UNIVERSITY OF SOUTHAMPTON

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

School of Chemistry

Volume 1 of 1

Evolution of Cyclic Peptide Inhibitors of the Gag-TSG101 Protein-Protein Interaction involved in HIV Budding

by

Katherine Rachel Lennard

Thesis for the degree of Doctor of Philosophy

January 2018

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

School of Chemistry

Thesis for the degree of Doctor of Philosophy

EVOLUTION OF A CYCLIC PEPTIDE INHIBITOR OF THE GAG-TSG101 PROTEIN-PROTEIN INTERACTION INVOLVED IN HIV BUDDING

Katherine Rachel Lennard

In 2015, the World Health Organisation estimated that there were approximately 36.7 million people worldwide living with HIV. Although progress has been made in finding a cure, current treatment options rely on antiretroviral therapies that target different stages of the HIV lifecycle to control viral replication and reduce the risk of transmission. One stage of the HIV lifecycle that is not targeted by current antiretroviral therapies is the process by which nascent HIV virions assemble and bud from the plasma membrane. This process is governed by interactions between host and viral proteins and involves the recruitment the host cells' endosomal sorting complexes required for transport (ESCRT) machinery, through an interaction between the ubiquitin E2 variant (UEV) domain of TSG101 in ESCRT-1 and the p6 domain of the viral Gag protein.

Previous research identified a cyclic peptide inhibitor (CP11) of the UEV-p6 interaction through a screening platform that combines the production of a randomised cyclic peptide library with a reverse two-hybrid system in *E. coli*. This research aimed to further improve the activity of CP11 through alanine-scanning to determine the active motif and to determine the binding partner biophysical assays, such as microscale thermophoresis. A more potent inhibitor was developed through the incorporation of unnatural amino acids that demonstrated improved inhibition of virus-like particle production in a cell-based assay with improved cell permeability over CP11. This work also provides a starting point for the development of small molecule inhibitors of a protein-protein interaction that is vital to the HIV lifecycle.

Table of Contents

Table of Contents .... **iii**

List of Tables .... **ix**

List of Figures .... **xi**

List of Accompanying Materials .... **xix**

DECLARATION OF AUTHORSHIP .... **xxi**

Acknowledgements .... **xxiii**

Definitions and Abbreviations .... **xxv**

Amino Acids .... *xxviii*

Chapter 1 Introduction **1**

1.1 Discovery of Human Immunodeficiency Virus (HIV) *1*

1.1.1 HIV Genome 2

1.2 HIV Lifecycle *3*

1.3 Recruitment of the Endosomal Sorting Complexes Required for Transport by the Viral Gag Polyprotein *6*

1.3.1 The ESCRT Machinery 6

1.3.2 Recruiting the ESCRT Machinery for Successful Viral Budding 7

1.4 Treatments for HIV *9*

1.5 Protein-Protein Interactions *12*

1.5.1 Discovering Inhibitors of PPIs 14

1.5.2 Cyclic Peptides as Potential PPI Inhibitors 15

1.5.3 Discovering Cyclic Peptide PPI Inhibitors 17

1.5.4 Split-Intein Circular Ligation of Proteins and Peptides (SICLOPPS) and Reverse Two-Hybrid System (RTHS) 20

1.5.5 Identification of an Inhibitor of TSG101-Gag PPI 22

1.6 Project Aims *24*

Chapter 2 Investigation into Small Motifs of CP11 .... **27**

2.1 Introduction *27*

2.2 Development of an Enzyme-Linked Immunosorbent Assay (ELISA) *27*

2.2.1 Recombinant Protein Expression 28

2.2.2 Determining the Concentration of Protein for ELISA 32

2.2.3 Assessing Binding Specificity 33

2.2.4 Testing the Effects of Increasing Concentrations of DMSO 34
2.3 Testing Synthesised Peptides for Inhibitory Activity in the p6-UEV ELISA 35
2.3.1 Testing CP11 and Capped Tri-Peptides in an ELISA..37
2.4 Conformational Analysis of Tri-Peptide Motif..38
2.4.1 Scalar Coupling Coupling Constants..39
2.4.2 Cross-Relaxation Data..40
2.4.3 Ramachandran and Janin Plots...41
2.4.4 Chemical Shift Assignment of Acetyl-NWY-Diethylamine and Structural Restraint Measurement...41
2.4.5 Development of a Computational Model 47
2.4.6 Inputting Data into the Model to Determine Conformation of Acetyl-NWY-Diethylamine 49
2.4.7 Determination of the Main Conformational Shapes of Acetyl-NWY-Diethylamine.....54
2.4.8 Comparison of the Determined Conformation of Acetyl-NWY-Diethylamine with an Structural Model of CP11 55
2.5 Synthesis of Sterically Hindered Tri-Peptides 57
2.6 Summary...59
Chapter 3 Determining the Active Motif of CP11 61
3.1 Introduction 61
3.2 Alanine Scan using the p6-UEV RTHS..61
3.2.1 Generation of SICLOPPS Plasmids for Alanine Scan in a RTHS...................................61
3.2.2 Optimisation of the Conditions for Drop-Spotting 63
3.3 Alanine Scan using an ELISA...67
3.3.1 Synthesis of Alanine Analogues of CP11 by Solid-Phase Peptide Synthesis 67
3.3.2 Assessing the Effect of the Alanine Analogues Using the p6-UEV ELISA 67
3.4 Synthesis and Testing of Acetyl-IYW-Diethylamine and Di-Peptide Truncations in an ELISA 69
3.5 Summary...71
Chapter 4 Investigating the Structure-Activity Relationship of CP11 73
4.1 Introduction 73
4.2 SPPS of Non-Natural Analogues...74
4.3 Testing in the p6-UEV ELISA 75
4.3.1 Testing Cysteine Derivatives of Most Active Peptides..81
4.4 Development of Biophysical Assays to Determine the Activity of CP11 Derivatives In Vitro 82
4.4.1 Thermal Shift Assay...82

4.4.2 Microscale Thermophoresis 88
4.4.3 Assessing the Binding of Most Active Peptides to UEV 93
4.5 Summary........ 95

Chapter 5 Assessing the Efficacy of a CP11 Derivative in a Cell-Based Assay of HIV-Budding.97

5.1 Introduction 97
5.2 Testing Peptides in a Virus-Like Particle Assay 97
5.2.1 Testing Peptide Cytotoxicity........ 100
5.3 Testing the Effect of the Most Active CP11 Derivative on EGFR Degradation 101
5.4 Creation of a Stable Cell Line to Express Cyclic Peptide Inhibitors of HIV Budding 102
5.4.1 Performing a VLP Assay with T-REx293-CP11A4 Cells 106
5.5 Summary........ 108

Chapter 6 Conclusions........109

Chapter 7 Future Directions 114

Chapter 8 Experimental........117

8.1 Materials and Equipment 117
8.1.1 Molecular Biology........ 117
8.1.2 Peptide Synthesis 117
8.1.3 HPLC Purification........ 117
8.1.4 NMR Analysis........ 117
8.1.5 Biophysical Assays 118
8.1.6 Mammalian Cell Culture........ 118
8.1.7 Curve Fitting using GraphPad Prism7 Software 118
8.2 General Methods 118
8.2.1 Cyclic Peptide Synthesis 118
8.3 General Molecular Biology Methods 120
8.3.1 Luria Broth (LB) Medium 120
8.3.2 Making Overnight Cultures 120
8.3.3 Plasmid Miniprep 120
8.3.4 LB Agar Plates........ 120
8.3.5 Agarose Gels........ 121
8.3.6 PCR for Amplification of UEV and p6........ 121
8.3.7 PCR Product Purification 122
8.3.8 Restriction Digestion 122
8.3.9 Ligation Procedure for Cloning Genes into Vectors 122
8.3.10 Making Chemically Competent Cells........ 123

8.3.11 Transformation into Chemically Competent Cells 124
8.3.12 Procedure for Colony PCR 124
8.3.13 Expression and Purification of His$_6$-tagged UEV and GST-tagged p6 125
8.3.14 Procedure for Running SDS Gels 126
8.4 Experimental for Chapter 2 127
8.4.1 CP11 127
8.4.2 Liquid Phase Synthesis of Tri-Peptides 128
8.4.3 Solid Phase Synthesis of Linked Tri-Peptides 140
8.5 Experimental for Chapter 4 146
8.5.1 Generation of pARCBD-CP11 146
8.5.2 Site Directed Mutagenesis for Generation of Alanine Analogues 148
8.5.3 Drop-Spotting Alanine Analogues 149
8.5.4 Enzyme-Linked Immunosorbent Assay 150
8.5.5 CP11 Analogue 1 (A1) 150
8.5.6 CP11A2 151
8.5.7 CP11A3 152
8.5.8 CP11A4 153
8.5.9 CP11A5 154
8.5.10 CP11A6 155
8.5.11 CP11A7 156
8.5.12 CP11A8 157
8.5.13 Synthesis of Tri- and Di-Peptide Motifs 158
8.6 Experimental for Chapter 5 171
8.6.1 CP11Y* 171
8.6.2 CP11A4Y* 182
8.6.3 CP11A5Y* 194
8.6.4 CP11A8Y* 205
8.6.5 Solid Phase Peptide Synthesis of Cysteine Derivatives 217
8.6.6 Thermal Shift 220
8.6.7 Microscale Thermophoresis 221
8.7 Experimental for Chapter 6 221
8.7.1 Plasmids for Cell Culture 221
8.7.2 Mammalian Cell Lines 222
8.7.3 Preparation of Mammalian Cell Stocks 222
8.7.4 Thawing Frozen Mammalian Cell Stocks 222
8.7.5 General Cell Culture Procedures 222
8.7.6 Western Blotting Procedure 223

8.7.7 Virus-Like Particle Assay 224
8.7.8 Cytotoxicity Assay 225
8.7.9 EGFR Down-Regulation Assay 225
8.7.10 Creation of a Stable Cell Line 225

Appendices **228**

Bibliography **229**

List of Tables

Table 1: Approved Novel Antiretroviral Drugs. 11

Table 2: Ranking of Hit Peptide Sequences Identified Through Screening SICLOPPS Libraries SGWX$_5$ and SGWXXPXXPXX in the p6-UEV RTHS. 23

Table 3: The Correlation Between Vicinal $^3J_{HH}$ and $^3J_{CH}$ Coupling Constants and Their Dihedral Angles[125]. 39

Table 4: $^3J_{HH}$ Coupling Constants for Acetyl-NWY-Diethylamine in Solution. 45

Table 5: Definition of Torsions of Acetyl-NWY-Diethylamine 48

Table 6: Table of Restraints used in the Model 50

Table 7: Names and Sequences of Alanine Analogues 62

Table 8: IC$_{50}$ of CP11 and Alanine Analogues 69

Table 9: Comparison of IC$_{50}$ Values of Cysteine Derivatives and the Original Analogues 82

Table 10: Working Concentrations of Antibiotics. 120

Table 11: Composition of TAE buffer. 121

Table 12: Composition of Solution for Gene Amplification by PCR 121

Table 13: Thermo Cycler Set Up 121

Table 14: Primers and Sequences 122

Table 15: Composition of Solutions for Restriction Digestion 122

Table 16: Composition of Ligation Solutions 123

Table 17: Composition of TBF I buffer. 123

Table 18: Composition of TBF II buffer. 123

Table 19: Composition of SOC medium. 124

Table 20: Composition for Colony PCR 124

Table 21: Thermo Cycler Set Up for Colony PCR. 124

Table 22: Reagents for Making Gels for SDS-PAGE Analysis of Proteins 126

Table 23: SDS-PAGE Running Buffer (5 ×)..126

Table 24: Reagents for PCR with pARCBD to generate pARCBD-CP11 146

Table 25: Table of Primers for Creation of pARCBD-CP11..147

Table 26: Thermocycler Program for Creation of pARCBD-CP11 147

Table 27: Primer Sequences for Zipper Primer PCR 147

Table 28: Reagents for Restriction Digestion of pARCBD-CP11 148

Table 29: Reagents for Ligation into pARCBD 148

Table 30: Reagents for Site Directed Mutagenesis 149

Table 31: Thermocycler Program for Site Directed Mutagenesis 149

Table 32: Sequences of Primers for Site Directed Mutagenesis...149

Table 33: Transfer Buffer Composition 224

Table 34: Amount of Plasmid Transfected into T-REx293 Cells for Creation of a Stable Cell Line225

List of Figures

Figure 1: HIV Genome and Viral Structure. 2

Figure 2: Schematic Representation of the HIV Lifecycle. 4

Figure 3: A Simplified Depiction of the Interactions of the ESCRT Machinery. 6

Figure 4: Interaction Between the UEV Domain of TSG101 and a Fragment of the p6 Domain of Gag. 8

Figure 5: Schematic of HIV Budding Mechanism. 9

Figure 6: Common Systems for the Creation of Macrocyclic Libraries. 18

Figure 7: Proposed SICLOPPS Mechanism for Production of Cyclic Peptides. 20

Figure 8: The Reverse Two-Hybrid System. 22

Figure 9: Testing Cyclic Peptides in a Mammalian VLP assay. 24

Figure 10: A. General Scheme for an ELISA investigating Inhibitors of the UEV-p6 PPI. B. Oxidation Mechanism of TMB. 28

Figure 11: Plasmid Map of pET28a-UEV. 29

Figure 12: Representative SDS Gel of Fractions Collected During Purification of His$_6$-UEV by FPLC using a HisTrap Column. 30

Figure 13: Plasmid Map of PGEX2TK-p6. 31

Figure 14: Representative SDS Gel of Fractions Collected During Purification of p6-GST by FPLC using a GSTrap Column. 32

Figure 15: Optimisation of Concentration of His$_6$-UEV to be used in ELISA. 33

Figure 16: ELISA Investigating a Blocking Step to Prevent Non-Specific Binding of p6-GST to the Nickel-Coated Plate. 33

Figure 17: ELISA Investigating the Non-Specific Binding of GST versus p6-GST. 34

Figure 18: ELISA with His$_6$-UEV and p6-GST with Increasing Concentrations of DMSO. 34

Figure 19: Mechanism of Fmoc Deprotection. 36

Figure 20: Results of an ELISA with CP11. 37

Figure 21: Results of an ELISA with A. Acetyl-NWY-Diethylamine and B. Acetyl-YWN-Diethylamine. 38

Figure 22: Graph Showing the Difference Between Cross-Relaxation Rate Constants Obtained by NOESY and ROESY. 40

Figure 23: A. A Typical Ramachandran Plot. B. A Representative Janin Plot, Based Upon the Janin Plot for Tyrosine. C. Amino Acid Dihedral Bond Torsion Nomenclature. 41

Figure 24: Atom Nomenclature for Acetyl-NWY-Diethylamine 42

Figure 25: NMR Spectra of a pH Series of Acetyl-NWY-Diethylamine..43

Figure 26: NMR spectra of a Temperature Series of Acetyl-NWY-Diethylamine. 44

Figure 27: Karplus Graphs of the Relationship Between Backbone Amide and α-Protons. 46

Figure 28: Atom Nomenclature for Acetyl-NWY-Diethylamine, with Torsion Nomenclature. ...48

Figure 29: A. Ensemble of all Possible Conformations of Acetyl-NWY-Diethylamine. B. Idealised Conformations of Acetyl-NWY-Diethylamine..50

Figure 30: Three Major Conformational Shapes of Acetyl-NWY-Diethylamine. 51

Figure 31: Dihedral Angles of N1, Plotted on A. Ramachandran Plot. B. Janin Plot. 52

Figure 32: Dihedral Angles of W2, Plotted on A. Ramachandran Plot. B. Janin Plot. 53

Figure 33: Dihedral Angles of Y3, Plotted on A. Ramachandran Plot. B. Janin Plot....................53

Figure 34: The Five Main Conformational Shapes Acetyl-NWY-Diethylamine Forms in Solution Organised by the Backbone Conformation. 54

Figure 35: Structural Model of CP11. 56

Figure 36: Overlaid Conformations of αβ Backbone of CP11 and Acetyl-NWY-Diethylamine Conformers. 57

Figure 37: Results of an ELISA with Sterically Hindered Cyclo-NWY Derivatives........................58

Figure 38: Plasmid Map of pARCBD-CP11. 62

Figure 39: Screening Alanine Analogues in the p6-UEV RTHS by Drop-Spotting onto Minimal Media Agar. 64

Figure 40: Optimisation of ʟ-(+)-Arabinose Concentrations for Drop-Spotting. 65

Figure 41: Results of a Screen with Lower Concentrations of ʟ-(+)-Arabinose (0.5-1.5 μM) With and Without IPTG (30 μM). 66

Figure 42: Results of Alanine Scan when Tested in an ELISA. 68

Figure 43: Results of an ELISA Testing Acetyl-IYW-Diethylamine, Acetyl-IY-Diethylamine and Acetyl-YW-Diethylamine. 70

Figure 44: Structure of the Four Backbones and the Unnatural Aromatic Analogues 74

Figure 45: Graphs of ELISA Results with Tyrosine Analogues Based upon the CP11 Backbone. 76

Figure 46: Graphs of ELISA Results with Tyrosine Analogues Based upon CP11A4 Backbone. .. 77

Figure 47: Graphs of ELISA Results with Tyrosine Analogues Based upon CP11A5 Backbone. .. 78

Figure 48: Graphs of ELISA Results with Tyrosine Analogues Based upon CP11A8 Backbone. .. 79

Figure 49: ELISA Results of CP11 SARs Investigating the Tyrosine Residue. 81

Figure 50: Results of an ELISA testing A. CP11A4Y3C, B. CP11A4Y4C and C. CP11A4Y10C. 82

Figure 51: Schematic Showing the Typical Melting Profile of a Protein. 83

Figure 52: Thermal Shift assay with Increasing Concentrations of UEV and p6. 84

Figure 53: Thermal Shift Assays with UEV and p6 and Increasing Concentrations of DMSO. 86

Figure 54: Thermal Shift Assays with UEV and p6 and Increasing Concentrations of CP11. 87

Figure 55: Schematic Showing the Five Main Events in an MST Signal. 90

Figure 56: Structure of p6 Drawn in MacPyMOL. 91

Figure 57: Surface of UEV Protein Showing Cysteine (red) and Lysine (blue) Side-Chains and p6 PTAP peptide (purple). 92

Figure 58: Capillary Scan of Fluorescently Labelled UEV in a Premium Capillary (Left) and Standard Capillary (Right) in HEPES Buffer. 93

Figure 59: Graphs of MST Analysis of Most Active Peptides with UEV labelled with NHS Fluorescent Dye. 94

Figure 60: Fluorescence Microscopy of HEK293T Cells Transfected with pMET7-Gag-eGFP Showing VLPs in Solution. 98

Figure 61: Western Blots of HEK293T Cells Transfected with Gag-eGFP and Treated with CP11A4Y3. 99

Figure 62: Assessing the Cytotoxicity of CP11A3Y4 on HEK293T Cells. 101

Figure 63: Western Blot Analysis of the Effect of CP11A4Y3 on EGFR Downregulation. 102

Figure 64: Schematic Showing the General Process for Generation of a Stably Integrated Cell Line Using Flippase/Flippase Recognition Target Recombination. 104

Figure 65: Western Blot Showing the Production of CBD-tagged N-Intein at 18 kDa. 105

Figure 66: CP11A4 Isolated from T-REx293-CP11A4 Cell Lysate. 106

Figure 67: Western Blot Analysis of VLP Assay Performed with T-REx293-CP11A4 Cells. 107

Figure 68: CP11 127

Figure 69: Synthesis of Fmoc-Tyr(tBu)-Diethylamine 128

Figure 70: Synthesis of H_2N-Tyr(tBu)-Diethylamine 129

Figure 71: Synthesis of Fmoc-Trp(Boc)-Tyr(tBu)-Diethylamine 129

Figure 72: Synthesis of H_2N-Trp(Boc)-Tyr(tBu)-Diethylamine 130

Figure 73: Synthesis of Fmoc-Asn(Trt)-Trp(Boc)-Tyr(tBu)-Diethylamine 130

Figure 74: Synthesis of H_2N-Asn(Trt)-Trp(Boc)-Tyr(tBu)-Diethylamine 131

Figure 75: Synthesis of Acetyl-Asn(Trt)-Trp(Boc)-Tyr(tBu)-Diethylamine 131

Figure 76: Synthesis of Acetyl-Asn-Trp-Tyr-Diethylamine 133

Figure 77: Synthesis of Fmoc-Asn(Trt)-Diethylamine 135

Figure 78: Synthesis of H_2N-Asn(Trt)-Diethylamine 135

Figure 79: Synthesis of Fmoc-Trp(Boc)-Asn(Trt)-Diethylamine 136

Figure 80: Synthesis of H_2N-Trp(Boc)-Asn(Trt)-Diethylamine 136

Figure 81: Synthesis of Fmoc-Tyr(tBu)-Trp(Boc)-Asn(Trt)-Diethylamine 137

Figure 82: Synthesis of H_2N-Tyr(tBu)-Trp(Boc)-Asn(Trt)-Diethylamine 137

Figure 83: Synthesis of Acetyl-Tyr(tBu)-Trp(Boc)-Asn(Trt)-Diethylamine 138

Figure 84: Synthesis of Acetyl-Tyr-Trp-Asn-Diethylamine 139

Figure 85: Cyclo-Ahx-NWY 141

Figure 86: Cyclo-Ava-NWY 143

Figure 87: Cyclo-Gaba-NWY 145

Figure 88: CP11A1 151

Figure 89: CP11A2 152

Figure 90: CP11A3 153

Figure 91: CP11A4 154

Figure 92: CP11A5 155

Figure 93: CP11A6 156

Figure 94: CP11A7 157

Figure 95: CP11A8 158

Figure 96: Mechanism of the Synthesis of Fmoc-Trp(Boc)-Diethylamine 159

Figure 97: Mechanism of the Synthesis of H_2N-Trp(Boc)-Diethylamine 159

Figure 98: Mechanism of the Synthesis of Fmoc-Tyr(tBu)-Trp(Boc)-Diethylamine 160

Figure 99: Mechanism of the Synthesis of H_2N-Tyr(tBu)-Trp(Boc)-Diethylamine 160

Figure 100: Mechanism of the Synthesis of Fmoc-Ile-Tyr(tBu)-Trp(Boc)-Diethylamine 161

Figure 101: Mechanism of the Synthesis of H_2N-Ile-Tyr(tBu)-Trp(Boc)-Diethylamine 161

Figure 102: Mechanism of the Synthesis of Acetyl-Ile-Tyr(tBu)-Trp(Boc)-Diethylamine 162

Figure 103: Mechanism of the Synthesis of Acetyl-Ile-Tyr-Trp-Diethylamine 163

Figure 104: Synthesis of Fmoc-Ile-Tyr(tBu)-Diethylamine 165

Figure 105: Synthesis of H_2N-Ile-Tyr(tBu)-Diethylamine 165

Figure 106: Synthesis of Acetyl-Ile-Tyr(tBu)-Diethylamine 166

Figure 107: Mechanism of the Synthesis of Acetyl-Ile-Tyr-Diethylamine 166

Figure 108: Synthesis of Acetyl-Tyr(tBu)-Trp(Boc)-Diethylamine..168

Figure 109: Synthesis of Acetyl-Tyr-Trp-Diethylamine 169

Figure 110: CP11Y1..171

Figure 111: CP11Y2..172

Figure 112: CP11Y3..173

Figure 113: CP11Y4..174

Figure 114: CP11Y5..175

Figure 115: CP11Y6..176

Figure 116: CP11Y7..177

Figure 117: CP11Y8..178

Figure 118: CP11Y9..179

Figure 119: CP11Y10..180

Figure 120: CP11Y11..181

Figure 121: CP11A4Y1 183

Figure 122: CP11Y2..184

Figure 123: CP11A4Y3 185

Figure 124: CP11A4Y4 186

Figure 125: CP11A4Y5 187

Figure 126: CP11A4Y6 188

Figure 127: CP11A4Y7 189

Figure 128: CP11A4Y8 190

Figure 129: CP11A4Y9 191

Figure 130: CP11A4Y10 192

Figure 131: CP11A4Y11 193

Figure 132: CP11A5Y1 195

Figure 133: CP11A5Y2 196

Figure 134: CP11A5Y3 197

Figure 135: CP11A5Y4 198

Figure 136: CP11A5Y5 199

Figure 137: CP11A5Y6 200

Figure 138: CP11A5Y7 201

Figure 139: CP11A5Y8 202

Figure 140: CP11A5Y9 203

Figure 141: CP11A5Y10 204

Figure 142: CP11A5Y11 205

Figure 143: CP11A8Y1 206

Figure 144: CP11A8Y2 207

Figure 145: CP11A8Y3 208

Figure 146: CP11A8Y4 209

Figure 147: CP11A8Y5 210

Figure 148: CP11A8Y6 211

Figure 149: CP11A8Y7 212

Figure 150: CP11A8Y8 213

Figure 151: CP11A8Y9 214

Figure 152: CP11A8Y10 215

Figure 153: CP11A8Y11 216

Figure 154: CP11A4Y3C 217

Figure 155: CP11A4Y4C 218

Figure 156: CP11A4Y10C 219

List of Accompanying Materials

One CD containing raw data, mass spec traces and analytical HPLC traces.

DECLARATION OF AUTHORSHIP

I, ...[please print name]

declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

EVOLUTION OF A CYCLIC PEPTIDE INHIBITOR OF THE GAG-TSG101 PROTEIN-PROTEIN INTERACTION INVOLVED IN HIV BUDDING

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. [Delete as appropriate] None of this work has been published before submission [or] Parts of this work have been published as: [please list references below]:

Signed: ....

Date: ....

Acknowledgements

Firstly, I would like to thank Professor Ali Tavassoli for the opportunity to complete this research and for all his help and advice throughout my PhD. I would also like to thank our collaborators at C4X Discovery for their support and for allowing me to complete a placement at their company, particularly Sadia Mohammed for her patience, and Thorsten Nowak.

Thanks, of course, must go to members of the Tavassoli group, in particular to Eilidh Leitch for her support and who always knew when more coffee was needed; to Andy Foster for making the pcDNA5/FRT/CP11A4 plasmid and always being willing to listen to a rant; to my running buddy, Eliot Osher; and to the best Master's student anyone could ask for, Cyrielle Doigneaux – good luck with your own PhD!

Finally, I would like to thank my family and friends, and especially Alex Byrne - without their unwavering support and encouragement I'm not sure that I would've had gotten this far.

Definitions and Abbreviations

3-AT	3-amino-1,2,4-triazole
Ahx	6-aminohexanoic acid
AICAR	5-Aminoimidazole-4-carboxamide ribonucleotide
AIDS	Acquired Immunodeficiency Syndrome
ALIX	Apoptosis Linked gene-2-Interacting protein X
ART	Antiretroviral Therapy
ARV	Antiretroviral
ATIC	aminoimidazole carboxamide ribonucleotide transformylase/inosine mono-phosphate cyclohydrolase
Ava	5-aminovaleric acid
CA	Capsid
cART	Combination Antiretroviral Therapy
CBD	Chitin Binding Domain
CDC	US Centres for Disease Control and Prevention
DIC	*N,N'*-diisopropylcarbodiimide
DCM	Dichloromethane
DMF	Dimethylformamide
DNA	Deoxyribonuecleic acid
dNTP	Deoxyribonucleotide triphosphates
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal Growth Factor
eGFP	Enhanced Green Fluorescent Protein
EGFR	Epidermal Growth Factor Receptor
ELISA	Enzyme-Linked Immunosorbent Assay
Env	Envelope
ESCRT	Endosomal Sorting Complexes Required for Transport
FBDD	Fragment Based Drug Design
FDA	Food and Drug Administration
FIT	*F*lexible *I*n-vitro *T*ranslation
fmoc	9-fluorenylmethoxycarbonyl
FPLC	Fast Protein Liquid Chromatography
Gaba	γ-aminobutyric acid
GFP	Green Fluorescent Protein
GST	Glutathione S-Transferase
HATU	2-(7-Aza-1H-benzotriazole-1-yl)-1,1,3,3,-tetramethyluronium
HIS	Histidine
HIS3	Imidazole Glycerol Phophotase Dehydratase

HIV	Human Immunodeficiency Virus
HOAt	1-Hydroxy-7-azabenzotriazole
HOBt	Hydroxybenzotriazole
HPLC	High Performance Liquid Chromatography
Hrs	Hepatocyte growth factor-regulated tyrosine kinase substrate
HTS	High Throughput Screening
IN	Integrase
IL-2	Interleukin-2
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal Titration Calorimetry
KanR	Aminoglycoside 3'-phosphotransferase
KS	Kaposi's Sarcoma
*lac*Z	β-Galactosidase
LB	Luria Broth
LC-MS	Liquid Chromatography Coupled to Mass Spectrometry
LTR	Long Terminal Repeats
MA	Matrix
MOPS	3-(N-morpholino)propanesulfonic acid
MST	Microscale Thermophoresis
MVB	Multivesicular body
NMR	Nuclear Magnetic Resonance
NRTIs	Nucleoside Reverse Transcriptase Inhibitors
NNRTIs	Non-Nucleoside Reverse Transcriptase Inhibitors
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
Npu	*Nostoc punctiforme*
NRTIs	Nucleoside/Nucleotide Reverse Transcriptase Inhibitors
OD	Optical Density
OIs	Opportunistic Infections
PCP	*Pneumocystis Carinii* Pneumonia
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
$PI(4,5)P_2$	phosphatidyl inositol (4,5) biphosphate
PPI	Protein- Protein Interactions
PR	Protease
PrEP	Pre Exposure Prophylaxis
Rev	Regulator of Virion
RNA	Ribonucleic Acid
ROE	Rotating frame nuclear Overhauser Effect
ROESY	Rotating frame nuclear Overhauser Effect Spectroscopy
RRE	Rev Response Element

RT	Reverse Transcriptase
RTHS	Reverse Two-Hybrid System
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SICLOPPS	Split Intein-Mediated Circular Ligation of Peptides and Proteins
siRNA	Small Interfering RNA
SIV	Simian Immunodeficiency Virus
SOC	Super Optimal broth with Catabolite repression
SPPS	Solid-Phase Peptide Synthesis
SPR	Surface Plasmon Resonance
Ssp	*Synechocystis* sp PCC6803
T-ROESY	Transverse-Rotating frame nuclear Overhauser Effect Spectroscopy
TAE	Tris-acetate-EDTA
Tat	Trans-Activator of Transcription
TEMED	Tetramethylethylenediamine
TFA	Trifluoroacetic acid
TIS	Triisopropylsilane
TOCSY	Total Correlation Spectroscopy
T_m	Melting temperature
TMB	3,3',5,5'-Tetramethylbenzidine
TSAP	Thermosensitive alkaline phosphatase
TSG101	Tumour Susceptibility Gene 101
UEV	Ubiquitin E2 Variant
UV	Ultraviolet
VLP	Virus-Like Particle
WHO	World Health Organisation

Amino Acids

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

Chapter 1 Introduction

1.1 Discovery of Human Immunodeficiency Virus (HIV)

In 1981, it was noted that there was a rise in the number of cases of an opportunistic infection, *Pneumocystis carinii* pneumonia (PCP), among young homosexual men and drug users in the US[1]. PCP was a rare infection that usually only occurs in individuals that are severely immunocompromised or suffering from cancer. In the same year, there were cases of other homosexual men developing the rare cancer, Kaposi's sarcoma (KS)[2]. KS was a type of cancer that predominantly affected elderly men, however in these cases it was found to be much more aggressive in younger males and in some cases proved fatal. The number of reported cases of both PCP and KS rose quickly along with a rise in the number of cases of other opportunistic infections. It was noted that all patients had a low $CD4^+$ cell count, a marked swelling of the lymph nodes and an opportunistic infection that often proved life threatening. As the patients did not have a history of other immune deficiency diseases, it was called Acquired Immune Deficiency Syndrome (AIDS). It was soon determined that AIDS was not solely isolated to men who have sex with men and intravenous drug users, as sex workers and haemophiliacs were also found to suffer from the disease.

It took two years for the causative virus behind AIDS to be discovered. Two research groups simultaneously managed to isolate the virus from patients exhibiting symptoms of AIDS[3-4]. It was initially believed that the virus was a member of the human T-cell Leukaemia Virus (HTLV) family, due to the low $CD4^+$ cell count in patients, and therefore named the virus HTLV-III[4]. However, Montagnier *et al.* determined that the core proteins were not immunologically related to HTLV, and therefore named the virus they had isolated as lymphadenopathy-associated virus (LAV)[3], in reference to the most common symptoms of swollen lymph glands. In 1984, another group managed to establish a stable cell line of T-cells from a patient with AIDS, and confirmed that the virus was not related to HTLV[5], but named the virus AIDS-associated Retrovirus (ARV). Eventually it was confirmed that HTLV-III, LAV and ARV were all variants of the same virus and related to the *Lentivirus* genus of the *Retroviridae* family; HTLV-III and LAV shared approximately 98% amino acid homology, while ARV had a slightly higher variation with approximately 90% amino acid homology[6]. The International Committee on the Taxonomy of Viruses then proposed to rename the virus human immunodeficiency virus (HIV) to conform to common nomenclature for retroviruses[7].

There are two major types of HIV: HIV-1 and HIV-2. The most common and pathogenic strain is HIV-1. HIV-1 can be subdivided into four major groups (M, N, O, and P) based upon the supposed species of origin. Group M is the most prevalent of the four groups and is thought to be of chimpanzee origin[8]. Groups N, O and P are far less prevalent than group M and are thought to be of either gorilla or chimpanzee origin. HIV-2 is less virulent than HIV-1 and is mainly restricted to West Africa. It is distantly related to HIV-1, and is thought to be caused by zoonotic transfer from the sooty mangabey species of monkey as it closely resembles a form of simian immunodeficiency virus (SIV) that is prevalent in this species. HIV-2 is thought to have a much lower transmission rate than HIV-1 and to be less pathogenic, although the mechanisms involved are not fully understood. HIV-1 is the pandemic form of the virus and its mechanism of action is far better understood, therefore HIV-1 will be referred to as HIV for simplicity throughout this thesis.

1.1.1 HIV Genome

The HIV virion is roughly spherical in structure and approximately 120 nm in diameter[9]. It's genetic information is stored on two copies of positive-sense RNA surrounded by a conical capsid, Figure 1. The HIV genome encodes 16 proteins, flanked by two long terminal repeats (LTRs). These proteins include major structural proteins (Gag, Pol and Env), accessory proteins (Vif, Vpr, Vpu and Nef) and regulatory proteins (Tat and Rev). The Gag gene encodes the structural proteins: matrix (MA), capsid (CA) and nucleocapsid (NC), but is also vital for assembly and budding of viral particles from the plasma membrane. The Pol gene encodes for the viral enzymes: protease (PR), reverse transcriptase (RT) and integrase (IN). The Env gene encodes for the envelope proteins: gp120 and gp41.

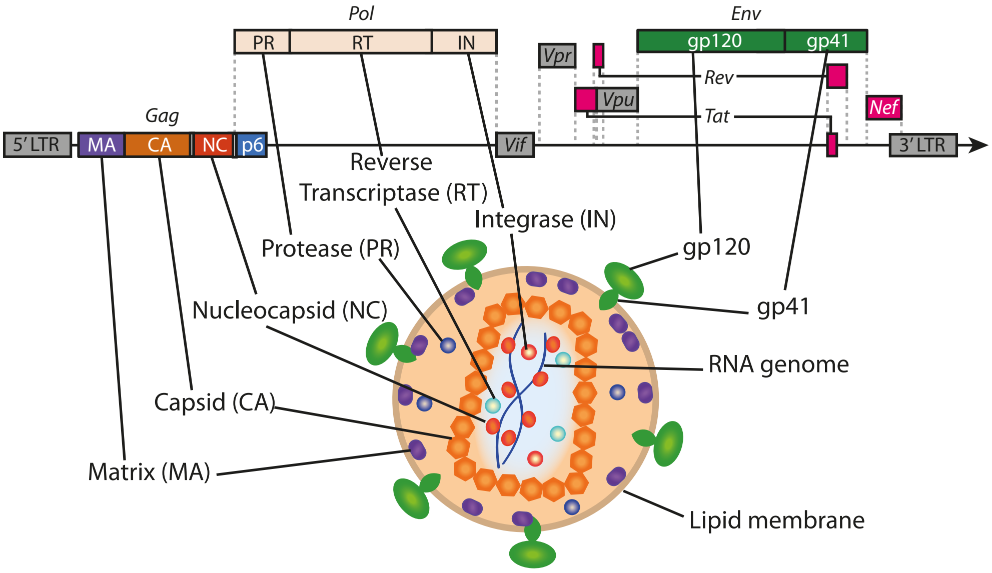


Figure 1: HIV Genome and Viral Structure. Information adapted from *Watts et al.* [10]

1.2 HIV Lifecycle

The HIV lifecycle can be described in eight main stages: binding, fusion, reverse transcription, integration, transcription and translation, assembly, budding, and maturation, Figure 2. At each stage in its life cycle, HIV hijacks the host cell's machinery through a series of interactions between viral and host proteins.

HIV primarily infects CD4$^+$ T-cells and macrophages through a complex, multistep process. The initial interaction is the binding of the gp120 protein on the HIV envelope to the CD4$^+$ surface receptor[11]. This initiates the formation of a bridging sheet between the inner and outer domains of gp120 and exposes a second cell surface-binding site[12]. The engagement of the co-receptor site leads to the insertion of a fusion peptide into the cell membrane by the amino terminus of gp41.

Once the virus has fused to the cell, the viral CA protein partially un-coats[12] to allow the single-stranded HIV RNA to be copied into double-stranded DNA. This process occurs via the viral RT, as the host cell does not have the cellular machinery for this conversion. As RT is essential for the HIV lifecycle and is not endogenous to the host cells, it is an attractive target for drug design however, due to RT's low fidelity, and therefore high mutation rate[13-14], drugs targeting this enzyme are highly susceptible to acquired resistance.

Once generated, the viral DNA enters the CD4 cell nucleus and is then incorporated into the host genomic DNA in a series of reactions mediated by the HIV IN enzyme. As is true for RT, IN is not common to the host and is essential to the HIV lifecycle, making it another attractive target for potential drug development. In the first step of integration, the viral IN cleaves two nucleotides from the 3' end of each blunt-ended DNA strand, to expose terminal hydroxyl groups. The viral DNA remains bound to the IN and is transported into the cell nucleus where it is inserted into the host DNA by strand transfer, catalysed by the viral IN.

After the viral DNA has been integrated into the host genome, it is transcribed by a combination of viral and host proteins. The viral protein Tat, up-regulates transcription of the viral DNA by binding to cellular kinases, triggering the phosphorylation of the C-terminal domain of the host RNA polymerase and causing a rapid increase in transcription. The viral mRNA is spliced by the host's splicing factors. However, to produce both partially spliced and full length viral mRNA that can translocate into the cytoplasm, the virus has to suppress the host's nuclear retention mechanism. This is achieved by the Regulator of Virion (Rev) viral protein, which binds to the Rev Response Element (RRE) and enhances the export of both types of viral mRNA from the nucleus.

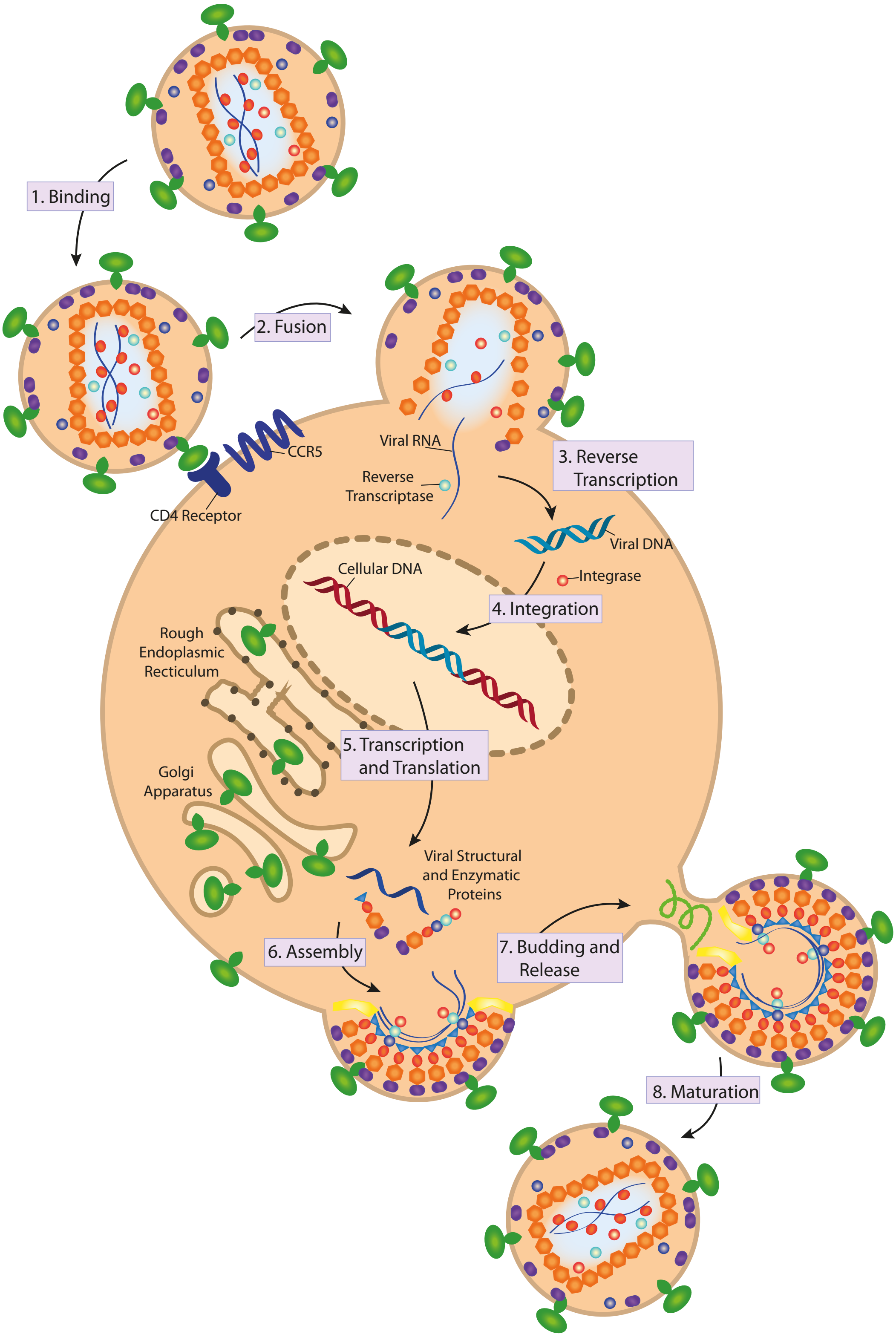


Figure 2: Schematic Representation of the HIV Lifecycle. The HIV lifecycle can be described in eight stages: 1) binding to the CD4$^+$ cell; 2) fusion and release of the viral material into the cell; 3) reverse transcription of the viral RNA by the viral enzyme reverse transcriptase into DNA; 4) integration of viral DNA into the host's cellular DNA by the viral enzyme integrase; 5) transcription and translation of viral proteins; 6) assembly and organisation of the viral proteins at the plasma membrane; 7) budding of the immature virion from the plasma membrane; 8) maturation into the infectious virion by the viral enzyme protease cleaving the precursor Gag polyprotein to form structural proteins resulting in the mature virion.

The Gag and Gag-Pol polyproteins are translated on cytosolic polysomes, and it is here that the N-terminal MA domain is myristoylated by cellular N-myristoyl transferase[15]. The Env polyprotein is translated in the rough endoplasmic reticulum.

After translation, the Env protein is assembled and translocated to the plasma membrane via the cellular secretory pathway. The intracellular tail of Env interacts with the MA domain of Gag to ensure its incorporation into the forming virions[16]. The two copies of viral RNA dimerise through the formation of a 'kissing-loop' structure, which is vital for RNA packaging through interactions with the two copies of zinc finger motifs in the NC domain of Gag[17].

After the Gag and Gag-Pol polyproteins are translated, they translocate to the plasma membrane. Although the exact pathway for this trafficking is uncertain and suggestions of cytoskeleton mediated transport are unconfirmed; Hübner *et al.* reported a diffusion of GFP-Gag fusion proteins through the cytoplasm to reach the plasma membrane[18]. The myristoylated N-terminal MA domain of Gag contains a basic region which allows binding to a plasma membrane specific lipid called phosphatidyl inositol (4,5) biphosphate ($PI(4,5)P_2$)[19]. When the MA domain binds, the myristoyl group becomes exposed and subsequently anchors the Gag polyprotein by inserting into the lipid bilayer of the plasma membrane[20]. Gag then induces the recruitment of cholesterol and sphingolipids to form enriched microdomains called lipid rafts that serve as a target for viral assembly[21]. After the Gag and Gag-Pol polyproteins assemble at the plasma membrane, they multimerise due to interactions within the CA domain that forms an immature lattice, containing irregular shaped defects to allow the membrane to curve[22]. However, to achieve budding from the host cell, the virus requires the use of the host's Endosomal Sorting Complexes Required for Transport (ESCRT) machinery, which are involved in the formation of multivesicular bodies (MVB)[12]. The viral late domain, p6, recruits ESCRT-I to the cell membrane through an interaction with its Tumour Suppressor Gene 101 (TSG101) site[23-24]. Then, ESCRT-III is recruited, which is involved with the scission and egress of the nascent virion[25]. The TSG101-p6 interaction and viral budding will be discussed in more detail in Section 1.3.

The final stage of the HIV lifecycle is the maturation of the immature, non-infectious, viral particles; this occurs concomitant to or just after budding and is highly regulated by the viral enzyme, PR[26]. The process by which PR is activated to form the homodimer is unknown however, this enzyme sequentially cleaves Gag at five specific sites[26] and Gag-Pol to yield the structural components MA, CA and NC as well as the enzymes PR, IN and RT[12, 27]. The cleaved components then reassemble to form infectious mature virus particles. The PR homodimer is an attractive drug target, as mutations that affect its activity can prove detrimental to viral activity[28].

1.3 Recruitment of the Endosomal Sorting Complexes Required for Transport by the Viral Gag Polyprotein

1.3.1 The ESCRT Machinery

The ESCRT machinery is comprised of four complexes (ESCRT-0, ESCRT-I, ESCRT-II, and ESCRT-III), Figure 3, that work in combination with the ATPase Vps4 to sort ubiquitinated membrane proteins into MVBs for lysosomal degradation[29-30]. The ESCRT machinery complexes are involved in many processes due to their ability to mediate membrane scission events. The first of the ESCRT machinery, ESCRT-0, is a heterodimer composed of the hepatocyte growth factor-regulated tyrosine kinase substrate (Hrs) and Vps27. ESCRT-0 localises to the endosome membrane through an endosome enriched lipid that then recruits ESCRT-I through a Pro-Ser-Ala-Pro (PSAP) motif in Hrs. ESCRT-I recruits ESCRT-II which in turn recruits the ESCRT-III subunits, although ESCRT-I can also directly interact with ESCRT-III[31]. The first three members of the ESCRT machinery have at least one domain each that can interact specifically with ubiquitin and serve as receptors for ubiquitin-tagged proteins. ESCRT-III is comprised of four core subunits (CHMP2, 3, 4 and 6) and three accessory components, which upon recruitment to the membrane multimerise to form filaments that constrict and cause the fission of the MVBs.

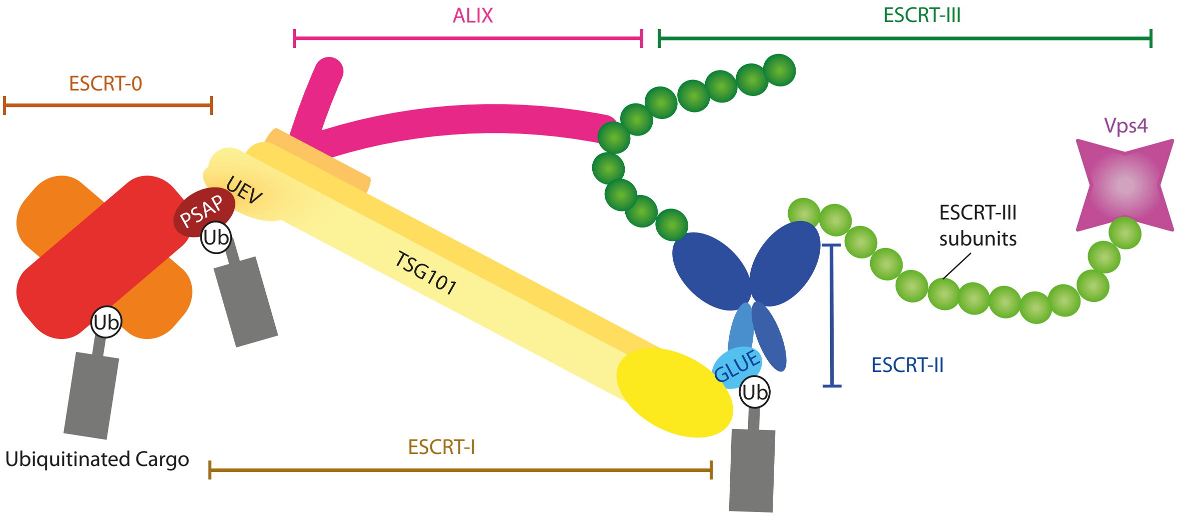


Figure 3: A Simplified Depiction of the Interactions of the ESCRT Machinery. It is comprised of ESCRT-0, ESCRT-I, ESCRT-II, and ESCRT-III. They also interact with the Bro1 domain containing protein ALIX and the ATPase VPS4.

The ESCRT machinery is vital to the homeostasis of the cell, through regulation of growth factor receptors and degradation of transmembrane and misfolded proteins. A key role of the ESCRT machinery is the targeted degradation of growth factor receptors, one of which is the epidermal growth factor receptor (EGFR), belonging to the ErbB family of cell surface receptor kinases. EGFR plays an important role in normal function of the cell as well as pathogenesis of human cancers[32] and is activated by a number of growth factors, including epidermal growth factor (EGF). This causes autophosphorylation of the receptor through activation of the EGFR tyrosine kinase activity and subsequently internalisation into the cell. The receptor is then either recycled back to the plasma membrane if it has not been ubiquitinated or is not bound to a ligand, or it is endocytically trafficked to the lysosome by the endocyctic machinery for degradation[33]. The ESCRT machinery is vital for regulation of EGFR, as it binds to ubiquitinated EGFR and packages it into MVBs to be trafficked to lysosomes.

1.3.2 Recruiting the ESCRT Machinery for Successful Viral Budding

As mentioned in Section 1.2, the recruitment of the host's ESCRT machinery is vital for successful viral budding[23]. Gag recruits the ESCRT machinery through two motifs in its late domain, p6. The first is a Tyr-Pro-X_n-Leu (YPX_nL, where X_n can be any number or type of amino acids) motif that binds to a Bro1 domain containing protein called Apoptosis Linked gene-2-Interacting protein X (ALIX) that interacts with ESCRT-I and ESCRT-III in the MVB pathway[34], Figure 3. Mutating YP to SR has been found to reduce the binding affinity of p6 to ALIX more than 15-fold, and demonstrated a moderate (approximately 3-fold) reduction in HIV particle release[35].

The second motif is a Pro-Thr-Ala-Pro (PTAP) sequence that binds to the TSG101 subunit of ESCRT-I[36-37]. This motif mimics the PSAP motif in Hrs, which recruits ESCRT-I and the rest of the ESCRT machinery to the endosomal membrane[38]. The PTAP motif in p6 binds specifically to a bifurcated groove in the ubiquitin E2 variant (UEV) domain of TSG101[39], Figure 4A; this motif has been found to have significant interactions with key residues in this groove of UEV, Figure 4B.

The interaction between p6 and TSG101 has been shown to be pivotal to viral budding; in a study investigating the effect of mutations on the p6 domain of Gag, Gottlinger *et al.* found that deletion of p6 from the HIV-1 Gag precursor (Pr55Gag) resulted in an accumulation of virus particles tethered to the cell membrane[40]. Garrus *et al.* also investigated inhibiting viral budding by using a small interfering RNA (siRNA) to deplete TSG101, which successfully arrested HIV budding[37]. Mutations in residues of either the PTAP motif of p6 or the TSG101 domain of ESCRT-I were also found to disrupt the TSG101-PTAP interaction[39]. Interestingly, when the PTAP motif was mutated to Leu-Ile-Arg-Leu (LIRL), this resulted in an over 100-fold reduction in HIV particle however, when intracellular concentrations of ALIX were increased, the reduction in particle release was moderately reversed[35]. This indicates that although the PTAP motif is more important for successful budding, the recruitment of ALIX also plays a significant role.

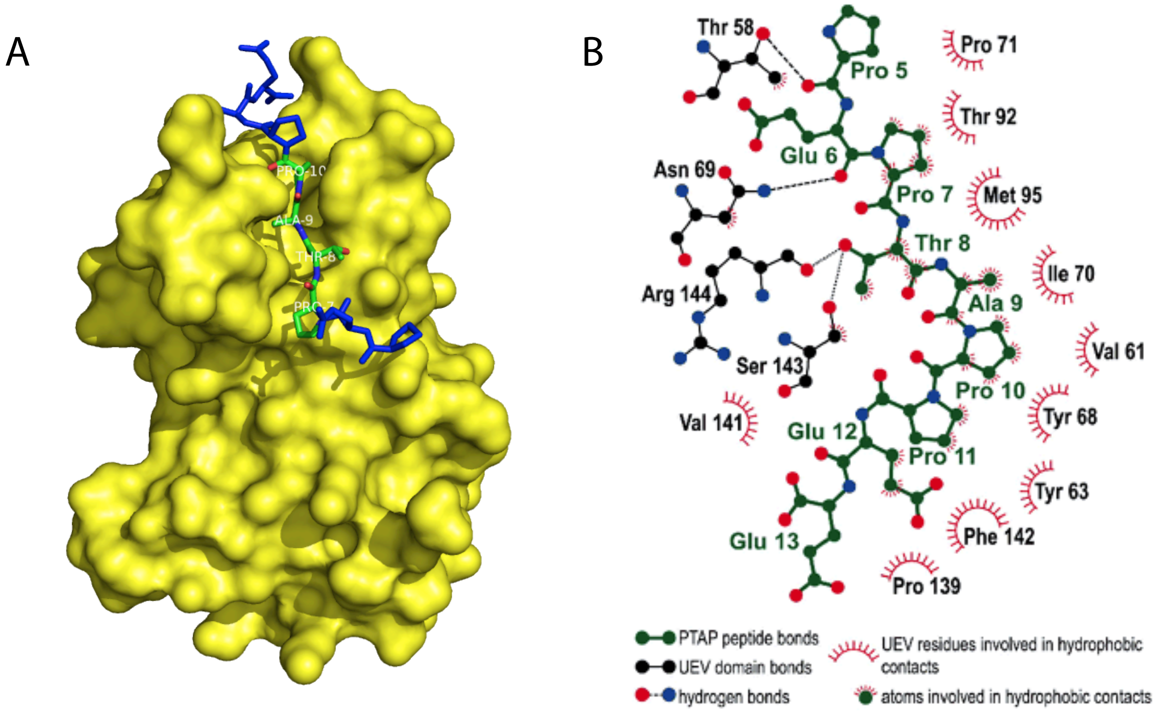


Figure 4: Interaction Between the UEV Domain of TSG101 and a Fragment of the p6 Domain of Gag. A. The UEV domain of TSG101 is shown in yellow and the p6 domain of Gag is shown in blue, with the PTAP residues labelled and coloured according to atom (C – green, N – Blue, O – Red). PDB: 3OBU[41]. B. Close up schematic of the key interactions between p6 and UEV, taken from Pornillos *et al.*[39].

It has been determined that not all members of the ESCRT machinery are required for viral budding. The p6 late domain of Gag recruits ESCRT-I or ALIX which in turn recruits ESCRT-III and Vps4, bypassing ESCRT-II[23]. ESCRT-III refers to a family of proteins with similar structures and features, but out of this family only two proteins (CHMP2 and CHMP4) are necessary for viral budding[42].

Once ESCRT-I has been recruited to the plasma membrane, the virion induces a curvature of the plasma membrane[22] to form a 'neck' caused by packing defects in the Gag polyprotein lattice. Upon recruitment by ALIX and ESCRT-I, CHMP2 begins to polymerise into filaments within the virion 'neck', that interact with CMP4 proteins[42] and the plasma membrane through basic amino acid residues, Figure 5. The CHMP proteins form flexible, helical spirals in the membrane 'neck' that gradually constrict until membrane scission occurs[43]. This scission is aided by the recruitment of Vps4. Although the exact role of Vps4 in viral budding is unknown, it is speculated that it may drive the budding process by destabilising intermediate structures[44]. After budding has occurred, Vps4 disassembles the ESCRT-III protein filaments by ATP hydrolysis[44].

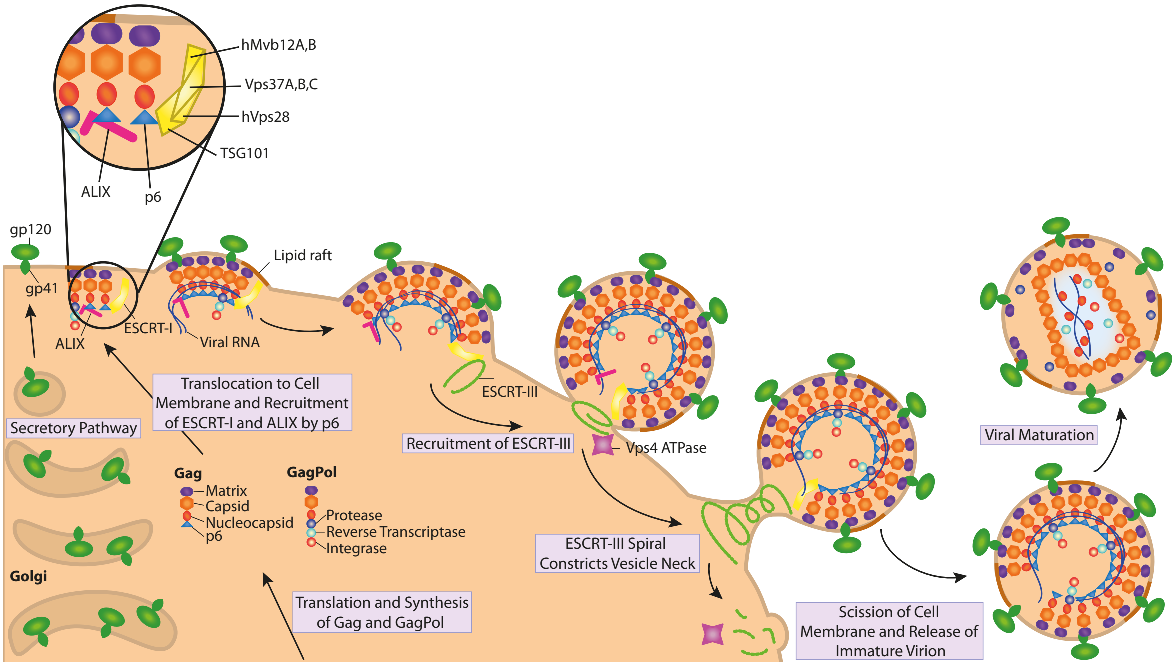


Figure 5: Schematic of HIV Budding Mechanism. The Env proteins are transported to the plasma membrane via the cellular secretory pathway. The Gag and Gag-Pol polyproteins translocate to the plasma membrane where the myristolated MA domain binds to the plasma membrane. The Gag and Ga-Pol proteins multimerise to form an imperfect lattice that causes curvature of the membrane. The p6 domain of Gag recruits ALIX and ESCRT-I through YPX$_n$L and PTAP motifs, respectively, which in turn recruit ESCRT-III proteins. The ESCRT-III proteins polymerise to form helical filaments in the forming ‘neck’ and gradually constrict. Vsp4 is recruited by ESCRT-III and after scission has occurred, recycles the ESCRT-III proteins by ATP hydrolysis. Adapted from Freed[45].

1.4 Treatments for HIV

Since its discovery, it is estimated that HIV infection has claimed more than 35 million lives worldwide. In 2016, there was approximately 36.7 million people worldwide living with HIV, of which 1.8 million had been newly infected[46]. There is currently no cure for HIV, however significant progress has been made with potential vaccines[47-49] and pre-exposure prophylaxis (PrEP) has been shown to effectively block transmission of HIV to at risk individuals[50]. Current treatment options involve the use of combinatorial antiretroviral therapy (cART) consisting of three or more antiretroviral (ARV) drugs that can manage the disease by controlling viral replication and thereby slowing the disease progression however, of the 36.7 million people living with HIV, only 19.5 million people were receiving cART in 2016[46].

There are five major types of ARV drugs on the market, which are grouped by how they interfere with the HIV life cycle: entry inhibitors, fusion inhibitors, RT inhibitors (of which there are two types: nucleoside/nucleotide reverse transcriptase inhibitors [NRTIs] and non-nucleoside reverse transcriptase inhibitors [NNRTIs]), integrase inhibitors and protease inhibitors. The entry inhibitor, Maraviroc (Celsentri) is an allosteric modulator of the CCR5 receptor that prevents gp120 from binding by inducing a conformational change in the receptor. The fusion inhibitor, Enfurvitide (Fuzeon) is a biomimetic peptide that works by mimicking the HIV fusion machinery and competitively binds to gp41, preventing the formation of the entry pore that allows the capsid to enter the cell. The two types of RT inhibitors both act on RT however, NNRTI's bind allosterically, whereas NRTIs are nucleotide or nucleoside analogues that when incorporated into the viral DNA prevent transcription. Integrase and protease inhibitors both function by targeting their respective enzymes through binding to their enzymatic active sites. Multi-class combination products that combine HIV drugs from two or more other types are being used to overcome resistance to treatment.

Since the identification of HIV as the causative agent for development of AIDS, the European Medicines Agency (EMA) has approved of 22 novel ARV drugs as well as 15 novel cARTs; the Food and Drug Administration (FDA) has approved of 24 novel ARV drugs and 12 cARTs, Table 1. However, it is clear to see that there are classes of ARV drugs that are far less populated than others.

Current cART treatment options are highly effective at suppressing the virus and preventing the development of AIDS, however drug resistance is emerging. In 1993, patients were found to have resistance to zidovudine (Retrovir) only six years after it was introduced as the first ARV drug[51]. HIV drug resistance can be split into two categories: transmitted resistance (where uninfected individuals are infected with a drug-resistant form of HIV) or acquired resistance (where resistance occurs due to drug-selective pressure caused by ART regimens). In 2005, drug resistance rates of up to 19% to some HIV inhibitors[52] and transmitted drug resistance of over 9% were reported[53].

HIV-infected cells have a very short half-life ($t_{½}$ = 1-2 days)[54-55], and combined with the high error rate of reverse transcription of viral RNA into DNA (on average one mutation for each viral genome transcribed)[13-14], leads to a high mutation rate per cycle of infection. Therefore patients have a complex and diverse mixture of quasi-species of HIV, which can differ by one or more mutations. In some cases the substitution of one amino acid can produce high levels of resistance, for example, a single mutation in M184V of the viral RT enzyme leads to high levels of resistance to the NRTI drug, lamivudine (Epivir)[56]. To limit resistance, a combination of ARV drugs from at least two different classes are prescribed in an approach called highly active antiretroviral therapy (HAART) as it is unlikely that resistant variants are resistant to all types of ARV drugs.

Once a person has been diagnosed with a HIV infection, they will need to take cART for the remainder of their lives, even if their blood plasma viral load is undetectable. This is due to the latency of HIV (the ability of the virus to remain dormant in various tissues in the form of reservoirs) that, if cART is stopped, can cause the detectable viral load to rebound. It is due to these reservoirs that a potential cure for HIV is elusive.

Table 1: Approved Novel Antiretroviral Drugs. *Drug only approved by FDA. **Drug only approved by EMA. Details obtained through FDA and EMA websites.

Class of Antiretroviral Drug	Medicine Trade Name	Active Substance
Entry Inhibitor	Celsentri (EMA)/Selzentry (FDA)	Maraviroc
Fusion Inhibitor	Fuzeon	Enfuvirtide
Integrase Strand Transfer Inhibitor	Tivicay	Dolutegravir
	Isentress	Raltegravir
	Vitekta*	Elvitegravir*
NNRTI	Stocrin/Sustiva (FDA)	Efavirenz
	Intelence	Etravirine
	Delavirdine*	Rescriptor*
	Viramune	Nevirapine
	Edurant	Rilpivirine hydrochloride
NRTI	Ziagen	Abacavir
	Emtriva	Emtricitabine
	Epivir	Lamivudine
	Zerit	Stavudine
	Viread	Tenofovir disoproxil fumarate
	Videx	Didanosine
Protease Inhibitor	Reyataz	Atazanavir sulphate
	Prezista	Darunavir
	Vitekta**	Elvitegravir**
	Telzir (EMA)/ Lexiva (FDA)	Fosamprenavir calcium
	Crixivan	Indinavir sulphate
	Viracept*	Nelfinavir mesylate*
	Norvir	Ritonavir
	Invirase	Saquinavir
	Aptivus	Tipranavir

One potential treatment for HIV involves dosing with immune activating agents, to induce the expression of HIV from the latently infected cells. This was hypothesised to cause the death of the infected cells, either by cytopathic effects caused by the virus or by the host's immune system, while the use of cART would prevent the infection of new cells. Unfortunately this method has failed to reduce viral reservoirs and patients that stopped cART suffered from a rapid viral load rebound[57], although this method does have potential. There is only one patient who has been 'cured' of HIV; the Berlin Patient has had undetectable plasma viral load levels since he underwent several rounds of intensive chemotherapy followed by a stem cell transplant from a donor whose CD4 cells contained a polymorphism resulting in the lack of expression of CD4, a receptor used by HIV to fuse with the cell, leading to a resistant to HIV infection[58].

If a patient begins ART as soon as possible after the initial HIV infection, this could prevent the formation of viral reservoirs, and could preserve the patients' immunity. Babies who are born to HIV positive mothers are given ART soon after birth. A baby, called the Mississippi Baby, was given ART for 18 months following birth and demonstrated a viral load below the limit of detection, and did not show signs of infection for 2 years[59]. However, initial hopes that the baby was 'cured' were quashed when the viral load rebounded after 2 years and cART had to be restarted[59].

1.5 Protein-Protein Interactions

Most ARV drugs target the enzymatic processes of the HIV lifecycle, such as reverse transcription, integration and protein cleavage to generate the mature virion, but with rising resistance, alternative modes of action must be investigated. HIV relies on a series of PPI throughout its lifecycle and inhibition of these interactions could be a target for new therapies.

PPIs govern nearly all cellular functions and are often dysregulated in disease; this makes PPIs an attractive target for potential therapeutics. However PPIs are often challenging targets, and were once deemed "undruggable" due to the large surface areas involved in interactions between proteins (typically around 1000-2000 $Å^2$) that were believed to be flat and featureless especially in comparison to the deep cavities of typical drug targets (approximately 300-500 $Å^2$), such as enzymes and G protein-coupled receptors[60]. This makes PPIs difficult to target by traditional small molecule techniques, as the PPI is usually stabilised through the interaction of residues across these large interfaces, rather than deep cavities that have many potential points of contact and do not have a natural substrate.

There are many different types of PPI, from homomeric to heteromeric with differing strengths of affinity that range across nearly six orders of magnitude[61], from picomolar (obligate interactions) to micromolar (non-obligate interactions)[62]. These different types of interaction can be further defined by the nature of the proteins involved in the PPI; globular proteins can interact with other globular proteins, single peptide chains can interact with other peptide chains, or globular peptides can interact with single peptide chains[63]. In some cases, one or both of the PPI binding partners may undergo a conformational change upon binding.

A pivotal moment in understanding PPIs was when it was determined that specific residues on the PPI interface contributed more to the overall binding affinity than others[64]. This was determined by a technique called alanine scanning, whereby individual residues in a PPI interface are substituted for alanine in turn and the change in binding affinity is determined. The residues that are the most important for binding are called 'hot-spot' residues and can be distributed non-contiguously across the whole PPI interface. Hot-spot residues were found to account for, on average, 9.5% of all interfacial residues and in a high number of cases the hot-spot regions were enriched with tryptophan (21%), arginine (13.3%) and tyrosine (12.3%) residues[65]. Hot-spots can cover several sub-pockets that bind to hydrophobic residues on the partner protein, and often the key residues are not continuous in the primary sequence.

It was discovered that the interfaces were much more fluid and convoluted in solution than indicated from crystal structures[66], and in some cases the interfaces have been shown to be adaptive, for example, Arkin *et al.* found that a small molecule bound to interleukin-2 (IL-2), in the same site as the natural receptor, however it was found to bind in a new groove that was not present in the free-structure[67]. This is not an isolated example; other proteins such as B-cell lymphoma-extra large (Bcl-xʟ) and HDM2 have also demonstrated a plasticity in their interfaces upon binding small molecules[68].

The incredibly complex and varied binding interactions that occur between types of PPIs make drug discovery very challenging, as the nature of each PPI might not be fully understood. Potential inhibitors can be designed to bind to hot-spots, although, as multiple hot-spots can contribute to the binding interaction these hot-spots can be located over a large area. Therefore, PPI inhibitors are typically larger than more traditional drugs and do not conform to the Lipinski Rule of Five (MW < 500 kDa, less than 5 hydrogen bond donors, less than 10 hydrogen bond acceptors, log P < 5). PPI inhibitors, on the other hand, usually have a MW > 400 kDa, are more hydrophobic (log P > 4), contain more hydrogen bond acceptors (> 4) and contain more ring structures (usually more than 4 rings per inhibitor)[69].

1.5.1 Discovering Inhibitors of PPIs

PPIs do not have natural ligands and the residues that are responsible for the binding interaction are not always contiguous in the polymer chain, meaning that mimetics may not be an ideal starting point in drug discovery. The first step in developing inhibitors for PPIs is a detailed understanding of the interaction and identification of hot-spots. There are many methods for identifying PPI interacting partners that range from *in vitro* techniques such as coimmunoprecipitation, to *in vivo* techniques such as yeast two-hybrid (where the yeast cell survival is linked to a successful binding interaction), to *in silico* methods that use computational techniques to predict binding partners. Once a PPI has been identified, the structure of the binding partners can be determined through techniques such as X-ray crystallography or NMR and hot-spot residues can be identified, as mentioned previously, by alanine scanning.

Due to the complexities involved with targeting PPIs, the traditional methods for drug discovery may not identify high binding compounds due to the large chemical space that potential inhibitors must occupy. A number of strategies have proven successful at identifying small molecule PPI inhibitors, which can be divided into three main types: screening strategies, design strategies and synthetic strategies.

Screening strategies include the use of large libraries of compounds, which are screened for activity against the chosen target and include high-throughput screening (HTS), fragment-based drug design (FBDD) and virtual screening. HTS is a well-established technique for traditional drug discovery, however typical libraries often include compounds that are very flexible and two-dimensional, which often leads to very low hit rates against PPI targets, but can provide a starting point to developing more potent inhibitors. Several inhibitors have been identified this way, for example an inhibitor of human papilloma virus transcription factor E2-helicase E1 interaction (IC_{50} of 6 nM)[70] and an inhibitor of thyroid hormone receptor-coregulator interaction (IC_{50} 0.58 µM)[71].

FBDD uses a fragment-based library, which is screened against the target to identify weak binders. It is often used in combination with X-ray crystallography or NMR to provide structural information on the site of binding and then the fragments can be linked together to produce highly potent binders. FBDD has proved highly successful at identifying PPI inhibitors, as fragments can be identified that bind to specific sub pockets within the hot-spot region and then linked together[72]. However FBDD can only be used in targets where the structures are known, so that the binding can be optimised through target-ligand structural information.

Virtual screening technologies use a computational screening approach where potential inhibitors are docked to the target protein and ranked based on their calculated affinities. Virtual screening can be very high throughput and a relatively cheap way to screen libraries of compounds, although the challenges lie with modelling the adaptive PPI interfaces[68]. Another computational approach uses a similar strategy to FBDD, called multiple ligand simultaneous docking, wherein many fragments are docked to a target protein to identify potential binders, which can be linked together to create highly potent inhibitors[73].

The design strategy utilises structural information about the PPI interaction to design molecule mimetics of the secondary structural elements (α-helix, β-turn, β-strand) that comprise of the interacting epitopes. The most common structural element in PPIs was found to be the α-helix, which accounts for 62% of the PPIs identified in the Protein Data Bank (PDB)[74] which has since lead to the development of many α-helix mimetics.

The synthetic strategy involves the synthetic creation of libraries with a greater diversity than traditional HTS libraries[63]. These libraries are often inspired by natural products and have inspired many classes of PPI inhibitors. There are also many libraries that utilise macrocycles and macrocyclic peptides due to their favourable properties that will be discussed in more detail in Section 1.5.2.

1.5.2 Cyclic Peptides as Potential PPI Inhibitors

There has been increasing interest in macrocyclic peptides as potential PPI inhibitors over the last decade as they have properties that make them ideally suited to these targets. Cyclic peptides are typically hydrophobic which means that they will be able to interact with the interfacial residues of PPIs, which are more likely to be hydrophobic[65]. Cyclic peptides are also typically much larger than small molecules, which means that they are more suited to cover the large surfaces of the protein interfaces to interact with more hot-spots. Cyclic peptides fall outside of the typical Lipinski rule of five for drug-likeness, however they follow the pattern of other PPI inhibitors, as described in Section 1.5. Another key property of peptides is that they contain a degree of chemical flexibility making them ideal for targeting protein interfaces, which often change depending upon their environment. Peptides can also be designed to mimic key features of the protein interfaces such as peptide loops.

Unfortunately, peptides do have some disadvantages. Peptides typically have very poor oral availability and are rapidly degraded by peptidases and gastric acid[75-76]. However, cyclisation can improve the undesirable characteristics of the peptide by increasing resistance to proteolysis and making them more stable to chemical and thermal degradation. This stability is conferred by the reduced conformational constraint introduced by the cyclisation[77-78]. Cyclic peptides also bind with higher affinity to their targets than their linear counterparts[79-80], as the increased rigidity requires a lower entropy penalty upon association to their target[81]. Some cyclic peptides demonstrate a lack of membrane permeability due to the hydration of their backbone N-H bonds, as to pass through the cell membrane these water molecules must be lost. However, cyclic peptides have greater cell membrane permeability than linear peptides due to the lack of C- and N-terminal charges and the formation of intermolecular hydrogen bonds during the passage through the lipid bilayer[82].

Cyclic peptides can also be a starting point to the development of compounds with improved activity and pharmacokinetic properties, for example peptidomimetics – small protein-like molecules designed to have the same biological effects as their natural product counter parts but with enhanced properties. Most protein-peptide interactions take place between the residue side chains, which means that in most cases the peptide backbone is not essential for binding efficacy. Therefore, the backbone of the peptide can readily be altered as long as the side chains remain in their active conformation. Depending on the method of altering the peptide backbone, a higher degree of rigidity can be imposed on the peptide thus decreasing the flexibility and potentially improving the activity. There are a number of ways of modifying the peptide backbone including: methylating the backbone amide; isosteric replacement of backbone amino groups to form depsipeptides or thiodepsipeptides; replacement of the α-carbon to form azapeptides or azatides; extending the backbone by inserting extra carbon, oxygen or nitrogen atoms; or using completely synthetic backbones designed to mimic protein secondary structures[83]. Although cyclic peptides already have a degree of rigidity imposed by the cyclisation, these backbone modifications could also be used to decrease the flexibility of a linear peptide motif to mimic the active conformation found in the original cyclic peptide.

Another method of developing cyclic peptides can include investigations into improving the peptide sidechains. Firstly, an alanine scan can be completed to determine the residues that are most important for binding activity, much like an alanine scan to determine the hot-spots of PPI interfaces. Once the most important residues have been identified, experiments can be completed into improving the activity by inclusion of unnatural amino acids, D-amino acids, β-substituted amino acids, or substituted aromatic amino acids[84].

1.5.3 Discovering Cyclic Peptide PPI Inhibitors

There is an obstacle to discovering potential cyclic peptide PPI inhibitors; traditional methods such as rational design are not suitable due to the cyclic nature of cyclic peptides so combinatorial library based approaches are more commonly used. The most common libraries are created synthetically or using biological techniques and include: one-bead-two-compound, phage display, mRNA display and *in*ternal pro*tein* (intein) based libraries, Figure 6.

The one-bead-two-compound method of library generation synthetically creates cyclic peptide libraries of approximately 10^8 members using split and pool synthesis[85]. The peptides are gradually synthesised on a bead that is spatially segregated to have two layers, Figure 6A. The cyclic peptide is synthesised on the outer layer of the bead via a linker, while the corresponding linear peptide is synthesised on the inner layer to act as the encoding sequence. The library of cyclic peptides is then screened against the protein of interest and only the cyclic peptide, and not the linear version, will be able to interact with the protein as the protein will be too large to diffuse into the inner layers of the bead. Once the most active peptide has been isolated, the sequence can be identified by sequence determination of the linear peptide by Edman degradation mass spectrometry. Peptide libraries generated by this method can be cyclised by head-to-tail cyclisation, although more recently this has been used to create bicyclic peptides using a central scaffold[86].

In mRNA display, cyclic peptides covalently linked to their corresponding mRNA sequence are created *in vitro* from a DNA library, Figure 6B. Firstly, the library is transcribed and the mRNA sequences are modified on the 3’ end to contain puromycin, as this mimics the 3’ end of an aminoacyl tRNA, the natural substrate of the ribosome[87]. Therefore, the ribosome will translate the mRNA until puromycin on the 3’ end, which due to the similarity to the natural substrate will catalyse the transfer of the peptide sequence to the puromycin of the mRNA. The linear peptides can be cyclised in a number of ways between the peptide side chains and the N-terminus or by disulfide bridge formation, but cannot be head-to-tail cyclised as the C-terminus is bound to puromycin of the mRNA. After cyclisation, the peptides can be subjected to multiple rounds of affinity-based screening against the protein of interest until a hit compound is isolated. The hit can then be identified by reverse transcription and then by PCR sequencing.

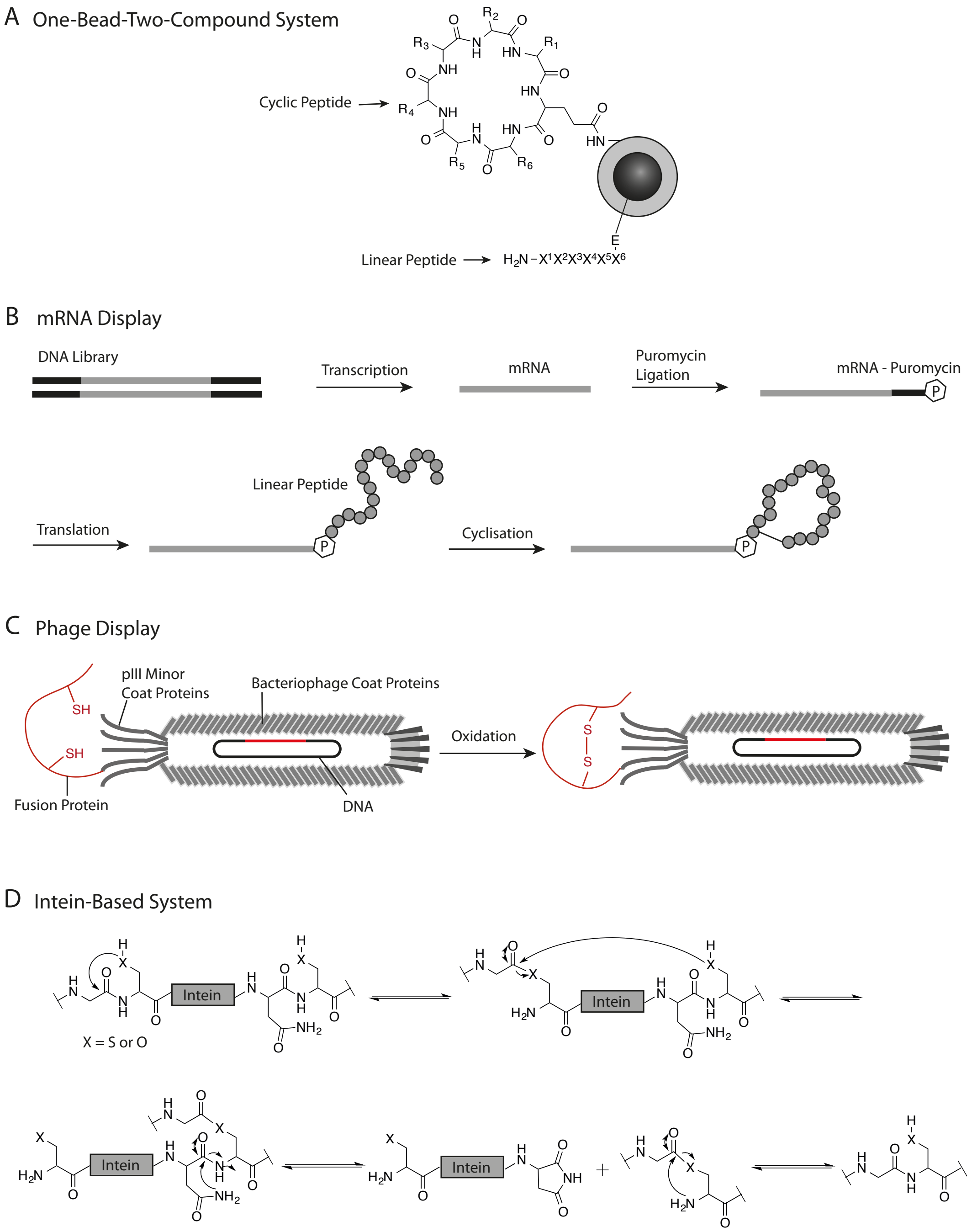


Figure 6: Common Systems for the Creation of Macrocyclic Libraries. A. One-bead-two-compound system whereby the cyclic peptide is presented on the outside of a bead and the corresponding linear sequence is protected in the inner layer of the bead. B. mRNA display. The DNA library is transcribed and the resulting mRNA is ligated to puromycin so that when it is translated the ribosome will catalyse the transfer of the peptide chain to the mRNA. The peptides are then cyclised by post-translational modification. C. Phage display expresses peptides as fusions to the N- or C-terminal coat proteins, which can be cyclised by disulfide bridge formation between two cysteine residues. D. The intein is preceded by the N-extein and followed by the C-extein, which self-excises to produce the extein. Firstly, a thioester bond is formed through an N-S/O acyl shift at the junction between the N-extein and the intein. Then a lariat intermediate is formed by the nucleophilic attack of the thioester bond. The cyclisation of the asparagine residue releases the intein and the extein is formed by a subsequent S/O-N acyl shift.

As mRNA display is an *in vitro* technique, it can easily be adapted to include non-canonical amino acids. Suga *et al.* successfully combined mRNA display with an artificial ribozyme called a flexizyme, which is capable of charging tRNAs with non-canonical amino acids, to create the *F*lexible *I*n-vitro *T*ranslation (FIT) system[88]. When the FIT system was used to incorporate an N^{α}-chloroacetyl-amino acid, the peptides generated spontaneously cyclised through a thioether linkage between a C-terminal cysteine side chain and the unnatural amino acid. The FIT system was then modified further to create libraries of between 10^{12}-10^{14} member macrocyclic libraries by the *Ra*ndom non-standard *P*eptides *I*ntegrated *D*iscovery (RaPID) system[89].

Phage display can generate libraries of up to 10^{9} members, by inserting DNA sequences encoding the desired peptide sequenced into the genome of engineered *E. coli* bacteriophage. The peptides are then expressed as fusion proteins at either the N- or C-terminus of the virion coat proteins, Figure 6C. Typically peptides are cyclised by disulphide bridge formation, although bridged bicyclic peptides have been generated by using *tris*(bromomethyl)benzene as a linker between the side chains of three cysteine residues[90]. Potential hits are isolated through affinity-based screening and then deconvoluted by DNA sequencing. Phage display library size is limited by the bacterial transformation efficiency however, they are not limited to natural amino acids; Tian *et al.* successfully used amber stop codon suppression to incorporate non-canonical amino acids into a phage display library[91].

*In*ternal pro*teins* (inteins) are protein sequences that post-translationally auto-catalyse the splicing of a functionally mature protein sequence from a larger peptide sequence through the cleavage of two peptide bonds and the ligation of the flanking extein (*ex*ternal pro*tein*) to form a new peptide bond[92]. The precursor protein typically takes the form of N-extein – intein – C-extein, which after the intein has excised itself generates the N-extein linked to the C-extein via a peptide bond, Figure 6D. Inteins have been used *in vitro* in combination with synthetic and semi-synthetic precursors to create macrocyclic organic-peptide hybrids (MOrPHs) that can have varying ring sizes[93]. This methodology was further adapted to create bicyclic organo-peptide hybrids (BOrPHs) by including two cysteine residues capable of forming a disulfide bridge[94]. Inteins can also be split into N-intein and C-intein fragments surrounding a central extein, which upon translation can self-excise the extein as a cyclic peptide. This methodology is called Split-Intein Circular Ligation of Proteins and Peptides (SICLOPPS) and will be discussed in more detail in Section 1.5.4.

Each method of creating combinatorial libraries has associated advantages and disadvantages[95] and can be adapted depending upon the desired type of macrocycle.

1.5.4 Split-Intein Circular Ligation of Proteins and Peptides (SICLOPPS) and Reverse Two-Hybrid System (RTHS)

SICLOPPS is a robust and flexible method for the intracellular synthesis of large libraries (up to a 10^8 members)[96] of cyclic peptides. This procedure utilises split-inteins from the cyanobacteria *Synechocystis* sp PCC6803 (*Ssp*) DnaE which are capable of *trans*-splicing and excising the target peptide or protein[97]. Split-inteins are non-contiguous inteins that are transcribed and translated as two separate peptides (N-intein [I_N] and C-intein [I_C]) each fused to an individual extein. In SICLOPPS, the intein structure is altered so that the inteins are both fused to the same central extein, however the inteins are reversed so that the C-intein precedes the N-intein (I_C-extein-I_N). The extein sequence in this case encodes for the cyclic peptide sequence, Figure 7. These intein fragments then spontaneously fold into the active cis-intein upon translation, which undergo a series of rearrangements to generate the target cyclic peptide, Figure 7. In the first step, the enzyme catalyses an N-to-S acyl shift to form a thioester intermediate, which undergoes transesterification at the I_C-end to form a lariat intermediate. The cyclic by-product is liberated first as a lactone before the thermodynamically favoured lactam product is generated by an Z-to-N acyl shift *in vivo*[98].

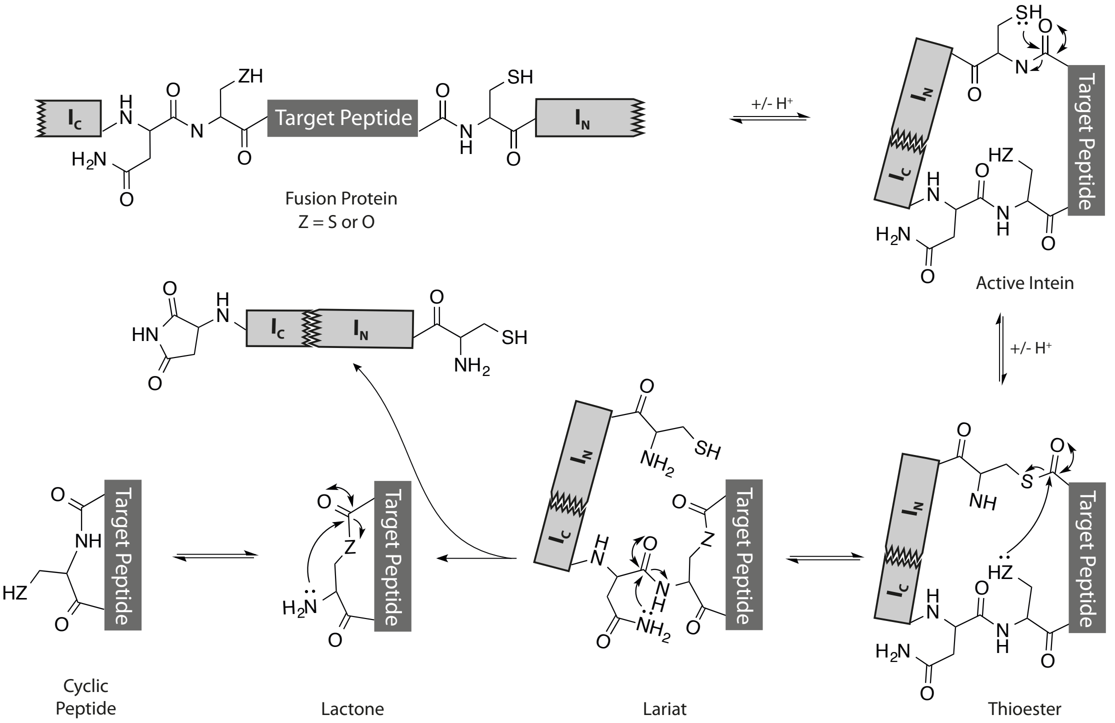


Figure 7: Proposed SICLOPPS Mechanism for Production of Cyclic Peptides. The fusion protein consists of the N-intein [I_N] and C-intein [I_C] flanking the target peptide sequence. This folds to form the active intein, which under goes an N to S acyl shift to produce the thioester. This then undergoes a transesterification reaction with the nucleophilic side chain of the cysteine or serine (Z=S or O) to form the lariat intermediate. The intein is released by a cyclisation reaction of the asparagine side chain and liberates the lactone form of the cyclic peptide, which undergoes a rearrangement to form the desired peptide.

The extein can encode for cyclic peptides of any size, the only requirement is the presence of either a nucleophilic serine or cysteine in the first position to allow for transesterification to form the lariat intermediate[78]. The rest of the sequence can be designed to encode for the incorporation of a random amino acid by using repeats of the NNS codon, where N represents any of the DNA bases (A, C, G or T) and S represents only C or G. This covers 32 codons and all 20 canonical amino acids but means that the ochre (UAA) and opal (UGA) stop codon will be excluded. This minimises the possibility of truncations in the library. Although SICLOPPS plasmids can be designed to encode for cyclic peptides of any size, they are usually restricted to less than six randomised positions to ensure that the total possible number of library members (6.4×10^7 members) will be covered by the number of *E. coli* transformants that can be obtained by standard biological techniques (10^8-10^9). SICLOPPS was initially limited to the 20 canonical amino acids, however in some instances it has been combined with an orthogonal aminoacyl-tRNA synthetase/tRNA_{CUA} pair so that non-canonical amino acids can also be included[99].

SICLOPPS can be used in combination with most cell-based screens, but it has been used most successfully in combination with the bacterial reverse two-hybrid system (RTHS). The RTHS selects against the association of two proteins, the opposite of the initial yeast two-hybrid systems that were used as a tool for studying PPIs *in vivo*. In the yeast two-hybrid system, a pair of proteins would interact to bring the activation and DNA binding domains of a transcription factor together, which initiates the expression of adjacent reporter genes. In the bacterial RTHS, the target proteins are fused to the 434 and P22 bacteriophage DNA binding proteins. If these proteins successfully bind, an active repressor complex will form which can bind to the corresponding DNA-binding sites on the *E. coli* chromosome. This represses the downstream transcription of the essential reporter genes, causing the inhibition of growth on selective media, Figure 8A. If the interaction is inhibited, the transcription and translation of the reporter genes is allowed and the bacteria will grow on selective media, Figure 8B. The reporter genes act as conditionally selective markers: *HIS3* (imidazoleglycerol-phosphate dehydratase, enables the biosynthesis of histidine), Kan^R (aminoglycoside 3'-phosphotransferase, confers kanamycin resistance) and *lac*Z (β-galactosidase, which is used to quantify the PPI through quantitative, colorimetric β-galactosidase assays). When used in combination with SICLOPPS, this methodology provides a streamlined way of screening for inhibitors of PPIs using a phenotypic readout. The isolated hit compounds can be determined by DNA sequencing. To ensure that specific inhibitors are isolated, the plasmids coding for the hit cyclic peptides are transformed into another, unrelated RTHS that is then grown on selective media. If a similar pattern of growth is observed with the unrelated RTHS, it can be inferred that the isolated hit is not a specific inhibitor of the PPI of interest and may be instead preventing the repressor complex from forming or binding to the DNA-binding site and therefore allowing the transcription of the downstream reporter genes. The hit cyclic peptides that have been determined to be specific inhibitors to the PPI of interest can then be taken further for future screening assays and ranked.

The combination of SICLOPPS and RTHS have been used previously to identify cyclic peptide inhibitors of a variety of PPI targets implicated in many disease states, from AICAR transformlyase/IMP cyclohydrolase (ATIC) homodimerisation involved in *de novo* purine biosynthesis[100], to cancer targets such as C-terminal binding protein homodimerisation[101] and hypoxia inducible factor-1 (HIF-1) heterodimerisation[102], to a novel cyclic peptide inhibitor of the TSG101-Gag interaction involved in HIV budding[103].

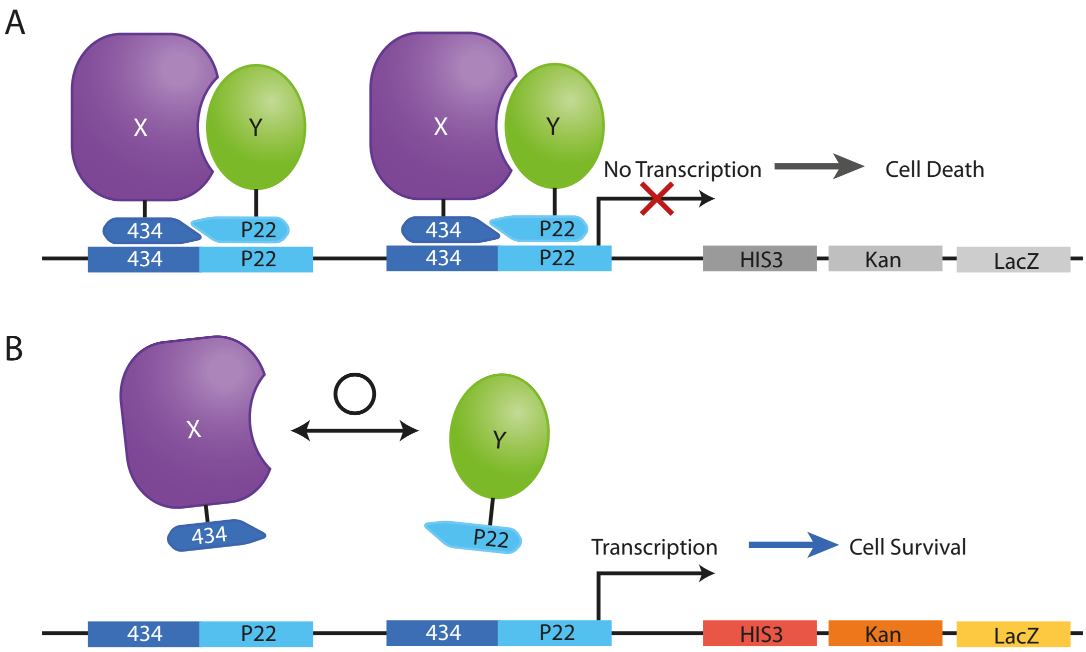


Figure 8: The Reverse Two-Hybrid System. The proteins of interest (X and Y) are fused to the bacteriophage DNA proteins. A. If the two proteins successfully bind the active repressor complex is formed which binds to the corresponding operator sequences and represses the transcription of downstream reporter genes, leading to cell death on selective media. B. If the proteins do not bind due to the presence of an inhibitor, the repressor complex cannot form, downstream genes can be successfully transcribed and cells will grow on selective media.

1.5.5 Identification of an Inhibitor of TSG101-Gag PPI

A SICLOPPS library was constructed to encode for 8-member cyclic peptides of the type SGWXXXXX ($SGWX_5$), where X = any amino acid, to screen for potential inhibitors of the TSG101-Gag PPI. The invariable motif was designed to contain serine (required for intein processing), glycine (to avoid racemisation during chemical synthesis), and tryptophan (as a chromophore for HPLC purification). An alternative library was also designed to investigate if a peptide with a

'PTAP'-like motif (SGWXXPXXPXX), similar to that found in Gag would produce better inhibitors[103]. However, the randomised cyclic peptides that were obtained from the PTAP targeted library were demonstrated to be far poorer inhibitors than the hits identified from the SGWX$_5$ library, Table 2, suggesting that the disruption of the TSG101-Gag interaction is conformationally based, rather than simply acting as a peptidomimetic[103]. It was interesting to note that the top two inhibitors, CP11 and CP6 contained a similar motif in the variable region (YWN in CP11 and NWY in CP6).

Table 2: Ranking of Hit Peptide Sequences Identified Through Screening SICLOPPS Libraries SGWX$_5$ and SGWXXPXXPXX in the p6-UEV RTHS. The invariable motif of each peptide is underlined. Taken from Tavassoli *et al.*[103]

Activity Rank	Name	Peptide Sequence
1	11	IYWNVSGW
1	6	TNWYGSGW
2	16	TLLVYSGW
3	8	VLRVHSGW
3	120	PGPVTPGFSGW
3	112	DGPRGPSTSGW
3	127	PGPCSPVGSGW
3	126	LVPWMPRPSGW
3	119	GCPFPPSYSGW
3	122	ARPNRPCRSGW

The top three peptides (CP11, CP6 and CP16) were found to be specific inhibitors to the p6-UEV interaction, as they did not demonstrate a growth advantage in an unrelated RTHS. Cysteine derivatives of the peptides were chemically synthesised and then tagged with a cysteine-modified HIV Tat peptide to ensure cellular permeability. The Tat-tagged peptides were then tested in mammalian cells in a virus-like particle (VLP) assay to assess the inhibitory activity of each peptide, Figure 9A. This assay utilises the viral Gag protein, which is able to induce the formation of VLPs that are morphologically similar to HIV virions, without requiring any other viral proteins[104], making it ideal for assessing the inhibitory activity. Of the three peptides, only CP11 was found to inhibit VLP release at 10 μM and this inhibitory effect was found to be dose dependent, with an IC$_{50}$ of 7 μM, Figure 9B. This demonstrates the first identification of a cyclic peptide inhibitor targeting a PPI that is essential for viral infection.

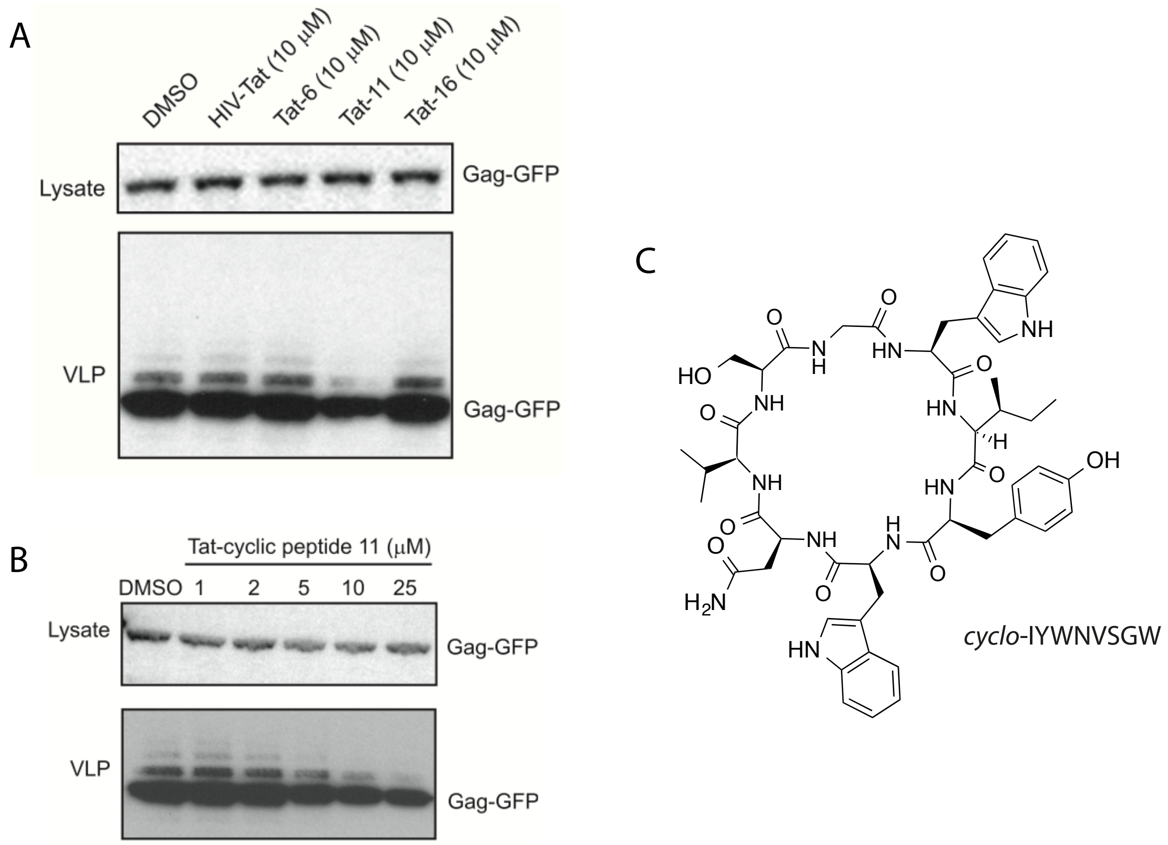


Figure 9: Testing Cyclic Peptides in a Mammalian VLP assay. A. Testing hit peptides CP11, CP6 and CP16 at a concentration of 10 µM. CP11 shows a 60% reduction in VLP release compared to Tat peptide alone. B. Effect of CP11 treatment on VLP release. C. Structure of the cyclic peptide inhibitor 11 (CP11), *cyclo*-IYWNVSGW. Figures A and B taken from Tavassoli *et al.*[103].

1.6 Project Aims

Although current ARTs are highly effective at suppressing the progression of HIV infection, there are a growing number of cases of drug resistance emerging. The interaction between TSG101 and the viral Gag polyprotein is pivotal to the HIV lifecycle and a currently unexploited target for treating HIV infection. A cyclic peptide inhibitor of TSG101-Gag (CP11) was identified through a combination of SICLOPPS and a RTHS and is a first-in-class inhibitor for this interaction. The aim of this project was to investigate small peptidic motif in CP11 as the potential starting point of developing small molecule inhibitors of the UEV-p6 interaction.

Two different approaches are going to be taken as starting points for developing small molecule inhibitors; firstly, a repeating motif that was identified in the top two hits determined by SICLOPPS/RTHS will be investigated for inhibitory activity as a small peptide fragment. Secondly, CP11 will be investigated through alanine scanning to see if the active motif identified is different to the repeating motif.

Once an active motif has been determined, methods into improving the activity via incorporation of unnatural amino acids will be investigated and experiments with biophysical assays to determine the binding partner of the PPI will be completed. Then if a more potent inhibitor is developed, to test its inhibitory activity in a cell-based assay.

Chapter 2 Investigation into Small Motifs of CP11

2.1 Introduction

The top two peptide inhibitors of the p6-UEV interaction identified in the SICLOPPS screen[103] were found to contain a similar motif; CP11 contained the amino acid motif YWN and in CP6 this motif was reversed as NWY. It was deemed significant that the top two peptides out of the 3.2 × 10^6 members of the library screened contained a similar motif and so these tri-peptides, YWN and NWY, were identified as the potential starting point for the development of small molecule-like inhibitors of the p6-UEV interaction.

In order to test the inhibitory activity of the peptides, an Enzyme-Linked Immunosorbent Assay (ELISA) was developed.

2.2 Development of an Enzyme-Linked Immunosorbent Assay (ELISA)

An ELISA is a solid-phase assay that uses one or two antibodies to detect the presence of a substrate. The term was first coined in 1971[105], and has since has been applied to many immunodiagnostic tests for different diseases, including HIV, malaria and Lyme disease, as well as other conditions that cause the presence of antigens or antibodies in the blood, such as food allergies and coeliac disease.

ELISAs have also been used to determine potential inhibitors of PPIs, where the primary protein of interest, in this case, UEV is tagged using a polyhistitidine (His) tag and added to a 96-well plate coated with activated nickel, to which the His tag binds, Figure 10A. Any unbound protein is washed off the plate and then the second protein of interest, in this case, p6 tagged with a glutathione S-transferase (GST) tag is added alongside a potential ligand inhibitor. If the ligand, binds to either of the proteins of interest and disrupts the interaction between them, the second protein will be washed off the plate. A primary antibody, in this case a mouse anti-GST antibody, which will bind to any bound GST-tagged protein is added to the plate. Any unbound antibody is washed off the plate and a secondary antibody, sheep anti-mouse, conjugated to horseradish peroxidase (HRP) is added which will recognise any mouse anti-GST antibody. The presence of sheep anti-mouse peroxidase will catalyse the oxidation of 3,3',5,5'-tetramethylbenzidine (TMB) to form the cation-radical, which is in equilibrium with the charge transfer complex and has λ_{max} = 370, 652 nm[106-107], resulting in the development of a blue solution, Figure 10B. To stop the reaction, sulfuric acid is added which further oxidises the cation-radical of TMB to form the diimine derivative, which is yellow in colour and the absorbance at 450 nm can be measured using a plate reader.

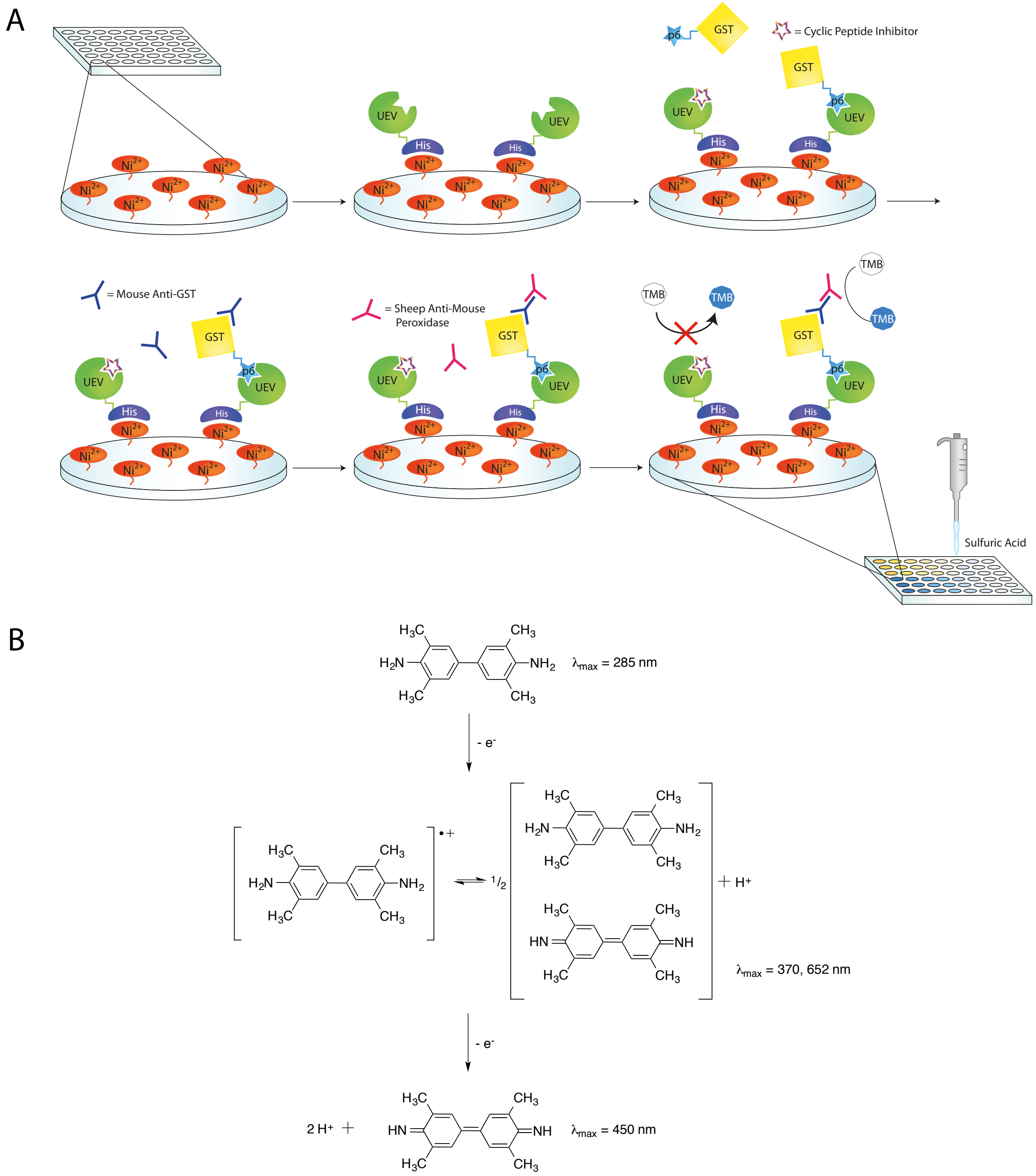


Figure 10: A. General Scheme for an ELISA investigating Inhibitors of the UEV-p6 PPI. B. Oxidation Mechanism of TMB.

