Perfluoroheteroaromatics as Tools for Peptide Modification

MOONEY, CAITLIN,ANNE

Perfluoroheteroaromatics as Tools for Peptide Modification

A Thesis Presented for the Degree of Doctor of Philosophy

Author: Caitlin Anne Mooney

Department of Chemistry, Durham University

February 2018

Publication and Conferences

The research carried out in this thesis and additional work not covered in this thesis has contributed to the publication of the following articles.

- I. R. Logan, L. Gaughan, V. Shuttleworth, L. Nawafa, **C. A. Mooney**, S. L. Cobb, N. Sheeran, The Methyltranferase SET9 Regulates TGF B-1 Activation of Renal Fibroblasts *via* Interaction with SMAD3, *J. Cell. Sci*, 2017, 131 (1)
- D. Gimenez, **C. A. Mooney**, A. Dose, G. Sandford, C. R. Coxon, S. L. Cobb, The application of pentafluoroheteroaromatic reagents in the preparation of modified peptide systems, *Org. Bio. Chem.,* 2017, **15***,* 4086-4095
- M. Jain, J. Harburn, J. H. Gill, P. M. Loadman, R. A. Falconer, **C. A. Mooney**, S. L. Cobb, D. J. Berry, Rationalised Compuer-Aided Design of Matrix-Metalloprotease-Selective Prodrugs, *J. Med. Chem*., 2017, 60(10), 4496-4502

Aspects of this work have also been presented *via* oral and poster presentations at both national and international conferences. Aspects of **Chapter 3** and **4** have been presented *via* poster presentation at Early Stage Peptide Meeting 2015 (Durham), Belgian Peptide Conference 2016 (Brussels), Chemistry and Biology of Peptides 2016 (Oxford) and 2017 (Durham). The results of **Chapter 3** and some aspects of **Chapter 2** were selected for oral presentation at SCI Postgraduate Symposium 2017 (Newcastle), RSC Chemical Biology and Bio-Organic Chemistry Postgraduate Symposium 2017 (Glasgow), Early Stage Peptide Meeting 2017 (Durham).

Abbreviations

Ac	Acetyl
Acm	Acetamidomethyl
AcOH	Acetic acid
Boc	*tert*-Butyl carbamate
CD$_3$CN	Acetonitrile-d$_3$
CDCl$_3$	Chloroform-d
DBPD	Dibromo-1,2-dihydro-pyridazine-3,6-dione
DCM	Dichloromethane
Dha	Dehydroalanine
DIPEA	*N,N*-Diisopropylethylamine
DMF	Dimethylformamide
DNA	Deoxyribonucleic acid
DPPIV	Dipeptidyl peptidase IV
EGFR	Epidermal growth factor receptor
EtOAc	Ethyl Acetate
Fmoc	Fluorenylmethyloxycarbonyl
GIP	Gastric inhibitory peptide
GLP-1	Glucagon-like peptide 1
GPRC	G-protein coupled receptor
GSH	Glutathione
HATU	1-[Bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-*b*]-pyridinium 3-oxid hexafluorophosphate
HER-2	Human epidermal growth factor-2
HFB	Hexafluorobenzene

HIV	Human immunodeficiency virus
HPLC	High pressure liquid chromatography
IgG	Immunoglobulin G
LCMS	Liquid chromatography mass spectroscopy
MALDI-ToF	Matrix assisted laser desorption ionization-time of flight
Me	Methyl
MeCN	Acetonitrile
MeOH	Methanol
MS-MS	Tandom mass spectrometry
mW	Microwave
NMM	*N*-Methylmorpholine
NMR	Nuclear magnetic resonance
Nu	Nucleophile
OBn	*O*-Benzyl ether
OT	Oxytocin
Pbf	2,2,4,6,7-Pentamethyldihydrobenzofuran-5-sulfonyl
PEG	Poly(ethylene glycol)
PEGA	Poly(ethyl glycol) acrylamide
PFP	Pentafluoropyridine
PhSNa	Sodium thiophenolate
PTM	Post translational modification
PyBOP	Benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate
Q-ToF	Quadropole time of flight
RGD	Arginylglycylaspartic acid

Rha	L-Rhamnose
S_N2	Nucleophilic substitution
S_NAr	Aromatic nucleophilic substitution
SPPS	Solid phase peptide synthesis
T2DM	Type 2 diabetes mellitus
Tat	Transactivaor of transcription
tBu	*Tert*-Butyl
TFA	Trifluoroacetic acid
TFE	Trifluorethanol
THF	Tetrahydrofuran
TIPS	Triisopropylsilane
TLC	Thin layer chromatography
Tris	Tris(hydroxymethyl)aminomethane
Trt	Triphenylmethyl
VP	Vasopressin

Contents

Abstract

The following represents a variety of projects which were undertaken throughout the course of a 3-year PhD degree course. This thesis focuses on using perfluoro-heteroaromatics as tools for peptide modification.

Chapter 2 discusses the efforts towards exploring the reactivity and modification of pentafluoropyridine (PFP) rings in terms of peptide modification. Thiol-containing dipeptides were synthesised, tagged using PFP and further functionalised using thiophenolate resulting in a surprising substitution pattern. These compounds were analysed using ^{19}F NMR, ^{1}H NMR and LCMS.

Chapter 3 focuses on the modification of Oxytocin and Vasopressin using PFP. The peptides were synthesised using Fmoc-SPPS and protecting groups exploited to tag the peptides with PFP resulting in mono- and di-tagged species. The enzymatic stability of these peptides tested, and the results discussed in Chapter 3.

The final experimental **Chapter 4** focuses on tagging longer and biologically active peptides with PFP. This chapter focuses on efforts made to synthesise Cys-modified GLP-(7-36) which can be modified using PFP and testing the stability of this compared to native GLP-1. GLP-1 is an interesting peptide target which has gained popularity recently due to the success of GLP-1 analogue Liraglutide in the treatment of diabetes.

Chapter 1: Introduction

1.1 Peptides as Drug Molecules

In general, there are two main classes of therapeutics currently marketed; small molecule drugs which are usually less than 500 Da and have good bioavailability and biologics which are typically more than 5000 Da and have poor bioavailability.[1]

Aspirin (**1**)	Herceptin® (**2**)
Mw: 180.157 Da	Mw: 145531.5 Da
Molecular formulae: $C_9H_8O_4$	Molecular formulae: $C_{6470}H_{10012}N_{1726}O_{2013}S_{42}$
Oral administration	Administration *via* injection
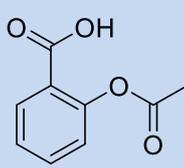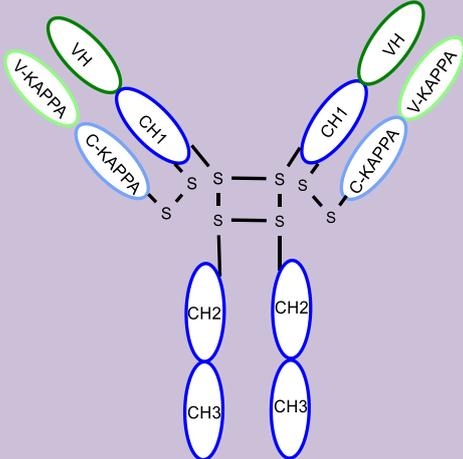
Table 1.1: i) Examples of small drug molecule: aspirin (**1**) ii) Example of a biologic drug molecule: Herceptin® (**2**).[1]

The difference in size between the two classes of therapeutic also leads to differences in mode of action and biological properties. For example, small molecule drugs often lack selectivity, leading to unwanted off-target effects, whereas biologics such as Herceptin (**2**) are much more selective towards their desired target and consequently display low toxicity.[1] Peptide therapeutics attempt to exploit the selectivity and potency of biologics but also harness the bioavailability and membrane permeability of small molecules.

It is estimated that peptide therapeutics account for 10% of the total pharmaceutical market and are worth upwards of $40 billion a year.[2] Between 1970 and 2003, the number of peptide drugs approved for therapeutic use was 30, with most of these being simple linear peptides (e.g. Lupron). In 2014/2015 the US FDA approved nine peptide drug candidates (**Table 1.2**) highlighting the increased interest in this area. In addition,

in 2015, there were 400-600 peptide drug candidates in various stages of clinical development with a variety of biological applications such as cancer treatments, cardiovascular treatments, pain medication and allergy treatments.[4]

Peptide name	Brand name (Manufacturer)	Treatment	Date of approval (FDA)
Metreleptin	Myalept (Astra Zeneca)	Complications arising from leptin deficiency	Feb. 2014
Albiglutide	Albiglutide (GSK)	Type 2 diabetes- GLP-1	Apr. 2014
Dulaglutide	Trulicity (Eli Lilly)	Type 2 diabetes- GLP-1	Sep. 2014
Insulin degludec	Tresiba (Novo Nordisk)	Type 2 diabetes GLP-1	Sep. 2015
Parathyroid hormone	Natpara (NPS. Pharma)	Hypocalcemia treatment for patients with hypoparathyroidism	Jan. 2015

Table 1.2: Examples of peptide therapeutics gaining market approval 2014/2015.[4]

2017 represented a particularly successful year for the peptide therapeutic market with the GLP-1 agonist Ozempic (for the treatment of Type 2 diabetes) and Tymlos (for the treatment of Osteoporosis) being approved for commercial distribution.[4] The emergence of peptides as viable therapeutics in recent years lends itself to the development of more effective methods of peptide synthesis and better understanding the stability of peptides within biological systems.[5]

There are multiple areas of peptide drug development research including antimicrobial peptides (AMPs) and G-protein-coupled receptor-targeting peptides (GPCR). Approximately 39% of peptides in clinical trials are GPCR-targeting peptides which is mainly due to the diversity of GPCRs within the body. GPCRs are a family of cell surface receptors which play important roles in the regulation of cell growth, metabolism and cell signalling. While GPCRs are vital in regulating normal bodily functions such as the immune system and the nervous system, abnormal GPCR function can cause tumour cell growth and metastasis. As **Table 1.2** shows, one of the most interesting GPCR targets for drug research is Glucagon-like Peptide-1 (GLP-1) and Glucagon-like Peptide-2 (GLP-2) which will be discussed in **Chapter 4.** [5]

AMPs represent an important field of peptide therapeutics due to the occurrence of antimicrobial drug resistance. As of 2014, there were 10 AMPs at various stages of clinical development against a range of microbial targets including parasites and bacteria.[5]

PPI's (Protein-Protein Interactions) are responsible for a large range of biological processes within the body ranging from catalysis to cell division.[6] Consequently, the ability to disrupt or modulate PPIs offers the opportunity to provide highly selective treatment of a broad range of diseases including cancer.[7] Finding molecules that can target PPIs with high affinity has proven to be challenging given that they often require binding to a large, relatively flat protein surface. While some small molecules have been developed to act as potent PPI inhibitors,[8] in recent years larger molecules, such as peptides, have seen increasing interest within both the academic and industrial communites.[9]

1.1.2 Challenges with Peptide Drug Development

Historically the use of peptides as therapeutics has been challenging, mainly due to poor bioavailability and poor stability within biological systems.[10] Due to their size and hydrophilicity it is often difficult for peptide drugs to cross cell membranes including the blood brain barrier and this affects a drug's ability to interact with its desired target.[11] Once administered, peptides are often rapidly degraded by proteases which recognise certain residues or sequences of residues within the peptide. Once the peptide chain is degraded or truncated it normally loses its biological activity.

Peptide Stability

Peptides are susceptible to chemical degradation, for example peptides containing Cys and Met residues readily undergo oxidation (as seen below), which in the case of Met, is very difficult to reverse (**Scheme 1.1i**) and the oxidation of Cys can result in the formation of disulphide bonds (**Scheme 1.1ii**). Asp-containing peptides can lead to aspartimide formation and subsequent altered peptide sequences with the formation of β-peptide bonds and peptide bond degradation (**Scheme 1.1iii**).
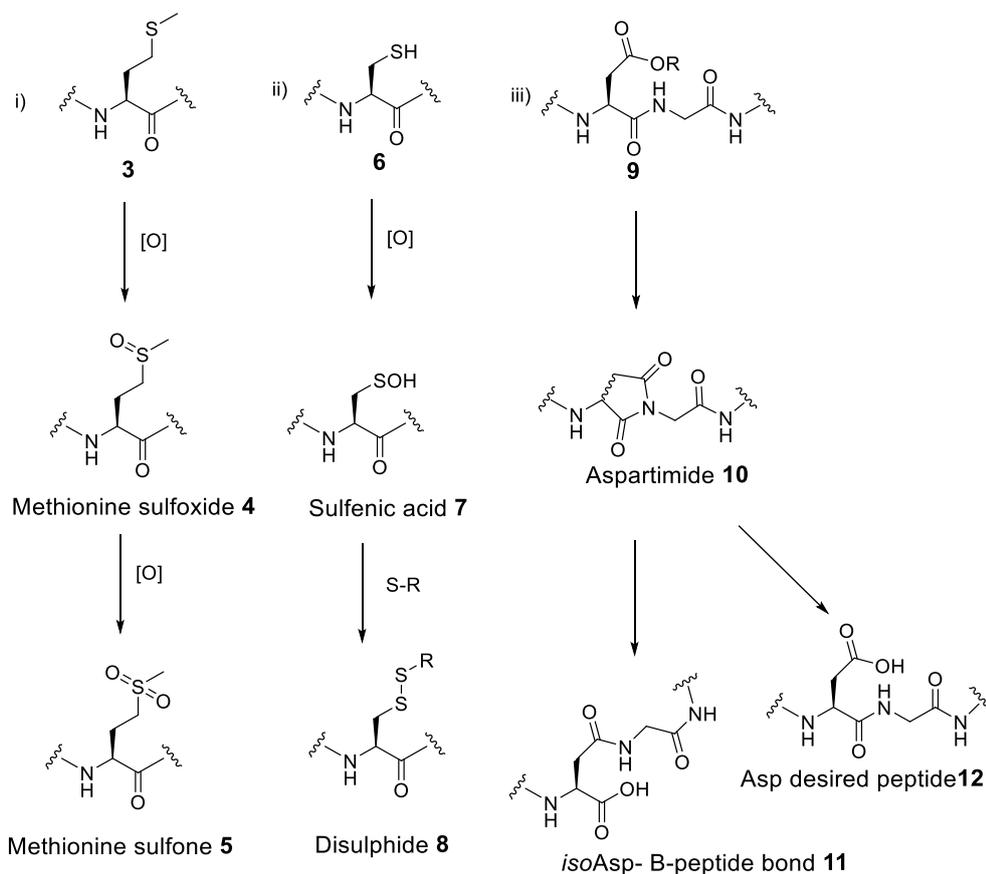
Scheme 1.1: Potential chemical degradation pathways: i) oxidation of methionine (**3**) to methionine sulfone (**5**) ii) oxidation of cysteine (**6**) to sulfenic acid (**7**) and disulphide bridge (**8**) and iii) formation of aspartimide intermediate (**10**) resulting in β-aspartic acid (**11**).[12]

Recent studies have sought to improve peptide stability by modifying residues within a peptide sequence.[13] Several studies in this area have focused on the substitution of both natural and non-natural amino acids into the native sequence of a biologically active peptide. Deacon and co-workers[13] demonstrated that by carrying out substitutions at position 8 of GLP-1(7-36), the properties of the original peptide could be modulated (**Table 1.3**). For example, substitution with various natural amino acids including, serine (**Entry 2**), threonine (**Entry 3**), glycine (**Entry 4**) and amino-isobutyric acid (**Entry 5**), increased the stability of the GLP-1 analogues against DPPIV.

Entry	Structure	Modification
1	HAEGTFTSDVSSYLEGQAAKEFIAWLVKGR	Native GLP-1
2	H**S**EGTFTSDVSSYLEGQAAKEFIAWLVKGR	Ser substiution
3	H**T**EGTFTSDVSSYLEGQAAKEFIAWLVKGR	Thr substitution
4	H**G**EGTFTSDVSSYLEGQAAKEFIAWLVKGR	Gly substitution
5	H**Aib**EGTFTSDVSSYLEGQAAKEFIAWLVKGR	Aib substitution

Table 1.3: Modification of native GLP-1 resulting in increased stability to enzymatic breakdown.

They found that when substituted with serine, threonine or glycine, the enzymatic cleavage rates were approximately 5% less than the native GLP-1 and when substituted with amino-isobutyric acid (Aib) there was no detectable degradation.[13] Further strategies that have been utilised for GLP-1 modification will be discussed in more detail in **Chapter 4**.

AMPs have provided many lead compounds in recent years for the design and development of new peptide antibiotics.[14] Researchers have explored the possibility of preparing all D-amino acid isomers of natural peptides as a means to increase stability as well as reduce toxicity. For example, Carmona and co-workers[15] investigated the biological activity and stability of the antimicrobial peptide, Pin-2 (FWGALAKGALKLIPSLFSSFSKKD). In this study, the 24 residue polycationic Pin-2 peptide was synthesised in both the D- and L-isomeric forms and tested against a variety of bacterial strains for biological activity and against bovine pancreatic trypsin and human elastase for proteolytic stability. In the stability studies against these two enzymes, the D-analogue was not degraded while the L-analogue was digested to smaller peptide fragments. In studies using human serum, the D-analogue was found to degrade at a reduced rate compared to the L-analogue (**Figure 1.1**).[15]
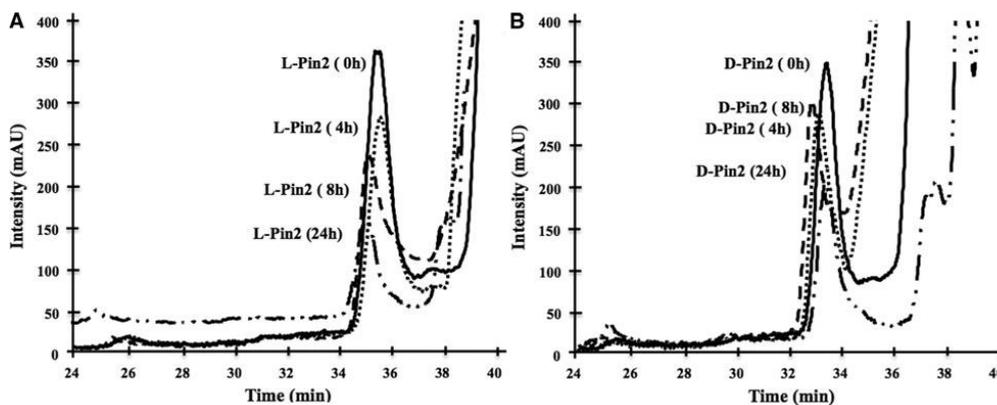
Figure 1.1: Analytical HPLC traces of stability profile of A) L-Pin-2 and B) D-Pin-2, to human serum at t = 0, 4, 8 and 24 h. (Figure reproduced with permission).

As demonstrated by the case of Pin-2, conversion of peptides from their natural L-amino acid form to their D-amino acid analogues can provided an effective method for improving the stability of a peptide. The disadvantages of such an approach are that it results in a more expensive peptide and will only work in cases where the peptide does not bind to a chiral target.

Peptide Targeting and Cellular Delivery

As well as the challenges presented by an inherent susceptibility towards enzymatic degradation, poor cell permeability due to their hydrophilic nature also presents a hurdle to overcome in peptide drug development. Until recently, the concept of delivering peptide drug cargo into cytoplasmic and nuclear components of a cell was under investigated and seemed an unattainable target. Translocating the cellular membrane has proven a problem with conventional drug discovery and often means that large quantities of a drug must be administered, resulting in side effects and toxicity.[16]

There have been a number of studies carried out to investigate the translocation of peptide fragments across cellular membranes[17] and from this work short peptide fragments that could cross cellular membranes without assistance from transporter proteins were identified. This group of peptides are now known within the community as a "cell penetrating peptides" (CPP). Since their discovery CPPs have been exploited as drug delivery agents given their ability to penetrate cells and deliver drugs to intracellular targets.[18]

An example of one of these CPP is the Tat-derived peptide of the HIV-1 virus.[19] The peptide itself is synthesised in HIV-infected cells and is responsible for HIV cell replication. The cell penetrating part of the peptide is only 10-16 amino acids long and contains 6 Arg residues so is overall cationic.[20] These Arg (and also Lys residues to a lesser extent) play a pivotal role in Tat-peptide cell penetrating abilities[19] and are thought to function by forming bidentate hydrogen bonds between the guanidinium groups of Arg and phosphates present in cell membranes.[21] There have been a number of studies concentrating on conjugating Tat fragments to a variety of biomolecules including proteins.[22] Dowdy and co-workers synthesised a Tat-caspase 3 conjugate which was shown to selectively target HIV infected cells. Once within the infected cells, HIV protease cleaves the Tat peptide from the peptide-protein conjugate resulting in active caspase-3 to effectively induce apoptosis.[23]

A related field of peptide research is cell targeting peptides (CTP's) which are small peptides that are cell specific and have a high affinity for their target. **Table 1.4** illustrates peptides sequences-including their lengths- which are shown to target specific tissues. [2]

Peptide sequence	Sequence length	Target tissue
TSPLNIHNGQKL	12	Head and neck solid tumours (Andrea, 2006)
CGKRK	5	Tumour neovasclature (Portoghese, 1982)
CGNKRTRGC	7	Breast tumours (Ye, 2006)
RGD	3	αVβ3 (Dijjkgraaf, 2007)
VHSPKNN	7	Endothelial VCAM-1 expressing cells (Harris , 2007)
EDYELMDLLAYL	12	Various tissue tumours (Janssen, 2002)
LTVSPWY	7	Breast tumours (Dromi, 2007)

Table 1.4: Cell targeting peptide (CTP) sequences, lengths and targets. [18]

The concept of CTPs has been applied to cancer research and an example of this is Arginylglycylaspartic acid (RGD). RGD analogues are used to target glycoprotein integrin $\alpha_v\beta_3$/ $\alpha_v\beta_5$ which is expressed in tumour cells. Integrins are a family of transmembrane glycoproteins that are expressed in a variety of cells and play a role in a number of physiological processes but integrin $\alpha_v\beta_3$ plays a significant role in tumour growth. Chen and co-workers[24] synthesised a cyclic RGD peptide linked to a ^{18}F-fluorobenzoate tracer *via* the amino moiety of the lysine residue (**Figure 1.2**). They found that using the ^{18}F-labelled RGD adduct (**13**) allowed for imaging of the selective inhibition of tumorous tissue cells compared to healthy tissue.
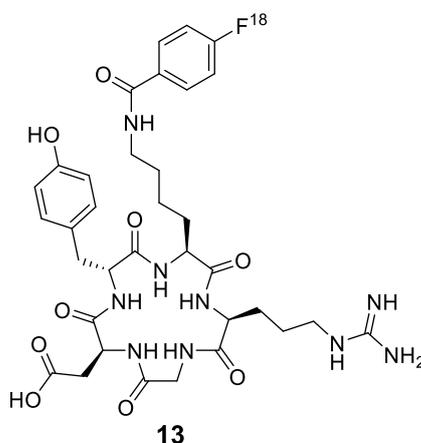
Figure 1.2: The structure of a cyclic RGD peptide labelled with [^{18}F]-fluorobenzoate (**13**).

1.2 Peptide Conjugate Applications

A variety of strategies have been used to improve the pharmokinetic properties of peptides including conjugation to small molecular probes, polyethylene glycol (PEG) groups and carbohydrates.[5]
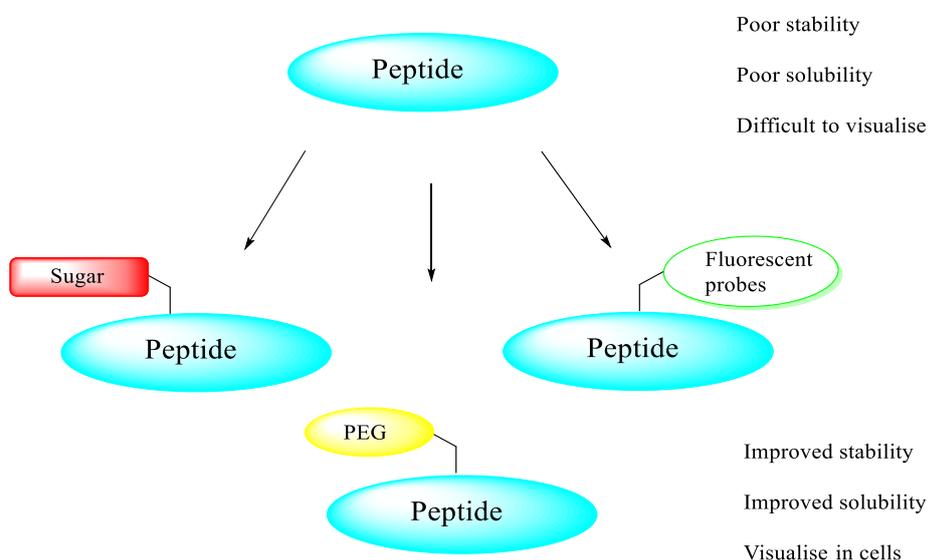
Scheme 1.2: Native peptide exhibiting poor physiochemical properties and peptides modified with sugars, PEG and fluorescent probes to improve peptide properties.

For example, incorporating sugars or lipids can improve the stability of a peptide and allow for drug delivery, but incorporating electron deficient alkynes into a peptide and coupling to fluorescent proteins can be used in imaging studies.[25] Peptides are also increasingly being conjugated to nanoparticles including silica nanoparticles,[26] gold nanoparticles[27] and quantum dots.[28] Bioconjugation to molecules such as these can improve targeting and cell penetration of a peptide.

1.2.1 Improving Lipophilicty

It has been previously discussed that poor oral bioavailability plays a crucial role in the breakdown of peptides within a biological system.[29] Therefore, improving the bioavailability of peptides is an effective strategy for ensuring that a peptide reaches its desired target.[30] Strategies for the addition of aliphatic moieties (e.g. fatty acid conjugation)[31] to a peptide system have been developed and have been shown to improve the lipophilicity of peptides.

Additions of PEG groups have also been shown to increase the stability of a peptide.[32] PEGylation is shown to dramatically improve various pharmacokinetic properties of peptides due to the alteration of binding patterns compared to non-PEGylated peptide therapeutics and has also shown to improve absorption rates due to reduced breakdown of the peptide in the renal system.[33] This provides promise in using PEGylation as a method to improve peptide properties and make it a more viable therapeutic option. This concept will be discussed using GLP-1 as an example in **Section 4.1** "Case Study: GLP-1: Improving Pharmokinetics *via* Bionconjugation".

Investigations into cyclic peptides have found that the lipophilicity of cyclic peptides are increased compared to their linear counterparts. This is due to the masking of the C- and *N*-terminus charges which protects the peptide from site specific degradation and improves cell permeability of the cyclic peptide.[34] Cilengitide (**14**) is a small RGD pentapeptide which is currently in clinical trials for the treatment of glioblastomas and has been the subject of bioconjugation studies in the past to improve lipophilicity.[35] Analogues of Cilengitide have been synthesied where the *N*-Me group was replaced with PEG chains of various sizes (**15** and **16**).

14   **15**

16
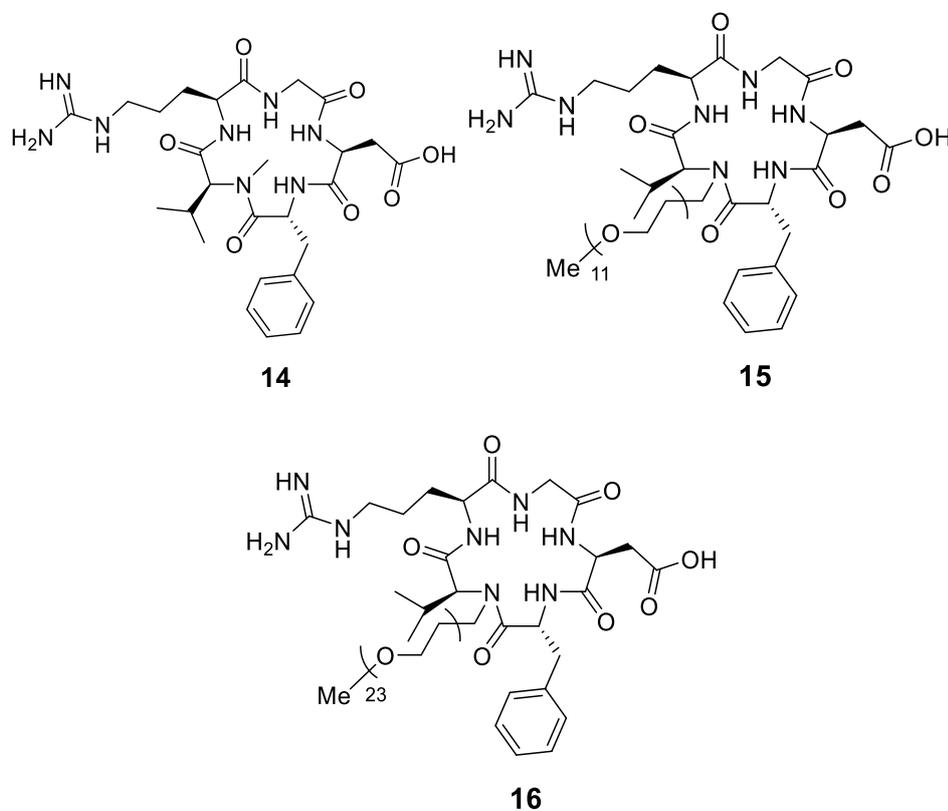
Figure 1.3: Structure of cilengitide (**14**), *N*-PEG$_2$ substituted cilengitide (**15**) and *N*-PEG$_{23}$ substituted cilengitide (**16**).[35]

It was found that by incorporating a short polyethylene glycol (PEG) chain into the cyclic peptides, as in compound **15**, the lipophilicity was improved compared to Cilengitide (**14**) with similar biological activity. When longer chains, as in compound **16**, were incorporated, the lipophilicity was improved, but biological activity was lower compared to the native peptide due to steric hindrance.[36]

1.2.2 Improving Cell Permeability

Using a targeting group within drug design allows for direct delivery of the drug molecule to the desired target, improved cell permeability and reduced toxicity within a biological system.

There have been a number of studies in recent years investigating the effect of peptide conjugation to nanoparticles such as silica[37] and gold.[38] Conjugation of Tat-peptide, which is a HIV-derived peptide, to mesopourous silica nanoparticles (MSN) was shown to target cell nuclei and successfully allowed for the delivery of DOX (doxorubicin) - an anticancer drug- into the nucleus.[39] MSN are nanomaterials which have proven effective at delivering drug cargoes due to their large surface areas, pore sizes and high stability while Tat-peptide has been shown to translocate cellular membranes. By combining these with a DOX drug cargo, the peptide-drug conjugate can cross cellular membranes and exert anticancer properties.[39]

Bioconjugation of peptides to gold nanoparticles (GNP) has become a well-documented method for synthesising cell permeable and potent drug candidates for the treatment of diseases such as cancer, as well as biosensing and diagnostics.[40] GNP are biologically stable and therefore non-toxic and easy to synthesise. Due to GNP reactivity, thiol and amine containing peptides can be conjugated to the GNP surface and delivered selectively to diseased cells. Rotello and co-workers[41] synthesised a variety of cationic tetra-alkyl ammonium modified GNP (**17**) to target the enzyme β-galactosidase (**Figure 1.4**).

17
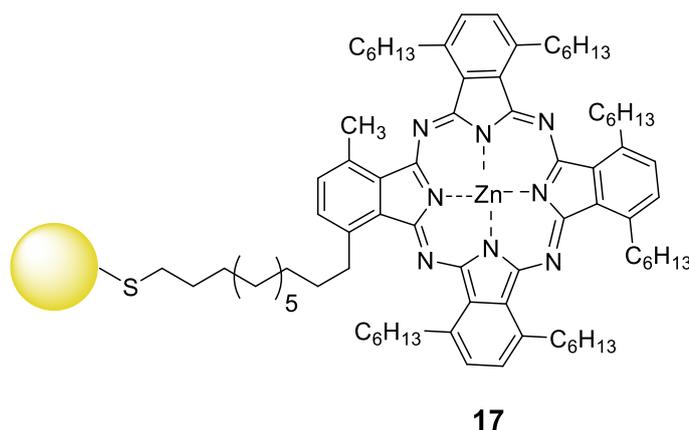
Figure 1.4: Tetra-alkyl ammonium modified GNP (**17**) developed by Rotello.[41]

This inhibition was reversed upon addition of glutathione (GSH), which is a reducing agent and functions by reducing disulphides to free cysteine and has potential to act as a releasing agent within cells due to differing concentrations of GSH inside and outside of a cell.[41] GNP-peptide analogues have become a well-established and successful method for improving cell permeability[42] and targeting.[43]

1.2.3. Molecular Imaging

Molecular imaging has proven a vital diagnostic tool for a variety of diseases such as cancer and Alzheimer's disease and has also proven to be an invaluable tool in drug development. Molecular imaging in drug development has allowed for determination of the mechanism of action of a peptide, determination of the target site of various molecules within a biological system as well as any breakdown products that may be produced.[44]

Tagged peptides have interesting properties which allow for imaging experiments and have proven an effective way to visualise the reaction of peptides within a biological system. They not only allow for the visualisation of an interaction but also quantification of an interaction between peptide and its target which can give an indication of the efficacy of the peptide. The peptide in question can be tagged by a variety of substrates including fluorescent tags such as fluorescein isothiocyanate (FITC) (**Scheme 1.3**) and phosphorescent tags such as iridium complexes (**Scheme 1.4**).

Fluorescein isothiocyanate (FITC) is a fluorescein analogue which is used routinely in a variety of labelling studies such as fluorescent microscopy, equilibrium studies and flow cytometric studies. [44] Due to its versatility, FITC (**18**) has featured in many imaging studies including animal and plant endocytosis studies. The isothiocyanate group of the fluorophore reacts with free amine groups and can therefore conjugate to peptides containing an amine in their sidechains. Zheng and co-workers used FITC coupled to RGD tripeptide (**Scheme 1.3**) to stain tumour cells. [45]
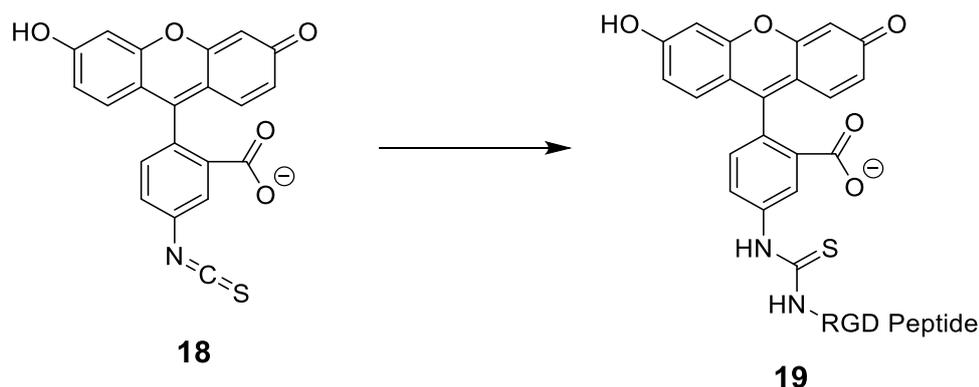
Scheme 1.3: Bioconjuagtion of RGC pepide and FITC. [45]

Techniques used to image integrin $\alpha_v\beta_3$ have included immunohistochemical staining with anti-integrin antibodies, but this proved to be a complicated method of imaging due to the instability of the antibodies in aqueous solution. Zheng and co-workers[45] were able to synthesis cyclic RGD peptides conjugated to FITC which were proven to selectively and efficiently target integrin $\alpha_v\beta_3$/ $\alpha_v\beta_5$.

The iridium complex (**21**) below has proven an effective tag in phosphorescence studies due to long half-life, stability and a substantial difference between absorption and emission wavelengths.[46] The alkynoic moiety of the complex reacted specifically with free cysteine, leaving unprotected serine and lysine unchanged, to give a phosphorescent peptide system conjugated by a vinyl sulphide linkage (**Scheme 1.4**).
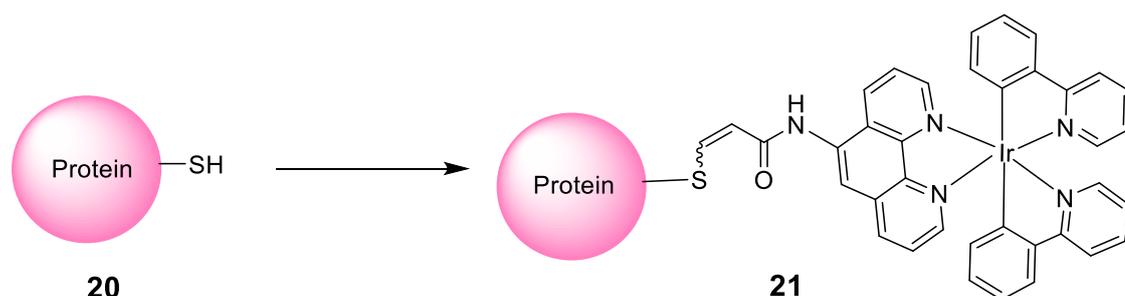
Scheme 1.4: Selective modification of cysteine containing protein using [Ir(ppy)$_2$(phen-propiolamide). [47]

Shui and colleagues[47] demonstrated the specificity of this complex by coupling the phosphorescent tag with a BCArg protein (**20**) which has been shown to exhibit anti-cancer activity and has a variety of unprotected side chains such as Lys, Tyr, His and Met and a single free Cys. The phosphorescent tag attached specifically at the free cysteine and exhibited a longer half-life and greater stability compared to fluorescent tags which allowed a distinction to be made between the tagged peptide and scattering light from the sample.[47]

1.3   Preparative Methods of Peptide Conjugates

1.3.1   Bioconjugation at Cysteine (Cys)

Cys is present in a range of proteins as it plays an important role in a large number of biological processes ranging from metabolism to apoptosis.[48] Given its high reactivity many research groups have focused on developing approaches that would allow the chemical modification of Cys residues within peptide and protein systems. Many of the approaches look to exploit the nucleophilic nature of the thiol within the Cys residue (**Figure 1.5**). Tagging studies based on Cys have previously concentrated on the synthesis of disulphides and the reaction of Cys with Michael acceptors (**Scheme 1.5i**).[49]
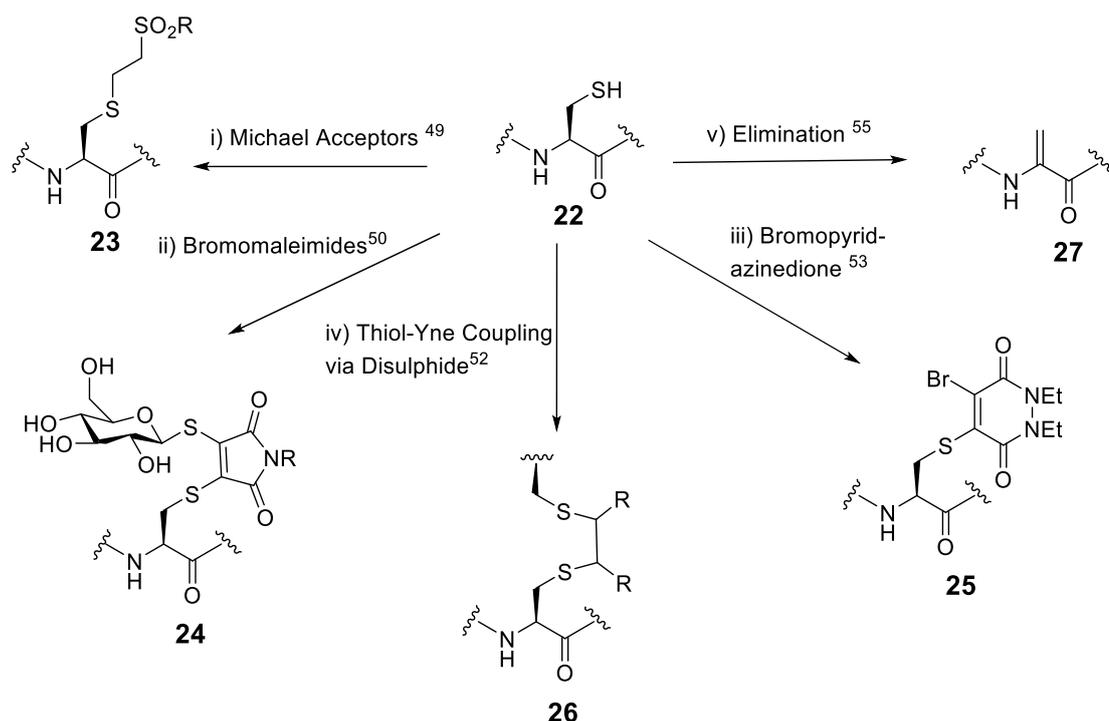
Scheme 1.5: An overview of Cys bioconjugation approaches.

Ryan *et. al*[50] focused on the bioconjuagation of Cys containing peptides with bromomaleimides (**Scheme 1.5ii**) and found that the reversibility of these conjugates could be controlled depending on reaction conditions. In theory this approach could be applied to peptide therapeutics which could allow for the conjugation of peptides to thiosugars which would have greater stability within a biological system. A popular method adopted for Cys modification is the use of disulphides in bioconjugation. A simple and effective method for the synthesis of disulphides is the use of disulphide exchange reagents such as pyridyl dithiol. This reaction can be carried out over a wide pH range and the resulting leaving group is stable, therefore driving disulphide exchange to completion.[51] Griebrenow and co-workers[52] were able to synthesise alkylated Terlipressin analogues which were accessed *via* "rebridging" of the disphulphide bond using substituted alkynes. These reactions were selective for the Cys over other nucleophilic side chains and allowed the synthesis of PEGylated Terlipressin analogues.

Chudasama[53] investigated the modification of Cys with bromopyridazinedione (**Scheme 1.5iii**) and found that these bioconjugated peptides were more stable and the reaction specific for Cys. Bromopyridazinedione (**29**) was used to conjugate thio-glucose (**31**) to cysteine and disulphide containing peptides/proteins (**28**) (**Scheme 1.6**).
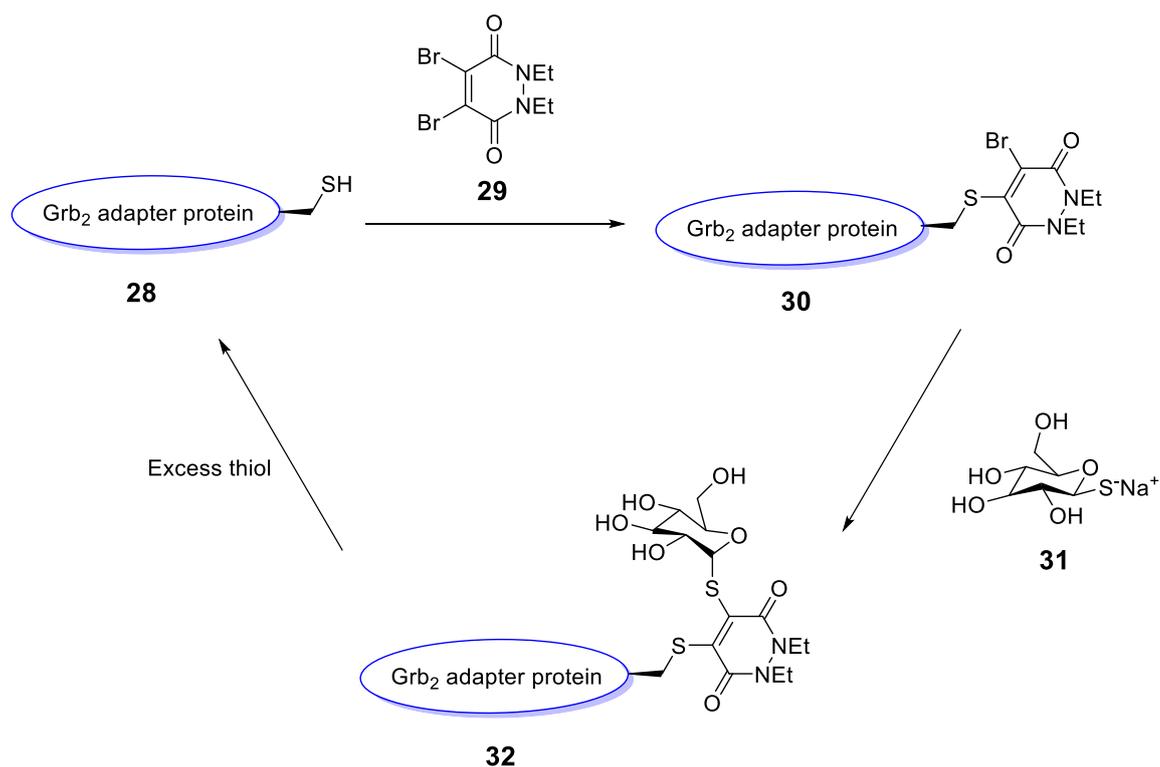
Scheme 1.6: Bromopyridazinedione-mediated peptide modification of Grb2 adapter protein (**28**).[53]

In this study Chudasama and co-workers used a Cys mutated variant of the SH2 domain of Grb2 adapter protein (**28**) and when reacted with dibromo-1,2-dihydro-pyridazine-3,6-diones (**29**) (DBPD), the bromine is selectively displaced by the free thiol of the mutated protein resulting in a mono-brominated intermediate (**30**). Thioglucose (**31**) is added to the intermediate which displaces the second bromine of the DBPD to give a protein-sugar conjugate (**32**), exhibiting increased hydrolytic stability compared to the native protein. In an excess of 2-mercaptoethanol (BME), a thiol alcohol, the conjugate is disassembled to yield the starting protein (**Scheme 1.6**). To further probe protein-conjugate release, glutathione (GSH), which is found in various concentrations in a number of living cells, was added to the protein-conjugate, resulting in disassembly and release of the free protein.[53] This gave the peptide-conjugate scope to be used in drug delivery to selectively deliver a therapeutic peptide to its target with reduced enzymatic breakdown.

Recent work from the Chudasama group has concentrated on using pyridazinediones (PD) to access disulphide bonds for modification.[54] They focused on the modification of the antibody Herceptin (**33**) with anti-cancer compound monomethyl aurisatain E (MMAE) using pyridazinediones. Analogue **34** is an ADC with a MMAE payload conjugated *via* a non-cleavable linker and compound **35** utilises a cathepsin B labile cleavable linker (**Scheme 1.7**).
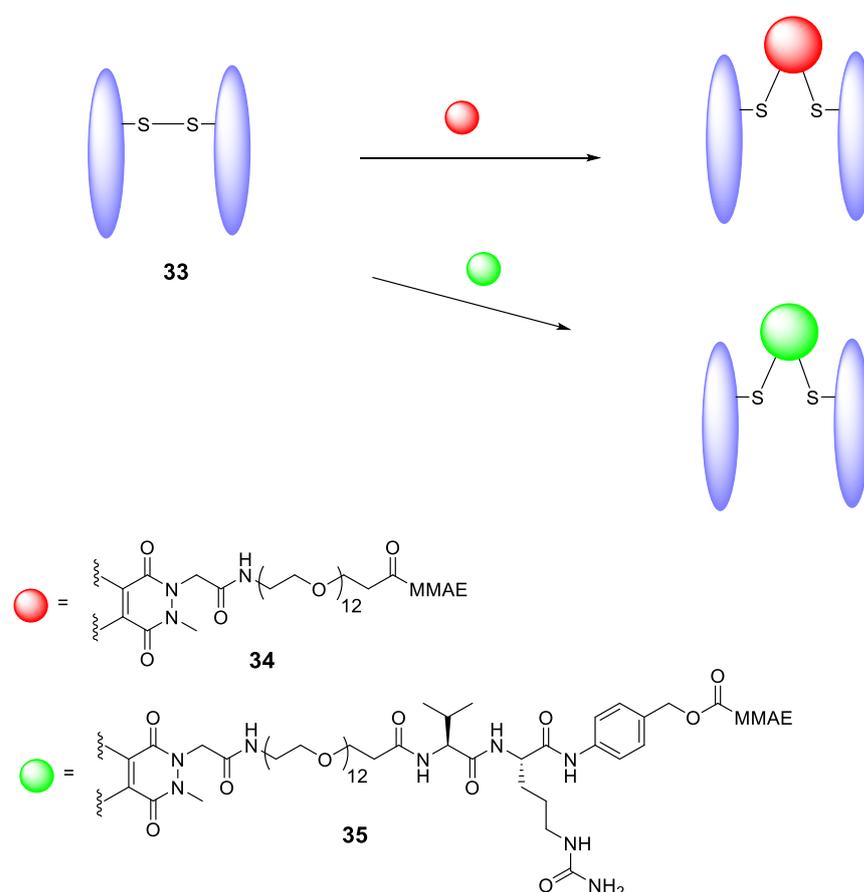
Scheme 1.7: Formation of Herceptin-MMAE conjugated *via* PD with cleavable linker (**35**) and non-cleavable linker (**34**).[54]

Analogues **34** and **35** were found to be stable in human serum and also exhibited activity against HER2-positive cell lines in *in vitro* assays and mouse xenograph models. These analogues were particularly active due to the fact they have the ability to conjugate MMAE to 4 different disulphide bridges and deliver a high drug payload to its target.[53] Davis and co-workers[55] have investigated a number of strategies for Cys specific tagging including the synthesis of selenocysteine conjugates, elimination to produce dehydroalanine (Dha) (**Scheme 1.5v**) and nucleophilic and electrophilic disulphide formation. By adopting these methods, Davis and co-workers were able to conjugate various groups such as thiosugars and thioprenyls into a range of peptide and

protein systems in a selective manner. The modification of Cys *via* Dha (**27**) (**Scheme 1.8**) has been an interesting and well documented method for incorporation of sugars and imaging labels into a peptide sequence selectively through Cys.[53]
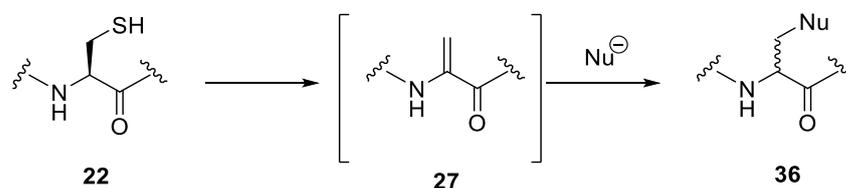
Scheme 1.8: Bioconjugation of Cys (**22**) *via* Dha intermediate (**27**) (Nu= PEG, glyscosugars, imaging labels etc).[56]

This method has allowed for the synthesis of a diverse range of modified Cys residues including modified histones, PEGylated proteins and glycoproteins.[57] In the studies involving modified histones it was established that Cys modification *via* Dha allowed for the production of a number of new analogues with interesting activities in imunoblot and enzymatic assays.[57] Post-translational modifications (PTM) of histones occur at a number of different reactive sites, including alkylation of Lys residues and are responsible for a number of important biological functions including DNA replication. Davis and co-workers aimed to alkylate histone Cys residues to produce a variety of histone mimetics *via* a Dha intermediate (**Scheme 1.10**).

Treatment of Cys modified H3 histone **37** with dibromide **38** resulted in the formation of the mono substituted compound **39**. Further incubation of intermediate **39** resulted in the cyclised intermediate **40**, followed by the elimination of the substituted furan leaving group, to give the Dha containing histone **41** (**Scheme 1.9**).
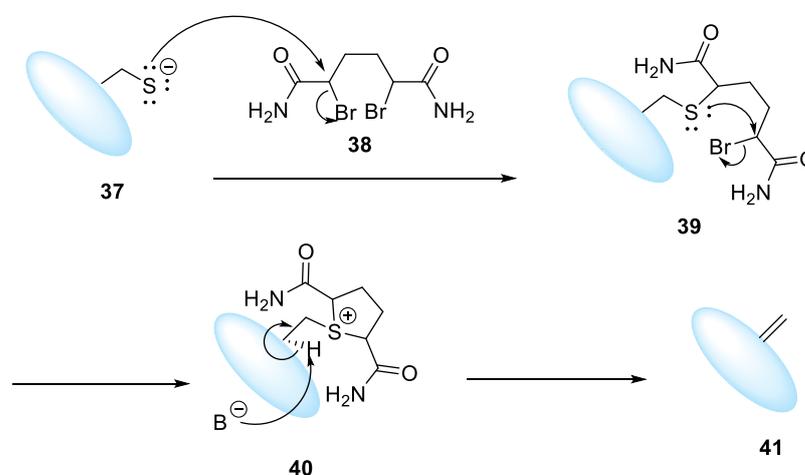
Scheme 1.9: Mechanism of formation of Dha from Cys *via* dibromide **38**.

The histone used in this study was a H3 mutant (**37**) with a Cys at position 110 and was found to be converted to Dha using the dibromide shown in **Scheme 1.10**. The Dha intermediate (**41**) did not react with Ellmans reagent, confirming the conversion of Cys to Dha. The Dha intermediate was then reacted with a variety of thiols (**42, 43, 44** and **45**) resulting in a series of methylated lysine mimics.[57]
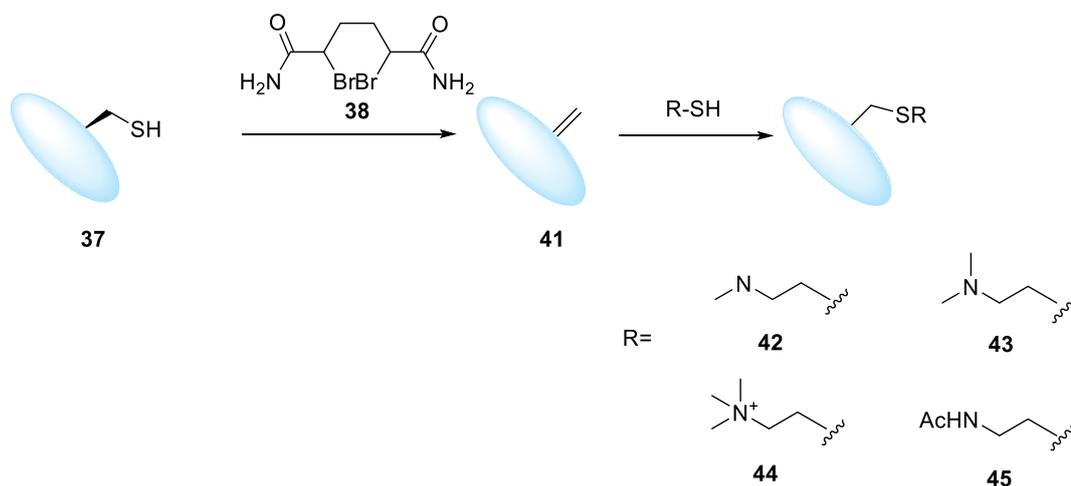
Scheme 1.10: PTM of Cys (**37**) residues utilising Dha intermediate (**41**).[57]

The resulting mono-(**42**), di-(**43**), and tri-methylated Lys (**44**) and acylated Lys (**45**) analogues were shown to exhibit activity in immunoblot and enzymatic assays as well as selectivity towards the antibody used in the study. As well as the methylated Lys analogues, Davies and co-workers[57] also synthesised phosphorylated Ser and glycosolated Ser analogues by utilising the reaction condition described above, *via* the Dha intermediate. These modified histone analogues also exhibited activity in immunoblot and enzymatic assays as well as selectivity towards the antibodies tested.[57]

The unique reactivity of hexafluorobenzene (HFB) (**47**) was utilised by Spokoyny *et al.*[58] in which substitution of HFB (**47**) using the thiols of Cys residues as the nucleophile was achieved. It was determined that HFB (**47**) could react selectively with Cys to form the 1,4-disubstituted aromatic product (**49** and **50**). A useful feature of this reaction is that it can be followed using ^{19}F NMR spectroscopy to observe the formation of products.
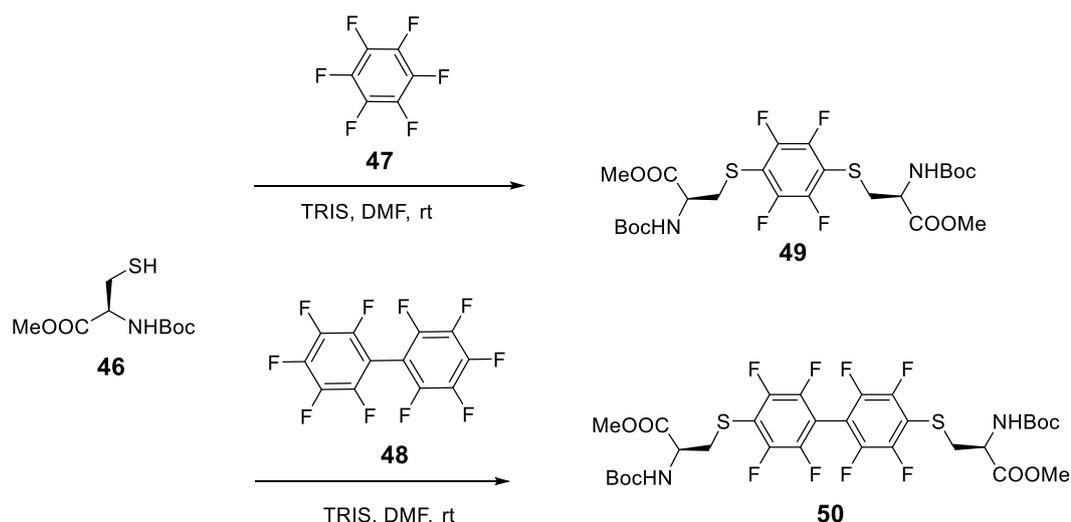
Scheme 1.11: The model preparation of the 1,4-disubstituted perfluoroaryls, **49** and **50**, from **47** and **48** respectively.

The reactivity of HFB (**47**) with Cys residues inspired the investigation of HFB for peptide stapling. The reaction allows for quick and efficient synthesis of stapled peptides without the synthesis of other isomers (**Scheme 1.11**).[59, 60]

Further work carried out by Pentalute and co-workers demonstrated the use of the 'π-clamp' to induce the selective modification of cysteine.[61] The 'π-clamp' is a sequence of aromatic amino acid residues (H$_2$N-FCPF-R) (**51**) which can be incorporated into a protein and increases the reactivity of the Cys in the sequence. The incorporation of this small peptide sequence into a protein allows for the selective modification of Cys using a decafluorodiphenyl derivative (**52**).[61] (**Scheme 1.12**)
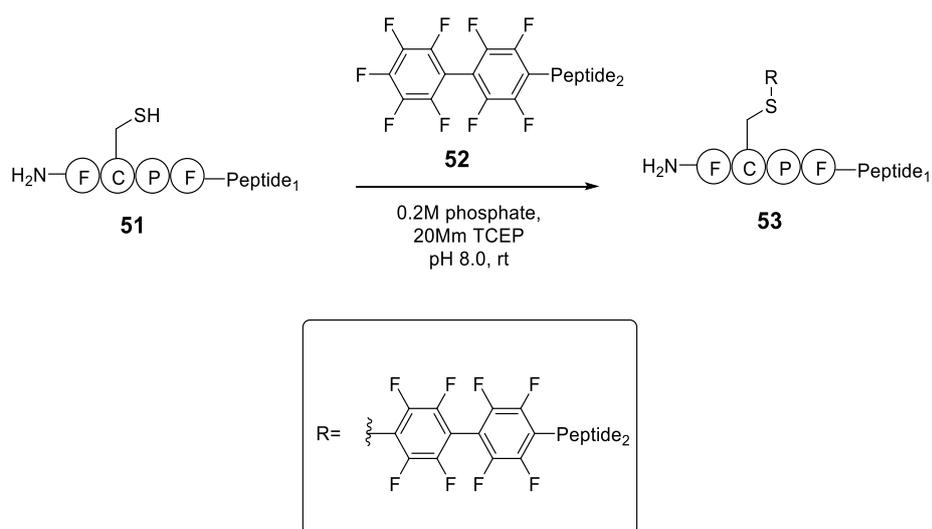
Scheme 1.12: Reaction of π-clamp peptide (**51**) with decafluorodiphenyl derivative (**52**).[61]

33

1.3.2   Bioconjugation at Lysine (Lys)

Targeting Lys is also a popular method for bioconjugation because, as with Cys, Lys has a nucleophilic side chain. The nature of the Lys side chain, a free primary amine, allows a degree of selectivity can be applied when modifying peptides with multiple active sites, with an exception of the peptide *N*-terminus.

Modification at Lys can be carried out using a range of reaction conditions to give a variety of products for example, the simple reaction between isothiocyanate and the free amine of Lys to give a thiourea product (**Scheme 1.13i**).[62] These can act as a precursor to further modifications such as incorporation of alkyl groups or a method to conjugate imaging labels into the peptide backbone.
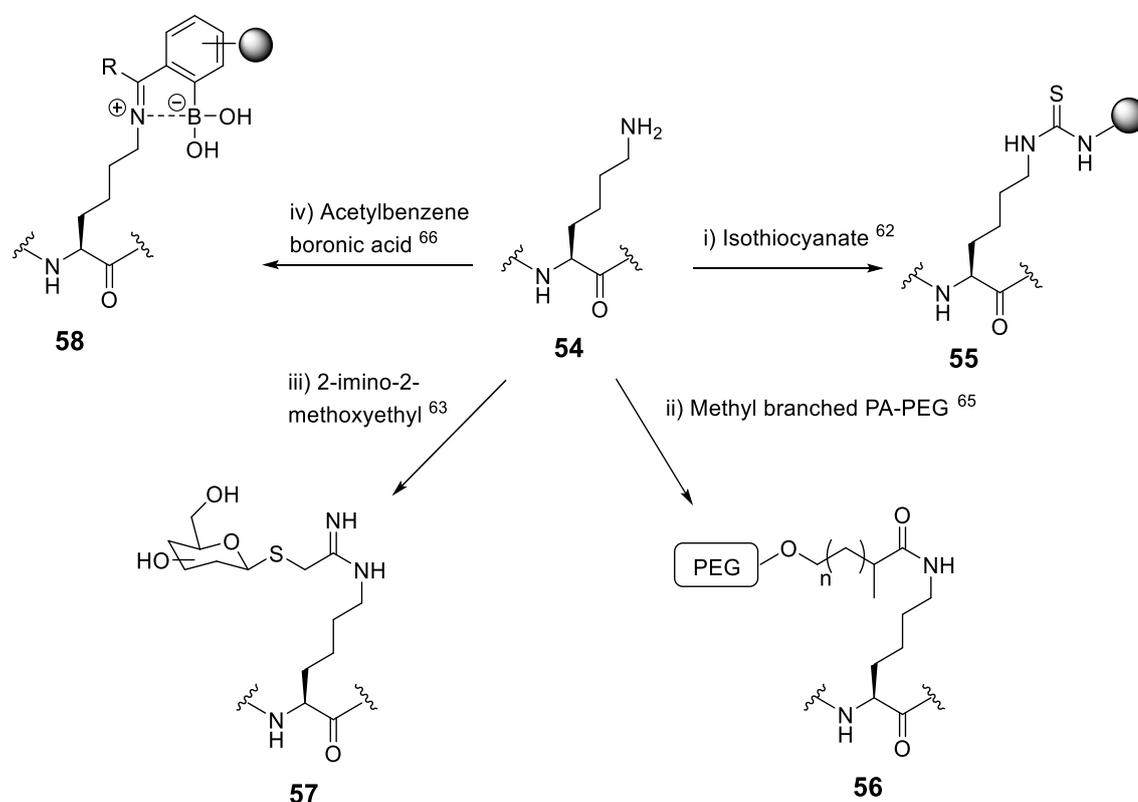
Scheme 1.13: Some examples of approaches utilised to carry out bioconjugation at Lys residues.

More complex modification directly at Lys includes the formation of an amidine linkage using 2-imino-2-methoxyethyl reagents (**Scheme 1.13iii**) more specifically 2-imino-2-methoxy-1-thioglycosides. This strategy was adopted to synthesise α-L-rhamnopyranosidase analogues (**Scheme 1.14**) which were shown to selectively target liver tumour cells.[63]
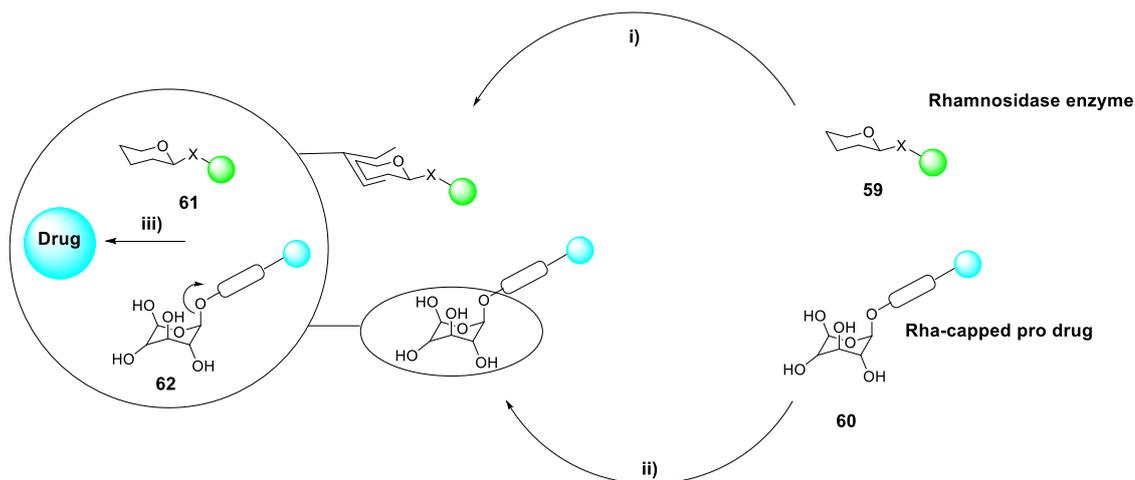
Scheme 1.14: Selective delivery of drug using α-L-rhamnopyranosidase.[63]

The above reaction **Scheme 1.14**, details biocatalytic delivery system which was used by Davis and co-workers[64] to selectively deliver L-rhamnose (Rha) capped doxorubicin (DOX) (**60**) to HepG2 disease model. The first step (**Scheme 1.14i**) in this reaction scheme is delivery of the Rha cleaving enzyme rhamnosidase (**59**) by receptor mediated endocytosis (RME). Conjugating a sugar moiety to this site allows for selective targeting of the diseased cell. **Scheme 1.14ii** describes the delivery of the Rha-capped drug (**60**) to the reaction site, whereupon its reaction with rhamnosidase, the drug cargo (DOX) is released (**Scheme 1.14iii**). These analogues were found to exhibit selective delivery of the drug cargo to the desired target, have enhanced stability compared to the non-conjugated drug cargo and exhibit an increased level of cargo delivery compared to the non-conjugated cargo.[64]

There have also been studies focusing on the PEGylation of peptides selectively at the free amine of Lys (**Scheme 1.13ii**). Initial studies by Harris and co-workers[65] centred on PEGylatation using PEG dichlorotriazine (**63**) (**Scheme 1.15i**) which was able to react with multiple side chains, but unfortunately led to cross-linked products. PEG-triflorate (**65**) (**Scheme 1.15ii**) was found to be a more selective PEGylating reagent towards Lys and Cys and was used to form linkages to a variety of proteins.
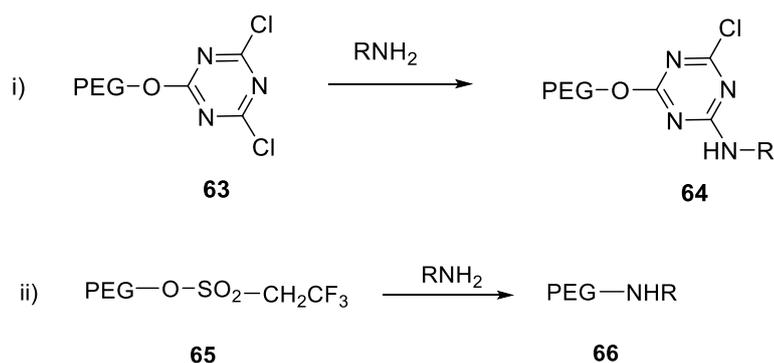
Scheme 1.15: Reaction of R-NH$_2$ with "First generation" PEGylating reagents.

These examples (**63** and **65**) are referred to as "first generation PEGylating reagents" and although the methods adopted are simple and efficient, PEGylation of proteins is inefficient due to contamination caused by diols in high molecular weight PEGs and side reactions. This has led to the synthesis of second-generation PEG reagents such as α-methyl branched PA-PEG analogue (**67**), as shown in **Scheme 1.16**.[65] Analogues such as this and other second generation PEGylatin reagents were shown to have increased stability to hydrolysis.[65]
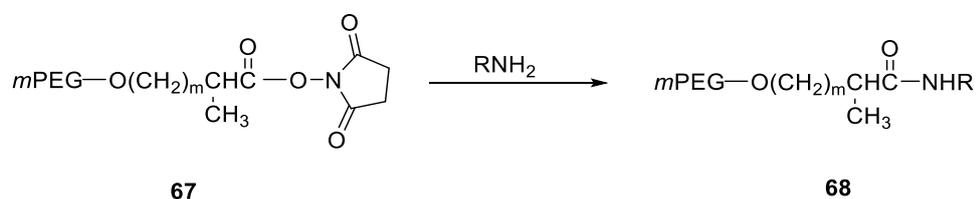
Scheme 1.16: Reaction of R-NH$_2$ with "Second-generation" PEGylating agents.

Finally, modification of Lys *via* the formation of iminoboronates (**71**) (**Scheme 1.13iv**) has been investigated as a method for synthesising cancer imaging conjugates by Gois and co-workers (**Scheme 1.17**).[66]
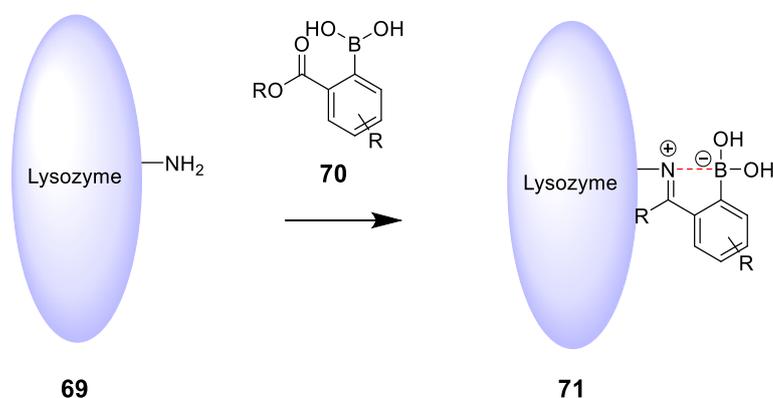
Scheme 1.17: Protein modification at Lys utilising iminoboronates.[66]

This modification proceeds *via* a B-N linkage which showed no detectable degradation after 5 h, yet was reversible upon addition of glutathione, fructose and dopamine. By selectively conjugating fluorescent analogues of acetylbenzeneboronic acid to the free amine of Lys in lysozyme, Gois and co-workers were able to use these analogues (**71**) to target folic acid receptors of cancer cells which allowed for selective internalisation of the described analogues.[66]

Cys and Lys modifications are, by far, the most common type of peptide side chain modification due to their nucleophilicity and their high abundance in nature. In the following sections, modification strategies developed towards other naturally occurring amino acids will be discussed (**Section 1.3.3**) along with approaches that utilise non-native amino acids (**Section 1.3.4**).

