affect the structure of the protein, and should therefore have little effect on the catalytic activity of the enzyme.

The validity of the technique was tested using a version of RNase A mutated to contain an alanine in place of a cysteine, removing a single disulfide bond. The superposition of mutated and wildtype RNase A from the MD simulations gave similar RMSD values to published experimental RMSD values from NMR studies.

MD simulations on proteins containing chemoselective bonds have not been performed before, and so this study provides an important foundation for this type of study. The results of the MD simulations suggest that the method provides a realistic model of the structure and dynamics of a protein containing chemoselective bond. Further work is required, for example to look at the effect of positioning chemoselective bonds in positions that are predicted to affect the structure of the protein. This is discussed in more detail in **Chapter 6**.

6.0 Conclusions and Future Work

Chemoselective ligation is a technique that can be used to join peptides together, and enable the total chemical synthesis of proteins that cannot be made in a single chain by solid phase peptide synthesis. The work presented in this thesis evaluates sequential chemoselective ligation as a method of joining more than two peptides, which can allow the production of larger peptides and proteins. A large number of different methods of chemoselective ligation have been reported, but this investigation has focused on the most common method, which is the reaction between an aldehyde and either hydroxylamine or hydrazine. In **Chapter 1**, the area of solid phase peptide synthesis for the total synthesis of peptides and proteins was introduced, and several of the chemoselective ligation methods that have been used in recent years for the synthesis of proteins have been described.

The basis of initial studies was an extension of the work of Sharma and Broadbridge (Broadbridge *et al.* (2001)), which used a 1, 2-diol as a masked aldehyde, for the sequential ligation of several peptide fragments. This investigation started with the aim of increasing the yields both of individual peptide fragments and of the ligated peptide products, which would allow the technique to be used for the commercial production of peptides and proteins. A 1, 2-amino alcohol was investigated as a replacement for the 1, 2-diol, as an amine is more reactive than an alcohol towards the solid support used during synthesis, which was predicted to improve peptide yields. This required development of a synthetic route for conversion of the carboxylic acid group of an amino acid to an amino alcohol. In **Chapter 2** the use of a 1, 2-diol as a masked aldehyde was reviewed, and then a synthetic route for the synthesis of 1, 2- amino alcohol derivatives of Fmoc protected amino acids was described, for use in the solid phase synthesis of peptides bearing C-terminal amino alcohols. The method of synthesis developed was successful, and was used to synthesise amino alcohol derivatives of several Fmoc protected amino acids in good yield. Unfortunately, a

number of problems, detailed in **Chapter 2,** meant that these derivatives were not suitable for the synthesis of peptides bearing C-terminal amino alcohols or aldehydes, and so an alternative approach was investigated, swapping the amino alcohol to the N-terminus of the peptides, by using an N-terminal serine as a masked glyoxal.

In **Chapter 3** the historical use of serine or threonine as an N-terminal masked aldehyde was reviewed. The structures of the chemoselective bond were compared between the initial approach of a C-terminal aldehyde and N-terminal hydrazine or hydroxylamine, and the alternative approach of an N-terminal aldehyde and C-terminal hydrazide. Short test peptides with either an N-terminal glyoxal or C-terminal hydrazide were synthesised in high yield and purity, and ligation of the peptides demonstrated the suitability of the technique for use in sequential chemoselective ligation. This investigation demonstrated the first use of aniline as a nucleophilic catalyst in hydrazone chemoselective ligation to dramatically increase ligation reaction rates. Similar results from experiments performed at the same time as this investigation have been published by Dirksen *et al.* (2008).

Bovine pancreatic ribonuclease A (RNase A) was chosen as an example protein to extend the method of ligation between an N-terminal glyoxal and C-terminal hydrazide to the total synthesis of a protein by sequential chemoselective ligation. A background to RNase A structure and catalysis was described in **Chapter 1**. In **Chapter 4** the synthesis of peptide fragments of RNase A bearing N-terminal masked glyoxal and C-terminal hydrazides in high yield was presented. Periodate oxidation of the C-terminal peptide section of RNase A showed multiple oxidation, which was demonstrated to be due to the presence of unprotected cysteine residues in the peptide. The use of cysteine protection prevented the unwanted oxidation reaction occurring but the protecting groups increased peptide hydrophobicity, which caused problems with solubilising the peptides and prevented ligation from occurring.

Poor peptide solubility during solution phase reactions is a common problem and a possible solution to this could be use of a solubilising peptide tag. A penta-lysine tag attached to the C-terminus via a base cleavable linker was recently used by Hossain *et al.* (2009) to increase peptide solubility of the A chain of human insulin glargine

during disulfide bond formation. This peptide has known solubility problems and the presence of the five lysine residues was sufficient to solubilise the peptide and allow disulfide bond formation to occur. A similar approach could be applied for the solubilisation of peptide fragments for sequential ligation, but the N and C-termini of the peptides are required for ligation. Therefore the solubilising tag would be required at an alternative location, for example on the side chain of an aspartate or glutamate residue instead of the C-terminal carboxyl. This would require a minor modification to the tag so that instead of being used as a linker between the peptide and the solid support it could be incorporated onto the side chain of aspartate or glutamate during solid phase synthesis.

Alternatively the solubility issues could be resolved by using a different cysteine protection that instead of increasing the hydrophobicity of the peptide could be used to increase peptide solubility. All the commercially available cysteine protections used in solid phase peptide synthesis are more hydrophobic than an unprotected cysteine and so a new protecting group would have to be designed for this purpose. The S-tBu protection used initially in this study to protect the cysteines forms a disulfide bond between the thiol side chain of cysteine and the thio-tertbutyl group. One solution could be to protect the cysteines though formation of a disulfide bond to a hydrophilic thiol such as a polyethylene glycol (PEG) thiol. Woghiren *et al.* (1993) used an activated PEG thiol to reversibly modify a protein through disulfide formation and reduction to reform the native protein. In this example a long 5000MW PEG was used, but short PEG molecules from 2 ethylene glycol units up to 24 ethylene glycol units are commercially available, pre activated with pyridyl-2-disulfide. It should only be necessary to add a short PEG-thiol as a protecting group, as the peptide fragments with unprotected cysteine were soluble and it was only the presence of hydrophobic cysteine protecting groups that caused solubility problems. There are two possible ways that the PEG-thiol protection could be used. An N-alpha Fmoc protected cysteine could be synthesised with the cysteine side chain protected with PEG-thiol, for use in solid phase peptide synthesis. Alternatively the peptide could be synthesized with the normal trityl protection on cysteine and a PEG-thiol could be used to protect the cysteine on the fully deprotected and purified peptide, as a post synthesis modification.

The presence of a chemoselective bond in replacement of a natural peptide bond would be expected to have an effect on the structure of the peptide or protein, which could have implications for the biological activity of that protein. An important aspect of this study was to investigate the structures formed through the presence of chemoselective bonds, and to compare these to the native protein structure. Molecular modelling using molecular dynamics simulations was chosen as a method to study the bond structures and to compare protein structures. Molecular dynamic simulations were performed on native RNase A containing only natural peptide bonds and RNase A with one or more peptide bonds replaced with chemoselective bonds. Final structures of the protein after the molecular dynamics simulations were superposed and calculation of root mean square deviation (RMSD) values was used to evaluate the effect of chemoselective bonds on the structure of the protein. **Chapter 5** described the background to the theory of molecular dynamics simulations, and the experiments that were performed. The computational methods used are described in **Chapter 7**.

The results of the dynamics experiments gave RMSD values of less than 3 Å for the superposition of native RNase A and RNase A containing three hydrazone chemoselective bonds. A value of under 3 Å means that the two structures share significant structural similarity and suggest that the presence of these hydrazone bonds have little effect on the structure of ribunuclease, and therefore on the activity of the enzyme. The experiments presented here represent the first dynamics simulations performed on a protein containing chemoselective bonds. These initial studies gave promising results for the potential of dynamics simulations to accurately predict protein structure when natural peptide bonds are replaced with alternative non-amide bonds. More investigation is required to simulate the effect of placing chemoselective bonds in different locations within RNase A and other proteins and to compare the results to experimental data. This will allow further evaluation of this computational method for the prediction of the effect of positioning chemoselective bonds within proteins.

An original aim of this investigation had been to compare the enzyme activity of native RNase A with that of ligated RNase A using a simple kinetic assay such as the

hydrolysis of cytidine 2':3'-phosphate (Crook *et al.* (1960)) or hydrolysis of poly(C) (delCardayré *et al.* (1994)). This would measure the effect of the chemoselective bonds on RNase A enzyme activity and allow comparison to the predictions of the dynamics simulations. Total synthesis of ligated RNase A was not achieved and so it was not able for this part of the investigation to be performed. If a successful synthesis of RNase A using this method of ligation can be achieved in future then these activity experiments will give valuable information about the effect of the bonds on enzyme activity.

It may be found that the synthesis of RNase A is not possible using this ligation technique, due to the side reactions and solubility issues related to cysteine that were discovered during this investigation. A large number of proteins exist which do not contain cysteine and some of these would be suitable proteins for testing the technique. One example could be Barnase, which is a 110 amino acid protein with ribonuclease activity that contains no cysteines. A version synthesised to contain ligated bonds could be compared to the native protein using an enzyme assay in a similar manner to RNase A, and also using molecular dynamics simulations to investigate the effect of the ligated bonds to the structure of the protein.

The work presented in this thesis evaluated the use of sequential chemoselective ligation for the synthesis of an example protein RNase A, and demonstrated the use of molecular dynamics simulations for the prediction of the effect of these bonds on the protein structure. In this chapter suggestions have been made to improve the solubility of peptide fragments while protecting cysteine residues, which may allow successful sequential ligation of RNase A. Additional dynamics experiments will be required to further evaluate the technique as a method of predicting protein structure enabling its use as a tool in peptide ligation.

Chapter 7: Experimental Methods

7.0 Introduction

This chapter lists materials and suppliers, and experimental detail for experiments performed. Method development is described in **Chapters 2, 3, 4** and **5**. The equipment and instrumentation used is listed and described. All work was carried out at Peptide Protein Research Ltd unless specified.

7.1 Materials

Chemicals were purchased from Novabiochem, Nottingham, UK; Sigma Aldrich, Dorset, UK; Alfa Aesar, Lancashire, UK; Fisher Scientific, Leicestershire, UK; BOC, UK, Iris Biochem, Waldershofer, Germany, and National Diagnostics, Hull, UK, as specified below.

Table 7.1: Chemicals and Suppliers

Chemical	**Abbreviation / Formula**	**Supplier**
2-Chlorotrityl chloride resin	2-CTR	Iris Biochem
Acetic anhydride	Ac_2O	Sigma Aldrich
Acetone		Fisher
Acetonitrile	MeCN	National Diagnostics / Sigma Aldrich
Amberlyst A26		Sigma Aldrich
Ammonium formate		Alfa Aesar
Aniline		Sigma Aldrich
Dichloromethane	DCM	Fisher
Diethyl ether	Et_2O	Fisher
Diisopropyl ethylamine	DIPEA	Alfa Aesar
Dimethyl formamide (anhydrous)	DMF	Sigma Aldrich

Dimethyl formamide	DMF	National Diagnostics
Di-tert-butyl dicarbonate	Boc_2O	Sigma Aldrich
Dry ice		BOC
Ethyl acetate	EtOAc	Sigma Aldrich
Ethyl chloroformate		Sigma Aldrich
Fmoc amino acids		Novabiochem
2-(1 H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate	HBTU	Novabiochem
Hydrochloric acid	HCl	Fisher
Hydrogen bromide	HBr	Sigma Aldrich
Imidazole hydrochloride		Sigma Aldrich
N-Hydroxybenzotriazole	HOBt	Novabiochem
Magnesium sulfate	$MgSO_4$	Alfa Aesar
Methanol	MeOH	Fisher
Methylamine hydrochloride		Alfa Aesar
Ninhydrin		Alfa Aesar
Palladium carbon (10%)	PdC	Sigma Aldrich
Piperidine		Alfa Aesar
Phenol		Sigma Aldrich
Potassium cyanide	KCN	Alfa Aesar
Potassium hydroxide	KOH	Sigma Aldrich
Benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate	PyBOP	Novabiochem
Pyridine		Alfa Aesar
Silica Gel 60		Alfa Aesar
Sodium acetate	NaOAc	Alfa Aesar
Sodium azide	NaN_3	Alfa Aesar
Sodium borohydride	$NaBH_4$	Alfa Aesar
Sodium hydrogen carbonate	$NaHCO_3$	Sigma Aldrich
Sodium metaperiodate	$NaIO_4$	Sigma Aldrich
Sodium nitrite		Alfa Aesar
Sodium sulfate	NaS_2O_4	Sigma Aldrich
Tetrahydrofuran	THF	Sigma Aldrich
Triethylamine	TEA	Sigma Aldrich
Trifluoroacetic acid	TFA	Apollo Scientific
Triisopropyl silane	TIS	Alfa Aesar

Urea		Alfa Aesar
Water	H_2O	Fisher

All solvents and reagents used were AnalR grade or peptide synthesis grade. Solvents for HPLC purification and analysis were Chromasolv grade from Sigma Aldrich.

7.2 Experimental Methods

All general methods used are detailed in this section, and are refered to in subsequent chapters.

7.2.1 Solid Phase Peptide Synthesis

All peptides were synthesised either on the Symphony or Prelude automated peptide synthesisers (Protein Technologies) or by manual peptide synthesis. Fmoc chemistry was used throughout. For more detailed descriptions of peptide synthesis using Fmoc chemistry see **Chapter 1 Section 1.2.1** and also "Solid phase peptide synthesis. A practical approach" (Atherton and Sheppard (1989)). Peptides were synthesised either on Wang resin pre-loaded with an Fmoc amino acid (Novabiochem), or on 2-chlorotrityl chloride resin (Iris Biochem) as detailed below.

7.2.1.1 Fmoc Amino Acid Side Chain Protecting Groups

The amino acids used for peptide synthesis and as starting material for the various amino acid derivatives were Fmoc protected on the alpha amino group, and side chain protected where necessary, as shown in **Table 7.2**.

Table 7.2: The standard side chain protecting groups used in Fmoc peptide synthesis

Amino acid	Side chain protection	Amino acid	Side chain protection
Alanine	None	Leucine	None
Arginine	Pbf	Lysine	Boc
Asparagine	Trt	Methionine	None
Aspartic acid	*t*-Bu	Phenylalanine	None
Cysteine	Trt	Proline	none
Glutamic acid	*t*-Bu	Serine	*t*-Bu
Glutamine	Trt	Threonine	*t*-Bu
Gycine	None	Tyrosine	*t*-Bu
Histidine	Trt	Tryptophan	Boc
Isoleucine	None	Valine	None

For synthesis of long peptide fragments of ribonuclease, cysteine was protected with either thio-tertbutyl (S-tBu) or acetamidomethyl (ACM). The use of pseuodoproline dipeptides (Mutter *et al.* (1995)) at regular invervals throughout long peptide fragments enabled the synthesis of these long peptides in high crude purity.

7.2.1.2 Fmoc Deprotection

The N-terminal Fmoc protecting group was removed by agitating (shaking or nitrogen bubbling) the Fmoc protected peptidyl-resin (100-200mg) with 20% (v/v) piperidine in DMF (5 min and 20 min respectively, 2 ml each). The resin was then washed with DMF (3 x 2 ml, 2 mins each). The resin was tested for a free primary amine by using qualitative ninhydrin assay (**Section 7.2.1.6**).

7.2.1.3 Coupling Reaction

An Fmoc protected amino acid (3 equivalents (equiv.) to the resin bound amine), HBTU (3 equiv.) and HOBt (3 equiv.) were added to the resin in DMF (2 ml), and DIPEA (9 equiv.) was added. The resin was agitated at room temperature (45-60 mins) and washed with DMF (2 x 2 ml, 2 mins each). A portion of the resin was tested by qualitative ninhydrin assay (**Section 7.2.1.6**), and if a blue colour was produced, indicating incomplete coupling, the coupling step was repeated using the same conditions as before. Amino acids succeeding a proline were coupled twice, as the ninhydrin assay only produces a blue colour in the presence of amino and not imino groups.

7.2.1.4 Acetylation

Difficult couplings (coupling reactions which were not negative by ninhydrin assay after 2 couplings) were capped with Ac_2O (100 μl) and pyridine (100 μl) in DMF (2 ml, 10 mins), then the resin washed with DMF (2 x 2 ml, 2 mins each).

7.2.1.5 Cleavage

The peptidyl resin was washed with DCM (2 x 2 ml) and Et_2O (2 x 2 ml) and allowed to air dry prior to cleavage. Peptides were cleaved in TFA/TIS/H_2O (96:2:2) or for peptides containing Cys or Met were cleaved in TFA/TIS/H_2O/EDT (94:2:2:2). The peptidyl resin (300 mg) was stirred at room temperature in the appropriate cleavage mix (2 ml) for 2-4 hours. The filtrate was collected and the resin washed with TFA (2 x 2 ml). The filtrate and washings were combined and the peptide precipitated by the addition of diethyl ether (10ml), and centrifuged to form a pellet. The pellet was resuspended in MeCN/H_20 (1:5, 5ml) and freeze dried.

7.2.1.6 Qualitative Ninhydrin Assay

7.2.1.6.1 Preparation of Ninhydrin Assay Solutions

Solution A: Phenol (40 g) was dissolved in MeOH (10 ml) and stirred at room temperature. A solution of KCN (6.5 mg) in H_20 (100 ml) was prepared and 1ml of this solution was added to pyridine (49 ml). This was added to the phenol/MeOH solution.

Solution B: Ninhydrin (500 mg) was dissolved in MeOH (10 ml)

7.2.1.6.2 Ninhydrin Assay

A small portion of peptidyl-resin was placed into a 2 ml test tube. Solution A (40 μl) and ninhydrin B (20 μl) were added to the resin, and the tubes heated for 3 mins at 100 °C. A blue colour (positive result) indicates the presence of a primary amine.

7.2.1.7 Automated Peptide Synthesis

Peptide fragments for ligation were synthesised by automated peptide synthesiser on either a Symphony or Prelude (Protein Technologies). The standard program cycle used is shown in **Table 2.3**:

Table 7.3: Standard Automated Synthesis Program

Reagant(s)	Time (mins)	Replicants
DMF	0.5	3
20% piperidine in DMF	2.5	2
DMF	0.5	6
Amino acid/HBTU/ DPEA	20	1
DMF	0.5	1
Amino acid/HBTU/ DPEA	20	1
DMF	0.5	3

The Symphony peptide synthesiser was used for synthesis of peptide containing only the 20 natural amino acids with standard protecting groups for Fmoc chemistry as listed in **Table 7.2**. The Prelude peptide synthesiser has spaces for 27 amino acids and so this was used for synthesis of peptides containing pseudoprolines and cysteine with S-tBu and ACM protection.

7.2.1.8 Calculation of Resin Substitution Level

The substitution level of polystyrene resins was measured using the absorbance at 304 nm for removal of the Fmoc protecting group, using the method recommended by Novabiochem [6]. A portion of dry resin (approx 10 mg, accurately weighed) was treated with DBU (2% in DMF, 2 ml) for 30 mins, then diluted to 10 ml with MeCN. A 2ml sample of this solution was diluted to 25 ml with MeCN and the optical density of this solution measured in a spectrophotometer at 304 nm using a 1 ml silica cell. A reference sample prepared in the same manor, but without the resin was measured and the substitution level calculated using the following equation (derived from the Beer Lambert equation $A = \varepsilon cl$):

$$\text{Resin loading(mmol/g)} = Abs_{sample} - Abs_{ref} \times 16.4 / \text{mg of resin}$$

7.2.2 Synthesis of N-nitrosomethyl Urea

Methylamine hydrochloride (60 g, 0.9 mol) and H_2O (to a final weight of 300 g) was stirred at room temperature, and urea (180 g, 3 mol) was added. The solution was refluxed for 5 hr, allowed to cool and left at room temperature overnight. The solution was cooled to 0°C and 98% sodium nitrite (66 g, 0.9 mol) was added with stirring. Separately, a mixture of ice (360 g) and concentrated H_2SO_4 was stirred, and the cold methylurea-nitrite solution was added over a period of 3 hr, adding dry ice as necessary to maintain the temperature below 0 °C. The resultant solid N-nitrosomethyl urea (60 g) was collected via filtration under reduced pressure through a sintered glass funnel, washed with ice-cold water (100 ml), and desiccated *in vacuo* over silica gel. The product was stored at 4 °C.

7.2.3 Generation of Diazomethane

Nitrosomethyl urea (6 g) was added to diethyl ether (20 ml) on ice, and KOH (50% aq) was added dropwise over 10-15 mins until the solid nitrosomethyl urea had completely dissolved. The yellow ether layer was separated and dried over KOH pellets, then used immediately.

7.2.4 Synthesis of Fmoc-Phe-epoxide

The epoxide derivative of phenylalanine **[3]** was synthesised starting with the commercially available Fmoc-phenylalanine.

7.2.4.1 Preparation of Fmoc-Phe-diazomethylketone [1]

Fmoc-phenylalanine (3.8 g, 10 mmol) was dissolved in dry THF, and TEA (1.01 g, 2 eq) was added. This was added dropwise to a solution of ethyl chloroformate (1 g, 2 eq) in dry THF, whilst stirring on ice. Stirring was continued for 3 hours. A saturated solution of diazomethane in ether (**Section 7.2.3**) was added, and the reaction stirred for a further 3 hours. H_2O (20 ml) was added slowly, and the reaction stirred until effervescence had stopped. The organic layer was separated and washed with saturated $NaHCO_3$ (20 ml) and H_2O (2 x 20 ml), then dried over $NaSO_4$, and the solvent removed by evaporation *in vacuo*. The crude product was purified by column chromatography (65% hexanes / ethyl acetate), to yield a pale yellow solid.

Yield of Fmoc-Phe-diazomethylketone **[1]** 3.2 g, 76 %, IR: 2105 cm^{-1} (s, diazo), ^{13}C 300MHz NMR ($CDCl_3$ 20°C) δ 192.8 (CO ketone), 155.6 (CO amide), (143.8, 141.4, 129.4, 128.7, 127.8, 127.1, 125.1, 120.0 (C, CH aromatic)), 66.8 (CH2 Fmoc), 58.7 (CH Phe), 54.7 (CH2 diazo methyl ketone), 47.3 (CH Fmoc), 38.4 (CH2 Phe). ^1H 300mHz NMR ($CDCl_3$ 20°C) δ 7.8-7.2 (m, 13H), 5.4 (d, 1H), 5.1 (s, 1H), 4.5 (m, 1H), 4.4 (m, 2H), 4.2 (m, 2H), 3.0, (m, 1H). CHN: expected: C, 72.80; H, 5.38; N, 10.19, found: C, 72.78; H, 5.01; N, 9.16. 1H NMR, 1H COSY, 13C NMR and FTIR data in **Appendix 11a.**

7.2.4.2 Preparation of Fmoc-Phe-bromomethylketone [2]

The diazomethylketone **[1]** (3.2 g) was dissolved in acetone (20 ml), and HBr (48% aq, 440 μl, 1 eq) was added dropwise. The reaction was stirred for 5 minutes, then saturated $NaHCO_3$ solution (20 ml) added. The product was extracted into ethyl acetate (2 x 10 ml), the layers combined and washed with H_2O (2 x 20 ml), and then dried over $NaSO_4$. The solvent was removed by evaporation *in vacuo* to yield a white solid.

Yield of Fmoc-Phe-bromomethylketone **[2]** 2.4 g, 98 %, LCMS: Purity 97.9% ESMS [M+H] 466, IR 1735 cm^{-1}(s, bromomethylketone). ^{13}C 300MHz NMR ($CDCl_3$ δ ppm

20°C) 200.4 (CO ketone), 155.7 (CO amide), (143.6, 141.4, 135.5 129.1, 129.0, 127.8, 127.4, 127.1, 125.0, 120.1 (C, CH aromatic)), 67.0 (CH2 Fmoc), 58.8 (CH Phe), 47.2 (CH Fmoc), 37.8 (CH2 Phe), 32.9 (CH2 bromomethylketone). ^{1}H 300mHz NMR ($CDCl_3$ 20°C) δ 7.8-7.1 (m, 13H), 5.3 (d, 1H), 4.8 (m, 1H), 4.4 (m, 2H), 4.2 (m, 2H), 3.9 (d *J* = 13.5, 1H) 3.7 (d *J* = 13.5, 1H), 3.0, (m,1H). CHN: expected: C, 62.66; H, 4.78; N, 3.02, found: C, 64.68; H, 4.65; N, 2.80. 1H NMR, 1H COSY, 13C NMR, FTIR and LCMS data in **Appendix 11a.**

7.2.4.3 Preparation of Fmoc-Phe-epoxide [3]

The bromomethylketone **[2]** (100 mg) was dissolved in MeOH (20 ml), and $NaBH_4$ (12 mg, 1.4 eq) was added. The reduction of the ketone was monitored by TLC (ethyl actetate / hexanes 7:3) and was complete in 10 minutes. The reaction was then refluxed for 15 minutes and then cooled on ice. The methanol was removed *in vacuo*, and the residue redissolved in ethyl acetate (10 ml). This was washed with dilute HCl (10 ml), H_2O (2 x 10 ml) and dried over $NaSO_4$. The solvent was removed by evaporation *in vacuo* to yield a white solid. Yield of Fmoc-Phe-epoxide **[3]** 60 mg, 72 %, ESMS [M+H] 386, [M+Na] 408.

7.2.5 Synthesis of 1-amino, 2-alcohol Derivatives

Amino alcohol derivatives were synthesised starting from the appropriate Fmoc protected amino acid (see **Table 2.2**). Phenylalanine is used here as an example. Fmoc-Phe-diazomethylketone **[1]** and Fmoc-Phe-bromomethylketone **[2]** were synthesised as above (**7.2.4.1** and **7.2.4.2**).

7.2.5.1 Preparation of Fmoc-Phe-azidomethylketone [4]

The bromomethylketone derivative of Fmoc-phenylalanine **[2]** (1.7 g) was dissolved in ethyl acetate (25 ml). Sodium azide (2.3 g 10 eq) was dissolved in H_2O and this was added to the solution of **[2]** with vigorous stirring. The reaction was stirred at room temperature overnight, then the ethyl acetate layer separated and washed with H_2O (2 x 20 ml). This was dried over sodium sulfate, and the solvent removed by evaporation *in vacuo*. The crude product was purified by column chromatography (65% hexanes / ethyl acetate), to yield a white solid.

