Semi-synthesis of glycoproteins

Bhavesh Premdjee

A thesis submitted to University College London in accordance with the requirements of the degree of Doctor of Philosophy in Chemistry

Declaration

I, Bhavesh Premdjee confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Bhavesh Premdjee
June 2014

Acknowledgements

I would like to thank Derek Macmillan for giving me the opportunity to work on this exciting interdisciplinary project. I am extremely grateful for his guidance (and patience) throughout my time as a student in his group. I would also like to express my sincere gratitude to Anna Adams, Ben Cowper and Samantha Gibson for their help and advice over the last few years. It has been enjoyable working alongside you. I also wish to thank Lisa Haigh, Vincent Gray, Kersti Karu, John Hill and Abil Aliev for support with mass spectrometry and NMR.

My time at UCL has been especially enjoyable thanks to 'Alpha Team'. Niral, Ahmed, Elena and Atif, it has truly been a privilege to share the journey of the last 8 years with you all. Thank you all for being there through the ups and downs, I could not ask for better friends.

I would like to thank Mum, Dad, Mayuri, Gary, Anita, Vijay, Antrix and Phoenix who have been supportive and understanding throughout my studies. Without you, I wouldn't have the determination to make it so far.

Sheetal, I could not have survived this journey without your love, support and encouragement. I am still amazed at the level of patience and understanding you have shown over the last few years. Thank you.

Abstract

Glycosylation is a prevalent form of post translational modification, believed to occur on over 50% of human proteins. Homogeneous forms of glycoproteins are essential for developing an understanding of how activity is mediated at a structural level. As biological origins of glycoproteins give rise to complex mixtures of glycoforms, homogeneous glycoprotein production has become an important goal.

As chemical protein synthesis is often limited to sequences of 30-50 residues, access to large native glycoproteins is currently restricted to fragment based approaches. Protein semi-synthesis enables the preparation of larger proteins which can be difficult to obtain through chemical synthesis alone. Consequently, a general semi-synthetic strategy towards *N*-glycoproteins has been proposed and demonstrated on Interferon-β-1 (IFNβ), a 166 residue glycoprotein. A three fragment strategy was designed, relying on the chemical synthesis of a short glycopeptide segment and recombinant expression of the two flanking domains. Homogeneity was established through the chemical synthesis of a glycopeptide containing a natively linked *N*-acetylglucosamine (GlcNAc), also enabling the selective transfer of complex oligosaccharides. After cloning and expression, the recombinant fragments were functionalised to allow assembly of the protein using Native Chemical Ligation. These desired protein modifications were achieved through the application of highly chemoselective reactions. These reactions were also applied towards the generation of *N*-glycopeptides compatible with the ligation strategy. Further to this, existing methods enabling the direct synthesis of functionalised *N*-glycopeptides were also explored. After glycopeptide synthesis, endoglycosidase A enabled the transfer of oligosaccharides to the *N*-acetylglucosamine motif. This has allowed the preparation of the desired IFNβ glycopeptide as well as a glycosylated variant of glucagon like peptide-1.

To expand the utility of endoglycosidase methodology, a novel sugar nucleotide was synthesised to facilitate the incorporation of a sialyl galactose mimic onto *N*-glycans. The resulting oligosaccharides may serve as novel substrates for endoglycosidases in the preparation of *N*-glycoprotein mimics.

Current publications from work presented in this thesis:

Bhavesh Premdjee, Anna Adams and Derek Macmillan: **Native *N*-glycopeptide thioester synthesis through *N→S* acyl transfer,** Bioorganic and Medicinal Chemistry Letters, September 2011; 21(17), 4973–4975

Derek Macmillan, Anna Adams and Bhavesh Premdjee: **Shifting Native Chemical Ligation into Reverse through *N→S* Acyl Transfer (Review),** Israel Journal of Chemistry Special Issue: Chemical Protein Synthesis, November 2011; 51(8-9), 885–899

Anna Adams, Ben Cowper, Rachel Morgan, Bhavesh Premdjee, Stephen Caddick and Derek Macmillan: **Cysteine Promoted C-Terminal Hydrazinolysis of Native Peptides and Proteins,** Angewandte Chemie International Edition, December 2013; 52(49), 13062–13066

Contents

List of Figures

List of Schemes

Abbreviations

Ac	acetyl
Acm	acetamidomethyl
BAL	backbone amide linker
Boc	*tert*-butyloxycarbonyl
BOP	benzotriazole-1-yl-oxy-tris(dimethylamino)phosphonium hexafluorophosphate
CBD	chitin binding domain
CG	carboxymethyl galactose
CMP	cytidine monophosphate
CNX	calnexin
CRT	calreticulin
DCC	*N,N'*-dicyclohexylcarbodiimide
DCM	dichloromethane
DIC	*N,N'*-diisopropylcarbodiimide
DIPEA	*N,N*-diisopropylethylamine
DMF	*N,N*-dimethylformamide
DMC	2-chloro-1,3-dimethylimidazolinium chloride
Dmmb	2-mercapto-4,5-dimethoxybenzyl
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
Dol-P	dolichyl phosphate
DTT	dithiothreitol
EDC	*N*-ethyl-*N*'-(3-dimethylaminopropyl)carbodiimide
EEDQ	*N*-ethoxycarbonyl-2-ethoxy-1,2-dihydroquinoline
Endo A	*endo*-β-*N*-acetylglucosaminidase isolated from *Arthrobacter Protophormiae*
EPL	expressed protein ligation
EPO	erythropoietin
ER	endoplasmic reticulum
ERGIC	ER-Golgi intermediate compartment
ESI	electrospray ionisation
FAB	fast atom bombardment
FMDV	foot-and-mouth disease virus

Fmoc	9-fluorenylmethoxycarbonyl
Gal	galactose
GDP	guanosine diphosphate
Glc	glucose
GlcNAc	*N*-acetylglucosamine
GLP-1	glucagon like peptide-1
GlyCAM-1	glycosylation dependant cell adhesion molecule-1
GST	glutathione-*S*-transferase
GT	glycosyltransferase
Gu.HCl	guanidine hydrochloride
HATU	*O*-(7-azabenzotriazol-1-yl)-*N,N,N',N'*-tetramethyluronium hexafluorophosphate
hBNP	human B-type natriuretic peptide
HBTU	*O*-(Benzotriazol-1-yl)-*N,N,N',N'*-tetramethyluronium hexafluorophosphate
HFIP	hexafluoroisopropanol
HOAt	1-hydroxy-7-azabenzotriazole
HOBt	1-hydroxybenzotriazole
IFNβ	Interferon β-1
IPTG	*iso*-propyl-β-D-thiogalactopyranoside
LC-MS	liquid chromatography mass spectrometry
Lev	levulinoyl
LIC	ligation independent cloning
Man	mannose
MAP	methionylaminopeptidase
MCP-3	monocyte chemotactic protein-3
MESNa	sodium 2-mercaptoethanesulfonate
MFD	metal-free dethiylation
MPA	3-mercaptopropionic acid
MPAA	4-mercaptophenylacetic acid
MS	mass spectrometry
NBS	*N*-bromosuccinimide
NCL	native chemical ligation
NDP	nucleoside diphosphate
NMP	*N*-methyl-2-pyrrolidone

NMR	nuclear magnetic resonance
Npys	3-nitro-2-pyridinesulfenyl
OST	oligosaccharyltransferase
Pbf	2,2,4,6,7-pentamethyl-dihydrobenzofuran-5-sulfonyl
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PG	protecting group
PMB	*para*-methoxybenzyl
PyBOP	benzotriazole-1-yl-oxy-tris-pyrrolidinophosphoniumhexafluorophosphonate
PyBrOP	bromo-*tris*-pyrrolidino phosphonium hexafluorophosphate
(RP)HPLC	reversed phase high performance liquid chromatography
RPM	revolutions per minute
RVG	rabies virus glycoprotein
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SPPS	solid phase peptide synthesis
Su	succinimide
TBAB	tetrabutylammonium bromide
TBAHS	tetrabutylammonium hydrogensulfate
TBPP	tetrabenzyl pyrophosphate
^{t}Bu	*tert*-butyl
TCEP	*tris*(2-carboxyethyl)phosphine
TES	triethylsilane
TEV	tobacco etch virus
TFA	trifluoroacetic acid
TfOH	triflic acid
TIPS	triisopropylsilane
TLC	thin layer chromatography
TMS	trimethylsilyl
Tris	*tris*(hydroxymethyl)aminomethane
Trt	trityl
UDP	uridine diphosphate
UGGT	UDP-glucose:glycoprotein glucosyltransferase

Amino acid single and three letter codes:

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

1 *N*-linked glycoproteins

1.1 Introduction

Glycosylation is a co and post-translational modification of proteins believed to occur on over 50% of proteins in humans, giving rise to important biological properties.[1] This modification serves to assist protein folding, mediate both cell-cell recognition and immune defence whilst improving the therapeutic potential of the protein.[2, 3] Carbohydrate structures also contribute to the hydrodynamic volume of glycoproteins, consequently prolonging circulation *in vivo.* In the absence of glycosylation, proteins can aggregate and fold incorrectly, rendering them unable to fulfil their functional role.[3] Glycans also enhance the thermal stability of proteins, improve solubility and prevent degradation by proteolysis.[4]

Glycosylation forms a covalent linkage between a carbohydrate and protein, most commonly to the amide nitrogen of an asparagine side chain. This *N*-linkage is observed at Asn-X-Ser or Asn-X-Thr motifs (where X is any amino acid except proline) and is largely abundant in eukaryotic cells. *O*-linked glycosylation is another prevalent form of glycosylation, formed through hydroxyl groups of serine or threonine residues.[5]

Conversely to protein biosynthesis where assembly is template mediated, the oligosaccharide structures observed on complex *N*-linked glycoproteins are built up from smaller carbohydrate building blocks through a series of enzymatic steps.[6] The vast assortment of building blocks and linkages can give rise to a number of possible glycan structures. Assembly of the glycan through a cascade of enzymatic steps is subject to substrate/acceptor competition, leading to varied incorporation of building blocks. As a consequence, glycoproteins are isolated as heterogeneous mixtures containing different ‘glycoforms’, exhibiting the same protein backbone but different glycosylation patterns.[7]

1.2 *N*-linked glycoprotein biosynthesis

N-linked glycosylation is the attachment of a carbohydrate through the amido group of an asparagine and is found to be more prominent than *O*-linked glycosylation.[8] In mammalian cells, the synthesis of the glycan commences on the cytosolic surface of the ER (Figure 1.1). Individual monosaccharide building blocks are added in a stepwise fashion to the lipid carrier, dolichylphosphate. The first seven sugars (two *N*-acetylglucosamine and five mannose residues) are enzymatically transferred by glycosyltransferases from sugar nucleotide precursors UDP-GlcNAc and GDP-Man. The next seven residues (four mannose and three glucose) are attached in the ER lumen from lipid derived intermediates, Dol-P-Man and Dol-P-Glc.

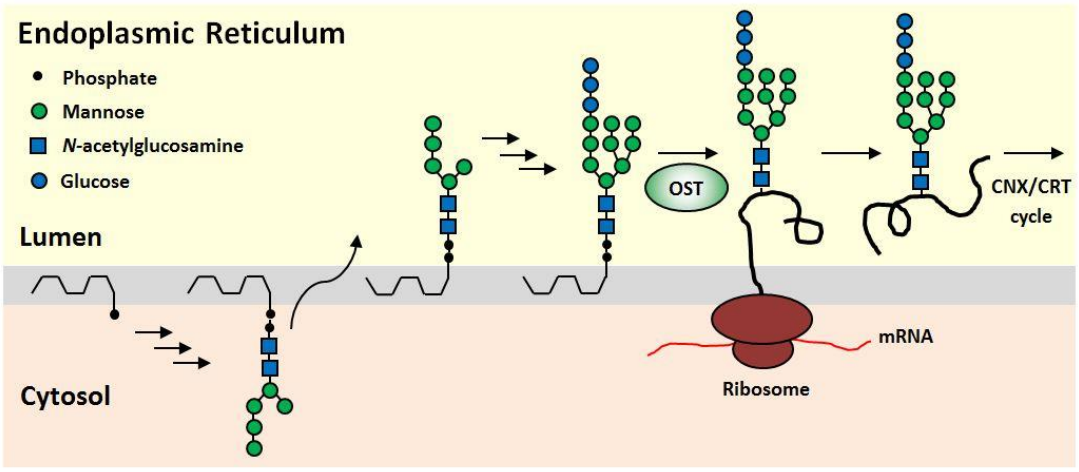


Figure 1.1: *N*-glycoprotein biosynthesis in the ER

The resulting oligosaccharide is then transferred to a growing polypeptide by an oligosaccharyltransferase (OST) and covalently binds to the asparagine side chain.[6] In a recent study, the X-ray crystal structure of a bacterial OST has elucidated the critical residues involved in this reaction. The co-crystal structure with a model hexapeptide 'DQNATF' suggests that the β-hydroxyl of threonine forms three hydrogen bonds to a 'WWD' motif in the catalytic subunit of the OST (Figure 1.2 a).[9] Critical to the function of the OST, a divalent cation coordinates the catalytic residues D56, D154 and E319 to activate the acceptor asparagine (Figure 1.2 b and c). Interestingly, the amide is activated sufficiently to act as a nucleophile and attacks C-1 of the oligosaccharide to form the *N*-glycosidic bond. It was

consequently proposed that the unique reactivity of the amide arises from the twisting of the C-N bond. The rotation out of planarity is facilitated by hydrogen bonding with catalytic residues D56 and E319.

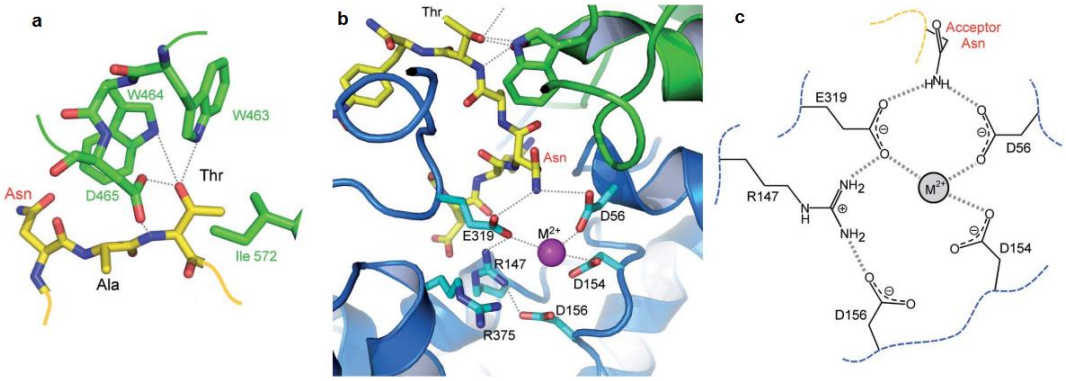


Figure 1.2: Mechanism for glycan transfer in the ER[9]

Following attachment of the glycan to the growing polypeptide, terminal glucose residues are trimmed by glucosidases I and II, to form a high mannose intermediate, typically $GlcMan_9GlcNAc_2$.[10] The resulting intermediates are bound by specific lectins calnexin (CNX) and/or calreticulin (CRT) (Figure 1.3), bringing the glycoprotein into close proximity with cochaperone oxidoreductase ERp57.[11-13] The oxidoreductase forms disulfide bonds with the substrate, enabling the formation of correctly paired disulfides through subsequent oxidation and isomerisation reactions.[14] The cleavage of the final Glc residue by glucosidase II causes the complex to dissociate, releasing the glycoprotein. Correct folding is controlled by UDP-glucose:glycoprotein glucosyltransferase (UGGT), which transfers glucose units to misfolded glycoproteins, forcing re-entry to the CNX/CRT cycle for refolding.[15] This process is repeated until the protein is correctly folded or degraded. The correctly folded, deglucosylated high mannose intermediates are bound by lectins ERGIC-53 or VIPL, which facilitate transport to the Golgi complex.[16]

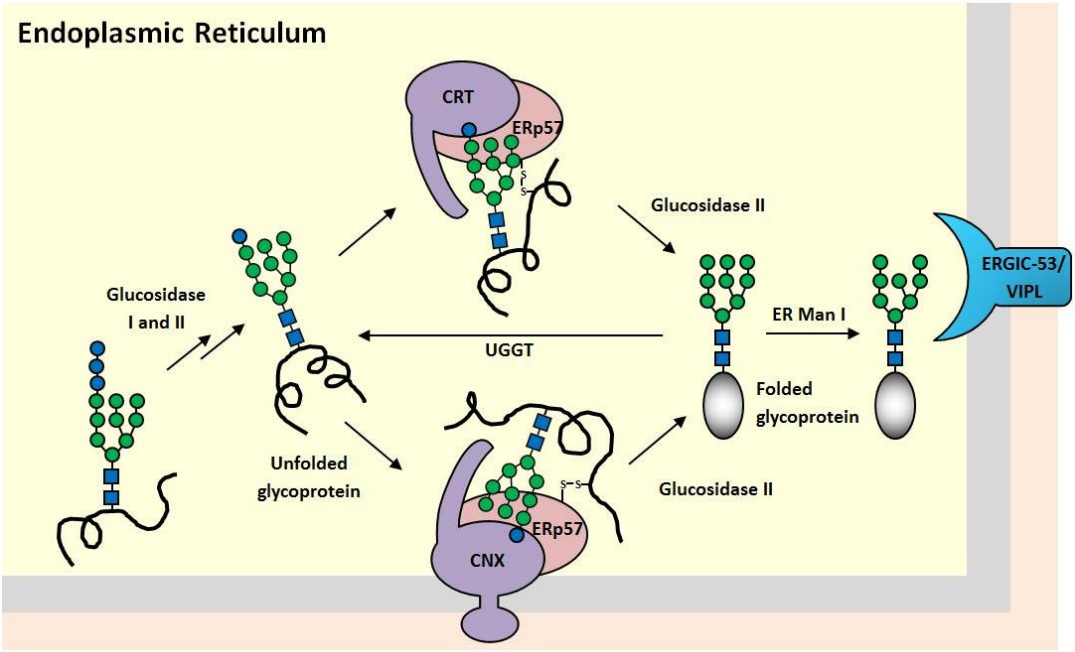


Figure 1.3: The Calreticulin-Calnexin cycle

At the Golgi complex (Figure 1.4), the glycoprotein undergoes trimming of mannose residues. An additional *N*-acetylglucosamine is added, after which further mannose trimming takes place. During terminal glycosylation, additional sugar residues are added onto the glycan structure, including *N*-acetylglucosamine, galactose, sialic acid and fucose. Figure 1.4 shows one pathway in the final stages of glycosylation, but alternative terminal glycosylation patterns are known to occur as a result of varied incorporation of sugars, furnishing a range of glycans (Figure 1.5).[17, 18]

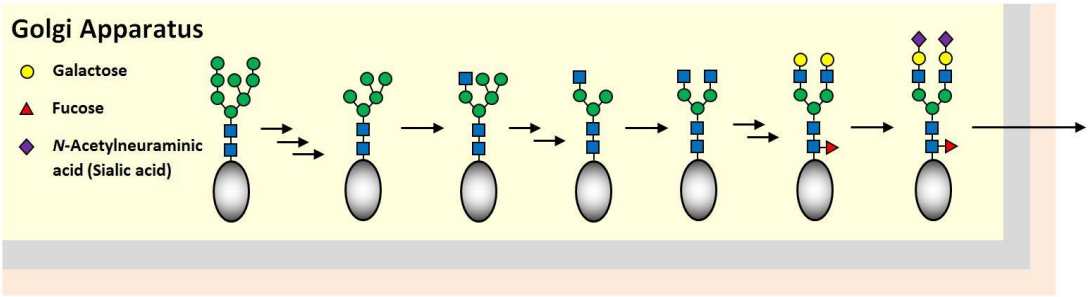


Figure 1.4: *N*-glycoprotein biosynthesis in the Golgi

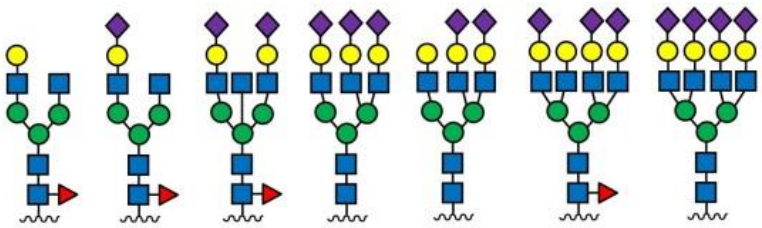


Figure 1.5: Examples of *N*-glycans isolated from human cells[18]

To gain a reliable understanding of how glycoprotein activity is mediated at a structural level, it is critical that studies are performed on single homogeneous forms. Glycoproteins from natural biological sources are largely unsuitable, as the phenomenon of microheterogeneity gives rise to inseparable mixtures of glycoforms.[19] Accordingly, much interest has grown in the chemical synthesis of homogenous glycoproteins.

1.3 Strategies towards the synthesis of homogeneous *N*-linked glycoproteins

To meet the demand of homogeneous *N*-glycoprotein synthesis, it is essential that glycans are accessible as defined single structures. The synthesis of such elaborate structures has been achieved by total chemical synthesis,[20-23] chemoenzymatic synthesis[24, 25] and technologies have also been developed to allow isolation of complex glycans from natural sources.[26] Once obtained, the incorporation of oligosaccharides into native *N*-glycoproteins is commonly achieved by one of three strategies (Figure 1.6).

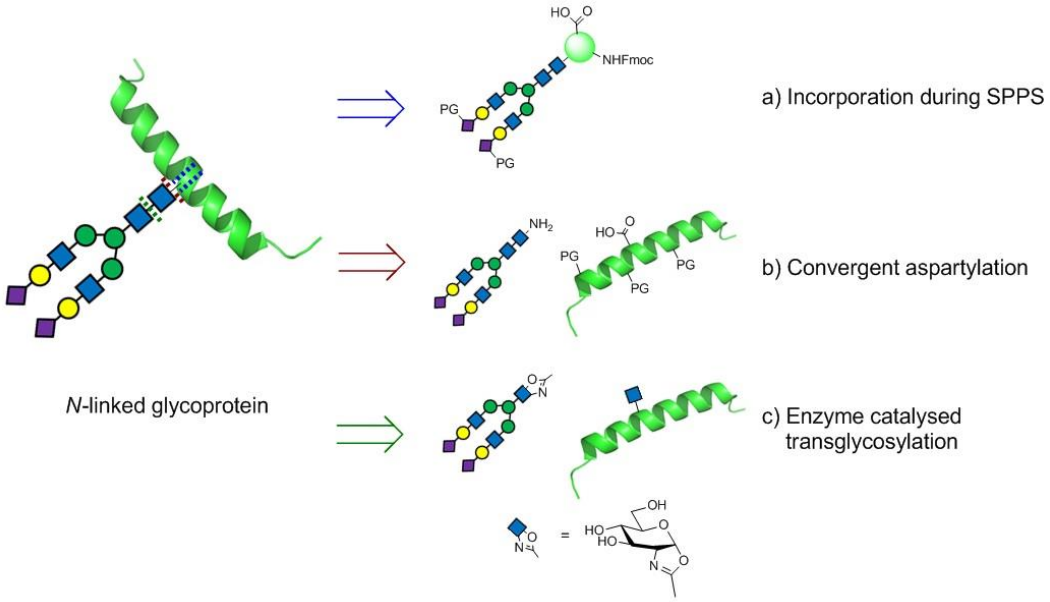
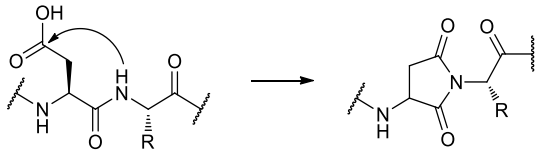


Figure 1.6: Retrosynthetic analysis highlighting synthetic routes towards *N*-glycoproteins

The incorporation of complex glycans can be achieved by the insertion of suitably protected glycoamino acid "cassettes" during solid phase peptide synthesis (SPPS) (Figure 1.6 a). Pioneered by the Kajihara and Unverzagt groups, this has been developed to enable compatibility with Boc and Fmoc methods.[27-29] During subsequent elongation steps, it is often necessary to protect carbohydrate hydroxyls by acetylation to prevent unwanted side reactions. However, it has been found that large acetylated glycans can impede peptide synthesis.[30] As an alternative, the GlcNAc-Asn linkage is formed after the completion of peptide synthesis. This convergent strategy relies on the orthogonal deprotection of the aspartic acid residue at the desired site of glycosylation, followed by its reaction

with a glycosyl amine (Figure 1.6 b). Originally developed by Lansbury,[31] this strategy has been applied to the synthesis of a number of *N*-glycoproteins.[32-37] Until recently,[38, 39] a major drawback encountered with this method was the formation of aspartimide by-products (Scheme 1.1).[40, 41]



Scheme 1.1: Formation of aspartimide by-products from the cyclisation reaction with protein backbone

Figure 1.6 c highlights a strategy that relies on the enzyme catalysed formation of the GlcNAcβ(1-4)GlcNAc chitobiose core present in *N*-glycans.[42] Endohexosaminidases have been shown to process sugar oxazolines as activated donors to promote transglycosylation reactions to GlcNAc.[43-46]

Chemical assembly of proteins is most commonly achieved using solid phase synthesis, although is not routinely performed for sequences larger than 30 residues. Therefore the synthesis of larger proteins relies heavily on methods which allow the conjugation of smaller fragments. The most distinguished method, Native Chemical Ligation (NCL) allows the chemoselective reaction of unprotected protein segments, resulting in the formation of a native amide bond. Fragment based approaches allow the incorporation of complex glycopeptides and have been successfully applied towards the synthesis of large glycoproteins which were otherwise inaccessible from traditional chemical methods. Despite impressive efforts,[30, 47-49] homogeneous glycoprotein synthesis still presents a significant challenge and has not advanced at the same rate as unglycosylated protein synthesis.

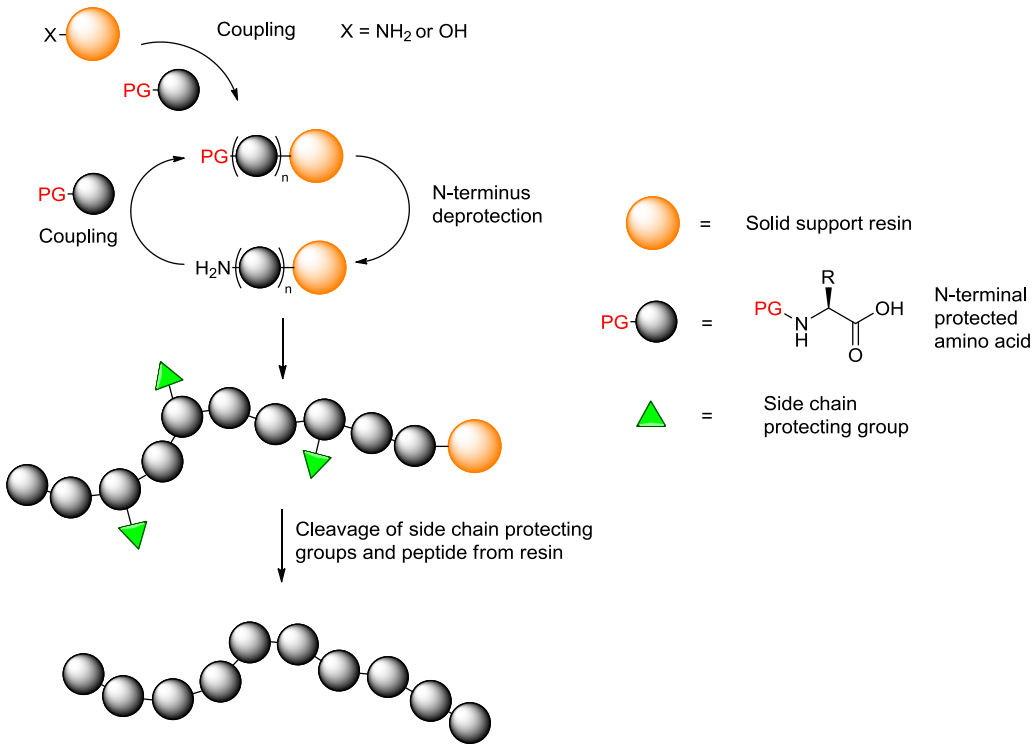
NCL can be extended to the semi-synthesis of proteins by ligating chemically synthesised fragments with proteins obtained from biosynthetic methods. Recombinant expression of unmodified proteins can obviate the laborious synthesis of large fragments, which can be difficult to obtain by chemical synthesis alone. Therefore a semi-synthetic strategy towards native *N*-glycoproteins may provide a significant advantage to existing methodology.

1.4 Chemical Protein Synthesis

Traditionally, peptide syntheses were carried out using fully protected amino acids in organic solvents. However, during attempts to synthesise longer polypeptide chains, many shortcomings of the 'solution phase' strategy were realised. With increasing chain length, fully protected peptides suffered from poor solubility, subsequently leading to poor coupling reactions. Another difficulty encountered was the inability to purify and characterise the protected peptides.[50] Multiple recrystallisations were not reliable enough to guarantee a pure product and characterisation would require deprotection of the peptide. Activation of the C-terminal amino acid under basic conditions was also accompanied by racemisation. Despite the problems associated with solution phase peptide synthesis, short peptide sequences have been successfully synthesised employing this method.[51]

1.5 Solid phase peptide synthesis

In 1963, Bruce Merrifield introduced a remarkable concept which overcame difficulties encountered with solution phase synthesis.[52] His strategy utilised an insoluble polymeric support to which a C-terminal amino acid was covalently attached. Removal of the N(α) amino protecting group could be achieved selectively, and the resin bound amino acid could be washed and isolated by filtration. The next N(α) amino protected, carboxyl activated residue could then be added, resulting in the formation of a peptide bond. After completion of the coupling, the excess amino acid and soluble by-products could simply be washed away to essentially yield a pure peptide. The process could then be repeated until the target peptide had been constructed. In the final step the peptide could be cleaved from the solid support, simultaneously removing any side chain protecting groups. This method has proved to be an extremely useful strategy, providing access to peptides of higher purity with better yields. Conveniently, this method is also suitable for automation by specialised machines. The most commonly utilised methods employ Boc (*tert*-butyloxycarbonyl) or Fmoc (9-fluorenylmethoxycarbonyl) for temporary protection of the N(α) amino group, with each strategy having advantages and limitations.



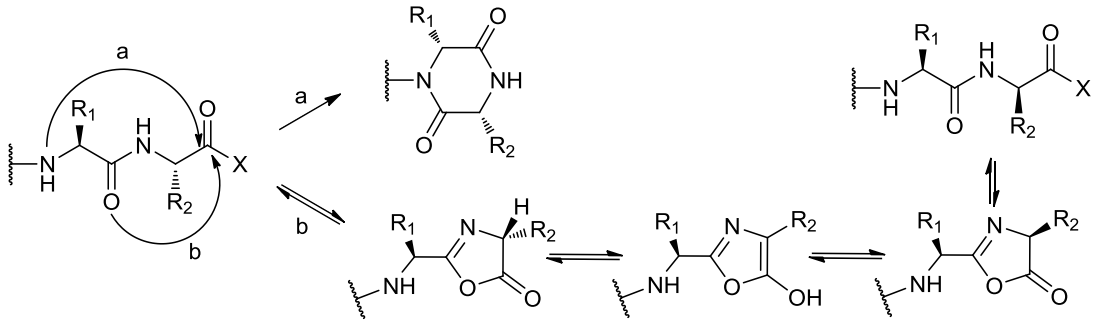
Scheme 1.2: Solid phase peptide synthesis

Merrifield's evolved SPPS method relied on N(α) Boc protection and benzyl-based protection for side chain functionalities. Boc removal is achieved with TFA and after neutralisation of the resulting ammonium salt, peptide elongation can be continued. Cleavage from the resin and concomitant deprotection of side chains is most commonly carried out by acidolysis using anhydrous hydrofluoric acid. This step highlights the major drawback of the Boc method, as the extreme toxicity and volatility of HF limits its use in laboratories without specialised equipment. Furthermore, repeated exposures to TFA and finally HF renders peptides with acid sensitive functionalities such as glycosidic linkages incompatible with this approach.

Due to the complications of Boc SPPS, researchers sought an alternate strategy that would obviate the repeated exposure of peptides to acidic conditions and the use of hazardous reagents. This demand was met by the Fmoc group,[53] and was first applied in SPPS by Meienhofer.[54] The Fmoc strategy employs the use of orthogonal protecting groups, where the N(α) amino group is deprotected by mild treatment with piperidine. The peptide-resin linkage and side chain protecting groups are stable to these conditions and are only cleaved by TFA on completion of peptide synthesis.

1.6 Peptide bond formation

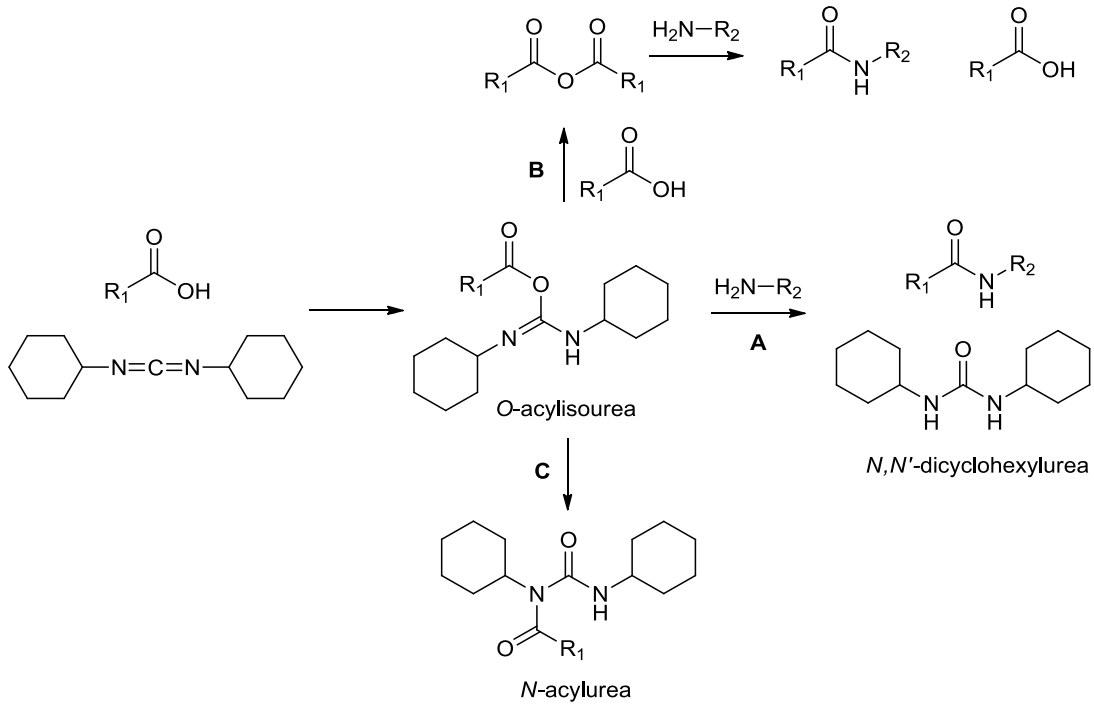
The progression of SPPS has been closely reliant on reactions which allow rapid coupling of amino acids. Carboxylic acid and amine functionalities do not react with each other spontaneously, and in extreme cases can require high temperatures.[55] Therefore, in most cases amide bond formation is facilitated by the activation of carboxylic acids to promote acylation of the free amine. A vast number of reagents and methods have been developed, a selection of which will be elaborated upon in this chapter. First and foremost, it is important to appreciate that SPPS is usually carried out in the *C* → *N* direction to prevent undesired side reactions resulting from the activation of resin bound carboxyl groups (Scheme 1.3).[56] Peptide synthesis in the *N* → *C* direction can result in attack of the activated carboxyl group by the adjacent amide bond, resulting in the formation of a diketopiperazine. Another possible side reaction commences with cyclisation to an oxazolone, which is followed by enolisation and the loss of the proton from the α-carbon. Accordingly, this pathway is capable of promoting racemisation, as the reprotonation step can take place from either face.



Scheme 1.3: Diketopiperazine formation and racemisation during SPPS in the N → C direction

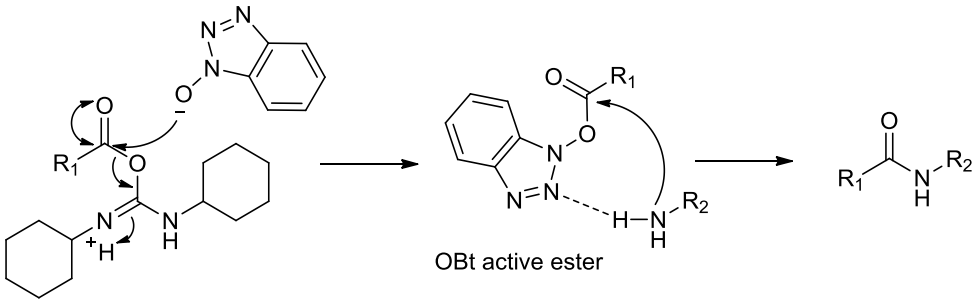
N,N'-dicyclohexylcarbodiimide (DCC) was one of the first coupling reagents to find use in peptide synthesis.[57] The reaction of DCC with carboxylic acids gives rise to a number of possible reaction pathways (Scheme 1.4). Initially an *O*-acylisourea is formed and this highly reactive intermediate is susceptible to direct aminolysis (path A). An excess of carboxylic acid leads to the formation of a symmetrical anhydride which is also capable of acylating an amine (path B). A drawback of this activation method is the rearrangement to the *N*-acylurea. This results in the

irreversible consumption of the carboxylic acid (path C). The by-product of the coupling reaction is *N,N'*-dicyclohexylurea, which is solubilised with TFA or methanol. Alternative carbodiimides have also been developed including *N,N'*-diisopropylcarbodiimide (DIC) and *N*-ethyl-*N'*-(3-dimethylaminopropyl)carbodiimide (EDC).[58]



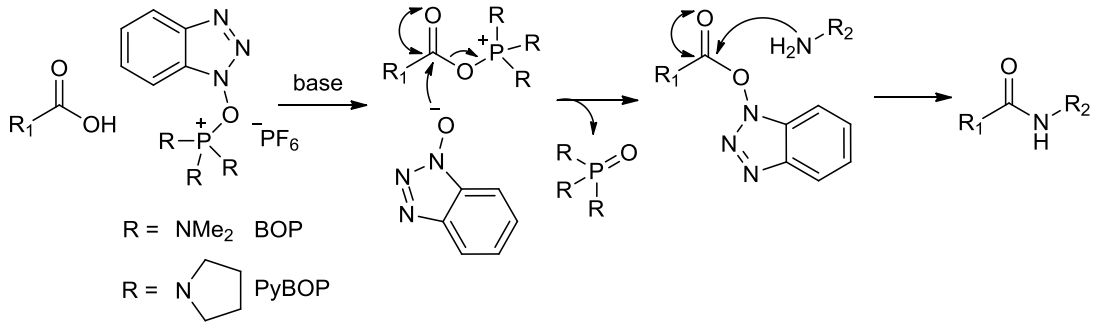
Scheme 1.4: Carbodiimides in peptide coupling reactions

It was later discovered that the *O*-acylisourea intermediates were prone to oxazolone formation and subsequently gave rise to racemisation.[59] This was shown to be minimised by the addition of 1-hydroxybenzotriazole (HOBt), which also improved yields.[60] The protonation of the *O*-acylisourea by HOBt prevents the intramolecular rearrangement to the *N*-acylurea and generates the corresponding active ester. The reactivity is further enhanced by the ability to stabilise the approaching amino group by hydrogen bonding. A derivative of HOBt, 1-Hydroxy-7-azabenzotriazole (HOAt) was found to be superior in terms of yield and ability to reduce racemisation.[61] A number of alternative additives have also been developed for carbodiimide coupling.[62]



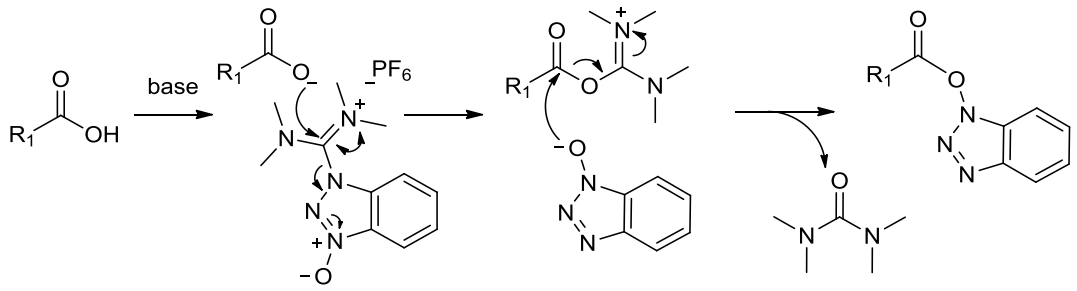
Scheme 1.5: HOBt assisted bond formation in DCC coupling

The use of HOBt and HOAt inspired the development of other benzotriazole reagents for generation of active esters. Castro reported Benzotriazole-1-yl-oxy-tris(dimethylamino)phosphonium hexafluorophosphate (BOP), which initially produces the acylphosphonium species before interception by HOBt to form the active ester (Scheme 1.6).[63] The resulting hexamethylphosphoric triamide by-product is extremely toxic and consequently BOP has been superseded by Benzotriazole-1-yl-oxy-tris-pyrrolidinophosphonium hexafluorophosphonate (PyBOP).[64]



Scheme 1.6: BOP and PyBOP assisted amide bond formation

The formation of active esters is also achieved through the use of uronium/aminium salts such as *O*-(Benzotriazol-1-yl)-*N,N,N′,N*′-tetramethyluronium hexafluorophosphate (HBTU) (Scheme 1.7).[65] A number of derivatives have been reported,[62] though it has been suggested that their use does not result in significant differences.[66]



Scheme 1.7: HBTU activation of carboxylic acid

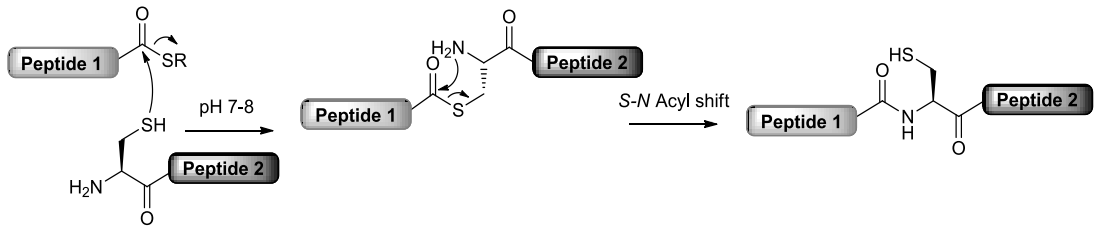
1.7 Peptide ligation

Despite being an excellent tool, the efficiency of SPPS for peptide assembly decreases for longer sequences due to the accumulation of by-products and impurities. With many peptides and proteins of interest containing over 30 amino acids, researchers sought a method to overcome the shortcomings of SPPS. It was envisaged that the most promising strategy would enable smaller unprotected peptide fragments to be joined together chemoselectively.

Schnolzer and Kent were able to first demonstrate chemoselective ligation of two unprotected peptides via a thioester linkage in the synthesis of HIV-1 protease.[67] Other chemical ligation strategies also emerged, all able to conjugate two unprotected peptides through a non-native peptide bond.[68, 69] However peptides containing non-native structures were criticised and deemed unreliable for studies on protein structure and function. Hence, further investigations were made into a ligation strategy that would furnish a native peptide bond.

1.8 Native Chemical Ligation

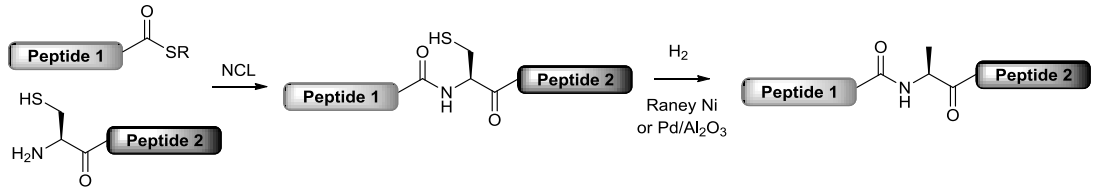
In 1994, the Kent group reported the synthesis of proteins by Native Chemical Ligation.[70] Based on a reaction first described by Wieland,[71] this significant innovation allowed ligation of two unprotected peptide segments through a native peptide bond. Traditionally in the first step of NCL, a C-terminal peptide thioester undergoes transthioesterification with the thiol group of an N-terminal cysteine. The intermediate rearranges via a spontaneous intramolecular *S*→*N* acyl transfer, resulting in the formation of an amide bond (Scheme 1.8). The reaction is typically performed at neutral pH under denaturing conditions and can be catalytically enhanced by aryl thiols.[72] The widespread use and success of NCL has also been accompanied by the development of a number of alternative ligation strategies.[73-79] However, the employment of mild aqueous conditions and remarkable chemoselective reactivity in the absence of protecting groups has distinguished NCL as the most popular method for assembling proteins from smaller fragments.



Scheme 1.8: Native Chemical Ligation

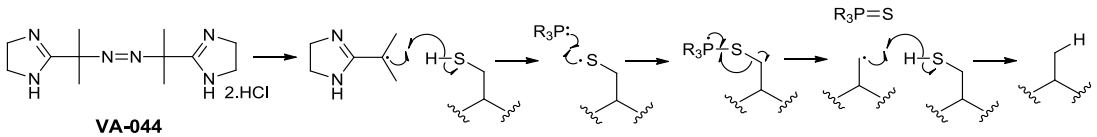
An important consideration for NCL strategies is the relatively low abundance of cysteine in protein sequences. Furthermore, the position of the Xaa-Cys junction is not always at a strategically favourable position. Although the substitution of cysteine for non-critical residues can be well tolerated without loss of protein activity,[80] methods have been developed to overcome this limitation.

One of the simplest innovations was the use of cysteine surrogates. After participation in NCL, these residues can be converted to native structures through desulfurisation or alkylation.[81, 82] This potential of desulfurisation was first demonstrated by Yan and Dawson, who converted cysteine to alanine after ligation using Raney nickel or palladium/aluminium oxide (Scheme 1.9).[83]



Scheme 1.9: Desulfurisation of cysteine to alanine after NCL

Such conditions however, can lead to the hydrogenation of tryptophan and demethythiolisation of methionine. To avoid the use of metal-based catalysts, Danishefsky developed a free radical approach to desulfurisation using TCEP and water soluble VA-044 (2,2'-azobis(2-(2-imidazolin-2-yl)propane) dihydrochloride) as a radical initiator (Scheme 1.10).[84]



Scheme 1.10: Proposed mechanism for VA-044/TCEP mediated desulfurisation

Metal-free dethiolation (MFD) was considered to be a milder treatment with improved selectivity and vastly expanded the applicability of ligation-desulfurisation methods. Consequently, various thiolated amino acids have been introduced to broaden the scope of NCL (Figure 1.7). The method was further extended by Dawson, who showed selenocysteine could be selectively deselenised to alanine in the presence of cysteine residues.[85] A number of other selenol containing amino acid derivatives have also been developed for NCL.[86]

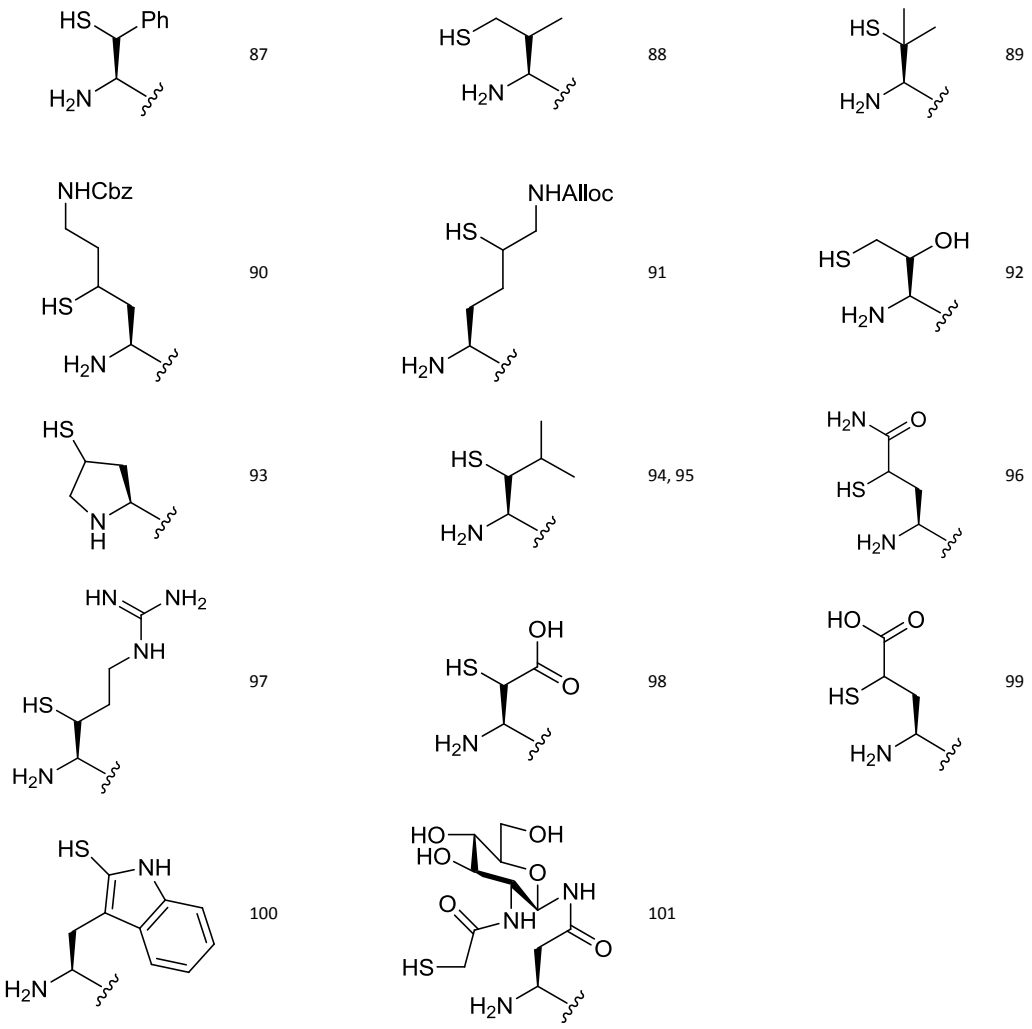
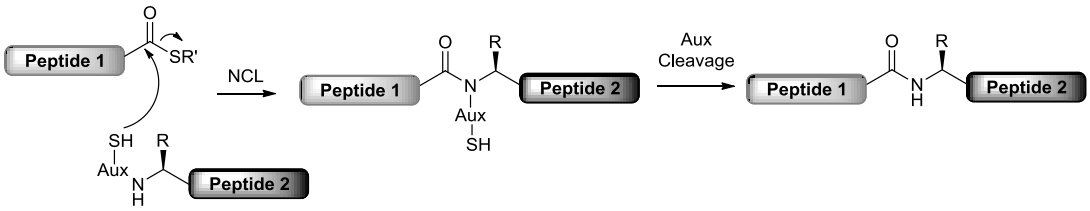


Figure 1.7: Thiolated amino acids used for ligation desulfurisation strategies

The low abundance of cysteine residues has also driven the investigation of thiol auxiliaries for NCL. Participation in the ligation reaction results in the formation of an amide linked product, after which the auxiliary is amenable to removal (Scheme 1.11). A number of cleavable auxiliaries have been developed,[102-106] although their use is limited to certain junctions.[107]



Scheme 1.11: Auxiliary mediated Native Chemical Ligation

The development of desulfurisation and auxiliary strategies has led to the widespread use of NCL. Most importantly these enabling technologies are compatible with glycosylated fragments and have been vital in the total chemical synthesis of glycoproteins.[108]

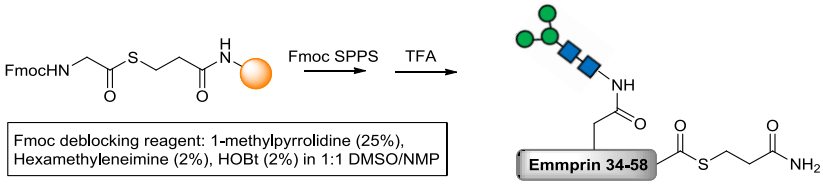
The task of obtaining the C-terminal fragment for NCL is comparatively straightforward, only requiring an N-terminal amino acid (or auxiliary) with a thiol moiety. These building blocks can be incorporated during SPPS, although in the case of recombinant fragments are limited to cysteine. More often the challenge encountered when attempting to synthesise proteins through NCL, is obtaining the N-terminal fragment adorned with a C-terminal thioester.

1.9 Chemical peptide thioester synthesis

Despite Boc SPPS facilitating the direct synthesis of peptide thioesters, the highly acidic conditions used in the final deprotection are often incompatible with glycosidic appendages.[109] Until recently,[48] this problem remained largely unresolved and consequently the Fmoc method has gained popularity for the synthesis of glycopeptides and their respective thioesters. However, thioesters are susceptible to nucleophilic attack during removal of the Fmoc group with piperidine, rendering direct synthesis unsuitable by routine protocols. Therefore, an important goal in the synthesis of glycoproteins has been the development of methods which enable the synthesis of glycopeptide thioesters via the Fmoc strategy.

1.9.1 Direct synthesis of peptide thioesters via Fmoc method

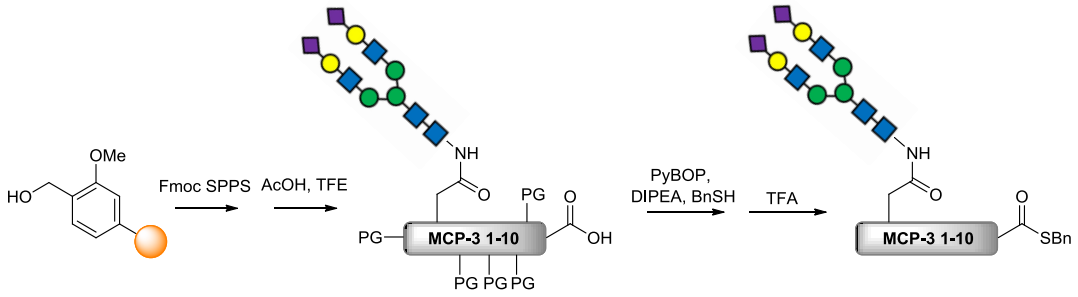
In an attempt to circumvent the cleavage of the thioester linkage, a number of researchers investigated the replacement of piperidine with less nucleophilic bases for Fmoc cleavage.[110-112] Hojo demonstrated the applicability of this strategy towards the synthesis of a glycopeptide thioester bearing an *N*-linked pentasaccharide (Scheme 1.12).[113]



Scheme 1.12: Fmoc SPPS of an N-linked glycopeptide thioester using bases with lower nucleophilicity

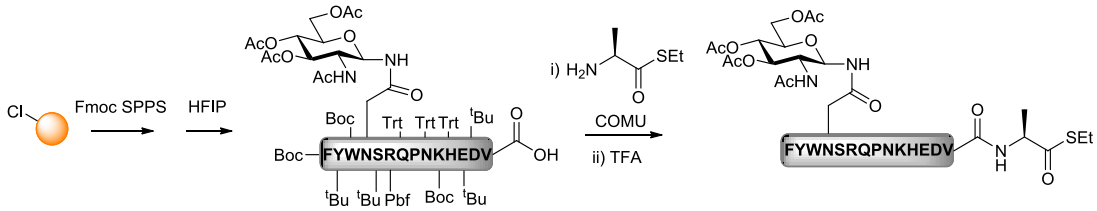
However, the use of bases with lower nucleophilicity can lead to racemisation of amino acids adjacent to the thioester. The thioester itself is also susceptible to aminolysis from adjacent residues during the first few cycles of peptide synthesis. This led researchers to investigate the direct thioesterification of fully protected peptides after assembly and cleavage from the resin. The use of highly acid-sensitive resins facilitates the release of protected peptide acids, which can subsequently be activated and thioesterified.[114-116] Kajihara successfully applied this concept towards the synthesis of a sialylglycopeptide thioester of monocyte chemotactic protein-3 (MCP-3).[117, 118] The glycopeptide was synthesised on HMPB-

PEGA resin, followed by treatment with acetic acid and trifluoroethanol to release the partially protected peptide. Subsequent activation of the C-terminal carboxylic acid and condensation with benzyl mercaptan afforded the thioester (Scheme 1.13).