1.3.3   Bioconjugation at Serine (Ser)/ Threonine (Thr) and Tyrosine (Tyr)

The hydroxyl group, which as a functionality, is present in a large number of proteins, natural products and drugs is an interesting and challenging target for chemoselective modification and bioconjugation. Hydroxyl group bioconjugation is not an area that has been widely researched and in part this is due to the hydroxyl group displaying a reduced nucleophilicity compared to thiol and amine containing residues. This reduced reactivity makes it difficult to carry out chemoselective reactions of hydroxyl containing amino acids (e.g. Ser or Thr) in the presence of Cys and Lys.[67]

The concept of Ser tagging is a very interesting one due to the fact that approximately 1% of the human genome contains a serine residue. A family of such enzymes are Ser hydrolases which are characterised by the presence of a serine residue within the active site which is involved in the hydrolysis of substrates. Utilising tagging strategies to investigate the function and mechanism of these enzymes could be useful in developing possible Ser hydrolase inhibitors. The ability to target these residues specifically has promise due to the fact Ser hydrolases are involved in the causation of a number of diseases such as diabetes, cancer and Alzheimer's disease.[68]

Levin and co-workers[69] were able to successfully incorporate PEG-salicylaldehyde (**73**) to the *N*-terminal Ser of a variant of parathyroid hormone (PTH). PEG-salicylaldehyde was reacted with the protein fragment in the presence of pyridine-acetyl acid to form a mono-PEGylated species conjugated *via* amide bond (**74**) (**Scheme 1.18**).
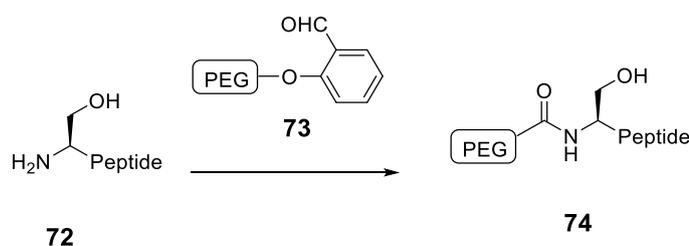
Scheme 1.18: Reaction of Ser (**72**) with PEG-salicylaldehyde (**73**) at the *N*-terminal amide bond.[68]

The formation of the mono-PEG compound **74** occurs *via* an imine **75** formation followed by O-N shift, producing compound **77**. Compound **77** is only formed in 1-amino-2-hydroxyl containing compounds such as Ser and Thr, meaning this PEGylation is selective for Ser and Thr. TFA is used to cleave the acetal moiety **78**, resulting in a mono PEGylated Ser compound **74** (**Scheme 1.19**).
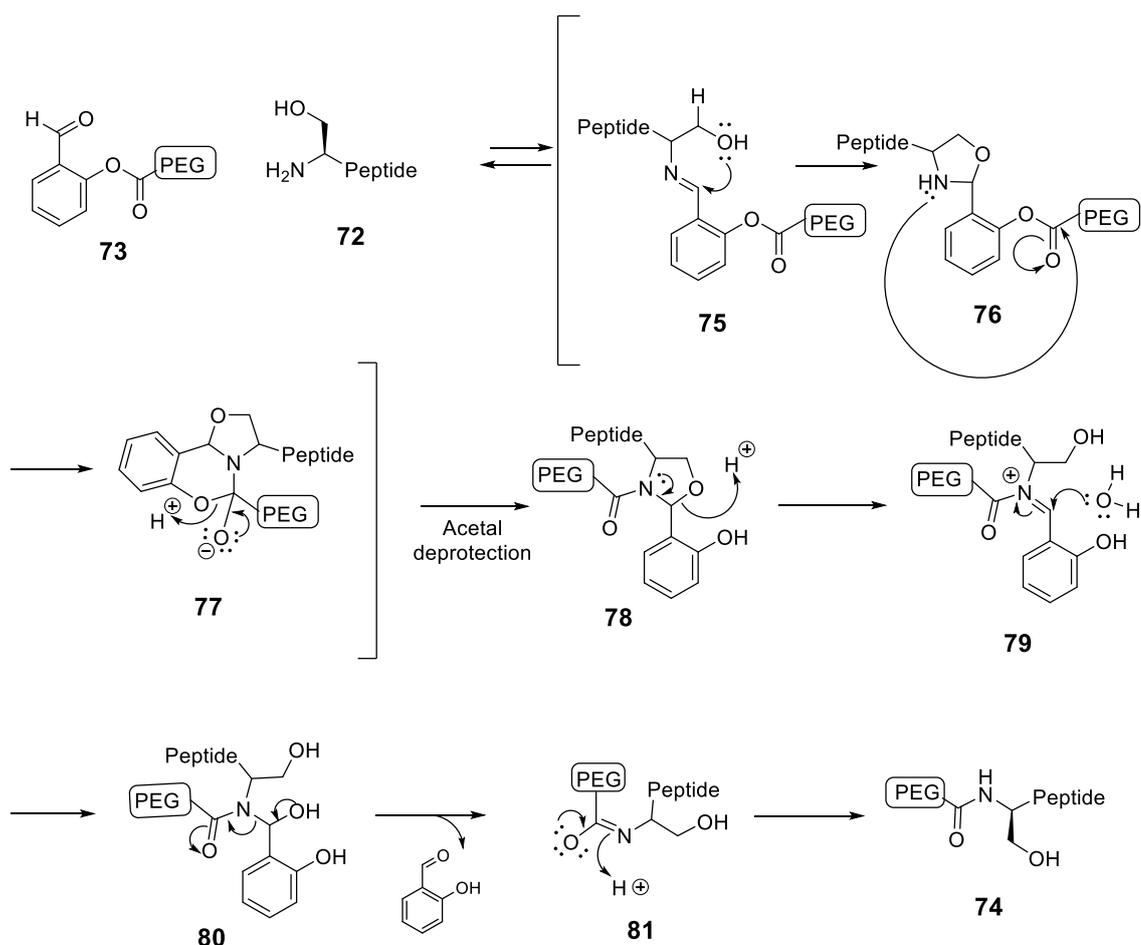
Scheme 1.19: Mechanism of PEGylation of *N*-terminal Ser (**72**).[69]

This method allowed for conjugation at the *N*-terminal Ser selectively and did not occur at other Lys residues within the sequence. PTH is a polypeptide which has shown promise in treating osteoporosis due to its ability to increase bone density, but due to is poor half-life has limited use as a therapeutic agent. Therefore, the reaction procedure mentioned above has promise in exploiting PTH as a viable therapeutic by improving stability.[69]

Bioconjugation at Ser and Thr residues has proven to be a challenging task due to the reasons previously described, but bioconjugation at Tyr has met with more success. For example, Barbas and co-workers[70] synthesised a 1,2,4-triazoldine-3,5-dione analogue (**82**) *via* an ene-type reaction at Tyr (**Scheme 1.20**).
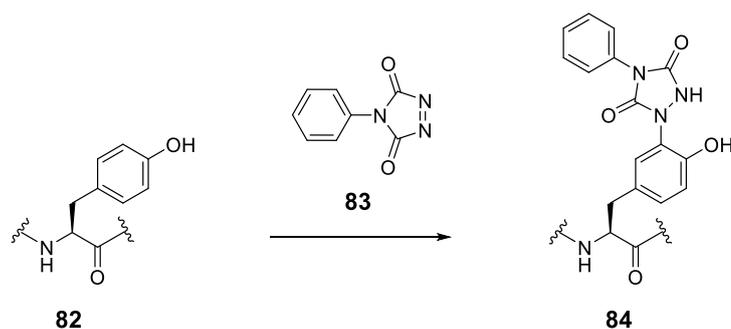
Scheme 1.20: Reaction of 1, 2, 4-triazoldine-3,5-dione (**83**) and Tyr (**82**).[70]

Due to the nature of the hydroxyl groups in Ser, Thr and Tyr residues studies have concentrated on modification of hydroxyl to more reactive species including conversion of the hydroxyl to a halide leaving group, oxidation to the corresponding aldehyde/ketone or conversion to a more reactive chemical species e.g. azide or thiol. Davis and co-workers[71] were able to utilise Lawesson's reagent (**86**) (LR) to convert hydroxyl of carbohydrates to thiols.
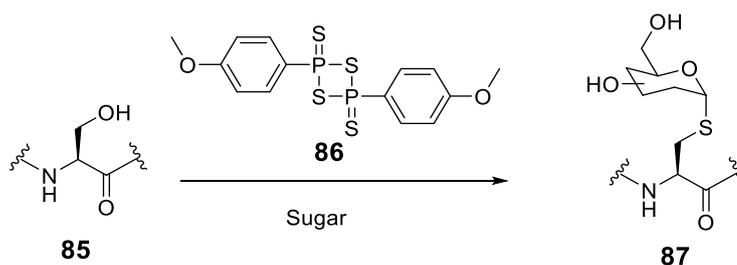
Scheme 1.21: Glycosylation using LR (**86**) *via* the hydroxyl of Ser (**85**).

LR has been shown to convert several chemically different hydroxyls to the corresponding thiol and has promise in glycosylating proteins in a one-pot reaction (**Scheme 1.21**).[71]

1.3.4   Bioconjugation at UAA using Click Chemistry

Click Chemistry commonly refers to the quick and efficient reaction between two small molecules to produce a larger structure. A variant of click chemistry, and perhaps the most widely used, is Cu(I) catalysed [3+2] cycloaddition (a variant of Huisgen cycloaddition reaction) and was first conceived by Sharpless, Meldal and colleagues in 2001.[72] In these reactions dipolarophiles react with 1,3-dipoles to form five-membered heterocycles. The reaction conditions implemented by Sharpless and co-workers suggested a quick, efficient and high yielding method of synthesising regio-specific heterocycles.
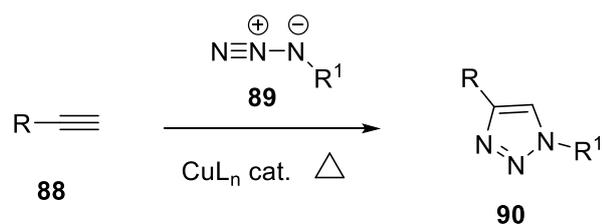
Scheme 1.22: Typical "click chemistry" reaction.[72]

The 1,2,3-triazole systems produced from a typical click chemistry reaction, as shown above (**Scheme 1.22**), has many interesting properties which make them useful tools in peptide synthesis. The triazole system is more stable than a typical amide bond present with a peptide sequence and has been implemented as a surrogate for amide bonds (**Figure 1.5i**) to help decrease enzymatic degradation.[73] Triazoles have also been investigated as linking agents in studies aimed at stabilising α-helices. Kawamoto and co-workers[74] synthesised triazole stapled peptides (**Figure 1.5ii**) and found that these stapled peptides had improved potency and stability compared to the non-stapled analogue.

i)   **91**   ii)   **92**
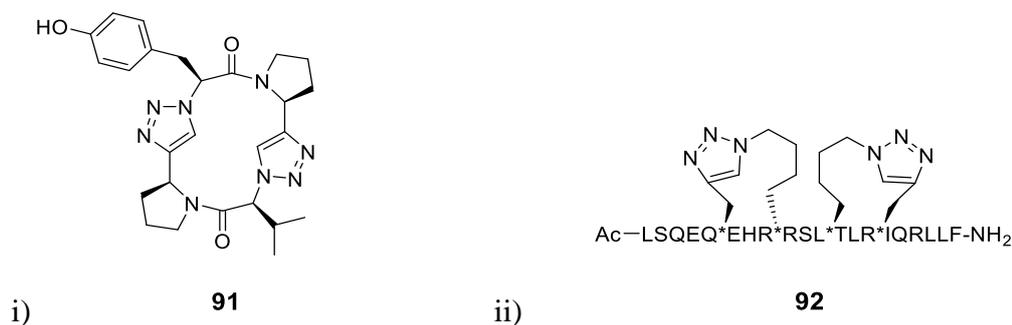
Figure 1.5: i) Triazole acting as a substitute for amide bond. ii) Triazole acting as a linker.[73]

In recent years, click chemistry conditions have been implemented as a method of conjugating peptides and a variety of groups including PEG groups[75] and imaging labels[76] as a method of investigating biological systems.

Liskamp and co-workers[75] investigated the use of click chemistry to bioconjugate biodegradable PEG groups to peptide-based hydrogels in an attempt to synthesise an effective drug delivery system. Hydrogels are hydrophilic polymer networks that have applications in drug delivery. By incorporating PEG groups into the polymeric network, it becomes biodegradable and allows for the release of a drug molecule when hydrolysed. The utilisation of click chemistry for the synthesis of the polymeric network

allowed the reaction to be carried out in mild conditions, without the need for protecting groups and with high yields of desired product.[75]

Molecules labelled with ^{18}F allow for positron emission spectroscopy (PET) which is a non-invasive and efficient method of visualising biological systems at a molecular level. Chen and co-workers[76] used click chemistry to incorporate ^{18}F labelled moieties into a peptide chain which can be traced *in vivo* (**Scheme 1.23**).
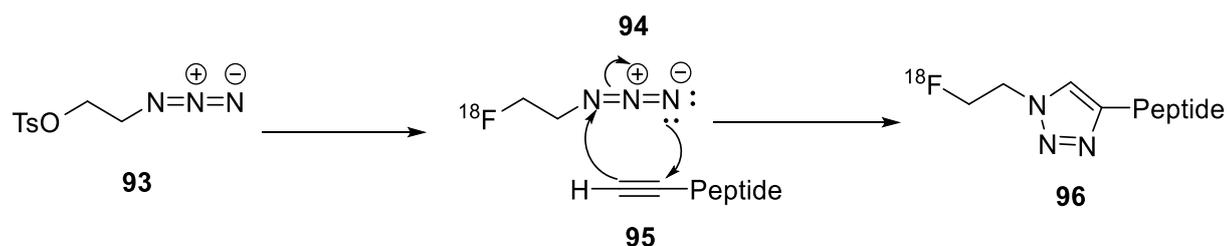
Scheme 1.23: Incorporation of labelled ^{18}F into peptide using click chemistry.

Several different radiolabels and peptides have been synthesised and it was found that PET imaging provided an effective method for visualising target peptides specifically and their mode of action. By synthesising a range of conjugated peptides and ^{18}F-labelled probes, several different targets can be investigated. These include integrin receptor[77] - overexpression of which has been shown to facilitate metastasis in cancer cells, somatostatin receptor[78] - which is known to facilitate insulin and glucagon release and neurotensin receptor- which has been implicated in many different cancers such as breast and pancreatic cancer.[79]

Another method of click chemistry is "Cu-free click chemistry" which utilises the 1,3-dipolar cycloaddition of azides and cyclooctynes and has been extensively researched by Bertozzi and co-workers.[80] This method does not use toxic Cu in its synthesis so has applications in *in vivo* studies. The reaction between the cyclooctyne ring and azides is rapid due to the release of ring strain which means there is no need for the use of a catalyst in this reaction. Antibody-drug-conjugates (ADC) synthesised using Cu-free click chemistry result in site specific conjugation to a variety of substrates including imaging labels and chemotherapeutic agents.[80]

Sato and co-workers[81] synthesised unnatural amino acid (UAA)-containing proteins (**93**) which were then used in the synthesis of ADC containing *p*-azidomethylphenylalanine (*p*AMF). *p*AMF altered IgG antibodies were conjugated to a cyclooctyne tags (**94**) which showed potency in *in vitro* studies (**Scheme 1.24**).

This reaction is particularly successful due to the highly reactive cyclooctyne used in the reaction. Upon investigation of cyclooctyne bond angles, Bertozzi and co-workers[82] determined the alkyne bond angle of cyclooctyne was ~160°, which is the same as the bond angle of the transition state of the cyclo-addition reaction. The release of ring strain upon reaction of cyclooctyne with azides results in a rapid reaction rate acceleration.
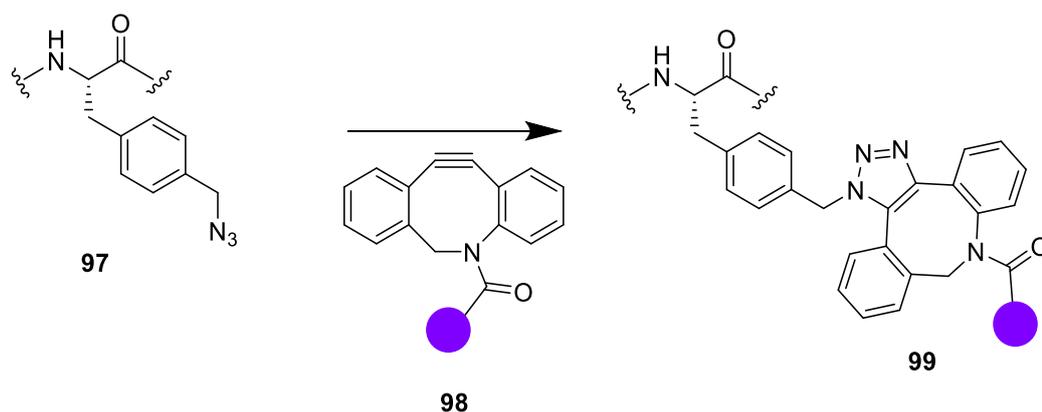
Scheme 1.24: Conjugation of UUA containing trastuzumab with cyclooctyne tag.[81]

The antibody used in this study was a Her2- binding IgG Trastuzumab and was altered to include a pAMF moiety into the protein backbone. Using Cu-free click chemistry, a DBCO-PEG-monomethyl auristatin (DBCO-PEG-MMAF) drug cargo was conjugated to the antibody which, when tested, exhibited high potency in cytotoxicity assays using a SKBR3 model system.[81]

As well as conjugation of drug cargo, Cu-free click chemistry has also been used to conjugate imaging agents to peptides. Bertozzi and co-workers[82] used Cu-free click chemistry to investigate zebrafish glycoproteins. Sialic acid has proven an important component in cellular processes such as cell adhesion and receptor signalling. Previous studies to investigate sialic acid have proven difficult due to the lack of chemical tools.

In this study, cell permeable (N-azidoacetylmannosamine) ManNAz was changed to a sialic acid derivative (SiaNAz) which was incorporated into glycans on the cell surface. Difluorinated cyclooctyne (DIFO) conjugated to an imaging probe reacted with the azido-sialic acid derivative *via* a click reaction resulting in a glycoprotein conjugated to an imaging tag. Bertozzi and co-workers used this strategy to visualise zebrafish embryos and found that they were able to successfully visualise zebrafish embryo cell surface sialoside biosynthesis and also visualise sialogylconjugate expression at various

levels of embryo development.[82] Due to its versatility, mild conditions and high yields, click chemistry has proven a useful tool in the bioconjugation of peptides to a vast range of different molecules.

1.3.5 Organometallic Mediated Peptide Bioconjugation

Although the field of organometallic compounds in biology is a well-researched field, the conjugation of transition metals (TM) to peptides for biomedical research is a relatively new field of research. Transition metals can be used in multiple scenarios to alter peptide function including being incorporated into a peptide backbone using unnatural amino acids[83], direct conjugation with the nucleophilic side chains of natural amino acids[84], conjugation *via* the *C*- or *N*-termini[85], and also act as catalysts to incorporate R groups into a peptide.[86]

There have been a number of studies into exploiting the reactivity of natural amino acids to conjugate a TM to the peptide backbone. This could be achieved by exploiting the electron accepting ability of TM to conjugate to π-electron rich residues such as Phe and Tyr. In 1998, Grotjahn and co-workers[87] synthesised a Ru-modified secretin analogue (**100**). Secretin is a peptide that controls water regulation throughout the body and Grotjahn and co-workers synthesised a secretin analogue containing multiple sites for possible TM conjugation. They were able to selectively modify the Phe side chain as opposed to the free *N*-terminus or *N*-terminal His with a reported conversion of 70-100%. (**Figure 1.6**).
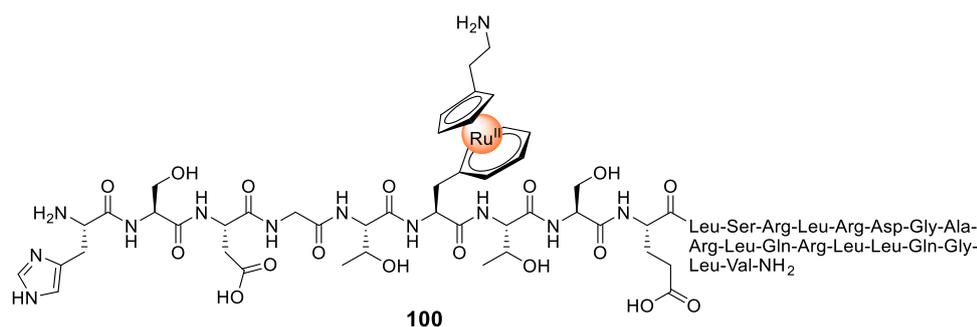
Figure 1.6: Ru-modified secretin analogue synthesised by Grotjahn and co-workers.[87]

Organometallic compounds can also be used to catalyse cross-coupling reactions of peptides and small molecules.[88] Davies and co-workers developed a Pd-pyrimidine catalyst which was used to conjugate R groups to a protein. First the Pd-pyrimidine coupling agent was tested on a series of small halogenated amino acids with various boronic acids using mild reaction conditions (**Table 1.5**).

Entry	ArX	Condition	Coupling Product	Yield (%)
1	101	37 °C, 4 h	102	95
2	103	37 °C, 4 h	104	98
3	105	37 °C, 4 h	106	0
4	107	37 °C, 4 h	108	94
5	109	37 °C, 4 h	110	95
6	111	37 °C, 6 h	112	0
7	113	37 °C, 6 h	114	92
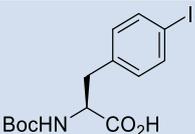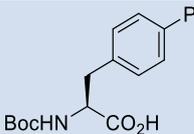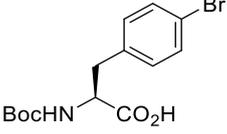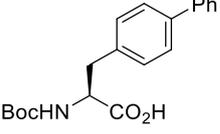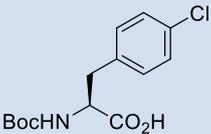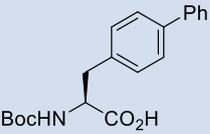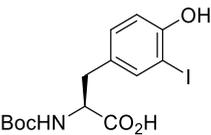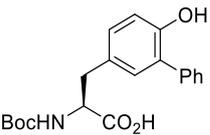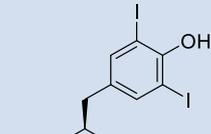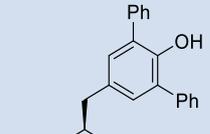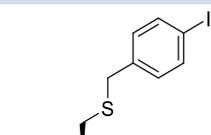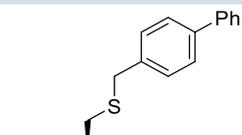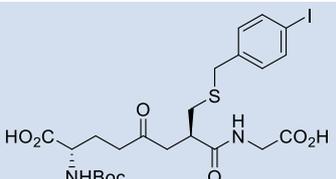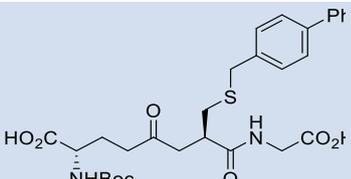
Table 1.5: Pd-pyrimidine mediated Suzuki-Miyaura cross coupling of halogenated aryl amino acids with Ph.[88]

The results from this exploratory work were promising with high yields for most of the test reactions, with these reactions carried out without the use of organic solvents. It was suggested that **Entry 3** low yield was due to the inactivity of the Cl leaving group. **Entry 6**- which contains a *C*-terminal *p*-iodobenzyl cysteine (Pic)-had a yield of 0 %, believed to be due to the active site inhibition of the Pd-catalyst, yet **Entry 7**- which contains an internal *p*-iodobenzyl cysteine- was successful with a yield of 92 %.[88]

Davies and co-workers proceeded to cross couple various R groups to a Pic mutated SBL protein (**115**) using the Pd-pyrimidine catalyst (**116**) and boronic acid, under mild conditions (**Scheme 1.25**).
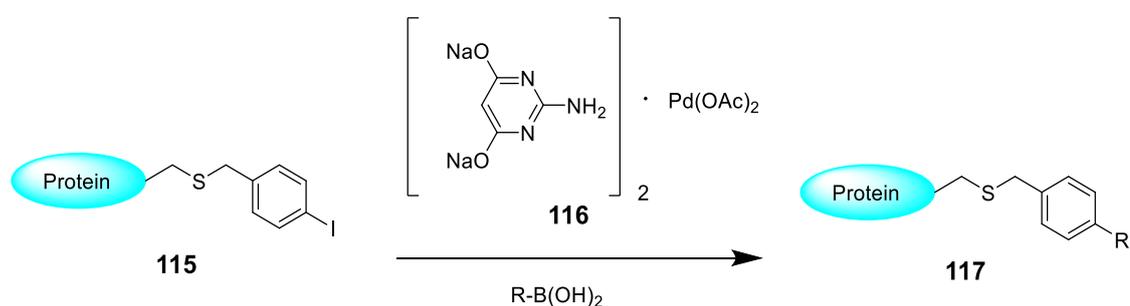
Scheme 1.25: Pd-pyrimidine mediated Suzuki-Miyaura cross coupling using a Pic mutated protein to couple various R groups.

A variety of aryl and vinylboronic acids were used in the cross-coupling reactions with conversions of >95%. Using this strategy, Davies and co-workers were able to synthesise a glycoprotein (**117**) using aqueous and mild conditions in high yield.[88]

With the significant advancements that have been made in the field of TM-peptide conjugation, it is expected that future research will concentrate on applying the techniques described to improve existing peptide drugs and to further explore the scope of this technique of peptide modification.[89]

1.4 Project Aims

Historically the use of peptides as therapeutics has been hampered due to poor bioavailability and poor stability within biological systems. As discussed in **Section 1.3.2-1.3.5**, there have been numerous studies into improving the chemical properties of peptides which include conjugation of small molecules *via* reactive side chain residues. A number of these studies have been successful in improving properties such as stability and solubility compared to the native peptides.

The aim of this project is to investigate the role of fluorinated building blocks in peptide modification. As mentioned in **Section 1.3.1** the role of HFB in peptide stapling and modification has been investigated by Pentalute and co-workers[58] and these studies yielded interesting results in improving the helicity of peptides. They found that HFB was reactive towards Cys residues, but not reactive towards other nucleophilic side chains on residues such as Tyr and Lys and also that the HFB could not be further modified once conjugated to the peptide.

Previous work in the Cobb group has investigated pentafluoropyridine (PFP) as a tool for peptide modification with Hudson and co-workers[90] synthesising a PFP containing tripeptide (**120**). PFP initially undergoes nucleophilic aromatic substitution at the *para* position to the nitrogen, but it can then undergo further nucleophilic aromatic substitution reactions at the *ortho* positions leading to a multi-functionalised scaffold (**Scheme 1.26**).
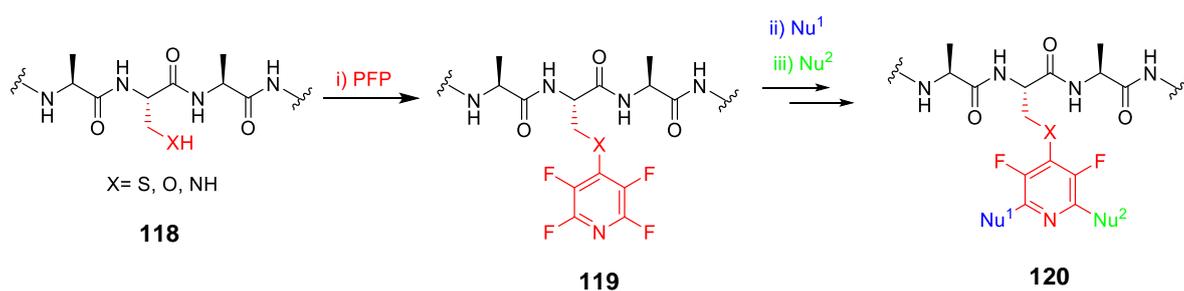
Scheme 1.26: i) Tagging of peptide **118** with PFP ii) Nucleophilic aromatic substitution at *ortho* position of compound **119**. iii) Further nuclopehilic aromatic substitution at *ortho* position of compound **119**.

The substitution pattern of PFP with nucleophiles is particularly interesting when applied to the modification of peptides as it can allow for conjugation of multiple PEG chains or thio-sugars into a peptide which may influence stability and lipophility.

Further work carried out within the Cobb group found that PFP was reactive towards Cys, Tyr and Lys side chains, whereas there was no such reaction between these side chains and HFB making PFP a more versatile tool for peptide modification (**Scheme 1.27**).[91]
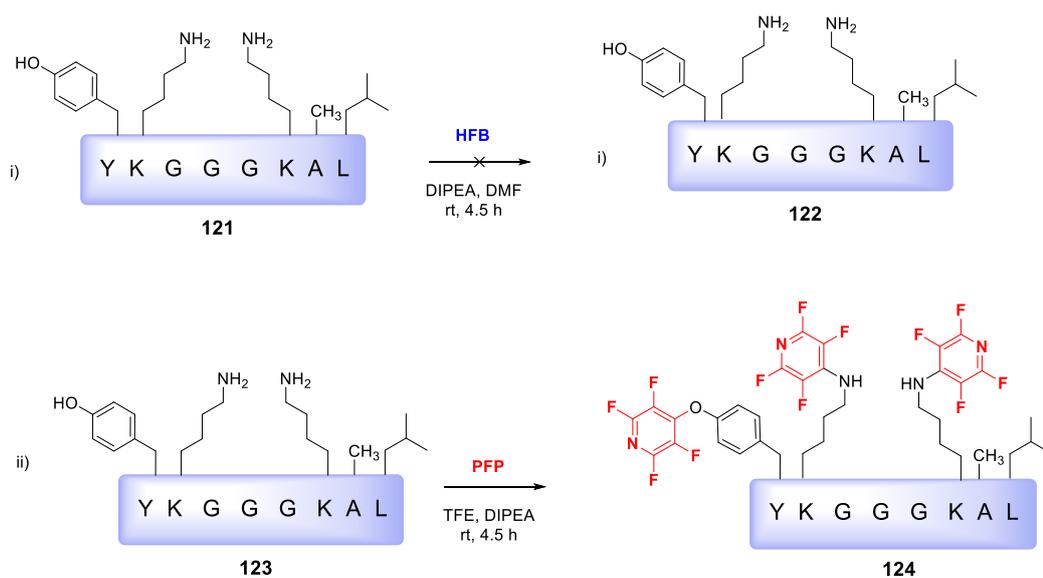
Scheme 1.27: i) Unselective tagging of Tyr and Cys residues using DIPEA, DMF, rt, 4 h ii) Selective tagging of Cys residues using DIPEA, TFE, DMF, rt, 4 h.

This work provided encouragement to pursue this area of research and to investigate the effect of PFP modification on peptide stability. The aim of this study was to investigate the role of PFP itself in altering peptide properties as well as using thio-sugars and thio-PEGs to further functionalise a peptide *via* aromatic substitutuion of the PFP.

As part of the presented work the investigation of tagging of various peptides with PFP starting with small dipeptides will be discussed (**Chapter 2: Synthesis, Tagging and Investigation of Dipeptides using PFP**), before investigation of the effect of PFP on stability of Oxytocin and Vasopressin (**Chapter 3: Synthesis, Tagging and Stability Studies of PFP-modified Oxytocin and Vasopressin**). The final aspect of this work will concentrate on the synthesis and modification of GLP-1 with PFP (**Chapter 4**: **Introduction to GLP-1: Improving Pharmokinetics *via* Bioconjugation**).

References

[1] D. W Miles, *Breast Cancer Res.*, 2001, **3,** 380-384.

[2] D. J. Craik, D. P, Fairlie, S. Liras and D. Price, *Chem. Biol. Drug Des.,* 2013, **81**,136–147.

[3] B. Leader, Q. J. Baca, D. E. Golan*, Nat. Rev. Drug Discov.*, 2008, **7**, 21-39.

[4] S. Sachdevi, *Int. J. Pept. Res. Ther.*, 2016, **5**, 1-12.

[5] A. A. Kaspar, J. M. Reichert., *Drug Discov. Today*, 2013, **18**, 807-817.

[6] O. Keskin, N. Tuncbag, A. Gursoy, *Chem. Rev*., 2016, **116**, 4884-4909.

[7] W. Iyer, *Curr. Med. Chem*., 2016, **23**, 3025-3043.

[8] M. R. Arkin, Y. Tang, J. A. Wells, *Chem. Biol*., 2014, **21**, 1102-1114.

[9] P. Wojcik, L. Berlicki, *Bioorg. Med. Chem. Lett*., 2016, **26**, 707-713.

[10] J. L. Fox., *Nat Biotechnol*., 2013, **5**, 379-382.

[11] A. A. Bahar, *Pharmaceuticals*., 2013, **12**, 1543-1575.

[12] W. S. Horne, M. K. Yadav, C. D. Stout, M. R. Ghadin, *J. Am. Chem. Soc*., 2004, **126**, 15366-15367.

[13] C. F. Deacon, L. B. Knudsen, K. Madsen, F. C. Wiberg, O. Jacobsen, J. J. Holst, *Diabetologia*, 1998, **41**, 271-278.

[14] R. E. Hancock, H. G. Sahl, *Nat. Biotechnol*., 2006, **24**, 1551-1557.

[15] G. Carmona, A Rodriguez, D Juarez, G. Corzo, E. Vilegas, *Prot. J.*, 2013, **32**, 456-466.

[16] T. J. Sanborn, C. W. Wu, R. N. Zuckermann, A. E Barron, *Biopolymers*, 2002, **63**, 12-20.

[17] M. Kristensen, D. Birch, H. M. Nielson, *Int. J. Mol. Sci*., 2016, **17**, 185-202.

[18] E. Vives, J. Schmidt, A. Pelegrin, *Biochim. Biophys. Acta*, 2008, **1786**, 126-138.

[19] M. Lindgren, M. Hallbrink, A. Prochiantz, U. Langel, *Trends Pharmacol. Sci.,* 2000, **3**, 99-103.

[20] S. Fukati, *Int. J. Pharma.*, 2002, **245**, 1-7.

[21] N. Schmidt, A. Mishra, G. H. Lai, G. C. L. Wong, *FEBS Lett.*, 2010, **584**, 1806-1813.

[22] J. B. Rothbard, E. Kreider, C. L. VanDeusen, L. Wright, B. L. Wylie, P. A. Wender, *J. Med. Chem.*, 2002, **45,** 3612-3618.

[23] S. R. Schwarze, A. Ho, A. Vocero-Akbani, S.F. Dowdy, *Science,* 1999, **285**, 1569-1572.

[24] X. Chen, R. Park, A. H. Shahanian, M. Tohme, V. Khankaldyyah, M. H. Bozorgzadeh, J. R. Bading, R. Moats, W. E. Laug, P. S. Coni, *Nuc. Med. and Bio.*, 2004, **31**, 179-189.

[25] H. Y. Shui, H. C. Chong, Y. C. Leung, T. Zou, C. M. Che, *Chem. Commun.*, 2014, **50**, 4375-4378.

[26] S. Santra, H. Yang, D. Dutta, J. T. Stanley, P. H. Holloway, W. Tan, B. M. Moudgil, R. A. Mericle, *Chem. Commun.*, 2004, 2810-2811.

[27] Y. Lui, M. K. Shipton, J. Ryan, E. D. Kaufman, S. Franzen, D. L. Feldheim, *Anal. Chem.*, 2007, **79**, 2221-2229.

[28] M. Zhou, I. Ghosh, *Biopolymers*, 2007, **88**, 325-339.

[29] J. F. Woodley, *Crit. Rev. Ther. Drug Carrier Syst.*, 1994, **11**, 61-95.

[30] E. Birion, J. Chatterjee, O. Ovadia, D. Langenegger, J. Bruggen, D. Hoyer, H. A. Schmidt, R. Jelinek, C. Gilon, A Hoffman, H. Kessler, *Angew. Chem.*, 2008, **14**, 2595-2599.

[31] A. F. Chu-Kung, K. N. Bozzelli, N. A. Lockwood, J. R. Haseman, K. H. Mayo, M. V. Tirrel, *Bioconjugate Chem.*, 2004, **15**, 530-535.

[32] M. J. Roberts, M. D. Bentley, J. M. Harris, *Adv. Drug. Disc. Rev.*, 2002, **54**, 459-476.

[33] J. M. Harris, N. E. Martin. M. Modi, *Clin. Pharmokinet.*, 2001, **40**, 539-551.

[34] A. T. Bockus, C. M. McEwan, R. S. Lokey, *Curr. Top. Med. Chem.*, 2013, **7**, 821-836.

[35] C. M. Moruno, F. Rechenmacher, H. Kessler, *Anti-Cancer Agents in Med. Chem.*, 2010, **10**, 753-768.

[36] A. I Fernandez-Llamazares, J. Adan, F. Mitjans, J. Spengler, F. Albericio, *Bioconjugate Chem.*, 2013, **25**, 11-17.

[37] M. Conti, V. Tazzari, C. Baccini, G. Pertici, L. P. Serino, U. De Giorgi, *In Vivo*, 2006, **20,** 697-702.

[38] A. G. Tkachenko, H. Xie, D. Coleman, W. Glomm, J. Ryan, M. F. Anderson, S. Franzen, D. L. Feldheim, *J. Am. Chem. Soc.*, 2003, **125**, 4700-4701.

[39] L. Pan, Q. He, J. Lui, Y. Chen, M. Ma, L. Zhang, J. Shi, *J. Am. Chem. Soc.*, 2012, **134**, 5722-5725.

[40] J. Zong, S. L. Cobb, N. R. Cameron, *Biomater. Sci.*, 2017, **5**, 872-886.

[41] P. Ghosh, G. Han, M. De, C. K. Kim, V. M. Rotello, *Adv. Drug Deliv. Rev.*, 2008, **60**, 1307-1315.

[42] D. Bartczak, S. Nitti, T. M. Millar, A. G. Kanaras, *Nanoscale*, 2012, **4**, 4470-4472.

[43] R. Shukla, E. Hill, X. Shi, J. Kim, M. C. Muniz, K. Sun, J. R. Baker, *Soft Matter*, 2008, **4**, 2160-2163.

[44] P. K. A. Weber, J. E. Bader, G. Folkers, A. G. Beck-Sickinger, *Bioorg. And Med. Chem. Lett.*, 1998, **8**, 597-600.

[45] Y. Zheng, S. Ji, A. Czerwinski, F. Valenzuela, M. Pennington, S. Lui, *Bioconjugate Chem.*, 2014, **25**, 1925-1941.

[46] L. M. Jungbauer, C. Yu, K. J. Laxton, M. J. LaDu, *J. Mol. Recognit.*, 2009, **22**, 403-413.

[47] H. Y. Shui, H. C. Chong, Y. C. Leung, T. Zou. C. M. Che, *Chem. Commun.*, 2014, **50**, 4375-4378.

[48] P. Giron, L. Dayon, J. C. Sanchez, *Mass Spec. Rev.*, 2011, **30**, 366-395.

[49] S. Faggiano, A. Pastore, *Cells,* 2014, **2**, 639-656.

[50] M. E. B. Smith, F. F. Schumacher, C. P. Ryan, L. M. Tedali, D. Papaioannuo, G. Waksman, S. Caddick, J. R. Baker, *J. Am. Chem. Soc.*, 2010 **132,** 1960-1965.

[51] G. T. Hermanson, *Bioconjugation Techniques* (Third Edition), New York, 2013, 229-258.

[52] N. Griebenow, A. M. Dilmac, S. Greven, S. Brase, *Bioconjugate Chem.*, 2016, **27**, 911-917.

[53] V. Chudasama, M. E. B. Smith, F. F. Schumacher, D. Papaioannuo, G. Waksman, S. Caddick, J. R. Baker, *Chem. Commun.*, 2011, **31**, 8781-8783.

[54] E. Robinson, J. P. M. Nunes, V. Vassileva, A. Maruani, J. C. F. Nogueira, M. E. B. Smith, R. B. Pedley, S. Caddick, J. R. Baker, V, Chudasama, *RSC Adv.*, 2017, **7**, 9073-9077.

[55] J. M. Chalker, G. J. L. Bernardes, Y. A. Lin, B. G. Davis, *Chem. Asian. J.*, 2009, **4**, 630-640.

[56] J. M. Chalker, S. B. Gunnoo, O. Boutureira, S. C. Gerstberger, M. Fernandez-Gonzalez, G. J. L. Bernardes, L. Griffin, H. Hailu, C. J. Scholfield, B.G. Davis, *Chem. Sci.,* 2011, **2**, 1666-1676.

[57] J. M Chalker, L. Lercher, N. R. Rose, C. J Schofield, B. G. Davis, *Angew. Chem. Int. Ed.*, 2012, **124**, 1871-1875.

[58] A. M. Spokoyny, Y. Zou, J. J. Ling, H. Yu, Y. Lin and B. L. Pentelute, *J. Am. Chem. Soc.*, 2013, **135**, 5946-5949.

[59] S. J. Miller, H. E. Blackwell and R. H. Grubbs, *J. Am. Chem. Soc.*, 1996, **118**, 9606-9614.

[60] J. L. Stymiest, B. F. Mitchell, S. Wong and J. C. Vederas, *Org. Lett.*, 2003, **5**, 47-49.

[61] C. Zhang, M. Welborn, T. Zhu, N. J. Yang, M. S. Santos, T. Van Voorhis and B. L. Pentelute, *Nature Chemistry*, 2015, **8**, 120-128.

[62] F. M. Veronese and G. Pasut, *Drug Discovery Today*, 2005, **10**, 1451-1458.

[63] P. R. Banks, D. M. Paquette, *Bioconjugate Chem*., 1995, **6,** 447-458.

[64] M. A. Robinson, S.T. Charlton, P. Garnier, X. Wang, S. S. Davis, A. C. Perkins, M. Frier, R. Duncan, T. J. Savage, D. A. Wyatt, S. A. Watson, B. G. Davies, *PNAS,* 2004, **101**, 14527-14532.

[65] M. J. Roberts, M. D. Bentley, J. M. Harris, *Adv. Drug Discov. Rev*., 2002, **54**, 459-476.

[66] P. M. S. D. Cal, R. F. M. Frade, V. Chudasama, C. Cordeiro, S. Caddick, P. M. P. Gois, *Chem. Commun*., 2014, **50**, 5261-5263.

[67] D. J. Trader, E. E. Carlson, *Biol. Biosys.,* 2012, **8**, 2484-2493.

[68] J. Z. Long, B. F. Cravatt, *Chem. Rev*., 2011, **111**, 6022-6063.

[69] P. M. Levin, T. W. Craven, R. Bonneau, K. Kirshenbaum, *Chem. Commun.,* 2014, **50**, 6909-6912.

[70] H. Ban, J. Gavrilyuk, C. F. Barbas, *J. Am. Chem. Soc.,* 2010, **132**, 1523-1525.

[71] G. J. L. Bernardes, D. P. Gamblin, B. G. Davis, *Angew. Chem. Int. Ed.*, 2006, **45**, 4007-4011.

[72] H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem. Int. Ed.,* 2001, **40**, 2004-2021.

[73] K. New, M. W. Brechniel, *Cancer Biother. Radiopharm*., 2009, **24**, 289-302.

[74] S. A Kawamoto, A Coleska, X. Ran, H. Yi, C. Y, Yamg, S. Wang, *J. Med. Chem*., 2012, **55**, 1137-1146.

[75] M. van Dijk, C. F. van Nostrum, W. E Hennink, D. T. S. Rijkers, R. M. J. Liskamp, *Biomacromol.,* 2010, **11**, 1608-1614.

[76] K. Chen, P. S. Conti, *Adv. Drug Deliv. Rev*., 2010, **62**, 1005-1022.

[77] R. Haubner, W. A Weber, A. J. Beer, E. Vabuliene, D. reim, M. Sarbia, K. F. Becker, M. Goebel, R. Hein, H. J. Wester, H. Kessler, M. Schwaiger, *PloS Med*., 2005, **2**, 244-252.

[78] J. Jeon, B. Shen. L. Xiong, Z. Mioa, K. H. Lee, J. Rao, F. T. Chin, *Bioconjugate Chem.,* 2012, **23**, 1902-1908.

[79] Z. B. Li, Z. Wu, K. Chen, E. K. Ryu, X. Chen, *J. Nuc. Med.*, 2008, **49**, 453-461.

[80] P. Agarwal, C. R. Bertozzi, *Bioconjugate Chem.*, 2015, **26**, 176-192.

[81] E. S. Zimmerman, T. H. Heibeck, A. Gill, X. Li, C. J. Murray, M. R. Madlansacav, C. Tran, N. T. Uter, G. Yin, P. J. Rivers, A. Y. Yam, W. D. Wang, A. R Steiner, S. U Bajad, K. Penta, W. Yang, T. J. Hallam, C. D. Thanos, A. K. Sato., *Bioconjugate Chem.,* 2014, **25**, 351-361.

[82] K. W. Dehnert, J. M. Baskin, S. T. Laughlin, B. J. Beahm, N. N. Naidu, S. L. Amacher, C. R. Bertozzi, *Chem. Biochem.*, 2012, **13**, 353-357.

[83] J. Lemke, N. Metzler-Nolte, *J. Organomet. Chem.*, 2011, **696**, 1018-1022.

[84] M. Salmain, K. L. Malisza, S. Top, G. Jaouen, M. C. Senechal-Tocquer, D. Senechal, B. Caro, *Bioconjugate Chem.*, 1994, **5**, 656-659.

[85] K. Splith, I. Neundorf, W. Hu, H. W. P. N'Dongo, V. Vasyyeva, K. Merz, U. Schatzschneider, *Dalton Trans.*, 2010, **39**, 2536-2545.

[86] D. F. Shriver, P. W. Aitken, *Inorg. Chem.*, 1991, **3**, 957-961.

[87] D. B. Grotjahn, *Chem. Rev.*, 1999, **190**, 1125-1141.

[88] J. M. Chalker, C. S. C. Wood, B. G. Davies, *J. Am. Chem. Soc.*, 2009, **131**, 16346-16347.

[89] B. Albada, N. Meltzer-Nolte, *Chem. Rev.*, 2016, **116**, 11797-11839.

[90] A. S. Hudson, A. Hoose, C. R. Coxon, G. Sandford and S. L. Cobb, *Tetrahedron Lett.*, 2013, **54**, 4865-4867.

[91] D. Gimenez, C. A. Mooney, A. Dose, G. Sandford, C. R. Coxon, S. L. Cobb, *Org. Biomol. Chem.*, 2017, **15**, 4086-4095.

Chapter 2: Synthesis, Tagging and Modification of Dipeptides using Perfluoroaromatics

2.1 Fluorine in Medicinal Chemistry

The use of fluorine and fluorinated building blocks as tools for the modification of pharmaceuticals has been extensively researched. This is due to the fact they offer highly versatile routes to modulate the physical, electronic and chemical properties of drug molecules.[1]

One of the earliest examples of a fluorinated pharmaceutical is 5-fluorouracil (**126**) the fluorinated analogue of uracil (**125**) (**Figure 2.1**). The fluorine atom is situated on the 5-position of the pyrimidine ring, and this simple switch from H to F has little steric impact, but it significantly alters the biological properties of the molecule. Uracil (**125**) is an essential building block for DNA transcription, but the fluorinated analogue (**126**) has proven to be an effective anti-cancer agent which acts as a pyrimidine antagonist. Fluorouracil (**126**) can act as a substrate for thymidylate synthase and is converted to fluorouridine. This is then converted to fluorouridine triphosphate and incorporated into RNA as a false nucleotide, disrupting translation.
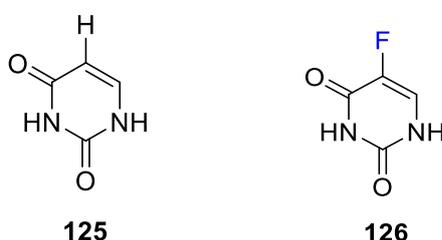
Figure 2.1: Structure of uracil (**125**) and 5-fluorouracil (**126**).

Substitution of a hydrogen atom by fluorine can also be used to alter the chemical properties of a drug, increasing its stability towards oxidation and unwanted metabolism by P450 enzymes. Ezetimibe (**128**) is a successful cholesterol drug that provides a good example of how fluorine can be used to prevent unwanted metabolic breakdown. The pro-drug SCH 48461 (**127**) was shown to have four possible sites of metabolism. By incorporating fluorine atoms at these specific sites, the resulting compound was found to be more stable to metabolism and this eventually led to the synthesis of Ezetimibe (**128**) which had 50 times higher potency than SCH 48461 (**127**).[3]
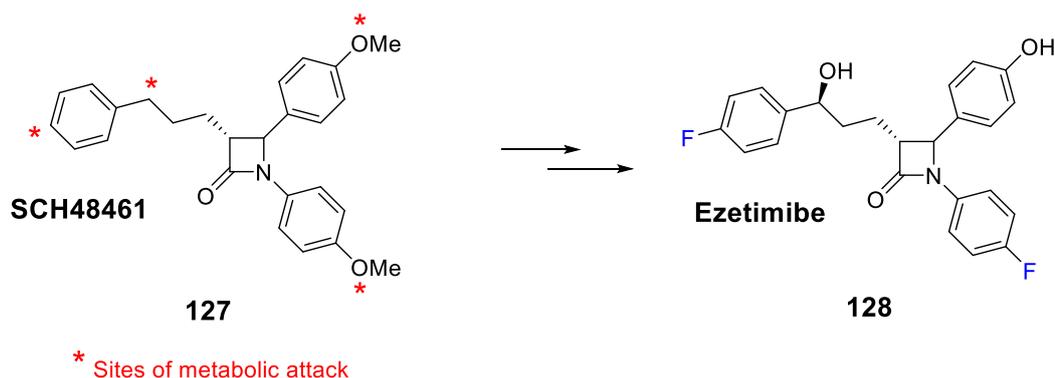
Scheme 2.1: Structure of SCH 48461(127) including sites of possible metabolic breakdown and Ezetimibe (128).[3]

2.1.1 Nucleophilic Aromatic Substitution of Pentafluoropyridine and Hexafluorobenzene

The electron withdrawing character of fluorine in hexafluorobenzene (HFB) (47) and pentafluoropyridine (PFP) (129) lowers the energy of their π^* orbitals. This results in a decrease in the electron density on the aromatic ring in both PFP (129) and HFB (47), making them susceptible to nucleophilic aromatic substitution (S_NAr).[5]
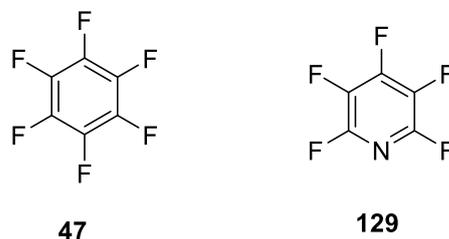
Figure 2.2: Structures of hexafluorobenzene (HFB) (47) and pentafluoropyridine (PFP) (129).

Moreover, the electron withdrawing power of the fluorine atoms in PFP (129) and HFB (47) stabilise the anionic intermediate formed during S_NAr further increasing the reactivity between PFP/HFB and nucleophiles.[6] The reactivity of HFB (47) towards nucleophilic peptides has been discussed in **Chapter 1** (**Section 1.3.1**) and here, the reactivity of PFP (129) with respect to S_NAr, and in particular sulphur-containing nucleophiles, will be briefly discussed.

Studies into the rate and regioselectivity of the S_NAr reactions of fluorinated heteroaromatics have been investigated previously with reactions of 2,4,6-trifluoropyridine (130) and pentafluoropyridine (129) being reported in the literature.[7,8] Comparison of the products of reaction between these two compounds found that

reaction between 2,4,6-trifluoropyridine (**130**) and the methanolate nucleophile afforded the *para* and *ortho* substituents with a ratio of 78:22 (**Scheme 2.2i**). In contrast, the reaction of pentafluoropyridine **129** and methanolate afforded the *para* substituent only (**Scheme 2.2ii**), suggesting that the position of the fluorine on the heteroaromatic ring plays an important role in regioselectivity.[9]
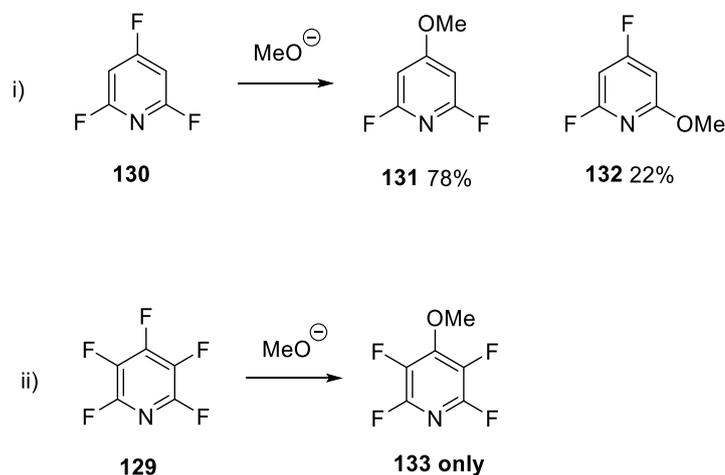
Scheme 2.2: (i) S_NAr of 2,4,6-trifluoropyrdine **130** with methanolate resulting in *para* and *ortho* substituted products **131** and **132** (ii) S_NAr of pentafluoropyridine **129** with methanolate resulting in *para* substituted product **133** only.

The difference in the ratio of products between 2,4,6-trifluoropyridine (**130**) and PFP (**129**) can be explained *via* investigation of the Meisenheimer intermediates and the activating/deactivating effects of fluorine in heteroaromatic environments.
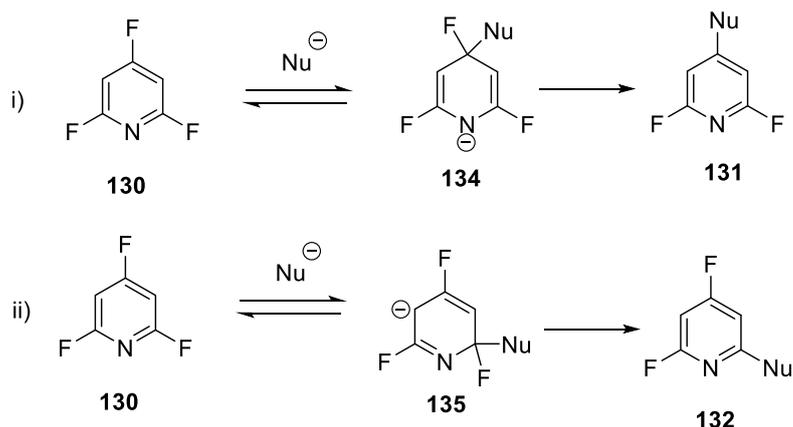
Scheme 2.3: S_NAr of 2,4,6-trifluoropyridine (**130**) with a Nu at the i) *para* position and ii) *ortho* position.

57

When nucleophilic attack occurs at the *para* position of **130** (**Scheme 2.3i**), the resulting Meisenheimer intermediate has a negative charge located on the nitrogen and a fluorine positioned *ortho* to the charged nitrogen, which has a stabilising effect. In intermediate **135**, there is no fluorine attached to the negatively charged carbon, meaning this intermediate is tolerated more than the corresponding PFP analogue **138** (**Scheme 2.4ii**). Therefore, nucleophilic attack at the *para* position is preferable to attack at the *ortho* position, for compound **130**.

Looking at the structures of the Meisenheimer intermediates for the nucleophilic attack of PFP (**129**) at *ortho*, *meta* and *para* fluorine atoms, it would be expected that the fluorine at the *ortho* position would react similarly to the *para* position (**Scheme 2.4**). In the case of nucleophilic attack at the *ortho* (**Scheme 2.4ii**) and *meta* positions (**Scheme 2.4iii**), the negative charge of the Meisenheimer intermediate lies on the carbon of the carbon-fluorine bond. The interaction between the negative charge on the carbon and fluorine lone pairs has a destabilising effect and is therefore unfavourable. Whereas in the case of nucleophilic attack at the *para* position (**Scheme 2.4i**), the negative charge is on the nitrogen and the fluorine *ortho* to this position has a stabilising effect.
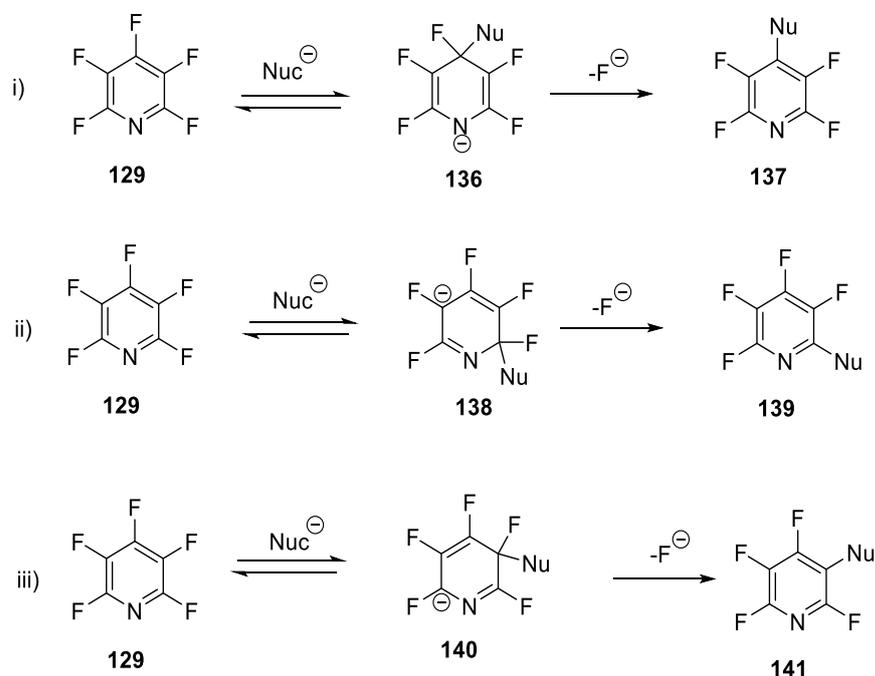
Scheme 2.4: S$_N$Ar of PFP (**129**) with Nu at the i) *para* position ii) *ortho* position and iii) *meta* position.

Kinetic studies have been carried out to determine the importance of fluorine at the *ortho*, *meta* and *para* positions. It was shown that the presence of fluorine at the *ortho* and *meta* positions have a strong activating effect whereas fluorine at the *para* position was deactivating.[7] In the case of perfluoroheteroaryl compounds, the number of activating fluorine atoms dictates the initial site of attack.
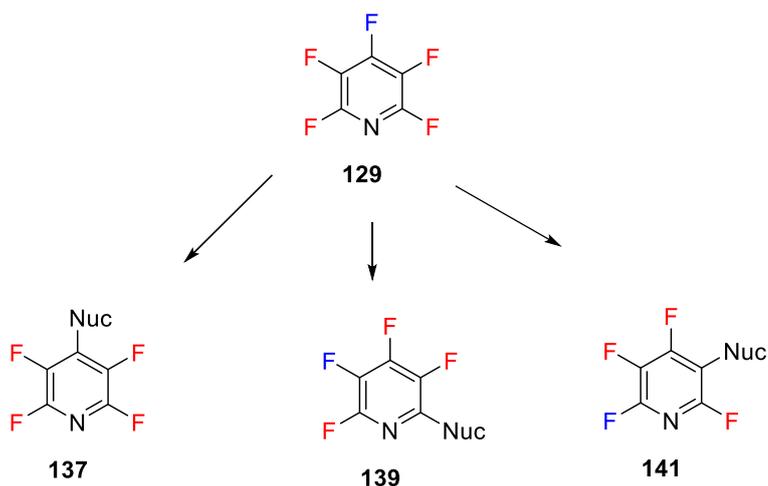
Scheme 2.5: PFP with activating (*ortho* and *meta*) fluorine atoms highlighted in red and deactivating (*para*) highlighted in blue.

The effect of fluorine in activation/deactivation of heteroaromatic systems helps to explain the difference in the ratio of products between 2,4,6-trifluoropyridine (**130**) and PFP (**129**). In the case of 2,4,6-trifluoropyridine (**130**), there are no deactivating fluorine atoms *para* to the site of attack, which means attack occurs at both *para* and *ortho* positions to the nitrogen (**Scheme 2.3**). The aromatic nucleophilic substitution of pentafluoropyridine (**129**), in general, occurs at the *para* position which maximises the number of activating fluorine atoms that are *ortho* and *meta* to the site of attack and minimises deactivating fluorine atoms *para* to the site of attack (**Scheme 2.5**).[9]

Shown in **Table 2.1** are the products resulting from the nucleophilic aromatic substitution of PFP (**129**) with various nucleophiles.[10,11,12,13] All of these reactions resulted in products with the nucleophile reacting *para* to the nitrogen.

Nucleophile	Product	Reaction Condition
 142	 **143**	1) MeCN, rt 2) NaBF₄, MeCN, rt
 144	 **145**	CsF, MeCN, 20°C
 146	 **147**	CsF, 3h
 148	 **149**	*n*-BuLi, TMEDA, THF

Table 2.1: Reaction of PFP (**129**) with various nucleophiles.[10,11,12,13]

Upon further treatment of PFP (**129**) with the same nucleophile, it cannot react *para* to the initially added nucleophile, as is the case for HFB (**47**), since the *para* position is blocked by a nitrogen. Therefore, the second S$_N$Ar reaction occurs *ortho* to the nitrogen (**Scheme 2.6**).[5, 9] It has also been demonstrated that all fluorine atoms in PFP can be substituted when using an excess of nucleophile. However, this takes a significant amount of time (~6 days).[14]

Scheme 2.6: Order of substitution of fluorine in S_NAr reactions with initial substitution at the *para* position followed by *ortho* substitution.

The versatility of the reactivity of PFP (**129**) is further illustrated in **Table 2.2**. A series of 3,5-difluoro-triayloxypyridines were synthesised using the sequential addition of nucleophiles to PFP. Compound **154** was synthesised using PFP as a starting material and stepwise nucleophilic aromatic substitution using the nucleophiles **151**, **152** and finally **153**.[15]

Table 2.2: Synthesis of 3,5-difluoro-triayloxypyridine and 2,6-diphenoxypyridine analogues (R=alkyl).[15]

These compounds were of particular interest as they proved to be biologically active and inhibitors of Xa and factors VII/TF which are involved in blood clotting pathways.[15]

2.1.2 Pentafluoropyridine and its Applications in Peptide Chemistry

In 2013, the Cobb group reported the synthesis of novel tetrafluoropyridine-containing amino acids (**157**, **158**).[16] The nucleophilic nature of oxygen and the predictable regioselectivity of PFP was used to tag an orthogonally protected serine.[16] Boc-Ser-OBn (**156**) was used as the protected amino acid and was prepared in good yield (71%) from the commercially available Boc-Ser-OH. Subsequent reaction of (**156**) with PFP (**129**) gave the the tetrafluoropyridine-containing amino acid (**158**) in a 55% yield (**Scheme 2.7**).[16] It was also possible to prepare **157** in a 30% yield using Boc-Ser-OMe (**155**) as the nucleophile.
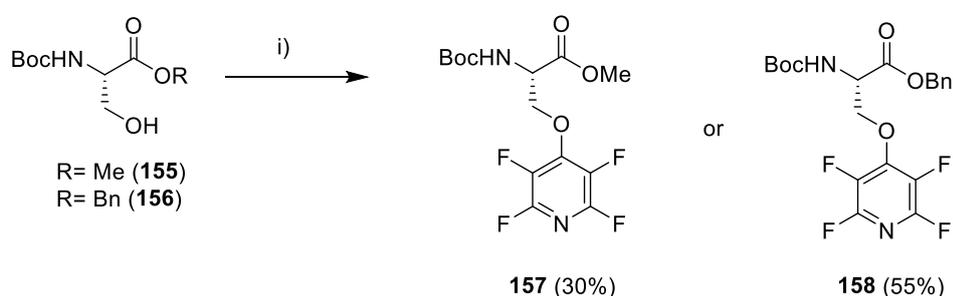
Scheme 2.7: Transformation of Boc-Ser-OBn (**156**) and Boc-Ser-OMe (**155**) when in the presence of PFP.[16] Reagents and conditions: i) PFP, K_2CO_3, MeCN, rt, 20 h.

The next step was to successfully incorporate the PFP-tagged amino acid (**158**) into a peptide. A test tetrafluoropyridine-containing peptide was synthesised using a solution phase peptide synthesis protocol (**Scheme 2.8**). It was found that the tetrafluoropyridine-containing amino acid (**158**) was robust to peptide coupling reactions, and resistant to Boc deprotection (e.g. strong acid).
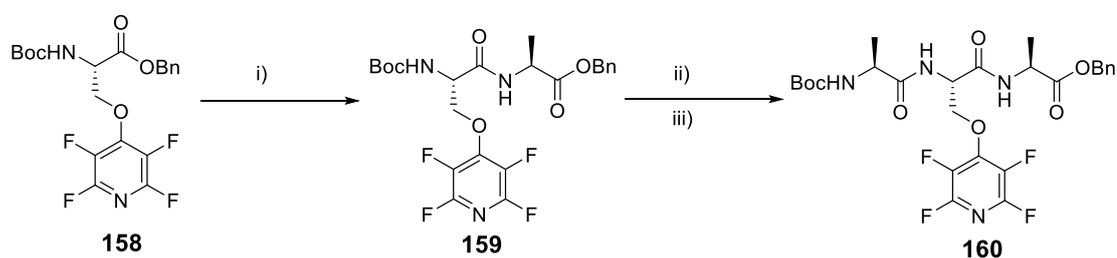
Scheme 2.8: Synthesis of PFP-tagged tripeptide (**160**). Reagents and conditions: i) NH_2-Ala-OBn (**158**), NMM, CH_2Cl, rt, 18 h ii) TFA, CH_2Cl, rt, 4 h iii) Boc-Ala-OH, PyBOP, rt, 18 h.

The tetrafluoropyridyl group has the potential to act as a leaving group due to the electron withdrawing power of fluorine (**Scheme 2.9ii**), or, additional groups could be attached *ortho* to the ring nitrogen in order to functionalise the peptide (**Scheme 2.9iv**). This would offer a new approach for peptide modification and bioconjugation.
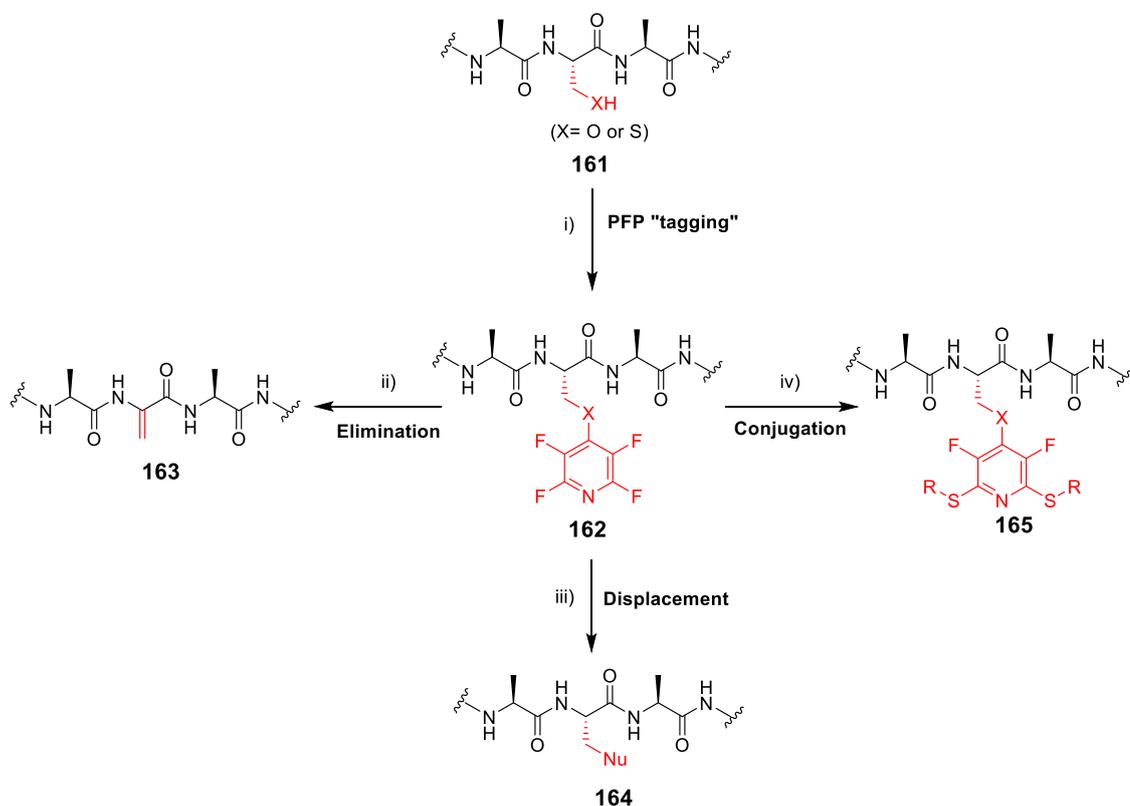
Scheme 2.9: i) Tagging of peptide **161** with PFP and further reaction pathways for fluoropyridine-tagged peptide **162** *via* ii) Elimination resulting in a Dha containing peptide **163** iii) Displacement of fluoropyridine using Nu iv) Aromatic substitution on PFP ring using thiol nucleophiles.

However, before the PFP group could be fully exploited in peptide chemistry it would be important to establish the mechanism by which nucleophilic displacement (**Scheme 2.9iii**) would occur.

2.2 Chapter Aims

As discussed in Chapter 1 (**Section 1.3.1**), Davies and co-workers have shown that reaction between cysteine residues and nucleophilic moieties can proceed *via* an elimination/addition reaction (**Scheme 2.10**).[17] This reaction pathway results in a mixture of isomeric products which, in theory, can be distinguished *via* NMR spectroscopy.
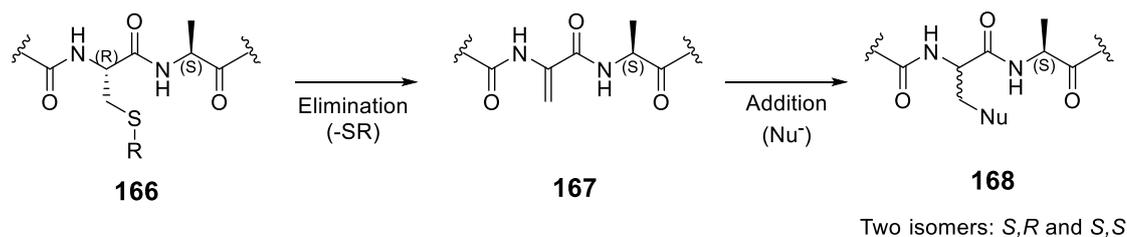
Scheme 2.10 - Elimination/addition reaction pathway and expected stereochemistry of products **168**.

We reasoned that by conjugating PFP (**129**) to a Cys-containing dipeptide, the reaction between the resulting "activated" di-peptide (**169**) and a nucleophile would proceed *via* an S_N2 reaction resulting in only one isomeric product (**170**). If the reaction does not proceed *via* an S_N2 pathway and instead proceeds *via* an elimination/addition reaction, there will be two isomeric products, which will have distinct ^{1}H NMR patterns. The nucleophile used in this study is thiophenolate which allows the reaction to be traced by ^{19}F NMR with the disappearance of peaks from the ^{19}F NMR suggesting the reaction has reached completion.
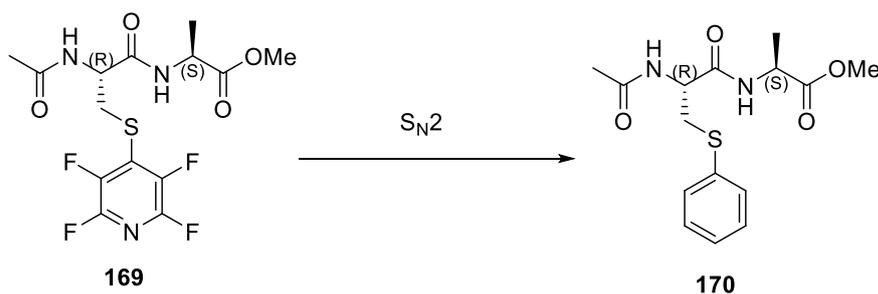
Scheme 2.11: Proposed displacement reaction of activated dipeptide **169** showing the expected stereochemistry of the product **170**.

We envisaged that it would also be of value to determine whether or not changing the amino acid residue would change the reaction mechanism. The target compounds which will be utilised in the mechanistic studies are shown in **Figure 2.3**. These include the PFP (**129**) tagged dipeptides containing cysteine (**169**), penicillamine (**171**) and homocysteine (**172**). It was expected that, upon addition of a nucleophile, attack will occur at the carbon of the cysteine and homocysteine residue in compound **169** and **172** but not in **171,** due to the fact it contains a tertiary carbon which will not readily undergo S_N2 reaction.
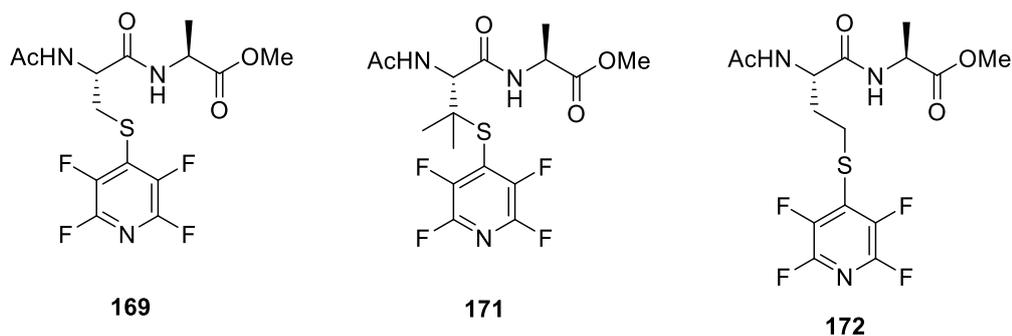
Figure 2.3: Structures of pentafluoropyridine tagged di-peptides targets (**169**, **171** and **172**).

The reaction of **169** and **172** with a nucleophile, paired with ^{1}H NMR data and the non-reaction of compound **171** with a nucleophile would confirm that the reaction proceeds *via* S_N2 and not an elimination-addition pathway.

2.3 Synthesis of PFP-Tagged Dipeptide Targets

2.3.1 Solution Phase Dipeptide Synthesis of Dipeptide Targets

Preparation of the pentafluoropyridine-tagged dipeptides (**169, 171** and **172**) was first investigated *via* solution-phase peptide synthesis. Using standard reaction conditions below, the acetyl-capped dipeptide (**177**) was synthesised.
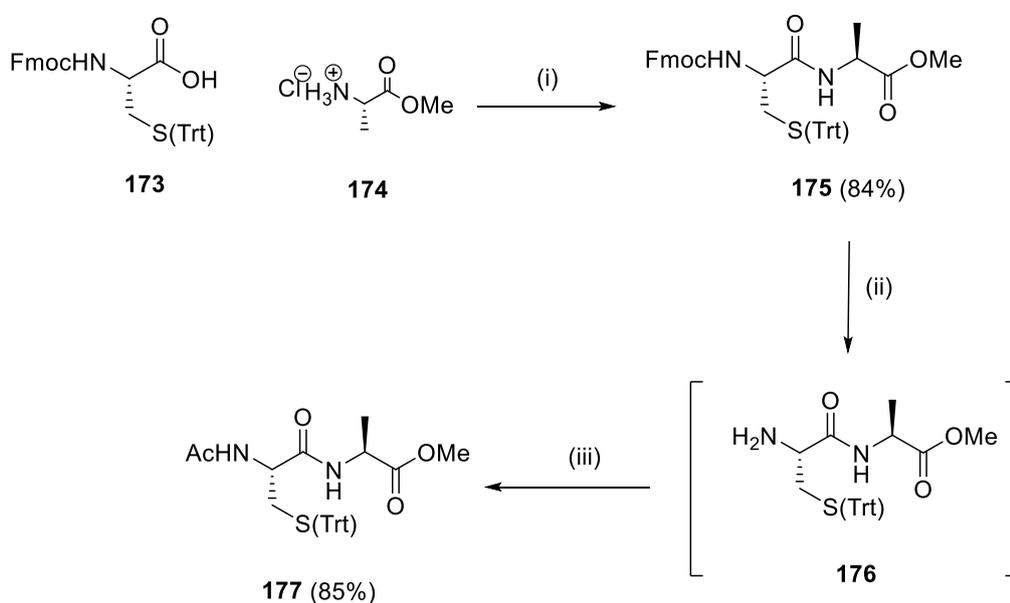
Scheme 2.12: Solution phase synthesis of dipeptide **177**. Reagents and conditions (i) PyBOP, NMM, DCM, rt, 18 h (ii) Piperidine, THF, rt, 3 h (iii) Acetyl chloride, NEt$_3$, DCM, rt, 1 h.

The initial coupling of Fmoc-Cys(trt)-OH (**173**) and H$_2$N-Ala-OMe (**174**) was carried out using benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (PyBOP) and *N*-methylmorpholine (NMM). After separation *via* an aqueous work up, the presence of the desired material in the crude reaction mixture was confirmed *via* ^{1}H NMR spectroscopy. LCMS analysis also identified a peak at *m/z* 692.1 which corresponds to the [M+Na]$^+$ peak of the desired material. A pure sample of **175** was obtained using column chromatography on silica gel.

Deprotection of the Fmoc protecting group in **175** was carried out using piperidine and acetyl chloride was used to acetylate the resultant free amine. This reaction was monitored by TLC and purified *via* column chromatography (Yield: 85%). ^{1}H NMR spectroscopy of the pure reaction mixture confirmed the presence of the capped dipeptide (**177**) with the presence of a signal of 3H at 2.10 ppm which corresponds to the CH$_3$ of the AcNH.

Upon reaction of the acetyl-capped dipeptide (**177**) with PFP (**129**) there was the expected peaks in the ^{19}F NMR of -94.2 ppm (m, 2F) and -164.0 ppm (m, 2F) and expected mass in the LCMS [M+H]$^+$ 398. However, the ^{1}H NMR and analytical HPLC of the crude reaction mixture also showed a number of side products.
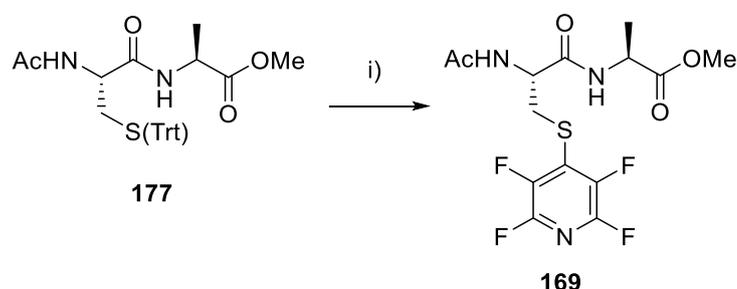
Scheme 2.13: Solution phase synthesis of the activated dipeptide **169**. Reagents and conditions: (i) PFP, DIPEA, DMF, rt, 4 h.