2.2.1 Recombinant Protein Expression

In order to investigate potential inhibitors of the p6-UEV PPI, it was necessary to express and purify the two proteins of interest, so that an ELISA to assess novel inhibitors could be developed. The proteins required different affinity tags in order for the interaction to be studied using an ELISA so that one tag could be used to bind to an activated surface in a plate and the other to be detected by an antibody. Affinity tags also aid in the purification of proteins from cellular lysate. High-protein binding plates, which would not require the proteins to be tagged, could have been utilised, however, it was chosen to use plates with an activated nickel surface in conjunction with a His-tag for specificity.

The gene encoding human UEV was cloned into the bacterial expression vector, pET28a (Novagen) using the *NheI* and *XhoI* restriction sites, Figure 11. The plasmid was sent for sequencing to ensure that the gene had successfully been incorporated without any mutations.

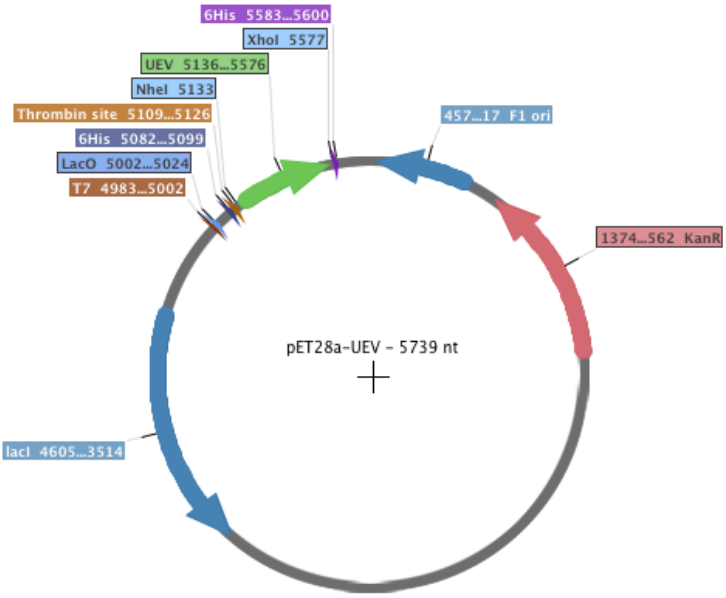


Figure 11: Plasmid Map of pET28a-UEV.

The plasmid was then transformed into chemically competent BL21 cells and from these cells a culture was grown and induced to express the UEV protein by the addition of IPTG (0.1 mM) then incubated overnight at 16°C. The following morning the culture was harvested and frozen at -80°C prior to purification. The pellet was thawed and resuspended in lysis buffer (20 mM Tris-HCl, 100 mM NaCl, 10% glycerol, pH 7.4, supplemented with 1 % (v/v) Triton X-100, Pierce Protease Inhibitor tablets (ThermoFisher Scientific, 1 tablet per 20 mL) and lysozyme (100 µg/ mL)). The lysate was sonicated (12 cycles of 10s on followed by 10s off) on ice to ensure complete lysis. Insoluble cell debris was pelleted via centrifugation at 22,700 × g for 45 min at 4°C. The supernatant was filtered using 0.45 µm syringe filters prior to purification by fast-protein liquid chromatography (FPLC) using a HisTrap column (GE Healthcare), Figure 12. The protein was loaded onto the column at a flow rate of 1 mL/min. The column was washed with 30 mL of Buffer A (20 mM Tris-HCl, 100 mM NaCl, 10% glycerol, pH 7.4) and then with a gradient of 8% Buffer B (500 mM imidazole, 20 mM Tris-HCl, 100 mM NaCl, pH 7.4) in Buffer A at a flow rate of 1 mL/min, then the protein was eluted in 100% Buffer B. The fractions containing His_6-UEV were collected and dialysed overnight at 4°C using SnakeSkin™ Dialysis Tubing 10K MWCO 22 mm (ThermoFisher Scientific) into Buffer A (1 L/ 2 mL protein). The buffer was exchanged for fresh and allowed to dialyse for another 3 hours. The final protein concentration of approximately 1 mg/mL for each purification was determined by NanoDrop analysis (approximately 80% purity).

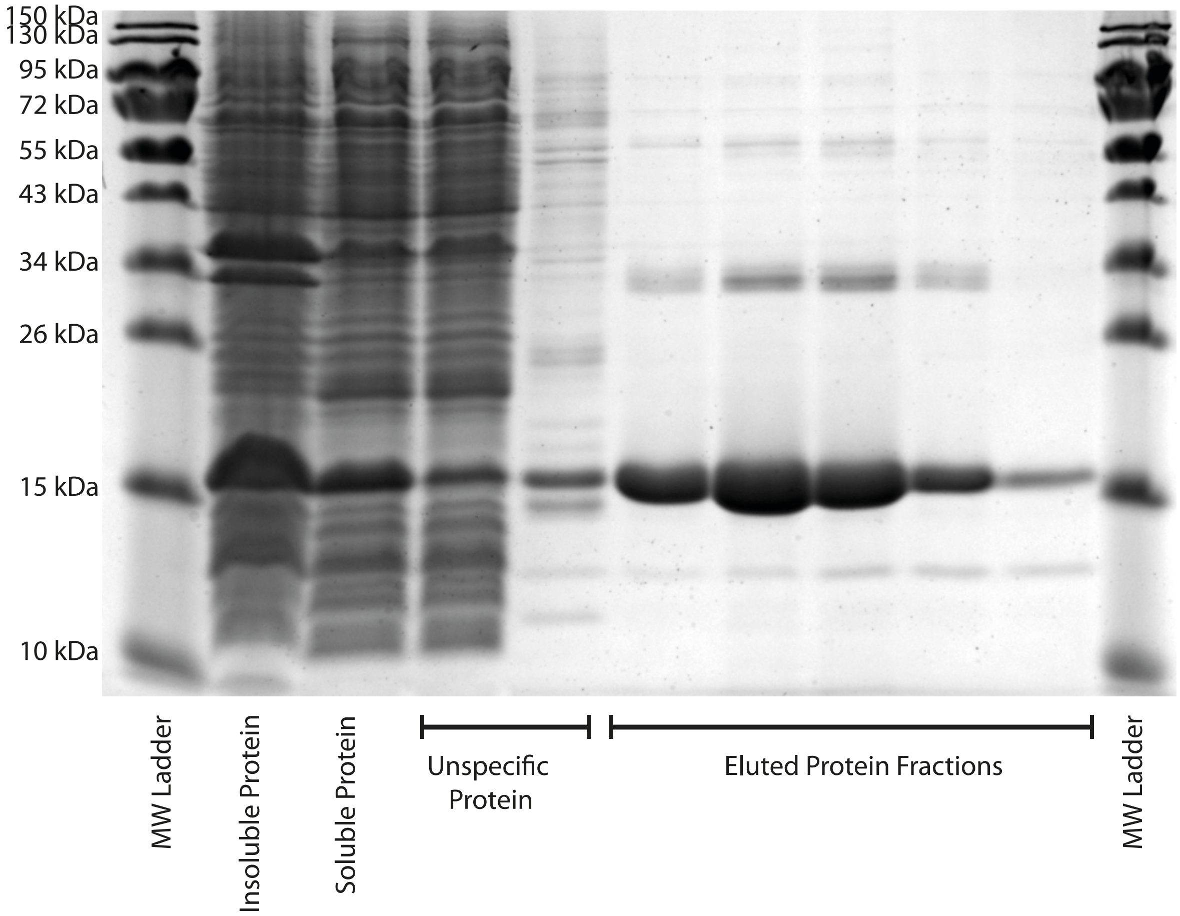


Figure 12: Representative SDS Gel of Fractions Collected During Purification of His$_6$-UEV by FPLC using a HisTrap Column. The insoluble fraction was resuspended bacterial contents that remained after centrifugation. The soluble protein was a fraction of the lysate taken after filtration. The unspecific protein fractions were the fractions of protein that flowed directly from the column. The eluted fractions were collected after elution from the column in buffer containing imidazole (500 mM). The His$_6$-UEV protein was visualised on a 10% SDS gel.

The viral p6 gene was cloned into the bacterial expression vector, PGEX-2TK (GE Healthcare), using the *EcoRI* and *BamHI* restriction sites, Figure 13. Successful incorporation into the plasmid was confirmed by DNA sequencing.

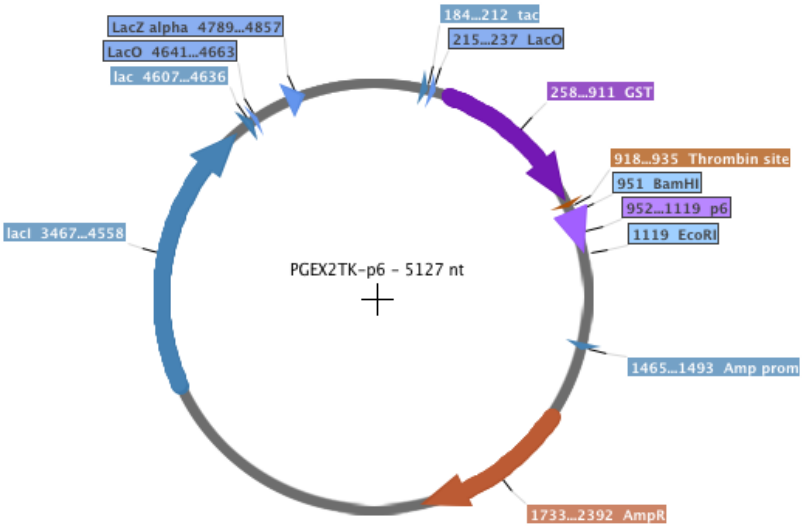


Figure 13: Plasmid Map of PGEX2TK-p6.

The plasmid was transformed into chemically competent BL21 cells and a culture was grown and induced to express the recombinant protein by the addition of IPTG (0.1 mM) then incubated overnight at 16°C. The following morning the culture was harvested and frozen at -80°C prior to purification. The pellet was thawed and resuspended in lysis buffer. The lysate was sonicated (12 cycles of 10s on followed by 10s off) on ice to ensure complete lysis. Insoluble cell debris was pelleted via centrifugation at 22,700 × g for 45 min at 4°C. The supernatant was filtered using 0.45 µm syringe filters prior to purification by FPLC using a GSTrap column (GE Healthcare), Figure 14. The column was pre-equilibrated in Buffer A at a flow rate of 1 mL/min. The protein was loaded onto the column at a flow rate of 0.5 mL/min. The column was washed with 30 mL of Buffer A, and the GST-p6 was eluted with 10 mL GST Elution buffer (50 mM Tris-HCl, 150 mM NaCl, 15 mM reduced glutathione, pH 7.4) at a flow rate of 1 mL/min. The fractions containing the desired protein were combined and dialysed overnight at 4°C using SnakeSkin™ Dialysis Tubing 10K MWCO 22 mm (ThermoFisher Scientific) into Buffer A (1 L/ 2 mL protein). The buffer was exchanged for fresh and allowed to dialyse for another 3 hours. The final protein concentration of approximately 1 mg/mL for each purification was determined by NanoDrop analysis (approximately 75% purity).

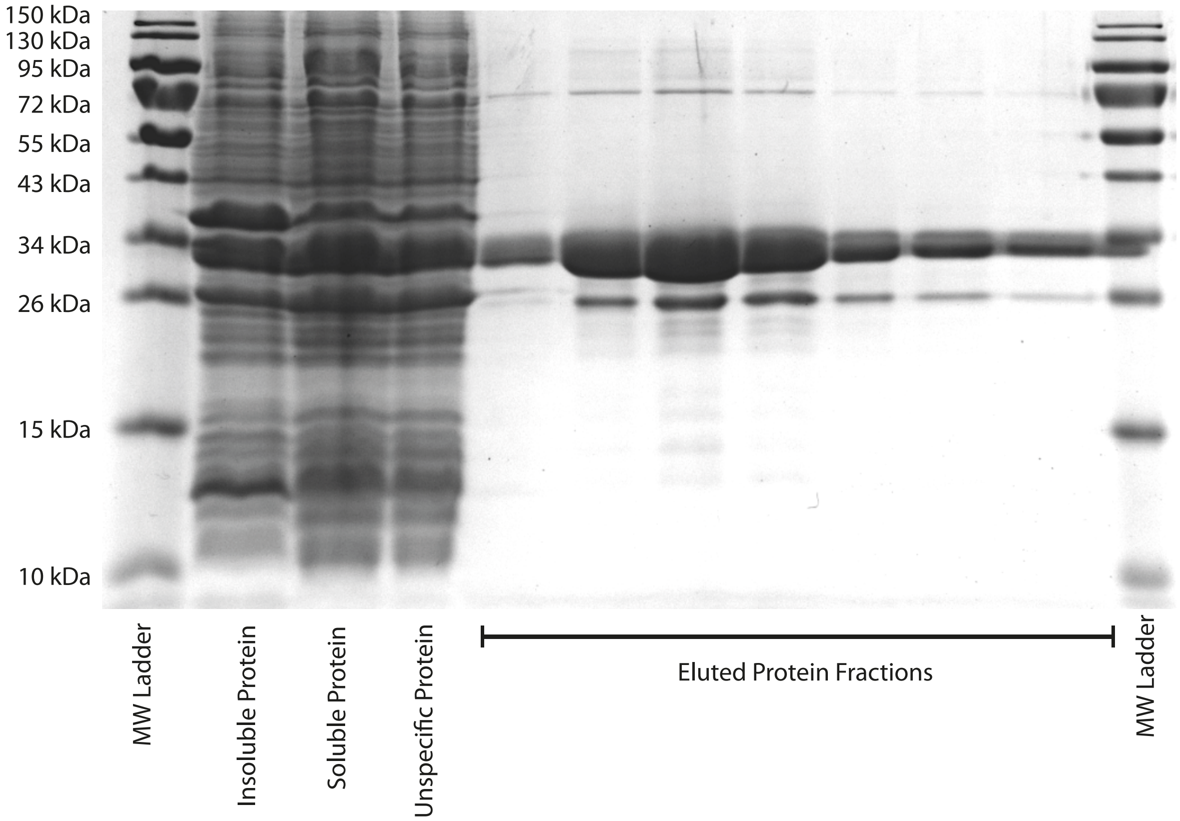


Figure 14: Representative SDS Gel of Fractions Collected During Purification of p6-GST by FPLC using a GSTrap Column. The insoluble fraction was resuspended bacterial contents that remained after centrifugation. The soluble protein was a fraction of the lysate taken after filtration. The unspecific protein fractions were the fractions of protein that flowed directly from the column. The eluted fractions were collected after elution from the column in buffer containing reduced glutathione (15 mM). The p6-GST protein was visualised on a 15% SDS gel.

2.2.2 Determining the Concentration of Protein for ELISA

It was important to determine the ideal concentration of His$_6$-UEV to be used in the ELISA. The concentration needed to be sufficient to observe a good signal above any background, but not excessive to use a wasteful amount of protein. Increasing concentrations of His$_6$-UEV (0- 1.8 µM) were bound to the plate and then detected with a mouse anti-His antibody, Figure 15. A good signal was observed at and above 1 µM, so for further ELISAs this concentration of UEV-His (1 µM) was chosen.

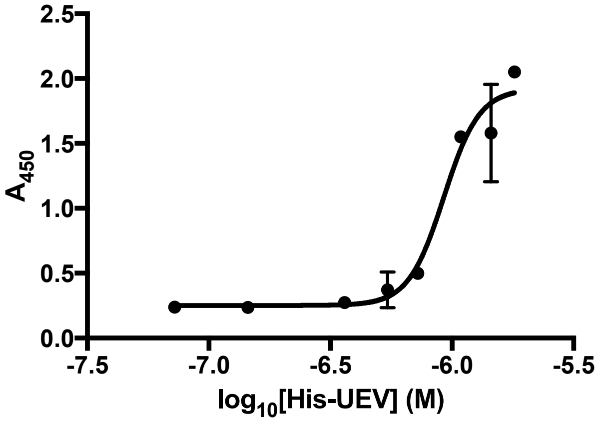


Figure 15: Optimisation of Concentration of His$_6$-UEV to be used in ELISA. Error bars represent the standard deviation of two repeats. Curve was fitted using GraphPad Prism7 Software using a Nonlinear Sigmoidal, 4PL curve fit where X is log(concentration of peptide).

2.2.3 Assessing Binding Specificity

To ensure that p6-GST was not binding non-specifically to the nickel-coated plates, a blocking step was investigated. Without previously binding His$_6$-UEV to the plate, 2% milk solution (w/v in deionised water) or PBS Buffer containing the detergent TWEEN-20 was added to the plate and incubated for 1 hour at room temperature. The milk solution or buffer was removed from the wells and then increasing concentrations of p6-GST were added. The presence of p6-GST was detected with an anti-GST antibody, Figure 16.

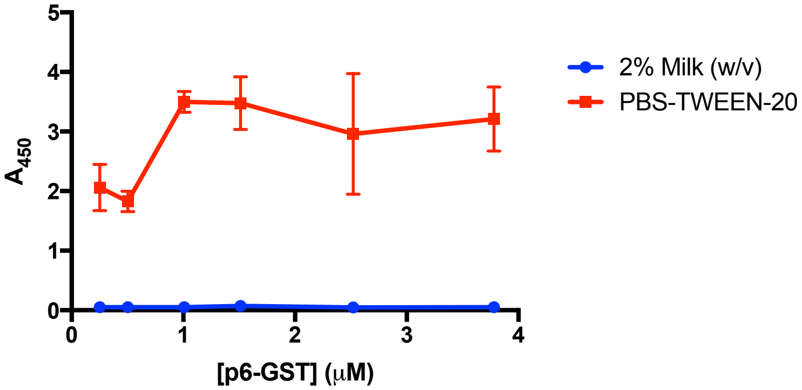


Figure 16: ELISA Investigating a Blocking Step to Prevent Non-Specific Binding of p6-GST to the Nickel-Coated Plate. The error bars represent the standard deviation of two repeats.

From this ELISA it can be seen that there is a high level of non-specific binding of p6-GST to the nickel-coated plate, even in the presence of low concentrations of detergent. However, this non-specific binding is almost completely prevented by previously incubating the wells with 2% milk solution (w/v) prior to the addition of the p6-GST protein.

Next the binding specificity of p6 to UEV was assessed as the relative size of the p6 protein is small when compared with the size of the GST tag (5.92 kDa versus 27.70 kDa, respectively), it was important to confirm that GST was not binding non-specifically to His_6-UEV within the assay. Increasing concentrations of p6-GST and GST were incubated with His_6-UEV and the assay completed as previously described, Figure 17. It can be seen that p6-GST has a higher affinity for His_6-UEV than GST alone, indicating that the binding interaction is specific between p6 and UEV.

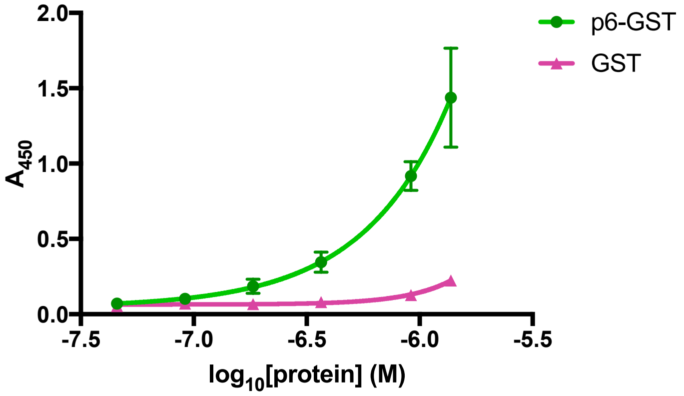
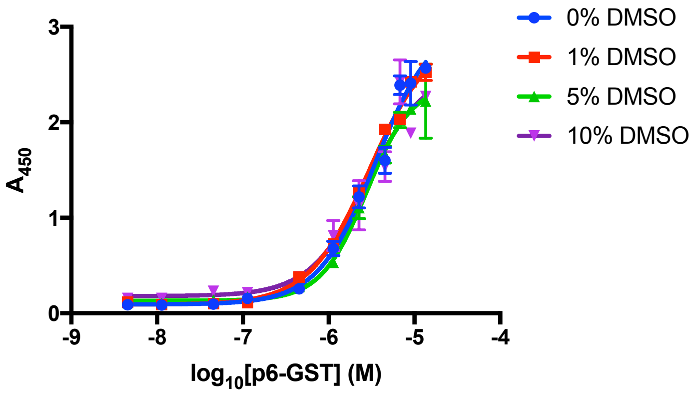


Figure 17: ELISA Investigating the Non-Specific Binding of GST versus p6-GST. The error bars represent the standard deviation of two repeats. Curve was fitted using GraphPad Prism7 Software using a Nonlinear Sigmoidal, 4PL curve fit where X is log(concentration of peptide).

2.2.4 Testing the Effects of Increasing Concentrations of DMSO

As some peptides can be sparingly soluble in aqueous buffers at high concentrations, but are soluble in DMSO, it was important to determine the DMSO tolerance of the ELISA as this reagent is known to negatively affect protein stability[108-109]. An ELISA was completed with increasing concentrations of p6-GST along with increasing concentrations of DMSO (up to 10% (v/v)) incubated with a constant concentration of plate-bound His_6-UEV, Figure 18.



DMSO Concentration (v/v, %)	Kd (μM)
0	3.7 ± 1
1	3.4 ± 1
5	3.2 ± 1
10	3.0 ± 1

Figure 18: ELISA with His_6-UEV and p6-GST with Increasing Concentrations of DMSO. Error bars represent the standard deviation of three repeats. Curve was fitted using GraphPad Prism7 Software using a Nonlinear Sigmoidal, 4PL curve fit where X is log(concentration of peptide).

The Kd of the p6-UEV interaction was determined to be 3.7 ± 1 µM by this assay. This value falls within the range of reported dissociation constants for this interaction as the affinity is dependent on the buffer conditions; Garrus *et al.* reported p6 to bind to UEV with a Kd of 27 ± 5 µM under physiological conditions containing low concentrations of salts[37], although a tighter affinity (Kd of 4.3 ± 1.6 µM) was reported by Pornillos *et al.*[110] under conditions similar to those used in this assay.

In the presence of increasing concentrations of DMSO, the assay was found to be unaffected by the presence of 10% DMSO (Kd 3.7 ± 1 µM versus Kd 3.0 ± 1 µM with 10% DMSO (v/v)). This confirms the tolerance of the assay with the use of DMSO up to 10% (v/v), which would be necessary for solubilising cyclic peptides.

2.3 Testing Synthesised Peptides for Inhibitory Activity in the p6-UEV ELISA

Once the p6-UEV ELISA had been developed, the next step was to synthesise the capped tri-peptide motifs, YWN and NWY, so that they could be tested for inhibitory activity. The cyclic peptide, CP11, was also synthesised so that the tri-peptides could be compared to the initial hit.

The peptides were synthesised using two different techniques; the tri-peptides were synthesised by liquid phase peptide synthesis, although this method is more laborious, it readily allows for the addition of an acetyl cap on the N-terminus and a diethylamine cap on the C-terminus to mimic the continuation of the peptide backbone. This masks the polar carboxyl and amino terminal groups and allows for the effect of the side chains to be observed. However, CP11 was synthesised by solid-phase peptide synthesis (SPPS), as this method of synthesis would be less time consuming due to the length of the peptide.

Peptide synthesis is the method by which individual amino acids are linked together by coupling the N-terminus of one amino acid to the C-terminus of another amino acid via amide bonds. This coupling reaction does occur spontaneously therefore, the carboxylic acid group must first be activated with a carbodiimide derivative, such as *N,N'*-diisopropylcarbodiimide (DIC), to produce an *O*-acylisourea moiety[111]. The addition of hydroxybenzotriazole (HOBt) also prevents racemisation of non-chiral centres[112-113] during the coupling step.

A way of synthesising peptides systematically through attachment to a porous bead was first conceived by Robert Merrifield in 1963[114], and called SPPS. This method revolutionised the way in which peptides were created synthetically and allows for the formation of unnatural peptides, or peptides that are difficult to produce by other methods.

In SPPS, an amino acid is attached to a porous bead by chemical linkers, which protects the C-terminus of the peptide while also linking the peptide to the bead. The linker used in the synthesis of the peptides in this project is an acid-labile Wang linker, which is compatible with fmoc α-amine protecting groups. The fmoc group is acid stable but can be cleaved under certain basic conditions, in this case: 20% piperidine in dimethylformamide (DMF); this means that the amine terminus of the peptide can be deprotected without removing the growing peptide from the resin, Figure 19.

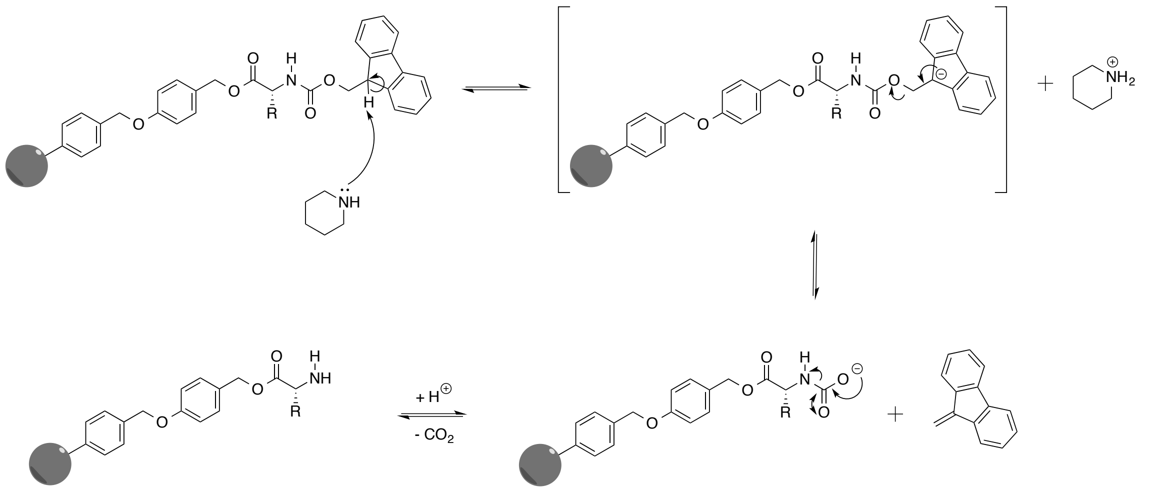


Figure 19: Mechanism of Fmoc Deprotection.

The progress of the reaction is monitored at each step via a Kaiser test[115-116], in which ninhydrin is used to detect primary amines, the abundance of which will decrease during the process of coupling to the next amino acid. In the presence of primary amines, ninhydrin reacts to form a deep purple colour known as Ruhemann's purple, however, if no primary amine is present, no colour change will occur.

Once the linear peptide has been synthesised, it can be cleaved from the resin bead by exposing it to a cleavage cocktail of trifluoroacetic acid (TFA), triisoproplysilane (TIS) and water[117]. The presence of TIS and water serve to minimise side reactions, while TFA cleaves the Wang linker as well as the acid labile side chain protecting groups.

The peptide is then cyclised using 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) and HOBt in DMF. EDC is used for cyclisation as the urea byproduct is water soluble and can readily be removed by aqueous extraction[118]. The cyclisation reaction is completed under high dilution of linear peptide in order to promote intramolecular rather than intermolecular reactions.

2.3.1 Testing CP11 and Capped Tri-Peptides in an ELISA

CP11 was synthesised by SPPS, cleaved from the resin and then cyclised in solution. The cyclic peptide was tested in the p6-UEV ELISA and demonstrated a moderate inhibitory activity, with an IC_{50} of 48.6 ± 2 μM, Figure 20.

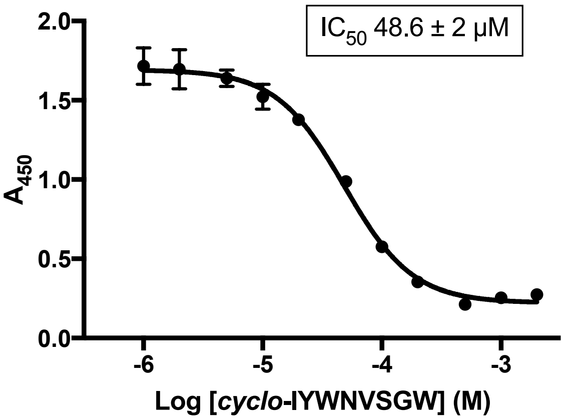
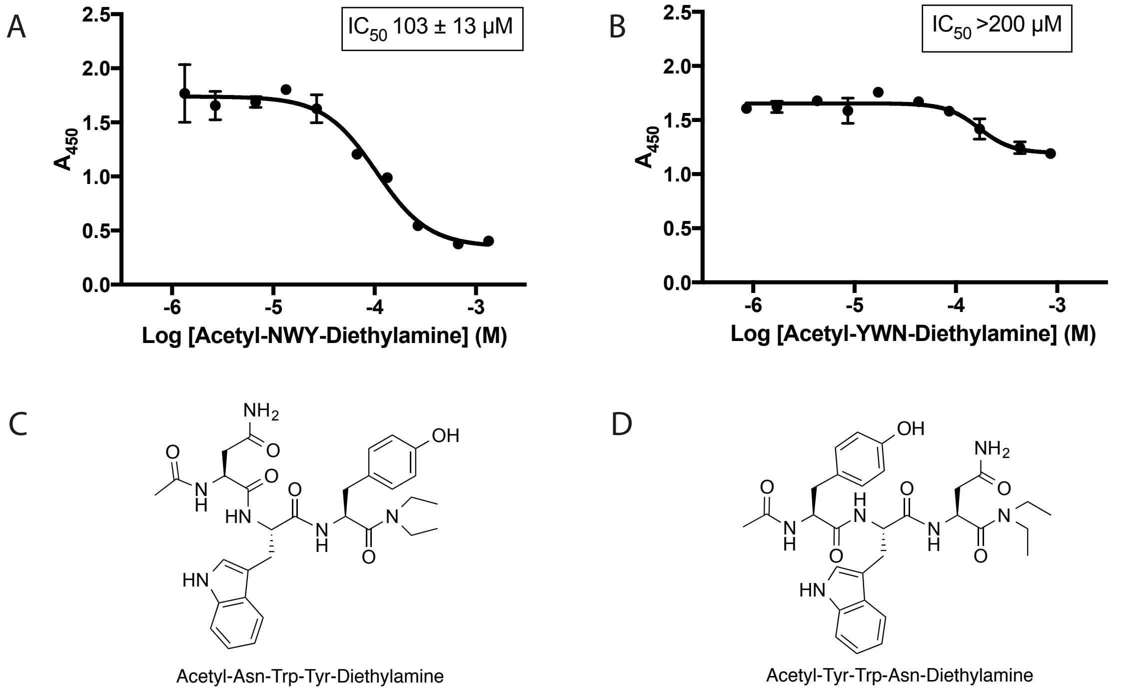


Figure 20: Results of an ELISA with CP11. Error bars represent the standard deviation of a triplicate repeat. Curve was fitted using GraphPad Prism7 Software using a Nonlinear Sigmoidal, 4PL curve fit where X is log(concentration of peptide).

The capped tri-peptides, acetyl-NWY-diethylamine and acetyl-YWN-diethylamine, found in the top two hits CP6 and CP11 respectively, were synthesised by liquid-phase peptide synthesis and then tested in an ELISA for inhibitory activity of the UEV-p6 PPI, Figure 21. Acetyl-NWY-diethylamine demonstrated weak inhibitory activity with an IC_{50} of 103 ± 13 μM, whereas acetyl-YWN-diethylamine had weaker activity that could not be measured as the dose response curve did not reach a plateau at higher concentrations. This could be due to the solubility of the peptide, or the peptide non-specifically binding to the activated nickel surface of the ELISA plate. A counter-screen could have been completed to assess the binding of the peptide to the nickel plate, however was not attempted in this case due to time constraints and resources.



Acetyl-Asn-Trp-Tyr-Diethylamine

Acetyl-Tyr-Trp-Asn-Diethylamine

Figure 21: Results of an ELISA with A. Acetyl-NWY-Diethylamine and B. Acetyl-YWN-Diethylamine. Each peptide was assayed in triplicate. Error bars represent the standard deviation. Curve was fitted using GraphPad Prism7 Software using a Nonlinear Sigmoidal, 4PL curve fit where X is log(concentration of peptide). C. Structure of Acetyl-NWY-Diethylamine. D. Structure of Acetyl-YWN-Diethylamine.

Both tri-peptides demonstrated weaker inhibitory activity in the ELISA compared to the IC_{50} of CP11, 48.6 ± 2 µM. It was interesting to observe that acetyl-YWN-diethylamine demonstrated weaker inhibitory activity even though this motif was observed in the lead compound CP11, than acetyl-NWY-diethylamine, the motif in the second best hit, CP6. This could be attributed to the loss of steric effects imposed by cyclisation that confer a conformational binding advantage, or there may be other residues in CP11 that also contribute to the peptide binding to the protein and hence the inhibitory activity.

2.4 Conformational Analysis of Tri-Peptide Motif

Following on from the discovery that the tri-peptide motifs were 2-3 fold less active than CP11 it was decided to investigate the conformation of the most active tri-peptide, acetyl-NWY-diethylamine as understanding the conformation in solution could provide an insight into the requirements for inhibitory activity.

There are many different experimental methods that can be used for conformational analysis of small molecules, the most popular of which are: Raman, microwave, infrared and UV-vis spectroscopies depending on the time-scale of interest. Another popular method for conformational analysis of small ligands is solution NMR.

NMR spectroscopy can give detailed information about the conformation of molecules through direct analysis of the data; specifically scalar spin-spin coupling constants and cross-relaxation data. These experimentally determined values can be used in combination with quantum mechanical calculations to determine the flexible conformations that a molecule can occupy in solution[119].

2.4.1 Scalar Coupling Coupling Constants

The Karplus Equation describes the relationship between ^3J coupling values and the dihedral torsions between vicinal hydrogens[120-121], H-C-C-H. The equation is:

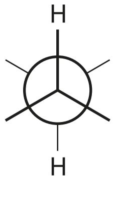
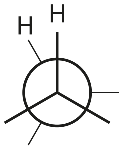
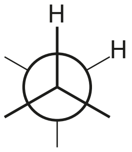
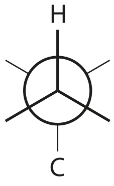
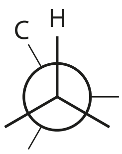
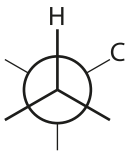
$$^3J_{HH'} = A + B\cos\phi + C\cos 2\phi$$

Where J refers to the coupling constant, ϕ refers to the dihedral angle and A, B and C are constants that vary based upon the substituents on the carbon atoms.

Using this equation, the coupling constants are largest when the dihedral angle is 0° or 180° and the coupling constant is 0 when the dihedral angle is 90°. This information can then be directly applied to conformers to calculate individual $^3J_{HH}$ coupling constants[122-123].

However, some systems do not have vicinal $^3J_{HH}$ coupling constants, therefore other scalar coupling constants are required for conformational analysis. Similar to the correlation between $^3J_{HH}$ and dihedral angles, vicinal $^3J_{CH}$ coupling constants were found to have a defined relationship to C-C-C-H dihedral angles[124], Table 3.

Table 3: The Correlation Between Vicinal $^3J_{HH}$ and $^3J_{CH}$ Coupling Constants and Their Dihedral Angles[125].

	Trans	Eclipsed	Gauche
$^3J_{HH}$	 Large (10-16 Hz)	 Medium (5-9 Hz)	 Small (0-4 Hz)
$^3J_{CH}$	 Large (6-12 Hz)	 Medium (4-6 Hz)	 Small (0-3 Hz)

2.4.2 Cross-Relaxation Data

Nuclear spin polarisation can be transferred from one nuclear spin population to another via cross-relaxation data. This effect is called the Nuclear Overhauser Effect (NOE) and it occurs through space in a quantifiable, distance-dependent manner[126-128]. Nuclear Overhauser Effect Spectroscopy (NOESY) is a 2D NMR technique that includes a mixing time to allow NOEs to build up to the extent that measurements of their magnitude between nuclei can be taken. NOEs are dependent upon three factors: the distance between nuclei, the magnetic field (Larmor procession frequency, ω) and the molecular tumbling (rotational correlation time, τ_c). Although NOEs provide important information about interactions through space rather than through bond, there are some limitations. For molecules with a molecular weight of approximately 500, $\omega\tau_c \approx 1.19$, the NOE becomes equal to zero, Figure 22, and no distance measurement can be taken. A few different aspects of the experiment can be changed to try to circumvent this problem; the solvent viscosity can be altered to try to change the rate of molecular tumbling (either by addition of a co-solvent or by changing the temperature), the magnetic field strength can be changed, or the cross-relaxation can be measured under spin-locked conditions which ensures that there is correlation between the spins. This final technique is known as Rotating frame nuclear Overhauser Effect Spectroscopy (ROESY)[129-130] and, unlike in NOESY, the cross-relaxation rate constant is always positive, Figure 22. ROESY is not without its own flaws though; in some cases, TOCSY (Total Correlation Spectroscopy) artefacts can be observed due to scalar coupling during the mixing state in the spin-lock conditions. To minimise this effect, transverse-ROESY (T-ROESY) was created that used off-resonance spin-locking[131].

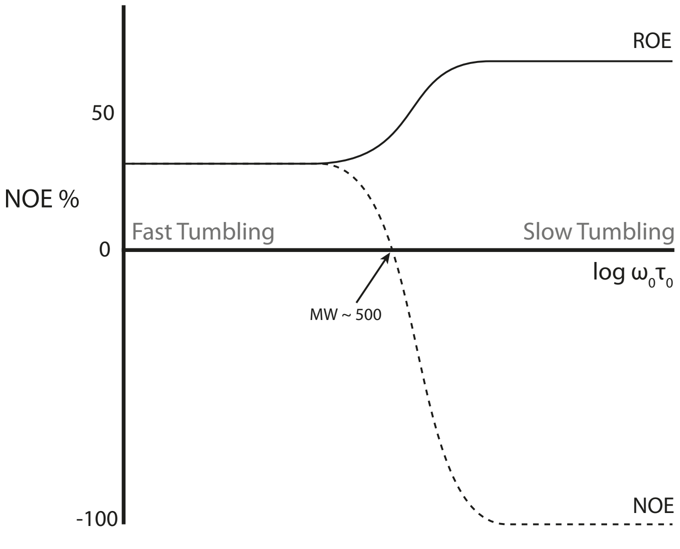


Figure 22: Graph Showing the Difference Between Cross-Relaxation Rate Constants Obtained by NOESY and ROESY.

2.4.3 Ramachandran and Janin Plots

A standard tool for estimating the conformational preference of the dihedral angles (ϕ, ψ) of amino acids is a Ramachandran Plot, Figure 23A. These plots are used to predict the dihedral angles of an individual amino acid when it is left- or right-handed α-helical or β-sheet like. They were first estimated by Ramachandran in 1963 based upon van der Waals forces in di-peptides[132-133], but were then validated using protein crystal structures in the early 2000s[134-135].

Similar to the Ramachandran plot, a Janin plot can be used to predict the most likely sidechain dihedral angles (χ^1 and χ^2) based upon the observed angles seen residues in proteins and compared to energy calculations[136]. Most amino acids were found to be trimodal in χ^1 and although the χ^2 angle was dependent upon the nature of the side chain, approximately 60% favoured one or two configurations, Figure 23B.

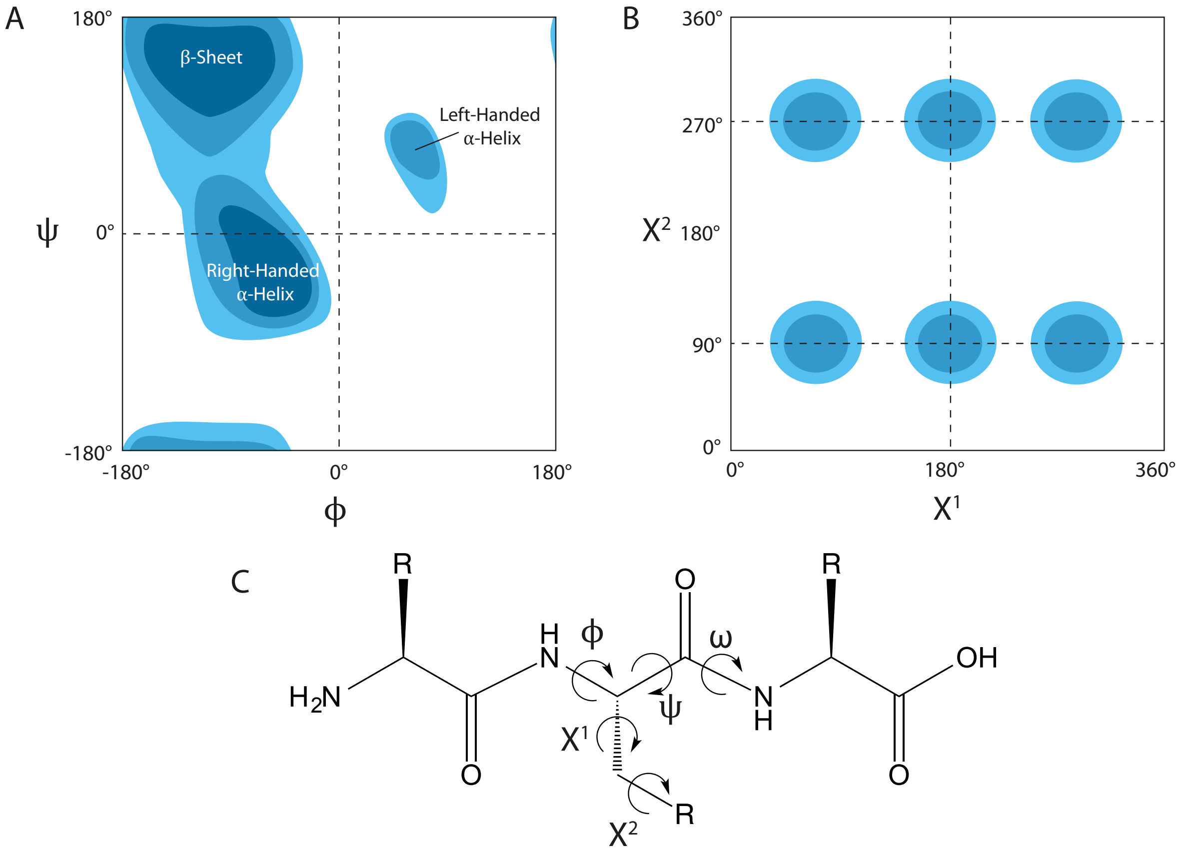


Figure 23: A. A Typical Ramachandran Plot. B. A Representative Janin Plot, Based Upon the Janin Plot for Tyrosine. C. Amino Acid Dihedral Bond Torsion Nomenclature.

2.4.4 Chemical Shift Assignment of Acetyl-NWY-Diethylamine and Structural Restraint Measurement

In collaboration with C4X Discovery (Manchester, UK), the ^{1}H and ^{13}C spectra of acetyl-NWY-diethylamine were recorded at 600 MHz. The atom nomenclature can be seen in Figure 24. To determine the optimum conditions for recording the spectra a few different series of spectra were recorded. The peptide was found to be readily soluble in water (90% H_2O, 10% D_2O) at the concentrations required. Comparison of spectra of two stocks of 0.88 mM and 4.7 mM (90% H_2O, 10% D_2O) showed that the peptide did not aggregate at high concentrations, as there was no change in scalar couplings or in chemical shifts, so a final concentration of 4.7 mM was chosen, as the spectral resolution was improved.

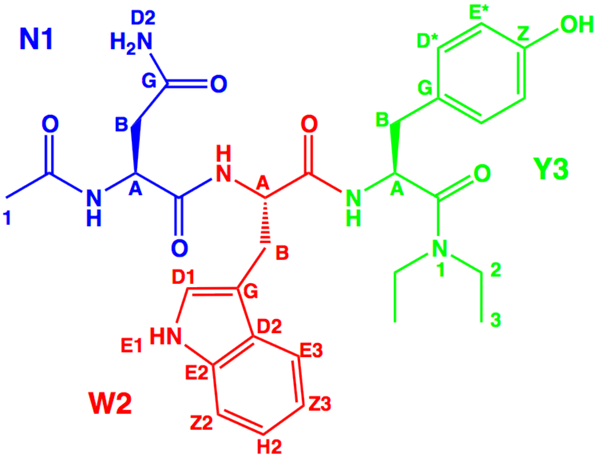


Figure 24: Atom Nomenclature for Acetyl-NWY-Diethylamine

A pH titration was completed and the pKa of the tyrosine hydroxy group was determined to be 10.47 ± 0.09 (measured at 25 °C) using the Henderson Hasselback equation:

$$pH = pK_a + log\frac{[A^-]}{[HA]}$$

Where [HA] is the molar concentration of the undissociated weak acid and $[A^-]$ is the molar concentration of the conjugate base. This value correlates with the known pKa value of the tyrosine hydroxyl group.

At pH 3-4, there was a reduction in the exchange rate of the backbone amide protons, Figure 25. Therefore a low pH was chosen for future experiments so that the peaks associated with the backbone amide proteins would be visible and provide information about the possible backbone configurations of the tri-peptide.

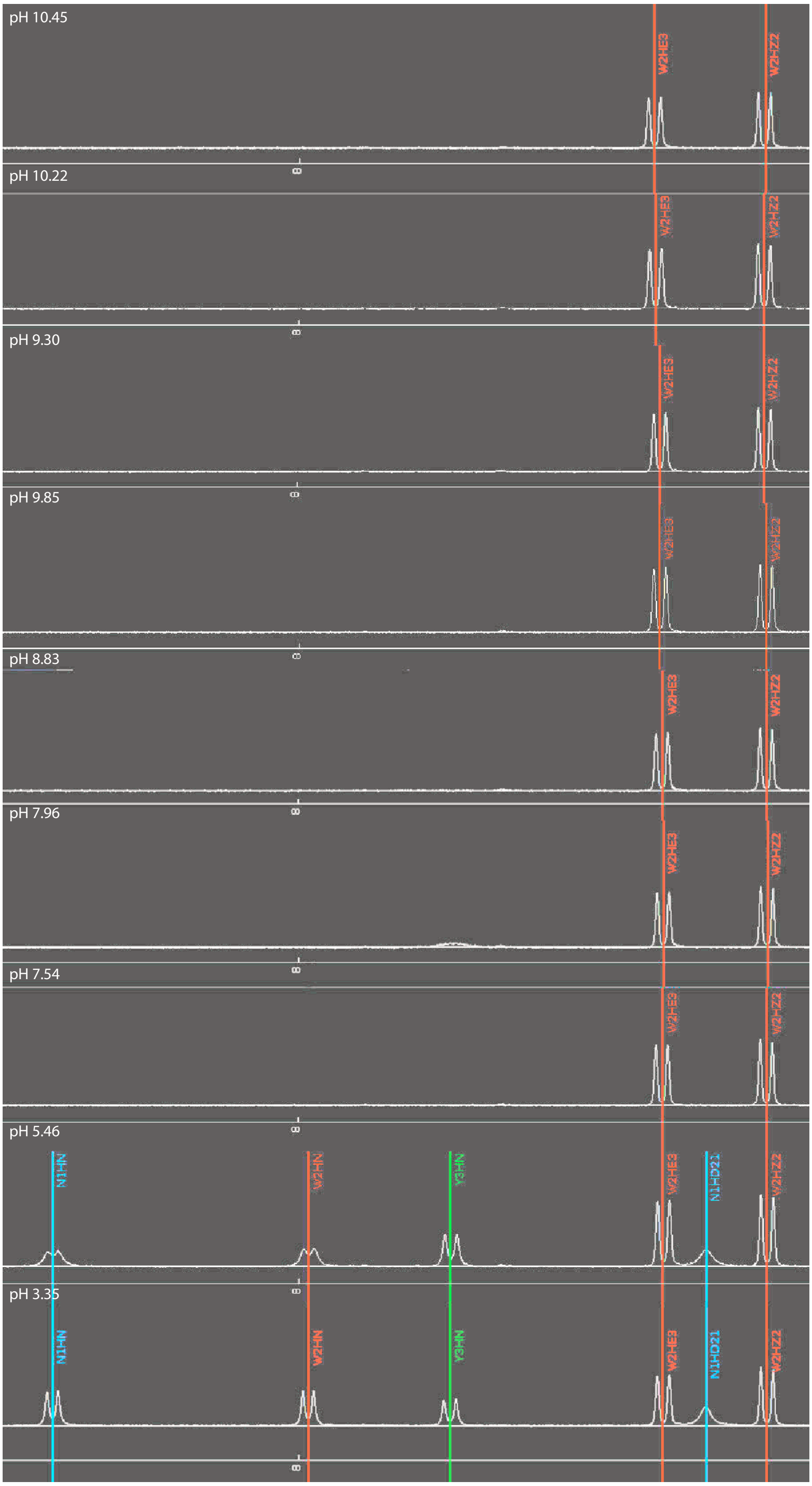


Figure 25: NMR Spectra of a pH Series of Acetyl-NWY-Diethylamine. The signals for the backbone amide protons are distinct at lower pH as proton exchange is minimised. NMR sample conditions: 600 MHz, WATERGATE water suppression was recorded (3-9-19 sequence, 180/-180 pulse train spin-lock, 400ms mixing time) 0.88 mM, 90% H_2O, 10% D_2O, 25°C, 600 MHz spectrometer.

Finally, a temperature series was completed (10 °C, 25 °C and 35 °C) and the temperature coefficients of the backbone amines for Asn-Trp-Tyr were found to be -7.9, -7.7 and -5.2 ppb/C respectively. The signals of the α-protons became less distinct at the higher temperatures as the water signal shifted, Figure 26. The lower temperature of 10 °C was chosen for the final condition as, at this temperature, the signal for the indole NH was visible and all the alpha protons were resolved. A temperature series was also completed as changes in temperature can affect the viscosity of the solution and therefore the τ_c. The information from the temperature series can be used to work out the τ_c when using the finished model to determine the conformation.

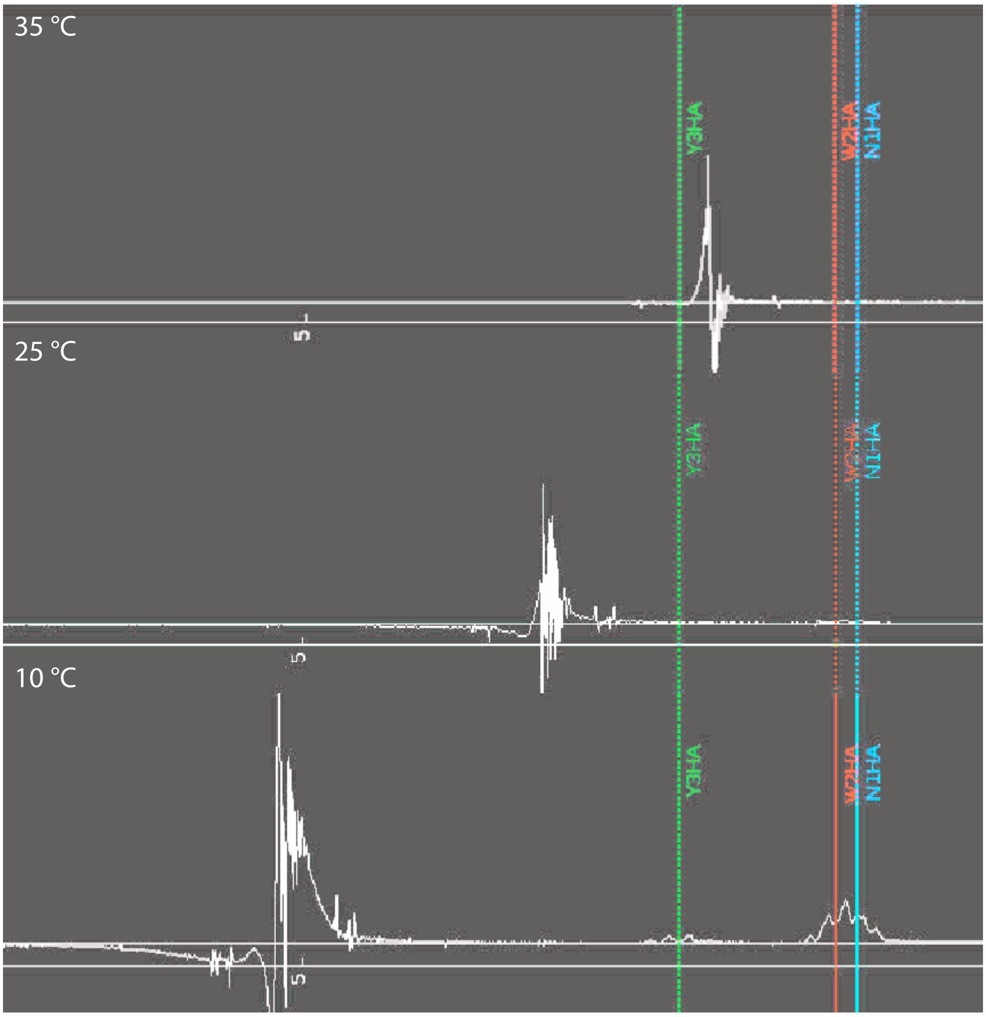


Figure 26: NMR spectra of a Temperature Series of Acetyl-NWY-Diethylamine. The water signal can be seen to shift upfield as the temperature increases and at the higher temperatures completely masks the signals from the α-protons. NMR conditions: 600 MHz, WATERGATE water suppression was recorded (3-9-19 sequence, 180/-180 pulse train spin-lock, 400ms mixing time), 0.86 mM, 100% D_2O, pH 4, 25°C.

Following on from these experiments, the optimal conditions for obtaining 4D NMR data were determined to be: 4.7 mM (90% D_2O, 10% H_2O), 10 °C at pH 3.84. These conditions were chosen as at the lower pH value the peaks relating to the backbone amines could be observed and at lower temperatures the water signal was sufficiently shifted so that the α-proton signals were fully resolved. It is important to ensure that these signals are visible as they provide information about the backbone configuration of the peptide, which is required to drive the computational model. At a higher pH or at higher temperatures more akin to physiological conditions, these signals would not be visible which would mean that the conformational derivatisation would be much more difficult and could lead to computer searching issues. These conditions were chosen by C4X Discovery as this methodology was found to be reproducible and would be easier for a student to complete the modelling required in a shorter time frame.

The ^1H and ^{13}C chemical shifts were assigned from 1D and 2D-NMR experiments. The scalar coupling constants were measured by a resonance line-fitting algorithm[137] and can be seen in Table 4.

Table 4: $^3J_{HH}$ Coupling Constants for Acetyl-NWY-Diethylamine in Solution. NMR conditions: 600 mHz, WATERGATE water suppression was recorded (3-9-19 sequence, 180/-180 pulse train spin-lock, 4.7 mM, 90% D_2O, 10% H_2O 10 °C, pH 3.84, 400ms mixing time).

Residue	Coupling Constant	Value (Hz)
N1	$^3J_{\text{N1HN-N1HA}}$	7.368
	$^3J_{\text{N1HA-N1HB1}}$	5.340
	$^3J_{\text{N1HA-N1HB2}}$	8.576
W2	$^3J_{\text{W2HN-W2HA}}$	6.990
	$^3J_{\text{W2HA-W2HB1}}$	6.737
	$^3J_{\text{W2HA-W2HB2}}$	7.535
Y3	$^3J_{\text{Y3HN-Y3HA}}$	8.211
	$^3J_{\text{Y3HA-Y3HB1}}$	8.219
	$^3J_{\text{Y3HA-Y3HB2}}$	7.016

The coupling constants between the backbone amide protons and the α-protons can be used to determine the dihedral angle using the Karplus Equation. The coupling constants can be plotted across the Karplus graph (generated using the Haasnoot Equation[122], which uses empirical coefficients dependent on the electronegativity of the substituent as well as the phase angle relative to the torsion) for each amino acid to show the predicted dihedral angle, Figure 27. Using these graphs it can be seen that the NH-HA dihedral angle of Asn could be either ± 149° or ± 7°, the NH-HA dihedral angle of Trp could be either ± 146° or ± 15°, and the NH-HA dihedral angle of Tyr could be ± 156°. By linking this information with the Ramachandran Plots for each amino acid residue, it can be seen that Asn and Trp can be either α-helical or β-strand-like, whereas Tyr is more likely to be β-strand-like.

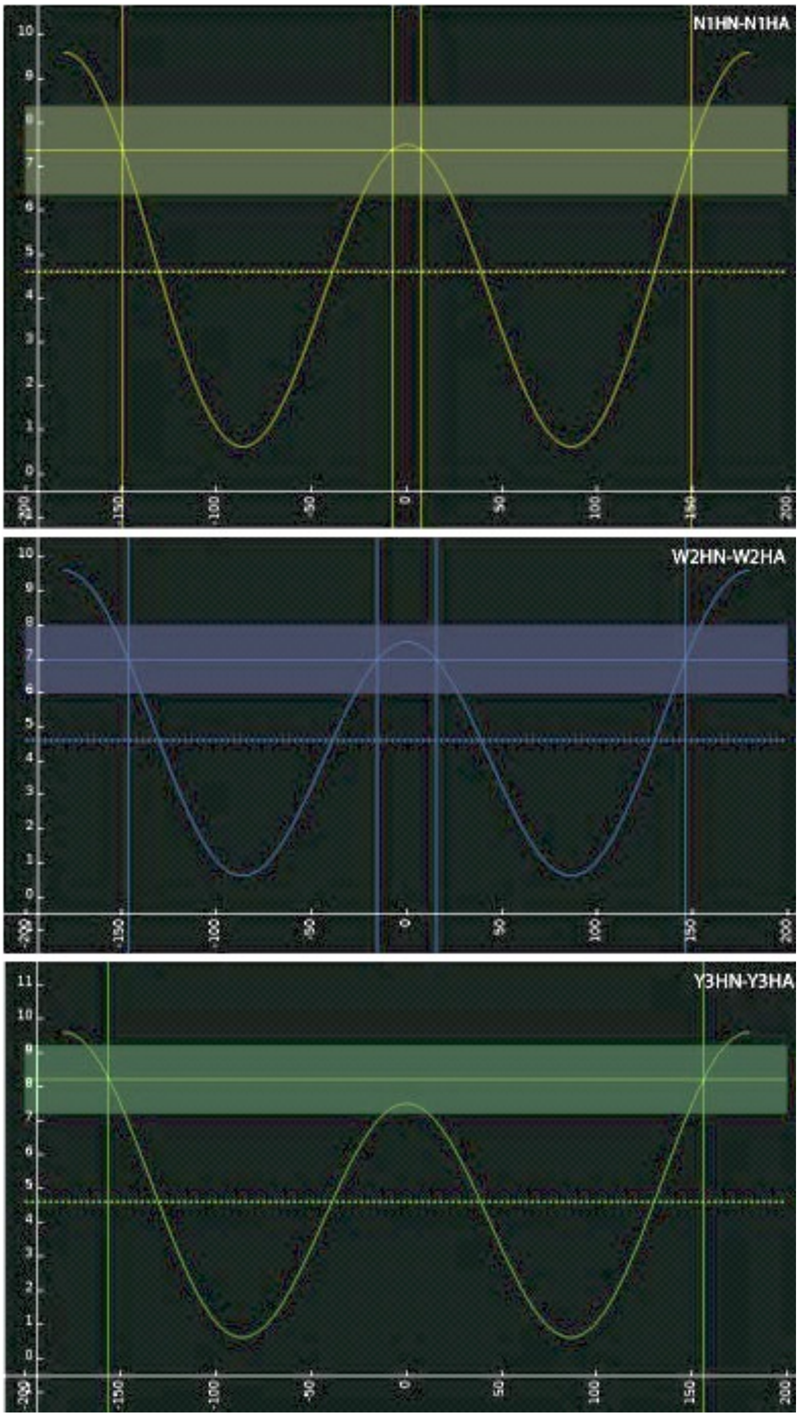


Figure 27: Karplus Graphs of the Relationship Between Backbone Amide and α-Protons. The graphs show the relationship between the calculated coupling constant, shown here as a horizontal line across each graph, and the calculated Karplus graphs specific to the backbone amide and α-protons of each amino acid.

A 2D-T-ROESY spectrum with WATERGATE water suppression was recorded (3-9-19 sequence, 180/-180 pulse train spin-lock, 4.7 mM, 90% D_2O, 10% H_2O 10 °C, pH 3.84, 400ms mixing time) and from this spectrum the absolute values of the cross-peak heights were recorded. It was possible to record the absolute values of the cross-peaks, as the resolution of the spectra was sufficient to see each component of the peaks. To ensure that the value of each cross-peak height compensated for any differences in line shapes, each component was corrected according to the background and rescaled according to the correct equivalent peak-height value determined by the splitting pattern equivalent for one mole of protons (for example, components in a doublet were scaled by 2, components in a triplet were scaled by 4, 2, 4). The mean cross-peak heights were then calculated from each scaled component and this value was used as a structural restraint (ROE) in the computational model. The standard error of each cross-peak was determined to be 40% of the measured height to include both the measurement and prediction errors. In positions where there was no signal intensity above the background, a 'no-ROE' was assigned with a value of 0. The standard error for these values was calculated from the background at the cross-peak coordinates (treating the background as a cross-peak) as a third of the peak value. Each ROE and no-ROE was compared to the [^{1}H-^{1}H]-TOCSY to ensure that any peaks were due to NOE and not just artefacts of directly coupled protons.

2.4.5 Development of a Computational Model

Before the computational model could be developed, a 3-D structure of acetyl-NWY-diethylamine had to be built. The structure was built in Maestro and the structure minimised using the MMFFs forcefield. All nitrogens and rings were checked for planarity and amide bond dihedral angles were set to 180°. The finished structure was then compared with the Cambridge Structural Database for a Mogul *Z*-score of <2.

In the model, each rotatable bond is defined as a set of macrostates (Σ), each described by a mean torsional angle (μ), a spread (σ) and a probability (π). Each rotatable bond can favour different numbers of macrostates allowing it to adopt individual conformations. For example, an sp^3-sp^3 bond will adopt a tri-modal conformation with three macrostates that can be described as $\Sigma_1[\mu_1\sigma_1\pi_1]$, $\Sigma_2[\mu_2\sigma_2\pi_2]$, and $\Sigma_3[\mu_3\sigma_3\pi_3]$. For this tri-mode, the sum of $\pi_1 + \pi_2 + \pi_3$ must be equal to 1, and using this information individual bond microstates can be generated by assigning a random number to each macrostate and then using standard stoichiometric computational techniques[119] and the spread and mean to randomly select the torsional angle. Each bond is assigned a set of macrostates, which when combined, comprise of a dynamic model which represents all the possible conformations that acetyl-NWY-diethylamine could form in solution. Fifteen rotatable bonds were modelled to determine the conformation of the tri-peptide; the two methyl groups on the diethylamine cap were not defined as rotatable bonds, as they will not affect the overall conformation and would complicate the model unnecessarily. The nomenclature of the rotatable bonds in acetyl-NWY-diethylamine is shown in Figure 28 and defined in Table 5.

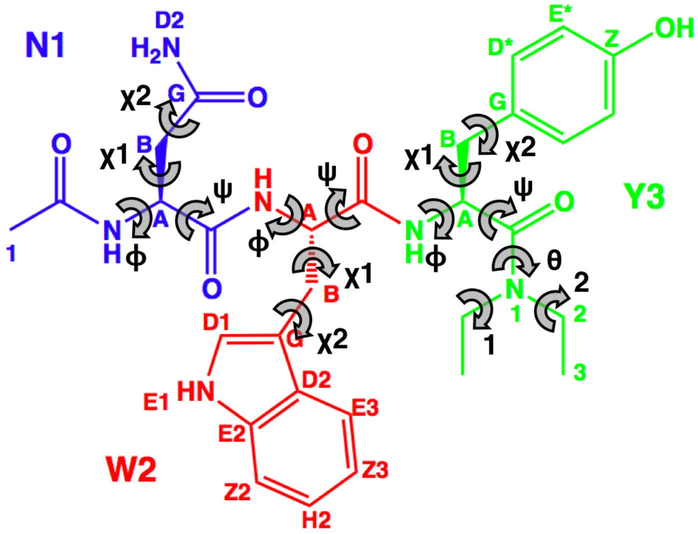


Figure 28: Atom Nomenclature for Acetyl-NWY-Diethylamine, with Torsion Nomenclature.

Table 5: Definition of Torsions of Acetyl-NWY-Diethylamine

Degree of Freedom (Torsion)	Atoms Defining Degree of Freedom
N1 Φ	N1 CO - NH - CA - CO
N1 χ1	N1 NH - CA - CB - CG
N1 χ2	N1 CA - CB - CG - ND2
N1 ψ	N1 NH - CA - CO - W2 NH
W2 Φ	N1 CO - W2 NH - CA - CO
W2 χ1	W2 NH - CA - CB - CG
W2 χ2	W2 CA - CB - CG - CD1
W2 ψ	W2 NH - CA - CO - Y3 NH
Y3 Φ	W2 CO - Y3 NH - CA - CO
Y3 χ1	Y3 - NH - CA - CB - CG
Y3 χ2	Y3 - CA - CB - CG - CD
Y3 ψ	Y3 - NH - CA - CO - N1
Y3 θ	Y3 - CA - CO - N1 - C21
1	Y3 - CO - N1 - C21 - C31
2	Y3 - CO - N1 - C22 - C32

From the inputted possible macrostates, the model could make theoretical predictions of experimental data based upon the theoretical distance between protons and the τ_c. The strength of the model for predicting the experimental data was determined by a Chi-square least-squares measure (X^2) on each experimental measure. The X^2 is defined as the sum of the square distances between predicted and experimental measures, divided by the square of the estimated error on each experimental measurement. After each iteration of the model, new restraints were added, and the X^2 was monitored to ensure that the X^2 per restraint was <1. If the model is too complex, a large X^2 can be predicted as there can be data searching errors. In these cases, the model was refined and the flexibility of the variables restricted, until the restraints allowed for the model to be further refined.

2.4.6 Inputting Data into the Model to Determine Conformation of Acetyl-NWY-Diethylamine

The structural restraints were added gradually to the model and iterative rounds of calculations were completed. After each round of calculations, the X^2 was monitored, and the model was refined by restricting the dihedral angles and the population of each microstate based upon the results generated. Each dihedral angle was given a libration amplitude, which determined how much the model could move the dihedral angle from the set value. In some cases, the macrostates relating to one torsion could be linked to the macrostate of another torsion, for example to prevent the model predicting conformations that would contain steric clashes. By linking the macrostate torsions in this way, the model can be simplified and any data searching errors can be avoided.

Once all the structural restraints were included, the model was tested for robustness by making small changes to see if the overall X^2 could be decreased. Once the X^2 could not be lowered any further, the model was defined as the final model, and a final calculation was completed to generate 9600 acetyl-NWY-diethylamine structures as determined by the model and the inputted structural restraints. These structures were used to generate an ensemble of all the possible conformations that the tri-peptide could form and included all the librations of the dihedral angles, Figure 29A. From this ensemble, an averaged or idealised version of all the conformations can be generated, Figure 29B, where the librations of each dihedral angle are not shown. The idealised possible structures shows the 136 possible macrostates that the model predicts the tri-peptide could form based upon the experimental data.

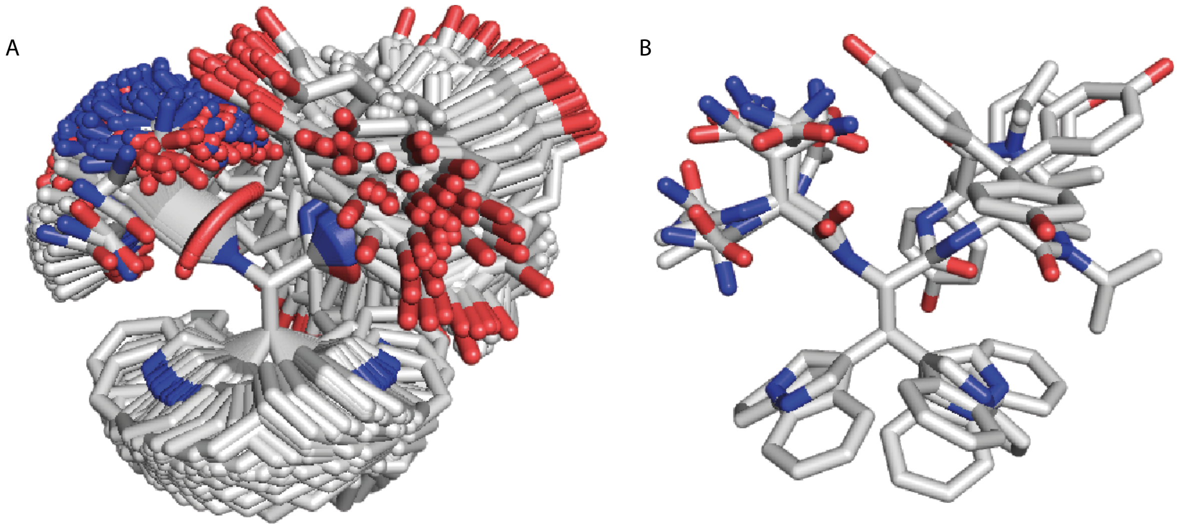


Figure 29: A. Ensemble of all Possible Conformations of Acetyl-NWY-Diethylamine. B. Idealised Conformations of Acetyl-NWY-Diethylamine.

A total of 558 structural restraints were included, Table 6, with an average of 37.2 restraints for each degree of freedom. The X^2 per restraint was calculated as less than 0.4, which is significantly less than the predefined cut-off of 1.0 for each type of restraint, indicating that the conformation generated by the model is well predicted by the experimental restraints.

Table 6: Table of Restraints used in the Model

Type of Restraint	Number of Each Restraint	X^2	X^2 per Restraint
ROE	26	10.2	0.39
No-ROE	523	11.8	0.02
^3J Coupling	9	3.4	0.38
Total	558	25.4	0.05

A number of restraints that had small errors were rejected, as these would skew the final model. The rejected data included restraints that were to W2HE1, as any peaks that lie too close to the water chemical shift and would be affected by the d19 null WATERGATE suppression and would therefore have a much smaller error and would be poorly predicted by the model.

The model predicts that acetyl-NWY-diethylamine can form three major conformations based upon the different backbone conformations as predicted by the $^3J_{HH}$ NH-HA coupling constants: where all the amino acids are in a β-strand like conformation (βββ), where Trp is more α-helical (βαβ) and where Asn is more α-helical (αββ), Figure 30. The primary conformation that acetyl-NWY-diethylamine conforms to in solution is βββ which is populated by 55 ± 6% of the population, whilst βαβ is populated 36 ± 2% and αββ is only minorly populated at 9 ± 2% of the population.

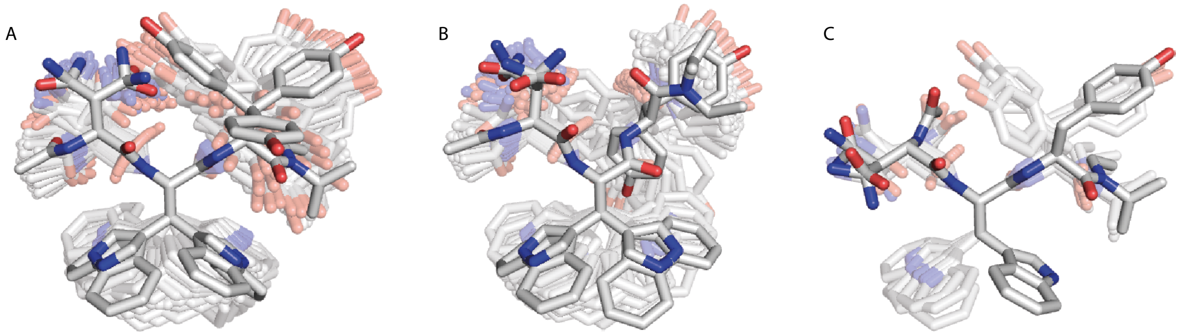


Figure 30: Three Major Conformational Shapes of Acetyl-NWY-Diethylamine. A. NWY (βββ). B. NWY (βαβ). C. NWY (αββ). The averaged shapes are shown in bold and all the possible shapes are shown slightly more faded.

From analysis of these major conformational shapes, information about the conformation of each residue can be analysed in more detail and will be discussed in Sections 2.4.6.1-2.4.6.3.

2.4.6.1 Conformation of Asparagine

The asparagine residue will be referred to as N1, according to the nomenclature defined in Figure 24. The backbone of N1 can adopt two conformations: α-helical or β-sheet like, with the β-sheet like conformation more favoured (91 ± 6% versus 9 ± 2%), Figure 31A. This confirms the conclusions that were drawn from the results of the coupling constant ($^3J_{\mathrm{N1HN\text{-}N1HA}}$) plotted in the Karplus Graph, Figure 27.

The angle of X^1 in the N1 sidechain was found to be trimodal, however the angle of X^2 was found to be bimodal, Figure 31B. The populations and angles of both X^1 and X^2 were found to be dependent on both the backbone and the position of the two other amino acid sidechains so as to avoid any potential steric clashes. In the βββ and βαβ families of shapes, X^1 is either 180° or -60°, but in the αββ family X^1 can either be 60° or -60°. The differences in these angles and how they affect the overall shape of N1 can be seen in Figure 30.

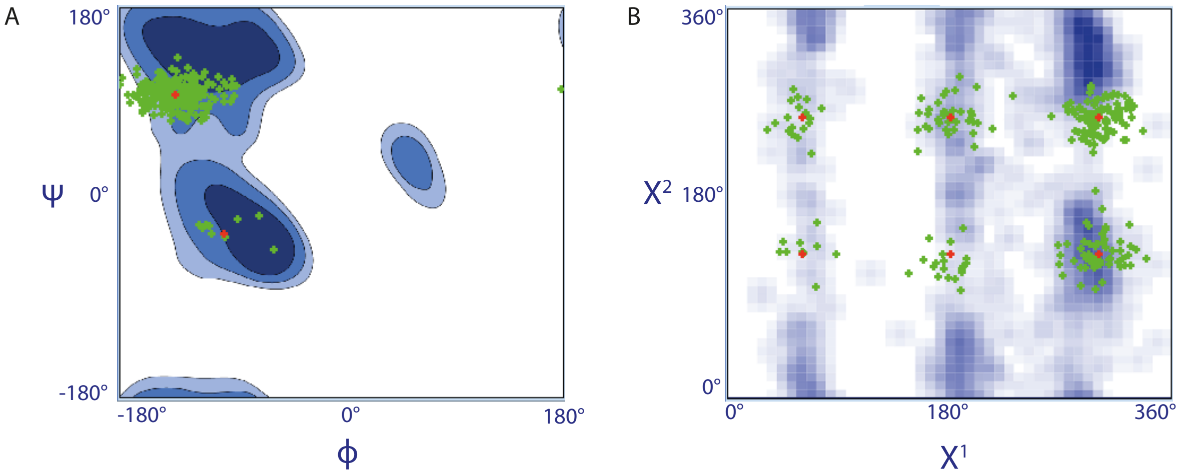


Figure 31: Dihedral Angles of N1, Plotted on A. Ramachandran Plot. B. Janin Plot. The red dots show the dihedral angles and the green dots show the libration range.

2.4.6.2 Conformation of Tryptophan

The tryptophan residue will be referred to as W2, according to the nomenclature defined in Figure 24. Similar to the N1 backbone, the W2 backbone can adopt two conformations: α-helical or β-sheet like. The α-helical conformation is slightly more favoured (64 ± 6%) than the β-sheet like conformation (36 ± 6%), Figure 32A. Again, this is in agreement with the conclusions drawn from the values of the coupling constant ($^{3}J_{\mathrm{W2HN\text{-}W2HA}}$) plotted on the Karplus Graph that W2 could form two backbone conformations, Figure 27. As W2 is in the centre of the tri-peptide, changes in its backbone conformation affects the position of the other residues. This can be seen in Figure 30A and B as the Y3 sidechain sits either towards the "front" or "back" of the molecule depending on the nature of W2.

The X^1 angle of the W2 sidechain is bimodal with approximately equal populations at either -60° or 180° (58 ±8% versus 42 ± 5%, respectively). The X^2 angle is bimodal, however the angles are slightly different when X^1 is 180°, leading to a slightly twisted shape to avoid steric clashes with the Y3 sidechain, especially in the βαβ family, Figure 30A and B.

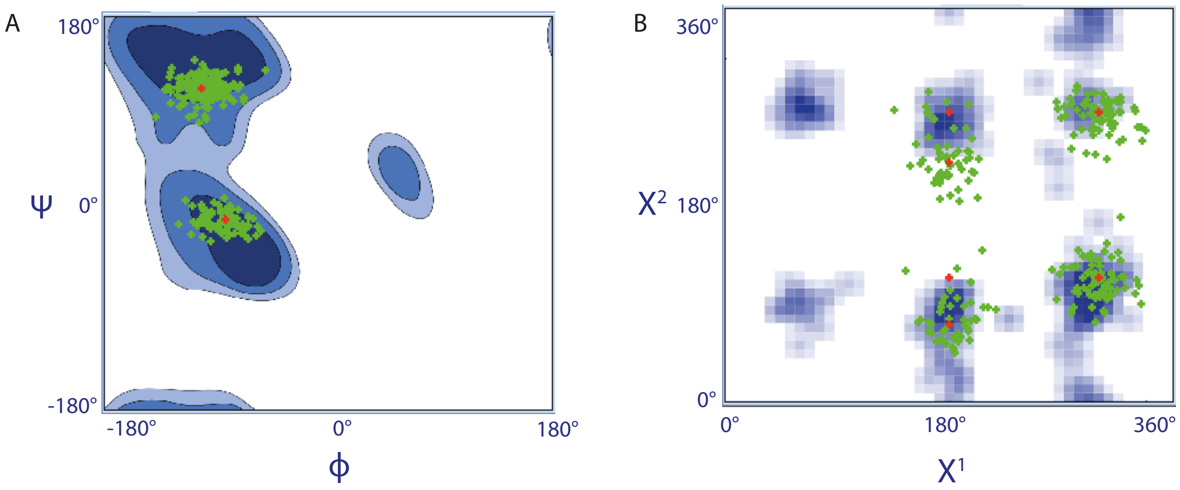


Figure 32: Dihedral Angles of W2, Plotted on A. Ramachandran Plot. B. Janin Plot. The red dots show the dihedral angles and the green dots show the libration range.

2.4.6.3 Conformation of Tyrosine

The tyrosine residue will be referred to as Y3, according to the nomenclature defined in Figure 24. As predicted by the coupling constant ($^3J_{\mathrm{Y3HN\text{-}Y3HA}}$) and Karplus Graph, Figure 27, the Y3 backbone solely forms an α-helical conformation, Figure 33A. The X^1 angle of Y3 is trimodal, Figure 33B, however the populations and torsions are dependent on the backbone conformation and the X^1 angles of N1 and W2. The X^2 angle is bimodal, however it can be seen that the angle is slightly offset from ± 90° when the phenol ring sits over the diethylamine.

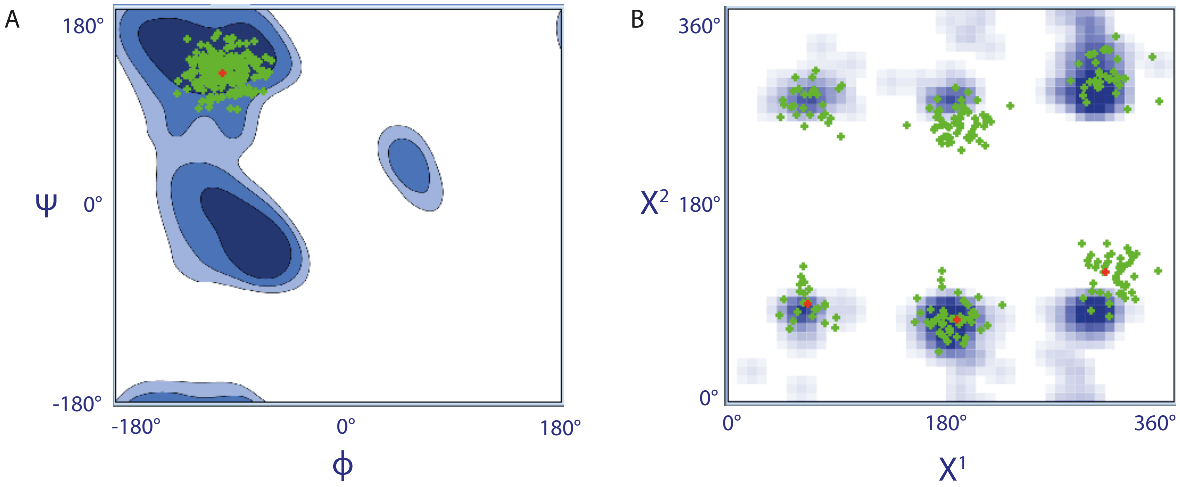


Figure 33: Dihedral Angles of Y3, Plotted on A. Ramachandran Plot. B. Janin Plot. The red dots show the dihedral angles and the green dots show the libration range.

2.4.7 Determination of the Main Conformational Shapes of Acetyl-NWY-Diethylamine

By combining the derived conformations of the individual residues, the main conformational shapes of acetyl-NWY-diethylamine were determined through analysis of the 9600 structures generated by the model. The capped tri-peptide is flexible in solution forming three families of conformations and within these families it can be seen to form five main shapes, Figure 34. These differences in shapes are firstly due to the backbone configurations and then the angles of X^1 for each residue are interdependent to avoid steric clashes within the molecule.

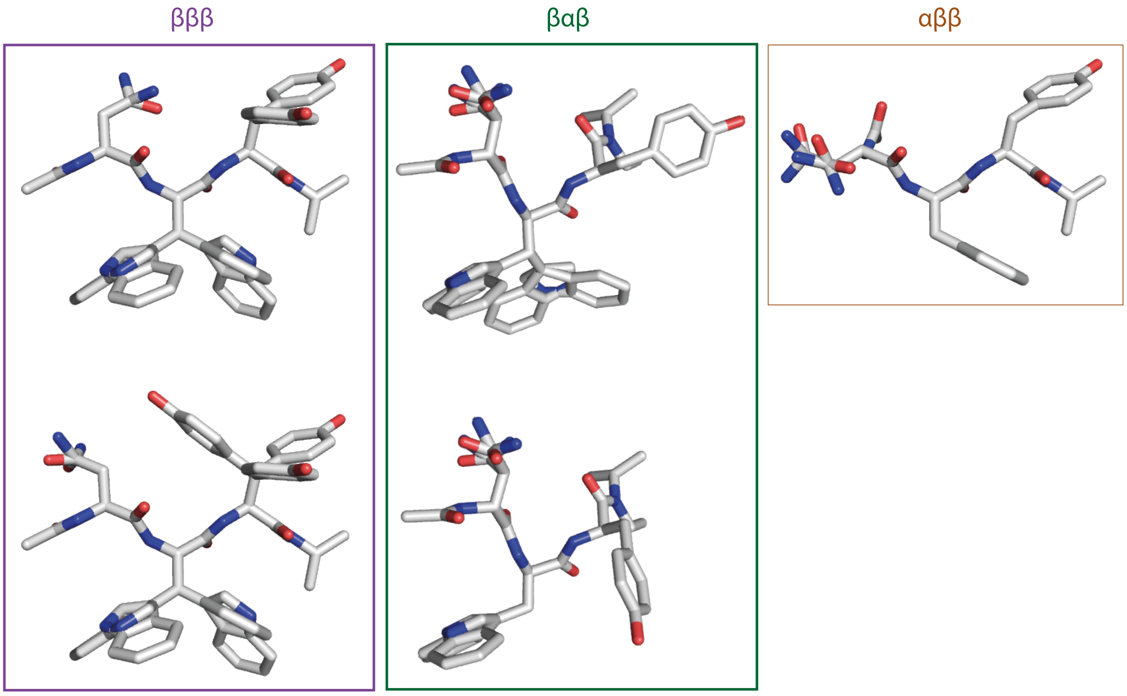


Figure 34: The Five Main Conformational Shapes Acetyl-NWY-Diethylamine Forms in Solution Organised by the Backbone Conformation.

It is not unsurprising that the tri-peptide does not form a fixed conformation in solution, as it is not cyclic nor has any other linking structure that would impose any rigidity. Small peptides are known to rapidly interchange between ensembles of conformations[138]. To circumvent this problem, it is possible to stabilise the peptide in a non-aqueous solution or fix it in a matrix and complete solid-state NMR analysis. However this was not attempted, as this would not be representative of how the peptide would behave under experimental conditions and could generate a skewed impression of the actual conformation.

This experiment was conducted using very specific conditions (4.7 mM, 90% D_2O, 10% H_2O 10 °C, pH 3.84) to ensure all possible signals were visible in the spectra. This would allow for the maximum amount of information to be gained as well as allow for the use as a training tool. However, these conditions are not physiologically relevant, therefore the tri-peptide conformations determined in this experiment may not be the conformations that would be formed or the populations of the families could be shifted, when under physiological conditions, for example with high concentrations of salts, additives, pH 7, 37°C. These conformations, therefore, can not be used in understanding potential binding conformations, especially as peptides are known to undergo changes of conformation upon binding. However, future work could be carried out investigating the conformation of the peptide under physiological conditions and in a bound state and then compared to these findings to assess any similarities or key differences.

No further calculations were carried out to correlate solvation energies, enthalpy or entropy considerations, as these were outside of the scope of this research. The aim of the experiments described in this section was to gain an understanding of the methodology employed by C4X Discovery as part of a collaboration. Especially as the experiments were not completed under physiological conditions, further calculations would not provide any additional information that could be important to this research topic. Future work to investigate the bound conformations of the peptides under physiologically relevant conditions ideally would be carried out and then further calculations about the physiochemical properties of the peptides could be calculated from these data.

2.4.8 Comparison of the Determined Conformation of Acetyl-NWY-Diethylamine with an Structural Model of CP11

As CP11 was more active than acetyl-NWY-diethylamine in the ELISA assay, it was hypothetically assumed that the most populated conformation of CP11 would be the active motif. Therefore the determined conformations of the tripeptide could be compared to the determined conformation of CP11 to see if any of the conformations show significant overlap. Although NMR spectra were not obtained under physiological conditions, or in the presence of the protein binding partner (as the conformation could change upon binding), the actual bound conformation could have no relevance to the determined conformation. However, these experiments were completed in collaboration with C4X Discovery as a simplified teaching tool to be used as an introduction to their software and the techniques used by the company.

The three conformational families of acetyl-NWY-diethylamine were compared to a structural model of CP11 generated by C4X Discovery. The model of CP11 was determined by analysis of the scalar coupling of the backbone and NOE restraints under similar conditions (low pH, low salt and low temperature), but was limited to the -YWN- backbone section to minimise the computational power required. From the scalar couplings measured in the NMR, the tryptophan residue existed solely in an extended conformation, whereas the tyrosine residue can adopt both helical and extended conformations, Figure 35. There were two NOE restraints measured between the tyrosine phenol ring and the tryptophan side chain, which suggests that the αβ model is significantly more favoured in solution as this is the only form where the two side chains are close enough.

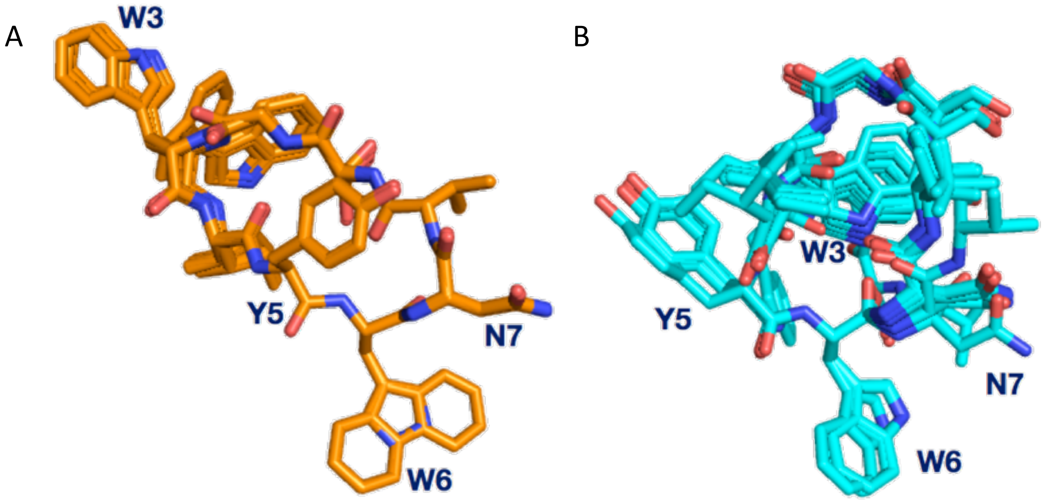


Figure 35: Structural Model of CP11. A. Backbone of YW in ββ conformation. B. Backbone of YW in αβ conformation.

When the conformational shapes of acetyl-NWY-diethylamine were overlaid with the αβ structural model of CP11, only some conformers within the βββ backbone had a significant overlay. These conformers only accounted for 5% of the overall population of acetyl-NWY-diethylamine forms. Hypothetically this could explain the findings described in Section 2.3.1, as to why acetyl-NWY-diethylamine is 2-fold less active than CP11 in an ELISA, based on the aforementioned assumption that the most populated conformation of CP11 determined by NMR would be indicative of the active conformation upon binding.

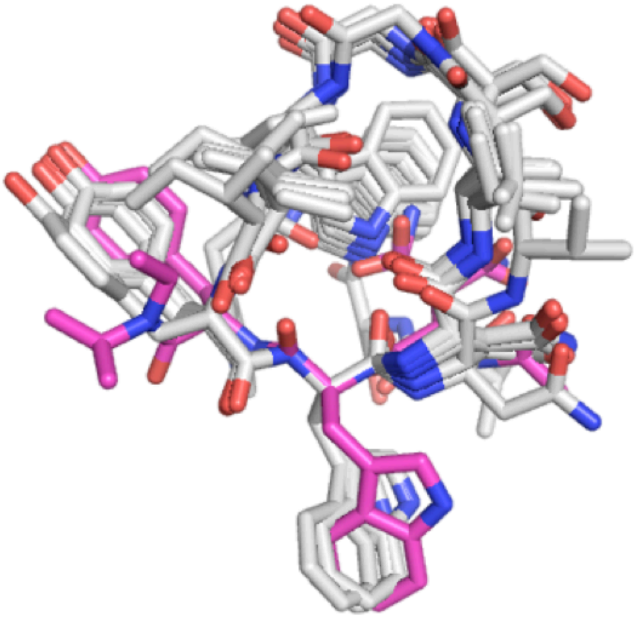


Figure 36: Overlaid Conformations of αβ Backbone of CP11 and Acetyl-NWY-Diethylamine Conformers.

There is no correlation between the peptide sequence of CP11 (*cyclo*-IYWNVSGW) or of acetyl-NWY-diethylamine and the active motif of p6 (PTAP) and without completing NMR analysis of the peptides in the presence of the protein binding partner it is not possible to hypothesise potential binding sites or compare any of the determined conformations with the known structural data for the p6-UEV binding interaction. As the NMR data were not gained under physiological conditions, comparison of the determined conformations to any known structural information would not provide a feasible comparison of the actual conformation of the peptides to any known structural data for the p6-UEV interaction.

2.5 Synthesis of Sterically Hindered Tri-Peptides

Following on from the determination of the conformation of acetyl-NWY-diethylamine, it was hypothesised that the loss in activity was due to the flexibility of the linear peptide. Therefore, in an attempt to improve the inhibitory effect of the more active tri-peptide, acetyl-NWY-diethylamine, sterically hindered cyclic derivatives were synthesised by SPPS, to try and reduce the flexibility of the peptide. The peptides were synthesised in bulk until the asparagine residue and then split into three different batches. Each batch was coupled to either 6-aminohexanoic acid (Ahx), 5-aminovaleric acid (Ava) or γ-amionbutyric acid (Gaba). The peptides were then cleaved from the resin and head-to-tail cyclised in solution under high dilution before purification by preparative-HPLC. These amino acid derivatives were chosen as they could readily be coupled onto the tri-peptide and allow for head-to-tail cyclisation of the peptides. The different lengths of each of the amino acid derivatives allowed for the peptides to be cyclised with varying ring sizes to include a degree of steric hindrance and with the intention of locking the peptides in a conformation that was more conducive to protein-binding and therefore improving the inhibitory activity.

The three tri-peptides were then tested in the p6-UEV ELISA, Figure 37. Unfortunately, the cyclic derivatives demonstrated no inhibitory activity in the ELISA. This could be due to the linkers imposing conformational constraint upon the tri-peptides that forces the motif away from the active conformation.

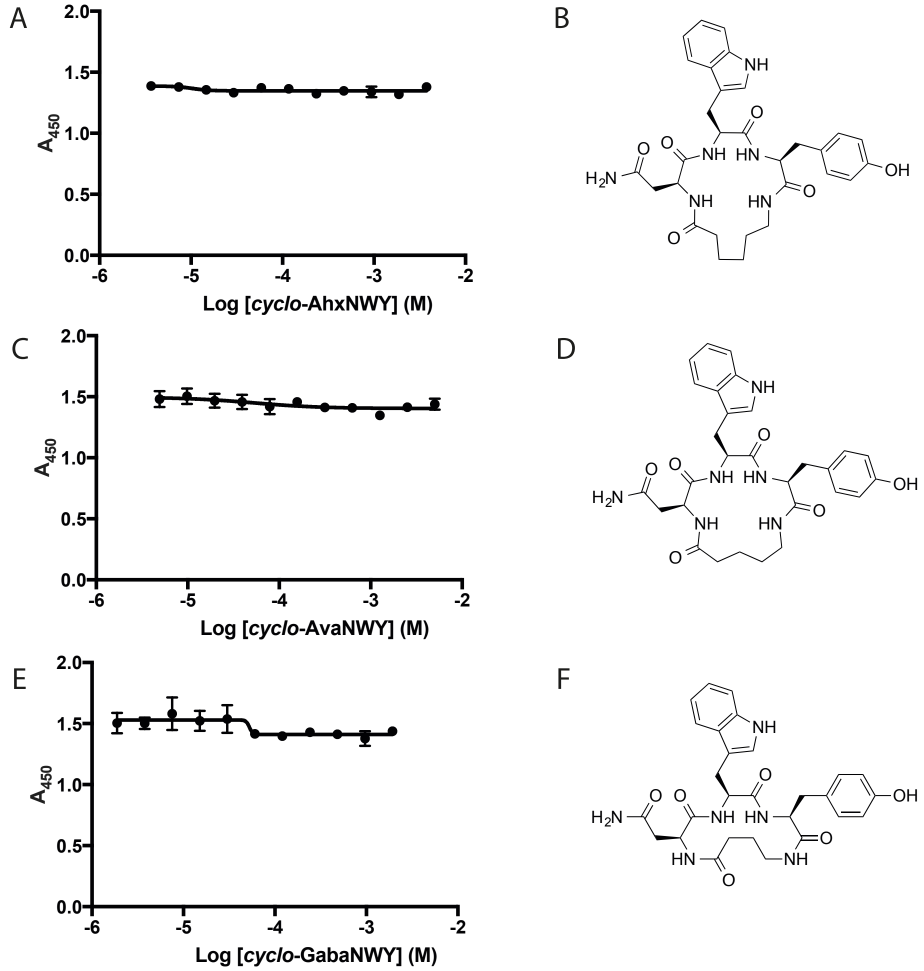


Figure 37: Results of an ELISA with Sterically Hindered Cyclo-NWY Derivatives. Each peptide was assayed in triplicate. The error bars represent the standard deviation. Curve was fitted using GraphPad Prism7 Software using a Nonlinear Sigmoidal, 4PL curve fit where X is log(concentration of peptide). A. Graph of effect of Cyclo-AhxNWY on p6-UEV interaction. B. Structure of Cyclo-AhxNWY. C. Graph of effect of Cyclo-AvaNWY on p6-UEV interaction. D. Structure of Cyclo-AvaNWY. E. Graph of effect of Cyclo-GabaNWY on p6-UEV interaction. F. Structure of Cyclo-GabaNWY.

Another method was attempted to enforce steric hindrance on the tri-peptides; N-methylation of the backbone amines is a popular method for introducing steric hindrance and has also been shown to improve some properties of peptides, such as cell permeability and solubility[139]. The methodology proposed by Chatterjee *et al.* for selective N-methylation using SPPS[140] was followed, whereby the peptide bond is protected first by twice repeated exposure (15 minutes) to four equivalents of 4-nitrobenzenesulfonyl chloride. Then N-methylation was attempted by exposing the protected amino acid to three equivalents of 1,8-diazabicyclo[5.4.0]undec-7-ene (3 minutes) followed by twice repeated exposure (2 minutes) to ten equivalents of dimethyl sulfide. The 4-nitrobenzenesulfonyl chloride protecting group was then removed by exposure to ten equivalents of 2-mercaptoethanol and five equivalents of 1,8-diazabicyclo[5.4.0]undec-7-ene (5 minutes). Then coupling of the next amino acid was attempted using three equivalents of the desired fmoc-protected amino acid with three equivalents each of 1-hydroxy-7-azabenzotriazole (HOAt) and 2-(7-Aza-1H-benzotriazole-1-yl)-1,1,3,3,-tetramethyluronium (HATU) (3 hours). After each step the resin was washed well with DMF. However, coupling proved unsuccessful due to the incomplete coupling reactions to the N-methylated amino acids. It was hypothesised that the failed coupling reactions were due to the additional methyl group significantly reducing the reactivity of the terminal nitrogen combined with the steric hindrance of the bulky tryptophan side chain when attempting to couple to the tyrosine residue. Future attempts could utilise microwave radiation to improve the coupling efficiencies[141], but this technology was not available for use during the time of this research.

2.6 Summary

In summary, two tri-peptide motifs (YWN and NWY), identified from the top two hits in the initial SICLOPPS/RTHS screen[103], were assessed for inhibitory activity in the p6-UEV ELISA developed for this purpose. They were found to be weakly active (IC_{50} of >200 µM and 103.8 ± 13 µM, respectively) at inhibiting the p6-UEV PPI in the ELISA, however they were more than 2-fold less active than the cyclic peptide, CP11 (IC_{50} of 48.6 ± 2 µM). This loss of activity was hypothesised to be due to the increased flexibility of the linear tri-peptides compared with CP11.

NMR analysis of the more active tri-peptide, acetyl-NWY-diethylamine, in combination with computational modelling was used as a learning tool for determination of peptide conformation in collaboration with C4X Discovery. The experiments were completed at specific conditions to ensure a good signal to noise and to obtain structural information about all the residues. These conditions were not physiologically relevant so do not provide any information about possible conformations but did highlight the intrinsic conformational flexibility of linear peptides.

An attempt was made to increase the steric hindrance of acetyl-NWY-diethylamine, to hopefully force the motif to adopt the more active conformation, however it caused a complete loss of inhibitory activity in the ELISA.

It was interesting to discover that the capped-NWY peptide, which is the motif from the second best hit compound identified in the original SICLOPPS screen, CP6, demonstrated slightly stronger inhibitory activity than the capped-YWN tri-peptide taken from CP11. It was therefore hypothesised that other residues in CP11 are part of the active motif and confer a binding advantage.

Chapter 3 Determining the Active Motif of CP11

3.1 Introduction

Although the inhibitory effect of CP11 had already been investigated[103], and also determined by ELISA in Chapter 1, the active motif of the peptide had not been determined, therefore an alanine scan was completed. Alanine scanning is a methodology whereby each residue in turn is exchanged for alanine and the effect of this substitution is determined. By this method, an indication of the residues that are most important for activity can be determined, as substitution for alanine at these positions will result in a decrease in potency. Alanine is used over other residues as it maintains the secondary structure of the peptide, but is chemically inert and lacks a bulky side chain.

In this chapter, two methods are described that were used to complete an alanine scan. The first makes use of the methodology by which CP11 was identified (SICLOPPS combined with a bacterial RTHS), whilst the second involves the synthesis of each analogue using SPPS and then testing in an ELISA.

3.2 Alanine Scan using the p6-UEV RTHS

3.2.1 Generation of SICLOPPS Plasmids for Alanine Scan in a RTHS

As CP11 was identified using a combination of SICLOPPS and a RTHS, this technology could be used to complete an alanine scan of the peptide. Firstly, the sequence for CP11 had to be incorporated into the pARCBD plasmid. This was achieved by PCR with a forward primer that coded for the CP11 peptide sequence and the chitin binding domain (CBD)-reverse primer to amplify the CP11 DNA sequence from the pARCBD plasmid. To prevent any mismatching between the DNA strands, this PCR was followed by a second PCR reaction that used a ‘zipper’ primer which anneals to the 3’ end of the I_c intein[96]. The PCR product was then digested using the *HindIII* and *BglI* restriction sites and then ligated back into a newly digested pARCBD backbone. This ligation was transformed into chemically competent DH5α and successful construction of the pARCDB-CP11 plasmid, Figure 38, was confirmed by DNA sequencing.

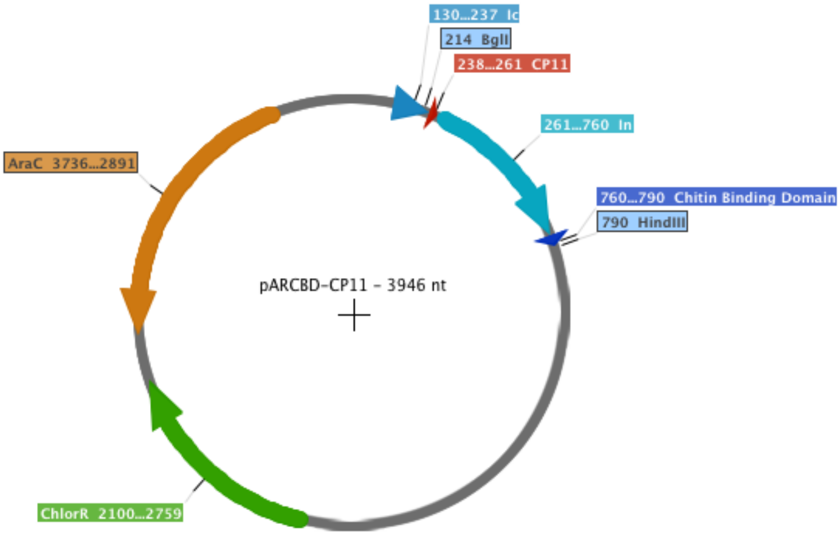


Figure 38: Plasmid Map of pARCBD-CP11.

The alanine analogues were created using site-directed mutagenesis to mutate the DNA sequence of the pARCBD-CP11 plasmid at each residue to alanine, Table 7.

Table 7: Names and Sequences of Alanine Analogues

Peptide Name	Peptide Sequence							
CP11	I	Y	W	N	V	S	G	W
CP11A1	A	Y	W	N	V	S	G	W
CP11A2	I	A	W	N	V	S	G	W
CP11A3	I	Y	A	N	V	S	G	W
CP11A4	I	Y	W	A	V	S	G	W
CP11A5	I	Y	W	N	A	S	G	W
CP11A6	I	Y	W	N	V	A	G	W
CP11A7	I	Y	W	N	V	S	A	W
CP11A8	I	Y	W	N	V	S	G	A

After the PCR with the site-directed mutagenesis primers, the PCR reactions were incubated at 37 °C with DpnI. As this restriction enzyme digests only at methylated nucleotides, which are produced during replication of the plasmid with *E. coli*, the template will be cleaved leaving only the PCR products containing the alanine mutations. The success of the site-directed mutagenesis could not be confirmed by colony PCR or by agarose gel analysis as the mutated plasmid would be the same size as the template plasmid, so sequences were checked by DNA sequencing.

An alanine analogue for serine could not be screened for by using SICLOPPS as the split inteins require a serine or cysteine in the first position to ensure splicing and generation of the cyclic peptide product as the first step of the process is a transesterification reaction to form the lariat intermediate, Figure 7. An analogue could have been created that contained a cysteine instead of a serine to ensure splicing efficiency, however, this might not be a significant enough exchange sterically to observe any effects on the potency of the analogue. It would also not be a true 'alanine-scan' as both residues are nucleophilic in nature and a false result could be obtained that would only be indicative of the difference in splicing efficiency between cysteine and serine residues.

The pARCBD plasmids containing the CP11, alanine analogues, as well as pARCBD-CP11 were transformed into chemically competent cells containing the p6-UEV RTHS and plated onto agar containing kanamycin (25 µg/mL), chloramphenicol (35 µg/mL) and spectinomycin (25 µg/mL), to select for successful transformants. Individual colonies from each transformation were picked and then grown overnight in LB broth with antibiotics to create bacterial stocks.

3.2.2 Optimisation of the Conditions for Drop-Spotting

Overnight cultures of the bacterial stocks were serially diluted ten-fold in 10% glycerol (v/v; 20 µL of culture in 180 µL glycerol solution) then drop-spotted (2.5 µL/drop) onto minimal media agar containing ʟ-(+)-arabinose (to induce expression of the cyclic peptides from the pARCBD vectors), 3-amino-1,2,4-triazole (3-AT, a competitive inhibitor of HIS3 gene production), kanamycin and IPTG (to induce expression of the target proteins, in this case UEV and p6).