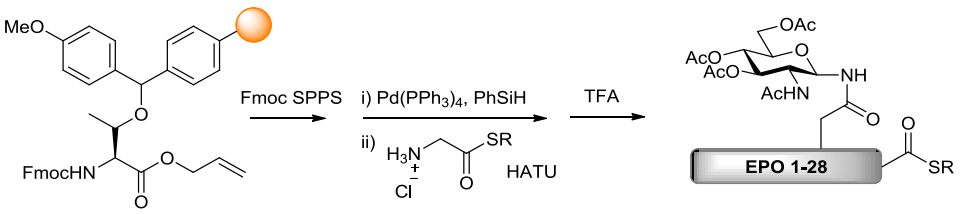
Scheme 1.13: Kajihara's thioesterification of a side chain protected sialylglycopeptide

This method has been further extended to obviate the use of thiols, by coupling amino acid thioesters (Scheme 1.14).[119] However, these approaches can be complicated by the poor solubility of fully protected peptides.



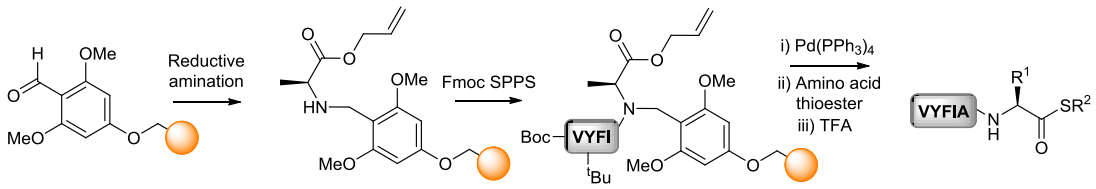
Scheme 1.14: Protected peptide coupling to amino acid thioesters

Another approach is the conversion of resin bound peptides to their corresponding thioester. This strategy relies on peptide attachment through a side chain and orthogonal protection of the C-terminal carboxylic acid in the form of an allyl ester.[120-122] Following peptide synthesis, the removal of the allyl group allows thioesterification by coupling of a thiol or an amino acid thioester. Wong and co-workers demonstrated this approach towards a glycosylated thioester corresponding to erythropoietin (EPO) residues 1-28 (Scheme 1.15).[123]



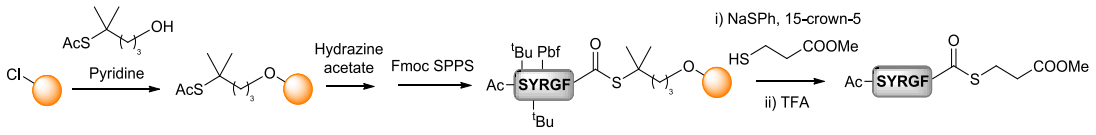
Scheme 1.15: Side chain anchoring strategy at threonine for the synthesis of EPO 1-28 thioester

The restriction of this methodology to amino acids with appropriate side chains is overcome by the use of the backbone amide linker (BAL). Originally applied to Fmoc SPPS by Barany,[124] the amino group of the penultimate C-terminal residue was initially attached to the tris(alkoxy)benzaldehyde resin via a reductive amination. Peptide extension in the *C* → *N* direction followed by allyl deprotection allowed coupling of an amino acid thioester (Scheme 1.16). Difficulties encountered with this strategy are largely due to the inefficient acylation of the secondary amine.



Scheme 1.16: BAL strategy applied to the synthesis of peptide thioesters via Fmoc SPPS

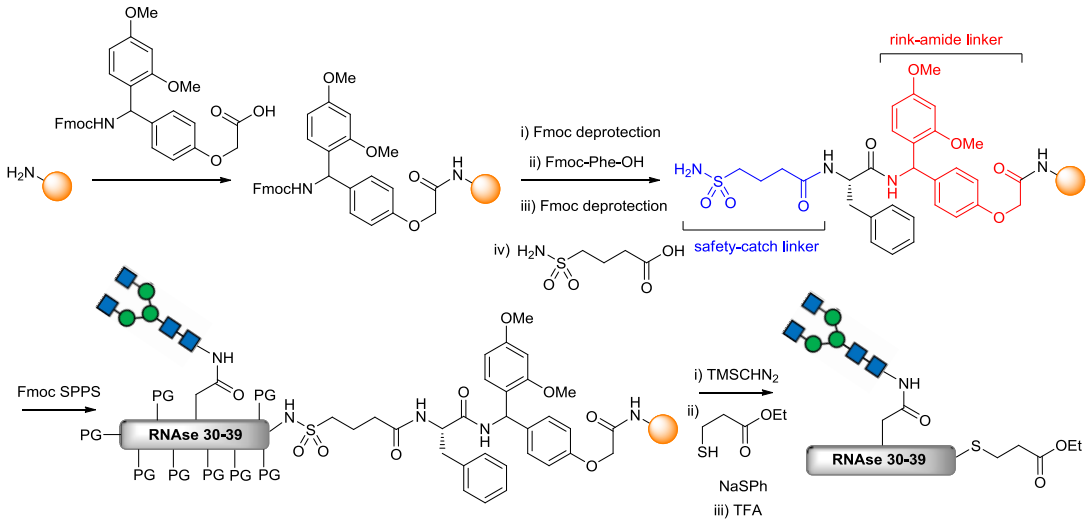
More recently, a linker developed by Raz and Rademann, showed that by using a *tert*-butyl thiol moiety it was possible to synthesise peptide thioesters that were stable to secondary amines (Scheme 1.17).[125] Steric hindrance and electron donation contribute to the stabilisation of the structure, although is still able to undergo thiolysis to afford the thioester.



Scheme 1.17: Peptide thioester synthesis on a tert-butyl thiol linker

1.9.2 Fmoc peptide thioester synthesis enabled by post SPPS activation of specialised linkers

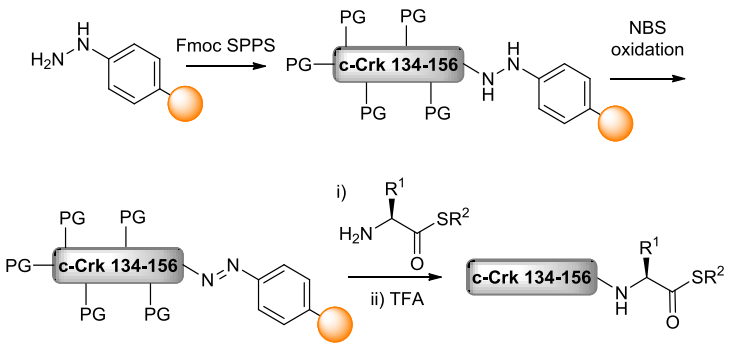
An alternative to the synthesis of peptide thioesters via the Fmoc approach utilises specialised linkers, which upon activation are susceptible to nucleophilic attack. One example of this is the use of a modified sulfonamide 'safety catch' linker in which the C-terminus of the peptide is attached to an *N*-acyl sulfonamide resin.[126] This functionality exhibits stability to acid and base, and is activated for cleavage with iodoacetonitrile or trimethylsilyldiazomethane prior to thiolysis. The method was extended by Unverzagt, who incorporated a double linker to allow convenient monitoring of peptide synthesis in the preparation of a glycopeptide derived from RNAse (Scheme 1.18).[127]



Scheme 1.18: Unverzagt's safety-catch linker in the synthesis of a glycopeptide thioester

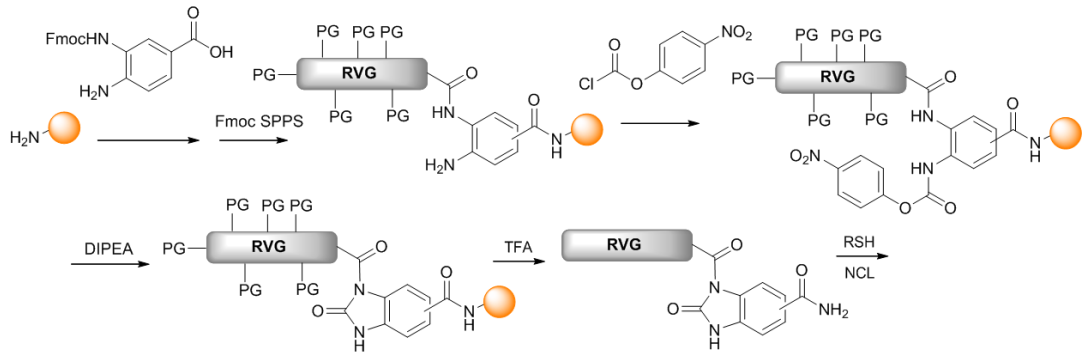
The use of the 'safety catch' strategy has successfully been applied to the synthesis glycopeptide thioesters,[127-129] although low loading of the first amino acid can lead to decreased yields.[130]

Camarero demonstrated that Fmoc SPPS could be carried out on an aryl hydrazide linker and subsequent oxidation to the acyl diazene could be followed by displacement with an amino acid thioester (Scheme 1.19).[131] Although successfully applied to the synthesis of a peptide thioester corresponding to the N-terminal SH3 domain of the c-Crk adaptor protein, undesired oxidation is observed at cysteine and methionine residues.



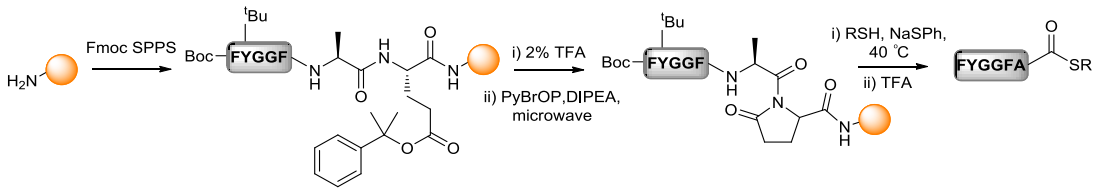
Scheme 1.19: Camarero's aryl hydrazine linker for Fmoc SPPS of peptide thioesters

Blanco-Canosa and Dawson later showed the utility of aromatic *N*-acyl ureas to thioesterify peptides.[132] Peptides could be assembled by Fmoc SPPS on a 3,4-diaminobenzoyl linker and formation of the *N*-acylbenzimidazolone was induced by 4-nitrophenyl chloroformate. Following ring closure under basic conditions, thiolysis of the activated peptide allows the formation a thioester (Scheme 1.20).



Scheme 1.20: Fmoc synthesis of rabies virus glycoprotein thioester with Dawson's Dbz linker

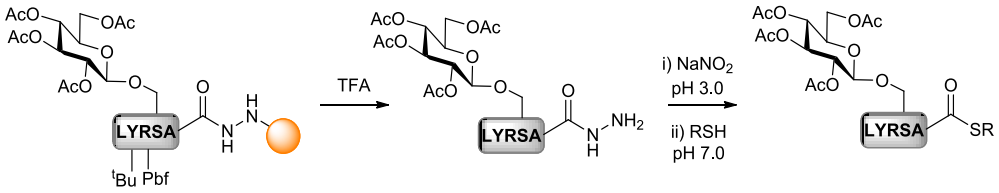
The backbone pyroglutamylimide linker was also shown to facilitate thioester formation by Tofteng *et al.*[133] Deprotection of the C-terminal glutamic acid side chain followed by activation with PyBrOP and microwave irradiation, allows pyroglutamylimide formation prior to thiolysis (Scheme 1.21).



Scheme 1.21: Pyroglutamylimide strategy for the synthesis of peptide thioesters

1.9.3 Fmoc SPPS of latent thioesters

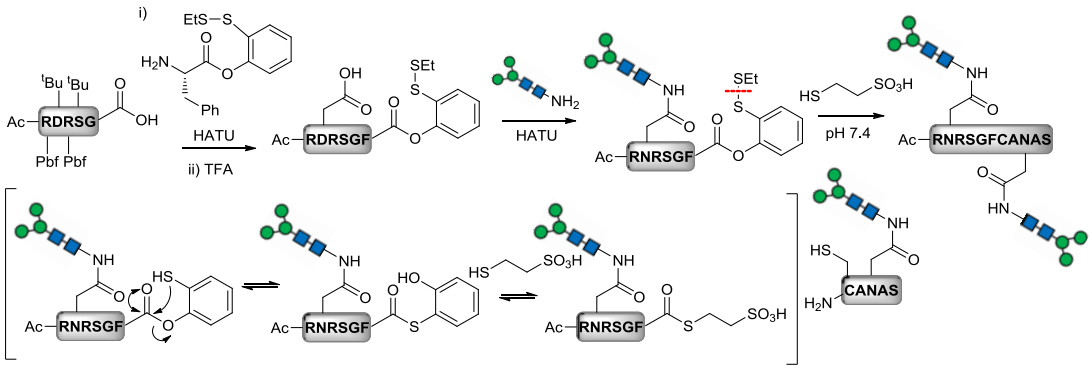
Another recent development in Fmoc SPPS of thioesters, is the use of surrogates in the form of acyl hydrazides. The utility of peptide hydrazides for ligation was originally investigated by Ramage,[134] although necessitated the protection of primary amino groups. The use of peptide hydrazides was further investigated by Liu, who showed that it was possible to selectively oxidise unprotected peptide hydrazides and convert the corresponding acyl azide to the thioester (Scheme 1.22).[135] Peptide hydrazides can be synthesised by Boc and Fmoc methods and also exhibit enhanced stability compared to thioesters. The ability to use peptide hydrazides as latent thioesters can be highly advantageous in convergent strategies.[136]



Scheme 1.22: Fmoc SPPS of a peptide hydrazide and conversion to a thioester

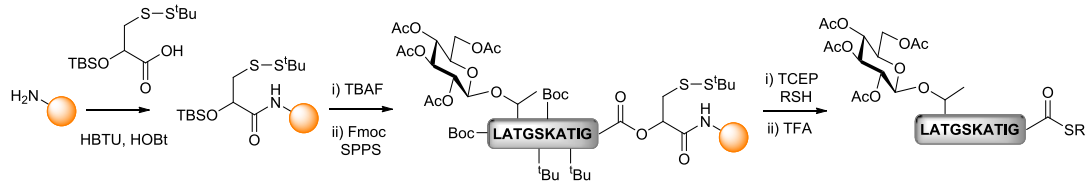
1.9.4 Peptide thioester formation via *O→S* acyl transfer

In recent years, the use of acyl transfer has gained considerable interest for the synthesis of thioesters. In 2004, Danishefsky showed that installation of a 2-mercaptophenyl ester at the C-terminus of a peptide could facilitate thioester formation via *O→S* acyl transfer and be used directly in ligation reactions (Scheme 1.23).[34]



Scheme 1.23: The use of α-2-mercaptophenyl esters for thioester formation via *O→S* acyl transfer

Botti also showed *O→S* acyl transfer of peptide esters to generate the corresponding thioester under ligation conditions.[137] This methodology was applied to glycopeptides by Zheng *et. al,* where thiolysis allows aryl thioester formation directly from the solid support (Scheme 1.24).[138]

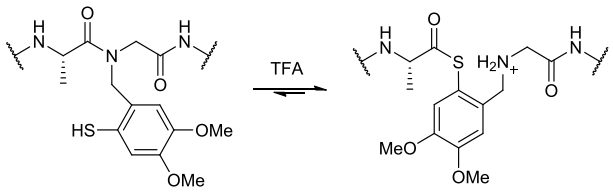


Scheme 1.24: Synthesis of glycopeptide thioesters via *O→S* acyl transfer

1.9.5 Peptide thioester formation via *N→S* acyl transfer

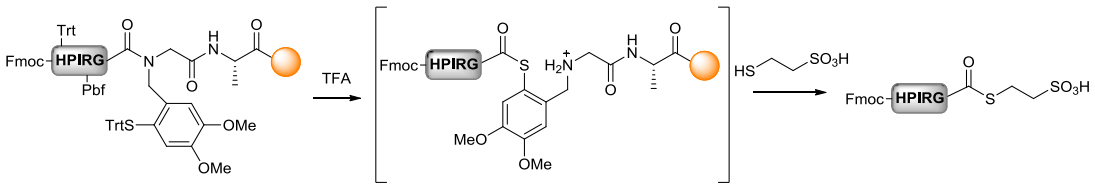
Many research groups have shown the use of *N→S* acyl transfer in the formation of peptide thioesters. The *N→S* acyl transfer mechanism has been shown to occur in nature, facilitating splicing reactions between inteins and N-exteins (see section 1.10.1.1). The resulting *S*-acyl intermediates can subsequently be intercepted by a thiol or used directly in ligation.

In a study investigating the use of the 2-mercapto-4,5-dimethoxybenzyl (Dmmb) group as a glycine auxiliary, it appeared that removal under acidic conditions triggered an *N→S* acyl shift to generate an internal thioester bond (Scheme 1.25).[139]



Scheme 1.25: Vorherr's observation of *N→S* acyl transfer on attempted auxiliary removal

Inspired by this observation, Vorherr and Aimoto evaluated the use of a C-terminal Dmmb group to form thioesters. After peptide assembly, TFA cleavage allowed removal of the side chain protecting groups and initiation of the *N→S* acyl transfer. This was followed by treatment with 2-mercaptoethanesulfonic acid to intercept the *S*-acyl intermediate resulting in the formation of the thioester (Scheme 1.26).[140]

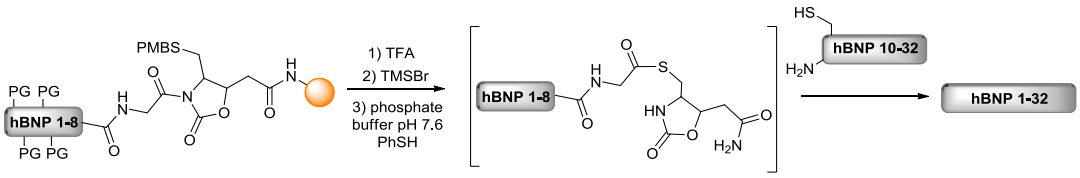


Scheme 1.26: Dmmb mediated thioester formation via *N→S* acyl transfer

The reaction was shown to proceed without any evidence of racemisation and can be performed at 37 °C to increase the rate of thioester formation.

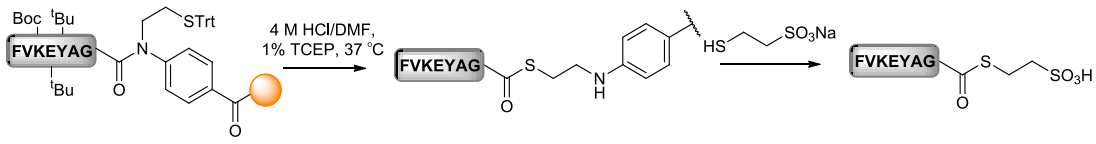
In 2004, Muir observed that a protein splicing reaction was enhanced by ground state destabilisation of a non-planar scissile amide bond.[141] Encouraged by this and a previous report which suggested that *N*-acyloxazolidinones were capable of

distorting amide planarity,[142] Otaka attempted the preparation of thioesters via a mercaptomethylated *N*-acyloxazolidinone derivative.[143] Fmoc SPPS required the use of Aimoto's reagent mixture[112] to prevent breakdown of the *N*-acyloxazolidinone. After cleavage from the resin and removal of the PMB group, *N*→*S* acyl transfer was induced in the ligation buffer (pH 7.6) in the presence of an N-terminal cysteine peptide affording full length hBNP (human B-type natriuretic peptide) (Scheme 1.27).



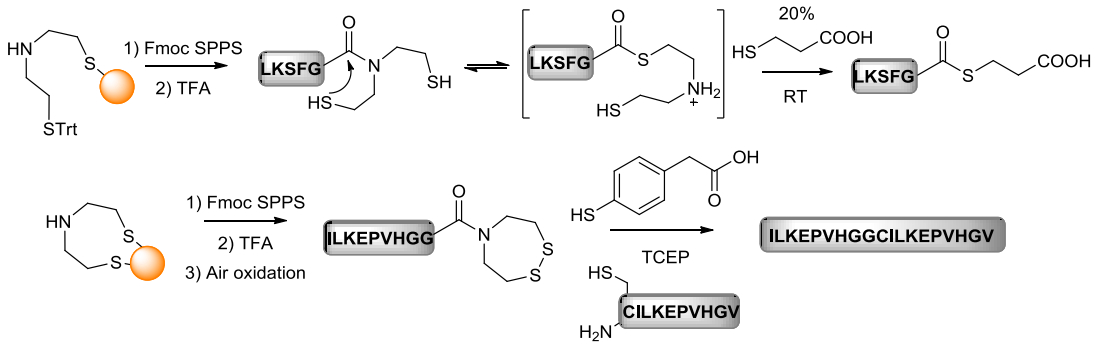
Scheme 1.27: *N*-acyloxazolidinone synthesis of thioesters and use in NCL

A drawback of this method is its incompatibility with piperidine during Fmoc deprotection. This led to the pursuit of more robust acyl transfer devices, which were compatible with Fmoc SPPS. Peptide *N*-sulfanylethylanilides were investigated as it was suspected that adjacent positioning of an aromatic ring would lead to activation of the amide bond. *N*→*S* acyl transfer was promoted with 4 M HCl/DMF and TCEP (1% w/v) at 37 °C for 8 hours, before thiolysis with MESNa. Further developments of this method allowed the acyl transfer reaction to be performed on-resin, allowing the crude release of peptide thioesters upon treatment with a thiol in aqueous buffer (Scheme 1.28).[144]



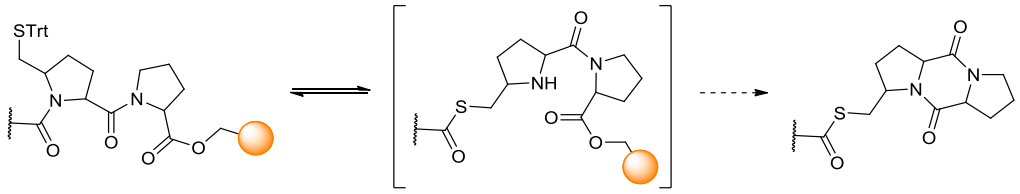
Scheme 1.28: *N*-sulfanylethylanilides mediating on-resin thioester formation

In an extension of this work, Melnyk and Liu independently reported the use of Bis(2-sulfanylethyl) amides to synthesise peptides for ligation.[145-147] The installation of two thiol groups obviates an isomerisation step to form the energetically unfavourable *cis* conformation. Therefore, this motif undergoes *N*→*S* acyl transfer more readily, as one of the sulfhydryl groups is always *anti* to the carbonyl oxygen and hence able to promote the intramolecular reaction (Scheme 1.29).



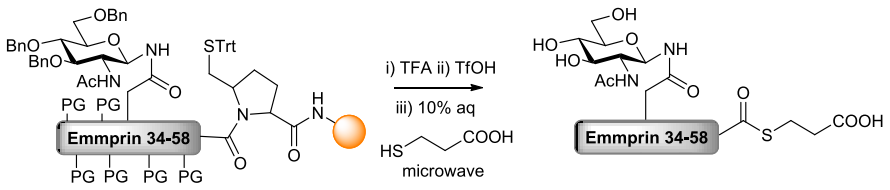
Scheme 1.29: Application of bis(2-sulfanylethyl) amides for thioester synthesis and direct use in NCL

Hojo and Nakahara later investigated the use of thiolated proline derivative as a post SPPS thioesterification device.[148] The reaction was expected to proceed via the formation of a diketopiperazine, which is particularly known to occur at C-terminal Pro-Pro dipeptides (Scheme 1.30).



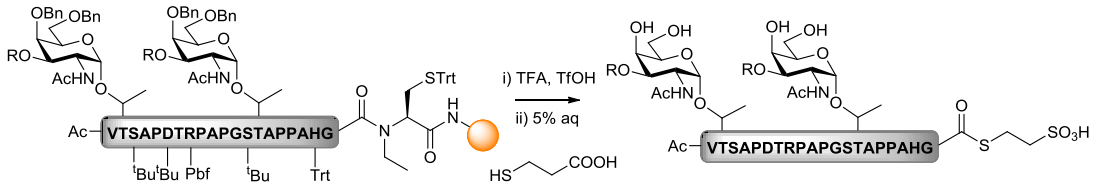
Scheme 1.30: Diketopiperazine formation towards enhanced thioester formation

However, the expected transformation did not occur under the reaction conditions and it was suspected that the reverse *S*→*N* acyl shift was favoured over diketopiperazine formation. Consequently, the reaction was attempted in the presence of 3-mercaptopropionic acid (MPA) in an attempt to form the thioester via intermolecular exchange. Although this method successfully generated the thioester, acceptable reaction rates necessitated the use of microwave irradiation. The reaction was subsequently applied towards the synthesis of a glycopeptide thioester (Scheme 1.31).



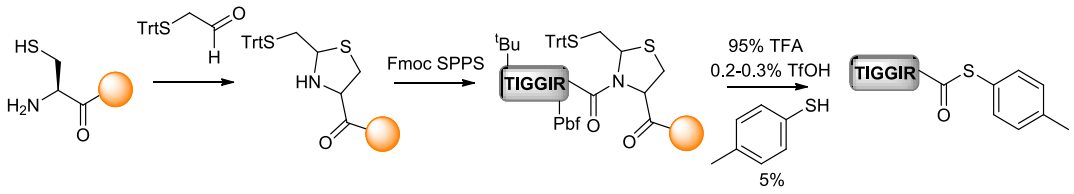
Scheme 1.31: *N*-glycopeptide thioester formation using a mercaptomethylated proline derivative

The need for microwave irradiation and the laborious preparation of the mercaptomethyl proline derivative led the authors to investigate alternative devices for *N*→*S* acyl transfer.[149] It was postulated that the intramolecular rearrangement was favoured by the increased lability of the tertiary amide bond when the C-terminal residue was proline. Accordingly, a number of *N*-alkyl cysteine derivatives were prepared and evaluated for thioester formation. Encouragingly, the *N*-alkyl cysteine derivatives could be thioesterified at room temperature with lower concentrations of MPA without the need of microwave irradiation. Conversion to the thioester was found to proceed quicker when bulky groups were used to alkylate the nitrogen, however steric hindrance resulted in poor coupling of the next amino acid. *N*-ethyl cysteine provided the best overall yield and has consequently been utilised for the synthesis of glycopeptide thioesters (Scheme 1.32).[150]



Scheme 1.32: *N*-alkyl cysteine mediated thioesterification of glycopeptides

Tam later developed an alternative mercaptomethylated proline surrogate in the form of a thiazolidine.[151] The shorter preparation of the acyl transfer device relied on the intermolecular cyclisation of a resin bound cysteine with 2-(tritylthio) acetaldehyde. After Fmoc SPPS, the peptide was cleaved from the resin with TFA, TfOH and thiocresol, leading to concomitant thioester formation (Scheme 1.33).

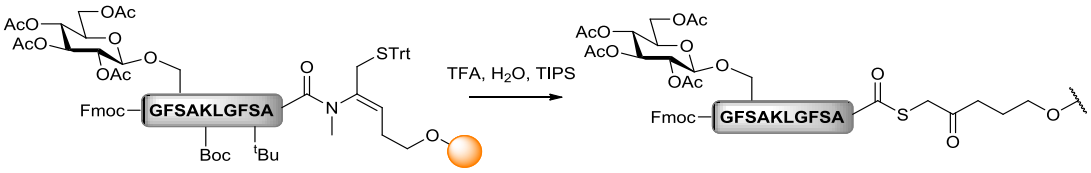


Scheme 1.33: Mercaptomethylated thiazolidine mediated thioester formation

Kawakami and Aimoto showed that it was possible to form thioesters through use of an autoactivating cysteine-proline ester. Originally reported by Zanotti *et al.*[152] *N→S* acyl shift is followed by diketopiperazine formation, thereby preventing the reverse reaction. The resulting diketopiperazine can be transthioesterified to generate peptide thioesters, or can be used directly in NCL reactions (Scheme 1.34).[153] The reaction occurs under neutral conditions at room temperature and presents a mild method of peptide thioesterification.

Scheme 1.34: Cysteine-proline ester mediated ligation

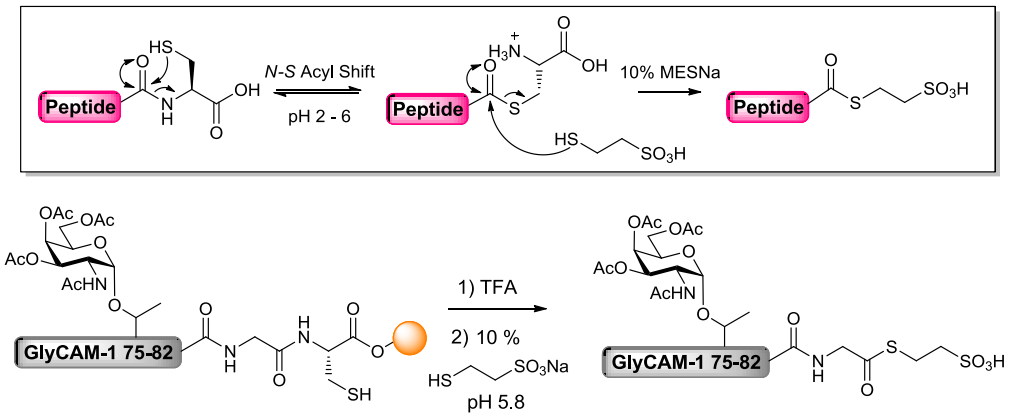
Recently, Liu introduced the concept of Fmoc synthesis of thioesters using a C-terminal enamide.[154] The clever design of the linker in this strategy allows thioester formation to take place via *N→S* acyl transfer, with the reverse reaction prevented by enamine hydrolysis. This essentially allows thioester formation to take place during cleavage from the resin without the need for any non-standard manipulations and has been successfully used to generate glycopeptide thioesters (Scheme 1.35).



Scheme 1.35: Enamide mediated thioester formation

Crucially, the methylation of the C-terminal amide nitrogen prevents thiazole formation, which was found to predominate in earlier studies. The major limitation of this method, is the multistep procedure required for the synthesis of the enamide building block.

The use of *N→S* acyl transfer for the synthesis of peptide thioesters is usually facilitated by incorporation of unnatural motifs preinstalled during SPPS. However, Macmillan observed that this transformation could also be achieved by a C-terminal cysteine residue (Scheme 1.36).[155] Thioester formation via *N→S* acyl shift was found to be efficient at Gly-Cys, His-Cys and Cys-Cys junctions and has been shown to be compatible with thioester synthesis of *O*-linked glycopeptides.[156]



Scheme 1.36: Cysteine promoted thioester formation via *N→S* acyl shift

The commercial availability of preloaded cysteine resins makes this a straightforward and accessible strategy towards synthesising peptide thioesters via the Fmoc method. The methodology also allows the convenient synthesis of cyclic peptides, allowing thioester formation and irreversible cyclisation to take place concomitantly under the same reaction conditions.[157] The concept was also extended to C-terminal selenocysteine residues, where it was shown that *N→Se* acyl transfer could also lead to accelerated formation of thioesters.[158]

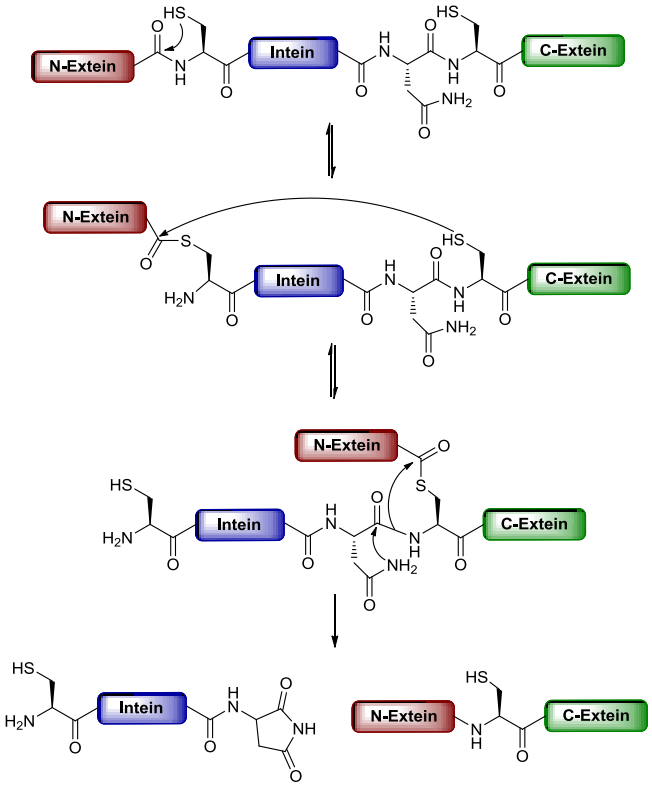
1.10 Production of recombinant proteins compatible with NCL

The use of recombinant DNA technology in the overexpression of proteins has provided an indispensable tool towards the study of protein structure and function. One of the most widely used hosts for protein expression is *E. coli,* which can be grown rapidly at high density to provide unmodified proteins.[159] The ability to access large recombinant fragments led to an interest in their use for protein semi-synthesis. However, the most commonly available expression systems are restricted to unmodified protein structures, prompting researchers to investigate techniques which enable their selective functionalisation. For protein semi-synthesis, particular emphasis has been geared towards methods which enable their compatibility with Native Chemical Ligation.

1.10.1 Production of recombinant protein thioesters for NCL

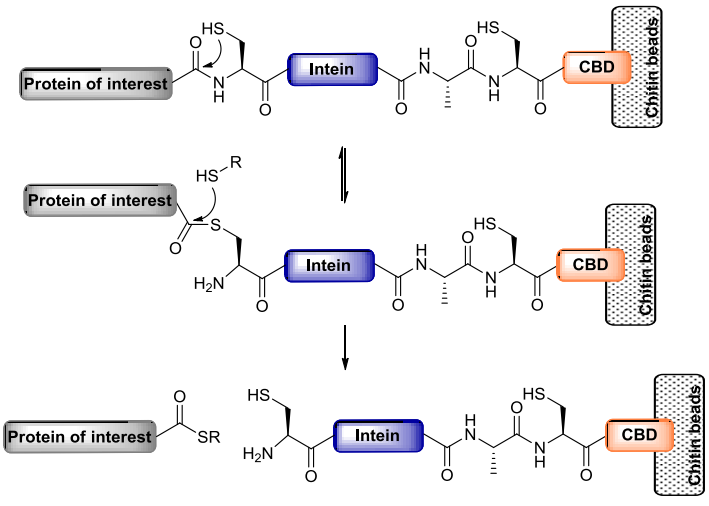
1.10.1.1 Intein mediated production of recombinant protein thioesters

One of the first methodologies explored for the thioesterification of recombinant proteins exploited the unique reactivity of protein splicing elements. These proteins are able to undergo intramolecular rearrangements to excise an internal sequence (intein), whilst subsequently linking the resulting external sequences (exteins). The first step in the splicing mechanism most commonly involves an *N*→*S* acyl transfer at the junction between the N-extein and intein. The resulting "*S*-peptide" intermediate can undergo transthioesterification by the side chain of cysteine at the intein-C-extein junction. Finally, the intein is excised from the branched intermediate through asparagine mediated succinimide formation and the thioester linked exteins undergo an *S*→*N* acyl transfer to form a native amide linkage (Scheme 1.37).[160]



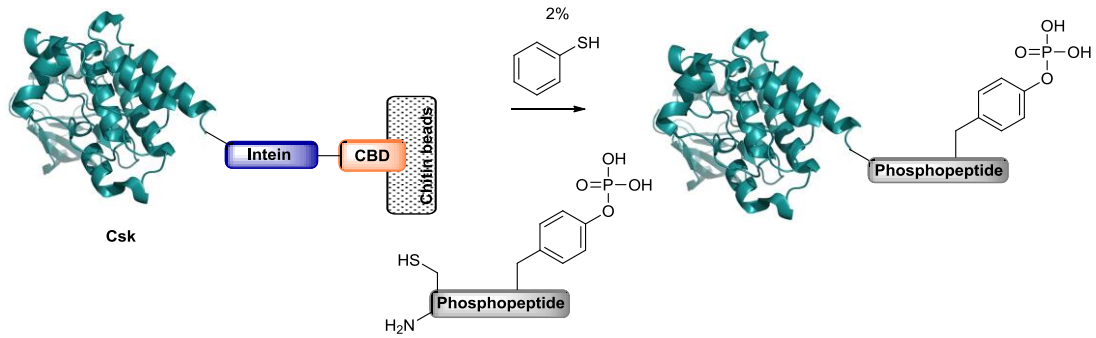
Scheme 1.37: Mechanism of protein *cis*-splicing

The use of intein technology was initially exploited for the single step purification of recombinant proteins using a C-terminal chitin binding domain (CBD) for chitin affinity chromatography. Mutated versions of inteins were designed to only allow the initial *N*→*S* acyl transfer to proceed, with the resulting intermediate susceptible to intermolecular transthioesterification by a thiol (Scheme 1.38). This effectively allows the protein of interest to be cleaved at the splice junction.[161]



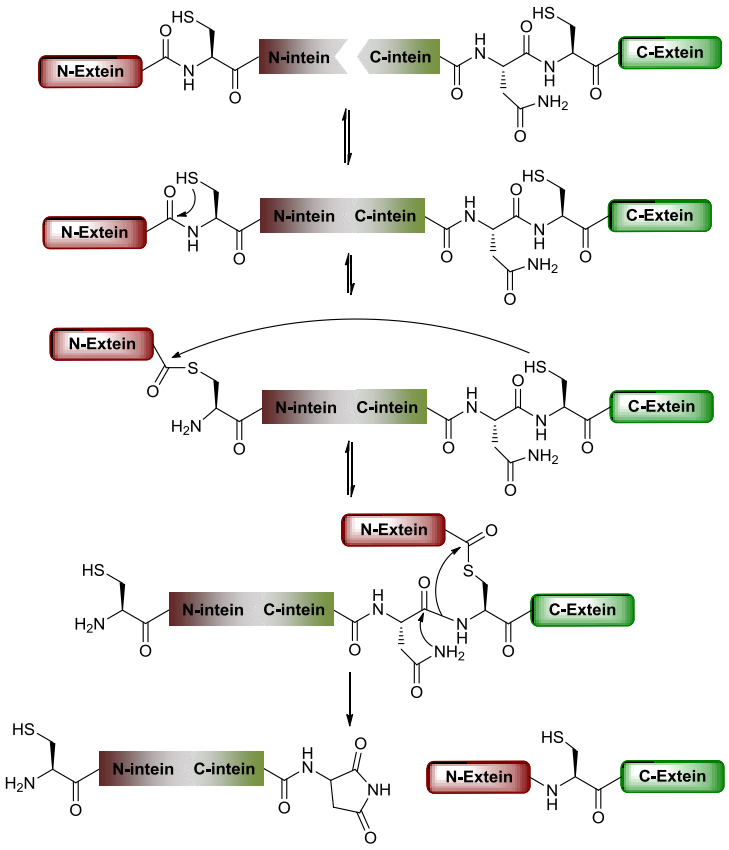
Scheme 1.38: Intein mediated purification of proteins via thiolysis

Cleavage of the intein fusion with β-mercaptoethanol or dithiothreitol ultimately results in a C-terminal carboxylic acid via hydrolysis of the thioester. Inspired by the fact that the reaction proceeded through a transthioesterification, Muir and co-workers explored the possibility of intercepting the *S*-peptide intermediate with synthetic peptides through NCL. This was successfully demonstrated when thiophenol was included in the cleavage buffer, suggesting formation of the aryl thioester was necessary before ligation with a cysteinyl peptide.[162] This methodology was successfully used to generate semi-synthetic tail-phosphorylated C-terminal Src kinase (Csk) (Scheme 1.39) and the technology was accordingly named "Expressed Protein Ligation" (EPL). Alternatively, the intein fusion can also be cleaved with a thiol such as MESNa, which provides a thioester with improved stability to hydrolysis and can hence be used in NCL at a later stage. EPL has since been applied towards a number of semi-synthetic strategies, enabling access to modified proteins.[91, 128, 130, 163-169]



Scheme 1.39: Expressed protein ligation - semi-synthesis of tail-phosphorylated Csk

Further developments of EPL included the discovery of "split" inteins. In this subclass, inteins are expressed as two fragments (N-intein and C-intein), which are able to spontaneously and non-covalently assemble to initiate protein *trans*-splicing (Scheme 1.40).[170]



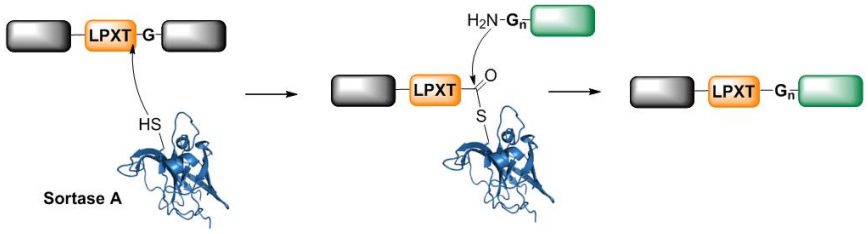
Scheme 1.40: Split intein mediated protein *trans*-splicing

In 2011, Liu and Cotton independently reported the cleavage of intein fusions with hydrazine.[135, 171] This reaction affords the corresponding acyl hydrazide, a latent thioester surrogate which can subsequently participate in EPL. Hydrazinolysis of C-terminal intein fusions has also shown to proceed more efficiently, with improved yields compared to thioester formation.

Although the application of EPL has been effective in a number of studies, it is important to note that efficient cleavage of the fusion protein relies on a correctly folded intein structure. In the case of hydrophobic sequences or unstructured fragments, the poor solubility of the target protein can render the intein fusion insoluble. As a consequence, certain sequences may be unsuitable for EPL, although it has been demonstrated that after successful refolding of an insoluble fusion, the resulting protein can be susceptible to thiolysis.[172]

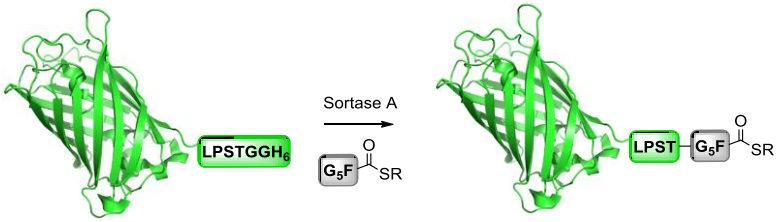
1.10.1.2 Protein thioester synthesis enabled by sortase

Sortase's are transpeptidases involved in the anchoring of surface proteins to the cell wall and found in most Gram-positive bacteria.[173] Sortase A recognises the motif 'LPXTG' (where X is any amino acid), and a cysteine residue within the active site acts as a nucleophile cleaving at the Thr-Gly junction. The resulting *S*-acyl intermediate can be intercepted by an N-terminal amino group of a sequence containing one to five glycine residues, affording the ligated product (Scheme 1.41).



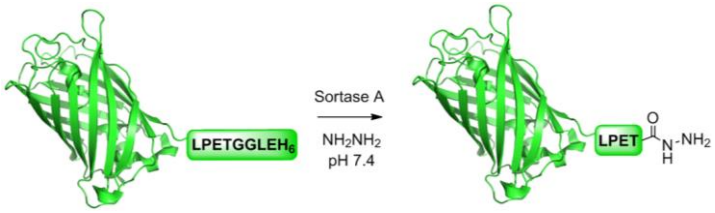
Scheme 1.41: Sortase A catalysed ligation

Recently, Pentelute reported that by using Sortase A, it was possible to append synthetic oligoglycine thioesters to recombinant proteins, potentially allowing ligation of fragments at a later stage (Scheme 1.42).[174] This strategy could be particularly beneficial for ligations involving insoluble C-terminal sequences, which are unsuitable for direct sortase ligation.



Scheme 1.42: Recombinant thioester synthesis enabled by sortase A

A further extension of sortase A in the production of recombinant proteins for use in NCL was demonstrated by Liu. It was shown that recombinant sequences bearing the LPXTG motif could be cleaved with hydrazine at the Thr-Gly junction in the presence of sortase A to form the acyl hydrazide (Scheme 1.43).[175] The irreversible reaction prevents further recognition by sortase and hence limits background hydrolysis of the thioester linked intermediate.

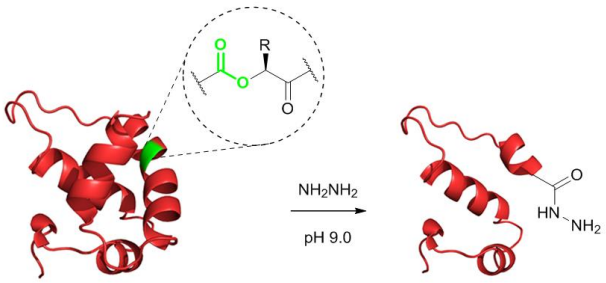


Scheme 1.43: Recombinant hydrazide synthesis enabled by sortase A

Although successfully utilised for peptide ligation[79, 176] and protein semi-synthesis,[177] the broad application of sortase is currently limited by the requirement of recognition motifs. Future investigations may yield evolved forms of the enzyme, which may exhibit wider tolerance for recognition sequences, consequently expanding its utility.

1.10.1.3 Recombinant production of protein hydrazides via oxoester moiety

In 2008, Schultz showed that the use of genetic code expansion with unique codons and a corresponding orthogonal tRNA–aminoacyl-tRNA synthetase pair[178, 179] could be used to introduce α-hydroxy acids into proteins expressed in *E. coli* with site specificity.[180] This results in the incorporation of a backbone ester linkage, which was shown to be vulnerable to selective hydrolysis under slightly basic conditions. Inspired by this, Liu showed that it was also possible to cleave such linkages with aqueous hydrazine to provide thioester synthons in the form of acyl hydrazides (Scheme 1.44).[181]



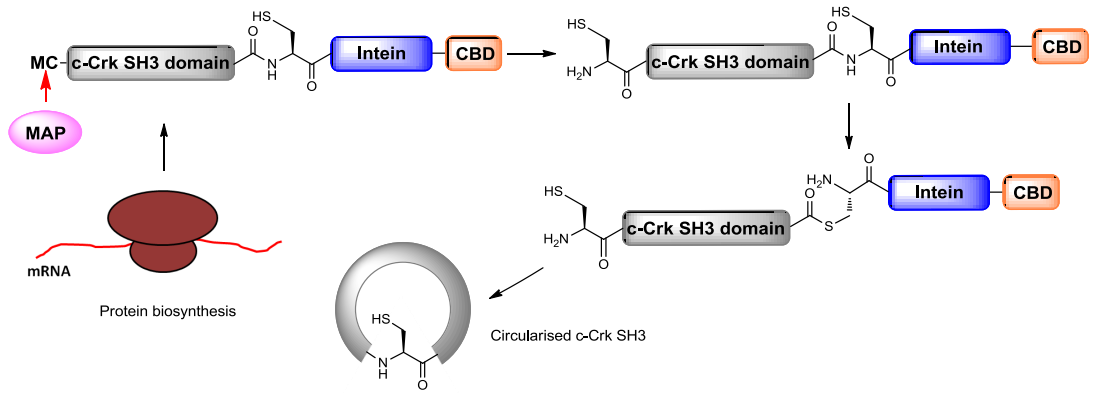
Scheme 1.44: Genetic incorporation of oxoesters to facilitate site selective hydrazinolysis

This strategy may exceed some of the limitations found with intein based technologies, as successful cleavage should not depend on the expression of correctly folded structures. However, the encoding of unnatural amino acids can significantly lower yields of the expressed protein.

1.10.2 Production of recombinant proteins functionalised with an N-terminal cysteine for NCL

1.10.2.1 Methionylaminopeptidase generation of an N-terminal cysteine

In prokaryotes and eukaryotes, protein translation is initiated by the incorporation of a methionine residue. As a consequence, the generation of C-terminal fragments for NCL relies on the cleavage of a leader sequence. In *E. coli,* most cytosolic proteins do not exhibit an N-terminal methionine, as it is excised by methionylaminopeptidase (MAP).[182, 183] With this in mind, Muir showed that inclusion of a Met-Cys leader motif could allow the expression of proteins containing an N-terminal cysteine and applied this to the circularisation of the c-Crk SH3 domain *in vivo* (Scheme 1.45).[184]



Scheme 1.45: Expression of a protein with an N-terminal cysteine followed by NCL *in vivo*

However, the expression of proteins with a Met-Cys leader sequence can result in diminished translation efficiency.[185] Furthermore, MAP cleavage of the N-terminal methionine does not always proceed.[186] This led researchers to investigate the use of proteases to unmask N-terminal cysteine residues.

1.10.2.2 Protease generation of an N-terminal cysteine

Verdine showed that insertion of a Factor Xa protease recognition sequence (IEGR) preceding the protein of interest, could facilitate enzymatic cleavage to reveal an N-terminal cysteine (Figure 1.8 a).[187] An extension of this method was introduced by Wong, who showed that the use of Tobacco Etch Virus (TEV) protease could be utilised in a similar manner.[188] TEV protease offers a significant advantage, as its

extended recognition sequence ENLYFQ(G/S) improves selectivity and is relatively inexpensive as it can be produced "in house" by expression in *E. coli.* TEV cleavage takes place between the final two amino acids of the recognition sequence and tolerates cysteine as the seventh residue (Figure 1.8 b).[189]

Leatherbarrow later showed that the 3C protease from Foot-and-mouth disease virus could also cleave between the final two residues in the recognition sequence PAKQX to generate N-terminal cysteinyl proteins for NCL (Figure 1.8 c).[190]

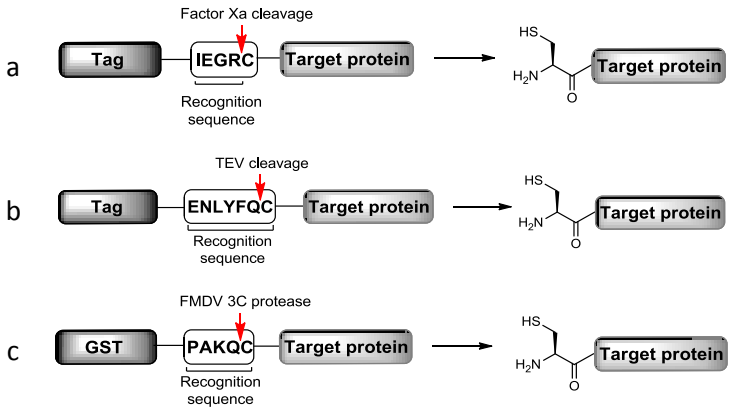
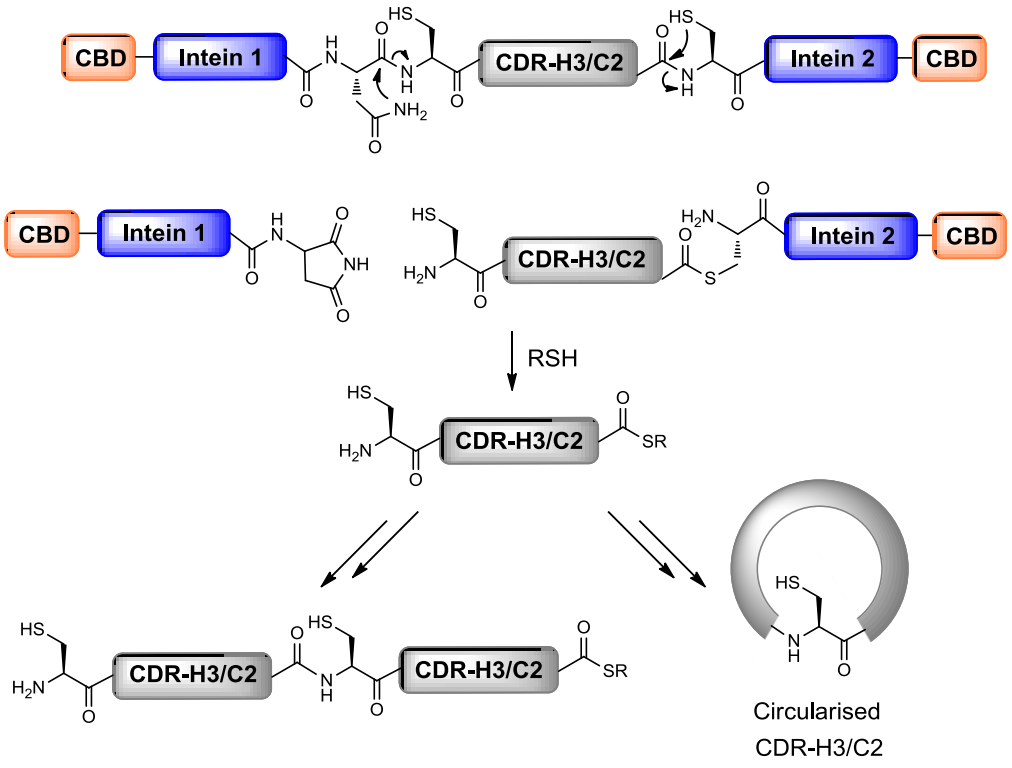


Figure 1.8: Protease cleavage for release of N-terminal cysteines

1.10.2.3 Intein mediated generation of an N-terminal cysteine

As discussed earlier, the natural ability of an intein to splice at its N-terminal junction with an extein is useful for the generation of recombinant protein thioesters. Similarly, the splicing activity at the C-terminus of an intein can be exploited to functionalise proteins with an N-terminal cysteine. A study by Evans *et al.* demonstrated this approach to be feasible and consequently showed that addition of a C-terminal intein could lead to cyclisation or polymerisation of the expressed protein corresponding to CDR-H3 (Scheme 1.46).[191]



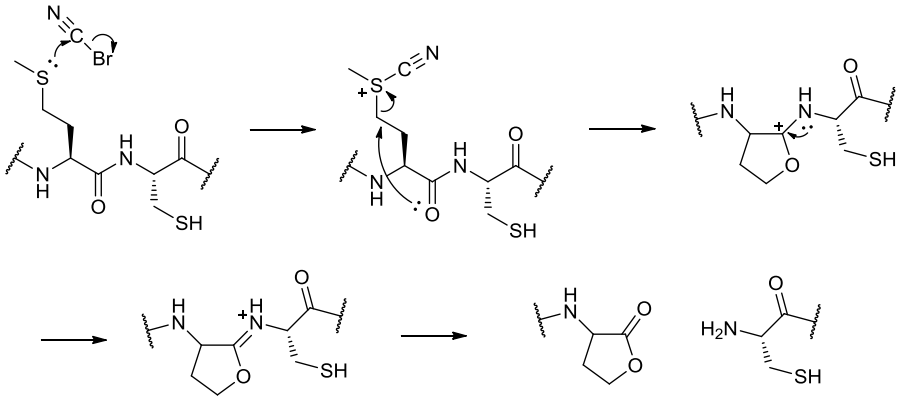
Scheme 1.46: N-terminal intein fusion for the expression of N-terminal cysteinyl proteins

1.10.2.4 Chemical approaches for the generation of an N-terminal cysteine

The use of a protease or an intein fusion to generate recombinant fragments with an N-terminal cysteine, rely on the expressed proteins to be soluble in buffers compatible with cleavage. In 2004, Macmillan designed a semi-synthetic strategy towards erythropoietin (EPO).[192] Residues 29-166 were expressed with a polyhistidine tag and Factor Xa recognition sequence preceding the desired N-terminal cysteine. The resulting protein could only be solubilised through the use of denaturants or detergents. It was found that the employment of these reagents compromised protease activity and formation of the desired product was not observed. Instead, a chemical approach was investigated adopting the use of cyanogen bromide, which had previously been demonstrated to selectively cleave proteins at the C-terminus of methionine residues (Scheme 1.47).[193]

Scheme 1.47: Liberation of an N-terminal cysteine via CNBr cleavage

The reaction typically employs the use of formic acid or TFA, which efficiently solubilises difficult sequences and masks the reactivity of basic residues. Selectivity is also ensured by the mutation of native methionine residues within the protein of interest.