The tagging reaction was carried out on a 50 mg scale with a resulting purified yield of 12%. The purity of the collected fractions was not satisfactory, with the presence of starting material signals (Trt group) at 7.26- 7.44 ppm being apparent (**Figure 2.4**).
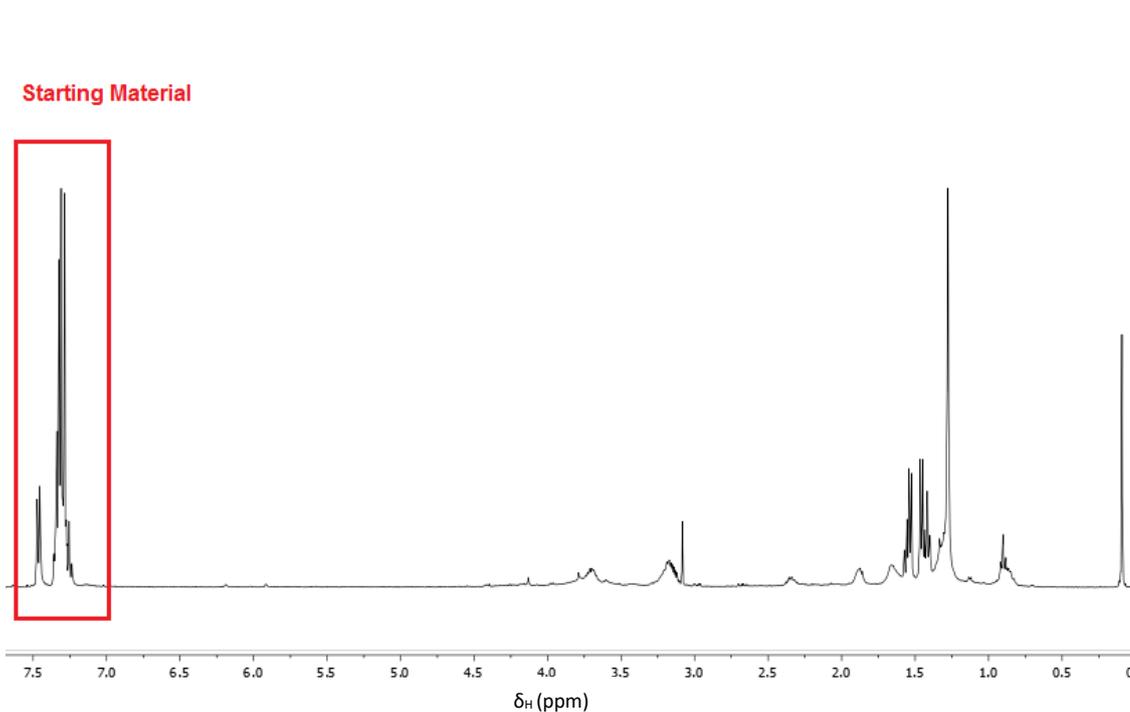
Figure 2.4: ^{1}H NMR of dipeptide **169**, synthesised using solution phase dipeptide synthesis.

Due to the unsatisfactory purity of the PFP-tagged peptide (**169**) and low-yielding reaction conditions, it was decided to investigate alternative reaction conditions for the synthesis of the pentafluoropyridine-tagged dipeptides.

2.3.2 Solid Phase Dipeptide Synthesis of Dipeptide Targets

It was decided that synthesis of the dipeptide analogues would be attempted using solid-phase peptide synthesis. The resin that would be used in the synthesis was RINK amide (RAM) resin and peptide couplings were to be carried out at rt using PyBOP, DIPEA in DMF. Shown in **Scheme 2.14** is the proposed synthesis of the dipeptide analogue using SPPS.
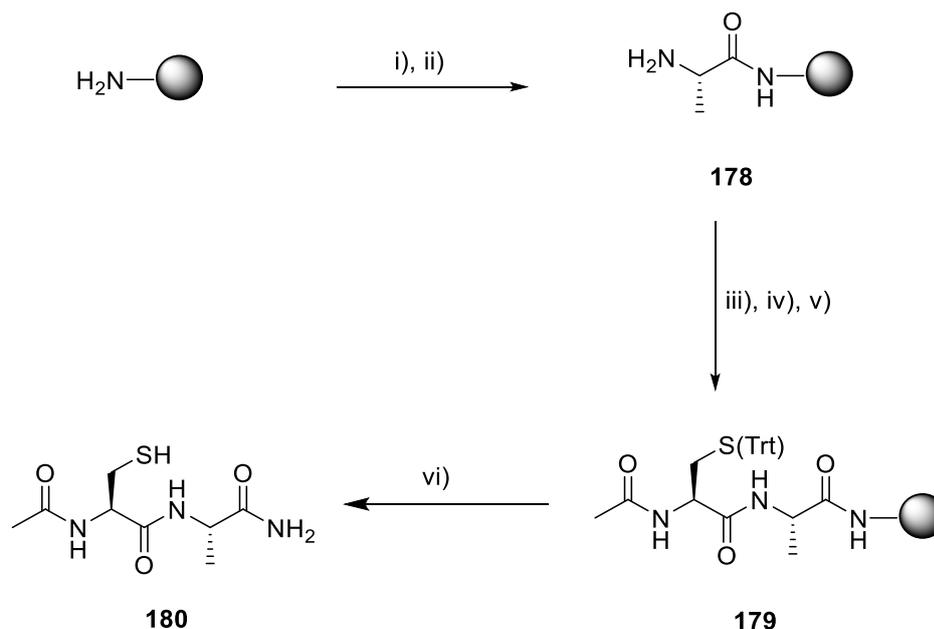
Scheme 2.14: Solid phase synthesis of dipeptide **180**. Reagents and conditions i) Manual Fmoc SPPS- Fmoc-Ala-OH, PyBOP, DIPEA and DMF, rt, (2 × 1 h) couplings ii) Fmoc deporotection- 20% piperidine/DMF, rt, 2 × 30 min iii) Manual Fmoc SPPS- Fmoc-Cys(Trt)-OH, PyBOP, DIPEA and DMF, rt, (2 × 1 h) couplings iv) Fmoc deprotection- 20% piperidine/DMF, rt, 2 × 30 min v) Capping of *N*-terminus- 20% acetic anhydride/DMF, rt, 2 × 30 min vi) Cleavage from resin- 95% TFA: 2.5% TIPS: 2.5 % H_2O.

The first stage in the synthesis was manual Fmoc SPPS of the first dipeptide analogue (**180**), using Fmoc-Ala-OH and Fmoc-Cys(Trt)-OH and the conditions described in **Chapter 6** (**Section 6.4**). Once the peptide was capped at the *N*-terminus it was cleaved from the resin. The resulting dipeptide (**180**) was characterised using LCMS and NMR. The LCMS data with peaks at $[M+H]^+$ 234.2 and $[M+Na]^+$ 256.0 supported the successful synthesis of the target dipeptide (**180**).

^{1}H NMR of the resulting crude peptide agreed with the expected spectra with a quartet at 4.29 ppm corresponding to the C-Hα of Ala, a triplet at 4.59 ppm corresponding to the C-Hα of Cys and singlet at 2.01 ppm with an integral 3H corresponding to the CH_3 of the acetyl group.

The solid-phase synthesis of **180** resulted in fewer side products compared to the solution-phase when analytical HPLC traces and ^{1}H NMR were compared. It was possible to carry out the synthesis on a 0.3 mmol scale which yielded 143 mg of the target dipeptide (**180**). The dipeptide was used directly in the next stage of the synthesis which was the PFP tagging (**Section 2.3.3**).

With the successful preparation of the cysteine dipeptide (**180**), the penicilamine and homocysteine analogues (**183** and **186**) were then also synthesised using a solid-phase peptide synthesis approach. The penicilamine analogue (**183**) was synthesised on a 0.2 mmol scale, using the same reaction conditions as the cysteine analogue (**Scheme 2.15**).

Scheme 2.15: Solid phase synthesis of dipeptide (**183**). Reagents and conditions i) Manual Fmoc SPPS- Fmoc-Ala-OH, PyBOP, DIPEA and DMF, rt, (2 × 1 h) couplings ii) Fmoc deporotection- 20% piperidine/DMF, rt, 2 × 30 min iii) Manual Fmoc SPPS- Fmoc-Pen(Trt)-OH, PyBOP, DIPEA and DMF, rt, (2 × 1 h) couplings iv) Fmoc deprotection- 20% piperidine/DMF, rt, 2 × 30 min v) Capping of *N*-terminus- 20% acetic anhydride/DMF, rt, 2 × 30 min vi) Cleavage from resin- 95% TFA: 2.5% TIPS: 2.5% H_2O.

The LCMS analysis of the penicilamine analogue **183** showed the expected mass of $[M+H]^+$ 262.8 and $[M+Na]^+$ 284.2. The ^{1}H NMR of the crude dipeptide **183** was shown to contain the relevant peaks- a doublet at 1.31 ppm of 3H corresponding to CH_3 of the Ala, a singlet at 1.36 ppm of 3H corresponding to one of the CH_3 of the Pen side chain, a singlet at 1.42 ppm of 3H corresponding to the other Pen side chain, a singlet at 3.77 ppm of 1H corresponding to the C-Hα of Pen side chain and quartet at 4.26 ppm of 1H corresponding to C-Hα of Ala side chain.

The homocysteine analogue (**186**) was also synthesised on a 0.2 mmol scale (**Scheme 2.16**).
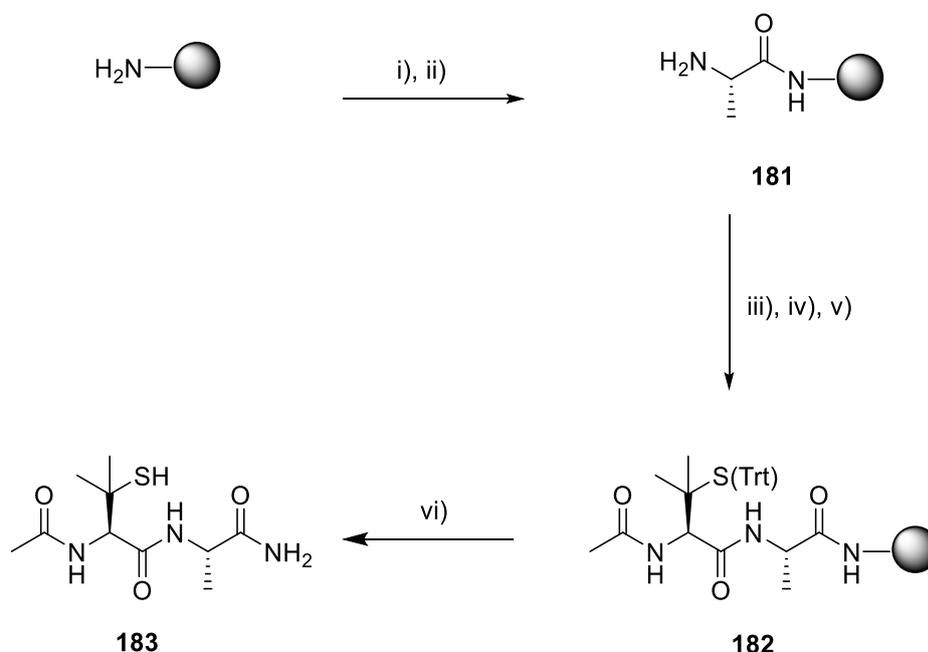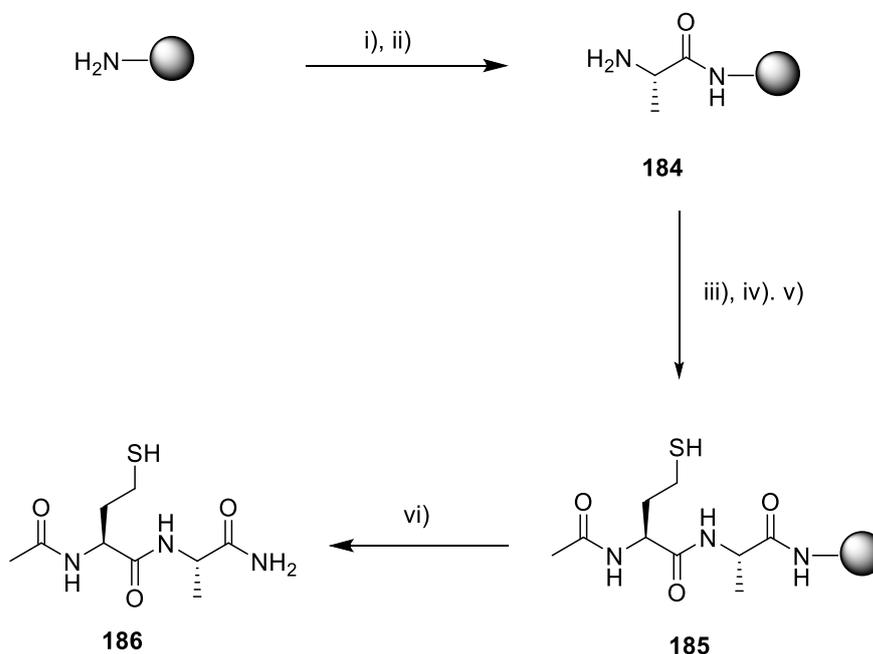

Scheme 2.16: Solid phase synthesis of dipeptide (**186**). Reagents and conditions i) Manual Fmoc SPPS- Fmoc-Ala-OH, PyBOP, DIPEA and DMF, rt, (2 × 1 h) couplings ii) Fmoc deporotection- 20% piperidine/DMF, rt, 2 × 30 min iii) Manual Fmoc SPPS- Fmoc-hCys(Trt)-OH, PyBOP, DIPEA and DMF, rt, (2 × 1 h) couplings iv) Fmoc deprotection- 20% piperidine/DMF, rt, 2 × 30 min v) Capping of *N*-terminus- 20% acetic anhydride/DMF, rt, 2 × 30 min vi) Cleavage from resin- 95% TFA: 2.5% TIPS: 2.5% H_2O.

The LCMS of the homocysteine analogue **186** showed the expected mass $[M+H]^+$ 248.1 $[M+Na]^+$ 270.3. The ^1H NMR of the crude dipeptide (**186**) was shown to contain the relevant peaks- singlet at 1.31 ppm and 1.34 ppm of 3H corresponding to the CH_3 of Ala side chain, singlet at 1.97 ppm of 3H corresponding to terminal acetyl CH_3, quartet at 2.59 ppm of 2H corresponding to CH_2 of Hcys side chain and multiplet at 4.33 ppm of 2H corresponding C-Hα of the Ala and Hcys side chains.

In summary, the cysteine (**180**), penicilamine (**183**) and homocysteine (**186**) dipeptides were successfully synthesised on a multi-hundred milligram scale and after cleavage from the resin it was determined by analysis of the ^1H NMR that further purification was not required.

2.3.3 Tagging of Dipeptides with Pentafluoropyridine

The crude cysteine dipeptide (**180**) was dissolved in DMF and reacted with PFP (**129**) and DIPEA, on a 0.1 mmol scale resulting in the tagged dipeptide analogue **187** (**Scheme 2.17**). The incorporation of pentafluoropyridine into the dipeptide was supported by ^{1}H NMR and ^{19}F NMR as well as LCMS.
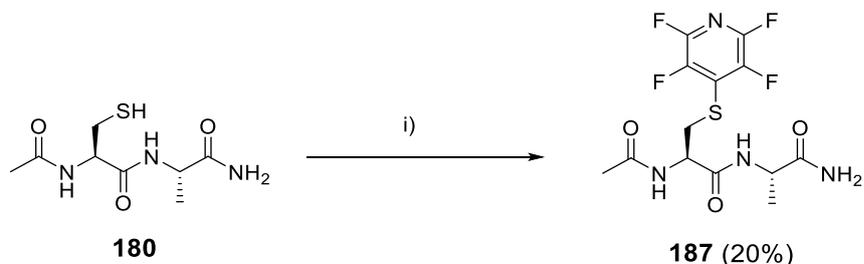
Scheme 2.17: Tagging reaction of cysteine dipeptide (**187**) with PFP. Reagents and conditions i) Pentafluoropyridine, DIPEA, DMF, rt, 4 h.

The LCMS of the crude dipeptide analogue **187** is shown below (**Figure 2.5**) with the $[M+H]^+$ 383.4, $[M+Na]^+$ 405.8 and $[2M+H]^+$ 765.4 all clearly visible.
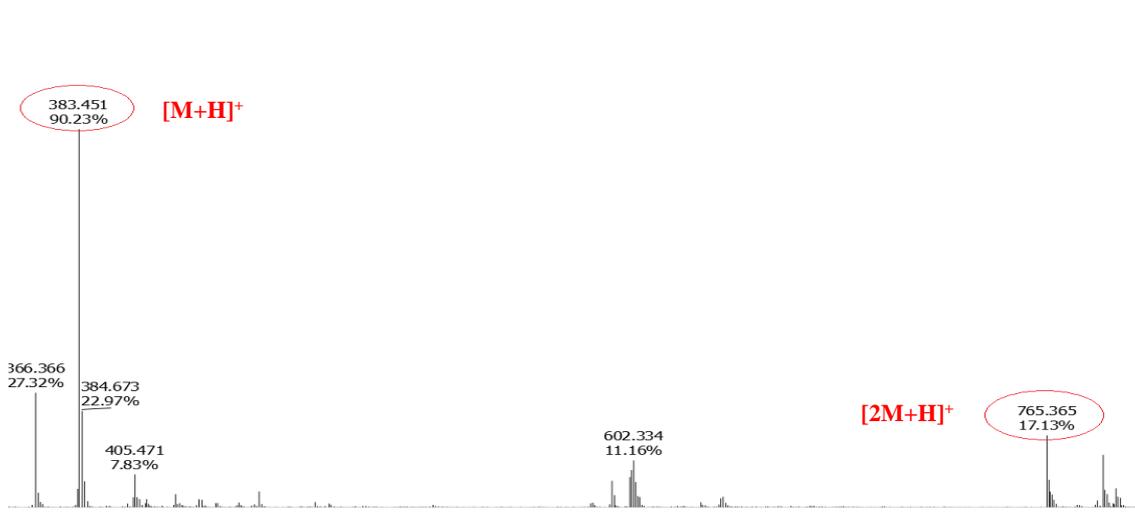
Figure 2.5: MS of tagged dipeptide **187**.

^{1}H NMR analysis of the crude dipeptide was carried out and signals relating to the desired material such as the quartet of 1H at 4.41 ppm corresponding to C-Hα of the Ala side chain, triplet of 1H at 4.22 ppm corresponding to C-Hα of the Cys side chain, singlet of 3H at 1.91 ppm corresponding to terminal acetyl CH_3 and doublet of 3H at 1.23 ppm corresponding to CH_3 of Ala side chain are all present in the reaction mixture.

Upon closer inspection of the ^{1}H NMR there appeared to be doubling of the signals. Shown below is an expansion of the ^{1}H NMR for **187** in the region between 3.30 ppm and 4.60 ppm which illustrates the presence of a possible second dipeptide product in the reaction mixture. Signals at 4.19 ppm, 4.21 ppm, 4.23 ppm, 1.39 ppm, 1.41 ppm, 1.43 ppm, and 1.45 ppm illustrate the presence of a second product.
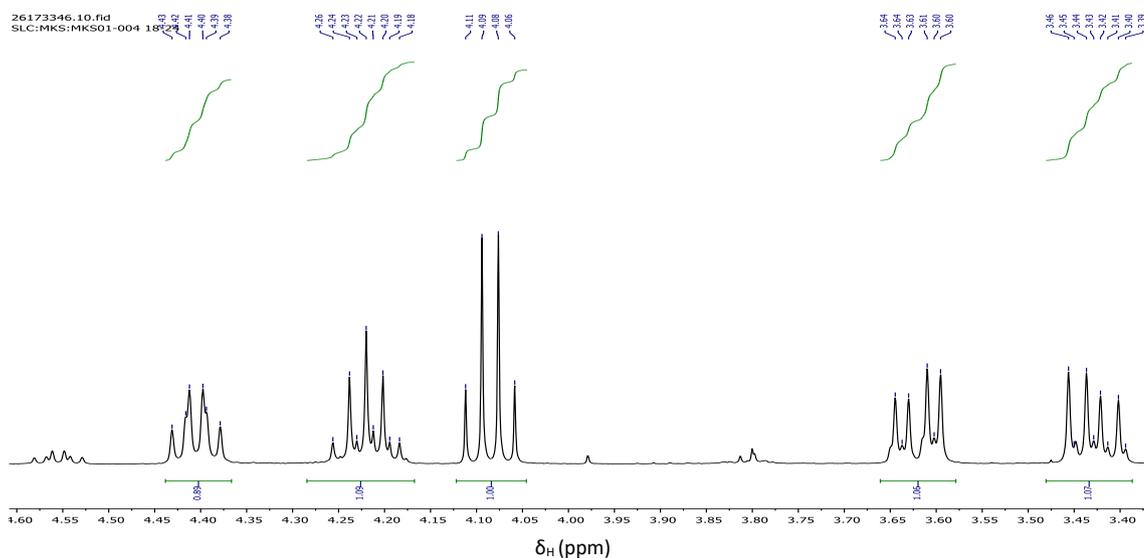
Figure 2.6: Expansion of ^{1}H NMR region 3.30 ppm- 4.60 ppm of **187**.

It was hypothesised that the additional signals in the NMR could either be due to the formation of diastereomers or due to the presence of conformers. It is possible that the tetrafluoropyridyl group could act as a leaving group, causing the formation of a Dha residue. Addition could then occur between Dha dipeptide and the $C_5F_4NS^-$ ion producing diastereomers. However, this series of events was deemed to be highly unlikely due to the low nucleophilicity of the $C_5F_4NS^-$ ion.

In order to determine if the additional peaks from the ^{1}H NMR spectrum were due to the presence of diastereomers or conformers, a varied temperature (VT) ^{1}H NMR experiment was conducted from rt to 70 °C in CD_3CN. If conformers were present, then convergence of peaks would be observed and upon heating of the sample, convergence was observed. This indicated the presence of confomers as opposed to diastereomers. **Figure 2.7** shows how the peaks in **Figure 2.6** start to converge upon heating the sample. At lower temperature, the thermal energy barrier has not been reached and the molecule is stuck in two confomers. As temperature increases, the molecule gains

energy, bond rotation occurs, and the protons become equivalent resulting in a convergence of the observed peaks in the NMR spectra.
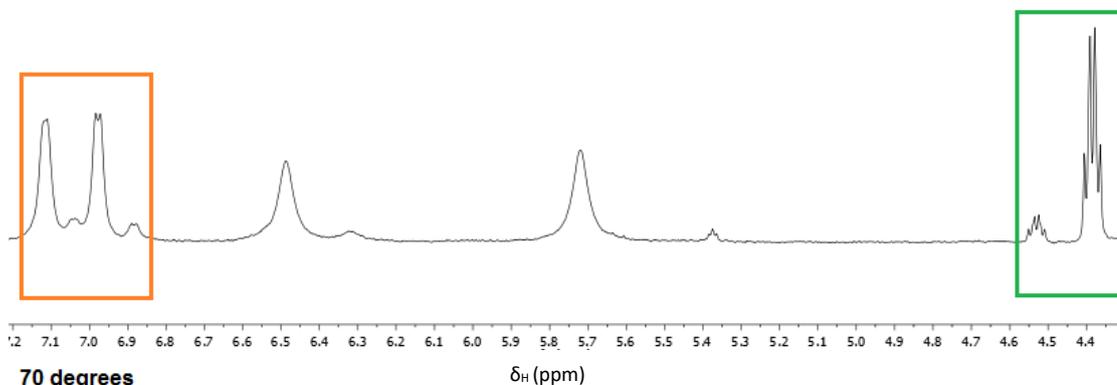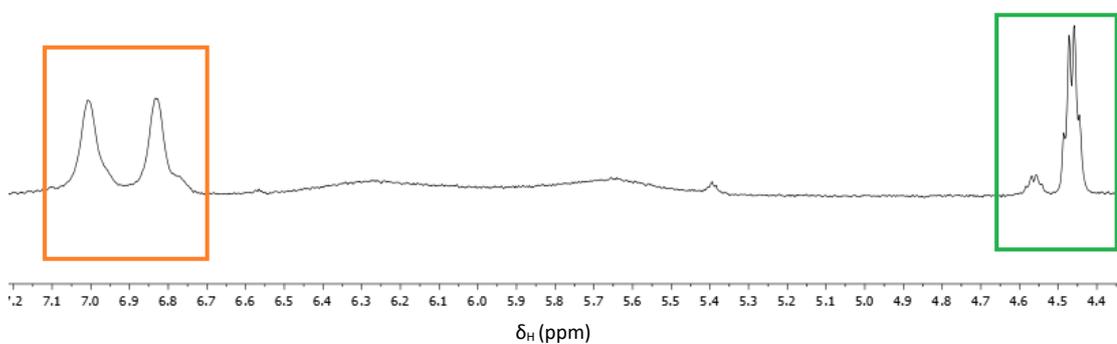
Figure 2.7: VT ^{1}H NMR spectra of compound **187** at 25 °C and 70 °C, illustrating the convergence of signals and the presence of confomers as opposed to diastereomers.

In an effort to separate the desired material from the reaction mixture, the tagged cysteine dipeptide was purified *via* column chromatography on silica gel and analysed *via* ^{1}H NMR and ^{19}F NMR. Shown in **Figure 2.8** is the same expansion region that was illustrated in **Figure 2.6** and clearly shows there are no additional signals that were present before purification.
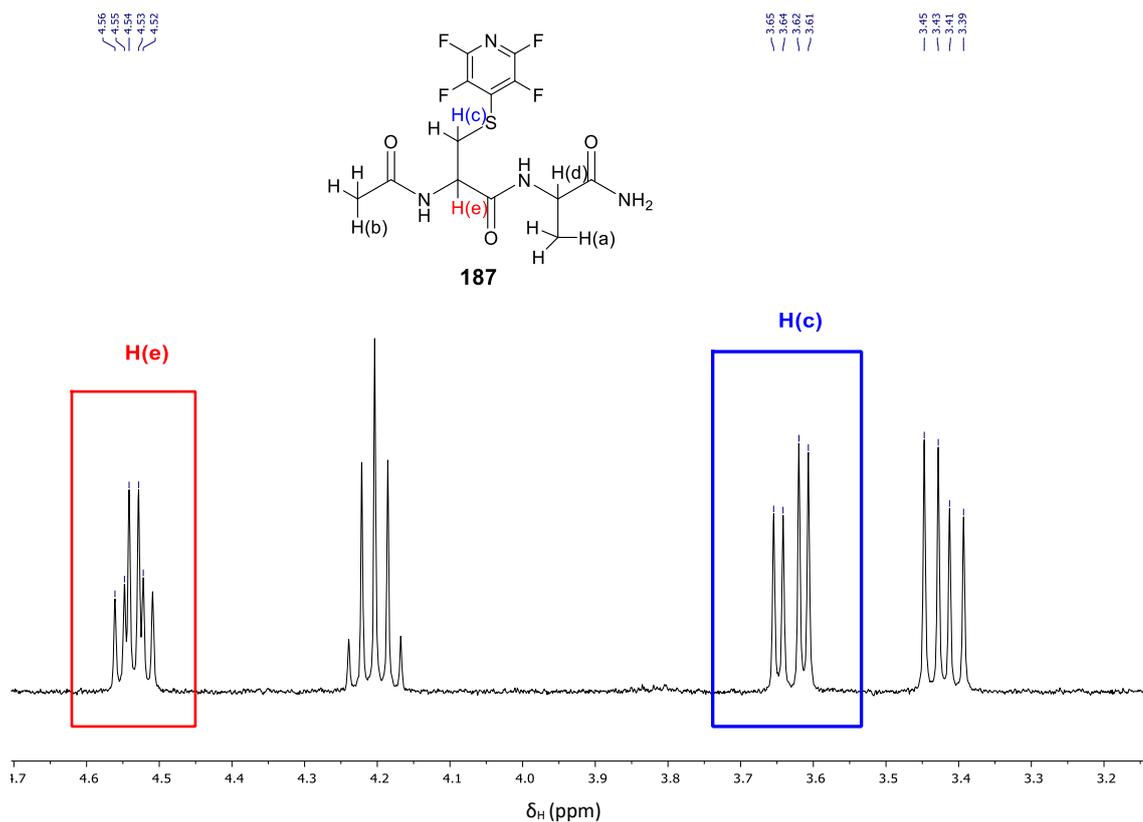
Figure 2.8: Expansion of ^{1}H NMR region 3.30 ppm- 4.60 ppm of purified PFP tagged dipeptide **187**.

The ^{1}H NMR of compound **187** shown in **Figure 2.8** has a multiplet at 4.53 ppm corresponding to the proton labelled H(e) and doublet at 3.62 ppm corresponding to proton labelled H(c) in **Figure 2.9**. It is expected that upon displacement of the tetrafluoropyridine moiety these signals will shift considerably.
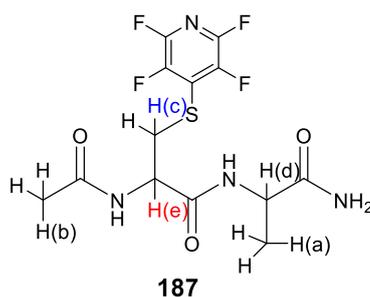
Figure 2.9: Labelled protons of compound **187.**

[19]F isotope is highly sensitive to NMR and is often used in conjunction with ^{1}H NMR in analysis of organofluoro compounds. The multiplicity of ^{19}F NMR provides information on the chemical environment and [19]F-^{19}F coupling constants are often a lot larger than observed in ^{1}H NMR. ^{19}F NMR of compound **187** was carried out and the resulting spectrum shown below with the relevant areas expanded. Analysis of the ^{19}F NMR spectra of the purified product showed signals at -94.0 ppm and -137.9 ppm (**Figure 2.10**).
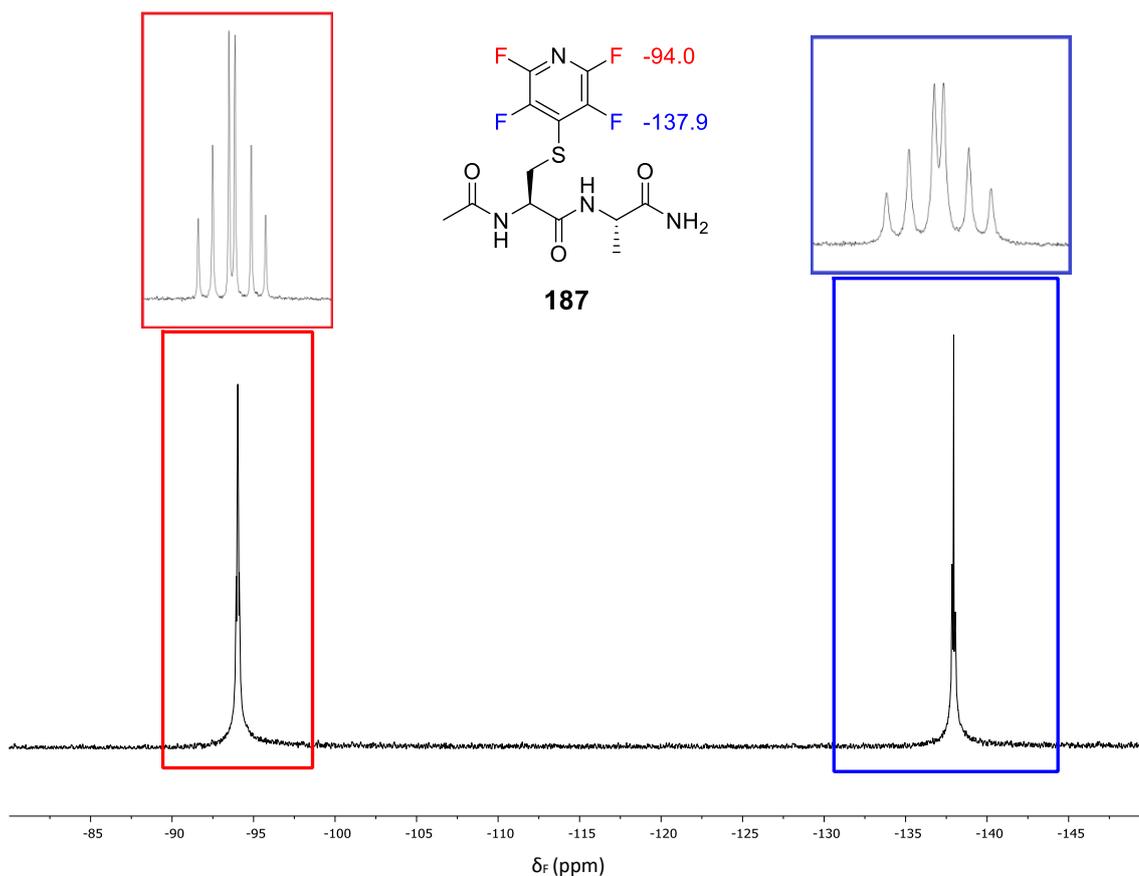
Figure 2.10: ^{19}F NMR of purified tagged dipeptide **187** (solvent CD$_3$CN).

Analaysis of the splitting pattern of the ^{19}F NMR of **187** is complicated due to the apparent overlap of signals. Examples of PFP-tagged amino acids have previously been reported as multiplets when analysed by ^{19}F NMR due to the complex splitting pattern.[16,19] Each fluorine of **187** is represented in the ^{19}F NMR as a doublet of doublet of doublets (ddd) and, due to the fact these fluorines are inequivalent (caused by the proximity of the fluorines to stereocentres), signal overlap occurs resuting in the splitting pattern and integration pattern shown in **Figure 2.10**.

For the peaks highlighted in blue (**Figure 2.10**), the splitting pattern of both fluorines are ddd with *J* values 7.5 Hz, 11.2 Hz and 26.3 Hz (**Figure 2.11**) and chemical shifts -137.9 ppm and -138.0 ppm. The proximity of these signals results in overlap and the non-typical integration pattern of 1:2:3:3:2:1 (**Figure 2.10**). The peaks highlighted in red (**Figure 2.10**) are also in an overlapping ddd formation with *J* values 7.5 Hz, 15.0 Hz and 30.0 Hz and chemical shifts -94.0 ppm and -94.1 ppm (**Figure 2.11**).
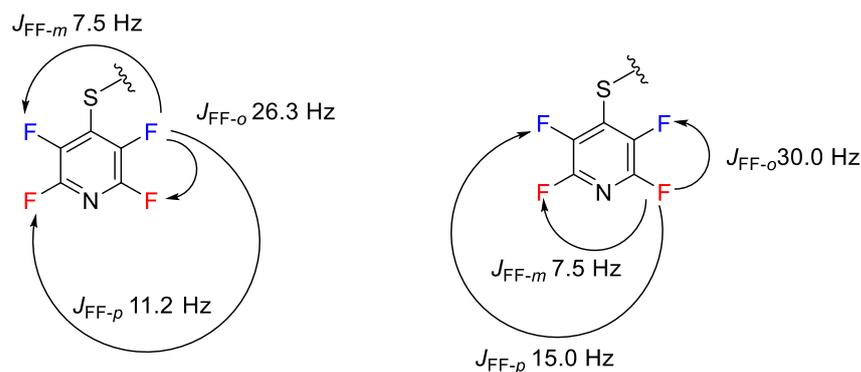
Figure 2.11: Flourine-fluorine coupling constants of PFP moiety of compound **187**.

The presence of only two signals in the ^{19}F NMR support that PFP is conjugated to the dipeptide *para* to the nitrogen. If PFP (**129**) was attached to the thiol with an alternative regiochemistry as shown below, then the splitting pattern in the ^{19}F NMR would not be the same as the one observed in **Figure 2.12**.
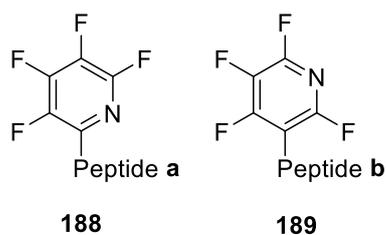
Figure 2.12: Alternative conformation of PFP-tagged peptide.

With reaction and purification conditions optimised it was decided to synthesise the penicillamine (**190**) and homocysteine (**191**) PFP tagged di-peptides. The crude penicillamine dipeptide (**183**) was dissolved in DMF and reacted with pentafluoropyridine and DIPEA on a 0.5 mmol scale resulting in the tagged dipeptide analogue **190** (**Scheme 2.18**).
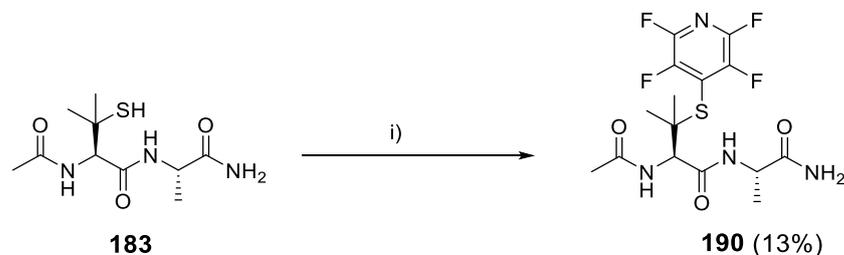
Scheme 2.18: Tagging reaction of penicillamine dipeptide **193** with pentafluoropyridine. Reagents and conditions i) Pentafluoropyridine, DIPEA, DMF, rt, 4 h.

Peptide **190** was purified *via* column chromatography and analysed using LCMS, ^{1}H NMR and ^{19}F NMR. The LCMS of the purified dipeptide analogue **190** is shown below with the [M+H]$^+$ 411.1.
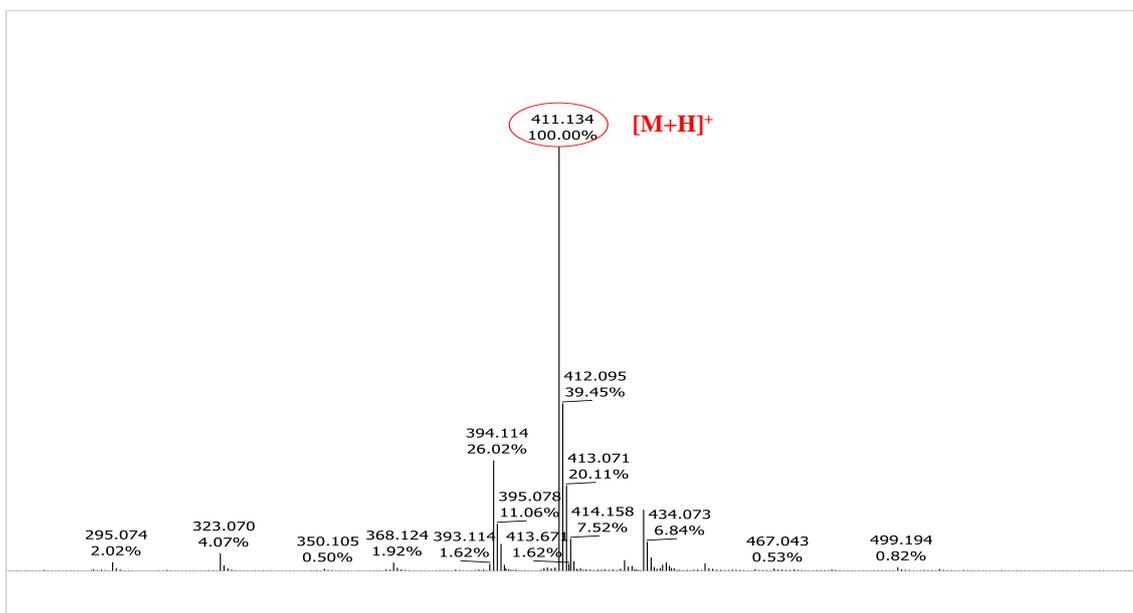
Figure 2.13: MS of tagged dipeptide **190.**

The ¹H NMR of the purified peptide **190** is shown below.
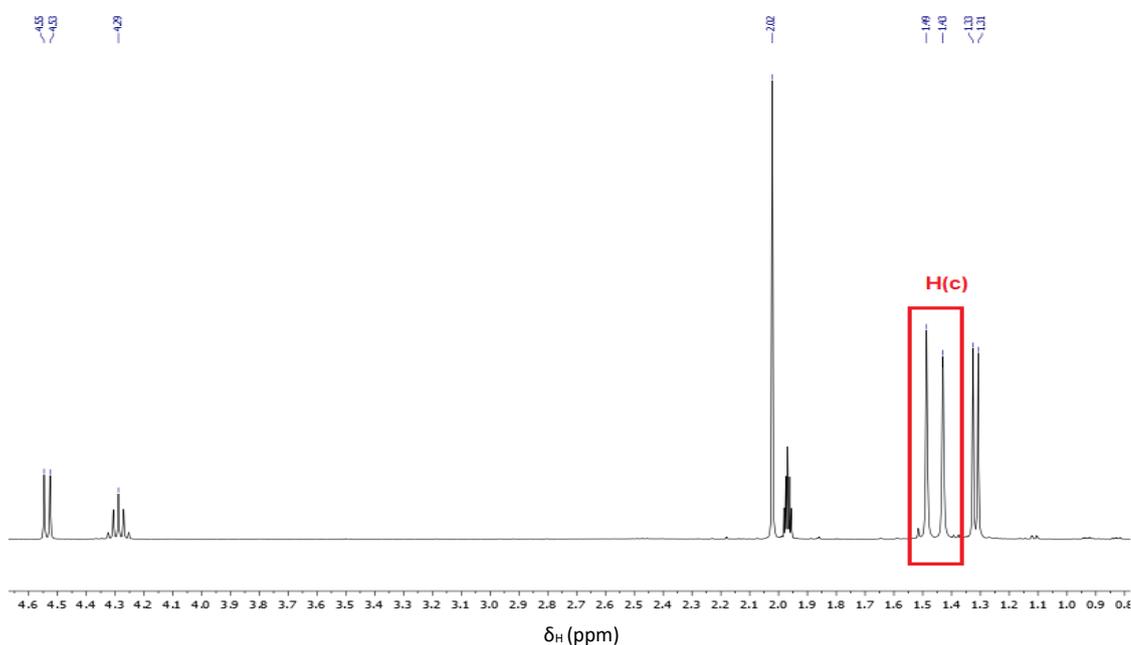
Figure 2.14: ¹H NMR of pure dipeptide **190** (solvent CD₃CN).

The ¹H NMR obtained for purified **190** confirmed its structure. At 1.43 ppm and 1.49 ppm there are singlets corresponding to the protons labelled H(c) and H(c') which are not present in the spectra of compound **187** and corresponds to the protons of the CH₃ of the Pen side chain. There is also no doublet at 3.62 ppm which corresponds to the CH₂ group of the cysteine side chain which can be seen in **Figure 2.8** but not in **Figure 2.14**.
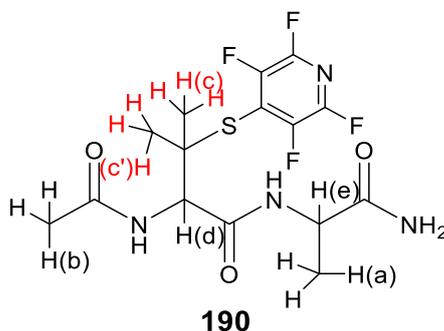
Figure 2.15: Labelled protons of compound **190**.

The peaks highlighted in **Figure 2.16** are displayed as overlapping ddd patterns (as in **Figure 2.10**). The ^{19}F NMR of the purified peptide **190** has distinct signals at -132.8 ppm and -92.7 ppm. The peaks highlighted in red (**Figure 2.16**) are in an overlapping ddd formation with *J* values 7.5 Hz, 15.0 Hz and 33.5 Hz and chemical shifts -92.6 ppm and -92.7 ppm. The peaks highlighted in blue (**Figure 2.16**) are also in an overlapping ddd formation with *J* values 7.5 Hz, 15.0 Hz and 33.5 Hz and chemical shifts -132.7 ppm and -132.9 ppm.
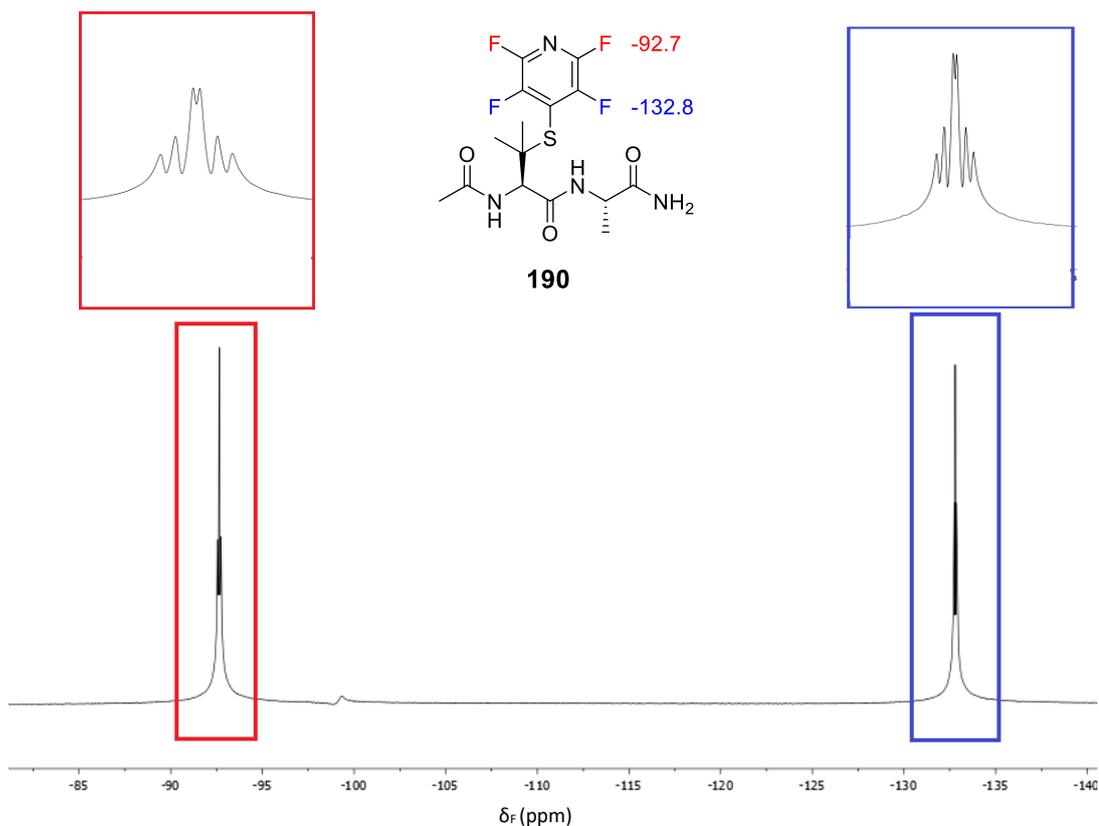
Figure 2.16: ^{19}F NMR of pure dipeptide **190** (solvent CD$_3$CN).

79

The crude homocysteine dipeptide (**186**) was dissolved in DMF and reacted with pentafluoropyridine and DIPEA on a 0.6 mmol scale resulting in the tagged dipeptide analogue **191** (**Scheme 2.19**).
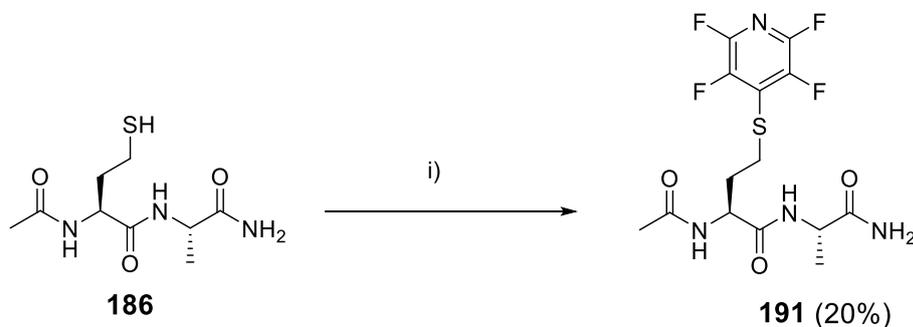
Scheme 2.19: Tagging reaction of homocysteine dipeptide **186** with pentafluoropyridine. Reagents and conditions i) Pentafluoropyridine, DIPEA, DMF, rt, 4 h.

Peptide **191** was purified *via* column chromatography and analysed using LCMS, ^{1}H NMR and ^{19}F NMR. The LCMS of the purified dipeptide analogue **191** is shown below with the $[M+H]^+$ 397.0.
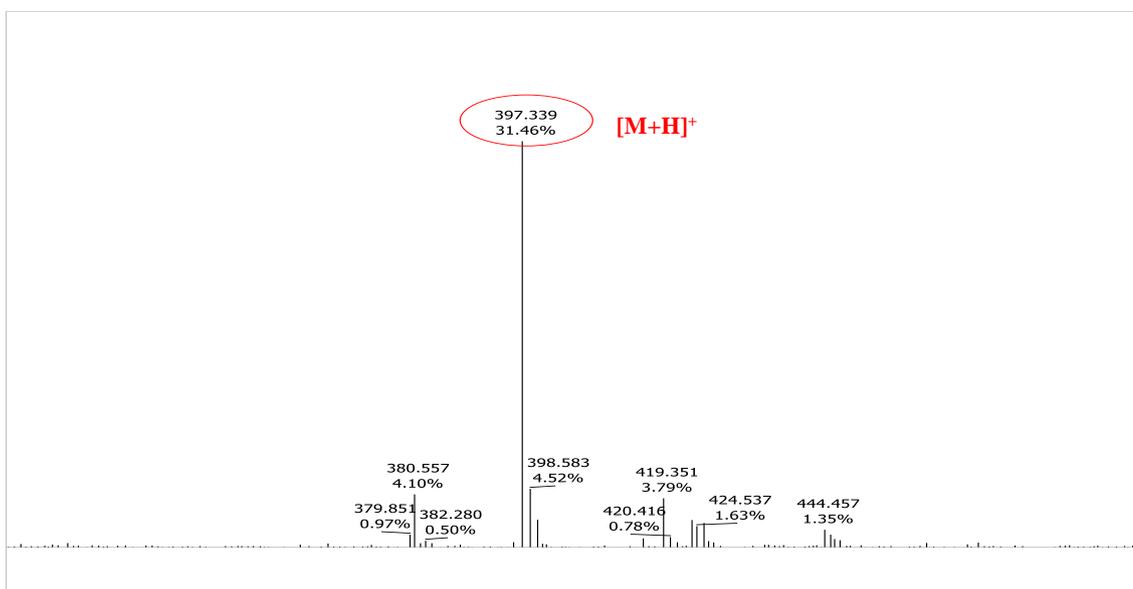
Figure 2.17: MS of PFP-tagged dipeptide **191.**

The ^{1}H NMR of the purified peptide **191** is shown below.
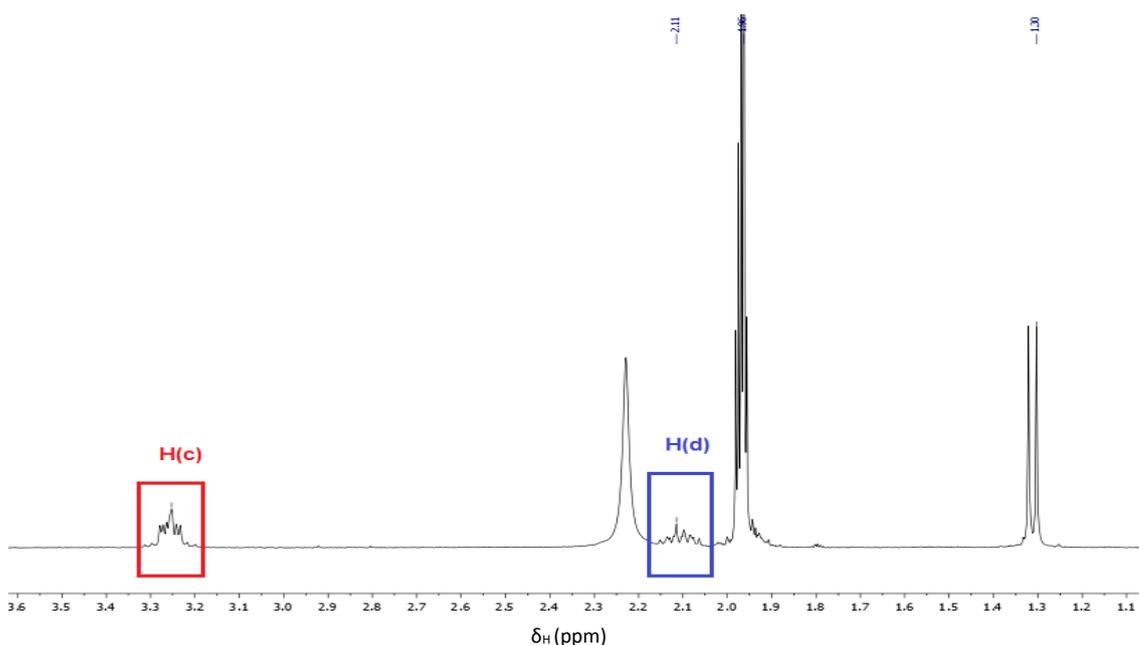
Figure 2.18: ^{1}H NMR of purified tagged dipeptide **191** (solvent CD$_3$CN).

The ^{1}H NMR of homocysteine analogue **191** also confirmed that the expected product had been prepared. At 2.11 ppm there is a multiplet corresponding to the proton labelled H(d) and the multiplet at 3.25 ppm which corresponds to H(c). The upfield shift and change in multiplicity of the proton H(e)corresponds to the insertion of a CH$_2$ group.

191
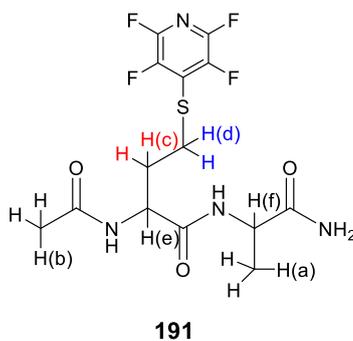
Figure 2.19: Labelled protons of compound **191**.

The peaks highlighted in **Figure 2.20** are displayed as overlapping ddd patterns (as in **Figure 2.10**). The ^{19}F NMR of the purified peptide **191** has distinct signals at -139.1 ppm and -94.2 ppm. The peaks highlighted in red (**Figure 2.20**) are in an overlapping ddd formation with *J* values 7.5 Hz, 15.0 Hz and 30.0 Hz and chemical shifts -94.1 ppm and -94.3 ppm. The peaks highlighted in blue (**Figure 2.20**) are also in an overlapping ddd formation with *J* values 7.5 Hz, 11.2 Hz and 30.0 Hz and chemical shifts -139.1 ppm and -139.2 ppm.
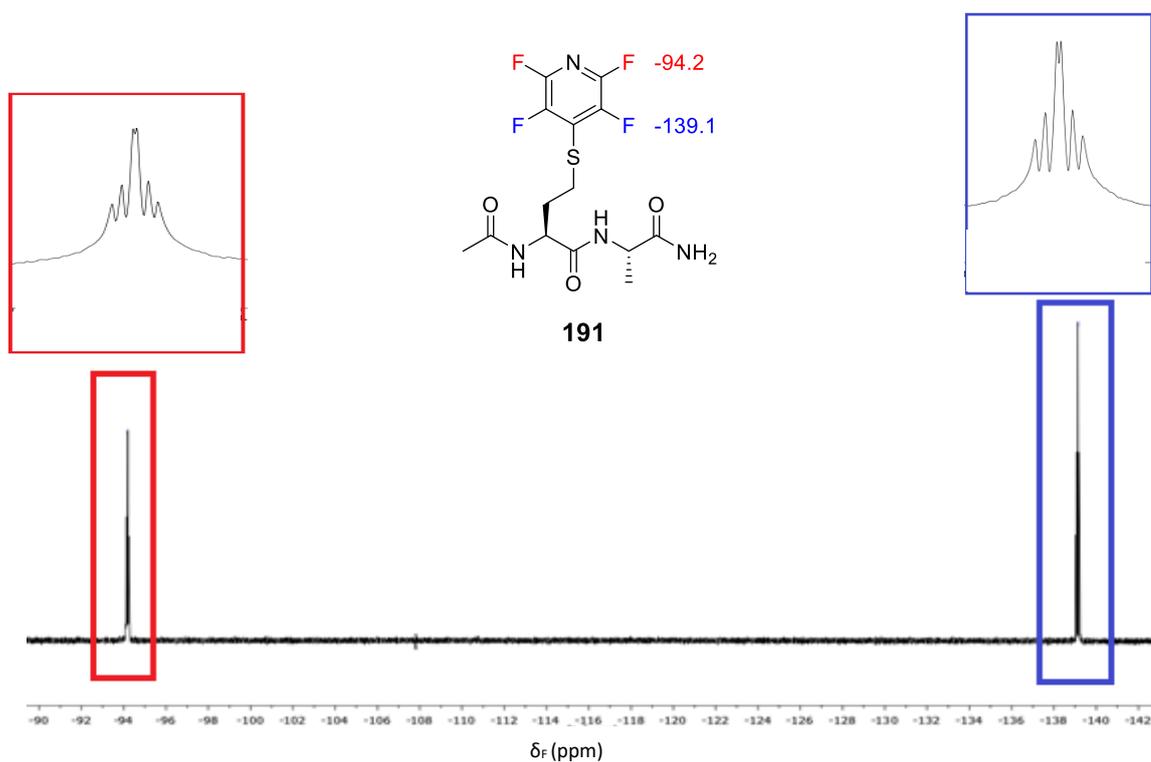

Figure 2.20: ^{19}F NMR of purified dipeptide **191** (solvent CD$_3$CN).

2.4 Displacement and Conjugation Reactions

The conjugation of peptides to small molecules has been well investigated (**Chapter 1**) and the benefits of such analogues can include increased solubility, increased stability and increased cell permeability.

By adopting the experimental procedures described within this chapter, it may be possible to selectively conjugate a variety of molecules such as thio-PEG and thiosugars to improve stability and lipophilicity of biologically active peptides. This approach is beneficial as the tagging reaction with PFP (**129**) followed by displacement using a thiol (**Scheme 2.20i**) is believed to occur *via* S_N2 and therefore produce only one stereoisomer. It would also allow the direct conjugation of the nucleophile to the peptide.

Before this reaction was carried out using thio-PEG and thiosugars on biologically active peptides, the displacement reaction was carried out on the small pentafluoropyridine-tagged dipeptide (**187, 190** and **191**) using a sodium thiophenolate. These model systems were characterised using ^{1}H NMR, ^{19}F NMR and MS to determine if the reaction had been successful.

2.4.1 Determination of Displacement Reaction Mechanism

There are four possible outcomes under the reaction conditions shown **Scheme 2.20** summarises all possible outcomes. The first outcome is the nucleophilic substitution reaction, which would occur *via* an S_N2 mechanism (**Scheme 2.20i**). This is the preferred mechanism for reaction because it would retain the stereochemistry at cysteine.
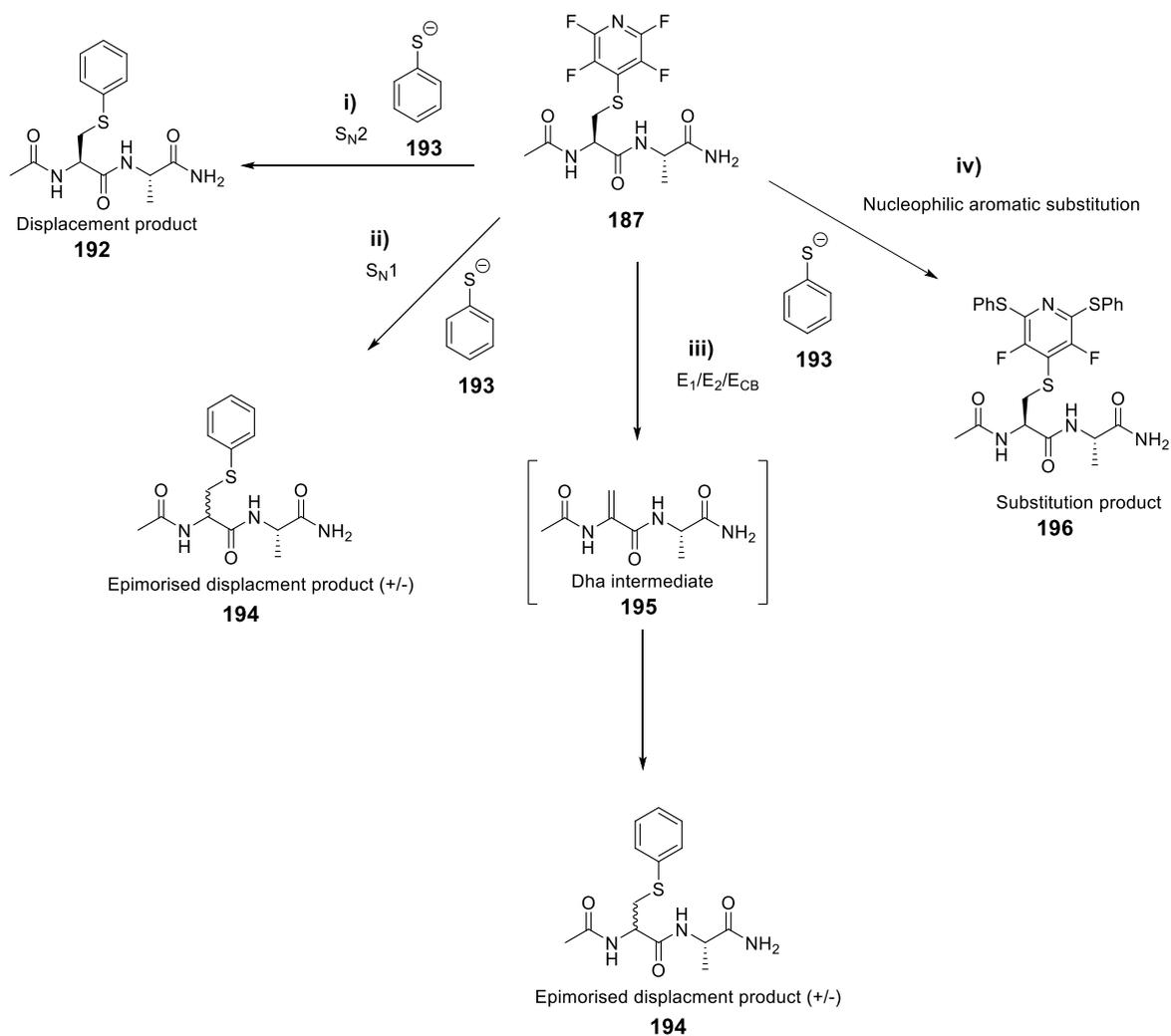
Scheme 2.20: Scheme to show the possible reaction pathways which could occus upon reaction of compound **187** with thiophenolate (**193**) i) S_N2 reaction, **ii)** S_N1 reaction, **iii)** $E_1/E_2/E_{CB}$ reaction, **iv)** S_NAr reaction.

The second possible outcome is an elimination-addition mechanism which occurs *via* the Dha intermediate (**Scheme 2.20iii**). Conjugate addition then occurs at the alkene in the Dha residue. The elimination-addition mechanism would remove the stereochemistry at cysteine and diastereomers would be formed. Chalker *et al.* have already demonstrated that conversion of the thiol in cysteine to Dha is one way of selectively conjugating functional groups to cysteine.[18] Finally, it is possible that further S_NAr reactions at the tetrafluoropyridyl group could occur *ortho* to the nitrogen. (**Scheme 2.20iv**). This could occur instead of the intended displacement reaction since the carbon-fluorine bonds *ortho* to the ring nitrogen are easily accessible and highly polarised. It is only by experiment that the preferred mechanism could be determined.

The tagged dipeptide analogue **187** was reacted with sodium thiophenolate (**193**) and DIPEA in MeCN/H$_2$O (**Scheme 2.21**).
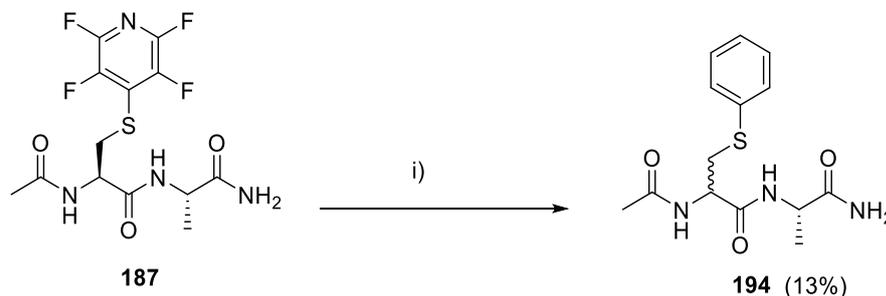
Scheme 2.21: Displacement reaction of PFP-tagged dipeptide **187** with thiophenolate. Reagents and conditions i) Thiophenolate, DIPEA, MeCN/H2O, rt, 4 h.

The crude reaction mixture was analysed using LCMS and the desired product [M+H]$^+$ 310.0 and [2M+H]$^+$ 618.6 was observed (**Figure 2.21**).
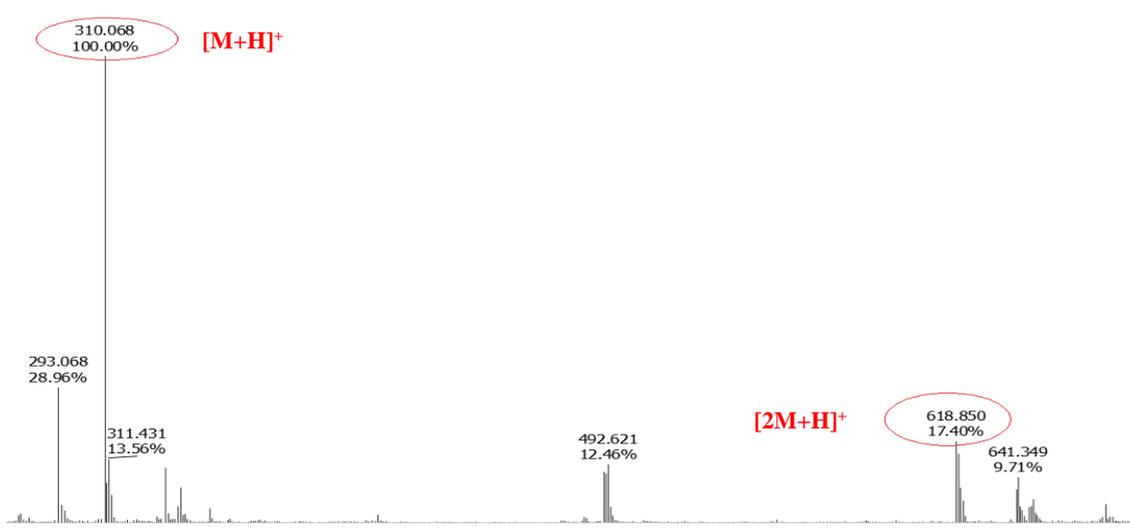
Figure 2.21 LCMS of crude dipeptide **194.**

In an attempt to limit the number of side products formed in the reaction, different reaction conditions were tested (**Table 2.3**). In all entries the starting peptide is peptide **187** and the solvent is MeCN and H$_2$O.

Entry	Nucleo-phile	Base	Temp.	Product(s)
1	PhSNa	DIPEA	rt	194 (*R,S and S,S*) [M+H]$^+$ 310.1 196 [M+H]$^+$ 563.4
2	PhSNa	–	rt	196 [M+H]$^+$ 563.4
3	PhSNa	K$_2$CO$_3$	rt	196 [M+H]$^+$ 563.4
5	PhOH	–	rt	No reaction
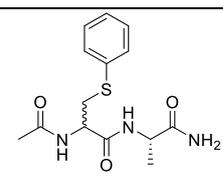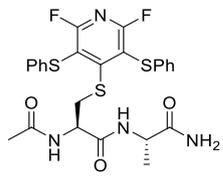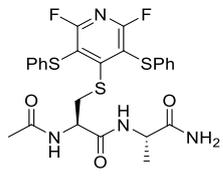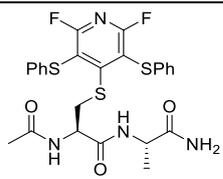
Table **2.3**: Reaction conditions of displacement reaction.

The only entry in **Table 2.3** in which the desired material was formed, is **Entry 1**. The reaction was purified prep TLC and compound **194** analysed. Shown below is an expansion of the ^{1}H NMR in the region of 3.00 ppm- 4.50 ppm.
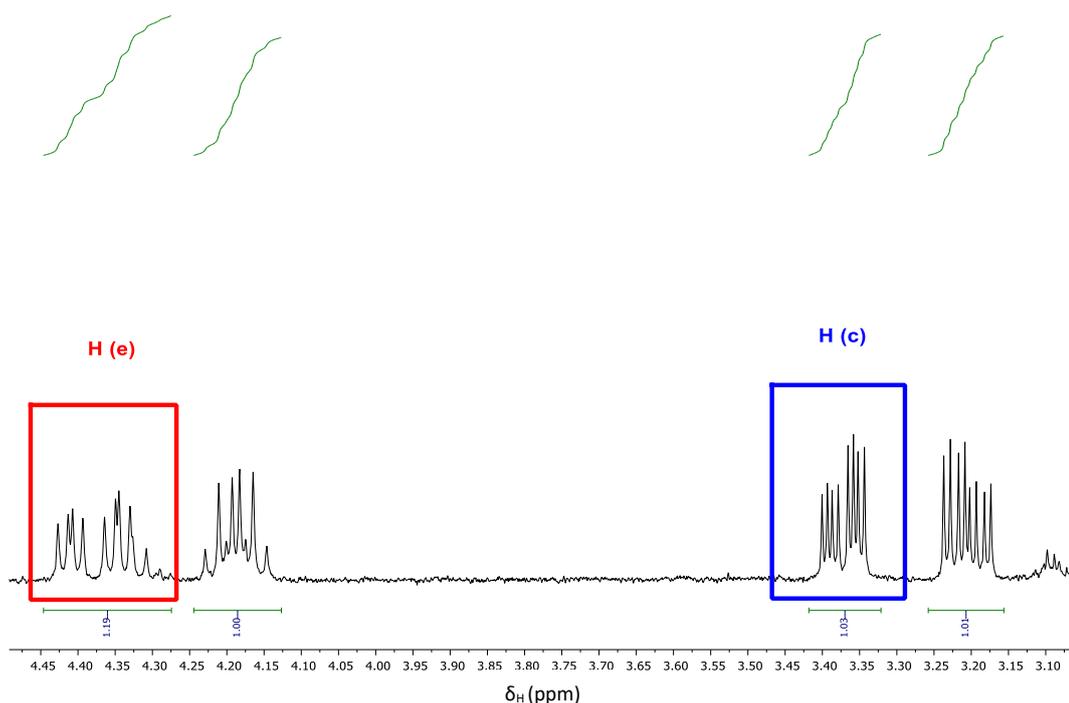
Figure 2.22: Expansion of region 3.00- 4.50 ppm in the ^{1}H NMR of compound **194**.

When this spectrum is compared to the spectrum of the starting material there are distinct differences. In the spectrum of compound **187** there is a multiplet between 4.51-4.56 ppm corresponding to proton H(e) whereas in the spectra of compound **194** there are 2 of these signals at 4.30- 4.42 ppm which correlates to proton H(e) of compound **194**. Another difference between the spectra of compounds **187** and **194** is the signals at 3.61- 3.65 ppm (compound **187**) and 3.34-3.40 ppm (compound **194**). In compound **187** this signal is a doublet of doublets whereas in the spectra of compound **194** there are 2 sets of doublet of doublets in this region which corresponds to proton H(c).

194
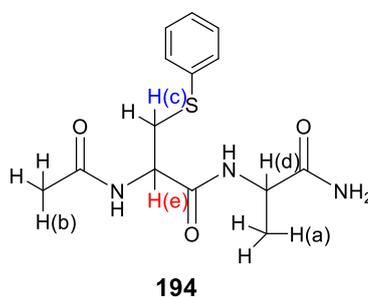
Figure 2.23: Labelled protons of compound **194**.

It is possible to hypothesise, due to the doubling of signals in **Figure 2.22** compared to the starting material of **Figure 2.8**, that there is a diastereomeric mixture of compound **194**. This suggests that of the 4 reaction pathways shown in **Scheme 2.20**, the reaction of fluoropyridine-tagged dipeptide and thiophenolate proceeds *via* an E_{CB} (**Scheme 2.20iii**) reaction pathway. E_{CB} reactions proceed when the α-carbon is acidic and leaving group is poor. E_{CB} mechanisms proceed *via* the extraction of an acidic α-carbon with base to form a carbanion followed by elimination of the leaving group and formation of Dha (**Scheme 2.22**). In the case of compound **187**, the Dha intermediate **198** undergoes nucleophilic addition of the thiophenolate, resulting in diasteromers of substituted analogue **194**.
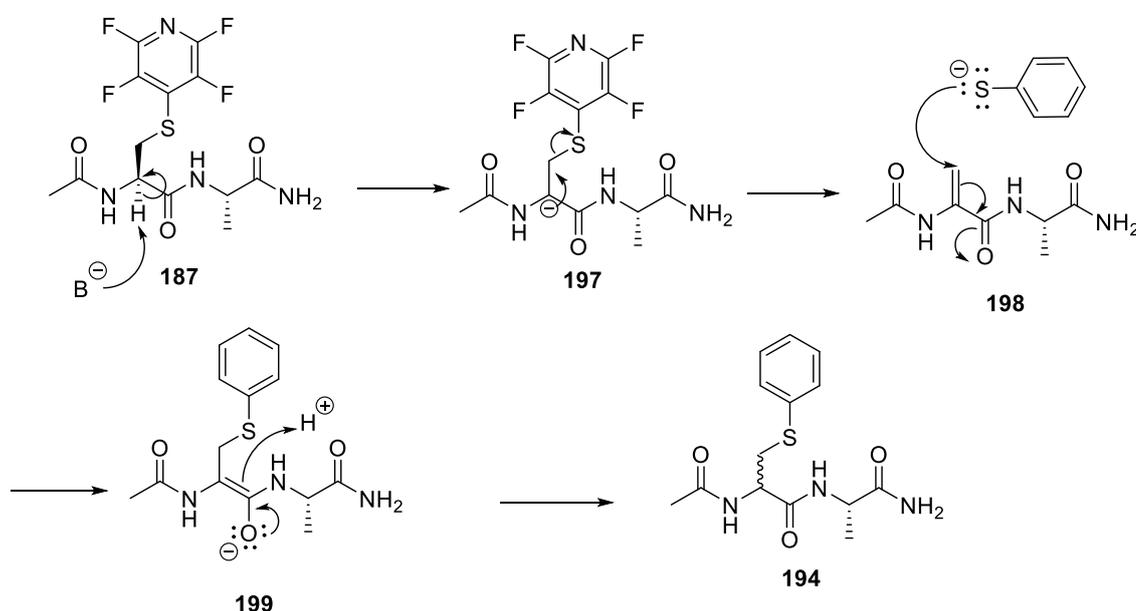
Scheme 2.22: E_{CB} mechanism of reaction of compound **187** to form the Dha intermediate **198** followed by addition of the thiophenolate nucleophile.

