

Improving the potency of biomaterials  
for tissue growth

Laetitia Raynal

PhD

University of York

Chemistry

July 2023

## Abstract

In recent years, it has been shown that synthetic biomaterials are not optimal supports for tissue growth due to their lack of intrinsic biochemical signalling potential. To improve the biological potency of these biomaterials and increase the reach of tissue engineering, key signalling proteins can be attached. However, when these proteins are attached via non-specific conjugation chemistries, they suffer a drastic loss of activity. There is therefore a need for site-selective approaches for protein-biomaterial conjugation.

In this PhD, ligand-directed chemistry was studied to achieve this goal, with two new approaches to achieve site selective modification of proteins without the need for protein engineering. The first approach was based on using peptides as protein-binding ligands, while the second approach was based on the use of 2-pyridinecarboxaldehydes as *N*-terminal targeting ligands.

This thesis will first introduce the peptide approach in Chapter 2, with details of the synthesis and modification of the peptides required, then will focus in Chapter 3 on the binding of these modified probes to their protein of interest. In Chapter 4, the use of these binding peptides for the formation of ligand-directed probes will be discussed. Chapter 5 will develop similar chemistry on 2-pyridinecarboxaldehydes, and finally Chapter 6 will describe the results using these two types of probes for the modification of proteins.

Our results demonstrate that we can achieve site-selective modification of model proteins, representing the first traceless modification of protein using pyridinium oxime and *N*-acyl-*N*-sulphonamide (PyOx/NASA) system on a peptide in a traceless manner. With more optimization still needed, these preliminary results are important for the biomedical field in which protein-biomaterial conjugation is key for the formation of *in-vitro* tissue models for the study of various disease such as myocardial infarction, diabetes, and osteoarthritis.

## Table of Contents

<b>Abstract</b> .....	ii
<b>Table of Contents</b> .....	iii
<b>Acknowledgments</b> .....	viii
<b>Declaration</b> .....	x
<b>Glossary</b> .....	xi
<b>Chapter 1 – Introduction</b> .....	1
1.1 Modification of proteins, a general background .....	3
1.1.1 <i>Chemoselective approaches, the use of natural amino acids</i> .....	3
1.1.2. <i>Regioselective approaches, the use of “tags” or ligands</i> .....	14
1.1.3 <i>A conclusion on both chemoselective and regioselective chemistries for protein labelling</i> .....	34
1.2 Proteins of interest .....	37
1.2.1 <i>A clinical need for improved biomaterials</i> .....	37
1.2.2 <i>Fibroblast growth factor 2, FGF-2</i> .....	38
1.2.3 <i>Insulin</i> .....	45
1.3 Overview of the goals of the PhD .....	50
1.4 Bibliography .....	52
<b>Chapter 2 – Peptides, the core of our binding proximity approach: synthesis and modification</b> ....	64
2.1 General introduction of SPPS technique.....	64
2.2 General introduction of the targeted peptides .....	68
2.3 Modification of peptides after SPPS .....	69
2.3.1 <i>Modification using a solution-phase approach</i> .....	69
2.3.2 <i>Modification using a solid-phase approach</i> .....	75
2.4 Orthogonal protection of side-group lysine .....	81
2.4.1 <i>An attempt with Cbz group</i> .....	82
2.4.2 <i>An attempt with Alloc group</i> .....	84
2.4.3 <i>DDE as the best candidate</i> .....	87
2.5 Summary .....	91

2.6 Experimental .....	92
2.6.1 General considerations .....	92
2.6.2 General conditions for SPPS .....	92
2.6.3 General conditions for LCMS .....	93
2.6.4 General conditions for Combiflash .....	93
2.6.5 General conditions for resin cleavage .....	94
2.6.6 Protocols for on-resin deprotection of protecting groups .....	94
2.6.7 Protocols for peptide modification .....	96
2.6.8 Chemical synthesis .....	96
2.6.9 List of the core peptides synthesized by SPPS present in this Chapter .....	99
2.6.10 LCMS graphs of the naïve peptides synthesized by SPPS present in this Chapter .....	99
2.6.11 LCMS graphs of the modified peptides .....	102
2.6.12 Structures of the peptides mentioned in this Chapter .....	104
2.7 Bibliography .....	107
<b>Chapter 3 – Approaches to study peptide-protein binding .....</b>	<b>110</b>
3.1 Quartz Crystal Microbalance with Dissipation monitoring (QCM-D) as a first approach to monitor binding .....	110
3.1.1. Underlying mechanism .....	110
3.1.2. The choice of sensor .....	111
3.1.3. Au surface functionalisation .....	113
3.1.4. SiO <sub>2</sub> surface, different optimisations .....	120
3.1.5 QCM-D, a conclusion .....	124
3.2 Fluorescence Polarization (FP) as a more sensitive technique for peptide-protein binding assays .....	125
3.2.1 The technique .....	125
3.2.2 Different fluorophores .....	126
3.2.3 Synthesis of fluorescently labelled peptides .....	129
3.2.4 Fluorescence polarisation for measuring protein binding .....	135
3.2.5 Fluorescence Polarisation, a conclusion .....	138

3.3 Diazirine photochemistry for peptide binding and cross-linking assays.....	139
3.3.1 <i>The use of diazirines in bioconjugation</i> .....	139
3.3.2 <i>Synthesis of diazirine probes</i> .....	140
3.3.3 <i>Modification of peptides with diazirine probes – testing reactivity</i> .....	143
3.3.4 <i>Labelling of proteins using diazirine labelled peptides.</i> .....	145
3.3.5 <i>Diazirine, a conclusion</i> .....	147
3.4 Summary .....	149
3.5 Experimental .....	150
3.5.1 <i>Chemical synthesis</i> .....	150
3.5.2 <i>List of the unfunctionalized peptides synthesized by SPPS in this Chapter</i> .....	163
3.5.3 <i>LCMS of peptides used in this Chapter</i> .....	164
3.5.4 <i>Structures of peptides used in this Chapter</i> .....	169
3.5.5 <i>QCM-D protocol</i> .....	172
3.5.5 <i>FP supplementary information</i> .....	173
3.5.6 <i>UV curing supplementary information</i> .....	174
3.6 Bibliography .....	175
<b>Chapter 4 – Synthesis of ligand-directed peptide probes</b> .....	<b>178</b>
4.1 Synthesis of LDAI probes for <i>OpLD</i> .....	179
4.1.1. <i>Synthesis of LDAI probes</i> .....	179
4.1.2. <i>LDAI chemistry on peptides</i> .....	182
4.1.3. <i>Conclusion on LDAI chemistry</i> .....	186
4.2 Synthesis of LD methylcyano-NASA probes for <i>OpLD</i> .....	187
4.2.1 <i>Background and synthesis plan</i> .....	187
4.2.2 <i>A general small molecule approach for building methylcyano-NASA</i> .....	188
4.2.3 <i>A solid-phase approach to build methylcyano-NASA</i> .....	190
4.2.4 <i>Conclusion methylcyano-LDNASA chemistry</i> .....	194
4.3 Synthesis of pyridinium oxime – catalyst for <i>CatLD</i> .....	195
4.3.1 <i>Synthesis of PyOx peptides using literature procedure</i> .....	195
4.3.2 <i>Synthesis of PyOx peptides using click-chemistry</i> .....	198

4.3.3 Conclusion on LDNASA chemistry.....	200
4.4 Synthesis of <i>p</i> -nitrobenzyl-NASA probes – substrates for CatLD.....	201
4.4.1 Synthesis of <i>p</i> -nitrobenzyl-NASA using literature procedure. ....	201
4.4.2 Synthesis of <i>p</i> -nitrobenzyl-NASA using new intermediate .....	203
4.5 Summary .....	205
4.6 Experimental .....	206
4.6.1 Chemical synthesis .....	206
4.6.2 LCMS of peptides used in this Chapter .....	224
4.6.3 Structures of peptides used in this Chapter not presented in discussion .....	229
4.7 Bibliography .....	230
<b>Chapter 5 – Synthesis of new PyOx-2-PCA probes for ligand-directed chemistry .....</b>	<b>232</b>
5.1 An introduction to 2-PCA probe .....	232
5.1.1. Chemoselective approach, the use of natural amino acids .....	232
5.1.2. Building blocks for 2-PCA head .....	234
5.2 Synthesis of alkyne-2PCA probes.....	236
5.2.1 A direct coupling on 2-PCA to an alkyne probe.....	236
5.2.2 The need to introduce protecting group to the 2-PCA .....	236
5.3 Synthesis of azide-PyOx probes .....	239
5.3.1 Synthesis of azido-diPyOx .....	239
5.3.2 Synthesis of azido-monoPyOx .....	240
5.3.3 Synthesis of PEG-linker-PyOx .....	241
5.4 Click-chemistry and deprotection.....	243
5.4.1 Click-chemistry, background and choice of reagents.....	243
5.4.2 Route A: click-chemistry before deprotection .....	244
5.4.3 Route B: deprotection before click-chemistry .....	247
5.5 Summary .....	248
5.6 Experimental .....	249
5.6.1 Chemical synthesis .....	249
5.6.2 Combiflash purification .....	261
5.7 Bibliography .....	262

<b>Chapter 6 – Using ligand-directed chemistry for protein modification.....</b>	<b>264</b>
6.1 Introduction .....	264
6.2 The use of PyOx peptide probes on insulin .....	265
6.2.1 Probes and set-up of the experiment .....	265
6.2.2 Experiment and LCMS analyses.....	268
6.2.3 SDS-PAGE and western blot analyses .....	272
6.2.4 MALDI analyses .....	275
6.2.5 Futher observations on peptide .....	278
6.2.6 Conclusion .....	283
6.3 The use of PyOx-2-PCA probes on proteins .....	284
6.4 Summary .....	288
6.5 Experimental .....	289
6.5.1 General protocols for protein modification.....	289
6.5.2 General protocols for analysis of protein modification .....	289
6.5.3 Supplementary LCMS protein spectra .....	291
6.6 Bibliography .....	293
<b>General Conclusions and Suggestions fur Further Work.....</b>	<b>294</b>
General Conclusions.....	294
Suggestions for Futher Work .....	295
<b>Annex.....</b>	<b>297</b>

## Acknowledgments

My PhD has been an incredible journey, which would not have been possible without the support of several people and Institutions, to whom I am very grateful. Herein, I express my acknowledgment to all those that have supported me in achieving this important goal.

I start by expressing my warmest gratitude and thanks to my supervisor, Dr. Christopher Spicer. Thank you for entrusting me in being one of your first PhD students; it has been incredible to see the Spicer group grow bigger over these three years with you. Thank you for always being so available and for the inspiration you gave me and the freedom you granted me to follow my own ideas even if they weren't all successful. I am so grateful for your scientific support, without which I would never have been able to grow into the chemist I am today, but also thank you so much for your personal support. Additionally, I have to thank the rest of Spicer group Nicholas Rose, Lydia Barber, Reuben Breetveld, Dr. James Donald, Dr. Nico Seling. I am really proud to have been part of this group filled with such good scientists and friends. I also had the opportunity to grow in myself mentoring different Master's and Bachelor's students, Angela, Taylor, Jack and Hannah, thank you. To everyone in E-block who made my time special, with a special note to Fraser, Grant and Alfiya, thank you!

I also have to thank all the people who helped me and performed experiments with me during the few last months of my PhD which led to critical results ensuring the practicality of the chemistry discussed in this thesis. Chapter 6 is dedicated to them. Firstly, thank you to Lydia Barber for performing protein labelling with the 2-PCA probes. Thank you also to the four postdoctoral researchers from B-block Chemistry and the Department of Physics, Dr. Saeed Akkad, Dr. Tessa Keenan, Dr. Amanda Noble and Dr. Lisa Miller, for their time and explanation on the recording of SDS-PAGE, western blot and QCM-D data. I am so glad that despite the time challenge of the last two months of my PhD, I managed to wrap up so many positive results and its thanks to your availability!

My thanks also goes out to the other faculty members of the University of York Chemistry and Biology Department, including Dr Ed Bergstrom, Dr. Karl Heaton and Dr. Adam Dowle, all of whom have provided guidance and support with MS and LCMS knowledge.

I extend my gratitude to my examiners, Dr Megan Wright and Dr Martin Fascione, who have kindly agreed to take the time to read through this work, and to my funding body, Rosetrees Trust, for enabling me to do a PhD in the first place as it was such an opportunity is a tremendous privilege.

I could not finish without mentioning my beloved family who not only inspired me but were a constant support of my career choices. Thank you for always being curious about my research which means so

much to me and that hopefully one day will also impact the medical world. Thank you for the constant enthusiasm you shared with me during my moments of success and happiness.

I would also like to thank my new work family at Bicycle Therapeutics. Thank you for giving me the opportunity to start such a fantastic job when I was still in progress of writing this thesis. Though it has been quite a challenge, I would not have done it differently as I am greatly enjoying my new life in Cambridge. While everyone has been so welcoming, I want to give a special thank you to Alessandra, Georgina and Nisha who listened to me on so many occasions when speaking about the writing process and kindly accepted to read through it giving me valuable critics! Moreover, thank you to Nicolas who kindly helped me with PDB for the protein structures represented in this thesis.

Finally, to the person who literally made me become a whole complete happy being. I am so lucky to have met you and I will never have the words to thank you enough for being by my side for these past three years. You have endured my complaining, have been happy about my positive results, have made me invent a whole Nemo world for the understanding of my work, have cooked so many delicious meals when I was working hard on the writing... Por todas essas coisas e muito mais, obrigada Mariela!

## Declaration

I, Laetitia Raynal, declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

A handwritten signature in black ink that reads "Laetitia". The signature is written in a cursive style with a large, decorative initial 'L'.

Laetitia Raynal

July the 3<sup>rd</sup> 2023

## Glossary

AcOH	Acetic acid
Alloc	Allyloxycarbonyl
APTES	3-Aminopropyltriethoxysilane
aq.	Aqueous
Boc	<i>tert</i> -Butyloxycarbonyl
br.	Broad
<i>t</i> Bu	<i>tert</i> -Butyl
calcd	Calculated
cat.	Catalytic
<i>Cat</i> LD	Catalyst based ligand-directed chemistry
CDCl <sub>3</sub>	Chloroform-deuterated
CH <sub>2</sub> Cl <sub>2</sub>	Dichloromethane
CuAAC	Copper-catalyzed alkyne-azide cycloaddition
CuSO <sub>4</sub>	Copper sulfate
d	Doublet
DDE	1-(4,4-Dimethyl-2,6-dioxocyclohex-1-ylidene)ethyl)
DIC	<i>N,N'</i> -Diisopropylcarbodiimide
DIPEA	<i>N,N</i> -Diisopropylethylamine
DMAP	4-Dimethylaminopyridine
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DMSO- <i>d</i> <sub>6</sub>	Deuterated dimethyl sulfoxide
DMTMM	4-(4,6-Dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium chloride
DTT	Dithiothreitol
EDC.HCl	<i>N</i> -ethyl- <i>N'</i> -(3-dimethylaminopropyl)carbodiimide hydrochloride
eq.	Equivalents
ESI	Electrospray ionisation
EtOAc	Ethyl acetate
EtOH	Ethanol
Et <sub>3</sub> N	Triethylamine
Et <sub>2</sub> O	Diethyl ether
FGF	Fibroblast growth factor
Fmoc	Fluorenylmethoxycarbonyl
FP	Fluorescence polarization

H <sub>2</sub>	Hydrogen
HATU	2-(7-Aza-1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HRMS	High resolution mass spectrometry
HCl	Hydrochloric acid
H <sub>2</sub> O	Distilled water
H <sub>2</sub> SO <sub>4</sub>	Sulfuric acid
HOBt	Hydroxybenzotriazole
Hz	Hertz
<i>J</i>	Coupling constant
LCMS	Liquid chromatograph mass spectrometry
LDAI	Ligand-directed acyl imidazole
LDNASA	Ligand-directed <i>N</i> -acyl- <i>N</i> -sulphonamide
<i>m</i>	multiplet
<i>m/z</i>	Mass to charge ratio
<i>M</i>	Molarity mol/L
MALDI	Matrix Assisted Laser Desorption/Ionisation
MeCN	Acetonitrile
MeOH	Methanol
MeOD	Methanol-deuterated
MgSO <sub>4</sub>	Magnesium sulfate
MPTES	Mercaptopropyltriethoxysilane
Na <sub>2</sub> CO <sub>3</sub>	Sodium carbonate
NaH	Sodium hydride
NaHCO <sub>3</sub>	Sodium bicarbonate
NaOH	Sodium hydroxyde
NASA	<i>N</i> -acyl- <i>N</i> -sulphonamide
Na <sub>2</sub> SO <sub>4</sub>	Sodium sulfate
NBS	<i>N</i> -Bromosuccinimide
NHS	<i>N</i> -Hydroxy succinimide
NH <sub>2</sub> NH <sub>2</sub> .H <sub>2</sub> O	Hydrazine hydrate
NH <sub>2</sub> OH.HCl	Hydroxylamine hydrochloride
nm	Nanometer
NMP	<i>N</i> -Methyl-2-pyrrolidone
NMR	Nuclear Magnetic resonance
<i>Op</i> LD	One-pot ligand-directed chemistry

PAL	Photo-Affinity Labeling
2-PCAs	2-pyridinecarboxyaldehydes
Pd/C	Palladium on carbon
PEG	Polyethylene glycol
PET	Petroleum ether 40-60 °C
Ph	Phenyl
POI	Protein of interest
PBS	Phosphate buffer solution
PPh <sub>3</sub>	Triphenylphosphine
PyOx	Pyridinium oxime
q	Quartet
QCM-D	Quartz Crystal Microbalance with Dissipation monitoring
qtt	Quantitative
rpm	Revolutions per minute
rT	Room temperature
s	singlet
SDS-PAGE	Sodium dodecyl sulfate – polyacrylamide gel electrophoresis
SiO <sub>2</sub>	Silicon dioxide
S <sub>N</sub> 2	Substitution Nucleophilic Bimolecular
SPPS	Solid Phase Peptide Synthesis
SSR	Sum of squares
t	Triplet
TBTU	O-(Benzotriazol-1-yl)-N,N,N',N'-tetramethyluronium tetrafluoroborate
TCEP	Tris(2-carboxyethyl)phosphine
TFA	Trifluoroacetic acid
Th	Theoretical mass
THF	Tetrahydrofuran
TIPS	Triisopropyl silane
Trt	Triphenylmethyl
<i>p</i> TsCl	<i>p</i> -toluenesulfonyl chloride

## Chapter 1 – Introduction

Proteins are key biomolecules that have been long studied by chemists and biologists to understand their multiple functions within live cells.<sup>1-4</sup> An existing challenge in protein science is to modify them chemically to create engineered proteins. There are many reasons to modify proteins, the most common being to attach reporters on their surfaces to track them within cells, or to modify their activity either by activating or inhibiting their biological activity.<sup>1-4</sup> Many methods for chemical protein modification have been developed over decades of research and many comprehensive reviews have been written on the subject.<sup>5-13</sup> Selective modifications of proteins can be grouped in three main classes of reaction, represented in Figure 1:

- Reactions based uniquely on reactive canonical amino acids as reactive handles within the protein sequence. This strategy is the oldest. It is based on a broad range of reagents developed to target specific amino acids with characteristic reactive handles that lead to chemoselective modification, such as NHS esters for lysine or acrylates for cysteine.<sup>1</sup>
- Reactions using genetic codon-expansion techniques and other techniques to introduce non-natural functionality into proteins that can then undergo functionalisation or derivatisation.

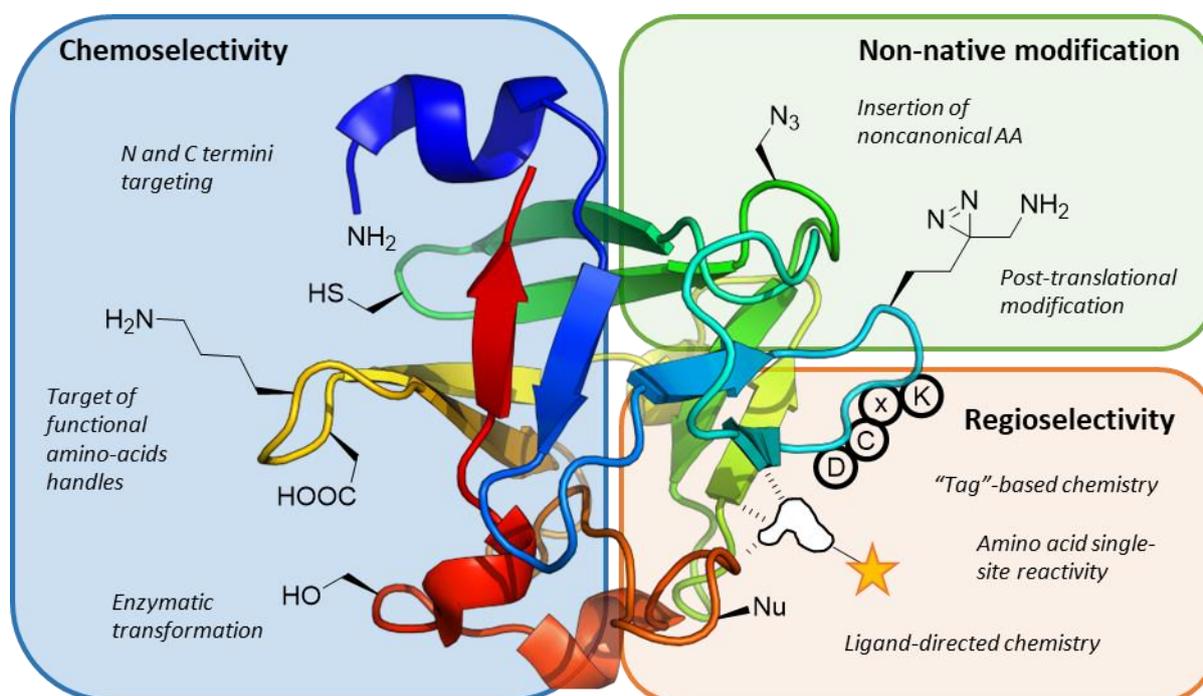


Figure 1: Outline of different strategies for the chemical modification of proteins. Chemoselective reactions, where the reactivity of one type of natural amino acid is targeted. Modification using non-canonical amino acid residues, allowing unnatural functional groups to be introduced and targeted. Regioselective chemical modification of proteins, where ligand binding induces modification. Protein image from PDB 4OEE of human fibroblast growth factor 2. The positions of the amino acids are illustrative.

- Reactions based on regioselective modification linked to specific environment. The main class of reaction fitting within this category exploit ligands to bind at a specific location.

The chemoselective modification of natural amino acids is the oldest of these methods and a large number of chemical reactions have been optimised over the decades to target different residues, such as through the condensation of the  $\epsilon$ -amino group of lysine with activated esters, or the Michael addition of  $\alpha,\beta$ -unsaturated carbonyls with cysteine residues. Whilst in this thesis no chemoselective modifications on proteins were performed, modification of peptides was an important part of the PhD and thus this chemoselective approach was used to synthesise and modify peptide probes. The targeting of unnatural amino acids has been widely developed in recent years, but since it has not been a focus of this PhD it falls outside the scope of this introduction. Instead, the reader is directed to a number of excellent reviews on the subject.<sup>7,8,10</sup>

A major focus of this PhD has instead been on the development of regioselective methods of protein modification. This last approach can be applied to native, non-engineered proteins which is advantageous because they can then be targeted more easily in their natural environment and don't require extra post-synthetic modification of the sequence to enable modification. The goal of this PhD was to develop a generalisable chemical strategy to modify native proteins by exploiting binding peptides for ligand-directed chemistry. This Introduction therefore gives a review of existing chemical modification strategies that exploit both chemoselective and regioselective approaches. Following this, a brief introduction to the protein targets used within this PhD will be given, and with this literature in mind, the fourth part will present an overview of the goals of this PhD.

## 1.1 Modification of proteins, a general background

Before going into the details of the reactions developed over the years to modify specific amino acids, three important general points have to be raised. The first aspect is on the reactions themselves: the reaction conditions used for protein modification must be as close to biological, ambient conditions as possible, i.e. pH 6–8,  $\leq 37$  °C, aqueous solvent. These conditions are often critical to ensure protein stability. If not observed, proteins can undergo degradation by unfolding and thus lose their biological activity. Moreover, unlike standard reaction conditions used for the synthesis of small molecules in organic synthesis (mM to M concentrations of substrates), proteins are generally handled as substrates at low concentrations ( $\mu\text{M}$  or less) which can be a kinetic obstacle as lower concentrations translate to slower reactions. The second aspect is the need to consider the effect of modification on the activity of the protein. It is important that modification results does not impair the inherent function and structure of the protein, which as described later can require significant optimisation.<sup>11</sup> Finally, the ideal modification would be able to take place within complex biological environments, such as in live cells. Due to the numerous biological molecules and reactive functional groups present in such a complex environment, both chemoselectivity and regioselectivity are incredibly challenging to achieve.

In this section, we will review methods to target natural amino acids chemoselectively, with an emphasis on the two most widely used targets, lysine and cysteine. Work on other amino acids such as tyrosine, and protein *N*-termini will also be raised. We will then go on to discuss methods for regioselective protein modification, presenting two sub-approaches with the use of “tag”-selective approach and the ligand-directed technique.

### 1.1.1 Chemoselective approaches, the use of natural amino acids

#### 1.1.1.a Lysine

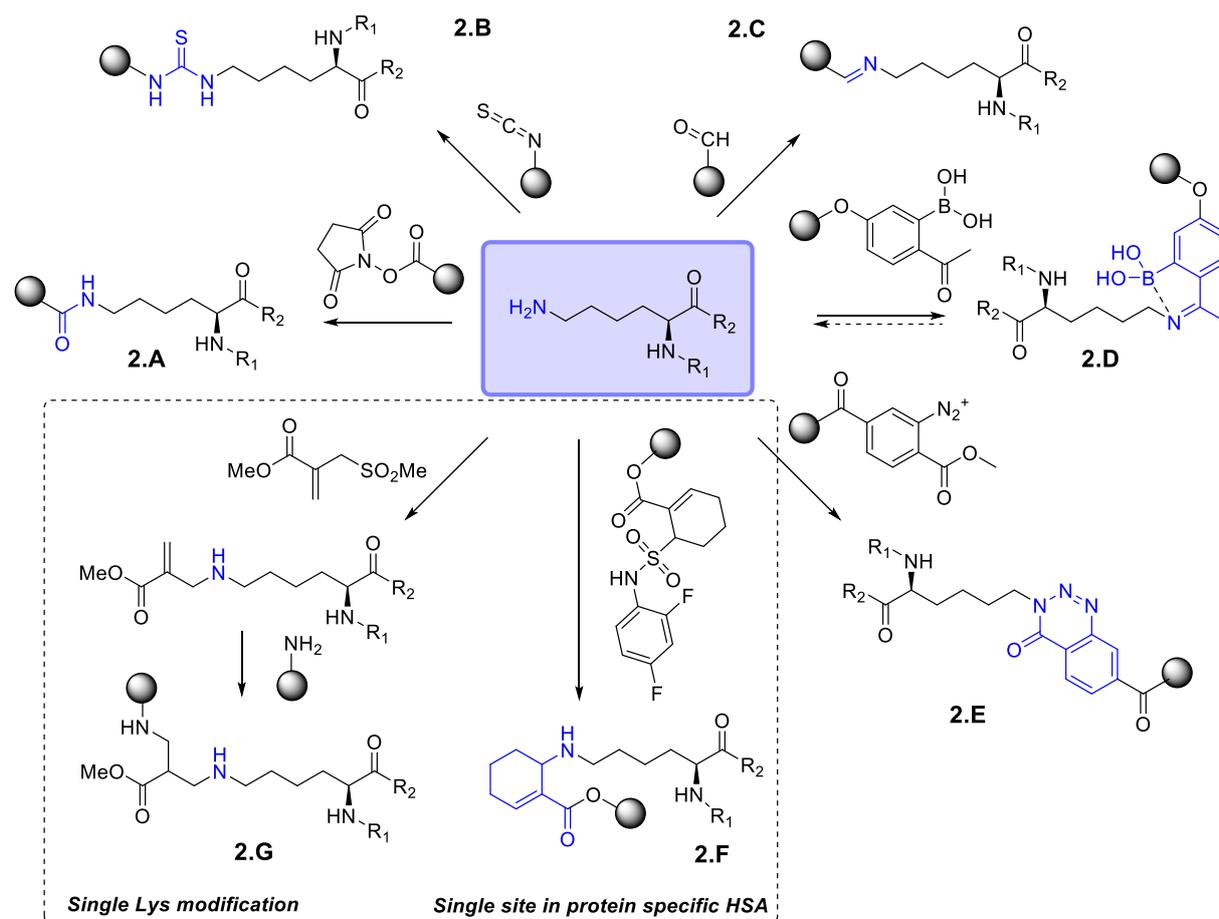
Lysine has been the most widely targeted canonical amino acid due to the intrinsic chemical nature of its primary amine. Due to the nucleophilicity of a primary amine, this amino acid has been a target for modification with different electrophilic reagents such as activated esters,<sup>14</sup> isothiocyanates<sup>15</sup> or aldehydes<sup>16</sup> to cite only the most common reagents (respectively routes **2.A**, **2.B**, **2.C** in Figure 2). These reactions are respectively condensations, urea formation and reductive alkylations. A specific application of isothiocyanates is the classical application in *N*-terminal protein sequencing: Edman degradation.

In order to achieve the best yield, reactions targeting lysine residues are generally performed at basic pH (pH > 8) because of the high  $pK_b$  of its side-chain amine group ( $pK_b \sim 10$ ).<sup>11</sup> Despite being a good handle for modification, lysine's high natural abundance on protein surfaces makes site-selective

modification generally difficult, leading to a mixture of heterogeneously modified products. To emphasise this point, Cravatt and co-workers recently studied lysine reactivity within the human proteome and reported more than 9000 lysines using a pentynoic acid sulfotetrafluoro-phenyl ester probe for site-specific chemical protein profiling, method termed 'isoTOP-ABPP' (isotopic tandem orthogonal proteolysis-activity-based protein profiling).<sup>17</sup> However only a few hundred of these lysines had the required reactivity to be directly targeted by amine-reactive electrophilic fragments at neutral pH in their native environment. Many examples could be given on non-specific lysine modification using the most common reagents, activated esters, isothiocyanates or aldehydes, but those approaches will not be discussed further in this Introduction.<sup>14-16</sup>

Two newer approaches which required less specific pH conditions for lysine modification were reported both based on the stabilisation of the new conjugate through kinetically favourable cyclisation. First, Gois and co-workers developed an approach using a reversible labelling reaction between 2-acetylphenylboronic acid and the primary amine of the lysine side chain to form a reversible iminoboronate bond (route **2.D** in Figure 2).<sup>18</sup> In this system, B-N interactions serve to stabilise an imine that would otherwise be unstable. The authors used different proteins including lysozyme, cytochrome c, ribonuclease A, and myoglobin and each of those proteins showed multiple additions, with up to five residues modified. Moreover, the authors also noticed reactivity at the *N*-terminus of some of those proteins. Another drawback of the approach was the reversible aspect of the iminoboronate bond upon the addition of biologically-relevant small molecules, such as fructose or dopamine, most likely resulting from the disruption of the B-N bond. In an alternative approach, Carreira and co-workers used newly designed diazonium terephthalate reagents that can form stable triazin-4(3H)-one rings for lysine modification (route **2.E**). They targeted myoglobin on which up to six additions was observed.<sup>19</sup> Such high levels of modification could impact drastically the capacity of the protein to fold properly and thus to maintain its activity after modification. Thus, despite some new work on the matter, the specificity of the reaction toward lysine sidechains remains a major challenge.

Due to this challenge, over the last decade, a number of groups have developed new chemistries to target a single lysine on a target protein. The differences in reactivity between different lysine groups is driven primarily by their local environment. Indeed highly solvent-exposed lysine residues are typically the primary target as they can react faster.<sup>1</sup> Examples of rapidly reacting reagents that can be used stoichiometrically for specific lysine modification were designed by Barbas and co-workers using cyclohexene sulfonamides to target single lysines with the protein of target being human serum albumin, HSA (route **2.F**).<sup>20</sup> While the single modification achieved in this work is impressive, the approach is highly protein dependant and not generalisable as some proteins might have lysine residues only in important structural position and that are not easily accessible.



**Figure 2: Most common lysine targeting modifications.** Route 2.A: use of activated NHS-esters for the formation of amide bond. Route 2.B: use of isothiocyanate for the formation of a thiourea bond. Route 2.C: use of aldehyde residues for the formation of a reversible imine. Route 2.D: use of acetylboronates for the formation of a stabilised iminoboronate.<sup>18</sup> Route 2.E: use of diazonium terephthalate reagents.<sup>19</sup> Route 2.F: use of cyclohexene sulfonamide derivatives to modify selective lysine within HSA.<sup>20</sup> Route 2.G: use of sulfonyl acrylate reagents for selective lysine modification.<sup>21</sup>

A more recent approach developed by Bernardes and co-workers in 2018 used computer-assisted design of sulfonyl acrylate reagents for the modification of a single lysine residue in native protein sequences (route **2.G** in Figure 2).<sup>21</sup> The unique aspect of this work was based on the additional regioselective considerations that took into account the local microenvironment surrounding each lysine which could lead to subtle reactivity differences between them. It was first predicted, computationally, which lysine had the lowest pK<sub>a</sub> within the structure and that at a slightly basic pH, modification of this specific lysine residue would be favoured. It is also important to mention that their work was also chemoselective as it was observed that the reagent reacted preferentially at lysine, even in those cases when other nucleophilic residues such as cysteine were present. Finally, and making this work even more important towards the future of protein modification, they modified successfully five different proteins at a single lysine residue, including the clinically used therapeutic antibody Trastuzumab without prior sequence engineering, with for each target the native secondary

structure and functionality retained after the modification.<sup>21</sup> The only limitation would be to ensure that the structure is known sufficiently well to enable effective computational modelling.

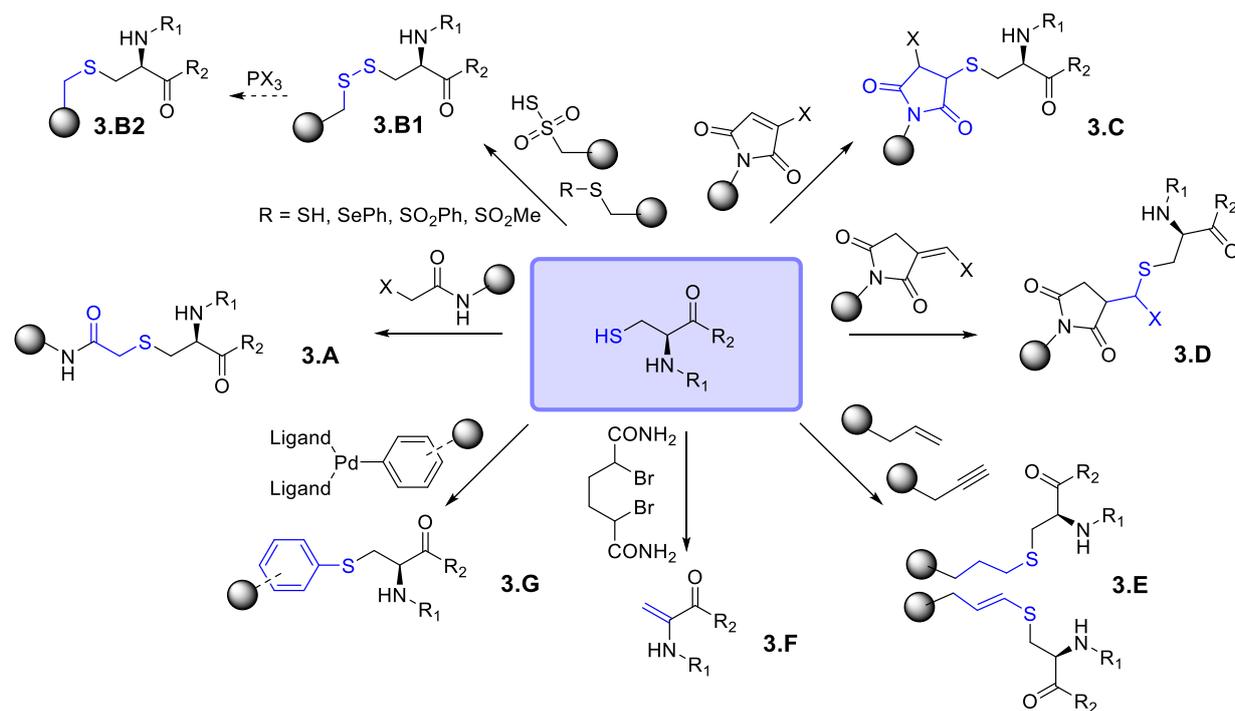
### 1.1.1.b Cysteine

The second most commonly targeted canonical amino acid for protein modification is cysteine. This amino acid is the most nucleophilic present in native proteins at neutral pH, ( $pK_a \sim 8.3$ ) making it a strong candidate for modification.<sup>7,22</sup> It is worth noting that at higher pH, lysine which was introduced in the previous paragraph then has stronger nucleophilicity. The thiol group of cysteine is also unique which allows a new range of reactions such as arylation, 1,4-addition, or radical reactions not achievable on a lysine residue. Finally compared to lysine, its low abundance, less than 2%, within protein sequences makes it an even stronger candidate for selective modification. It will be discussed in a further paragraph the importance of the presence of cysteine in disulfide bond and the impact between being part of such a bond or a free cysteine.

The first type of probe used for almost a century for cysteine modification is  $\alpha$ -halocarbonyl reagents (iodoacetyl, bromoacetyl, or chloroacetyl), route **3.A** in Figure 3, leading to alkylation reaction.<sup>23</sup> Iodoacetamide is used routinely for capping cysteines before digestion for protein sequencing. A drawback of these type of reagents is that they react in a similar manner with methionine side chains making this chemistry not very chemo-selective.<sup>1</sup>

Another common method to modify cysteine is based on the formation of disulfide bonds (route **3.B1** in Figure 3). A lot of different reagents have been used for this type of reaction such as 2-thiopyridines,<sup>24</sup> methanesulfonates,<sup>25</sup> phenylthiosulfonates<sup>26</sup> and selenyl sulfides.<sup>27</sup> However, the disulfide linkages formed are typically unstable under reductive conditions and quite slow to form. To limit this, it was shown that disulfide bonds could be converted to an irreducible thioether bond via desulfurization using phosphorous triamides  $PX_3$  such as tris(dimethylamino)phosphine (route **3.B2**).<sup>28</sup>

Alternatively, cysteine can react in a very fast manner with Michael acceptors. While  $\alpha$ ,  $\beta$ -unsaturated carbonyls encompass a variety of probes the most common which have been used for cysteine modification are maleimide probes (route **3.C** in Figure 3).<sup>29</sup> While classical maleimides are sensitive to hydrolysis, and the thioether bonds can undergo thiol exchange via retro-Michael addition, more recently bromo- and dibromomaleimides have been shown to overcome this reversibility.<sup>30,31</sup> However it is important to mention that at higher pH, Michael acceptors also react with amines (particularly *N*-termini) lowering chemoselectivity. Another interesting alternative approach to modify cysteine residues was reported by Kalia and co-workers using exocyclic olefins instead of classical Michael acceptors resulting in stable thio-Michael products, route **3.D**, which did not undergo thiol exchange.<sup>32</sup>

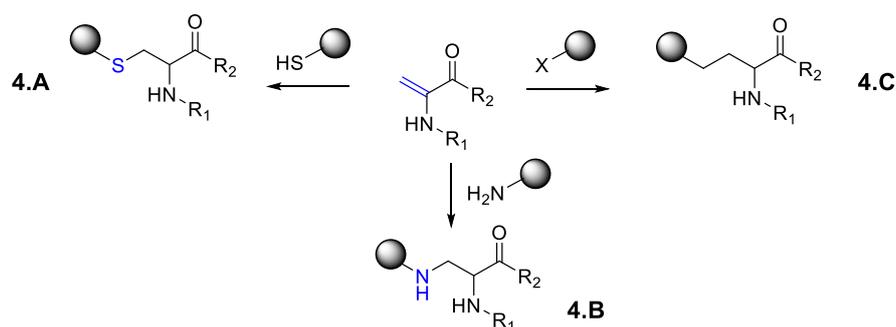


**Figure 3: Most common cysteine targeting modifications.** Route 3.A: use of  $\alpha$ -halocarbonyl reagents for alkylation.<sup>23</sup> Route 3.B1: use of methanesulfonate, phenylthiosulfonate and selenyl sulfides substrates for disulfide formation. Route 3.B2: use of phosphorous triamide  $PX_3$  for desulfurization of disulfide bond. Route 3.C: use of maleimide, bromo- and dibromomaleimides<sup>29–31</sup>. Route 3.D: use of alkyne and alkene for photoinduced reactions.<sup>33–36</sup> Route 3.E: use of  $\alpha, \alpha'$ -di-bromo-adipyl(bis)amide for the formation of electrophilic dehydroalanine<sup>37,38</sup> Route 3.F: use of aryl-palladium(II)-complexes for arylation reactions.<sup>39,40</sup>

With  $\alpha$ ,  $\beta$ -unsaturated carbonyls being good substrates for cysteine modification, it was further shown that the reaction of cysteine with electron-deficient alkynes and alkenes is also favorable. A broad range of substrates have been developed such as alkynones, alkynoic amides, alkynoic esters, or 3-arylpropionitriles.<sup>33,34</sup> This last example from 2014 showed particularly promising results with faster reaction kinetics than maleimides or  $\alpha$ -halocarbonyl reagents and reaching greater conversions. An alternative strategy for thiol-ene conjugation is via a radical pathway, often initiated by light, so called photoinduced thiol-ene and thiol-yne reactions (route **3.E** in Figure 3).<sup>35,36</sup> These routes have been used widely for protein modification, and the reader is directed to a number of excellent reviews on the subject and direct applications.<sup>41–44</sup> The main elements which are worth mentioning in this introduction are: the high specificity of these routes for targeting cysteine only; the full compatibility with water and oxygen; and the fast kinetics observed in most cases.

A more recent two-step approach for cysteine modification was developed by the Davis group, forming in a first step the unnatural electrophilic residue dehydroalanine (Dha) via a bis-alkyl elimination procedure with  $\alpha, \alpha'$ -di-bromo-adipyl(bis)amide (route **3.F** in Figure 3). This group can then react in three main ways with either nucleophiles such as thiols (route **4.A**),<sup>37,38</sup> or amines via aza-Michael ligations (route **4.B**)<sup>45</sup> or finally using radical reactions with reactive halogenated species

(route **4.C**, Figure 4).<sup>46,47</sup> These products of these reactions are stable and can be used to mimic native post-translational modifications within the human body such as glycosylation,<sup>37</sup> phosphorylation,<sup>48</sup> acetylation,<sup>49</sup> and methylation.<sup>49,50</sup> Recently, radical reactions were used to form stable C(sp<sup>3</sup>)-C(sp<sup>3</sup>) bonds (route **4.C**), mimicking natural post-synthetic modifications of protein within the human body, an improvement compared to the unnatural thioether or *sec/tert*-amine linkage formed in route **4.A** and **4.B** respectively. The main drawback of Dha chemistry is the racemization and loss of stereochemistry that occurs following formation of Dha leading to diastereomeric mixtures of products.



**Figure 4: Reactivity of electrophilic Dha group.** Route 4.A: Michael addition with thiol nucleophiles.<sup>37,38</sup> Route 4.B: aza-Michael ligation with amine nucleophiles.<sup>45</sup> Route 4.C: C-C bond formation with alkyl halides.<sup>46,47</sup>

A more contemporary approach by Pentelute, Buchwald, and co-workers uses transition metals to promote arylation of cysteine residues under mild conditions (route **3.G** in Figure 3).<sup>39</sup> The authors reported the use of aryl-palladium(II)-complexes that react specifically with cysteine. The reaction kinetics were shown to be comparable to the rates of maleimide–thiol couplings ( $k_2 = 10^3\text{--}10^4 \text{ M}^{-1} \text{ s}^{-1}$ ), leading to reactions proceeding within 30 min at micromolar concentrations with almost 100% yield. Moreover, the final *S*-arylated products obtained were more stable than the corresponding maleimide–thiol conjugate products under *in vitro* conditions. Even more importantly, the authors succeeded in attaching the anticancer drug vandetanib to cysteine residues of the Trastuzumab antibody while maintaining its biological activity. The same group continued pushing this chemistry and showed that it could also be used to generate peptide-antibody conjugates, protein-antibody conjugates and protein homodimers.<sup>51</sup> In an extension of this work, this chemistry was adapted by Messaoudi and co-workers to guide palladium-mediated arylation onto metal-binding proteins, showing the advantage of those new type of reactions.<sup>40</sup>

As mentioned in previous paragraphs, it is often essential that any cysteine modification does not disrupt the natural function of the protein. However, due to their low abundance in native proteins, free cysteine residues are often critical for function. In contrast, selective targeting of active-site cysteines has shown more recently to be vital for the development of covalent inhibitors. Moreover, another approach directly related to cysteine modification, despite not targeting free cysteine

residues, is the use of disulfide re-bridging strategies. The re-bridging strategy consists of two steps, first the reduction of the natural disulfide bond, generally in the presence of tris(2-carboxyethyl)phosphine (TCEP), and then using the reactive moiety such as dibromomaleimides<sup>31,52</sup> (route **5.A**) or diallyl-ureas (route **5.B**)<sup>53</sup> as shown in Figure 5. The advantage of using already existing disulfide bridge compared to cysteine is that the new bridge generally has less effect on the protein activity.

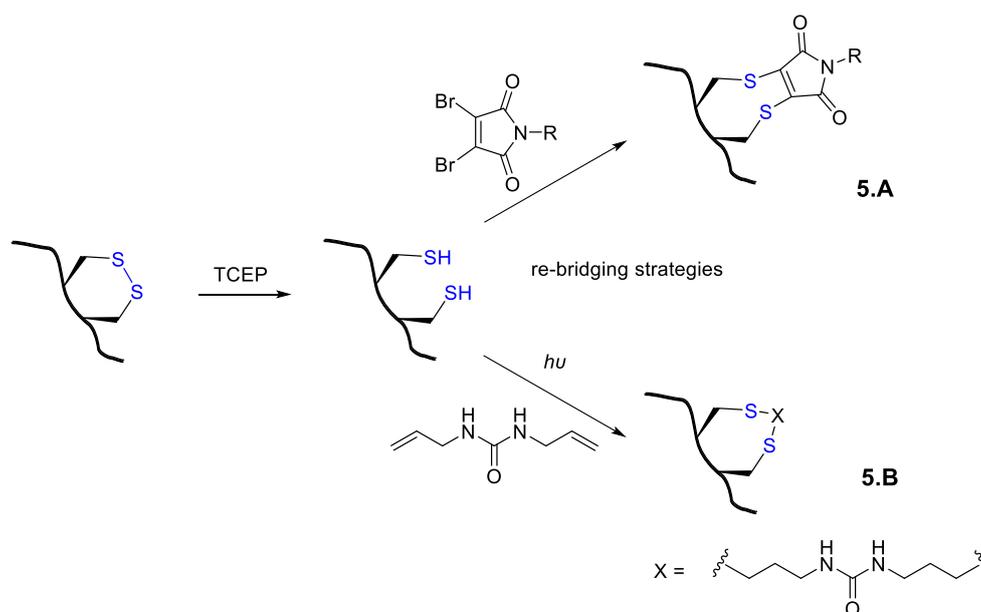


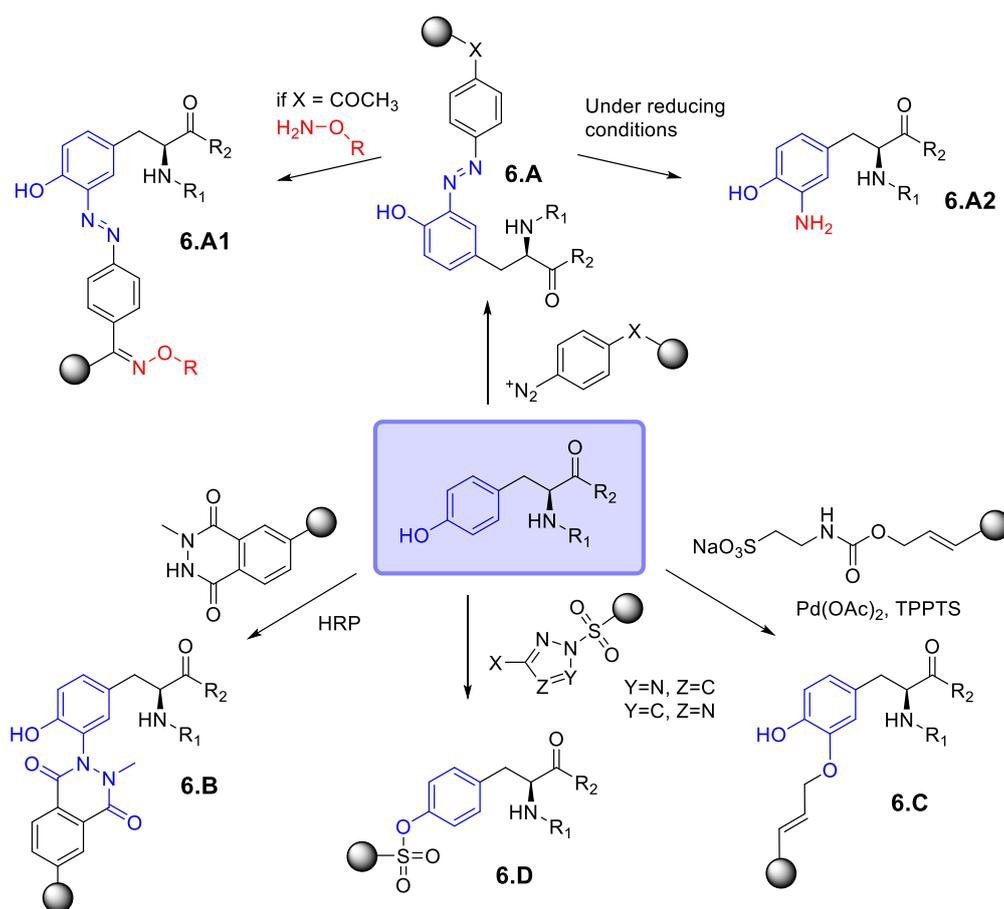
Figure 5: Two common re-bridging strategies of disulfide bonds after reducing conditions in presence of tris(2-carboxyethyl)phosphine (TCEP) using dibromomaleimides (route **5.A**)<sup>31,52</sup> or diallyl-urea under UV light excitation (route **5.B**).<sup>53</sup>

### 1.1.1.c Tyrosine

Another attractive amino acid for site-specific modification of proteins is tyrosine. Its aromatic nature, presence of a phenol group, and its low frequency on the surface of proteins make it an interesting target. Due to the  $pK_a$  of phenol ( $\sim 10$ ), different reactions can be induced depending on the protonation state of the side chain. When at a slightly acidic to neutral pH, reactions on the aromatic carbons *ortho* to the hydroxyl group, such as via ene-type reactions, diazonium couplings, or Mannich-type condensation are favoured. When the pH is more basic and closer to the  $pK_a$  of phenol, alkylation or acylation reactions of the hydroxyl group can be induced.<sup>7</sup> Since protein modification is preferable at neutral pH, only methods that are viable at that pH will be discussed in more detail.

A first approach for tyrosine modification was demonstrated using azo-coupling reactions with diazonium compounds to form diazobenzene species, at the *ortho* position of the phenol group (route **6.A**). Francis and co-workers pioneered this work showing the importance of electron-withdrawing groups (EWGs) at the *p*-position of the diazonium group for higher conversion.<sup>54</sup> While they showed

*p*-bromo and *p*-nitro functionalised diazoniums gave the highest conversions (> 90%), the use of *p*-acetophenone groups allowed further modification due to the ketone introduced on the protein surface. Indeed using alkoxyamines they showed the use of this handle for attachment of a series of reagents, such as biotin, via oxime formation (route **6.A1**).<sup>54</sup> Moreover it was shown that the diazonium group could be reduced to generate the corresponding aniline, (route **6.A2**), which could then be either acylated or be used as a handle for further functionalisation.<sup>55</sup>



**Figure 6: Most common tyrosine targeting modifications.** From top to bottom: diazonium reagents used to form diazobenzene species (route **6.A**) which can undergo further modification using hydroxylamines<sup>54</sup> or by being reduced to an amine,<sup>55</sup> horseradish peroxidase catalyzed addition of *N*-methyl luminol derivatives (route **6.B**), sulfur-triazole exchange chemistry (route **6.D**),<sup>56,57</sup> palladium catalyzed alkylation using  $\pi$ -allylpalladium complexes (route **6.C**).<sup>58</sup>

Two main other types of modification have been shown to target tyrosine residues. Sato and coworkers recently developed a unique reaction catalyzed by horseradish peroxidase for the modification of surface exposed tyrosine residues (route **6.B**). Under these conditions they could modify proteins with *N*-methyl luminol derivatives using a minimum amount of  $\text{H}_2\text{O}_2$  as an oxidant.<sup>59</sup> Alternatively, Francis and coworkers reported the modification of tyrosine using palladium catalyzed alkylation, using  $\pi$ -allylpalladium complexes (route **6.C**). Using this chemistry they reported successful alkylation of a Tyr residue on the surface of chymotrypsinogen A.<sup>58</sup>

Sulfur-triazole exchange (SuTEx) chemistry has showed to target tyrosine on its hydroxyl group.<sup>56,57</sup> The protein SuFEx rate is dependant on several factors such as protein binding affinity as well as chemical reactivity of the functional group but also the solution pH which we already mentioned is critical for tyrosine modification. This way it was shown that when nearby cationic residue (e.g. arginine or lysine) were present leading to a lower  $pK_a$  of the tyrosine phenol, SuTEx reactions are catalyzed and showed selective modification of tyrosine residues.<sup>60</sup> However, at slightly higher pH other residues such as lysine could be modified using this chemistry thus limiting its range.<sup>61</sup>

#### 1.1.1.d *N*-terminus

Different modifications of the *N*-termini of proteins have recently been investigated. Indeed, after lysine, the *N*-terminus is the only other primary amine of a protein. The main distinctive characteristic between these two primary amine groups is the lower  $pK_a$  of the *N*-terminus ( $\sim 8$ ) in comparison to lysine residue ( $\sim 10$ ).<sup>4</sup> Moreover, the *N*-terminus is the only alpha-amido amine of a protein. Another important characteristic of targeting the *N*-terminus is that this site is rarely involved in the protein active site or important to folding. Different reviews highlight the research based on *N*-terminal modification and compare the reactivity of different substrates used for its modification.<sup>5,62,63</sup>

An overview of the main reactions for targeting the *N*-terminus will be presented. Two main categories can be described for *N*-terminal modification, the first class relying on specific *N*-terminal amino acid residues, and the second class which can modify the *N*-terminus regardless of the residue.

Within the first category, *N*-terminal cysteines have been widely modified through use of thioesters for acyl transfer (route **8.A**).<sup>64,65</sup> This reaction consists of a two-step process, first the intermolecular formation of an intermediate thioester at the thiol, followed by a rapid *S* to *N*-acyl shift, resulting in the formation of a native peptide bond, Figure 7. This approach not only allows protein modification but also the conjugation of two peptide fragments to build complex proteins, a process known as native chemical ligation.

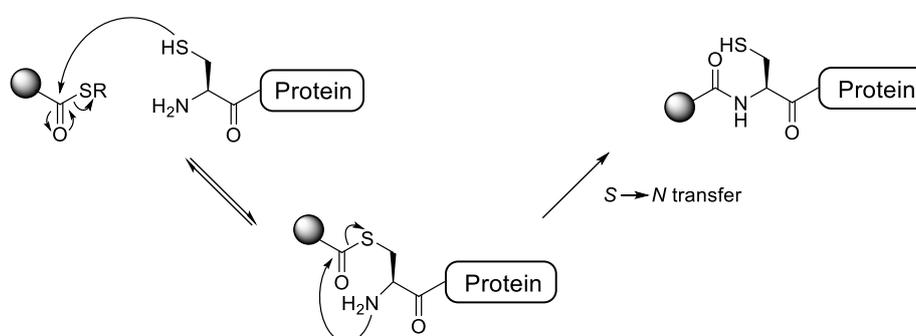
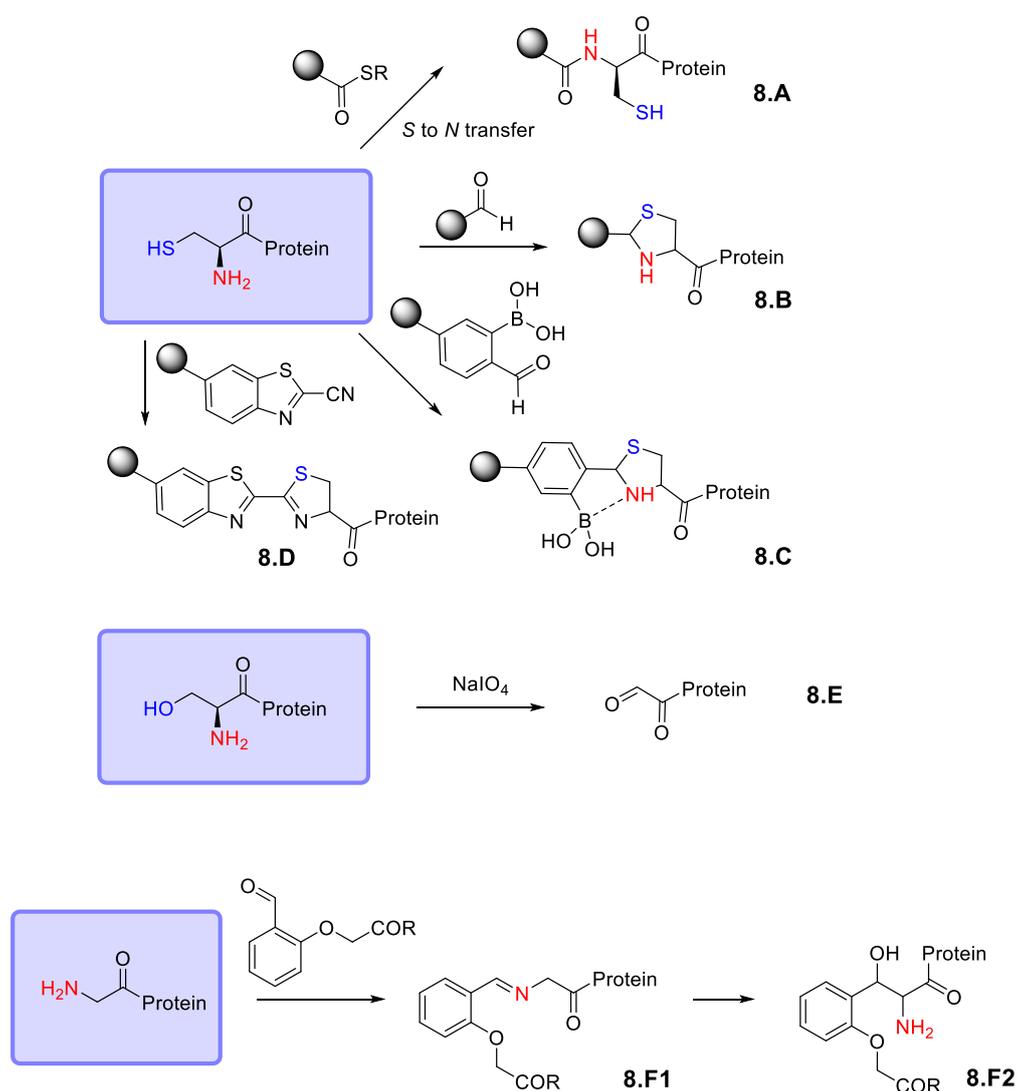


Figure 7: Mechanism of a native chemical ligation at a *N*-terminal cysteine of a protein.

Another common approach to *N*-terminal cysteine modification is to use aldehydes which react first with the cysteine side-chain thiolate followed by intramolecular cyclization. The capture of latent electrophile generates the five-membered stable thiazolidine complex. (route **8.B**).<sup>66</sup> More recently, acetylboronates were used to stabilise thiazollidones thanks to the presence of the boronic group (route **8.C**).<sup>67</sup> However, these reagents also react with lysine residues, limiting specificity. Another example for the modification of cysteine is the use of cyanobenzothiazoles (route **8.D**).<sup>68</sup> This approach was shown to proceed rapidly forming a symmetrical five-membered ring system, inspired by the last step of luciferin formation.

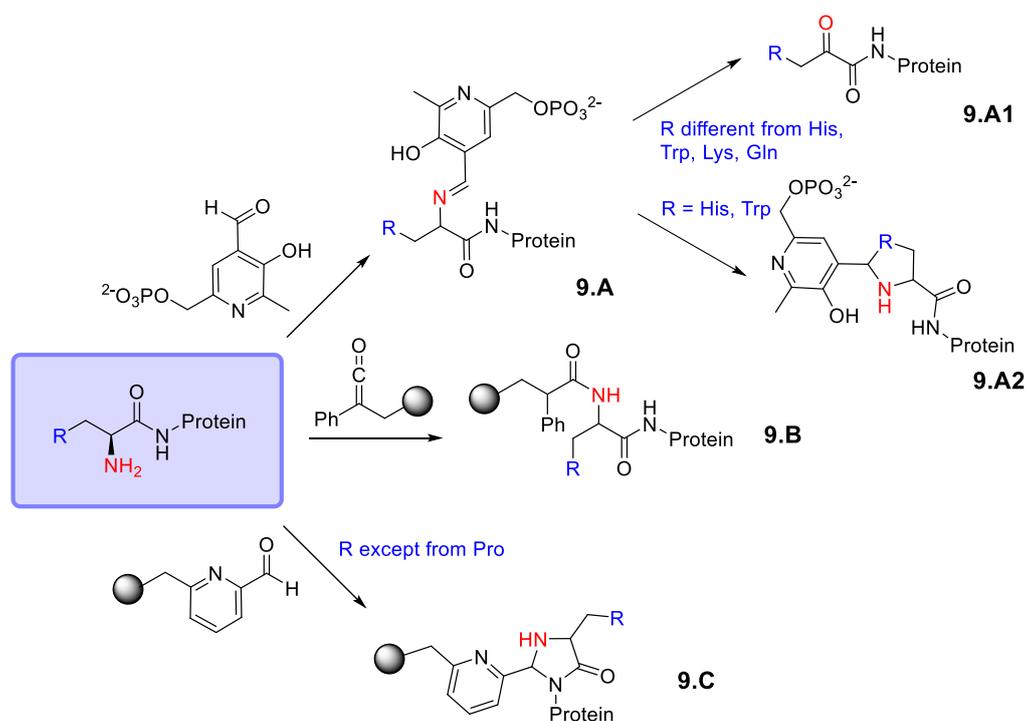


**Figure 8: Most common *N*-terminal amino acid targeted modifications.** Route 8.A: modification of Cys *N*-terminus using *S* to *N* transfer reaction.<sup>64,65</sup> Routes 8.B and 8.C: use of aldehyde 2-formylboronic acid reagent for boronate-stabilised thiazolidine formation.<sup>66,67</sup> Route 8.D: use of cyanobenzothiazole reagents.<sup>68</sup> Route 8.E: modification of Ser *N*-terminus via oxidative formation of aldehyde functionality.<sup>69,70</sup> Route 8.F: modification of Gly *N*-terminus using 2-formylphenoxy acid probes.<sup>71</sup>

Another targeted residue at the *N*-terminus is serine which can be oxidized to generate an aldehyde which can then react with hydrazides and aminoxy groups to form hydrazones and oximes

respectively (route **8.E**).<sup>69,70</sup> Recently, Rawale and co-workers reported the first use of single-site labeling of *N*-terminal glycine residues using 2-formylphenoxy acid probes. After the formation of an intermediate imine (route **8.F1**), rearrangements leads to the formation of an amino alcohol which can undergo further modification (route **8.F2**).<sup>71</sup>

The second category of reagents for *N*-terminal modification are the substrates which are not dependent on the first residue. A first example is pyridoxal-5'-phosphate which can undergo a transamination reaction leading to the generation of a reactive ketone on the surface of the protein (intermediate **9.A** to ketone **9.A1**).<sup>72,73</sup> The ketone can undergo subsequent modification via similar routes to the aldehydes generated via serine oxidation. However, it was shown that some *N*-terminal residues such Lys and Gln give side reactions which limit the generality of this chemistry. Interestingly though, it was also shown that when using specific aromatic residues such as Trp and His, side-reactions to form Pictet–Spengler adducts took place in higher yields than ketone formation, believed to result from the stabilised ring (**9.A2**).



**Figure 9: Most common *N*-terminal specific modifications.** Route 9.A: use of pyridoxal phosphate, leading to the formation of aldehyde moiety by transamination 9.A.1, or to the promoted Pictet–Spingler reaction in the case of *N*-terminal Trp or His 9.A.2.<sup>72,73</sup> Route 9.B: use of ketenes for protein *N*-terminus modification.<sup>74</sup> Route 9.C: use of 2-pyridinecarboxaldehyde for protein *N*-terminus modification.<sup>75,76</sup>

Che, and co-workers developed another interesting *N*-terminal modification protocol using ketenes (**9.B**).<sup>74</sup> They showed excellent selectivity (*N*-terminal amino : ε-amino of up to >99:1) under neutral conditions and at moderate temperatures (room temperature or 37 °C) and pH between 6.3–9.2. When using alkyne-functionalized ketenes they showed another reactive handle could be introduced on different proteins including insulin, lysozyme, RNase A, which could then be modified using click

chemistry. Another approach was explored by Francis and co-workers using 2-pyridinecarboxaldehydes (2-PCA, **9.C**).<sup>75,76</sup> This chemistry was shown to be highly versatile, tolerating all amino acids with the exception of prolines at the second position. This specific chemistry, which will be exploited in this PhD thesis, will be introduced in further detail in Chapter 5, where we will present the range of 2-PCA probes synthesised within this PhD. This last chemistry is the most generalised approach discussed in this introduction, which can be widely applied to many proteins without the need of engineering the protein sequence beforehand and is therefore highly promising for chemo-selective protein modification.

### *1.1.2. Regioselective approaches, the use of “tags” or ligands*

The different strategies presented in Section 1.1 showed only a small amount of the whole “toolbox” for protein modification. While efficient, these techniques targeting natural amino acids have also some drawbacks, such as the limited chemical functionality of the canonical twenty amino acids, and as mentioned earlier, the common occurrence of the same amino acid within a particular protein. To overcome these issues, recognition-driven protein labelling has been developed as an alternative method to modify proteins. This approach was inspired by enzymatic reactions which can take place selectively in live cells with some of the fastest kinetics observed for protein labelling.<sup>2</sup> Selectivity is mainly due to molecular recognition which enforces the close proximity of the reaction substrates facilitating a rapid reaction with high selectivity due to the increased local concentration (the proximity effect).<sup>77</sup> In order to mimic these type of reactions, two main recognition-driven protein labelling pathways have been studied by researchers. First, the “tag” approach consists of targeting a special sequence of amino acids for which a specific reactive probe can be formed in order to modify the protein in proximity of that tag. Second, the ligand-directed approach consists of using a ligand binding to the protein of interest, POI. This ligand is modified beforehand to include either a reactive group able to attach to the protein once close to the binding pocket (known as the one-pot ligand-directed approach), or a catalyst able to induce modification close by using specific substrate (known as the catalyst-based ligand-directed approach). While in literature these two ligand approaches have been respectively called ligand-directed chemistry and affinity guided chemistry, this thesis will use the term “One-pot ligand directed chemistry” and “Catalyst based ligand directed chemistry” as we believe these terms show clearer distinctions between the methods, as used in a previous review.<sup>78</sup>

While only a brief description of the tag approach will be given in the next paragraph, ligand-directed approaches will be given more attention as they were key to the goals of this PhD. Progress in this field has mainly been based on work undertaken by Hamachi and co-workers since 2009 and has developed significantly over the last two decades.

The advantage of regioselective techniques over the chemoselective methods described in the previous section is that the chemistry is not anymore solely dependent on the functional group but also on the actual POI. Indeed while using a “tag” or a ligand, a pseudo-intramolecular environment is created which triggers modification based on location rather than reactive group, allowing a wider range of substrates to be modified.<sup>2</sup> This approach also greatly enhances the specificity of modifications as the reaction happens solely within the pseudo-intramolecular environment formed. Moreover, due to the high specificity of binding and labelling, the amount of probe required for labeling can be drastically reduced. To date, chemical modification using recognition-driven approaches is so far the only technique known that allows modification of endogenous proteins without any genetic manipulation in live cells and organisms.<sup>1,2,77</sup> However, compared to chemoselective strategies targeting single amino acids through their reactive handles, for which a wide range of reagents are known, this proximity-mediated strategy requires more careful design of ligands and reagents for each particular POI.

#### 1.1.2.a “Tag”

Two main types of “tag” for protein modification have been described in literature. The first type of “tag” uses protein scaffolds such as self-modifying enzymes or fluorogen binding proteins for labelling, such as SNAP-tag,<sup>79</sup> CLIP-tag,<sup>80</sup> BL-tag.<sup>81</sup> Often these approaches have shown that fast labelling kinetics could be achieved. However, a number of drawbacks exist. First, the need to modify genetically the POI, which can be both time consuming and technically demanding. In addition, the expression of non-native proteins may suffer from poor yield.<sup>77</sup> Finally, the added protein “tag”, which can be big relative to the POI, can alter the protein function and structure. These factors make this technique outside the scope of this thesis. More literature on these protein-based labelling techniques can be found, but are not addressed any further in this Chapter.<sup>82-84</sup>

The second type of “tag” is the use of short peptides for recognition with a specific substrate for protein selective labelling. While a number of peptide tags have been described, such as  $\pi$ -clamp by Pentelute’s group<sup>85</sup>, the FLASH tag by Tsien’s group,<sup>86,87</sup> or the FLAG-tag by Hamachi’s group,<sup>88</sup> these methods have both advantages and drawbacks. The main direct advantage of using smaller size tags (0.6-6 kDa) is that it minimises the influence on the structure and function of the POI targeted relative to protein tags. Moreover, the different peptide tags mentioned above are generally prepared by organic syntheses before subsequent incorporation onto the protein, which helped to vary the functionalities inserting fluorophore, photo-affinity probe, offering an attractive method. However, many require the tag to be installed at the protein *N*-terminus leading to the artificial expression of the corresponding fusion protein which can be limiting. Here we will focus only on modification happening only on native proteins.

Schepartz and coworkers pioneered the field of native peptide tags finding a rhodamine derived bisboronic dye that specifically recognizes the tetraserine motif (SSPGSS), with this motif being present in more than 100 human proteins. More than being already a game-changer in the tag approach for protein labelling, they also showed the strong affinity of their substrate to the SSPGSS motif with a  $K_d$  in the micromolar range.<sup>89</sup> The main drawback of this method is its generality as it is dependent on the peptide sequence being present or not within the POI.

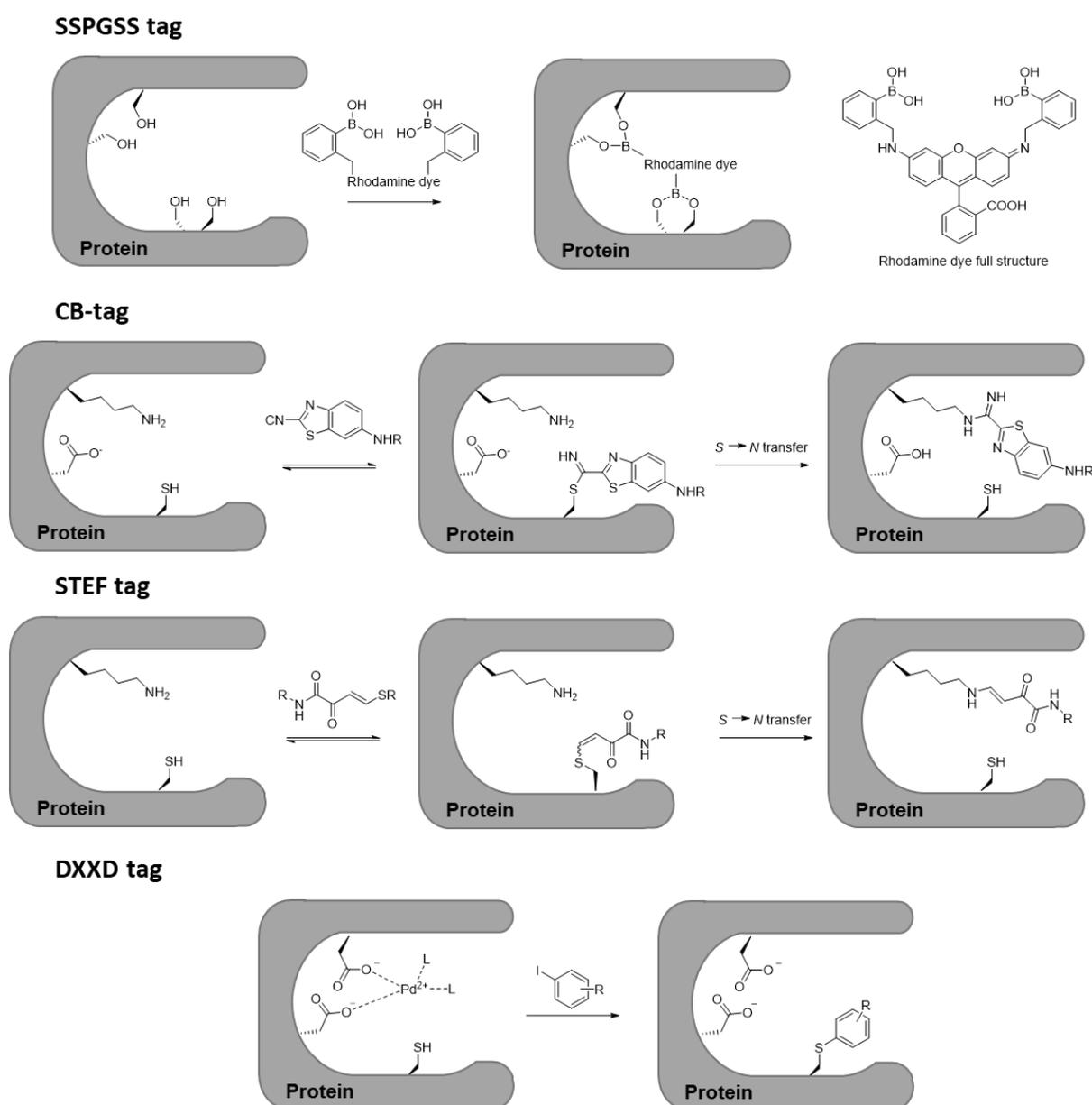


Figure 10: Scheme for different tag regioselective approaches, SSPGSS<sup>89</sup>, CB<sup>90</sup>, STEF<sup>91</sup> and DXXD<sup>92</sup> tags from top to bottom

A more recent approach was reported by Keyser et al, who developed a peptide sequence known as a CB-tag, that reacted specifically with 2-cyanobenzothiazole (CBT) substrates.<sup>90</sup> They based their work on the well-known native chemical ligation reaction (NCL), already mentioned earlier during the N-

terminal cysteine section and presented in Figure 7. The CBT substrate reacts first with the available cysteine to form an activated species which is then disrupted by the lysine in proximity to lead to the equivalent *S* to *N* transfer forming a stronger bond.

The first example of modification using a CB-tag was on a 11 amino acid peptide (GGHPDPCPKGG) using a 6-amino-2-cyanobenzothiazole derivative (CBTNac) derivative to modify the lysine. Different criteria were shown to play a crucial role for optimal modification, including the presence of a proline residue in between the lysine and cysteine to fix the overall conformation, and the use of naturally occurring catalytic triads, with histidine and aspartic acid residues hypothesised to increase the nucleophilicity of the cysteine and lysine residues.<sup>90</sup> Similar to this approach, Hansen and co-worker developed a new acrylate reactive electrophile known as STEF reagent, named after Stachel and Effenberger who provided the first examples of their preparation.<sup>91</sup>

Another example of site-selective tag was reported by Davis and co-workers.<sup>92</sup> They used the motif Asp-Xaa-Asp (DxD), present in mannosyl-glycerate synthase to promote regioselective transition-metal catalysis. Indeed, the close proximity of two aspartic acid makes an ideal metal coordination site to coordinate a palladium catalyst which could then lead to the arylation using aryl iodide reagents of a nearby Cys residue. Those reagents allowed the labelling of the protein with biotin and carbohydrates. Despite the requirement for relatively elevated temperatures, 65 °C, this novel approach highlights the potential of transition-metal chemistry utilizing endogenous metal binding site.

A final example of chemistry bridging the gap between the tag approach and the ligand-directed approach is the method of linchpin directed modification (LDM). This chemistry relies on a two-step process. First, a fast, reversible bond is formed with nucleophilic residues via the formation of an imine; second an irreversible, proximity-driven, intramolecular reaction to a second nearby residue occurs. The selectivity is thus not dictated during the first imine-forming step but during the second. The selectivity is directly linked to the linker shape between the electrophilic tail and head, which can direct which nucleophile on the protein surface can react in a proximity-driven manner. In the first report, the first targeted nucleophiles were lysines residues using salicylaldehyde for imine formation directed to histidine modification using an epoxide tail, Figure 11.<sup>93</sup> The salicylaldehyde moiety could further be modified via oxime ligation. The authors demonstrated that two different His residues could be site-selectively alkylated through a slight change in the spacing between the electrophilic moieties of the reagent. This chemistry was then extended to a lysine-to-lysine strategy using dibromophenyl esters,<sup>94</sup> similar chemistry which will be discussed later on for ligand-directed modification. In this approach the main element relies on the fact that one Lys residue reacts preferentially to form the reversible imine much faster in order to avoid two distinct products. This method was successfully

applied to the selective modification of ribonuclease A (RNase A), and trastuzumab IgG. The main drawback of LDM is the importance of proximity of two specific residues which leads to a need for optimization on a case-by-case basis.

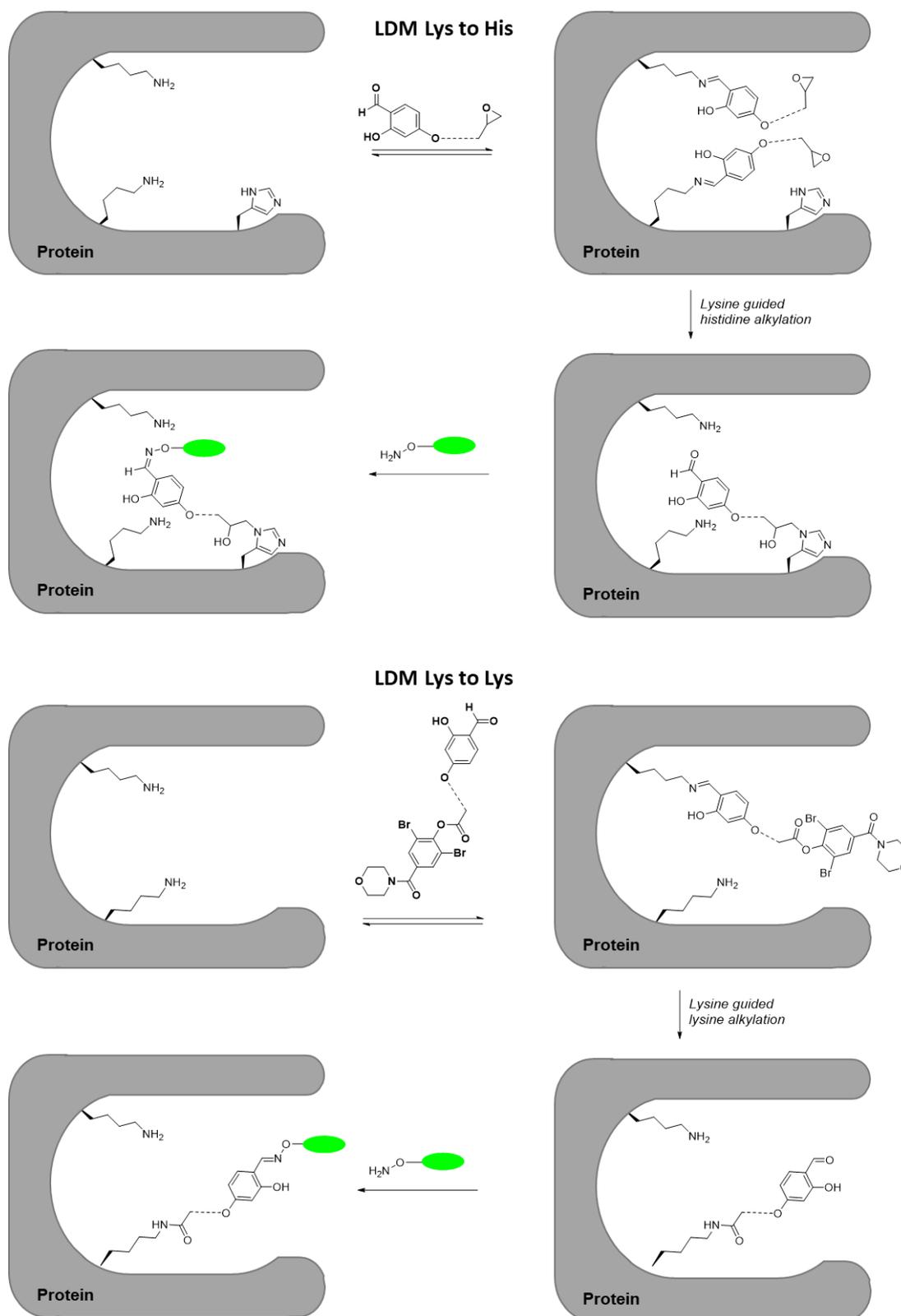


Figure 11: Lynchpin reagents for Lys-guided, single site-specific modification. At the top, linchpin directed modification using salicylaldehyde moiety for lysine-to-histidine modification. At the bottom, linchpin directed modification using dibromophenyl esters for lysine-to-lysine modification

### 1.1.2.b One-pot ligand directed (*OpLD*) chemistry

The first description of one-pot ligand directed chemistry for protein modification was in 2009 by Hamachi, Tsukiji, and co-workers. This approach consists of using a ligand to direct an electrophile to the surface of a protein which can react with a nucleophile present on the surface of the POI. Upon reaction, the ligand is cleaved giving a traceless reaction, a major distinction compared to traditional approaches such as covalent inhibitors.<sup>2,95</sup> This minimizes the size of the added group on the protein surface, leaving the binding site, often also active site, free. This chemistry exploits poor electrophiles which would not react under normal conditions, but reactivity is enhanced by the creation of a pseudo-intramolecular environment when the ligand is bound to the protein, increasing the 'concentration' of the electrophilic moiety at a very specific position.

This chemistry has been described as a three-step process as pictured in Figure 12. First the binding between the POI and its specific ligand (micromolar to sub-micromolar dissociation constants are typically required); then the reaction between a nucleophilic amino acid on the surface of the POI in close proximity to the electrophilic group within the reagent; and finally, the dissociation of the now cleaved ligand from the newly covalently modified protein. The major advantage of this technique is being traceless as the ligand is removed, allowing the active site to be free, ensuring its activity even after labelling.

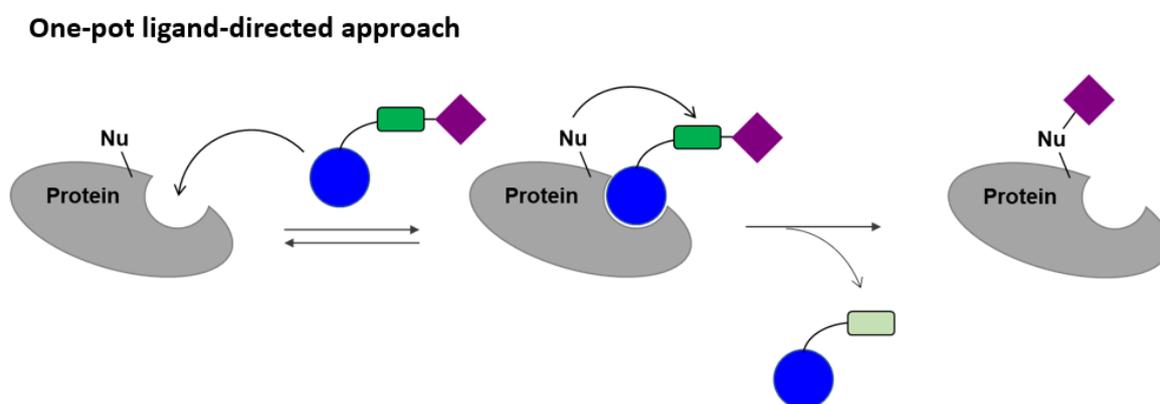


Figure 12: One-pot ligand-directed approach scheme. The ligand (blue circle) bearing the electrophile moiety (green rectangle) able to create a pseudo-environment for the transfer of the substrate (purple) from the ligand toward the protein surface.

Different *OpLD* chemistry methods have been developed changing the reactive electrophile and thus the reactivity towards different nucleophiles on the surface of the protein and the kinetics of the reaction.<sup>2</sup> While other approaches have been reported, the specific groups which have been fully kinetically and regioselectivity studied are presented in Table 1. More information in literature can be found about other reagents that have been used such as phenyl esters,<sup>96</sup> ethyl 3-(2-hydroxy-4-methoxyphenyl)propiolates<sup>97</sup> and 5-sulfonyl tetrazoles.<sup>98</sup>

*Ligand-directed tosylate (LDT) chemistry*<sup>95,99–103</sup>

LDT chemistry was the first *Op*LD to be reported based on the electrophilicity of a phenyl sulfonate (tosylate) group. This can undergo  $S_N2$  reaction with nucleophilic amino acids, such as His, Tyr, Glu, Cys, and Asp, Figure 13. Given that the selective modification of Asp and Glu side chains remains challenging, the LDT method is unique as a chemical modification tool targeting carboxyl groups exposed on a POI.<sup>1</sup> However, LDT chemistry has shown to have slow reaction kinetics which are not ideal for protein modification despite being promoted within the pseudo-intramolecular environment.

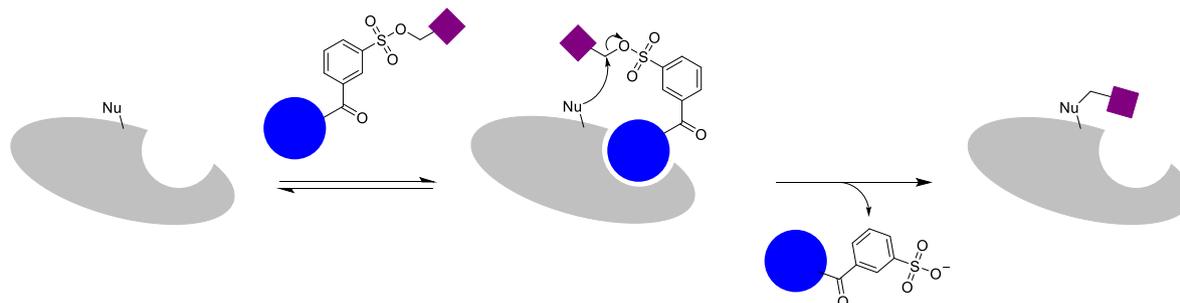


Figure 13: Ligand-directed tosylate chemistry. A ligand (blue) binds to the protein of interest leading to the substrate (purple) being transferred onto the surface of the protein after  $S_N2$  reaction with a nearby nucleophile.

Using proximity-driven approaches, the interaction between the POI and the ligation reagent will play a crucial role in the reaction kinetics. An in-depth study was conducted using peptidyl-prolyl cis-trans isomerase FKBP12 as the POI with various LDT reagents and various spacers to evaluate their reactivity.<sup>99</sup> The main result of this study showed that not only the length but also the rigidity of the spacer between the ligand and the tosylate was a key determinant of overall yield and labelling site. Indeed, using a piperazine spacer the main modification was observed on residue Glu57 (96%) with only 4% modification of His90, Figure 14. Using a linear spacer led to a drastic change with the conversion reaching 43% for His90, only 53% for Glu57 and even a newly modified residue Tyr85 (4%). These characteristics therefore offer the ability to control the labelling site. Another study with carbonic anhydrase II (CAII) shows a big difference with the piperazine spacer, highlighting the fact that every single protein is unique.

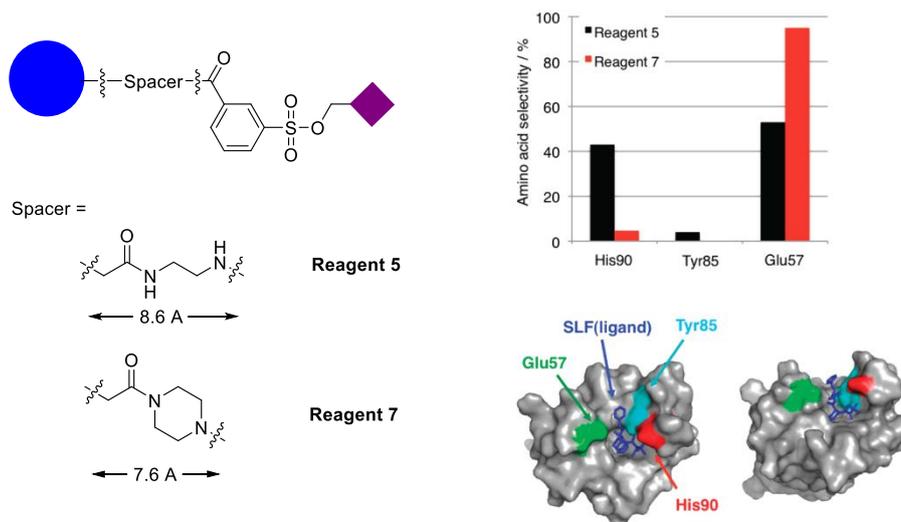


Figure 14: Representation of the FKBP12 ligand in blue modified with LDT probes using a linear spacer, reagent 5, or a piperazine spacer, reagent 7, and its influence on the localisation of the modification between three residues His90, Tyr85 and Glu57. Graph and protein illustration from literature.<sup>99</sup>

A range of studies have shown that LDT can modify proteins with a wide range of synthetic probes such as fluorophores, affinity tags and <sup>19</sup>F-NMR probes.<sup>99–103</sup> Moreover the ability of LDT to be used to label endogenous proteins in various biological systems is a great strength.<sup>99</sup> However, the slow reaction kinetics and low labelling yield (less than 30% for FKBP12 over 30 h labelling time) has hindered its wider use in the regioselective modification of target proteins.

#### Ligand-directed acyl imidazole (LDAI) chemistry<sup>104–109</sup>

LDAI chemistry is based on a different type of reaction, via an acyl transfer, using a moderately reactive alkyloxyacyl imidazole, for protein acylation, Figure 15. LDAI chemistry has shown higher kinetic rates and higher labelling efficiency than LDT chemistry. It has been shown that the modified amino acids varied with lysine and serine also modified by LDAI reagents. It is worth noting that these two groups are more nucleophilic than glutamic acid or aspartic acid and thus could also influence the kinetics of the modification.

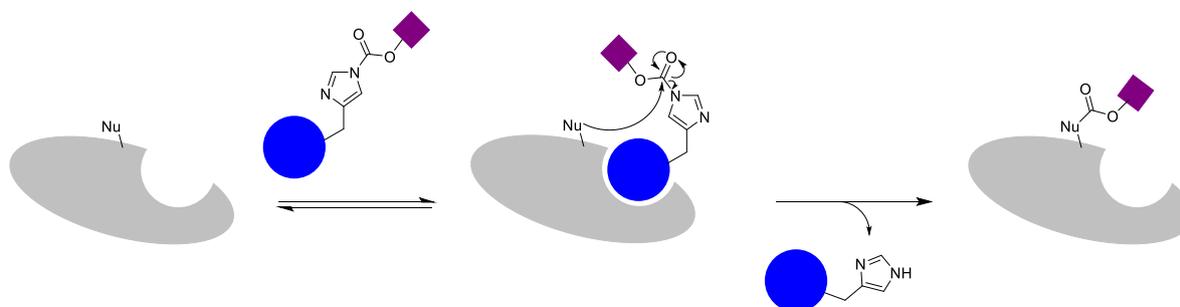
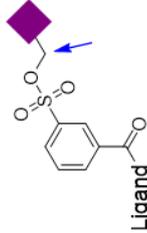
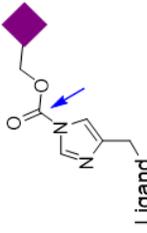
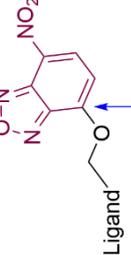
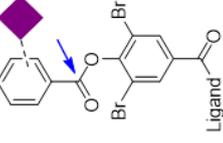
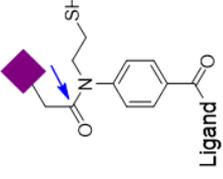
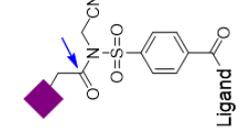


Figure 15: Ligand-directed acyl imidazole chemistry. A ligand (blue) binds to the protein of interest leading to the substrate (purple) being transferred onto the surface of the protein after acyl transfer reaction with a nearby nucleophile.

Table 1: Summary of one-pot ligand-directed chemistries, in blue in the structure an arrow showing the electrophilic group that will undergo reaction with the nucleophile present on the surface of the POI.

Type of ligand-directed chemistry	LDT (2009)	LDAI (2012)	LD-O-NBD (2013)	LDBB (2015)	SEAlide (2016)	LDNASA (2018)
						
	Tosylate group	N-acyl imidazole group	O-nitrobenzoxadiazoole	Dibromobenzyl benzoate	N-sulfanyl-ethylamide group	N-acyl-N-alkyl sulfonamide group
Labelled amino acid	His, Tyr, Glu, Asp, Cys	Lys, Ser, Thr, Tyr	Lys	Lys, His	Lys	Lys
Labelling time (h)	> 10	3-6	1	1-3	48	< 1
Binding affinity observed for modification	$K_d < 43.5 \mu\text{M}$	$K_d < 5.3 \mu\text{M}$	$K_d < 1.4 \text{ nM}$	$K_d < 0.26 \mu\text{M}$	$K_i < 3.4 \mu\text{M}$	N.A.
Labelling conditions	Purified protein, cell surface, mucus tissue and more	Purified protein, cell surface, mouse brain tissue	Purified protein, cell surface, inside cells	Purified protein, cell surface, inside cells	Purified protein, cell surface, inside cells	Purified protein, cell surface, inside cells
Protein targeted	hCAI, hCAII, FKBP12, Hsc70, HER2, NADH-quinone oxidoreductase	hCAII, hCAIX, hCAXII, NMDAR, AMPAR, GABAR, mGluR	Avidin, VDAC	hCAII, hCAXII, eDHFR	hCAI, COX-1	FKBP12, Hsp90s and folate receptors
Applications	Several type of biosensors, FRET analysis, Photocrosslink, Ligand binding site identification	Fluorescent biosensor, Live cell imaging of membrane proteins, Photo controlling of enzyme affinity and more	Fluorescent biosensor, Live cell imaging of membrane proteins.	Live cell imaging of intracellular protein		Covalent inhibition of HSP90
References	95,99–103	104–109	118,119	117	114–116	110–113

While a range of applications were possible thanks to this new chemistry,<sup>105–108</sup> the key breakthrough of LDAI was due to the small size of the reactive core. Indeed, the reagents could penetrate further into tissue and reach solid tumour cells allowing precise imaging of tumours using fluorescent substrate.<sup>109</sup> However, a major drawback was reached when it was shown that the acyl imidazole group could be easily degraded by intracellular enzymes (e.g. esterases, proteases) and was unstable to pH changes leading to the hydrolysis of the electrophile group and its inactivation.

#### Ligand-directed *O*-nitrobenzoxadiazole (LD-*O*-NBD) chemistry<sup>118,119</sup>

A new type of reagent has been developed by Sodeoka and co-workers using an alkoxy-substituted 7-nitrobenz-2-oxa-1,3-diazole (*o*-NBD). This compound shows no fluorescence compared to its *N*-linked-analogue but has the great advantage of showing high electrophilicity. If attacked by nucleophilic lysines on the surface of a POI via an  $S_NAr$  reaction, the new *N*-bonded species will exhibit high fluorescence at specific wavelengths ( $\lambda_{em}$  520–550 nm). Thus, fluorescence can be used as a measurement of modification efficiency. This technique has therefore been described as a turn-on LD chemistry due to the generation of the fluorophore.<sup>118</sup>

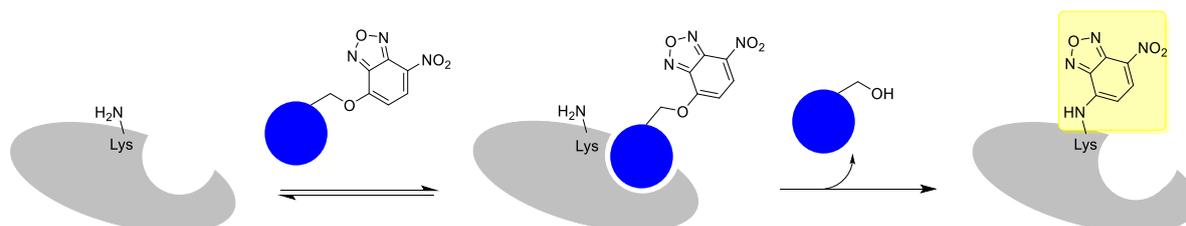


Figure 16: Ligand-directed fluorescence turn-on chemistry. A ligand (blue) binds to the protein of interest leading to the transferred of 7-nitrobenz-2-oxa-1,3-diazole (*o*-NBD) to a nearby lysine forming the fluorophore *n*-NBD.<sup>118</sup>

#### Ligand-directed dibromobenzyl benzoate (LDBB) chemistry<sup>117</sup>

Following the drawbacks of initial reports of one-pot ligand-directed chemistries, dibromophenyl benzoate was adopted as a new probe with improved biological stability due to steric hindrance making the reactive group less susceptible to enzymatic hydrolysis. Different modifications of the phenol moiety were explored and it was found that the dibromo motif has superior bioconjugation efficiency compared to both single *o*-nitro- or dichloro-modifications. Using only 2 equivalents of a dibromo-phenylbenzoate probe, nearly full conversion was achieved over 12 h *in vitro*, showing faster kinetics than LDT and LDAI chemistry. This probe was then used to label selectively intracellular proteins within cultured cells with fast kinetics and modification was shown to target lysine and histidine. However, while being bulky was an important factor for biological stability it also made the probe more hydrophobic which has inherent limitations for use in aqueous media.

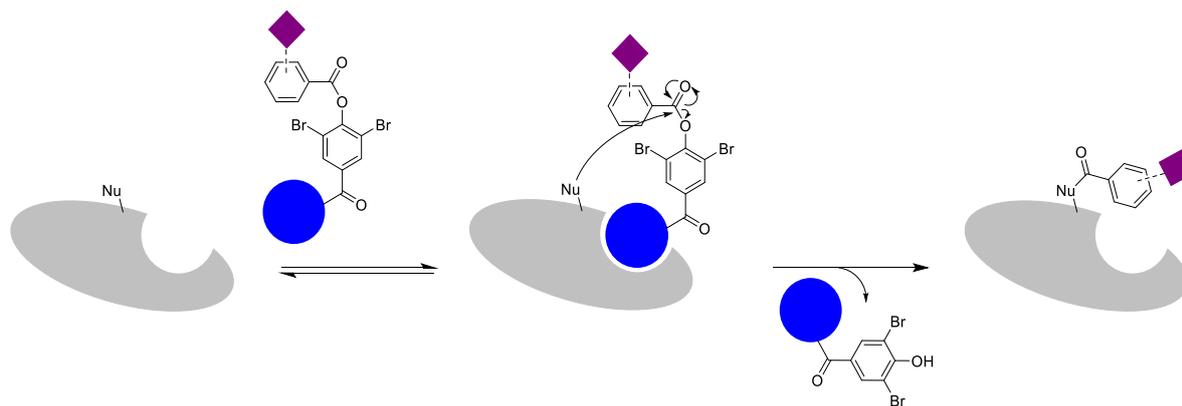


Figure 17: Ligand-directed dibromobenzoyl benzoate chemistry. A ligand (blue) binds to the protein of interest leading to the substrate (purple) being transferred onto the surface of the protein after acyl transfer reaction with a nearby nucleophile.

#### Ligand-directed SEALide chemistry<sup>114–116</sup>

Otaka, Shigenaka and coworkers proposed a unique LD chemistry using *N*-SulfanylEthylAnilide (SEALide) as an electrophile.<sup>114</sup> The SEALide unit is converted to a reactive thioester via an *N*–*S* acyl transfer shift, in a similar way to the CB-tag approach mentioned in the previous section, forming a thioester which is much more electrophilic, Figure 18. Compared to previous *Op*LD chemistries, SEALide chemistry is pH sensitive. The conjugation of the probe to the two model proteins human carbonic anhydrase (hCA) and cyclooxygenase 1 (COX-1) could be performed at both 37 and 4 °C. Moreover the probes showed high protein selectivity in complex mixtures with only little off-target labelling. Although the use of this group has yet to be described in further details, it was used in the original report to modify successfully endogenous carbonic anhydrases in human red blood cells.

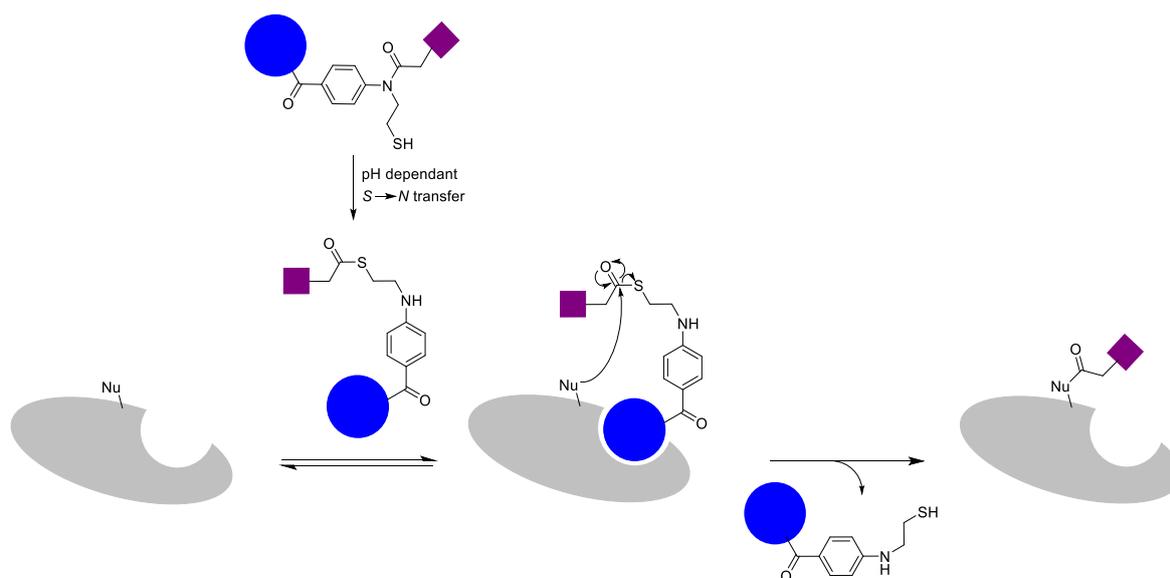


Figure 18: Top: *N*-SulfanylEthylAnilide (SEALide) reaction through *N*–*S* acyl transfer shift to form an activated thio-ester compound.<sup>114</sup> Bottom: Ligand-directed dibromobenzoyl benzoate chemistry. A ligand (blue) binds to the protein of interest leading to the substrate (purple) being transferred onto the surface of the protein after acyl transfer reaction with a nearby nucleophile.

### Ligand-directed *N*-acyl-*N*-alkyl sulfonamide (LDNASA) chemistry<sup>110–113</sup>

The most recent one-pot ligand-directed approach used *N*-acyl-*N*-alkyl sulphonamides and revealed to be much faster than any of the other chemistries described in the previous paragraphs. The second-order rate constant reaction is  $\sim 10^4 \text{ M}^{-1}\text{s}^{-1}$  which can be compared to prevalent self-labeling protein tag methods (e.g., SNAP-tag, Halo-tag) and to the fastest bio-orthogonal cycloadditions (between tetrazines and trans-cyclooctenes).<sup>11</sup> However, to ensure efficient labelling it has been shown that sub-micromolar  $K_d$  for ligand binding is required. On FKBP12, the authors reported two sites of modification both on lysine residues Lys44 and Lys34.<sup>110</sup> The selectivity toward lysine was consistent across all protein targets, which could lead to drawbacks if no lysines were available in proximity of the binding site.

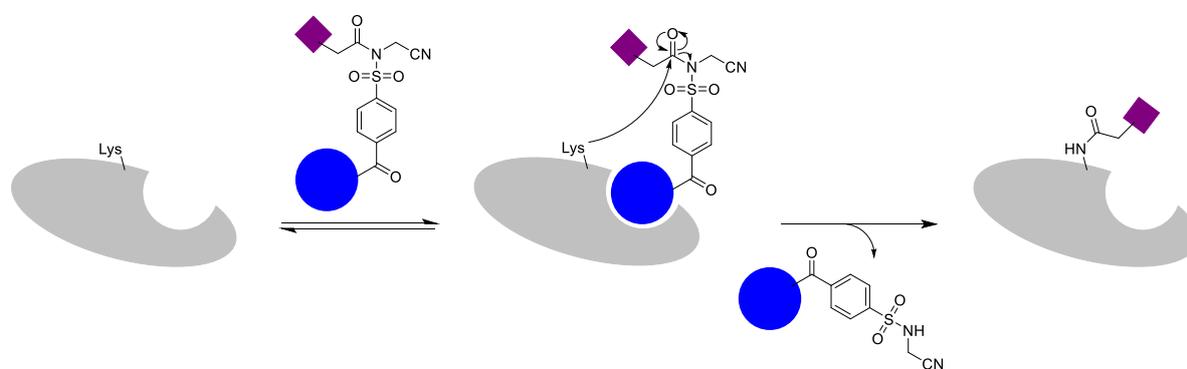


Figure 19: Ligand-directed NASA chemistry. A ligand (blue) binds to the protein of interest leading to the substrate (purple) being transferred onto the surface of the protein after acyl transfer reaction with a nearby lysine.

Moreover, it was also shown that NASA group exhibits good stability in biological aqueous conditions and are not prone to enzymatic degradation. The use of NASA to modify proteins allowed remarkably rapid and selective labeling of FKBP12, with 78% conversion of the entire population of intracellular FKBP12 labelled within one hour. The application of NASA probes to other proteins, Hsp90s and folate receptors, show the potential of this new ligand-directed species. Finally, the use of LDNASA was the broadest *Op*LD chemistry in terms of applications allowing for selective and efficient labelling of intracellular endogenous proteins, as well as membrane proteins, and within living cells.

### Conclusion on one-pot ligand-directed chemistry

As a conclusion on the one-pot ligand-directed chemistry, it has been shown that in terms of reactivity the probes over time have reached faster kinetics, with the latest LDNASA comparable to some of the fastest reactions for protein modification as shown in Figure 20.

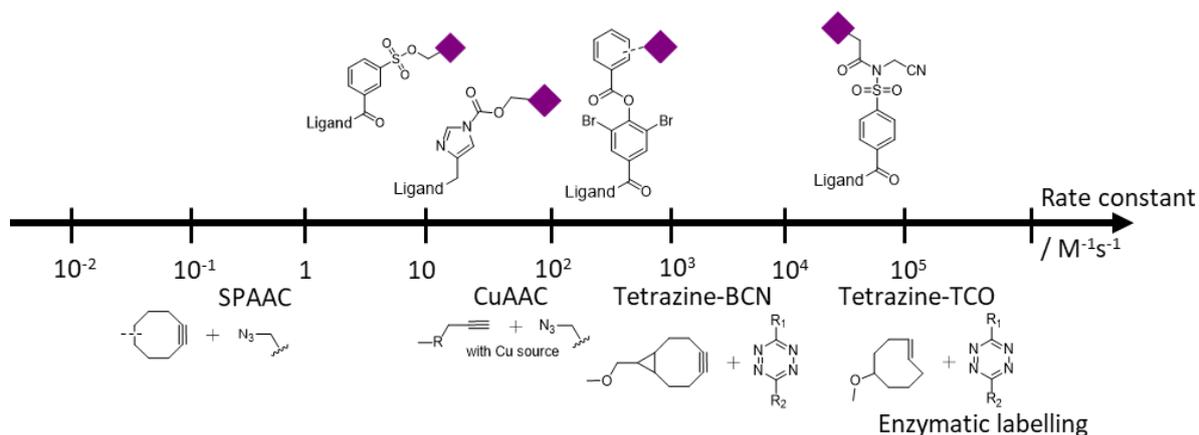


Figure 20: Comparison of the rate constants of representative bioconjugation reactions and OpLD chemistries for protein modification.

While OpLD chemistry has shown great progress, and great possibilities, it also requires a significant amount of optimisation, from the choice of ligand/protein of interest, to the spacer of the reagent and its inherent characteristics (flexibility and length). Indeed, even if a reagent shows great specificity, a new ligand is required for each POI, which can be time consuming and costly. Another main drawback of OpLD probes is that they cannot have reactive nucleophiles within their structures because of intramolecular deactivation of the cleavable electrophile. With respect to these limitations, catalyst-mediated ligand-directed chemistry is highly complementary and will be described in more detail in the following paragraph.

### 1.1.2.c Catalyst-based ligand directed (CatLD) chemistry

As mentioned earlier, the introduction of catalyst-based ligand-chemistry was proposed to address the limitations of the OpLD strategy due to its intrinsic electrophilic group which could undergo intramolecular decomposition. Once again, this catalyst-based ligand-directed chemistry has been described as a three-step process as pictured in Figure 21. First, the binding between the POI and a ligand bearing a catalyst occurs; then the activation of a substrate by the catalyst, within a pseudo-intramolecular environment at the surface of the protein; finally, the reaction between an amino acid on the surface of the POI with the activated substrate group, with the dissociation of the ligand from the now covalently modified protein. Similarly to the OpLD strategy, the major advantage of this technique is that it is traceless ensuring that the protein activity is maintained even after labelling.

### Catalyst based ligand-directed approach

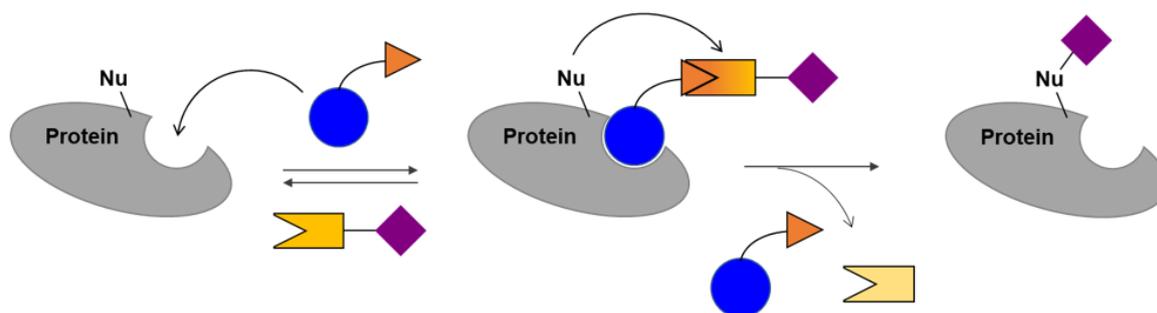


Figure 21: Catalyst based ligand-directed approach scheme. The ligand (blue circle) bearing the catalytic moiety (orange triangle). The catalyst activates its specific complementary moiety (yellow) leading to the creation of a pseudo-environment for the transfer of the substrate (purple) from the ligand toward the protein surface.

Different *CatLD* chemistry methods have been developed, with different catalytic mechanisms and substrates. Two groups of catalysts have been created, based on transition metal or organo-catalysts.<sup>6,11,77</sup> The key reactions are summarised in Table 3.

#### *DMAP CatLD chemistry*<sup>120</sup>

The first example of *CatLD* chemistry was introduced with the catalyst (4-dimethylaminopyridine) (DMAP). In organic chemistry, DMAP is a commonly used catalyst for acyl transfer reactions, which can further activate activated esters to react with a nucleophilic probe. In *CatLD* the first substrate was a thioester acyl donor, able to form a highly reactive acylpyridinium intermediate once reacted with DMAP, Figure 22. Then, an acyl transfer reaction would happen between a nucleophilic amino acid on the POI and the activated acylpyridinium group. The nucleophilic amino acids able to react with the electrophilic intermediate have been shown to be lysine, serine and tyrosine. It was shown that the ratios between POI, catalyst and substrate were important to reach maximum labeling. A 1:1 ratio between catalyst and substrate, with a catalyst to protein ratio of 5:1, resulted in full modification of a lectin after 3h at 25 °C.<sup>121</sup>

Compared to *OpLD* where only one activated species could be present at the ligand binding site, another interesting aspect of *CatLD* was investigated by Hamachi's group: the number of DMAP groups present on the ligand could be increased, having two or three catalyst heads per ligand. The authors showed that both reaction rate and efficiency of protein labeling were increased up to 10-fold as a result.<sup>122</sup> Moreover it is believed that DMAP catalysts act not only as an activator but also as a base to accelerate the acyl transfer reactions.<sup>78</sup>



of nearby residues such as tryptophan, tyrosine. The authors showed that this method could modify a range of antibodies in a relatively simple process, by creating a peptide-binder by standard solid-phase synthesis techniques and one-step metalation.

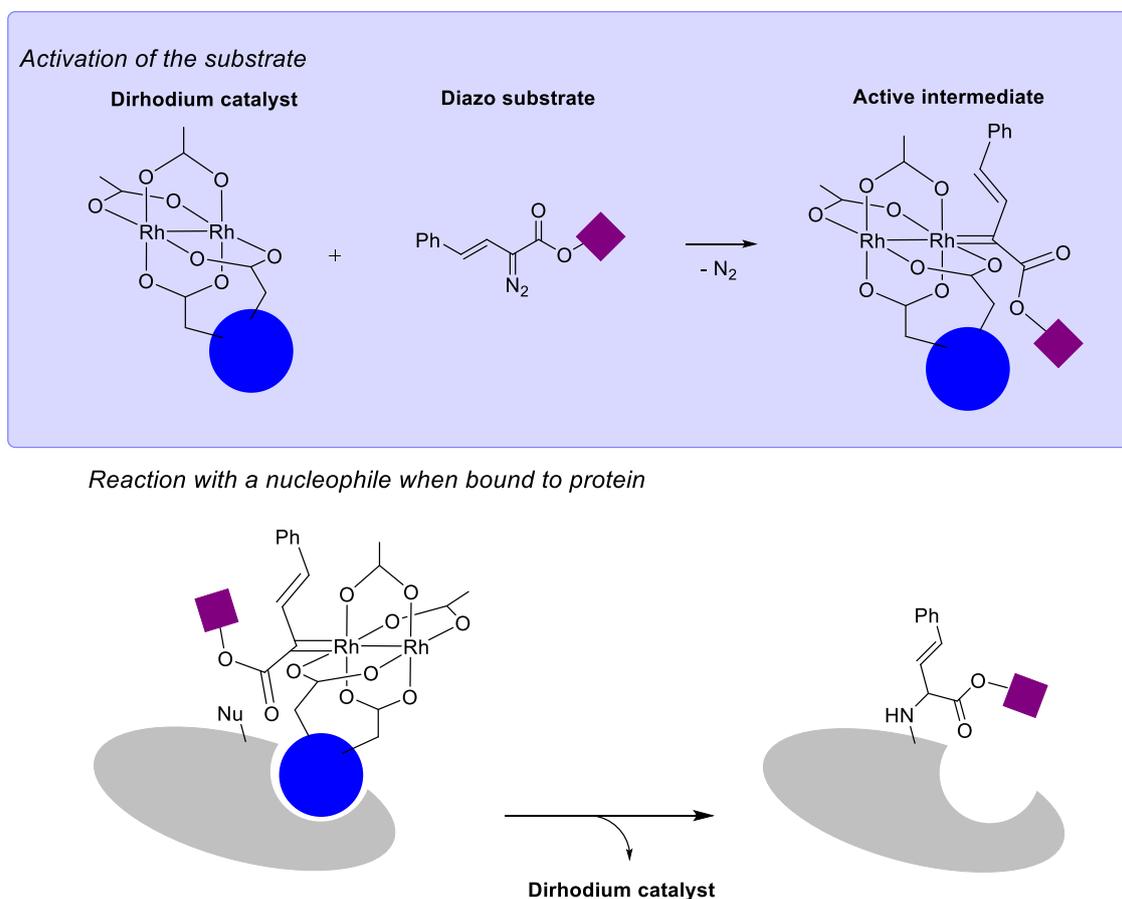
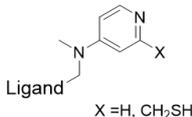
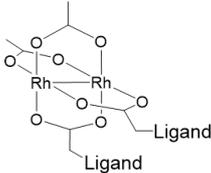
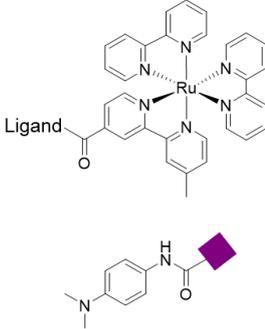
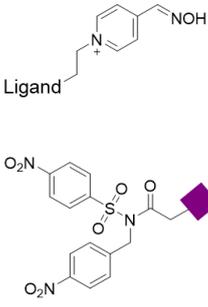
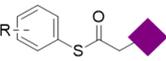
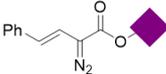
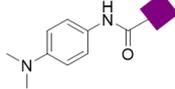
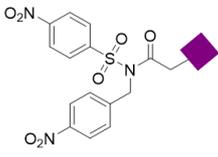


Figure 23: Formation of highly reactive intermediates from the reaction between one dirhodium catalyst and diazonium substrate. At the bottom, scheme of the reaction between the electrophilic activated moiety and a nucleophile moiety present on the surface of the protein in proximity to the binding site of the ligand.

Table 2: Summary of catalyst-based ligand-directed chemistries

	<b>DMAP (2008)</b>	<b>LD-diRhodium (2012)</b>	<b>LD-Ru(pby)<sub>3</sub> BB (2013)</b>	<b>LDOx (2017)</b>
<b>Catalyst and substrate of CatLD chemistry</b>				
				
	DMAP and DMAP-SH + thioester substrate	Dirhodium catalyst + diazo substrate	Ru(pby) <sub>3</sub> catalyst + dimethylaniline substrate	Pyridinium oxime catalyst + NASA substrate
<b>Labelled amino acid</b>	Lys, Ser, Tyr	Trp, Tyr, Phe, Asn, Gln, and others	Tyr	Lys, Ser
<b>Labelling time (h)</b>	< 3	16	1	1
<b>Binding affinity observed for modification</b>	$K_d < 43.5 \mu\text{M}$	$K_d = 0.65 \mu\text{M}$	Not known	$K_d < 0.26 \mu\text{M}$
<b>Labelling conditions</b>	Purified protein, cell surface, nuclei, <i>in-vivo</i>	Purified protein, cell lysate	Purified protein, cell lysate	Purified protein, cell surface, cell lysate, live tissue
<b>Protein targeted</b>	Congerin II, FKBP12, HER2, EGFR, nucleosome and more	Fyn SH3 domain	Erythrocyte lysate	hCAII, hCAXII, FKBP12, AMPAR, folate receptor
<b>References</b>	78,120,122,123	126–128	129–131	110,132

### Ruthenium CatLD chemistry<sup>129–131</sup>

Sato and coworkers described a unique catalyst-based ligand-directed approach using for the first time a photoredox catalyst based on a ruthenium complex Ru(bpy)<sub>3</sub>. Its unique feature is that the catalyst tethered to the ligand can be excited under light irradiation and subsequently activate not its substrate but the targeted amino acid. Indeed, this chemistry generates tyrosyl radicals at tyrosine residues in proximity to the ligand binding site of the POI, Figure 24. These radicals could be then trapped with dimethylaniline derivatives acting as substrates and be used as fluorophores for biosensing application. While the mechanism has not yet been fully detailed, and different pathways are being studied,<sup>129</sup> Figure 24 represents the most likely route in which the radical is formed on the tyrosine

residue. This approach was not only developed with dimethylaniline moiety but also phthalic hydrazide moieties as substrate. After optimisation, it was found that using 2 equivalents of the ligand catalyst under irradiation with visible light for 15 min, the modification was exclusively found at tyrosine residues. The authors reported successful selective labeling of erythrocyte lysate.

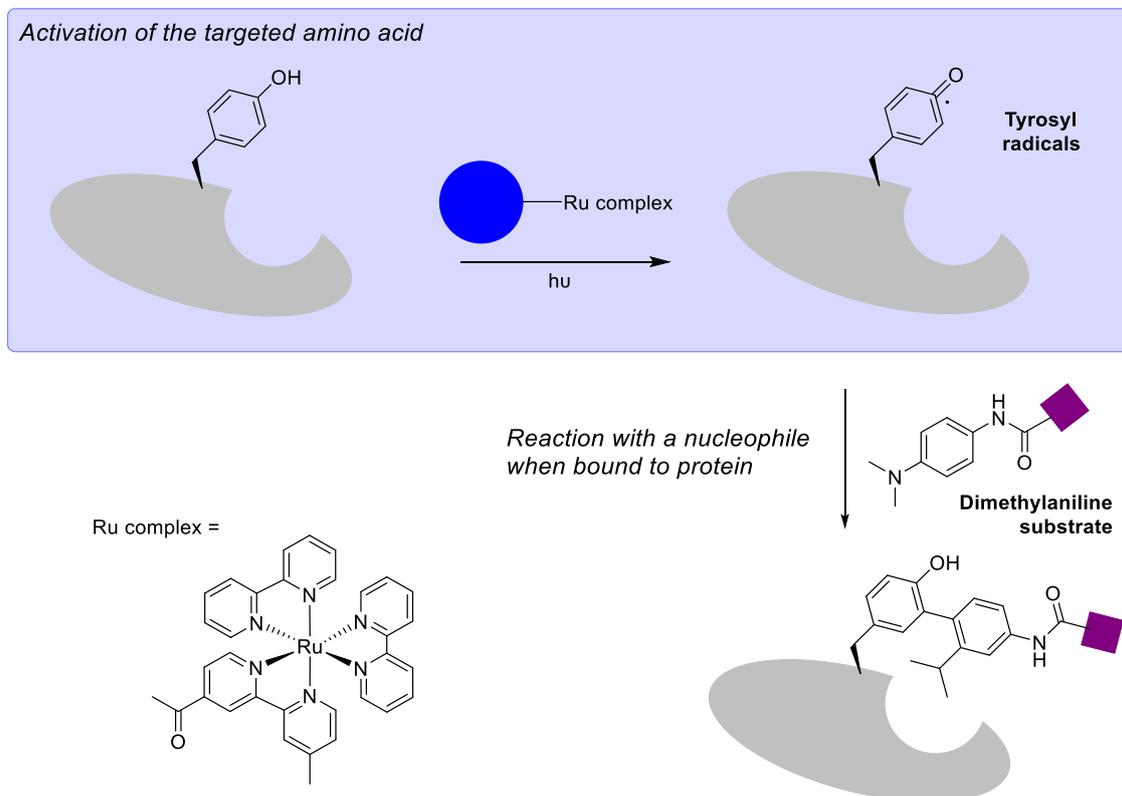


Figure 24: Formation of highly reactive tyrosyl radical from the excitation state of the Ruthenium complex. Followed by reaction with dimethylaniline substrate trapping the radical and modifying selectively tyrosine residues on the surface of the protein of interest.

#### *Oxime CatLD chemistry*<sup>132</sup>

The last *CatLD* described is based on a pyridinium oxime catalyst which catalyses acyl transfer from poorly electrophile *N*-acyl-*N*-alkylsulfonamides (NASA) to nucleophilic residues. While NASA reagents had already been mentioned in the previous section in *OpLD*, the main differentiation between the two is the type of electron-withdrawing group used on the sulfonamide. Indeed while previously used methylcyano substituents provided good substrates for *OpLD*, *CatLD* required a less reactive substrate found with nitrobenzyl group.<sup>110</sup> Indeed, the nitrobenzyl-NASA was shown to be much more stable than its analogous methylcyano-NASA, and must be activated by an oxime catalyst for protein labelling. The newly activated bond is shown in Figure 25. It was shown that using nitrobenzyl-NASA in a *OpLD* without the presence of the oxime catalyst resulted in no labeling showing the importance of the activation of this compound. This chemistry was applied not only for the labeling of isolated proteins but has also been used for the modification of cell lysates and mouse hippocampal and

cerebellar slices showing the diversity of NASA *CatLD*. On a critical note, no in-depth studies have compared reaction kinetics and their reactivity toward amino acids using the analogues in both *CatLD* and *OpLD* with the same ligand and same protein to properly compare the two chemistries. In this PhD, these probes were used as a major component, see part 1.3 Overview of the goals of the PhD.

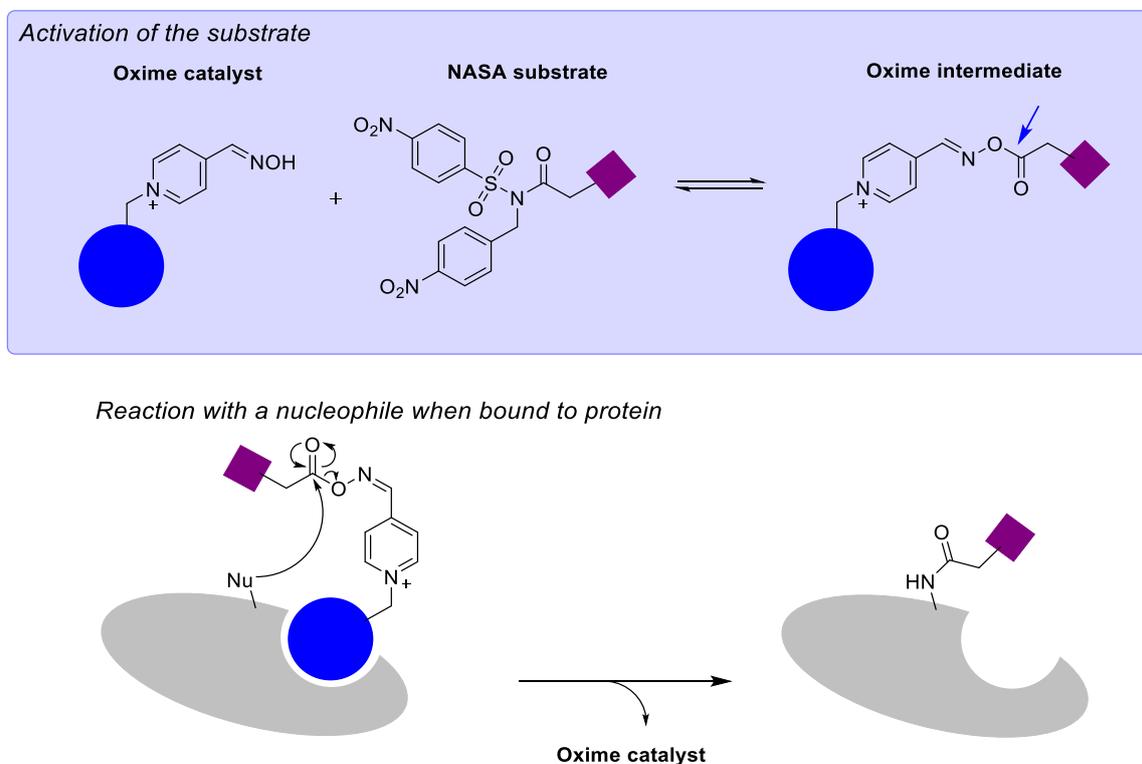


Figure 25: Formation of highly reactive intermediate from the reaction between an oxime catalyst and NASA substrate for Oxime *CatLD* chemistry.<sup>132</sup> At the bottom, scheme of the reaction between the electrophilic activated carbonyl and a nucleophile moiety present on the surface of the protein in proximity of the binding site of the ligand.

### 1.1.3 A conclusion on both chemoselective and regioselective chemistries for protein labelling

The first section of this introduction highlighted the main chemoselective and regioselective protein modifications. Chemoselective chemistries target specific amino acids such as lysine, cysteine. The reliability and robustness of these approaches make them still today the methods of choice for worldwide biotechnological applications such as immobilization protocols for enzyme-linked immunosorbent assays (ELISAs), microarrays, or polymer beads. Other less frequently targeted residues include tyrosine and the *N*-terminus mentioned in previous sections but also methionine (being the second least abundant of the twenty amino acids), and tryptophan (being the least abundant of the twenty amino acids).<sup>1,4</sup> The different approaches highlight diversity and different levels of selectivity.<sup>1,2</sup> Table 3 categorizes the amino acid residue described in this section and their corresponding types of chemical transformation. It highlights the fact that some chemistries have been used now for almost a century with historical successes, but also the new range of chemistries developed more recently which have allowed new types of amino acids to be targeted, and finally several remaining challenges, or rather, opportunities for invention.

A bright general note, which will probably lead to much more potential for chemoselective chemistry, is the use of new type of chemistries, such as computer-assisted design, photo-chemistry, transition-metal chemistry, for which some examples have been given above.

The second approach using regioselective protein modification methodologies enables selective modification of non-genetically modified proteins both on purified proteins and in natural cells. These strategies have led to different applications which have contributed greatly to the understanding of protein biology, such as through visualization of endogenous proteins using fluorophore substrates,<sup>119</sup> the study of protein-protein interactions,<sup>110</sup> the design of covalent inhibitors and even drug screening in live cell environments.<sup>108</sup> However, these approaches still have limitations.

Table 3: Summary of the different residues easy to target through chemoselective chemistry for protein modification and their different known chemistry approaches.  $\odot$  : Fully compatible (protein modification achieved in aqueous mild conditions, multiple modes of reactivity and peer-reviewed).  $\star$  : Partially compatible (reactions which have shown less selectivity or stability).  $\blacktriangle$  : Preliminary results  $\emptyset$  Unknown (no literature on the matter)

Method vs Residue	Cysteine	Lysine	Methionine	Tryptophan	Tyrosine	Dha
Alkylation	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\emptyset$
Arylation	$\odot$	$\star$	$\emptyset$	$\blacktriangle$	$\blacktriangle$	$\emptyset$
Acylation	$\emptyset$	$\odot$	$\emptyset$	$\blacktriangle$	$\blacktriangle$	$\emptyset$
Halogenation	$\blacktriangle$	$\emptyset$	$\emptyset$	$\blacktriangle$	$\blacktriangle$	$\blacktriangle$
Oxidation	$\odot$	$\blacktriangle$	$\odot$	$\star$	$\star$	$\emptyset$
1,4-Addition	$\odot$	$\blacktriangle$	$\emptyset$	$\emptyset$	$\emptyset$	$\odot$
Condensation	$\emptyset$	$\odot$	$\emptyset$	$\star$	$\star$	
Radical reaction	$\odot$	$\emptyset$	$\emptyset$	$\odot$	$\star$	$\emptyset$
Transition-metal functionalisation	$\odot$	$\blacktriangle$	$\emptyset$	$\odot$	$\odot$	$\odot$

First, higher labelling efficiency, and improved kinetics are necessary to make these approaches more widely usable to label biologically interesting intracellular proteins in their native environment. The benefit of specificity is in a way also a detriment as to maximise labelling protein-specific probes are required. This means that a new probe for every protein of interest has to be designed, which can be time consuming and challenging. In order to target specific amino acids, the linker moiety between the ligand and reactive species plays a major role in controlling the proximity between the reactive moiety and the reactive amino acid residue in the POI, adding another parameter to study for each protein. In addition, key parameters for any type of protein labeling such as water solubility, stability, and cell permeability/retention need to be considered in the ligand design. Indeed, it was shown that some of the probes were too easily hydrolysed, e.g. LDAI probes, or too bulky, e.g. LDBB probes, to provide high labelling efficiencies.<sup>104,117</sup> Moreover as emphasized in both Table 1 and Table 2, the type of nucleophilic amino acids in proximity to the ligand-binding pocket of the POI are not necessarily known in advance if no crystal structure has been produced, thus researchers need to carefully select a suitable method that tailors the choice of chemistry, the ligand and the linker to be able to target effectively the POI. Despite these limitations, regioselective chemistries offer a powerful way to

modify native proteins. A final criterion for any protein modification, is to observe how the modification impacts the protein, on both structure and biological functions.