Scheme 1.48: Proposed mechanism of CNBr cleavage

An inherent drawback of this approach, is the highly toxic nature of cyanogen bromide and requirement to mutate native methionine residues. However, this is the only method which can be universally applied to any protein of interest.

1.11 Summary

Chemical protein synthesis relies on the platforms of solid phase peptide synthesis and Native Chemical Ligation, which have enabled the synthesis of large proteins through convergent strategies. The chemoselective reaction of NCL requires the synthesis of C-terminal thioesters, which can be accessed directly through Boc SPPS. However, posttranslational modifications are often incompatible with harshly acidic hydrogen fluoride used during this method, which is also accompanied by significant hazards of its own.

The Fmoc method has consequently gained popularity, although thioesters are prone to nucleophilic attack during piperidine treatment used to remove the Fmoc group. As a result, much effort has aimed to address the general incompatibility of Fmoc SPPS with direct thioester synthesis. A number of strategies have consequently been developed to facilitate the synthesis of peptide thioesters using Fmoc SPPS. These methods have been extended to the synthesis of glycopeptide thioesters, enabling homogeneous glycoprotein synthesis to be attained through NCL. It is important to note that in all cases, homogeneity is established by the employment of chemical methods to access defined glycosylated protein fragments.

The compatibility of NCL with expressed proteins has provided a useful extension of this methodology. Recombinant expression of unmodified proteins allows access to fragments which are not limited in size and can present a significant advantage over chemical synthesis. However, for recombinant proteins to participate in expressed protein ligation, they usually require a processing step which enables the formation of the required functionality. The exploitation of various biological technologies can provide access to the desired functionality, but often fall short in the case of particularly hydrophobic sequences. On the other hand, chemoselective treatments can overcome these limitations, as reactions can be performed under conditions which are usually incompatible with biological systems.

1.12 Project Aims

This study has aimed to establish a routine method of homogeneous native glycoprotein assembly through semi-synthesis. For demonstration of this strategy, efforts were focussed on Interferon β-1, a 166 residue glycoprotein currently used for the treatment of multiple sclerosis. Due to the position of glycosylation, a three fragment semi-synthetic strategy for IFNβ was designed (Figure 1.9).

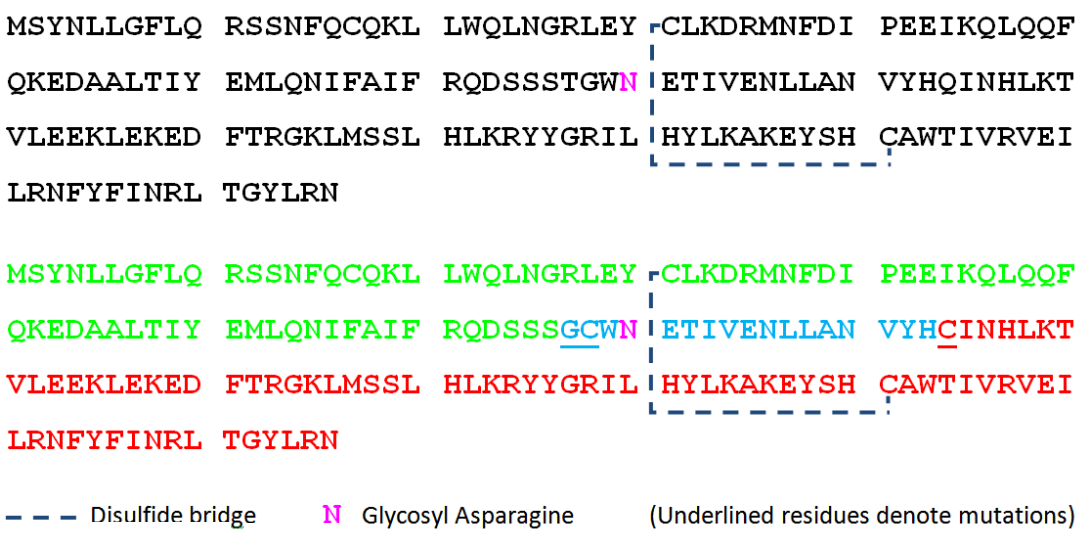
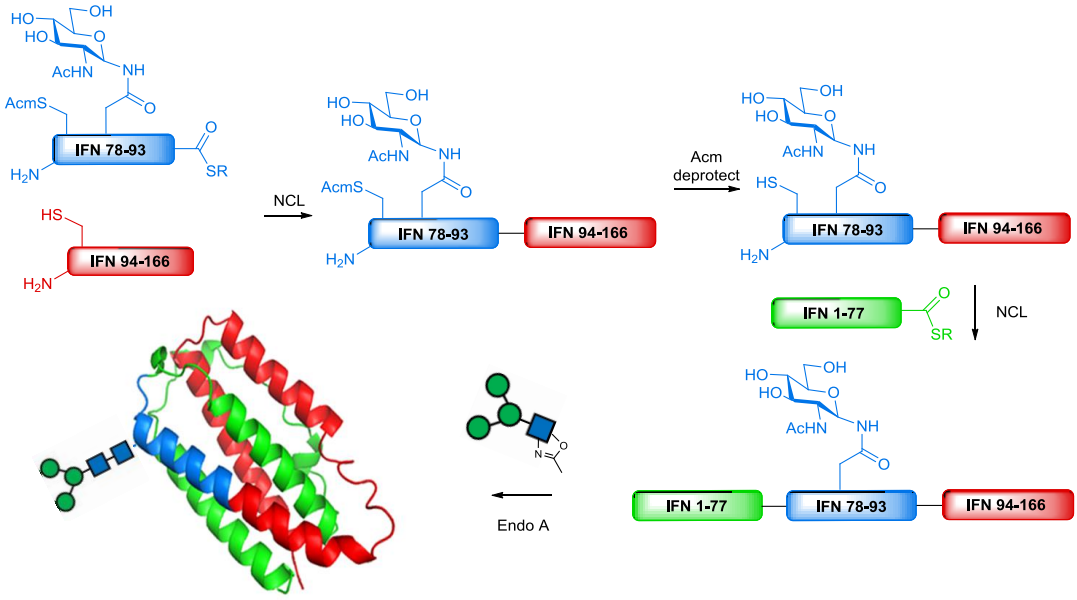


Figure 1.9: Sequence analysis of IFNβ and design of a semi-synthetic strategy

To facilitate the NCL strategy, cysteine residues were introduced by substitution at residues 78 and 94. A T77G mutation was also incorporated as NCL is known to be more efficient with glycine thioesters.[194] It was envisaged that the chemical synthesis of a glycopeptide corresponding to residues 78-93 would establish homogeneity. The incorporation of *N*-acetylglucosamine at the site of glycosylation also enables a selective enzymatic transfer of a glycan with endoglycosidase A.

After expression of recombinant fragments corresponding to residues 1-77 and 94-166 in *E. coli,* functionalisation of these proteins enables participation in NCL reactions. Sequential ligation reactions in the *N* → *C* or *C* → *N* direction would allow access to the complete protein backbone, allowing investigation into enzymatic transfer of a glycan in the final stages of glycoprotein assembly (Scheme 1.49).

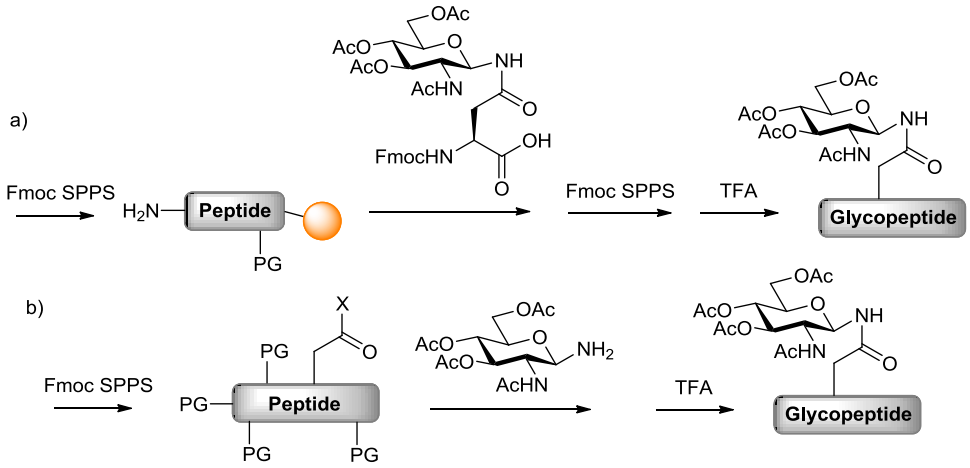


Scheme 1.49: Proposed assembly of semi-synthetic IFNβ

2 Synthesis of *N*-linked glycopeptides for semi-synthesis of Interferon β-1

2.1 Strategies for incorporation of GlcNAc into *N*-glycopeptides

A critical step in the synthesis of homogeneous *N*-glycoprotein synthesis is the formation of the GlcNAc-Asparagine linkage. For the purpose of this study where the GlcNAcβ(1-4)GlcNAc chitobiose core is formed by endoglycosidase A, methods allowing the introduction of *N*-acetylglucosamine into peptides were of particular interest. The stepwise strategy relies on the incorporation of suitably protected glycoamino acids during SPPS (Scheme 2.1 a),[195, 196] whereas the convergent strategy relies on coupling between GlcNAc and the side chain of a peptide (Scheme 2.1 b). A complication often observed with this approach is the formation of aspartimide by-products.[31, 197, 198]

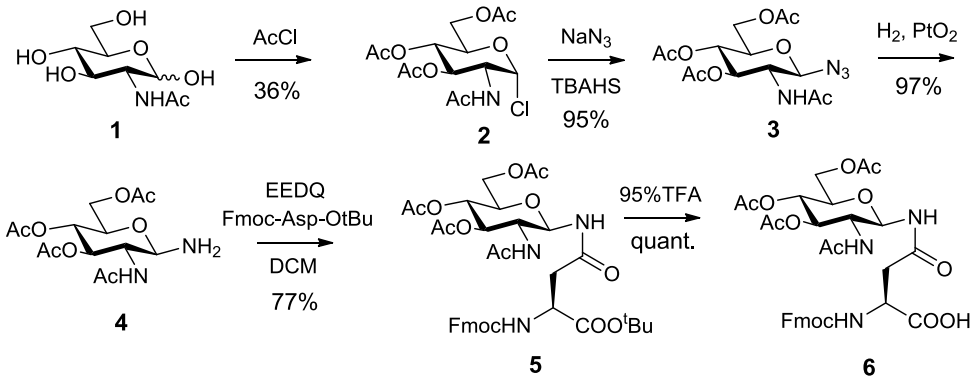


Scheme 2.1: Methods of GlcNAc incorporation into peptides

It has also recently been shown that the *N*-glycosylation enzyme Pg1B can be used to incorporate GlcNAc at certain sequences.[199, 200] However, the requirement of the D/E-X-N-X-S/T motif prevents broader application and is further complicated by the need for complex unnatural substrates.

2.2 Synthesis of glycoamino acid for Fmoc SPPS

The glycoamino acid was synthesised from *N*-acetylglucosamine, initially treated with acetyl chloride according to the procedure reported by Horton.[201] Acetylation at all the hydroxyl groups of the sugar is followed by loss of acetic acid from the anomeric carbon. The reaction proceeds via an oxocarbenium ion intermediate, which is intercepted by chloride ions to form the α-glycosyl chloride **2**, favoured by the anomeric effect. Following this step, the α-chloride was converted to the glycosyl azide under phase transfer conditions to give the β-azide **3** in excellent yield.[202] Subsequent reduction of the azide with a platinum (IV) catalyst yielded the free amine **4**, which could then be coupled to the side chain of aspartic acid.[203] This process of forming the native *N*-glycosidic linkage was assisted by the coupling reagent EEDQ.[204] After *tert*-butyl deprotection of the α-carboxyl group with TFA, the glycoamino acid **6** was ready to be used in Fmoc SPPS (Scheme 2.2).



Scheme 2.2: Glycoamino acid synthesis

2.3 Synthesis of glycopeptides for NCL

2.3.1 Native *N*-glycopeptide thioester synthesis via *N*→*S* acyl transfer

In a model study for thioester formation of *N*-glycopeptides, the peptide corresponding to EPO residues 22-28 was assembled with a C-terminal cysteine on NovaSyn TGT resin. The glycoamino acid **6** was incorporated at position 24, the native site of glycosylation. The peptide was assembled in a stepwise fashion using 10 equivalents of protected amino acids (except 5 equivalents for the glycoamino acid), employing 0.45 M HBTU/HOBt in DMF to assist peptide bond formation. After cleavage from the resin, the glycopeptide **7** was purified by (RP)HPLC (Figure 2.1).

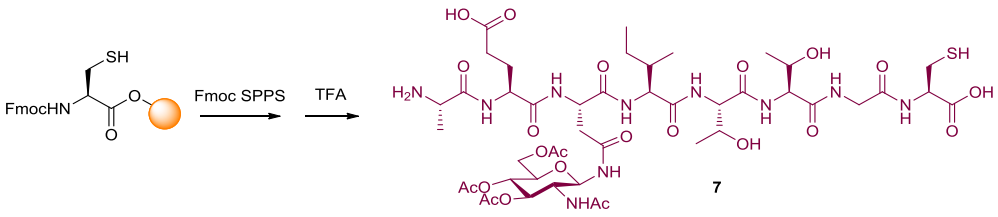
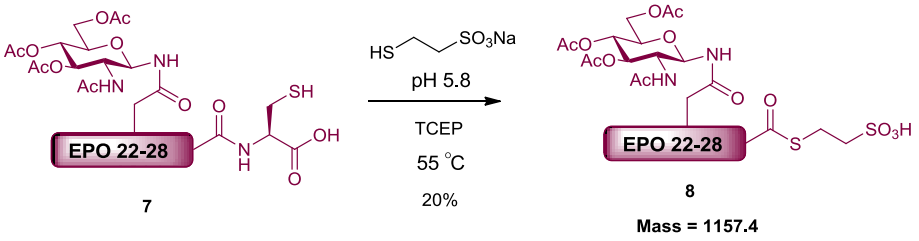


Figure 2.1: Fmoc SPPS of model glycopeptide corresponding to EPO residues 22-28-Cys

The glycopeptide **7** was treated with MESNa (10% w/v) and TCEP (0.5% w/v) in 0.1 M sodium phosphate buffer (pH 5.8) at 55 °C for 48 hours to afford the C-terminal thioester **8** (Scheme 2.3)**.** The reaction was monitored by analytical (RP)HPLC and indicated a smooth transition to the thioester (Figure 2.2), confirmed by LC-MS (Figure 2.3) and NMR after purification (Figure 2.4).



Scheme 2.3: Thioester formation of model glycopeptide via *N→S* acyl transfer

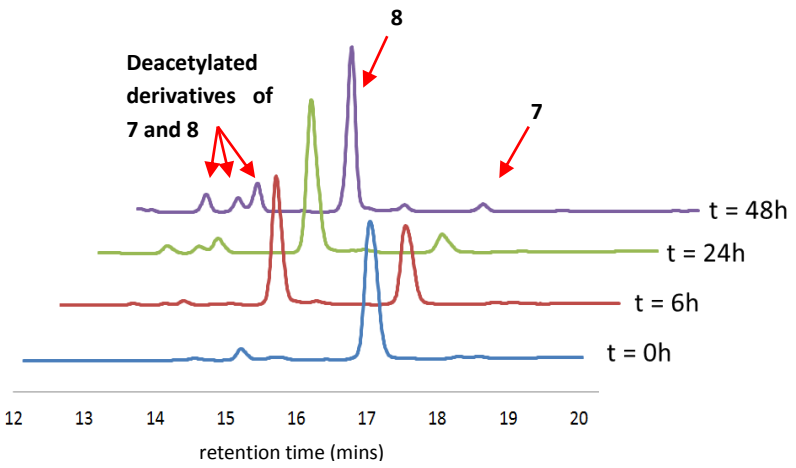


Figure 2.2: Analytical (RP)HPLC data with UV absorption at 230nm monitoring formation of thioester (spectra timepoints offset by increments of 0.5 mins)

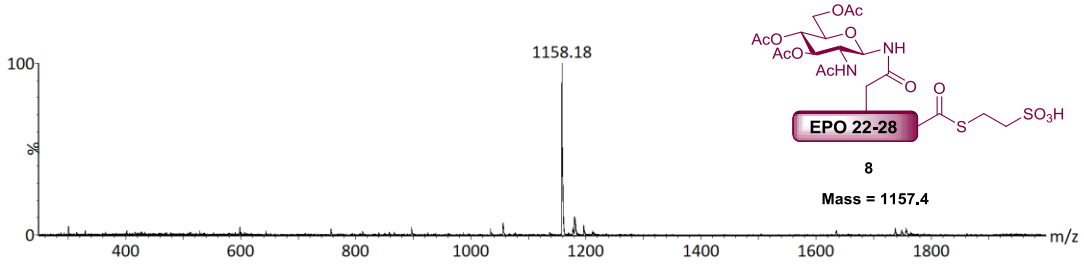


Figure 2.3: ES+ mass spectrum of glycopeptide thioester 8

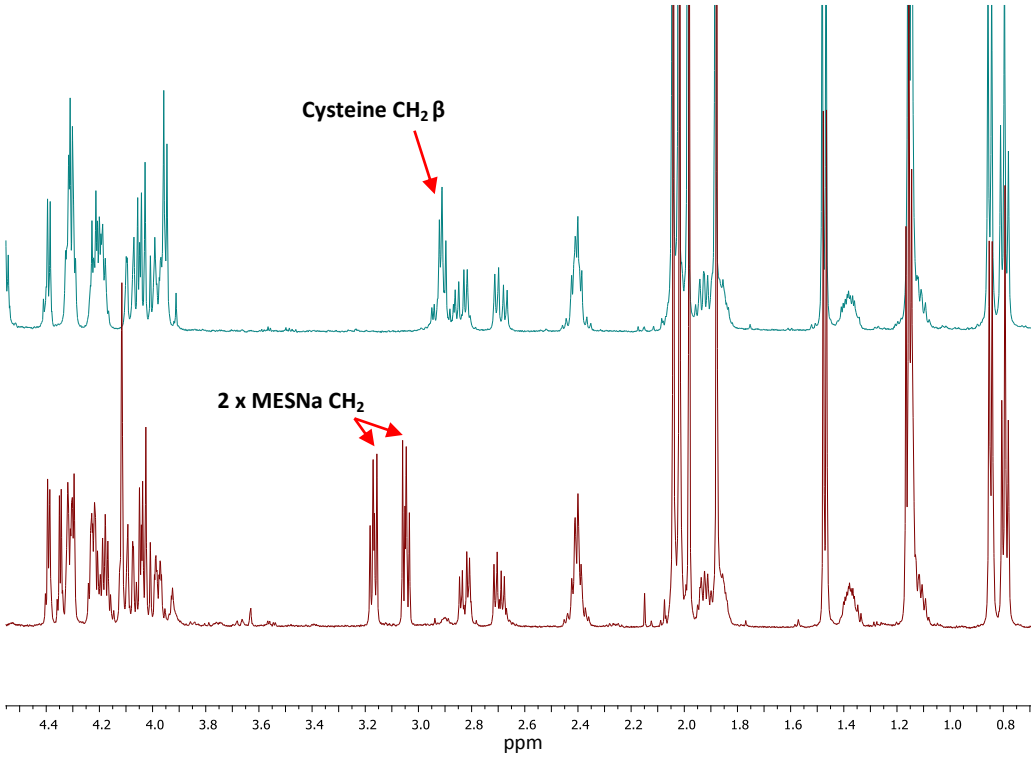


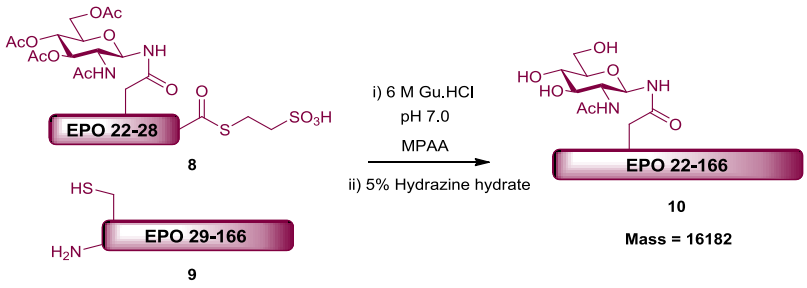
Figure 2.4: ^{1}H NMR (D_2O) comparison of the cysteinyl peptide 7 (upper trace) and MESNa thioester 8 (lower trace) showing the appearance of MESNa CH_2's and the absence of the cysteine CH_2 β

During thioester formation, it was postulated that the free cysteine released could reverse the reaction through NCL back to the starting material. To test this theory, the peptide concentration in the reaction was lowered from 1 mg/mL to 0.2 mg/mL. It was envisaged that the cysteine residue released upon thioester formation would remain at a relatively lower concentration, reducing the extent of the reverse reaction. The other reaction components were kept constant to ensure thioester formation could take place at the same rate and after purification, improved yields were attained (~ 40%).

2.3.2 Native Chemical Ligation of model glycopeptide thioester

To confirm the successful formation of the desired thioester, the purified glycopeptide was ligated to a recombinant protein corresponding to EPO residues 29-166 (donated by Dr Derek Macmillan). The ligation reaction was carried out in 6 M Guanidine hydrochloride (Gu.HCl) buffered with 0.3 M sodium phosphate (pH 7.0). The inclusion of mercaptophenyl acetic acid (100 mM) catalyses the reaction by forming the corresponding aryl thioester *in situ* and LC-MS indicated completion

of the ligation reaction within 3 hours. A final treatment of hydrazine hydrate in the presence of DTT afforded the deacetylated glycoprotein **10** (Scheme 2.4).



Scheme 2.4: Model glycopeptide thioester ligation

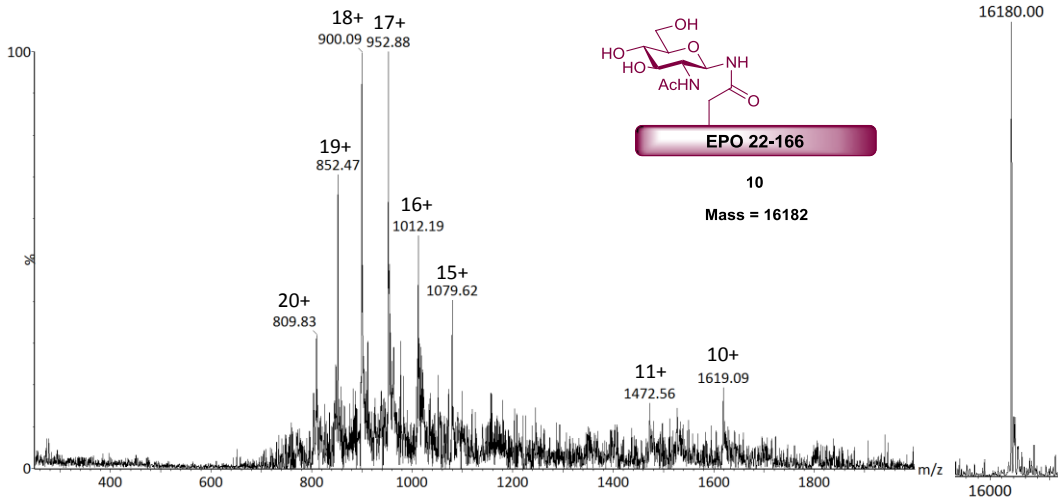
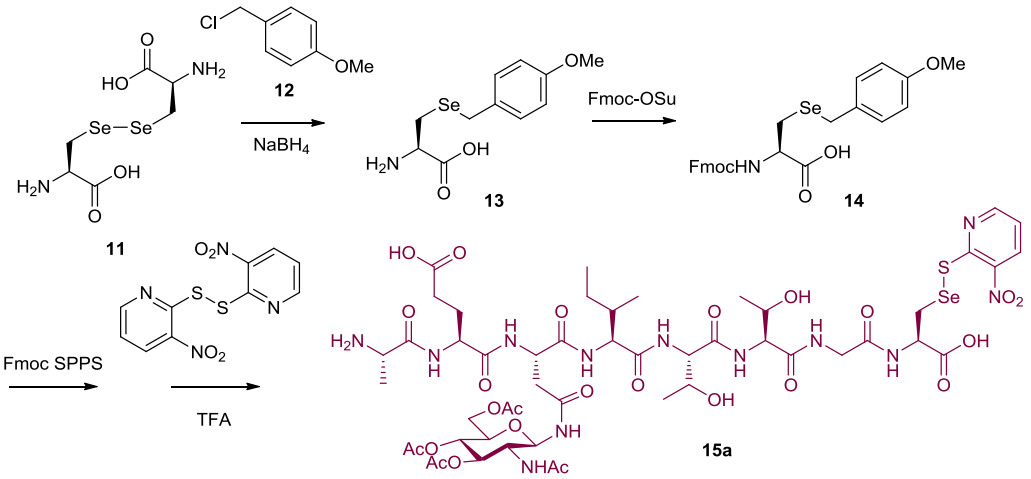


Figure 2.5: ES+ mass spectrum and deconvoluted mass of EPO 22-166

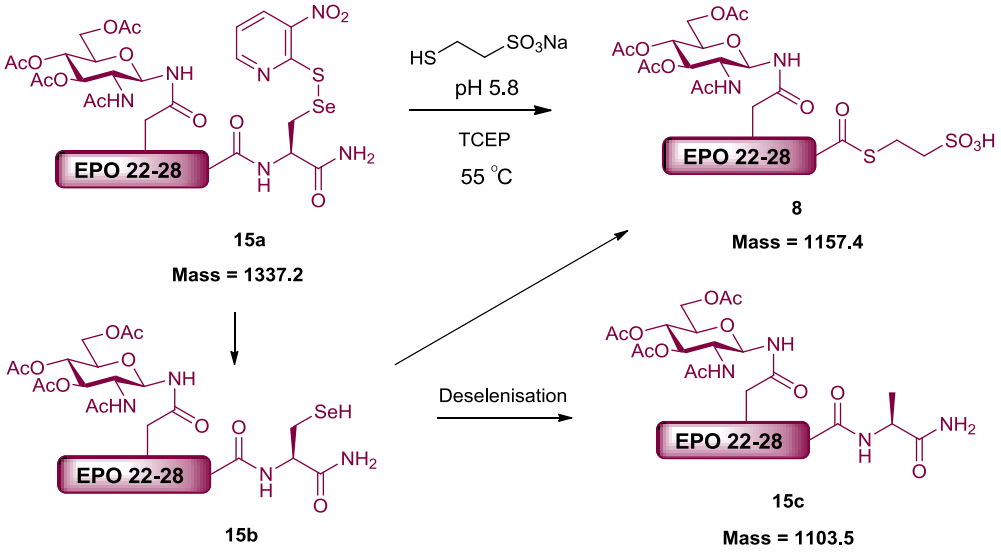
2.3.3 Native *N*-glycopeptide thioester synthesis via *N→Se* acyl transfer

With the successful application of *N→S* acyl transfer towards glycopeptide thioesters, the use of selenocysteine was explored to generate thioesters via *N→Se* acyl shift.[158] The model glycopeptide was resynthesised with a C-terminal selenocysteine, which was concomitantly protected with the 3-nitro-2-pyridinesulfenyl (Npys) group during cleavage from solid support (Scheme 2.5).[205]



Scheme 2.5: Synthesis of Fmoc-Sec(PMB)-OH and incorporation into model glycopeptide

N→Se acyl shift mediated thioester formation of the glycopeptide was attempted as before with MESNa (10% w/v) and TCEP (0.5% w/v) in 0.1 M sodium phosphate buffer (pH 5.8) at 55 °C (Scheme 2.6). From previous experiments with selenocysteine, it was suspected that oxygen could promote TCEP mediated deselenisation and hence the reaction mixture was degassed prior to the addition of TCEP.[85] Although a small amount of the thioester was detected by LC-MS after 48 hours, a significant amount of peptide was still found to have irreversibly deselenised to **15c** (Figure 2.6).



Scheme 2.6: Thioester formation of model glycopeptide via *N→Se* acyl transfer

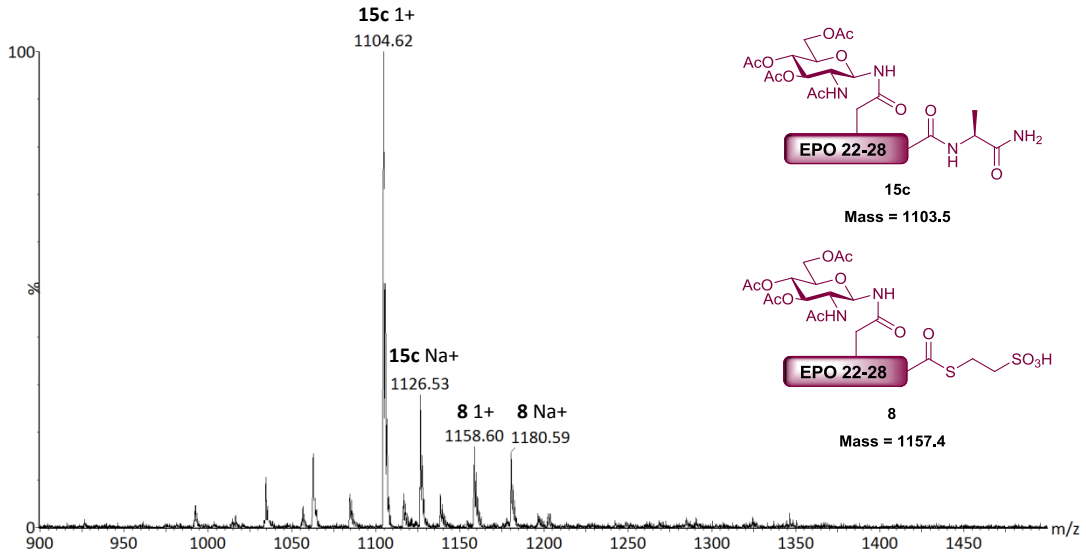


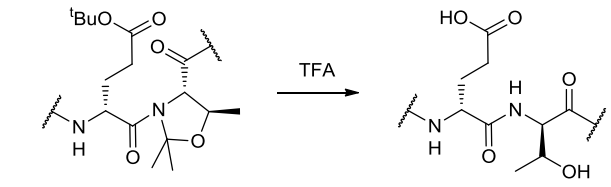
Figure 2.6: ES+ mass spectrum showing progress of thioester formation via *N→Se* acyl transfer after 48 hours. The major species corresponds to the deselenised peptide **15c.**

The use of ascorbate, which had recently been shown to prevent TCEP-mediated desulfurisation, was also unable to improve the outcome of the reaction.[206]

2.3.4 Synthesis of IFNβ 78-93 glycopeptide

With the *N→S* acyl transfer method still a promising method of generating glycopeptide thioesters, attention was focussed on the synthesis of the IFNβ 78-93. The peptide was synthesised using Fmoc SPPS on preloaded Cys-NovaSyn TGT resin (0.05 mmol scale) and acetamidomethyl protection was employed for the N-terminal cysteine. Residues 81-93 were assembled through automated synthesis, using 10 equivalents for each coupling. A small amount of resin was cleaved, allowing the efficiency of peptide synthesis to be determined by LC-MS. Encouragingly, the major species corresponded to the expected peptide.

The manual coupling of the glycoamino acid was attempted and subsequently followed by a microscale cleavage to assess coupling efficiency. LC-MS revealed that the coupling had only been partially efficient and so the coupling of the glycoamino acid was repeated. Once again, analytical scale cleavage indicated poor coupling of the glycoamino acid. In an attempt to progress this coupling step, the use of stronger coupling reagents PyBOP and HATU were evaluated. Disappointingly, neither seemed to improve coupling efficiency and it was considered that peptide aggregation could be the cause of poor coupling reactions. To overcome this, a commercially available pseudoproline dipeptide was incorporated instead of Glu-Thr at positions 81-82. Pseudoprolines have been reported to prevent the formation of secondary structures during peptide synthesis by introducing kinks into the peptide chain. Upon cleavage with TFA, the structure reverts to the native dipeptide (Scheme 2.7).[207]



-Glu(O^{t}Bu)-Thr(psiMe,Mepro)- -Glu-Thr-

Scheme 2.7: Pseudoproline dipeptides and their conversion to native structures upon treatment with TFA

Pleasingly, the incorporation of the Fmoc-Glu(O^{t}Bu)-Thr(psiMe,Mepro)-OH pseudoproline alleviated the difficulties encountered earlier. Following the final couplings and cleavage from the resin, the crude glycopeptide was deacetylated with aqueous hydrazine (5% v/v). (RP)HPLC afforded the desired product **16** (6%), confirmed by mass spectrometry (Figure 2.7).

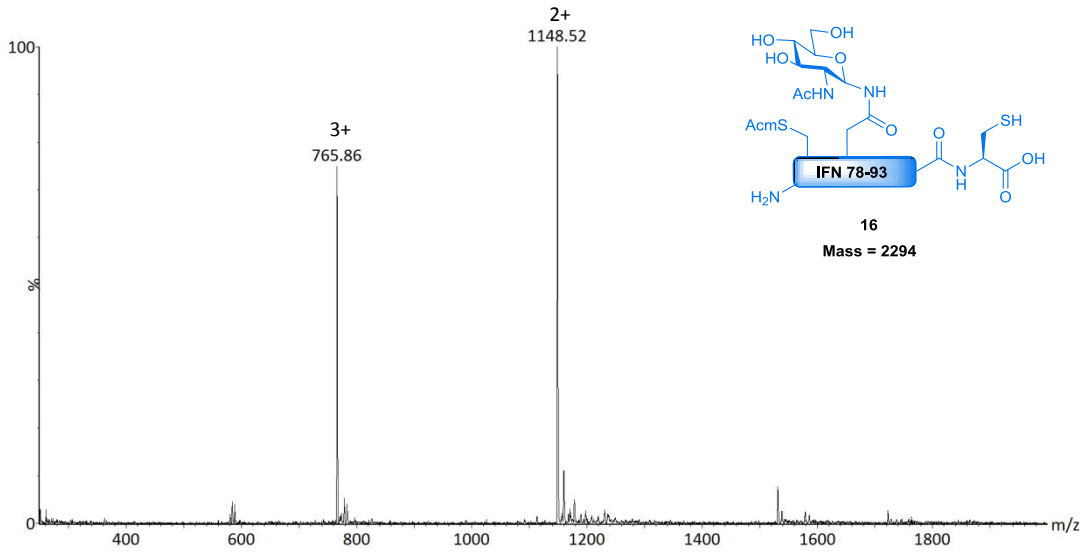
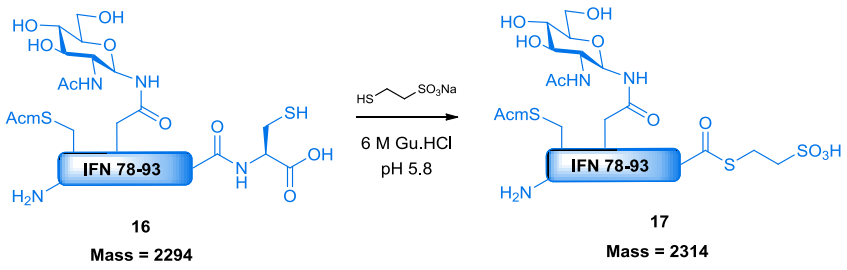


Figure 2.7: ES+ mass spectrum confirming successful synthesis of IFNβ 78-93 glycopeptide 16

2.3.5 Application of *N→S* acyl transfer towards IFNβ 78-93 glycopeptide thioester

Following glycopeptide assembly, the application of the thioester formation via *N→S* acyl transfer was investigated. Unlike the model EPO sequence, this peptide necessitated the use of chaotropes to remain in solution, although previous studies had shown that 6 M Gu.HCl was compatible with thioester forming reactions.[130, 208] Peptide **16** was subjected to a treatment of MESNa (10% w/v) and TCEP (0.5% w/v) in 6 M Gu.HCl/0.1 M sodium phosphate buffer (pH 5.8) at 55 °C and monitored via LC-MS over 48 hours (Figure 2.8).



Scheme 2.8: Thioester formation of IFNβ 78-93 glycopeptide via *N→S* acyl transfer

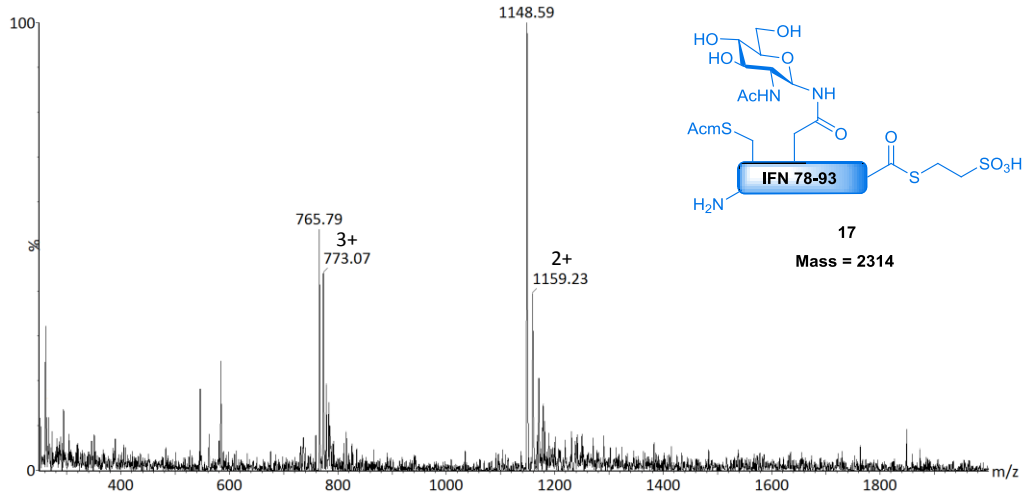


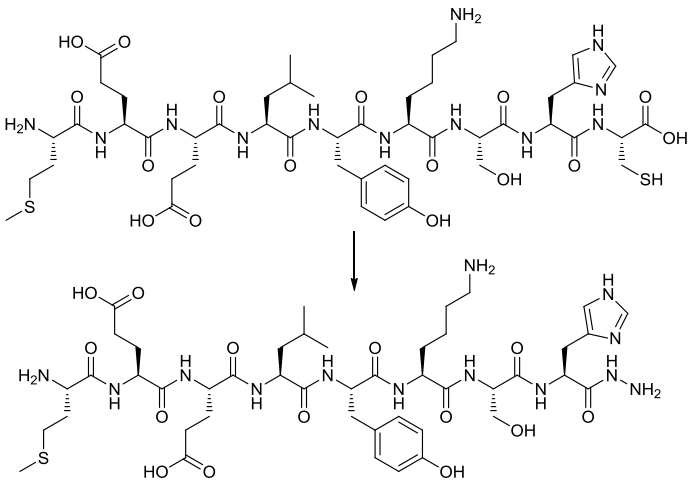
Figure 2.8: ES+ mass spectrum analysis of thioester formation of IFNβ 78-93 after 48 hours

LC-MS analysis suggested that thioester formation over 48 hours was slower compared to previous experiments. This was consistent with previous observations within the research group, where the use of 6 M Gu.HCl was found to slow the rate of reaction. Peptide aggregation was also observed over the course of the reaction and was believed to contribute to slower conversion. Although the reaction was slow, the product was isolated by (RP)HPLC, albeit in low yield (<5%).

2.3.6 *N→S* acyl transfer in the synthesis of glycopeptide hydrazides

At the time of this study, independent investigations by the groups of Liu and Cotton suggested that hydrazine could intercept *S*-acyl intermediates to form acyl hydrazides.[135, 171] Peptide hydrazides exhibit improved stability towards hydrolysis and can be converted to the corresponding thioester under mild conditions. This led to an investigation of whether the inclusion of hydrazine in an *N→S* acyl transfer reaction could allow the generation of the corresponding peptide hydrazide. It was hypothesised that the improved stability of the product would allow it to accumulate under the reaction conditions, thus enabling the reaction to be carried out over extended periods of time.

Preliminary studies were performed on model peptide "MEELYKSHC" (Scheme 2.9) using hydrazinium acetate as the source of hydrazine, in the presence and absence of MESNa (Figure 2.9).[209]



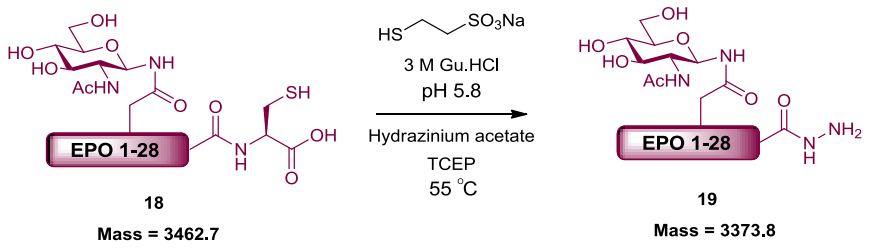
Scheme 2.9: Hydrazinolysis of model peptide MEELYKSHC

N_2H_4·HOAc [% w/v]	T (°C)	MESNa [% w/v]	MEELYKSHC (%)* 24 h	48 h
5	60	10	5	0
5	50	10	27	5
5	40	10	28	19
0	50	10	33	30
5	50	5	33	21
2.5	50	10	-†	5
5	50	0	-†	30

Figure 2.9: Conditions used to investigate hydrazinolysis via *N→S* acyl transfer
*Reaction progress was judged by the quantity of starting material remaining, determined by (RP)HPLC
†Not determined

Although reactions were conducted in 0.1 M sodium phosphate buffer (pH 5.8), it is important to note that the final pH of the reaction mixture was found to be approximately 7, potentially slowing the process of *N→S* acyl transfer. Hydrazinolysis proceeded smoothly over a range of temperatures and was found to be more efficient compared to direct thioester formation. Interestingly, when MESNa was excluded from the reaction there was no evidence of hydrazide formation by LC-MS. Instead, the predominant species indicated the formation of peptide dimers and oxidation of methionine.

To ensure compatibility of this reaction with *N*-glycopeptides, the hydrazinolysis of a glycopeptide corresponding to EPO residues 1-28 was investigated. The peptide was treated with MESNa (10% w/v), hydrazinium acetate (5% w/v) and TCEP (0.5% w/v) in 3 M Gu.HCl/0.1 M sodium phosphate buffer (pH 5.8) at 55 °C. Conversion to the hydrazide **19** was monitored by LC-MS and took place over 48 hours (Figure 2.10).



Scheme 2.10: Hydrazinolysis of EPO 1-28 glycopeptide via *N→S* acyl transfer

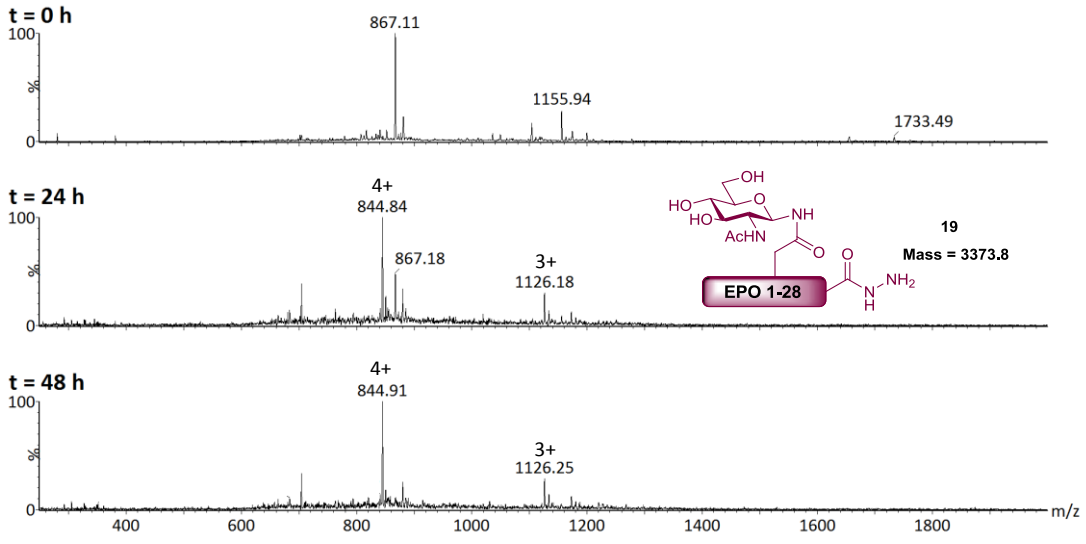
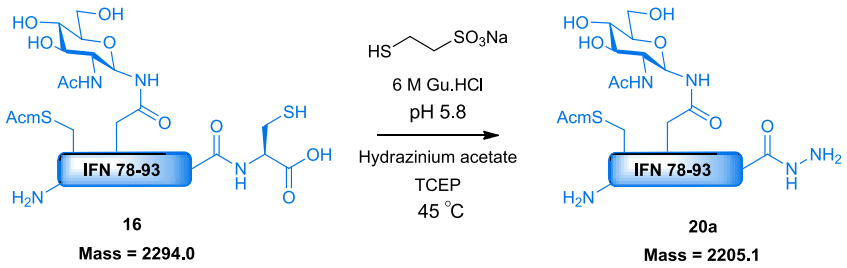


Figure 2.10: ES+ mass spectrum analysis of hydrazinolysis of EPO 1-28 over 48 hours

2.3.7 Application of *N→S* acyl transfer towards IFNβ 78-93 glycopeptide hydrazide

Following this success, the hydrazide forming reaction was applied to the glycopeptide corresponding to IFNβ 78-93. The reaction was conducted in 6 M Gu.HCl/0.1 M sodium phosphate buffer (pH 5.8) in the presence of MESNa (10% w/v), hydrazinium acetate (5% w/v) and TCEP (0.5% w/v). To minimise peptide aggregation, the reaction temperature was lowered to 45 °C. LC-MS analysis showed transition to the hydrazide **20a** over 120 hours and the product was successfully isolated by (RP)HPLC.



Scheme 2.11: Hydrazinolysis of IFNβ 78-93 glycopeptide via *N→S* acyl transfer

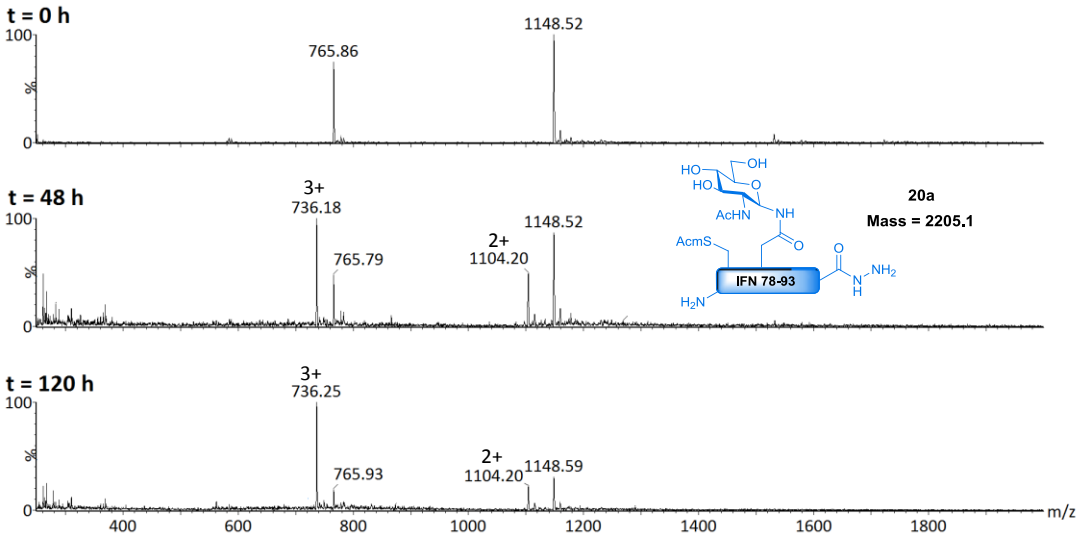
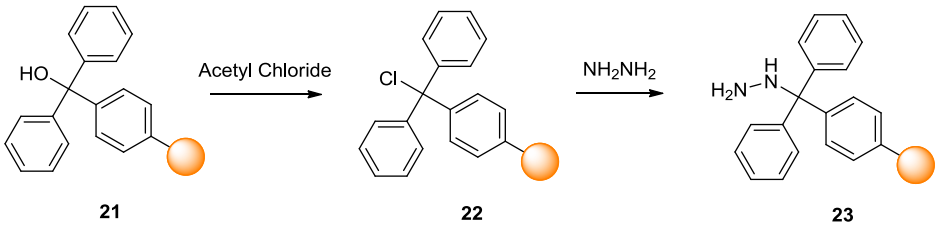


Figure 2.11: ES+ mass spectrum analysis of hydrazinolysis of IFNβ 78-93 over 120 hours

2.3.8 Direct synthesis of IFNβ 78-93 glycopeptide hydrazide by SPPS

Despite the novel hydrazinolysis reaction providing the desired product, existing methods allowed the direct synthesis of peptide hydrazides by Fmoc SPPS.[135, 210] Often during the synthesis of peptide thioesters with an N-terminal cysteine, protection is necessary to prevent the undesired cyclisation of the peptide.[157] However, for the synthesis a latent peptide thioester in the form of a hydrazide, it was envisaged that N-terminal cysteine protection would not be necessary, consequently avoiding an additional deprotection and purification step during protein assembly.

In the first instance, NovaSyn TGT alcohol resin **21** (0.21 mmol/g) was converted to the chloride derivative **22** followed by displacement with hydrazine (Scheme 2.12).

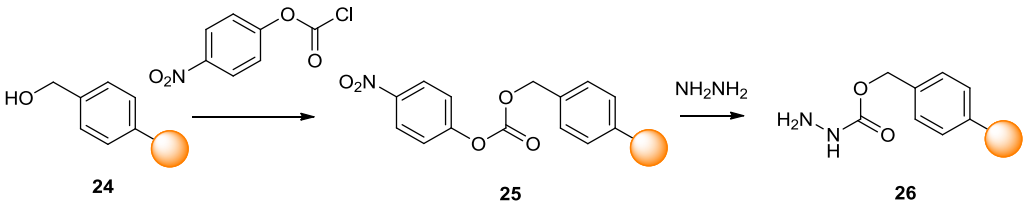


Scheme 2.12: Synthesis of NovaSyn TGT hydrazine resin

Fmoc-His(Trt)-OH was double-coupled with HBTU/HOBt and loading was determined by Fmoc analysis to be 0.04 mmol/g. The remainder of the sequence was completed by manual synthesis as before, using 10 equivalents of each Fmoc amino acid and HBTU/HOBt (except for Fmoc-Glu(O^{t}Bu)-Thr(psiMe,Mepro)-OH and the glycoamino acid where 5 equivalents were used). Cleavage from the resin was enabled by a treatment of TFA/thioanisole/water/phenol/EDT (82.5:5:5:5:2.5) and after precipitation in ether, the crude glycopeptide was deacetylated with aqueous hydrazine (5%). Disappointingly, the overall yield after purification by (RP)HPLC was ~ 1 mg (5%).

At this point, the preparation of a hydrazine functionalised Wang (para-alkoxybenzyl alcohol) resin was investigated according to a procedure reported by Liu.[135] Commercially available Wang resin (1.3 mmol/g loading) was initially treated

with para-nitrophenyl chloroformate, followed by hydrazine hydrate to afford the hydrazine functionalised Wang resin **26** (Scheme 2.13).



Scheme 2.13: Synthesis of hydrazine functionalised Wang resin

Fmoc-His(Trt)-OH was coupled to the resin and loading was determined to be 0.67 mmol/g. In the first attempt to synthesise the IFNβ 78-93 glycopeptide, analytical cleavage was performed after coupling of residue 85. Unexpectedly, LC-MS analysis suggested that peptide synthesis was not entirely efficient, with a number of deletions observed (Figure 2.12).