Yield of Fmoc-Phe-azidomethylketone **[4]** 1.5 g, 96 %, LCMS: Purity 98.0% ESMS [M+H] 427. ^{13}C 300MHz NMR ($CDCl_3$ δ ppm 20°C) 203.4 (CO ketone), 155.8 (CO amide), (143.6, 143.5, 141.3, 135.2, 129.2, 129.1, 127.8, 127.5, 127.1, 125.0, 120.1 (C, CH aromatic)), 67.0(CH2 Fmoc), 58.7 (CH Phe), 56.5 (CH2 azidomethyl ketone), 47.2 (CH Fmoc), 37.5 (CH2 Phe). 1H NMR ($CDCl_3$) δ 7.8-7.1 (m, 13H), 5.2 (d, 1H), 4.5 (m, 1H), 4.4 (m, 2H), 4.2 (m, 1H), 3.9 (d, *J* = 16 Hz, 1H) 3.7 (d, *J* = 16 Hz, 1H) 3.0 (d *J* = 7.5, 1H). CHN: expected: C, 70.41; H, 5.20; N, 13.14, found: C, 68.64; H, 5.09; N, 12.19. 1H NMR, 1H COSY, 13C NMR, FTIR and LCMS data in **Appendix 11a.**

7.2.5.2 Preparation of Fmoc-Phe-azidomethylalcohol [5]

The azidomethylketone **[4]** (1.5 g) was dissolved in MeOH (20 ml), and $NaBH_4$ (200 mg, 1.4 eq) was added. The reduction was monitored by TLC (ethyl actetate / hexanes 7:3) and was complete in 10 minutes. The methanol was removed *in vacuo*, and the residue redissolved in ethyl acetate (30 ml). This was washed with dilute HCl (20 ml), H_2O (2 x 20 ml) and dried over $NaSO_4$. The solvent was removed by evaporation *in vacuo.* The crude product was purified by column chromatography (65% hexanes / ethyl acetate), to yield a white solid.

Yield of Fmoc-Phe-azidomethylalcohol **[5]** 1.3 g, 87 % LCMS: Purity 90.7% ESMS [M+H] 429 ^{13}C 300MHz NMR ($CDCl_3$δ ppm 20°C) 156.5 (CO amide), (143.8,

141.4, 137.4, 137.0, 129.3, 129.2, 128.7, 127.8, 127.1, 126.8, 125.0, 120.0 (C, CH aromatic)), (72.2, 70.3 (CH alcohol)), 66.5(CH2 Fmoc), (55.0, 54.3, (CH2 azidomethylalcohol)), 47.3 (CH Fmoc), 38.3(CH), 35.9 (CH2 Phe).^{1}H NMR: (500 MHz CDCl3) δ 7.8-7.1 (m, 13H), 5.1 (d, 1H) 4.8 (m, 1H), 4.4 (m, 2H), 4.2 (m, 1H), 3.9 (m, 1H) 3.7 (m, 1H), 3.3 (m, 2H), 2.9 (m, 1H). Stereochemistry of hydroxyl shown by HPLC to be approximately 60:40. CHN: expected: C, 70.08; H, 5.65; N, 13.08, found: C, 69.92; H, 5.54; N, 12.79. 1H NMR, 1H COSY, 13C NMR, FTIR and LCMS data in **Appendix 11a.**

7.2.5.3 Preparation of Fmoc-Phe-amino-alcohol [6]

Fmoc-Phe-azidomethylalcohol **[5]** (500 mg) was dissolved in MeOH (15 ml) and ammonium formate (290 mg, 4 eq) was added. This was stirred at 30°C, and Pd-C (10%, 25 mg) was added with continual stirring. After 30 minutes the reaction was allowed to cool to room temperature, and the reaction was filtered through celite. The MeOH was removed by evaporation *in vacuo*. Yield of Fmoc-Phe-amino-alcohol **[6]** 420 mg, 85%, LCMS: Purity 84.1% ESMS [M+H] 403, LCMS data in **appendix IIa**. Fmoc-Phe-amino-alcohol was immediately used, due to its instability.

7.2.5.4 Synthesis of Amino Alcohol Derivatives of Other Amino Acids

Amino alcohol derivatives of Fmoc-leucine, Fmoc-isoleucine and Fmoc-alanine were synthesised following the same method used for the amino alcohol derivative of phenylalanine. These derivatives were not analysed by NMR due to a lack of access to NMR facilities when this work was performed, however the methods used were identical to those used for phenylalanine. All compounds were analysed by ESMS and gave the expected results.

7.2.5.4.1 Preparation of Fmoc-Leu-diazomethylketone [7]

Fmoc-Leu-diazomethylketone was synthesised using the method in **Section 7.2.4.1**, starting from Fmoc-leucine (1.8 g, 5 mM). Yield of Fmoc-Leu-diazomethylketone **[7]** 1.6 g, 83 %, IR (thin film) 2100 cm^{-1}(s, diazo).

7.2.5.4.2 Preparation of Fmoc-Leu-bromomethylketone [8]

Fmoc-Leu-bromomethylketone was synthesised using the method in **Section 7.2.4.2**, starting from Fmoc-Leu-diazomethylketone **[7]** (1.6 g). Yield of Fmoc-Leu-bromomethylketone **[8]** 1.7 g, 93 %, ESMS [M+H] 430/432.

7.2.5.4.3 Preparation of Fmoc-Leu-azidomethylketone [9]

Fmoc-Leu-azidomethylketone was synthesised using the method in **Section 7.2.5.1**, starting from Fmoc-Leu-bromomethylketone **[8]** (1.7 g). Yield of Fmoc-Leu-azidomethylketone [9] 1.4 g, 90 %, ESMS [M+H] 393.

7.2.5.4.4 Preparation of Fmoc-Leu-azidomethylalcohol [10]

Fmoc-Leu-azidomethylalcohol was synthesised using the method in **Section 7.2.5.2**, starting from Fmoc-Leu-azidomethylketone **[9]** (1.4 g). Yield of Fmoc-Leu-azidomethylketone **[10]** 1.3 g, 90 %, ESMS [M+H] 395.

7.2.5.4.5 Preparation of Fmoc-Leu-amino alcohol [11]

Fmoc-Leu-amino alcohol was synthesised using the method in **Section 7.2.5.3**, starting from Fmoc-Leu-azidomethylketone **[10]** (500 mg). Yield of Fmoc-Leu-azidomethylketone **[11]** 430 mg, 92 %, ESMS [M+H] 369.

7.2.5.4.6 Preparation of Fmoc-Ile-diazomethylketone [12]

Fmoc-Ile-diazomethylketone was synthesised using the method in **Section 7.2.4.1**, starting from Fmoc-isoleucine (1 g, 2.7 mM). Yield of Fmoc-Ile-diazomethylketone **[12]** 1 g, 93 %, IR (thin film) 2100 cm^{-1}(s, diazo).

7.2.5.4.7 Preparation of Fmoc-Ile-bromomethylketone [13]

Fmoc-Ile-bromomethylketone was synthesised using the method in **Section 7.2.4.2**, starting from Fmoc-Ile-diazomethylketone **[12]** (1 g). Yield of Fmoc-Ile-bromomethylketone **[13]** 950 mg, 83 %, ESMS [M+H] 430/432.

7.2.5.4.8 Preparation of Fmoc-Ile-azidomethylketone [14]

Fmoc-Ile-azidomethylketone was synthesised using the method in **Section 7.2.5.1**, starting from Fmoc-Ile-bromomethylketone **[13]** (950 mg). Yield of Fmoc-Ile-azidomethylketone **[14]** 800 mg, 92 %, ESMS [M+H] 393.

7.2.5.4.9 Preparation of Fmoc-Ile-azidomethylalcohol [15]

Fmoc-Ile-azidomethylalcohol was synthesised using the method in **Section 7.2.5.2**, starting from Fmoc-Ile-azidomethylketone **[14]** (800 mg). Yield of Fmoc-Ile-azidomethylketone **[15]** 800 mg, 100 %, ESMS [M+H] 395.

7.2.5.4.10 Preparation of Fmoc-Ile-amino alcohol [16]

Fmoc-Ile-amino alcohol was synthesised using the method in **Section 7.2.5.3**, starting from Fmoc-Ile-azidomethylketone **[15]** (500 mg). Yield of Fmoc-Ile-azidomethylketone **[16]** 450 mg, 96 %, ESMS [M+H] 369.

7.2.5.4.11 Preparation of Fmoc-Ala-diazomethylketone [17]

Fmoc-Ala-diazomethylketone was synthesised using the method in **Section 7.2.4.1**, starting from Fmoc-alanine (1.65 g, 5 mM). Crude yield of Fmoc-Ala-diazomethylketone **[17]** 1.7 g, 88 %. 1 g purified by column chromatography (65% hexanes / ethyl acetate), to yield a pale yellow solid, 800 mg IR (thin film) 2100 cm^{-1} (s, diazo).

7.2.5.4.12 Preparation of Fmoc-Ala-bromomethylketone [18]

Fmoc-Ala-bromomethylketone was synthesised using the method in **Section 7.2.4.2**, starting from Fmoc-Ala-diazomethylketone **[17]** (800 mg). Yield of Fmoc-Ala-bromomethylketone **[18]** 800 mg, 94 %, ESMS [M+H] 388/390.

7.2.5.4.13 Preparation of Fmoc-Ala-azidomethylketone [19]

Fmoc-Ala-azidomethylketone was synthesised using the method in **Section 7.2.5.1**, starting from Fmoc-Ala-bromomethylketone **[18]** (800 mg). Yield of Fmoc-Ala-azidomethylketone **[19]** 650 mg, 90 %, ESMS [M+H] 351.

7.2.5.4.14 Preparation of Fmoc-Ala-azidomethylalcohol [20]

Fmoc-Ala-azidomethylalcohol was synthesised using the method in **Section 7.2.5.2**, starting from Fmoc-Ala-azidomethylketone **[19]** (650 mg). Yield of Fmoc-Ala-azidomethylketone **[20]** 600 mg, 92 %, ESMS [M+H] 353.

7.2.5.4.15 Preparation of Fmoc-Ala-amino alcohol [21]

Fmoc-Ala-amino alcohol was synthesised using the method in **Section 7.2.5.3**, starting from Fmoc-Ala-azidomethylketone **[20]** (200 mg). Yield of Fmoc-Ala-azidomethylketone **[21]** 450 mg, 96 %, ESMS [M+H] 327. Due to stability issues this was immediately attached to 2-CTR resin (using the method in **Section 7.2.7.3** for Fmoc-Phe amino alcohol).

7.2.6 Stability of 1, 2 Amino Alcohol derivatives of Fmoc Protected Amino Acids

1, 2 amino alcohol derivatives of α-Fmoc protected amino acids showed signs of Fmoc removal upon storage. A stability study was performed to assess the extent of Fmoc removal.

A 5 mg sample of Fmoc-Phe, Fmoc-Leu and Fmoc-Ala azido alcohol and amino alcohol derivatives were stored at 4º C. A 1 mg sample of each was dissolved in MeCN / H_2O (1:1, 300 μl) and analysed after 24 h, 1 week and 1 month by ESMS.

7.2.7 Synthesis of Peptides with a C-terminal 1, 2 Amino Alcohol

Peptides were synthesised on 2-chlorotrityl chloride resin, loaded either with the commercially available 1, 3-amino-2-hydroxypropane (glycine amino alcohol), or the 1, 2 amino alcohol derivatives of Fmoc protected amino acids.

7.2.7.1 Substitution of 2-Chlorotrityl Resin with 1, 3-amino-2-hydroxypropane

2-chlorotrityl resin was substituted with 1, 3-amino-2-hydroxypropane for the synthesis of peptides containing a C-terminal glycine amino alcohol. Chlorotrityl

resin (200 mg) was swelled in dry DMF/DCM (1:1, 2 ml). 1, 3-amino-2-hydroxypropane (30 μl) and DIPEA (125 μl) was added and the resin agitated for 2 hours. The solution was drained and the resin capped with MeOH /DMF (1:1, 2ml) containing DIPEA (125 μl) for 30 minutes, then the resin washed with DMF (3 x 2ml). The resin was tested by ninhydrin rection and gave a positive (blue) result.

7.2.7.2 Synthesis of Peptides on 1, 3-amino-2-hydroxypropane Substituted 2-chlorotrityl Resin

The C-terminal amino acid was loaded onto the 1, 3-amino-2-hydroxypropane substituted resin (200 mg). The Fmoc amino acid (125 mg), PyBOP (1 M equiv to Fmoc amino acid) and DIPEA (3 M eq) were added in NMP, and the resin agitated for 1 hour. The solution was drained, and the resin washed with DMF (3 x 2ml). A sample of the resin was tested by nihydrin assay, and if positive the reaction was repeated.

The Fmoc-amino acid-1, 3-amino-2-hydroxypropane functionalised resin was loaded onto an automated peptide synthesiser (Symphony or Prelude).

7.2.7.3 Substitution of 2-Chlorotrityl Resin with Fmoc-Phe amino alcohol [6]

2-chlorotrityl chloride resin (200 mg) was swelled in dry DMF/DCM (1:1, 2ml). Fmoc-Phe amino alcohol **[6]** (50 mg) and DIPEA (125 μl) was added and the resin agitated for 48 hours. The solution was drained and the resin capped with MeOH /DMF (1:1, 2 ml) containing DIPEA (125 μl) for 30 minutes, then the resin washed with DMF (3 x 2 ml). The side chain hydroxyl was protected by adding t-butyl-2,2,2-trichloroacetamidate (100μl) and DIPEA (125 μl) in DMF. The N-terminal amine was deprotected by removing the Fmoc group using the method given in **Section 7.2.1.2**.

7.2.7.3.1 Synthesis of Z-RLF-amino alcohol

Phe-amino alcohol-CTR (200mg, **Section 7.2.7.3**) was reacted with Fmoc-Leu-OH (140 mg, 3eq), HBTU (150 mg, 3 eq) and DIPEA (125 μl) for 1 h and the reaction monitored by ninhydrin assay (**Section 7.2.1.6**). The N-terminal amine was deprotected by removing the Fmoc group using the method given in **Section 7.2.1.2**, and the reaction again monitored by ninhydrin assay. Fmoc-Arg-OH (260 mg, 3eq), HBTU (150 mg, 3 eq) and DIPEA (125 μl) were reacted for 1 h and the reaction monitored by ninhydrin assay (**Section 7.2.1.6**). The N-terminal amine was deprotected by removing the Fmoc group using the method given in **Section 7.2.1.2**, and the reaction again monitored by ninhydrin assay. Z-chloroformate (55 μl, 3eq), and DIPEA (125 μl) in DMF were reacted for 1 hour and the reaction monitored by ninhydrin assay. The resin was washed with DCM (3 x 2ml) and allowed to dry before cleavage. The peptide was cleaved using the method in **Section 2.7.5.1,** and purified by RP-HPLC (**Section 7.3.1.2**), to yield Z-RLF-amino alcohol (7.1 mg), LCMS: purity 100% ESMS [M + H] 584 (data in **appendix IIb**).

7.2.7.3.2 Synthesis of Z-RLF-aldehyde

Z-RLF-amino alcohol (7.1 mg) was dissolved in NaOAc buffer (7ml, 0.1M, pH5.5) and NaIO4 (10 μl, from a stock of 10 mg in 100 μl NaOAc buffer (7ml, 0.1M, pH5.5)) was added. The reaction was stirred for 30 mins and the recation monitored by ESMS. The peptide was immediately purified by RP-HPLC (**Section 7.3.1.2**), freeze dried and stored at -20 ºC. ESMS [M + H] 553, [M + H_30] 571 (data in **appendix IIb**).

7.2.8 Synthesis of Peptides with a C-terminal Hydrazide

Peptides bearing a C-terminal hydrazide were synthesised on 2-chlorotrityl chloride resin, substituted with hydrazine.

7.2.8.1 Hydrazine Substituted 2-chlorotrityl Chloride Resin

Chlorotrityl resin (200 mg) was swelled in dry DMF/DCM (1:1, 2ml). Hydrazine monohydrate (40 μl) and DIPEA (125 μl) was added and the resin agitated for 45 minutes. The solution was drained and the resin capped with MeOH /DMF (1:1, 2 ml) containing DIPEA (125 μl) for 30 minutes, then the resin was washed with DMF (3 x 2 ml).

7.2.8.2 Synthesis of Peptides on Hydrazine Substituted 2-chlorotrityl Resin

The C-terminal amino acid was loaded onto hydrazine substituted resin (200 mg). The Fmoc amino acid (125 mg), PyBOP (1M eq to Fmoc amino acid) and DIPEA (3M eq) were added in NMP, and the resin agitated for 1 hour. The solution was drained, and the resin washed with DMF (3 x 2 ml).

The Fmoc-amino acid-hydrazide functionalised resin was loaded onto an automated peptide synthesiser (Symphony or Prelude), and the peptides synthesised as detailed below. As the starting resin substitution level was not measured final yields are in mg only, not as a percentage.

7.2.8.3 Synthesis of 8Aoa-fQWAVGHL-NH-NH2

Hydrazine substituted 2-chlorotrityl chloride resin (300 mg) was prepared using the method in **Section 7.2.8.1**. The C-terminal leucine was loaded using Fmoc-Leu-OH (125 mg), Boc-Gly-OH (6.5 mg), PyBOP (185 mg, 1 eq) and DIPEA (125 μl), and the reisn agitated for 1 hour. The rsin was transferred to the Synpony eptide synthesiser and the peptide elongated to make QWAVGHL-NH-NH2-CTR resin. The final two residues D-Phe and 8Aoa were added by manual peptide synthesis

using the method in **Section 7.2.1**. The peptide was cleaved (**Section 7.2.1.5**) and purified (**Section 7.3.1.2**) to yield 48 mg of peptide.

7.2.8.4 Synthesis of KETAAAKFERQHMDSSTSA-NH-NH2 (Rib 1)

Peptide KETAAAKFERQHMDSSTSA-NH-NH2 (RNase peptide 1) was made using hydrazine substituted CTR (**Section 7.2.8.2**). All amino acids, pseudoprolines and HBTU were used in a 3 fold excess, using the standard program detailed in Section 7.2.1.7. The peptide was cleaved from the resin using the method in **Section 2.7.5.1**, and purified by RP-HPLC (**Section 7.3.1.2**), to yield the final peptide (125 mg, 71% pure by HPLC).

7.2.8.5 Synthesis of SSSSNYCNQMMKSRNLTKDRCKPVNTFVHESL-NH-NH2 (Rib 2)

Peptide SSSSNYCNQMMKSRNLTKDRCKPVNTFVHESL-NH-NH2 (RNase peptide 2) was made using hydrazine substituted CTR (**Section 7.2.8.2**) and pseudoprolines LT and ES. All amino acids, pseudoprolines and HBTU were used in a 3 fold excess, using the standard program detailed in **Section 7.2.1.7**. The peptide was cleaved from the resin using the method in **Section 2.7.5.1,** and purified by RP-HPLC (**Section 7.3.1.2**), to yield the final peptide (65 mg, 94% pure by HPLC)

7.2.8.6 Synthesis of SDVQAVCSQKNVACKNGQTNCYQSYSTMSI-NH-NH2 (Rib 3)

Peptide SDVQAVCSQKNVACKNGQTNCYQSYSTMSI-NH-NH2 (RNase peptide 3) was made using hydrazine substituted CTR (**Section 7.2.8.2**) and pseudoprolines QT and ST. All amino acids, pseudoprolines and HBTU were used in a 3 fold excess, using the standard program detailed in **Section 7.2.1.7,** apart from the three asparagines reactions which were triple coupled instead of the standard double

coupling reaction. The peptide was cleaved from the resin using the method in **Section 2.7.5.1,** and purified by RP-HPLC (**Section 7.3.1.2**), to yield the final peptide (38 mg, 84% pure by HPLC).

7.2.8.7 Synthesis of SSSSNYC(S-tBu)NQMMKSRNLTKDRC(S-tBu)KPVNTFVHESL-NH-NH2 (Rib 2 StBu)

Peptide SSSSNYC(S-tBu)NQMMKSRNLTKDRC(S-tBu)KPVNTFVHESL-NH-NH2 was made using hydrazine substituted CTR (**Section 7.2.8.2**) and pseudoprolines LT and ES. All amino acids, pseudoprolines and HBTU were used in a 3 fold excess, using the standard program detailed in **Section 7.2.1.7**. The peptide was cleaved from the resin using the method in **Section 2.7.5.1,** and purified by RP-HPLC (**Section 7.3.1.2**), to yield the final peptide (24 mg).

7.2.8.8 Synthesis of SDVQAVC(S-tBu)SQKNVAC(StBu)KNGQTNC(S-tBu)YQSYSTMSI-NH-NH2 (Rib 3 StBu)

Peptide SDVQAVC(S-tBu)SQKNVAC(S-tBu)KNGQTNC(S-tBu)YQSYSTMSI-NH-NH2 was made using hydrazine substituted CTR (**Section 7.2.8.2**) and pseudoprolines QT and ST. All amino acids, pseudoprolines and HBTU were used in a 3 fold excess, using the standard program detailed in **Section 7.2.1.7,** apart from the three asparagines reactions which were triple coupled instead of the standard double coupling reaction. The peptide was cleaved from the resin using the method in **Section 2.7.5.1,** and purified by RP-HPLC (**Section 7.3.1.2**), to yield the final peptide (17 mg).

7.2.8.9 Synthesis of SSSSNYC(ACM)NQMMKSRNLTKDRC(ACM) KPVNTFVHESL-NH-NH2 (Rib 2 ACM)

Peptide SSSSNYC(ACM)NQMMKSRNLTKDRC(ACM)KPVNTFVHESL-NH-NH2 was made using hydrazine substituted CTR (**Section 7.2.8.2**) and pseudoprolines LT and ES. All amino acids, pseudoprolines and HBTU were used in a 3 fold excess, using the standard program detailed in **Section 7.2.1.7**. The peptide was cleaved from the resin using the method in **Section 2.7.5.1,** and purified by RP-HPLC (**Section 7.3.1.2**), to yield the final peptide (34 mg).

7.2.8.10 Synthesis of SDVQAVC(ACM)SQKNVAC(ACM)KNGQT NC(ACM)YQSYSTMSI-NH-NH2 (Rib 3 ACM)

Peptide SDVQAVC(ACM)SQKNVAC(ACM)KNGQTNC(ACM)YQSYSTMSI-NH-NH2 was made using hydrazine substituted CTR (**Section 7.2.8.2**) and pseudoprolines QT and ST. All amino acids, pseudoprolines and HBTU were used in a 3 fold excess, using the standard program detailed in **Section 7.2.1.7,** apart from the three asparagines reactions which were triple coupled instead of the standard double coupling reaction. The peptide was cleaved from the resin using the method in **Section 2.7.5.1,** and purified by RP-HPLC (**Section 7.3.1.2**), to yield the final peptide (19 mg).

7.2.9 Synthesis of N-terminally modified peptide fragments for use in ligation

Not all peptides required for ligation needed C-terminal modification, and so these peptides were synthesised using commercially available preloaded Wang polystyrene resins. The N-terminal modifications were either hydrazine actetic acid, or serine for subsequent oxidation to the glyoxal.

7.2.9.1 Synthesis of H2N-HN-DCRETGSSKYPNCAYKTTQANKHIIVACEGNPYVPVHFDASV (Rib 4)

Peptide DCRETGSSKYPNCAYKTTQANKHIIVACEGNPYVPVHFDASV was synthesised on the Prelude peptide synthesiser using Fmoc-Val Wang resin (300 mg, 0.27 mmol/g) and using the pseudoprolines GS, KT and AS. All amino acids and pseudoprolines and HBTU were used in a 3 fold excess, using the standard program detailed in **Section 7.2.1.7**. The N-terminus was modified using hydrazinoacetic acid (100mg), PyBop (135 mg) and DIPEA (125 μl) in DMF. The reaction was moitored by pervorming a small scale cleavage of the resin and monitoring by MS. The peptide was cleaved from the resin using the method in **Section 2.7.5.1,** and purified by RP-HPLC (**Section 7.3.1.2**), to yield the final peptide (76 mg, 20% from initial resin substitution)

7.2.9.2 Synthesis of SDCRETGSSKYPNCAYKTTQANKHIIVACEGNPYVPVHFDASV (Rib 4)

Peptide SDCRETGSSKYPNCAYKTTQANKHIIVACEGNPYVPVHFDASV was synthesised on the Prelude peptide synthesiser using Fmoc-Val Wang resin (300 mg, 0.27 mmol/g) and using the pseudoprolines GS, KT and AS. All amino acids and pseudoprolines and HBTU were used in a 3 fold excess, using the standard program detailed in **Section 7.2.1.7**. The peptide was cleaved from the resin using the method in **Section 2.7.5.1,** and purified by RP-HPLC (**Section 7.3.1.2**). The peptide was repurified by RP-HPLC to achieve the highest purity possible, to yield the final peptide (54 mg, 15% from initial resin substitution, 100% purity by HPLC (data in **appendix II**)).

7.2.9.3 Synthesis of SDC(S-tBu)RETGSSKYPNC(StBu)AYKTTQANKHIIVAC(S-tBu)EGNPYVPVHFDASV (Rib 4 StBu)

Peptide SDC(S-tBu)RETGSSKYPNC(S-tBu)AYKTTQANKHIIVAC(S-tBu)EGNP

YVPVHFDASV was synthesised on the Prelude peptide synthesiser using Fmoc-Val Wang resin (300 mg, 0.27 mmol/g) and using the pseudoprolines GS, KT and AS. All amino acids and pseudoprolines and HBTU were used in a 3 fold excess, using the standard program detailed in **Section 7.2.1.7**. The peptide was cleaved from the resin using the method in **Section 2.7.5.1,** and purified by RP-HPLC (**Section 7.3.1.2**), to yield the final peptide (45 mg, 11% from initial resin substitution)

7.2.9.4 Synthesis of SDC(ACM)RETGSSKYPNC(ACM)AYKTTQANKHIIVAC(ACM)EGNPYVPVHFDASV (Rib 4 ACM)

Peptide SDC(ACM)RETGSSKYPNC(ACM)AYKTTQANKHIIVAC(ACM)EGNP YVPVHFDASV was synthesised on the Prelude peptide synthesiser using Fmoc-Val Wang resin (300 mg, 0.27 mmol/g) and using the pseudoprolines GS, KT and AS. All amino acids and pseudoprolines and HBTU were used in a 3 fold excess, using the standard program detailed in **Section 7.2.1.7**. The peptide was cleaved from the resin using the method in **Section 2.7.5.1,** and purified by RP-HPLC (**Section 7.3.1.2**), to yield the final peptide (52 mg, 13% from initial resin substitution)

7.2.9.5 Synthesis of SDCRETGSSK

Peptide SDCRETGSSK was synthesised on the Prelude peptide synthesiser using Fmoc-Lys Wang resin (300 mg, 0.33 mmol/g). All amino acids and HBTU were used in a 3 fold excess, using the standard program detailed in **Section 7.2.1.7**. The peptide was cleaved from the resin using the method in **Section 2.7.5.1,** and purified by RP-HPLC (**Section 7.3.1.2**), to yield the final peptide (54 mg, 51% from initial resin substitution).