Initially, the drop-spotting conditions as described by Tavassoli *et al.*[103]*,* in which CP11 was identified, were used (minimal media supplemented with 13 µM ʟ-(+)-arabinose, 2.5 mM 3-AT, 25 µM kanamycin and 30 µM IPTG). However, in the presence of arabinose and independent of any IPTG, toxicity was observed for the alanine analogues and no growth was observed. Therefore, the concentration of ʟ-(+)-arabinose was halved to 6.5 µM and the growth of bacterial transformants was assessed by drop-spotting, Figure 39.

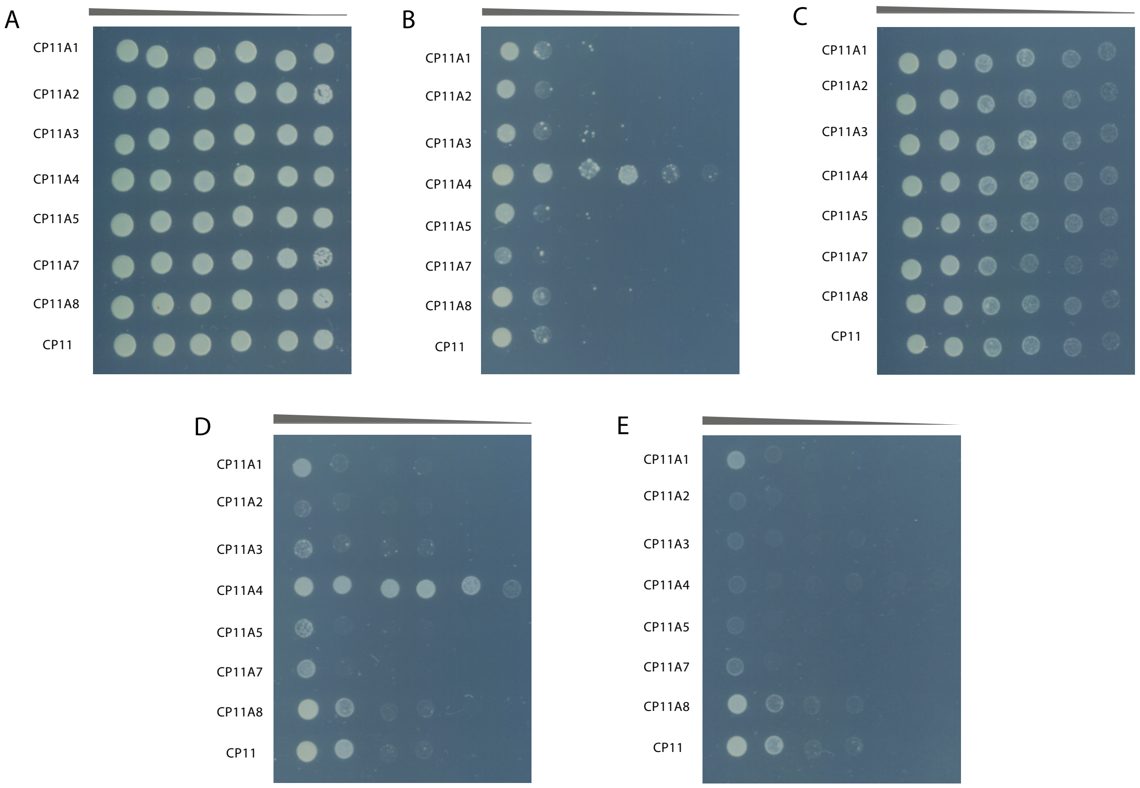


Figure 39: Screening Alanine Analogues in the p6-UEV RTHS by Drop-Spotting onto Minimal Media Agar. Containing: A. No additives. B. Arabinose (6.5 µM). C. 3-AT (2.5 µM), Kanamycin (25 µM) and IPTG (30 µM). D. 3-AT (2.5 µM), Kanamycin (25 µM) and Arabinose (6.5µM). E. 3-AT (2.5 µM), Kanamycin (25 µM), Arabinose (6.5µM) and IPTG (30 µM). Concentration of bacteria decreases from left to right across the agar plate.

Toxicity was still observed for most of the alanine analogues apart from CP11A4 even at this lower concentration of arabinose, so a further screen was completed with a range of lower concentrations of ʟ-(+)-arabinose (0.1-3 µM) in the absence of other additives to optimise the conditions of the screen, Figure 40.

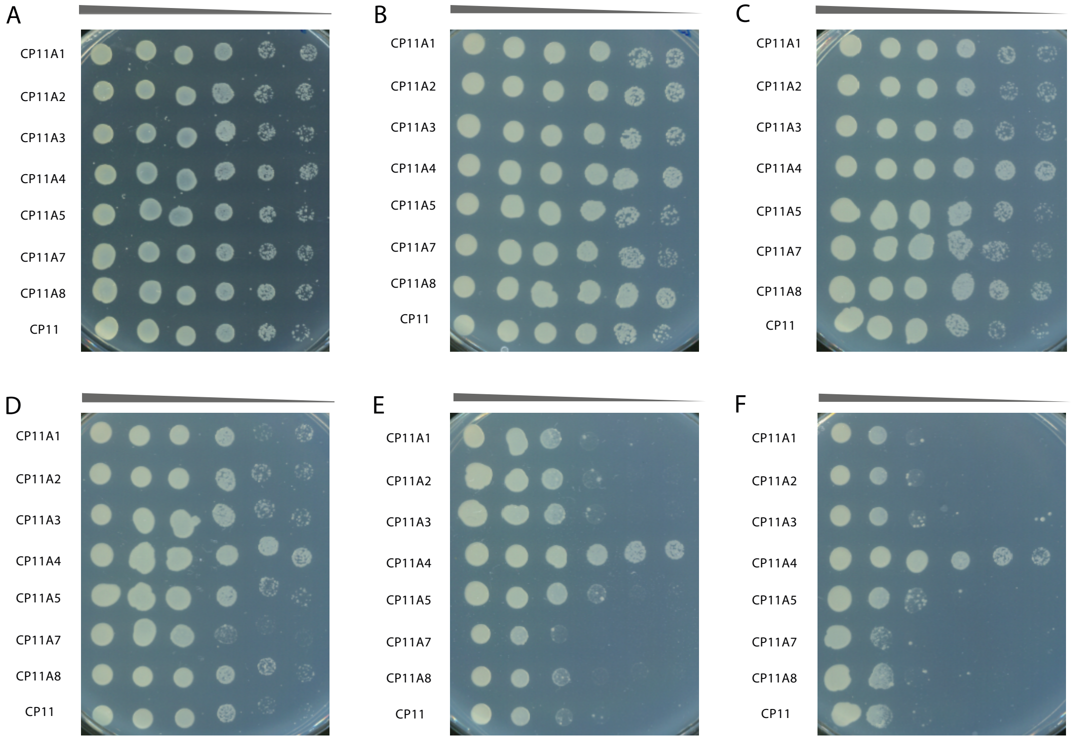


Figure 40: Optimisation of ι-(+)-Arabinose Concentrations for Drop-Spotting. Minimal media plates contain: A. no additives. B. 0.1 µM ι-(+)-arabinose. C. 0.5 µM ι-(+)-arabinose. D. 1 µM ι-(+)-arabinose. E. 2 µM ι-(+)-arabinose. F. 3 µM ι-(+)-arabinose. Concentration of bacteria decreases from left to right across the agar plate.

From the results of the screen with a range of lower concentrations of ι-(+)-arabinose, toxicity was still observed at 2-3 µM, however, moderate to low toxicity was observed at 0.1-1 µM. This could indicate that at the low concentrations of ι-(+)-arabinose the expression of the peptides is low enough that the *E. coli* cells are unaffected by the toxicity. Conversely, it could also mean that the concentration of ι-(+)-arabinose is insufficient to induce the expression of the cyclic peptides so no toxicity can be observed.

A screen was therefore completed with a smaller range of concentrations of ι-(+)-arabinose (0.5-1 µM) and IPTG to see if any effect was observable, Figure 41.

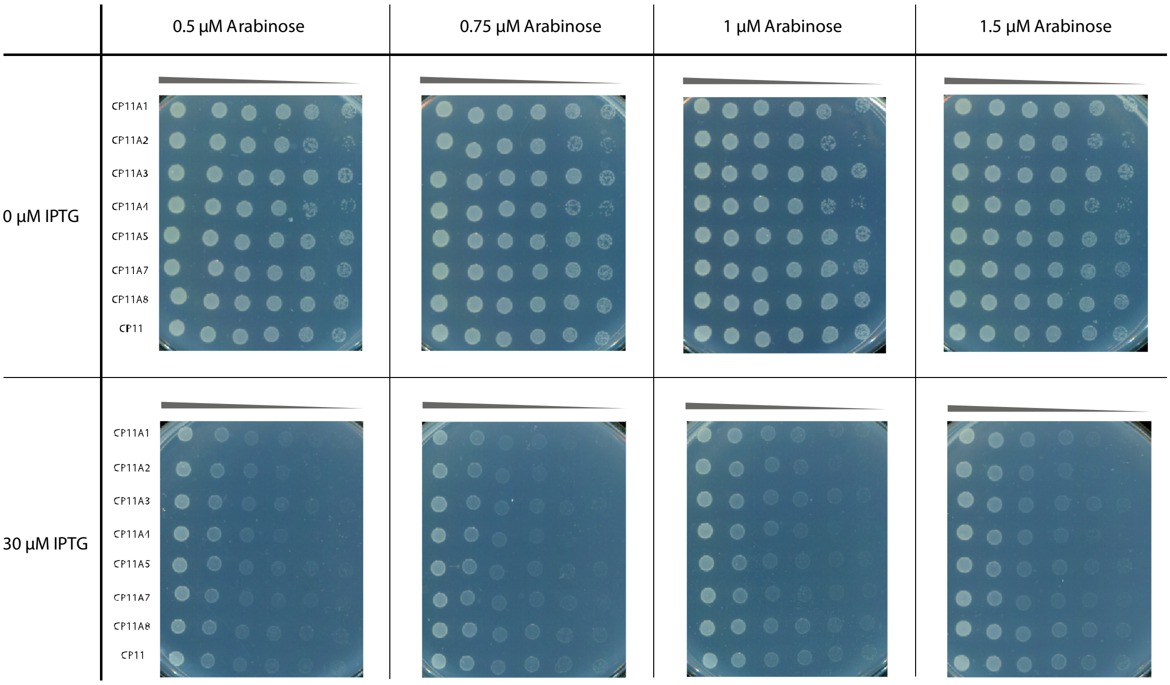


Figure 41: Results of a Screen with Lower Concentrations of ʟ-(+)-Arabinose (0.5-1.5 µM) With and Without IPTG (30 µM). Concentration of bacteria decreases from left to right across the agar plates.

Minimal toxicity was observed at lower concentrations of ʟ-(+)-arabinose (0.5-1.5 µM), however, this lack of toxicity could be due to the concentration of ʟ-(+)-arabinose being too low to induce the splicing of the cyclic peptide or the peptides must not have been expressed in a high enough concentration to allow for the PPI to be inhibited in the presence of IPTG and thus allowing the bacterial cells to grow. This would also explain why no obvious regrowth is observed in the presence of IPTG for CP11, as this peptide is a known inhibitor of the p6-UEV RTHS. The alanine analogues must therefore be toxic to the *E. coli* even at low concentrations. Therefore, this was not an appropriate assay to determine the active motif of CP11. For this reason an alternative method was used whereby the alanine analogues were synthesised by SPPS and then screened using the p6-UEV ELISA.

3.3 Alanine Scan using an ELISA

As the alanine analogues were toxic to *E. coli,* it was not possible to screen the alanine analogues using the p6-UEV RTHS and an alternative method had to be used. Therefore, the alanine analogues were synthesised by SPPS and their inhibitory effect was assessed using the ELISA that was developed as described in Chapter 1.

3.3.1 Synthesis of Alanine Analogues of CP11 by Solid-Phase Peptide Synthesis

Eight alanine analogues of CP11 were synthesised by SPPS and then purified using preparative-HPLC. The sequences and names of the peptides can be seen in Table 7; the residues that have been substituted for alanine in each peptide are shown in magenta.

3.3.2 Assessing the Effect of the Alanine Analogues Using the p6-UEV ELISA

The eight alanine analogues were assayed in triplicate in the ELISA to evaluate changes in inhibitory activity, Figure 42A-H. The IC_{50} of each peptide was calculated and can be seen in Table 8. From the results of the ELISA, the most important residue for activity in CP11 is the tyrosine residue in position 2, as substitution for alanine resulted in a 5-fold loss in activity compared to the IC_{50} of CP11. This IC_{50} value was calculated assuming the two data points presented the beginning of the final plateau of the sigmoidal curve; this may not be the case as it would be expected that the final plateau should be approximately the same as for the other peptides if maximal inhibition occurs. This could mean that the actual IC_{50} value for this alanine analogue may be much higher and may be greater than 1 mM; further assays would need to be completed at higher concentrations to accurately determine the IC_{50}, however this was not possible due to the insolubility of the peptide in highly aqueous conditions.

Substituting any of the other residues for alanine did not result in any change in IC_{50}. The ELISA data obtained for the peptides with tryptophan and glycine substituted for alanine, Figure 42C and G respectively, have significantly larger error bars than the other data. These large errors could be due to issues with peptide purity or due to slight differences between the ELISA plates used, as the triplicate repeats of these peptides were completed over two different ELISA plates. Although every attempt was made to incubate the plates with the TMB-ELISA solution for the same amount of time before adding sulfuric acid to stop the colorimetric development, there would be deviation due to pipetting errors. These errors make it difficult to draw any conclusions about the importance of these residues, however, as they do not demonstrate the large loss in activity shown when tyrosine was substituted for alanine and their IC_{50} values were within the range of the IC_{50} of CP11 they were assumed to be unimportant for inhibitory activity.

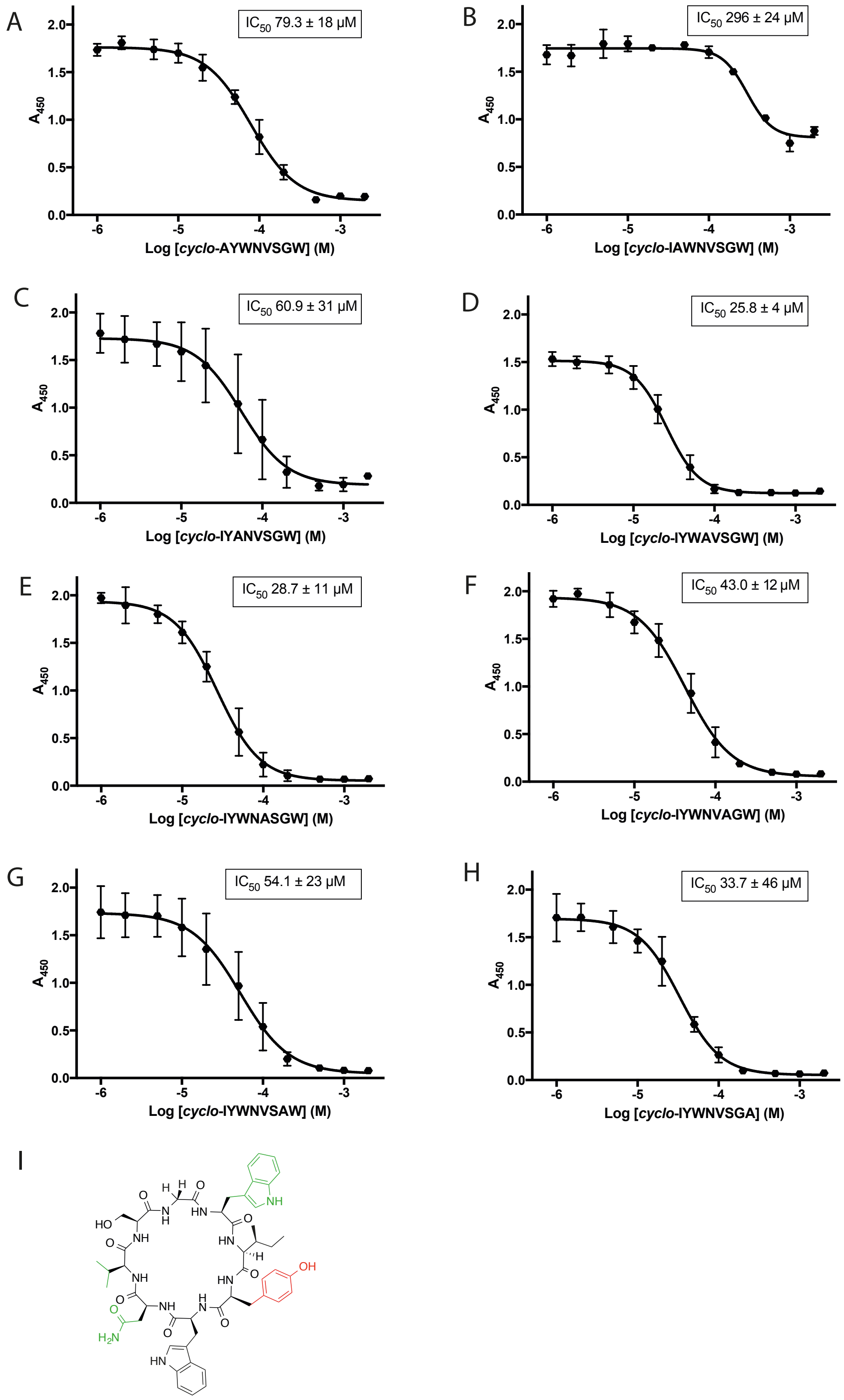


Figure 42: Results of Alanine Scan when Tested in an ELISA. Each peptide was assayed in triplicate. Error bars represent the standard deviation. Curves was fitted using GraphPad Prism7 Software using a Nonlinear Sigmoidal, 4PL curve fit where X is log(concentration of peptide). Graphs showing inhibitory effect on p6-UEV PPI of A. CP11A1. B. CP11A2. C. CP11A3. D. CP11A4. E. CP11A5. F. CP11A6. G. CP11A7. H. CP11A8. I. Structure of CP11 highlighting the most important residues according to the results of the alanine scan. Red is the most important residue, black is no change in inhibitory activity, green shows an improvement in inhibitory activity when substituted for alanine.

Interestingly, when some of the residues were substituted for alanine (asparagine in position 4, valine in position 5 and tryptophan in position 8) the analogues showed a slight improvement in IC_{50} when compared to CP11. This improvement in IC_{50} might be due to a slight change in conformation of the cyclic peptide, as alanine has less steric bulk than valine or asparagine.

Table 8: IC_{50} of CP11 and Alanine Analogues

Peptide	P11 Alanine Analogues	IC_{50} (µM)
CP11	Cyclo-IYWNVSGW	48.6 ± 2
CP11A1	Cyclo-AYWNVSGW	79.3 ± 18
CP11A2	Cyclo-IAWNVSGW	296 ± 24 (>1 mM)
CP11A3	Cyclo-IYANVSGW	60.9 ± 31
CP11A4	Cyclo-IYWAVSGW	25.8 ± 4
CP11A5	Cyclo-IYWNASGW	28.7 ± 11
CP11A6	Cyclo-IYWNVAGW	43.0 ± 12
CP11A7	Cyclo-IYWNVSAW	54.1 ± 23
CP11A8	Cyclo-IYWNVSGA	33.7 ± 4

3.4 Synthesis and Testing of Acetyl-IYW-Diethylamine and Di-Peptide Truncations in an ELISA

As the alanine scan showed that the active motif of the cyclic peptide centred around the tyrosine residue, not the residues YWN as hypothesised previously, it was decided to synthesise a motif surrounding the tyrosine residue as a tri-peptide and as two smaller truncations IY and YW to see if only these motifs were sufficient for inhibitory activity. The di- and tri-peptides were synthesised by liquid phase peptide synthesis and capped as before with an acetyl group on the N-terminus and a diethylamine group at the C-terminus to mimic the continuation of the peptide backbone. These peptides were tested using the p6-UEV ELISA, Figure 43.

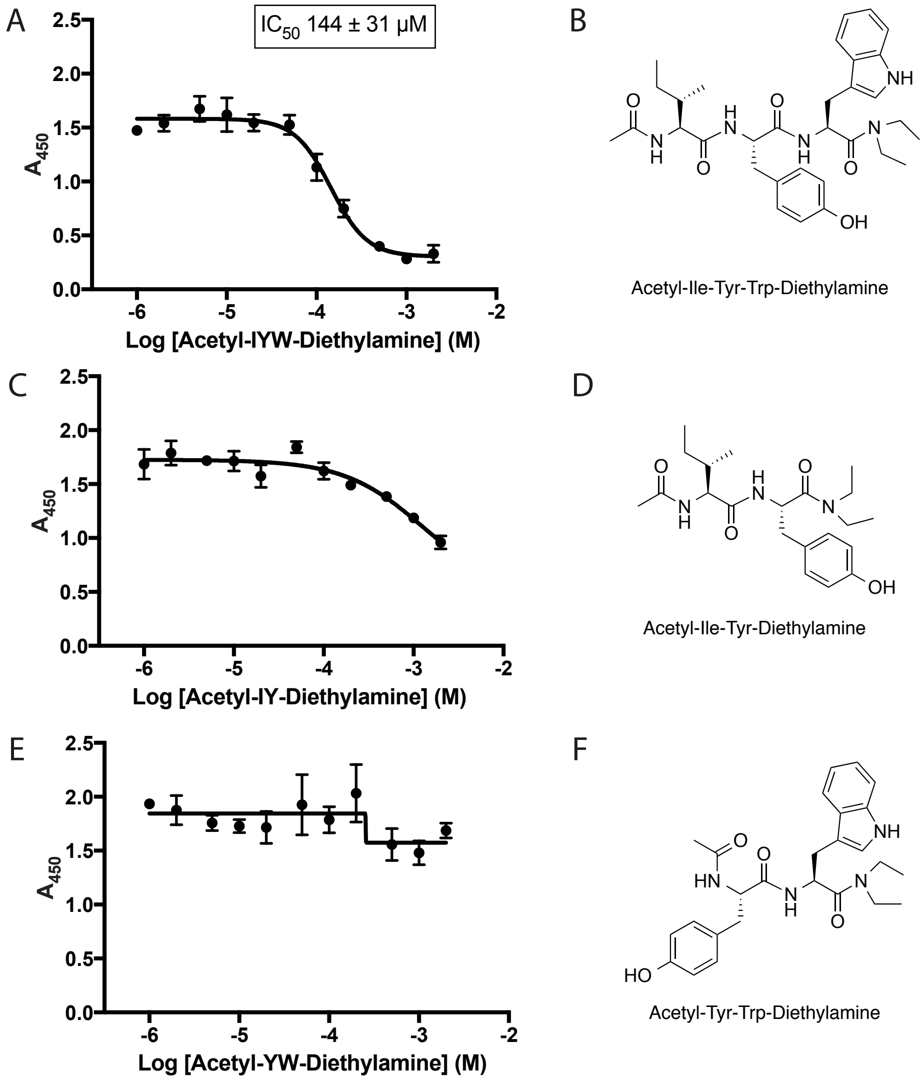


Figure 43: Results of an ELISA Testing Acetyl-IYW-Diethylamine, Acetyl-IY-Diethylamine and Acetyl-YW-Diethylamine. Each peptide was assayed in triplicate. Curves was fitted using GraphPad Prism7 Software using a Nonlinear Sigmoidal, 4PL curve fit where X is log(concentration of peptide). Error bars represent the standard deviation. A. Graph of ELISA results with Acetyl-IYW-Diethylamine. B. Structure of Acetyl-IYW-Diethylamine. C. Graph of ELISA results with Acetyl-IY-Diethylamine. D. Structure of Acetyl-IY-Diethylamine. E. Graph of ELISA results with Acetyl-YW-Diethylamine. F. Structure of Acetyl-YW-Diethylamine.

The capped tri-peptide acetyl-IYW-diethylamine demonstrated moderate inhibitory activity in the ELISA that was slightly stronger than acetyl-YWN-diethylamine (165 ± 34 µM), but was almost 3-fold less active than CP11. The capped di-peptide truncations IY and YW were weak or insufficiently potent enough to observe any inhibitory activity in the ELISA. This decrease in activity could be due to the increased flexibility of the tri-peptide, as observed with YWN, and so the tri-peptide must occupy the more active conformation for a smaller proportion of the time than the more sterically hindered cyclic peptide. This also indicates that the tri-peptide must be the minimal possible motif for conferring inhibitory activity as neither of the di-peptides demonstrated any inhibitory activity.

3.5 Summary

In an attempt to determine the active motif of CP11, an alanine scan was performed using two different techniques: the SICLOPPS/RTHS from which CP11 was initially discovered, and screening synthesised alanine analogues for inhibitory activity in the p6-UEV ELISA. Unfortunately the alanine analogues were toxic to *E. coli* and therefore it was not possible to screen for activity in the p6-UEV RTHS, despite efforts to optimise the drop-spotting conditions.

The results of the alanine scan in the ELISA demonstrated that the active motif of CP11 is solely the tyrosine residue not the YWN motif as was initially suspected. This contradicts the hypothesis that additional residues other than YWN are important for inhibitory activity but instead suggests that conformation of the peptide must be of importance, otherwise any peptide with a tyrosine residue would demonstrate an inhibitory effect. The capped IYW motif was synthesised by liquid phase peptide synthesis and tested in the p6-UEV ELISA, however, it was found to be 3-fold less active than CP11, which, again, was attributed to the increased flexibility of linear peptides.

These findings highlight that the conformation of the active motif is highly important for peptide binding and therefore inhibitory activity. It was decided to further investigate CP11 to gain potential insights into the binding properties of the cyclic peptide through the inclusion of unnatural aromatic analogues in the position of the tyrosine residue in position 2 as its substitution for alanine resulted in a 5-fold loss in activity when compared to CP11.

Chapter 4 Investigating the Structure-Activity Relationship of CP11

4.1 Introduction

Following on from the results of Chapter 3, where tyrosine was identified as the critical residue for CP11 activity, it was decided to investigate the effect of varying the substituents on the phenol ring of tyrosine would have on the activity of the cyclic peptide. It was hypothesised that by using substituents with different ring substituent effects it would provide an insight into the binding properties of the peptide. There are many unnatural aromatic analogues commercially available that have differing ring substitutions, meaning that analogues of CP11 containing unnatural aromatic residues could be easily synthesised by SPPS and then cyclised in solution as before.

In Chapter 3, it was shown that substituting certain residues, particularly asparagine, valine and tryptophan (positions 4, 5 and 8, respectively), for alanine resulted in an improvement of inhibition over CP11. It was hypothesised that this improvement in activity is due to the alanine substitutions causing a favourable change in conformation in the cyclic peptides, resulting in greater binding to the target protein. To further investigate this hypothesis, analogues of CP11A4, CP11A5 and CP11A8 were synthesised with unnatural aromatic analogues.

Unnatural aromatic analogues can provide information about the interaction between the protein and the peptide. Strongly deactivating aromatic ring substituents are known to enhance π-π stacking interactions by decreasing the electronic repulsion between interacting π-systems by withdrawing electron density from the aromatic ring[142]. Therefore, if the tyrosine aromatic ring was binding via a π-π stacking interaction it would be expected that analogues with strong electron withdrawing effects would improve binding. On the other hand, analogues that have aromatic rings with strongly electron donating groups such as methyltyrosine or tyrosine itself would have an opposing effect to π-π stacking interactions; by donating electron density to the aromatic ring, the electronic repulsion between π-systems will be greater and the π-π stacking interaction will be destabilised. Halogen and methyl substituents are known to have differing effects on aromatic rings; halogens are weakly deactivating whilst methyl groups are weakly activating, although both are ortho/para directing and so could be influencing a cationic- or anionic- (respectively) aromatic non-covalent interaction in a similar manner. Substituent groups on aromatic rings are known to have an additive effect on aromatic interactions depending on the nature of the substituent[143]. The unnatural phenylalanine analogue, 4-pyridylalanine, contains an aromatic ring with an electrophilic nitrogen that is slightly less electrophilic than benzene due to the electron withdrawing effects of the nitrogen, this slight altering of the properties of the aromatic ring could provide information about the binding interaction, if the peptides containing this analogue demonstrate an improvement in binding.

Non-aromatic analogues cannot bind through π-π interactions, however aliphatic-aromatic stacking interactions have been shown to be stronger than π-stacking interactions[144], which could be used to develop a more potent inhibitor if the peptide is binding through this interaction.

4.2 SPPS of Non-Natural Analogues

A library of 44 cyclic peptides was synthesised using SPPS based on four backbones: CP11, CP11A4, CP11A5 and CP11A8. For each backbone, the tyrosine residue was substituted for one of 11 unnatural aromatic analogues, Figure 44. The peptides were synthesised in bulk for each backbone until the position of the tyrosine analogue at which point the resin was split and then the different aromatic analogues were synthesised in parallel. The CP11 analogues were synthesised in bulk on a glycine Wang resin until after the tryptophan residue (*resin*-GSVNW), at which point the resin was split and then the tyrosine analogues and the remainder of the amino acids were coupled on the resin.

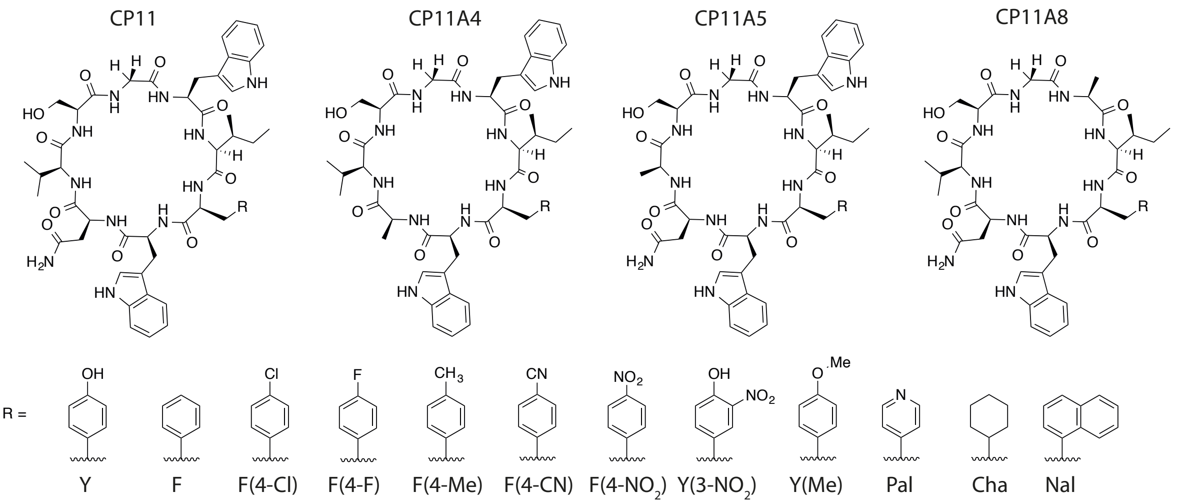


Figure 44: Structure of the Four Backbones and the Unnatural Aromatic Analogues.

The CP11A4 analogues were synthesised in bulk using an isoleucine Wang resin until the tryptophan residue (*resin*-IWGSVAW) and then the tyrosine analogues were coupled onto the peptide. The CP11A5 analogues were synthesised in bulk using an isoleucine Wang resin until the tryptophan residue (*resin*-IWGSANW) and then the tyrosine analogues were coupled onto the peptide. The CP11A8 analogues were synthesised in bulk using a glycine Wang resin until after the tryptophan residue (*resin*-GSVNW) and then the tyrosine analogues and the remainder of the amino acids were coupled onto the peptide. The linear peptides were cleaved from the resins using a cleavage cocktail of TFA, TIS and water. The peptides were then cyclised in solution under high dilution in the presence of HOBt and EDC and then purified using preparative HPLC.

Although the peptides were purified to the highest possible purity by collecting only one peak identified on the preparative HPLC trace, upon analytical HPLC analysis some of the purified peptides were found to be a mixture of epimers with peaks that greatly overlapped. It was not possible to purify the peptides further as the epimers co-eluted using the preparative HPLC column, and purification was not possible using the analytical column as it would be too time and resource-consuming to purify approximately 20 mg of each peptide, as each injection should contain less than 1 mg of peptide to avoid the possibility of the peptide precipitating on the column under aqueous conditions. Although HOBt can decrease the likelihood of epimerisation, it is still possible to synthesise epimers, it can occur at any of the coupling step or at the cyclisation step. The presence of epimers could affect the inhibitory effect of the peptides, however, it was decided to attempt to test the peptides in the ELISA to see if any of the analogues had improved activity compared to CP11.

4.3 Testing in the p6-UEV ELISA

The library of unnatural aromatic analogues was then assayed in triplicate using the p6-UEV ELISA. The cyclic peptides containing unnatural aromatic analogues based upon the CP11 backbone all demonstrated a similar or improved IC_{50} value compared to CP11, Figure 45. The cyclic peptide containing phenylalanine in the place of tyrosine showed a slight decrease in IC_{50} compared to CP11 (59.6 ± 5 µM), whilst the cyclic peptide containing cyclohexyl-alanine (Cha) demonstrated a 3.5 fold improvement in inhibitory activity (IC_{50} 13.6 ± 0.8 µM), indicating that the loss of activity seen with the phenylalanine analogue may not be due to a loss in hydrogen-bonding interaction. The analogues with the most improvement in inhibitory activity contained the following residues: 4-fluorophenylalanine (F(4-F)), 4-chlorophenylalanine (F(4-Cl)), and 4-methylphenylalanine (F(4-Me)) with IC_{50}s of 18.9 ± 2 µM, 16.7 ± 6 µM, and 12.8 ± 1 µM, respectively.

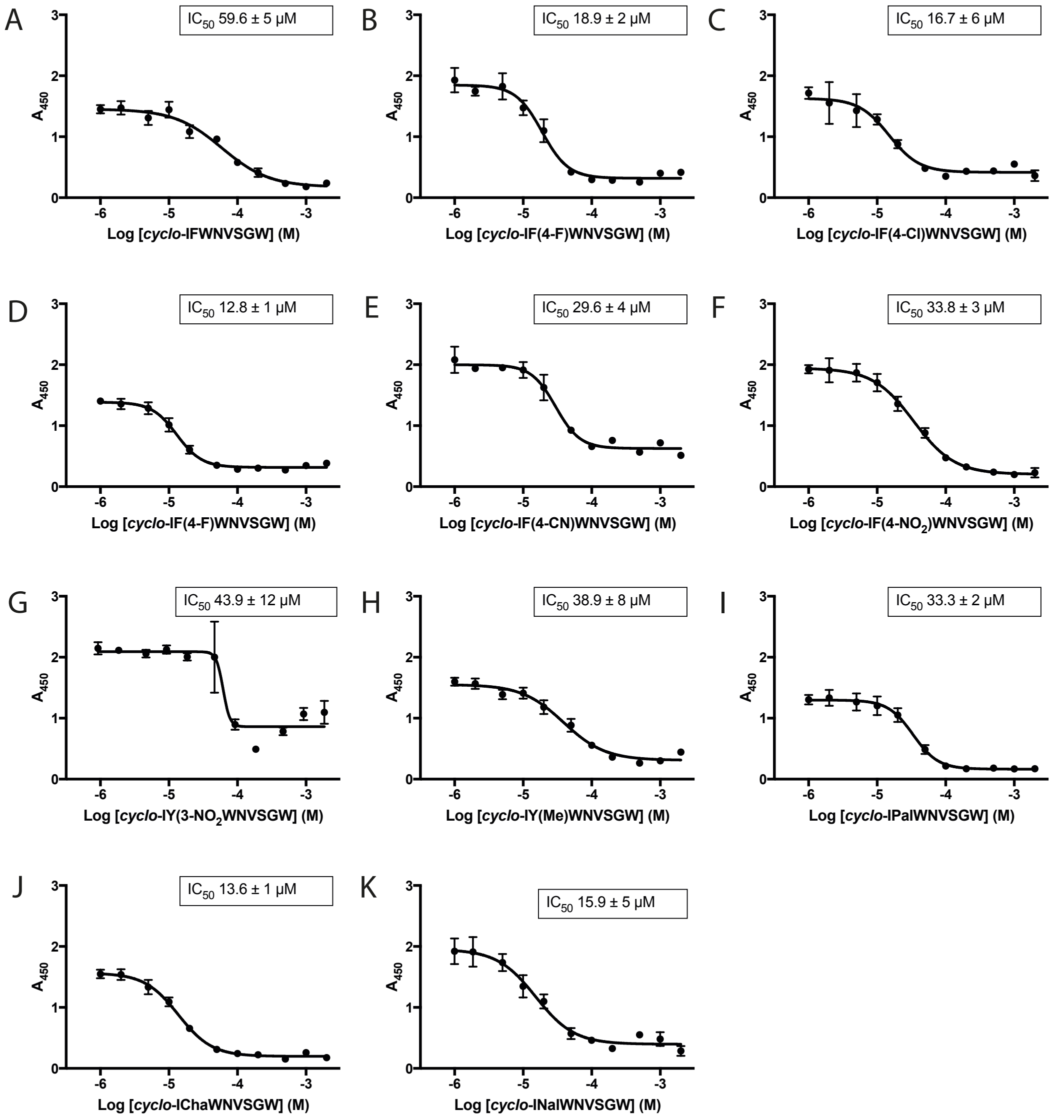


Figure 45: Graphs of ELISA Results with Tyrosine Analogues Based upon the CP11 Backbone. Each peptide was assayed in triplicate. Curves was fitted using GraphPad Prism7 Software using a Nonlinear Sigmoidal, 4PL curve fit where X is log(concentration of peptide). Error bars represent the standard deviation. In each case the tyrosine residue was exchanged for A. Phenylalanine (F). B. 4-fluorophenylalanine (F(4-F)). C. 4-chlorophenylalanine (F(4-Cl)). D. 4-methylphenylalanine (F(4-Me)). E. 4-cyanophenylalanine (F(4-CN)). F. 4-nitrophenylalanine (F(4-NO$_2$)). G. 3-nitrotyrosine (Y(3-NO$_2$)). H. methyltyrosine (Y(Me)). I. 4-pyridylalanine (Pal). J. cyclohexyl-alanine (Cha). K. (1-naphthyl)-L-alanine (Nal).

The cyclic peptide analogues based upon the CP11A4 backbone all demonstrated an improvement in IC$_{50}$ compared to CP11, but remained comparable to the inhibitory activity of CP11A4 (25.8 ± 4 µM). However, the analogue containing the 4-pyridylalanine (Pal) residue, had a similar IC$_{50}$ to CP11 (50.1 ± 3 µM), which is a 2-fold loss in activity compared to CP11A4, Figure 46. The analogues with the most improvement in inhibitory activity contained the following residues: F(4-Cl), F(4-Me), or Cha with IC$_{50}$s of 6.2 ± 0.2 µM, 8.2 ± 0.20 µM, and 6.9 ± 0.6 µM, respectively. There is a large initial error bar for the first point in the ELISA data for the peptide containing the

3-nitrotyrosine (Y(3-NO$_2$)) residue, Figure 46G, this is due to a pipetting error where the multichannel pipette did not dispense the correct volume of TMB-ELISA, but it was not noticed in time to rectify. The data for the peptide containing the 4-pyridylalanine (Pal) residue do not demonstrate a true sigmoidal curve as the final plateau starts to increase with the three highest concentrations. This could be due to the insolubility of the peptide; if the peptide has started to precipitate out of solution this could allow the proteins to interact, thus the signal starts to increase at higher concentrations.

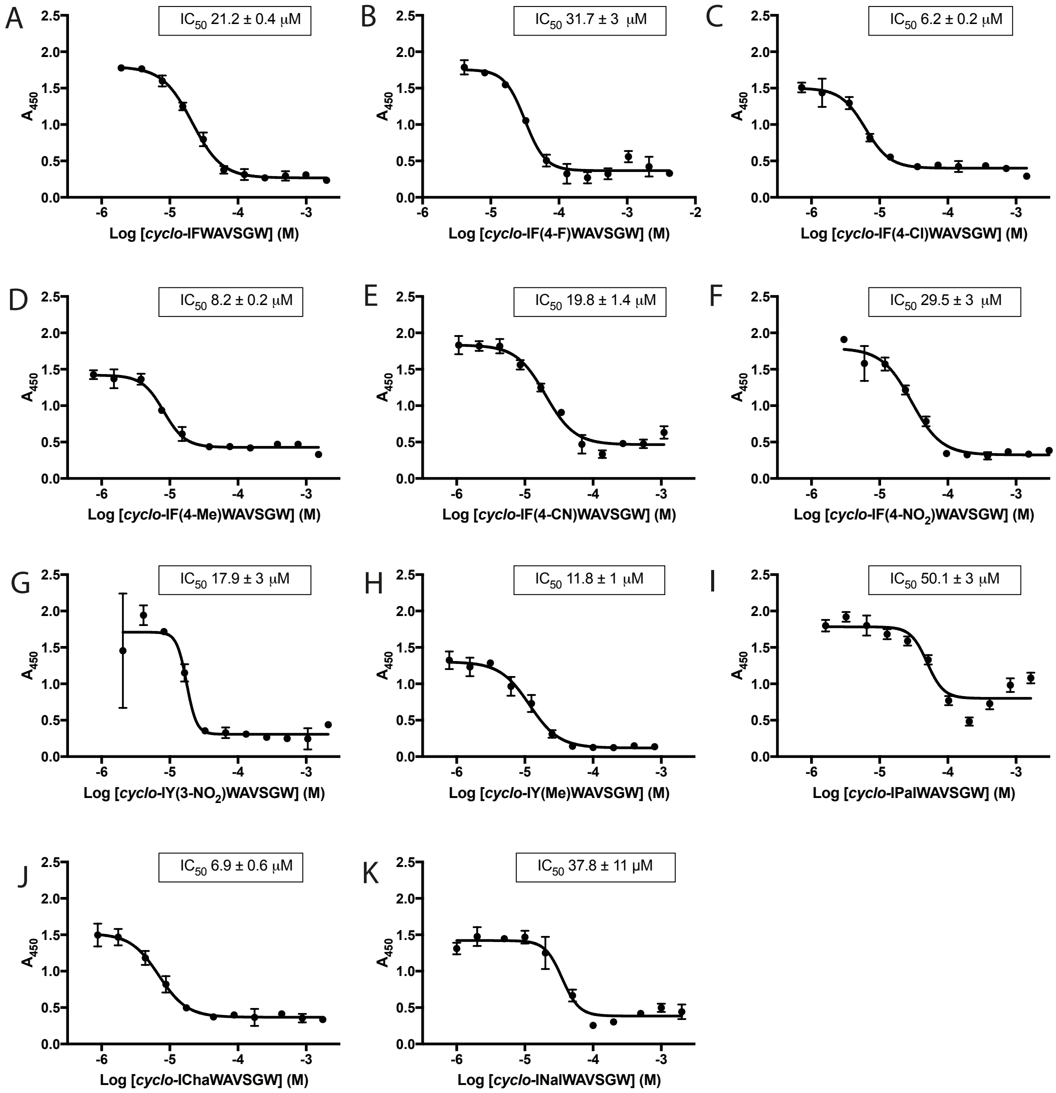


Figure 46: Graphs of ELISA Results with Tyrosine Analogues Based upon CP11A4 Backbone. Each peptide was assayed in triplicate. Curves was fitted using GraphPad Prism7 Software using a Nonlinear Sigmoidal, 4PL curve fit where X is log(concentration of peptide). Error bars represent the standard deviation. In each case the tyrosine residue was exchanged for A. Phenylalanine (F). B. 4-fluorophenylalanine (F(4-F)). C. 4-chlorophenylalanine (F(4-Cl)). D. 4-methylphenylalanine (F(4-Me)). E. 4-cyanophenylalanine (F(4-CN)). F. 4-nitrophenylalanine (F(4-NO$_2$)). G. 3-nitrotyrosine (Y(3-NO$_2$)). H. methyltyrosine (Y(Me)). I. 4-pyridylalanine (Pal). J. cyclohexyl-alanine (Cha). K. (1-naphthyl)-L-alanine (Nal).

Most of the peptides based upon the CP11A5 backbone demonstrated an improvement in IC_{50} compared to CP11 and comparable IC_{50} to CP11A5 (28.7 ± 11 μM), except for those containing 4-nitrophenylalanine (F(4-NO_2)), methyltyrosine (Y(Me)), and 4-pyridylalanine (Pal), Figure 47.

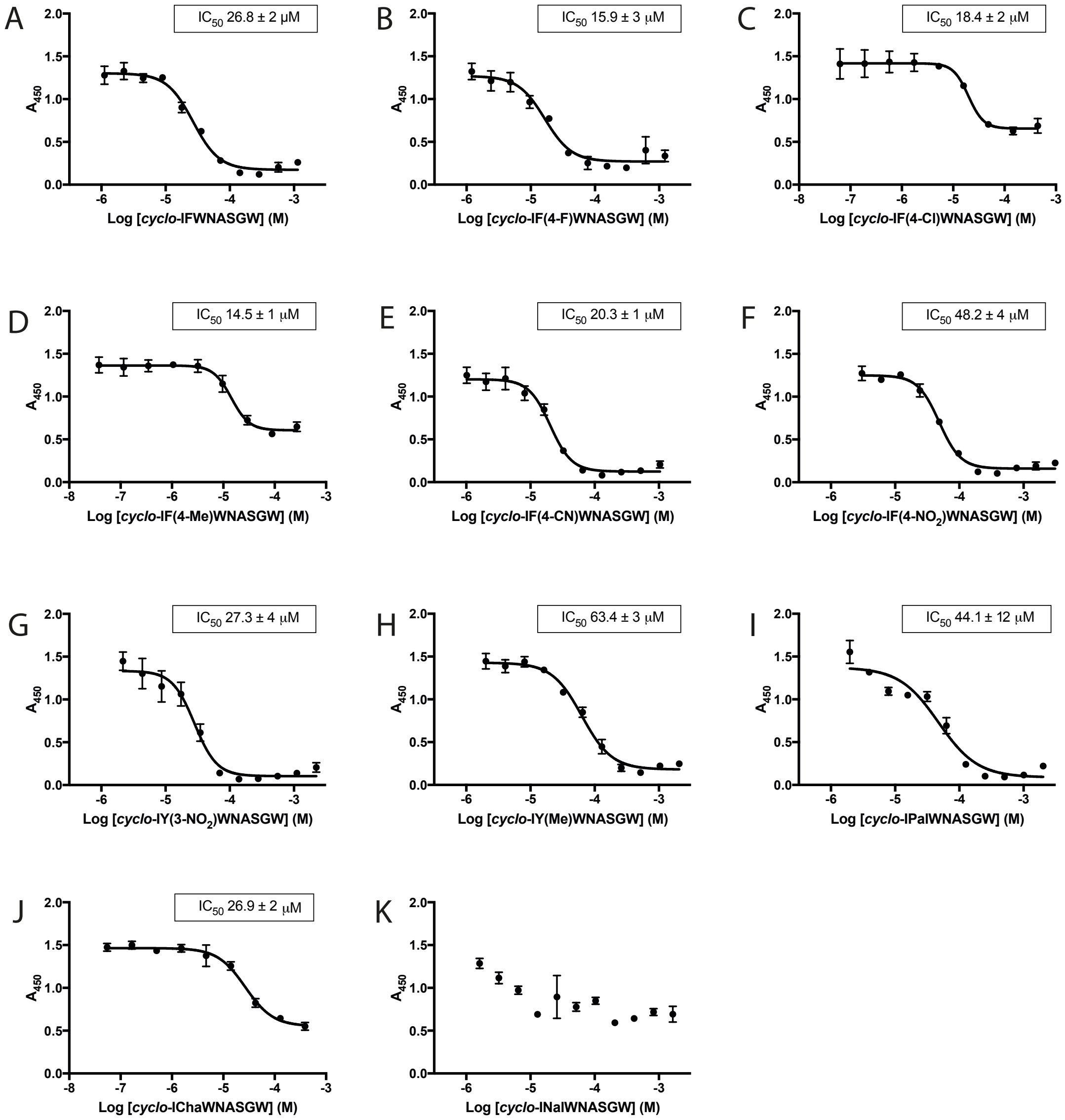


Figure 47: Graphs of ELISA Results with Tyrosine Analogues Based upon CP11A5 Backbone. Each peptide was assayed in triplicate. Curves was fitted using GraphPad Prism7 Software using a Nonlinear Sigmoidal, 4PL curve fit where X is log(concentration of peptide). Error bars represent the standard deviation. In each case the tyrosine residue was exchanged for A. Phenylalanine (F). B. 4-fluorophenylalanine (F(4-F)). C. 4-chlorophenylalanine (F(4-Cl)). D. 4-methylphenylalanine (F(4-Me)). E. 4-cyanophenylalanine (F(4-CN)). F. 4-nitrophenylalanine (F(4-NO_2)). G. 3-nitrotyrosine (Y(3-NO_2)). H. methyltyrosine (Y(Me)). I. 4-pyridylalanine (Pal). J. cyclohexyl-alanine (Cha). K. (1-naphthyl)-L-alanine (Nal).

Three analogues with the most improvement in inhibitory activity contained the following residues: 4-fluorophenylalanine (F(4-F)), 4-chlorophenylalanine (F(4-Cl)), and 4-methylphenylalanine (F(4-Me)) with IC_{50}s of 15.9 ± 3 μM, 18.4 ± 2 μM, and 14.5 ± 1 μM, respectively. The peptide containing the Nal residue demonstrated no inhibitory activity of the

PPI. The analogues containing the 4-chlorophenylalanine (F(4-Cl)), and 4-methylphenylalanine (F(4-Me)) and cyclohexyalanine residues, Figure 47C, D and J, were completed on a different occasion to the other analogues and therefore their lower plateaus are at higher values than the other analogues as they were incubated with the TMB-ELISA solution for a slightly longer time, causing a darker colour to develop and therefore the absorbance to be more intense.

The analogues based on CP11A8 demonstrated little to no inhibitory activity in the ELISA assay, Figure 48. This was thought to be due to a lack of solubility of these analogues at higher concentrations.

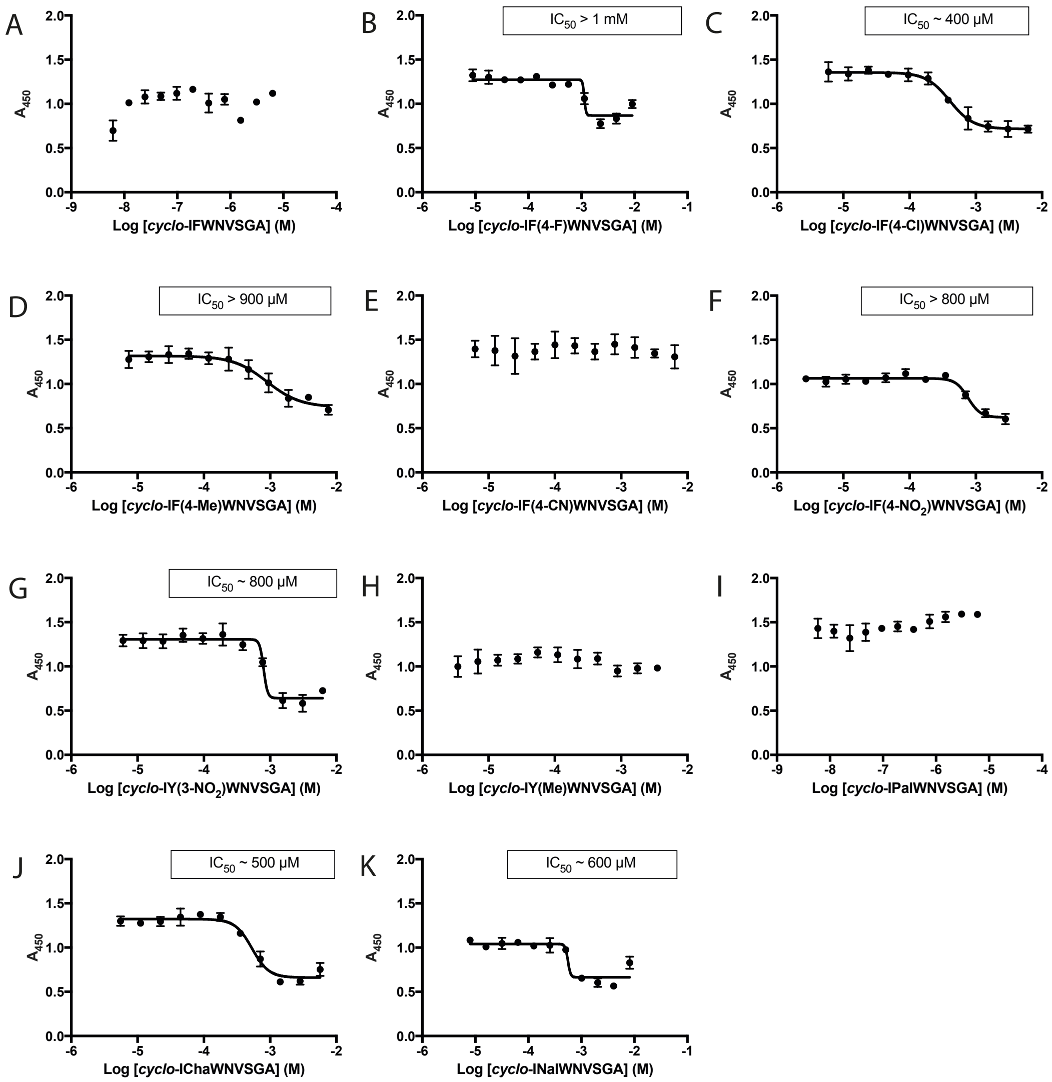


Figure 48: Graphs of ELISA Results with Tyrosine Analogues Based upon CP11A8 Backbone. Each peptide was assayed in triplicate. Curves was fitted using GraphPad Prism7 Software using a Nonlinear Sigmoidal, 4PL curve fit where X is log(concentration of peptide). Error bars represent the standard deviation. In each case the tyrosine residue was exchanged for A. Phenylalanine (F). B. 4-fluorophenylalanine (F(4-F)). C. 4-chlorophenylalanine (F(4-Cl)). D. 4-methylphenylalanine (F(4-Me)). E. 4-cyanophenylalanine (F(4-CN)). F. 4-nitrophenylalanine (F(4-NO_2)). G. 3-nitrotyrosine (Y(3-NO_2)). H. methyltyrosine (Y(Me)). I. 4-pyridylalanine (Pal). J. cyclohexyl-alanine (Cha). K. (1-naphthyl)-L-alanine (Nal).

It was observed that a high proportion of the analogues synthesised were sparingly soluble in aqueous buffer supplemented with 10% DMSO, and the CP11A8 tyrosine analogues were less soluble than the analogues based upon the other backbones. Although both tryptophan and alanine have hydrophobic side chains, tryptophan is able to form hydrogen bonds via the indole nitrogen, which when combined with the decreased polarity of some of the phenylalanine analogue side chain could explain why the peptides were less soluble and therefore much less potent than expected. Although the concentrations of the CP11 analogues chosen for ELISAs were concentrations at which the peptides were soluble in the buffer, there is the possibility that the peptides were only sparingly soluble and therefore the concentrations of peptide were far less than calculated and gave rise to poor IC_{50} values. The concentration of DMSO was not increased further for these peptides as higher concentrations have been shown to have a detrimental effect on proteins[108] and the results would not be comparable without in depth analysis of the effects of DMSO on each protein. Alternatively, this loss of inhibitory activity could be attributed to a significant change in conformation causing the active motif to be altered away from the desirable conformation.

Almost all of the analogues based on the CP11A4, CP11A5 and CP11 backbones exhibited greater inhibitory activity than CP11 itself, Figure 49. There was not always correlation between the backbones with the same unnatural tyrosine residue, indicating that the difference in conformation of the backbone has as much of an effect as the unnatural aromatic residue on the activity of the peptide. The loss of the hydroxyl group on peptides containing a phenylalanine residue in the place of tyrosine did not result in a significant loss of activity compared to the original cyclic peptide. This suggests that the peptide is not interacting with the protein through a hydrogen bonding interaction.

For most of the analogues, there were only slight differences in IC_{50} that were less than 2-fold different from the IC_{50} of CP11. These are not significant and could be due to differences between the purities of the peptides, or due to the presence of epimers. These differences in IC_{50} fall within the error range of the ELISA assay itself, or could be due to slight differences in the solubilities of the peptides causing a hydrophobic effect on the protein interfaces rather than a specific interaction.

As the fold difference is less than 2-fold, no quantitative discussion can be made about these data. It is not possible to gain any information about the thermodynamic nature of this interaction nor thermodynamic binding information pertaining to peptides from the results of the experiments due to the nature of this assay. Other experiments would have to be completed to find out this kind of information.

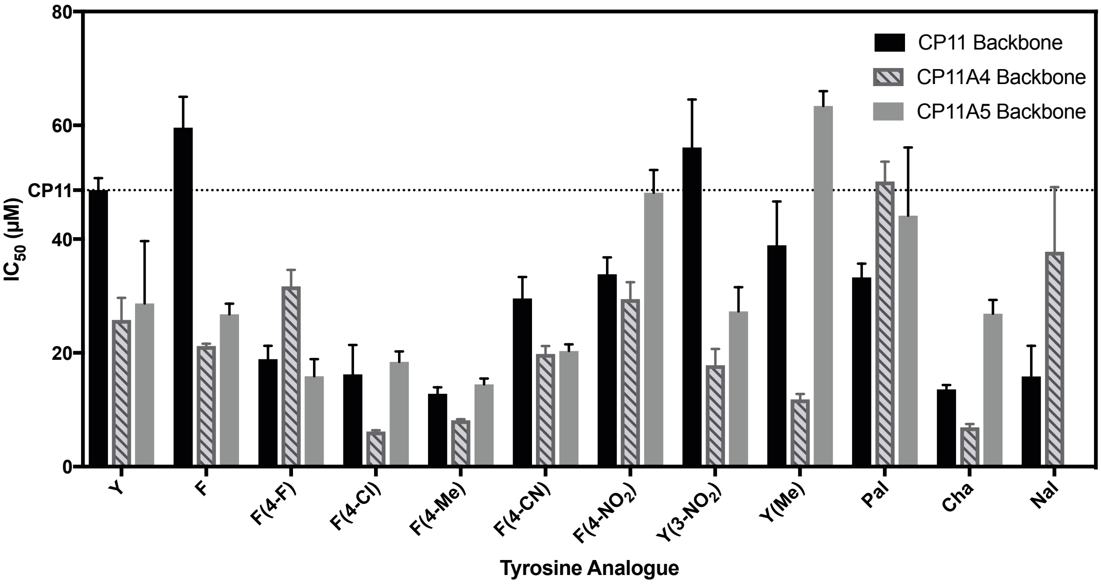


Figure 49: ELISA Results of CP11 SARs Investigating the Tyrosine Residue. Graph showing the IC_{50}s of each tyrosine analogue calculated from triplicate ELISA results. IC_{50}s of tyrosine analogues of CP11A8 were excluded from this table, as the calculated IC_{50}s were sufficiently large enough that the results of the other analogues could not be viewed.

Overall no conclusions can be drawn from these data about potential interactions between the protein interface and the peptide, however, three cyclic peptides were identified that had a 5-7 -fold improvement in inhibitory activity when compared to CP11. These analogues were all based upon the CP11A4 backbone and contained F(4-Cl) (CP11A4Y3), F(4-Me) (CP11A4Y4) or Cha (CP11A4Y10) in the place of the tyrosine residue.

4.3.1 Testing Cysteine Derivatives of Most Active Peptides

Cysteine derivatives of CP11A4Y3 (*cyclo*-IF(4-Cl)WAVSGW), CP11A4Y4 (*cyclo*-IF(4-Me)WAVSGW and CP11A4Y10 (*cyclo*-IChaWAVSGW) were synthesised by SPPS for easy tagging with a Tat-peptide, or other cell-penetrating peptide for improved cell permeability in future cell-based assays. Tat is a commonly used 12 amino acid long cell penetrating peptide that is derived from the Tat sequence of HIV. It facilitates the uptake of non-permeable molecules through the plasma membrane by endocytosis[145]. As the serine residue of CP11 was not determined to be important for inhibiting the p6-UEV interaction during the alanine scan, it suggested that it could be substituted for cysteine without resulting in a loss of inhibitory activity. The cysteine derivatives were found to have comparable IC_{50}s to the original analogues by ELISA, Figure 50 and Table 9, indicating that they could be substituted for the original peptides in future assays that might require tagged peptide with Tat or fluorophores.

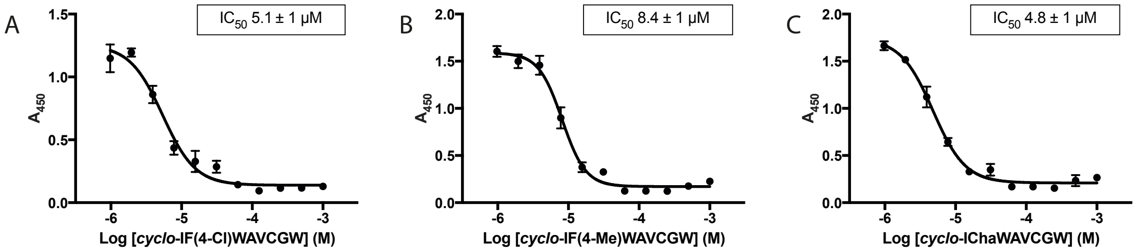


Figure 50: Results of an ELISA testing A. CP11A4Y3C, B. CP11A4Y4C and C. CP11A4Y10C. Curves were fitted using GraphPad Prism7 Software using a Nonlinear Sigmoidal, 4PL curve fit where X is log(concentration of peptide). Each peptide was assayed in triplicate. Error bars represent the standard deviation.

Table 9: Comparison of IC_{50} Values of Cysteine Derivatives and the Original Analogues

Unnatural Aromatic Residue	IC_{50} of CP11A4 Analogues (µM)	IC_{50} of CP11A4 Cysteine (CP11A4C) Analogues (µM)
F(4-Cl)	6.2 ± 0.2	5.1 ± 1
F(4-Me)	8.2 ± 0.20	8.4 ± 1
Cha	6.9 ± 0.5	4.8 ± 1

4.4 Development of Biophysical Assays to Determine the Activity of CP11 Derivatives *In Vitro*

After more potent inhibitors of the p6-UEV interaction had been developed, the next step was to determine which protein of the PPI the peptides were binding to and to investigate the strength of the binding interaction. There are many biophysical techniques that can provide information such as the affinity, on (K_{on}) and off (K_{off}) rate constants and the thermodynamics of the binding interaction. This information is important to fully understand the effect the inhibitor might have when used in *in vivo* or in further drug development.

4.4.1 Thermal Shift Assay

The Gibbs free energy of unfolding (ΔG_u) of a protein is temperature dependent and related to the protein's stability. As a protein is gradually heated, it becomes less stable and therefore its ΔG_u decreases until it reaches zero when half the protein is unfolded. The temperature at which this occurs is called the melting temperature (T_m). The ΔG_u of the protein can be affected by the free energy contribution of ligand binding which results in a change in the T_m.

Assays that investigate the changes in melting temperature of a protein are called thermal shift or differential scanning fluorimetry assays. These assays can be used to measure the binding of a ligand to a protein in the protein's native or unfolded state. The assay monitors the protein unfolding using a fluorescent dye (e.g., SYPRO Orange, ThermoFisher), which binds non-specifically to hydrophobic regions that are exposed as the protein is heated and subsequently unfolds, and this can be used to calculate the T_m of a protein, in the absence or presence of a ligand, Figure 51. The dye fluoresces upon binding to hydrophobic protein regions due to exclusion of water molecules, which have a quenching effect when the dye is unbound. The fluorescence gradually increases as the protein melts, and more molecules of the dye can bind, until a maximum fluorescence is reached. The point at which half the protein has melted is the T_m. After the maximum fluorescence has been reached the signal decreases as the protein starts to aggregate and the dye dissociates at higher temperatures[146-147].

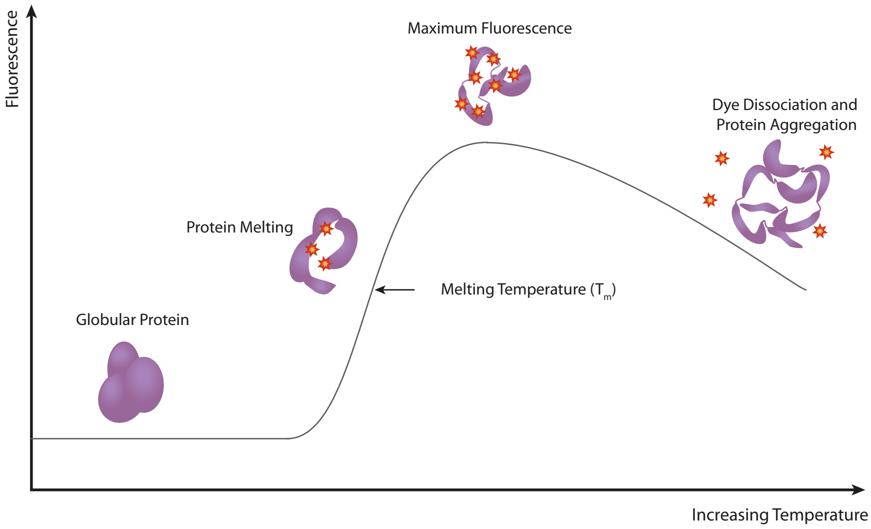


Figure 51: Schematic Showing the Typical Melting Profile of a Protein. As the protein melts, the fluorescent dye binds to the exposed hydrophobic regions of the protein. This causes an increase in fluorescence, as the dye is no longer quenched by the presence of water molecules, until the protein is completely unfolded, and the maximum amount of dye is bound. The temperature at which half the protein has melted is called the T_m. As the protein is heated past the maximum fluorescence, the fluorescence decreases as the protein aggregates and the dye dissociates.

By performing the assay with varying concentrations of ligand, information can be gained about the binding interaction. If a ligand binds to the native state of the protein, it might have a stabilising effect on the protein, which would be indicated by an increased ΔT_m[148], whereas if a ligand binds to the denatured state of the protein, it may destabilise the protein and lead to a decreased ΔT_m, although this is not always applicable to all interactions[149].

4.4.1.1 Optimisation of Thermal Shift Assay

The first step in optimising the thermal shift assay was to determine the concentration of each protein to be used in the assay. Increasing concentrations of each protein (0-10 µM) were diluted in the assay buffer and then mixed with SYPRO Orange (1 µL, 15X). This concentration of SYPRO Orange was chosen as it was a standard concentration utilised in the laboratory to give a good signal on the plate reader and fell within the recommended concentration as suggested by the supplier. The fluorescence of the proteins was measured using a BioRad CFX Connect Real-Time System and the system set up to heat from 30 °C to 95 °C at a rate of 1 °C/ 30 seconds with a fluorescence reading measured at $\lambda_{ex}/\lambda_{em}$: 300, 470 nm/ 570 nm every 30 seconds. The results were analysed using GraphPad Prism and firstly the fluorescence was plotted, Figure 52A-B. The first derivative of the fluorescence was then plotted, Figure 52C-D, to show the rate of fluorescence change against the temperature, which provides a clear visual representation of the T_m as the peak of the graph. The T_m was actually calculated more precisely by fitting the melting curve to the Boltzmann equation using GraphPad Prism, where the generated V_{50} value equates to the T_m[150].

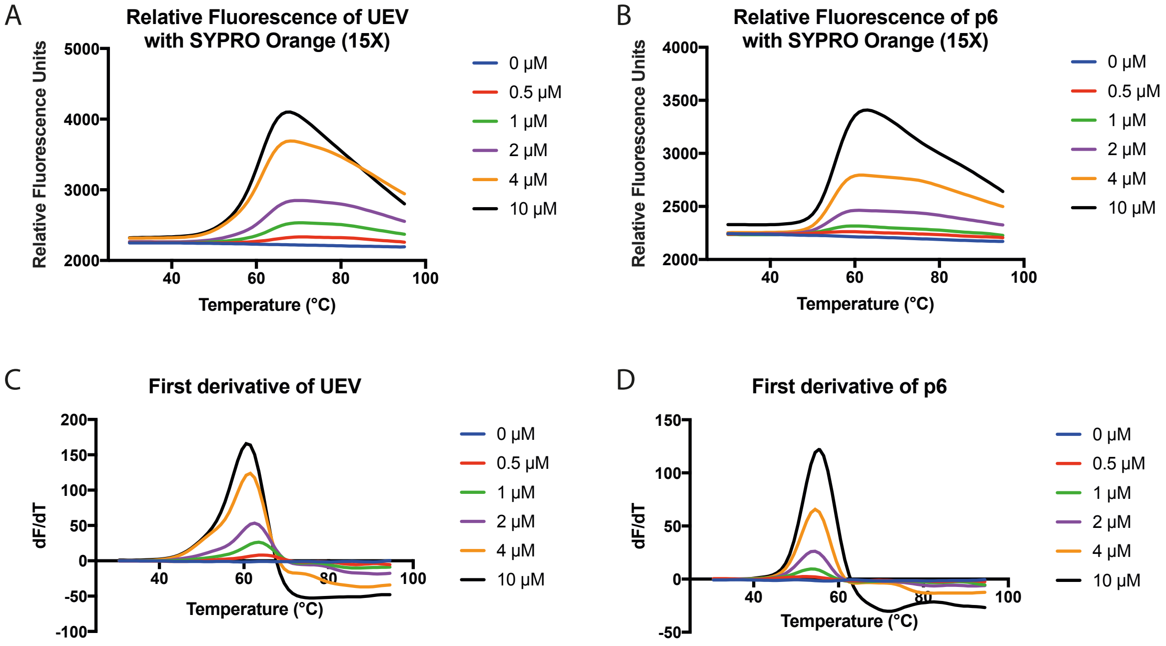


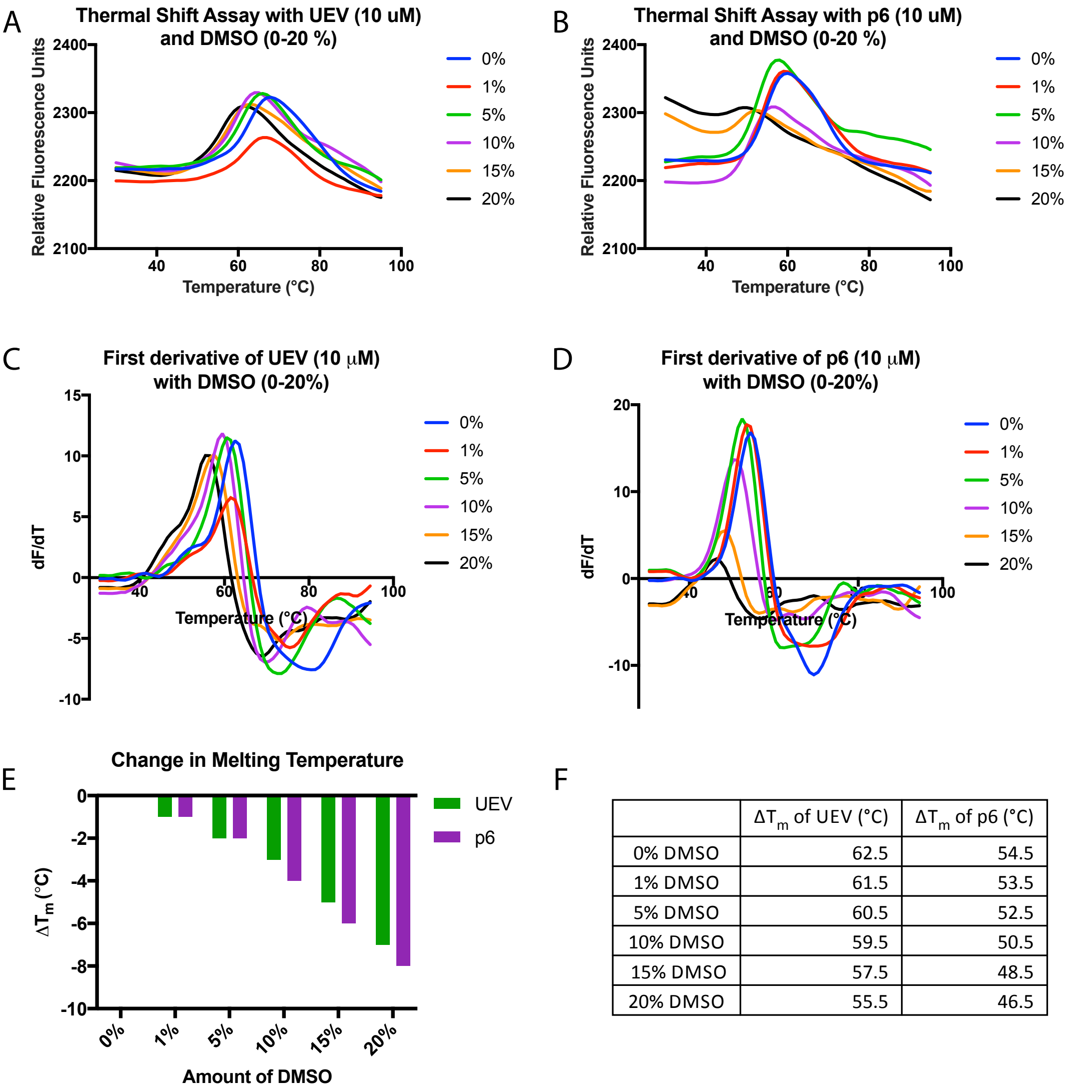
Figure 52: Thermal Shift assay with Increasing Concentrations of UEV and p6. Assay was completed with SYPRO Orange (15X). First derivative plots were generated using GraphPad Prism software. A. Relative fluorescence of thermal shift results with UEV (0-10 µM). B. Relative fluorescence of thermal shift results with p6 (0-10 µM). C. First derivative of thermal shift results with UEV (0-10 µM). B. First derivative of thermal shift results with p6 (0-10 µM).

Following on from this experiment, the concentrations of UEV and p6 were chosen for future assays to be 10 µM. This concentration of each protein gave the best fluorescence signal above the baseline. From close inspection of the first derivative plot of the thermal shift assay with UEV, Figure 52C, the melting profile appears to be biphasic, which could be due to impurities within the protein sample, and could cause the slight differences between the melting temperatures observed, as the impurity will have a melting temperature as well as the protein of interest. The melting profile for p6, Figure 52B and D, is most likely to be due solely to the GST tag, as p6 is more peptidic in nature. p6 was not cleaved from GST for this assay due to concerns about purifying the small protein from the cleaved tag using the equipment available. If the peptides demonstrate binding to the recombinant p6-GST protein, the assay would need to be completed screening the peptides against the GST-tag alone, to ensure that the peptides are specifically binding to the protein rather than the GST-tag.

Due to the poor solubility of the cyclic peptides in aqueous buffer, DMSO is required to aid solubilisation at high concentrations. In the ELISA, the p6-UEV PPI was found to be stable in the presence of up to 10% DMSO, however a range of DMSO concentrations (0-20%) were tested with each protein individually by thermal shift, Figure 53, to determine the DMSO tolerance of the proteins in this assay.

In the presence of DMSO, both p6 and UEV proteins showed a decrease in melting temperature. This means that increasing concentrations of DMSO exerts a destabilising effect on both of the proteins. In the case of p6, Figure 53B and D, high concentrations of 15-20% DMSO also resulted in a decrease in fluorescent signal, which could be due to the protein being destabilised to such a degree that the dye binds to the denatured protein before it is heated, hence causing high background levels of fluorescence and minimising the overall signal. In the presence of 10% DMSO, UEV and p6 were found to have ΔT_m of 3 °C and 4 °C, respectively. This is a significant change in melting temperature; a change in melting temperature over 1 °C is usually considered significant to indicate a change in T_m[151-152]. This demonstrates that the DMSO tolerance of the thermal shift assay is lower than that of the ELISA, which could be due to a difference in sensitivity between the assays caused by increasing temperature. Following these observations, it was decided to use an assay concentration of 10% DMSO in order to maximise peptide solubility, as the peptides are inherently insoluble at lower concentrations.

As discussed previously, when determining the optimal concentration of UEV, the melting curves appear to be biphasic, Figure 53A and C, and as these experiments were completed with the same batch of purified protein, it was concluded that this was due to the same impurities that have a slightly different melting temperature.



	ΔT_m of UEV (°C)	ΔT_m of p6 (°C)
0% DMSO	62.5	54.5
1% DMSO	61.5	53.5
5% DMSO	60.5	52.5
10% DMSO	59.5	50.5
15% DMSO	57.5	48.5
20% DMSO	55.5	46.5

Figure 53: Thermal Shift Assays with UEV and p6 and Increasing Concentrations of DMSO. Assays were completed with SYPRO Orange (15X) and 10 µM each of either UEV or p6. First derivative plots were generated using GraphPad Prism software. Melting temperatures were determined by fitting the melting curve to the Boltzmann equation and the V_{50} value was taken to be the T_m. A. Relative fluorescence melting curves of UEV (10 µM) with 0-20% DMSO. B. Relative fluorescence melting curves of p6 (10 µM) with 0-20% DMSO. C. First derivative of thermal shift results with UEV and 0-20% DMSO. D. First derivative of thermal shift results with p6 and 0-20% DMSO. E. Graph showing the destabilisation effect of increasing concentrations of DMSO on UEV and p6. F. Table showing the change in melting temperature of UEV and p6 with increasing concentrations of DMSO.

4.4.1.2 Thermal Shift Assay with CP11 Titrated Against UEV or p6

CP11 was titrated against a constant concentration of either UEV or p6 in the presence of SYPRO Orange. The melting temperatures of the proteins were then determined, Figure 54.

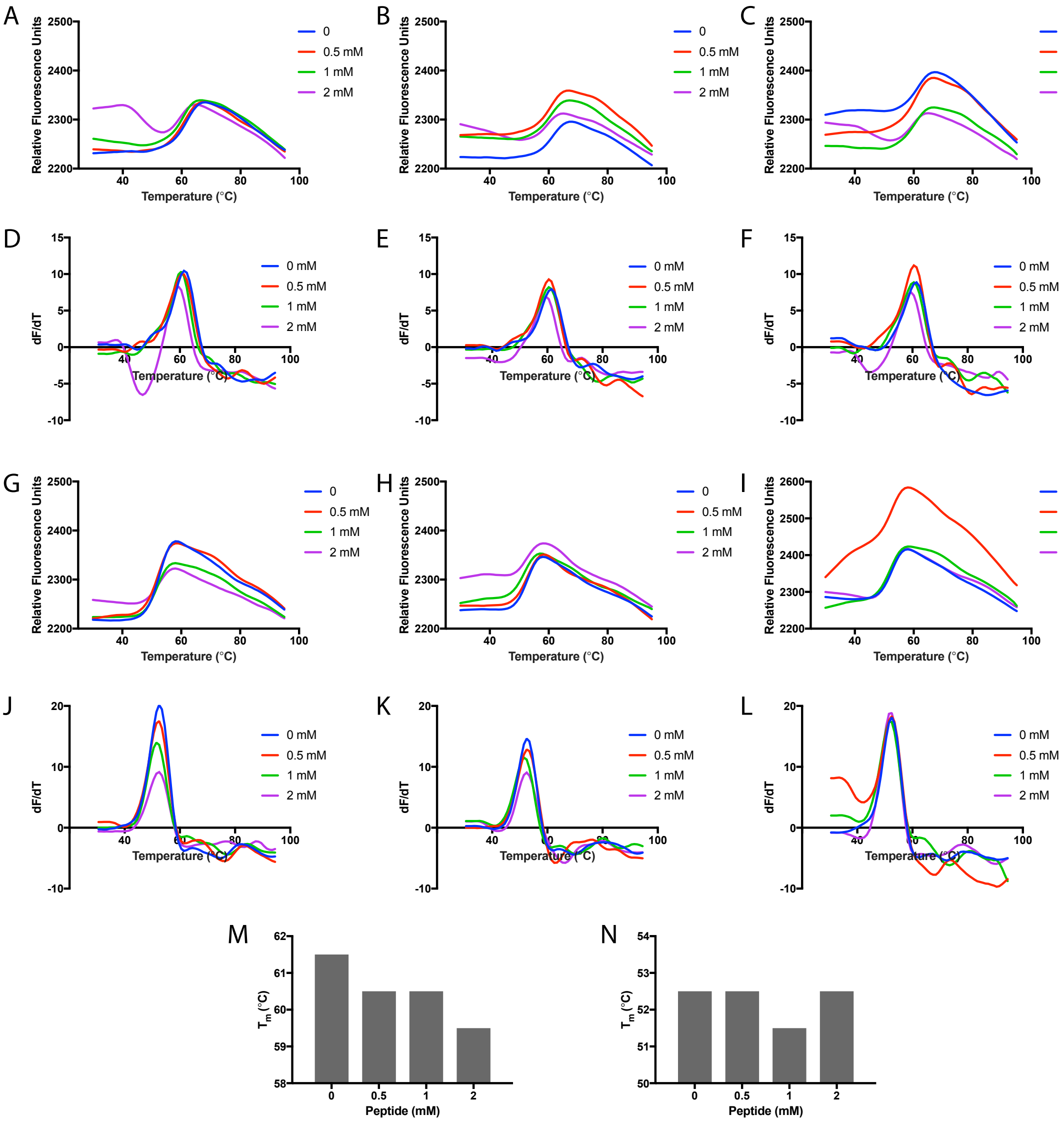


Figure 54: Thermal Shift Assays with UEV and p6 and Increasing Concentrations of CP11. Assays were completed with SYPRO Orange (15X) and 10 µM each of either UEV or p6. First derivative plots were generated using GraphPad Prism software. Melting temperatures were determined by fitting the melting curve to the Boltzmann equation and the V_{50} value was taken to be the T_m. A. Relative fluorescence melting curves of UEV (10 µM) with CP11, repeat 1. B. Relative fluorescence melting curves of UEV (10 µM) with CP11, repeat 2. C. Relative fluorescence melting curves of UEV (10 µM) with CP11, repeat 3. D. First derivative of thermal shift results with UEV and CP11 (0-2 mM), repeat 1. E. First derivative of thermal shift results with UEV and CP11 (0-2 mM), repeat 2. F. First derivative of thermal shift results with UEV and CP11 (0-2 mM), repeat 3. G. Relative fluorescence melting curves of p6 (10 µM) with CP11, repeat 1. H. Relative fluorescence melting curves of p6 (10 µM) with CP11, repeat 2. I. Relative fluorescence melting curves of p6 (10 µM) with CP11, repeat 3. J. First derivative of thermal shift results with p6 and CP11 (0-2 mM), repeat 1. K. First derivative of thermal shift results with p6 and CP11 (0-2 mM), repeat 2. L. First derivative of thermal shift results with p6 and CP11 (0-2 mM), repeat 3. M. Graph showing the dose dependent destabilisation effect of increasing concentrations of CP11 on UEV, triplicate results. N. Graph showing the minimal effect of increasing concentrations of CP11 on p6, triplicate results. There are no visible error bars in this graph due to the similarities of the data.

A high relatively fluorescence was observed in the third repeat of the thermal shift assay with p6 and CP11 (0.5 µM), Figure 54I. This was due to a pipetting error when adding the SYPRO Orange to the plate. In all cases there is a high initial fluorescence in the assays completed with 2 mM CP11, this is due to the natural fluorescence of CP11. No effect was observed that was greater than 1 °C on the melting temperature of p6, however a small dose-dependent temperature change (2 °C with CP11 at 2 mM) was observed when the peptides were titrated against UEV. This indicates that the peptides are affecting UEV in some way, rather than p6. However, protein destabilisation is not indicative of a binding interaction and could be caused by a non-specific destabilising interaction. If this is the case, this assay is not suitable for fully understanding this interaction. Equally, only very small changes in melting temperature were observed when CP11 was titrated against either UEV or p6 which could indicate that this assay is not suitable for investigating potential binders of these proteins and therefore a different assay was investigated.

4.4.2 Microscale Thermophoresis

Microscale thermophoresis (MST) is a biophysical technique that can be used to measure interactions of proteins with small molecules or other biomolecular interactions including DNA, RNA and enzymatic activity[153-156]. Thermophoresis is the movement of molecules in aqueous solution induced by a temperature gradient. This phenomenon was first described by Ludwig in 1856 and since then has been the subject of much research. The thermophoretic effect of molecules is greatly affected by particle charge and changes in the solvent particle interface. This can be explained by the Soret coefficient, which assumes that the movement of molecules in a temperature gradient can be described using linear drift response. Therefore, the molecular velocity of a molecule (v) is linearly dependent on the temperature gradient (gradT):

$$v = -D_{Ti} gradT$$

Where D_{Ti} is a proportionality constant that describes the thermophoretic mobility. In a steady state, the linear drift will be counterbalanced by mass diffusion of molecules, therefore for the low concentrations that will be used in MST assays the balance of molecular flow densities in a temperature gradient can be described as:

$$j_i = -c_i D_{Ti} gradT - D_i gradc_i$$

Where D_i refers to the diffusion coefficient and c_i refers to the concentration of molecules. If this equation is fully integrated in terms of D_{Ti} and D_i the steady state concentration of molecules becomes:

$$c_{Ti} = c_i \exp\left(-S_{Ti} dT\right)$$

Where the temperature is increased by dT above the ambient temperature. The Soret coefficient, S_{Ti}, is therefore defined as:

$$S_{Ti} = \frac{D_{Ti}}{D_i}$$

In most cases, the concentration of molecules in a temperature gradient is less than the initial concentration, which gives a positive D_{Ti} value and hence a positive Soret coefficient. However, the Soret coefficient is dependent on the molecule size, charge and conformation as it can change based on the nature of the interaction between the molecule and the solvent. Therefore, when the changes in thermophoretic motion are measured by MST, they provide information about interactions between biomolecules and how they are affected by binding.

The binding constant is determined by titrating the binding partner into a constant concentration of the fluorescently labelled molecule of interest. Any binding event will cause a change in the amplitude of the MST signal, which can be quantified using:

$$F_{norm} = (1 - x)F_{norm}(unbound) + xF_{norm}(bound)$$

Where F_{norm}(unbound) refers to the normalised fluorescence of unbound, labelled molecules, F_{norm}(bound) refers to the normalised fluorescence of the complex formed by the binding of the ligand to the molecule of interest, and x refers to the fraction of labelled molecules bound to the ligand. By measuring the difference between F_{norm}(bound) and F_{norm}(unbound), x and the dissociation constant can be determined.

In an MST assay, the samples of interest are loaded into capillary tubes and the temperature gradient is provided by an IR-Laser focused on the sample. The IR-Laser is coupled into the path of fluorescence/emission using an IR dichroic mirror to allow for the excitation of the fluorescent sample. The IR-Laser is able to focus on the sample and provide a localised temperature increase in a microscopic area. After an initial increase in temperature in the area (MST temperature-Jump or T-Jump), thermophoresis occurs and a concentration gradient forms in the sample. Potential inhibitors can be titrated into the sample of interest and changes in MST signal can be measured, with these data being used to provide information about the binding interaction.

The MST signal can be described as five main events: initial state, T-Jump, thermophoresis, steady state and back-diffusion, Figure 55. The initial state is the fluorescence of the sample without heating. The T-jump state is the change in fluorescence yield of the dye observed in the sample induced by heating before thermophoresis begins to take effect. Thermophoresis is the observable change in fluorescence caused by thermophoretic effect. The final state is the diffusion of the biomolecules to regain a homogeneous concentration after the laser has been switched off. Small changes in each of these states can provide information about the mechanism of binding.

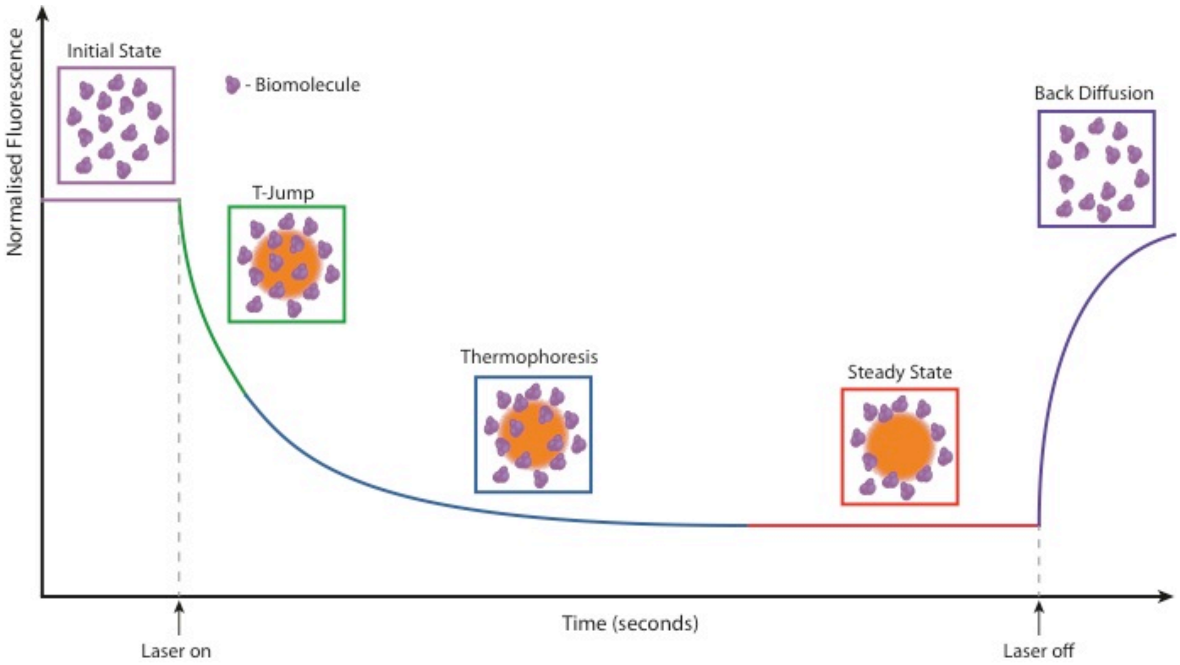


Figure 55: Schematic Showing the Five Main Events in an MST Signal. Prior to the start of the experiment, the fluorescence of the sample is constant in the initial state. The IR laser is switched on and the decrease in fluorescence is due to the change in fluorescent yield of the dye, prior to the thermophoretic movement of the molecules during the thermophoresis state. Eventually a steady state is reached. Then the IR laser is switched off and the increase in fluorescence indicates the diffusion of the molecules back to a homogenous state akin to the initial state.

Changes in the T-jump state can indicate a binding mode close to the position of the fluorescent dye or it could be due to an induced conformational change in the biomolecule. If a change is observed in the thermophoresis state, this is more likely to be due to a binding mode that affects the thermophoretic mobility of the biomolecule, for example affecting the charge, overall size or the solvation shell. If changes are observed in both T-jump and thermophoresis states, these states can be analysed in combination. Changes in the back diffusion can be indicative of aggregation if the size of the biomolecule is not expected to change to a large extent.

4.4.2.1 Optimisation of MST Assay

It was assumed that the peptides would be interacting with p6 rather than UEV as p6 is a significantly smaller protein than UEV. p6 is one of the smallest lentiviral proteins at only 52 amino acids in length, reducing the surface area available for binding any ligands. It has been found to adopt a helix-flexible-helix conformation[157], Figure 56, with the second defined helix important for binding to the virus budding factor AIP-1/ALIX. However, the PTAP motif that is important for binding to TSG101 is located at the N-terminal, which does not exhibit a secondary structure. It would be highly unlikely for the peptide to be interacting with this motif to prevent binding rather than binding to UEV. For this reason, it was decided to investigate the effects the peptides had on UEV by MST.

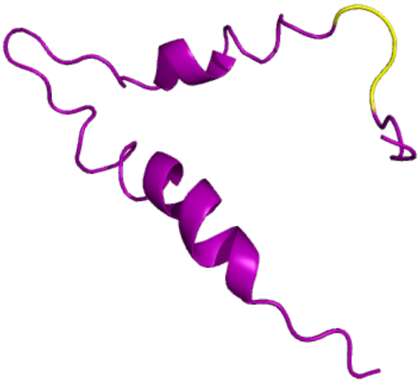


Figure 56: Structure of p6 Drawn in MacPyMOL. PTAP motif is highlighted in yellow and can be seen at the N-terminus. PDB: 2C55[157].

4.4.2.1.1 Labelling Optimisation

In an MST assay, one of the binding partners must be fluorescently labelled so that the changes in fluorescence can be observed. This can be either the protein of interest or the ligand. As fluorophores can be quite bulky, and if used to label the cyclic peptide, could potentially inhibit the binding or influence the conformation either adversely or favourably and provide a false result. Therefore it was decided to fluorescently label the UEV protein. The two dye options for fluorescent labelling provided with the instrument (NanoTemper) were an NHS-dye for labelling lysine side-chains or a maleimide dye for labelling cysteine side-chains. The dyes were purchased from NanoTemper, but the structures of the dyes are undisclosed. The use of MacPyMOL software to analyse the UEV protein (PDB: 3OBU)[41] for surface-exposed lysine and cysteine residues showed that there were many lysine side-chains, but there were only a two surface cysteine residues that could be labelled, Figure 57. The main issue with labelling the lysine side-chains is that due to the number of solvent-exposed lysine residues it would be more likely that you would obtain a mixed sample of single-, double- and multiple-labelled protein that could behave differently and mask any effect seen in the MST-signal. Several of the lysine side-chains are also in close proximity to the p6 binding-pocket, therefore labelling of these lysine residues could prevent the inhibitors from binding if they bind in the same pocket. Therefore, it was decided to label the surface cysteine side-chains as the likelihood of producing multiple-labelled protein is reduced through the use of less equivalents of maleimide dye. Additionally, cysteine side-chains are situated further from the binding groove of p6, and therefore the process of labelling should not prevent the inhibitors from binding.

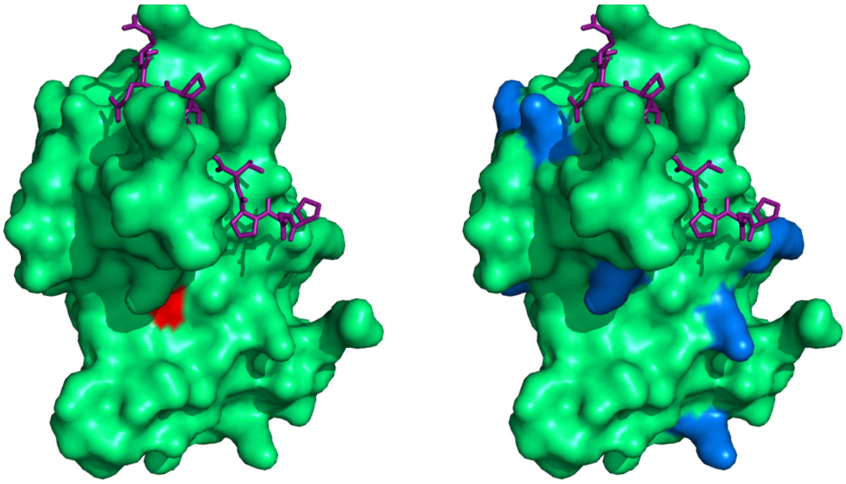


Figure 57: Surface of UEV Protein Showing Cysteine (red) and Lysine (blue) Side-Chains and p6 PTAP peptide (purple). PBD: 3OBU[41].

4.4.2.1.2 Buffer Optimisation

An important part of any assay development is optimisation of the assay buffer. Different types of buffer can have different ionic strengths, for example a phosphate buffer can contribute approximately seven times more ions to a solution than a zwitterionic Tricine buffer at the same pH[158]. The ionic strength of the buffer can affect the solubility of the protein as can the pH, which should be at least 1 unit away from the pI of the protein. In this case, UEV has a pI of 9.04 (ExPASY Bioinformatics Resource Portal) so a pH of 7.4 was chosen, as this is approximately physiological pH.

Three different buffers were produced at a pH of 7.4 with 100 mM NaCl and 5% glycerol: Tris-HCl, HEPES and MOPS, all of which have an effective pH range that covers the desired pH. When the protein was buffer-exchanged into MOPS buffer using a PD-10 desalting column, it immediately formed a visible precipitate, therefore MOPS was considered an inappropriate buffer system. The protein samples in Tris-HCl buffer and HEPES buffers showed comparable traces on the MST, however the HEPES buffer was chosen as the pH is less affected by small changes in temperature, an important consideration as the samples are heated using the IR-Laser and this might cause a change in the pH, which could affect the stability of the protein and disrupt any potential binding.

Next the capillary type for the assay was optimised; UEV was loaded into two different types of capillary (NanoTemper), standard and premium, in the same concentration, Figure 58. The premium capillary has a coating to minimise the protein sticking to the plastic walls.

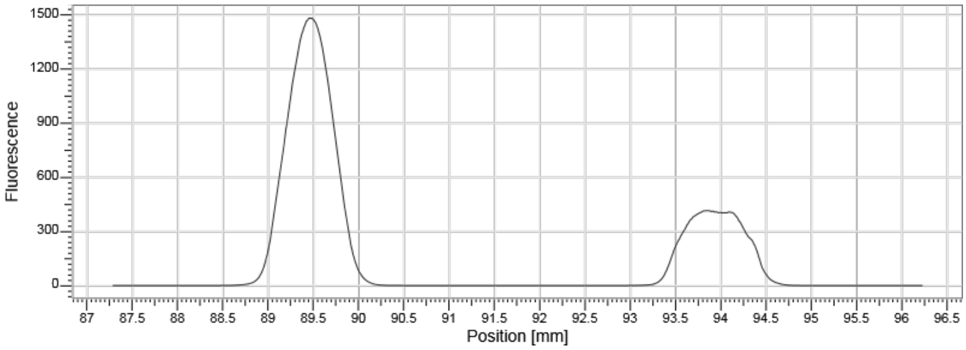


Figure 58: Capillary Scan of Fluorescently Labelled UEV in a Premium Capillary (Left) and Standard Capillary (Right) in HEPES Buffer.

It can be seen that there is a significant difference in the peak shape and measured fluorescence between the two different capillaries. The peak for the standard capillary is irregular indicating that the protein is sticking to the walls of the capillary, which is causing a decrease in fluorescence signal. The peak for the premium capillary appears uniform indicating that this is the best capillary to use for further experiments. As there is a slight irregularity to the peak for the premium capillary, 0.05% Tween-20 was added as a detergent in an attempt to minimise sticking of the sample to the capillary walls.

A reducing agent was not included in the MST buffer in order to keep the conditions similar to buffers used in previous assays and in buffers used in the literature with this protein.

4.4.3 Assessing the Binding of Most Active Peptides to UEV

After the conditions for the assay were optimised, CP11 and the most active peptides (CP11A4Y3, CP11A4Y4 and CP11A4Y10) as well as the cysteine analogue, CP11A4Y3C, were tested for binding to UEV (4.6 nM NHS-labelled UEV-His$_6$, MST power 40%, excitation power 15%), Figure 59. CP11 was found to have a Kd 142.00 ± 1.04 µM by analysis of the change in the MST T-Jump. A change in the T-Jump indicates that CP11 may bind in close proximity to the fluorophore or cause a conformational change in the protein close to the fluorophore that affects the intrinsic temperature dependence of the fluorescence. The Δ F$_{norm}$ [$^0/_{00}$] is small (less than 10 units) indicating that the peptide does not have a large effect upon the protein and confirming that the binding is weak.

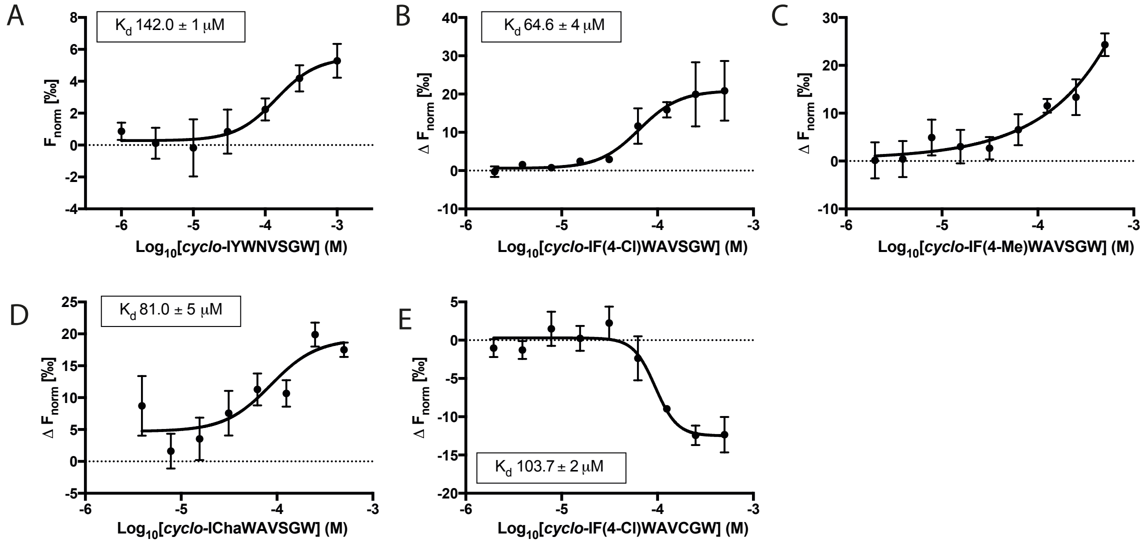


Figure 59: Graphs of MST Analysis of Most Active Peptides with UEV labelled with NHS Fluorescent Dye. Each peptide was assayed in triplicate. Curves was fitted using GraphPad Prism7 Software using a Nonlinear Sigmoidal, 4PL curve fit where X is log(concentration of peptide). Data was excluded if fluorescence fell outside of 10% threshold. Error bars represent the standard error about the mean of three independent measurements. A. Graph showing the binding curve of CP11 binding to UEV, using T-Jump analysis. B. Graph showing the binding curve of CP11A4Y3 binding to UEV using thermophoresis analysis. C. Graph showing the binding curve of CP11A4Y4 binding to UEV, using thermophoresis analysis. D. Graph showing the binding curve of CP11A4Y10 binding to UEV, using thermophoresis analysis. E. Graph showing the binding curve of CP11A4Y3C binding to UEV, using thermophoresis analysis.

The data do not reach final plateaus due to problems with solubility at high concentrations. This makes it difficult to have confidence in the calculated Kd values. However, there have been published data using MST to calculate Kd values, which do not completely reach a final plateau[154-156], which provides merit to the calculated values. From the calculated Kds, CP11A4Y3 was found to bind to UEV moderately well (Kd 64.6 ± 4 µM) by analysis of the thermophoresis state and demonstrated a 2.2-fold improvement over CP11 (Kd 142.0 ± 1 µM), although this is a small fold change which could be influenced by experimental error or discrepancies in peptide purity. As this Kd was determined from the thermophoresis state, there is an indication that the peptide is binding in such a way as to affect the size, the charge or the solvation shell of the protein and therefore the diffusion coefficient. As the peptide is quite small (molecular weight approximately 1000), it would be unlikely to change the size of the protein upon binding to significantly cause a change in the diffusion coefficient, therefore it must be changing the charge or solvation shell of the protein.

CP11A4Y4 was not soluble enough under these conditions to assay at sufficiently high concentrations to achieve a binding curve by MST analysis and therefore a Kd could not be calculated. CP11A4Y10 demonstrated a slightly lower affinity than CP11A4Y3 (Kd 81.0 ± 5 µM), but again showed a change in the thermophoretic motion of UEV. Interestingly, CP11A4Y3C had a lower affinity than CP11A4Y3 (Kd 103.7 ± 2 µM), but again affected the thermophoretic motion of UEV, however, in a different manner to the serine peptides. Although it was interesting to note that the ΔF_{norm} [$^{0}/_{00}$] of CP11A4Y3C was negative rather than positive like the other peptides with a serine residue. This could be linked to the difference in effects seen in the thermal shift assay where CP11 demonstrated a destabilising effect on UEV whereas the cysteine derivatives showed a stabilising effect. A negative ΔF_{norm} [$^{0}/_{00}$] indicates a positive Soret coefficient as the fluorescent molecules have positive thermophoretic mobility and move away from the laser heated area. Conversely, a positive ΔF_{norm} [$^{0}/_{00}$] indicates a negative Soret coefficient as the fluorescent molecules have negative thermophoretic mobility and move towards the laser heated area. The Soret coefficient is dependent on molecule size, charge and molecular interactions with the solvent. As the peptides CP11A4Y3 and CP11A4Y3C are similar in size, this finding confirms that the cysteine derivative either has a different binding conformation than the serine peptides or affects the solvation shell or charge of the protein in a different way, hence affecting the thermophoretic motion in opposing manners. As the assay buffers did not contain a reducing agent, the cysteine derivative could be binding to the surface cysteines in UEV through a disulfide bridging interaction, which could explain the difference in Kd values between CP11A4 and CP11A4C. This could have been investigated further through inclusion of a reducing agent in the buffer conditions, but was not investigated due to time constraints.

The destabilising effect seen in the thermal shift and the negative Soret coefficient could indicate that CP11 and its derivatives are disrupting the UEV-p6 PPI by inducing a destabilising conformational change in UEV, rather than binding to the active site of UEV. Some ligands are known to be efficacious by causing conformational changes in their targets and therefore preventing the natural ligand from binding[159]. The relatively weak binding of the cyclic peptides to UEV could also be due in part to the fluorescent labelling of the protein, which could hinder the cyclic peptides from binding.

4.5 Summary

Three peptides were identified from the unnatural aromatic analogue library that had a 5-7 -fold improvement in inhibitory activity when compared to CP11. These peptides were all from the same family of cyclic peptides that were based upon the CP11A4 backbone, where the asparagine in the fourth position was substituted for alanine. The three peptides (CP11A4Y3, CP11A4Y4 and CP11A4Y10) contained the unnatural tyrosine analogues F(4-F), F(4-Cl) and Cha, respectively.

CP11 was found to have a destabilising effect on UEV by thermal shift. However, the change in melting temperature at 1 mM was only 2 °C, which is close to the level of error for this assay, indicating that this assay was not the ideal method for assessing the interaction between the peptide and proteins.

The cyclic peptides were then tested in an MST assay to determine the strength of binding of the peptide to UEV. CP11A4Y3 was found to have a Kd 64.56 ± 4.24 µM, which is a 2-fold improvement in Kd compared to CP11. The other two analogues, CP11A4Y4 and CP11A4Y10, either did not show significant improvement over CP11 or were not soluble enough to obtain a full binding curve. This moderate binding activity is weaker than expected as the IC_{50}s obtained from the ELISA were below 10 µM. It was hypothesised that the cyclic peptides could be inhibited from binding by the presence of the fluorescent label of UEV.

Despite the relatively weak binding determined by MST, the most active peptide, CP11A4Y3 still demonstrated a slightly tighter binding constant than CP11, which was hypothesised would correlate to an improved inhibition of HIV VLPs, than CP11 in a cell-based assay.

Chapter 5 Assessing the Efficacy of a CP11 Derivative in a Cell-Based Assay of HIV-Budding

5.1 Introduction

After the binding of the peptide was determined by MST as described in 4.4, it was decided to attempt a virus-like particle (VLP) assay with the most active peptide to see if the peptide was able to inhibit viral budding *in vivo.*

It is known that the viral Gag protein is sufficient to be able to produce VLPs that can bud from the plasma membrane of mammalian cells, even in the absence of other viral proteins[104]. This is due to the p6 late domain, which recruits the ECSRT machinery through its PTAP motif, which binds to the UEV domain of TSG101 that makes up the polyprotein, ESCRT-1[39-40, 110]. This makes it possible to study the effect of inhibitors on viral budding *in vivo*, without needing to use the intact virus.