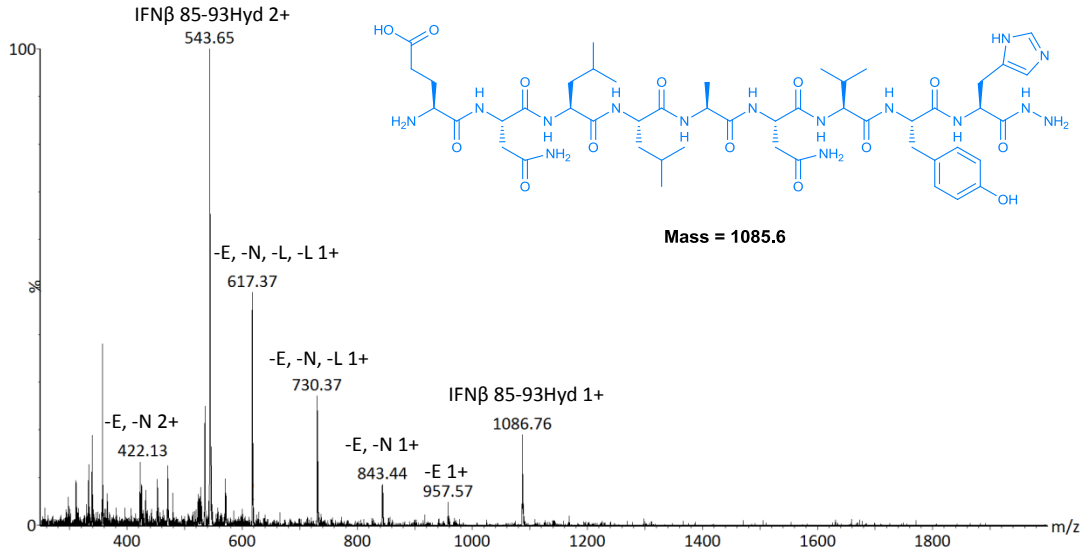


Figure 2.12: ES+ mass spectrum of crude IFNβ 85-93 hydrazide

The poor coupling efficiencies were attributed to the high loading of the resin. During the second attempt to synthesise the peptide, the hydrazine resin was loaded with a mixture of Fmoc-His(Trt)-OH and Boc-Ala-OH (1:2). The inclusion of Boc protected alanine in the first coupling prevents further elongation during Fmoc SPPS, thus effectively lowering the loading of the resin. The new resin loading was determined by Fmoc analysis to be 0.15 mmol/g and peptide elongation was continued as before. Finally, the peptide hydrazide was cleaved from the resin, deacetylated and purified by (RP)HPLC with a final yield of 6% (Figure 2.13).

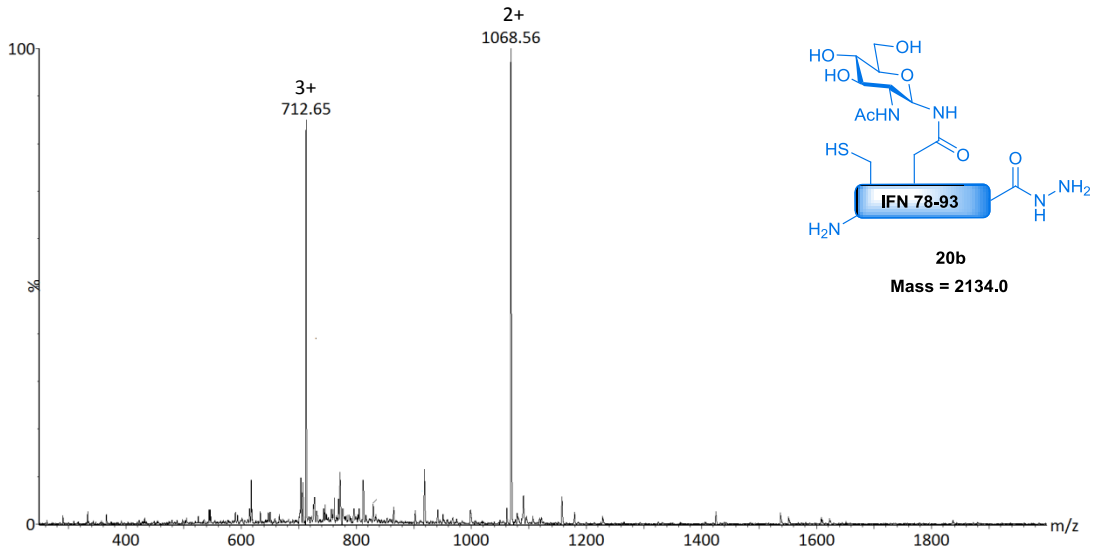


Figure 2.13: ES+ of purified IFNβ 78-93 glycopeptide hydrazide

To enable a flexible approach to IFNβ assembly, the glycopeptide was also synthesised with an Acm protected N-terminal cysteine. For the synthesis of this glycopeptide hydrazide, commercially available trityl chloride resin (1.4 mmol/g) was functionalised according to an updated protocol reported by Liu.[210] After treatment with hydrazine hydrate and coupling of Fmoc-His(Trt)-OH, the loading was found to have decreased to 0.45 mmol/g. Wary of previous difficulties associated with higher resin loading, the second coupling was performed with a mixture of Fmoc-Tyr(O^{t}Bu)-OH and Boc-Ala-OH (1:1). Resin loading was determined to have decreased to 0.18 mmol/g, and the synthesis of the peptide was completed as before.

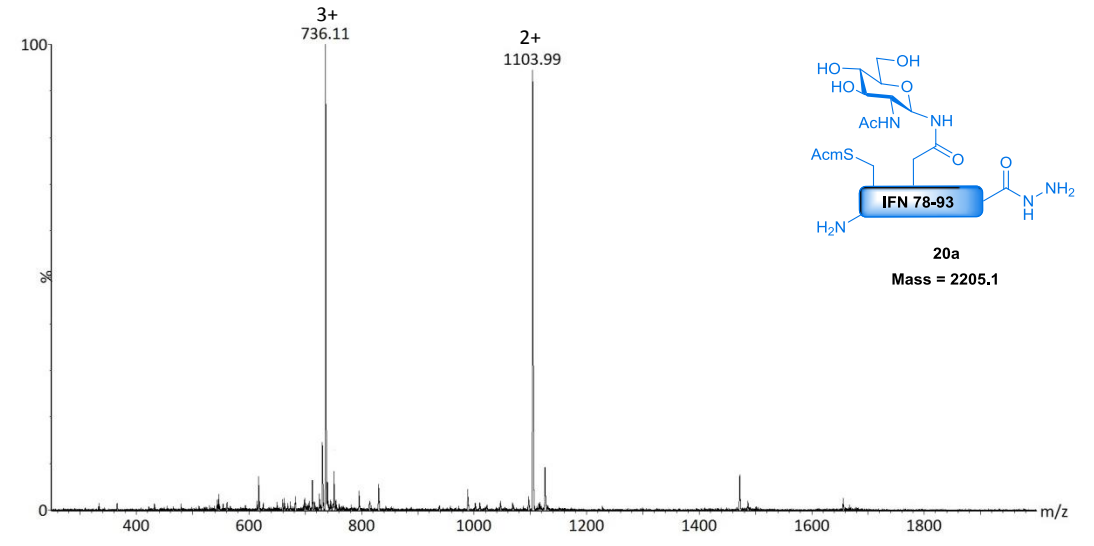


Figure 2.14: ES+ of purified IFNβ 78-93 glycopeptide hydrazide (Acm protected)

2.4 Conclusions

In summary, the novel application of *N*→*S* acyl transfer towards the synthesis of native *N*-linked glycopeptide thioesters and hydrazides has been demonstrated. This methodology is particularly amenable to soluble peptide sequences, but difficulties were encountered with hydrophobic segments. In the case of such sequences, hydrazinolysis via *N*→*S* acyl transfer offers a significant advantage. The peptide hydrazide product can accumulate without breakdown over extended periods of time, which is often necessitated due to lower reaction temperatures and the use of denaturants.

The expanding use of peptide hydrazides as thioester surrogates has gained significant popularity due to its compatibility with Fmoc SPPS. The direct synthesis of peptide hydrazides is facilitated by the use of functionalised resins and was shown to be compatible with the Fmoc synthesis of native *N*-glycopeptides. Importantly, this platform provides a convenient and attractive route towards peptide thioester surrogates and will undoubtedly broaden the scope of protein assembly through NCL.

3 Expression and functionalisation of proteins for the semi-synthesis of Interferon β-1

3.1 Introduction

The development of recombinant methods allowing expression of protein in a bacterial host has provided an indispensable tool for obtaining large unmodified proteins. Recombinant DNA technology allows a gene corresponding to a desired protein to be inserted into a bacterial vector. Following transformation into a bacterial host, the protein can be overexpressed in a controlled manner. For the semi-synthesis of Interferon β-1 through NCL, both an N-terminal fragment with a C-terminal thioester and a C-terminal fragment with an N-terminal cysteine (Figure 3.1) were required.

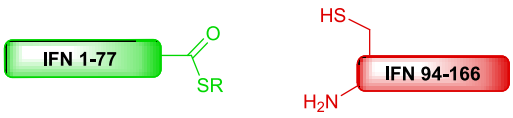


Figure 3.1: Recombinant fragments required for the semi-synthesis of IFNβ

3.2 Cloning and expression of IFNβ fragments for semi-synthesis

3.2.1 Cloning of IFNβ 1-77 intein fusion

With the previously mentioned technologies in mind (section 1.10.1.1), the proteins corresponding to IFNβ 1-77 and 94-166 were cloned into the relevant expression vectors. The sequence corresponding to IFNβ residues 1-77 was cloned into vector pTYB1 (Figure 3.2), which allows expression of IFNβ 1-77 with a C-terminal intein-chitin binding domain fusion.[161]

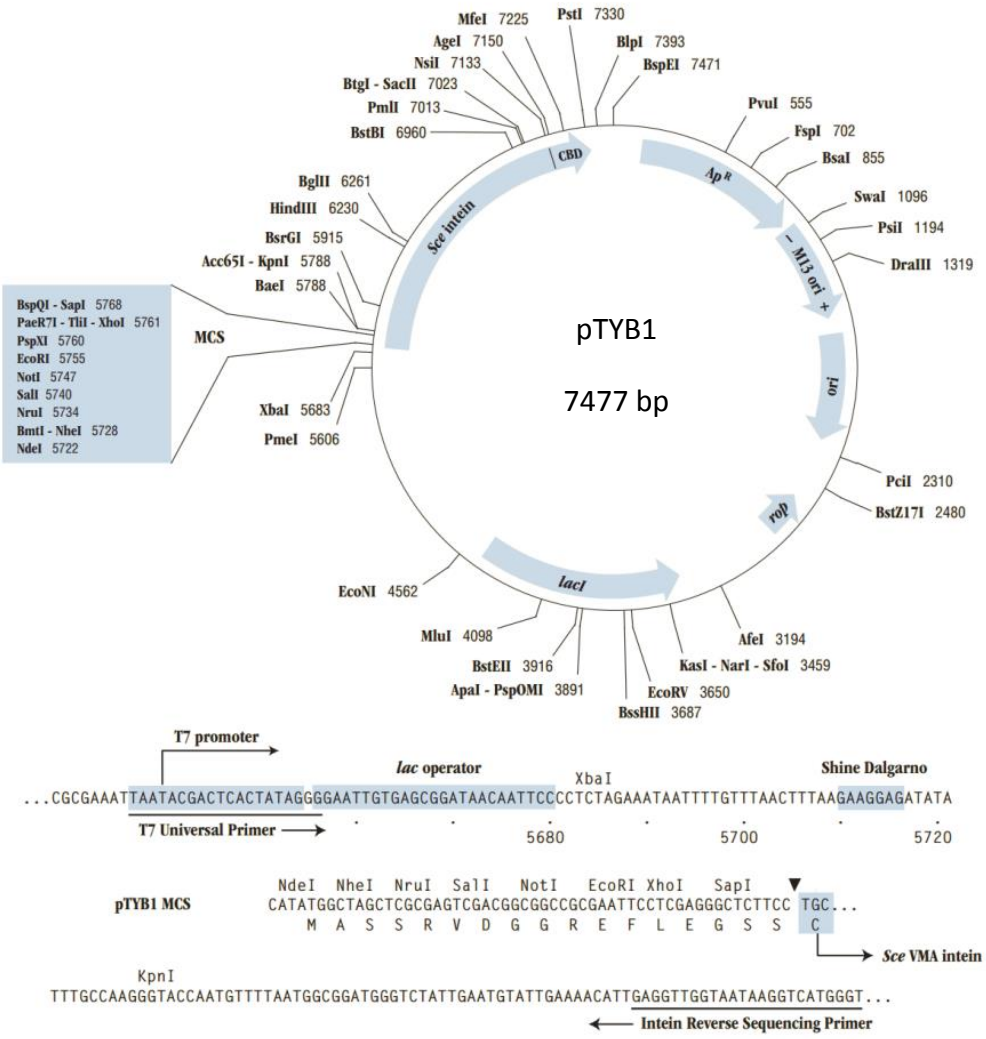


Figure 3.2: pTYB1 vector map

The gene corresponding to IFNβ 1-77 was cloned between the NdeI and SapI restriction sites (Figure 3.3) present in the multiple cloning site.

Figure 3.3: NdeI and SapI recognition sequences and cleavage positions

This strategy ensured that the expressed protein would be capable of splicing at the correct position, resulting in a product free of any vector encoded amino acids. NdeI and SapI are restriction enzymes that cleave double stranded DNA to produce sticky ends, enabling efficient DNA ligation (Figure 3.4). Primers were designed for the amplification of the gene corresponding to IFNβ 1-77 and also to introduce the T77G mutation (see section 8.3 for details).

pTYB1 cloning site:

```
5′                                   Intein                              3′
…  CAT  ATG  GCT  …  GGC  TCT  TCC  TGC  TTT  GCC  AAG  GGT  ACC  …
…  GTA  TAC  CGA  …  CCG  AGA  AGG  ACG  AAA  CGG  TTC  CCA  TGG  …
3′    NdeI                SapI                              KpnI    5′
```

pTYB1 cloning site after treatment with NdeI/SapI:

```
5′                  Intein                    3′
…  CA               TGC  TTT  GCC  AAG  GGT  ACC  …
…  GTA  T               AAA  CGG  TTC  CCA  TGG  …
3′                                    KpnI    5′
```

PCR amplified gene corresponding to IFNβ 1-77:

```
5′                IFNβ 1-77                    3′
…  CAT  ATG  AGC  …  GGC  TGC  GGA  AGA  GCA  …
…  GTA  TAC  TCG  …  CCG  ACG  CCT  TCT  CGT  …
3′    NdeI                          SapI    5′
```

PCR amplified gene corresponding to IFNβ 1-77 after treatment with NdeI/SapI:

```
5′              IFNβ 1-77          3′
   T  ATG  AGC  …  GGC
      AC  TCG  …  CCG  ACG
3′                               5′
```

DNA ligation product:

```
5′                IFNβ 1-77   Intein                          3′
…  CAT  ATG  AGC  …  GGC  TGC  TTT  GCC  AAG  GGT  ACC  …
…  GTA  TAC  TCG  …  CCG  ACG  AAA  CGG  TTC  CCA  TGG  …
3′    NdeI                                        KpnI    5′
```

Figure 3.4: NdeI/SapI cloning strategy of IFNβ 1-77 gene into vector pTYB1

PCR was conducted from a plasmid containing the IFNβ gene cloned into pKYB1 (donated by Dr Derek Macmillan). The vector and PCR product were treated with restriction endonucleases NdeI and SapI to prepare sticky ends for subsequent ligation. After purification by preparative agarose gel, DNA ligation was attempted, followed by transformation into competent cells. Disappointingly, multiple attempts to clone the gene corresponding to IFNβ 1-77 into pTYB1 were unsuccessful, determined by DNA sequencing and agarose gel analysis. It was postulated that poor ligation efficiency could be attributed to the unusual activity of SapI, which cleaves outside of its recognition sequence. As a consequence, the

recognition site remains intact after cleavage, possibly resulting in a poor turnover rate and inefficient cleavage by SapI. At this stage, an alternative ligation at the KpnI site was considered. The reverse primer was redesigned to encode the KpnI recognition sequence (Figure 3.5) and the N-terminal residues of the intein lost during KpnI digest of the vector, resulting in the originally desired construct (Figure 3.6).

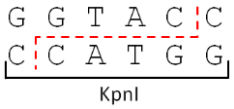


Figure 3.5: KpnI recognition sequence and cleavage position

pTYB1 cloning site:

```
5′                                        Intein                                  3′
…   CAT  ATG  GCT  …  GGC  TCT  TCC  TGC  TTT  GCC  AAG  GGT  ACC  …
…   GTA  TAC  CGA  …  CCG  AGA  AGG  ACG  AAA  CGG  TTC  CCA  TGG  …
3′    NdeI                  SapI                              KpnI    5′
```

pTYB1 cloning site after treatment with NdeI/KpnI:

```
5′                          3′
…   CA                 C    …
…   GTA  T        CA   TGG  …
3′                          5′
```

Intein codons lost from KpnI digest are accounted for in the reverse primer, leading to their incorporation in the PCR product

PCR amplified gene corresponding to IFNβ 1-77:

```
5′              IFNβ 1-77    Intein                          3′
…   CAT  ATG  AGC  …  GGC  TGC  TTT  GCC  AAG  GGT  ACC  …
…   GTA  TAC  TCG  …  CCG  ACG  AAA  CGG  TTC  CCA  TGG  …
3′    NdeI                                      KpnI    5′
```

PCR amplified gene corresponding to IFNβ 1-77 after treatment with NdeI/KpnI:

```
5′              IFNβ 1-77    Intein                          3′
      T  ATG  AGC  …  GGC  TGC  TTT  GCC  AAG  GGT  AC
         AC   TCG  …  CCG  ACG  AAA  CGG  TTC  C
3′                                                          5′
```

DNA ligation product:

```
5′              IFNβ 1-77    Intein                          3′
…   CAT  ATG  AGC  …  GGC  TGC  TTT  GCC  AAG  GGT  ACC  …
…   GTA  TAC  TCG  …  CCG  ACG  AAA  CGG  TTC  CCA  TGG  …
3′    NdeI                                      KpnI    5′
```

Figure 3.6: NdeI/KpnI cloning strategy of IFNβ 1-77 gene into vector pTYB1

After PCR amplification, the product and vector were treated with NdeI/KpnI. The DNA ligation was attempted and subsequently transformed into cloning hosts. After colony screening, successful cloning was confirmed by sequencing.

3.2.2 Expression of IFNβ 1-77 intein fusion

The plasmid was transformed into BL21(DE3) cells to enable protein expression, which was carried out at 30 °C for 5 hours. Analysis of the soluble fraction and whole cell lysate (WCL) prepared by Bugbuster Protein Extraction Reagent (Millipore) was carried out by SDS-PAGE (Figure 3.7). A number of conditions for sample preparation were evaluated, to ensure that cleavage of the intein was not induced by DTT or heat.

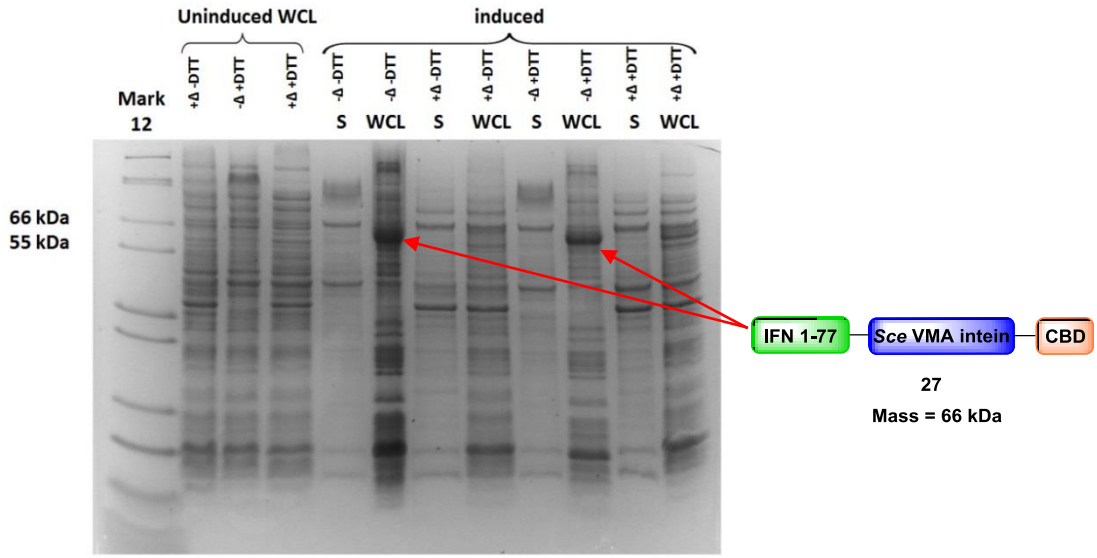


Figure 3.7: SDS-PAGE analysis after expression of IFNβ 1-77 intein fusion

A band corresponding to the expected mass of the IFNβ 1-77 intein fusion **27** (~ 66 kDa) was observed, but it was found that the majority of this protein was not soluble. Nonetheless, the soluble fraction was passed down a chitin column and thiolytic cleavage was attempted with MESNa. After 24 hours at 4 °C, the column was eluted and analysed by SDS-PAGE. Unfortunately, product isolation was unsuccessful and instead the decision was made to investigate the refolding of the IFNβ 1-77 intein fusion from the insoluble fraction.

Insoluble material remaining from the initial lysis and centrifugation step was resuspended in binding buffer containing 6 M Gu.HCl, 500 mM NaCl and 20 mM Tris (pH 8.0) and stirred vigorously at room temperature for 30 minutes. The remaining insoluble cell debris was pelleted by centrifugation and the supernatant containing the solubilised protein was dialysed against 8 M urea containing 10 mM DTT, 500 mM NaCl, 20 mM Tris (pH 8.5). Refolding was attempted by slowly reducing the concentration of urea according to a protocol reported by New England Biolabs, with all buffers containing 500 mM NaCl, 20 mM Tris (pH 8.5) allowing at least 12 hours for each step. Dialysis into 6 M urea, 1 mM DTT was followed by 4 M urea, 1 mM DTT and finally 2 M urea, 1 mM reduced glutathione and 1 mM oxidised glutathione. At this stage formation of a precipitate was observed, which was separated by centrifugation and separated into two batches. The first batch was kept in 2 M urea to allow the investigation of *Sce* VMA intein cleavage, which has been reported to take place under these conditions.[211] The remainder was dialysed into the final renaturation buffer containing 1 mM reduced glutathione and 0.1 mM oxidised glutathione. Once again precipitate was formed and removal was achieved by centrifugation. The remaining fractions were analysed by SDS-PAGE (Figure 3.8).

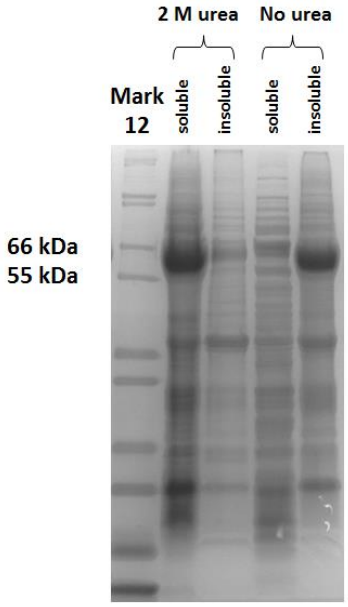


Figure 3.8: SDS-PAGE analysis after attempted refolding

It appeared that in the absence of urea, a significant amount of the fusion protein had precipitated out. Therefore, purification and cleavage was attempted in 2 M urea, initially passing the fraction through a chitin column. The column was washed and then flushed with buffer containing MESNa. The immediate flow through was collected and then the resin was incubated with MESNa for 24 hours. Flow through wash fractions and chitin beads were analysed by SDS-PAGE (Figure 3.9).

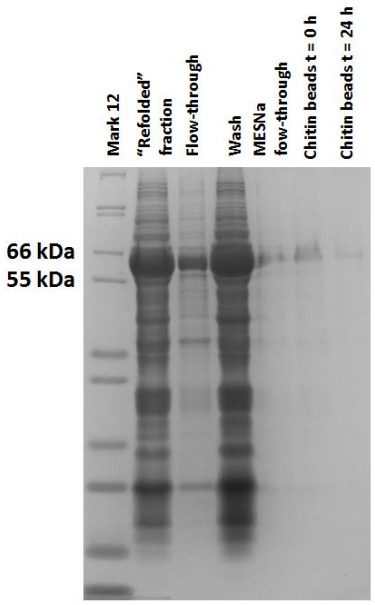


Figure 3.9: Attempted purification and cleavage of IFNβ 1-77 intein fusion in 2 M urea

From SDS-PAGE analysis, it was evident that the protein did not bind effectively to the resin. The poor affinity can be attributed to the presence of urea, which is likely to denature the chitin binding domain to some extent.

3.2.3 Cloning and expression of IFNβ 1-77 thioredoxin intein fusion

In another attempt to express a soluble and active form of the IFNβ 1-77 intein fusion, an N-terminal fusion protein was considered. Attention was drawn to the thioredoxin, which has previously been shown to improve the solubility and yield of proteins expressed in *E. coli*.[212] The pTYB1 construct was subcloned with the intein-chitin binding domain fusion into a pET32 Xa/LIC vector (Figure 3.10) using ligation independent cloning (LIC).[213]

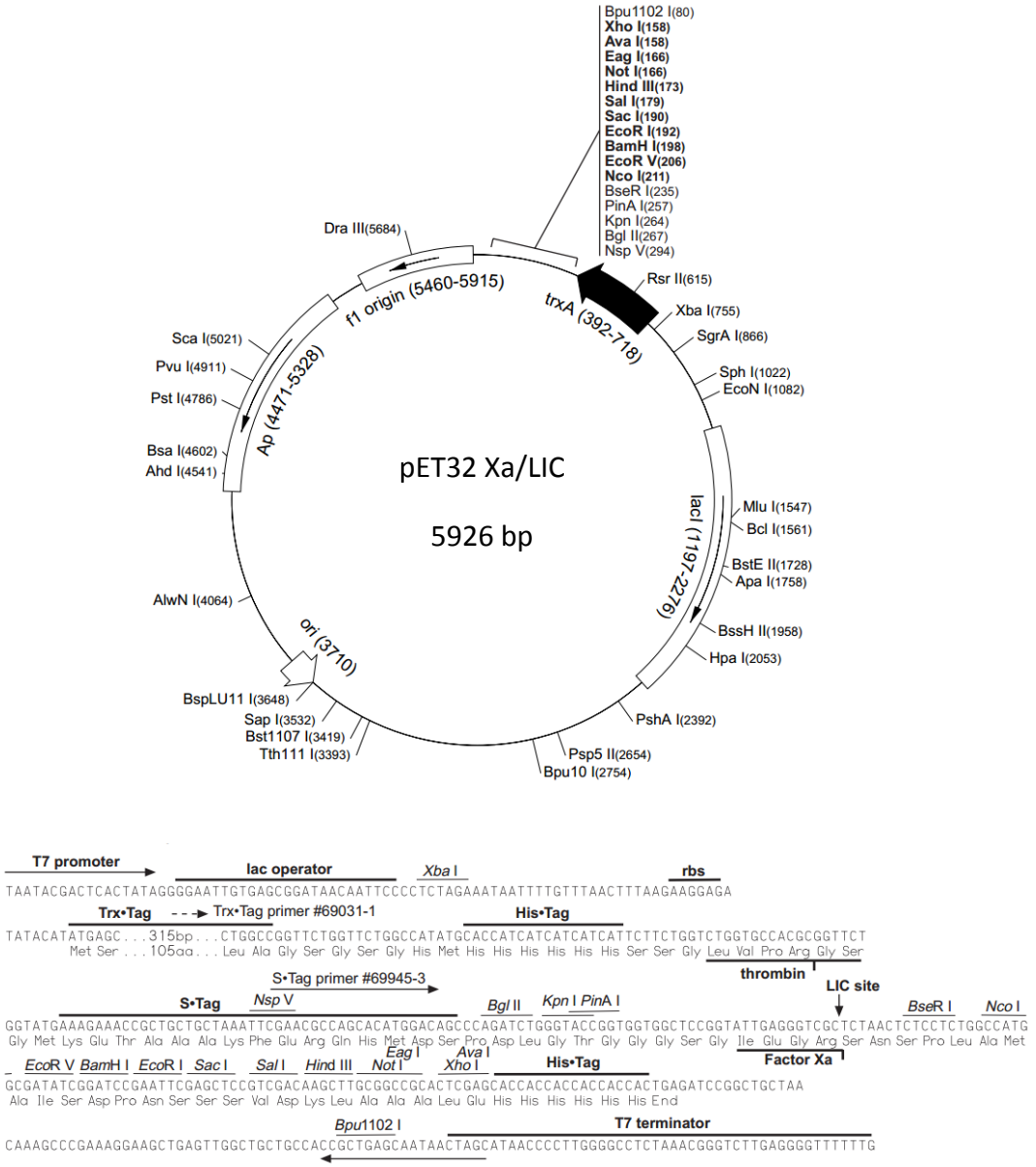


Figure 3.10: pET32 Xa/LIC vector map

For ligation independent cloning into pET32 Xa/LIC, the direct annealing of the PCR product into the vector is facilitated by GTP-free overhangs of 15 nucleotide bases. This strategy obviates the requirement for restriction digests and ligation reactions and overhangs are generated by T4 DNA polymerase which exhibits 3'-5' exonuclease activity. After annealing of the insert, the plasmid is transformed into *E. coli* and circularisation of the plasmid via formation of phosphodiester bonds, takes place within the cell.

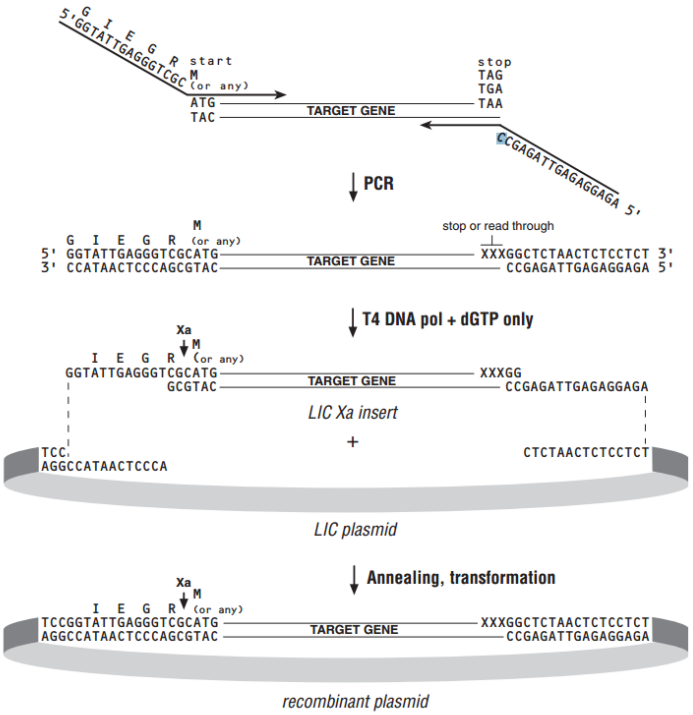


Figure 3.11: Ligation independent cloning with Xa/LIC strategy

After verifying the new construct, the protein was expressed at 30 °C for 5 hours at an analytical scale. The whole cell lysate and soluble fraction were analysed by SDS-PAGE (Figure 3.12), which unfortunately indicated that the thioredoxin fusion **28** was also insoluble.

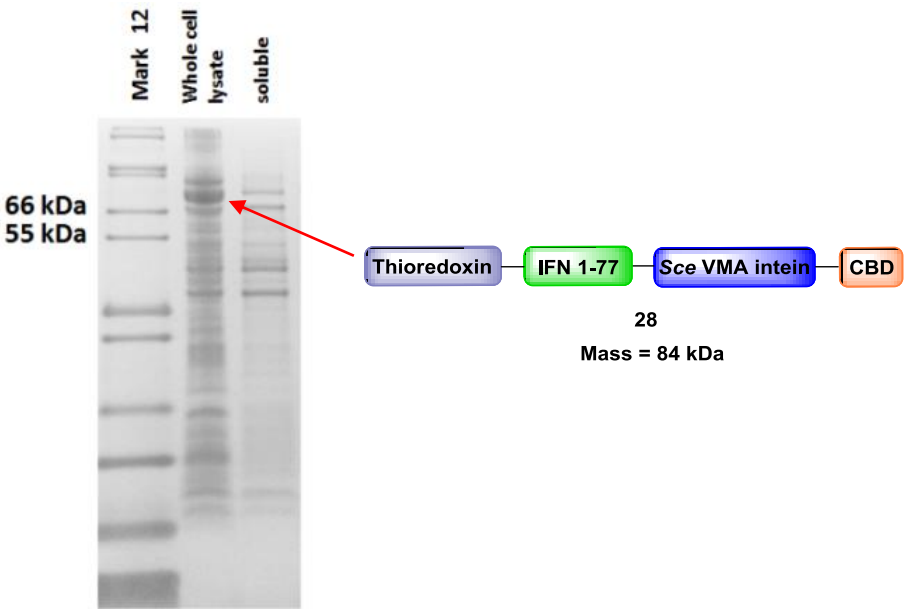


Figure 3.12: SDS-PAGE analysis after expression of IFNβ 1-77 thioredoxin intein fusion 84. Although the desired protein had been expressed, it appeared to be insoluble.

3.2.4 Cloning of IFNβ 1-77 with C-terminal Gly-Cys

It was concluded that an intein based strategy was unlikely to be successful towards obtaining the thioester of IFNβ 1-77. At this point, it was decided to investigate the formation of a thioester via an *N*→*S* acyl transfer reaction. The gene corresponding to IFNβ 1-77 was cloned into pET16b (Figure 3.13) which introduces an N-terminal His_{10} tag, allowing the protein of interest to be purified using Ni^{2+} affinity chromatography.[214]

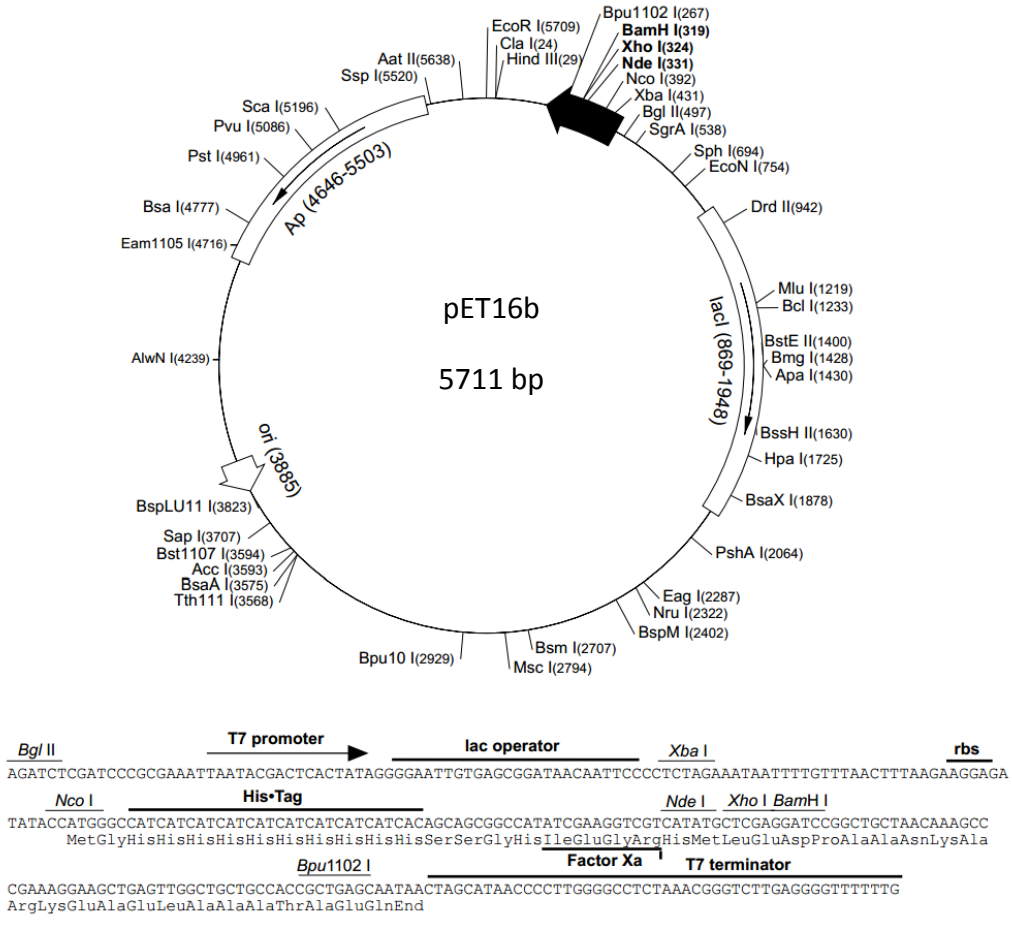


Figure 3.13: pET16b vector map

The gene was inserted between NdeI and BamHI (Figure 3.14), with the reverse primer designed to encode a C-terminal Gly-Cys motif followed by a stop codon.

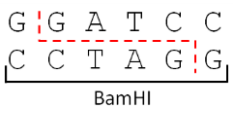


Figure 3.14: BamHI recognition sequence and cleavage position

pET16b cloning site:

```
5′  His10 tag        Factor Xa                                   3′
…   …   …   ATC GAA GGT CGT CAT ATG CTC GAG GAT CCG …
…   …   …   TAG CTT CCA GCA GTA TAC GAG CTC CTA GGC …
3′          Ile Glu Gly Arg   NdeI          BamHI   5′
```

pET16b cloning site after treatment with NdeI/BamHI:

```
5′  His10 tag        Factor Xa                          3′
…   …   …   ATC GAA GGT CGT CA        GAT CCG …
…   …   …   TAG CTT CCA GCA GTA T         GC …
3′          Ile Glu Gly Arg                    5′
```

PCR amplified gene corresponding to IFNβ 1-77:

```
5′                IFNβ 1-77                    3′
…   CAT ATG AGC … GGC TGT TAA GGA TCC …
…   GTA TAC TCG … CCG ACA ATT CCT AGG …
3′    NdeI        Gly Cys Stop  BamHI   5′
```

PCR amplified gene corresponding to IFNβ 1-77 after treatment with NdeI/BamHI:

```
5′              IFNβ 1-77                  3′
…   T ATG AGC … GGC TGT TAA G              …
…     AC TCG … CCG ACA ATT CCT AG          …
3′                Gly Cys Stop             5′
```

DNA ligation product:

```
5′  His10 tag        Factor Xa                    IFNβ 1-77                          3′
…   …   …   ATC GAA GGT CGT CAT ATG AGC … GGC TGT TAA GGA TCC …
…   …   …   TAG CTT CCA GCA GTA TAC TCG … CCG ACA ATT CCT AGG …
3′          Ile Glu Gly Arg   NdeI          Gly Cys Stop  BamHI   5′
```

Figure 3.15: NdeI/BamHI cloning strategy of IFNβ 1-77 gene into vector pET16b

3.2.5 Expression of IFNβ 1-77 with C-terminal Gly-Cys

After sequencing, the pET16b plasmid encoding IFNβ 1-77 was transformed into BL21(DE3) cells for protein expression. The His_{10} tagged protein was found to be insoluble, but was nonetheless purified under denaturing conditions (Figure 3.16) and identified by LC-MS (Figure 3.17).

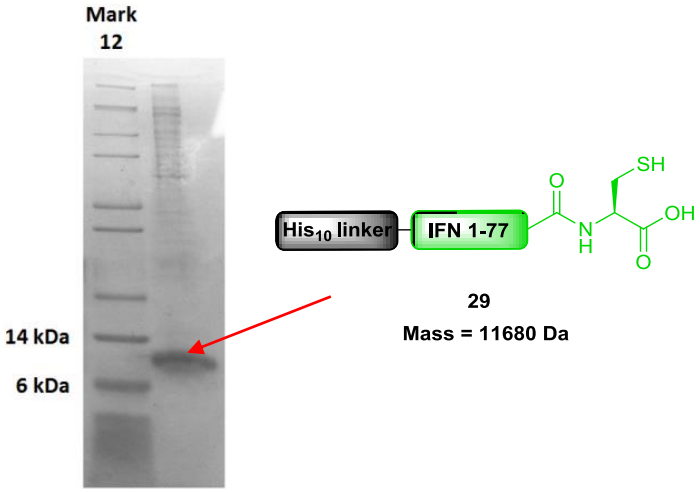


Figure 3.16: SDS-PAGE of His_{10} tagged IFNβ 1-77 after purification

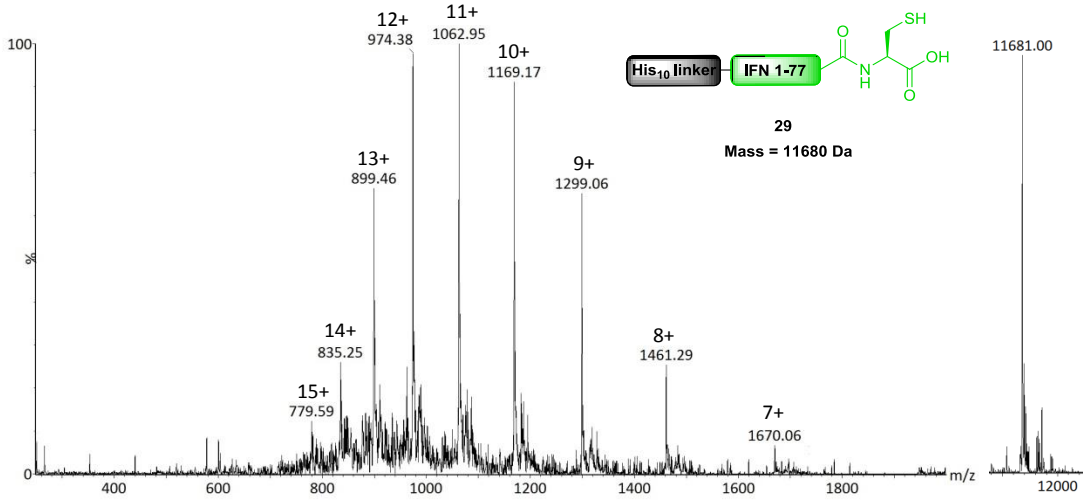
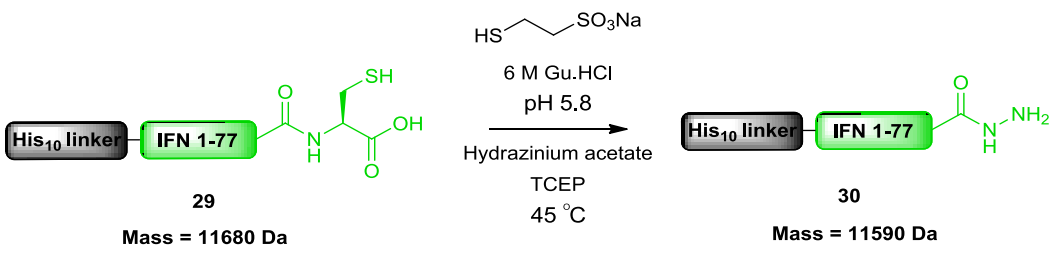


Figure 3.17: ES+ mass spectrum and deconvoluted mass of His_{10} tagged IFNβ 1-77

3.2.6 C-terminal chemical functionalisation of IFNβ 1-77 via *N→S* acyl transfer

From previous experience functionalising the C-terminus of insoluble proteins (Section 2.3.7), it was decided to apply the hydrazide forming reaction to His_{10} tagged IFNβ 1-77 (Scheme 3.1). The reaction was followed by LC-MS over 72 hours and formation of the hydrazide **30** was observed (Figure 3.18).



Scheme 3.1: Hydrazinolysis of His_{10} tagged IFNβ 1-77 via *N→S* acyl transfer

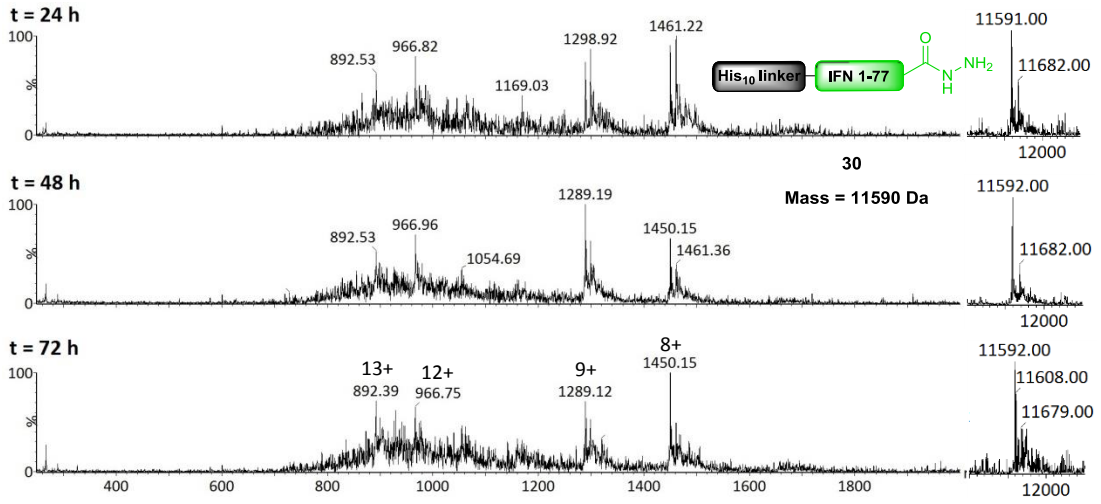


Figure 3.18: ES+ mass spectrum analysis of hydrazinolysis of His_{10} tagged IFNβ 1-77 over 72 hours

Although hydrazide formation occurred, there was an evident loss of quality in the mass spectrum signal over time. Consistent with previous observations, it also appeared that the rate of reaction decreased after the initial 48 hours of the reaction. The decreased reaction rate was attributed to the oxidation of the protein and reagents, where resulting disulfide formation blocks *N→S* acyl transfer.

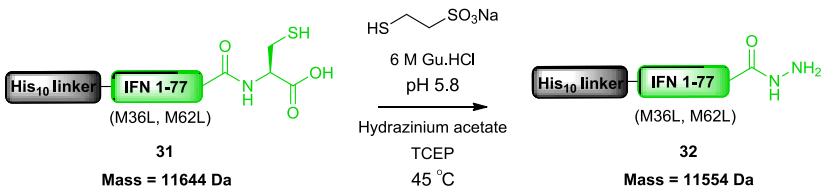
Due to the protein's insolubility, removal of the His_{10} tag would be ineffective with Factor Xa and hence the possibility of CNBr cleavage was considered. As this strategy necessitates the removal of any native methionine residues, it was also reasoned that this could reduce the formation of an additional species corresponding to mass difference of +16, which was attributed to methionine oxidation.

GHHHHHHHHH HSSGHIEGRH **MSYNLLGFLQ RSSNFQCQKL LWQLNGRLEY**
CLKDRMNFDI PEEIKQLQQF QKEDAALTIY EMLQNIFAIF RQDSSSGC

GHHHHHHHHH HSSGHIEGRH **MSYNLLGFLQ RSSNFQCQKL LWQLNGRLEY**
CLKDRLNFDI PEEIKQLQQF QKEDAALTIY ELLQNIFAIF RQDSSSGC

Figure 3.19: Sequence analysis of His_{10} tagged IFNβ 1-77 highlighting mutations for CNBr strategy

The internal methionine residues were mutated to leucine through site directed mutagenesis and the protein was expressed as before. After purification, the protein was subjected to hydrazinolysis conditions and checked by LC-MS after 48 hours (Figure 3.20).



Scheme 3.2: Hydrazinolysis of His_{10} tagged IFNβ 1-77 (CNBr mutant) via *N→S* acyl transfer

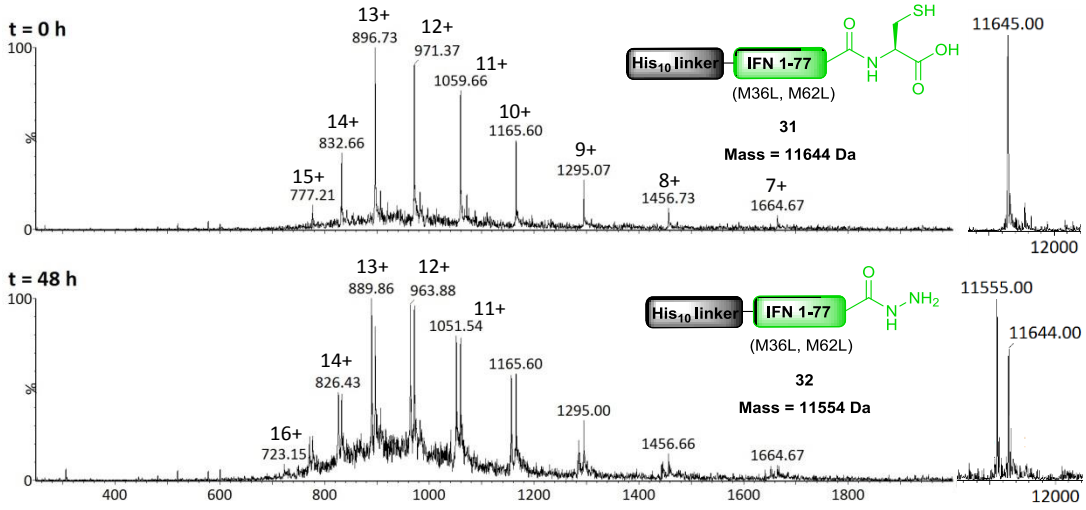


Figure 3.20: ES+ mass spectrum analysis of hydrazinolysis of His_{10} tagged IFNβ 1-77 (CNBr mutant) over 48 h

Encouragingly, conversion to the hydrazide **32** appeared to have proceeded by approximately 60% after 48 hours, with less evidence of competing side reactions. According to the previous rationale that reaction progress was slowed by oxidation of MESNa and the protein, the use of fresh reagents was considered to enable further hydrazinolysis. The sample was exhaustively buffer-exchanged to ensure maximum removal of expended reagents and then resubjected to the hydrazide forming conditions.

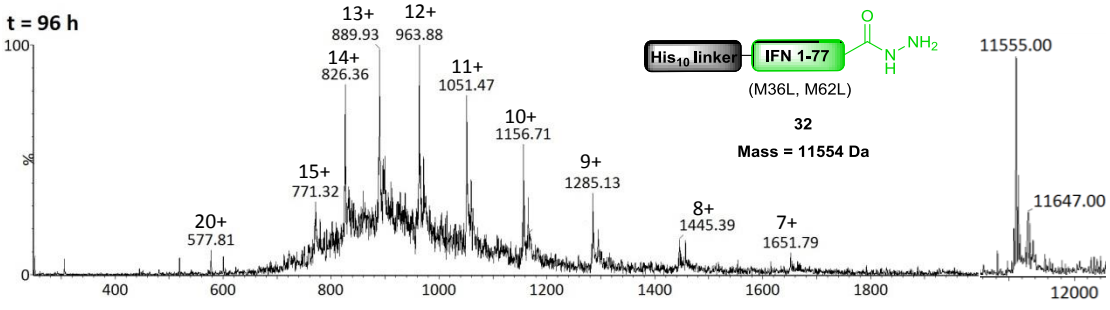


Figure 3.21: ES+ mass spectrum analysis of hydrazinolysis of His_{10} tagged IFNβ 1-77 (CNBr mutant) after 96 h

LC-MS analysis revealed that further hydrazinolysis appeared to take place, although a small amount of inseparable starting material remained. However, it was reasoned that any residual cysteinyl protein would not interfere with forthcoming ligation steps. The His_{10} tag was retained to aid subsequent purification steps.

3.2.7 Cloning and expression of IFNβ 94-166 allowing formation of N-terminal cysteine

The gene corresponding to IFNβ 94-166 was also cloned into pET16b to allow expression with a His_{10} tag. As pET16b encodes a Factor Xa cleavage site preceding a methionine residue, an additional IEGR recognition sequence was introduced before the desired N-terminal cysteine during PCR.

pET16b cloning site:

5′	His_{10} tag		Factor Xa									3′
…	…	…	…	…	CAT	ATG	CTC	GAG	GAT	CCG	…	
…	…	…	…	…	GTA	TAC	GAG	CTC	CTA	GGC	…	
3′						NdeI			BamHI			5′

pET16b cloning site after treatment with NdeI/BamHI:

5′	His_{10} tag		Factor Xa						3′
…	…	…	…	…	CA		GAT	CCG	…
…	…	…	…	…	GTA	T		GC	…
3′									5′

PCR amplified gene corresponding to IFNβ 94-166:

5′			Additional Factor Xa site				IFNβ 94-166							3′
…	CAT	ATG	**ATC**	**GAA**	**GGT**	**CGT**	**TGT**	…	**AAC**	TAA	GGA	TCC	…	
…	GTA	TAC	**TAG**	**CTT**	**CCA**	**GCA**	**ACA**	…	**TTG**	ATT	CCT	AGG	…	
3′		NdeI	**Ile**	**Glu**	**Gly**	**Arg**	**Cys**			Stop		BamHI		5′

PCR amplified gene corresponding to IFNβ 94-166 after treatment with NdeI/BamHI:

5′			Additional Factor Xa site				IFNβ 94-166						3′
…	T	ATG	**ATC**	**GAA**	**GGT**	**CGT**	**TGT**	…	**AAC**	TAA	G		…	
…		AC	**TAG**	**CTT**	**CCA**	**GCA**	**ACA**	…	**TTG**	ATT	CCT	AG	…	
3′			**Ile**	**Glu**	**Gly**	**Arg**	**Cys**			Stop			5′

DNA ligation product:

5′	His_{10} tag		Factor Xa				Additional Factor Xa site				IFNβ 94-166							3′
…	…	…	…	…	CAT	ATG	**ATC**	**GAA**	**GGT**	**CGT**	**TGT**	…	**AAC**	TAA	GGA	TCC	…	
…	…	…	…	…	GTA	TAC	**TAG**	**CTT**	**CCA**	**GCA**	**ACA**	…	**TTG**	ATT	CCT	AGG	…	
3′						NdeI	**Ile**	**Glu**	**Gly**	**Arg**	**Cys**			Stop		BamHI		5′

Figure 3.22: NdeI/BamHI cloning strategy of IFNβ 94-166 gene into vector pET16b

The protein was expressed in BL21(DE3) cells and SDS-PAGE analysis revealed that His_{10} tagged IFNβ 94-166 was also insoluble. In an attempt to solubilise this fragment, the gene was subcloned into pET32 Xa/LIC in the hope that an N-terminal thioredoxin fusion would be soluble. Disappointingly, after expression at 30 °C it appeared that the thioredoxin fusion **33** was also insoluble (Figure 3.23).

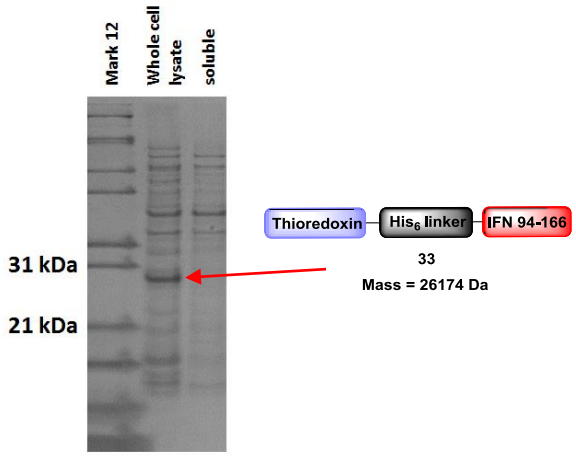


Figure 3.23: SDS-PAGE of thioredoxin His_6 tagged IFNβ 94-166 after expression

With the fusion protein insoluble and unsuitable for Factor Xa treatment, the pET16 plasmid was redesigned to produce a protein suitable for CNBr cleavage. The methionine residues were mutated to leucine and a methionine substitution was introduced preceding the desired N-terminal cysteine (Figure 3.24) through site directed mutagenesis.