7.2.9.6 Synthesis of SDC(S-tBu)RETGSSK

Peptide SDC(S-tBu)RETGSSK was synthesised on the Prelude peptide synthesiser using Fmoc-Lys Wang resin (300 mg, 0.33 mmol/g). All amino acids and HBTU were used in a 3 fold excess, using the standard program detailed in **Section 7.2.1.7**. The peptide was cleaved from the resin using the method in **Section 2.7.5.1,** and purified by RP-HPLC (**Section 7.3.1.2**), to yield the final peptide (38 mg, 34% from initial resin substitution).

7.2.10 Oxidation of Peptide Fragments Bearing an N-terminal Serine or C-terminal Amino Alcohol

Peptides were oxidised using sodium metaperiodate, at an N-terminal serine to yield a glyoxal or a C-terminal amino alcohol to yield an aldehyde. Oxidation was either performed using solid supported $NaIO_4$ or in solution as described below.

7.2.10.1 Polymer Supported Oxidation

Polymer supported periodate resin was prepared from Amberlyst A26 by ion exchange with sodium metaperiodate. Amberlyst A26 (20 g) was treated with an aqueous solution of $NaIO_4$ (1 M, 100 ml), by slowly passing the solution through the resin in a PD 10 column.

The peptide to be oxidised was dissolved in MeOH (5-10 mg/ml) and this was added to polymer supported periodate resin. The slurry was stirred at room temperature for 30-60 mins then the solution filtered away from the resin and analysed by MS. If the oxidation reaction was not complete the solution was returned to the resin and stirring continued for a further 30-60 mins. Once the reaction was complete as determined by MS the MeOH was removed by evaporation, and the peptide aldehyde stored at -20 ºC.

7.2.10.1.1 Polymer Supported Oxidation of KPVNTFVHESL-amino alcohol

KPVNTFVHESL-amino alcohol (1.3 mg) was oxidised to KPVNTFVHESL-aldehyde using polymer supported $NaIO_4$ using the method in **Section 7.2.10.1**. The peptide was analysed by ESMS [M + H] 1254.

7.2.10.1.2 Polymer Supported Oxidation of YQSYSTMSI-amino alcohol

YQSYSTMSI-amino alcohol (2.4 mg) was oxidised to YQSYSTMSI -aldehyde using polymer supported $NaIO_4$ using the method in **Section 7.2.10.1**. The peptide was analysed by ESMS [M + H] 1063.

7.2.10.2 Solution Phase Oxidation

The peptide to be oxidised was dissolved in imidazole buffer (50 mM, pH 6.9) at a concentration of 5mg/ml. $NaIO_4$ (2 eq) was added and the reaction stirred at room temperature. The reaction was monitored using LC-MS until oxidation was complete, usually 10-30 mins. The reaction was quenched using ethylene glycol (2 eq) and the product immediately purified by RP-HPLC. The purified peptide aldehyde or glyoxal was lyophilised and stored at -20 ºC.

7.2.10.2.1 Oxidation of SLIGRL-NH2

Peptide SLIGRL-NH2 (50 mg) was dissolved in imidazole buffer (50 mM, pH 6.9) at a concentration of 5mg/ml. $NaIO_4$ (32 mg, 2 eq) was added and the reaction stirred at room temperature for 20 minutes. The reaction was quenched using ethylene glycol (8.5 μl, 2 eq) and the product immediately purified by RP-HPLC, and analysed by MS. [M + H] 626, [M + H_2O] 64

7.2.10.2.2 Oxidation of SDCRETGSSK

Peptide SDCRETGSSK (20 mg) was oxidised in solution using the method in **Section 7.2.10.2**. The reaction was monitored by MS and gave multiple unidentifiable products.

7.2.10.2.3 Oxidation of SDC(S-tBu)RETGSSK

Peptide SDC(S-tBu)RETGSSK (15 mg) was oxidised in solution using the method in **Section 7.2.10.2**. The reaction was monitored by MS and gave the expected glyoxal-DC(S-tBu)RETGSSK and the hydrated version. [M + H] 1026 [M + H_3O] 1144

7.2.10.2.4 Oxidation of SDCRETGSSKYPNCAYKTTQANKHIIVACEGNPYVPVHFDASV, and S-tBu and ACM protected versions

Peptide SDCRETGSSKYPNCAYKTTQANKHIIVACEGNPYVPVHFDASV was oxidised in solution was oxidised in solution using the method in **Section 7.2.10.2**. The reaction was monitored by MS and gave multiple peaks separated by 6 mass units corresponding to multiple oxidation products.

Peptide SDC(S-tBu)RETGSSKYPNC(S-tBu)AYKTTQANKHIIVAC(S-tBu)EGNP YVPVHFDASV was oxidised in solution by the same method and analysed by MS, which gave the expected mass for the hydrated glyoxal. The peptide was purified by HPLC and stored at -20 ºC until required. ESMS [M + H_3O] 4689.

Peptide SDC(ACM)RETGSSKYPNC(ACM)AYKTTQANKHIIVAC(ACM)EGN PYVPVHFDASV was oxidised in solution by the same method and analysed by MS, which gave the expected mass for the glyoxal and hydrated glyoxal. The peptide was purified by HPLC and stored at -20 ºC until required ESMS [M+ H] 4882 [M + H_3O] 4900.

7.2.11 Chemoselective Ligation of Peptides

The peptides to be ligated were dissolved in NaOAc buffer (0.1 M, pH 4.6), at a concentration of 1 mg/ml unless specified otherwise, and the reaction stirred at room temperature. The reaction was monitored by LC-MS, using the HPLC to remove the buffer, allowing analysis of the reaction by ESMS. When ligation was compete, the product was purified by RP-HPLC, and the ligated product was lyophilised and stored at 4 ºC.

7.2.11.1 Initial Ligations of 8Aoc-fQWAVGHL-NH2-NH2 to glyoxal-LIGRL-NH2

8Aoc-fQWAVGHL-NH2-NH2 (7 mg) and glyoxal-LIGRL-NH2 (4 mg) were dissolved in NaOAc buffer (11 ml, 0.1 M, pH 4.6). The reaction was stirred at room temperature for 3 days and monitored by ESMS after 24 hours and 3 days. The reaction had proceeded to completion so the reaction mixture was purified by HPLC and the purified peptide was lyophilised and stored at 4 ºC.

8Aoc-fQWAVGHL-NH2-NH2 (7 mg) and glyoxal-LIGRL-NH2 (4 mg) were dissolved in NaOAc buffer (5.5 ml, 0.1 M, pH 4.6). The reaction was stirred at room temperature for 3 days and monitored by ESMS after 24 hours, 48 hours and 3 days. The recation as not complete after 48 hours. The reaction had proceeded to completion after 3 days, so the reaction mixture was purified by HPLC and the purified peptide was lyophilised and stored at 4 ºC.

7.2.11.2 Optimised Ligation of 8Aoc-fQWAVGHL-NH2-NH2 to glyoxal-LIGRL-NH2

8Aoc-fQWAVGHL-NH2-NH2 (7 mg) and glyoxal-LIGRL-NH2 (4 mg) were dissolved in NaOAc buffer (11 ml, 0.1 M, pH 4.6 containing 10 mM aniline). The reaction was stirred at room temperature for 18 hours and monitored by ESMS to

ensure it had reached completion. The reaction mixture was purified by HPLC and the purified peptide was lyophilised and stored at 4 ºC.

7.3 Equipment and Instrumentation

Manual peptide synthesis was performed using an orbital shaker (IKA). Empty PD 10 columns (Amersham Bioscience) were used as peptide synthesis reaction vessels. Alternatively Teflon peptide synthesis reaction vessels (Activotec) were used. Automated peptide synthesis was carried out on either the Symphony or Prelude peptide synthesisers (Protein Technologies Inc.). Solvent was evaporated under reduced pressure using a rotary evaporator (Buchii) and Laboport® pump (KNF).

Column chromatography was performed using silica gel 60 (Alfa Aesar). Column chromatography was monitored by thin layer chromatography (TLC) on Keiselgel 60 F_{254} pre coated plates (Sigma Aldrich). Compounds were observed under a UV lamp at 254 nm.

7.3.1 Reverse Phase High Performance Liquid Chromatography

Reverse phase high performace liquid chromatography (RP-HPLC) was used to purify peptides and to analyse crude and purified peptides. RP-HPLC was also used to analyse derivatives of amino acids.

7.3.1.1 HPLC Analysis

Peptides and amino acid derivatives were analysed by reverse phase high pressure liquid chromatography (RP-HPLC) using a Gilson HPLC machine running either 712 or Unipoint software. The system consisted of two slave Gilson 306 pumps, 805 manometric module, 118C dynamic mixer, and 155 UV/Vis detector. Columns used were C18 primesphere or Waters symmetry analytical columns, 5 μm, 4.6 mm x 150

mm 100 Å pore size. Analysis was performed using a linear gradient of two solvents, A (H_2O containing 0.1% v/v TFA) and B (acetonitrile containing 0.1% v/v TFA), running from 0%-80% B over 20 mins or 10%-90% B over 20 mins at 1 ml/min.

7.3.1.2 HPLC Purification

Peptides were purified by preparative RP-HPLC using a Varian Prostar 210 system, consisting of two slave pumps and UV/Vis detector running Varian Star software. The column used was a C18 Gemini column, 5 μm, 10.0 mm x 250 mm 300 Å pore size. Purification was performed using a linear gradient of two solvents, A (H_2O containing 0.1% v/v TFA) and B (acetonitrile containing 0.1% v/v TFA), typically running from 10%-60% B over 20 mins at 5 ml/min.

7.3.2 Mass Spectrometry

Mass spectrometry (MS) of amino acid derivatives and peptides were recorded at Peptide Protein Research Ltd, on a Micromass ZMD 4000 mass spectrometer. Liquid Chromatography-Mass Spectrometry (LC-MS) was performed using an Agilent 1100 series Liquid Chromatography-Mass Spectrometer, running Chemstation software. Mass spectra were recorded for all amino acid derivatives and peptides. LC-MS was used to monitor oxidation and ligation reactions, and for analysis of amino acid derivatives and peptides.

7.3.2.1 Electrospray Mass Spectrometry of Peptides

All peptide mass spectra were recorded using an electrospray mass spectrometer. Electrospray ionisation is particularly suitable for peptides as it allows the peptide to be ionised without fragmentation. Ionised peptides are often multiply charged, which means that the spectra recorded show multiple peaks on a mass/charge (m/z) scale,

correlating to the different ionisation states of the peptide. A computer program can then be used to calculate the actual mass of the peptide from the original spectrum.

7.3.3 Nuclear Magnetic Resonance Spectroscopy

Nuclear Magnetic Resonance (NMR) Spectroscopy was recorded at Surrey University using a Brucker DRX-300 pulse Fourier transform spectrometer operating at 300 MHz, or at Reading University Chemical Analysis Facility (CAF) using a Brucker DPX400 Fourier transform spectrometer operating at 400 MHz. Spectra were recorded for final compounds and key intermediates.

7.3.4 Optical Spectroscopy

Fourier Transform Infra-red Spectroscopy (FTIR) was recorded at Reading University CAF using a PerkinElmer Spectrum 100 FTIR spectrometer.

7.4 Molecular Dynamics Simulations

The software package used for molecular dynamics (MD) simulations at the University of Surrey was Molecular Operating Environment (MOE 2007.09, Chemical Computing Group Inc.).

7.4.1 Running a Molecular Dynamics Simulation

A number of individual processes were performed in order to carry out a protein molecular dynamics simulation using MOE. The relative positions of all the atoms were first loaded into the software from a Protein Data Bank (pdb) file, which has the co-ordinates of every atom position taken from a solved crystal structure. This structure was placed inside a virtual box of water, and then an energy minimisation

was performed, Once the energy minimisation was performed, the molecular dynamics process was started. The simulation generated a database of entries after each individual energy calculation, which are essentially snapshots of the system as an experimentally derived timecourse through the simulation.

7.4.2 Molecular Dynamics Simulation of Bovine Ribonuclease A

A pdb file for bovine ribonuclease A was loaded into MOE. The crystal structure that was chosen for the dynamic simulation was a pdb file called 3rn3, from the publication by Howlin *et al*. (1989). An energy minimisation was run on this structure using the AMBER 99 forcefield. The energy minimised structure was saved and this was placed inside a box of water by running a function called water soak. The size of the box was set to just larger than the ribonuclease molecule, and periodic boundary conditions were applied (**Chapter 5, Section 5.1.4.3**). The system was constrained to NTP (**Chapter 5, Section 5.1.4.4**) and the molecular dynamics simulation was run using a timestep of 0.5 ps and the number of timesteps set to 200.

7.4.2.1 Replacement of Ala20 with an Oxime Bond

Using the original crystal structure of bovine ribonuclease A, 3rn3 (Howlin *et al*. (1989)) loaded into MOE, the site for a single chemoselective ligation replacement for Ala20 was identified. The software has the facility to delete parts of the structure and draw replacement structures atom by atom. Using the software to hide the entire molecule except Ala20 and the immediately adjacent amino acids, the atoms in the native peptide bond were deleted and replaced with the structure of an oxime bond, formed through reaction of an N-terminal hydroxylamine with a C-terminal aldehyde. An energy minimisation using the AMBER 99 forcefield was performed, firstly just on the replaced atoms, and then using a proximity tool to identify all atoms within a 10 Å distance of the changed atoms. The whole system was then energy minimised.

7.4.2.2 Replacement of Ala20, Ala52 and Thr82 with a hydrazide chemoselective bond

Using the original crystal structure of bovine ribonuclease A, 3rn3 (Howlin *et al*. (1989)), loaded into MOE, the site for a single chemoselective ligation replacement for Ala20 was identified. The atoms in the native peptide bond were deleted and replaced with the structure of a hydrazide bond formed through reaction between an N-terminal glyoxal and a C-terminal hydrazide. After the bond was replaced, the structure was run through an energy minimisation using the AMBER 99 forcefield, firstly just on the replaced atoms, and then using the proximity tool to identify all atoms within a 10 Å distance of the changed atoms. The process was repeated for Ala52 and Thr82, replacing these with the hydrazide bond. Once all three bonds had been replaced and the energy minimised, the whole system was energy minimised once more.

7.4.2.3 Molecular Dynamics Simulation of RNase A containing ligated bonds

The energy minimised structure of ribonuclease containing three hydrazide bonds (**Section 7.4.1.2**) was placed inside a box of water using the water soak function in MOE. The same constraints were applied as used for the MD simulation of native ribonuclease (**Section 7.4.1.1**). The MD simulation was run using a timestep of 0.5ps and the number of timesteps set to 400.

7.4.2.4 Replacement of Cys65 with Ala

Using the original crystal structure of bovine ribonuclease A, 3rn3 (Howlin *et al*. (1989)), loaded into MOE, the amino acid Cys65 was located. The disulfide between Cys65 and Cys72 was deleted and the side chain of Cys65 was replaced with a methyl group to produce the mutant Ala65. The structure was run through an energy minimisation using the AMBER 99 forcefield, firstly using the proximity tool to

identify all atoms within a 10 Å distance of the changed atoms, and then on the whole system.

7.4.2.5 Molecular Dynamics Simulation of Cys65Ala RNase A

The energy minimised structure of ribonuclease containing Cys65Ala was placed inside a box of water using the water soak function in MOE. The same constraints were applied as used for the MD simulation of native RNase A (**Section 7.4.1.1**). The MD simulation was run using a timestep of 0.5ps and the number of timesteps set to 400.

Bibliography

A

Al-Gharabli, S. I., Shah, S. A., Weik, S., Schmidt, M. F., Mesters, J. R., Kuhn, D., Klebe, G., Hilgenfeld, R. and Rademann, J. (2006) An Efficient Method for the Synthesis of Peptide Aldehyde Libraries Employed in the Discovery of Reversible SARS Coronavirus Main Protease (SARS-CoV Mpro) Inhibitors *ChemBioChem* **7** 1048-1055

Anfinsen, C. B. and Haber, E. (1960) Studies on the Reduction and Re-formation of Protein Disulfide Bonds *J. Biol.Chem* **236** 1361-1363

Anfinsen, C, B., Haber, E., Sela, M. and White Jr, F. H. (1961) The Kinetics of Formation of Native Ribonuclease During Oxidation of the Reduced Polypeptide Chain *Proc. Natl. Acad. Sci.* **47** 1309-1314.

Atherton, E. and Sheppard, R.C. (1989). Solid Phase peptide synthesis: a practical approach. Oxford, England: IRL Press. ISBN 0199630674

B

Backes, J. B., Virgilio, A. A., Ellman, J.A. (1996) Activation method to prepare a highly reactive acylsulfonamide "safety-catch" linker for solid-phase synthesis *J. Am. Chem. Soc.* **118** 3055-3056

Balvert-Geers, I. C. and Tesser G. I. (1975) The methylsulfonylethyloxycarbonyl group, a new and versatile amino protective function *Int. J. of Pept. Pro. Res.* **4** 295-305

Boerema, D. L. (2007) Total Chemical Synthesis of Ribonuclease A and Designed Analogues with Enhanced Catalytic Activity. *PhD Thesis, University of Chicago*

Botstein, D., Falco, S.C., Stewart, S.E., Brennan, M., Scherer, S., Stinchcomb, D.T., Struhl, K. and Davis R.W. (1979) Sterile host yeasts (SHY): A eukaryotic system of biological containment for recombinant DNA experiments *Gene* **8** 17-24

Broadbridge, R. J. (1998) Design and Synthesis of Novel Inhibitors to the SH2 Domain of the Protein Kinase Inhibitor p56lck. *PhD Thesis, University of Southampton*

Broadbridge, R. J., Winsor, C. E., and Sharma, R. P. (2001) 1-Amino 2,3-Diols: Applications in Peptide Aldehydes and Chemical Ligation, *"Innovation & Perspectives in*

Solid Phase Synthesis & Combinatorial Libraries" (Proceedings of the 7th International Symposium, Southampton) 33-38

Bruckdorfer. T., Marder. O., and Albericio. F. (2004) From production of peptides in milligram amounts for research to multi-tons quantities for drugs of the future *Curr Pharm Biotechnol.* **1** 29-43

C

Chatani, E., Tanimizu, N., Ueno, H., and Hayashi, R. (2001) Structural and functional changes in bovine pancreatic ribonuclease A by the replacement of Phe120 with other hydrophobic residues. *J Biochem* **129** 917–922

Chesney, A.(1999) Selected highlights in the application of ion-exchangers as supports for reagents in organic synthesis *Green Chemistry* **1** 209-219

Clamp, J. R. and Hough, L. (1965) The Periodate Oxidation of Amino Acids with Reference to Studies on Glycoproteins *Biochem. J.* **94**, 17-24

Chini, M., Croti, P. and Macchia, F. (1990) Metal salts as new catalysts for mild and efficient aminolysis of oxiranes *Tetrahedron Lett.* **31** 4661-4664

Chini, M., Crotti, P., Favero, L., Macchia, F. and Pineschi, M. (1994) Lanthanide (III) Trifluoromethanesulfonates as Extraordinarily Effective New Catalysts for the Aminolysis of 1,2-Epoxides *Tetrahedron Lett.***35** 433-436

Cohen, S., Chang, A., Boyer, H. and Helling, R. (1973) Construction of biological functional bacterial plasmids in vitro *PNAS* **70** 3240-3244

Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M. Jr, Ferguson. D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules *J. Am. Chem. Soc.* **117** 5179–5197

Crook, E. M., Mathias, A. P., and Rabin, B. R. (1960) Spectrophotometric Assay of Bovine Pancreatic Ribonuclease by the use of Cytidine 2':3'-Phosphate *Biochem J.* **74** 234-238

D

Dawson, P. E., Muir, T. W., Clark-Lewis, I and Kent, S.B.H. (1994). "Synthesis of Proteins by Native Chemical Ligation". *Science* **266** 776–779

delCardayré S. B. and Raines. R.T. (1994) Structural determinants of enzymatic processivity *Biochemistry* **33** 6031-7

Dirksen, A., Hackeng, T. M. and Dawson, P. E., (2006) Nucleophilic Catalysis of Oxime Ligation Angew. *Chem. Int. Ed.* **45** 7581 –7584

Dirksen, A. and Dawson, P. E. (2008) Rapid Oxime and Hydrazone Ligations with Aromatic Aldehydes for Biomolecular Labeling *Bioconjugate Chem.* **19** 2543–2548

E

Ede, N. J., Eagle, S. N., Wickham, G., Bray, A. M. and Warne, B. (2000) Solid Phase Synthesis of Peptide Aldehyde Protease Inhibitors. Probing the Proteolytic Sites of Hepatitis C Virus *J. Peptide Sci.* **6** 11–18

Englebretsen, D. R., Garnham, B. G., Bergman, D. A. and Alewood, P. F (1995) A novel thioether linker: Chemical synthesis of a HIV-1 protease analogue by thioether ligation *Tetrahedron Lett*, **48** 8871-8874

F

Fehrentz , J., Paris, M. Heitz, A., Velek, J., Liu, C., Winternitz, F., and Martinez, J. (1995) Improved solid phase synthesis of C-terminal peptide aldehydes *Tetrahedron Lett.* **36** 7871-7874

Findlay D., Herries D.G., Mathias A.P., Rabin B.R. and Ross C.A. (1961) The active site and mechanism of action of bovine pancreatic ribonuclease. *Nature* **190** 781–784

G

Gartiser, T. Selve, C. and Delpuech J. -J. (1983) Reduction d'azides en amines par le formiate d'ammonium par "Transfert d'Hydrogene Catalyse" (CTH) *Tetrahedron Lett.* **24** 1609-1610

Geoghegan, K. F. and Stroh J. G.(1992) Site-directed conjugation of nonpeptide groups to peptides and proteins via periodate oxidation of a 2-amino alcohol. Application to modification at N-terminal serine *Bioconjugate Chem.* **3** 138-146

Glacon, V., El Meslouti, A., Uzan, R., Demailly, G. and Beaupere, D. (1996) A new efficient way to α,o-diaminoitols by direct azidation of unprotected itols, *Tetrahedron Lett.* **37** 3683-3686

Gutte, B and Merrifield, R. B (1971) The Synthesis of Ribonuclease A *J. Biol. Chem* **246** 1922

H

Hackeng T. M., Griffin, J. H. and Dawson P. E., (1999) Protein synthesis by native chemical ligation: Expanded scope by using straightforward methodology *Proc. Natl. Acad. Sci.* **96** 10068-10073

Hirs, C. H. W. (1960) Studies on the Structure of Ribonuclease. Enzymatic Hydrolysis of the Four Large Peptides Formed in the Tryptic Hydrolysis of the Oxidized Protein *J. Biol. Chem.* **235** 625-632

Hirs, C. H. W., Moore, S. and Stein, W. H. (1960) The Sequence of the Amino Acid Residues in Performic Acid-oxidized Ribonuclease *J. Biol. Chem.* **235** 633-647

Hockney, R. W. (1970) The potential calculation and some applications *Methods in Computational Physics* **9** 135-211

Hossain, M. A., Belgi, A., Lin, F., Zhang, S., Shabanpoor, F., Chan, L., Belyea, C., Truong, H., Blair, A. R., Andrikopoulos, S, Tregear, G. W., and Wade , J.D. (2009) Use of a Temporary "Solubilizing" Peptide Tag for the Fmoc Solid-Phase Synthesis of Human Insulin Glargine via Use of Regioselective Disulfide Bond Formation *Bioconjugate Chemistry* **20** 1390-1396

Howlin B, Moss DS, Harris GW. (1989) Segmented anisotropic refinement of bovine ribonuclease A by the application of the rigid-body TLS model. *Acta Crystallogr. A.* **45** 851-861

I

Ingenito, R., Bianchi, E., Fattori, D., and Pessi, A. (1999) Solid Phase Synthesis of Peptide C-Terminal Thioesters by Fmoc/t-Bu Chemistry *J. Am. Chem. Soc.*, **121** 11369–11374

J

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79** 926

K

Kenner G. W., McDermott J. R. and Sheppard R. C. (1971) The safety catch principle in solid phase peptide synthesis *J. Chem. Soc.* D, 636-637

Kent, S.B.H. and Dawson, P. (2000) Synthesis of native proteins by chemical ligation, *Annual Review of Biochemistry* **69** 925-962

Kunitz, M. (1940) Crystalline Ribonuclease *J. Gen. Physiol.* **24** 15-32

L

Lennard-Jones, J. E. (1924) On the Determination of Molecular Fields *Proc. R. Soc. Lond. A* **106** 463–477

Liu, C. F., Rao, C. and Tam, J. P., (1996) Orthogonal Ligation of Unprotected Peptide Segments through Pseudoproline Formation for the Synthesis of HIV-1 Protease Analogs *J. Am. Chem. Soc*. **188** 307-312

Liu, C.F. and Tam, J.P. (1994) Chemical Ligation Approach to Form a Peptide Bond Between Unprotected Peptide Segments. Concept and Model Study. *J. Am. Chem. Soc*. **116** 4149-53

Low, D.W., Hill, M. G., Carrasco M.R., Kent S.B.H. and Botti, P. (2001) Total synthesis of cytochrome b562 by native chemical ligation using a removable auxiliary *PNAS* **98** 6554-6559

Lue, R. Y., Chen, G. Y., Hu, Y. Zhu Q. and Yao, S. Q. (2004) Versatile Protein Biotinylation Strategies for Potential High-Throughput Proteomics *J. Am.Chem. Soc.* **126** 1055–1062

M

Merrifield, R. B. (1963) Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide *J. Am. Chem. Soc.* **85** 2149