To confirm that the signals in the ^{1}H NMR are due to diastereoisomers, analytical HPLC was carried out and the spectra of compound **194** (**Figure 2.24ii**) was compared to starting material **187** (**Figure 2.24i**).
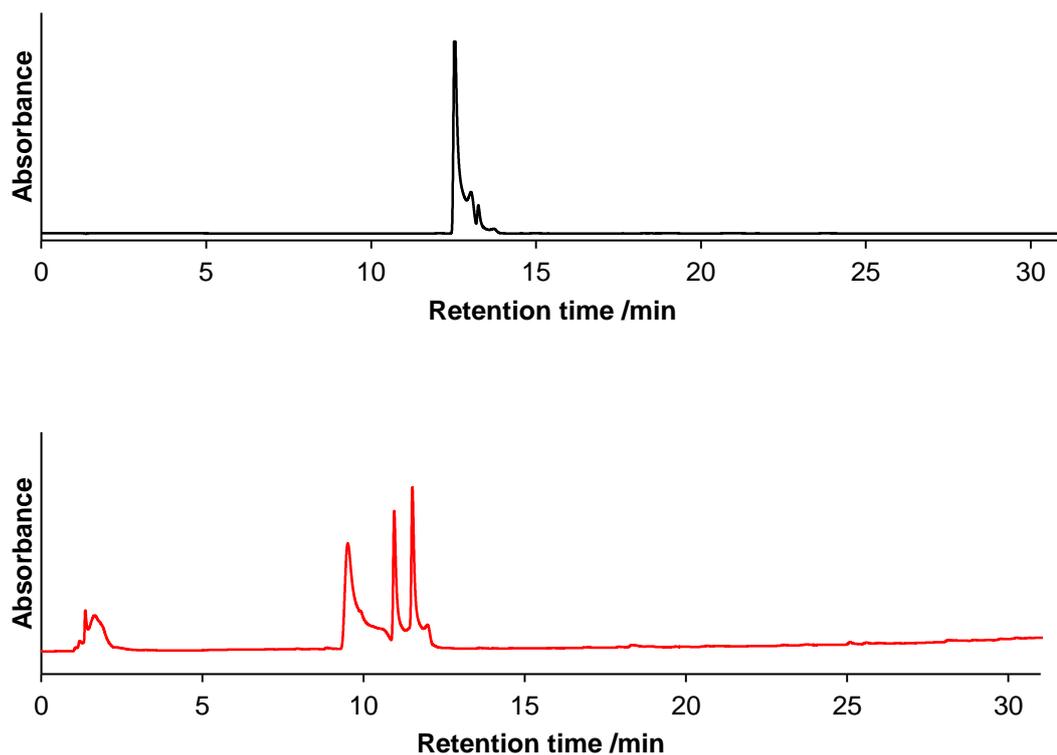
Figure 2.24: i) Analytical HPLC trace of starting material compound **187** ii) Analytical HPLC trace of compound **194**.

The peak at 9.5 min (**Figure 2.24ii**) is believed to correspond to an unknown side product with *m/z* 289.3. The peak 12.5 min in **Figure 2.24i** corresponds to compound **187** and one major peak is observed, suggesting that there is only one stereoisomer in the compound mixture. The peaks at 11.0 mins and 11.5 mins in **Figure 2.24ii** correspond to the displaced compound **194**. The presence of two signals that are separable by HPLC suggests that there are diastereoisomers within the sample.

The presence of multiple signals in the analytical HPLC and the doubling of the signals in the ^{1}H NMR suggests that the reaction between dipeptide **187** and thiophenolate proceeds *via* an E_{CB} reaction.

Upon inspection of the crude LCMS of **Scheme 2.21**, another product was also present in the reaction mixture with a mass of $[M+H]^+$ 563.2. It was assumed that this corresponds to compound **196**, which could arise due to nucleophilic attack on the fluoropyridine ring as shown in **Scheme 2.23**.
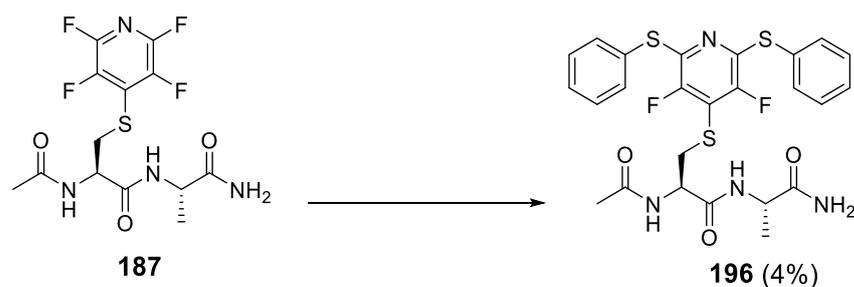
Scheme 2.23: Displacement reaction of tagged dipeptide **187** with thiophenolate. Reagents and conditions. Thiophenolate, DIPEA, MeCN/H$_2$O, rt, 4 h.

It was noted that there are a number of possible structures that could correspond to the LCMS signal of mass [M+H]$^+$ 563.2 and these are shown below (where R represents the dipeptide).
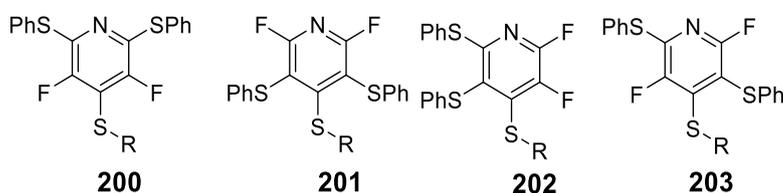
Figure 2.25: Possible product structures from S$_N$Ar reaction with compound **187**.

Since LCMS analysis could not give any elucidation of the structure of products, NMR spectroscopy was consulted. The crude material obtained from the reaction of compound **196** in **Scheme 2.23** was purified *via* a preparitive TLC plate. ^{1}H NMR and ^{19}F NMR analysis were carried out on purified **196** to deduce the substitution pattern of the thiophenolate on the fluoropyridine ring (**Figure 2.25**). Shown in **Figure 2.26** is an expansion of the ^{1}H NMR spectrum between 3.00 and 4.50 ppm.
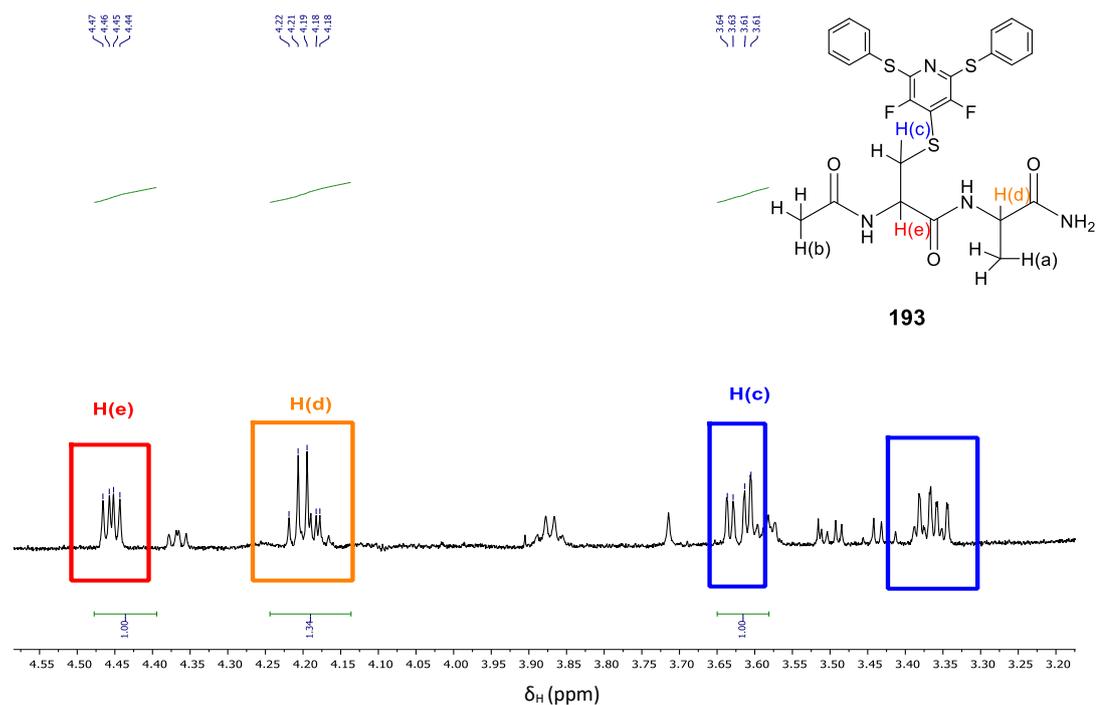
Figure 2.26: ^{1}H NMR of compound **196** region 3.00- 4.50ppm.

The ^{1}H NMR spectrum proved that the sample was not completely pure. The sample was submitted for high resolution ^{1}H NMR, but due to the dilute nature of the sample no improvement in resolution could be obtained. What can be observed from **Figure 2.26** are the signals at 3.34- 3.38 ppm and 3.61- 3.64 ppm which correlate to the protons labelled H(c) adjacent to the sulphur, 4.18- 4.21 ppm which correspond to the proton labelled H(d) and 4.44- 4.47 ppm which corresponds to the proton labelled H(e). The multiplet signal between 6.50- 7.90 ppm corresponds to the aromatic protons conjugated to the fluoropyridine ring, with an integration of 10H.

Analysis of the ^{1}H NMR spectrum could not conclusively determine the structure of compound **196**. ^{19}F NMR of compound **196** was carried out and the resulting spectrum obtained is shown in **Figure 2.27**. There are two distinct signals that relate to the structure of compound **196** with a doublet at -66.45 ppm and -119.42 ppm.
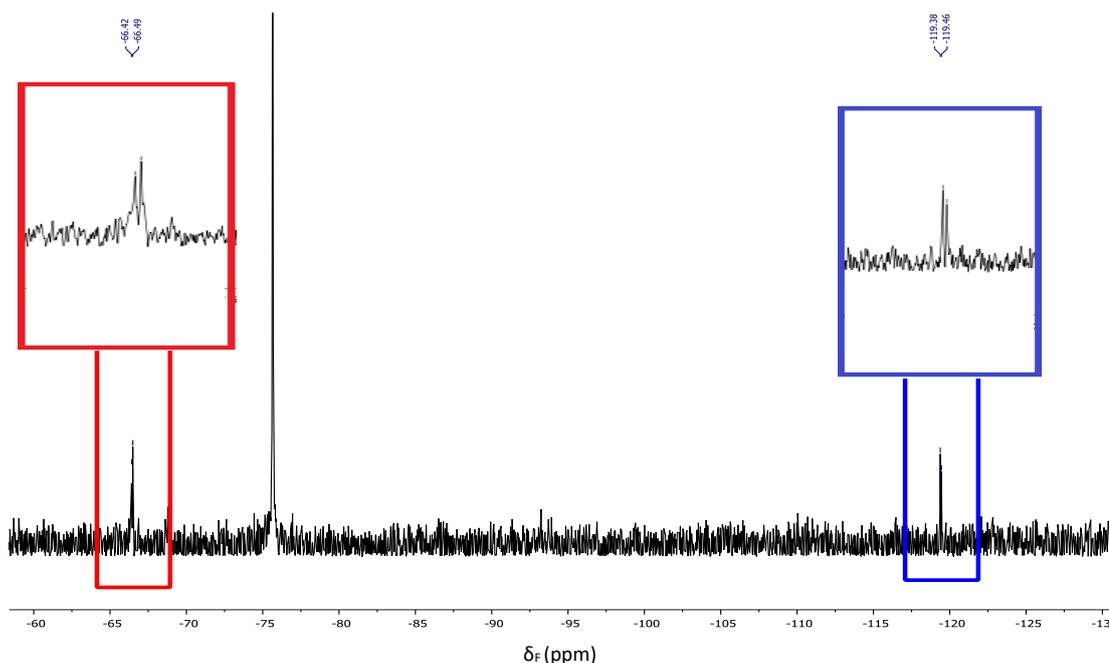
Figure 2.27: ^{19}F NMR of compound **196** with expansion of regions -65-67 ppm and -114-122 ppm.

Of the 4 configurations shown in **Figure 2.25**, there are 2 that could relate to the spectrum shown above. If compound **196** had the configuration shown in **200** or **201** then there would only be one signal in the ^{19}F NMR spectrum, therefore the structure of **196** must either be **202** or **203**. Density functional theory (DFT) calculations were carried out and the estimated chemical shifts shown in **Figure 2.28**.
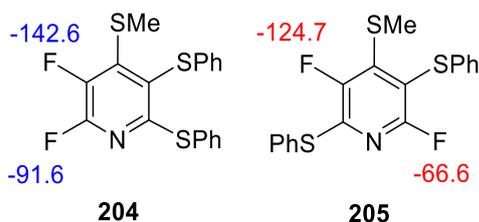
Figure 2.28: DFT calculations of the ^{19}F NMR shifts of compound **204** and **205.**

According to the chemical shifts shown in **Figure 2.28**, the predicted structure of the di-substituted dipeptide is **206**. The ^{19}F NMR (**Figure 2.27**) has a chemical shift of -66.5 ppm which matches the predicted chemical shift -66.6 pm of the *ortho*-F. The ^{19}F NMR in **Figure 2.27** also has a signal at -119.4 ppm which closely matches the *meta*-F signal at -124.7 ppm of the predicted structure. Therefore, the structure of the di-substituted dipeptide is as shown below (**Figure 2.29**).

206
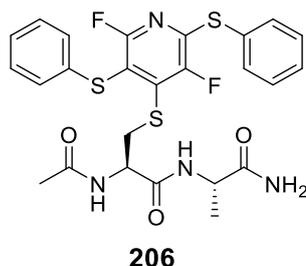
Figure 2.29: Structure of compound **206** according to ^{1}H NMR, ^{19}F NMR and molecular modelling calculations.

Compound **206** is the result of displacement of two fluorine atoms from fluoropyridine by thiophenolate as opposed to complete displacement of the fluoropyridine moiety. This side reaction product is an interesting development as it could provide possible scope for pentafluoropyridine acting as a scaffold which could allow for the conjugation of multiple nucleophiles to a peptide backbone.

With analysis of the cysteine analogue carried out, focus was shifted towards investigation of the displacement reaction between thiophenolate and the penicilamine (**190**) and homocysteine (**191**) PFP-tagged dipeptides.

Compound **190** was reacted with thiophenolate using the reaction conditions described in **Chapter 6**, **Section 6.6** (**Scheme 2.24**). LCMS analysis of the crude material confirmed the presence of the desired material [M+H]$^+$ 591.2.

190   **207** (33.5%)
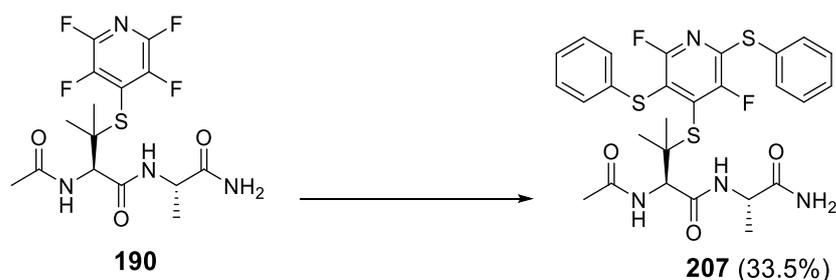
Scheme 2.24: Reaction of compound **190** with thiophenolate resulting in the di-substituted analogue **207**. Reagents and conditions thiophenolate, DIPEA, MeCN/H$_2$O, rt, 4 h.

The peptide was purified *via* preparative TLC plate and analysed using LCMS, ^{1}H NMR, ^{19}F NMR. **Figure 2.30** illustrates the ^{1}H NMR of compound **207** with no signals for CH$_2$ protons in the region that would be expected as well as the signals that correspond to H(d) and H(e) in the regions that are expected.
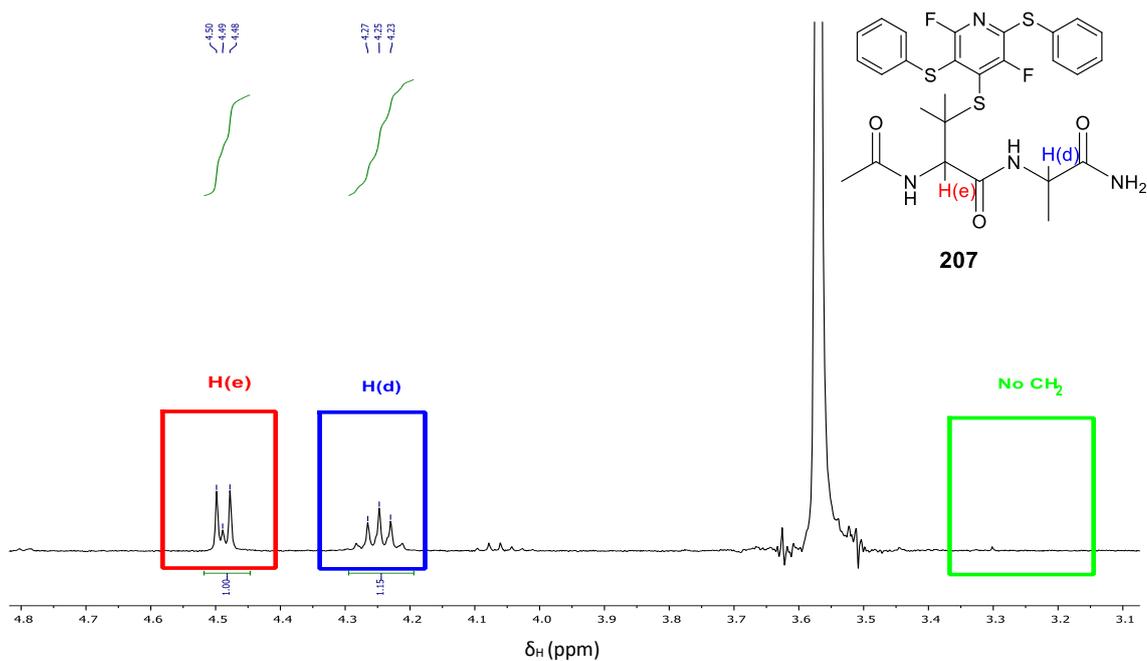
Figure 2.30: ^{1}H NMR spectrum of compound **207**- expansion of region 3.10- 4.80 ppm.

Also present in the ^{1}H NMR spectrum of compound **207** are 2 CH$_3$ signals corresponding to the methyl groups of the penicilamine residue and aromatic signals which integrate to 10H.

^{19}F NMR was also carried out on compound **207** and the resulting spectra shown below (**Figure 2.31**).
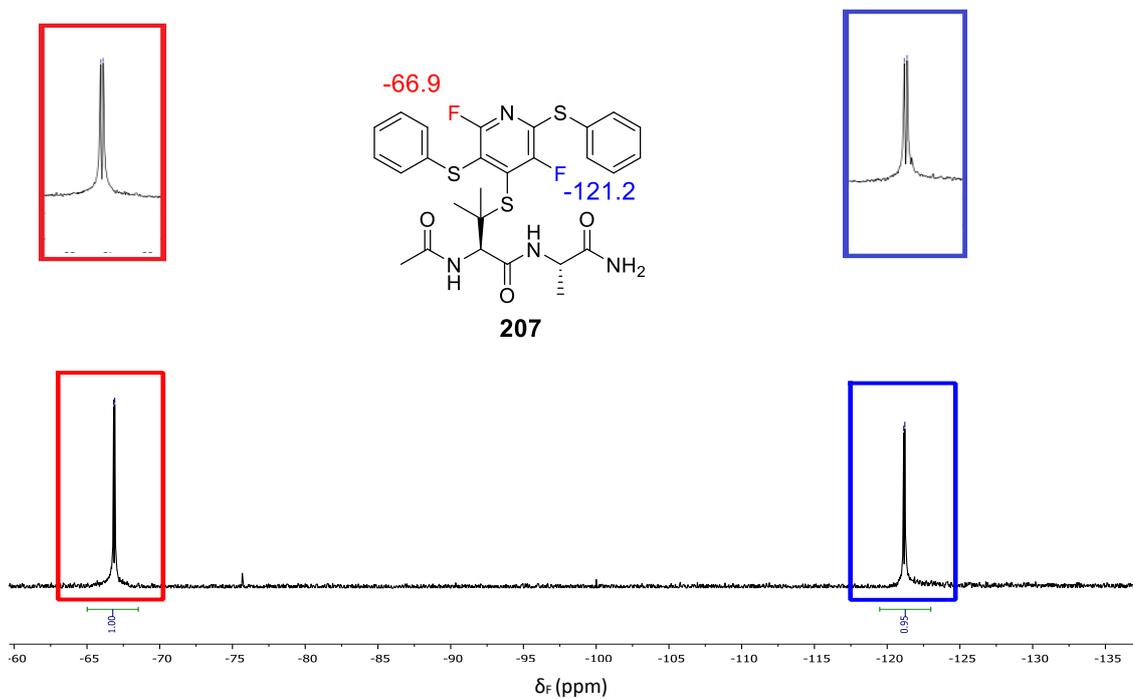
Figure 2.31: ^{19}F NMR spectrum of compound **207.**

The ^{19}F NMR spectrum shows 2 distinct peaks with integration of 1F at -66.9 ppm and -121.2 ppm. These signals match the calculated chemical shifts from the molecular modelling calculations carried out previously.

The homocysteine analogue (**191**) was also reacted with thiophenolate, purified *via* prep TLC and the resulting fractions analysed using LCMS, ^{1}H NMR and ^{19}F NMR. The LCMS contained the expected mass of [M+H]$^+$ 577.2, relating to the di-substituted product (**208**).
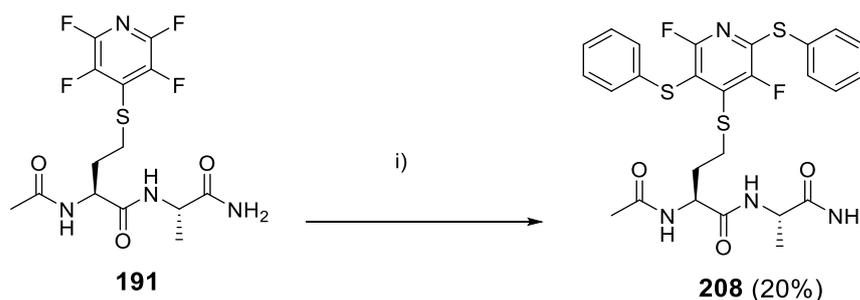
Scheme 2.25: Reaction of compound **191** with thiophenolate resulting in the di-substituted analogue **208.** Reagents and conditions i) Thiophenolate, DIPEA, MeCN/H$_2$O, rt, 4 h.

Figure 2.32 illustrates the region of the ^{1}H NMR between 3.00 and 4.40 ppm of compound **208**. In the spectra, the protons that relate to H(e) and H(d) have merged into one multiplet. An additional signal was observed at 3.19 ppm that relates to H(f) of the CH$_2$ adjacent to the S of the homocysteine analogue. The proton signal for the other CH$_2$ of the homocysteine is hidden under the MeCN solvent peak and is therefore not evident below.
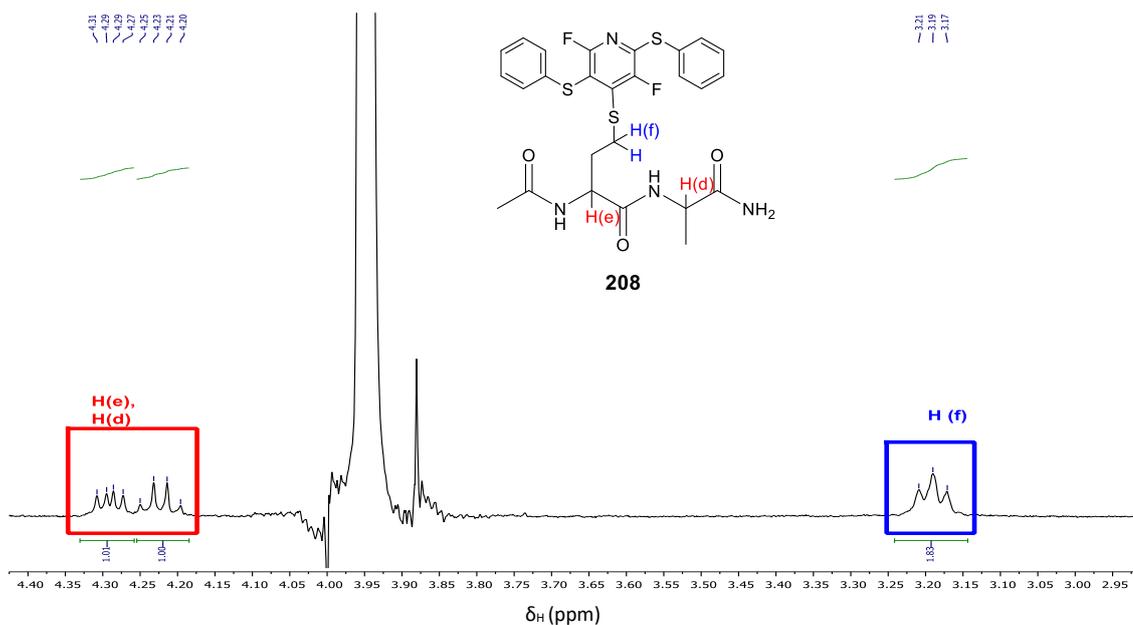
Figure 2.32: ^{1}H NMR spectrum region 2.90- 4.40 ppm of compound **208.**

There are also signals in the aromatic region which integrate to 10H which have similar splitting patterns to both the cysteine and penicilamine analogues, further suggesting that the substitution pattern on the fluoropyridine ring is as shown in compound **206, 207** and **208**.

^{19}F NMR was also carried out on compound **208** and the resulting spectra shown below (**Figure 2.33**).
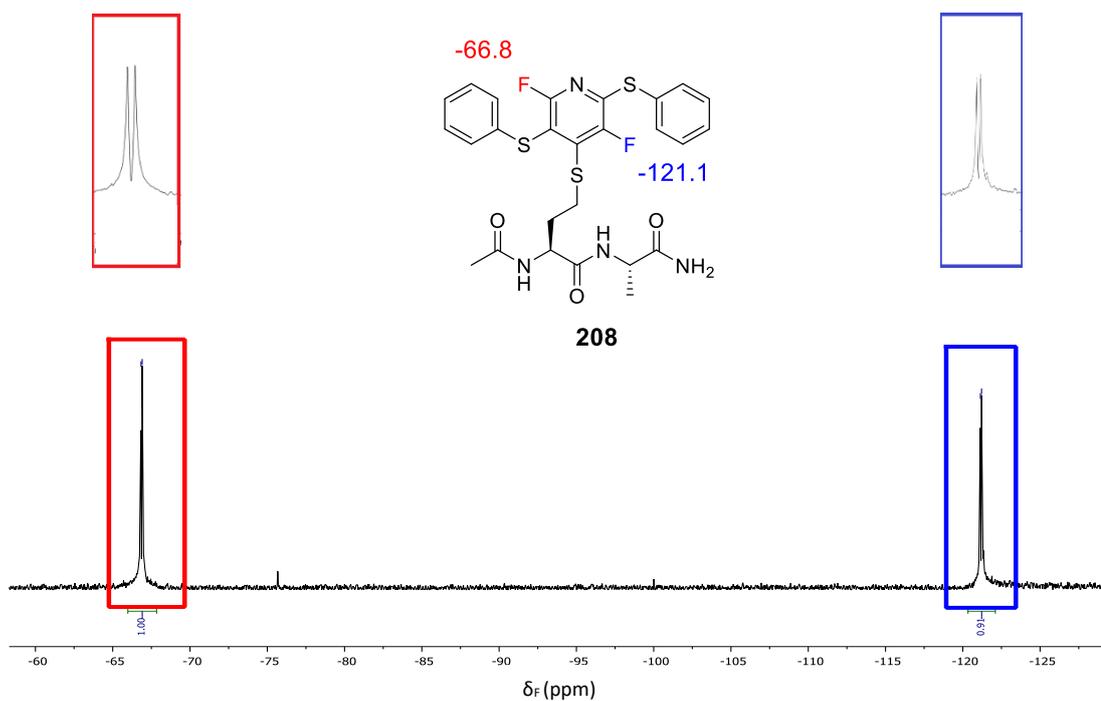
Figure 2.33: ^{19}F NMR of compound **208.**

The ^{19}F NMR shows 2 distinct peaks with integration of 1F at -66.9 and -121.2 ppm. The chemical shifts shown in **Figure 2.33** match the predicted chemical shifts of the molecular modelling calculation and therefore compound **208** has the structure shown in **Scheme 2.23**.

2.5 Conclusions

Efforts into the synthesis and mechanism investigation of sulphur-containing dipeptides and pentafluoropyridine have been carried out. A series of dipeptides were successfully synthesised and reacted with pentafluopyridine resulting in modified peptides (**187, 190** and **191**) which were fully analysed and characterised. Upon further reaction of peptides (**187, 190** and **191**), a number of products were observed. It was expected upon reaction of the fluopyridine-tagged peptide (**187**) and thiophenolate nucleophile that the peptide (**194**) would be observed. This product was indeed observed and ^{1}H NMR was carried out to determine the reaction mechanism. It was determined that due to the doubling of signals in the ^{1}H NMR and the analytical HPLC that the reaction had progressed *via* an E$_{CB}$ reaction pathway and resulted in the formation of diastereomers.
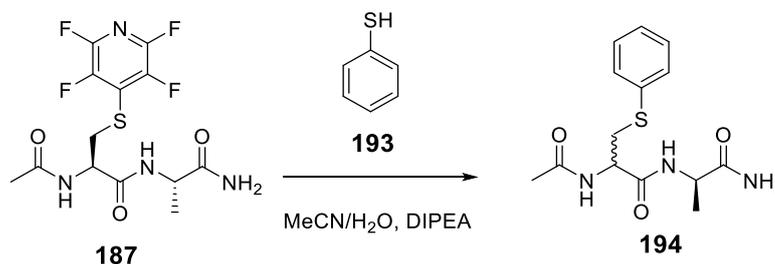
Scheme 2.26: Reaction of peptide **187** with thiophenolate (**193**) *via* an E$_{CB}$ pathway.

An unexpected peptide (**206**) was observed upon reaction of peptide (**187**) with the thiophenolate nucleophile. The data obtained suggests an aromatic substitution reaction on the fluoropyridine ring had occurred which was confirmed *via* interpretation of the ^{19}F and ^{1}H NMR spectrum. It was also confirmed *via* ^{19}F NMR spectroscopy that further substitution occurred at the *ortho* and *meta* positions which was unexpected (**Scheme 2.27**).
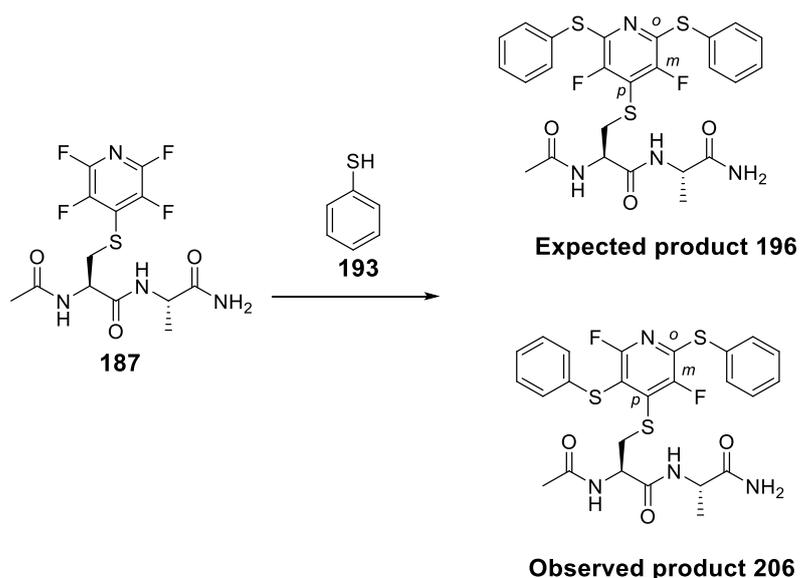
Scheme 2.27: Reaction of peptide **187** with thiophenolate and the expected **196** and observed di-substituted products **206**.

The aromatic substitution of Pen (**190**) and hCys (**191**) dipeptides with thiophenolate were also investigated and it was determined, from the ^{19}F NMR and LCMS data, that the primary product in those reactions were the *o-m*-disubstituted products (**Figure 2.34**).
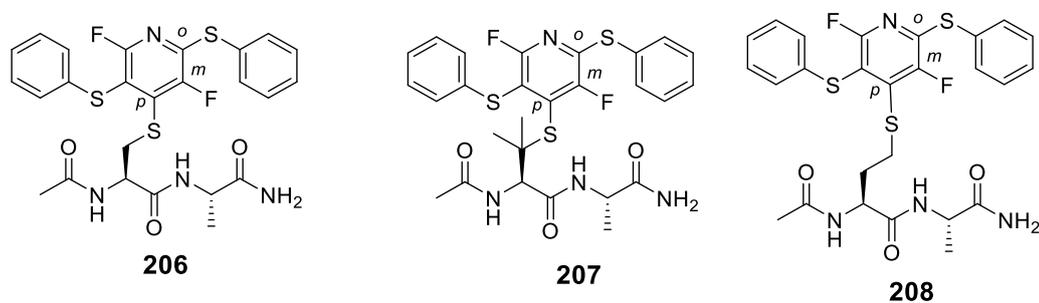
Figure 2.34: *Ortho-meta*-disubstituted i) Cys dipeptide (**206**), Pen dipeptide (**207**) and hCys dipeptide (**208**).

References

[1] W. K. Hagmann, *J. Fluor. Chem*., 2008, **51**, 4359-4369.

[2] D. B. Longley, D. P. Harkin, P. G. Johnston, *Nature*, 2003, **3**, 330-338.

[3] P. Shah, A. D. Westwell, *J. Enzym. Inhib. Med. Ch*., 2007, **22**, 527-540.

[4] J. Wang, M. S. Rosello, J. L. Acena, C. Pozo, A. E. Sorochinsky, S. Fustero, V. A. Soloshonok, H. Lui, *Chem. Rev*., 2014, **114**, 2432-2506.

[5] H. Amii and K. Uneyama, *J. Am. Chem. Soc.*, 2009, **109**, 2119-2183.

[6] M. W. Cartwright, L. Convery, T. Kraynck, G. Sandford, D. S. Yufit, J. A. K. Howard, J. A. Christopher and D. D. Miller, *Tetrahedron,* 2010, **66**, 519-529.

[7] M. Schlosser, T. Rausis, C. Bobbio, *Org. Lett*., 2004, **7**, 127-129.

[8] K. Beyki, R. Haydari, M. T. Maghsoodluo, *SpringerPlus*, 2015, **4**, 757-762.

[9] R. D. Chambers, *Fluorine in Organic Chemistry*, Blackwell Publishing Ltd, Oxford, 1st edn, 2004, ch. 9, pp. 307-315.

[10] C. B. Murray, G. Sandford, S. R. Korn, D. S. Yufit, J. A. K. Howard, *J. Fluor. Chem*., 2005, **126**, 571-576.

[11] E. Lork, R. Mews, M. M. Shakirov, P. G. Watson, A. V. Zibarev, *J. Fluor. Chem*., 2002, **115**, 165-168.

[12] G. A. Artamkina, E. A. Tarasenko, N. V. Lukashev, J. P. Beletskaya, *Tettrahedron Lett*., 1998, **39,** 901-904.

[13] Y. Wang, S. R. Parkin, M. D. Watson, *Org. Lett*., 2008, **10**, 4421-4424.

[14] C. J. Gilmore, D. D. MacNicol, A. Murphy and M. A. Russell, *Tetrahedron Lett.*, 1984, **25**, 4303-4306.

[15] H. P. Ng, B. O. Buckman, K. A. Eagen, W. J. Guilford, M. J. Kochanny, R. Mohan, K. J. Shaw, S. C. Wu, D. Lentz, A. Liang, L. Trinh, E. Ho, D. Smith, B. Subramanyam, R. Vergona, J. Walters, K. A. White, M. E. Sullivan, M. M. Morrisey, G. B. Phillips, *Bioorgan. Med. Chem*., 2001, **10**, 657-666.

[16] A. S. Hudson, A. Hoose, C. R. Coxon, G. Sandford and S. L. Cobb, *Tetrahedron Lett.*, 2013, **54**, 4865-4867.

[17] J. M. Chalker, S. B. Gunnoo, O. Boutureira, S. C. Gerstberger, M. Fernandez-Gonzalez, G. J. L. Bernardes, L. Griffin, H. Hailu, C. J. Scholfield, B.G. Davis, *Chem. Sci.* 2011, **2**, 1666-1676.

[18] J. M. Chalker, G. J. L. Bernades and B. G. Davis, *Acc. Chem. Res.*, 2011, **44**, 730-741.

[19] A. M Webster, C. R. Coxon, A. M. Kenright, S. L. Cobb, *Tetrahedron,* **70**, 4661-4667.

Chapter 3: Synthesis, Tagging and Stability Studies of PFP-modified Oxytocin and Vasopressin

3.1 Oxytocin (OT) and Vasopressin (VP)

Oxytocin (OT) (**209**) and Vasopressin (VP) (**210**) are both cyclic peptides consisting of 9 amino acid residues and a disulphide bridge between cysteine residues. OT (**209**) has the sequence **C**YIQN**C**PLG and VP (**210**) has the sequence **C**YFQN**C**PRG and both have *C*-terminal amides (**Figure 3.1**).

209   **210**
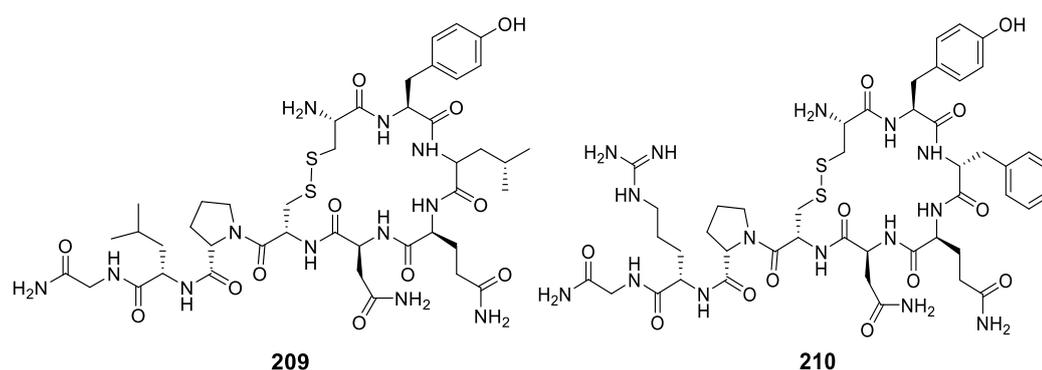
Figure 3.1: Structure of OT (**209**) and VP (**210**).

OT (**209**) was discovered in 1906 by Henry Dale[1] and the crystal structure was determined in 1952. OT (**209**) is a neuropeptide which is released from the posterior pituitary into the bloodstream and has been investigated in a number of different roles. OT is released during childbirth and into the bloodstream after childbirth and plays a role in social bonding between mother and child.[3] OT is used during childbirth to promote uterine contractions which increases the speed of labour and also reduces bleeding after childbirth.[4]

OT has also been shown to have an effect on depressive behaviour.[5] Many animal studies[4,5] have found that lower levels of OT in animal examples can lead to depressive behaviour, although the studies of this behaviour in humans has been largely disputed and there are multiple studies to suggest both arguments.[6,7] It is believed that any perceived depressive behaviour in humans as a result of lower OT levels may be related to the impaired social bonding and negative early life experiences.[5]

VP (**210**) was first synthesised by Vincent du Vigneaud, who was awarded the Nobel Prize in Chemistry in 1955 for his work in synthesising the first polypeptide hormone.[1] VP (**210**) and OT have different target sites and it is believed the difference in polarity between OT and VP is responsible for the difference in response biological response.[2] VP (**210**) is an antidiuretic hormone (ADH) and is used as a treatment against abdominal bleeding and diabetes as well as having a number of off-target effects which can be beneficial therapeutically. VP targets a number of GPCRs accounting for the range of effects it has on the body, which include mediation of blood vessels and water absorbance.[11]

3.1.1 Oxytocin Modification Studies

OT (**209**) has a relatively short half-life ($t_{1/2}$= 2-5 min)[8] *in vivo* and is readily degraded *via* proteolysis. Collins[9] and co-workers synthesised a peptide-polymer analogue with a maleimide linker *via* the disulphide bond (**Scheme 3.1**) in an effort to improve the stability.
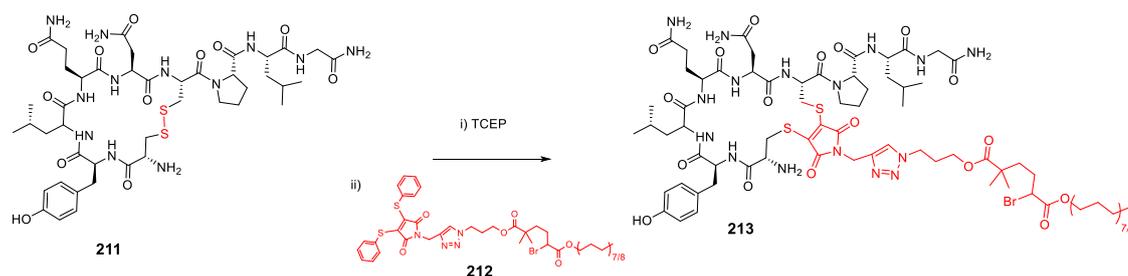
Scheme 3.1: Synthesis of OT peptide-polymer (**213**) *via* a maleimide linker. i) Reduction of disulphide bond using TCEP ii) Conjugation using excess dithiophenol maleimide (**212**).

The stabilities of the polymer-peptide analogues in water were compared with the stability of native OT in water at pH 7.6. After 28 days, only 2.5% of the native OT was present whereas for the polymer-peptide analogues OT-(mPEGA)$_{20}$ and OT-(mPEGA)$_{100}$ the amount of peptide left after 28 days was 94% and 87% respectively (**Figure 3.2**).
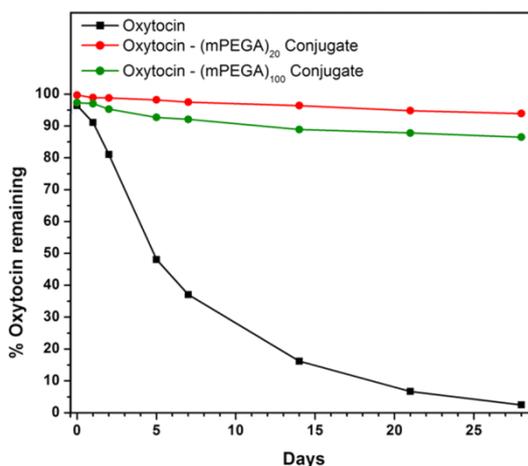
Figure 3.2: Stability test of native OT, OT-(mPEGA)$_{20}$, OT-(mPEGA)$_{100.}$[9]

Carbetocin (**214**) is an OT analogue that was designed to have enhanced stability. It is structurally very similar to OT, but the disulfide bridge has been replaced with a thioether linkage (**Figure 3.3** shown in red) and the Tyr residue contains a methyl ether rather than a free phenol (**Figure 3.3** in blue).

214
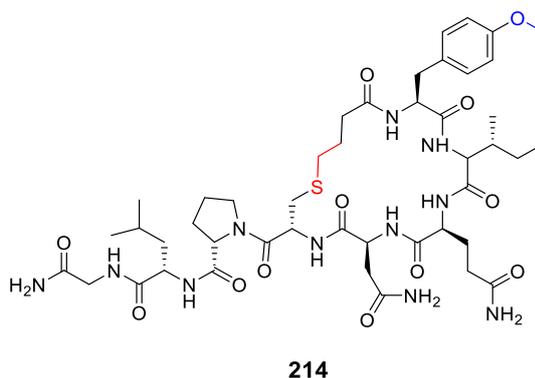
Figure 3.3: Chemical structure of Carbetocin (**214**).

The *in vivo* stability of Carbetocin (**214**) was shown to be higher than OT with it having a half-life *in vivo* of 85-100 min compared to approximately 3 min for OT.[8]

Further work to probe the stability of OT (**209**) was carried out by Barth and co-workers[10] who investigated the degradation of Carbetocin (**214**) using chymotrypsin and found that after 30 min it had been degraded to deglycineamide Carbetocin (**216**) (**Scheme 3.2ii**). Chymotrypsin primarily degrades amide bonds where the *N*-terminal amino acid is an aromatic residue, suggesting that the enzymatic degradation of an OT analogue would occur at the Tyr residue (**Scheme 3.2i**). Barth and co-workers were able to illustrate that incorporation of the methyl group onto the Tyr residue improved

stability as there was no degradation of the peptide at the Tyr-Leu amide bond and the only degradation product observed was the Gly deletion (**Scheme 3.2ii**).[10]
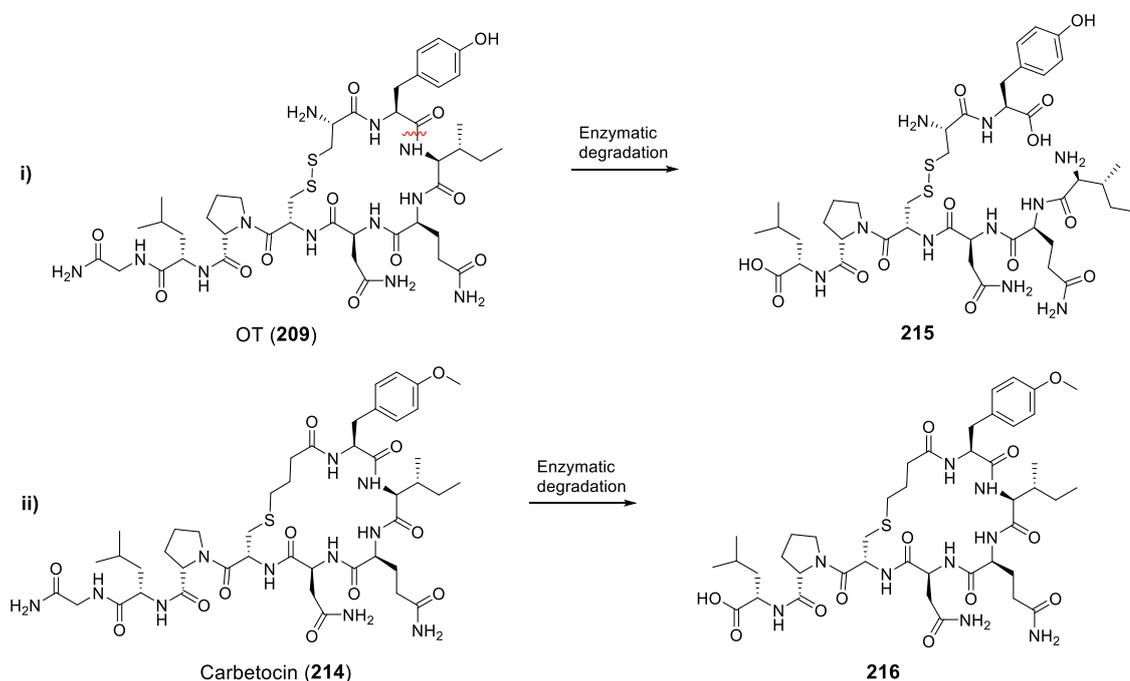
Scheme 3.2: Enzymatic degradation using chymotrypsin of i) OT (**209**) to give Tyr-Leu cleavage product (**215**) and ii) Carbetocin (**214**) to give the Gly deletion product (**216**).

3.1.2 Vasopressin Modification Studies

Lammek and co-workers investigated the effects of amino acid substitution on the potency and selectivity of VP.[11] One of the most interesting analogues of this study was **217** which not only contains a 1-aminocyclohexane-1-carboxylic acid (Acc) substitution at position 2 but also deamination at position 1, Val substitution at position 4 and inversion of stereochemistry at position 8 (**Figure 3.4**). This pattern of substitution was found to be the optimal analogue (**217**) which was a potent V_2 agonist.[12] Further studies into the substitution at position 2 of VP concentrated on the incorporation of 2-aminoindane-2-carboxylic acid (Aic) at position 2 (**218**) (**Figure 3.4**).[11] Incorporating these building blocks into VP resulted in analogues that had potent biological activity, believed to be due to the conformational restriction imposed by the new building blocks.
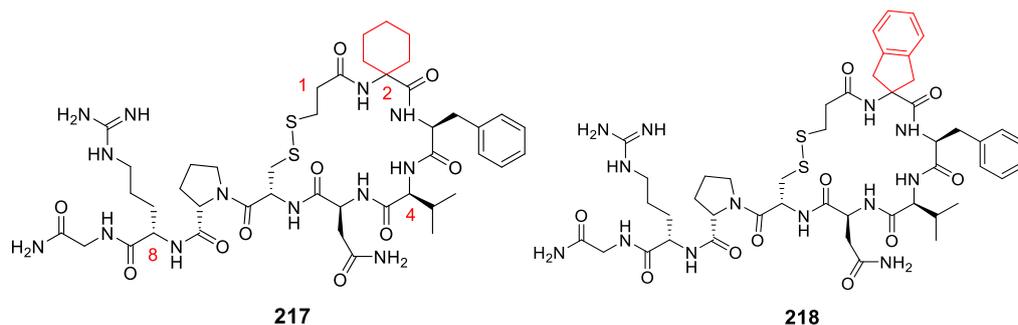
Figure 3.4: Structures of 1-aminocyclohexane-1-carboxylic acid (Acc) (**217**)[12] 2- aminoindane-2-carboxylic acid (Aic) (**218**).[11]

Desmopressin (**219**) is a VP analogue which has been approved for the treatment of diabetes inspidus, haemophilia and other bleeding disorders. Desmopressin (**219**) is structurally similar to native VP apart from the incorporation of D-Arg at position 8 and deamination of the first Cys residue (**Figure 3.5**).[13]
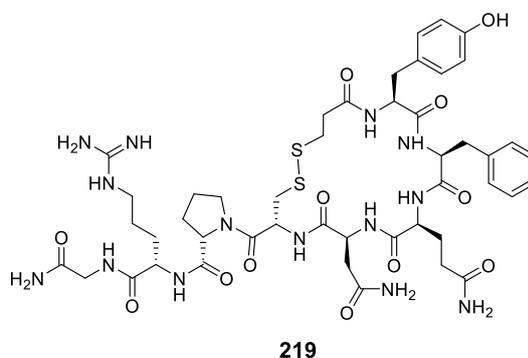
Figure 3.5: Chemical structure of Desmopressin (**219**) a VP analogue.

Deamination at position 1 and the inversion of stereochemistry of Arg-8 increases antidiuretic activity. Desmopressin (**219**) is shown to have antidiuretic activity close to 2000 units per milligram compared to 400 units per milligram for VP (**210**) which suggests Desmopresssin is both more biologically active and has higher specificity than the native VP.[14]

3.1.3 Chapter Aims

The aim of this chapter was to synthesise a series of OT and VP analogues tagged with pentafluoropyridine (PFP), in an effort to improve their stability towards enzymatic degradation. The peptides will be synthesised using Fmoc SPPS and tagging of the peptides with PFP carried out using both on-resin and solution phase techniques. By utilising the orthogonal protecting group strategies, the nucleophilic side chains (e.g. Tyr, NH_2 of *N*-terminus) can be selectively reacted with PFP.

Once the VP and OT analogues have been synthesised, the stability of these peptides to enzymatic degradation will be tested using chymotrypsin and HPLC analysis. Peptides **209** and **210** in **Figure 3.6** represent the structures of native OT and native VP, peptides **220** and **222** represent the singular tagged OT and VP analogues and peptides **221** and **223** represent the double-tagged OT and VP analogues which will be synthesised and form part of this enzymatic stability study.

OT (**209**)

PFP-OT (**220**)

2PFP-OT (**221**)
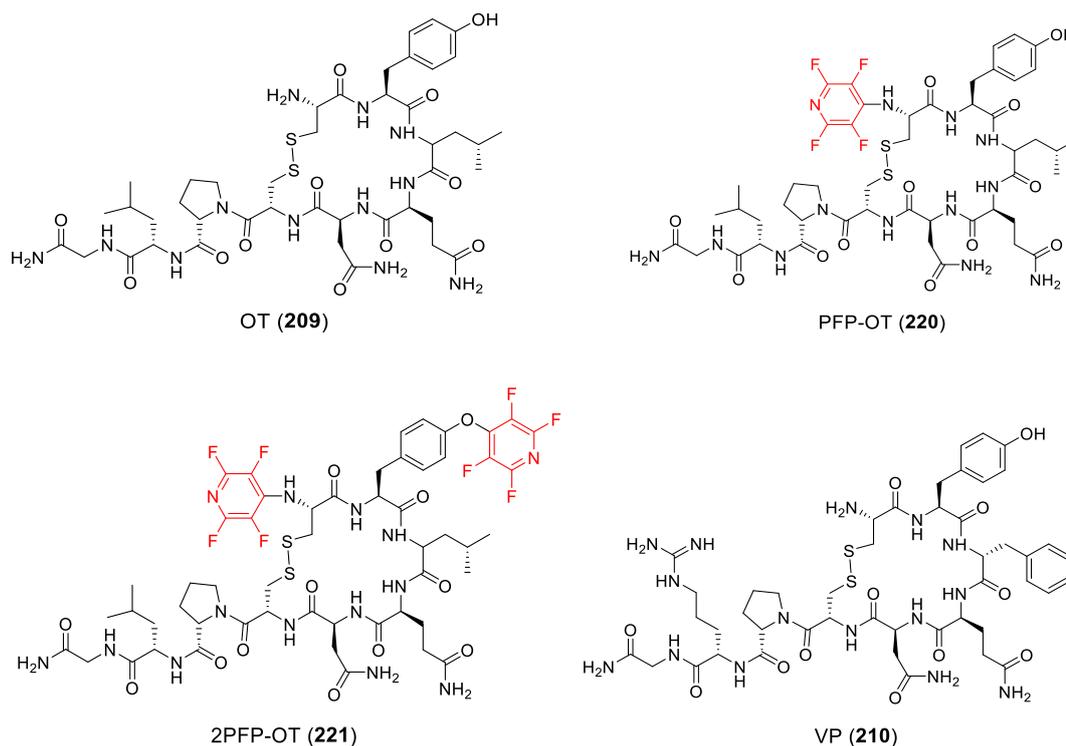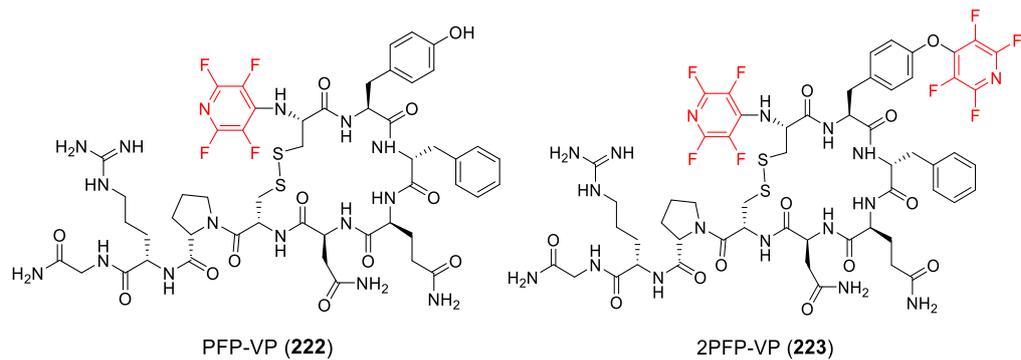
VP (**210**)

Figure 3.6: Chemical structure of native peptides OT (**209**) and VP (**210**), singular PFP-tagged peptides (**220** and **222**) and doubly PFP-tagged peptides (**221** and **223**).

3.2 Synthesis of OT and VP and PFP-tagged Analogues using Fmoc SPPS

Native OT (**209**) was synthesised *via* manual SPPS using RINK Amide resin. Amino acids were protected using standard acid-labile protecting groups with the exception of Cys which was protected using the acetamidomethyl (Acm) group (**Scheme 3.3**). This protecting group was used in place of the normal trityl (Trt) group for the Cys thiol as it would allow for the selective PFP tagging of the peptide.
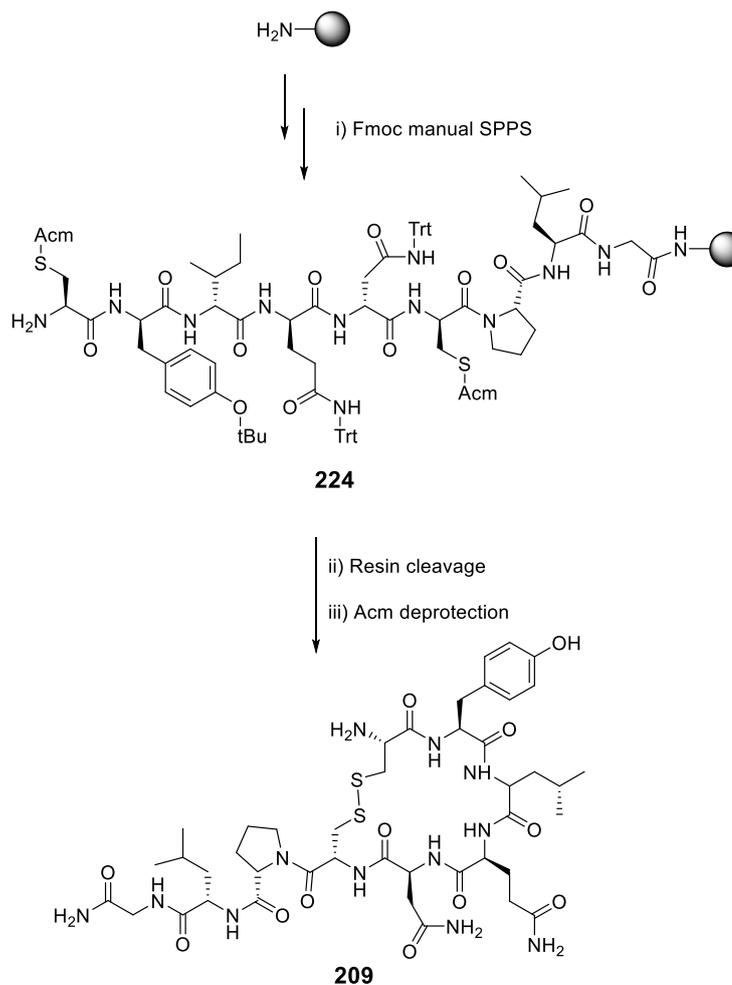
Scheme 3.3: Manual SPPS of OT (**209**) using RINK amide resin. Reagents and conditions: i) Manual Fmoc SPPS- Fmoc-Gly-OH, Fmoc-Leu-OH, Fmoc-Pro-OH, Fmoc-Cys(Acm)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Gln(Trt)-OH, Fmoc-Ile-OH, Fmoc-Tyr(tBu)-OH, PyBOP, DIPEA and DMF, rt, (2 × 1 h) couplings. Fmoc deporotection- 20% piperidine/DMF, rt, 2 × 30 min ii) Cleavage from resin- 95% TFA: 2.5% TIPS: 2.5% H_2O, rt, 4 h iii) Acm cleavage- I_2, AcOH (40%)/DMF, rt, 1.5 h.

Upon completion of the synthesis and deprotection of the *N*-terminal Fmoc group, a test cleavage of the crude linear peptide (**224**) from the resin was carried out and analysed *via* LCMS (**Figure 3.7**). The linear peptide **224** was present in the test cleave as indicated by 1152.5 [M+H]$^+$ and 576.4 [M+2H]$^{2+}$.
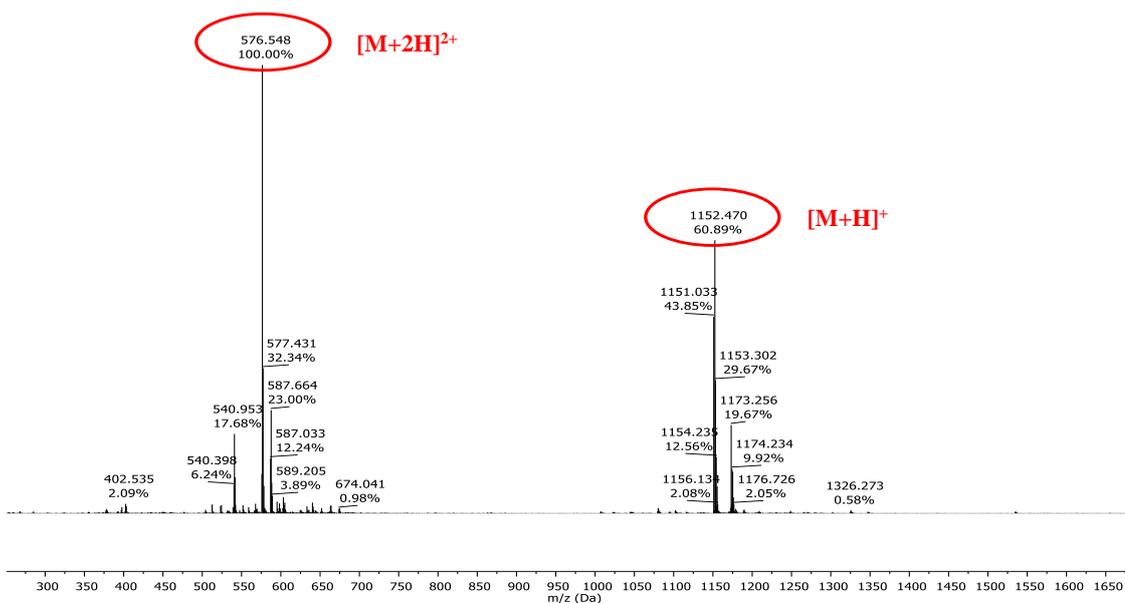
Figure 3.7: MS of linear peptide cleaved peptide **224** with desired mass 1152.5 [M+H]$^+$ and 576.5 [M+2H]$^{2+}$.

To access native OT (**209**), the Acm protected linear peptide (**224**) underwent an Acm deprotection and disulphide bond formation using I$_2$ (3 eq), 40% AcOH/DMF solution at rt. After 1.5 h the reaction was quenched using a 1 M solution of ascorbic acid and diluted using H$_2$O/MeCN until the final volume of AcOH was less than 10%. The crude peptide was analysed *via* LCMS (**Figure 3.8**).
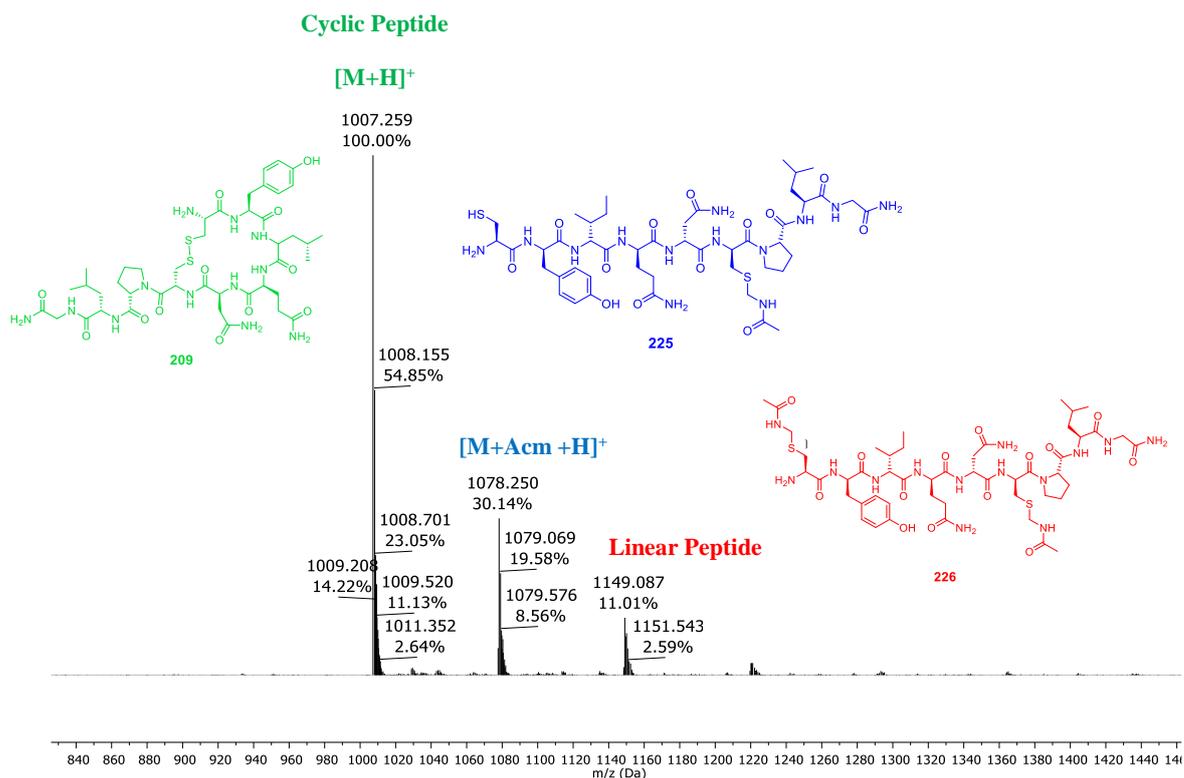
Figure 3.8: MS of native OT (**209**) with desired mass [M+H]$^+$ 1007.2, [M+Acm+H]$^+$ 1078.3 (**225**) [M+2Acm+H]$^+$ 1149.0 (**226**).

As shown in the MS above, the desired material (**209**) was present along with two side products with 1078.3 [M+Acm+H]$^+$ (**225**) and 1149.0 [M+2Acm+H] (**226**) which relate to incomplete Acm deprotection of the Cys residues. The Acm deprotection was carried out using alternative reaction conditions to drive the reaction to completion. The Acm protected linear peptide (**224**) was dissolved in 1 ml of AcOH (40%)/DMF and the I$_2$ was added. In this reaction 25 eq of I$_2$ were used instead of the 3 eq previously used while the percentage of AcOH remained the same. The reaction mixture was transferred to an Eppendorff tube and left at rt, under agitation for 1.5 h. The reaction mixture was quenched using 1 M ascorbic acid and diluted using H$_2$O/MeCN until the final volume of AcOH was less than 10%.

The OT peptide (**209**) was purified using reverse phase HPLC (RP-HPLC) and analysed using LCMS. As shown in **Figure 3.9**, the full deprotection of the Acm group and cyclisation has been achieved as illustrated by the LCMS peak of 1008.2 [M+H]$^+$.
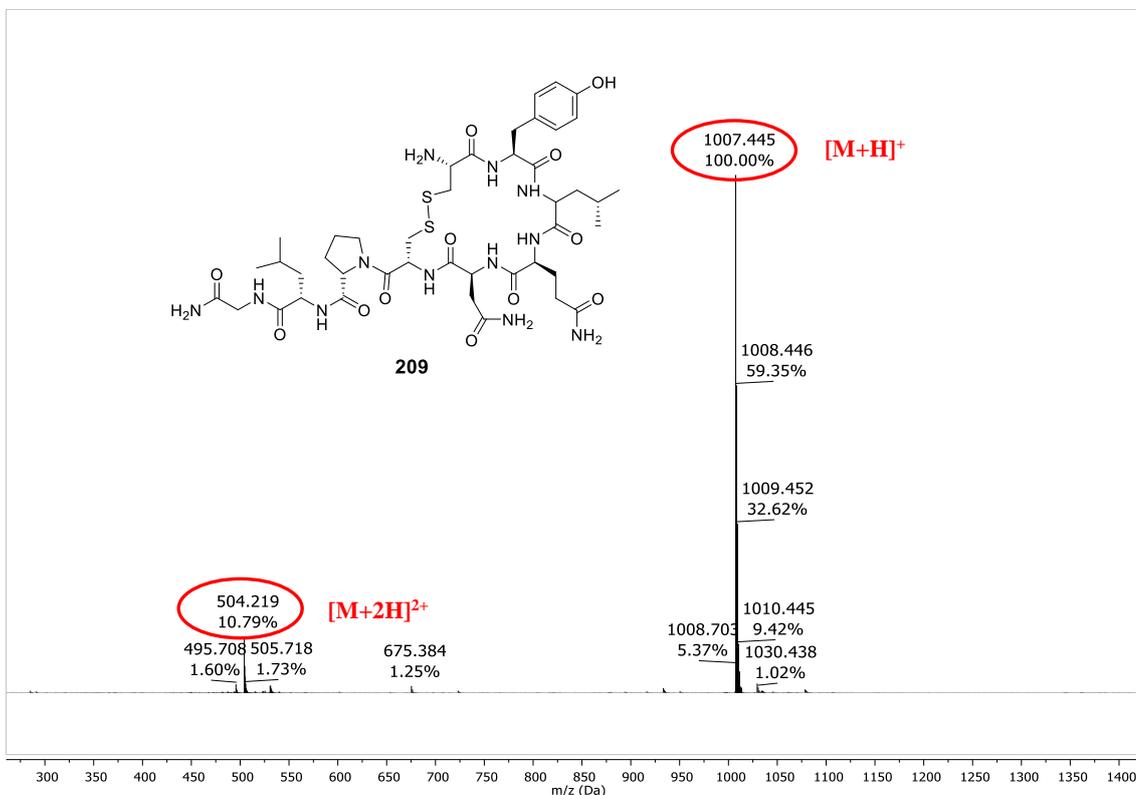
Figure 3.9: MS analysis of the Acm deprotection/cyclisation reaction **209** using 25 eq of I$_2$, 40% AcOH/DMF for 1.5 h [M+H]$^+$1007.4.

Native VP (**210**) was synthesised *via* manual Fmoc SPPS using RINK Amide resin. The side chains of the Fmoc amino acids were protected using standard acid-labile protecting groups with the exception of Cys, which was protected using Acm (**Scheme 3.4**).

227

210
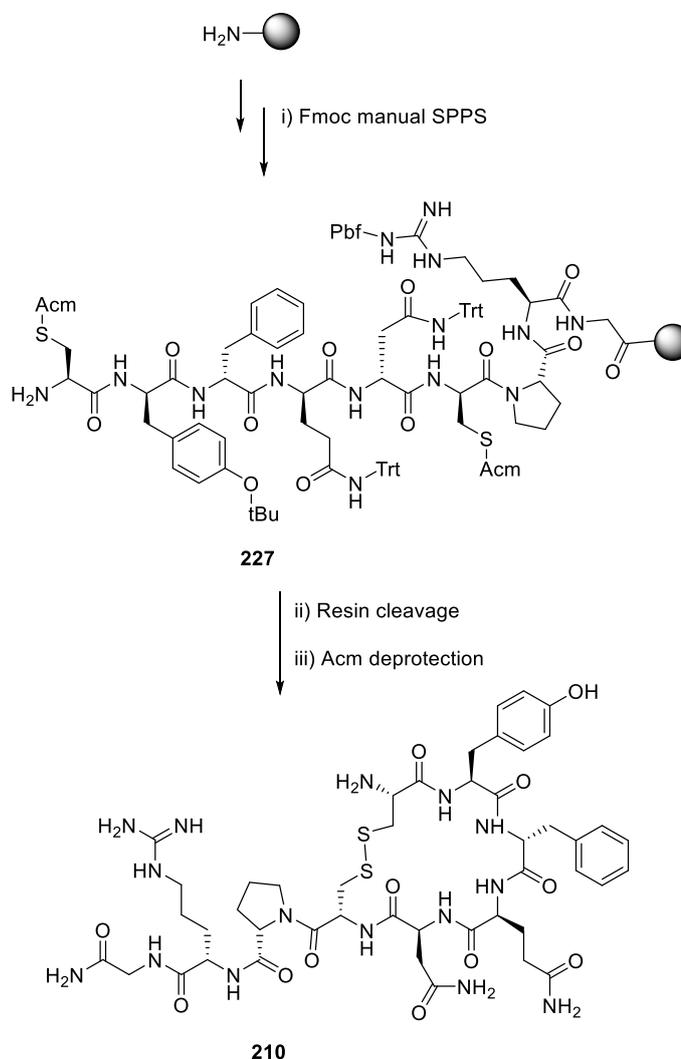
Scheme 3.4: Manual SPPS of VP (**210**) using RINK amide resin. Reagents and conditions: i) Manual Fmoc SPPS- Fmoc-Gly-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Pro-OH, Fmoc-Cys(Acm)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Gln(Trt)-OH, Fmoc-Phe-OH, Fmoc-Tyr(tBu)-OH, PyBOP, DIPEA and DMF, rt, (2 × 1 h) couplings. Fmoc deporotection- 20% piperidine/DMF, rt, 2 × 30 min ii) Cleavage from resin- 95% TFA: 2.5% TIPS: 2.5% H_2O, rt, 6 h iii) Acm cleavage- I_2, AcOH (40%)/DMF, rt, 1.5 h.

Upon completion of the synthesis and deprotection of the *N*-terminal Fmoc group, a test cleavage of the peptide from the resin was carried out and analysed *via* LCMS. To synthesise the cyclic VP analogue (**210**), the Acm protected peptide underwent a one-pot Acm deprotection and then disulphide bond formation using I_2, AcOH/DMF solution at rt. After 1.5 h the reaction was quenched using a 1 M solution of ascorbic acid and diluted using H_2O/MeCN until the final volume of AcOH was less than 10%. The peptide was purified using HPLC and analysed using LCMS (**Figure 3.10**).
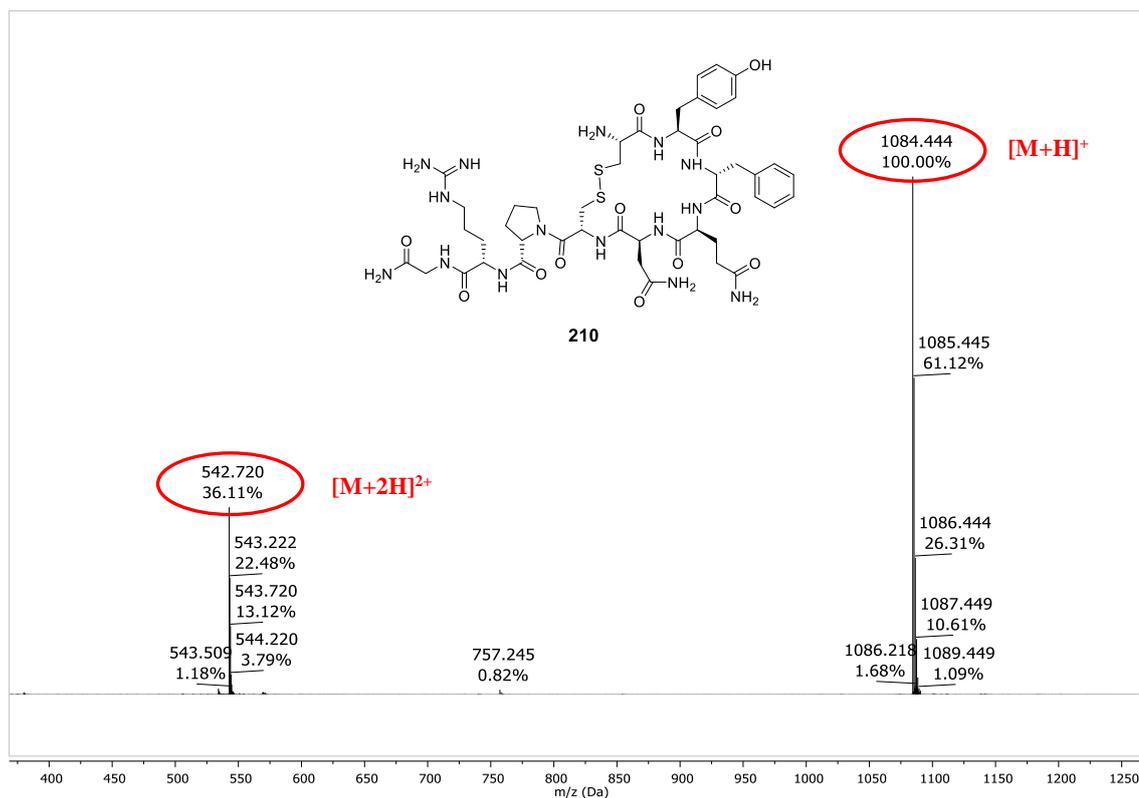
Figure 3.10: LCMS of native VP (**210**) with desired mass 1084.4 [M+H]$^+$.