GHHHHHHHHH HSSGHIEGRH MIEGR**CINHL KTVLEEKLEK EDFTRGKLMS**
SLHLKRYYGR ILHYLKAKEY SHCAWTIVRV EILRNFYFIN RLTGYLRN

GHHHHHHHHH HSSGHIEGRH LIEG**MCINHL KTVLEEKLEK EDFTRGKLLS**
SLHLKRYYGR ILHYLKAKEY SHCAWTIVRV EILRNFYFIN RLTGYLRN

Figure 3.24: Sequence analysis of His_{10} tagged IFNβ 94-166 highlighting mutations for CNBr strategy

The modified construct was used to express the His_{10} tagged protein, which was purified by Ni^{2+} affinity chromatography under denaturing conditions as before. LC-MS confirmed successful expression of His_{10} tagged IFNβ 94-166 (Figure 3.25).

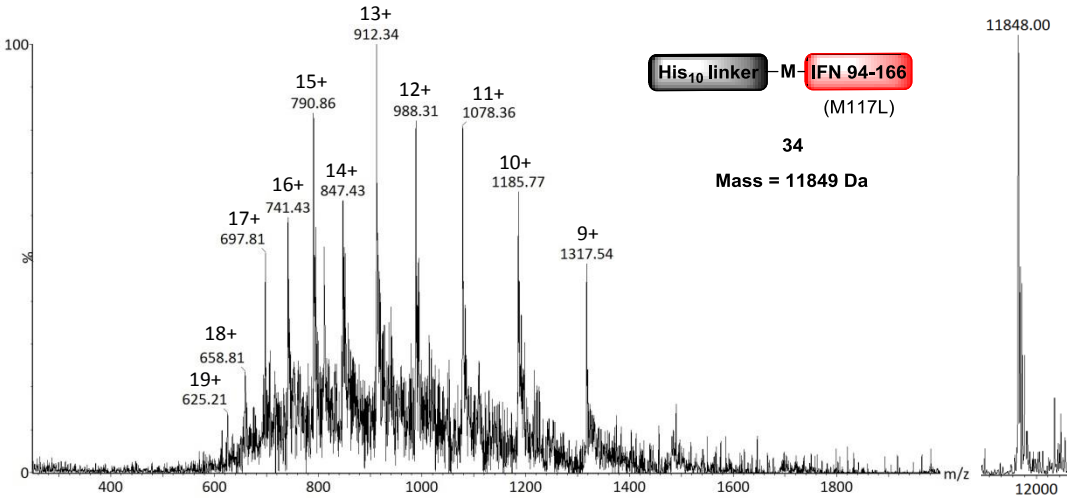
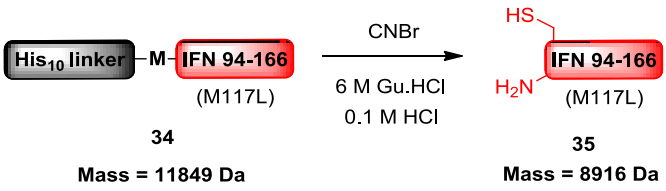


Figure 3.25: ES+ mass spectrum and deconvoluted mass of His_{10} tagged IFNβ 94-166 (CNBr mutant)

3.2.8 N-terminal chemical functionalisation of IFNβ 94-166 via CNBr treatment

CNBr treatment of the protein was performed according to the protocol reported by Richardson and Macmillan.[215] Although the use of 80% formic acid enabled the complete solubilisation of the fusion protein, LC-MS disappointingly indicated deterioration of the sample after treatment with CNBr overnight. After a number of unsuccessful attempts, an alternative protocol for CNBr cleavage reported by Andreev *et. al* was investigated.[216] This method relies on solubilisation of proteins in 6 M Gu.HCl, with 0.1 M HCl to prevent reactions at basic residues (Scheme 3.3). Pleasingly, LC-MS of the crude sample after CNBr treatment suggested quantitative conversion to the desired product **35** (Figure 3.26). The crude mixture was subjected to another round of Ni^{2+} affinity chromatography, enabling the separation of the product from any His_{10} tagged proteins.



Scheme 3.3: CNBr treatment of His_{10} tagged IFNβ 94-166

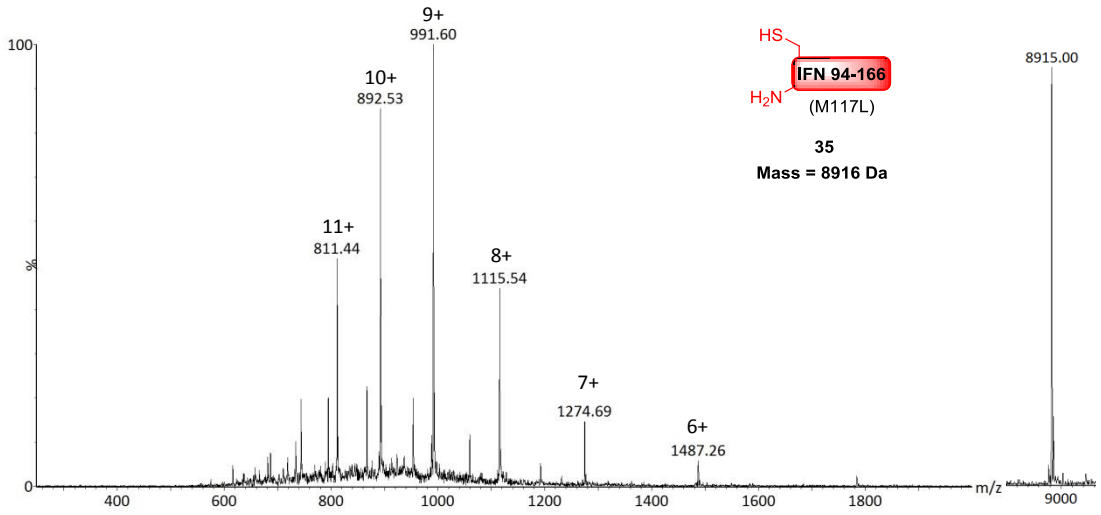


Figure 3.26: ES+ mass spectrum and deconvoluted mass of IFNβ 94-166 (CNBr mutant)

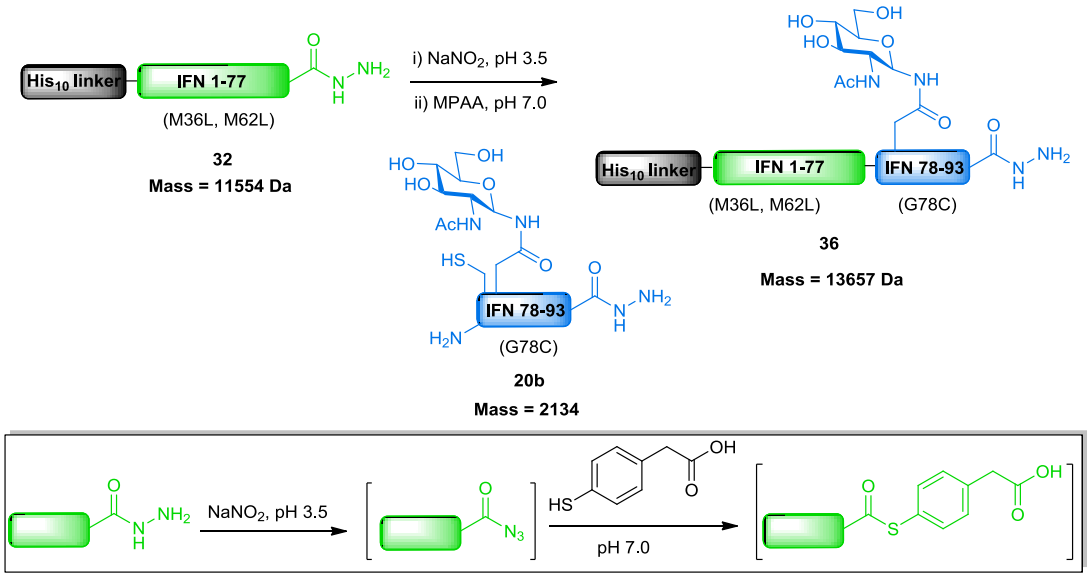
3.3 Conclusions

A number of methods have been explored towards the functionalisation of IFNβ fragments to enable a semi-synthetic strategy. Unfortunately, the highly hydrophobic sequences in both recombinant segments have rendered some of these strategies incompatible. Consequently, novel and existing chemoselective reactions have been applied to achieve the desired protein modifications. The novel application of *N→S* acyl transfer towards the C-terminal functionalisation of recombinant protein hydrazides was demonstrated. Importantly, this methodology represents the first demonstration of C-terminal hydrazinolysis to an insoluble recombinant protein. This may expand the applicability of semi-synthetic strategies towards particularly hydrophobic sequences, where intein and sortase technologies are largely unsuitable. As the desired protein fragments were armed with the appropriate functional groups, investigations were able to proceed towards protein assembly via NCL.

4 Application of Native Chemical Ligation towards the assembly of semi-synthetic IFNβ

4.1 NCL of His_{10} tagged IFNβ 1-77 hydrazide with 78-93 glycopeptide

In the first approach to assemble semi-synthetic IFNβ, a ligation between the N-terminal fragment hydrazide was conducted with the glycopeptide segment containing a free N-terminal cysteine and a C-terminal hydrazide (Scheme 4.1).



Scheme 4.1: NCL of His_{10} tagged IFNβ 1-77 hydrazide with 78-93 glycopeptide hydrazide

It was envisaged that this strategy would allow the resulting product to be used in the next ligation without the need for an additional deprotection step. The ligation was attempted following the protocol reported by Liu,[135] where the treatment of the N-terminal hydrazide with sodium nitrite in 6 M Gu.HCl, 0.2 M Na Phosphate (pH 3.0-4.0) results in the formation of the acyl azide. After the oxidation step, the pH was raised to 7.0 by addition of neutral ligation buffer containing mercaptophenyl acetic acid and the glycopeptide hydrazide. This step allows the thioester to be formed *in situ,* with subsequent participation in the ligation reaction.

The reaction was allowed to proceed overnight, followed by concentration in a molecular weight cut off device and was subsequently reduced with TCEP. Discouragingly, LC-MS analysis of the crude reaction mixture showed an abundance of the hydrolysed N-terminal fragment, with little accumulation of the ligated product (Figure 4.1). Purification of the ligated product **36** was attempted by (RP)HPLC, but isolation of the semi-synthetic glycoprotein was not successful.

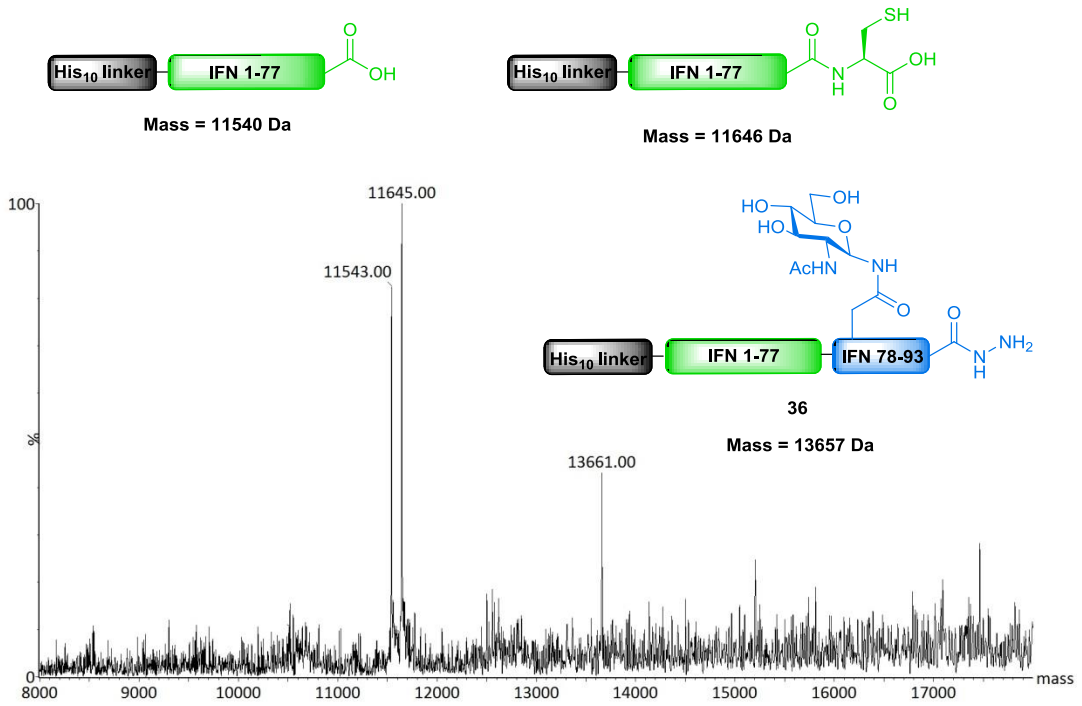
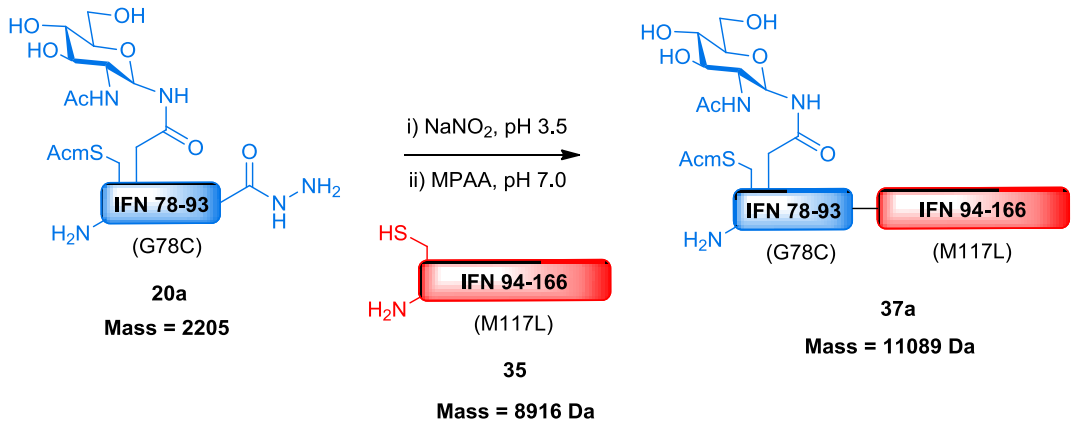


Figure 4.1: ES+ deconvoluted mass spectrum of attempted His_{10} IFNβ 1-93 formation. Ligated product was detected, although the major species correspond to cysteinyl protein **31** and hydrolysis of the hydrazide.

4.2 NCL of IFNβ 78-93 glycopeptide hydrazide with 94-166

In an alternate strategy towards semi-synthetic IFNβ, the ligation between the C-terminal fragment and the Acm protected glycopeptide hydrazide (Scheme 4.2) was performed. The ligation was carried out as before and LC-MS analysis revealed that complete consumption of the C-terminal fragment **35** took place within 3 hours. The semi-synthetic glycoprotein **37a** was subsequently purified by (RP)HPLC.



Scheme 4.2: NCL of IFNβ 78-93 glycopeptide hydrazide with 94-166 (CNBr mutant)

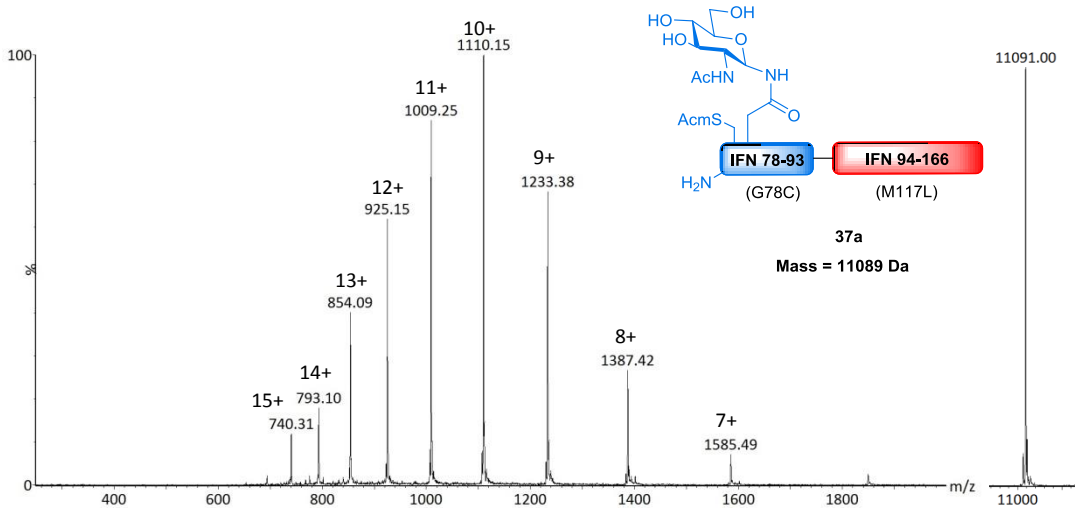


Figure 4.2: ES+ mass spectrum and deconvoluted mass of IFNβ 78-166 glycopeptide (Acm protected)

4.2.1 Acm deprotection of IFNβ 78-166 glycopeptide

In preparation for the next ligation, the N-terminal cysteine was unmasked by Ag^{+} mediated Acm deprotection according to a procedure reported by Danishefsky (performed by Dr Derek Macmillan).[217] After reduction with DTT, the protein **37b** was purified by (RP)HPLC and confirmed to be the product by LC-MS (Figure 4.3).

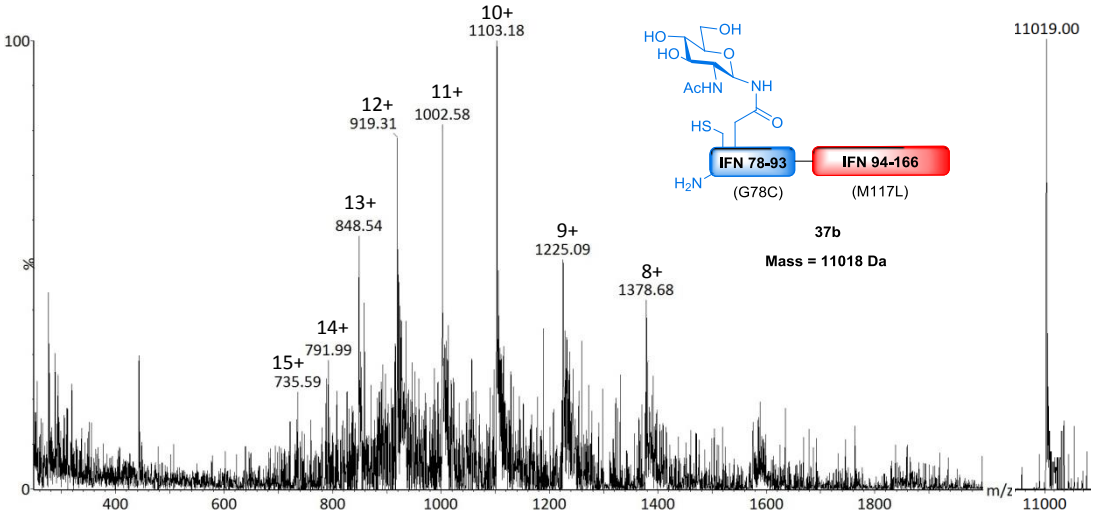
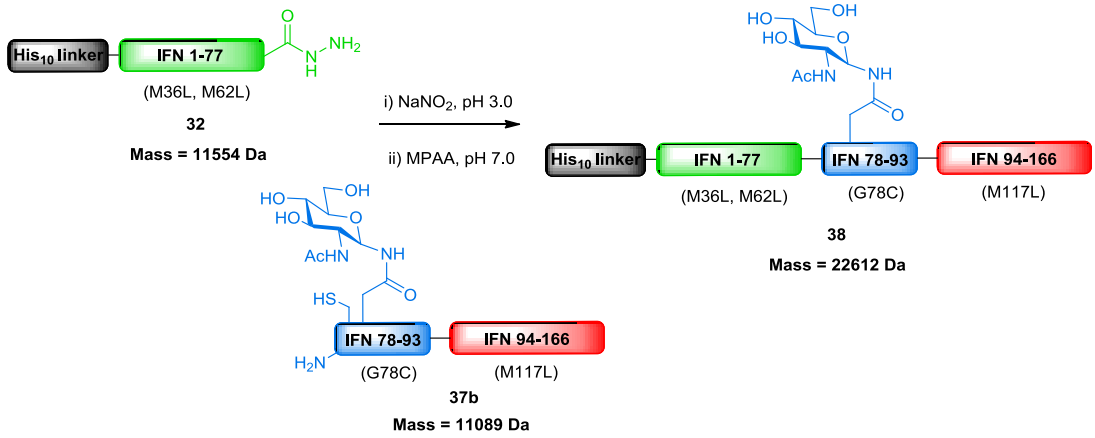


Figure 4.3: ES+ mass spectrum and deconvoluted mass of IFNβ 78-166 glycopeptide

4.3 NCL of His_{10} tagged IFNβ 1-77 hydrazide with 78-166 glycopeptide

The final ligation was attempted, with initial diazotisation of the N-terminal hydrazide as before. Addition of MPAA and the C-terminal fragment was followed by adjustment of the final pH to 7.0 and the ligation was allowed to proceed overnight (performed by Dr Ben Cowper).



Scheme 4.3: NCL of His_{10} tagged IFNβ 1-77 hydrazide with semi-synthetic glycopeptide 94-166

After analysis by LC-MS no trace of the expected product was found, but instead the major species was found to correspond to the N-terminal fragment MPAA thioester (Figure 4.4). Purification was attempted by (RP)HPLC, but disappointingly attempts to isolate the product **38** were unsuccessful. The failed ligation was attributed to an insufficient concentration of the C-terminal fragment.

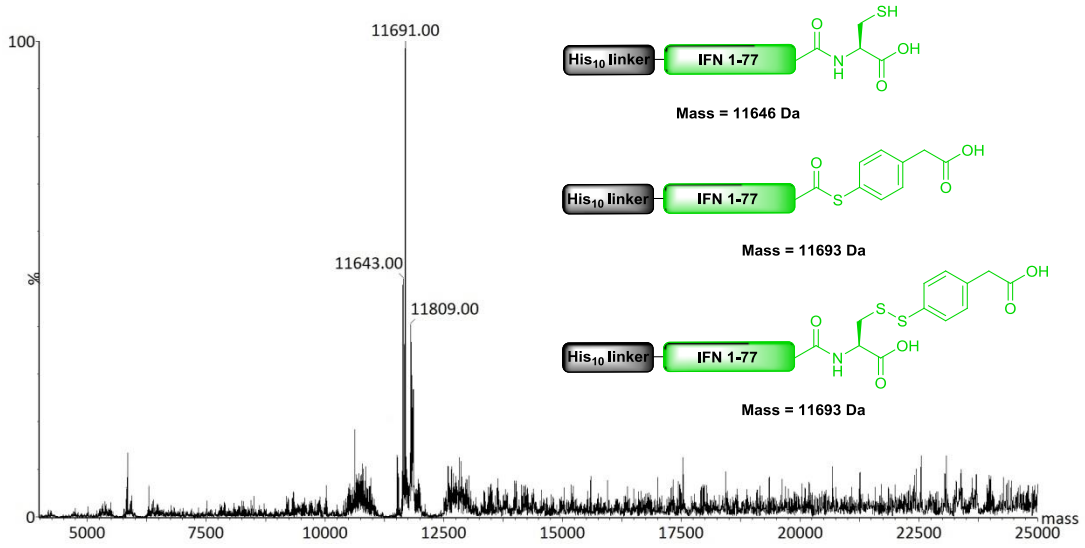


Figure 4.4: Deconvoluted mass spectrum analysing attempted NCL of His_{10} tagged IFNβ 1-77 hydrazide with semi-synthetic glycopeptide

4.4 Conclusions

Unfortunately, the first attempt to assemble semi-synthetic IFNβ by Native Chemical Ligation was ineffective. Nonetheless, the foundations have been laid to enable this goal to be achieved in the near future. For these studies, it is critical that large amounts of material are accessible. The low expression efficiency of IFNβ fragments in *E. coli* may be overcome by the use of a fermenter, which will allow protein expression to be conducted with much larger culture volumes.

Most importantly this study has shown that the necessary fragments were accessible, overcoming the limitations of existing methodology. Hence, this strategy should still enable a universal application to virtually any *N*-glycoprotein of interest.

5 Enzymatic transfer of oligosaccharides to *N*-glycopeptides

5.1 Introduction

In the early 1970's, a family of enzymes were identified with the capability of selectively cleaving the GlcNAcβ(1-4)GlcNAc chitobiose core of *N*-glycans.[218-220] *endo*-β-*N*-acetyl-glucosaminidases were initially exploited for truncating *N*-glycoproteins and for the release of glycans for analysis.[221] However, it was later discovered that the enzyme catalysed cleavage could also be accompanied by a transglycosylation reaction, wherein the released glycan is transferred to another substrate.[222-224] The ability to transfer large oligosaccharides *en bloc* in a single step with remarkable regio- and stereoselectivity, has led researchers to investigate the use of endoglycosidases in the synthesis of *N*-glycoproteins (Figure 5.1). In particular, Endo M from *Mucor hiemalis* and Endo A from *Arthrobacter protophormiae* have been exploited in such endeavours, due to their synthetic transglycosylation efficiency. Endo M recognises both complex-type and high mannose *N*-glycans as native substrates, whereas Endo A is specific for the high mannose subtype.

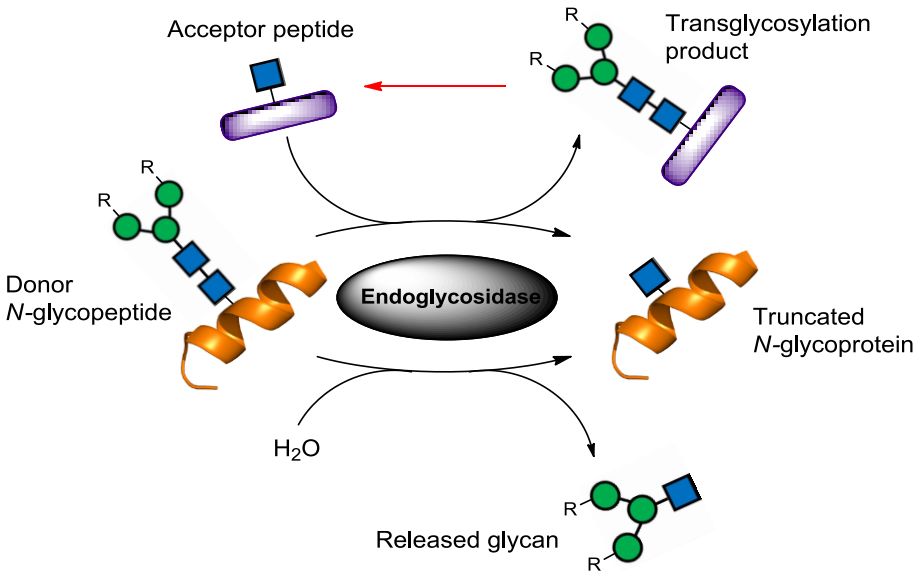
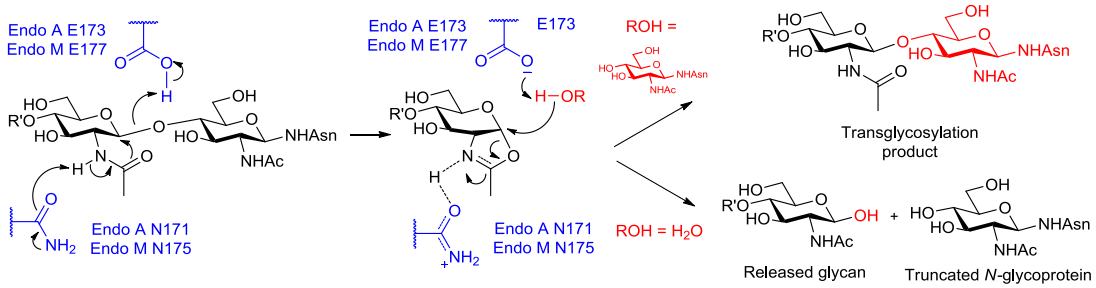


Figure 5.1: Endoglycosidase catalysed reactions

Both Endo A and Endo M fall into the GH85 subfamily of glycosyl hydrolases, and are understood to operate via a two-step acid/base catalysed reaction involving a glutamic acid and an asparagine residue within the active site (Scheme 5.1). The mechanism of transglycosylation is presumed to take place through an oxazoline or oxazolinium intermediate, which is formed by neighbouring group participation from the acetamide of the second GlcNAc residue.

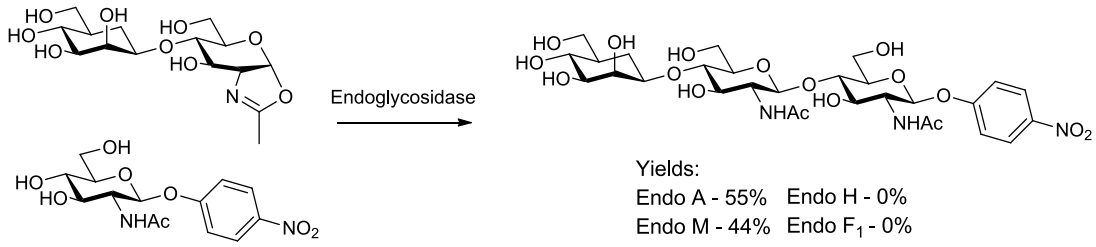


Scheme 5.1: Proposed mechanism for GH85 endoglycosidase mediated hydrolysis and transglycosylation

A complication encountered with endoglycosidase mediated synthesis is the natural hydrolytic activity of these enzymes, which often leads to poor overall yields as the transglycosylated product is also recognised as a substrate.[225] Furthermore, transglycosylation reactions with un-activated donors (in the form of *N*-glycopeptides) are generally inefficient and require high concentrations of the acceptor to be effective.[225, 226] In an attempt to favour the synthetic reaction, researchers have sought methodologies which overcome the low transglycosylation efficiency of these enzymes, whilst minimising product hydrolysis. Consequently, two major routes have been explored, utilising activated glycosyl donors in the form of oxazolines and the evaluation of mutant enzymes.

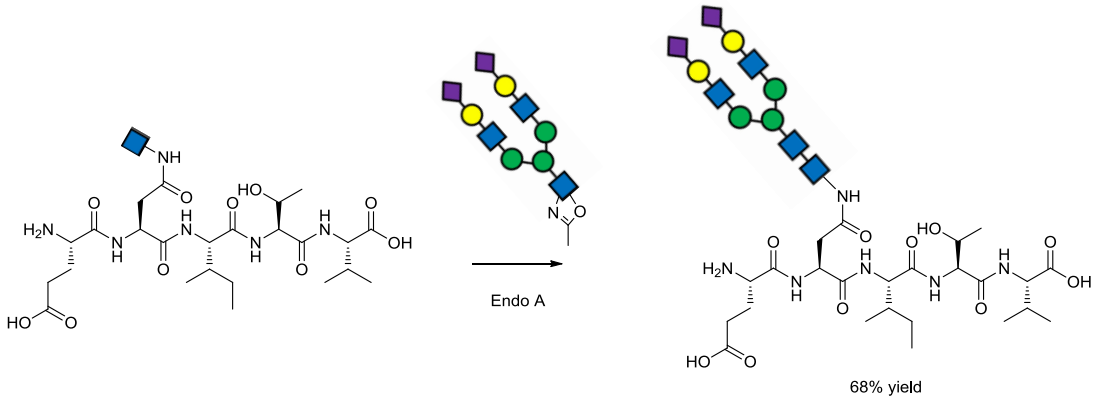
5.1.1 Oxazolines as activated glycosyl donors for "transglycosylation" reactions

With the assumption that transglycosylation mediated reactions proceed via an oxazolinium intermediate, Shoda surveyed the ability of various endoglycosidases to process a disaccharide oxazoline as an activated substrate for glycosylation.[43] Interestingly, transfer to a GlcNAc acceptor (Scheme 5.2) was observed with Endo A and Endo M, with the desired regio- and stereoselectivity confirmed by NMR. This prompted the investigation of a number of other glycosyl oxazolines as substrates for endoglycosidase mediated synthesis.[45, 227-231]



Scheme 5.2: Shoda's investigation of a disaccharide oxazoline as a substrate for transfer

Importantly, these studies have revealed that endoglycosidases are able to transfer a wide range of glycosyl oxazolines, regardless of their native substrate specificity. Accordingly, application of this methodology can result in reduced product hydrolysis, as the transglycosylated product is not recognised by the enzyme. A noteworthy example of this was demonstrated by Wang, where it was shown that Endo A could transfer a complex sialoglycan to a GlcNAc-pentapeptide without product hydrolysis (Scheme 5.3).[231]

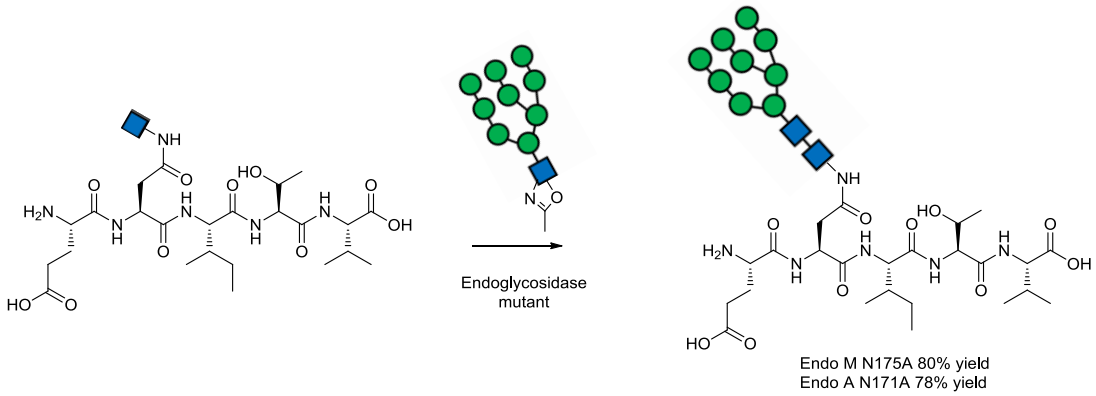


Scheme 5.3: Transfer of a complex type glycan oxazoline with Endo A

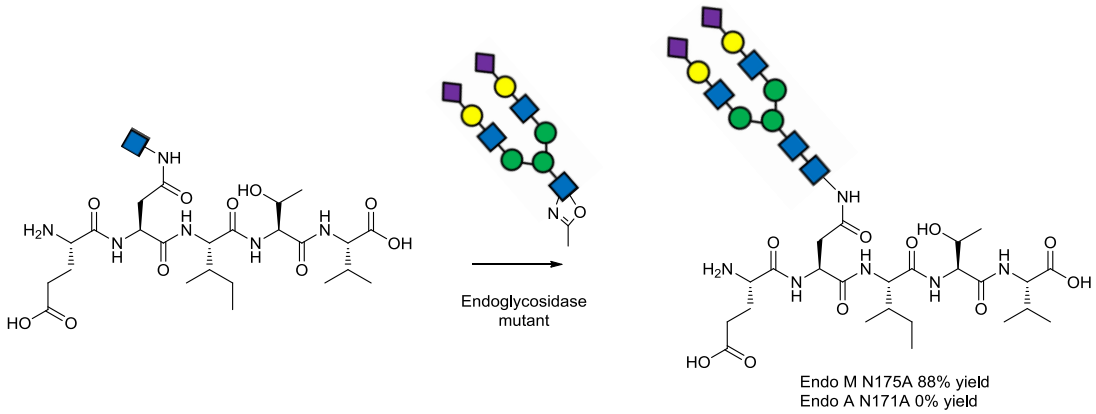
5.1.2 Endoglycosidase mutants

In an alternative approach to abolish the hydrolytic activity of these enzymes, several research groups have examined endoglycosidase mutants which are able to transglycosylate sugar oxazolines but are unable to hydrolyse the chitobiose core. These investigations have also yielded mutants that have improved transglycosylation efficiency compared to the wild type. A common starting point for these studies involves the mutagenesis of the critical residues involved in endoglycosidase mediated hydrolysis.

Wang and Yamamoto evaluated the mutation of asparagine at residue 175 in Endo M, which was understood to be critical to the hydrolysis reaction by assisting formation of the oxazoline intermediate (Scheme 5.1). It was found that an N175A mutation allowed the transfer of a high mannose glycan oxazoline to a GlcNAc-pentapeptide to take place without significant hydrolysis of the product (Scheme 5.4).[232] The enzyme also permitted transfer of a complex type oxazoline with excellent conversion (Scheme 5.5).[231] Encouraged by these findings, Wang subsequently considered the mutation of the analogous residues in Endo A.[230] The N171A mutant was found to transfer high mannose glycan oxazolines without product hydrolysis, but disappointingly was unable to process complex type glycan oxazolines.

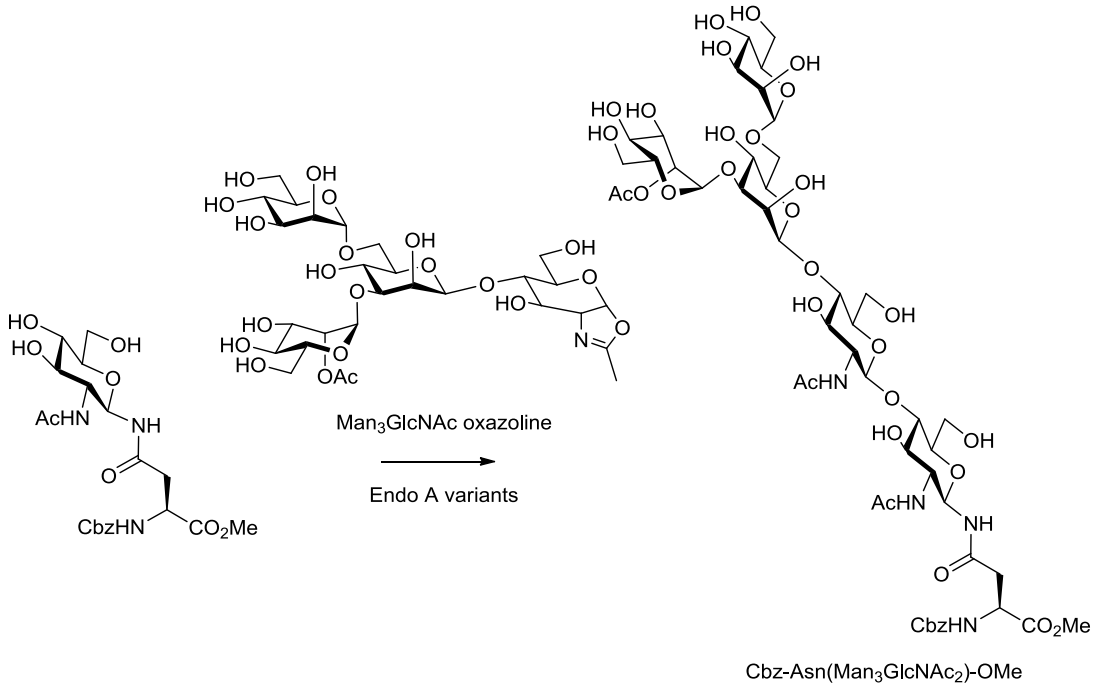


Scheme 5.4: Transfer of high mannose glycan oxazolines with endoglycosidase mutants



Scheme 5.5: Transfer of complex type glycan oxazolines with endoglycosidase mutants

Fairbanks postulated that minimised Endo A hydrolysis could also be attained through removal of a glutamic acid residue from the active site. It was reasoned that the acidic functionality promotes formation of the oxazolinium intermediate by protonation of the outgoing anomeric oxygen in the first step of hydrolysis. The glutamate is then capable of acting as a general base, deprotonating water to promote hydrolysis (Scheme 5.1). Accordingly, an E173Q mutation was introduced under the rationale that non-acidic glutamine would be unable to catalyse the hydrolytic reaction. It was hoped that the synthetic reaction could still be accommodated by hydrogen bonding or deprotonation of the incoming nucleophile. An E173H mutation was also investigated, as histidine can act as a general acid or base. The pK_a of glutamic acid is ~ 4.1, whereas histidine is ~ 6.0 and hence it was supposed that the E173H mutant could potentially exhibit reduced hydrolytic activity.[233] The mutant enzymes were subjected to glycosylation assays with a tetrasaccharide oxazoline donor and Cbz-Asn(GlcNAc)-OMe (Scheme 5.6).



Scheme 5.6: Tranfer of tetrasaccharide oxazoline to Cbz-Asn(GlcNAc)-OMe with Endo A variants

Results showed that whilst oxazolines were processed with the wild type and both mutants, the synthetic reaction proceeded at different rates with varied amounts of product hydrolysis (Figure 5.2). The wild type enzyme rapidly transferred the tetrasaccharide, although was followed by gradual hydrolysis of the product. The E173Q mutant was able to catalyse the reaction, albeit slowly but without evidence of product hydrolysis. Finally, the E173H mutant was capable of processing the tetrasaccharide oxazoline efficiently and although product hydrolysis took place, it was significantly slower compared to the wild type.

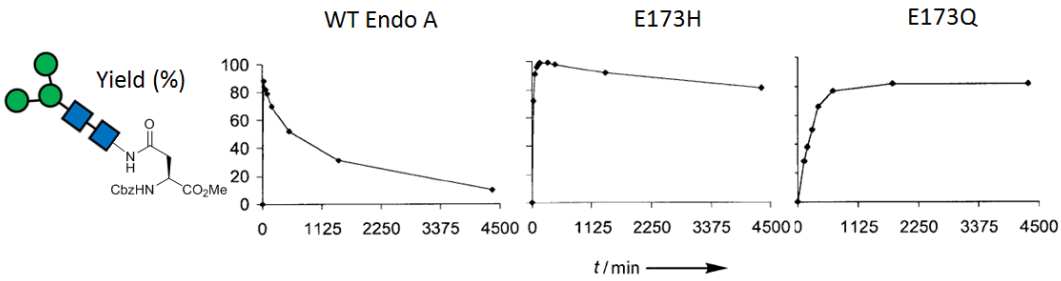
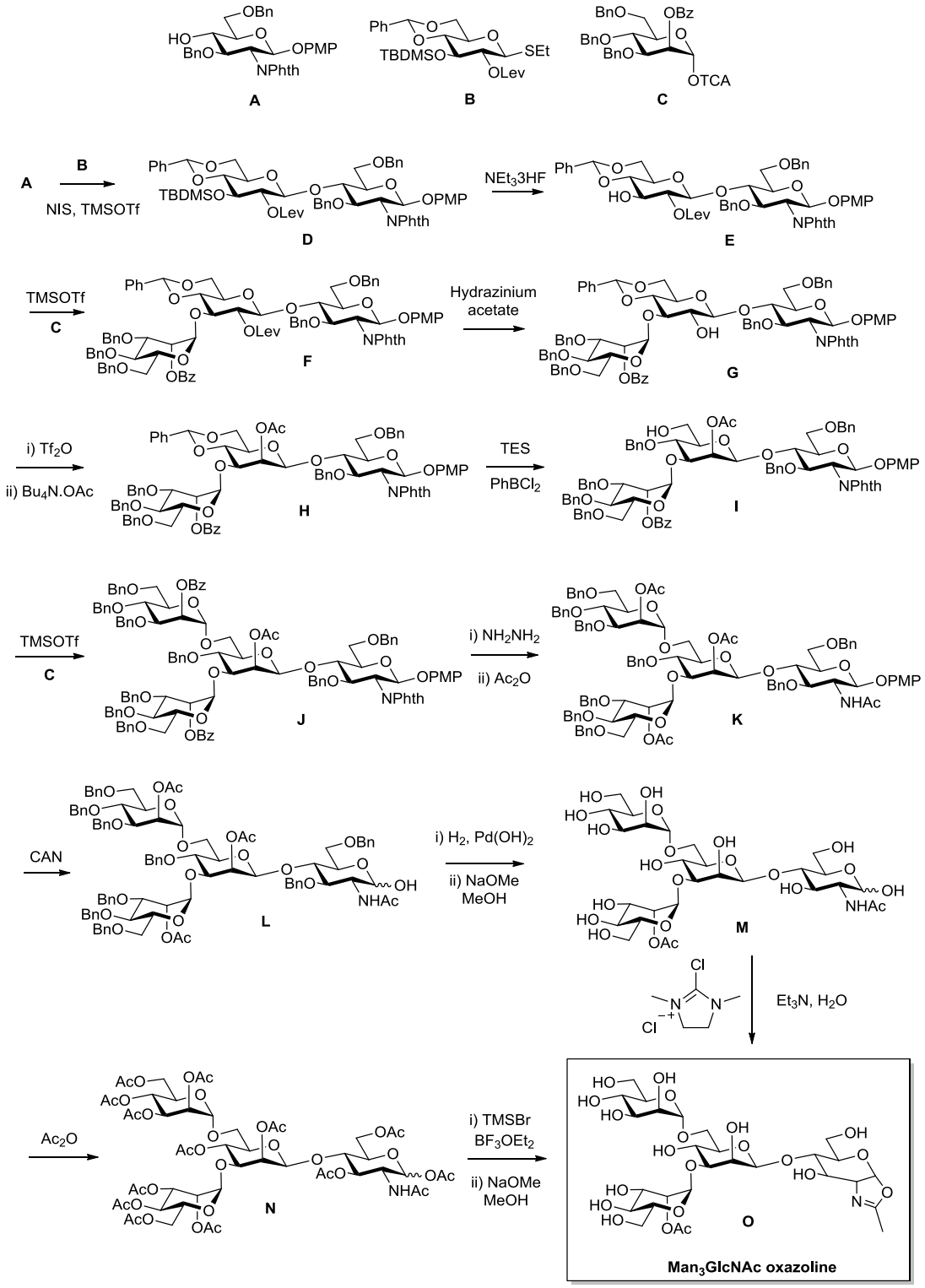


Figure 5.2: Time correlations of product yield for glycosylations of Cbz-Asn(GlcNAc)-OMe with tetrasaccharide oxazoline and Endo A variants

5.2 Synthesis of a tetrasaccharide oxazoline donor

As part of a collaborative investigation, the glycosyl donor was provided by Dextra Laboratories, a specialist in the manufacturer of complex carbohydrates. The oxazoline was synthesised in a convergent manner, initially synthesising suitably protected monosaccharides **A-C** (Scheme 5.7). To begin with, the disaccharide core **D** was synthesised from GlcNAc acceptor **A** and Glc donor **B**. The donor was orthogonally protected as a levulinic acid ester at the 2-position, facilitating selective deprotection and epimerisation to convert glucose to mannose later in the synthesis. This strategy was adopted from a procedure described by Ajisaka and obviates a difficult step in the direct formation of a β-mannoside.[234-236] Selective deprotection of the silyl protecting group was followed by glycosidation with Donor **C** to construct the Manα(1-3)Man linkage. After Lev removal, the free hydroxyl group was converted to the triflate and nucleophilically displaced with tetrabutylammonium acetate, resulting in inversion of configuration at C-2 to provide the desired trisaccharide **H** containing the Manβ(1-4)GlcNAc linked core. The benzylidene acetal was regioselectively opened using triethylsilane and $PhBCl_2$, leaving the 6-hydroxyl group free to act as an acceptor.[237] Glycosidation of **I** with trichloroacetimidate **C** allowed construction of the Manα(1-6)Man linkage to afford **J**. After a series of steps to remove the protecting groups, the fully deprotected tetrasaccharide **M** was converted to the corresponding oxazoline **O** using 2-chloro-1,3-dimethylimidazolinium chloride (DMC).[238] The resulting product was found to contain significant amounts of an inseparable triethylammonium species and subsequently as a precaution, an alternative method was investigated from the unprotected tetrasaccharide. Reacetylation to **N** was followed by treatment with an equimolar mixture of TMSBr and BF_3OEt_2 to form the protected oxazoline[239] and the final deacetylation was achieved by a treatment of sodium methoxide in methanol.



Scheme 5.7: Synthesis of tetrasaccharide oxazoline

5.3 Preparation of Endo A for transfer reactions

5.3.1 Cloning of Endo A E173Q mutant

With the tetrasaccharide oxazoline as the first substrate for glycosylation reactions, interest was drawn to the methodology reported by Fairbanks who had previously investigated a similar transformation.[233] The E173Q mutant was of particular interest, as it was reported that no hydrolysis of the product was observed. Site directed mutagenesis was used to introduce the E173Q substitution into a plasmid containing the Endo A gene in vector pGEX2T (donated by Dr Derek Macmillan). After DNA sequencing verified that the desired mutation was introduced, both the wild type and mutant plasmids were transformed into BL21(DE3) cells.

5.3.2 Expression of Endo A wild Type and E173Q mutant

Expression of the GST (glutathione *S*-transferase) tagged proteins was carried out at 30 °C for 5 hours and the enzymes were purified using glutathione affinity chromatography. SDS-PAGE analysis confirmed that both fusion proteins had been purified successfully although also suggested that a significant amount of protein did not bind to the column (Figure 5.3).

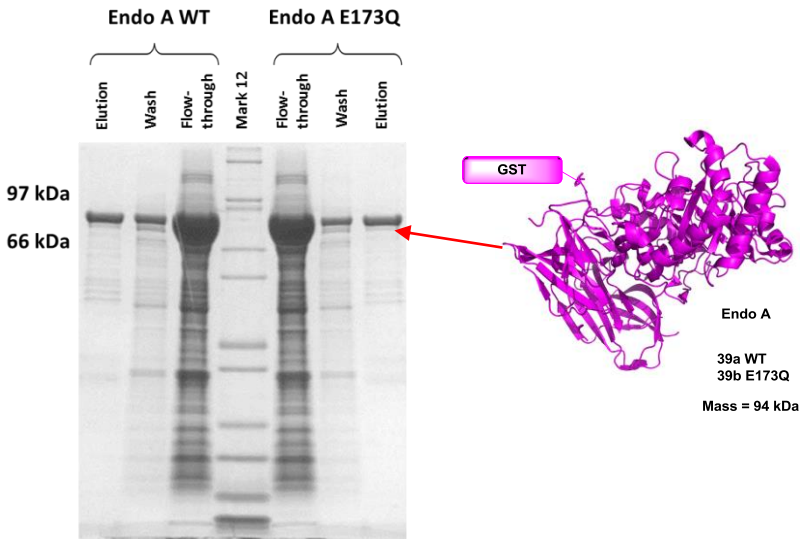


Figure 5.3: SDS-PAGE analysis of column fractions after purification of GST tagged Endo A WT and E173Q

5.3.3 Removal of GST tag

Proteins expressed from vector pGEX2T encode a Thrombin recognition sequence (LVPRGS) between GST and the target protein. After expression and the first round of purification, the Endo A fusion proteins were treated with thrombin. Following protease cleavage, the mixtures were passed down a glutathione column to bind GST and any remaining uncleaved protein. The flow-through was collected and after analysis by SDS-PAGE (Figure 5.4), the purified enzymes were dialysed into transglycosidation assay buffer and concentrated in a molecular weight cut off device.

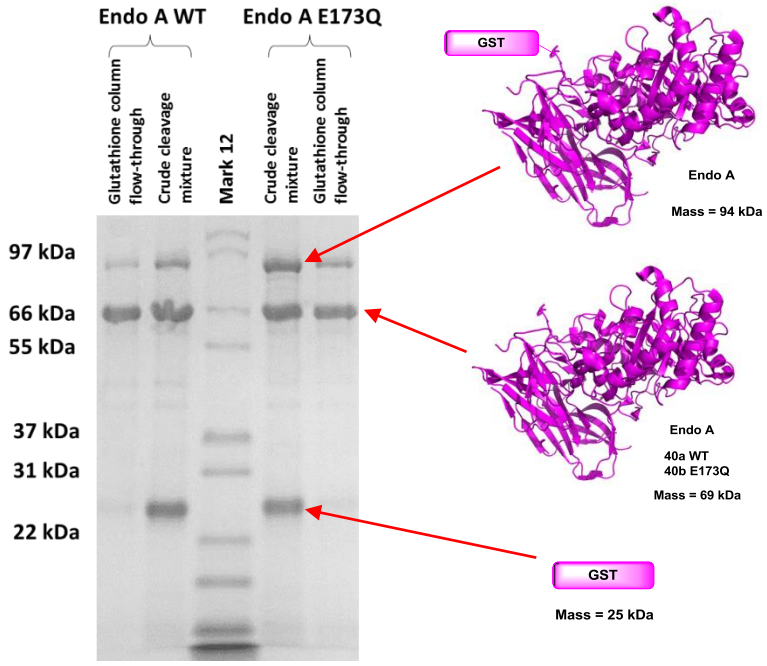
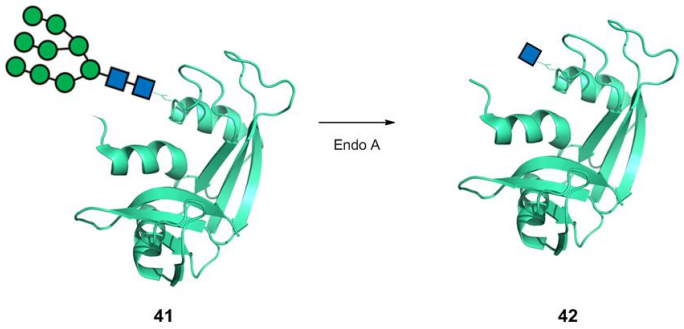


Figure 5.4: Thrombin cleavage of GST-Endo A fusions

5.4 Endo A transfer assays

5.4.1 Endo A activity assays with RNAse B

With the endoglycosidase enzymes in hand, their hydrolytic activity was evaluated against commercially available RNAse B, a high mannose glycoprotein (Scheme 5.8).



Scheme 5.8: Deglycosylation of RNAse B to assess hydrolytic activity of Endo A variants

Endo A wild type, E173Q and their GST variants were tested at different concentrations, followed by SDS-PAGE to assess the extent of deglycosylation (Figure 5.5).