With this in mind, this PhD is based on furthering regioselective approaches by exploiting chemoselectively modified peptides as ligands for ligand-directed protein modification. In the next section, we will go on to discuss which proteins were targeted within this PhD.

## 1.2 Proteins of interest

### 1.2.1 *A clinical need for improved biomaterials*

For decades now the tissue engineering science described by Langer and Vacanti as "an interdisciplinary field that applies the principles of engineering and life sciences toward the development of biological substitutes that restore, maintain, or improve [Biological tissue] function or a whole organ" has progressed giving more and more possibility to cure diseases worldwide. One of the approaches from tissue engineering is to culture cells within a matrix (often referred to as a scaffold) forming a biomaterial. In recent years, it has been shown that despite incredible progress, synthetic biomaterials are not optimal scaffolds for tissue growth. This is mainly due to their lack of intrinsic biological properties. However, if it was possible to extend the reach of tissue engineering by functionalising biomaterials with proteins and thus enhancing the biochemical signalling capabilities of these materials, this would represent a major advance. Progress in this area could potentially revolutionise how we treat diseases such as heart disease, diabetes, and osteoarthritis. While research in this area is widespread, this PhD aims to improve on biomaterials designed for the treatment of heart disease. To do so a first paragraph will detail the scope and importance of this worldwide challenge and the following paragraphs will introduce the protein of interest to this PhD and its relevance to the treatment of heart disease.

#### 1.2.1.a A fight against heart diseases

It is important to understand the impact of common cardiovascular diseases that could be prevented or reduced if better biomaterials able to control new blood vessel formation and tissue development were available. The first condition out of the range of common cardiovascular diseases is coronary heart disease. It was estimated that in 2022, more than 200 million people were living with this disease globally.<sup>133</sup> This disease is generally caused by the blockage of arteries due to a build-up of fat inside the vessels, a process known as atherosclerosis. This effect leads to a reduction in the volume of the artery, decreasing the quantity of blood delivered to the heart. When the whole surface of the vessel is eventually covered, blood cannot be transferred anymore. Thus, the heart which does not receive any blood supply will begin to suffer and the longer the shortage of blood, the larger the area of damaged heart tissue will be. The second condition is heart attack also known as myocardial infarction, a condition which is, in many cases, caused by coronary heart disease.<sup>134</sup> Heart attacks are common as in the United States alone every forty seconds someone suffers from a heart attack and they are responsible for approximately 13% of deaths in the US in 2016. In the UK every five minutes a hospital admission is due to heart attack. Additionally, each year around 60 million people across the world develop a heart or circulatory disease, a number equivalent to the entire population of the UK.<sup>133</sup>

A direct result of these diseases is a weakened heart. The heart even more so than many other organs critically depends on good blood circulation and needs a strong network of blood vessels to function properly. Thus, being able to create a biomaterial in which potent growth factor proteins could induce faster and more effective growth of blood vessels, essential for the heart activity, would impact massively the medical world allowing patients to recover more rapidly after heart attack. Moreover, the direct and indirect costs linked to people suffering from heart disease for the period 2014-2015 were estimated to be \$218.7 billion.<sup>135</sup> This makes heart disease one of the 10 most expensive conditions treated in US hospitals.<sup>135</sup> Thus new biomaterials would not only have a major impact in the medical field but would also make a financial impact.

#### 1.2.1.b Angiogenesis

The process of forming and expanding blood vessels within a tissue is called angiogenesis. *In vivo* the majority of tissues need blood vessels which supply cells with nutrients and oxygen through a highly branched network of capillaries. This network has very specific and controlled properties such as the distance between vessels, which has to be around 200  $\mu\text{m}$  correlating to the diffusion limit of oxygen through tissue.<sup>136</sup> When oxygen transfer is inefficient or under hypoxic conditions the result leads to poor cell growth and ultimately the death of surrounding tissue. Thus, for successful tissue engineering, it is not only sufficient that materials promote the growth of new tissues, but they also have to enhance the growth of blood vessels to supply oxygen to this tissue. The first report of angiogenic therapy of coronary heart disease in humans was published in 1998 using free protein however still no biomaterial has shown good efficiency.<sup>137</sup>

Different growth factor proteins are known to promote angiogenesis such as vascular endothelial growth factor (VEGF), platelet-derived growth factor (PDGF) and fibroblast growth factor (FGF). These growth factors all promote endothelial cell proliferation and their physical organization into tube-like structures, leading to the growth of new blood vessels from the pre-existing vasculature. However, it has been shown that FGF-2 was more potent than VEGF and PDGF used for this application.<sup>138</sup> FGF-2 has shown great potential for cardiac tissue engineering and was thus picked as the first target of interest for this PhD project. The coming paragraphs gives a brief introduction on the discovery of FGF-2, its main characteristics, and its use in tissue engineering.

### *1.2.2 Fibroblast growth factor 2, FGF-2*

#### 1.2.2.a FGF, an introduction to the fibroblast growth factors

FGF-2 is a member of the fibroblast growth factor (FGF) family. FGFs were discovered in the late 1970s by Armelin and Gospodarowicz who observed in a bioassay that a protein was causing a specific type of cell, fibroblasts, to proliferate, leading to the naming of this protein as "fibroblast growth factor".<sup>139</sup>

Indeed, FGF is a mitogen of fibroblasts, inducing and accelerating their cell division upon binding to specific receptors (FGFR) on the surface of the cells as shown in Figure 26. The role of FGFs in tissue biology is critical as fibroblasts are among the cells responsible for foetal development and wound healing, producing the extracellular matrix that is key to the formation of tissues such as muscle and skin.<sup>140</sup>

Following this first observation of FGF activity, the same groups were able to identify two different forms of the protein using either acidic or basic pH which led to the distinction of two members of the FGF family, the "acidic fibroblast growth factor" (FGF-1, aFGF) and the "basic fibroblast growth factor" (FGF-2, bFGF).<sup>141</sup> While these two proteins share a high degree of similarity within their sequence, they are distinct proteins. While Armelin and Gospodarowicz were the first to describe the effect of FGFs on fibroblast proliferation, other characteristics of FGF-1 and FGF-2 were subsequently shown by other groups, mainly Burgess and Maciag, such as the fact that these two proteins are both heparin-binding growth factors, and that they both enhance proliferation of blood vessel endothelium cells.<sup>142</sup> Thus FGF-1 and FGF-2 were given alternative names based on these distinctive characteristics such as HBGF-1 (heparin relative), ECGF-1 (endothelium cell) and HBGF-2, ECGF2 respectively.<sup>143</sup> Though it was clear that these two proteins were important signalling proteins for many cell types, the name fibroblast growth factor was kept despite its limitations. To avoid confusion, basic and acidic fibroblast growth factor will be referred to in this introduction as FGF-2 and FGF-1 respectively. Subsequently, 21 other members of the FGF family have been identified with diverse roles in biologic processes such as invasion, morphogenesis, tumor growth, and hormonal effects. Only the basic fibroblast growth factor, FGF-2, which has been described as the most potent for angiogenesis will be discussed further in detail in this section and will be targeted for modification in the following chapters.

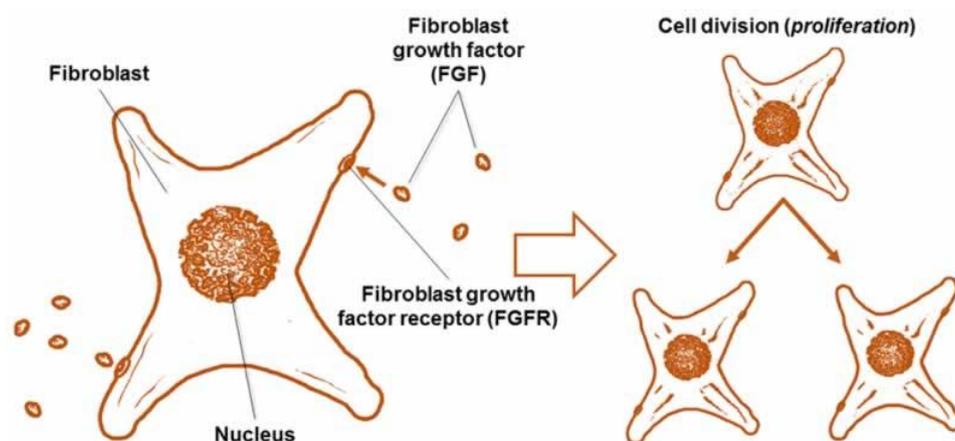


Figure 26: Picture from the online article "Basic-fibroblast-growth-factor-good-bad-stable" showing the action of mitosis from FGF on fibroblast cell when bound to fibroblast growth factor receptor<sup>144</sup>

### 1.2.2.b FGF-2

#### Structure

FGF-2 can be produced as a protein of varying length depending on alternative start codons, extending the *N*-terminus by 41, 46, 55, or 133 amino acids, mainly depending on the cell of origin. However, its primary structure is based on a 155 amino acid polypeptide, resulting in an 18 kDa protein.<sup>142,145,146</sup> This main polypeptide chain is considered to be cytoplasmic and can be secreted from the cell, whereas the higher molecular weight forms are directed to the cell's nucleus. It has been shown that the variation throughout the sequence and the length of the chain can vary between species but more remarkably can also vary in humans depending on the tissue of origin.<sup>146</sup> However a minimum of 24 residues are completely conserved.<sup>147</sup>

The main characteristic of the 155 AAs sequence structure is shown in the representations in Figure 27. 12 Anti-parallel  $\beta$ -sheets are organized into a trigonal pyramidal structure, with three trefoils all containing four  $\beta$ -sheets.<sup>148</sup> Specific domains have been shown to play major roles, e.g. residues 13–30 and 106–129 as receptor-binding sites; serine 64 and threonine 112 as potential phosphorylation sites.<sup>148–150</sup> Moreover FGF-2 contains four cysteines, residues 34, 78, 96, 101. While it was originally hypothesised that an intramolecular disulfide bond was present between residues 34, 101, it has subsequently been shown that native FGF2 does not contain any disulfide bonds.<sup>147</sup> It was also shown that the receptor binding site was physically separated from the heparan sulfate binding site.<sup>151</sup> Its interaction with heparan sulfate proteoglycans in the extracellular matrix<sup>143</sup> leads to its main characteristic being an angiogenic growth factor.<sup>152</sup>

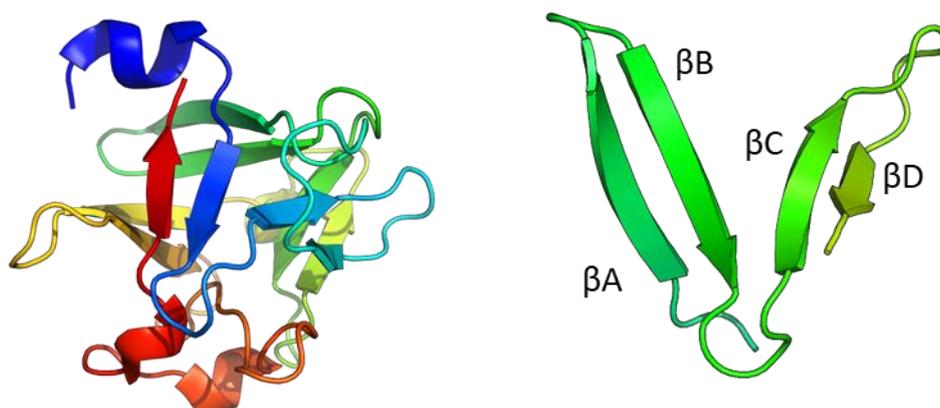


Figure 27 Crystal structure of FGF-2, PDB 4OEE, adopting a pseudo three-fold axis of symmetry from the  $\beta$ -trefoil, with on the left a representation of one of the three sets of four  $\beta$  strands. The first ascending strand ( $\beta$ A) is connected to a descending strand ( $\beta$ B). The following “horizontal” strand ( $\beta$ C) finishes by return strand ( $\beta$ D).<sup>150</sup>

#### Receptors

FGF-2, like other members of the FGF family, is a multifunctional protein, which can affect multiple cell types in different ways. All actions are triggered by the binding of the protein to the cell surface

FGF receptors (FGFRs), themselves a family of four closely related proteins. While FGF-1 was recognised as a universal ligand binding to each member of the receptor family, FGF-2 primarily interacts with FGFR-1. Binding of the receptor leads to intracellular signal transduction through activation of the receptor tyrosine kinase (RTK) pathway. When FGF-2 binds to its receptor, the FGFR dimerises enabling the phosphorylation of tyrosine residues in the intracellular kinase activation loop, as the first step of the RTK pathway. FGFRs also bind to extracellular heparan sulfate, leading to a ternary FGF-FGFR-heparan sulfate complex which can activate alternative intracellular signalling pathways.<sup>138,151,153</sup> More details on the structure and interaction of FGFRs and their downstream effects can be found in the literature, but it is important to be aware that a full understanding of FGF signalling is still far from being complete.

### Function

FGF-2's key role as a mitogen has already been introduced earlier and was one of the first characteristics discovered about this protein. A wide range of literature explains the diverse roles of FGF2 in the development of various organs<sup>146</sup> and its mechanism of action.<sup>138,151</sup> A summary of its different functions can be found in Table 4 and Table 5. In this thesis we will not focus extensively on FGF-2 function due to a primary focus on chemical modification. However, the next paragraphs will discuss briefly how this protein has already been used for tissue engineering including in clinical trials, and the particular interest in its angiogenic characteristics.

Table 4: Different functions of FGF2 and the targeted cells showing its multifunctionality within the body<sup>138,151</sup>

Function	Target Cell
Cell proliferation	Endothelial cell, epithelial cell, fibroblast cell, neural stem cell
Cell migration	Astrocyte, myogenic cell
Cell differentiation	Neuroepithelial
Angiogenesis	Endothelial cell

#### 1.2.2.c FGF-2 in engineering applications and method of applications

FGF-2 has already been used in several clinical tests as a repair/regeneration factor.<sup>140,146,151</sup> A summary of some of the clinical trials and tissue engineering data of FGF-2 is presented in Table 6.<sup>153-162</sup> The table also emphasizes the different application modes of FGF-2 within these technologies, ranging from being in soluble form (in either a gauze or through injection) or contained within a solid-support (mainly polymeric in the form of microparticles, scaffolds or hydrogels).

Table 5: Putative functions of FGF2 in various organs<sup>138,146,151</sup>

<b>Organ</b>	<b>Putative function</b>
<b>Brain</b>	Neuronal differentiation and survival
<b>Blood vessel</b>	Angiogenesis, muscle cell proliferation Atherogenesis, blood pressure control
<b>Lung</b>	Branching morphogenesis, fibrosis
<b>Limb</b>	Limb development
<b>Muscle</b>	Myogenesis
<b>Bone</b>	Osseous healing, chondrogenesis
<b>Reproductive system</b>	Spermatogenesis
<b>Eye</b>	Photoreceptor survival and transduction
<b>Skin</b>	Melanogenesis, tissue repair

A number of studies have shown the advantages of using FGF-2 containing gauzes for the treatment of burns and chronic wounds leading to significantly reduced times of treatment, and increased quality of scars with a major impact on the quality of life of patients.<sup>151,154,155,163</sup> A major review published in 2020 summarized 27 Japanese full-text articles including: 11 laboratory studies, 7 case reports, 4 clinical studies, and 5 randomized controlled trials using of FGF-2 in wound healing.<sup>164</sup> Moreover, a 0.01% FGF-2 preparation (brand name: Fiblast® Spray) obtained marketing approval in April 2001 to treat bedsores and skin ulcers. This was the first drug in the world to contain FGF-2.

It was shown that FGF-2 also enhanced bone growth. For the treatment of tibial bone fractures, two concentrations of FGF-2 in hydrogels have been tested against a placebo.<sup>165</sup> The dose was injected percutaneously twice with each time 0.5 mL of hydrogel (containing 0, 0.4, or 1.2 mg of FGF-2). Both doses (0.8 mg and 2.4 mg total) showed that bone union was significantly higher in the presence of the protein. Surprisingly though, the difference in healing for both doses of FGF-2 were very similar, showing that the smallest quantity was already sufficient in this specific application. It would have been interesting following this study to lower the dose further, as the quantity of FGF-2 needed will greatly impact the cost of the final biomaterial. Other studies showed the importance of FGF-2 for periodontal regeneration, and the company Kaken Pharmaceutical launched their second medicinal product, Regroth, containing FGF-2 in 2016.<sup>166,167</sup>

A number of studies also showed the impact of FGF-2 on the treatment of different types of ulcers, including diabetic ulcers, pressure ulcers, and aphthous ulcers. The protein was shown to shrink ulcers in a considerably shorter amount of time compared to placebo. For the case of diabetic ulcers, application was by spraying over the wound areas either 0.01% and 0.001% FGF-2 preparations.

Thanks to the application at proximity of the surface of the ulcer, it was shown that the area of the ulcer decreased by more than 75% after eight weeks versus 57.5% within the placebo group.

Moreover and more importantly for this PhD project, results have shown that FGF-2 can be delivered via extravascular and intra-vascular approaches leading to promising results with clear reduction of infarct size after heart attack.<sup>151</sup> A study showed the use of FGF-2 encapsulated in 10 µm microspheres of biodegradable poly(lactic-co-glycolic acid) containing each an average of 1 ng protein.<sup>160</sup> The authors showed that after 14 days, the number of blood vessels formed after injection were significantly higher when using the microsphere compared to the negative control, with a 4-fold increase in penetrating capillaries compared to the FGF-2-free group. They interestingly also showed the difference in using free FGF-2 compared to the use of microsphere to deliver the FGF-2. After 48 h, they showed that more cells were viable when treated with free FGF-2 compared to the use of FGF-2 encapsulated microspheres. However after 14 days post injection in mice, and despite FGF-2 being released from the microspheres mainly within the first day (96% in the first 24 hours, and the remaining 4% over the following 13 days), the group using the microspheres reported greater increase of both number of viable cells and blood vessels over the free FGF-2.<sup>160</sup> It is also important to note that injecting high concentrations of free protein at a specific place, such as in the first intra-vascular approach, can cause supraphysiological signalling, leading to diffusional loss and/or enzymatic inactivation/degradation and impact the channel of communication between cells and protein which can be a significant danger for the body.<sup>140,151</sup> Thus while the authors showed great potential the results are not yet optimal having such a high release of 96% within one day.

Table 6: Summary of tissue engineering examples, with target tissues and application modes, using FGF-2

Target tissues	FGF-2 application mode	References
Burns and chronic wounds	Soaked gauze, spray, intradermal injection	154,155,163
Bone repair	Hydrogel, polymeric scaffold, active glass	165,168,169
Ulcers	Protein sponge, spray	156–158
Limb ischaemia	Gelatin hydrogel	159
Vascularization	Gelatin hydrogel, embedded microspheres, polymeric scaffold	153,160–162

Another group showed the benefits of star polyethylene glycol (PEG)-heparin hydrogels functionalized with FGF-2 for encouraging angiogenic response in vitro using human endothelial cells from the

umbilical cord vein (HUVECs) cultured on a fibronectin surface.<sup>161</sup> The immobilisation of FGF-2 on the hydrogel was achieved using a cyclic RGD peptide as a binding tool. While this system shows promising results two elements have to be emphasised. First the attachment of FGF-2 was not protein specific as RGD is a known sequence to bind to a broad range of proteins and second compared to the previous report, no release over time can be done in this system.<sup>161</sup> Another similar approach was taken by Li and co-workers using heparin-immobilized porous biodegradable poly(lactic-co-glycolic acid) scaffolds.<sup>153</sup> The interaction between the scaffold and protein was mediated through ionic interaction between both *O*- and *N*-sulfate groups of heparin molecules and certain lysine and arginine residues in FGF-2. Depending on the loading of heparin the gel could contain from 10 to 30 ng of growth factor which could undergo slow release over up to 25 days. They showed the beneficial advantage of a controlled and sustained release using “immobilized” protein onto a biomaterial compared to the first study using loaded FGF-2 in microspheres. These results enhanced the formation of blood vessels in the vicinity of the subcutaneous pockets implant site of the mice.

Another attempt to immobilize FGF-2 onto a biomaterial was reported by Perets and co-workers. This approach is more selective as it depends on the inherent ability of FGF-2 to bind to heparin. The authors used an alginate biomaterial conjugated to heparin to embed FGF-2 into the microspheres.<sup>162</sup> It was shown that delivery took place via a slow release over 12 days. Interestingly this study showed not only an increase in the number of blood vessels while using the growth factor, but also the differences between the capillary's diameters. Most of the blood vessels from the FGF-2 releasing groups appeared to be large (>10 µm in diameter) and matured while when in the control composites, the majority of the blood vessels were micro-capillaries (< 10 µm in diameter).

Thus, it has been shown that FGF-2 can be of a significant help for the recovery of a number of tissues and shows great promises in the medical field with at least two products on the market. However, to date, no medical products have been developed using FGF-2 for vascularization. Moreover, no study has been reported to have used chemo-specific immobilization of the protein onto materials in clinical trials. This PhD will attempt to specifically modify and tether this protein covalently to a biomaterial scaffold, while maintaining its activity.

#### 1.2.2.d Known binding peptides to FGF-2

In this PhD, we aim to show that we can modify covalently FGF-2 using a ligand-directed approach (see part 2 of this Introduction) and then immobilise the protein onto a biomaterial using the new handle. To target this protein, the ligands used will be based on known binding peptides of FGF.<sup>170,171</sup>

Two studies in the literature have used phage display to find two libraries of hepta- and hexa-peptides binding to FGF-2.<sup>170,171</sup> Both these studies sought to identify peptides that could block FGF-2's

interactions with its receptor in cell membranes. While these approaches were used for an alternative application (inhibiting FGFR which has been shown to have a role in cancer progression),<sup>172</sup> the binding peptides identified that do not have inhibitory properties are attractive ligands for the modification of FGF-2 away from its active binding region.

However, FGF-2 is a complex, unstable, and expensive growth factor. Thus, in order to develop the underlying methodology, insulin was also studied as a cheap and readily accessible model protein.

### 1.2.3 Insulin

#### 1.2.3.a Insulin, an introduction

Insulin was isolated for the first time at the University of Toronto in 1921 and was purified the following year with the work of four scientists of that time, Sir Frederick G Banting, Charles H Best, John Macleod and James B Collip.<sup>173,174</sup> Very shortly after the start of their experiment, in January 1922, they injected insulin into the first human, and the second injection showed how important their discovery was due to the unprecedented decrease in blood sugar levels, becoming near-normal, with no obvious side effects. For the first time in history, type 1 diabetes was not a death sentence. In recognition of their life-saving discovery, Banting and Macleod were jointly awarded the 1923 Nobel Prize in Physiology or Medicine.<sup>173</sup> This discovery is considered one of the greatest medical breakthroughs in history, which went on to save millions of lives and triggered many more discoveries by many research groups across the world using insulin in the medical field.<sup>175</sup>

#### *Sequence and structure*

Following the discovery of insulin, much more work was done to understand the protein. A key moment was the determination of the protein sequence by Dr. Frederick Sanger who was awarded his first Nobel Prize in Chemistry in 1958 for this achievement. His pioneering study showed that human insulin B-chain consisted of 30 amino acids and the A-chain of 21 amino acids, with the B- and A-chains being held together by two interchain disulphide bonds (CysB7 to CysA7 and CysB19 to CysA20) and a third intra-chain disulphide bond (CysA6 to CysA11), Figure 28.<sup>176-179</sup> These three disulphide bonds bridging the chains together result in the production of three  $\alpha$ -helices (residues A1–A8, A12–A18 and B9–B19). Later on, it was shown that the insulin sequence was largely conserved between bovine, porcine and human insulin, suggesting that it has been conserved across much of animal evolutionary history.

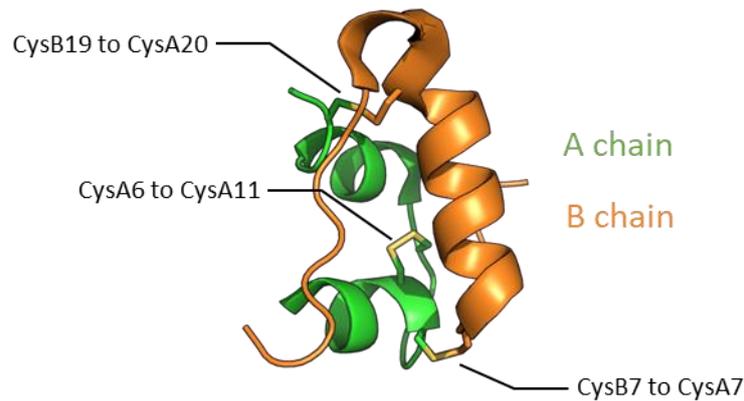


Figure 28: Crystal structure of insulin, PDB 3I40, highlighting the two chains and the three disulphide bonds of the protein. Its molecular mass is 5808 Da.

Another key discovery over the following decades was the work of Dorothy Hodgkin, a founder of X-ray crystallography, who determined insulin's 3D structure in 1969 after 35 years of research,<sup>173</sup> paving the way for even more scientific breakthroughs in diabetes research. Not only did this work enhance the understanding of insulin's mechanism of action but also promoted further research into structure-based drug design and enzyme-substrate interactions. Dr Hodgkin was awarded the 1964 Nobel Prize in Chemistry for her work on structure determination, not only for insulin but also penicillin and vitamin B<sub>12</sub>. She still remains the only British woman to have won this Award to date.

On an interesting note, researchers in the York Structural Biology Laboratory (YSBL) at the University of York, Department of Chemistry, were the first to establish the three-dimensional interactions between insulin and its receptor on cell surfaces. The study, led by Prof. Guy Dodson, himself a student of Dorothy Hodgkin, was published in *Nature*.<sup>180,181</sup> More information on the interaction between insulin and its receptor allowed a better understanding of insulin modification but also impacted the understanding of its role in cell growth and cancers.

### Function

Insulin is a highly specific protein produced by beta cells of the pancreatic islets. Its secretion and function are directly correlated to the amount of sugar in the body and play a crucial role in the regulation of the metabolism of carbohydrates and fats, by promoting the absorption of glucose from the blood into liver, fat and skeletal muscle cells. The action starts through the binding of the insulin to its receptor on the cell membrane of these cells. During the binding process, it has been shown that both insulin and its receptor undergo a conformational switch, Figure 29.<sup>181</sup>

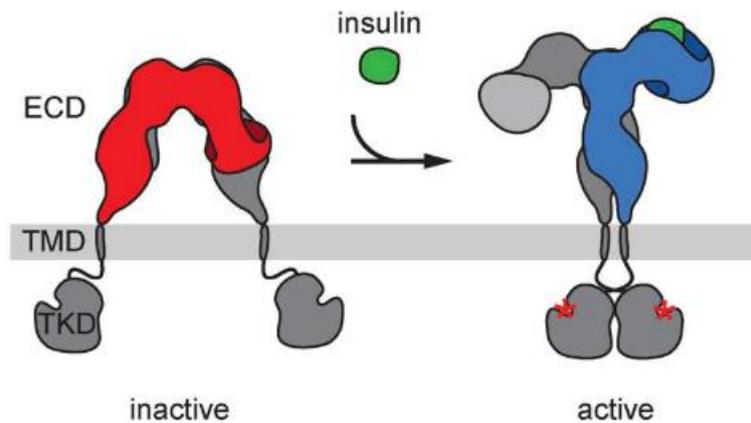


Figure 29: Schematic adapted from literature, showing the change of conformation of insulin receptor upon binding with insulin.<sup>182</sup>

The receptor protein contains  $\alpha$ - and  $\beta$  subunits. Two receptors are joined to form a homodimer. When insulin binds to the extracellular face of the  $\alpha$ -subunits of the homodimer, the  $\beta$  subunits undergo a conformational switch leading to a series of signal transduction cascades which are represented in the schematic Figure 30. The main action after binding is the intake of glucose from the blood through the Glut-4 transporter (3). The glucose is then used in different processes such as glycogen synthesis via glycogenesis (4), glycolysis (5) or triglyceride synthesis via lipogenesis (6) all necessary for the functioning of cells particularly during body activity as the need for energy increases. This mechanism has led to insulin being considered the main anabolic hormone of the body.

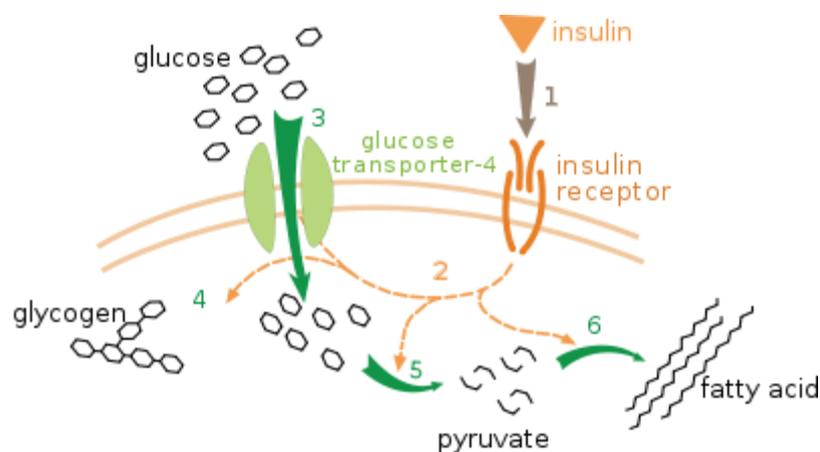


Figure 30: Effect of insulin on glucose uptake and metabolism. Insulin binds to its receptor (1), activating different pathways (2). First the intake of glucose through the Glut-4 transporter (3), the glucose itself then takes part in different processes including glycogen synthesis (4), glycolysis (5) and triglyceride synthesis (6). Scheme from XcepticZP released into the public domain ©

Therefore, when the insulin levels are low in the blood, the opposite effect happens called catabolism, leading to the reservation of body fat. Beta cells are able to respond to blood sugar levels, secreting insulin into the blood when high levels of glucose are detected, and inhibiting insulin secretion when glucose levels are low. Thus, the lack of secretion of insulin leads to catabolism being favoured over

anabolism, resulting in diabetes mellitus, a condition of high blood sugar level (hyperglycaemia). Today there are an estimated 463 million people living with this damaging disease.<sup>183</sup>

#### 1.2.3.b Insulin, its use in medicine

As introduced earlier, insulin has been mainly used to target diabetes through a range of products over the last century by two main companies, Novo Therapeutisk in Denmark and Eli Lilly and Co. in the USA. Its use in tissue engineering was also extended in recent work, such as insulin-loaded poly(D,L-lactide-co-glycolide) PLGA microspheres by Andreas and co-workers.<sup>184</sup> They compared different microencapsulation techniques, their effect on the loading of insulin, its release over time for drug delivery and the activity of the insulin after delivery. Through suitable design, microspheres showed sustained release of structurally intact and biologically active insulin that promoted the formation of cartilage-specific extracellular matrix thanks to the action of insulin promoting the secretion of proteoglycans and collagen type II. Thus, they showed the potential for delivery devices for insulin-loaded microspheres to find use in cartilage tissue engineering. A few years later, Ramos and co-workers showed the use of insulin for another application in tendon tissue engineering using a polycaprolactone scaffold.<sup>185</sup> They immobilized insulin onto the polymer and showed that cells exposed to insulin in a range of 100- or 200-ng/mL showed better expression of scleraxis and collagen after 14 days. This research demonstrated the potential for using alternative smaller proteins compared to the traditional use of growth factors in tissue engineering, with good results for cell stimulation and differentiation towards desired lineages. It also showed the benefits of covalent conjugation of proteins to material supports, which can extend the half-life of the protein compared to the addition of soluble insulin, providing a more extended exposure. However, in this process, attachment of insulin to the material was uncontrolled and likely led to a significant drop off in signalling activity, though no studies were performed to determine this. Thus, further research on the specific immobilization of proteins such as insulin would be highly beneficial with the potential to maintain signalling activity if more controlled modification could be achieved.

#### 1.2.3.c Known peptides to insulin

Similarly to FGF-2, peptides that are known to bind insulin have already been reported in the literature. Four of these sequences come from three distinct studies which will be used in the following chapter. The earliest study showed the use of a six amino acid peptide binding to insulin. Their work was based on the interaction between insulin and its receptor, using the complementary strand of the insulin domain binding to the receptor leading to the sequence CVEEAS.<sup>186</sup> The authors showed that binding took place near the carboxyl terminus of the  $\beta$ -chain with high specificity and a  $K_d$  of 3 nM. While it was shown that when in presence of this peptide, the binding of insulin to its receptor was

inhibited by approximately 50%, they showed very interestingly that glucose uptake was the same, providing evidence that of the two domains of insulin one is responsible for receptor binding and one for signal translocation across the membrane. A following study used a heptapeptide found through phage display to bind insulin. The authors showed that after three rounds of screening, the last 15 clones showing highest ELISA signals all had the same sequence, HWWWPAS. They then used this peptide for affinity chromatography of insulin by immobilising the peptide on a Sepharose column chromatography.<sup>187</sup> Finally, a more recent study identified two peptides of which each sequence is actually derived from the insulin primary sequence itself, LVEALYL (fragment B11–B17) and RGFFYT (fragment B22-27). These peptides have been shown to play a role in the self-assembly of insulin.<sup>188</sup> These four peptides were synthesised and used in the following chapters.

### 1.3 Overview of the goals of the PhD

This PhD takes place in a context in which, despite the huge promise of the ‘tissue engineering’ field, very few biomaterial-based technologies have so far reached patients. The main roadblock to clinical translation is the inability to design materials which provide cells with the potent biological stimuli they require to form functional new tissue. Thus, the covalent modification of biomaterials with growth factors, powerful signalling proteins, would allow the science community to make a real impact on the treatment of disease. However, as has been mentioned at the start of this Introduction, modifications of proteins are difficult to achieve in a selective manner and if not controlled can lead to a dramatic loss of protein activity (often >95 % loss). As an alternative, this project will develop an innovative strategy for functionalising biomaterials with growth factors using ligand-directed approaches. In doing so, we hope to create a novel technology that enables us to produce the next generation of bioactive materials, with far-reaching impact in the treatment of disease.

Both target proteins FGF-2 and insulin have now been introduced in Part 2 of this Introduction, as well as their respective binding peptides. In this PhD, two main approaches will be undertaken to modify the surface of these two proteins. Each will be based on ligand-directed chemistry depending on two different types of ligands, Figure 31. The first approach will use peptides that can be evolved to bind any target protein, using cutting-edge phage display evolution techniques that were awarded the 2018 Nobel Prize in Chemistry. The main advantage of using peptides is that in a very short manner of time, a wide range of sequences can be identified, and they can bind at very specific, but at the same time varied, places on the protein surface, providing different potential local environments for modification. The second approach will develop a novel class of reagents for modifying proteins in proximity to their *N*-termini, using 2-PCA functional groups to direct catalysis to this site.

While a wide range of ligand-directed approaches have been developed in the last decade, two main groups were picked for exploration during this PhD. LDAI was a first choice due to the simplicity of probe design and synthesis. Second the use of NASA reagents was studied. As mentioned during the introduction, methylcyano-NASA and nitrobenzyl-NASA acyl donors can be exploited for respectively one-pot ligand-directed chemistry and catalyst-based ligand-directed chemistry. Those two probes, while very similar, have never been fully compared providing an opportunity to study in-depth the difference of action between *OpLD* and *CatLD*.

We anticipated using these chemistries to introduce diverse functional groups such as fluorophores, biotin or click-chemistry handles for targeting and further tethering of the modified protein onto a biomaterial.

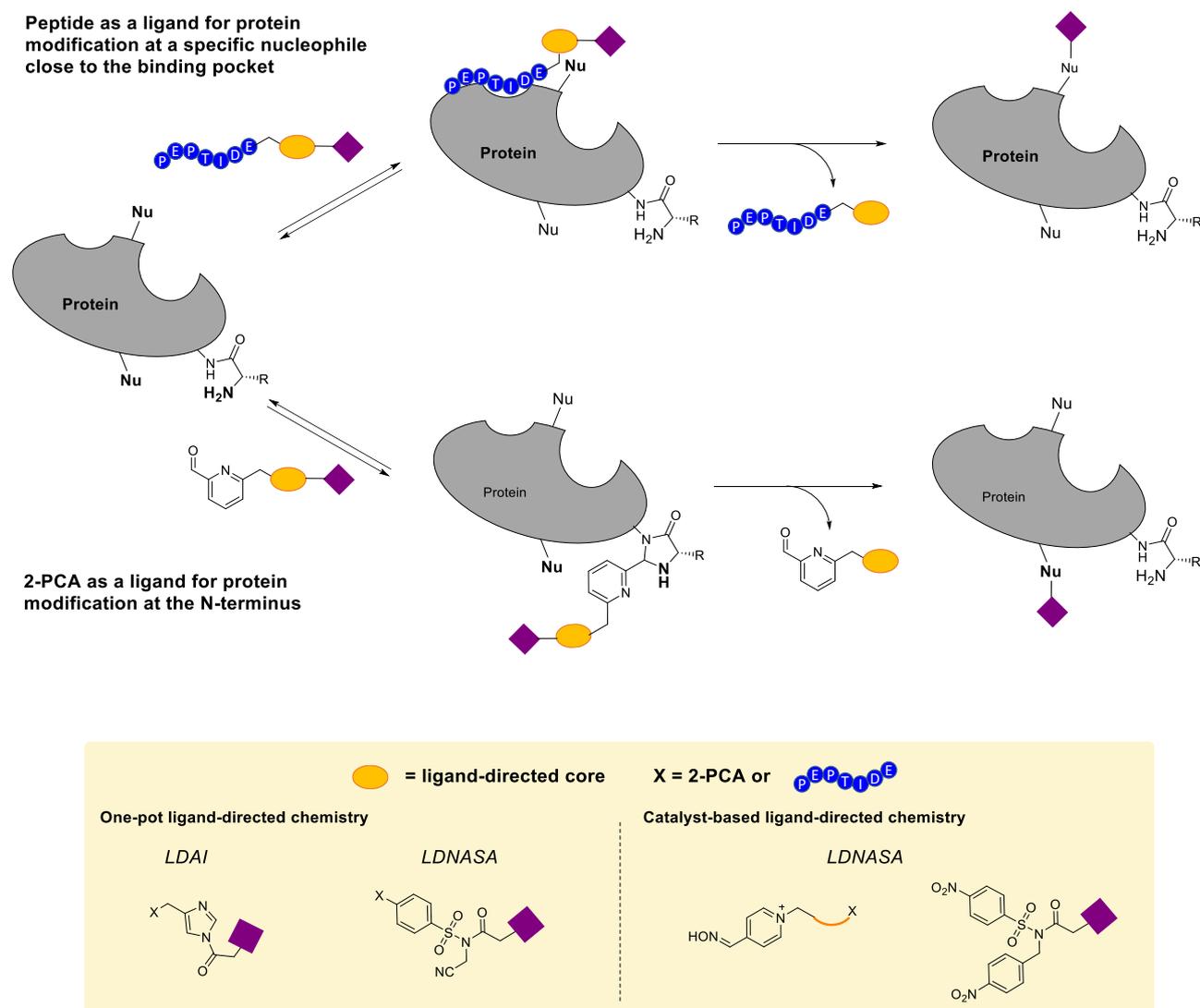


Figure 31: The two approaches considered in this PhD to modify protein via ligand-directed chemistry using a peptide for a ligand in the top scheme, or a 2-PCA reagent in the bottom scheme. In yellow the ligand-directed species detailed at the bottom and in purple the substrate. In the yellow box, probes picked for one-pot ligand directed-chemistry with both LDAI and LDNASA active electrophile reagents and for catalyst-based ligand-directed chemistry with the couple PyOx-NASA for both using either 2-PCA or peptide ligand.

Thus, this thesis will first present in Chapter 2 the synthesis and modification of peptide probes using solid-phase peptide synthesis and amidation reactions. Chapter 3 will then focus on characterising the binding affinity of the modified peptides using different techniques such as Quartz Crystal Microbalance with Dissipation monitoring (QCM-D), and Fluorescence Polarization (FP). This chapter will also introduce photo cross-linking probes developed in an attempt to decipher protein binding sites. Chapter 4 will describe in detail the synthesis of the substrates for ligand-directed chemistry, with the building blocks of both LDAI and LDNASA, and the modification of the binding peptides with these building blocks. Chapter 5 will then present the synthesis of *N*-terminus targeting 2-PCA ligands for use with LDNASA warheads. Finally, Chapter 6 will give an overview of the promising results obtained for the modification of proteins with both 2-PCA and peptide probes.

## 1.4 Bibliography

- (1) Akamoto, S. S.; Amachi, I. H. Recent Progress in Chemical Modification of Proteins. *Anal. Sci.* **2019**, *35* (January), 5–27. <https://doi.org/10.2116/analsci.18R003>.
- (2) Amaike, K.; Tamura, T.; Hamachi, I. Recognition-Driven Chemical Labeling of Endogenous Proteins in Multi-Molecular Crowding in Live Cells. *Chem. Commun.* **2017**, *53* (88), 11972–11983. <https://doi.org/10.1039/c7cc07177a>.
- (3) Spicer, C. D.; Pashuck, E. T.; Stevens, M. M. Achieving Controlled Biomolecule–Biomaterial Conjugation. *Chem. Rev.* **2018**, *118*, 7702–7743. <https://doi.org/10.1021/acs.chemrev.8b00253>.
- (4) Rawale, D. G.; Thakur, K.; Adusumalli, S. R.; Rai, V. Chemical Methods for Selective Labeling of Proteins. *European J. Org. Chem.* **2019**, *2019* (40), 6749–6763. <https://doi.org/10.1002/ejoc.201900801>.
- (5) Tantikanjanaporn, A.; Wong, M.-K. Development and Recent Advances in Lysine and N-Terminal Bioconjugation for Peptides and Proteins. *Molecules* **2023**, *28* (3), 1083. <https://doi.org/10.3390/molecules28031083>.
- (6) Hymel, D.; Liu, F. Proximity-Driven, Regioselective Chemical Modification of Peptides and Proteins. *Asian J. Org. Chem.* **2021**, *10* (1), 38–49. <https://doi.org/10.1002/ajoc.202000328>.
- (7) Justine, N. deGruyter; Malins, L. R.; Baran, P. S. Residue-Specific Peptide Modification: A Chemist's Guide. *Biochemistry* **2017**, *3*, 3863–3873. <https://doi.org/10.1021/acs.biochem.7b00536>.
- (8) Spicer, C. D.; Davis, B. G. Selective Chemical Protein Modification. *Nat. Commun.* **2014**, *5*. <https://doi.org/10.1038/ncomms5740>.
- (9) Hoyt, E. A.; Cal, P. M. S. D.; Oliveira, B. L.; Bernardes, G. J. L. Contemporary Approaches to Site-Selective Protein Modification. *Nat. Rev. Chem.* **2019**, *3* (3), 147–171. <https://doi.org/10.1038/s41570-019-0079-1>.
- (10) Boutureira, O.; Bernardes, J. L. Advances in Chemical Protein Modification. *Chem. Rev.* **2015**, *115*, 2174–2195. <https://doi.org/10.1021/cr500399p>.
- (11) Tamura, T.; Hamachi, I. Chemistry for Covalent Modification of Endogenous/Native Proteins: From Test Tubes to Complex Biological Systems. *J. Am. Chem. Soc.* **2019**. <https://doi.org/10.1021/jacs.8b11747>.
- (12) Shiraiwa, K.; Cheng, R.; Nonaka, H.; Tamura, T.; Hamachi, I. Chemical Tools for Endogenous Protein Labeling and Profiling. *Cell Chem. Biol.* **2020**, *27* (8), 970–985. <https://doi.org/10.1016/j.chembiol.2020.06.016>.
- (13) Krall, N.; Cruz, F. P.; Boutureira, O.; Bernardes, G. J. L. Site-Selective Protein-Modification Chemistry for Basic Biology and Drug Development. *Nat. Publ. Gr.* **2016**, *8* (February), 103–113. <https://doi.org/10.1038/nchem.2393>.
- (14) Kalkhof, S.; Sinz, A. Chances and Pitfalls of Chemical Cross-Linking with Amine-Reactive N-Hydroxysuccinimide Esters. *Anal. Bioanal. Chem.* **2008**, *392* (1–2), 1–8. <https://doi.org/10.1007/s00216-008-2231-5>.
- (15) Nakamura, T.; Kawai, Y.; Kitamoto, N.; Osawa, T.; Kato, Y. Covalent Modification of Lysine Residues by Allyl Isothiocyanate in Physiological Conditions: Plausible Transformation of Isothiocyanate from Thiol to Amine. *Chem. Res. Toxicol.* **2009**, *22* (3), 536–542. <https://doi.org/10.1021/tx8003906>.
- (16) Tanaka, K.; Masuyama, T.; Hasegawa, K.; Tahara, T.; Mizuma, H.; Wada, Y.; Watanabe, Y.; Fukase, K. A Submicrogram-Scale Protocol for Biomolecule-Based PET Imaging by Rapid 6 $\pi$ -Azaelectrocyclization: Visualization of Sialic Acid Dependent Circulatory Residence of Glycoproteins. *Angew. Chem.* **2008**, *120* (1), 108–111. <https://doi.org/10.1002/ange.200702989>.
- (17) Hacker, S. M.; Backus, K. M.; Lazear, M. R.; Forli, S.; Correia, B. E.; Cravatt, B. F. Global Profiling of Lysine Reactivity and Ligandability in the Human Proteome. *Nat. Chem.* **2017**, *9* (12), 1181–1190.

- <https://doi.org/10.1038/nchem.2826>.
- (18) Cal, P. M. S. D.; Veiros, L. F.; Cordeiro, C.; Vicente, B.; Pires, E.; Coelho, A. V.; Gois, P. M. P. Iminoboronates : A New Strategy for Reversible Protein Modification. *J. Am. Chem. Soc.* **2012**, *134*, 10299–10305. <https://doi.org/10.1021/ja303436y>.
- (19) Diethelm, S.; Schafroth, M. A.; Carreira, E. M. Amine-Selective Bioconjugation Using Arene Diazonium Salts. *Org. Lett.* **2014**, *16* (15), 3908–3911. <https://doi.org/10.1021/ol5016509>.
- (20) Asano, S.; Patterson, J. T.; Gaj, T.; Barbas, C. F. Site-Selective Labeling of a Lysine Residue in Human Serum Albumin. *Angew. Chem.* **2014**, *53*, 11783–11786. <https://doi.org/10.1002/anie.201405924>.
- (21) Matos, M. J.; Oliveira, B. L.; Martínez-Sáez, N.; Guerreiro, A.; Cal, P. M. S. D.; Bertoldo, J.; Maneiro, M.; Perkins, E.; Howard, J.; Deery, M. J.; Chalker, J. M.; Corzana, F.; Jiménez-Osés, G.; Bernardes, G. J. L. Chemo- and Regioselective Lysine Modification on Native Proteins. *J. Am. Chem. Soc.* **2018**, *140* (11), 4004–4017. <https://doi.org/10.1021/jacs.7b12874>.
- (22) Bischoff, R.; Schlüter, H. Amino Acids: Chemistry, Functionality and Selected Non-Enzymatic Post-Translational Modifications. *J. Proteomics* **2012**, *75* (8), 2275–2296. <https://doi.org/10.1016/j.jprot.2012.01.041>.
- (23) Lundell, N.; Schreitmüller, T. Sample Preparation for Peptide Mapping - A Pharmaceutical Quality-Control Perspective. *Anal. Biochem.* **1999**, *266* (1), 31–47. <https://doi.org/10.1006/abio.1998.2919>.
- (24) Chatterjee, C.; McGinty, R. K.; Fierz, B.; Muir, T. W. Disulfide-Directed Histone Ubiquitylation Reveals Plasticity in HDot1L Activation. *Nat. Chem. Biol.* **2010**, *6* (4), 267–269. <https://doi.org/10.1038/nchembio.315>.
- (25) Davis, B. G.; Maughan, M. A. T.; Green, M. P.; Ullman, A.; Jones, J. B. Glycomethanethiosulfonates: Powerful Reagents for Protein Glycosylation. *Tetrahedron* **2000**, *11* (1), 245–262. [https://doi.org/10.1016/S0957-4166\(99\)00497-8](https://doi.org/10.1016/S0957-4166(99)00497-8).
- (26) Gamblin, D. P.; Garnier, P.; Ward, S. J.; Oldham, N. J.; Fairbanks, A. J.; Davis, B. G. Glycosyl Phenylthiosulfonates (Glyco-PTS): Novel Reagents for Glycoprotein Synthesis. *Org. Biomol. Chem.* **2003**, *1* (21), 3642–3644. <https://doi.org/10.1039/b306990g>.
- (27) Davis, B. G.; Van Kasteren, S. I.; Kramer, H. B.; Gamblin, D. P. Site-Selective Glycosylation of Proteins: Creating Synthetic Glycoproteins. *Nat. Protoc.* **2007**, *2* (12), 3185–3194. <https://doi.org/10.1038/nprot.2007.430>.
- (28) Bernardes, G. J. L.; Grayson, E. J.; Thompson, S.; Chalker, J. M.; Errey, J. C.; El Oualid, F.; Claridge, T. D. W.; Davis, B. G. From Disulfide- to Thioether-Linked Glycoproteins. *Angew. Chem.* **2008**, *120* (12), 2276–2279. <https://doi.org/10.1002/ange.200704381>.
- (29) Chalker, J. M.; Bernardes, G. J. L.; Lin, Y. A.; Davis, B. G. Chemical Modification of Proteins at Cysteine: Opportunities in Chemistry and Biology. *Chem. - An Asian J.* **2009**, *4* (5), 630–640. <https://doi.org/10.1002/asia.200800427>.
- (30) Nathani, R. I.; Chudasama, V.; Ryan, C. P.; Moody, P. R.; Morgan, R. E.; Fitzmaurice, R. J.; Smith, M. E. B.; Baker, J. R.; Caddick, S. Reversible Protein Affinity-Labeling Using Bromomaleimide-Based Reagents. *Org. Biomol. Chem.* **2013**, *11* (15), 2408–2411. <https://doi.org/10.1039/c3ob40239h>.
- (31) Smith, M. E. B.; Schumacher, F. F.; Ryan, C. P.; Tedaldi, L. M.; Papaioannou, D.; Waksman, G.; Caddick, S.; Baker, J. R. Protein Modification, Bioconjugation, and Disulfide Bridging Using Bromomaleimides. *J. Am. Chem. Soc.* **2010**, *132* (6), 1960–1965. <https://doi.org/10.1021/ja908610s>.
- (32) Kalia, D.; Malekar, P. V.; Parthasarathy, M. Exocyclic Olefinic Maleimides: Synthesis and Application for Stable and Thiol-Selective Bioconjugation. *Angew. Chem., Int. Ed. Engl.* **2016**, *55* (4), 1432–1435. <https://doi.org/10.1002/anie.201508118>.
- (33) Koniev, O.; Leriche, G.; Nothisen, M.; Remy, J. S.; Strub, J. M.; Schaeffer-Reiss, C.; Van Dorsselaer,

- A.; Baati, R.; Wagner, A. Selective Irreversible Chemical Tagging of Cysteine with 3-Arylpropionitriles. *Bioconjug. Chem.* **2014**, *25* (2), 202–206. <https://doi.org/10.1021/bc400469d>.
- (34) Shiu, H. Y.; Chong, H. C.; Leung, Y. C.; Zou, T.; Che, C. M. Phosphorescent Proteins for Bio-Imaging and Site Selective Bio-Conjugation of Peptides and Proteins with Luminescent Cyclometalated Iridium(III) Complexes. *Chem. Commun.* **2014**, *50* (33), 4375–4378. <https://doi.org/10.1039/c3cc48376b>.
- (35) Conte, M. Lo; Staderini, S.; Marra, A.; Sanchez-Navarro, M.; Davis, B. G.; Dondoni, A. Multi-Molecule Reaction of Serum Albumin Can Occur through Thiol-Yne Coupling. *Chem. Commun.* **2011**, *47* (39), 11086–11088. <https://doi.org/10.1039/c1cc14402b>.
- (36) Li, F.; Allahverdi, A.; Yang, R.; Lua, G. B. J.; Zhang, X.; Cao, Y.; Korolev, N.; Nordenskiöld, L.; Liu, C. F. A Direct Method for Site-Specific Protein Acetylation. *Angew. Chem., Int. Ed. Engl.* **2011**, *50* (41), 9611–9614. <https://doi.org/10.1002/anie.201103754>.
- (37) Lercher, L.; Raj, R.; Patel, N. A.; Price, J.; Mohammed, S.; Robinson, C. V.; Schofield, C. J.; Davis, B. G. Generation of a Synthetic GlcNAcylated Nucleosome Reveals Regulation of Stability by H2A-Thr101 GlcNAcylation. *Nat. Commun.* **2015**, *6*. <https://doi.org/10.1038/ncomms8978>.
- (38) Raj, R.; Lercher, L.; Mohammed, S.; Davis, B. G. Synthetic Nucleosomes Reveal That GlcNAcylation Modulates Direct Interaction with the FACT Complex. *Angew. Chem., Int. Ed. Engl.* **2016**, *55* (31), 8918–8922. <https://doi.org/10.1002/anie.201603106>.
- (39) Vinogradova, E. V.; Zhang, C.; Spokoyny, A. M.; Pentelute, B. L.; Buchwald, S. L. Organometallic Palladium Reagents for Cysteine Bioconjugation. *Nature* **2015**, *526* (7575), 687–691. <https://doi.org/10.1038/nature15739>.
- (40) Al-Shuaeeb, R. A. A.; Kolodych, S.; Koniev, O.; Delacroix, S.; Erb, S.; Nicolaÿ, S.; Cintrat, J. C.; Brion, J. D.; Cianféran, S.; Alami, M.; Wagner, A.; Messaoudi, S. Palladium-Catalyzed Chemoselective and Biocompatible Functionalization of Cysteine-Containing Molecules at Room Temperature. *Chem. - A Eur. J.* **2016**, *22* (32), 11365–11370. <https://doi.org/10.1002/chem.201602277>.
- (41) Xiao, Q.; Tong, Q. X.; Zhong, J. J. Recent Advances in Visible-Light Photoredox Catalysis for the Thiol-Ene/Yne Reactions. *Molecules* **2022**, *27* (3). <https://doi.org/10.3390/molecules27030619>.
- (42) Choi, H.; Kim, M.; Jang, J.; Hong, S. Visible-Light-Induced Cysteine-Specific Bioconjugation: Biocompatible Thiol–Ene Click Chemistry. *Angew. Chem., Int. Ed. Engl.* **2020**, *59* (50), 22514–22522. <https://doi.org/10.1002/anie.202010217>.
- (43) Gunnoo, S. B.; Madder, A. Chemical Protein Modification through Cysteine. *ChemBioChem* **2016**, *17* (7), 529–553. <https://doi.org/10.1002/cbic.201500667>.
- (44) Bottecchia, C.; Noël, T. Photocatalytic Modification of Amino Acids, Peptides, and Proteins. *Chem. - A Eur. J.* **2019**, *25* (1), 26–42. <https://doi.org/10.1002/chem.201803074>.
- (45) Freedy, A. M.; Matos, M. J.; Boutureira, O.; Corzana, F.; Guerreiro, A.; Akkapeddi, P.; Somovilla, V. J.; Rodrigues, T.; Nicholls, K.; Xie, B.; Jiménez-Osés, G.; Brindle, K. M.; Neves, A. A.; Bernardes, G. J. L. Chemoselective Installation of Amine Bonds on Proteins through Aza-Michael Ligation. *J. Am. Chem. Soc.* **2017**, *139* (50), 18365–18375. <https://doi.org/10.1021/jacs.7b10702>.
- (46) Yang, A.; Ha, S.; Ahn, J.; Kim, R.; Kim, S.; Lee, Y.; Kim, J.; Söll, D.; Lee, H.; Park, H. A Chemical Biology Route to Site-Specific Authentic Protein Modifications. **2016**, *354* (6312), 623–626. <https://doi.org/10.1126/science.aah4428>.
- (47) Wright, T. H.; Davis, B. G. Post-Translational Mutagenesis for Installation of Natural and Unnatural Amino Acid Side Chains into Recombinant Proteins. *Nat. Protoc.* **2017**, *12* (10), 2243–2250. <https://doi.org/10.1038/nprot.2017.087>.
- (48) Gunnoo, S. B.; Finney, H. M.; Baker, T. S.; Lawson, A. D.; Anthony, D. C.; Davis, B. G. Creation of a Gated Antibody as a Conditionally Functional Synthetic Protein. *Nat. Commun.* **2014**, *5* (4388), 1–8. <https://doi.org/10.1038/ncomms5388>.

- (49) Guo, J.; Wang, J.; Lee, J. S.; Schultz, P. G. Site-Specific Incorporation of Methyl- and Acetyl-Lysine Analogues into Recombinant Proteins. *Angew. Chem., Int. Ed. Engl.* **2008**, *47* (34), 6399–6401. <https://doi.org/10.1002/anie.200802336>.
- (50) Bernardes, G. J. L.; Chalker, J. M.; Errey, J. C.; Davis, B. G. Facile Conversion of Cysteine and Alkyl Cysteines to Dehydroalanine on Protein Surfaces: Versatile and Switchable Access to Functionalized Proteins. *J. Am. Chem. Soc.* **2008**, *130* (15), 5052–5053. <https://doi.org/10.1021/ja800800p>.
- (51) Dhanjee, H. H.; Buslov, I.; Windsor, I. W.; Raines, R. T.; Pentelute, B. L.; Buchwald, S. L. Palladium-Protein Oxidative Addition Complexes by Amine-Selective Acylation. *J. Am. Chem. Soc.* **2020**, *142* (51), 21237–21242. <https://doi.org/10.1021/jacs.0c09180>.
- (52) Jones, M. W.; Strickland, R. A.; Schumacher, F. F.; Caddick, S.; Baker, J. R.; Gibson, M. I.; Haddleton, D. M. Polymeric Dibromomaleimides as Extremely Efficient Disulfide Bridging Bioconjugation and Pegylation Agents. *J. Am. Chem. Soc.* **2012**, *134* (3), 1847–1852. <https://doi.org/10.1021/ja210335f>.
- (53) Wang, Y.; Bruno, B. J.; Cornillie, S.; Nogueira, J. M.; Chen, D.; Cheatham, T. E.; Lim, C. S.; Chou, D. H. C. Application of Thiol–Yne/Thiol–Ene Reactions for Peptide and Protein Macrocyclizations. *Chem. - A Eur. J.* **2017**, *23* (29), 7087–7092. <https://doi.org/10.1002/chem.201700572>.
- (54) Schlick, T. L.; Ding, Z.; Kovacs, E. W.; Francis, M. B. Dual-Surface Modification of the Tobacco Mosaic Virus. *J. Am. Chem. Soc.* **2005**, *127* (11), 3718–3723. <https://doi.org/10.1021/ja046239n>.
- (55) Hooker, J. M.; Kovacs, E. W.; Francis, M. B. Interior Surface Modification of Bacteriophage MS2. *J. Am. Chem. Soc.* **2004**, *126* (12), 3718–3719. <https://doi.org/10.1021/ja031790q>.
- (56) Hahm, H. S.; Toroitich, E. K.; Borne, A. L.; Brulet, J. W.; Libby, A. H.; Yuan, K.; Ware, T. B.; McCloud, R. L.; Ciancone, A. M.; Hsu, K. L. Global Targeting of Functional Tyrosines Using Sulfur-Triazole Exchange Chemistry. *Nat. Chem. Biol.* **2020**, *16* (2), 150–159. <https://doi.org/10.1038/s41589-019-0404-5>.
- (57) Brulet, J. W.; Borne, A. L.; Yuan, K.; Libby, A. H.; Hsu, K. L. Liganding Functional Tyrosine Sites on Proteins Using Sulfur-Triazole Exchange Chemistry. *J. Am. Chem. Soc.* **2020**, *142* (18), 8270–8280. <https://doi.org/10.1021/jacs.0c00648>.
- (58) Tilley, S. D.; Francis, M. B. Tyrosine-Selective Protein Alkylation Using  $\pi$ -Allylpalladium Complexes. *J. Am. Chem. Soc.* **2006**, *128* (4), 1080–1081. <https://doi.org/10.1021/ja057106k>.
- (59) Sato, S.; Nakamura, K.; Nakamura, H. Horseradish-Peroxidase-Catalyzed Tyrosine Click Reaction. *ChemBioChem* **2017**, *18* (5), 475–478. <https://doi.org/10.1002/cbic.201600649>.
- (60) Borne, A. L.; Brulet, J. W.; Yuan, K.; Hsu, K. L. Development and Biological Applications of Sulfur-Triazole Exchange (SuTEx) Chemistry. *RSC Chem. Biol.* **2021**, *2* (2), 322–337. <https://doi.org/10.1039/d0cb00180e>.
- (61) Chen, Y.; Craven, G. B.; Kamber, R. A.; Cuesta, A.; Zherish, S.; Moroz, Y. S.; Bassik, M. C.; Taunton, J. Direct Mapping of Ligandable Tyrosines and Lysines in Cells with Chiral Sulfonyl Fluoride Probes. *Nat. Chem.* **2023**. <https://doi.org/10.1038/s41557-023-01281-3>.
- (62) Barber, L. J.; Yates, N. D. J.; Fascione, M. A.; Parkin, A.; Hemsworth, G. R.; Genever, P. G.; Spicer, C. D. Selectivity and Stability of N-Terminal Targeting Protein Modification Chemistries. *RSC Chem. Biol.* **2022**, No. i, 56–64. <https://doi.org/10.1039/d2cb00203e>.
- (63) Rosen, C. B.; Francis, M. B. Targeting the N Terminus for Site-Selective Protein Modification. *Nat. Chem. Biol.* **2017**, *13* (7), 697–705. <https://doi.org/10.1038/nchembio.2416>.
- (64) Dawson, P.; Muir, T.; Clark-Lewis, I.; Kent, S. B. H. Synthesis of Proteins by Native Chemical Ligation. *Science (80-. )*. **1994**, *266*, 776–779. <https://doi.org/10.1126/science.7973629>.
- (65) Dawson, P. E.; Kent, S. B. H. Synthesis of Native Proteins by Chemical Ligation. *Annu. Rev. Biochem.* **2000**, *69*, 923–960. <https://doi.org/10.1146/annurev.biochem.69.1.923>.
- (66) Bernardes, G. J. L.; Steiner, M.; Hartmann, I.; Neri, D.; Casi, G. Site-Specific Chemical Modification

- of Antibody Fragments Using Traceless Cleavable Linkers. *Nat. Protoc.* **2013**, *8* (11), 2079–2089. <https://doi.org/10.1038/nprot.2013.121>.
- (67) Faustino, H.; Silva, M. J. S. A.; Veiros, L. F.; Bernardes, G. J. L.; Gois, P. M. P. Iminoboronates Are Efficient Intermediates for Selective, Rapid and Reversible N-Terminal Cysteine Functionalisation. *Chem. Sci.* **2016**, *7* (8), 5052–5058. <https://doi.org/10.1039/c6sc01520d>.
- (68) Ren, H.; Xiao, F.; Zhan, K.; Kim, Y. P.; Xie, H.; Xia, Z.; Rao, J. A Biocompatible Condensation Reaction for the Labeling of Terminal Cysteine Residues on Proteins. *Angew. Chem., Int. Ed. Engl.* **2009**, *48* (51), 9658–9662. <https://doi.org/10.1002/anie.200903627>.
- (69) Ning, X.; Temming, R. P.; Dommerholt, J.; Guo, J.; Ania, D. B.; Debets, M. F.; Wolfert, M. A.; Boons, G.-J.; van Delft, F. L. Protein Modification by Strain-Promoted Alkyne-Nitrone Cycloaddition. *Angew. Chem.* **2010**, *122* (17), 3129–3132. <https://doi.org/10.1002/ange.201000408>.
- (70) Geoghegan, K. F.; Stroh, J. G. Site-Directed Conjugation of Nonpeptide Groups to Peptides and Proteins Via Periodate Oxidation of a 2-Amino Alcohol. Application to Modification at N-Terminal Serine. *Bioconjug. Chem.* **1992**, *3* (2), 138–146. <https://doi.org/10.1021/bc00014a008>.
- (71) Purushottam, L.; Adusumalli, S. R.; Singh, U.; Unnikrishnan, V. B.; Rawale, D. G.; Gujrati, M.; Mishra, R. K.; Rai, V. Single-Site Glycine-Specific Labeling of Proteins. *Nat. Commun.* **2019**, *10* (1), 1–9. <https://doi.org/10.1038/s41467-019-10503-7>.
- (72) Scheck, R. A.; Dedeo, M. T.; Iavarone, A. T.; Francis, M. B. Optimization of a Biomimetic Transamination Reaction. *J. Am. Chem. Soc.* **2008**, *130* (35), 11762–11770. <https://doi.org/10.1021/ja802495w>.
- (73) Gilmore, J. M.; Scheck, R. A.; Esser-Kahn, A. P.; Joshi, N. S.; Francis, M. B. N-Terminal Protein Modification through a Biomimetic Transamination Reaction. *Angew. Chem., Int. Ed. Engl.* **2006**, *45* (32), 5307–5311. <https://doi.org/10.1002/anie.200600368>.
- (74) Chan, A. O. Y.; Ho, C. M.; Chong, H. C.; Leung, Y. C.; Huang, J. S.; Wong, M. K.; Che, C. M. Modification of N-Terminal  $\alpha$ -Amino Groups of Peptides and Proteins Using Ketenes. *J. Am. Chem. Soc.* **2012**, *134* (5), 2589–2598. <https://doi.org/10.1021/ja208009r>.
- (75) Macdonald, J. I.; Munch, H. K.; Moore, T.; Francis, M. B. One-Step Site-Specific Modification of Native Proteins with 2-Pyridinecarboxaldehydes. *Nat. Chem. Biol.* **2015**, *11* (5), 326–331. <https://doi.org/10.1038/nchembio.1792>.
- (76) Lee, J. P.; Kassianidou, E.; MacDonald, J. I.; Francis, M. B.; Kumar, S. N-Terminal Specific Conjugation of Extracellular Matrix Proteins to 2-Pyridinecarboxaldehyde Functionalized Polyacrylamide Hydrogels. *Biomaterials* **2016**, *102*, 268–276. <https://doi.org/10.1016/j.biomaterials.2016.06.022>.
- (77) Mortensen, M. R.; Skovsgaard, M. B.; Gothelf, K. V. Considerations on Probe Design for Affinity-Guided Protein Conjugation. *ChemBioChem* **2019**, *20* (21), 2711–2728. <https://doi.org/10.1002/cbic.201900157>.
- (78) Hayashi, T.; Hamachi, I. Traceless Affinity Labeling of Endogenous Proteins for Functional Analysis in Living Cells. *Acc. Chem. Res.* **2012**, *45* (9), 1460–1469. <https://doi.org/10.1021/ar200334>.
- (79) Keppler, A.; Gendreizig, S.; Gronemeyer, T.; Pick, H.; Vogel, H.; Johnsson, K. A General Method for the Covalent Labeling of Fusion Proteins with Small Molecules in Vivo. *Nat. Biotechnol.* **2003**, *21* (1), 86–89. <https://doi.org/10.1038/nbt765>.
- (80) Gautier, A.; Juillerat, A.; Heinis, C.; Corrêa, I. R.; Kindermann, M.; Beauflis, F.; Johnsson, K. An Engineered Protein Tag for Multiprotein Labeling in Living Cells. *Chem. Biol.* **2008**, *15* (2), 128–136. <https://doi.org/10.1016/j.chembiol.2008.01.007>.
- (81) Mizukami, S.; Watanabe, S.; Hori, Y.; Kikuchi, K. Covalent Protein Labeling Based on Noncatalytic  $\beta$ -Lactamase and a Designed FRET Substrate. *J. Am. Chem. Soc.* **2009**, *131* (14), 5016–5017. <https://doi.org/10.1021/ja8082285>.

- (82) Crivat, G.; Taraska, J. W. Imaging Proteins inside Cells with Fluorescent Tags. *Trends Biotechnol.* **2012**, *30* (1), 8–16. <https://doi.org/10.1016/j.tibtech.2011.08.002>.
- (83) Li, C.; Tebo, A. G.; Gautier, A. Fluorogenic Labeling Strategies for Biological Imaging. *Int. J. Mol. Sci.* **2017**, *18* (7). <https://doi.org/10.3390/ijms18071473>.
- (84) Dean, K. M.; Palmer, A. E. Advances in Fluorescence Labeling Strategies for Dynamic Cellular Imaging. *Nat. Chem. Biol.* **2014**, *10* (7), 512–523. <https://doi.org/10.1038/nchembio.1556>.
- (85) Zhang, C.; Welborn, M.; Zhu, T.; Yang, N. J.; Santos, M. S.; Van Voorhis, T.; Pentelute, B. L.  $\pi$ -Clamp-Mediated Cysteine Conjugation. *Nat. Chem.* **2016**, *8* (2), 120–128. <https://doi.org/10.1038/nchem.2413>.
- (86) Adams, S. R.; Campbell, R. E.; Gross, L. A.; Martin, B. R.; Walkup, G. K.; Yao, Y.; Llopis, J.; Tsien, R. Y. New Biarsenical Ligands and Tetracysteine Motifs for Protein Labeling in Vitro and in Vivo: Synthesis and Biological Applications. *J. Am. Chem. Soc.* **2002**, *124* (21), 6063–6076. <https://doi.org/10.1021/ja017687n>.
- (87) Griffin, B. A.; Adams, S. R.; Tsien, R. Y. Specific Covalent Labeling of Recombinant Protein Molecules inside Live Cells. *Science* (80-. ). **1998**, *281* (5374), 269–272. <https://doi.org/10.1126/science.281.5374.269>.
- (88) Nonaka, H.; Fujishima, S. hei; Uchinomiya, S. hei; Ojida, A.; Hamachi, I. FLAG-Tag Selective Covalent Protein Labeling via a Binding-Induced Acyl-Transfer Reaction. *Bioorganic Med. Chem. Lett.* **2009**, *19* (23), 6696–6699. <https://doi.org/10.1016/j.bmcl.2009.09.122>.
- (89) Halo, T. L.; Appelbaum, J.; Hobert, E. M.; Balkin, D. M.; Schepartz, A. Selective Recognition of Protein Tetraserine Motifs with a Cell-Permeable, pro-Fluorescent Bis-Boronic Acid. *J. Am. Chem. Soc.* **2009**, *131* (2), 438–439. <https://doi.org/10.1021/ja807872s>.
- (90) Keyser, S. G. L.; Utz, A.; Bertozzi, C. R. Computation-Guided Rational Design of a Peptide Motif That Reacts with Cyanobenzothiazoles via Internal Cysteine-Lysine Relay. *J. Org. Chem.* **2018**, *83* (14), 7467–7479. <https://doi.org/10.1021/acs.joc.8b00625>.
- (91) Hansen, B. K.; Loveridge, C. J.; Thyssen, S.; Wørmer, G. J.; Nielsen, A. D.; Palmfeldt, J.; Johannsen, M.; Poulsen, T. B. STEFs: Activated Vinylogous Protein-Reactive Electrophiles. *Angew. Chem., Int. Ed. Engl.* **2019**, *58* (11), 3533–3537. <https://doi.org/10.1002/anie.201814073>.
- (92) Willwacher, J.; Raj, R.; Mohammed, S.; Davis, B. G. Selective Metal-Site-Guided Arylation of Proteins. *J. Am. Chem. Soc.* **2016**, *138* (28), 8678–8681. <https://doi.org/10.1021/jacs.6b04043>.
- (93) Adusumalli, S. R.; Rawale, D. G.; Singh, U.; Tripathi, P.; Paul, R.; Kalra, N.; Mishra, R. K.; Shukla, S.; Rai, V. Single-Site Labeling of Native Proteins Enabled by a Chemoselective and Site-Selective Chemical Technology. *J. Am. Chem. Soc.* **2018**, *140* (44), 15114–15123. <https://doi.org/10.1021/jacs.8b10490>.
- (94) Adusumalli, S. R.; Rawale, D. G.; Thakur, K.; Purushottam, L.; Reddy, N. C.; Kalra, N.; Shukla, S.; Rai, V. Chemoselective and Site-Selective Lysine-Directed Lysine Modification Enables Single-Site Labeling of Native Proteins. *Angew. Chem.* **2020**, *132* (26), 10418–10422. <https://doi.org/10.1002/ange.202000062>.
- (95) Tsukiji, S.; Miyagawa, M.; Takaoka, Y.; Tamura, T.; Hamachi, I. Ligand-Directed Tosyl Chemistry for Protein Labeling in Vivo. *Nat. Chem. Biol.* **2009**, *5* (5), 341–343. <https://doi.org/10.1038/nchembio.157>.
- (96) Hughes, C. C.; Yang, Y. L.; Liu, W. T.; Dorrestein, P. C.; La Clair, J. J.; Fenical, W. Marinopyrrole A Target Elucidation by Acyl Dye Transfer. *J. Am. Chem. Soc.* **2009**, *131* (34), 12094–12096. <https://doi.org/10.1021/ja903149u>.
- (97) Kojima, H.; Fujita, Y.; Takeuchi, R.; Ikebe, Y.; Ohashi, N.; Yamamoto, K.; Itoh, T. Cyclization Reaction-Based Turn-on Probe for Covalent Labeling of Target Proteins. *Cell Chem. Biol.* **2020**, *27* (3), 334–349.e11. <https://doi.org/10.1016/j.chembiol.2020.01.006>.

- (98) Otsuki, S.; Nishimura, S.; Takabatake, H.; Nakajima, K.; Takasu, Y.; Yagura, T.; Sakai, Y.; Hattori, A.; Kakeya, H. Chemical Tagging of a Drug Target Using 5-Sulfonyl Tetrazole. *Bioorganic Med. Chem. Lett.* **2013**, *23* (6), 1608–1611. <https://doi.org/10.1016/j.bmcl.2013.01.092>.
- (99) Tamura, T.; Tsukiji, S.; Hamachi, I. Native FKBP12 Engineering by Ligand-Directed Tosyl Chemistry: Labeling Properties and Application to Photo-Cross-Linking of Protein Complexes in Vitro and in Living Cells. *J. Am. Chem. Soc.* **2012**, *134* (4), 2216–2226. <https://doi.org/10.1021/ja209641t>.
- (100) Masuya, T.; Murai, M.; Ifuku, K.; Morisaka, H.; Miyoshi, H. Site-Specific Chemical Labeling of Mitochondrial Respiratory Complex i through Ligand-Directed Tosylate Chemistry. *Biochemistry* **2014**, *53* (14), 2307–2317. <https://doi.org/10.1021/bi500205x>.
- (101) Tamura, T.; Kioi, Y.; Miki, T.; Tsukiji, S.; Hamachi, I. Fluorophore Labeling of Native FKBP12 by Ligand-Directed Tosyl Chemistry Allows Detection of Its Molecular Interactions in Vitro and in Living Cells. *J. Am. Chem. Soc.* **2013**, *135* (18), 6782–6785. <https://doi.org/10.1021/ja401956b>.
- (102) Takaoka, Y.; Sun, Y.; Tsukiji, S.; Hamachi, I. Mechanisms of Chemical Protein 19F-Labeling and NMR-Based Biosensor Construction in Vitro and in Cells Using Self-Assembling Ligand-Directed Tosylate Compounds. *Chem. Sci.* **2011**, *2* (3), 511–520. <https://doi.org/10.1039/c0sc00513d>.
- (103) Takaoka, Y.; Kioi, Y.; Morito, A.; Otani, J.; Arita, K.; Ashihara, E.; Ariyoshi, M.; Tochio, H.; Shirakawa, M.; Hamachi, I. Quantitative Comparison of Protein Dynamics in Live Cells and in Vitro by In-Cell 19F-NMR. *Chem. Commun.* **2013**, *49* (27), 2801–2803. <https://doi.org/10.1039/c3cc39205h>.
- (104) Fujishima, S. H.; Yasui, R.; Miki, T.; Ojida, A.; Hamachi, I. Ligand-Directed Acyl Imidazole Chemistry for Labeling of Membrane-Bound Proteins on Live Cells. *J. Am. Chem. Soc.* **2012**, *134* (9), 3961–3964. <https://doi.org/10.1021/ja2108855>.
- (105) Mortensen, M. R.; Nielsen, N. L.; Palmfeldt, J.; Gothelf, K. V. Imidazole Carbamate Probes for Affinity Guided Azide-Transfer to Metal-Binding Proteins. *Org. Biomol. Chem.* **2019**, *17* (6), 1379–1383. <https://doi.org/10.1039/C8OB03017K>.
- (106) Mino, T.; Sakamoto, S.; Hamachi, I. Recent Applications of N-Acyl Imidazole Chemistry in Chemical Biology. *Biosci. Biotechnol. Biochem.* **2021**, *85* (1), 53–60. <https://doi.org/10.1093/bbb/zbaa026>.
- (107) Matsuo, K.; Kioi, Y.; Yasui, R.; Takaoka, Y.; Miki, T.; Fujishima, S.; Hamachi, I. One-Step Construction of Caged Carbonic Anhydrase I Using a Ligand-Directed Acyl Imidazole-Based Protein Labeling Method. *Chem. Sci.* **2013**, *4*, 2573–2580. <https://doi.org/10.1039/c3sc50560j>.
- (108) Miki, T.; Fujishima, S.; Komatsu, K.; Kuwata, K.; Kiyonaka, S.; Hamachi, I. LDAO-Based Chemical Labeling of Intact Membrane Proteins and Its Pulse-Chase Analysis under Live Cell Conditions. *Chem. Biol.* **2014**, *21* (8), 1013–1022. <https://doi.org/10.1016/j.chembiol.2014.07.013>.
- (109) Wakayama, S.; Kiyonaka, S.; Arai, I.; Kakegawa, W.; Matsuda, S.; Iyata, K.; Nemoto, Y. L.; Kusumi, A.; Yuzaki, M.; Hamachi, I. Chemical Labelling for Visualizing Native AMPA Receptors in Live Neurons. *Nat. Commun.* **2017**, *8*. <https://doi.org/10.1038/ncomms14850>.
- (110) Tamura, T.; Ueda, T.; Goto, T.; Tsukidate, T.; Shapira, Y.; Nishikawa, Y.; Fujisawa, A.; Hamachi, I. Rapid Labelling and Covalent Inhibition of Intracellular Native Proteins Using Ligand-Directed N-Acyl-N-Alkyl Sulfonamide. *Nat. Commun.* **2018**, *9* (1), 1–12. <https://doi.org/10.1038/s41467-018-04343-0>.
- (111) Ueda, T.; Tamura, T.; Kawano, M.; Shiono, K.; Hobor, F.; Wilson, A. J.; Hamachi, I. Enhanced Suppression of a Protein-Protein Interaction in Cells Using Small-Molecule Covalent Inhibitors Based on an N-Acyl- N-Alkyl Sulfonamide Warhead. *J. Am. Chem. Soc.* **2021**, *143* (12), 4766–4774. <https://doi.org/10.1021/jacs.1c00703>.
- (112) Teng, M.; Jiang, J.; Ficarro, S. B.; Seo, H.-S.; Bae, J. H.; Donovan, K. A.; Fischer, E. S.; Zhang, T.; Dhe-Paganon, S.; Marto, J. A.; Gray, N. S. Exploring Ligand-Directed N -Acyl- N -Alkylsulfonamide-Based Acylation Chemistry for Potential Targeted Degradation Development. *ACS Med. Chem. Lett.* **2021**, *12* (8), 1302–1307. <https://doi.org/10.1021/acsmchemlett.1c00285>.

- (113) Thimaradka, V.; Hoon Oh, J.; Heroven, C.; Radu Aricescu, A.; Yuzaki, M.; Tamura, T.; Hamachi, I. Site-Specific Covalent Labeling of His-Tag Fused Proteins with N-Acyl-N-Alkyl Sulfonamide Reagent. *Bioorganic Med. Chem.* **2021**, *30*, 115947. <https://doi.org/10.1016/j.bmc.2020.115947>.
- (114) Denda, M.; Morisaki, T.; Kohiki, T.; Yamamoto, J.; Sato, K.; Sagawa, I.; Inokuma, T.; Sato, Y.; Yamauchi, A.; Shigenaga, A.; Otaka, A. Labelling of Endogenous Target Protein via N-S Acyl Transfer-Mediated Activation of N-Sulfanylethylamide. *Org. Biomol. Chem.* **2016**, *14* (26), 6244–6251. <https://doi.org/10.1039/c6ob01014h>.
- (115) Shigenaga, A. Theoretical Study on Reaction Mechanism of Phosphate-Catalysed N-S Acyl Transfer of: N -Sulfanylethylamide (SEAlide). *Org. Biomol. Chem.* **2020**, *18* (47), 9706–9711. <https://doi.org/10.1039/d0ob01968b>.
- (116) Kohiki, T.; Kato, Y.; Nishikawa, Y.; Yorita, K.; Sagawa, I.; Denda, M.; Inokuma, T.; Shigenaga, A.; Fukui, K.; Otaka, A. Elucidation of Inhibitor-Binding Pockets of D-Amino Acid Oxidase Using Docking Simulation and N-Sulfanylethylamide-Based Labeling Technology. *Org. Biomol. Chem.* **2017**, *15* (25), 5289–5297. <https://doi.org/10.1039/c7ob00633k>.
- (117) Takaoka, Y.; Nishikawa, Y.; Hashimoto, Y.; Sasaki, K.; Hamachi, I. Ligand-Directed Dibromophenyl Benzoate Chemistry for Rapid and Selective Acylation of Intracellular Natural Proteins. *Chem. Sci.* **2015**, *6* (5), 3217–3224. <https://doi.org/10.1039/c5sc00190k>.
- (118) Yamaguchi, T.; Asanuma, M.; Nakanishi, S.; Saito, Y.; Okazaki, M.; Dodo, K.; Sodeoka, M. Turn-ON Fluorescent Affinity Labeling Using a Small Bifunctional o-Nitrobenzoxadiazole Unit. *Chem. Sci.* **2014**, *5*, 1021–1029. <https://doi.org/10.1039/c3sc52704b>.
- (119) Xue, D.; Ye, L.; Zheng, J.; Wu, Y.; Zhang, X.; Xu, Y.; Li, T.; Stevens, R. C.; Xu, F.; Zhuang, M.; Zhao, S.; Zhao, F.; Tao, H. The Structure-Based Traceless Specific Fluorescence Labeling of the Smoothed Receptor. *Org. Biomol. Chem.* **2019**, *17* (25), 6136–6142. <https://doi.org/10.1039/c9ob00654k>.
- (120) Koshi, Y.; Nakata, E.; Miyagawa, M.; Tsukiji, S.; Ogawa, T.; Hamachi, I. Target-Specific Chemical Acylation of Lectins by Ligand-Tethered DMAP Catalysts. **2008**, No. c, 2796–2801.
- (121) Koshi, Y.; Nakata, E.; Miyagawa, M.; Tsukiji, S.; Ogawa, T.; Hamachi, I. Target-Specific Chemical Acylation of Lectins by Ligand-Tethered DMAP Catalysts. *J. Am. Chem. Soc.* **2008**, *130* (1), 245–251. <https://doi.org/10.1021/ja075684q>.
- (122) Wang, H.; Koshi, Y.; Minato, D.; Nonaka, H.; Kiyonaka, S.; Mori, Y.; Tsukiji, S.; Hamachi, I. Chemical Cell-Surface Receptor Engineering Using Affinity-Guided, Multivalent Organocatalysts. *J. Am. Chem. Soc.* **2011**, *133* (31), 12220–12228. <https://doi.org/10.1021/ja204422r>.
- (123) Hayashi, T.; Yasueda, Y.; Tamura, T.; Takaoka, Y.; Hamachi, I. Analysis of Cell-Surface Receptor Dynamics Through Covalent Labeling by Catalyst-Tethered Antibody. *J. Am. Chem. Soc.* **2015**, *137* (16), 5372–5380. <https://doi.org/10.1021/jacs.5b02867>.
- (124) Amamoto, Y.; Aoi, Y.; Nagashima, N.; Suto, H.; Yoshidome, D.; Arimura, Y.; Osakabe, A.; Kato, D.; Kurumizaka, H.; Kawashima, S. A.; Yamatsugu, K.; Kanai, M. Synthetic Posttranslational Modifications: Chemical Catalyst-Driven Regioselective Histone Acylation of Native Chromatin. *J. Am. Chem. Soc.* **2017**, *139* (22), 7568–7576. <https://doi.org/10.1021/jacs.7b02138>.
- (125) Ishiguro, T.; Amamoto, Y.; Tanabe, K.; Liu, J.; Kajino, H.; Fujimura, A.; Aoi, Y.; Osakabe, A.; Horikoshi, N.; Kurumizaka, H.; Yamatsugu, K.; Kawashima, S. A.; Kanai, M. Synthetic Chromatin Acylation by an Artificial Catalyst System. *Chem* **2017**, *2* (6), 840–859. <https://doi.org/10.1016/j.chempr.2017.04.002>.
- (126) Chen, Z.; Vohidov, F.; Coughlin, J. M.; Stagg, L. J.; Arold, S. T.; Ladbury, J. E.; Ball, Z. T. Catalytic Protein Modification with Dirhodium Metallopeptides: Specificity in Designed and Natural Systems. *J. Am. Chem. Soc.* **2012**, *134* (24), 10138–10145. <https://doi.org/10.1021/ja302284p>.
- (127) Vohidov, F.; Coughlin, J. M.; Ball, Z. T. Rhodium(II) Metallopeptide Catalyst Design Enables Fine Control in Selective Functionalization of Natural SH3 Domains. *Angew. Chem.* **2015**, *127* (15), 4670–

4674. <https://doi.org/10.1002/ange.201411745>.
- (128) Ohata, J.; Ball, Z. T. A Hexa-Rhodium Metallopeptide Catalyst for Site-Specific Functionalization of Natural Antibodies. *J. Am. Chem. Soc.* **2017**, *139* (36), 12617–12622. <https://doi.org/10.1021/jacs.7b06428>.
- (129) Sato, S.; Hatano, K.; Tsushima, M.; Nakamura, H. 1-Methyl-4-Aryl-Urazole (MAUra) Labels Tyrosine in Proximity to Ruthenium Photocatalysts. *Chem. Commun.* **2018**, *54* (46), 5871–5874. <https://doi.org/10.1039/c8cc02891e>.
- (130) Sato, S.; Nakamura, H. Ligand-Directed Selective Protein Modification Based on Local Single-Electron-Transfer Catalysis. *Angew. Chem.* **2013**, *125* (33), 8843–8846. <https://doi.org/10.1002/ange.201303831>.
- (131) Sato, S.; Morita, K.; Nakamura, H. Regulation of Target Protein Knockdown and Labeling Using Ligand-Directed Ru(Bpy)<sub>3</sub> Photocatalyst. *Bioconjug. Chem.* **2015**, *26* (2), 250–256. <https://doi.org/10.1021/bc500518t>.
- (132) Tamura, T.; Song, Z.; Amaike, K.; Lee, S.; Yin, S.; Kiyonaka, S.; Hamachi, I. Affinity-Guided Oxime Chemistry for Selective Protein Acylation in Live Tissue Systems. *J. Am. Chem. Soc.* **2017**, *139* (40), 14181–14191. <https://doi.org/10.1021/jacs.7b07339>.
- (133) British Heart Foundation. Global Heart & Circulatory Diseases. *Br. Hear. Found.* **2022**, No. January, 1–12.
- (134) NHS. Heart attack <https://www.nhs.uk/conditions/heart-attack/>.
- (135) Benjamin, E. J.; Muntner, P.; Alonso, A.; Bittencourt, M. S.; Callaway, C. W.; Carson, A. P.; Chamberlain, A. M.; Chang, A. R.; Cheng, S.; Das, S. R.; Delling, F. N.; Djousse, L.; Elkind, M. S. V.; Ferguson, J. F.; Fornage, M.; Jordan, L. C.; Khan, S. S.; Kissela, B. M.; Knutson, K. L.; Kwan, T. W.; Lackland, D. T.; Lewis, T. T.; Lichtman, J. H.; Longenecker, C. T.; Loop, M. S.; Lutsey, P. L.; Martin, S. S.; Matsushita, K.; Moran, A. E.; Mussolino, M. E.; O'Flaherty, M.; Pandey, A.; Perak, A. M.; Rosamond, W. D.; Roth, G. A.; Sampson, U. K. A.; Satou, G. M.; Schroeder, E. B.; Shah, S. H.; Spartano, N. L.; Stokes, A.; Tirschwell, D. L.; Tsao, C. W.; Turakhia, M. P.; VanWagner, L. B.; Wilkins, J. T.; Wong, S. S.; Virani, S. S. *Heart Disease and Stroke Statistics-2019 Update: A Report From the American Heart Association*; 2019; Vol. 139. <https://doi.org/10.1161/CIR.0000000000000659>.
- (136) Novosel, E. C.; Kleinhans, C.; Kluger, P. J. Vascularization Is the Key Challenge in Tissue Engineering. *Adv. Drug Deliv. Rev.* **2011**, *63* (4), 300–311. <https://doi.org/10.1016/j.addr.2011.03.004>.
- (137) Schumacher, B.; Peecher, P.; Von Specht, B. U.; Stegmann, T. Induction of Neoangiogenesis in Ischemic Myocardium by Human Growth Factors: First Clinical Results of a New Treatment of Coronary Heart Disease. *Circulation* **1998**, *97* (7), 645–650. <https://doi.org/10.1161/01.CIR.97.7.645>.
- (138) Javerzat, S.; Auguste, P.; Bikfalvi, A. The Role of Fibroblast Growth Factors in Vascular Development. *Mol. Med.* **2002**, *8* (10), 483–489. [https://doi.org/10.1016/S1471-4914\(02\)02394-8](https://doi.org/10.1016/S1471-4914(02)02394-8).
- (139) Gospodarowicz, D. Localisation of a Fibroblast Growth Factor and Its Effect Alone and with Hydrocortisone on 3T3 Cell Growth. *Nature* **1974**, *249*, 123–127. <https://doi.org/10.1038/249123a0>.
- (140) Yun, Y.-R.; Won, J. E.; Jeon, E.; Lee, S.; Kang, W.; Jo, H.; Jang, J.-H.; Shin, U. S.; Im, H.-W. Fibroblast Growth Factors: Biology, Function, and Application for Tissue Regeneration. *J Tissue Eng* **2010**, *210*. <https://doi.org/10.4061/2010/218142>.
- (141) Abraham, J. A.; Whang, J. L.; Tumolo, A.; Mergia, A.; Friedman, J.; Gospodarowicz, D.; Fiddes, J. C. Human Basic Fibroblast Growth Factor: Nucleotide Sequence and Genomic Organization. *EMBO J.* **1986**, *5* (10), 2523–2528. <https://doi.org/10.1002/j.1460-2075.1986.tb04530.x>.
- (142) Dionne, C. A.; Crumley, G.; Bellot, F.; Kaplow, J. M.; Searfoss, G.; Ruta, M.; Burgess, W. H.; Jaye, M.; Schlessinger, J. Cloning and Expression of Two Distinct High-Affinity Receptors Cross-Reacting with

- Acidic and Basic Fibroblast Growth Factors. *EMBO J.* **1990**, *9* (9), 2685–2692. <https://doi.org/10.1002/j.1460-2075.1990.tb07454.x>.
- (143) Burgess, W. H.; Maciag, T. The Heparin-Binding (Fibroblast) Growth Factor Family of Proteins. *Annu. Rev. Biochem.* **1989**, *58*, 575–602. <https://doi.org/10.1146/annurev.bi.58.070189.003043>.
- (144) <https://cellculturedish.com/basic-fibroblast-growth-factor-good-bad-stable/>.
- (145) Florkiewicz, R. Z.; Shibata, F.; Barankiewicz, T.; Baird, A.; Gonzalez, A.-M.; Florkiewicz, E.; Shah, N. Basic Fibroblast Growth Factor Gene Expression. *Ann. N. Y. Acad. Sci.* **1991**, *638* (1), 109–126. <https://doi.org/10.1111/j.1749-6632.1991.tb49022.x>.
- (146) Bikfalvi, A.; Klein, S.; Pintucci, G.; Rifkin, D. B. Biological Roles of Fibroblast Growth Factor-2. *Endocr. Rev.* **1997**, *18* (1), 26–45. <https://doi.org/10.1210/edrv.18.1.0292>.
- (147) Thompson, S. A. The Disulfide Structure of Bovine Pituitary Basic Fibroblast Growth. *J. Biol. Chem.* **1992**, *267* (4), 2269–2273. [https://doi.org/10.1016/S0021-9258\(18\)45873-5](https://doi.org/10.1016/S0021-9258(18)45873-5).
- (148) Kastrop, J. S.; Eriksson, S.; Dalbtge, H.; Flodgaard, H. X-Ray Structure of the 154-Amino-Acid Form of Recombinant Human Basic Fibroblast Growth Factor . Comparison with the Truncated 146-Amino-Acid Form. **1997**, 160–168. <https://doi.org/10.1107/S0907444996012711>.
- (149) Eriksson, A. E.; Cousens, L. S.; Weaver, L. H.; Matthews, B. W. Three-Dimensional Structure of Human Basic Fibroblast Growth Factor. **1991**, *88* (April), 3441–3445. <https://doi.org/10.1073/pnas.88.8.3441>.
- (150) Zhang, J.; Cousens, L. S.; Barr, P. J.; Sprang, S. R. Three-Dimensional Structure of Human Basic Fibroblast Growth Factor, a Structural Homolog of Interleukin 1 $\beta$ . *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88* (8), 3446–3450. <https://doi.org/10.1073/pnas.88.8.3446>.
- (151) Nunes, Q. M.; Li, Y.; Sun, C.; Kinnunen, T. K.; Fernig, D. G. Fibroblast Growth Factors as Tissue Repair and Regeneration Therapeutics. **2016**, 1–31. <https://doi.org/10.7717/peerj.1535>.
- (152) Stegmann, T. J. Heart Disease Induction of Neovascularisation by Growth Factors. *BioDrugs* **1999**, *11* (5), 301–308. <https://doi.org/10.2165/00063030-199911050-00002>.
- (153) Li, M.; Mondrinos, M. J.; Chen, X.; Gandhi, M. R.; Ko, F. K.; Lelkes, P. I. Heparin-Immobilized Biodegradable Scaffolds for Local and Sustained Release of Angiogenic Growth Factor. *J. Biomed. Mater. Res. Part A* **2006**, *79* (4), 963–973. <https://doi.org/10.1002/jbm.a>.
- (154) Liu, X.; Jiang, N.; Tan, C.; Tang, M. The Clinical Study of Recombinant Human Basic Fibroblast Growth Factor in Stimulating the Wound Surface Healing. *Acta Acad.* **2005**.
- (155) Akita, S.; Akino, K.; Imaizumi, T.; Hirano, A. Basic Fibroblast Growth Factor Accelerates and Improves Second-Degree Burn Wound Healing. *Wound Repair Regen.* **2008**, *16* (5), 635–641. <https://doi.org/10.1111/j.1524-475X.2008.00414.x>.
- (156) Uchi, H.; Igarashi, A.; Urabe, K.; Koga, T.; Nakayama, J.; Kawamori, R.; Tamaki, K.; Hirakta, H.; Ohura, T.; Furue, M. Clinical Efficacy of Basic Fibroblast Growth Factor (BFGF) for Diabetic Ulcer. *Eur. J. Dermatology* **2009**, *19* (5), 461–468. <https://doi.org/10.1684/ejd.2009.0750>.
- (157) Jiang, X. W.; Zhang, Y.; Zhang, H.; Lu, K.; Yang, S. K.; Sun, G. L. Double-Blind, Randomized, Controlled Clinical Trial of the Effects of Diosmectite and Basic Fibroblast Growth Factor Paste on the Treatment of Minor Recurrent Aphthous Stomatitis. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **2013**, *116* (5), 570–575. <https://doi.org/10.1016/j.oooo.2013.07.003>.
- (158) Robson, M. C.; Phillips, L. G.; Lawrence, W. T.; Bishop, J. B.; Youngerman, J. S.; Hayward, P. G.; Broemeling, L. D.; Heggers, J. P.; Mannick, J. A.; Cohen, I. K.; Friedmann, P.; Kaplan, E. L.; Jurkiewicz, M. J.; Trunkey, D.; Silen, W. The Safety and Effect of Topically Applied Recombinant Basic Fibroblast Growth Factor on the Healing of Chronic Pressure Sores. *Ann. Surg.* **1992**, *216* (4), 401–408. <https://doi.org/10.1097/0000658-199210000-00002>.
- (159) Kumagai, M.; Marui, A.; Tabata, Y.; Takeda, T.; Yamamoto, M.; Yonezawa, A.; Tanaka, S.; Yanagi, S.;

- Ito-Ihara, T.; Ikeda, T.; Murayama, T.; Teramukai, S.; Katsura, T.; Matsubara, K.; Kawakami, K.; Yokode, M.; Shimizu, A.; Sakata, R. Safety and Efficacy of Sustained Release of Basic Fibroblast Growth Factor Using Gelatin Hydrogel in Patients with Critical Limb Ischemia. *Heart Vessels* **2016**, *31* (5), 713–721. <https://doi.org/10.1007/s00380-015-0677-x>.
- (160) Bennett, J.; Rubin, J. P. FGF-2 Enhances Vascularization for Adipose. 1153–1164. <https://doi.org/10.1097/01.prs.0000305517.93747.72>.
- (161) Zieris, A.; Prokoph, S.; Levental, K. R.; Welzel, P. B.; Grimmer, M.; Freudenberg, U.; Werner, C. FGF-2 and VEGF Functionalization of StarPEG-Heparin Hydrogels to Modulate Biomolecular and Physical Cues of Angiogenesis. *Biomaterials* **2010**, *31* (31), 7985–7994. <https://doi.org/10.1016/j.biomaterials.2010.07.021>.
- (162) Perets, A.; Baruch, Y.; Weisbuch, F.; Shoshany, G.; Neufeld, G.; Cohen, S. Enhancing the Vascularization of Three-Dimensional Porous Alginate Scaffolds by Incorporating Controlled Release Basic Fibroblast Growth Factor Microspheres. *J. Biomed. Mater. Res. - Part A* **2003**, *65* (4), 489–497. <https://doi.org/10.1002/jbm.a.10542>.
- (163) Ono, I.; Akasaka, Y.; Kikuchi, R.; Sakemoto, A.; Kamiya, T.; Yamashita, T.; Jimbow, K. Basic Fibroblast Growth Factor Reduces Scar Formation in Acute Incisional Wounds. *Wound Repair Regen.* **2007**, *15* (5), 617–623. <https://doi.org/10.1111/j.1524-475X.2007.00293.x>.
- (164) Abdelhakim, M.; Lin, X.; Ogawa, R. The Japanese Experience with Basic Fibroblast Growth Factor in Cutaneous Wound Management and Scar Prevention: A Systematic Review of Clinical and Biological Aspects. *Dermatol. Ther. (Heidelb)*. **2020**, *10* (4), 569–587. <https://doi.org/10.1007/s13555-020-00407-6>.
- (165) Kawaguchi, H.; Oka, H.; Jingushi, S.; Izumi, T.; Fukunaga, M.; Sato, K.; Matsushita, T.; Nakamura, K. A Local Application of Recombinant Human Fibroblast Growth Factor 2 for Tibial Shaft Fractures: A Randomized, Placebo-Controlled Trial. *J. Bone Miner. Res.* **2010**, *25* (12), 2735–2743. <https://doi.org/10.1002/jbmr.146>.
- (166) Murakami, S. Periodontal Tissue Regeneration by Signaling Molecule(s): What Role Does Basic Fibroblast Growth Factor (FGF-2) Have in Periodontal Therapy? *Periodontol.* **2000** **2011**, *56* (1), 188–208. <https://doi.org/10.1111/j.1600-0757.2010.00365.x>.
- (167) Murakami, S.; Takayama, S.; Ikezawa, K.; Sltimabukuro, Y.; Kitamura, M.; Nozaki, T.; Terashima, A.; Asano, T.; Okada, H. Regeneration of Periodontal Tissues by Basic Fibroblast Growth Factor. *J. Periodont Res* **1999**, *34*, 425–430. <https://doi.org/10.1111/j.1600-0765.1999.tb02277.x>.
- (168) Hong, K. S.; Kim, E. C.; Bang, S. H.; Chung, C. H.; Lee, Y. Il; Hyun, J. K.; Lee, H. H.; Jang, J. H.; Kim, T. Il; Kim, H. W. Bone Regeneration by Bioactive Hybrid Membrane Containing FGF2 within Rat Calvarium. *J. Biomed. Mater. Res. - Part A* **2010**, *94* (4), 1187–1194. <https://doi.org/10.1002/jbm.a.32799>.
- (169) Maehara, H.; Sotome, S.; Yoshii, T.; Torigoe, I.; Kawasaki, Y.; Sugata, Y.; Yuasa, M.; Hirano, M.; Mochizuki, N.; Kikuchi, M.; Shinomiya, K.; Okawa, A. Repair of Large Osteochondral Defects in Rabbits Using Porous Hydroxyapatite/Collagen (HAp/Col) and Fibroblast Growth Factor-2 (FGF-2). *J. Orthop. Res.* **2010**, *28* (5), 677–686. <https://doi.org/10.1002/jor.21032>.
- (170) Wu, X.; Yan, Q.; Huang, Y.; Huang, H.; Su, Z.; Xiao, J.; Zeng, Y.; Wang, Y.; Nie, C.; Yang, Y.; Li, X. Isolation of a Novel Basic FGF-Binding Peptide with Potent Antiangiogenic Activity. *J. Cell. Mol. Med.* **2010**, *14* (1–2), 351–356. <https://doi.org/10.1111/j.1582-4934.2008.00506.x>.
- (171) Yayon, A.; Aviezer, D.; Safran, M.; Gross, J. L.; Heldman, Y.; Cabilly, S.; Givol, D.; Katchalski-Katzir, E. Isolation of Peptides That Inhibit Binding of Basic Fibroblast Growth Factor to Its Receptor from a Random Phage-Epitope Library. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90* (22), 10643–10647. <https://doi.org/10.1073/pnas.90.22.10643>.
- (172) Segal, D. H.; Germano, I. M.; Bederson, J. B. Effects of Basic Fibroblast Growth Factor on in Vivo Cerebral Tumorigenesis in Rats. *Neurosurgery* **1997**, *40* (5), 1027–1033.

- <https://doi.org/10.1097/00006123-199705000-00028>.
- (173) Vecchio, I.; Tornali, C.; Bragazzi, N. L.; Martini, M. The Discovery of Insulin : An Important Milestone in the History of Medicine. *Front. Endocrinol. (Lausanne)*. **2018**, *9* (October), 1–8. <https://doi.org/10.3389/fendo.2018.00613>.
- (174) Banting, F. G.; Best, C. H.; Collip, J. B.; Campbell, W. R.; Fletcher, A. A.; Noble, E. C. The Effect Produced on Diabetes By Extracts of Pancreas. *Trans Assoc Am Physicians* **1922**, *37*, 337–347.
- (175) Ward, C. W.; Lawrence, M. C. Landmarks in Insulin Research. *Front. Endocrinol. (Lausanne)*. **2011**, *2* (NOV), 1–11. <https://doi.org/10.3389/fendo.2011.00076>.
- (176) Sanger, F.; Tuppy, H. The Amino-Acid Sequence in the Phenylalanyl Chain of Insulin. 2. The Investigation of Peptides from Enzymic Hydrolysates. *Biochem. J.* **1951**, *49* (4), 481–490. <https://doi.org/10.1042/bj0490481>.
- (177) Sanger, F.; Thompson, E. O. The Amino-Acid Sequence in the Glycyl Chain of Insulin. II. The Investigation of Peptides from Enzymic Hydrolysates. *Biochem. J.* **1953**, *53* (3), 366–374. <https://doi.org/10.1042/bj0530366>.
- (178) Sanger, F.; Thompson, E. O. The Amino-Acid Sequence in the Glycyl Chain of Insulin. I. The Identification of Lower Peptides from Partial Hydrolysates. *Biochem. J.* **1953**, *53* (3), 353–366. <https://doi.org/10.1042/bj0530353>.
- (179) Sanger, F.; Tuppy, H. The Amino-Acid Sequence in the Phenylalanyl Chain of Insulin 1. *Biochem. J.* **1951**, *49*, 463–481.
- (180) Dodson, G. G. Dorothy Hodgkin, Protein Crystallography and Insulin. *Curr. Sci.* **1997**, *72* (7), 466–468.
- (181) Menting, J. G.; Whittaker, J.; Margetts, M. B.; Whittaker, L. J.; Kong, G. K.-W.; Smith, B. J.; Watson, C. J.; Zakova, L.; Kletvikova, E.; Jiracek, J.; Chan, S. J.; Steiner, D. F.; Dodson, G. G.; Brzozowski, A. M.; Weiss, M. A.; Ward, C. W.; Lawrence, M. C. How Insulin Engages Its Primary Binding Site on the Insulin Receptor. *Nature* **2013**, *492*, 241–245. <https://doi.org/10.1038/nature11781>.
- (182) Gutmann, T.; Kim, K. H.; Grzybek, M.; Walz, T.; Coskun, Ü. Visualization of Ligand-Induced Transmembrane Signaling in the Full-Length Human Insulin Receptor. *J. Cell Biol.* **2018**, *217* (5), 1643–1649. <https://doi.org/10.1083/jcb.201711047>.
- (183) Rostène, W.; De Meyts, P. Insulin: A 100-Year-Old Discovery with a Fascinating History. *Endocr. Rev.* **2021**, *42* (5), 503–527. <https://doi.org/10.1210/endrev/bnab020>.
- (184) Andreas, K.; Zehbe, R.; Kazubek, M.; Grzeschik, K.; Sternberg, N.; Bäumlner, H.; Schubert, H.; Sittinger, M.; Ringe, J. Biodegradable Insulin-Loaded PLGA Microspheres Fabricated by Three Different Emulsification Techniques: Investigation for Cartilage Tissue Engineering. *Acta Biomater.* **2011**, *7* (4), 1485–1495. <https://doi.org/10.1016/j.actbio.2010.12.014>.
- (185) Ramos, D. M.; Abdulmalik, S.; Arul, M. R.; Rudraiah, S.; Laurencin, C. T.; Mazzocca, A. D.; Kumbar, S. G. Insulin Immobilized PCL-Cellulose Acetate Micro-Nanostructured Fibrous Scaffolds for Tendon Tissue Engineering. *Polym. Adv. Technol.* **2019**, *30* (5), 1205–1215. <https://doi.org/10.1002/pat.4553>.
- (186) Knutson, V. P. Insulin-Binding Peptide. *J. Biol. Chem.* **1988**, *263* (28), 14146–14151. [https://doi.org/10.1016/S0021-9258\(18\)68197-9](https://doi.org/10.1016/S0021-9258(18)68197-9).
- (187) Yu, H. Q.; Dong, X. Y.; Sun, Y. Affinity Chromatography of Insulin with a Heptapeptide Ligand Selected from Phage Display Library. *Chromatographia* **2004**, *60* (7–8), 379–383. <https://doi.org/10.1365/s10337-004-0407-5>.
- (188) Chiang, H. L.; Ngo, S. T.; Chen, C. J.; Hu, C. K.; Li, M. S. Oligomerization of Peptides LVEALYL and RGFYYT and Their Binding Affinity to Insulin. *PLoS One* **2013**, *8* (6). <https://doi.org/10.1371/journal.pone.0065358>.

## Chapter 2 – Peptides, the core of our binding proximity approach: synthesis and modification

The core of this PhD approach is the design and synthesis of peptides which have protein binding affinity and then to modify these peptides into ligand-directed probes for functionalisation of the protein. To do so the peptides needed to be synthesised, their binding affinity studied, and then functionalised with the handles needed for ligand-directed chemistry. This chapter will first give a background on peptide chemistry and explain the optimisation achieved over the PhD for the functionalisation of the peptides used, with their functions then being discussed in more detail in Chapters 3, and 4, and their use for protein modification discussed in Chapter 6.

### 2.1 General introduction of SPPS technique

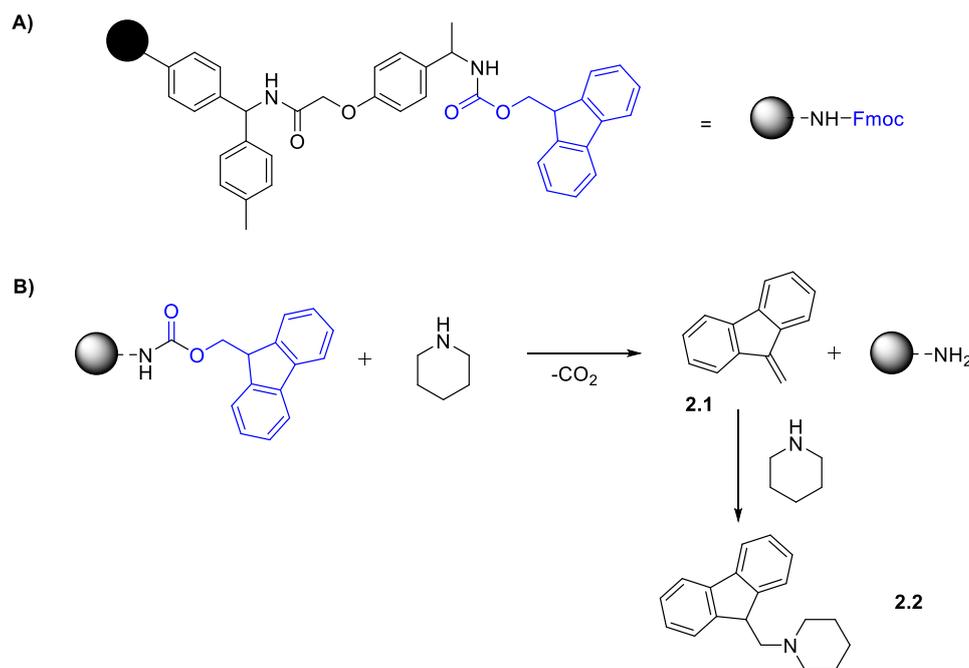
All the peptides used in this PhD thesis were synthesized in-house using a **Solid Phase Peptide Synthesis (SPPS)** approach using a CEM Liberty Lite automated microwave peptide synthesizer (CEM Corporation). This first paragraph describes the background of SPPS and the general workflow used throughout the synthesis process within this thesis.

SPPS was developed by Bruce Merrifield who was awarded the Nobel Prize for it in 1984.<sup>1</sup> It was a revolutionary change for chemists, and biologists over the world to be able to obtain larger quantities of peptides and with higher purity in an automated process, thanks to the anchoring of the C-terminal amino acid of the peptide on an insoluble polystyrene resin support.<sup>2</sup> The drastic change was on the purification aspect, compared to in solution synthesis, where while the peptide was grown on resin all the wash steps were achieved in a straight forward manner without need of purification. Moreover, due to the later-on automation, the time of a synthesis was drastically reduced and that allowed far longer sequences to be achieved compared to in solution synthesis. Despite the need to use reagents in excess to drive reactions to completion, SPPS has become the go-to technique for peptide synthesis. SPPS was then greatly improved thanks to the introduction of microwave-reactions accelerating drastically the process. The first fully microwave automated peptide synthesizer was put on the market in 2003 by CEM.<sup>3</sup> While the field of peptide chemistry is very broad, depending on, for example the choice of protecting groups or the nature of the peptide C-terminus,<sup>1-4</sup> only the chemistry used during this PhD is described in more detail in the coming paragraphs.

The peptide synthesiser used during this thesis, the Liberty Lite by CEM, uses Fmoc-chemistry, meaning that all the N-termini of the amino acids are protected by a Fmoc group. This protection is critical to ensure the formation of the correct amide bond, ensuring only a single coupling at each amino acid coupling cycle. While a few non-polar amino acids such as Gly, Ala, and Leu, require only

protection at their *N*-termini, the majority of amino acids are trifunctional meaning that their side-chain also requires a protecting group to protect functional groups such as carboxyls (Asp, Glu), amines (Lys), thiols (Cys), hydroxyls (Ser, Thr, Tyr), imidazoles (His), and indoles (Trp). These side-chain protecting groups must be stable to the basic conditions used for Fmoc deprotection, but must be easily removed at the end of the synthesis. All the groups used for side-chain protection for classic Fmoc synthesis used in this PhD (except when specified) all shared the characteristic of being acid sensitive and are generally being deprotected at the end of the synthesis via an acid-based cleavage cocktail which will be described in more detail later.

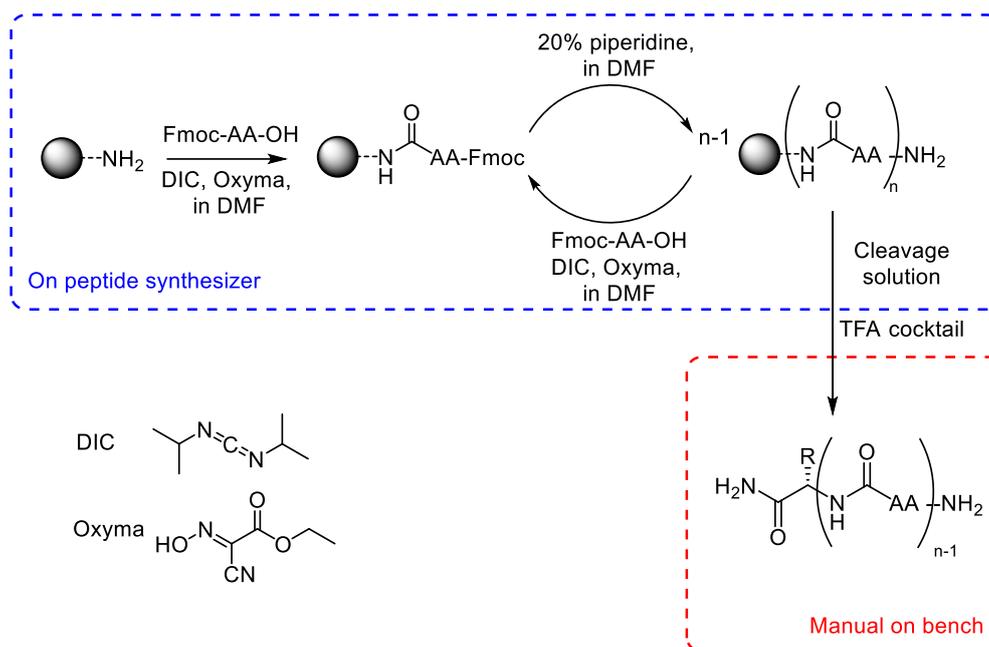
During peptide synthesis, the first step consists of deprotecting the Fmoc-group present on the resin-support to provide a free amine on which the first amino acid will be coupled. Rink Amide MBHA resin was used throughout the thesis, leading thus to peptide with amide *C*-termini. To initiate Fmoc deprotection, a solution of 20% piperidine in dimethylformamide (DMF) is used liberating carbon dioxide and 9-methylene-fluorene **2.1** as shown in Scheme 1. Side product **2.1** can also react with piperidine giving a stabilised compound **2.2**.



Scheme 1: A) Structure of the rink-amide resin, B) Scheme of the deprotection of Rink Amide resin using piperidine to liberate a free amine on the resin and two side products **2.1** and **2.2**.

Following this step, an iterative cycle of adding amino acids and deprotecting their Fmoc-protected amine is initiated. For the coupling of amino acids, *N,N'*-diisopropylcarbodiimide (DIC) is used as an activator of the *C*-terminus of an Fmoc-protected amino acid, and Oxyma base is used as an extra coupling reagent. Oxyma not only is used as a base ensuring the change of pH but is also able to activate the newly formed bond between DIC and the *C*-carboxylic terminus of the amino acid

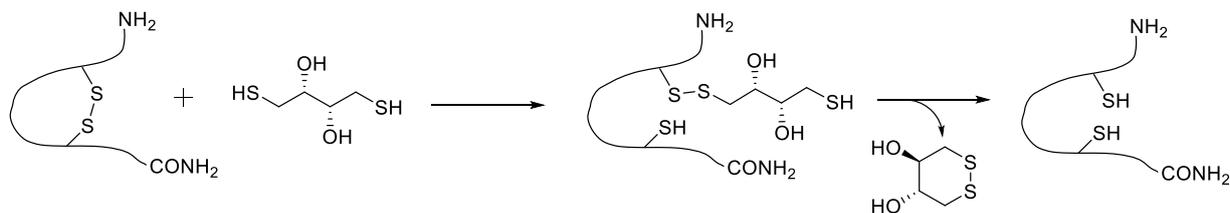
enhancing its reactivity with an amine. Following the coupling, excess reagents are washed away from the solid-supported peptide and the Fmoc-group is deprotected by a 20% piperidine solution, as in the initial deprotection step, to generate a new *N*-terminal amine, while keeping the side-chain protecting groups on. The liberated amine can then be coupled with a second Fmoc-protected amino acid in an iterative fashion, Scheme 2. Once the last amino acid has been added, it is possible to leave the last Fmoc-group on the *N*-terminus or not, depending on the application of the peptide, which will be explained in more detail in a later section.



Scheme 2: General scheme of the synthesis of peptides with iterative coupling via SPPS with Fmoc protected amino acids and coupling agents DIC and Oxyma in DMF. Final manual deprotection and cleavage step in TFA solution.

After the peptide synthesis is complete, the next step consists of cleaving the peptide from the resin and deprotecting the side chain protecting groups. To do that a cleavage mixture is used. While once again a range of possibilities exist in literature the mixture used throughout the PhD was the following: 90% trifluoroacetic acid (TFA), 5%  $\text{H}_2\text{O}$ , 3% triisopropyl silane (TIPS), with 2% in mass dithiothreitol (DTT) when the sequence contained a cysteine moiety. TFA is a strong acid, able to cleave and deprotect the acid-sensitive groups used to protect the peptide sidechains such as *tert*-butyloxycarbonyl (Boc), triphenylmethyl (Trt), and *tert*-butyl (tBu) esters and ethers. TIPS is used as a scavenger able to react with the cleaved protecting group avoiding them to react with the deprotected amino-acid side chains and go back onto the peptide. DTT is used as a reducing agent for cysteine

containing peptides. It ensures the cysteine on the peptide is in a reduced state, as shown in Scheme 3; once oxidized, DTT forms a stable six-membered ring with an internal disulfide bond.



Scheme 3: Scheme for the reduction of disulfide bridge in presence of DTT.

After the cleavage deprotection step, run for a time that is sequence dependent, typically between 4 to 6 h, the acid mixture is added to  $-20\text{ }^{\circ}\text{C}$  diethyl ether ( $\text{Et}_2\text{O}$ ) which results in the precipitation of the fully deprotected peptide. The full experimental conditions of the coupling on the peptide synthesizer and the cleavage on bench are fully described in the Experimental section at the end of this chapter. After cleavage, the peptides were analysed by LCMS and purified if necessary, with data presented throughout this chapter using the general format shown in Figure 1, with the LCMS UV trace (wavelength always specified in caption as peptide dependant) of the peptide and the mass spectrum with observed and theoretical mass. Every peptide synthesised, including those kept on resin, were stored until further use in the freezer.

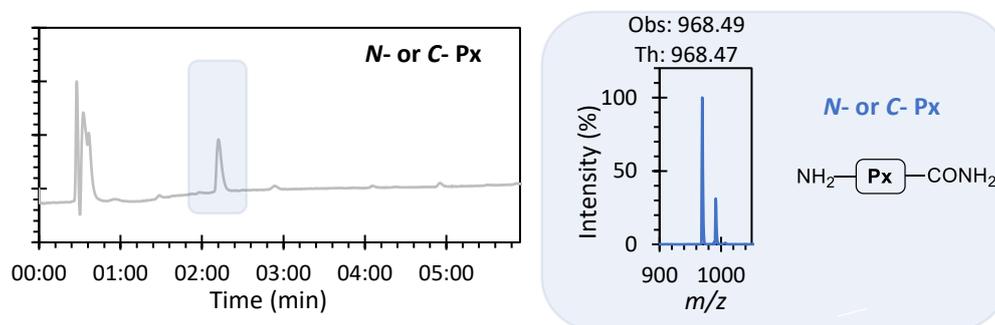


Figure 1: Generic layout of LC-MS data to be used throughout this thesis. On the left-hand side the LC UV-trace at 220 nm, plotted against time, and the compound MS spectrum extracted from the boxed area in the second graph with theoretical mass (*Th*) and observed mass (*Obs*) compared.

## 2.2 General introduction of the targeted peptides

Throughout the next sections functionalisation of the peptides used at both the *N*- and *C*-termini will be discussed. Indeed, once the core peptides were synthesised via SPPS, different functional groups were attached onto them to be able to perform binding assays, such as with the addition of fluorophores, or to use the peptide for protein modification by inserting catalytical pyridinium oxime group onto the peptide. While each of these modifications and their use will be discussed in more detail within subsequent chapters, in this chapter we will describe the optimisations performed to maximise functionalisation yields and to develop general methods that could be feasibly applied toward any peptide sequence. Several peptides were synthesised to bind the two main protein targets throughout the PhD, insulin and FGF. Those two proteins were introduced in Introduction and from literature it was possible to obtain previously reported binding peptide sequences.<sup>5-9</sup>

For simplicity the peptide sequences taken from the literature were renamed during this Chapter to allow a clearer discussion of the general procedures developed, which were applicable to all of the sequences given in Table 1. As a general rule, when a peptide **Px** is modified at the *N*-terminus with a compound R, the name given would be **N-R-Px** and when at the *C*-terminus **Px-R-C**. All structures and specific data of the synthesised peptides are given in the Experimental section.

Table 1: List of core peptides used throughout the PhD.

Protein targeted	Peptide sequence	Name use in text	Reference
<b>Insulin</b>	HWWWPAS	P1	5
<b>Insulin</b>	RGFFYT	P2	6
<b>Insulin</b>	LVEALYL	P3	6
<b>Insulin</b>	CVEEAS	P4	7
<b>FGF</b>	HTTHMYL	P5	8
<b>FGF</b>	TLHSAQA	P6	8
<b>FGF</b>	HRNPRNN	P7	8
<b>RNase-A</b>	KETAAAKFERQHMDSTSA	P8	9

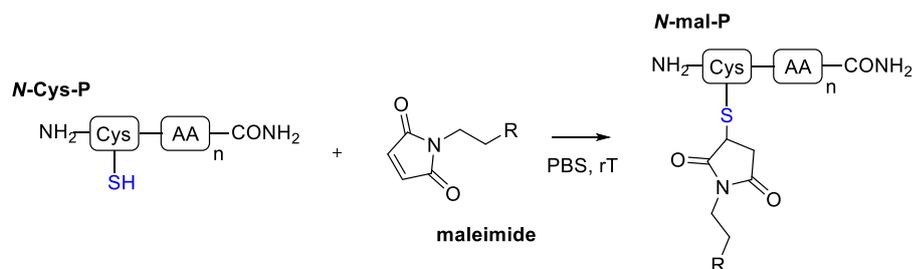
In each following chapter, the peptides synthesised and their specific modification will be listed in the Experimental section of the corresponding chapter.

## 2.3 Modification of peptides after SPPS

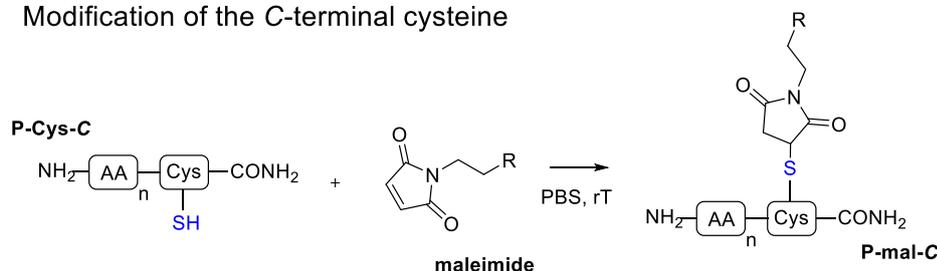
### 2.3.1 Modification using a solution-phase approach

To modify peptides with the functional groups needed for subsequent chapters, two different approaches were explored depending on the sequence of the peptide. In a first approach, the peptides were fully cleaved from the resin and deprotected after synthesis, meaning that different chemical functions were potentially available to react in subsequent steps. Having the future of the project in mind, whereby a large number of diverse peptides might be generated, we initially chose to use thiol-maleimide chemistry for selective modification of the peptides in the solution-phase. Indeed, very few of the peptides we planned to explore had a cysteine in their sequence, making it possible to insert an extra cysteine, providing a thiol-handle, at the peptide *N* or *C*-terminus as a handle for further modification Scheme 4.<sup>10-12</sup>

#### Modification of the *N*-terminal cysteine



#### Modification of the *C*-terminal cysteine



Scheme 4: General scheme for modification of cysteine containing peptides A) at the *N*-terminus and B) at the *C*-terminus in presence of maleimide in PBS 7.4 pH at room temperature. *R* = functional group such as diazine, fluorophore, see Figure 2.

This thiol-maleimide chemistry was attempted on two peptide models, **Cys-P1** and **Cys-P2** using maleimides, **Dz-mal** and **BODIPY-mal**, represented in Figure 2 in phosphate buffer solution (PBS) buffer pH 7.4 at room temperature, conditions similar to literature.<sup>13</sup> Those two maleimide probes were synthesised during the PhD, more information will be given later as this part focus on peptide modification. By testing two peptides and maleimides we hoped to understand what influence either

the peptide sequence or its functionalisation at *N* or *C*-termini would have, and the effects of the maleimide bulk and their intrinsic chemical properties.

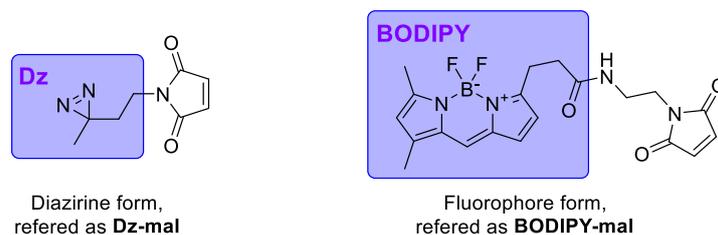


Figure 2: Structures of the two maleimide structures tested one with a diazirine moiety **Dz-mal** and one with a fluorophore **BODIPY-mal**

A first attempt at solution-phase cysteine modification with peptide **Cys-P1** in PBS solution at a concentration of 10 mM was achieved through first treating the peptide with the reducing agent tris(2-carboxyethyl)phosphine TCEP (500 mM), ensuring that the thiol would not be in the form of a disulphide bridge, and then adding 1 eq of **Dz-mal** (100 mM in PBS). LCMS, Figure 3, showed the appearance of a new peak matching the mass for the modified peptide with the addition of maleimide **Dz-P1** after 45 min of reaction.

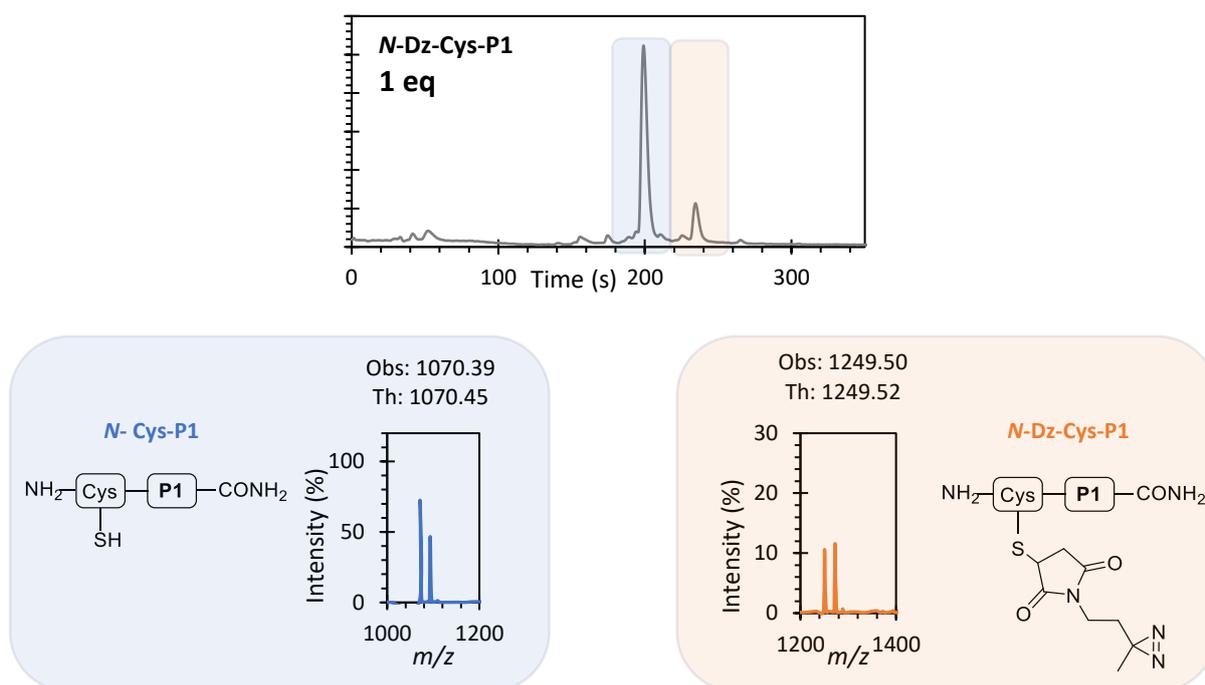


Figure 3: LCMS 280 nm UV absorbance trace and MS of the two highlighted peaks, starting material peptide **N-Cys-P1**, and the modified peptide **Dz-P1**.

Integration of the UV absorbance allowed us to estimate conversion to be 17%. Following that, it was clear that the maleimide needed to be at a higher stoichiometry to achieve higher conversions. The relatively low conversion could be explained by the steric effect of the peptide and by increasing the ratio of the maleimide to the peptide it may then increase conversion. This was achieved by

introducing one extra equivalent of **Dz-mal** every 45 min, with aliquots taken and analysed by LCMS which showed an increase in conversion, as represented in Figure 4.