LCMS of the HPLC fraction confirmed the presence of the desired linear peptide 1008.2 [M+H]$^+$ and charged species 504.9 [M+2H]$^{2+}$. The native OT (**209**) and VP (**210**) used as controls in later enzymatic degradation studies (**Section 3.3.2**).

3.2.1 Tagging of Oxytocin with Pentafluoropyridine

The linear peptide **C**(Acm)YIQN**C**(Acm)PLG (**224**) was synthesised on RINK Amide resin using Fmoc SPPS on a 0.2 mmol scale according to **Section 6.4.1**.

224 0.2 mmol
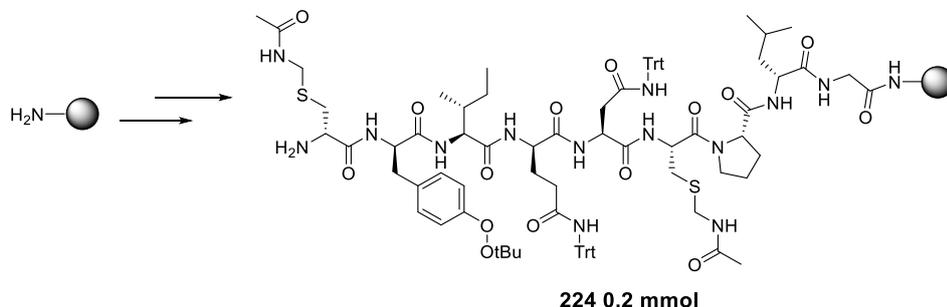
Scheme 3.5: Synthesis of linear peptide **224** *via* Fmoc manual SPPS on a 0.2 mmol scale using RINK amide resin. Reagents and conditions: i) Manual Fmoc SPPS- Fmoc-Gly-OH, Fmoc-Leu-OH, Fmoc-Pro-OH, Fmoc-Cys(Acm)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Gln(Trt)-OH, Fmoc-Ile-OH, Fmoc-Tyr(tBu)-OH, PyBOP, DIPEA and DMF, rt, (2 × 1 h) couplings. Fmoc deprotection- 20% piperidine/DMF, rt, 2 × 30 min.

For the synthesis of the singular PFP-tagged OT (**220**), the linear peptide (**224**) was tagged on-resin using PFP, DIPEA in DMF for 4 h on a 0.1 mmol scale followed by resin cleavage (**Scheme 3.6i and ii**). By carrying out the tagging reaction on resin, the only nucleophilic site available to react with PFP was the free terminal amine as the Tyr still contains a *t*-Butyl protecting group on the aromatic OH group (**Scheme 3.6i**).
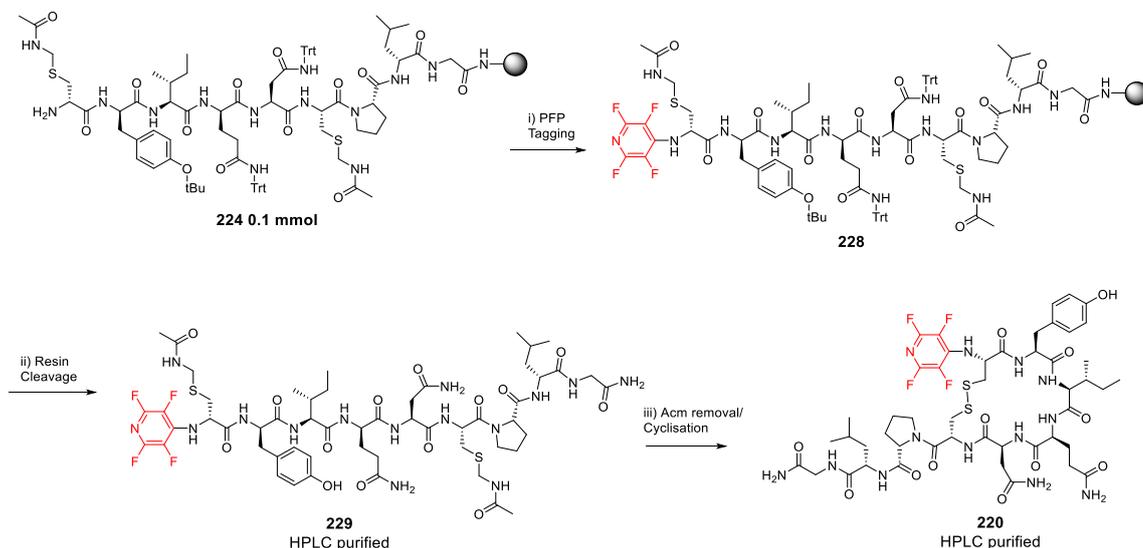
Scheme 3.6: Synthesis of singular-tagged OT (**220**). Reagents and conditions: i) On-resin PFP tagging- PFP, DIPEA, DMF, rt, 4 h ii) Cleavage from resin- 95% TFA: 2.5% TIPS: 2.5% H_2O, rt, 4 h iii) Acm cleavage- I_2, AcOH (40%)/DMF, rt, 1.5 h.

114

The presence of the PFP-tagged peptide (**229**) (with Acm groups attached) was confirmed using LCMS and the crude material was purified *via* RP-HPLC before it was used in the last stage of the synthesis. The Acm deprotection/disulphide bond formation was carried out using the reaction conditions described previously (e.g. synthesis of OT and VP). The crude reaction mixture was purified *via* RP-HPLC and the MS analysis of the purified PFP-tagged-peptide (**220**) is given below (**Figure 3.11**).
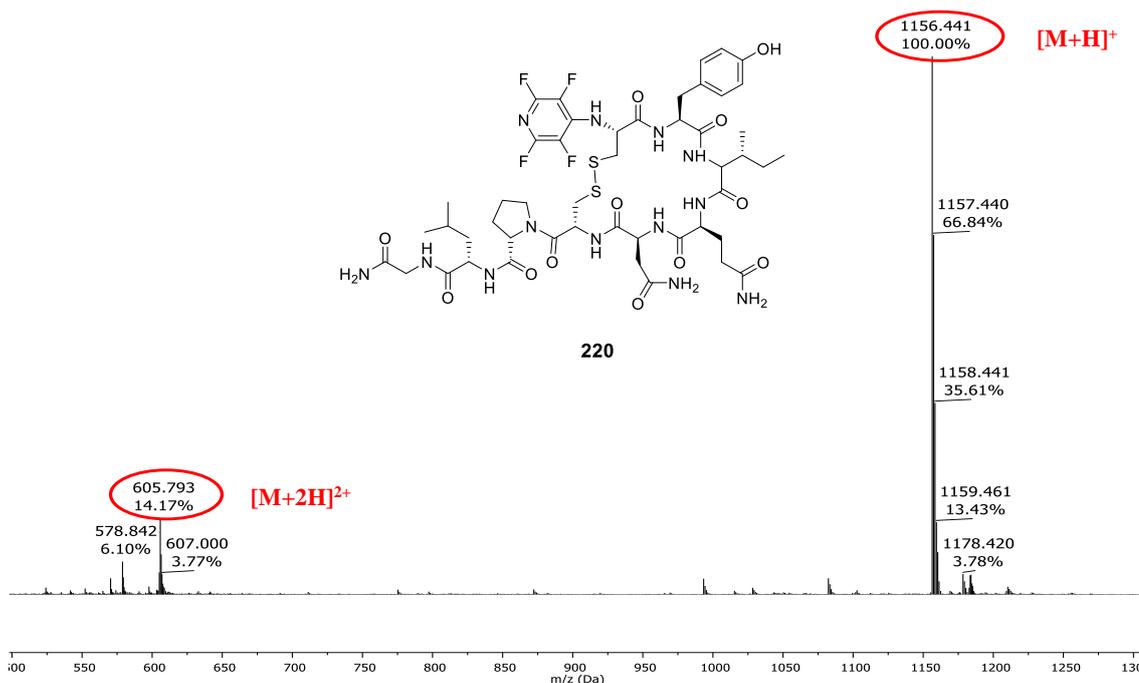
Figure 3.11: MS of purified PFP-tagged OT (**220**) with desired mass [M+H]$^+$ 1156.4 and [M+2H]$^{2+}$ 578.8.

The LCMS for purified tagged-peptide **220** confirmed the presence of the desired cyclic peptide 1156.4 [M+H]$^+$ and charged species 578.8 [M+2H]$^{2+}$. To confirm that the tagging of OT had occurred at the *N*-terminus MS-MS analysis was carried out. The MS-MS spectrum obtained for **220** is shown in **Figure 3.12.**

MS-MS (also referred to as tandem mass spectrometry) is a method of mass spectrometry used in the analysis of peptides and involves multiple stages of mass spectrometry with fragmentation of the peptide occurring between these stages (**Scheme 3.7**).
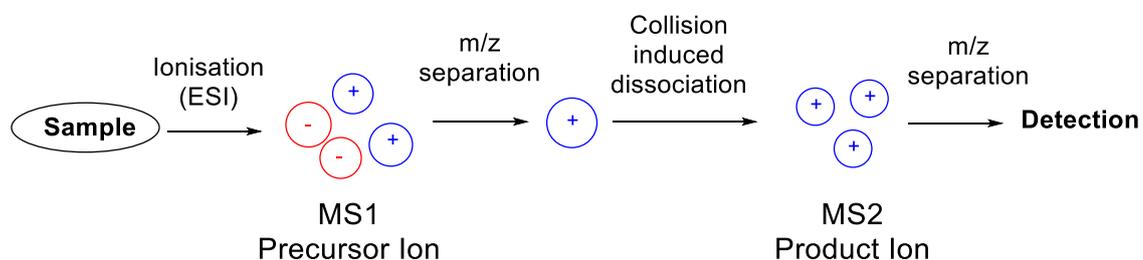
Scheme 3.7: Process of tandem mass spectrometry.[15]

There is an intial ionisation step *via* electrospray ionisation and the resulting ions are separated before undergoing a second step of mass spectrometry *via* collision induced dissociation. This results in product ions which are detected *via* time-of-flight spectrometry and analysed. Analysis of the resulting spectra allows for the determination of peptide structure due to the fact amino acid fragments will correspond to a specific pattern in the spectra.[15]
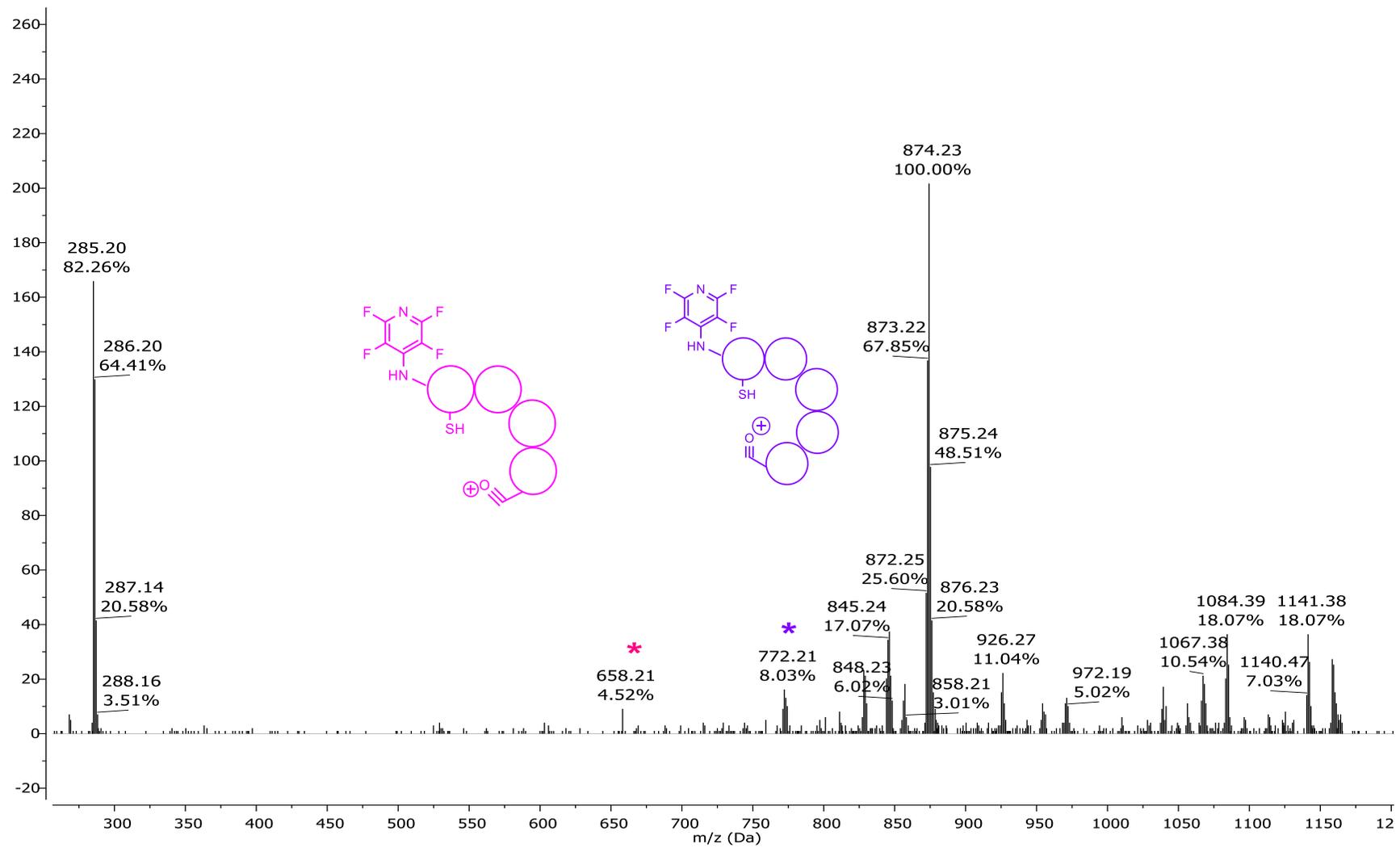
Figure 3.12: MS-MS spectrum of the singularly PFP tagged OT (**220**).

Entry	Calculated Mass	Observed Mass	Deletion
1	1156.2	1156.4	[M+H]- No deletion
2	771.7	772.2 *	Gly, Leu, Pro, Cys
3	657.7	658.2 *	Gly, Leu, Pro, Cys, Asn

Table 3.1: MS-MS data of singular PFP-tagged OT (**220**) detailing amino acid deletion.

The MS-MS data obtained (**Figure 3.12**) confirmed the presence of a singularly tagged OT analogue (**220**). Both fragments observed corresponds to a peptide which is single tagged, suggesting that PFP is attached to the *N*-terminus of the peptide only (**Figure 3.12**).

It is predicted that the fluorine *para* to the nitrogen will primarily react with the free amine of the *N*-terminus (as discussed in **Chapter 2**). ^{19}F NMR of the PFP-tagged peptide (**220**) was carried out to analyse the splitting pattern of the fluorines of the PFP-tagged peptide (**Figure 3.13**).
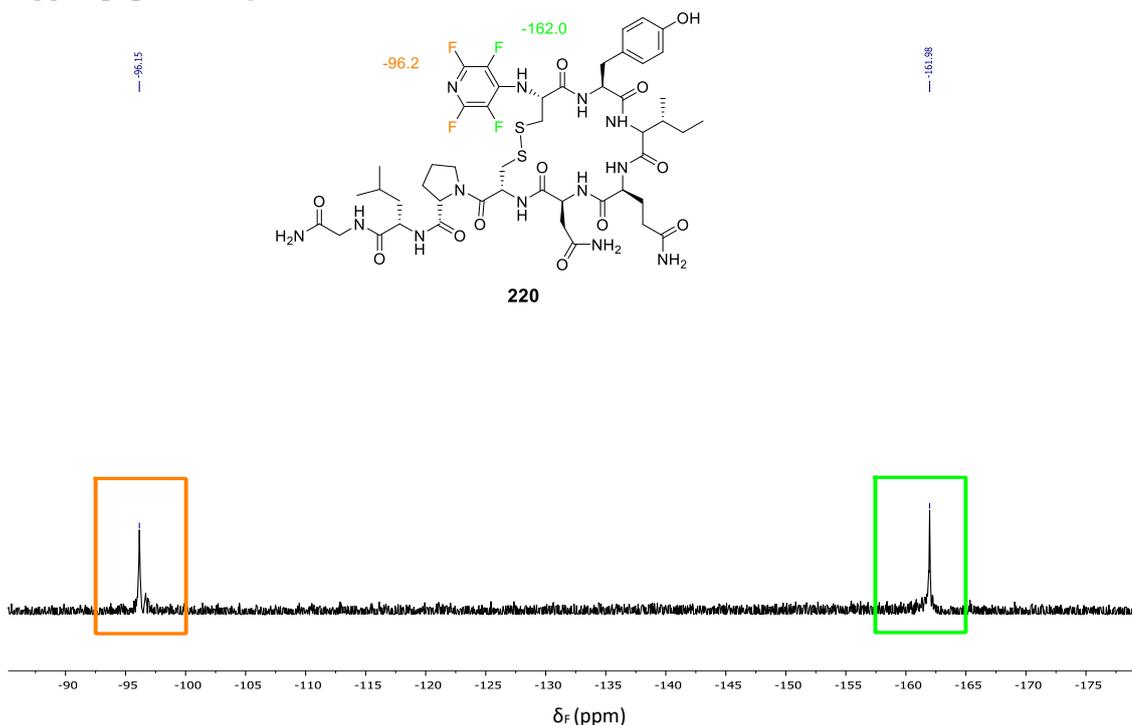
Figure 3.13: ^{19}F NMR spectrum of singular PFP-tagged OT (**220**).

It can be observed that there are two distinct fluorine environments at -96.2 ppm and -162.0 ppm. The fluorine signal at -96.2 ppm corresponds to the fluorine atoms which are *ortho* to the nitrogen and the -162.0 ppm signal corresponds to the fluorine atoms which are *meta* to the nitrogen as shown in **Figure 3.13.** According to previous studies carried out within the group, PFP attached to different amino acids exhibits distinct chemical shifts when analysed *via* ^{19}F NMR (**Table 3.2**).[16]
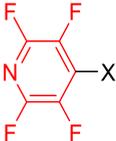
	^{19}F NMR chemical shifts (ppm)
X= Cys	-93.42 and -137.49
X= Lys	-98.17 and -165.54
X= Tyr	-91.32 and -155.98
X= Ser	-97.10 and -165.92

Table 3.2: ^{19}F NMR signals of pentafluoropyrdine attached to various nucleophilic amino acids.

Comparing the data in **Table 3.2** to the ^{19}F NMR data in **Figure 3.13**, it is possible to hypothesise that the PFP is attached the amine of the *N*-terminus. Although these chemical shifts are not reported in the above table, the data collected in the ^{19}F NMR in **Figure 3.13** (-96.2 and –162.0 ppm) closely resemble the data for PFP attached to a Lys (-98.2 and -165.5 ppm) rather than the values for Tyr or Cys. Tagging at the *N*-terminus is also supported by the MS-MS data as previously discussed.

For the synthesis of the doubly PFP-tagged OT (**221**), the linear peptide was prepared using Fmoc-SPPS and then it was cleaved from resin before PFP-tagging was carried out (**Scheme 3.8ii** and **iii**). By cleaving the peptide from resin, the nucleophilic phenol of Tyr is exposed and available for reaction with PFP as well as the free amine on the *N*-terminus (**Scheme 3.8**).
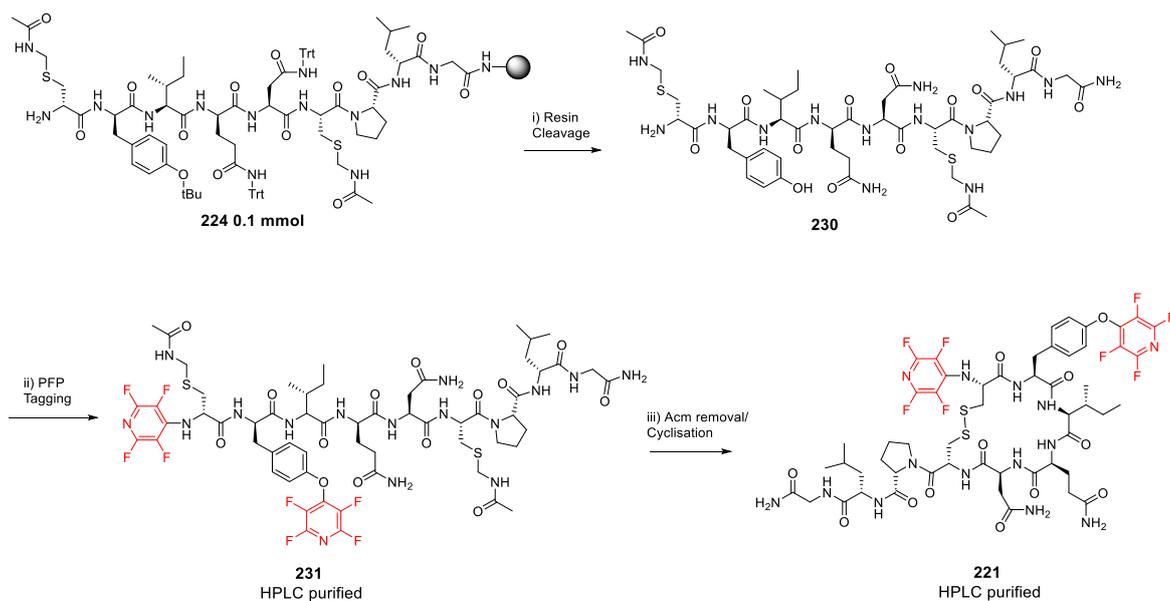
Scheme 3.8: Synthesis of double-tagged OT (**221**). Reagents and conditions: i) Cleavage from resin- 95% TFA: 2.5% TIPS: 2.5% H_2O, rt, 4 h ii) Solution phase tagging- pentafluoropyridine, DIPEA, DMF, rt, 4 h iii) Acm cleavage- I_2, AcOH (40%)/DMF, rt, 1.5 h.

The presence of the double PFP-tagged peptide (**231**) (with Acm groups attached) was confirmed using LCMS and **231** was purified by HPLC from the crude reaction mixture before being used in the final stage of the synthesis. The Acm deprotection/disulphide bond formation was carried out using the reaction conditions described in **Section 6.4.4.** The crude reaction mixture was purified *via* RP-HPLC and the MS analysis of the purified PFP-tagged-peptide (**221**) is given below (**Figure 3.14**).
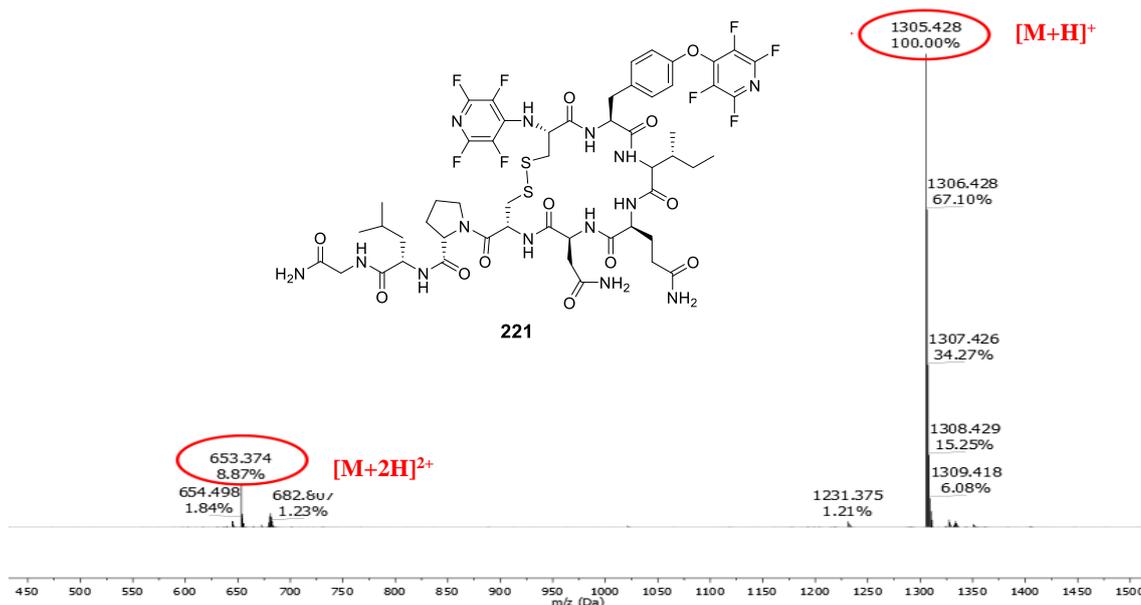
Figure 3.14: MS of purified doubly PFP-tagged OT (**221**) with desired mass $[M+H]^+$ 1305.4 and $[M+2H]^{2+}$ 653.3.

Shown in **Figure 3.14** is the MS for the purified, double PFP-tagged OT (**221**) illustrating the presence of the desired cyclic peptide 1305.4 $[M+H]^+$ and charged species 65.33 $[M+2H]^{2+}$. To confirm the tagging of OT at the site hypothesised in **Scheme 3.8**, MS-MS experiments were carried out to confirm the position of PFP tag within the peptide (**Figure 3.15**).
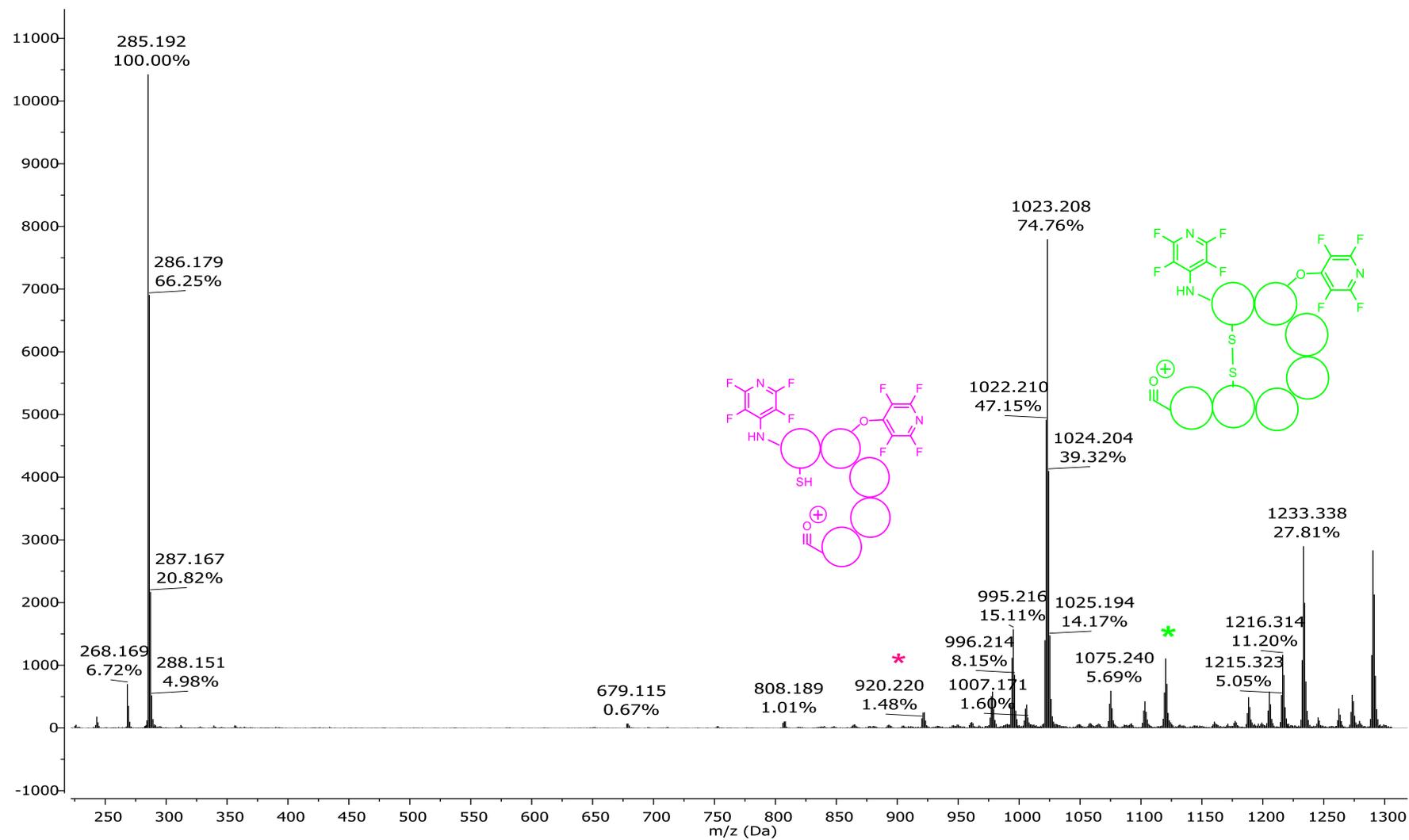
Figure 3.15: MS-MS analysis of the double PFP-tagged OT (**221**).

Entry	Calculated Mass	Observed Mass	Deletion
1	1305.2	1305.4	[M+H]-No deletion
2	1120.1	1120.4 *	Gly, Leu
3	920.7	920.2 *	Gly, Leu, Pro, Cys

Table 3.3: MS-MS data of double PFP-tagged OT (**221**) detailing amino acid deletion.

The MS-MS data obtained (**Figure 3.15**) confirmed the presence of a double-tagged OT analogue (**221**). Each fragment observed corresponds to a peptide which is double tagged (**Figure 3.15**), suggesting there are modifications sites at the phenol of Tyr and free amine of the *N*-terminus.

^{19}F NMR was carried out in order to analyse the splitting pattern of the fluorine of the pentafluoropyridine-tagged peptide. It is predicted that the fluorine *para* to the nitrogen will react with the free amine of the *N*-terminus and the phenol of the Tyr side chain (as discussed in **Chapter 2**). ^{19}F NMR spectroscopy was carried out using peptide (**218**) and the resulting ^{19}F NMR spectrum is shown below (**Figure 3.16**).
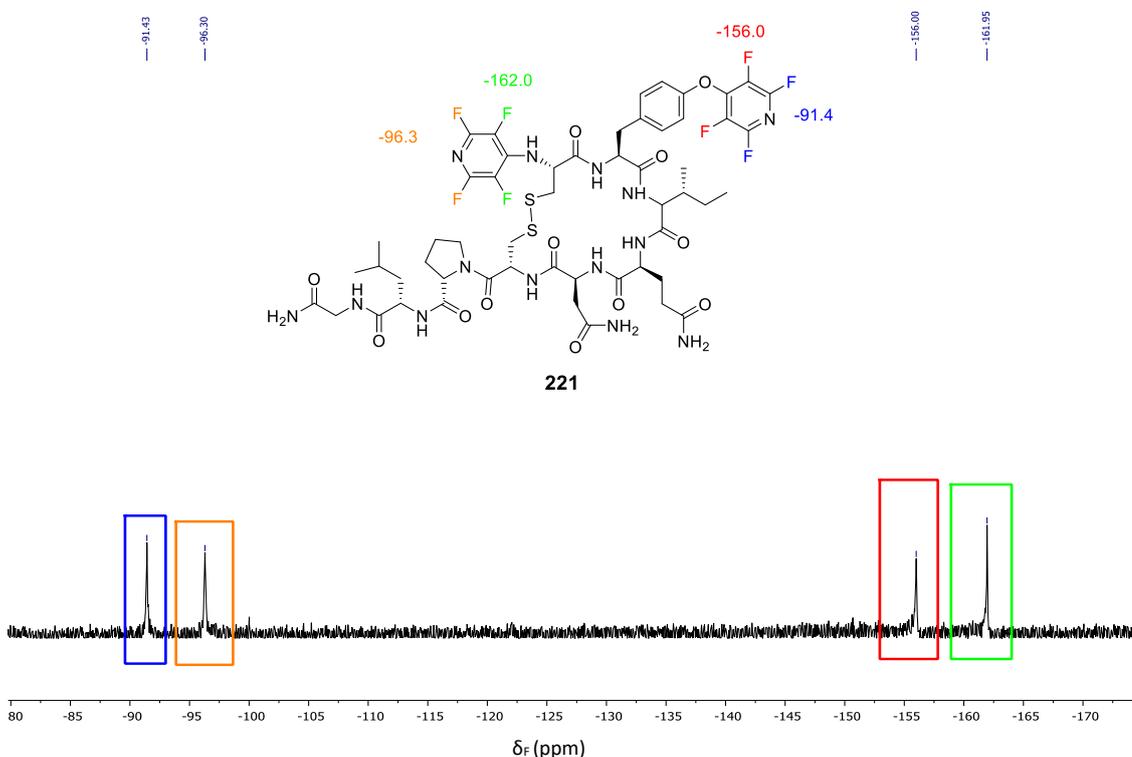
Figure 3.16: ^{19}F NMR spectrum of double-tagged OT (**221**).

123

It can be seen from the ^{19}F NMR analysis of **221** (**Figure 3.16**) that there are 4 distinct fluorine environments with signals at -91.4, -96.3, -156.0 and -162.0 ppm. The signals at -96.3 and -162.0 ppm correspond to the fluorine atoms attached to the fluoropyridine ring on the amine of the *N*-terminus. These signals match the data for the same constitution within the singular-tagged OT (**220**) (**Figure 3.13**). The signals at -91.4 and -156.0 ppm correspond to the fluorine atoms of fluoropyridine which are attached to the Tyr side chain. This aligns with data observed in **Table 3.2** which suggests that the signals observed when fluoropyridine is attached to a Tyr are -91.4 and -156.0 ppm.

3.2.2 Attempted Tagging of Oxytocin using Hexafluorobenzene

The selective tagging with PFP (**129**) was shown to be an efficient and quick method to alter the structure of OT and so further investigations using other fluorinated aromatic compounds were carried out. We chose to explore the tagging of OT using HFB (**47**) which has been extensively researched by Pentalute and co-workers[16] for its ability to staple proteins by utilising the nucleophilic nature of cysteine.

We hypothesised that HFB (**47**) could react with the nucleophilic residues in OT if the same reaction conditions were used as in the PFP tagging. This would give rise to novel HFB-tagged peptides as shown in **Figure 3.17**, if the same on-resin and solution phase tagging techniques were adopted (as with the PFP tagging of OT).

Chemical Formula: $C_{61}H_{76}F_{10}N_{14}O_{14}S_2$
Exact Mass: 1482.49
Molecular Weight: 1483.47

Chemical Formula: $C_{55}H_{77}F_5N_{14}O_{14}S_2$
Exact Mass: 1316.5105
Molecular Weight: 1317.4170
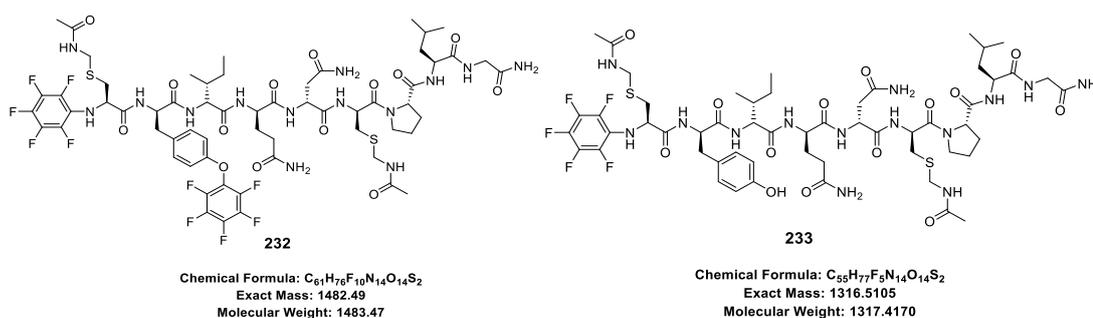
Figure 3.17: Theoretical singular and double-HFB tagged peptides (**232** and **233**).

Peptide **230** 1152.2 [M+H]$^+$ (1 mg) was dissolved in DMF (1 ml) and DIPEA and HFB (**47**) added. The reaction was left under agitation for 4.5 h and after this time DMF was removed under reduced pressure. The resulting crude mixture was analysed using LCMS and analytical HPLC. The MS trace of the resulting peptide is shown below in **Figure 3.18**.
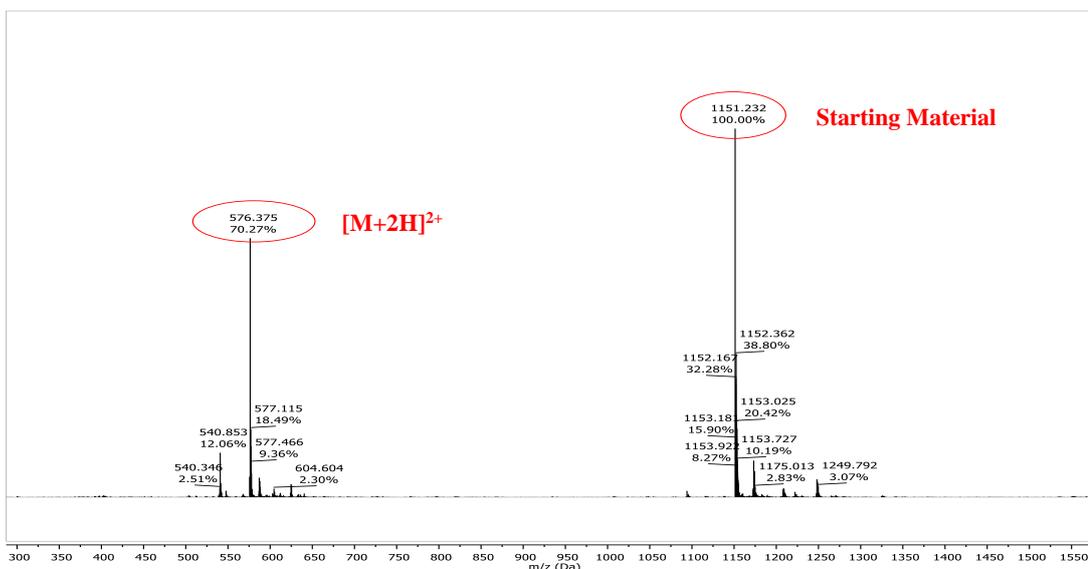
Figure 3.18: MS of the crude HFB-tagged peptide with expected 1483 [M+H]$^+$.

As illustrated in **Figure 3.18**, there is no peak for the desired mass of 1483 and only presence of starting material (**230**) 1151 [M+H]$^+$ and charged species 576 [2M+H]$^{2+}$. This reaction was repeated, using fresh reagents, but the same outcome was observed. Given the lack of reactivity of HFB (**47**) towards OT tagging the decision was made to precede with the preparation of the VP analogues only using PFP.

3.2.3 Tagging of Vasopressin with Pentafluoropyridine

The linear peptide **C**(Acm)YFQN**C**(Acm)PRG was synthesised on resin using manual Fmoc SPPS on a 0.2 mmol scale (**Scheme 3.9**). Once the target sequence was prepared the resin was split into two equal batches.

227 0.2 mmol
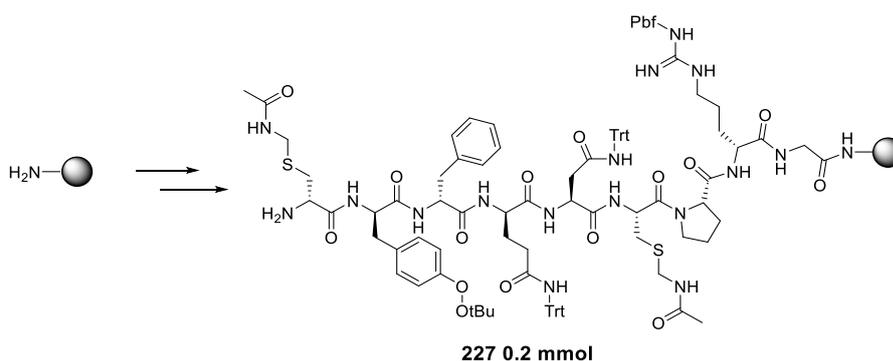
Scheme 3.9: Synthesis of linear peptide **227** synthesised *via* Fmoc SPPS on a 0.2 mmol scale using RINK amide resin. Reagents and conditions: Manual Fmoc SPPS- Fmoc-Gly-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Pro-OH, Fmoc-Cys(Acm)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Gln(Trt)-OH, Fmoc-Phe-OH, Fmoc-Tyr(tBu)-OH, PyBOP, DIPEA and DMF, rt, (2 × 1 h) couplings. Fmoc deporotection- 20% piperidine/DMF, rt, 2 × 30 min.

125

As with the synthesis of the OT analogues, the singular tagged VP analogue was synthesised using an on-resin PFP tagging reaction. By tagging the peptide on-resin it was possible to selectively react PFP with the amine of the *N*-terminus of VP while the hydroxyl of the Tyr remained protected by *t*-butyl group (**Scheme 3.10**).
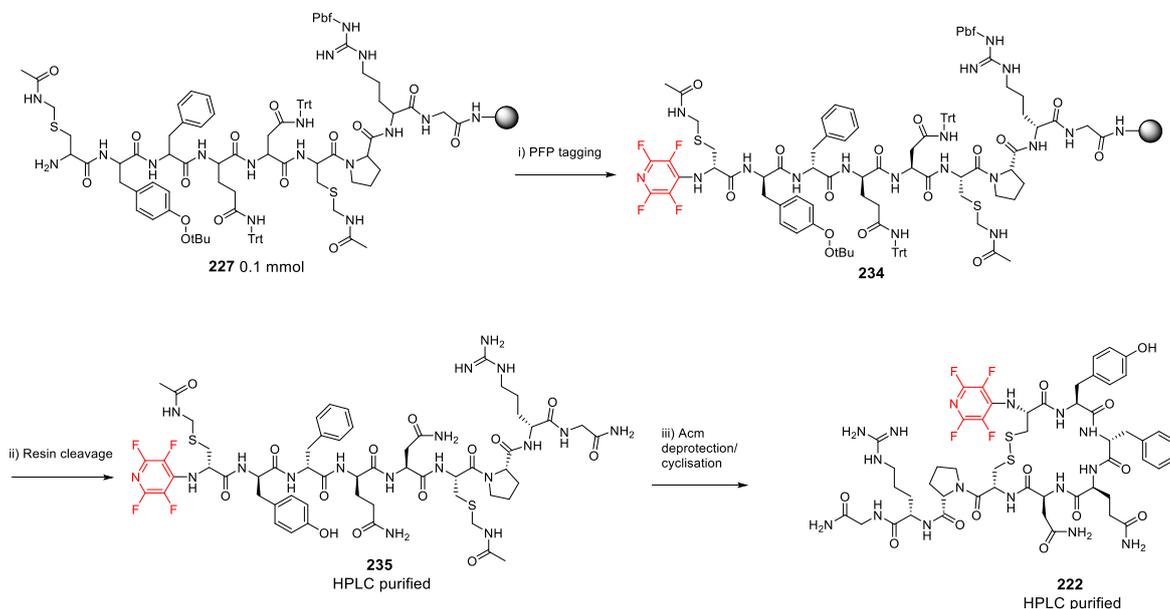
Scheme 3.10: Synthesis of singular tagged VP (**222**). Reagents and conditions: i) On-resin PFP tagging- PFP, DIPEA, DMF, rt, 4 h ii) Cleavage from resin- 95% TFA: 2.5% TIPS: 2.5% H_2O, rt, 4 h iii) Acm cleavage- I_2, AcOH (40%)/DMF, rt, 1.5 h.

The presence of the PFP-tagged peptide (**235**) (with Acm groups attached) was confirmed using LCMS and purified *via* preperative HPLC before being used in the last stage of the synthesis. The Acm deprotection/disulphide bond formation was carried out using the reaction conditions described in **Section 6.4.4**. The crude reaction mixture was purified *via* RP-HPLC and the LCMS analysis of the purified PFP-tagged-peptide (**222**) is given below (**Figure 3.19**).
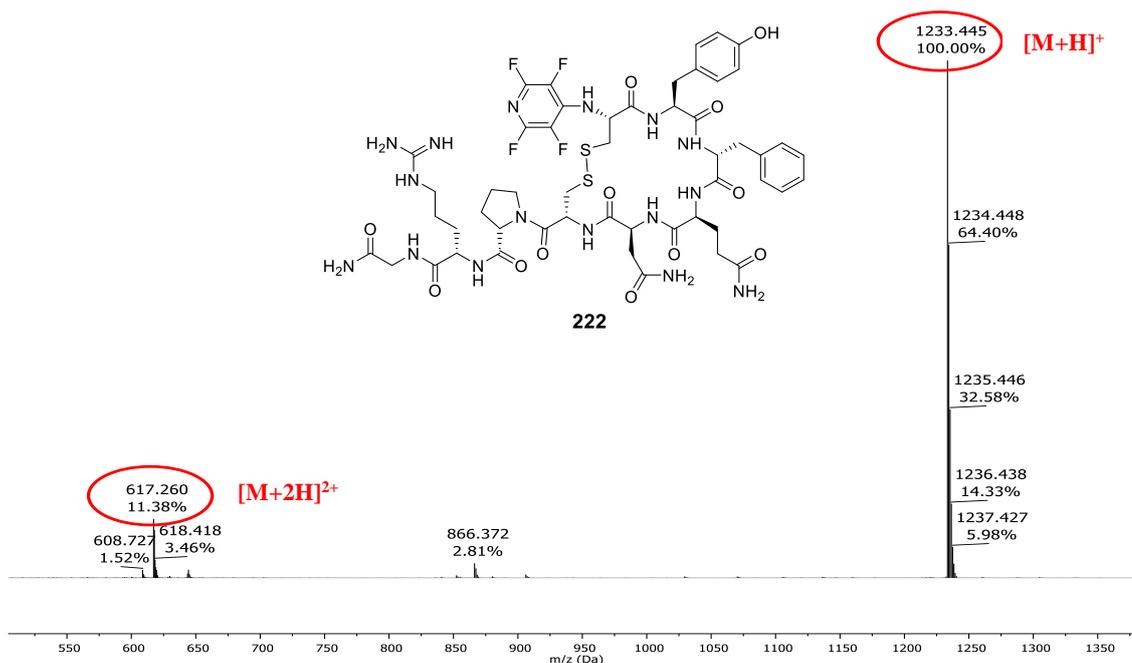
Figure 3.19: MS of purified PFP-VP (**222**) with desired mass [M+H]$^+$ 1233.4 and [M+2H]$^{2+}$ 617.3.

Shown in **Figure 3.19** is the MS for the purified PFP-tagged VP illustrating the presence of the desired cyclic peptide 1233.4 [M+H]$^+$ and charged species 617.3 [M+2H]$^{2+}$. To confirm the tagging of VP at the site hypothesised in **Scheme 3.10**, MS-MS experiments were carried out to confirm the amino acid sequence of the peptide and the resulting spectra shown in **Figure 3.20**. In this figure, the relevant peaks have been highlighted and the corresponding structure of the fragment overlaid on the spectra.
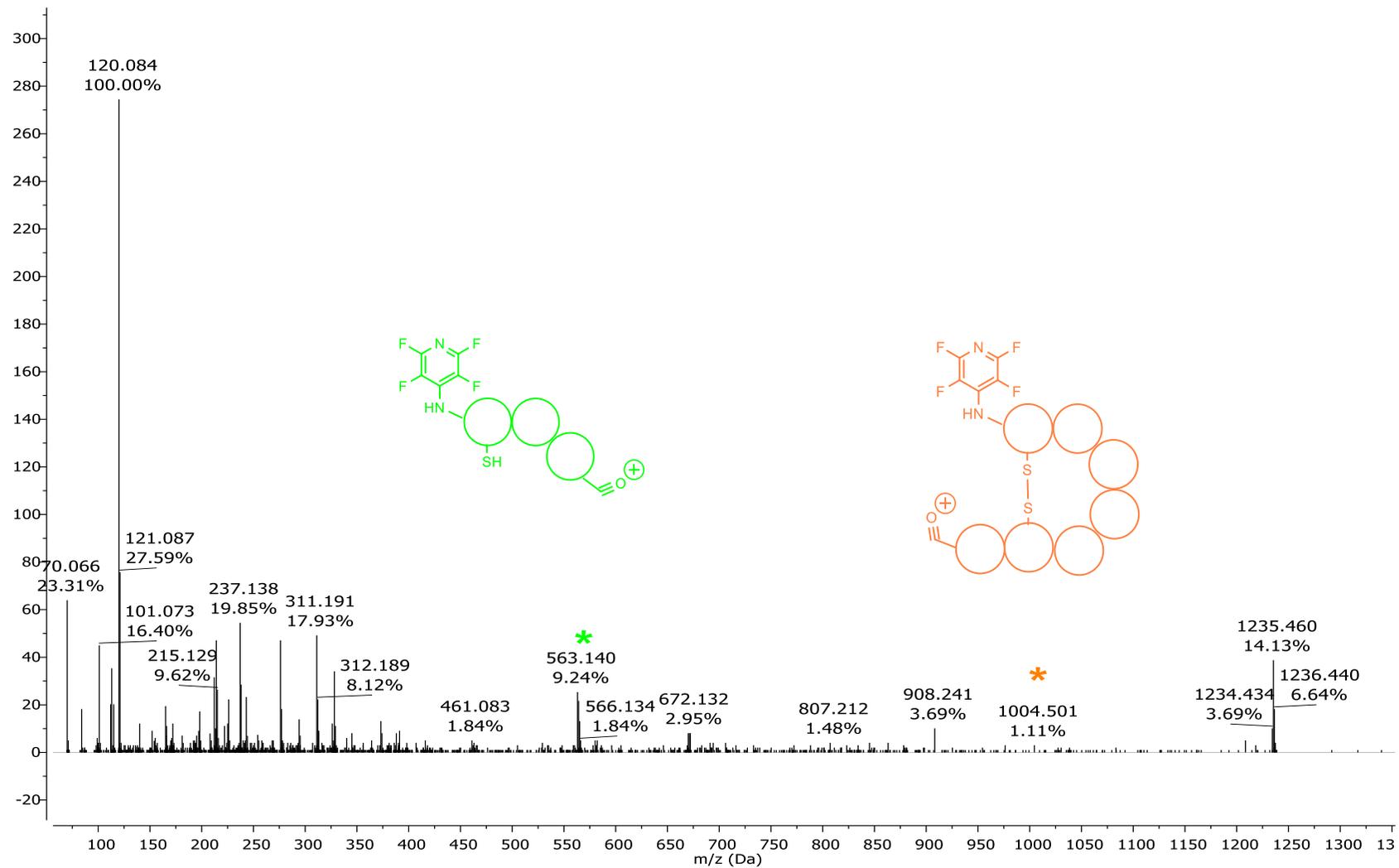
Figure 3.20: MSMS data of singularly tagged VP analogue (**222**).

Entry	Calculated Mass	Observed Mass	Deletion
1	1233.3	1233.5	No deletion [M+H]$^+$
2	1004.2	1004.5 *	Gly, Arg
3	563.1	563.1 *	Gly, Leu, Pro, Cys, Asn, Gln

Table 3.4: MSMS analysis of singular-PFP-tagged VP (**222**) detailing amino acid deletion.

The MS-MS data obtained (**Figure 3.20**) confirmed the presence of a single-tagged VP analogue (**222**). Each fragment observed corresponds to a peptide which is single tagged, suggesting that PFP is attached to the *N*-terminus of the peptide only (**Figure 3.20**).

^{19}F NMR analysis was carried out to analyse the splitting pattern of the fluorines of the PFP-tagged peptide. It is predicted that the fluorine *para* to the nitrogen will primarily react with the free amine of the *N*-terminus (as discussed in **Chapter 2**). ^{19}F NMR analysis was carried out on peptide (**222**) and the resulting ^{19}F NMR spectrum is shown below (**Figure 3.21**).
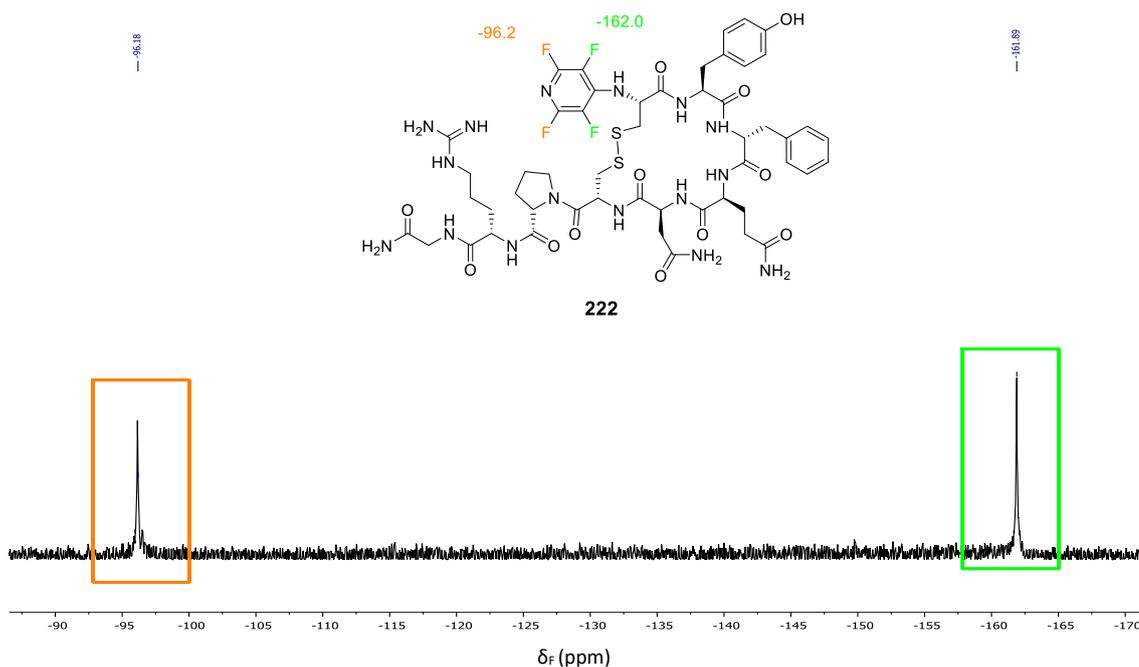
Figure 3.21: ^{19}F NMR spectrum of singularly tagged VP (**222**).

As seen in **Figure 3.21**, there are 2 distinct chemical shifts and therefore 2 distinct chemical environments. The signal at -96.2 ppm corresponds to the fluorine atoms which are *ortho* to the nitrogen and -162.0 ppm signal corresponds to the fluorine atoms which are *meta* to the nitrogen. The similarity between ^{19}F NMR spectra of singular tagged OT (**Figure 3.13**) and VP (**Figure 3.20**) suggests that both structures contain a fluoropyrdine ring attached to the amine of the *N*-terminus *via* displacement of the fluorine *para* to the nitrogen.

As with the synthesis of the double tagged OT analogue (**221**), the double tagged VP analogue (**223**) was synthesised using solution phase reaction conditions. By carrying out the tagging reaction in solution phase, both the nucleophilic Tyr side chain and amine of the *N*-terminus are available for reaction with PFP and this was confirmed *via* LCMS of the crude reaction material (**Scheme 3.11**).
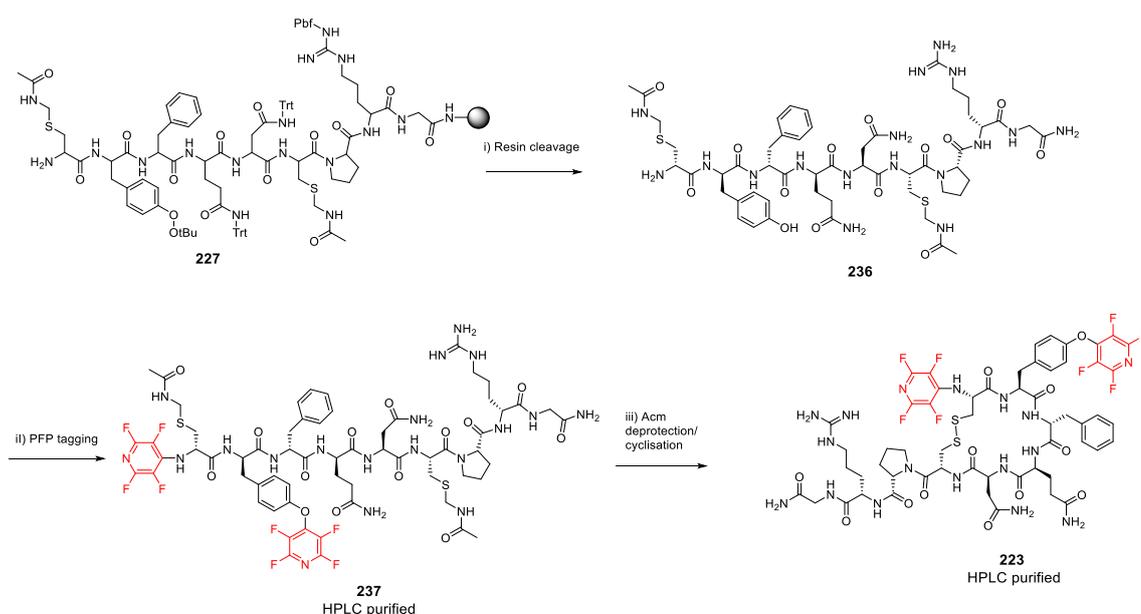
Scheme 3.11: Synthesis of double tagged VP (**223**) i) Cleavage from resin- 95% TFA: 2.5% TIPS: 2.5% H$_2$O, rt, 4 h ii) Solution phase tagging- PFP, DIPEA, DMF, rt, 4 h iii) Acm cleavage- I$_2$, AcOH (40%)/DMF, rt, 1.5 h.

The presence of the PFP-tagged peptide (**237**) (with Acm groups attached) was confirmed using LCMS and the crude peptide was purified *via* preparative HPLC before it was used in the last stage of the synthesis. The Acm deprotection/disulphide bond formation (**Scheme 3.11**) was carried out using the reaction conditions described in **Section 6.4.4**. The crude reaction mixture was purified *via* RP-HPLC and the MS analysis of the purified PFP-tagged-peptide (**223**) is given below (**Figure 3.22**).
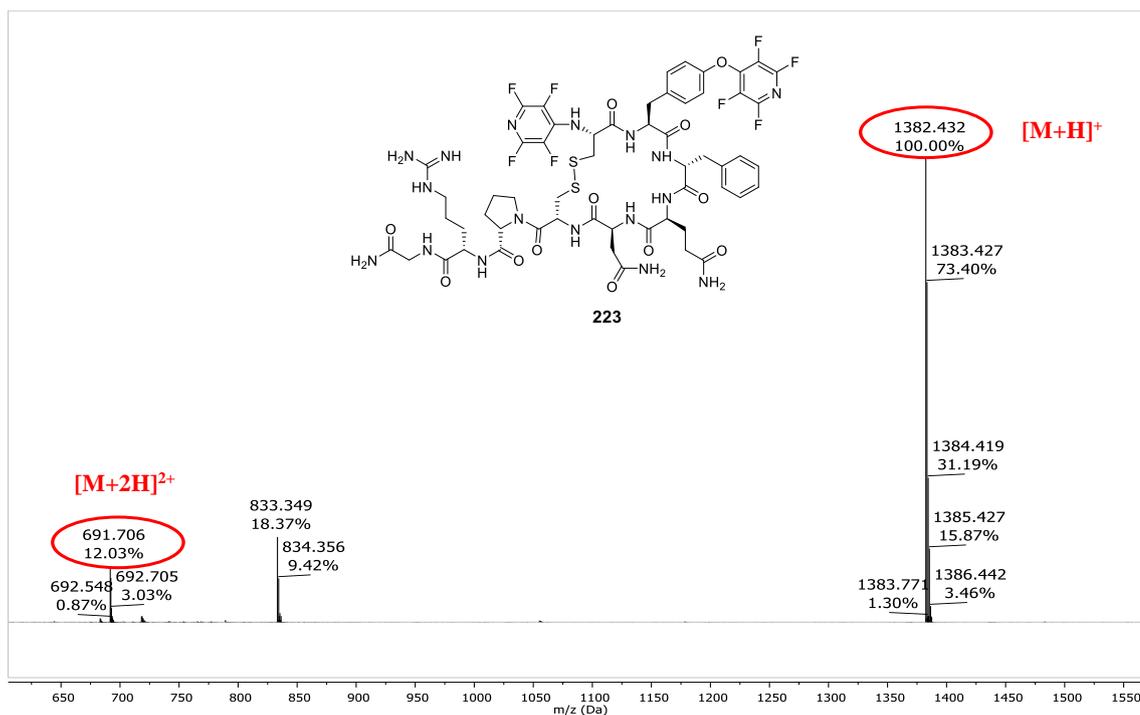
Figure 3.22: MS of purified doubly tagged PFP tagged VP (**223**) with desired material 1382.4 [M+H]⁺ and 691.7 [M+2H]²⁺.

Shown in **Figure 3.22** is the MS for the purified doubly PFP-tagged VP illustrating the presence of the desired cyclic peptide $[M+H]^+$ 1382.9 and charged species $[M+2H]^{2+}$ 692.3. To confirm the tagging of VP at the site hypothesised in **Scheme 3.11**, MS-MS experiments were carried out to confirm the amino acid sequence of the peptide and resulting spectra shown in **Figure 3.23**. In this figure, the relevant peaks have been highlighted and the corresponding structure of the fragment overlaid on the spectra.
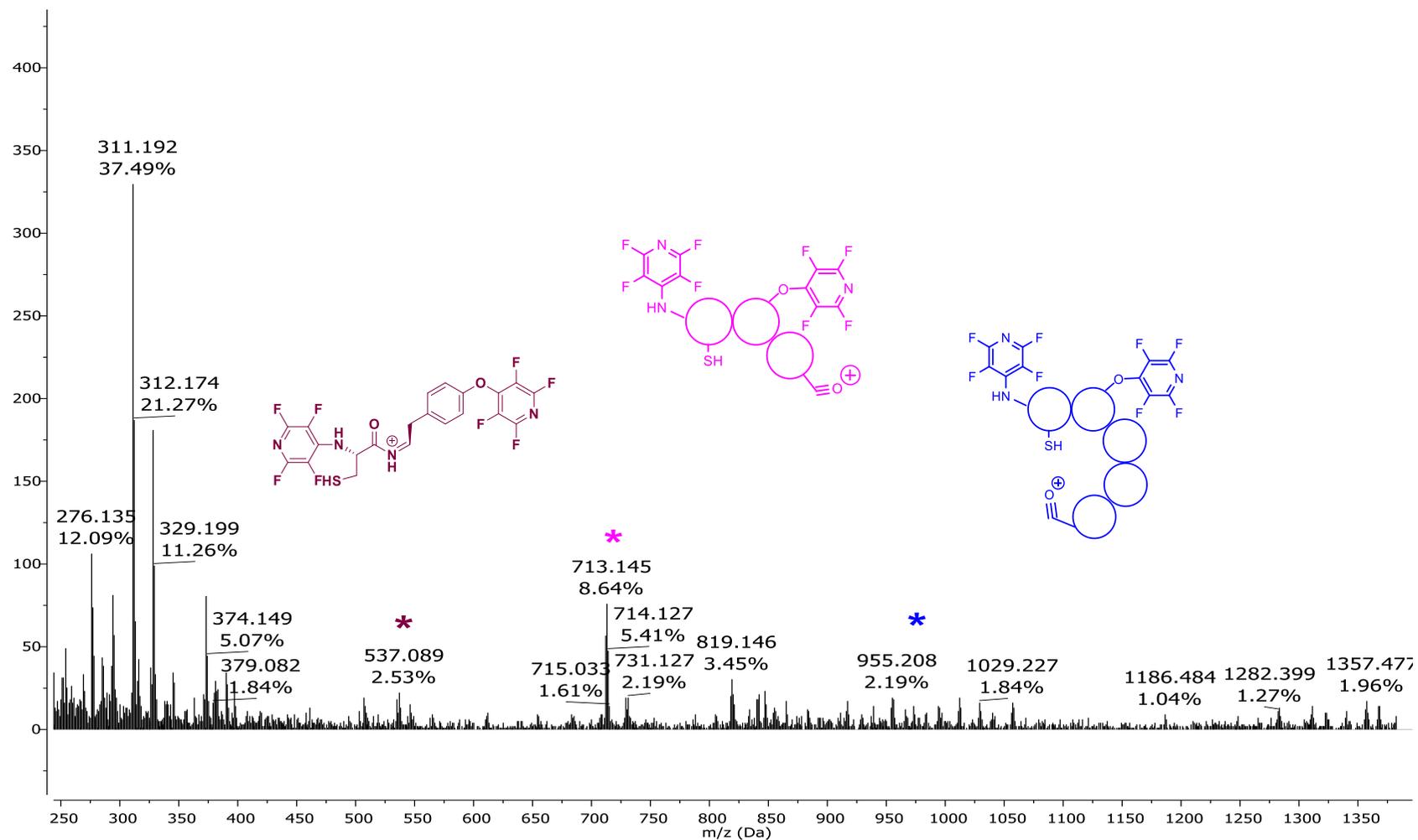
Figure 3.23: MSMS analysis of double PFP-tagged VP (**223**).

Entry	Calculated Mass	Observed Mass	Deletion
1	1382.3	1382.4	No deletion [M+H]
2	954.8	955.2 *	Gly, Arg, Pro, Cys
3	712.6	713.1 *	Gly, Leu, Pro, Cys, Asn, Gln
4	537.4	537.1 *	Gly, Leu, Pro, Cys, Asn, Gln, Phe

Table 3.5: Analysis of MS-MS data of double-tagged VP (**223**) detailing amino acid deletion.

The MS-MS data obtained (**Figure 3.23**) confirmed the presence of a double-tagged VP analogue (**223**). Each fragment observed corresponds to a peptide which is double tagged (**Figure 3.23**), suggesting there are modifications sites at the phenol of Tyr and free amine of the *N*-terminus.

^{19}F NMR spectroscopy was carried out in order to analyse the splitting pattern of the fluorines of the pentafluoropyridine-tagged peptide. It is predicted that the fluorine *para* to the nitrogen will react with the free amine of the *N*-terminus and the phenol of the Tyr side chain (as discussed in **Chapter 2**). ^{19}F NMR spectroscopy was carried on peptide (**223**) and the resulting ^{19}F NMR spectrum is shown below (**Figure 3.24**).
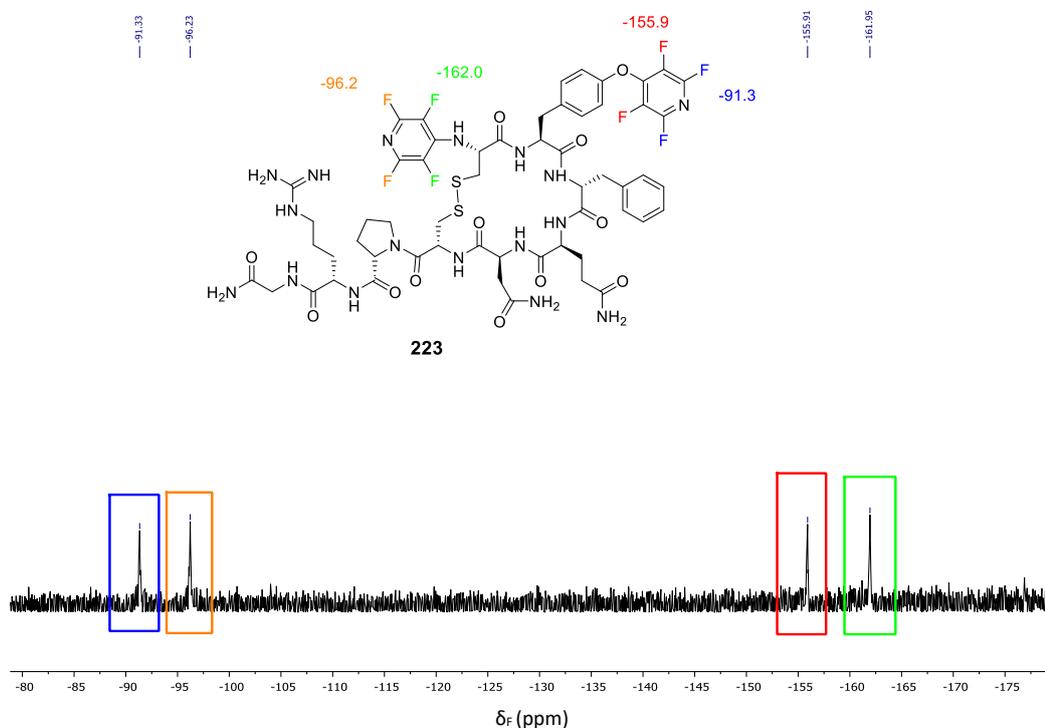
Figure 3.24: ^{19}F NMR spectrum of double-PFP-tagged VP (**223**).

The signals highlighted in **Figure 3.24** show there are 4 distinct fluorine signals and therefore 4 distinct chemical environments with signals at -91.3, -96.3, -155.9 and -162.0 ppm. The signals at -96.3 and -162.0 ppm correspond to the fluorine atoms attached to the fluoropyridine ring on the amine of the *N*-terminus which matches the ^{19}F NMR data collected from the singularly tagged VP (**223**). The signals at -91.3 and -155.9 ppm correspond to the fluorine atoms of fluoropyridine which are attached to the Tyr side chain. The similarity in ^{19}F NMR spectra of double-tagged OT (**Figure 3.16**) and VP (**Figure 3.24**) suggest that both structures contain a fluoropyrdine ring attached to the amine of the *N*-terminus and phenol of the Tyr side chain *via* the fluorine *para* to the nitrogen.

Using both solid phase and solution phase tagging techniques has proven an effective way of synthesising a series of tagged peptide analogues. These analogues were all purified *via* preparative HPLC to yield individual products which have been analysed by mass spectroscopy and analytical HPLC traces. With the synthesis, purification and analysis of the singular and doubly tagged OT and VP analogues carried out, investigation into the enzymatic stability of these compounds was carried out.

3.3 Enzymatic Degradation Studies

The instability of peptides *in vivo* presents a major obstacle in developing viable peptide-based drugs. OT (**209**) has a $t_{1/2}$= 5 min and VP (**210**) has $t_{1/2}$= 10 min[17] and although both are used as therapeutics at present, the possibility of improving the half-lives and the overall chemical properties of OT (**209**) and VP (**210**) is extremely attractive. If peptide stability could be improved *via* the modification with PFP (**129**), this could provide an interesting tool for improving stability as well as further modification of other therapeutically relevant peptides.

The first test of stability for the PFP-tagged OT (**220** and **221**) and VP analogues (**222** and **223**) was an enzymatic degradation test using chymotrypsin. Chymotrypsin is a digestive enzyme which facilitates the breakdown of proteins and peptides in the duodenum and functions by cleaving amide bonds at the carboxyl group of hydrophobic amino acids such as tyrosine (**Scheme 3.12**) by using Ser, His and Asp residues on the enzyme surface.[18]
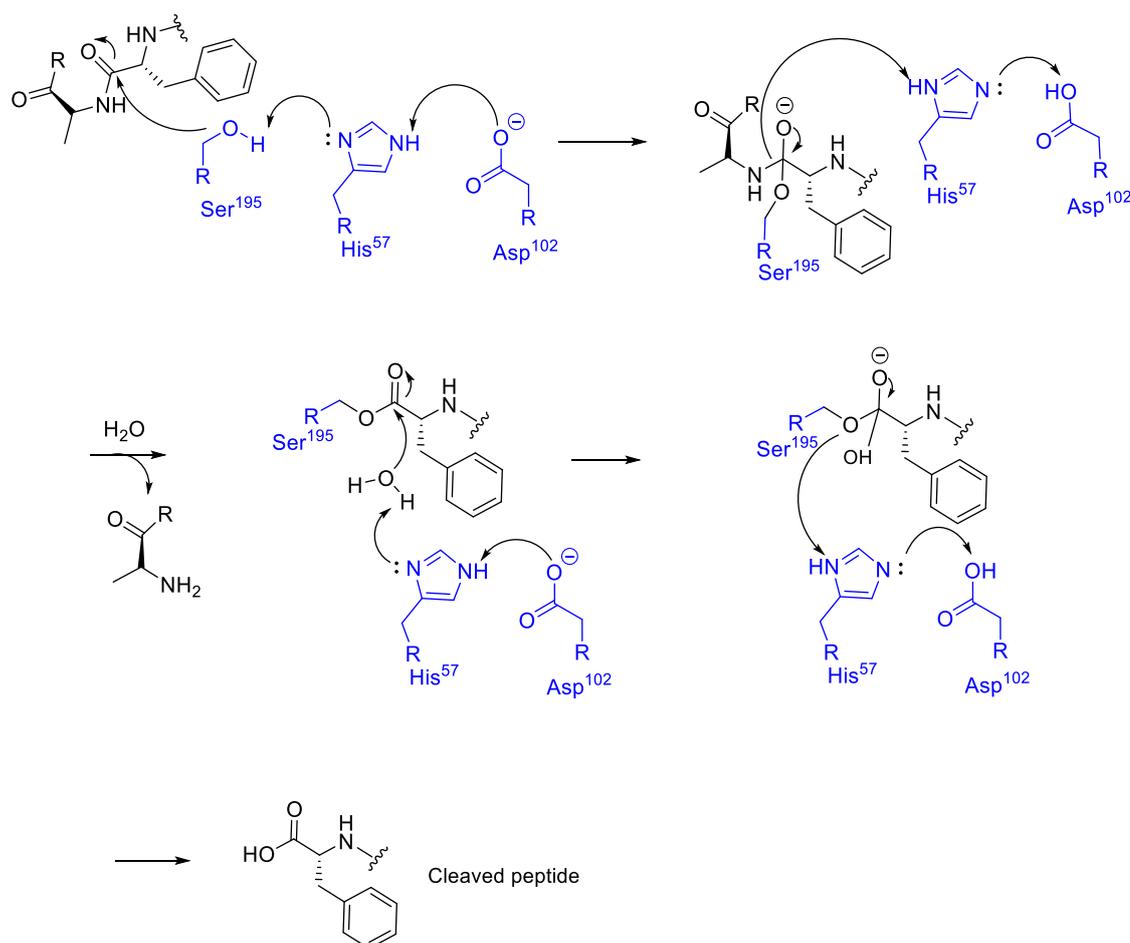
Scheme 3.12: Chymotrypsin cleavage mechanism.