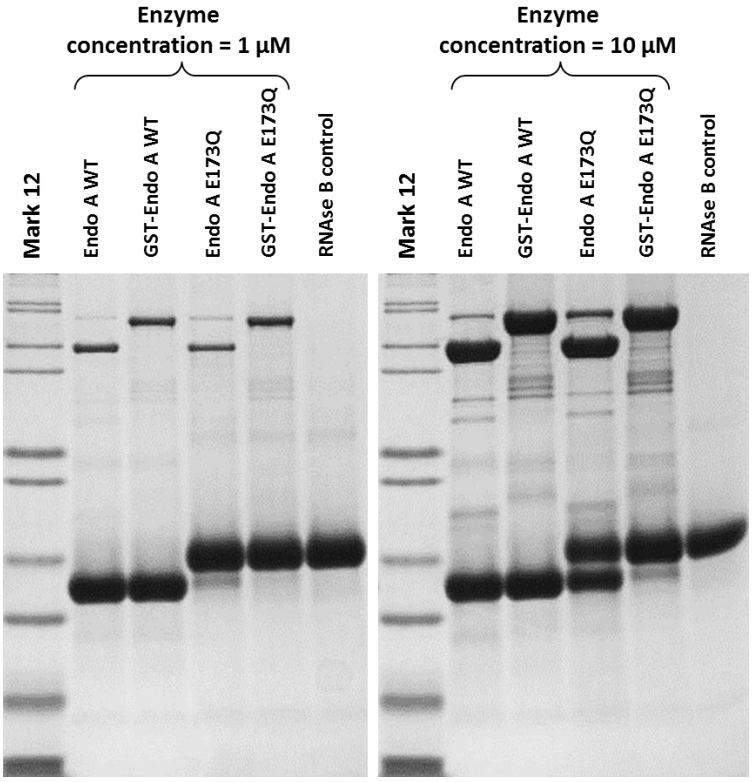


Figure 5.5: SDS-PAGE analysis of deglycosylation assay with RNAse B

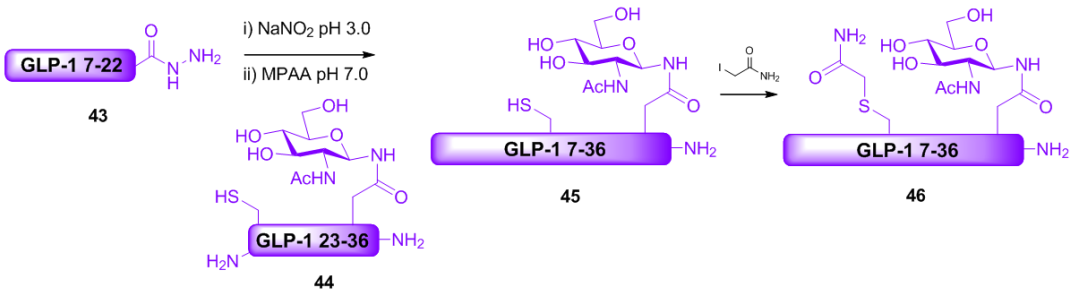
As expected, the wild type Endo A appeared exceptionally capable of hydrolysing the substrates, even at a relatively lower concentration. The GST-tagged protein was also able to promote the deglycosylation reaction, concurrent with previous findings.[240] Unexpectedly, the E173Q mutants also appeared to catalyse hydrolysis of the high mannose glycan to some extent, which contradicted earlier reports.[233, 241]

5.4.2 Model transfer assays

To initially survey the activity of the expressed endoglycosidases, it was decided to investigate the transfer of the tetrasaccharide oxazoline to the glycopeptide corresponding to IFNβ 78-93 (Chapter 2). However, the insolubility of this peptide in glycosylation buffer prompted the investigation of an alternative model system. Glucagon like peptide-1 (GLP-1) 7-36 has recently gained widespread attention as a potential therapeutic for the treatment of type 2 diabetes and affects glucose control by stimulating insulin secretion, suppressing glucagon secretion and lowering appetite. However, it has a relatively short half-life of two minutes as it is rapidly degraded *in vivo* by dipeptidyl peptidase-4.[242] Although GLP-1 is not naturally glycosylated, a recent study showed that the addition of carbohydrate moieties prolonged the half-life of the peptide *in vivo*, thus increasing its potential for therapeutic applications.[243]

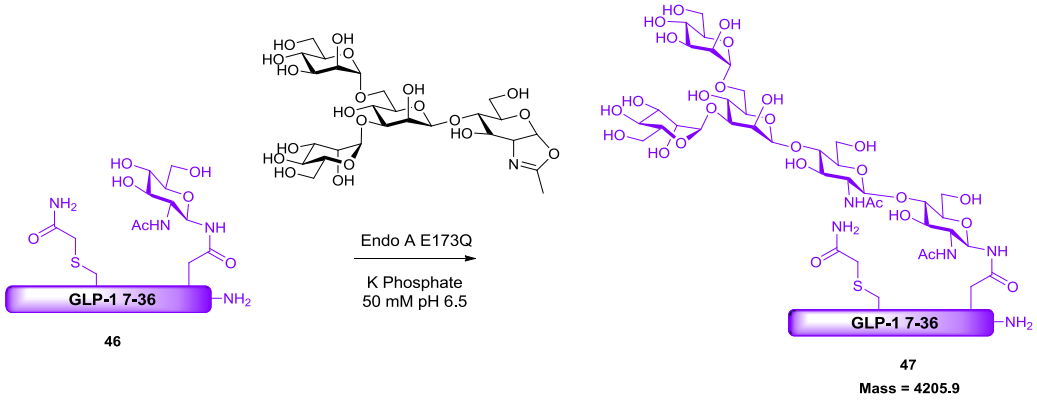
Therefore a study on the endoglycosidase mediated transfer of the $Man_3GlcNAc$ oxazoline was performed with an analogue of GLP-1 7-36, incorporating GlcNAc-Asn in place of lysine at residue 34. This introduction was made in accordance with the previous study by Ueda *et al*. which showed that glycosylation at this position was the most beneficial.[243]

The glycopeptide was synthesised by Native Chemical Ligation from two fragments using Liu's hydrazide activation method.[135] The glutamine residue at position 23 was chosen as the ligation junction and was substituted with cysteine. It was envisaged that chemoselective treatment of the peptide with iodoacetamide after the ligation would convert the cysteine to a glutamine like residue (Scheme 5.9).



Scheme 5.9: Synthesis of a glycosylated GLP-1 analogue

The transfer of the glycan was performed according to a protocol described by Wang, which was also attempted using a tetrasaccharide oxazoline (Scheme 5.10).[45] Initially, the activity of the E173Q mutant was assessed, following the reaction progress by analytical (RP)HPLC (Figure 5.6).



Scheme 5.10: Endo A E173Q mediated transfer to GLP-1 with $Man_3GlcNAc$ oxazoline

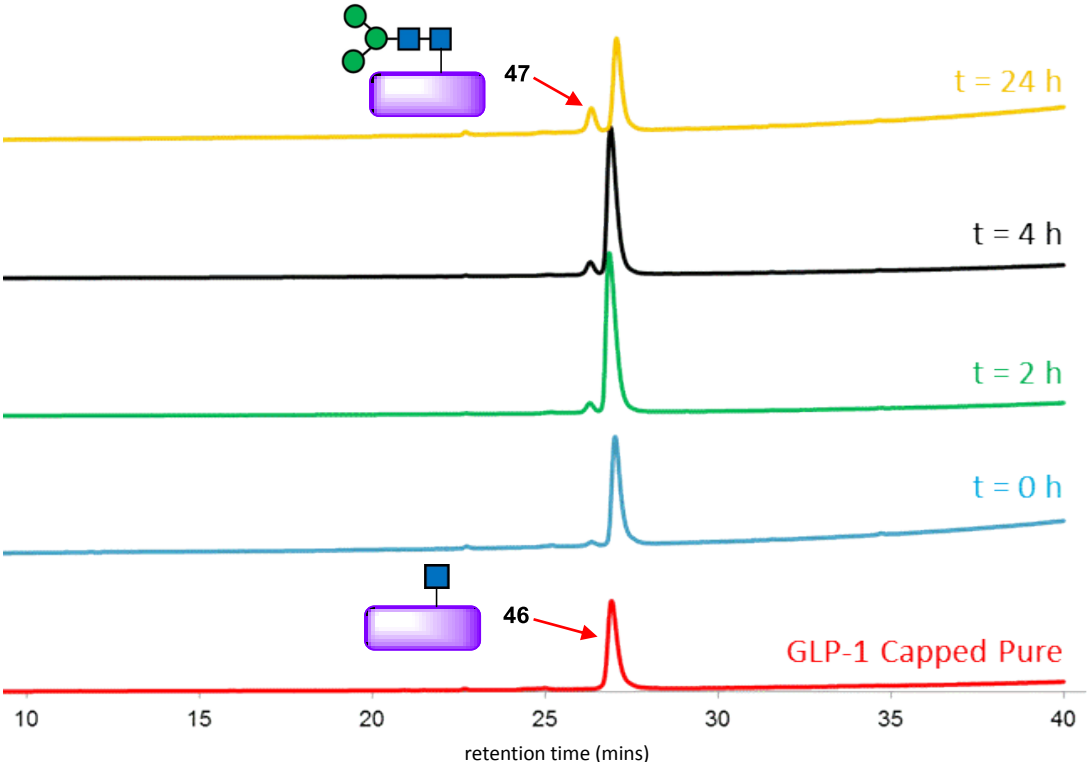


Figure 5.6: Analytical (RP)HPLC data with UV absorption at 230nm monitoring Endo A E173Q mediated transfer to GLP-1

Over 24 hours, a peak which eluted 1.0 minute earlier than the starting material was observed and assumed to be the more hydrophilic product. However, the rate of reaction was disappointing and it was concluded that background hydrolysis of the donor oxazoline would overall limit the maximum obtainable yield. This prompted the investigation of wild type Endo A, which was carried out under the same conditions and monitored by (RP)HPLC (Figure 5.7).

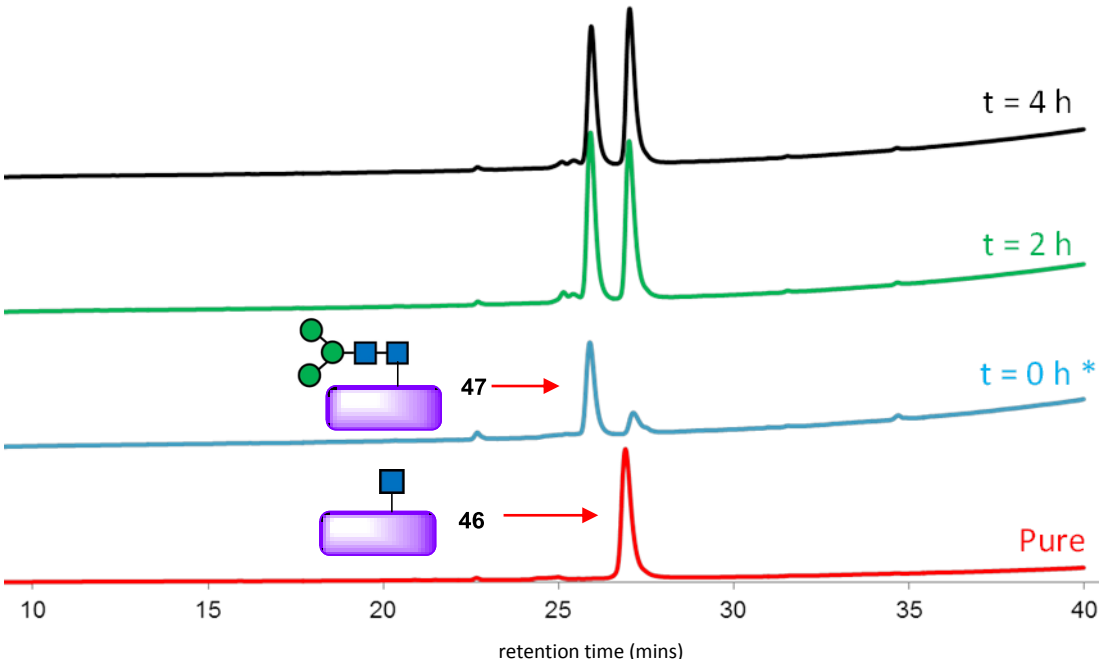


Figure 5.7: Analytical (RP)HPLC data with UV absorption at 230nm monitoring Endo A (wild type) mediated transfer to GLP-1. * The time point corresponding to t = 0 h was injected shortly after addition of the reaction components.

Remarkably, the first analytical time point (which was taken immediately after the addition of reaction components) suggested an almost instantaneous transformation to the desired pentasaccharide. Over the next 4 hours, it appeared that the newly formed product slowly reverted back to the starting material. LC-MS analysis confirmed successful transformation to the desired product, which was subsequently purified by RP(HPLC) (Figure 5.8).

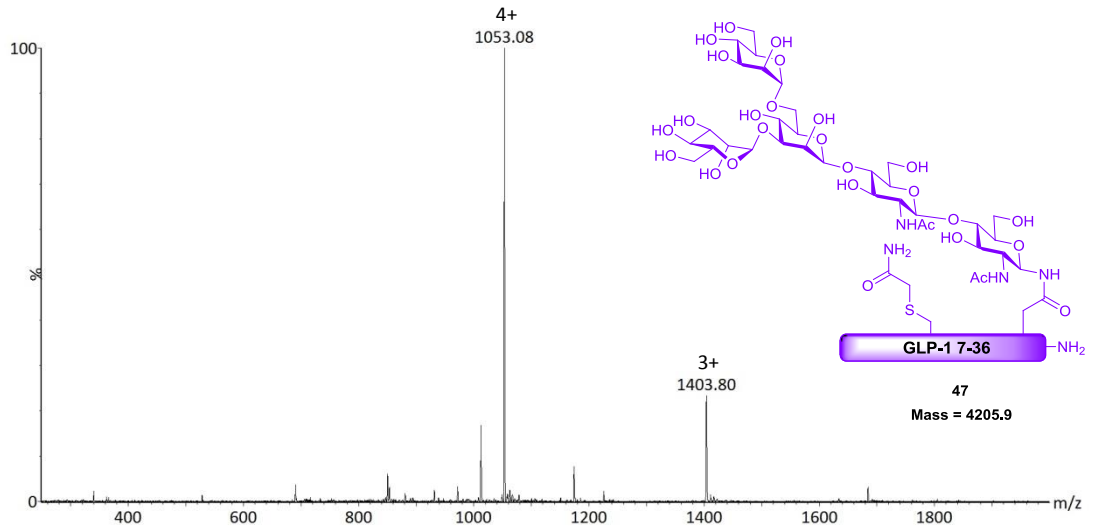
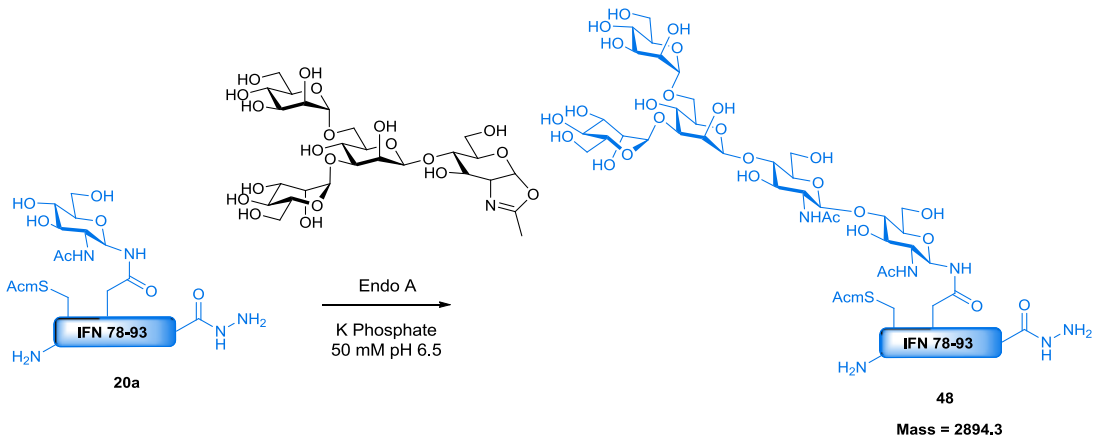


Figure 5.8: ES+ mass spectrum of GLP-1 pentasaccharide analogue

5.4.2.1 IFNβ transfer assays

Inspired by the rapid transfer observed in model studies, the procedure was applied towards the IFNβ 78-93 glycopeptide (Scheme 5.11).



Scheme 5.11: Endo A (wild type) mediated transfer to IFNβ 78-93 with $Man_3GlcNAc$ oxazoline

With the previous experiment suggesting that maximum yield was obtained in the early stages, the glycosylation reaction was allowed to proceed for 15 minutes before analysis by LC-MS (Figure 5.9). Once again, formation of the pentasaccharide was observed and the mixture was immediately purified by (RP)HPLC.

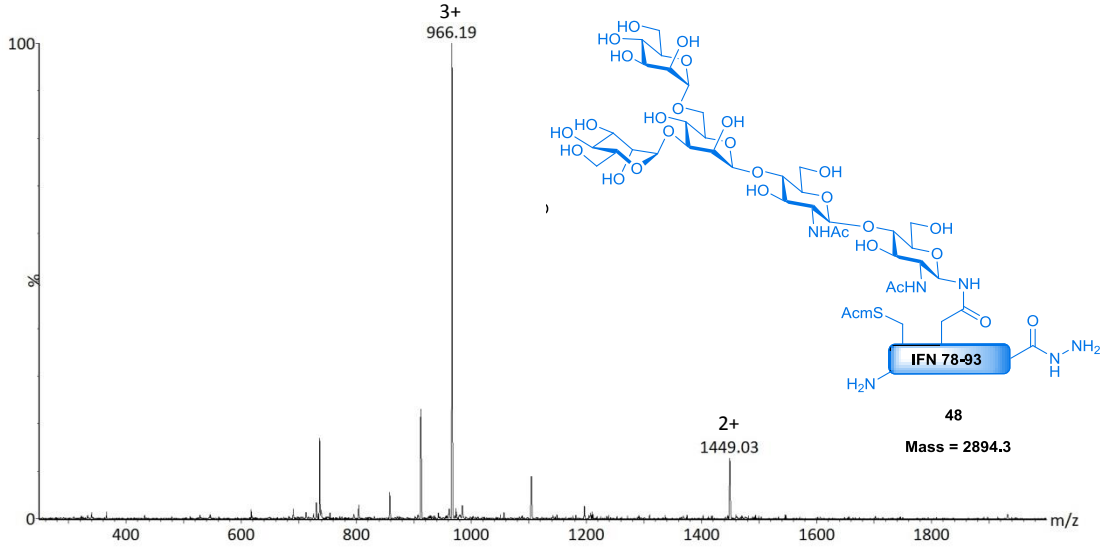


Figure 5.9: ES+ mass spectrum of IFNβ 78-93 pentasaccharide analogue

5.5 Conclusions

Endoglycosidase A was successfully used to form a novel analogue of GLP-1 bearing an *N*-linked pentasaccharide. Although biological activity has yet to be evaluated, the new analogue may extend the half-life of GLP-1 *in vivo* and may hold therapeutic potential. The endoglycosidase strategy was also applied towards an IFNβ glycopeptide, demonstrating the potential compatibility of this approach with the semi-synthetic strategy described in Chapter 4.

Disappointingly, the E173Q mutant of Endo A did not appear to catalyse glycosylation efficiently. Subsequent investigations may benefit from the employment of alternative mutants such as the E173H mutant variant. Nonetheless, the wild type enzyme was capable of furnishing the desired products, albeit with accompanied hydrolysis after extended reaction times.

By demonstrating the feasibility of this strategy, it is envisaged that future studies will allow the extension of this methodology to a number of alternative glycosyl donors and enzymes, allowing the production of a range of glycosylated peptides and proteins.

6 Synthesis of 3-carboxymethyl galactose uridine diphosphate

6.1 Introduction

To further expand the utility of endoglycosidase mediated transfer of complex glycans for the semi-synthesis of *N*-glycoproteins, it was decided to investigate the compatibility of this methodology with novel glycoprotein analogues. Given the importance of sialic acid in complex glycans to prolong the activity of glycoproteins *in vivo*, it was decided to investigate non-hydrolysable mimics of the sialyl galactose. In the absence of sialic acid, glycoproteins are removed from circulation via the asialoglycoprotein receptor in the liver and are subsequently degraded.[244] Therefore it was postulated that a non-hydrolysable mimic may present a method of extending the half-life of *N*-glycoproteins. Previous studies had indicated that sialyl Lewis^X (SLe^x) analogues retaining the negative carboxylate moiety were still able to bind to *E*-selectin.[245-247] This observation was consistent with the crystal structure of SLe^x bound to *E*-selectin, which revealed the critical residues involved in this interaction (Figure 6.1).[248]

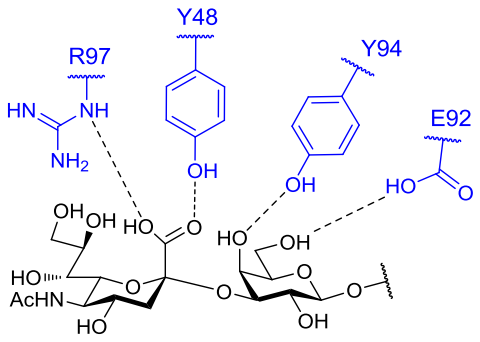


Figure 6.1: Interaction between sialyl galactose of SLe^x and *E*-selectin revealed by an X-ray crystal structure[248]

Based on these findings, a non-hydrolysable sialyl galactose mimic was designed, retaining the important carboxyl motif (Figure 6.2). The incorporation of this motif into glycosyl oxazolines provides a novel substrate for Endo A, potentially generating novel *N*-glycoproteins which may exhibit prolonged activity *in vivo.*

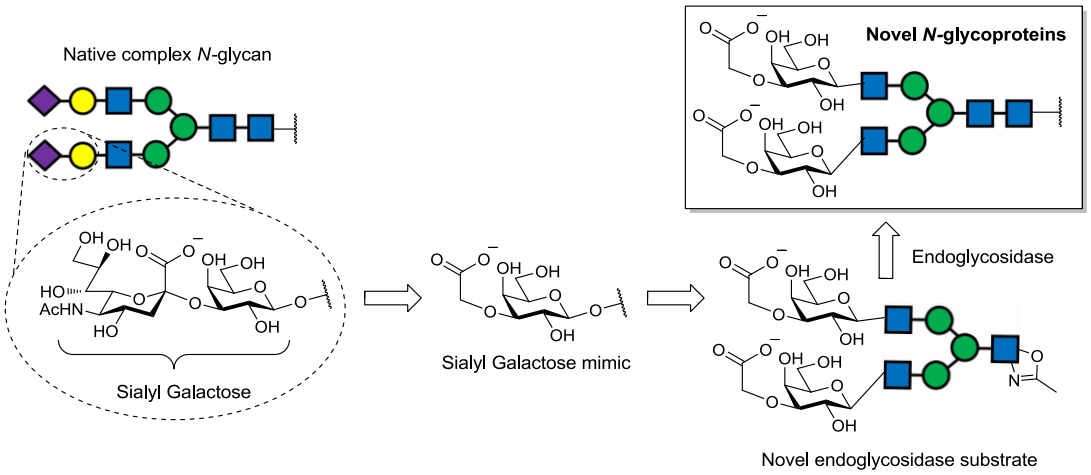


Figure 6.2: Rationale for 3-carboxymethyl galactose as a sialyl galactose mimic

It was reasoned that the most convenient method of mimic incorporation would be through the use of a commercially available glycosyltransferase, thus avoiding the laborious synthesis of the complex glycan. This strategy could also allow for glycoproteins to be "remodelled" by treating complex type glycans with sialidases and galactosidases before transfer of the mimic to the terminal *N*-acetylglucosamine acceptor. The formation of the lactosamine linkage usually relies on a β-1,4 galactosyltransferase, which processes galactose in the activated form of a UDP sugar nucleotide. Accordingly, the initial investigation focussed on the synthesis of 3-carboxymethyl galactose uridine diphosphate (Figure 6.3) to facilitate the incorporation of the mimic through the use of a galactosyltransferase.

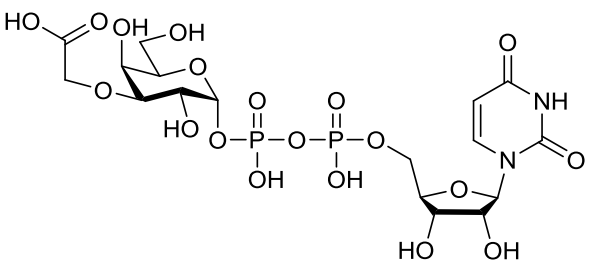
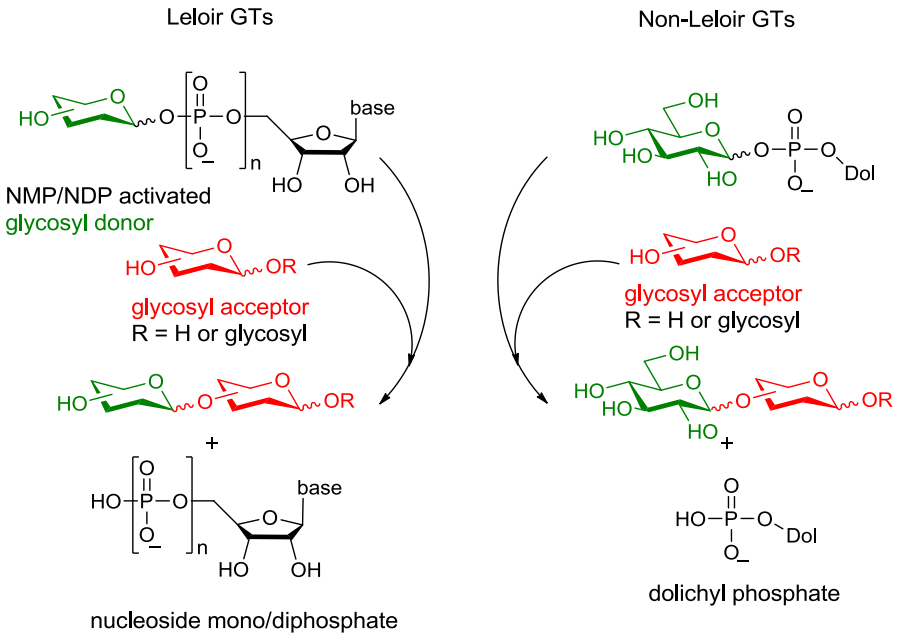


Figure 6.3: 3-carboxymethyl galactose uridine diphosphate

6.1.1 Glycosyltransferases

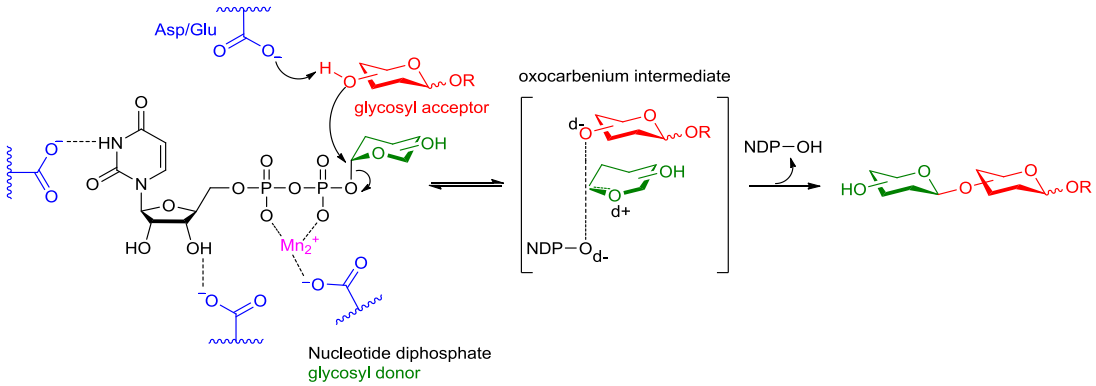
Glycosyltransferases (GTs) are a class of enzymes capable of forming glycosidic bonds with precise regio- and stereoselectivity, playing a vital role in the biosynthesis of oligosaccharides and glycoconjugates.[249] GTs utilise activated glycosyl donors containing a good leaving group at the anomeric position to facilitate transfer. Depending on the nature of this leaving group, eukaryotic GTs fall into two families – the Leloir type and the non-Leloir type (Scheme 6.1).[250] Most mammalian GTs belong to the Leloir type subfamily and require nucleotide donors such as UDP-Glc, UDP-Gal, UDP-GlcNAc, GDP-Man and CMP-Sialic acid. Not only does the nucleotide serve the role of a leaving group, it is also important for GT recognition.[251] The active site of a GT binds the reacting donor and acceptor in a very specific manner, ensuring regio- and stereoselectivity in the glycosylation reaction. For certain GTs, a divalent cation such as Mn^{2+} is also important in aiding the loss of the leaving group.[252]



Scheme 6.1: Glycosidic bond formation using GTs

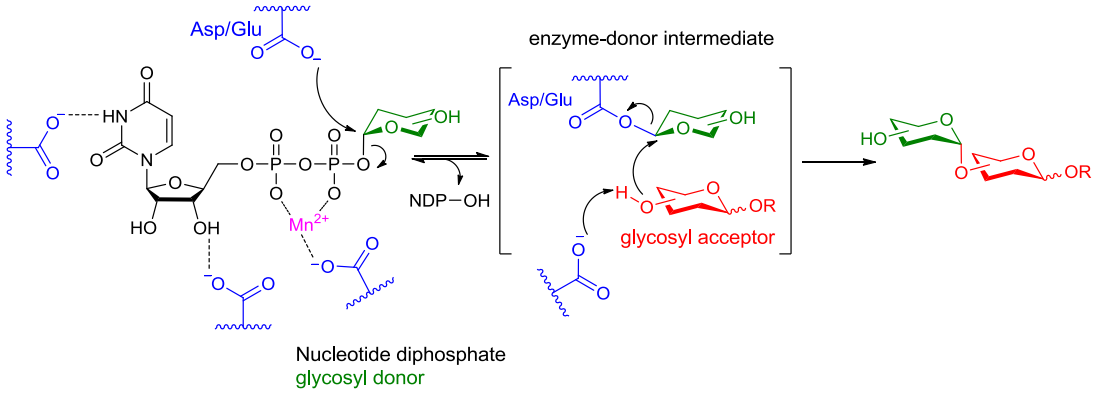
GTs are also classified as inverting or retaining, according to the stereochemistry observed at the anomeric centre after glycosidic bond formation. Whilst the mechanism of inverting GTs is well established (Scheme 6.2),[253, 254] the mechanism for retaining GTs is less understood.

For inverting GTs, basic carboxylate side chains from Asp or Glu facilitate nucleophilic attack by the glycosyl acceptor. The leaving group of the sugar nucleotide facilitates the formation of an oxocarbenium intermediate, enabling the glycosidation reaction to take place. The result of this process gives rise to inversion of stereochemistry at the anomeric position of the donor.



Scheme 6.2: Mechanism for inverting GTs

For retaining GTs, it has been suggested that the reaction occurs via a double displacement S_N2 with net retention of configuration (Scheme 6.3).[251, 255, 256] This reaction is initiated by direct attack of a carboxylate residue from Asp or Glu at the anomeric centre of the donor. The enzyme-donor intermediate is then displaced by another S_N2 process by the glycosyl acceptor, giving net retention of stereochemical configuration in the product.

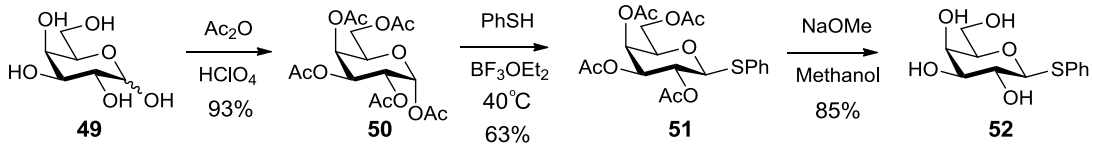


Scheme 6.3: Suggested double displacement mechanism for retaining GTs[41]

6.2 Synthesis of UDP 3-carboxymethyl galactose

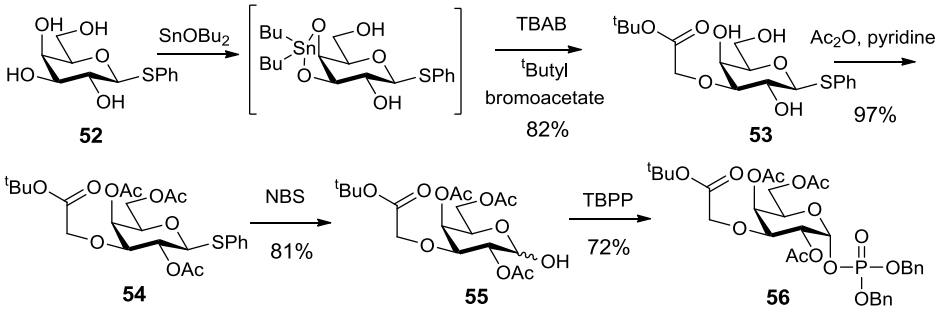
The most popular method of preparing nucleoside diphosphate (NDP) sugars exploits the formation of the pyrophosphate bond. Coupling a sugar phosphate and a nucleoside monophosphate preserves the anomeric configuration of the sugar, whereas glycosylation of NDPs can lead to loss of stereocontrol at the anomeric centre.[257] Activated nucleoside monophosphates are commercially available in the form of morpholidates and hence efforts have been driven towards the synthesis of a 3-carboxymethyl galactose (3-CG) phosphate.

In the first step of 3-CG phosphate synthesis, the thiophenyl glycoside of galactose was formed. The thioglycoside enables protection of the anomeric carbon, exhibiting stability to a wide range of conditions and can be selectively removed by mild hydrolysis.[258] Initially, D-Galactose **49** was peracetylated with acetic anhydride in the presence of catalytic perchloric acid, providing the α-configured anomer. Thioglycosidation of the peracetylated sugar was performed using thiophenol and catalytic boron trifluoride diethyl etherate to give the β-thioglycoside in moderate yield.[259] Deacetylation with a catalytic amount of sodium methoxide in methanol afforded the deprotected thioglycoside **52**.



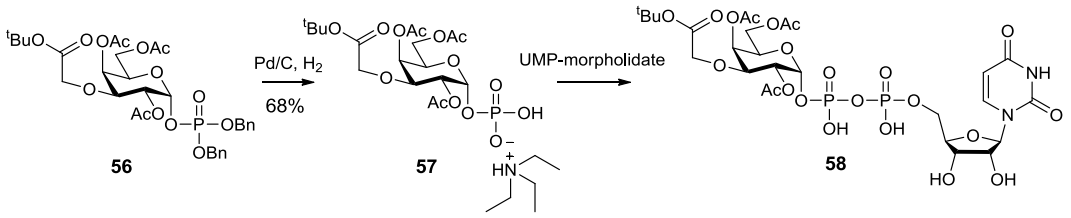
Scheme 6.4: Synthesis of thioglycoside 52

The hydroxyl group at C-3 was selectively activated for alkylation by reaction of the thioglycoside with dibutyl tin oxide. The *cis* relationship of the hydroxyl groups at C-3 and C-4, allows the formation of a stannylene acetal and the more reactive equatorial C-3 can react with *tert*-butyl bromoacetate. The free hydroxyls at C-2, C-4 and C-6 were subsequently reacetylated with acetic anhydride and pyridine, before treatment of **54** with *N*-bromosuccinimide in aqueous acetone to selectively cleave the thiophenyl glycoside. The anomeric mixture was deprotonated with LDA prior to treatment with tetrabenzyl pyrophosphate, resulting in formation of the α-phosphate **56** exclusively (Scheme 6.5).



Scheme 6.5: Synthesis of 3-carboxymethyl galactose phosphate

Following phosphorylation, the benzyl groups were removed by hydrogenation in the presence of a palladium catalyst. As the resulting phosphate **57** was expected to be highly acidic, the reaction was quenched with triethylamine to prevent the unwanted cleavage of the *tert*-butyl ester. The formation of the pyrophosphate bond was conducted in DMF over 24 hours with the commercially available UMP morpholidate, followed by ion exchange to the sodium salt (Scheme 6.6).[260, 261] The presence of the pyrophosphate linkage was confirmed by ^{31}P NMR (Figure 6.4).



Scheme 6.6: Deprotection of sugar monophosphate 56 and formation of the pyrophosphate bond

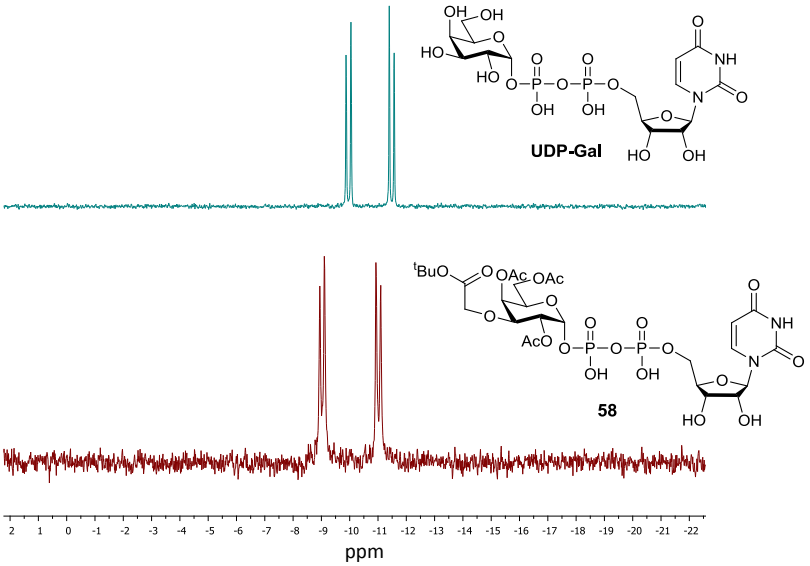
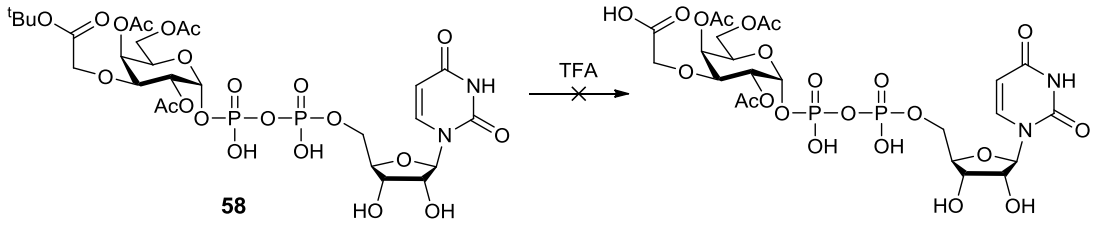


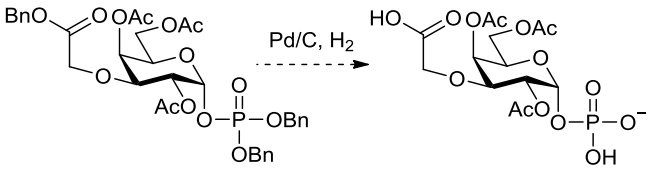
Figure 6.4: ^{31}P NMR (D_2O) comparison of commercially purchased UDP-Gal and sugar nucleotide 58 confirming successful pyrophosphate bond formation. The pyrophosphate linkage gives rise to two doublets which arise from phophorus-phosphorus coupling.

Subsequently, the removal of the *tert*-butyl protecting group was investigated using various concentrations of aqueous TFA. Initial experiments using 5% or 20% TFA showed that cleavage of the ester was sluggish. Over extended periods of time, ^{31}P NMR revealed cleavage of the pyrophosphate bond, indicated by the presence of a singlet corresponding to monophosphate **57**. This led to attempted *tert*-butyl ester cleavage with higher concentrations of TFA. Indeed the use of 95% TFA resulted in rapid removal of the *tert*-butyl group, although hydrolysis of the pyrophosphate was still observed. With the incompatibility of the pyrophosphate bond with *tert*-butyl ester cleavage, an alternative protecting group strategy was investigated.



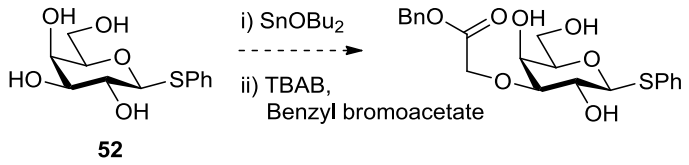
Scheme 6.7: Unsuccessful cleavage of *tert*-butyl ester with TFA

It was postulated that carboxylate protection with a benzyl group would allow simultaneous deprotection during debenzylation of the protected sugar monophosphate (Scheme 6.8). Despite the presence of a free carboxyl group, previous studies have shown that this functionality is compatible with pyrophosphate formation.[262, 263]



Scheme 6.8: Alternative protection strategy of the carboxylate as a benzyl ester facilitates concomitant removal during deprotection of the sugar monophosphate

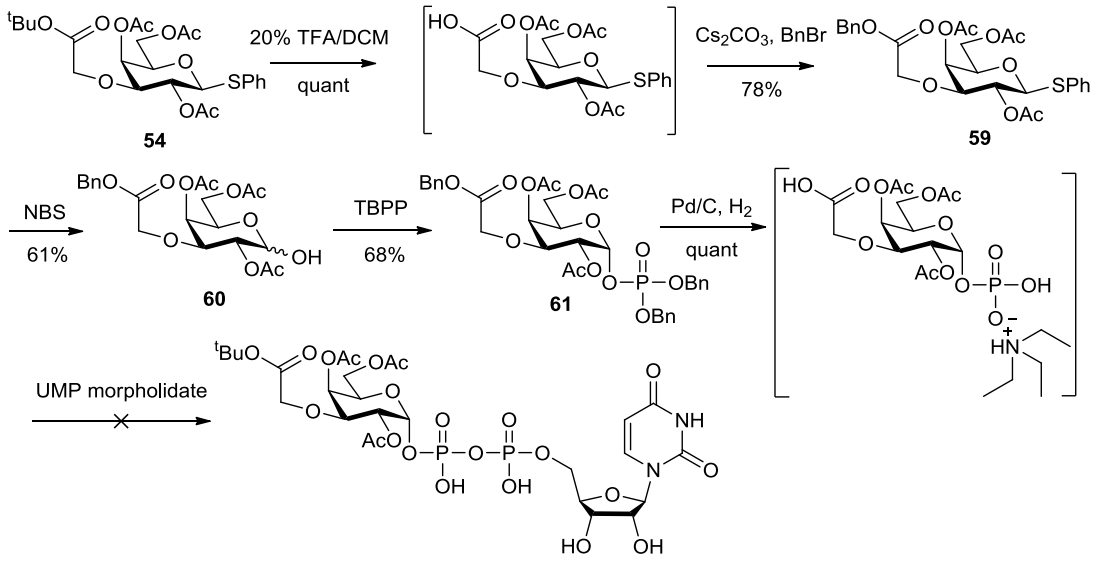
In an initial attempt to synthesise the benzyl protected derivative, it was decided to alkylate thioglycoside **52** with benzyl bromoacetate via the stannylene acetal as before (Scheme 6.9).



Scheme 6.9: Proposed synthesis of benzyl protected thioglycoside

Surprisingly, the use of benzyl bromoacetate as an alkylating reagent did not successfully lead to the formation of the desired product. However, this observation was consistent with a previous study which also reported inefficient alkylation of a thiogalactoside with benzyl bromoacetate.[247]

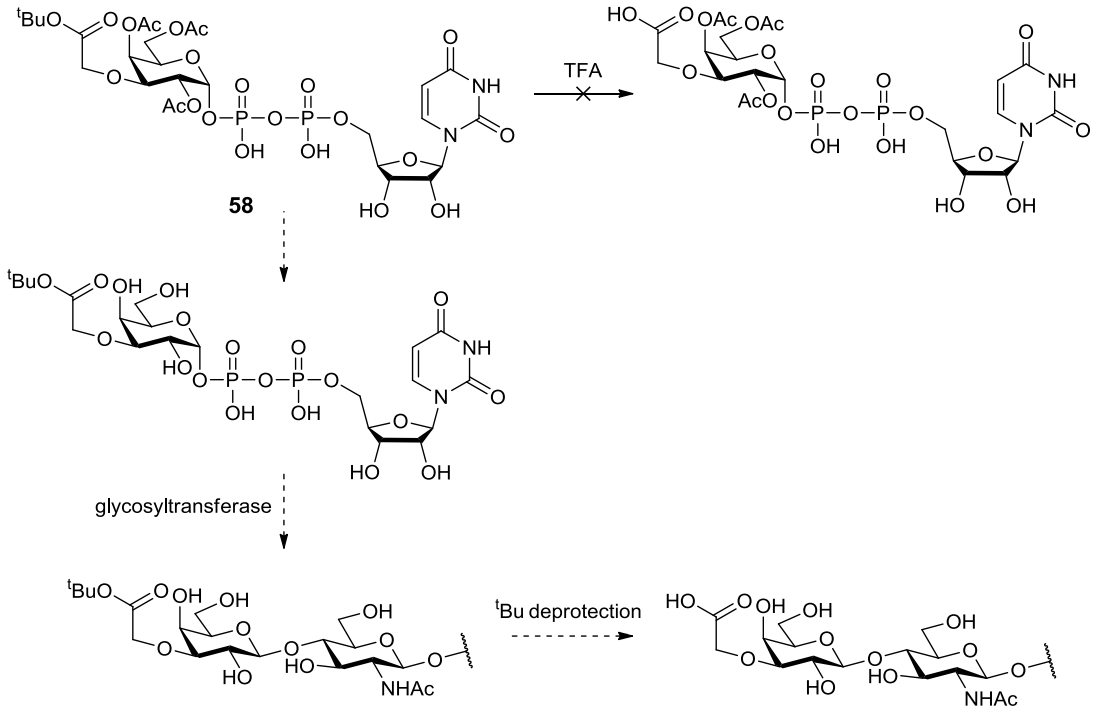
As a result, the benzyl protected thioglycoside was synthesised from derivative **54**, substituting *tert*-butyl protection in the form of a benzyl ester. Cleavage of the *tert*-butyl group took place quantitatively in 20% TFA and the free carboxylate was esterified by treatment with caesium carbonate and benzyl bromide.[264] Following hydrolysis of thioglycoside **59** and subsequent phosphorylation, debenzylation of **61** was achieved by hydrogenation as before.



Scheme 6.10: Synthesis of sugar nucleotide via benzyl protection strategy

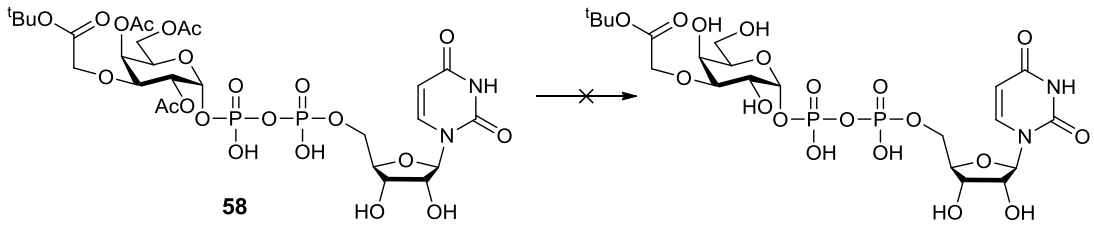
Unfortunately, the critical formation of the pyrophosphate linkage proved unsuccessful using the debenzylated sugar monophosphate and UMP morpholidate. Despite a number of attempts, no product was detected by TLC or crude ^{31}P NMR, even after extensive drying of reaction components by co-evaporation with dry pyridine and addition of molecular sieves.

At this point, it was hypothesised that a strategy relying on *tert*-butyl ester removal after glycosyl transfer (Scheme 6.11) could prove more feasible and the previous strategy was revisited.



Scheme 6.11: Alternative strategy relying on tert-butyl deprotection following glycosyl transfer

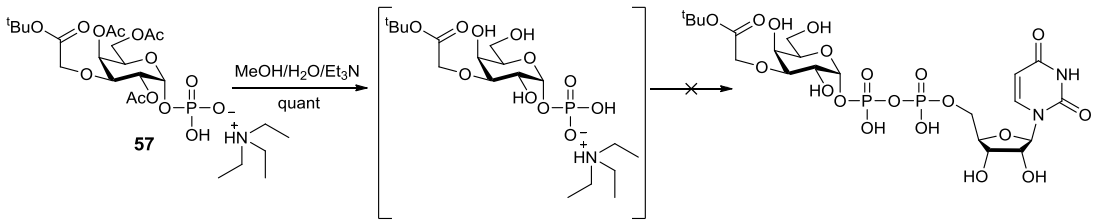
Consequently, efforts were directed towards the removal of the remaining sugar acetate groups from sugar nucleotide **58**. The sensitivity of the pyrophosphate motif led to the exploration of protocols which had previously been shown to be compatible with sugar nucleotides. The deprotection was first attempted in a mixture of methanol/water/triethylamine (5:2:1) according to the procedure described by Marchesan and Macmillan.[265] These relatively mild conditions are popular for deacetylation, as they are typically high yielding and product purification is achieved by extraction followed by lyophilisation of the aqueous layer. Unfortunately, cleavage of the acetate groups proceeded much slower than expected and was accompanied by pyrophosphate cleavage. Zemplen deacetylation with sodium methoxide in methanol was also found to be ineffective due to the poor solubility of the sugar nucleotide in methanol. Later, the use of aqueous lithium hydroxide and hydrazine hydrate was attempted, but these conditions also resulted in the breakdown of the sugar nucleotide or inefficient deacetylation.



Conditions	Observations
MeOH/H_2O/Et_3N (5:2:1)	Inefficient deacetylation after 24 hours, pyrophosphate cleavage occurs slowly
MeOH/H_2O/Et_3N (7:3:1)	Inefficient deacetylation after 5 days, pyrophosphate cleavage occurs slowly
0.1 M LiOH	Inefficient deacetylation after 24 hours, pyrophosphate cleavage occurs slowly
1 M LiOH	Quantitative deacetylation accompanied by pyrophosphate cleavage within 4 hours
1% Hydrazine hydrate	Inefficient deacetylation after 24 hours, pyrophosphate cleavage occurs slowly
5% Hydrazine hydrate	Quantitative deacetylation accompanied by pyrophosphate cleavage within 2 hours
NaOMe in methanol	Sugar nucleotide insoluble in methanol, inefficient deacetylation over 24 hours

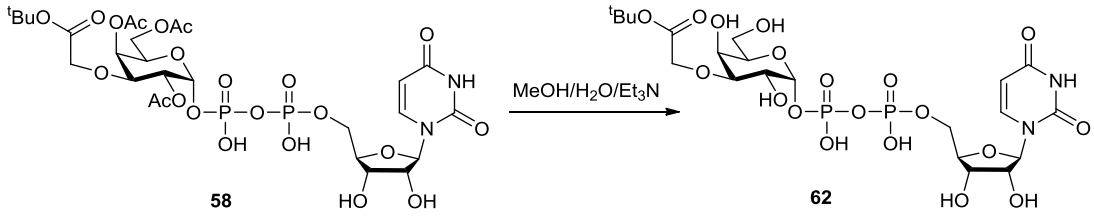
Figure 6.5: Reaction conditions and outcomes for attempted deacetylation of sugar nucleotide 58

This further prompted a redesign of the approach to the final product. Removal of the protecting groups before pyrophosphate formation could potentially obviate the difficulties encountered previously. However, it was also found when attempting the coupling in the absence of acetate groups, formation of the desired product was not observed.



Scheme 6.12: Attempted synthesis of sugar nucleotides with deacetylated sugar monophosphate

Finally, the previous attempts to deacetylate the sugar nucleotide using methanol/water/triethylamine (5:2:1) were re-examined. Although previous experiments had shown the reaction was slow, pyrophosphate cleavage did not occur to the same extent as observed under the alternative conditions. The reaction was allowed to proceed for 7 days, resulting in quantitative deacetylation albeit accompanied with hydrolysis of the pyrophosphate linkage. Removal of the acetate groups was confirmed by ^{1}H NMR (Figure 6.6), and ^{31}P NMR indicated the presence of **62** (Figure 6.7).



Scheme 6.13: Deacetylation of sugar nucleotide 58

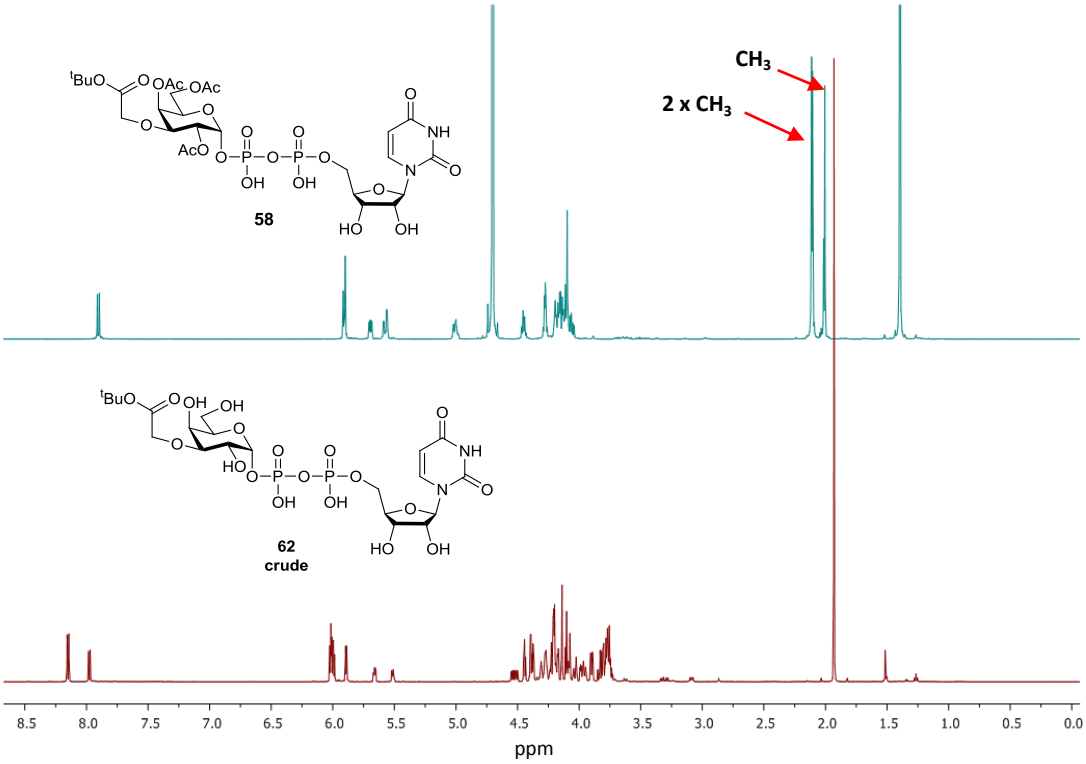


Figure 6.6: ^{1}H NMR (D_2O) comparing sugar nucleotide 58 (upper) with crude deacetylated mixture (lower) after treatment with MeOH/H_2O/Et_3N (5:2:1) for 7 days. From comparison of the spectra, efficient deacetylation was confirmed by the absence of 3 x CH_3 singlets.