Merrifield, R. B., Barany, G., Cosand, W. L., Engelhard, M. and Mojsov, S. (1977) *Pept. Proc. Am. Pept. Symp. 5th* 488

Muir, T.W., Dawson, P.E., and Kent, S.B.H. (1997) Protein Synthesis by Chemical Ligation of Unprotected Peptides in Aqueous-Solution. *Methods Enzymol.* **289** 266-298

Murdock, A.L. ,Grist, K. L. and Hirs, C.H.W. (1966) On the dinitrophenylation of bovine pancreatic ribonuclease A. Kinetics of the reaction in water and 8 M urea. *Arch Biochem Biophys* **114** 375–390

Mutter, M., Nefzi, A., Sato, T., Sun, X., Wahl, F. and Wohr, T., (1995) Pseudo-Prolines (Psi-Pro) for Accessing Inaccessible Peptides, *J. Pept. Res*, **8** 145-153

N

Nahm, S. and Weinreb, S. M. (1981**)** N-methoxy-n-methylamides as effective acylating agents *Tetrahedron Lett.* **22** 3815

Nilsson, B. L., Hondal, R. J., Soellner, M., N., and Raines, R. T. (2003) Protein assembly by orthogonal chemical ligation methods. *J. Am. Chem. Soc.* **125** 5268-5269

Nilsson, B.L., Kiessling, L.L. and Raines, R.T. (2000) Staudinger ligation: A peptide from a thioester and azide *Org Lett.* **2** 1939-1941

Nogues, M. V., Vilanova M. and Cuchillo C. M. (1995) Bovine pancreatic ribonuclease A as a model of an enzyme with multiple substrate binding sites *Biochimica et Biophysica Acta* **1253** 16-24

Nyfeler, R., Wixmerten, U., Seidel, C., and Mergler, M. (1992) Peptide synthesis by a combination of solid phase and solution methods Peptides, *Proc. 12th APS 1991* (Smith, J A. and Rivier, J. E., eds.), Escom, Leiden 661-663

Nyfeler, R. (1994) Peptide synthesis via fragment condensation *Methods in Molecular Biology, Vol 35 Peptide Synthesis Protocols* **Chapter 15** 303-316 Edited by: M. W. Pennington and B M Dunn Copyright 01994 Humana Press Inc , Totowa, NJ

P

Paléus, S., and Neilands, J. B., (1950) Preparation of Cytochrome c with the Aid of Ion Exchange Resin *Acta Chem. Scand.* **4** 1024-1030

Parr, R, G., Craig D. P. and Ross, I. G (1950) Molecular Orbital Calculations of the Lower Excited Electronic Levels of Benzene, Configuration Interaction included. *Journal of Chemical Physics* **18** 1561-1563

R

Raines, R. (2004) Active Site of Ribonuclease A, *Nucleic Acids and Molecular Biology* **13** 19-32

Rose, K. (1994) Facile synthesis of homogenous artificial proteins, *J. Am. Chem. Soc.* **116** 30-33

Rothwarf, D. M., and Scheraga, H. A. (1991) Regeneration and reduction of native bovine pancreatic ribonuclease A with oxidized and reduced dithiothreitol *J. Am. Chem. Soc.* **113** 6293–6294

S

Schultz, L. W., Quirk, D. J., and Raines, R. T. (1998) His..Asp Catalytic Dyad of Ribonuclease A: Structure and Function of the Wild-Type, D121N, and D121A Enzymes *Biochemistry* **37** 8886-8898

Schilling, C. I., Jung, M., Biskup, M. and Schepers, U. (2011) Bioconjugation via azide–Staudinger ligation: an overview *Chem. Soc. Rev.* Advance Article First published on the web 17 Jun 2011

Sharma, R. P., Gore, M. G. and Akhtar, M. (1979) Stereospecific synthesis of α-aminoacyl aldehydes and their interaction with α-chymotrypsin *J. Chem. Soc. Chem. Comm.* 875-877

Smyth, D. G., Stein, W. H. and Moore, S. (1962) On the Sequence of Residues 11 to 18 in Bovine Pancreatic Ribonuclease The Sequence of Amino Acid Residues in Bovine Pancreatic Ribonuclease: Revisions and Confirmations *J. Biol. Chem.* **237** 1845-1850

Smyth, D. G., Stein, W. H. and Moore, S. (1963) The sequence of amino acid residues in bovine pancreatic ribonuclease: revisions and confirmations *J. Biol. Chem.* **238** 227-234

Soellner, M. B., Nilsson, B. L., and Raines R. T. (2002) Staudinger Ligation of α-Azido Acids Retains Stereochemistry *JOC* **67** 4993–4996

Spackman, D. H., Stein, W. H. and Moore, S. (1960) The Disulfide Bonds of Ribonuclease *J. Biol. Chem.* **235** 648-659

Stavropoulos, G., Gatos, D., Magafa, V., and Barlos, K. (1995) Preparation of polymer-bound trityl-hydrazines and their application in the solid phase synthesis of partially protected peptide hydrazides *Lett Peptide Sci* **2** 315–318

Staudinger, H. M., J. (1919) *Helvetica Chimica Acta* **2** 635-646

Swope, W. C., Andersen, H. C., Berens, P. H. and Wilson, K. R. (1982) *J. Chem. Phys.* **76** 637

T

Talluri. S., Rothwarf D., M. and Scheraga, H. A. (1994) Structural Characterization of a Three-Disulfide Intermediate of Ribonuclease A Involved in Both the Folding and Unfolding Pathways. *Biochemistry* **33** 10437-10449

Tam, J. P., Rao, C., Liu, C. and Shao, J. (1995) Specificity and formation of unusual amino acids of an amide ligation strategy for unprotected peptides, *Int J. Peptide Protein Res.* **45** 209-216

Tam, J. P. and Yu, Q. (1998) Methionine ligation strategy in the biomimetic synthesis of parathyroid hormones, *Biopolymers* **46** 319-327

Tanaka, M., Oishi, S., Ohno, H. and Fujii, N. (2007) A Novel Oxazolidine Linker for the Synthesis of Peptide Aldehydes *Int. J. of Peptide Research and Therapeutics* **13** 271-279

Tanimizu N, Ueno H. and Hayashi R (1998) Role of Phe120 in the activity and structure of bovine pancreatic ribonuclease A. *J Biochem* **124** 410–416

Tarragona-Fiol, A., Eggelte, H.J., Harbron, S., Sanchez, E., Taylorson, C.J. , Ward, J.M. and Rabin, B.R. (1993) Identification by site-directed mutagenesis of amino acids in the B2 subsite of bovine pancreatic ribonuclease A *Protein Engineering* **6** 901-906

Thompson, J.E. and Raines. R.T. (1994) Value of general acid–base catalysis to ribonuclease A. *J Am Chem Soc.* **116** 5467–5468

Trautwein, K., Holliger, P., Stackhouse, J. and Benner, S.A. (1991) Site-directed mutagenesis of bovine pancreatic ribonuclease: Lysine-41 and aspartate-121. *FEBS Lett* **281** 275–277

Tuchscherer, G. (1993) Template assembled synthetic proteins: condensation of a multifunctional peptide to a topological template via chemoselective ligation, *Tetrahedron Lett.* **52** 8419-8422

V

Verlet, L. (1967). Computer Experiments on Classical Fluids I. Thermodynamical Properties of Lennard−Jones Molecules. *Physical Review* **159** 98–103

W

Wahl, F. and Mutter, M. (1996) Analogues of oxytocin with an oxime bridge using chemoselectively addressable building blocks, *Tetrahedron Lett.* **37** 6861-6864

Wang, J.M, Cieplak, P. and Kollman, P.A. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comp.Chem.* **21** 1049-1074

Wang, S.S. and Merrifield, R. B. (1969) Preparation of a Tert-alkyloxycarbonylhydrazide Resin and Its Application to Solid Phase Peptide Synthesis *J Am Chem Soc.* **91** 6488–6491

Wang, S.S. (1973) p-alkoxybenzyl alcohol resin and p-alkoxybenzyloxycarbonylhydrazide resin for solid phase synthesis of protected peptide fragments *J Am Chem Soc.* **95** 1328–1333

Wieland, T., Bokelmann, E., Bauer, L., Lang, H. U. and Lau, H (1953) Über Peptidsynthesen. 8. Mitteilung Bildung von S-haltigen Peptiden durch intramolekulare Wanderung von Aminoacylresten *Liebigs Ann. Chem.* **583** 129

Woghiren, C., Sharma, B. and Stein, S. (1993) Protected thiol-polyethylene glycol: A new activated polymer for reversible protein modification *Bioconjugate Chem.* **4** 314–318

Y

Yan L.Z. and Dawson P.E. (2001) Synthesis of Peptides and Proteins without Cysteine Residues by Native Chemical Ligation Combined with Desulfurization *J. Am. Chem. Soc*. **123** 526–533

Z

Zhang, L, Torgerson, T. R., Liu, X., Timmons, S., Colosia, A. D., Hawiger , J. and Tam, J. P., (1998) Preparation of functionally active cell-permeable peptides by single-step ligation of two peptide modules *Proc. Natl. Acad. Sci.* **95** 9184–9189

1. http://www.ornl.gov/sci/techresources/Human_Genome/glossary/glossary_p.shtml accessed on 07-12-05
2. http://www.harolda_scheraga-biophysicalchemistry-cornelluniversity.html/ accessed on 09/12/03
3. http://http:/www.people.virginia.edu/~rjh9u/ accessed on 29/11/03
4. http://www.lonza.com/group/en/products_services/Custom_Manufacturing/peptides_oligo/technical_literature.-ParSys-0001-ParSysdownloadlist-0001-DownloadFile.pdf/Presentation%20-%20Peptides%20At%20a%20Glance%20-%20Lonza%20Custom%20Manufacturing%202009.pdf accessed on 07/04/10
5. Novabiochem catalogue 2011 Method 2-8 Section 2.29
6. Novabiochem catalogue 2011 Method 3-6 Section 3.6
7. http://www.chemcomp.com/

Appendix I

Conferences

Oral Presentations

Winsor, C. E. and Broadbridge R. J. (2006) A 1, 2-Amino Alcohol Derivative of Fmoc Amino Acids: A Masked Aldehyde for use in Sequential Chemoselective Ligation *Dr Bert L Schram Young Investigators' Mini-Symposium 29th European Peptide Symposium, Gdansk Poland*

Winsor, C. E. and Broadbridge R. J. (2007) Synthesis of C-Terminal Peptide Aldehydes Via Oxidation of a 1, 2- Amino Alcohol *Young Investigator Symposium 9th International Symposium on Solid Phase Synthesis, Complementary Solution Methods and Combinatorial Libraries, Norwich, UK*

Poster Presentations

Winsor, C. E. and Broadbridge R. J. (2006) A 1, 2-Amino Alcohol Derivative of Fmoc Amino Acids: A Masked Aldehyde for use in Sequential Chemoselective Ligation (2006) *Festival of Research, University of Surrey, UK*

Winsor, C. E., Broadbridge R. J. and Howlin, B. (2009) Molecular Dynamics Simulations of Chemoselective Bonds in Bovine Ribonuclease A *21st American Peptide Symposium, Bloomington, Indiana, USA*

Appendix IIa

Analytical data for derivatives of Phenylalanine

Fmoc-phenylalanine diazomethylketone

^{1}H NMR, ^{1}H COSY, ^{13}C NMR, FTIR

Fmoc-phenylalanine bromomethylketone

LCMS, ^{1}H NMR, ^{1}H COSY, ^{13}C NMR, FTIR

Fmoc-phenylalanine azidomethylketone

LCMS, ^{1}H NMR, ^{1}H COSY, ^{13}C NMR, FTIR

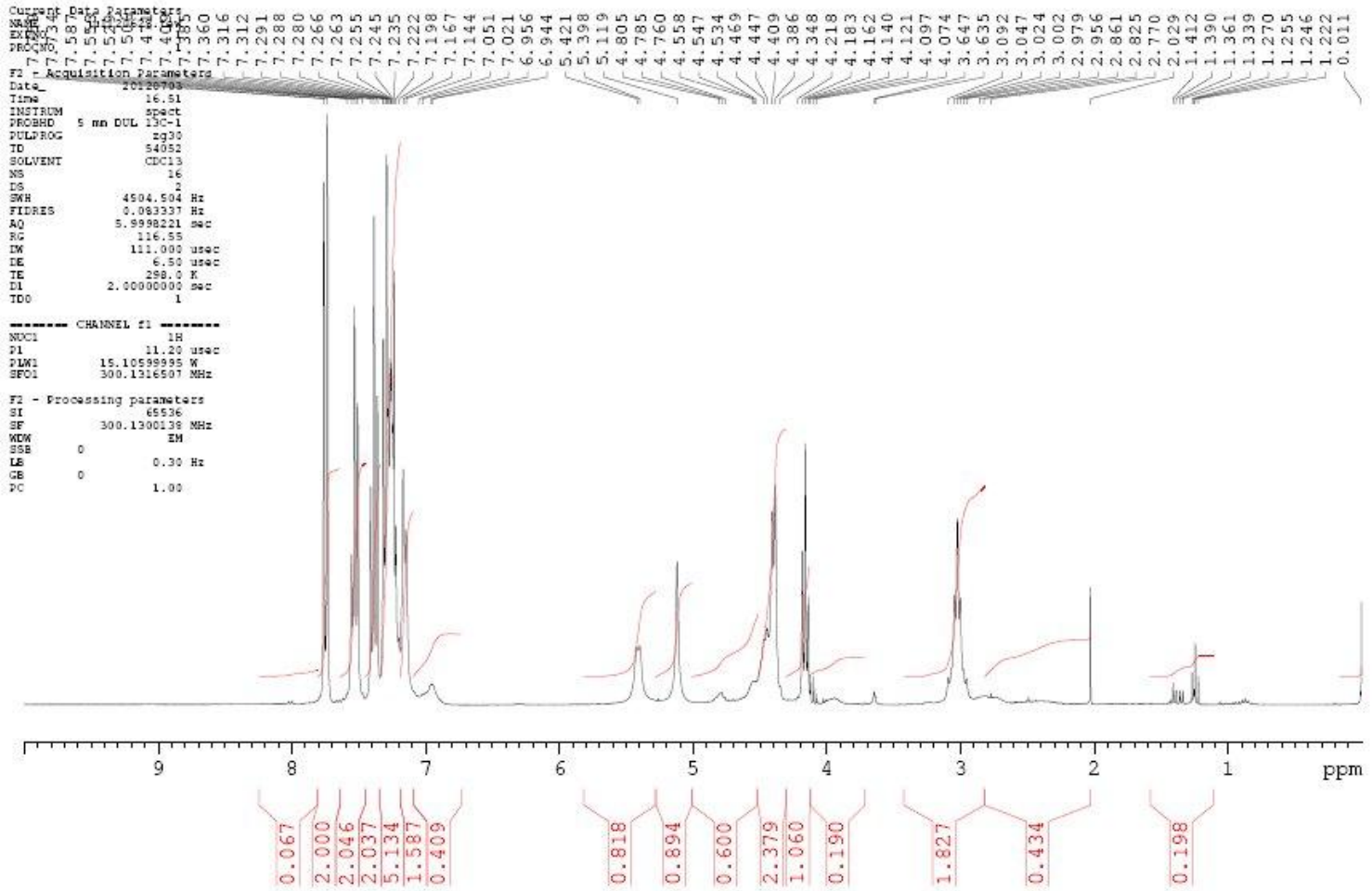
Fmoc-phenylalanine azidomethylalcohol

LCMS, ^{1}H NMR, ^{1}H COSY, ^{13}C NMR, FTIR

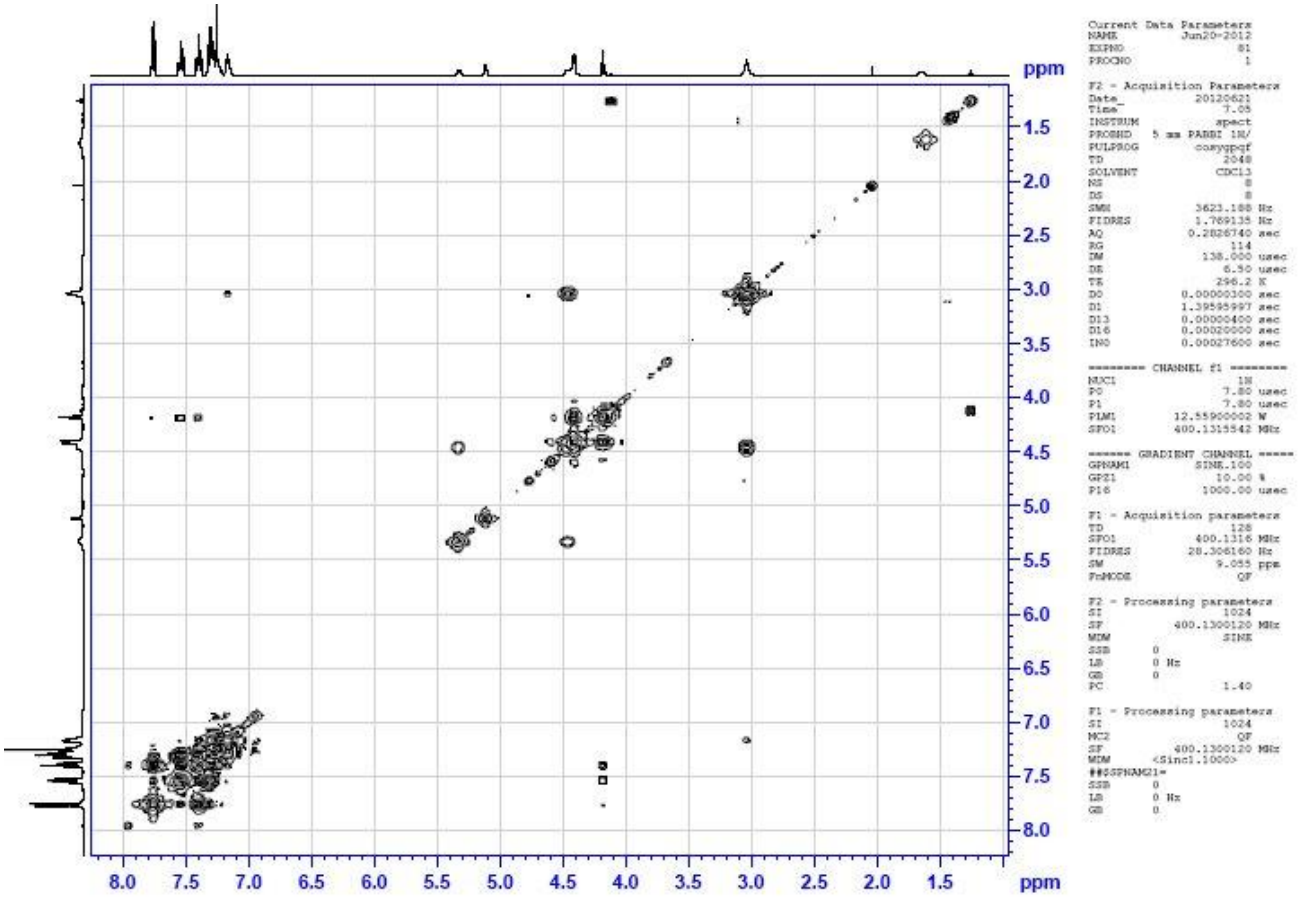
Fmoc-phenylalanine amino alcohol

LCMS , ^{1}H NMR

1H Fmoc-phenylalanine diazomethylketone in CDCl3 300MHz proton spectrum

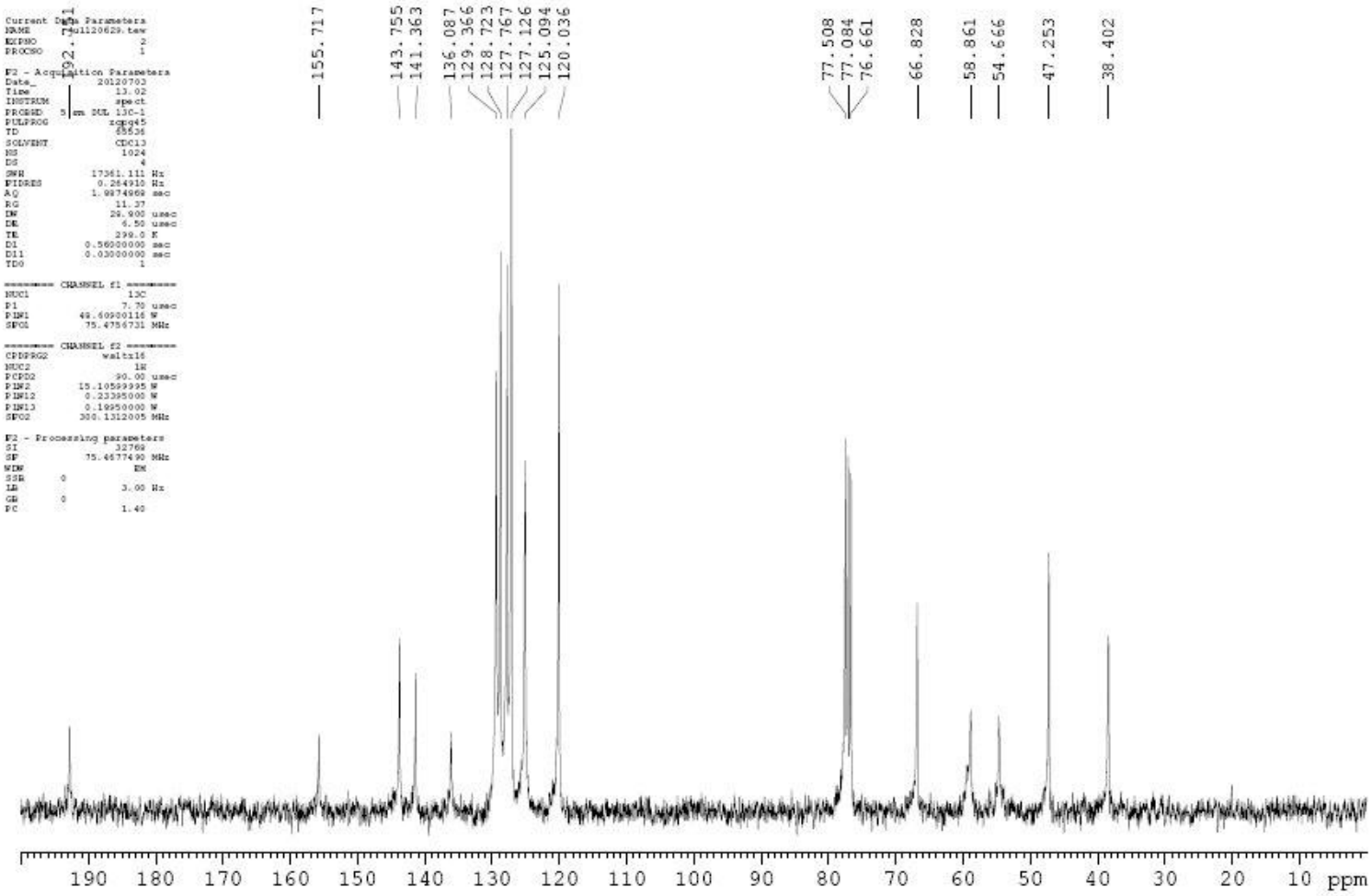


Fmoc-phenylalanine diazomethylketone COSY spectrum



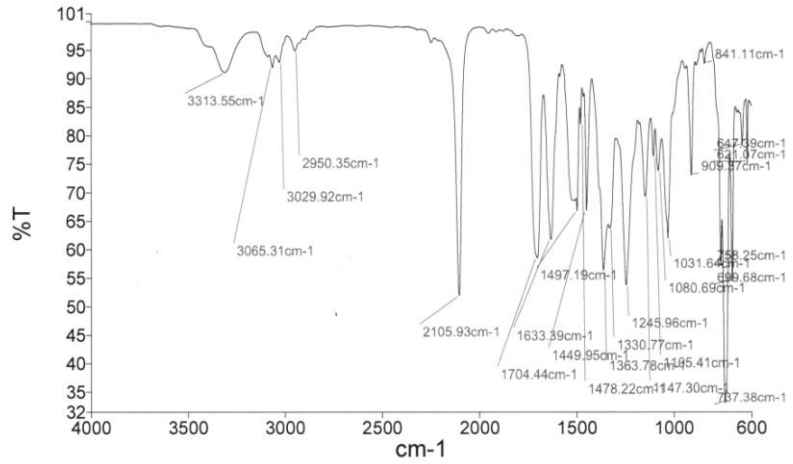

```
Current Data Parameters
NAME                 Jun20-2012
EXPNO                        81
PROCNO                        1

F2 - Acquisition Parameters
Date_              20120621
Time               7.05
INSTRUM                 spect
PROBHD    5 mm PABBI 1H/
PULPROG              conyypgf
TD                       2048
SOLVENT                  CDCl3
NS                           8
DS                           8
SWH                3623.188 Hz
FIDRES             1.769135 Hz
AQ                 0.2826740 sec
RG                         114
DW                 138.000 usec
DE                   6.50 usec
TE                   296.2 K
D0                 0.00000300 sec
D1                 1.39595997 sec
D13                0.00000400 sec
D16                0.00020000 sec
IN0                0.00027600 sec

======== CHANNEL f1 ========
NUC1                       1H
P0                   7.80 usec
P1                   7.80 usec
PLW1              12.55900002 W
SFO1             400.1315542 MHz

====== GRADIENT CHANNEL =====
GPNAM1                SINE.100
GPZ1                 10.00 %
P16                1000.00 usec

F1 - Acquisition parameters
TD                         128
SFO1             400.1316 MHz
FIDRES            28.306160 Hz
SW                  9.055 ppm
FnMODE                     QF

F2 - Processing parameters
SI                        1024
SF               400.1300120 MHz
WDW                       SINE
SSB          0
LB           0 Hz
GB           0
PC                        1.40

F1 - Processing parameters
SI                        1024
MC2                        QF
SF               400.1300120 MHz
WDW          <Sinc1.1000>
##$SPNAM21=
SSB          0
LB           0 Hz
GB           0
```