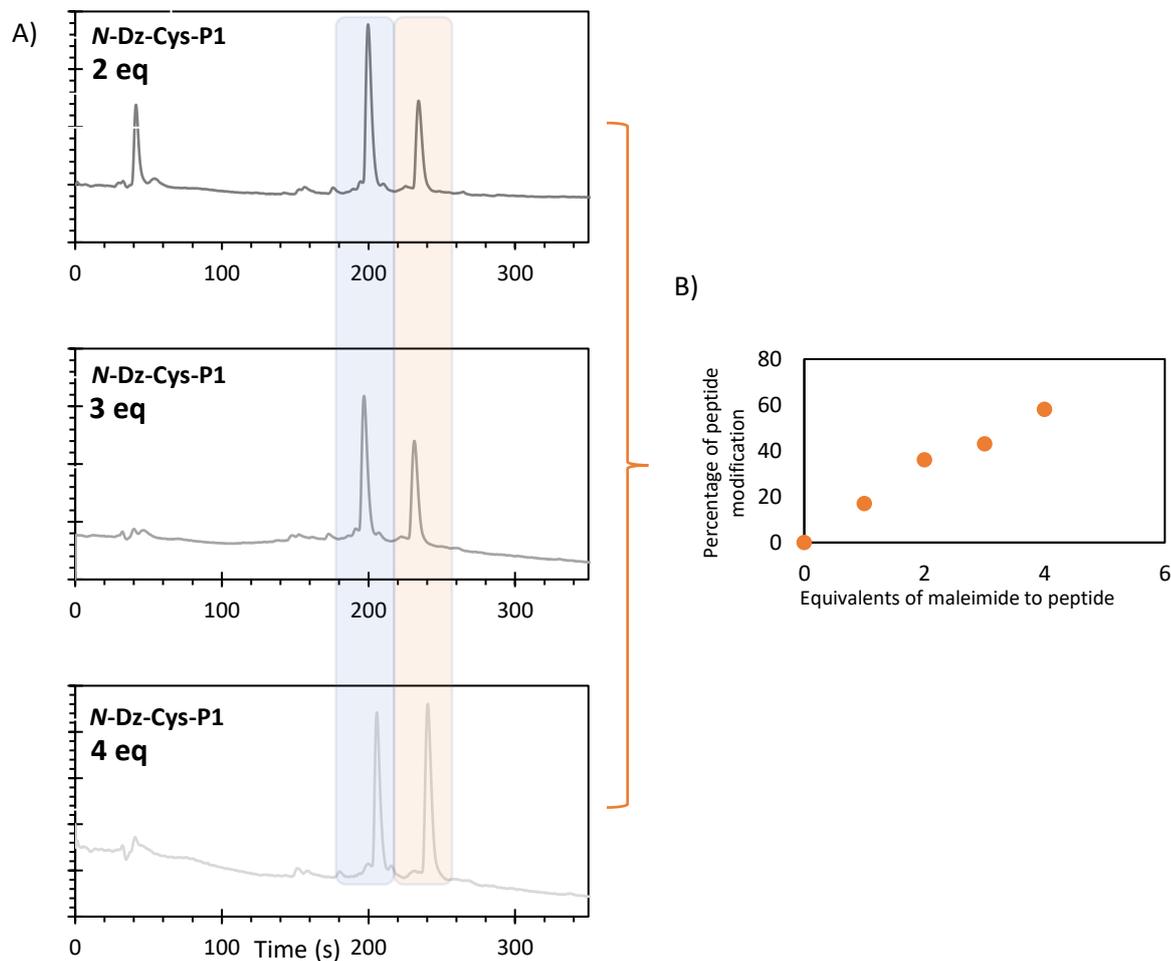


Figure 4: A) LCMS 280 nm UV absorbance traces of the reaction between *N*-Cys-P1 and *Dz*-mal at different equivalents from 2 to 4 eq. Retention time 200 sec in blue for the starting peptide P1 decreasing over time with an increase of the modified peptide mal-P1, at retention time 240 sec. B) Plot of peptide conversion using the 280 nm chromatogram for integration giving the estimated ratio between *N*-Cys-P1 and *N*-Dz-Cys-P1 depending on the equivalents of maleimide introduced in solution.

A gradual increase in conversion was seen up to 4 equivalents, reaching a maximal conversion of 58%. While higher equivalents could have been added to reach higher conversion, we hypothesised that the maleimide could undergo Retro-Michael deconjugation competing with the expected reaction.

As in the first attempt the equivalents of maleimide were added stepwise, and low equivalents showed to be insufficient, another assay was done directly adding 4 equivalents of **Dz-mal** thus also limiting the time of reaction hoping to reduce potential Retro-Michael deconjugation. The conversion reached after 45 min was lower compared to *N*-Cys-P1 and depending on the position of the cysteine the yield differed importantly, reaching 48% for *N*-Cys-P2 and only 20% for P2-Cys-C, as shown in Figure 5. Moreover, another peak at 1.2 min could be observed too, which matched the mass of the reaction between TCEP and the maleimide, compound **2.3**, as shown in Figure 5. This reactivity was unexpected

as TCEP as a reducing agent is often considered as being chemically compatible with Michael acceptors. However, it was shown recently by Watts and all,<sup>14</sup> that when TCEP is used as a reducing agent but not removed during protein conjugation reactions, two rapid and irreversible reactions take place with maleimides, leading to the production of either ylens or succidimidyl derivatives which are both inactive for further modification, decreasing drastically the potential for protein- or in our case peptide-conjugation.

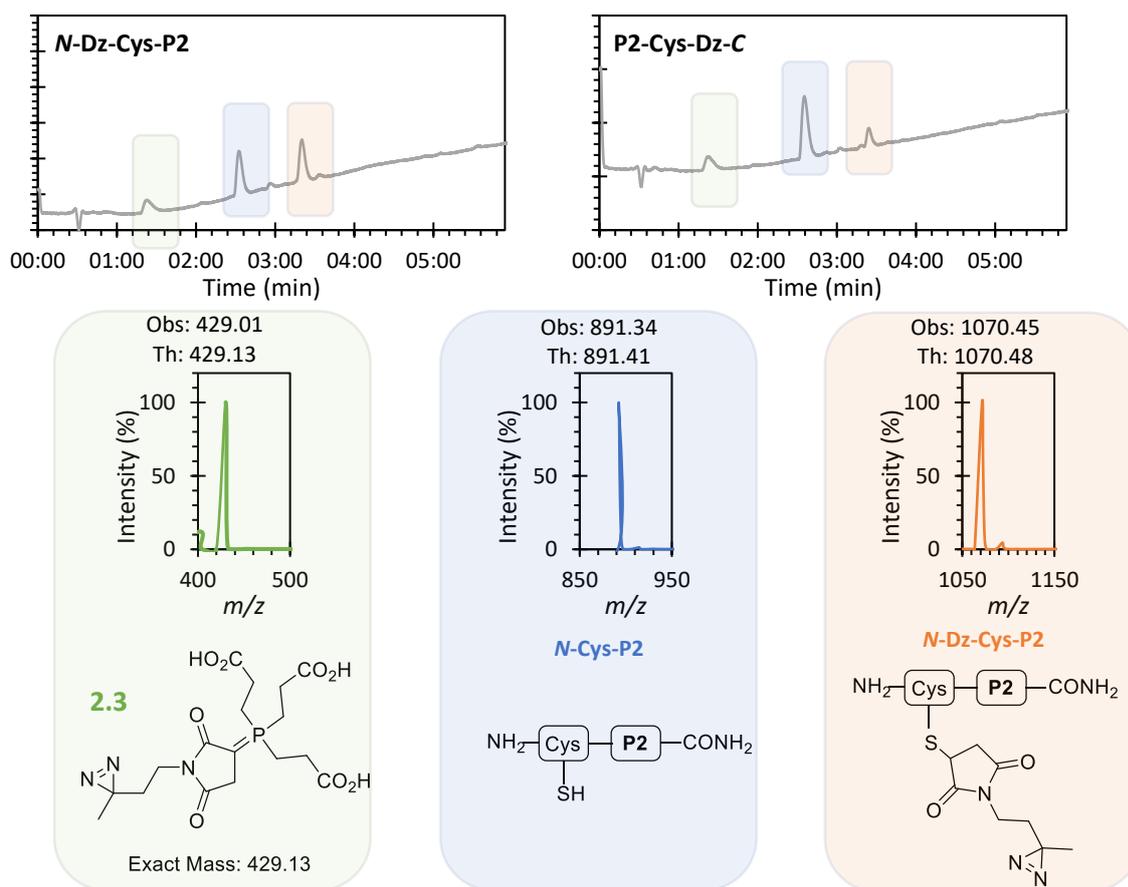


Figure 5: LCMS 280 nm UV absorbance traces and MS data for the modification of peptide **N-Cys-P2** (left) and **P2-Cys-C** (right) with **Dz-mal** under TCEP conditions

To check another important parameter of the method, a second type of maleimide, **BODIPY-mal** see Figure 2, was introduced for modification of **N-Cys-P1**. Having shown a peptide sequence dependency it was also important to see if the change of maleimide structure would interfere too. This time a 19% conversion was reached for the modified-peptide on C-terminus and again a peak corresponding to the reaction between TCEP and **BODIPY-mal**, compound **2.4**, was observed, though the intensity of the UV absorbance may be impacted due to the absorbance of the fluorophore itself, Figure 6.

Once again it seemed that the reaction between the maleimide and TCEP was more efficient than the reaction between the maleimide and the cysteine of the peptide. This could be potentially explained

by first that both these reagents were in excess to the peptide, and second that the thiol could be more sterically hindered due to the 3D structure of the peptide.

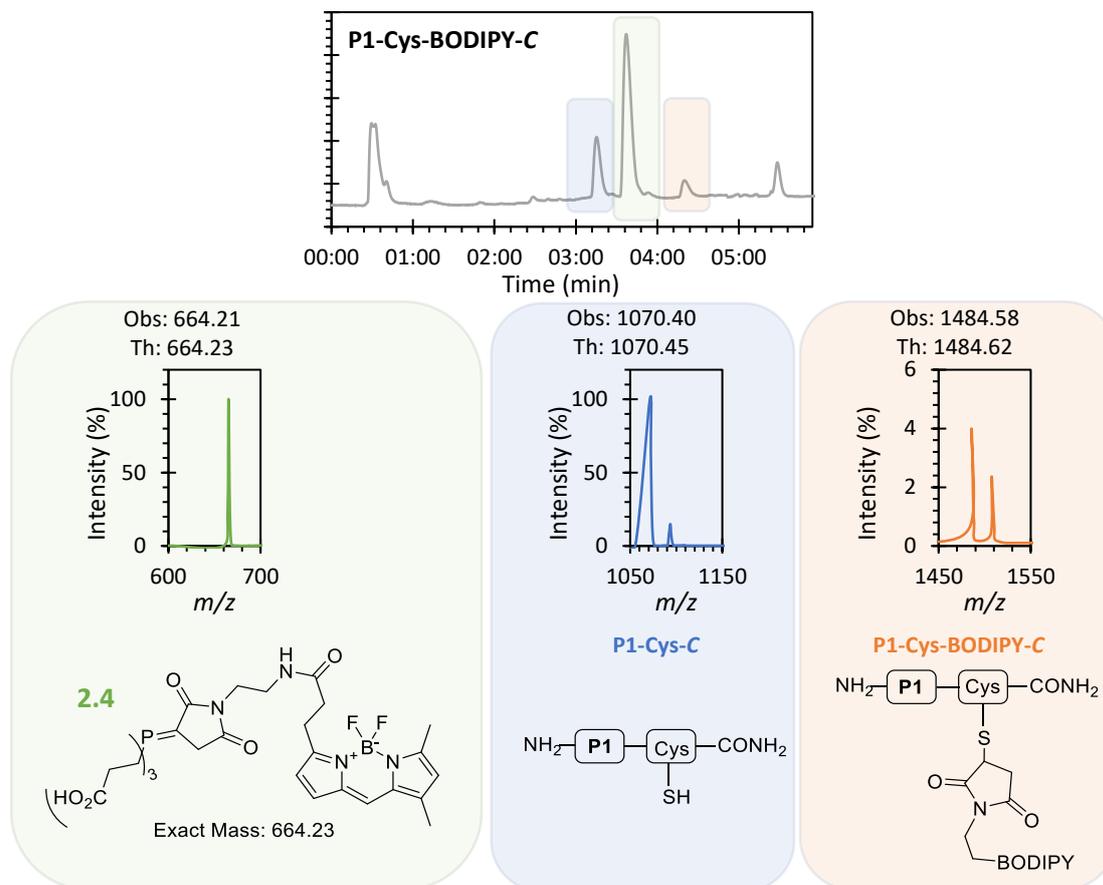


Figure 6: LCMS 280 nm UV absorbance traces and MS data for the modification of peptide **P1-Cys-C** with maleimide **BODIPY-mal** under TCEP conditions

Following this observation, and having observed little formation of disulphide bridges over extended periods, another approach without the use of TCEP for thiol reduction was next tried. Run as described for the standard protocol, but omitting TCEP, 5 equivalents of maleimide were introduced directly to a solution of peptide, see Experimental section for full protocol. Again, this was attempted on both peptides **P1** and **P2** with *N*- or *C*-terminal cysteines. The conversions obtained were considerably higher compared to the attempts done with TCEP, with conversions superior to 80% using **BODIPY-mal** as summarized in Table 2. The LCMS data are presented in the Experimental section.

Table 2: Summary of the conversion reached for peptide modification via solution-phase cysteine modification with the use of TCEP or not. All the LCMS and MS data can be found in the Experimental paragraph. N.A. = reaction not ran.

Peptide sequence	Maleimide use	TCEP use	Conversion (%)
<b>N-Cys-P1</b>	BODIPY-mal	N.A.	N.A.,
<b>P1-Cys-C</b>	BODIPY-mal	YES	19%
<b>N-Cys-P1</b>	BODIPY-mal	NO	86%
<b>P1-Cys-C</b>	BODIPY-mal	NO	93%
<b>N-Cys-P2</b>	Dz-mal	YES	48%
<b>P2-Cys-C</b>	Dz-mal	YES	20%
<b>N-Cys-P2</b>	BODIPY-mal	NO	84%
<b>P2-Cys-C</b>	BODIPY-mal	NO	95%

Though these results were promising, the crude mixtures generated still contained some unmodified peptide, the desired product and the excess of maleimide. Two approaches were then taken at this point to purify the crude mixture. First, a simple organic extraction was put in place between EtOAc and the aqueous solution showing success in the removal of the excess maleimide. However, it was still necessary to separate the non-modified peptide from the product. Second, an attempt for purification was done using Supelclean™ LC-18 6mL column. This type of column works by solid phase extraction on a silica gel base material to extract nonpolar from moderately polar analytes from aqueous samples. These columns were run under gravity and thus took significant time for elution with a manual gradient starting with 100% H<sub>2</sub>O and then gradually adding MeCN (20%, 30%, 40%, 60%, 80%) to finish with a MeOH wash. This led to the collection of a lot of fractions which all needed analysis and which showed poor separation between the species. Thus, these methods due to their manual aspect were not reliable and took a significant amount of time. Moreover, at this point time, HPLC purification was not possible at the University due to the impact of COVID with no training or access possible over a long period of time.

Therefore, to conclude on this maleimide-cysteine organic coupling approach, we saw that first the conversion to the expected peptide was not only peptide sensitive but also regio-dependent, as *N* and *C* terminal cysteines of otherwise identical sequences showed different reactivity. Moreover, the maleimide species also had an important influence. While the last attempt using 5 eq without the presence of TCEP showed great improvement, full conversion was never reached, and purification was necessary. To homogenise the results, the ratio maleimide/peptide could be potentially increased to reach the maximum of conversion, however this would be costlier and would potentially make subsequent purification even more difficult due to the excess of maleimide. We therefore sought an approach that would maximise peptide conjugation efficiency and aid purification.

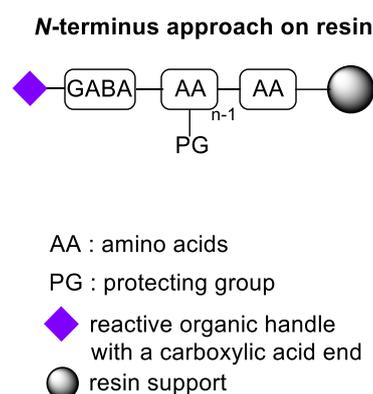
## 2.3.2 Modification using a solid-phase approach

2.3.2.a Use of the primary amine from the N-terminus

Due to the challenges associated with solution-phase modification another approach on-resin was considered. Due to the inherent process of the peptide synthesiser the last step of the synthesis leads to the deprotection of the *N*-terminus freeing an amine. This new handle can be easily modified with functional groups bearing carboxylic acids through the same process of amidation used throughout the synthesis. The modification would only take place at the *N*-terminus as all other functional groups at that stage would be protected by their specific protecting group, ensuring a single modification. Thus, the new strategy to modify the *N*-terminus consisted of using a new type of reactive handle on our functional motifs, from maleimide to a carboxylic acid. We decided at the *N*-terminus to introduce an extra amino acid, the unnatural amino acid GABA, to extend the functional group away from the main protein-binding sequence.

The different steps to release the free amine and the amidation step are:

- 1) Synthesis of the targeted peptide with an extra unnatural GABA at its *N*-terminus
- 2) Deprotection of the Fmoc at the *N*-terminus
- 3) Amide coupling with a carboxylic acid functional handle
- 4) Final deprotection of the peptide and cleavage in TFA



To insert an amine at the *C*-terminus, more effort was required. A later section will be dedicated to that aspect.

Using this solid-phase approach, if good conditions could be found to fully modify the peptide, even if using excess reagents was necessary, the excess could be easily washed out, and consequently would reduce the need for purification after cleavage, expediting the process. As mentioned earlier and using similar chemistry to that used for peptide bond formation, the free amine of the peptide could be reacted on resin with a carboxylic acid to form a new amide bond. During peptide synthesis, the coupling agents DIC and Oxyma were used as an activator and a base respectively, but different conditions were tried to see what the easiest and most effective coupling method for on-resin peptide functionalisation was. All the optimisation was done on the *N*-terminus and the final application on the *C*-terminus will be described in a later section.

### 2.3.2.b First amidation protocol using *N*-hydroxysuccinimide (NHS) activated ester

The first idea was to use the well-known NHS-activated form of the carboxylic acids. Similarly to the modification in solution, different peptides were tested to ensure the generalisability of the method, with two different probes synthesised during the PhD **Dz-ONSu** (see Chapter 3 for carbonate synthesis) and **NBD-NHS**, Figure 7.

A first experiment was done on **N-NH<sub>2</sub>-P1** and **N-NH<sub>2</sub>-P2** using **Dz-ONSu**. 5 equivalents of **Dz-ONSu** were initially used, in the presence of 20 equivalents of base *N,N*-diisopropylethylamine (DIPEA) in dimethylformamide (DMF) for a reaction time between 4 to 5 h. In each case some starting material was observed during LC-MS analysis, however the majority was converted successfully with conversion being respectively 85% for P1 and 90% for P2, as shown in Figure 8. These conditions were also tried with **NBD-NHS** showing very similar conversion of 72% and 95% reached for P1 and P2 respectively (spectra in Experimental section), summary presented in Table 3.

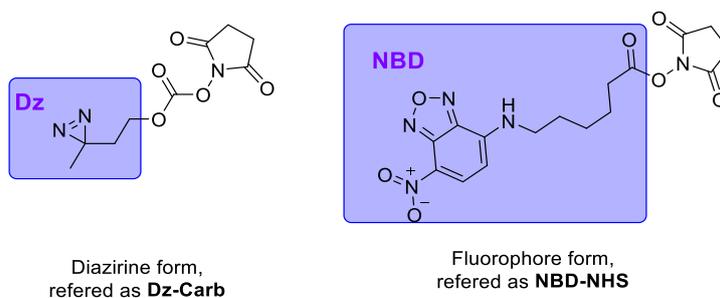


Figure 7: Structures of the two NHS-activated carboxylic structures tested one with a diazirine moiety **Dz-ONSu** and one with a fluorophore **NBD-NHS**

Table 3: Summary of the conversion reached for peptide modification with the use of NHS-ester.

Peptide sequence	NHS use	Conversion (%)
<b>N-NH<sub>2</sub>-P1</b>	Dz-ONSu	85%
<b>N-NH<sub>2</sub>-P1</b>	NBD-NHS	72%
<b>N-NH<sub>2</sub>-P2</b>	Dz-ONSu	90%
<b>N-NH<sub>2</sub>-P2</b>	NBD-NHS	95%

While generally the results were promising, we could see once again that the conversion rate seemed to depend on the peptide sequence. **P1** contained more aromatic residues, thus potentially being more sterically hindered and it could therefore be hypothesised that the *N*-terminus of the sequence would be potentially harder to reach and thus modify. However, while promising, this approach required the formation of an NHS-activated ester. While a lot of compounds can be bought in their carboxylic acid form, the NHS-analogues often require further synthesis adding extra time, and we therefore sought to couple acids directly onto the *N*-terminus.

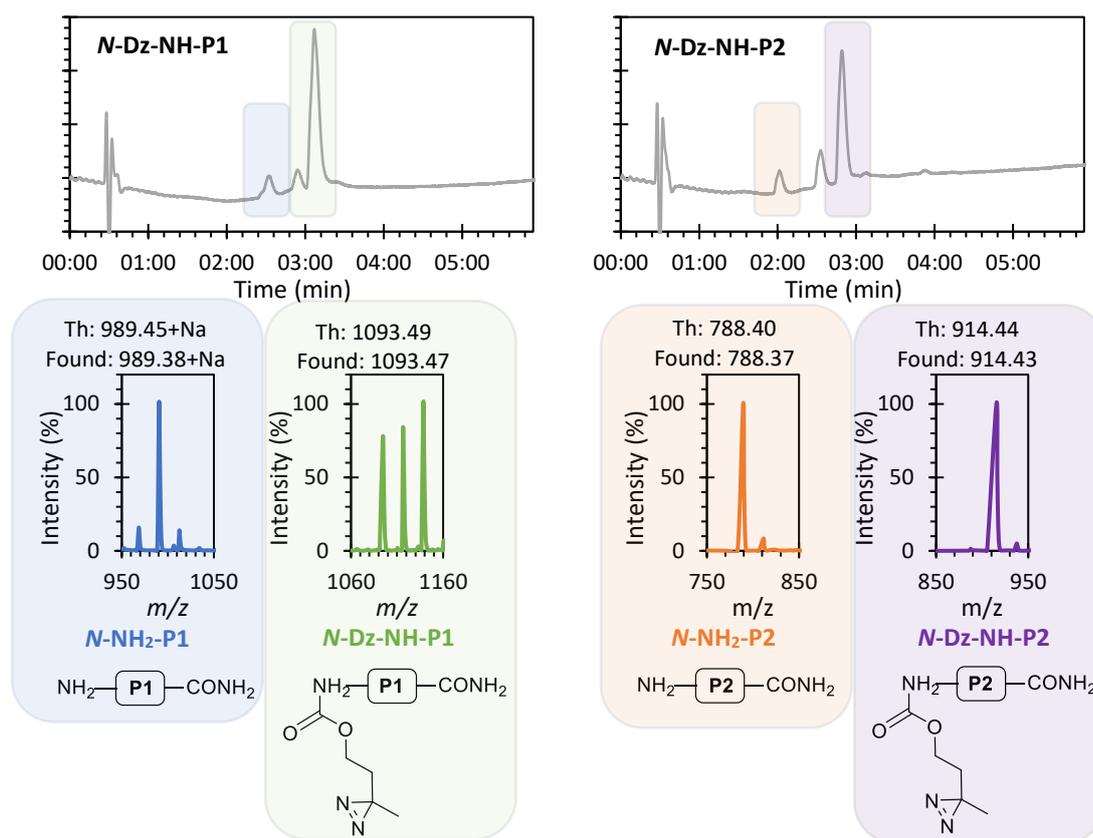


Figure 8: LCMS 280 nm UV absorbance traces and MS data for the modification of peptides **N-NH-P1** and **N-NH-P2** with **Dz-ONSu**

### 2.3.2.c Solid-phase coupling of carboxylic acids

It was then considered to attempt another type of reaction more similar to the conditions used during peptide synthesis using coupling agents to activate carboxylic acids *in situ*. While a lot of conditions for amide couplings have been described,<sup>15–18</sup> two different conditions were tried.

The first used 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium chloride (DMTMM) as a triazine derivative. This condition was picked as it was the conditions used from the literature article used for coupling pyridinium oxime probes to SLF (synthetic ligand for FKBP12)<sup>19</sup> which will be discussed in a following chapter. This coupling agent seemed to be also a good candidate as it was shown that using DMTMM during the synthesis of the 65–74 segment of ACP each coupling went faster (15 min) than with other coupling agents such as O-(Benzotriazol-1-yl)-N,N,N',N'-tetramethyluronium tetrafluoroborate (TBTU) (45 min) or 2-(7-Aza-1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU) (30 min) and gave higher purities (84%) than TBTU (69%).<sup>17</sup> The second approach used a standard carbodiimide, EDC.HCl, in conjunction with 1-hydroxy-1H-benzotriazole (HOBt), which was shown to give higher yields and lower epimerisation levels than EDC.HCl on its own. A reported example showed that when coupling Z-Gly-Phe-OH to H-Val-OMe, the epimerisation levels dropped from 35% to 1.5% in the presence of HOBt. HOBt is believed to work by

initially reacting with the intermediate obtained after carbodiimide addition to give the OBt activated ester, which enhances even greater the reactivity of the “activated ester”.<sup>20</sup> Moreover, while we introduced earlier the reagents DIC and Oxyma for SPPS, those reagents are respectively more toxic and more expensive than EDC.HCl and HOBt. Therefore DIC and Oxyma were not used for post-synthesis modification. The reagents DMTMM, HOBt and EDC.HCl are represented in Figure 9.

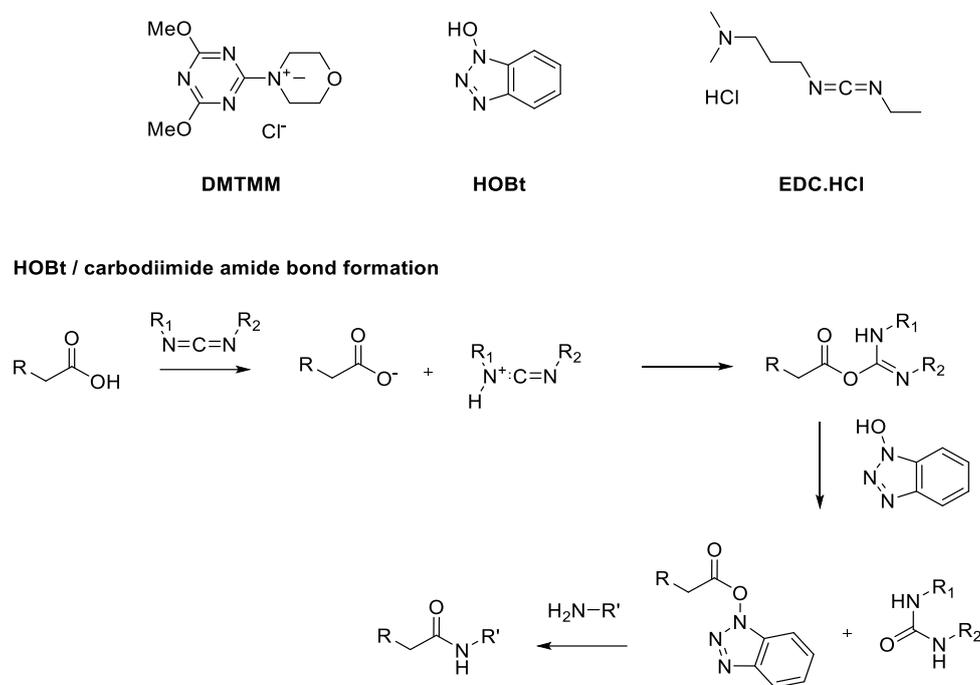


Figure 9: Structures of different coupling agents tested for amide bond formation between a carboxylic acid residue and an amine, with the scheme of amide bond formation using a couple between HOBt and carbodiimide reagent

### 2.3.2.c1 DMTMM

Similarly to the protocol using NHS, DMTMM was introduced in excess (13 equiv.) to the resin-bound peptide, with the carboxylic acid (11 equiv.) and the base DIPEA (15 equiv.) (see Experimental paragraph). After 1 h of reaction and cleavage to recover the peptide, it was shown that using 7-bromoheptanoic acid as the carboxylic acid, 100% conversion was reached for the conversion of P2, as shown in Figure 10. Indeed, while two peaks could be observed, the second matched the mass of a TFA adduct of the expected modified peptide, coming from the cleavage step and not the modification itself. This result was the first time that full conversion was observed showing the good ratio of those compounds.

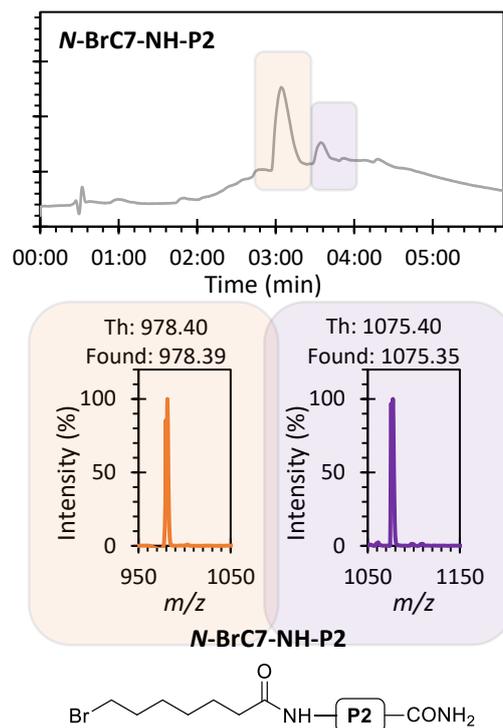


Figure 10: LCMS 280 nm UV absorbance traces and MS data for the modification of peptide **N-NH-P2** with DMTMM protocol using 7-bromoheptanoic acid.

### 2.3.2.c2 HOBT/EDC.HCl

To balance a need for excess reagents and reaction times, another approach with HOBT and EDC.HCl (cheaper reagents compared to DMTMM) were tried. As the time of modification was not the main parameter or the limiting aspect, it was observed that using carboxylic acid (5 equiv.) and HOBT/EDC.HCl (respectively at 10 equiv.) also allowed us to reach 100% conversion after 24 h in solution in DMF, as shown in Figure 11 using 5,6-carboxyfluorescein on peptide **N-NH-P2**. This method which is the most price sustainable was then picked as the general choice for amidation using a free amine on solid-support.

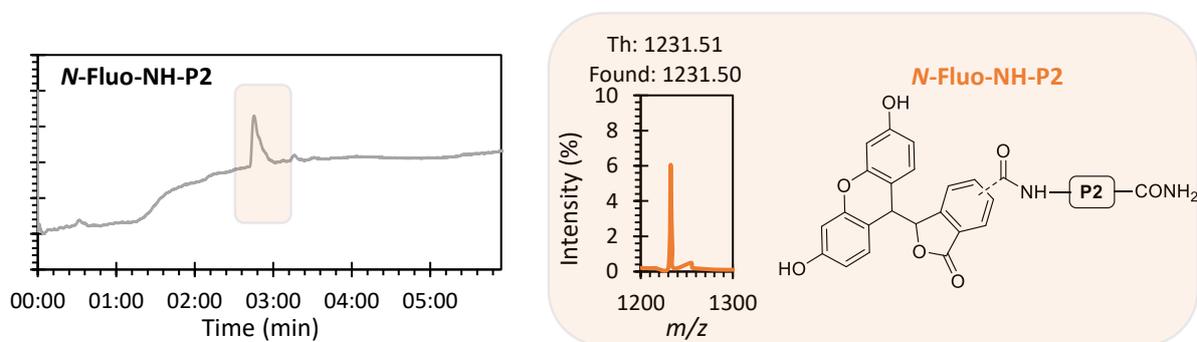


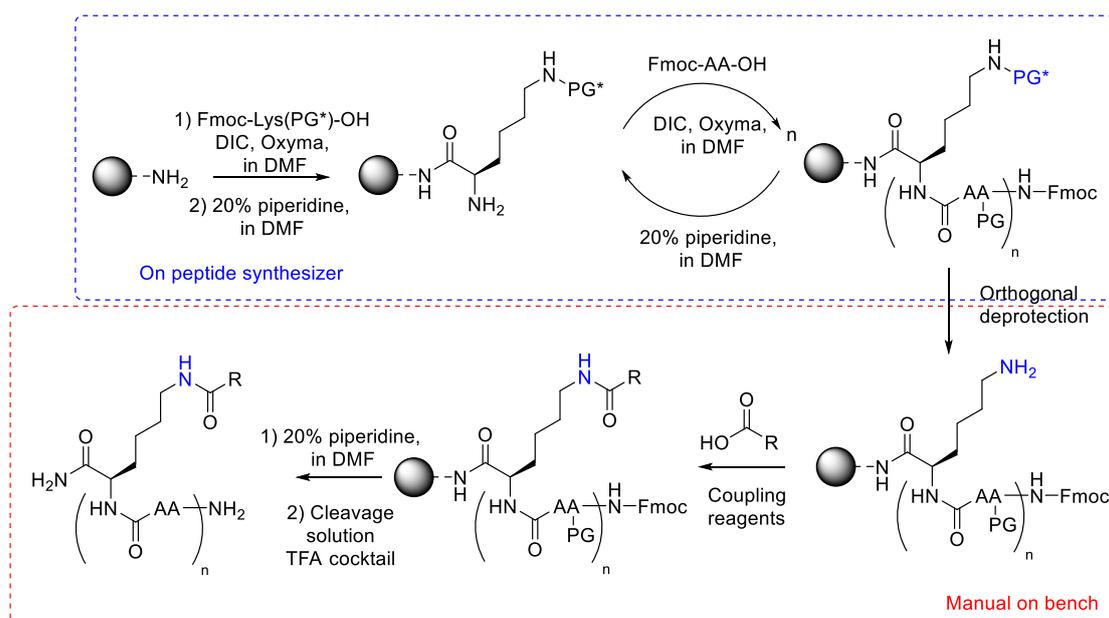
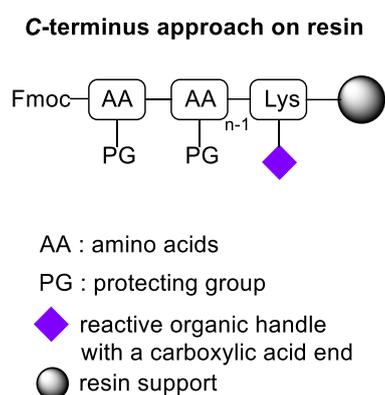
Figure 11: LCMS 280 nm UV absorbance traces and MS data for the modification of peptide **N-NH-P2** with HOBT/EDC.HCl protocol using 5,6-carboxyfluorescein.

It is worth noting that this protocol did not reach full conversion in some cases, often due to the solubility of the carboxylic acid. Sometimes additional peaks were observed after LCMS analysis. In certain cases, deviations from the protocol were made and where this is the case, it is discussed in the following chapters on a case-to-case approach. When coupling was incomplete, or impurities were present, modified peptides were purified by reverse-phase flash column chromatography as mentioned in the Experimental section.

## 2.4 Orthogonal protection of side-group lysine

As mentioned in the previous section, in order to modify a peptide on resin using an amidation process on the C-terminus more steps were necessary. A lysine residue bearing an orthogonally protected side chain was inserted as the first amino acid on the C-terminus. This lysine could be selectively deprotected on resin under conditions that do not induce cleavage or deprotection of the other side chains from both the traditional acid-labile sidechain protecting groups and the Fmoc group (present on the last amino acid of the sequence). The orthogonal deprotection will allow to unhide selectively the amine group from the lysine using an appropriate method which could then be modified via amidation as discussed in the previous section. Finally, the *N*-terminal Fmoc deprotection and cleavage/deprotection step under acidic conditions will be undertaken to obtain the final peptide with its *N*-terminus non-modified and its C-terminus bearing the carboxylic acid of choice. The different steps, general structure and a schematic for the different steps mentioned are summarized below and represented in Scheme 5.

- 1) Synthesis of the targeted peptide with an orthogonally-protected lysine at its C-terminus
- 2) Deprotection of the orthogonal side-chain lysine protecting group
- 3) Reaction with a carboxylic acid functional handle
- 4) Deprotection of the Fmoc-*N*-terminus
- 5) Final deprotection and cleavage in TFA



Scheme 5: Scheme representing the different steps of a modification on resin via amidation due to the insertion of an orthogonal protecting group on a lysine at the C-terminus.

While orthogonal protecting groups are numerous, three of them were used in this PhD, structure shown in Figure 12, and the following paragraphs will explain each of their ease of introduction into peptides and their deprotection conditions. As Cbz and Alloc group are both very common reagents for orthogonal protection not only in peptide but also in organic chemistry, starting materials are relatively cheap and both those reagents only required only a one-step process to be installed on lysine. DDE is a much more recent approach and has been used mainly in peptide chemistry. While it was shown that DDE is orthogonal to acid-labile protecting group, its common method of deprotection uses hydrazine which can also deprotect Fmoc group. As using Fmoc was necessary for our strategy, a new approach described by Bradley and co-workers was used to deprotect DDE.<sup>21</sup>

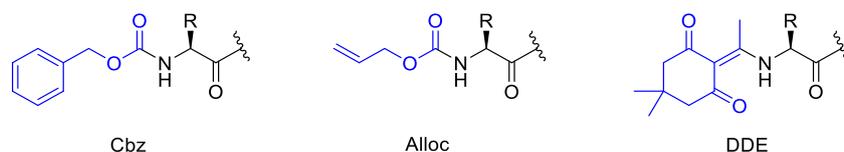


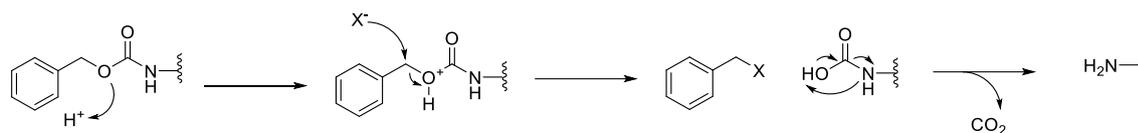
Figure 12: Representation in blue of the different investigated protecting groups of lysine, Cbz, Alloc, and DDE.

#### 2.4.1 An attempt with Cbz group

The Cbz (Benzyloxycarbonyl) group was one of the first used orthogonal groups to Boc-chemistry due to its acid-resistant characteristic. It was introduced as a general protecting group for organic chemistry as early as 1932 by Leonidas Zervas.<sup>1</sup> While the Cbz group has been extensively used in organic synthesis, very few reports with full experimental procedures exist for deprotection on peptides on resin. We therefore explored this protecting group and had different attempts to deprotect it while present on resin.

The sequence Fmoc-P3-K(Cbz)-CONH<sub>2</sub> was synthesised, using the conditions described in Experimental for SPPS procedure, by Jack Askew, a BSc student in the group under my supervision. We first attempted to demonstrate the successful incorporation of Cbz-Lys by cleaving a portion of the synthesised peptide in a TFA cocktail mixture for 18 h. However the results showed that using these conditions the Cbz was also deprotected obtaining the sequence Fmoc-P3-K(NH<sub>2</sub>) instead of Fmoc-P3-K(Cbz). When under strong acid conditions for too long, the Cbz was also able to be cleaved as shown in Scheme 6 which was also reported in literature.<sup>22</sup> Therefore the time of cleavage was reduced to 2 hr, which showed a considerably smaller degree of Cbz deprotection, with less than 10% of compound Fmoc-P3-K(NH<sub>2</sub>) formed, which showed that this group is much more acid-sensitive than planned, Figure 13.

In presence of strong acid XH



Scheme 6: Mechanism for the deprotection of Cbz under acidic condition with XH the acid agent.

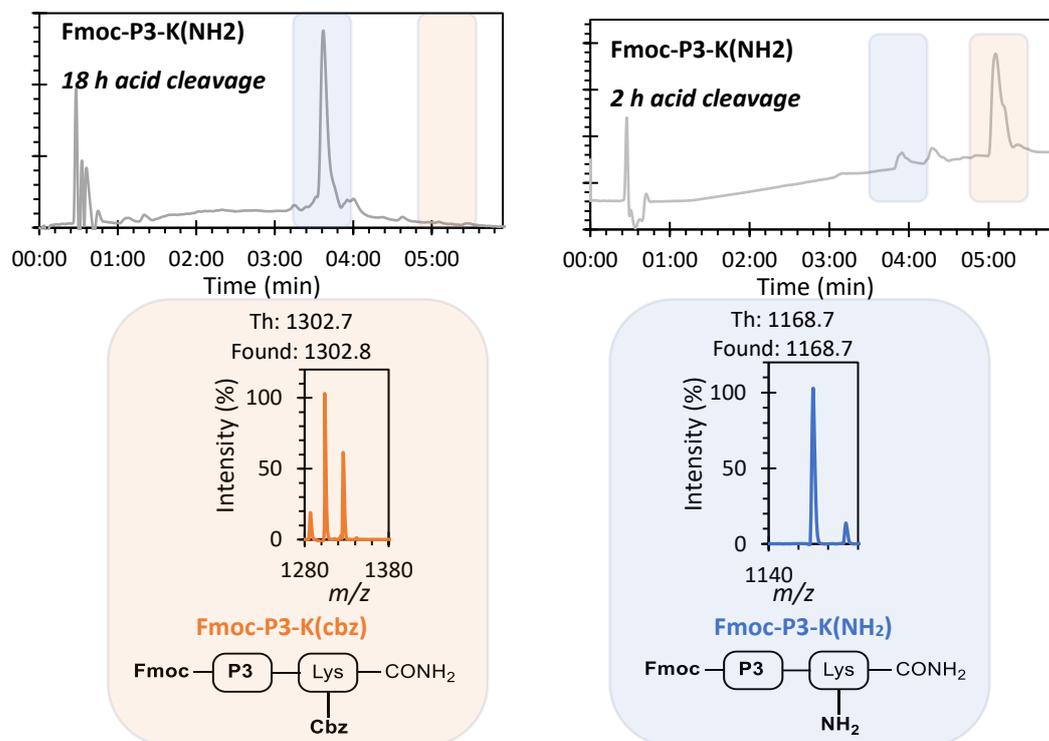


Figure 13: Graph showing the deprotection of **Fmoc-P3-K(Cbz)** in a TFA cocktail for either 18 h or 2 h, LCMS UV 220 nm trace recorded on a function of time.

The main protocols for the deprotection of Cbz found in literature were then applied. Protocol 1 corresponded to classic hydrogenation reactions using palladium on carbon as a catalyst with a source of hydrogen. Less than 10% of deprotected peptide was obtained, Figure 14. Protocol 2 used NaBH<sub>4</sub> (1 eq) to generate H<sub>2</sub> in-situ with the solvent MeOH. This was tried to negate the need for gaseous H<sub>2</sub>, reducing safety concerns, and improving the ease of reaction. This however yielded only a very small conversion and so 10 equivalents of NaBH<sub>4</sub> were then used. Again, this gave very low conversions. We considered that the small conversion could be again coming from the TFA deprotection/cleavage as the increase of equivalents of NaBH<sub>4</sub> did not increase the deprotection. Following these failed attempts, another source of palladium was used with palladium acetate and ammonium formate but once again without success, Figure 14.

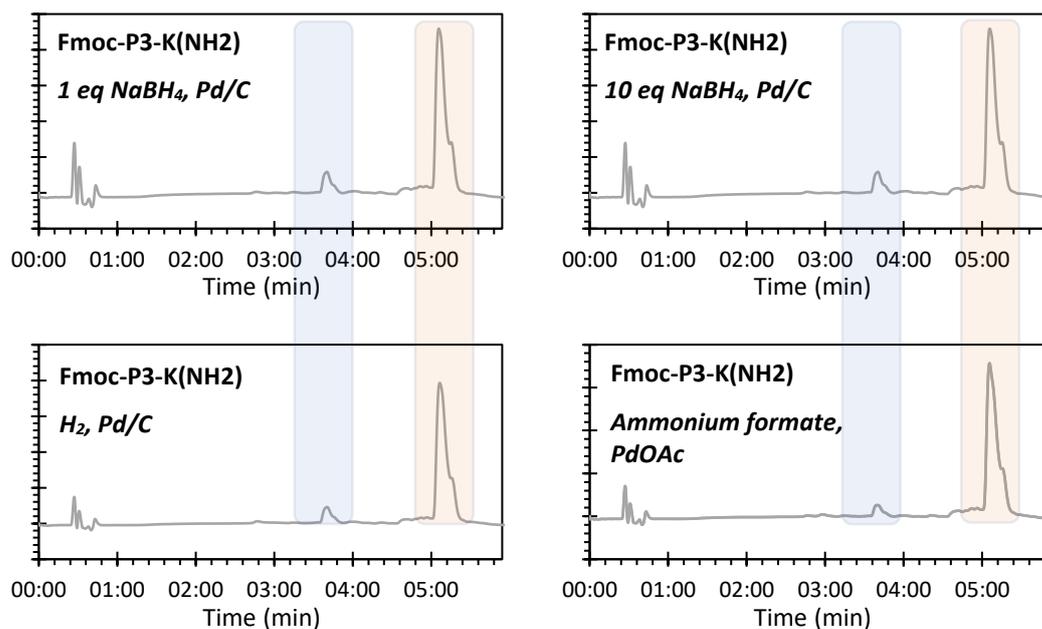


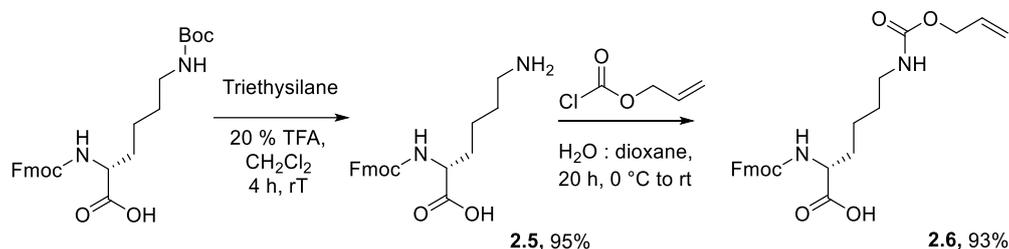
Figure 14: Graphs of the different conditions tested for **Fmoc-P3-K(Cbz)** deprotection on resin, LCMS UV 280 nm trace recorded on a function of time. See Experimental section for full procedure.

No full understanding of these results has been concluded but one main hypothesis has been developed. As the Cbz group is at the C-terminus of the peptide, and thus very close in proximity to the resin, the bulk of the resin and the peptide could have sterically hindered the access of hydrogenated palladium preventing reaction. We tried to overcome this problem by increasing the pressure of hydrogen in the first protocol with Pd/C by directing the hydrogen directly into the solution with the use of a needle instead of having just a balloon on top of the reaction vessel. This needed to be balanced against safety concerns. While not on resin, Sultane et al.<sup>23</sup> reported high yields and fast kinetics when using higher equivalences of higher % Pd/C. However, this would increase the cost and once again the safety with the higher loadings of material. Finally between the sensitivity to acid solution and the poor results for deprotection under different conditions, the Cbz route for protection of the lysine was stopped and other protecting groups were explored.

#### 2.4.2 An attempt with Alloc group

The second type of protecting group which was investigated during this PhD to selectively deprotect a C-terminal lysine was the Alloc group. Compared to the Cbz group, Alloc has been much more used on resin and two different protocols were followed.<sup>24</sup> Fmoc-Lys(Alloc)-OH was synthesised in house due to its high commercial cost. The synthesis was achieved in a two-step process from commercially available Fmoc-Lys(Boc)-OH. The Boc group was removed in a TFA mixture and different conditions for the work-up were tried to obtain **2.5** in the best yield. Precipitation in 1 M sodium hydroxide was found to give better yields (95%) than precipitation in diethyl ether (< 50%). The second step consisted

of the conversion of **2.5** to its Alloc-protected form using allyl chloroformate, which had a very good yield to product **2.6** (93%), Scheme 7.



Scheme 7: Scheme for the synthesis of Fmoc-Lys(Alloc)-OH

**2.6** could then be used in SPPS, and no significant difference could be observed for the purity and yield of the peptides generated compare to those generated containing the standard Fmoc-Lys(Boc)-OH amino acids.

After having a first attempt to deprotect the Alloc group using protocol A,<sup>2</sup> using glacial acetic acid, N-methylmorpholine and Pd(PPh<sub>3</sub>)<sub>4</sub> for 16 h, cleavage was done, and the product analysed by LCMS. The results were surprising giving a relatively clean UV spectrum with two main peaks but none matching the wanted product, Figure 15. The first peak (highlighted in blue) matched the mass of **H<sub>2</sub>N-P4-K(Alloc)** and the second peak (highlighted in orange) matched the starting material **Fmoc-P4-K(Alloc)**. Instead of observing the Alloc group being deprotected, we were observing primarily the deprotection of the Fmoc group. As a partial cleave was performed straight after synthesis it was impossible to know if this was due to this protocol or that the last step of the synthesis affected the deprotection of the Fmoc group. Therefore, this protocol was repeated on a fresh batch of peptide on resin resulted in a mixture of products which showed that this protocol was not adaptable for the use needed in this PhD. While unclear on the results, we hypothesized similarly to previously due to the presence of the Alloc group on the C-terminus of the peptide it could have been sterically hindered preventing reaction with a bulky palladium catalyst.

A second set of conditions for deprotection were therefore attempted using Pd(PPh<sub>3</sub>)<sub>4</sub> and phenylsilane. A first attempt was done at room temperature as inspired by literature for 4.5 h,<sup>25,26</sup> resulting in 13% deprotection which was greater than had been observed until now. To increase the deprotection efficiency, three parameters were studied, time, temperature, and the quantity of catalyst. When increasing this last parameter, from 0.2 equivalent to 0.5 equivalent and when Pd(PPh<sub>3</sub>)<sub>4</sub> was added in two portions, once at the beginning and the 0.3 extra equivalents at the middle time point of the reaction, a considerable increase was observed with 64% conversion rate, Figure 16.

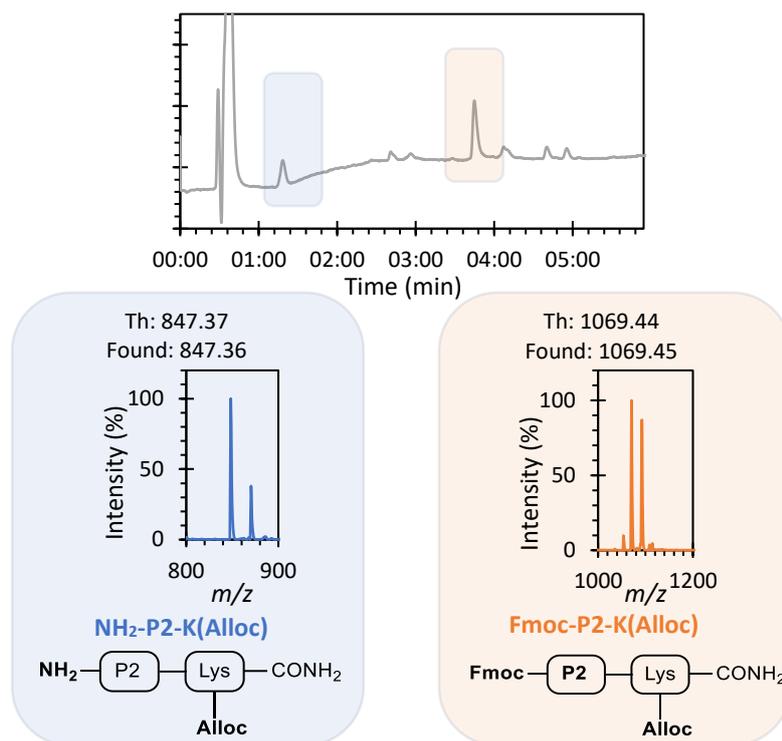


Figure 15: Graph of the deprotection of Alloc group using AAPTEC protocol with glacial acetic acid, *N*-methylmorpholine and  $\text{Pd}(\text{PPh}_3)_4$  after an overnight reaction. LCMS UV 280 nm trace recorded on a function of time.

To increase this yield even more, the reaction was left for 18 h. This did not seem to increase significantly the deprotection efficiency with 74% conversion reached. However, two new peaks were present too showing Fmoc-deprotection. Then the last parameter, temperature was therefore increased from room temperature to 40 °C, with shorter time as it was shown that at longer times Fmoc deprotection took place. This last attempt showed that Fmoc-group was also sensitive to temperature as the same extra peaks were observed for Fmoc deprotection (peaks highlighted in green and purple in Figure 16).

In conclusion while this approach was significantly better than those used for Cbz deprotections, no conditions were ideal for full deprotection of the Alloc group without any other side reaction.

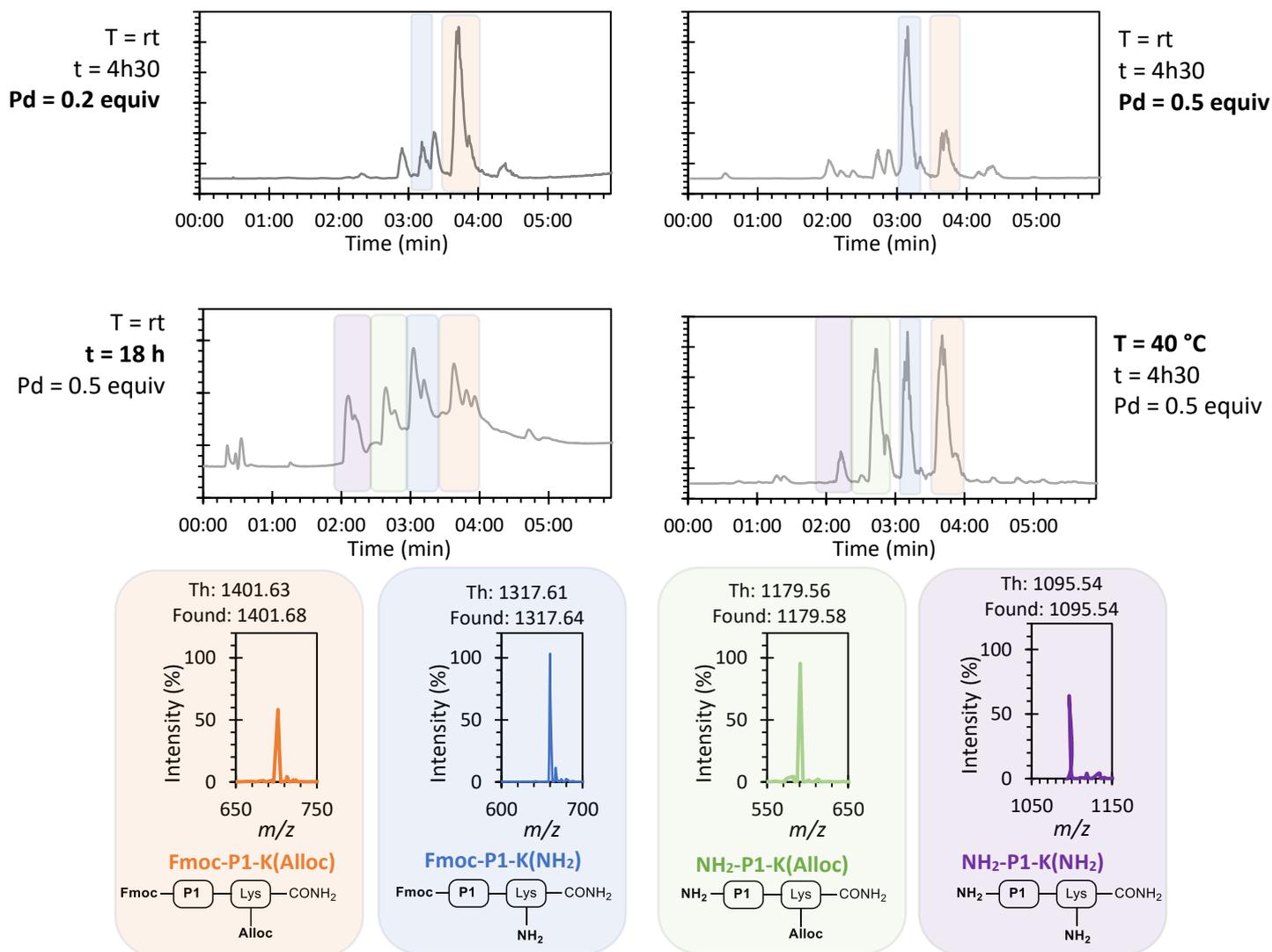
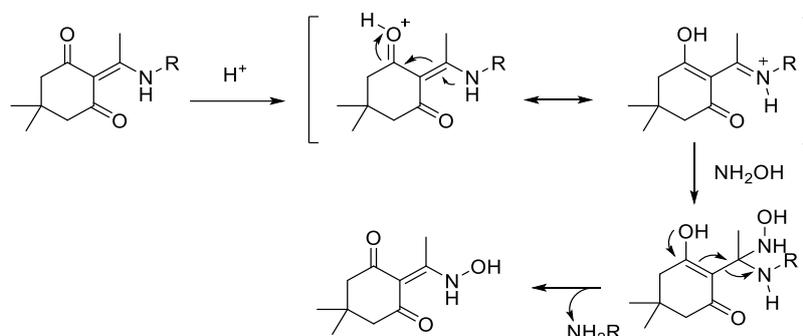


Figure 16: Graphs of the deprotection of Alloc group using  $\text{Pd}(\text{PPh}_3)_4$  and phenylsilane using different conditions of time, temperature and palladium catalyst quantity. LCMS UV 220-280 nm traces recorded on a function of time

### 2.4.3 DDE as the best candidate

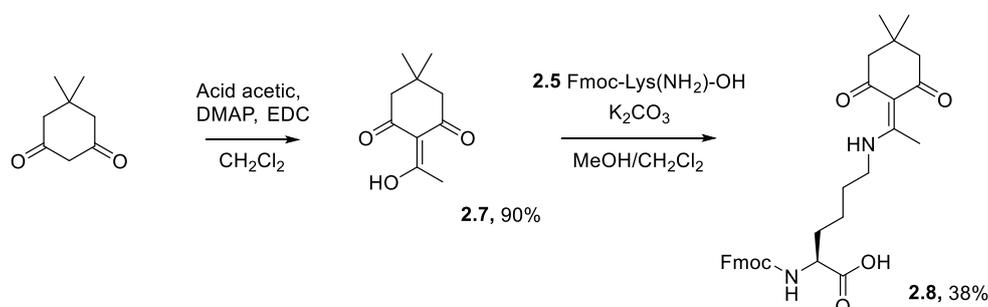
Finally a new type of protecting group DDE (1-(4,4-dimethyl-2,6-dioxocyclohex-1-ylidene)ethyl) was tried. This group was first reported in 1998,<sup>27,28</sup> and was adapted very quickly for the synthesis of peptide containing lysine which required orthogonal protecting group to the Boc chemistry. Moreover, DDE showed to be orthogonal not only to Boc chemistry but also Alloc and Fmoc chemistry which made it a very attractive candidate when complex peptide requiring additional modification of amine where needed. The first report about the deprotection was using hydrazine hydrate ( $\text{NH}_2\text{NH}_2 \cdot \text{H}_2\text{O}$ ).<sup>27</sup> While this protocol showed to be extremely efficient showing more than 99% deprotection in a short time (protocol from literature using 2%  $\text{NH}_2\text{NH}_2 \cdot \text{H}_2\text{O}$  in two cycles of 10 min each)<sup>2</sup> this showed that Fmoc was also deprotected. Therefore, the *N*-terminal residue needed to be

protected with a Boc group. While this is possible using *N*-terminal Boc-amino acid or by reacting the *N*-terminal residue with Boc anhydride this would add either an extra step or require a complete set of *N*-terminal Boc-amino acid which would have been too costly for the project. Only a few years later in 2004 Bradley and co-workers,<sup>21</sup> showed full orthogonality between Fmoc and DDE using hydroxylamine hydrochloride (NH<sub>2</sub>OH.HCl) and imidazole in *N*-methyl-2-pyrrolidone (NMP), the mechanism of deprotection being shown in Scheme 8. The DDE deprotection occurs in presence of a nucleophile, hydroxylamine hydrochloride, with a base which can act as a hydrogen transfer doner, imidazole. There is first a transamination reaction following the protonation of one of the carbonyl from the DDE group, as shown in Scheme 8 followed by an elimination process, releasing the primary amine protected in the first place. Compared to the other protocols the advantage of using this deprotection mechanism is the ease of access of the smaller nucleophile such as hydroxylamine to the sterically hindered C-terminus making the protocol more likely to be successful.



Scheme 8: Mechanism for DDE deprotecting using hydroxylamine hydrochloride and imidazole

DDE, compound **2.7**, was then synthesized in-house for a first time starting with dimedone to ensure that the conditions would work before buying Fmoc-Lys(DDE)-OH a compound much more expensive than the common Fmoc-Lys(Boc)-OH use for SPPS. After a very straight forward first step yielding **2.6** in a 90% yield, the coupling with the deprotected Fmoc-Lys(NH<sub>2</sub>)-OH compound **2.5** gave the expected Fmoc-Lys(DDE)-OH **2.8** in a 38% yield after purification.



Scheme 9: Synthesis steps for the formation from dimedone of DDE (compound **2.7**) and then Fmoc-Lys(DDE)-OH **2.8**

Similarly as before, the addition of this orthogonal protected lysine to peptides was successful using standard microwave peptide synthesis conditions. Following the addition of **2.7** on resin on a first peptide Fmoc-P2-K(DDE)-CONH<sub>2</sub> (once again synthesised by Jack Askew), the deprotection was attempted using Bradley and co-workers protocol with hydroxylamine hydrochloride and imidazole.<sup>21</sup> While a first attempt to the deprotection of the DDE group was done for 3 hr, after cleavage and LCMS, it was shown that some starting material was still present. The protocol was then adapted differently with lower equivalents of reagents from 110 eq of NH<sub>2</sub>OH.HCl to 24 eq, 85 eq of imidazole to 20 eq (full protocol is described in the Experimental section) but a longer time of reaction from 3 hr to an overnight reaction, showing full deprotection of the DDE group (Figure 17). This was found to be reproducible on a series of peptides, showing generalisability of this method. This method was by far the simplest to set up with the advantage of using reasonably cheap deprotection reagents without the need of any metal catalyst. Thus, DDE protection was therefore pursued as the method of choice for any C-terminal orthogonal protection.

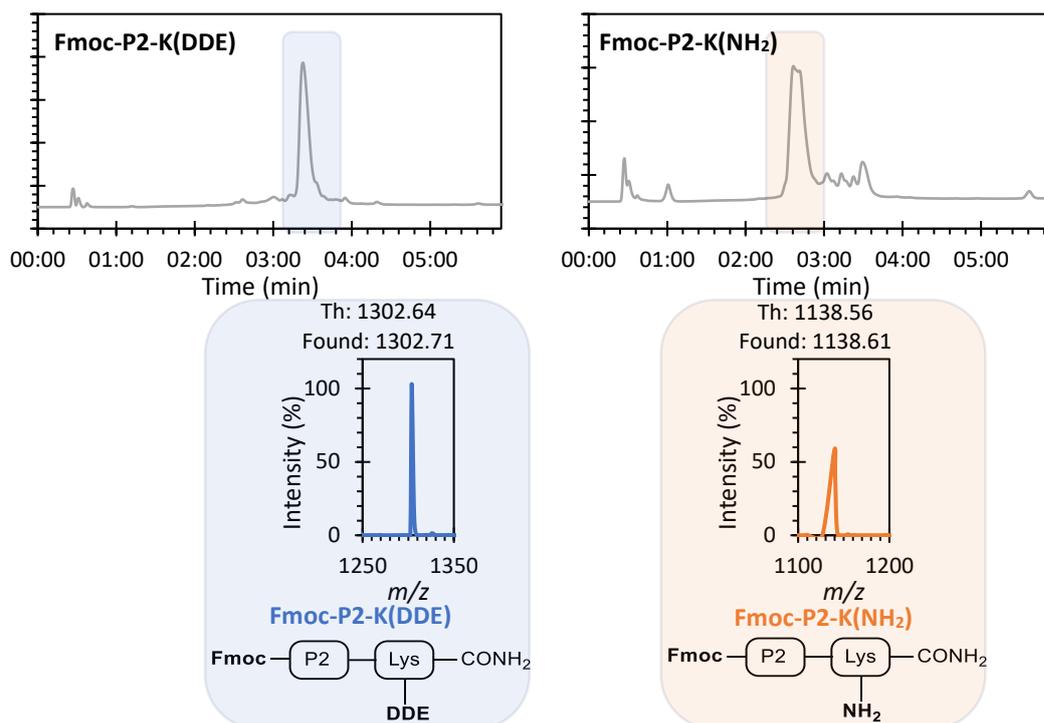


Figure 17: Graphs of the deprotection of Fmoc-P2-K(DDE)-CONH<sub>2</sub> using NH<sub>2</sub>OH.HCl and imidazole resulting in the sequence Fmoc-P2-K(NH<sub>2</sub>)-CONH<sub>2</sub>. LCMS UV 220-280 nm traces recorded on a function of time.

Finally, the optimised conditions from paragraph 2.3.2.c1 EDC/HOBt showed full conversion to modify the C-terminal lysine residue with 5,6-carboxyfluorescein showing the good generality of this method to functionalise amine on resin, Figure 18. The broadness of the peak was due to the presence of fluorescein isomers and the peak at 3.8 min is 5,6-carboxyfluorescein excess.

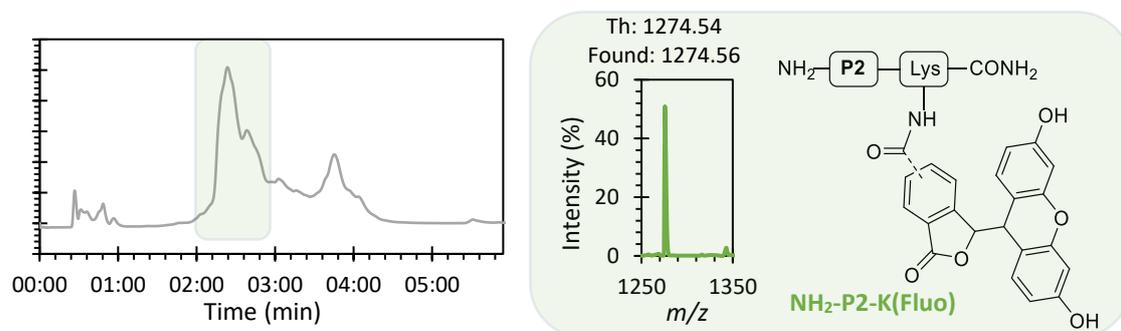


Figure 18: Graphs of the modification of Fmoc-P2-K(NH<sub>2</sub>)-CONH<sub>2</sub> using 5,6-carboxyfluorescein. LCMS UV 220-280 nm traces recorded on a function of time.

## 2.5 Summary

In this Chapter, we first saw the ease to synthesis a broad range of peptides thanks to SPPS. Two strategies were then discussed to modify the peptide post-synthesis using either a thiol-maleimide approach by inserting on the sequence an extra cysteine and modifying this residue in solution or an amidation approach by using a free amine from the sequence and then modifying it while the peptide was still on solid support. While the first approach showed some potential in solution as thiol-maleimide chemistry was specific to the only cysteine no conditions were found to fully modify the peptides and problems of purification of the final peptides were encountered. The second strategy was done on resin using either the free *N*-terminus or by inserting an extra lysine at the *C*-terminus to each time have access to a free primary amine. While adding the lysine required to find orthogonal conditions to deprotect only this group on resin and a few conditions were tested a final protecting group, DDE, was found to be easily synthesised and easily deprotected on resin. This candidate was the best and therefore pursued as the method of choice for any *C*-terminal orthogonal protection. Moreover, while the variety of amidation process known in literature is very broad, and while a few of them were tested to achieve the best conversion rate, the final optimised conditions with EDC.HCl and HOBt showed the highest potential throughout several sequences. Those conditions were then the standard conditions used in all later chapters unless cited otherwise.

The work done during this Chapter is setting up every modification of all the peptides synthesised and modified throughout the whole PhD. The several optimisations were critical to ensure first that the highest conversion rates were obtained to simplify or in the best case avoid any purification step, and second to ensure the generality of the method by using a range of sequences with different characteristics such as hydrophilicity and choice of amino acids.

While this work was essential to the main goal of the PhD, building new peptide probes for close-affinity protein modification as explained in the Introduction, it was at that point necessary to ensure that the peptides once modified showed still affinity to their protein of interest which will be demonstrated in the next Chapter.

## 2.6 Experimental

### 2.6.1 General considerations

Deionized water was used for chemical reactions and for protein manipulations. All other solvents were used as supplied (Analytical or HPLC grade), without prior purification. Reagents were purchased from Sigma-Aldrich, VWR, or Fluorochem and used as supplied, unless otherwise indicated. Brine refers to a saturated solution of sodium chloride. PET refers to the fraction of petroleum ether boiling in the range 40-60 °C. H<sub>2</sub>O refers to distilled water. Anhydrous magnesium sulfate (MgSO<sub>4</sub>) was used as the drying agent after reaction workup unless otherwise stated. RNase A from bovine pancreas (powder, 50 units/mg protein) and insulin (powder) were purchased from Sigma Aldrich.

Proton nuclear magnetic resonance (<sup>1</sup>H NMR) spectra were recorded on a Bruker AVII (300 MHz), Jeol ECX-400 (400 MHz), Bruker AVIIIHD (500 MHz) or Bruker AVIIIHD (600 MHz) spectrometer. Carbon nuclear magnetic resonance (<sup>13</sup>C NMR) spectra were recorded on a Jeol ECX-400 (100 MHz) or Bruker AVII (75 MHz) spectrometer. NMR shifts were assigned using COSY, HSQC and HMBC spectra. All chemical shifts are quoted on the  $\delta$  scale in ppm using residual solvent as the internal standard (<sup>1</sup>H NMR: CDCl<sub>3</sub> = 7.26; CD<sub>3</sub>OD = 3.31; D<sub>2</sub>O = 4.69; DMSO-*d*<sub>6</sub> = 2.50 and <sup>13</sup>C NMR: CDCl<sub>3</sub> = 77.16, CD<sub>3</sub>OD = 49.00, DMSO-*d*<sub>6</sub> = 39.52). Coupling constants (*J*) are reported in Hz with the following splitting abbreviations: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, app = apparent, br = broad.

Melting points (m.p.) were recorded on a Gallenkamp melting point apparatus. Infrared (IR) spectra were recorded on a Perkin Elmer UATR Two FT-IR spectrometer or a Bruker Alpha II ATR spectrometer with Opus build 8.5.29. High resolution electrospray ionisation (ESI) mass spectra (HRMS) were recorded on a Bruker Compact TOF-MS or a Jeol AccuTOF GCx-plus spectrometer. Nominal and exact *m/z* values are reported in Daltons (Da).

Thin layer chromatography (TLC) was carried out using aluminium backed sheets coated with 60 F254 silica gel (Merck). Visualization of the silica plates was achieved using a UV lamp ( $\lambda_{\text{max}}$  = 254 nm), potassium permanganate (5% KMnO<sub>4</sub> in 1M NaOH with 5% potassium carbonate), or ninhydrin (1.5% ninhydrin, 3% AcOH in *n*-butanol). Flash column chromatography was carried out using Geduran Si 60 (40-63  $\mu\text{m}$ ) (Merck). Mobile phases are reported as % volume of more polar solvent in less polar solvent. Anhydrous solvents were dried over a PureSolv MD 7 Solvent Purification System.

### 2.6.2 General conditions for SPPS

Solid-phase peptide synthesis (SPPS) was performed on a CEM Liberty Lite Automated Microwave Peptide Synthesiser, according to the manufacturer's standard protocols. Briefly, Fmoc-protected amino acids (5 equiv., 0.2 M in DMF) were coupled in the presence of *N,N'*-Diisopropylcarbodiimide (DIC, 5 equiv., 1.0 M in DMF) and Oxyma Pure (5 equiv., 1.0 M in DMF), as coupling agent and base

respectively, under microwave irradiation at a temperature of 90 °C for 2 minutes. Arginine and cysteine amino acids were coupled at 25 °C for 25 minutes and 50 °C for 10 minutes respectively. Fmoc deprotection was performed using 20% piperidine in DMF (4 mL) at 90 °C for 60 seconds. Syntheses were performed on either a 0.05, 0.1 or 0.25 mmol scale, using Rink Amide MBHA resin (C-terminal amide, 0.5 mmol/g, 1% DVB, 100-200 mesh, Fluorochem). Prior to cleavage, the resin was washed sequentially with dichloromethane (2 × 15 mL), methanol (3 × 15 mL) and again dichloromethane (2 × 15 mL) and was then either used for cleavage or stored at - 20 °C.

### 2.6.3 General conditions for LCMS

Liquid chromatography-mass spectrometry (LCMS) was performed on a HCTultra ETD II ion trap spectrometer, coupled to an Ultimate300 HPLC using an Accucore C18 column (150 × 2.1 mm, 2.6 µm particle size). Water (solvent A) and acetonitrile (solvent B, MeCN), both containing 0.1% (v/v) formic acid, were used as the mobile phase at a flow rate of 0.3 mL.min<sup>-1</sup>. The solvent gradient was programmed as a gradient of 10-100% (v/v) MeCN in water over 6 min as shown in Figure 19. Samples were prepared at a concentration of ~0.1 mg/ml in 1:1 MeCN/water (v/v). 10 µL injections were used in each run. LC traces were measured via UV absorption at 220, 270, 280, and 220-400 nm and the mass spectra (in positive and negative mode,  $m/z = 200-2000$ ) were both recorded as a function of time. Spectra were analysed using the Bruker Data Analysis 4.1 software. When given in text, MS results are obtained from an average of all the measurements done within the time of elution of one compound.

Conversions were determined from the integration of the UV trace at 220-400 nm.

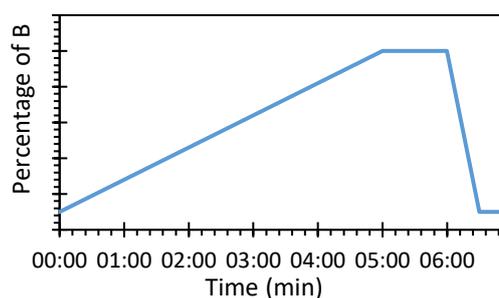


Figure 19: Percentage of solvent B used during LCMS over the duration of analysis.

### 2.6.4 General conditions for Combiflash

Reverse-phase purification of peptides was achieved on a Combiflash NextGen from Teledyne using C18 columns (4.3 gram RediSep Rf Reversed-phase C18 Columns – 69-2203-410 – or 5.5 gram RediSep Gold Rf Reversed-phase C18 Columns – 69-2203-328). Water (solvent A) and acetonitrile (solvent B),

both containing 0.1% trifluoroacetic acid were used as the mobile phase at a flow rate of 15 mL/min. LC traces were measured via UV absorption at 214, and 254 nm with the average measured from 200-300 nm, with a signal gain of x3. The solvent gradient was programmed with B going from 0 to 100% and then a clean at A/B 50%/50% for 1 min. When a peak was detected, the program held the gradient for the length necessary. Every fraction was collected and analysed by LCMS.

### 2.6.5 General conditions for resin cleavage

#### Full scale 0.1 mmol

Peptides were cleaved from the resin in 20 mL of cleavage cocktail (90% TFA, 5% H<sub>2</sub>O, 3% TIPS, 2% DTT for Cys-containing sequences) for 6 hours (18 hours for Arg-containing sequences). After filtration, the resin was washed extensively with dichloromethane (3 × 20 mL) and the filtrate concentrated *in vacuo* to ~5 mL volume. The residue was dropped into diethyl ether stored at -20 °C (~ 80 mL), and the resultant precipitate collected by centrifugation (3000 rpm, 6 min). The residual solid was allowed to air dry for 5 min, then dissolved in deionized water (~ 15 mL) – a few drops of 2 M HCl or methanol could be added to help solubilisation if needed - and then dried by lyophilisation.

#### Small scale from 0.01 to 0.025 mmol

Peptides were cleaved from the resin in 5 mL of cleavage cocktail (90% TFA, 5% H<sub>2</sub>O, 3% TIPS, 2% DTT for Cys-containing sequences) for 2-3 hours. After filtration, the resin was washed with dichloromethane (3 × 5 mL) and the filtrate concentrated *in vacuo* to ~ 0.5 mL volume. The residue was dropped into diethyl ether stored at -20 °C (~ 15 mL), and the resultant precipitate collected by centrifugation (3000 rpm, 6 min). The residual solid was allowed to air dry for 5 min, then dissolved in deionized water (5 mL) – a few drops of 2 M HCl or methanol could be added to help solubilisation if needed - and then dried by lyophilisation.

### 2.6.6 Protocols for on-resin deprotection of protecting groups

#### Fmoc-deprotection

To the resin (0.1 mmol scale), 20% piperidine in DMF (5 mL) was added and the reaction mixture was shaken for 1 h. The resin was filtered on paper and washed with DMF (3 × 5 mL), MeOH (3 × 10 mL) and CH<sub>2</sub>Cl<sub>2</sub> (3 × 5 mL).

Cbz-deprotection (no positive outcome)**Protocol A**

To a stirred solution of Cbz protected amine on resin (0.05 mmol) and 5 % Pd-C (10 mg) in MeOH (2 mL) was added NaBH<sub>4</sub> sodium borohydride (1 eq 2mg, or 10 eq 20 mg ) in portion wise using solid addition funnel or dispenser. The solution was stirred for 1 h before cleavage.

**Protocol B**

To a stirred solution of Cbz protected amine on resin (0.05 mmol) and 5 % Pd-C (10 mg) in MeOH (2 mL) a H<sub>2</sub> balloon was added. The solution was stirred for 2 h before cleavage.

**Protocol C**

To a stirred solution of Cbz protected amine on resin (0.05 mmol) and palladium acetate (1 eq 11.2 mg) in DMF (2 mL) was added ammonium formate (5 eq 16mg in 0.2 ml water). The solution was stirred for 2 h before cleavage.

Alloc-deprotection**Protocol A**

The resin (0.1 mmol) was swelled in chloroform (7 mL), then glacial acetic acid (0.1 mL), *N*-methylmorpholine (0.4 ml) and Pd(PPh<sub>3</sub>)<sub>4</sub> (347 mg, 0.3 mmol) were added to the solution in this specific order. The mixture was left to stir for 4 hours, before being filtered, and washed with CH<sub>2</sub>Cl<sub>2</sub>. The same deprotection procedure was conducted again, to increase the time with the mixture left to stir for another 12 hours.

**Protocol B**

To a stirred solution of Alloc protected amine on resin was added phenylsilane (24 eq) and Pd(PPh<sub>3</sub>)<sub>4</sub> (0.2 eq) in 10 mL CH<sub>2</sub>Cl<sub>2</sub>. The mixture was stirred either at rT or 40 °C for a varying time. The resin was then recovered by filtration washed successively by DMF (10 mL) and CH<sub>2</sub>Cl<sub>2</sub> (3 x 10 mL). The dry resin was stored in freezer before used.

DDE-deprotection

The following protocol is describing how to deprotect DDE on resin at a 0.05 mmol scale. It is important to say that the deprotection solution is not stable and was prepared fresh each time.

NH<sub>2</sub>OH.HCl (83 mg, 1.2 mmol, 24 eq) and imidazole (65 mg, 1.0 mmol, 20 eq) were suspended in *N*-methyl-2-pyrrolidone (1.5 mL), and the mixture was sonicated until complete dissolution. The resin

was then added to this solution which was further diluted with 0.5 mL of CH<sub>2</sub>Cl<sub>2</sub>. The mixture was left to stir for 17 h before being filtered, and washed with MeOH (2 × 10 mL) and CH<sub>2</sub>Cl<sub>2</sub> (2 × 10 mL).

### 2.6.7 Protocols for peptide modification

#### In-solution protocol using thiol-maleimide

A solution of peptide **P** was prepared in PBS (180 μL, 10 mM), to this solution TCEP (20 μL, 500 mM) was added and mixed during 45 min at 800 rpm, rT. Then the maleimide compound was introduced (20 μL, 500 mM for 5 equiv., or 100 mM for 1 equiv.) and the solution mixed again during 45 min at 800 rpm, rT. Aliquots of 50 μL were dissolved in 450 μL PBS for LCMS analysis.

The stock solutions were stored in freezer after each use.

#### On-resin protocol using NHS-activated ester or carbamate

To the resin in solution in DMF (2 mL, for 0.05 mmol scale), 5 eq of NHS-activated carboxylic acid and 20 eq of DIPEA were added. The solution mixture was then stirred for 4-5 h at room temperature. The resin was then recovered by filtration after washing\* with CH<sub>2</sub>Cl<sub>2</sub> (2 × 5 mL) then MeOH (2 × 5 mL) and CH<sub>2</sub>Cl<sub>2</sub> (5 mL). The resin was then either stored in the freezer for further modifications or the peptide was cleaved from the resin with the protocol described in a previous paragraph.

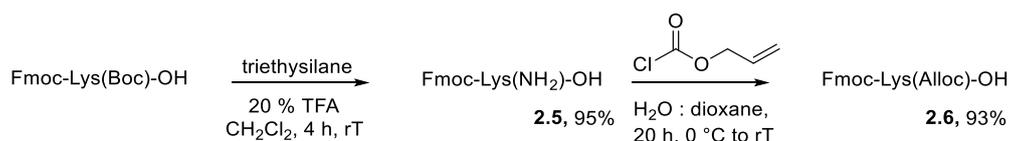
*\*If the NHS-activated carboxylic acid was a fluorophore, an extra first wash with DMF was done (5 mL)*

#### On-resin protocol using in-situ activation of carboxylic acid

To the resin in solution in DMF (2 mL, for 0.05 mmol scale), 5 eq of carboxylic acid and 10 eq of coupling agents HOBt/EDC.HCl were added. The solution mixture was then stirred for 24 h at room temperature. The resin was then recovered by filtration after washing\* with CH<sub>2</sub>Cl<sub>2</sub> (2 × 5 mL) then MeOH (2 × 5 mL) and CH<sub>2</sub>Cl<sub>2</sub> (5 mL). The resin was then either stored in the freezer for further modifications or the peptide was cleaved from the resin with the protocol described in a previous paragraph.

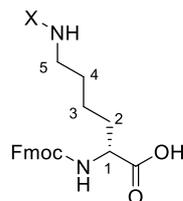
*\*If the carboxylic acid was 5,6-carboxyfluorescein, an extra first wash with DMF was done (5 mL)*

### 2.6.8 Chemical synthesis



Numbering for proton  
and carbon for NMR  
of lysine compounds

x = H or protecting group



### Compound 2.5

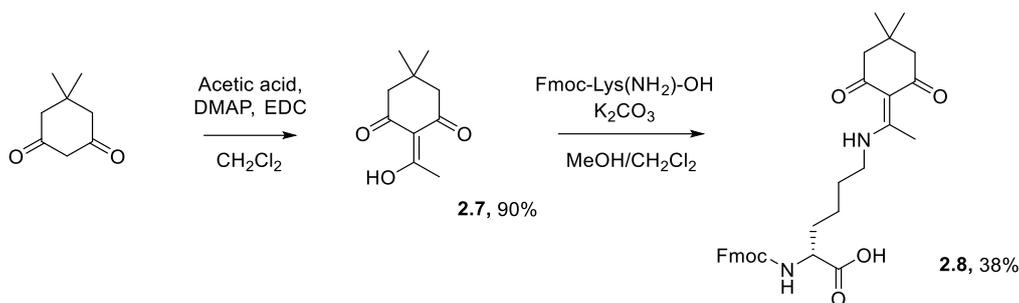
A solution of TFA in  $\text{CH}_2\text{Cl}_2$  (50% v/v, 16 mL) was added to Fmoc-Lys(Boc)-OH (4.68 g, 10 mmol) and then triethylsilane (1.6 mL, 10 mmol) was added to the reaction mixture. The mixture was stirred for 4 h at room temperature. The solvent was then removed under reduced pressure to 5 mL. A solution of sodium hydroxide (1 M) was added dropwise to the residue until pH reached 6. The precipitate was collected by filtration and washed with water ( $2 \times 5$  mL) and dried affording **2.5** as a white solid (3.5 g, 7.7 mmol, 77%). Data were consistent with those previously reported.<sup>29</sup>

**$^1\text{H}$  NMR** (400 MHz, MeOD):  $\delta$  7.77 (d, 2H,  $J = 7.2$  Hz,  $-\text{CH}_{\text{arom}}$ ), 7.64 (t, 2H,  $J = 7.2$  Hz,  $-\text{CH}_{\text{arom}}$ ), 7.36 (dd, 2H,  $J_1 = J_2 = 7.4$  Hz,  $-\text{CH}_{\text{arom}}$ ), 7.28 (dd, 2H,  $J_1 = 7.4$  Hz,  $J_2 = 1.1$  Hz,  $-\text{CH}_{\text{arom}}$ ), 4.41-4.27 (m, 2H,  $-\text{CH}_2(\text{Fmoc})$ ), 4.19 (t, 1H,  $J = 7.0$  Hz,  $-\text{CH}(\text{Fmoc})$ ), 4.10 (m, 1H,  $-\text{CH}(1)$ ), 2.87 (t, 2H,  $J = 7.4$  Hz,  $-\text{CH}_2(5)$ ), 1.90-1.81 (m, 1H,  $-\text{CH}(2)$ ), 1.73-1.58 (m, 3H,  $-\text{CH}(2)$ ,  $-\text{CH}_2(4)$ ), 1.49-1.40 (m, 2H,  $-\text{CH}_2(3)$ ). **HRMS** (ESI<sup>+</sup>):  $m/z$  calcd for  $\text{C}_{21}\text{H}_{25}\text{N}_2\text{O}_4$  369.1809 [M+H]<sup>+</sup>; found: 369.1810

### Compound 2.6

$\text{Na}_2\text{CO}_3$  (2.3 g, 22 mmol) was added at 0 °C to a solution of Fmoc-Lys(NH<sub>2</sub>)-OH **2.5** (1.6 g, 4.3 mmol) in a mixture of  $\text{H}_2\text{O}$  (48 mL) and dioxane (32 mL). A solution of allylchloroformate (0.5 mL, 4.7 mmol) in dioxane (16 mL) was then added over 15 min. The mixture was warmed to room temperature and stirred 17 h, then the dioxane was evaporated under reduced pressure. The aqueous phase was acidified with 1 M HCl to pH 4, then extracted with EtOAc ( $3 \times 50$  mL). The solution was dried with  $\text{MgSO}_4$ , filtered, and concentrated under reduced pressure to afford the product **2.6** as a light brown viscous oil (1.8 g, 4.0 mmol, 93%), which was used without further purification. Data were consistent with those previously reported.<sup>30</sup>

**$^1\text{H}$  NMR** (400 MHz, MeOD):  $\delta$  7.76 (dd, 2H,  $J = 7.5, 5.1$  Hz,  $-\text{CH}_{\text{arom}}$ ), 7.65 (m, 2H,  $-\text{CH}_{\text{arom}}$ ), 7.36 (t, 2H,  $J = 7.5$  Hz,  $-\text{CH}_{\text{arom}}$ ), 7.31-7.25 (m, 2H,  $-\text{CH}_{\text{arom}}$ ), 5.88 (ddd, 1H,  $J = 16.4, 11.0, 5.5$  Hz,  $-\text{CH}=\text{CH}_2$ ), 5.25 (d, 1H,  $J = 16.4$  Hz,  $-\text{CH}=\text{CH}_2$  trans), 5.13 (d, 1H,  $J = 11.0$  Hz,  $-\text{CH}=\text{CH}_2$  cis), 4.50-4.47 (d, 2H,  $J = 5.5$  Hz,  $-\text{CH}_2\text{CH}=\text{CH}_2$ ), 4.36-4.30 (dd, 2H,  $J = 7.0, 2.8$  Hz,  $-\text{CHCH}_2(\text{Fmoc})$ ), 4.20 (t, 1H,  $J = 7.0$  Hz,  $-\text{CHCH}_2(\text{Fmoc})$ ), 4.11 (dd, 1H,  $J = 9.3, 4.5$  Hz,  $-\text{CH}_2(1)$ ), 3.09 (t, 2H,  $J = 6.5$  Hz,  $-\text{CH}_2(5)$ ), 1.88-1.77 (m, 1H,  $-\text{CH}(2)$ ), 1.73-1.64 (m, 1H,  $-\text{CH}_2(2)$ ), 1.52-1.46 (m, 2H,  $-\text{CH}_2(4)$ ), 1.45-1.31 (m, 2H,  $-\text{CH}_2(3)$ ). **HRMS** (ESI<sup>+</sup>):  $m/z$  calcd for  $\text{C}_{21}\text{H}_{25}\text{N}_2\text{O}_4$  453.2020 [M+H]<sup>+</sup>; found: 453.2021



### Compound 2.7

A mixture of dimedone (5.00 g, 35.7 mmol), EDC.HCl (8.0 g, 41.7 mmol), 4-dimethylaminopyridine (DMAP) (5.1 g, 41.8 mmol), and acetic acid (2.5 mL, 43.8 mmol) in  $\text{CH}_2\text{Cl}_2$  (50 mL) was stirred at room temperature for 4 h. Then,  $\text{H}_2\text{O}$  (150 mL) was added, and the aqueous phase was extracted by  $\text{CH}_2\text{Cl}_2$  ( $3 \times 100$  mL). The combined organic layers were washed by 1 M HCl (100 mL) then by brine ( $2 \times 100$  mL), dried over anhydrous  $\text{MgSO}_4$  filtered and concentrated under reduced pressure to afford DDE-OH **2.7** as a yellow oil (1.6 g, 31.3 mmol, 90%). Data were consistent with those previously reported.<sup>31</sup>

**<sup>1</sup>H NMR** (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  2.59 (s, 3H,  $-\text{COHCH}_3$ ), 2.52 (d,  $J = 0.8$  Hz, 2H,  $-\text{CH}_2$ ), 2.38-2.32 (d,  $J = 0.8$  Hz, 2H,  $-\text{CH}_2$ ), 1.06 (s, 6H,  $-\text{C}(\text{CH}_3)_2$ ). **HRMS** (ESI<sup>+</sup>):  $m/z$  calcd for  $\text{C}_{10}\text{H}_{15}\text{O}_3$ : 183.1016 [M+H]<sup>+</sup>; found: 183.1017

### Compound 2.8

Fmoc-Lys(NH<sub>2</sub>)-OH **2.5** (1.2 g, 3.26 mmol) was dispersed in MeOH (10 mL) containing  $\text{K}_2\text{CO}_3$  (0.48 g, 3.47 mmol). Then, a solution of DDE **2.7** (0.9 g, 4.94 mmol) in  $\text{CH}_2\text{Cl}_2$  (10 mL) was added slowly over 10 min. The reaction was stirred for 18 h at room temperature. After concentration under vacuum, 1 M HCl (30 mL) was added to the white paste, and the aqueous phase was extracted with EtOAc ( $3 \times 50$  mL). The organic phases were combined, washed with  $\text{H}_2\text{O}$  (50 mL) and Brine (50 mL) before being dried on  $\text{Na}_2\text{SO}_4$ . After filtration, the solvent was removed under reduced pressure. The crude product was purified by flash chromatography using a gradient 1-10% MeOH in  $\text{CH}_2\text{Cl}_2$ . Pure fractions were concentrated under reduced pressure to give Fmoc-Lys(DDE)-OH **2.8** as a white solid (0.62 g, 1.16 mmol, 38%). Data were consistent with those previously reported.<sup>31</sup>

**<sup>1</sup>H NMR** (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  7.75 (d,  $J = 7.6$  Hz, 2H,  $-\text{CH}_{\text{arom}}$ ), 7.59 (t,  $J = 7.9$  Hz, 2H,  $-\text{CH}_{\text{arom}}$ ), 7.39 (t,  $J = 7.4$  Hz, 2H,  $-\text{CH}_{\text{arom}}$ ), 7.30 (ddd,  $J = 7.5, 2.6, 1.2$  Hz, 2H,  $-\text{CH}_{\text{arom}}$ ), 5.77 (d,  $J = 8.1$  Hz, 1H,  $-\text{NH}_{\text{DDE}}$ ), 4.47 (q,  $J = 7.0$  Hz, 1H,  $-\text{CH}(1)$ ), 4.37 (d,  $J = 7.3$  Hz, 2H,  $-\text{CH}_2(\text{Fmoc})$ ), 4.20 (t,  $J = 7.1$  Hz, 1H,  $-\text{CH}(\text{Fmoc})$ ), 3.46-3.34 (m, 2H,  $-\text{CH}_2(5)$ ), 2.55 (s, 3H,  $-\text{CH}_3_{\text{DDE}}$ ), 2.36 (m, 4H,  $-\text{CH}_2_{\text{DDE}}$ ), 1.97 (d,  $J = 10.9$  Hz, 1H,  $-\text{CH}(2)$ ), 1.87-1.64 (m, 3H,  $-\text{CH}(2)$ ,  $-\text{CH}_2(4)$ ), 1.55 (m, 2H,  $-\text{CH}_2(3)$ ), 1.01 (s, 6H,  $-\text{C}(\text{CH}_3)_2$ ). **HRMS** (ESI<sup>+</sup>):  $m/z$  calcd for  $\text{C}_{31}\text{H}_{37}\text{N}_2\text{O}_6$ : 533.2646 [M+H]<sup>+</sup>; found: 533.2660

## 2.6.9 List of the core peptides synthesized by SPPS present in this Chapter

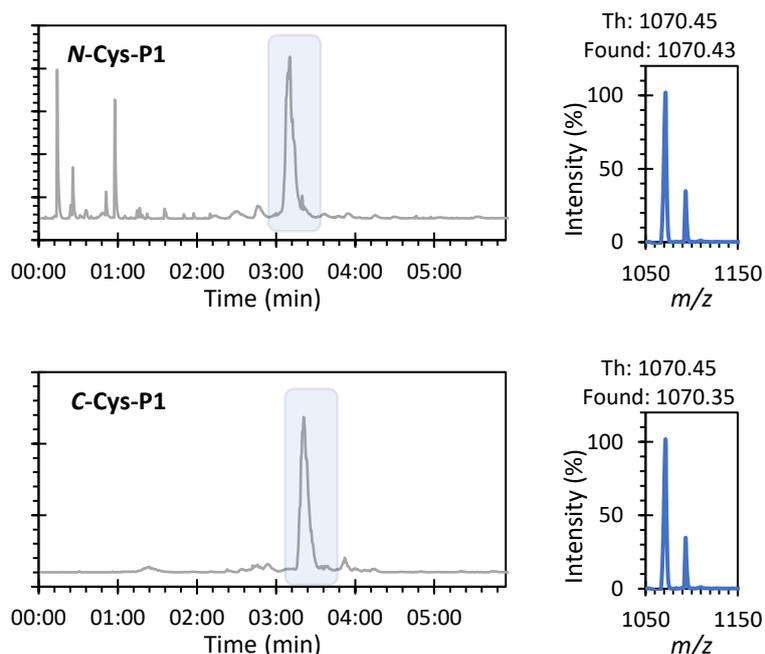
As a reminder, all the peptides are synthesised by SPPS on Rink amide resin leading to an amide terminus. If not specified no protecting are present leaving the amino acid side chain free.

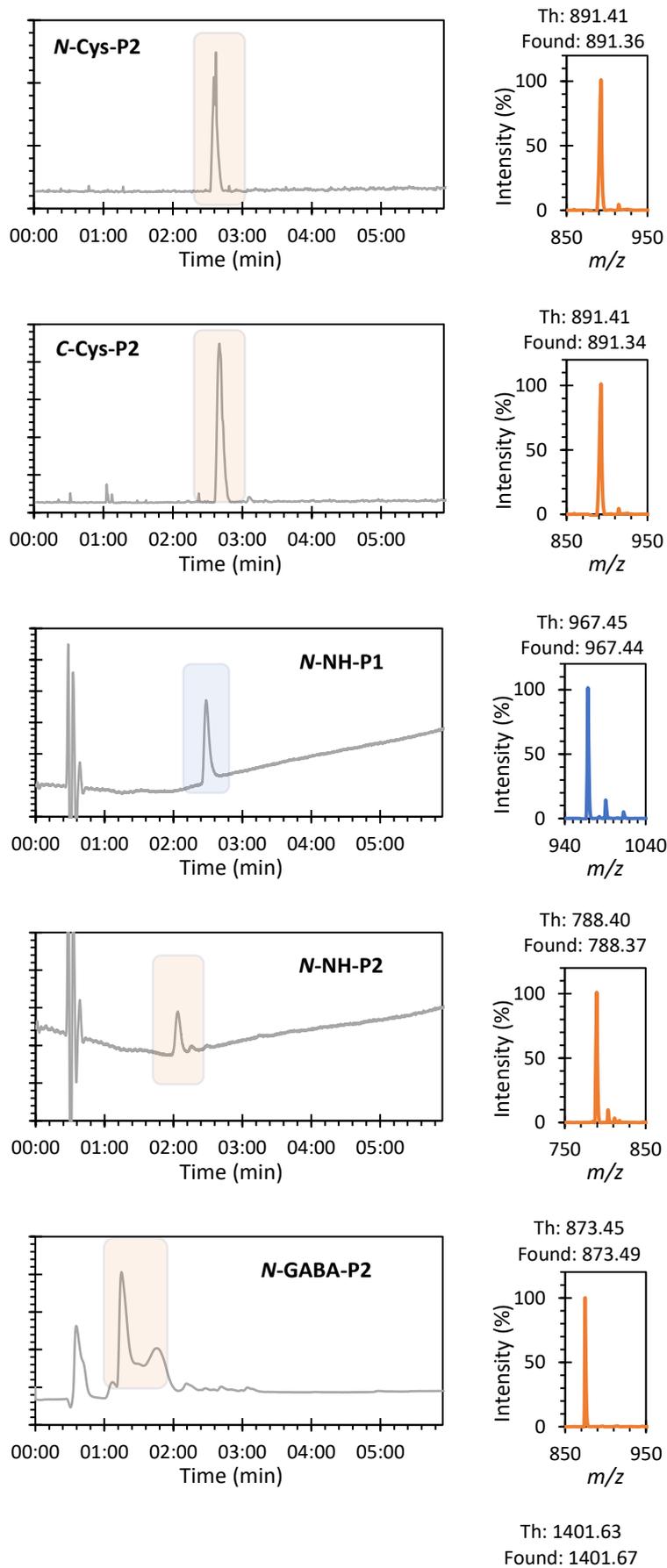
Name in text	Peptide sequence	Yield (%)
<i>N</i> -Cys-P1	CHWWWPAS	65 mg, 61%
<i>C</i> -Cys-P1	HWWWPASC	181 mg, 31%
<i>N</i> -Cys-P2	CRGFFYT	60 mg, 67%
<i>C</i> -Cys-P2	RGFFYTC	75 mg, 84%
<i>N</i> -NH-P1	HWWWPAS	NA <sup>x</sup>
<i>N</i> -NH-P2	RGFFYT	NA <sup>x</sup>
<i>N</i> -GABA-P2	GABA-RGFFYT	NA <sup>x</sup>
Fmoc-P2-K(Alloc)	Fmoc-HWWWPAS-K(Alloc)	NA <sup>x</sup>
Fmoc-P4-K(Alloc)	Fmoc-CVEEAS-K(Alloc)	NA <sup>x</sup>
Fmoc-P3-K(cbz) <sup>†</sup>	Fmoc-LVEALYL-K(cbz)	NA <sup>x</sup>
Fmoc-P2-K(DDE) <sup>†</sup>	Fmoc-RGFFYT-K(DDE)	NA <sup>x</sup>
Fmoc-P3-K(DDE)	Fmoc-LVEALYL-K(DDE)	NA <sup>x</sup>
Fmoc-P4-K(DDE)	Fmoc-CVEEAS-K(DDE)	NA <sup>x</sup>

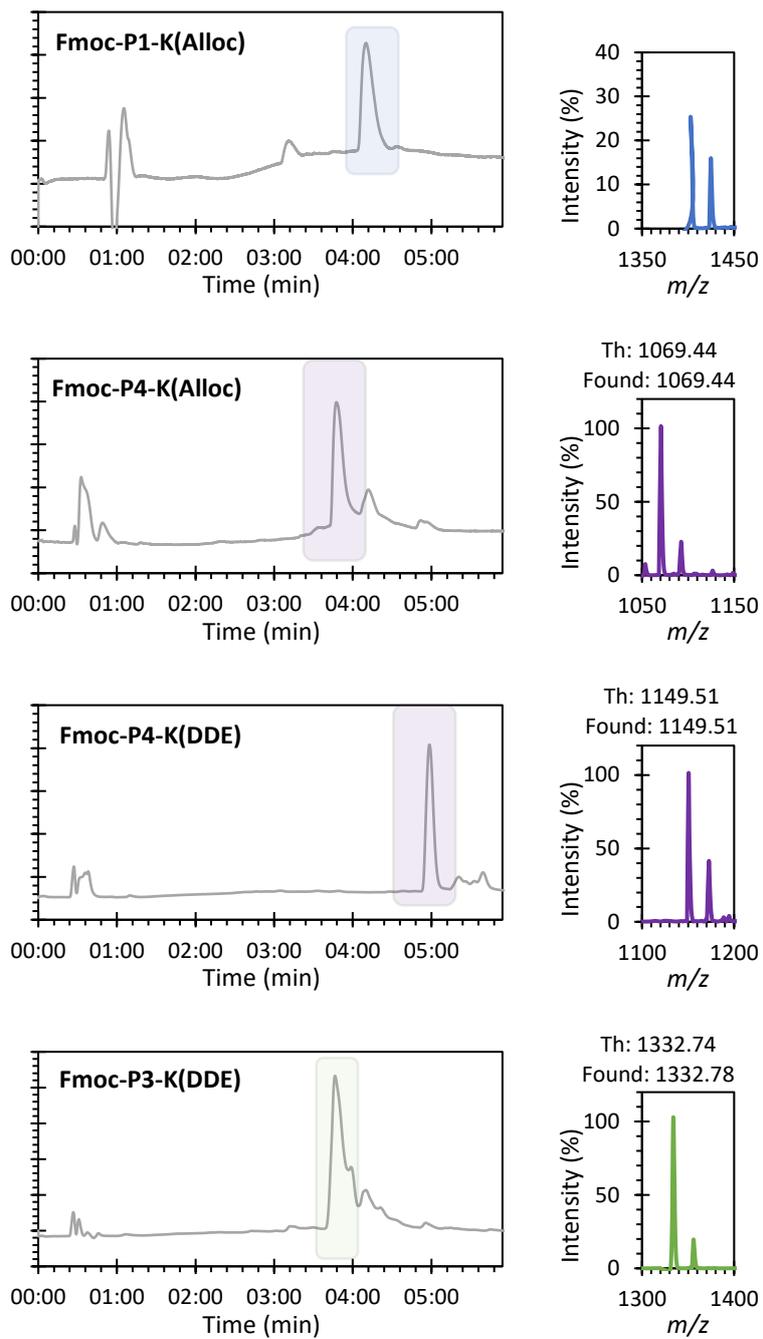
<sup>x</sup>Yield non-applicable due to modification on resin with resin split between different reactions

<sup>†</sup>Graphs already in main text

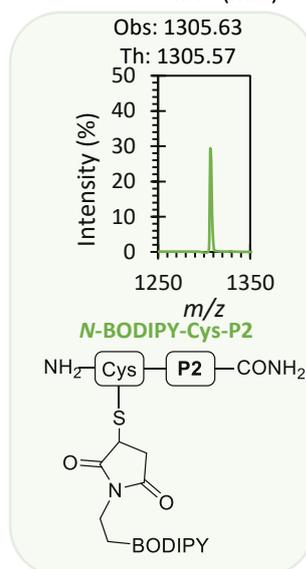
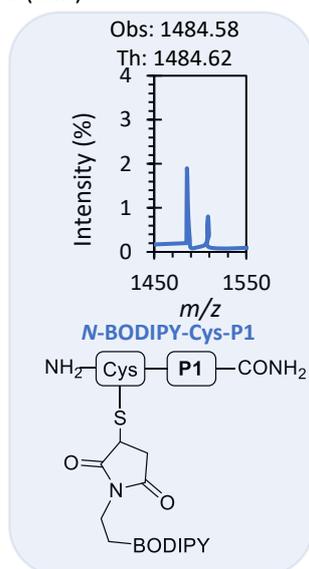
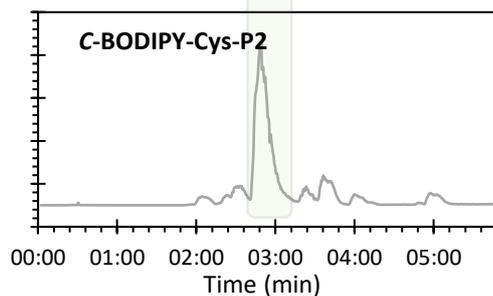
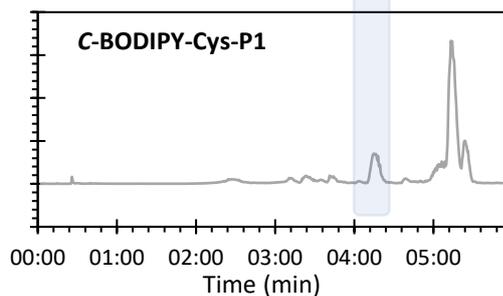
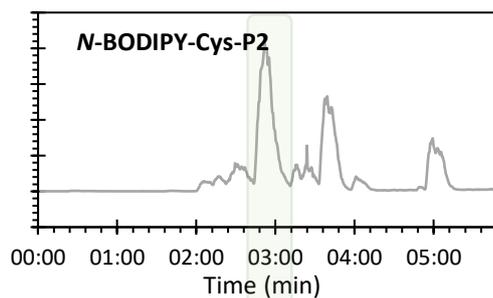
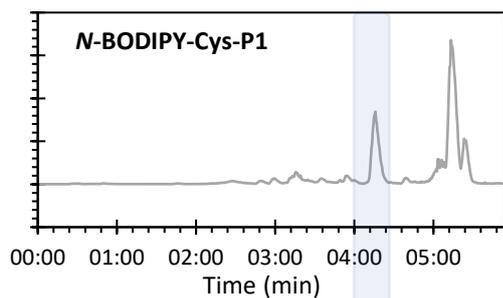
## 2.6.10 LCMS graphs of the naïve peptides synthesized by SPPS present in this Chapter

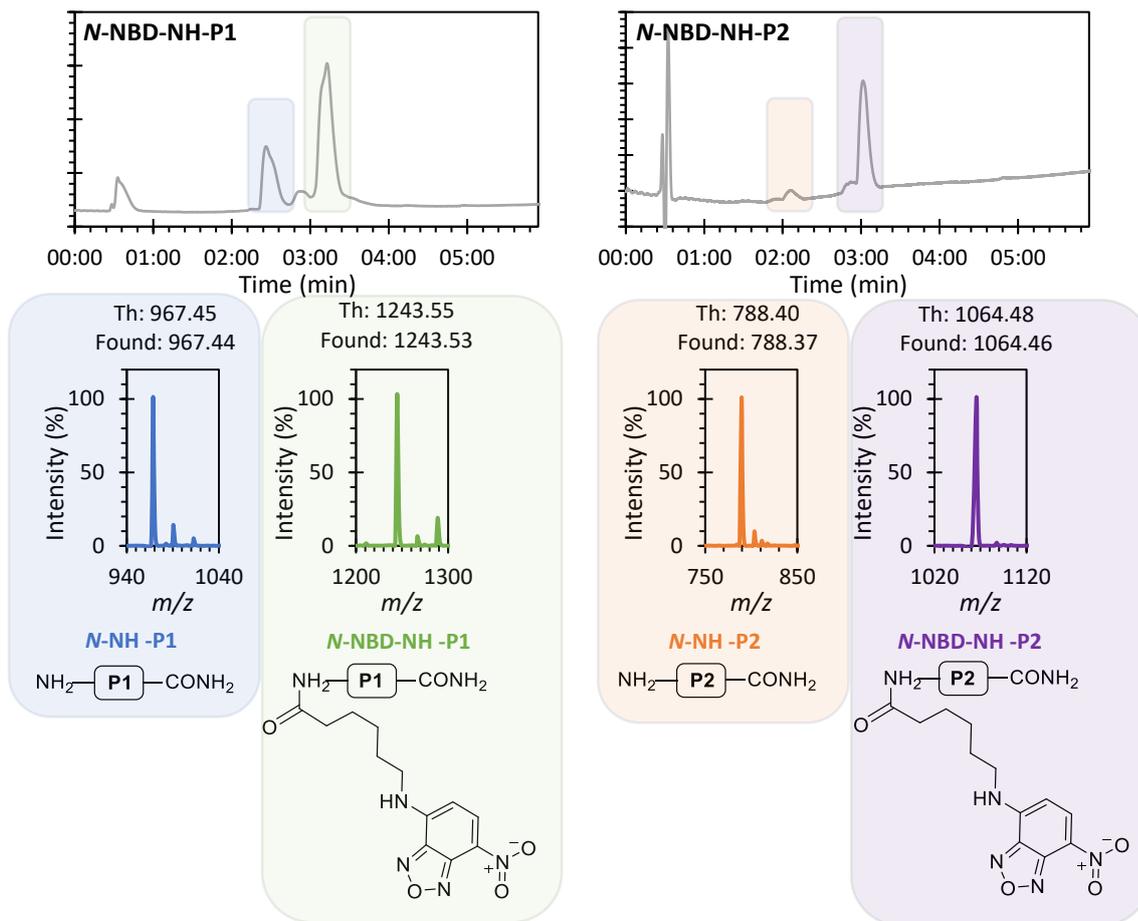






## 2.6.11 LCMS graphs of the modified peptides

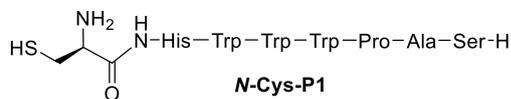




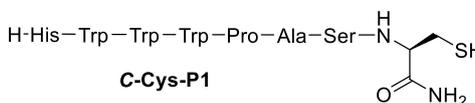
## 2.6.12 Structures of the peptides mentioned in this Chapter

## Cysteine modification

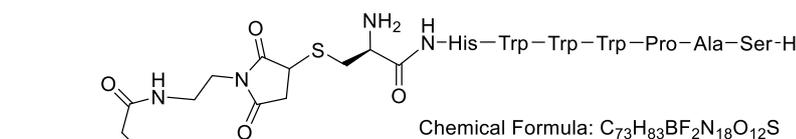
P1 : HWWWPAS



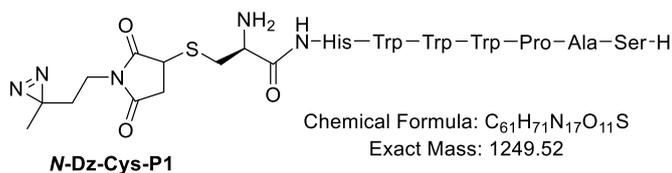
Chemical Formula:  $C_{53}H_{62}N_{14}O_9S$   
Exact Mass: 1070.45



Chemical Formula:  $C_{53}H_{62}N_{14}O_9S$   
Exact Mass: 1070.45

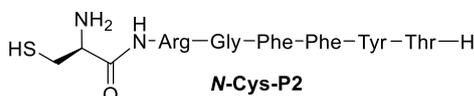


Chemical Formula:  $C_{73}H_{83}BF_2N_{18}O_{12}S$   
Exact Mass: 1484.62

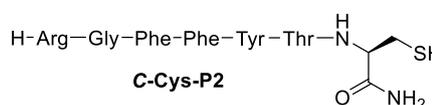


Chemical Formula:  $C_{61}H_{71}N_{17}O_{11}S$   
Exact Mass: 1249.52

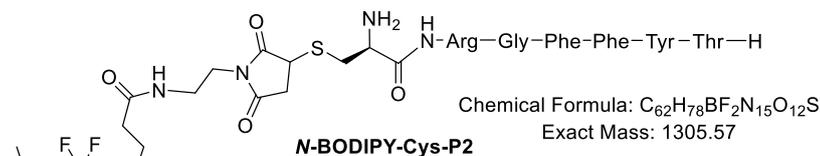
P2 : RGFFYT



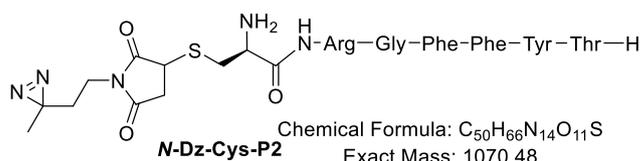
Chemical Formula:  $C_{42}H_{57}N_{11}O_9S$   
Exact Mass: 891.41



Chemical Formula:  $C_{42}H_{57}N_{11}O_9S$   
Exact Mass: 891.41



Chemical Formula:  $C_{62}H_{78}BF_2N_{15}O_{12}S$   
Exact Mass: 1305.57



Chemical Formula:  $C_{50}H_{66}N_{14}O_{11}S$   
Exact Mass: 1070.48

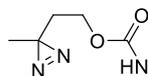
C-BODIPY-Cys-P1 and P2 as well as C-Dz-Cys-P1 and P2 are not represented due to the same chemical formula and mass due to the symmetry of those peptides between N and C termini.

**N-terminus modification via amidation coupling**

P1 : HWWWPAS

**H<sub>2</sub>N-His-Trp-Trp-Trp-Pro-Ala-Ser-H**  
**N-NH-P1**

Chemical Formula: C<sub>50</sub>H<sub>57</sub>N<sub>13</sub>O<sub>8</sub>  
Exact Mass: 967.45



HN-His-Trp-Trp-Trp-Pro-Ala-Ser-H

Chemical Formula: C<sub>55</sub>H<sub>63</sub>N<sub>15</sub>O<sub>10</sub>  
Exact Mass: 1093.49

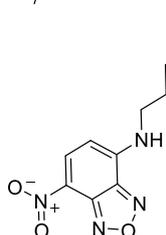
**N-Dz-NH-P1**

H-N-His-Trp-Trp-Trp-Pro-Ala-Ser-H

Chemical Formula: C<sub>64</sub>H<sub>70</sub>BF<sub>2</sub>N<sub>15</sub>O<sub>9</sub>  
Exact Mass: 1241.55

**N-BODIPY-NH-P1**

Chemical Formula: C<sub>62</sub>H<sub>69</sub>N<sub>17</sub>O<sub>12</sub>  
Exact Mass: 1243.53



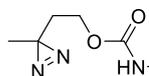
N-His-Trp-Trp-Trp-Pro-Ala-Ser-H

**N-NBD-NH-P1**

P2 : RGGFFYT

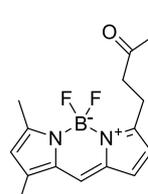
**H<sub>2</sub>N-Arg-Gly-Phe-Phe-Tyr-Thr-H**  
**N-NH-P2**

Chemical Formula: C<sub>39</sub>H<sub>52</sub>N<sub>10</sub>O<sub>8</sub>  
Exact Mass: 788.40



HN-Arg-Gly-Phe-Phe-Tyr-Thr-H

Chemical Formula: C<sub>44</sub>H<sub>58</sub>N<sub>12</sub>O<sub>10</sub>  
Exact Mass: 914.44

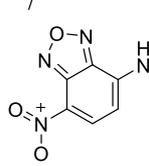
**N-Dz-NH-P2**

H-N-Arg-Gly-Phe-Phe-Tyr-Thr-H

Chemical Formula: C<sub>53</sub>H<sub>65</sub>BF<sub>2</sub>N<sub>12</sub>O<sub>9</sub>  
Exact Mass: 1062.51

**N-BODIPY-NH-P2**

Chemical Formula: C<sub>51</sub>H<sub>64</sub>N<sub>14</sub>O<sub>12</sub>  
Exact Mass: 1064.48



N-Arg-Gly-Phe-Phe-Tyr-Thr-H

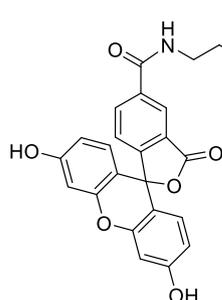
**N-NBD-NH-P2**

Chemical Formula: C<sub>46</sub>H<sub>63</sub>BrN<sub>10</sub>O<sub>9</sub>  
Exact Mass: 978.40

Br-N-Arg-Gly-Phe-Phe-Tyr-Thr-H

**N-BrC7-NH-P2**

Chemical Formula: C<sub>64</sub>H<sub>69</sub>N<sub>11</sub>O<sub>15</sub>  
Exact Mass: 1231.50



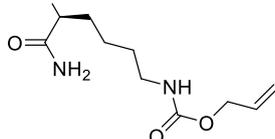
N-Arg-Gly-Phe-Phe-Tyr-Thr-H

**N-5,6Fluo-NH-P2**

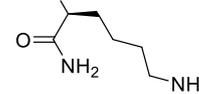
## C-terminus modification via amidation coupling

## P1 : HWWWPAS

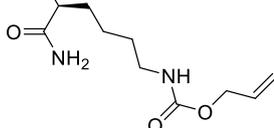
Fmoc-His-Trp-Trp-Trp-Pro-Ala-Ser-NH

**Fmoc-P1-K(Alloc)**Chemical Formula:  $C_{75}H_{83}N_{15}O_{13}$   
Exact Mass: 1401.63

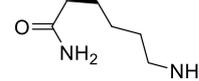
Fmoc-His-Trp-Trp-Trp-Pro-Ala-Ser-NH

**Fmoc-P1-K(NH<sub>2</sub>)**Chemical Formula:  $C_{71}H_{79}N_{15}O_{11}$   
Exact Mass: 1317.61

H-His-Trp-Trp-Trp-Pro-Ala-Ser-NH

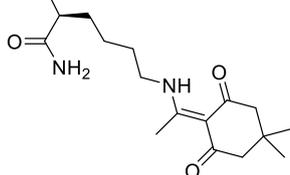
**NH<sub>2</sub>-P1-K(Alloc)**Chemical Formula:  $C_{60}H_{73}N_{15}O_{11}$   
Exact Mass: 1179.56

H-His-Trp-Trp-Trp-Pro-Ala-Ser-NH

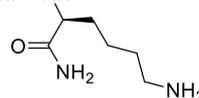
**NH<sub>2</sub>-P1-K(NH<sub>2</sub>)**Chemical Formula:  $C_{56}H_{69}N_{15}O_9$   
Exact Mass: 1095.54

## P2 : RGFFYT

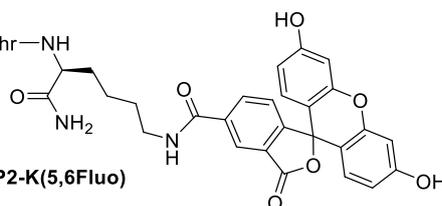
Fmoc-Arg-Gly-Phe-Phe-Tyr-Thr-NH

**Fmoc-P2-K(DDE)**Chemical Formula:  $C_{70}H_{86}N_{12}O_{13}$   
Exact Mass: 1302.64

Fmoc-Arg-Gly-Phe-Phe-Tyr-Thr-NH

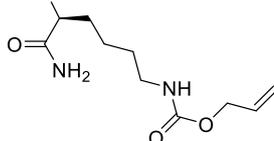
**Fmoc-P2-K(NH<sub>2</sub>)**Chemical Formula:  $C_{60}H_{74}N_{12}O_{11}$   
Exact Mass: 1138.56

H-Arg-Gly-Phe-Phe-Tyr-Thr-NH

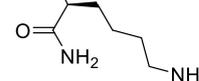
Chemical Formula:  $C_{66}H_{74}N_{12}O_{15}$   
Exact Mass: 1274.54**NH<sub>2</sub>-P2-K(5,6Fluo)**

## P4 : CVEEAS

Fmoc-Cys-Val-Glu-Glu-Ala-Ser-NH

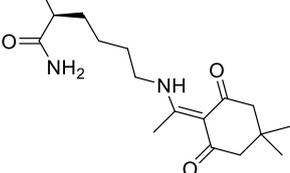
**Fmoc-P4-K(Alloc)**Chemical Formula:  $C_{49}H_{67}N_9O_{16}S$   
Exact Mass: 1069.44

Fmoc-Cys-Val-Glu-Glu-Ala-Ser-NH

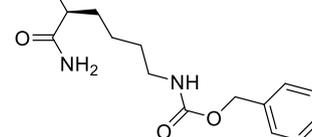
**Fmoc-P4-K(NH<sub>2</sub>)**Chemical Formula:  $C_{45}H_{63}N_9O_{14}S$   
Exact Mass: 985.42

## P3 : LVEALYL

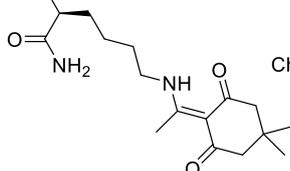
Fmoc-Cys-Val-Glu-Glu-Ala-Ser-NH

**Fmoc-P4-K(DDE)**Chemical Formula:  $C_{55}H_{75}N_9O_{16}S$   
Exact Mass: 1149.51

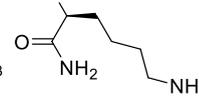
Fmoc-Leu-Val-Glu-Ala-Leu-Tyr-Leu-NH

**Fmoc-P3-K(Cbz)**Chemical Formula:  $C_{69}H_{94}N_{10}O_{15}$   
Exact Mass: 1302.69

Fmoc-Leu-Val-Glu-Ala-Leu-Tyr-Leu-NH

**Fmoc-P3-K(DDE)**Chemical Formula:  $C_{71}H_{100}N_{10}O_{15}$   
Exact Mass: 1332.74

Fmoc-Leu-Val-Glu-Ala-Leu-Tyr-Leu-NH

**Fmoc-P3-K(NH<sub>2</sub>)**Chemical Formula:  $C_{61}H_{88}N_{10}O_{13}$   
Exact Mass: 1168.65

## 2.7 Bibliography

- (1) Jensen, K. J. *Solid-Phase Peptide Synthesis: An Introduction*; **2013**; Vol. 1047. [https://doi.org/10.1007/978-1-62703-544-6\\_1](https://doi.org/10.1007/978-1-62703-544-6_1).
- (2) Aapptec. Practical Synthesis Guide to Solid Phase Peptide Chemistry. *Aapptec, Synth. Notes* **2011**, 1–76.
- (3) Pedersen, S. .; Tofteng, A. P.; Malik, L.; Jensen, K. J. Microwave Heating in Solid-Phase Peptide Synthesis. *Chem. Soc. Rev.* **2012**, *41*, 1826–1844. <https://doi.org/10.1039/c1cs15214a>.
- (4) Kreutzer, A. G.; Salveson, P. J. *Standard Practices for Fmoc-Based Solid-Phase Peptide Synthesis in the Nowick Laboratory*; 2015.
- (5) Yu, H. Q.; Dong, X. Y.; Sun, Y. Affinity Chromatography of Insulin with a Heptapeptide Ligand Selected from Phage Display Library. *Chromatographia* **2004**, *60* (7–8), 379–383. <https://doi.org/10.1365/s10337-004-0407-5>.
- (6) Chiang, H. L.; Ngo, S. T.; Chen, C. J.; Hu, C. K.; Li, M. S. Oligomerization of Peptides LVEALYL and RGFFYT and Their Binding Affinity to Insulin. *PLoS One* **2013**, *8* (6). <https://doi.org/10.1371/journal.pone.0065358>.
- (7) Knutson, V. P. Insulin-Binding Peptide. *J. Biol. Chem.* **1988**, *263* (28), 14146–14151. [https://doi.org/10.1016/S0021-9258\(18\)68197-9](https://doi.org/10.1016/S0021-9258(18)68197-9).
- (8) Wu, X.; Yan, Q.; Huang, Y.; Huang, H.; Su, Z.; Xiao, J.; Zeng, Y.; Wang, Y.; Nie, C.; Yang, Y.; Li, X. Isolation of a Novel Basic FGF-Binding Peptide with Potent Antiangiogenic Activity. *J. Cell. Mol. Med.* **2010**, *14* (1–2), 351–356. <https://doi.org/10.1111/j.1582-4934.2008.00506.x>.
- (9) Jankovic, B.; Ruf, J.; Zanobini, C.; Bozovic, O.; Buhrke, D.; Hamm, P. Sequence of Events during Peptide Unbinding from RNase S: A Complete Experimental Description. *J. Phys. Chem. Lett.* **2021**, *12* (21), 5201–5207. <https://doi.org/10.1021/acs.jpcllett.1c01155>.
- (10) Gunnoo, S. B.; Madder, A. Chemical Protein Modification through Cysteine. *ChemBioChem* **2016**, *17* (7), 529–553. <https://doi.org/10.1002/cbic.201500667>.
- (11) Rawale, D. G.; Thakur, K.; Adusumalli, S. R.; Rai, V. Chemical Methods for Selective Labeling of Proteins. *European J. Org. Chem.* **2019**, *2019* (40), 6749–6763. <https://doi.org/10.1002/ejoc.201900801>.
- (12) Shadish, J. A.; DeForest, C. A. Site-Selective Protein Modification: From Functionalized Proteins to Functional Biomaterials. *Matter* **2020**, *2* (1), 50–77. <https://doi.org/10.1016/j.matt.2019.11.011>.
- (13) Hayashi, T.; Yasueda, Y.; Tamura, T.; Takaoka, Y.; Hamachi, I. Analysis of Cell-Surface Receptor Dynamics through Covalent Labeling by Catalyst-Tethered Antibody. *J. Am. Chem. Soc.* **2015**, *137* (16), 5372–5380. <https://doi.org/10.1021/jacs.5b02867>.
- (14) Kantner, T.; Watts, A. G. Characterization of Reactions between Water-Soluble Trialkylphosphines and Thiol Alkylating Reagents: Implications for Protein-Conjugation Reactions. *Bioconjug. Chem.* **2016**, *27* (10), 2400–2406. <https://doi.org/10.1021/acs.bioconjchem.6b00375>.
- (15) Santos, A. S.; Silva, A. M. S.; Marques, M. M. B. Sustainable Amidation Reactions – Recent Advances. *European J. Org. Chem.* **2020**, *2020* (17), 2501–2516. <https://doi.org/10.1002/ejoc.202000106>.
- (16) Massolo, E.; Pirola, M.; Benaglia, M. Amide Bond Formation Strategies: Latest Advances on a Dateless Transformation. *European J. Org. Chem.* **2020**, *2020* (30), 4641–4651. <https://doi.org/10.1002/ejoc.202000080>.

- (17) Valeur, E.; Bradley, M. Amide Bond Formation: Beyond the Myth of Coupling Reagents. *Chem. Soc. Rev.* **2009**, *38* (2), 606–631. <https://doi.org/10.1039/b701677h>.
- (18) Pattabiraman, V. R.; Bode, J. W. Rethinking Amide Bond Synthesis. *Nature* **2011**, *480* (7378), 471–479. <https://doi.org/10.1038/nature10702>.
- (19) Tamura, T.; Song, Z.; Amaike, K.; Lee, S.; Yin, S.; Kiyonaka, S.; Hamachi, I. Affinity-Guided Oxime Chemistry for Selective Protein Acylation in Live Tissue Systems. *J. Am. Chem. Soc.* **2017**, *139* (40), 14181–14191. <https://doi.org/10.1021/jacs.7b07339>.
- (20) Ma, Y.; Liu, Y.; Wang, J.; Chen, X.; Yin, H.; Chi, Q.; Jia, S.; Du, S.; Qi, Y.; Wang, K. DIC/Oxyma Based Efficient Synthesis and Activity Evaluation of Spider Peptide Toxin GsMTx4. *Chinese J. Org. Chem.* **2022**, *42* (2), 498–506. <https://doi.org/10.6023/cjoc202109003>.
- (21) Díaz-Mochón, J. J.; Bialy, L.; Bradley, M. Full Orthogonality between Dde and Fmoc: The Direct Synthesis of PNA-Peptide Conjugates. *Org. Lett.* **2004**, *6* (7), 1127–1129. <https://doi.org/10.1021/ol049905y>.
- (22) 6 Substituted 2-Heterocyclamino Pyrazine Compounds as CHK-1 Inhibitors. WO2010016005, 2010.
- (23) Sultane, P. R.; Mete, T. B.; Bhat, R. G. A Convenient Protocol for the Deprotection of N-Benzyloxycarbonyl (Cbz) and Benzyl Ester Groups. *Tetrahedron Lett.* **2015**, *56* (16), 2067–2070. <https://doi.org/10.1016/j.tetlet.2015.02.131>.
- (24) Thieriet, N.; Alsina, J.; Giralt, E.; Guibé, F.; Albericio, F. Use of Alloc-Amino Acids in Solid-Phase Peptide Synthesis. Tandem Deprotection-Coupling Reactions Using Neutral Conditions. *Tetrahedron Lett.* **1997**, *38* (41), 7275–7278. [https://doi.org/10.1016/S0040-4039\(97\)01690-0](https://doi.org/10.1016/S0040-4039(97)01690-0).
- (25) Schmidt, J.; Garambois, V.; Rocheblave, L.; Martinez, J.; Pèlerin, A.; Cavelier, F.; Vivès, E. Cyclization of Peptides through a Urea Bond: Application to the Arg-Gly-Asp Tripeptide. *ChemBioChem* **2010**, *11* (8), 1083–1092. <https://doi.org/10.1002/cbic.201000062>.
- (26) Roos, E. C.; Bernabé, P.; Hiemstra, H.; Speckamp, W. N.; Kaptein, B.; Boesten, W. H. J. Palladium-Catalyzed Transprotection of Allyloxycarbonyl-Protected Amines: Efficient One-Pot Formation of Amides and Dipeptides. *J. Org. Chem.* **1995**, *60* (6), 1733–1740. <https://doi.org/10.1021/jo00111a035>.
- (27) Rohwedder, B.; Mutti, Y.; Dumy, P.; Mutter, M. Hydrazinolysis of Dde: Complete Orthogonality with Alloc Protecting Groups. *Tetrahedron Lett.* **1998**, *39* (10), 1175–1178. [https://doi.org/10.1016/S0040-4039\(97\)10810-3](https://doi.org/10.1016/S0040-4039(97)10810-3).
- (28) ITERS, L. E.; Chhabra, S. R.; Hothi, B.; Evans, D. J.; Whitea, P. D.; Bycroft, B. W.; Chan, W. C. An Appraisal of New Variants of Dde Amine Protecting Group for Solid Phase Peptide Synthesis. *Tetrahedron Lett.* **1998**, *39* (12), 1603–1606.
- (29) de Andrade, P.; Mantoani, S. P.; Gonçalves Nunes, P. S.; Magadán, C. R.; Pérez, C.; Xavier, D. J.; Hojo, E. T. S.; Campillo, N. E.; Martínez, A.; Carvalho, I. Highly Potent and Selective Aryl-1,2,3-Triazolyl Benzylpiperidine Inhibitors toward Butyrylcholinesterase in Alzheimer's Disease. *Bioorganic Med. Chem.* **2019**, *27* (6), 931–943. <https://doi.org/10.1016/j.bmc.2018.12.030>.
- (30) Crivici, A.; Lajoie, G. A Procedure for the Large Scale Preparation of Nε-Alloc-Lysine and Nε-Alloc-Nα-Fmoc-Lysine. *Synth. Commun.* **1993**, *23* (1), 49–53. <https://doi.org/10.1080/00397919308020400>.
- (31) Li, J.; Guan, D.; Chen, F.; Shi, W.; Lan, L.; Huang, W. Total and Semisyntheses of Polymyxin Analogues with 2-Thr or 10-Thr Modifications to Decipher the Structure-Activity Relationship

and Improve the Antibacterial Activity. *J. Med. Chem.* **2021**, *64* (9), 5746–5765.  
<https://doi.org/10.1021/acs.jmedchem.0c02217>.

## Chapter 3 – Approaches to study peptide-protein binding

The visualization and monitoring of proteins without disturbing their biological function is a major challenge in chemical biology.<sup>1</sup> This is especially true when studying protein/protein or peptide/protein interactions.<sup>2,3</sup> Indeed, these interactions are key to a wide range of biological signalling pathways and in these systems. It was thus essential that after the production of the peptides introduced in Chapter 2, we verified that the modified sequences were still able to bind their target protein. Different techniques were used to study these interactions. Depending on the technique, different peptide modifications were required. To perform QCM-D (**Q**uartz **C**rystal **M**icrobalance with **D**issipation monitoring) extra cysteines were integrated into the sequences, while fluorophores were inserted to undertake FP (**F**luorescence **P**olarization), and diazirines were installed for photo-affinity labelling. QCM-D and FP provided understanding of the strength of the binding while the diazine photo-affinity labelling strategy could also add information about the localisation of the modification. In this Chapter 3, the peptides synthesised and modified in Chapter 2 were studied via these techniques.

### 3.1 Quartz Crystal Microbalance with Dissipation monitoring (QCM-D) as a first approach to monitor binding

#### 3.1.1. Underlying mechanism

QCM-D is a sensitive technique that measures mass changes at the surface of a quartz receptor with nanogram resolution. The technique works through the excitation of a quartz piezoelectric sensor via the application of an alternating voltage. Two parameters are measured, the resonance frequency and the dissipation. The resulting resonance frequency,  $f$ , is linked to the mass of the surface,  $m$ , through the Sauerbrey equation, presented in Figure 1, which also correlates inherently to the dissipation of the loss of energy.<sup>4,5</sup> Thus depending on the mass present at a certain time, the resonance frequency and the voltage vary. Both dissipation and resonance frequency are monitored as a function of time. The frequency changes,  $\Delta f$ , are calculated and give information on changes in mass, which occur due to the interaction between the molecule studied and the surface, see Figure 1.