5.2 Testing Peptides in a Virus-Like Particle Assay

The VLP assay chosen was based upon work originally used to assess CP11[103] *in vivo* and an assay developed by Pornillos *et al.*[38]. Human embryonic kidney (HEK293T) cells were used as they express a mutant SV40 large T-antigen enabling a consistently high level of recombinant protein expression and are consistently used in assays requiring the production of retroviral proteins. The cells were transfected with pMET7-GAG-eGFP using Lipofectamine 2000 (4:1 plasmid:Lipofectamine 2000, Life Technologies) according to the manufacturers protocol. After 5 hours, the Lipofectamine solution was removed, the cells gently washed and then dosed with increasing concentrations of peptide or with DMSO (0.4%) and incubated for a further 18 hours. The cells were washed prior to dosing to ensure removal of the transfection reagent and to remove any VLPs that might have formed during transfection. Transfection efficiency was assessed by fluorescence microscopy and was typically about 70% efficient and VLPs were visible in the supernatant of the cells, Figure 60.

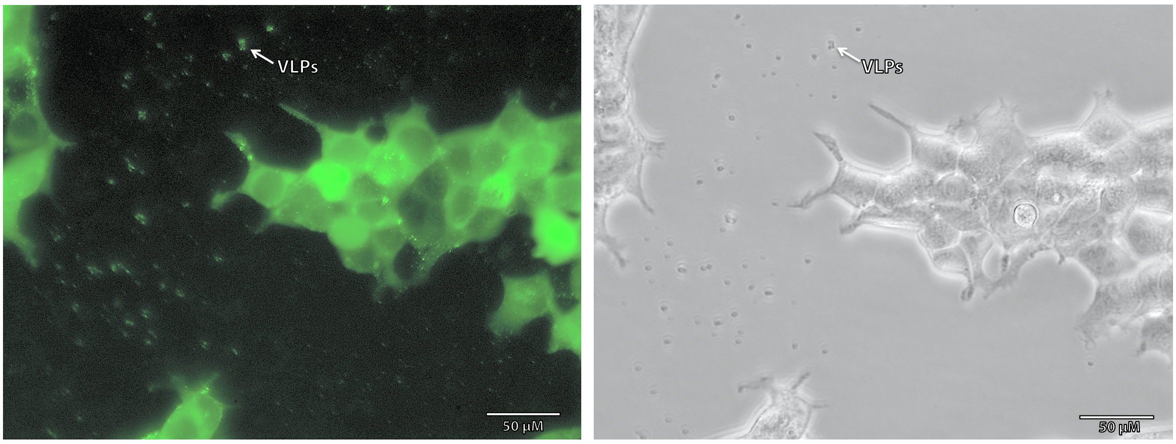


Figure 60: Fluorescence Microscopy of HEK293T Cells Transfected with pMET7-Gag-eGFP Showing VLPs in Solution.

The supernatants were centrifuged (4000 rpm, rt, 4 min) to remove any cell debris, followed by filtration through a 20% (w/v) sucrose cushion (13 000 rpm, 4 °C, 90 minutes) to purify the VLPs and then resuspended in Laemmli buffer to produce the supernatant fraction. Cells were lysed in RIPA buffer by sonication (12 cycles of 30 seconds on, 30 seconds off) and resuspended in Laemmli buffer to produce the cell lysate sample. Both supernatant and cell lysate fractions were analysed by immunoblotting, Figure 61A-B.

A dose-dependent decrease in Gag-eGFP intensity was observed in the cell supernatant for both repeats of this assay. Although there was an anomaly in the first repeat at the 2.5 µM dose, Figure 61A, which has a much more intense band that is to be expected. In the second repeat of the assay, Figure 61B, there is not a consistent amount of Gag-eGFP detected in each of the cell lysate samples. The increased amount of Gag-eGFP in the samples of the cells dosed with higher concentrations of peptide (5-10 µM) could be due to the peptide preventing the budding of VLPs and therefore leading to an accumulation of Gag-eGFP in the cell cytoplasm. An IC_{50} of 1.67 ± 0.27 µM was determined by comparison of the intensities of the bands detected in the supernatants, relative to the untreated control samples. The anomalous intense band for the sample dosed with 2.5 µM from the first repeat was discounted.

It was found that CP11A4Y3 was efficacious in preventing VLP release without the need to Tat-tag the cysteine derivative. This could be due to the inclusion of the (4-chloro)-phenylalanine, as other research has shown that inclusion of this unnatural amino acid enhanced passage of peptides across the blood-brain barrier[160] or the peptide could be interacting with the cell membrane in order to produce an intracellular response. In the work by Tavassoli *et al.*[103], CP11 was not tested in the VLP assay without Tat-tagging the cysteine derivative, so it is not possible to know if CP11 would be as efficacious without tagging with Tat peptide. Due to time constraints it was not possible to test CP11 in this assay to be able to compare the IC_{50} values.

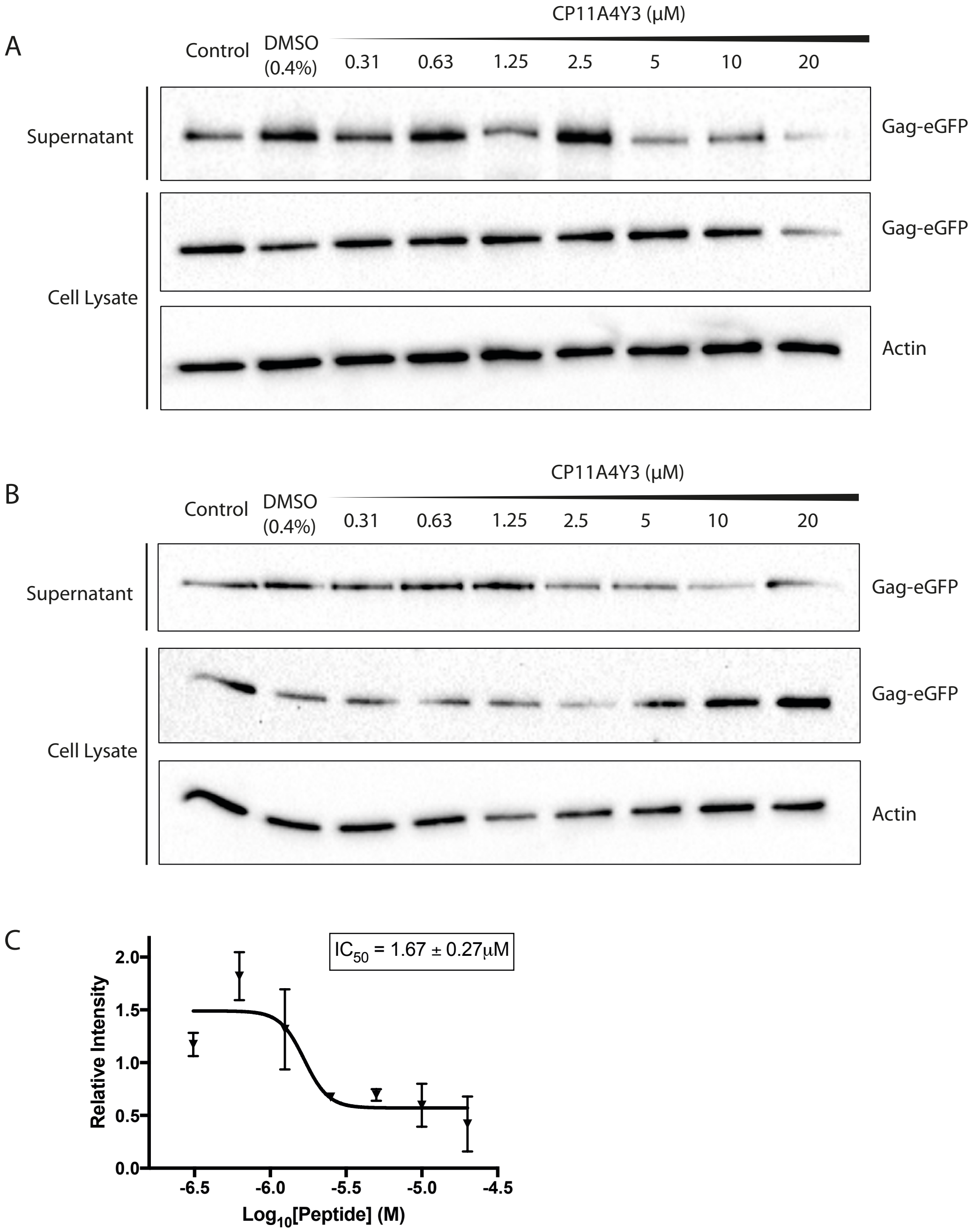


Figure 61: Western Blots of HEK293T Cells Transfected with Gag-eGFP and Treated with CP11A4Y3. A. First repeat of HEK293T cells dosed with CP11A4Y3 (0.31 – 20 μM). B. Second repeat of HEK293T cells dosed with CP11A4Y3 (0.31 – 20 μM). C. Graph of the intensities of the bands relative to the untreated control samples imaged in the supernatant samples after blotting first with anti-eGFP (ThermoFisher) and then anti-mouse conjugated-horse radish peroxidase (GE Healthcare) and detected using ECL Prime Western Blotting reagent (GE Healthcare). The relative intensities were measured using the ImageLab 4.0 Software (BioRad). Error bars show the standard error about the mean of 2 repeats.

Due to time constraints it was not possible to repeat the VLP assay with positive and negative controls other than the DMSO negative control completed at the time of the assay. The ideal negative control would be an analogue of CP11A4Y3 where the tyrosine residue had been exchanged for alanine (*cyclo*-IAWAVSGW), as this analogue would be structurally similar to

CP11A4Y3 (*cyclo*-IF(4-Cl)WAVSGW), but would not contain the tyrosine residue that was determined to be important as described in previous chapters.

As there are no confirmed inhibitors of this interaction in the literature, it is not possible to have a dosed positive control for this assay. A peptide based on the PTAP motif may be a possible positive control; work has been completed investigating variations on peptides derived from the PEPTAPPEE sequence in p6 as well as cyclised versions that have approximately 2-3-fold improved affinities to the linear wild type peptide and when tagged with the fluorescent tag, FITC, demonstrated good cellular uptake in HeLa cells[161]. However, these peptides have yet to be tested in a VLP assay and their efficacy in cells is unknown.

There is an interferon-induced protein, called tetherin, which has been found to prevent the release of enveloped viral particles such as HIV, but its mechanism of action is as yet unknown[162]. This could be a potential positive control, although optimisation of the VLP assay would need to be completed to ensure co-transfection would not induce a level of cellular stress that would affect the results of the assay. HIV virus production has been found to be inhibited by microRNA in radioimmunoprecipitation assays and this could be another option as a positive control in future experiments[163].

5.2.1 Testing Peptide Cytotoxicity

To ensure that the decrease in VLP budding was not due to cell death, the cellular toxicity of the peptides was assessed using the CytoTox 96® Non-Radioactive Cytotoxicity Assay Kit (Promega) to ensure that the inhibition of VLP formation by the peptide was due to specific inhibition of the p6-UEV interaction and not due to any cellular toxicity. This kit quantitatively measures release of lactate dehydrogenase, which is a stable cytosolic enzyme that is released upon cell lysis. The amount of lactate dehydrogenate released into the supernatant is proportional to the number of lysed cells in the sample. The assay utilises a coupled enzymatic assay, whereby lactate dehydrogenase converts a tetrazolium salt into a red formazan product that can be detected using a plate reader. A sample of the cell supernatant of each well of cells at each dose was taken in duplicate prior to harvesting for the VLP assay. The samples were tested according to the manufacturers protocol, using the cell lysate (1 µL diluted in 39 µL RIPA buffer) of the untreated cells as a positive control. The negative control is the supernatant of the untreated cells. The data were normalised to a sample of buffer solution that was incubated for the same length of time as the cells in the assay. CP11A4Y3 was found not to be cytotoxic to HEK293T cells at the concentrations used in the VLP assay (0-20 µM), compared to the untreated cells, Figure 62. There was also no cytotoxicity observed with the cells treated with DMSO (0.4%).

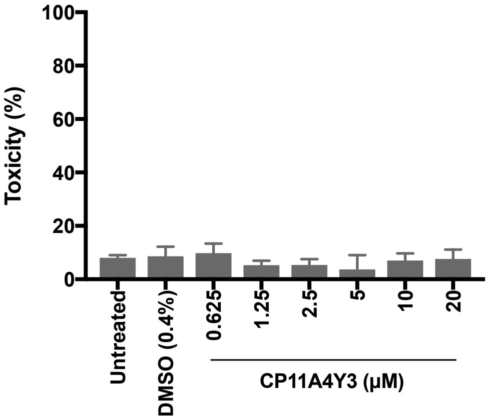


Figure 62: Assessing the Cytotoxicity of CP11A3Y4 on HEK293T Cells. The supernatants of HEK293T cells treated with CP11A3 in increasing concentrations were collected after incubation for 18 hours. The supernatants were assessed for toxicity using the CytoTox 96® Non-Radioactive Cytotoxicity Assay Kit (Promega) and compared to a positive control of cell lysate from untreated cells. The negative control is the supernatant of the untreated cells. Error bars show the standard error about the mean of a duplicate of duplicates.

5.3 Testing the Effect of the Most Active CP11 Derivative on EGFR Degradation

Under normal conditions, ESCRT-1 plays an important role in the MVB pathway, which is responsible for targeting integral membrane proteins, such as growth factor receptors, for degradation in the lysosome[30-31, 164]. One such transmembrane protein is EGFR, which upon activation by EGF is internalised and either recycled back to the plasma membrane or targeted for lysosomal degradation by the ESCRT machinery[165].

It was described in Section 4.4 that the peptides were binding to UEV, which could mean that the normal function of the host protein, such as the MVB pathway, could be affected. To assess the impact of CP11A4Y3 on the MVB pathway, an EGFR downregulation assay was performed. In this assay, HeLa cells were transfected with pcDNA3.1CAT-EGFR using Lipofectamine 2000 (1:4, Life Technologies), and dosed after 5 hours with peptide or control solutions for 20 hours. The cells were then starved in serum-free media for 1 hour and either mock-treated or treated with EGF (150 ng/mL) for 1.5 hours. Cells were harvested, lysed and analysed by Western Blotting, Figure 63.

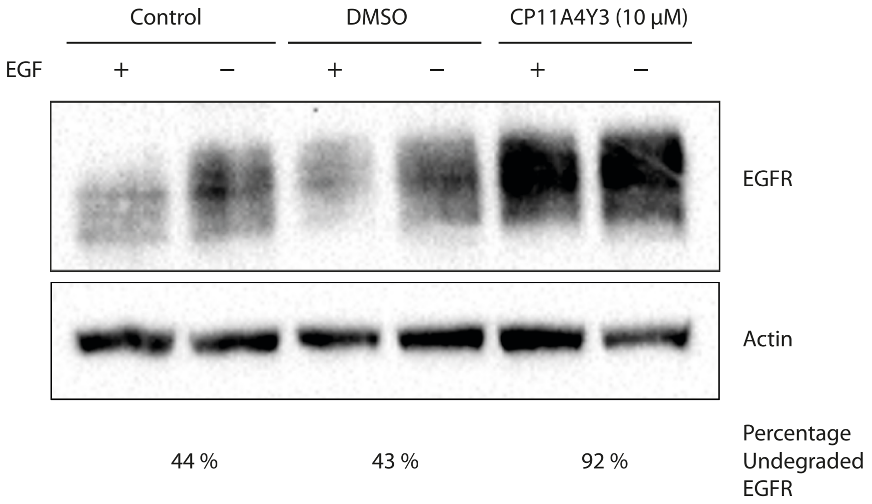


Figure 63: Western Blot Analysis of the Effect of CP11A4Y3 on EGFR Downregulation.

CP11A4Y4 (10μM) was found to prevent the downregulation of EGFR when dosed with EGF, compared to the control and the cells treated with DMSO (0.4%). When densities of the bands for EGFR were compared between the control lysate and the lysate of cells treated with CP11A4Y3, but not treated with EGF, a 4-fold increase in intensity was observed for the lysate of the cells treated with peptide. This could indicate that the peptide is preventing the normal regulation of EGFR by the ESCRT machinery in the absence of the target growth factor, causing an accumulation of EGFR at the cell membrane.

This finding is in contrast to the previous work that showed that CP11 did not affect the degradation of EGFR[103] under the same assay conditions, and this difference in effect could be due to the improved potency of CP11A4Y3.

5.4 Creation of a Stable Cell Line to Express Cyclic Peptide Inhibitors of HIV Budding

Previous work has shown the ability of SICLOPPS *Nostoc punctiforme* DnaE (*Npu*) inteins to successfully splice and generate a cyclic peptide in human HEK293 cells[166] and have been found to splice at a much faster rate than the original *Synechocystis sp* PCC6803 DnaE (*Ssp*) split inteins[167]. In this work, a stable cell line was generated that produced a known cyclic peptide inhibitor of HIF-1α/HIF-1β dimerisation[102] (a PPI that has been shown to be upregulated in certain cancers[168]) in response to hypoxic conditions, showing targeted production of an inhibitor without the need for chemical synthesis. Using the same methodology, it was hypothesised that a stable cell line could be created that would be able to produce a cyclic peptide that could prevent VLP budding when induced by treatment with compounds such as doxycycline.

Although cyclic peptides containing unnatural amino acids have been generated by SICLOPPS combined with an orthogonal aminoacyl-tRNA synthetase/ $tRNA_{CUA}$ pair in *E. coli*[99], this has yet to be replicated in human cells; therefore the peptide CP11A4 was used as a proof of concept as it is the original backbone of CP11A4Y3. A stable cell line was created by integrating the DNA sequence for CP11A4, flanked by the N- and C-inteins at the 5' and 3' ends respectively, into the genome of the T-REx293 cell line. The T-REx293 cell line has been produced from HEK293 cells that have been specifically designed to be suitable for using flippase-flippase recognition target (Flp-FRT) recombination[169]. This form of stable cell line generation uses a *Sacchoaromyces cerevisiae* derived DNA recombination system to integrate the gene of interest into a specific site on the genome[170]. The T-REx293 cell line constitutively expresses a *lacZ*-Zeocin fusion gene; however just downstream of the ATG start codon, a FRT site has been inserted, Figure 64A. The FRT site is the binding and cleavage site for the Flp recombinase[171-172], which is encoded on a pOG44 plasmid, and can be cotransfected into the cell line with a pcDNA5/FRT plasmid that contains the gene of interest, as well as a hygromycin resistance cassette to allow for the screening of successful integrants, Figure 64B. The Flp recombinase then catalyses the recombination between the two FRT sites in the cell genome and the pcDNA5/FRT vector, via a Holliday junction intermediate, resulting in the integration of the gene of interest into the cell genome, Figure 64C. In this case the gene of interest was the sequence for CP11A4 along with the flanking intein regions, tagged with a CBD. The plasmid contained the CMV promoter followed by two copies of the tetracycline operator (tetO), which can be bound by the tetracycline repressor protein encoded within the T-REx293 genome, and allow for the regulation of cyclic peptide production by treatment with doxycycline, resulting in derepression of genes regulated by the tetracycline repressor. The pcDNA5/FRT/CP11A4 plasmid was kindly generated by another member of the Tavassoli group, Andrew Foster.

Once the pOG44 and pcDNA5/FRT/CP11A4 plasmids had been transfected into the T-REx293 cells, the cells were treated with hygromycin (200 µg/mL) to select for successful integration of the pcDNA5/FRT/CP11A4 vector. After 3 weeks, the integration was deemed to be successful as colonies of cells were observed and no cell growth was seen on control plates that were integrated with pOG44 only. The new cell line was subcultured and maintained with hygromycin (200 µg/mL) and blasticidin (15 µg/mL) to ensure that the SICLOPPS genes were retained.

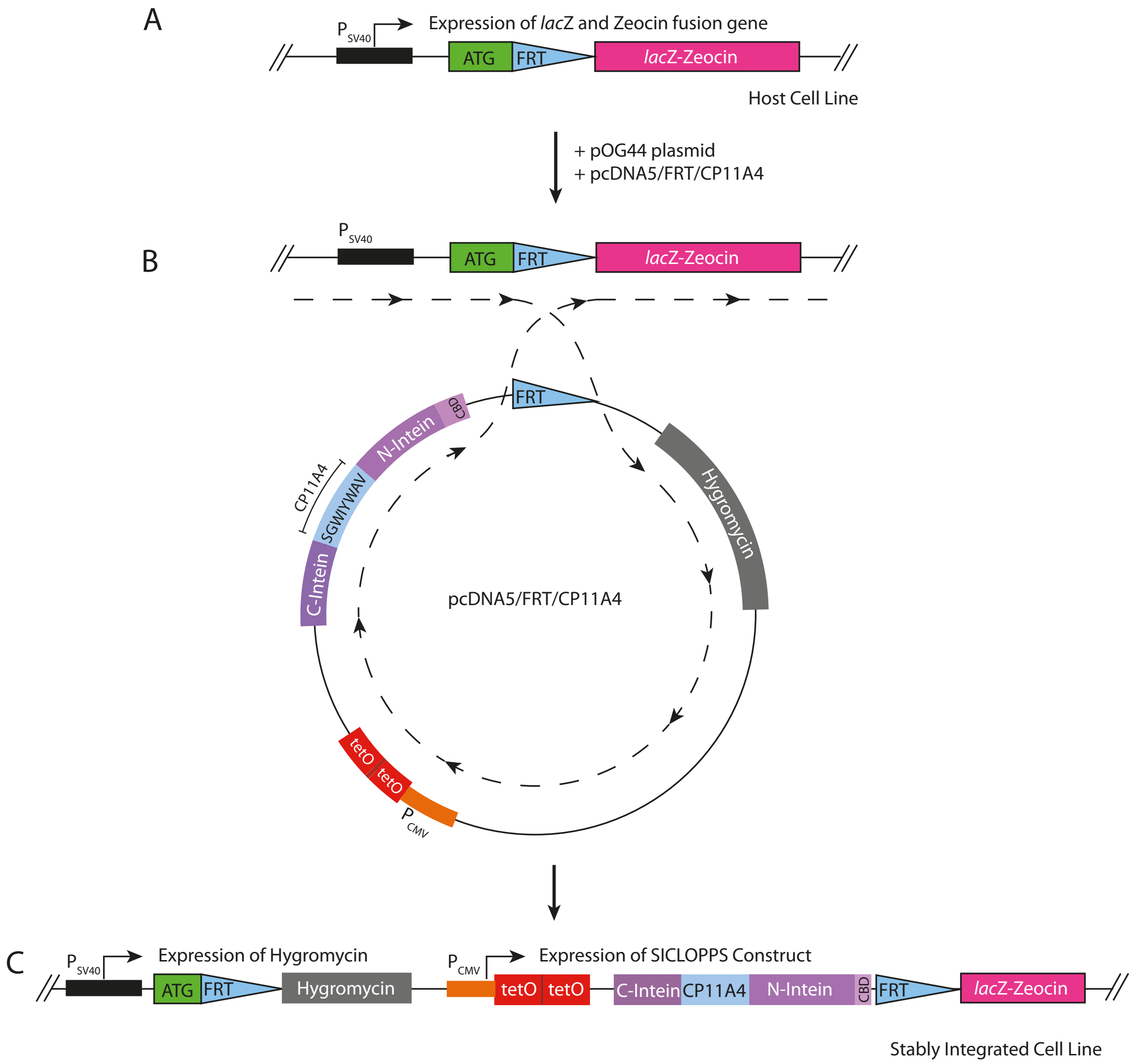


Figure 64: Schematic Showing the General Process for Generation of a Stably Integrated Cell Line Using Flippase/Flippase Recognition Target Recombination. A. The T-REx293 cell line contains a FRT site between the ATG start codon and the *lac*Z-Zeocin fusion gene. B. When the T-REx293 cells are cotransfected with pOG44 and pcDNA5/FRT/CP11A4. pOG44 encodes Flp recombinase which catalyses a recombination event between the FRT site in the host cells and the pcDNA5/FRT/CP11A4 plasmid to form C. a stably integrated cell line that inducibly expresses the cyclic peptide CP11A4 when treated with doxycycline.

To ensure successful integration into the T-REx293 cells, a time course assay was performed where cells were treated with doxycycline (1 µg/mL) to induce expression of the SICLOPPS construct and then harvested at regular intervals over a 24 hour time course. Cells were lysed in RIPA buffer by sonication and immediately immunoblotted for CBD, as this would show the production of the N-intein over time and hence an indication of the production of CP11A4, Figure 65A. Only one band at approximately 18 kDa was observed, and no higher bands were observed on the gel which indicates that there are no intact inteins and that the SICLOPPS construct is splicing to form CP11A4, as this is the size of the N-intein without the peptide or C-intein.

To test expression of the SICLOPPS construct, T-REx293 and T-REx293-CP11A4 cells were either mock-treated or treated with doxycycline (1 µg/mL) for 24 hours and then harvested and immediately lysed in RIPA buffer by sonication and immunoblotted for CBD, Figure 65B. As expected, the N-intein was not expressed in T-REx293 cells. The N-intein was selectively expressed in T-REx293-CP11A4 cells treated with doxycycline, but no expression was observed without doxycycline, indicating that expression of the SICLOPPS construct is doxycycline-dependent.

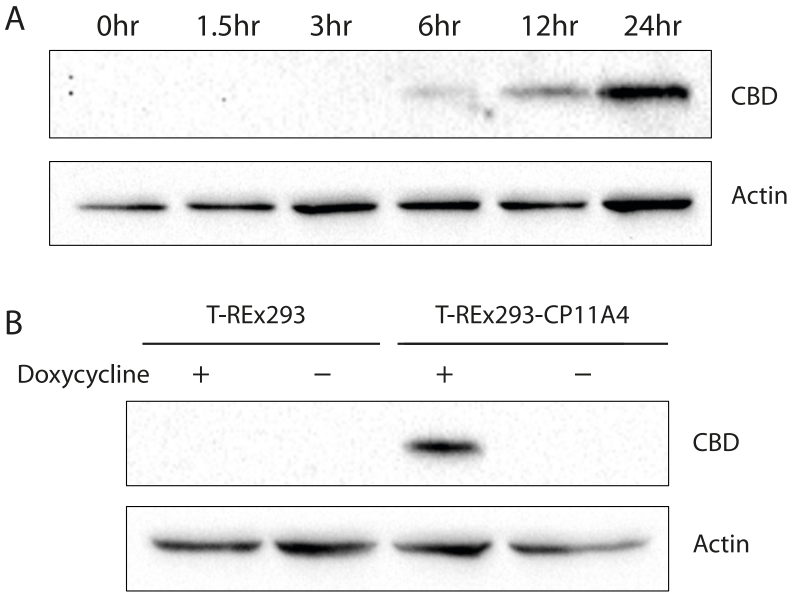


Figure 65: Western Blot Showing the Production of CBD-tagged N-Intein at 18 kDa. A. Time course assay with T-REx293-CP11A4 cells treated with doxycycline (1 µg/mL) over 24 hours. B. Western blot showing that the parent cell line (T-REx293) does not express N-intein and that the N-intein is not expressed in T-REx293-CP11A4 cells without being induced with doxycycline.

To ensure that CP11A4 was successfully cyclised and spliced from the inteins, T-REx293-CP11A4 cells were grown in high volume and induced with doxycycline (1 µg/mL) for 24 hours. Cells were then lysed in PMSF lysis buffer (5 mM EDTA, 2 mM, EGTA, 0.4 mM PMSF in PBS) containing a protease inhibitor cocktail by three freeze/thaw cycles in liquid nitrogen. PMSF Buffer was used as it contains a lower concentration of buffer salts and is therefore more compatible with HPLC technology. The lysate was then filtered through a 10 kDa cut-off filter and analysed by analytical HPLC, Figure 66A. A small peak was observed in the analytical HPLC trace at 19.6 minutes, which is the same retention time as the synthetic CP11A4. The fractions corresponding to the retention time of CP11A4 were collected and analysed by high resolution LC-MS (ESI+), Figure 66B. The mass of the peptide ($[M+H]^+$, $C_{50}H_{63}N_{10}O_{10}$, calculated mass 963.4704) was found (963.4723, 2.0 ppm error), as well as the mass of the cyclised peptide plus a sodium ion ($[M+Na]^+$, $C_{50}H_{62}N_{10}NaO_{10}$, calculated mass 985.4532, found mass 985.4543, 1.1 ppm error). This confirms that the SICLOPPS gene is being expressed in the integrated cell line and that the inteins are splicing to form CP11A4.

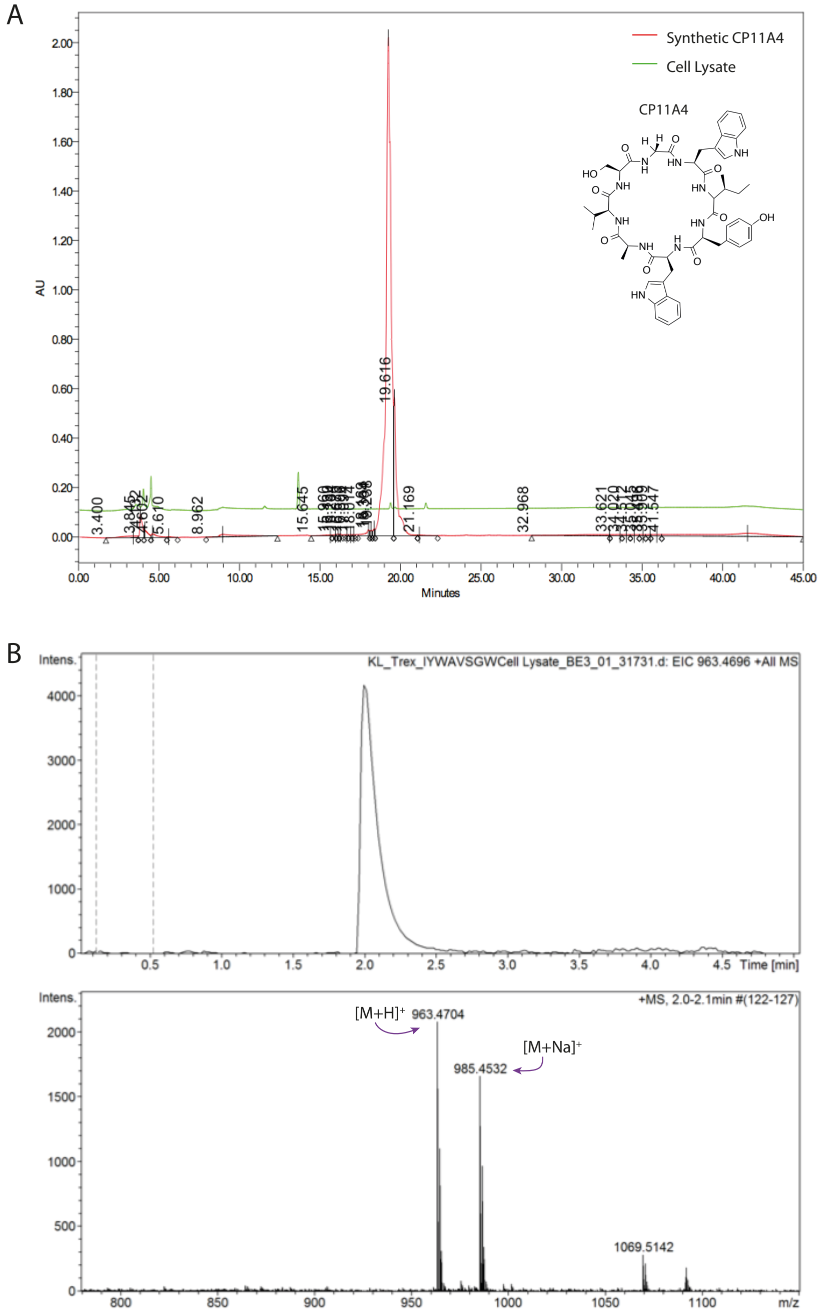


Figure 66: CP11A4 Isolated from T-REx293-CP11A4 Cell Lysate. A. Analytical HPLC trace comparing synthetic CP11A4 (*cyclo*-IF(4-Cl)WAVSGW) and cell lysate. A small peak can be seen in the cell lysate that has the same retention time as the synthetic peptide. B. High resolution MS (ESI+) analysis of the cell lysate peak at 19.6 minutes, showing the presence of CP11A4.

5.4.1 Performing a VLP Assay with T-REx293-CP11A4 Cells

The T-REx293-CP11A4 cells were then assessed using a VLP assay to see if budding of VLPs was inhibited by the expression of the cyclic peptide stably integrated into the cell genome. T-REx293 cells are derived from HEK293 cells and therefore it was assumed that the transfection procedure described in 5.2 for transfection of the pMET7-Gag-eGFP plasmid using Lipofectamine 2000

(LifeTechnologies) would not need further optimisation. The T-REx293-CP11A4 cells were grown to 60% confluency and either mock-treated or treated with doxycycline (1µg/mL) for 12 hours before the transfection with pMET7-Gag-eGFP using Lipofectamine 2000 (4:1, LifeTechnologies) according to the manufacturers instructions or mock transfected with water. Cells were either dosed with doxycycline (1 µg/mL) to ensure consistent expression of the SICLOPPS gene or mock-dosed as before. After 12 hours, the supernatants and cells were harvested. The supernatants were centrifuged (4000 rpm, 4 minutes, rt) to remove any cell debris, and then filtered through a 20% (w/v) sucrose cushion (13 000 rpm, 4 °C, 90 minutes) before resuspension in Laemmli buffer to produce the supernatant fraction. Cells were lysed in RIPA buffer by sonication (12 cycles of 30 seconds on, 30 seconds off) then resuspended in Laemmli buffer to produce the cell lysate fraction. Both cell lysate and supernatant fractions were analysed by immunoblotting, Figure 67.

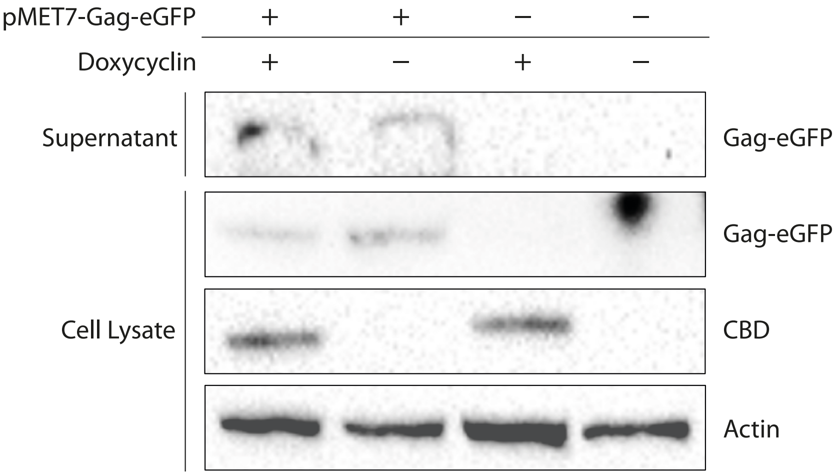


Figure 67: Western Blot Analysis of VLP Assay Performed with T-REx293-CP11A4 Cells. The membrane containing the cell lysate fractions was first exposed to anti-CBD monoclonal antibody, then anti-mouse HRP-conjugated antibody and detected using ECL reagents. The membrane was stripped using a mild glycine buffer and then incubated with anti-eGFP, followed by anti-mouse conjugated-HRP antibody and detected using ECL reagents. Actin was detected using a conjugated actin-HRP antibody.

The N-intein was expressed in cells treated with doxycycline (1 µg/mL) to similar levels and were not observed in the absence of doxycycline treatment. VLPs were seen in the supernatant fraction of cells transfected with pMET7-Gag-eGFP, although there was not a visible difference in intensity of the detected bands between the sample dosed with or without doxycycline to induce CP11A4 production. This could be due to the lower expression of the peptide in the cells at an inefficacious concentration so inhibiting the budding of VLPs. When cells are transiently transfected with a plasmid, it non-selectively expresses the gene of interest until the plasmid is eventually lost from the cells. Stably transfected cells will selectively express a gene only in the presence of the effector. This may explain why the cyclic peptide was not produced in a high enough concentration to prevent VLP formation. Further optimisation is required for this assay to optimise the doxycycline treatment of the cells prior to transfection with PMET7-Gag-eGFP.

Although the desired inhibition of viral budding was not observed in this case, this is an important proof of concept for the assessment of cyclic peptide libraries for inhibition of PPIs involved in disease.

5.5 Summary

The effect of CP11A4Y3 on HIV VLP production was tested in HEK293T cells transfected with the HIV Gag protein. It was found that CP11A4Y3 did not need to be Tat-tagged for cell permeability, as it was found to prevent VLP release upon dosing. CP11A4 was found to have an IC_{50} of 1.67 ± 0.27 µM in this assay, although further work is required to repeat the assay with additional positive and negative controls and alongside CP11. CP11A4 was not cytotoxic to HEK293T cells at the range of concentrations used in this assay.

CP11A3Y4 was found to prevent the degradation of EGFR in HeLa cells dosed with EGF compared to the control and cells treated with DMSO (0.4%), which was not seen in previous research with CP11[103]. This prevention of EGFR degradation could be due to the improvement in binding seen with CP11A4Y3, which means that the peptide is binding sufficiently tightly to UEV so as to affect the normal function of the ESCRT machinery.

As a proof of concept, a stable cell line was created that selectively expresses a SICLOPPS construct to generate a cyclic peptide in response to doxycycline. The methodology to express unnatural amino acids in cyclic peptides using SICLOPPS in human cells has not been fully realised, therefore the cyclic peptide CP11A4 was chosen as this is the backbone of the most active peptide as well as the most active peptide without unnatural amino acids. CP11A4 was selectively expressed upon induction with doxycycline and its presence confirmed by high resolution MS analysis. However, when this cell line was utilised in a HIV VLP assay, no effect was observed on the production of VLPs, indicating that further work is still required to optimise this assay.

Chapter 6 Conclusions

Ever year HIV affects millions of people worldwide, with 1.8 million newly infected people in 2016 alone[46]. In 2016, only 54% of adults and 43% of children with HIV were receiving cART although this number has increased since 2000 and the number of deaths due to the development of the disease has decreased by a third[46]. As there is currently no cure, individuals with HIV will have to take cART for the rest of their lives to suppress the development of the disease even though their viral load might become undetectable. This is due to the virus' ability to remain dormant in various tissues until ART is stopped and then rebound. Although current cART is highly effective, there are a growing number of cases of resistance and the levels of resistance to first-line treatments, such as efavirenz or nevirapine in cART-naïve patients has reached over 10% in six out of eleven countries conducting research into pre-treatment HIV drug resistance[173]. cART targets multiple stages of the HIV lifecycle, although one stage is not targeted by current therapies. Viral budding is an important stage in the HIV lifecycle as it involves the egress of the assembled nascent virions from the plasma membrane. This process is governed by the recruitment of the host's ESCRT machinery by a PPI between the viral late domain p6 and the TSG101 protein of ESCRT-I. Studies have shown that if this PPI is disrupted, the virions are unable to bud from the cell and remain tethered to the membrane[37, 40]. A cyclic peptide inhibitor of the TSG101-p6 PPI was identified through a combined SICLOPPS and RTHS screen and found to inhibit viral budding in a VLP assay, with an IC_{50} of 7 µM. The aims of this project were to investigate the binding of this inhibitor, first through an alanine scan to determine the active motif, then to improve the inhibitor and determine the binding partner of the PPI.

The top two hits initially identified in the SICLOPPS/RTHS screen both contained the same motif, YWN in CP11 and reversed as NWY in CP6, so this was thought to be a good starting point for the project. The capped tri-peptide, acetyl-NWY-diethylamine was synthesised by liquid phase peptide synthesis and then analysed by NMR to generate constraints to drive a computational model to determine the conformation. The tri-peptide was found to be very flexible and formed five main conformations in solution. When these conformations were overlaid with a structural model of CP11, it was found that only 5% of the overall shapes of acetyl-NWY-diethylamine overlaid sufficiently with the predominant CP11 backbone. This confirms the finding that the tri-peptides, acetyl-NWY-diethylamine and acetyl-YWN-diethylamine, were significantly less active than CP11 at inhibiting the p6-UEV interaction in an ELISA (103.8 ± 13 µM and 164.8 ± 34 µM versus 48.6 ± 2 µM, respectively) as they occupy the potential bioactive conformation very rarely due to their inherent flexibility in solution.

To determine the active motif of CP11, an alanine scan was completed by two different methods. The first method involved the use of the SICLOPPS/RTHS that was utilised to originally identify CP11. SICLOPPS plasmids were generated for each alanine analogue and then transformed into the p6-UEV RTHS and dropspotted on minimal media. The active motif of CP11 could not be determined by this technique as high levels of toxicity were observed, even when the ʟ-(+)-arabinose concentrations were decreased to 1 µM. When the concentrations of ʟ-(+)-arabinose were decreased below 1 µM, toxicity was not observed, however the cells were not seen to regrow in the presence of IPTG indicating that the low concentration of ʟ-(+)-arabinose may not be sufficient to produce high enough concentrations of peptide to inhibit the PPI. The *Ssp* split-inteins utilised in this system are known to be intolerant of certain amino acids at the junction between the extein and inteins[167] as well as being relatively slow at splicing. Recently, *Nostoc punctiforme* (*Npu*) DnaE inteins have been utilised in SICLOPPS as they are more tolerant of amino acid variations as well as having a faster rate of splicing[167]. Although, there were initially high levels of toxicity observed, when an SsrA tag was incorporated in the inteins to target them for degradation, the levels of toxicity were decreased. This modified system could be used in future investigations when there is this level of toxicity observed and could provide a better system for screening peptides that is less limited by intein-related toxicities.

The second method for determining the active motif was more successful. Each alanine analogue was synthesised chemically and tested for changes in inhibitory activity in the p6-UEV ELISA. The most important residue was determined to be the tyrosine residue. A capped tri-peptide containing an amino acid flanking the tyrosine residue (IYW), was synthesised and assessed using the p6-UEV ELISA to see if activity was retained. This approach was used when investigating the ATIC transformylase interaction, and successfully showed that a capped di-peptide version of the active motif was more potent than the original cyclic hexa-peptide (K_i 685 nM versus 17 µM)[174]. Unfortunately, the capped IYW tri-peptide was 2-fold less active than CP11, indicating that the small peptide truncations are possibly less active than the cyclic peptides as they lack the structural hindrance to remain in the bioactive conformation and therefore infer a strong inhibition of the PPI.

The alanine scan also revealed that some of the alanine analogues were found to be slightly more active than CP11, which may be due to the removal of the steric bulk from the substituted residues (for example asparagine, tryptophan and valine), by substitution to alanine. This would allow the sidechains of the other residues more conformational freedom without the possibility of steric clashing and therefore allow the peptide to occupy the bioactive conformation for a higher proportion of time. It is interesting that the original sequence isolated from the SICLOPPS/RTHS screen did not contain these alanine substitutions, although this could be due to the aforementioned observed toxicity.

The inhibitory activity of the peptide was improved by investigating substitution of the tyrosine residue with unnatural analogues. This approach has been utilised in our research group previously to develop a more potent inhibitor for AICAR transformylase homodimerisation[174]. Unnatural analogues of tyrosine were used that had substituents of differing electron-donating or withdrawing properties to allow for the investigation of the electronic as well as steric requirements of binding. Following on from the alanine scan results, three of the alanine analogues were determined to be better inhibitors of the PPI than CP11 via ELISA. It was therefore decided to further investigate the substitution of the tyrosine residues of these alanine analogues by the same method. These tyrosine analogues were synthesised by SPPS and then tested using the p6-UEV ELISA and three were found to have significantly improved IC_{50} values when compared to CP11. There was no overall correlation between peptides with similar types of substituent groups on their tyrosine analogues, which prevented the further understanding of the potential type of non-covalent binding that was occurring between the peptide and the protein. There were no clear comparisons to be drawn between peptides containing tyrosine analogues of similar sizes as bulky analogues such as (1-naphthyl)-L-alanine (Nal) showed a similar potency to much smaller tyrosine analogues such as 4-fluorophenylalanine (F(4-F)). The most important aspect for binding appeared to be the overall backbone configuration conferred by the alanine analogues as all the most potent analogues were found in the CP11A4 backbone family of peptides, where alanine had replaced the asparagine residue at the fourth position.

Even though the most active peptide (CP11A4Y3) did not exhibit a strong binding affinity in an MST assay, it was found to be a potent inhibitor of VLP production in HEK293T cells expressing the HIV Gag protein, although it was found to also prevent the normal function of TSG101 by disrupting the recycling of membrane receptors. It has been shown that changes in the abundance of transmembrane proteins can have an effect on the tumourigenicity of cells, where alterations in transmembrane proteins can affect the epithelial cell polarity, leading to the cells converting to a fibroblastoid state with increased motility and the potential to facilitate metastasis[175]. This could present a problem when treating with CP11A4Y3, however, a novel way of flushing out viral reservoirs involves the treatment of the patient with immune activating agents to induce the expression of viral proteins from latently infected cells while maintaining cART to prevent infection of healthy tissues. This could be a potential mode of action for this peptide as a short term rather than a life-long treatment, to avoid the potential development of cancers.

The number of HIV drugs approved by the FDA and EMA has decreased in recent years, the majority of which are generic or different combinations of previously approved treatments. There are few novel drugs reaching approval and few are for novel targets. This is worrying as the number of cases of viral resistance is rising, which highlights a need for new drugs and novel targets. The p6-TSG101 PPI is untargeted by current ARTs and has been shown to be vital to the HIV lifecycle[37-38]. Due to the high mutation rate in viral genetic material, resistance occurs very quickly. Cases have been found where treatment-naïve patients have a strain of HIV that contains a double PTAP motif in the p6 gene to ensure for successful budding[176]. Duplications in this area of the p6 have also been linked with cART failure due to treatment pressures and can lead to protease inhibitor drug resistance[177]. By binding to UEV, the host protein, rather than p6, the risk of resistance developing is reduced.

HIV is not the only virus to hijack the ESCRT machinery to enable successful viral budding; the Ebola virus contains a conserved overlapping PTAP and PPEY motif in its matrix VP40 protein that binds to TSG101 and Nedd4, respectively to facilitate virus budding[178-179]. The Ebola virus is a member of the Filoviridae family of viruses and causes a serious and often fatal haemorrhagic illness upon infection. This crossover between viral species in mode of action demonstrates the possibility of successful inhibitors having use across different viral infections.

An attempt was made to investigate if a potent small molecule-like inhibitor could be developed from the cyclic peptide, CP11. The approach was similar to that used previously within the group[174], where the peptide motif was capped with an N-terminal acetyl group and a C-terminal diethylamine group. As determined by NMR conformational analysis, the capped peptides, although they demonstrated slight inhibitory activity, had a high degree of conformational freedom. To decrease this flexibility, the tri-peptide with the sequence NWY was synthesised with the unnatural amino acids Ahx, Ava and Gaba and then cyclised to afford different ring sizes and therefore different levels of steric hindrance. These hindered peptides demonstrated no inhibitory activity against the p6-UEV PPI by ELISA, further confirming that the potency of the inhibitors is strongly conformational based, and changes away from the preferred conformation can be detrimental to the inhibitory action. The tri-peptide, acetyl-IYW-diethylamine, determined to be the active motif by ELISA, was found to be 2-fold less active than CP11, although it demonstrated significantly more activity than either of the two di-peptides (acetyl-IY-diethylamine or acetyl-YW-diethylamine), indicating that there is a minimum requirement for the size of the motif. This provides important information that can be used in future research into developing small molecule inhibitors of this interaction.

In conclusion, as a result of the combination of determining the active motif through alanine scanning and the incorporation of unnatural amino acids, a more potent inhibitor of the UEV-p6 interaction required for HIV budding has been developed. This cyclic peptide has demonstrated iinhibition of VLP release and doesn't require tagging with a cell penetrating peptide such as Tat to ensure efficacy. This research has been completed to provide a better understanding of the mode of binding of CP11 and other cyclic peptide derivatives, however, more biophysical analysis needs to be completed to fully understand the mechanism of action. Although the cyclic peptide has the undesirable side effect of preventing EGFR downregulation, it could still have the potential to be developed further into small molecules for short-term treatments for flushing out HIV reservoirs or to be co-dosed with a recovery drug.

Chapter 7 Future Directions

Although this work presents research into determining the mode of action of the developed inhibitors of the p6-UEV interaction, there is still work to be done to fully understand the intricacies of targeting this interaction.

The most potent peptides determined by ELISA were found to bind to UEV rather than p6, which was not wholly unexpected when the properties of the two proteins are compared; p6 is a small, flexible protein lacking in secondary structure apart from two alpha-helices which are not involved in the binding to UEV[157]. The peptides were found to bind relatively weakly to UEV by MST, which could be due to the presence of the fluorescent label, potentially hindering the peptide binding to the protein in its favoured site. As UEV is a small protein, the surface area is relatively small, so a bulky fluorophore could potentially disrupt binding to a significant proportion of the protein surface. Future work could involve the use of label-free MST assays using the intrinsic fluorescence of the protein as long as the protein contains tryptophan residues[180], however, this could prove problematic in this case, as UEV only contains two tryptophan residues and CP11 and the most potent inhibitors also contain two tryptophan residues. Therefore further investigations of such methods is required.

As the cyclic peptides have been shown to exhibit inhibitory activity in ELISAs when the His-tagged UEV is bound to a nickel-coated plate, it can be assumed that binding to the His-tag does not prevent the peptides from binding. This could be exploited by using fluorescently conjugated antibodies that recognise the His-tag as an equivalent to fluorescently labelling the protein using a maleimide dye. Equally it would be possible to use a technique such as Surface Plasmon Resonance (SPR). SPR is a technique that measures the change in intensity of reflected light when a potential ligand is flowed across a protein that has been immobilised upon a gold-coated surface. His-tagged UEV could be bound to the surface and then the peptide could be flowed across to allow for the calculation of affinity constants. Another potential technique is isothermal titration calorimetry (ITC), which can be used to determine the binding affinity as well as the thermodynamics parameters by measuring the changes in heat that occur during a binding event. ITC does not require labelling and the reaction is measured in solution, making it ideal for measuring binding affinities of protein-ligand or protein-protein interactions.

The peptide was seen to destabilise UEV in a thermal shift assay and in the MST assay demonstrated a negative Soret coefficient, which could indicate that the peptide is an allosteric inhibitor. To confirm this hypothesis, the interaction would need to be investigated by either X-ray crystallography or NMR analysis. The structure of UEV has previously been confirmed by NMR spectroscopy in complex with the p6 PTAP motif[39] and the structure has also been determined by X-ray crystallography[181]. If a complex could be isolated by either NMR or X-ray crystallography and the site of cyclic peptide binding confirmed, this could then be used as a tool for developing more potent inhibitors as well as providing important information about the mode of binding.

One of the main disadvantages with peptides is their poor bioavailability combined with low metabolic stability. This can lead to poor pharmacokinetic properties as they are often metabolised before they can reach the site of action. Cyclisation of peptides can improve their resistance to proteolysis, however, cyclic peptides are inherently insoluble in aqueous solutions and require DMSO or another organic solvent to aid dissolution. This can be problematic in certain assays as DMSO can be denaturing to proteins in high concentration. The CP11 peptide consists of half non-polar and half polar amino acids, however this balance is disrupted when the polar tyrosine residue is substituted for unnatural analogues, which have varying solubilities in solution. The solubility of the peptide could be improved by the substitution of more polar residues, such as arginine, for the residues which were found to not be required for inhibition of the PPI.

Cyclic peptides can be a useful starting point to developing small molecules that have improved pharmacokinetic properties and efficacy. One of the disadvantages of peptides is the high level of flexibility along the peptide backbone, which can result in decreased activity. Cyclisation does impose a level of steric hindrance, however, this is not always compatible with developing small molecules. Other methods of generating steric hindrance in peptides can be included either by modifying the amino acid side chains or the peptidic backbone.

Modified amino acids include the use of β-substituted amino acids to decrease the torsional freedom within the side-chain, or analogues of amino acids with modified substituents[83]. In previous research, the activity of a smaller peptide motif was improved by the addition of unnatural aromatic analogues[174]; this methodology could be combined with the results already obtained whereby substitution of tyrosine with unnatural aromatic analogues improved the activity of CP11A4, to determine if the activity of this tri-peptide motif could be improved in a similar way. Investigations into β-substituted tyrosine would also be worthwhile to see if this could improve the activity of the tri-peptide motif.

Modifications to the peptide backbone including methylation of the backbone amines, isosteric replacement of the amino functionality, replacement of the α-carbon with nitrogen, use of D-amino acids or synthetic scaffolds can also impart desirable characteristics[84]. N-methylation can not only increase the steric hindrance but also have the additional advantage of improving the cell-permeability of peptides and improve peptide stability[139]. Further investigation into these modifications could be an interesting starting point in transitioning the identified tri-peptide motifs into peptidomimetics or potentially further into small molecules through the use of synthetic backbone scaffold that mimic peptidic characteristics such as α-helices or β-turns.

Although it is not as vital for viral budding as the UEV-p6 interaction, the ALIX-p6 PPI still plays an important role in the process[35]. This could be a potential target for future investigations into budding inhibitors as p6 binds to ALIX through an α-helical loop along one face of ALIX rather than a linear PTAP sequence in the TSG101-p6 interaction. Although there is potential for high levels of genetic variation within the YPX_nL motif of p6, mimetics of ALIX could be designed to mimic the ALIX binding site and therefore sequester the viral p6 protein. Studies have shown that overexpression of an ALIX fragment ($ALIX_{364-716}$) inhibited virus release 10-fold[182], demonstrating that targeting this interaction could be important for future research.

This work presents further understanding of how cyclic peptide inhibitors affect the p6-UEV interaction and demonstrates a methodology for improving hits discovered using the SICLOPPS/RTHS. It also provides an important starting point for the development of novel small molecule inhibitors of an untargeted interaction in HIV budding, which could be vital to investigating future treatments of this disease.

Chapter 8 Experimental

8.1 Materials and Equipment

8.1.1 Molecular Biology

Molecular biology reagents were purchased from ThermoScientific or Promega (UK). Restriction enzymes and polymerase were purchased from New England Biolabs (UK). DNA purifications were carried out using GeneJet Plasmid Miniprep Kit or GeneJet PCR Purification Kit, both purchased from ThermoScientific. PCR reactions were completed using either an Eppendorf Mastercycler or BioRad MyCycler Thermocycler Machine. DNA concentrations were measured using a NanoDrop ND-1000 Spectrophotometer. Agarose and SDS gels electrophoresis, and Western Blotting transfers were completed using a BioRad Power Pack Basic. Gels and immunoblots were imaged using a BioRad Universal Hood II Imager and analysed using Quantity One Software (BioRad, UK). Sequencing reactions were completed by Eurofins MWG (DE). Primers and gene fragments were purchased from Integrated DNA Technologies.

8.1.2 Peptide Synthesis

Reagents for peptide synthesis were purchased from either: SigmaAldrich (UK), ThermoScientific (UK) or Matrix Innovation Inc. (Quebec, CAN).

8.1.3 HPLC Purification

The linear and cyclic forms of the peptide were purified via manual injection into a Waters Flex inject HPLC system connected to a Waters 1525 binary pump and a Waters 2998 photodiode array detector set to measure at 220 nm and 280 nm. The system utilised a Waters Atlantis prep T3, amide capped C18, 5 μm diameter, 19 × 100 mm column. The HPLC was set to use a binary solvent system of A (water and 0.1 % TFA) and B (acetonitrile and 0.1 % TFA) at a flow rate of 17 mL/ min. The fractions were collected by a Waters Fraction Collector III.

8.1.4 NMR Analysis

NMR spectra were recorded on Bruker AVII400 or Bruker AVIIIHD400 FT-NMR spectrometers in the indicated solvent at 298 K. Chemical shifts for proton and carbon spectra are reported on the delta scale in ppm and were referenced to residual solvent references or internal TMS reference.

8.1.5 Biophysical Assays

Nickel coated 96-well plates were purchased from ThermoScientific. The thermal shift assay was performed on a BioRad CFX Connect Real-Time System. SYPRO Orange was purchased from ThermoScientific. The MST samples were analysed on a Monolith NT.115 MST Instrument (NanoTemper Technologies). MST capillaries and NT- 647-Maleimide dye were purchased from NanoTemper Technologies.

8.1.6 Mammalian Cell Culture

All cell culture reagents were purchased from Life Technologies unless otherwise stated. Mammalian cells were incubated in a HeraCell 150I Incubator (FisherScientific, UK), at 5% CO_2, 37 °C. Cells were visualised using a Trinocular Inverted Microscope (FisherScientific, UK).

8.1.7 Curve Fitting using GraphPad Prism7 Software

Data was analysed using the GraphPad Prism7 Software. ELISA data was analysed by first transforming the peptide concentrations to gain logarithmic concentrations. The curve was then fitted using a Nonlinear Sigmoidal, 4PL curve fit where X is log(concentration of peptide). This curve fitting software also generated IC_{50} values. Thermal Shift data was analysed by first transforming the data to plot the first derivate of the data, then using the software to calculate the area under the curve and therefore the peak height, which correlated to the T_m value.

8.2 General Methods

8.2.1 Cyclic Peptide Synthesis

Every effort was made to synthesise peptides that were as pure as possible, however in some cases epimers were generated at some point in the synthesis. This most likely occurred during the cyclisation step but could have occurred at any coupling step even in the presence of HOBt. During preparative HPLC, only one peak was collected in each case, however the epimers were found to co-elute when using the preparative column and were only slightly separable by analytical HPLC, however it was not feasible to purify the peptides (approximately 20 mg per peptide) using analytical HPLC as the maximal injection amount due to peptide solubility issues is <1 mg of peptide (greater amounts could cause problems if the peptide precipitated on the column or in the HPLC machinery), which would be too laborious and a waste of resources. In most cases the epimers visibly co-eluted even by analytical HPLC to such a degree that further purification would not be feasible. Therefore the peptides were synthesised in differing purities.

8.2.1.1 Fmoc Solid Phase Peptide Synthesis (SPPS)

The primary amino acid coupled to the Wang resin (1 mmol) was swollen in the presence of DMF for approximately 20 minutes, and then deprotected by exposure to 20 % piperidine in DMF (v/v) (10 mL) for 20 minutes. All coupling and deprotections were confirmed via a Kaiser Test (8.2.1.2). The resin was washed with DMF (3 × 10 mL), followed by alternating washes with DCM and diethyl ether (3 × 10 mL DCM, followed by 10 mL diethyl ether). Coupling was achieved via exposure to Fmoc protected amino acid (weight dependent on amino acid, 3 mmol), HOBt (0.406 g, 3 mmol), DIC (464 µL, 3 mmol) in DMF (6 mL) for 2 hours. The resin was washed as before. The linear peptide was synthesised by systematic coupling and decoupling steps until the full-length linear peptide was completed.

8.2.1.2 Kaiser Test

A few resin beads were added to 8 drops of 0.01 mM potassium cyanide in pyridine and 2 drops of 0.3 M ninhydrin in ethanol and before heating to 130 °C for 4 minutes. Deprotected primary amines result in a deep blue colour, while Fmoc protected amines do not react and result in a colourless solution.

8.2.1.3 Peptide Cleavage from the Wang Resin

The linear peptide was deprotected by exposure to 20 % piperidine in DMF (10 mL) for 20 minutes, and deprotection confirmed by a Kaiser Test (8.2.1.2). The peptide was washed as described in 8.2.1.1, and then cleaved from the Wang resin by exposure to a cleavage solution of: TFA (9.5 mL), TIS (0.25 mL), and water (0.25 mL), for 2.5 hours under argon. The peptide was filtered and then dried to minimum volume by rotary evaporation. The peptide was washed three times with DCM (10 mL) and then precipitated in cold diethyl ether (20 mL). The supernatant was removed and then the peptide was washed with additional diethyl ether (20 mL). The peptide was dried under argon and then stored at -20°C.

8.2.1.4 Peptide Cyclisation

The pure linear peptide was dissolved in DMF (approximately 1mg/ mL) with EDC (6 equivalents) and HOBt (6 equivalents) and stirred under argon for 24 hours. The solvent was then removed *in vacuo* and purified by HPLC (8.1.3).

8.3 General Molecular Biology Methods

8.3.1 Luria Broth (LB) Medium

To 250 mL deionised water, 6.25 g of LB powder was added. This solution was autoclaved at 121°C for 20 minutes on a media ballast cycle and allowed to cool prior to use.

8.3.2 Making Overnight Cultures

The *E. coli* strain recovered from frozen stock or picked from an agar plate was added to 10 mL of fresh LB medium containing the appropriate antibiotic (Table 10), and incubated at 37 °C overnight with shaking.

Table 10: Working Concentrations of Antibiotics. Antibiotics were made up using autoclaved, deionised water (or sterile ethanol for chloramphenicol) and filtered through a 0.22 μM sterile filter (Milipore, UK) and stored at 4 °C.

Antibiotic	Stock Concentration (mg/ mL)	Working Concentration (μg/ mL)
Carbenicillin	20	100
Kanamycin	10	50
Chloramphenicol	10	35
Spectinomycin	35	10

8.3.3 Plasmid Miniprep

An overnight culture containing the required strain, was centrifuged (1332 g, 15 minutes, 4 °C), the supernatant discarded and the resulting pellet purified using the GeneJET Miniprep Purification Kit (ThermoFisher) in accordance to the manufacturers guidelines.

8.3.4 LB Agar Plates

To 250 mL deionised water, 8.75 g of LB agar powder was added. This solution was autoclaved prior to use. To make agar plates, the autoclaved solution was allowed to cool to 50 °C prior to the addition of the relevant antibiotics. The liquid agar was poured into sterile plates (at a volume of approximately 25 mL per plate) using aseptic techniques. The plates were allowed to set and then dried at 37 °C.

8.3.5 Agarose Gels

To make a 1% agarose gel for gel electrophoresis, 1 g of agarose was added in 100 mL Tris-acetate-EDTA (TAE) (Table 11) and dissolved in the microwave. After leaving to cool, 1-2 drops of ethidium bromide was added, then the solution was poured into moulds and allowed to set. Once set the gel was placed in a running tank and TAE buffer added to sufficiently cover the gel. Samples of 5 μL were mixed with 1 μL 5× Green GoTaq® Buffer (Promega) and loaded onto the agarose gel. A 2-log DNA ladder (0.1-10 kb) (NEB) was loaded as a molecular weight marker. The gel electrophoresis was run for 30 minutes under an electric field of 100 V and then imaged under UV using GelDoc software.

Table 11: Composition of TAE buffer. Made up to 1 L with deionised water.

Reagent	Concentration
Tris base, molecular biology grade	2 M
Glacial acetic acid	1 M
Ethylenediaminetetraacetic acid (EDTA)	50 mM

8.3.6 PCR for Amplification of UEV and p6

To amplify the required gene, the reagents in Table 12 were mixed together and loaded onto the thermal cycler which was set to the run as described in Table 13. The primers used can be seen in Table 14.

Table 12: Composition of Solution for Gene Amplification by PCR

Reagent	Quantity (μL)	Final Concentration
GoTaq® Polymerase buffer	10	1×
Deoxyribonucleotide triphosphates (dNTP) (10 mM)	2	200 μM
Forward primer (10 pmol/μL)	2	0.2 μM
Reverse primer (10 pmol/μL)	2	0.2 μM
Template	1	Variable
DNA polymerase	0.25	1 unit
Deionised water	32.75	To make up to 50 μL

Table 13: Thermo Cycler Set Up

Cycle Number	Denaturation	Annealing	Extension
1	95 °C for 2 mins	-	-
2-30	95 °C for 2 mins	55 °C for 30 secs	72 °C for 30 secs
31	-	-	72 °C for 10 mins

Table 14: Primers and Sequences

Primer Name	Primer Sequence
Forward p6 with *Bam*HI	GCCGCCGGATCCATGCTTCAGAGCAGACCAGAGC
Reverse p6 with *Eco*RI	GCCGCCGAATTCTTATTGTGACGAGGGGTCGTTGC
Forward UEV with *Nhe*I	GCCGCCGCTAGCATGGCGGTGTCGGAGAGCCAGC
Reverse UEV with *Xho*I	GCCGCCCTCGAGTTAAGGACGAGAGAAGACTGG
Forward T7	TAATACGACTCACTATAGGG
Forward pGEX-2TK	GGGCTGGCAAGCCACGTTTGGTG

8.3.7 PCR Product Purification

The product obtained from the PCR amplification was purified using the GeneJET PCR Purification Kit (Thermo Scientific) according to the manufacturers guidelines. The concentration of the purified PCR product was obtained via the NanoDrop.

8.3.8 Restriction Digestion

Digestion of vectors and PCR products were carried out by incubating the solution in Table 15 for 4 hours at 37 °C. After 4 hours, the enzymes were heat inactivated by heating at 65 °C for 20 minutes. Thermosensitive alkaline phosphatase (TSAP) (1 μL) was added and the solutions incubated at 37 °C for a further 1 hour and then the TSAP was heat inactivated by heating at 70 °C for 20 minutes. The restriction digestions were then analysed on a 1% agarose gel for completion of digestion. The digested vectors and PCR products were purified using the GeneJET PCR Purification Kit (Thermo Scientific) according to the manufacturers guidelines.

Table 15: Composition of Solutions for Restriction Digestion

Reagent	Quantity (μL)
Vector/ PCR product	20
CutSmart buffer	5
Water	23
Enzyme 1	1
Enzyme 2	1

8.3.9 Ligation Procedure for Cloning Genes into Vectors

The digested vectors and inserts were ligated at 4 °C for 18 hours. The ligation was set up as in Table 16 with molar ratios of 1:3 and 1:6. The amount of vector and insert was calculated using Equation 1. A ligation was set up with vector only as a control to see the amount of re-ligation occurring. After 18 hours, the ligase was inactivated by heating the solution to 70 °C for 10 minutes.

$$\boldsymbol{Amount\ of\ insert\ (ng/\mu L)} = \frac{\boldsymbol{(amount\ of\ vector\ (ng) \times size\ of\ insert\ (bp) \times ratio)/size\ of\ vector\ (bp)}}{\boldsymbol{concentration\ of\ insert\ (ng/\mu L)}}$$

Equation 1

Table 16: Composition of Ligation Solutions

Chemical	Quantity (µL)		
	Vector Alone	1:3	1:6
Ligase buffer	1	1	1
T4 DNA ligase	1	1	1
Vector	90 ng (dependent on concentration)		
Insert	-	Calculated using Equation 1	
Water	To make up to 10 µL		

8.3.10 Making Chemically Competent Cells

A 1% culture was made from an overnight culture of *Escherichia coli* (*E. coli*) (strains DH5α or BL21) in 25 mL LB medium containing the relevant antibiotic. The culture was grown at 37 °C until an optical density at 600 nm (OD_{600}) of 0.5-0.7 was reached. The culture was then centrifuged (1332 g, 15 minutes, 4 °C) and the supernatant discarded. The pellet was resuspended in 5 mL TBF I buffer, Table 17, and then centrifuged (1332 g, 15 minutes, 4 °C). The supernatant was discarded and the pellet resuspended in TBF II buffer (Table 18). This resulting solution was divided into 100 µL aliquots and frozen over dry ice before storing at -80 °C until required.

Table 17: Composition of TBF I buffer. Solution was adjusted to pH 5.8 with 1% acetic acid and then made up to 200 mL and stored at -80 °C.

Chemical	Final Concentration
Potassium acetate	30 mM
Rubidium chloride	100 mM
Calcium chloride	10 mM
Manganese chloride	50 mM
Glycerol	15% (v/v)

Table 18: Composition of TBF II buffer. Solution was adjusted to pH 6.5 and then made up to 100 mL with autoclaved deionised water and stored at -80 °C.

Chemical	Final Concentration
3-(N-morpholino)propandesulfonic acid (MOPS)	10 mM
Calcium chloride	10 mM
Rubidium chloride	75 mM
Glycerol	15% (v/v)

8.3.11 Transformation into Chemically Competent Cells

Successfully ligated plasmids (5 µL) were transformed directly into one 100 µL aliquot of chemically competent cells and incubated on ice for 30 minutes. A water-only negative control and a positive control of vector were completed in parallel. The cells were then heat shocked for 30 seconds in a 42 °C water bath. The cells were then transferred to 895 µL of prewarmed SOC (Super Optimal broth with Catabolite repression), Table 19, and incubated at 37°C with agitation for 1 hour. The culture (100 µL) was then spread over LB agar containing the relevant antibiotic (See 8.3.4) and incubated at 37 °C for 18 hours.

Table 19: Composition of SOC medium. The solution was filtered through a 0.22 µM sterile filter (Millipore, UK) and stored in 10 mL aliquots at 4 °C.

Chemical	Quantity
1 M Magnesium chloride	500 µL
2 M Magnesium sulphate	500 µL
20% glucose solution (w/v)	500 µL
Autoclaved LB broth	50 mL

8.3.12 Procedure for Colony PCR

Colonies were picked from the LB agar plates of successful transformations and added to the reagents in Table 20. The thermocycler was set up for the run detailed in, Table 20. The reactions were then analysed on a 1% agarose gel at 100 V for 30 minutes.

Table 20: Composition for Colony PCR

Reagent	Quantity (µL)
GoTaq® 5× Green buffer	2
dNTP mix (10 mM)	1
Forward primer (10 pmol/µL)	0.25
Reverse primer (10 pmol/µL)	0.25
GoTaq® polymerase	0.05
Deionised water	6.45
Template	1 colony

Table 21: Thermo Cycler Set Up for Colony PCR. Where X = 51°C for pET-28a primers and X = 56°C for pGEX-2TK primers.

Cycle Number	Denaturation	Annealing	Extension
1	95 °C for 2 mins	-	-
2-30	95 °C for 30 secs	X °C for 30 secs	72 °C for 30 mins
31	-	-	72 °C for 10 mins

8.3.13 Expression and Purification of His$_6$-tagged UEV and GST-tagged p6

Both recombinant proteins were expressed in BL21 (DE3) *Escherichia coli* cells. Bacterial cultures were incubated at 37 °C with shaking until an OD$_{600}$ of between 0.5 and 0.7 was reached. Protein expression was induced with 0.1 mM IPTG overnight at 16°C. After harvesting, the cell pellets were suspended at 1 g/mL in Buffer A (20 mM Tris-HCl, 100 mM NaCl, 10% glycerol, pH 7.4) supplemented with 1 % (v/v) Triton X-100, Pierce Protease Inhibitor tablets (ThermoFisher Scientific, 1 tablet per 20 mL) and lysozyme (100 µg/ mL), and the cells were incubated on ice for 30 minutes. The lysate was sonicated (12 cycles of 10s on followed by 10s off) on ice to ensure complete lysis. Insoluble cell debris was pelleted via centrifugation at 22,700 × g for 45 min at 4 °C. The supernatant was filtered using 0.45 µm syringe filters. For His$_6$-UEV, the supernatant was loaded onto a 1 mL HisTrap column (GE Healthcare) equilibrated in Buffer A coupled to an ÄKTA Prime fast protein liquid chromatography (FPLC) system (GE Healthcare). The protein was loaded onto the column at a flow rate of 1 mL/min. The column was washed with 30 mL of Buffer A and then with a gradient of 8% Buffer B (500 mM imidazole, 20 mM Tris-HCl, 100 mM NaCl, pH 7.4) in Buffer A at a flow rate of 1 mL/min, then the protein was eluted in 100% Buffer B. UV absorbance at 280 nm was monitored and fractions were collected and analyzed by SDS-PAGE. For GST-p6, the filtered supernant after cell lysis was loaded onto a 1 mL GSTrap column (GE Healthcare) coupled to an ÄKTA Prime FPLC system (GE Healthcare) pre-equilibrated in Buffer A at a flow rate of 1 mL/min. The protein was loaded onto the column at a flow rate of 0.5 mL/min. The column was washed with 30 mL of Buffer A, and the GST-p6 was eluted with 10 mL GST Elution buffer (50 mM Tris-HCl, 150 mM NaCl, 15 mM reduced glutathione, pH 7.4) at a flow rate of 1 mL/min. For both proteins, the fractions containing the desired protein were combined and dialysed overnight at 4°C using SnakeSkin™ Dialysis Tubing 10K MWCO 22 mm (ThermoFisher Scientific) into Buffer A (1 L/ 2 mL protein). The buffer was exchanged for fresh and allowed to dialyse for another 3 hours. His$_6$-UEV was obtained with a yield of 20 mg per litre cell culture, and GST-p6 was obtained with a yield of 10 mg per litre cell culture. Protein aliquots were flash-frozen in liquid nitrogen and stored at −80 °C.

8.3.14 Procedure for Running SDS Gels

Gels of differing percentages for protein analysis by SDS-PAGE analysis were made using the amounts detailed in, Table 22 (the resolving gels were made to be 10, 12 or 15%, and then the stacking gel was layered on top).

Table 22: Reagents for Making Gels for SDS-PAGE Analysis of Proteins

Reagent	Resolving Gel Percentage			Stacking Gel
	10%	12%	15%	
1.5 M Tris-Base (pH 8.8)	7.5 mL	7.5 mL	7.5 mL	-
0.5 M Tris-Base (pH 6.8)	-	-	-	1.25 mL
10% (w/v) SDS Solution	0.3 mL	0.3 mL	0.3 mL	0.05 mL
Acrylamide/bisacrylamine	10.0 mL	12.0 mL	15.0 mL	0.84 mL
10% Ammonium Persulfate	0.15 mL	0.15 mL	0.15 mL	0.05 mL
Tetramethylethylenediamine	0.02 mL	0.02 mL	0.02 mL	0.005 mL
dH_2O	12.03 mL	10.03 mL	7.03 mL	2.88 mL

The protein samples were diluted 1:1 in 2 × Laemmli Buffer (100 mM Tris-HCl (pH 6.8), 4% SDS Solution (w/v), 20% Glycerol, 50 mM dithiothreitol, trace of bromophenol blue) and then heated to 95 °C for 10 minutes then allowing to cool to 4 °C for 5 minutes prior to loading into the wells of the gel. The gels were run for 60-70 minutes at 150 V in SDS-PAGE Running Buffer (diluted from 5 × to 1 ×), Table 23.

Table 23: SDS-PAGE Running Buffer (5 ×)

Reagent (Final Concentration)	Amount
Tris-Base (120 mM)	15.1 g
Glycine (1.25 M)	94 g
20% (w/v) SDS Solution (0.5%)	25 mL
dH_2O	Up to 1 L

8.3.14.1 Visualising SDS-PAGE Gels

SDS-PAGE gels were visualised by staining with Coomassie Blue Solution (400 mL Methanol, 100 mL Glacial acetic acid, 1 g Coomassie, 500 mL dH_2O) for 10 minutes. The gels were then destained using Destain Buffer (200 mL MeOH, 100 mL acetic acid, 700 mL dH_2O) for 1hr to reveal the protein bands.

8.4 Experimental for Chapter 2

8.4.1 CP11

CP11 (IYWNVSGW) was synthesised as 8.2.1, on a Glycine Wang resin and cleaved to form linear CP11 (LCP11A1; pink solid, 180 mg crude). This peptide (180 mg, 0.176 mmol) was cyclised to form CP11, Figure 68, and purified via preparative HPLC yielding a pale purple solid (62 mg, 0.062 mmol, 35.0% yield). LR MS (ESI+) *m/z* (%) 504.0 (½$[M+2H]^{2+}$, 100%), 1028.7 ($[M+Na]^{+}$, 62.7%), 1006.7 ($[M+H]^{+}$, 46.6%). Analytical HPLC (280 nm) 21.3 min (85%).

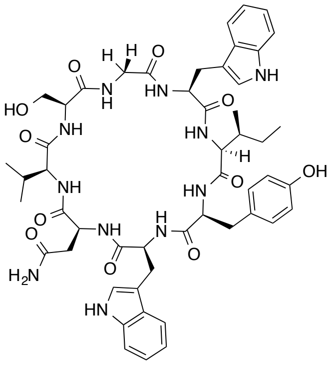
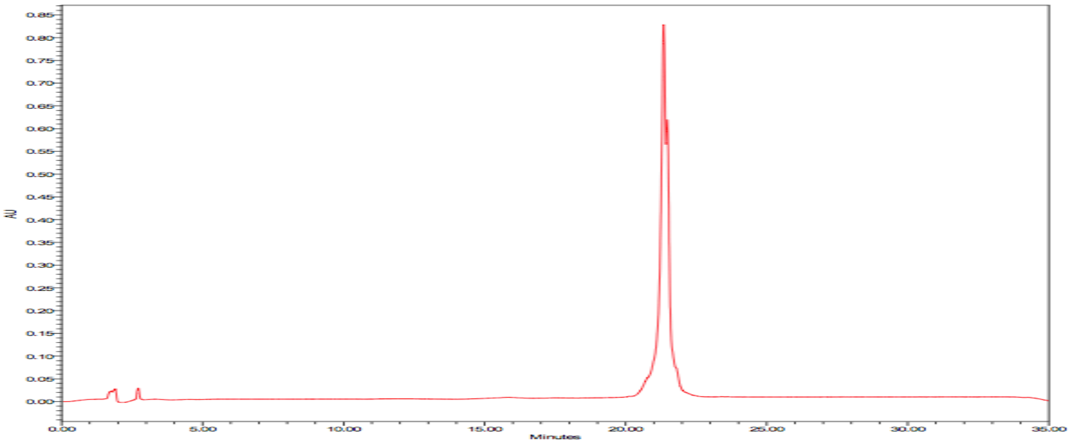
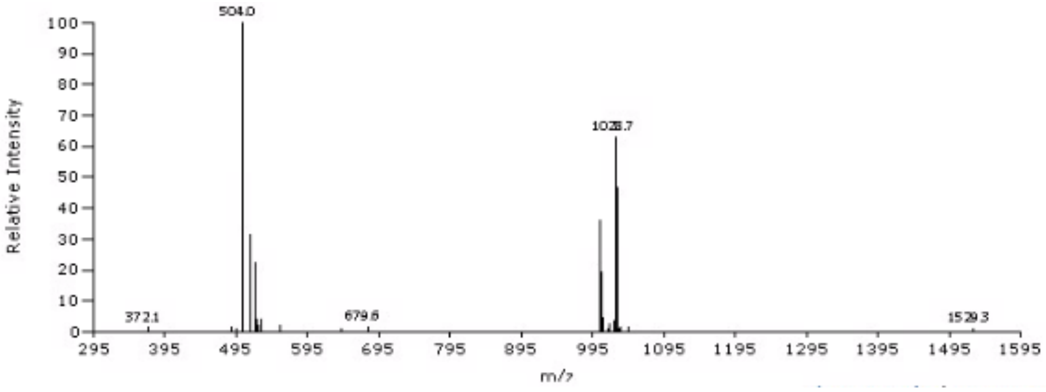


Figure 68: CP11





8.4.2 Liquid Phase Synthesis of Tri-Peptides

8.4.2.1 Synthesis of Acetyl-NWY-Diethylamine

8.4.2.1.1 Synthesis of Fmoc-Tyr(tBu)-Diethylamine

Fmoc-Tyr(tBu)-OH (1.55 g, 3.22 mmol) was dissolved in DCM (15 mL) with one equivalent each of HOBt (0.44 g, 3.22 mmol) and EDC (0.62 g, 3.22 mmol). Diethylamine (0.7 mL, 6.5 mmol) was added dropwise and the reaction left stirring under argon for 4 hours at room temperature, Figure 69. The reaction was stopped and then diluted with DCM (20 mL). The organic layer was washed with sodium bicarbonate solution (10% w/v, 20 mL), citric acid (20 mL) and brine (20 mL). It was then dried over anhydrous sodium sulphate. The solution was filtered and the solvent removed by rotary evaporation. The crude yellow oil was purified via column chromatography (10% MeOH in DCM) yielding a white solid (1.38 g, 2.68 mmol, 83% yield, R_f 0.74).

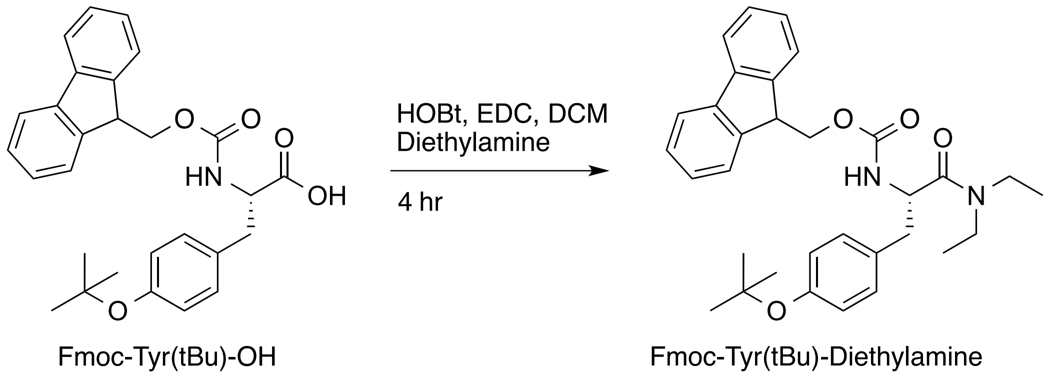


Figure 69: Synthesis of Fmoc-Tyr(tBu)-Diethylamine

8.4.2.1.2 Synthesis of H_2N-Tyr(tBu)-Diethylamine

Fmoc-Tyr(tBu)-diethylamine (1.38 g, 2.68 mmol) was dissolved in ethyl acetate (20 mL). 1,8-Diazabicyclo[5.4.0]undec-7-ene (DBU, 0.41 mL, 3.0 mmol) was added dropwise and the reaction left stirring under argon at room temperature for 3 hours, Figure 70. The solvent was removed by rotary evaporation and the crude oil was purified by column chromatography (10% MeOH in DCM). Following on from purification a yellow oil was obtained (0.45 g, 1.54 mmol, 58% yield, R_f 0.32).

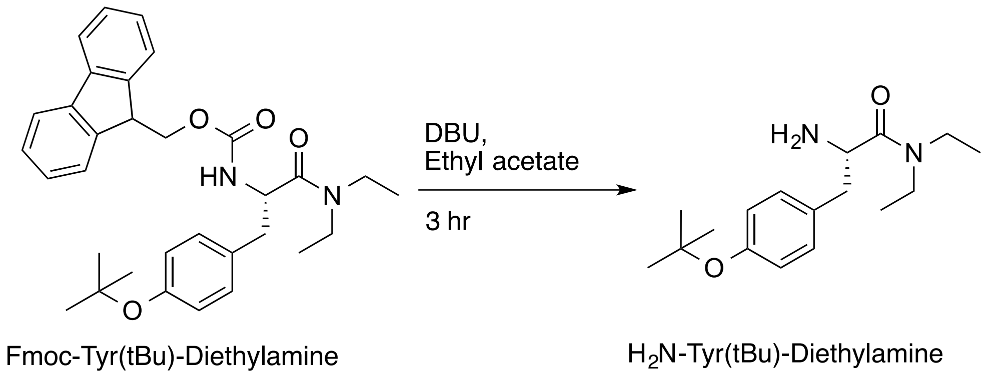


Figure 70: Synthesis of H_2N-Tyr(tBu)-Diethylamine

8.4.2.1.3 Synthesis of Fmoc-Trp(Boc)-Tyr(tBu)-Diethylamine

H_2N-Tyr(tBu)-diethylamine (0.83 g, 2.84 mmol) was dissolved in DCM (15 mL) with two equivalents of Fmoc-Trp(Boc)-OH (3.0 g, 5.7 mmol), EDC (1.1 g, 5.7 mmol) and HOBt (0.78 g, 5.7 mmol). The reaction was left to stir for 18 hours, under argon at room temperature, Figure 71. The organic layer was washed with sodium bicarbonate solution (10% w/v, 20 mL), citric acid (20 mL) and brine (20 mL). It was then dried over anhydrous sodium sulphate. The solution was filtered and the solvent removed by rotary evaporation. The crude yellow oil was purified via column chromatography (9:1, DCM:MeOH) yielding a white solid (2.16 g, 2.70 mmol, 95% yield, R_f 0.76).

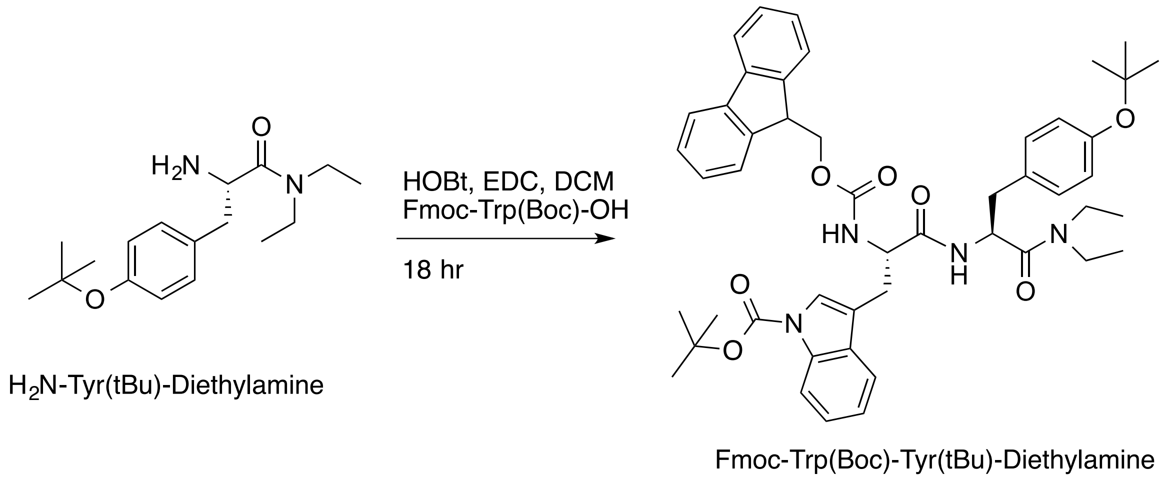


Figure 71: Synthesis of Fmoc-Trp(Boc)-Tyr(tBu)-Diethylamine

8.4.2.1.4 Synthesis of H_2N-Trp(Boc)-Tyr(tBu)-Diethylamine

Fmoc-Trp(Boc)-Tyr(tBu)-diethylamine (2.16 g, 2.70 mmol) was dissolved in ethyl acetate (20 mL). DBU (0.41 mL, 3.0 mmol) was added dropwise and the reaction left stirring under argon at room temperature for 3 hours, Figure 72. The solvent was removed by rotary evaporation and the crude oil was purified by column chromatography (10% MeOH in DCM). Following on from purification a crude, yellow oil was obtained (2.06 g, 3.56 mmol, crude, R_f 0.53). This oil was taken directly onto the next step so as to avoid loss of material.

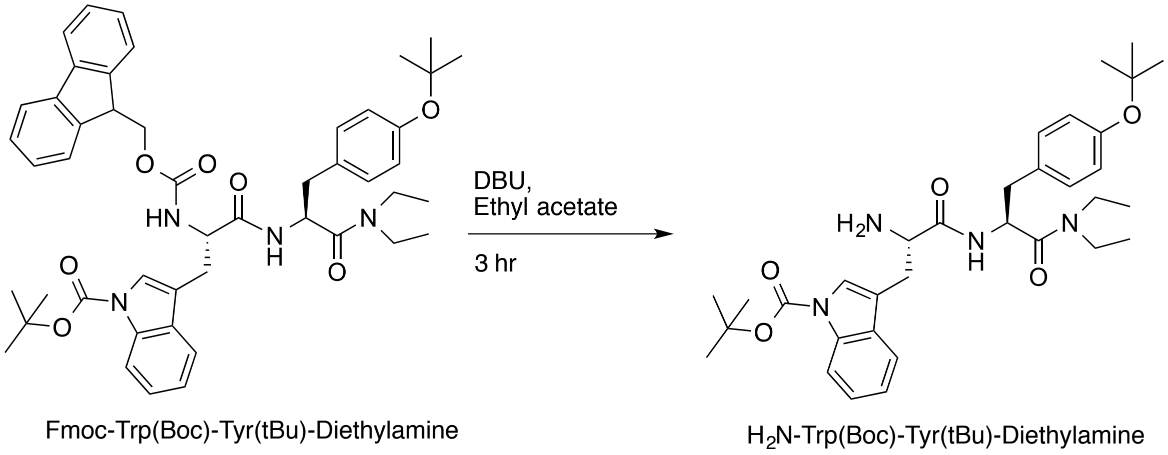


Figure 72: Synthesis of H_2N-Trp(Boc)-Tyr(tBu)-Diethylamine

8.4.2.1.5 Synthesis of Fmoc-Asn(Trt)-Trp(Boc)-Tyr(tBu)-Diethylamine

H_2N-Trp(Boc)-Tyr(tBu)-diethylamine (2.06 g, 2.84 mmol) was dissolved in DCM (15 mL) with two equivalents of Fmoc-Asn(Trt)-OH (3.4 g, 5.7 mmol), EDC (1.1 g, 5.7 mmol) and HOBt (1.78 g, 5.7 mmol). The reaction was left to stir for 18 hours, under argon at room temperature, Figure 73. The organic layer was washed with sodium bicarbonate solution (10% w/v, 20 mL), citric acid (20 mL) and brine (20 mL). It was then dried over anhydrous sodium sulphate. The solution was filtered and the solvent removed by rotary evaporation. The crude yellow oil was purified via column chromatography (9:1, DCM:MeOH) yielding a white solid (3.24 g, 2.80 mmol, 98 % yield, R_f 0.79).

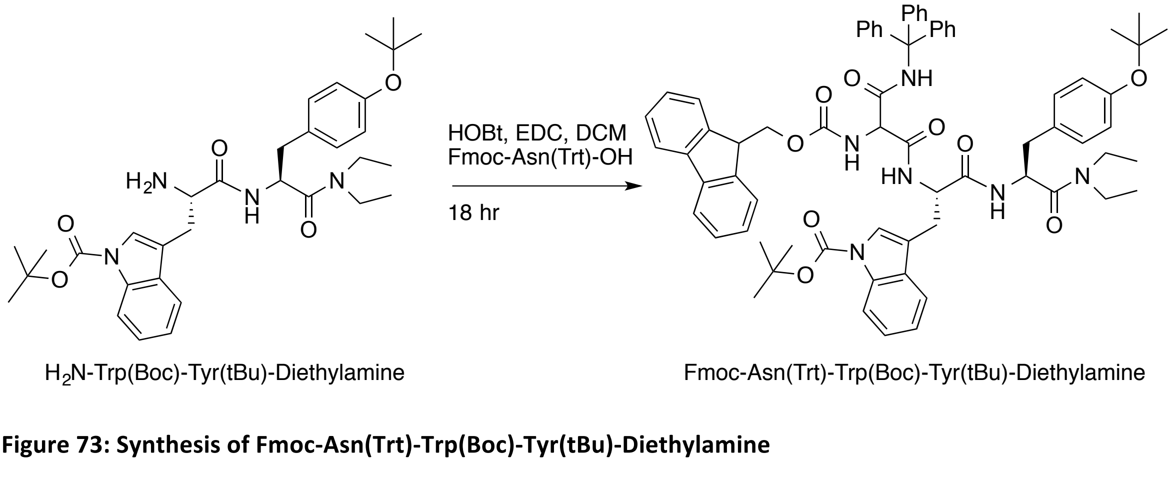
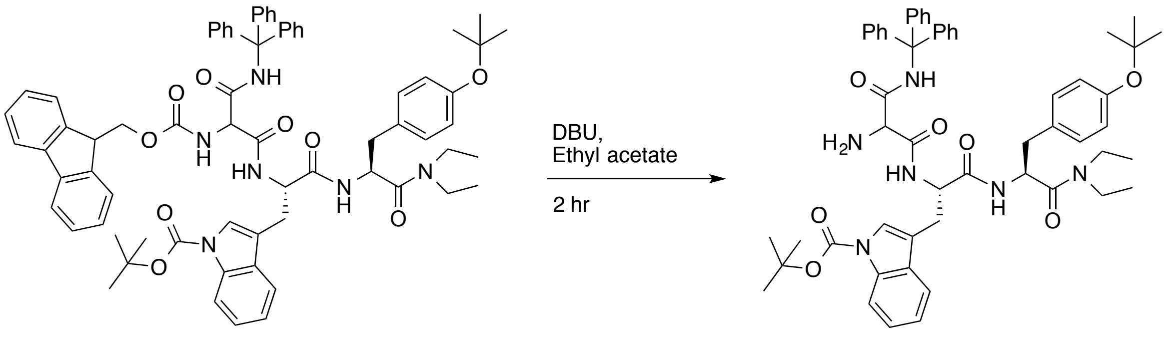


Figure 73: Synthesis of Fmoc-Asn(Trt)-Trp(Boc)-Tyr(tBu)-Diethylamine

8.4.2.1.6 Synthesis of H_2N-Asn(Trt)-Trp(Boc)-Tyr(tBu)-Diethylamine

Fmoc-Asn(Trt)-Trp(Boc)-Tyr(tBu)-diethylamine (3.24 g, 2.80 mmol) was dissolved in ethyl acetate (20 mL). DBU (0.42 mL, 2.8 mmol) was added dropwise and the reaction left stirring under argon at room temperature for 2 hours, Figure 74. The solvent was removed by rotary evaporation and the crude oil was purified by column chromatography (10% MeOH in DCM). Following on from purification a crude, yellow oil was obtained (3.15 g, 3.37 mmol, crude, R_f 0.55). This oil was taken directly onto the next step so as to avoid loss of material.



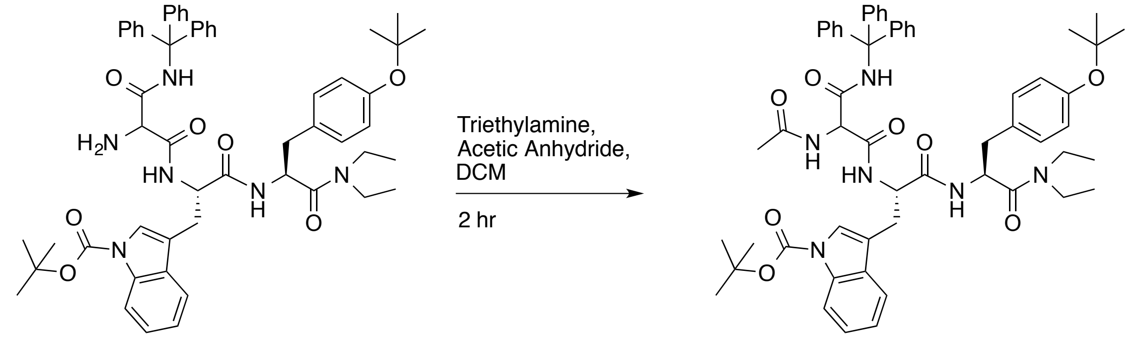
Fmoc-Asn(Trt)-Trp(Boc)-Tyr(tBu)-Diethylamine

H_2N-Asn(Trt)-Trp(Boc)-Tyr(tBu)-Diethylamine

Figure 74: Synthesis of H_2N-Asn(Trt)-Trp(Boc)-Tyr(tBu)-Diethylamine

8.4.2.1.7 Synthesis of Acetyl-Asn(Trt)-Trp(Boc)-Tyr(tBu)-Diethylamine

H_2N-Asn(Trt)-Trp(Boc)-Tyr(tBu)-diethylamine (3.15 g, 3.37 mmol crude) was dissolved in DCM (10 mL). One equivalent each of triethylamine (0.5 mL, 3.4 mmol) and acetic anhydride (0.3 mL, 3.4 mmol) were added dropwise and the reaction stirred under argon for 2 hours at room temperature, Figure 75. The solvent was removed by rotary evaporation and the crude oil was purified by column chromatography (10% MeOH in DCM) yielding a yellow oil (2.4 g, 2.45 mmol, 87% yield [calculated from 8.4.2.1.5], R_f 0.53).



H_2N-Asn(Trt)-Trp(Boc)-Tyr(tBu)-Diethylamine

Acetyl-Asn(Trt)-Trp(Boc)-Tyr(tBu)-Diethylamine

Figure 75: Synthesis of Acetyl-Asn(Trt)-Trp(Boc)-Tyr(tBu)-Diethylamine

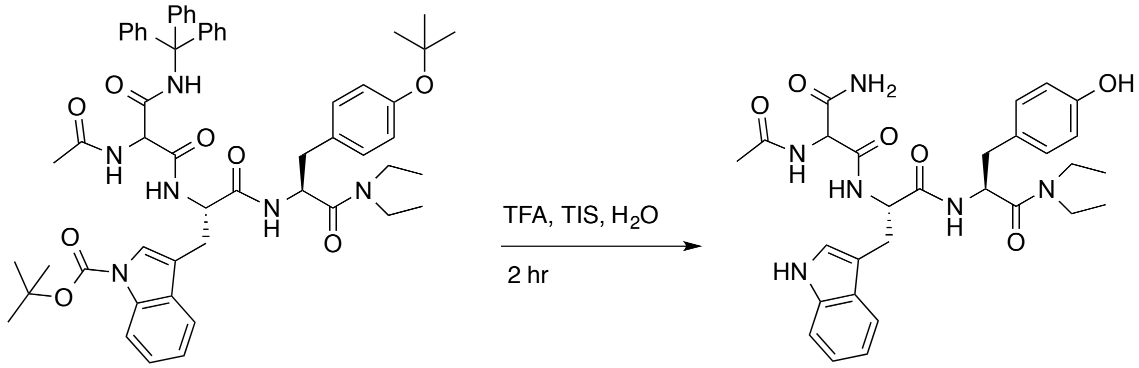
8.4.2.1.8 Synthesis of Acetyl-Asn-Trp-Tyr-Diethylamine

Acetyl-Asn(Trt)-Trp(Boc)-Tyr(tBu)-diethylamine (2.4 g, 2.45 mmol) was dissolved in a cleavage cocktail of TFA, TIS and water (9.5:0.25:0.25, 10 mL) and stirred for 2 hours, Figure 76. The solvent was removed by rotary evaporation, and the peptide was washed with DCM (3 × 10 mL) and then precipitated in ice cold diethyl ether (10 mL). The peptide was filtered and allowed to dry yielding an orange power (1.35 g, 2.33 mmol, 95% yield). The peptide was purified by RP-HPLC and lyophilised to yield a pale cream powder (0.875 g, 1.55 mmol, 48% overall yield). LR MS (ESI+) *m/z* (%) 579.4 ($[M+H]^+$, 100%), 601.5 ($[M+Na]^+$, 84.55%). HR HPLC-MS (ESI+) for $C_{30}H_{38}N_6O_6$ $(M+H)^+$ Calculated mass 579.2930, found mass 579.2926 (-0.7 ppm error). Analytical HPLC (280 nm) 16.37 min (95%).

^{1}H NMR (400 MHz, DMSO-d_6) δ = 10.81 (1H, d, J = 1.5 Hz, Trp Indole NH), 8.26 (1H, br d, J6,2 = 8.3 Hz, Asn backbone NH), 8.05 (1H, d, J2,6 = 7.9 Hz, Trp backbone NH), 7.74 (1H, d, J3,25 = 8.1 Hz, Tyr backbone NH), 7.51 (1H, d, J25,3 = 7.8 Hz, Trp E3-H), 7.32 (1H, s, Asn NH_2), 7.30 (1H, br d, J22,35, 33 = 7.9 Hz, Trp Z2-H), 7.04 (1H, d, J = 5.4 Hz, Trp HD1-H), 7.04 (1H, t, J = 7.5 Hz, Trp H2-H), 6.98 (2H, d, J = 8.6 Hz, Tyr D*-H), 6.95 (1H, t, J = 7.0 Hz, Trp Z3-H), 6.88 (1H, br s, Asn NH_2), 6.64 (2H, d, J = 8.4 Hz, Tyr E*-H), 4.71 (1H, td, J = 8.5 Hz, J = 6.0 Hz, Tyr α-H), 4.55 (1H, td, J = 7.7 Hz, J = 6.5 Hz, Trp α-H), 4.50 (1H, td, J = 7.7 Hz, J = 5.6 Hz, Asn α-H), 3.35 (1H, sxt, J = 6.4 Hz, overlapped diethylamine CH_2), 3.09 - 3.19 (1H, m, overlapped diethylamine CH_2), 2.89 - 3.09 (4H, m, overlapped diethylamine CH_2, Trp β-H), 2.84 (1H, br dd, J = 3.1 Hz, J = 8.4 Hz, Tyr β-H), 2.65 (1H, br dd, J = 3.1 Hz, J = 6.2 Hz, Tyr β-H), 2.39 (1H, br dd, J = 6.4 Hz, J = 8.2 Hz, Asn β-H), 2.31 (1H, dd, J = 15.3 Hz, J = 7.8 Hz, Asn β-H), 1.78 (3H, s, acetyl CH_3), 0.94 (3H, t, J = 7.0 Hz, diethylamine CH_3), 0.84 (3H, t, J = 7.0 Hz, diethylamine CH_3)

^{1}H NMR obtained at C4X Discovery. Coupling constants were not calculated as were not necessary for the computational model.

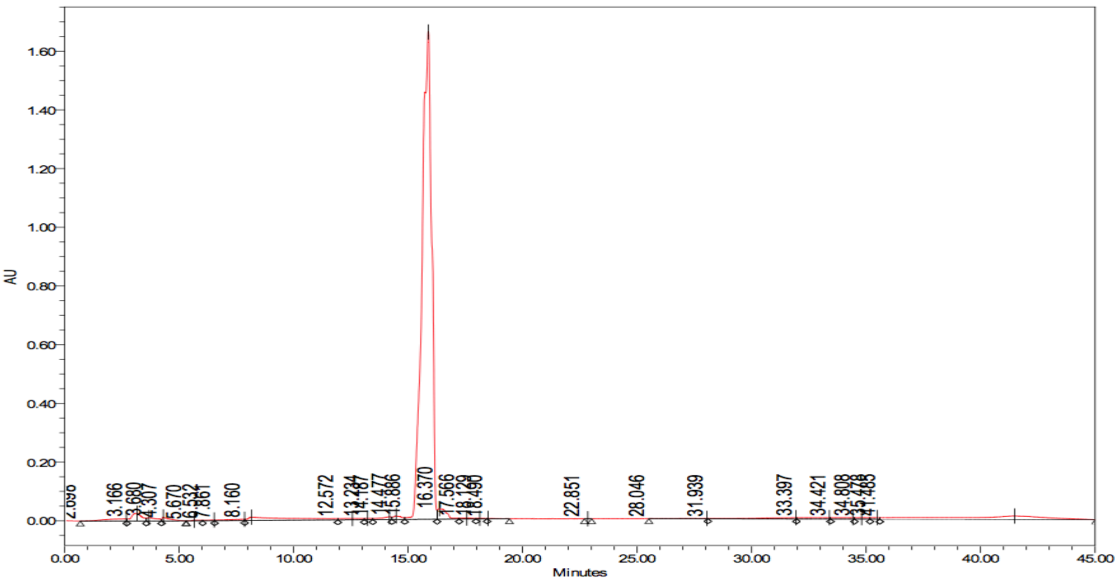
^{1}H NMR (800 MHz, H_2O:D_2O, 9:1, 10°C) δ = 10.220 (1H, d, J = Hz, Trp indole NH), 8.393 (1H, d, Asn backbone NH), 7.904 (1H, d, Trp backbone NH), 7.904 (1H, d, Tyr backbone NH), 7.631 (1H, s, Asn HD21), 7.587 (1H, d, Trp HE3), 7.466 (1H, d, Trp HZ2), 7.224 (1H, t, Trp HH2), 7.163 (1H, s, Trp HD1), 7.141 (1H, t, Trp HZ3), 7.020 (2H, d, Tyr HD*), 6.990 (2H, s, Asn HD22), 6.773 (2H, d, Tyr HE*), 4.753 (1H, dt, Tyr α-H), 4.649 (1H, dt, Trp α-H), 4.640 (1H, dt, Asn α-H), 3.321 (1H, dt, diethylamine H221), 3.216 (1H, dd, Trp Hβ1), 3.177 (1H, dd, Trp Hβ2), 3.020 (1H, dt, diethylamine H211), 2.993 (1H, dt, diethylamine H222), 2.961 (1H, dt, diethylamine H212), 2.806 (1H, dd, Tyr Hβ1), 2.695 (1H, dd, Tyr Hβ2), 2.686 (1H, dd, Asn Hβ1), 2.616 (1H, dd, Asn Hβ2), 1.929 (3H, s, acetyl CH_3), 0.938 (3H, t, diethylamine 32*CH_3), 0.922 (3H, t, diethylamine 31*CH_3).