13C Fmoc-Phenylalanine diazomethyl ketone in CDCl3 75.48 MHz carbon spectrum



Fmoc-phenylalanine diazomethylketone FTIR

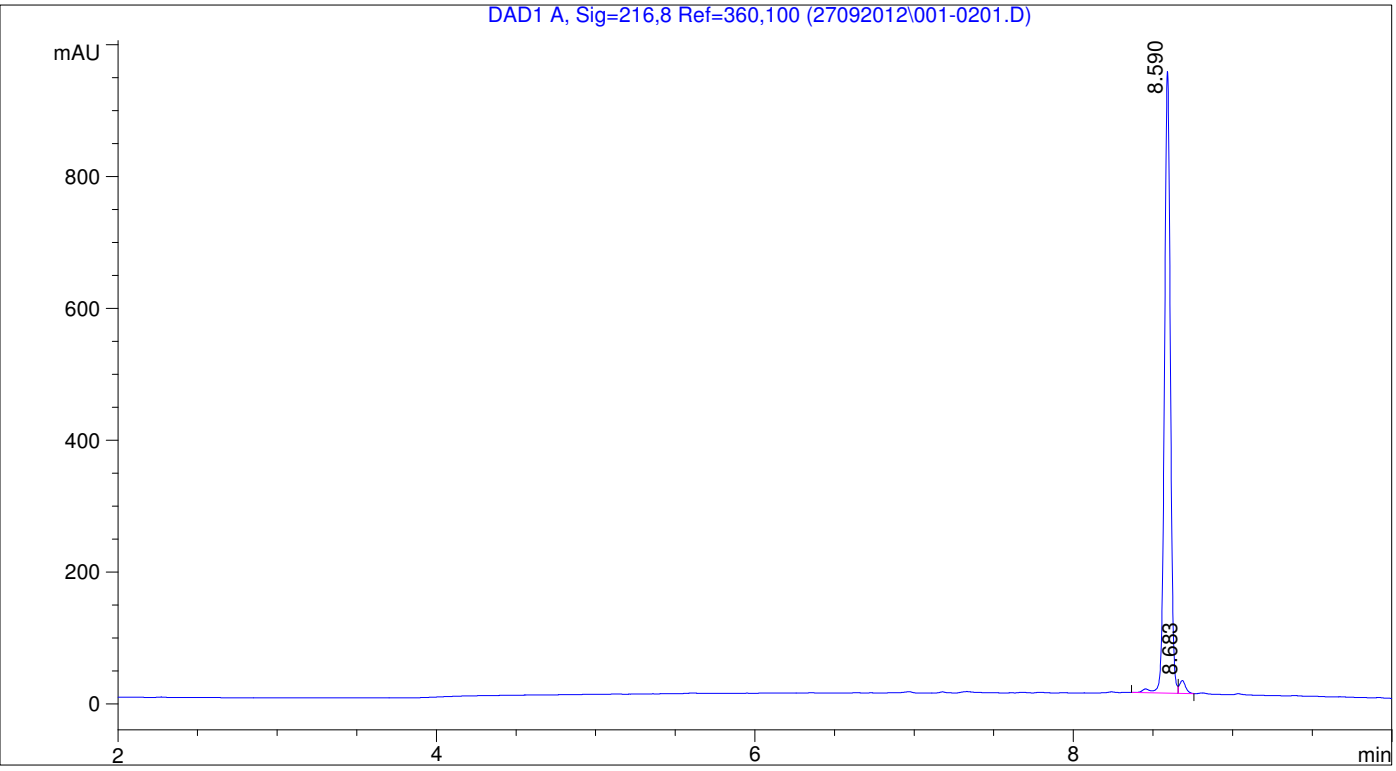
Analyst user
Date 20 June 2012 13:46



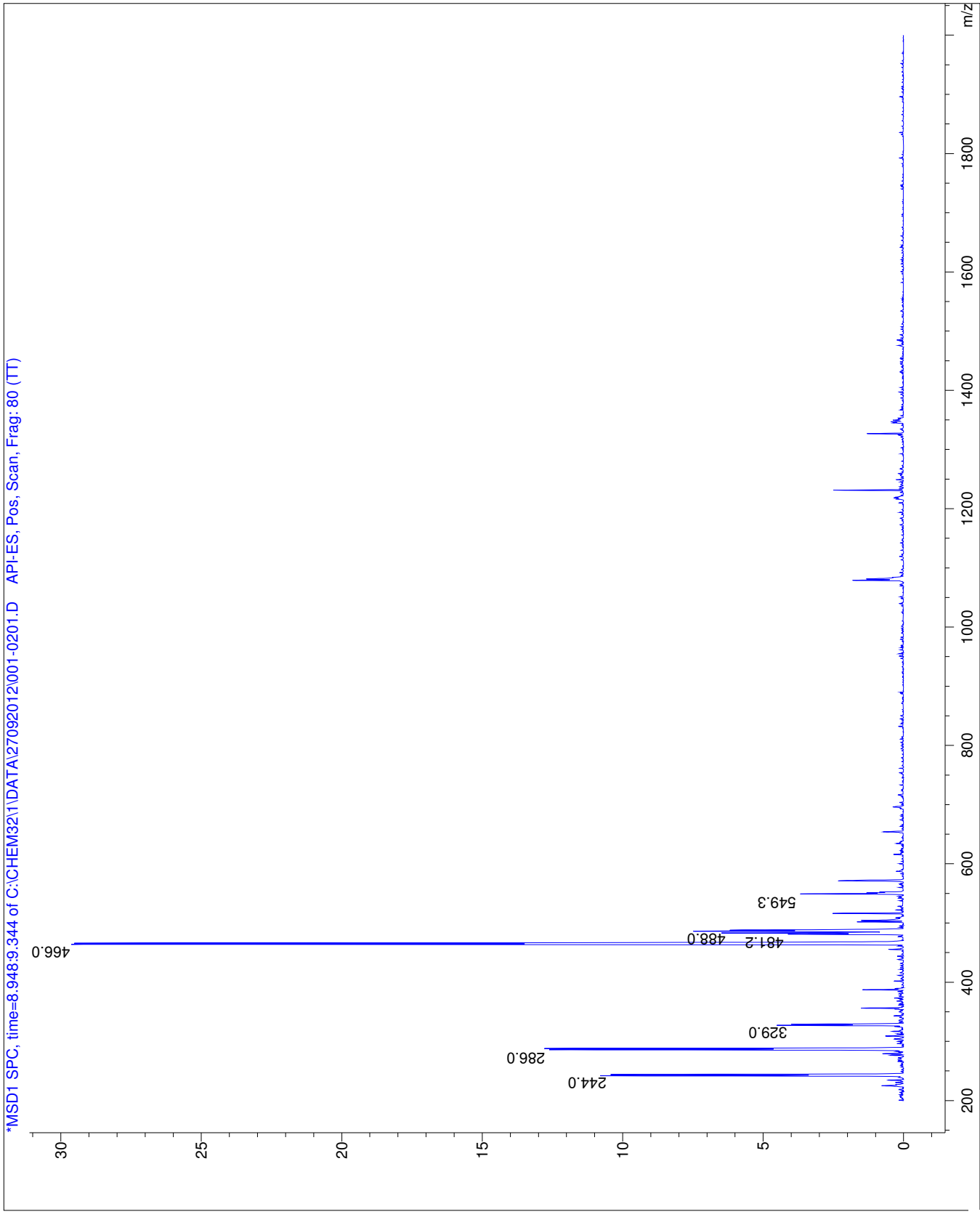
Sample Name	Description	Quality Checks
AA-Fmoc Phen Diazo Methyl Ketone	Fmoc Phen Diazo Methyl Ketone	The Quality Checks do not report any warnings for the sample.

Sample No:	Fmoc-Phe-brom
File:	C:\CHEM32\1\DATA\27092012\001-0201.D
Operator:	Phil

Column:	Kinetex 2.6u C18 100A
Method:	C:\CHEM32\1\METHODS\METHODS\AMINO ACID ANALYSIS.M
Instrument:	Instrument 1
Flow Rate:	1.5mls/min
Injection Volume:	60ul
Method Info:	Analysis carried out using a 100A 4.6 x 50mm column, gradient from 2% – 80% Acetonitrile, in 8 minutes. At 60°C.



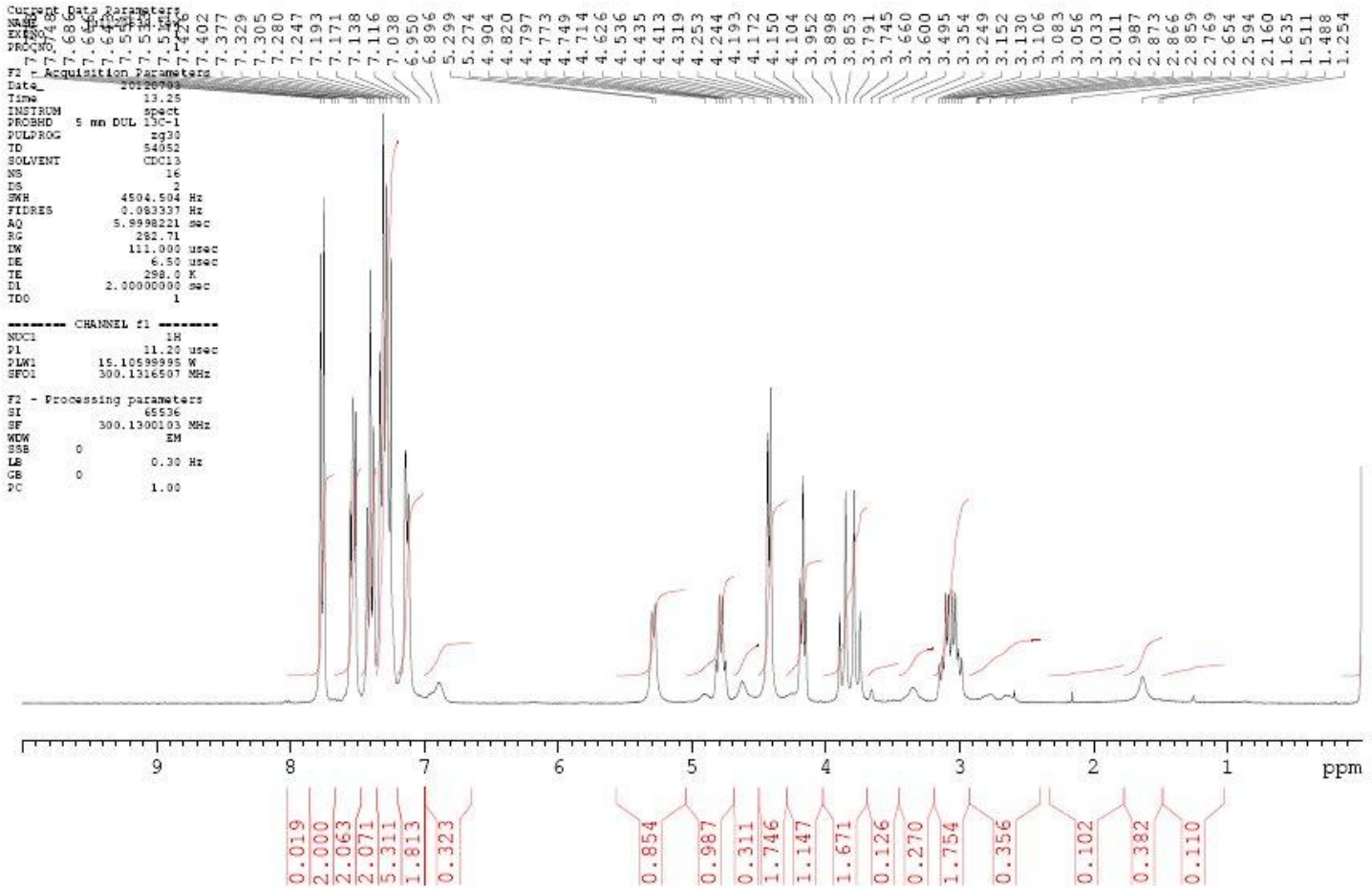
#	Meas. Ret.	Area	Area %
1	8.590	2432.007	97.913
2	8.683	51.831	2.087



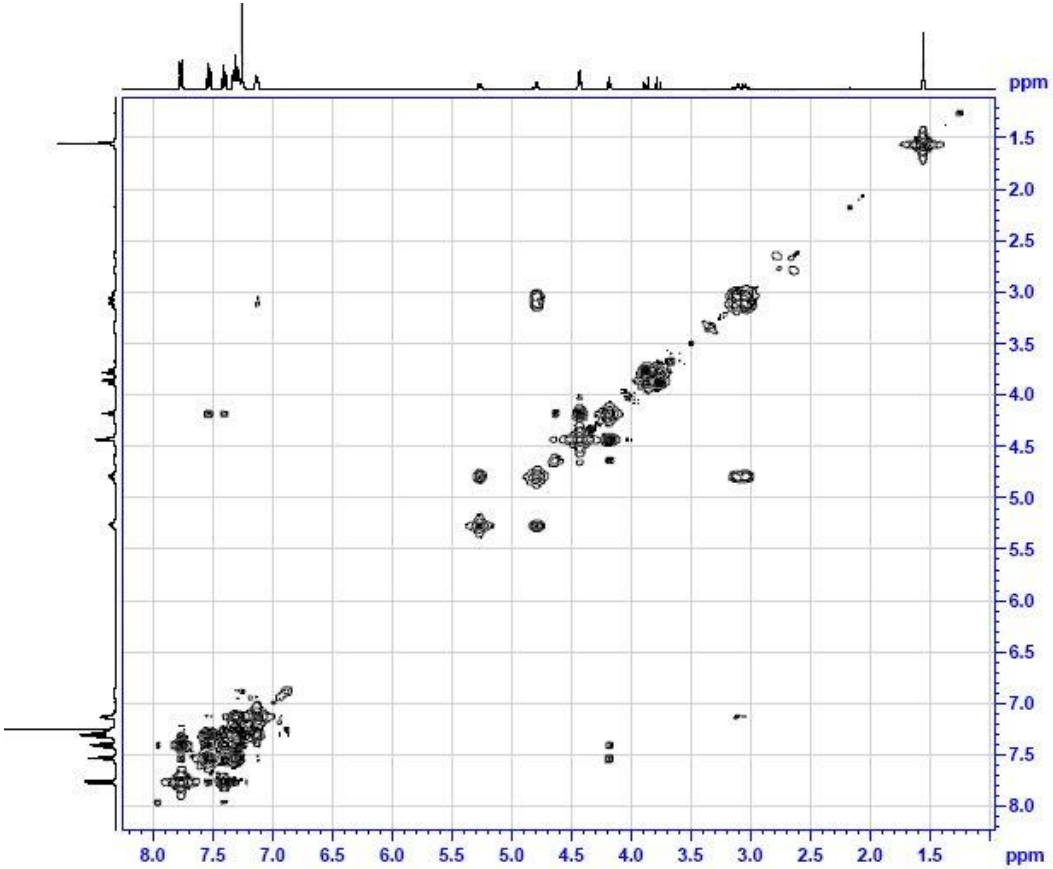
File: C:\CHEM32\1\DATA\27092012\001-0201.D

Sample No: Fmoc-Phe-brom

1H Fmoc-Phenylalanine bromomethyl ketone in CDCl3 300MHz proton spectrum

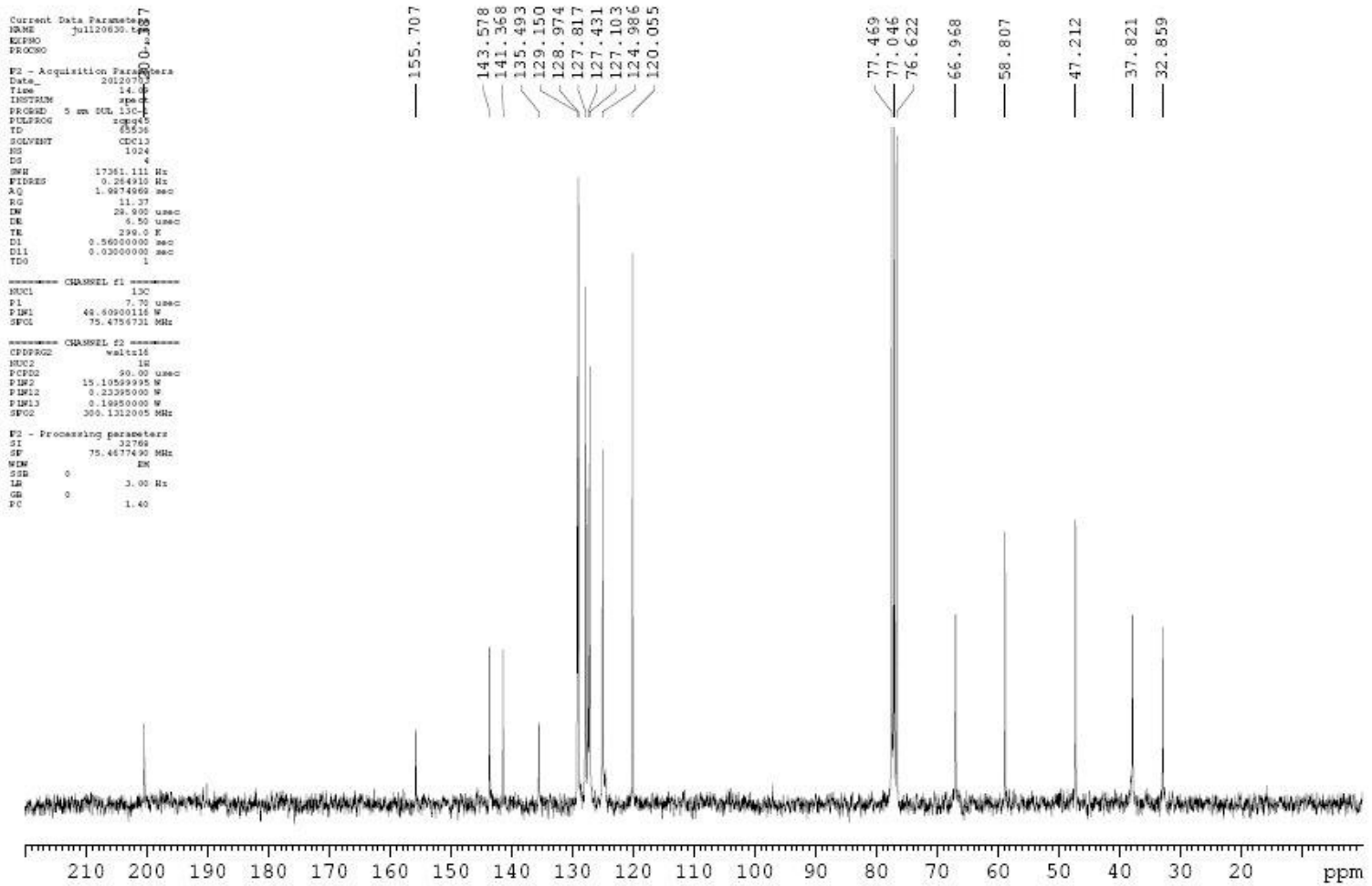


Fmoc-phenylalanine bromomethylketone COSY spectrum




```
Current Data Parameters
NAME                  Jun20-2012
EXPNO                         71
PROCNO                         1

F2 - Acquisition Parameters
Date_              20120621
Time               5.04
INSTRUM            spect
PROBHD   5 mm PABBI 1H/
PULPROG            cosygpqf
TD                 2048
SOLVENT            CDCl3
NS                 8
DS                 8
SWH                3571.428 Hz
FIDRES             1.743862 Hz
AQ                 0.2867700 sec
RG                 114
DW                 140.000 usec
DE                 6.50 usec
TE                 296.2 K
D0                 0.00000300 sec
D1                 1.39186394 sec
D13                0.00004400 sec
D16                0.00020000 sec
IN0                0.00028000 sec

======== CHANNEL f1 ========
NUC1                    1H
P0                 7.80 usec
P1                 7.80 usec
PLW1          12.55900002 W
SFO1          400.1315695 MHz

====== GRADIENT CHANNEL =====
GPNAM1           SINE.100
GPZ1                10.00 %
P16              1000.00 usec

F1 - Acquisition parameters
TD                      128
SFO1             400.1316 MHz
FIDRES          27.901785 Hz
SW                 8.926 ppm
FnMODE                  QF

F2 - Processing parameters
SI                      1024
SF            400.1300112 MHz
WDW                     SINE
SSB          0
LB           0 Hz
GB           0
PC                      1.40

F1 - Processing parameters
SI                      1024
MC2                     QF
SF            400.1300112 MHz
WDW       <Sinc1.1000>
##$SPNAM21=
SSB          0
LB           0 Hz
GB           0
```

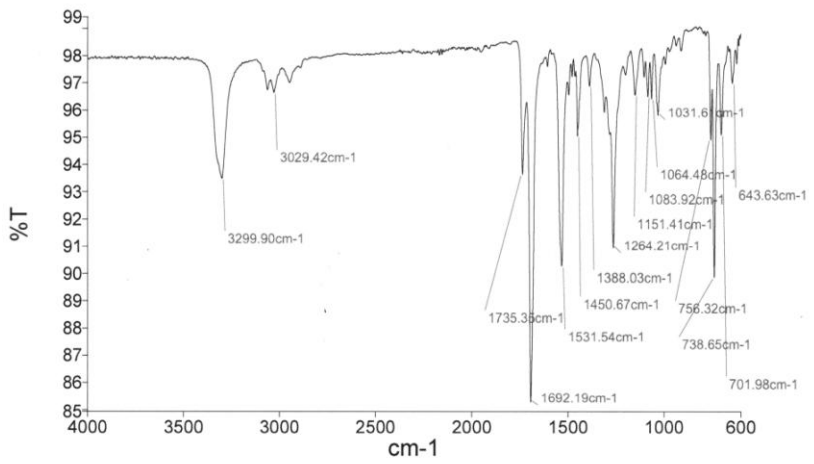
13C Fmoc-Phenylalanine bromomethyl ketone in CDCl3 75.48 MHz carbon spectrum




```
Current Data Parameters
NAME       jul120630.337
EXPNO                337
PROCNO               337

F2 - Acquisition Parameters
Date_          20120703
Time               14.08
INSTRUM             spect
PROBHD   5 mm DUL 13C-H
PULPROG          zgpg45
TD                65536
SOLVENT           CDCl3
NS                 1024
DS                    4
SWH           17361.111 Hz
FIDRES        0.264910 Hz
AQ            1.9874969 sec
RG                11.37
DW               28.800 usec
DE                6.50 usec
TE              299.0 K
D1            0.56000000 sec
D11           0.03000000 sec
TD0                   1

======== CHANNEL f1 ========
NUC1                 13C
P1                  7.70 usec
PL1           48.60900116 W
SFO1          75.4756731 MHz

======== CHANNEL f2 ========
CPDPRG2          waltz16
NUC2                  1H
PCPD2              90.00 usec
PL2           15.10599995 W
PL12           0.23395000 W
PL13           0.19950000 W
SFO2          300.1312005 MHz

F2 - Processing parameters
SI                  32768
SF            75.4677490 MHz
WDW                    EM
SSB       0
LB                 3.00 Hz
GB        0
PC                 1.40
```

Fmoc-phenylalanine bromomethylketone FTIR

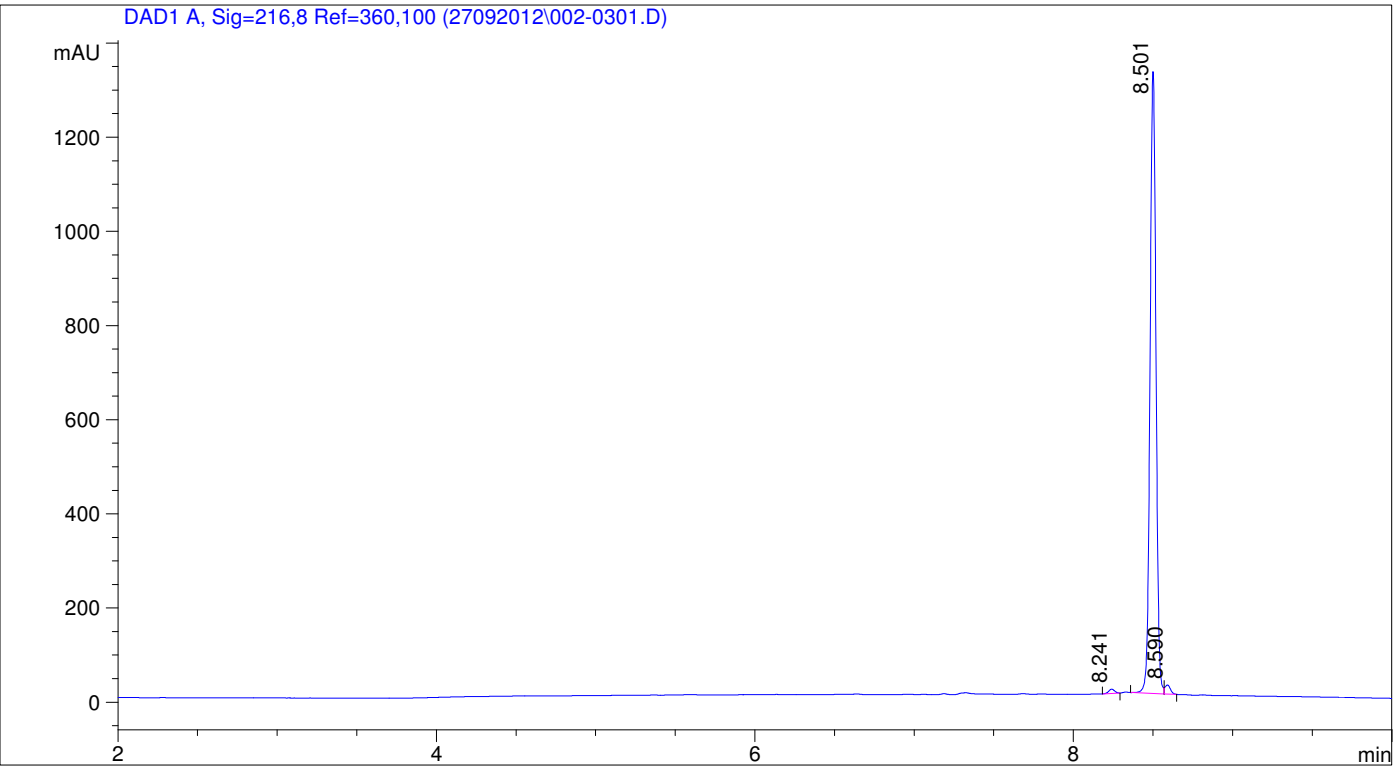
Analyst user
Date 20 June 2012 13:29



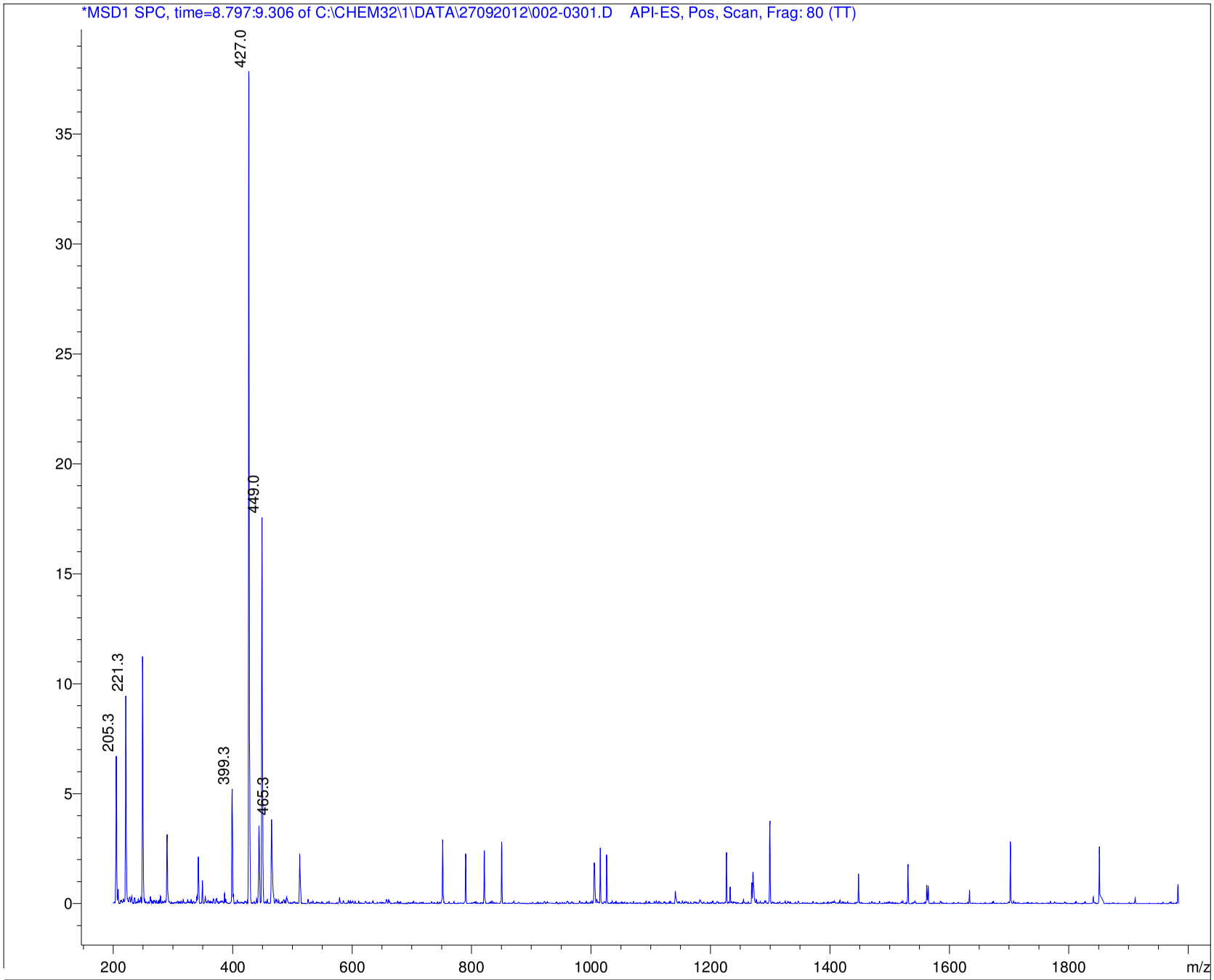
Sample Name	Description	Quality Checks
AA-Fmoc Phen Bromo Methyl Ketone	Fmoc Phen Bromo Methyl Ketone	The Quality Checks give rise to a Weak Bands warning for the sample.