3.3.1 Enzymatic Breakdown Using Chymotrypsin

Both the OT (**209**) and VP (**210**) sequences contain a Tyr which is susceptible to chymotrypsin cleavage as well as undergo cleavage of the amide bonds in general. The enzymatic degradation of the OT and VP analogues were carried out in a water bath set at 23 °C. The peptides OT (**209**), PFP-OT (**220**), 2PFP-OT (**221**), VP (**210**), PFP-VP (**222**) and 2PFP-VP (**223**) were dissolved in Tris buffer (pH 7.6) with a resulting concentration of 1 mg of peptide per 1 ml of buffer. The total volume of peptide/buffer solution needed for the experiment was 120 µL so a total of 0.12 mg of each peptide was required for the degradation studies. α-Chymotrypsin from bovine pancreas was purchased from Sigma Aldrich and dissolved in Tris buffer (pH 7.6) to a final concentration of 1 mg/ml. The peptide and enzyme solutions were added to an Eppendorf so that the ratio of peptide to enzyme solution was 25:1 and the Eppendorf was placed in a water bath (set at 23 °C) (**Scheme 3.13**).
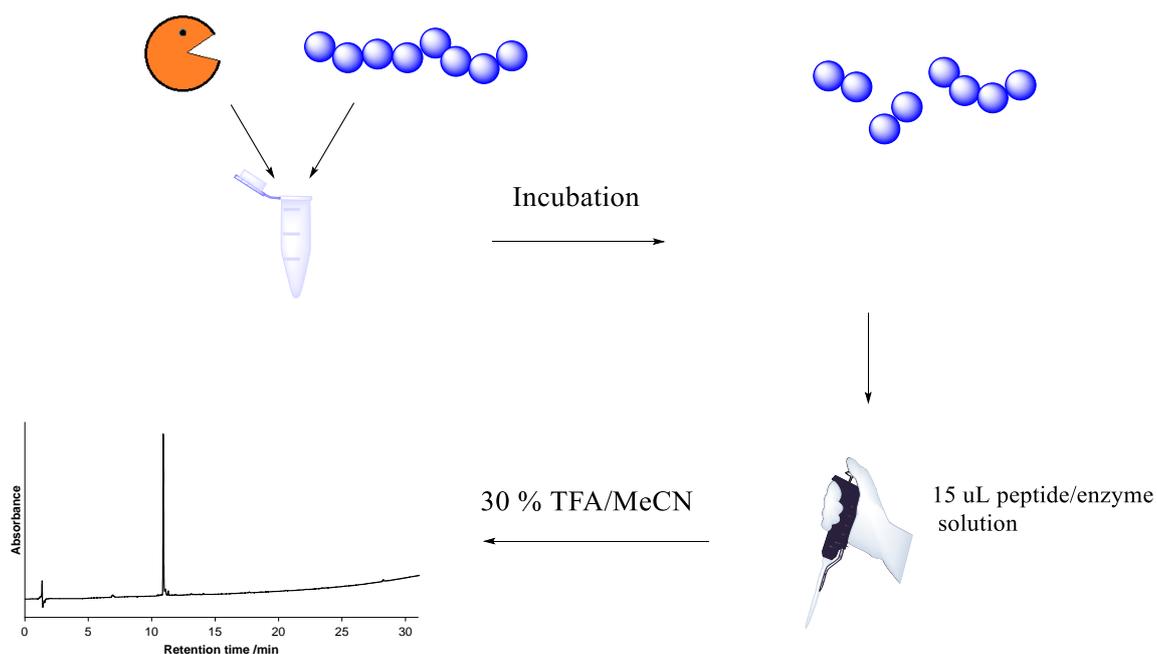
Scheme 3.13: Enzymatic degradation procedure using analytical HPLC to trace peptide fragments.

At this point, an initial aliquot of 15 µL was removed from the reaction mixture and diluted with MeCN and H_2O and analysed *via* analytical HPLC to give a reading for t = 0. At t = 1, 2, 3, 4, 5, 6 and 24 h, 15 µL aliquots of the peptide/enzyme mixture were removed and 15µL of TFA (30%)/MeCN added, diluted with MeCN and H_2O and analysed by analytical HPLC.

3.3.2 Stability Studies of Oxytocin and Pentafluoropyridine-Tagged Analogues

The enzymatic stability of native OT (**209**) and VP (**210**) were analysed as controls. Shown in **Figure 3.25** are the analytical HPLC traces for the enzymatic degradation of native OT. At t = 0 there is one major peak which corresponds to native OT and this was confirmed *via* LCMS. The deletion products, as determined by LCMS, for the enzymatic degradation of compound **209** are shown in **Table 3.6**.

t=0

t=4

t=1

t=5

t=2

t=6

t=3

t=24
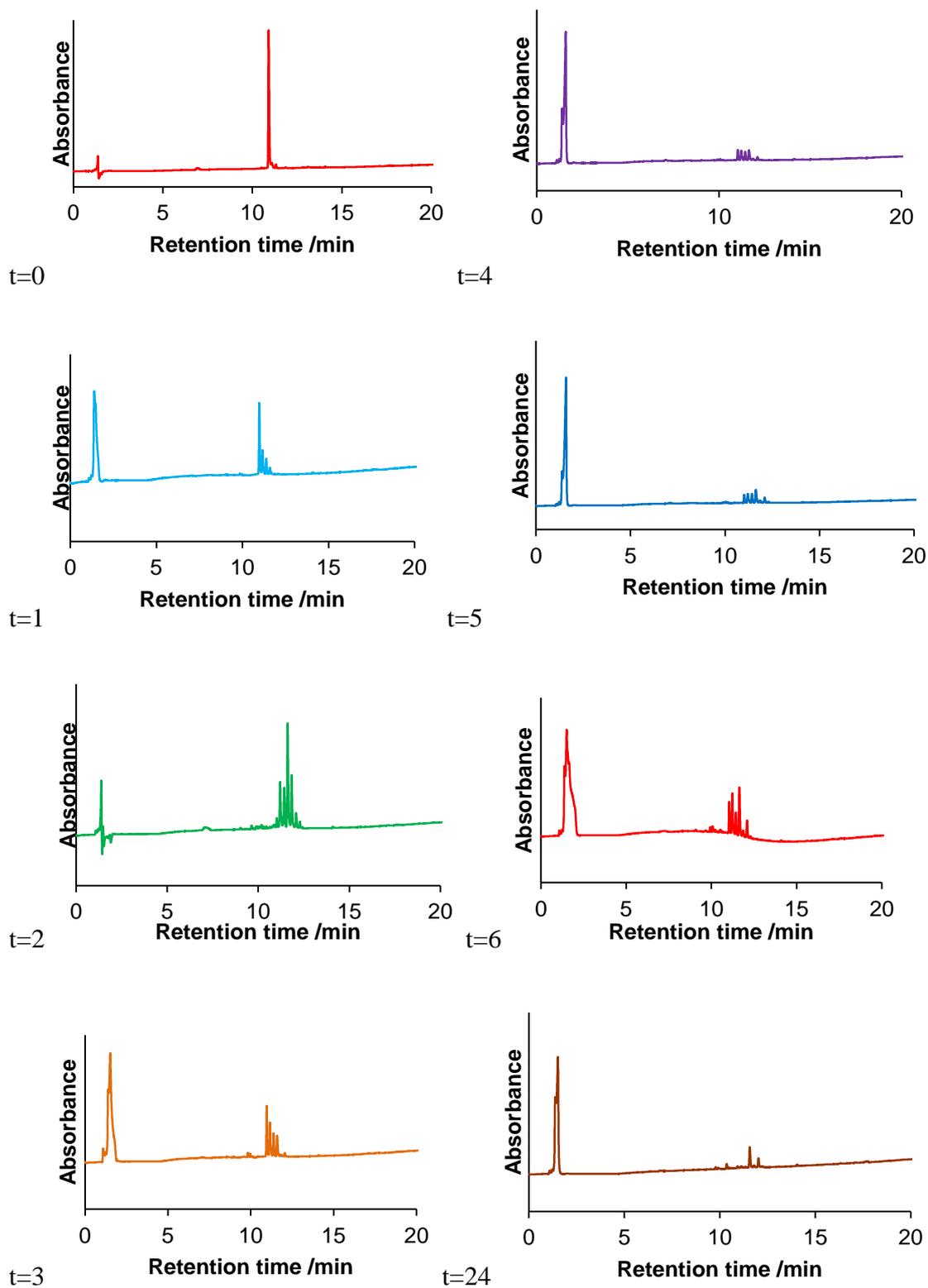
Figure 3.25: Analytical HPLC traces of enzymatic breakdown products of native OT (**209**) taken over a 24 h period.

Entry	Time	Mass (as observed in Q-ToF)	Cleavage	Main product
1	t = 0	1008.2	-	No deletion- [M+H] 1008.2 (**209**)
2	t = 1	1008.2, 951.5, 934.2, 820.1	G, L, P	[M+H] 1008.2
3	t = 2	1008.2, 951.1, 934.2, 821.1, 724, 644	G, L, P, C	[M+H] 1008.2
4	t = 3	1008.2, 951.1, 934.2, 821.1, 724, 644	G, L (side chain), L, P, C	[M+H] 1008.2
5	t = 4	1008.2, 951.1, 934.2, 821.1, 724.5, 567.2	G, L (side chain), L, P, C, C	[M-Gly] 952.5 (**238**)
6	t = 5	1008.2, 951.1, 934.2, 821.1, 724.5	G, L (side chain), L, P, C,	[M-Gly] 951.1
7	t = 6	1008.2, 951.1, 840, 725.5, 639.7	G, L, P, C, N(NH_2)	[M-Gly-Leu-Pro-Cys] 639.7 (**239**)
8	t = 24	408.3, 397.5, 291.7, 186.8, 146.6	Significant degradation of peptide to small fragments	[M-Gly-Leu-Pro-Cys-Asn-Gln] 397.5 (**240**)

Table 3.6: Deletion products resulting from enzymatic degradation of native OT (**209**).

Shown in **Figure 3.25** are the analytical HPLC traces from the enzymatic degradation of the native OT (**209**) with chymotrypsin. At t = 0 there is one single peak corresponding to the native OT (**209**) which was confirmed by LCMS. At t = 1 there is a total of 4 peaks- one corresponding to the native OT, which is the most abundant and other that corresponds to a Gly deletion (**238**) of *m/z* 951.1, Gly + Leu deletion *m/z* 934.2, Gly+ Leu + Pro deletion *m/z* 820.1. From t = 0 to t = 3 the main product is the native OT analogue but the signal corresponding to this mass decreases over time until t = 4, where the most abundant peak becomes the Gly deletion (**238**) (*m/z* 934.2). At t = 6 the most abundant peak then becomes the Gly+ Leu + Pro + Cys + Asn (NH$_2$) fragment (**239**) (*m/z* 408.3) with large amounts of other deletion products. After 24 h, the initial peptide has completely degraded to small peptide fragments (**240**) of under *m/z* 500.0.

To ensure that the degradation observed was enzymatic and not chemical, a control reaction was run without enzyme. The analytical HPLC traces resulting from this study at t = 0 and t = 24 are shown in **Figure 3.26**.

t=0

t=24
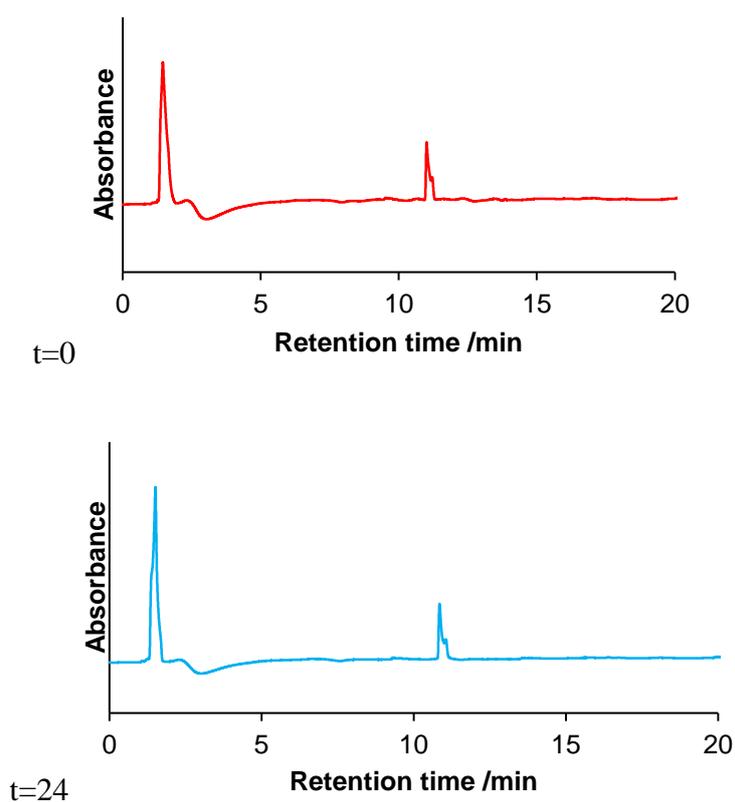
Figure 3.26: Analytical HPLC traces of the non-enzymatic degradation experiment of native OT (**209**) at t = 0 and t = 24 (UV wavelength 220 nm).

This experiment demonstrated that over the course of 24 h there was no observable degradation of native OT (**209**). This suggests that the degradation observed in **Figure 3.25** is purely due to enzymatic degradation and not any form of chemical degradation.

After degradation of the native OT (**209**) was carried out, the singular PFP-tagged OT analogue (**220**) underwent the same enzymatic degradation. Samples were removed and analysed using analytical HPLC and the traces of samples t = 0, 1, 2, 3, 4, 5, 6 and 24 are show in **Figure 3.27**.

t=0

t=1

t=2

t=3

t=4

t=5

t=6

t=24

Figure 3.27: Analytical HPLC traces of enzymatic breakdown products of singular PFP-tagged OT (**220**) (UV wavelength 220 nm) taken over a 24 h period.

Entry	Time	Mass (as observed in Q-ToF)	Cleavage	Structure
1	t = 0	1156.1	-	[M+H] 1156.1 (**220**)
2	t = 1	1156.1, 1100.0	G	[M-Gly] 1100.0 (**241**)
3	t = 2	1156.1, 1100.0	G	
4	t = 3	1156.1, 1100.0	G	
5	t = 4	1156.1, 1100.0	G	
6	t = 5	1156.1, 1100.0	G	
7	t = 6	1156.1, 1100.0	G	
8	t = 24	1100.0	G	

Table 3.7: Deletion products resulting from enzymatic degradation of singularly tagged OT (**220**).

Although over time the singularly tagged OT compound (**220**) is degraded to the Gly cleaved peptide (**241**), the degree of fragmentation is decreased compared to the native OT. In the enzymatic degradation of native OT (**209**), multiple peaks were observed in the HPLC- corresponding to multiple deletions- from t = 1 onwards whereas the PFP-tagged OT HPLC traces show 2 peaks throughout the experiment until full conversion to the Gly cleaved peptide at t = 24 (e.g. compare **Figure 3.25** (t = 24) and **Figure 3.27** (t = 24)). This is an improvement of stability of the tagged peptide to enzymatic degradation compared to native OT as the incorporation of PFP into OT has stabilised degradation of the cyclic moiety of OT.

Shown in **Figure 3.28** are the analytical HPLC traces from the enzymatic degradation of the doubly PFP-tagged OT (**221**) with chymotrypsin. At t = 0 there is a single peak corresponding to **221** and the degradation products that arise over 24 hrs incubation are shown in **Table 3.8**, as analysed by LCMS.

t=0

t=4

t=1

t=5

t=2

t=6

t=3

t=24

Figure 3.28: Analytical HPLC traces of enzymatic breakdown products of double PFP-tagged OT (**221**) (UV wavelength 220 nm) taken over a 24 h period.

Entry	Time	Mass (as observed in Q-ToF)	Cleavage	Structure
1	t = 0	1305.2	-	[M+H] 1305.2
2	t = 1	1305.2	-	
3	t = 2	1305.2	-	
4	t = 3	1305.2	-	
5	t = 4	1305.2	-	
6	t = 5	1305.2	-	
7	t = 6	1305.2	-	(**221**)
8	t = 24	1100.1	G and PFP	[M-Gly-PFP] 1100.1 (**242**)
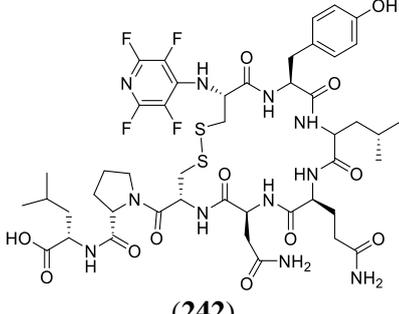
Table 3.8: Deletion products resulting from enzymatic degradation of doubly tagged OT (**221**).

From analyses of the analytical HPLC, it is evident that the double PFP-tagged OT (**221**) is the more stable of the two analogues synthesised. Over the first 6 h, the peptide did not undergo any noticeable degradation, with intact **221** being the major product present in the LCMS analysis. After 24 h there was degradation and the peptide converted to the Gly + PFP deletion product (**242**). This is considerably more stable than the native OT and even the single PFP-tagged analogue.

3.3.3 Stability Studies of Vasopressin and Pentafluoropyridine-Tagged Analogues

VP is structurally similar to OT with two nucleophilic sites available for reaction with PFP. Native VP (**210**) has a half-life of 10-35 min and is degraded *via* vasopressinases in the human body.[17] VP has a longer half-life than OT which may be due to the proximity of the Tyr and Phe which prevents chymotrypsin degradation of that site. The method used for the enzymatic degradation studies of VP was the same used for the studies of OT (**Section 3.3.1**).

Figure 3.29 shows the analytical HPLC traces of the enzymatic degradation of native VP and **Table 3.9** describes the LCMS data of these samples and suggested degradation structures. At t = 0 there is one peak which was determined by LCMS to be the native VP with a mass of 1085.3.
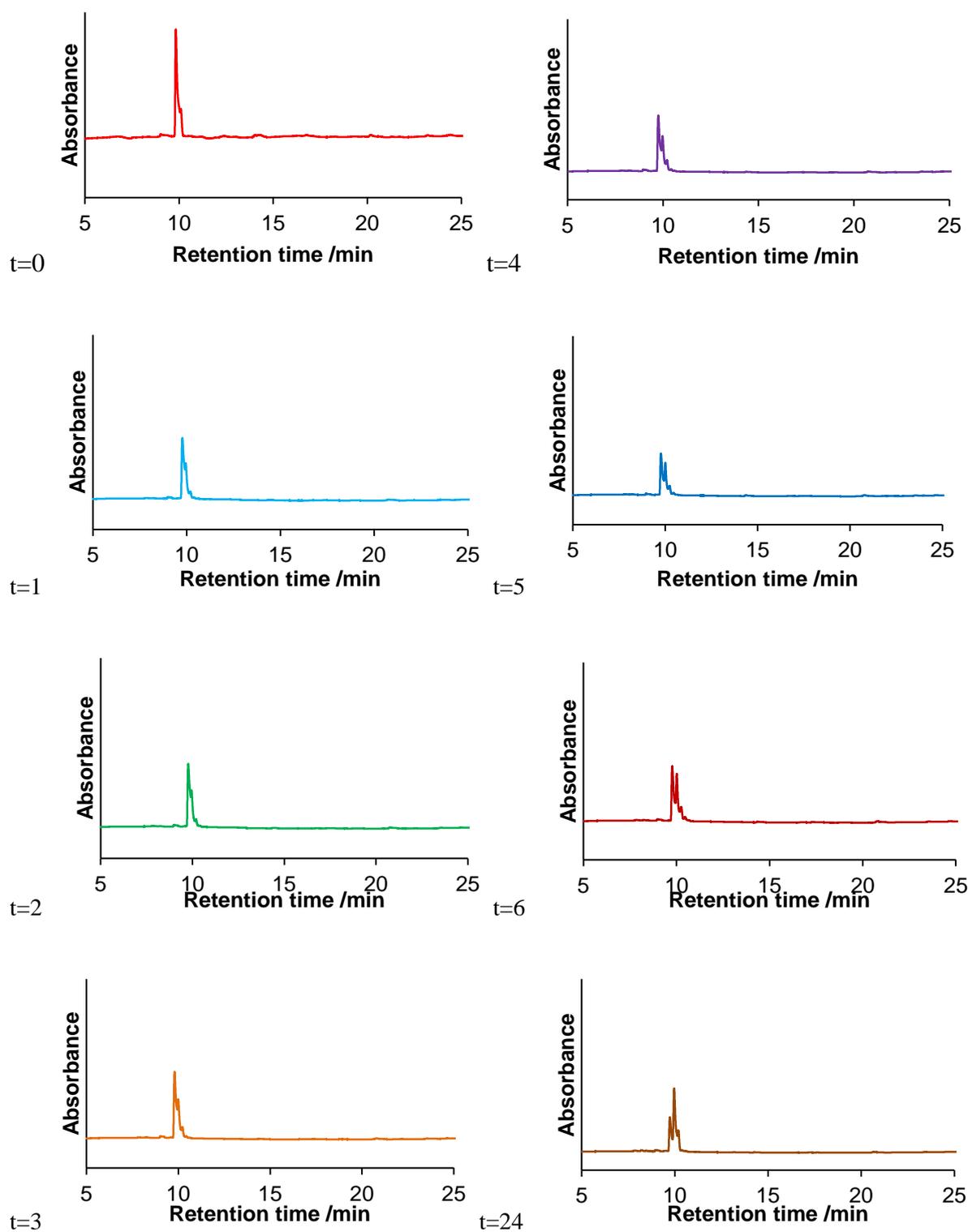
Figure 3.29: Analytical HPLC traces of enzymatic breakdown products of native VP (**210**) (UV wavelength 220 nm) taken over a 24 h period.

Entry	Time	Mass (as observed in Q-ToF)	Cleavage	Main product
1	t = 0	1085.3	-	[M+H] 1085.3 (**210**)
2	t = 1	1085.3, 1028.1	G	[M+H] 1085.3
3	t = 2	1085.3, 1028.1	G	[M+H] 1085.3
4	t = 3	108.35, 1028.1	G	[M+H] 1085.3
5	t = 4	1085.3, 1028.1, 956.2	G	[M+H] 1085.3
6	t = 5	1085.3, 1028.1, 956.2, 824.5	G, R (Side cahin)	[M+H] 1085.3
7	t = 6	1085.3, 1028.1, 956.2, 824.5, 757.7	G, R, P,	[M-Gly] 1028.1 (**243**)
8	t = 24	1085.3, 1028.1, 956.2, 757.7, 704.1	G, R, P, C	[M-Gly] 1028.1

Table 3.9: Deletion products resulting from enzymatic degradation of native VP (**210**).

At t = 0 there is one major peak corresponding to the native peptide (**210**) [M+H]$^+$ 1085.3 and this remains the major product until t = 6 where the major product becomes the Gly deletion (**243**) (*m/z* 1028.1). From t = 1 there is only 2 degradation products until t = 6 where there are now 5 other degradation products. Even though native VP has a half-life which is double that of OT, efforts were still made to try and improve the stability of native VP.

Again, to confirm that the degradation observed in this study was due to enzymatic degradation and not chemical degradation, a reaction void of enzyme was carried out. The same reaction conditions were used as discussed previously but the enzyme was not added. The analytical HPLC traces resulting from this study at t = 0 and t = 24 are shown in **Figure 3.30**.
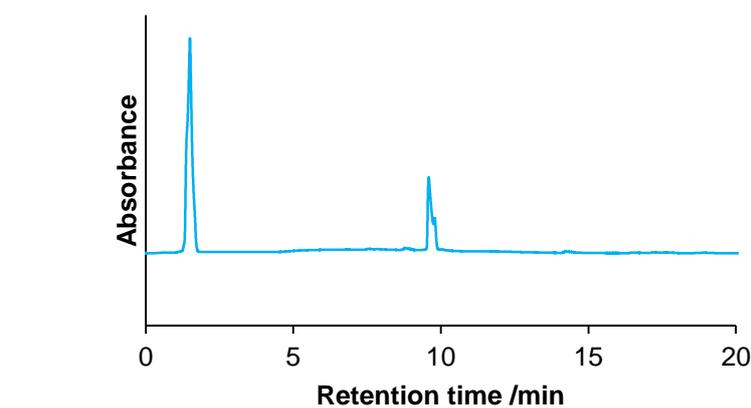
t = 0
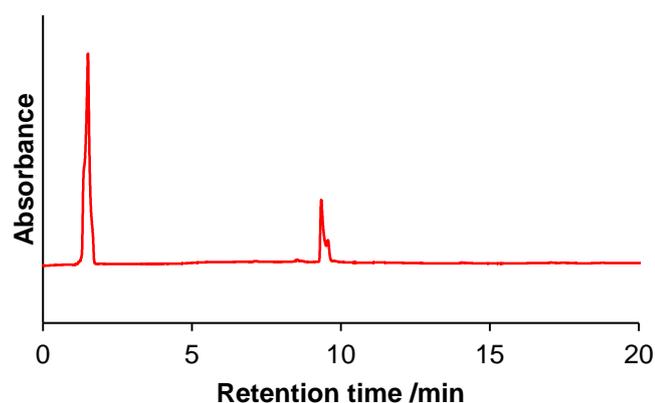
t = 24

Figure 3.30: Analytical HPLC traces of the non-enzymatic degradation experiment of native VP (**210**) at t = 0 and t = 24 (UV wavelength 220 nm).

Figure 3.30 illustrates that over the course of 24 h there was no observable degradation of the native VP (**210**) in the absence of chymotrypsin, suggesting that any degradation observed in **Figure 3.29** is due to enzymatic degradation and not chemical degradation.

Figure 3.31 shows the analytical HPLC data from the enzymatic degradation of single PFP-tagged VP (**222**). At t = 0 there is one single peak which LCMS proved was the singular PFP-tagged VP (**222**) with a mass of 1233.3. **Figure 3.31** shows the analytical HPLC traces of the enzymatic degradation and **Table 3.10** describes the LCMS data of these samples and suggested degradation structures.
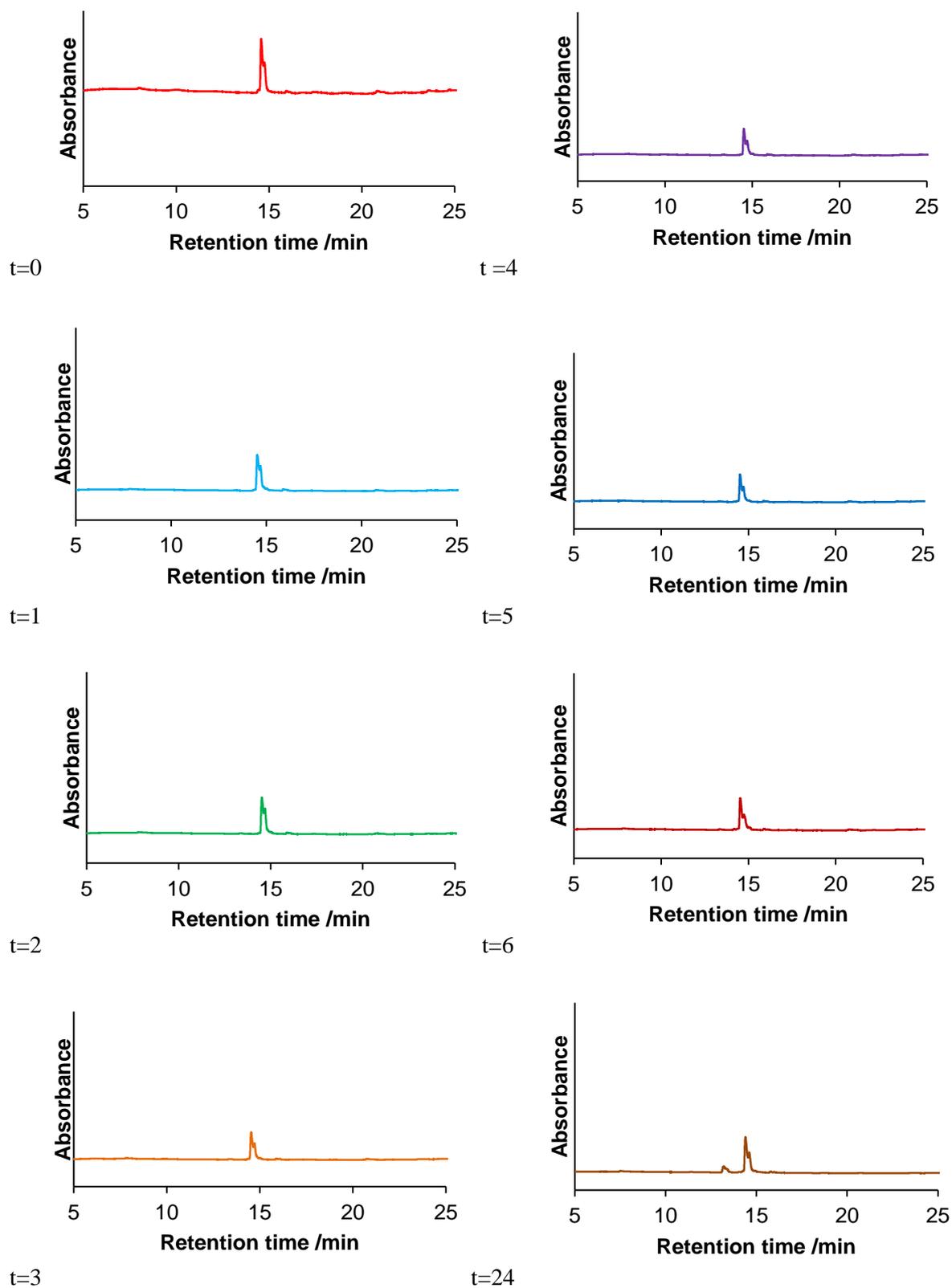
Figure 3.31: Analytical HPLC traces of enzymatic breakdown products of singular tagged VP (**222**) (UV wavelength 220 nm) taken over a 24 h period.

150

Entry	Time	Mass (as observed in Q-ToF)	Cleavage	Main product
1	t = 0	1233.3	-	[M+H] 1233.3 (**222**)
2	t = 1	1233.3, 1177.4	G	[M+H] 1232.3
3	t = 2	1233.3, 1177.4	G	[M+H] 1232.3
4	t = 3	1233.3, 1177.4	G	[M+H] 1232.3
5	t = 4	1233.3, 1177.4	G	[M+H] 1232.3
6	t = 5	1233.3, 1177.4	G	[M+H] 1232.3
7	t = 6	1233.3, 1177.4	G	[M+H] 1232.3
8	t = 24	1233.3, 1177.4	G	[M+H] 1232.3

Table 3.10: Deletion products resulting from enzymatic degradation of singular PFP-tagged VP (**222**).

At t = 0 there is one major peak corresponding to singular PFP-tagged VP (**222**) and remains the major product for the full 24 h. At t = 2 a second peak appears which corresponds to the Gly deletion of *m/z* 1177.4, but this remains a minor side product throughout the enzymatic degradation study. Compared to native VP, the singular PFP-tagged VP analogue is significantly more stable to enzymatic degradation.

Figure 3.32 shows the analytical HPLC traces from the enzymatic degradation of double PFP-tagged VP (**223**). At t = 0 there is a singular peak corresponding to the whole double tagged peptide which was confirmed *via* LCMS to have a mass of 1382.1. **Figure 3.32** shows the analytical HPLC data from the degradation study and **Table 3.11** lists the LCMS data of these samples as well as suggested fragment structures.
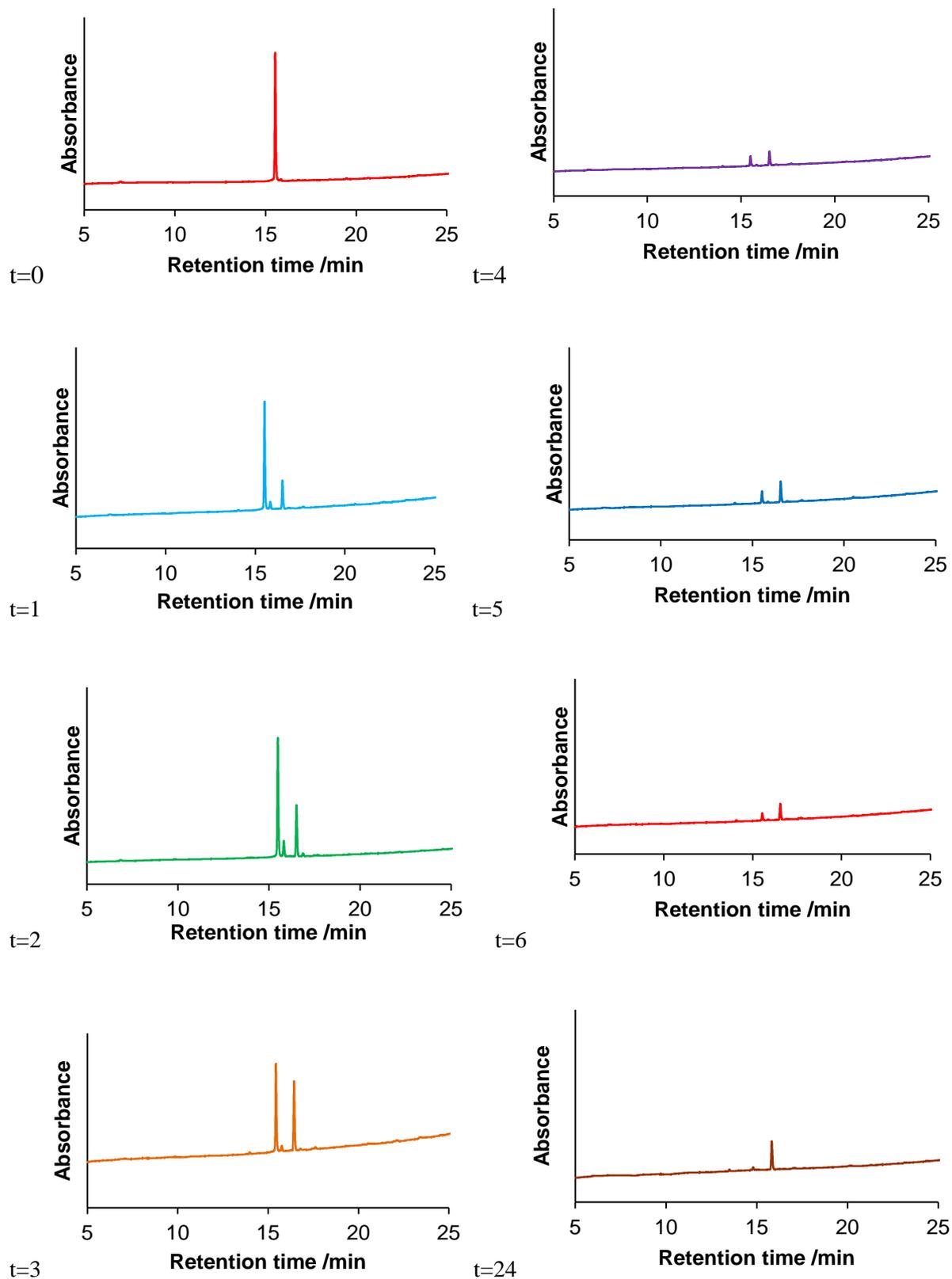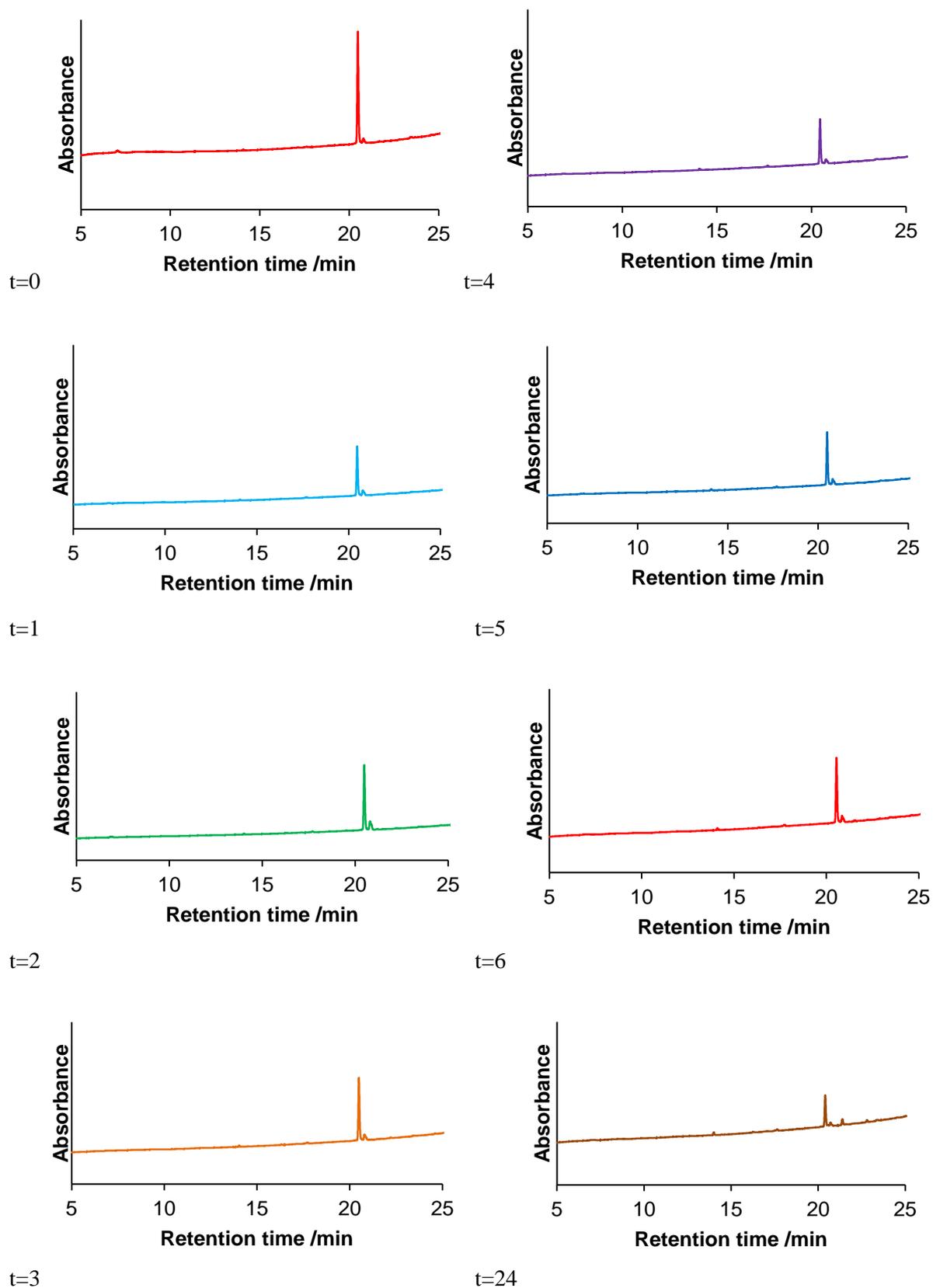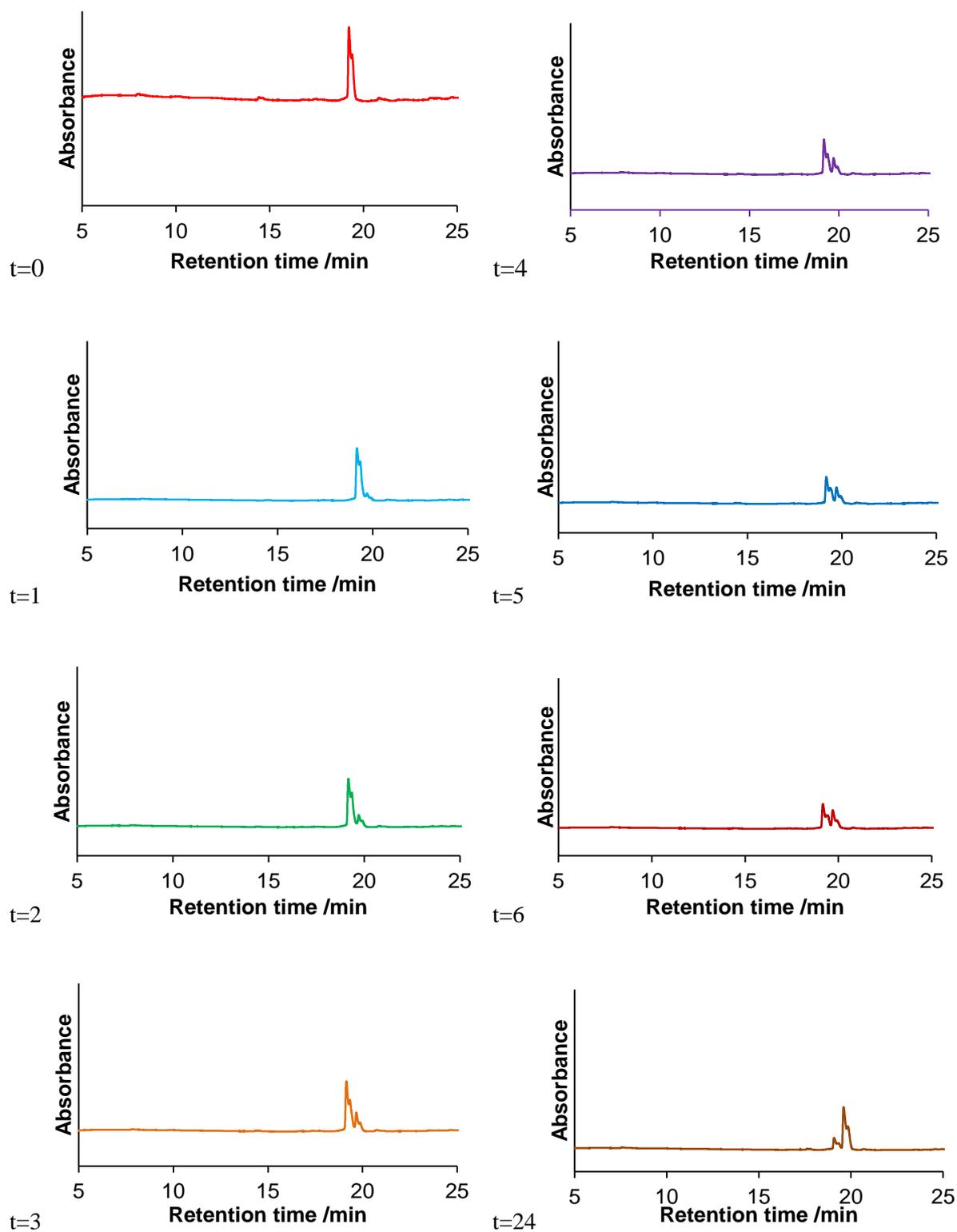
Figure 3.32: Analytical HPLC traces of enzymatic breakdown products of double tagged VP (**223**) (UV wavelength 220 nm) taken over a 24 h period.
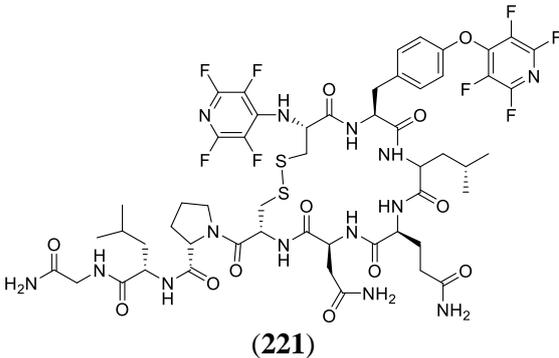
Entry	Time	Mass (as observed in Q-ToF)	Cleavage	Main product
1	t = 0	1382.1	-	[M+H] 1382.1 (**223**)
2	t = 1	1382.1	-	[M+H] 1382.1
3	t = 2	1382.1, 1326.0	G	[M+H] 1382.1
4	t = 3	1382.1, 1326.0	G	[M+H] 1382.1
5	t = 4	1382.1, 1326.0	G	[M+H] 1382.1
6	t = 5	1382.1, 1326.0	G	[M+H] 1382.1
7	t = 6	1382.1, 1326.0	G	[M+H] 1382.1
8	t = 24	1326.0	G	[M-Gly] 1326.0 (**244**)
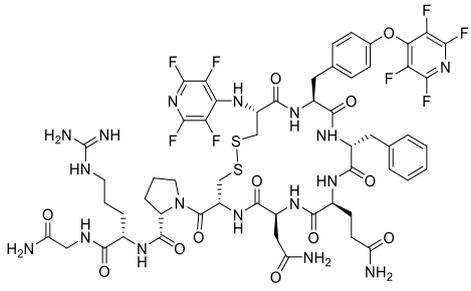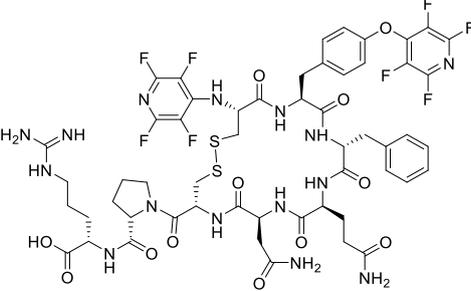
Table 3.11: Deletion products resulting from the enzymatic degradation of doubly PFP-tagged VP (**223**).

At t = 0 and t = 1 there is a single peak corresponding to the mass 1382 and over the course of 24 h there is full conversion from the intact double-tagged peptide (**223**) to the Gly deletion product (*m/z* 1326.0) (**244**). At t = 2 there is the appearance of the Gly deletion product (**244**) which gradually increases over time as the intact double tagged peptide decreases until full degradation at t = 24.

153

It is possible to hypothesise that the double-tagged VP analogue (**223**) is more stable to enzymatic degradation than the native VP but is not as stable as the singular tagged analogue since there was no native peptide present at t = 24, whereas in the study of the singular-tagged peptide **222** the intact peptide *m/z* 1233.3 remained the major product throughout.

3.4 Conclusions

The synthesis of native OT (**209**) and VP (**210**) was carried out using manual Fmoc SPPS. PFP tagged analogues of OT (**220** and **221**), and VP (**222** and **223**) were also prepared. For the synthesis of the singularly PFP tagged analogues **220** and **222**, an on-resin tagging strategy was utilised. By carrying out the tagging reaction on resin, the *t*-Bu protecting group of the Tyr could be exploited to mask the nucleophilic phenol leaving only the amine of the *N*-terminus available for reaction with PFP (**Scheme 3.6**).
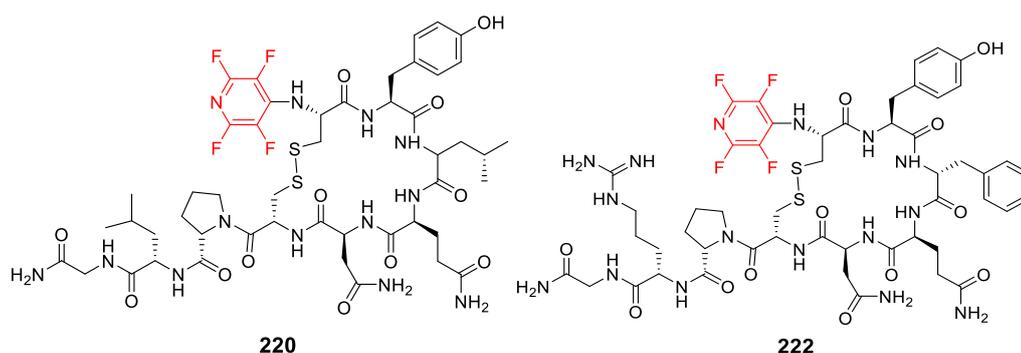
Figure 3.33: Single-tagged OT (**220**) and VP (**222**).

220 and **222** were purified *via* semi-preparative HPLC and analysed using LCMS. Analysis of the MS-MS data confirmed that the PFP was attached to the *N*-terminal NH$_2$ group and not at any other site. ^{19}F NMR analysis also confirmed that the NH$_2$ attacks PFP at the *para* position for the OT analogue (**220**) (**Figure 3.13**, -96.1 ppm and -161.9 ppm) and the VP analogue (**222**) (**Figure 3.22**, -96.2 ppm and -161.9 ppm).

Synthesis of the doubly PFP-tagged peptides (**221** and **223**) was carried out using solution phase tagging conditions. By adopting this technique, PFP (**129**) was able to attach to both the NH$_2$ of the *N*-terminus and hydroxyl group of Tyr (**Scheme 3.8**).
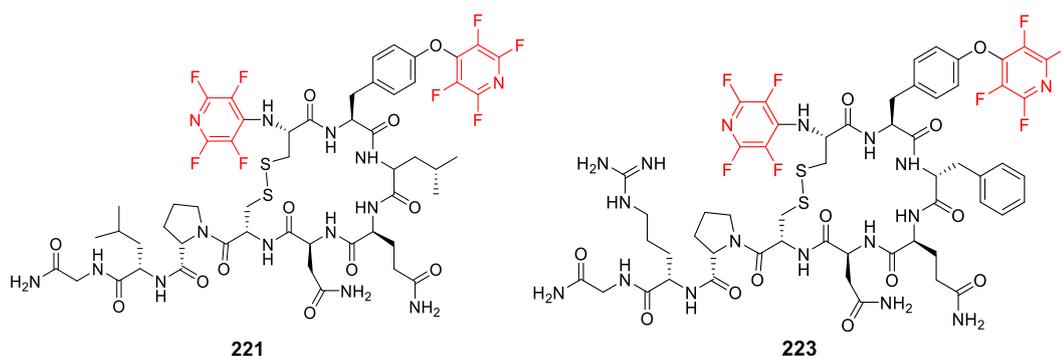
Figure 3.34: Double-tagged OT (**221**) and VP (**223**).

The peptides were, again, analysed using MS-MS and ^{19}F NMR. The MS-MS data illustrated that PFP was now attached to the two positions hypothesised and was visibly different to the singular-tagged analogue data. ^{19}F NMR also suggested that there were now 2 distinct PFP environments in both the OT analogue (**221**) (**Figure 3.16** -91.4 ppm, -96.3 ppm, -156.0 ppm and -162.9 ppm) and VP analogue (**223**) (**Figure 3.24** -91.3 ppm, -96.2 ppm, -155.9 ppm and -162.0 ppm).

The native OT (**209**) and VP peptides (**210**) were also synthesised to act as control compounds in the enzymatic stability studies. These were synthesised using Fmoc SPPS, purified by preparative HPLC and analysed using LCMS (**Section 3.2.1** and **3.3.3**).
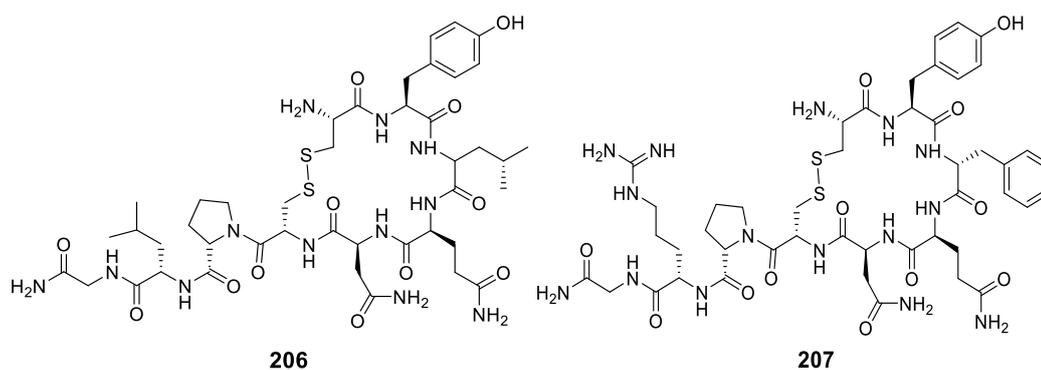
Figure 3.35: Structures of native (**209**) and VP (**210**).

Once synthesised, the enzymatic stability of the PFP-tagged analogues and native peptides were tested. For each analogue, a peptide/enzyme solution was made up and incubated in a water bath for 24 h. Aliquots were removed and the stability of each analogue and analysed using analytical HPLC.

For the OT analogues (**Section 3.3.2**), native OT **209** showed significant degradation within 1 h and over the course of 24 h there was no native OT left in the sample (**Figure 3.25**). For the singular-tagged analogue **220**, stability towards chymotrypsin was significantly improved. After 1 h there was only one side product which remained the only side product (**241**) throughout the study (**Figure 3.27**). For the double-tagged analogue **221**, the stability was further improved as over the course of 6 h there was no degradation of the whole peptide (**Figure 3.28**). In this case, modifying OT with either one or two PFP moieties significantly improves the stability of the peptides to enzymatic breakdown compared to the native peptide.

In the case of the VP analogues (**222** and **223**), a similar trend was observed (**Section 3.3.3**). For the native VP (**210**) there was degradation of the native peptide over the course of 24 h but the native peptide was still present along with approximately another 4 side products (**Figure 3.29**). For the singular-tagged analogue **222**, there was slight degradation with the singular tagged peptide **222** remaining the principle product throughout the course of 24 h (**Figure 3.31**). In the case of the double-tagged species **223**, there was complete degradation of the native peptide to the Gly deletion side product (**244**) over the course of 24 h (**Figure 3.32**). Surprisingly in the case of VP, the singular-tagged species (**222**) was the most stable analogue, as the principle product was the singular-tagged species after 24 h. Although degradation of native VP resulted in 4 side products, the native peptide (**210**) was still present as opposed to the double-tagged species (**223**) which was completely degraded after 24 h. This could be due to fact that chymotrypsin is known to target aromatic side chains and the presence of the PFP-tagged Tyr and Phe side chain promote degradation compared to the native VP or due to poor solubility of the double-tagged peptide in H_2O.

The chemical stabilities of the native peptides were also tested to ensure that any degradation observed was the result of interaction between enzyme and peptide and not chemical degradation occurring over time. For both OT (**209**) and VP (**210**) it was determined that the inclusion of either a singular or multiple PFP moieties into the peptide could positively impact the enzymatic stability compared to the native peptides. Stability tests in human serum of the tagged analogues (**220, 221, 222** and **223**) should be carried out in future to provide a more robust test of stability of the synthesised analogues.

References

[1] Corey E, *Oxytocin.Molecules and Medicine,* 2012, John Wiley & Sons.

[2] G. Gimpl, F. Fahrenholz, *Physiol. Rev.*, 2001, **81**, 629-683.

[3] R. Feldman, A. Weller, O. Zagoory-Sharon, A, Levine, *Psychol. Sci.*, **18**, 965-970.

[4] A. R. Fuchs, F Fuchs, P. Husslein. M. S. Soloff, M. J. Fernstrom, *Science*, 1982, **215**, 1396-1398.

[5] R. J. McQuaid, O. A. McInnis, A. Abizaid, H. Anisman, *Neurosci. Biobehav. R.*, 2014, **45**, 305-322.

[6] J. A. Amico, S. M. Seiff, A. G. Robinson, *J. Clin. Endocrinol. Metab.*, 1981, **52**, 988-993.

[7] A. J. Grippo, D. M. Trahanas, R. R. Zimmerman, S. W. Porges, C. S. Carter, *Psychoneuroendocrino,* 2009, **34**, 1542-1553.

[8] G. Gimpl, F, Farenholz., *Physiol. Rev.*, 2001, **81**, 629-683.

[9] J. Collins, J. Tanka, P. Wilson, K. Kempe, T. P. Davis, M. P. McIntosh, M. R. Whittaker, D. M. Haddleton, *Bioconjugate Chem.*, 2015, **26**, 633-638.

[10] T. Engstrom, T, Barth, P. Melin, H. Vilhardt, *Eur. J. Pharmacol.*, 1998, **355**, 203-210.

[11] W. Kowalczyk, D. Sobolewski, A. Prahl, I. Derdowska, L. Borovickova, J. Slaninova, B. Lammek, *J. Med. Chem.*, 2007, **50**, 2926-2929.

[12] W. Kowalczyk, A. Prahl, I. Derdowska, J. Slaninova, B. Lammek, *J. Med. Chem*, 2004, **47**, 6020-6024.

[13] R. J. Kim, C. Malattia, M. Allen, T. Jr, Moshang, M. Mahgnie, *Pediatr. Endocr. Rev.*, 2004, **6**, 115-123.

[14] A. G. Robinson, *J. Med.*, 1976, **294**, 507-511.

[15] J. S. Cottrell, *J. Proteomics*, 2011, **74**, 1842-1851.

[16] D. Gimenez, C. A. Mooney, A. Dose, G. Sandford, C. R. Coxon, S. L. Cobb, *Org. Biomol. Chem.*, 2017, **15**, 4086-4095.

Chapter 4: Improving the Pharmokinetics of GLP-1 *via* Bionconjugation

4.1 GLP-1 Modifications and their role as Future Therapeutics

There are several challenges in developing peptides as drug molecules. An example of this is the use of Glucagon-like peptide-1 receptor agonists (GLP-1 RA) as treatment against Type 2 diabetes mellitus (T2DM). T2DM is caused by decreased insulin secretion and/or insulin resistance which leads to hyperglycaemia and in long term can lead to heart disease, blindness and amputation.[1] T2DM is caused by a variety of factors including poor diet and an unhealthy lifestyle while some individuals may be genetically predisposed to developing T2DM. With levels of obesity rising dramatically over the last 30 years throughout the world, the levels of T2DM have also risen dramatically.[2] It is estimated by 2030, the number of adults diagnosed with diabetes will rise to 4.6 million with 90% of these being T2DM and thus, there is an urgent need for new and effective treatments.[3]

GLP-1 (**245**) (**Figure 4.1**) is an incretin hormone which binds to the GLP-1 receptor and is released from intestinal cells. Once in circulation GLP-1(**245**) has a very short life ($t_{1/2}$= 1.5min) and undergoes enzymatic degradation by dipeptidyl peptidase IV (DPPIV).

$$\text{His}^7\text{-Ala-Glu-Gly-Thr-Phe-Thr-Ser-Asp-}$$

$$\text{Val-Ser-Ser-Tyr-Glu-Gly-Gln-Ala-Lys}^{26}\text{-Glu-}$$

$$\text{Phe-Ile-Ala-Trp-Leu-Val-Lys}^{34}\text{-Gly-Arg-NH}_2$$

245
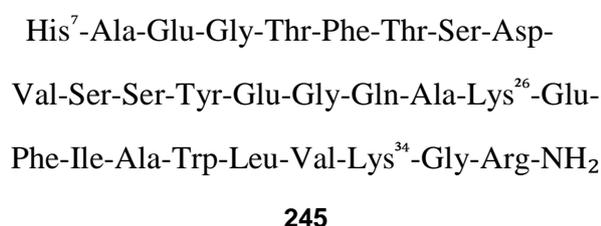
Figure 4.1: Primary Structure of GLP-1 (**245**).[4]

When excreted within the body, GLP-1 modualtes a number of biological functions including increasing secretion of insulin, decreasing secretion of glucagon and limiting glucose absorption into the bloodstream. It is these properties that make it an interesting target molecule for the treatment of T2DM, but due to its short half-life *in vivo* (less than 2 min) developing GLP-1 as a therapeutic is not starightforward.[5]

Improving Pharmokinetics via Bioconjugation

There have been a number of studies focusing on improving the pharmacokinetic properties of GLP-1 including PEGylation of the GLP-1 backbone to increase stability, GLP residue substitution to decrease susceptibility to enzymatic breakdown and bioconjugation to proteins and biopolymers to increase solubility and cell permeability.[6] Due to this and the work of others, there are a number of GLP-1 receptor agonists (GLP-1RA) currently in various stages of clinical trials (**Table 4.1**).[5]

Company sponsoring clinical study	Brand name/Drug code	Molecular type	Status	Treatment
Bristol-Myers Squibb	Exenatide	39 aa-peptidase-resistant peptide	Approved	T2 diabetes, obesity
Novo Nordisk	Liraglutide	31 aa peptide linked to lipid	Approved	T2 diabetes, obesity
Sanofi	Lixisenatide	44 aa peptidase-resistant peptide	Approved	T2 diabetes
GSK	Albiglutide	Tandem repeat of 30 aa peptide	Regulatory review	T2 diabetes
Eli Lilly	Dulaglutide	46 aa peptide fused with IgG4 Fc	Phase III	T2 diabetes
Novo Nordisk	Semaglutide	37 aa acetylated peptide	Phase III	T2 diabetes
PhaseBio Phamaceuticals	PB1023	Peptide fused with ELP biopolymer	Phase II	T2 diabetes
Hanmi Pharmaceuticals	HM11260C, LAPS-Exendin	Exendin-4 analog conjugated to human Ig fragment	Phase I	T2 diabetes
Hoffman-La Roche	RO6811135, MAR709	16 aa pegylated peptide; dual agonist	Phase I	T2 diabetes

Table 4.1: GLP-1 peptide receptor agonists in clinical studies.[5]

PEGylation of peptides was shown to improve various properties including enzymatic stability and absorption.[7] Lee and co-wrokers[8] investigated the PEGylation of GLP-1 and the effect this had on the pharmokinetic properties compared to native GLP-1. PEGylation of GLP-1 is possible at a number of sites within the sequence including the amines of His at position 7 and Lys at postions 26 and 34 (**Figure 4.1**). Two different approaches were used by Lee and co-workers in their PEGylation studies to produce mono-PEGylated GLP-1 analogues, using PEG-SPA to conjugate the mPEG$_{2k}$ to Lys residues and reductive alkylation of His with mPEG$_{2k}$-ALD and NaCNBH$_3$ reduction.[8] These analogues were tested against rat islet cells in a high glucose concentration environment and the rate of insulin secretion was measured to evaluate biological activity. It was found that the PEG-Lys-GLP analogue showed significant insulin secretion, but the PEG-His-GLP analogue did not show significant biological activity. This is due to the fact that the free *N*-terminus of His in GLP is believed to be responsible for important interactions with the GLP-1 receptor and therefore important in biological activity. Conjugation of PEG to the His at the *N*-terminus results in a disruption of ligand-receptor binding and therefore affects biological activity. Stability studies of the PEG-Lys^{26}-GLP analogue (half-life = 770 min) also showed significant resistance to DPPIV (dipeptidyl peptidase IV) breakdown compared to GLP-1 (half-life= 15.5 min). Overall Lee and co-workers were able to synthesise a PEGylated GLP-1 analogue with potent biological activity, increased stability to enzymatic breakdown and an improvement in plasma exposure.[7]

There have been a number of studies that have concentrated on improving the properties of GLP by altering the amino acid sequence. Liraglutide (**246**) (**Figure 4.2**), developed by Novo Nordisk and approved for treatment of T2DM in 2009, is structurally similar to GLP-1 hormone itself and has an Arg substitution at position 34 and palmitic acid joined to Lys *via* a glutamate spacer at position 26.[1]

246
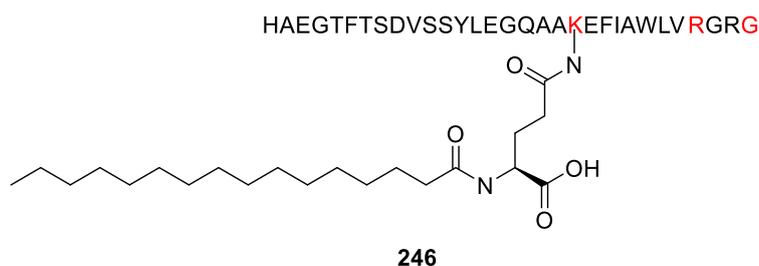
Figure 4.2 Structure of Liraglutide (**246**) substituted residues highlighted in red.

Liraglutide (**246**) exhibits prolonged stability in vivo compared to GLP-1[9] and is administered *via* subcutaneous injection with a half-life of 11-13 h. Once injected the peptide is released *via* self-dissociation and is bound to serum albumin in the bloodstream which increases its stability towards DDPIV breakdown and reduces renal clearance.[1]

Studies have also been carried out where the conjugation of GLP-1 analogues to biological molecules such as human serum albumin and immunoglobulin G (IgG) has been investigated. Albiglutide (**247**), developed by GSK, contains two repeating GLP-1 strands linked to the *N*-terminus of albumin and a Gly substitution at position 2 of the peptide.[1]

247
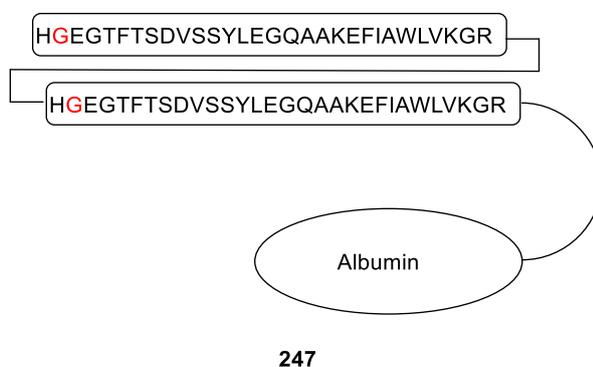
Figure 4.3: Structure of Albiglutide **247.**

The repeating GLP-1 unit in **247** acts as a spacer which improves potency compared to native GLP-1, the Gly substitution of Ala at position 8 allows DDPIV resistance and conjugation to albumin improves stability and half-life due to the increased molecular size that fusion to albumin contributes.[1] Similair to Albiglutide (**247**), Dulaglutide (**248**) (**Figure 4.4**) consists of 2 GLP-1 chains and various residue substitutions which limit enzyme degradation, but differs due to the fact the GLP-1 peptides are joined to a IgG-Fc heavy chain variant. This results in a long half-life and reduced renal clearance. Dulagutide (**248**) has shown superior action to Exanatide (an approved GLP-1RA) in a 6-month phase 3 trial period and has shown great promise as a possible therapy for T2DM treatment.[1]

248
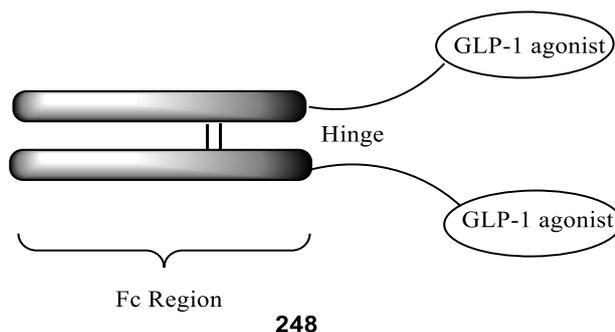
Figure 4.4: Structure of Dulaglutide **248**- GLP-1 analogue linked to Fc region of immunoglobulin (IgG).

Studies have concentrated on improving the pharmacokinetic properties of GLP-1 receptor analogues and a number of these studies have been successful with 3 GLP-1RA currently available and a large number in various stages of development.[5]

Improving Efficacy via Hybridisation of GLP-1

Work by Matthais Tschop and co-workers[10] has concentrated on synthesising a series of GLP-1 co-agonists to improve targeting and reduce possible side effects. One of these co-agonists is a GLP-1/glucagon analogue. Glucagon (GCGR) is structurally related to GLP-1 and has been shown to improve metabolism and lower body weight. A GLP-1/GCGR co-agonist peptide was synthesised and tested in mice and shown to restore leptin responsiveness in obese mice and aid weight loss.[10] Leptin is a hormone released by adipose tissue to regulate energy balance and functions by inhibiting hunger. It has been shown that obese patients exhibit a decreased sensitivity to leptin which results in an inability to control appetite and food consumption.[11] Tschop and co-workers were able to illustrate that co-administration of PEG-GLP-1/GCGR restored leptin responsiveness and promoted weight loss when administered alongside PEG-leptin, in mice models. Another GLP-1 co-agonist which has been recently synthesised is a GLP-1/GIP analogue. GIP, like GLP-1, is an incretin peptide, but functions by promoting insulin secretion.[12]

Using the promising results from the studies into GLP-1, co-agonists with glucagon and GIP, Tschop and co-workers[13] developed a "peptide triagonist"-synthesising a single peptide containing active sequences of GLP-1, glucagon and GIP. GIP was discounted as a therapeutic agent against T2DM due to instability and was initially reported to promote obesity. Recent data, however, have shown that this conclusion was not fully

understood, and in fact using GIP analogues to regulate glucose levels could be a promising area of research for anti-obesity and anti-diabetes drugs.[12] Shown in **Table 4.2** are sequences of GLP-1 (**Entry 1**), glucagon (**Entry 2**) and GIP (**Entry 3**) with the sequence of the triagonist shown in **Entry 4.**[13]

Entry	Name	Sequence
1	GLP-1	HAEGTFTDSVSSYLEGQAAKEFIAWLVKGR
2	Glucagon	HSQGTFTSDYSKYLDSRRAQDFVQWLMNT
3	GIP	YAEGTFISDYSIAMDKIHQQDFNVWLLAQKGKKNDWKHNIT
4	Triagonist	HXQGTFTSDKSKYLDERAAQDFVQWLLGGPSSGAPPPS-NH$_2$

Table 4.2: Sequence of GLP-1 (**Entry1**), glucagon (**Entry 2**), GIP (**Entry 3**) and Triagonist (**Entry 4**)- **X** represents Aib and _K_ represents Lys(γE-C$_{16}$).

From **Table 4.2**, it is evident that the peptides listed are of similar sizes and sequences but are varied enough to ensure that, individually, they are selective for their own receptor. The tri-agonist was shown to be selective for GLP-1 receptors, GIP receptor and glucagon receptors over other GPCRs and more potent than the individual peptides. This work was particularly interesting because it was the first synthesis of a potent anti-obesity drug with the ability to bind with three different metabolic receptors.[13]

Improving Targeting via Hormone Bioconjugation of GLP-1

A method for improving targeting of GLP-1 is *via* conjugating to oestrogen, which aids selective tissue targeting (**Figure 4.5**). Oestrogen is a hormone which has been used previously as a treatment for obesity and T2DM, but its use as a drug has been extremely limited as it promotes tumour growth.

Figure 4.5: Structure of GLP-1-Oestrogen analogue **249.**

Compound **249** functions by exploiting the anti-obesity properties of oestrogen and using GLP-1 as a targeting peptide to allow intracellular transport, therefore eliminating side effects previously observed when oestrogen is used on its own. The GLP-1-estrogen conjugate (**249**) was tested in GLP-1 receptor knockout mice to determine if

the activity observed for the conjugate relied on GLP-1 targeting. It was found that in these mice, no weight loss was observed, but in wild type mice, an 8.5% decrease in body weight was observed. This helped prove that the GLP-1 targeting effect of the conjugate was a key component to its activity.[14] A later study into the GLP-1-estrogen conjugate (**249**) mode of action used single-photon emission commuted tomography imaging (SPECT-imaging). In this study, distribution of the **249** was monitored using SPECT-imaging and it was revealed that the supramammillary nucleus (SUM) area of the brain is the main target site for the conjugate.[15]

Despite the advances that have been made there are still challenges in synthesising novel stable GLP-1 analogues. As discussed, previous efforts have concentrated on modification of the GLP-1 backbone *via* Lys. Somewhat surprisingly, there have been no studies to date that have looked at the incorporation of Cys residues into the GLP-1 backbone and its use as a chemical handle for modification. Future work into the modification of GLP-1 could concentrate on further improving the bioavailability of GLP-1 analogues to possibly reduce dosage or introduce an orally available GLP-1 analogue.

4.1.1   Aims

As mentioned in **Section 4.1**, the GLP-1 receptor is an interesting pharmaceutical target with a number of new GLP-1 receptor agonists in various stages of development (**Table 4.1**). Comparisons between the native GLP-1 (**245**) and modified analogues currently being studied show that the main area of modification is at Lys[26]- as in Liraglutide (**245**).

245
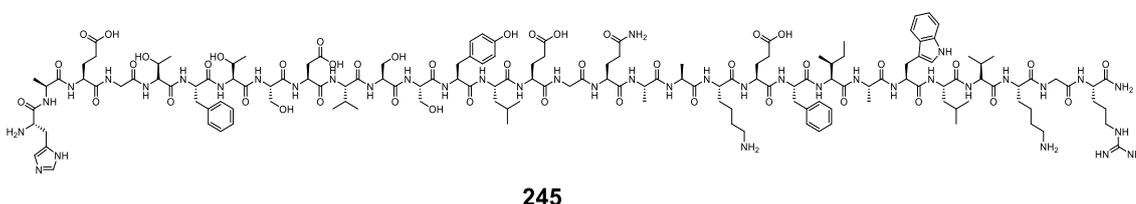
Figure 4.6: Native GLP-1(7-36) (**245**) sequence.

The native GLP-1(7-36) **245** sequence will be synthesised to act as a control compound in enzymatic degradation studies of the PEG-modified GLP-1 analogue **252**. Comparing the degradation of the native sequence and PEG-modified analogue will illustrate the effectiveness of PEG to decrease enzymatic degradation.

The aim of this aspect of the project was to synthesise a Cys-modified GLP-1 analogue **250** which could be subsequently tagged with pentafluoropyridine (PFP) resulting in an activated GLP-1(7-36) analogue **251**.

This activated GLP-1 analogue (**251**) could undergo a displacement reaction with thio-PEG or thiosugar analogues, resulting in a PEG-modified GLP-1 analogue **252**. It is hypothesised that the modified GLP-1 analogue **252** will exhibit improved stability and solubility compared to the native GLP-1 analogue **245**.
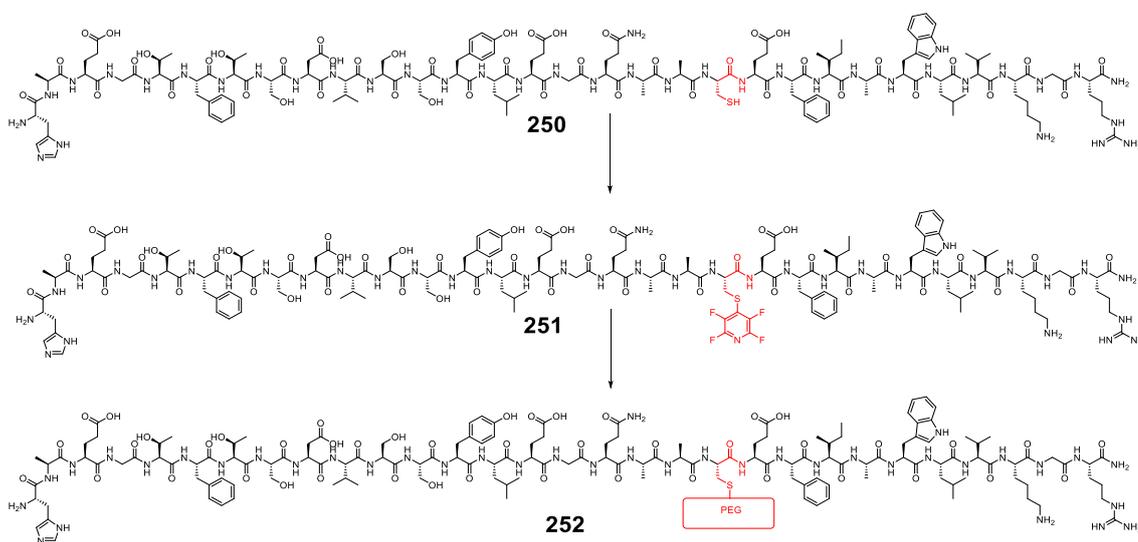
Scheme 4.1: PEG-modified Cys GLP-1(7-36) **252**.