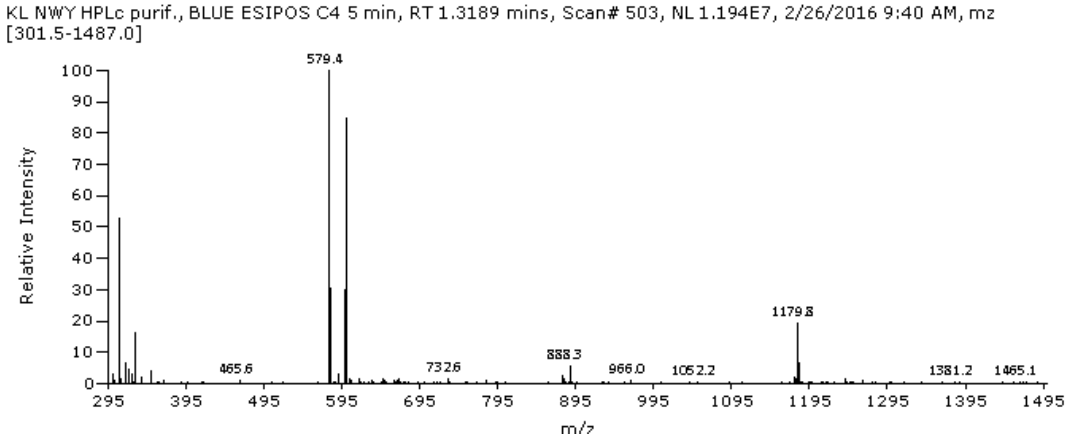


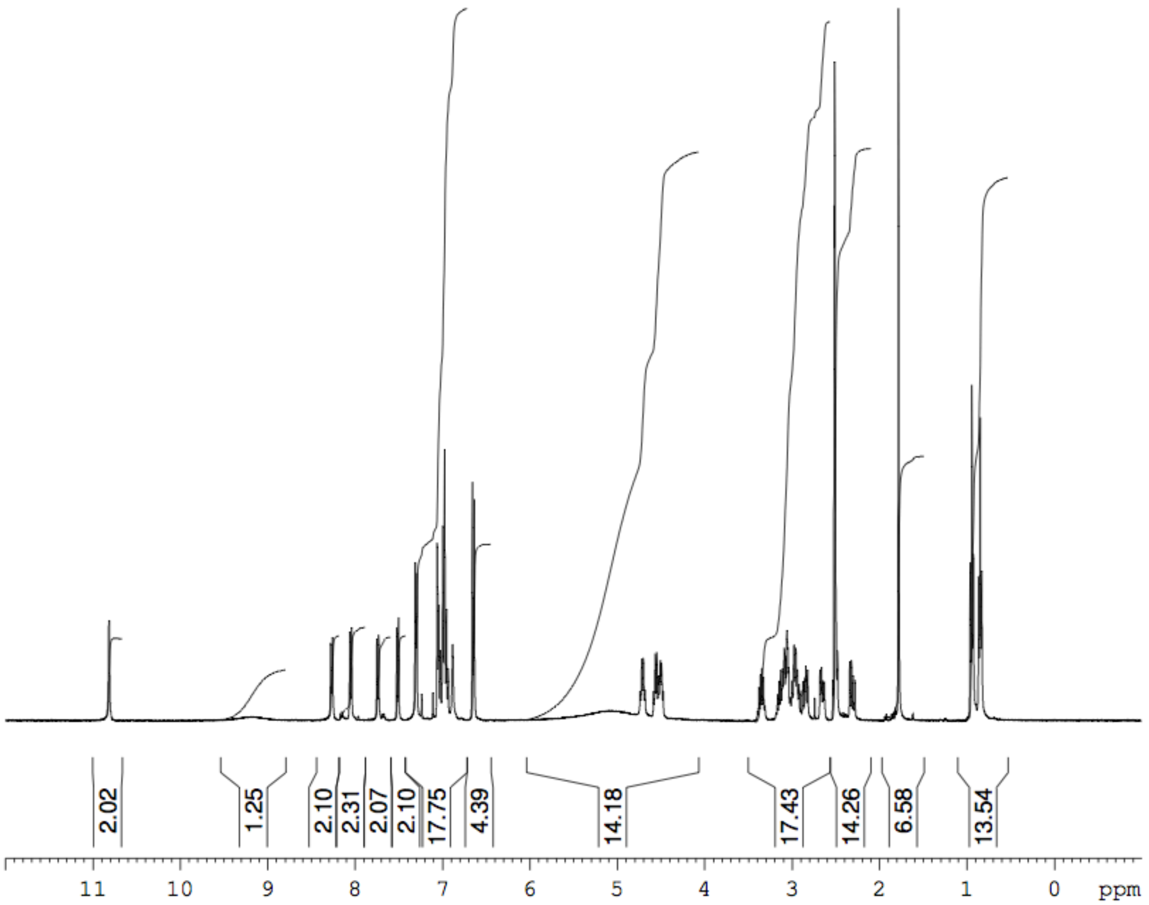
Acetyl-Asn(Trt)-Trp(Boc)-Tyr(tBu)-Diethylamine

Acetyl-Asn-Trp-Tyr-Diethylamine

Figure 76: Synthesis of Acetyl-Asn-Trp-Tyr-Diethylamine







8.4.2.2 Synthesis of Acetyl-YWN-Diethylamine

8.4.2.2.1 Synthesis of Fmoc-Asn(Trt)-Diethylamine

Fmoc-Asn(Trt)-OH (2.01 g, 3.35 mmol) was dissolved in DCM (15 mL) with one equivalent each of HOBt (0.44 g, 3.3 mmol) and EDC (0.62 g, 3.3 mmol). Diethylamine (0.7 mL, 6.5 mmol) was added dropwise and the reaction left stirring under argon for 4 hours at room temperature, Figure 77. The reaction was stopped and then diluted with DCM (20 mL). The organic layer was washed with sodium bicarbonate solution (10% w/v, 20 mL), citric acid (20 mL) and brine (20 mL). It was then dried over anhydrous sodium sulphate. The solution was filtered and the solvent removed by rotary evaporation. The crude yellow oil was purified via column chromatography (10% MeOH in DCM) yielding a white solid (1.59 g, 2.44 mmol, 73% yield, R_f 0.81).

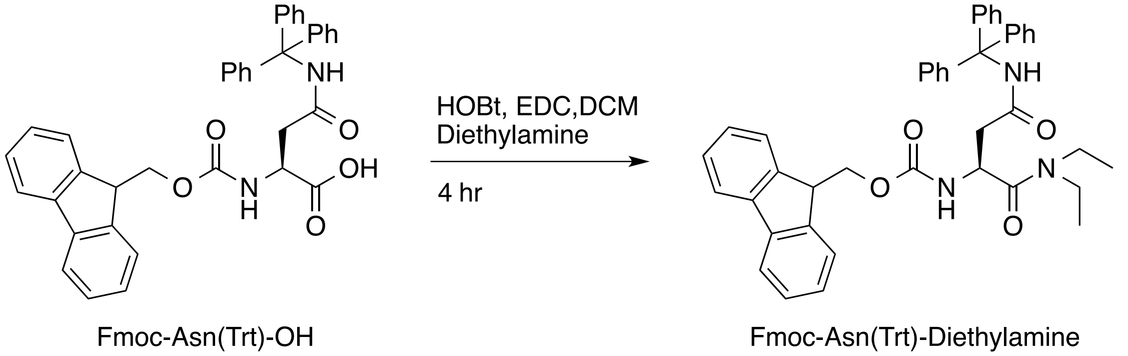


Figure 77: Synthesis of Fmoc-Asn(Trt)-Diethylamine

8.4.2.2.2 Synthesis of H_2N-Asn(Trt)-Diethylamine

Fmoc-Asn(Trt)-diethylamine (1.59 g, 2.44 mmol) was dissolved in ethyl acetate (20 mL). DBU (0.41 mL, 2.5 mmol) was added dropwise and the reaction left stirring under argon at room temperature for 3 hours, Figure 78. The solvent was removed by rotary evaporation and the crude oil was purified by column chromatography (10% MeOH in DCM). Following on from purification a white powder was obtained (0.84 g, 1.96 mmol, 80% yield, R_f 0.39).

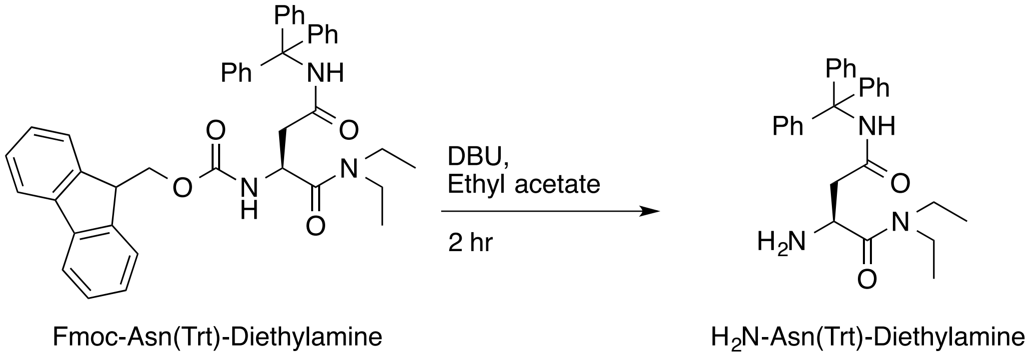


Figure 78: Synthesis of H_2N-Asn(Trt)-Diethylamine

8.4.2.2.3 Synthesis of Fmoc-Trp(Boc)-Asn(Trt)-Diethylamine

H_2N-Asn(Trt)-diethylamine (0.84 g, 1.96 mmol) was dissolved in DCM (15 mL) with two equivalents of Fmoc-Trp(Boc)-OH (2.1 g, 3.9 mmol), EDC (0.8 g, 3.9 mmol) and HOBt (0.53 g, 3.9 mmol). The reaction was left to stir for 18 hours, under argon at room temperature, Figure 79. The organic layer was washed with sodium bicarbonate solution (10% w/v, 20 mL), citric acid (20 mL) and brine (20 mL). It was then dried over anhydrous sodium sulphate. The solution was filtered and the solvent removed by rotary evaporation. The crude yellow oil was purified via column chromatography (9:1, DCM:MeOH) yielding a yellow oil (1.61 g, 1.71 mmol, 87% yield, R_f 0.79).

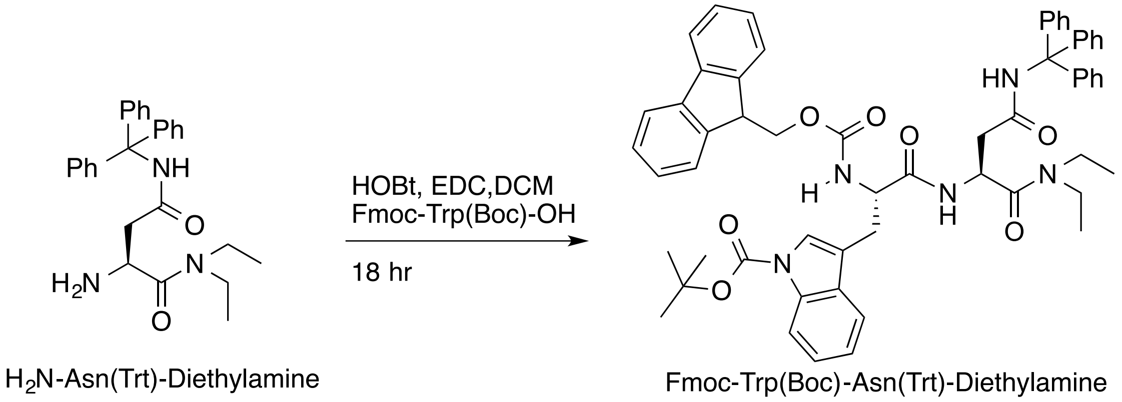


Figure 79: Synthesis of Fmoc-Trp(Boc)-Asn(Trt)-Diethylamine

8.4.2.2.4 Synthesis of H$_2$N-Trp(Boc)-Asn(Trt)-Diethylamine

Fmoc-Trp(Boc)-Asn(Trt)-diethylamine (1.61 g, 1.71 mmol) was dissolved in ethyl acetate (20 mL). DBU (0.41 mL, 2.5 mmol) was added dropwise and the reaction left stirring under argon at room temperature for 2 hours, Figure 80. The solvent was removed by rotary evaporation and the crude oil was purified by column chromatography (10% MeOH in DCM). Following on from purification a yellow oil was obtained (1.20 g, 1.71 mmol, crude, R$_f$ 0.47), this was crude however to avoid loss of product it was taken directly onto the next step.

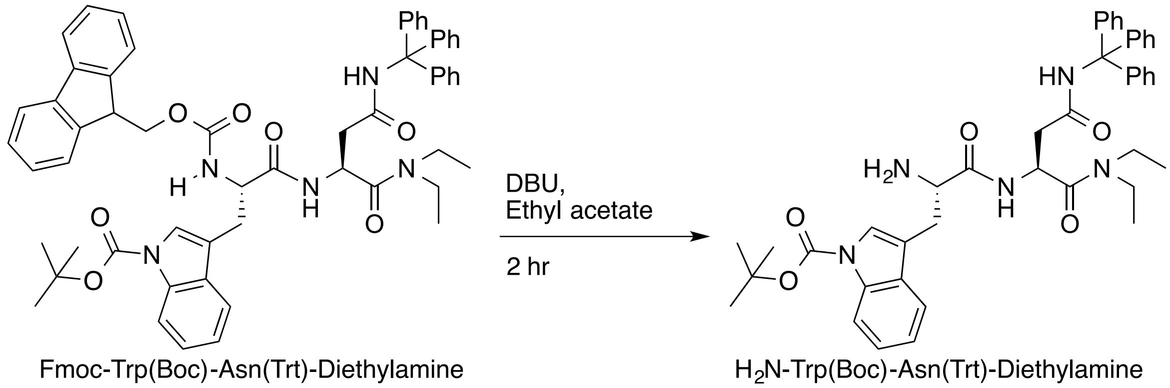
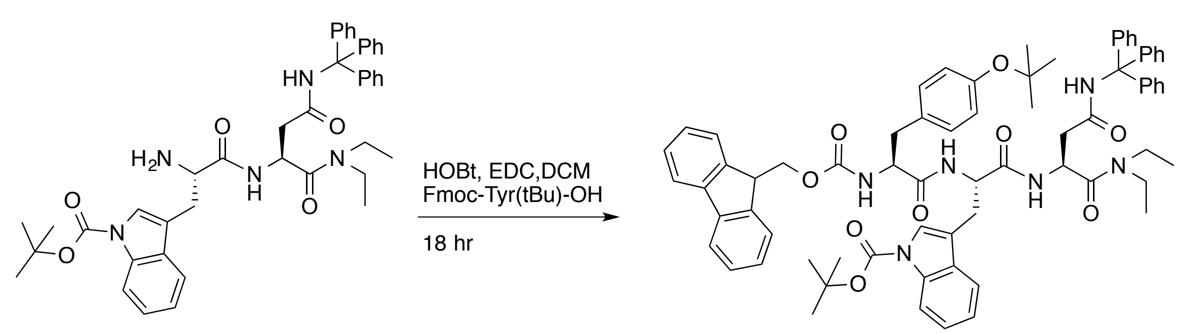


Figure 80: Synthesis of H$_2$N-Trp(Boc)-Asn(Trt)-Diethylamine

8.4.2.2.5 Synthesis of Fmoc-Tyr(tBu)-Trp(Boc)-Asn(Trt)-Diethylamine

H$_2$N-Trp(Boc)-Asn(Trt)-diethylamine (1.20 g, 1.71 mmol) was dissolved in DCM (15 mL) with two equivalents of Fmoc-Tyr(tBu)-OH (1.6 g , 3.4 mmol), EDC (0.7 g, 3.4 mmol) and HOBt (0.5 g, 3.4 mmol). The reaction was left to stir for 18 hours, under argon at room temperature, Figure 81. The organic layer was washed with sodium bicarbonate solution (10% w/v, 20 mL), citric acid (20 mL) and brine (20 mL). It was then dried over anhydrous sodium sulphate. The solution was filtered and the solvent removed by rotary evaporation. The crude yellow oil was purified via column chromatography (9:1, DCM:MeOH) yielding a yellow solid (1.69 g, 1.46 mmol, 86% yield, R_f 0.79).



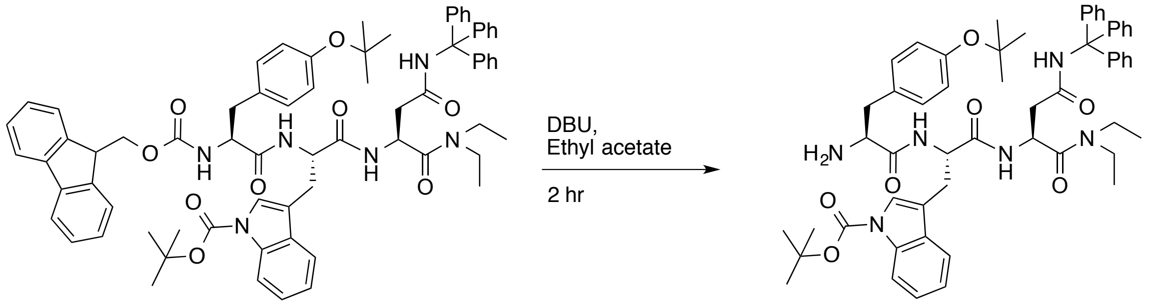
H$_2$N-Trp(Boc)-Asn(Trt)-Diethylamine

Fmoc-Tyr(tBu)-Trp(Boc)-Asn(Trt)-Diethylamine

Figure 81: Synthesis of Fmoc-Tyr(tBu)-Trp(Boc)-Asn(Trt)-Diethylamine

8.4.2.2.6 Synthesis of H$_2$N-Tyr(tBu)-Trp(Boc)-Asn(Trt)-Diethylamine

Fmoc-Tyr(tBu)-Trp(Boc)-Asn(Trt)-diethylamine (1.69 g, 1.46 mmol) was dissolved in ethyl acetate (20 mL). DBU (0.41 mL, 2.5 mmol) was added dropwise and the reaction left stirring under argon at room temperature for 2 hours, Figure 82. The solvent was removed by rotary evaporation and the crude oil was purified by column chromatography (10% MeOH in DCM). Following on from purification a pale yellow solid was obtained (1.27 g, 1.36 mmol, 93% yield, R_f 0.54).



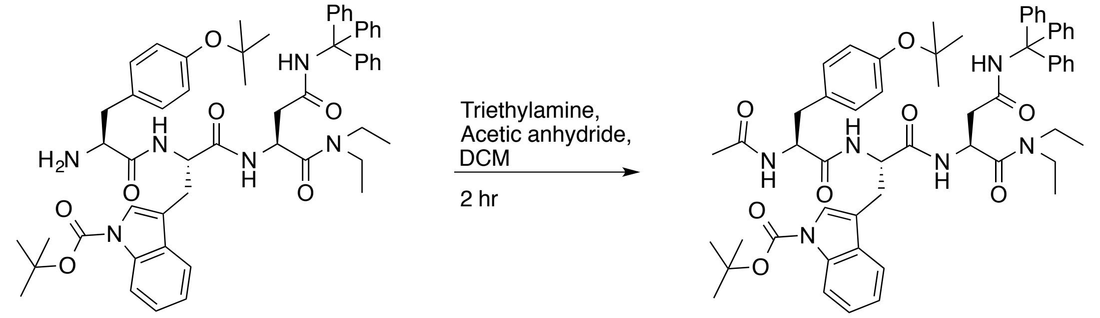
Fmoc-Tyr(tBu)-Trp(Boc)-Asn(Trt)-Diethylamine

H$_2$N-Tyr(tBu)-Trp(Boc)-Asn(Trt)-Diethylamine

Figure 82: Synthesis of H$_2$N-Tyr(tBu)-Trp(Boc)-Asn(Trt)-Diethylamine

8.4.2.2.7 Synthesis of Acetyl-Tyr(tBu)-Trp(Boc)-Asn(Trt)-Diethylamine

H$_2$N-Tyr(tBu)-Trp(Boc)-Asn(Trt)-diethylamine (1.27 g, 1.36 mmol crude) was dissolved in DCM (10 mL). One equivalent each of triethylamine (0.1 mL, 1.4 mmol) and acetic anhydride (0.2 mL, 1.4 mmol) were added dropwise and the reaction stirred under argon for 2hours at room temperature, . The solvent was removed by rotary evaporation and the crude oil was purified by column chromatography (10% MeOH in DCM) yielding a crude yellow solid (1.36 g, 1.38 mmol, crude, R_f 0.56). This solid was taken directly onto the next step so as to avoid excess loss of material.



H_2N-Tyr(tBu)-Trp(Boc)-Asn(Trt)-Diethylamine

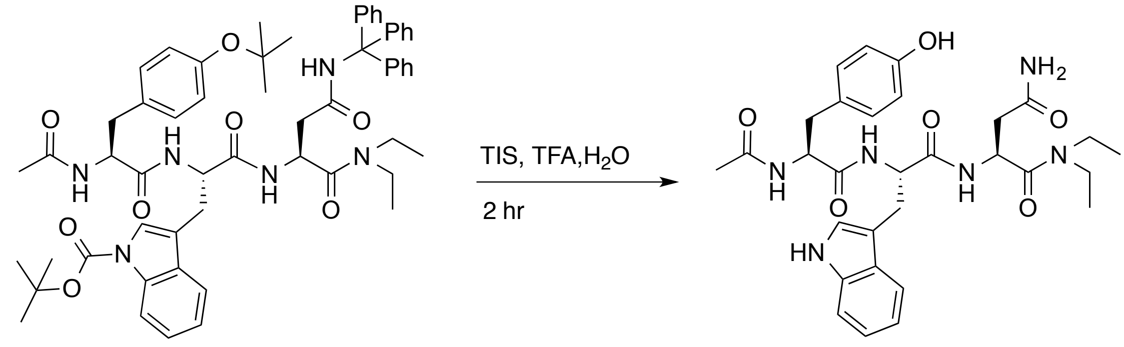
Acetyl-Tyr(tBu)-Trp(Boc)-Asn(Trt)-Diethylamine

Figure 83: Synthesis of Acetyl-Tyr(tBu)-Trp(Boc)-Asn(Trt)-Diethylamine

8.4.2.2.8 Synthesis of Acetyl-Tyr-Trp-Asn-Diethylamine

Acetyl-Tyr(tBu)-Trp(Boc)-Asn(Trt)-diethylamine (1.36 g, 1.38 mmol) was dissolved in a cleavage cocktail of TFA, TIS and water (9.5:0.25:0.25, 10 mL) and stirred for 2 hours, Figure 84. The solvent was removed by rotary evaporation, and the peptide was washed with DCM (3 × 10 mL) and then precipitated in ice cold diethyl ether (10 mL). The peptide was filtered and allowed to dry yielding a pink power (0.72 g, 1.28 mmol, 93% yield). The peptide was purified by RP-HPLC and lyophilised to yield an off white solid (0.25 g, 0.43 mmol, 34%). LR MS (ESI+) *m/z* (%) 601.5 ($[M+Na]^+$, 100%), 579.5 ($[M+H]^+$, 79.67%). HR HPLC-MS (ESI+) for $C_{30}H_{38}N_6O_6$ $(M+H)^+$ Calculated mass 579.2926, found mass 579.2924 (0.3 ppm error). Analytical HPLC (280 nm) 15.97 min (96%).

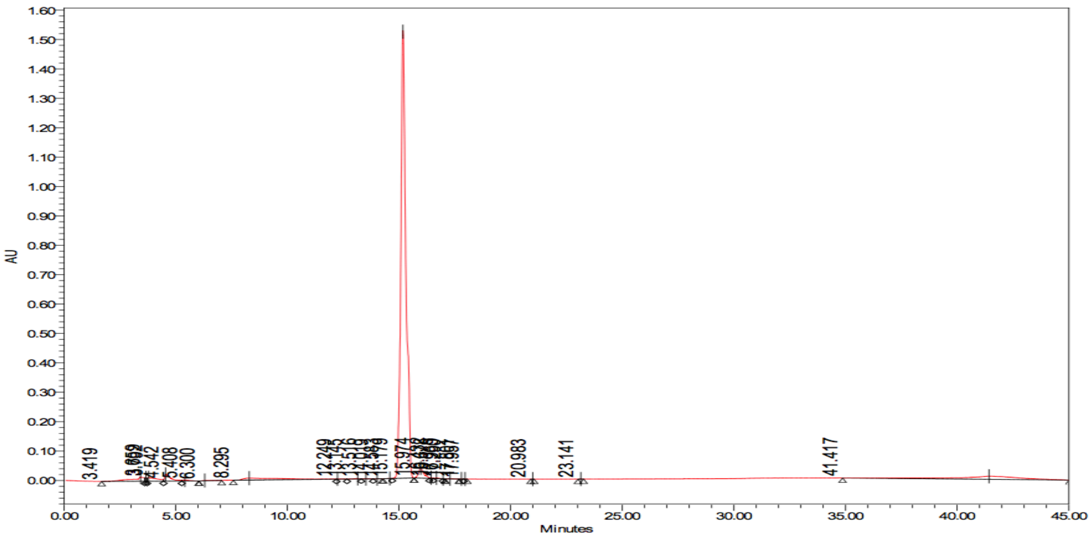
^1H NMR (400 MHz, DMSO-d_6) δ = 10.82 (1H, d, J = 2.0 Hz, Trp indole NH), 8.28 (1H, d, J3,30 = 8.8 Hz, Asn backbone NH), 7.94 (2H, d, J = 8.1 Hz, Trp backbone NH), 7.94 (2H, d, J = 8.3 Hz, Tyr backbone NH), 7.57 (1H, d, J = 7.8 Hz, Trp E3-H), 7.31 (1H, br d, J26,15, 13 = 8.2 Hz, Trp Z2-H), 7.27 - 7.30 (1H, overlapped s, Tyr phenol OH), 7.10 (1H, d, J = 1.8 Hz, Trp H2-H), 7.05 (1H, t, J = 7.1 Hz, Tyr D*-H), 6.94 - 7.00 (3H, m, Overlapped Tyr D*-H, Trp Z3-H), 6.78 (1H, br s, Asn NH_2), 6.61 (2H, d, J15, 13,26 = 8.4 Hz, Tyr E*-H), 4.99 (1H, td, J30,33 = 8.6 Hz, J30,33 = 5.2 Hz, Asn α-H), 4.54 (1H, td, J = 7.7 Hz, J = 5.8 Hz, Trp α-H), 4.38 (1H, ddd, J = 9.8 Hz, J = 8.7 Hz, J = 4.0 Hz, Tyr α-H), 3.16 - 3.31 (2H, m, diethylamine CH_2), 3.13 - 3.24 (2H, m, diethylamine CH_2), 3.09 (1H, dd, J = 1.7 Hz, J = 5.4 Hz, Trp β-H), 2.94 (1H, dd, J = 1.8 Hz, J = 8.1 Hz, Trp β-H), 2.84 (1H, dd, J = 1.1 Hz, J = 3.9 Hz, Tyr β-H), 2.65 (1H, dd, J33,33 = 1.2 Hz, J33,30 = 8.8 Hz, Asn β-H), 2.52 - 2.58 (1H, m, Tyr β-H), 2.21 (1H, dd, J33,33 = 1.2 Hz, J33,30 = 5.2 Hz, Asn β-H), 1.71 (3H, s, Acetyl CH_3), 1.07 (3H, t, J40,41 = 7.1 Hz, Diethylamine CH_3), 0.96 (3H, t, J41,40 = 7.0 Hz, Diethylamine CH_3).

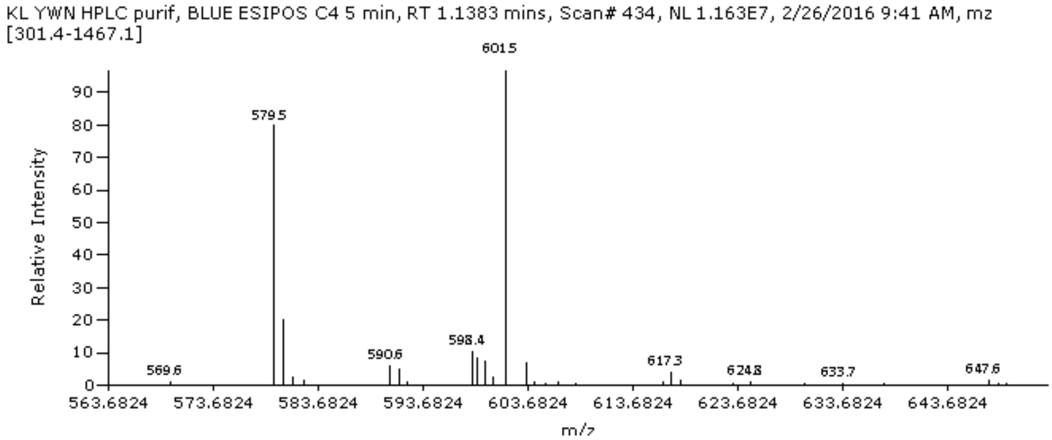


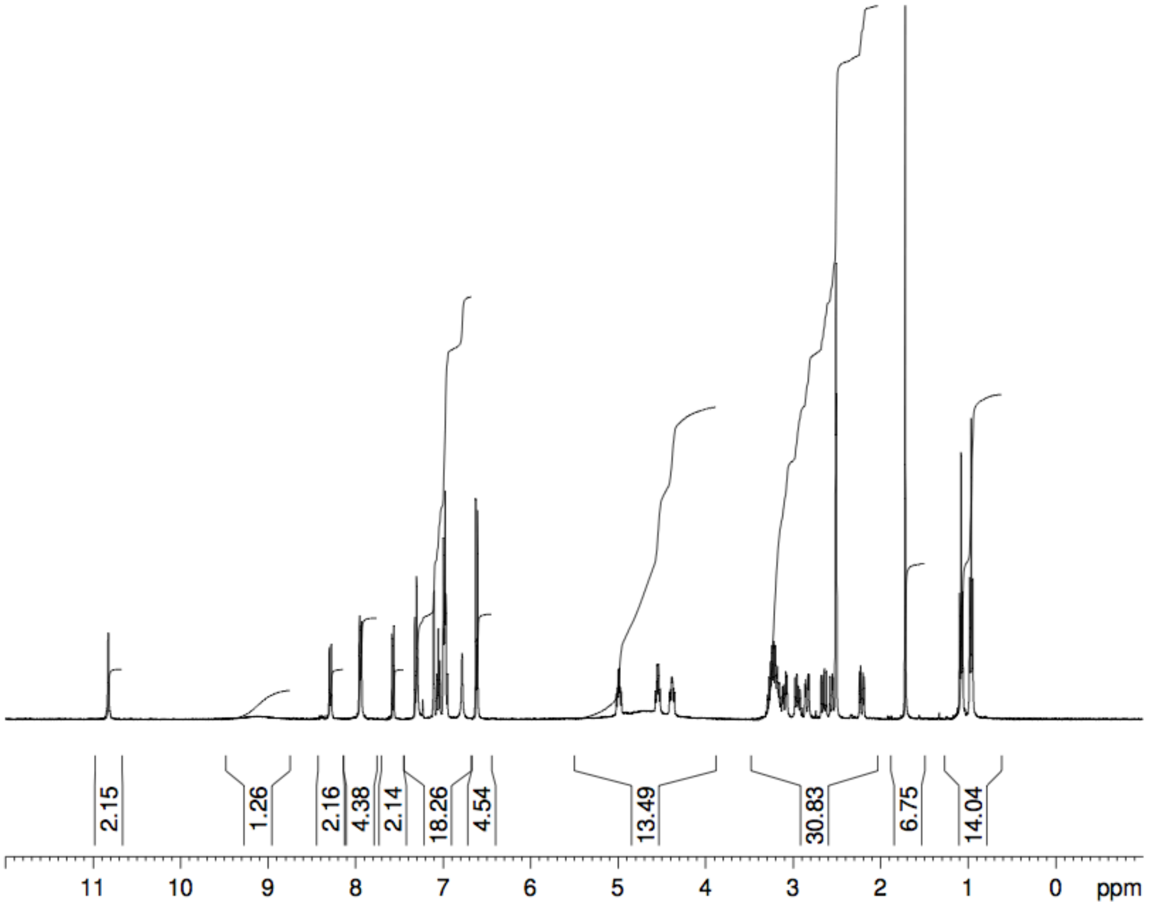
Acetyl-Tyr(tBu)-Trp(Boc)-Asn(Trt)-Diethylamine

Acetyl-Tyr-Trp-Asn-Diethylamine

Figure 84: Synthesis of Acetyl-Tyr-Trp-Asn-Diethylamine




KL YWN HPLC purif, BLUE ESIPOS C4 5 min, RT 1.1383 mins, Scan# 434, NL 1.163E7, 2/26/2016 9:41 AM, mz [301.4-1467.1]



8.4.3 Solid Phase Synthesis of Linked Tri-Peptides

NWY was synthesised on a Tyrosine Wang resin (0.3 mmol scale) as 8.2.1, and then split three ways (~0.1 mmol each).

8.4.3.1 Cyclo AhxNWY

Fmoc-Ahx-OH was coupled onto the resin-bound NWY. It was deprotected as 8.2.1 and then cleaved from the resin to form linear AhxNWY (white solid, 149.0 mg crude). This peptide (149.0 mg, 0.25 mmol) was cyclised to form cyclo-AhxNWY, Figure 85, and purified via preparative HPLC yielding a white solid (34.4 mg, 0.060 mmol, 24.0% yield). LR MS (ESI+) *m/z* (%) 577.5 ($[M+H]^+$, 100%), 599.5 ($[M+Na]^+$, 55.8%). HR HPLC-MS (ESI+) for $C_{30}H_{36}N_6O_6$ $(M+H)^+$ Calculated mass 577.2766, found mass 577.5769 (0.6 ppm error). $C_{30}H_{35}N_6NaO_6$ $(M+Na)^+$ Calculated mass 599.2581, found mass 599.2589 (1.2 ppm error).

^{1}H NMR (400 MHz, DMSO-d$_6$) δ = 10.77 (1H, br d, J = 9.3 Hz, Trp indole NH), 9.13 (1H, br s, Tyr OH), 8.21 (1H, dd, J = 22.0 Hz, J = 7.7 Hz, Tyr backbone NH), 8.11 (1H, dd, J = 8.1 Hz, J = 2.3 Hz, Trp backbone NH), 7.96 (1H,d, J = 7.3 Hz, Asn backbone NH), 7.83 (1H, d, J = 6.8 Hz, Ahx backbone NH) 7.60 (1H, br dd, J = 6.7 Hz, J = 4.2 Hz, Trp E3-H), 7.50 (1H, d, J = 7.8 Hz, Trp Z2-H), 7.27 - 7.40 (3H, m, Trp H2-H), 7.03 - 7.10 (1H, m, Trp D1-H), 6.94 - 7.03 (2H, m, Tyr D*-H), 6.88 - 6.94 (2H, m, Trp Z3-H), 6.61 (2H, d, J = 8.4 Hz, Tyr E*-H), 4.29 - 4.36 (1H, m, Trp α-H), 4.24 (1H, dt, J = 2.0 Hz, J = 7.0 Hz, Asn α-H), 4.06 - 4.15 (1H, m, Tyr α-H), 3.19 - 3.32 (1H, m, Ahx α-H), 3.05 (1H, br d, J = 7.3 Hz, Asn β-H), 2.88 - 3.02 (1H, m, Trp β-H), 2.76 - 2.88 (1H, m, Trp β-H), 2.68 (1H, dd, J = 3.4 Hz, J = 7.3 Hz, Tyr β-H), 2.53 - 2.58 (1H, m, Tyr β-H), 2.32 (1H, ddd, J = 1.7 Hz, J = 1.4 Hz, J = 4.2 Hz, Asn β-H), 2.00 - 2.22 (2H, m, Ahx E-H), 1.54 (1H, br d, J = 3.4 Hz, Ahx D-H), 1.29 - 1.46 (2H, m, Ahx B-H), 1.08 - 1.29 (2H, m, Ahx G-H).

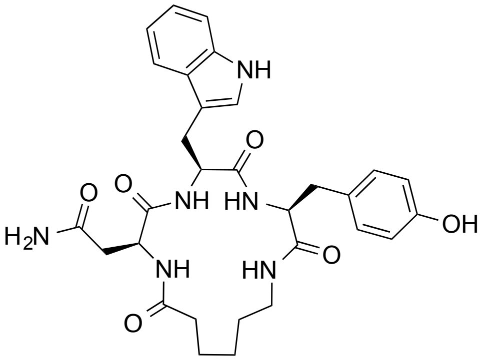
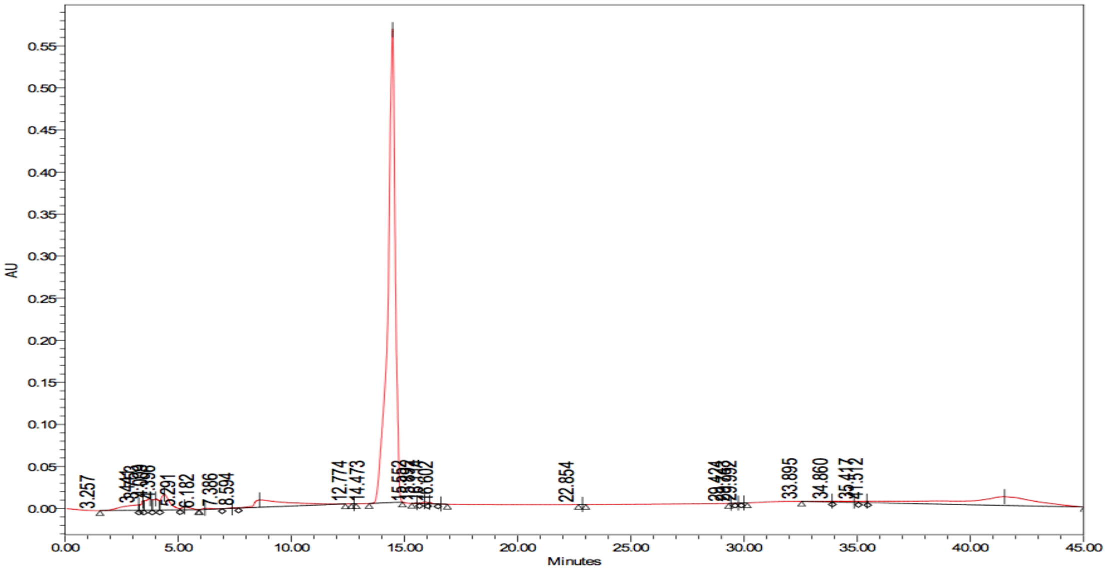
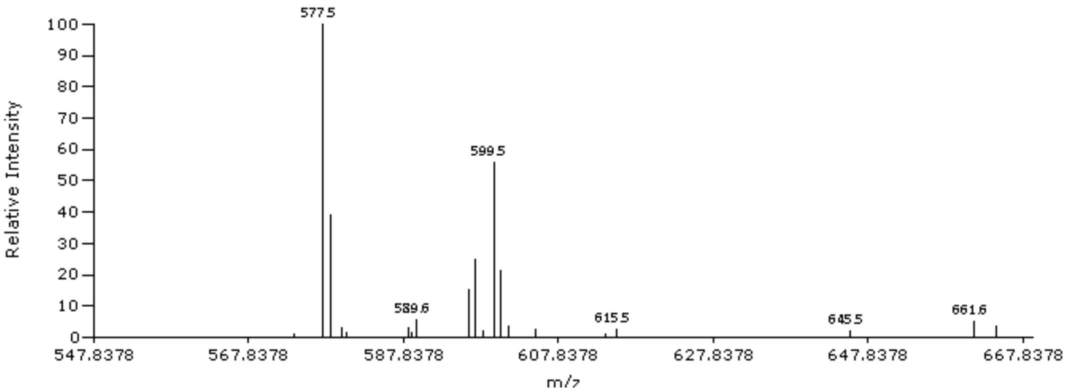
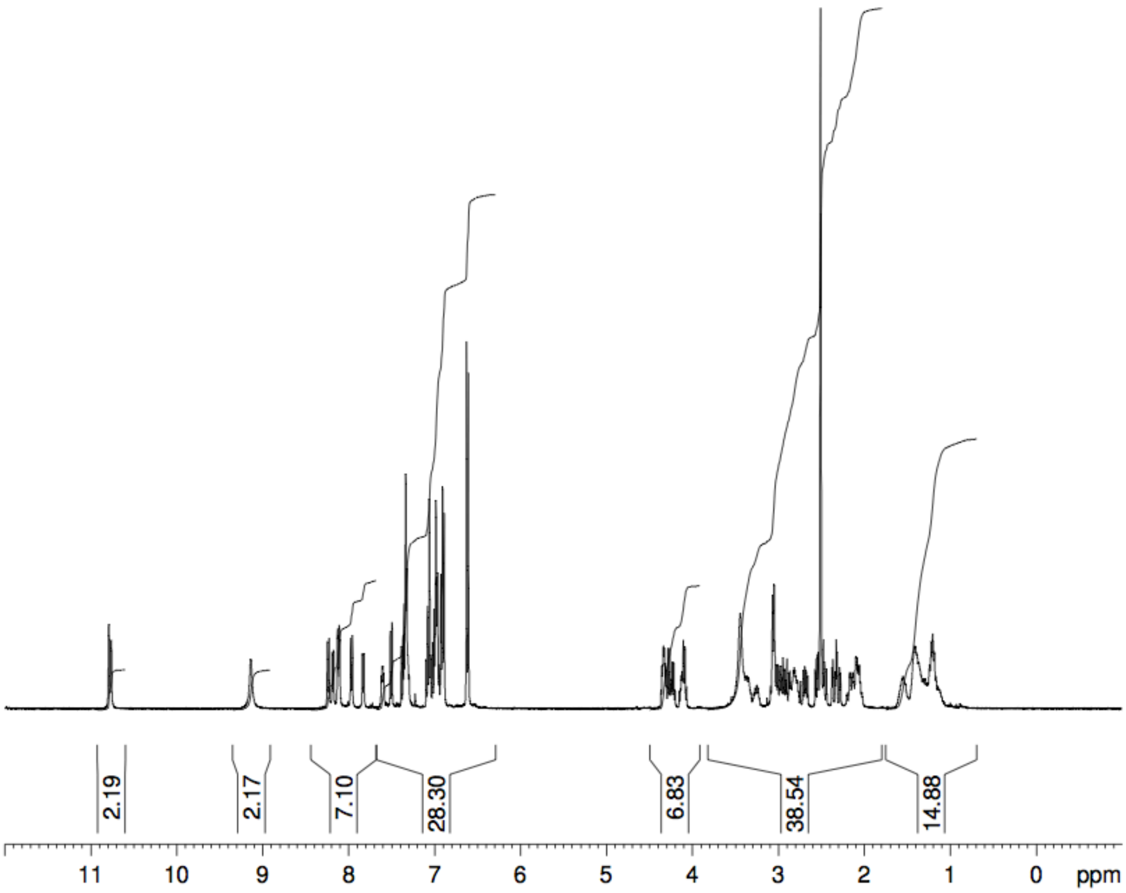


Figure 85: Cyclo-Ahx-NWY



KL cyclo AhxNWY, BLUE ESIPOS C18 5 min , RT 0.7946 mins, Scan# 226, NL 1.200E7, 11/29/2016 1:51 PM, mz [167.3-1462.2]





8.4.3.2 Cyclo AvaNWY

Fmoc-Ava-OH was coupled onto the resin-bound NWY. It was deprotected as 8.2.1 and then cleaved from the resin to form linear AvaNWY (white solid, 83.7 mg crude). This peptide (83.7 mg, 0.144 mmol) was cyclised to form cyclo-AvaNWY, Figure 86, and purified via preparative HPLC yielding a pale yellow solid (24.3 mg, 0.043 mmol, 30% yield). LR MS (ESI+) *m/z* (%) 585.5 ($[M+Na]^+$, 100%), 563.5 ($[M+H]^+$, 97.4%). HR HPLC-MS (ESI+) for $C_{29}H_{35}N_6O_6$ $(M+H)^+$ Calculated mass 563.2604, found mass 563.2613 (1.6 ppm error). $C_{29}H_{35}N_6NaO_6$ $(M+Na)^+$ Calculated mass 585.2432, found mass 585.2420 (2.1 ppm error).

^1H NMR (400 MHz, DMSO-d_6) δ = 10.72 - 10.82 (1H, m, Trp indole NH), 9.15 (1H, br s, Tyr phenol-OH), 8.33 (1H, d, J = 6.2 Hz, Asn backbone NH), 8.29 (1H, br dd, J = 35.0 Hz, J = 6.2 Hz, Ava backbone NH), 8.01 - 8.14 (1H, m, Tyr backbone NH), 7.95 (1H, d, J = 8.7 Hz, Trp backbone NH), 7.48 (1H, d, J = 7.9 Hz, Trp Z2-H), 7.37 - 7.44 (1H, m, Trp E3-H), 7.34 (1H, dd, J = 8.0 Hz, J = 5.7 Hz, Trp Z3- H), 6.98 - 7.10 (1H, m, Trp D1-H), 6.91 - 6.98 (3H, overlapped, m, Tyr D*-H, Trp H2-H), 6.64 (3H, dd, J = 11.1 Hz, J = 8.8 Hz, Tyr E*-H), 4.34 (1H, ddd, J = 10.3 Hz, J = 6.4 Hz, J = 3.7 Hz, Asn α-H), 4.14 - 4.29 (2H, overlapped, m, Tyr α-H, Trp α-H), 3.26 - 3.45 (1H, m, Trp β-H), 3.19 (1H, dd, J = 14.6 Hz, J = 6.2 Hz, Trp β-H), 2.94 - 3.13 (2H, m, Tyr β-H), 2.64 - 2.80 (2H, m, Asn β-H), 2.38 - 2.48

(2H, m, Ava α-CH_2), 2.13 - 2.27 (1H, m, Ava δ-CH_2), 1.89 - 2.02 (2H, m, Ava δ-CH_2), 1.61 (1H, br s, Ava γ-CH_2), 1.41 - 1.51 (2H, m, Ava β-CH_2), 1.27 - 1.41 (1H, m, Ava γ-CH_2).

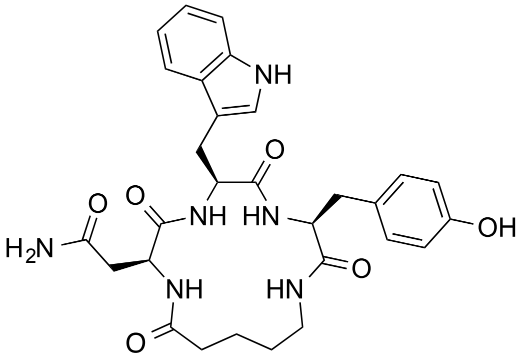
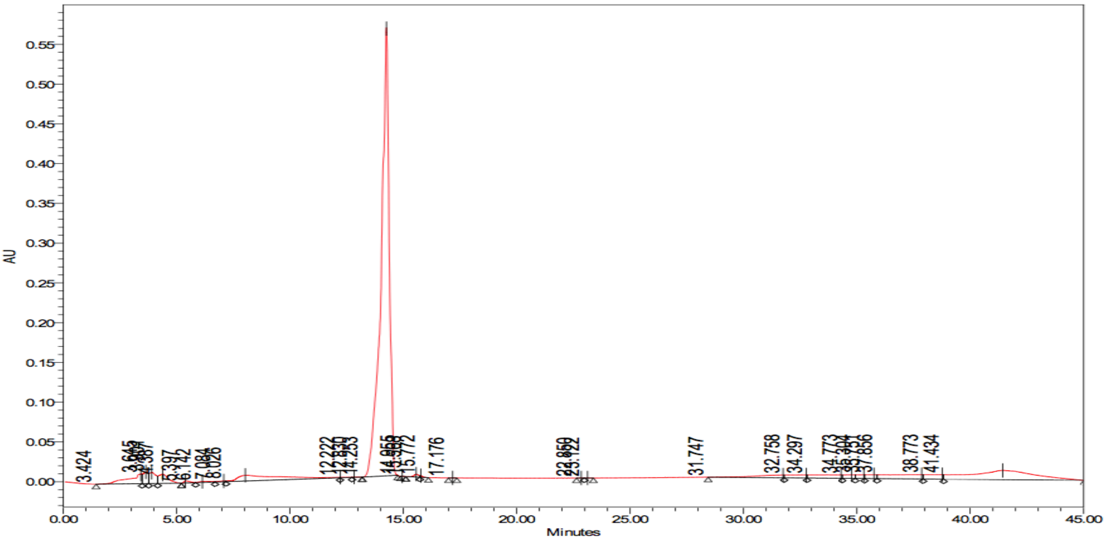
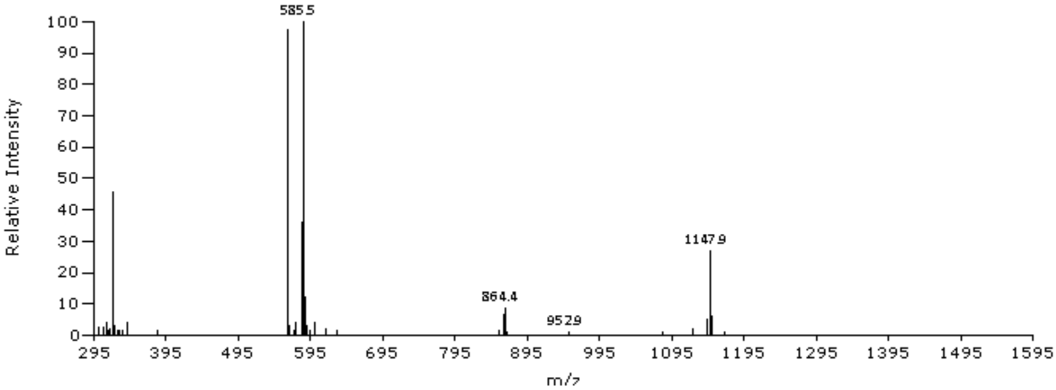
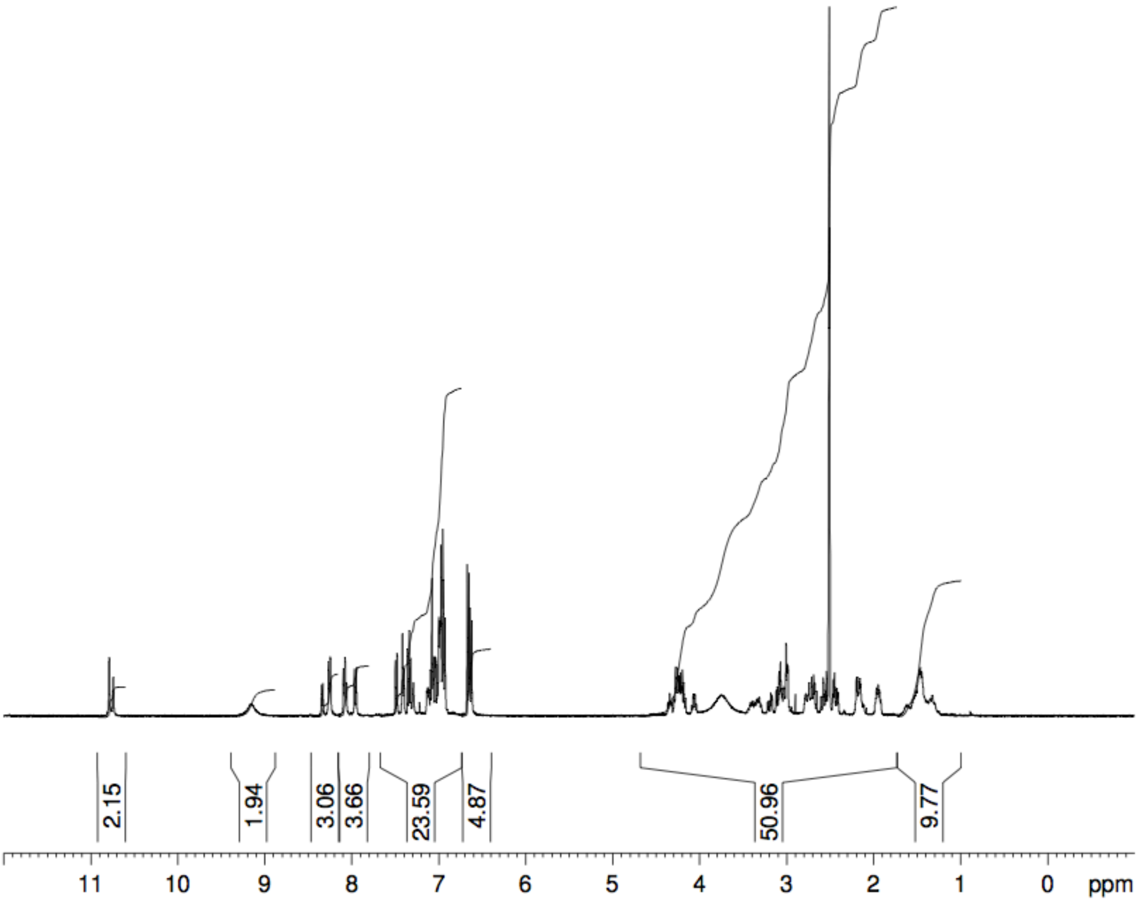


Figure 86: Cyclo-Ava-NWY



KL Cyclo AvaNWY, BLUE ESIPOS C4 10 min, RT 0.5072 mins, Scan# 146, NL 9.880E6, 1/12/2017 9:50 AM, mz [302.5-1169.2]





8.4.3.3 Cyclo GabaNWY

Fmoc-Gaba-OH was coupled onto the resin-bound NWY. It was deprotected as 8.2.1 and then cleaved from the resin to form linear GabaNWY (white solid, 123.2 mg crude). This peptide (123.2 mg, 0.217 mmol) was cyclised to form cyclo-GabaNWY, Figure 87, and purified via preparative HPLC yielding a white, shiny solid (26.3 mg, 0.048 mmol, 22% yield). LR MS (ESI+) *m/z* (%) 571.5 ($[M+Na]^+$, 100%), 549.6 ($[M+H]^+$, 20.9%). HR HPLC-MS (ESI+) for $C_{28}H_{33}N_6O_6$ $(M+H)^+$ Calculated mass 549.2449, found mass 549.2456 (1.3 ppm error). $C_{28}H_{32}N_6NaO_6$ $(M+Na)^+$ Calculated mass 571.2264, found mass 571.2276 (2.0 ppm error).

^{1}H NMR (400 MHz, DMSO-d_6) δ = 10.75 (1H, d, J = 27.3 Hz, Trp indole NH), 9.12 (1H, br s, Tyr phenol OH), 8.27 - 8.42 (1H, m, Trp backbone NH), 7.97 (1H, dd, J = 5.0 Hz, J = 8.2 Hz, Gaba backbone NH), 7.40 - 7.51 (1H, m, Asn backbone NH), 7.26 - 7.39 (3H, m, overlapped, Tyr backbone NH, Tyr D*-H), 7.15 (1H, d, J = 9.2 Hz, Asn NH_2), 7.07 (1H, q, J = 6.7 Hz, Asn NH_2), 6.93 - 7.04 (4H, m, overlapped, Trp D1, Z2, H2, Z3), 6.87 (1H, d, J = 8.3 Hz, Trp E3), 6.61 (2H, dd, J = 8.1 Hz, J = 5.5 Hz, Tyr E*-H), 4.29 - 4.41 (1H, m, Tyr α-H), 4.14 - 4.29 (2H, m, Trp α-H), 4.10 (1H, dt, J = 9.2 Hz, J = 4.8 Hz, Asn α-H), 3.36 (1H, br dd, J = 5.4 Hz, J = 3.9 Hz, Gaba α-H), 3.18 (1H, br dd, J = 3.8 Hz, J = 4.5 Hz, Gaba α-H), 2.88 - 3.04 (2H, m, Trp β-H), 2.59 - 2.77 (1H, m, Tyr β-H), 2.45 (1H, br t, J = 9.0 Hz, Gaba γ-H), 2.30 - 2.41 (2H, m, Asn β-H), 2.21 - 2.30 (1H, m, Gaba γ-H), 1.92 - 2.07 (1H, m, Gaba β-H), 1.80 - 1.92 (1H, m, Gaba β-H).

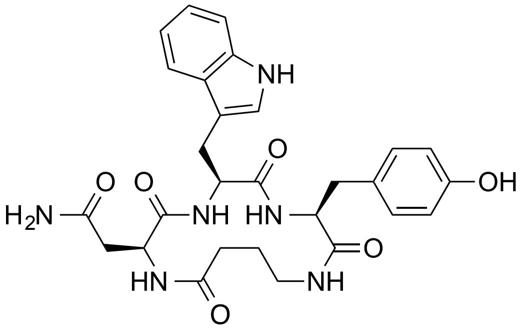
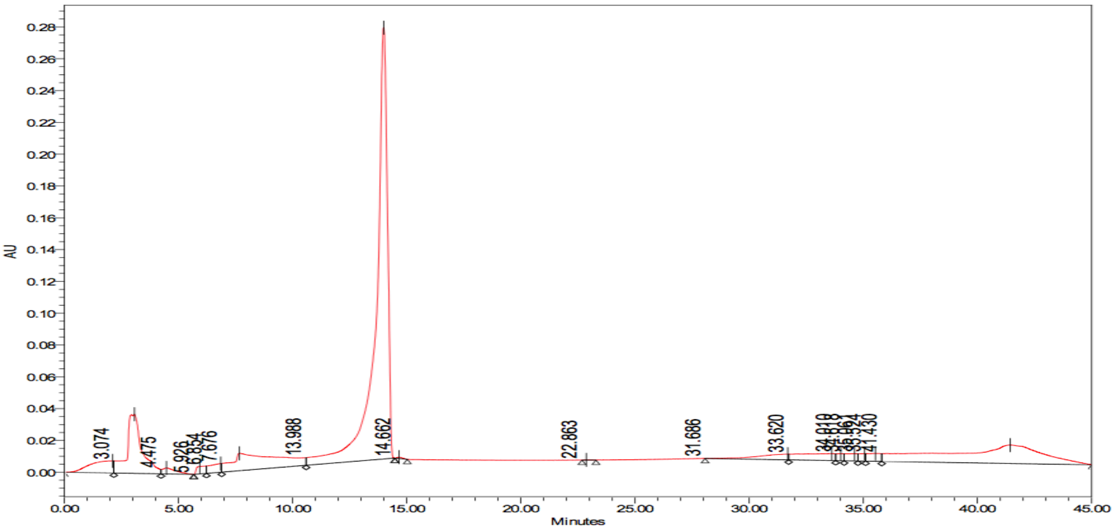
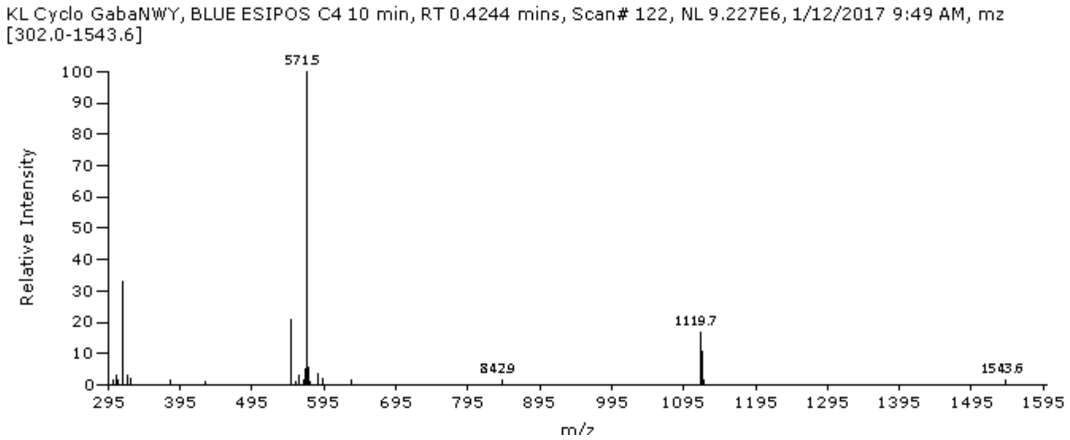
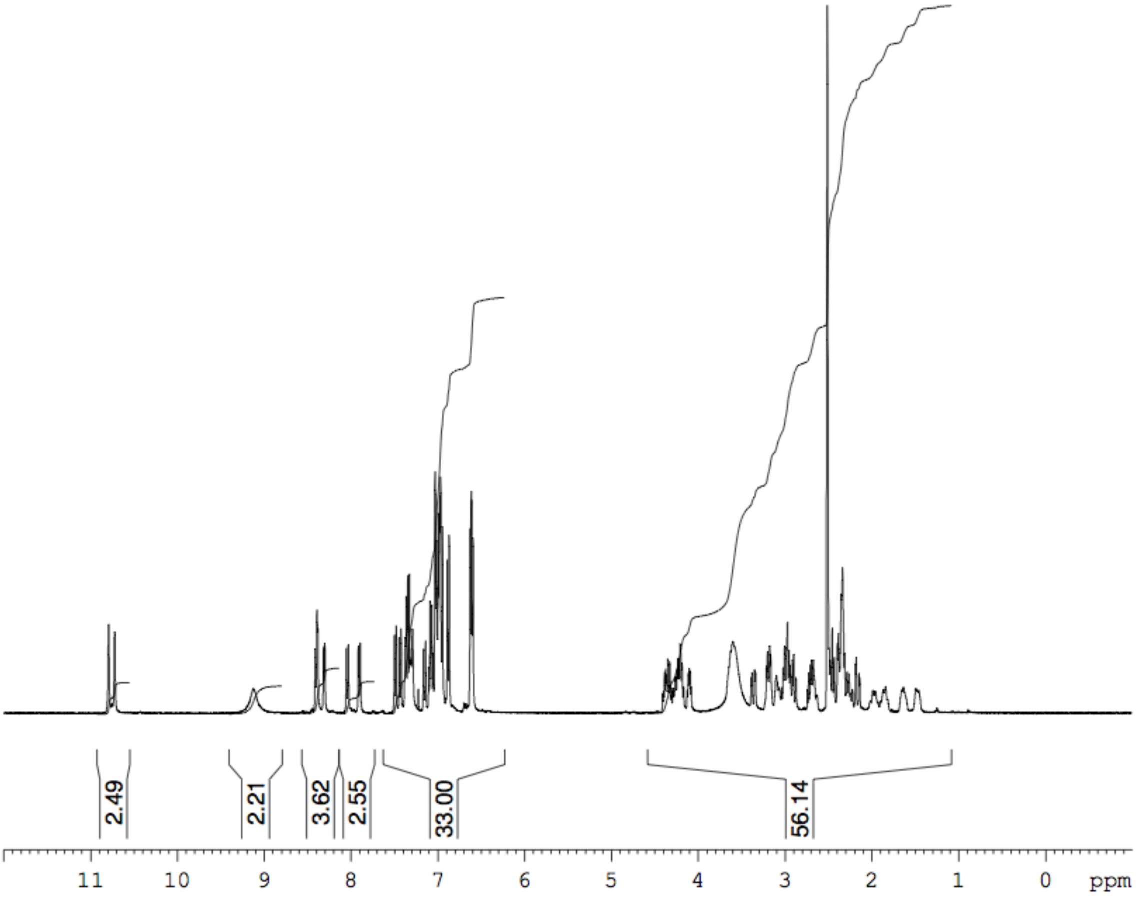


Figure 87: Cyclo-Gaba-NWY







8.5 Experimental for Chapter 4

8.5.1 Generation of pARCBD-CP11

8.5.1.1 PCR with pARCBD

The reagents in Table 24 were added in duplicate to PCR tubes. The forward and reverse primers can be seen in Table 25.

Table 24: Reagents for PCR with pARCBD to generate pARCBD-CP11

Reagents	Amount (μL)
5 × GoTaq Buffer	10
Forward Primer	1
Reverse Primer	1
dNTPs	1
Template	0.5
GoTaq Enzyme	0.25
dH_2O	31.25

Table 25: Table of Primers for Creation of pARCBD-CP11

Primer Name	Primer Sequence
CP11 Forward	GGAATTCGCCAATGGGGCGATCGCCCACAATAGCGGCTGGATCTACTGGAATGTGTGCTTAAGTTTTGGC
pARCBD Reverse	GGA ATT CAA GCT TTC ATT GAA GCT GCC ACA AGG

The PCR tubes were loaded onto the thermocycler which was set to the following program, Table 26.

Table 26: Thermocycler Program for Creation of pARCBD-CP11

Cycle Number	Denaturation	Annealing	Extension
1	95 °C for 2 mins	-	-
2-30	95 °C for 2 mins	55 °C for 30 secs	72 °C for 1 min
31	-	-	72 °C for 5 mins

The PCR products were analysed on a 1% agarose gel and then purified as 8.3.7.

8.5.1.2 Zipper Primer PCR

The purified PCR products from 8.5.1.1 were used as the templates in the zipper primer PCR. The reagents for the PCR were as Table 24. The primers used can be seen in Table 27. The thermocycler was set up as, Table 26.

Table 27: Primer Sequences for Zipper Primer PCR

Primer Name	Primer Sequence
Zipper Forward	GGA ATT CGC CAA TGG GGC GAT CGC C
pARCBD Reverse	GGA ATT CAA GCT TTC ATT GAA GCT GCC ACA AGG

The zipper primer PCR products were analysed on a 1% agarose gel and purified as 8.3.7.

8.5.1.3 Restriction Digestion, Ligation and Transformation into Chemically Competent Cells

The purified zipper primer PCR products, 8.5.1.2 were digested from the plasmid using the following reagents, Table 28. The reagents were incubated at 37 °C for 18 hours and then purified as 8.3.7 and combined.

Table 28: Reagents for Restriction Digestion of pARCBD-CP11

Reagents	Amount (µL)
Zipper Primer PCR Product	28.5
NEB Buffer 2	5
BglI Restriction Enzyme	2
HindIII Restriction Enzyme	1
dH_2O	13.5

The purified digestion products were mixed with the reagents as Table 29 and ligated into a digested pARCBD backbone in a 1:6 ratio by incubating at 4 °C for 18 hours. The ligation mix was then heated to 70 °C for 10 minutes.

Table 29: Reagents for Ligation into pARCBD

Reagents	Quantity (µL)	
	Vector Alone	1:6
Ligase buffer	1	1
T4 DNA ligase	1	1
Vector	2	2
Insert	-	0.7
Water	6	5.3

The ligation mix was transformed into DH5α chemically competent cells as 8.3.11. Colonies were subjected to Colony PCR as 8.3.12 and success was confirmed by sequencing the plasmids.

8.5.2 Site Directed Mutagenesis for Generation of Alanine Analogues

Site directed mutagenesis of pARCBD-CP11 was conducted to created alanine analogues of CP11. The reagents in Table 30 were added to PCR tubes and the thermocycler was programmed as Table 31. The template was pARCBD-CP11 and the primers are as Table 32.

Table 30: Reagents for Site Directed Mutagenesis

Reagents	Amount (µL)
10 × Pfu Buffer	5
Forward Primer	1.5
Reverse Primer	1.5
dNTPs	1
Template	0.5
Pfu Enzyme	1
dH_2O	39.5

Table 31: Thermocycler Program for Site Directed Mutagenesis

Cycle Number	Denaturation	Annealing	Extension
1	95 °C for 2 mins	-	-
2-30	95 °C for 30 sec	55 °C for 1 min	68 °C for 6 min
31	-	-	68 °C for 20 mins

Table 32: Sequences of Primers for Site Directed Mutagenesis

Primer Name	Primer Sequence
CP11A1 Forward	AATTGCTCGGGATGGGCGTACTGGAATGTGTGC
CP11A2 Forward	TGCTCGGGATGGATCGCGTGGAATGTGTGCTTA
CP11A3 Forward	TCGGGATGGATCTACGCGAATGTGTGCTTAAGT
CP11A4 Forward	CGGGATGGATCTACTGGGCGGTGTGCTTAAGTTTTGG
CP11A5 Forward	AGGATGGATCTACTGGAATGCGTGCTTAAGTTTTGGCACCG
CP11A7 Forward	GCCCACAATTGCTCGGCGTGGATCTACTGGAAT
CP11A8 Forward	CACAATTGCTCGGGAGCGATCTACTGGAATGTG

After the thermocycler program had finished, the restriction enzyme DpnI (1 µL) was added and the reactions were incubated at 37 °C for 3 hours and then purified as 8.3.7.

8.5.3 Drop-Spotting Alanine Analogues

Frozen stocks of DH5α *E. coli* were cultured in LB broth (10 mL) containing antibiotics (carbenicillin (100 µg/mL), spectinomycin (25 µg/mL) and chloramphenicol (35 µg/mL)) and incubated at 37°C overnight with agitation. The overnight cultures were diluted 10-fold to provide 6-point serial dilutions. The cultures were serially diluted in 10% glycerol (v/v, in sterile dH_2O; 20 µL culture diluted in 180 µL 10% glycerol) and then dropped (2.5 µL/drop) onto minimal media selection plates (minimal media supplemented with 6.5-13 µM ʟ-(+)-arabinose, 2.5 mM 3-AT, 25 µM kanamycin and 30 µM IPTG) and then incubated at 37°C for 2-3 days.

8.5.4 Enzyme-Linked Immunosorbent Assay

To each well in a clear, nickel coated 96-well plate (Thermo Scientific), purified His_6-UEV protein diluted in PBS (Thermo Scientific; 1 tablet dissolved in 100 mL deionized water; at a final concentration of 1 µM of His_6-UEV; 100 µL per well), was added and incubated for 1 h at room temperature with rocking. The wells were then washed with PBS containing 0.05% TWEEN 20 (Sigma; 3 × 200 µL, 5 min incubation per wash). The wells were blocked for 1 h with 2% milk solution in deionized water (150 µL per well) at room temperature with rocking. Without washing the wells, descending concentrations of peptide (serially diluted in 100% DMSO, 10 µL) with 0.5 µM of purified GST-tagged p6 protein in PBS (90 µL) were added and then incubated for 1 h at room temperature with rocking (10% final concentration of DMSO). The wells were then washed as before, and mouse anti-GST Ab-1 antibody (Neomarkers; 1:1000 made up in 2% milk in deionized water; 100 µL per well) was added and incubated for 1 h at room temperature with rocking. The wells were washed, then sheep anti-mouse IgG, peroxidase-linked whole antibody (GE Healthcare; 1:3000 made up in 2% milk in deionized water; 100 µL per well) was added and the plate incubated for 1 h at room temperature with rocking. The wells were washed and 1-Step™ Ultra TMB ELISA solution (Thermo Scientific; 100 µL) was added to each well and the plate incubated for 15 min at room temperature with rocking. To quench the reaction, 20 mM sulphuric acid (100 µL) was added to each well and the absorbance at 450 nm measured using a BMG Clariostar Plate reader.

8.5.5 CP11 Analogue 1 (A1)

CP11 Analogue 1 (CP11A1 = AYWNVSGW) was synthesised as 8.2.1, on a Glycine Wang resin and cleaved to form linear CP11A1 (LCP11A1; pink solid, 294.8 mg crude). This peptide (114.7 mg, 0.116 mmol) was cyclised (CP11A1), Figure 88: CP11A1, and purified via preparative HPLC yielding a pale orange solid (49.6 mg, 0.0051 mmol, 44% yield). MS (ESI+) *m/z* (%) 986.4 ($[M+Na]^+$, 100%), 482.9 (½$[M+2H]^{2+}$, 90.2%), 964.0 ($[M+H]^+$, 69%). Analytical HPLC (280 nm) 20.8 min (96%).

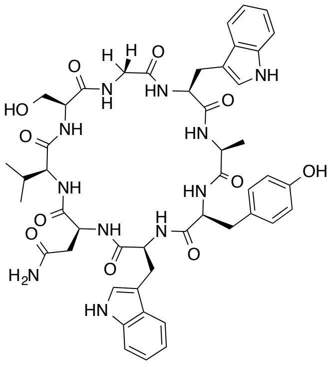
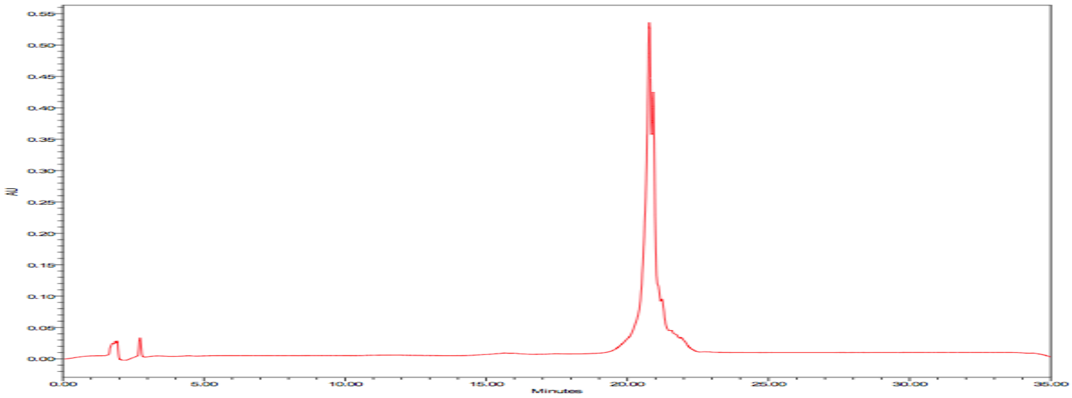
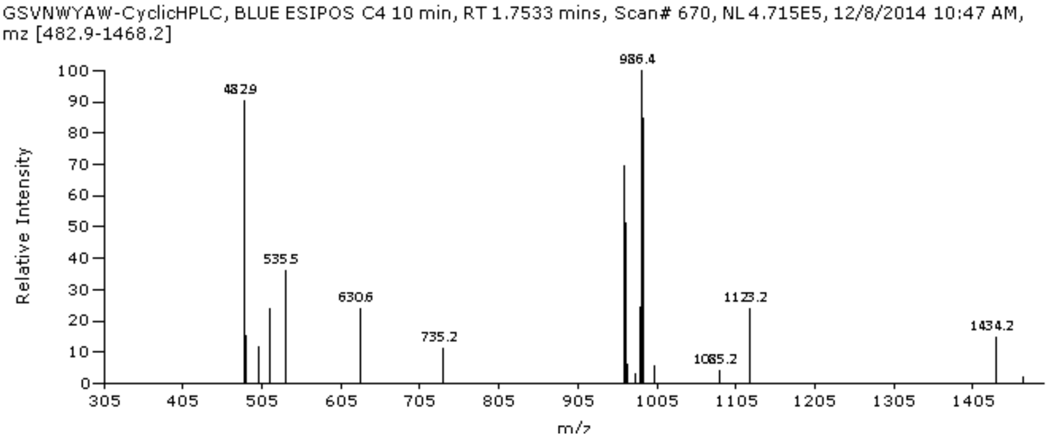


Figure 88: CP11A1





8.5.6 CP11A2

CP11Analogue 2 (CP11A2 = IAWNVSGW) was synthesised as 8.2.1, on a Glycine Wang resin and cleaved to form linear CP11A2 (LCP11A2; pale yellow solid, 144 mg crude). This peptide (94.1 mg, 0.101 mmol) was cyclised (CP11A2), Figure 89, and purified via preparative HPLC yielding a creamy, white solid (16.1 mg, 0.018 mmol, 17% yield). MS (ESI+) *m/z* (%) 914.5 ($[M+H]^+$, 100%), 457.8 (½$[M+2H]^{2+}$, 49%), 936.6 ($[M+Na]^+$, 17%). Analytical HPLC (280 nm) 20.9 min (96%).

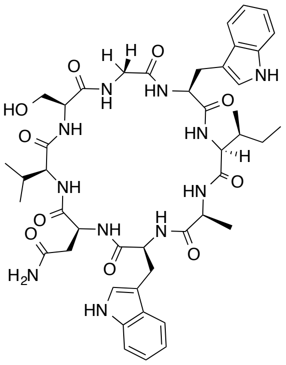
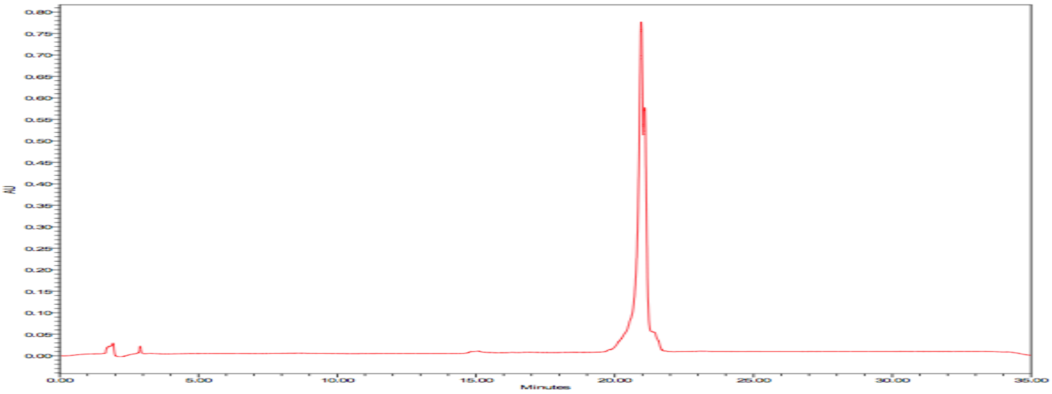
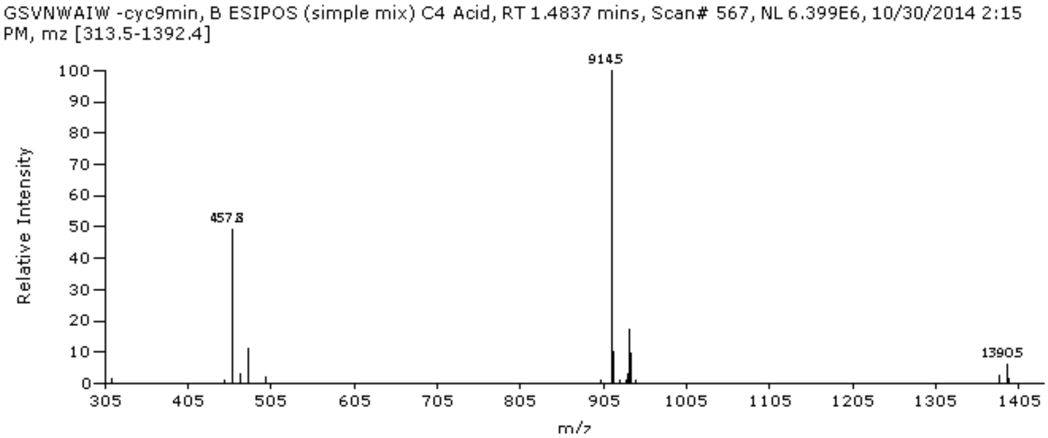


Figure 89: CP11A2





8.5.7 CP11A3

CP11Analogue 3 (CP11A3 = IYANVSGW) was synthesised as 8.2.1, on a Glycine Wang resin and cleaved to form linear CP11A3 (LCP11A3; cream solid, 156 mg crude). This peptide (98.7 mg, 0.109 mmol) was cyclised (CP11A3), Figure 90, and purified via preparative HPLC yielding a creamy, white solid (7.8 mg, 0.009 mmol, 8% yield). MS (ESI+) *m/z* (%) 913.4 ($[M+H+H_2O]^+$, 100%), 891.5 ($[M+H]^+$, 80%), 465.4 (½$[M+2H]^{2+}$, 64%). Analytical HPLC (280 nm) 19.8 min (94%).

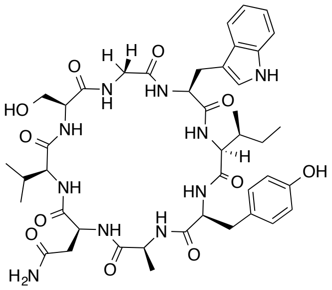
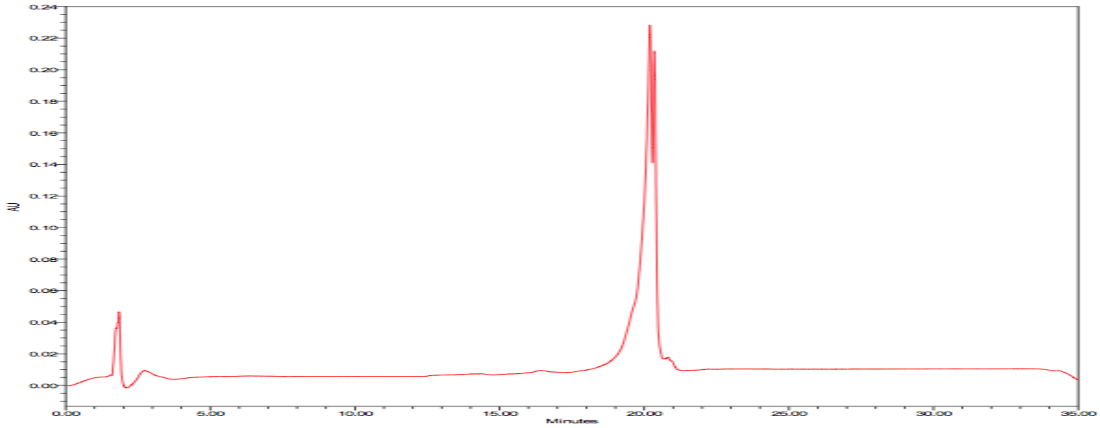
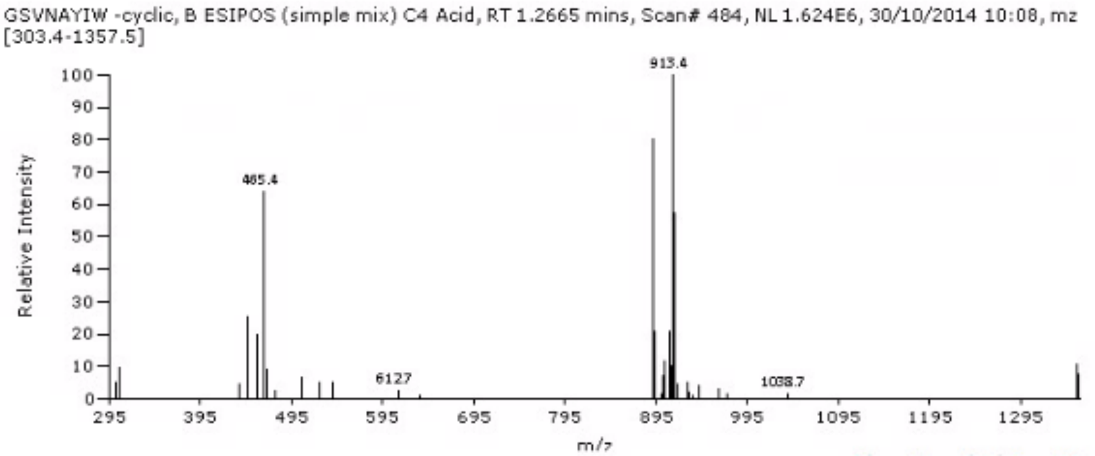


Figure 90: CP11A3





8.5.8 CP11A4

CP11 Analogue 4 (CP11A4 = IYWAVSGW) was synthesised as 8.2.1, on a Glycine Wang resin and cleaved to form linear CP11A4 (LCP11A4; cream solid, 126 mg crude). This peptide (82 mg, 0.084 mmol) was cyclised (CP11A4), Figure 91, and purified via preparative HPLC yielding a pale orange solid (8.7 mg, 0.009 mmol, 10% yield). MS (ESI+) *m/z* (%) 501.3 (½$[M+2H]^{2+}$, 100%), 964.6 ($[M+H]^{+}$, 25%), 985.3 ($[M+H+H_2O]^{+}$, 30%). Analytical HPLC (280 nm) 22.6 min (75%).

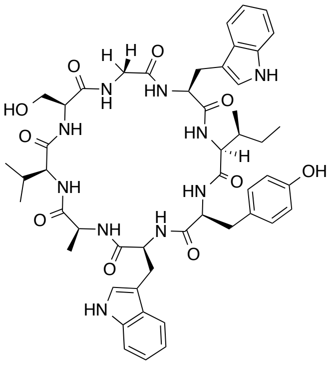
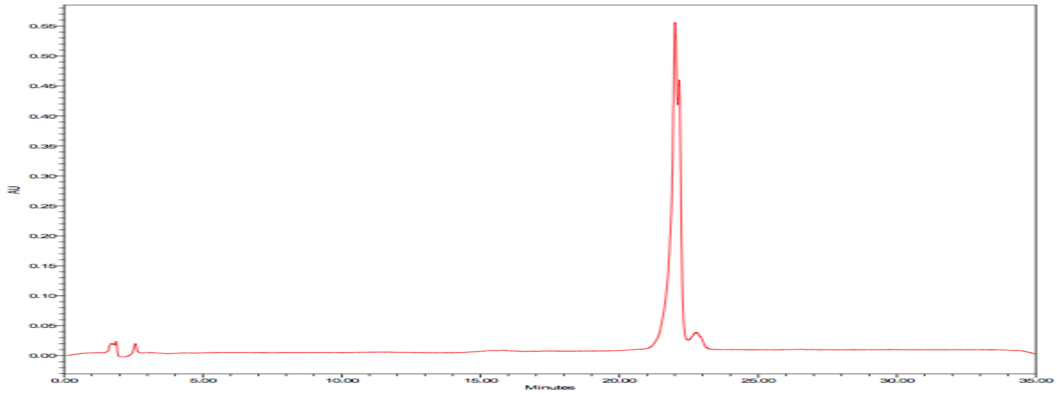
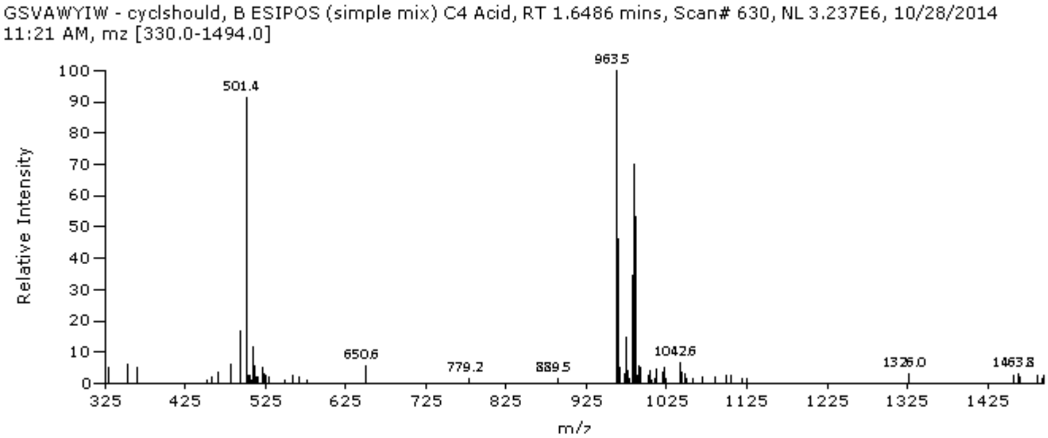


Figure 91: CP11A4





8.5.9 CP11A5

CP11 Analogue 5 (CP11A5 = IYWNASGW) was synthesised as 8.2.1, on a Glycine Wang resin and cleaved to form linear CP11A5 (LCP11A5; pale orange, 138 mg crude). This peptide (98 mg, 0.098 mmol) was cyclised (CP11A5), Figure 92, and purified via preparative HPLC yielding a pale orange solid (41.6 mg, 0.043 mmol, 41% yield). MS (ESI+) *m/z* (%) 508.5 (½$[M+K]^{2+}$, 100%), 999.9 ($[M+Na]^{+}$, 40%). Analytical HPLC (280 nm) 21.8 min (96%).

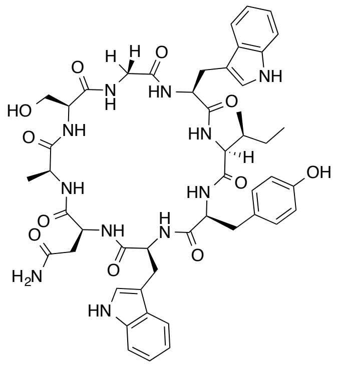
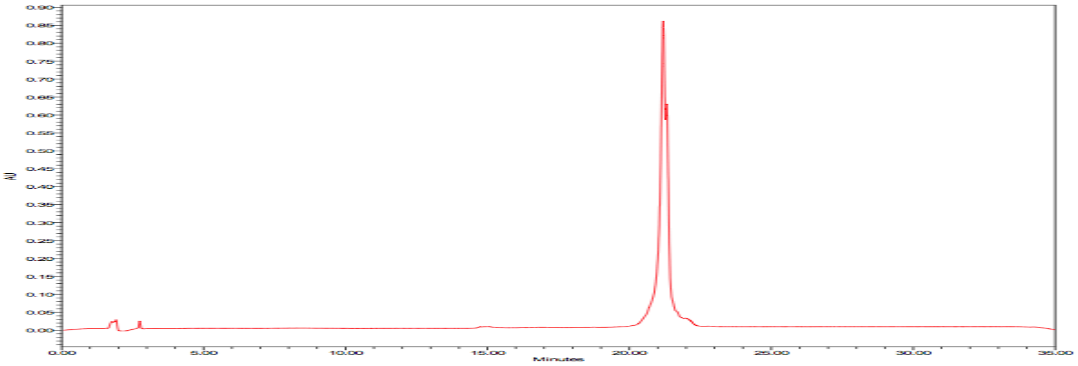
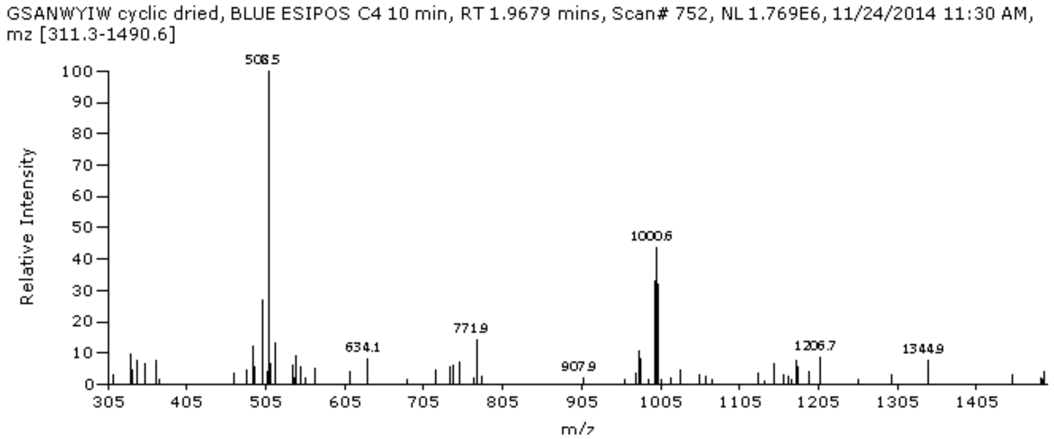


Figure 92: CP11A5





8.5.10 CP11A6

CP11Analogue 6 (CP11A6 = IYWNVAGW) was synthesised as 8.2.1, on a Glycine Wang resin and cleaved to form linear CP11A6 (LCP11A6; pale orange, 266.7 mg crude). This peptide (112.6 mg, 0.112 mmol) was cyclised (CP11A6), Figure 93, and purified via preparative HPLC yielding an orange solid (41.9 mg, 0.042 mmol, 37% yield). MS (ESI+) *m/z* (%) 1013.2 ($[M+Na]^+$, 100%), 495.7 (½$[M+2H]^{2+}$, 48%), 990.0 ($[M]^+$, 45%), 514.1 (½$[M+K]^{2+}$, 37%). Analytical HPLC (280 nm) 21.6min (97%).

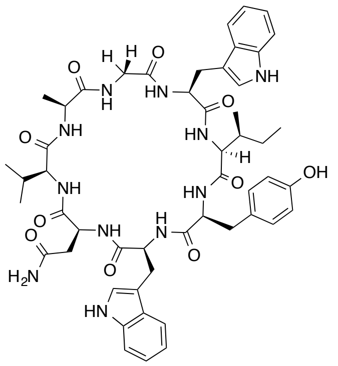
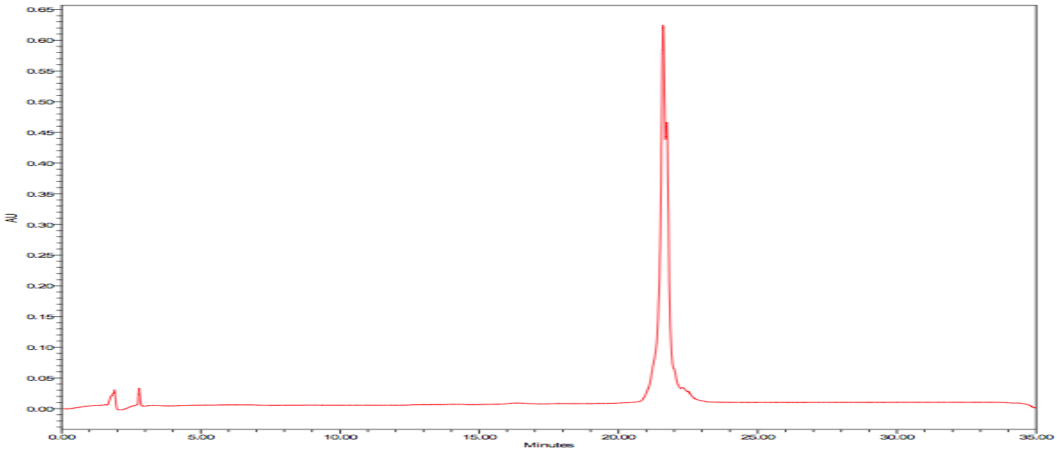
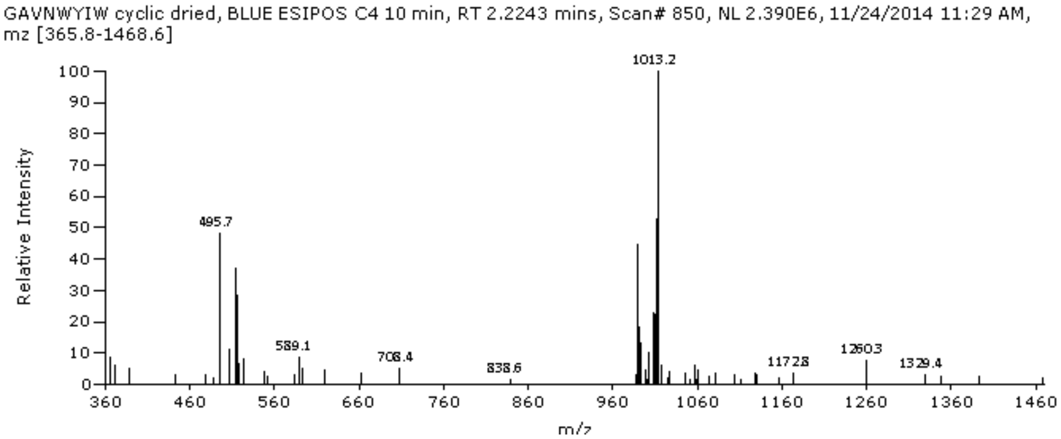


Figure 93: CP11A6





8.5.11 CP11A7

CP11Analogue 7 (CP11A7 = IYWNVSAW) was synthesised as 8.2.1, on a Valine Wang resin and cleaved to form linear CP11A7 (LCP11A7; pale pink solid, 657.8 mg crude). This peptide (130.3 mg, 0.126 mmol) was cyclised (CP11A7), Figure 94, and purified via preparative HPLC yielding a pale orange solid (41.6 mg, 0.041 mmol, 32% yield). MS (ESI+) *m/z* (%) 1042 ($[M+Na]^+$, 100%), 529.7 ($½[M+H+H_2O]^{2+}$, 93%), 510.7 ($½[M+2H]^{2+}$, 64%), 1020.1 ($[M]^+$, 44%). Analytical HPLC (280 nm) 22.4 min (97%).

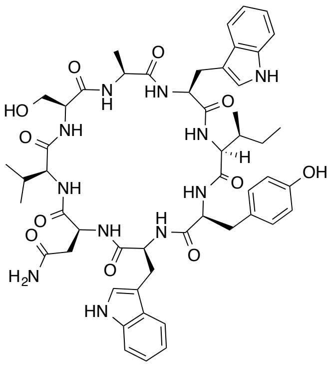
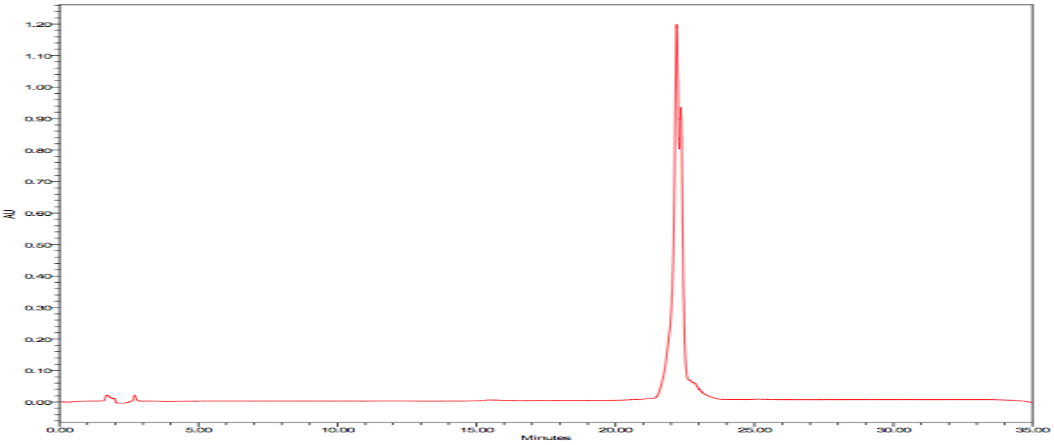
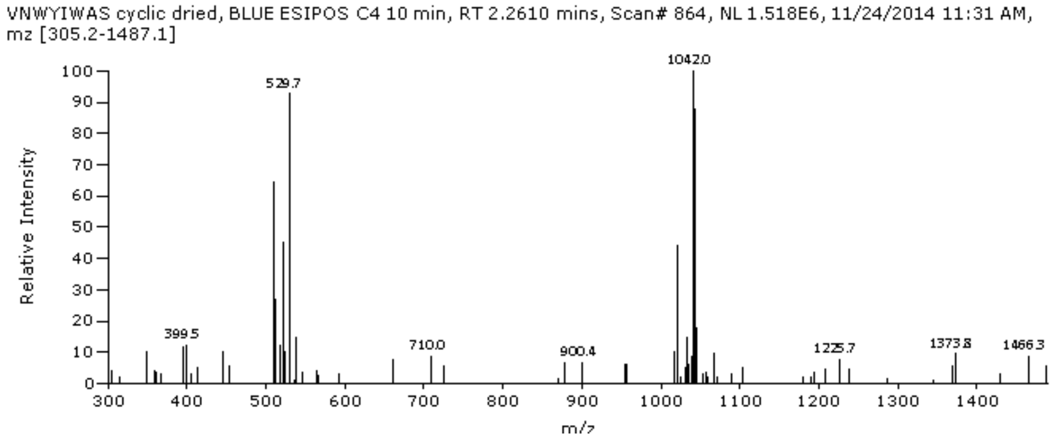


Figure 94: CP11A7





8.5.12 CP11A8

CP11Analogue 8 (CP11A8 = IYWNVSGA) was synthesised as 8.2.1, on a Glycine Wang resin and cleaved to form linear CP11A8 (LCP11A8; pale orange solid, 266.7 mg crude). This peptide (117.5 mg, 0.13 mmol) was cyclised (CP11A8), Figure 95, and purified via preparative HPLC yielding a pale pink solid (50.6 mg, 0.057 mmol, 43% yield). LR MS (ESI+) *m/z* (%) 913.2 ([M+Na]$^+$, 100%), 891.3 ([M+H]$^+$, 85.1%). Analytical HPLC (280 nm) 19.4 min (79%).

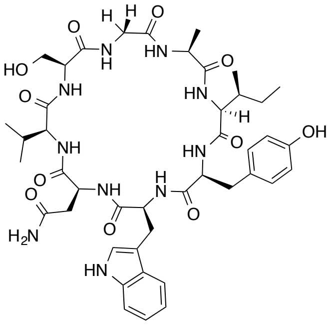
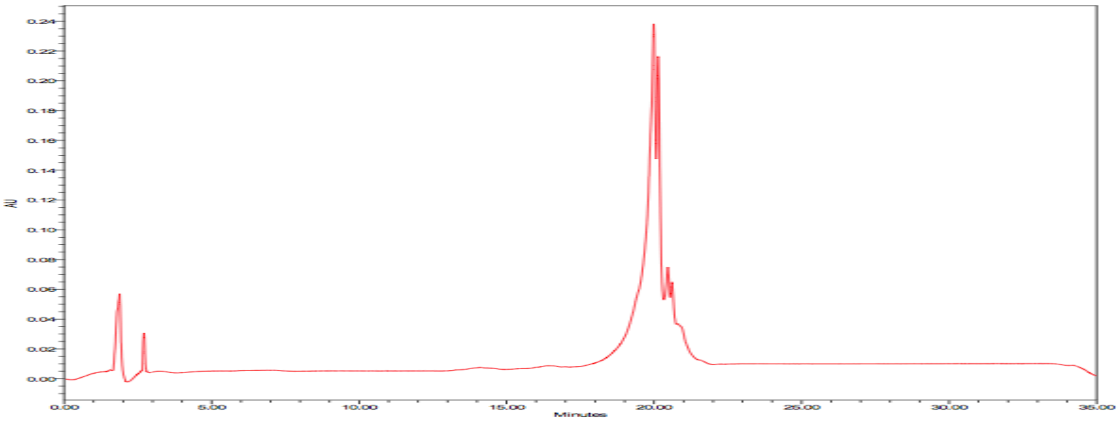
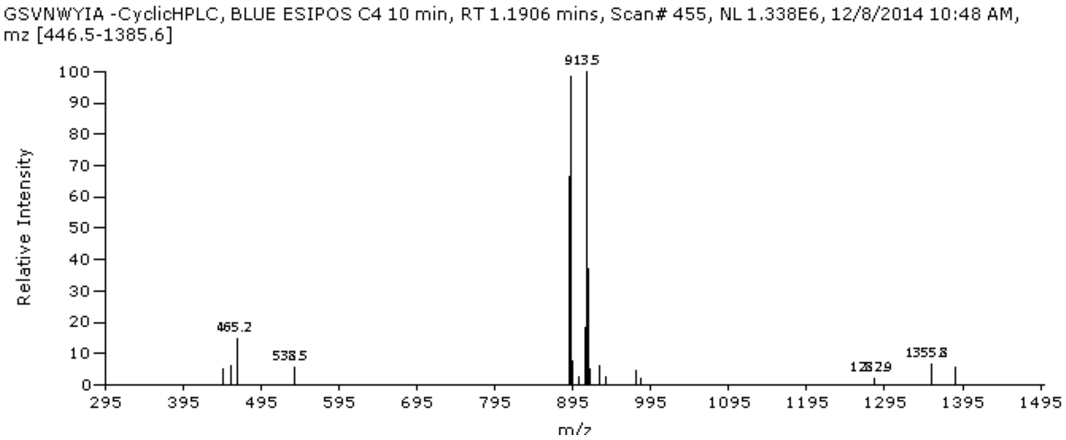


Figure 95: CP11A8





8.5.13 Synthesis of Tri- and Di-Peptide Motifs

8.5.13.1 Synthesis of Acetyl-IYW-Diethylamine

8.5.13.1.1 Synthesis of Fmoc-Trp(Boc)-Diethylamine

Fmoc-Trp(Boc)-OH (1.75 g, 3.22 mmol) was dissolved in DCM (15 mL) with one equivalent each of HOBt (0.44, 3.22 mmol) and EDC (0.62 g, 3.2 mmol). Diethylamine (0.7 mL, 6.5 mmol) was added dropwise and the reaction left stirring under argon for 18 hours at room temperature, Figure 96. The reaction was stopped and then diluted with DCM (20 mL). The organic layer was washed with sodium bicarbonate solution (10% w/v, 20 mL), citric acid (20 mL) and brine (20 mL). It was then dried over anhydrous sodium sulphate. The solution was filtered and the solvent removed by rotary evaporation. The crude yellow oil was purified via column chromatography (9:1, DCM:MeOH) yielding a pale yellow oil (1.3194 g, 2.27 mmol, 71% yield, R_f 0.8).

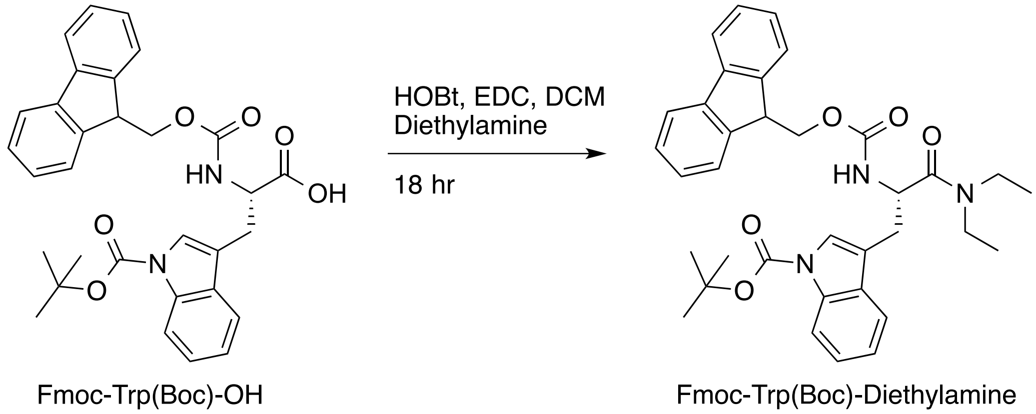


Figure 96: Mechanism of the Synthesis of Fmoc-Trp(Boc)-Diethylamine.

8.5.13.1.2 Synthesis of H_2N-Trp(Boc)-Diethylamine

Fmoc-Trp(Boc)-diethylamine (1.32 g, 2.27 mmol) was dissolved in ethyl acetate (20 mL). 1,8-Diazabicyclo[5.4.0]undec-7-ene (DBU; 0.366 mL, 2.45 mmol) was added dropwise and the reaction left stirring under argon at room temperature for 3 hours, Figure 97. The solvent was removed by rotary evaporation and the crude oil was purified by column chromatography over a solvent gradient (0-5% MeOH in DCM). Following on from purification a yellow oil was obtained (0.67 g, 1.87 mmol, 76% yield, R_f 0.08).

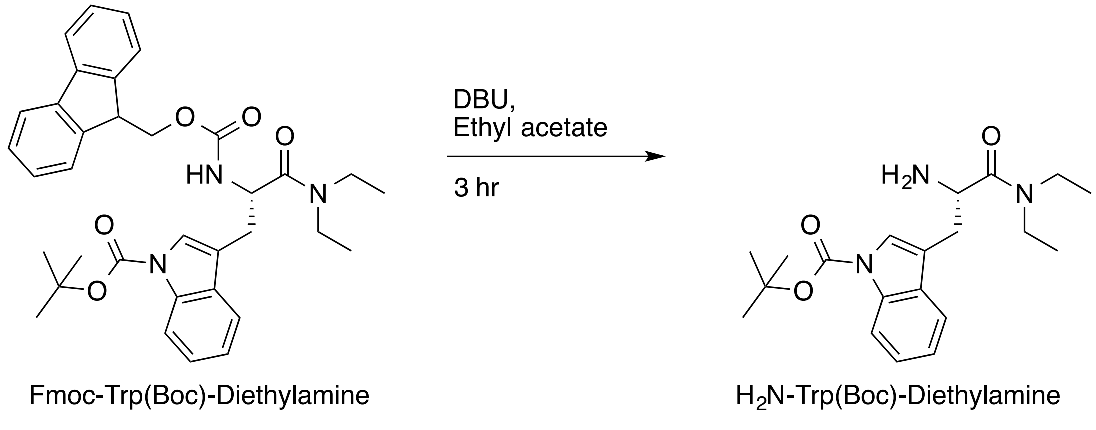


Figure 97: Mechanism of the Synthesis of H_2N-Trp(Boc)-Diethylamine

8.5.13.1.3 Synthesis of Fmoc-Tyr(tBu)-Trp(Boc)-Diethylamine

H_2N-Trp(Boc)-diethylamine (0.67 g, 1.86 mmol) was dissolved in DCM (15 mL) with two equivalents of Fmoc-Tyr(tBu)-OH (1.7 g, 3.7 mmol), EDC (0.5 g, 3.7 mmol) and HOBt (0.7 g, 3.7 mmol). The reaction was left to stir for 18 hours, under argon at room temperature, Figure 98.

The organic layer was washed with sodium bicarbonate solution (10% w/v, 20 mL), citric acid (20 mL) and brine (20 mL). It was then dried over anhydrous sodium sulphate. The solution was filtered and the solvent removed by rotary evaporation. The crude yellow oil was purified via column chromatography (9:1, DCM:MeOH) yielding a white solid (1.22 g, 1.52 mmol, 81% yield, R_f 0.79).