Sample No: fmoc-phe-azido
File: C:\CHEM32\1\DATA\27092012\002-0301.D
Operator: Phil

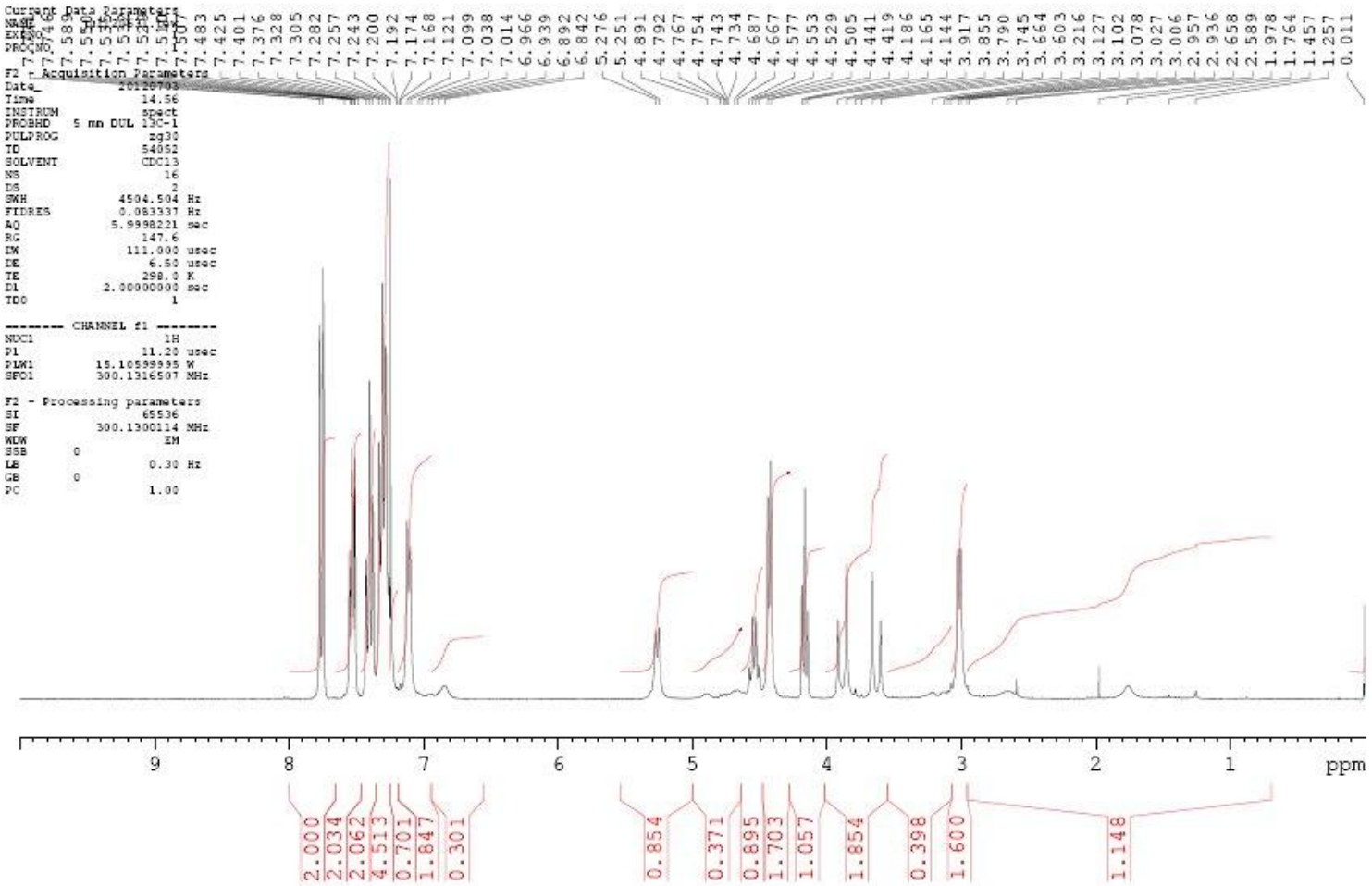
Column: Kinetex 2.6u C18 100A
Method: C:\CHEM32\1\METHODS\METHODS\AMINO ACID ANALYSIS.M
Instrument: Instrument 1
Flow Rate: 1.5mls/min
Injection Volume: 60ul
Method Info: Analysis carried out using a 100A 4.6 x 50mm column, gradient from 2% - 80% Acetonitrile, in 8 minutes. At 60°C.



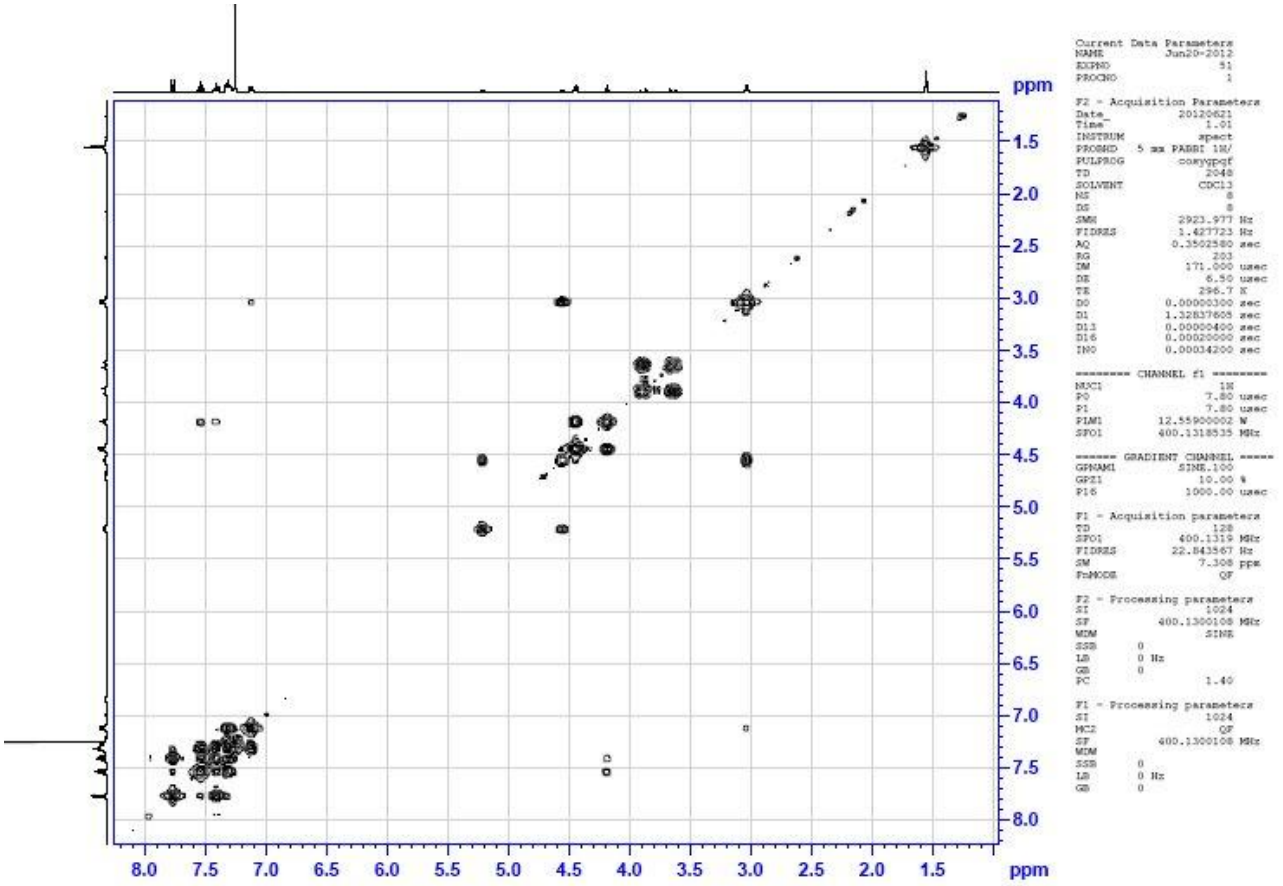
#	Meas. Ret.	Area	Area %
1	8.241	23.501	0.672
2	8.501	3428.647	98.038
3	8.590	45.129	1.290


File: C:\CHEM32\1\DATA\27092012\002-0301.D
Sample No: fmoc-phe-azido
*MSD1 SPC, time=8.797:9.306 of C:\CHEM32\1\DATA\27092012\002-0301.D API-ES, Pos, Scan, Frag: 80 (TT)
427.0
449.0
221.3
205.3
399.3
465.3
0
5
10
15
20
25
30
35
200
400
600
800
1000
1200
1400
1600
1800
m/z

1H Fmoc-Phenylalanine azidomethyl ketone in CDCl3 300MHz proton spectrum

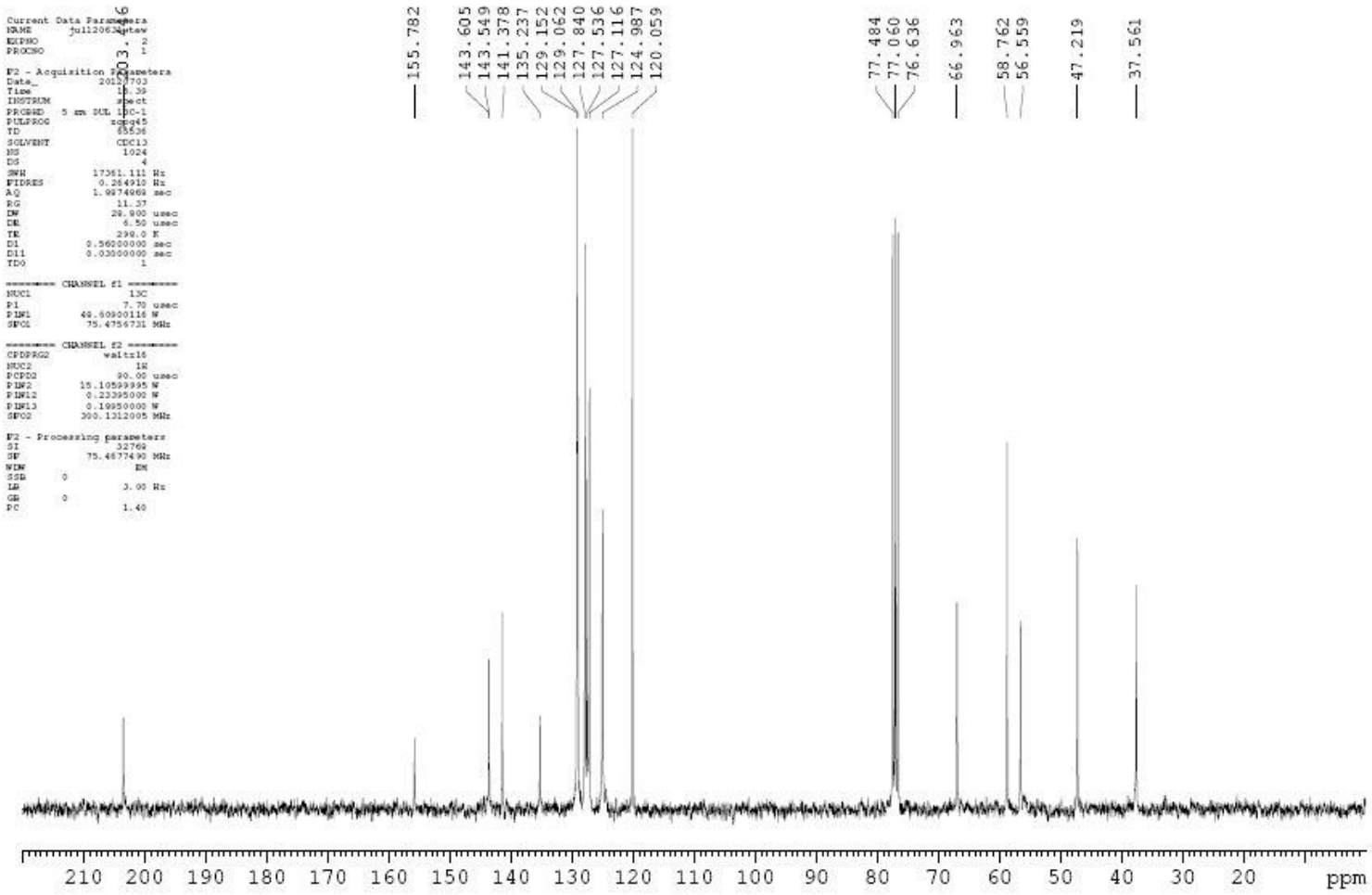


Fmoc-phenylalanine azidomethylketone COSY spectrum

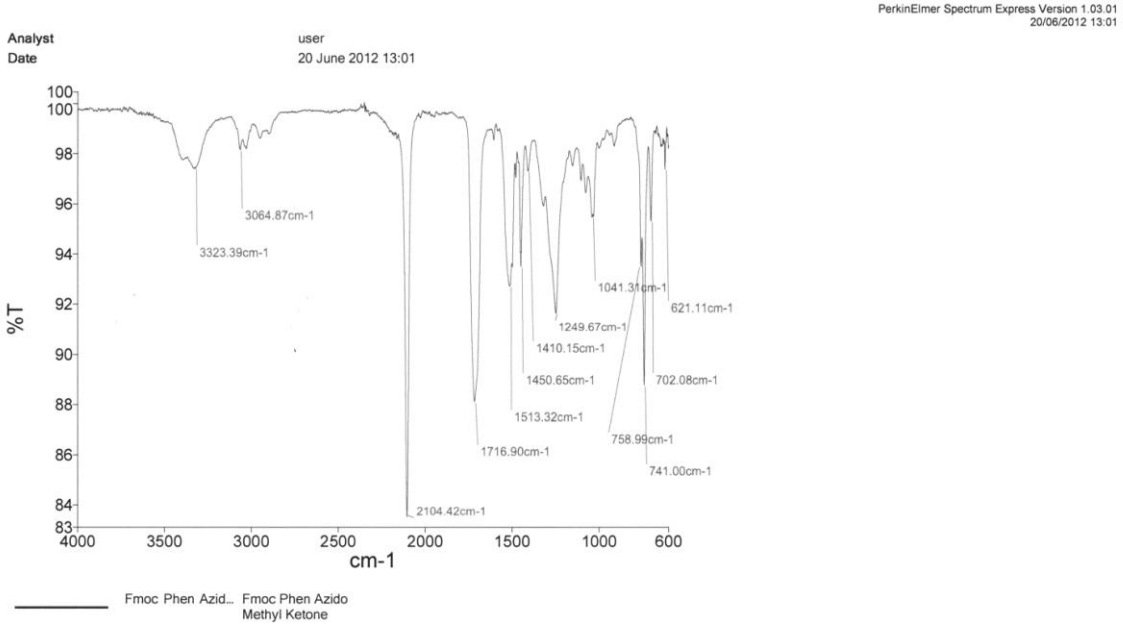



```
Current Data Parameters
NAME                 Jun20-2012
EXPNO                        51
PROCNO                        1

F2 - Acquisition Parameters
Date_               20120621
Time                    1.01
INSTRUM                 spect
PROBHD   5 mm PABBI 1H/
PULPROG              cosygpqf
TD                       2048
SOLVENT                  CDCl3
NS                           8
DS                           8
SWH                  2923.977 Hz
FIDRES               1.427723 Hz
AQ                   0.3502580 sec
RG                         203
DW                   171.000 usec
DE                     6.50 usec
TE                     296.7 K
D0                   0.00000300 sec
D1                   1.32837608 sec
D13                  0.00000400 sec
D16                  0.00020000 sec
IN0                  0.00034200 sec

======== CHANNEL f1 ========
NUC1                       1H
P0                       7.80 usec
P1                       7.80 usec
PLW1                12.55900002 W
SFO1               400.1318535 MHz

====== GRADIENT CHANNEL =====
GPNAM1                SINE.100
GPZ1                     10.00 %
P16                    1000.00 usec

F1 - Acquisition parameters
TD                         128
SFO1               400.1319 MHz
FIDRES              22.843567 Hz
SW                       7.308 ppm
FnMODE                     QF

F2 - Processing parameters
SI                         1024
SF                 400.1300108 MHz
WDW                        SINE
SSB                 0
LB                  0 Hz
GB                  0
PC                         1.40

F1 - Processing parameters
SI                         1024
MC2                        QF
SF                 400.1300108 MHz
WDW
SSB                 0
LB                  0 Hz
GB                  0
```

13C Fmoc-Phenylalanine azidomethyl ketone in CDCl3 75.48 MHz carbon spectrum

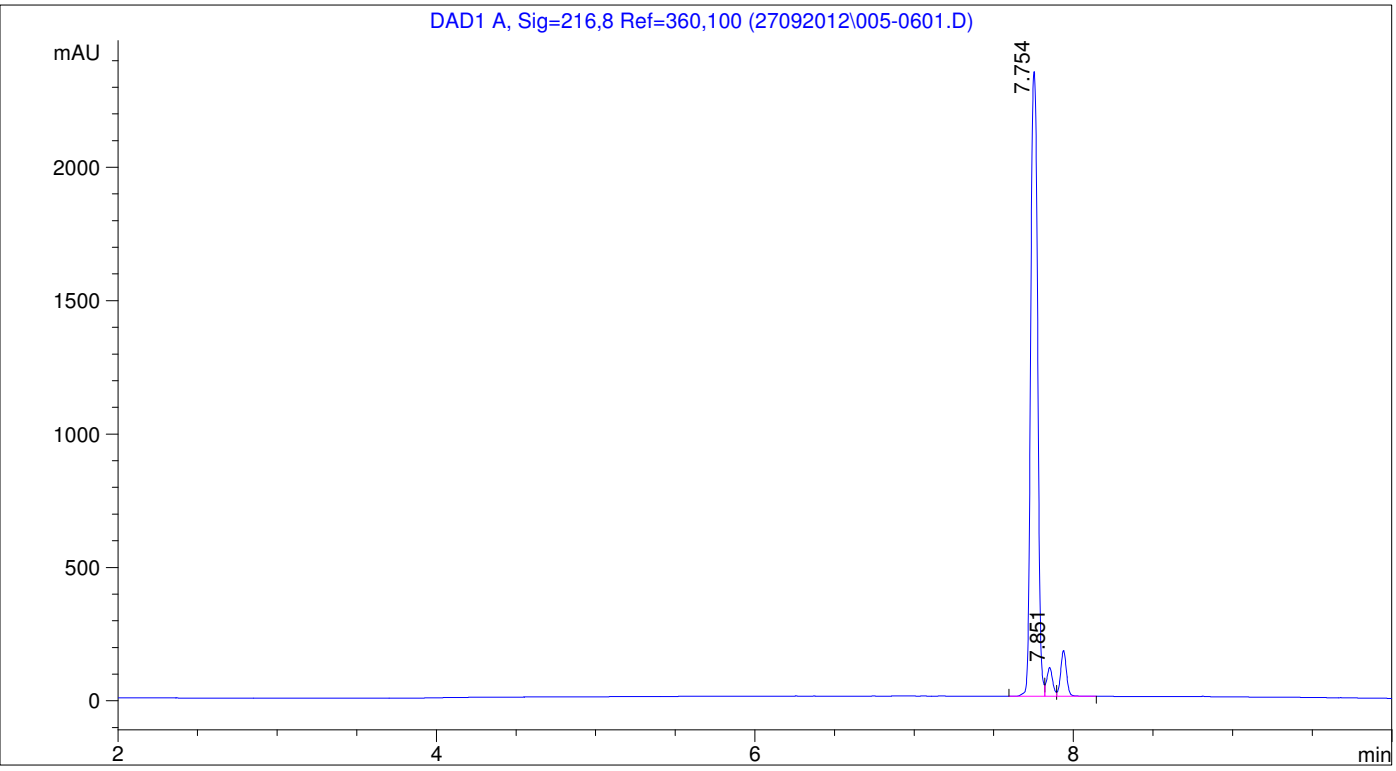


Fmoc-phenylalanine azidomethylketone FTIR

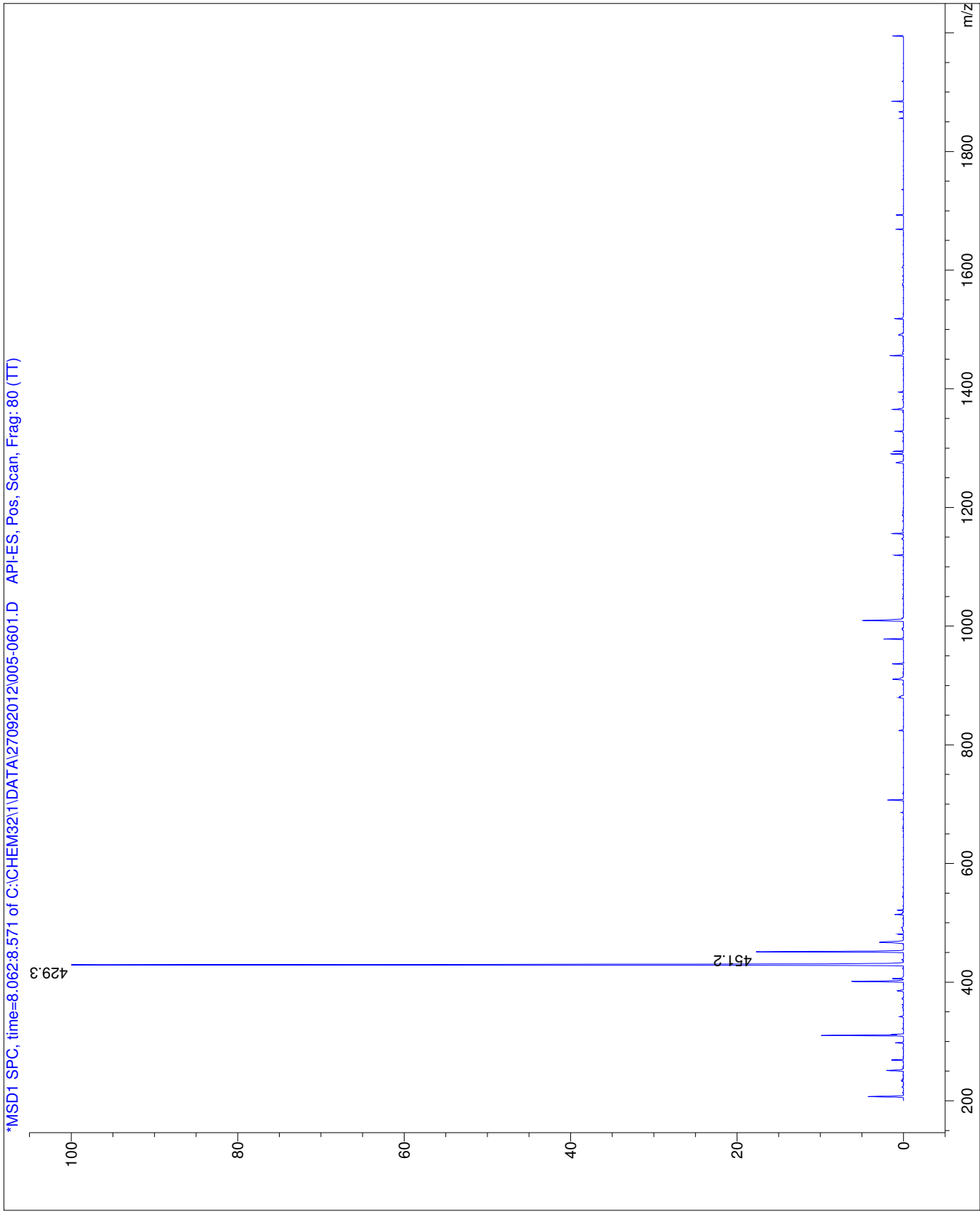


Sample No: FMOC-Phe-azido OH
File: C:\CHEM32\1\DATA\27092012\005-0601.D
Operator: Phil

Column: Kinetex 2.6u C18 100A
Method: C:\CHEM32\1\METHODS\METHODS\AMINO ACID ANALYSIS.M
Instrument: Instrument 1
Flow Rate: 1.5mls/min
Injection Vc : 60ul
Method Info: Analysis carried out using a 100A 4.6 x 50mm column, gradient from 2% - 80% Acetonitrile, in 8 minutes. At 60°C.