Different applications of this technique can be used either with rigid layers in a gas phase, or in liquid phase, or using a viscoelastic layer in liquid phase. This solution-phase approach was used in our studies for the analysis of peptide-protein binding.<sup>4,5</sup>

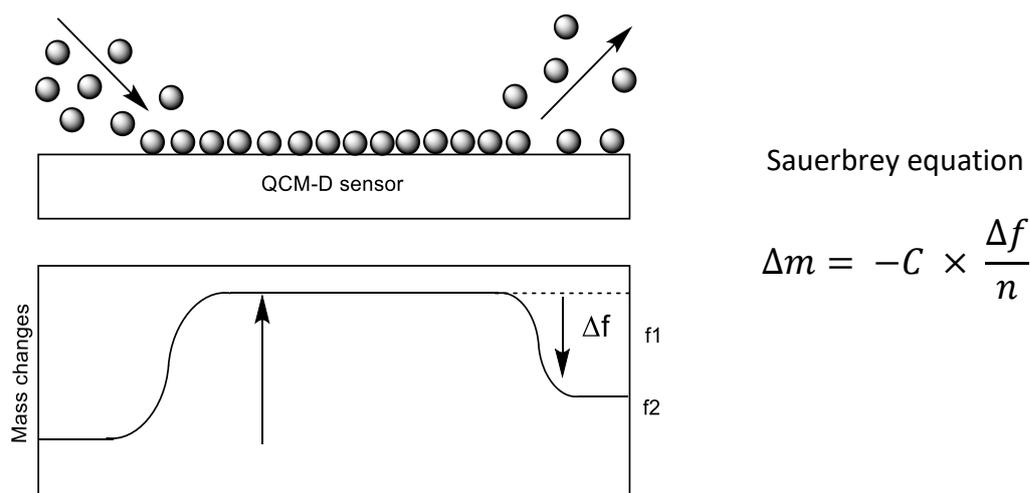


Figure 1: A schematic illustration of mass uptake and mass loss, as molecules adsorb to, and desorb from, the QCM-D sensor surface adapted from [What is QCM-D? \(biolinscientific.com\)](http://biolinscientific.com). Sauerbrey equation  $\Delta m$  change of mass,  $C$  constant depending on the quartz sensor,  $\Delta f$  change of frequency,  $n$  overtone number.

During this PhD, QCM-D was used to determine the binding between a peptide and its target protein. As such, the surface was functionalised with the peptides and subsequently protein solution was added and the  $\Delta f$  was measured to confirm binding. Figure 2 illustrates the changes expected between the different steps.

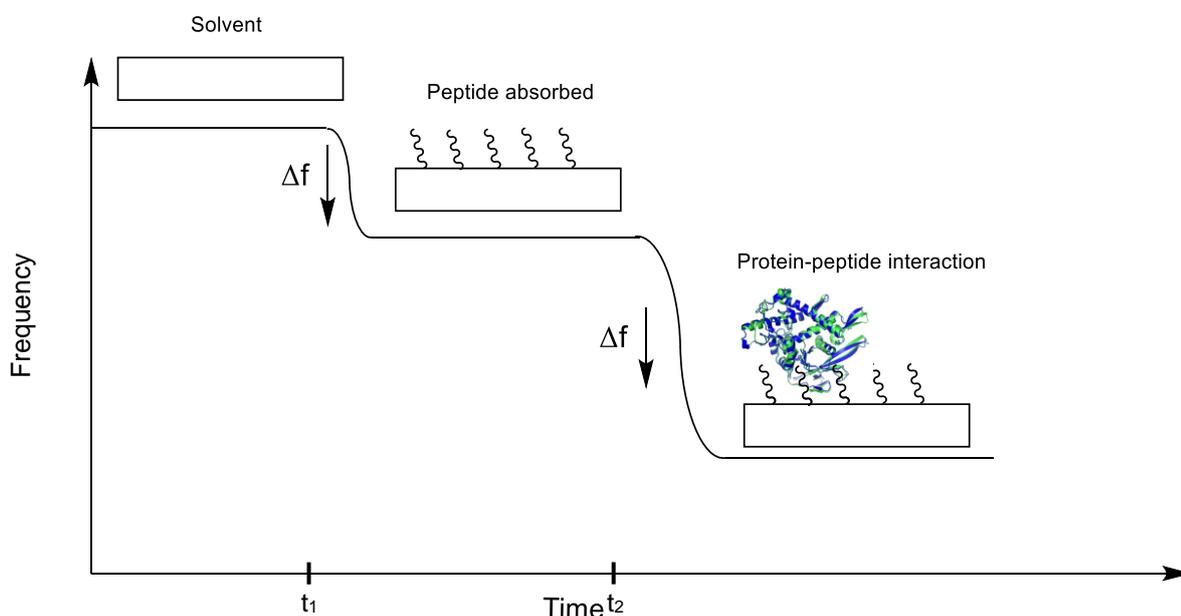


Figure 2: A schematic illustration of the absorption of peptide on a quartz sensor to study protein/peptide interactions over time.  $t_1$  addition of peptide solution onto the sensor,  $t_2$  addition of a protein solution

### 3.1.2. The choice of sensor

In order to attach peptides on to a quartz sensor, two possibilities were investigated either using gold or silicon dioxide surfaces. While gold surfaces are able to interact directly with thiols, making the absorption of a peptide relatively straightforward with the only requirement being the addition of a cysteine residue into the peptide, see Figure 4, silicon dioxide surfaces need more preparation steps.

To utilise the same peptides, the surface is first functionalised with a maleimide layer which can then react with the cysteine of the peptide. The conditions to insert this first solid layer of maleimide moieties are explained in more detail in the Experimental Section, while the final set-up is represented in Figure 4. As seen in this last Figure, two types of linkers were introduced before the maleimide layer, the reasons and results obtained will be described in following paragraph.

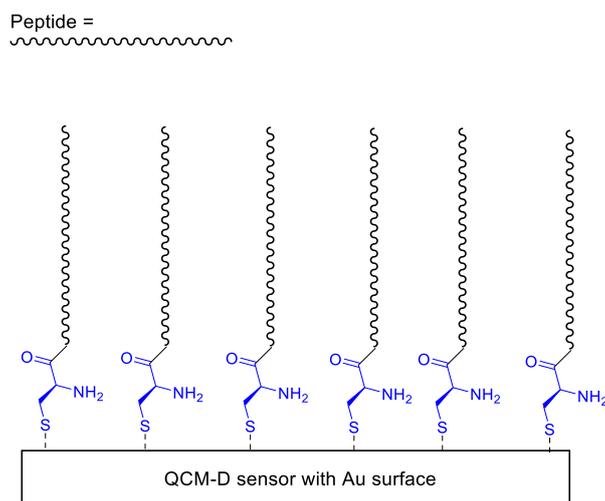


Figure 3: A schematic illustration of the installation of cysteine modified peptide layer on to a gold surface sensor.

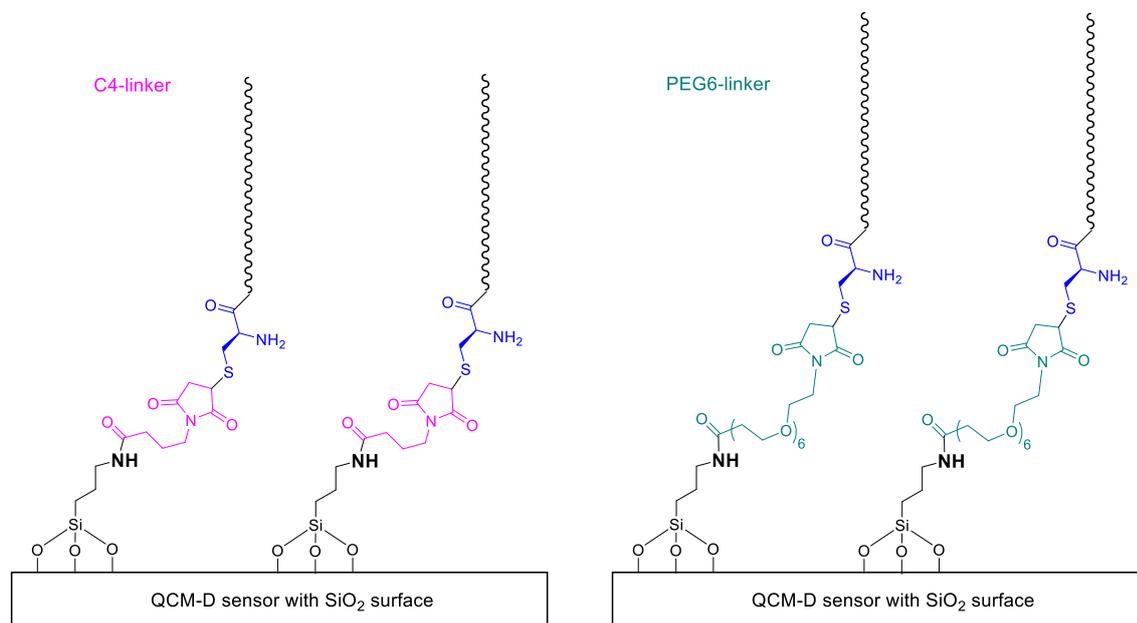


Figure 4: A schematic illustration of the installation of cysteine modified peptide (in blue) layer onto a silicon dioxide SiO<sub>2</sub> sensor via two different linkers. The surface is first modified with 3-aminopropyltriethoxysilane (APTES) leaving a layer of amine which then reacted with two types of carboxylic acid: left: 4-maleimidobutyric acid N-succinimidyl ester (= C4 linker in pink); Right: Mal-PEG<sub>6</sub>-NHS ester (= PEG<sub>6</sub> linker in green).

A third set-up was also investigated. Instead of using APTES to functionalise the silicon surface with amines, mercaptopropyltriethoxysilane (MPTES) was used providing a thiol-functionalised surface, which could then be reacted with a Mal-PEG<sub>6</sub>-NHS ester, leaving an NHS-ester cap able to react with

nucleophiles within the studied peptide sequences. It was hypothesised that at pH 7.4 the free primary amine of the GABA inserted at the *N*-terminus would be the strongest nucleophile present when no lysine was present in the peptide sequence and would be the primary site of peptide attachment through an amide linkage, Figure 5.<sup>6-8</sup> However, this approach would not be as selective as using cysteine.

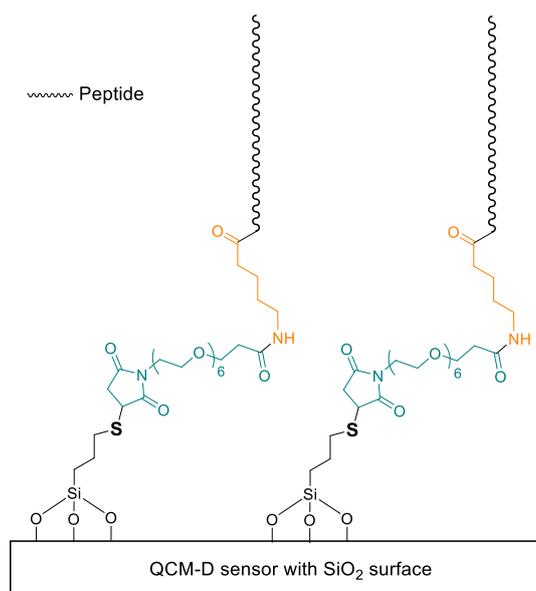


Figure 5: A schematic illustration of the installation of free amine *N*-terminus peptide (in orange) layer onto a silicon dioxide SiO<sub>2</sub> sensor. The surface is first modified with mercaptopropyltriethoxysilane (MPTES) leaving a layer of thiol which then react with Mal-PEG6-NHS ester (in green).

### 3.1.3. Au surface functionalisation

In order to investigate the binding of peptides to their target protein, gold surfaces were first chosen as they require less preparation steps. Functionalisation of these surfaces required only the synthesis of peptides with an extra cysteine at their *N* or *C*-termini (the exception would be when the known binding peptide sequence already contains a cysteine, but as mentioned in Chapter 2, this is not common). In order to ensure protein binding was specific to the peptide, rather than non-specific binding to the surface, an extra step was added after the installation of the peptide on the surface. A small thiol-PEG-OMe was synthesised (compound **3.1**, Figure 6) and was introduced to saturate the gold surface, forming a dense, compact, and inert monolayer. The synthesis was straight forward with only two steps from a commercial PEG with a tosylation step followed by the transformation to a thiol.

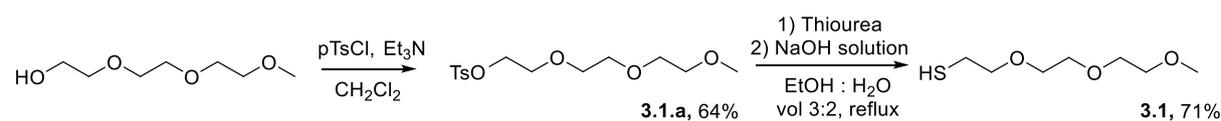


Figure 6: Synthesis of a small PEG-thiol blocker

Before protein binding studies were undertaken, the concentration of the peptide solution, the introduction of blocker **3.1** and its respective concentration, and the range of protein concentrations needed were optimised. While these results are not discussed in this thesis, for clarity, optimisation data is presented in the Annex of this thesis. The results for two peptides sequences for insulin are presented as well as the work on FGF-2.

### 3.1.3.a Peptide P2, RGFFYT, insulin binder

The first peptide sequence tested was peptide P2, RGFFYT, which we proposed would bind to insulin, modified at either the *N* and-*C* termini by a cysteine. A negative control with only blocker **3.1** and no peptide was also ran at the same time to ensure that the data obtained was a result of the interaction between peptide and protein.

The results are shown in Figure 7, with the time of different additions of solutions represented by black arrows, at  $t_1$ , the addition of the peptide solution (10  $\mu\text{M}$ ), at  $t_2$ , the addition of blocker **3.1**; at  $t_3$  a washing step to release any species not bound to the surface and then from  $t_4$  to the end the step wise addition of increasing equivalents of insulin. Notable differences in frequency were observed for the highest concentrations of insulin, marked for the *C*-terminal cysteine sequence as a blue capped linker.

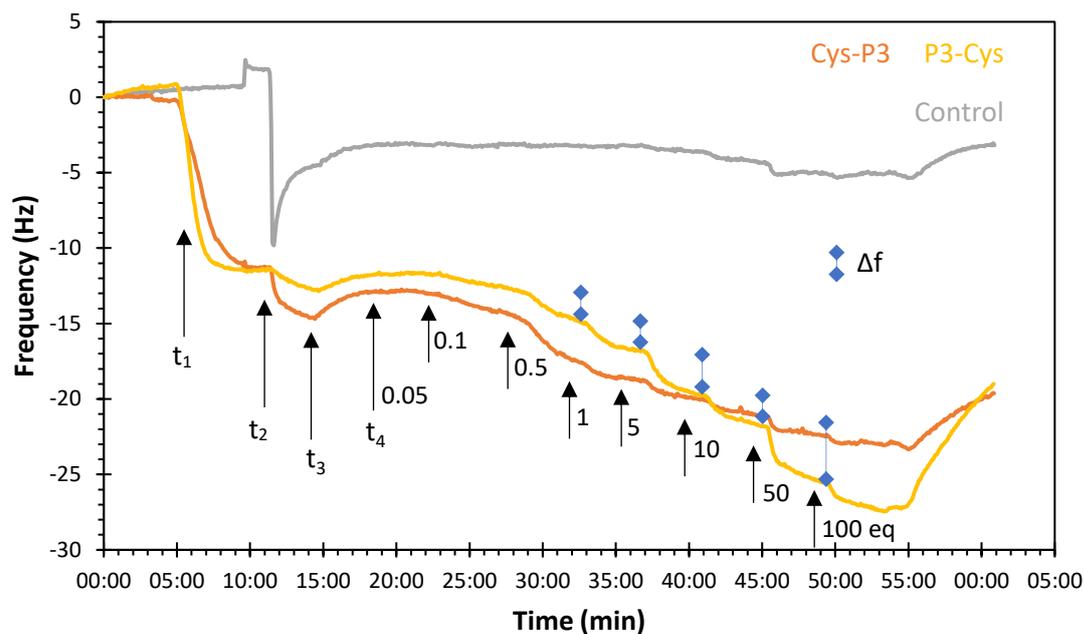


Figure 7: QCM-D curves obtained for the binding assay between peptide RGFYF with cysteine at its N-terminus (orange) or C-terminus (yellow) and insulin on a gold surface and the control with no peptide (grey).  $t_1$  addition of peptide ( $10 \mu\text{M}$  in PBS),  $t_2$  addition of compound **3.1** to prevent non-specific binding,  $t_3$  washing step with PBS,  $t_4$  addition of more and more concentrated protein concentration (0.05, 0.1, 0.5, 1, 5, 10, 50,  $100 \mu\text{M}$ )

The addition of the peptide at  $t_1$  on the surface showed a clear shift in frequency (-12 Hz) compared to the control (in grey) after stabilisation, indicating successful surface functionalisation. The blocker **3.1** was added at  $t_2$  and after a short time, around 2 min, an equilibrium was reached. A following buffer wash was done, during which the frequency went back up showing some desorption from  $t_3$  to  $t_4$ . While the frequency dropped to -4 Hz for the control between  $t_1$  to  $t_4$ , showing the saturation of the surface with compound **3.1**, a smaller shift is observed for the sensors with peptide (-3 Hz for the N-terminus and -0.5 Hz for the C-terminus functionalised peptides) showing the importance of adding the blocker to remove any available surface for the protein to bind to. The key step was then the addition of a solution of insulin from  $t_4$  until the end at different concentrations (pictured by black arrows in Figure 7) 0.05, 0.1, 0.5, 1, 5, 10, 50,  $100 \mu\text{M}$ . The results showed a decrease in frequency only with the peptide-functionalised sensors. While the control showed a small amount of biofouling at the highest concentrations of insulin (50 and  $100 \mu\text{M}$ ) with a shift of 1-2 Hz, the total shift for the N-terminally modified peptide was around 10 Hz and up to 17 Hz for the C-terminally modified. From the change of frequencies at different concentrations, represented by the blue capped lines in Figure 7, binding data were obtained and are represented in Figure 8. The data processing and the obtained values can both be found in the Experimental section. A sum of squares (SSR) showing the difference between the fitting model and the points obtained was calculated and is present in Figure 8.

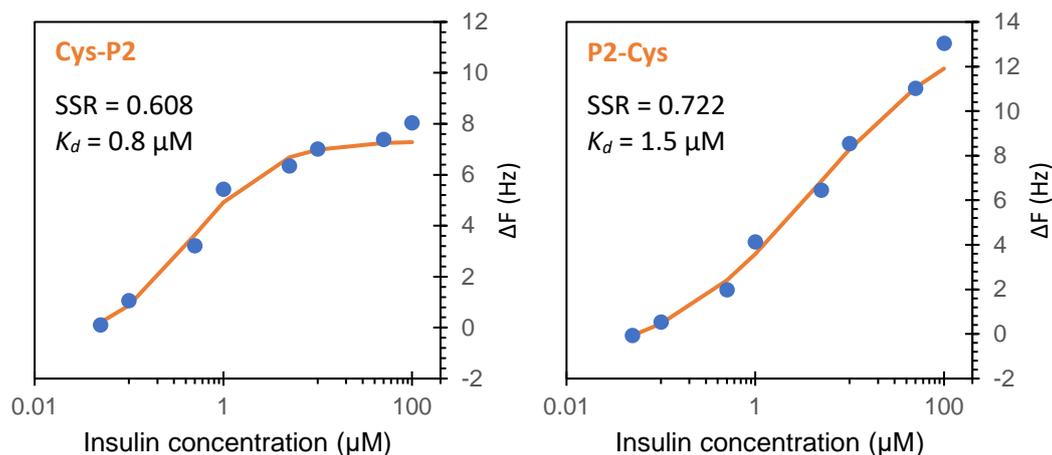


Figure 8: Binding curves for peptides Cys-P2,  $K_d$  obtained 0.8  $\mu\text{M}$  (left), and P2-Cys,  $K_d$  obtained 1.5  $\mu\text{M}$  (right) to insulin

While the  $K_d$  was obtained from the full set of data for the *N*-terminally modified peptide Cys-P2, it seemed like higher insulin concentrations could have been tested for the *C*-terminally modified peptide P2-Cys to reach the end of the curve. Thus, when still calculated as binding was observed, the presented  $K_d$  is likely to be in a range of error due to missing data point. However, increasing insulin concentration could also impact the biofouling onto the surface itself limiting this methodology for binding assay. The article used for those sequences only gave binding energy (-5.9 kcal/mol) which made comparison difficult.<sup>9</sup>

### 3.1.3.b Peptide P3, LVEALYL, insulin binder

The second sequence for insulin binding was then studied in the same way. Frequency plots for the binding of peptide P3 LVEALYL can be found in Figure 9, with the extrapolated binding curves represented in Figure 10. Similarly to P2, this sequence showed strong binding affinity to insulin with  $K_d$  of 2.0  $\mu\text{M}$  for both modified sequences. In similar manner to P2-Cys, extra data points at higher concentration could have allowed us to observe the highest plateau of the curve. However, it can already be seen that the two last points do not fit the theoretical curve as well which could be due to the increase in biofouling happening at greater concentrations of insulin. Again, this inaccuracy at high protein concentrations was a limiting factor of the QCM-D technology.

As a conclusion of these first tests, the two peptides immobilised at their *N* or *C* termini, which we hypothesise would work similar as future modification of either these termini, showed strong binding to the protein in the  $\mu\text{M}$   $K_d$  range.

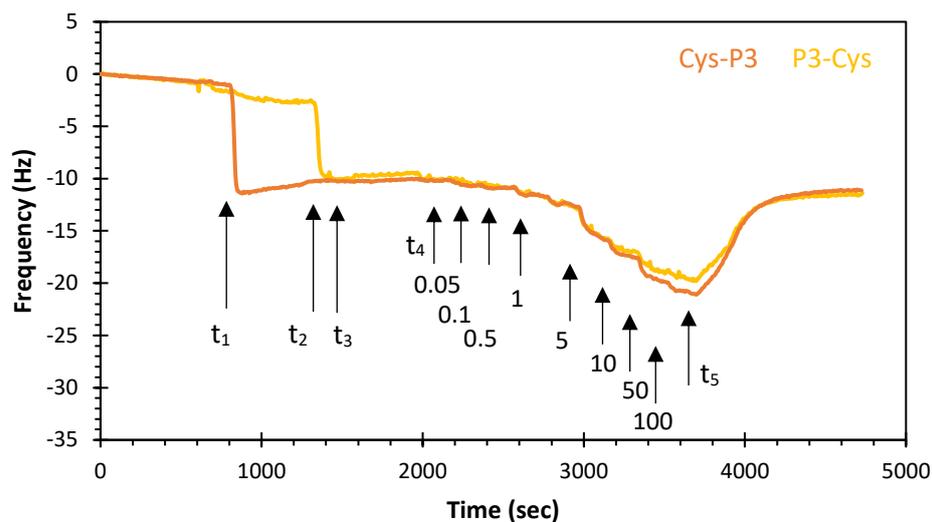


Figure 9: QCM-D curves obtained for the binding assay between peptide P3 LVEALYL with cysteine at its N-terminus (orange) or C-terminus (yellow) and insulin on a gold surface.  $t_1$  addition of peptide ( $10 \mu\text{M}$  in PBS),  $t_2$  addition of compound **3.1** to prevent non-specific binding,  $t_3$  washing step with PBS,  $t_4$  addition of more and more concentrated protein concentration ( $0.05$ ,  $0.1$ ,  $0.5$ ,  $1$ ,  $5$ ,  $10$ ,  $50$ ,  $100 \mu\text{M}$ ),  $t_5$  final washing with PBS

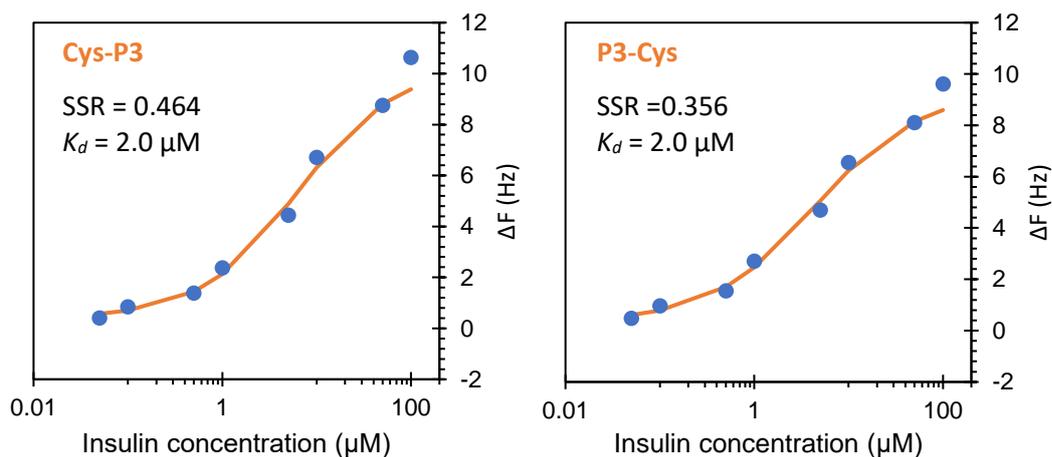


Figure 10: Binding curves for peptide Cys-P3,  $K_d$  obtained  $2.0 \mu\text{M}$  (right), and P3-Cys,  $K_d$  obtained  $2.0 \mu\text{M}$  (left) to insulin

### 3.1.3.c Another model protein RNase A

To verify the successful measurement of binding, another model protein, RNase-A, was also studied prior to working with the expensive and challenging FGF-2. This protein was picked as it was easily accessible and a known binding peptide had been reported in the literature.<sup>10,11</sup> This peptide was identified after putting RNase-A in presence of subtilisin which can cleave a single peptide bond yielding the RNase S complex composed of the S-peptide (residues 1–20) and the S-protein (residues 21–124).<sup>10,11</sup> The known binding peptide, the S-peptide, KETAAKFERQHMSSTSA, will be referred to P8. RNase A has a higher molecular weight than insulin and shows more potential for protease digestion which will be discussed later on. An additional cysteine was introduced at both termini and

QCM-D data were obtained for both peptides. RNase-A showed to have no affinity with control surfaces as shown in Figure 11, with small changes in frequency on protein addition being quickly reversible upon washing. When the peptides are immobilised two different profiles could be seen: for Cys-P8, the curve was constant until the addition of 50 and 100 equivalents of RNase A, indicating poor binding. On the contrary, peptide P8-Cys showed a slow increase of  $\Delta f$  upon RNase A addition.

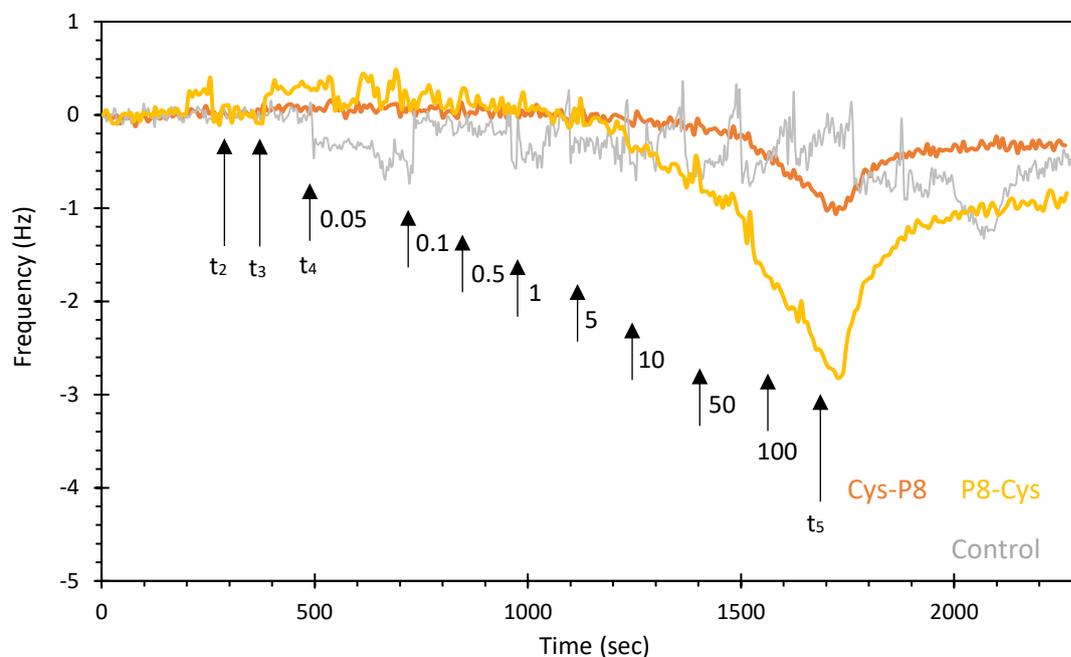


Figure 11: QCM-D curves obtained for the binding assay between peptide P8 modified with cysteine at its N-terminus (orange) or C-terminus (yellow) and RNase-A on a gold surface and the control with no peptide (grey).  $t_1$  addition of peptide ( $10 \mu\text{M}$  in PBS),  $t_2$  addition of compound **3.1** to prevent non-specific binding,  $t_3$  washing step with PBS,  $t_4$  addition of more and more concentrated protein concentration (0.05, 0.1, 0.5, 1, 5, 10, 50, 100  $\mu\text{M}$ ),  $t_5$  final washing step with PBS

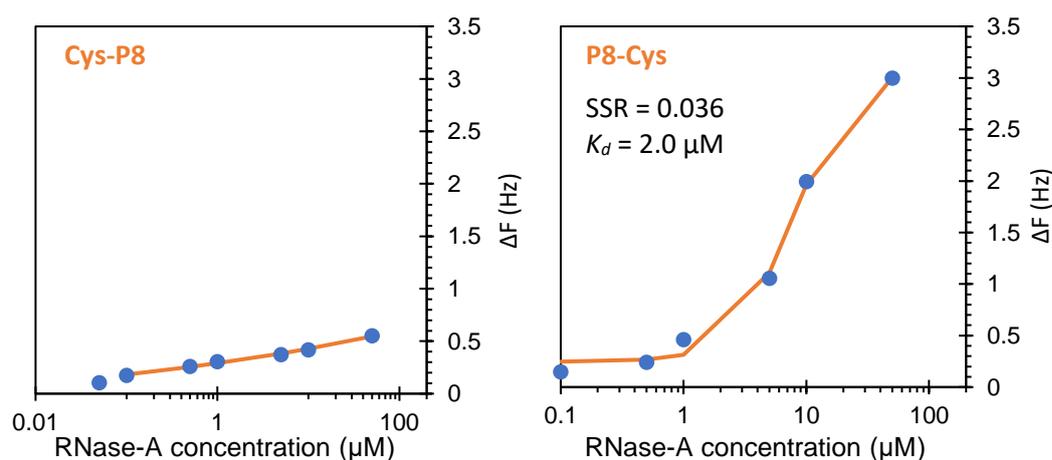


Figure 12: Binding curves for peptide Cys-P8 (right), and P8-Cys,  $K_d$  obtained  $10.0 \mu\text{M}$  (left) to RNase-A

After extrapolation, the binding curves were obtained, Figure 12, showing that only the C-terminal modified peptide was still able to bind to RNase-A. Although some binding of N-terminal modified peptide was observed at the very highest protein concentrations, these concentrations are impractical

and likely non-specific. These results demonstrate the importance of validating that peptide modification does not disrupt target binding.

#### 3.1.3.d An attempt with FGF-2

After optimisation of the QCM-D technique with insulin and RNase A, we moved onto the study of FGF-2 with its respective binding peptides which were analysed in similar way. However, since FGF-2 (£371 for 50 µg) is ~ x1000 times more expensive than insulin (£683 for 100 mg), and due to the volume of protein required at each step (around 1 mL per data point), it was decided that the concentration of peptides and protein used would be decreased by a factor of ten from a range of 0.05-100 µM, to 0.005-10 µM for the protein.

The protein used within this thesis was provided through a collaboration with Dr Marko Hyvönen at the Cambridge Stem Cell Institute and was the zebrafish version of FGF-2. This analogue has very similar structure (more than 110 amino residues in common on a 155 amino acid sequence, > 70%) and activity to the human version and can be used to activate human cells. Human FGF-2 is patented making its use in many settings challenging, thus the zebrafish analogue sharing same biological activity was a suitable choice for this thesis.

A first negative control was performed to ensure that FGF-2 would not biofoul the surface. As shown in Figure 13, it could be seen that even after the addition of low concentrations from 10 nM of FGF-2 that the frequency was decreasing on control surfaces, with a shift of -2 Hz, and that this shift increased to 7-8 Hz for a concentration of 100 nM FGF-2, showing high biofouling toward the surface despite the presence of the blocker **3.1** inserted at  $t_1$ . Gold surfaces using blocker **3.1** therefore seemed to not be suitable for studying FGF-2 as biofouling of the surface was observed rapidly

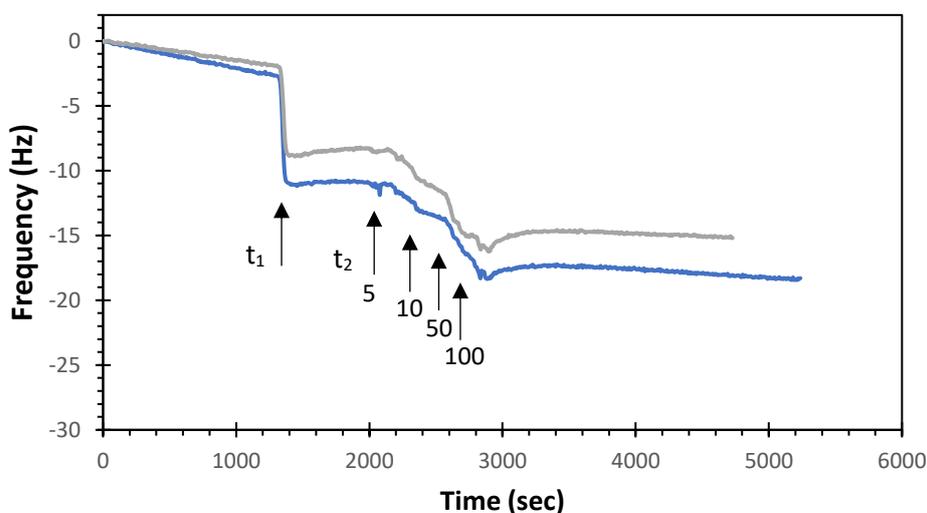


Figure 13: QCM-D curves obtained for the negative control assay on gold surface of FGF-2. Addition of compound **3.1** at  $t_1$  and then at  $t_2$  addition of more and more concentrated protein concentration (5, 10, 50, and 100 nM)

### 3.1.4. SiO<sub>2</sub> surface, different optimisations

#### 3.1.4.a An attempt with an NHS-layer

An alternative approach to gold sensors was the use of SiO<sub>2</sub> surface. Compared to the work on gold surfaces, this type of surface needed extra functionalisation steps. As a first attempt, a system using MPTES as shown in Figure 5 was carried out. This approach to peptide immobilisation is based on an amidation reaction by reacting a surface layer of activated NHS-ester and a free amine from the peptide. While for the *N*-terminus the primary amine could be used directly for attachment, modification of the *C*-terminus necessitated inserting an extra lysine and blocking the *N*-terminus with a protecting group. As this requires a significant number of extra steps in terms of peptide preparation, and steps on the surface, it would add difficulty each step of the way. Thus, the *N*-terminal immobilisation approach was tested first. For this study, a new blocker was used to saturate the surface after addition of the peptide to the surface, 2-(2-methoxyethoxy)ethanamine **3.2**, to react with excess NHS esters, Figure 14.

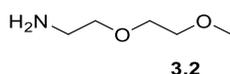


Figure 14: Compound **3.2**, blocker used for the NHS-ester rigid layer for QCM-D measurement

Insulin binding was first tested as a positive control. Two peptides were tested for conjugation to these NHS-functionalised surfaces, Cys-P2, introduced previously, and a P2 sequence capped with an *N*-terminal GABA unit, GABA-P2. As shown in Figure 15, both of these peptides bound insulin as decreased frequencies were observed when adding more and more protein. The data showed that the change in frequency is much stronger for Cys-P2, with a  $\Delta f$  of almost 11 Hz compared to 3.5 Hz for GABA-P2. After extrapolation of the data to have the binding curves, we could see that the curve obtained had similar trends for both peptides. From the equation used and detailed in the Experimental section, the curve of the sigmoidal, the *n* factor, was obtained with value of 0.38 and 0.42 for the sigmoidal fits respectively for GABA-P2 and Cys-P2, Figure 16. While this trend is different to the previous results obtained from the same peptide on gold sensors (*n* value of 1.0), the surface is also drastically different, and the quantity of peptide linked to the surface not controlled. Thus, while 100  $\mu$ M insulin was enough to reach saturation for the first assay with the gold surface it seems that a greater concentration would be needed in this case to allow  $K_d$  determination. While such high concentrations could be conceivable for model proteins, this is not possible for FGF-2 and so this strategy wasn't pursued further.

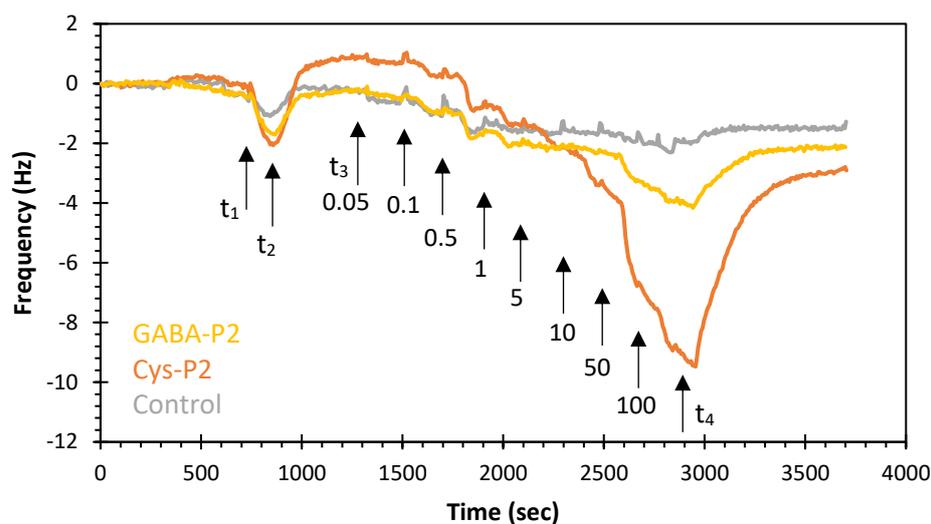


Figure 15: QCM-D curves obtained for the binding of peptide P2 with insulin on an NHS-ester PEG6 modified sensor surface. Addition of peptide overnight not represented in figure, addition at  $t_1$  of blocker 3.2,  $t_2$  PBS wash,  $t_3$  addition of more and more concentrated protein concentration from  $t_3$  (0.05, 0.1, 0.5, 1, 5, 10, 50, 100  $\mu\text{M}$ ) and  $t_4$  final PBS wash.

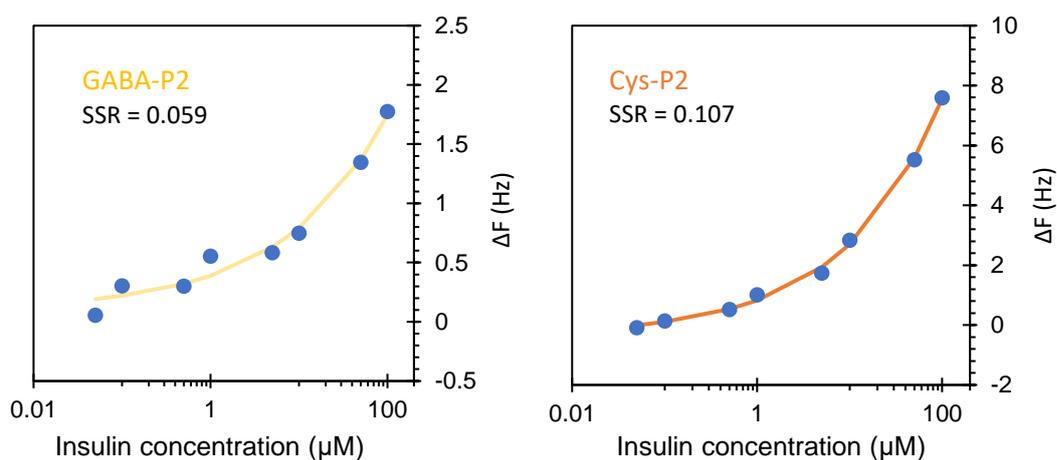


Figure 16: Binding curves for peptides GABA-P2 (left), and Cys-P2 (right) to insulin when using a  $\text{SiO}_2$  sensor with a rigid layer of NHS-ester

### 3.1.4.b An attempt with a maleimide-layer

Another attempt was done using  $\text{SiO}_2$  sensors with this time the surface being modified by 3-aminopropyltriethoxysilane (APTES) and then a 4-maleimidobutyric acid *N*-succinimidyl ester (C4 linker, as represented in Figure 4). In this case the surface could be functionalised with cysteine-modified peptides as described above. It was shown using the C4-linker  $\text{SiO}_2$  sensors that only a  $-3$  Hz shift was observed compared to the original  $-8$  Hz shift on the gold sensor upon adding FGF-2 at the same 100 nM concentration, which is an improvement in non-specific binding. However higher concentrations of proteins still led to biofouling, as shown by the two control repeats in Figure 17.

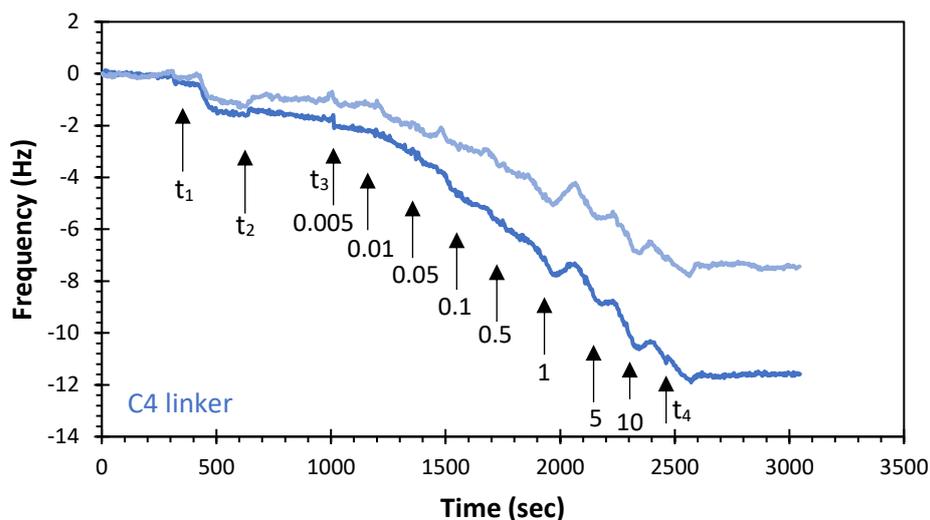


Figure 17: QCM-D curves obtained for the binding negative control assay with FGF-2 on a SiO<sub>2</sub> surface using a short C4 linker. Addition at  $t_1$  of blocker **3.1**,  $t_2$  PBS wash, from  $t_3$  addition of more and more concentrated protein concentration (0.005, 0.01, 0.05, 0.1, 0.5, 1, 5, 10  $\mu$ M) and  $t_4$  final PBS wash.

A longer linker (PEG6 as represented in Figure 4) was therefore used to increase the distance between the surface and the protein, decreasing the chance of non-specific interactions between the sensor and the protein. As shown in Figure 18, two repeats of these conditions minimised biofouling with a change in frequency of around 1Hz observed only for the highest concentration of FGF-2 introduced, 10  $\mu$ M.

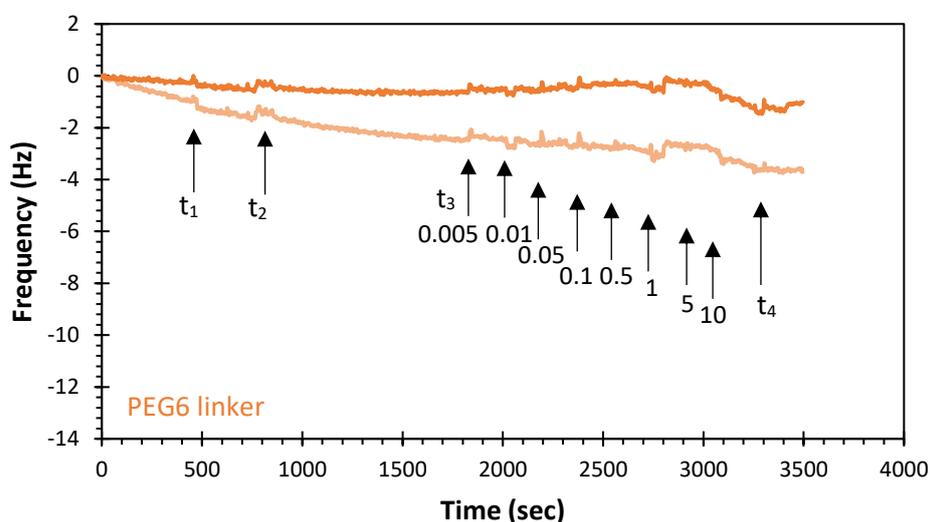


Figure 18: QCM-D curves obtained for the binding negative control assay with FGF-2 on a SiO<sub>2</sub> surface using a PEG6 linker. Addition at  $t_1$  of blocker **3.1**,  $t_2$  PBS wash, from  $t_3$  addition of more and more concentrated protein concentration (0.005, 0.01, 0.05, 0.1, 0.5, 1, 5, 10  $\mu$ M) and  $t_4$  final PBS wash.

We next introduced the cysteine-modified FGF-2 binding peptides onto the PEG6 SiO<sub>2</sub> sensors and then added FGF-2 at the same range of concentrations. Three sequences were picked for FGF-2

binding, all known from literature as discussed in Chapter 2, sequence HTTHMYL known as P5, TLHSAQA known as P6 and HRNPRNN P7.<sup>12</sup> Sequence P6 was studied first.

When adding the free cysteine peptide onto the surface at  $t_1$  no shift was observed, unlike previous studies with insulin and RNase A binding peptides, but this could have been due to the fact that the size of the peptide was not significant compared to the size of the already existing maleimide layer on the surface. Similarly, no peak was observed for the addition of the blocker at  $t_2$  supporting this hypothesis. However, more significantly no clear shift could be observed from the addition of the protein from  $t_3$  onward (Figure 19) despite the increase of concentration from 5 nM to 10  $\mu$ M FGF-2. These results seemed to show that no binding could be observed between the first sequence P6 and FGF, for both C and N terminally modified peptides.

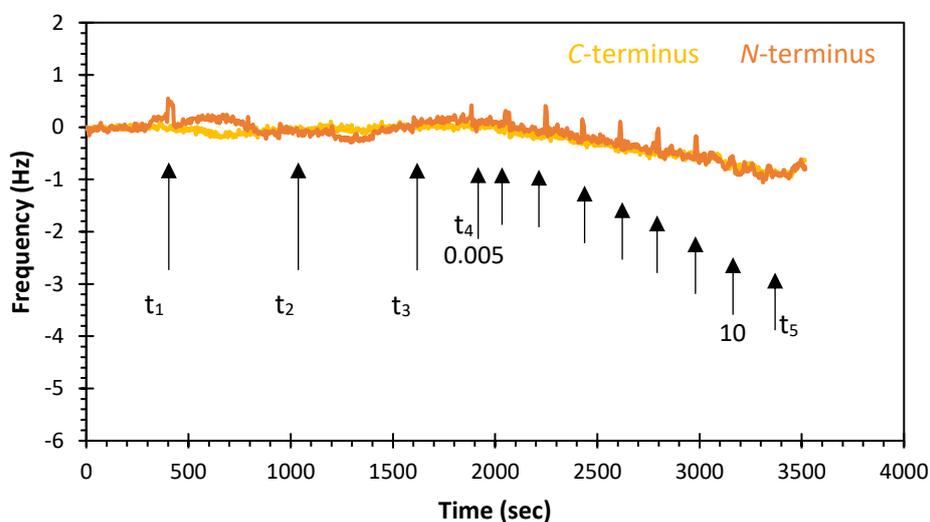


Figure 19: QCM-D curves obtained for the binding of peptide P6 with FGF on a maleimide PEG6 modified sensor surface. Addition of peptide at  $t_1$  and then at  $t_2$  of blocker **3.1** and finally addition of more and more concentrated protein concentration from  $t_3$  (0.005, 0.01, 0.05, 0.1, 0.5, 1, 5, 10  $\mu$ M),  $t_5$  final PBS wash.

In order to confirm these results, it was decided that the set-up of this sensor would be validated using the P2/insulin system as a positive control. This experiment using insulin showed again that no peptide shift was observed using P2, when this system showed earlier successful binding. More importantly even with adding the protein no curve was obtained, with an overall shift under -2 Hz, Figure 20.

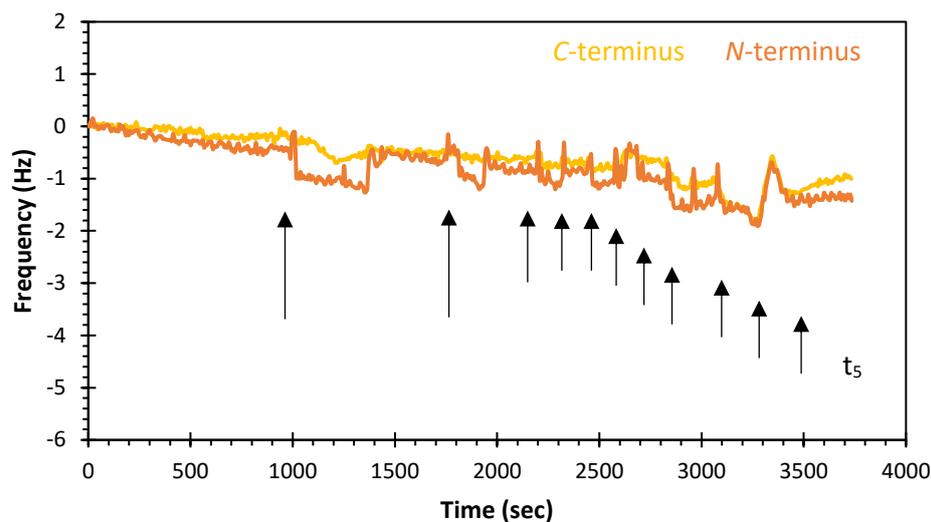


Figure 20: QCM-D curves obtained for the binding of peptide P2 with insulin on a maleimide PEG6 modified sensor surface. Addition of peptide at  $t_1$  and then at  $t_2$  of blocker **3.1** and finally addition of more and more concentrated protein concentration from  $t_3$  (0.05, 0.1, 0.5, 1, 5, 10, 50, 100  $\mu$ M).

Thus, these results contradicted those on the gold surface, and indicated additional challenges in the construction of the rigid layer of maleimide, with the stability of the different reagents needed to form the layer, the reactivity of the maleimide moiety, the hydrolysis of the maleimide at any of those steps, and the accessibility towards the surface due to the presence of the thick PEG6 layer being potentially problematic. This again made this surface unsuitable for measuring FGF-2 binding.

### 3.1.5 QCM-D, a conclusion

To conclude this part, while QCM-D measurements were found to be straight forward to set-up on gold surfaces, the surface does not work for all types of protein. While it has been seen that a lot of optimisations can be done in order to obtain valuable information about the affinity between the protein and peptide, this method was also found to have a lot of parameters that were dependent on the protein used, which limits its use as a quick and widely applicable test. In addition, it required around 1 mL of each concentration of protein for one data point which was another limiting factor when using valuable proteins. We therefore investigated other techniques for assessing protein-peptide binding which will be discussed in the following section.

## 3.2 Fluorescence Polarization (FP) as a more sensitive technique for peptide-protein binding assays

### 3.2.1 The technique

Fluorescence polarization (FP) was first described in 1926 by Perrin and has since been used widely for the study of small molecule-protein, antigen-antibody, and hormone-receptor binding. It has been shown to be one of the most sensitive, robust, and high-throughput screening methods available for such studies.<sup>13,14</sup> The technique is based on the principle of polarization, inherently linked to the mass of a species, known as the Brownian molecular rotation. The technique uses two sets of polarising filters, one polarizing the light entering the solution vertically, and two filters at the end with one vertical ( $I_{\parallel}$ ) and one horizontal ( $I_{\perp}$ ), the second being positioned at a  $90^{\circ}$  of the first, Figure 21.

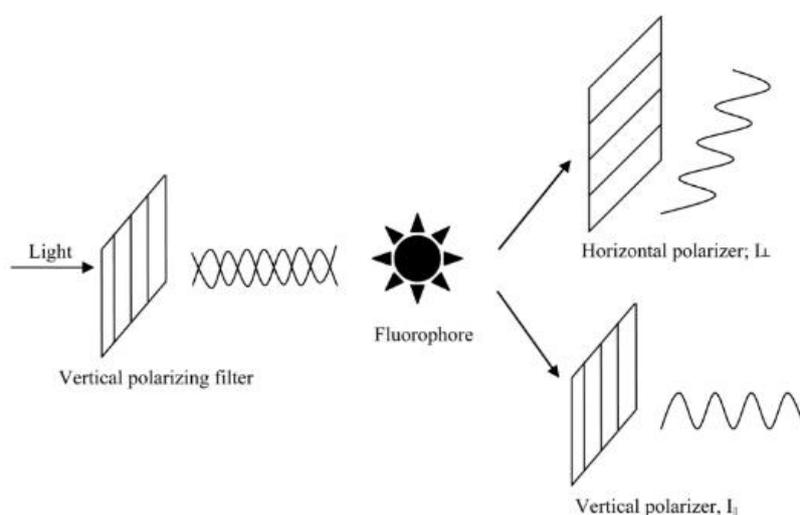


Figure 21: Schematic representation of fluorescence polarization. A fluorophore is excited through an excitation polarizing filter; and two measurements are obtained: parallel or perpendicular emission intensity which can then be used for the calculation of FA and FP, reproduced from Simeonov's group reference.<sup>13</sup>

When a fluorophore, of a small molecular weight, is excited via the polarized light its inherent Brownian rotation is fast, and the resulting emission is split equally between both vertical and perpendicular filters. On the contrary when a protein is excited, due to a considerably higher molecular weight resulting in a lower Brownian rotation, the emissions will be different between both filters. Therefore, when a fluorophore is added to a protein, its Brownian rotation is altered through dissociation/breakdown or association/binding events, thus affecting the degree of depolarization of the emitted light. This gives an FP value which is based on the difference between the parallel ( $I_{\parallel}$ ) and perpendicular ( $I_{\perp}$ ) polarized light intensity obtained after the excitation of the target compound with polarised light. The difference between these intensities is normalized by the total fluorescence

emission intensity, and from this, two equations can be given, one for fluorescence polarization (FP, equation 1) and one for anisotropy (FA, equation 2). This technique is also widely used to perform competition kinetic studies but this will not be used in this thesis.<sup>2</sup>

$$\text{Equation 1 } FP = \frac{I_{//} - I_{\perp}}{I_{//} + I_{\perp}}$$

$$\text{Equation 2 } FA = \frac{I_{//} - I_{\perp}}{I_{//} + 2I_{\perp}}$$

An important point for later is that it can be seen from Equations 1 and 2 that the FP value is independent of fluorophore concentration.<sup>13</sup> Thus, compared to QCM-D considerably smaller concentrations of protein can be used for binding assays, depending only on the sensitivity of the instrument.

The equations mentioned depend on a free system where the molecules studied can tumble, thus a number of assumptions must be made. First, aggregation of the fluorescent probe is ignored (which can result in a considerable decrease in Brownian rotation) as this can initiate deviation from the original mathematical relationship. Second, as instruments may have unequal sensitivity in detecting light in the perpendicular and the parallel orientations, a grating factor (commonly referred to as G factor) can be introduced to correct for this bias in order to calculate absolute polarization values leading to equation 3.

$$\text{Equation 3 } FP = \frac{I_{//} - G * I_{\perp}}{I_{//} + G * I_{\perp}}$$

It is also important to mention that the intrinsic fluorescence intensity (i.e., quantum yield) of a fluorophore may change upon binding to its cognate partner, thus resulting in significantly different contributions of the bound versus free forms of the fluorophore to the total fluorescence intensity of the sample. This phenomena can complicate the interpretation of FP measurements and thus was hypothesised of not happening.<sup>13</sup>

### 3.2.2 Different fluorophores

A wide range of fluorophores have already been used for FP measurements such as nitrobenzofurazan (NBD), fluorescein, rhodamine, Texas Red, Alexa Fluor, and BODIPYs which all have different wavelengths of excitation and emission.<sup>2,15</sup> In order to compare which fluorophore would be the easiest to work with for the modification of peptides on resin, three of them were picked: BODIPY-FI (**3.3**) NBD (**3.4**), and, 5,6-carboxyfluorescein (**3.5**) represented in Figure 22. The applications of these dyes in biomolecular labelling have been largely exploited due to their high molar extinction coefficients ( $\epsilon$ ), high fluorescence quantum yields ( $\Phi_f$ ), and narrow absorption and emission spectra

in the visible light range.<sup>16,17</sup> On the other hand, NBD has suitable fluorescence for excitation at 488 nm, a wavelength commonly used in the FP instrument we had available at the outset of the PhD (see Experimental section) so was suitable for the assay, and its small size and flexibility, due to the 6-carbon chain linker, make it a non-disruptive fluorophore which is less likely to affect peptide binding. Carboxy-fluorescein was also picked due to its known excitation at 488 nm and its widespread use in many studies of protein-protein interactions.<sup>18,19</sup>

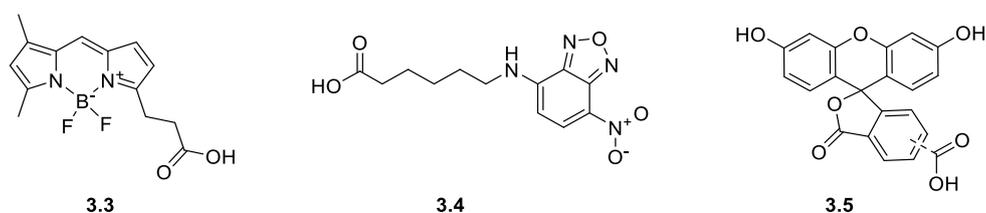
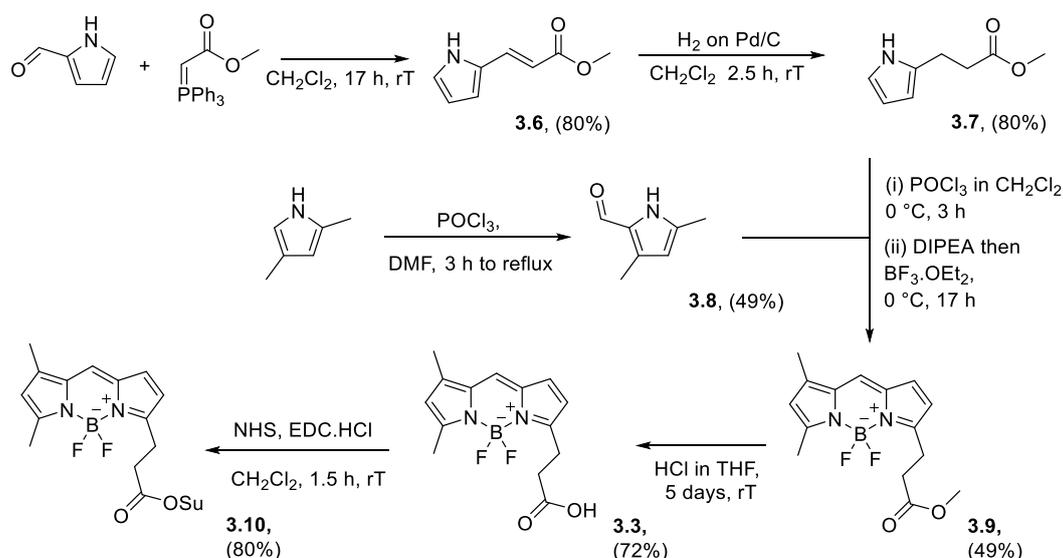
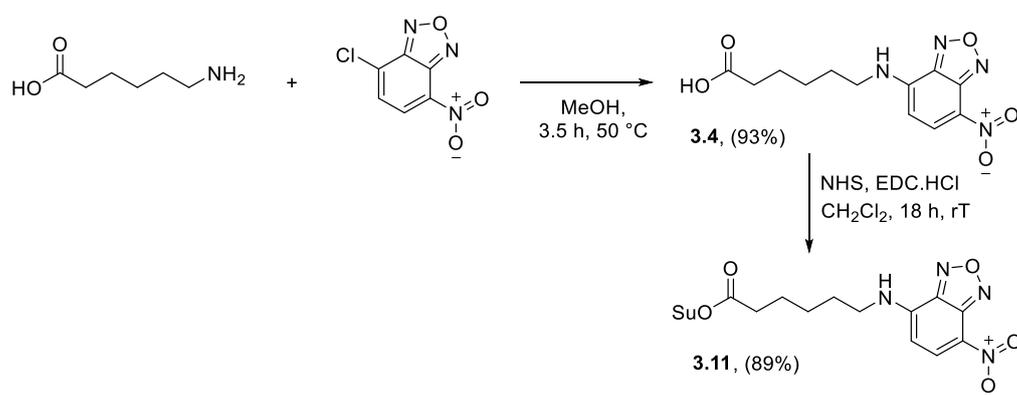
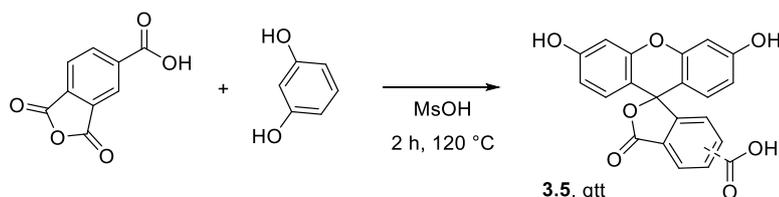


Figure 22: Structures of BODIPY-FI (**3.3**), NBD (**3.4**) and 5,6-carboxyfluorescein (**3.5**).

### 3.2.2.a Synthesis of the fluorophores

BODIPY-FI, **3.3**, was synthesised successfully in a five-step route described by Krajcovicova's group as shown in Scheme 1.<sup>20</sup> The final hydrolysis was modified as the reported use of lithium hydroxide did not seem to form the desired carboxylic acid with no consumption of starting material observed. Instead, a solution of hydrochloric acid in THF was used giving good yield, 72%, of **3.3** as a red crystalline solid, leading to an overall yield of 13% over the five-step process.<sup>20</sup>

Both **3.4** and **3.5** were synthesised in a single step from commercially available materials, as represented in Scheme 1, see part 3.5 Experimental.

**BODIPY-FI = 3.3****NBD = 3.4****5,6-carboxyfluorescein = 3.5**

Scheme 1: Synthesis of the different fluorophores, BODIPY-FI (**3.3**), NBD (**3.4**) and 5,6-carboxy fluorescein (**3.5**).

**3.2.2.b Fluorescence intensity of the fluorophores**

The fluorescence intensity of the fluorophores was assessed after excitation at 480 nm, a common wavelength in fluorescence protocol, (30 nm excitation width) and the detection was set at 530 nm (30 nm excitation width). A range of concentrations from 5 nM to 0.5 mM were tested for the different compounds BODIPY-NHS (**3.10**), 5,6-carboxyfluorescein (**3.5**), and NBD-NHS (**3.11**) in PBS, Figure 23. It was clear that compound **3.10** exhibited the strongest fluorescence at this wavelength with an intensity 8.5 times stronger than **3.11** and 2.5 times stronger than **3.5** at 50  $\mu\text{M}$ . For both compound **3.10** and **3.5**, the fluorescence intensity detected at 5 nM was 10 times higher than the background

noise. To ensure that fluorescence intensity is not affected by the background, a general factor of being three times greater is a common standard. Thus, both of these compounds respect this criterion even at these lowest concentrations. For compound **3.11** to have a similar margin the concentration would need to be significantly greater, between 5  $\mu\text{M}$  and 10  $\mu\text{M}$ . This was a limitation of this fluorophore as the use of high concentrations would also necessitate the use of high concentrations of protein.

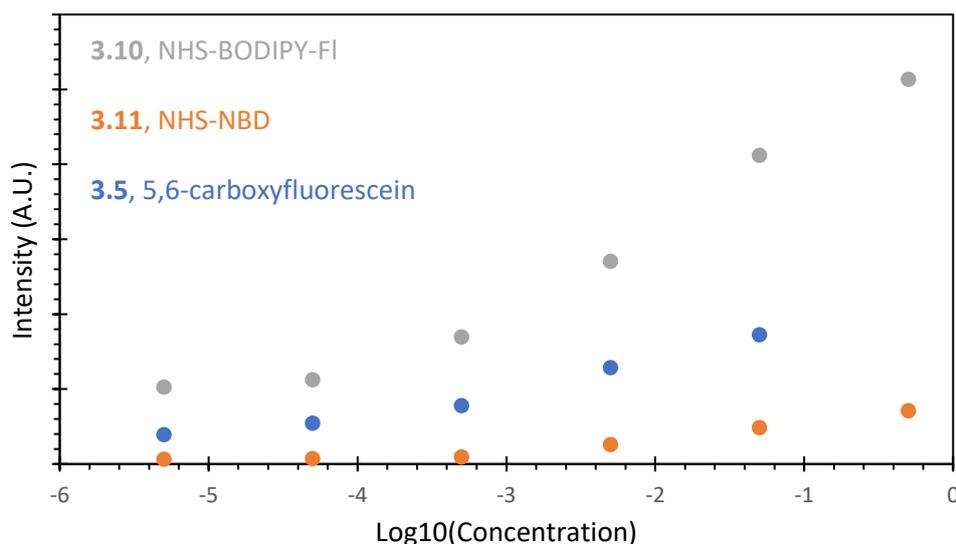


Figure 23: Fluorescence intensity measured from 5 nM to 500  $\mu\text{M}$  of compounds **3.10** (NHS-BODIPY-FI, in grey), **3.11** (NHS-NBD, in orange) and **3.5** (5,6-carboxyfluorescein, in blue) at excitation of 480 nm and detection at 530 nm

It is important to note that this first experiment was run just to measure the fluorescence of the compounds as free fluorophores. These intensities could be impacted not only by interactions with the protein, but also by the attachment to the peptides, which will be discussed in the following paragraph.

### 3.2.3 Synthesis of fluorescently labelled peptides

#### 3.2.3.a Using BODIPY **3.10**

Following the first results obtained in section 3.2.1, compound **3.10** seemed to be the best candidate for peptide labelling as it showed the greatest intensity after excitation at 480 nm leading to the highest signal to noise. The compound was attached to the peptides at the *N*-terminus using the protocols optimised in Chapter 2, for on-resin modification and cleavage of peptides.

While the first attempts to modify peptides P1 and P2 on their respective *N*-terminus showed full consumption of the starting material, the mass of the resulting peak did not match the expected compound. This phenomenon was explained by the instability of the  $\text{BF}_2$  BODIPY core in the presence of acid during the cleavage of the peptide from the resin, confirmed by LCMS, Figure 24. This loss of

the BF<sub>2</sub> group also led to total loss of fluorescence of the fluorophore as the BF<sub>2</sub> group is essential for the conjugation of the dye. Precedent for the acid-sensitive character of BODIPY dyes could be found in literature supporting our observations.<sup>21</sup> The loss of BF<sub>2</sub> was not related to ionisation during LCMS analysis, as controls with both starting material **3.3** and peptide coupled with compound **3.37**, see Experimental section, showed expected mass.

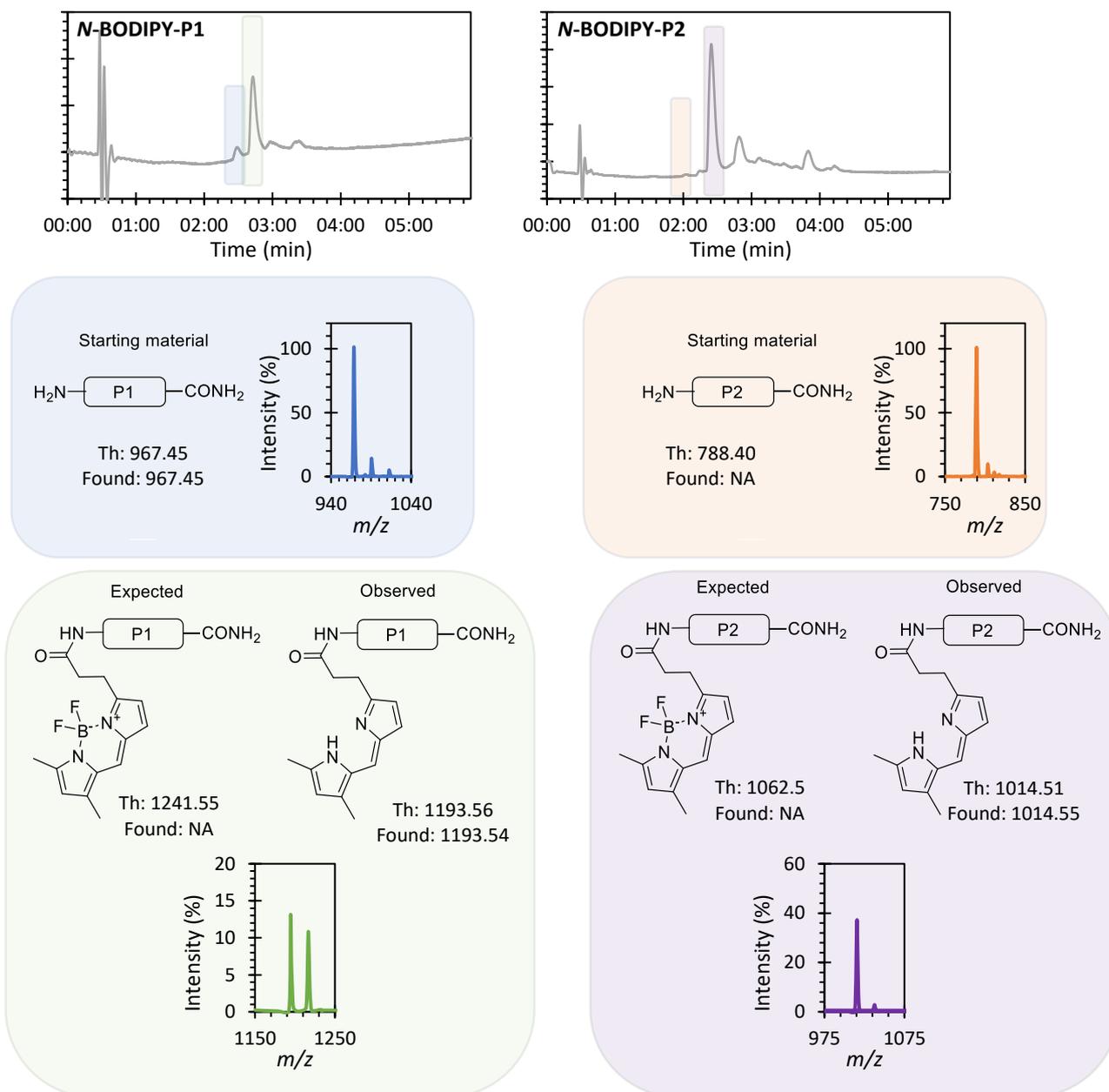


Figure 24: LC-MS traces of the modification of P1 and P2 on their respective N-terminus with compound **3.10**. LCMS UV 220-280 nm traces recorded as a function of time, with the MS for each highlighted peak and their matching structures.

As the BODIPY fluorophore had the strongest fluorescence, alternative conditions for peptide cleavage were attempted. Two aspects could be changed in order to reduce the  $\text{BF}_2$  loss during acid cleavage, first lowering the concentration of acid used by diluting it within a  $\text{CH}_2\text{Cl}_2$  solution, or second lowering the time of cleavage.

In a first attempt, going from a 95% TFA solution to a 10% solution in  $\text{CH}_2\text{Cl}_2$ , with the cleavage time also reduced from 5 h to 30 min, showed that no more loss of the  $\text{BF}_2$  was occurring. However, the chromatogram obtained after cleavage showed a lot of different peptide peaks corresponding to different combinations of BODIPY-modified peptide with protecting groups still attached (such as Trt, Boc, *t*Butyl). While this was encouraging, it was necessary to achieve full peptide deprotection. Another attempt with a 20% solution of TFA for 1 h seemed to help as only the Pbf group was still present on the peptide sequence, however a small peak matching the loss of the  $\text{BF}_2$  was observed again. Thus, despite its interesting properties as a fluorophore, the BODIPY did not seem to be the ideal candidate for peptide modification on-resin due to the instability to acid conditions.

#### 3.2.3.b Using NBD 3.11

The next fluorophore to be introduced on peptide was compound **3.11**, using an NHS-coupling to the peptide *N*-terminus, see general conditions in the Experimental Part. **3.11** was used first despite lower fluorescence as easier to synthesise in bigger scale and more soluble than 5,6-carboxyfluorescein. It was shown that the conversion to the fluorescent peptide was almost complete for sequence P2, 94%, however a significant amount of starting material for peptide P1 was observed, with only 66% conversion to the expected fluorophore-modified peptide, as shown in Figure 25.

In order to verify if the NBD fluorescence was affected once the fluorophore was conjugated to the peptides, the fluorescence of different concentrations of purified *N*-NBD-P2 were measured and compared to compound **3.11**, to observe the difference in intensity at fixed concentrations. The results obtained are represented in Figure 26. Despite the first data point at 500  $\mu\text{M}$  being out of the trend, all the other data points showed a decrease of fluorescence intensity for the labelled peptide by a factor of 2-3. In order to have a strong signal for fluorescence polarization, using these NBD-peptide would require a minimum concentration of 10  $\mu\text{M}$  peptide which is quite significant and would limit the range of protein concentrations for fluorescence polarization, due to both the cost of using more protein and solubility problems at high concentrations.

Thus, both **3.10** and **3.11** exhibited limitations for the applications needed throughout this PhD, to be able to modify peptide on resin in high yields with high fluorescence intensities at reasonable concentrations for FP measurements.

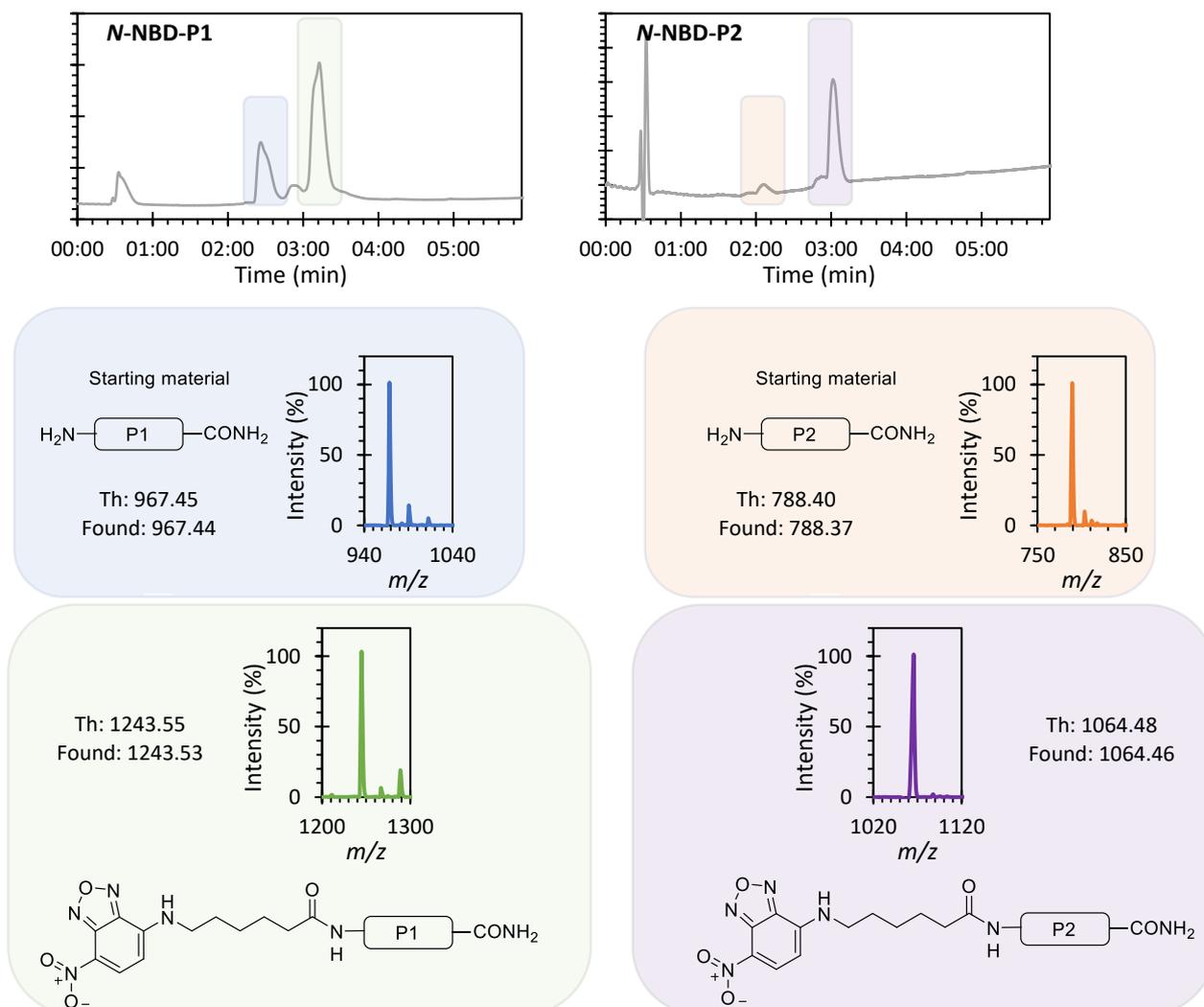


Figure 25: Graphs of the modification of P1 and P2 on their respective N-terminus with compound **3.11**. LCMS UV 220–280 nm traces recorded as a function of time, with the MS for each highlighted peak and their matching structures.

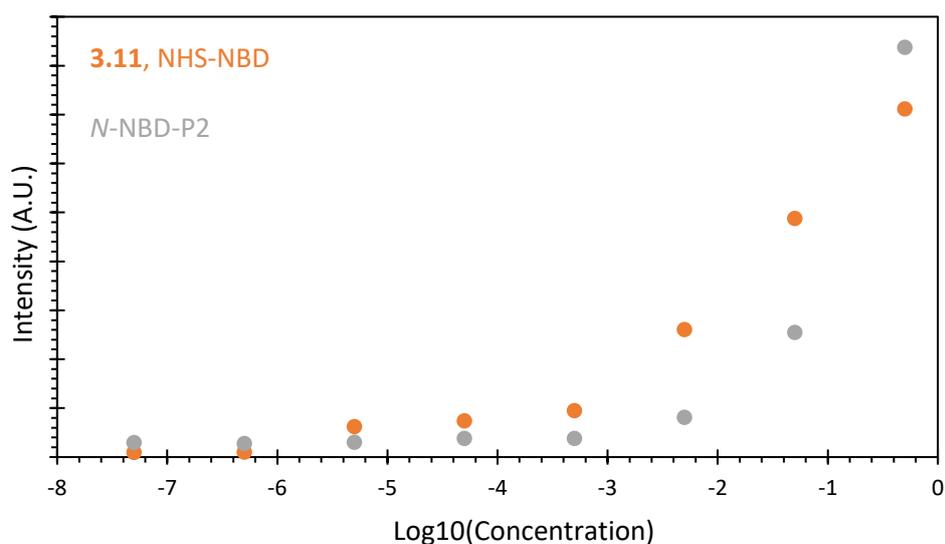


Figure 26: Fluorescence intensity measured from 5 nM to 500  $\mu$ M of compound **3.11** (orange) and N-NBD-P2 (grey) after excitation at 480 nm and detection at 530 nm

### 3.2.3.c Using 5,6-carboxyfluorescein 3.5

**3.5** was the last of the synthesised fluorophores to be investigated. Compared to the first two compounds, 5,6-carboxyfluorescein was introduced via the direct amide coupling of the carboxylic acid as outlined in Chapter 2. This protocol was found to be efficient with conversions >90% for all attempts (see LCMS traces in the Experimental section). However during the precipitation process of the modified peptides following resin cleavage, and despite the previous washes done before cleavage, it was found that due to its inherent hydrophobicity, compound **3.5** was present in the final solution, Figure 27, likely due to some retention on the resin. Thus, extra washing steps with DMF were done before peptide cleavage until no colour was observed in the wash, indicating all fluorescein had been effectively removed, followed by washes with CH<sub>2</sub>Cl<sub>2</sub> and MeOH. However, in some cases peptides needed purification after cleavage via reverse-phase column chromatography as described in the experimental section, or by size exclusion chromatography. The final LCMS traces of the purified peptides are presented in the 3.5 Experimental section.

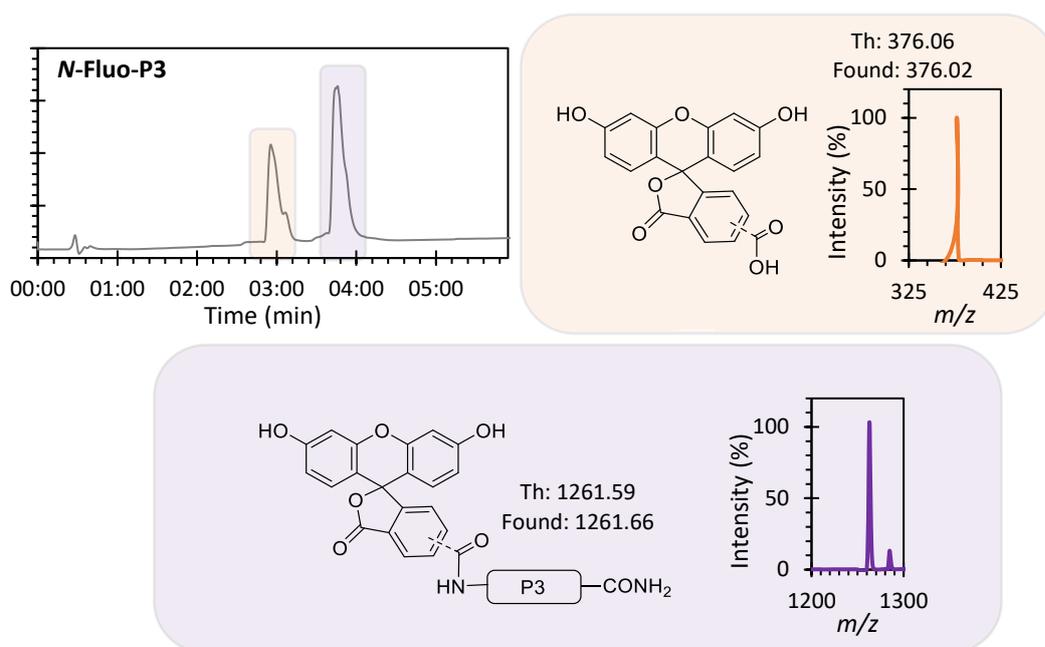


Figure 27: Graph of the modification of P3 at the N-terminus with compound **3.5**. LCMS UV 220 nm traces recorded as a function of time, with the MS for each highlighted peak and their matching structures

After obtaining pure fluorescently labelled peptides, the fluorescence intensity was measured again across the range 0.5 nM-500 nM. Three peptides were tested to determine if fluorescence intensity was largely independent of the peptide sequence. The results, Figure 28, showed that the fluorescence intensity observed was more dependent on the sequence than we had predicted as, for example, at a concentration of 100 nM, the sequence Fluo-P6 already reached the detection limit of the instrument while Fluo-P7 was only at 34% of the total intensity and Fluo-P5 at 74%.

As mentioned earlier, for FP measurements the fluorescence intensity of the probe, here the peptide, needs to be greater than three times the background measurement of the instrument for accurate measurements. By zooming in on the region 5 to 50 nM, it can be seen that all of the peptides at 50 nM are at least five times the background and so this concentration was used as a minimum moving forward.

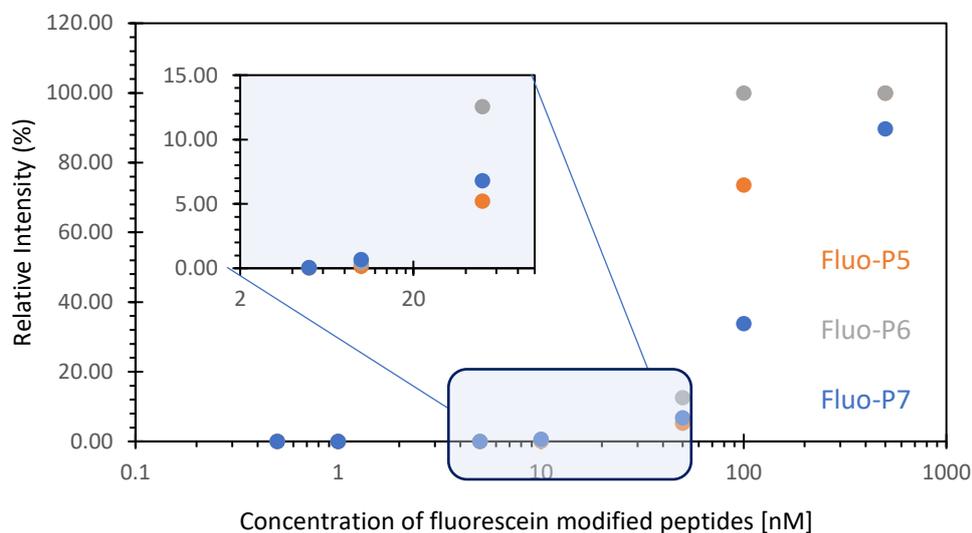


Figure 28: Fluorescence intensity as a function of concentration for 5,6-carboxyfluorescein modified peptides (in orange Fluo-P5; in grey Fluo-P6; in blue Fluo-P7)

Following this experiment, a second set of concentration ranges of 5-200 nM was tested by applying a 70% gain on the instrument, to avoid reaching the detection limit as in the first experiment, with the results shown in Figure 29. The difference between the two peptides was consistent with a factor 1.4 between the values of peptide P6 fluorescence compared to P5.

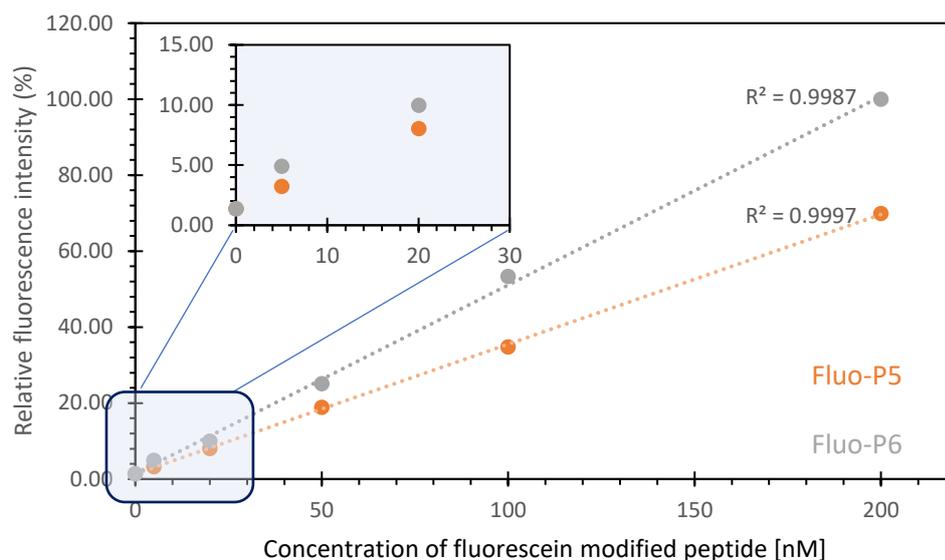


Figure 29: Second fluorescence intensity measured in the nM range for 5,6carboxyfluorescein modified peptides (in orange Fluo-P5; in grey Fluo-P6) with a 70% gain adjustment

It was possible to show that lower concentrations of peptide fulfilled the signal to noise criterion, resulting in the use of 20 nM as a target peptide concentration for FP measurement. This concentration was drastically lower than needed for QCM-D detection (10  $\mu$ M) showing the sensitivity of FP technique for peptide-protein binding assay. This would in turn impact the amount of protein needed making this method more advantageous by first allowing us to analyse a wider range of protein concentrations, and second avoiding solubility and precipitation risks.

### 3.2.4 Fluorescence polarisation for measuring protein binding

For a first experiment using FP, the three FGF-2- peptides mentioned in the previous sections, *N*-Fluo-P5, *N*-Fluo-P6 and *N*-Fluo-P7 were used. A peptide concentration of 100 nM was used first to ensure good signal, as we were not sure whether the fluorescence intensity would be impacted by the presence of the protein. The concentration range of the protein was 0.02-20  $\mu$ M (0.1-100 eq to peptide). The full protocol, data obtained, and data processing details are presented in section 3.5 Experimental. After processing the data the following binding curves were obtained, Figure 30.

From this first set of data, two observations could be made. First peptide *N*-Fluo-P7 seemed to be a poor binder of FGF-2 once modified on its *N*-terminus. This result indicated that modification of the *N*-terminus of the sequence P7 impacted the previously reported binding somehow.<sup>12</sup> This could be due to the lack of hydrogen bonding capability at the *N*-terminus which is lost during modification, or due to the steric bulk of the fluorescein positioned at the beginning of the sequence which could block access to the binding site. This result showed how peptide modification can dramatically impact the binding of a peptide, and the need for target binding to be validated following peptide modification. Second, while it was promising that the two other sequences *N*-Fluo-P5 and *N*-Fluo-P6 could still bind

to FGF-2, the binding curves obtained were non optimal. While a plateau was reached for *N*-Fluo-P5 at higher equivalents (from 10 eq), this was not observed for *N*-Fluo-P6, and the bottom plateau for lower concentration could also be increased with additional data points. It was therefore decided to run a second FP assay with the two binding peptides.

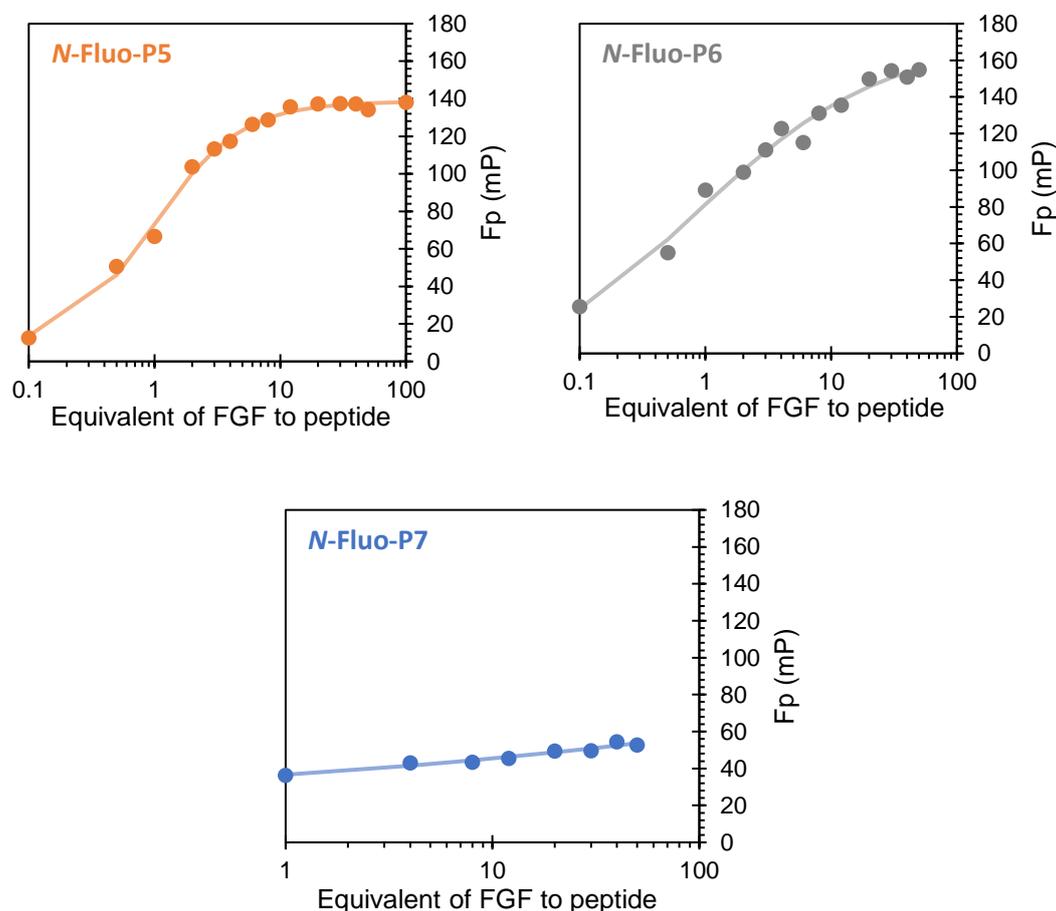


Figure 30: FP measurements of *N*-Fluo-P5 (top left in orange) and *N*-Fluo-P6 (top right in grey) and *N*-Fluo-P7 (bottom in blue) at a fixed peptide concentration of 100 nM in presence of FGF-2 zebrafish from 0.1 to 100 equivalents

In this second experiment the fluorescent peptide concentration was fixed at 25 nM allowing us to record more data points for the protein and to perform the measurements in triplicate to ensure the reproducibility of the results. Moreover, it was shown that good signal was obtained in the previous experiment showing that the binding was not overly impacting the fluorescence of the peptides. The new range of protein concentrations studied was from 0.5 nM (0.01 eq) to 10  $\mu$ M (200 eq). Nineteen data points were covered within this range, with the table of results in section 3.5 Experimental. After analysis, the two new binding curves presented in Figure 31 were obtained. It can be seen that the variation between the triplicates were highest within the slope, and this time both plateaus at high and low concentration were resolved.

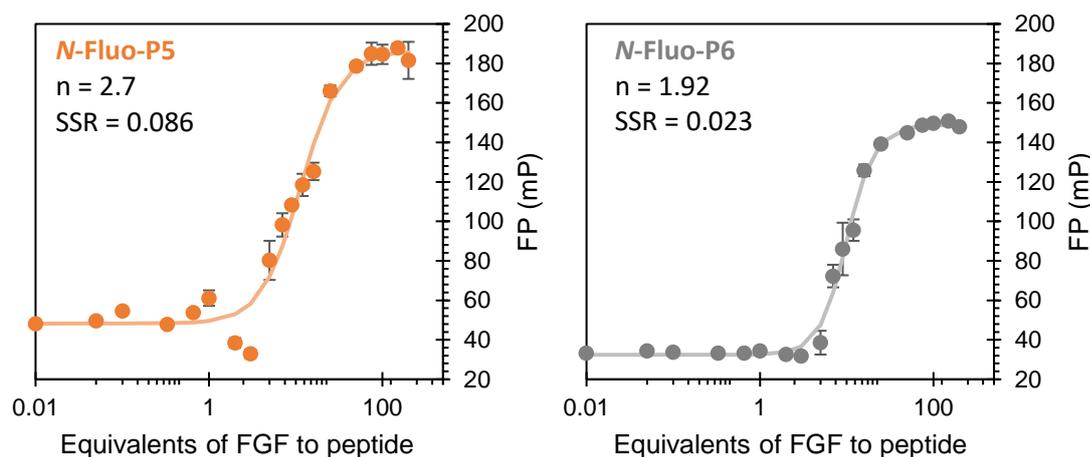


Figure 31: Binding curves obtained from FP measurements for N-Fluo-P5 (left in orange), with its respective sigmoidal fit (orange line), and N-Fluo-P6 (right in grey), with its respective sigmoidal fit (grey line) at a fixed concentration of 25 nM in presence of FGF-2 zebrafish from 0.01 to 200 equivalents

Using a sigmoidal fitting, it could be observed that both of the curves presented in Figure 31 had  $n > 1$ , unlike the usual properties of a standard sigmoidal fit. Due to this, more analysis and interpretation of this set of data was performed. Nobel Prize winner Dr Hill described the possibility of cooperative binding to a multi-site protein by integrating the factor  $n$ , known as the “Hill coefficient” into the usual ligand binding equation, equation 4, where  $[C]$  denotes the concentration of protein, and  $K$  denotes an apparent association constant.

$$\text{Equation 4} \quad Y = \frac{K \times [C]^n}{1 + K \times [C]^n}$$

From his theorem, when  $n < 1$ , the system exhibits negative cooperativity, whereas cooperativity is positive if  $n > 1$  meaning that for the system ligand/protein two or more ligand molecules can bind to the same protein and binding of the second is more favourable after binding of the first. The Hill equation can be linearized, in order to plot the equation 5 as described below. This equation was used to obtain the following graphs for each of N-Fluo-P5 and N-Fluo-P6, Figure 32.

$$\text{Equation 5} \quad \log\left(\frac{Y}{1-Y}\right) = n \times \log([C]) + n \times \log(K)$$

From the trends of the graphs, it was possible to conclude that a positive cooperative binding mechanism was happening during peptide-protein binding, with factor  $n$  being close to 2. This could be linked to the fact that FGF-2 can dimerize and thus once one peptide was actively binding to one part of the dimer, the structural aspect of this dimer help the second binding which could be accelerated. Moreover it was possible to determine binding constants which were 289 nM for N-Fluo-P5 and 246 nM for N-Fluo-P6 respectively. These peptides therefore exhibit strong binding to the target FGF-2.

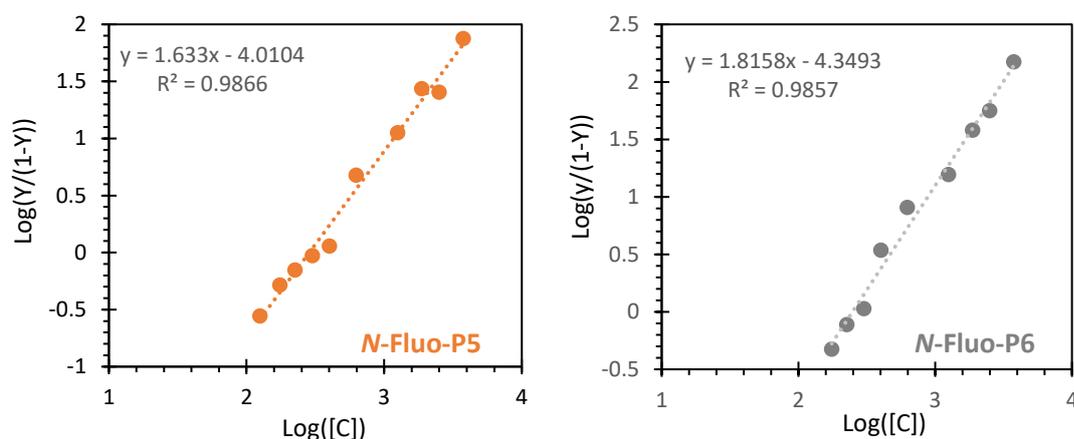


Figure 32: Hill plot of the FP values of N-Fluo-P5 (left in orange) and N-Fluo-P6 (right in grey) at a fixed concentration of 25 nM in presence of FGF-2 zebrafish

Cummulatively, these results showed that of the three modified sequences, two retained target binding after modification, P5-6, while P7 show no significant binding, showing the importance of measuring binding after each peptide modification step. While this last sequence P7 does not show binding it still could be used further down the line as a negative control to ensure that the modifications of the protein are coming from the interaction between the peptide probes and not from background reactivity.

### 3.2.5 Fluorescence Polarisation, a conclusion

Using FP as a method to determine peptide binding helped significantly in terms of time (higher throughput method) and expense (requiring much diluting samples) compared to QCM-D method. A first set of data on FGF-2 were successful. Indeed, these results showed that despite N-terminus modification two of the three known sequences were still strong binders to the protein with  $K_d$  in the nM range. While this technique was tried on insulin for comparison with QCM-D, no data was able to be collected. We hypothesised that the small size of this protein made it not appropriate for FP.

While these two methods, QCM-D and FP, allowed us to study the strength of binding, they didn't provide information on the localisation of the binding at the surface of the protein. The use of photoaffinity labelling was therefore tried to identify the site of binding.

### 3.3 Diazirine photochemistry for peptide binding and cross-linking assays

#### 3.3.1 The use of diazirines in bioconjugation

The development of **Photo-Affinity Labeling (PAL)** since its first report in 1962 was able to considerably change the way protein-protein, and ligand-protein interactions were studied. Combined with mass spectrometry, PAL became a high-throughput method for measuring biomolecular interactions, protein structure, and the identification of small molecule targets in cells. While many photo-reactive molecules have been used for PAL, diazirines have been particularly widely used because of their high reactivity.<sup>22</sup> Figure 33 shows how a diazirine moiety can covalently attach a ligand to a protein, and then how this can be used to identify the binding site. This system was used in this PhD to try and identify the site of binding of the peptides to the protein of interest.

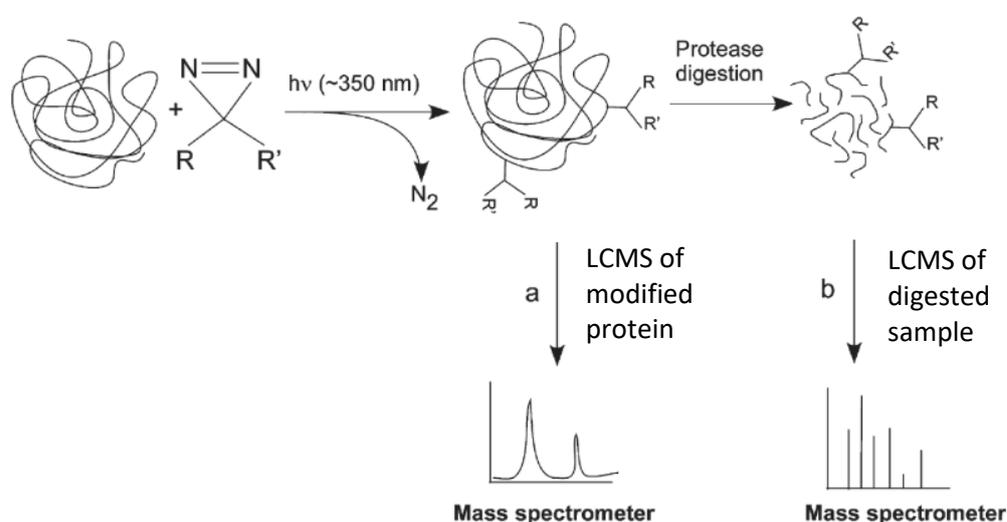


Figure 33: Simplified scheme for the detection and analysis of modified protein by diazirine under irradiation light. The labelled protein can be analysed by ESI mass spectrometer to identify incorporation (a) and then following protease digestion, sites of photoincorporation are detected by acquiring LCMS data from the digested sample in ESI mass spectrometer (b)<sup>22</sup>

On a related note, during the COVID-19 pandemic we produced a review article on the main reactions that have been used for photo-cyclisation of peptides to date.<sup>23</sup> While the aims of this PhD were different, writing this review was greatly beneficial in understanding the underlying photochemistry of the reactions used for PAL.

Two main classes of diazirine have been used in the literature, depending on the targeted amino acids, protein or at a more complex level whole cell proteome: alkyl or aryl diazirines. While it has been found that alkyl diazirines exhibit preferential labeling of acidic amino acids in a pH-dependent manner, the aryl-fluorodiazirine labeling pattern showed higher stability, explained in more details next paragraph.<sup>24</sup> While more than 30 alkyl diazirine probes have been reported in literature, in this

PhD only two type of probes were picked and compared to an aryl probe for their ease of handling, and reactivity towards the protein targets of this project.

### 3.3.2 Synthesis of diazirine probes

Two diazirine probes were synthesised during the PhD one with an alkyl chain **alkyl-diazirine (3.12 Alk-Dz)** and one with a **trifluoromethyl aromatic core (3.13 TFM-Ar-Dz)**, see Figure 34, to be able to compare their reactivity.<sup>25</sup> While **3.12** has more flexibility and is less bulky than **3.13**, when comparing the carbene intermediates formed upon irradiation compounds **3.14** has been shown to be less stable than its analogue **3.15**, due to the lack of electron-withdrawing group, and undergo quenching leading to internal rearrangement to an alkene, **3.18**. In contrast, though **3.13** is bulky but is also more stable.<sup>24</sup> While literature procedures were used to synthesize **3.12**,<sup>26</sup> a new synthetic pathway was devised to achieve the formation of **3.13**, reducing the number of steps from nine to six.<sup>27</sup>



Figure 34: Scheme of the two diazirine probes

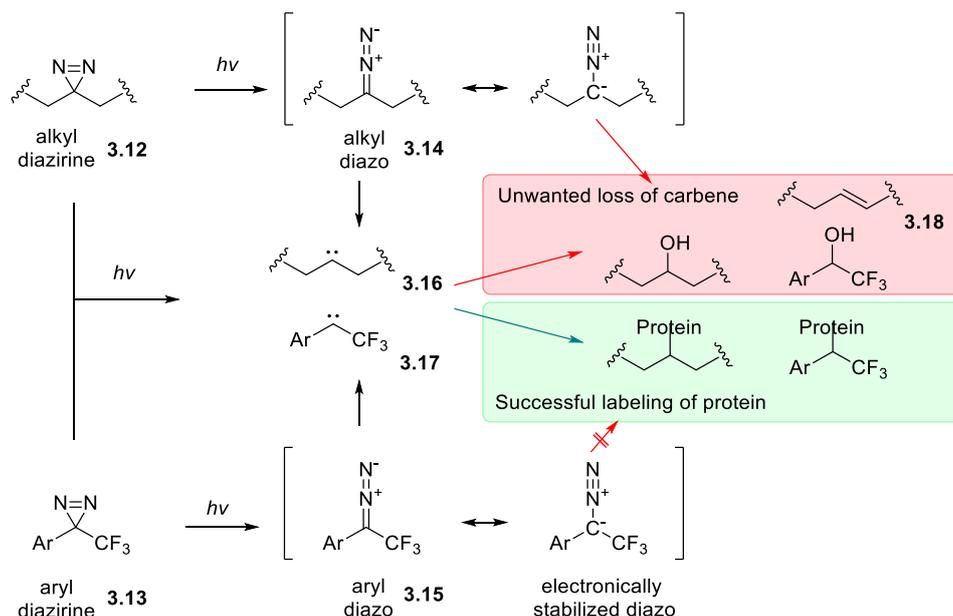
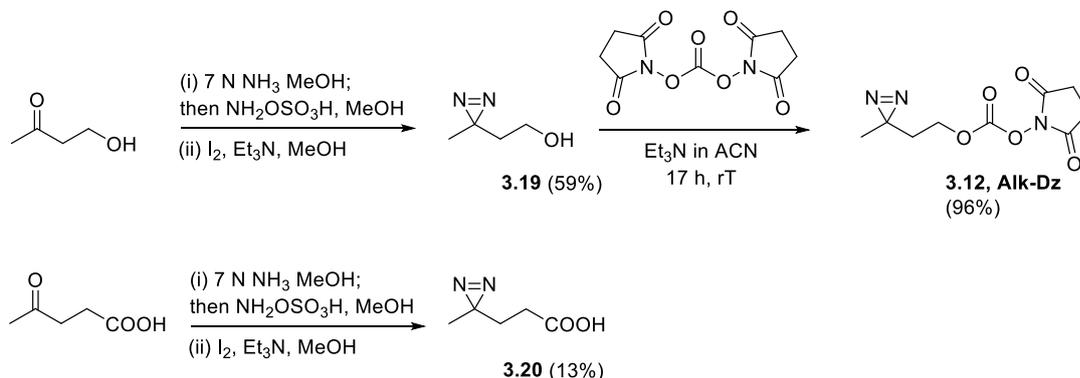


Figure 35: Scheme adapted from literature,<sup>24</sup> showing an overview of diazirine reactivity pathways. Alkyl and aryl diazirines form carbene and diazo intermediates upon irradiation. Carbenes label nearby proteins but are rapidly quenched if no protein substrate is nearby by water or for alkyl diazirines by internal rearrangement.

#### 3.3.2.a Synthesis of alkyl-diazirines

The synthesis of **3.12** was achieved in a two-step process from commercially available 4-hydroxy-2-butanone in a yield of 59% higher than reported in literature (22%).<sup>26,28</sup> The second step to form the activated carbonate **3.12** proceeded in 96% yield. This compound was already activated to amine

attack and could be directly attached to peptide substrates. Probe **3.20** was also synthesized from levulinic acid (in a 13% yield in a single step) for amide coupling.<sup>29</sup>

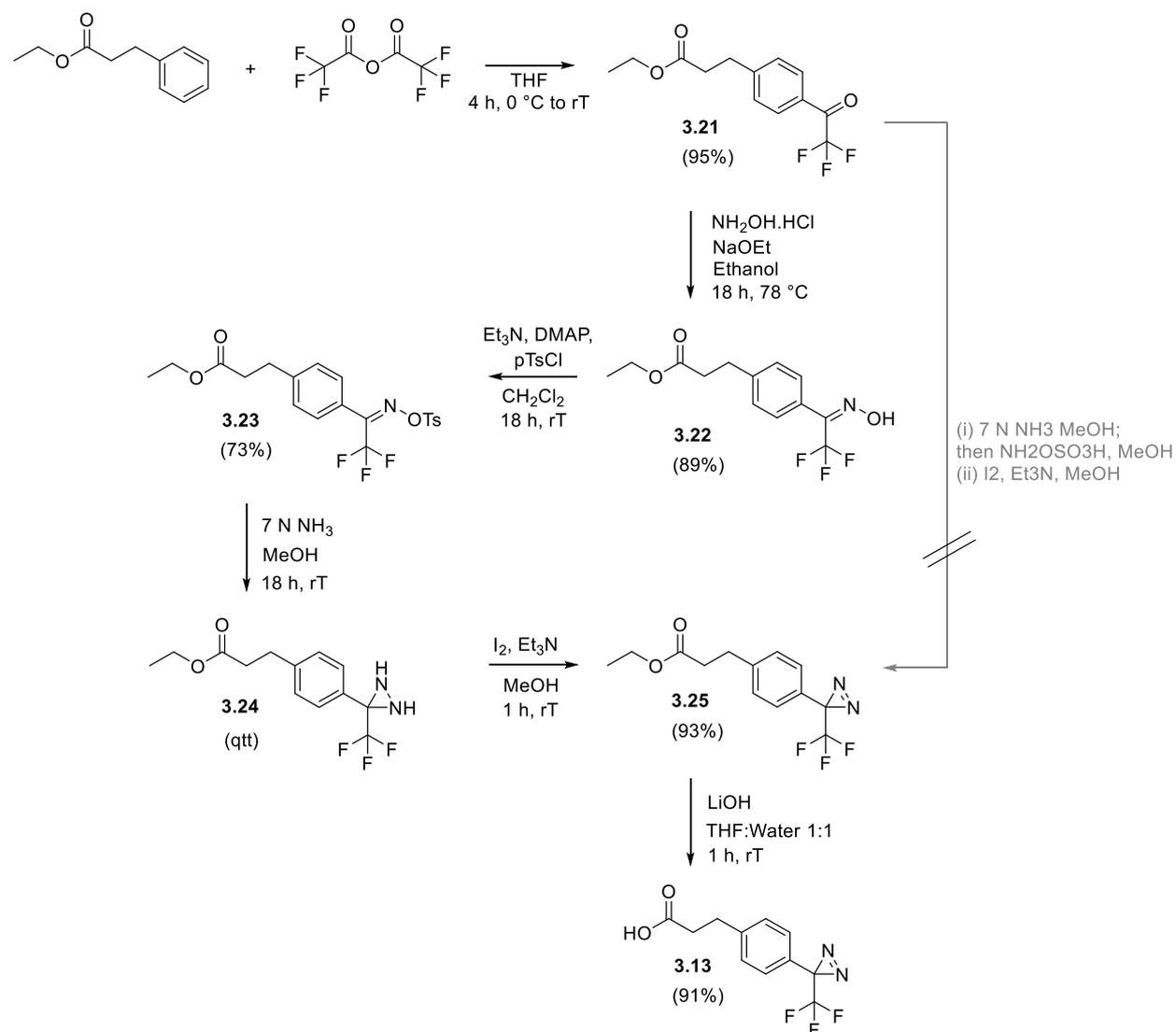


Scheme 2: Synthesis of Alk-Dz probe **3.12**, and probe **3.20**

### 3.3.2.b Synthesis of trifluoromethyl aromatic diazirine (TFM-Ar-Dz, **3.13**)

The synthesis of **3.13** was performed starting from ethyl hydrocinnamate as shown in Scheme 3. While compound **3.13** was known in literature,<sup>27</sup> this new pathway was inspired by reports on the syntheses of alternative diazirines, to decrease the number of synthetic steps.<sup>30</sup> The first step initially consisted of a Friedel-Crafts acylation which gave 24% of compound **3.21** with 40% starting material recovered. The conditions were therefore slightly changed by increasing both the equivalents of TFAA and AlCl<sub>3</sub> and leaving the reaction stir for 5 h instead of 1.5 h, leading to an increase in yield from 24% to 95%. The two next steps were the formation of the oxime **3.22** and then its tosylate form **3.23**, which were successful with high yield (respectively 89% and 73%). In literature, the formation of **3.24** and **3.25** were done in one-pot without isolating the diaziridine compound **3.24**. However, when following similar conditions, the analyses were not sufficient to confirm which species was obtained between **3.24** and **3.25**. By isolating compound **3.24** we were able to make the distinction possible and confirm the generation of **3.25** after further addition of iodine. The last step to form **3.13** was done using classic hydrolysis conditions in a 1 M LiOH solution in THF:H<sub>2</sub>O to give **3.13** in an overall 52% yield over 6 steps.

While literature for synthesis of aryl diazirine seems to always use tosylate activation, the condition for modification of alkyl ketones used during the synthesis of **3.12** were tried too. No modification of compound **3.21** was observed in this case which we hypothesised was due to the presence of the CF<sub>3</sub> group.



Scheme 3: Synthesis of aryl diazirine **3.13** starting with ethyl hydrocinnamate

Our new route had the benefit of using a Friedel Crafts reaction to replace 3 steps in the previously reported route to **3.13**. Moreover, we showed that diazirine is stable to basic conditions, which was avoided in the original protocol in which they use an acidolysis of a TBS ether followed by biphasic TEMPO/BAIB oxidation of the resulting alcohol to provide the final carboxylic acid. While the overall yield was comparable, the decrease of steps reduced the time of synthesis significantly (57% overall yield).<sup>27</sup>

### 3.3.3 Modification of peptides with diazirine probes – testing reactivity

#### 3.3.3.a Solvent and set-up optimisation

Both diazirine probes were not easily observed by standard mass spectrometry techniques, both ESI and APCI. This would make the assignment of possible product peaks formed following irradiation potentially challenging. We therefore performed a study of diazirine reactivity directly on peptides that had a diazirine attached at either *N*-terminus or a *C*-terminal lysine.

For diazirine **3.12**, using the NHS/carbamate protocol, attachment to the peptide on resin was successful, with the diazirine shown to be stable to the peptide cleavage conditions, on both peptides P1 and P2, LCMS traces and structures of the modified peptides are present in section 3.5 Experimental. For Alk-Dz-P2, it was shown that even when stored at -20 °C and wrapped in foil, a new peak appeared over time in LCMS analysis matching the oxidation of the diazirine to a ketone. As shown in Figure 36, when alkyl diazirine **3.26** is excited the carbene formed is the singlet **3.27**. However, a rearrangement is possible due to the intersystem crossing of the position of the electrons in different orbitals from the singlet carbene **3.27** to the triplet carbene **3.28**. While both carbenes species are very reactive, they react differently. While **3.27** undergoes reactions with O-H bond such as water, **3.28** can be oxidized to the corresponding ketone **3.29** due to the reaction with O<sub>2</sub> as shown in Figure 36. Thus, for any further samples used and made, storing under inert atmosphere was ensured to maximise the stability of alkyl diazirine compounds.

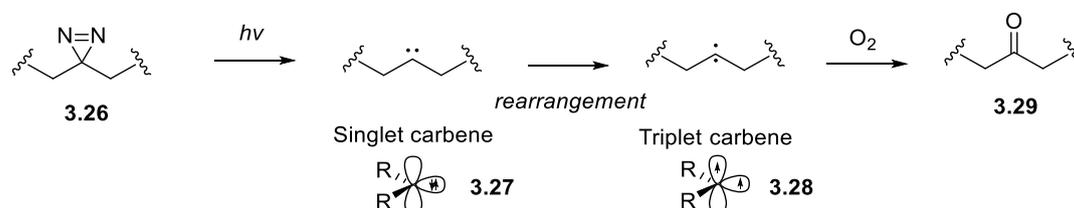


Figure 36: Degradation route due to the intersystem crossing of electrons between singlet and triplet carbenes resulting in degradation of diazirine to its corresponding ketone

The first tests of diazirine reactivity were performed by irradiating the labelled peptides under UV light, at 365 nm, using two types of instruments, and three solvent mixtures in order to follow their respective reactivity. The three solvents used were 100% MeOH, 1:1 MeOH-H<sub>2</sub>O and 100% H<sub>2</sub>O. In a first test the samples were irradiated for 1 h. The two instruments were either a UV lamp from a standard TLC visualisation apparatus or a UV-curing nail instrument from Rio®, see Experimental section. We found that only the UV-curing nail instrument was able to modify the diazirine. Thus, only those results are presented. Different peaks were observed by LCMS matching the structures proposed in Figure 37, following reaction with the hydroxyl of water, structure **3.30**, or with the methoxy of methanol, structure **3.31**. Independent of the solvent, a smaller mass which could reflect

either a cyclisation, **3.32**, or the formation of an alkene, **3.33**, was also observed. The intensity of the peaks was solvent and peptide dependent, with in MeOH the methoxy adduct dominating, while in H<sub>2</sub>O, the ratios showed some variability depending on the sequence of the peptide. Finally, in MeOH, activation was found to be less efficient with some residual diazine still present after 1 hr irradiation.

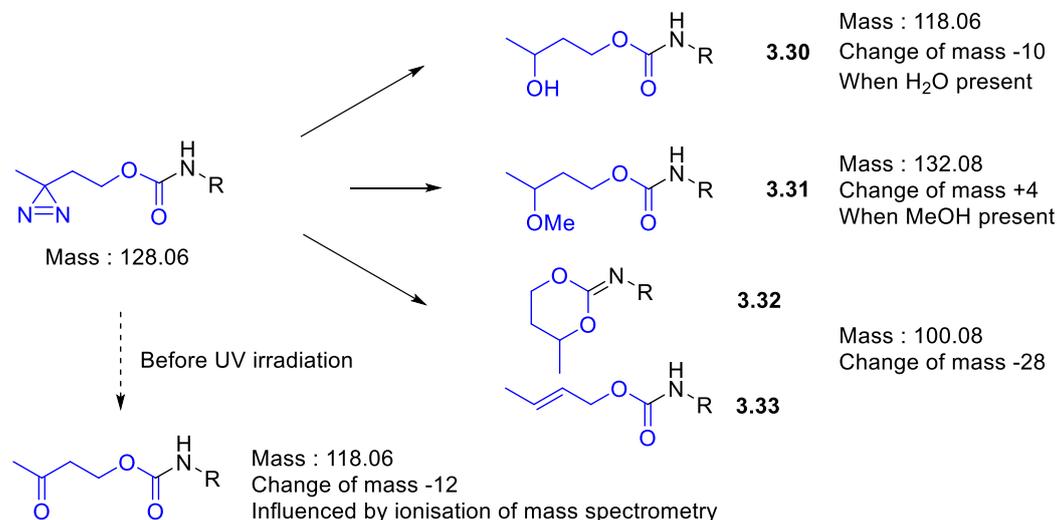


Figure 37: Photoreactivity of alkyl diazine under UV depending on solvents.

### 3.3.3.b Reactivity over time

A second test was then performed over time to find an optimal time to reach high conversion for alkyl diazine labelled peptides. This test was achieved using the mixture MeOH/H<sub>2</sub>O allowing better solubility of the labelled peptide. The peptide use for this experiment was peptide P2 labelled at its C-terminus. Analysis was performed after irradiation for 5 min, 15 min and 25 min, and results are shown in Figure 38, representing the mass changes from the diazine compound (Mw 1042.5 g/mol, in purple in Figure 38) to the elimination side product, either **3.32** or **3.33** (Mw 1014.5 g/mol, in orange).

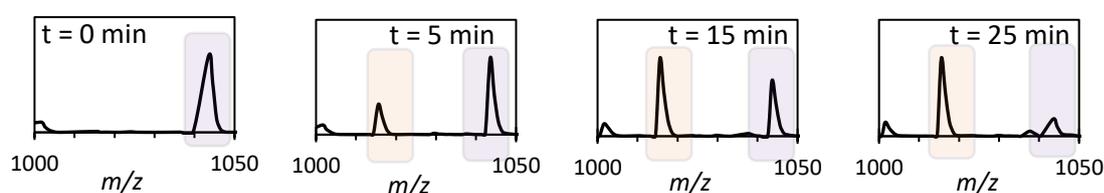


Figure 38: LCMS zoomed-in spectra showing the conversion of peptide P2-Alk-Dz (in purple) over 25 min under irradiation in a mixture 50/50 MeOH/H<sub>2</sub>O leading to the formation of elimination side product, either **3.32** or **3.33** (in orange).

Full consumption of the diazine was reached at 25 min. For future use of peptides containing diazirines, to ensure good cross-linking with the protein 30 min irradiation was used.

In these tests, we found that the volume of the solution was decreasing over time due to the heat generated from the lamp. Therefore, another approach to diazine irradiation was tried in order to

have more control and consistency between the samples. The samples were directly prepared and sealed in a 100  $\mu\text{L}$  LCMS insert vial, and the vials laid down on a support as shown in the Experimental Section. Measurements were taken after 10, 20, 30 and 40 min, as it was hypothesized that due to the double layer of glass within the vials it might take longer for the diazirine to be fully consumed. However, the LCMS results showed that at 10 min, less than 5% of the starting material was present. The solution remaining was diluted again to add more points at 2, 4, 6, and 8 min and the conversion rate is plotted in Figure 39.

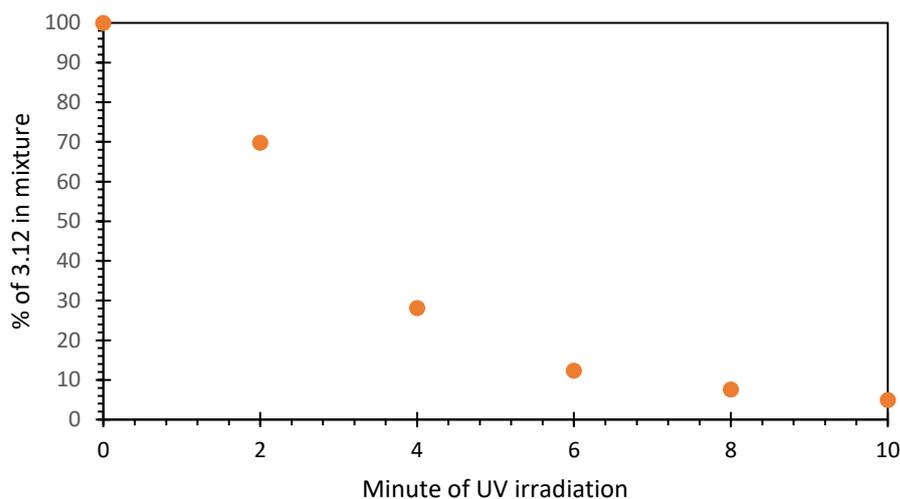


Figure 39: Consumption of diazirine probe **3.12** on peptide P2 over time using a Rio UV curing instrument.

A similar test was then run on a peptide labelled with the aromatic diazirine, **3.13**, sequence P6 on its *N*-terminus. At 10 min, it was shown that 68% of the diazirine was consumed with only the hydroxyl product being formed. This preferential hydroxyl labelling supports the hypothesis that compound **3.13** was much more stable than **3.12** on peptide which could be an advantage to ensure the reactivity toward the protein and avoid loss of reactive reagents due to reactions with solvents.

#### 3.3.4 Labelling of proteins using diazirine labelled peptides.

The final step was to use the diazirine labelled peptides to attempt to label the target proteins. The first attempt was done using P2 and insulin, as this was the strongest binding peptide. P2 was labelled with compound **3.12** on its *N*-terminus. Two first attempts with 2 or 20 equivalents of peptide in PBS relative to insulin (25  $\mu\text{M}$ ) were achieved. However, no difference in protein mass was observed between the sample before and after UV irradiation. The equivalents were then increased to 60, leading to the appearance of new MS peaks within the co-eluting peptide/protein peaks at 958 and 1115  $m/z$ , Figure 40. While those two peaks are less than 1% of the highest intense peak (peptide peak), their masses correspond to what would be expected after the addition of the peptide onto the

protein, with a shift of 888 as presented in Table 1. The rest of the peptide reacted with water or underwent elimination, with the presence of no residual Dz-P2 peaks.

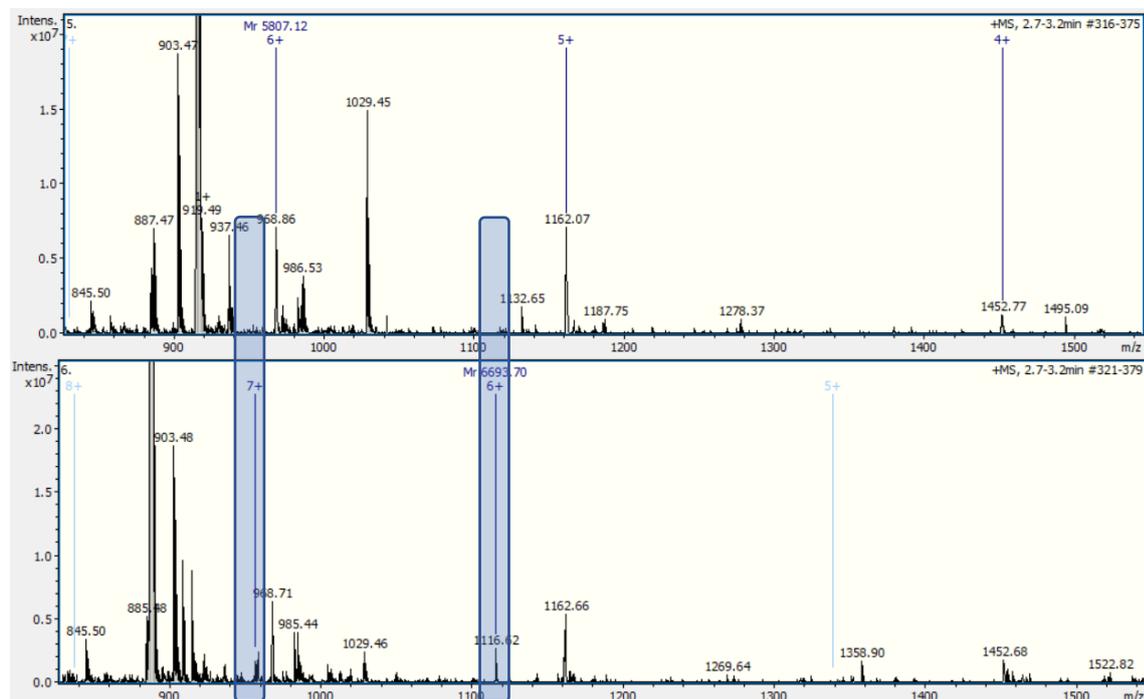


Figure 40: Spectrograms of co-eluting peptide and protein peak showing before UV (top) and after 15 min UV irradiation (bottom) showing the deconvoluted masses of the unmodified and modified insulin.

Table 1: Summary of the observed insulin peaks observed when in presence of peptide P2 before and after UV irradiation (15 min).

m/z	Unmodified insulin	Modified insulin	Delta
+7	Not seen	957.18	N.A
+6	968.86	1116.62	N.A
+5	1162.07	Not seen	N.A
<b>Deconvoluted</b>	<b>5805.3</b>	<b>6693.7</b>	<b>888.4</b>

While these first attempts were very inefficient the cross-linking seemed to be possible and further optimisation was conducted.

A second attempt was used with P2 peptide modified with compound **3.20** again at the N-terminus. 100 equivalents of peptide were used to try to increase the cross-coupling to insulin (50  $\mu$ M in PBS). While the peptide and protein peaks co-eluted again it was possible to see better protein peaks due to an increased protein concentration in this experiment. Again after 15 minutes irradiation three new peaks could be detected showing the change from a deconvoluted mass of 5807.06 to 6762.97, Figure 41, which matched the expected peptide addition (delta 955.9).

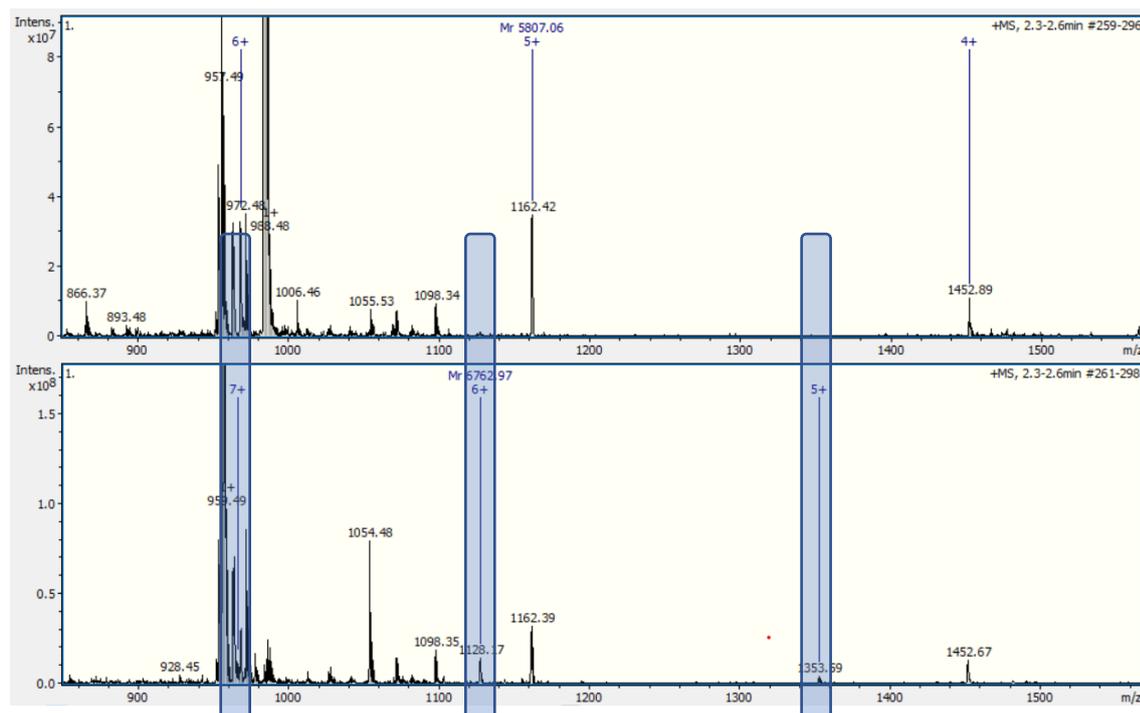


Figure 41: Spectrograms of co-eluting peptide and protein peak showing before UV (top) and after 15 min UV irradiation (bottom) showing the deconvoluted masses of the unmodified and modified insulin.