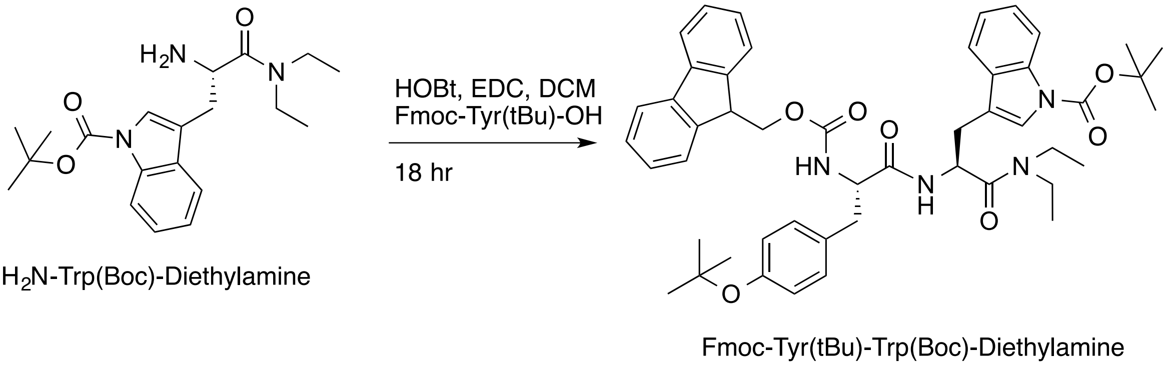


Figure 98: Mechanism of the Synthesis of Fmoc-Tyr(tBu)-Trp(Boc)-Diethylamine

8.5.13.1.4 Synthesis of H_2N-Tyr(tBu)-Trp(Boc)-Diethylamine

Fmoc-Tyr(tBu)-Trp(Boc)-diethylamine (1.22 g, 1.52 mmol) was dissolved in ethyl acetate (15 mL). DBU (0.224 mL, 1.5 mmol) was added dropwise and the reaction left stirring under argon at room temperature for 3 hours, Figure 99. The solvent was removed by rotary evaporation and the crude oil was purified by column chromatography over a solvent gradient (0-5% MeOH in DCM). Following on from purification a yellow oil was obtained (0.72 g, 1.25 mmol, 81% yield, R_f 0.37).

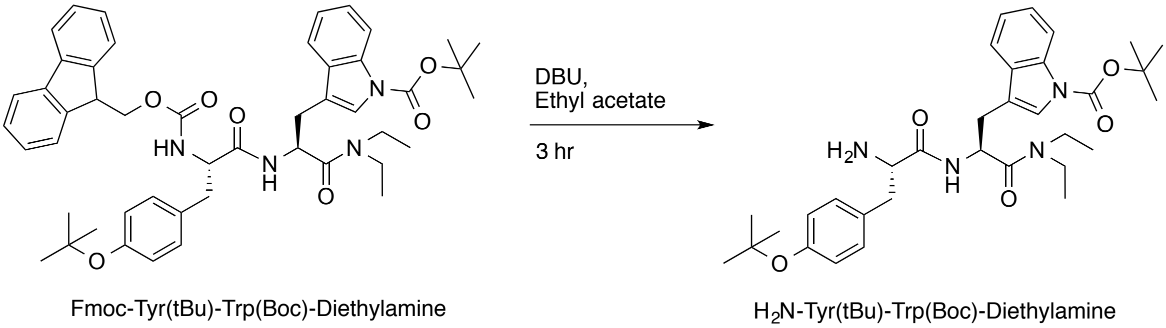


Figure 99: Mechanism of the Synthesis of H_2N-Tyr(tBu)-Trp(Boc)-Diethylamine

8.5.13.1.5 Synthesis of Fmoc-Ile-Tyr(tBu)-Trp(Boc)-Diethylamine

H_2N-Tyr(tBu)-Trp(Boc)-diethylamine (0.72 g, 1.25 mmol) was dissolved in DCM (15 mL) with two equivalents of Fmoc-Ile-OH (0.9 g, 2.5 mmol) , EDC (0.4 g, 2.5 mmol) and HOBt (0.5 g, 2.5 mmol). The reaction was left to stir for 18 hours, under argon at room temperature, Figure 100. The organic layer was washed with saturated sodium bicarbonate solution (20 mL), citric acid (20 mL) and brine (20 mL). It was then dried over anhydrous sodium sulphate. The solution was filtered and the solvent removed by rotary evaporation. The crude yellow oil was purified via column chromatography (9:1, DCM:MeOH) yielding a yellow oil (1.07 g, 1.17 mmol, 93% yield, R_f 0.68).

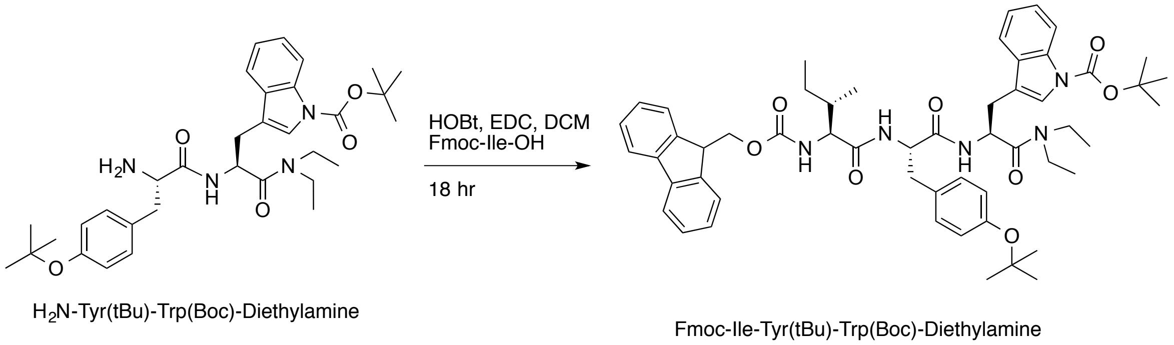


Figure 100: Mechanism of the Synthesis of Fmoc-Ile-Tyr(tBu)-Trp(Boc)-Diethylamine

8.5.13.1.6 Synthesis of H_2N-Ile-Tyr(tBu)-Trp(Boc)-Diethylamine

Fmoc-Ile-Tyr(tBu)-Trp(Boc)-diethylamine (1.07 g, 1.17 mmol) was dissolved in ethyl acetate (15 mL). DBU (0.175 mL, 1.2 mmol) was added dropwise and the reaction left stirring under argon at room temperature for 3 hours, Figure 101. The solvent was removed by rotary evaporation and the crude oil was purified by column chromatography over a solvent gradient (5% MeOH in DCM). Following on from purification a yellow oil was obtained (0.46 g, 0.67 mmol, 57% yield, R_f 0.16).

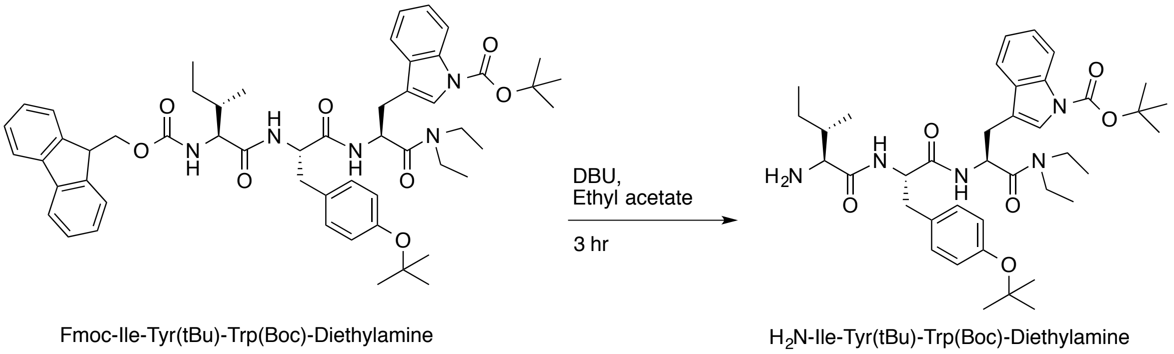


Figure 101: Mechanism of the Synthesis of H_2N-Ile-Tyr(tBu)-Trp(Boc)-Diethylamine

8.5.13.1.7 Synthesis of Acetyl-Ile-Tyr(tBu)-Trp(Boc)-Diethylamine

H_2N-Ile-Tyr(tBu)-Trp(Boc)-diethylamine (0.46 g, 0.67 mmol) was dissolved in DCM (10 mL). One equivalent each of triethylamine (0.1 mL, 0.7 mmol) and acetic anhydride (0.1, 0.7 mmol) were added dropwise and the reaction stirred under argon for 3 hours at room temperature, Figure 102. The solvent was removed by rotary evaporation and the crude oil was purified by column chromatography (5% MeOH in DCM) yielding a yellow oil (0.74 g, 1.02 mmol, crude, R_f 0.26).

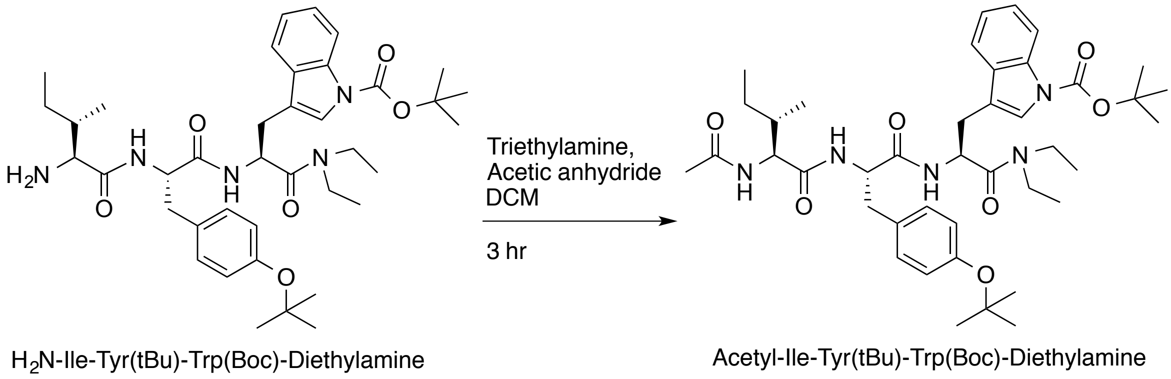
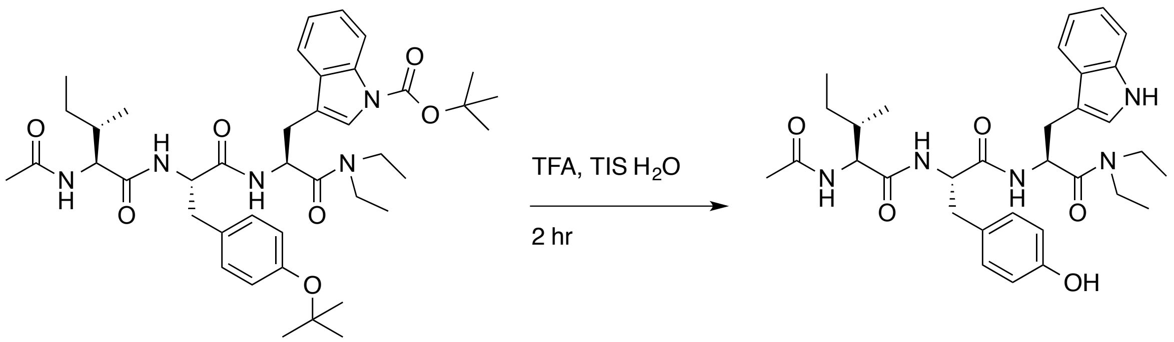


Figure 102: Mechanism of the Synthesis of Acetyl-Ile-Tyr(tBu)-Trp(Boc)-Diethylamine

8.5.13.1.8 Synthesis of Acetyl-Ile-Tyr-Trp-Diethylamine

Acetyl-Ile-Tyr(tBu)-Trp(Boc)-diethylamine (0.74 g, 1.02 mmol) was dissolved in a cleavage cocktail of TFA, TIS and water (9.5:0.25:0.25, 10 mL) and stirred for 2 hours, Figure 103. The solvent was removed by rotary evaporation, and the peptide was washed with DCM (3 × 10 mL) and then precipitated in ice cold diethyl ether (10 mL). The peptide was filtered and allowed to dry yielding a white power (0.43 g, 0.74 mmol, 73% yield). The peptide was purified by RP-HPLC and lyophilised to yield a white powder (0.263 g, 0.46 mmol, 61.6%). LR MS (ESI+) *m/z* (%) 578.5 ($[M+Na]^+$, 100%), 600.5 ($[M+H]^+$, 37.5%). HR HPLC-MS (ESI+) for $C_{32}H_{43}N_5O_5$ $(M+H)^+$ Calculated mass 578.3337, found mass 578.3335 (0.4 ppm error). Analytical HPLC (280 nm) 18.94 min (97%).

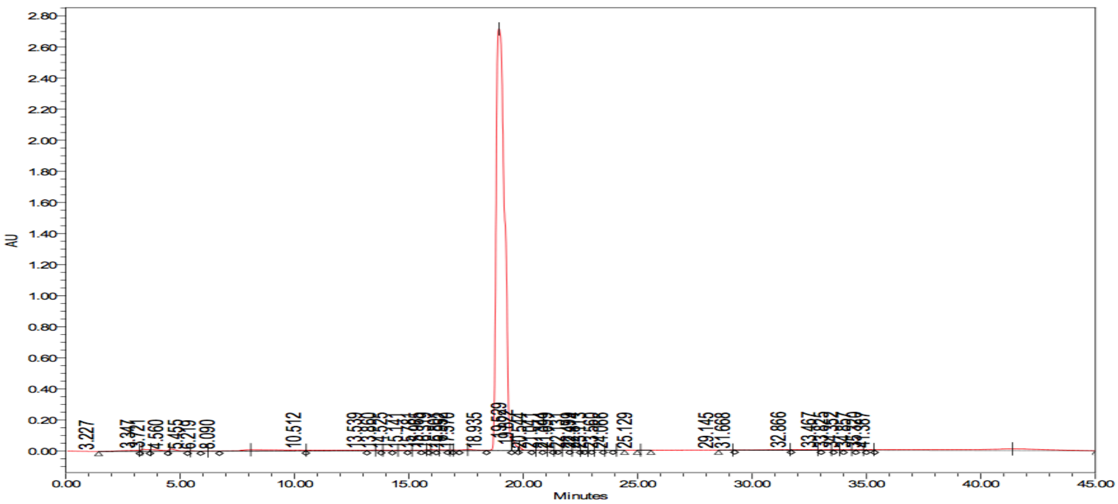
^{1}H NMR (400 MHz, DMSO-d_6) δ = 10.81 (1H, d, J = 1.7 Hz, Trp indole NH), 8.18 (1H, d, J3,7 = 8.3 Hz, Trp backbone NH), 7.89 (1H, br d, J = 8.3 Hz, Tyr backbone NH), 7.86 (1H, d, J = 8.9 Hz, H-6), 7.58 (1H, d, J = 7.7 Hz, Ile backbone NH), 7.32 (1H, d, J = 7.9 Hz, Trp E3-H), 7.07 (1H, d, J = 8.7 Hz, Trp Z2-H), 7.06 (1H, t, J = 7.5 Hz, Trp D1-H), 7.00 (1H, d, J = 7.5 Hz, Trp H2-H), 6.95 (2H, d, J = 8.6 Hz, Tyr D*-H), 6.59 (2H, d, J = 8.4 Hz, Tyr E*-H), 4.90 (1H, td, J = 8.3 Hz, J = 5.9 Hz, Trp α-H), 4.49 (1H, td, J = 8.4 Hz, J = 5.4 Hz, Tyr α-H), 4.15 (1H, t, J7,3 = 8.4 Hz, Ile α-H), 3.19 (1H, dt, J = 3.5 Hz, J = 6.9 Hz, Diethylamine CH_2-H), 3.12 (1H, dd, J = 3.8 Hz, J = 2.7 Hz, Diethylamine CH_2-H), 3.10 (2H, br d, J = 3.0 Hz, Diethylamine CH_2-H), 3.00 (1H, dt, J = 4.6 Hz, J = 7.2 Hz, Diethylamine CH_2-H), 2.88 - 2.95 (1H, m, Tyr β-H), 2.83 (1H, br d, J = 5.0 Hz, Trp β-H), 2.71 (1H, br dd, J = 13.9 Hz, J16,7 = 8.5 Hz, Trp β-H), 1.85 (3H, s, Acetyl CH_3), 1.68 (1H, td, J = 6.8 Hz, J = 6.0 Hz, Ile β-H), 1.08 (1H, quin, J = 7.3 Hz, Ile G1-H), 0.88 (3H, t, J = 7.0 Hz, Acetyl CH_3), 0.76 - 0.81 (3H, m, Acetyl CH_3), 0.71 - 0.76 (3H, m, Ile D-H), 0.76 (3H, dq, J = 3.7 Hz, J = 7.1 Hz, Ile G2-H).

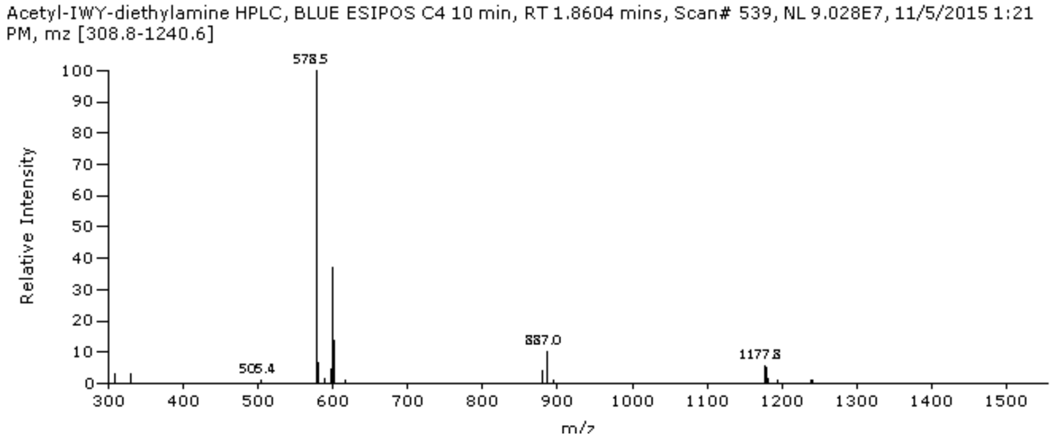


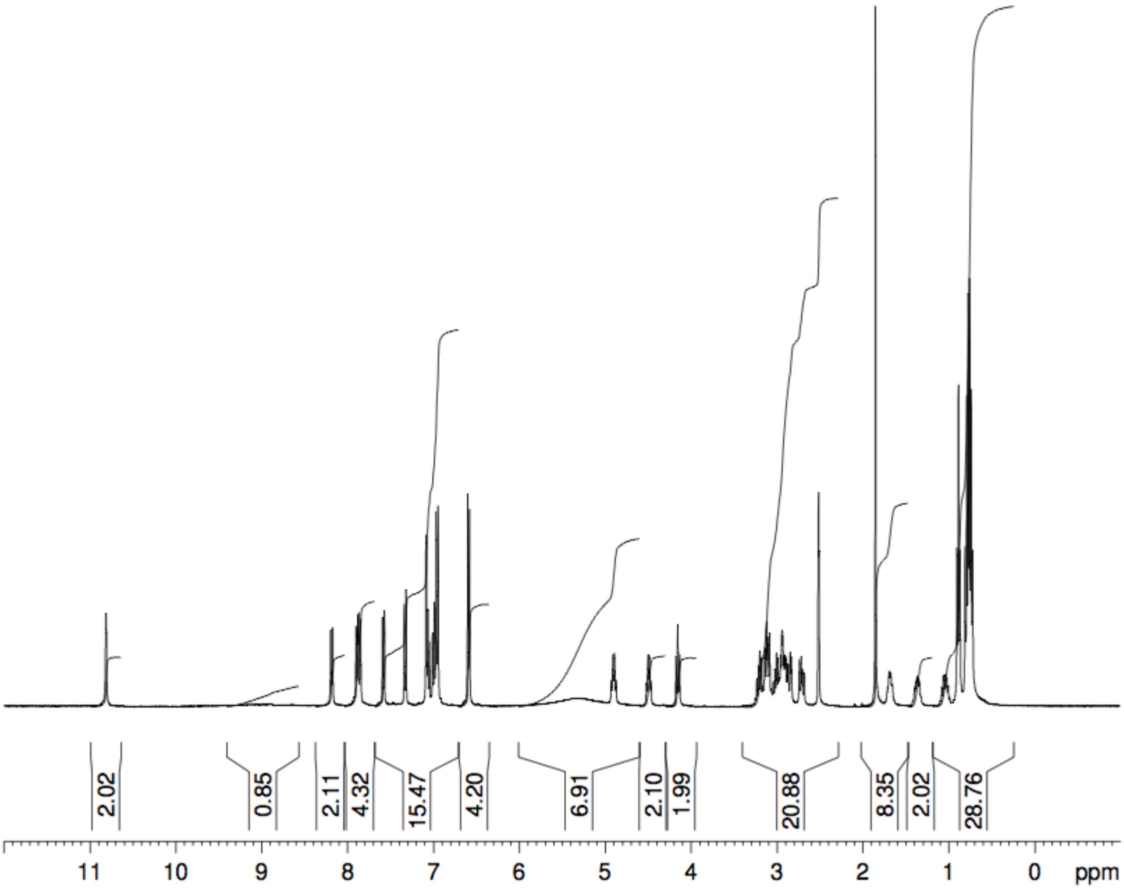
Acetyl-Ile-Tyr(tBu)-Trp(Boc)-Diethylamine

Acetyl-Ile-Tyr-Trp-Diethylamine

Figure 103: Mechanism of the Synthesis of Acetyl-Ile-Tyr-Trp-Diethylamine







8.5.13.2 Synthesis of Acetyl-IY-Diethylamine

8.5.13.2.1 Synthesis of Fmoc-Tyr(tBu)-Diethylamine

Fmoc-Tyr(tBu)-OH (1.53 g, 3.2 mmol) was dissolved in DCM (15 mL) with one equivalent each of HOBt (0.4 g, 3.2 mmol) and EDC (0.6 g, 3.2 mmol). Diethylamine (0.7 mL, 6.5 mmol) was added dropwise and the reaction left stirring under argon for 4 hours at room temperature, Figure 69 .The reaction was stopped and then diluted with DCM (20 mL). The organic layer was washed with saturated sodium bicarbonate solution (20 mL), citric acid (20 mL) and brine (20 mL). It was then dried over anhydrous sodium sulphate. The solution was filtered and the solvent removed by rotary evaporation. The crude yellow oil was purified via column chromatography (5% MeOH in DCM) yielding a white solid (1.40 g, 2.716 mmol, 84% yield, R_f 0.76).

8.5.13.2.2 Synthesis of H_2N-Tyr(tBu)-Diethylamine

Fmoc-Tyr(tBu)-diethylamine (1.40 g, 2.716 mmol) was dissolved in ethyl acetate (15 mL). DBU (0.41 mL, 3.0 mmol) was added dropwise and the reaction left stirring under argon at room temperature for 3 hours. The solvent was removed by rotary evaporation and the crude oil was purified by column chromatography over a solvent gradient (0-5% MeOH in DCM). Following on from purification a yellow oil was obtained (0.60 g, 2.05 mmol, 75% yield, R_f 0.31).

8.5.13.2.3 Synthesis of Fmoc-Ile-Tyr(tBu)-Diethylamine

H_2N-Tyr(tBu)-diethylamine (0.60 g, 2.05 mmol) was dissolved in DCM (15 mL) with two equivalents of Fmoc-Ile-OH (1.5 g , 4.1 mmol), EDC (0.6 g , 4.1 mmol) and HOBt (0.8 g , 4.1 mmol). The reaction was left to stir for 18 hours, under argon at room temperature, Figure 104. The organic layer was washed with sodium bicarbonate solution (10% w/v, 20 mL), citric acid (20 mL) and brine (20 mL). It was then dried over anhydrous sodium sulphate. The solution was filtered and the solvent removed by rotary evaporation. The crude yellow oil was purified via column chromatography (5% MeOH in DCM) but purification was unsuccessful yielding a crude yellow oil (1.43 g, 2.3 mmol, crude, R_f 0.79) which was carried straight onto the next step.

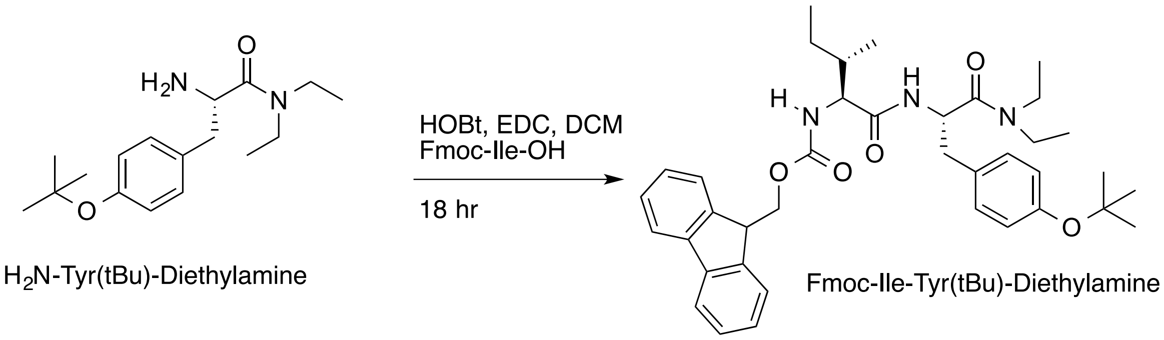


Figure 104: Synthesis of Fmoc-Ile-Tyr(tBu)-Diethylamine

8.5.13.2.4 Synthesis of H_2N-Ile-Tyr(tBu)-Diethylamine

Fmoc-Ile-Tyr(tBu)-diethylamine (1.43 g, 2.3 mmol) was dissolved in ethyl acetate (10 mL). DBU (0.41 mL, 3.0 mmol) was added dropwise and the reaction left stirring under argon at room temperature for 3 hours, Figure 105. The solvent was removed by rotary evaporation and the crude oil was purified by column chromatography over a solvent gradient (10% MeOH in DCM). Following on from purification a yellow oil was obtained 0.70 g, 1.73 mmol, 75% yield, R_f 0.47).

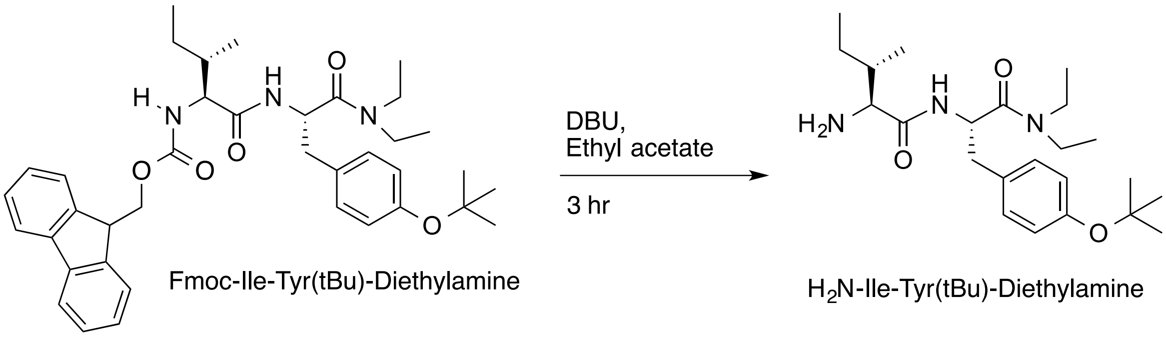


Figure 105: Synthesis of H_2N-Ile-Tyr(tBu)-Diethylamine

8.5.13.2.5 Synthesis of Acetyl-Ile-Tyr(tBu)-Diethylamine

H_2N-Ile-Tyr(tBu)-diethylamine (0.70g, 1.73 mmol) was dissolved in DCM (5 mL). One equivalent each of triethylamine (0.5 mL , 1.7 mmol) and acetic anhydride (0.2 mL , 1.7 mmol) were added dropwise and the reaction stirred under argon for 2 hours at room temperature, Figure 106. The solvent was removed by rotary evaporation and the crude oil was purified by column chromatography (5% MeOH in DCM) yielding a yellow oil (0.73 g, 1.63 mmol, 94% yield, R_f 0.66).

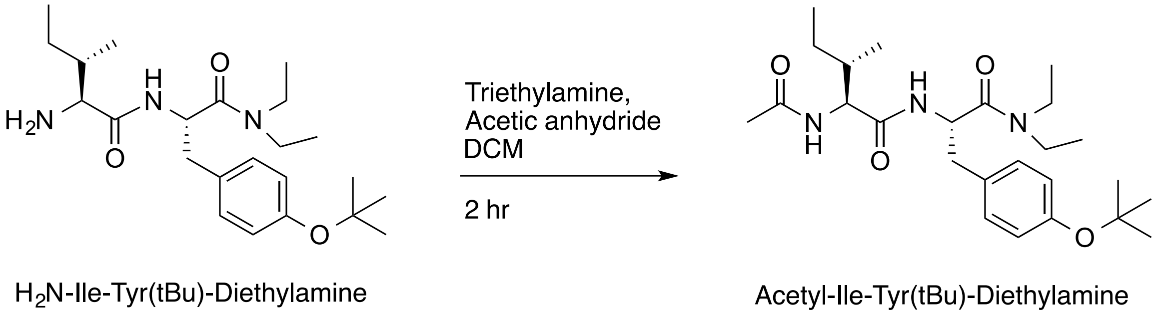


Figure 106: Synthesis of Acetyl-Ile-Tyr(tBu)-Diethylamine

8.5.13.2.6 Synthesis of Acetyl-Ile-Tyr-Diethylamine

Acetyl-Ile-Tyr(tBu)-diethylamine (0.73 g, 1.63 mmol) was dissolved in a cleavage cocktail of TFA, TIS and water (9.5:0.25:0.25, 10 mL) and stirred for 2 hours, Figure 107. The solvent was removed by rotary evaporation, and the peptide was washed with DCM (3 × 10 mL) and then precipitated in ice cold diethyl ether (10 mL). The peptide was filtered and allowed to dry yielding an orange power (0.43 g, 1.10 mmol, 67% yield). The crude peptide was purified by RP HPLC, then lyophilised to yield a cream solid (0.24 g, 0.61 mmol, 55%). LR MS (ESI+) *m/z* (%) 664.5 ($[2M+H]^+$, 100%). HR HPLC-MS (ESI+) for $C_{45}H_{50}N_3O_4$ $(M+H)^+$ Calculated mass 664.3878, found mass 664.3898 (2.9 ppm error). Analytical HPLC (280 nm) 22.84 min (97%).

Insufficient material for determining a clear NMR spectrum.

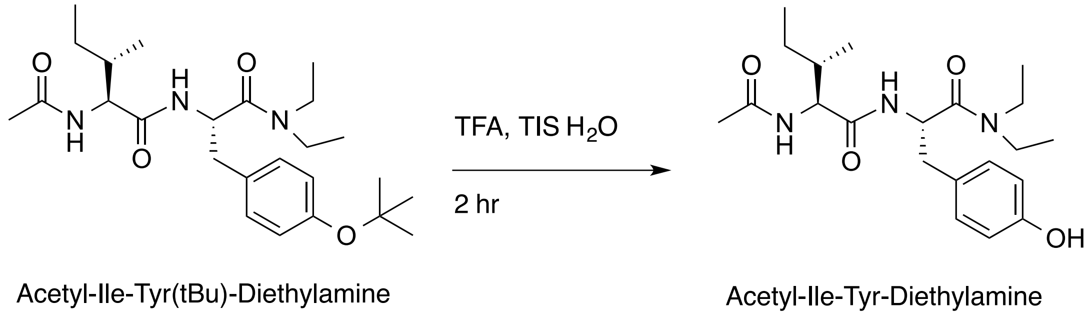
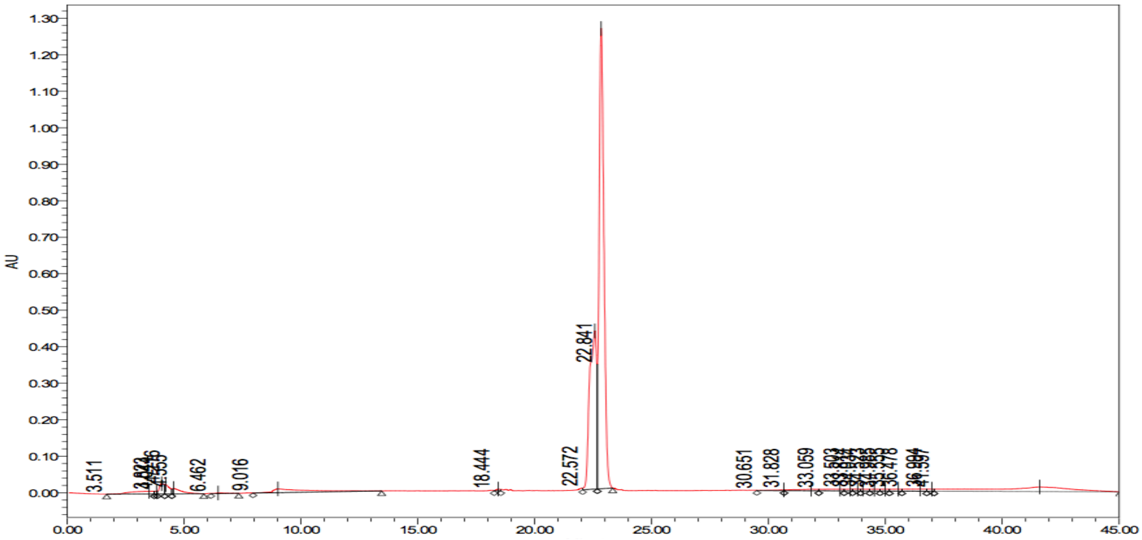
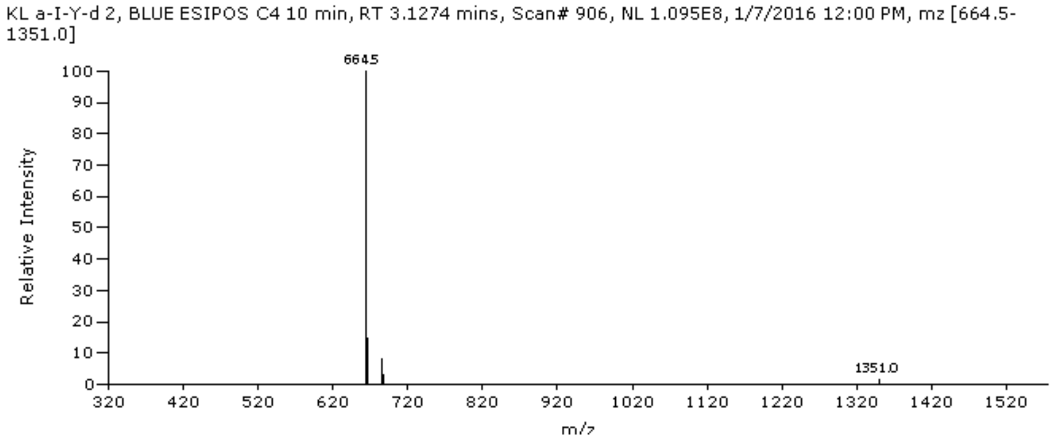


Figure 107: Mechanism of the Synthesis of Acetyl-Ile-Tyr-Diethylamine





8.5.13.3 Synthesis of Acetyl-YW-Diethylamine

8.5.13.3.1 Synthesis of Fmoc-Trp(Boc)-Diethylamine

Fmoc-Trp(Boc)-OH (1.81 g, 3.22 mmol) was dissolved in DCM (15 mL) with one equivalent each of HOBt (0.4 g, 3.2 mmol) and EDC (0.6 g, 3.2 mmol). Diethylamine (0.7 mL, 6.5 mmol) was added dropwise and the reaction left stirring under argon for 18 hours at room temperature, Figure 96. The reaction was stopped and then diluted with DCM (20 mL). The organic layer was washed with sodium bicarbonate solution (10% w/v, 20 mL), citric acid (20 mL) and brine (20 mL). It was then dried over anhydrous sodium sulphate. The solution was filtered and the solvent removed by rotary evaporation. The crude yellow oil was purified via column chromatography (9:1, DCM:MeOH) yielding a white solid (1.64 g, 2.82 mmol, 87% yield, R_f 0.55).

8.5.13.3.2 Synthesis of H_2N-Trp(Boc)-Diethylamine

Fmoc-Trp(Boc)-diethylamine (1.64 g, 2.82 mmol) was dissolved in ethyl acetate (20 mL). DBU (0.42 mL, 2.8 mmol) was added dropwise and the reaction left stirring under argon at room temperature for 3 hours, Figure 97. The solvent was removed by rotary evaporation and the crude oil was purified by column chromatography (2% MeOH in DCM). Following on from purification a yellow oil was obtained (0.72 g, 2.0 mmol, 70% yield, R_f 0.13).

8.5.13.3.3 Synthesis of Fmoc-Tyr(tBu)-Trp(Boc)-Diethylamine

H_2N-Trp(Boc)-diethylamine (0.72 g, 2.00 mmol) was dissolved in DCM (15 mL) with two equivalents of Fmoc-Tyr(tBu)-OH (1.8 g , 4.0 mmol), EDC (0.5 g , 4.0 mmol) and HOBt (0.8 g , 4.0 mmol). The reaction was left to stir for 18 hours, under argon at room temperature, Figure 98. The organic layer was washed with sodium bicarbonate solution (10% w/v, 20 mL), citric acid (20 mL) and brine (20 mL). It was then dried over anhydrous sodium sulphate. The solution was filtered and the solvent removed by rotary evaporation. The crude yellow oil was purified via column chromatography (9:1, DCM:MeOH) yielding a white solid (1.33 g, 1.66 mmol, 83% yield, R_f 0.55).

8.5.13.3.4 Synthesis of H_2N-Tyr(tBu)-Trp(Boc)-Diethylamine

Fmoc-Tyr(tBu)-Trp(Boc)-diethylamine (1.33 g, 1.66 mmol) was dissolved in ethyl acetate (15 mL). DBU (0.298 mL, 2.0 mmol) was added dropwise and the reaction left stirring under argon at room temperature for 3 hours, Figure 99. The solvent was removed by rotary evaporation and the crude oil was purified by column chromatography over a solvent gradient (2% MeOH in DCM). Following on from purification a yellow oil was obtained (0.97 g, 1.6 mmol, 96% yield, R_f 0.47).

8.5.13.3.5 Synthesis of Acetyl-Tyr(tBu)-Trp(Boc)-Diethylamine

H_2N-Tyr(tBu)-Trp(Boc)-diethylamine (0.97 g, 1.6 mmol) was dissolved in DCM (10 mL). One equivalent each of triethylamine (0.1 mL, 1.6 mmol) and acetic anhydride (0.1 mL, 1.6 mmol) were added dropwise and the reaction stirred under argon for 3 hours at room temperature, Figure 108.The solvent was removed by rotary evaporation and the crude oil was purified by column chromatography (5% MeOH in DCM) yielding a yellow oil (0.93 g, 1.50 mmol, 93% yield, R_f 0.61).

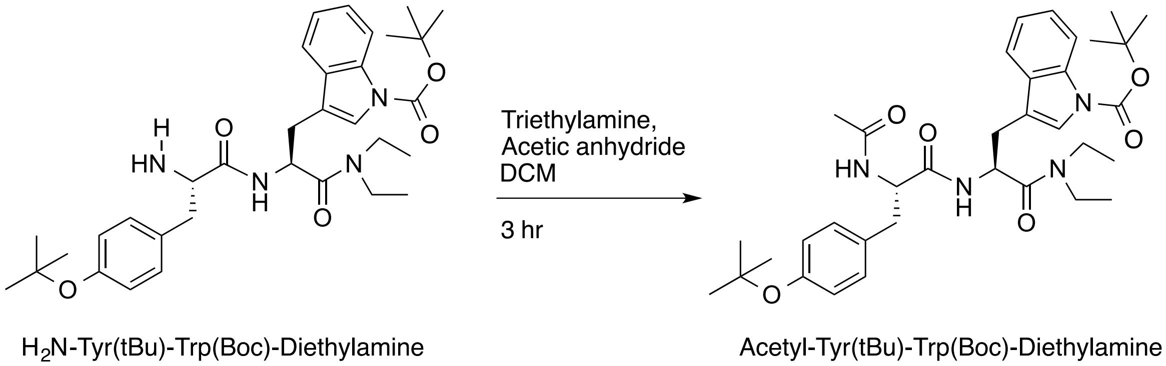


Figure 108: Synthesis of Acetyl-Tyr(tBu)-Trp(Boc)-Diethylamine

8.5.13.3.6 Synthesis of Acetyl-Tyr-Trp-Diethylamine

Acetyl-Tyr(tBu)-Trp(Boc)-diethylamine (0.93 g, 1.50 mmol) was dissolved in a cleavage cocktail of TFA, TIS and water (9.5:0.25:0.25, 10 mL) and stirred for 2 hours, Figure 109. The solvent was removed by rotary evaporation, and the peptide was washed with DCM (3 × 10 mL) and then precipitated in ice cold diethyl ether (10 mL). The peptide was filtered and allowed to dry yielding a pale orange solid (0.38 g, 1.26 mmol, 84% yield). LR MS (ESI+) *m/z* (%) 601.4 ([2M+H]$^+$, 100%). HR HPLC-MS (ESI+) for $C_{26}H_{32}N_4O_4$ (M+H)$^+$ Calculated mass 601.3184, found mass 601.3192 (1.3 ppm error). Analytical HPLC (280 nm) 19.72 min (95%).

^{1}H NMR (400 MHz, ACETONITRILE-d_3) δ = 9.09 (1H, br s, Trp indole NH), 7.85 (2H, dd, J5,28, 25 = 7.6 Hz, J = 2.4 Hz, Tyr backbone NH), 7.64 (1H, d, J = 7.3 Hz, Trp backbone NH), , 7.47 (2H, t, J28, 25,5 = 7.6 Hz, Trp E3, Z2 -H), 7.14 (1H, d, J = 2.3 Hz, Trp D1-H), 7.09 (2H, ddd, J = 9.3 Hz, J27, 26,28, 25 = 7.6 Hz, J = 1.2 Hz, Trp H2, Z3 -H), 6.80 - 6.83 (2H, m, Tyr E*-H), 6.63 - 6.65 (2H, m, Tyr D*-H), 5.02 - 5.09 (1H, m, Trp α-H), 3.80 (1H, dd, J = 8.9 Hz, J = 5.6 Hz, Tyr α-H), 3.66 - 3.89 (1H, m, Diethylamine CH_2), 3.38 - 3.48 (2H, m, Diethylamine CH_2), 3.31 (1H, dt, J = 14.9 Hz, J = 7.4 Hz, Diethylamine CH_2), 3.10 (1H, dd, J = 4.5 Hz, J = 5.4 Hz, Trp β-H), 2.99 (1H, br dd, J = 4.6 Hz, J = 9.1 Hz, Tyr β-H), 2.72 (1H, dd, J = 3.6 Hz, J = 8.9 Hz, Tyr β-H), 2.50 (1H, dd, J = 2.7 Hz, J = 5.7 Hz, Tyr β-H), 1.11 (3H, t, J = 7.1 Hz, Diethylamine CH_3), 1.05 (3H, t, J = 7.1 Hz, Diethylamine CH_3)

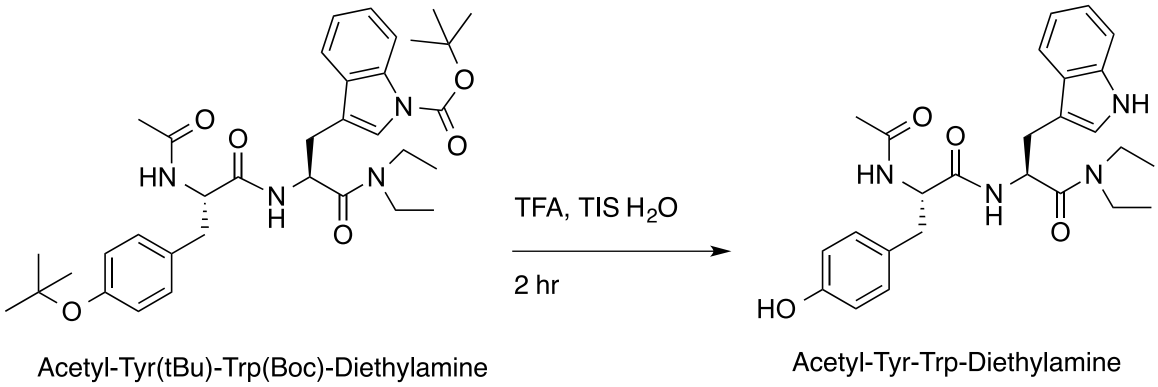
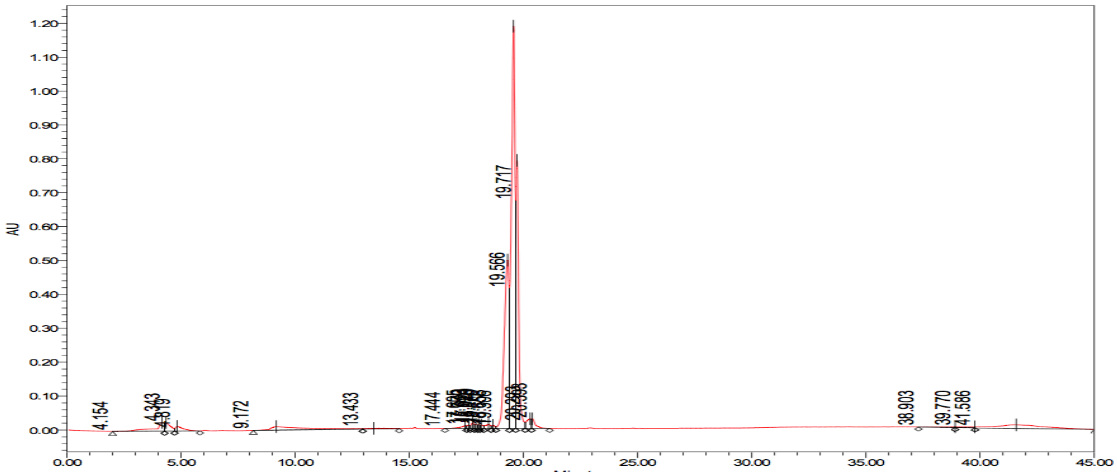
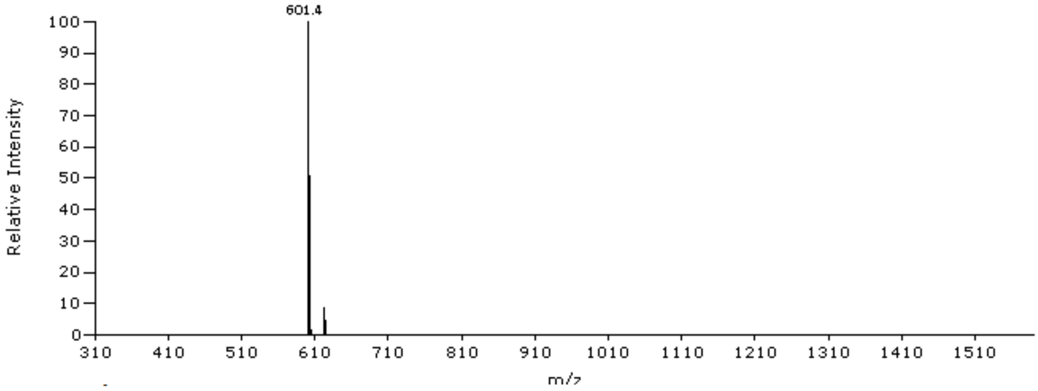
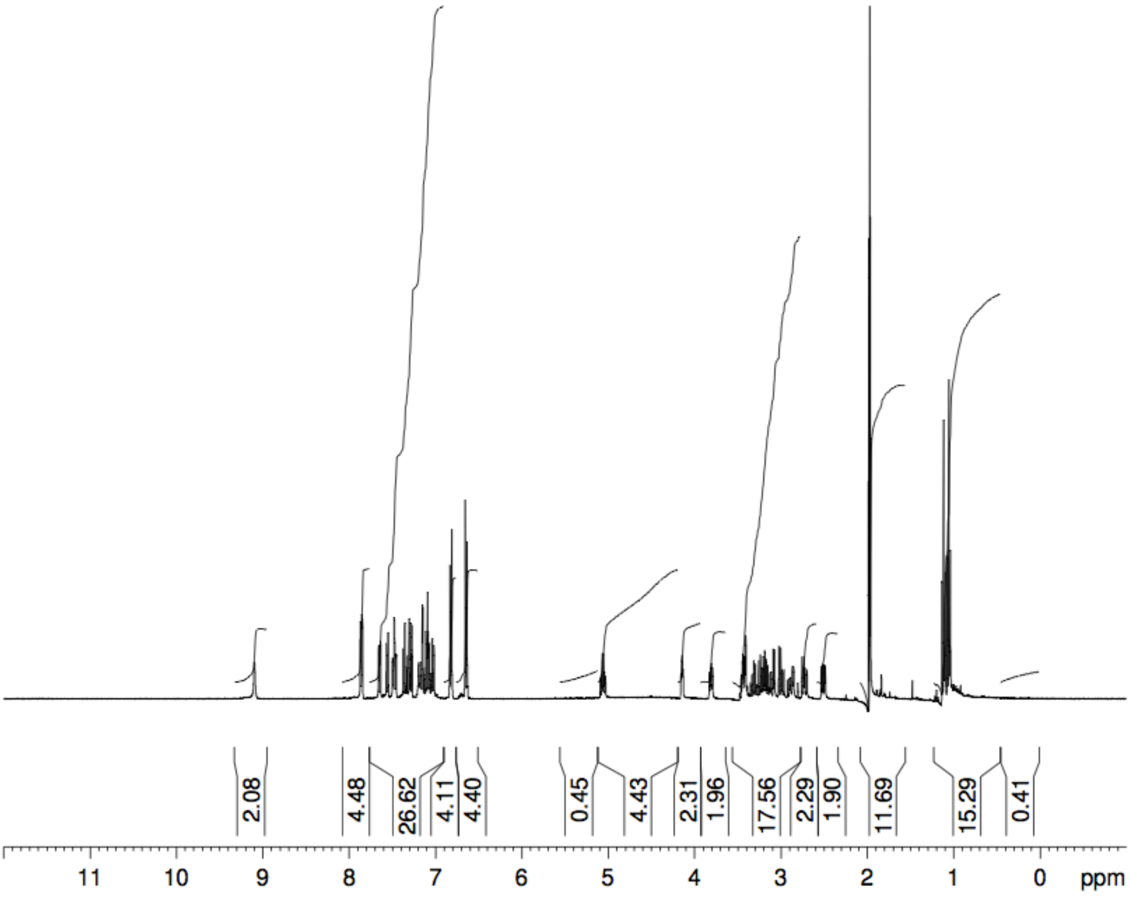


Figure 109: Synthesis of Acetyl-Tyr-Trp-Diethylamine



KL a-Y-W-d 2, BLUE ESIPOS C4 10 min, RT 2.2368 mins, Scan# 648, NL 9.724E7, 1/7/2016 12:01 PM, mz [601.4-624.4]





8.6 Experimental for Chapter 5

8.6.1 CP11Y*

This series of cyclic peptides have the same backbone as CP11, however the tyrosine residue has been substituted for phenylalanine or tyrosine analogues. The analogues were synthesised in bulk on a Glycine Wang resin (2.5 mmol scale) as 8.2.1, until the first tryptophan residue. The resin was then split into 0.6 g batches, the tyrosine analogue, isoleucine and final tryptophan residue were then coupled to the linear resin bound peptide.

8.6.1.1 CP11Y1

CP11Y1 (IFWNVSGW) was cleaved from the resin to form the linear peptide (pale pink solid, 119.5 mg crude). This peptide (119.5 mg, 0.12 mmol) was cyclised, Figure 110, and purified via preparative HPLC yielding a pale pink solid (33.8 mg, 0.034 mmol, 28% yield). LR MS (ESI+) *m/z* (%) 990.8 ($[M+H]^+$, 100%), 496.0 (½$[M+2Na]^{2+}$, 87.13%), 1012.8 ($[M+Na]^+$, 62.78%). Analytical HPLC (280 nm) 19.8 mins (97%).

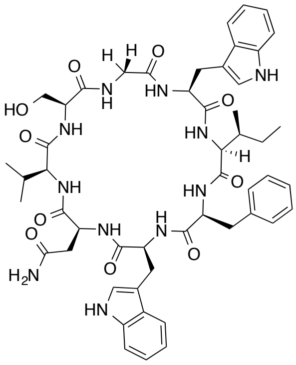
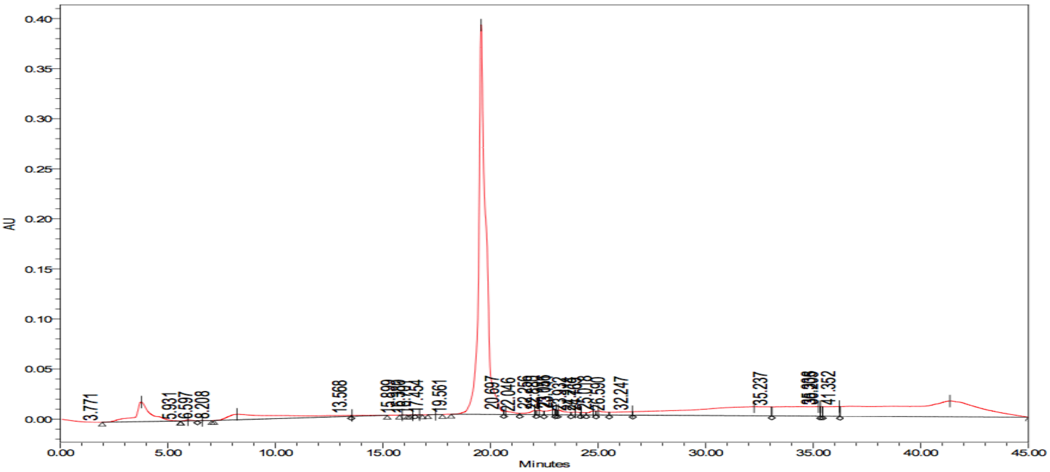
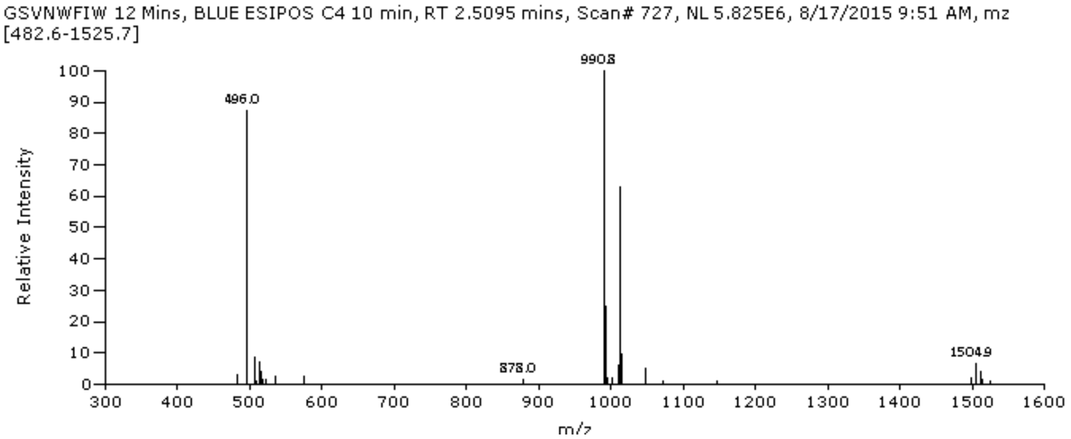


Figure 110: CP11Y1





8.6.1.2 CP11Y2

CP11Y2 (IF(4-F)WNVSGW) was cleaved from the resin to form the linear peptide (beige solid, 110.3 mg crude). This peptide (110.3 mg, 0.11 mmol) was cyclised, Figure 111, and purified via preparative HPLC yielding a pale pink solid (24.9 mg, 0.025 mmol, 23% yield). LR MS (ESI+) *m/z* (%) 505.1 (½$[M+2Na]^{2+}$, 100%), 1030.8 ($[M+Na]^{+}$, 55.29%), 1009.9 ($[M+H]^{+}$, 51.69%). Analytical HPLC (280 nm) 19.4 mins (98%).

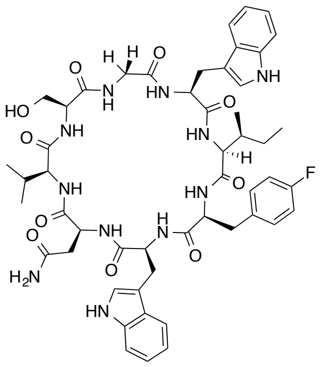
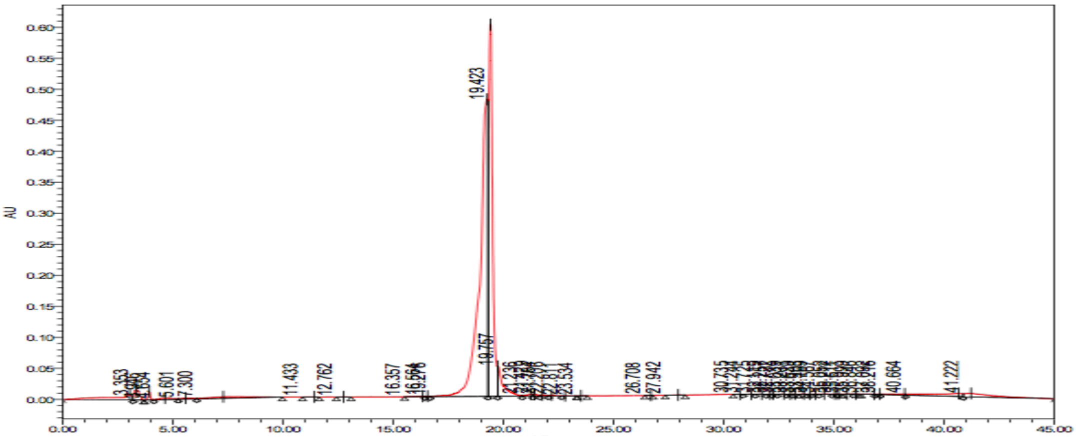
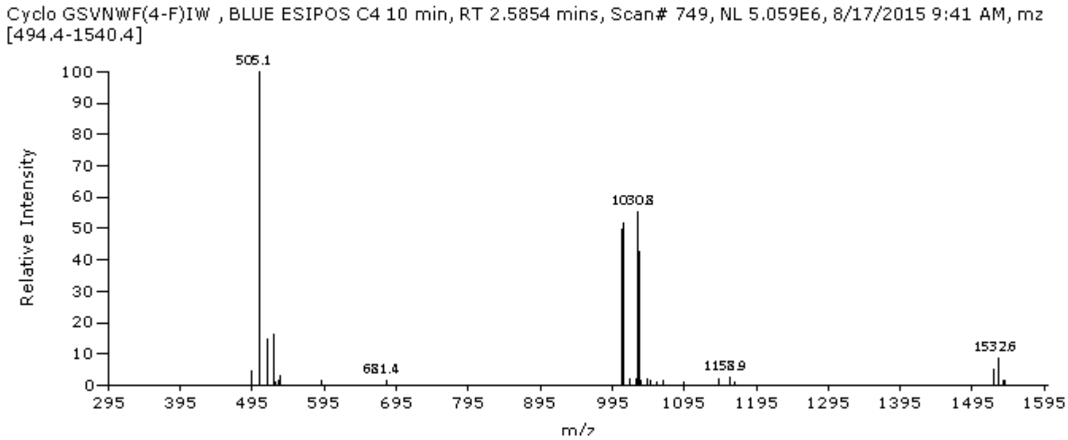


Figure 111: CP11Y2





8.6.1.3 CP11Y3

CP11Y3 (IF(4-Cl)WNVSGW) was cleaved from the resin to form the linear peptide (pale yellow solid, 276.3 mg crude). This peptide (276.3 mg, 0.27 mmol) was cyclised, Figure 112, and purified via preparative HPLC yielding a pale cream solid (27.8 mg, 0.027 mmol, 10% yield). LR MS (ESI+) *m/z* (%) 1024.8 ($[M+H]^+$, 100%), 1046.9($[M+Na]^+$, 90.93%), 514.1 (½$[M+2Na]^{2+}$, 53.07%). Analytical HPLC (280 nm) 21.1 mins (98%).

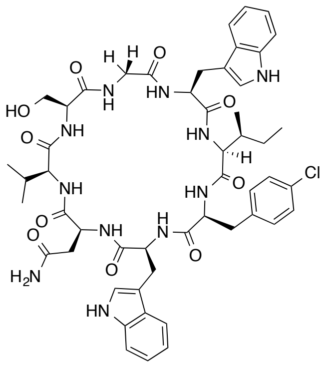
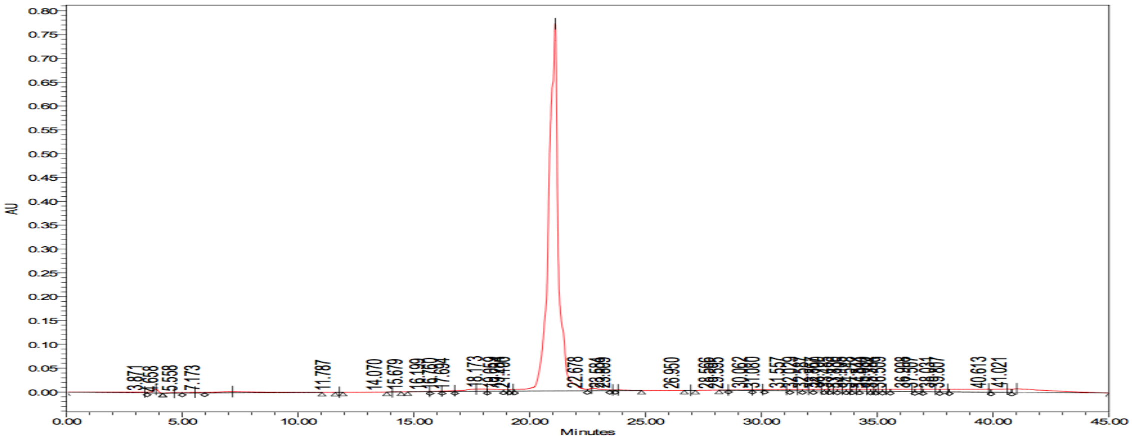
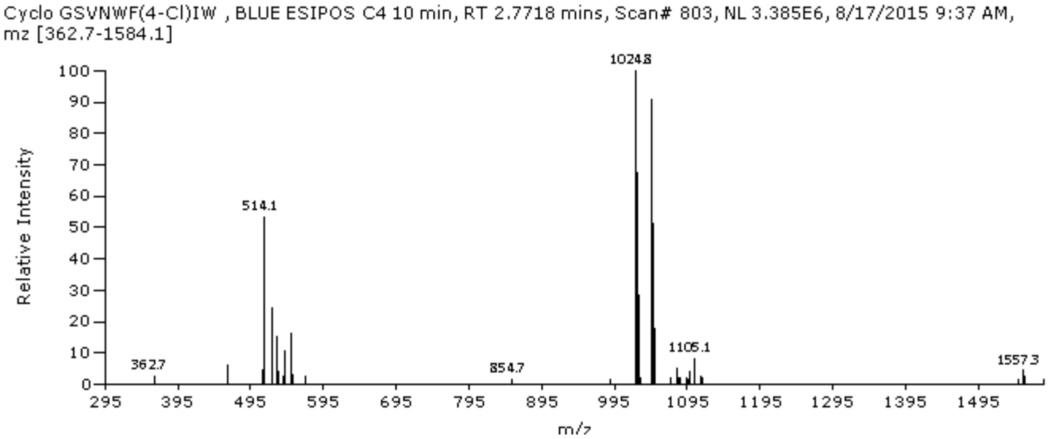


Figure 112: CP11Y3





8.6.1.4 CP11Y4

CP11Y4 (IF(4-Me)WNVSGW) was cleaved from the resin to form the linear peptide (pale pink solid, 150.6 mg crude). This peptide (150.6 mg, 0.15 mmol) was cyclised, Figure 113, and purified via preparative HPLC yielding a pale pink solid (18 mg, 0.018 mmol, 12% yield). LR MS (ESI+) *m/z* (%) 1027.8 ([M+Na]$^+$, 100%), 1004.8 ([M+H]$^+$, 63.39%), 503.0 (½[M+2Na]$^{2+}$, 48.35%). Analytical HPLC (280 nm) 21.9 mins (95%).

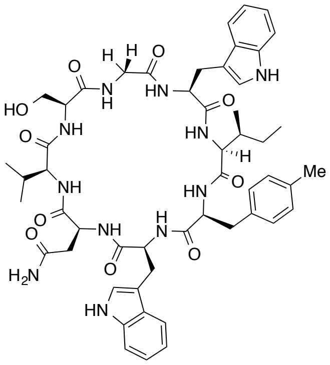
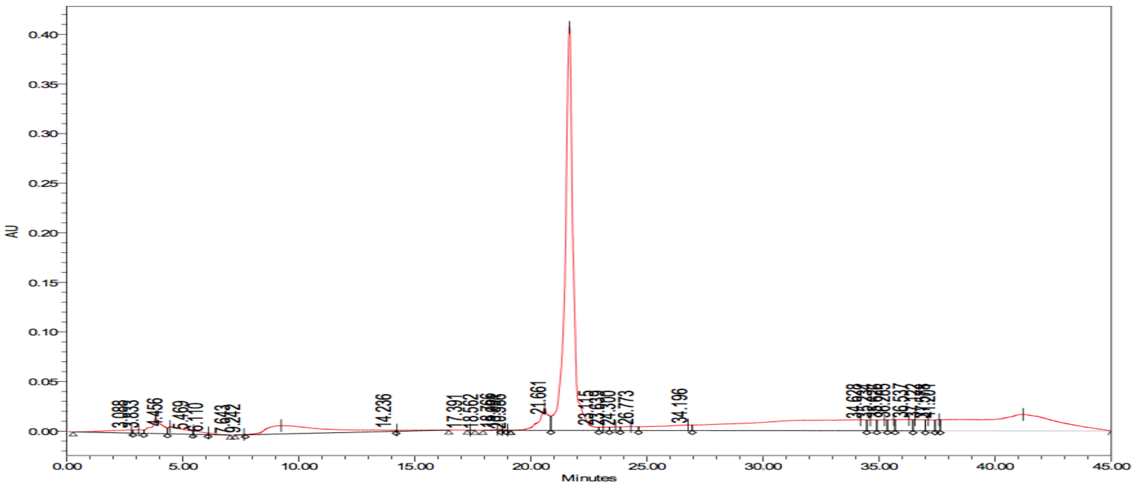
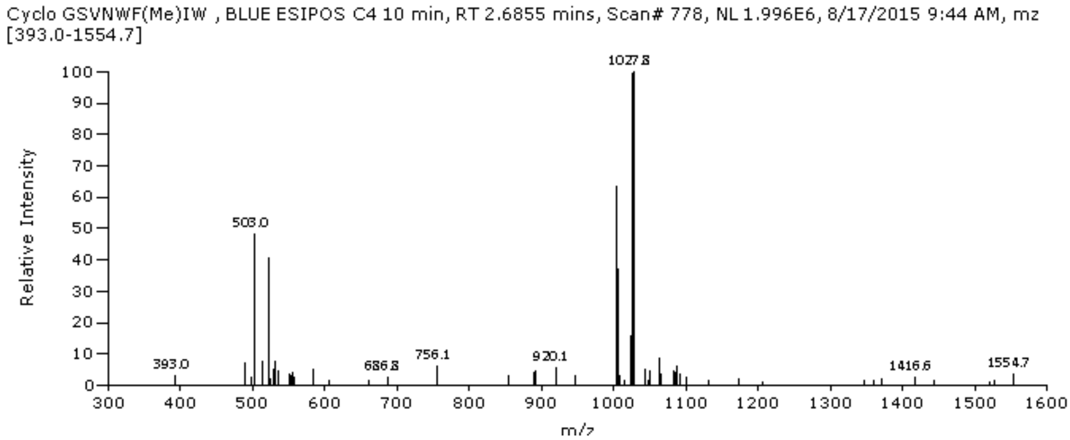


Figure 113: CP11Y4





8.6.1.5 CP11Y5

CP11Y5 (IF(4-CN)WNVSGW) was cleaved from the resin to form the linear peptide (pale orange solid, 124.4 mg crude). This peptide (124.4 mg, 0.12 mmol) was cyclised, Figure 114, and purified via preparative HPLC yielding a pale orange solid (19.6 mg, 0.019 mmol, 16% yield). LR MS (ESI+) *m/z* (%) 1015.9 ($[M+H]^+$, 100%), 1037.8 ($[M+Na]^+$, 74.32%), 508.5 ($½[M+2Na]^{2+}$, 43.65%). Analytical HPLC (280 nm) 18.8 mins (92%).

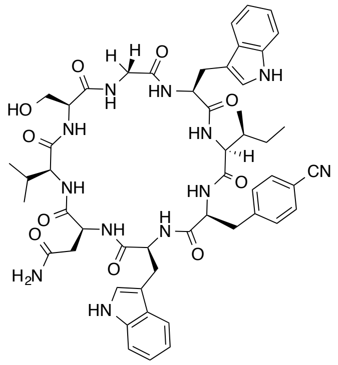
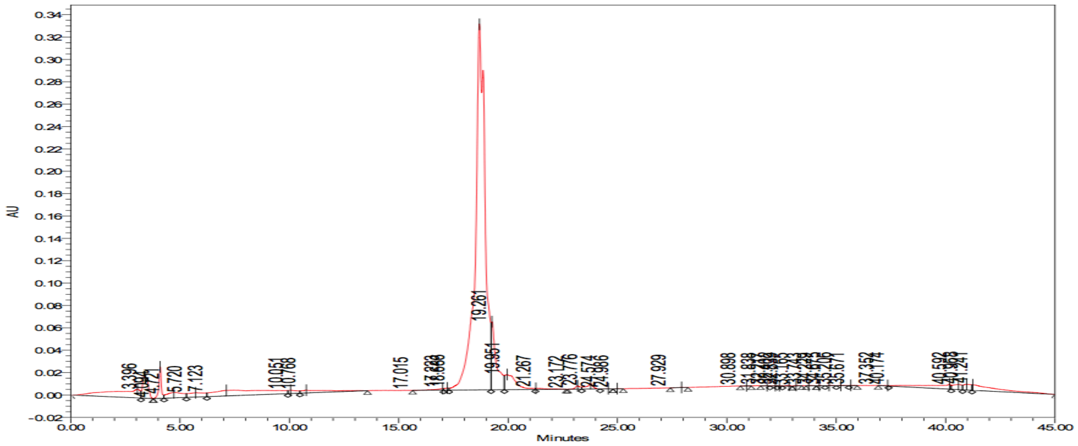
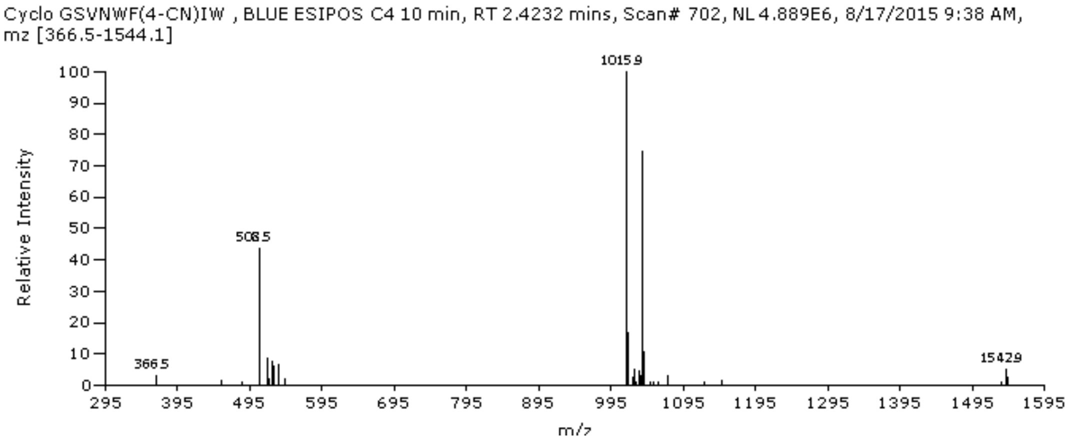


Figure 114: CP11Y5





8.6.1.6 CP11Y6

CP11Y6 (IF(4-NO$_2$)WNVSGW) was cleaved from the resin to form the linear peptide (pale yellow solid, 180 mg crude). This peptide (115 mg, 0.109 mmol) was cyclised, Figure 115, and purified via preparative HPLC yielding a pale yellow solid (25.2 mg, 0.023 mmol, 22% yield). LR MS (ESI+) *m/z* (%) 1035.9 ([M+H]$^+$, 100%), 1057.9 ([M+Na]$^+$, 77.63%), 518.7 (½[M+2Na]$^{2+}$, 35.45%). Analytical HPLC (280 nm) 19.4 mins (89%).

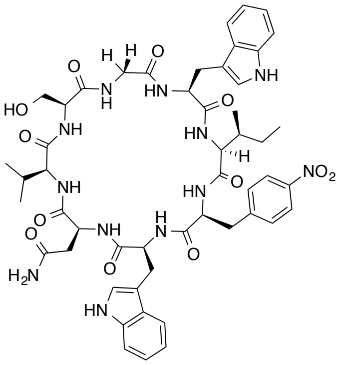
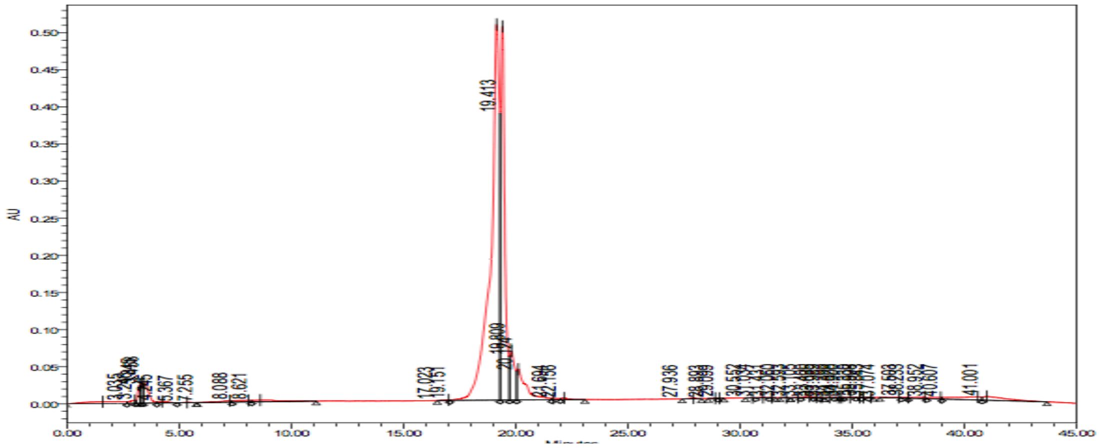
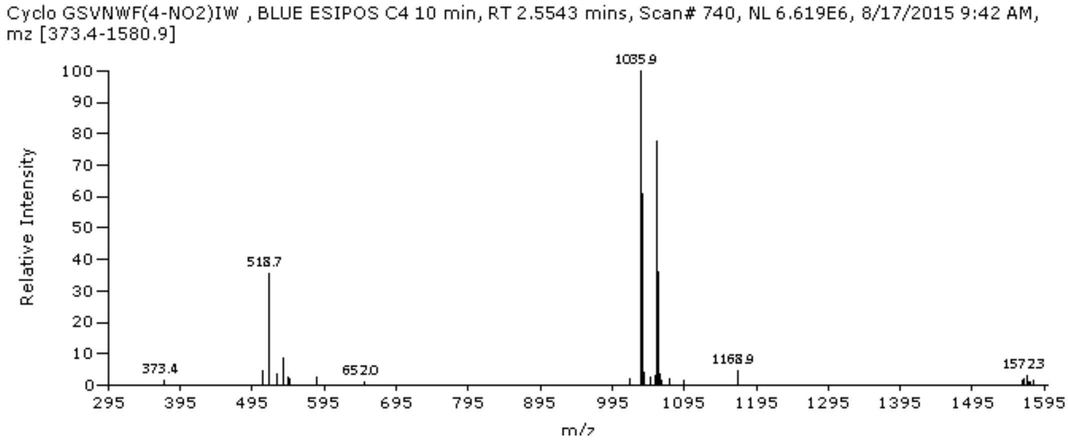


Figure 115: CP11Y6





8.6.1.7 CP11Y7

CP11Y7 (IY(3-NO$_2$)WNVSGW) was cleaved from the resin to form the linear peptide (yellow solid, 316 mg crude). This peptide (108.6 mg, 0.10 mmol) was cyclised, Figure 116, and purified via preparative HPLC yielding a yellow solid (20.8 mg, 0.02 mmol, 19% yield). LR MS (ESI+) *m/z* (%) 1051.9 ([M+H]$^+$, 100%), 1073.9 ([M+Na]$^+$, 77.63%), 526.4 (½[M+2Na]$^{2+}$, 70.89%). Analytical HPLC (280 nm) 18.9 mins (94%).

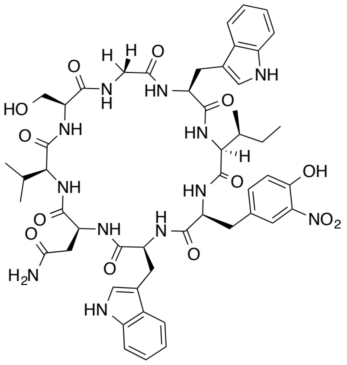
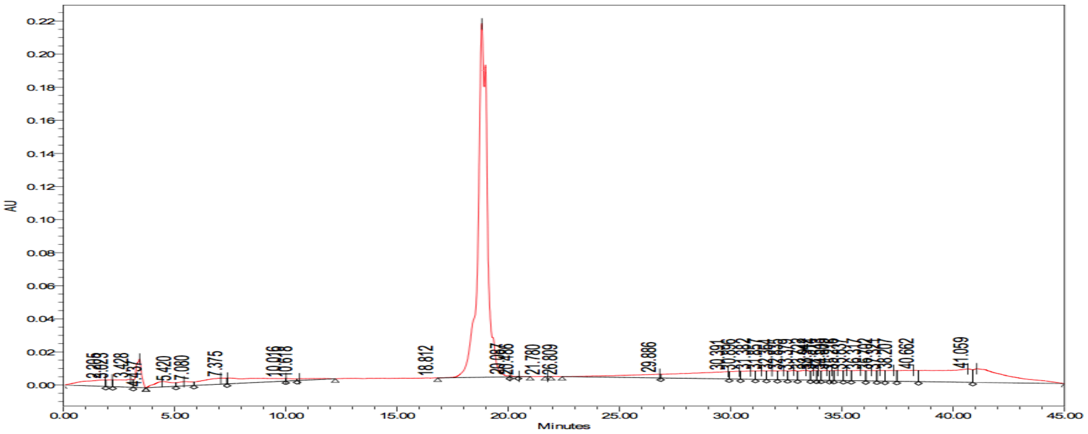
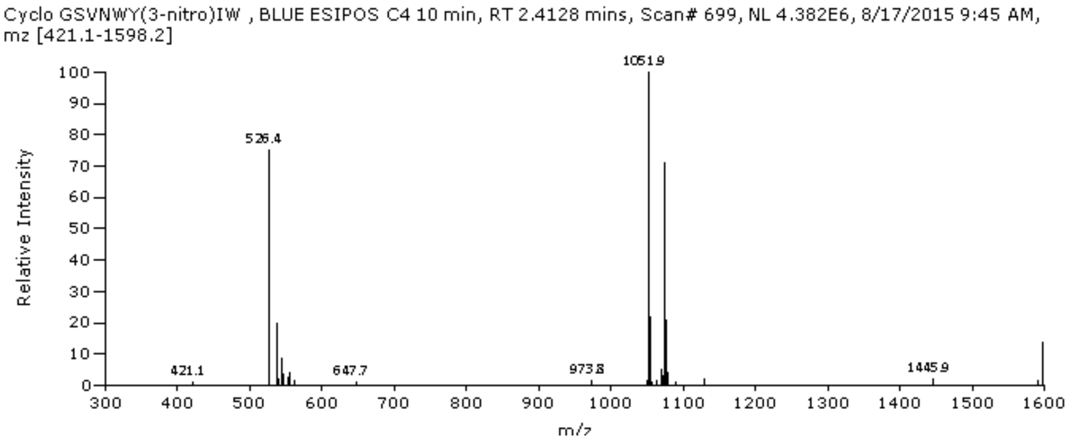


Figure 116: CP11Y7





8.6.1.8 CP11Y8

CP11Y8 (IY(Me)WNVSGW) was cleaved from the resin to form the linear peptide (pale pink solid, 165.3 mg crude). This peptide (103.6 mg, 0.10 mmol) was cyclised, Figure 117, and purified via preparative HPLC yielding a pale pink solid (16.5 mg, 0.016 mmol, 16% yield). LR MS (ESI+) *m/z* (%) 1020.9 ([M+H]$^+$, 100%), 1042.9 ([M+Na]$^+$, 97.10%), 511.0 (½[M+2Na]$^{2+}$, 72.17%). Analytical HPLC (280 nm) 19.9 mins (95%).

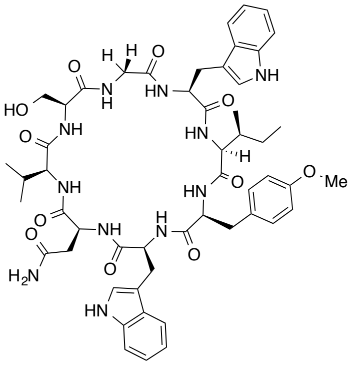
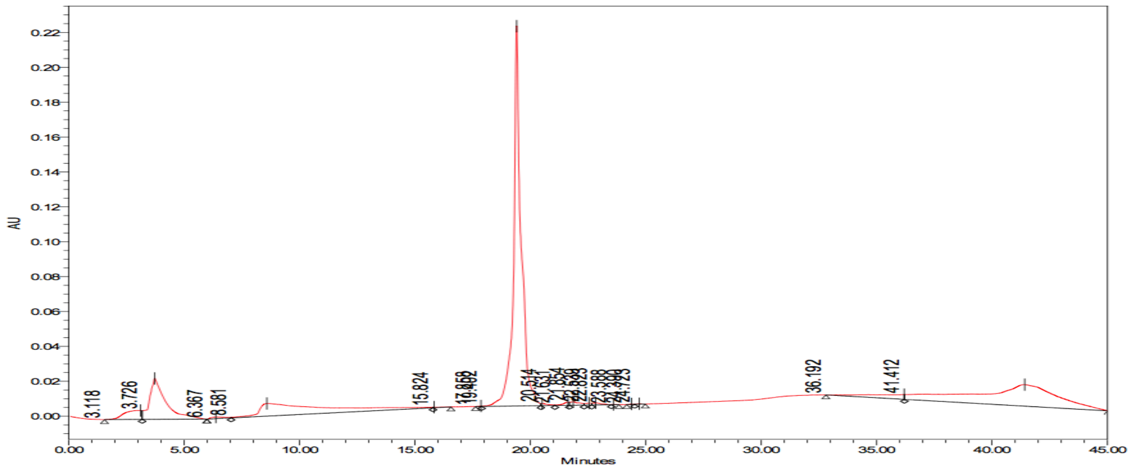
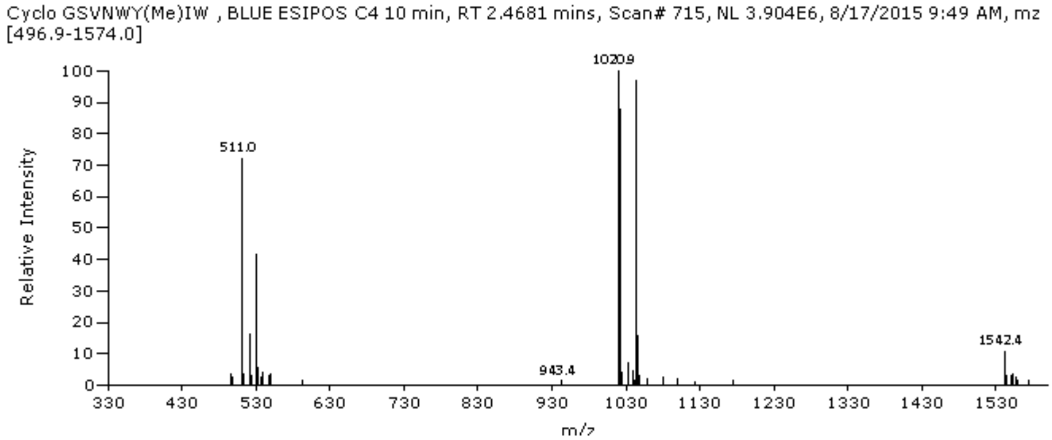


Figure 117: CP11Y8





8.6.1.9 CP11Y9

CP11Y9 (IPalWNVSGW) was cleaved from the resin to form the linear peptide (pale cream solid, 105.2 mg crude). This peptide (105.2 mg, 0.10 mmol) was cyclised, Figure 118, and purified via preparative HPLC yielding a pale cream solid (21.1 mg, 0.021 mmol, 21% yield). LR MS (ESI+) *m/z* (%) 996.7 ([M+H]$^+$, 100%), 1018.7 ([M+Na]$^+$, 64.49%), 499.1 (½[M+2Na]$^{2+}$, 43.97%). Analytical HPLC (280 nm) 15.5 mins (96%).

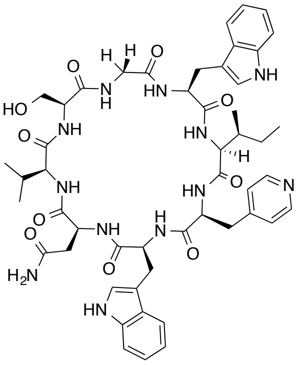
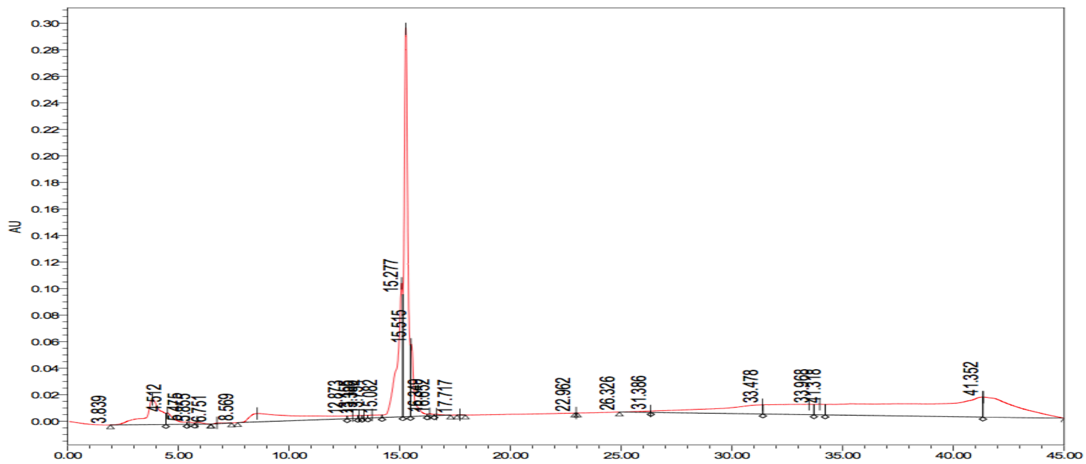
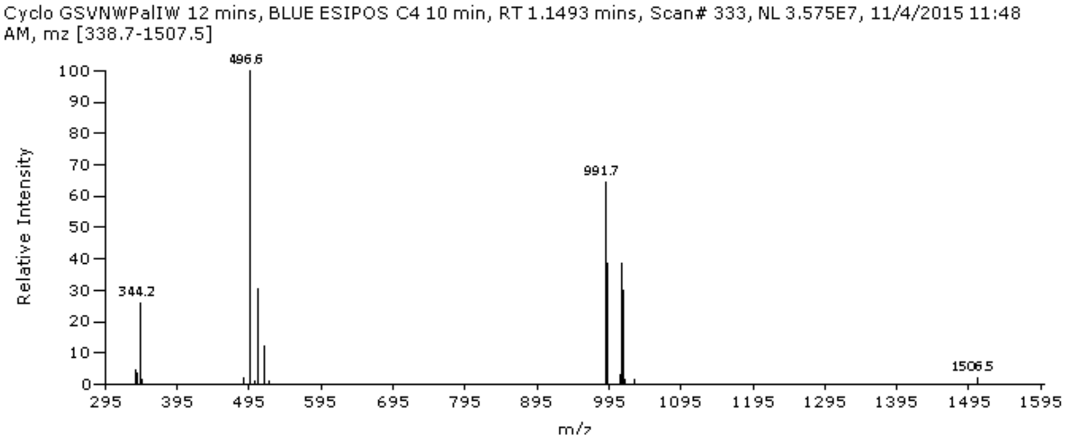


Figure 118: CP11Y9





8.6.1.10 CP11Y10

CP11Y10 (IChaWNVSGW) was cleaved from the resin to form the linear peptide (off white solid, 163.6 mg crude). This peptide (101.5 mg, 0.10 mmol) was cyclised, Figure 119, and purified via preparative HPLC yielding a pale cream solid (11.6 mg, 0.012 mmol, 11% yield). LR MS (ESI+) *m/z* (%) 496.6 (½[M+2Na]$^{2+}$, 100%), 991.7 ([M+H]$^{+}$, 64.52%), 1013.6 ([M+Na]$^{+}$, 38.48%). Analytical HPLC (280 nm) 20.2 min (98%).

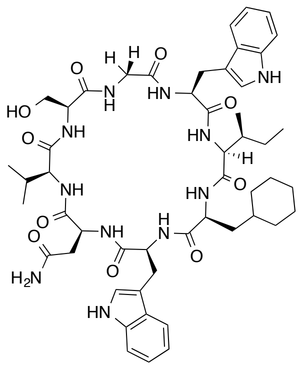
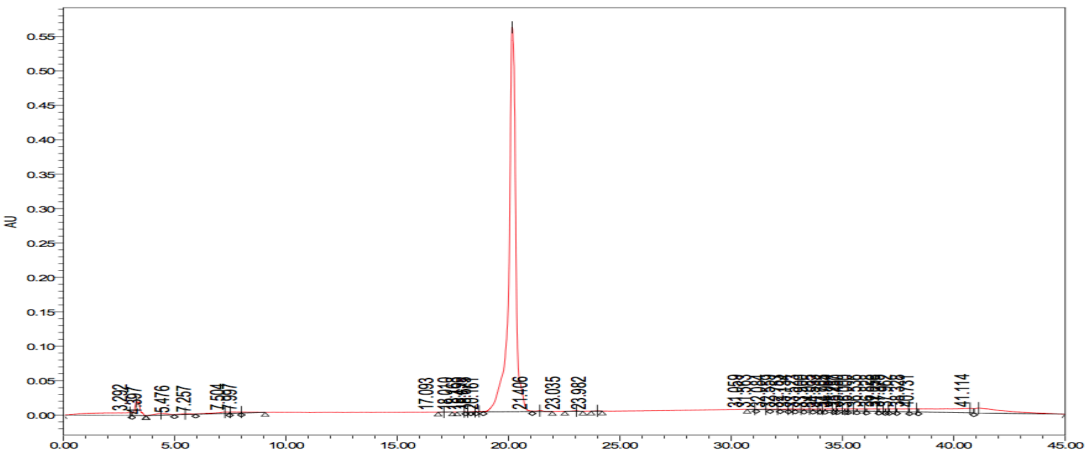
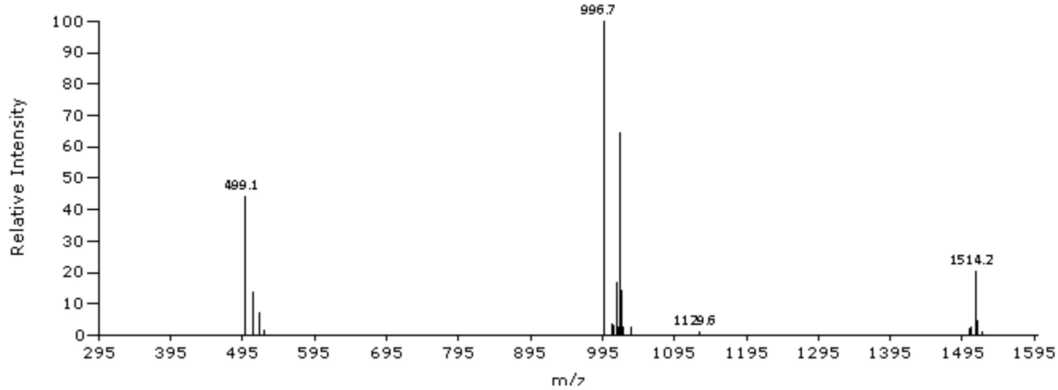


Figure 119: CP11Y10



Cyclo GSVNWChaIW 12 mins, BLUE ESIPOS C4 10 min, RT 2.7890 mins, Scan# 808, NL 5.095E7, 11/4/2015 11:48 AM, mz [499.1-1522.3]



8.6.1.11 CP11Y11

CP11Y11 (INalWNVSGW) was cleaved from the resin to form the linear peptide (white solid, 367 mg crude). This peptide (121.4 mg, 0.11 mmol) was cyclised, Figure 120, and purified via preparative HPLC yielding a pale pink solid (27.7 mg, 0.027 mmol, 24% yield). LR MS (ESI+) *m/z* (%) 1040.8 ($[M+H]^+$, 100%), 1063.8 ($[M+Na]^+$, 56.44%), 521.5 (½$[M+2Na]^{2+}$, 62.62%). Anaytical HPLC (280 nm) 21.2 min (98%).

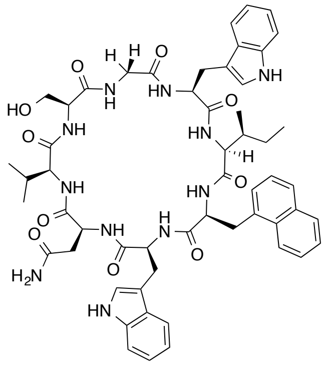
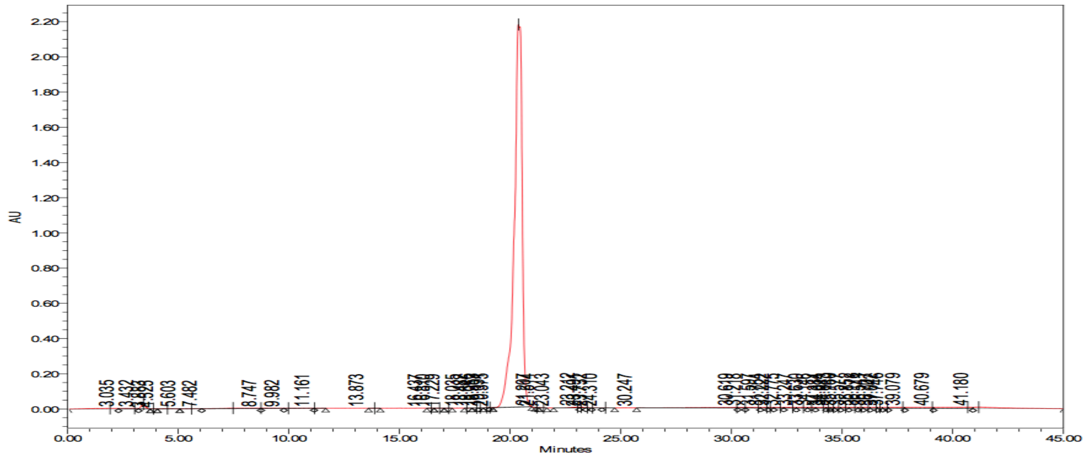
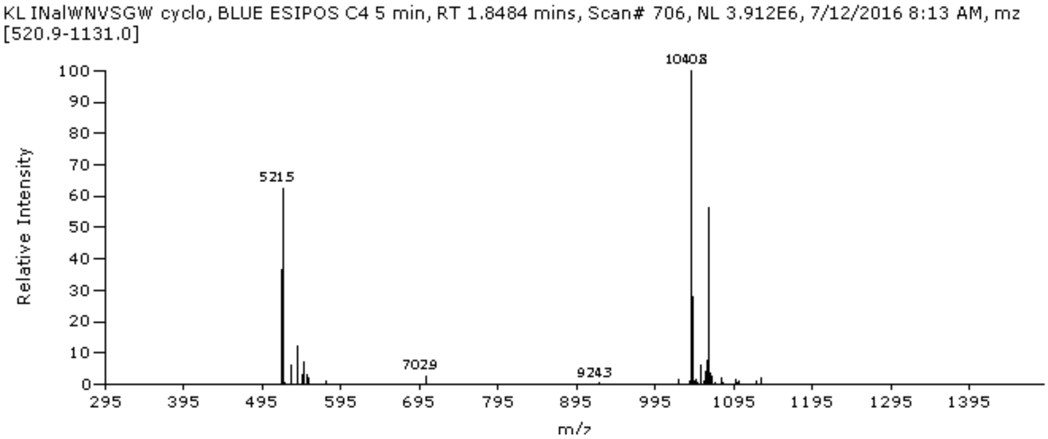


Figure 120: CP11Y11





8.6.2 CP11A4Y*

This series of cyclic peptides have the same backbone as CP11A4, however the tyrosine residue has been substituted for phenylalanine or tyrosine analogues. The analogues were synthesised in bulk on an Isoleucine Wang resin (3.3 mmol scale) as 8.2.1, until the second tryptophan residue. The resin was then split into 0.6 g batches, then the tyrosine analogues were coupled to complete the linear resin bound peptides.

8.6.2.1 CP11A4Y1

CP11A4Y1 (IFWAVSGW) was cleaved from the resin to form the linear peptide (pale pink solid, 113.3 mg crude). This peptide (113.3 mg, 0.12 mmol) was cyclised, Figure 121, and purified via preparative HPLC yielding a pale pink solid (26.5 mg, 0.03 mmol, 23% yield). LR MS (ESI+) *m/z* (%) 969.8 ([M+H]$^{+}$, 100%), 948.7 ([M+Na]$^{+}$, 51.64%), 493.7 (½[M+2Na]$^{2+}$, 31.31%). Analytical HPLC (280 nm) 20.4 min (91%).

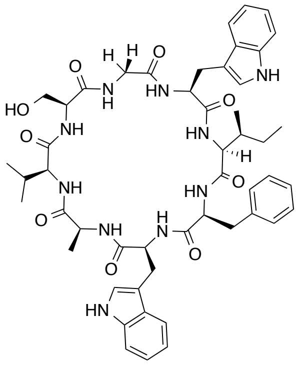
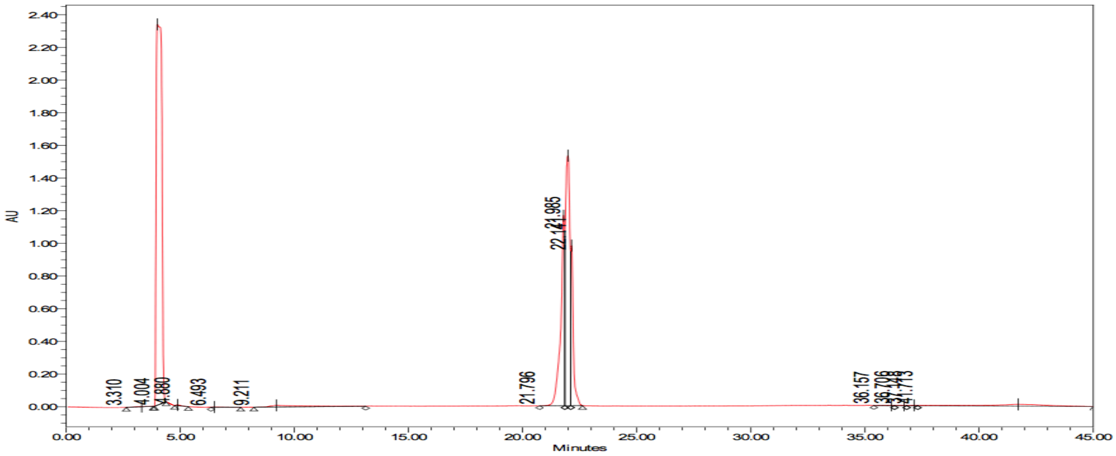
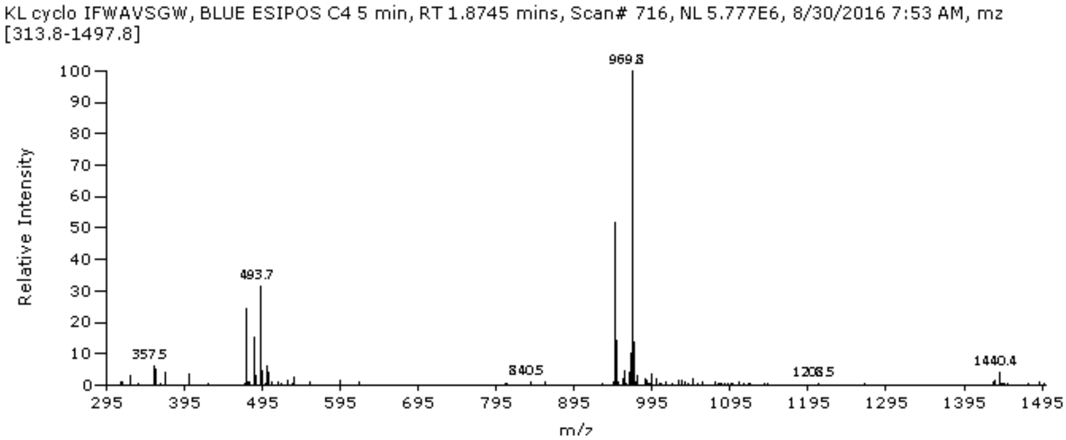
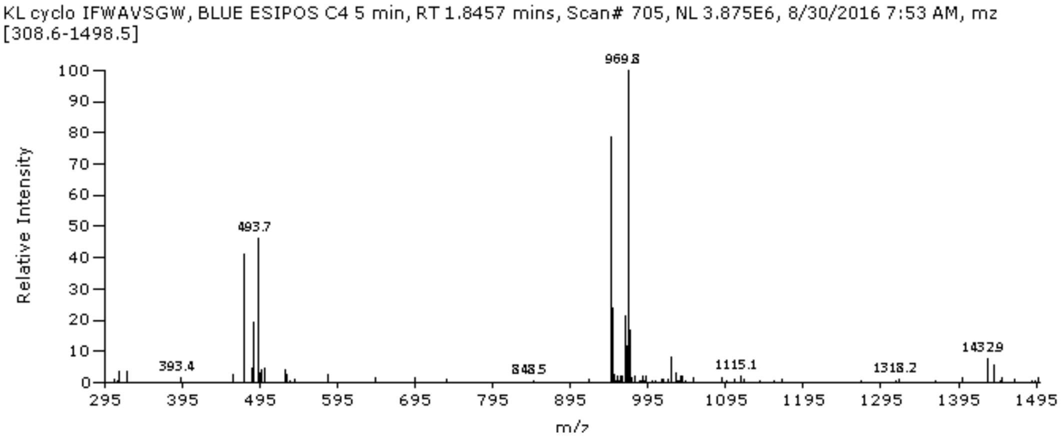


Figure 121: CP11A4Y1



The first peak is due to DMSO present in the stock solution. There were epimers present that could be detected separately in the Mass spectrum.





8.6.2.2 CP11A4Y2

CP11A4Y2 (IF(4-F)WAVSGW) was cleaved from the resin to form the linear peptide (pink solid, 70 mg crude). This peptide (70 mg, 0.07 mmol) was cyclised, Figure 122, and purified via preparative HPLC yielding a pale pink solid (4.8 mg, 0.005 mmol, 27% yield). LR MS (ESI+) *m/z* (%) 965.8 ($[M+H]^+$, 100%), 988.9 ($[M+Na]^+$, 41.47%), 502.2 (½$[M+2Na]^{2+}$, 17.68%). Analytical (280 nm) 22.01 min (98%).

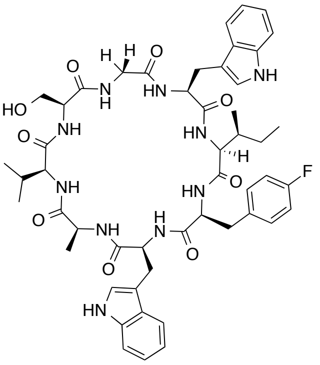
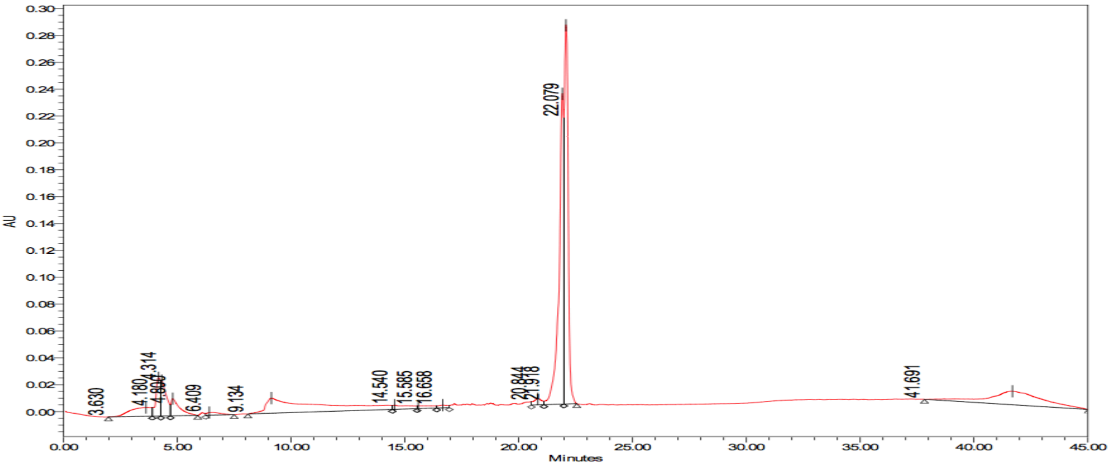
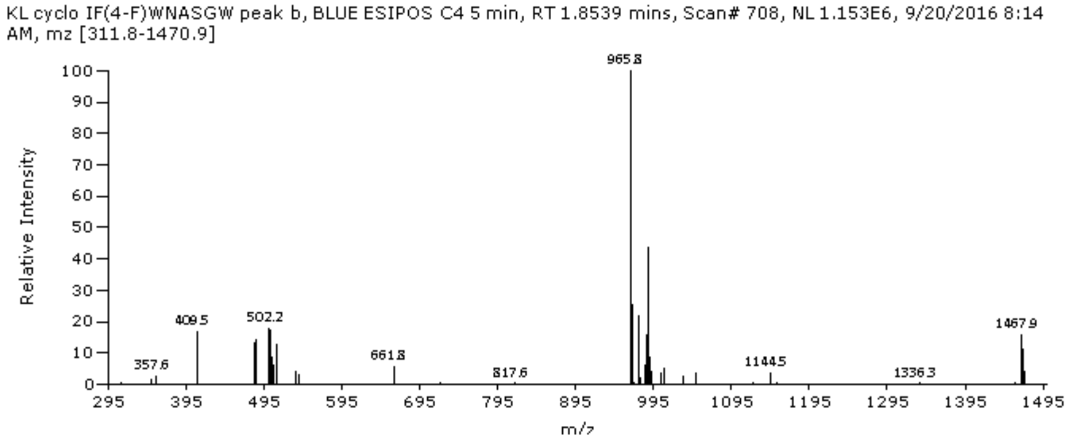


Figure 122: CP11Y2





8.6.2.3 CP11A4Y3

CP11A4Y3 (IF(4-Cl)WAVSGW) was cleaved from the resin to form the linear peptide (red solid, 271.4 mg crude). This peptide (118 mg, 0.12 mmol) was cyclised, Figure 123, and purified via preparative HPLC yielding a pale pink solid (18.2 mg, 0.049 mmol, 15% yield). LR MS (ESI+) *m/z* (%) 1003.8 ([M+Na]$^{+}$, 100%), 981.8 ([M+H]$^{+}$, 72.47%), 510.8 (½[M+2Na]$^{2+}$, 66.60%). Analytical HPLC (280 nm) 23.8 min (98%).