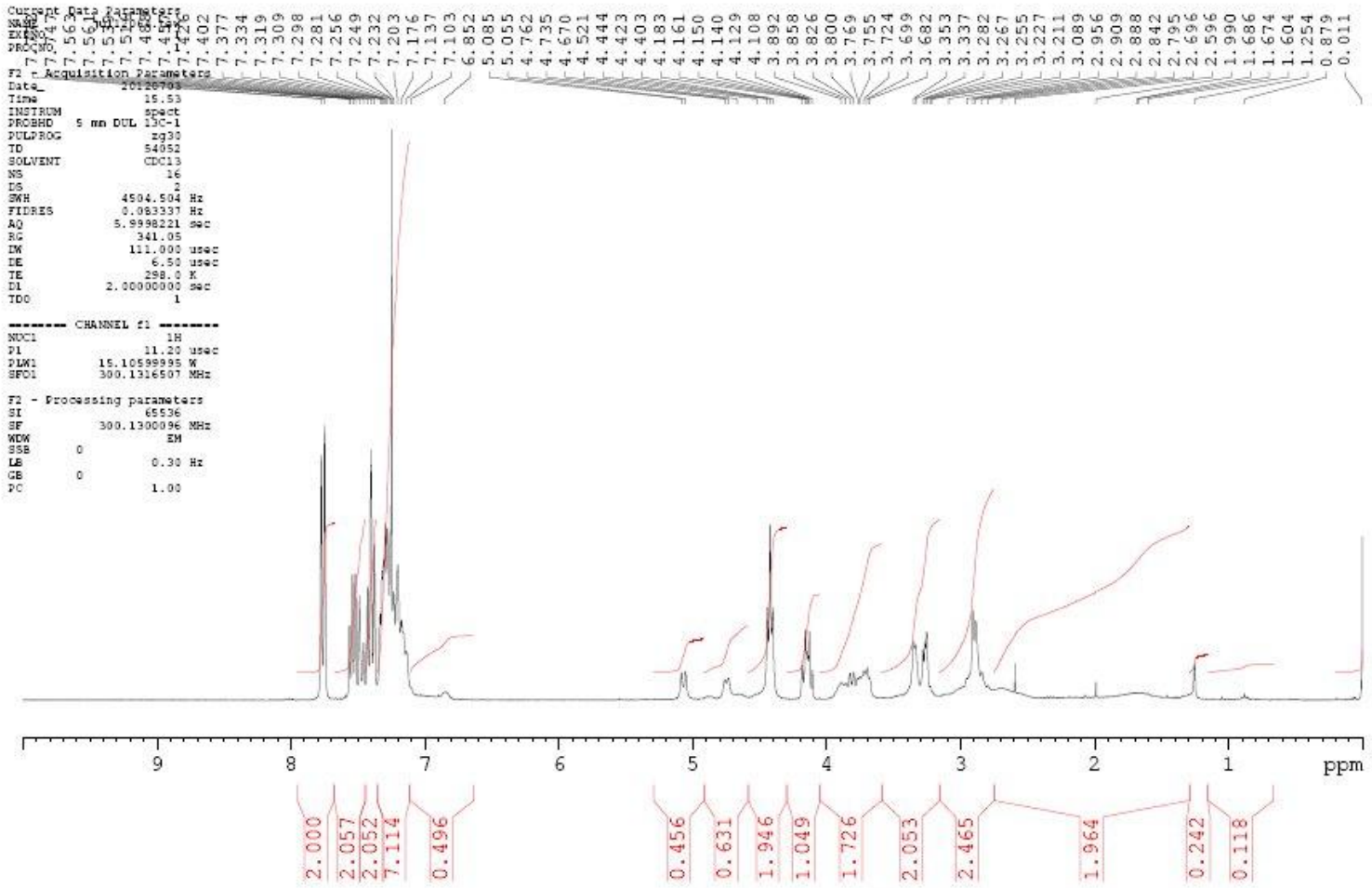


#	Meas. Ret.	Area	Area %
1	7.754	7070.183	90.742
2	7.851	280.175	3.596
3	7.939	441.131	5.662

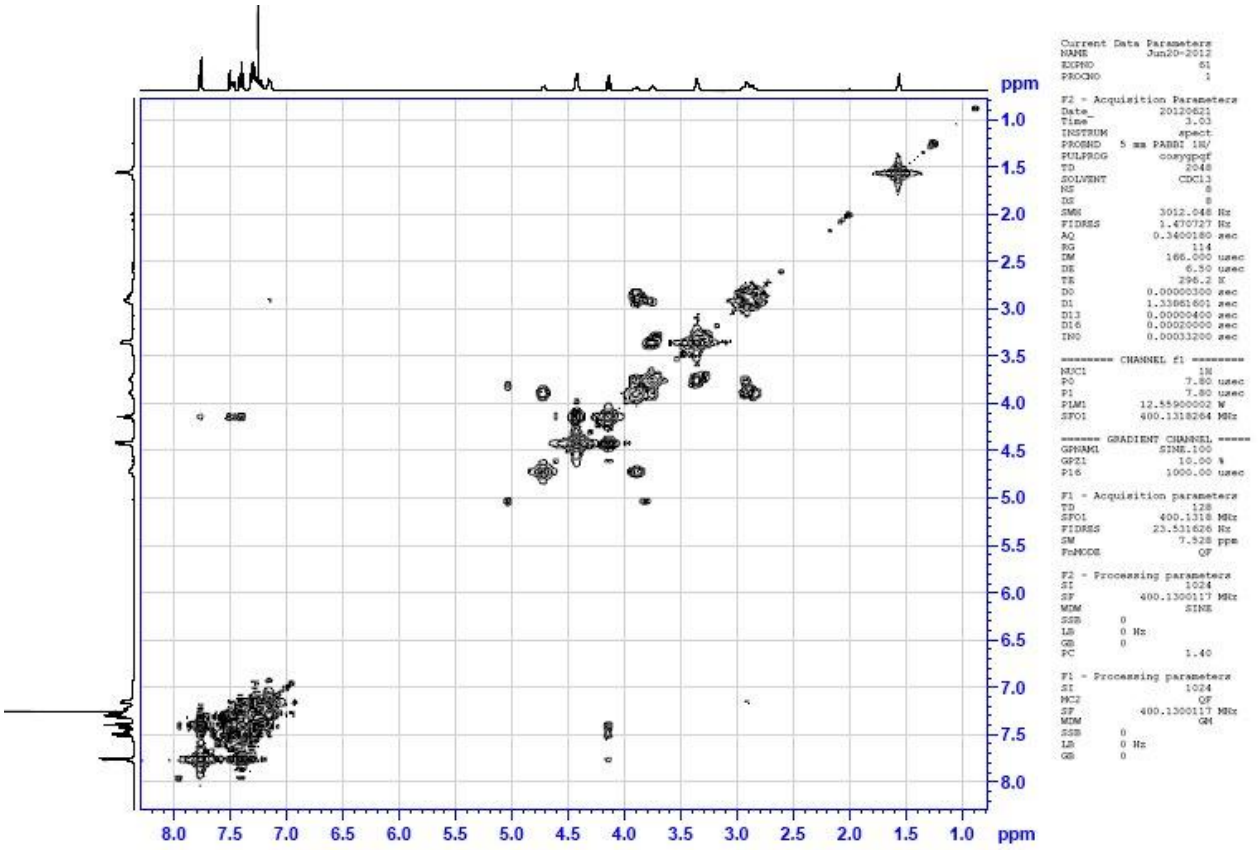


File: C:\CHEM32\1\DATA\27092012\005-0601.D
Sample No: FMOC-F-azido OH

1H Fmoc-Phenylalanine azidomethyl alcohol in CDCl3 300MHz proton spectrum

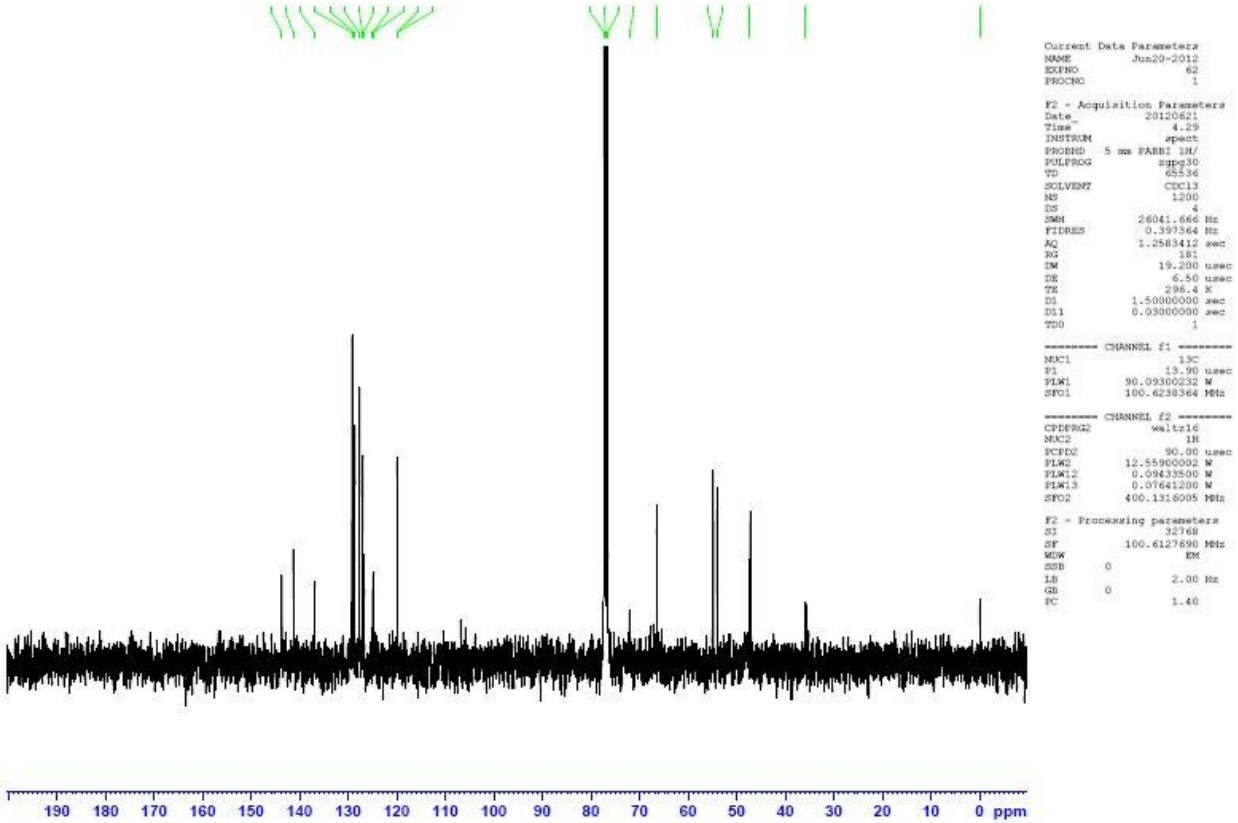


Fmoc-phenylalanine azidomethylalcohol COSY spectrum




```
Current Data Parameters
NAME          Jun20-2012
EXPNO                 61
PROCNO                 1

F2 - Acquisition Parameters
Date_          20120621
Time              1.03
INSTRUM           spect
PROBHD   5 mm PABBI 1H/
PULPROG       cosygpqf
TD                2048
SOLVENT           CDCl3
NS                   8
DS                   8
SWH           3012.048 Hz
FIDRES        1.470727 Hz
AQ            0.3400180 sec
RG                 114
DW              166.000 usec
DE                6.50 usec
TE              296.2 K
D0           0.00000300 sec
D1           1.33861601 sec
D13          0.00000400 sec
D16          0.00020000 sec
IN0          0.00033200 sec

======== CHANNEL f1 ========
NUC1               1H
P0                7.80 usec
P1                7.80 usec
PLW1         12.55900002 W
SFO1        400.1318264 MHz

====== GRADIENT CHANNEL =====
GPNAM1          SINE.100
GPZ1               10.00 %
P16              1000.00 usec

F1 - Acquisition parameters
TD                 128
SFO1        400.1318 MHz
FIDRES       23.531626 Hz
SW            7.528 ppm
FnMODE             QF

F2 - Processing parameters
SI                1024
SF          400.1300117 MHz
WDW               SINE
SSB       0
LB        0 Hz
GB        0
PC                1.40

F1 - Processing parameters
SI                1024
MC2                QF
SF          400.1300117 MHz
WDW                GM
SSB       0
LB        0 Hz
GB        0
```

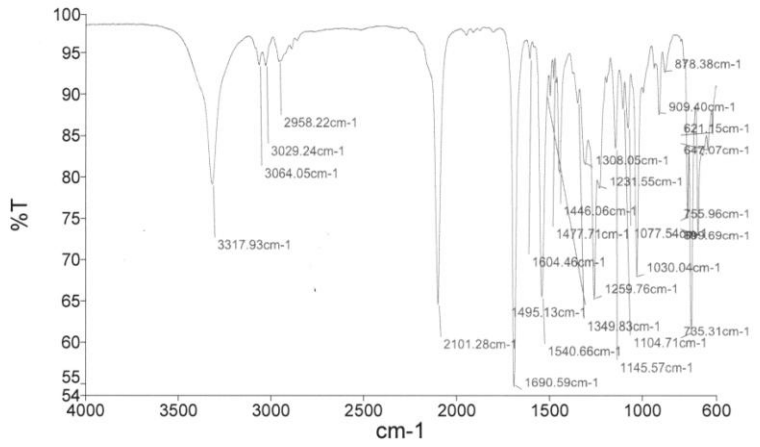
13C Fmoc-phenylalanine azidomethylalcohol




```
Current Data Parameters
NAME                  Jun20-2012
EXPNO                         62
PROCNO                         1

F2 - Acquisition Parameters
Date_                   20120621
Time                        4.29
INSTRUM                    spect
PROBHD    5 mm PABBI 1H/
PULPROG                 zgpg30
TD                           65536
SOLVENT                   CDCl3
NS                            1200
DS                               4
SWH                   26041.666 Hz
FIDRES                 0.397364 Hz
AQ                     1.2583412 sec
RG                             181
DW                      19.200 usec
DE                       6.50 usec
TE                       296.4 K
D1                    1.50000000 sec
D11                   0.03000000 sec
TD0                            1

======== CHANNEL f1 ========
NUC1                         13C
P1                         13.90 usec
PLW1                90.09300232 W
SFO1               100.6238364 MHz

======== CHANNEL f2 ========
CPDPRG2                  waltz16
NUC2                         1H
PCPD2                      90.00 usec
PLW2                12.55900002 W
PLW12                0.09435500 W
PLW13                0.07641200 W
SFO2               400.1316005 MHz

F2 - Processing parameters
SI                           32768
SF                 100.6127690 MHz
WDW                            EM
SSB       0
LB                         2.00 Hz
GB        0
PC                         1.40
```

Fmoc-phenylalanine azidomethylalcohol FTIR

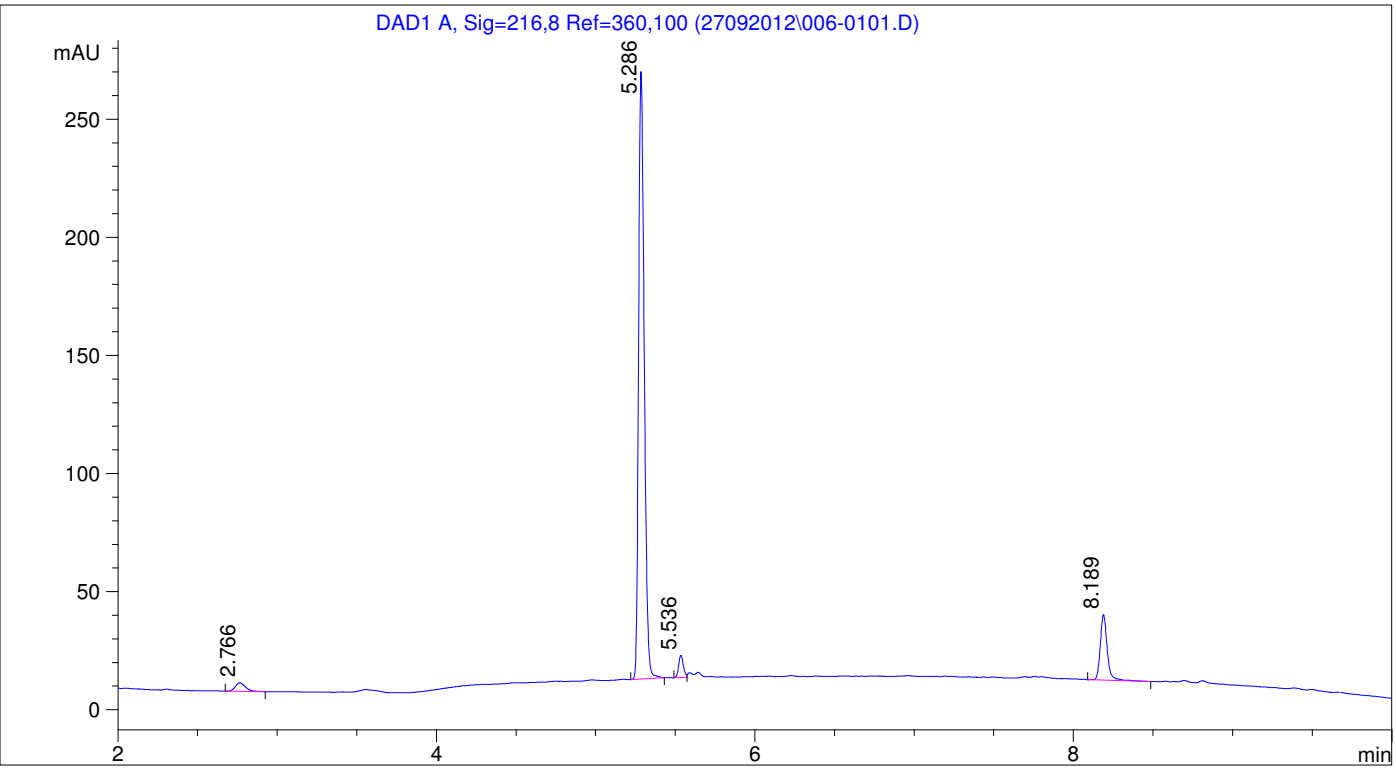
Analyst user
Date 20 June 2012 13:12



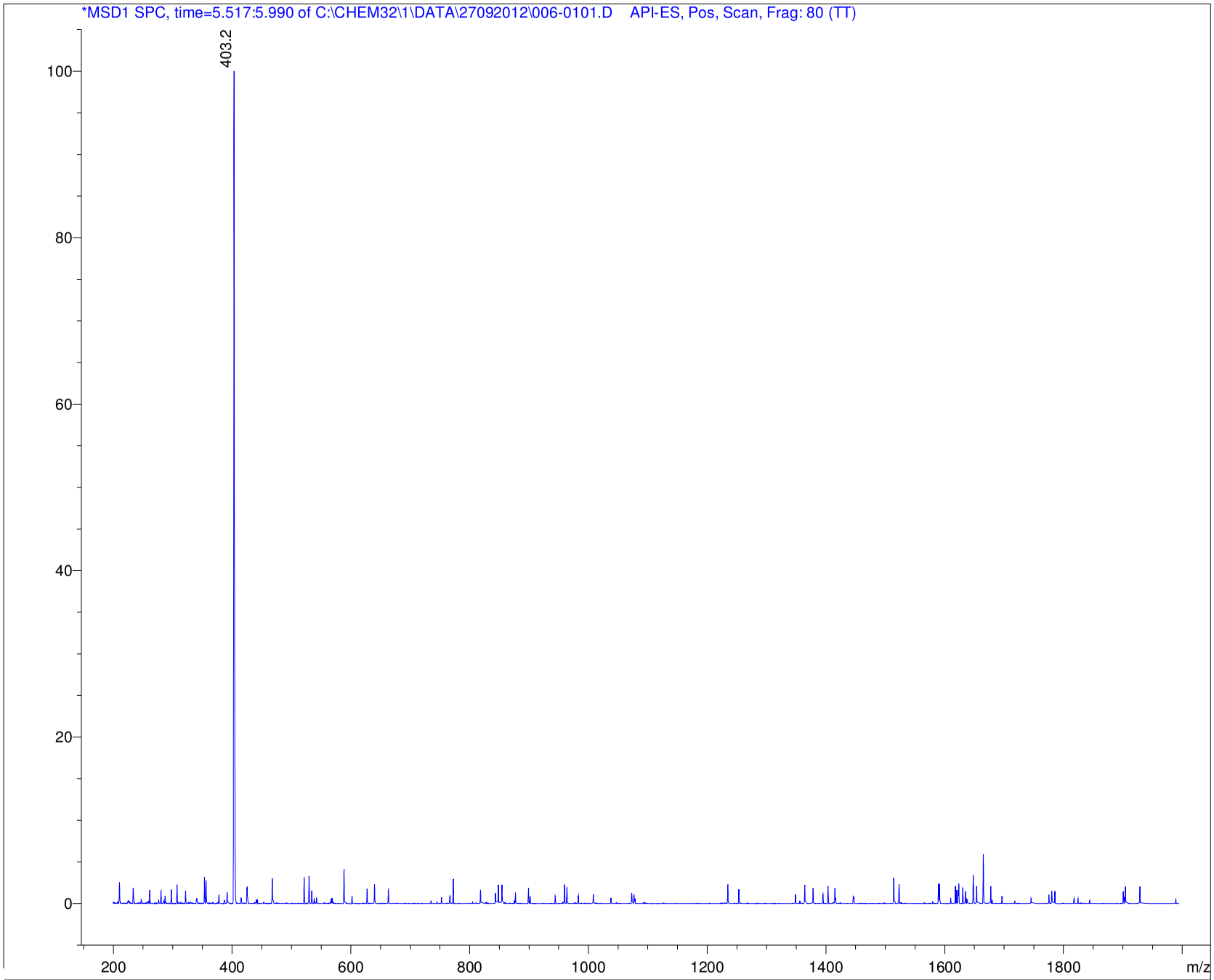
Sample Name	Description	Quality Checks
AA-Fmoc Phen Azido Methyl Alcohol_001	Fmoc Phen Azido Methyl Alcohol	The Quality Checks do not report any warnings for the sample.

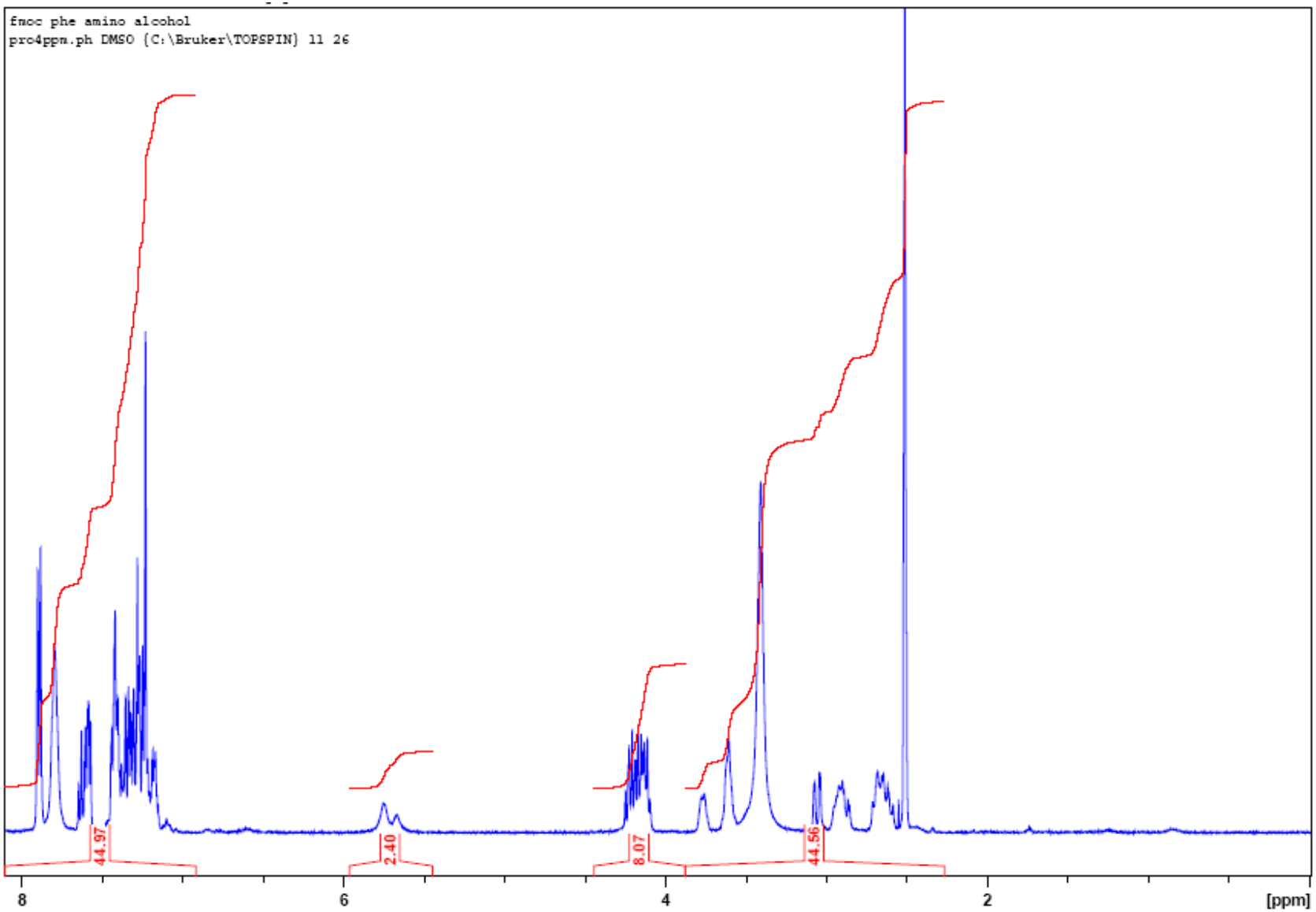
Sample No: Fmoc-F-Amino OH
File: C:\CHEM32\1\DATA\27092012\006-0101.D
Operator: Phil

Column: Kinetex 2.6u C18 100A
Method: C:\CHEM32\1\METHODS\METHODS\AMINO ACID ANALYSIS.M
Instrument: Instrument 1
Flow Rate: 1.5mls/min
Injection Vo` : 60ul
Method Info: Analysis carried out using a 100A 4.6 x 50mm column, gradient from 2% - 80% Acetonitrile, in 8 minutes. At 60°C.