4.2 Synthesis of Native GLP-1(7-36) (245)

GLP-1(7-36) **245** was synthesised *via* automated Fmoc solid phase peptide synthesis (SPPS) using a Liberty1 automated microwave synthesiser. GLP-1(15-36) **253** was initially synthesised on RINK amide resin (0.1 mmol scale) using 30 min single couplings rt + 10 min mW and the remaining amino acids were coupled using single 60 min couplings + 10 min mW to give the target peptide **245** (**Section 6.3.2**).
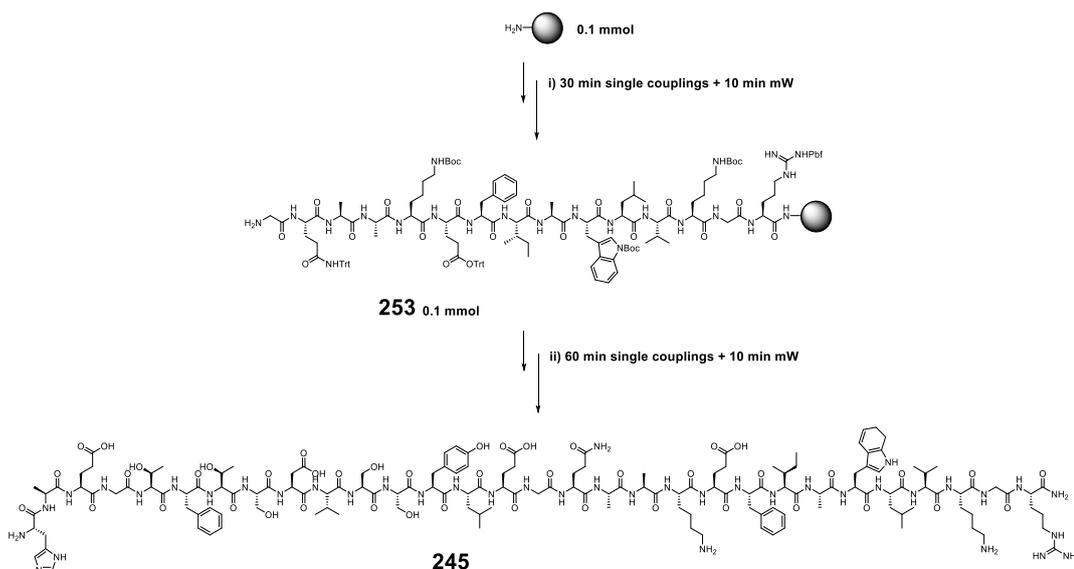
Scheme 4.2: Automated SPPS of GLP-1(7-36) **245** synthesised on 0.1 mmol scale using i) 30 min single couplings at rt + 10min mW ii) 60 min single couplings rt + 10 min mW.

The resulting crude 30-mer peptide (**245**) was cleaved from resin and was analysed using MALDI-ToF (**Figure 4.7**).
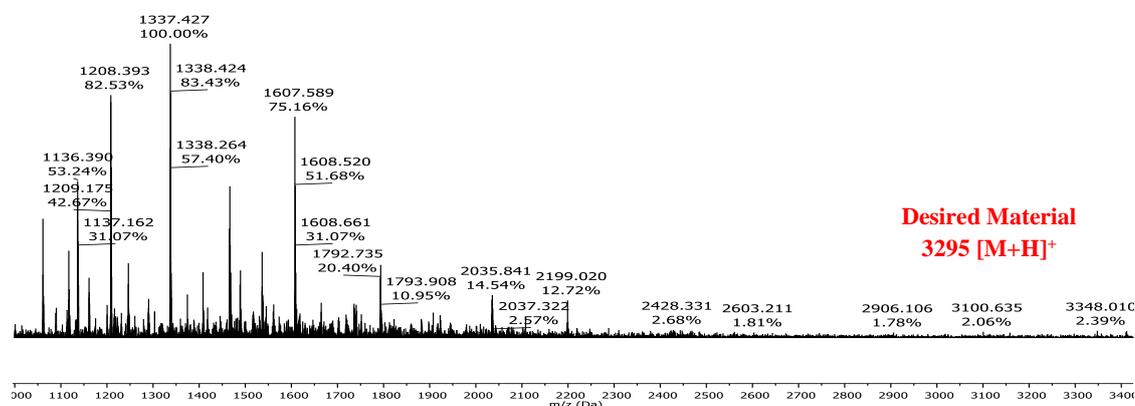
Figure 4.7: MALDI-ToF analysis of peptide (**245**).

As demonstrated in the above spectra, there is no noticeable peak for the desired target peptide **245** at 3295.7 [M+H]$^+$. Analytical HPLC also confirmed that the synthesis had not proceed as expected with several peaks being observed.

It was decided that the peptide would be resynthesised using multiple coupling conditions and longer reaction times. Due to the length of **245** it was decided to synthesise it on a 0.1 mmol scale for the first 10 amino acid residues (GLP-1(27-37)) and then the resin would be split into 2 equal portions and the synthesis of the last 20 amino acids would be carried out on a 0.05 mmol scale. The advantage of using this method is that aggregation of the peptide on resin is minimised, resulting in a homogenous peptide synthesis.

Peptide **254** (EFIAWLVKGR-amide) was synthesised using the conditions described in **Section 6.3.2** on RINK amide resin using 30 min single couplings + 10 min mW for all amino acids, except for Arg (2 × 60 min couplings rt + 10 min mW) and Leu (1 × 60 min coupling + 10 min mW). LCMS of the resulting crude peptide carried out and the desired material 1217.1 [M+H]$^+$ was present. The rest of the synthesis was carried out using approximately 0.05 mmol of resin from the last step using single 60 min coupling rt + 10 min mW.
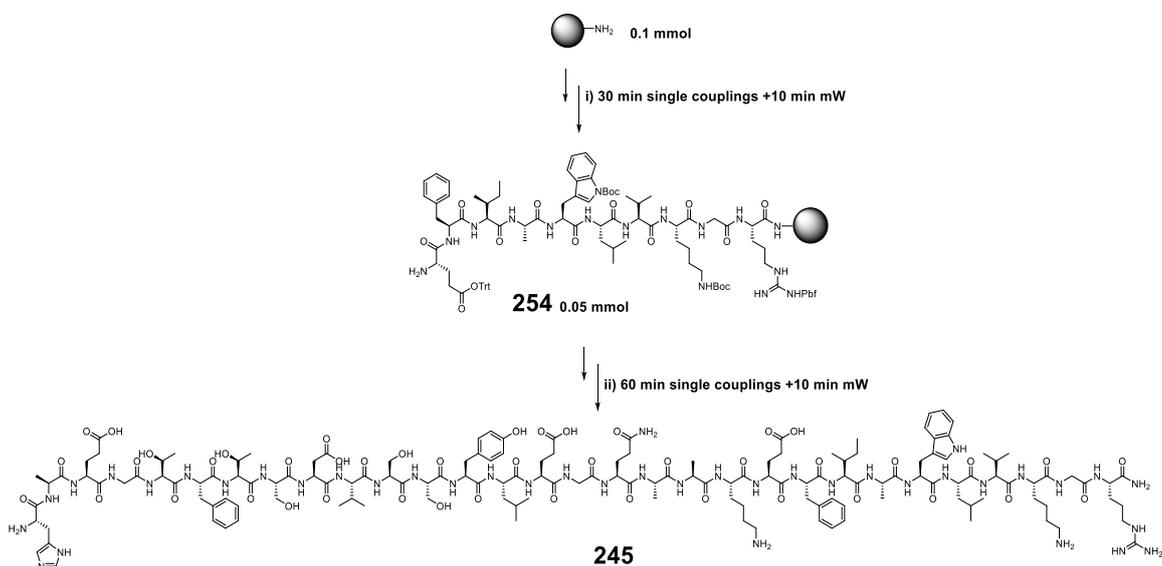
Scheme 4.3: Automated SPPS of GLP-1(7-36) **245** synthesised on 0.1 mmol scale using i) 30 min single couplings at rt +10 min mW and **254** synthesisied on a 0.5 mmol scale using ii) 60 min single couplings rt + 10 min mW.

The resulting crude peptide **245** (HAEGTFTSDVSSYLEGQAAKEFIAWLVKGR-amide) was cleaved from resin and analysed using MALDI-ToF (**Figure 4.8**). The deletion products that have occurred during the synthesis are highlighted in red.
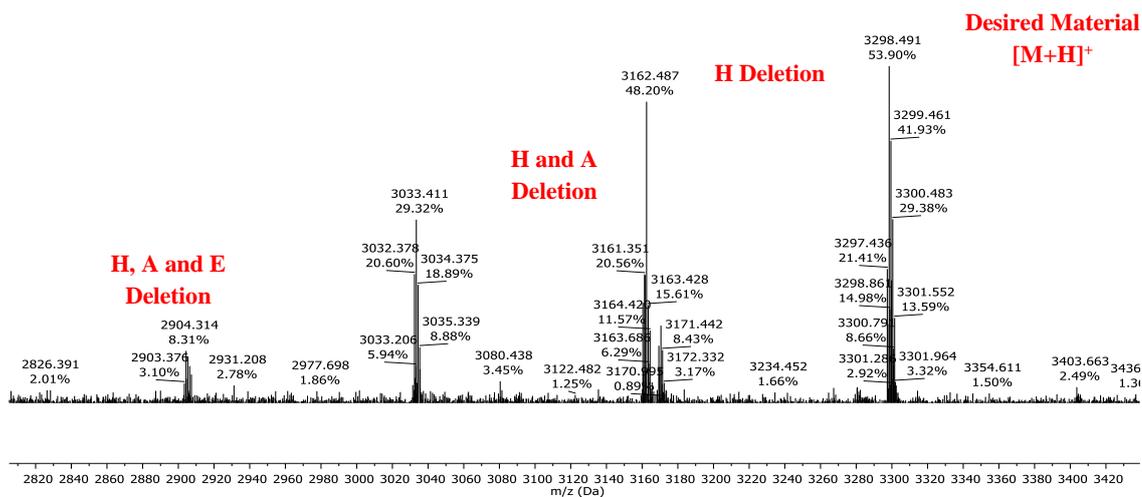
Figure 4.8: MALDI-ToF of crude peptide **245**.

The main deletions observed (*m/z* 2904.3, 3033.4, 3162.5) correspond to the final amino acids of the sequence. To try and eliminate these deletions in the synthesis, double couplings were adopted for the final four amino acids of the sequence (**Scheme 4.4**).
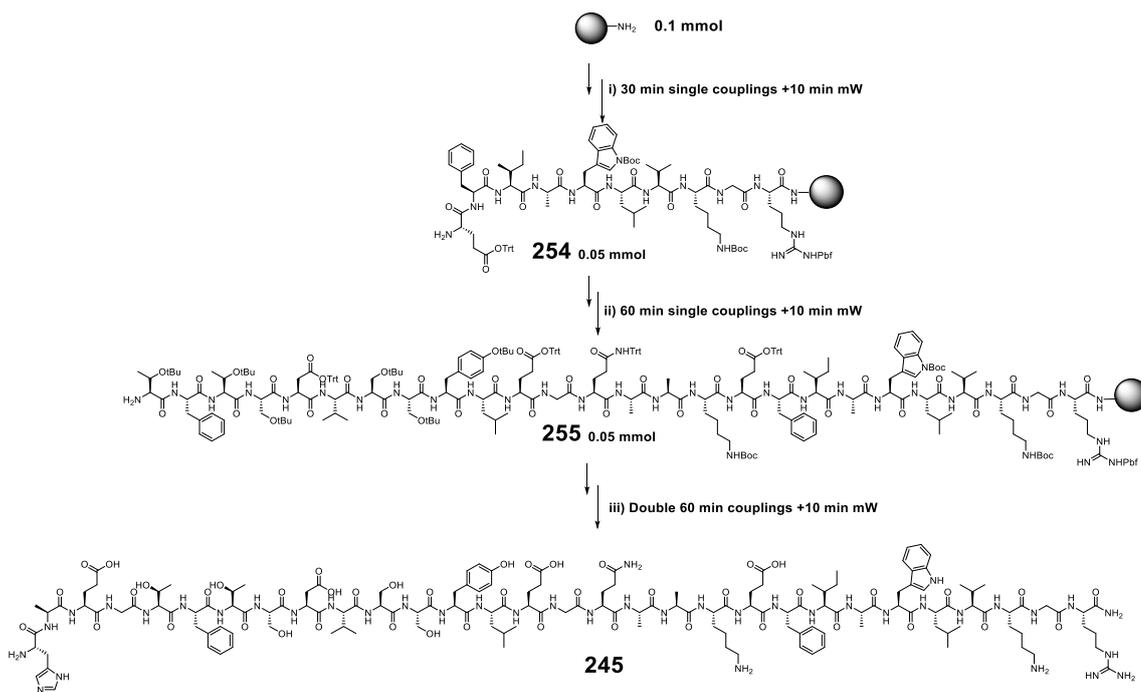
Scheme 4.4: Automated SPPS of GLP-1(7-36) **254** synthesised on 0.1 mmol scale using i) 30 min single couplings at rt +10 min mW, **255** synthesised on a 0.5 mmol scale using ii) 60 min single couplings rt + 10 min mW and **245** synthesised using double 60 min coupling rt + 10 min mW.

The resulting crude peptide **245** (HAEGTFTSDVSSYLEGQAAKEFIAWLVKGR-amide) was cleaved from resin and analysed using MALDI-ToF (**Figure 4.9**).
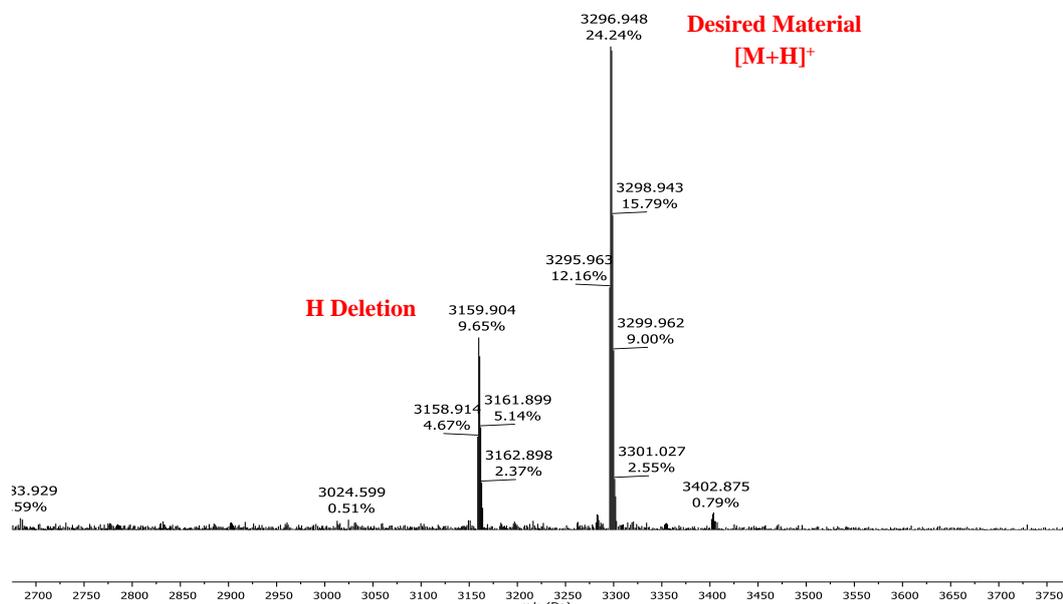
Figure 4.9: MALDI-ToF of peptide **245**.

Figure 4.9 shows the desired material (**245**) 3296.9 [M+H]$^+$ and also a side product corresponding to the His deletion (*m/z* 3159.9) at the *N*-terminus. Upon initial HPLC purification, the His deletion side product could not be successfully separated from the target peptide so alternative solvent gradients were used. Instead of the standard gradient of solvent A (95% H$_2$O, 5% MeCN, 0.01% TFA) 100-50% over 60 min, a more gradual gradient of solvent A 70%- 30% over 60 min was used and the resulting purified peptide (**245**) obtained was analysed using MALDI-ToF and analytical HPLC (**Figure 4.10** and **4.11**).
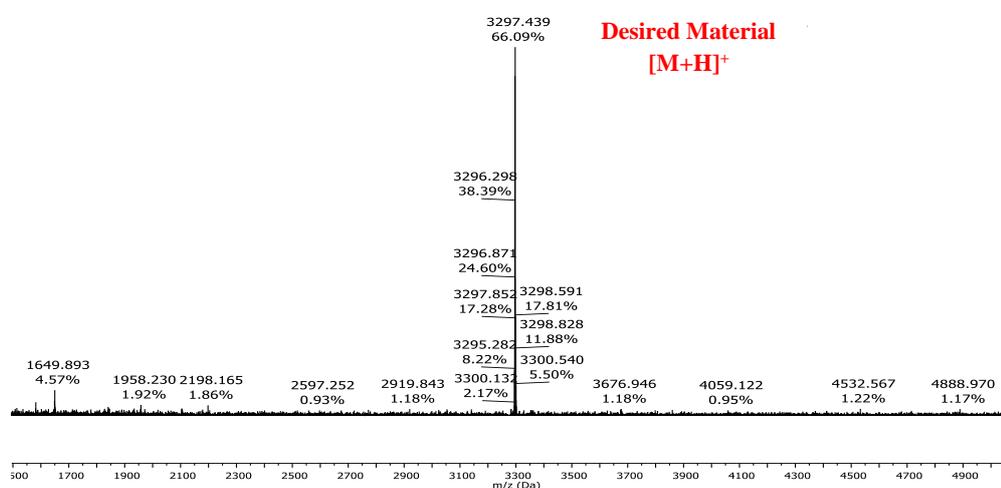
Figure 4.10: MALDI-ToF of **245** (HAEGTFTSDVSSYLEGQAAKEFIAWLVKGR-amide).
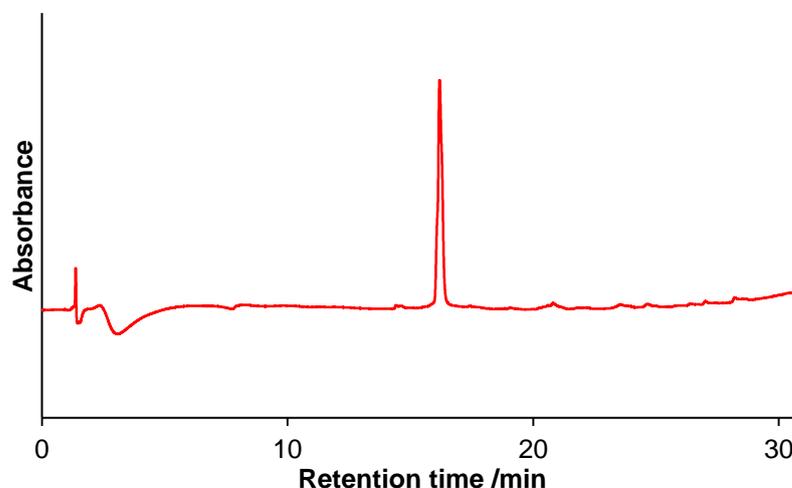
Figure 4.11: Analytical HPLC of native GLP-1 **245** (HAEGTFTSDVSSYLEGQAAKEFIAWLVKGR-amide).

4.3 Synthesis of Cys-modified GLP-1 (250)

With a method for the synthesis of the native GLP-1 developed, the same approach was applied to prepare a Cys modified analogue of GLP-1.

As discussed in **Section 4.1**, it has been shown that conjugation of a peptide to PEG increases its stability towards enzymatic degradation. The objective of this section of the study is to incorporate Cys into position 26 of the native GLP-1 backbone (**250**) and carry out tagging and thiol displacement illustrated in earlier sections (**Chapter 2**).

Figure 4.12: Cys-modified GLP-1(7-36) analogue **250**.

Once the modified GLP-1 sequence was synthesised and displacement reactions with PEG have been successful the modified peptide sequence will be tested for enzymatic breakdown. The stability of the modified sequence will be compared to the unmodified sequence which will be synthesised using automated Fmoc SPPS.

4.3.1 Synthesis of Cys-modified GLP-1 fragments

To demonstrate that the tagging of the Cys-modified GLP-1 could be successful, a short 15-mer fragment of the GLP-1 peptide was first synthesised. The peptide fragment (GQAACEFIAWLVKGR-amide) (**258**) was synthesised on RINK amide resin on a 0.1 mmol scale *via* automated Fmoc SPPS. With the synthesis complete the crude peptide was cleaved from resin and analysed using LCMS (**Scheme 4.5**).

Scheme 4.5: Automated SPPS of **256** synthesised on 0.1 mmol scale using i) 30 min single couplings at rt +10 min mW, **257** using ii) manual HATU coupling of the Fmoc-Cys(Trt)-OH and continuation of the synthesis of **258** using iii) 30 min single couplings at rt +10 min mW followed by acetyl capping of the *N*-terminus using 20% acetic anhydride/DMF.

Peptide **258** was synthesised by adopting 30 min single couplings + 10 min mw and the Cys residue coupled manually using HATU as a coupling reagent (**Section 6.4.2**). Peptide **258** was cleaved from the resin and deprotected using TFA resulting in peptide **259**. The reactivity of the cleaved peptide **259** was investigated and further reacted with PFP.

Scheme 4.6: Reaction of peptide **259** using i) PFP, DIPEA and DMF, 4 h, rt and ii) PFP, DIPEA, TFE, 4 h, rt.

Initial investigation into the reaction of peptide **259** with PFP (**129**), concentrated on using DMF to react PFP unselectively at both the Cys and Lys residues within the peptide sequence (**Scheme 4.6i**). Shown below is the MS of purified peptide **260** with the desired mass of 1970.8 [M+H]$^+$.
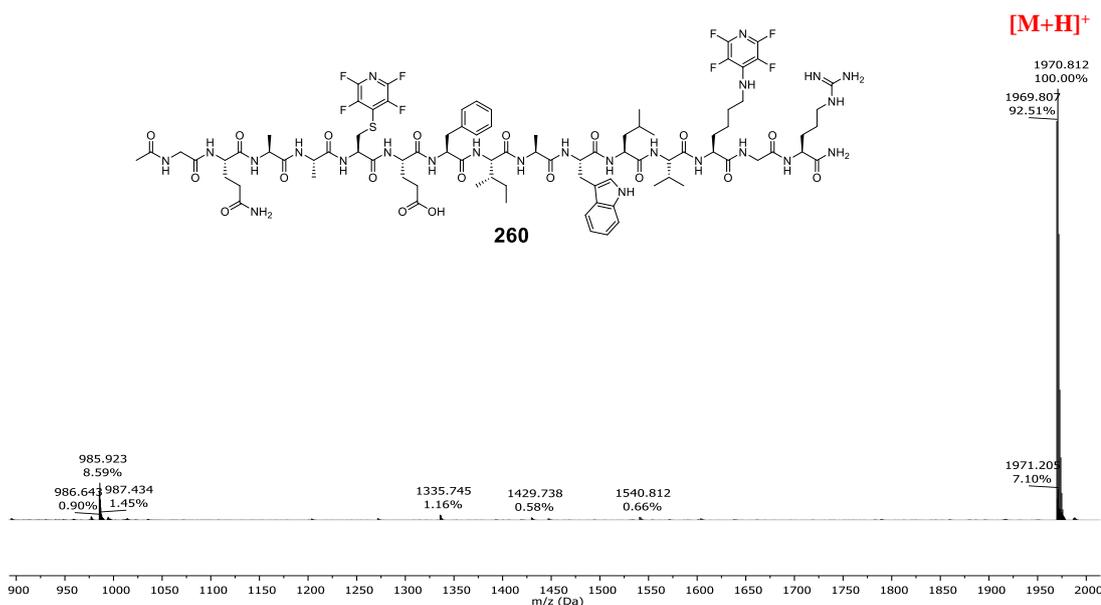
Figure 4.13: MS of purified peptide **260** with [M+H]+ 1970.8.

Selective tagging of peptide **259** with PFP was then carried out using TFE as the solvent (**Scheme 4.6ii**). TFE reduces the reactivity of nucleophilic side chains such as Lys but has a lesser effect on Cys side chains thus allowing the PFP to react selectively.[17] This was confirmed using MS with mass of 1839.8 [M+H]+, as shown in **Figure 4.14.**
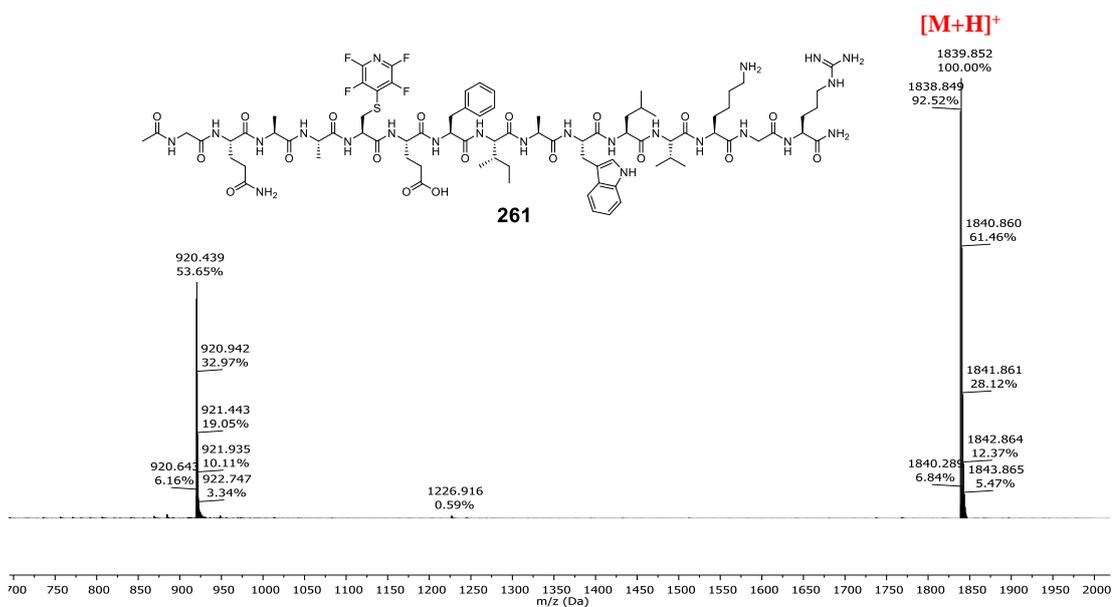
Figure 4.14: MS of purified peptide **261** with [M+H]+ 1839.9.

To confirm the structure and position of PFP tagging MS-MS experiments were carried out on both **260** and **261**. The masses shown in **Table 4.3** confirm the presence of a double PFP-tagged peptide **260** with PFP tagged at both the Lys and Cys.

Entry	Mass	Fragment
1	1971.1	**260** Ac-GQAAC(Pfp)EFIAWLVK(Pfp)GR-amide
2	1764.0	**260** Ac-GQAAC(Pfp)EFIAWLVK(Pfp) GR-amide
3	1082.2	**260** Ac-GQAAC(Pfp)EFIA WLVK(Pfp)GR-amide
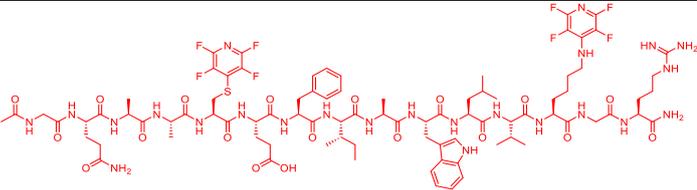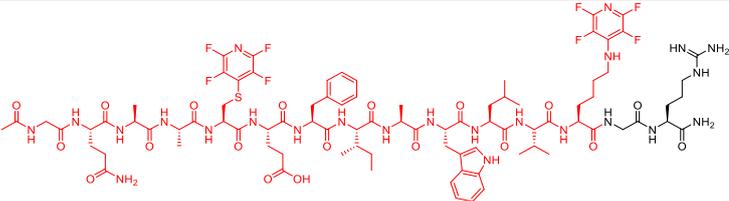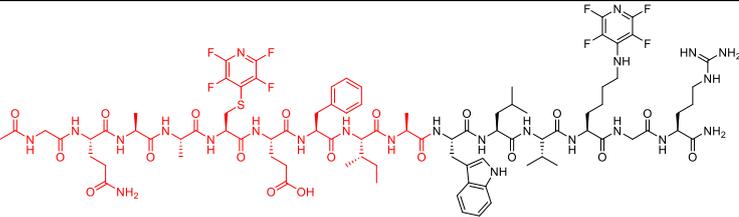
Table 4.3: MS-MS data of peptide **260** (Fragment observed in MS in red).

Entry **1** and **2** are significant in analysis of the structure of peptide **260** as the masses observed *m/z* 1971.1 and *m/z* 1764.0 correspond to peptides containing 2 PFP units and are not observed in the data for the singularly tagged peptide. **Entry 1** corresponds to the whole peptide without any fragmentation and **Entry 2** (*m/z* 1764.0) corresponds to a deletion of terminal Arg and Gly. The peak at *m/z* 1082.2 corresponds to a fragmentation of Arg, Gly, Lys(PFP), Val, Leu, Trp, Ala, Ile and Phe (**Entry 3**). This mass is also present in the MS-MS spectra of the singular tagged peptide which suggests that the second PFP has been cleaved.

To confirm the structure of peptide **261** MS-MS experiments were carried out. Shown in **Table 4.4** are the MS-MS data for peptide **261** which was believed to be singularly tagged at the Cys residue only.

Entry	Mass	Fragment
1	1610.5	 **261** Ac-GQAAC(Pfp)EFIAWLVK GR-amide
2	1089.7	 **261** Ac-GQAAC(Pfp)EFIA WLVKGR-amide
3	829.0	 **261** Ac-GQAAC(Pfp)E FIAWLVKGR-amide
4	623.2	 **261** Ac-GQAAC(Pfp) EFIAWLVKGR-amide
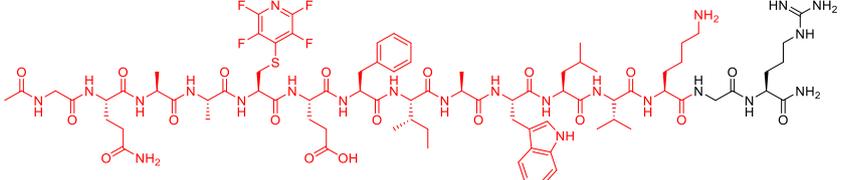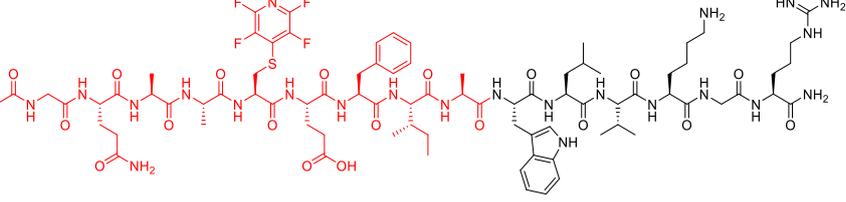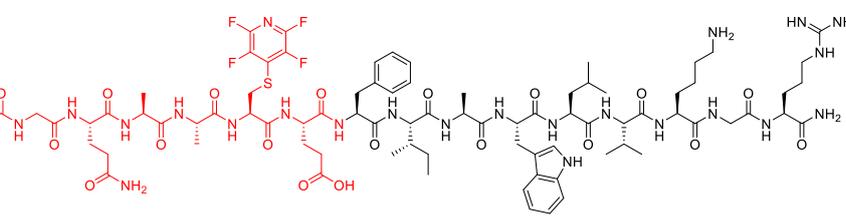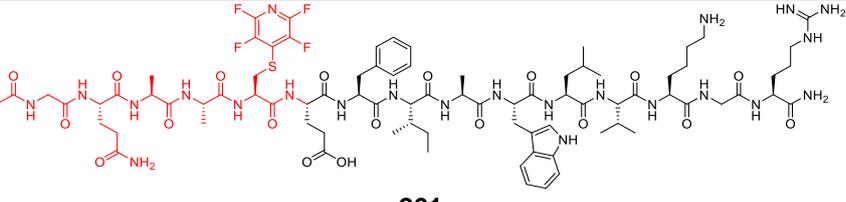
Table 4.4: MS-MS data of peptide **261** (Fragment observed in MS in red).

Entry **1** in **Table 4.4** corresponds to a deletion of Arg and Gly (*m/z* 1610.5). Entry **2**, **3** and **4** correspond to peptides which have further deletions, but still contain the Cys(PFP) residue. Further small fragment peptides are observed which suggest the Cys(PFP) has been removed.

The data from these MS-MS experiments suggest that for the reaction of peptide **259** using PFP, DIPEA and DMF as a solvent, a double-tagged peptide **260** is formed with tagging taking place at the Lys and Cys residues. For the reaction of peptide **259** with PFP, DIPEA and using TFE as a solvent, the MS-MS data suggest that a mono-tagged peptide **261** is formed with selective tagging occurring at the Cys residue.

4.4 Synthesis of Cys-modified GLP-1 30-mer (250)

Having established a selective tagging method towards Cys, efforts were made to synthesise the full 30-mer Cys modified GLP-1 (**250**). If **250** could be synthesised, then selective tagging of Cys with PFP could be carried out using the reaction conditions described in **Section 4.3.1** (**Scheme 4.7**).
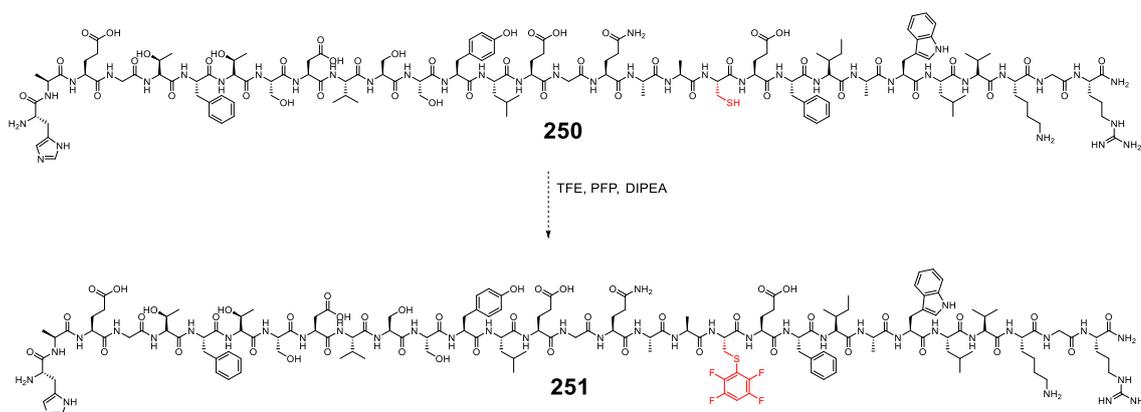
Scheme 4.7: Selective tagging of peptide **250** with PFP using TFE.

Analysis of the crude reaction mixture of peptide **259** revealed a significant amount of Cys deletion by products. It was decided that this amino acid would be manually coupled using HATU in an attempt to improve the overall synthesis (**Scheme 4.8**).
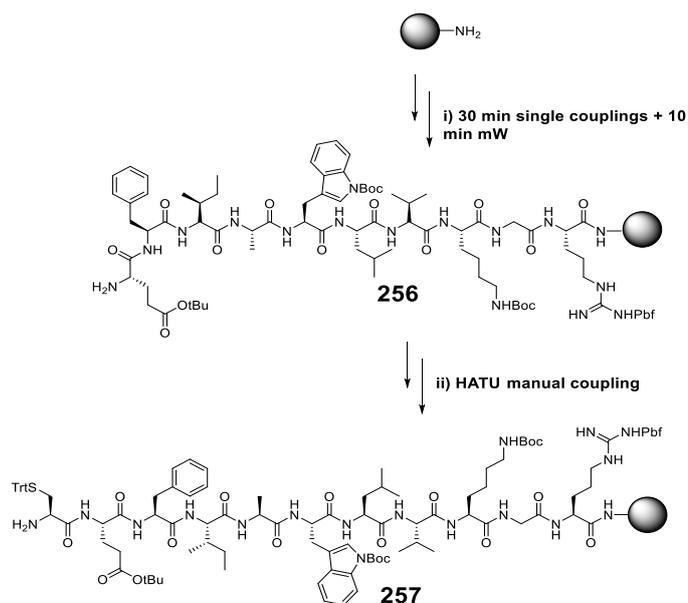
Scheme 4.8: i) Automated SPPS synthesis of peptide **256** and ii) HATU manual coupling resulting in **257**.

The synthesis of peptide **257** was successful with a test from resin showing the desired material 1319.7 [M+H]$^+$ had been synthesised (**Figure 4.15**).
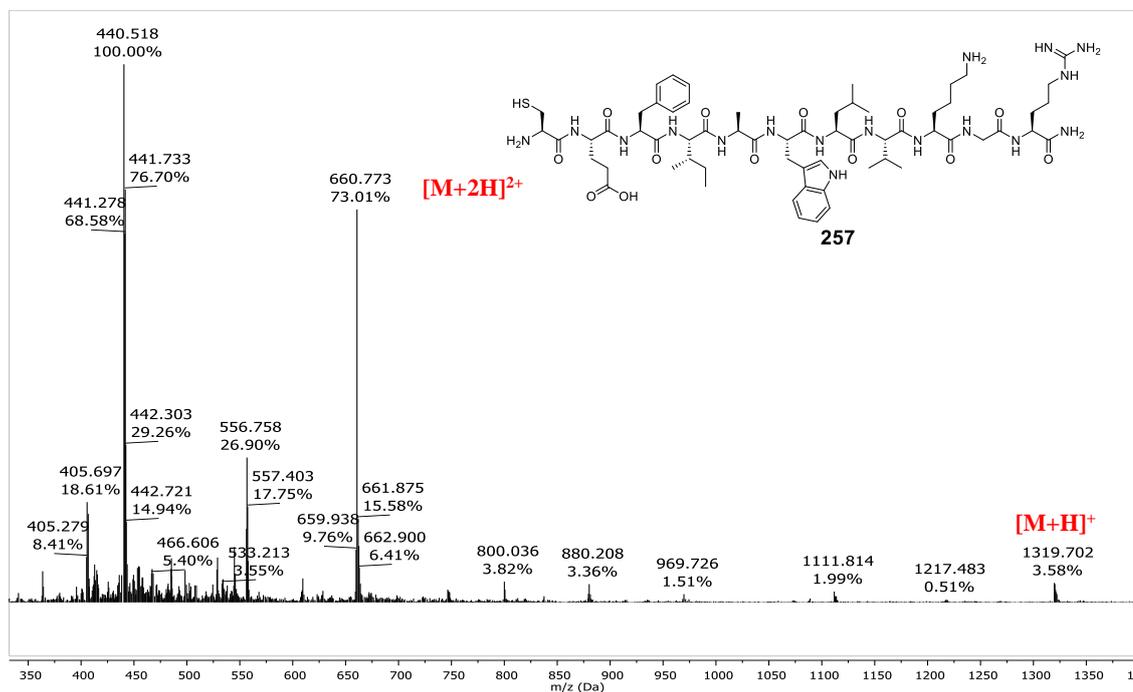
Figure 4.15: MS of crude peptide **257** with [M+H]$^+$ 1319.7.

Peptide **257** was then used to synthesise the 30-mer Cys modified GLP-1 27 (**Scheme 4.9**). The rest of the peptide sequence was continued using automated SPPS using 60 min single couplings at rt + 10 min mW and double 60 min couplings at rt +10 min mW for the final 4 amino acids (His, Ala, Glu and Gly) (**Scheme 4.9**).
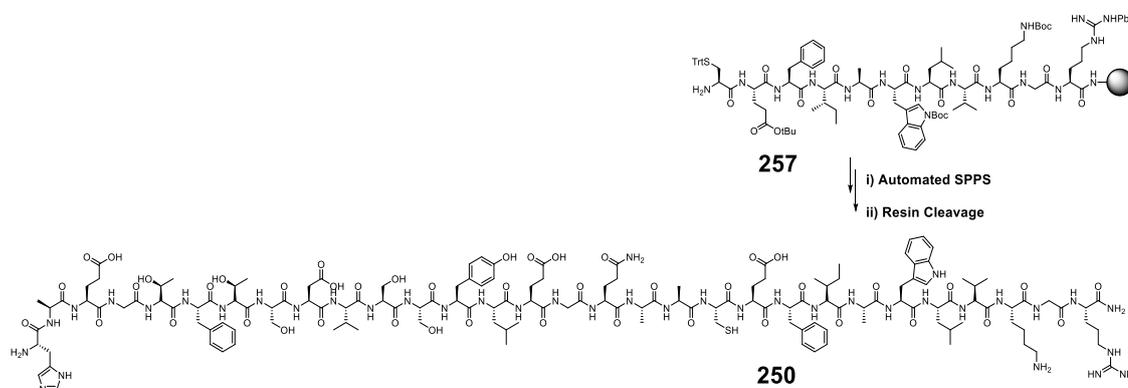
Scheme 4.9: Automated SPPS synthesis of peptide **250** and cleavage of peptide from resin.

The resulting peptide **250** (HAEGTFTSDVSSYLEGQAACEFIAWLVKGR-amide) was cleaved from resin as described in **Section 6.3.5**, analysed using MALDI-ToF and was found to contain the desired peak of 3274.0 [M+H]$^+$, His deletion of m/z 3136.9 aswell as the signal at m/z 3033.0 which corresponds to Cys and His deletion.
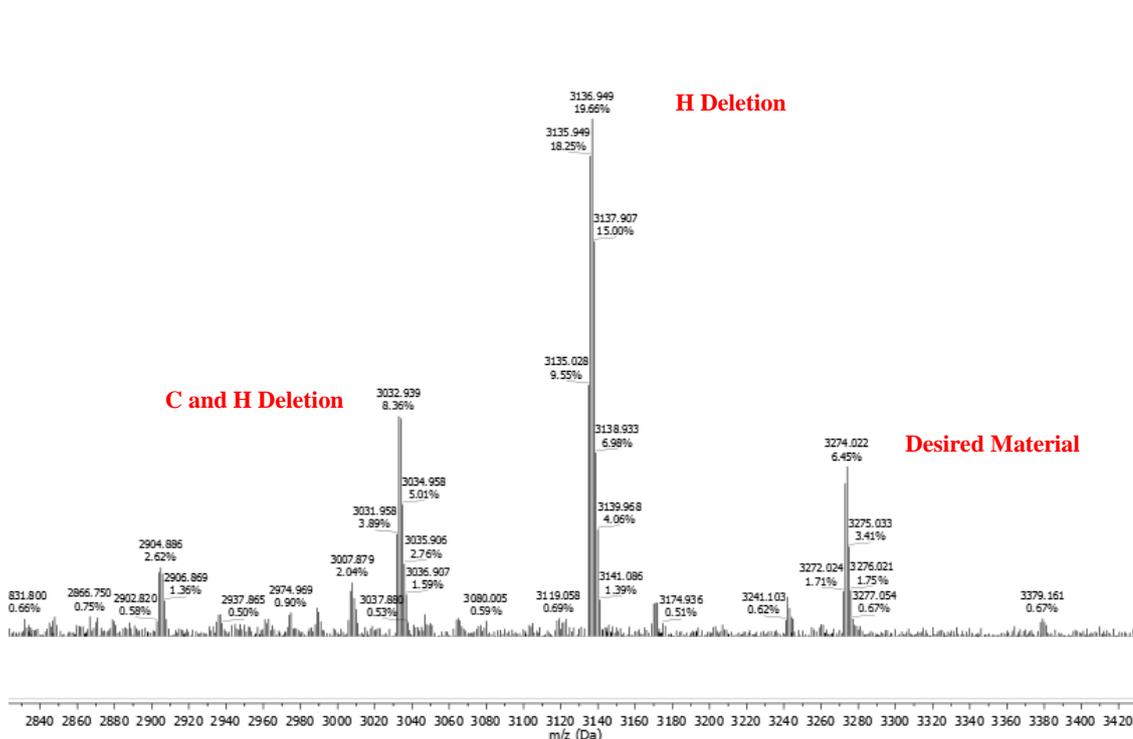
Figure 4.16: MALDI-ToF of peptide **250** with desired m/z 3274.0.

Due to the significant amount of His deletion product that was obtained, the target peptide was resynthesized using 0.5 mmol peptide **250** and the same coupling conditions as in the previous synthesis of peptide **245,** except that the final His coupling was carried out manually using HATU as a coupling agent (**Scheme 4.10**).
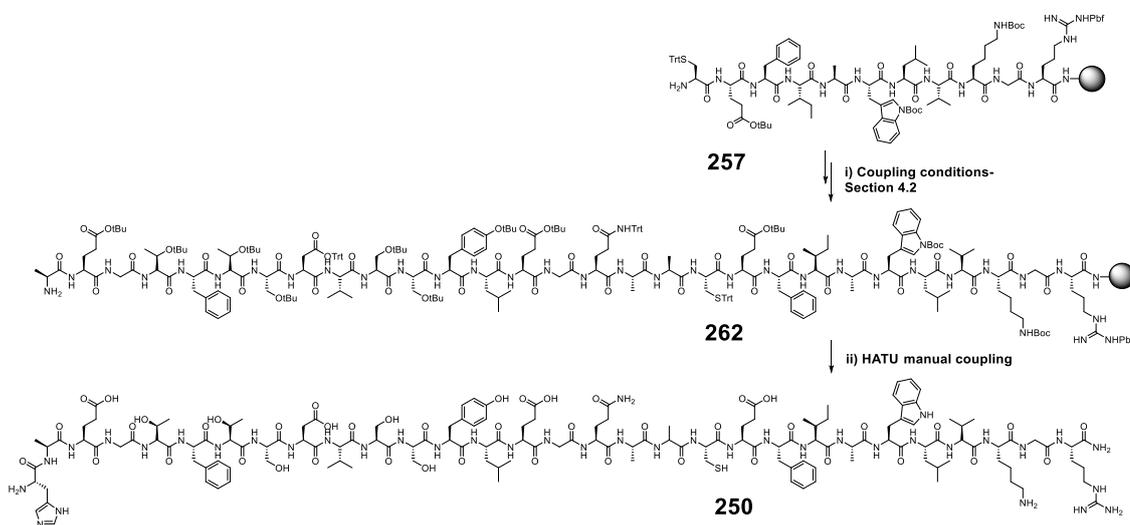
Scheme 4.10: i) Automated SPPS synthesis of peptide **262** and ii) HATU manual synthesis of Fmoc-His(Trt)-OH resulting in peptide **250**.

The resulting peptide **250** (HAEGTFTSDVSSYLEGQAACEFIAWLVKGR-amide) was cleaved from resin as described in **Section 6.3.5**, analysed using MALDI-ToF and the desired mass of m/z 3272.6 [M+H]$^+$ (**Figure 4.17**).
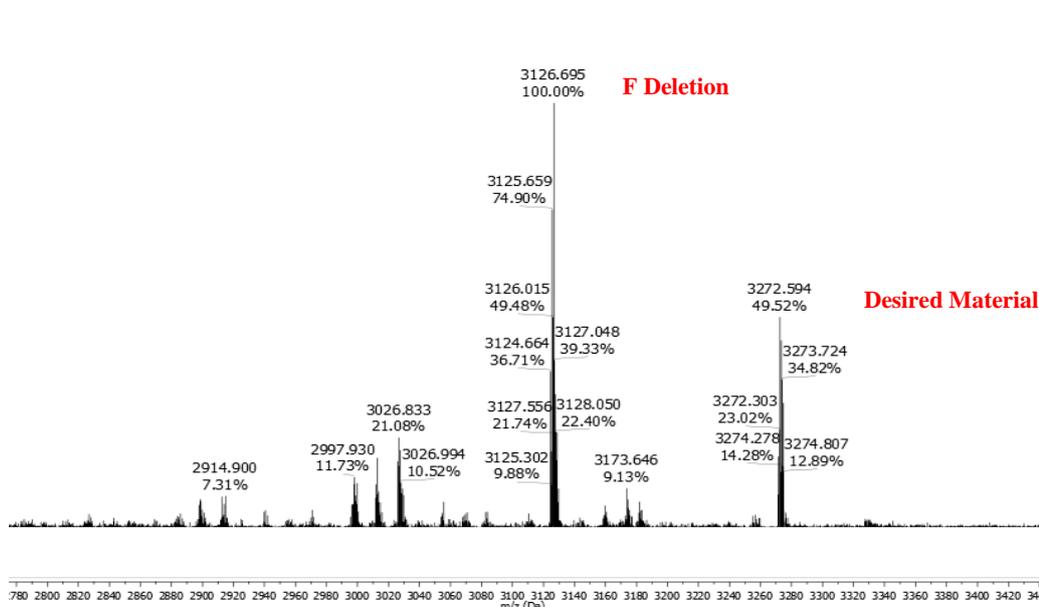
Figure 4.17: MALDI-ToF of peptide **250**.

HPLC purification of the peptide **250** was carried out using the conditions described in **Section 6.7.1** and the resulting purified peptide was analysed using MALDI-ToF (**Figure 4.18**).
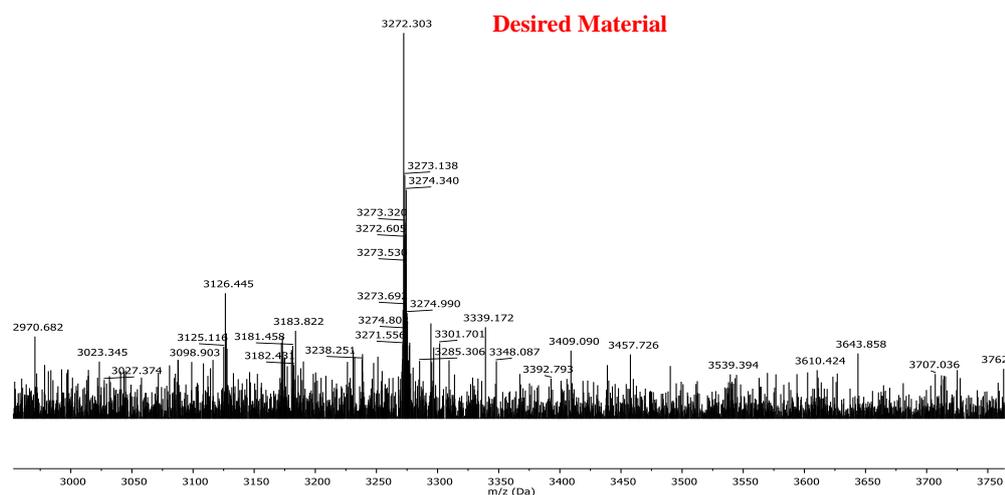
Figure 4.18: MALDI-ToF of purified peptide **250**.

Although the peak for the desired material *m/z* 3272.3 [M+H]$^+$ was present in the MALDI-ToF (**Figure 4.18**), the purification of peptide **250** resulted in a significant loss of product. Therefore, the synthesis and purification of peptide **250** needs to be optimised in future to allow for viable amounts of peptide to be produced for future PEGylation studies.

4.5 Enzymatic Degradation Studies

With synthesis of the native GLP-1 **245** and PFP-modified GLP-1 fragments **261** and **260** carried out successfully, the stabilities of the peptides towards enzymatic degradation were investigated. The degradation of GLP-1 (**245**) *in vivo* is extremely rapid and is carried out primarily by DPPIV which cleaves the *N*-terminal amino acids of the sequence resulting in an extremely short half-life.[16] For the enzymatic degradation studies the same reaction conditions that were used for the previous work on OT and VP (**Section 3.3.2 and 3.3.3**) were applied.

A peptide solution in Tris buffer and enzyme solution of 1 mg/ml were made up and transferred to an eppendorf tube with a ratio of 25:1 (peptide: enzyme solution). The peptide/enzyme solution was incubated in a water bath for 24 h and aliquots removed every hour, which were analysed using analytical HPLC.

Shown in **Figure 4.19** are the analytical HPLC traces from the enzymatic degradation of native GLP-1 (HAEGTFTSDVSSYLEGQAAKEFIAWLVKGR-amide) using chymotrypsin. At t = 0, there is a single peak which corresponds to the peptide **245** which has not undergone any degradation. Within 1 h, peptide **245** has been completely degraded to smaller fragments, the masses of which are shown in **Table 4.5**. Over the course of 24 h, there was no further degradation beyond the level seen for t = 1.

t=0

t=1

t=2

t=3

t=4

t=5

t=6

t=24

Figure 4.19: Analytical HPLC traces (220 nm) following the enzymatic degradation of native GLP-1 (**245**). Note - Analytical HPLC analysis was carried out for a total run time of 30 min but for formatting reasons only t = 5- 25 min are shown in the traces above.

Each analytical HPLC fraction underwent Q-ToF analysis and the results are given in **Table 4.5**. As can be seen at t = 0 the analytical HPLC trace showed a single peak which correspond to the intact peptide (e.g. masses of 1649.3 $[M+2H]^{2+}$, 1099.2 $[M+3H]^{3+}$, 825.7 $[M+4H]^{4+}$ and 600.1 $[M+6H]^{6+}$ in the Q-ToF). After 1 h, **242** has underwent significant degradation which is observed in the Q-ToF with the presence of many deletion products (**Entry 2**). Most of these deletion products are small fragments which have been cleaved at or near an aromatic side chain such as the fragments with *m/z* 758.5 which has been cleaved at the Trp residue and 496.0 which has been cleaved at Trp and Phe residues (**Table 4.5, Entry 2**). This would be expected as chymotrypsin cleaves amide bonds at the carbonyl of an aromatic side chain. Over the course of 24 h there was no further degradation with the same masses observed in all samples.

Mass in MALDI-ToF

389.3	496.5	777.8	992.2 571.0	496.0	758.5

Entry	Time	Mass (as observed in Q-ToF)	Structure
1	t = 0	1649.3 [2M+H], 1099.2 [3M+H], 825.7 [4M+H], 660.1[6M+H]	HAEGTFTSDVSSYLEGQAAKEFIAWLVKGR-NH$_2$
2	t = 1	992.2, 777.8, 758.5, 661.1, 571.0, 496.0, 496.5, 389.3	YLEGQAAKE [992.2] TSDVSSY [777.8] WLVKG [758.5] HAEGTF [661.1] GQAAKEF [571.0] FIAW[496.0] HAEGT [496.5] HAE [389.3]
3	t = 2	Same as above	Same as above
4	t = 3	Same as above	Same as above
5	t = 4	Same as above	Same as above
6	t = 5	Same as above	Same as above
7	t = 6	Same as above -	Same as above
8	t = 24	Same as above	Same as above

Table 4.5: MS analysis of enzymatic degradation of compound **245**.

Enzymatic degradation of the native GLP-1 **245**, as expected, was found to be quick and efficient with complete degradation of the peptide occurring within 1 h in this experiment. It was thought (based on the work in **Chapter 3**) that incorporation of the PFP tag into GLP-1 could increase the stability of the peptide. To test this theory, before committing to the tagging reaction of the full length Cys modified GLP-1 (**250**) (as this material was not available on a large scale-See **Section 4.4**) enzymatic degradation of the small peptide analogues **263** and **261** were tested. Peptide **263** was selected to act as a control in the stability study as it corresponds to the first 15 amino acids of native GLP-1. Peptide **261** is the single PFP-tagged GLP-1 analogue and was expected to have a positive effect on the stability of the peptide.
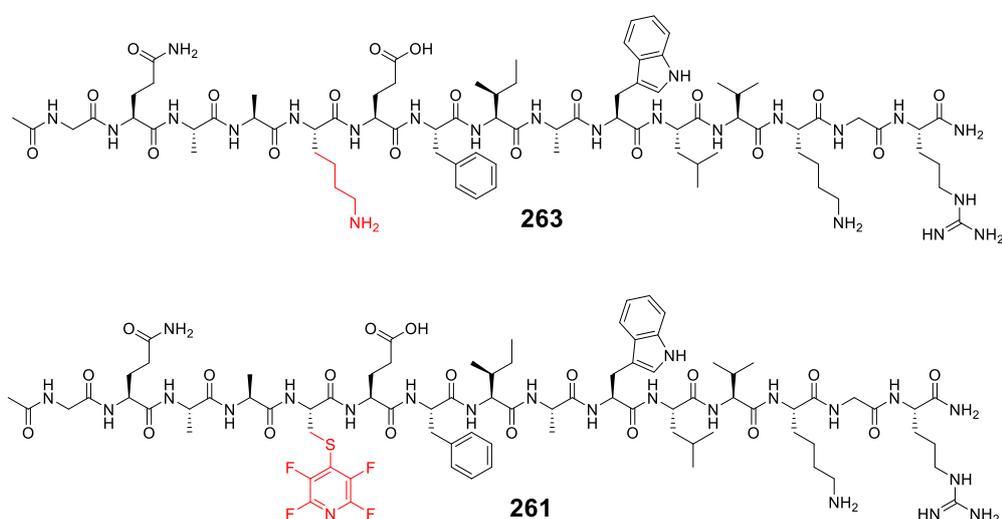
Figure 4.20: 15-mer fragment of native GLP-1 acting as a control (**263**) and singular PFP-tagged analogue (**261**).

Figure 4.21 show the analytical HPLC traces of the peptide **263**. This was carried out to act as a control to compare the stability to the PFP-modified peptide and determine if the incorporation of a singular PFP into the peptide could improve stability.
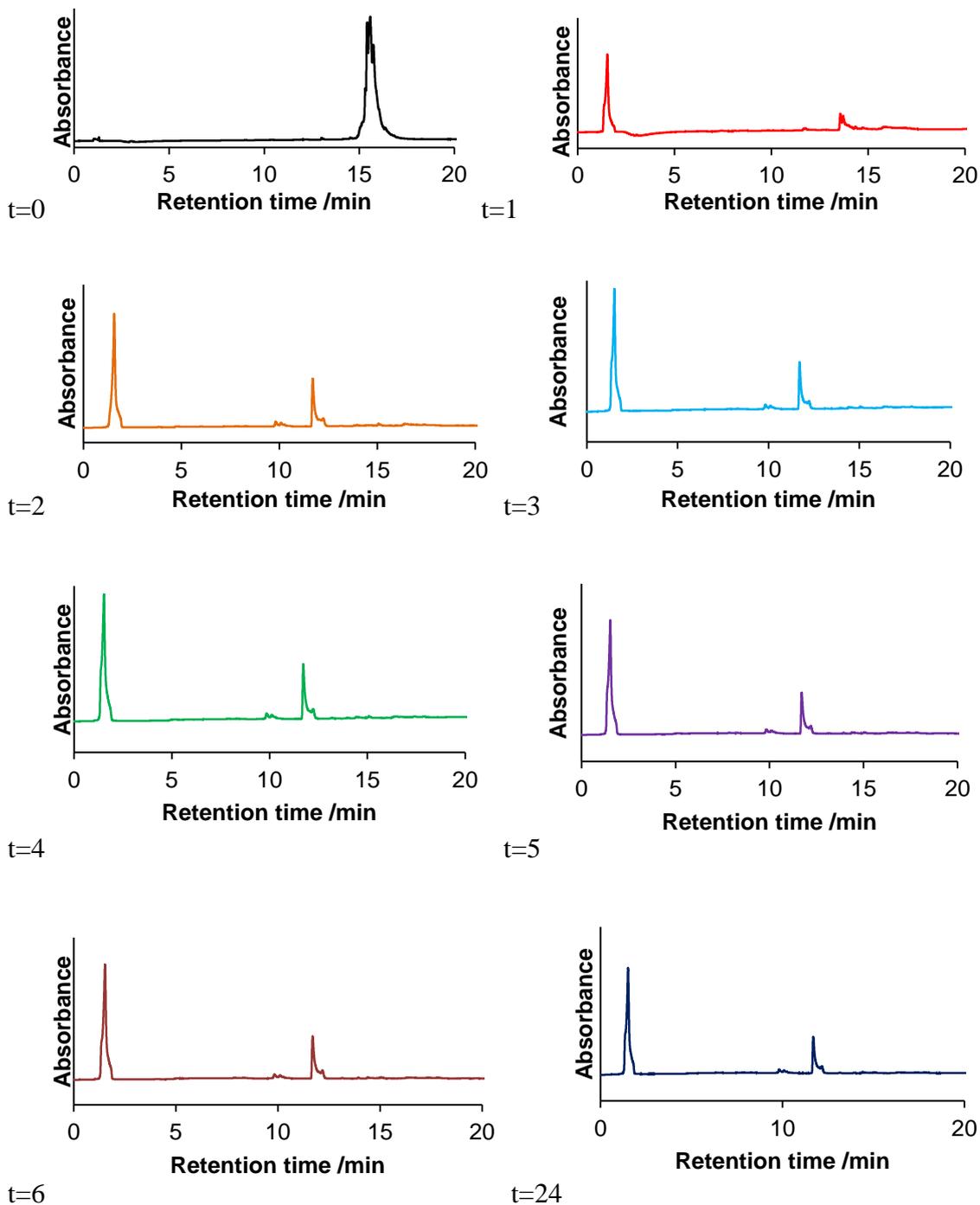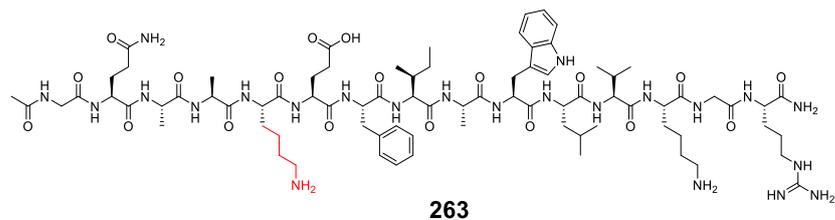
Figure 4.21: Analytical HPLC traces (220 nm) following the enzymatic degradation of peptide **263**. Note - Analytical HPLC analysis was carried out for a total run time of 30 min but for formatting reasons only t = 0-20 min are shown in the traces above.

Mass observed in LCMS 286.5 397.1 644.0 777.4 959.2 1163.3
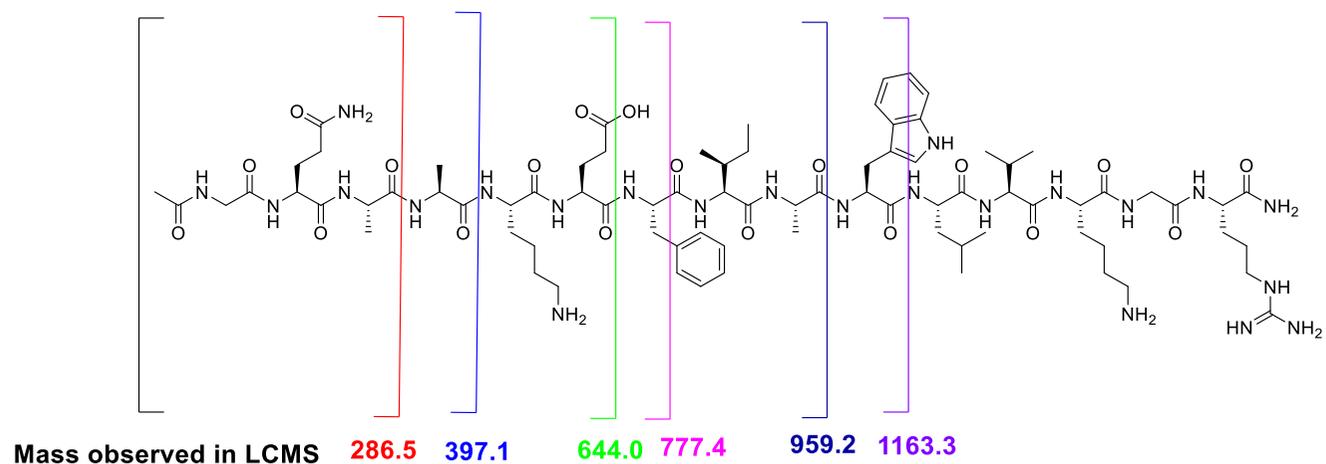
Entry	Time	Mass observed	Structure
1	t = 0	1715.0	Ac-GQAAKEFIAWLVKGR [1715.0]
2	t = 1	1163.3, 959.2, 777.4	Ac-GQAAKEFIAW [1163.3] Ac-GQAAKEFIA [959.2] Ac-GQAAKEF [777.4]
3	t = 2	777.4, 644.0, 397.1, 286.5	Ac-GQAAKEF [777.4] Ac-GQAAKE [644.0] Ac-GQAA[397.1] Ac-GQA [286.5]
4	t = 3	Same as above	Same as above
5	t = 4	Same as above	Same as above
6	t = 5	Same as above	Same as above
7	t = 6	Same as above	Same as above
8	t = 24	Same as above	Same as above

Table 4.6: MS data of enzymatic degradation of peptide **263**.

At t = 0 there is more than one peak but the main peak present corresponds to the original peptide (**263**) 1715.0 [M+H]$^+$. At t = 1 there has been significant degradation of the native peptide with the appearance of 1163.3. At t = 2 there has been complete degradation of the native peptide (**263**) to peptides with *m/z* 777.4, which is the fragmentation product relating to a small peptide fragments cleaved near the *N*-terminus (**Table 4.6, Entry 3**). Over the course of 24 h there is no further degradation from this fragment. Although the analytical HPLC trace at t = 0 suggests the peptide was not pure, it is reasonable to conclude that over the course of 2 h there has been complete degradation of the peptide *m/z* 1715.0 to smaller fragments.

Figure 4.22 shows the analytical HPLC trace of the singular PFP tagged peptide **261**. This was expected to be more stable to enzymatic degradation than the control peptide **263**.
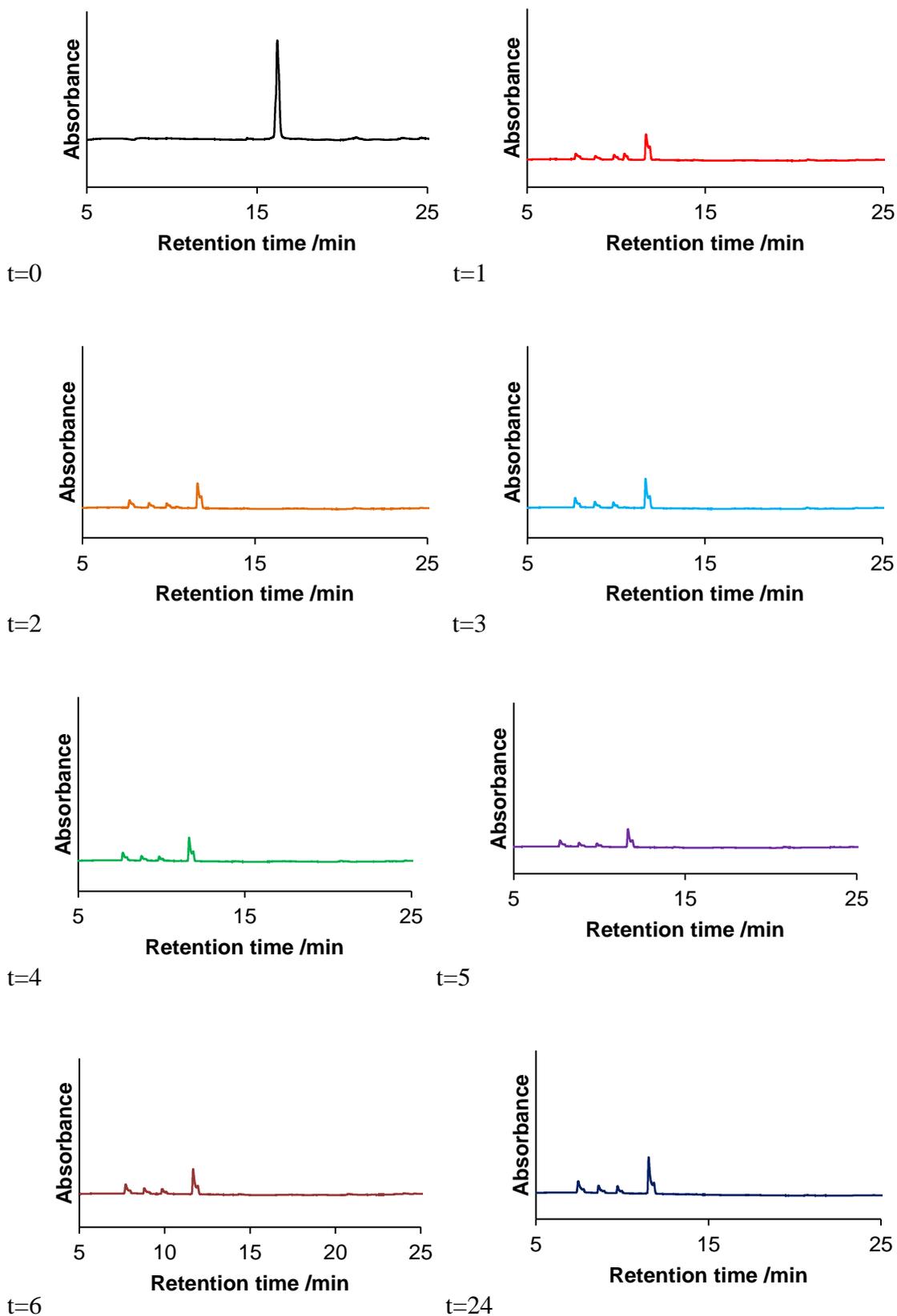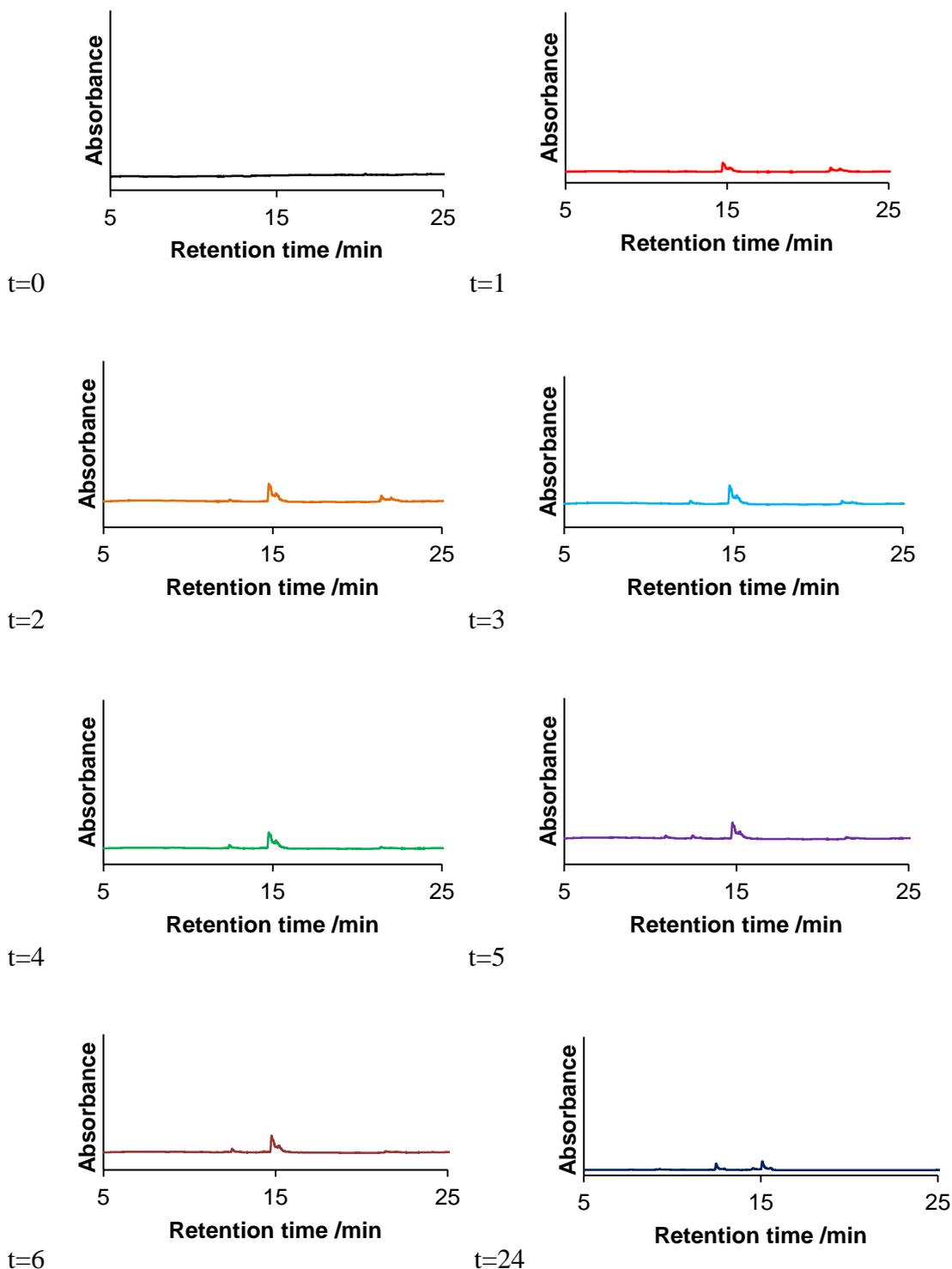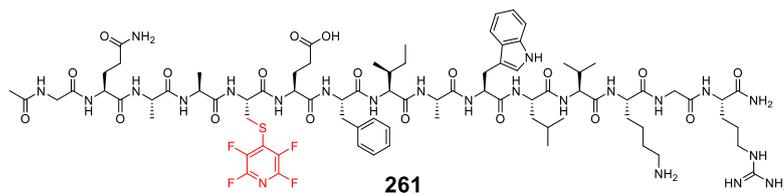
Figure 4.22: Analytical HPLC traces (220 nm) following the enzymatic degradation of peptide **261**. Note - Analytical HPLC analysis was carried out for a total run time of 30 min but for formatting reasons only t = 5-25 min are shown in the traces above.

From the analytical HPLC data, it is difficult to fully visualise any degradation that is occurring due to the poor absorbance of the samples. This is due to the small amount of peptide used in the experiment (0.1 mg of peptide used in this study whereas 1 mg used in the previous study (**Figure 4.21**). For a better understanding of any degradation, mass spectroscopy was carried out and the data are shown in **Table 4.7**.

Mass observed in LCMS   289.0   929.6 1034.2   854.7
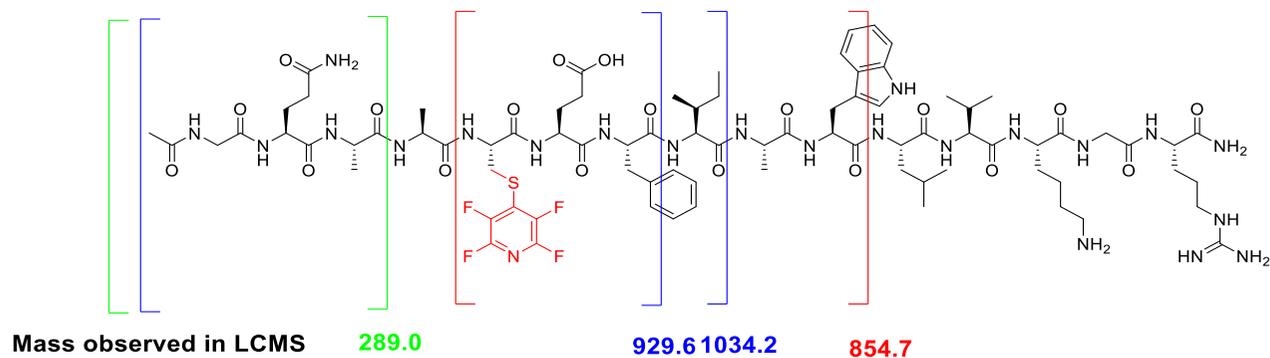
Entry	Time	Mass observed	Structure
1	t = 0	1839.1 [M+H]$^+$	Ac-GQAAC(PFP)EFIAWLVKGR-NH$_2$
2	t = 1	1839.1, 1034.2, 929.6, 845.7, 518.0, 412.1, 289.0	Ac-GQAAC(PFP)EFIAWLVKGR-NH$_2$ [1839.1] Ac-GQAAC(PFP)EFI [1034.2] Ac-GQAAC(PFP)EF [929.6] C(PFP)EF [518.0] C(PFP)EFIAW [845.7] C(PFP)E [412.1] Ac-GQA [289.0]
3	t = 2	Same as above	-
4	t = 3	Same as above	-
5	t = 4	Same as above	-
6	t = 5	1034.2, 929.6, 845.7, 518.0, 412.1, 289.0	Ac-GQAAC(PFP)EFI [1034.2] Ac-GQAAC(PFP)EF [929.6] C(PFP)EF [518.0] C(PFP)EFIAW [845.7] C(PFP)E [412.1] Ac-GQA [289.0]
7	t = 6	Same as above	-
8	t = 24	Same as above	-

Table 4.7: Mass spectroscopy data of enzymatic degradation of peptide **261**.