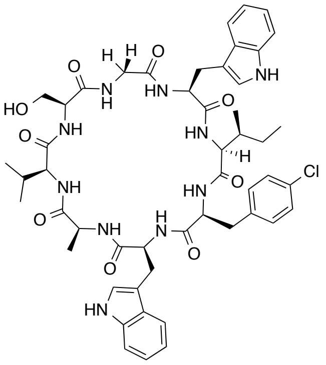
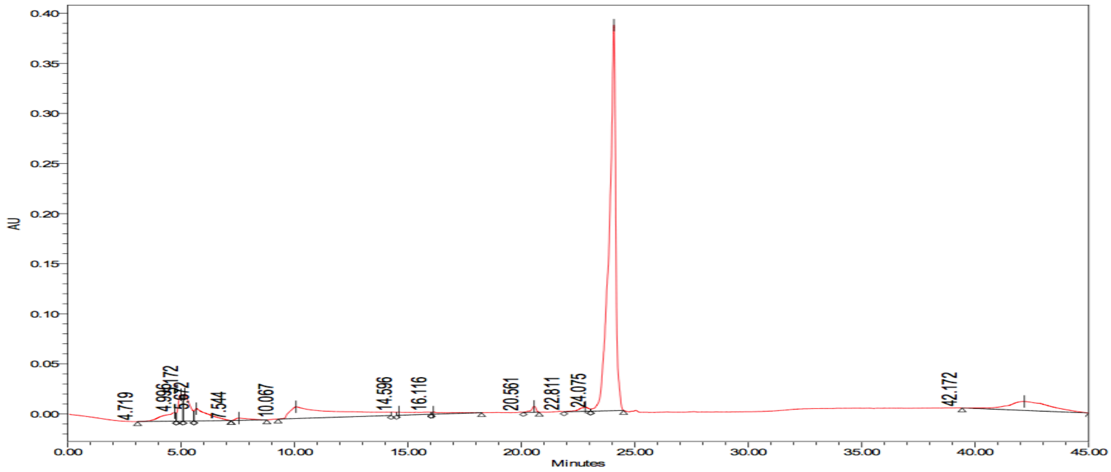
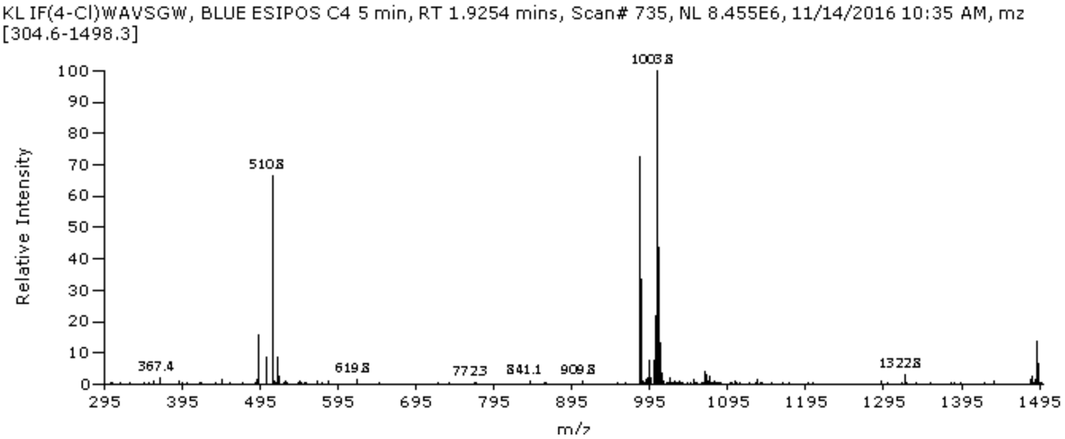


Figure 123: CP11A4Y3





8.6.2.4 CP11A4Y4

CP11A4Y4 (IF(4-Me)WAVSGW) was cleaved from the resin to form the linear peptide (orange solid, 332.1 mg crude). This peptide (118 mg, 0.12 mmol) was cyclised, Figure 124, and purified via preparative HPLC yielding a pale orange solid (11.2 mg, 0.012 mmol, 10 % yield). LR MS (ESI+) *m/z* (%) 983.7 ([M+Na]$^+$, 100%), 961.7 ([M+H]$^+$, 57.71%), 500.7 (½[M+2Na]$^{2+}$, 38.41%). Analytical HPLC (280 nm) 28.2 min (99%).

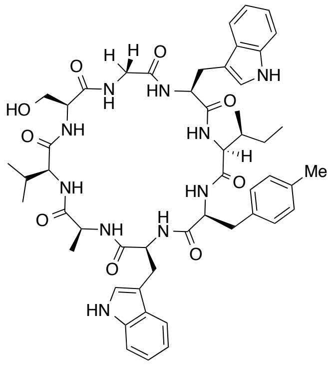
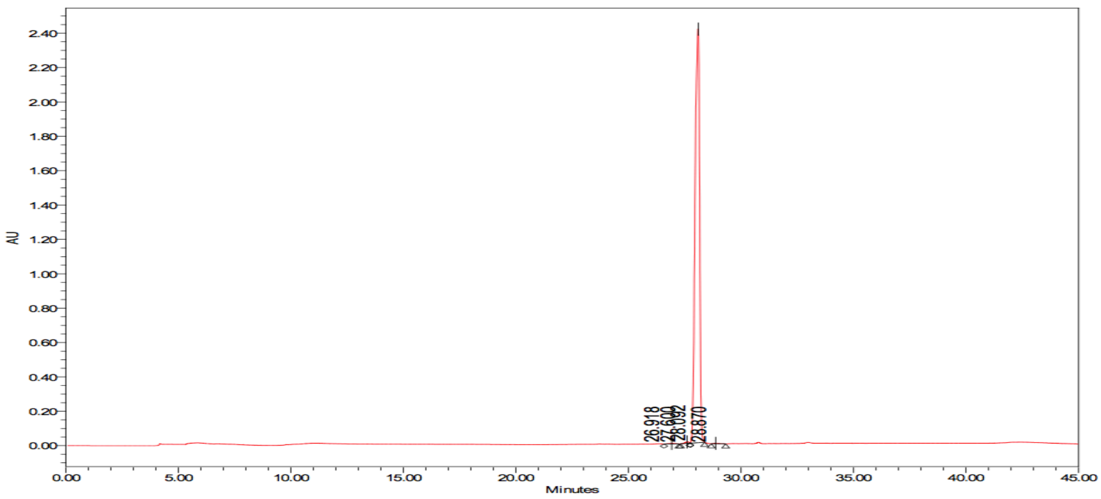
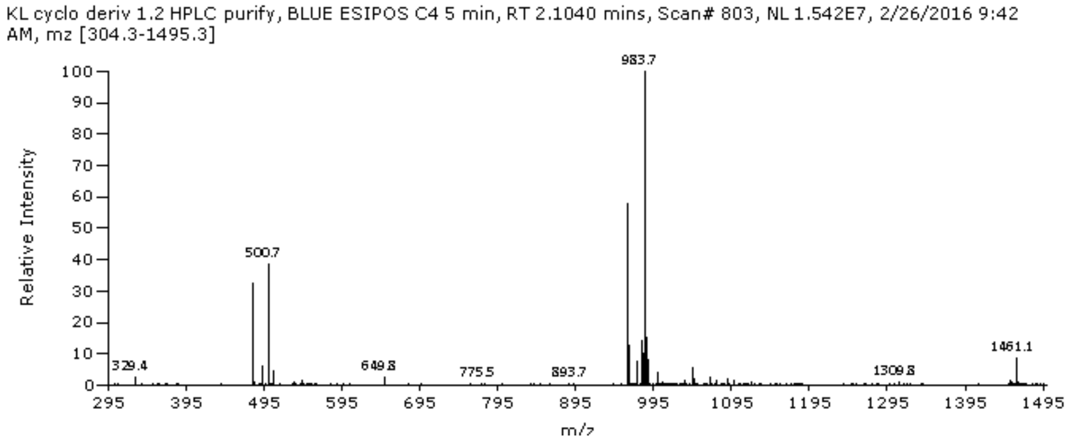


Figure 124: CP11A4Y4





8.6.2.5 CP11A4Y5

CP11A4Y5 (IF(4-CN)WAVSGW) was cleaved from the resin to form the linear peptide (orange solid, 106.3 mg crude). This peptide (106.3 mg, 0.10 mmol) was cyclised, Figure 125, and purified via preparative HPLC yielding a pale orange solid (16.4 mg, 0.017 mmol, 17% yield). LR MS (ESI+) *m/z* (%) 506.5 (½$[M+2Na]^{2+}$, 100%), 995.7 ($[M+Na]^{+}$, 87.46%), 972.8 ($[M+H]^{+}$, 84.01%). Analytical HPLC (280 nm) 25.1 min (98%).

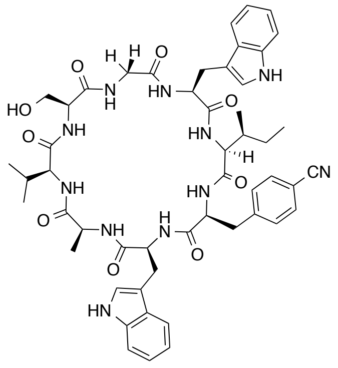
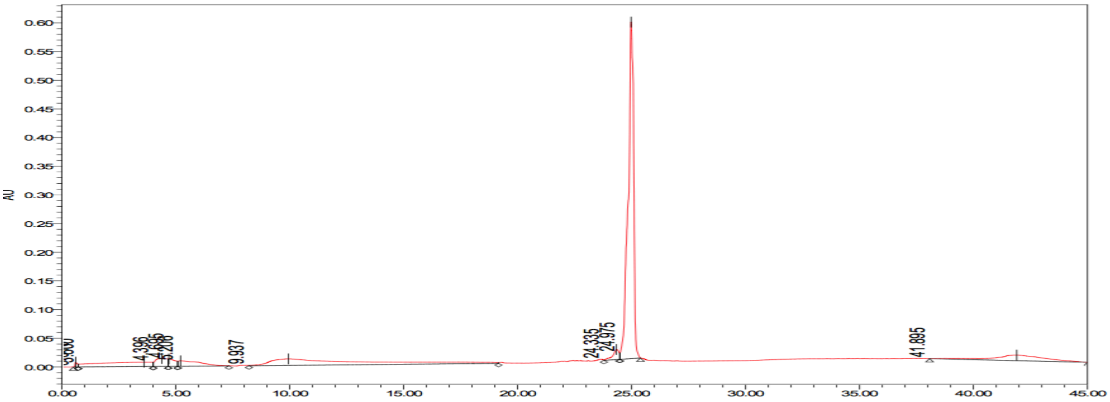
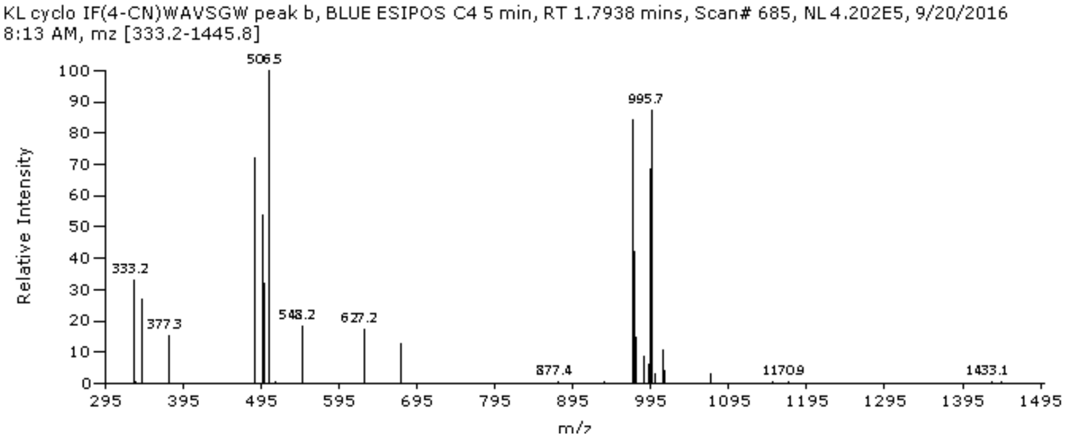


Figure 125: CP11A4Y5





8.6.2.6 CP11A4Y6

CP11A4Y6 (IF(4-NO_2)WAVSGW) was cleaved from the resin to form the linear peptide (dark yellow solid, 97.5 mg crude). This peptide (97.5 mg, 0.10 mmol) was cyclised, Figure 126, and purified via preparative HPLC yielding a pale orange solid (9.0 mg, 0.009 mmol, 9% yield). LR MS (ESI+) *m/z* (%) 1014.8 ($[M+Na]^+$, 100%), 992.7 ($[M+H]^+$, 46.91%), 515.9 (½$[M+2Na]^{2+}$, 36.09%). Analytical HPLC (280 nm) 22.1 min (95%).

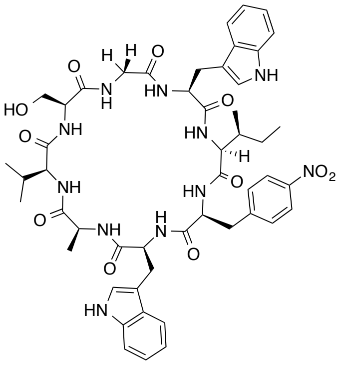
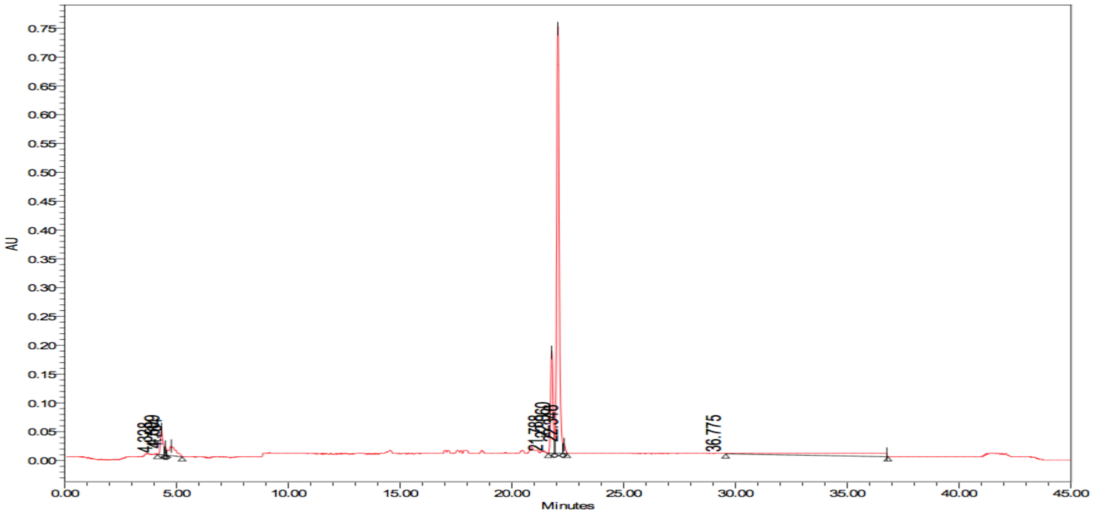
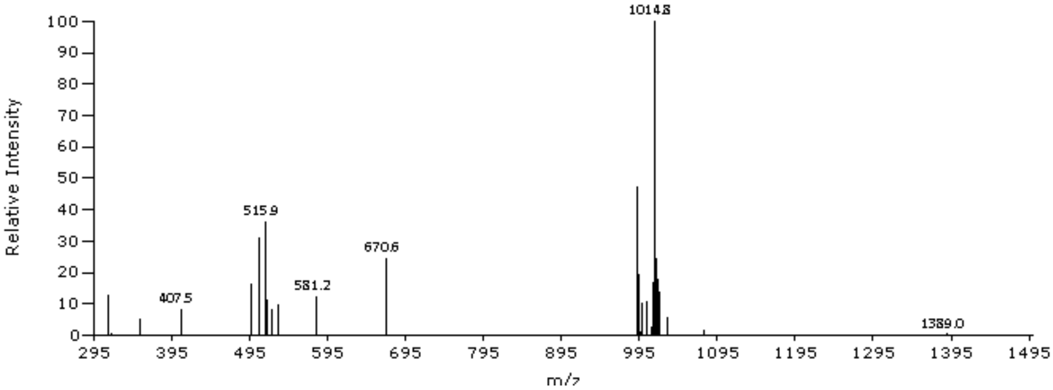


Figure 126: CP11A4Y6





8.6.2.7 CP11A4Y7

CP11A4Y7 (IY(3-NO$_2$)WAVSGW) was cleaved from the resin to form the linear peptide (yellow solid, 162.3 mg crude). This peptide (102.0 mg, 0.1 mmol) was cyclised, Figure 127, and purified via preparative HPLC yielding a pale orange solid (6.9 mg, 0.007 mmol, 7% yield). LR MS (ESI+) *m/z* (%) 1030.8 ([M+Na]$^+$, 100%), 1008.9 ([M+H]$^+$, 92.43%), 525.3 (½[M+2Na]$^{2+}$, 53.01%). Analytical HPLC (280 nm) of DMSO stock due to lack of material, hence lower purity, 21.5min (85%). First peak in analytical HPLC trace is due to DMSO.

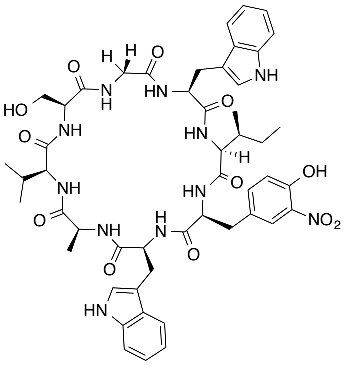
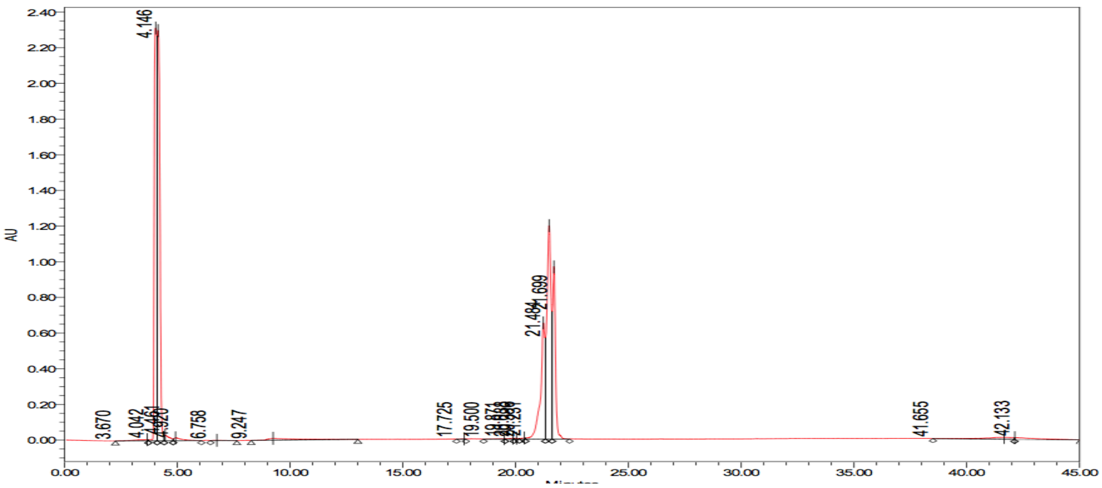
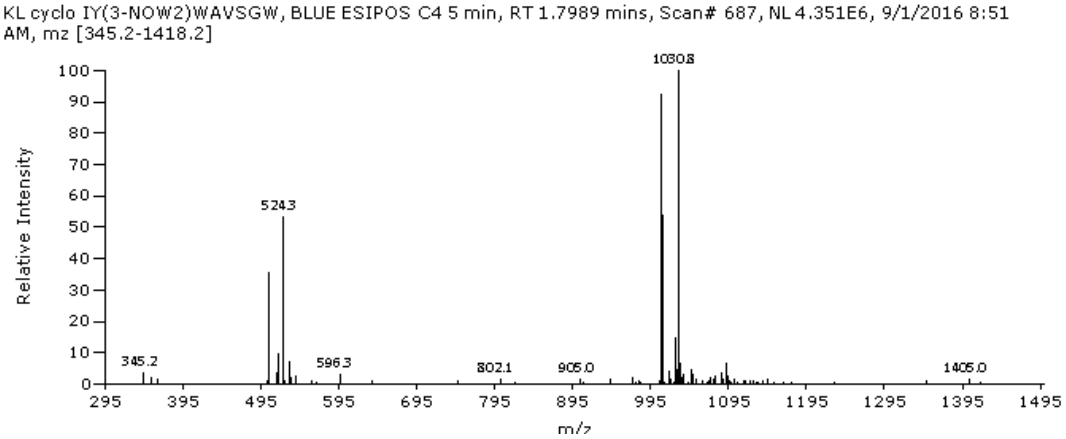


Figure 127: CP11A4Y7





8.6.2.8 CP11A4Y8

CP11A4Y8 (IY(Me)WAVSGW) was cleaved from the resin to form the linear peptide (pale orange solid, 145.2 mg crude). This peptide (109.0 mg, 0.11 mmol) was cyclised, Figure 128, and purified via preparative HPLC yielding a white solid (10.7 mg, 0.01 mmol, 10% yield). LR MS (ESI+) *m/z* (%) 508.9 (½$[M+2Na]^{2+}$, 100%), 978.2 ($[M+H]^{+}$, 91.93%), 1000.5 ($[M+Na]^{+}$, 60.04%). Analytical HPLC (280 nm) 20.0 min (98%).

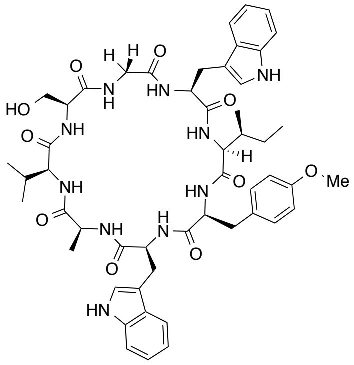
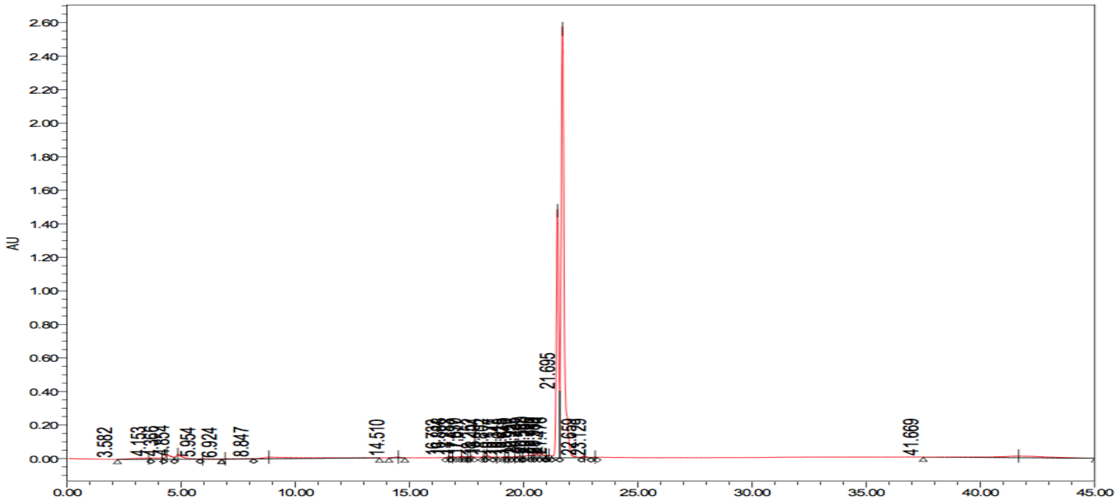
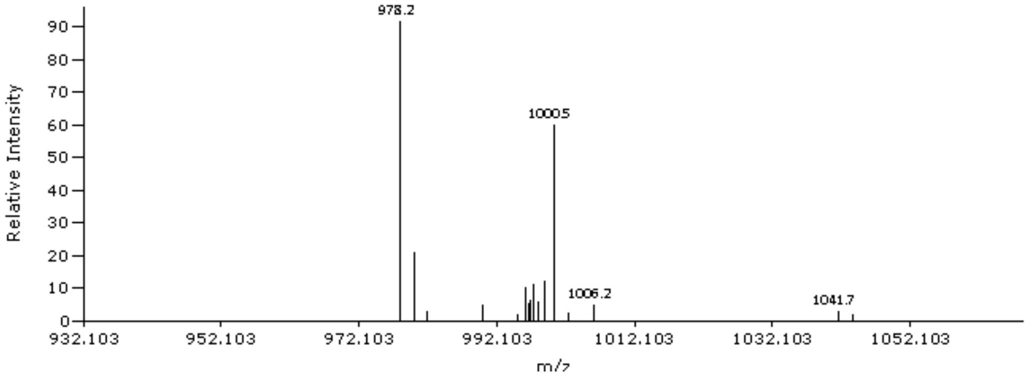


Figure 128: CP11A4Y8





8.6.2.9 CP11A4Y9

CP11A4Y9 (IPalWAVSGW) was cleaved from the resin to form the linear peptide (yellow solid, 171.3 mg crude). This peptide (117.0 mg, 0.12 mmol) was cyclised, Figure 129, and purified via preparative HPLC yielding a white solid (4.8 mg, 0.01 mmol, 4% yield). LR MS (ESI+) *m/z* (%) 329.9 (⅓[M+3Na]$^{3+}$, 100%), 475.5 (½[M+2Na]$^{2+}$, 56.55%), 970.8 ([M+Na]$^{+}$, 46.27%), 949.0 ([M+H]$^{+}$, 19.98%). Analytical HPLC (280 nm), due to low amount of material DMSO stock hence lower purity 16.0 min (85%). The first peak in the analytical HPLC trace is due to the presence of DMSO.

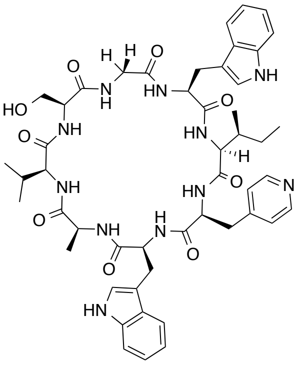
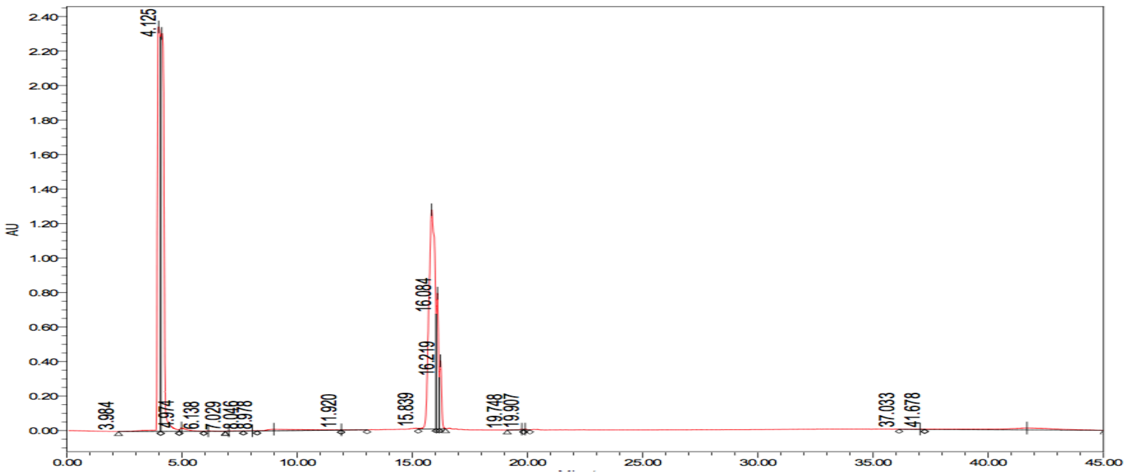
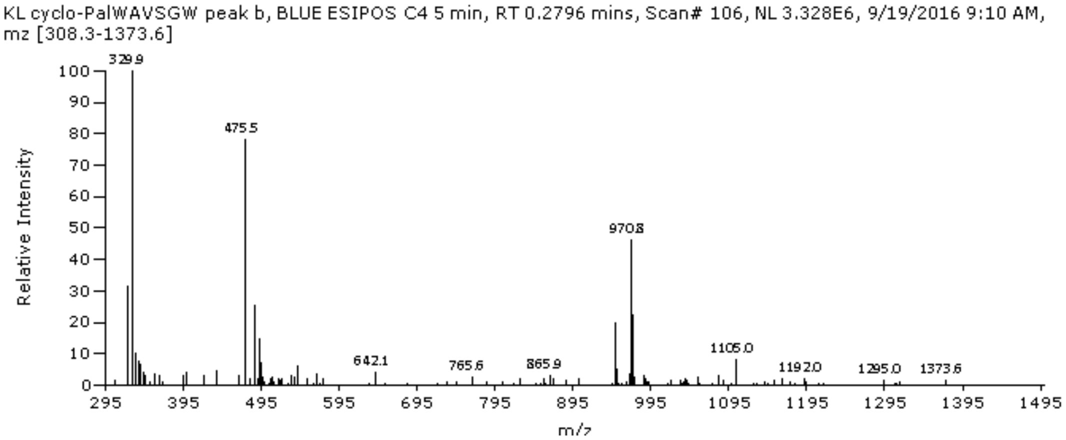


Figure 129: CP11A4Y9





8.6.2.10 CP11A4Y10

CP11A4Y10 (IChaWAVSGW) was cleaved from the resin to form the linear peptide (dark red solid, 385.3 mg crude). This peptide (120 mg, 0.12 mmol) was cyclised, Figure 130, and purified via preparative HPLC yielding a pale orange solid (16.3 mg, 0.17 mmol, 14% yield). LR MS (ESI+) *m/z* (%) 975.8 ($[M+Na]^+$, 100%), 953.7 ($[M+H]^+$, 85.05%), 477.7 ($½[M+2Na]^{2+}$, 34.60%). Analytical HPLC (280 nm) 23.6 min (92%).

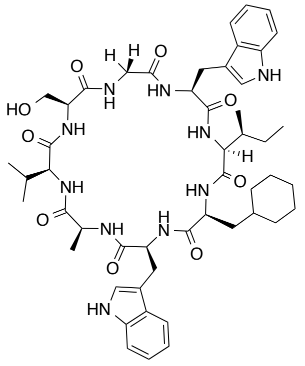
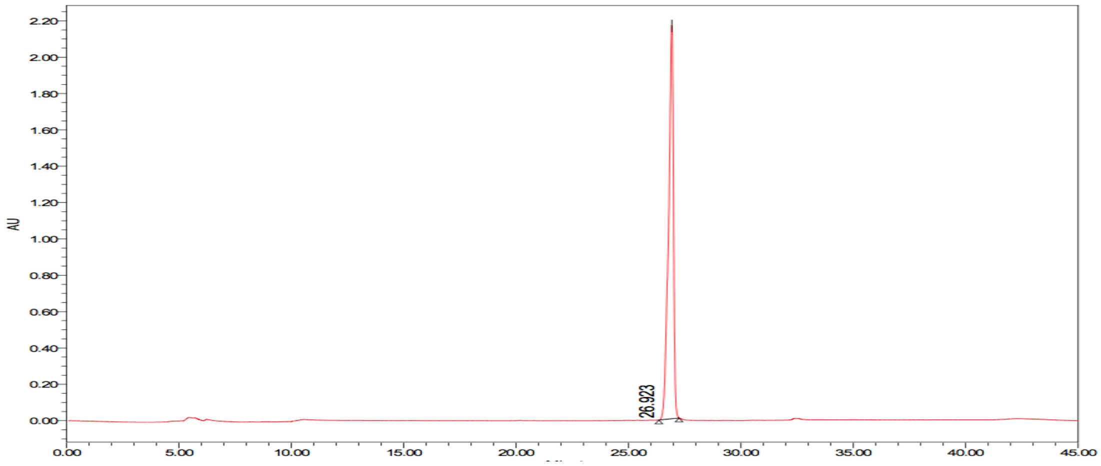
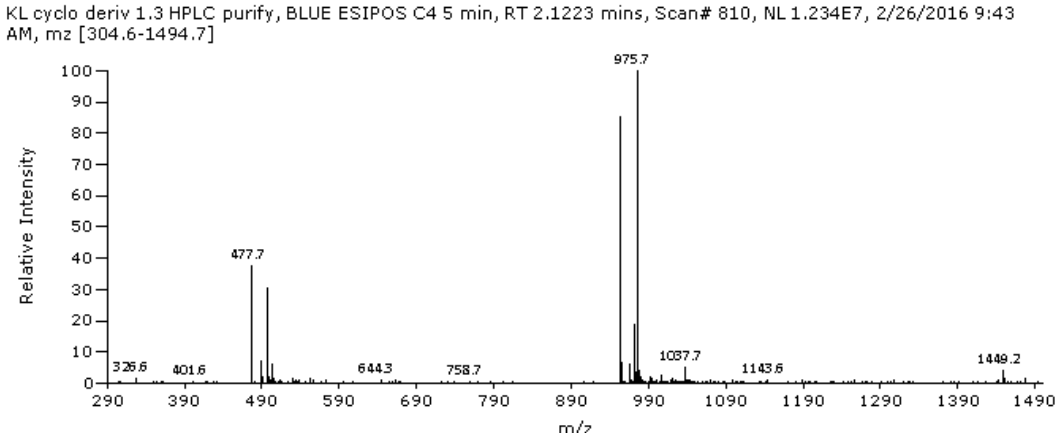


Figure 130: CP11A4Y10





8.6.2.11 CP11A4Y11

CP11A4Y11 (INalWAVSGW) was cleaved from the resin to form the linear peptide (pale orange solid, 108.7 mg crude). This peptide (108.7 mg, 0.107 mmol) was cyclised, Figure 131, and purified via preparative HPLC yielding a pale orange solid (14.2 mg, 0.014 mmol, 13% yield). LR MS (ESI+) *m/z* (%) 997.8 ($[M+H]^+$, 100%), 1019.8 ($[M+Na]^+$, 54.40%), 499.5 (½$[M+2Na]^{2+}$, 39.94%). Analytical HPLC (280 nm) 24.9 min (98%).

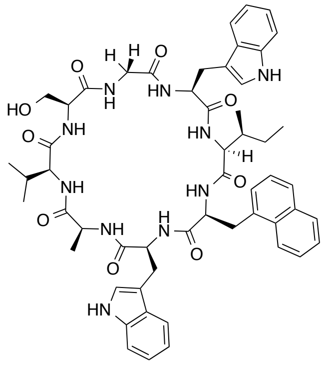
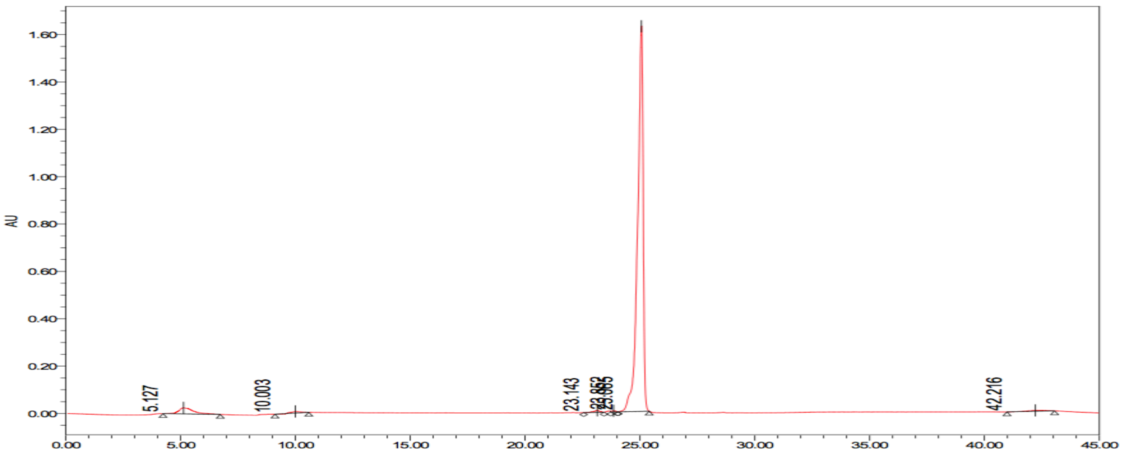
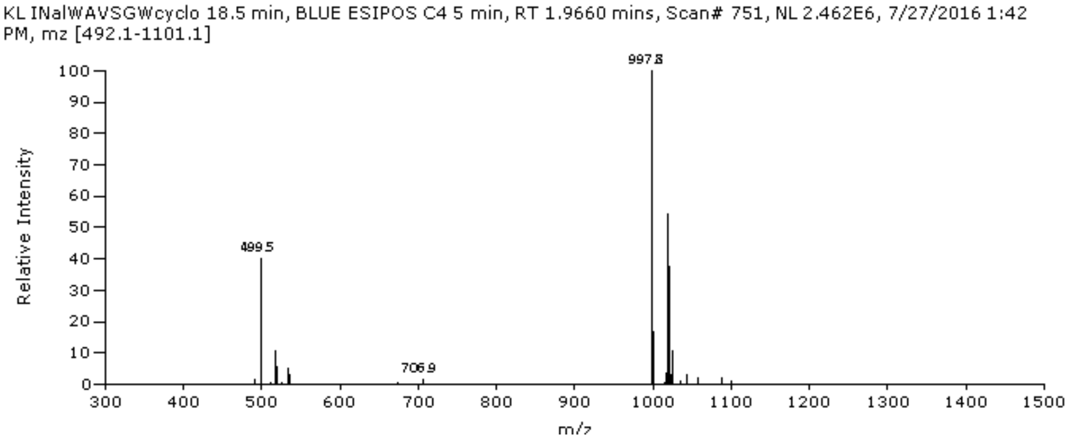


Figure 131: CP11A4Y11





8.6.3 CP11A5Y*

This series of cyclic peptides have the same backbone as CP11A5, however the tyrosine residue has been substituted for phenylalanine or tyrosine analogues. The analogues were synthesised in bulk on an Isoleucine Wang resin (3.3 mmol scale) as 8.2.1, until the second tryptophan residue. The resin was then split into 0.6 g batches, then the tyrosine analogues were coupled to complete the linear resin bound peptides.

8.6.3.1 CP11A5Y1

CP11A5Y1 (IFWNASGW) was cleaved from the resin to form the linear peptide (pale pink solid, 107.4 mg crude). This peptide (107.4 mg, 0.11 mmol) was cyclised, Figure 132, and purified via preparative HPLC yielding a white solid (18 mg, 0.02 mmol, 17% yield). LR MS (ESI+) *m/z* (%) 984.8 ($[M+Na]^+$, 100%), 962.9 ($[M+H]^+$, 97.14%), 482.3 (½$[M+2Na]^{2+}$, 47.95%). Analytical HPLC (280 nm) 20.2 min (95%).

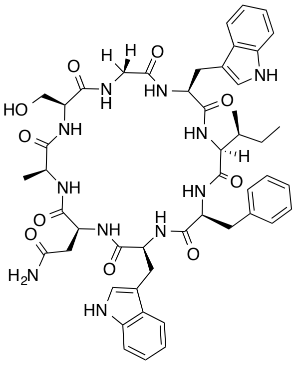
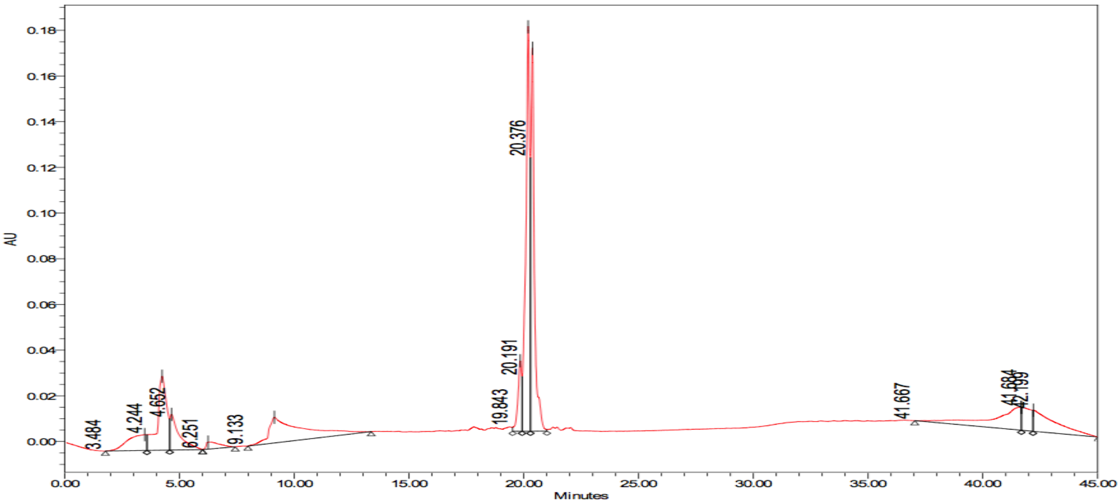
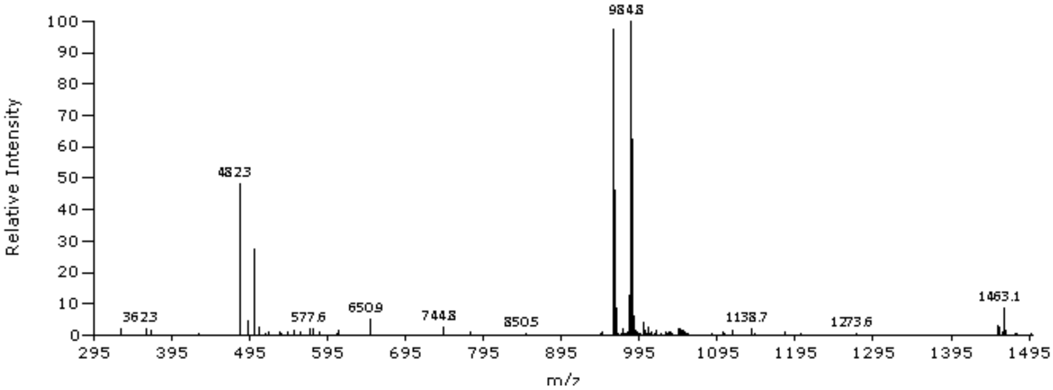


Figure 132: CP11A5Y1



KL cyclo IFWNASGW, BLUE ESIPOS C4 5 min, RT 1.6967 mins, Scan# 648, NL 5.519E6, 8/25/2016 3:07 PM, mz [330.5-1497.5]



8.6.3.2 CP11A5Y2

CP11A5Y2 (IF(4-F)WNASGW) was cleaved from the resin to form the linear peptide (pink solid, 70 mg crude). This peptide (70 mg, 0.07 mmol) was cyclised, Figure 133, and purified via preparative HPLC yielding a white solid (11.5 mg, 0.01 mmol, 17% yield). LR MS (ESI+) *m/z* (%) 1002.9 ($[M+Na]^+$, 100%), 491.3 (½$[M+2Na]^{2+}$, 63.03%), 980.8 ($[M+H]^+$, 59.04%). Analytical HPLC (280 nm) 20.9 min (89%).

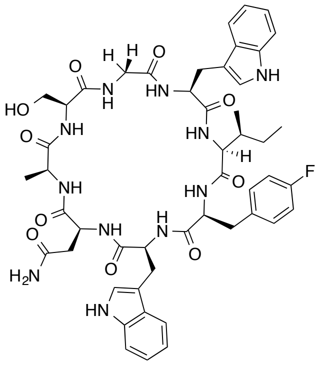
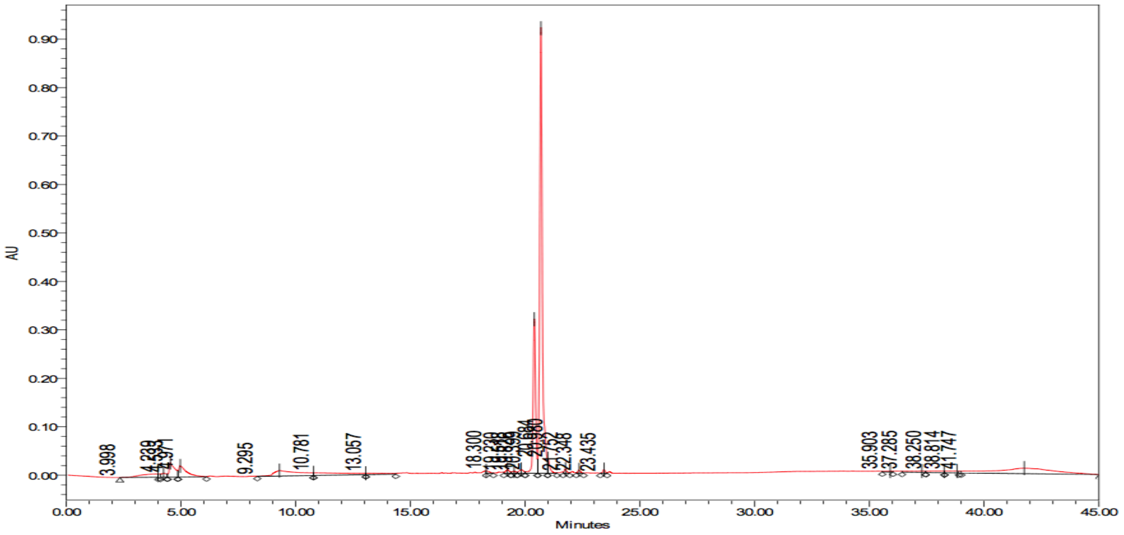
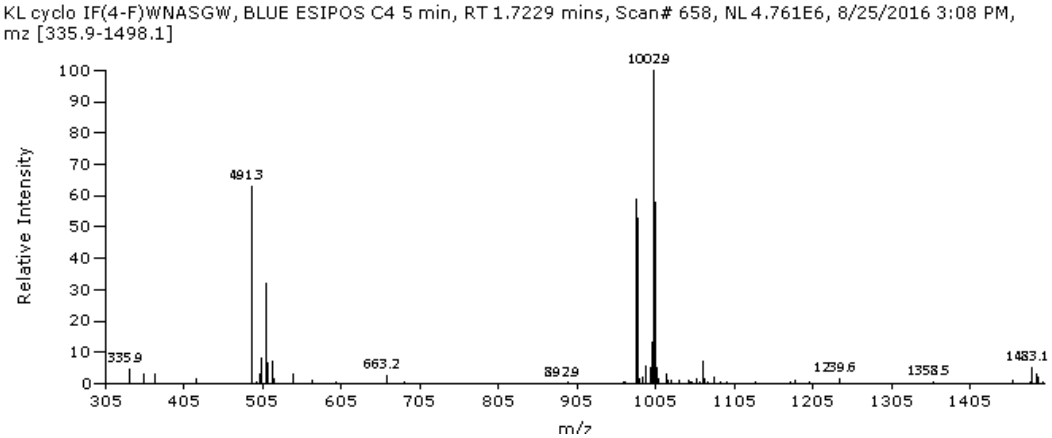


Figure 133: CP11A5Y2





8.6.3.3 CP11A5Y3

CP11A5Y3 (IF(4-Cl)WNASGW) was cleaved from the resin to form the linear peptide (orange solid, 272.2 mg crude). This peptide (115.3 mg, 0.11 mmol) was cyclised, Figure 134, and purified via preparative HPLC yielding a pale orange solid (18.3 mg, 0.018 mmol, 16% yield). LR MS (ESI+) *m/z* (%) 1018.6 ([M+Na]$^+$, 100%), 499.3 (½[M+2Na]$^{2+}$, 52.01%). Analytical HPLC (280 nm) 23.2 min (96%).

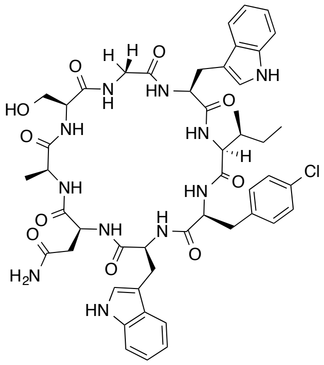
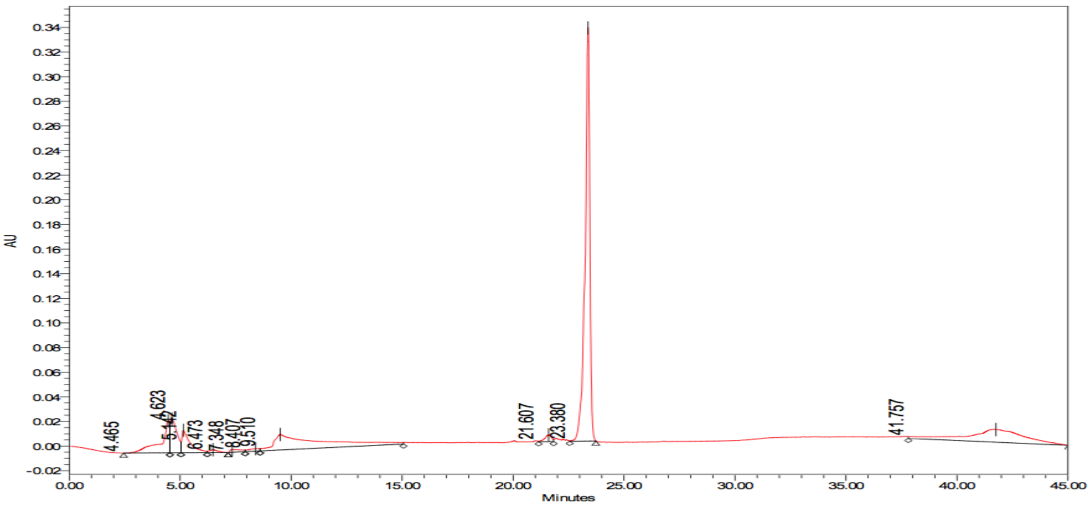
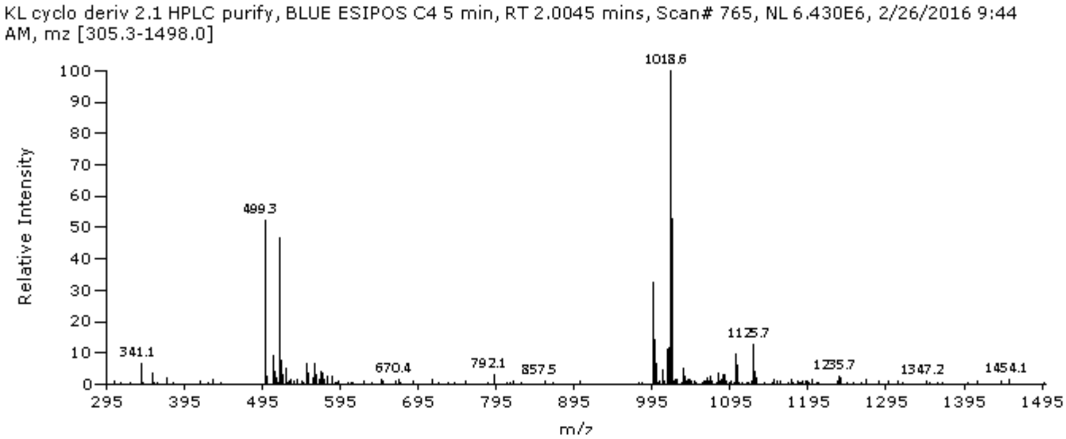


Figure 134: CP11A5Y3





8.6.3.4 CP11A5Y4

CP11A5Y4 (IF(4-Me)WNASGW) was cleaved from the resin to form the linear peptide (pale orange solid, 276.3 mg crude). This peptide (123.0 mg, 0.12 mmol) was cyclised, Figure 135, and purified via preparative HPLC yielding a pale cream solid (11.4 mg, 0.012 mmol, 10% yield). LR MS (ESI+) *m/z* (%) 998.6 ($[M+Na]^+$, 100%), 976.6 (½$[M+2Na]^{2+}$, 52.79%), 508.8 ($[M+H]^+$, 52.28%). Analytical HPLC (280 nm) 23.0 min (89%).

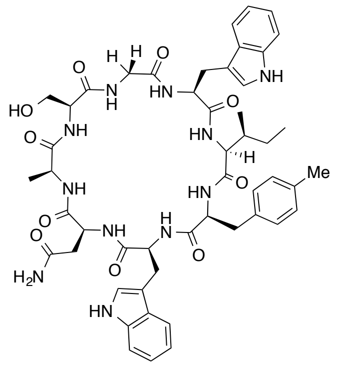
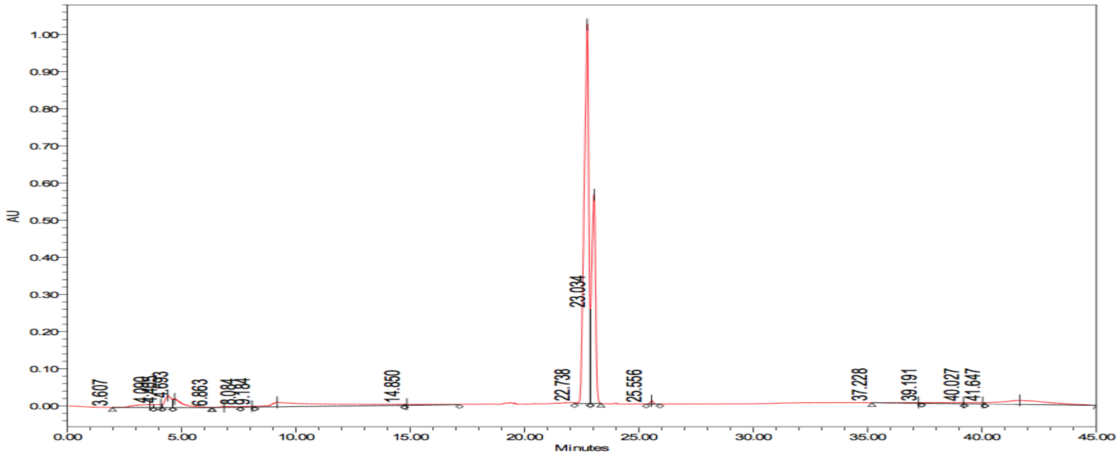
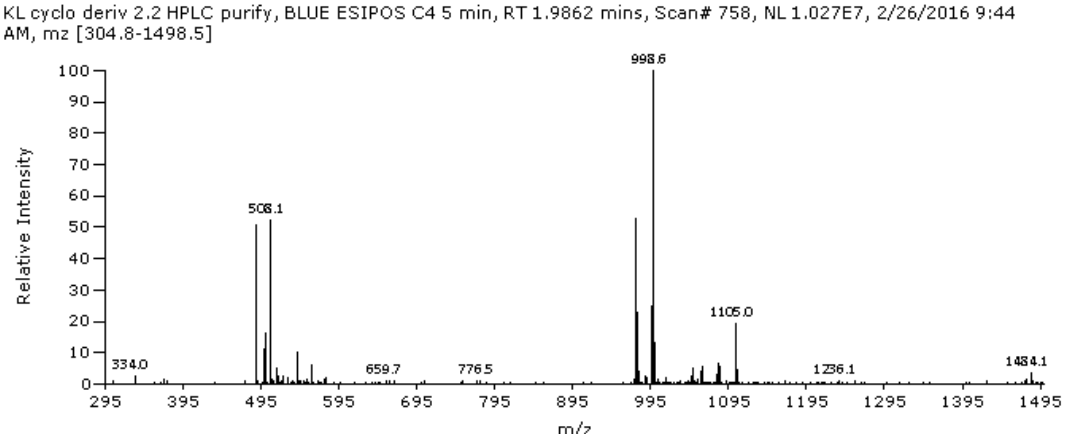


Figure 135: CP11A5Y4





8.6.3.5 CP11A5Y5

CP11A5Y5 (IF(4-CN)WNASGW) was cleaved from the resin to form the linear peptide (pale orange solid, 180.0 mg crude). This peptide (100.9 mg, 0.10 mmol) was cyclised, Figure 136, and purified via preparative HPLC yielding a pale pink solid (37.7 mg, 0.04 mmol, 38% yield). LR MS (ESI+) *m/z* (%) 994.9 ($[M+Na]^+$, 100%), 506.1 (½$[M+2Na]^{2+}$, 54.55%), 973.0 ($[M+H]^+$, 48.62%). Analytical HPLC (280 nm) 20.2 min (75%), due to low amount of material analytical HPLC run with DMSO stock, hence peak at 4.4 mins and appears lower purity. The first peak on the analytical HPLC trace is due to the presence of DMSO in the stock.

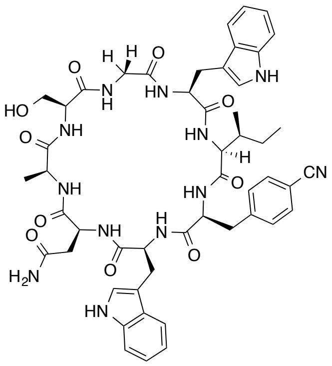
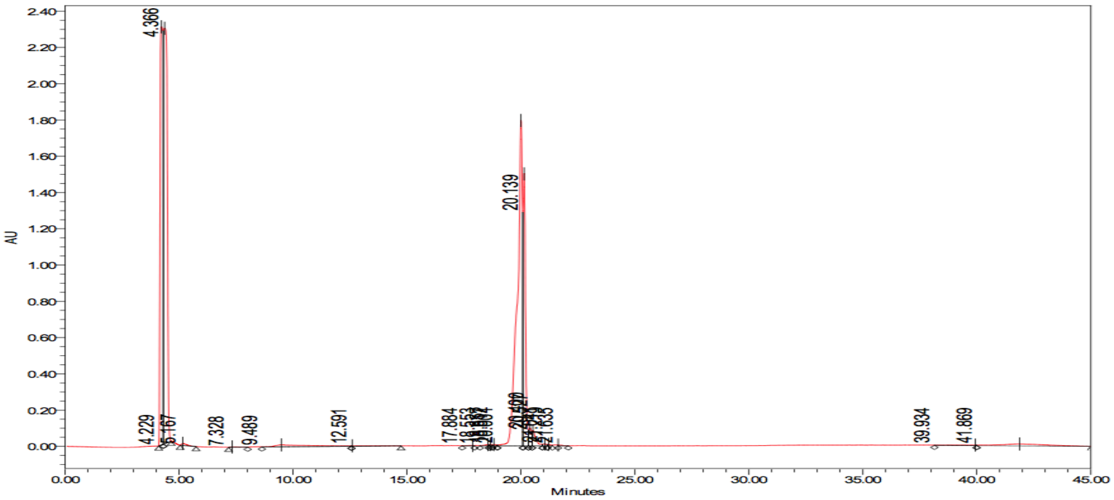
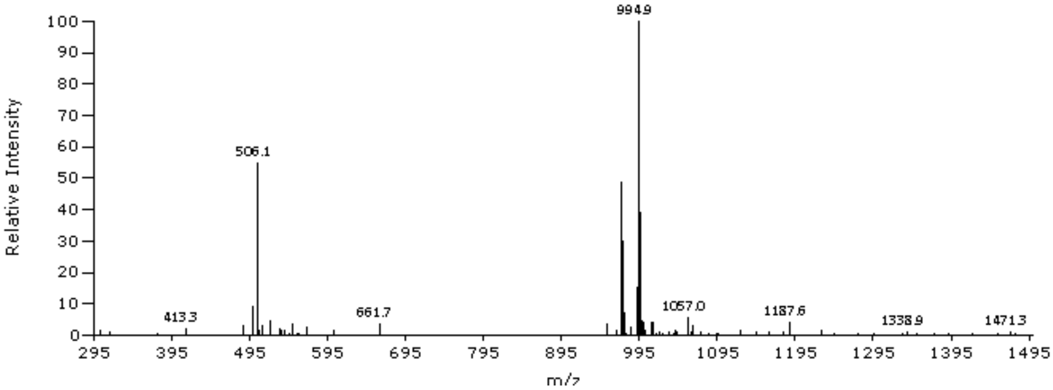


Figure 136: CP11A5Y5



KL cyclo IF(4-CN)WAVSGW, BLUE ESIPOS C4 5 min, RT 1.8118 mins, Scan# 692, NL 3.582E6, 8/30/2016 7:51 AM, mz [304.3-1477.5]



8.6.3.6 CP11A5Y6

CP11A5Y6 (IF(4-NO$_2$)WNASGW) was cleaved from the resin to form the linear peptide (yellow solid, 200.0 mg crude). This peptide (112.8 mg, 0.11 mmol) was cyclised, Figure 137, and purified via preparative HPLC yielding a pale yellow solid (39.4 mg, 0.04 mmol, 36% yield). LR MS (ESI+) *m/z* (%) 1029.9 ([M+H]$^+$, 100%), 1007.8 ([M+Na]$^+$, 69.83%), 523.6 (½[M+2Na]$^{2+}$, 35.13%). Analytical HPLC (280 nm) 20.5min (70%), due to low amount of material analytical HPLC run with DMSO stock, hence peak at 4.2 mins and appears lower purity. The first peak on the analytical HPLC trace is due to the presence of DMSO in the stock.

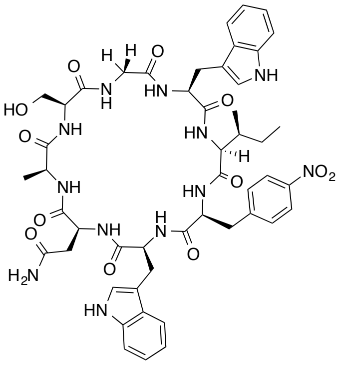
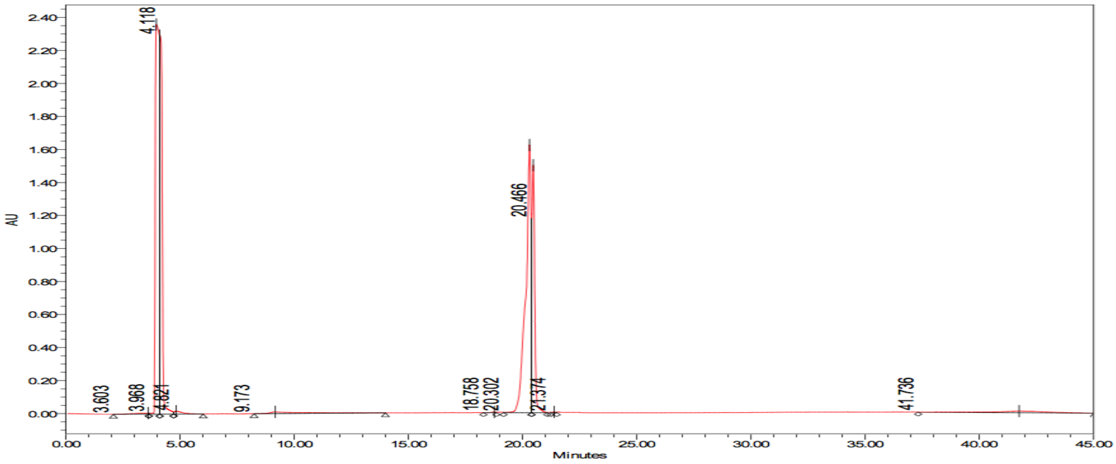
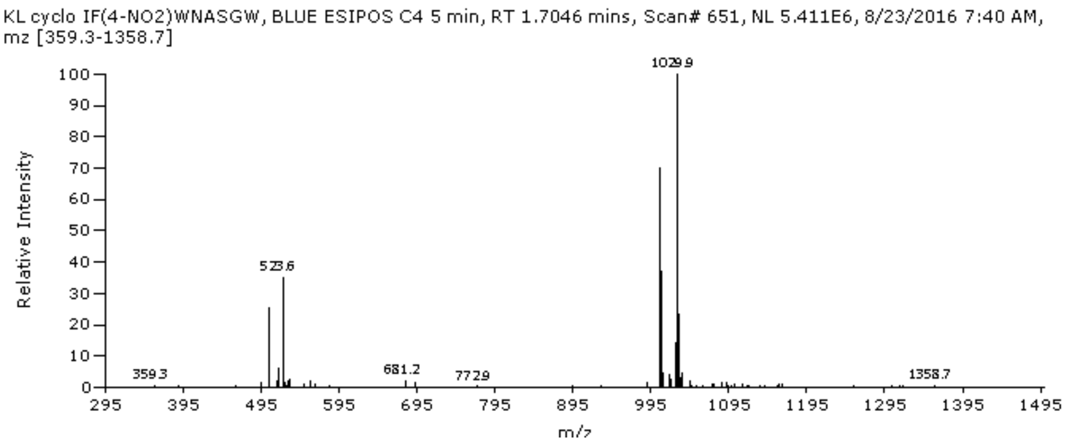


Figure 137: CP11A5Y6





8.6.3.7 CP11A5Y7

CP11A5Y7 (IY(3-NO$_2$)WNASGW) was cleaved from the resin to form the linear peptide (yellow solid, 85.9 mg crude). This peptide (85.9 mg, 0.08 mmol) was cyclised, Figure 138, and purified via preparative HPLC yielding a pale yellow solid (15.6 mg, 0.02 mmol, 19% yield). LR MS (ESI+) *m/z* (%) 1024.0 ($[M+H]^+$, 100%), 1046.7 ($[M+Na]^+$, 79.20%), 531.6 (½$[M+2Na]^{2+}$, 63.79%). Analytical (280 nm) 19.9 min (75%), due to low amount of material analytical HPLC run with DMSO stock, hence peak at 4.2 mins and appears lower purity.

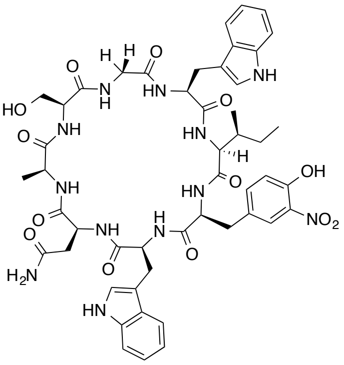
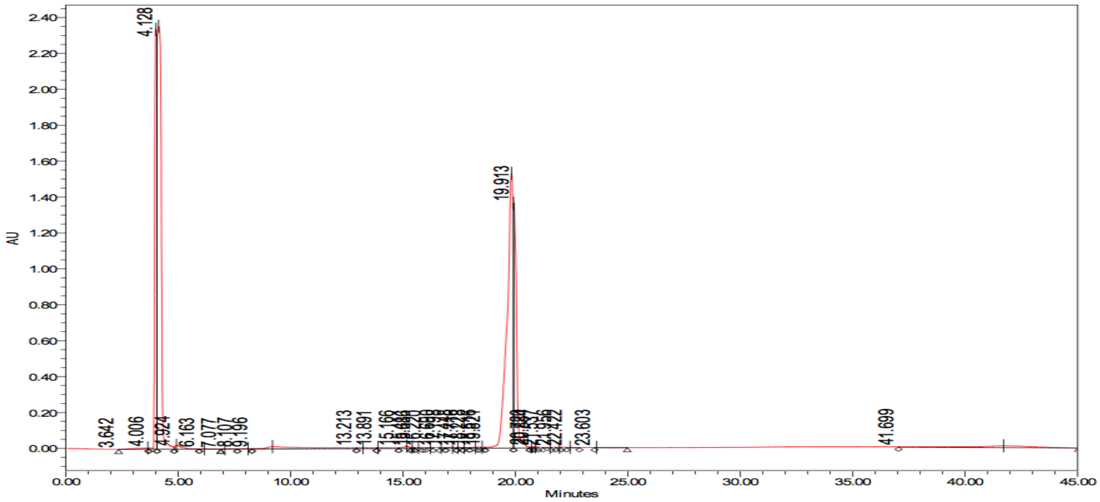
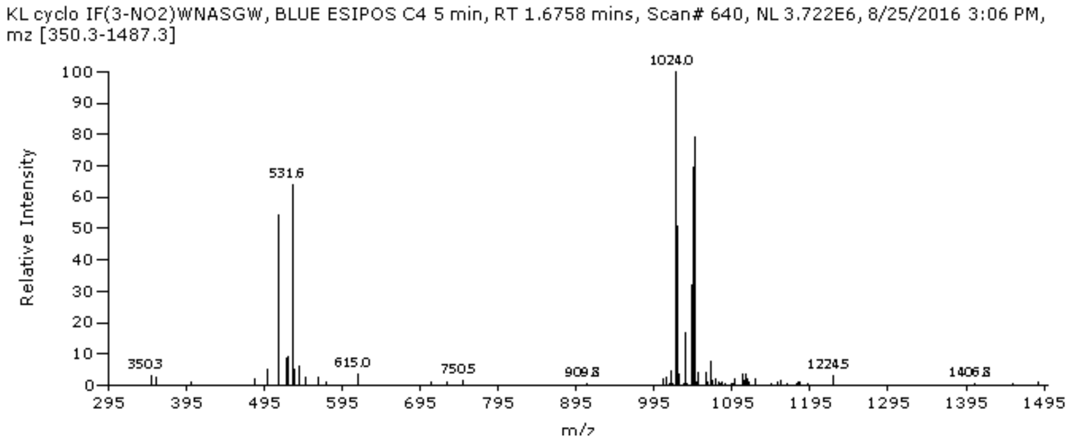


Figure 138: CP11A5Y7





8.6.3.8 CP11A5Y8

CP11A5Y8 (IY(Me)WNASGW) was cleaved from the resin to form the linear peptide (pale cream solid, 141.5 mg crude). This peptide (93.6 mg, 0.09 mmol) was cyclised, Figure 139, and purified via preparative HPLC yielding a cream solid (13.2 mg, 0.01 mmol, 14% yield). LR MS (ESI+) *m/z* (%) 1015.1 ($[M+Na]^+$, 100%), 993.1 ($[M+H]^+$, 48.16%), 497.2 (½$[M+2Na]^{2+}$, 44.25%). Analytical HPLC (280 nm) 20.1 min (92%).

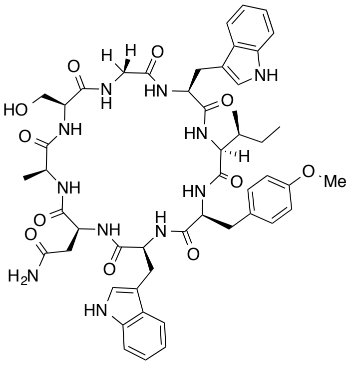
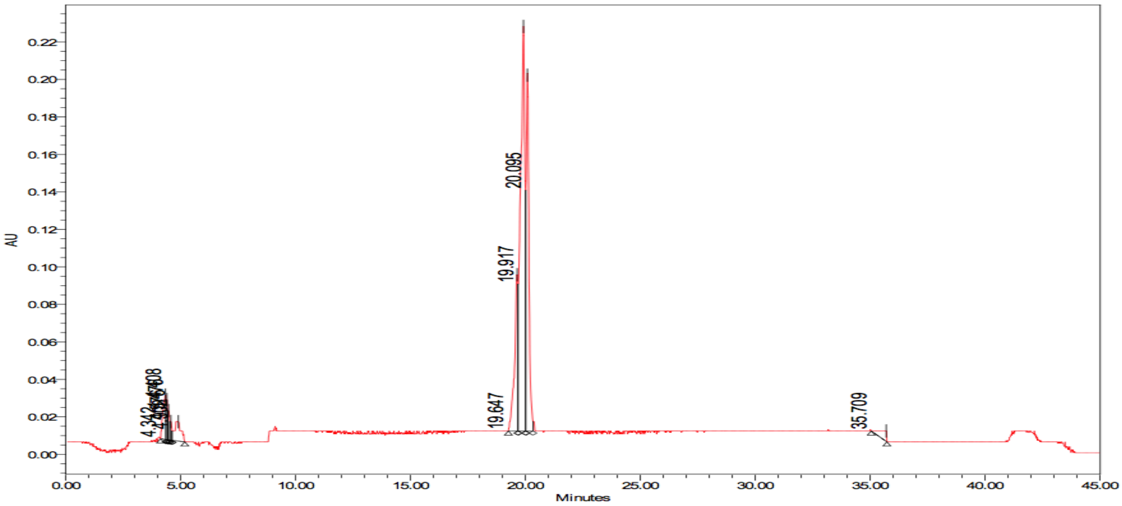
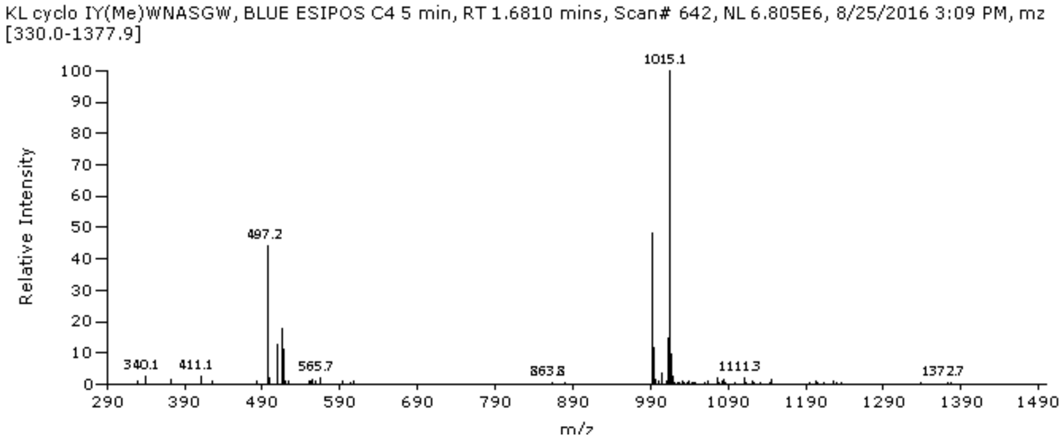


Figure 139: CP11A5Y8





8.6.3.9 CP11A5Y9

CP11A5Y9 (IPalWNASGW) was cleaved from the resin to form the linear peptide (yellow solid, 398.7 mg crude). This peptide (111.1 mg, 0.11 mmol) was cyclised, Figure 140, and purified via preparative HPLC yielding a cream solid (12.8 mg, 0.01 mmol, 12% yield). LR MS (ESI+) *m/z* (%) 330.0 (⅓[M+3Na]$^{3+}$, 100%), 475.1 (½[M+2Na]$^{2+}$, 64.59%), 971.0 ([M+Na]$^{+}$, 36.13%), 948.8 ([M+H]$^{+}$, 21.84%). There wasn’t enough material to run an analytical HPLC to determine purity, however purity from UV trace (280 nm) of LR MS estimated (89%).

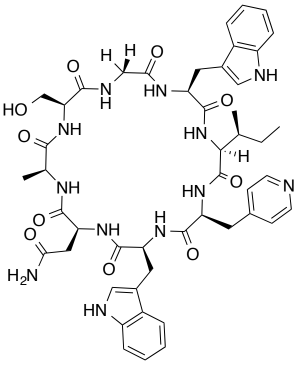
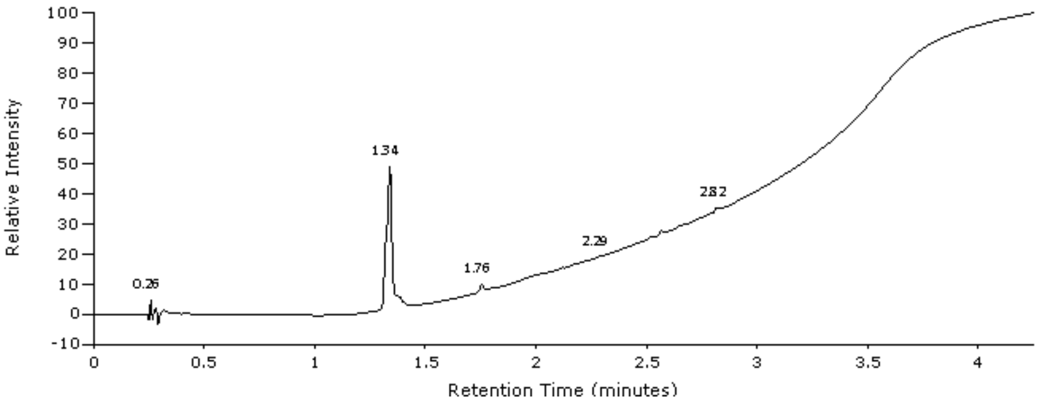
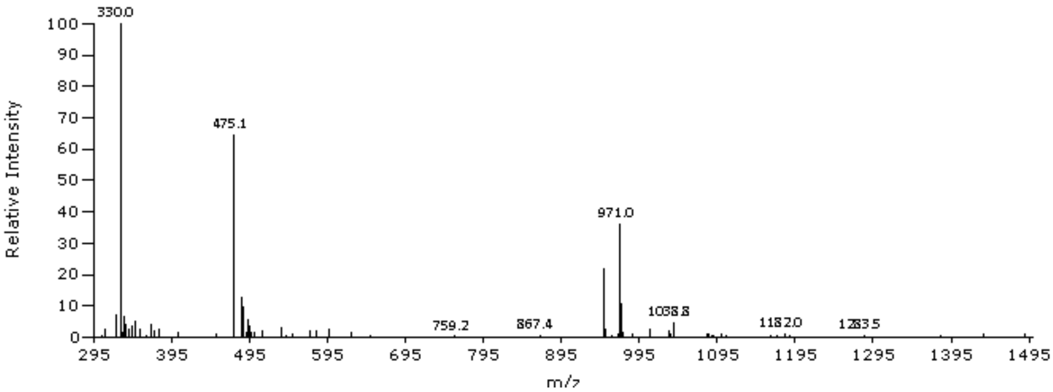


Figure 140: CP11A5Y9

UV Absorbance, KL cyclo IPalWAVSGW, BLUE ESIPOS C4 5 min, S9, NL 2.448E4, 8/30/2016 7:53 AM



KL cyclo IPalWAVSGW, BLUE ESIPOS C4 5 min, RT 1.3464 mins, Scan# 514, NL 4.789E6, 8/30/2016 7:53 AM, mz [304.9-1490.2]



8.6.3.10 CP11A5Y10

CP11A5Y10 (IChaWNASGW) was cleaved from the resin to form the linear peptide (pale orange solid, 318.6 mg crude). This peptide (123 mg, 0.12 mmol) was cyclised, Figure 141, and purified via preparative HPLC yielding a pale orange solid (11.6 mg, 0.017 mmol, 13% yield). LR MS (ESI+) *m/z* (%) 990.8 ($[M+Na]^+$, 100%), 485.2 (½$[M+Na]^{2+}$, 47.00%), 968.8 ($[M+H]^+$, 37.45%). Analytical HPLC (280 nm) 23.6 min (92%).

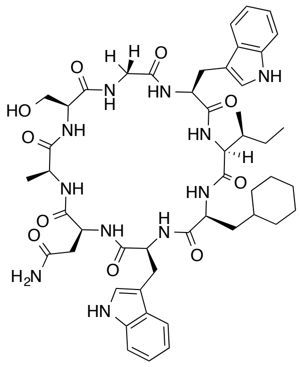
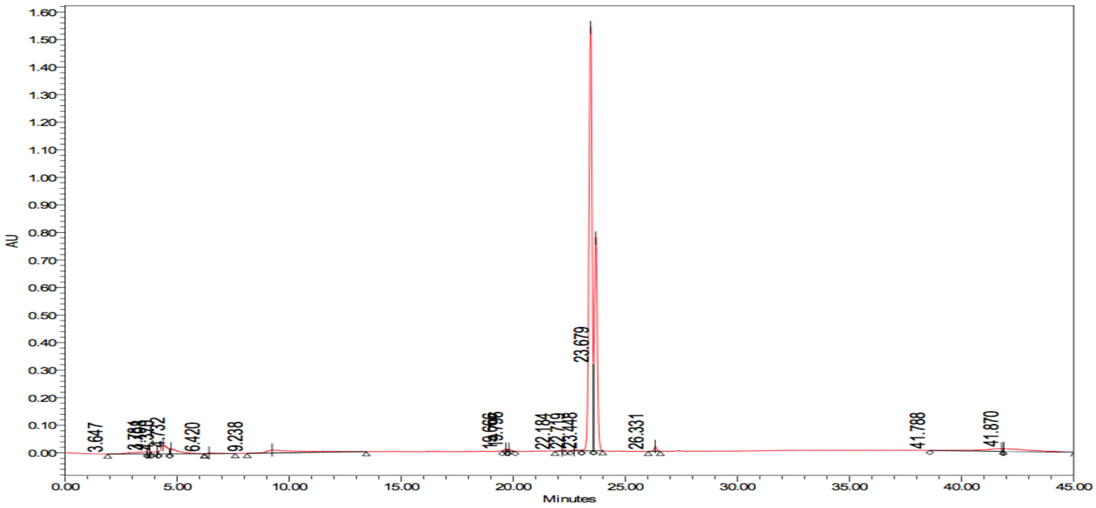
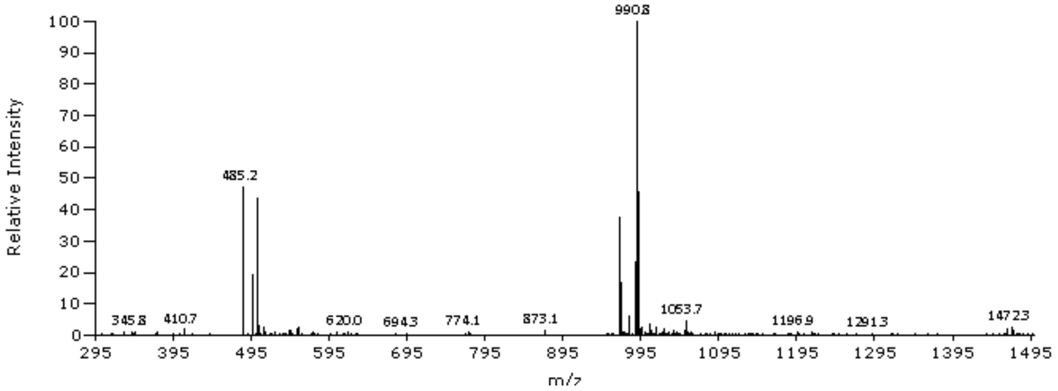


Figure 141: CP11A5Y10



KL cyclo deriv 2.3 HPLC purify, BLUE ESIPOS C4 5 min, RT 2.0124 mins, Scan# 768, NL 1.206E7, 2/26/2016 9:45 AM, mz [303.0-1498.3]



8.6.3.11 CP11A5Y11

CP11A5Y11 (INalWNASGW) was cleaved from the resin to form the linear peptide (yellow solid, 255.5 mg crude). This peptide (118.0 mg, 0.11 mmol) was cyclised, Figure 142, and purified via preparative HPLC yielding a pale orange solid (17.5 mg, 0.017 mmol, 16% yield). LR MS (ESI+) *m/z* (%) 1035.4 ($[M+Na]^+$, 100%), 1013.5 ($[M+H]^+$, 39.83%), 507.2 (½$[M+Na]^{2+}$, 36.35%). There wasn't enough material to run an analytical HPLC to determine purity, however purity from UV trace (280 nm) of LR MS estimated (91%).

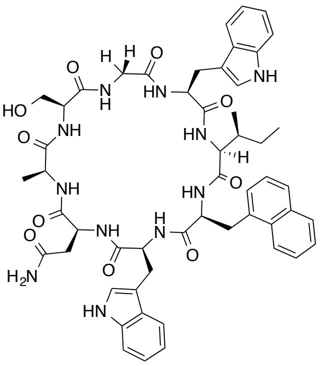
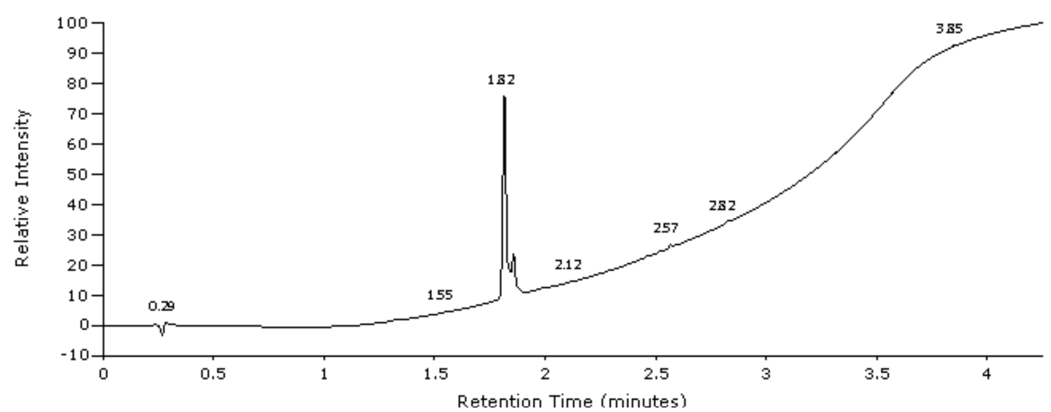
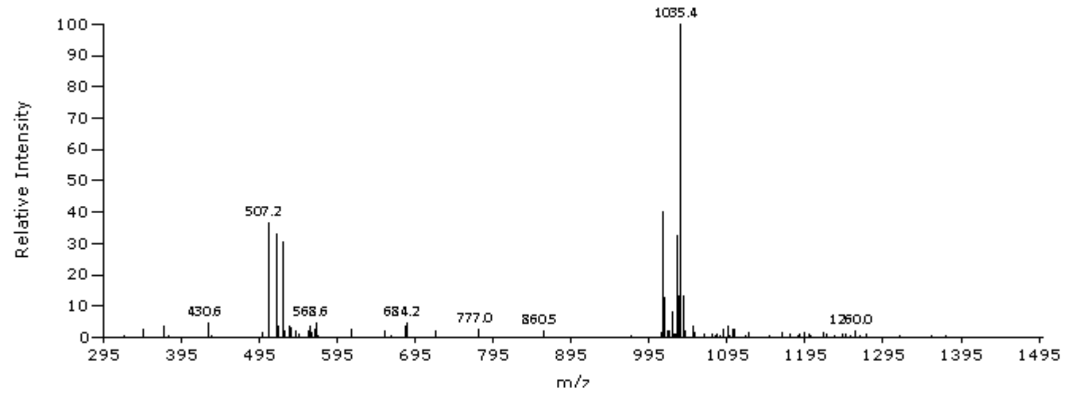


Figure 142: CP11A5Y11

UV Absorbance, KL cyclo INalWNASGW, BLUE ESIPOS C4 5 min, S9, NL 2.613E4, 9/22/2016 2:27 PM



KL cyclo INalWNASGW, BLUE ESIPOS C4 5 min, RT 1.8301 mins, Scan# 699, NL 3.473E6, 9/22/2016 2:27 PM, mz [322.6-1376.4]



8.6.4 CP11A8Y*

This series of cyclic peptides have the same backbone as CP11A8, however the tyrosine residue has been substituted for phenylalanine or tyrosine analogues. The analogues were synthesised in bulk on a Glycine Wang resin (3 mmol scale) as 8.2.1, until the first tryptophan residue. The resin was then split into 0.4 g batches, then the tyrosine analogues were coupled and then the isoleucine and alanine were coupled to complete the linear resin bound peptides.

8.6.4.1 CP11A8Y1

CP11A8Y1 (IFWNVSGA) was cleaved from the resin to form the linear peptide (purple solid, 147.0 mg crude). This peptide (100.0 mg, 0.11 mmol) was cyclised, Figure 143, and purified via preparative HPLC yielding a pale pink solid (22.7 mg, 0.026 mmol, 23% yield). LR MS (ESI+) *m/z* (%) 875.9 ([M+H]$^+$, 100%), 897.7 ([M+Na]$^+$, 26.33%), 438.5 (½[M+Na]$^{2+}$, 20.97%). Analytical HPLC (280 nm) 18.9 min (96%).

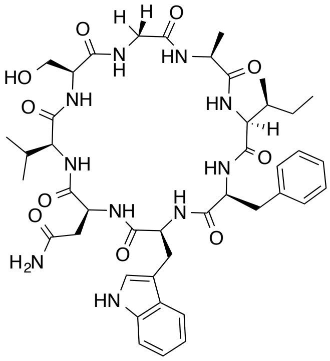
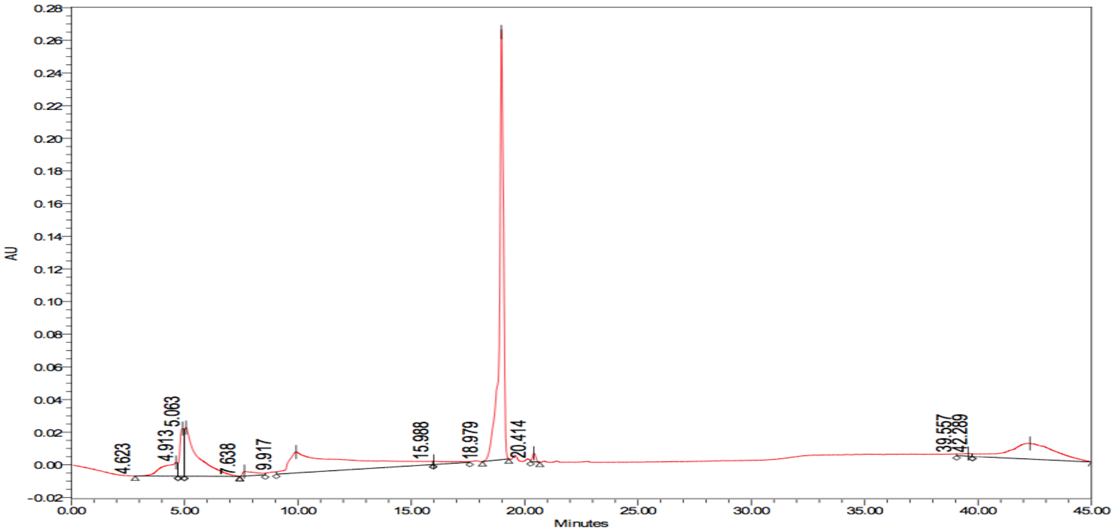
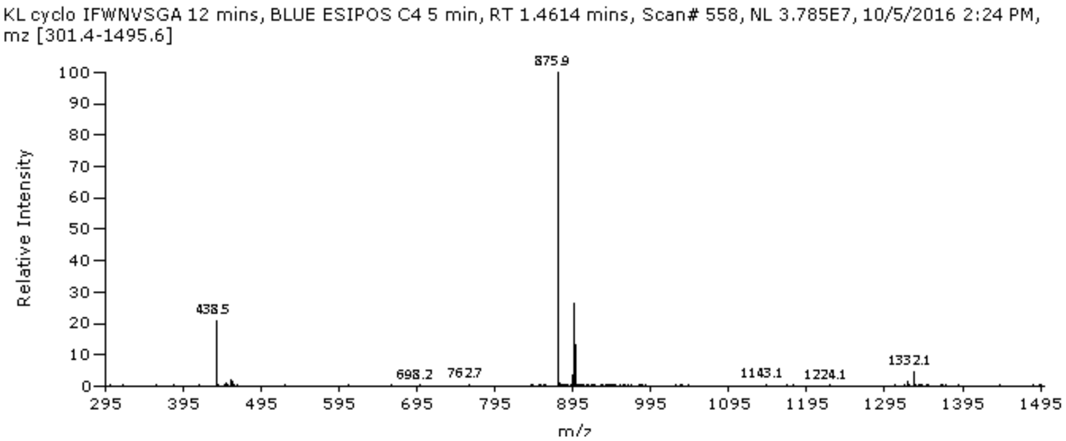


Figure 143: CP11A8Y1




KL cyclo IFWNVSGA 12 mins, BLUE ESIPOS C4 5 min, RT 1.4614 mins, Scan# 558, NL 3.785E7, 10/5/2016 2:24 PM, mz [301.4-1495.6]

8.6.4.2 CP11A8Y2

CP11A8Y2 (IF(4-F)WNNSGA) was cleaved from the resin to form the linear peptide (purple solid, 189.4 mg crude). This peptide (100.0 mg, 0.11 mmol) was cyclised, Figure 144, and purified via preparative HPLC yielding a pale pink solid (38.7 mg, 0.043 mmol, 39% yield). LR MS (ESI+) *m/z* (%) 915.9 ($[M+Na]^+$, 100%), 893.6 ($[M+H]^+$, 40.80%), 447.3 (½$[M+H]^{2+}$, 36.25%). Analytical HPLC (280 nm) 20.1 min (96%).

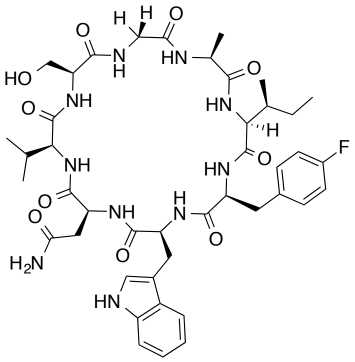
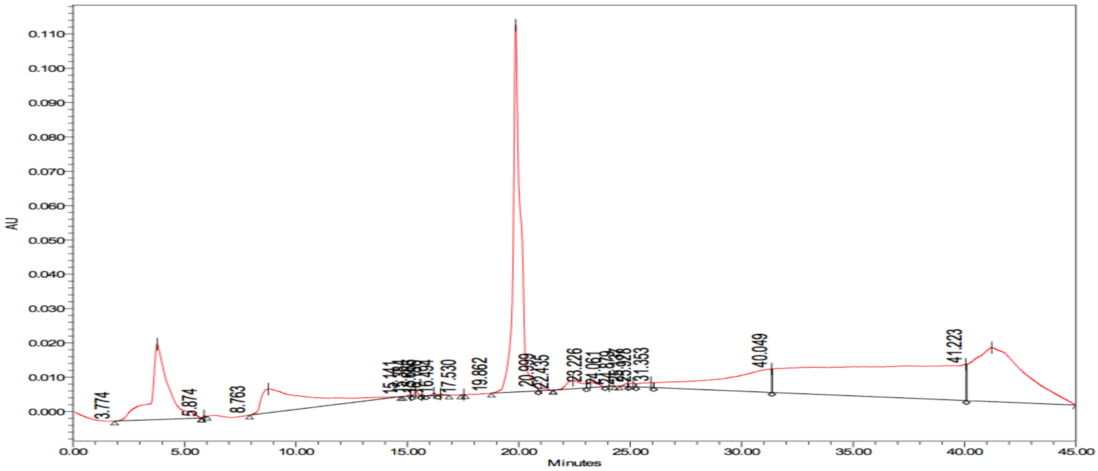
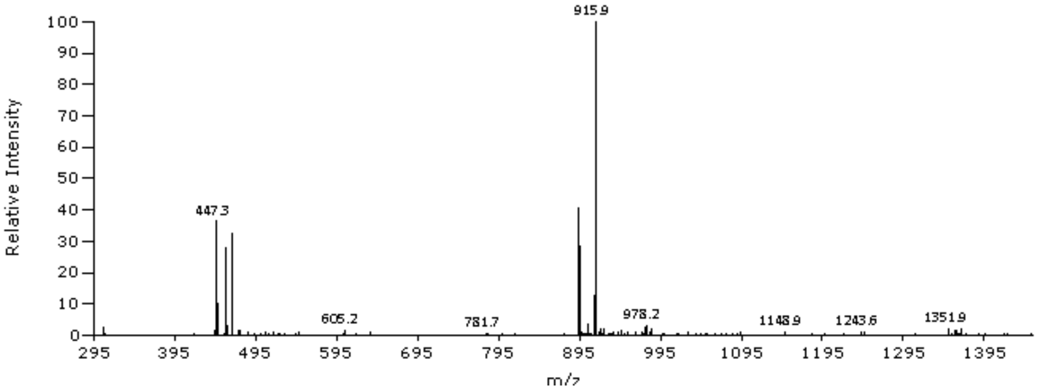


Figure 144: CP11A8Y2



KL cyclo IF(4-F)WNVSGA, BLUE ESIPOS C4 5 min, RT 1.5007 mins, Scan# 573, NL 9.371E6, 10/10/2016 7:37 AM, mz [306.7-1452.2]



8.6.4.3 CP11A8Y3

CP11A8Y3 (IF(4-Cl)WNVSGA) was cleaved from the resin to form the linear peptide (purple solid, 189.4 mg crude). This peptide (100.0 mg, 0.11 mmol) was cyclised, Figure 145, and purified via preparative HPLC yielding a cream solid (35.4 mg, 0.040 mmol, 35% yield). LR MS (ESI+) *m/z* (%) 909.8 ($[M+H]^+$, 100%), 931.7 ($[M+Na]^+$, 65.59%), 455.7 ($½[M+Na]^{2+}$, 43.11%). Analytical HPLC (280 nm) 19.1 min (97%).

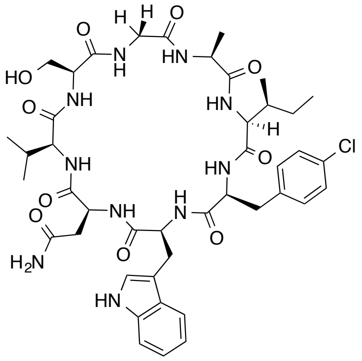
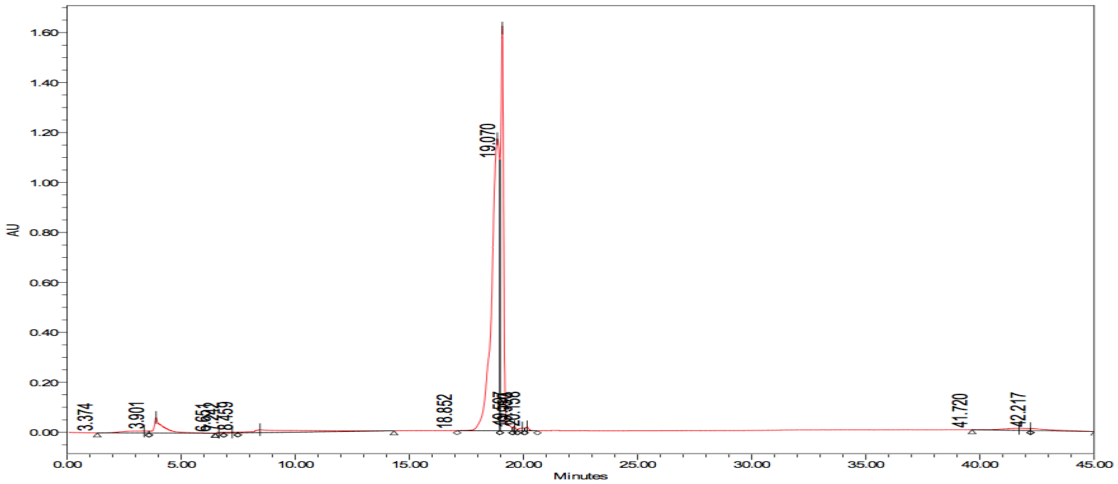
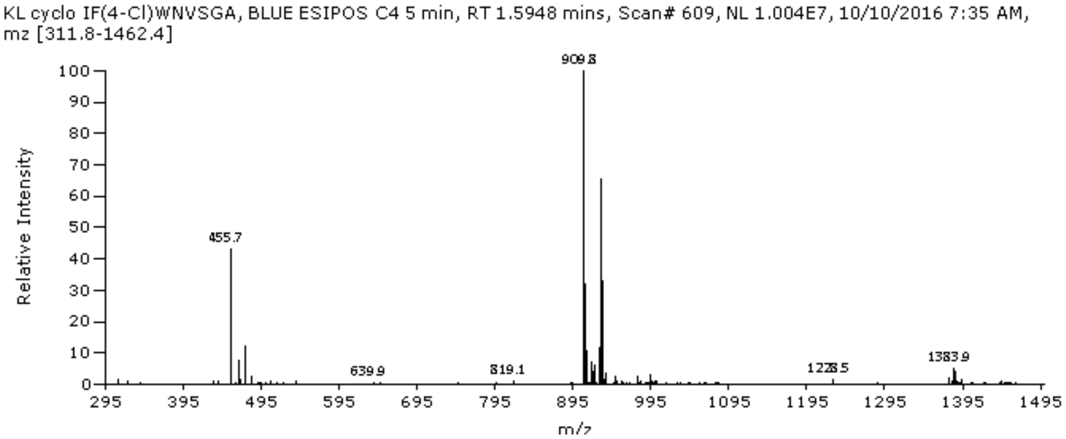


Figure 145: CP11A8Y3





8.6.4.4 CP11A8Y4

CP11A8Y4 (IF(4-Me)WNVSGA) was cleaved from the resin to form the linear peptide (pale pink solid, 212.1 mg crude). This peptide (100.0 mg, 0.11 mmol) was cyclised, Figure 146, and purified via preparative HPLC yielding a white solid (49.6 mg, 0.056 mmol, 50% yield). LR MS (ESI+) *m/z* (%) 911.7 ($[M+Na]^+$, 100%), 889.7 ($[M+H]^+$, 83.22%), 445.5 (½$[M+H]^{2+}$, 74.42%). Insufficient material to run analytical HPLC trace so purity assessed by UV trace (280 nm) of LR MS trace (80%).

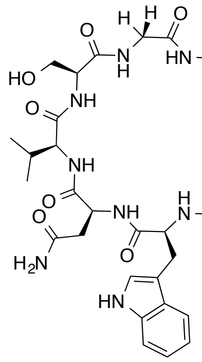
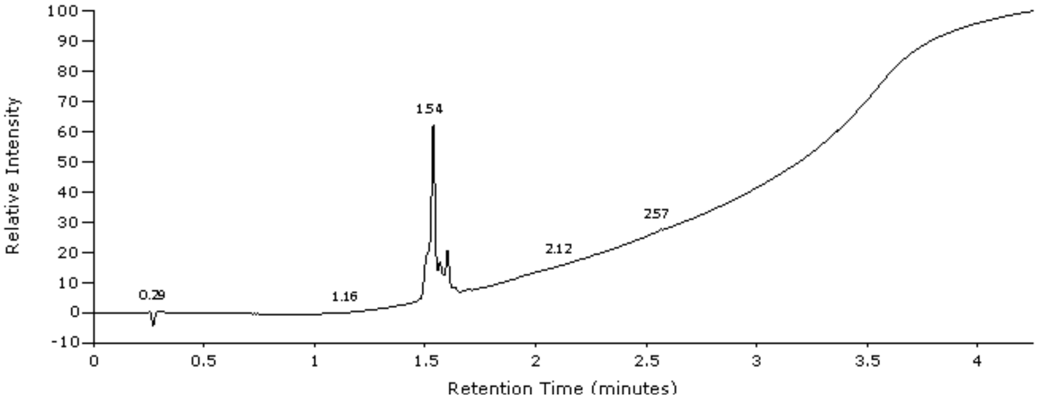
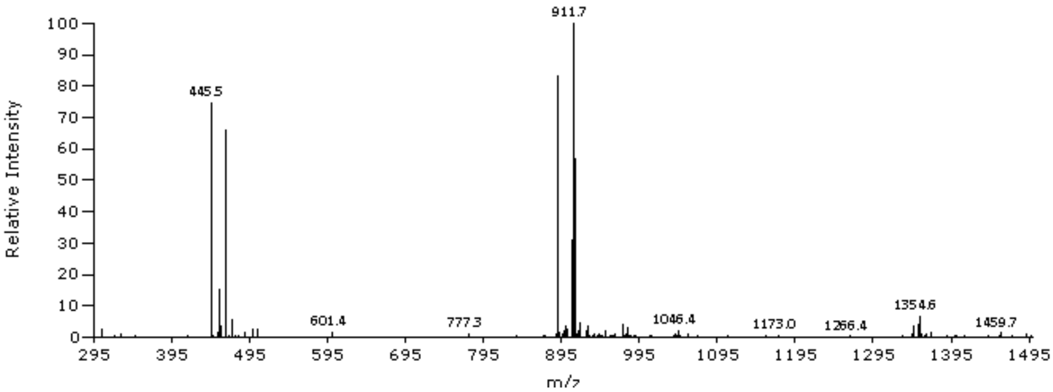


Figure 146: CP11A8Y4

UV Absorbance, KL cyclo IF(4-Me)WNVSGA, BLUE ESIPOS C4 5 min, S9, NL 2.561E4, 10/10/2016 7:37 AM



KL cyclo IF(4-Me)WNVSGA, BLUE ESIPOS C4 5 min, RT 1.5477 mins, Scan# 591, NL 7.245E6, 10/10/2016 7:37 AM, mz [305.6-1498.0]



8.6.4.5 CP11A8Y5

CP11A8Y5 (IF(4-CN)WNVSGA) was cleaved from the resin to form the linear peptide (pale pink solid, 173.5 mg crude). This peptide (102.1 mg, 0.11 mmol) was cyclised, Figure 147, and purified via preparative HPLC yielding a white solid (42.4 mg, 0.047 mmol, 43% yield). LR MS (ESI+) *m/z* (%) 922.9 ([M+Na]$^+$, 100%), 900.8 ([M+H]$^+$, 56.42%), 470.1 (½[M+H]$^{2+}$, 28.45%). Analytical HPLC (280 nm) 17.8 min (98%).