This second probe also allowed cross-linking onto the protein. Using the intensities of the mass peaks it is possible to assume ~30% of the protein was successful modified with peptide P2. However, this makes the assumption that ionisation potential is not significantly affected by crosslinking and is likely to be a major simplification.

A third experiment was set-up with the aromatic diazirine, **3.13**, to be able to compare its reactivity, however despite using the same concentrations of peptide as the second experiment, no protein peak could be observed in both samples (before and after irradiation) which was probably due to a mistake during manipulation.

Following crosslinking, the protein samples were sent for proteomic analysis to determine the site of labelling, results obtained thanks to Dr. Adam Dowle. However, at the time of writing this thesis this had not been successfully achieved, with precipitation of the protein prior to analysis leading to very weak signals in which neither modified or unmodified insulin could be detected.

### 3.3.5 Diazirine, a conclusion

It was shown that two different types of diazirine probes could be synthesised using ketone precursors as starting material which led to **3.12**, **3.20** for alkyl probes and **3.13** for an aromatic probe. Peptides were successfully modified with these probes using either the amidation protocol or the NHS-ester activated protocol from Chapter 2. The reactivity of each diazirine probe was tested on peptides and with the products produced being dependent on the solvent as suggested by literature. Time courses

suggested that the aromatic probe was more stable which could reduce unwanted side reactions with solvent. Moreover it was shown that alkyl diazirines could be oxidised by O<sub>2</sub> on storage, necessitating the use of inert atmosphere during storage to extend their lifetime.

Some first experiments were conducted to see the possibility of cross-linking to protein targets. Only work on insulin was undertaken due to time constraints, optimistic first results with alkyl diazirine probes showing cross-linking onto the protein. Further work was needed but due to lack of time at the end of the PhD, the focus was put on the ligand-directed chemistry which will be discussed in Chapter 6. In the future, if these studies were to be extended, a dialysis step after cross-linking would help to provide purer labelled protein samples that could be more readily analysed to determine the site of labelling.

### 3.4 Summary

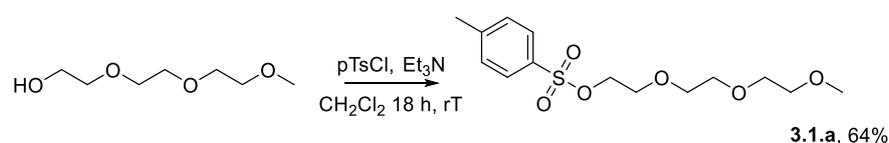
In this chapter, further analyses were done on the peptides generated in Chapter 2 to ensure their affinity to the protein they were synthesised to target. While QCM-D was the ideal assay for stable proteins which were available in reasonably high quantity, due to the ease of the preparation and as there was no need to modify the peptide prior to use, it was found that each protein exhibited different levels of biofouling and so this technique was not universally applicable. A second test, FP, decreased drastically the quantity of protein needed to assay binding, and was shown to be a faster method to perform due to the use of 96-well plates to test multiple peptides at the same time. Finally, using photo-chemistry with diazirine probes, we have shown the synthesis of three probes which were all successful tethered onto peptides. Following studies on their reactivity, initial attempts were conducted to cross-link the peptide to the protein. A highlight of this chapter was to learn so many different techniques, to develop a new route for achieving the synthesis of compound **3.13** in a reduced number of steps, which could be a real asset for bioconjugation, and showing the first protein modification of insulin via photocrosslinking.

### 3.5 Experimental

All the general conditions for peptide synthesis, analysis and purification (SPPS, LCMS and SEC/Combiflash) used throughout the thesis are fully described in Chapter 2 in the Experimental Section 2.5.1 to 2.5.6. This experimental section details the synthesis and testing of probes used in Chapter 3. The same UV lamp was used for TLC and for assay for diazirine its characteristics are UVGL-58 Handheld UV Lamp,  $\lambda_{\text{max}} = 254 \text{ nm}$ , 6 W, 230 V, 50-60 Hz.

#### 3.5.1 Chemical synthesis

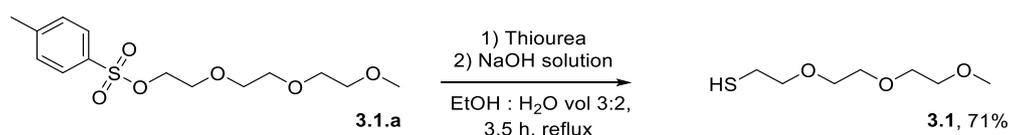
##### Compound 3.1.a



To a solution of triethylene glycol (2.0 g, 12.2 mol) in  $\text{CH}_2\text{Cl}_2$  (30 mL) was added  $\text{Et}_3\text{N}$  (2.5 mL, 17.9 mmol) and the mixture was cooled to 0 °C. A solution of *para*-toluene sulfonyl chloride (2.5 g, 13.1 mmol) in  $\text{CH}_2\text{Cl}_2$  (5 mL) was then added slowly over 5 min. The mixture was allowed to warm to room temperature and stirred for 18 h. The reaction was then quenched with hydrochloric acid (4 M, 20 mL) and extracted with  $\text{CH}_2\text{Cl}_2$  (3 × 40 mL). The organic layers were combined, dried over anhydrous  $\text{MgSO}_4$  and concentrated under reduced pressure to afford the crude product. The residue was purified by flash chromatography on silica gel, using 33% EtOAc in PET. Pure fractions were concentrated under reduced pressure to afford the product 2-(2-(2-hydroxyethoxy)ethyl)-4-methylbenzenesulfonate **3.1.a** (2.5 g, 7.86 mmol, 64%) as a pale-yellow liquid. Data were consistent with those previously reported.<sup>31</sup>

$^1\text{H NMR}$  (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  7.66 (d,  $J = 8.0 \text{ Hz}$ , 2H,  $-(\text{ArCH})_2\text{SO}_2$ ), 7.22 (d,  $J = 8.0 \text{ Hz}$ , 2H,  $-(\text{ArCH})_2\text{SO}_2$ ), 4.06-4.01 (m, 2H,  $-\text{CH}_2\text{OSO}_2$ ), 3.59-3.51 (m, 2H,  $-\text{CH}_2\text{PEG}$ ), 3.49-3.44 (m, 6H,  $-\text{CH}_2\text{PEG}$ ), 3.41-3.37 (m, 2H,  $-\text{CH}_2\text{PEG}$ ), 3.23 (s, 3H,  $-\text{OCH}_3$ ), 2.32 (s, 3H,  $-\text{CH}_3$  tosylate). **HRMS** (ESI<sup>+</sup>):  $m/z$  calcd for  $\text{C}_{14}\text{H}_{22}\text{NaO}_6\text{S}$ : 341.1029  $[\text{M}+\text{Na}]^+$ ; found 341.1022

##### Compound 3.1

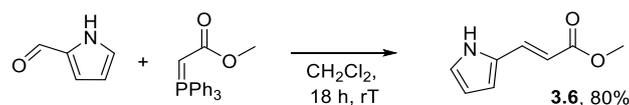


To a reaction flask were added 2-(2-(2-hydroxyethoxy)ethyl)-4-methylbenzenesulfonate (1.50 g, 4.72 mmol), EtOH (3 mL), thiourea (360 mg, 4.74 mmol) and  $\text{H}_2\text{O}$  (2 mL). The reaction was refluxed for 3 h. A solution of NaOH (230 mg, 5.75 mmol) in  $\text{H}_2\text{O}$  (2 mL) was added, and reflux continued for 1.5 h.

The reaction mixture was neutralized with concentrated HCl and extracted with CH<sub>2</sub>Cl<sub>2</sub> (3 × 5 mL). The organic layers were combined, dried over anhydrous MgSO<sub>4</sub> and concentrated under reduced pressure to afford the crude product **3.1** (0.60 g, 3.33 mmol, 71%) yield as a colorless liquid used without further purification. Data were consistent with those previously reported.<sup>32</sup>

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>): δ 3.67-3.51 (m, 10H, -CH<sub>2</sub> PEG), 3.37 (s, 3H, -OCH<sub>3</sub>), 2.68 (dt, *J* = 8.1, 6.3 Hz, 2H, -CH<sub>2</sub>SH), 1.57 (t, *J* = 8.1 Hz, 1H, -SH). **HRMS** (ESI<sup>+</sup>): *m/z* calcd for C<sub>7</sub>H<sub>16</sub>NaO<sub>3</sub> 203.0712 [M+Na]<sup>+</sup>; found 203.0711

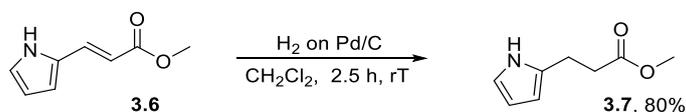
### Compound 3.6



To a stirred solution of pyrrole-2-carbaldehyde (1.5 g, 15.8 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (20 mL) was added a solution of methyl (triphenylphosphoranylidene)acetate (10.6 g, 31.7 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (50 mL). The resulting mixture was stirred at room temperature for 18 h. The residual solvent was evaporated under reduced pressure and the crude product was purified by flash chromatography using 33% EtOAc in PET. Pure fractions were concentrated under reduced pressure to afford the product **3.6** (2.0 g, 13.2 mmol, 80%) as a pale-pink solid. Data were consistent with those previously reported.<sup>20</sup>

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>): δ 9.60 (app s, 1H, -NH), 7.63 (d, *J* = 16.0 Hz, 1H, -CH=CHCOOMe), 6.93 (dd, *J* = 3.0, 1.5 Hz, 1H, -NHCH), 6.58 (t, *J* = 3.0, 1H, -NHCHCH), 6.28 (d, *J* = 3.0, 1H, -NHCHCHCH), 6.15 (d, *J* = 16.0 Hz, 1H, -CH=CHCOOMe), 3.78 (s, 3H, -Me). **HRMS** (ESI<sup>+</sup>): *m/z* calcd for C<sub>8</sub>H<sub>9</sub>NNaO<sub>2</sub> 174.0525 [M+Na]<sup>+</sup>; found 174.0527

### Compound 3.7

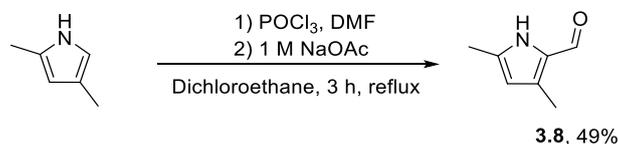


**3.6** (0.19 g, 1.26 mmol) was dissolved in MeOH (10 mL) and Pd/C (5% loading, 91 mg, 0.04 mmol) was added. The reaction was then put under a H<sub>2</sub> atmosphere using two balloons for 2.5 h. After this time the reaction mixture was diluted with MeOH (20 mL) and filtered through a bed of Celite, which was washed thoroughly with MeOH (3 × 15 mL). The solvent was concentrated under reduced pressure and the crude pale-yellow product was then used without further purification (0.15 g, 1.01 mmol, 80%). Data were consistent with those previously reported.<sup>20</sup>

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>): δ 8.55 (s, 1H, -NH), 6.67 (dt, *J* = 2.8, 1.5 Hz, 1H, -NHCH), 6.10 (dd, *J* = 2.8 Hz, 1H, -NHCHCH), 5.91 (m, 1H, -NHCHCHCH), 3.70 (s, 3H, -Me), 2.92 (t, *J* = 6.8 Hz, 2H, -CH<sub>2</sub>CH<sub>2</sub>COOMe),

2.65 (t,  $J = 6.8$  Hz, 2H,  $-\text{CH}_2\text{COOMe}$ ). **HRMS** (ESI<sup>+</sup>):  $m/z$  calcd for  $\text{C}_8\text{H}_{11}\text{NNaO}_2$  176.0682 [M+Na]<sup>+</sup>; found 176.0683

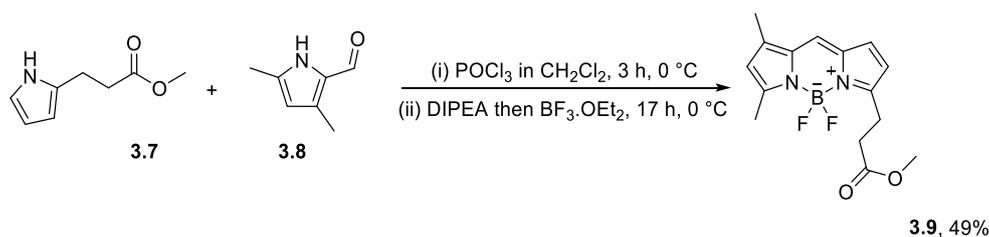
### Compound 3.8



$\text{POCl}_3$  (0.9 mL, 10.0 mmol) was added dropwise to DMF (0.5 mL) at 0 °C. The mixture was warmed to room temperature and stirred for 15 min. The mixture was then cooled back to 0 °C, and 1,2-dichloroethane (5 mL) was added to the mixture. A solution of 2,4-dimethyl pyrrole (0.8 mL, 7.4 mmol) in 1,2-dichloroethane (5 mL) was added dropwise over 20 min at 0 °C. After the addition was complete, the reaction mixture was refluxed for 3 h and then cooled to room temperature. A solution of NaOAc (1 M, 10 mL) was added dropwise. The reaction mixture was then again refluxed for 30 min. The cooled mixture was washed with  $\text{H}_2\text{O}$  (10 mL), then saturated  $\text{Na}_2\text{CO}_3$  (10 mL) and finally brine (5 mL). The organic layer was dried over  $\text{Na}_2\text{SO}_4$  and the solvents were concentrated under reduced pressure. The resulting crude material was purified by flash chromatography using 30% EtOAc in PET. Pure fractions were concentrated under reduced pressure to afford the product **3.8** (0.44 g, 3.6 mmol, 49%) as a green solid. Data were consistent with those previously reported.<sup>33</sup>

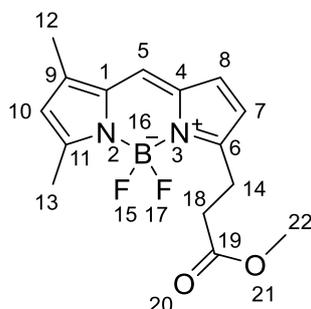
**<sup>1</sup>H NMR** (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  10.9 (s, 1H,  $-\text{NH}$ ), 9.44 (s, 1H,  $-\text{COH}$ ), 5.84 (s, 1H,  $-\text{ArH}$ ), 2.31 (s, 3H,  $-\text{ArCH}_3$ ), 2.30 (s, 3H,  $-\text{ArCH}_3$ ). **HRMS** (ESI<sup>+</sup>):  $m/z$  calcd for  $\text{C}_7\text{H}_9\text{NNaO}$  146.0576 [M+Na]<sup>+</sup>; found 146.0578

### Compound 3.9



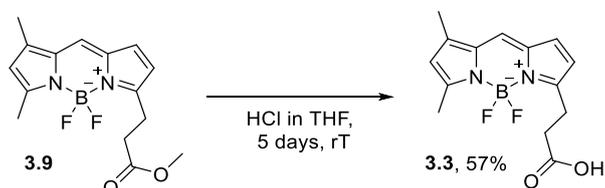
Methyl 3-(1H-pyrrol-2-yl)propanoate **3.7** (155 mg, 1.0 mmol) was dissolved in  $\text{CH}_2\text{Cl}_2$  (10 mL) at 0 °C. 3,5-dimethylpyrrole-2-carboxaldehyde **3.8** (138 mg, 1.1 mmol) was then added slowly to the mixture. After 10 minutes  $\text{POCl}_3$  (0.15 mL, 1.6 mmol) was added dropwise and then the reaction mixture was allowed to warm to room temperature for 2.5 h. The reaction mixture was then cooled down to 0 °C again and *N,N*-diisopropylethylamine (0.8 mL, 4.5 mmol) was added in a series of 0.1 mL additions, followed by stirring for 30 min at 0 °C. Then,  $\text{BF}_3\cdot\text{OEt}_2$  (0.5 mL, 4.0 mmol) was added, and the reaction was warmed to room temperature and stirred for 17 h. The mixture was diluted with  $\text{CH}_2\text{Cl}_2$  (10 mL) and brine (10 mL). The mixture was filtered through a bed of Celite and then the organic phase was

washed with brine (3 x 20 mL). The organic layers were combined, dried over anhydrous MgSO<sub>4</sub> and concentrated under reduced pressure to afford the crude product. The resulting crude material was purified by flash chromatography using 100% CH<sub>2</sub>Cl<sub>2</sub>. Pure fractions were concentrated under reduced pressure to afford the product **3.9** (150 mg, 0.49 mmol, 49%) as a green-red solid. Data were consistent with those previously reported.<sup>20</sup>



**<sup>1</sup>H NMR** (500 MHz, CDCl<sub>3</sub>): δ 7.08 (s, 1H, -ArH<sub>5</sub>), 6.89 (d, *J* = 4.1 Hz, 1H, -ArH<sub>8</sub>), 6.27 (d, *J* = 4.1 Hz, 1H, -ArH<sub>7</sub>), 6.11 (s, 1H, -ArH<sub>10</sub>), 3.69 (s, 3H, -OCH<sub>3</sub>), 3.29 (t, *J* = 7.6 Hz, 2H, -COCH<sub>2</sub>CH<sub>2</sub>), 2.77 (t, *J* = 7.6 Hz, 2H, -COCH<sub>2</sub>), 2.56 (s, 3H, -Ar11CH<sub>3</sub>), 2.25 ppm (s, 3H, -Ar9CH<sub>3</sub>); assignment above used for all further BODIPY FL core structures. **<sup>19</sup>F NMR** (400 MHz, CDCl<sub>3</sub>): δ 145.19 (d, *J* = 35.4 Hz), 145.37 ppm (d, *J* = 35.4 Hz). **HRMS** (ESI<sup>+</sup>): *m/z* calcd for C<sub>15</sub>H<sub>17</sub>BF<sub>2</sub>N<sub>2</sub>NaO<sub>2</sub> 329.1243 [M+Na]<sup>+</sup>; found: 329.1243

### Compound 3.3

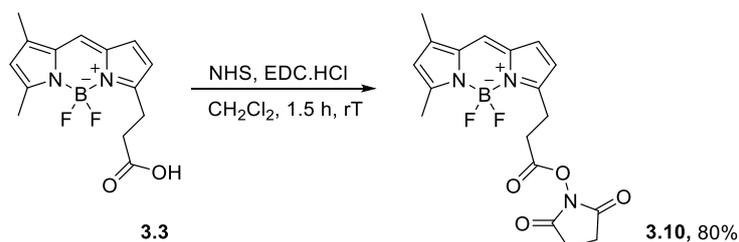


To a stirred solution of BODIPY-ester **3.9** (27 mg, 88 μmol) in THF (3 mL) at 0 °C was added water (2 mL) and then concentrated HCl (1 mL). The reaction mixture was stirred at room temperature for 5 days. The reaction mixture was diluted with water (20 mL) and extracted with CH<sub>2</sub>Cl<sub>2</sub> (3 x 10 mL). The organic layers were combined, dried over anhydrous MgSO<sub>4</sub>, and concentrated under reduced pressure to afford the crude product. The resulting crude material was purified by flash chromatography using 1% AcOH in CH<sub>2</sub>Cl<sub>2</sub>. Pure fractions were concentrated under reduced pressure to afford the product **3.3** (15 mg, 51 μmol, 58%) as a bright red crystalline solid. Data were consistent with those previously reported.<sup>20</sup>

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>): δ 7.09 (s, 1H, -ArH<sub>5</sub>), 6.87 (d, *J* = 4.1 Hz, 1H, -ArH<sub>8</sub>), 6.27 (d, *J* = 4.1 Hz, 1H, -ArH<sub>7</sub>), 6.11 (s, 1H, -ArH<sub>10</sub>), 3.29 (t, *J* = 7.6 Hz, 2H, -COCH<sub>2</sub>CH<sub>2</sub>), 2.82 (t, *J* = 7.6 Hz, 2H, -COCH<sub>2</sub>), 2.56 (s, 3H, -Ar11CH<sub>3</sub>), 2.25 ppm (s, 3H, -Ar9CH<sub>3</sub>); **<sup>19</sup>F NMR** (400 MHz, CDCl<sub>3</sub>): δ 145.16 (d, *J* = 35.4 Hz),

145.35 ppm (d,  $J = 35.4$  Hz); **HRMS** (ESI<sup>+</sup>):  $m/z$  calcd for C<sub>14</sub>H<sub>15</sub>BF<sub>2</sub>N<sub>2</sub>NaO<sub>2</sub> 315.1087 [M+Na]<sup>+</sup>; found: 315.1084

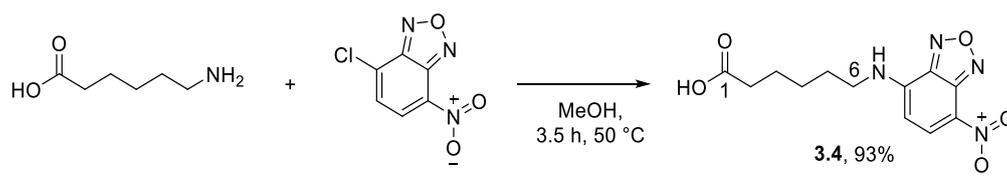
### Compound 3.10



To a solution of **3.3** (10.8 mg, 0.037 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (5 mL) was added *N*-hydroxysuccinimide (6.5 mg, 0.056 mmol) and then EDC.HCl (12.8 mg, 0.066 mmol). The reaction mixture was stirred for 2 h at room temperature. The mixture was diluted with CH<sub>2</sub>Cl<sub>2</sub> (20 mL) then washed with water (3 x 20 mL). The organic phase was collected, dried over MgSO<sub>4</sub> and concentrated under reduced pressure to afford product **3.10** (11 mg, 0.028 mmol, 80%) as a red solid. Data were consistent with those previously reported.<sup>34</sup>

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>): δ 7.09 (s, 1H, -ArH<sub>5</sub>), 6.88 (d,  $J = 4.1$  Hz, 1H, -ArH<sub>8</sub>), 6.27 (d,  $J = 4.1$  Hz, 1H, -ArH<sub>7</sub>), 6.12 (s, 1H, -ArH<sub>10</sub>), 3.37 (t,  $J = 7.3$  Hz, 2H, -COCH<sub>2</sub>CH<sub>2</sub>), 3.07 (t,  $J = 7.3$  Hz, 2H, -COCH<sub>2</sub>), 2.82 (s, 4H, -OSu), 2.56 (s, 3H, -Ar11CH<sub>3</sub>), 2.24 ppm (s, 3H, -Ar9CH<sub>3</sub>); **<sup>19</sup>F NMR** (400 MHz, CDCl<sub>3</sub>): δ 146.01 (d,  $J = 34$  Hz), 146.18 ppm (d,  $J = 34$  Hz); **HRMS** (ESI<sup>+</sup>):  $m/z$  calcd for C<sub>18</sub>H<sub>18</sub>BF<sub>2</sub>N<sub>3</sub>NaO<sub>4</sub> 412.1251 [M+Na]<sup>+</sup>; found: 412.1251.

### Compound 3.4

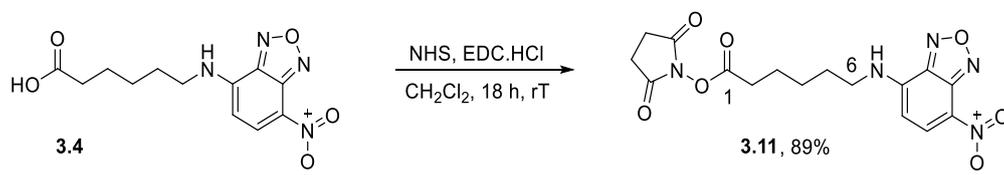


To a solution of 6-aminohexanoic acid (602 mg, 4.6 mmol) in methanol (32 mL) in an amber flask for light exclusion, was added 4-chloro-7-nitrobenzo-2-oxa-1,3-diazole (924 mg, 4.6 mmol) and NaHCO<sub>3</sub> (1.16 g, 13.8 mmol). The reaction mixture was stirred under light exclusion at 50 °C for 3.5 h. After cooling to room temperature, the organic layers were combined, dried over anhydrous MgSO<sub>4</sub> and concentrated under reduced pressure to afford **3.4** (1.25 g, 4.3 mmol, 93%) as a brown solid used without further purification. Data were consistent with those previously reported.<sup>35</sup>

**<sup>1</sup>H NMR** (400 MHz, *d*-DMSO): δ 9.55 (app s, 1H, -NH), 8.50 (d,  $J = 9.0$  Hz, 1H, -CH arom), 6.40 (d,  $J = 9.0$  Hz, 1H, -CH arom), 3.46 (d,  $J = 6.9$  Hz, 2H, -CH<sub>2</sub>(6)), 2.21 (t,  $J = 7.3$  Hz, 2H, -CH<sub>2</sub>(2)), 1.68 (tt,  $J_1 = J_2 = 7.4$

Hz, 2H, -CH<sub>2</sub>(5)), 1.54 (tt,  $J_1 = J_2 = 7.4$  Hz, 2H, -CH<sub>2</sub>(3)), 1.37 (tt,  $J = 7.4$  Hz, 2H, -CH<sub>2</sub>(4)). **HRMS** (ESI<sup>+</sup>):  $m/z$  calcd for C<sub>12</sub>H<sub>13</sub>N<sub>4</sub>O<sub>5</sub>: 293.0891 [M+H]<sup>+</sup>; found: 293.0892

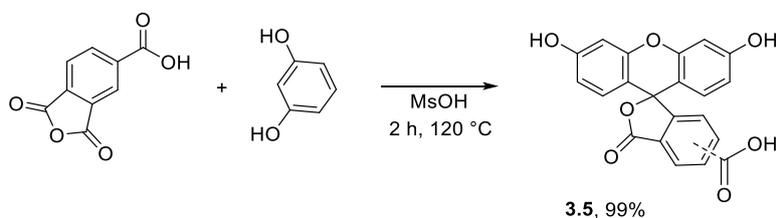
### Compound 3.11



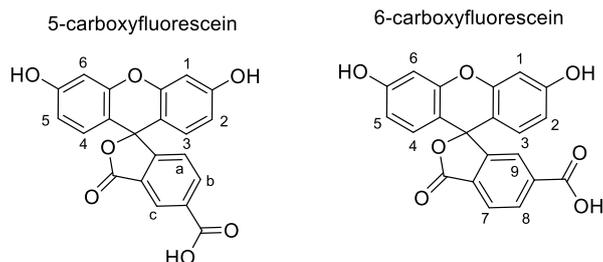
To a solution of **3.4** (0.53 g, 1.8 mmol) and *N*-hydroxysuccinimide (0.25 g, 2.2 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (15 mL) was added 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (0.52 g, 2.7 mmol). The resultant solution was allowed to stir 18 h at room temperature, after which time the mixture was washed with water (3 x 50 mL). The organic layer was then dried over anhydrous MgSO<sub>4</sub> and under reduced pressure to afford the crude product. The crude residue **3.11** (0.62 g, 1.6 mmol, 89%) was used without further purification. Data were consistent with those previously reported.<sup>36</sup>

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>): δ 8.48 (d,  $J = 8.6$  Hz, 1H, -CH arom), 6.45 (s, 1H, -NH), 6.17 (d,  $J = 8.6$  Hz, 1H, -CH arom), 3.53 (q,  $J = 6.5$  Hz, 2H, -CH<sub>2</sub>(6)), 2.86 (s, 4H, -OSu), 2.66 (t,  $J = 6.9$  Hz, 2H, -CH<sub>2</sub>(2)), 1.91-1.80 (m, 4H, -CH<sub>2</sub>(5), CH<sub>2</sub>(3)), 1.65-1.60 (m, 2H, -CH<sub>2</sub>(4)). **HRMS** (ESI<sup>+</sup>):  $m/z$  calcd for C<sub>16</sub>H<sub>18</sub>N<sub>5</sub>O<sub>7</sub>: 392.1201 [M+H]<sup>+</sup>; found: 392.1214

### Compound 3.5

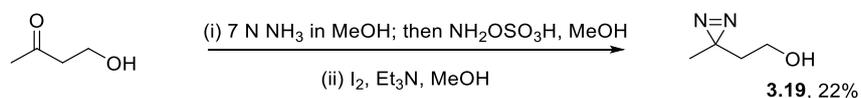


1,2,4-Benzoxycarboxylic anhydride (1.94 g, 10.0 mmol) and resorcinol (2.32 g, 21.0 mmol) were refluxed in methanesulfonic acid (15 mL) for 2 h at 120 °C. After cooling to room temperature, the mixture was added to 300 mL of rapidly stirred ice-water. The resultant brown precipitate was recovered by vacuum filtration and washed with water (50 mL). The precipitate was dissolved in NaOH (4 M, 25 mL) and acidified with conc. H<sub>2</sub>SO<sub>4</sub> until a light brown precipitate formed. 5(6)-carboxyfluorescein was recovered by vacuum filtration. The product was air-dried for 4 h to give **3.5** (4.1 g, 9.9 mmol, 99%) as a brown solid. Data were consistent with those previously reported.<sup>37,38</sup>



$^1\text{H NMR}$  (400 MHz,  $\text{DMSO-}d_6$ ): mixture of 5-6 isomers  $\delta$  8.35 (s, 1H,  $-\text{CH}(c)$ ), 8.29-8.22 (m, 1H,  $-\text{CH}(b)$ ), 8.18 (d,  $J = 8.0$  Hz, 1H,  $-\text{CH}(8)$ ), 8.05 (d,  $J = 8.0$  Hz, 1H,  $-\text{CH}(7)$ ), 7.57 (s, 1H,  $-\text{CH}(9)$ ), 7.31 (d,  $J = 8.0$  Hz, 1H,  $-\text{CH}(a)$ ), 6.65 (t,  $J = 2.5$  Hz, 4H,  $-\text{CH}(1)$  and  $-\text{CH}(6)$ ), 6.52 (qd,  $J = 8.7, 2.5$  Hz, 8H,  $-\text{CH}(5)\text{CH}(4)$  and  $-\text{CH}(2)\text{CH}(3)$ ). **HRMS** (ESI<sup>+</sup>):  $m/z$  calcd for  $\text{C}_{21}\text{H}_{13}\text{O}_7$ : 377.0656 [M+H]<sup>+</sup>; found: 377.0660.

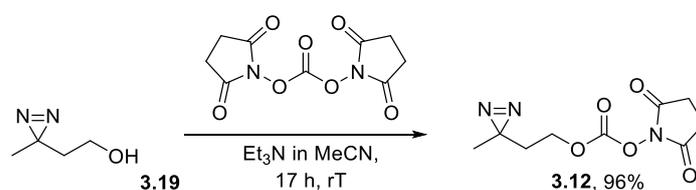
### Compound 3.19



4-hydroxy-2-butanone (2 mL, 23.2 mmol) was cooled to  $-78$  °C. 7 N  $\text{NH}_3$  in MeOH (14 mL) was added slowly and stirred for 3 h warming up to  $-60$  °C. The solution was cooled again to  $-78$  °C before the addition of a solution of hydroxylamine-*O*-sulfonic acid (3.0 g, 26.5 mmol) in MeOH (10 mL). The reaction mixture was allowed to warm up to room temperature for 18 h. The mixture was then filtered, and the solid residue washed with MeOH (2 x 20 mL). Half of the MeOH was then evaporated under reduced pressure.  $\text{Et}_3\text{N}$  (3.6 mL, 25.6 mmol) was added to the resulting solution followed by iodine in several portions. After adding 3.0 g (11.8 mmol) of  $\text{I}_2$ , the red colour of the iodine persisted for more than 10 min, indicating the end of the reaction. The solvents were carefully removed from the reaction mixture under reduced pressure (25 °C and 120 mbar) and the residue was partitioned between  $\text{Et}_2\text{O}$  (20 mL) and  $\text{NaHSO}_3$  (20 mL, 10 g / 100 mL solution). The organic layer was separated and the aqueous layer was extracted with  $\text{Et}_2\text{O}$  (3 x 10 mL). The organic layers were combined, dried over  $\text{Na}_2\text{SO}_4$  and concentrated under reduced pressure (25 °C and 120 mbar) to afford product **3.19** as an orange liquid (511 mg, 5.1 mmol, 22%). Data were consistent with those previously reported.<sup>26</sup>

$^1\text{H NMR}$  (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  3.40 (t,  $J = 6.2$  Hz, 2H,  $-\text{CH}_2\text{OH}$ ), 1.49 (t,  $J = 6.2$  Hz, 2H,  $-\text{CH}_2\text{CN}_2$ ), 0.95 (s, 3H,  $-\text{Me}$ ).

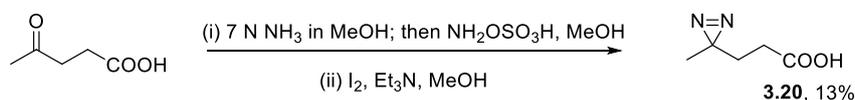
### Compound 3.12



To a solution of **3.19** (172 mg, 1.7 mmol) in anhydrous acetonitrile (10 mL) was added *N,N'*-disuccinimidyl carbonate (660 mg, 2.6 mmol) and Et<sub>3</sub>N (1.5 mL). The mixture was stirred at room temperature under N<sub>2</sub> atmosphere for 17 h. The solvents were carefully removed from the reaction mixture under reduced pressure (30 °C, 90 mbar). The resulting crude material was purified by flash chromatography using 5% EtOAc in CH<sub>2</sub>Cl<sub>2</sub>. Pure fractions were concentrated under reduced pressure to afford the product **3.12** (439 mg, 1.64 mmol, 96% yield) as a light yellow oil. Data were consistent with those previously reported.<sup>26</sup>

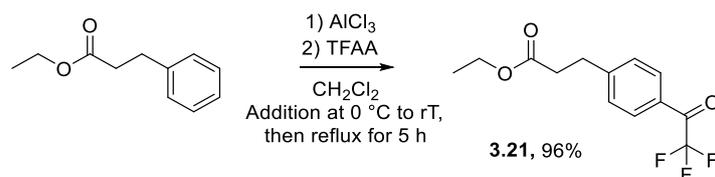
<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>): δ 4.23 (t, *J* = 6.4 Hz, 2H, -CH<sub>2</sub>O), 2.80 (s, 4H, -OSu), 1.72 (t, *J* = 6.4 Hz, 2H, -CH<sub>2</sub>CH<sub>2</sub>O), 1.04 (s, 3H, -CH<sub>3</sub>); HRMS (ESI<sup>+</sup>): *m/z* calcd for C<sub>9</sub>H<sub>11</sub>N<sub>3</sub>NaO<sub>5</sub>: 264.0591 [M+Na]; found: 264.0583

### Compound 3.20

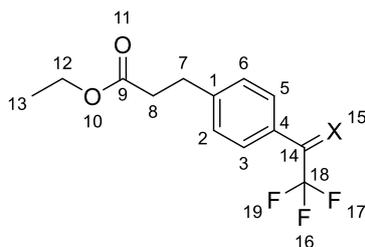


Levulinic acid (2.0 g, 17.2 mmol) was cooled to -78 °C before 7 N NH<sub>3</sub> in MeOH (14 mL) was added slowly and stirred for 4 h, warming up to -60 °C. The solution was cooled again to -78 °C before the addition of a solution of hydroxylamine-*O*-sulfonic acid (3.0 g, 26.5 mmol) in MeOH (10 mL) over 20 min. The reaction mixture was allowed to warm to room temperature for 18 h. The mixture was then filtered, and the solid residue washed with MeOH (2 x 20 mL). Half of the MeOH was then evaporated under reduced pressure. Et<sub>3</sub>N (4.0 mL, 28.7 mmol) was added to the resulting solution at 0 °C, followed by iodine in several portions. After adding 2.3 g (9.2 mmol) of I<sub>2</sub>, the red colour of the iodine persisted for more than 10 min, indicating the end of the reaction. The solvents were carefully removed from the reaction mixture under reduced pressure (30 °C and 120 mbar) and the residue was diluted in EtOAc (20 mL). The solution was washed with 1 M HCl (2 x 20 mL) and then NaHSO<sub>3</sub> (2 x 20 mL, 10 g / 100 mL solution). The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated under reduced pressure (30 °C and 80 mbar) to afford product **3.20** as a yellow oil (280 mg, 2.2 mmol, 13%). Data were consistent with those previously reported.<sup>28</sup>

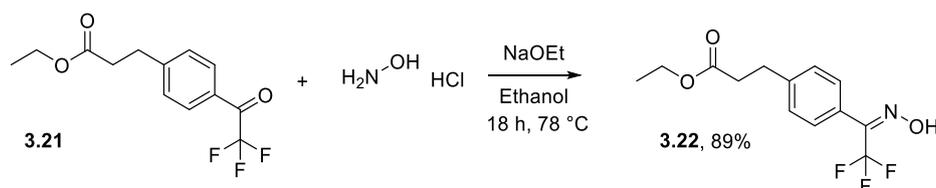
<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>): δ 2.22 (t, *J* = 7.6 Hz, 2H, -CH<sub>2</sub>COOH), 1.71 (t, *J* = 7.6 Hz, 2H, -CH<sub>2</sub>CN<sub>2</sub>), 1.06 (s, 3H, -Me).

**Compound 3.21**

Ethyl hydrocinnamate (3.0 mL, 17.0 mmol) was dissolved in dry  $\text{CH}_2\text{Cl}_2$  (10 mL) under argon. The flask was then cooled to  $0\text{ }^\circ\text{C}$  and  $\text{AlCl}_3$  (8.0 g, 60.2 mmol) was added, followed by trifluoroacetic anhydride (5 mL, 35.7 mmol). Dry  $\text{CH}_2\text{Cl}_2$  (10 mL) was then added to ensure dissolution of the foam forming. The reaction mixture was allowed to stir for 30 min then refluxed for 5 hours. The mixture was then poured onto ice water (17 g of ice / 10 mL of water) slowly and extracted with  $\text{CH}_2\text{Cl}_2$  (2 x 20 mL). The combined organic phases were washed with brine (20 mL) then dried over anhydrous  $\text{MgSO}_4$ , and concentrated under reduced pressure to afford the crude product. The resulting crude material was purified by flash chromatography using a gradient of 10-20% EtOAc in PET. Pure fractions were concentrated under reduced pressure to afford the product **3.21** (4.5 g, 16.3 mmol, 96%) as a light-yellow oil.



$^1\text{H NMR}$  (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  7.99 (d,  $J = 8.0$  Hz, 2H, -CH(3 and 5)), 7.38 (d,  $J = 8.0$  Hz, 2H, -CH(2 and 6)), 4.12 (q,  $J = 7.1$  Hz, 2H, -CH(12)<sub>2</sub>), 3.04 (t,  $J = 7.6$  Hz, 2H, -CH(7)<sub>2</sub>), 2.66 (t,  $J = 7.6$  Hz, 2H, -CH(8)<sub>2</sub>), 1.22 (t,  $J = 7.1$  Hz, 3H, -CH<sub>3</sub>).  $^{13}\text{C NMR}$  (101 MHz,  $\text{CDCl}_3$ ):  $\delta$  172.31 (-C<sub>9</sub>), 149.38 (-C<sub>1</sub>), 130.56 (-C<sub>3</sub> and 5), 129.23 (-C<sub>2</sub> and 6), 124.11 (-C<sub>4</sub>), 60.79 (-C<sub>12</sub>), 35.06 (-C<sub>8</sub>), 31.08 (-C<sub>7</sub>), 14.25 (-C<sub>13</sub>). Figure above used for all further BODIPY FL core structures.  $^{19}\text{F NMR}$  (376 MHz,  $\text{CDCl}_3$ ):  $\delta$  -71.25. **HRMS** (ESI<sup>+</sup>):  $m/z$  calcd for  $\text{C}_{13}\text{H}_{14}\text{F}_3\text{O}_3$ : 275.0890 [M+H]; found: 275.0895

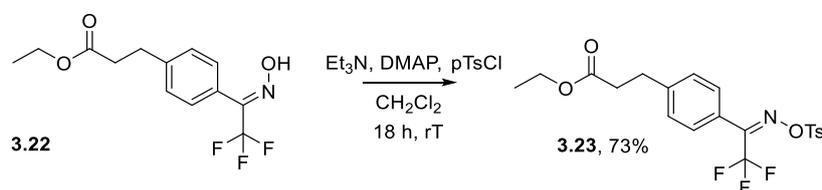
**Compound 3.22**

Hydroxylamine hydrochloride (260 mg, 3.74 mmol) in ethanol (40 ml) at room temperature was neutralized with sodium ethanolate in ethanol (260 mg, 3.82 mmol, in 5 mL). Product **3.21** (340 mg, 1.24 mmol) was then added to the reaction mixture and the mixture heated to  $78\text{ }^\circ\text{C}$  for 18 h. After

cooling to room temperature, Et<sub>2</sub>O (150 mL) was added. The organics were washed successively with HCl (0.01 M, 2 x 50 mL) and water (2 x 50 mL), dried over MgSO<sub>4</sub>, filtered and concentrated under reduced pressure to afford the oxime **3.22** (320 mg, 3.33 mmol, 89%) as a yellow oil.

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>): δ 8.50 (br s, 1H, -NOH), 7.47-7.41 (d, *J* = 8.2 Hz, 2H, -CH(3 and 5)), 7.30 (d, *J* = 8.2 Hz, 2H, -CH(2 and 6)), 4.12 (q, *J* = 7.1 Hz, 2H, -CH(12)<sub>2</sub>), 2.99 (t, *J* = 7.8 Hz, 2H, -CH(7)<sub>2</sub>), 2.64 (t, *J* = 7.8 Hz, 2H, -CH(8)<sub>2</sub>), 1.22 (t, *J* = 7.1 Hz, 3H, -CH<sub>3</sub>). <sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>): δ 172.83 (-C<sub>9</sub>), 143.43 (-C<sub>1</sub>), 128.91 (-C<sub>3</sub> and 5), 128.63 (-C<sub>2</sub> and 6), 124.11 (-C<sub>4</sub>), 60.71 (-C<sub>12</sub>), 35.51 (-C<sub>8</sub>), 30.88 (-C<sub>7</sub>), 14.25 (-C<sub>13</sub>). <sup>19</sup>F NMR (376 MHz, CDCl<sub>3</sub>): δ -66.45. HRMS (ESI<sup>+</sup>): *m/z* calcd for C<sub>13</sub>H<sub>15</sub>F<sub>3</sub>NO<sub>3</sub>: 290.0992 [M+H]; found: 290.0992

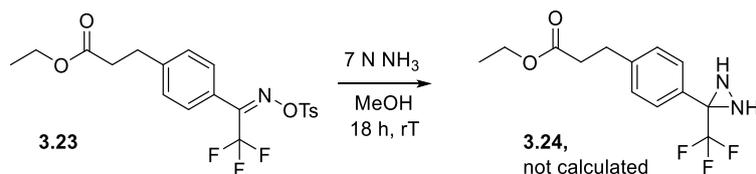
### Compound 3.23



To a solution of compound **3.22** (410 mg, 1.42 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (10 mL) was added successively triethylamine (0.4 mL, 2.87 mmol), DMAP (200 mg, 1.64 mmol) and pTsCl (300 mg, 1.57 mmol). The reaction mixture was stirred for 18 h at room temperature, and quenched with a saturated aqueous solution of NH<sub>4</sub>Cl (10 mL). The phases were separated and the organic phase was washed with H<sub>2</sub>O (30 mL), dried over MgSO<sub>4</sub>, filtered and concentrated under reduced pressure. The resulting crude material was purified by flash chromatography using 20% EtOAc in PET. Pure fractions were concentrated under reduced pressure to afford the product **3.23** (360 mg, 1.04 mmol, 73%) as a colorless oil.

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>): δ 7.87 (d, *J* = 8.3 Hz, 2H, -CH(3 and 5)), 7.37 (d, *J* = 8.3 Hz, 2H, -CH(2 and 6)), 7.34-7.28 (m, 4H, -CH<sub>arom</sub>, tosylate), 4.12 (q, *J* = 7.1 Hz, 2H, -CH(12)<sub>2</sub>), 2.98 (t, *J* = 7.8 Hz, 2H, -CH(7)<sub>2</sub>), 2.63 (dd, *J* = 8.2, 7.1 Hz, 2H, -CH(8)<sub>2</sub>), 2.46 (s, 3H, -CH<sub>3</sub> tosylate), 1.21 (t, *J* = 7.1 Hz, 3H, -CH(13)<sub>3</sub>). <sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>): δ 172.60 (-C<sub>9</sub>), 153.9 (-SO<sub>2</sub>C<sub>q</sub>), 146.24 (-C<sub>q</sub>CH<sub>3</sub> tosylate), 144.87 (-C<sub>1</sub>), 129.83 (-C<sub>3</sub> and 5), 129.36 (-C<sub>2</sub> and 6), 130.22-128.60 (m, -CH<sub>arom</sub> (tosyl)), 119.74 (-C<sub>4</sub>), 60.71 (-C<sub>12</sub>), 35.51 (-C<sub>8</sub>), 30.88 (-C<sub>7</sub>), 20.88 (-TsCH<sub>3</sub>), 13.9 (-C<sub>13</sub>). <sup>19</sup>F NMR (376 MHz, CDCl<sub>3</sub>): δ -66.58. HRMS (ESI<sup>+</sup>): *m/z* calcd for C<sub>20</sub>H<sub>21</sub>F<sub>3</sub>NO<sub>5</sub>S: 444.1087 [M+H]; found: 444.1101

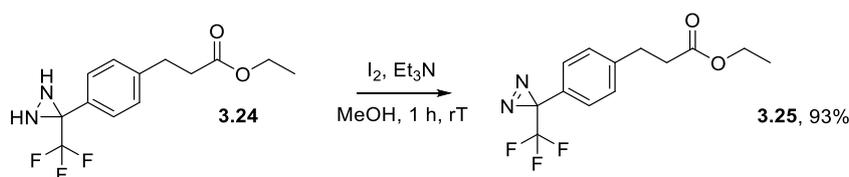
### Compound 3.24



To a solution of compound **3.23** (580 mg, 1.31 mmol) in THF (2 ml) was added ammonia (7 M in MeOH, 2.0 mL, 14.0 mmol) at room temperature. The reaction mixture was stirred 18 h, filtered, washed with MeOH (2 x 2 mL) and the filtrate concentrated under reduced pressure to give a yellow oil, **3.24** use directly for following step without full characterisation.

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>): δ 7.23 (d, *J* = 7.8 Hz, 2H, -CH(3 and 5)), 7.10 (d, *J* = 7.8 Hz, 2H, -CH(2 and 6)), 4.12 (q, *J* = 7.1 Hz, 2H, -CH(12)<sub>2</sub>), 2.95 (t, *J* = 7.8 Hz, 2H, -CH(7)<sub>2</sub>), 2.60 (dd, *J* = 7.8, 7.1 Hz, 2H, -CH(8)<sub>2</sub>), 1.22 (t, *J* = 7.1 Hz, 3H, -CH(13)<sub>3</sub>). **<sup>19</sup>F NMR** (376 MHz, CDCl<sub>3</sub>): δ -64.93. **HRMS** (ESI<sup>+</sup>): *m/z* calcd for C<sub>13</sub>H<sub>16</sub>F<sub>3</sub>N<sub>2</sub>O<sub>2</sub>: 289.1158 [M+H]; found: 289.1158

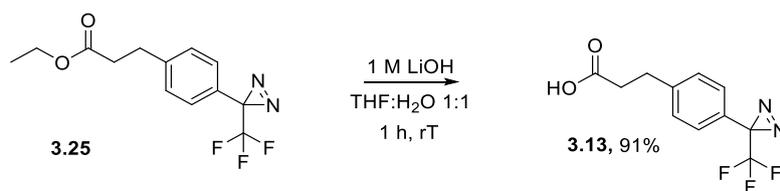
### Compound 3.25



To compound **3.24** (assumed 1.31 mmol) was added methanol (2 mL), triethylamine (280 μL, 2.01 mmol) and then small portions of iodine until the purple dark red remained. The reaction mixture was stirred at room temperature for 1 hour, Et<sub>2</sub>O (10 mL) was added, and the organic phase was washed successively with saturated solutions of sodium thiosulfate (10 mL) and ammonium chloride (10 mL), dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and concentrated under reduced pressure. The resulting crude material was purified by flash chromatography using 20% EtOAc in PET. Pure fractions were concentrated under reduced pressure to afford the product **3.25** (350 mg, 1.22 mmol, 93%) as a yellow oil. Data were consistent with literature.<sup>39</sup>

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>): δ 7.19 (d, *J* = 7.8 Hz, 2H, -CH(3 and 5)), 6.85 (d, *J* = 7.8 Hz, 2H, -CH(2 and 6)), 4.12 (q, *J* = 7.1 Hz, 2H, -CH(12)<sub>2</sub>), 2.91 (t, *J* = 7.8 Hz, 2H, -CH(7)<sub>2</sub>), 2.60 (dd, *J* = 7.8, 7.1 Hz, 2H, -CH(8)<sub>2</sub>), 1.22 (t, *J* = 7.1 Hz, 3H, -CH(13)<sub>3</sub>). **<sup>19</sup>F NMR** (376 MHz, CDCl<sub>3</sub>): δ -66.54. **HRMS** (ESI<sup>+</sup>): *m/z* calcd for C<sub>13</sub>H<sub>14</sub>F<sub>3</sub>N<sub>2</sub>O<sub>2</sub>: 287.1013 [M+H]; found: 289.1005

### Compound 3.13

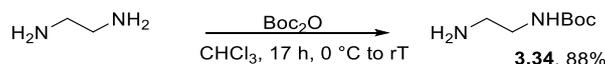


To a solution of compound **3.25** (220 mg, 0.77 mmol) in THF (5 mL) was added 1 M LiOH (5 mL) and the mixture was stirred at room temperature for 1 h. The THF was removed under reduced pressure at 150 mbar, 30 °C. The residue was dissolved in EtOAc (15 mL) which was washed by 1 M HCl (10 mL). The organic layer was then washed with water (10 mL). The organic phase was dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and concentrated under reduced pressure (30 °C) to give compound **3.13** (180 mg, 0.70 mmol, 91%) as a white solid used without further purification. Data were consistent with literature.<sup>27</sup>

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>): δ 7.24 (d, *J* = 8.1 Hz, 2H, -CH<sub>arom</sub>), 7.11 (d, *J* = 8.1 Hz, 2H, -CH<sub>arom</sub>), 2.95 (td, *J* = 7.7, 2.9 Hz, 2H, -CH<sub>2</sub>CH<sub>2</sub>COOH), 2.66 (td, *J* = 7.7, 2.5 Hz, 2H, -CH<sub>2</sub>CH<sub>2</sub>COOH). <sup>19</sup>F NMR (376 MHz, CDCl<sub>3</sub>): δ -65.16.

The following compounds were not discussed in the main discussion but used for peptide modification.

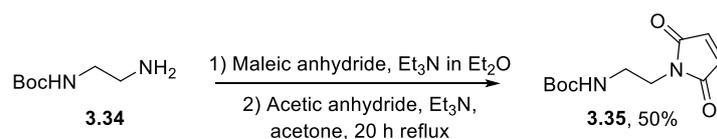
#### Compound 3.34



A solution of Boc-anhydride (4.35 g, 20 mmol) in CHCl<sub>3</sub> (100 mL) was added dropwise to an ice-cold solution of 1,2-diaminoethane (13.5 mL, 200 mmol) in CHCl<sub>3</sub> (200 mL) over 2.5 h. After complete addition, the mixture was stirred over an ice bath for 30 min and afterwards allowed to warm to room temperature overnight. The mixture was then separated in two and each part washed with water (2 x 100 mL), brine (100 mL) and finally with water again (100 mL). The organic phase was then dried over MgSO<sub>4</sub> and the solvent was evaporated to provide **3.34** as a yellow oil (2.81 g, 17.6 mmol, 88%). Data were consistent with those previously reported.<sup>40</sup>

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>): δ 4.87 (broad s, 1H, -NH), 3.07 (m, 2H, -CH<sub>2</sub>NHBoc), 2.79 (t, *J* = 6.0 Hz, 2H, -CH<sub>2</sub>NH<sub>2</sub>), 1.44 (s, 9H, -C(CH<sub>3</sub>)<sub>3</sub>).

#### Compound 3.35

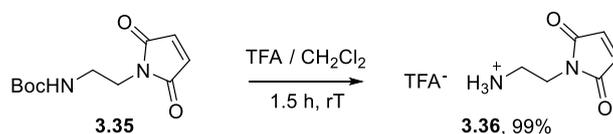


**3.34** (0.99 g, 6.2 mmol) and Et<sub>3</sub>N (1.3 mL, 9.3 mmol) were dissolved in Et<sub>2</sub>O (20 mL) at 0 °C. Maleic anhydride (0.61 g, 6.2 mmol) dissolved in Et<sub>2</sub>O (20 mL) was added dropwise over 0.5 h. Then the

reaction mixture was warmed to room temperature and stirred for 3 h. After concentrating the solution, the residue was dissolved in acetone (50 mL). Et<sub>3</sub>N (1.7 mL, 12.4 mmol) was added, and the mixture was heated to reflux. Acetic anhydride (1 mL, 9.3 mmol) was then added and the solution was heated to reflux for 20 h. After the solvent was concentrated under reduced pressure, the resulting crude material was purified by flash chromatography using 40% EtOAc in PET. Pure fractions were concentrated under reduced pressure to afford the product **3.35** (0.75 g, 3.1 mmol, 50%) as a white solid. Data were consistent with those previously reported.<sup>40</sup>

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>): δ 6.71 ppm (s, 2H, -CH=CH), 4.73 (br s, 1H, -NH), 3.66 (t, *J* = 6.0 Hz, 2H, -CH<sub>2</sub>Mal), 3.32 (td, *J* = 5.5 Hz, 2H, -CH<sub>2</sub>NH), 1.4 (s, 9H, -C(CH<sub>3</sub>)<sub>3</sub>). HRMS (ESI<sup>+</sup>): *m/z* calcd for C<sub>11</sub>H<sub>16</sub>N<sub>2</sub>NaO<sub>4</sub>: 263.1002 [M+Na]<sup>+</sup>; found: 263.1006.

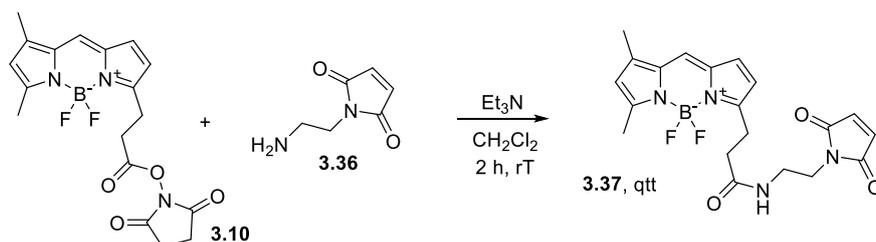
### Compound 3.36



To a stirred solution of **3.35** (123 mg, 0.51 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (4 mL) was added TFA (1 mL). The mixture was stirred for 1 h at room temperature. After evaporation of the solvent, the product was azeotroped with Et<sub>2</sub>O (3 x 5 mL). The product **3.36** was obtained as a white-pale yellow solid (73.8 mg, 0.51 mmol, 99%). Data were consistent with those previously reported.<sup>41</sup>

<sup>1</sup>H NMR (400 MHz, MeOD): δ 6.85 (s, 2H, -CH=CH), 3.78 (t, *J* = 5.7 Hz, 2H, -CH<sub>2</sub>Mal), 3.13 (t, *J* = 5.7 Hz, 2H, -CH<sub>2</sub>NH<sub>2</sub>). HRMS (ESI<sup>+</sup>): *m/z* calcd for C<sub>6</sub>H<sub>9</sub>N<sub>2</sub>O<sub>2</sub>: 141.0659 [M+H]<sup>+</sup>; found: 141.0658

### Compound 3.37



BODIPY **3.10** (11 mg, 28 μmol) was dissolved in CH<sub>2</sub>Cl<sub>2</sub> (1 mL). Then Et<sub>3</sub>N (30 μL) was added to the mixture. After few minutes, maleimide **3.36** (4 mg, 30 μmol) was added, and the mixture was stirred for 2 h. Then CH<sub>2</sub>Cl<sub>2</sub> (10 mL) was added and the organics washed with HCl (0.5 M, 3 x 10 mL). The organic phase was dried over anhydrous MgSO<sub>4</sub> and concentrated under reduced pressure to afford the crude product **3.37** (13 mg, 28 μmol, 100%) as a red solid used without further purification.

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>): δ 7.08 (s, 1H, -ArH<sub>5</sub>), 6.87 (d, *J* = 4.1 Hz, 1H, -ArH<sub>8</sub>), 6.66 (s, 2H, -CH=CH maleimide), 6.28 (d, *J* = 4.1 Hz, 1H, -ArH<sub>7</sub>), 6.12 (s, 1H, -ArH<sub>10</sub>), 5.98 (br s, 1H, -NH), 3.65 (t, *J* = 5.6 Hz, 2H, -NHCH<sub>2</sub>CH<sub>2</sub>), 3.44 (q, *J* = 5.6 Hz, 2H, -NHCH<sub>2</sub>CH<sub>2</sub>), 3.22 (t, *J* = 7.4 Hz, 2H, -COCH<sub>2</sub>CH<sub>2</sub>), 2.59 (t, *J* = 7.4 Hz, 2H, -COCH<sub>2</sub>), 2.55 (s, 3H, -Ar11CH<sub>3</sub>), 2.25 ppm (s, 3H, -Ar9CH<sub>3</sub>); **<sup>13</sup>C NMR** (400 MHz, CDCl<sub>3</sub>): δ 134.03 (CH=CH maleimide), 128.40 (C<sub>8</sub>), 123.76 (C<sub>5</sub>), 120.58 (C<sub>10</sub>), 117.15 (C<sub>7</sub>), 58.20 (NHCH<sub>2</sub>CH<sub>2</sub>), 37.41 (NHCH<sub>2</sub>CH<sub>2</sub>), 35.70 (C<sub>18</sub>), 24.69 (C<sub>14</sub>), 14.91 (C<sub>13</sub>), 11.24 (C<sub>12</sub>) ppm; **<sup>19</sup>F NMR** (400 MHz, CDCl<sub>3</sub>): δ 144.35 (d, *J* = 35 Hz), 144.5 ppm (d, *J* = 35 Hz); **HRMS** (ESI<sup>+</sup>): *m/z* calcd for C<sub>20</sub>H<sub>21</sub>BF<sub>2</sub>N<sub>4</sub>NaO<sub>3</sub> 437.1567 [M+Na]<sup>+</sup>; found: 437.1576.

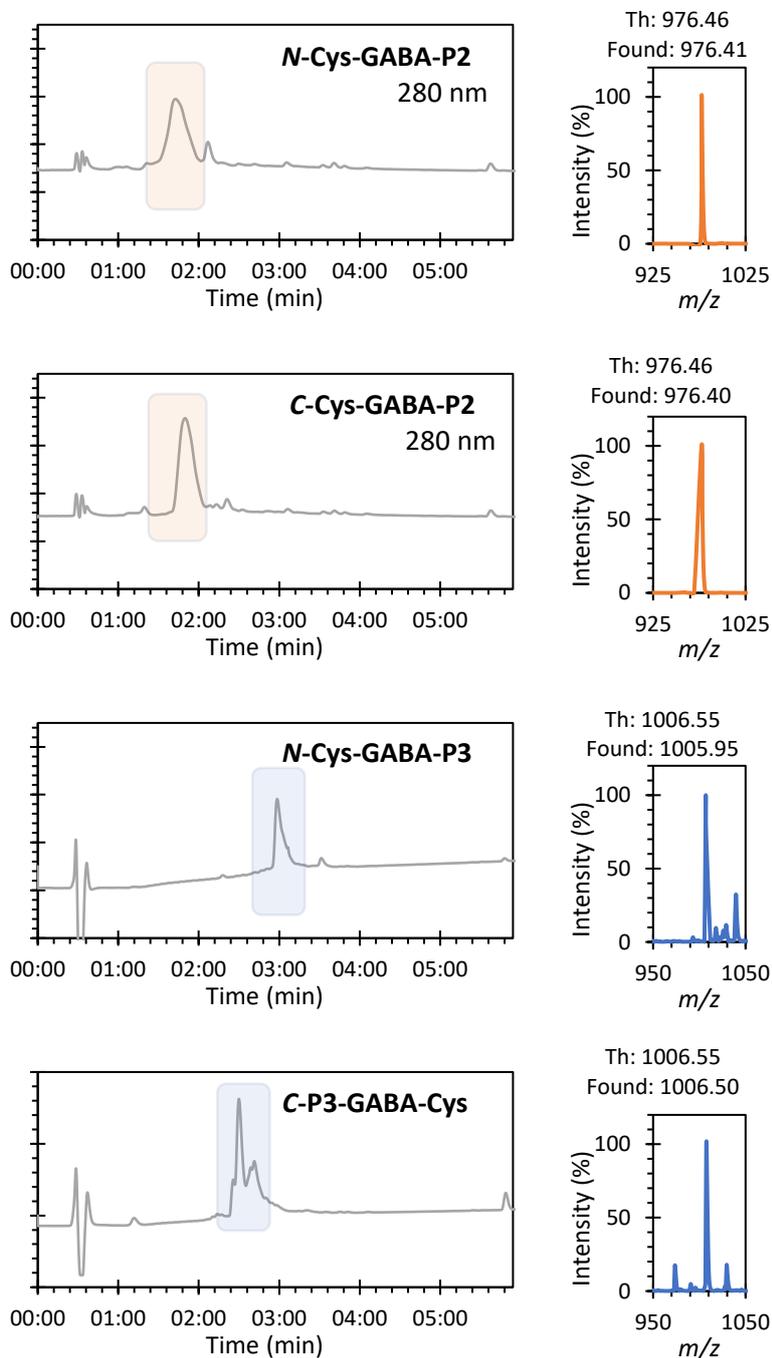
### 3.5.2 List of the unfunctionalized peptides synthesized by SPPS in this Chapter

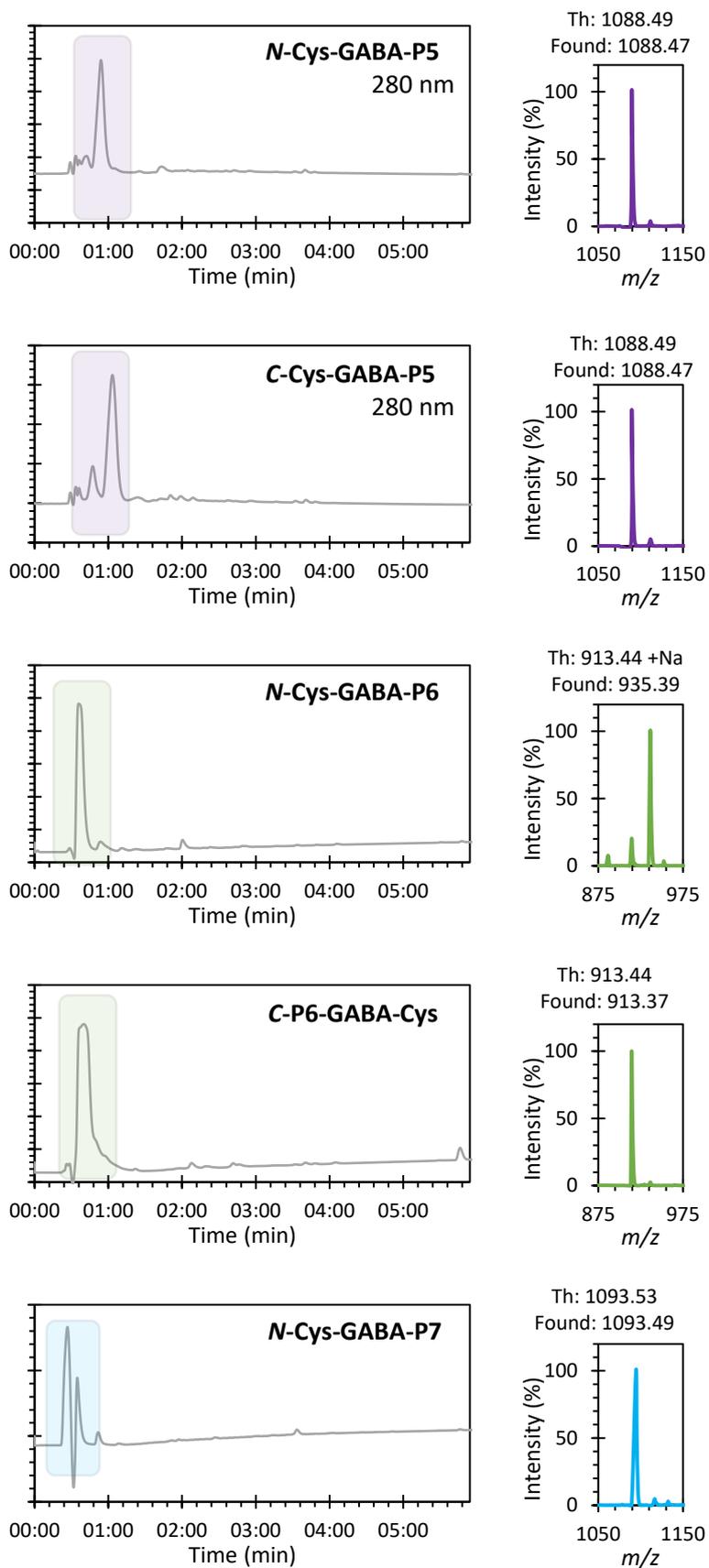
Name in text	Peptide sequence
<i>N</i> -Cys-GABA-P2	C-GABA-RGFFYT
<i>N</i> -GABA-P2	GABA-RGFFYT
<i>C</i> -P2-GABA-Cys	RGFFYT-GABA-C
<i>C</i> -Lys-P2	Fmoc-RGFFYT-K(DDE)
<i>N</i> -Cys-GABA-P3	C-GABA-LVEALYL
<i>C</i> -P3-GABA-Cys	LVEALYL-GABA-C
<i>N</i> -Cys-GABA-P5	C-GABA-HTTHMYL
<i>N</i> -GABA-P5	GABA-HTTHMYL
<i>C</i> -P5-GABA-Cys	HTTHMYL-GABA-C
<i>N</i> -Cys-GABA-P6	C-GABA-TLHSAQA
<i>N</i> -GABA-P6	GABA-TLHSAQA
<i>C</i> -P6-GABA-Cys	TLHSAQA-GABA-C
<i>N</i> -Cys-GABA-P7	C-GABA-HRNPRNN
<i>N</i> -GABA-P7	GABA-HRNPRNN
<i>C</i> -P7-GABA-Cys	HRNPRNN-GABA-C
<i>N</i> -Cys-P8	C-KETAAAKFERQHMSSTSA
<i>C</i> -Cys-P8	KETAAAKFERQHMSSTSA-C

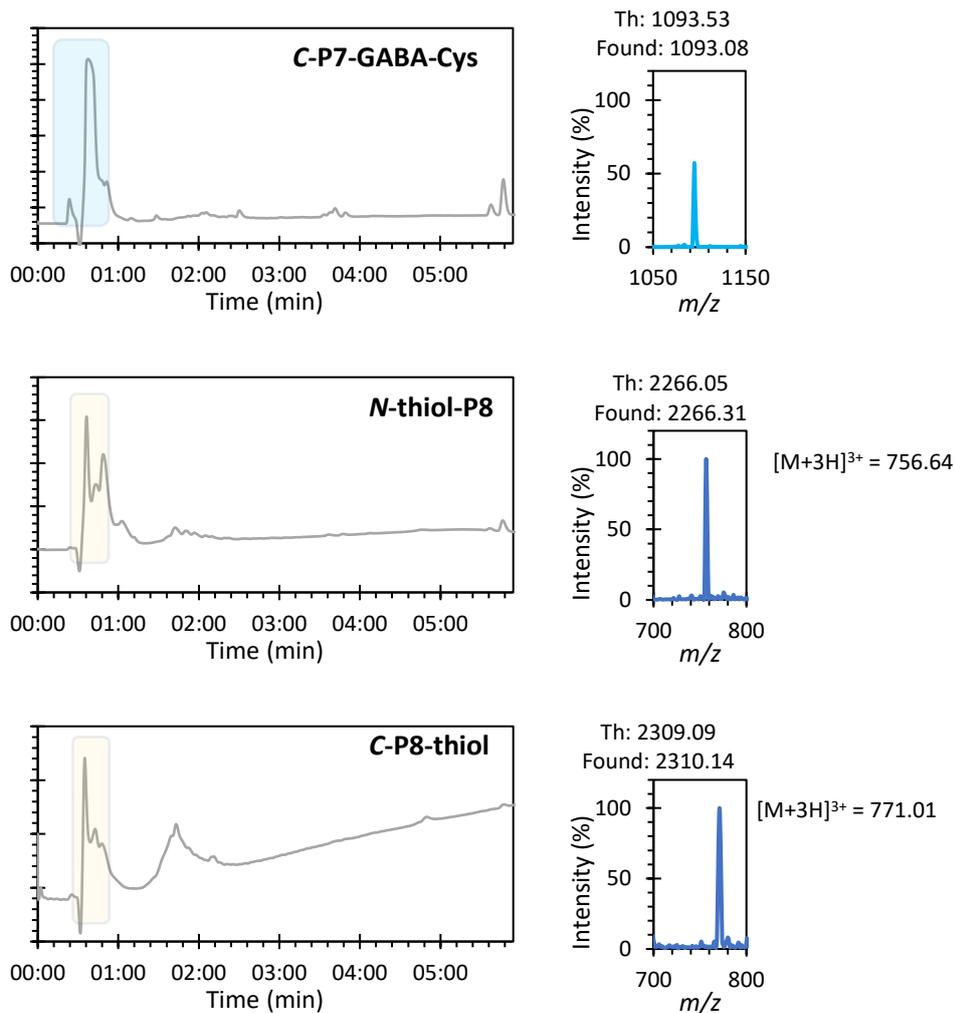
## 3.5.3 LCMS of peptides used in this Chapter

All the chromatograms are of absorbance at 220 nm except when noted otherwise.

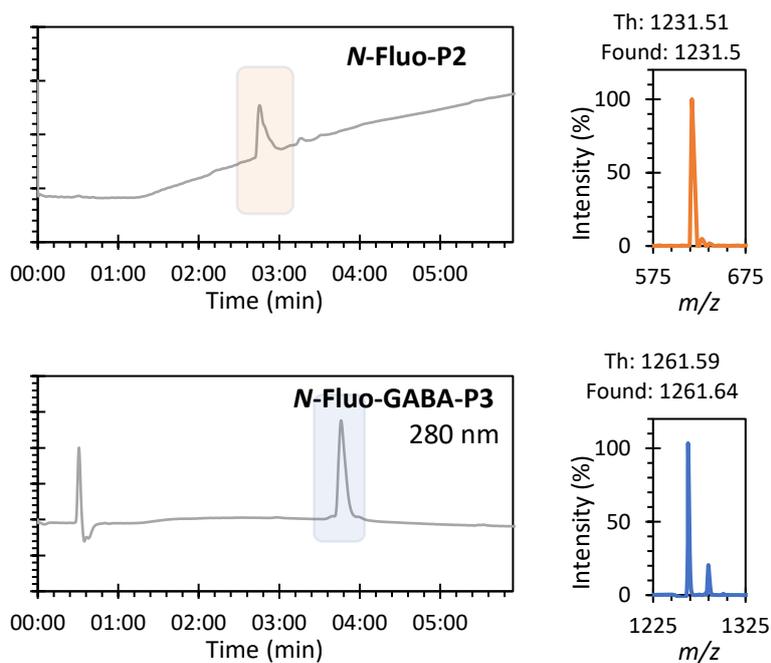
## 3.5.3.a QCM-D peptides

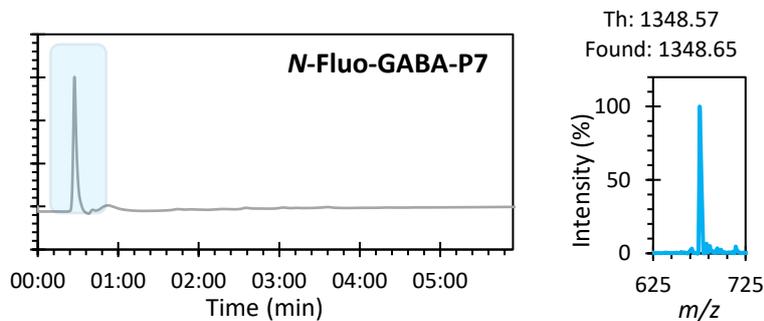
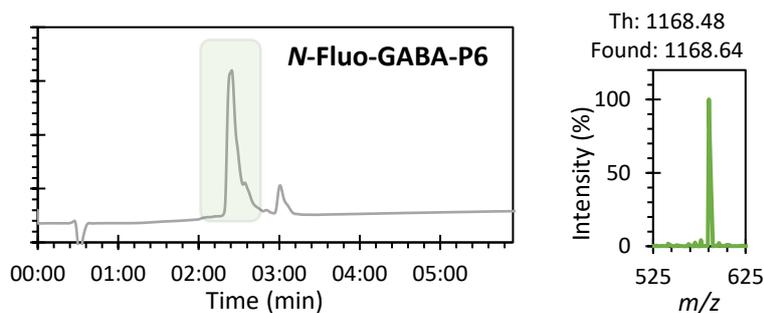
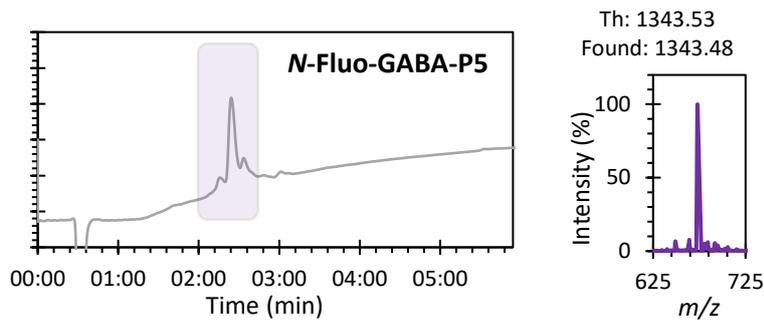




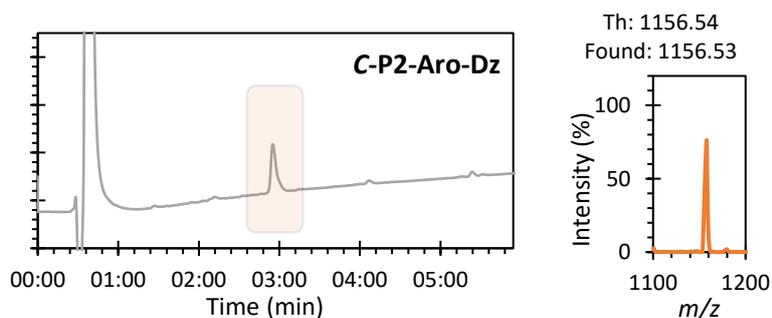
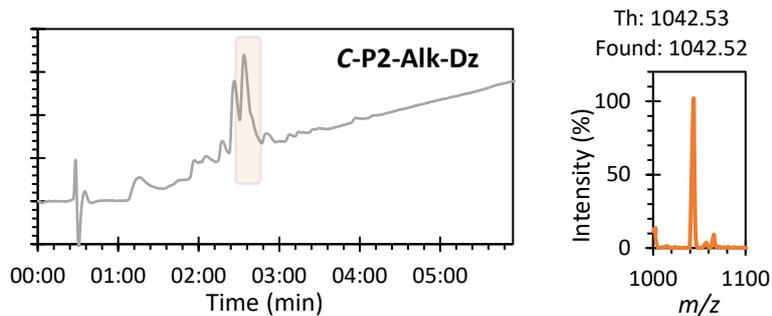


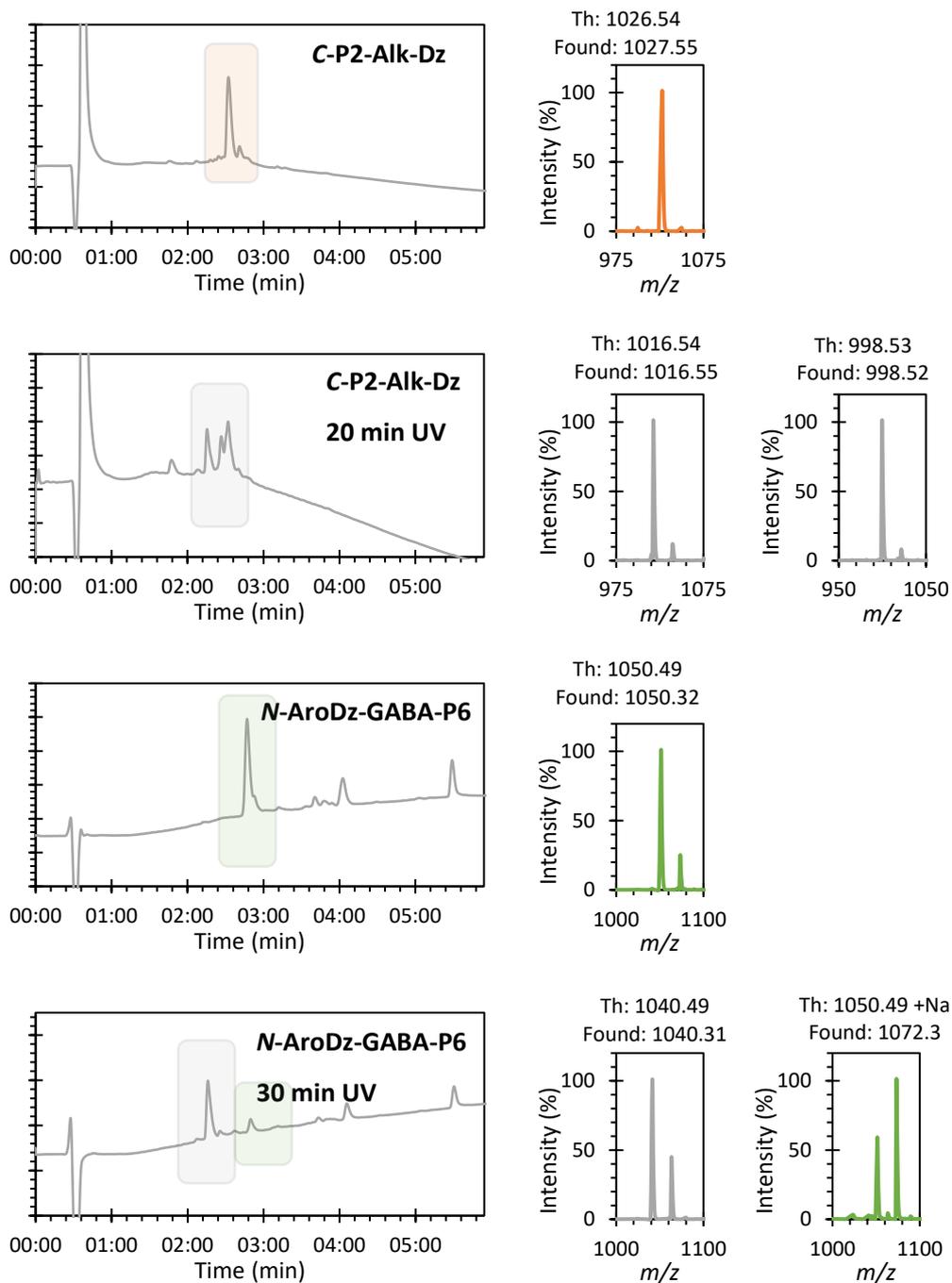
### 3.5.3.b 5,6-carboxyfluorescein 3.5 labelled peptides





### 3.5.3.c Diazirine labelled peptides

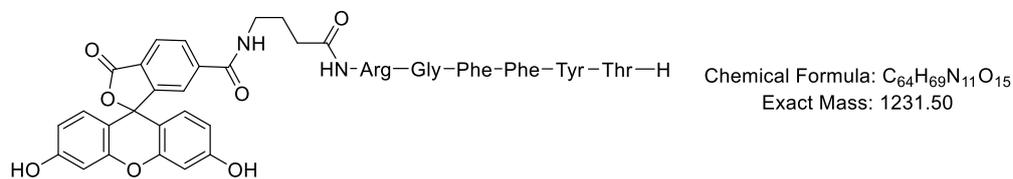




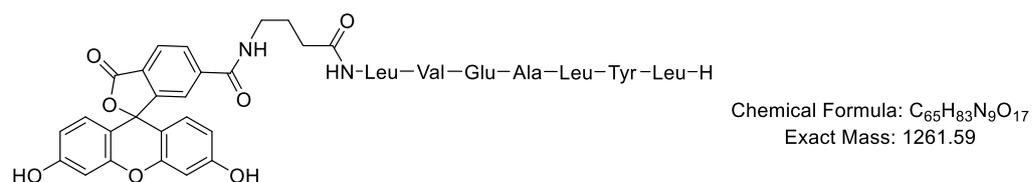
## 3.5.4 Structures of peptides used in this Chapter

**Fluorophore modified peptides**

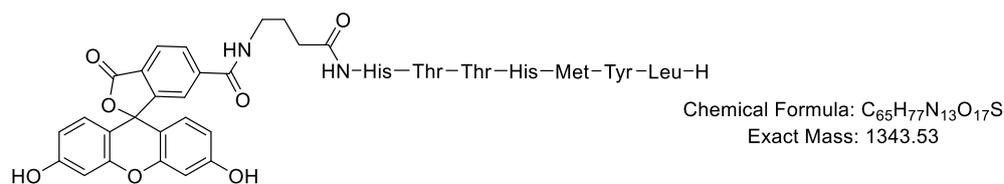
P2 : RGFFYT



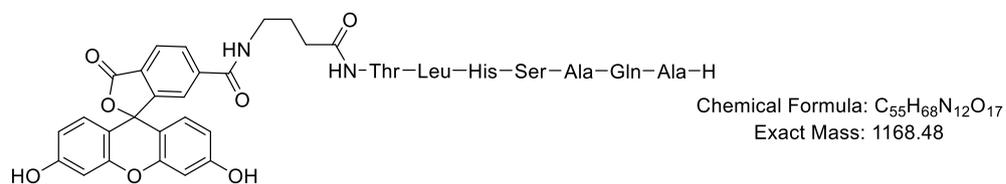
P3 : LVEALYL



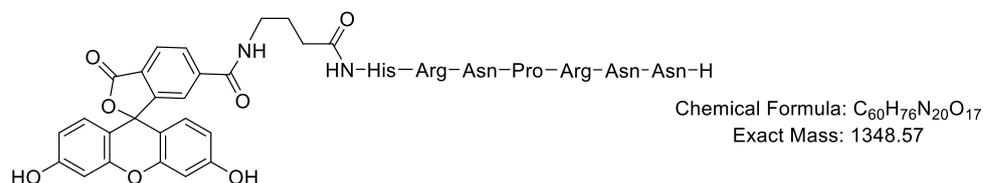
P5 : HTTHMYL



P6 : TLHSAQA

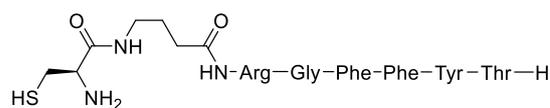


P7 : HRNPRNN

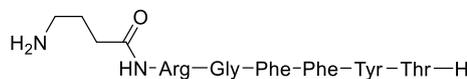


**QCM-D peptides**

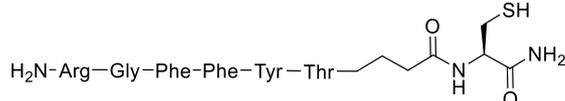
P2 : RGFFYT



Chemical Formula: C<sub>46</sub>H<sub>64</sub>N<sub>12</sub>O<sub>10</sub>S  
Exact Mass: 976.46

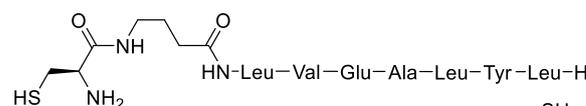


Chemical Formula: C<sub>43</sub>H<sub>59</sub>N<sub>11</sub>O<sub>9</sub>  
Exact Mass: 873.45

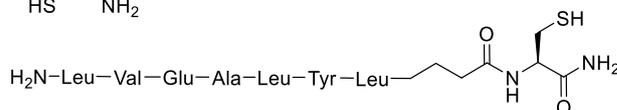


Chemical Formula: C<sub>46</sub>H<sub>64</sub>N<sub>12</sub>O<sub>10</sub>S  
Exact Mass: 976.46

P3 : LVEALYL

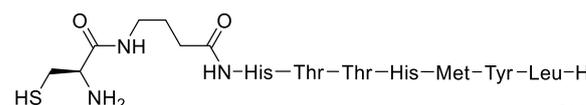


Chemical Formula: C<sub>47</sub>H<sub>78</sub>N<sub>10</sub>O<sub>12</sub>S  
Exact Mass: 1006.55

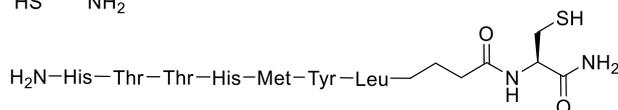


Chemical Formula: C<sub>47</sub>H<sub>78</sub>N<sub>10</sub>O<sub>12</sub>S  
Exact Mass: 1006.55

P5 : HTHMYL

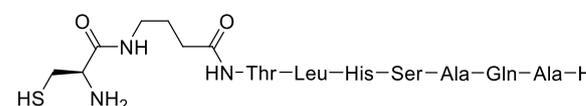


Chemical Formula: C<sub>47</sub>H<sub>72</sub>N<sub>14</sub>O<sub>12</sub>S<sub>2</sub>  
Exact Mass: 1088.49

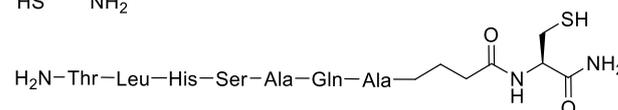


Chemical Formula: C<sub>47</sub>H<sub>72</sub>N<sub>14</sub>O<sub>12</sub>S<sub>2</sub>  
Exact Mass: 1088.49

P6 : TLHSAQA

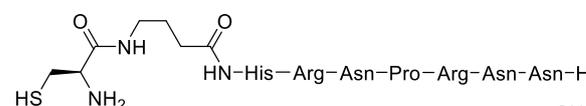


Chemical Formula: C<sub>37</sub>H<sub>63</sub>N<sub>13</sub>O<sub>12</sub>S  
Exact Mass: 913.44

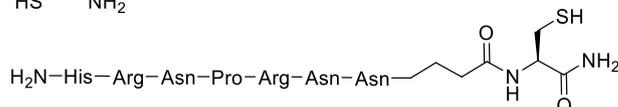


Chemical Formula: C<sub>37</sub>H<sub>63</sub>N<sub>13</sub>O<sub>12</sub>S  
Exact Mass: 913.44

P7 : HRNPRNN

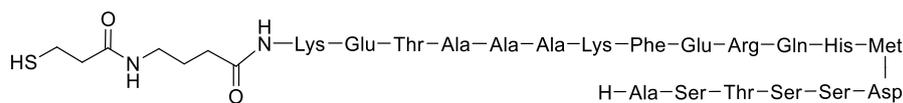


Chemical Formula: C<sub>42</sub>H<sub>71</sub>N<sub>21</sub>O<sub>12</sub>S  
Exact Mass: 1093.53

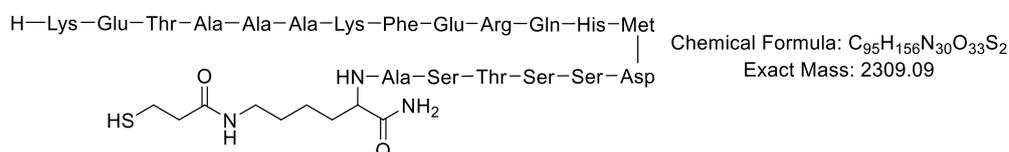


Chemical Formula: C<sub>42</sub>H<sub>71</sub>N<sub>21</sub>O<sub>12</sub>S  
Exact Mass: 1093.53

P8 : KETAAAKFERQHMSSTSA



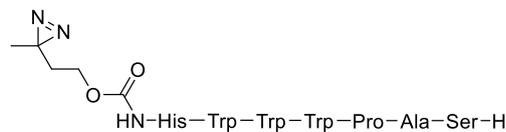
Chemical Formula: C<sub>93</sub>H<sub>151</sub>N<sub>29</sub>O<sub>33</sub>S<sub>2</sub>  
Exact Mass: 2266.05



Chemical Formula: C<sub>95</sub>H<sub>156</sub>N<sub>30</sub>O<sub>33</sub>S<sub>2</sub>  
Exact Mass: 2309.09

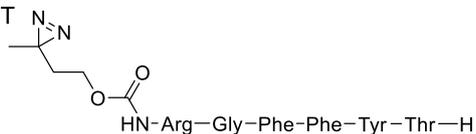
## Diazirine modified peptides

P1 : HWWWPAS

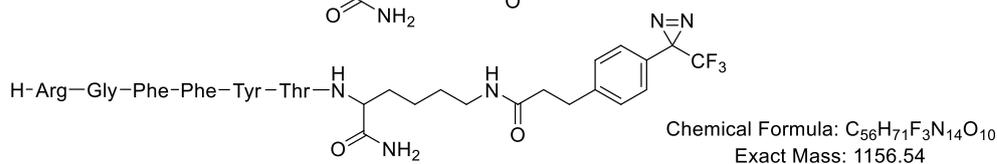
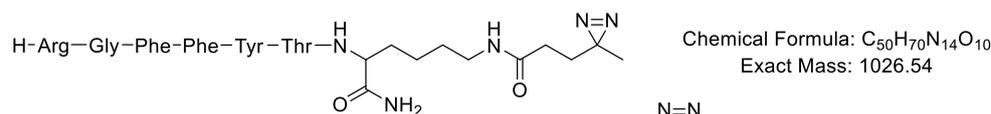
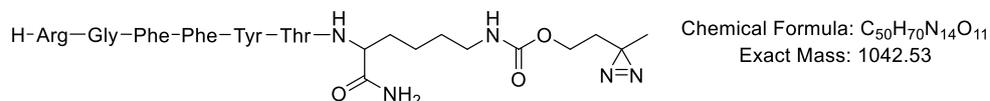


Chemical Formula: C<sub>55</sub>H<sub>63</sub>N<sub>15</sub>O<sub>10</sub>  
Exact Mass: 1093.49

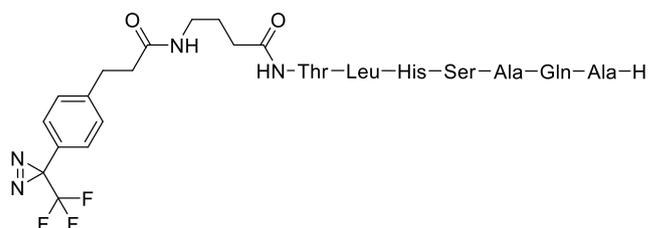
P2 : RGFFYT



Chemical Formula: C<sub>44</sub>H<sub>58</sub>N<sub>12</sub>O<sub>10</sub>  
Exact Mass: 914.44

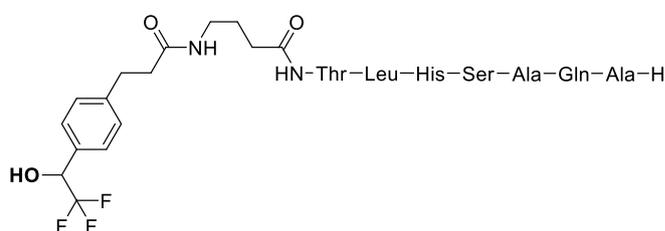


P6 : TLHSAQA



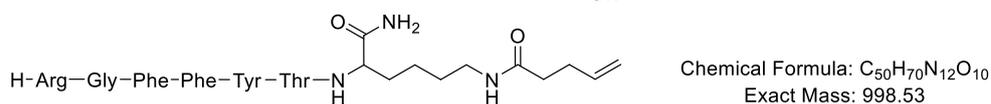
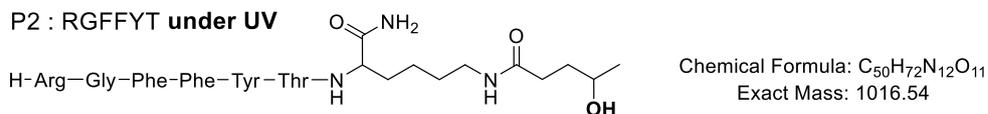
Chemical Formula: C<sub>45</sub>H<sub>65</sub>F<sub>3</sub>N<sub>14</sub>O<sub>12</sub>  
Exact Mass: 1050.49

P6 : TLHSAQA under UV



Chemical Formula: C<sub>45</sub>H<sub>67</sub>F<sub>3</sub>N<sub>12</sub>O<sub>13</sub>  
Exact Mass: 1040.49

P2 : RGFFYT under UV



### 3.5.5 QCM-D protocol

#### **Sensor preparation**

QCM-D sensors were cleaned by UV–ozone treatment (30 min), followed by sonication in a 2% Hellmanex III solution (5 min) and then sonication in ultrapure water (5 min). Cleaned sensors were functionalized according to the type of surface employed. Cleaned gold sensors were loaded into the instrument and functionalized with the cysteine labelled peptide in flow using solution at 10  $\mu\text{M}$  in PBS (1 mL). For maleimide surfaces: Cleaned silicon dioxide samples were functionalized using a three-step procedure to produce a rigid layer of maleimide on the surface: (i) immersion in freshly prepared 1% v/v APTES/anhydrous EtOH solution for 1 hour; (ii) maleimide-PEG6-NHS (50  $\mu\text{M}$ , DMSO) overnight; and finally (iii) the cysteine-modified peptide (10  $\mu\text{M}$ , PBS) overnight. Sensors were then rinsed with PBS and dried with  $\text{N}_2$  gas before use. For NHS-ester surfaces: Cleaned silicon dioxide samples were functionalized using a three-step procedure to produce a rigid layer of NHS-ester on the surface: (i) immersion in 1% v/v MPTES/anhydrous EtOH solution for 1 hour; (ii) maleimide-PEG6-NHS (50  $\mu\text{M}$ , DMSO) overnight; and finally (iii) the free amine peptide (10  $\mu\text{M}$ , PBS) overnight. Sensors were then rinsed with PBS and dried with  $\text{N}_2$  gas before use.

#### **QCM-D surface studies**

Following formation of the functionalized surfaces, each sensor was installed into the flow modules of a Q-Sense E4 system by Biolin Scientific. Each chamber was then filled with Milli-Q water at a flow rate of 100  $\mu\text{L}/\text{min}$  controlled by a four-channel peristaltic pump. To achieve a stable baseline, PBS was left to flow through the modules at 100  $\mu\text{L}/\text{min}$  until the drift in frequency was  $<\pm 1$  Hz over 10 min. For all experiments, the temperature of the modules was kept at 20  $^\circ\text{C}$  (standard deviation  $5 \times 10^{-3}$   $^\circ\text{C}$ ) and the flow rate was kept constant at 100  $\mu\text{L}/\text{min}$ . 1 mL of each concentration of protein was inserted at the same flow rate after stabilisation. The range of concentration for insulin and RNase-A was 0.05, 0.1, 0.5, 1, 5, 10, 50, 100  $\mu\text{M}$ , while for FGF-2 the range was 0.005, 0.01, 0.05, 0.1, 0.5, 1, 5, 10  $\mu\text{M}$ . QCM-D values showing both the frequency and dissipation shifts for each experiment are provided in Annex of this thesis.

#### **QCM-D data processing**

The data was processed in the following way. 4 data points for each concentration were averaged to calculate the delta frequency versus a control using PBS. The solver function in excel was then used to generate the best sigmoidal fit using the following equation:

$$y_{calc} = \min + \frac{\max - \min}{1 + 10^{n \cdot (\log(C) - \log(ec50))}}$$

C = concentration. Max, min, n and ec50 are generated by the solver function to ensure that the difference between  $y_{calc}$  and C is the minimal value possible. The following table represents the delta

frequency for each concentration of protein C. In the second part of the table are the values obtained for the sigmoidal fit.

Table 2: Data for Figure 8, Figure 10, Figure 12 and Figure 16

$\Delta F$ (Hz) at C [nM]	Au sensor					
	Cys-P2	P2-Cys	Cys-P3	P3-Cys	Thiol-P8	P8-thiol
0.05	0.1083131	-0.071818	0.4009851	0.469273	0.106478294	-0.04970982
0.1	1.0569878	0.5383773	0.8369196	0.9578793	0.174228427	0.149092797
0.5	3.218969	1.9779933	1.3730259	1.5452228	0.259757537	0.242358417
1	5.4302758	4.1338487	2.3717243	2.694455	0.305748345	0.461853039
5	6.3504695	6.4493745	4.4348969	4.6929392	0.371885484	1.055892924
10	7.0054375	8.5453041	6.699322	6.5375715	0.419073518	1.996655926
50	7.3842105	11.02579	8.7410521	8.0942722	0.552051192	2.999318923
100	8.0337198	13.037933	10.627377	9.6022866		
Sigmoidal fit						
min	7.3134225	14.289058	10.211088	9.2799395	No fit	3.020954866
max	-0.609252	-1.39738	0.4124178	0.3657438		0.236245108
n	1.0033988	0.5391779	0.8534846	0.7959578		1.882432457
ec50	0.4333366	4.1160295	6.1331272	4.4037843		7.630565843
SSR	0.6076389	0.7224246	0.4644592	0.3562835		0.041281773

$\Delta F$ (Hz) at C [nM]	SiO <sub>2</sub> sensor	
	GABA-P2	Cys-P2
0.05	0.0527815	0.091758945
0.1	0.3026862	0.131608663
0.5	0.2999337	0.513690197
1	0.55319	1.001476692
5	0.5824119	1.725266613
10	0.7452717	2.83076185
50	1.345384	5.518616779
100	1.7721015	7.588129632
Sigmoidal fit		
min	211.61589	1364.8012
max	0.1000305	-0.344349
n	0.3799448	0.4161654
ec50	34787604	23486106
SSR	0.0594314	0.1073917

### 3.5.5 FP supplementary information

All FP data was recorded on a BMG Labtech Clariostar Plate Reader instrument from the technology facility of the Department of Biology. Two protocols were used one for fluorescence intensity and one for fluorescence polarization. A pre-set program adapted for fluorescein fluorophore was selected

from the options of the instruments and the different parameters are summarized in the following table for both protocols.

	Fluorescence (FI)	Fluorescence polarization (FP)
<b>Basic settings</b>		
Microplate name:	PE OptiPlate 96	PE OptiPlate 96
<b>Endpoint settings</b>		
No. of flashes per well:	20	200
<b>Optic settings</b>		
Presetname:	Fluorescein (FITC)	FP-EndPoint
Excitation:	483-14	F: 482-16
Dichroic filter:	auto 502.5	F: LP 504
Emission:	530-30	F: 530-40
Gain A		Depend on plate
Gain B	Depend on plate	Depend on plate
Well(s) used for gain adjustment:	Depend on plate	D
Focal height [mm]:	6.4	6
Target mp value [mP]:		35
<b>General settings</b>	Top optic used	Top optic used

The raw FP data was processed by first subtraction of the background for both parallel and perpendicular emissions and then applying the following formula to obtain anisotropy, with  $I_{\parallel}$  as the parallel intensity and  $I_{\perp}$ . Standard deviation was calculated from the three replicates.

$$r = \frac{(I_{\parallel} - I_{\perp})}{I_{\parallel} + 2I_{\perp}}$$

### 3.5.6 UV curing supplementary information

All UV irradiations for diazirine studies were done using the nail machine Rio UV 36W (product code RIO03373, using 1-UVLP6-BULB-SET bulb, wavelength not found in product description). As shown in the picture below, to raise the sample and place it closer to the bulb, a plastic TipOne box (height 4.5 cm) was used as support to hold either open 2mL Eppendorf lids or LCMS vials in a horizontal position to be irradiated for different times. Then a box was positioned in front of the opening to ensure a close environment and to contain the irradiation of UV.



### 3.6 Bibliography

- (1) Murale, D. P.; Hong, S. C.; Haque, M. M.; Lee, J. S. Photo-Affinity Labeling (PAL) in Chemical Proteomics: A Handy Tool to Investigate Protein-Protein Interactions (PPIs). *Proteome Sci.* **2017**, *15* (1), 1–34. <https://doi.org/10.1186/s12953-017-0123-3>.
- (2) Kenny, C. H.; Ding, W.; Kelleher, K.; Benard, S.; Dushin, E. G.; Sutherland, A. G.; Mosyak, L.; Kriz, R.; Ellestad, G. Development of a Fluorescence Polarization Assay to Screen for Inhibitors of the FtsZ/ZipA Interaction. *Anal. Biochem.* **2003**, *323* (2), 224–233. <https://doi.org/10.1016/j.ab.2003.08.033>.
- (3) Moerke, N. J. Fluorescence Polarization (FP) Assays for Monitoring Peptide-Protein or Nucleic Acid-Protein Binding. *Curr. Protoc. Chem. Biol.* **2009**, *1* (1), 1–15. <https://doi.org/10.1002/9780470559277.ch090102>.
- (4) What is QCM-D? <https://www.3t-analytik.de/technologies/qcm-d/what-qcm-d>.
- (5) Quartz crystal microbalance <https://www.nanoscience.com/techniques/quartz-crystal-microbalance/>.
- (6) Spicer, C. D.; Davis, B. G. Selective Chemical Protein Modification. *Nat. Commun.* **2014**, *5*. <https://doi.org/10.1038/ncomms5740>.
- (7) Boutureira, O.; Bernardes, J. L. Advances in Chemical Protein Modification. *Chem. Rev.* **2015**, *115*, 2174–2195. <https://doi.org/10.1021/cr500399p>.
- (8) Akamoto, S. S.; Amachi, I. H. Recent Progress in Chemical Modification of Proteins. *Anal. Sci.* **2019**, *35* (January), 5–27. <https://doi.org/10.2116/analsci.18R003>.
- (9) Chiang, H. L.; Ngo, S. T.; Chen, C. J.; Hu, C. K.; Li, M. S. Oligomerization of Peptides LVEALYL and RGFFYT and Their Binding Affinity to Insulin. *PLoS One* **2013**, *8* (6). <https://doi.org/10.1371/journal.pone.0065358>.
- (10) Jankovic, B.; Ruf, J.; Zanobini, C.; Bozovic, O.; Buhrke, D.; Hamm, P. Sequence of Events during Peptide Unbinding from RNase S: A Complete Experimental Description. *J. Phys. Chem. Lett.* **2021**, *12* (21), 5201–5207. <https://doi.org/10.1021/acs.jpcllett.1c01155>.
- (11) Dwyer, J. J.; Dwyer, M. A.; Kosiakoff, A. High Affinity RNase S-Peptide Variants Obtained by Phage Display Have a Novel “Hot-Spot” of Binding Energy. *Biochem. Pharmacol.* **2001**, *40*, 13941–13500.
- (12) Wu, X.; Yan, Q.; Huang, Y.; Huang, H.; Su, Z.; Xiao, J.; Zeng, Y.; Wang, Y.; Nie, C.; Yang, Y.; Li, X. Isolation of a Novel Basic FGF-Binding Peptide with Potent Antiangiogenetic Activity. *J. Cell. Mol. Med.* **2010**, *14* (1–2), 351–356. <https://doi.org/10.1111/j.1582-4934.2008.00506.x>.
- (13) Lea, W. A.; Simeonov, A. Fluorescence Polarization Assay in Small Molecule Screening. *Expert Opin Drug Discov.* **2011**, *6* (1), 17–32. <https://doi.org/10.1517/17460441.2011.537322.Fluorescence>.
- (14) Roehrl, M. H. A.; Wang, J. Y.; Wagner, G. A General Framework for Development and Data Analysis of Competitive High-Throughput Screens for Small-Molecule Inhibitors of Protein-Protein Interactions by Fluorescence Polarization. *Biochemistry* **2004**, *43* (51), 16056–16066. <https://doi.org/10.1021/bi048233g>.
- (15) Tan, H. Y.; Eskandari, R.; Shen, D.; Zhu, Y.; Liu, T. W.; Willems, L. I.; Alteen, M. G.; Madden, Z.; Voadlo, D. J. Direct One-Step Fluorescent Labeling of O-GlcNAc-Modified Proteins in Live Cells Using Metabolic Intermediates. *J. Am. Chem. Soc.* **2018**, *140* (45), 15300–15308. <https://doi.org/10.1021/jacs.8b08260>.
- (16) Wang, D.; Fan, J.; Gao, X.; Wang, B.; Sun, S.; Peng, X. Carboxyl BODIPY Dyes from Bicarboxylic

- Anhydrides: One-Pot Preparation, Spectral Properties, Photostability, and Biolabeling. *J. Org. Chem.* **2009**, *74* (20), 7675–7683. <https://doi.org/10.1021/jo901149y>.
- (17) Pakhomov, A. A.; Kononevich, Y. N.; Stukalova, M. V.; Svidchenko, E. A.; Surin, N. M.; Cherkaev, G. V.; Shchegolikhina, O. I.; Martynov, V. I.; Muzafarov, A. M. Synthesis and Photophysical Properties of a New BODIPY-Based Siloxane Dye. *Tetrahedron Lett.* **2016**, *57* (9), 979–982. <https://doi.org/10.1016/j.tetlet.2016.01.059>.
- (18) Reindl, W.; Yuan, J.; Krämer, A.; Strebhardt, K.; Berg, T. Inhibition of Polo-like Kinase 1 by Blocking Polo-Box Domain-Dependent Protein-Protein Interactions. *Chem. Biol.* **2008**, *15* (5), 459–466. <https://doi.org/10.1016/j.chembiol.2008.03.013>.
- (19) Xu, M.; Liu, C.; Zhou, M.; Li, Q.; Wang, R.; Kang, J. Screening of Small-Molecule Inhibitors of Protein-Protein Interaction with Capillary Electrophoresis Frontal Analysis. *Anal. Chem.* **2016**, *88* (16), 8050–8057. <https://doi.org/10.1021/acs.analchem.6b01430>.
- (20) Krajcovicova, S.; Stankova, J.; Dzubak, P.; Hajduch, M.; Soural, M.; Urban, M. A Synthetic Approach for the Rapid Preparation of BODIPY Conjugates and Their Use in Imaging of Cellular Drug Uptake and Distribution. *Chem. - A Eur. J.* **2018**, *24* (19), 4957–4966. <https://doi.org/10.1002/chem.201706093>.
- (21) Mendive-Tapia, L.; Miret-casals, L.; Barth, N.; Wang, J.; Bray, A. De; Beltramo, M.; Robert, V.; Ampe, C.; Hodson, D.; Madder, A.; Vendrell, M. Acid-Resistant BODIPY Amino Acids for Peptide-Based Fluorescence Imaging of GPR54 Receptors in Pancreatic Islets. *Angew. Chem., Int. Ed. Engl.* **2023**. <https://doi.org/10.1002/anie.202302688>.
- (22) Das, J. Aliphatic Diazirines as Photoaffinity Probes for Proteins: Recent Developments. *Chem. Rev.* **2011**, *111* (8), 4405–4417. <https://doi.org/10.1021/cr1002722>.
- (23) Raynal, L.; Rose, N. C.; Donald, J. R.; Spicer, C. D. Photochemical Methods for Peptide Macrocyclisation. *Chem. - A Eur. J.* **2021**, *27* (1), 69–88. <https://doi.org/10.1002/chem.202003779>.
- (24) West, A. V.; Muncipinto, G.; Wu, H. Y.; Huang, A. C.; Labenski, M. T.; Jones, L. H.; Woo, C. M. Labeling Preferences of Diazirines with Protein Biomolecules. *J. Am. Chem. Soc.* **2021**, *143* (17), 6691–6700. <https://doi.org/10.1021/jacs.1c02509>.
- (25) Horne, J. E.; Walko, M.; Calabrese, A. N.; Levenstein, M. A.; Brockwell, D. J.; Kapur, N.; Wilson, A. J.; Radford, S. E. Rapid Mapping of Protein Interactions Using Tag-Transfer Photocrosslinkers. *Angew. Chem., Int. Ed. Engl.* **2018**, No. 57, 16688–16692. <https://doi.org/10.1002/anie.201809149>.
- (26) Tian, Y.; Jacinto, M. P.; Zeng, Y.; Yu, Z.; Qu, J.; Liu, W. R.; Lin, Q. Genetically Encoded 2-Aryl-5-Carboxytetrazoles for Site-Selective Protein Photo-Cross-Linking. *J. Am. Chem. Soc.* **2017**, *139* (17), 6078–6081. <https://doi.org/10.1021/jacs.7b02615>.
- (27) Geurink, P. P.; Klein, T.; Prèly, L.; Paal, K.; Leeuwenburgh, M. A.; Van der Marel, G. A.; Kauffman, H. F.; Overkleeft, H. S.; Bischoff, R. Design of Peptide Hydroxamate-Based Photoreactive Activity-Based Probes of Zinc-Dependent Metalloproteases. *European J. Org. Chem.* **2010**, 2100–2112. <https://doi.org/10.1002/ejoc.200901385>.
- (28) Kambe, T.; Correia, B. E.; Niphakis, M. J.; Cravatt, B. F. Mapping the Protein Interaction Landscape for Fully Functionalized Small-Molecule Probes in Human Cells. *J. Am. Chem. Soc.* **2014**, *136* (30), 10777–10782. <https://doi.org/10.1021/ja505517t>.
- (29) Wang, L.; Tachrim, Z. P.; Kurokawa, N.; Ohashi, F.; Sakihama, Y.; Hashidoko, Y.; Hashimoto, M. Base-Mediated One-Pot Synthesis of Aliphatic Diazirines for Photoaffinity Labeling. *Molecules* **2017**, *22* (8), 2–9. <https://doi.org/10.3390/molecules22081389>.

- (30) Spillier, Q.; Ravez, S.; Dochain, S.; Vertommen, D.; Thabault, L.; Feron, O.; Frederick, R. Unravelling the Allosteric Targeting of PHGDH at the ACT-Binding Domain with a Photoactivatable Diazirine Probe and Mass Spectrometry Experiments. *Molecules* **2021**, *26* (2), 477. <https://doi.org/10.3390/molecules26020477>.
- (31) Kohmoto, S.; Mori, E.; Kishikawa, K. Room-Temperature Discotic Nematic Liquid Crystals over a Wide Temperature Range: Alkali-Metal-Ion-Induced Phase Transition from Discotic Nematic to Columnar Phases. *J. Am. Chem. Soc.* **2007**, *129* (44), 13364–13365. <https://doi.org/10.1021/ja073962f>.
- (32) Snow, A. W.; Foos, E. E. Conversion of Alcohols to Thiols via Tosylate Intermediates. *Synthesis (Stuttg)*. **2003**, No. 4, 509–512. <https://doi.org/10.1055/s-2003-37650>.
- (33) Wu, L.; Burgess, K. A New Synthesis of Symmetric Boraindacene (BODIPY) Dyes. *Chem. Commun.* **2008**, No. 40, 4933–4935. <https://doi.org/10.1039/b810503k>.
- (34) Yao, H.; Gunawan, Y. F.; Liu, G.; Tse, M. K.; Zhu, G. Optimization of Axial Ligands to Promote the Photoactivation of BODIPY-Conjugated Platinum(IV) Anticancer Prodrugs. *Dalt. Trans.* **2021**, *50* (39), 13737–13747. <https://doi.org/10.1039/d1dt02362d>.
- (35) Wenskowsky, L.; Schreuder, H.; Derdau, V.; Matter, H.; Volkmar, J.; Nazaré, M.; Opatz, T.; Petry, S. Identification and Characterization of a Single High-Affinity Fatty Acid Binding Site in Human Serum Albumin. *Angew. Chem., Int. Ed. Engl.* **2018**, *57* (4), 1044–1048. <https://doi.org/10.1002/anie.201710437>.
- (36) Novotný, J.; Pospěchová, K.; Hrabálek, A.; Čáp, R.; Vávrová, K. Synthesis of Fluorescent C24-Ceramide: Evidence for Acyl Chain Length Dependent Differences in Penetration of Exogenous NBD-Ceramides into Human Skin. *Bioorganic Med. Chem. Lett.* **2009**, *19* (24), 6975–6977. <https://doi.org/10.1016/j.bmcl.2009.10.047>.
- (37) Lee, M.; Grissom, C. B. Design, Synthesis, and Characterization of Fluorescent Cobalamin Analogues with High Quantum Efficiencies. *Org. Lett.* **2009**, *11* (12), 2499–2502. <https://doi.org/10.1021/ol900401z>.
- (38) Ueno, Y.; Jiao, G.; Burgess, K. Preparation of 5- and 6-Carboxyfluorescein. *Synthesis (Stuttg)*. **2004**, *15*, 2591–2593. <https://doi.org/10.1055/s-2004-829194>.
- (39) Hashimoto, M.; Kato, Y. H.; Hatanaka, Y. Selective Hydrogenation of Alkene in (3-Trifluoromethyl) Phenyl diazirine Photophor with Wilkinson's Catalyst for Photoaffinity Labeling. *Chem. Pharm. Bull.* **2007**, *55* (10), 1540–1543. <https://doi.org/10.1248/cpb.55.1540>.
- (40) Richter, M.; Chakrabarti, A.; Ruttekolk, I. R.; Wiesner, B.; Beyermann, M.; Brock, R.; Rademann, J. Multivalent Design of Apoptosis-Inducing Bid-BH3 Peptide-Oligosaccharides Boosts the Intracellular Activity at Identical Overall Peptide Concentrations. *Chem. - A Eur. J.* **2012**, *18* (52), 16708–16715. <https://doi.org/10.1002/chem.201202276>.
- (41) Hayashi, T.; Yasueda, Y.; Tamura, T.; Takaoka, Y.; Hamachi, I. Analysis of Cell-Surface Receptor Dynamics through Covalent Labeling by Catalyst-Tethered Antibody. *J. Am. Chem. Soc.* **2015**, *137* (16), 5372–5380. <https://doi.org/10.1021/jacs.5b02867>.

## Chapter 4 – Synthesis of ligand-directed peptide probes

As mentioned throughout the introduction of this thesis, ligand-directed chemistries have been increasingly used to modify proteins in a traceless manner. This approach relies on the specific interaction between a protein and ligand, which due to the proximity effect creates a pseudo-intramolecular micro-environment that facilitates the reaction between the probe attached to the ligand and the protein. Two classes of ligand-directed chemistries were introduced: the one-pot ligand directed approach (*OpLD*) and the catalyst-based approach (*CatLD*). The reader is referred to the introduction section for an in-depth description of these two approaches.<sup>1</sup> It was chosen that in this thesis, two types of reagents for *OpLD* would be studied: ligand-directed *N*-acyl imidazole chemistry (LDAI) and ligand-directed methylcyano-*N*-acyl-*N*-sulfonamide chemistry (LDNASA). These species were picked as LDAI requires only a few synthetic steps to access the required probes and we predicted it would be easy to apply to protein substrate due to the similarity between the imidazole core and a histidine residue. The duo needed for *CatLD* chemistry was chosen to be pyridinium oxime catalysts able to catalyse *p*-nitrobenzyl *N*-acyl-*N*-sulfonamide substrates. The two NASA probes were picked to allow comparison between the two approaches with similar structures in term of ease of use, kinetic, and targeted residues. As a reminder, the main core for the NASA and imidazole probes are represented in Figure 1.

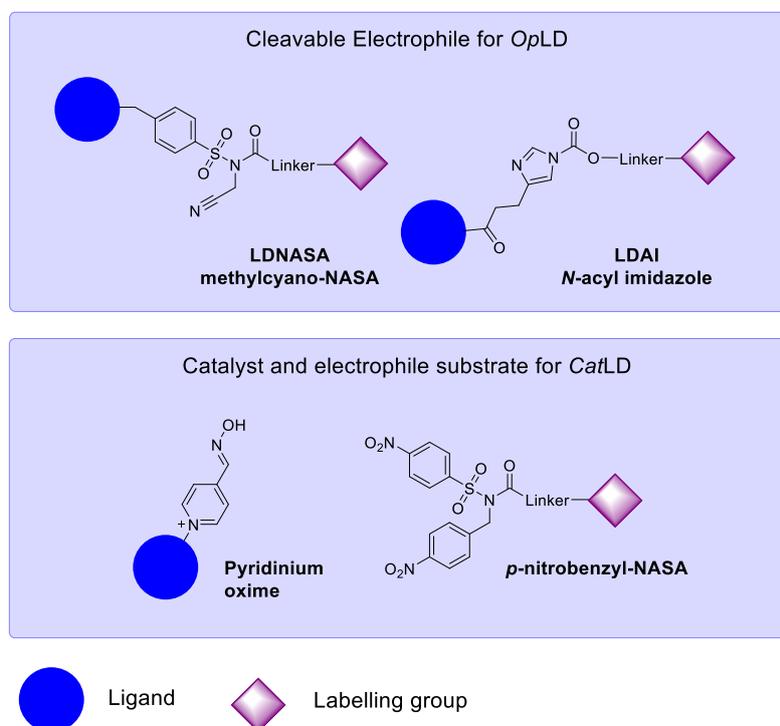


Figure 1: Structure of the *OpLD* probes and *CatLD* probes investigated in this chapter.

While these probes have to date been used on a number of small molecule ligands for protein modification, in this PhD, we will study the use of peptide ligands. Using phage display, it has become

recently much easier to find binding peptides to any protein of choice reducing the time and cost to find an appropriate ligand. Moreover, as mentioned in the introduction, the use of peptides is also beneficial as it means that native proteins can be used directly without the need of post-translation modifications. Thus, while we focused on the modification of peptides for the model proteins insulin and FGF-2 in this chapter, we believe that this new technique could be translational to many protein substrates later on.

In this chapter, we will detail the synthesis of the core elements for each LDAI, and LDNASA chemistries and the insertion of those reagents onto peptides.

#### 4.1 Synthesis of LDAI probes for *Op*LD

*N*-Acyl imidazoles probes (LDAI) were the first type of ligand-directed chemistry probe to be synthesised in this PhD. As a small reminder from the introduction, *N*-acyl imidazole units are synthesised through an alkyloxyacyl bond from the reaction between an imidazole core and an activated carbamate, as presented in Figure 2.<sup>2-4</sup> They have been shown to be reactive with a number of residues on a protein surface such as Lys, Ser and Tyr.<sup>2</sup> The second order rate constants for these type of probes have already been investigating showing a range between  $\sim 10$ - $100 \text{ M}^{-1}.\text{s}^{-1}$  which is comparable to those of copper-catalyzed azide-alkyne cycloaddition (CuAAC), a common approach used for protein modification.<sup>2</sup> The reader is referred to the Introduction for further detail on this chemistry and its known applications. The following paragraph will describe the synthesis of LDAI probes and their insertion in to peptide ligands.

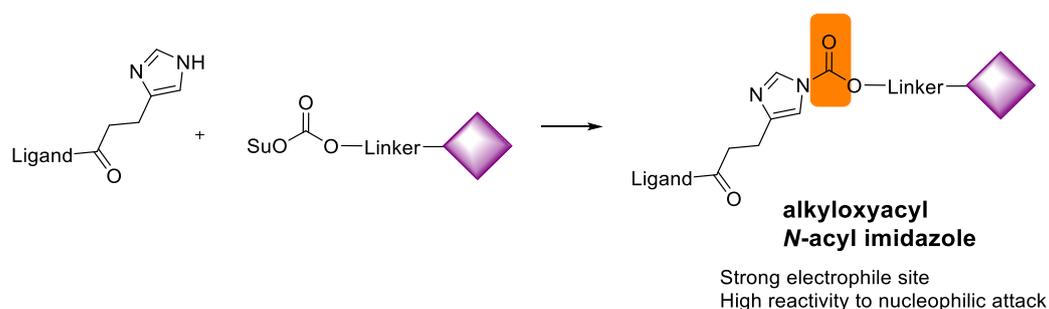


Figure 2: Difference between *N*-acyl imidazole and alkyloxy *N*-acyl imidazole

##### 4.1.1 Synthesis of LDAI probes

Different starting material were used in literature in order to create LDAI probes such as urocanic acid **4.1**, 2-(1H-imidazol-4-yl)acetic acid **4.2**, and histamine dihydrochloride **4.3**.<sup>3,5-7</sup> As our final goal was to insert the probe onto a peptide we chose to first investigate the use of **4.1** and **4.2**. Due to the availability of compound **4.1** in the laboratory the first synthesis was based on **4.1** as the core centre for LDAI synthesis.

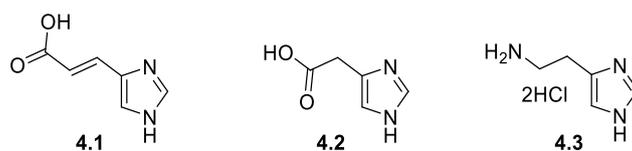
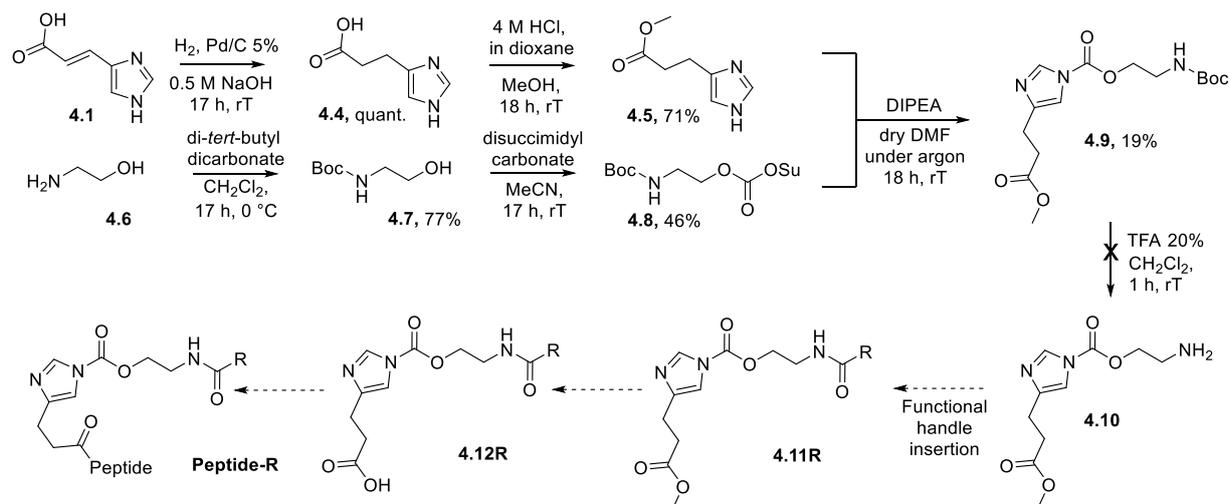


Figure 3: Structure of the starting material for LDAI chemistry.

Starting from **4.1**, compound **4.5** was obtained in a two-step literature process, Figure 4.<sup>8</sup> First a hydrogenation step was done to obtain **4.4** in quantitative yield and second an esterification to form **4.5**. This step was first carried out with H<sub>2</sub>SO<sub>4</sub> (cat) in MeOH, but the yield was relatively low, 26% yield. Thus, it was then repeated using 4 M HCl in a mixture of dioxane and methanol to form this time compound **4.5** with a 71% yield. This second step was done to facilitate reactivity and purification of the following compounds due to the presence of an ester instead of a carboxylic acid. It was thought that then a small linker **4.8** could be introduced onto the imidazole core to produce once deprotected an amine which could be used as a generic handle for any further modification, similarly to literature procedure.<sup>5</sup> Ethanolamine, **4.6**, was used as a starting material leading to **4.8** obtained with satisfying yields compared to literature (77% and 46% for the respective steps).



#### Literature example for acid-labile group deprotection on LDAI

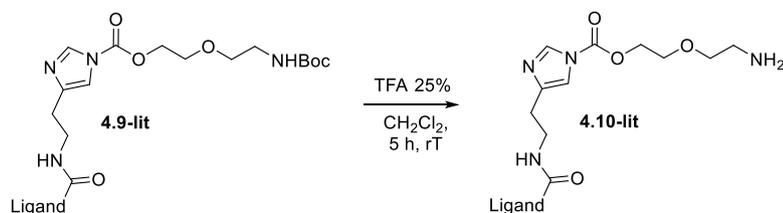


Figure 4: Scheme for the synthesis of imidazole probe for LDAI using fluorophore and biotin as R groups, with at the bottom the literature precedent deprotection step on LDAI probes.

From compound **4.5** and **4.8**, the first alkyloxyacyl N-acyl imidazole probe **4.9** as a carbamate was formed in 19%. The yield of the reaction was significantly lower compared to literature in which this

type of coupling showed yields usually between 45-70%. However, having isolated enough of material **4.9** for future steps, the following deprotection step in TFA was carried out. Interestingly, **4.10** was not obtained, despite usual reaction set up for Boc deprotection, showing that **4.9** was unstable during the acid treatment, despite what has been achieved in literature as shown in the bottom of Figure 4.<sup>5</sup> The analysis showed that instead **4.9** was decomposed back to product **4.5**. We hypothesised that this phenomenon was not necessarily due to the instability of the alkyloxyacyl imidazole towards acid, but the formation of a free amine at a relatively convenient position away. Indeed, compared to the literature example, where the nitrogen was placed seven atoms away from the electrophilic centre, in compound **4.10**, only four atoms was separating the electrophilic centre from the free amine which could favour an intramolecular nucleophilic attack as represented in Figure 5 following Baldwin's rule.

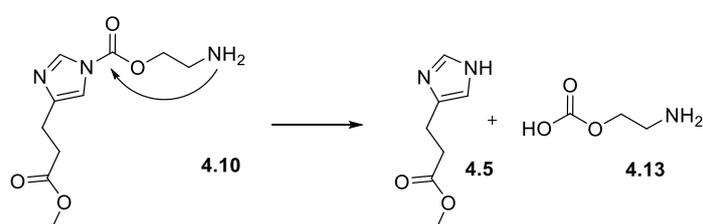


Figure 5: Proposed intramolecular decomposition of **4.10**.

To investigate this further, we synthesised a benzyl protected amine that would not be cleaved in acid conditions. Using again ethanolamine as the linker, compound **4.15** was synthesised and modified to form carbamate, **4.16**, in better yield than its Boc-protected predecessor (87% compared to 46%), Figure 6. The last step was the alkyloxyacyl bond formation between **4.16** and **4.5**. The protocol was slightly changed using DMAP as a nucleophilic catalyst which significantly improved the yield of compound **4.17** to 64%.

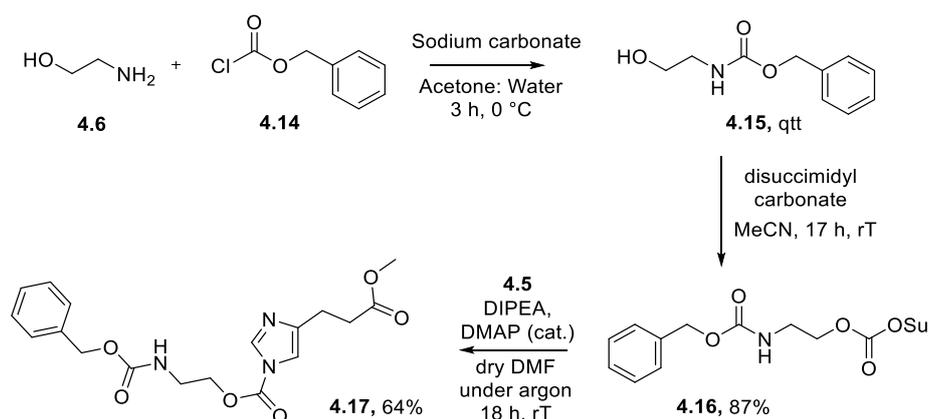


Figure 6: Scheme to develop new imidazole protected LDAI probes

Compound **4.17** was then put in an acidic solution, 20% TFA in  $\text{CH}_2\text{Cl}_2$ , and after 5 h less than 10% of degradation was observed, implying that the new alkyloxyacyl *N*-acyl bond is stable to acidic condition,

Figure 7 and the result obtained with compound **4.10** during deprotection is likely to be due to the presence of the now unprotected amine, as mentioned above, reinforcing the ring closure hypothesis.

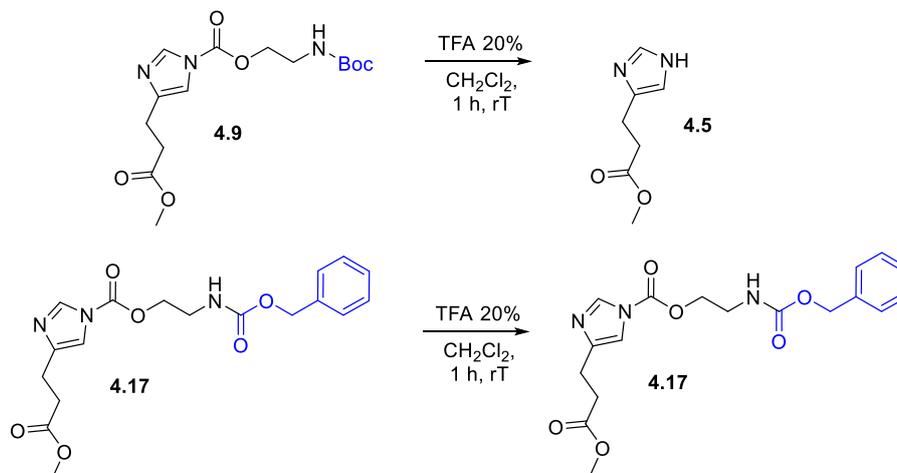


Figure 7: Scheme demonstrating the impact of acid on LDAI probes **4.9** and **4.17**

Following these results, we decided that linkers containing residual nucleophilic group would be unsuitable. Having those considerations in mind, a new probe was synthesised, compound **4.18**, Figure 8, which had the advantage of providing a convenient handle for CuAAC, and in the future could also act as an anchor for material conjugation when transferred onto the protein for further modification. Instead of coupling **4.18** to **4.5** as performed previously we considered doing the coupling onto the peptide directly to improve conversion to the alkyloxyacyl *N*-acyl bond.

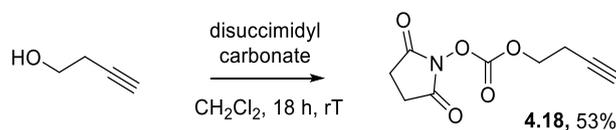


Figure 8: Scheme for the synthesis of a click chemistry probe for LDAI application.

#### 4.4.2 LDAI chemistry on peptides

In order to apply the chemistry detailed in the previous section to peptides, compound **4.4** was reacted with Fmoc-Cl to give compound **4.19** which could be used directly in peptide synthesis as an unnatural amino acid to introduce an imidazole group at the peptide *N*-terminus. However, **4.19** was not soluble enough in DMF for automated peptide synthesis and couplings were instead done manually using general conditions for peptide modification detailed in Chapter 2 with the coupling agents EDC.HCl and HOBt. The addition showed full conversion of the **P5** and **P6** with compound **4.19** added successfully to the *N*-terminus and Fmoc-deprotected using piperidine solution.