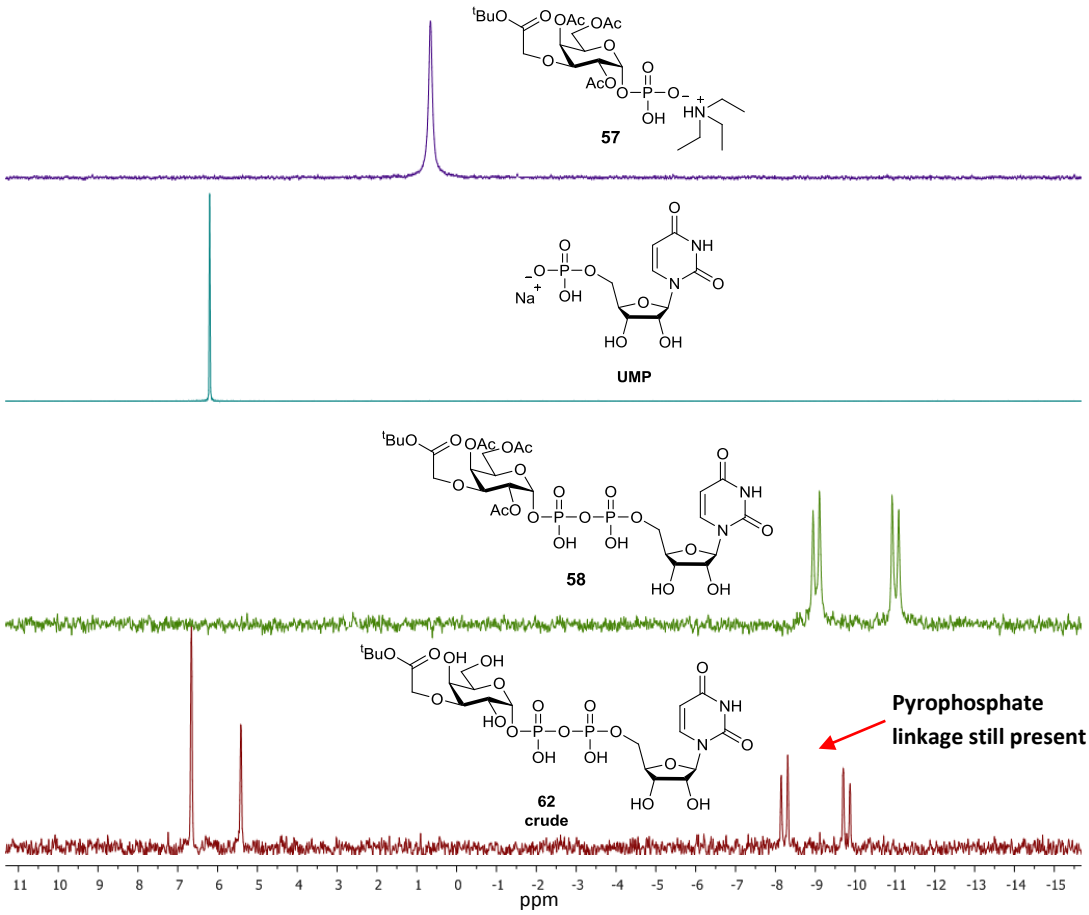
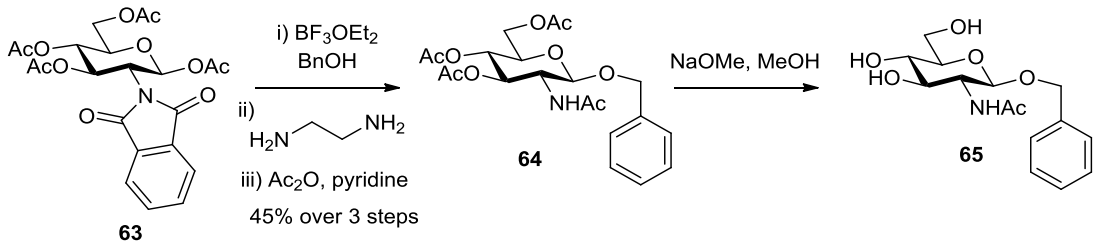


Figure 6.7: ^{31}P NMR comparing (upper to lower) sugar monophosphate 57 ($CDCl_3$), uridine monophosphate (D_2O), sugar nucleotide 58 (D_2O) and crude deacetylated mixture (D_2O). Comparison of the spectra confirms the presence of a pyrophosphate linkage in the crude deacetylated mixture, although appears to be accompanied by species corresponding to monophosphates.

6.3 Synthesis of GlcNAc benzyl glycoside

To initially survey the ability of a glycosyltransferase to accept UDP 3-carboxymethyl galactose **62** as a substrate, the benzyl glycoside of *N*-acetylglucosamine was chosen as a glycosyl acceptor. It was envisaged that the benzyl group incorporated a UV chromophore, allowing visualisation during (RP)HPLC analysis of transfer assays. The glycoside was synthesised from phthalamido protected GlcNAc **63**[266] (provided by Dr Derek Macmillan) furnishing the benzyl glycoside **64** in 45% yield over 3 steps. Subsequent Zemplen deacetylation afforded the desired unprotected glycoside **65**.



Scheme 6.14: Synthesis of GlcNAc benzyl glycoside 65

6.4 Conclusions

The latter stages of sugar nucleotide synthesis have presented significant problems when investigating protecting group removal as the pyrophosphate linkage appears vulnerable to both acidic and basic conditions. The coupling of unprotected sugar phosphates was accordingly investigated, however disappointingly failed to yield the product.

The objective of this study was to synthesise a novel sugar nucleotide analogue and assess its ability to be transferred by a galactosyltransferase. With this in mind, this study has still provided a source of sugar nucleotide **62** which can be tested as a substrate. A suitable glycosyl acceptor **65** has also been synthesised for use in the initial transfer assays. Future studies may yield a more efficient synthesis of the sugar nucleotide and access to purer samples. In time, this should allow the crucial evaluation of a sialyl galactose mimic on glycoproteins *in vivo.*

7 Final conclusions and outlook

7.1 Semi-synthesis of *N*-linked glycoproteins

A semi-synthetic strategy towards *N*-glycoproteins was directed towards Interferon β-1, a 166 residue cytokine bearing a single site of *N*-glycosylation. IFNβ represented an ideal target for the demonstration of this methodology, as the central position of glycosylation necessitated the expression of two flanking protein sequences, each requiring distinct functionalisation. By meeting this criterion, the developed methodology may potentially be applied to any *N*-glycoprotein of interest.

The application of *N*→*S* acyl transfer has allowed the preparation of *N*-glycopeptide thioesters for participation in protein assembly through Native Chemical Ligation. This was successfully demonstrated on a model glycopeptide corresponding to EPO residues 22-28, which was subsequently able to ligate with a recombinant protein. However, the elevated temperature required to drive formation of a thioester was found to be incompatible with hydrophobic sequences. When applied to the thioester formation of the glycopeptide corresponding to IFNβ 78-93, aggregation was observed, ultimately leading to poor conversions. Although temperature could be lowered to minimise aggregation, longer reaction times would result in hydrolysis of the thioester. The utility of peptide hydrazides and their use in NCL led to the exploration of hydrazine in the acyl transfer reaction. Pleasingly, this reaction afforded the respective hydrazide, which exhibited stability over longer reaction periods. This allowed the functionalisation of hydrophobic sequences, overcoming the earlier encountered difficulties. Whilst these studies were intriguing, the impact of this reaction for the functionalisation of synthetic peptides was limited, as peptide hydrazides can be prepared directly on specialised resins.

The true potential of this reaction was unlocked when exploring the functionalisation of recombinant proteins. Despite intein and sortase technologies offering access to recombinant protein thioesters (and hydrazides), these methods are typically limited to soluble protein sequences. The developed hydrazinolysis reaction can be performed in the presence of chaotropes and is potentially applicable to any sequence of interest. Accordingly, the highly insoluble N-terminal fragment (IFNβ 1-77) was converted to the corresponding C-terminal hydrazide.

The C-terminal fragment (IFNβ 94-166) was also found to be highly insoluble, excluding the protein from a number of existing functionalisation methods. The liberation of an N-terminal cysteine was eventually achieved using a highly chemoselective treatment of cyanogen bromide, although this required mutation of methionine residues from within the sequence.

With all the appropriately functionalised fragments in hand, IFNβ assembly was attempted using the hydrazide ligation methodology developed by Liu. The ligation was successfully conducted between the glycopeptide and C-terminal fragment, allowing the isolation of a semi-synthetic glycoprotein corresponding to IFNβ 78-166. Although the final ligation step was attempted, the final product corresponding to full length IFNβ was not observed. Future studies with larger amounts of material may enable optimisation of this step, allowing the preparation of full length IFNβ.

In spite of this outcome, the study has demonstrated the potential of the hydrazinolysis reaction, importantly overcoming the limitations of current methodologies. This should permit a widespread application of the semi-synthetic strategy towards any protein of interest.

7.2 Glycosylation with Endohexosaminidases

In recent years, the use of endoglycosidases for the glycosylation of proteins has proved an invaluable tool. Nonetheless, the use of this methodology requires access to complex glycans, where preparation is typically restricted to specialist laboratories. As a consequence, the use of this technology has been limited to a small number of research groups around the world. In this study, a tetrasaccharide oxazoline was successfully synthesised and transferred to glycopeptides bearing a natively linked *N*-acetylglucosamine acceptor. This led to the preparation of a glycosylated analogue of GLP-1, as well as IFNβ 78-93 for use in our semi-synthetic strategy. The first demonstration of endoglycosidase technology within this research group should undoubtedly pave the way towards future studies with a range of glycans and enzyme variants to generate homogenous *N*-glycoproteins.

7.3 3-carboxymethyl galactose as a sialyl galactose mimic

The synthesis of 3-carboxymethyl galactose uridine diphosphate for the preparation of sialyl galactose mimics through glycosyl transfer has proved challenging in the final stages, as the pyrophosphate bond was found to be compromised upon removal of the protecting groups. Attempting formation of the pyrophosphate in the absence of protecting groups was met with considerable difficulty and unsuccessful isolation of the product. Consequently, carboxylate protection was retained with the intention that deprotection could be performed after glycosyl transfer. However, it is still unknown if the presence of the *tert*-butyl group will be tolerated in this reaction, in which case it may be necessary to investigate an alternative method of protection. The deacetylation step was examined under a number of conditions, but a clean conversion to the desired product without breakdown of the pyrophosphate has not been discovered. Future studies may uncover an improved procedure for deacetylation, accompanied by an efficient method for purification of the sugar nucleotide. In the meantime, the crude deacetylated mixture may provide a suitable source of the glycosyl donor. It is envisaged that successful transfer of the mimic will provide novel glycan substrates for Endo A and the preparation of *N*-glycoproteins with enhanced stability *in vivo*.

8 Experimental

8.1 General experimental

Chemicals were obtained from Novabiochem, Sigma-Aldrich, Fischer Scientific and AGTC Bioproducts, used without further purification unless mentioned otherwise. Petroleum Spirit refers to petroleum ether that boils in the range 40-60 °C. Flash chromatography was carried out with silica gel 40-63 µm (VWR). Analytical TLC was performed on Merck Silica gel 60 F_{254} aluminium sheets (0.25 mm layer coating) and was visualised with short wavelength (366 nm) ultraviolet light. Staining was performed with a solution of anisaldehyde, using heat to visualise spots. Anhydrous magnesium sulfate was used to dry organic extracts.

Proton NMR (^{1}H NMR) were recorded at 500 MHz or 600 MHz on Bruker AMX500 and AMX600 instruments respectively. ^{13}C NMR experiments were run on the same instruments at 125 MHz or 150 MHz respectively. ^{31}P NMR was recorded at 121 on a Bruker AMX300. All NMR experiments were conducted at room temperature. Chemical shifts (δ) are recorded in parts per million (ppm) and reported with reference to the solvent peak. $CDCl_3$ refers to fully deuterated chloroform, MeOD (CD_3OD) refers to fully deuterated methanol, D_2O to fully deuterated water and DMSO-d_6 refers to fully deuterated dimethyl sulfoxide. Coupling constants (*J*) are reported in hertz, calculated from observed chemical shifts. LC-MS was conducted on a Waters UPLC/SQD-LC mass spectrometer with a linear gradient from 95% of solvent A (water with 0.1% formic acid) to 95% solvent B (acetonitrile with 0.1% formic acid) over 4 minutes at a flow rate of 0.6 mL/min. FAB and ESI mass spectrometry experiments were performed Dr Lisa Haigh or Dr Vincent Gray from the Department of Chemistry at UCL. Melting points were determined using an SRS Digimelt melting point apparatus. Infrared spectra were recorded on a Bruker Alpha Platinum ATR FT-IR spectrometer.

8.1.1 HPLC purification methods

Analytical (RP)HPLC was carried out using a Phenomenex SphereClone™ 5 µm ODS(2) 80 Å, LC Column (diameter = 250 mm × 4.6 mm): flow rate 1.0 mL/min, detection at 230, 254 and 280 nm using a gradient of 95% water (0.1% TFA)/5% acetonitrile (0.1% TFA), to 5% water (0.1% TFA)/95% acetonitrile (0.1% TFA), over 45 min, using a Dionex Ultimate 3000.

Semi-preparative (RP)HPLC method 1 was carried out using a Phenomenex Luna 5 µm C18 100 Å, LC Column (diameter = 250 mm × 10.0 mm): flow rate 4.0 mL/min, detection at 230, 254 and 280 nm (Dionex UVD170U) using a gradient of 95% water (0.1% TFA)/5% acetonitrile (0.1% TFA), to 40% water (0.1% TFA)/60% acetonitrile (0.1% TFA), over 45 min, using a P680 HPLC pump and a Foxy Jr fraction collector.

Semi-preparative (RP)HPLC method 2 was carried out using a Phenomenex Jupiter 4 µm Proteo 90 Å, LC Column (diameter = 250 mm × 10.0 mm): flow rate 4.0 mL/min, detection at 230, 254 and 280 nm (Dionex UVD170U) using a gradient of 95% water (0.1% TFA)/5% acetonitrile (0.1% TFA), to 70% water (0.1% TFA)/30% acetonitrile (0.1% TFA), over 5 min, to 30% water (0.1% TFA)/70% acetonitrile (0.1% TFA), over 35 min, using a P680 HPLC pump and a Foxy Jr fraction collector.

Preparative (RP)HPLC was carried out using a Phenomenex Jupiter 4 µm Proteo 90 Å, LC Column (diameter = 250 mm × 21.2 mm): flow rate 10.0 mL/min, detection at 230, 254 and 280 nm using a gradient of 95% water (0.1% TFA)/5% acetonitrile (0.1% TFA), to 40% water (0.1% TFA)/60% acetonitrile (0.1% TFA), over 45 min, using a Dionex Ultimate 3000 equipped with a fraction collector unless specified otherwise.

Solvents were removed on an Edwards Modulyo freeze dryer equipped with an RV12 vacuum pump for 24-48 hours.

8.1.2 General microbiology

Standard techniques were performed as described in Molecular Cloning - A Laboratory Manual (Sambrook and Russell, 3rd edition, 2001, Cold Spring Harbour Press). Standard sterile technique was maintained during all microbiological work. All media and glassware were sterilised in an autoclave at 121 °C for 15 minutes. Agarose gels were prepared by dissolving agarose (1.0 g) in TAE buffer (100 mL) with microwave heating. Ethidium bromide (5 µL) was added to the agarose solution, which was subsequently poured into the electrophoresis chamber to set. Gels were run at 110 volts and visualised with UV light. SDS-PAGE was carried out using the Novex NuPAGE gel system (Life technologies) using 4% polyacrylamide for the stacking gel and 12-14% for the resolving gel. Samples in Gu.HCl were precipitated by a 10 fold addition of acetone/methanol (1:1) and storage at -20 °C for 5 minutes. The proteins were pelleted by centrifugation at 13000 RPM before resuspension in SDS-PAGE loading buffer. DNA sequences were verified by the sequencing service at the Wolfson Institute of Biomedical Research. Cell cultures were grown in shaker incubators at 175 RPM.

Preparation of *Escherichia coli* competent cells

LB medium 10 mL was seeded with the appropriate cell line and grown overnight with shaking at 37 °C. An aliquot of the overnight culture (100 µL) was used to seed fresh LB medium (10 mL) and grown to an absorbance of $OD_{600} \sim 0.2$. The cells were collected by centrifugation at 3000 RPM for 15 minutes (4 °C). The supernatant was discarded and the cells were suspended in fresh ice cold transformation buffer (50 mM $CaCl_2$, 10 mM Tris, pH 8.0, 4.0 mL). The cells were left on ice for 30 minutes, with occasional agitation. The cells were then collected by centrifugation at 3000 RPM for 15 minutes (4 °C). The supernatant was discarded and the cells were resuspended in 400 µL of ice cold transformation buffer.

Transformation of plasmids into *Escherichia coli* cells

A 100 µL aliquot of competent cells was transferred to a sterile microcentrifuge tube (1.5 mL) and left on ice for 30 minutes. The respective DNA (2.5 µL) was added; the tube was gently tapped and then left on ice for another 30 minutes. The sample was then heat shocked for 30 seconds at 42 °C on an Eppendorf thermomixer and then LB medium (1 mL) was added. The sample was then incubated for 1 hour at 37 °C and then collected by centrifugation at 3000 RPM for 15 minutes (4 °C). The supernatant was discarded and the cells were resuspended in fresh LB medium (100 µL). The cells were then spread to dryness on LB-agar plates with 100 µg/ml ampicillin and grown inverted at 37 °C overnight.

Restriction endonuclease digestions

Unless specified otherwise, the typical DNA digest was performed as follows:

PCR product or vector 40 µL
Compatible 10 x buffer 5 µL
Restriction endonuclease I 1 µL
Restriction endonuclease II 1 µL
Distilled sterile water 3 µL

DNA digests were performed at 37 °C overnight

Protein and DNA quantification

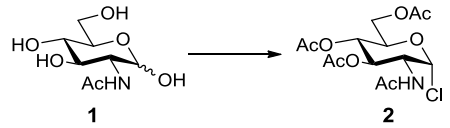
Single stranded or double stranded DNA concentrations were determined on an Eppendorf BioPhotometer using the appropriate function. Protein absorbance was measured on the same machine and quantified by the Beer-Lambert Law where:

$$A_{280} = \varepsilon.c.l$$

The molar extinction coefficient was predicted using the ProtParam tool from Expasy (web.expasy.org/protparam).

8.2 Chapter 2 experimental details

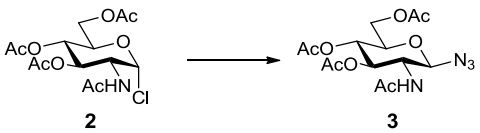
2-acetamido-2-deoxy-3, 4, 6-tri-*O*-acetyl-α-D-glucopyranosyl chloride 2



1 2

N-acetyl glucosamine **1** (10.2 g, 46.0 mmol) was added to stirring acetyl chloride (20 mL) in a flask equipped with a drying tube. The suspension was stirred for 18 hours and formed an amber coloured solution, to which chloroform (80 mL) was added. The resulting solution was poured into ice (80 g) and water (20 mL) with stirring. The organic layer was run into saturated aqueous sodium bicarbonate (80 mL) with neutralisation being completed in a separating funnel. The organic layer was then dried over magnesium sulfate for 15 minutes before being filtered with suction. The yellow solution was then evaporated to dryness *in vacuo* to afford the crude product as a light brown solid, which was subsequently purified by flash chromatography over silica (100% ethyl acetate) to give the product as an off white solid (5.96 g, 36%). R_f 0.69 (100% ethyl acetate). ^{1}H NMR (500 MHz, $CDCl_3$) δ 6.18 (1H, d, *J* = 3.7, H1), 5.83 (1H, d, *J* = 8.5, NH), 5.32 (1H, t, *J* = 10.4, H3), 5.21 (1H, t, *J* = 9.5, H4), 4.55 - 4.51 (1H, m, H2), 4.30 - 4.25 (2H, m, $H6_a$ and H5), 4.13 (1H, dd, *J* = 12.5, 2.0, $H6_b$), 2.10 (3H, s, CH_3), 2.05 (6H, s, 2 x CH_3), 1.98 (3H, s, CH_3). ^{13}C NMR (125 MHz, $CDCl_3$) δ 171.6, 170.7, 170.2, 169.2 (4 x C=O), 93.7 (C1), 71.0 (C5), 70.2 (C3), 67.0 (C4), 61.2 (C6), 53.6 (C2), 23.2, 20.8, 20.6, 20.5 (4 x CH_3). IR υ_{max} (cm^{-1}) neat 3240 (NH), 1737 (C=O), 1641 (C=O). FAB$^+$ MS (*m/z*) calculated for $C_{14}H_{20}NO_8Cl$ 365.09 found $[MNa]^+$ 388.08. m.p. 122-124 °C (lit. 124-125 °C).[203]

2-acetamido-2-deoxy-3, 4, 6-tri-*O*-acetyl-β-D-glucopyranosyl azide 3

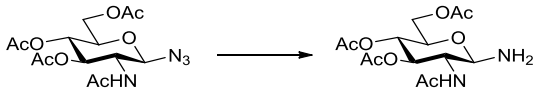


2 3

A mixture of glycosyl chloride **2** (5.75 g, 15.7 mmol), TBAHS (5.35 g, 15.7 mmol) and sodium azide (3.07 g, 47.2 mmol) in dichloromethane (50 mL) with saturated aqueous sodium bicarbonate (50 mL) were stirred vigorously at room temperature

for 4 hours. To the biphasic mixture was added ethyl acetate (125 mL) and the organic layer was separated. The aqueous layer was extracted with ethyl acetate (2 x 100 mL) before the combined organic extracts were washed with saturated aqueous sodium bicarbonate (100 mL) and water (100 mL). It was then washed with saturated aqueous sodium chloride (100 mL) before being dried over magnesium sulfate. The organic mixture was then collected under vacuum and the volatile solvents were removed in a rotary evaporator. The pure product was recrystallised from diethyl ether to afford the products as an off white solid (5.58 g, 95%). R_f 0.41 (100% ethyl acetate). ^{1}H NMR (500 MHz, $CDCl_3$) δ 5.73 (1H, d, *J* = 8.8, NH), 5.24 (1H, t, *J* = 9.6, H3), 5.09 (1H, t, *J* = 9.6, H4), 4.75 (1H, d, *J* = 9.3, H1), 4.26 (1H, dd, *J* = 12.4, 4.8, $H6_a$), 4.16 (1H, dd, *J* = 12.4, 2.0, $H6_b$), 3.92 (1H, q, *J* = 9.2, H2), 3.80 - 3.77 (1H, m, H5), 2.10 (3H, s, CH_3), 2.04 (3H, s, CH_3), 2.03 (3H, s, CH_3), 1.97 (3H, s, CH_3). ^{13}C NMR (125 MHz, $CDCl_3$) δ 171.1, 170.8, 170.5, 169.4 (4 x C=O), 88.5 (C1), 74.1 (C5), 72.2 (C3), 68.1 (C4), 62.0 (C6), 54.2 (C2), 23.3, 20.8, 20.7, 20.6 (4 x CH_3). IR υ_{max} (cm^{-1}) neat 3350 (NH), 2103 (N_3), 1742 (C=O), 1662.6 (C=O). ESI^+ MS (*m/z*) calculated for $C_{14}H_{20}N_4O_8$ 372.13 found $[MNa]^+$ 395.12. m.p. 170-172 °C (lit. 166-168 °C).[267]

2-acetamido-2-deoxy-3, 4, 6-tri-*O*-acetyl-β-D-glucopyranosyl amine 4

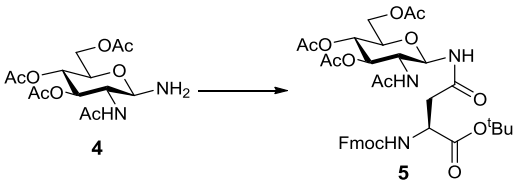


3 **4**

Glycosyl azide **3** (2.01 g, 5.39 mmol) and platinum (IV) oxide (100 mg) were added to a 3 necked flask, which was evacuated and purged with nitrogen gas 3 times. Dry tetrahydrofuran (36 mL) was added and flask was evacuated and purged twice with hydrogen gas. The mixture was stirred under hydrogen at room temperature and monitored by thin layer chromatography. After 3 hours TLC indicated complete consumption of the azide and the reaction mixture was filtered through celite. The product was evaporated to dryness to yield the amine as a grey powder (1.43 g, 97%). ^{1}H NMR (500 MHz, $CDCl_3$) δ 5.66 (1H, d, *J* = 8.8, NH), 5.13 - 5.01 (2H, m, H3, H4), 4.21 (1H, dd, *J* = 12.3, 4.9, $H6_a$), 4.12 - 4.09 (2H, m, H1, $H6_b$), 4.02 (1H, q, *J* = 9.4, H2), 3.65 - 3.62 (1H, m, H5), 2.09 (3H, s, CH_3), 2.04 (3H, s, CH_3), 2.03 (3H, s, CH_3), 1.99 (3H, s, CH_3). ^{13}C NMR (125 MHz, $CDCl_3$) δ 171.7, 170.9, 170.8, 169.4 (4 x C=O),

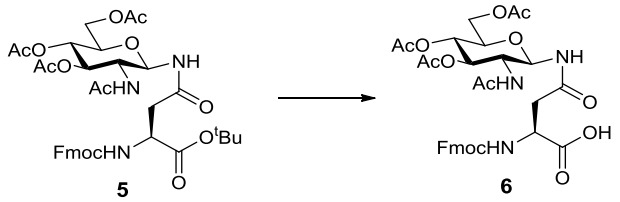
86.5 (C1), 73.5 (C3), 72.9 (C5), 68.5 (C4), 62.5 (C6), 55.0 (C2), 23.4, 20.9, 20.8, 20.7 (4 x CH_3). IR υ_{max} (cm^{-1}) neat 3400 (NH), 3343 (NH), 1736 (C=O), 1659 (C=O). ESI^+ MS (*m/z*) calculated for $C_{14}H_{22}N_2O_8$ 346.14 found $[MNa]^+$ 369.13. m.p. 155-157 °C (lit. 159 °C).[203]

_N_-α-fluorenylmethoxycarbonyl-_N_-β-(2-_N_-acetylamido-2-deoxy-β-D-glucopyranosyl)-_tert_-butyloxy-L-asparagine 5



Glycosyl amine **4** (500 mg, 1.44 mmol), Fmoc-Asp-O^{t}Bu (594 mg, 1.44 mmol) and *N*-ethoxycarbonyl-2-ethoxy-1, 2-dihydroquinoline (357 mg, 1.44 mmol) were stirred in dichloromethane (45 mL) under nitrogen at room temperature for 4 hours. The crude mixture was concentrated *in vacuo* to afford a grey solid to which a mixture (1:1) of hot ethyl acetate and petroleum sprit (20 mL) was added. The suspension was filtered and the product was collected as a grey solid which was dried under vacuum to yield 821 mg (77%). R_f 0.5 (100% ethyl acetate). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.75 (2H, d, *J* = 7.5, 2 x Ar-H), 7.59 (2H, d, *J* = 7.4, 2 x Ar-H), 7.39 (2H, t, *J* = 7.3, 2 x Ar-H), 7.31 - 7.28 (3H, m, 2 x Ar-H, NH), 6.21 (1H, d, *J* = 8.3, NH), 5.96 (1H, d, *J* = 8.6, NH), 5.11 - 5.03 (3H, m, H1, H3, H4), 4.55 - 4.49 (1H, m, CH_α), 4.43 - 4.40 (1H, m, CH), 4.30 - 4.17 (3H, m, 2 x CH, $H6_a$), 4.14 - 4.12 (2H, m, CH, H2), 3.74 (1H, dd, *J* = 9.7, 2.0, H5), 2.84 (1H, dd, *J* = 16.4, 4.2, CH_β), 2.70 (1H, dd, *J* = 16.3, 3.9, CH_β), 2.06 (3H, s, CH_3), 2.05 (3H, s, CH_3), 2.04 (3H, s, CH_3), 1.94 (3H, s, CH_3), 1.44 (9H, s, $C(CH_3)_3$). ^{13}C NMR (125 MHz, $CDCl_3$) δ 172.4, 172.0, 171.2, 170.8, 170.1, 169.4, 156.2 (7 x C=O), 144.0, 141.3 (4 x Ar-*q*C), 127.8, 127.2, 125.2, 120.0 (8 x Ar-C), 82.3 ($\underline{C}(CH_3)$), 80.2, 73.6, 73.0, 67.7 (4 x CH), 67.2 (CH_2), 61.8 (CH_2), 60.5, 53.4 (2 x CH), 51.1 (CH_α), 47.2 (CH), 38.1 ($CH_{2\beta}$), 28.0 ($C(\underline{C}H_3)$), 23.1, 20.8, 20.7, 20.6 (4 x CH_3). IR υ_{max} (cm^{-1}) neat 3308 (NH), 1738 (C=O), 1658 (C=O). FAB^+ MS (*m/z*) calculated for $C_{37}H_{45}N_3O_{13}$ 739.30 found $[MNa]^+$ 762.29. m.p. 208-211 °C (lit. 214-215 °C).[204]

_N_-α-fluorenylmethoxycarbonyl-_N_-β-(2-_N_-acetylamido-2-deoxy-β-D-glucopyranosyl)-L-asparagine 6



Glycoamino acid **(5)** (500 mg, 680 µmol) was treated with 95% trifluoroacetic acid (10 mL) and stirred at room temperature for 3 hours. TLC analysis indicated completion of the reaction and the solvent was removed *in vacuo* to yield the product as a grey solid (462 mg, 100%). R_f 0.18 (100% ethyl acetate). ^{1}H NMR (500 MHz, DMSO-d_6) δ 8.57 (1H, d, *J* = 9.2, NH), 7.87 (3H, d, *J* = 8.0, 2 x Ar-H, NH), 7.69 (2H, d, *J* = 7.4, 2 x Ar-H), 7.45 (1H, d, *J* = 11.5, NH), 7.40 (2H, t, *J* = 7.4, 2 x Ar-H), 7.31 (2H, t, *J* = 7.4, 2 x Ar-H), 5.17 (1H, t, *J* = 9.9, H1), 5.08 (1H, t, *J* = 9.9, H3), 4.80 (1H, t, *J* = 9.8, H4), 4.38 (1H, q, *J* = 8.7, CH_α), 4.30 - 4.15 (4H, m, $H6_a$, CH, CH_2), 3.93 (1H, dd, *J* = 10.7, 1.6, $H6_b$), 3.86 (1H, q, *J* = 9.7, H2), 3.80 (1H, dd, *J* = 9.9, 1.5, H5), 2.64 (1H, dd, *J* = 16.3, 5.4, CH_β), 2.48 (1H, dd, overlapped by DMSO signal CH_β), 1.98 (3H, s, CH_3), 1.95 (3H, s, CH_3), 1.89 (3H, s, CH_3), 1.70 (3H, s, CH_3). ^{13}C NMR (125 MHz, DMSO-d_6) δ 173.0, 171.5, 170.0, 169.8, 169.5, 169.3, 155.8 (7 x C=O), 143.8, 140.7 (4 x Ar-*q*C), 127.6, 127.1, 125.2, 120.1 (8 x Ar-C), 78.1 (C1), 73.4 (C3), 72.3 (C5), 68.4 (C4), 65.7 (C6), 52.1 (C2), 50.0 (CH), 46.6 (CH), 36.8 (CH_β), 22.6, 20.5, 20.4, 20.3 (4 x CH_3). IR υ_{max} (cm^{-1}) neat 3308 (NH), 1743 (C=O), 1693 (C=O), 1659 (C=O). ESI^+ MS (*m/z*) calculated for $C_{33}H_{37}N_3O_{13}$ 683.23 found $[MNa]^+$ 706.22. m.p. 209-211 °C (lit. 210 °C).[204]

General peptide synthesis

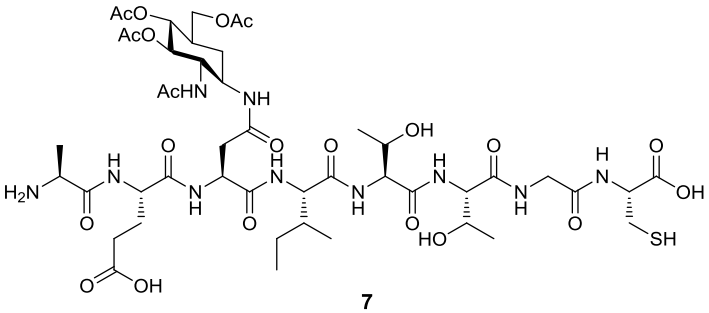
Protected amino acids and resins for Fmoc peptide synthesis were purchased from Novabiochem or AGTC Bioproducts as: Fmoc-Ala-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Asp(O^tBu)-OH, Fmoc-Cys(Trt)-OH, Fmoc-Cys(Acm)-OH Fmoc-Glu(O^tBu)-OH, Fmoc-Gln(Trt)-OH, Fmoc-Gly-OH, Fmoc-His(Trt)-OH, Fmoc-Ile-OH, Fmoc-Leu-OH, Fmoc-Lys(Boc)-OH, Fmoc-Phe-OH, Fmoc-Pro-OH, Fmoc-Ser(O^tBu)-OH, Fmoc-Thr(O^tBu)-OH, Fmoc-Trp(Boc)-OH, Fmoc-Tyr(O^tBu)-OH, Fmoc-Val-OH unless specified otherwise.

For manual syntheses, steps were performed in sintered 12 mL ISOLUTE SPE syringe tubes (Biotage). For a 0.05 mmol scale synthesis, couplings were performed using 10 equivalents of the amino acid dissolved in DMF (1.1 mL), 10 equivalents of HBTU/HOBt (1.1 mL from a 0.45 M stock in DMF) and 17.5 equivalents of DIPEA (150 µL) for at least 4 hours with agitation at 430 RPM. Deprotection was carried out with 3.0 mL of 20% piperidine in DMF (v/v) for 15 minutes. Washings between each step were carried out with DMF (3 x 5.0 mL) and DCM (2 x 5.0 mL). Automated peptide synthesis was carried out on an ABI 433A synthesiser (Applied Biosystems) using the FastMoc protocol. Resin loading could be determined by quantification of the Fmoc-fulvene adduct at 290 nm. Deprotection of a small but accurately weighed amount of resin was conducted with 20% piperidine in DMF (3.0 mL, v/v). The fulvene adduct was quantified by the equation:

$$\text{Resin loading (mmol/g)} = \frac{\text{Abs}_{290}}{1.75 \times \text{resin weight (mg)}}$$

Peptides were cleaved from the resin using the Reagent K cocktail (TFA/phenol/water/thioanisole/EDT, 82.5:5:5:5:2.5, v/w/v/v/v) for 4-5 hours unless specified otherwise. After filtration of the resin, peptides were precipitated in cold diethyl ether (60 mL). Crude peptides were obtained by centrifugation at 3000 RPM for 15 minutes at 4 °C. The ethereal layer was decanted and the crude peptide was resuspended in fresh diethyl ether before being pelleted again by centrifugation. Crude peptide pellets were typically resuspended in the minimum volume of water with up to 35% acetonitrile before purification by (RP)HPLC.

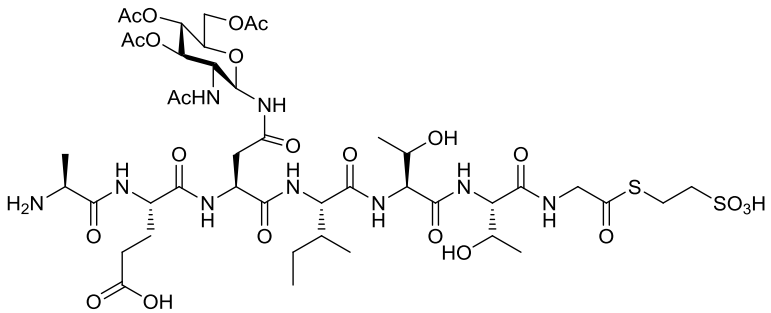
AEN(GlcNAc(OAc)$_3$)ITTGC (EPO 22-28 Cys) 7



7

Glycopeptide **7** was assembled on Fmoc-Cys(Trt)-NovaSyn TGT resin (0.20 mmol/g loading) (500 mg, 0.10 mmol) according to the general peptide synthesis protocols. An exception to this was the glycoamino acid **6** coupling, for which only 5 equivalents were employed. The peptide was cleaved with TFA/EDT/Water (95:2.5:2.5, v/v/v, 4.0 mL) for 5 hours, filtered and washed with neat TFA (1.0 mL) and the crude peptide was precipitated. The peptide was purified using semi-preparative (RP)HPLC method 1, (t_r 22.3 min). The collections were lyophilised to afford **7** as a fluffy white solid (30.2 mg, 53%). ESI$^+$ LC-MS (*m/z*) calculated for $C_{45}H_{72}N_{10}O_{22}S$ 1136.45 found $[MH]^+$ 1137.25.

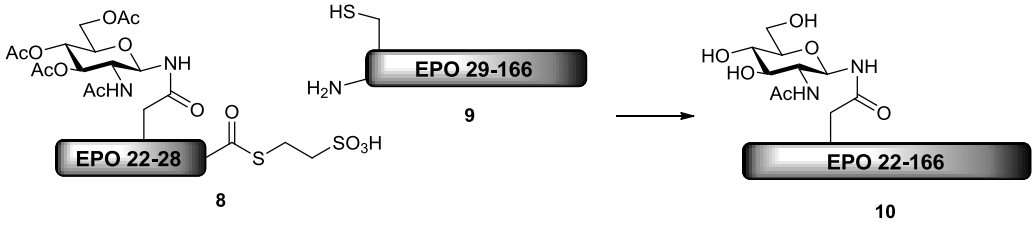
AEN(GlcNAc(OAc)$_3$)ITTG-SCH$_2$CH$_2$SO$_3$H (EPO 22-28 thioester) 8



8

Glycopeptide **7** (5 mg) was dissolved in sodium phosphate buffer (0.1 M, pH 5.8), containing MESNa (10% v/v) and TCEP (0.5% v/v). The samples were agitated at 500 RPM in 5 x 1.5 mL microcentrifuge tubes in an Eppendorf Thermomixer at 55 °C for 48 hours. The reaction was monitored by analytical (RP)HPLC at set intervals of 0, 6, 24 and 48 hours before purification by semi-preparative (RP)HPLC method 1 (t_r 20.8 min) to yield a white solid (1 mg, 20%). ESI$^+$ LC-MS (*m/z*) calculated for $C_{44}H_{71}N_9O_{23}S_2$ 1157.41 found $[MH]^+$ 1158.39.

EPO22-166 (GlcNAc at residue 24) 10

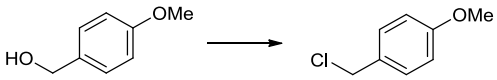


EPO 29-166 **9** (2.5 mg) was dissolved in 6 M Gu.HCl containing 0.3 M sodium phosphate buffer (pH 5.8, 0.5 mL). To the protein solution was added 20 µL of 1 M TCEP (neutralised to pH 7.0 with NaOH) and 50 µL of 1 M MPAA (neutralised to pH 7.0 with NaOH) to give final concentrations of 40 mM and 100 mM respectively. The resulting mixture was added to glycopeptide thioester **8** (2 mg) in a microcentrifuge tube and shaken at 500 RPM under N_2 at room temperature on an Eppendorf Thermomixer. After 3 hours, LC-MS indicated consumption of the protein EPO 29-166 and formation of the ligation product. The reaction mixture was subsequently diluted by the addition of cold distilled water (8.0 mL) and stored at 4 °C overnight. A white precipitate formed and was collected by centrifugation at 3000 RPM for 10 minutes. The process was repeated, resuspending the protein in cold distilled water (8.0 mL) allowing precipitation at 4 °C overnight. The crude protein was then collected by centrifugation and subsequently treated with hydrazine hydrate (5%) and DTT (50 mM) in 6 M Gu.HCl (3.0 mL) at room temperature with gentle agitation. After 1 hour, LC-MS indicated complete deacetylation of the starting material and the protein **10** was isolated through precipitation as before. ESI^+ LC-MS (*m/z*) average mass calculated for $C_{720}H_{1164}N_{204}O_{213}S_3$ 16182.38 found $[MH_{17}]^{17+}$ 952.81.

Product protein sequence:

AEN(GlcNAc)ITTGCCAEHCSLNENITVPDTKVNFYAWKRLEVGQQAVEVWQGLALLSEAVLR GQALLVKSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKEAIKPPDAASAAPLRTITADTFRKL FRVYSNFLRGKLKLYTGEACRTGDR

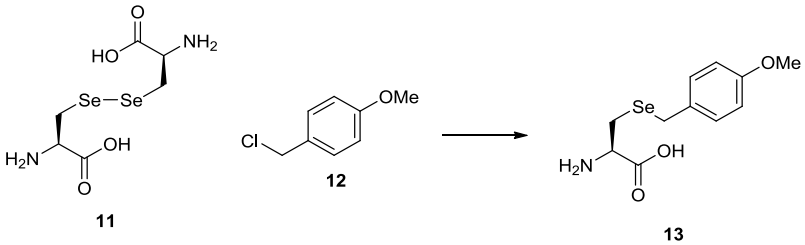
Para-methoxybenzyl chloride 12



12

Para-methoxybenzyl alcohol (13.82 g, 0.1 mol) was stirred in dry diethyl ether (150 mL) at 0 °C. To the stirred solution was added thionyl chloride (15 mL, 0.2 mol) dropwise and after complete addition was stirred vigorously at room temperature. After 3 hours, TLC indicated completion of the reaction and saturated sodium bicarbonate solution was added until the pH reached 8. The mixture was extracted with chloroform (3 x 100 mL) and washed with water (100 mL) followed by brine (100 mL). The organic phase was dried over magnesium sulfate before concentration *in vacuo* to yield a clear oil (11.24 g, 72%). R_f 0.82 (3:1 petroleum spirit/ethyl acetate). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.34 (2H, d, *J* = 8.6, Ar-H), 6.91 (2H, d, *J* = 8.6, Ar-H), 4.60 (2H, s, CH_2), 3.84 (3H, s, CH_3). ^{13}C NMR (125 MHz, $CDCl_3$) δ 130.5 (Ar-CH), 114.5 (Ar-CH), 55.7 (CH_3), 46.7 (CH_2). ESI^+ MS (*m/z*) calculated for C_8H_9OCl 156.03 found $[M\text{-}Cl]^+$ 121.03.

Para-methoxybenzyl selenocysteine 13

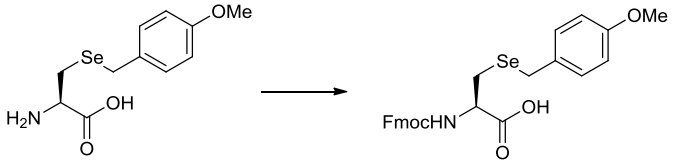


11 12 13

Selenocystine **11** (480 mg, 1.44 mmol) was stirred in 0.5 M NaOH (1.3 mL) in an ice bath. Sodium borohydride (450 mg, 11.9 mmol) was added in small portions cautiously and the mixture was stirred until the yellow colour had disappeared. Further 2 M NaOH (3.9 mL) was added, followed by dropwise addition of *para*-methoxybenzyl chloride **12** (1.22 g, 7.82 mmol) and the resulting mixture was stirred vigorously at 4 °C for 4 hours. The mixture was acidified with concentrated HCl and the precipitated product was collected by filtration. The crude product was washed with diethyl ether (25 mL) and water (5.0 mL) before being dried in a desiccator overnight to yield a white solid (620 mg, 75%). ^{1}H NMR (600 MHz,

DMSO-d_6) δ 7.25 (2H, d, *J* = 8.6 Ar-H), 6.86 (2H, d, *J* = 8.6, Ar-H), 6.64 (2H, br, NH_2), 3.97 (1H, br, CH), 3.83 - 3.81 (2H, m, CH_2), 3.73 (3H, s, CH_3), 2.88 (2H, m, CH_2). ESI^+ MS (*m/z*) calculated for $C_{11}H_{15}NO_3Se$ 288.20 found $[MH]^+$ 289.16. m.p. 169-172 °C (lit. 173-176 °C).[268]

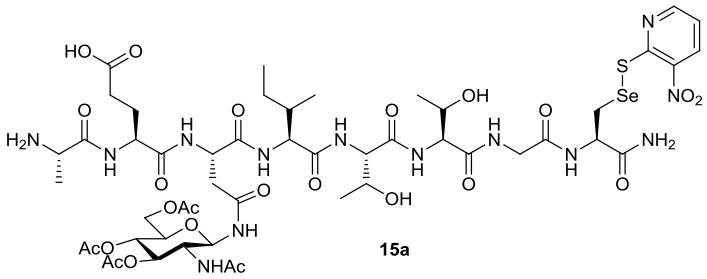
_N_-α-fluorenylmethoxycarbonyl _para_-methoxybenzyl selenocysteine 14



13 **14**

Fmoc succinimide (724 mg, 2.15 mmol) was dissolved in acetonitrile (3.0 mL) and added to an ice cooled suspension of *para*-methoxybenzyl selenocysteine **13** in water (4.0 mL) and triethylamine (300 µL, 2.15 mmol). Further triethylamine (300 µL) was added to maintain a pH above 8 and the mixture was stirred at room temperature for 2 hours. The mixture was acidified with 1 M HCl dropwise, and the product was extracted with ethyl acetate (50 mL). The organic extract was washed with 1 M HCl (2 x 30 mL) and brine (30 mL) before being dried over magnesium sulfate. After concentration *in vacuo* the crude product was triturated with a minimum volume of diethyl ether and dried under vacuum to yield an off-white solid (827 mg, 75%). R_f 0.2 (19:1 chloroform/methanol). ^1H NMR (500 MHz, $CDCl_3$) δ 7.90 (2H, d, *J* = 7.5, Ar-H), 7.82 - 7.69 (3H, m, 2 x Ar-H, NH), 7.42 (2H, t, *J* = 7.5, Ar-H), 7.38 - 7.27 (2H, m, Ar-H), 7.22 (2H, d, *J* = 8.6, Ar-H), 6.84 (2H, d, *J* = 8.7, Ar-H), 4.41 - 4.13 (3H, m, CH_2, CH), 3.81 (2H, s, CH_2), 2.84 (1H, dd, *J* = 12.7, 4.7, CH_β), 2.74 (1H, dd, *J* = 12.6, 9.4, CH_β). ESI^+ MS (*m/z*) calculated for $C_{26}H_{25}NO_5Se$ 511.09 found $[MNa]^+$ 534.08. m.p. 135-138 °C (lit. 138-140 °C).[268]

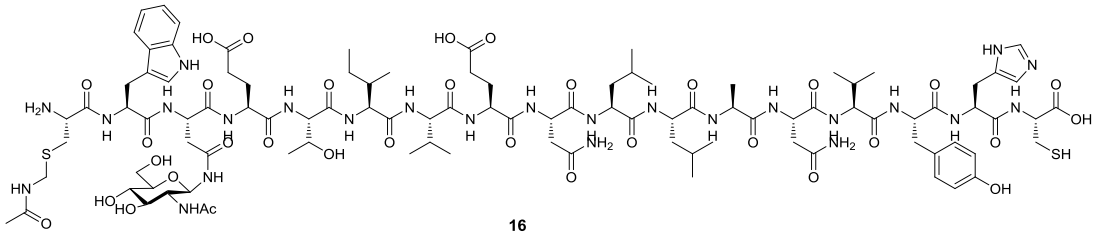
AEN(GlcNAc(OAc)$_3$)ITTGU(Npys)-NH$_2$ (EPO 22-28 Sec) 15a



15a

Glycopeptide **15a** was assembled on Rink amide MBHA resin (0.64 mmol/g loading) (156 mg, 0.10 mmol) according to the general peptide synthesis protocols. An exception to this was the coupling of selenocysteine **14** and the glycoamino acid **6**, for which only 5 equivalents were employed. A small sample of the resin (40 mg) was treated with TFA/thioanisole (97.5:2.5, v/v, 2.0 mL) in the presence of 2,2'-dithiobis(5-nitropyridine) (20 mg, 65 µmol, 1.3 equivalents) for 5 hours, filtered and washed with neat TFA (1.0 mL) and the crude peptide was precipitated. The peptide was purified using semi-preparative (RP)HPLC method 1, (t_r 29.0 min). The collections were lyophilised to afford **15a** as a fluffy white solid (1.5 mg, 8%). ESI$^+$ LC-MS (*m/z*) calculated for $C_{50}H_{75}N_{13}O_{23}SSe$ 1337.23 found $[MH]^+$ 1338.48.

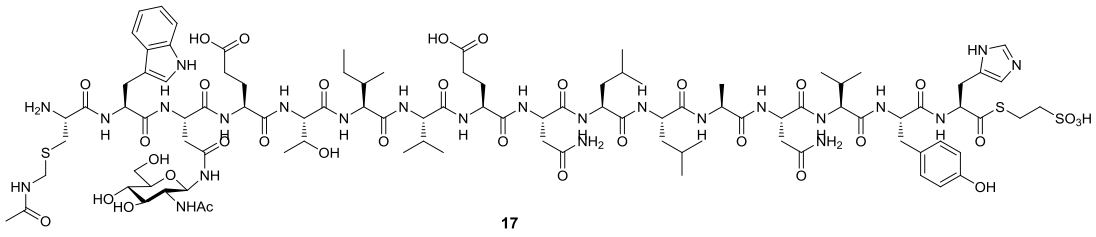
C(Acm)WN(GlcNAc)ETIVENLLANVYHC (IFNβ 78-93 Cys) 16



16

Fmoc-Cys(Trt)-NovaSyn TGT resin (0.19 mmol/g loading) (263 mg, 0.05 mmol scale) was elongated to IVENLLANVYHC by automated synthesis. The pseudoproline Fmoc-Glu(O^{t}Bu)-Thr(psi^{Me,Me}pro)-OH, glycoamino acid **6**, tryptophan and cysteine (Acm protected) were coupled manually using 5 equivalents. Peptide cleavage was conducted with Reagent K (4.0 mL) for 5 hours, and precipitated in ether. The crude peptide was freeze dried before being resuspended in aqueous hydrazine (3.0 mL, 5% v/v) with DTT (50 mM). After stirring at room temperature for an hour, the mixture was subjected to semi-preparative (RP)HPLC method 1 (t_r 36.3 min), yielding the glycopeptide **16** as a white fluffy solid (7 mg, 6%). ESI$^+$ LC-MS (*m/z*) calculated for $C_{100}H_{151}N_{25}O_{33}S_2$ 2294.03 found $[MH_2]^{2+}$ 1148.73.

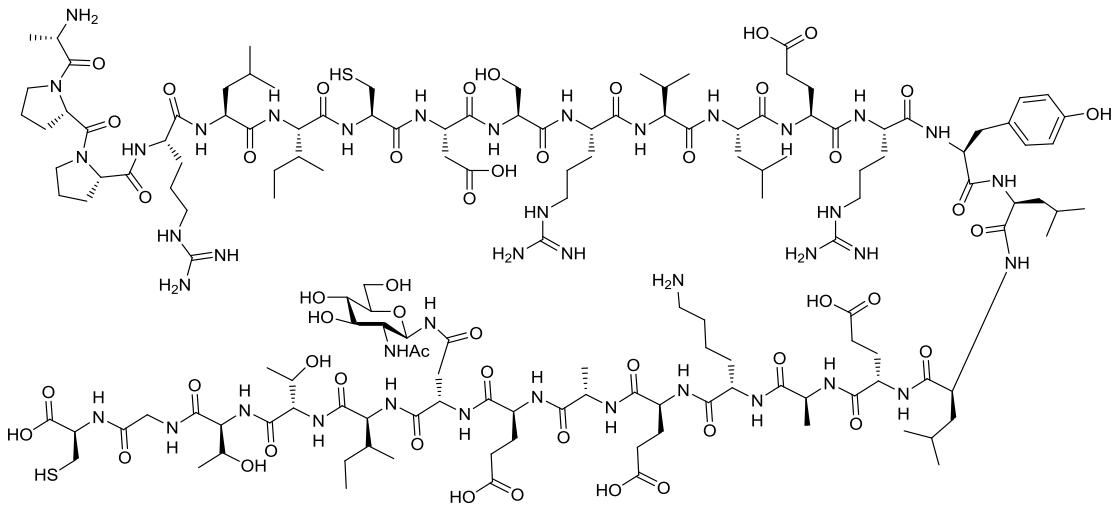
C(Acm)WN(GlcNAc)ETIVENLLANVYH-SCH$_2$CH$_2$SO$_3$H (IFNβ 78-93 thioester) 17



17

Glycopeptide **16** (7 mg) was dissolved in 6 M Gu.HCl (4.0 mL) containing sodium phosphate buffer (0.1 M, pH 5.8), MESNa (10% v/v) and TCEP (0.5% v/v). The samples were agitated at 600 RPM in 4 x 1.5 mL microcentrifuge tubes in an Eppendorf Thermomixer at 55 °C for 48 hours. The reaction was monitored by LC-MS at set intervals 0, 24 and 48 hours. The product was purified by semi-preparative (RP)HPLC method 1 (t_r 36.4 min) to yield a white solid (<1 mg). ESI$^+$ LC-MS (*m/z*) calculated for $C_{99}H_{150}N_{24}O_{34}S_3$ 2314.99 found $[MH_2]^{2+}$ 1159.16.

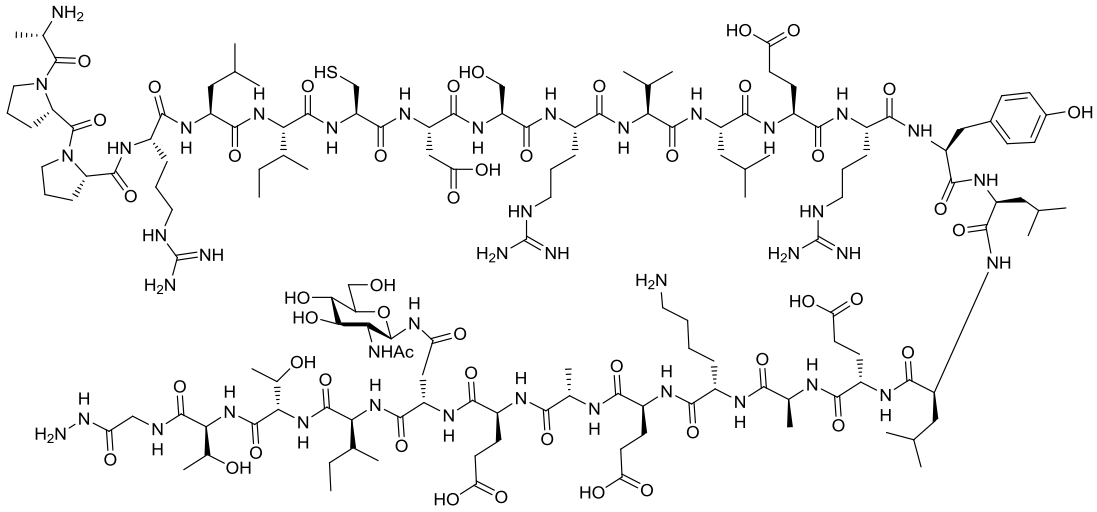
APPRLICDSRVLERYLLEAKEAEN(GlcNAc)ITTGC (EPO 1-28 Cys) 18



18

Glycopeptide **18** was assembled manually on Fmoc-Cys(Trt)-NovaSyn TGT resin (0.20 mmol/g loading) (250 mg, 0.05 mmol) according to the general peptide synthesis protocols. An exception to this was the glycoamino acid **6** coupling, for which only 5 equivalents were employed. After the glycoamino acid coupling, the remainder of the peptide was completed by automated synthesis. The peptide was cleaved with TFA/EDT/Water (95:2.5:2.5, v/v/v, 4.0 mL) for 5 hours, filtered and washed with neat TFA (1.0 mL) and the crude peptide was precipitated. The peptide was purified using semi-preparative (RP)HPLC method 1, (t_r 32.5 min). The collections were lyophilised to afford the glycopeptide as a fluffy white solid (17.2 mg, 10%). Deacetylation was conducted in aqueous hydrazine (5.0 mL, 5% v/v) with DTT (50 mM) before purification of **18** by semi-preparative (RP)HPLC method 1, (t_r 31.5 min) yielding 3.4 mg (35%). ESI^+ LC-MS (m/z) calculated for $C_{148}H_{247}N_{41}O_{50}S_2$ 3462.75 found $[MH_4]^{4+}$ 867.04.