At t = 0 the main product observed in the LCMS is the [M+H]$^+$ 1839.1 peptide **261** and over the course of 1 h significant degradation has occurred (**Table 4.7, Entry 2**). The main product observed at t = 1 is still 1839.1, but there are approximately 5 other side products all relating to smaller peptide fragments (with mass 1034.2, 929.6, 518.0, 412.1 and 289.0) (**Table 4.7, Entry 2**). At t = 5 the main product becomes the peptide corresponding to mass 1034 which is a fragment corresponding to a loss of Arg, Gly, Lys, Val, Leu, Trp, Ala, Phe and Glu. At this point the peptide corresponding to mass 1839.1 has been completely degraded (**Table 4.7, Entry 6**). As with the degradation of GLP-1, it appears the enzyme has cleaved the amide bonds at a number of aromatic side chain sites, which is what would be expected with using chymotrypsin as an enzyme.

Enzymatic degradation of peptide **263** showed that the peptide was fully degraded to smaller fragments after 2 h of the experiment and peptide **261** was fully degraded after 5 h. Comparing the stabilities of peptide **263** and peptide **261**, it is reasonable to hypothesise that the modified peptide **261** is more stable to enzymatic breakdown, which would align with what was observed in **Chapter 3**.

4.6 Conclusions

Synthesis of GLP-1(7-36) **245** was carried out using automated Fmoc SPPS and analysed using MALDI-ToF (**Figure 4.10**). Several different coupling conditions were adopted to optimise the synthesis of **245**. The most problematic residues in the peptide synthesis proved to be the last 4 residues, with these deletion products reappearing in multiple syntheses. To overcome these synthetic problems, longer reaction times and double and triple couplings were adopted. The His residue at position 7 in the sequence proved the most difficult of all to attach to the peptide, even when manual coupling with HATU was carried out (**Section 4.2**). It is hypothesised this was due to the bulky Trt protecting group of the His and using a Boc protected His may allow for easier coupling of the final His amino acid in the future.

Due to the poor stability of GLP-1 (**245**) *in vivo,* various efforts have been made by various research groups[7,8,12] to improve the enzymatic stability of peptides. Based on our work in **Chapter 3** we hypothesised that by attaching a singular PFP to a Cys-modified GLP-1 (**250**), the enzymatic stability may be improved. To achieve this aim and develop the selective tagging chemistry a smaller 15-mer fragement of the Cys-modified peptide (**250**) was synthesised and reaction conditions were optimised to achieve a singular tagging of PFP using TFE as a solvent (**Section 4.3.1**).

Before committing to synthesis of the full Cys-modified GLP-1, the enzymatic stability of the smaller modified peptide fragment **261** was compared to the native peptide **263**.
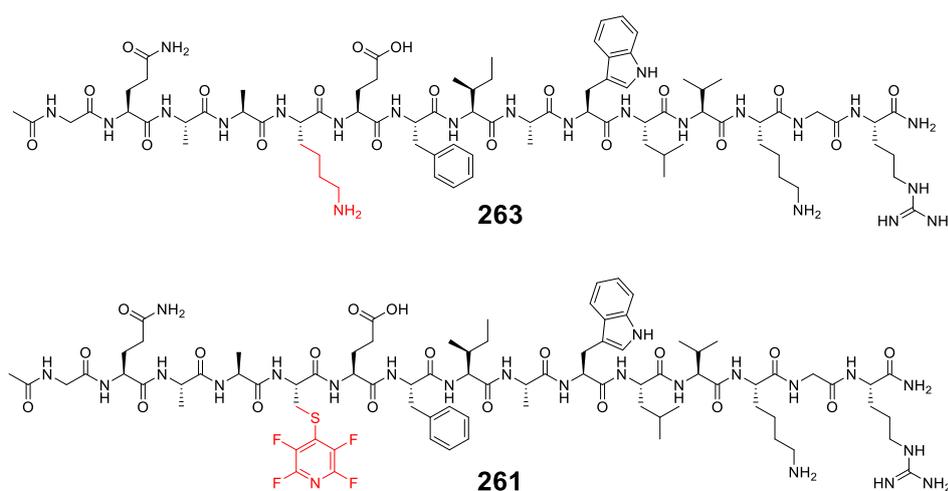
Figure 4.23: Structure of control peptide **263** and PFP-tagged peptide **261**.

It was found that the modified peptide **261** was more stable to enzymatic breakdown than the control peptide **263**. For the enzymatic degradation of peptide **263**, the peptide was fully degraded within 2 h (**Figure 4.21**). The PFP-tagged peptide **261** was found to be more stable as this was degraded over the course of 5 h (**Figure 4.22**). This provided confirmation that modifying a peptide at a singular site with PFP does improve enzymatic stability. We hypothesised that a similar effect may be observed with the full PFP-modified GLP-1, so efforts were made towards the synthesis of peptide **250** (**Section 4.4**).

Figure 4.24: Structure of peptide Cys-modofied GLP-1 (**250**).

The synthesis of this peptide proved challenging (as was the native GLP-1) with a variety of coupling conditions being adopted. Eventually peptide **250** was successfully synthesised and purified, but not enough of the resulting peptide could be recovered to react further.

References

[1] M. Lorenz, A. Evers, M. Wagner, *Bioorg. Med. Chem. Lett.*, 2013, **23,** 4011-4018.

[2] F. Selis, R. Schrepfer, R. Sanna, S. Scaramuzza, G. Tonon, S. Dedoni, P. Onali, G Orsini, S. Genovese, *Pharma. Sci.*, 2012, **2**, 58-65.

[3] R. J. Rodgers, M. H. Tschop, J. P. H. Wilding, *Dis. Model. Mech.*, 2012, **5**, 621-626.

[4] J. M. Chalker, L. Lercher, N. R. rose, C. J. Schofield, B. G. Davis, *Angew. Chem. Int. Ed.*, 2012, **124**, 1871-1875.

[5] A. Kaspar, J. M. Reichert., *Drug Discov. Today*, 2013, **18**, 807-817.

[6] L. Prasad-Reddy, D. Isaacs, *Drugs in Context*, 2015, **4**, 212283-212302.

[7] M. J. Roberts, M. D. Bentley, J. M. Harris, *Adv. Drug Deliver. Rev.*, 2012, **64**, 116-127.

[8] S. H. Lee, S. Lee, Y. S. Youn, D. H. Na, S. Y. Chae, Y. Byun, K. C. Lee, *Bioconjugate Chem.*, 2005, **16,** 377-382.

[9] K. Madsen, L. B. Knudsen, H. Agersoe, P. F. Neilson, H. Thogersen, M. Wilken, N. L. Johansen*, J. Med. Chem.*, 2007, **50,** 6126-6132.

[10] C. Clemmensen, J. Chabanne, B. Finan, L. Sullivan, K. Fischer, D. Kuchler, L. Sehrer, T. Ograjsek, S. M. Hoffmann, S. C. Schriever, P. T. Pfluger, J. Pinkstaff, M. H. Tschop, R. DiMarchi, T. D. Muller, *Diabetes*, 2014, **63**, 1422-1427.

[11] H. Pan, J. Guo, Z. Su, *Physiol. Behav*., 2014, **130**, 157-169.

[12] B. Finan, T. D. Muller, C. Clemmensen, D. Perrez-Tilve, R. D. DiMarchi, M. H. Tschop, *Trends in Mol. Med.*, 2016, **22**, 359-376.

[13] B. Finan, B. Yang, N. Ottaway, D. L. Smiley, T. Ma, C. Clemmensen, J. Chabenne, L. Zhang, K. M. Habegger, K. Fischer, J. E. Campbell, D. Sandoval, R. J. Seeley, K. Bleicher, S. Uhles, W. Riboulet, J. Fuck, C. Hertel, S. Belli, E. Sebokova, K. Conde-Knape, A. Konkar, D. J. Drucker, V. Gelfanov, P. T. Pfluger, T. D. Muller, D. erez-Tilve, T. D. DiMarchi, M. H. Tschop, *Nat. Med.,* 2015, **21**, 27-39.

[14] B. Finan, B. Yamg, N. Ottaway, K. Stemmer, T. D. Muller, C. X. Yi, K. Habegger, S. C. Schreiver, C. Garcia-Caceres, D. G. Kabra, J. Hembree, J. Holland, C. Raver, R. J. Seeley, W. Hans, M. Irmler, J. Beckers, M. Hrabe de Angelis, J. P. Tiano, F. Mauvias-Jarvis, D. Perez-Tilve, P. Pfluger, L. Zhang, V. Gelfanov, R. D. DiMarchi, M. H. Tschop, *Nat. Med.*, 2012, **18**, 1847-1859.

[15] H. Vogel, S. Wolf, C. Rabasa, F. Rodriguez-Paceco, C. S. Babaei, F. Stober, J. Goldschmidt, R. D. DiMarchi, B. Finan, M. H. Tschop, S. L. Dickson, A. Schurmann, K. P. Shibicka, *Neuroparm.*, 2016, **110**, 396-406.

[16] D. Gimenez, C. A. Mooney, A. Dose, G. Sandford, C. R. Coxon, S. L. Cobb, *Org. Biomol. Chem.*, 2017, **15**, 4086-4095.

[17] B. Gallwitz, T. Ropeter, C. Morys-Wortmann, R. Mentlein, E. G. Siegel, W. E. Schmidt, *Regul. Peptides*, 2000, **86**, 103-111.

Chapter 5: Conclusions and Future Work

The aim of this study was to investigate the role of perfluoro-heteroaromatics in peptide modification and the effects of perfluoro-heteromatics on peptide stability. Small dipeptides have been successfully synthesised and modified using PFP (**129**) and the substitution pattern of PFP upon nucleophilic aromatic substitution with thiophenolate was investigated. Biologically active peptides OT, VP and GLP-1 were synthesised and modified using PFP and the stabilities of these modified peptides towards chymotrypsin tested and compared to the native peptides.

We have successfully illustrated during this study that PFP (**129**) is an interesting and versatile tool for peptide modification. It is highly reactive towards nucleophilic side chains within a peptide and has been shown to react unselectively with Cys, Lys, Tyr residues (**Scheme 5.1i**), but can also react selectively with Cys residues by altering the reaction conditions (**Scheme 5.1ii**). By utilising protecting group chemistry it is also possible to selectively tag reactive side chain residues (**Scheme 5.1iii**).
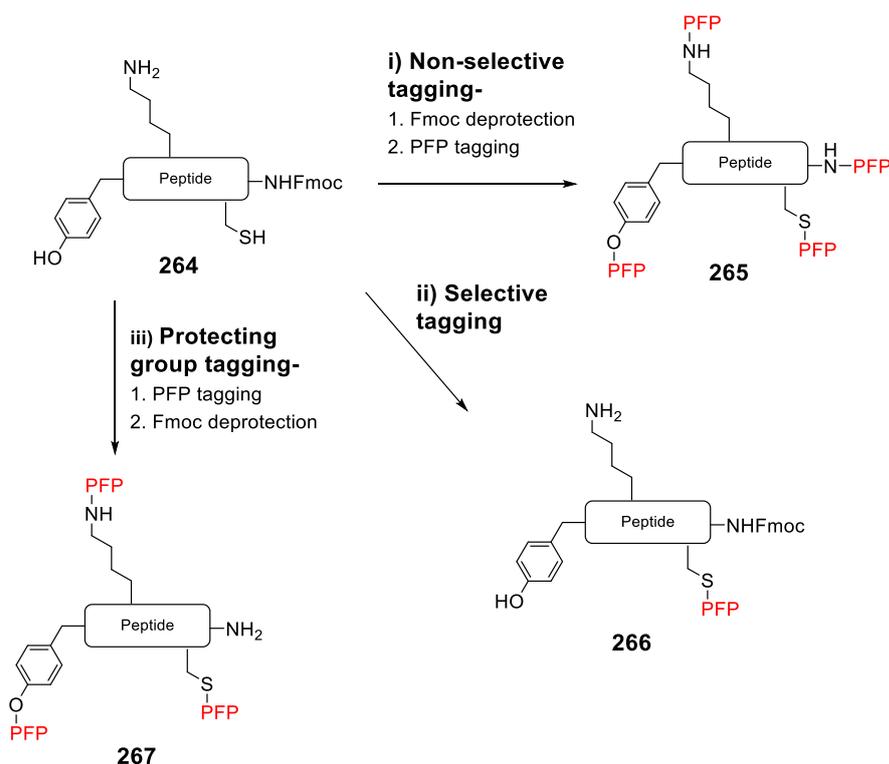
Scheme 5.1: i) Non-selective tagging of peptide substrates with PFP ii) Selective tagging of peptide substrates with PFP iii) Selective tagging of peptide with PFP.

Once peptides were synthesised and tagged using PFP, the stabilities of these peptides were tested and compared to the native peptides. It was determined that in all cases investigated in this study incorporation of the PFP resulted in an improvement of enzymatic stability compared to the native peptide (**Section 3.3.2, 3.3.3 and 4.5**). Even mono-tagged peptides were found to be stable to enzymatic stability which could prove a useful tool in improving the stabilities of peptides without losing biological activity.

In **Chapter 2**, it was shown that it is possible to tag dipeptides with PFP and further functionalise these activated peptides *via* nucleophilic aromatic substitution reaction. In this study we used thiophenolate to functionalise the attached fluoropyridine ring. It was determined using ^{19}F NMR spectroscopy that upon reaction of the PFP-tagged dipeptide with thiophenolate, that the nucleophile reacted at both the *ortho* and *meta* positions of the PFP ring resulting in a di-substituted peptide.

Building on from this initial study, future work in this area could investigate using thio-sugars or thio-PEG as a nucleophile to give access to modified peptides (**Scheme 5.2**)
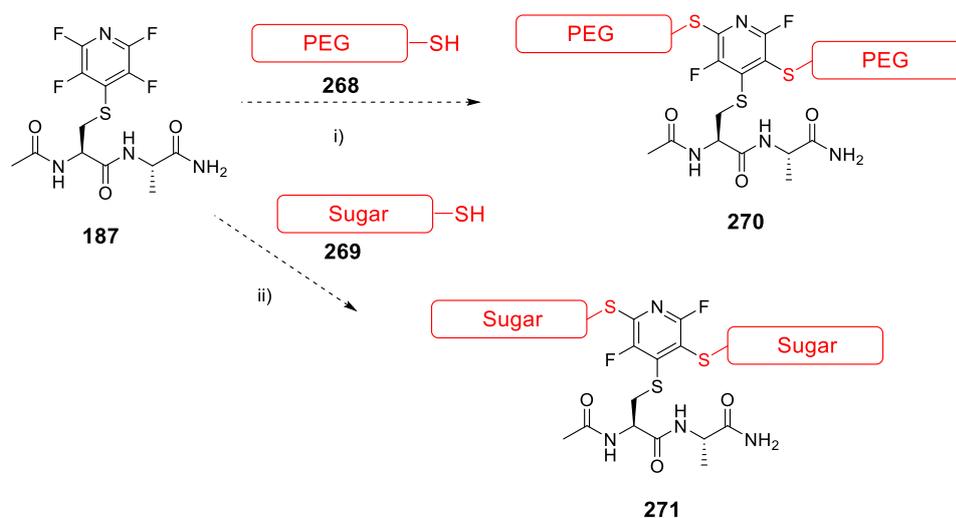
Scheme 5.2: Potential applications of PFP tagging- Aromatic substitution reaction between pentafluorpyridine tagged peptide **187** with i) thio-PEG (**268**) and ii) thio-sugar (**269**).

Using PFP as a scaffold could allow for the conjugation of thio-PEG or thio-sugars resulting in multiply tagged analogues. This could be beneficial over the singularly tagged analogues as it could lead to possibly further increased stability and solubility. Examples of multiply PEGylated peptides can be found in work examining branched PEG analogues. These analogues cause increased peptide stability due to an "umbrella effect".[1]

Examples of using small molecules as scaffolds in peptides can be found in the work of Liskamp and coworkers. Here, Liskamp and co-workers conjugated cyclic peptides to a small molecule scaffold in order to synthesise protein mimetics. They were able to build a library of protein surface epitope mimetics of HIV—gp120 using triazacyclophane as the scaffold and click chemistry to conjugate cyclic peptides to the scaffold (**Scheme 5.3**).[1] By conjugating the cyclic peptides to the small molecule scaffold, binding affinity to HIV CD4 cells was increased compared to the individual cyclic peptides or linear forms.[1]
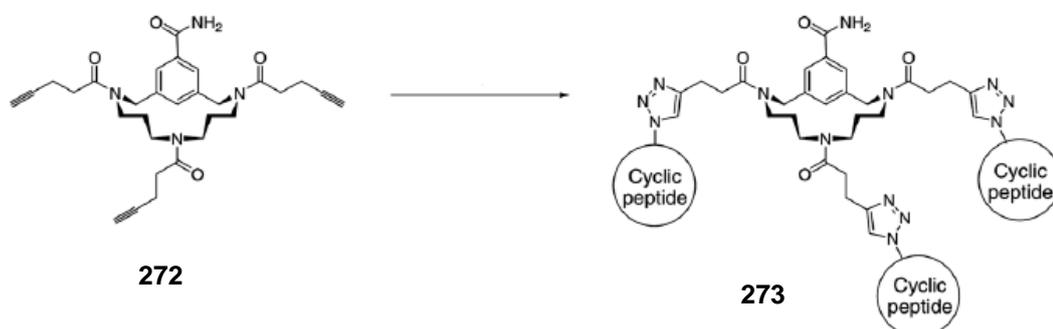
Scheme 5.3: Click reaction of triazacyslophane scaffold and cyclic peptides [1]

In **Chapter 3**, PFP-modified peptides (**220, 221, 222** and **223**) were synthesised using Fmoc-SPPS and modified with PFP using on-resin and solution phase tagging.
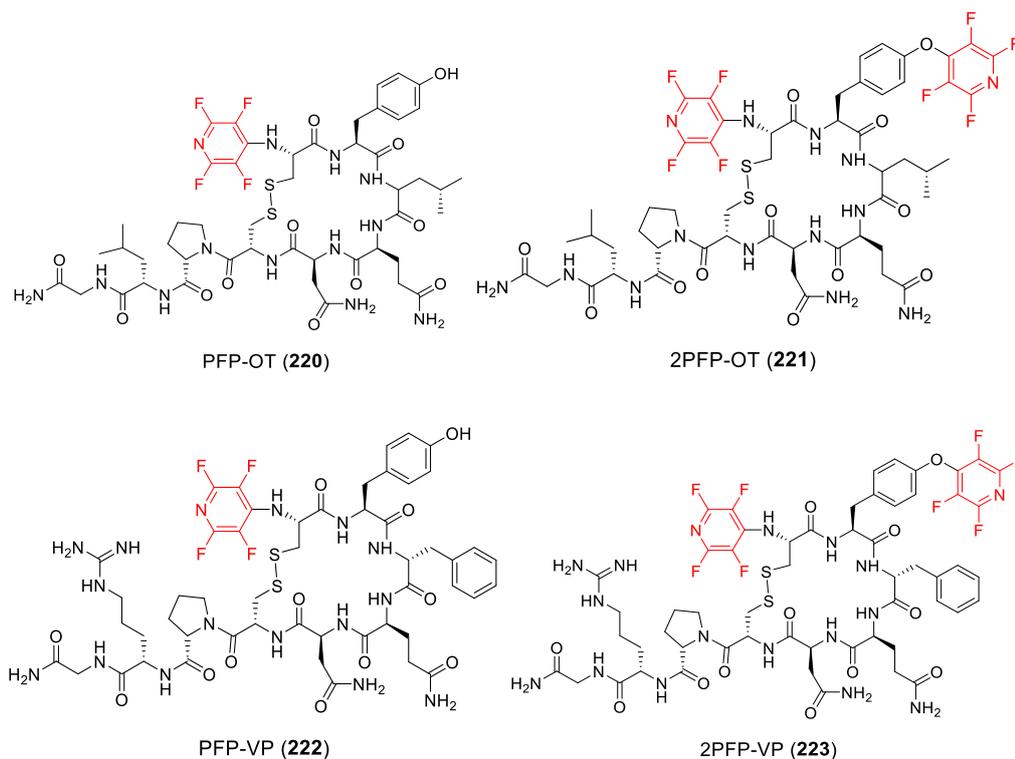
Figure 5.1: Structure of PFP-modified peptides PFP-OT (**220**), 2PFP-OT (**221**), PFP-VP (**222**) and 2PFP-VP (**223**).

The *in vitro* stability of the PFP-tagged OT (**220** and **221**) and VP analogues (**222** and **223**) were tested and provided promising results. In the case of the OT analogues, both PFP analogues were found to be more stable than the native OT and for VP the singly tagged peptide was found to be more stable than the native peptide (**Section 3.3.2 and 3.3.3**). Over the course of 24 h compounds **220** and **221** were degraded to the Gly deletion product, whereas native OT was fully degraded to small peptide fragments over the same course of time. A similar trend was observed with the VP PFP-tagged analogue (**222**) with significant improvement in peptide stability observed compared to the native peptide.

Future work into the stability of these peptides will concentrate on analysing the *in vivo* stabilities of these analogues by using human serum stability studies. The method for these studies is much the same as the *in vitro* study, but mouse or human serum is used instead of chymotrypsin. This *in vivo* study would give an indication of the stability of the analogues in a biological system.

Future work in this area may also concentrate on obtaining crystal structures of the PFP-tagged analogues and native OT and VP peptides. The ^{19}F NMR and MS-MS data provide reasonable proof that the PFP-tagged analogues have the structure shown in **Chapter 3,** but crystal structures of these compounds would provide definitive proof of the structures of these compounds.

In **Chapter 4**, we investigated the effect of PFP modification on a larger peptide, the 30-residue peptide GLP-1 (**245**). The aim of this chapter was to tag a Cys-modified peptide (**250**) with PFP (**129**) and compare the stability of this peptide to native GLP-1 (**245**). A small 15-mer fragment of the Cys-modified GLP-1 (**261**) was successfully synthesised and tagged using PFP and the stability of this peptide was compared to that of the 15-mer fragment of native GLP-1 (**263**). It was determined that the PFP-tagged peptide **261** exhibited reduced enzymatic degradation compared to the native peptide **263** (**Section 4.5**).
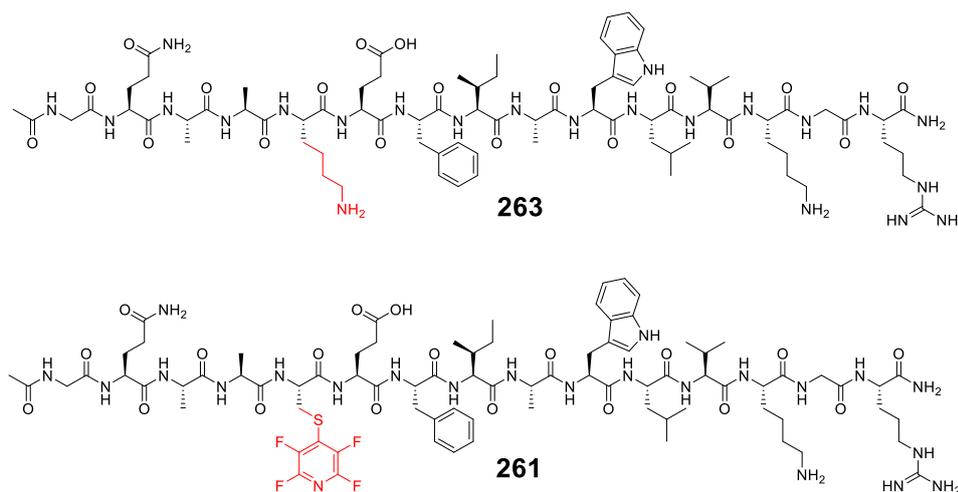
Figure 5.2: Structure of singularly PFP-tagged peptide **261** and the corresponding native GLP-1 fragment **263**.

Future work in this area of research should continue to try and firstly optimise the synthesis of peptide **250,** which was attempted in **Section 4.3**. The selective tagging conditions investigated in the tagging of the 15-mer GLP-1 fragment (**Section 4.3.1**) should also be adopted to selectively tag at the Cys residue to give peptide **251**.
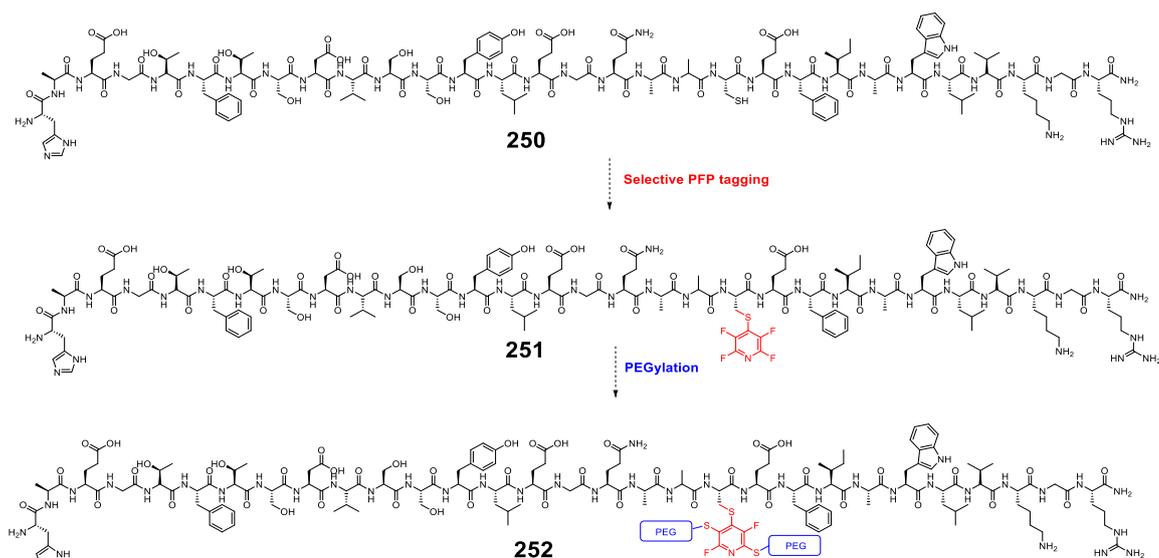
Scheme 5.5: Selective tagging of Cys of peptide **250** with PFP and further modification with PEG resulting in a PEGylated GLP-1 analogue **252**.

The stability of this peptide **251** compared to native GLP-1 **245** should be investigated as promising results were observed when smaller fragments of GLP-1 **263** and PFP-tagged peptide **261** were compared (**Section 4.5**). The possible modification of peptide **251** *via* the PFP moiety with PEG should also be investigated and stability tested (**Scheme 5.5**).

Possible synthetic strategies that may be adopted in the future for the synthesis of the single tagged peptide **251** could concentrate on incorporating PFP-tagged amino acids into the peptide during the synthesis. An Fmoc-protected PFP-tagged Cys amino acid could be synthesised and manually coupled in the same position as peptide **250** (**Scheme 5.6**).
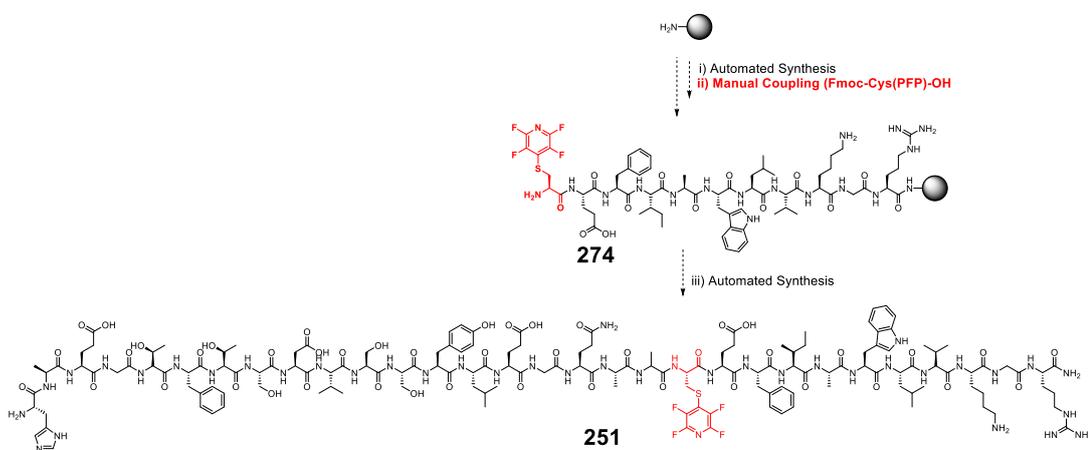
Scheme 5.6: Synthesis of peptide **251** utilising automated and manual SPPS conditions.

If the stability of the single tagged peptide **251** is not significantly better than native GLP-1 **245**, there are other analogues which could be synthesised, and their stabilities tested. By adopting non-selective tagging conditions (as discussed in **Section 4.3.1**), it should be possible to tag GLP-1 at multiple nucleophilic sites (**Scheme 5.7**). Using non-selective tagging to react PFP at Tyr, Ser, Lys and Cys residues may result in a peptide more stable to enzymatic degradation, but possibly less biologically active due to modification at multiple sites.

Scheme 5.7: Non-selective tagging of peptide **247** with PFP.

References

[1] G. E. Mulder, J. A. W. Kruijtzer and R. M. J. Liskamp, *Chem. Commun.*, 2012, **48**, 10007-10009.

Chapter 6: Experimental Details and Characterisation

6.1 General Procedures

All reagents were purchased from Sigma-Aldrich unless stated otherwise. Peptide synthesis grade DMF was purchased from Fisher Scientific and amino acids purchased from Novabiochem(Merck) or Iris Biotech. PyBOP® was purchased from Apollo Scientific (Stockport, UK) and resins purchased from Novabiochem. The general procedure for Fmoc SPPS synthesis is detailed in the following sections. All reactions were carried out in fritted polypropylene reaction vessels and DMF used as a solvent. The amino acids were protected in the following manner, Fmoc-Thr(O^{t}Bu)-OH, Fmoc-Arg(Pbf)-Oh, Fmoc-Ser(O^{t}Bu)-OH Fmoc-Cys(Trt)-OH, Fmoc-Cys(Acm)-OH, Fmoc-His(Trt)-OH, Fmoc-Gln(Trt)-OH, Fmoc-Glu(O^{t}Bu)-OH, Fmoc-Trp(Pbf)-OH, Fmoc-Asp(O^{t}Bu) and Fmoc-Lys(Boc)-OH.

6.2 Analytical Equipment and Methods

6.2.1 Analytical liquid chromatography mass spectroscopy (LCMS)

Analytical LCMS was carried out using an Acquity UPLC system (Waters Ltd, UK) with a photodiode array detector. Samples were injected onto an Acquity UPLC BEH C8 column and a gradient of 5–95% B (solvent A = H$_2$O, 0.1% TFA; B = MeCN, 0.1% TFA) a flow rate of 0.6 ml/min. The solvent flow of the UPLC system was introduced into the electrospray ion source of an Aquity TQD mass spectrometer.

6.2.2 Q-ToF

Q-ToF MS was carried out using a QtoF Premier MS with Acquity UPLC system (Waters Ltd, UK). The sample was injected into a 0.2 ml/min flow of MeCN which was introduced *via* electrospray ion source.

6.2.3 Matrixassisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS)

MALDI-TOF mass analysis was carried out using an Autoflex II ToF/ToF mass spectrometer (Bruker Daltonik GmBH) with a 337 nm nitrogen laser. Peptides were dissolved in deionized H$_2$O/MeCN for analysis. Sample solution (1 mg/ml) was mixed with matrix solution (cyano-4-hydroxcinnamic acid, 50 mg/mL) in a ratio of 1:9. MS data was processed using FlexAnalysis 2.0 (Bruker Daltonik GmBH).

6.2.4 Nuclear Magnetic Resonance (NMR) Spectroscopy

NMR spectra were carried out using Bruker Avance 400 MHz. Multiplicities s= singlet, d= doublet, dd= doublet of doublets, t= triplet, quint= quintet, m= multiplet. NMR solvent peaks: $CDCl_3$ (^{1}H NMR 7.26 ppm, ^{13}C NMR 77.0 ppm), CD_3CN (^{1}H NMR 1.94 ppm, ^{13}C NMR 1.3, 118.3 ppm) and D_2O (^{1}H NMR 4.79 ppm).

6.3 Automated Solid Phase Peptide Synthesis

6.3.1 Amino Acid Loading

Loading of the intial amino acid onto the resin was carried out in a fritted polypropylene reaction chamber. For every 0.1 mmol of resin, 2 ml of DMF was used for resin swelling. After 2 h, DMF was removed and coupling/deprotection steps continued.

6.3.2 Automated SPPS

Automated SPPS was carried using a CEM Liberty1 microwave peptide synthesizer with Discover microwave unit. All reactions were carried out using a 30 ml PTFE reaction vessel, with microwave heating and agitation carried out *via* nitrogen gas. Couplings were carried out using Fmoc-protected amino acid (5 eq), activator (10 eq) and DIPEA (20 eq). For double and triple couplings, the reaction vessel was drained after each cycle and fresh reagents were added.

6.3.4 Peptide Couplings

Peptide couplings were carried out at rt for 30 min followed by 10 min microwave assisted at 75 °C and 25 W power. Arginine residues were double coupled, with 30 min at rt and 10 min at 75 °C. Leucine couplings were carried at rt for 60 min with 10 min at 75 °C. Removal of the Fmoc group was carried out using a solution of 20% piperidine in DMF for 5 + 10 min.

6.3.5 Resin Cleavage

Peptide-resin was incubated with a cleavage cocktail of TFA 95%, H_2O 2.5% and TIPS 2.5% for 4 h at rt. The resin was removed by filtration and the solvent removed under reduced pressure before precipitation using ether and decanting of the liquid. The resulting solid peptide was dissolved in deionized water and lyophilized.

6.4 Manual Solid Phase Peptide Synthesis Procedure

6.4.1 Peptide Coupling

Single couplings were carried out using 5 eq of amino acid (compared to resin), 5 eq of PyBOP® , 10 eq of DIPEA and 2 ml of DMF under agitation for 1 h. Unless stated otherwise, double peptide couplings were carried out: 2×1 h couplings for each residue and reaction drained after each coupling and fresh reagent added. After each amino acid coupling, the reaction solution was drained, and resin washed with 5×2 ml DMF. Removal of the Fmoc group was carried out using a solution of 20% piperidine in DMF for 5 + 10 min under agitation. Piperidine solution was drained and resin rinsed using 5×2 ml of DMF.

6.4.2 HATU Peptide Coupling

Single couplings were carried out using 5 eq of amino acid (compared to resin), 5 eq of HATU, 10 equivalents of DIPEA and 2 ml of DMF under agitation for 1 h. Unless stated otherwise, double peptide couplings were carried out: 2×1 h couplings for each residue and the reaction drained after each coupling and fresh reagent added. After each set of coupling reactions, the reaction solution drained, and resin washed with 5×2 ml portions of DMF. Removal of the Fmoc group was carried out using a solution of 20% piperidine in DMF for 5 + 10 min under agitation. Piperidine solution was drained and resin rinsed using 5×2 ml of DMF.

6.4.3 Resin Cleavage

Peptide-resin was incubated with a cleavage cocktail of TFA 95%, H_2O 2.5% and TIPS 2.5% for 4 h at rt. The resin was removed by filtration and the solvent removed under reduced pressure before precipitation using ether and decanting of the liquid. The resulting solid peptide was dissolved in deionized water and lyophilized.

6.4.4 Acm protecting group removal

Solid peptide was dissolved in 0.5 ml of 40% AcOH/DMF solution in an Eppendorf tube. To the peptide solution solid iodine (25 eq) was added and left under agitation at rt. After 45 min, 1M ascorbic acid solution was added dropwise until the reaction mixture turned colourless. The peptide was fully dissolved in MeCN/H_2O to a total volume of 2.0 ml for purification and analysis.

6.5 Tagging Reactions

6.5.1 Selective Solution Phase Tagging Reaction

Solid peptide was dissolved in TFE (0.5 ml) in a 1.5 ml plastic Eppendorf tube, to which DIPEA (2.5 eq) was added. Pentafluoropyridine (10 eq) was added and the tube was shaken at rt for 4.5 h. After removal of volatiles under reduced pressure, the sample was dissolved in a 1:1 mixture of H_2O/MeCN (1 ml) and analysed by Q-ToF-MS.

6.5.2 Non-Selective Solution Phase Tagging Reaction

Solid peptide was dissolved in DMF (0.5 ml) in a 1.5 ml plastic Eppendorf tube, to which DIPEA (2.5 eq) was added. Pentafluoropyridine (10 eq) was added and the tube was shaken at rt for 4.5 h. After removal of volatiles under reduced pressure, sample dissolved in a 1:1 mixture of H_2O/MeCN (1 ml) and analysed by Q-ToF-MS.

6.5.3 On-resin Tagging Reaction

Peptide-resin was treated with DMF (2 ml), DIPEA (2.5 eq) and PFP (10 eq) for 4.5 h under agitation. After 4.5 h, the peptide-resin was filtered from reaction solution and rinsed with DMF (5 × 2 ml).

6.6 Displacement Reactions

To a solution of tagged peptide in MeCN (0.5 ml) and water (0.5 ml) in a 1.5 ml plastic Eppendorf tube, DIPEA (2.5 eq) was added. Sulphur nucleophiles were added (5 eq) and the tube shaken at rt for 4 h. After removal of volatiles under reduced pressure, each reaction mixture dissolved in a mixture of H_2O/MeCN (1 ml) and analysed by Q-ToF-MS.

6.7 Purification

6.7.1 Preparative High-Pressure Liquid Chromatography

Samples were dissolved in H_2O/MeCN mix, centrifuged and injected onto a Speck and Burke Analytical C18 Column. Purification of peptides was carried out using Perkin Elmer Series 200 LC pump and 785A UV/Vis detector. A gradient of 0-100% solvent B (solvent A= 95% H_2O, 5% MeCN, 0.1% TFA, solvent B = 95% MeCN, 5% H_2O, 0.1% TFA) with a flow rate of 2.0 ml/min was used and absorbance data collected at 220 nm, unless stated otherwise.

## 6.7.2 Analytical High-Pressure Liquid Chromatography

Samples (10 µl) were injected onto a Waters XBridge BEH C18 Column using a Perkin Elmer Series 200 Autmosampler and Perkin Elmer Series 200 LC Pump and detected using a Series 200 UV/Vis detector. A gradient of 0-100% solvent B (solvent A= 95% H_2O, 5% MeCN, 0.05% TFA, solvent B = 95% MeCN, 5% H_2O, 0.03% TFA) with a flow rate of 1 ml/min was used and absorbance data collected at 220 nm and processed using TotalChem software.

6.8 Stability Testing

Peptide was dissolved in Tris buffer (pH 7.6) resulting in a solution with a concentration of 1 mg/ml in a 0.5 ml Eppendorf tube. α-Chymotryspin from bovine pancreas was dissolved in Tris buffer (pH 7.6) to a concentration of 1 mg/ml and added to the peptide solution so that the resulting ratio of peptide to enzyme solution is 10:1. The Eppendorf was incubated at 23 °C using a Clifton NE5 Series Water Bath and 10 µl aliquots removed at t = 0, 1, 2, 3, 4, 5, 6 and 24 h. To each 10 µl aliquot, 30% TFA/MeCN solution was added and injected onto a Waters XBridge BEH C18 Column using a Perkin Elmer Series 200 Autosampler and Perkin Elmer Series 200 LC Pump and detected using a Series 200 UV/Vis detector. A gradient of 0-100% solvent B (solvent A= 95% H_2O, 5% MeCN, 0.05% TFA, solvent B = 95% MeCN, 5% H_2O, 0.03% TFA) over 30 min with a flow rate of 1 ml/min was used and absorbance data collected at 280 nm and processed using TotalChem software.

6.9 Synthesis of Dipeptide Analogues

6.9.1 Fmoc-Cys(Trt)-Ala-OMe (175)

175
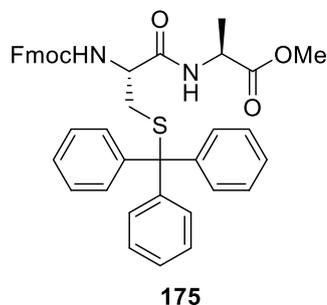
L-Alanine methyl ester hydrochloride (200 mg, 1.44 mmol) was dissolved in DCM (30 ml). Fmoc-L-Cys(trt)-OH (842 mg, 1.44 mmol) and PyBOP (749 mg, 1.44 mmol) were added and stirred. NMM (0.37 ml, 4.32 mmol) was added drop-wise and the reaction mixture stirred at rt for 18 h. After 18 h, DCM was removed under reduced pressure and the resulting yellow oil purified *via* column chromatography to yield **175** as a white solid (816 mg, 84%). δ_H (400 MHz, CDCl$_3$) 1.37 (d, 3H, $J = 8.0$ Hz), 2.75 (dd, 2H, $J = 8.0$ Hz, 12.0 Hz), 3.72 (s, 3H), 4.20 (t, 1H, $J = 8.0$ Hz), 4.51 (q, 1H, $J = 8.0$ Hz), 7.22- 7.77 (m, 28H). δ_C (100 MHz, CDCl$_3$) 18.2, 47.0, 48.2, 52.4, 67.3, 199.9, 125.0, 127.0, 128.1, 129.6, 141.3, 143.6, 144.3, 169.5, 172.7. LCMS found *m/z* 692.1 [M+Na]$^+$, calcd 670.2 [M+H]$^+$ Experimental data agrees with literature data.[1]

6.9.2 Ac-Cys(Trt)-Ala-OMe (177)

177
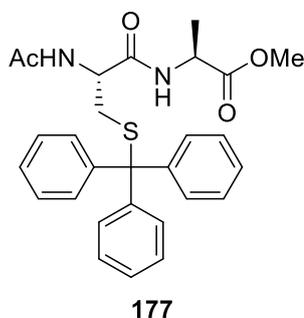
Fmoc protected dipeptide (335 mg, 0.5 mmol) was dissolved in THF (7.5 ml) at rt and under argon. Piperidine (1.5 ml) was added and the reaction was run for 3 h at rt and under argon. The reaction was traced by TLC. Once the reaction had reached completion solvent was removed under reduced pressure to yield an off-white solid.

This crude deprotected dipeptide was directly used in the next stage of the synthesis and was dissolved in DCM (7.5 ml) and cooled to 0 °C. Tritheylamine (0.08 ml, 0.55 mmol) and acetyl chloride (0.04 ml, 0.5 mmol) were added and the reaction mixture warmed to rt and stirred for 1 h under argon. The reaction was monitored by TLC and once it had reached completion solvent was removed to yield **177** an off-white solid.

The crude product was purified *via* column chromatography (100% EtOAc to 10% MeOH in EtOAc). NMR of the collected fractions was carried out and it was determined the desired material **177** was present (207 mg, 0.42 mmol, 85%). δ_H (400

MHz, CDCl$_3$) 1.41 (d, 3H, J = 8.0 Hz), 2.10 (s, 3H), 2.97-3.14 (m 2H), 3.27 (s, 3H), 3.56 (t, 1H, J = 4.0 Hz), 3.95 (q, 1H, J = 8.0 Hz), 6.21 (s, 1H), 7.23-7.79 (m, 14H), δ_C (100 MHz, CDCl$_3$) 14.2, 19.7, 24.4, 35.4, 50.7, 53.1, 67.5, 127.1, 128.2, 129.5, 142.1, 166.5, 167.9. LCMS mass: found m/z 491.1 [M+H], calcd 490.1 [M+H]$^+$

6.9.3 Ac-Cys-Ala-NH$_2$ (180)

180
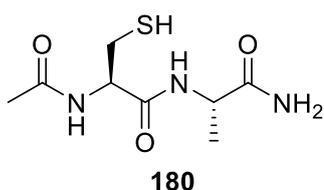
Compound **180** was synthesised *via* manual Fmoc-SPPS at rt on a 0.3 mmol scale (according to **Section 6.4.1**) on RINK amide resin (0.82 mg/mmol loading). All couplings were double couplings and carried out using 5 eq of amino acid. Capping and cleavage of peptide from the resin was carried out following the procedure detailed in **Section 6.4.3**. ^{1}H NMR δ_H (400 MHz, CD$_3$CN) 1.34 (d, 3H, J = 4.0 Hz, CH$_3$), 1.97 (s, 3H, CH$_3$), 3.77 (d, 2H, J = 4.0 Hz, CH$_2$), 4.29- 4.40 (m, 1H, H-α), 4.44-4.74 (m, 1H, H-α), LCMS mass: found m/z [M+Na]$^+$ 256.1.

6.9.4 Ac-Pen-Ala-NH$_2$ (183)

183
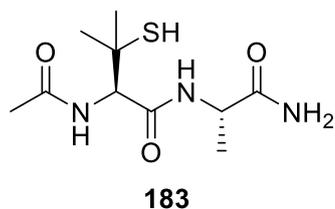
Compound **183** was synthesised *via* manual Fmoc-SPPS at rt on a 0.2 mmol scale (according to **Section 6.4.1**) on RINK amide resin (0.82 mg/mmol loading). All couplings were double couplings and carried out using 5 eq of amino acid. Capping and cleavage of peptide from the resin was carried out following the procedure detailed in **Section 6.4.3**. ^{1}H NMR δ_H (400 MHz, CD$_3$CN) 1.31 (d, 3H, J = 4.0 Hz, CH$_3$), 1.36 (s, 6H, CH$_3$ ×2), 2.01 (s, 3H, CH$_3$), 4.25-4.37 (m, 1H, H-α), 4.43- 4.58 (m, 1H, H-α), LCMS mass: found m/z [M+H]$^+$ 262.8.
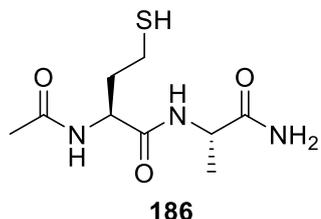
6.9.5 Ac-hCys-Ala-NH₂ (186)

Compound **186** was synthesised *via* manual Fmoc-SPPS at rt on a 0.2 mmol scale (according to **Section 6.4.1**) on RINK amide resin (0.82 mg/mmol loading). All couplings were double couplings and carried out using 5 eq of amino acid. Capping and cleavage of peptide from the resin was carried out following the procedure detailed in **Section 6.4.3**. ^{1}H NMR δ_H (400 MHz, CD₃CN) 1.33 (d, 3H, J = 4.0 Hz, CH₃), 1.97 (s, 3H, CH₃), 2.12-2.15 (m, 2H, CH₂), 2.55- 2.61 (m, 2H, CH₂), 4.23- 4.37 (m, 2H, H-α), LCMS mass: found m/z [M+H]$^+$ 248.0.

6.9.6 Ac-Cys(PFP)-Ala-NH₂ (187)
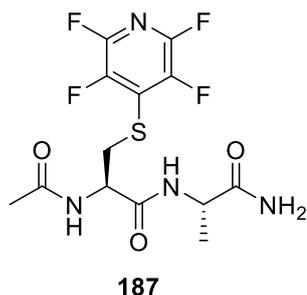
Compound **187** was synthesised using the reaction conditions described in **Section 6.5.2** using 192 mg of **177**. The resulting crude material was purified using column chromatography to yield **187** as a white solid (62 mg, 20%) δ_H (400 MHz, CD₃CN) 1.30 (dd, 3H, J = 7.1, 2.9 Hz, CH₃), 1.91 (d, 3H, J = 4.2 Hz, CH₃), 3.69- 3.40 (m, 2H, CH₂), 4.27- 4.14 (m, 1H, H-α), 4.53 (td, 1H, J = 7.8, 5.2 Hz, H-α), 6.34 (s, 1H, NH), 6.90 (s, 1H, NH), δ_C (100 MHz, CD₃CN) 17.3, 21.9, 34.5, 48.9, 48.9, 53.0, 117.3, 168.9, 170.4, 173.5, 175.3, δ_F (376 MHz, CD₃CN) -94.0 (ddd, 1F, J = 7.5 Hz, 15.0 Hz, 30.0 Hz), -94.1 (ddd, 1F, J = 7.5 Hz, 15.0 Hz, 30.0 Hz), -137.9 (ddd, 1F, J = 7.5 Hz, 11.2 Hz, 26.3 Hz), -138.0 (ddd, 1F, J = 7.5 Hz, 11.2 Hz, 26.3 Hz). Acc Mass: found m/z [M+H]$^+$ 383.0815, calcd 383.0801.
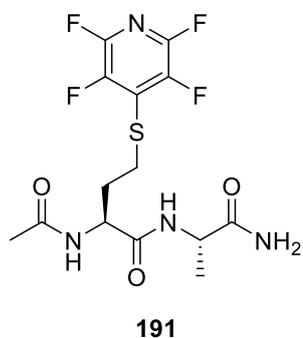
213

6.9.7 Ac-Pen(PFP)-Ala-NH₂ (190)
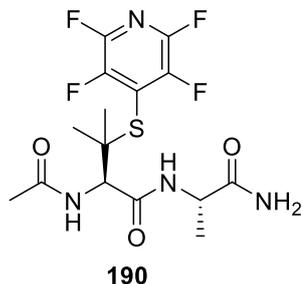
Compound **190** was synthesised using the reaction conditions described in **Section 6.5.2** using 125 mg of **183**. The resulting crude material was purified using column chromatography to yield **190** as a white solid (23 mg, 13%) δ_H (400 MHz, CD$_3$CN) 1.37 (d, 3H, J = 8.0 Hz, CH$_3$), 1.43 (s, 3H, CH$_3$), 1.49 (s, 3H, CH$_3$), 2.02 (s, 3H, CH$_3$), 4.29 (q, 1H, J = 8.0 Hz, H-α), 4.53 (s, 1H, H-α). 5.94 (s, 1H, NH), 6.49 (s, 1H, NH), δ_C (100 MHz, CD$_3$CN) 17.4, 24.3, 24.5, 24.9, 46.5, 50.2, 57.6, 142.2, 154.3, 169.8, 175.2, δ_F (376 MHz, CD$_3$CN) -92.6 (ddd, 1F, J = 7.5 Hz, 15.0 Hz, 33.5 Hz), -92.7 (ddd, 1F, J = 7.5 Hz, 15.0 Hz, 33.5 Hz) -132.7 (ddd, 1F, J = 7.5 Hz, 15.0 Hz, 33.5 Hz) -132.9 (ddd, 1F, J = 7.5 Hz, 15.0 Hz, 33.5 Hz), Acc Mass: found m/z [M+H]$^+$ 411.1100, calcd 411.1103.

6.9.8 Ac-hCys(PFP)-Ala-NH₂ (191)

Compound **191** was synthesised using the reaction conditions described in **Section 6.5.2** on a using 143 mg of **186**. The resulting crude material was purified using column chromatography to yield **191** as a white solid (46 mg, 20%) δ_H (400 MHz, CD$_3$CN), 1.30 (d, 3H, J = 8.0 Hz, CH$_3$), 1.95 (s, 1H, CH$_3$), 2.14 (q, 2H, J = 8.0 Hz, CH$_2$), 3.25-3.43 (m, 2H, CH$_2$), 4.25 (t, 1H, J = 8.0 Hz, H-α), 4.34-4.74 (m, 1H, H-α), 6.52 (s, 1H, NH), 6.92 (d, 1H, J = 4.0 Hz, NH), δ_C (100 MHz, CD$_3$CN) 17.0, 22.0, 29.6, 31.9, 48.8, 52.5, 170.6, 171.3, 174.7 δ_F (376 MHz, CD$_3$CN) -94.1 (ddd, 1F, J = 7.5 Hz, 15.0 Hz, 30.0 Hz), -94.3 (ddd, 1F, J = 7.5 Hz, 15.0 Hz, 30.0 Hz), -139.1 (ddd, 1F, J = 7.5 Hz, 11.2 Hz, 30.0 Hz), -139.2 (ddd, 1F, J = 7.5 Hz, 11.2 Hz, 30.0 Hz), Acc Mass: found m/z [M+H]$^+$ 397.0948, calcd 397.0958.

6.9.9 Ac-Cys(SPh)-Ala-NH$_2$ (194)
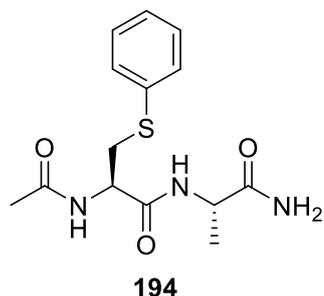
Compound **194** was synthesised using the reaction conditions described in **Section 6.6** on a using 10 mg of **187** and purified using preparative TLC plates using solvent system EtOAc:MeOH 9:1 to yield **194** as a white solid (8 mg, 13%); δ_H (400 MHz, CD$_3$CN) 1.30 (d, 3H, J = 4.0 Hz, CH$_3$), 1.92 (s, 3H, CH$_3$), 3.43 (dd, 1H, J = 8.0, 12.0 Hz, CH$_2$), 3.64 (dd, 1H, J = 4.0 Hz, 16.0 Hz, CH$_2$), 4.20- 4.42 (m, 1H, H-α), 4.55- 4.62 (m, 1H, H-α), 6.42 (s, 1H, NH), 7.04- 7.75 (m, 5H, CH-aromatic), δ_C (100 MHz, CD$_3$CN) 17.3, 21.9, 34.8, 48.9, 53.0, 127.56, 127.9, 128.0, 129.3, 129.4, 129.5, 135.2, 168.9, 170.4, 173.9; Acc Mass: found m/z [M+H]$^+$ 310.1233, calcd 310.1232
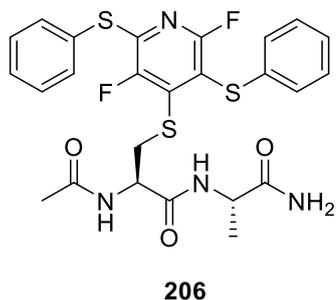
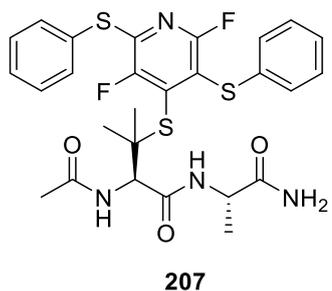### 6.9.10 Ac-Cys(SPh)$_2$-Ala-NH$_2$ (206)

Compound **206** was synthesised using the reaction conditions described in **Section 6.6** on a using 10 mg of **187** and purified using preparative TLC plates using solvent system EtOAc:MeOH 9:1 to yield **206** as a white solid (0.8 mg, 4%); δ_H (400 MHz, CD$_3$CN) 1.27 (d, 3H, J = 4.0 Hz, CH$_3$), 1.88 (s, 3H, CH$_3$), 3.50 (dd, 1H, J = 4.0 Hz, 8.0 Hz, CH$_2$), 3.61 (dd, 1H, J = 4.0 Hz, 8.0 Hz, CH$_2$) 4.20 (q, 1H, J = 4.0 Hz, H-α), 4.45 (dd, 1H, J = 4.0 Hz, 8.0 Hz, H-α), 7.15- 7.58 (m, 10H, CH-aromatic), δ_F (376 MHz, CD$_3$CN) -66.5 (d, 1F, J = 26.0 Hz), -199.4 (d, 1F, J = 30.0 Hz) Acc Mass: found m/z [M+H]$^+$ 563.1064, calcd 563.1079.

6.9.11 Ac-Pen(SPh)₂-Ala-NH₂ (207)

Compound **207** was synthesised using the reaction conditions described in **Section 6.6** using 10 mg of **190** and purified using preparative TLC plates using solvent system EtOAc:MeOH 9:1 to yield **207** as a white solid (4 mg, 33.5%); δ_H (400 MHz, CD₃CN) 1.30 (s, 6H, CH₃), 1.46 (s, 3H, CH₃), 1.95 (s, 3H, CH₃), 4.25 (q, 1H, J = 8.0 Hz, H-α), 4.50 (d, 1H, J = 8.0 Hz, H-α), 7.08 (d, 1H, J = 4.0 Hz, NH), 7.16- 7.59 (m, 10H, CH-aromatic), 7.80 (d, 1H, J = 4.0 Hz, NH), δ_C (100 MHz, CD₃CN) 16.9, 22.0, 48.9, 60.0, 127.0, 127.6, 128.0, 129.1, 129.4, 129.6, 129.9, 135.2, 168.9, 171.9, 175.5, δ_F (376 MHz, CD₃CN) -62.9 (d, 1F, J = 30.0 Hz), -110.8 (d, 1F, J = 26.0 Hz), Acc Mass: found m/z [M+H]⁺ 591.1373, calcd 591.1370.
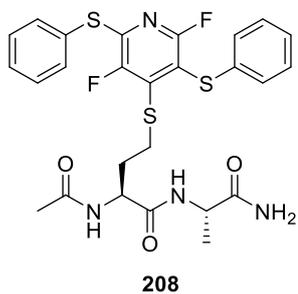
6.9.12 Ac-hCys(SPh)₂-Ala-NH₂ (208)

Compound **208** was synthesised using the reaction conditions described in **Section 6.6** on a using 10 mg of **191** and purified using preparative TLC plates using solvent system EtOAc:MeOH 9:1 to yield **208** as a white solid (3 mg, 20%); δ_H (400 MHz, CD₃CN) 1.30 (d, 3H, J = 8.0 Hz, CH₃), 1.94 (s, 3H, CH₃), 3.19 (t, 2H, J = 8.0 Hz, CH₂), 4.22 (q, 1H, J = 4.0 Hz, H-α), 4.29 (dd, 1H, J = 4.0 Hz, 8.0 Hz, H-α), 7.15- 7.58 (m, 10H, CH-aromatic), δ_C (100 MHz, CD₃CN) 17.0, 21.9, 48.9, 52.6, 129.9, 127.8, 129.5, 129.6, 129.9, 135.9, δ_F (376 MHz, CD₃CN) -66.9 (d, 1F, J = 30.0 Hz), -121.1 (d, 1F, J = 26.0 Hz), Acc Mass: found m/z [M+H]⁺ 577.1229, calcd 577.1202.

## 6.10 Synthesis of Native OT and VP and PFP-tagged OT and VP analogues

6.10.1 Peptide 209

209
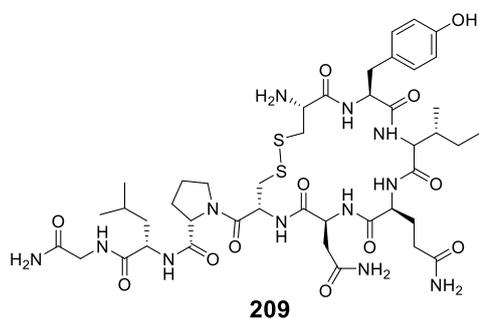
Sequence **CYIQNCPLG**-NH$_2$ was synthesised *via* manual Fmoc-SPPS at rt on a 0.1 mmol scale on RINK amide resin (0.82 mg/mmol loading). Amino acids CYIQNCPLG were coupled using 60 min single couplings using the reaction condition described in **Section 6.4.1**. Cleavage of peptide from the resin and removal of the Acm group/cyclisation were carried out according to **Section 6.4.3** and **6.4.4**. Accurate mass: found *m/z* 1007.4471 [M+H]$^+$, calcd 1006.4365, analytical HPLC retention time 10.9 min. Analytical HPLC 100% solvent A-100% solvent B gradient over 20 min, observed at 220 nm.

6.10.2 Peptide 220

220
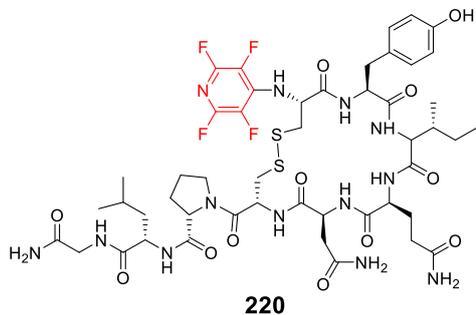
Sequence **(PFP)CYIQNCPLG**-NH$_2$ was synthesised *via* manual Fmoc-SPPS at rt on a 0.1 mmol scale on RINK amide resin (0.82 mg/mmol loading). Amino acids CYIQNCPLG were coupled using 60 min single couplings using the reaction condition described in **Section 6.4.1**. Tagging of the peptide on resin was carried out using the reaction conditions described in **Section 6.5.3**. Cleavage of peptide from the resin and removal of the Acm group/cyclisation were carried out according to **Section 6.4.3** and **6.4.4**. δ_F (376 MHz, CD$_3$CN) -96.2 (m, 2F), -161.9 (m, 2F), Accurate mass: found *m/z* 1156.4331 [M+H]$^+$, calcd 1155.4253, analytical HPLC retention time 15.5 min. Analytical HPLC 100% solvent A-100% solvent B gradient over 20 min, observed at 220 nm.

6.10.3 Peptide 221

221

Sequence **(PFP)CY(PFP)IQNCPLG**-NH$_2$ was synthesised *via* manual Fmoc-SPPS at rt on a 0.1 mmol scale on RINK amide resin (0.82 mg/mmol loading). Amino acids CYIQNCPLG were coupled using 60 min single couplings using the reaction condition described in **Section 6.4.1**. Tagging of the peptide in solution was carried out using the reaction conditions described in **Section 6.5.2**. Cleavage of peptide from the resin and removal of the Acm group/cyclisation were carried out according to **Section 6.4.3** and **6.4.4**. δ_F (376 MHz, CD$_3$CN) -91.4 (m, 2F), -96.3 (m, 2F), -156.0 (m, 2F), -161.9 (m, 2F), Accurate mass: found *m/z* 1305.4235 [M+H]$^+$, calcd 1304.4124, analytical HPLC retention time 20.4 min. Analytical HPLC 100% solvent A-100% solvent B gradient over 20 min, observed at 220 nm.
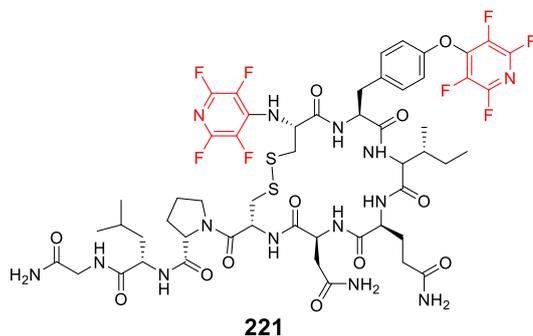
6.10.4 Peptide 210

210

Sequence **CYFQNCPRG**-NH$_2$ was synthesised was synthesised *via* manual Fmoc-SPPS at rt on a 0.1 mmol scale on RINK amide resin (0.82 mg/mmol loading). Amino acids CYFQNCPRG were coupled using 60 min single coupl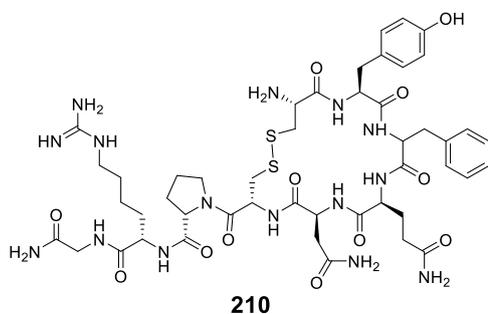ings using the reaction condition described in **Section 6.4.1**. Cleavage of peptide from the resin and removal of the Acm group/cyclisation were carried out according to **Section 6.4.3** and **6.4.4**. Accurate mass: found *m/z* 1084.4453 [M+H]$^+$, calcd 1084.2390, analytical HPLC retention time 9.8 min. Analytical HPLC 100% solvent A-100% solvent B gradient over 20 min, observed at 220 nm.

6.10.5 Peptide 222

222

Sequence **(PFP)CYFQNCPRG**-NH$_2$ was synthesised *via* manual Fmoc-SPPS at rt on a 0.1 mmol scale on RINK amide resin (0.82 mg/mmol loading). Amino acids CYFQNCPRG were coupled using 60 min single couplings using the reaction condition described in **Section 6.4.1**. Tagging of the peptide on resin was carried out using the reaction conditions described in **Section 6.5.3**. Cleavage of peptide from the resin and removal of the Acm group/cyclisation were carried out according to **Section 6.4.3** and **6.4.4**. δ_F (376 MHz, CD$_3$CN) -96.2 (m, 2F), -191.9 (m, 2F), Accurate mass: found *m/z* 1233.4437 [M+H]$^+$, calcd 1232.4267, analytical HPLC retention time 14.5 min. Analytical HPLC 100% solvent A-100% solvent B gradient over 20 min, observed at 220 nm.

6.10.6 Peptide 223

223

Sequence **(PFP)CY(PFP)FQNCPRG**-NH$_2$ was synthesised *via* manual Fmoc-SPPS at rt on a 0.1 mmol scale on RINK amide resin (0.82 mg/mmol loading). Amino acids CYFQNCPRG were coupled using 60 min single couplings using the reaction condition described in **Section 6.4.1**. Tagging of the peptide in solution was carried out using the reaction conditions described in **Section 6.5.2**. Cleavage of peptide from the resin and removal of the Acm group/cyclisation were carried out according to **Section 6.4.3** and **6.4.4**. δ_F (376 MHz, CD$_3$CN) -91.3 (m, 2F), -96.2 (m, 2F), -155.9 (m, 2F), -161.9 (m, 2F), Accurate mass: found *m/z* 1381.4156 [M+H]$^+$, calcd 1382.4240, analytical HPLC retention time 19.1 min. Analytical HPLC 100% solvent A-100% solvent B gradient over 20 min, observed at 220 nm.

6.11 Synthesis of GLP-1 and GLP-1 analogues

### 6.11.1 Peptide 245
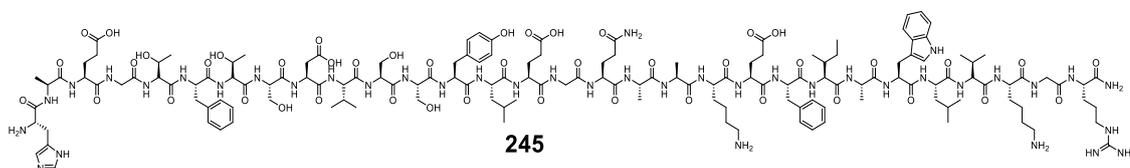
245

Sequence **HAEGTFTSDVSSYLEGQAAKEFIAWLVKGR**-NH$_2$ was synthesised *via* automated Fmoc-SPPS at rt on a 0.05 mmol scale (according to **Section 6.3**) on RINK Amide resin (0.82 mg/mmol loading). Amino acids EFIAWLVKGR were coupled using 30 min single couplings + 10 min mW, amino acids TFTSDVSSYLEGQAAK were coupled using 60 min single couplings + 10 min mW and amino acids HAEG were coupled using double 60 min couplings + 10 min mW. Cleavage of peptide from the resin was carried out following the procedure detailed in **Section 6.3.5** and the peptide purified using preparative HPLC as described in **Section 6.7.1**. MALDI-ToF mas: found *m/z* 3296.4 [M+H]$^+$ analytical HPLC retention time 16.1 min. Analytical HPLC 100% solvent A- 100% solvent B gradient over 20 min, observed at 220 nm.
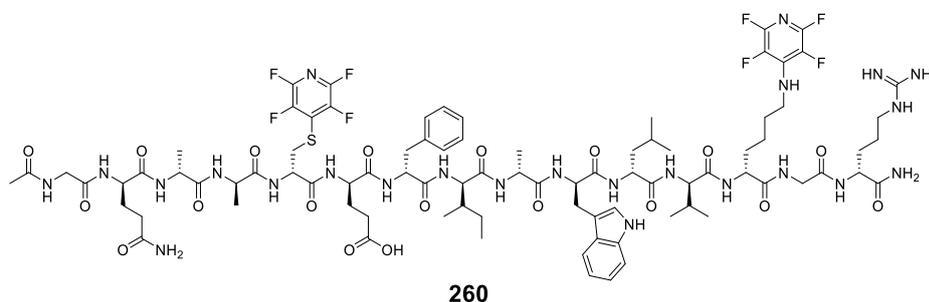
6.11.2 Peptide 259

259
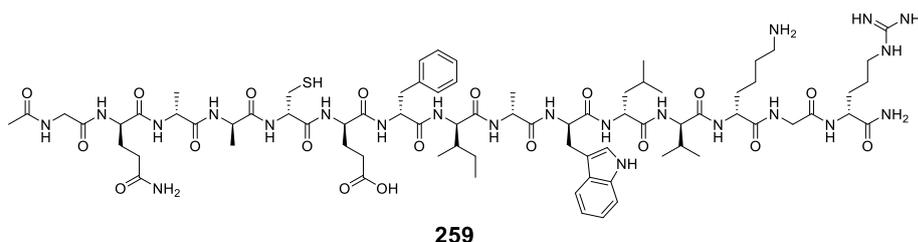
Sequence **GQAACEFIAWLVKGR**-NH$_2$ was synthesised *via* automated Fmoc-SPPS at rt on a 0.1 mmol scale (according to **Section 6.3**) on RINK Amide resin (0.82 mg/mmol loading). Amino acids GQAAEFIAWLVKGR were coupled using 30 min single couplings + 10 min mW. For the coupling of C, HATU manual coupling conditions were adopted according to **Section 6.4.2**. Cleavage of peptide from the resin was carried out following the procedure detailed in **Section 6.3.5** and the peptide purified using preparative HPLC as described in **Section 6.7.1**. LCMS mass: found *m/z* 1690.2 [M+H]$^+$.

6.11.3 Peptide 261
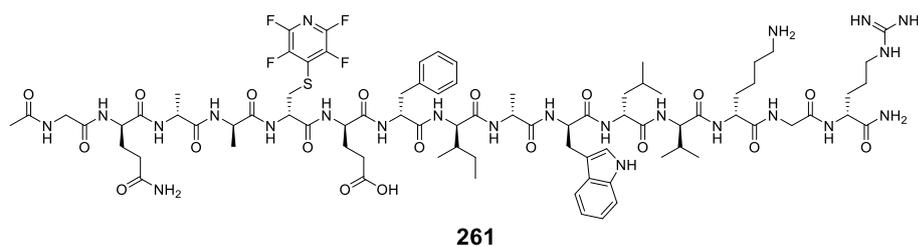
261

Sequence **GQAAC(Pfp)EFIAWLVKGR**-NH$_2$ was synthesised using peptide **259** (0.05 mmol) and following the procedure for selective tagging described in **Section 6.5.1** and the tube was shaken at rt for 4.5 h. After removal of volatiles under reduced pressure, dissolved in a 1:1 mixture of H$_2$O and MeCN (1 mL) and purified *via* preparative HPLC according to **Section 6.7.1**. LCMS mass: found 1840.4 [M+H]$^+$, analytical HPLC retention time 20.5 min. Analytical HPLC 100% solvent A- 100% solvent B gradient over 20 min, observed at 220 nm.

6.11.4 Peptide 260

260

Sequence **GQAAC(Pfp)EFIAWLVKGR**-NH$_2$ was synthesised using peptide **259** (0.05 mmol) and following the procedure for non-selective tagging described in **Section 6.5.2** and the tube was shaken at rt for 4.5 h. After removal of volatiles under vacuum, dissolved in a 1:1 mixture of H$_2$O and MeCN (1 ml) and purified *via* preparative HPLC according to **Section 6.7.1**. LCMS mass: found 1970. 0 [M+H]$^+$.

6.11.5 Peptide 263

263
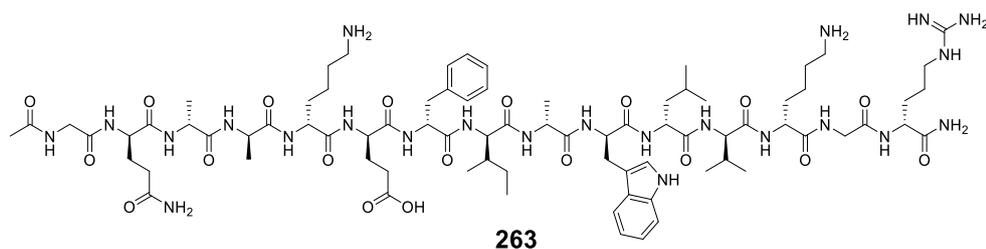
Sequence **GQAAKEFIAWLVKGR**-NH$_2$ was synthesised *via* automated Fmoc-SPPS at rt on a 0.1 mmol scale (according to **Section 6.3**) on RINK Amide resin (0.82 mg/mmol loading). Amino acids GQAAEFIAWLVKGR were coupled using 30 min single couplings + 10 min mW. Cleavage of peptide from the resin was carried out following the procedure detailed in **Section 6.3.5** and the peptide purified using preparative HPLC as described in **Section 6.7.1**. LCMS mass: found *m/z* 1715.3 [M+H]$^+$, analytical HPLC retention time 15.4 min. Analytical HPLC 100% solvent A-100% solvent B gradient over 20 min, observed at 220 nm.

6.11.6 Peptide 250
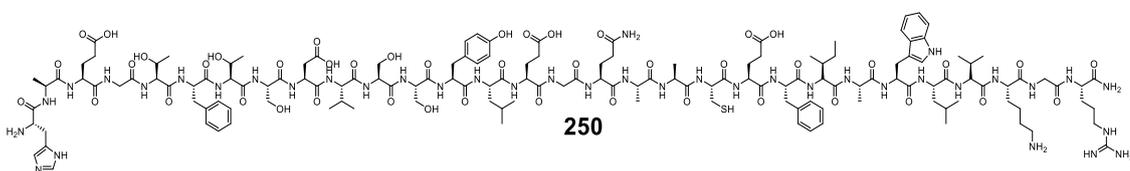
250

Sequence **HAEGTFTSDVSSYLEGQAACEFIAWLVKGR**-NH$_2$ was synthesised *via* automated Fmoc-SPPS at rt on a 0.05 mmol scale (according to **Section 6.3**) on RINK Amide resin (0.82 mg/mmol loading). Amino acids EFIAWLVKGR were coupled using 30 min single couplings + 10 min mW, amino acid C was coupled manually using reaction condition described in **Section 6.4.2** and using HATU, amino acids TFTSDVSSYLEGQAA were coupled using 60 min single couplings + 10 min mW, amino acids AEG were coupled using double 60 min couplings +10 min mW and amino acid H was coupled manually using HATU and the conditions described in **Section 6.4.2**. Cleavage of peptide from the resin was carried out following the procedure detailed in **Section 6.3.5** and the peptide purified using preparative HPLC as described in **Section 6.7.1**. MALDI-ToF mass: found *m/z* 3274.0 [M+H]$^+$

References

[1] Z. B. Kaminski, B. Kolesinska, J. Kolesinska, G. Sabatino, M. Chelli, P. Rovero, M. Blaszczyk, M. L. Glowka, A. M. Papini, *JACS*, 2005, **127**, 16912-16920).