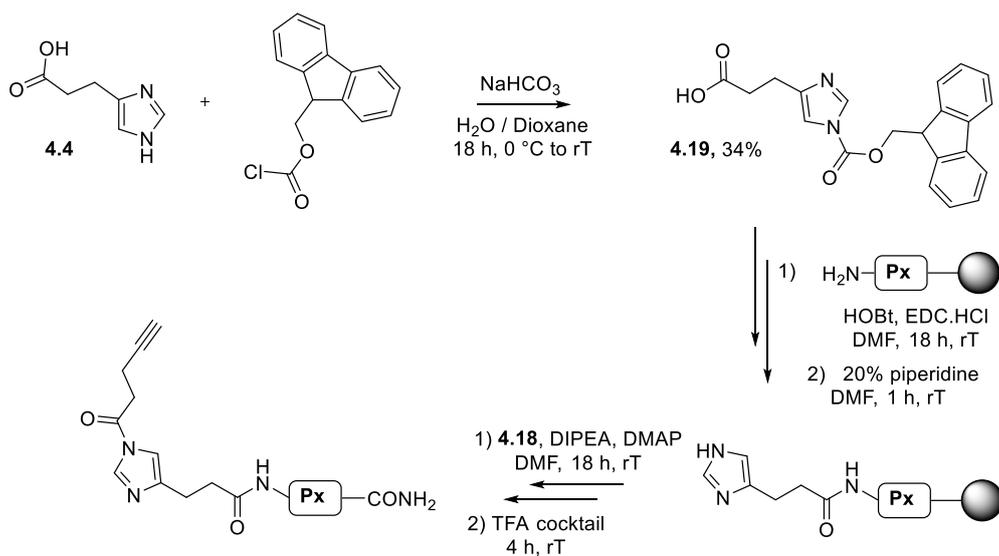


Figure 9: Reaction scheme for the formation of Fmoc-imidazole and its introduction on peptide on solid support. Compound **4.18** was subsequently inserted onto the *N*-terminus with full conversion by LCMS to the expected probe for both peptide **P5** and **P6** giving **Alk-Im-P5** and **Alk-Im-P6** respectively. However, after lyophilising the sample, further analysis revealed that hydrolysis had occurred regenerating **Im-P5**, Figure 10. Literature had previously reported LDAI reagents to have a half-life measured of 16 h at pH 7.2 and 37 °C,<sup>7</sup> and this instability was therefore surprising, but made the use of LDAIs on peptides potentially challenging.

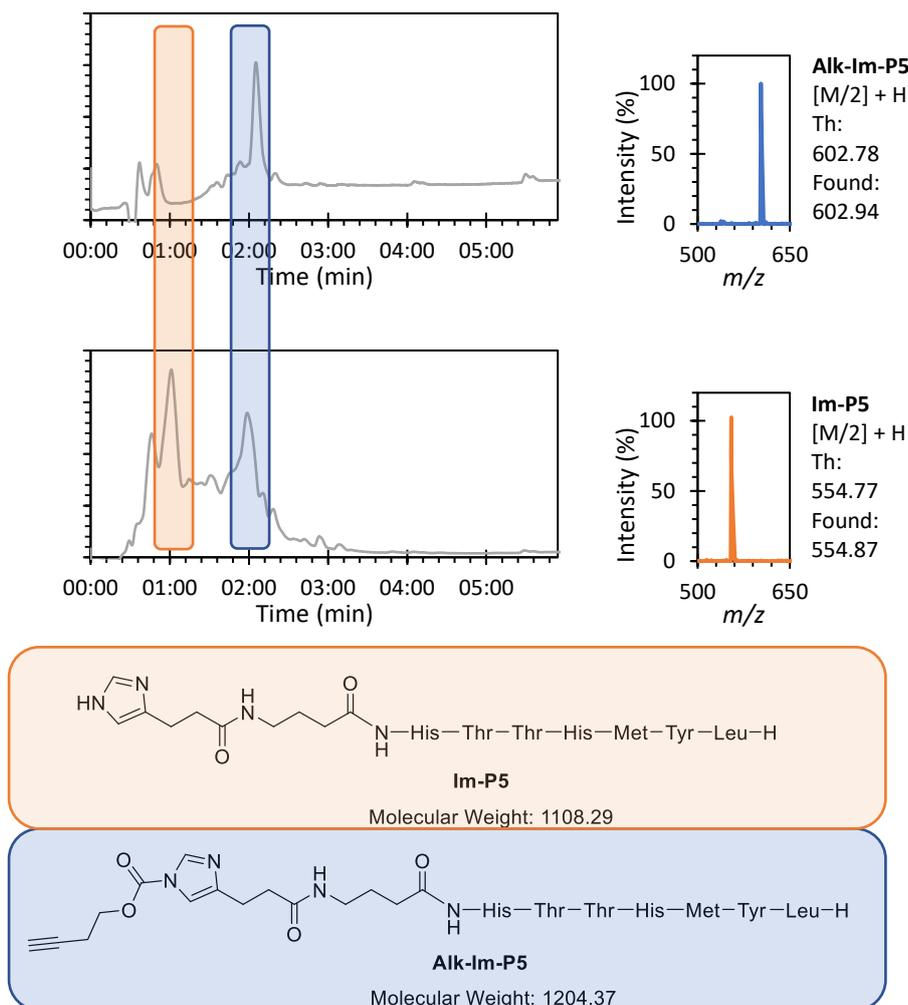


Figure 10: Chromatograms of modified peptide *Im-P5* with 4.18 before and after lyophilisation, and respective structures.

Thus, to study this aspect further, **Alk-Im-P6** was left in solution, in water at room temperature, and analysed by LCMS every hour over 24 h. Water was used instead of buffer at this point for most relevance to the purification process. It is expected at that stage that the resulting pH is relatively acidic due to the TFA salts coming from the cleavage steps. It was found that the degradation from **Alk-Im-P6** to **Im-P6** was happening rapidly, as in less than 3 h no more peptide peak for **Alk-Im-P6** could be observed, Figure 11. This experiment emphasised the fact that these type of probes are not suitable for the application due to a very short half-life.

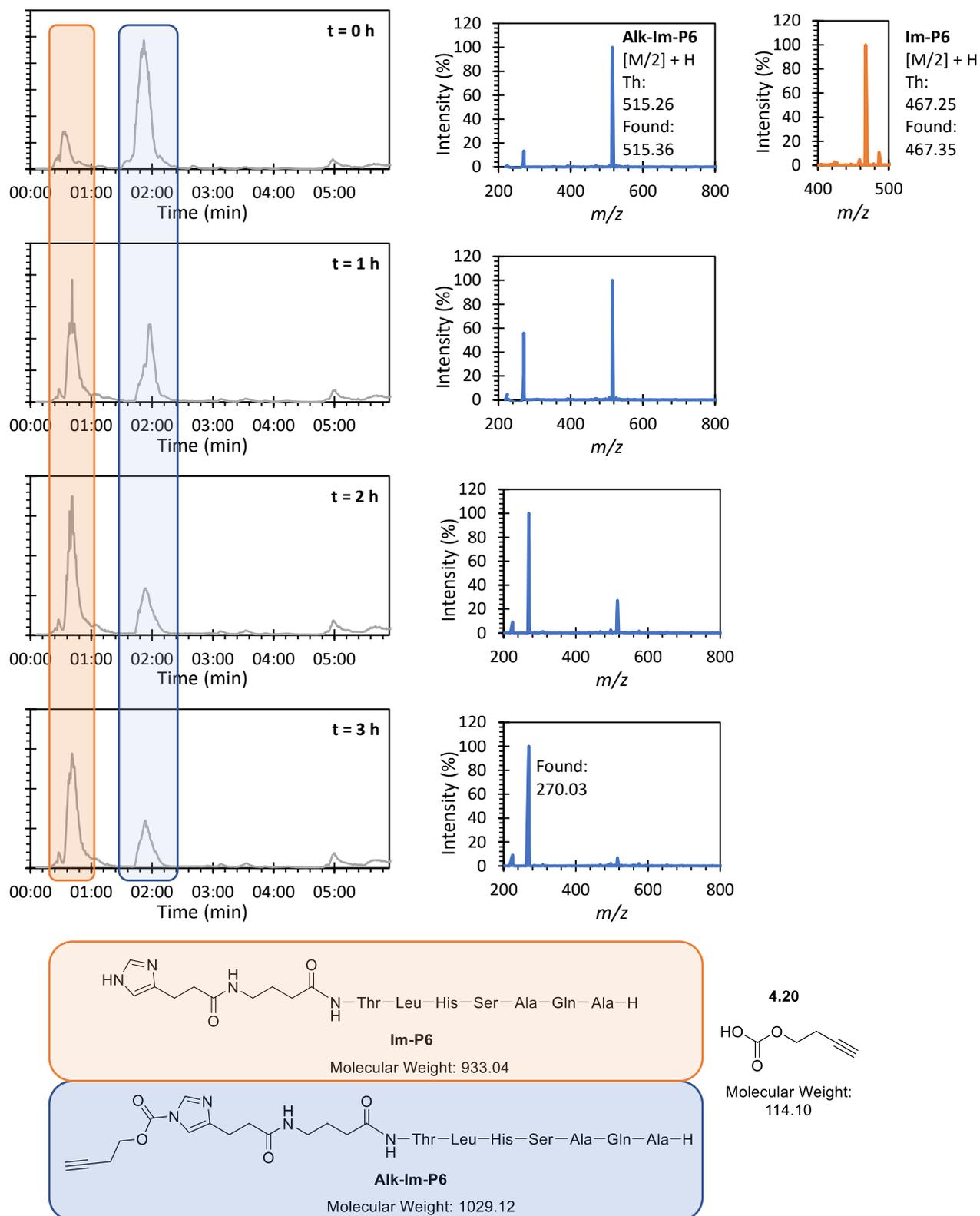


Figure 11: Chromatograms of modified peptide Im-P6 with 4.18 before and after lyophilisation, and respective structures.

#### 4.1.3. Conclusion on LDAI chemistry

To conclude on LDAI, it was shown that the alkyloxyacyl imidazole cores were reactive with nearby nucleophiles during organic synthesis and the products were also sensitive to hydrolysis. Moreover, manual modification of the *N*-terminus was necessary due to the poor solubility of compound **4.19** in DMF. While purification could have been attempted, or flash freezing in liquid nitrogen to limit the time in solution, the work on LDAI was not carried further to focus on LDNASA probes for *OpLD* and *CatLD*.

## 4.2 Synthesis of LD methylcyano-NASA probes for *Op*LD

### 4.2.1. Background and synthesis plan

Hamachi's group developed in 2018 new NASA probes for *Op*LD.<sup>9</sup> Their study compared different electron withdrawing alkyl groups on the NASA species and studied their reactivity. They reported that methylcyano-NASA esters showed higher reactivity and better aqueous solubility, two key factors for protein modification, Figure 12, than previously reported *p*-nitrobenzyl NASA.<sup>9,10</sup>

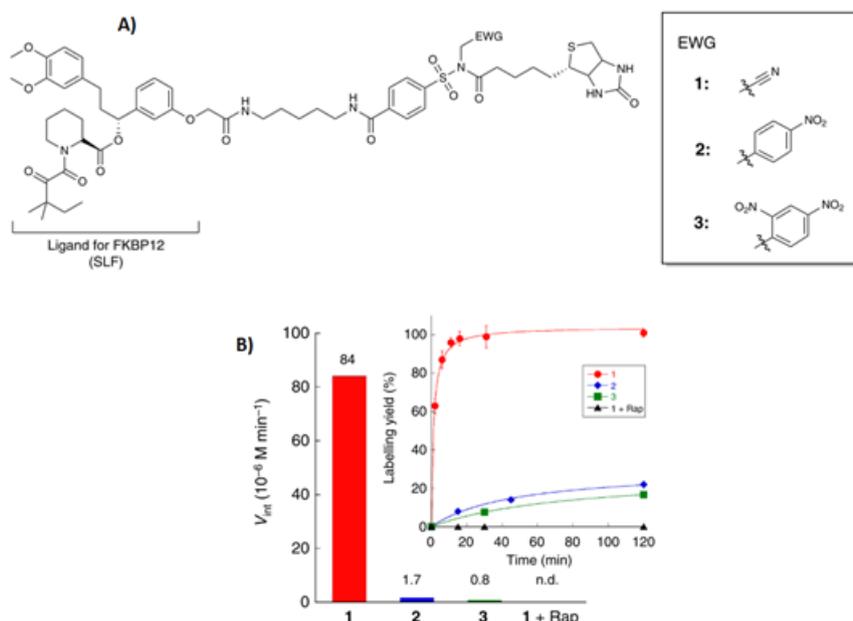


Figure 12: Figure taken from literature<sup>9</sup> showing *In vitro* FKBP12 labelling with LDNASA reagents 1–3. A) Molecular structures of LDNASA reagents 1–3. B) Initial rates ( $M \text{ min}^{-1}$ ) of FKBP12 labelling by 1–3.

In this thesis, in order to build methylcyano-NASA probes, building block 4-sulfanidobenzoic acid, was used as a starting point. This compound can be linked to functional ligands by amidation reactions at the carboxylic acid “tail” and the NASA species can be built onto the sulfonamide “head”, Figure 13.

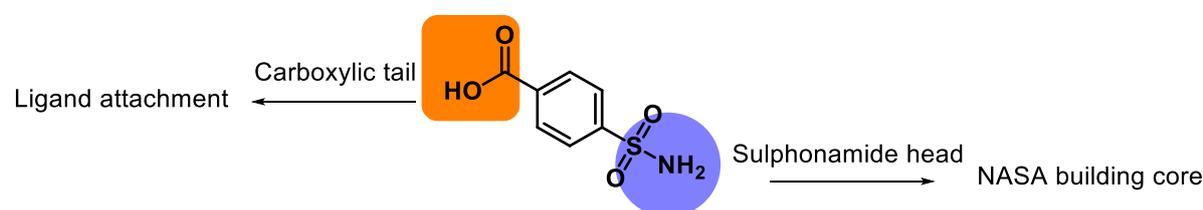


Figure 13: Construction of NASA species starting from 4-sulfanidobenzoic acid

Two approaches were considered to synthesise methylcyano-NASAs: first, small molecule organic synthesis to generate the NASA which could then be coupled to the peptide ligand; or second, use a peptide modification approach by doing reactions directly on the solid-supported peptide. Both methods have advantages. The first approach would build a bank of probes that could be coupled with any peptide ligand, focusing mainly on the sulfonamide head transformation. However, synthesis of

the probes in solution would necessitate purification. The second approach facilitates considerably the need for purification at each step as all the excess of reagents will be washed off before cleaving the peptide from the resin. However, this method requires bespoke synthesis for each ligand. Another drawback would be the risk that one of the reactions fails and that this is known only after cleavage. Both ways were investigated.

#### 4.2.2. A general small molecule approach for building methylcyano-NASA

Compared to literature in which 4-sulfanidobenzoic acid was first coupled with the linker of interest and finally modified to the methylcyano-NASA using iodoacetonitrile, in this PhD we wanted to build first an intermediate **4.21** which already contained the methylcyano component, Figure 14. This approach was investigated as we expected it to decrease the time needed for each new NASA to be synthesised in one step from a common intermediate. Moreover, this approach would avoid the risk that other amides in the backbone could react with the iodoacetonitrile reagent. Two approaches, were possible to synthesise **4.21**.

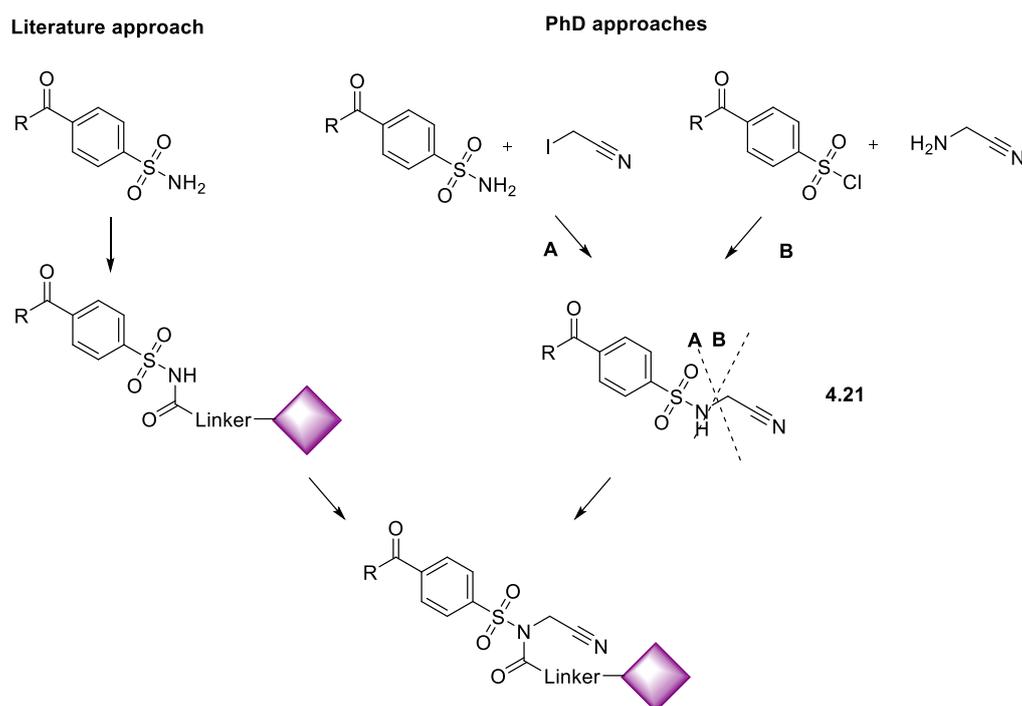


Figure 14: New building block for nitrile-NASA compounds and its retrosynthetic approach

Approach A would use a sulfonamide compound which could react with iodoacetonitrile, while approach B used aminoacetonitrile hydrochloride to react with a sulfonyl chloride.

First, approach A was attempted, Figure 15. From the commercially 4-sulfanidobenzoic acid, **4.22**, a functionalisation of the carboxylic acid using  $\beta$ -alanine methylester as a small linker was achieved first in a 50% yield. The second step was using iodoacetonitrile for the transformation to the methylcyano building block. At room temperature overnight only 9% conversion to product **4.25** was observed with

most of the starting material **4.24** being unmodified. The reaction was tried again at 60 °C to increase the reactivity of the sulfonamide, leading to the starting material being fully consumed however no compound **4.25** was formed. The mass of the product isolated would match compound **4.26** as if a double addition happened which could be explained as we used 4.5 eq of iodoacetonitrile to sulfonamide. This formation could be decreased by adding 1 equivalent of iodoacetonitrile or by decreasing the temperature for a compromise between the two attempts. However, this was not carried further as approach B was found to be more promising.

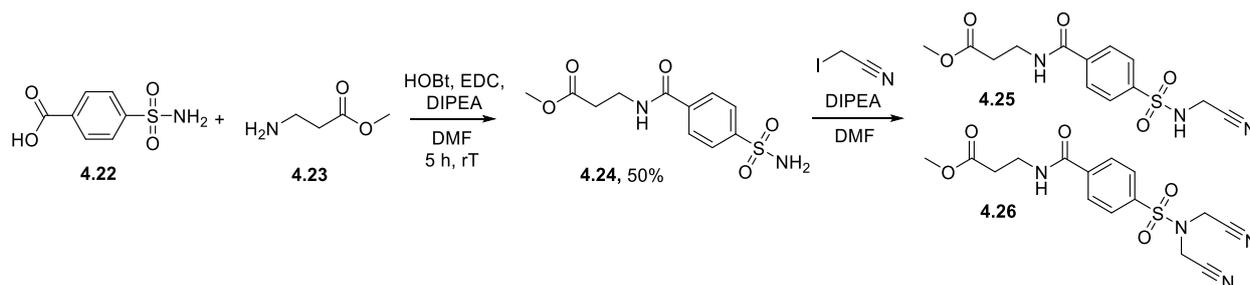


Figure 15: Introduction of methylcyano group onto the sulfonamide using approach A

Approach B was attempted using 4-(chlorosulfonyl)benzoic acid **4.27** as the starting material, Figure 16. The first step was similar to approach A, protecting the carboxylic acid tail. Instead of using an amidation protocol, in which we suspected that the sulfonylchloride could also react, a methylation was achieved in an 85% yield using MeOH after activation of the starting material with oxalyl chloride in DMF. Compound **4.28** could then react with propargyl amine to form **4.29** which was obtained in a reasonable 39% yield without formation of major side-product. The yield could be increased by keeping conditions as dry as possible to ensure compound **4.28** does not undergo hydrolysis of the sulfonylchloride, however enough material was obtained to pursue this approach without further optimisation.

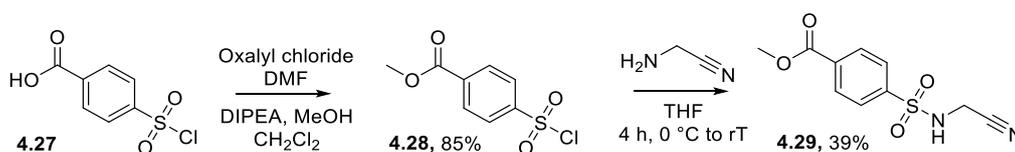


Figure 16: Introduction of methylcyano group onto the sulfonamide using approach B

The following step was the modification with a carboxylic acid in order to obtain a final NASA probe. A first attempt using a carboxylic moiety in DMF with the coupling agents HOBt/EDC.HCl/DIPEA and catalytic DMAP was not successful and compound **4.29** was recovered following column chromatography. In order to have a more reactive species, butyric acid was converted *in situ* to its activated acid chloride form which reacted then successfully with **4.29** to form compound **4.30** as

shown in Figure 17. The difficult purification of the product only led to 21% yield which could be improved in the future.

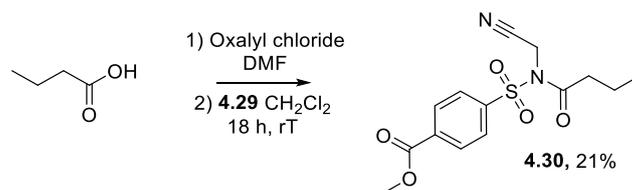


Figure 17: Final step to form a methylcyano-NASA probe using an organic approach.

Thus, to conclude, a new successful route was developed building an interesting new methylcyano sulfonamide building block compound **4.29** after some optimisation in a two-step process. The next step will be to broaden the strategy with more complex carboxylic acid compounds in the last step to achieve a library of NASA compounds in another two-step process, cleaving the ester and coupling it to the target peptide. In parallel we also investigated direct modification of the target peptide on solid-phase. This approach is described in the following section.

#### 4.2.3. A solid-phase approach to build methylcyano-NASA

The solid-phase approach was similarly started from 4-sulfanidobenzoic acid, compound **4.22**, as this species contains a carboxylic acid group able to react directly with a free amine at either the *N*- or *C*-termini of a peptide, Figure 18. Using the protocol from Chapter 2 for on-resin amide coupling, **4.22** was inserted onto peptides with full conversion, to generate **Sulf-P5** and **Sulf-P6** see Experimental section.

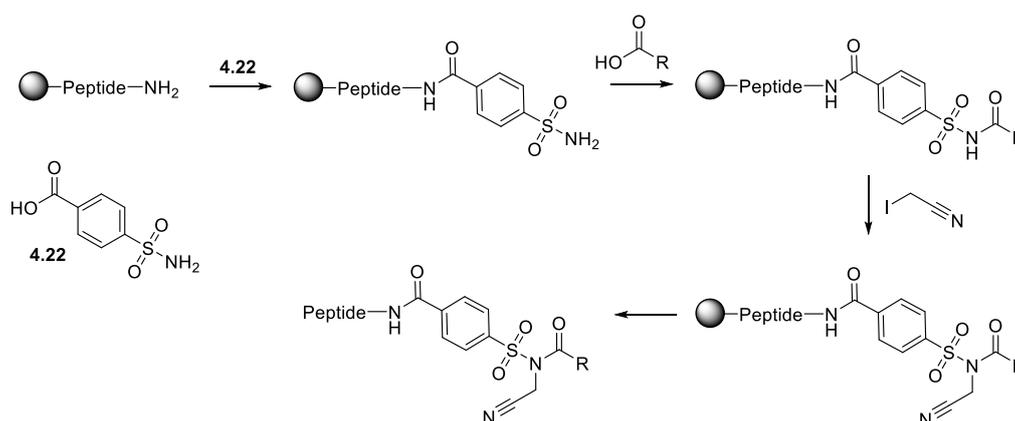


Figure 18: Representative approach to build methylcyano NASA on solid-phase

The next step was then the formation of the sulfonamide bond with a carboxylic acid of interest, pentynoic acid **4.31**, as a future handle for click-chemistry, represented in Figure 19. Adapting literature procedure for the formation of sulfonamide,<sup>11</sup> DMAP was introduced to act as a nucleophilic catalyst with the coupling agents HOBt/EDC.HCl. Even with 10 equiv. of DMAP at room temperature only around 30% of the products **Alk-Sulf-P5** and **Alk-Sulf-P6** were obtained after 4 hrs. While the data

for **P5** are presented below, the data for **P6** is presented in the Experimental section. The reaction was thus carried over for another 36 h and it was shown a 50% conversion had been achieved, Figure 19. While it is generally preferred to reach greater conversion, the fact that the product and starting material were easily separated by reverse phase chromatography allowed us to continue with the synthesis.

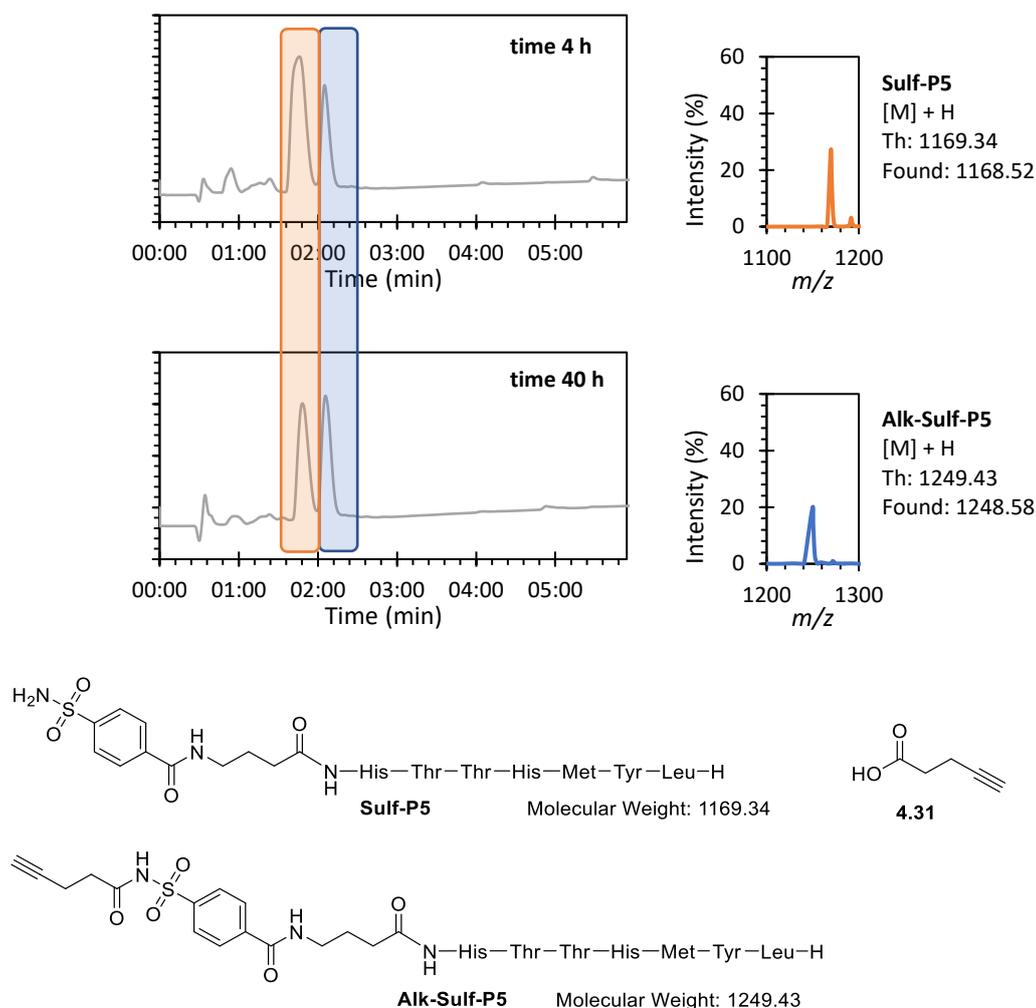


Figure 19: Chromatograms at 220 nm of modified peptide Sulf-P5 with **4.31** at 4 h (top) and 40 h (bottom) at rT with their respective structures.

The last step of the process was the most challenging, first due to the potential side-reactivity of the iodoacetonitrile with the different amide bonds present in the backbone of the peptide structure, and second due to the multiple optimisations required to reach full conversion on resin varying parameters such as equivalents and temperature. A first reaction at room temperature with 5 eq of base DIPEA and 5 eq of iodoacetonitrile overnight was performed and no modification of the peptide was observed. In order to avoid using too much peptide resin, this pathway was optimised on a dialanine model peptide on which the 4-sulfanidobenzoic acid was added during the peptide synthesis, forming

product **4.32**. The addition of **4.31** was achieved following the same protocol as for the previous peptide examples, with 62% of product **4.33** being obtained, Figure 20.

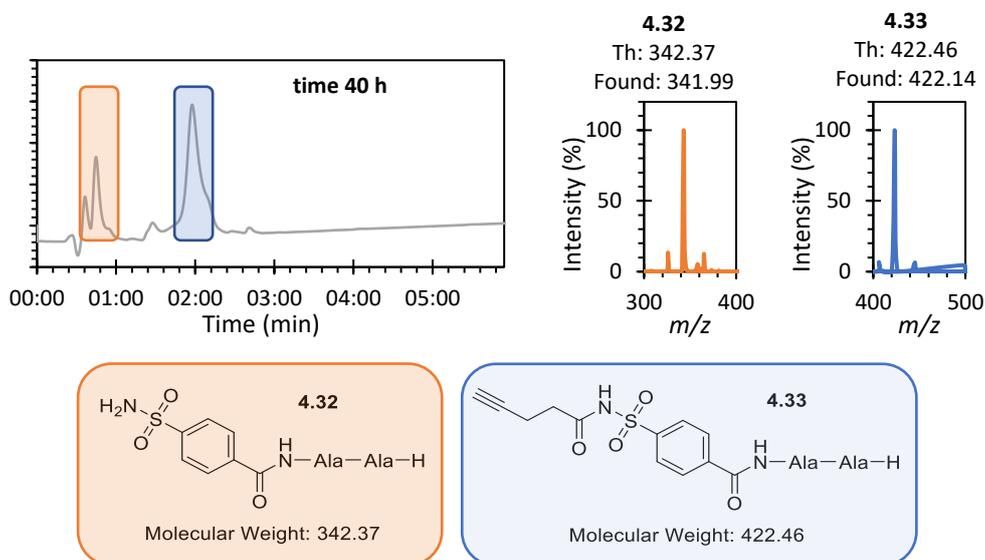


Figure 20: Chromatograms at 220 nm of modified **4.32** with **4.31** for 40 h at rT to compound **4.33** with their respective structures.

**4.33** was used to screen a set of conditions to give the final NASA compound **4.34**. The temperature was set to 30 °C for modification of **4.33** on resin, and both equivalents of iodoacetonitrile and DIPEA were set to either 10 equivalents (100 mM) or 100 equivalents (1 M) in a DMF solution. Four different time points were studied, 2 h, 4 h, and 6 h. Moreover due to the instability noticed during the formation of LDAI reagents, the same samples were studied over a day to ensure no hydrolysis or degradation of compound **4.34** was occurring.

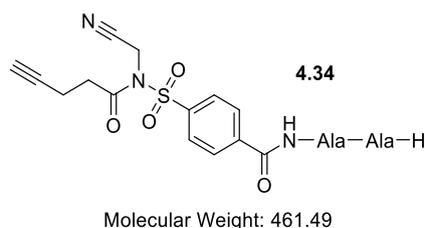


Figure 21: Structure of the expected final NASA compound **4.34**

A first observation at the first time point 2 h was the significant impact of the concentration. While for 10 eq of iodoacetonitrile, no peaks were observed other than the starting material at 1.2 min, using 100 eq, a new peak around 2.4 min appeared with a mass that matched the expected compound **4.34**, Figure 22. While the relative UV absorption of both compounds **4.33** and **4.34** were not known, the average spectra between 210 and 400 nm was used to estimate a conversion of ~40-50%. Another important observation was that at 4 h, no distinction could be made between the spectra obtained at

2 h. Indeed for 10 eq still no peak was observed for the expected product and using 100 eq the ratio between the two UV peaks showed a similar conversion.

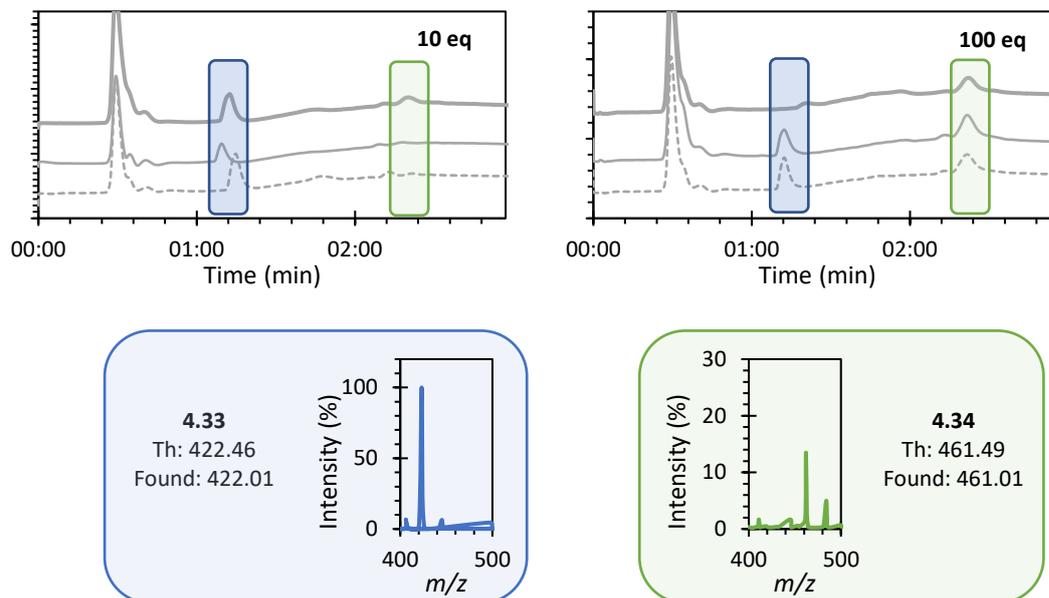


Figure 22: Chromatograms average 210-400 nm of modification of **4.33** with iodoacetoneitrile at 30 °C. Chromatogram for 10 eqs (left) and 100 eqs (right) with in dash  $t = 2$  h, plain  $t = 4$  h and bold  $t = 6$  h.

However, at 6 h more distinctive changes could be observed. First for 10 eq of iodoacetoneitrile, a first peak could be observed for the product, with around 15% conversion, and more importantly at 100 eq, no more starting material was observed suggesting full conversion to product. Moreover, as mentioned the stability was studied. The same sample showing full conversion was thus analysed after 20 h. LCMS revealed that the chromatograms were almost identical, Figure 23 and that product **4.34** could still be observed with no apparition of new masses showing that the compound could be kept in water for 20 h without undergoing hydrolysis or other degradation.

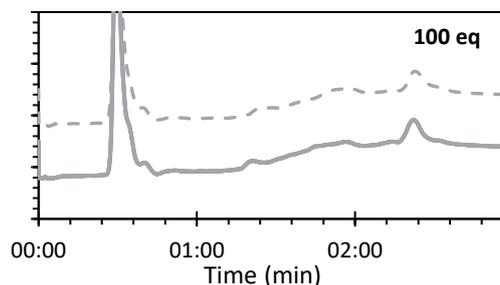


Figure 23: Chromatograms average 210-400 nm of modification of **4.33** with iodoacetoneitrile at 30 °C. Chromatogram in bold of the formation of product **4.34** using 100 eq of iodoacetoneitrile and chromatogram in long dash showing the stability of the product over 20 h.

To conclude this part on the synthesis of LDNASA on resin, we showed that literature conditions were not optimal. Indeed using room temperature and concentration of 100 mM,<sup>9</sup> were not sufficient to form the desired products. However, optimisation was performed to achieve the desired products in

relatively good yield using a temperature of 30 °C and a concentration of iodoacetonitrile of 1 M (100 eq). While the results obtained on the dialanine model peptide were promising, the time left in the PhD did not allow for the application of this chemistry to binding peptides.

#### 4.2.4. Conclusion methylcyano-LDNASA chemistry

As described in this section, methylcyano-LDNASA probes were synthesized for *Op*LD chemistry. Two approaches were tried first via small molecules and second via an on-solid support in order to synthesize probes which could be used on peptides. During the small molecule approach, we created a new route for more effective library building using a core intermediate **4.30**. This route could be broadened in the future to other carboxylic acids bearing reactive handles for further use once the handle has been transferred onto the POI. The on-solid support approach showed as well promising results with final conditions (6 h, 30 °C, 100 eq of iodoacetonitrile to peptide) showing full conversion forming compound **4.34**. Both routes could then be used to synthesise a wider range of probes for LDNASA chemistry, however the time constraint did not allow this, and the remainder of the PhD focused on developing probes for *Cat*LD chemistry described in the following section.

### 4.3 Synthesis of pyridinium oxime – catalyst for CatLD

Hamachi's group reported in 2017 a CatLD probe based on a pyridinium oxime (PyOx) catalyst which catalyses acyl transfer from mild electrophilic *p*-nitrobenzyl *N*-acyl-*N*-alkylsulfonamides (NASA) to nucleophilic residues present on the surface of proteins.<sup>11</sup> This section will focus on the synthesis of peptides bearing pyridinium oxime catalysts and a later section will describe the synthesis of the *p*-nitrobenzyl NASA substrates. The synthetic routes mentioned throughout the two following sections are based/inspired on literature introduced in Chapter 1.<sup>9–14</sup>

#### 4.3.1 Synthesis of PyOx peptides using literature procedure

Inspired by literature protocols,<sup>11</sup> PyOx probes could be synthesised in a two-step process with first the insertion of an alkyl bromide moiety on the peptide able to react in a second step with the commercial pyridine-4-aldoxime to form the desired pyridinium oxime PyOx, as represented in Figure 24. This process was carried out on resin.

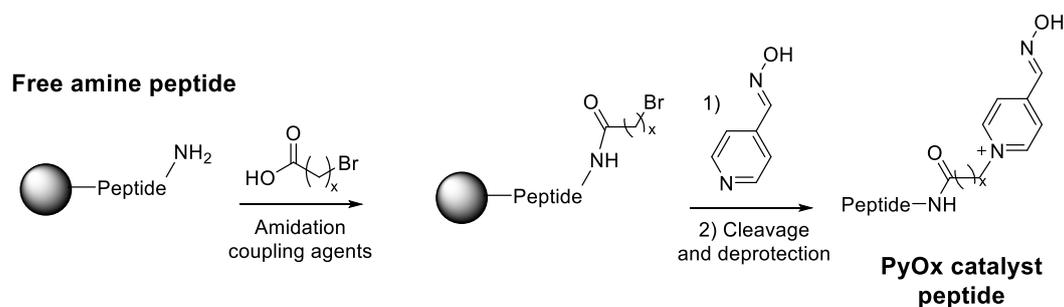


Figure 24: Scheme for preparation of pyridinium oxime (PyOx) catalyst on peptides

The first step was first carried out by using 7-bromoheptanoic acid which would react with the primary amine of the *N*-terminus of the peptide via amide coupling. While some processes for amidation were already detailed in Chapter 2, a new coupling agent used in literature for similar reactions was used, 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium chloride (DMTMM).<sup>11</sup> For on-resin modification, 10 equivalents of acid to the free amine of the peptide, with 15-20 equivalents of DMTMM and 15-20 equivalents of DIPEA were used. High conversions were reached in 2 h in DMF, **P2** was converted to **BrC7-P2** (C7 for heptanoic acid) with 72% conversion and **P4** to **BrC7-P4** in a 79% conversion. Following this, the reaction was also attempted at the *C*-terminus of **P2**. To do so the protocol for modification at a selectively deprotected lysine at the peptide *C*-terminus introduced in Chapter 2, was used. The modification of **Fmoc-P2-K** with 7-bromoheptanoic acid moiety gave a lower conversion of only 40%, but still enough for subsequent modification with the commercially available pyridine oxime. The LCMS of these results are presented in the Experimental section.

The second step was also adapted from literature.<sup>11</sup> After a first attempt, using 2 eq of pyridyl-oxime as a nucleophile at 80 °C for 20 h in MeCN, showed low conversion, a larger excess of oxime was used,

20 eq for 72 h, giving this time a 80% conversion to product **PyOx-C7-P4(trt)**. However, using standard peptide cleavage conditions, even after 5 h incomplete deprotection was observed via LCMS analysis. While this is problematic for further use of peptide **P4**, for other sequences such as **PyOx-C7-P2** the deprotected peptide was isolated effectively, suggesting this difficulty is sequence dependent.

Alkylation was then applied for modification of a peptide C-terminus. After coupling, two options were possible to recover the desired peptide from the resin as both Fmoc and acid-labile protecting groups needed deprotection, Figure 25.

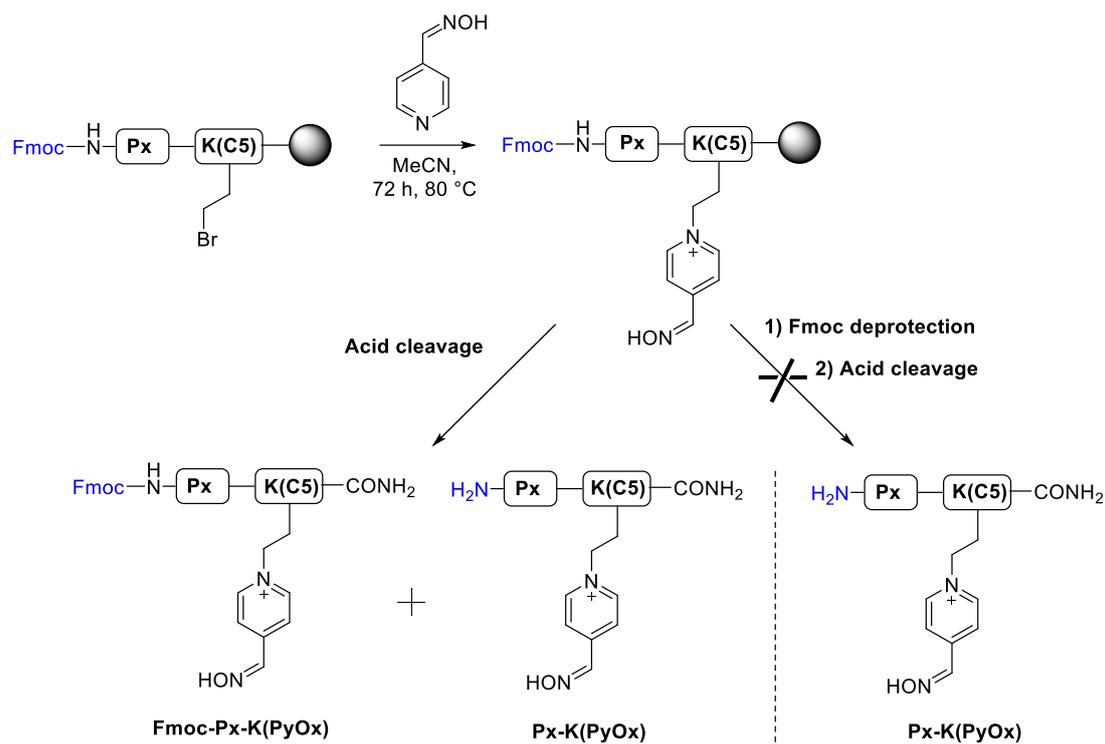


Figure 25: Scheme representative of the product obtained from the pyridinium oxime coupling step at the C-terminal lysine after 72 h at 80 °C. On the left hand side, use of acid cleavage only, and on the left hand side use of piperidine solution for Fmoc deprotection followed by acid cleavage.

In a first attempt a direct acid cleavage in TFA was attempted, with the LCMS analysis revealing a number of different peaks. We had hypothesized that the N-terminal Fmoc-group, under high temperature of 80 °C for 72 h required for the pyridinium coupling step, would be cleaved, which has also been reported in literature.<sup>15</sup> However, the main compound was the unprotected product **Fmoc-P2-K-PyOx** in a 50% conversion while the desired product **P2-K-PyOx** was present in a 15% conversion, Figure 26. Another peak with a mass of -1 compared to **P2-K-PyOx** was seen in a 12% conversion. At present the identity of this peak is unknown.

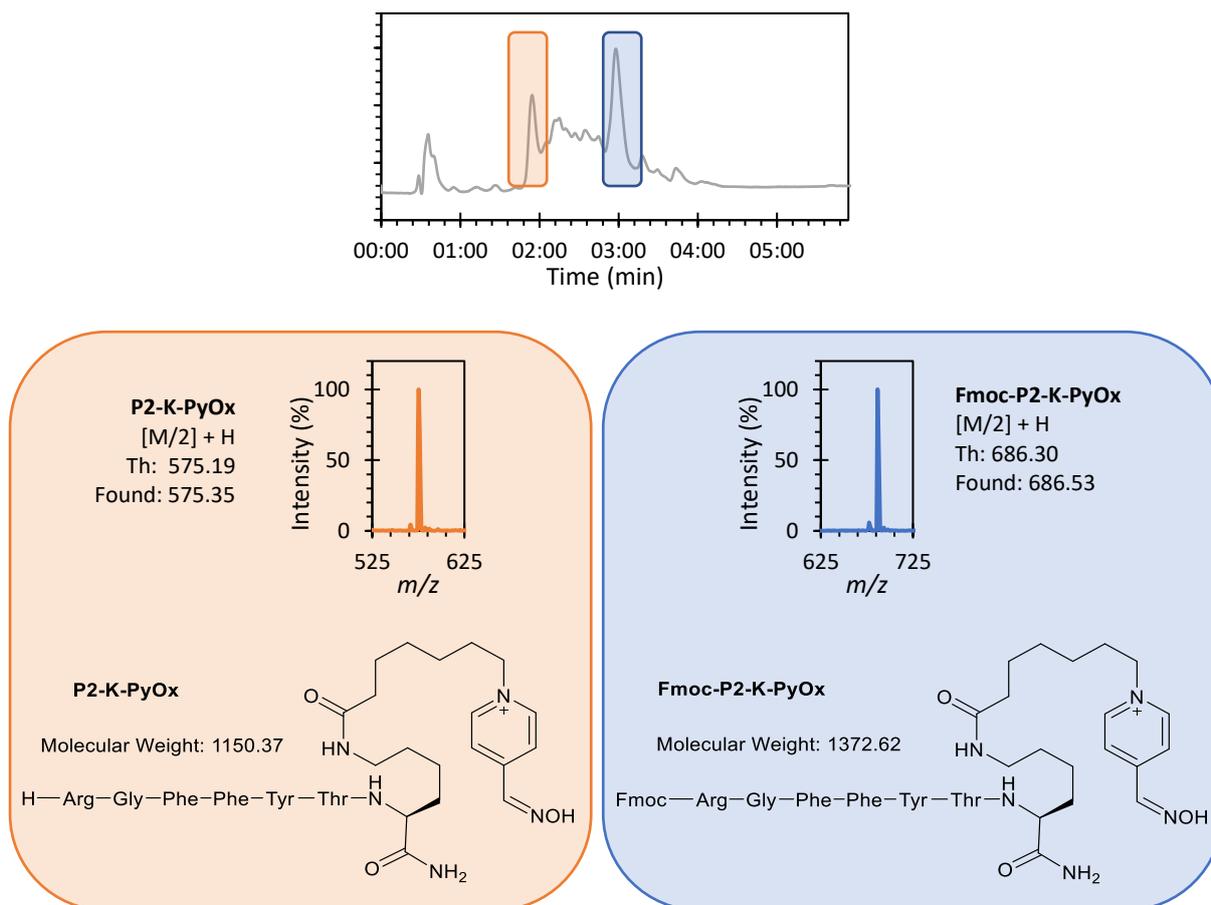


Figure 26: Chromatograms average 210-400 nm of the formation of **Fmoc-P2-K-PyOx** using pyridine-4-aldoxime and **Fmoc-P2-K-Br** in MeCN at 80 °C for 72 h followed by the respective structure of the peptides.

As an alternative, after coupling the PyOx group onto the peptide, Fmoc-deprotection was attempted using 20% piperidine in DMF, followed by acid cleavage. However, no peak for **P2-K-PyOx** could be observed after this treatment. The analysis showed that the main peak (60%) had a mass of 1063.6, Figure 27, significantly smaller (-86 m/z) than the one expected, showing a deletion or an elimination in the sequence. While the structure of the compound was not identified, this result showed that the PyOx was not stable to base and was most likely cleaved or modified during piperidine treatment.

To summarize this section, we showed that increasing equivalents was necessary to reach higher conversion to the PyOx peptide probes. However, while the chemistry works well on the *N*-terminus, it was shown that on the *C*-terminus, the need for Fmoc deprotection under basic conditions was not compatible with PyOx chemistry. While to avoid this problem *N*-terminal Boc-protection could be used, a new route using click-chemistry was attempted to limit the number of steps and thus the time to prepare PyOx peptides probes.

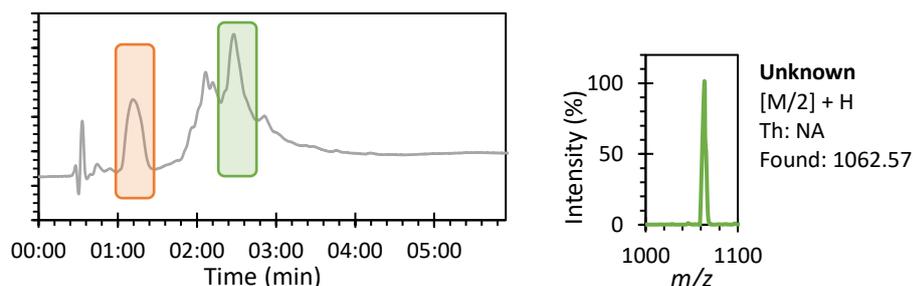


Figure 27: Chromatograms average 210-400 nm after basic deprotection of **Fmoc-P2-K-PyOx** leading to an unknown side product of Mw 1062.57.

#### 4.3.2 Synthesis of PyOx peptides using click-chemistry

As an alternative route we used copper catalyst-based click-chemistry to introduce the PyOx catalyst onto the peptide probes. This route was approached while working in parallel on the insertion of pyridinium oxime into 2-PCAs probes which will be discussed in detail in the following chapter. The azide-functionalised pyridinium oxime catalysts synthesised for 2-PCAs, presented in Figure 28, were used also to modify peptides, and their synthesis and characterisations are detailed in Chapter 5.

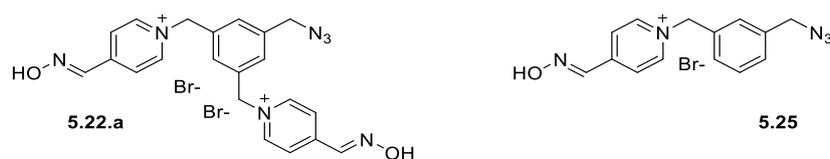


Figure 28: Structures of the two pyridinium oxime catalyst for click chemistry.

To achieve click-chemistry on peptides, alkyne residues had to be inserted into the sequences. This could be done relatively easily using commercially available Fmoc-propargyl glycine for C-terminal modification, or standard pentynoic acid for N-terminal functionalisation. Two additional peptides were synthesised using an extra PEG linker to modify the distance between the probe and the catalyst. While Fmoc-NH<sub>2</sub>-PEG<sub>6</sub>-CH<sub>2</sub>CH<sub>2</sub>COOH was used on the N-terminus, a carboxylic acid PEG bearing an alkyne, compound **4.36**, was synthesised from tetraethylene glycol in a two-step process, Figure 29.

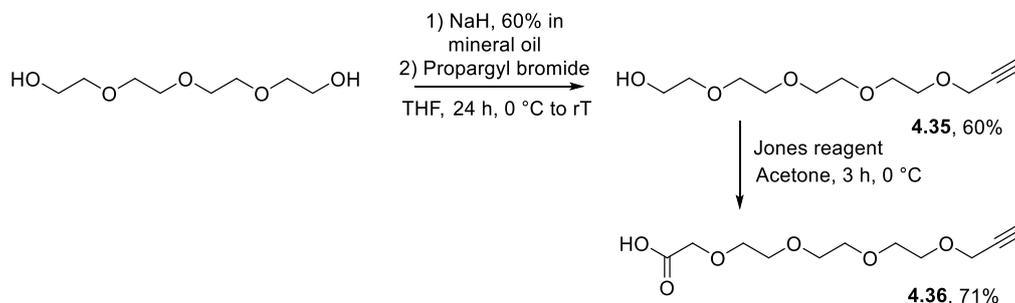


Figure 29: Scheme for the synthesis of a carboxylic acid PEG alkyne, compound **4.36**

Thus, four derivative peptides from sequence P2 were obtained, Figure 30, in a straight-forward manner using the SPPS protocols followed by one standard amidation coupling detailed in Chapter 2.

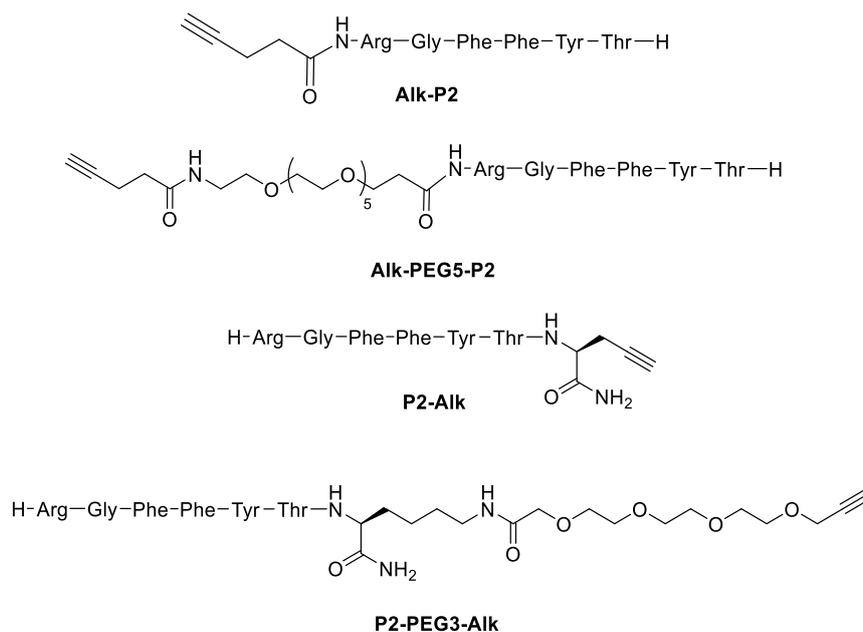


Figure 30: Structures of the modified P2 sequence with alkyne for click-chemistry

Each of these four peptides were then clicked with compounds **5.22.a** and **5.25** respectively to form eight peptide probes, Figure 31, which were all purified by reverse-phase chromatography. Most peptides were >90% pure after purification, but in some cases some impurities were present. Due to the time constraints of the project the probes showing higher purity were used in Chapter 6 to ensure their reactivity before broadening the scope to the less pure peptides which in future would require HPLC purification.

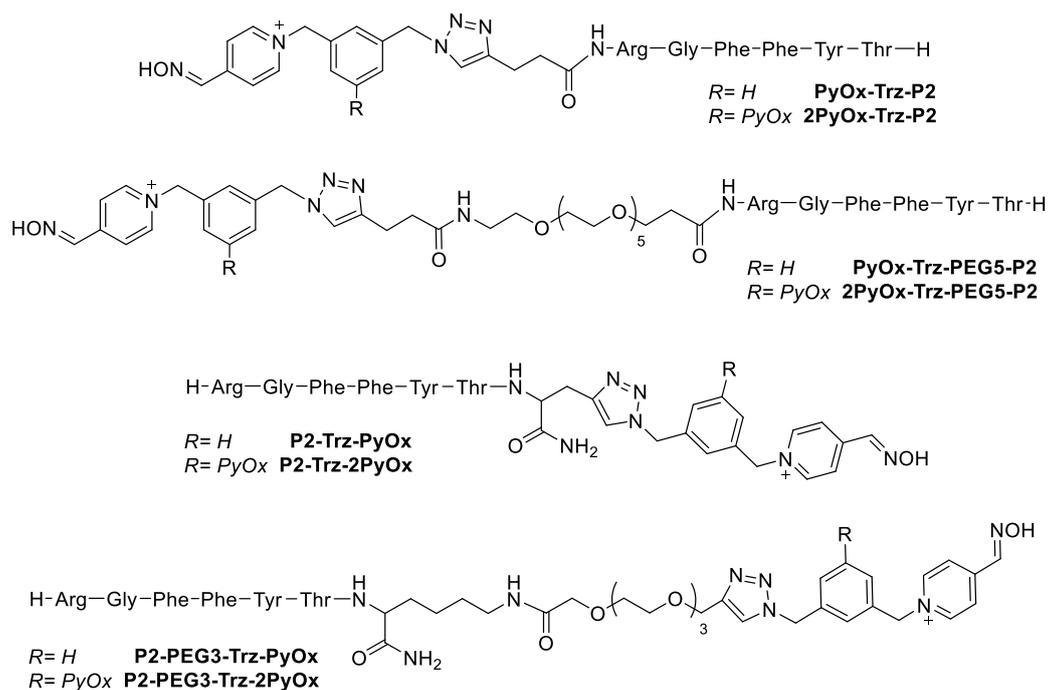


Figure 31: Structure of the PyOx-P2 modified peptide containing either a monovalent or divalent PyOx catalytical head

#### 4.3.3 Conclusion on LDNASA chemistry

To conclude this section, it was possible to form PyOx *N*-modified peptide both by adapting literature procedures or via click-chemistry, by inserting alkynes into the peptides and using azide PyOx probes. When the PyOx modification was desired at the *C*-terminus, the first approach showed limitations due to the sensitivity of the different orthogonal groups needed for specific modification of the terminus. For this reason the click-chemistry approach had advantages and was therefore used in subsequent work.

#### 4.4 Synthesis of *p*-nitrobenzyl NASA – substrates for *CatLD*

The last section of this chapter will describe the synthesis of *p*-nitrobenzyl NASAs, as substrates for *CatLD* chemistry. While NASA reagents had already been mentioned in the previous section in *OpLD*, the main differentiation between the reagents synthesised in this section is the type of electron-withdrawing group used on the sulfonamide. Indeed while methylcyano substituents provided good substrates for *OpLD*, *CatLD* requires a less reactive substrate bearing *p*-nitrobenzyl groups, Figure 32.<sup>9</sup>

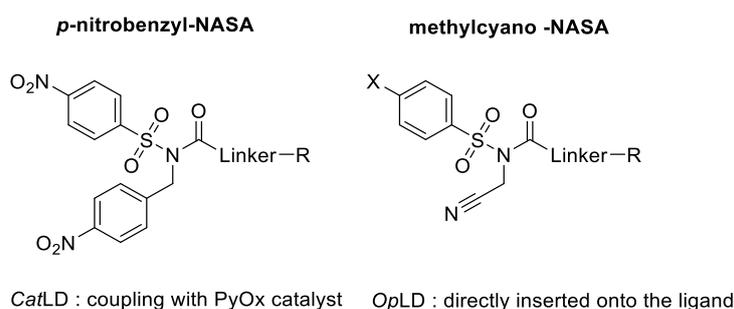


Figure 32: Distinction between the two types of NASA species

##### 4.4.1 Synthesis of *p*-nitrobenzyl-NASA using literature procedure

Chemical routes for NASA probes have been reported in literature<sup>11,12</sup> using as a starting material compound **4.37** 4-nitrobenzenesulfonamide. First carboxylic acids are able to react with the amine of the sulfonamide, and second once the amide bond is formed, the amide nitrogen can be alkylated with 4-nitrobenzyl bromide via a  $S_N2$  reaction to form a generic *p*-nitrobenzyl-NASA, Figure 33.<sup>11</sup> In the first step, due to the reduced nucleophilicity of the sulfonamide group relative to a standard amine, the nucleophilic catalyst DMAP is a necessary additive to reach good yields.

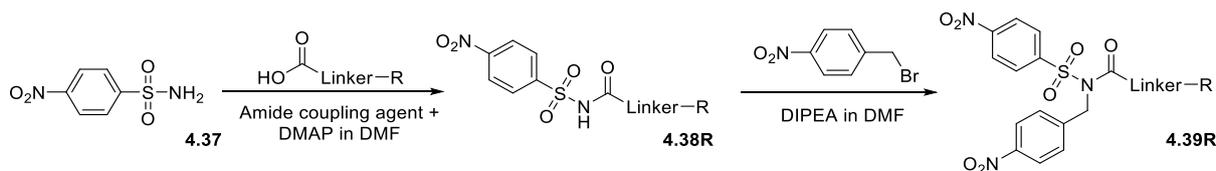


Figure 33: General scheme to form *p*-nitrobenzyl-NASAs from 4-nitrobenzenesulfonamide using literature procedure.<sup>11,12</sup>

Different R groups were investigated in this thesis to provide useful functional handles for protein modification. However to start, hexanoic acid was picked as a simple non-functional model.

Compound **4.38-hex** was obtained after purification in a 90% yield. The second step following literature conditions<sup>11</sup> with only DIPEA as a base in DMF at room temperature didn't give the expected compound **4.39-hex**. We introduced potassium iodide to activate the bromide *in situ* which successfully led to the product **4.39-hex** in a 39% yield, with an overall yield 35% over two steps.

A biotin moiety was also considered practical in order to enable western Blotting of the future modified proteins. The biotin-NASA **4.38-biot** was synthesized following literature procedures, without the need of the addition of KI, with 56% and 51% yields for first and second step respectively comparable to the literature, leading to an overall yield of 29%.<sup>9</sup> Following the success of these first two NASA syntheses, three more NASA with interesting handles were synthesised using the optimised protocol for the formation of NASA-hexanoic. The third NASA to be synthesised contained a PEG moiety, **4.39-PEG**, which was considered to maximise the solubility, while the last two contained two fluorophores which would help for imaging, compounds **4.39-NBD** and **4.39-BODIPY** from fluorophores **3.3** and **3.4** synthesised in Chapter 3.

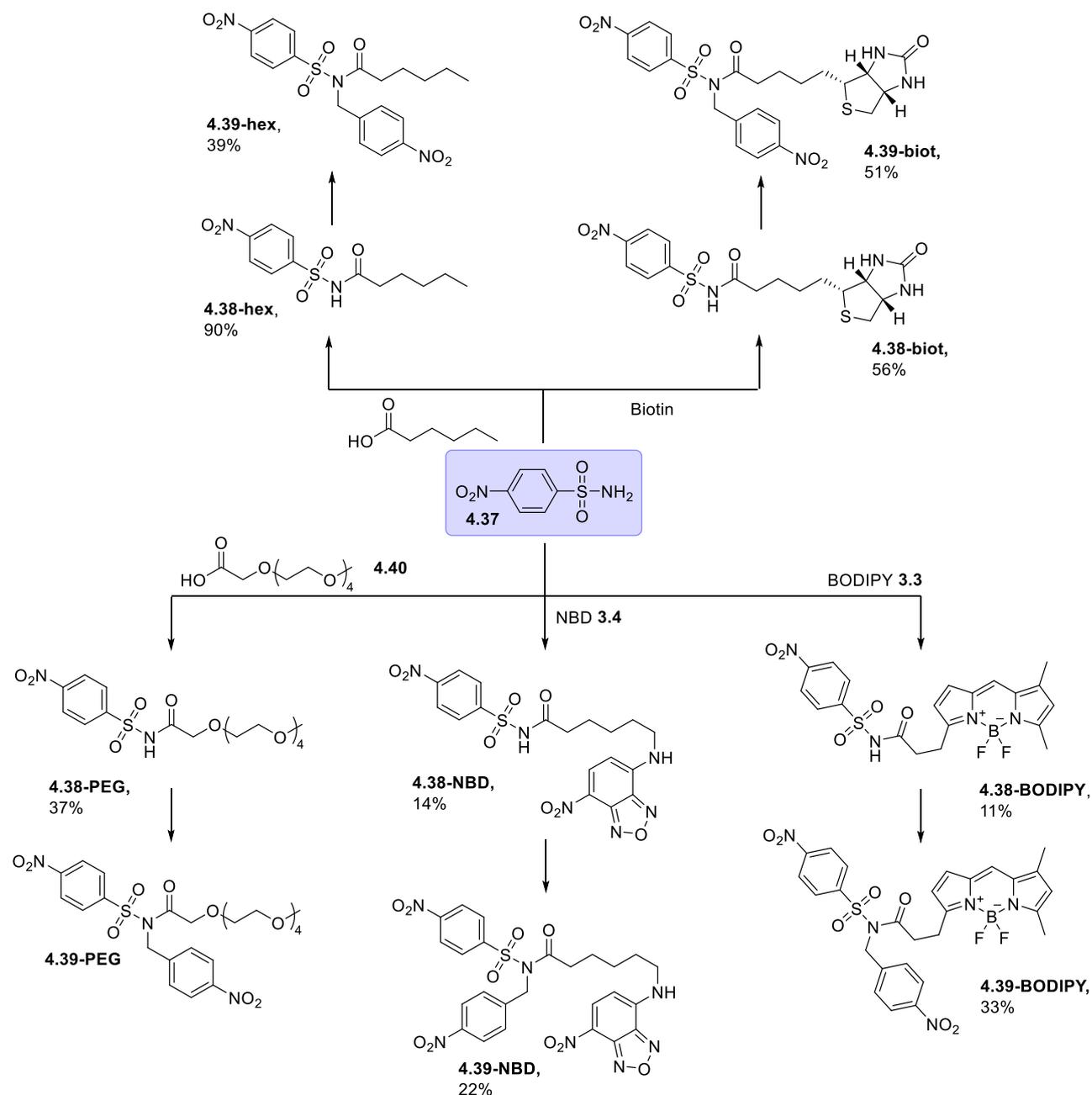


Figure 34: Targeted *p*-nitrobenzyl-NASA 4.39R compounds synthesised.

These three species were obtained in considerably lower overall yield than the first NASAs synthesised, with 3% for 4.39-NBD, and 4% for 4.39-BODIPY. Thus another approach was attempted to increase the yields.

#### 4.4.2 Synthesis of *p*-nitrobenzyl-NASA using new intermediate

While in the process of making the targeted compounds, another type of approach was considered. In the original route, every single *p*-nitrobenzyl-NASA needed will required two synthesis steps and two purification steps. Thus, to avoid this, a new building block from the coupling reaction between 4-nitrobenzenesulfonyl chloride 4.41 and 4-nitrophenyl)methanamine hydrochloride 4.42, compound

**4.43** was synthesised in 76% yield to act as an intermediate, Figure 35. In a second step, a carboxylic acid could be transformed *in-situ* to their acid chloride equivalent and react with **4.43**, providing the NASA reagents **4.39R** in a single step from the common intermediate.

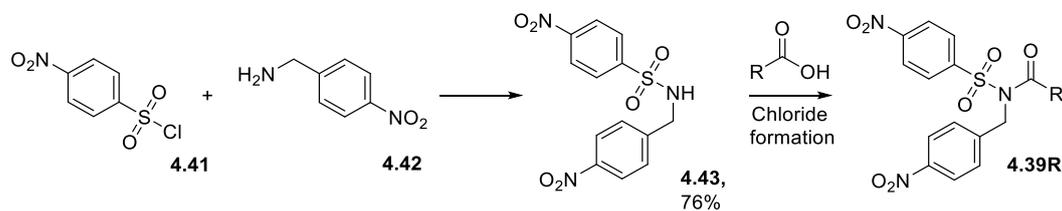


Figure 35: New approach to synthesize NASA compounds.

This strategy was applied for the formation of **4.39-PEG** and **4.39-BODIPY** and their yields were improved to 40% and 63% respectively, considerably increasing the overall yield to 30% and 48% respectively. This strategy had thus not only the advantage of requiring less purification, thanks to the use of a common building block, but also increased the overall yield compared to literature routes for the formation of NASA probes subsequently used in protein labelling.

## 4.5 Summary

In this Chapter, three different groups were targeted for ligand-directed chemistry. Two reagents were synthesized for the one-pot ligand directed approach: LDAI and LDNASA, and for the catalyst-based approach a pyridinium oxime catalyst was synthesised along with the required *p*-nitrobenzyl *N*-acyl-*N*-sulfonamide substrates.

While several synthetic routes were known for building LDAI reagents in a relatively straight forward approach from commercially available materials, we saw that its application to peptide chemistry was more challenging. Indeed, not only were the chemicals synthesised for SPPS not usable for automation synthesis due to their poor solubility, but also the final LDAI compound was shown to undergo hydrolysis during necessary steps in the peptide synthesis process. Thus, this approach was not further explored.

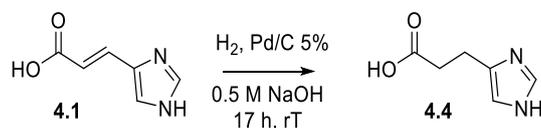
For LDNASA, the chemistry presented in literature needed optimisation. Two approaches were undertaken, first a small molecule approach and second an on-solid support approach based on peptide chemistry. The small molecule approach created a new route to a core intermediate **4.29** which could react with diverse carboxylic acids for further protein modification. For the on-solid support approach, different parameters were studied and by increasing reaction temperatures to 30 °C, the concentration of iodoacetonitrile and DIPEA from 100 mM to 1 M, and the time up to 6 h, full conversion to the expected methylcyano-LDNASA probe was obtained.

Finally for the two probes necessary for *CatLD* chemistry, a similar on-resin approach was used to build pyridinium oxime catalyst peptides in a two-step process using click-chemistry which had the advantage of allowing the formation of a wide range of catalyst probes. For the *p*-nitrobenzyl-NASA substrate, literature conditions were used for the synthesis of most of the probes with the only change being the addition of potassium iodide to accelerate the alkylation. We also designed a more concise route with the creation of a core intermediate **4.43** which could react with any carboxylic acid needed for LDNASA chemistry. This approach has the potential to reduce importantly the time of synthesis and increased the yield of the desired compounds as shown with the two probes synthesised **4.39-PEG** and **4.39-BODIPY**.

As the library of probes was primarily focused on *CatLD* chemistry, the next step would be to use the different optimised route for LDNASA *OpLD* in order to compare the reactivity of these groups which was a goal of the PhD. However, the time constraints principally due to COVID-19 were not compatible with the possibility to do further synthesis and also protein modification. Thus, our focus was placed on the use of the probes detailed in this chapter rather than their further synthesis.

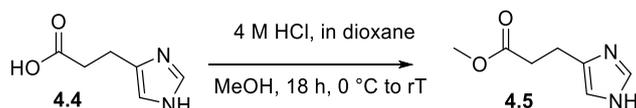
## 4.6 Experimental

## 4.6.1 Chemical synthesis

**Compound 4.4**

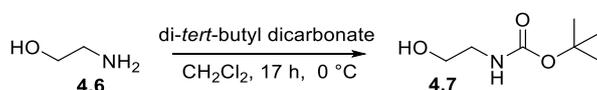
Urocanic acid, **4.1**, (1.56 g, 11.3 mmol) was dissolved in 0.5 M NaOH (60 mL), and Pd/C (5%, 0.40 g, 3.76 mmol) was added. The apparatus was evacuated under vacuum and flushed with argon once and then three times with H<sub>2</sub>. The mixture was stirred under 1 atmosphere of H<sub>2</sub> using a balloon for 15 h at room temperature. The mixture was filtered through Celite and washed with H<sub>2</sub>O (40 mL); the filtrate was adjusted to pH 2 with conc. HCl and concentrated under reduced pressure. Traces of H<sub>2</sub>O were co-evaporated with toluene (2 x 50 mL). The white residue was dissolved in EtOH (20 mL) and filtered; the filtrate was then evaporated under reduced pressure to yield **4.4** as a white solid (1.36 g, 9.71 mmol, 85%). Data were consistent with those previously reported.<sup>8</sup>

<sup>1</sup>H NMR (400 MHz, Methanol-*d*<sub>4</sub>): δ 8.80 (s, 1H, -NCHN), 7.34 (s, 1H, -CHN), 3.01 (dt, *J* = 7.2, 2.6 Hz, 2H, -CH<sub>2</sub>), 2.74 (dt, *J* = 7.2, 2.6 Hz, 2H, -CH<sub>2</sub>). HRMS (ESI<sup>-</sup>): *m/z* calcd for C<sub>6</sub>H<sub>7</sub>N<sub>2</sub>O<sub>2</sub>: 139.0513 [M-H]; found: 139.0515

**Compound 4.5**

**4.4** (550 mg, 3.93 mmol) was dissolved in MeOH (20 mL) and then HCl in dioxane (4 M, 3 mL) was added slowly at 0 °C. The mixture was allowed to warm to room temperature and then stirred for 18 h. The mixture was then concentrated in vacuo to give **4.5** as a yellow oil (430 mg, 2.79 mmol, 71%). Data were consistent with those previously reported.<sup>8</sup>

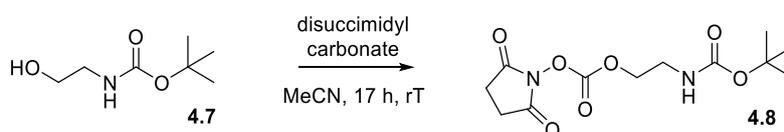
<sup>1</sup>H NMR (400 MHz, Chloroform-*d*): δ 7.54 (d, *J* = 1.1 Hz, 1H, -NCHN), 6.80 (d, *J* = 1.1 Hz, 1H, -CHN), 3.69 (s, 3H, -CH<sub>3</sub>), 2.92 (t, *J* = 7.4, 2H, -CH<sub>2</sub>), 2.68 (t, *J* = 7.4 Hz, 2H, -CH<sub>2</sub>CO<sub>2</sub>Me). HRMS (ESI<sup>+</sup>): *m/z* calcd for C<sub>7</sub>H<sub>11</sub>N<sub>2</sub>O<sub>2</sub>: 155.0815 [M+H]; found: 155.0816

**Compound 4.7**

In dry  $\text{CH}_2\text{Cl}_2$  (50 mL) was added di-*tert*-butyl dicarbonate (3.57 g, 16.4 mmol). Ethanolamine **4.6** (1.00 g, 16.4 mmol) was added dropwise to the solution at 0 °C over 10 min. The reaction was allowed to warm to room temperature and stirred for 2 h. The solvent was removed under reduced pressure and the resulting crude material was purified by flash chromatography using 40% EtOAc in PET. Pure fractions were concentrated under reduced pressure to afford the product **4.7** as a transparent oil (2.04 g, 12.7 mmol, 77%). Data were consistent with those previously reported.<sup>16</sup>

<sup>1</sup>H NMR (400 MHz, Chloroform-*d*):  $\delta$  3.69 (t,  $J$  = 5.0 Hz, 2H,  $-\text{CH}_2\text{OH}$ ), 3.27 (t,  $J$  = 5.0 Hz, 2H,  $-\text{CH}_2\text{NH}$ ), 1.43 (s, 9H,  $-\text{tBu}$ ). HRMS (ESI<sup>+</sup>):  $m/z$  calcd for  $\text{C}_7\text{H}_{15}\text{NNaO}_3$ : 184.0944 [M+Na]; found: 184.0943

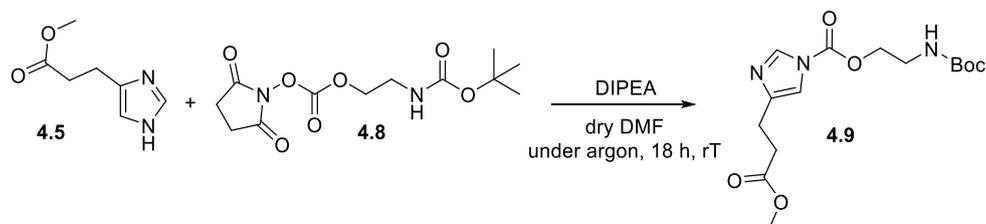
### Compound 4.8



To a solution of **4.7** (0.85 g, 5.27 mmol) in dry MeCN (50 mL), *N,N'*-disuccimidyl carbonate (1.76 g, 6.87 mmol) and  $\text{Et}_3\text{N}$  (1.5 mL, 10.76 mmol) were added and the resulting mixture was stirred at room temperature. After 17 h, the solvent was removed under reduced pressure, the residue was dissolved in EtOAc (50 mL) and the organics washed with saturated solution of  $\text{NaHCO}_3$  (10 mL). The aqueous phase was extracted with EtOAc (2 x 50 mL). The combined organic extracts were dried over anhydrous  $\text{Na}_2\text{SO}_4$ , filtered, and concentrated under reduced pressure to afford the crude product. The resulting crude material was purified by flash chromatography using 20% EtOAc in PET. Pure fractions were concentrated under reduced pressure to afford the product **4.8** as a white solid (0.74 g, 2.45 mmol, 46%). Data were consistent with those previously reported.<sup>17</sup>

<sup>1</sup>H NMR (400 MHz, Chloroform-*d*):  $\delta$  4.95 (app s, 1H,  $-\text{NH}$ ), 4.37 (t,  $J$  = 5.2 Hz, 2H,  $-\text{CH}_2\text{O}$ ), 3.48 (dt,  $J_1$  =  $J_2$  = 5.2 Hz, 2H,  $-\text{CH}_2\text{N}$ ), 2.84 (s, 4H,  $-\text{OSu}$ ), 1.44 (s, 9H,  $-\text{tBu}$ ). HRMS (ESI<sup>+</sup>):  $m/z$  calcd for  $\text{C}_{12}\text{H}_{18}\text{N}_2\text{NaO}_7$ : 325.1006 [M+Na]; found: 325.1006

### Compound 4.9

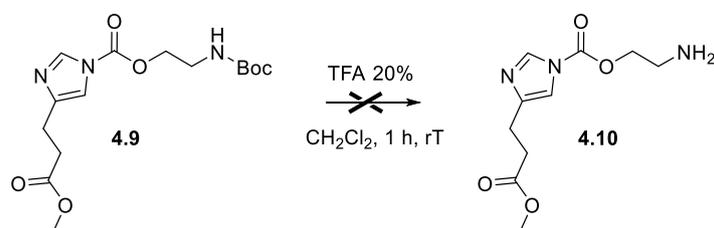


To a stirred solution of **4.5** (122 mg, 0.79 mmol) in dry *N,N*-dimethylformamide (DMF) (5 mL) was added **4.8** (360 mg, 1.19 mmol), and *N,N*-diisopropylethylamine (0.35 mL, 2.01 mmol) under argon atmosphere. The mixture was stirred for 18 h at room temperature. After reducing the amount of

solvent under reduced pressure to the minimum (< 500  $\mu$ L), H<sub>2</sub>O (100 mL) was added to the solution and was then extracted with EtOAc (3 x 100 mL). The combined organics were washed with brine (50 mL), dried over anhydrous MgSO<sub>4</sub>, and concentrated under reduced pressure. The resulting crude material was purified by flash chromatography using 50% EtOAc in PET. Pure fractions were concentrated under reduced pressure to afford the product **4.9** as a white solid (51 mg, 0.15 mmol, 19%).

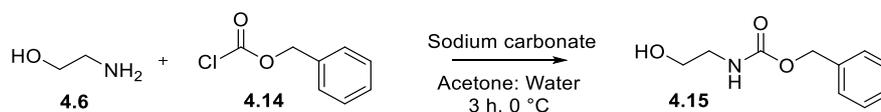
**<sup>1</sup>H NMR** (400 MHz, Chloroform-*d*):  $\delta$  8.03 (d,  $J$  = 1.2 Hz, 1H, -NCHN), 7.14 (d,  $J$  = 1.2 Hz, 1H, -CHN), 5.05 (t,  $J$  = 6.2 Hz, 1H, -NH<sub>2</sub>Boc), 4.40 (t,  $J$  = 5.2 Hz, 2H, -NC(O)OCH<sub>2</sub>), 3.64 (s, 3H, -CH<sub>3</sub>), 3.49 (dt,  $J_1 = J_2 = 5.2$  Hz, 2H, -CH<sub>2</sub>NHBoc), 2.84 (t,  $J$  = 7.3 Hz, 2H, -C(O)CH<sub>2</sub>CH<sub>2</sub>), 2.65 (t,  $J$  = 7.3 Hz, 2H, -C(O)CH<sub>2</sub>CH<sub>2</sub>), 1.39 (s, 9H, -<sup>t</sup>Bu). **HRMS** (ESI<sup>+</sup>):  $m/z$  calcd for C<sub>15</sub>H<sub>23</sub>N<sub>3</sub>NaO<sub>6</sub>: 364.1479 [M+Na]; found: 364.1480.

#### Compound 4.10



To a stirred solution of **4.9** (22 mg, 0.06 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (4 mL) was added TFA (1 mL). The mixture was stirred for 1 h at room temperature. After reducing the amount of solvent to 0.5 mL under reduced pressure, and co-evaporation of solvent with toluene (3 x 5 mL), the crude mixture was obtained as a transparent oil. It was found that instead of forming **4.10**, cleavage had occurred to form **4.5**.

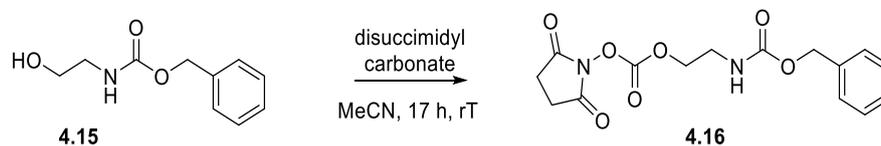
#### Compound 4.15



2-Aminoethanol (1.0 g, 16.5 mmol) and sodium carbonate (4.3 g, 40.2 mmol) were dissolved in acetone-H<sub>2</sub>O (1 : 1 v/v) (50 mL) and benzyl chloroformate (3.5 mL, 24.6 mmol) was added dropwise to the stirred solution at 0 °C over 10 minutes. The mixture was stirred at 0 °C for a further 3 h. The solids were collected by filtration and washed with acetone (2 x 5 mL) and the filtrate was evaporated to dryness. The resulting gum was taken up in chloroform (5 mL) and filtered. The filtrate was concentrated under reduced pressure and crystallized on standing to give **4.15** as a white solid (3.2 g, 16.5 mmol, qtt). Data were consistent with those previously reported.<sup>18</sup>

$^1\text{H NMR}$  (400 MHz, Chloroform-*d*):  $\delta$  7.39-7.27 (m, 5H, -ArH), 5.31 (br s, 1H, -NH), 5.09 (s, 2H, -OCH<sub>2</sub>Ar), 3.68 (t,  $J = 5.1$  Hz, 2H, -CH<sub>2</sub>OH), 3.33 (td,  $J_1 = J_2 = 5.1$  Hz, 2H, -NHCH<sub>2</sub>). **HRMS** (ESI<sup>+</sup>):  $m/z$  calcd for C<sub>10</sub>H<sub>13</sub>NNaO<sub>3</sub>: 218.0788 [M+Na]; found: 218.0782.

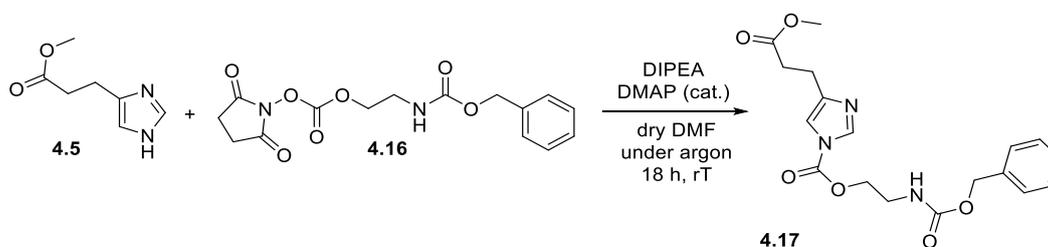
#### Compound 4.16



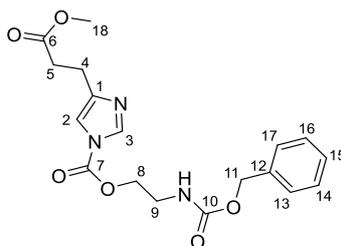
To a solution of **4.15** (1.00 g, 5.12 mmol) in dry MeCN (50 mL), *N,N'*-disuccimidyl carbonate (2.00 g, 7.81 mmol) and Et<sub>3</sub>N (3 mL, 21.52 mmol) were added, and the resulting mixture was stirred at room temperature overnight under a N<sub>2</sub> atmosphere. After removal of solvent under reduced pressure, the crude product was dissolved in H<sub>2</sub>O (50 mL) and then extracted three times by EtOAc (3 x 50 mL). The combined organic phases were then dried over MgSO<sub>4</sub> and concentrated under reduced pressure to afford product **4.16** as an orange viscous oil (1.5 g, 4.46 mmol, 87%). Data were consistent with those previously reported.<sup>18</sup>

$^1\text{H NMR}$  (400 MHz, Chloroform-*d*):  $\delta$  7.42-7.29 (m, 5H, -ArH), 5.22 (s, 1H, -NH), 5.12 (s, 2H, -OCH<sub>2</sub>Ar), 4.40 (t,  $J = 5.0$  Hz, 2H, -OCH<sub>2</sub>), 3.56 (td,  $J_1 = J_2 = 5.0$  Hz, 2H, -NHCH<sub>2</sub>), 2.83 (s, 4H, -OSu). **HRMS** (ESI<sup>+</sup>):  $m/z$  calcd for C<sub>15</sub>H<sub>16</sub>N<sub>2</sub>NaO<sub>7</sub>: 359.0850 [M+Na]; found: 359.0843.

#### Compound 4.17

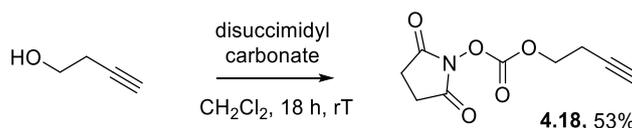


To a stirred solution of **4.5** (161 mg, 1.1 mmol) in dry *N,N*-dimethylformamide (DMF) (5 mL) was added **4.16** (483 mg, 1.4 mmol), and *N,N*-diisopropylethylamine (DIPEA) (0.5 mL, 2.9 mmol), and DMAP (31 mg, 0.25 mmol) under argon atmosphere. The mixture was stirred for 18 h at room temperature. After reducing the amount of solvent under reduced pressure to the minimum, H<sub>2</sub>O (50 mL) was added to the solution and was then extracted with EtOAc (3 x 50 mL). The combined organics were washed with brine (20 mL), dried over anhydrous MgSO<sub>4</sub>, and concentrated under reduced pressure. The resulting crude material was purified by flash chromatography using a gradient of 25-50% EtOAc in PET. Pure fractions were concentrated under reduced pressure to afford the product **4.17** as a yellow oil (230 mg, 0.67 mmol, 64%).



**<sup>1</sup>H NMR** (400 MHz, Chloroform-*d*):  $\delta$  8.05 (d,  $J = 1.3$  Hz, 1H, -NCHN), 7.34 (m, 5H, -ArH), 7.15 (d,  $J = 1.3$  Hz, 1H, -CHN), 5.11 (s, 2H, -OCH<sub>2</sub>Ar), 4.46 (t,  $J = 5.3$  Hz, 2H, -NC(O)OCH<sub>2</sub>CH<sub>2</sub>), 3.67 (s, 3H, -CH<sub>3</sub>), 3.60 (td,  $J_1 = J_2 = 5.3$  Hz, 2H, -NC(O)OCH<sub>2</sub>CH<sub>2</sub>), 2.88 (t,  $J = 7.4$  Hz, 2H, -C(O)CH<sub>2</sub>CH<sub>2</sub>), 2.68 (t,  $J = 7.4$  Hz, 2H, -C(O)CH<sub>2</sub>CH<sub>2</sub>). **<sup>13</sup>C NMR** (101 MHz, Chloroform-*d*)  $\delta$  173.36 (C<sub>6</sub>), 162.7 (C<sub>1</sub>), 148.60 (C<sub>10</sub>), 142.89 (C<sub>2</sub> and C<sub>3</sub>), 136.7 (C<sub>2</sub>, C<sub>3</sub>), 128.68 (C<sub>15</sub>), 128.41 (C benzyl), 128.29 (C benzyl), 67.52 (C<sub>11</sub>), 67.04 (C<sub>8</sub>), 51.80 (C<sub>18</sub>), 40.54 (C<sub>9</sub>), 39.99, 33.01 (C<sub>5</sub>), 23.40 (C<sub>4</sub>). **HRMS** (ESI<sup>+</sup>):  $m/z$  calcd for C<sub>18</sub>H<sub>21</sub>N<sub>3</sub>NaO<sub>6</sub>: 398.1323 [M+Na]; found: 398.1314.

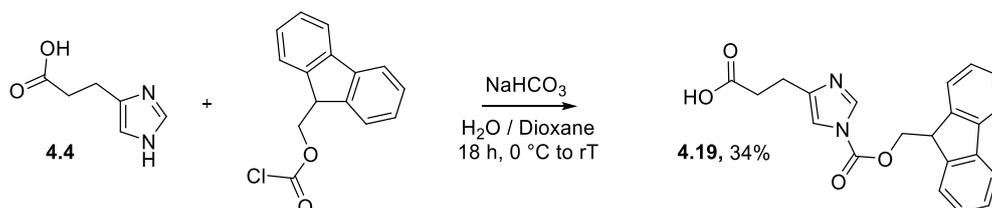
#### Compound 4.18



To a solution of 3-butyn-1-ol (1.00 mL, 13.3 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (100 mL), *N,N'*-disuccimidyl carbonate (4.50 g, 17.6 mmol) and Et<sub>3</sub>N (5 mL, 35.9 mmol) were added and the resulting mixture was stirred at room temperature for 18 h. After removal of solvent under reduced pressure, the crude product was dissolved in H<sub>2</sub>O (50 mL) and then extracted by EtOAc (3 x 50 mL). The organic layers were combined, dried over anhydrous MgSO<sub>4</sub> and concentrated under reduced pressure to afford the crude product **4.18** as an orange viscous oil (1.5 g, 7.1 mmol, 53%). Data were consistent with those previously reported.<sup>19</sup>

**<sup>1</sup>H NMR** (400 MHz, Chloroform-*d*):  $\delta$  4.40 (t,  $J = 6.9$  Hz, 2H, -OCH<sub>2</sub>), 2.83 (s, 4H, -Osu), 2.65 (td,  $J = 6.9$ , 2.7 Hz, 2H, -CH<sub>2</sub>), 2.05 (t,  $J = 2.7$  Hz, 1H, CCH). **HRMS** (ESI<sup>+</sup>):  $m/z$  calcd for C<sub>9</sub>H<sub>9</sub>NnaO<sub>5</sub>: 234.0373 [M+Na]; found: 234.0363.

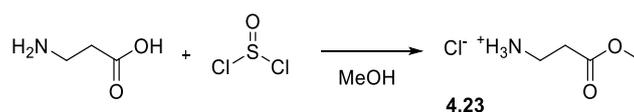
#### Compound 4.19



A mixture of **4.4** (1.5 g, 11.1 mmol) and NaHCO<sub>3</sub> (1.5 g, 17.9 mmol) in H<sub>2</sub>O (50 mL) was added to a solution of Fmoc-Cl (4.5 g, 17.4 mmol) in dioxane (100 mL) at 0 °C and the solution was stirred for 1 h. Then the mixture was stirred at room temperature for 17 h. After adding conc. HCl to reach pH 2, the aqueous was extracted with EtOAc (3 x 100 mL). The combined organic layers were then washed with brine (100 mL) and dried over MgSO<sub>4</sub>, and concentrated under reduced pressure to afford the crude product. The resulting crude material was purified by flash chromatography using 40% EtOAc in PET and then 100% EtOAc. Pure fractions were concentrated under reduced pressure to afford the product **4.19** as white sticky solid (1.36 g, 3.76 mmol, 34%).

**<sup>1</sup>H NMR** (400 MHz, Chloroform-*d*): δ 8.05 (d, *J* = 1.3 Hz, 1H, -NCHN), 7.79 (d, *J* = 7.5 Hz, 2H, -FmocArH), 7.55 (dd, *J* = 7.5, 1.3 Hz, 2H, -FmocArH), 7.43 (dd, *J*<sub>1</sub> = *J*<sub>2</sub> = 7.5 Hz, 2H, -FmocArH), 7.33 (ddd, *J*<sub>1</sub> = *J*<sub>2</sub> = 7.5, *J*<sub>3</sub> = 1.3 Hz, 2H, -FmocArH), 7.16 (d, *J* = 1.3 Hz, 1H, -CHN), 4.74 (d, *J* = 6.4 Hz, 2H, -OCH<sub>2</sub>), 4.35 (t, *J* = 6.4 Hz, 1H, -OCH<sub>2</sub>CH), 2.91 (t, *J* = 6.7 Hz, 2H, -COCH<sub>2</sub>CH<sub>2</sub>), 2.75 (t, *J* = 6.7 Hz, 2H, -COCH<sub>2</sub>CH<sub>2</sub>). **HRMS** (ESI<sup>+</sup>): *m/z* calcd for C<sub>21</sub>H<sub>18</sub>N<sub>2</sub>NaO<sub>4</sub>: 385.1159 [M+Na]; found: 385.1167.

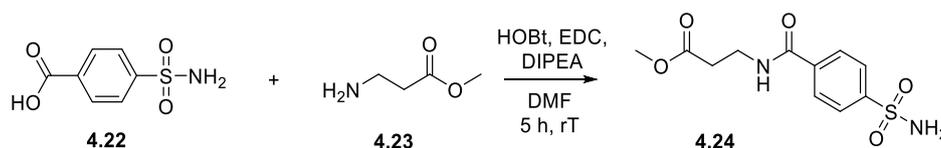
#### Compound 4.23



β-Alanine (2.03 g, 22.8 mmol) was dissolved in MeOH (30 mL) and cooled to 0 °C. Thionyl chloride (2.8 mL, 38.6 mmol) was then slowly added to the solution and the conversion was monitored by TLC which showed full conversion after 2.5 h. The solvent was removed under reduced pressure to give the hydrochloride salt of the β-alanine methyl ester **4.23** as a white solid (3.2 g, 22.9 mmol, qtt). Data were consistent with those previously reported.<sup>20</sup>

**<sup>1</sup>H NMR** (400 MHz, DMSO-*d*<sub>6</sub>): δ 8.15 (br s, 3H, -NH<sub>3</sub><sup>+</sup>), 3.63 (s, 3H, -CH<sub>3</sub>), 3.02-2.96 (m, 2H, -CH<sub>2</sub>NH<sub>2</sub>), 2.73-2.69 (m, 2H, -CH<sub>2</sub>CH<sub>2</sub>NH<sub>2</sub>). **HRMS** (ESI<sup>+</sup>): *m/z* calcd for C<sub>4</sub>H<sub>10</sub>NO<sub>2</sub>: 104.0706 [M+H]; found: 104.0706

#### Compound 4.24

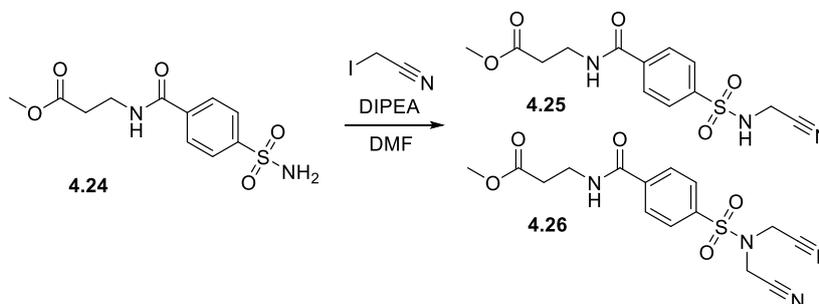


To a stirred solution of 4-sulfamoylbenzoic acid **4.22** (1.7 g, 8.5 mmol) in *N,N*-dimethylformamide (DMF) (25 mL) was added β-alanine acid methyl ester **4.23** (1.2 g, 11.7 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) (3.4 g, 17.7 mmol), 1-hydroxybenzotriazole

(HOBt) (2.3 g, 17.0 mmol) and *N,N*-diisopropylethylamine (DIPEA) (5 mL, 28.7 mmol). The mixture was stirred for 5 h at room temperature. After reducing the amount of solvent under reduced pressure to the minimum, H<sub>2</sub>O (100 mL) was added to the solution and was then extracted with EtOAc (3 x 100 mL). The combined organics were washed with brine (2 x 50 mL), dried over anhydrous MgSO<sub>4</sub>, and concentrated under reduced pressure. The resulting crude material was purified by flash chromatography using 100% EtOAc. Pure fractions were concentrated under reduced pressure to afford the product **4.24** as a white solid (1.20 g, 4.19 mmol, 50%).

<sup>1</sup>H NMR (400 MHz, MeOD-*d*<sub>4</sub>): δ 7.99-7.91 (m, 4H, -ArH), 3.69 (s, 3H, -COCH<sub>3</sub>), 3.65 (t, *J* = 6.8 Hz, 2H, -CH<sub>2</sub>NH), 2.68 (t, *J* = 6.8 Hz, 2H, -CH<sub>2</sub>CO). <sup>13</sup>C NMR (101 MHz, MeOD-*d*<sub>4</sub>) δ 172.51 (-COO), 167.56 (-CON), 146.41 (-CCON), 137.54 (-CSO<sub>2</sub>), 127.65 (-CHAR), 126.00 (-CHAR), 50.93 (-CH<sub>3</sub>), 35.79 (-CH<sub>2</sub>NH), 33.20 (-CH<sub>2</sub>CO). HRMS (ESI<sup>+</sup>): *m/z* calcd for C<sub>11</sub>H<sub>15</sub>N<sub>2</sub>O<sub>5</sub>S: 287.0696 [M+H]; found: 287.0706

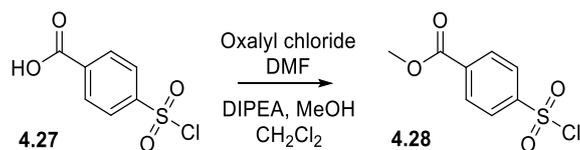
### Compounds 4.25 and 4.26



In 5 mL DMF, was added **4.24** (100 mg, 0.35 mol), DIPEA (0.6 mL, 3.45 mol) and finally iodoacetonitrile (0.13 mL, 1.75 mol).

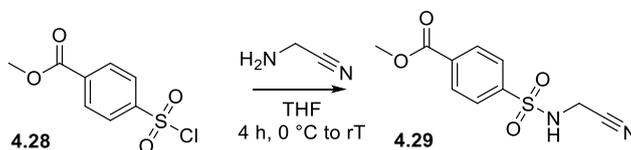
First attempt: The mixture was stirred for 18 h at room temperature. After removing the solvent under reduced pressure, the resulting crude material was purified by flash chromatography using 60 then 80% EtOAc in PET and finally 100% EtOAc. Pure fractions were concentrated under reduced pressure to afford product **4.25** (10 mg, 31 mmol, 9%). <sup>1</sup>H NMR (400 MHz, Chloroform-*d*): δ 8.06-7.77 (m, 4H, -CHAR), 7.03 (s, 1H, -NH), 3.73 (m, 5H, -OCH<sub>3</sub> and -CH<sub>2</sub>CO), 2.67 (t, *J* = 5.7 Hz, 2H, -CH<sub>2</sub>NH), 2.16 (app s, 2H, -CH<sub>2</sub>CN). HRMS (ESI<sup>-</sup>): *m/z* calcd for C<sub>13</sub>H<sub>14</sub>N<sub>3</sub>O<sub>5</sub>S: 324.0660 [M-H]; found: 324.0647

Second attempt: The mixture was stirred for 18 h at 60 °C. After removing the solvent under reduced pressure, the resulting crude material was purified by flash chromatography using 60 then 80% EtOAc in PET and finally 100% EtOAc. Pure fractions were concentrated under reduced pressure to afford product **4.26**. HRMS (ESI<sup>+</sup>): *m/z* calcd for C<sub>15</sub>H<sub>16</sub>N<sub>4</sub>O<sub>5</sub>S: 365.0873 [M+H]; found: 365.0888. No NMR analysis.

**Compounds 4.28**

To a suspension of 4-chlorosulfonylbenzoic acid **4.27** (1.1 g, 5.0 mmol) in DMF (0.05 mL) and dry CH<sub>2</sub>Cl<sub>2</sub> (15 mL) was added dropwise oxalyl chloride (1.6 mL, 8.4 mmol), and the reaction was stirred at room temperature for 2 h. The reaction mixture was then concentrated under reduced pressure at 40 °C, added dropwise to an ice cold mixture of MeOH (1 mL) and DIPEA (1.3 mL, 13.6 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (10 mL), then stirred at room temperature for 2 h. Solvents were removed under reduced pressure, and the reaction mixture was dissolved in EtOAc (50 mL) and the organics washed with a saturated aqueous solution of ammonium chloride (50 mL) and brine (2 x 50 mL). The organic layer was dried over anhydrous sodium sulfate, filtered, and concentrated under reduced pressure to afford product **4.28** as a brown solid (1.0 g, 4.3 mmol, 85%). Data were consistent with those previously reported.<sup>21</sup>

<sup>1</sup>H NMR (400 MHz, Chloroform-*d*): δ 8.27 (d, *J* = 8.9 Hz, 2H, -CH<sub>2</sub>CO<sub>2</sub>), 8.12 (d, *J* = 8.9 Hz, 2H, -CH<sub>2</sub>SO<sub>2</sub>Cl), 3.99 (s, 3H, -CH<sub>3</sub>). HRMS (ESI<sup>-</sup>): *m/z* calcd for C<sub>8</sub>H<sub>7</sub>O<sub>5</sub>S (seen as the hydroxyl): 215.0020 [M-H]; found: 215.0016

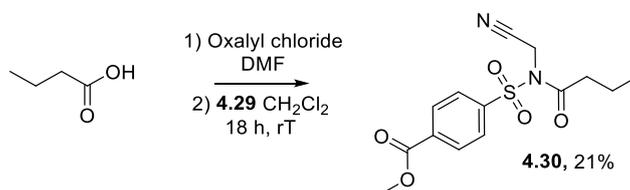
**Compounds 4.29**

**4.28** (123 mg, 0.48 mmol) was dissolved in THF (10 mL) with DIPEA (0.25 mL, 1.44 mmol) and stirred at 0 °C. Aminoacetonitrile hydrochloride (67 mg, 0.72 mmol) was then added. After 1.5 h, the reaction mixture was warmed to room temperature and stirred for another 5 hours. The reaction was concentrated under reduced pressure, and the residue dissolved in EtOAc (20 mL). The organics were washed with H<sub>2</sub>O (2 x 10 mL) and with brine (5 mL), dried over anhydrous Na<sub>2</sub>SO<sub>4</sub> and concentrated under reduced pressure to afford the crude product. The resulting crude material was purified by flash chromatography using 25% EtOAc in PET. Pure fractions were concentrated under reduced pressure to afford the product **4.29** as a white solid (45 mg, 0.19 mmol, 39%).

<sup>1</sup>H NMR (400 MHz, Chloroform-*d*): δ 8.26-8.15 (m, 2H, -CH<sub>2</sub>CO<sub>2</sub>), 8.01-7.93 (m, 2H, -CH<sub>2</sub>SO<sub>2</sub>), 5.22 (t, *J* = 6.5 Hz, 1H, -NH), 4.07 (d, *J* = 6.5 Hz, 2H, -CH<sub>2</sub>NH), 3.96 (s, 3H, -CH<sub>3</sub>). <sup>13</sup>C NMR (101 MHz, Chloroform-*d*): δ 165.48 (-C=O), 142.64 (-CH<sub>2</sub>CO<sub>2</sub>), 134.90 (-CH<sub>2</sub>SO<sub>2</sub>), 130.81 (-CH<sub>2</sub>CO<sub>2</sub>), 127.41 (-

CHCSO<sub>2</sub>), 114.43 (-C<sub>2</sub>NCN), 52.87 (-CH<sub>3</sub>), 31.12 (-CH<sub>2</sub>NH). HRMS (ESI<sup>-</sup>): *m/z* calcd for C<sub>10</sub>H<sub>9</sub>N<sub>2</sub>O<sub>4</sub>S: 253.0289 [M-H]; found: 253.0289

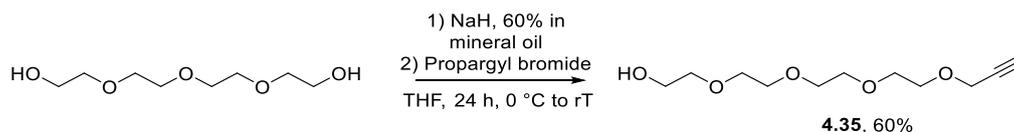
### Compounds 4.30



Butyric acid (10 mg, 109 μmol) was dissolved in CH<sub>2</sub>Cl<sub>2</sub> (4 mL) and two drops of DMF was added followed by oxalyl chloride (25 μL, 295 μmol). After 15 min, the solvent was reduced under reduced pressure. The residue was dissolved in CH<sub>2</sub>Cl<sub>2</sub> (1 mL) which was slowly added to a solution of **4.29** (18 mg, 71 μmol) and DIPEA (50 μL, 287 μmol) in CH<sub>2</sub>Cl<sub>2</sub> (10 mL). After 2 h, the solution was washed with brine (3 x 20 mL) then dried over anhydrous MgSO<sub>4</sub> and concentrated under reduced pressure to afford the crude product. The resulting crude material was purified by flash chromatography using 20% EtOAc in PET. Pure fractions were concentrated under reduced pressure to afford the product **4.30** as a white solid (6.0 mg, 15 μmol, 21%).

<sup>1</sup>H NMR (300 MHz, Chloroform-*d*): δ 8.25-8.16 (m, 2H, -CH<sub>2</sub> arom), 7.99 (d, *J* = 8.6 Hz, 2H, -CH<sub>2</sub> arom), 4.72 (s, 2H, -NCH<sub>2</sub>CN), 3.91 (s, 3H, -OCH<sub>3</sub>), 2.56 (t, *J* = 7.2 Hz, 2H, -COCH<sub>2</sub>), 1.67-1.41 (m, 2H, -CH<sub>2</sub>CH<sub>2</sub>), 0.89-0.72 (m, 3H, -CH<sub>2</sub>CH<sub>3</sub>). HRMS (ESI<sup>+</sup>): *m/z* calcd for C<sub>14</sub>H<sub>11</sub>N<sub>2</sub>NaO<sub>5</sub>S: 347.0672 [M+Na]; found: 347.0672

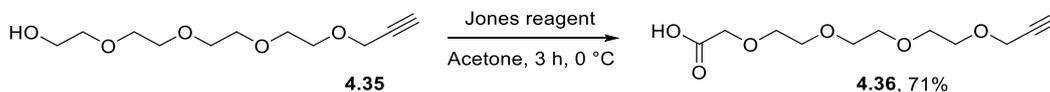
### Compound 4.35



NaH, 60% in mineral oil (500 mg, 10.3 mmol) was added to a solution of tetraethylene glycol (3.5 g, 5.15 mmol) in THF (10 mL) at 0 °C for 10 min. Then, a solution of propargyl bromide (1.6 mL, 7.2 mmol) was added dropwise and the resulting mixture was stirred for 24 h warming up to room temperature. After the solvent was concentrated under reduced pressure, the crude material was dissolved in CH<sub>2</sub>Cl<sub>2</sub> (100 mL) and washed with aqueous HCl (1 M, 50 mL). The organic layer was collected and dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated under reduced pressure to afford the crude product. The resulting crude material was purified by flash chromatography using 100% EtOAc. Pure fractions were concentrated under reduced pressure to afford the product **4.35** as an oil (715 mg, 3.09 mmol, 60%). Data were consistent with those previously reported.<sup>22</sup>

**<sup>1</sup>H NMR** (400 MHz, Chloroform-*d*):  $\delta$  4.19 (d,  $J$  = 2.3 Hz, 2H, -CHCCH), 3.74-3.63 (m, 14H, PEG), 3.61-3.57 (m, 2H, PEG), 2.76 (t,  $J$  = 6.2 Hz, 1H, -OH), 2.43-2.40 (t,  $J$  = 2.3 Hz, 1H, -CHCCH). **HRMS** (ESI<sup>-</sup>):  $m/z$  calcd for C<sub>11</sub>H<sub>20</sub>NaO<sub>5</sub>: 255.1203 [M+Na]; found: 255.1194

### Compound 4.36



Jones reagent (20 ml) was added to a solution of **4.35** (600 mg, 2.59 mmol) dissolved in acetone (20 mL) at 0 °C. After the reaction was stirred for 3 h, 2-propanol was added until the colour became green. Then, the reaction solution was filtrated using Celite and the filtrate washed by CH<sub>2</sub>Cl<sub>2</sub> (2 x 10 mL). The solvents were concentrated under reduced pressure to afford the product **4.36** as an oil (450 mg, 1.83 mmol, 71%). Data were consistent with those previously reported.<sup>22</sup>

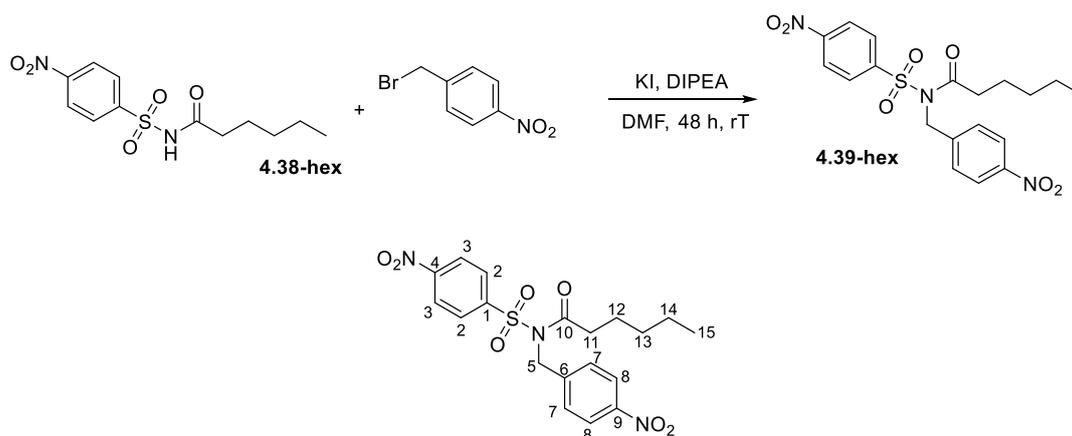
**<sup>1</sup>H NMR** (400 MHz, Chloroform-*d*):  $\delta$  9.78 (br s, 1H, -COOH), 4.32-4.11 (m, 6H, -CHCCH and -CH<sub>2</sub>COOH and PEG), 3.81-3.60 (m, 10H, PEG), 2.43 (m, 1H, -CHCCH). **HRMS** (ESI<sup>-</sup>):  $m/z$  calcd for C<sub>11</sub>H<sub>18</sub>NaO<sub>6</sub>: 269.0996 [M+Na]; found: 269.0994

### Compound 4.38-hex



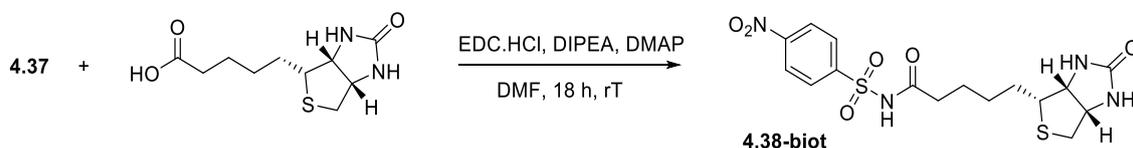
To a solution of 4-nitrobenzenesulfonamide **4.37** (101 mg, 0.50 mmol) in DMF (10 mL) was added hexanoic acid (95  $\mu$ L, 0.75 mmol), EDC (300 mg, 1.5 mmol), 4-dimethylaminopyridine (30 mg, 0.25 mmol) and DIPEA (260  $\mu$ L, 1.5 mmol). The mixture was allowed to stir at room temperature for 40 h. After the solvent was removed under reduced pressure, the resulting crude material was purified by flash chromatography using 5-30% EtOAc in PET. Pure fractions were concentrated under reduced pressure to afford the product **4.38-hex** as a white solid (140 mg, 0.45 mmol, 90%). Data were consistent with those previously reported.<sup>23</sup>

**<sup>1</sup>H NMR** (400 MHz, Chloroform-*d*):  $\delta$  8.35 (d,  $J$  = 9.0 Hz, 2H, -CH(*o*-nitro)), 8.25 (d,  $J$  = 9.0 Hz, 2H, -CH(*m*-nitro)), 2.27 (t,  $J$  = 7.4 Hz, 2H, -COCH<sub>2</sub>), 1.59-1.50 (m, 3H, -COCH<sub>2</sub>CH<sub>2</sub>, -CH<sub>2</sub>CH<sub>3</sub>), 0.94-0.85 (m, 3H, -COCH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>, -CH<sub>2</sub>CH<sub>3</sub>), 0.82 (t,  $J$  = 6.9 Hz, 3H, -CH<sub>2</sub>CH<sub>3</sub>). **HRMS** (ESI<sup>-</sup>):  $m/z$  calcd for C<sub>12</sub>H<sub>15</sub>N<sub>2</sub>O<sub>5</sub>S: 299.0707 [M-H]; found: 299.0707

**Compound 4.39-hex**

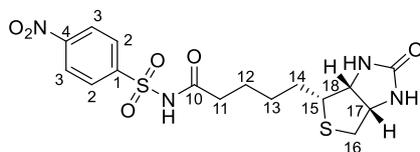
To a solution of *N*-((4-nitrophenyl)sulfonyl)hexanamide **4.38-hex** (70 mg, 0.23 mmol) in DMF (2 mL) was added 4-nitrobenzyl bromide (151 mg, 0.7 mmol), potassium iodide (115 mg, 0.7 mmol) and DIPEA (122  $\mu$ L, 0.7 mmol). The mixture was allowed to stir at room temperature for 48 h. After the solvent was removed under reduced pressure, the resulting crude material was purified by flash chromatography using 10-100% EtOAc in PET. Pure fractions were concentrated under reduced pressure to afford the product **4.39-hex** as a pale pink solid (39 mg, 89  $\mu$ mol, 39%).

<sup>1</sup>H NMR (500 MHz, Chloroform-*d*):  $\delta$  8.38 (d,  $J$  = 8.9 Hz, 2H, -H<sub>3</sub>), 8.23 (d,  $J$  = 8.7 Hz, 2H, -H<sub>8</sub>), 8.06 (d,  $J$  = 8.9 Hz, 2H, -H<sub>2</sub>), 7.55 (d,  $J$  = 8.7 Hz, 2H, -H<sub>7</sub>), 5.16 (s, 2H, -H<sub>5</sub>), 2.48 (t,  $J$  = 7.4 Hz, 2H, -H<sub>11</sub>), 1.49 (tt,  $J_1 = J_2 = 7.4$  Hz, 2H, -H<sub>12</sub>), 1.23-1.06 (m, 4H, -H<sub>13</sub> and H<sub>14</sub>), 0.84 (t,  $J$  = 7.4 Hz, 3H, -H<sub>15</sub>). <sup>13</sup>C NMR (100 MHz, Chloroform-*d*):  $\delta$  172.72 (C<sub>10</sub>), 150.70 (C<sub>1</sub>), 147.65 (C<sub>6</sub>), 144.66 (C<sub>4</sub>), 143.34 (C<sub>9</sub>), 129.25 (C<sub>2</sub> and 2'), 128.11 (C<sub>7</sub> and 7'), 124.42 (C<sub>3</sub> and 3'), 124.06 (C<sub>8</sub> and 8'), 49.26 (C<sub>5</sub>), 36.19 (C<sub>11</sub>), 30.86 (C<sub>13</sub>), 23.91 (C<sub>12</sub>), 22.18 (C<sub>14</sub>), 13.67 (C<sub>15</sub>). HRMS (ESI<sup>+</sup>):  $m/z$  calcd for C<sub>19</sub>H<sub>21</sub>N<sub>3</sub>O<sub>7</sub>S: 434.1027 [M-H]; found: 434.1019

**Compound 4.38-biot**

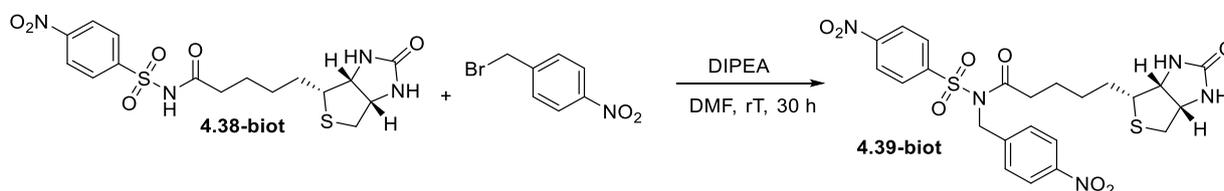
To a solution of 4-nitrobenzenesulfonamide **4.37** (376 mg, 1.86 mmol) in DMF (12 mL) was added biotin (502 mg, 2.05 mmol), EDC (720 mg, 3.75 mmol), 4-dimethylaminopyridine (DMAP) (458 mg, 3.75 mmol) and DIPEA (1.0 mL, 5.58 mmol). The mixture was allowed to stir at room temperature for 18 h. After the solvent was removed under reduced pressure, the resulting crude material was purified by flash chromatography using 0.25% AcOH, 2.5% MeOH in CH<sub>2</sub>Cl<sub>2</sub>. Pure fractions were concentrated

under reduced pressure to afford the product **4.38-biotin** as a white solid (540 mg, 1.05 mmol, 68%). Data were consistent with those previously reported.<sup>11</sup>

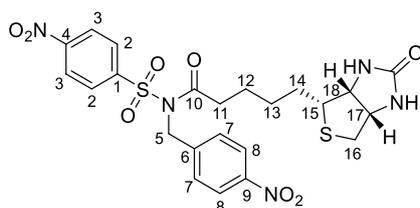


<sup>1</sup>H NMR (400 MHz, DMSO-*d*<sub>6</sub>): δ 8.39 (d, *J* = 7.2 Hz, 2H, -H<sub>3</sub>), 8.11 (d, *J* = 7.2 Hz, 2H, -H<sub>2</sub>), 6.34 (s, 1H, -NH biotin), 6.32 (s, 1H, -NH biotin), 4.27-4.23 (m, 1H, -H<sub>17</sub> or 18), 4.06-4.03 (m, 1H, -H<sub>17</sub> or 18), 3.01-2.96 (m, 1H, -H<sub>15</sub>), 2.76 (dd, *J* = 12.0, 5.2 Hz, 1H, -H<sub>16</sub>), 2.52 (d, *J* = 12.0 Hz, 1H, -H<sub>16</sub>), 2.18 (t, *J* = 7.4 Hz, 2H, -H<sub>11</sub>), 1.52-1.32 (m, 4H, -H<sub>12</sub> and -H<sub>14</sub>), 1.20-1.14 (m, 2H, -H<sub>13</sub>). HRMS (ESI<sup>-</sup>): *m/z* calcd for C<sub>16</sub>H<sub>20</sub>N<sub>4</sub>NaO<sub>6</sub>S<sub>2</sub>: 451.0716 [M-H]; found: 451.0723

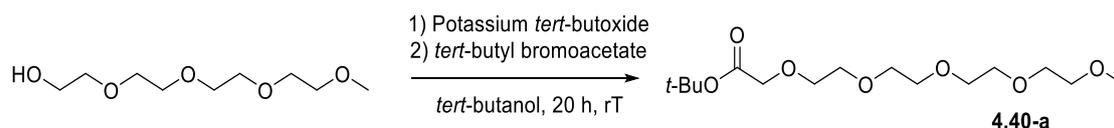
### Compound 4.39-biotin



To a solution of compound **4.38-biotin** (90.0 mg, 0.21 mmol) in dry DMF (3.0 mL) was added 4-nitrobenzyl bromide (272 mg, 1.27 mmol) and DIPEA (100 μL, 0.57 mmol). The mixture was allowed to stir at room temperature for 48 h. After the solvent was removed under reduced pressure, the resulting crude material was purified by flash chromatography using 1-10% MeOH in CH<sub>2</sub>Cl<sub>2</sub>. Pure fractions were concentrated under reduced pressure to afford the product **4.39-biotin** (60 mg, 0.11 mmol, 51%) as a white solid. Data were consistent with those previously reported<sup>11</sup>

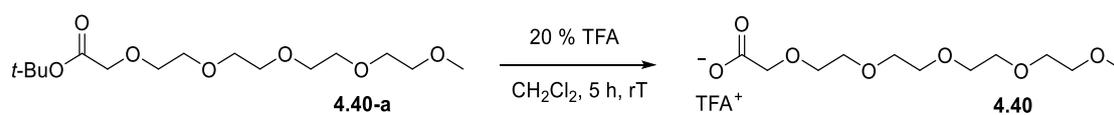


<sup>1</sup>H NMR (400 MHz, Chloroform-*d*): δ 8.41 (d, *J* = 8.8 Hz, 2H, -H<sub>3</sub>), 8.25 (d, *J* = 8.8 Hz, 2H, -H<sub>8</sub>), 8.10 (d, *J* = 8.8 Hz, 2H, -H<sub>2</sub>), 7.56 (d, *J* = 8.8 Hz, 2H, -H<sub>7</sub>), 5.18 (s, 2H, -H<sub>5</sub>), 4.50-4.47 (m, 1H, -H<sub>17</sub> or 18), 4.25-4.22 (m, 1H, -H<sub>17</sub> or 18), 3.07-3.02 (m, 1H, -H<sub>15</sub>), 2.87 (dd, *J* = 13.0, 5.2 Hz, 1H, -H<sub>16</sub>), 2.68 (d, *J* = 13.0 Hz, 1H, -H<sub>16</sub>), 2.54 (t, *J* = 7.0 Hz, 2H, -H<sub>11</sub>), 1.61-1.43 (m, 4H, -H<sub>12</sub> and -H<sub>14</sub>), 1.32-1.23 (m, 2H, -H<sub>13</sub>). HRMS (ESI<sup>+</sup>): *m/z* calcd for C<sub>23</sub>H<sub>25</sub>N<sub>5</sub>NaO<sub>8</sub>S<sub>2</sub>: 586.1037 [M+Na]; found: 586.1051

**Compound 4.40-a**

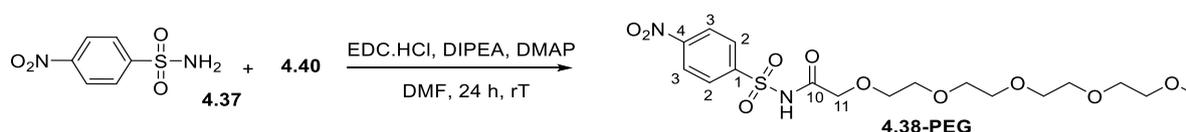
Triethylene glycol monomethyl ether (1 mL, 4.6 mmol) was dissolved in *tert*-butanol. Potassium *tert*-butoxide (1.1 g, 9.8 mmol) was added and the resulting reaction mixture was stirred for 2 h before the addition of *tert*-butyl bromoacetate (2.4 mL, 16.4 mmol). The mixture was stirred for 18 h at room temperature, filtered through Celite and the filtrate concentrated under reduced pressure to afford the product **4.40-a** as a transparent oil (1.4 g, 4.3 mmol, 94%). Data were consistent with those previously reported.<sup>24</sup>

<sup>1</sup>H NMR (400 MHz, Chloroform-*d*):  $\delta$  3.74-3.62 (m, 2H,  $-\text{CH}_2\text{COOH}$ ), 3.44-3.25 (m, 14H, PEG), 3.25-3.13 (m, 2H, PEG), 3.09 (s, 3H,  $-\text{OCH}_3$ ), 1.14 (m, 9H,  $-\text{O}t\text{-Bu}$ ). HRMS (ESI<sup>-</sup>):  $m/z$  calcd for  $\text{C}_{15}\text{H}_{30}\text{NaO}_7$ : 345.1884 [M+Na]; found: 345.1877

**Compound 4.40**

To a stirred solution of **4.40-a** (1.3 g, 4.0 mmol) in  $\text{CH}_2\text{Cl}_2$  (20 mL) was added TFA (5 mL). The mixture was stirred for 5 h at room temperature. After reducing the amount of solvent to 0.5 mL under reduced pressure, the crude mixture was extracted between  $\text{CH}_2\text{Cl}_2$  (50 mL) and  $\text{H}_2\text{O}$  (50 mL). The aqueous phase was concentrated under reduced pressure to afford the product **4.40** (0.6 g, 2.3 mmol, 56%) as a transparent oil. Data were consistent with those previously reported.<sup>25</sup>

<sup>1</sup>H NMR (400 MHz, Chloroform-*d*):  $\delta$  3.99 (s, 2H,  $-\text{CH}_2\text{COOH}$ ), 3.64 (app s, 14H, PEG), 3.52 (app s, 2H, PEG), 3.34 (s, 3H,  $-\text{OCH}_3$ ). HRMS (ESI<sup>-</sup>):  $m/z$  calcd for  $\text{C}_{11}\text{H}_{22}\text{O}_7$ : 265.1293 [M-H]; found: 265.1298

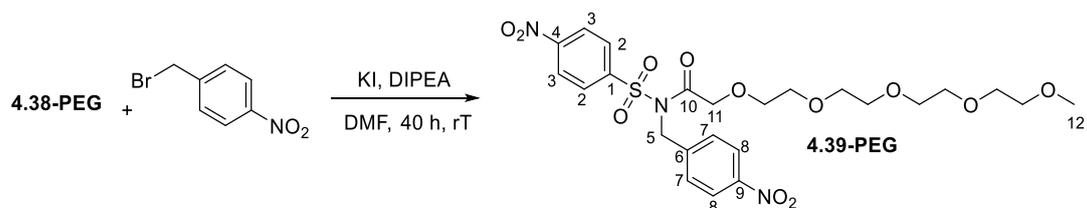
**Compound 4.38-PEG**

To a solution of 4-nitrobenzenesulfonamide **4.37** (36 mg, 0.18 mmol) in DMF (6 mL) was added PEG acid **4.40** (70 mg, 0.26 mmol), EDC.HCl (101 mg, 0.53 mmol), 4-dimethylaminopyridine (DMAP) (11 mg, 0.09 mmol) and DIPEA (90  $\mu\text{L}$ , 0.53 mmol). The mixture was allowed to stir at room temperature for 24 h. After the solvent was removed under reduced pressure, the resulting crude material was

purified by flash chromatography using 0-20% MeOH in CH<sub>2</sub>Cl<sub>2</sub>. Pure fractions were concentrated under reduced pressure to afford the product **4.38-PEG** (30 mg, 0.07 mmol, 37%) as an oil.

<sup>1</sup>H NMR (400 MHz, MeOD-*d*<sub>4</sub>): δ 8.34-8.29 (m, 2H, -H<sub>3</sub>), 8.16-8.12 (m, 2H, -H<sub>2</sub>), 3.99 (s, 2H, -COCH<sub>2</sub>O), 3.66-3.61 (m, 14H, PEG), 3.52-3.49 (m, 2H, -CH<sub>2</sub>OCH<sub>3</sub>), 3.29 (s, 3H, -OCH<sub>3</sub>). HRMS (ESI<sup>-</sup>): *m/z* calcd for C<sub>17</sub>H<sub>26</sub>N<sub>2</sub>O<sub>10</sub>S: 449.1235 [M-H]; found: 449.1241.

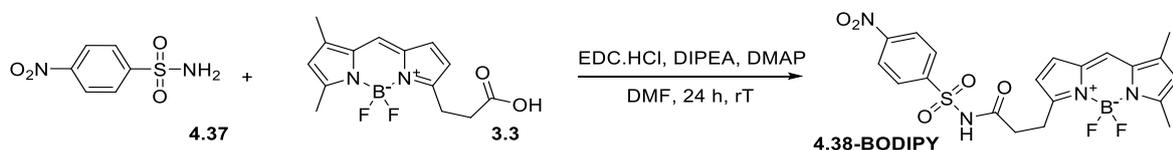
### Compound 4.39-PEG



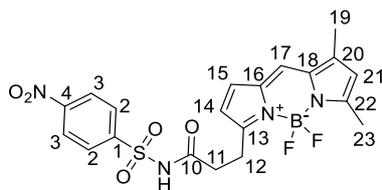
To a solution of **4.38-PEG** (22 mg, 49 μmol) in DMF (3 mL) was added 4-nitrobenzyl bromide (70 mg, 290 μmol), KI (26 mg, 150 μmol) and DIPEA (26 μL, 150 μmol). The mixture was allowed to stir at room temperature for 40 h. After the solvent was removed under reduced pressure, the resulting crude material was purified by flash chromatography using 0-20% MeOH in CH<sub>2</sub>Cl<sub>2</sub>. Pure fractions were concentrated under reduced pressure to afford the product **4.39-PEG** as a yellow oil.

<sup>1</sup>H NMR (400 MHz, Chloroform-*d*): δ 8.38 (d, *J* = 8.9 Hz, 2H, -H<sub>3</sub>), 8.20 (d, *J* = 8.7 Hz, 2H, -H<sub>8</sub>), 8.10 (d, *J* = 8.9 Hz, 2H, -H<sub>2</sub>), 7.55 (d, *J* = 8.7 Hz, 2H, -H<sub>7</sub>), 5.12 (s, 2H, -NCH<sub>2</sub>), 4.42 (s, 2H, -COCH<sub>2</sub>), 3.63-3.55 (m, 14H, PEG -CH<sub>2</sub>), 3.52-3.49 (m, 2H, PEG -CH<sub>2</sub>), 3.33 (s, 3H, -CH<sub>3</sub>). <sup>13</sup>C NMR (101 MHz, Chloroform-*d*): δ 170.04 (C10), 150.94 (C1), 147.76 (C6), 144.10 (C4), 143.00 (C9), 129.55 (C2), 128.66 (C7), 124.68 (C3), 124.11 (C8), 71.97 (C-PEG), 71.65 (C11), 71.19 (C-PEG), 70.69 (C-PEG), 70.66 (C-PEG), 70.63 (C-PEG), 70.60 (C-PEG), 70.57 (C-PEG), 59.11 (C12), 49.15 (C5). HRMS (ESI<sup>+</sup>): *m/z* calcd for C<sub>24</sub>H<sub>31</sub>NaN<sub>3</sub>O<sub>12</sub>S: 608.1521 [M+Na]; found: 608.1522

### Compound 4.38-BODIPY

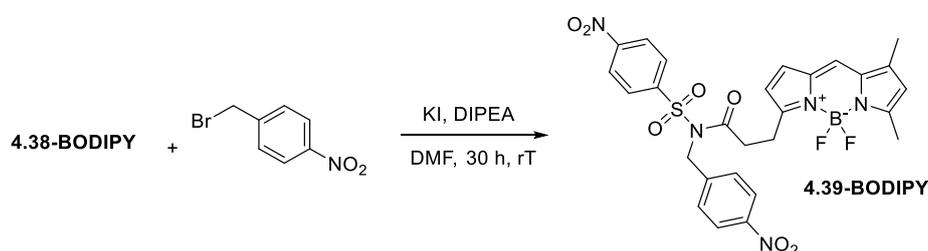


To a solution of 4-nitrobenzenesulfonamide **4.37** (93 mg, 0.46 mmol) in DMF (5 mL) was added BODIPY **3.3** (150 mg, 0.51 mmol), EDC.HCl (177 mg, 0.92 mmol), 4-dimethylaminopyridine (DMAP) (113 mg, 0.92 mmol) and DIPEA (242 μL, 1.39 mmol). The mixture was allowed to stir at room temperature for 24 h. After the solvent was removed under reduced pressure, the resulting crude material was purified by flash chromatography using 0.25-0.75% MeOH in CH<sub>2</sub>Cl<sub>2</sub>. Pure fractions were concentrated under reduced pressure to afford the product **4.38-BODIPY** as a red solid (25 mg, 0.05 mmol, 11%).