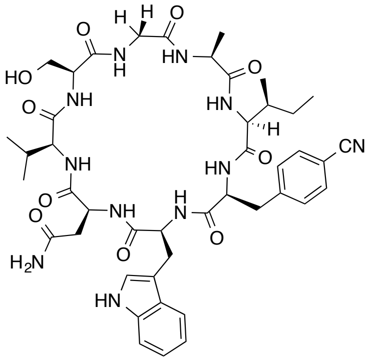
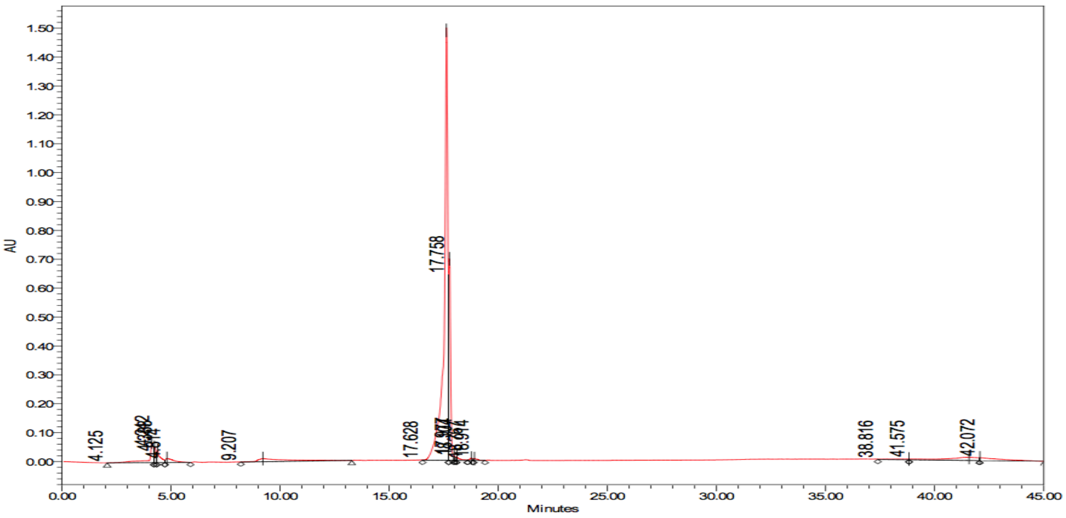
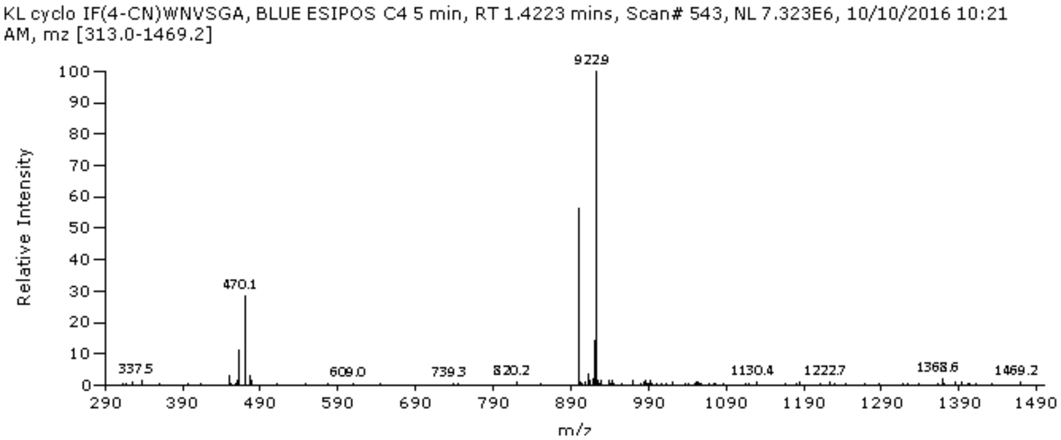


Figure 147: CP11A8Y5





8.6.4.6 CP11A8Y6

CP11A8Y6 (IF(4-NO_2)WNVSGA) was cleaved from the resin to form the linear peptide (purple solid, 143.2 mg crude). This peptide (103.2 mg, 0.11 mmol) was cyclised, Figure 148, and purified via preparative HPLC yielding a cream solid (39.8 mg, 0.043.3 mmol, 39% yield). LR MS (ESI+) *m/z* (%) 942.6 ($[M+Na]^+$, 100%), 920.60 ($[M+H]^+$, 92.46%), 470.1 (½$[M+H]^{2+}$, 40.98%). Analytical HPLC (280 nm) min (%). Analytical HPLC (280 nm) 19.8 min (98%).

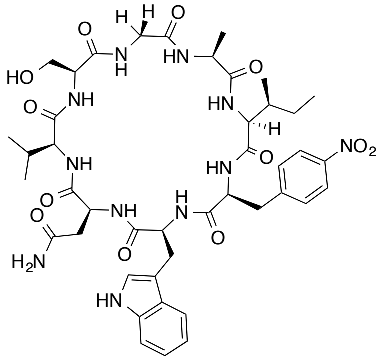
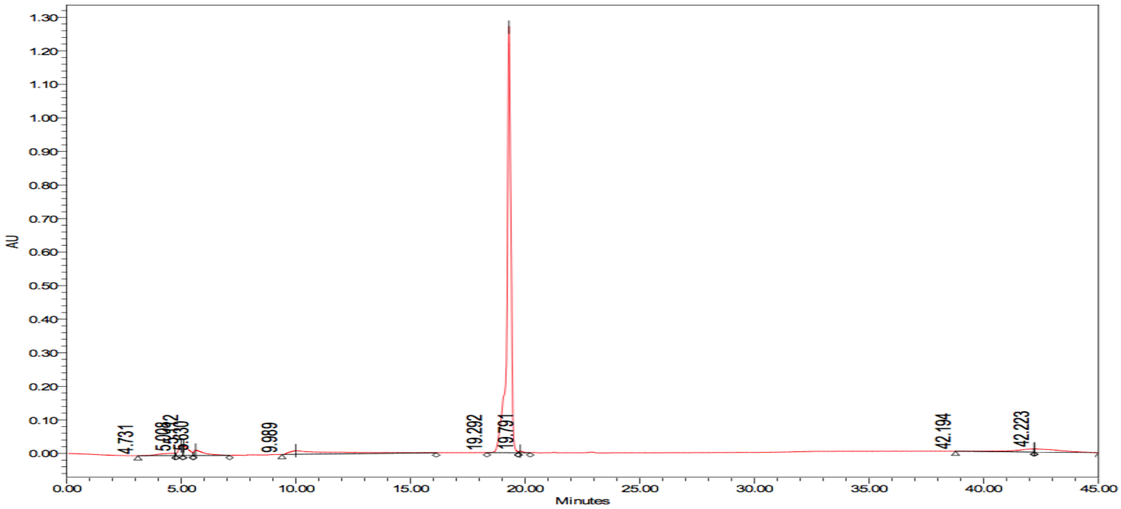
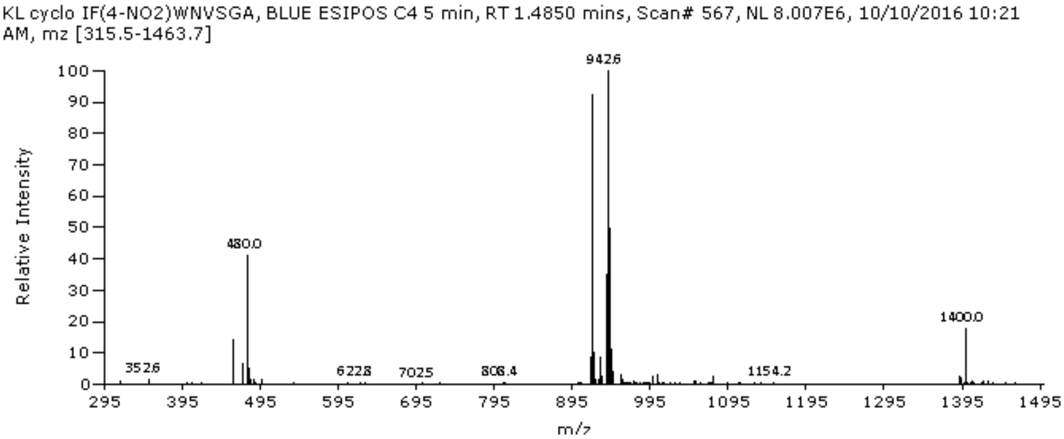


Figure 148: CP11A8Y6





8.6.4.7 CP11A8Y7

CP11A8Y7 (IY(3-NO_2)WNVSGA) was cleaved from the resin to form the linear peptide (yellow solid, 215.2 mg crude). This peptide (104.6 mg, 0.11 mmol) was cyclised, Figure 149, and purified via preparative HPLC yielding a yellow solid (24 mg, 0.026 mmol, 24% yield). LR MS (ESI+) *m/z* (%) 958.8 ($[M+Na]^+$, 100%), 936.8 ($[M+H]^+$, 56.97%), 488.5 (½$[M+H]^{2+}$, 29.25%). Analytical HPLC (280 nm) 19.6 min (95%).

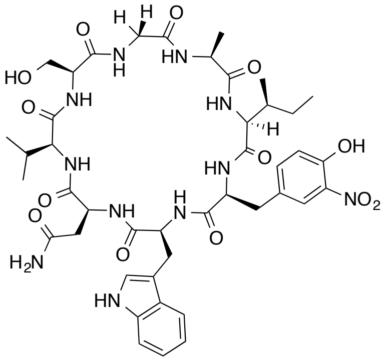
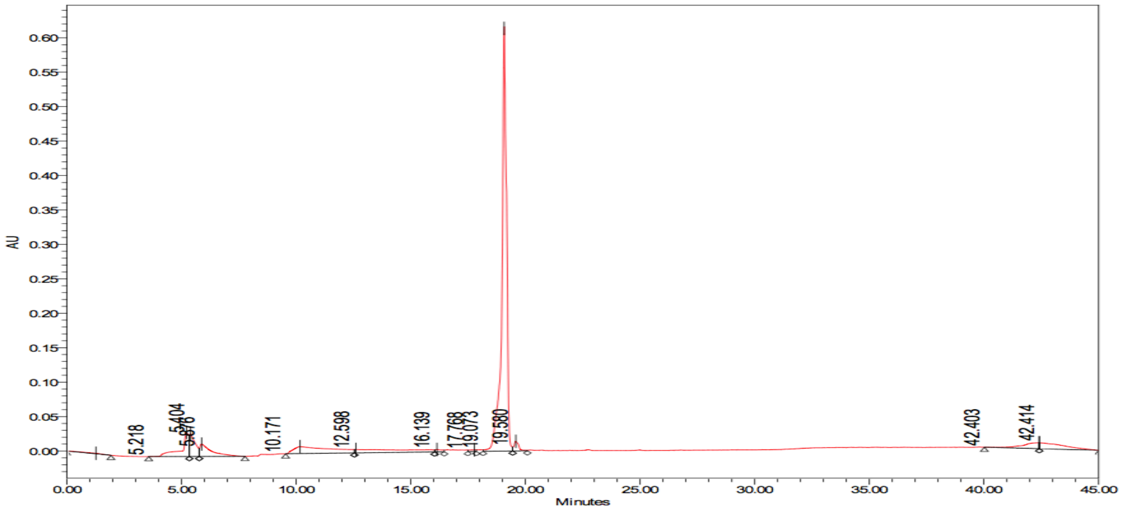
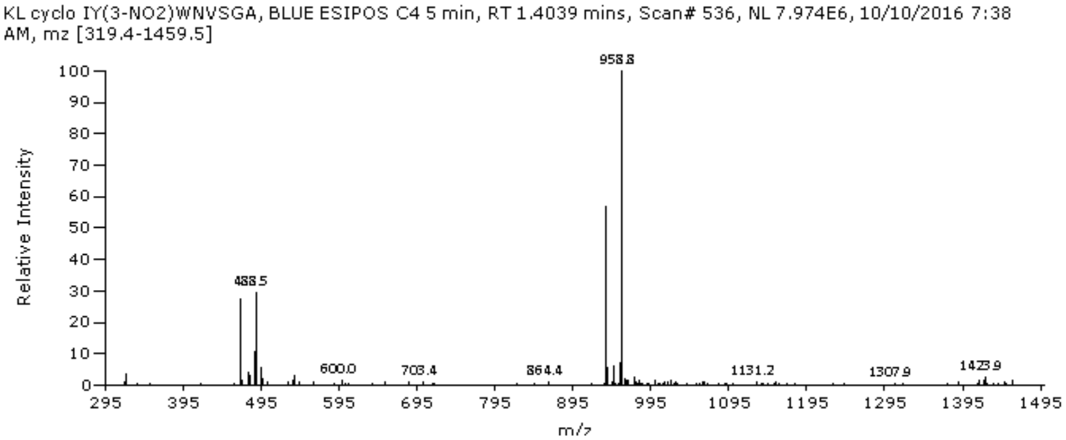


Figure 149: CP11A8Y7




KL cyclo IY(3-NO2)WNVSGA, BLUE ESIPOS C4 5 min, RT 1.4039 mins, Scan# 536, NL 7.974E6, 10/10/2016 7:38 AM, mz [319.4-1459.5]

8.6.4.8 CP11A8Y8

CP11A8Y8 (IY(Me)WNVSGA) was cleaved from the resin to form the linear peptide (pale pink solid, 181.6 mg crude). This peptide (104.2 mg, 0.11 mmol) was cyclised, Figure 150, and purified via preparative HPLC yielding a white solid (44.3 mg, 0.049 mmol, 43% yield). LR MS (ESI+) *m/z* (%) 927.5 ([M+Na]$^+$, 100%), 453.7 (½[M+H]$^{2+}$, 66.90%), 905.7 ([M+H]$^+$, 39.81%). Analytical HPLC (280 nm) 17.8 min (94%).

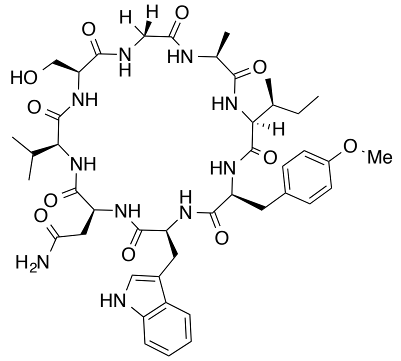
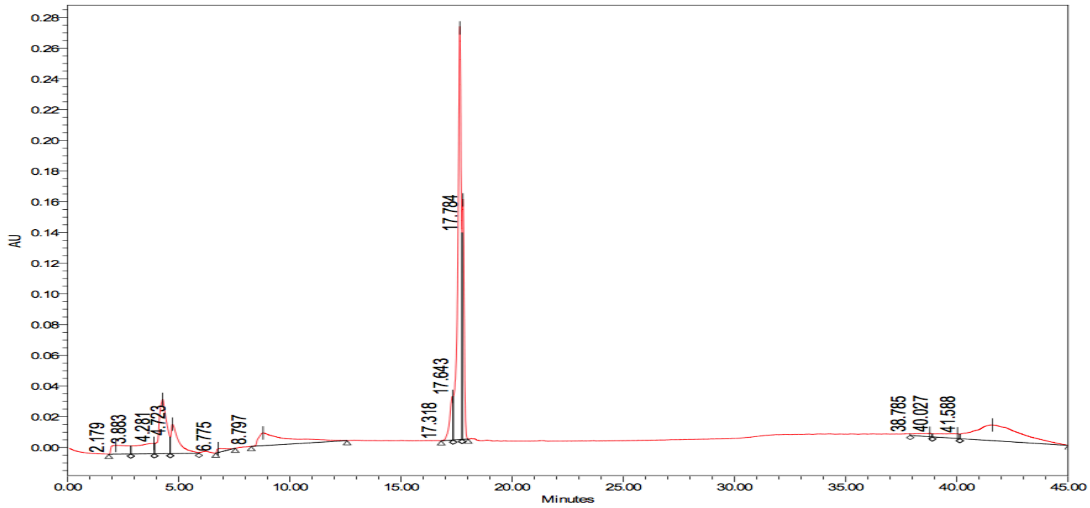
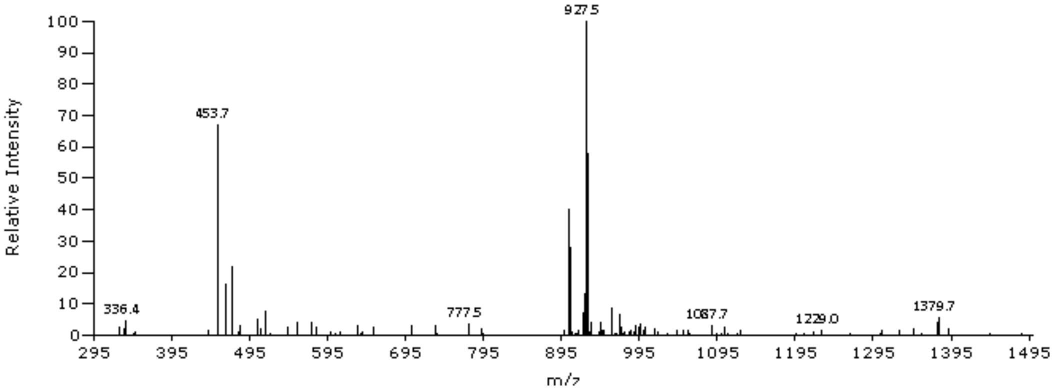


Figure 150: CP11A8Y8



KL cyclo IY(ME)WNVSGA, BLUE ESIPOS C4 5 min, RT 1.4405 mins, Scan# 550, NL 3.118E6, 10/10/2016 10:22 AM, mz [328.0-1485.9]



8.6.4.9 CP11A8Y9

CP11A8Y9 (IPalWNVSGA) was cleaved from the resin to form the linear peptide (cream solid, 194.0 mg crude). This peptide (102.1 mg, 0.11 mmol) was cyclised, Figure 151, and purified via preparative HPLC yielding a white solid (5.0 mg, 0.006 mmol, 5% yield). LR MS (ESI+) *m/z* (%) 306.1 (⅓[M+2Na]$^{3+}$, 100%), 898.7 ([M+Na]$^{+}$, 60.39%), 438.8 (½[M+H]$^{2+}$, 55.40%). Analytical HPLC (280 nm) 17.4 min (89%).

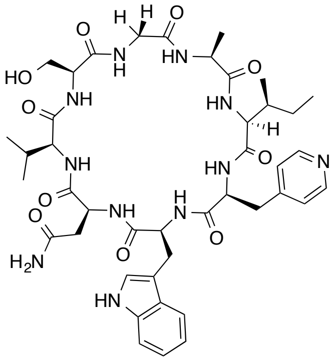
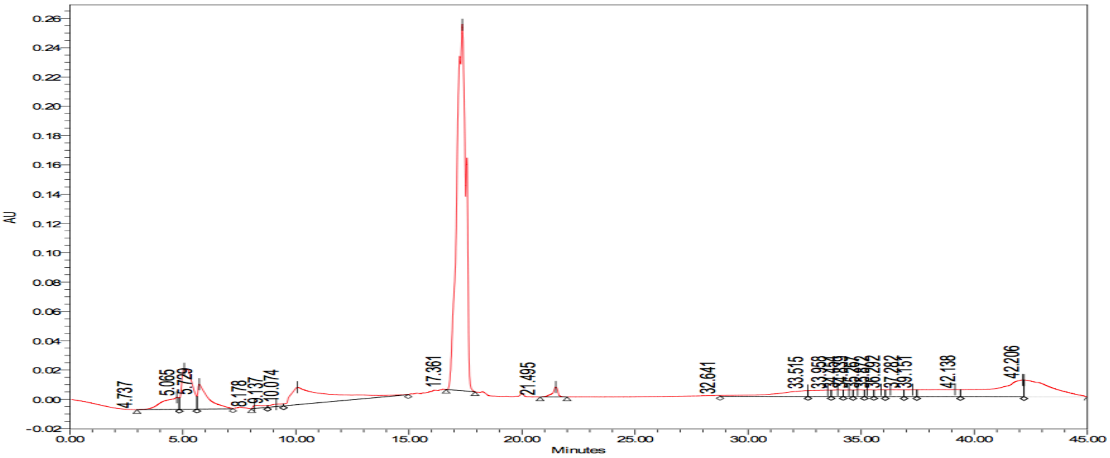
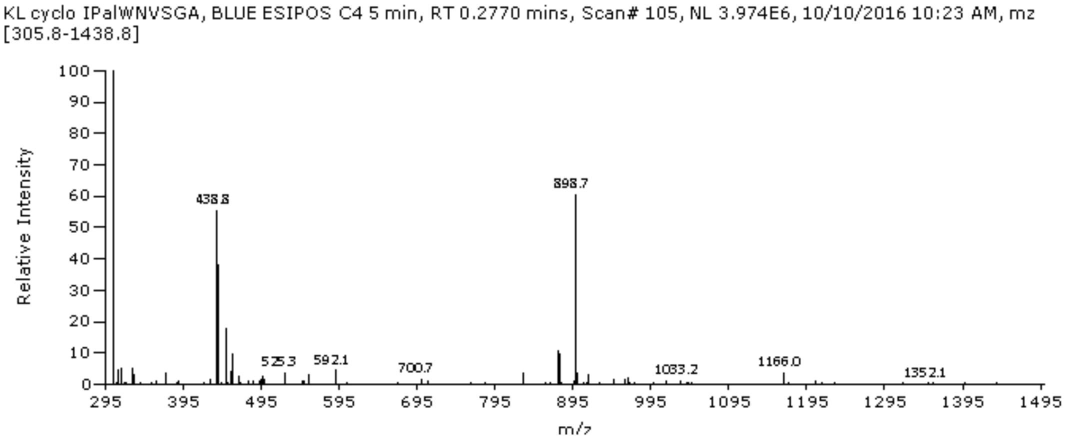


Figure 151: CP11A8Y9




KL cyclo IPalWNVSGA, BLUE ESIPOS C4 5 min, RT 0.2770 mins, Scan# 105, NL 3.974E6, 10/10/2016 10:23 AM, mz [305.8-1438.8]

8.6.4.10 CP11A8Y10

CP11A8Y10 (IChaWNVSGA) was cleaved from the resin to form the linear peptide (pale pink solid, 154.2 mg crude). This peptide (105.4 mg, 0.12 mmol) was cyclised, Figure 152, and purified via preparative HPLC yielding a white solid (36.1 mg, 0.041 mmol, 34% yield). LR MS (ESI+) *m/z* (%) 903.8 ($[M+Na]^+$, 100%), 455.7 (½$[M+H]^{2+}$, 56.04%), 881.7 ($[M+H]^+$, 40.22%). Analytical HPLC (280 nm) 19.9 min (98%).

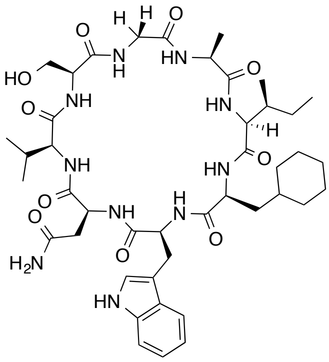
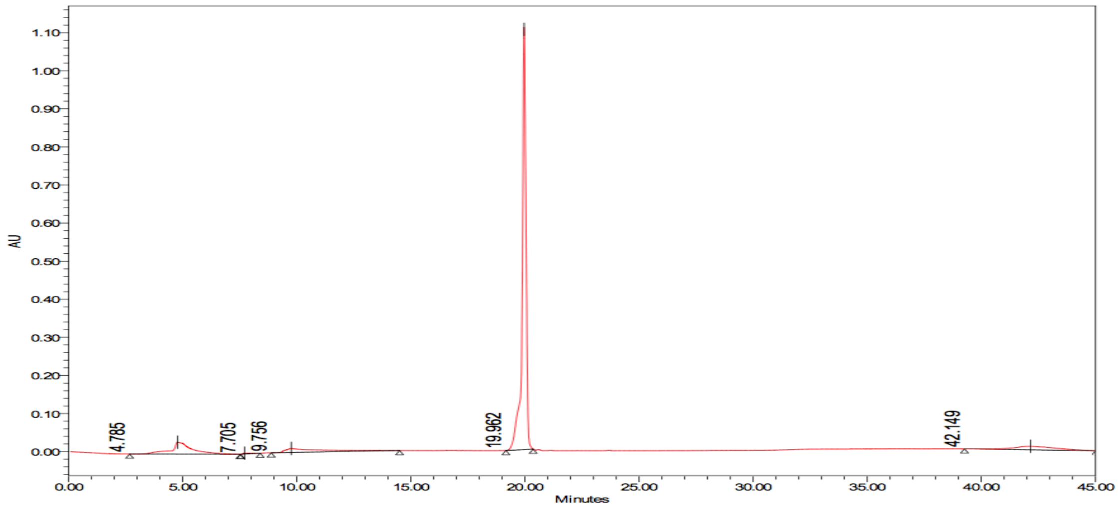
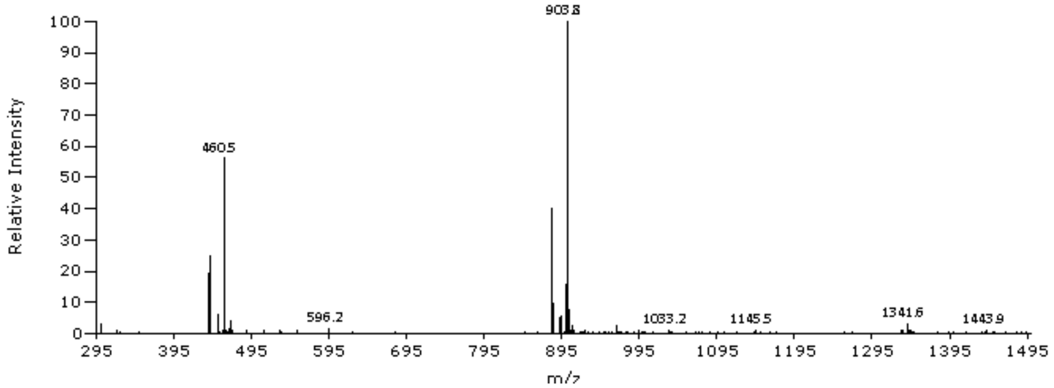


Figure 152: CP11A8Y10



KL cyclo IChaWNVSGA, BLUE ESIPOS C4 5 min, RT 1.6157 mins, Scan# 617, NL 9.896E6, 10/10/2016 7:38 AM, mz [302.4-1494.0]



8.6.4.11 CP11A8Y11

CP11A8Y11 (INalWNVSGA) was cleaved from the resin to form the linear peptide (white solid, 190.3 mg crude). This peptide (102.6 mg, 0.11 mmol) was cyclised, Figure 153, and purified via preparative HPLC yielding a pale cream solid (47.6 mg, 0.051 mmol, 47% yield). LR MS (ESI+) *m/z* (%) 947.7 ([M+Na]$^+$, 100%), 925.5 ([M+H]$^+$, 50.44%), 482.5 (½[M+H]$^{2+}$, 33.95%). Analytical HPLC (280 nm) 20.9 min (98%).

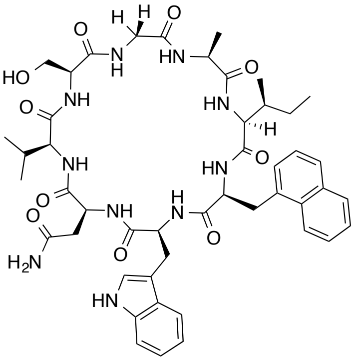
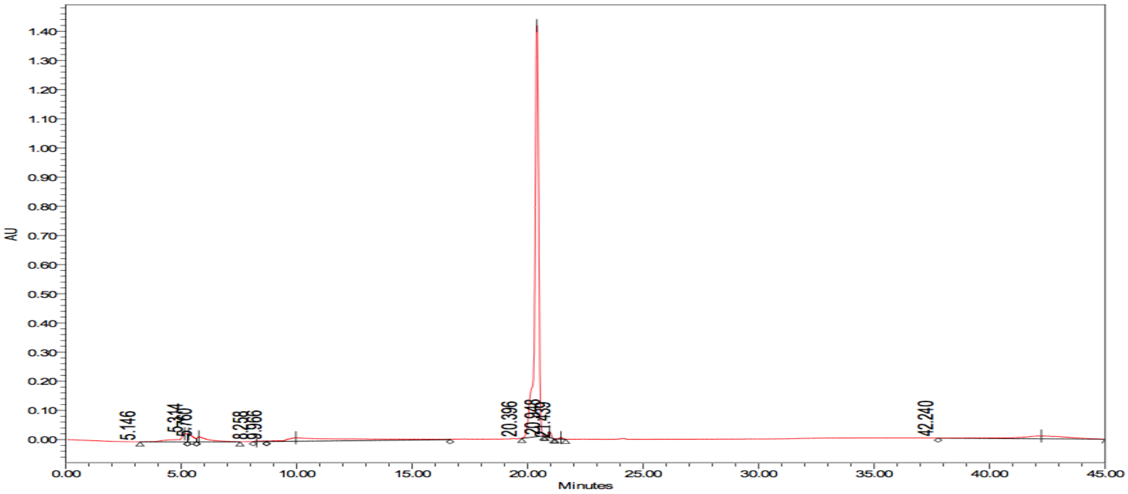
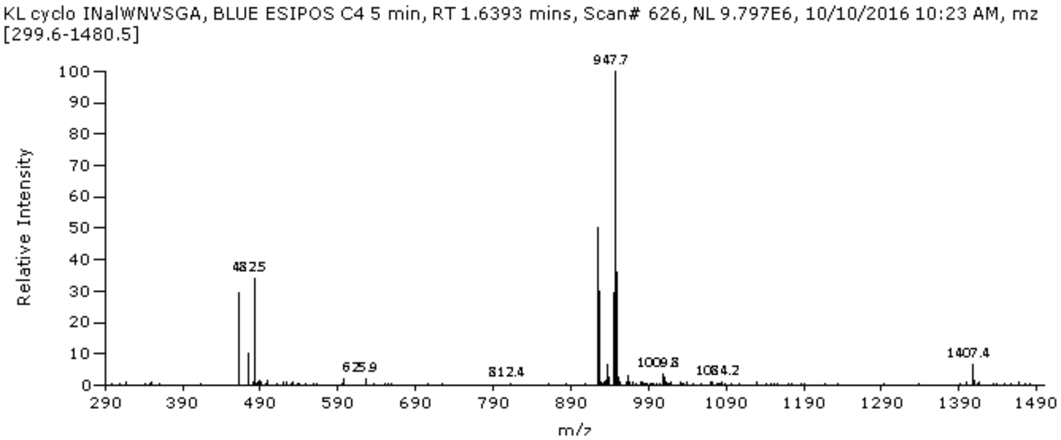


Figure 153: CP11A8Y11





8.6.5 Solid Phase Peptide Synthesis of Cysteine Derivatives

This series of cyclic peptides are the peptides that were found to be the most potent from the ELISA, however, the serine residue was substituted for cysteine so as to allow for tagging with the Tat peptide or with a fluorophore. The analogues were synthesised in bulk on a Glycine Wang resin (3 mmol scale) as 8.2.1, until the first tryptophan residue. The resin was then split into ~1mmol batches, then the tyrosine analogues were coupled and the rest of the peptide was finished.

8.6.5.1 CP11A4Y3C

CP11A4Y3C (IF(4-Cl)WAVCGW) was cleaved from the resin to form the linear peptide (red solid, 271.4 mg crude). This peptide (118 mg, 0.12 mmol) was cyclised, Figure 112, and purified via preparative HPLC yielding a pale yellow solid (18.4 mg, 0.019 mmol, 15% yield). LR MS (ESI+) *m/z* (%) 1003.8 ($[M+Na]^+$, 100%), 981.8 ($[M+H]^+$, 72.47%), 510.8 (½$[M+2Na]^{2+}$, 66.60%). There was insufficient material for running an analytical HPLC, however purity was estimated from the UV trace from MS to be 95%.

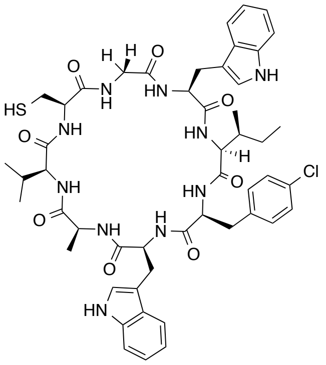
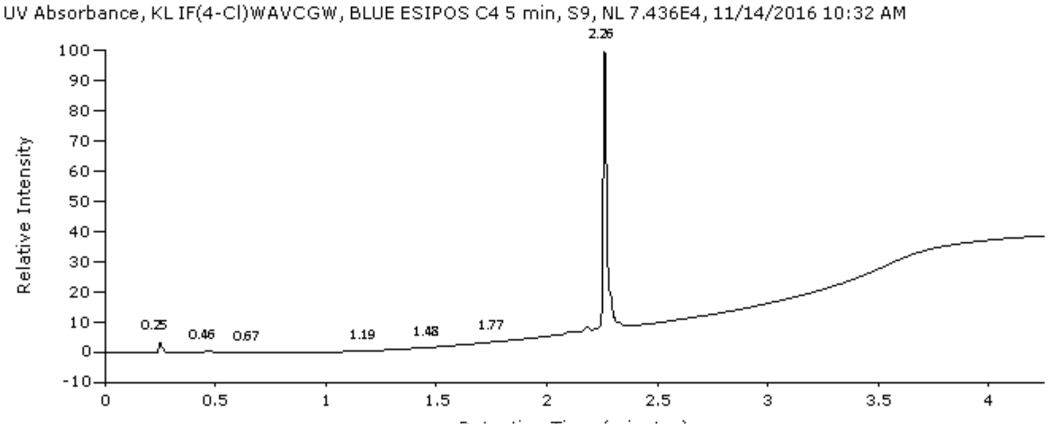
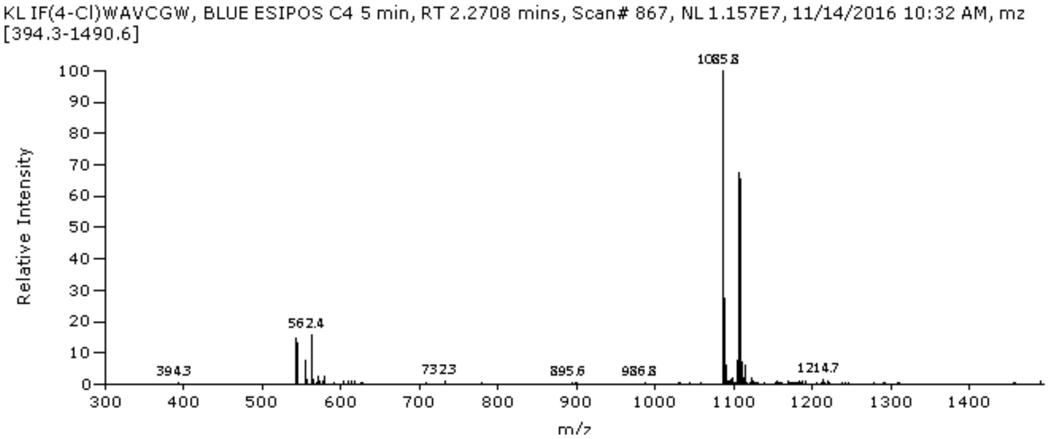


Figure 154: CP11A4Y3C





8.6.5.2 CP11A4Y4C

CP11A4Y4C (IF(4-Me)WAVCGW) was cleaved from the resin to form the linear peptide (cream, 59.0 mg crude). This peptide (59.0 mg, 0.058 mmol) was cyclised, Figure 155, and purified via preparative HPLC yielding an off white solid (8.6 mg, 0.008 mmol, 13% yield). LR MS (ESI+) *m/z* (%) 1065.9 ($[M+H]^+$, 100%), 1087.9 ($[M+Na]^+$, 76.37%), 552.8 (½$[M+2Na]^{2+}$, 33.68%). There was insufficient material for running an analytical HPLC, however purity was estimated from the UV trace from MS to be 95%.

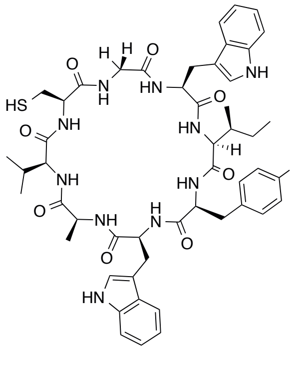
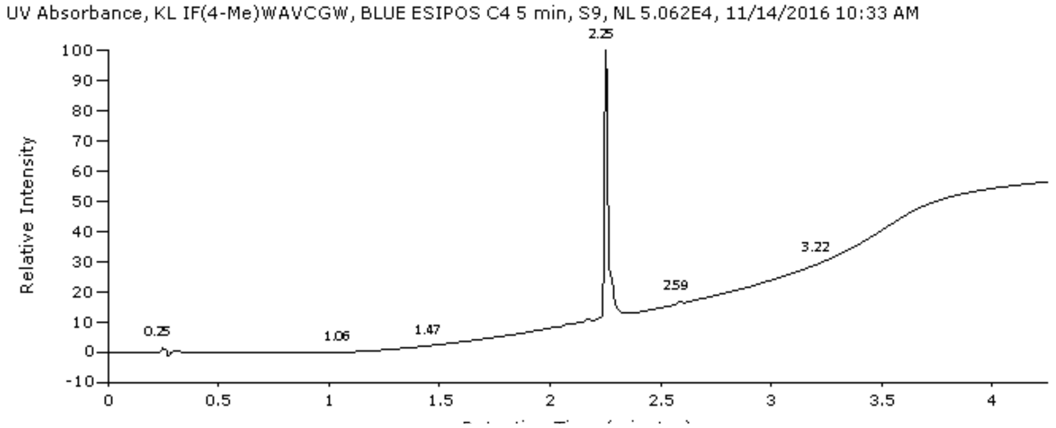
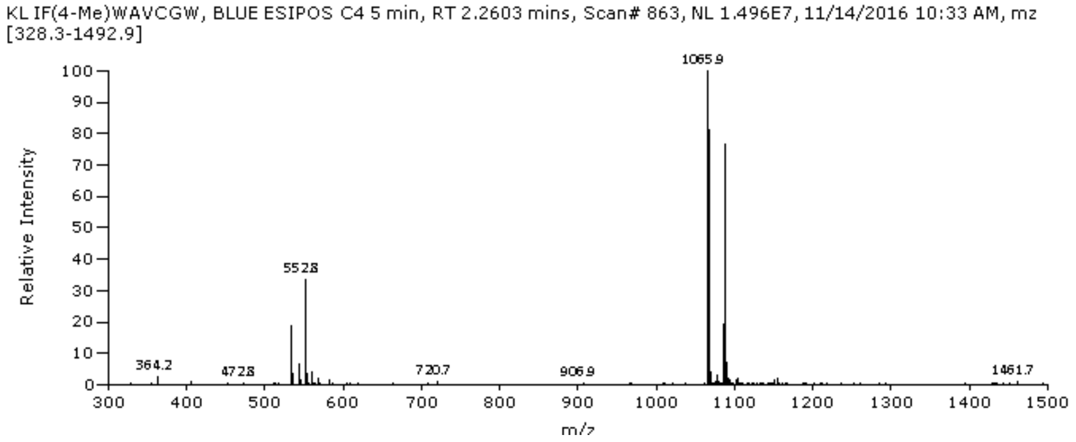


Figure 155: CP11A4Y4C





8.6.5.3 CP11A4Y10C

CP11A4Y4C (IF(4-Me)WAVCGW) was cleaved from the resin to form the linear peptide (pale yellow solid, 101.6 mg crude). This peptide (101.6 mg, 0.095 mmol) was cyclised, Figure 156, and purified via preparative HPLC yielding a cream solid (29.4 mg, 0.028 mmol, 29% yield). LR MS (ESI+) *m/z* (%) 1079.8 ([M+H]$^+$, 100%), 1057.8 ([M+Na]$^+$, 98.63%), 548.8 (½[M+2Na]$^{2+}$, 45.92%). There was insufficient material for running an analytical HPLC, however purity was estimated from the UV trace from MS to be 95%.

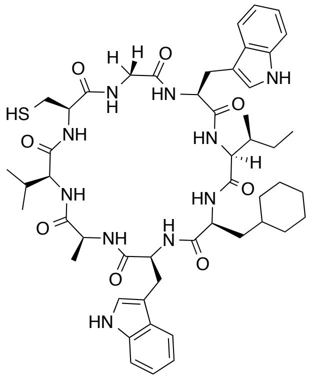
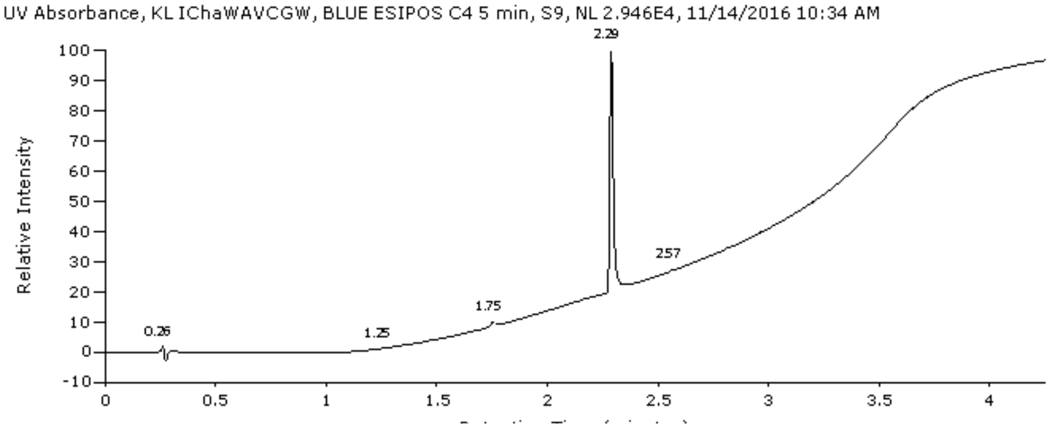
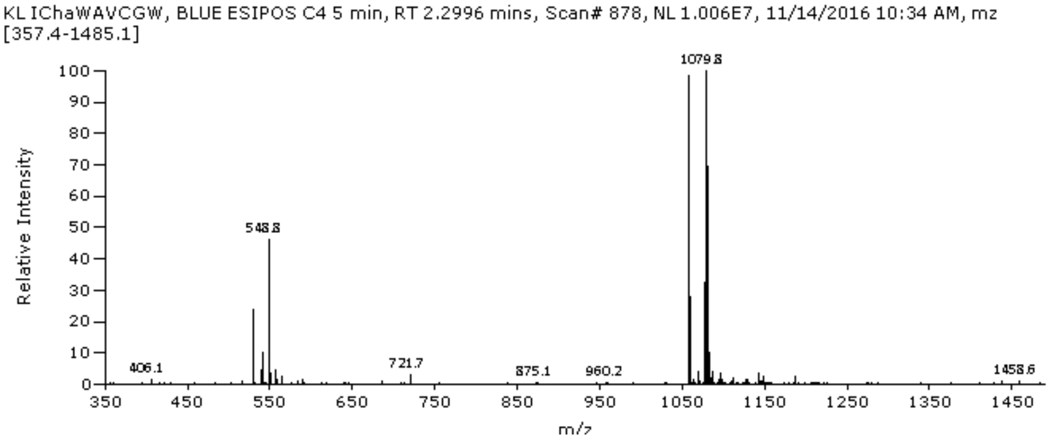


Figure 156: CP11A4Y10C





8.6.6 Thermal Shift

The melting temperature (T_m) of the protein is the midpoint of the thermal denaturing curve. It was calculated in this instance using the GraphPad Prism Software by firstly plotting the first derivative of the curve, so that the rate of fluoresence change is plotted against the temperature, and therefore the Tm becomes the point of the highest rate and therefore the peak of the trace. ΔT_m values were calculated as the difference between the T_m of the protein with 10% DMSO and the T_m of the protein incubated with ligand. Where applicable, the error bars represent the standard deviation from the mean of separate experiments.

8.6.6.1 Creating a melting curve

Solutions of protein (with final concentrations of 0-20 μM) and SYPRO® Orange (diluted from 5000× to a final concentration of 300× in DMSO; 1 μL; Life Technologies) were made up to 20 μL with Buffer A (20 mM Tris-HCl, 100 mM NaCl, 10% glycerol, pH 7.4) and added to a 96 well PCR plate (BioRad). The plate was gently centrifuged for 1 minute to mix the components. The thermal shift assay was performed on a BioRad CFX Connect Real-Time System and the system set up to heat from 30 °C to 95 °C at a rate of 1 °C/ 30 seconds with a fluorescence reading measured at $\lambda_{ex}/\lambda_{em}$: 300, 470 nm/ 570 nm every 30 seconds.

8.6.6.2 Analysing the effect of a peptide inhibitor on the melting temperature

Solutions of GST-tagged p6 and His-tagged UEV (with final concentrations of 5 µM of each) with varying concentrations of peptide inhibitor (final concentrations of 0-4 mM dissolved in DMSO, with a final concentration of DMSO per sample of 10%) and SYPRO® Orange (diluted from 5000× to a final concentration of 300× in DMSO; 1 µL; Life Technologies) were made up to 20 µL with Buffer A (20 mM Tris-HCl, 100 mM NaCl, 10% glycerol, pH 7.4) and added to a 96 well PCR plate (BioRad). The plate was gently centrifuged for 1 minute to mix the components. The thermal shift assay was performed on a BioRad CFX Connect Real-Time System and the system set up to heat from 30 °C to 95 °C at a rate of 1 °C/ 30 seconds with a fluorescence reading measured at $\lambda_{ex}/\lambda_{em}$: 300, 470 nm/ 570 nm every 30 seconds.

8.6.7 Microscale Thermophoresis

The purified His_6-UEV recombinant protein was labelled with the red NT-647-Maleimide dye using a protein-labelling kit (NanoTemper Technologies) according to the manufacturers instructions. The labelled protein was then purified using a gravity flow column into MST assay buffer (20 mM HEPES, pH 7.4, 100 mM NaCl, 5% Glycerol, 0.01% Tween-20). The inhibitor was dissolved in 100% DMSO and then diluted 5-fold into MST buffer. The inhibitor was serially diluted in MST assay buffer supplemented with 20% DMSO (5 µL) and then mixed with the labelled protein 1:1 (10 µL final volume, 0.01 µM final concentration of protein, final DMSO concentration 10%). The samples were incubated at room temperature for 5 minutes and then loaded into MST premium coated capillaries (NanoTemper Technologies). The samples were analysed on a Monolith NT.115 MST Instrument (NanoTemper Technologies, Excitation Power 20%, MST Power 40%).

8.7 Experimental for Chapter 6

8.7.1 Plasmids for Cell Culture

Expression vector pMET7-GAG-eGFP was a gift from Sven Eyckerman (Addgene plasmid ♯ 80605[183]. Expression vector pBABE-puro-EGFR-WT was a gift from Matthew Meyerson (Addgene plasmid ♯ 11011)[184]. EGFR was amplified with the primers 5’-CGCGCGCTAGCGCCACCATGGCGATGCGACCCTCCGGGACGGC-3’ and 5’-CGCGCGATATCTGCTCCAATAAATTCACTGCTTTGTGGCGC-3’ then digested with NheI-HF and EcoRV-HF (both New England Biolabs Inc.) and then ligated into pcDNA3.1 CAT. Plasmid vectors pcDNA5/FRT/TO and pOG44 were a kind gift from Dr. Noel Wortham (University of Southampton, UK).

8.7.2 Mammalian Cell Lines

HEK293T cells were a kind gift from Dr. Diego Gomez-Nicola (University of Southampton, UK). Flp-In T-REx293 (T-REx293) cells were a kind gift from Dr. Noel Wortham (University of Southampton, UK).

8.7.3 Preparation of Mammalian Cell Stocks

The culture medium was carefully removed from the flask, and then cells were gently washed with PBS (5 mL). The cells were incubated with trypsin (4 mL, 0.05%) at 37°C for approximately 5 minutes, or until the cells had detached from the flask surface. Fresh culture medium (6 mL) was added to the flask to neutralise the trypsin and the cells were transferred to a 15 mL centrifuge tube. The cells were centrifuged (1000 rpm, 4 min), the supernatant discarded and the cell pellet was resuspended in fresh culture medium (10 mL). The cells were counted using a Moxi Z Automated Cell Counter (ORFLO Technologies, VWR). An amount of culture was transferred to a fresh 15 mL centrifuge tube to ensure that there were 3×10^6 cells per stock. The tube was centrifuged (1000 rpm, 4 minutes), the supernatant discarded and the pelleted cells were resuspended in Freezing Medium (1 mL per stock; 50% FBS, 10% DMSO, 40% DMEM (v/v)) and transferred into individual cryotubes. The stocked were gradually frozen in an isopropanol box, placed in a -80°C freezer for 24 hours and then transferred to liquid nitrogen for long term storage.

8.7.4 Thawing Frozen Mammalian Cell Stocks

After, removing the cell stock from liquid nitrogen storage, it was thawed rapidly in a 37°C water bath and immediately transferred to culture medium (10 mL) in a 15 mL centrifuge tube. The cells were centrifuged (1000 rpm, 4 mins) and the supernatant removed. The cell pellet was resuspended in fresh culture medium (12 mL) and transferred to a sterile 75 cm^3 flask and incubated at 37°C, 5% CO_2

8.7.5 General Cell Culture Procedures

All cell culture reagents were purchased from Life Technologies unless otherwise stated. Mammalian cells were incubated in a HeraCell 150I Incubator (FisherScientific, UK), at 5% CO_2, 37 °C. HEK293T cells were maintained in DMEM Glutamax supplemented with antibiotics (penicillin/streptomycin 100 U/mL/100 µg/mL) and 10% Fetal Bovine Serum (FBS). HeLa cells were maintained in DMEM Glutamax supplemented with 10% FBS. T-REx293 cells were maintained in DMEM Glutamax supplemented with 10% FBS, hygromycin B (100 µg/mL) and blasticidin (15 µg/mL). Plasmid DNA transfections were performed using Lipofectamine 2000 (Invitrogen), as per the manufacturer's instructions.

8.7.5.1 Mammalian Cell Passaging

Mammalian cells were passaged when the cells were approximately 70-80% confluent. The culture medium was carefully removed from the flask, and then cells were gently washed with PBS (5 mL). The cells were incubated with trypsin (4 mL, 0.05%) at 37°C for approximately 5 minutes, or until the cells had detached from the flask surface. Fresh culture medium (6 mL) was added to the flask to neutralise the trypsin and the cells were transferred to a 15 mL centrifuge tube. The cells were centrifuged (1000 rpm, 4 min), the supernatant discarded and the cell pellet was resuspended in fresh culture medium. The desired concentration of cells was transferred to a sterile 75 cm^3 flask containing culture medium, up to a final volume of 12 mL.

8.7.5.2 Lysis of Harvested Cell Pellets

The frozen cell pellets were thawed on ice and then resuspended in 100 µL RIPA buffer containing protease inhibitor. The resuspended cell pellets were incubated on ice for 5-10 minutes and then sonicated in an ice water bath (10 × 30 s on, 30 s off). The samples were centrifuged (10 000 rpm, 10 minutes, 4 °C) and the supernatants were transferred to a clean tube. The protein concentration was estimated using a plate-based Bradford Assay, 8.7.5.3.

8.7.5.3 Plate-Based Bradford Assay for Estimating Protein Concentration

In a 96-well plate, bovine serum albumin (BSA) standards (0, 0.1, 0.25, 0.5, 1.0, 1.5 mg/mL) were added in duplicate (5 µL/well). The cell lysate samples were diluted 1:5 and then added to the same plate in duplicate (5 µL/well) as well as a 1:5 diluted RIPA buffer (5 µL/well). To each sample, Bradford Reagent (250 µL) was added and the absorbance measured at 595 nm.

8.7.6 Western Blotting Procedure

The cell lysate protein samples were diluted to approximately 80 ng and prepared as 8.3.14, and loaded into an SDS-PAGE gel of the desired percentage. The SDS-PAGE gel was run as 8.3.14. The SDS-PAGE gel was then loaded into the transfer cassette with blotting paper and transfer membrane (GE Healthcare Amersham Hybond-ECL), that was pre-soaked in ice-cold Transfer Buffer, Table 33.

Table 33: Transfer Buffer Composition

Reagent	Amount
Glycine	2.93 g
1.5 M Tris (pH 8.3)	32 mL
Methanol	200 mM
dH_2O	Make up to 1 L

The apparatus was set up with a cooling block and ice-cold Transfer Buffer and the transfer was run for 2 hours at 250 mA. The membrane was stained with Ponceau Stain (0.1% (w/v) Ponceau in 1% acetic acid) to confirm successful transfer onto the membrane. The Ponceau Stain was washed from the membrane using the Transfer Buffer and then the membrane was blocked with 5% (w/v) non-fat powdered milk with 0.1% Tween-20 (v/v) in PBS and then subjected to immunoblot analysis. The primary antibodies either anti-eGFP (MA1-952, 1:1000, ThermoFisher), or anti-EGFR (ab252894, 1:1000, Abcam) were diluted in 5% (w/v) non-fat powdered milk with 0.1% Tween-20 (v/v) in PBS and incubated with the membrane overnight at 4°C. Horseradish peroxidase conjugated anti-mouse (NA931, 1:100 000, GE Healthcare) or anti-rabbit (7074, 1:50 000, Cell Signalling) antibodies were used as the secondary antibodies for anti-eGFP and anti-EGFR respectively. Bound immunocomplexes were detected using ECL prime Western blot detection reagent (RPN2232, GE Healthcare) and analyzed using a ChemiDoc Imaging System (Bio-Rad) and Image Lab 4.0 Software (Bio-Rad).

To visualise β-actin, the membrane was washed with PBS-Tween as before and then incubated with α-actin HRP (1:50 000, in 5% milk, 20 mL) for 1 hour at room temperature. The membrane was then washed as before and exposed to Thermo SuperSignal West Pico ECL Kit (600 µL, 1:1 Reagent A: Reagent B) for 5 minutes in the dark. The excess ECL reagent was removed and the membrane visualised using a ChemiDoc Imaging System (Bio-Rad) and Image Lab 4.0 Software (Bio-Rad).

8.7.7 Virus-Like Particle Assay

HEK293T celled were seeded at 25 000 cells/well in a 24-well plate and incubated overnight such that the cells reached 60-70% confluency prior to transfection. The cells were transfected with Lipofectamine 2000 and pMET7-GAG-EGFP in a 4:1 ratio in Opti-MEM (Life Technologies). After 6 hours, the cells were either mock-treated with control or DMSO (final concentration 0.2% DMSO), or treated with increasing concentrations of peptide. After 18 hours, the supernatants were collected and the cells washed and lysed prior to immunoblotting analysis. The supernatants were centrifuged (1000 rpm, 4 minutes), and then filtered through a 20% (w/v) sucrose cushion (12 000 rpm, 90 minutes, 4°C). The filtered VLPs (final 15 µL) was resuspended with Laemelli Buffer (5 µL) and then subjected to SDS-PAGE and immunoblotting analysis.

8.7.8 Cytotoxicity Assay

The cytotoxicity was investigated using the CytoTox96® Non-Radioactive Cytotoxicity Assay Kit (Promega) used according to the Manufacturers guidelines and the absorbance at 490 nm measured using a BMG CLARIOstar Plate reader.

8.7.9 EGFR Down-Regulation Assay

HeLa cells were seeded at 150 000 cells/well in a 6-well plate and incubated overnight such that the cells reached 60-70% confluency prior to transfection. The cells were transfected with Lipofectamine 2000 and pcDNA3.1 CAT-EGFR in a 1:1 ratio in Opti-MEM. After 6 hours, the cells were either mock-treated or treated with DMSO (final concentration 0.2% DMSO) or Peptide (10 µM) in DMEM Glutamax supplemented with 10% FBS and incubated for a further 18 hours. The cells were then starved in serum-free DMEM Glutamax. After 1 hour, the cells were either mock-treated or treated with EGF (150 ng/mL, Sigma) for 90 minutes. The cells were washed with cold PBS and immediately lysed and subjected to SDS-PAGE and immunoblotting analysis.

8.7.10 Creation of a Stable Cell Line

T-REx293 cells were seeded at 800 000 cells per 6 cm diameter plate in triplicate and incubated overnight. The cells were then transfected using Lipofectamine 2000 at a 4:1 ratio according to the manufacturers guidelines, with the concentration of each plasmid as Table 34.

Table 34: Amount of Plasmid Transfected into T-REx293 Cells for Creation of a Stable Cell Line

Plasmid	Negative Control	No Enzyme Control	Positive Control
pOG44	0	0	2700 ng
pCDNA5/GOI	0	300 ng	300 ng

After transfection, the plates were incubated overnight. The medium was removed from the plates and the cells were washed with PBS (2 mL) and then incubated at 37°C with trypsin (0.05%, 1 mL) until the cells had become detached. To neutralise the trypsin, DMEM Glutamax supplemented with 10% FBS (2 mL) was added to the plate, and then the cells were collected in a 15 mL centrifuge tube. The cells were centrifuged (1000 rpm, 4 min) and the supernatant discarded. The cells were resuspended in Selection Medium (4 mL; DMEM Glutamax supplemented with 10% FBS, hygromycin (200 µg/mL) and blasticidin (15 µg/mL)) and then the resuspended cells (1 mL) were re-plated into 10 cm diameter plates with Selection Medium (9 mL). The cells were incubated at 37°C, 5% CO_2. Every 2 days the cells were gently washed with PBS (10 mL) and then fresh Selection Medium (10 mL) was added. This was repeated until there were no visible cells growing on the negative and no enzyme control plates. These plates were then discarded. The positive control plate was incubated until the cells were approximately 70-80% confluent and then they were split as 8.7.5.1 and maintained in a 75 cm^3 flask.

8.7.10.1 Expression of Peptide in Stable Cell Line

T-REx293_A4 cells were seeded at 2×10^5 cells/well in a 6 well plate with Selection Medium and incubated overnight. The following day, the selection medium was aspirated from the cells and the cells washed gently with PBS (1 mL) and Expression Medium was added (2 mL/well; DMEM supplemented with 10% FBS, hygromycin (200 µg/mL), blasticidin (15 µg/mL) and doxycycline (1 µg/mL)), this was the zero time point and the first cells were harvested. The cells were harvested at 1.5, 3, 6, 12, and 24 hours and lysed immediately and subjected to Western Blot analysis as 8.7.6 with an anti-CBD monoclonal antibody (1:000, E8034S, NewEnglandBioLabs).

8.7.10.2 Extraction of Peptide and Analysis by Accurate HPLC-ESI

T-REx293_A4 cells were seeded at 2×10^6 cells in a 10 cm diameter plate, one plate was incubated with selection medium and one was incubated with expression medium (10 mL) for 36 hours. The cells were harvested in ice cold PBS, and centrifuged (1000 rpm, 4 min). The cell pellets were frozen in liquid nitrogen, then thawed and resuspended in PMSF lysis buffer (500 µL per pellet; 5 mM EDTA, 2 mM EGTA, 0.4 mM PMSF in PBS with protease inhibitor). The cells were lysed by three freeze thaw cycles. The lysate was centrifuged (8000 rpm, 30 mins, 4°C). The supernatant was then passed through a 10 kDa cut-off filter (Amicon Ultra 0.5, ThermoFisher). The flow-through was collected and immediately analysed by manual injection into reverse-phase HPLC on a Waters HPLC system equipped with a Waters Atlantis T3, Amide capped C18 5 µM, 6 × 100 mm column. Fractions were collected in a 5 minute window around the elution time of the synthetic peptide. The fractions were lyophilised and then analysed by LC-MS. Samples were analysed using a MaXis (Bruker Daltonics, Bremen, Germany) mass spectrometer equipped with a Time of Flight (TOF) analyser. Samples were introduced to the mass spectrometer via a Dionex Ultimate 3000 autosampler and uHPLC pump. Gradient 20% acetonitrile (0.2% formic acid) to 100% acetonitrile (0.2% formic acid) in five minutes at 0.6 mL min. Column, Acquity UPLC BEH C18 (Waters) 1.7 micron 50 x 2.1 mm. High resolution mass spectra were recorded using positive ion electrospray ionisation.

8.7.10.3 Testing Expressed Peptide for Inhibition of Viral Budding

T-REx293-CP11A4 cells were seeded at 1 × 10^5 cells/well in 4 wells of a 6 well plate. Two of the wells (1, 2) were incubated with selection medium and the other two wells (3, 4) were incubated with expression medium for 12 hours. The cells were washed with PBS (2 mL) and then Wells 1 and 3 were transfected with Lipofectamine 2000 and pMET7-GAG-EGFP in a 4:1 ratio in Opti-MEM (Life Technologies) and Wells 2 and 4 were incubated with Opti-MEM for 6 hours. The transfection medium was removed from each well and selection medium or expression medium was added to the wells as before (2 mL/ well. After 24 hours, the supernatants were collected and the cells washed and lysed prior to immunoblotting analysis. The supernatants were centrifuged (1000 rpm, 4 minutes), and then filtered through a 20% (w/v) sucrose cushion (12 000 rpm, 90 minutes, 4°C). The filtered VLPs (final 15 µL) was resuspended with Laemelli Buffer (5 µL) and then subjected to SDS-PAGE and immunoblotting analysis. The membrane was stripped inbetween incubation with anti-CBD antibody and prior to incubation with anti-eGFP antibody by repeated exposure (2 × 10 minutes) to a mild stripping buffer (15 g glycine, 1g SDS, 10 mL Tween 20, pH 2.2, up to 1 L with deionised water). The membrane was washed with PBS-Tween 20 buffer (2 × 10 minutes).

Appendices

All raw data, mass spec traces and analytical HPLC traces can be found on the attached CD.

Bibliography

1. Masur, H.; Michelis, M. A.; Greene, J. B.; Onorato, I.; Stouwe, R. A.; Holzman, R. S.; Wormser, G.; Brettman, L.; Lange, M.; Murray, H. W.; Cunningham-Rundles, S., An outbreak of community-acquired Pneumocystis carinii pneumonia: initial manifestation of cellular immune dysfunction. *The New England journal of medicine* **1981,** *305* (24), 1431-8.

2. Hymes, K. B.; Cheung, T.; Greene, J. B.; Prose, N. S.; Marcus, A.; Ballard, H.; William, D. C.; Laubenstein, L. J., Kaposi's sarcoma in homosexual men-a report of eight cases. *Lancet* **1981,** *2* (8247), 598-600.

3. Barre-Sinoussi, F.; Chermann, J. C.; Rey, F.; Nugeyre, M. T.; Chamaret, S.; Gruest, J.; Dauguet, C.; Axler-Blin, C.; Vezinet-Brun, F.; Rouzioux, C.; Rozenbaum, W.; Montagnier, L., Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **1983,** *220* (4599), 868-71.

4. Gallo, R. C.; Sarin, P. S.; Gelmann, E. P.; Robert-Guroff, M.; Richardson, E.; Kalyanaraman, V. S.; Mann, D.; Sidhu, G. D.; Stahl, R. E.; Zolla-Pazner, S.; Leibowitch, J.; Popovic, M., Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science* **1983,** *220* (4599), 865-7.

5. Levy, J. A.; Hoffman, A. D.; Kramer, S. M.; Landis, J. A.; Shimabukuro, J. M.; Oshiro, L. S., Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. *Science* **1984,** *225* (4664), 840-2.

6. Ratner, L.; Gallo, R. C.; Wong-Staal, F., HTLV-III, LAV, ARV are variants of same AIDS virus. *Nature* **1985,** *313* (6004), 636-7.

7. Coffin, J.; Haase, A.; Levy, J. A.; Montagnier, L.; Oroszlan, S.; Teich, N.; Temin, H.; Toyoshima, K.; Varmus, H.; Vogt, P.; et al., What to call the AIDS virus? *Nature* **1986,** *321* (6065), 10.

8. Sharp, P. M.; Hahn, B. H., Origins of HIV and the AIDS pandemic. *Cold Spring Harbor perspectives in medicine* **2011,** *1* (1), a006841.

9. Gentile, M.; Adrian, T.; Scheidler, A.; Ewald, M.; Dianzani, F.; Pauli, G.; Gelderblom, H. R., Determination of the size of HIV using adenovirus type 2 as an internal length marker. *Journal of virological methods* **1994,** *48* (1), 43-52.

10. Watts, J. M.; Dang, K. K.; Gorelick, R. J.; Leonard, C. W.; Bess, J. W., Jr.; Swanstrom, R.; Burch, C. L.; Weeks, K. M., Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **2009,** *460* (7256), 711-6.

11. Kwong, P. D.; Wyatt, R.; Robinson, J.; Sweet, R. W.; Sodroski, J.; Hendrickson, W. A., Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **1998,** *393* (6686), 648-59.

12. Engelman, A.; Cherepanov, P., The structural biology of HIV-1: mechanistic and therapeutic insights. *Nature reviews. Microbiology* **2012,** *10* (4), 279-90.

13. Clavel, F.; Hance, A. J., HIV Drug Resistance. *New England Journal of Medicine* **2004,** *350* (10), 1023-1036.

14. Preston, B. D.; Poiesz, B. J.; Loeb, L. A., Fidelity of HIV-1 reverse transcriptase. *Science* **1988,** *242* (4882), 1168-71.

15. Gottlinger, H. G.; Sodroski, J. G.; Haseltine, W. A., Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. *Proc Natl Acad Sci U S A* **1989,** *86* (15), 5781-5.

16. Yu, X.; Yuan, X.; McLane, M. F.; Lee, T. H.; Essex, M., Mutations in the cytoplasmic domain of human immunodeficiency virus type 1 transmembrane protein impair the incorporation of Env proteins into mature virions. *Journal of virology* **1993,** *67* (1), 213-21.

17. Moore, M. D.; Nikolaitchik, O. A.; Chen, J.; Hammarskjold, M. L.; Rekosh, D.; Hu, W. S., Probing the HIV-1 genomic RNA trafficking pathway and dimerization by genetic recombination and single virion analyses. *PLoS pathogens* **2009,** *5* (10), e1000627.

18. Hubner, W.; Chen, P.; Del Portillo, A.; Liu, Y.; Gordon, R. E.; Chen, B. K., Sequence of human immunodeficiency virus type 1 (HIV-1) Gag localization and oligomerization monitored with live confocal imaging of a replication-competent, fluorescently tagged HIV-1. *Journal of virology* **2007,** *81* (22), 12596-607.

19. Ono, A.; Ablan, S. D.; Lockett, S. J.; Nagashima, K.; Freed, E. O., Phosphatidylinositol (4,5) bisphosphate regulates HIV-1 Gag targeting to the plasma membrane. *Proc Natl Acad Sci U S A* **2004,** *101* (41), 14889-94.

20. Saad, J. S.; Loeliger, E.; Luncsford, P.; Liriano, M.; Tai, J.; Kim, A.; Miller, J.; Joshi, A.; Freed, E. O.; Summers, M. F., Point mutations in the HIV-1 matrix protein turn off the myristyl switch. *Journal of molecular biology* **2007,** *366* (2), 574-85.

21. Hogue, I. B.; Grover, J. R.; Soheilian, F.; Nagashima, K.; Ono, A., Gag induces the coalescence of clustered lipid rafts and tetraspanin-enriched microdomains at HIV-1 assembly sites on the plasma membrane. *Journal of virology* **2011,** *85* (19), 9749-66.

22. Briggs, J. A.; Riches, J. D.; Glass, B.; Bartonova, V.; Zanetti, G.; Krausslich, H. G., Structure and assembly of immature HIV. *Proc Natl Acad Sci U S A* **2009,** *106* (27), 11090-5.

23. Morita, E.; Sundquist, W. I., Retrovirus budding. *Annu Rev Cell Dev Biol* **2004,** *20*, 395-425.

24. Demirov, D. G.; Freed, E. O., Retrovirus budding. *Virus research* **2004,** *106* (2), 87-102.

25. Chojnacki, J.; Muller, B., Investigation of HIV-1 assembly and release using modern fluorescence imaging techniques. *Traffic* **2013,** *14* (1), 15-24.

26. Briant, L., HIV-1 Assembly, Release and Maturation. *World Journal of AIDS* **2011,** *01* (04), 111-130.

27. Adamson, C. S.; Salzwedel, K.; Freed, E. O., Virus maturation as a new HIV-1 therapeutic target. *Expert Opin Ther Targets* **2009,** *13* (8), 895-908.

28. Kaplan, A. H.; Zack, J. A.; Knigge, M.; Paul, D. A.; Kempf, D. J.; Norbeck, D. W.; Swanstrom, R., Partial inhibition of the human immunodeficiency virus type 1 protease results in aberrant virus assembly and the formation of noninfectious particles. *Journal of virology* **1993,** *67* (7), 4050-5.

29. Katzmann, D. J.; Babst, M.; Emr, S. D., Ubiquitin-dependent sorting into the multivesicular body pathway requires the function of a conserved endosomal protein sorting complex, ESCRT-I. *Cell* **2001,** *106* (2), 145-55.

30. Schmidt, O.; Teis, D., The ESCRT machinery. *Current biology : CB* **2012,** *22* (4), R116-20.

31. Bache, K. G.; Brech, A.; Mehlum, A.; Stenmark, H., Hrs regulates multivesicular body formation via ESCRT recruitment to endosomes. *The Journal of cell biology* **2003,** *162* (3), 435-42.

32. Baldys, A.; Raymond, J. R., Critical role of ESCRT machinery in EGFR recycling. *Biochemistry* **2009,** *48* (40), 9321-3.

33. Raiborg, C.; Stenmark, H., The ESCRT machinery in endosomal sorting of ubiquitylated membrane proteins. *Nature* **2009,** *458* (7237), 445-52.

34. Strack, B.; Calistri, A.; Craig, S.; Popova, E.; Gottlinger, H. G., AIP1/ALIX is a binding partner for HIV-1 p6 and EIAV p9 functioning in virus budding. *Cell* **2003,** *114* (6), 689-99.

35. Fisher, R. D.; Chung, H. Y.; Zhai, Q.; Robinson, H.; Sundquist, W. I.; Hill, C. P., Structural and biochemical studies of ALIX/AIP1 and its role in retrovirus budding. *Cell* **2007,** *128* (5), 841-52.

36. VerPlank, L.; Bouamr, F.; LaGrassa, T. J.; Agresta, B.; Kikonyogo, A.; Leis, J.; Carter, C. A., Tsg101, a homologue of ubiquitin-conjugating (E2) enzymes, binds the L domain in HIV type 1 Pr55(Gag). *Proc Natl Acad Sci U S A* **2001,** *98* (14), 7724-9.

37. Garrus, J. E.; von Schwedler, U. K.; Pornillos, O. W.; Morham, S. G.; Zavitz, K. H.; Wang, H. E.; Wettstein, D. A.; Stray, K. M.; Cote, M.; Rich, R. L.; Myszka, D. G.; Sundquist, W. I., Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. *Cell* **2001,** *107* (1), 55-65.

38. Pornillos, O.; Higginson, D. S.; Stray, K. M.; Fisher, R. D.; Garrus, J. E.; Payne, M.; He, G. P.; Wang, H. E.; Morham, S. G.; Sundquist, W. I., HIV Gag mimics the Tsg101-recruiting activity of the human Hrs protein. *The Journal of cell biology* **2003,** *162* (3), 425-34.

39. Pornillos, O.; Alam, S. L.; Davis, D. R.; Sundquist, W. I., Structure of the Tsg101 UEV domain in complex with the PTAP motif of the HIV-1 p6 protein. *Nat Struct Biol* **2002,** *9* (11), 812-7.

40. Gottlinger, H. G.; Dorfman, T.; Sodroski, J. G.; Haseltine, W. A., Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release. *Proc Natl Acad Sci U S A* **1991,** *88* (8), 3195-9.

41. Im, Y. J.; Kuo, L.; Ren, X.; Burgos, P. V.; Zhao, X. Z.; Liu, F.; Burke, T. R., Jr.; Bonifacino, J. S.; Freed, E. O.; Hurley, J. H., Crystallographic and functional analysis of the ESCRT-I /HIV-1 Gag PTAP interaction. *Structure* **2010,** *18* (11), 1536-47.

42. Morita, E.; Sandrin, V.; McCullough, J.; Katsuyama, A.; Baci Hamilton, I.; Sundquist, W. I., ESCRT-III protein requirements for HIV-1 budding. *Cell host & microbe* **2011,** *9* (3), 235-42.

43. Hurley, J. H.; Hanson, P. I., Membrane budding and scission by the ESCRT machinery: it's all in the neck. *Nature reviews. Molecular cell biology* **2010,** *11* (8), 556-66.

44. Lata, S.; Schoehn, G.; Jain, A.; Pires, R.; Piehler, J.; Gottlinger, H. G.; Weissenhorn, W., Helical structures of ESCRT-III are disassembled by VPS4. *Science* **2008,** *321* (5894), 1354-7.

45. Freed, E. O., HIV-1 assembly, release and maturation. *Nature reviews. Microbiology* **2015,** *13* (8), 484-96.

46. UNAIDS, Global AIDS Update 2016. **2016**.

47. Sok, D.; Le, K. M.; Vadnais, M.; Saye-Francisco, K.; Jardine, J. G.; Torres, J.; Berndsen, Z. T.; Kong, L.; Stanfield, R.; Ruiz, J.; Ramos, A.; Liang, C. H.; Chen, P. L.; Criscitiello, M. F.; Mwangi, W.; Wilson, I. A.; Ward, A. B.; Smider, V. V.; Burton, D. R., Rapid elicitation of broadly neutralizing antibodies to HIV by immunization in cows. *Nature* **2017**.

48. Choi, E.; Michalski, C. J.; Choo, S. H.; Kim, G. N.; Banasikowska, E.; Lee, S.; Wu, K.; An, H. Y.; Mills, A.; Schneider, S.; Bredeek, U. F.; Coulston, D. R.; Ding, S.; Finzi, A.; Tian, M.; Klein, K.; Arts, E. J.; Mann, J. F.; Gao, Y.; Kang, C. Y., First Phase I human clinical trial of a killed whole-HIV-1 vaccine:

demonstration of its safety and enhancement of anti-HIV antibody responses. *Retrovirology* **2016,** *13* (1), 82.

49. Guardo, A. C.; Joe, P. T.; Miralles, L.; Bargallo, M. E.; Mothe, B.; Krasniqi, A.; Heirman, C.; Garcia, F.; Thielemans, K.; Brander, C.; Aerts, J. L.; Plana, M.; i, H. c., Preclinical evaluation of an mRNA HIV vaccine combining rationally selected antigenic sequences and adjuvant signals (HTI-TriMix). *Aids* **2017,** *31* (3), 321-332.

50. McCormack, S.; Dunn, D. T.; Desai, M.; Dolling, D. I.; Gafos, M.; Gilson, R.; Sullivan, A. K.; Clarke, A.; Reeves, I.; Schembri, G.; Mackie, N.; Bowman, C.; Lacey, C. J.; Apea, V.; Brady, M.; Fox, J.; Taylor, S.; Antonucci, S.; Khoo, S. H.; Rooney, J.; Nardone, A.; Fisher, M.; McOwan, A.; Phillips, A. N.; Johnson, A. M.; Gazzard, B.; Gill, O. N., Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from the pilot phase of a pragmatic open-label randomised trial. *Lancet* **2016,** *387* (10013), 53-60.

51. Erice, A.; Mayers, D. L.; Strike, D. G.; Sannerud, K. J.; McCutchan, F. E.; Henry, K.; Balfour, H. H., Jr., Brief report: primary infection with zidovudine-resistant human immunodeficiency virus type 1. *The New England journal of medicine* **1993,** *328* (16), 1163-5.

52. Cane, P.; Chrystie, I.; Dunn, D.; Evans, B.; Geretti, A. M.; Green, H.; Phillips, A.; Pillay, D.; Porter, K.; Pozniak, A.; Sabin, C.; Smith, E.; Weber, J.; Zuckerman, M.; Resist, U. G. T. H. D., Time trends in primary resistance to HIV drugs in the United Kingdom: multicentre observational study. *British Medical Journal* **2005,** *331* (7529), 1368-1371.

53. Wensing, A. M.; van de Vijver, D. A.; Angarano, G.; Asjo, B.; Balotta, C.; Boeri, E.; Camacho, R.; Chaix, M. L.; Costagliola, D.; De Luca, A.; Derdelinckx, I.; Grossman, Z.; Hamouda, O.; Hatzakis, A.; Hemmer, R.; Hoepelman, A.; Horban, A.; Korn, K.; Kucherer, C.; Leitner, T.; Loveday, C.; MacRae, E.; Maljkovic, I.; de Mendoza, C.; Meyer, L.; Nielsen, C.; Op de Coul, E. L.; Ormaasen, V.; Paraskevis, D.; Perrin, L.; Puchhammer-Stockl, E.; Ruiz, L.; Salminen, M.; Schmit, J. C.; Schneider, F.; Schuurman, R.; Soriano, V.; Stanczak, G.; Stanojevic, M.; Vandamme, A. M.; Van Laethem, K.; Violin, M.; Wilbe, K.; Yerly, S.; Zazzi, M.; Boucher, C. A.; Programme, S., Prevalence of drug-resistant HIV-1 variants in untreated individuals in Europe: implications for clinical management. *J Infect Dis* **2005,** *192* (6), 958-66.

54. Ho, D. D.; Neumann, A. U.; Perelson, A. S.; Chen, W.; Leonard, J. M.; Markowitz, M., Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* **1995,** *373* (6510), 123-6.

55. Perelson, A. S.; Neumann, A. U.; Markowitz, M.; Leonard, J. M.; Ho, D. D., HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* **1996,** *271* (5255), 1582-6.

56. Schuurman, R.; Nijhuis, M.; van Leeuwen, R.; Schipper, P.; de Jong, D.; Collis, P.; Danner, S. A.; Mulder, J.; Loveday, C.; Christopherson, C.; et al., Rapid changes in human immunodeficiency virus type 1 RNA load and appearance of drug-resistant virus populations in persons treated with lamivudine (3TC). *J Infect Dis* **1995,** *171* (6), 1411-9.

57. Chun, T. W.; Moir, S.; Fauci, A. S., HIV reservoirs as obstacles and opportunities for an HIV cure. *Nature immunology* **2015,** *16* (6), 584-9.

58. Hutter, G.; Nowak, D.; Mossner, M.; Ganepola, S.; Mussig, A.; Allers, K.; Schneider, T.; Hofmann, J.; Kucherer, C.; Blau, O.; Blau, I. W.; Hofmann, W. K.; Thiel, E., Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. *The New England journal of medicine* **2009,** *360* (7), 692-8.

59. Luzuriaga, K.; Gay, H.; Ziemniak, C.; Sanborn, K. B.; Somasundaran, M.; Rainwater-Lovett, K.; Mellors, J. W.; Rosenbloom, D.; Persaud, D., Viremic relapse after HIV-1 remission in a perinatally infected child. *The New England journal of medicine* **2015,** *372* (8), 786-8.

60. Fuller, J. C.; Burgoyne, N. J.; Jackson, R. M., Predicting druggable binding sites at the protein-protein interface. *Drug discovery today* **2009,** *14* (3-4), 155-61.

61. Smith, M. C.; Gestwicki, J. E., Features of protein-protein interactions that translate into potent inhibitors: topology, surface area and affinity. *Expert reviews in molecular medicine* **2012,** *14*, e16.

62. Jones, S.; Thornton, J. M., Principles of protein-protein interactions. *Proc Natl Acad Sci U S A* **1996,** *93* (1), 13-20.

63. Scott, D. E.; Bayly, A. R.; Abell, C.; Skidmore, J., Small molecules, big targets: drug discovery faces the protein-protein interaction challenge. *Nature reviews. Drug discovery* **2016,** *15* (8), 533-50.

64. Clackson, T.; Wells, J. A., A hot spot of binding energy in a hormone-receptor interface. *Science* **1995,** *267* (5196), 383-6.

65. Bogan, A. A.; Thorn, K. S., Anatomy of hot spots in protein interfaces. *Journal of molecular biology* **1998,** *280* (1), 1-9.

66. Mullard, A., Protein-protein interaction inhibitors get into the groove. *Nature reviews. Drug discovery* **2012,** *11* (3), 173-5.

67. Hyde, J.; Braisted, A. C.; Randal, M.; Arkin, M. R., Discovery and characterization of cooperative ligand binding in the adaptive region of interleukin-2. *Biochemistry* **2003,** *42* (21), 6475-83.

68. Sheng, C.; Dong, G.; Miao, Z.; Zhang, W.; Wang, W., State-of-the-art strategies for targeting protein-protein interactions by small-molecule inhibitors. *Chemical Society reviews* **2015,** *44* (22), 8238-59.

69. Bakail, M.; Ochsenbein, F., Targeting protein–protein interactions, a wide open field for drug design. *Comptes Rendus Chimie* **2016,** *19* (1-2), 19-27.

70. Goudreau, N.; Cameron, D. R.; Deziel, R.; Hache, B.; Jakalian, A.; Malenfant, E.; Naud, J.; Ogilvie, W. W.; O'Meara, J.; White, P. W.; Yoakim, C., Optimization and determination of the absolute configuration of a series of potent inhibitors of human papillomavirus type-11 E1-E2 protein-protein interaction: a combined medicinal chemistry, NMR and computational chemistry approach. *Bioorg Med Chem* **2007,** *15* (7), 2690-700.

71. Arnold, L. A.; Estebanez-Perpina, E.; Togashi, M.; Jouravel, N.; Shelat, A.; McReynolds, A. C.; Mar, E.; Nguyen, P.; Baxter, J. D.; Fletterick, R. J.; Webb, P.; Guy, R. K., Discovery of small molecule inhibitors of the interaction of the thyroid hormone receptor with transcriptional coregulators. *J Biol Chem* **2005,** *280* (52), 43048-55.

72. Rees, D. C.; Congreve, M.; Murray, C. W.; Carr, R., Fragment-based lead discovery. *Nature reviews. Drug discovery* **2004,** *3* (8), 660-72.

73. Li, H.; Xiao, H.; Lin, L.; Jou, D.; Kumari, V.; Lin, J.; Li, C., Drug design targeting protein-protein interactions (PPIs) using multiple ligand simultaneous docking (MLSD) and drug repositioning: discovery of raloxifene and bazedoxifene as novel inhibitors of IL-6/GP130 interface. *J Med Chem* **2014,** *57* (3), 632-41.

74. Bullock, B. N.; Jochim, A. L.; Arora, P. S., Assessing helical protein interfaces for inhibitor design. *J Am Chem Soc* **2011,** *133* (36), 14220-3.

75. Mahato, R. I.; Narang, A. S.; Thoma, L.; Miller, D. D., Emerging trends in oral delivery of peptide and protein drugs. *Critical Reviews in Therapeutic Drug Carrier Systems* **2003,** *20* (2-3), 153-214.

76. Khafagy, E. S.; Morishita, M., Oral biodrug delivery using cell-penetrating peptide. *Advanced Drug Delivery Reviews* **2012,** *64* (6), 531-539.

77. Kessler, H., Peptide Conformations .19. Conformation and Biological-Activity of Cyclic-Peptides. *Angewandte Chemie-International Edition in English* **1982,** *21* (7), 512-523.

78. Horswill, A. R.; Benkovic, S. J., Cyclic peptides, a chemical genetics tool for biologists. *Cell Cycle* **2005,** *4* (4), 552-5.

79. Horton, D. A.; Bourne, G. T.; Smythe, M. L., Exploring privileged structures: the combinatorial synthesis of cyclic peptides. *Mol Divers* **2002,** *5* (4), 289-304.

80. Ladner, R. C., Constrained peptides as binding entities. *Trends in biotechnology* **1995,** *13* (10), 426-30.

81. Delorbe, J. E.; Clements, J. H.; Whiddon, B. B.; Martin, S. F., Thermodynamic and Structural Effects of Macrocyclization as a Constraining Method in Protein-Ligand Interactions. *ACS medicinal chemistry letters* **2010,** *1* (8), 448-452.

82. Rezai, T.; Yu, B.; Millhauser, G. L.; Jacobson, M. P.; Lokey, R. S., Testing the conformational hypothesis of passive membrane permeability using synthetic cyclic peptide diastereomers. *J Am Chem Soc* **2006,** *128* (8), 2510-1.

83. Grauer, A.; König, B., Peptidomimetics - A Versatile Route to Biologically Active Compounds. *European Journal of Organic Chemistry* **2009,** *2009* (30), 5099-5111.

84. Avan, I.; Hall, C. D.; Katritzky, A. R., Peptidomimetics via modifications of amino acids and peptide bonds. *Chemical Society reviews* **2014,** *43* (10), 3575-94.

85. Joo, S. H.; Xiao, Q.; Ling, Y.; Gopishetty, B.; Pei, D. H., High-throughput sequence determination of cyclic peptide library members by partial Edman degradation/mass spectrometry. *J Am Chem Soc* **2006,** *128* (39), 13000-13009.

86. Lian, W. L.; Upadhyaya, P.; Rhodes, C. A.; Liu, Y. S.; Pei, D. H., Screening Bicyclic Peptide Libraries for Protein-Protein Interaction Inhibitors: Discovery of a Tumor Necrosis Factor-alpha Antagonist. *J Am Chem Soc* **2013,** *135* (32), 11990-11995.

87. Nemoto, N.; MiyamotoSato, E.; Husimi, Y.; Yanagawa, H., In vitro virus: Bonding of mRNA bearing puromycin at the 3'-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro. *Febs Lett* **1997,** *414* (2), 405-408.

88. Goto, Y.; Katoh, T.; Suga, H., Flexizymes for genetic code reprogramming. *Nature protocols* **2011,** *6* (6), 779-90.

89. Hipolito, C. J.; Suga, H., Ribosomal production and in vitro selection of natural product-like peptidomimetics: The FIT and RaPID systems. *Curr Opin Chem Biol* **2012,** *16* (1-2), 196-203.

90. Heinis, C.; Rutherford, T.; Freund, S.; Winter, G., Phage-encoded combinatorial chemical libraries based on bicyclic peptides. *Nat Chem Biol* **2009,** *5* (7), 502-7.

91. Tian, F.; Tsao, M. L.; Schultz, P. G., A phage display system with unnatural amino acids. *J Am Chem Soc* **2004,** *126* (49), 15962-3.

92. Shah, N. H.; Muir, T. W., Inteins: nature's gift to protein chemists. *Chem Sci* **2014,** *5* (2), 446-461.

93. Frost, J. R.; Vitali, F.; Jacob, N. T.; Brown, M. D.; Fasan, R., Macrocyclization of Organo-Peptide Hybrids through a Dual Bio-orthogonal Ligation: Insights from Structure-Reactivity Studies. *Chembiochem* **2013,** *14* (1), 147-160.

94. Smith, J. M.; Hill, N. C.; Krasniak, P. J.; Fasan, R., Synthesis of bicyclic organo-peptide hybrids via oxime/intein-mediated macrocyclization followed by disulfide bond formation. *Org Biomol Chem* **2014,** *12* (7), 1135-1142.

95. Foster, A. D.; Ingram, J. D.; Leitch, E. K.; Lennard, K. R.; Osher, E. L.; Tavassoli, A., Methods for the creation of cyclic Peptide libraries for use in lead discovery. *Journal of biomolecular screening* **2015,** *20* (5), 563-76.

96. Tavassoli, A.; Benkovic, S. J., Split-intein mediated circular ligation used in the synthesis of cyclic peptide libraries in E. coli. *Nature protocols* **2007,** *2* (5), 1126-33.

97. Wu, H.; Hu, Z.; Liu, X. Q., Protein trans-splicing by a split intein encoded in a split DnaE gene of Synechocystis sp. PCC6803. *Proc Natl Acad Sci U S A* **1998,** *95* (16), 9226-31.

98. Scott, C. P.; Abel-Santos, E.; Wall, M.; Wahnon, D. C.; Benkovic, S. J., Production of cyclic peptides and proteins in vivo. *Proc Natl Acad Sci U S A* **1999,** *96* (24), 13638-43.

99. Young, T. S.; Young, D. D.; Ahmad, I.; Louis, J. M.; Benkovic, S. J.; Schultz, P. G., Evolution of cyclic peptide protease inhibitors. *Proc Natl Acad Sci U S A* **2011,** *108* (27), 11052-6.

100. Tavassoli, A.; Benkovic, S. J., Genetically selected cyclic-peptide inhibitors of AICAR transformylase homodimerization. *Angew Chem Int Edit* **2005,** *44* (18), 2760-2763.

101. Birts, C. N.; Nijjar, S. K.; Mardle, C. A.; Hoakwie, F.; Duriez, P. J.; Blaydes, J. P.; Tavassoli, A., A cyclic peptide inhibitor of C-terminal binding protein dimerization links metabolism with mitotic fidelity in breast cancer cells. *Chem Sci* **2013,** *4* (8), 3046-3057.

102. Miranda, E.; Nordgren, I. K.; Male, A. L.; Lawrence, C. E.; Hoakwie, F.; Cuda, F.; Court, W.; Fox, K. R.; Townsend, P. A.; Packham, G. K.; Eccles, S. A.; Tavassoli, A., A cyclic peptide inhibitor of HIF-1 heterodimerization that inhibits hypoxia signaling in cancer cells. *J Am Chem Soc* **2013,** *135* (28), 10418-25.

103. Tavassoli, A.; Lu, Q.; Gam, J.; Pan, H.; Benkovic, S. J.; Cohen, S. N., Inhibition of HIV budding by a genetically selected cyclic peptide targeting the Gag-TSG101 interaction. *Acs Chem Biol* **2008,** *3* (12), 757-64.

104. Accola, M. A.; Strack, B.; Gottlinger, H. G., Efficient particle production by minimal Gag constructs which retain the carboxy-terminal domain of human immunodeficiency virus type 1 capsid-p2 and a late assembly domain. *Journal of virology* **2000,** *74* (12), 5395-402.

105. Engvall, E.; Perlmann, P., Enzyme-linked immunosorbent assay (ELISA) quantitative assay of immunoglobulin G. *Immunochemistry* **1971,** *8* (9), 871-874.

106. Bally, R. W.; Gribnau, T. C., Some aspects of the chromogen 3,3',5,5'-tetramethylbenzidine as hydrogen donor in a horseradish peroxidase assay. *Journal of clinical chemistry and clinical biochemistry. Zeitschrift fur klinische Chemie und klinische Biochemie* **1989,** *27* (10), 791-6.

107. Josephy, P. D.; Eling, T.; Mason, R. P., The horseradish peroxidase-catalyzed oxidation of 3,5,3',5'-tetramethylbenzidine. Free radical and charge-transfer complex intermediates. *J Biol Chem* **1982,** *257* (7), 3669-75.

108. Jackson, M.; Mantsch, H. H., Beware of proteins in DMSO. *Biochimica et biophysica acta* **1991,** *1078* (2), 231-5.

109. Arakawa, T.; Kita, Y.; Timasheff, S. N., Protein precipitation and denaturation by dimethyl sulfoxide. *Biophysical chemistry* **2007,** *131* (1-3), 62-70.

110. Pornillos, O.; Alam, S. L.; Rich, R. L.; Myszka, D. G.; Davis, D. R.; Sundquist, W. I., Structure and functional interactions of the Tsg101 UEV domain. *The EMBO journal* **2002,** *21* (10), 2397-406.

111. Jursic, B. S.; Zdravkovski, Z., A Simple Preparation of Amides from Acids and Amines by Heating of Their Mixture. *Synthetic Commun* **1993,** *23* (19), 2761-2770.

112. Konig, W.; Geiger, R., A New Method for Synthesis of Peptides - Activation of Carboxyl Group with Dicyclohexylcarbodiimide Using 1-Hydroxybenzotriazoles as Additives. *Chem Ber-Recl* **1970,** *103* (3), 788-&.

113. Carpino, L. A., 1-Hydroxy-7-Azabenzotriazole - an Efficient Peptide Coupling Additive. *J Am Chem Soc* **1993,** *115* (10), 4397-4398.

114. Merrifield, R. B., Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *J Am Chem Soc* **1963,** *85* (14), 2149-2154.

115. Kaiser, E.; Colescott, R. L.; Bossinger, C. D.; Cook, P. I., Color test for detection of free terminal amino groups in the solid-phase synthesis of peptides. *Anal Biochem* **1970,** *34* (2), 595-8.

116. Sarin, V. K.; Kent, S. B.; Tam, J. P.; Merrifield, R. B., Quantitative monitoring of solid-phase peptide synthesis by the ninhydrin reaction. *Anal Biochem* **1981,** *117* (1), 147-57.

117. King, D. S.; Fields, C. G.; Fields, G. B., A Cleavage Method Which Minimizes Side Reactions Following Fmoc Solid-Phase Peptide-Synthesis. *Int J Pept Prot Res* **1990,** *36* (3), 255-266.

118. Nozaki, S., Effects of amounts of additives on peptide coupling mediated by a water-soluble carbodiimide in alcohols. *The journal of peptide research : official journal of the American Peptide Society* **1999,** *54* (2), 162-7.

119. Blundell, C. D.; Packer, M. J.; Almond, A., Quantification of free ligand conformational preferences by NMR and their relationship to the bioactive conformation. *Bioorg Med Chem* **2013,** *21* (17), 4976-87.

120. Karplus, M., Contact Electron-Spin Coupling of Nuclear Magnetic Moments. *Journal of Chemical Physics* **1959,** *30* (1), 11-15.

121. Karplus, M., Vicinal Proton Coupling in Nuclear Magnetic Resonance. *J. Am. Chem. Soc.* **1963,** *85* (18), 2870-2871.

122. Haasnoot, C. A. G.; Deleeuw, F. A. A. M.; Altona, C., The Relationship between Proton-Proton Nmr Coupling-Constants and Substituent Electronegativities .1. An Empirical Generalization of the Karplus Equation. *Tetrahedron* **1980,** *36* (19), 2783-2792.

123. Altona, C.; Francke, R.; de Haan, R.; Ippel, J. H.; Daalmans, G. J.; Westra Hoekzema, A. J. A.; van Wijk, J., Empirical group electronegativities for vicinal NMR proton–proton couplings along a C-C bond: Solvent effects and reparameterization of the Haasnoot equation. *Magnetic Resonance in Chemistry* **1994,** *32*, 670-678.

124. Aydin, R. G., H., 13C, 1H spin–spin coupling. X—Norbornane: A reinvestigation of the karplus curve for 3J(13C, 1H). *Magnetic Resonance in Chemistry* **1990,** *28*, 448-457.

125. Tormena, C. F., Conformational analysis of small molecules: NMR and quantum mechanics calculations. *Progress in nuclear magnetic resonance spectroscopy* **2016,** *96*, 73-88.

126. Wüthrich, K., Wider, G., Wagner, G., Braun, W., Sequential Resonance Assignments as a Basis for Determination of Spatial Protein Structures by High Resolution Proton Nuclear Magnetic Resonance. *Journal of molecular biology* **1982,** *155*, 311-319.

127. Jeener, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R., Investigation of exchange processes by two-dimensional NMR spectroscopy. *The Journal of Chemical Physics* **1979,** *71* (11), 4546-4553.

128. Kaiser, R., Use of the Nuclear Overhauser Effect in the Analysis of High-Resolution Nuclear Magnetic Resonance Spectra. *The Journal of Chemical Physics* **1963,** *39* (10), 2435-2442.

129. Bax, A., Davis, D. G., Practical Aspectsof Two-Dimensional Transverse NOE Spectroscopy. *Journal of Magnetic Resonance* **1985,** *63*, 207-213.

130. Bothner-by, A. A.; Stephens, R. L.; Lee, J.; Warren, C. D.; Jealoz, R. W., Structure determination of a tetrasaccharide: transient nuclear Overhauser effects in the rotating frame. *J. Am. Chem. Soc.* **1984,** *106*, 811-813.

131. Hwang, T., Shaka, A. J., Cross relaxation without TOCSY: transverse rotating-frame Overhauser effect spectroscopy. *J. Am. Chem. Soc.* **1992,** *114*, 3157-3159.

132. Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V., Stereochemistry of polypeptide chain configurations. *Journal of molecular biology* **1963,** *7*, 95-9.

133. Ramachandran, G. N.; Sasisekharan, V., Conformation of Polypeptides and Proteins. **1968,** *23*, 283-437.

134. Ho, B. K.; Thomas, A.; Brasseur, R., Revisiting the Ramachandran plot: hard-sphere repulsion, electrostatics, and H-bonding in the alpha-helix. *Protein Sci* **2003,** *12* (11), 2508-22.

135. Hovmoller, S.; Zhou, T.; Ohlson, T., Conformations of amino acids in proteins. *Acta crystallographica. Section D, Biological crystallography* **2002,** *58* (Pt 5), 768-76.

136. Janin, J., Wodak, S., Conformation of amino acid side-chains in proteins. *Journal of molecular biology* **1978,** *125* (3), 357-386.

137. Blundell, C. D.; Reed, M. A.; Overduin, M.; Almond, A., NMR spectra of oligosaccharides at ultra-high field (900 MHz) have better resolution than expected due to favourable molecular tumbling. *Carbohydrate research* **2006,** *341* (12), 1985-91.

138. Bundi, A.; Wüthrich, K., 1H-NMR Parameters of the Common Amino Acid Residues Measured in Aqueous Solutions of the Linear Tetrapeptides H-Gly-Gly-X-L-Ala-OH. *Biopolymers* **1979,** *18*, 285-297.

139. Chatterjee, J.; Rechenmacher, F.; Kessler, H., N-methylation of peptides and proteins: an important element for modulating biological functions. *Angewandte Chemie* **2013,** *52* (1), 254-69.

140. Chatterjee, J.; Laufer, B.; Kessler, H., Synthesis of N-methylated cyclic peptides. *Nature protocols* **2012,** *7* (3), 432-44.

141. Roodbeen, R.; Jensen, K. J., Synthesis of N-methylated peptides: on-resin methylation and microwave-assisted couplings. *Methods in molecular biology* **2013,** *1047*, 141-9.

142. Wheeler, S. E., Understanding substituent effects in noncovalent interactions involving aromatic rings. *Accounts of chemical research* **2013,** *46* (4), 1029-38.

143. Hwang, J.; Li, P.; Carroll, W. R.; Smith, M. D.; Pellechia, P. J.; Shimizu, K. D., Additivity of substituent effects in aromatic stacking interactions. *J Am Chem Soc* **2014,** *136* (40), 14060-7.

144. Ninkovic, D. B.; Vojislavljevic-Vasilev, D. Z.; Medakovic, V. B.; Hall, M. B.; Brothers, E. N.; Zaric, S. D., Aliphatic-aromatic stacking interactions in cyclohexane-benzene are stronger than aromatic-aromatic interaction in the benzene dimer. *Physical chemistry chemical physics : PCCP* **2016,** *18* (37), 25791-25795.

145. Frankel, A. D.; Pabo, C. O., Cellular Uptake of the Tat Protein from Human Immunodeficiency Virus. *Cell* **1988,** *55*, 1189-1193.

146. Lo, M. C.; Aulabaugh, A.; Jin, G.; Cowling, R.; Bard, J.; Malamas, M.; Ellestad, G., Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal Biochem* **2004,** *332* (1), 153-9.

147. Niesen, F. H.; Berglund, H.; Vedadi, M., The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature protocols* **2007,** *2* (9), 2212-21.

148. Cimmperman, P.; Baranauskiene, L.; Jachimoviciute, S.; Jachno, J.; Torresan, J.; Michailoviene, V.; Matuliene, J.; Sereikaite, J.; Bumelis, V.; Matulis, D., A quantitative model of thermal stabilization and destabilization of proteins by ligands. *Biophysical journal* **2008,** *95* (7), 3222-31.

149. Waldron, T. T.; Murphy, K. P., Stabilization of proteins by ligand binding: application to drug screening and determination of unfolding energetics. *Biochemistry* **2003,** *42* (17), 5058-64.

150. Huynh, K.; Partch, C. L., Analysis of protein stability and ligand interactions by thermal shift assay. *Current protocols in protein science* **2015,** *79*, 28 9 1-14.

151. Kim, A.; Wolf, N. M.; Zhu, T.; Johnson, M. E.; Deng, J.; Cook, J. L.; Fung, L. W., Identification of Bacillus anthracis PurE inhibitors with antimicrobial activity. *Bioorg Med Chem* **2015,** *23* (7), 1492-9.

152. Sledz, P.; Lang, S.; Stubbs, C. J.; Abell, C., High-throughput interrogation of ligand binding mode using a fluorescence-based assay. *Angewandte Chemie* **2012,** *51* (31), 7680-3.

153. Wienken, C. J.; Baaske, P.; Duhr, S.; Braun, D., Thermophoretic melting curves quantify the conformation and stability of RNA and DNA. *Nucleic acids research* **2011,** *39* (8), e52.

154. Wienken, C. J.; Baaske, P.; Rothbauer, U.; Braun, D.; Duhr, S., Protein-binding assays in biological liquids using microscale thermophoresis. *Nature communications* **2010,** *1*, 100.

155. Seidel, S. A.; Dijkman, P. M.; Lea, W. A.; van den Bogaart, G.; Jerabek-Willemsen, M.; Lazic, A.; Joseph, J. S.; Srinivasan, P.; Baaske, P.; Simeonov, A.; Katritch, I.; Melo, F. A.; Ladbury, J. E.; Schreiber, G.; Watts, A.; Braun, D.; Duhr, S., Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. *Methods* **2013,** *59* (3), 301-15.

156. Jerabek-Willemsen, M.; Wienken, C. J.; Braun, D.; Baaske, P.; Duhr, S., Molecular interaction studies using microscale thermophoresis. *Assay and drug development technologies* **2011,** *9* (4), 342-53.

157. Fossen, T.; Wray, V.; Bruns, K.; Rachmat, J.; Henklein, P.; Tessmer, U.; Maczurek, A.; Klinger, P.; Schubert, U., Solution structure of the human immunodeficiency virus type 1 p6 protein. *J Biol Chem* **2005,** *280* (52), 42515-27.

158. Good, N. E.; Izawa, S., Hydrogen ion buffers. *Methods in enzymology* **1972,** *24*, 53-68.

159. Christopoulos, A., Allosteric binding sites on cell-surface receptors: novel targets for drug discovery. *Nature reviews. Drug discovery* **2002,** *1* (3), 198-210.

160. Weber, S. J.; Abbruscato, T. J.; Brownson, E. A.; Lipkowski, A. W.; Polt, R.; Misicka, A.; Haaseth, R. C.; Bartosz, H.; Hruby, V. J.; Davis, T. P., Assessment of an in vitro blood-brain barrier model using several [Met5]enkephalin opioid analogs. *The Journal of pharmacology and experimental therapeutics* **1993,** *266* (3), 1649-55.

161. Liu, F.; Stephen, A. G.; Waheed, A. A.; Freed, E. O.; Fisher, R. J.; Burke, T. R., Jr., Application of ring-closing metathesis macrocyclization to the development of Tsg101-binding antagonists. *Bioorganic & medicinal chemistry letters* **2010,** *20* (1), 318-21.

162. Perez-Caballero, D.; Zang, T.; Ebrahimi, A.; McNatt, M. W.; Gregory, D. A.; Johnson, M. C.; Bieniasz, P. D., Tetherin inhibits HIV-1 release by directly tethering virions to cells. *Cell* **2009,** *139* (3), 499-511.

163. Chen, A. K.; Sengupta, P.; Waki, K.; Van Engelenburg, S. B.; Ochiya, T.; Ablan, S. D.; Freed, E. O.; Lippincott-Schwartz, J., MicroRNA binding to the HIV-1 Gag protein inhibits Gag assembly and virus production. *Proc Natl Acad Sci U S A* **2014,** *111* (26), E2676-83.

164. Christ, L.; Raiborg, C.; Wenzel, E. M.; Campsteijn, C.; Stenmark, H., Cellular Functions and Molecular Mechanisms of the ESCRT Membrane-Scission Machinery. *Trends Biochem Sci* **2017,** *42* (1), 42-56.

165. Carter, R. E.; Sorkin, A., Endocytosis of functional epidermal growth factor receptor-green fluorescent protein chimera. *J Biol Chem* **1998,** *273* (52), 35000-7.

166. Mistry, I. N.; Tavassoli, A., Reprogramming the Transcriptional Response to Hypoxia with a Chromosomally Encoded Cyclic Peptide HIF-1 Inhibitor. *ACS synthetic biology* **2017,** *6* (3), 518-527.

167. Townend, J. E.; Tavassoli, A., Traceless Production of Cyclic Peptide Libraries in E. coli. *Acs Chem Biol* **2016,** *11* (6), 1624-30.

168. Giaccia, A.; Siim, B. G.; Johnson, R. S., HIF-1 as a target for drug development. *Nature reviews. Drug discovery* **2003,** *2* (10), 803-11.

169. Schlake, T.; Bode, J., Use of Mutated FLP Recognition Target (FRT) Sites for the Exchange of Expression Cassettes at Defined Chromosomal Loci. *Biochemistry* **1994,** *33*, 12746-12751.

170. Broach, J. R.; Hicks, J. B., Replication and recombination functions associated with the yeast plasmid, 2 mu circle. *Cell* **1980,** *21* (2), 501-8.

171. Sauer, B., Site-specific recombination: developments and applications. *Current opinion in biotechnology* **1994,** *5* (5), 521-7.

172. Gronostajski, R. M.; Sadowski, P. D., Determination of DNA sequences essential for FLP-mediated recombination by a novel method. *J Biol Chem* **1985,** *260* (22), 12320-7.

173. World Health Organization, G., HIV Drug Resistance Report 2017. **2017**.

174. Spurr, I. B.; Birts, C. N.; Cuda, F.; Benkovic, S. J.; Blaydes, J. P.; Tavassoli, A., Targeting Tumour Proliferation with a Small-Molecule Inhibitor of AICAR Transformylase Homodimerization. *Chembiochem* **2012,** *13* (11), 1628-1634.

175. Mattissek, C.; Teis, D., The role of the endosomal sorting complexes required for transport (ESCRT) in tumorigenesis. *Molecular membrane biology* **2014,** *31* (4), 111-9.

176. Sharma, S.; Aralaguppe, S. G.; Abrahams, M. R.; Williamson, C.; Gray, C.; Balakrishnan, P.; Saravanan, S.; Murugavel, K. G.; Solomon, S.; Ranga, U., The PTAP sequence duplication in HIV-1 subtype C Gag p6 in drug-naive subjects of India and South Africa. *BMC infectious diseases* **2017,** *17* (1), 95.

177. Neogi, U.; Rao, S. D.; Bontell, I.; Verheyen, J.; Rao, V. R.; Gore, S. C.; Soni, N.; Shet, A.; Schulter, E.; Ekstrand, M. L.; Wondwossen, A.; Kaiser, R.; Madhusudhan, M. S.; Prasad, V. R.; Sonnerborg, A., Novel tetra-peptide insertion in Gag-p6 ALIX-binding motif in HIV-1 subtype C associated with protease inhibitor failure in Indian patients. *Aids* **2014,** *28* (15), 2319-22.

178. Timmins, J.; Schoehn, G.; Ricard-Blum, S.; Scianimanico, S.; Vernet, T.; Ruigrok, R. W.; Weissenhorn, W., Ebola virus matrix protein VP40 interaction with human cellular factors Tsg101 and Nedd4. *Journal of molecular biology* **2003,** *326* (2), 493-502.

179. Martin-Serrano, J.; Zang, T.; Bieniasz, P. D., HIV-1 and Ebola virus encode small peptide motifs that recruit Tsg101 to sites of particle assembly to facilitate egress. *Nature medicine* **2001,** *7* (12), 1313-9.

180. Seidel, S. A.; Wienken, C. J.; Geissler, S.; Jerabek-Willemsen, M.; Duhr, S.; Reiter, A.; Trauner, D.; Braun, D.; Baaske, P., Label-free microscale thermophoresis discriminates sites and affinity of protein-ligand binding. *Angewandte Chemie* **2012,** *51* (42), 10656-9.

181. Palencia, A.; Martinez, J. C.; Mateo, P. L.; Luque, I.; Camara-Artigas, A., Structure of human TSG101 UEV domain. *Acta crystallographica. Section D, Biological crystallography* **2006,** *62* (Pt 4), 458-64.

182. Munshi, U. M.; Kim, J.; Nagashima, K.; Hurley, J. H.; Freed, E. O., An Alix fragment potently inhibits HIV-1 budding: characterization of binding to retroviral YPXL late domains. *J Biol Chem* **2007,** *282* (6), 3847-55.

183. Eyckerman, S.; Titeca, K.; Van Quickelberghe, E.; Cloots, E.; Verhee, A.; Samyn, N.; De Ceuninck, L.; Timmerman, E.; De Sutter, D.; Lievens, S.; Van Calenbergh, S.; Gevaert, K.; Tavernier, J., Trapping mammalian protein complexes in viral particles. *Nature communications* **2016,** *7*, 11416.

184. Greulich, H.; Chen, T. H.; Feng, W.; Janne, P. A.; Alvarez, J. V.; Zappaterra, M.; Bulmer, S. E.; Frank, D. A.; Hahn, W. C.; Sellers, W. R.; Meyerson, M., Oncogenic transformation by inhibitor-sensitive and -resistant EGFR mutants. *PLoS medicine* **2005,** *2* (11), e313.