#	Meas. Ret.	Area	Area %
1	2.766	16.040	2.127
2	5.286	633.975	84.064
3	5.536	19.639	2.604
4	8.189	84.504	11.205


*MSD1 SPC, time=5.517:5.990 of C:\CHEM32\1\DATA\27092012\006-0101.D API-ES, Pos, Scan, Frag: 80 (TT)
403.2
File: C:\CHEM32\1\DATA\27092012\006-0101.D
Sample No: Fmoc-F-Amino OH
100
80
60
40
20
0
200
400
600
800
1000
1200
1400
1600
1800
m/z


fmoc phe amino alcohol
pro4ppm.ph DMSO (C:\Bruker\TOPSPIN) 11 26
44.97
2.40
8.07
44.56
8
6
4
2
[ppm]

Appendix IIb

Analytical data for small peptide aldehyde

Z-Arg-Leu-Phe-amino alcohol

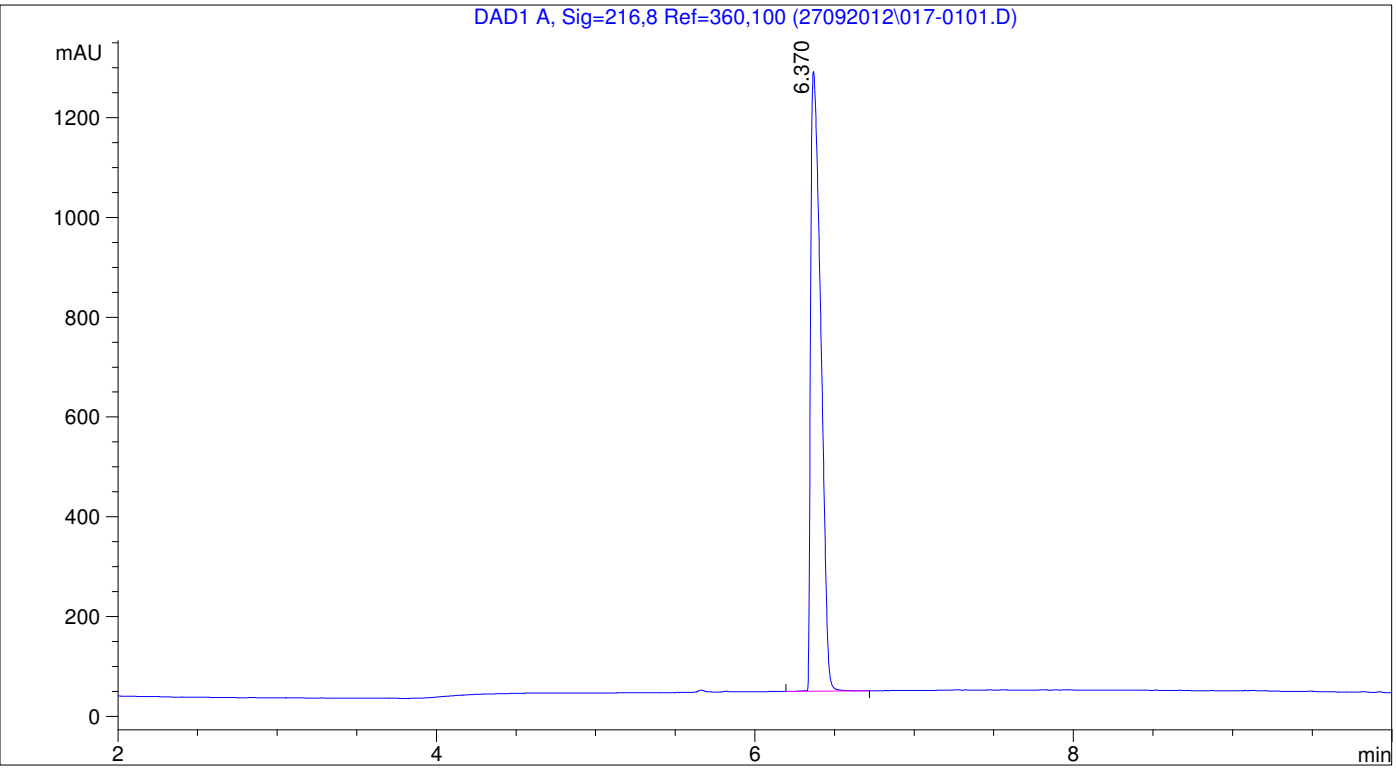
LCMS, ^{1}H NMR

Z-Arg-Leu-Phe-aldehyde

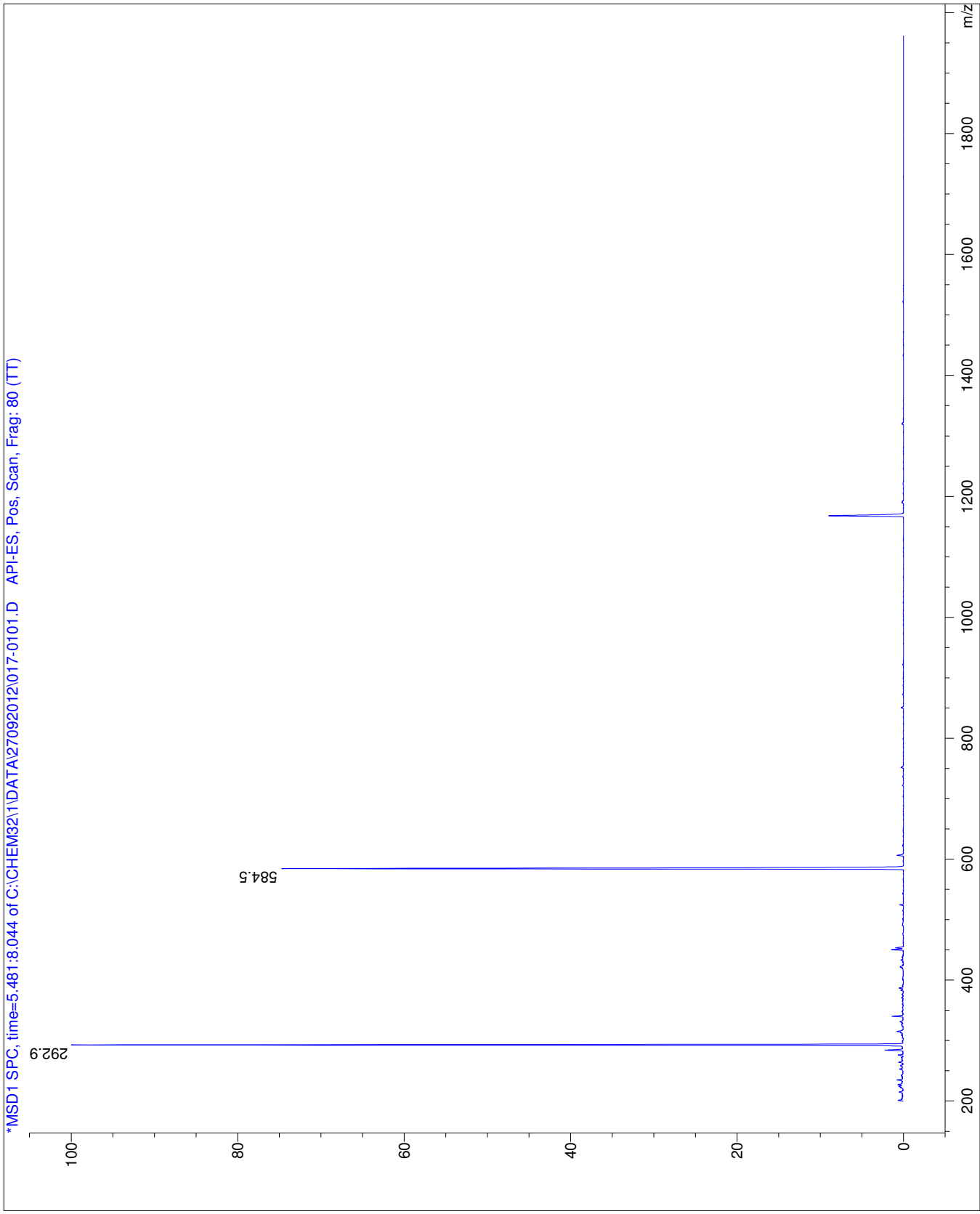
MS, ^{1}H NMR

Sample No: Z-RLF-Amino OH
File: C:\CHEM32\1\DATA\27092012\017-0101.D
Operator: Phil

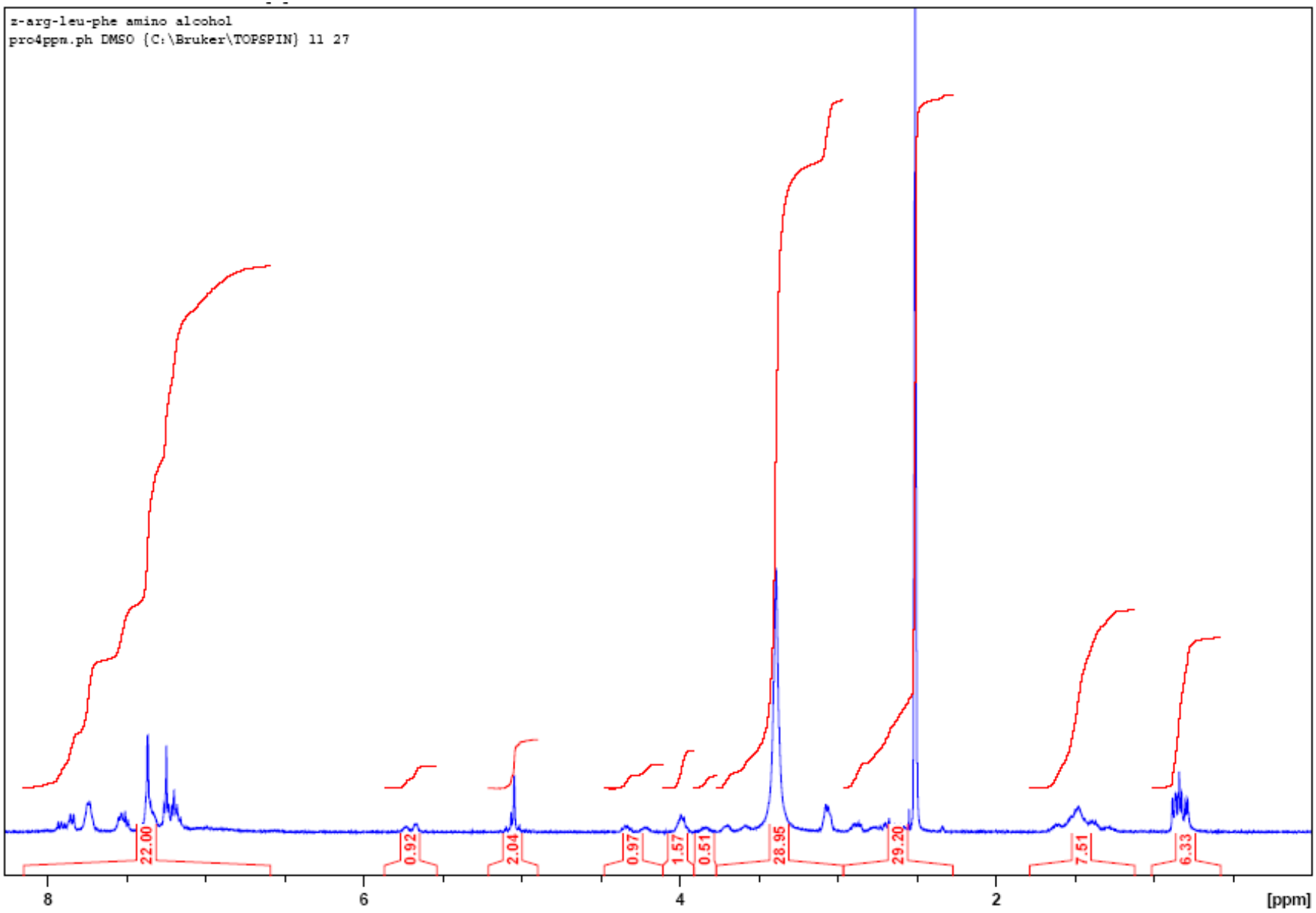
Column: XBridge BEH C18 2.5u
Method: C:\CHEM32\1\METHODS\METHODS\10-90 FINIAL ANALYSISDMSO.M
Instrument: Instrument 1
Flow Rate: 1.5mls/min
Injection Volume: 60ul
Method Info: Analysis carried out using a 100A 4.6 x 50mm column, gradient from 10% - 90% Acetonitrile, in 8 minutes.

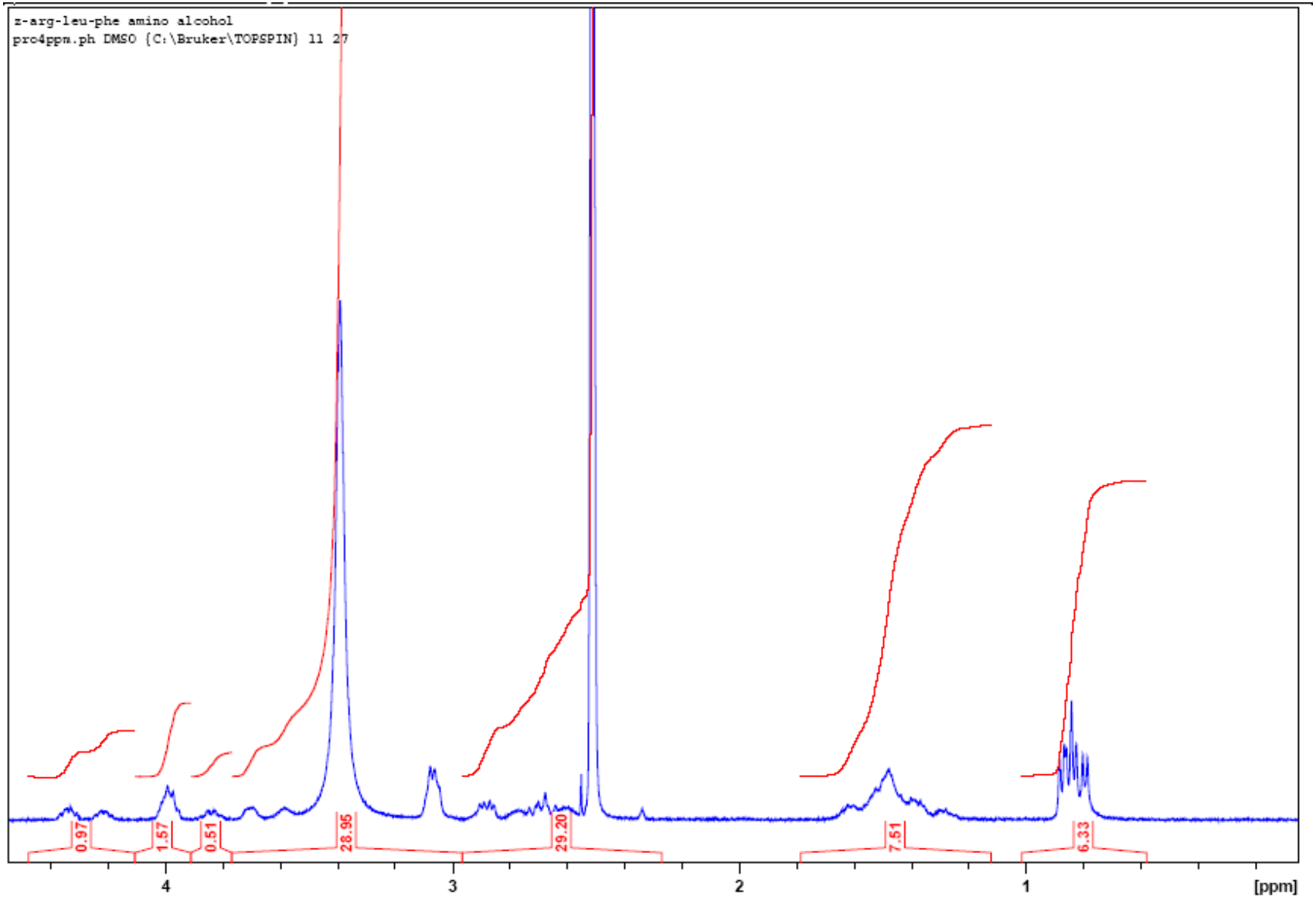


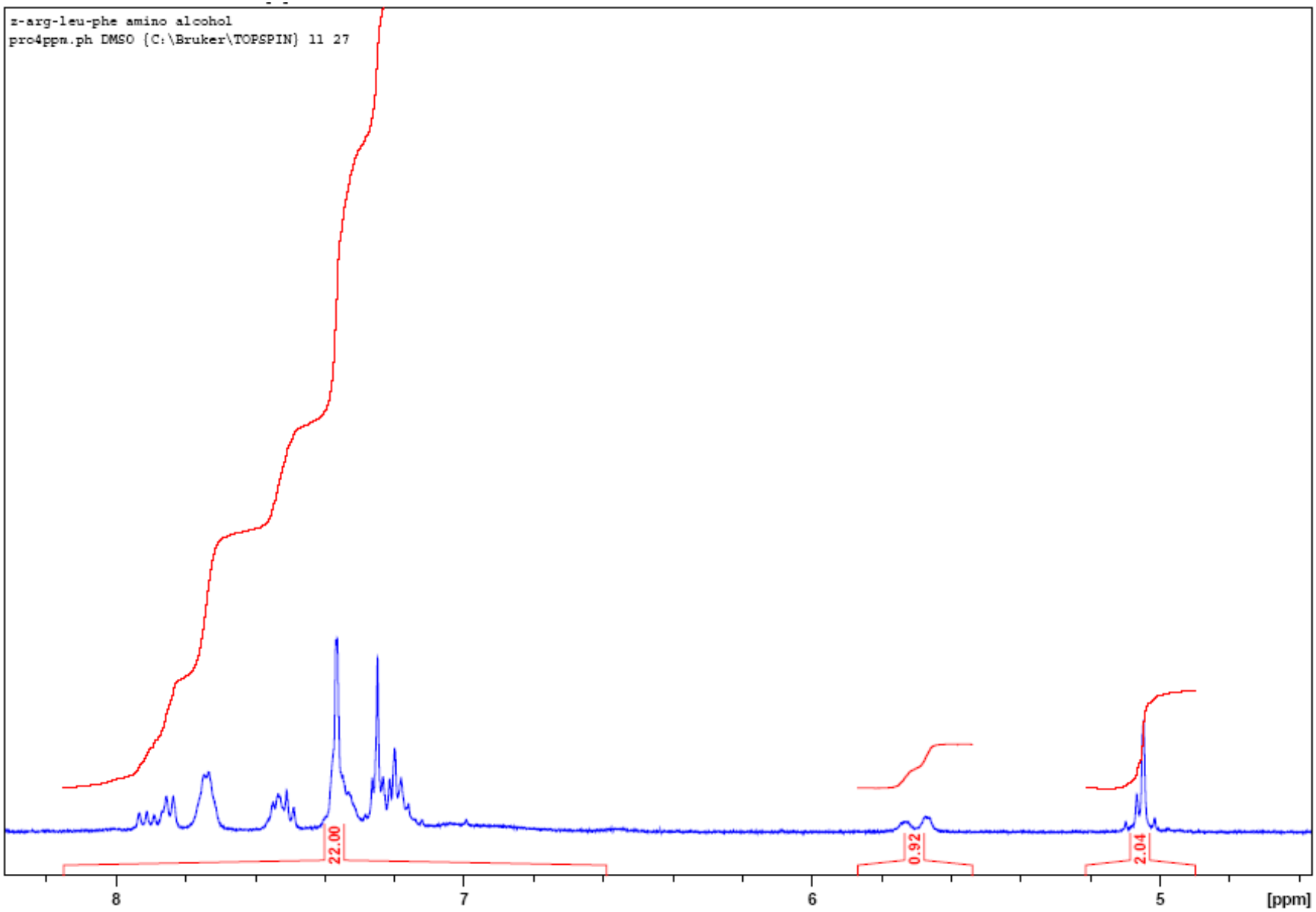
#	Meas. Ret.	Area	Area %
1	6.370	5487.156	100.000

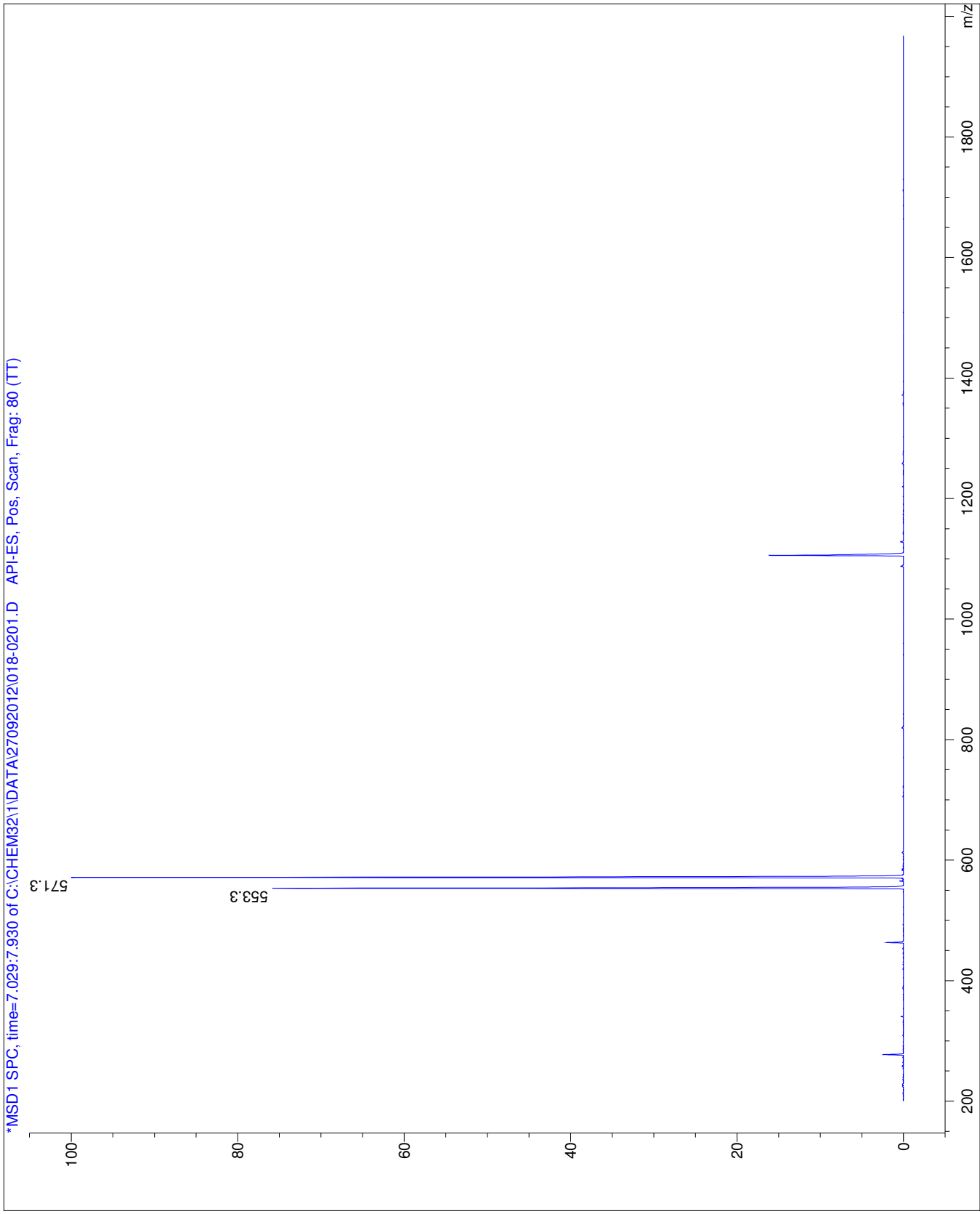


Sample No: Z-RLF-Amino OH
File: C:\CHEM32\1\DATA\27092012\017-0101.D


z-arg-leu-phe amino alcohol
pro4ppm.ph DMSO (C:\Bruker\TOPSPIN) 11 27
22.00
0.92
2.04
0.97
1.57
0.51
28.95
29.20
7.51
6.33
8
6
4
2
[ppm]


z-arg-leu-phe amino alcohol
pro4ppm.ph DMSO (C:\Bruker\TOPSPIN) 11 27
0.97
1.57
0.51
28.96
29.20
7.51
6.33
4
3
2
1
[ppm]


z-arg-leu-phe amino alcohol
pro4ppm.ph DMSO (C:\Bruker\TOPSPIN) 11 27
22.00
0.92
2.04
8
7
6
5
[ppm]



Sample No: Z-RLF-Aldehyde
File: C:\CHEM32\1\DATA\27092012\018-0201.D

Z-RLF aldehyde