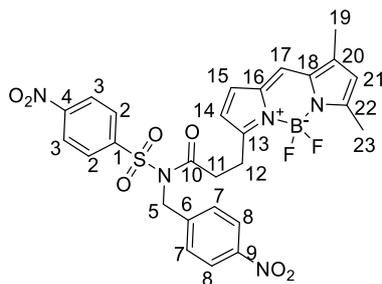


$^1\text{H NMR}$  (400 MHz, Chloroform-*d*):  $\delta$  8.70 (s, 1H, -NH), 8.29-8.20 (m, 2H, -H<sub>3</sub>), 8.15-8.09 (m, 2H, -H<sub>2</sub>), 7.07 (s, 1H, -H<sub>17</sub>), 6.72 (d,  $J = 4.0$  Hz, 1H, -H<sub>15</sub>), 6.17 (s, 1H, -H<sub>21</sub>), 6.03 (d,  $J = 4.0$  Hz, 1H, -H<sub>14</sub>), 3.13 (t,  $J = 7.3$  Hz, 2H, -H<sub>11</sub>), 2.72 (t,  $J = 7.3$  Hz, 2H, -H<sub>12</sub>), 2.56 (s, 3H, -H<sub>23</sub>), 2.28 (s, 3H, -H<sub>19</sub>). **HRMS** (ESI<sup>+</sup>):  $m/z$  calcd for C<sub>20</sub>H<sub>19</sub>BF<sub>2</sub>N<sub>4</sub>NaO<sub>5</sub>S: 499.1029 [M+Na]; found: 499.1048

### Compound 4.39-BODIPY



To a solution of **4.38-BODIPY** (10 mg, 0.021 mmol) in dry DMF (4.0 mL) was added 4-nitrobenzyl bromide (14 mg, 0.06 mmol), potassium iodide (11 mg, 0.06 mmol) and DIPEA (10  $\mu\text{L}$ , 0.06 mmol). The mixture was allowed to stir at room temperature for 30 h. After the solvent was removed under reduced pressure, the resulting crude material was purified by flash chromatography using 20-50% EtOAc in PET. Pure fractions were concentrated under reduced pressure to afford the product **4.39-BODIPY** (4.2 mg, 6.9  $\mu\text{mol}$ , 33%) as a red powder.

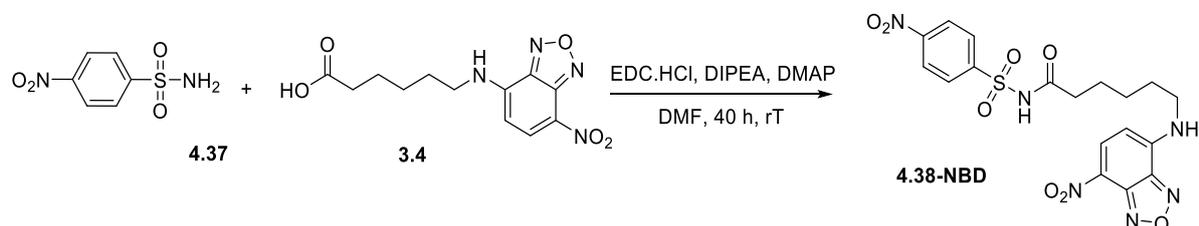


$^1\text{H NMR}$  (400 MHz, Chloroform-*d*):  $\delta$  8.23 (d,  $J = 9.0$  Hz, 2H, -H<sub>3</sub>), 8.17 (d,  $J = 8.8$  Hz, 2H, -H<sub>8</sub>), 7.97 (d,  $J = 9.0$  Hz, 2H, -H<sub>2</sub>), 7.51 (d,  $J = 8.8$  Hz, 2H, -H<sub>7</sub>), 7.02 (s, 1H, -H<sub>17</sub>), 6.74 (d,  $J = 4.0$  Hz, 1H, -H<sub>15</sub>), 6.14 (s, 1H, -H<sub>21</sub>), 6.05 (d,  $J = 4.0$  Hz, 1H, -H<sub>14</sub>), 5.18 (s, 2H, -NCH<sub>2</sub>), 3.14 (t,  $J = 6.9$  Hz, 2H, -H<sub>11</sub>), 2.99 (t,  $J = 6.9$  Hz, 2H, -H<sub>12</sub>), 2.49 (s, 3H, -CH<sub>3</sub>23), 2.27 (s, 3H, -CH<sub>3</sub>19).  $^{13}\text{C NMR}$  (101 MHz, Chloroform-*d*)  $\delta$  172.13 (C10), 161.57 (C22), 155.05 (C13), 150.54 (C1), 147.63 (C6), 145.24 (C18), 144.61 (C4), 143.44 (C9), 135.58 (C20), 133.19 (C16), 129.30 (C2), 128.50 (C7), 127.48 (C15), 124.44 (C3), 124.12 (C8),

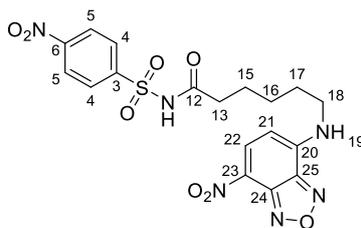
123.83 (C17), 121.13 (C21), 117.65 (C14), 49.40 (C5), 36.43 (C11), 23.70 (C12), 15.06 (C23), 11.47 (C19).

**HRMS** (ESI<sup>+</sup>):  $m/z$  calcd for C<sub>27</sub>H<sub>24</sub>BF<sub>2</sub>N<sub>5</sub>NaO<sub>7</sub>S: 634.1350 [M+Na]; found: 634.1368

### Compound 4.38-NBD

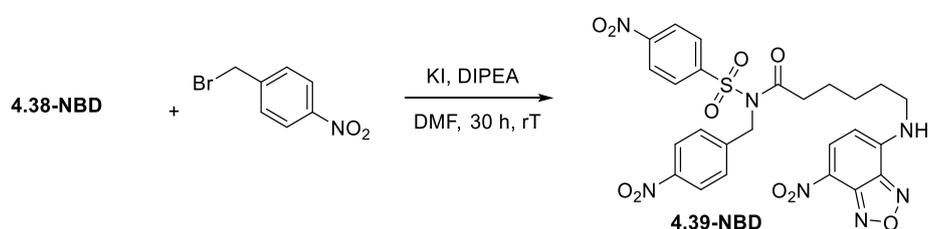


To a solution of 4-nitrobenzenesulfonamide **4.37** (101 mg, 0.50 mmol) in DMF (5 mL) was added NBD **3.4** (219 mg, 0.75 mmol), EDC.HCl (300 mg, 1.5 mmol), 4-dimethylaminopyridine (DMAP) (33.6 mg, 0.25 mmol) and DIPEA (260  $\mu$ L, 1.5 mmol). The mixture was allowed to stir at room temperature for 40 h. After the solvent was removed under reduced pressure, the resulting crude material was purified by flash chromatography using 0.25-5% MeOH in CH<sub>2</sub>Cl<sub>2</sub>. Pure fractions were concentrated under reduced pressure to afford the product **4.38-NBD** as an orange solid (36 mg, 0.07 mmol, 14%).



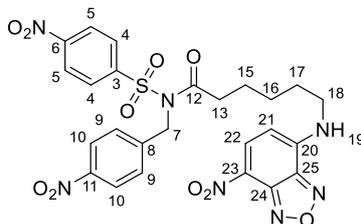
**<sup>1</sup>H NMR** (400 MHz, DMSO-*d*<sub>6</sub>):  $\delta$  8.50 (d,  $J$  = 9.0 Hz, 1H, -H<sub>21</sub>), 8.46-8.38 (m, 2H, -H<sub>5</sub>), 8.17-8.12 (m, 2H, -H<sub>4</sub>), 6.37 (d,  $J$  = 9.0 Hz, 1H, -H<sub>22</sub>), 3.45 (td,  $J_1 = J_2 = 6.8$  Hz, 2H, -CH<sub>2</sub>NH), 2.24 (t,  $J$  = 7.3 Hz, 2H, -CH<sub>2</sub>CO), 1.60 (tt,  $J_1 = J_2 = 7.3$  Hz, 2H, -CH<sub>2</sub>CH<sub>2</sub>NH), 1.45 (tt,  $J_1 = J_2 = 7.3$  Hz, 2H, -CH<sub>2</sub>CH<sub>2</sub>O), 1.29- 1.21 (m, 2H, -CH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>O). **HRMS** (ESI<sup>+</sup>):  $m/z$  calcd for C<sub>18</sub>H<sub>17</sub>N<sub>6</sub>O<sub>8</sub>S: 477.0834 [M+H]; found: 477.0840

### Compound 4.39-NBD



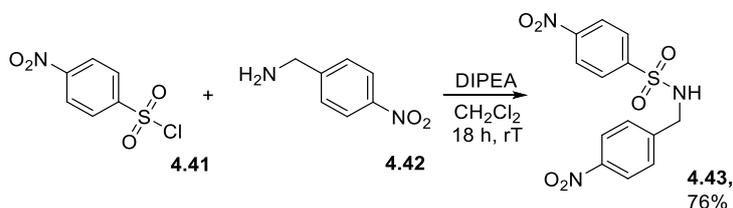
To a solution of **4.38-NBD** (36 mg, 0.075 mmol) in dry DMF (2.0 mL) was added 4-nitrobenzyl bromide (98 mg, 0.45 mmol), potassium iodide (46 mg, 0.22 mmol) and DIPEA (40  $\mu$ L, 0.22 mmol). The mixture was allowed to stir at room temperature for 30 h. After the solvent was removed under reduced pressure, the resulting crude material was purified by flash chromatography using 0-0.5% MeOH in

CH<sub>2</sub>Cl<sub>2</sub>. Pure fractions were concentrated under reduced pressure to afford the product **4.39-NBD** as an orange solid (10 mg, 16 μmol, 22%).

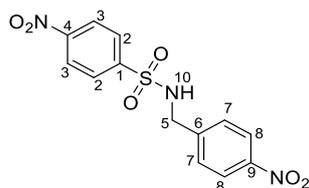


**<sup>1</sup>H NMR** (500 MHz, Chloroform-*d*): δ 8.47 (d, *J* = 8.5 Hz, 1H, -H<sub>21</sub>), 8.40 (d, *J* = 8.5 Hz, 2H, -H<sub>5</sub>), 8.24 (d, *J* = 8.2 Hz, 2H, -H<sub>10</sub>), 8.08 (d, *J* = 8.5 Hz, 2H, -H<sub>4</sub>), 7.56 (d, *J* = 8.2 Hz, 2H, -H<sub>9</sub>), 6.21 (s, 1H), 6.13 (d, *J* = 8.5 Hz, 1H, -H<sub>22</sub>), 5.15 (s, 2H, -H<sub>7</sub>), 3.45 (td, *J*<sub>1</sub> = *J*<sub>2</sub> = 6.8 Hz, 2H, -H<sub>18</sub>), 2.59 (t, *J* = 7.0 Hz, 2H, -H<sub>13</sub>), 1.76 (tt, *J*<sub>1</sub> = *J*<sub>2</sub> = 7.0 Hz, 2H, -H<sub>17</sub>), 1.65 (tt, *J*<sub>1</sub> = *J*<sub>2</sub> = 7.0 Hz, 2H, -H<sub>15</sub>), 1.39 (m, 2H, -H<sub>16</sub>). **<sup>13</sup>C NMR** (126 MHz, Chloroform-*d*): δ 207.28 (C<sub>20</sub>), 172.92 (C<sub>12</sub>), 150.94 (C<sub>3</sub>), 147.86 (C<sub>8</sub>), 144.59 (C<sub>24</sub> or C<sub>25</sub>), 144.35 (C<sub>24</sub> or C<sub>25</sub>), 143.65 (C<sub>6</sub>), 143.15 (C<sub>11</sub>), 136.37 (C<sub>22</sub>), 129.31 (C<sub>4</sub>), 128.37 (C<sub>9</sub>), 124.69 (C<sub>5</sub>), 124.24 (C<sub>10</sub>), 98.41 (C<sub>21</sub>), 48.64 (C<sub>7</sub>), 43.48 (C<sub>18</sub>), 36.16 (C<sub>13</sub>), 28.20 (C<sub>17</sub>), 26.78 (C<sub>15</sub>), 23.42 (C<sub>16</sub>). **HRMS** (ESI<sup>+</sup>): *m/z* calcd for C<sub>25</sub>H<sub>22</sub>N<sub>7</sub>O<sub>10</sub>S: 612.1154 [M+H]; found: 612.1145

### Compound 4.43



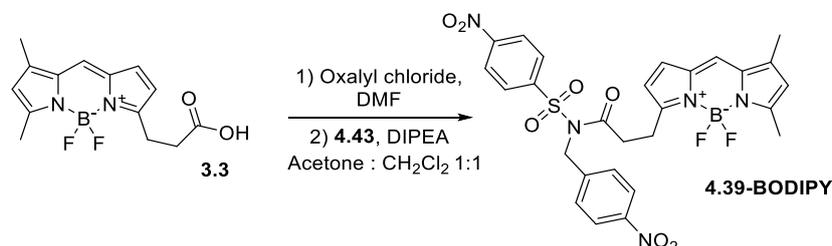
Nitrobenzylamine **4.42** (564 mg, 3.0 mmol) and chlorosulfonamide **4.41** (663 mg, 3.0 mmol) were dissolved in CH<sub>2</sub>Cl<sub>2</sub> (10 mL) and DIPEA (1.5 mL, 9.0 mmol) was added. The mixture was stirred overnight at room temperature and after the solvent was removed under reduced pressure, the resulting crude material was purified by flash chromatography using 30-50% EtOAc in PET. Pure fractions were concentrated under reduced pressure to afford the product **4.43** as a white solid (770 mg, 2.3 mmol, 76%).



**<sup>1</sup>H NMR** (400 MHz, DMSO-*d*<sub>6</sub>): δ 8.78 (1H, t, *J* = 5.9 Hz, -NH), 8.37 (2H, d, *J* = 8.9 Hz, -H<sub>3</sub>), 8.14 (2H, d, *J* = 8.8 Hz, -H<sub>8</sub>), 8.02 (2H, d, *J* = 8.9 Hz, -H<sub>2</sub>), 7.50 (d, *J* = 8.8 Hz, -H<sub>7</sub>), 4.23 (2H, d, *J* = 5.9 Hz, -CH<sub>2</sub>). **<sup>13</sup>C NMR** (101 MHz, DMSO-*D*<sub>6</sub>): δ 150.07 (C<sub>1</sub>), 147.18 (C<sub>4</sub>), 146.65 (C<sub>6</sub>), 145.90 (C<sub>9</sub>), 129.19 (C<sub>7</sub>), 128.7

and 128.6 (C2), 125.2 and 125.1 (C3), 124.0 and 123.9 (C8), 45.86 (C5). HRMS (ESI<sup>+</sup>):  $m/z$  calcd for C<sub>13</sub>H<sub>10</sub>N<sub>3</sub>O<sub>6</sub>S: 336.0296 [M-H]; found: 336.0292.

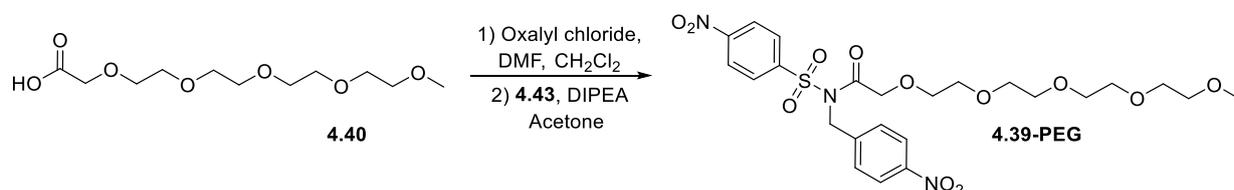
#### Compound 4.39-BODIPY using intermediate 4.43



BODIPY **3.3** (50.7 mg, 0.17 mmol) was dissolved in anhydrous CH<sub>2</sub>Cl<sub>2</sub> (10 mL) containing 3 drops of DMF. Oxalyl chloride (41  $\mu$ L, 0.48 mmol) was added to the solution and the reaction mixture was stirred for 15 min at room temperature. Then the solvent was removed under reduced pressure and the residue was re-evaporated from toluene (2 x 10 mL) to give the intermediate acid chloride as a red oil. The resulting material was dissolved in CH<sub>2</sub>Cl<sub>2</sub> (8 mL) and the solution was added dropwise to a solution of compound **4.43** (52.5 mg, 0.16 mmol) in CH<sub>2</sub>Cl<sub>2</sub>:Acetone (1:1, 10 mL) containing DIPEA (80  $\mu$ L, 0.47 mmol). The reaction mixture was stirred for 20 min at room temperature then diluted with CH<sub>2</sub>Cl<sub>2</sub> (10 mL) and washed with aqueous acetic acid (3% in H<sub>2</sub>O, 2 x 30 mL), H<sub>2</sub>O (30 mL), and sat. NaHCO<sub>3</sub> (2 x 30 mL), dried over sodium sulfate and the solvent removed under reduced pressure. The resulting crude material was purified by flash chromatography using 20-50% EtOAc in PET. Pure fractions were concentrated under reduced pressure to afford the product **4.39-BODIPY** as a red solid (62 mg, 0.11 mmol, 63%).

Data consistent with previous experiment.

#### Compound 4.39-PEG using intermediate 4.43

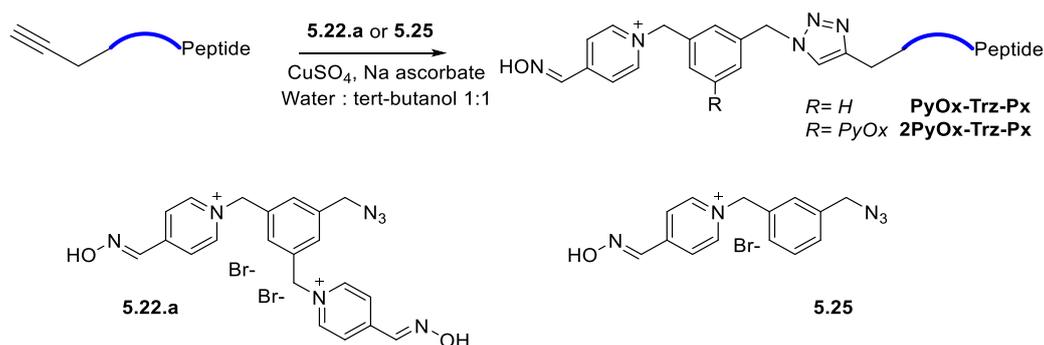


**4.40** (120 mg, 0.45 mmol) was dissolved in dry CH<sub>2</sub>Cl<sub>2</sub> (10 mL) containing 3 drops of DMF. Oxalyl chloride (110  $\mu$ L, 1.30 mmol) was added to the solution and the reaction mixture was stirred for 15 min at room temperature. Then the solvent was removed under reduced pressure and the residue was re-evaporated from toluene (2 x 10 mL) to give the intermediate acid chloride as an oil. The resulting material was dissolved in acetone (8 mL) and the solution was added dropwise to the solution of compound **4.43** (52.5 mg, 0.16 mmol) in acetone (10 mL) containing DIPEA (215  $\mu$ L, 1.23 mmol). The reaction mixture was stirred for 15 min at room temperature then diluted with CH<sub>2</sub>Cl<sub>2</sub> (30 mL) and

washed with acetic acid (3% in H<sub>2</sub>O, 2 x 30 mL), H<sub>2</sub>O (30 mL), and sat. NaHCO<sub>3</sub> (2x30 mL), dried over sodium sulfate and the solvent was removed under reduced pressure. The resulting crude material was purified by flash chromatography using 25-100% EtOAc in PET. Pure fractions were concentrated under reduced pressure to afford the product **4.39-PEG** as a light yellow oil (95 mg, 0.18 mmol, 40%).

Data consistent with previous experiment.

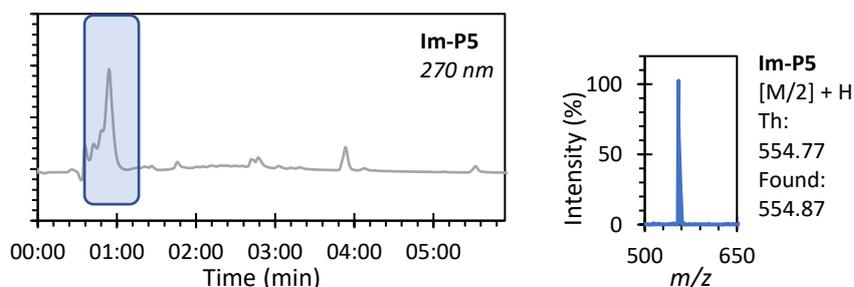
### General experimental for click-chemistry reaction

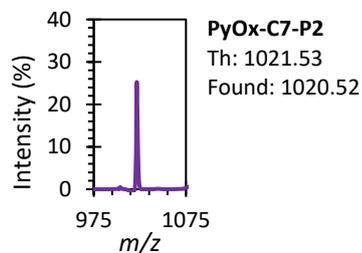
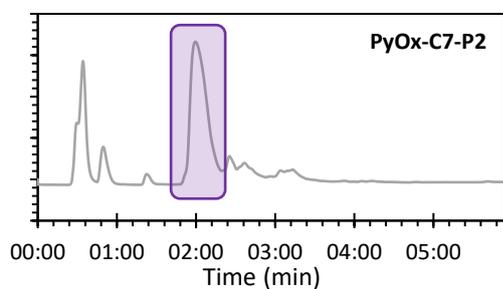
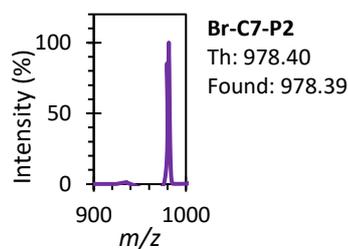
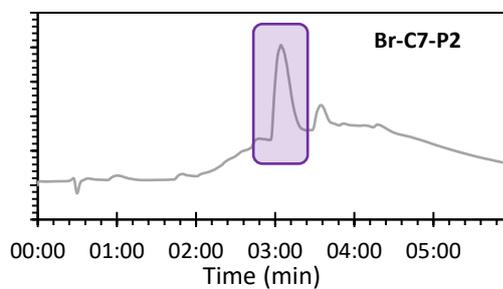
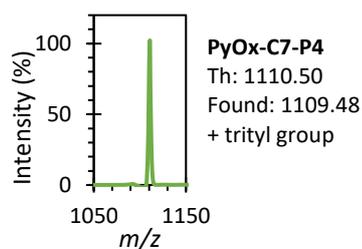
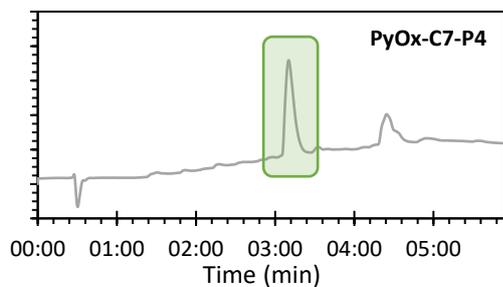
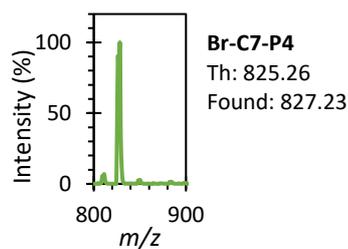
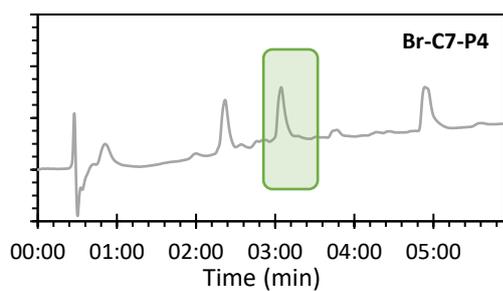
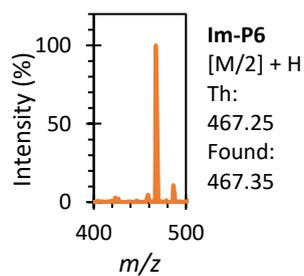
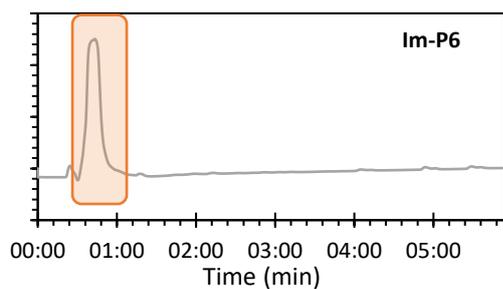


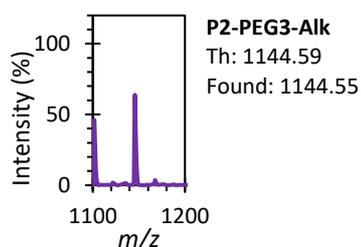
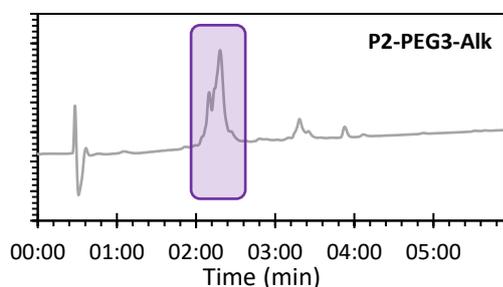
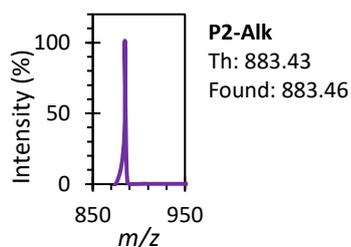
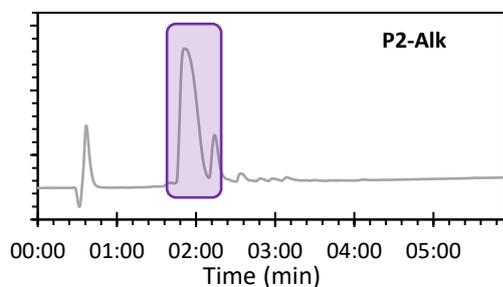
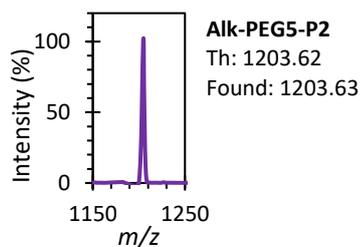
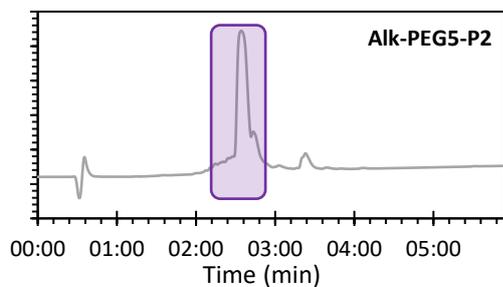
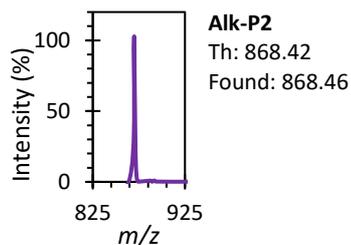
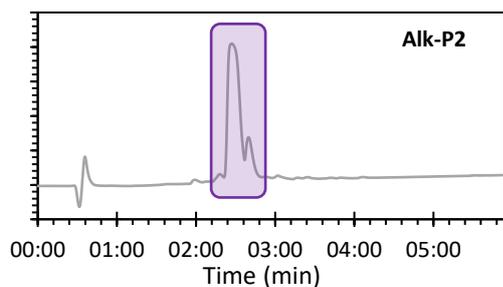
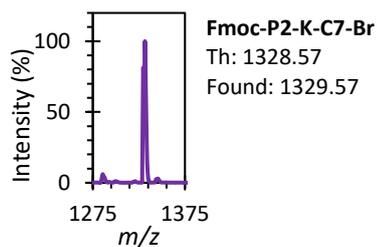
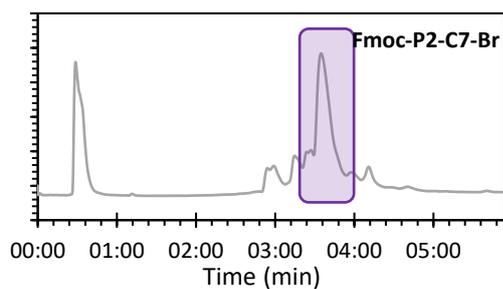
**5.22.a** or **5.25** (1.2 eq) and **alkyne-peptide** (1 eq) were dissolved in a mixture of *tert*-butanol (0.5 mL) and water (0.5 mL), then copper sulfate (1.5 eq) and sodium ascorbate (3 eq) were added. The resulting mixture was stirred for 1.5-2 h. The crude mixture was analysed by LCMS to ensure that the alkyne-peptide was fully transformed. If not complete, a further 0.5 eq of azide was added to ensure full conversion. Once fully converted, the product was purified by reverse-phase column chromatography, see details in Chapter 2. Pure fractions were lyophilised to give pure peptides as white solids. Due to the small scale of these reactions, compounds were analysed only via LCMS and high-resolution MS to confirm identity.

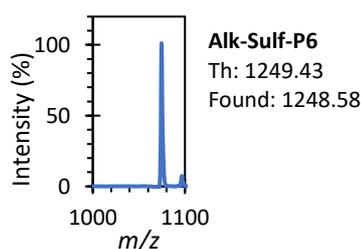
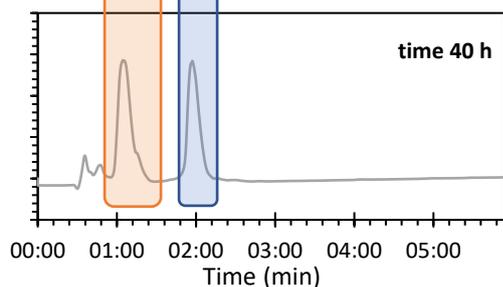
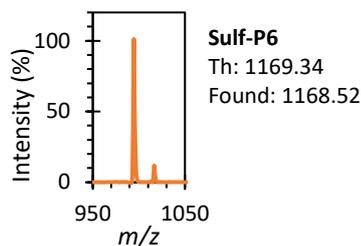
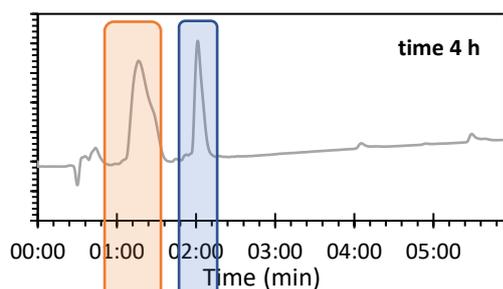
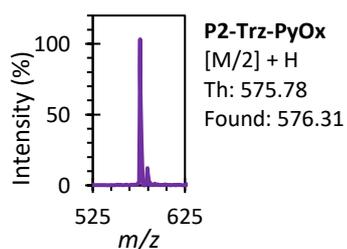
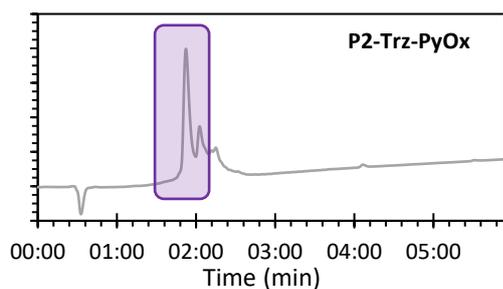
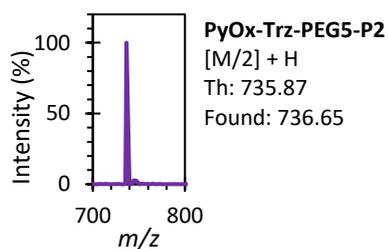
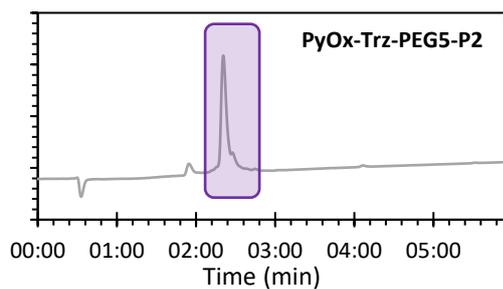
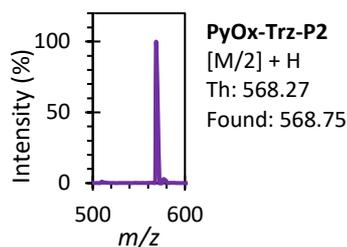
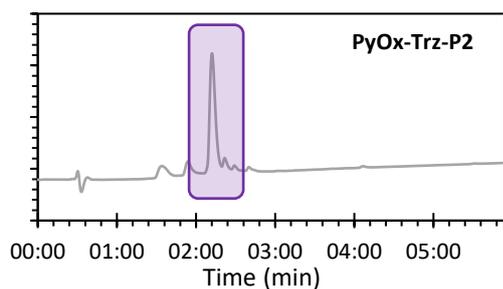
### 4.6.2 LCMS of peptides used in this Chapter

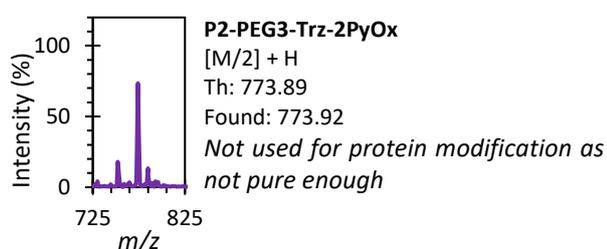
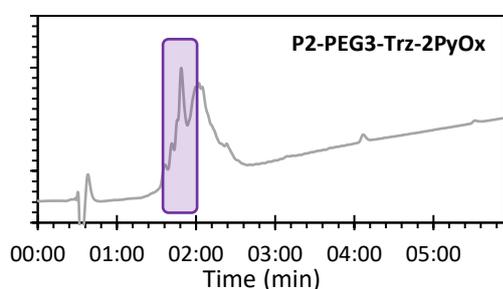
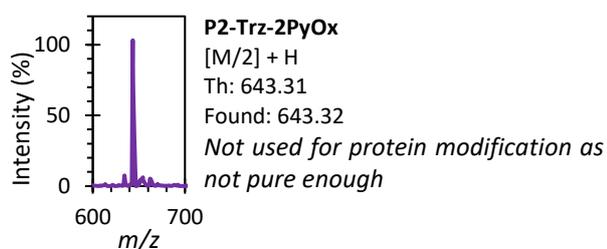
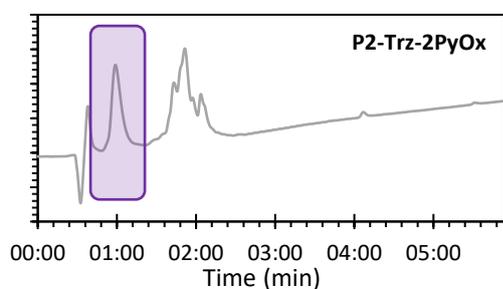
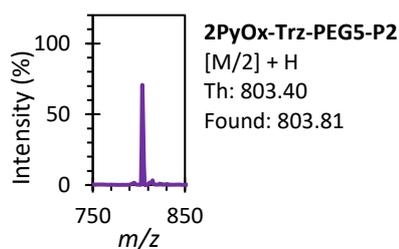
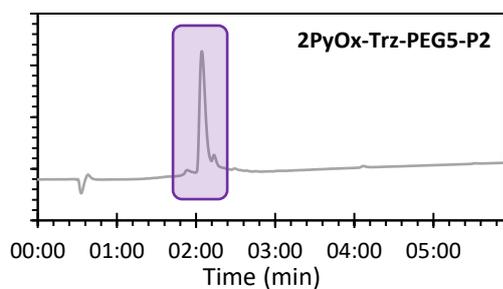
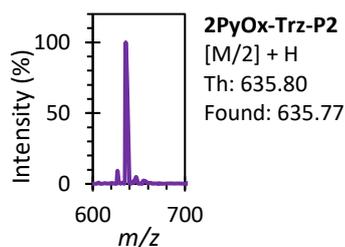
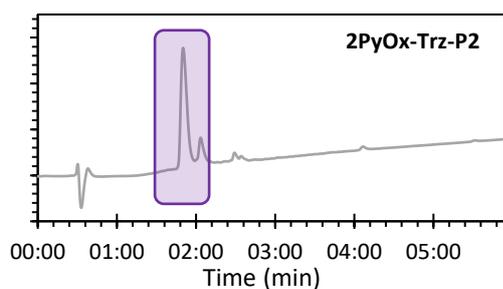
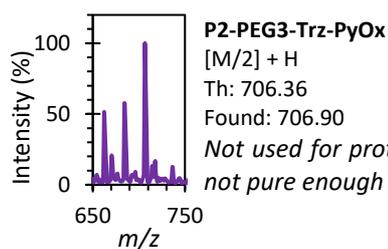
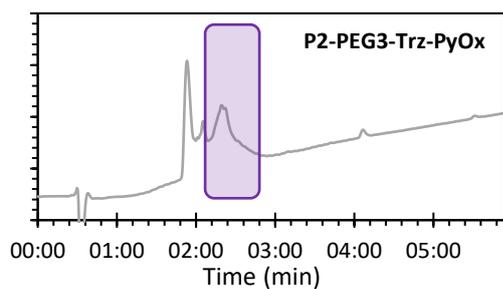
All the chromatograms are of absorbance at 220 nm except when noted otherwise. The goals of the LCMS are to show the presence of the expected peptides, most of them were carried for further modification thus none of those chromatograms are used for purity analysis.





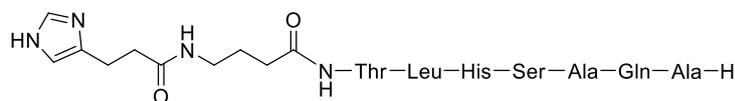


**Sulf-P6 modified with 4.31:****PyOx peptides after reverse phase chromatography:**

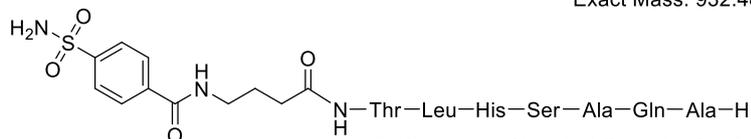


## 4.6.3 Structures of peptides used in this Chapter not presented in discussion

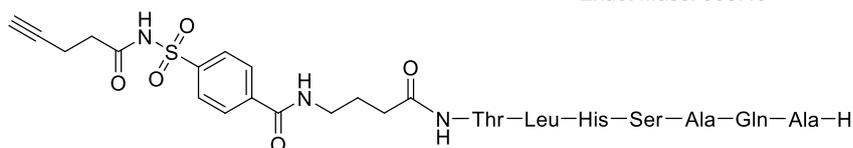
## Peptides for OpLD



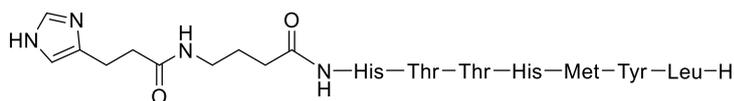
**Im-P6** Chemical Formula: C<sub>40</sub>H<sub>64</sub>N<sub>14</sub>O<sub>12</sub>  
Exact Mass: 932.48



**Sulf-P6** Chemical Formula: C<sub>41</sub>H<sub>63</sub>N<sub>13</sub>O<sub>14</sub>S  
Exact Mass: 993.43

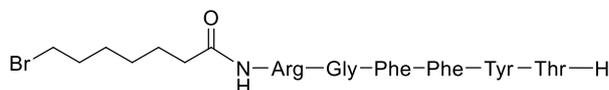


**Alk-Sulf-P6** Chemical Formula: C<sub>46</sub>H<sub>67</sub>N<sub>13</sub>O<sub>15</sub>S  
Exact Mass: 1073.46



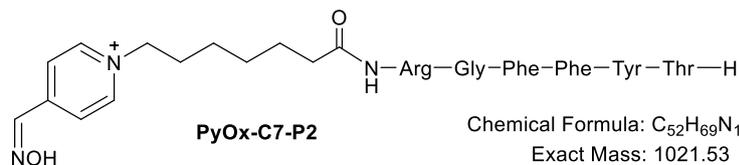
**Im-P5** Chemical Formula: C<sub>50</sub>H<sub>73</sub>N<sub>15</sub>O<sub>12</sub>S  
Exact Mass: 1107.53

## Peptides for CafLD



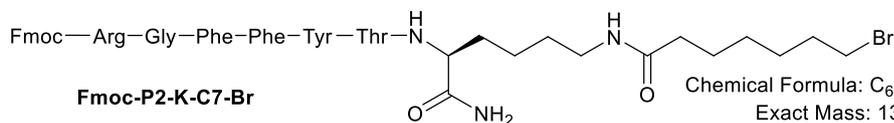
Chemical Formula: C<sub>46</sub>H<sub>63</sub>BrN<sub>10</sub>O<sub>9</sub>  
Exact Mass: 978.40

**Br-C7-P2**



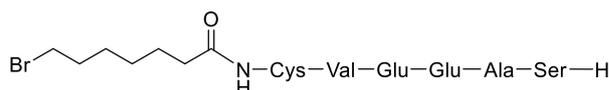
Chemical Formula: C<sub>52</sub>H<sub>69</sub>N<sub>12</sub>O<sub>10</sub><sup>+</sup>  
Exact Mass: 1021.53

**PyOx-C7-P2**



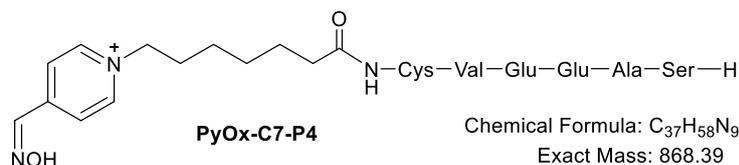
Chemical Formula: C<sub>67</sub>H<sub>85</sub>BrN<sub>12</sub>O<sub>12</sub>  
Exact Mass: 1328.56

**Fmoc-P2-K-C7-Br**



Chemical Formula: C<sub>31</sub>H<sub>52</sub>BrN<sub>7</sub>O<sub>12</sub>S  
Exact Mass: 825.26

**Br-C7-P4**



Chemical Formula: C<sub>37</sub>H<sub>58</sub>N<sub>9</sub>O<sub>13</sub>S<sup>+</sup>  
Exact Mass: 868.39

**PyOx-C7-P4**

## 4.7 Bibliography

- (1) Hayashi, T.; Hamachi, I. Traceless Affinity Labeling of Endogenous Proteins for Functional Analysis in Living Cells. *Acc. Chem. Res.* **2012**, *45* (9), 1460–1469. <https://doi.org/10.1021/ar200334>.
- (2) Mino, T.; Sakamoto, S.; Hamachi, I. Recent Applications of N-Acyl Imidazole Chemistry in Chemical Biology. *Biosci. Biotechnol. Biochem.* **2021**, *85* (1), 53–60. <https://doi.org/10.1093/bbb/zbaa026>.
- (3) Fujishima, S. H.; Yasui, R.; Miki, T.; Ojida, A.; Hamachi, I. Ligand-Directed Acyl Imidazole Chemistry for Labeling of Membrane-Bound Proteins on Live Cells. *J. Am. Chem. Soc.* **2012**, *134* (9), 3961–3964. <https://doi.org/10.1021/ja2108855>.
- (4) Mortensen, M. R.; Nielsen, N. L.; Palmfeldt, J.; Gothelf, K. V. Imidazole Carbamate Probes for Affinity Guided Azide-Transfer to Metal-Binding Proteins. *Org. Biomol. Chem.* **2019**, *17* (6), 1379–1383. <https://doi.org/10.1039/C8OB03017K>.
- (5) Wakayama, S.; Kiyonaka, S.; Arai, I.; Kakegawa, W.; Matsuda, S.; Iyata, K.; Nemoto, Y. L.; Kusumi, A.; Yuzaki, M.; Hamachi, I. Chemical Labelling for Visualizing Native AMPA Receptors in Live Neurons. *Nat. Commun.* **2017**, *8*. <https://doi.org/10.1038/ncomms14850>.
- (6) Miki, T.; Fujishima, S.; Komatsu, K.; Kuwata, K.; Kiyonaka, S.; Hamachi, I. LDAI-Based Chemical Labeling of Intact Membrane Proteins and Its Pulse-Chase Analysis under Live Cell Conditions. *Chem. Biol.* **2014**, *21* (8), 1013–1022. <https://doi.org/10.1016/j.chembiol.2014.07.013>.
- (7) Matsuo, K.; Kioi, Y.; Yasui, R.; Takaoka, Y.; Miki, T.; Fujishima, S.; Hamachi, I. One-Step Construction of Caged Carbonic Anhydrase I Using a Ligand-Directed Acyl Imidazole-Based Protein Labeling Method. *Chem. Sci.* **2013**, *4*, 2573–2580. <https://doi.org/10.1039/c3sc50560j>.
- (8) Bridger, G.; Kaller, A.; Harwig, C.; Skerlj, R.; Bogucki, D.; Wilson, T. R.; Crawford, J.; McEachern, E. J.; Atsma, B.; Nan, S.; Zhou, Y.; Schols, D.; Smith, C. D.; Rosaria Di Fluri, M. Chemokine Receptor Binding Heterocyclic Compounds with Enhanced Efficacy. U.S. Patent WO 2004106493, 2008.
- (9) Tamura, T.; Ueda, T.; Goto, T.; Tsukidate, T.; Shapira, Y.; Nishikawa, Y.; Fujisawa, A.; Hamachi, I. Rapid Labelling and Covalent Inhibition of Intracellular Native Proteins Using Ligand-Directed N-Acyl-N-Alkyl Sulfonamide. *Nat. Commun.* **2018**, *9* (1), 1–12. <https://doi.org/10.1038/s41467-018-04343-0>.
- (10) Thimaradka, V.; Hoon Oh, J.; Heroven, C.; Radu Aricescu, A.; Yuzaki, M.; Tamura, T.; Hamachi, I. Site-Specific Covalent Labeling of His-Tag Fused Proteins with N-Acyl-N-Alkyl Sulfonamide Reagent. *Bioorganic Med. Chem.* **2021**, *30*, 115947. <https://doi.org/10.1016/j.bmc.2020.115947>.
- (11) Tamura, T.; Song, Z.; Amaike, K.; Lee, S.; Yin, S.; Kiyonaka, S.; Hamachi, I. Affinity-Guided Oxime Chemistry for Selective Protein Acylation in Live Tissue Systems. *J. Am. Chem. Soc.* **2017**, *139* (40), 14181–14191. <https://doi.org/10.1021/jacs.7b07339>.
- (12) Keijzer, J. F.; Firet, J.; Albada, B. Site-Selective and Inducible Acylation of Thrombin Using Aptamer-Catalyst Conjugates. *Chem. Commun.* **2021**, *57* (96), 12960–12963. <https://doi.org/10.1039/d1cc05446e>.
- (13) Teng, M.; Jiang, J.; Ficarro, S. B.; Seo, H.-S.; Bae, J. H.; Donovan, K. A.; Fischer, E. S.; Zhang, T.; Dhe-Paganon, S.; Marto, J. A.; Gray, N. S. Exploring Ligand-Directed N -Acyl- N -

- Alkylsulfonamide-Based Acylation Chemistry for Potential Targeted Degradation Development. *ACS Med. Chem. Lett.* **2021**, *12* (8), 1302–1307. <https://doi.org/10.1021/acsmchemlett.1c00285>.
- (14) Ueda, T.; Tamura, T.; Kawano, M.; Shiono, K.; Hobor, F.; Wilson, A. J.; Hamachi, I. Enhanced Suppression of a Protein-Protein Interaction in Cells Using Small-Molecule Covalent Inhibitors Based on an N-Acyl- N-Alkyl Sulfonamide Warhead. *J. Am. Chem. Soc.* **2021**, *143* (12), 4766–4774. <https://doi.org/10.1021/jacs.1c00703>.
- (15) Jensen, K. J. *Solid-Phase Peptide Synthesis: An Introduction*; 2013; Vol. 1047. [https://doi.org/10.1007/978-1-62703-544-6\\_1](https://doi.org/10.1007/978-1-62703-544-6_1).
- (16) Delor, M.; Dai, J.; Roberts, T. D.; Rogers, J. R.; Hamed, S. M.; Neaton, J. B.; Geissler, P. L.; Francis, M. B.; Ginsberg, N. S. Exploiting Chromophore-Protein Interactions through Linker Engineering to Tune Photoinduced Dynamics in a Biomimetic Light-Harvesting Platform. *J. Am. Chem. Soc.* **2018**, *140* (20), 6278–6287. <https://doi.org/10.1021/jacs.7b13598>.
- (17) Parai, M. K.; Huggins, D. J.; Cao, H.; Nalam, M. N. L.; Ali, A.; Schiffer, C. A.; Tidor, B.; Rana, T. M. Design, Synthesis, and Biological and Structural Evaluations of Novel HIV-1 Protease Inhibitors to Combat Drug Resistance. *J. Med. Chem.* **2012**, *55* (14), 6328–6341. <https://doi.org/10.1021/jm300238h>.
- (18) Takaoka, Y.; Tsutsumi, H.; Kasagi, N.; Nakata, E.; Hamachi, I. One-Pot and Sequential Organic Chemistry on an Enzyme Surface to Tether a Fluorescent Probe at the Proximity of the Active Site with Restoring Enzyme Activity. *J. Am. Chem. Soc.* **2006**, *128* (10), 3273–3280. <https://doi.org/10.1021/ja057926x>.
- (19) Gao, Y.; Lin, Y.; Liu, T.; Zhang, X.; Xu, F.; Liu, P.; Du, L.; Li, M. A Specific and Selective Chemiluminescent Probe for Pd<sup>2+</sup> Detection. *Chinese Chem. Lett.* **2019**, *30* (1), 63–66. <https://doi.org/10.1016/j.ccl.2018.03.028>.
- (20) Nathanael, J. G.; White, J. M.; Richter, A.; Nuske, M. R.; Wille, U. Oxidative Damage of Proline Residues by Nitrate Radicals (NO<sub>3</sub>): A Kinetic and Product Study. *Org. Biomol. Chem.* **2020**, *18* (35), 6949–6957. <https://doi.org/10.1039/d0ob01337d>.
- (21) Ghazy, E.; Zeyen, P.; Herp, D.; Hügler, M.; Schmidtkunz, K.; Erdmann, F.; Robaa, D.; Schmidt, M.; Morales, E. R.; Romier, C.; Günther, S.; Jung, M.; Sippl, W. Design, Synthesis, and Biological Evaluation of Dual Targeting Inhibitors of Histone Deacetylase 6/8 and Bromodomain BRPF1. *Eur. J. Med. Chem.* **2020**, *200*, 112338. <https://doi.org/10.1016/j.ejmech.2020.112338>.
- (22) Assali, M.; Kittana, N.; Qasem, S. A.; Adas, R.; Saleh, D.; Arar, A.; Zohud, O. Combretastatin A4-Camptothecin Micelles as Combination Therapy for Effective Anticancer Activity. *RSC Adv.* **2019**, *9* (2), 1055–1061. <https://doi.org/10.1039/C8RA08794F>.
- (23) Nanjo, T.; De Lucca, E. C.; White, M. C. Remote, Late-Stage Oxidation of Aliphatic C-H Bonds in Amide-Containing Molecules. *J. Am. Chem. Soc.* **2017**, *139* (41), 14586–14591. <https://doi.org/10.1021/jacs.7b07665>.
- (24) Deng, T.; Mao, X.; Li, Y.; Bo, S.; Yang, Z.; Jiang, Z. X. Monodisperse Oligoethylene Glycols Modified Propofol Prodrugs. *Bioorganic Med. Chem. Lett.* **2018**, *28* (22), 3502–3505. <https://doi.org/10.1016/j.bmcl.2018.10.009>.
- (25) Tao, X.; Liu, K.; Li, W.; Zhang, A. Remarkable Structure Effects on Thermoresponsive Properties of Dendritic Macromolecules. *Polymer* **2014**, *55* (16), 3672–3679. <https://doi.org/10.1016/j.polymer.2014.06.010>.

## Chapter 5 – Synthesis of new PyOx-2-PCA probes for ligand-directed chemistry

Chapter 5 focusses on the second set of probes synthesised throughout the PhD for ligand-directed chemistry: 2-pyridinecarboxyaldehydes (2-PCAs). This species has shown recently to be an interesting group for protein modification due to its high selectivity for *N*-termini. In this chapter, after an introduction of the characteristic and reactivity of 2-PCAs, we will describe how alkyne functionalised 2-PCA, and compatible azide functionalised PyOx probes were synthesised to form, to date, the first ligand-directed 2-PCA ligand.

### 5.1 An introduction to 2-PCA probe

#### 5.1.1 Chemoselective approach, the use of natural amino acids

The first use of 2-pyridinecarboxyaldehydes (2-PCA) for modification of native proteins was achieved in 2015,<sup>1</sup> and has the potential to drastically impact the field of protein modification due to its precise single site targeting of protein *N*-termini. The intrinsic chemical properties of the *N*-terminus (only  $\alpha$ -amido amine group of a protein, lower  $pK_b$  (~8) versus that for the  $\epsilon$ -amino group of Lys (~10), generally more solvent accessible) makes this group a very attractive target for modification.<sup>2-4</sup> While a number of chemistries have been developed to target the *N*-terminus, discussed in the Introduction chapter, they previously relied on specific amino acid residues such as serine,<sup>5,6</sup> through subsequent oxime ligation; cysteine, through native chemical ligation with thioesters,<sup>7,8</sup> or tryptophans via Pictet-Spengler reactions.<sup>9,10</sup> The use of 2-pyridinecarboxyaldehydes was innovative in that it was not dependent on the *N*-terminal residue, having being tested on a broad range of peptides and proteins under mild conditions, thus promising a great step forward in the chemical biology field. In this thesis we wanted to develop a new approach for protein modification using 2-PCAs as an intermediate for ligand-directed chemistry in order to modify specifically a nucleophilic residue in proximity to the *N*-terminus of the proteins of interest.

2-PCA have been shown to react with protein *N*-termini, forming a mixture of imidazolidinone diastereomers after cyclisation to the  $\alpha$ -amide, see Figure 1, under standard conditions for protein modification (50 mM phosphate buffer at pH 7.5 and 37 °C). This cyclisation takes place as a result of the favourable steric arrangement of the  $\alpha$ -amide of the peptide backbone toward the electrophilic *N*-terminal imine species **B** initially formed, resulting in a stable 5-membered ring, **C**.<sup>1</sup> For lysine residues there is no amide nearby the  $\epsilon$ -amine, and the intermediate imine is rapidly hydrolysed and so this chemistry is highly selective for the *N*-terminus. The high reactivity of 2-PCA is also attributed to the nitrogen in the pyridine ring which can be involved in the cyclisation as a catalytic base. The

selectivity toward the *N*-terminus was also validated by doing a negative control when the *N*-terminal amino group was either blocked as pyroglutamate or when proline was the second amino acid residue (removing the amide NH group that is required for cyclization; see Figure 1), in which no modification was observed. The chemistry was shown to be reliable on diverse proteins such as aldolase, lysozyme, BSA with conversion greater than 40%.

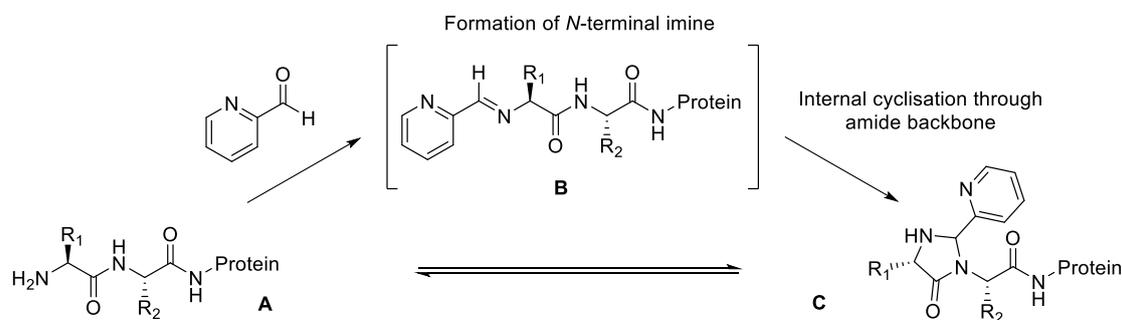


Figure 1: Scheme representing the modification of the *N*-terminus of a protein using a 2-PCA probe through a two-step process

Though a powerful method for site-selective modification, it has been shown that under incubation at 37 °C a 20–30% decrease of modified protein after 12 h was observed, indicating that 2-PCA modification is not stable.<sup>11</sup> While this is an issue for long-term *N*-terminal modification, we hypothesised that this instability could be advantageous for the goals of this PhD. Indeed, we envisaged synthesising a new bifunctional probe containing a 2-PCA head, and a tail which would be the probe for ligand-directed chemistry, as shown in Figure 2. After modification of the *N*-terminus with the 2-PCA, to form **D**, a NASA species could be added to form compound **E**, which would be then converted to **F** when a near-by nucleophile would react with the activated carboxylic acid. Finally, after finding optimal conditions for cleaving the 2-PCA, compound **G** would be obtained with a free *N*-terminus able to potentially react a second time. Following this pathway, single site modification would be achieved in close proximity to the *N*-terminus of the protein.

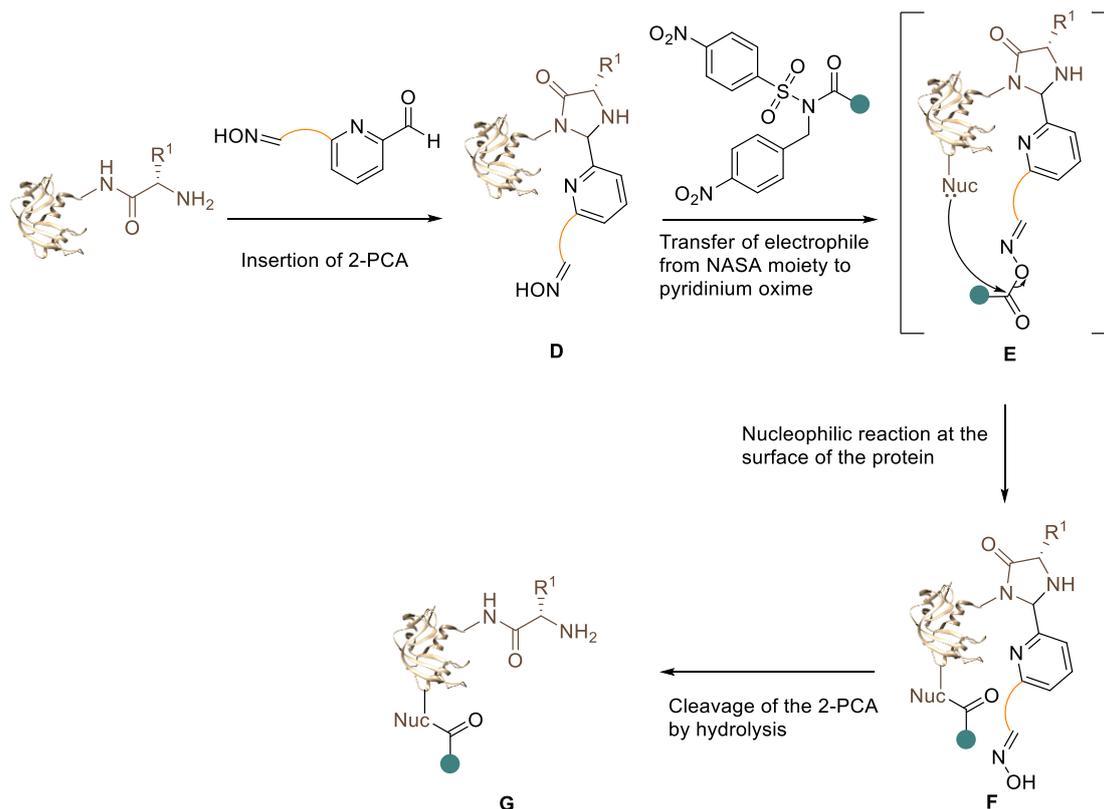


Figure 2: Proposed mechanism for 2-PCA-mediated ligand directed protein modification.

### 5.1.2 Building blocks for 2-PCA head

The simplest core of a 2-PCA species, 2-pyridinecarboxaldehyde, as shown in Figure 3, has no other groups than the one necessary for the reaction with the *N*-terminus (the aldehyde and pyridine part). Thus, to generate 2-PCA probes which can be used for various purposes, other commercially available starting materials have been used in literature, 2,6-pyridinedimethanol **5.1** and 6-(bromomethyl)-2-pyridinemethanol **5.2**, both represented in Figure 3.

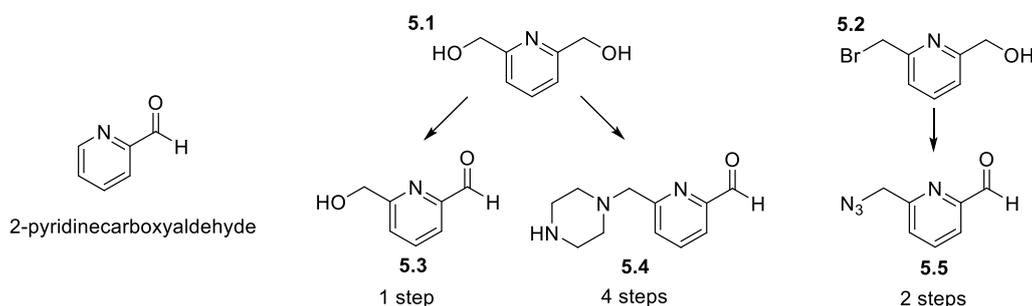


Figure 3: Known procedures in literature for the formation of functionalised 2PCA heads

Starting from 2,6-pyridinedimethanol, **5.1**, the Francis group showed that they could form two compounds **5.3** and **5.4**.<sup>1</sup> The first was obtained in one step consisting of an oxidation using SeO<sub>2</sub>, then after another three steps including the addition of a piperazine **5.4** could be obtained. The piperazine linker was used as it contains a secondary amine for functionalization and a tertiary amine for

increased water solubility. Both probes, **5.3** and **5.4**, can be further modified in a similar approach using either the hydroxyl or the amine group in a nucleophilic reaction with different useful electrophiles.

Starting from 6-(bromomethyl)-2-pyridinemethanol, **5.2**, Dimitrov's group formed a click-2PCA probe by first transforming the bromide to an azide, and only then oxidising the alcohol into the desired aldehyde, **5.5**. Later on, this chemistry was exploited for protein immobilization on polyacrylamide hydrogels,<sup>12</sup> or as an intermediate for click chemistry to modify antibodies.<sup>13</sup>

We considered that using click chemistry would be an advantage to make a library of compounds in a more modular fashion with, for example, variable type and length of linkers. Thus, the idea was to make an azide/alkyne functionalised PCA which could react with a compatible functionalised PyOx probe, see Figure 4.

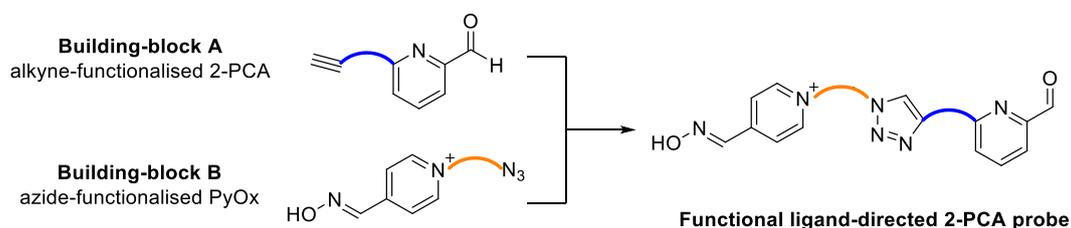


Figure 4: Proposed building blocks for the synthesis of functional ligand-directed 2-PCA

PyOx azide probe could be obtained in a two-step process based on literature precedent.<sup>11</sup> Thus, an alkyne functionalised PCA was needed. Having the starting material **5.1** in the lab, **5.3** was synthesised in a slightly lower yield than literature, 54-69% compared to 84% after purification. This compound could then react with a wide range of linkers to install an alkyne tail.

The following paragraphs will present the synthesis of the linkers and then the library of alkyne functionalised PCA / azide-functionalised PyOx synthesised.

## 5.2 Synthesis of alkyne-2-PCA probes

The first part of this Chapter will detail the synthesis of the building block A, the alkyne functionalised 2-PCA.

### 5.2.1 A direct coupling on 2-PCA to an alkyne probe

The reactive alcohol of **5.6** was directly used as a handle for  $S_N2$  reactions with different alkyne linkers to react later on with the azido-PyOx building blocks, as discussed in the previous paragraph. Two linker lengths were chosen in order to compare their solubility and also to potentially affect the location of reactions on protein surfaces later down the line, with a PEG linker giving more flexibility compared to the modification of **5.3** with propargyl bromide, Figure 5.

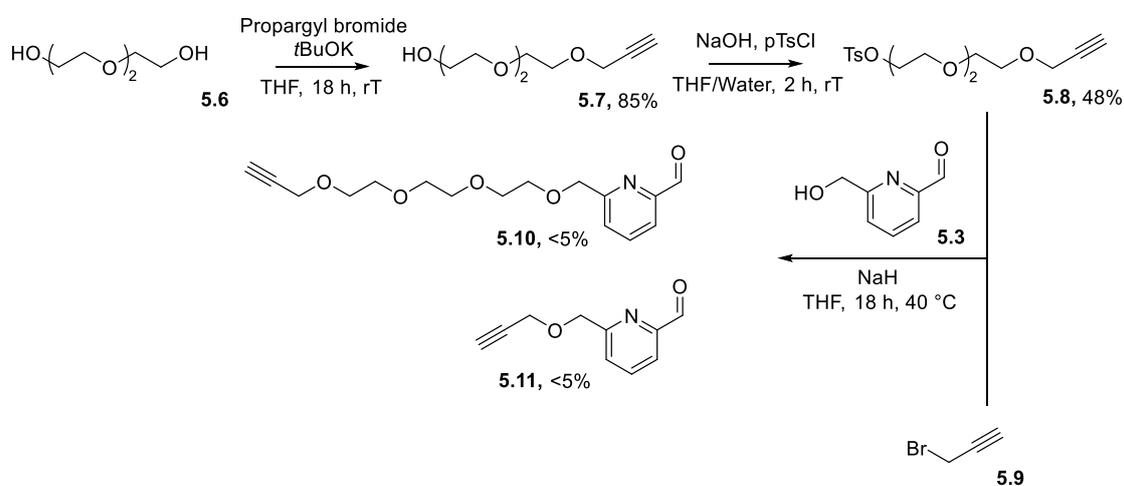


Figure 5: Scheme for the insertion of alkyne onto a 2-PCA head

The activated tosylate PEG alkyne **5.8** was produced in two steps in reasonable yields (85% and 48% respectively). Full conversion was not observed during the initial alkylation step as also double modification was observed and some starting material was recovered too. Despite using an excess of tosyl chloride during the second step only 48% of product **5.8** was recovered after purification. It is possible that either more equivalents or a change of base to activate better the hydroxyl of the PEG or a longer reaction time would have increased the yield. The last step to form both **5.10** and **5.11** from compound 6-(hydroxymethyl)-2-pyridinecarboxaldehyde **5.3** was unsuccessful, despite the use of sodium hydride as a strong base.

We considered that the free aldehyde of **5.3** may have been affecting the reaction, and so we decided to protect it and reattempt the synthesis.

### 5.2.2 The need to introduce protecting group to the 2-PCA

Two options were present in literature<sup>14,15</sup> to protect the aldehyde using either a cyclic acetal or a dimethyl acetal protecting group. The two reactions were tried to see which one would give the

highest conversion rate and also which one would be easier to be deprotected later on. Both protections were a one-step process using para-tolylsulphonic acid hydrate and either trimethyl orthoformate for compound **5.12** or diethylene glycol for compound **5.13**, Figure 6.

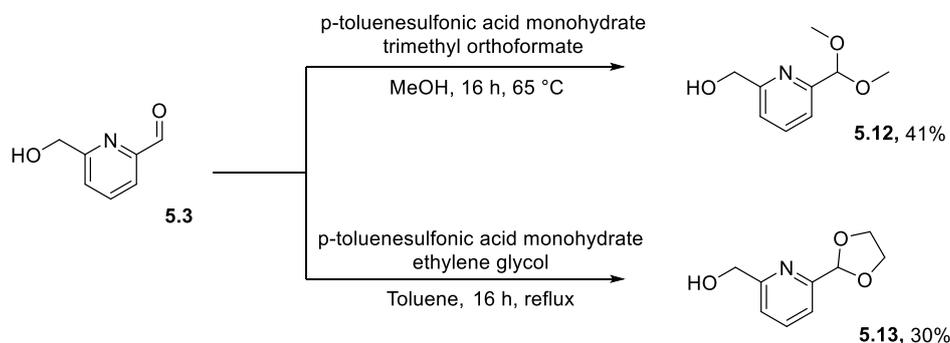


Figure 6: Scheme for the two versions of protected 2-PCA, **5.12** and **5.13**

Both conversions to the protected aldehydes **5.12** and **5.13** were reasonable with yields of 41% and 30%. In both cases an aqueous work-up was done and it is possible that the product was partly soluble in the aqueous phase and could have resulted in a partial loss of product, as for each reaction the starting material was fully consumed.

Following the protection, the  $S_N2$  reactions to insert the alkyne were performed. The conditions were kept the same as previously used with 1.1 equivalents of activated alkyne for both the propargyl and ethylene glycol-linked chains. While 1.5 equivalents of sodium hydride was used for the propargyl bromide, **5.9**, 3 equivalents were needed to push the reaction to completion for the ethylene-glycol linker. Large differences between the two linkers were obtained, with yields respectively of 39% and 91% for compounds **5.14** and **5.15** respectively as shown in Figure 7. However, even with only 39% for compound **5.14**, it was clear that protecting the aldehyde from **5.3** to **5.13**, impacted considerably the reaction.

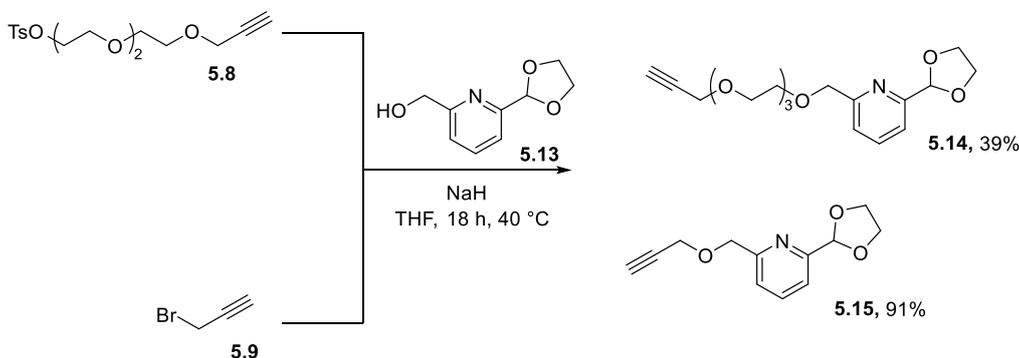


Figure 7: Scheme for the modification of diacetal 2-PCA protected head for the insertion of alkyne

Similarly, the reactions were done with the dimethylacetal protected version of the 2-PCA **5.12**. Due to the results obtained above, the tetraethylene glycol alkyne was synthesised as the bromide **5.17** in

an attempt to improve conjugation yields as shown in Figure 8. Starting from tetraethyleneglycol the  $S_N2$  with propargyl bromide was achieved in the same conditions as earlier on and gave compound **5.16** in a slightly lower yield (60% instead of 85%). The hydroxyl group of **5.16** was then converted to the equivalent bromide **5.17** using triphenylphosphine ( $PPh_3$ ) and *N*-bromosuccinimide (NBS) in 43% yield. Then compound **5.17** was reacted with **5.12** with 3 equivalents of NaH, following optimised conditions from the cyclic acetal reactions. The main compound formed was not identified and only 15% of desired product **5.18** was isolated, which was greater than when the aldehyde was unprotected (**5.10**, <5%) but lower than for the cyclic acetal version (**5.14**, 39%). However, all these species were obtained in enough quantity for further use without the need for further optimisation.

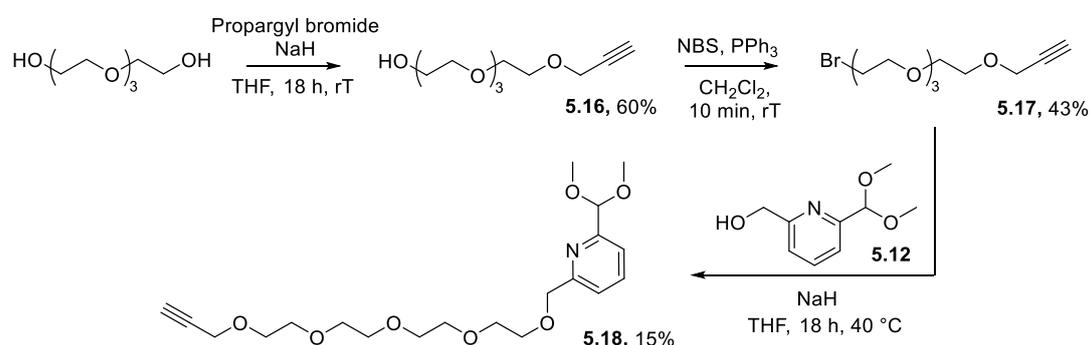


Figure 8: Scheme showing the modification of a PEG linker to a better leaving group for the modification of 2-PCA dimethyl acetal protected head

### 5.3 Synthesis of azide-PyOx probes

This second part will focus on the synthesis of the building block B, the azide functionalised PyOx.

#### 5.3.1 Synthesis of azido-diPyOx

Following literature procedures,<sup>16</sup> azido-diPyOx **5.22.a**, was synthesised in a two-step process from the commercially available 1,3,5-tris(bromomethyl)benzene, **5.19**, as shown in Figure 9.

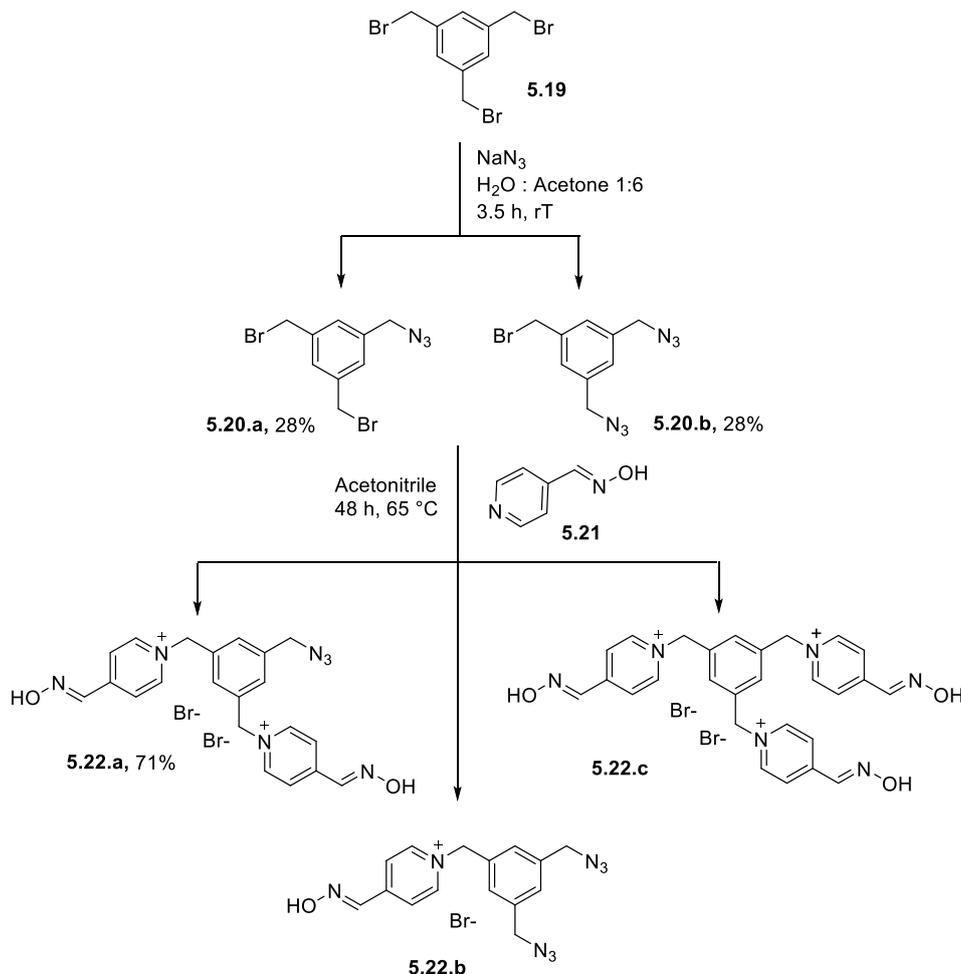


Figure 9: Scheme following literature process for the formation of a 2PyOX- $\text{N}_3$  probe

The first step led to a mixture of unmodified starting material, the desired compound **5.20.a**, and the bi-modified azide **5.20.b**. While purification was attempted and the ratio of **5.20.b** decreased, all three compounds eluted too closely to be separated. It was thus decided to carry forward with the mixture for a first attempt to react with pyridine-4-aldoxime **5.21** in acetonitrile at  $65^\circ\text{C}$  for 40 h. Three species were obtained after precipitation, the desired product, **5.22.a**, but also the bis-azido-mono-PyOx, **5.22.b**, coming from **5.20.b**, and the triPyOx, compound **5.22.c**, without any azide coming from the original starting material **5.19**. These three species were highly water soluble and so were purified via reverse phase column chromatography but although it was possible to separate the di-azido-mono-

PyOx **5.22.b**, the other two compounds **5.22.a** and **5.22.c** co-eluted during the earliest stages of the chromatography and could not be separated successfully.

On a second attempt, only one equivalent of sodium azide was used relative to the starting material **5.19**, and added stepwise to limit the formation of the di-azide, **5.20.b**. The mixture was this time purified in 10:90 CH<sub>2</sub>Cl<sub>2</sub>:PET which slowed down the elution. The fractions collected showed still a mixture with 90% of the wanted compound **5.20.a** and 10% being the di-azide **5.20.b**, considerable purer than the first attempt with only 50% of **5.20.a** in the fractions collected. Moreover, pure **5.20.b** was still obtained in a significant amount with 28% of the total yield. The second step using pyridine-4-aldoxime, **5.21**, was carried out again and only compounds **5.22.a** and **5.22.b** were obtained, which were easily separated by reverse phase chromatography to yield the pure product **5.22.a**.

### 5.3.2 Synthesis of azido-monoPyOx

Based on the literature procedures known for synthesising di-PyOx **5.22.a**, an analogue having only one PyOx group was synthesised in order to compare the reactivity between a monovalent or divalent catalyst. While Albada's group has shown that differences in reactivity are minor at higher pH (pH 8), the modification of thrombin at pH 7.2 using a divalent PyOx versus a monovalent aptamer led to an increase in protein modification of up to 5-fold.<sup>16</sup> Moreover they interestingly showed that modification was happening at different residues depending on the valency of the catalyst used. Thus, using the same core design as for the divalent catalyst, compound **5.25** was obtained from the starting material  $\alpha,\alpha'$ -dibromo-m-xylene, **5.23**, as shown in Figure 10.

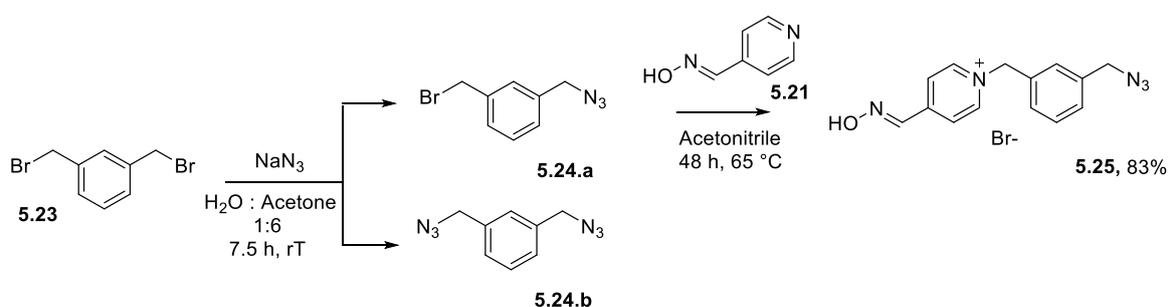


Figure 10: Scheme adapted from a literature process for the formation of a monovalent PyOX-N<sub>3</sub> probe **5.25**

1.5 equivalents of sodium azide was used to convert starting material **5.23** to the desired product **5.24.a** and some double modification product **5.24.b** was also obtained. These two species had very similar R<sub>f</sub> and after the column both were still present in a 1:1 ratio. The mixture was still used for the next step as **5.24.b** would not react with **5.21** and only the desired product would form a salt and therefore be prone to precipitation in organic solvents, which happened to give **5.25** in an overall yield of 36%.

## 5.3.3 Synthesis of PEG-linker-PyOx

Following the results discussed above, the majority of the synthetic steps led to good yields with over 50% conversion, except for the addition of a PEG linker onto the 2-PCA head (such as **5.14** or **5.18**). Moreover, preparation of the reactive PEG species always required more synthesis steps as described in previous paragraphs than using only the propargyl bromide which seemed to give the highest yield. It was therefore considered to redesign the synthetic route to introduce the PEG linker onto the second building block, the PyOx tail, as shown in Figure 11 with a protocol inspired from literature.<sup>17</sup>

This was achieved using the commercially available 11-azido-3,6,9-trioxaundecanol **5.26** in presence of sodium hydride as a base to activate, and the same starting blocks, **5.19** and **5.23** for PyOx synthesis which would undergo a  $S_N2$  type of reaction. Both mono **5.27** or di **5.28** substituted species were obtained in similar yields after purification with respectively 39% and 43% yield. Subsequent reaction with pyridine-4-aldoxime in acetonitrile at reflux gave the di-substituted species **5.30** after 20 hrs and the mono **5.29** after 48 h, as shown in Figure 11.

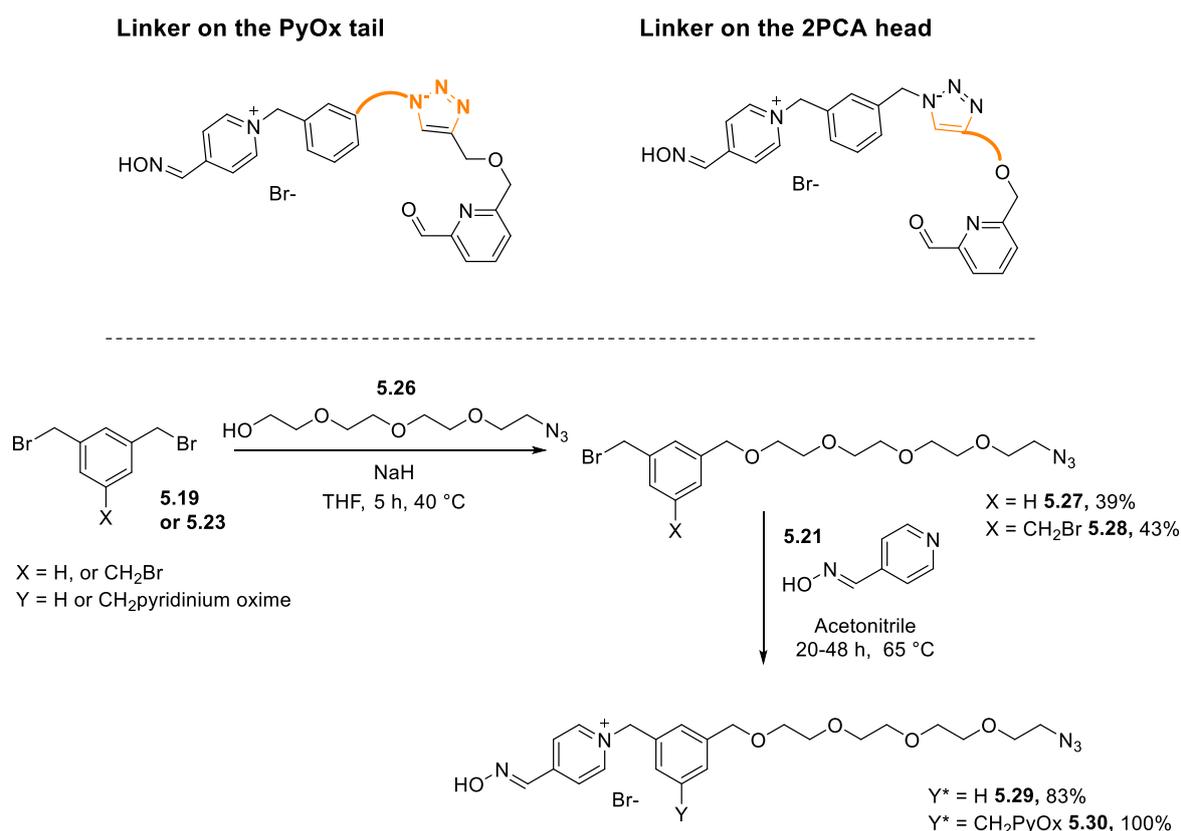


Figure 11: Scheme for the synthesis of PyOx probe with a PEG linker, PyOX-PEG4-N3 and 2PyOx-PEG4-N3. \*The yields are NMR yields

While **5.30** did precipitate in acetonitrile when the mixture was cooled down at room temperature at the end of the reaction like the other PyOx synthesised so far, **5.29** did not being an oil instead. This is likely due to the presence of the PEG linker increasing the solubility in acetonitrile. It was therefore

challenging to purify the product away from left over of pyridine-4-aldoxime. Reverse-phase chromatography was attempted on the crude mixture of sample **5.29** to separate the starting material **5.21** from the product, but due to similar polarity the two species co-eluted. The mixture was therefore used directly in the next step click-chemistry as the starting material **5.21** would be unreactive under click-chemistry conditions. In the future to avoid the presence of contaminating pyridine-4-aldoxime starting material, the excess could be decreased from 2.5 to 1.25 per bromine.

Despite this excess of starting material which could be optimised in future, this new route of building the PEG linker onto the PyOx tail compared to having the linker on the 2-PCA head, seemed to be simpler in terms of reactivity but gave higher yield than the attempt to put a PEG linker on the 2-PCA (<30%). However, it would be interesting to see if the position of the linker and thus the triazole impact the stability and reactivity of the new compounds synthesised in the future.

## 5.4 Click-chemistry and deprotection

To generate the final PyOx-2-PCA probes, two options were then possible from the two building blocks synthesised in the previous paragraphs. Either the compounds could be used straight away for click chemistry with acetal deprotection being undertaken after, as shown in route A Figure 12, or the order could be reversed with deprotection then followed by click-chemistry, as shown in route B. Once again to ensure the best conversion, both routes were attempted.

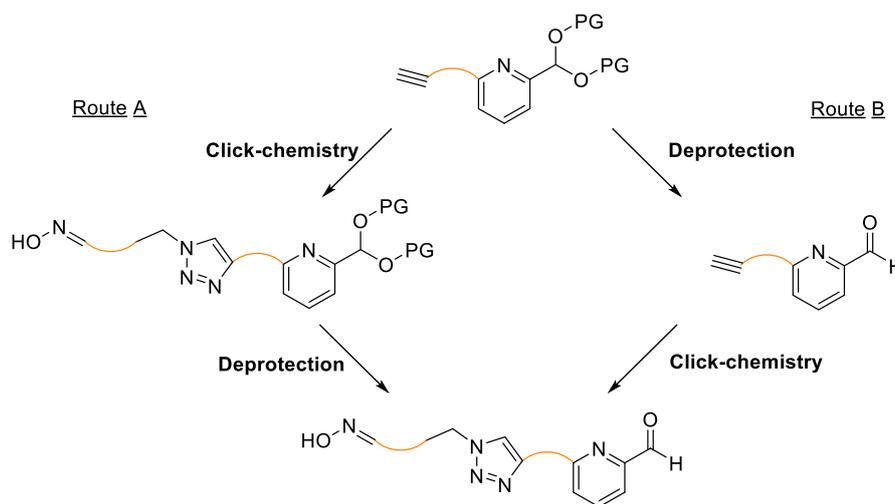


Figure 12: Two different pathways to obtain final 2-PCA pyridinium oxime probes. Route A: click chemistry then deprotection; route B: deprotection then click chemistry

### 5.4.1 Click-chemistry, background and choice of reagents

Click-chemistry was first introduced by the Sharpless group in 1999 and has been since developed to encompass a broad range of reaction types, with his first article on the subject cited more than 10 000 times.<sup>18</sup> In this PhD only the classical copper-catalysed Huisgen 1,3-dipolar cycloaddition of azides and terminal alkynes will be used which was later given the name **Cu alkyne-azide cycloaddition (CuAAC)**. This chemistry has shown to be wide in scope, stereospecific, and high yielding, becoming widely used in biochemistry due to its ease. The general mechanism of CuAAC chemistry is shown in Figure 13.

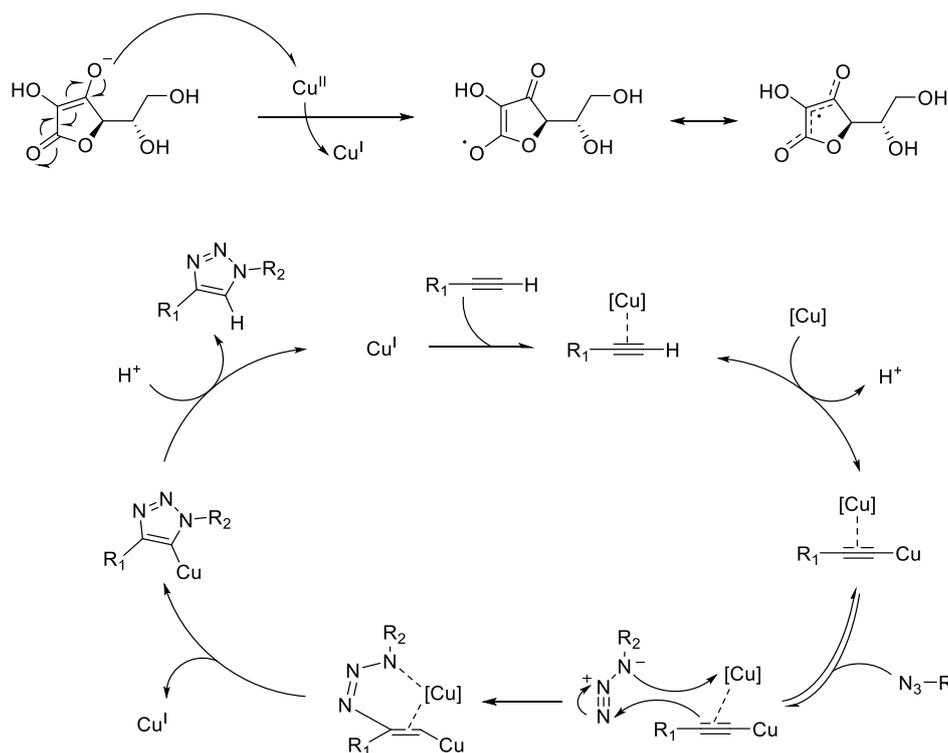


Figure 13: Mechanism of Cu activation with sodium ascorbate as the reducing agent, and mechanism of the Copper-catalyzed Azide-Alkyne Cycloaddition (CuAAC).<sup>19</sup>

The use of Cu(I) also presents the major advantages of being inexpensive and easy to handle with most of the protocols involving the *in-situ* reduction of stable sources of Cu(II), such as  $\text{CuSO}_4$ .<sup>20</sup>

#### 5.4.2 Route A: click-chemistry before deprotection

While many protocols have been reported, a first attempt at click chemistry was done using conditions inspired by the literature<sup>21</sup> due to the resemblance of its reported chemistry to the 2-PCA head, see Figure 14.

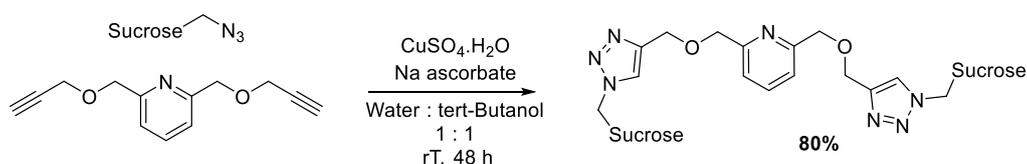


Figure 14: Inspiration for click-chemistry conditions from literature

1.2 equiv. of azide was chosen to convert fully the alkyne, as an excess of azide could be removed via reverse phase chromatography very easily due to its high solubility in water. Moreover, 3 equivalents of sodium ascorbate and 1.5 of  $\text{CuSO}_4$  were chosen, see full protocol in the Experimental. While  $\text{CuSO}_4$  is generally introduced in a catalytic amount and not in excess, this was not an issue for a first attempt working on small scale (10-50  $\mu\text{mol}$ ), but these parameters could be optimised at a later stage. These

conditions were applied to compound **5.25** with the two cyclic acetal protected 2-PCAs, **5.14** and **5.15**, resulting in full conversion after only 2 h, to form products **5.31** and **5.32**, Figure 15.

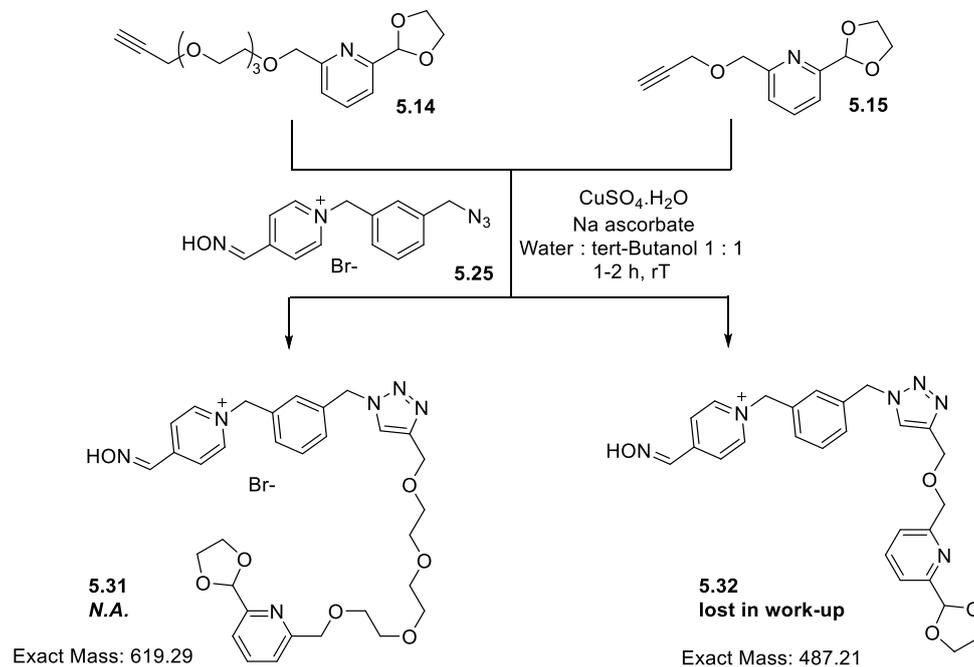


Figure 15: First attempt of click-chemistry with PyOx azide probe **5.25** on 2-PCA **5.14** and **5.15**

While in literature precedent that we based our conditions on, an aqueous work-up was done and the product was then recovered from the organic phase, the product **5.32** was initially lost in the aqueous due to its highly hydrophilic character. We therefore tried to purify the crude mixture of **5.31** using reverse phase chromatography straight after the 2 h of reaction without any other purification steps. This was successful, with LC-MS revealing clear separation between the product and other impurities. The following step was to deprotect the cyclic acetal. A first trial used impure fractions from the purification of compound **5.31** to save valuable material, adding HCl at different concentrations (6 M HCl or 1 M HCl). The solution was then stirred at 40 °C and 20  $\mu\text{L}$  aliquots were taken over time to follow the acetal deprotection by LCMS. As shown in Figure 16, the starting material peak disappeared over 9 hrs, shown by the blue line, and while no new UV peak was observed within the chromatogram, the spectrogram at the same retention time, rt 130 sec, showed the apparition of a new peak with a mass that was higher than we expected. The mass of this new species, Figure 17, matched the mass of the dimethyl acetal, **5.31.b**, forming from potential reaction with methanol. Generally, though, it seemed as if the material was degrading with a loss of signal intensity, with less than 10% of the original intensity by 4 hr.

Due to these unpromising results, it was decided to attempt the alternative route, trying to deprotect the aldehyde first prior to the click chemistry.

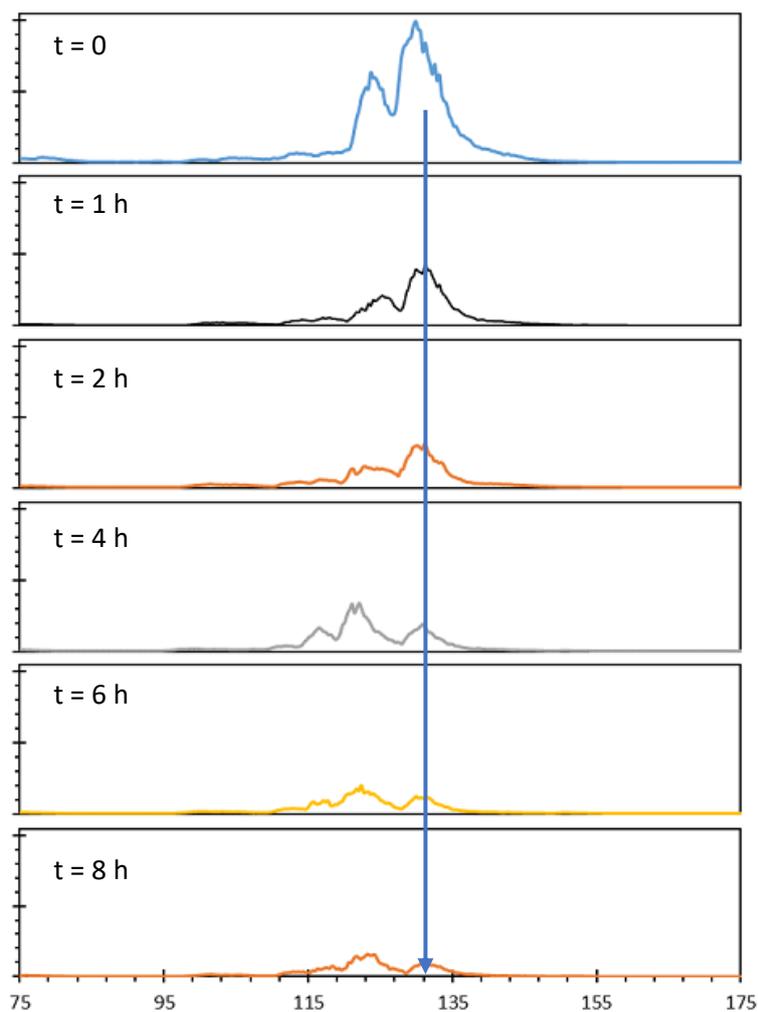


Figure 16: LCMS trace following the deprotection of diacetal **5.31**. Blue arrow shows the disappearance of compound **5.31** peak.

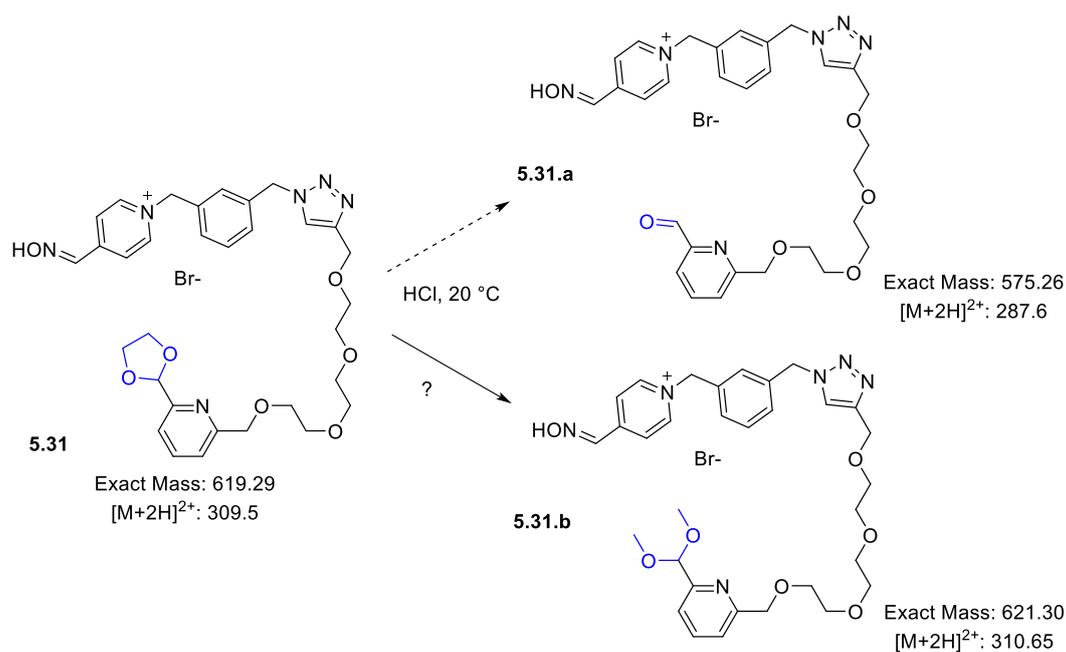
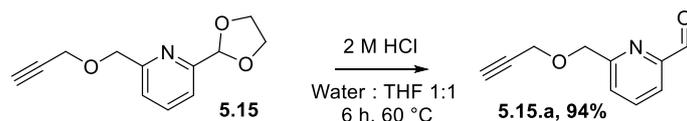


Figure 17: Structure and mass of the desired product **5.31.a** from the hydrolysis of compound **5.31** and hypothesized side-product **5.31.b**

## 5.4.3 Route B: deprotection before click-chemistry

Following the results of the last section and the use of most of compound **5.14**, we attempted to find good conditions for the deprotection of **5.15** to its aldehyde analogue **5.15.a**, Figure 18, which could subsequently be used with all types of azide functionalised PyOx probes. After trying a series of acidic conditions for deprotection followed by LCMS, the optimal conditions were found to be the use of 2M HCl in a 1:1 mixture of water and THF at 60 °C for 6 h, which resulted in full conversion to the species **5.15.a** after work up without the requirement of further purification.

Figure 18: Scheme for the deprotection of **5.15** to its aldehyde analogue

The conditions developed in the previous section for click chemistry (protocol in Experimental section), were then applied to **5.15.a** allowing the formation of the four final probes **5.33**, **5.34**, **5.35**, **5.36** as shown in Figure 19, which were purified on reverse phase chromatography (see Experimental paragraph), leading to the only examples of PyOx-2PCA probes developed to date.

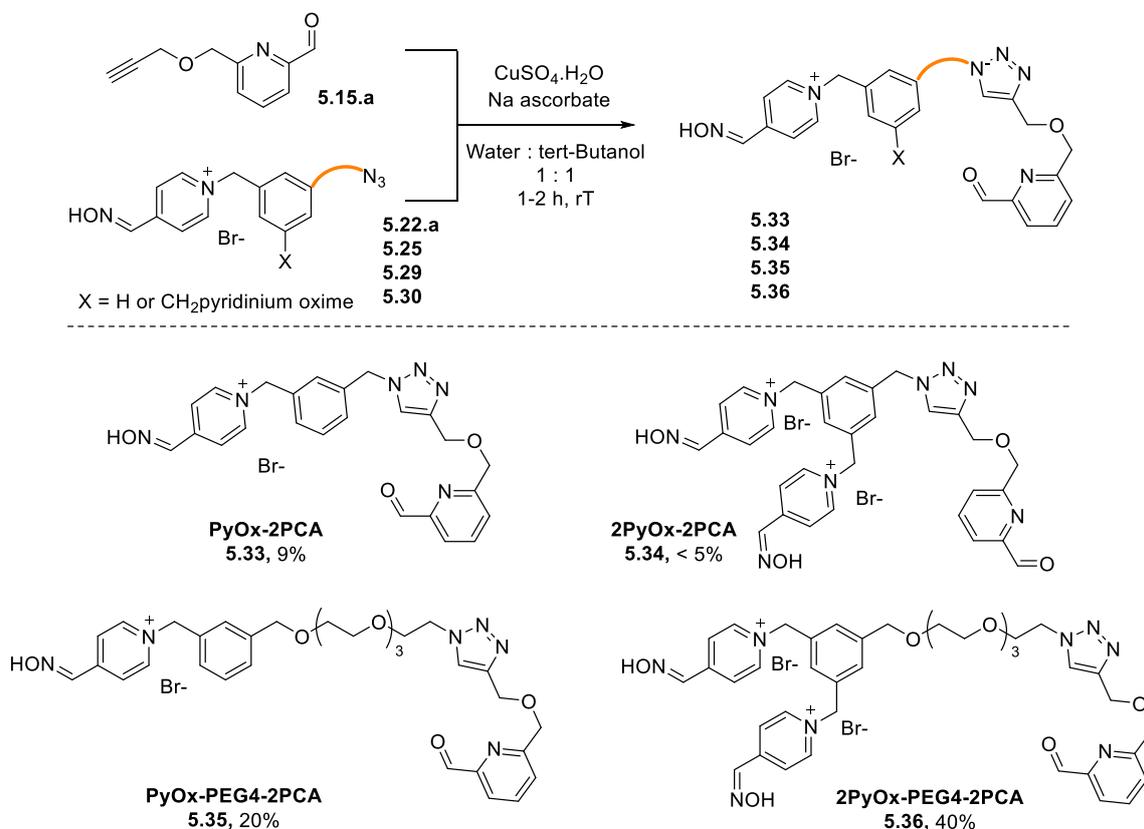


Figure 19: Final probes obtained after click chemistry with pyridinium oxime catalyst head and 2-PCA tails for protein modification

## 5.5 Summary

From this chapter, the main conclusion drawn is that 2PCA head can be modified using click-chemistry to insert a pyridinium oxime tail in relatively few steps. Optimisation of the chemistry allowed us to develop routes in which all the steps were relatively high yielding with the majority of them having a conversion rate greater than 60%. Moreover it is the first time that this approach of coupling 2-PCA and pyridinium oxime has been used, leading to a new class of probes which will be used for ligand-directed chemistry and protein modification as will be discussed in Chapter 6. With the synthetic route in place, the approach could also be broadened to a lot of other potential modifications for a wide range of applications. Figure 20 represents a generic scheme of the final best conditions obtained for each step, in order to synthesize PyOx-2PCAs.

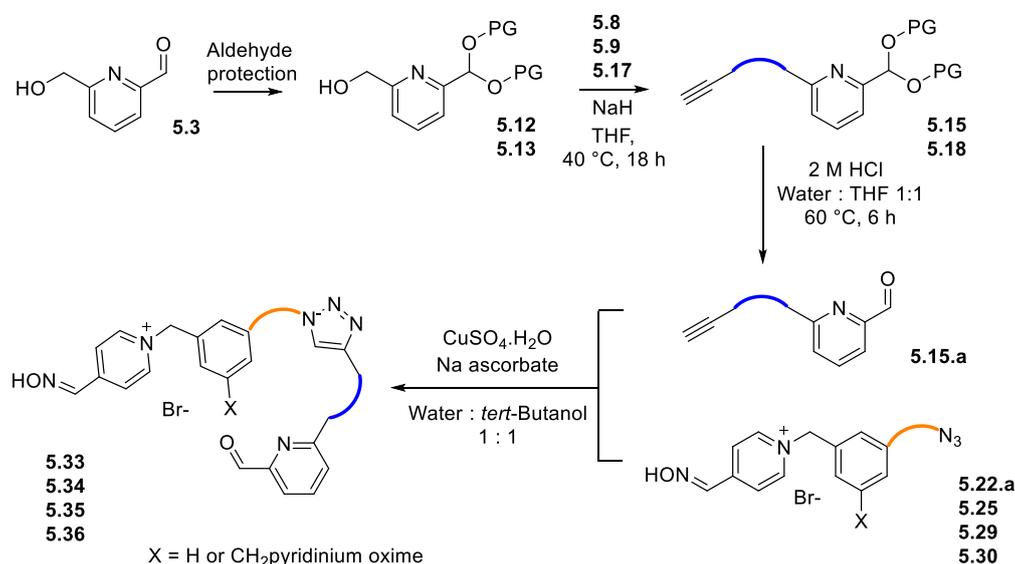
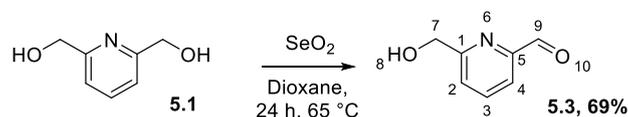


Figure 20: General scheme for the formation of PyOx-2PCAs probe from starting material 5.3

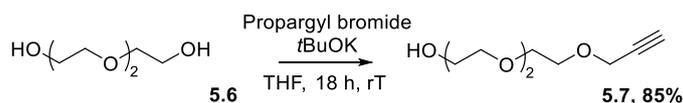
## 5.6 Experimental

## 5.6.1 Chemical synthesis

**Synthesis of 2-PCA head****Compound 5.3**

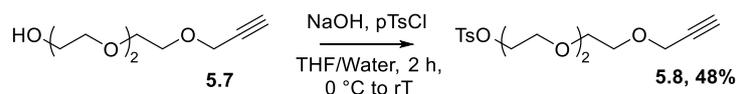
To a solution of 2,6-pyridinedimethanol, **5.1**, (3.0 g, 22 mmol) in 1,4-dioxane (60 ml) was added selenium dioxide (0.95 g, 8.6 mmol). The resulting mixture was sonicated for 2 min and then stirred at 65 °C, turning red after 30 min. After 21 h, the reaction was cooled to room temperature and diluted with CH<sub>2</sub>Cl<sub>2</sub> (40 mL). The mixture was filtered through Celite, washed with CH<sub>2</sub>Cl<sub>2</sub> (40 mL) and the filtrate was concentrated under reduced pressure. The resulting crude material was purified by flash chromatography using a gradient of 1-20% MeOH in CH<sub>2</sub>Cl<sub>2</sub>. Pure fractions were concentrated under reduced pressure to afford **5.3** as a pale viscous oil (2.1 g, 15 mmol, 69%). Data were consistent with those previously reported.<sup>1</sup>

<sup>1</sup>H NMR (400 MHz, Chloroform-*d*): δ 10.08 (s, 1H, -CH aldehyde), 7.92-7.84 (m, 2H, -ArH<sub>2</sub> and -ArH<sub>4</sub>), 7.51 (m, 1H, -ArH<sub>3</sub>), 4.87 (d, *J* = 3.5 Hz, 2H, -CH<sub>2</sub>), 3.54 (br s, 1H, -OH). HRMS (ESI<sup>+</sup>): *m/z* calcd for C<sub>7</sub>H<sub>8</sub>NO<sub>2</sub>: 138.0550 [M+H]<sup>+</sup>; found: 138.0546

**Compound 5.7**

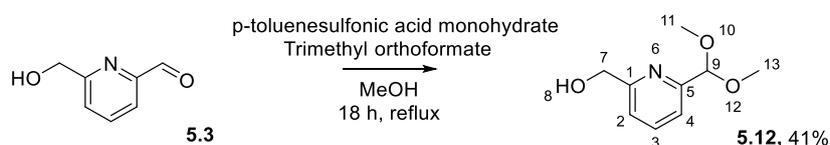
To a suspension of potassium tert-butoxide (0.75 g, 6.7 mmol) in dry THF (40 mL) was added tri(ethylene)glycol (1.2 g, 8.0 mmol) under nitrogen atmosphere. The reaction mixture was allowed to stir at room temperature for 10 min then propargyl bromide (80% in toluene, 0.7 mL, 6.3 mmol) was added dropwise. The resulting mixture was allowed to stir at room temperature for 17 h. The mixture was then diluted with THF (40 mL) and filtered through Celite®. The filtrate was concentrated under reduced pressure and the crude mixture was purified by flash chromatography using a gradient of 50-100% EtOAc in PET. Pure fractions were concentrated under reduced pressure to afford **5.7** as a pale-yellow oil (1.00 g, 5.32 mmol, 85%). Data were consistent with those previously reported.<sup>22</sup>

<sup>1</sup>H NMR (400 MHz, Chloroform-*d*): 4.19 (dd, *J* = 2.4, 0.8 Hz, 2H, -OCH<sub>2</sub>CCH), 3.74-3.64 (m, 10H, -CH<sub>2</sub> PEG), 3.61-3.57 (m, 2H, -CH<sub>2</sub> PEG), 2.42 (t, *J* = 2.4 Hz, 1H, -CCH). HRMS (ESI<sup>+</sup>): *m/z* calcd for C<sub>9</sub>H<sub>16</sub>NaO<sub>4</sub>: 211.0941 [M+Na]<sup>+</sup>; found: 211.0941

**Compound 5.8**

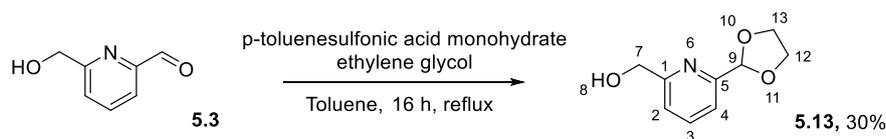
A solution of **5.7** (1.0 g, 5.3 mmol) in a mixture of THF (10 mL) and aqueous NaOH (1.6 M, 10 mL, 16.3 mmol) was added to a solution of TsCl (1.4 g, 7.35 mmol) in THF (15 mL) at 0 °C. After, the mixture was stirred warming up to room temperature for 2 h, and CH<sub>2</sub>Cl<sub>2</sub> (100 mL) was then added. The mixture was washed with water (50 mL), the organic layer was collected, dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated under pressure. The crude product was purified by flash chromatography using a gradient of 50-100% EtOAc in PET. Pure fractions were concentrated under reduced pressure to afford the pure product **5.8** (0.87 g, 2.5 mmol, 48%) as an oil. Data were consistent with those previously reported.<sup>23</sup>

**<sup>1</sup>H NMR** (400 MHz, Chloroform-*d*): δ 7.79 (d, *J* = 8.3 Hz, 2H, -(CH<sub>2</sub>)<sub>2</sub>CSO<sub>2</sub>), 7.33 (d, *J* = 8.3 Hz, 2H, -(CHCH)<sub>2</sub>CSO<sub>2</sub>), 4.18 (dd, *J* = 2.4, 0.5 Hz, 2H, -OCH<sub>2</sub>CCH), 4.16-4.13 (m, 2H, -CH<sub>2</sub>OTs), 3.70-3.61 (m, 6H, -CH<sub>2</sub> PEG), 3.58 (d, *J* = 0.5 Hz, 4H, -CH<sub>2</sub>OCH<sub>2</sub>CH<sub>2</sub>OCH<sub>2</sub>CCH), 2.43 (s, 3H, -CH<sub>3</sub>), 2.41 (td, *J* = 2.4, 0.5 Hz, 1H, -CCH). **HRMS** (ESI<sup>+</sup>): *m/z* calcd for C<sub>16</sub>H<sub>22</sub>NaO<sub>6</sub>S: 365.1029 [M+Na]<sup>+</sup>; found: 365.1030

**Compound 5.12**

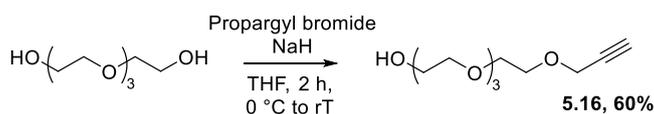
A solution of **5.3** (0.70 g, 5.11 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (700 μL) was combined with a catalytic amount of p-toluenesulfonic acid monohydrate (120 mg, 0.63 mmol) and trimethyl orthoformate (3.0 ml, 27.4 mmol) in MeOH (15 mL). This mixture was then refluxed for 18 h. After cooling the mixture to room temperature and removing most of the solvent under reduced pressure the green oil was purified by flash chromatography using a gradient of 30-50% EtOAc in PET. Pure fractions were concentrated under reduced pressure to afford the final product **5.12** as a transparent oil (380 mg, 0.41 mmol, 41%).

**<sup>1</sup>H NMR** (400 MHz, MeOD): δ 7.85 (m, 1H, -ArH<sub>3</sub>), 7.53-7.48 (m, 1H, -ArH<sub>4</sub>), 7.47-7.42 (m, 1H, -ArH<sub>2</sub>), 5.34-5.27 (m, 1H, -CH(OCH<sub>3</sub>)<sub>2</sub>), 4.68 (d, *J* = 5.2 Hz, 2H, -CH<sub>2</sub>), 3.37-3.31 (m, 6H, CH(OCH<sub>3</sub>)<sub>2</sub>). **<sup>13</sup>C NMR** (101 MHz, MeOD): δ 160.75 (C1), 156.19 (C5), 137.70 (C3), 120.43 (C2), 119.74 (C4), 103.58 (C9), 64.13 (C7), 52.57 (C13, C11). **HRMS** (ESI<sup>+</sup>): *m/z* calcd for C<sub>9</sub>H<sub>13</sub>NNaO<sub>3</sub>: 206.0788 [M+Na]<sup>+</sup>; found: 206.0783

**Compound 5.13**

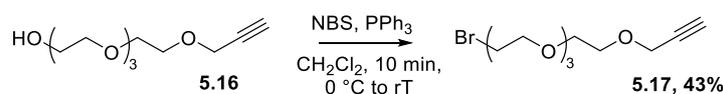
Ethylene glycol (2.8 mL, 51 mmol), p-toluenesulfonic acid monohydrate (0.16 g, 0.84 mmol) and **5.3** (1.0 g, 7.3 mmol) were refluxed in toluene (60 mL) under Dean-Stark conditions. The mixture was stirred for 18 h before letting the solution cool down to room temperature. After removing the solvent under reduced pressure, the crude mixture was dissolved in EtOAc (50 mL) and washed twice with saturated sodium bicarbonate (2 x 30 mL) and then twice with brine (2 x 30 mL). The organic phase was then dried on MgSO<sub>4</sub>, filtered and the solvent removed under reduced pressure. The crude mixture was then purified by flash chromatography using a gradient of 75-100% EtOAc in PET. Pure fractions were concentrated under reduced pressure to afford the pure product **5.13** as a yellow oil (400 mg, 2.21 mmol, 30%).

**<sup>1</sup>H NMR** (400 MHz, Chloroform-*d*): δ 7.73 (t, *J* = 7.7 Hz, 1H, -ArH<sub>3</sub>), 7.45 (d, *J* = 7.7 Hz, 1H, -ArH<sub>4</sub>), 7.25 (dd, *J* = 7.7, 0.7 Hz, 1H, -ArH<sub>2</sub>), 5.86 (s, 1H, -CH(OCH<sub>2</sub>)<sub>2</sub>), 4.77 (d, *J* = 0.7 Hz, 2H, -CH<sub>2</sub>OH), 4.22-4.02 (m, 4H, -CH(OCH<sub>2</sub>)<sub>2</sub>). **<sup>13</sup>C NMR** (101 MHz, Chloroform-*d*): δ 158.91 (C1), 156.17 (C5), 137.59 (C3), 120.89 (C2), 119.24 (C4), 103.65 (C9), 65.71 (C13, C12), 64.18 (C7). **HRMS** (ESI<sup>+</sup>): *m/z* calcd for C<sub>9</sub>H<sub>11</sub>NNaO<sub>3</sub>: 204.0631 [M+Na]<sup>+</sup>; found: 204.0630

**Compound 5.16**

Following the same protocol as compound **5.8**, the pure compound was obtained as a pale-yellow oil (1.24 g, 5.34 mmol, 60%). Data were consistent with those previously reported.<sup>24</sup>

**<sup>1</sup>H NMR** (400 MHz, Chloroform-*d*): δ 4.19 (d, *J* = 2.3 Hz, 2H, -OCH<sub>2</sub>CCH), 3.74-3.63 (m, 14H, --CH<sub>2</sub> PEG), 3.61-3.57 (m, 2H, -CH<sub>2</sub> PEG), 2.76 (t, *J* = 6.2 Hz, 1H, -OH), 2.41 (t, *J* = 2.3 Hz, 1H, -CCH). **HRMS** (ESI<sup>+</sup>): *m/z* calcd for C<sub>11</sub>H<sub>20</sub>NaO<sub>5</sub>: 255.1203 [M+Na]<sup>+</sup>; found: 255.1194

**Compound 5.17**

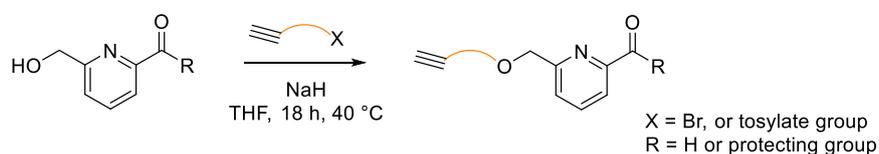
A solution of **5.16** (200 mg, 0.86 mmol) in dry CH<sub>2</sub>Cl<sub>2</sub> (5 mL) at 0 °C under N<sub>2</sub> atmosphere was treated with triphenylphosphine (350 mg, 1.33 mmol) and stirred at 0 °C for 5 minutes. NBS (240 mg,

1.35 mmol) was then added and the mixture was stirred at 0 °C for another 20 minutes then warmed to room temperature for 1 h. The reaction was quenched with H<sub>2</sub>O (20 mL) and extracted with CH<sub>2</sub>Cl<sub>2</sub> (20 mL). The organic layer was dried over anhydrous Na<sub>2</sub>SO<sub>4</sub>, filtered, and concentrated under pressure. The residue was purified by flash chromatography using 33% EtOAc in PET. Pure fractions were concentrated under reduced pressure to give the desired product **5.17** as a pale-yellow liquid (171 mg, 0.58 mmol, 43 %). Data were consistent with those previously reported.<sup>25</sup>

<sup>1</sup>H NMR (400 MHz, Chloroform-*d*): δ 4.21 (dd, *J* = 2.4, 0.5 Hz, 2H, -OCH<sub>2</sub>CCH), 3.81 (t, *J* = 6.3 Hz, 2H, -CH<sub>2</sub>Br), 3.72-3.64 (m, 12H, -CH<sub>2</sub> PEG), 3.47 (t, *J* = 6.3 Hz, 2H, -CH<sub>2</sub>CH<sub>2</sub>Br), 2.43 (dd, *J* = 2.4, 0.5 Hz, 1H).

HRMS (ESI<sup>+</sup>): *m/z* calcd for C<sub>11</sub>H<sub>19</sub>BrNaO<sub>4</sub>: 317.0359 [M+Na]<sup>+</sup>; found: 317.0360

#### **General procedure for alkyne addition to 2-PCAs**



**2-PCA** (1 eq) was dissolved in anhydrous THF (5 mL) and NaH (1.5 or 3 eq) was added slowly. After 5 min, a solution of **alkyne** (1.05 eq) in THF (2 mL) was added. The mixture was heated to 40 °C and stirred overnight. After cooling down, water (10 mL) was used to quench the reaction and the compound was extracted with EtOAc (3 x 20 mL). The combined organic phases were washed with brine (10 mL), dried over MgSO<sub>4</sub>, and filtered. The solvent was removed under reduced pressure to afford a crude yellow oil which was purified if necessary by flash chromatography, to afford the desired compound as an oil.

#### **Compound 5.10**

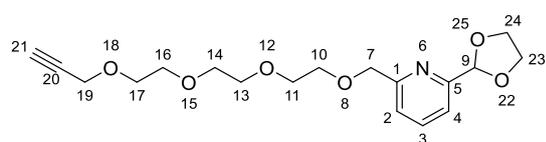
Less than <5%. Not purified. Not further analysed.

#### **Compound 5.11**

Less than <5%. Not purified. Not further analysed.

#### **Compound 5.14**

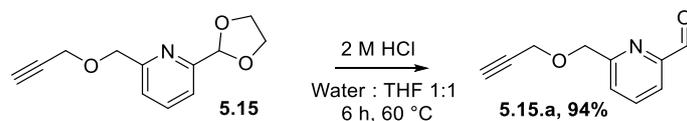
Purified by flash chromatography using a gradient of 75-100% EtOAc in PET.



(170 mg, 0.48 mmol, 44%)

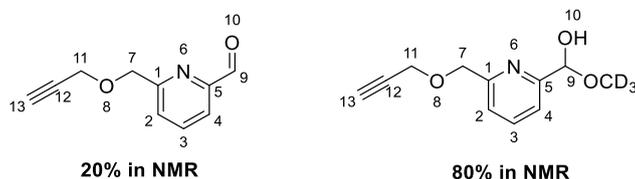
<sup>1</sup>H NMR (400 MHz, Chloroform-*d*): 7.74 (t, *J* = 7.7 Hz, 1H, -ArH<sub>3</sub>), 7.49 (d, *J* = 7.7 Hz, 1H, -ArH<sub>4</sub>), 7.43 (d, *J* = 7.7 Hz, 1H, -ArH<sub>2</sub>), 5.80 (s, 1H, -CH(OCH<sub>2</sub>)<sub>2</sub>), 4.71 (s, 2H, -NCCH<sub>2</sub>), 4.19 (dd, *J* = 2.4, 0.6 Hz, 2H, -



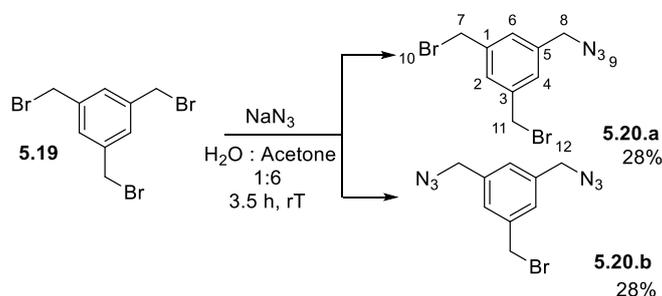
**Deprotection of 2-PCA head****Compound 5.15.a**

**5.15** (20 mg, 91  $\mu\text{mol}$ ) was dissolved in THF (1.5 mL) and aqueous HCl (4 M, 1.5 mL). The solution was then heated at 60 °C for 1.5 h and aliquots (20  $\mu\text{L}$ ) of the reaction were taken to be analysed by LCMS. After 6 h, full conversion was reached. The solvents were removed under reduced pressure to afford **5.15.a** as a yellow oil (15 mg, 86  $\mu\text{mol}$ , 94%) which was stored under Ar in the freezer without further purification.

In MeOD mixture between the two following species:



$^1\text{H NMR}$  (400 MHz, MeOD- $d_4$  NMR data is provided for the hemiacetal):  $\delta$  7.90-7.82 (m, 1H, -Ar3H), 7.56-7.50 (m, 1H, -Ar2H), 7.49-7.42 (m, 1H, -Ar4H), 5.50 (d,  $J = 5.2$  Hz, 1H, -C9H), 4.69 (d,  $J = 5.0$  Hz, 2H, -C7H<sub>2</sub>), 4.28 (dd,  $J = 2.2$  Hz, 2H, -C11H<sub>2</sub>), 2.92-2.86 (m, 1H, -CCH);  $^{13}\text{C NMR}$  (101 MHz, MeOD- $d_4$ ):  $\delta$  158.48 (C5), 156.71 (C1), 138.31 (C3), 121.56 (C4), 119.54 (C2), 97.20 (C9), 78.90 (C12), 75.13 (C13), 71.25 (C7), 57.60 (C11). **HRMS** (ESI<sup>+</sup>):  $m/z$  calcd for C<sub>10</sub>H<sub>10</sub>NO<sub>2</sub>: 176.0706 [M+H]<sup>+</sup>; found: 176.0710

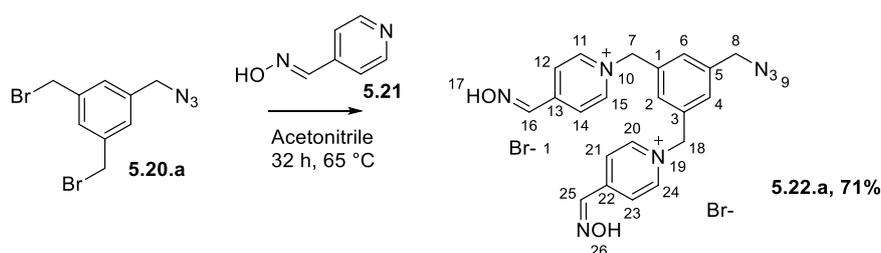
**Synthesis of PyOx tail****Compound 5.20**

To a solution of 1,3,5-tris(bromomethyl)benzene, **5.19**, (1.0 g, 2.8 mmol) in 6:1 acetone/H<sub>2</sub>O (30 mL in total) was added sodium azide (0.18 g, 2.8 mmol) portionwise. The reaction mixture was stirred for 3.5 h at room temperature. The mixture was then diluted with CH<sub>2</sub>Cl<sub>2</sub> (10 mL), and the organics washed with H<sub>2</sub>O (3 x 10 mL) and brine (10 mL), dried over Na<sub>2</sub>SO<sub>4</sub>, filtered, and concentrated under reduced

pressure. The crude material was purified by flash chromatography using a 10% CH<sub>2</sub>Cl<sub>2</sub> in PET. Pure fractions were concentrated under reduced pressure to afford the product **5.20.a** as an oil (220 mg, 0.79 mmol, 28%). Data were consistent with those previously reported.<sup>26</sup>

**<sup>1</sup>H NMR** (400 MHz, Chloroform-*d*): δ 7.38 (d, *J* = 1.7 Hz, 1H, -ArH<sub>2</sub>), 7.27 (d, *J* = 1.7 Hz, 2H, -ArH<sub>4</sub> and <sub>6</sub>), 4.46 (s, 4H, -CH<sub>2</sub>Br), 4.36 (s, 2H, -CH<sub>2</sub>N<sub>3</sub>). **HRMS** (APCI<sup>+</sup>): *m/z* calcd for C<sub>9</sub>H<sub>10</sub>Br<sub>2</sub>N: 289.9175[M+H]<sup>+</sup>; found: 289.9167 (Compound ionised as the amine, rather than azide)

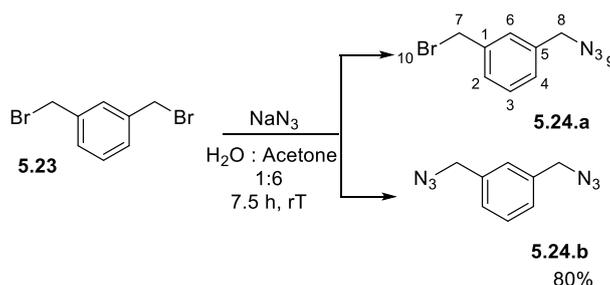
### Compound 5.22.a



**5.20.a** (400 mg, 1.13 mmol) and pyridine-4-aldoxime, **5.21**, (703 mg, 5.76 mmol) were dissolved in MeCN (10 mL) and the mixture was stirred at 65 °C for 32 h. The mixture was cooled down and the precipitate recovered by filtration. The precipitate was washed with MeCN (5 mL) and dried under vacuum. The crude mixture was purified by reverse-phase column chromatography using a gradient of 0-100% MeCN in Water. Pure fractions were concentrated under reduced pressure to afford the product **5.22.a** as a pasty yellow solid (450 mg, 0.80 mmol, 71%).

**<sup>1</sup>H NMR** (400 MHz, D<sub>2</sub>O): δ 8.88-8.76 (m, 4H, -ArH<sub>11, 15, 20, 24</sub>), 8.36 (d, *J* = 1.3 Hz, 2H, -CH<sub>16, -CH25</sub>), 8.19 (dd, *J* = 7.1, 1.3 Hz, 4H, -ArH<sub>12, 14, 21, 23</sub>), 7.53 (d, *J* = 1.7 Hz, 2H, -ArH<sub>4</sub> and <sub>6</sub>), 7.49 (d, *J* = 1.7 Hz, 1H, -ArH<sub>2</sub>), 5.84 (s, 4H, -CH<sub>2</sub>N<sup>+</sup>), 4.46 (s, 2H, -CH<sub>2</sub>N<sub>3</sub>). **<sup>13</sup>C NMR** (101 MHz, D<sub>2</sub>O): δ 149.48 (C13 or C22), 149.34 (C13 or C22), 146.21 (C16 or C25), 146.15 (C16 or C25), 144.81 (C12/C14 or C21/C23), 144.70 (C12/C14 or C21/C23), 135.64 (C1 and C3), 134.77 (C5), 130.29 (C6 or C4), 129.98 (C6 or C4), 129.06 (C2), 125.08 (C11, C15, C20, C24), 63.38 (C7 or C18), 63.05 (C7 or C18), 53.29 (C8). **HRMS** (ESI<sup>+</sup>): *m/z* calcd for C<sub>21</sub>H<sub>20</sub>N<sub>7</sub>O<sub>2</sub>: 402.1673 [M+H]<sup>+</sup>; found: 402.1668

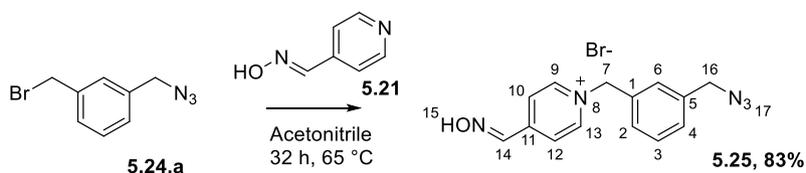
### Compound 5.24



To a stirred solution of  $\alpha,\alpha'$ -dibromo-*m*-xylene, **5.23**, (5.0 g, 18.9 mmol, 1 equiv.) in 6:1 acetone/H<sub>2</sub>O (35 mL in total) was added sodium azide (1.8 g, 28.4 mmol, 1.5 equiv.). The reaction mixture was stirred for 7.5 h at room temperature. The mixture was diluted with CH<sub>2</sub>Cl<sub>2</sub> (30 mL), and the organics washed with H<sub>2</sub>O (20 mL) and brine (20 mL), dried over Na<sub>2</sub>SO<sub>4</sub>, filtered, concentrated under reduced pressure. The resulting crude material was purified by flash chromatography using 10% EtOAc in PET. Pure fractions were concentrated under reduced pressure to afford the product as a mixture of compound **5.24.a** and diazido-*m*-xylene **5.24.b** (50%/50%) as a transparent liquid (1.9 g, 8.2 mmol, 43%). Data were consistent with those previously reported.<sup>27</sup>

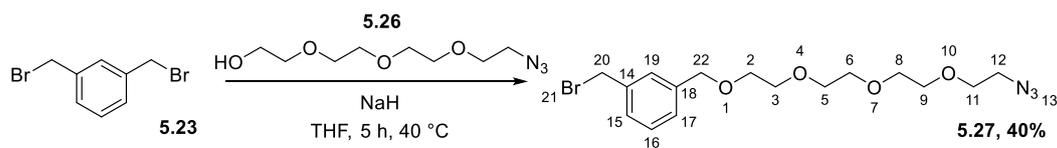
**<sup>1</sup>H NMR** (400 MHz, Chloroform-*d*):  $\delta$  7.42-7.25 (m, 8H, -CH<sub>arom</sub>), 4.49 (s, 2H, -CH<sub>2</sub>Br), 4.47 (s, 2H, -CH<sub>2</sub>Br), 4.36 (s, 2H, -CH<sub>2</sub>N<sub>3</sub>), 4.34 (s, 2H, -CH<sub>2</sub>N<sub>3</sub>). **HRMS** (APCI<sup>+</sup>): *m/z* calcd for C<sub>8</sub>H<sub>9</sub>BrN: 197.9913 [M+H]<sup>+</sup>; found: 197.9925

### Compound 5.25



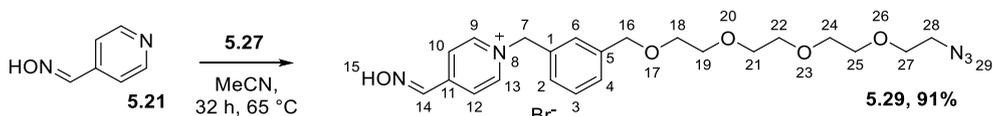
1-(azidomethyl)-3-(bromomethyl)benzene, **5.24.a**, (750 mg, 3.33 mmol) and pyridine-4-aldoxime, **5.21**, (800 mg, 6.56 mmol) were dissolved in MeCN (5 mL) and the mixture was stirred at 65 °C for 32 h. The mixture was cooled down and the precipitate recovered by filtration. The precipitate was washed with MeCN (5 mL) and dried under vacuum. The crude mixture was purified by reverse-phase column chromatography using a gradient of 0 to 100% MeCN in Water. Pure fractions were concentrated under reduced pressure to afford the product **5.25** as a white solid (450 mg, 1.30 mmol, 39%).

**<sup>1</sup>H NMR** (400 MHz, D<sub>2</sub>O):  $\delta$  8.83 (dd, *J* = 6.5, 1.5 Hz, 2H, -ArH<sub>9,13</sub>), 8.33 (d, *J* = 1.5 Hz, 1H, -CHNOH), 8.18-8.11 (m, 2H, -ArH<sub>10,12</sub>), 7.55-7.41 (m, 4H, -ArH benzyl), 5.79 (m, 2H, -CH<sub>2</sub>N<sup>+</sup>), 4.41 (d, *J* = 1.5 Hz, 2H, -CH<sub>2</sub>N<sub>3</sub>). **<sup>13</sup>C NMR** (101 MHz, D<sub>2</sub>O):  $\delta$  146.25 (C<sub>14</sub>), 144.67 (C<sub>10</sub> or C<sub>12</sub>), 144.56 (C<sub>10</sub> or C<sub>12</sub>), 130.13 (C benzyl), 129.90 (C benzyl), 129.28 (C benzyl), 128.97 (C benzyl), 128.81 (C benzyl), 125.00 (C<sub>9</sub> or C<sub>13</sub>), 124.98 (C<sub>9</sub> or C<sub>13</sub>), 63.62 (C<sub>7</sub>), 53.75 (C<sub>16</sub>). **HRMS** (APCI<sup>+</sup>): *m/z* calcd for C<sub>14</sub>H<sub>14</sub>N<sub>3</sub>O: 240.1131 [M+H]<sup>+</sup>; found: 240.1137

**Compound 5.27**

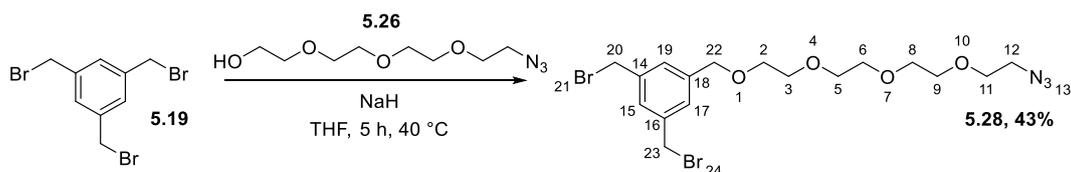
A mixture of  $\alpha,\alpha'$ -dibromo-*m*-xylene, **5.23**, (280 mg, 1.06 mmol) and sodium hydride (60% suspension in mineral oil, 50 mg, 1.25 mmol) in THF (4 mL) was stirred at 0 °C. A solution of 11-azido-3,6,9-trioxaundecan-1-ol, **5.26**, (150 mg, 0.68 mmol) in THF (1.5 mL) was then added to the suspension dropwise. The mixture was kept stirring at 0 °C for 2 h and then allowed to warm to room temperature for 2 h. The solvent was evaporated under reduced pressure. After addition of H<sub>2</sub>O (5 mL), the mixture was extracted with CH<sub>2</sub>Cl<sub>2</sub> (10 mL  $\times$  3). The combined organic phases were washed with H<sub>2</sub>O (5 mL) and then dried over anhydrous MgSO<sub>4</sub>, and the solvent was evaporated under reduced pressure. The crude product was purified flash chromatography using 33% EtOAc in PET. Pure fractions were concentrated under reduced pressure to afford the pure product, **5.27**, as an oil (0.40 g, 0.27 mmol, 40%).

<sup>1</sup>H NMR (400 MHz, Chloroform-*d*):  $\delta$  7.37 (s, 1H, -ArH<sub>19</sub>), 7.30 (dd,  $J = 4.7, 1.2$  Hz, 2H, -ArH<sub>15</sub> and -ArH<sub>17</sub>), 7.27 (dd,  $J_1 = J_2 = 4.7$  Hz, 1H, -ArH<sub>16</sub>), 4.55 (s, 2H, -CH<sub>2</sub>22), 4.48 (d,  $J = 1.2$  Hz, 2H, -CH<sub>2</sub>20), 3.71-3.61 (m, 14H, -CH<sub>2</sub> PEG), 3.36 (t,  $J = 5.1$  Hz, 2H, -CH<sub>2</sub>12). HRMS (ESI<sup>+</sup>):  $m/z$  calcd for C<sub>16</sub>H<sub>24</sub>BrN<sub>3</sub>NaO<sub>4</sub>: 424.0842 [M+Na]<sup>+</sup>; found: 424.0839

**Compound 5.29**

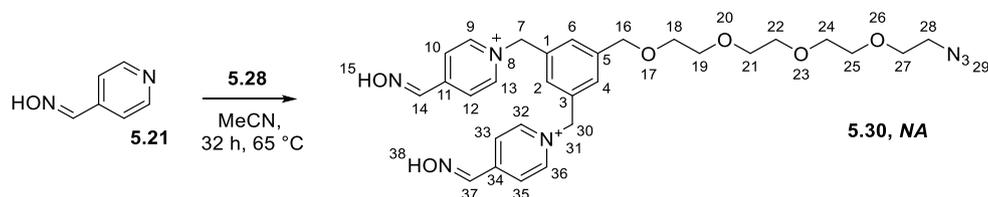
**5.27** (100 mg, 250  $\mu$ mol) and pyridine-4-aldoxime, **5.21**, (75 mg, 615  $\mu$ mol) were dissolved in MeCN (5 mL) and the mixture was stirred at 65 °C for 32 h. The mixture was cooled to room temperature and the mixture was concentrated under atmospheric pressure. The resultant oil was obtained as a 40%/60% mixture with remaining starting material **5.21** and used without further purification (81 mg, 155  $\mu$ mol, 62%).

<sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O):  $\delta$  8.72 (d,  $J = 6.8$  Hz, 2H, -ArH<sub>9, 13</sub>), 8.21 (s, 1H, -CH<sub>14</sub>), 8.03 (d,  $J = 6.8$  Hz, 2H, -ArH<sub>10, 12</sub>), 7.41-7.23 (m, 4H, -ArH benzyl), 5.65 (s, 2H, -CH<sub>2</sub>N<sup>+</sup>), 4.45 (s, 2H, -CH<sub>2</sub>O), 3.63-3.47 (m, 14H, -CH<sub>2</sub> PEG), 3.33-3.27 (m, 2H, -CH<sub>2</sub>N<sub>3</sub>). <sup>13</sup>C NMR (101 MHz, D<sub>2</sub>O)  $\delta$  149.19 (C<sub>11</sub>), 146.25 (C<sub>14</sub>), 144.58 (C<sub>10</sub>/C<sub>12</sub>), 133.05 (C<sub>q</sub> benzyl), 129.88 (C benzyl), 128.85 (C benzyl), 128.77 (C benzyl), 124.93 (C<sub>9</sub>/C<sub>13</sub>), 72.35 (C<sub>16</sub>), 69.61 (C PEG), 69.22 (C PEG), 69.06 (C<sub>27</sub>), 64.04 (C<sub>7</sub>), 50.15 (C<sub>28</sub>). HRMS (ESI<sup>+</sup>):  $m/z$  calcd for C<sub>22</sub>H<sub>30</sub>N<sub>5</sub>O<sub>5</sub>: 444.2241 [M+H]<sup>+</sup>; found: 444.2241

**Compound 5.28**

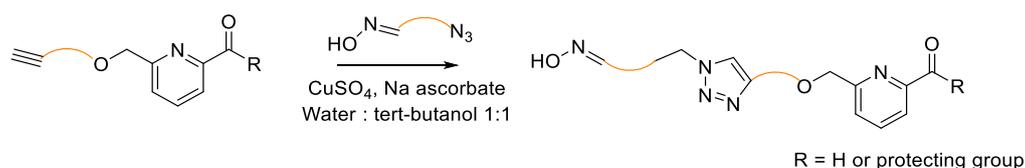
A mixture of 1,3,5-tris(bromomethyl)benzene, **5.19**, (700 mg, 0.99 mmol) and sodium hydride (60% suspension in mineral oil, 80 mg, 1.00 mmol) in dry THF (10 mL) was stirred at 0 °C. A solution of 11-azido-3,6,9-trioxaundecanol, **5.26**, (300 mg, 0.69 mmol) in THF (3 mL) was added to the suspension dropwise. The mixture was kept stirring at 0 °C for 1 h and then allowed to warm to room temperature for 3 h. The solvent was evaporated under reduced pressure. After addition of H<sub>2</sub>O (10 mL), the mixture was extracted with CH<sub>2</sub>Cl<sub>2</sub> (3 × 20 mL). The combined organic phases were dried over anhydrous MgSO<sub>4</sub>, and the solvent evaporated under reduced pressure. The crude product was purified by flash chromatography using 40% EtOAc in PET. Pure fractions were concentrated under reduced pressure to afford the pure product **5.28** as an oil (0.42 g, 0.85 mmol, 43%).

<sup>1</sup>H NMR (400 MHz, Chloroform-d): δ 7.32 (d, *J* = 1.8 Hz, 1H, -ArH<sub>15</sub>), 7.30 (d, *J* = 1.8 Hz, 2H, -ArH<sub>17</sub>, <sub>19</sub>), 4.54 (s, 2H, -CCH<sub>2</sub>O), 4.46 (s, 4H, -CH<sub>2</sub>Br), 3.72-3.61 (m, 14H, -CH<sub>2</sub> PEG), 3.36 (t, *J* = 5.0 Hz, 2H, -CH<sub>2</sub>N<sub>3</sub>). HRMS (ESI<sup>+</sup>): *m/z* calcd for C<sub>17</sub>H<sub>25</sub>Br<sub>2</sub>N<sub>3</sub>NaO<sub>4</sub>: 516.0104 [M+Na]<sup>+</sup>; found: 516.0104

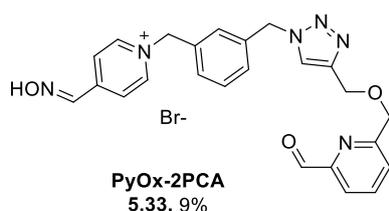
**Compound 5.30**

**5.21** (400 mg, 0.81 mmol) and pyridine-4-aldoxime (500 mg, 4.1 mmol) were dissolved in MeCN (10 mL) and the mixture was stirred at 65 °C for 32 h. The mixture was cooled down and the precipitate recovered by filtration. The precipitate was washed with MeCN (5 mL) and dried to give a pale-yellow sticky solid. The residue was allowed to dry overnight and used in the subsequent step without further purification. The product was obtained as a 85%/15% mixture with remaining starting material **5.21**.

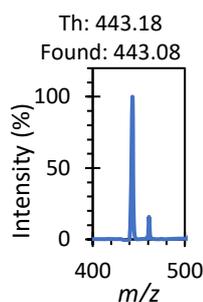
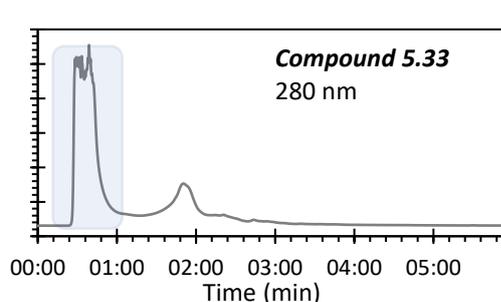
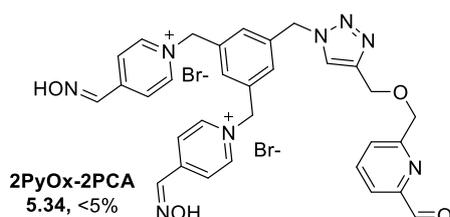
<sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O): δ 8.73-8.62 (m, 4H, -ArH<sub>9</sub>, <sub>13</sub>, <sub>32</sub>, <sub>36</sub>), 8.21 (d, *J* = 1.1 Hz, 2H, -CH<sub>14</sub> and -CH<sub>37</sub>), 8.06-7.99 (m, 4H, -ArH<sub>10</sub>, <sub>12</sub>, <sub>33</sub>, <sub>35</sub>), 7.37 (s, 2H, -ArH<sub>4</sub>, <sub>6</sub>), 7.32 (s, 1H, -ArH<sub>2</sub>), 5.67 (s, 4H, -CH<sub>2</sub>N<sup>+</sup>), 4.45 (s, 2H, -CH<sub>2</sub>O), 3.54-3.48 (m, 14H, -CH<sub>2</sub> PEG), 3.30-3.22 (m, 2H, -CH<sub>2</sub>N<sub>3</sub>). <sup>13</sup>C NMR (101 MHz, D<sub>2</sub>O): δ 149.42 (C<sub>11,34</sub>), 146.20 (C<sub>14,37</sub>), 144.65 (C<sub>10,12,33,35</sub>), 134.52 (C<sub>1,3</sub>), 129.78 (C<sub>4,6</sub>), 128.88 (C<sub>2</sub>), 124.94 (C<sub>9,13,32,36</sub>), 71.80 (C<sub>16</sub>), 69.60 (C PEG), 69.53 (C PEG), 69.20 (C PEG), 63.44 (C<sub>7,30</sub>), 50.15 (C<sub>28</sub>). HRMS (ESI<sup>+</sup>): *m/z* calcd for C<sub>29</sub>H<sub>36</sub>N<sub>7</sub>O<sub>6</sub>: 578.2722 [M+H]<sup>+</sup>; found: 578.2732

**General experimental for click-chemistry reaction**

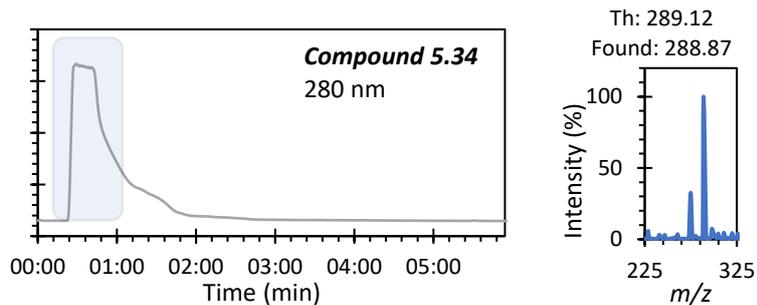
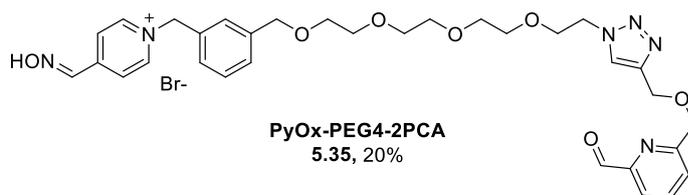
**Azide** moiety (1.2 eq) and **alkyne**-moiety (1 eq) were dissolved in a mixture of *tert*-butanol (0.5 mL) and water (0.5 mL), then copper sulfate (1.5 eq) and sodium ascorbate (3 eq) were added. The resulting mixture was stirred for 1.5-2 h. The crude mixture was analysed by LCMS to ensure that the alkyne-2PCA was fully transformed. If not complete, a further 0.5 eq of azide was added to ensure full conversion. Once fully converted, the product was purified by reverse-phase column chromatography see details in paragraph below. Pure fractions were lyophilised to give green oils. Due to the small scale of those reactions, compounds were analysed via LC-MS and high resolution MS to confirm identity.

**Compound 5.33**

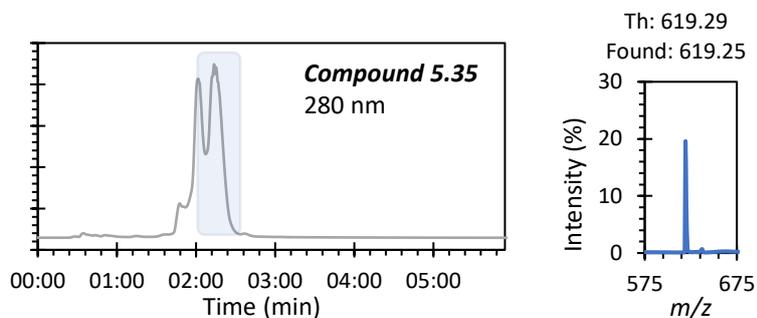
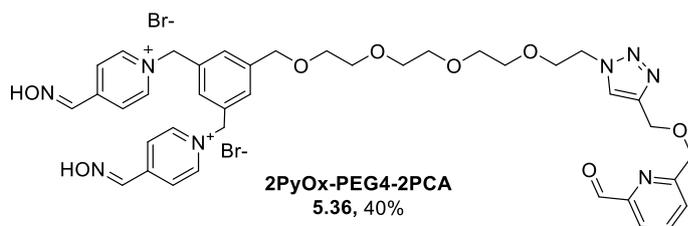
**HRMS (ESI<sup>+</sup>):**  $m/z$  calcd for C<sub>24</sub>H<sub>23</sub>N<sub>6</sub>O<sub>3</sub>: 443.1826 [M+H]<sup>+</sup>; found: 443.1838

**Compound 5.34**

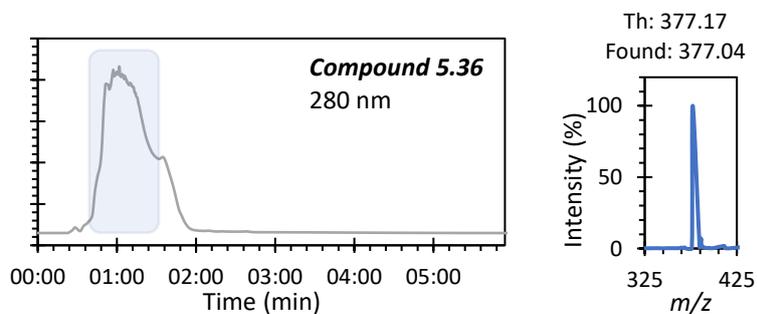
**HRMS (ESI<sup>+</sup>):**  $m/z$  calcd for C<sub>31</sub>H<sub>30</sub>N<sub>8</sub>O<sub>4</sub>: 289.1190 [M/2+H]<sup>+</sup>; found: 289.1170

**Compound 5.35**

HRMS (ESI<sup>+</sup>):  $m/z$  calcd for C<sub>32</sub>H<sub>39</sub>N<sub>6</sub>O<sub>7</sub>: 619.2875 [M+H]<sup>+</sup>; found: 619.2882

**Compound 5.36**

HRMS (ESI<sup>+</sup>):  $m/z$  calcd for C<sub>39</sub>H<sub>46</sub>N<sub>8</sub>O<sub>8</sub>: 377.1714 [M/2+H]<sup>+</sup>; found: 377.1691



### 5.6.2 Combiflash purification

Reverse-phase purification of 2-PCA probes was achieved on a Combiflash NextGen from Teledyne using C18 columns (4.3 gram RediSep Rf Reversed-phase C18 Columns – 69-2203-410 – or 5.5 gram RediSep Gold Rf Reversed-phase C18 Columns – 69-2203-328). Water (solvent A) and acetonitrile (solvent B), both containing 0.1% trifluoroacetic acid were used as the mobile phase at a flow rate of 15 mL/min. LC traces were measured via UV absorption at 214, and 254 nm with the average measured from 200-300 nm, with a signal gain of x3. The solvent gradient was programmed as shown in Figure 21. When a peak was detected, the program held the gradient for the length necessary. Every fraction was collected, with different volumes depending on the presence of a peak or not (5 mL or maximum volume).

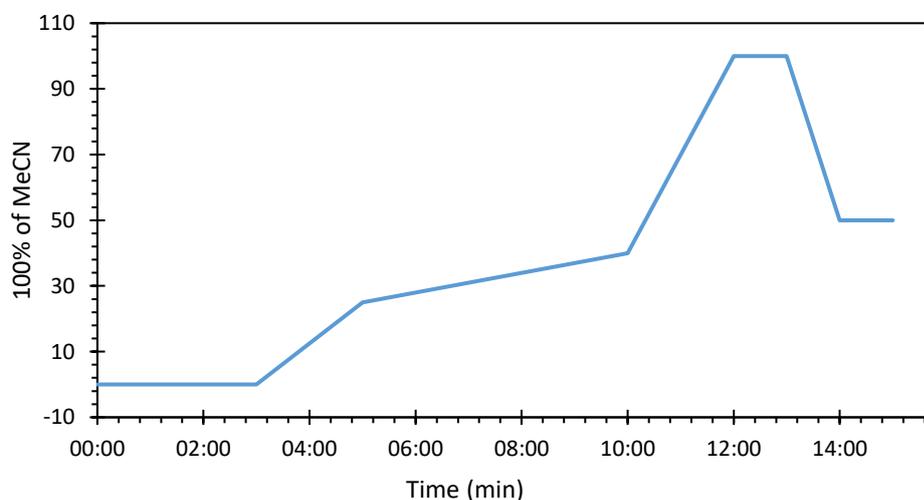


Figure 21: Percentage of MeCN + 0.1 TFA in H<sub>2</sub>O + 0.1% TFA for purification by reverse-phase chromatography on Combiflash used for purification of PyOx-2PCA probes.

## 5.7 Bibliography

- (1) Macdonald, J. I.; Munch, H. K.; Moore, T.; Francis, M. B. One-Step Site-Specific Modification of Native Proteins with 2-Pyridinecarboxaldehydes. *Nat. Chem. Biol.* **2015**, *11* (5), 326–331. <https://doi.org/10.1038/nchembio.1792>.
- (2) Boutureira, O.; Bernardes, J. L. Advances in Chemical Protein Modification. *Chem. Rev.* **2015**, *115*, 2174–2195. <https://doi.org/10.1021/cr500399p>.
- (3) Rosen, C. B.; Francis, M. B. Targeting the N Terminus for Site-Selective Protein Modification. *Nat. Chem. Biol.* **2017**, *13* (7), 697–705. <https://doi.org/10.1038/nchembio.2416>.
- (4) Tantipanjanorn, A.; Wong, M.-K. Development and Recent Advances in Lysine and N-Terminal Bioconjugation for Peptides and Proteins. *Molecules* **2023**, *28* (3), 1083. <https://doi.org/10.3390/molecules28031083>.
- (5) Ning, X.; Temming, R. P.; Dommerholt, J.; Guo, J.; Ania, D. B.; Debets, M. F.; Wolfert, M. A.; Boons, G.-J.; van Delft, F. L. Protein Modification by Strain-Promoted Alkyne-Nitrone Cycloaddition. *Angew. Chem.* **2010**, *122* (17), 3129–3132. <https://doi.org/10.1002/ange.201000408>.
- (6) Geoghegan, K. F.; Stroh, J. G. Site-Directed Conjugation of Nonpeptide Groups to Peptides and Proteins Via Periodate Oxidation of a 2-Amino Alcohol. Application to Modification at N-Terminal Serine. *Bioconjug. Chem.* **1992**, *3* (2), 138–146. <https://doi.org/10.1021/bc00014a008>.
- (7) Dawson, P.; Muir, T.; Clark-Lewis, I.; Kent, S. B. H. Synthesis of Proteins by Native Chemical Ligation. *Science (80-. )*. **1994**, *266*, 776–779. <https://doi.org/10.1126/science.7973629>.
- (8) Dawson, P. E.; Kent, S. B. H. Synthesis of Native Proteins by Chemical Ligation. *Annu. Rev. Biochem.* **2000**, *69*, 923–960. <https://doi.org/10.1146/annurev.biochem.69.1.923>.
- (9) Scheck, R. A.; Dedeo, M. T.; Iavarone, A. T.; Francis, M. B. Optimization of a Biomimetic Transamination Reaction. *J. Am. Chem. Soc.* **2008**, *130* (35), 11762–11770. <https://doi.org/10.1021/ja802495w>.
- (10) Gilmore, J. M.; Scheck, R. A.; Esser-Kahn, A. P.; Joshi, N. S.; Francis, M. B. N-Terminal Protein Modification through a Biomimetic Transamination Reaction. *Angew. Chem., Int. Ed. Engl.* **2006**, *45* (32), 5307–5311. <https://doi.org/10.1002/anie.200600368>.
- (11) Barber, L. J.; Yates, N. D. J.; Fascione, M. A.; Parkin, A.; Hemsworth, G. R.; Genever, P. G.; Spicer, C. D. Selectivity and Stability of N-Terminal Targeting Protein Modification Chemistries. *RSC Chem. Biol.* **2022**, No. i, 56–64. <https://doi.org/10.1039/d2cb00203e>.
- (12) Lee, J. P.; Kassianidou, E.; MacDonald, J. I.; Francis, M. B.; Kumar, S. N-Terminal Specific Conjugation of Extracellular Matrix Proteins to 2-Pyridinecarboxaldehyde Functionalized Polyacrylamide Hydrogels. *Biomaterials* **2016**, *102*, 268–276. <https://doi.org/10.1016/j.biomaterials.2016.06.022>.
- (13) Li, D. zhi; Han, B. nan; Wei, R.; Yao, G. yang; Chen, Z.; Liu, J.; Poon, T. C. W.; Su, W.; Zhu, Z.; Dimitrov, D. S.; Zhao, Q. N-Terminal  $\alpha$ -Amino Group Modification of Antibodies Using a Site-Selective Click Chemistry Method. *MAbs* **2018**, *10* (5), 712–719. <https://doi.org/10.1080/19420862.2018.1463122>.
- (14) Wang, F.; Tan, X.; Lv, H.; Zhang, X. New Ruthenium Complexes Based on Tetradentate Bipyridine Ligands for Catalytic Hydrogenation of Esters. *Chem. - An Asian J.* **2016**, *11* (15), 2103–2106. <https://doi.org/10.1002/asia.201600506>.
- (15) Rizeq, N.; Georgiades, S. N. Linear and Branched Pyridyl-Oxazole Oligomers: Synthesis and

- Circular Dichroism Detectable Effect on c-Myc G-Quadruplex Helicity. *European J. Org. Chem.* **2016**, 2016 (1), 122–131. <https://doi.org/10.1002/ejoc.201501269>.
- (16) Keijzer, J. F.; Firet, J.; Albada, B. Site-Selective and Inducible Acylation of Thrombin Using Aptamer-Catalyst Conjugates. *Chem. Commun.* **2021**, 57 (96), 12960–12963. <https://doi.org/10.1039/d1cc05446e>.
- (17) Chu, J. C. H.; Yang, C.; Fong, W. P.; Wong, C. T. T.; Ng, D. K. P. Facile One-Pot Synthesis of Cyclic Peptide-Conjugated Photosensitisers for Targeted Photodynamic Therapy. *Chem. Commun.* **2020**, 56 (80), 11941–11944. <https://doi.org/10.1039/d0cc05264g>.
- (18) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem., Int. Ed. Engl.* **2001**, 40, 2004–2021. [https://doi.org/10.1002/1521-3773\(20010601\)40:11<2004::AID-ANIE2004>3.0.CO;2-5](https://doi.org/10.1002/1521-3773(20010601)40:11<2004::AID-ANIE2004>3.0.CO;2-5).
- (19) Castro, V.; Rodríguez, H.; Albericio, F. CuAAC: An Efficient Click Chemistry Reaction on Solid Phase. *ACS Comb. Sci.* **2016**, 18 (1), 1–14. <https://doi.org/10.1021/acscmbosci.5b00087>.
- (20) Amblard, F.; Cho, J. H.; Schinazi, R. F. Cu(I)-Catalyzed Huisgen Azide-Alkyne 1,3-Dipolar Cycloaddition Reaction in Nucleoside, Nucleotide, and Oligonucleotide Chemistry. *Chem. Rev.* **2009**, 109 (9), 4207–4220. <https://doi.org/10.1021/cr9001462>.
- (21) Lewandowski, B.; Jarosz, S. Amino-Acid Templated Assembly of Sucrose-Derived Macrocycles. *Org. Lett.* **2010**, 12 (11), 2532–2535. <https://doi.org/10.1021/ol100749m>.
- (22) Percec, V.; Leowanawat, P.; Sun, H. J.; Kulikov, O.; Nusbaum, C. D.; Tran, T. M.; Bertin, A.; Wilson, D. A.; Peterca, M.; Zhang, S.; Kamat, N. P.; Vargo, K.; Moock, D.; Johnston, E. D.; Hammer, D. A.; Pochan, D. J.; Chen, Y.; Chabre, Y. M.; Shiao, T. C.; Bergeron-Briek, M.; André, S.; Roy, R.; Gabius, H. J.; Heiney, P. A. Modular Synthesis of Amphiphilic Janus Glycodendrimers and Their Self-Assembly into Glycodendrimersomes and Other Complex Architectures with Bioactivity to Biomedically Relevant Lectins. *J. Am. Chem. Soc.* **2013**, 135 (24), 9055–9077. <https://doi.org/10.1021/ja403323y>.
- (23) Lu, Y.; Song, S.; Hou, C.; Pang, S.; Li, X.; Wu, X.; Shao, C.; Pei, Y.; Pei, Z. Facile Fabrication of Branched-Chain Carbohydrate Chips for Studying Carbohydrate-Protein Interactions by QCM Biosensor. *Chinese Chem. Lett.* **2018**, 29 (1), 65–68. <https://doi.org/10.1016/j.ccllet.2017.08.003>.
- (24) Assali, M.; Kittana, N.; Qasem, S. A.; Adas, R.; Saleh, D.; Arar, A.; Zohud, O. Combretastatin A4-Camptothecin Micelles as Combination Therapy for Effective Anticancer Activity. *RSC Adv.* **2019**, 9 (2), 1055–1061. <https://doi.org/10.1039/C8RA08794F>.
- (25) Fernandes, A. E.; Riant, O.; Jensen, K. F.; Jonas, A. M. Molecular Engineering of Trifunctional Supported Catalysts for the Aerobic Oxidation of Alcohols. *Angew. Chem., Int. Ed. Engl.* **2016**, 55 (37), 11044–11048. <https://doi.org/10.1002/anie.201603673>.
- (26) Werkhoven, P. R.; Elwakiel, M.; Meuleman, T. J.; Quarles Van Ufford, H. C.; Kruijtzter, J. A. W.; Liskamp, R. M. J. Molecular Construction of HIV-Gp120 Discontinuous Epitope Mimics by Assembly of Cyclic Peptides on an Orthogonal Alkyne Functionalized TAC-Scaffold. *Org. Biomol. Chem.* **2016**, 14 (2), 701–710. <https://doi.org/10.1039/c5ob02014j>.
- (27) Peterson, Q. P.; Hsu, D. C.; Goode, D. R.; Novotny, C. J.; Totten, R. K.; Hergenrother, P. J. Procaspase-3 Activation as an Anti-Cancer Strategy: Structure-Activity Relationship of Procaspase-Activating Compound 1 (PAC-1) and Its Cellular Co-Localization with Caspase-3. *J. Med. Chem.* **2009**, 52 (18), 5721–5731. <https://doi.org/10.1021/jm900722z>.

## Chapter 6 –Using ligand-directed chemistry for protein modification

### 6.1 Introduction

The last two chapters presented the work achieved throughout the PhD to develop ligand-directed warheads on either peptides (see Chapter 4) or on 2-PCAs (see Chapter 5). In this chapter, both types of probe were used for protein modification in a “traceless” manner. While these results were obtained in a relatively short time at the end of the PhD and the results require further optimisation, some exciting first results were obtained and will be described. All the results presented on protein labelling using 2-PCAs came from experiments and LCMS analysis ran by Lydia, another PhD within Dr Spicer group.

As mentioned in Chapter 4, no final probes for LDAI were obtained. LDAI probes were found to be too electrophilic, not only reacting fast in a proximity mediated reaction, but also undergoing rapid hydrolysis. Second, the route found for *OpLD* NASA using methylcyano probes was challenging hindering their use. Although good reaction conditions were eventually found there was not sufficient time to fully study the complete set of probes. Thus, it was decided to focus on the approach using the *p*-nitrobenzyl-NASA probes for *CatLD* detailed in Chapter 4.

Due to the difficulty of working with FGF-2 as a protein, with two main issues, precipitation and dimerization, which led to difficulty in studying binding in a reproducible manner, we focussed on the use of insulin as a model, with two other proteins, myoglobin and RNase-A also introduced for additional final results. First, the use of PyOx labelled peptides will be described to modify insulin, and then in a second part, the use of 2-PCA probes will be described to modify insulin, myoglobin and RNase-A.



modification could happen with peptides modified at either terminus, thus probes **A** and **B** presented in Figure 2 were selected. For consistency, both linkers contained the short C4 alkyne linker, however due to the larger amount of the divalent-catalyst at the C-terminus available following purification, one of the probes was monovalent while the second was divalent. It is important also to emphasise that no literature to date has compared the effects of having one or two PyOx catalyst heads on catalysis, which would be interesting to investigate in the future. Three NASA probes were selected as shown in Figure 2, the NASA-hexanoic due to its low steric hindrance, and both the biotin and NBD to have the possibility to do more analysis on the modified protein via western blot for biotin and fluorescence imaging for NBD. The experimental route for those three species and their full characterisation are to be found in Chapter 4.

For all the following experiments in terms of ease, speed, and mass accuracy, LCMS was the analysis of choice. However, the drawback of this analysis was the restricted range of concentrations that could be used, requiring us to work in a range that would not overload the column but also provide us with high enough signal and peak resolution. This concentration would in turn affect the concentrations of the other reagents which also needed to be considered. After a range of protein concentrations were studied between 1-1000  $\mu\text{M}$ , 50  $\mu\text{M}$  was picked as the best compromise. A large number of parameters were considered for screening, as shown in Table 1. Indeed, reaction time and temperature, as well as concentrations of both PyOx probe and NASA reagent needed to be varied leading to a lot of experimental possibilities.<sup>1,2</sup>

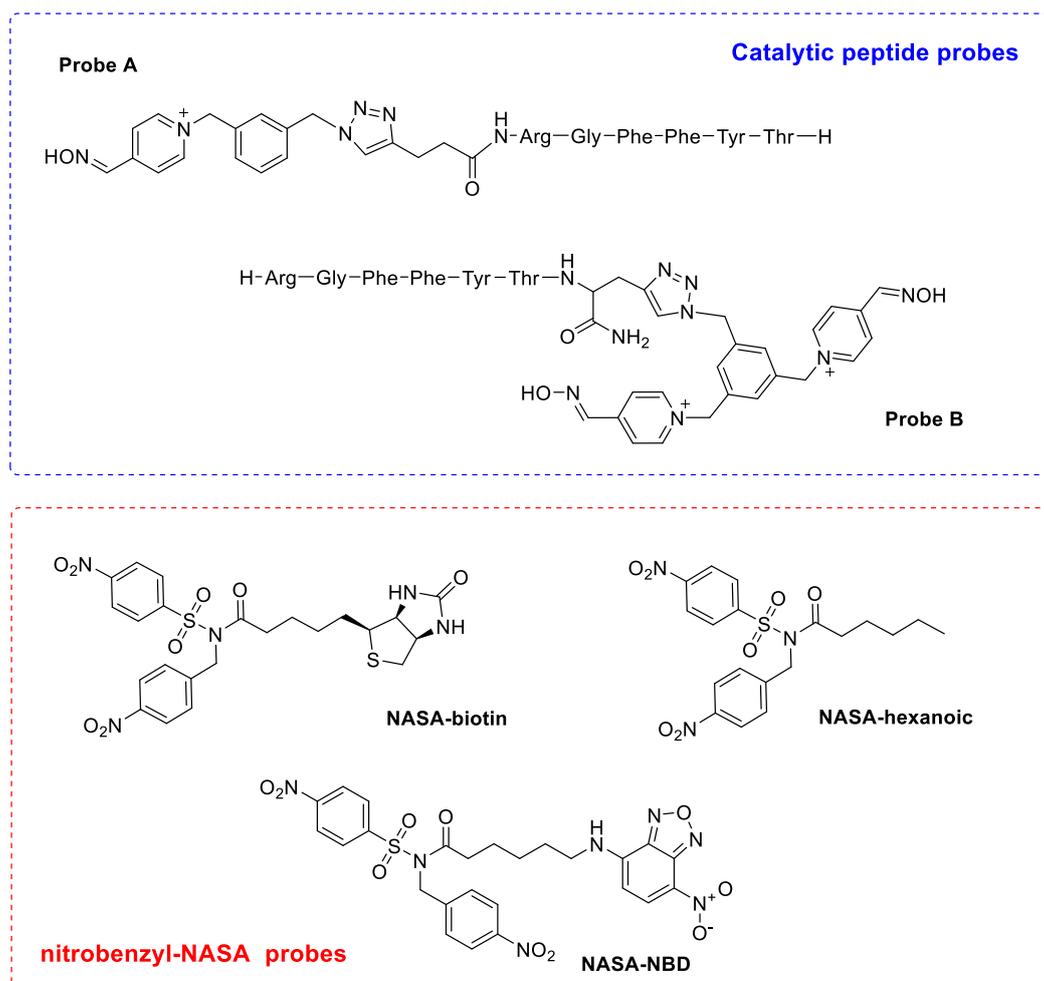


Figure 2: Figure of the final probes for ligand-directed chemistry using PyOx modified peptides binding to insulin, probe A and B, and different NASA reagents with biotin, hexanoic and NBD electrophile group

The poor solubility of the nitrobenzyl-NASA reagents which are relatively bulky and hydrophobic, was another limiting aspect, requiring pre-solubilisation in DMSO. In the literature, the use of 10 equivalents of NASA relative to catalyst was reported,<sup>1,2</sup> but this was not always possible in our system due to the poor solubility of the NASA species when using the highest concentration of PyOx probes. Moreover, due to the set-up of the experiment all reactions were run at room temperature for ease. Both solubility and kinetics will be affected by a higher temperature and this parameter would need to be investigated in the future.

Table 1: Parameters used in literature for PyOx catalytical ligand-directed chemistry

Reference	[protein] μM	[PyOx] μM (eq)	[NASA] μM (eq)	t, T	Main observations
<b>1</b>	10	30 ( <b>3</b> )	300 ( <b>30</b> )	6 h, 37 °C	Single addition between 18-28% with PyOx, or 70-90% with 2PyOx
<b>2</b>	5	5 ( <b>1</b> )	50 ( <b>10</b> )	6 h, 37 °C	Full conversion / Double and triple addition observed
<b>In this thesis</b>	50	50-500 ( <b>1-10</b> )	50-1000 ( <b>1-20</b> )	2-48 h, room temperature	Discussed below

### 6.2.2 Experiment and LCMS analyses

The first experiment was set up with both probe A and probe B, using NASA probes mentioned in the previous paragraph for insulin labelling. The equivalent of peptide probe to the protein was picked to be 1, 5 and 10 equivalents, and the equivalents of NASA species to the protein 1, 10 and 20. The stock concentration of peptide was 1 mM and the NASA reagent prepared as a stock solution in DMSO at high concentration then subsequently diluted to give a final DMSO content of 1.2%. Different controls were also performed which will be discussed in the following paragraphs. In the following section, we will be using the notation: A-1-10. This depicts that peptide probe A was used, with one equivalent of peptide to the protein and ten equivalents of NASA to the protein. The NASA used will always be specified. All samples were analysed by LCMS at two different time points, t = 2 hrs and t = 22 hrs. Not all the chromatograms will be presented in this thesis which will focus on the processed data.

The first observation was that when NASA reagents were incubated with insulin, without the peptide catalysts, no modification of the protein was observed. This showed the importance of the presence of the catalyst to activate the electrophilic reagent. The second observation was that when no NASA was added the two peaks of peptide and protein were similar to pure samples of each, without any changes as expected. Very interestingly as soon as both NASA and catalytic peptide were added, even at only one equivalent of each relative to the protein, modification of the protein could be seen with the appearance of two new peaks of +6 and +5 charge, Figure 3, crude chromatogram of mixture A-1-1.

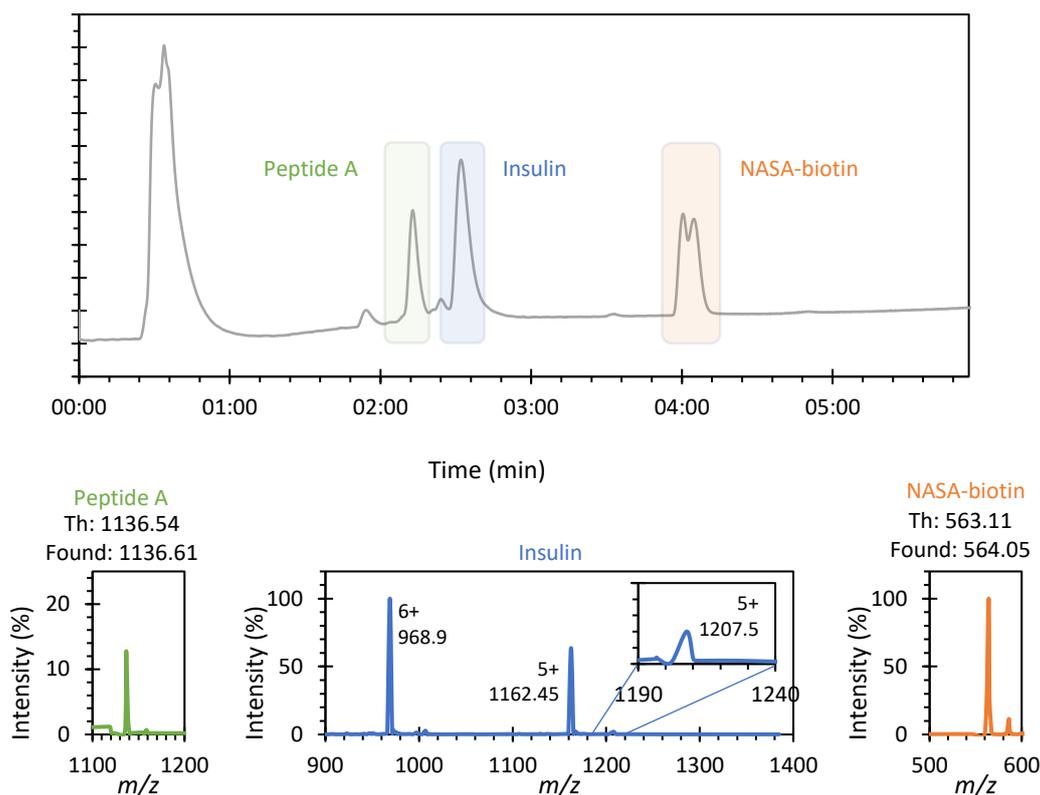


Figure 3: Chromatogram 210-400 nm of the crude sample A-1-1 after 2 hrs of reaction at room temperature of insulin in presence of probe A (one equivalent) and NASA-biotin (one equivalent)

Deconvolution of the mass spectra allowed us to confirm the mass of the modified protein being 6032.48 as summarized in Table 2, which was the mass expected due to the addition of a single biotin moiety onto the protein. Similar experiments were performed with the other two NASA reagents and the mass of each modified protein was observed, as shown in Table 2 and Figure 4. The percentage of modification was between 0.5-2% with the low intensities making it impossible to calculate percentages reliably.

Table 2: Mass detected for the modified and unmodified insulin linked to their total mass

m/z	Insulin	Insulin-hexanoic	Insulin-biotin	Insulin-NBD
+6	968.8		1006.6	1014.97
+5	1162.4	1182.11	1207.7	1217.71
+4	1452.8			
<b>Total mass</b>	<b>5806</b>	<b>5905</b>	<b>6033</b>	<b>6083</b>
<b>Delta</b>		<b>99</b>	<b>227</b>	<b>277</b>

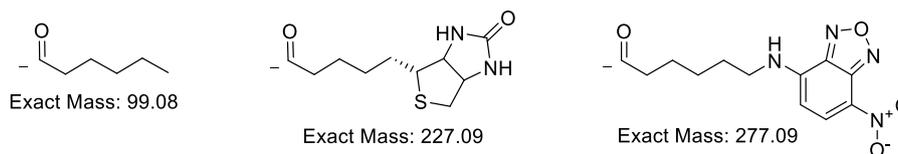


Figure 4: Mass corresponding to the three moieties tethered to be covalently attached to insulin through ligand-directed chemistry

With these promising results in hand, a range of equivalents was tested for both NASA-biotin and NASA-NBD. It is important to note that when increasing the concentration and using either 5 or 10 equivalents of any of the NASA reagents precipitation occurred impacting the success of the reaction. In the future it will be important to prevent that happening, however the amount of DMSO could not be increased too much for the protein to maintain its structure and activity. The development of new hydrophilic NASA esters is therefore attractive but outside the scope of this thesis.

The use of either NASA-biotin or NASA-NBD gave different results. First for NASA-biotin, it was shown that increasing the equivalents of either NASA or peptide led to an increase in conversion, up to 22% in the case of the crude mixture A-10-10. This increase in conversion allowed us to observe a shoulder peak in LCMS, with the retention time of the modified protein being slightly shifted Figure 5. This shape became more defined for higher equivalents of NASA-biotin and peptide.

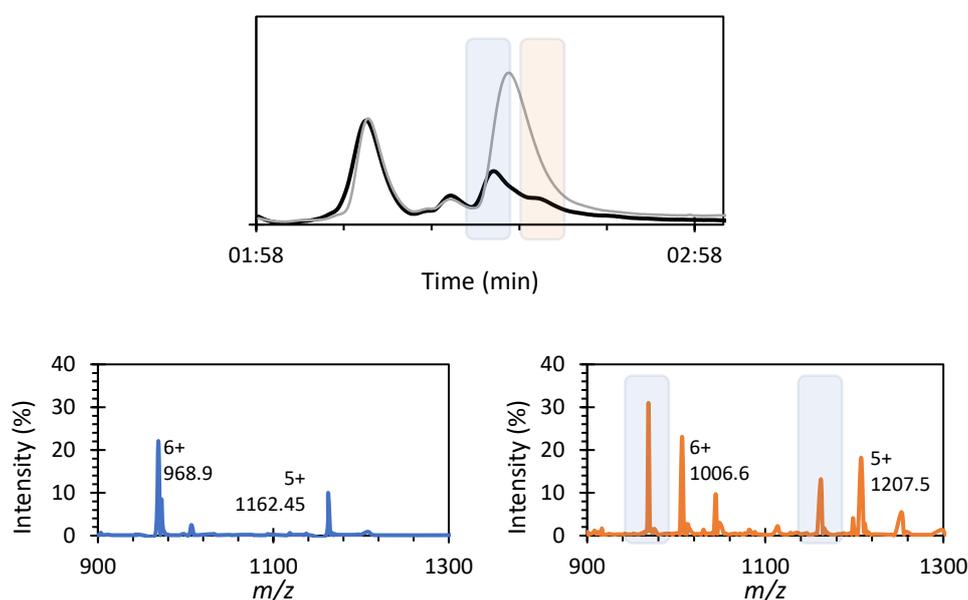


Figure 5: Overlapping of the chromatograms 210-400 nM of crude mixture A-1-1 (grey) and mixture A-5-20 (black). Corresponding spectrogram for the highlighted main insulin peak in blue and the modified insulin in orange, showing the apparition of the two new mass peaks at 1006.6 and 1207.5 with still presence of unmodified insulin mass highlighted in blue..

Moreover, it was shown that for the same ratio of NASA/peptide probes, the conversion was peptide dependant, with probe B being less effective, which could be due to a less ideal arrangement of nucleophiles in the vicinity to the C-terminus of the bound peptide. Emphasizing the difference in

reactivity between the two termini even more, the probe for the C-terminus was divalent compared to the monovalent probe at the N-terminus. It would be interesting in future to study the influence of both mono and divalent catalysts at both the N or C terminus. However, conversion was still observed up to 19% using the crude mixture B-10-20. Though similar trends in reactivity were observed for NASA-NBD, conversions were generally lower, and the respective spectra are presented in the Experimental section.

Only the conversions for reactions with NASA-biotin are presented in Table 3 (2 h) and Table 4 (22 h).

Table 3: Summary of the yield of modification of insulin using probe A - in bracket using probe B - with NASA-biotin at different equivalents after 2 h at room temperature. The percentage are calculated from the ratio of the intensity of modified protein versus unmodified protein hypothesising that both species ionised in a similar manner. <sup>‡</sup>Not calculated due to too much overlapping of peptide peaks

t1 2 h		Peptide (equivalent)			
		0	1	5	10
NASA-biotin (equivalent)	0	No modification of the protein			
	1	No modification of the protein	3% (0%)	3% (0%)	7% (7%)
	10	No modification of the protein	3% (4%)	7% (5%)	22% (8%)
	20	No modification of the protein	6% (9%)	11% (16%)	N.A. <sup>‡</sup> (19%)

Table 4: Summary of the yield of modification of insulin using probe A - in bracket using probe B - with NASA-biotin at different equivalents after 22 h at room temperature. The percentage are calculated from the ratio of the intensity of modified protein versus unmodified protein hypothesising that both species ionised in a similar manner. <sup>‡</sup>Not calculated due to too much overlapping of peptide peaks

t2 22 h		Peptide (equivalent) A (B)			
		0	1	5	10
NASA-biotin (equivalent)	0	No modification of the protein	1% (1%)	2% (1%)	2% (2%)
	1	No modification of the protein	5% (0%)	3% (0%)	3% (4%)
	10	No modification of the protein	7% (5%)	10% (1%)	18% (8%)
	20	No modification of the protein	7% (11%)	12% (8%)	N.A. <sup>‡</sup> (N.A. <sup>‡</sup> )

The main observation after overnight reaction was that some unselective modification was now present in the absence of the peptide. However, these conversions were not significant (< 2%) compared to when the probe was added. Though a small increase in conversion between 2 and 22 h

(up to 2%), this effect was minimal and within experimental error and the reaction can be considered to have neared completion at the first time point.

To conclude these results, we observed the first modification of insulin using ligand-directed peptide-based chemistry with peptide RGFFYT modified at either of its termini. While the conversion rate was dependent on the position of the catalyst on the peptide and the NASA used, different groups were able to be covalently tethered to insulin. A major conclusion was that the reaction happens in short time frames at room temperature. With longer reaction times there was not a significant increase in conversion but non-specific labelling started to be observed. Despite LCMS not being a quantitative method, promising results suggested conversions up to ~20% could be achieved. Different analyses were therefore performed to investigate these promising results further.

### 6.2.3 SDS-PAGE and western blot analyses

As mentioned earlier, the use of NASA-NBD to modify insulin allowed us to perform fluorescence imaging, while NASA-biotin allowed western blot analysis to demonstrate covalent labelling. After a preliminary test on insulin to determine good gel conditions, see Experimental section, the concentration of protein per well was picked to be 25  $\mu$ M.

As seen in the Coomassie gel, Figure 6, insulin runs very quickly due to its small size, with the band being close to the bottom of the gel, and conditions are not optimal. Better separation could be achieved using a trypsin gel or a gel with a higher percentage of acrylamide to slow down and separate better the bands at lower molecular weight. However, due to time constraints, these results, which confirm LCMS results of protein labelling, were not optimised. The second well contained insulin mixed only with NBD-NASA. As supported by LCMS, a small fluorescent band was observed showing a small degree of background labelling. The third well contain peptide probe A. While no band could be observed in the Coomassie, when under fluorescence a large band at around 20 kDa could be observed. This band could either come from peptide aggregation, or also from some precipitate.

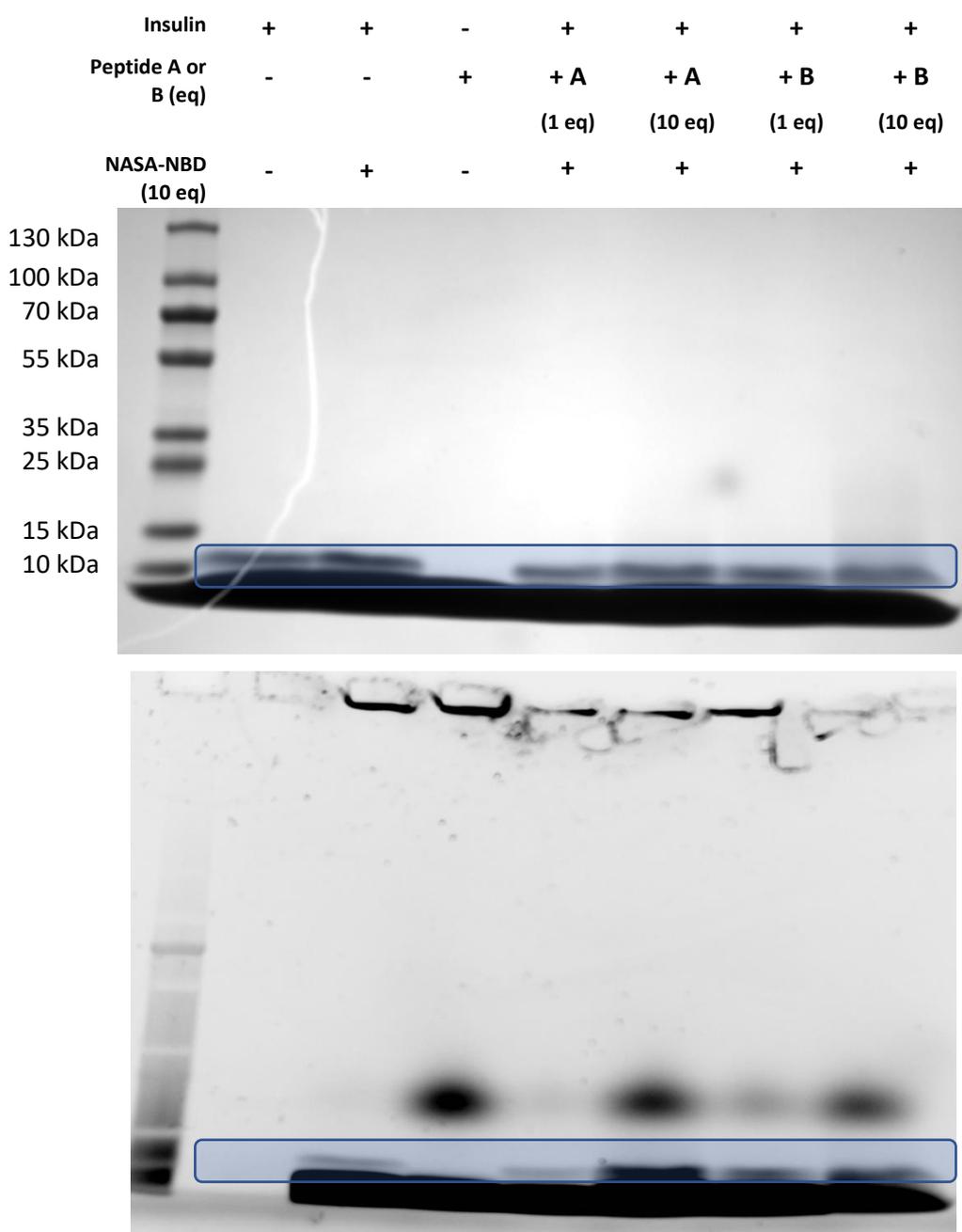


Figure 6: SDS-PAGE of insulin ( $25 \mu\text{M}$ ) in well 1 and then of insulin in presence of peptide probes A and B and NBD-NASA at different conditions. Top after rapid Coomassie stain, and bottom fluorescent gel with opposite saturation. Insulin band highlighted in the blue rectangle.

Following the different controls, four reaction mixtures were analysed from well four to seven: A-1-10, A-10-10, B-1-10 and B-10-10. An increase of fluorescence intensity could be observed when the equivalent of both peptide probe was increased, from well four to five for probe A and from well six to seven for probe B, highlighting again the importance of catalyst and confirming the conversions obtained by LCMS analysis.

Figure 7 represents the SDS-PAGE and subsequent anti-biotin western blot of the modification of insulin using NBD-biotin with probes A and B. Once again, some potential background labelling was observed.

Similarly to the gel run for fluorescence analysis, mixture A-1-10, A-10-10, B-1-10 and B-10-10 were used in the last four wells. While no significant change can be observed under Coomassie stain, the intensity of the band and its shape changed quite significantly after western blot, potentially showing an increase of modification of the protein. It is necessary to add that standard western blotting is not quantitative, and these results are presented to support the LCMS data. The diffusivity of the bands especially in the last two wells could result from lack of specificity of the antibody used for the blot.

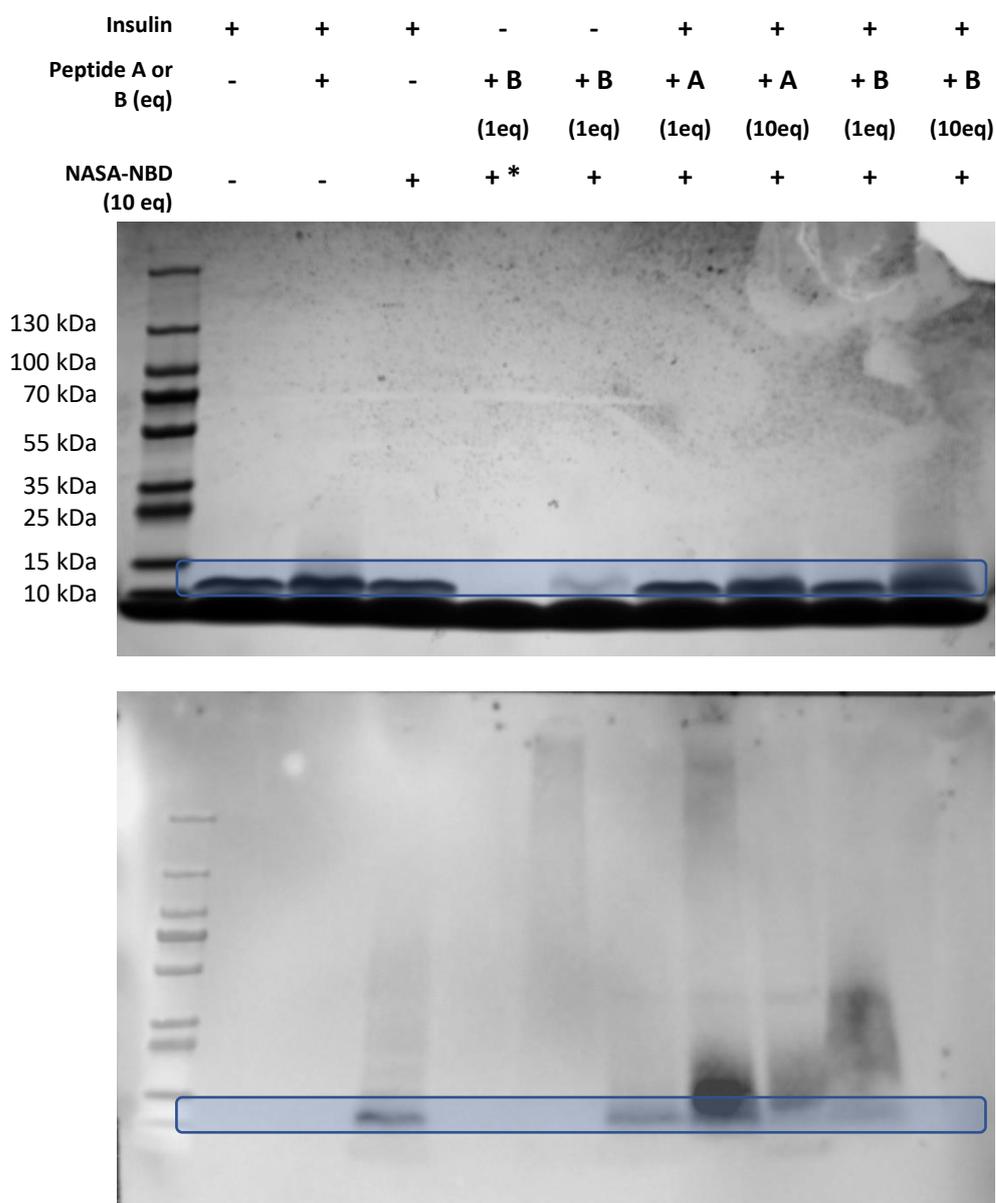


Figure 7: Top: SDS-PAGE of insulin with peptide probe A and B and NBD-biotin at different conditions after Coomassie stain. Bottom: western blot of the same conditions. \* 1eq of NASA-NBD

To conclude these gel results, it was positively confirmed that insulin was modified by two different types of NASA, the NASA-NBD and the NASA-biotin. Second, it was shown that adding more equivalents of catalyst increased conversion, confirmed by both LCMS and gel analysis. However, it was also shown that some low level of non-specific addition occurred when no peptide catalyst was present in solution.

#### 6.2.4 MALDI analyses

A final type of analysis was performed to confirm the masses observed by LCMS and more importantly to give information on the localisation of the modification using mass spectrometry. MALDI and MALDI-MS/MS were performed by Dr Adam Dowle in the Department of Biology to look at the fragmented protein.

A control experiment was performed using unmodified insulin to identify the peaks of interests, and to optimise procedures since the small size of insulin makes it difficult to study by mass spec. The acquired spectra for unmodified insulin showed that each obtained peak matched human insulin P01308, Figure 8. The peptide at 2601.3 m/z is the most prominent identifier. Based on literature,<sup>3</sup> the peptide we used within this chapter should bind to this fragment of the full-length insulin which will be more discussed in the following paragraphs.

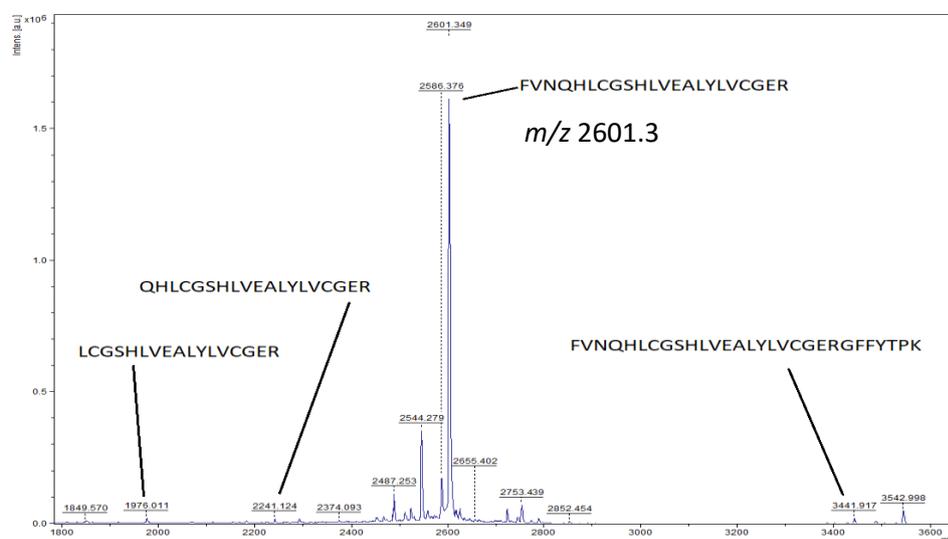


Figure 8: Representative results done for MALDI-MS/MS of insulin after reducing conditions

For protein analysis, no extra purification step of the crude samples used for LCMS was performed, and thus the samples still contained peptide and unreacted NASA species. The data was analysed by isolation through LIFT-MS/MS. It is worth noting that this method is classically used for samples with a mass up to 4000 Da. With insulin having a mass of 5806 Da, slightly out of this range, the quality of the results could have been impacted. However due to the complex structure of insulin presenting three cysteine bonds, analysis of non-reduced protein would be easier to analyse and identify.

The first results presented were obtained by MALDI-MS of the crude mixture. They confirmed with a high mass accuracy what was seen during LCMS and gel analysis, that insulin was successfully modified with two different types of NASA, hexanoic (shift of +98 Da) and biotin (shift of +228 Da), respectively in brown and green in Figure 9.

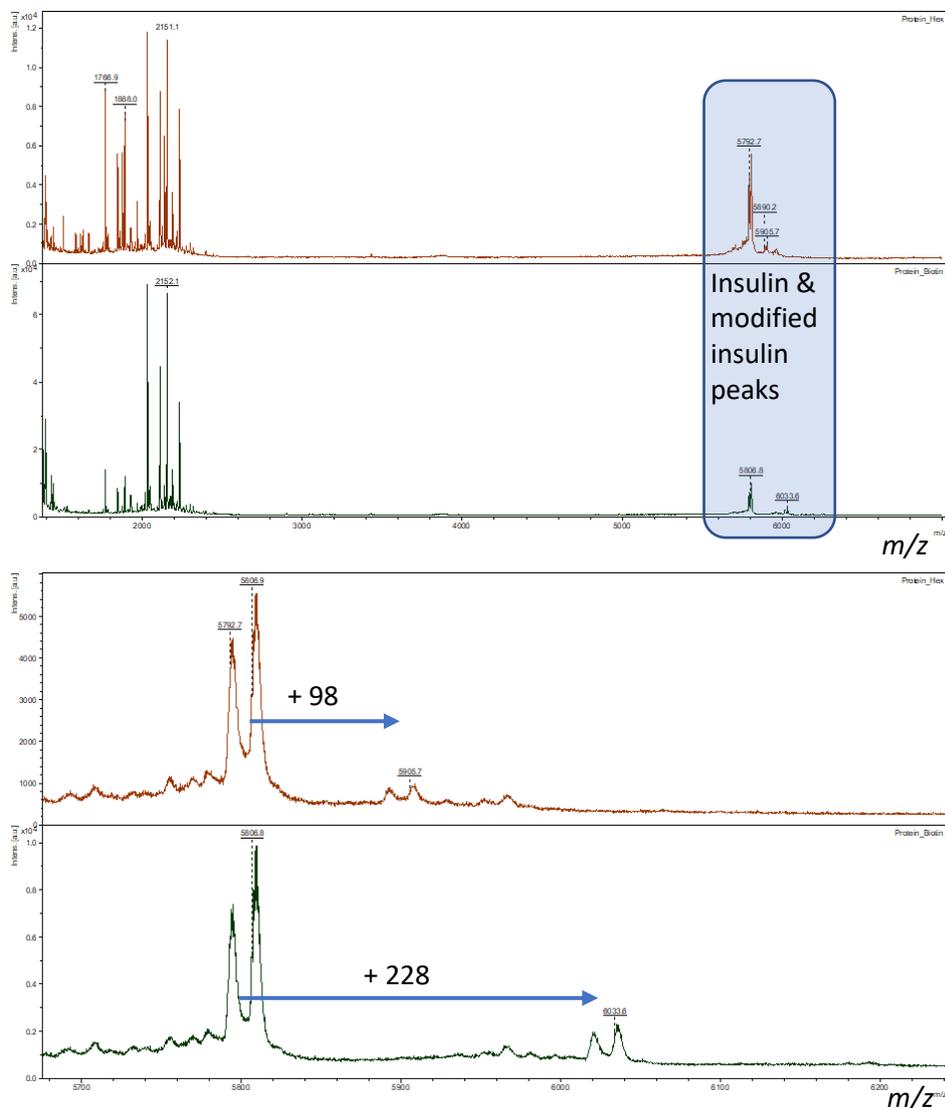


Figure 9: Image from MALDI-MS showing the full scan (top) and the zoomed in version (bottom) of the crude mixture with insulin in presence of NASA-hexanoic (brown) and NASA-biotin (green). Both unmodified insulin peaks and modified insulin peaks with either hexanoic (mass +98) or biotin (mass+228) could be observed.

Following this confirmation, the following step was to identify where on the protein the modification had occurred. For this, the crude mixture was first reduced with 2.5 mM tris(2-carboxyethyl)phosphine (TCEP), followed by heating at 60 °C for 1 h and then alkylation with 20 mM S-methyl methanethiosulfonate (MMTS) for 10 mins. The main peak observed after this breakdown of the two insulin chains was the peak at 3428.7 which matched chain A of insulin. Both addition of hexanoic acid or biotin were observed, showing that the modification was occurring on chain A, Figure 10.

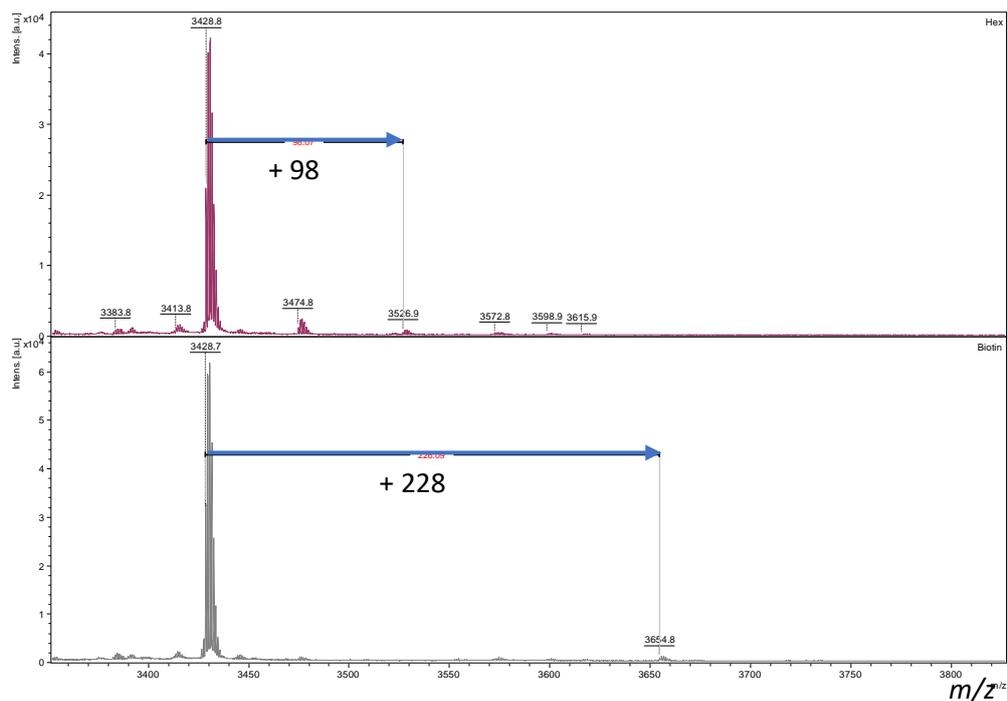


Figure 10: MALDI results for fragmented and reduced protein showing the addition of hexanoic acid and biotin on peak 3428  $m/z$  matching the sequence FVNQHLCGSHLVEALYLVCGERGFFYTPKT

Then both masses 3526.9 and 3654.8 were subjected to LIFT fragmentation Figure 11. While only a small number of peaks could be observed, three peaks matching b-ions could be identified at  $m/z$  1303, 1965 and 2586, matching the  $y_{11}$ ,  $y_{17}$  and  $y_{23}$  ions of the unmodified sequence.

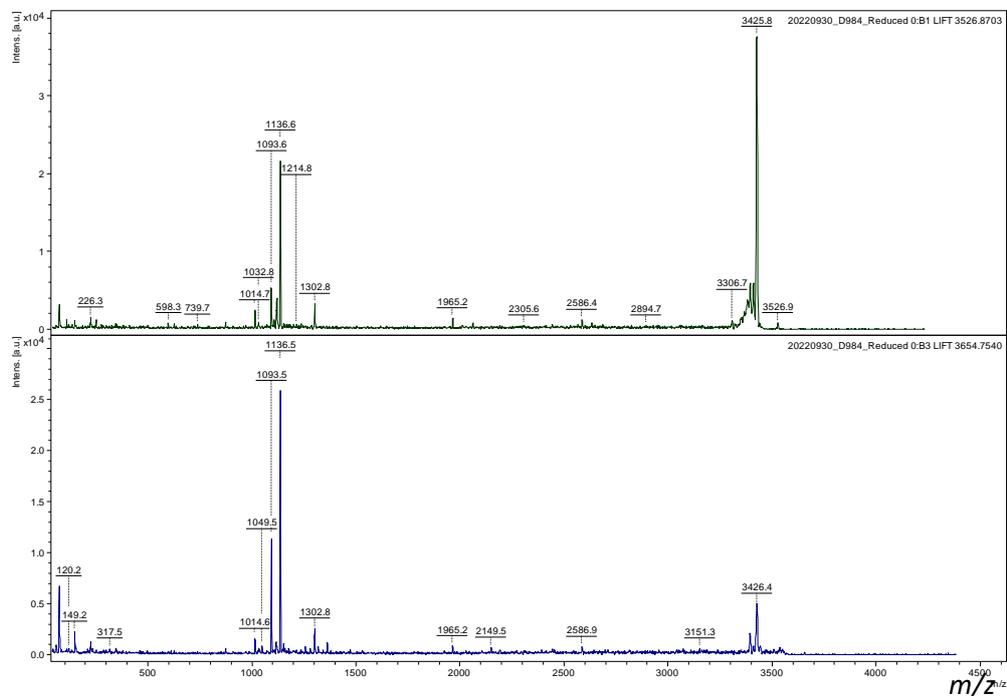


Figure 11: LIFT fragmentation of peptide FVNQHLCGSHLVEALYLVCGERGFFYTPKT with hexanoic and biotin addition

This suggests that the amino acids from residue 11 till the C-terminus are unmodified – i.e. modification with hexanoic acid and biotin is most likely happening at the *N*-terminus of chain A in the underlined part of the sequence **FVNQHLCGSHL**VEALYLVCGERGFFYTPKT.

From literature it is known that the peptide RGFFYT binds to the underlined part of the sequence FVNQHLCGSH**LVEALYL**VCGERGFFYTPKT,<sup>3</sup> shown in dark orange in Figure 12. When using catalyst probe A, being modified at its *N*-terminus, the modifying site on insulin correlates well with this proposed binding site. Of the first eleven residues of chain A, histidine, serine and cysteine are present and each of them could be a potentially modified nucleophilic amino acid. Moreover, it is important to not discard the possibility of reaction with the *N*-terminus of the chain. So far, we have not been able to further assign the true site of modification. Some computational modelling could be done to have a better idea of the binding pocket and which of these four residues would be more likely to react in terms of 3D arrangement and distance from the catalytical head. In literature,<sup>1,2</sup> only lysine and serine have been shown to serve as good nucleophiles for *CatLD* NASA modification chemistry, and thus serine emerges as a strong candidate.

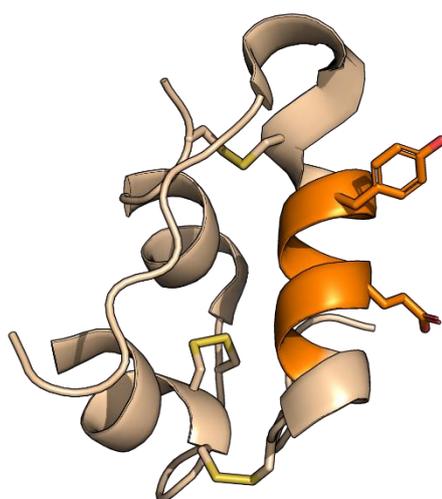


Figure 12: Protein structure of insulin highlighting the known segment to which the peptide RGFFYT binds to.

#### 6.2.5 Further observations on peptide

While the protein studies were carried out, a few interesting observations were found which we thought worth mentioning in this thesis.

The first observation was from the negative control in which NASA reagent and probe B were mixed together in the absence of the protein. While by LCMS, after short time (hours), the naked peptide peaks were still the main ones, a new peak was formed overtime with masses that would match the addition of the NASA onto peptide B, Figure 13 and Figure 14. The masses were seen for any of the NASA reagent used as summarized in Table 5 but only the example with hexanoic is showcased in the

main text of this thesis. This phenomenon was interesting to investigate further as following activation of the NASA by the oxime catalyst, the intermediate species formed is expected to be highly unstable.

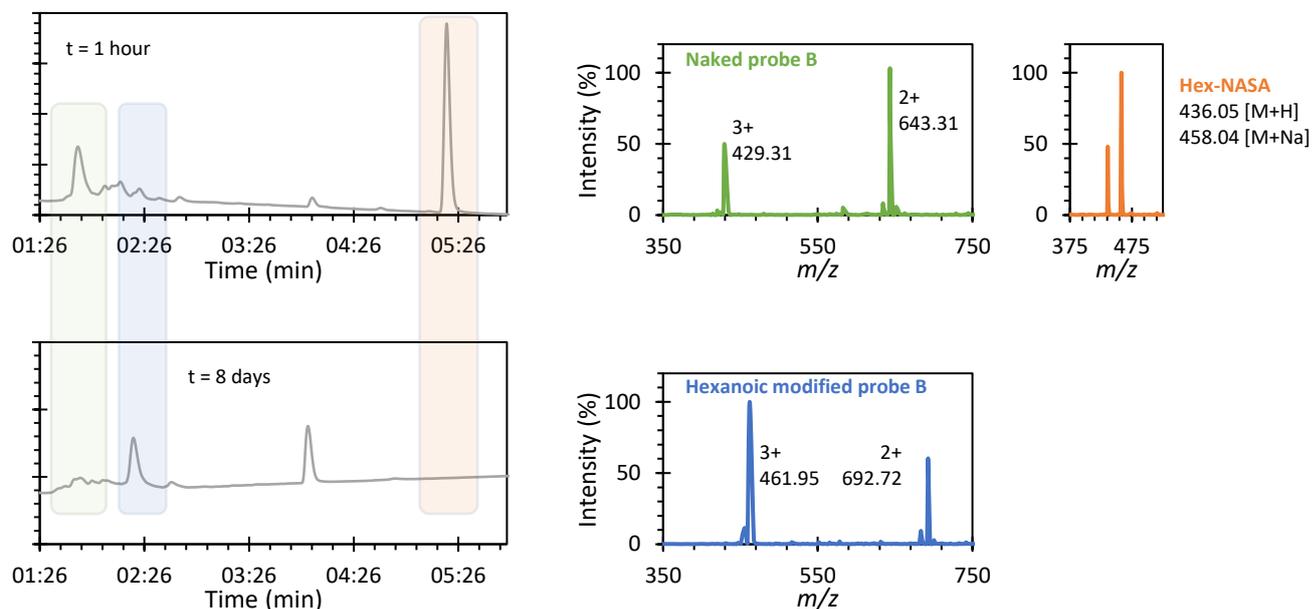


Figure 13: Chromatograms 220 nm of the crude sample containing probe B and NASA-hexanoic after 2 h of reaction at room temperature (top) and after 8 days (bottom)

Table 5: Summary of the observed peptide peaks using probe B with both NASA-biotin and NASA-NBD

m/z	Naked peptide B	B-hexanoic	B-biotin	B-NBD
<b>+3</b>	429.26	461.95	504.75	521.34
<b>+2</b>	643.29	692.72	756.75	781.71
<b>Deconvolute</b>	1284.57	1383.43	1511.49	1560.66
<b>Theoretical</b>	1286.61	1384.68	1512.69	1562.69

Interestingly, this phenomena was not observed for probe A. The main difference between probe A and probe B is the presence of a potent nucleophile with a free primary amine in probe B. Thus, we hypothesised that the NASA when first attached through the oxime bond for both probes can undergo another reaction while on probe B leading to being covalently attached to a residue and likely via the *N*-terminus. Indeed, when only present at the oxime bond, which could be the case for probe A, potential hydrolysis could happen overtime. It is worth noting that it is unknown yet if the reaction was intramolecular with possible cyclisation within one peptide or intermolecular while a robe modified the next.

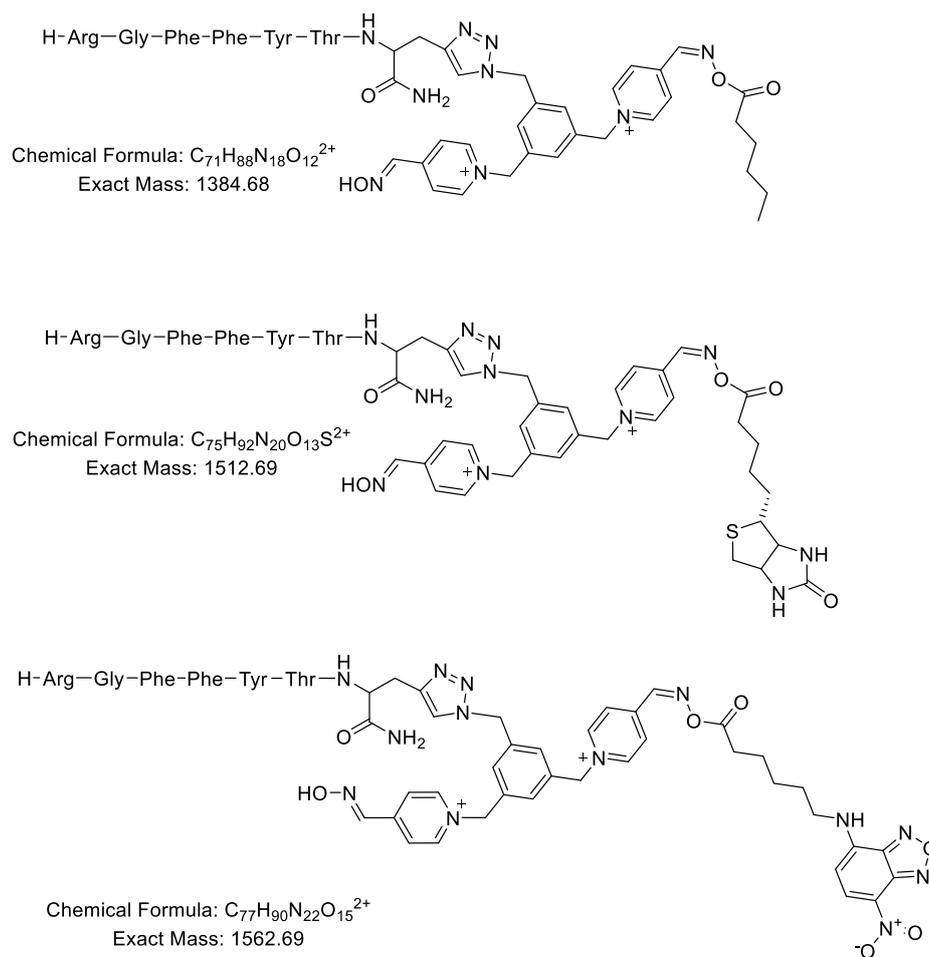


Figure 14: Structures using probe B showing the addition of NASA reagents onto peptide moiety

To study this further, MALDI-MS/MS results were undertaken on two samples using a simpler monovalent catalyst C-terminus peptide probe C in presence of both NASA-hexanoic and NASA-Biotin to identify where the modification happened.

#### Probe C

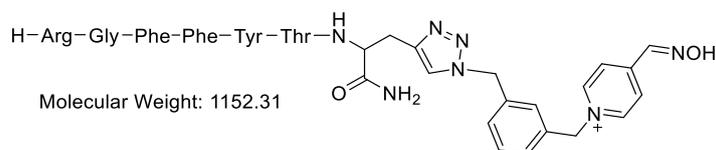


Figure 15: Probe C for analysis of addition of NASA substrate onto peptide moiety

As we can see from Figure 16, the main peak was the modified peptide probe at 1261.7  $m/z$  and 1389.6  $m/z$  for modification with NASA-hexanoic and NASA-biotin respectively. After MS/MS of this peak, further analysis could be done. Indeed, an interesting fragmentation pattern appeared in both sets of data, with the presence of three peaks. The delta shift between the first and second peak matched the mass of a phenylalanine residue while the delta shift between the second and third matched a

tyrosine residue. With the sequence being **RGFFYT**, it was possible to conclude that the modification was happening on the first three highlighted residues.

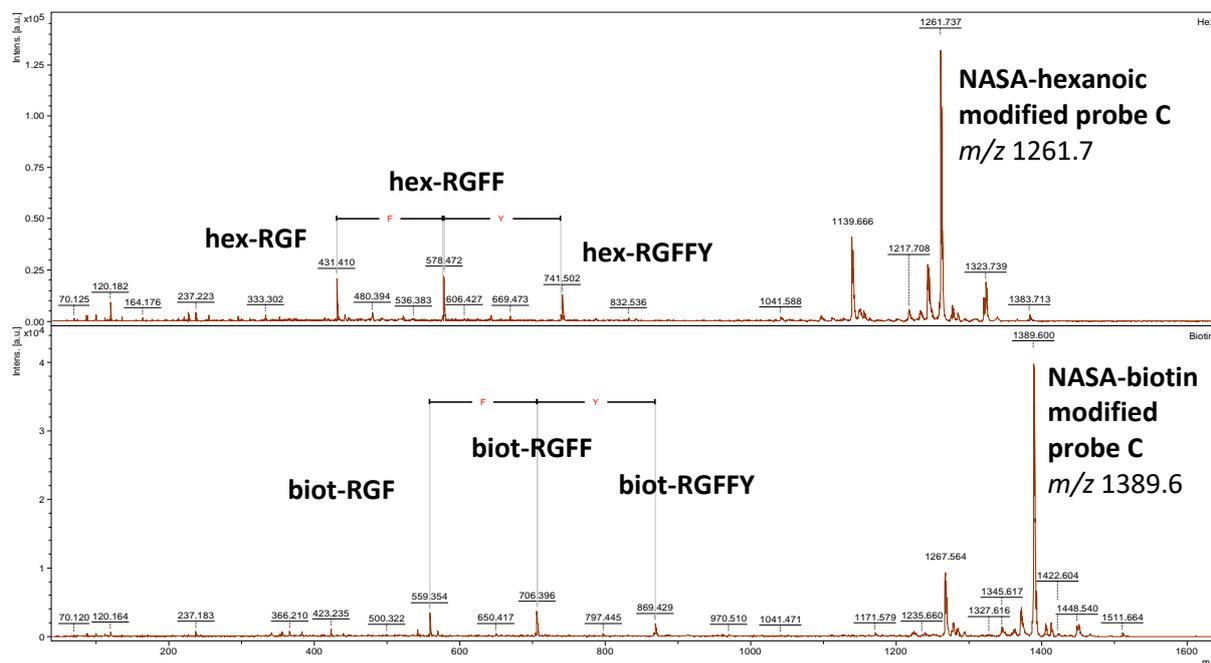


Figure 16: MALDI obtained after MS/MS of peptide peaks for probe C modified with NASA-hexanoic (top) and NASA-biotin (bottom), highlighting the three a-ions obtained for each, structure in Figure 17.

Having no other nucleophile other than the arginine on this sequence, it is very likely that the modification happened on this residue. Indeed, as shown in the representative scheme, Figure 17, of the a-ions expected for each peptide, the masses matched the fragmentation of the peptide peak by MALDI. However, it was not possible to demonstrate clearly whether the side chain or the free *N*-terminus primary amine was modified with this second option being more likely. While exciting to see these first results as a side-output of our protein studies, a full peptide study would need to be conducted to clarify and enlarge the results with further analysis. Some interesting parameters to study would be: which amino acids could be modified using this method; what linker length is optimal; if the *N*-terminus is really the main tethering point using capped peptides; what are the kinetics of the reaction. This would provide a useful platform to investigate the labelling preference of LDOp chemistry, as literature<sup>1,2</sup> currently speculates that only lysine and serine residues could be modified, but it is not possible to rule out modification at other sites. However, as an important note to conclude this part, while this phenomena was peculiar, our main reason to use PyOx-designed peptide is to modify proteins. Thus, the intramolecular reactions seen here should be limited as they could potentially impact the ability of the peptides and their ability to bind to the protein of interest, as well as reducing the amount of NASA available.

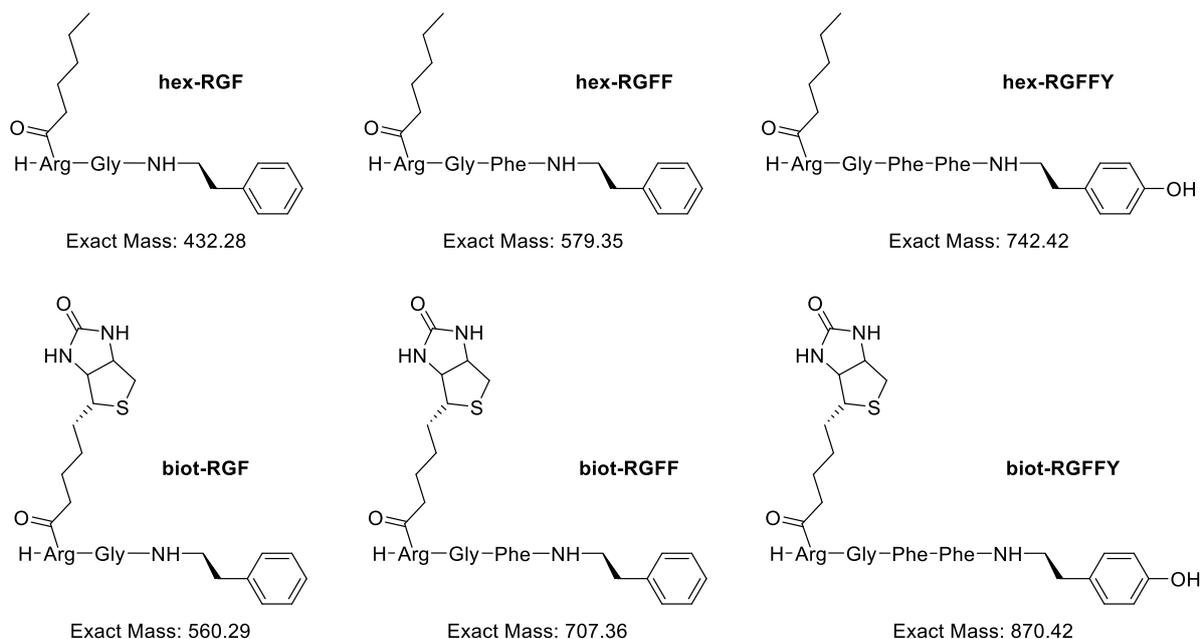


Figure 17: Structure of the ions observed under fragmentation of peptide peak by MALDI showing that the reaction between catalytic head oxime and NASA moieties lead to N-terminus modification of the peptide.

Another observation was found during LCMS studies, with the apparition of a new mass peak within the starting material matching a mass of -18 Da when in presence of the NASA species showing potential degradation of the probe. Moreover, this peak increased between the two time points, see Experimental section Figure 25 and Figure 26. This was similarly observed at the second time point. Indeed, at a fixed ratio of peptide to protein this new peak was increasing with the percentage of NASA within the solution. This will have to be investigated further.

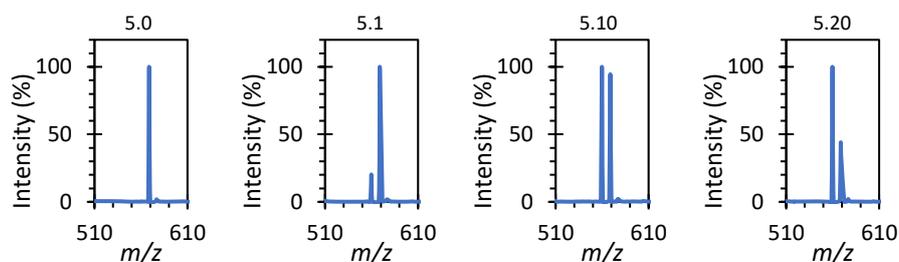


Figure 18: Chromatogram of the peptide A peak at 2.1 min from the crude mixture (left to right) 5.0, 5.1, 5.10, 5.20 of ratio peptide.NASA showing the increase of this second -18 Da peak.

### 6.2.6 Conclusion

To conclude the use of PyOx-peptide catalysts for protein modification, these first experiments using different peptides with either NASA-hexanoic, NASA-biotin and NASA-NBD showed promising results. Indeed, modification of insulin was observed with all three NASA species with up to 22% conversion when using peptide probe A. While some questions were answered such as an idea of where the insulin was modified, supporting the literature precedent for peptide binding site, a lot of other interesting observations require further investigation. First it would be interesting to discover if the protein was modified at another site when using the catalyst on the other terminus of the same peptide sequence. Second if the use of different peptide probes would be more efficient for *OpLD* on insulin. Third, while MALDI and LIFT fragmentation were crucial toward the conclusion of this section, tryptic digest could also be done to have more information on the localisation of the modification site. Again though, the small size of insulin with three disulfide bonds also leads to only very few cleavage points using trypsin. Another protease could be used in the future to aid analysis. Fourth, another interesting fact to look into in the future would be to see if the small amount of background modification was occurring at the same place or not. It was already shown that non-specific addition was happening mainly over longer time and this will have to be controlled and limited but it would also be interesting to see how the use of peptide catalysts impacts the specificity of the modification. Finally, as mentioned when comparing to literature, temperature and pH influence would also have to be studied.

### 6.3 The use of PyOx 2-PCA probes on proteins

In this last section, the use of the 2-PCA probes synthesised in Chapter 5 was attempted for protein modification at the *N*-terminus of different proteins. All the experimental set-up and the analysis of the LCMS was done by Lydia Barber and these results represent just a starting point of what will be possible using this new type of probes. Two different experiments were set up using two of the six 2PCAs synthesised for a proof-of-concept study, 2PCA PyOx-PEG3-2PCA, referred as probe D, and 2PCA 2PyOx-2PCA, referred as probe E, Figure 19. These two structures differed in both the linker used, and the number of catalyst heads from monovalent to divalent. Two NASA reagents presented earlier in this Chapter NASA-hexanoic and NASA-biotin were used. Three proteins were picked for this first study, myoglobin, RNase-A and insulin, as they were shown to be successfully modified using 2-PCA probes in the literature and our labs previous studies.<sup>4</sup>

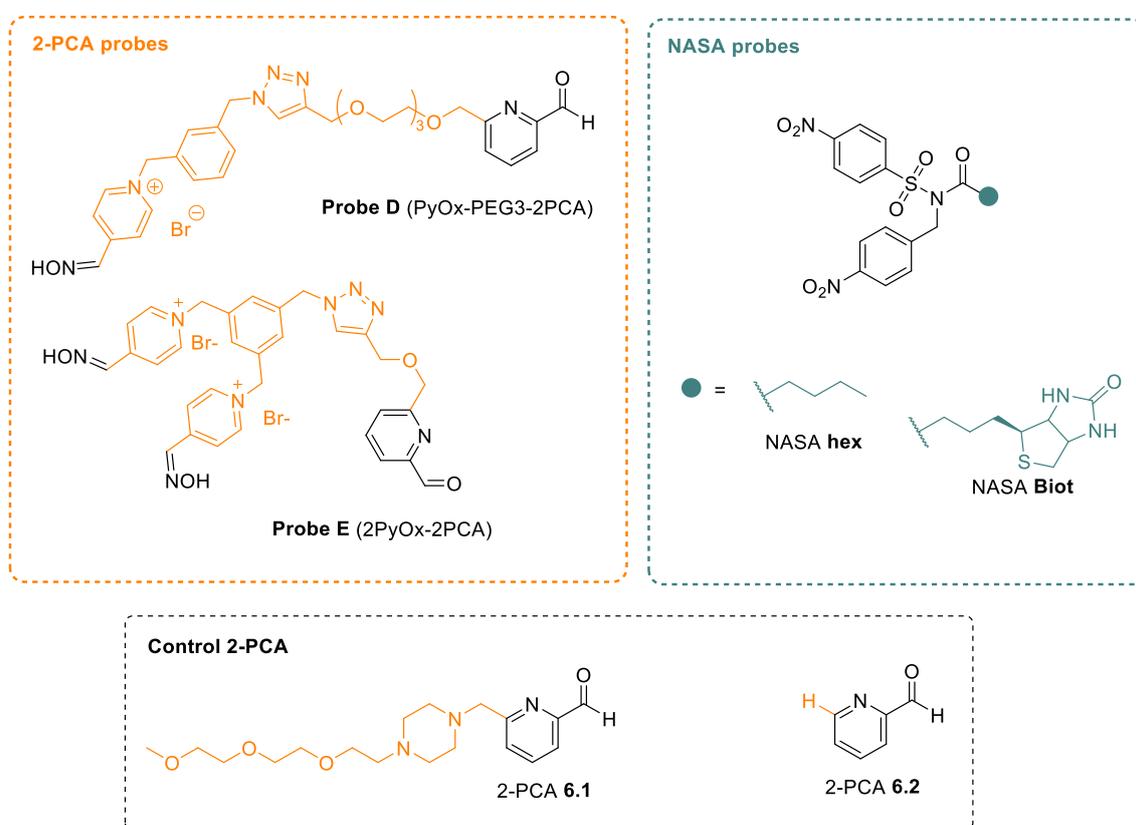


Figure 19: Representation of each probes used 2-PCA, NASA and control moieties

As a reminder from Chapter 5, after the addition of 2-PCA on to a protein, forming compound **I**, we hypothesised that after introduction of NASA, the formation of intermediate **II**, Figure 20, would happen which could then catalyse protein modification via ligand-directed chemistry if a nucleophile was in proximity to the *N*-terminus of the protein, forming **III**. The last step would be the hydrolysis of 2PCA from the protein leading to **IV**.

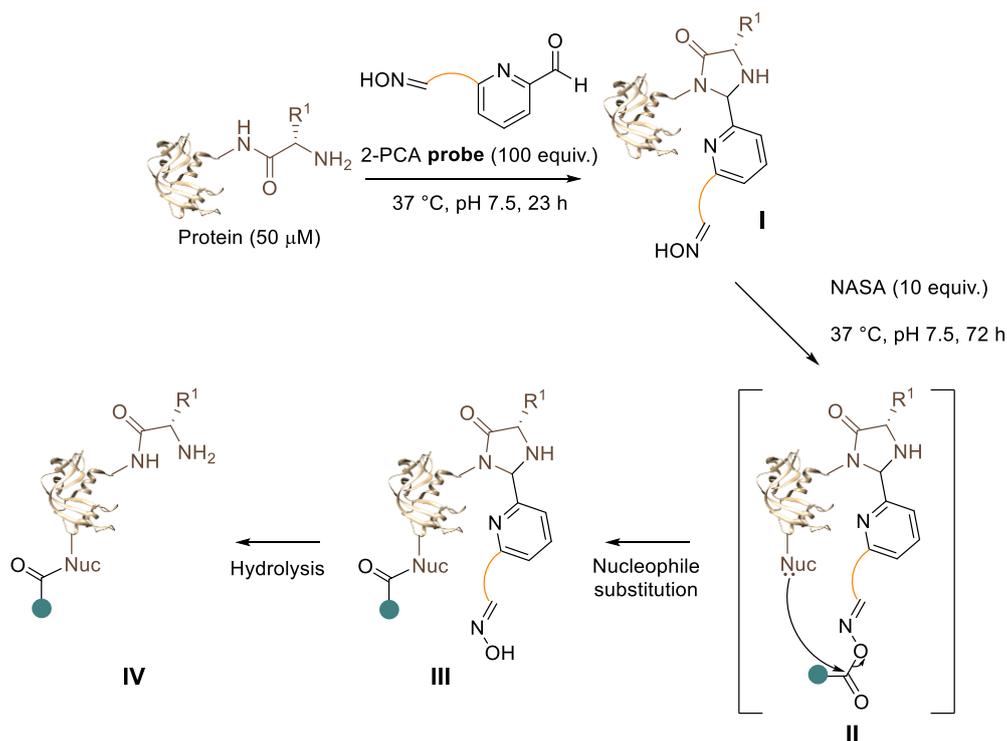


Figure 20: Scheme representing the different steps of the process of using PyOx-2PCA moiety to modify protein in a ligand-driven approach. Compound I is formed after cyclisation, then II is formed activating the NASA moiety which can then react with a nearby nucleophile to give III. Hydrolysis of the 2PCA overtime leaves the selectively single modified protein IV.

For analysis, Lydia looked at the formation of compound I for each attempt as well as compound II and III. These last two species have the same mass and so could not be distinguished. However, we hypothesised that oxime linked NASA II would be prone to hydrolysis and so be unlikely to be observed. Our analysis showed that the presence of excess probe D in the mixture made analysis challenging due to co-elution with the target proteins. Thus, for future attempts a dialysis step would be necessary prior to analysis. Thus, the calculation of the conversions given below are approximate.

Despite these challenges, and as highlighted in a dark blue in Table 6, the formation of II/III was observed (up to 14% for myoglobin) in some cases after 24 hrs. However, after 72 hrs modification was found to be lost. Three hypotheses for this observation are possible: i) no suitable nucleophilic groups are available on the protein *N*-termini and that the additions observed at 3 hrs and 24 hrs were more likely to be the transient intermediate II; ii) the modified protein precipitated over longer incubations; or iii) addition to the protein was unstable, e.g. by forming an ester at serine that would be prone to hydrolysis. Additionally, for an RNase A control when no 2PCA catalyst probe was inserted, a mass for compound III (<5%) could be observed showing some potential background acyl transfer from the NASA to the protein in a non-specific manner.

Table 6: Summary of the modification of proteins using Probe D and NASA-hexanoic and NASA-biotin. I = 2-PCA-modified protein; II = Acyl intermediate; III = 2-PCA-acyl-modified protein; s = single modification; d = double modification; t = triple modification; q = quadruple modification; qu = quintuple modification.

2-PCA	NASA	Protein	Conversion (%)			
			Before NASA addition	3 hrs after NASA addition	24 hrs after NASA addition	72 hrs after NASA addition
Probe D	Hex	Myoglobin	I (15)	I (26)	II/III (6)	I (17)
Control 6.1	Hex	Myoglobin	I (42)	I (51s, 6d)	I (43s, 5d)	I (39s, 11d)
Probe D	Biot	Myoglobin	I (15)	I (12), II/III (15)	I (9), II/III (14)	I (6)
Control 6.1	Biot	Myoglobin	I (42)	I (40)	I (44)	I (34s, 8d)
Probe D	Hex	RNase A	I (49)	-	-	I (69)
Control 6.1	Hex	RNase A	I (77s, 4d)	I (88)	I (86), III (4)	I (86s, 4d), III (4)
Probe D	Biot	RNase A	I (49)	-	-	I (70)
Control 6.1	Biot	RNase A	I (77s, 4d)	I (90)	I (84s, 5d)	I (93)
Probe D	Hex	Insulin	-	-	-	-
Control 6.1	Hex	Insulin	I (67t, 27q, 6qu)	I (33d, 45t, 23q)	I (22d, 52t, 26q)	I (6d, 60t, 30q, 4qu)
Probe D	Biot	Insulin	-	-	-	-
Control 6.1	Biot	Insulin	I (67t, 27q, 6qu)	I (32d, 42t, 25q)	I (100d)	I (22d, 41t, 32q, 4qu)

A second experiment was set-up in a similar manner with probe E. However, precipitation was observed during the protein modification with probe E; this led to no protein peak observed by LCMS in any of the samples. It could be possible that due to the bulk aspect of probe E the addition was happening but making the proteins insoluble. An SDS-PAGE was run of both the precipitate and supernatant for reactions run on myoglobin and RNase A. The SDS-PAGE confirmed that the protein had precipitated as no protein band was observed in the solution sample but only in the precipitate, Figure 21.

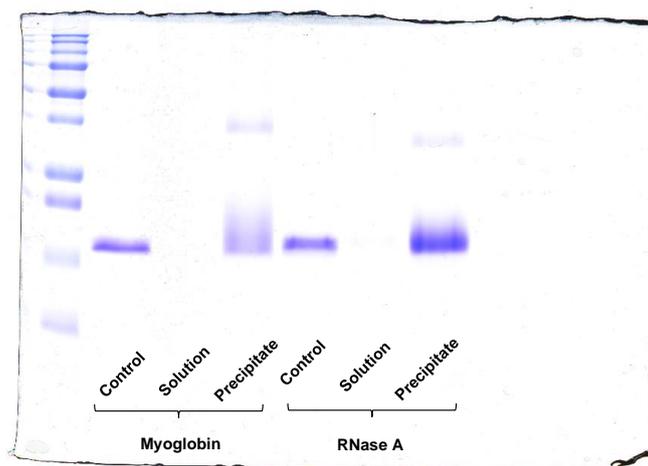


Figure 21: SDS-PAGE of myoglobin and RNase A mixture after the addition of probe E of both the solution and the precipitate formed

To conclude this part, it was shown that although some promising suggestions that PyOx-2PCA reagents could induce protein modification, the stability of these modifications, and the effect the probe had on protein stability, requires significant optimisation. First the 2PCA probe itself could be made more hydrophilic and less bulky such as the structure present in Figure 22. Second, fewer equivalents of the probe could be used to try and minimise precipitation. While only a few results were obtained, these preliminary results provide a direction of research to pursue further.

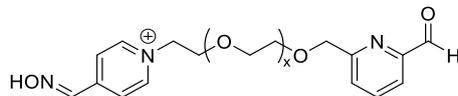


Figure 22: Potential new PyOx-2PCA probe

## 6.4 Summary

In this last Chapter, further study with PyOx-peptide probes was conducted. Very excitingly and for the first time reported to date, PyOx ligand-directed chemistry was achieved using peptide probes, modifying insulin with three types of NASA, hexanoic, biotin and NBD. It was also shown that the conversion depended on a lot of factors such as the NASA solubility, the position of the catalyst PyOx probe on the peptide, and the ratio between the different species. The results of modification were confirmed by three analysis techniques, LCMS, gel electrophoresis and MALDI. This last technique allowed much more in-depth analysis of where the modification took place. Only two samples were fully analysed, and this type of analysis could be used for a bigger range of samples to answer important remaining questions. On a general note, this first set of results showed up to 22% conversion which was promising, and further optimisation can be done to increase even higher the conversion. Some parameters to study would include the NASA hydrophilicity, the influence of temperature on modification, the catalyst valency, the location of modification depending on the orientation of the catalyst and the length of the linker used. A comparison to the background labeling to reinforce the importance of selectivity using the ligand-directed approach towards one specific amino-acid could be another important emphasis of distinction using specific modification via ligand-directed approach.

Finally, while very brief, the use of 2PCA-PyOx probes to modify protein was also attempted. While some potential was seen, especially with myoglobin, the peptide approach seemed more promising and more specific too, with more possibility as a range of peptide binders can be found via techniques such as phage display. These results though showed an important point mentioned in the introduction, that every protein has its own features and characteristic so one method will unlikely suit all targets.

## 6.5 Experimental

The spectrum of each probe described in this Chapter can be found in Chapter 4 for the NASA moieties and the different peptides, and in Chapter 5 for the 2PCA moieties. Similarly, methods for LCMS are detailed in Chapter 2.

### 6.5.1 General protocols for protein modification

#### Set-up for protein modification using peptide PyOx catalysts and NASA species

Insulin (final concentration: 50  $\mu$ M in phosphate buffer pH 7.4) was modified with PyOx-peptide probe at room temperature (initial concentration 1.0 mM in phosphate buffer pH 7.4) by adding specific volumes to finish with 1, 5 or 10 equivalents peptide relative to the insulin. NASA species, prepared as a 90 mM DMSO stock, was added to the mixture to finish with 1, 10 or 20 equivalents relative to insulin, final volume 102  $\mu$ L. The samples were incubated at room temperature and the conversion of the reaction was determined by LCMS at different time points. A deconvolution ladder from the BrukerAnalysis Software was used to identify different charged states and the relative deconvoluted mass.

#### Set-up for protein modification using 2PCA PyOx catalysts and NASA species

Proteins (myoglobin, RNase A, insulin; 100  $\mu$ M, 60  $\mu$ L) were modified with 2-PCA (10 mM, 100 equiv., 60  $\mu$ L) in Na phosphate buffer (50 mM, pH 7.5) and incubated at 37 °C for 23 h, and the conversion was determined by LCMS of the reaction mixture at this time point. Each sample was then split into four for each time point, 0 h, 3 h, 24 h and 72 h after NASA addition (10 mM prepared in 10% DMSO in Na phosphate buffer, 10 equiv., 2  $\mu$ L). The reaction mixtures were incubated at 37 °C. All aliquots for LCMS were diluted by a factor of 5 prior to injection (except for the 72 h timepoint, with no extra dilution).

### 6.5.2 General protocols for analysis of protein modification

#### Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis of proteins

Protein samples were used directly from the crude samples prepared in section 6.1 and diluted as indicated in Table 7 before being loaded onto pre-cast SDS-PAGE gels (BIO-RAD Mini-PROTEAN TGX Gel, 4-15% acrylamide). No heat or reducing was performed for insulin samples prior to loading. 5  $\mu$ L of a protein molecular weight marker (PageRuler Plus Prestained Protein Ladder, 10 to 250 kDa, Thermo Scientific) was loaded onto the gels, alongside the protein samples. The samples were separated by SDS-PAGE<sup>5</sup> at 200 V for 20 min. The gels were then either subjected to fluorescent imaging using a Syngene G:BOX Chemi XRQ equipped with a Synoptics 4.0 MP camera in line with

GeneSys software (Version 1.5.7.0) and then stained using InstaBlue protein stain for 30 min, or prepared for western blotting as detailed in following section.

Table 7: Assembly of protein samples for SDS-PAGE analysis

Reagent	Quantity ( $\mu\text{L}$ )
Protein ( $\mu\text{g}/\mu\text{L}$ )	5
Dye-free SDS non-reducing sample buffer (2% SDS, 4% glycerol, 40mM Tris-HCl pH 6.8)	2.5
H <sub>2</sub> O	5

#### Western blotting analysis of proteins

An Amersham Hybond P 0.2  $\mu\text{m}$  PVDF membrane (GE Healthcare Life Sciences) cut to the size of the gel was prepared by soaking in methanol for 30 s and then rinsing in ddH<sub>2</sub>O. The PVDF membrane and 3MM blotting papers were equilibrated in 1 x Transfer buffer for 15 min prior to blotting. Blotting was performed using an electroblot apparatus (BioRad, Hercules, CA) at 100 V, 350 mA for 1 h in cooled transfer buffer (25 mM Tris– HCl pH 8.3, 192 mM glycine, 0.1% SDS, 20% (v/v) methanol). Following transfer, the membrane was blocked for 1 h at room temperature in 5 ml of PBS + 0.1% v/v Tween 20 + 5% non-fat dry milk powder. The membrane was washed for 3 x 5 min in 5 ml of PBS + 0.1% v/v Tween 20. The membrane was incubated in anti-biotin HRP GTX77581 antibody (1/1000 diluted in PBS + 0.1% v/v Tween 20) for 1 h at rT, followed by washing in PBS + 0.1% v/v Tween 20. The membrane was then incubated with visualising substrate BCIP/NBT Alkaline Phosphatase Substrate Kit (Vector Labs, CA) until immunoreactive proteins on the membrane were visible (ca. 20 min). The reaction was stopped by washing the membrane in distilled water. The membrane was imaged using the Syngene G:BOX Chemi XRQ equipped with a Synoptics 4.0 MP camera, with GeneSys software (Version 1.5.7.0).

#### MALDI analysis of proteins

Protein or peptide solutions were diluted to 10  $\mu\text{M}$  (dilution by five) with aqueous water before reducing with 1  $\mu\text{L}$  aqueous 50 mM tris(2-carboxyethyl)phosphine (TCEP) and incubation at 60 °C for 1 h. Post reduction, proteins were alkylated with 0.5  $\mu\text{L}$  200 mM methyl methanethiosulfonate (MMTS), with incubation for 10 mins.

A 0.5  $\mu\text{L}$  aliquot of each peptide mixture was applied to a ground steel MALDI target plate, followed immediately by an equal volume of a freshly-prepared 10 mg/mL solution of 4-hydroxy- $\alpha$ -cyano-cinnamic acid (Sigma) in 50% aqueous (v:v) acetonitrile containing 0.1% , trifluoroacetic acid (v:v).

Positive-ion MALDI mass spectra were obtained using a Bruker ultrafleXtreme mass spectrometer in reflectron mode, equipped with a Nd:YAG smart beam laser. MS spectra were acquired over a range

of  $m/z$  900-7000. Final mass spectra were externally calibrated against an adjacent spot containing 6 peptides (des-Arg<sup>1</sup>-Bradykinin, 904.681; Angiotensin I, 1296.685; Glu<sup>1</sup>-Fibrinopeptide B, 1750.677; ACTH (1-17 clip), 2093.086; ACTH (18-39 clip), 2465.198; ACTH (7-38 clip), 3657.929.). Monoisotopic masses were obtained using the SNAP averaging algorithm (C 4.9384, N 1.3577, O 1.4773, S 0.0417, H 7.7583) and a S/N threshold of 2.

Precursors were manually selected to MS/MS fragmentation. Fragmentation was performed in LIFT mode without the introduction of a collision gas. The laser power and number of spectra summed was manually optimised for each precursor. The default calibration was used for MS/MS spectra, which were baseline-subtracted and smoothed (Savitsky-Golay, width 0.15  $m/z$ , cycles 4); monoisotopic peak detection used a SNAP averaging algorithm (C 4.9384, N 1.3577, O 1.4773, S 0.0417, H 7.7583) with a minimum S/N of 6. Bruker flexAnalysis software (version 3.3) was used to perform spectral processing and peak list generation.

### 6.5.3 Supplementary LCMS protein spectra

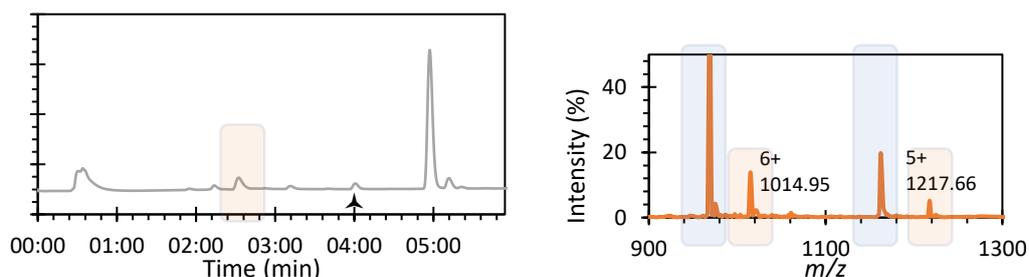


Figure 23: Chromatogram 210-400 nM of the crude mixture A-1-10 with NASA-NBD after 2 hrs reaction at room temperature. Spectrogram of the highlighted insulin peak with in blue the peak of the non-modified insulin and in orange the peaks with their corresponding masses of the modified-insulin with NBD.  $\blacktriangle$ UV peaks with no mass signals

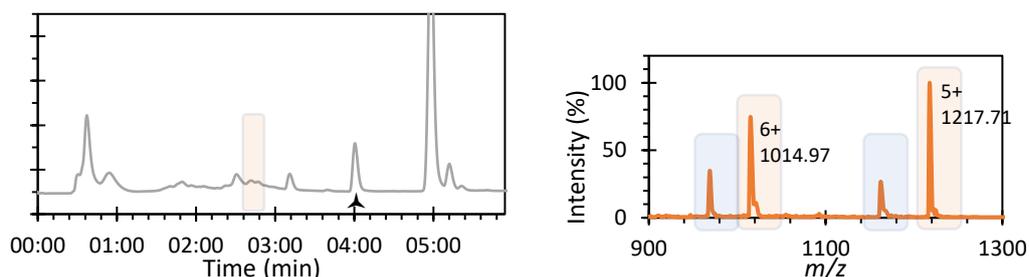


Figure 24: Chromatogram 210-400 nM of the crude mixture B-10-10 with NASA-NBD after 2 hrs reaction at room temperature. Spectrogram of the highlighted insulin peak with in blue the peak of the non-modified insulin and in orange the peaks with their corresponding masses of the modified-insulin with NBD.  $\blacktriangle$ UV peaks with no mass signals

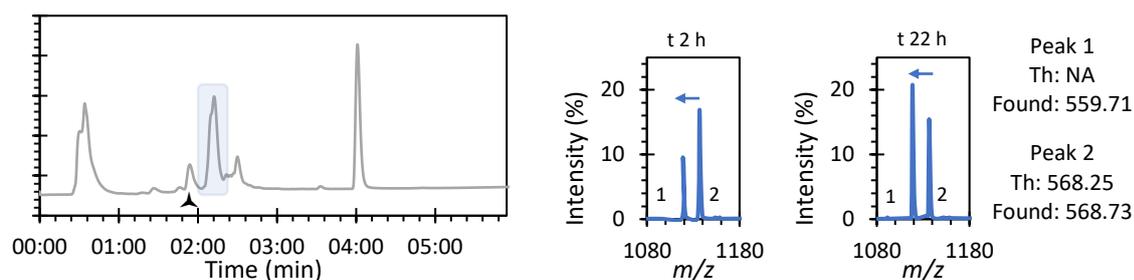


Figure 25: Chromatogram 210-400 nM of the crude mixture negative control-A-10-20 with NASA-biotin after 2 h reaction at room temperature. Spectrogram of the highlighted peptide A peak and the new peptide peaks appearing just before 3 min retention time. Peak 1: peptide peak minus 18 Da; Peak 2: peptide peak. Spectrogram showing the increase of the minus 18 Da species showing potential peptide degradation.  $\blacktriangle$  UV peaks with no mass signals

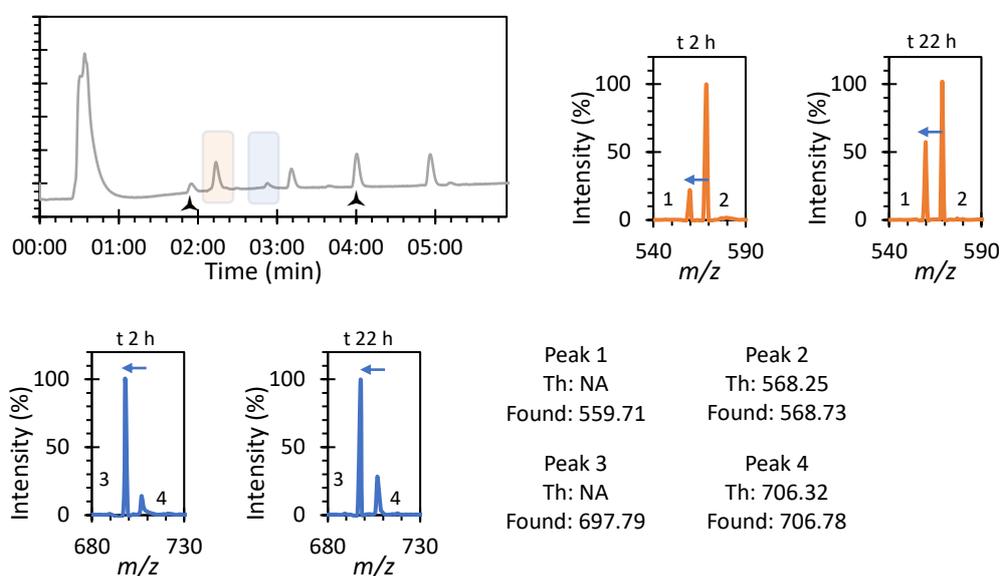


Figure 26: Chromatogram 210-400 nM of the crude mixture negative control-A-1-1 with NASA-NBD after 2 h reaction at room temperature. Spectrogram of the highlighted peptide A peak and the new peptide peaks appearing just before 3 min retention time. Peak 1: peptide peak minus 18 Da; Peak 2: peptide peak; Peak 3: peptide+NASA-NBD peak minus 18 Da; Peak 4: peptide+NASA-NBD peak. Spectrogram showing the increase of the minus 18 Da species showing potential peptide degradation.  $\blacktriangle$  UV peaks with no mass signals

## 6.6 Bibliography

- (1) Keijzer, J. F.; Firet, J.; Albada, B. Site-Selective and Inducible Acylation of Thrombin Using Aptamer-Catalyst Conjugates. *Chem. Commun.* **2021**, 57 (96), 12960–12963. <https://doi.org/10.1039/d1cc05446e>.
- (2) Tamura, T.; Song, Z.; Amaike, K.; Lee, S.; Yin, S.; Kiyonaka, S.; Hamachi, I. Affinity-Guided Oxime Chemistry for Selective Protein Acylation in Live Tissue Systems. *J. Am. Chem. Soc.* **2017**, 139 (40), 14181–14191. <https://doi.org/10.1021/jacs.7b07339>.
- (3) Chiang, H. L.; Ngo, S. T.; Chen, C. J.; Hu, C. K.; Li, M. S. Oligomerization of Peptides LVEALYL and RGFFYT and Their Binding Affinity to Insulin. *PLoS One* **2013**, 8 (6). <https://doi.org/10.1371/journal.pone.0065358>.
- (4) Barber, L. J.; Yates, N. D. J.; Fascione, M. A.; Parkin, A.; Hemsworth, G. R.; Genever, P. G.; Spicer, C. D. Selectivity and Stability of N-Terminal Targeting Protein Modification Chemistries. *RSC Chem. Biol.* **2022**, No. i, 56–64. <https://doi.org/10.1039/d2cb00203e>.
- (5) Laemmli, U. K. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **1970**, 227, 680–685.

## General Conclusions and Suggestions for Further Work

### General Conclusions

This last section of this thesis summarises the main conclusions and major outcomes of the work reported in the previous chapters. The strategies for modifying peptides into ligand-directed probes are highlighted with their application to modify protein of interest in a traceless manner. Similar work on 2-PCA was achieved for modifying protein selectively at their *N*-terminus. Different directions for further work not achieved in the time frame of this thesis are outlined in the following section.

Starting from known binding sequences to the two POI, the first requirement was to ensure that the peptide once modified at either of its terminus will still bind to the POI. Two different techniques were used, QCM-D and FP. They both showed that while most of the sequences of interest could be modified, two of them once modified at their *N*-terminus showed to be poor binders. While this result made the peptide in question not applicable for the direct application, it showcased very importantly how external modification can disrupt target binding presumably by blocking potential hydrogen bonding or changing the 3D structure of the peptide.

The main element of this thesis was then to establish routes to modify the still binding peptides into ligand-directed probes. Two types of probes were picked from the range of known chemistries detailed in Introduction. LDAI was a first choice due to the simplicity of probe design and synthesis. However, this chemistry showed to be challenging on peptide mainly due to its poor stability showing to be hydrolysable in a very short time frame. Second the use of NASA reagents was studied. Two routes were attempted to synthesis methylcyano-NASA, an organic based approach, and an on-solid support approach, both successful after optimisation. While the organic approach allowed to have an intermediate facilitating the general process, the on-solid support has the great advantage of requiring less purification steps using more equivalents of the required reagents. After optimisation of time, temperature, and concentration, we showed full conversion to the expected methylcyano-NASA on resin which has not yet been reported in literature. The only drawback of this chemistry on-solid support is that the process needs to be repeated for each sequence of interest, however automation could be easier to set-up on support. Finally, nitrobenzyl-NASA acyl donors exploited as a substrate for catalyst-based ligand-directed chemistry were successfully synthesised broadening the range of reagents known in literature. Their respective catalyst was once again synthesized on resin using click-chemistry in a straightforward approach leading to the synthesis of eight probes using one peptide sequence. Indeed the linker size between the catalyst and the peptide, as well as the valency of the catalyst, were varied to be able to look at their impact on protein labeling.

Following the work on peptide, *CatLD* chemistry was also applied to 2-PCA probe using click-chemistry leading to six final probes with once again different linker sizes.

We finished the work of this thesis by verifying the application of such modified probes for the labeling of protein. Very excitingly and for the first time reported to date, PyOx ligand-directed chemistry was achieved using peptide probes, modifying insulin with three types of NASA, hexanoic, biotin and NBD. The results of modification were confirmed by three analysis techniques, LCMS, gel electrophoresis and MALDI. This first set of results showed up to 22% conversion. The use of 2PCA-PyOx probes to modify protein was also attempted. While some potential was seen, especially with myoglobin, the peptide approach seemed more promising and more specific too, with more possibility as a range of peptide binders can be found via techniques such as phage display. These results though showed an important point mentioned in the introduction, that every protein has its own features and characteristic so one method will unlikely suit all targets.

### Suggestions for Further Work

While research on that matter is just at its beginning and we believe that the route ahead could impact massively the tissue engineering world, a few short-term and long-term suggestions are given below.

- 1) Different parameters such as the NASA solubility, the temperature, the pH should be envisaged to understand better the labelling of the protein and increase conversion rate. Kinetic study could also be undertaken for deeper knowledge.
- 2) MALDI analyses were completed only on RGFFYT probes modified by catalyst at their *N*-terminus. This technique allowed much more in-depth analysis and samples with the *C*-terminus being modified of the same sequence or samples using different linker size or different peptides should be run to answer important questions such as the localisation of the modification. A follow up point would be the comparison of the localisation of the background labeling observed which would emphasize the importance of the selectivity using peptides for ligand-directed approach.
- 3) While this point was a goal of this PhD, the timeframe did not allow for a full study comparing *OpLD* and *CatLD* chemistries. Several experiments could be envisaged to study in-depth the difference of action between those two chemistries and their respective kinetic and their amino acids preferential targets.
- 4) Similarly, to making sure that modified peptide could still bind their target, the modification of the protein will have to be studied to ensure that the protein is still functional keeping its activity within a biological environment.

- 5) Finally, two ultimate experiments could be run First, we will need to show that this type of reaction can happen in a complex environment showing the specificity of the designed peptide probe to its targeted protein. Second, after modification, we could use the new inserted handle on the protein for selective and control modification of a biomaterial to improve its biological function as represented in Figure 1.

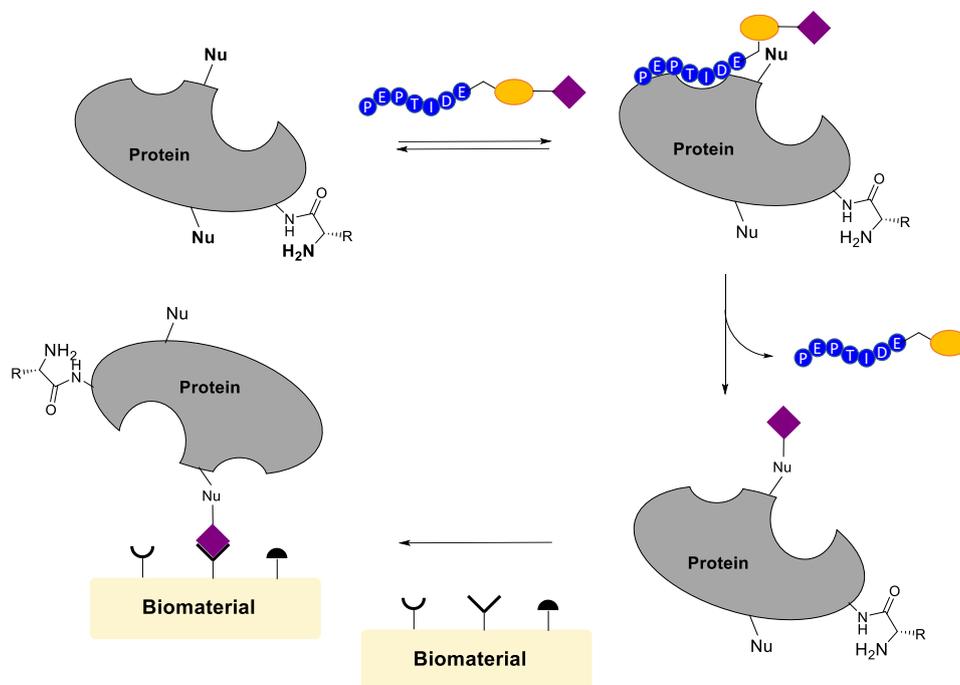


Figure 1: Tethering of a modified protein in a traceless manner using ligand-directed peptide probes to a biomaterial to subsequently enhance the biomaterial biological properties.

## Annex

Different attempts at QCM-D for insulin on gold sensor, temperature 20 °C:

	Rate μl/min	Peptide	Blocker	Concentration gradient								
<b>Exp 1 - S2</b>	50	Cys-P2	<b>3.1</b> (90%)	No								
<b>Exp 1 - S3</b>	50	None	<b>3.1</b> (100%)	No								
<b>Exp 1 - S4</b>	50	Cys-P2	None	No								
<b>Exp 3 -S1</b>	100	P2-Cys	None	Yes	1	5	10	15	25	50	100	
<b>Exp 3 -S2</b>	100	P2-Cys	<b>3.1</b>	Yes	1	5	10	15	25	50	100	
<b>Exp 3 -S3</b>	100	Cys-P2	None	Yes	1	5	10	15	25	50	100	
<b>Exp 3 -S4</b>	100	Cys-P2	<b>3.1</b>	Yes	1	5	10	15	25	50	100	
<b>Exp 4 -S1</b>	100	P2-Cys	<b>3.1</b>	Yes	0.05	0.1	0.5	1	5	10	50	100
<b>Exp 4 -S2</b>	100	Cys-P2	<b>3.1</b>	Yes	0.05	0.1	0.5	1	5	10	50	100
<b>Exp 4 -S3</b>	100	None	<b>3.1</b>	Yes	0.05	0.1	0.5	1	5	10	50	100