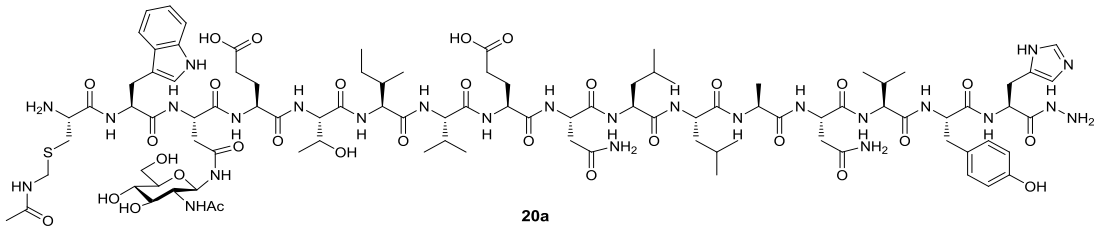
APPRLICDSRVLERYLLEAKEAEN(GlcNAc)ITT-NHNH$_2$ (EPO 1-28 hydrazide) 19



19

Glycopeptide **18** (1 mg) was dissolved in 3 M Gu.HCl (1.0 mL) containing sodium phosphate buffer (0.1 M, pH 5.8), MESNa (10% v/v), TCEP (0.5% v/v) and hydrazinium acetate. Hydrazinium acetate was prepared as a 50% w/v stock in 7 M Hydrochloric acid, added to a final concentration of 5%. The reaction mixture was agitated at 600 RPM in a 1.5 mL microcentrifuge tube in an Eppendorf Thermomixer at 55 °C for 48 hours. The reaction was monitored by LC-MS at set intervals 0, 24 and 48 hours. The product was purified by semi-preparative (RP)HPLC method 1 (t_r 31.5 min) to yield a white solid (<1 mg). ESI$^+$ LC-MS (*m/z*) calculated for $C_{145}H_{244}N_{42}O_{48}S$ 3373.76 found $[MH_4]^{4+}$ 844.77.

C(Acm)WN(GlcNAc)ETIVENLLANVYH-NHNH$_2$ (IFNβ 78-93 hydrazide) 20a

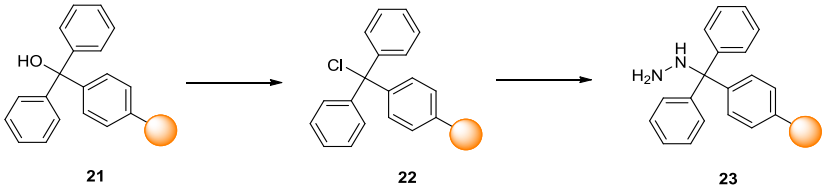


20a

Glycopeptide **16** (2 mg) was dissolved in 6 M Gu.HCl (1.0 mL) containing sodium phosphate buffer (0.1 M, pH 5.8), MESNa (10% v/v), TCEP (0.5% v/v) and hydrazinium acetate. Hydrazinium acetate was prepared as a 50% w/v stock in 7 M

Hydrochloric acid, added to a final concentration of 5%. The reaction mixture was agitated at 600 RPM in 4 x 1.5 mL microcentrifuge tubes in an Eppendorf Thermomixer at 45 °C for 120 hours. The reaction was monitored by LC-MS at set intervals 0, 24, 48, 72 and 120 hours. The product was purified by semi-preparative (RP)HPLC method 1 (t_r 36.3 min) to yield a white solid (<1 mg). ESI^+ LC-MS (*m/z*) calculated for $C_{97}H_{148}N_{26}O_{31}S$ 2205.05 found $[MH_3]^{3+}$ 736.04.

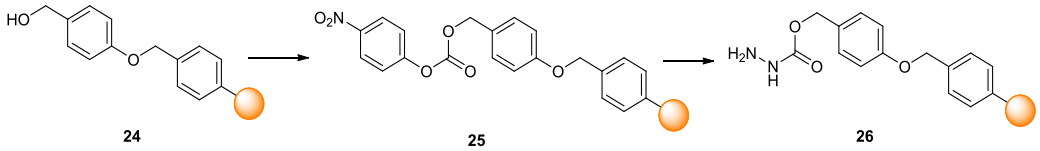
NovaSyn TGT hydrazine resin 23



Novasyn TGT alcohol resin **21** (500 mg, 0.105 mmol) (0.21 mmol/g loading) was washed with DMF (2 x 5.0 mL) , DCM (3 x 5.0 mL) and toluene (3 x 5.0 mL) before being transferred to a round bottomed flask equipped with a reflux condenser. Dry toluene was added until the resin was covered, followed by the addition of acetyl chloride (1.0 mL) and the mixture was heated at 70 °C for 3 hours. The resin was transferred to a sintered syringe tube and washed with dry toluene (3 x 5.0 mL) and dry DCM (3 x 5.0 mL). The chlorinated resin **22** was immediately treated with a solution of hydrazine hydrate (20 µL) in dry THF (5.0 mL) and stirred at room temperature for 2 hours. The resin was subsequently washed with DCM/MeOH/DIPEA (17:2:1, v/v/v, 3 x 5.0 mL), DCM (3 x 5.0 mL), DMF (2 x 5.0 mL) and DCM (2 x 5.0 mL). **23** was employed in peptide synthesis without any further purification.

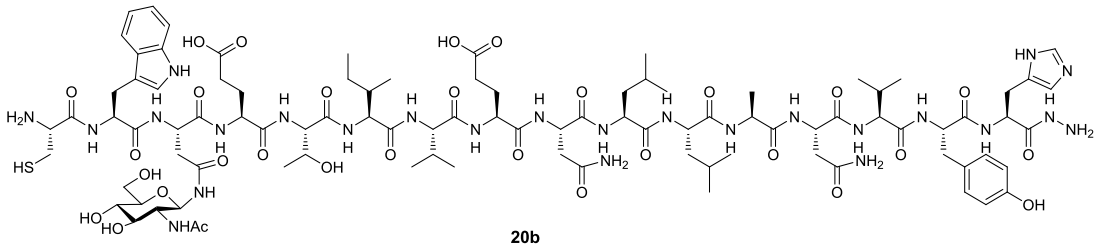
Alternatively, trityl chloride resin (1.4 mmol/g loading) was purchased directly from Novabiochem and functionalised as follows: Trityl chloride resin (500 mg, 0.7 mmol) was treated with a solution of hydrazine hydrate (140 µL, 2.8 mmol) in dry THF (5.0 mL) and stirred at room temperature. The resin was purified according to the same procedure used in the previously described method for **23**.

Hydrazine carboxylate Wang resin 26



Wang resin **24** (1.0 g, 1.3 mmol) (1.3 mmol/g loading) and *N*-methyl morpholine (467 µL, 4.25 mmol) were stirred at 0 °C in dry DCM (25 mL). 4-nitrophenyl chloroformate (840 mg, 3.9 mmol) was added, and the reaction mixture was allowed to warm to room temperature before being stirred overnight. The following day, the resin was collected by filtration and washed with DCM (2 x 5.0 mL), DMF (3 x 5.0 mL), MeOH, (3 x 5.0 mL) and DCM (3 x 5.0 mL). The resin **25** was allowed to dry in a vacuum desiccator overnight. Assuming quantitative conversion, **25** (1.3 mmol) was treated with hydrazine hydrate (442 µL, 8.9 mmol) in dry DCM/DMF (2:3, v/v, 50 mL) at 0 °C. The reaction was warmed to room temperature and stirred overnight under nitrogen. The resin was collected by filtration and washed with DCM (2 x 5.0 mL), DMF (3 x 5.0 mL), MeOH, (3 x 5.0 mL) and DCM (3 x 5.0 mL) before being dried in a vacuum desiccator. **26** was employed in peptide synthesis without any further purification.

CWN(GlcNAc)ETIVENLLANVYH-NHNH$_2$ (IFNβ 78-93 hydrazide) 20b



20b

Glycopeptide **20b** was manually synthesised on hydrazine resin **26** according to the general peptide synthesis protocols. An exception to this was the glycoamino acid **6** coupling, for which only 5 equivalents were employed. After cleavage with Reagent K (4.0 mL), the product was purified by preparative (RP)HPLC (t_r 36.4 min) yielding a fluffy white solid (6 mg, 6%). ESI$^+$ LC-MS (*m/z*) calculated for $C_{94}H_{143}N_{25}O_{30}S$ 2134.02 found $[MH_2]^{2+}$ 1068.77.

8.3 Chapter 3 experimental details

Cloning of IFNβ 1-77 intein-CBD fusion

Primers were designed to allow the cloning of the IFNβ 1-77 gene into vector pTYB1 between the NdeI/SapI or KpnI restriction sites.

Forward primer:
GGTGGTCATATGAGCTACAACTTG

Reverse primer (SapI cloning):
GGTGGTTGCTCTTCCGCAGCCGCTAGATGAATCTTG

Reverse primer (KpnI cloning):
GGTGGTGGTACCCTTGGCAAAGCAGCCGCTAGATGAATCTTG

The PCR mix was made up as follows:

Forward primer (10 µM stock) 5 µL
Reverse primer (10 µM stock) 5 µL
10 x ThermoPol reaction buffer 5 µL
dNTP mix (10 mM stock) 5 µL
Template DNA (IFN 1-166 pKYB1) 1 µL
Deep Vent Polymerase (New England Biolabs) 1 µL
Distilled sterile water 28 µL

PCR was conducted with the following cycling conditions:

Step	Temp	Time
Initial denaturation	95 °C	2 minutes
30 cycles	95 °C	30 seconds
	54 °C	30 seconds
	72 °C	1 minute
Final extension	72 °C	5 minutes
Hold	4 °C	

DNA ligation for IFNβ 1-77 intein-CBD fusion

The vector and PCR product were treated with NdeI/SapI or KpnI to generate sticky ends. DNA ligation was performed using a Rapid DNA Ligation Kit (Roche) according to the manufacturer's instructions.

Expression and purification of IFNβ 1-77 intein-CBD fusion 27

The relevant plasmid was transformed into BL21(DE3) cells and a colony was used to seed LB medium (200 mL) containing ampicillin (100 μg/ml) and grown at 37 °C overnight. The following day, the overnight culture was used to seed fresh LB medium (4 x 500 mL) containing ampicillin (100 μg/ml) and grown to an optical density (OD_{600}) of ~ 0.6. At this point, protein expression was induced by the addition of IPTG at a final concentration of 1 mM. Expression was carried out at 30 °C for 5 hours. Cells were harvested by centrifugation at 8000 RPM for 15 minutes at 4 °C (Beckman Coulter Avanti J-26XP). The combined cell pellets were resuspended in binding buffer (500 mM NaCl, 20 mM Tris pH 8.0, 0.005% PMSF w/v, 50 mL) and sonicated on ice for 10 bursts of 30 seconds, with 30 seconds of cooling time in between. Cell debris was pelleted at 10000 RPM for 15 minutes at 4 °C and the supernatant was loaded onto chitin resin (2 x 10 mL) which had been pre-equilibrated with 10 volumes of binding buffer). The columns were subsequently washed with a further 10 volumes of binding buffer, before being washed with 3 volumes of binding buffer containing 50 mM MESNa. Column flow was stopped and the cleavage was allowed to proceed over 24 hours at 4 °C. After unsuccessful attempts to isolate the product, refolding of the insoluble fusion was attempted according to New England Biolabs IMPACT manual (described in section 3.2.2).

Cloning of IFNβ 1-77 thioredoxin intein-CBD fusion

Primers were designed to allow the cloning of the IFNβ 1-77 intein fusion gene into vector pET32XaLIC for ligation independent cloning.

Forward primer:
GGTATTGAGGGTCGCATGAGCTACAACTTGCTTGGATTCC

Reverse primer:
AGAGGAGAGTTAGAGCCTTATTGAAGCTGCCACAA

The PCR mix was made up as follows:

Forward primer (10 µM stock) 5 µL
Reverse primer (10 µM stock) 5 µL
10 x ThermoPol reaction buffer 5 µL
dNTP mix (10 mM stock) 5 µL
Template DNA (IFN 1-77 pTYB1) 1 µL
Deep Vent Polymerase (New England Biolabs) 1 µL
Distilled sterile water 28 µL

PCR was conducted with the following cycling conditions:

Step	Temp	Time
Initial denaturation	95 °C	2 minutes
30 cycles	95 °C	30 seconds
	64 °C	30 seconds
	72 °C	2.5 minutes
Final extension	72 °C	10 minutes
Hold	4 °C	

Ligation independent cloning for IFNβ 1-77 thioredoxin intein fusion

Ligation independent cloning was carried out according to the manufacturer's instructions (Novagen).

Expression of IFNβ 1-77 thioredoxin intein-CBD fusion 28

The protein was expressed in BL21(DE3) cells as before for **27** (2 x 500 mL) at 30 °C for 5 hours. After collection of the cells by centrifugation, the whole cell lysate and soluble fraction were treated with Bug Buster protein extraction reagent (Novagen) and analysed by SDS-PAGE.

Cloning of His_{10} tagged IFNβ 1-77 with C-terminal Gly-Cys

Primers were designed to allow the cloning of the IFNβ 1-77 terminated with a Gly-Cys motif into vector pET16b for expression with an N-terminal polyhistidine tag.

Forward primer:
GGTGGTCATATGAGCTACAACTTG

Reverse primer:
GGTGGTGGATCCTTAACAGCCGCTAGATGAATCTTG

The PCR mix was made up as follows:

Forward primer (10 µM stock) 5 µL
Reverse primer (10 µM stock) 5 µL
10 x ThermoPol reaction buffer 5 µL
dNTP mix (10 mM stock) 5 µL
Template DNA (IFN 1-166 pKYB1) 1 µL
Deep Vent Polymerase (New England Biolabs) 1 µL
Distilled sterile water 28 µL

PCR was conducted with the following cycling conditions:

Step	Temp	Time
Initial denaturation	95 °C	2 minutes
30 cycles	95 °C	30 seconds
	54 °C	30 seconds
	72 °C	1 minute
Final extension	72 °C	5 minutes
Hold	4 °C	

DNA ligation for His_{10} tagged IFNβ 1-77 with C-terminal Gly-Cys

The vector and PCR product were treated with NdeI/BamHI to generate sticky ends. DNA ligation was performed using a Rapid DNA Ligation Kit (Roche) according to the manufacturer's instructions.

Expression and purification of His_{10} tagged IFNβ 1-77 with C-terminal Gly-Cys 29

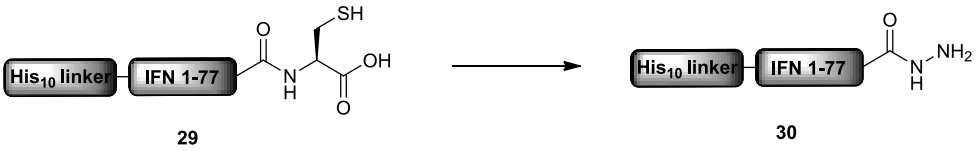
The relevant plasmid was transformed into BL21(DE3) cells and a colony was used to seed LB medium (200 mL) containing ampicillin (100 µg/ml) and grown at 37 °C overnight. The following day, the overnight culture was used to seed fresh LB medium (8 x 500 mL) containing ampicillin (100 µg/ml) and grown to an optical density (OD_{600}) of ~ 0.6. At this point, protein expression was induced by the addition of IPTG at a final concentration of 1 mM. Expression was carried out at 37 °C for 3 hours. Cells were harvested by centrifugation at 8000 RPM for 10 minutes at 4 °C. The combined cell pellets were resuspended in binding buffer (5 mM imidazole, 500 mM NaCl, 20 mM Tris pH 8.0, 0.005% PMSF w/v, 100 mL) and sonicated on ice for 10 bursts of 30 seconds, with 45 seconds of cooling time in between. Inclusion bodies and cell debris was collected at 10000 RPM for 15 minutes at 4 °C and the supernatant was saved for analysis. The cell pellets were resuspended in denaturing binding buffer (6 M Gu.HCl, 5 mM imidazole, 500 mM NaCl, 20 mM Tris pH 8.0, 100 mL) and stirred vigorously at 4 °C for 45 minutes. The remaining insoluble material was pelleted at 16000 RPM for 15 minutes at 4 °C. The supernatant was stored at -20 °C prior to affinity chromatography.

Ni-NTA Agarose resin (3.0 mL, Qiagen) was charged with $NiSO_4$ (0.1 M, 15 mL) and pre-equilibrated with denaturing binding buffer (30 mL). The cell extract was loaded onto the column and unbound proteins were eluted with denaturing wash buffer (6 M Gu.HCl, 25 mM imidazole, 500 mM NaCl, 20 mM Tris pH 8.0, 30 mL). The target protein was eluted with denaturing elution buffer (6 M Gu.HCl, 500 mM imidazole, 500 mM NaCl, 20 mM Tris pH 8.0, 30 mL), which was subsequently concentrated in an Amicon Ultra molecular weight cut off device (3 kDa, Millipore). This method of concentration also facilitated buffer-exchange into 6 M Gu.HCl to remove other components present in the elution buffer. ESI^+ LC-MS (*m/z*) calculated for $C_{514}H_{773}N_{157}O_{146}S_6$ 11680.12 found $[MH_{11}]^{11+}$ 1062.88.

Product protein sequence:

GHHHHHHHHHHSSGHIEGRHMSYNLLGFLQRSSNFQCQKLLWQLNGRLEYCLKDRMNFDIP
EEIKQLQQFQKEDAALTIYEMLQNIFAIFRQDSSSGC

His_{10} tagged IFNβ 1-77 Hydrazide 30



The His_{10} tagged protein **29** (~ 2 mg) was dissolved in 6 M Gu.HCl (3.0 mL) containing MESNa (10% v/v), TCEP (0.5% v/v) and hydrazinium acetate. Hydrazinium acetate was prepared as a 50% w/v stock in 7 M Hydrochloric acid, added to a final concentration of 5%. The reaction mixture was agitated at 600 RPM in 2 x 1.5 mL microcentrifuge tubes in an Eppendorf Thermomixer at 45 °C for 72 hours. The reaction was monitored by LC-MS at set intervals 0, 24, 48 and 72 hours and formation of **30** was observed. ESI^+ LC-MS (*m/z*) calculated for $C_{511}H_{770}N_{158}O_{144}S_5$ 11590.93 found $[MH_9]^{9+}$ 1289.19.

Product protein sequence:

GHHHHHHHHHHSSGHIEGRHMSYNLLGFLQRSSNFQCQKLLWQLNGRLEYCLKDRMNFDIP EEIKQLQQFQKEDAALTIYEMLQNIFAIFRQDSSSG-$NHNH_2$

Mutagenesis of His_{10} tagged IFNβ 1-77 with C-terminal Gly-Cys (M36L, M62L)

Primers were designed to introduce the mutations M36L and M62L to facilitate selective cleavage during treatment of the protein with CNBr.

M36L Forward primer:
TGCCTCAAGGACAGGCTGAACTTTGACATCCCT

M36L Reverse primer:
AGGGATGTCAAAGTTCAGCCTGTCCTTGAGGCA

The PCR mix was made up as follows:

Forward primer (10 µM stock) 5 µL
Reverse primer (10 µM stock) 5 µL
10 x ThermoPol reaction buffer 5 µL
dNTP mix (10 mM stock) 5 µL
Template DNA (IFN 1-77 pET16b) 1 µL
Deep Vent Polymerase (New England Biolabs) 1 µL
Distilled sterile water 27 µL
$MgSO_4$ (100 mM) 1 µL

PCR was conducted with the following cycling conditions:

Step	Temp	Time
Initial denaturation	95 °C	1.5 minutes
20 cycles	95 °C	30 seconds
	54 °C	30 seconds
	72 °C	8 minutes
Final extension	72 °C	10 minutes
Hold	4 °C	

M62L Forward primer:
TTGACCATCTATGAGCTGCTCCAGAACATCTTT

M62L Reverse primer:
AAAGATGTTCTGGAGCAGCTCATAGATGGTCAA

The PCR mix was made up as follows:

Forward primer (10 µM stock) 5 µL
Reverse primer (10 µM stock) 5 µL
10 x ThermoPol reaction buffer 5 µL
dNTP mix (10 mM stock) 5 µL
Template DNA (IFN 1-77 pET16b M36L) 2 µL
Deep Vent Polymerase (New England Biolabs) 1 µL
Distilled sterile water 26 µL
$MgSO_4$ (100 mM) 1 µL

PCR was conducted as above.

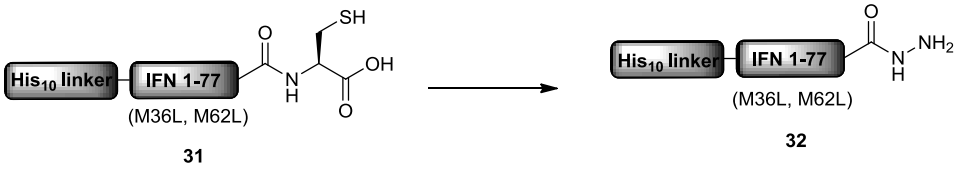
Expression and purification of His_{10} tagged IFNβ 1-77 (M36L, M62L) with C-terminal Gly-Cys 31

Protein was expressed and purified as before for **29.** ESI^+ LC-MS (*m/z*) calculated for $C_{516}H_{777}N_{157}O_{146}S_4$ 11643.91 found $[MH_{12}]^{12+}$ 971.51.

Product protein sequence:

GHHHHHHHHHHSSGHIEGRHMSYNLLGFLQRSSNFQCQKLLWQLNGRLEYCLKDRLNFDIPE
EIKQLQQFQKEDAALTIYELLQNIFAIFRQDSSSGC

His_{10} tagged IFNβ 1-77 (M36L, M62L) hydrazide 32



The His_{10} tagged protein **31** (~ 4 mg) was dissolved in 6 M Gu.HCl (3.0 mL) containing MESNa (10% v/v), TCEP (0.5% v/v) and hydrazinium acetate. Hydrazinium acetate was prepared as a 50% w/v stock in 7 M Hydrochloric acid, added to a final concentration of 5%. The reaction mixture was agitated at 750 RPM in 2 x 1.5 mL microcentrifuge tubes in an Eppendorf Thermomixer at 45 °C for 48 hours. After 48 hours, LC-MS showed partial conversion to the product **32**. The reaction mixture was diluted with 6 M Gu.HCl (10 mL) and transferred to an Amicon Ultra molecular weight cut off device (3 kDa, Millipore). The sample was rigorously buffer-exchanged by concentrating to ~ 1 mL and subsequently re-diluting with 6 m Gu.HCl (12 mL). This process was repeated 3 times and finally diluted into 6 M Gu.HCl (3.0 mL) containing MESNa (10% v/v), TCEP (0.5% v/v) and hydrazinium acetate (5% w/v) The reaction was continued for a further 48 hours at 45 °C. ESI^+ LC-MS (*m/z*) calculated for $C_{513}H_{774}N_{158}O_{144}S_3$ 11554.80 found $[MH_{12}]^{12+}$ 963.88.

Product protein sequence:

GHHHHHHHHHHSSGHIEGRHMSYNLLGFLQRSSNFQCQKLLWQLNGRLEYCLKDRLNFDIPE EIKQLQQFQKEDAALTIYELLQNIFAIFRQDSSSGC-$NHNH_2$

Cloning of His_{10} tagged IFNβ 94-166 preceded by Factor Xa recognition site

Primers were designed to allow the cloning of the IFNβ 94-166 preceded by a Factor Xa recognition site allowing the liberation of an N-terminal cysteine on treatment with the protease. The gene was cloned into vector pET16b for expression with an N-terminal polyhistidine tag.

Forward primer:
GGTGGTCATATGATCGAAGGTCGTTGTATAAACCATCTGAAG

Reverse primer:
GGTGGTGGATCCTTAGTTTCGGAGGTAACC

The PCR mix was made up as follows:

Forward primer (10 µM stock) 5 µL
Reverse primer (10 µM stock) 5 µL
10 x ThermoPol reaction buffer 5 µL
dNTP mix (10 mM stock) 5 µL
Template DNA (IFN 1-166 pKYB1) 1 µL
Deep Vent Polymerase (New England Biolabs) 1 µL
Distilled sterile water 28 µL

PCR was conducted with the following cycling conditions:

Step	Temp	Time
Initial denaturation	95 °C	2 minutes
30 cycles	95 °C	30 seconds
	54 °C	30 seconds
	72 °C	1 minute
Final extension	72 °C	5 minutes
Hold	4 °C	

DNA ligation for His_{10} tagged IFNβ 94-166

The vector and PCR product were treated with NdeI/BamHI to generate sticky ends. DNA ligation was performed using a Rapid DNA Ligation Kit (Roche) according to the manufacturer's instructions.

Expression and purification of His_{10} tagged IFNβ 94-166 preceded by Factor Xa recognition site

His_{10} tagged IFNβ 94-166 was expressed and purified as before for **29**. ESI^+ LC-MS (*m/z*) calculated for $C_{534}H_{819}N_{169}O_{136}S_4$ 11910.53 found $[MH_{16}]^{16+}$ 745.35.

Product protein sequence:

GHHHHHHHHHHSSGHIEGRHMIEGRCINHLKTVLEEKLEKEDFTRGKLMSSLHLKRYYGRILHY
LKAKEYSHCAWTIVRVEILRNFYFINRLTGYLRN

Cloning of IFNβ 94-166 thioredoxin fusion

Primers were designed to allow the cloning of the IFNβ 94-166 gene into vector pET32XaLIC for ligation independent cloning.

Forward primer:
GGTATTGAGGGTCGCTGTATAAACCATCTGAAGACAGTCC

Reverse primer:
AGAGGAGAGTTAGAGCCTTAGTTTCGGAGGTAACCTG

The PCR mix was made up as follows:

Forward primer (10 µM stock) 5 µL
Reverse primer (10 µM stock) 5 µL
10 x ThermoPol reaction buffer 5 µL
dNTP mix (10 mM stock) 5 µL
Template DNA (IFN 94-166 pET16b) 1 µL
Deep Vent Polymerase (New England Biolabs) 1 µL
Distilled sterile water 28 µL

PCR was conducted with the following cycling conditions:

Step	Temp	Time
Initial denaturation	95 °C	2 minutes
30 cycles	95 °C	30 seconds
	64 °C	30 seconds
	72 °C	2.5 minutes
Final extension	72 °C	10 minutes
Hold	4 °C	

Ligation independent cloning for IFNβ 94-166 thioredoxin fusion

Ligation independent cloning was carried out according to the manufacturer's instructions (Novagen).

Expression of IFNβ 94-166 thioredoxin intein 33

The protein was expressed in BL21(DE3) cells as before for **27** (2 x 500 mL) at 30 °C for 5 hours. After collection of the cells by centrifugation, the whole cell lysate and soluble fraction were treated with Bug Buster protein extraction reagent (Novagen) and analysed by SDS-PAGE.

Mutagenesis of His_{10} tagged IFNβ 94-166 (M117L and cleavage site)

Primers were designed to introduce the mutations M117L and methionine before the desired N-terminal cysteine for selective CNBr cleavage.

M117L Forward primer:
ACCAGGGGAAAACTCCTGAGCAGTCTGCACCTG

M117L Reverse primer:
CAGGTGCAGACTGCTCAGGAGTTTTCCCCTGGT

The PCR mix was made up as follows:

Forward primer (10 µM stock) 2.5 µL
Reverse primer (10 µM stock) 2.5 µL
10 x ThermoPol reaction buffer 5 µL
dNTP mix (10 mM stock) 5 µL
Template DNA (IFN 94-166 pET16b) 0.5 µL
Deep Vent Polymerase (New England Biolabs) 0.5 µL
Distilled sterile water 32.5 µL
$MgSO_4$ (100 mM) 0.5 µL

PCR was conducted with the following cycling conditions:

Step	Temp	Time
Initial denaturation	95 °C	1.5 minutes
20 cycles	95 °C	30 seconds
	54 °C	30 seconds
	72 °C	7 minutes
Final extension	72 °C	10 minutes
Hold	4 °C	

Cleavage site (**M**IEGR**C**→**L**IEG**M**C) Forward primer:
GAAGGTCGTCATCTGATCGAAGGTATGTGTATAAACCAT

Cleavage site (**M**IEGR**C**→**L**IEG**M**C) Reverse primer:
ATGGTTTATACACATACCTTCGATCAGATGACGACCTTC

The PCR mix was made up as follows:

Forward primer (10 µM stock) 5 µL
Reverse primer (10 µM stock) 5 µL
10 x ThermoPol reaction buffer 5 µL
dNTP mix (10 mM stock) 5 µL
Template DNA (IFN 94-166 pET16b M117L) 1 µL
Deep Vent Polymerase (New England Biolabs) 1 µL
Distilled sterile water 25.75 µL
$MgSO_4$ (100 mM) 1 µL
DMSO 1.25 µL

PCR was conducted with the following cycling conditions:

Step	Temp	Time
Initial denaturation	95 °C	1.5 minutes
20 cycles	95 °C	30 seconds
	54 °C	30 seconds
	72 °C	8 minutes
Final extension	72 °C	10 minutes
Hold	4 °C	

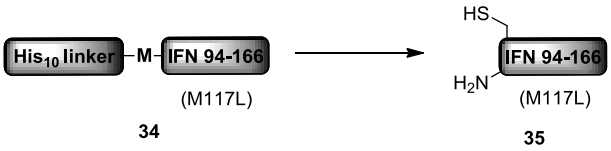
Expression and purification of His_{10} tagged IFNβ 94-166 CNBr mutant 34

Protein was expressed and purified as before for **29.** ESI^+ LC-MS (*m/z*) calculated for $C_{535}H_{820}N_{166}O_{136}S_3$ 11849.46 found $[MH_{13}]^{13+}$ 912.34.

Product protein sequence:

GHHHHHHHHHHSSGHIEGRHLIEGMCINHLKTVLEEKLEKEDFTRGKLLSSLHLKRYYGRILHYL
KAKEYSHCAWTIVRVEILRNFYFINRLTGYLRN

IFNβ 94-166 (M117L) 35



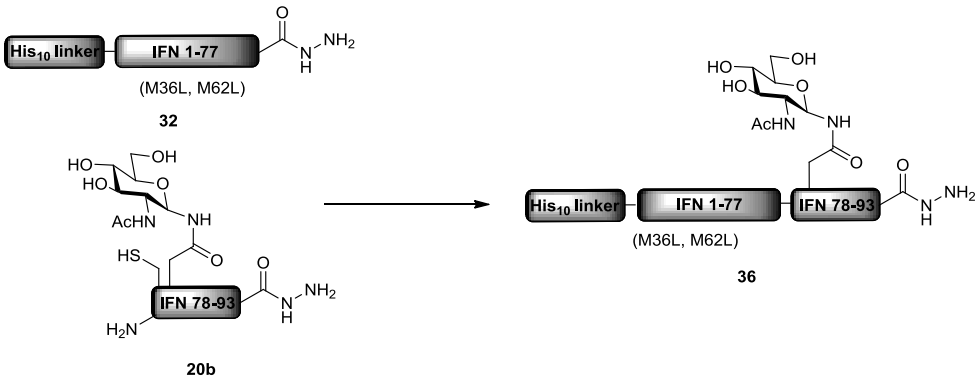
The His_{10} tagged protein **34** (7 mg) was dissolved in 6 M Gu.HCl (1.0 mL) containing 0.1 M HCl. CNBr was added (~ 5 mg) and the mixture was stirred overnight with the exclusion of light. The following day, the sample was treated with DTT (5.0 mg) and checked by LC-MS which indicated successful cleavage. The reaction mixture was diluted to 5.0 mL in denaturing binding buffer (6 M Gu.HCl, 5 mM imidazole, 500 mM NaCl, 20 mM Tris pH 8.0) containing 2-mercaptoethanol (5 mM). The mixture was adjusted to pH 8.0 with NaOH and applied to Ni-NTA resin (1.0 mL) pre-equilibrated with denaturing binding buffer (10 mL). The flow-through was collected and the resin was washed with further denaturing binding buffer (2 x 2.5 mL). The washes were combined with the flow-through and concentrated in an Amicon Ultra molecular weight cut off device (3 kDa, Millipore) to a final volume of 1.0 mL. The concentrated protein mixture was diluted with distilled water to a final volume of 15 mL and left to precipitate overnight at 4 °C. Precipitated protein was collected by centrifugation, resuspended in water/acetonitrile (3:1) and lyophilised to yield ~ 6 mg. ESI^+ LC-MS (*m/z*) calculated for $C_{410}H_{646}N_{114}O_{105}S_2$ 8916.35 found $[MH_9]^{9+}$ 991.60.

Product protein sequence:

CINHLKTVLEEKLEKEDFTRGKLLSSLHLKRYYGRILHYLKAKEYSHCAWTIVRVEILRNFYFINRLT GYLRN

8.4 Chapter 4 experimental details

His_{10} tagged IFNβ 1-93 (GlcNAc) hydrazide 36

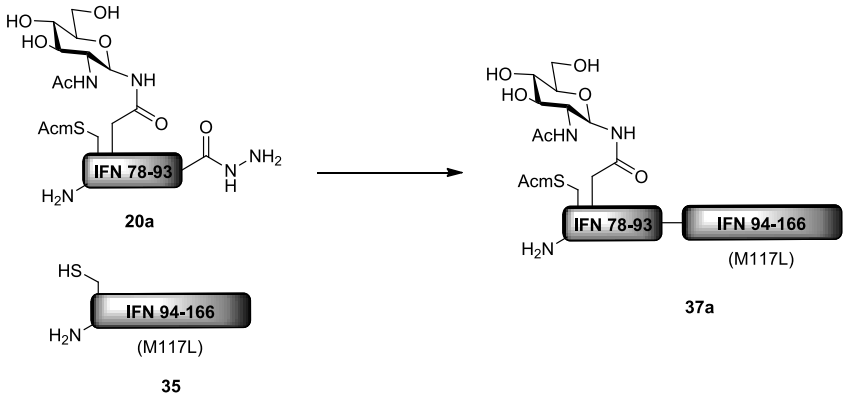


The N-terminal fragment **32** (6 mg) was dissolved in diazotisation buffer (6 M Gu.HCl, 0.2 M sodium phosphate pH 3.5, 600 µL) and stirred at -10 °C. $NaNO_2$ (0.2 M, 60 µL) was added dropwise, and the mixture was stirred at -10 °C for 20 minutes. Ligation buffer (6 M Gu.HCl, 0.2 M sodium phosphate pH 7.0, 600 µL) containing 0.2 M MPAA was added and the pH of the final reaction mixture was adjusted to 7.0. IFNβ glycopeptide **20b** (3 mg) was added to the stirred mixture which was allowed to warm to room temperature. The mixture was transferred to a 1.5 mL microcentrifuge tube and shaken at 25 °C, 800 RPM. After 2 hours the sample was treated with neutralised TCEP solution (200 mM, 120 µL) and the reaction was allowed to proceed overnight.

Product protein sequence:

GHHHHHHHHHHSSGHIEGRHMSYNLLGFLQRSSNFQCQKLLWQLNGRLEYCLKDRLNFDIPE EIKQLQQFQKEDAALTIYELLQNIFAIFRQDSSSGCWN(GlcNAc)ETIVENLLANVYH-$NHNH_2$

His_{10} tagged IFNβ 78-166 (GlcNAc) Acm 37a



Glycopeptide **20a** (4 mg) was dissolved in diazotisation buffer (6 M Gu.HCl, 0.2 M sodium phosphate pH 3.0, 400 µL) and stirred at -10 °C. $NaNO_2$ (0.2 M, 40 µL) was added dropwise, and the mixture was stirred at -10 °C for 15 minutes. To this mixture was added the C-terminal fragment **35** dissolved in ligation buffer (6 M Gu.HCl, 0.2 M sodium phosphate pH 7.0, 600 µL) containing 0.2 M MPAA and the pH of the final reaction mixture was adjusted to 7.0. The reaction was allowed to warm to room temperature and after 2.5 hours, TCEP was added from a neutral stock (1 M, 50 µL). After a further 30 minutes, LC-MS analysis revealed the complete consumption of **35** and the presence of **37a**. The product was purified by semi-preparative (RP)HPLC method 2 (t_r 27.5 min) to yield a white solid (<1 mg). ESI^+ LC-MS (*m/z*) calculated for $C_{507}H_{790}N_{138}O_{136}S_3$ 11090.74 found $[MH_{10}]^{10+}$ 1110.15.

Product protein sequence:

C(Acm)WN(GlcNAc)ETIVENLLANVYHCINHLKTVLEEKLEKEDFTRGKLLSSLHLKRYYGRILHY
LKAKEYSHCAWTIVRVEILRNFYFINRLTGYLRN

8.5 Chapter 5 experimental details

Mutagenesis of Endo A (E173Q)

Primers were designed to introduce the mutation E173Q.

E173Q Forward primer:
TGGTTTATTAACCAACAAACAGAAGGGGCAGAC

E173Q Reverse primer:
GTCTGCCCCTTCTGTTTGTTGGTTAATAAACCA

The PCR mix was made up as follows:

Forward primer (10 µM stock) 2.5 µL
Reverse primer (10 µM stock) 2.5 µL
10 x ThermoPol reaction buffer 5 µL
dNTP mix (10 mM stock) 5 µL
Template DNA (Endo A pGEX2T) 1 µL
Deep Vent Polymerase (New England Biolabs) 1 µL
Distilled sterile water 33 µL

PCR was conducted with the following cycling conditions:

Step	Temp	Time
Initial denaturation	95 °C	1.5 minutes
30 cycles	95 °C	30 seconds
	60 °C	30 seconds
	72 °C	15 minutes
Final extension	72 °C	10 minutes
Hold	4 °C	

Expression of Endo A variants 39a and 39b

The relevant plasmid was transformed into BL21(DE3) cells and a colony was used to seed LB medium (10 mL) containing ampicillin (100 µg/ml) and grown at 37 °C overnight. The following day, the overnight culture was used to seed fresh LB medium (500 mL) containing ampicillin (100 µg/ml) and grown to an optical density (OD_{600}) of ~ 0.6. At this point, protein expression was induced by the addition of IPTG at a final concentration of 1 mM. Expression was carried out at 25 °C for 16

hours. Cells were harvested by centrifugation at 8000 RPM for 15 minutes at 4 °C (Beckman Coulter Avanti J-26XP). The combined cell pellets were resuspended in binding buffer (PBS, 1 mM EDTA, 1 mM DTT, 20 mL) and sonicated on ice for 5 bursts of 30 seconds, with 30 seconds of cooling time in between. Cell debris was pelleted at 10000 RPM for 15 minutes at 4 °C and the supernatant was loaded onto a glutathione Superflow resin (3.0 mL, Qiagen) which had been pre-equilibrated with 10 volumes of binding buffer). The columns were subsequently washed with a further 10 volumes of binding buffer, before the proteins were eluted with elution buffer (50 mM Tris pH 8.0, 100 mM NaCl, 50 mM reduced glutathione, and 1 mM DTT). The eluted proteins were subsequently concentrated in an Amicon Ultra molecular weight cut off device (10 kDa, Millipore) to the required concentration.

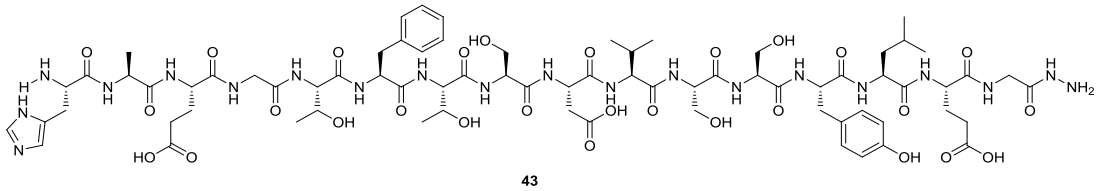
Preparation of Endo A variants without GST tag 40a and 40b

The GST tagged proteins **39a** and **39b** were dialysed into protease cleavage buffer (20 mM Tris pH 8.0, 100 mM NaCl, 2 mM $CaCl_2$) and concentrated to 1.0 mL. Thrombin (20 U, Calbiochem) was added to the mixture, which was incubated at 20 °C for 24 hours. After analysis by SDS-PAGE, further Thrombin (50 U) was added and the cleavage was allowed to proceed for a further 24 hours. The cleavage mixture was diluted to 5 mL in column buffer (100 mM NaCl, 20 mM Tris pH 7.2) and passed through glutathione Superflow resin (3.0 mL, Qiagen) which had been pre-equilibrated with 10 volumes of column buffer. The flow-through was collected and reapplied to the column for a total of five times. Finally, the resin was washed with further column buffer (2.5 mL) and the combined flow-through was rigorously buffer-exchanged in an Amicon Ultra molecular weight cut off device (10 kDa, Millipore) into potassium phosphate buffer (50 mM, pH 6.4).

Endo A activity assays with RNAse B

A mixture of RNAse B **41** (300 µM, 2 µL), denaturing buffer (5% SDS, 400 mM DTT, 1 µL) and distilled water (4 µL) was boiled at 99 °C for 10 minutes. This was followed by the addition of 10 x Protease Inhibitor cocktail (1 µL, Roche), sodium phosphate (500 mM, pH 7.5, 1 µL), NP-40 (10% w/v, 1 µL) and Endo A or dH_2O for control (1 µL). The mixture was incubated at 37 °C for an hour before analysis by SDS-PAGE.

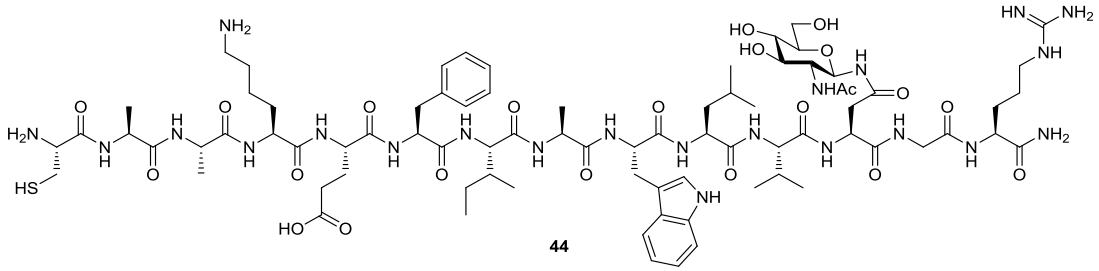
HAEGTFTSDVSSYLEG-NHNH$_2$ (GLP-1 7-22 hydrazide) 43



43

Hydrazine carboxylate Wang resin **26** (285 mg, 0.2 mmol) was treated with a mixture of Fmoc-Gly-OH (100 mg, 0.33 mmol) and Boc-Ala-OH (126 mg, 0.67 mmol) in DMF (2.2 mL), 0.45 HBTU/HOBt (2.2 mL) and DIPEA (300 µL). Fmoc analysis indicated a loading of approximately 0.35 mmol/g and a scale of 0.1 mmol. The remainder of the peptide was manually synthesised according to standard protocols before being cleaved with Reagent K (6.0 mL) for 5 hours and purified by preparative (RP)HPLC (t_r 29.7 min) yielding 39 mg of the desired peptide **43** (23%). ESI$^+$ (*m/z*) calculated for $C_{73}H_{108}N_{20}O_{28}$ 1712.76 found $[MH_2]^{2+}$ 857.86.

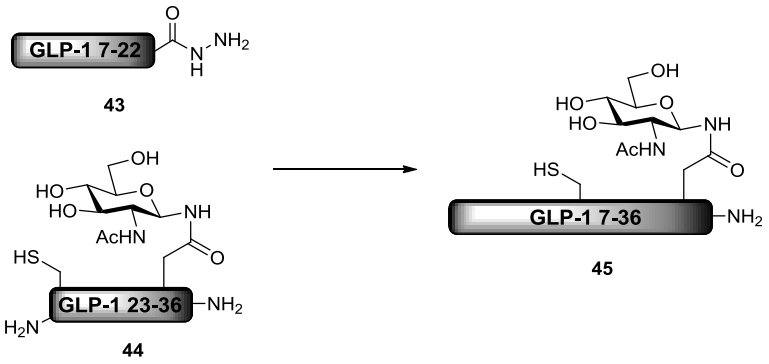
CAAKEFIAWLVN(GlcNAc)GR-NH$_2$ (GLP-1 23-36) 44



44

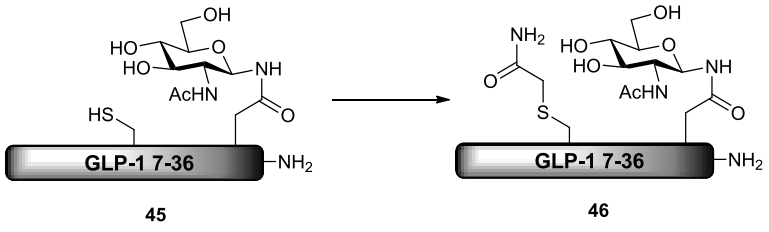
Glycopeptide **44** was manually assembled on Rink amide MBHA resin using the standard protocols. An exception to this was the glycoamino acid **6** coupling, for which only 3 equivalents were employed using HATU (3 equivalents) and DIPEA (6 equivalents). The peptide was cleaved with Reagent K (10 mL) for 4 hours, and precipitated in ether. The crude peptide was treated with aqueous hydrazine hydrate (5%, 15 mL) containing DTT (50 mM). The glycopeptide was purified by preparative (RP)HPLC (t_r 36.5 min) yielding 39 mg of the desired peptide **44** (22%). ESI$^+$ (*m/z*) calculated for $C_{80}H_{126}N_{22}O_{22}S$ 1778.91 found $[MH]^+$ 1779.68.

HAEGTFTSDVSSYLEGCAAKEFIAWLVN(GlcNAc)GR-NH_2 (GLP-1 7-36) 45



The GLP-1 peptide hydrazide **43** (19 mg) was dissolved in diazotisation buffer (6 M Gu.HCl, 0.2 M sodium phosphate pH 3.0, 2.5 mL) and stirred at -10 °C. $NaNO_2$ (0.2 M, 250 µL) was added dropwise, and the mixture was stirred at -10 °C for 20 minutes. Ligation buffer (6 M Gu.HCl, 0.2 M sodium phosphate pH 7.0, 2.5 mL) containing 0.2 M MPAA and glycopeptide **44** (29 mg) was added and the pH of the final reaction mixture was adjusted to 7.0. The reaction was allowed to warm to room temperature and stirred for 3 hours after which the sample was treated with neutralised TCEP solution (1 M, 250 µL) and checked by LC-MS. Analysis revealed the formation of the product **45** with no trace of the hydrazide **43** and hence the mixture was purified by preparative (RP)HPLC (t_r 40.7 min). The product co-eluted with MPAA yielding a white solid (~ 60 mg). ESI^+ LC-MS (*m/z*) calculated for $C_{153}H_{230}N_{40}O_{50}S$ 3459.64 found $[MH_4]^{4+}$ 866.27.

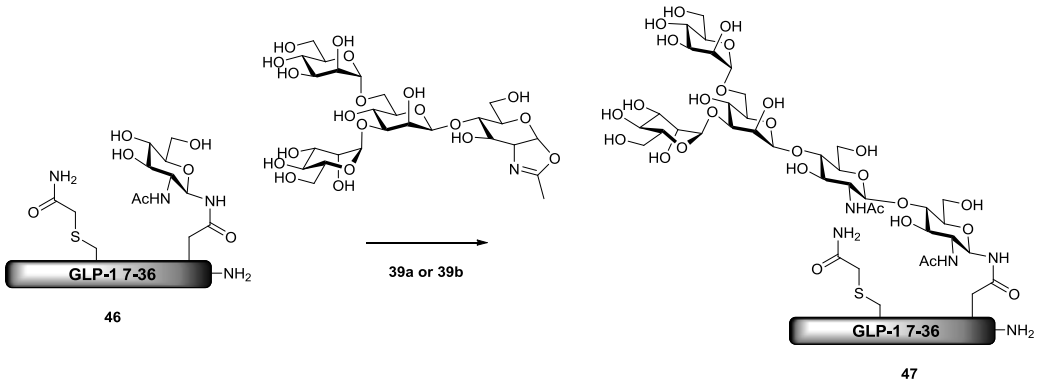
HAEGTFTSDVSSYLEGC(capped)AAKEFIAWLVN(GlcNAc)GR-NH_2 (GLP-1 7-36) 46



The glycopeptide **45** (~ 60 mg mixture with MPAA) was stirred in NH_4HCO_3 (50 mM, 30 mL) and treated with TCEP (1 M, 30 µL). The mixture was stirred under nitrogen and after 5 minutes, a solution of freshly prepared iodoacetamide (100 mM, 3.0 mL) was added. The mixture was stirred at room temperature, after which LC-MS

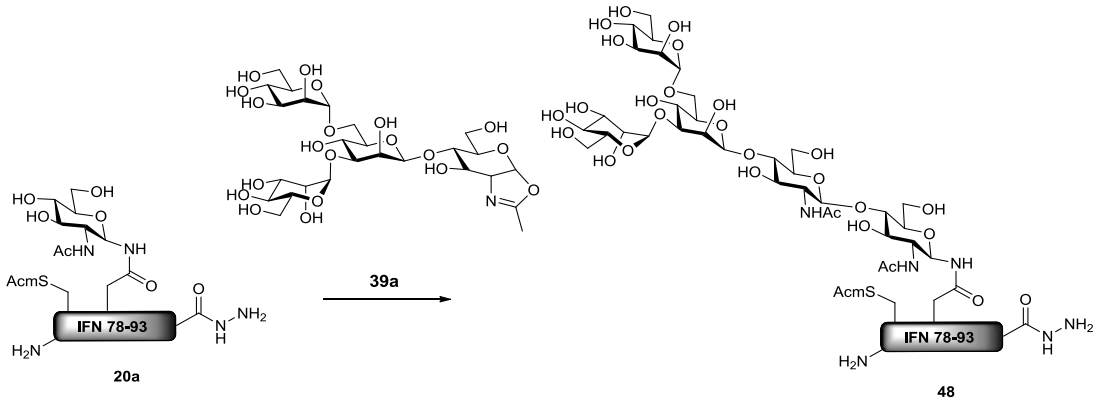
analysis suggested complete conversion to the product **46**. The product was purified by preparative (RP)HPLC (t_r 40.7 min) to yield 12 mg of the capped glycopeptide. ESI^+ LC-MS (*m/z*) calculated for $C_{155}H_{233}N_{41}O_{51}S$ 3516.66 found $[MH_4]^{4+}$ 880.62.

Transfer assays with GLP-1 7-36 and Endo A variants



GLP-1 7-36 **46** (700 µg) and $Man_3GlcNAc$ oxazoline (1.5 mg, 10 equivalents) were dissolved in potassium phosphate buffer (50 mM, pH 6.4, 375 µL). PMSF (100 mM in isopropanol) was added to a final concentration of 1 mM and Endo A **39a** or **39b** was added from a 100 µM stock to a final concentration of 1 µM. The reaction was incubated at 23 °C and monitored by analytical (RP)HPLC at various time points. In the case of the wild type Endo A **39a**, product formation was confirmed by LC-MS and the GLP-1 pentasaccharide **47** was purified by preparative (RP)HPLC using a Phenomenex Jupiter 4 µm Proteo 90 Å, LC Column (diameter = 250 mm × 21.2 mm): flow rate 7.0 mL/min, detection at 230, 254 and 280 nm using a gradient of 95% water (0.1% TFA)/5% acetonitrile (0.1% TFA), to 40% water (0.1% TFA)/60% acetonitrile (0.1% TFA), over 45 min (t_r 42.8 min). ESI^+ LC-MS (*m/z*) calculated for $C_{181}H_{276}N_{42}O_{71}S$ 4205.90 found $[MH_4]^{4+}$ 1053.08.

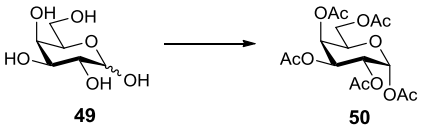
C(Acm)WN(Man$_3$GlcNAc$_2$)ETIVENLLANVYH-NHNH$_2$ (IFNβ 78-93 pentasaccharide hydrazide) Acm 48



IFNβ glycopeptide **20a** (1 mg) was suspended in potassium phosphate buffer (50 mM, pH 6.5, 600 µL) and Endo A **39a** was added from a 100 µM stock to a final concentration of 1 µM, followed by Man$_3$GlcNAc oxazoline (1 mg, 3 equivalents). The mixture was agitated at room temperature at 500 RPM and after 15 minutes was analysed by LC-MS. After confirmation that the desired glycopeptide **48** had been formed, the mixture was immediately subjected to purification by semi-preparative (RP)HPLC method 1 (t_r 34.9 min) yielding (<1 mg). ESI$^+$ LC-MS (*m/z*) calculated for $C_{123}H_{191}N_{27}O_{51}S$ 2894.29 found $[MH_3]^{3+}$ 966.19.

8.6 Chapter 6 experimental details

1, 2, 3, 4, 6-Penta-*O*-acetyl-α-D-galactopyranose 50



49 **50**

Perchloric acid (20 µL) was added to a stirred solution of acetic anhydride (50 mL) in an ice bath at 0 °C. Galactose **49** (5.00 g, 27.8 mmol) was added in small portions over an hour. TLC indicated completion of the reaction after 4 hours and the reaction mixture was concentrated to 20 mL. The reaction mixture was then diluted with chloroform (125 mL) and washed with saturated aqueous sodium bicarbonate (4 x 30 mL), water (30 mL) and brine (30 mL). The organic phase was dried over magnesium sulfate and evaporated under vacuum to afford the peracetylated sugar as a white solid (10.0 g, 93%). R_f 0.16 (3:1 petroleum spirit/ethyl acetate). ^{1}H NMR (500 MHz, $CDCl_3$) δ 6.36 (1H, d, *J* = 1.8, H1), 5.49 (1H, br, H4), 5.33 - 5.31 (2H, m, H2, H3), 4.33 (1H, apparent t, *J* = 6.6, H5), 4.12 - 4.06 (2H, m, $H6_a$, $H6_b$), 2.15 (6H, s, 2 x CH_3), 2.03 (3H, s, CH_3), 2.01 (3H, s, CH_3), 1.99 (3H, s, CH_3). ^{13}C NMR (125 MHz, $CDCl_3$) δ 170.5, 170.2, 170.2, 170.0, 169.0 (5 x C=O), 89.8 (C1), 68.8 (C5), 67.5 (C4), 67.4, 66.5 (C2, C3), 61.3 (C6), 21.0, 20.7, 20.7, 20.7, 20.6 (5 x CH_3). IR υ_{max} (cm^{-1}) neat 1737 (C=O). ESI^+ MS (*m/z*) calculated for $C_{16}H_{22}O_{11}$ 390.12 found $[MNa]^+$ 413.11. m.p. 94-96 °C (lit. 96-97°C).[269]

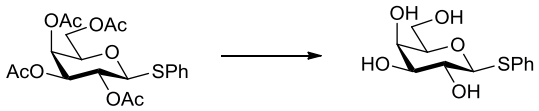
Phenyl 2,3,4,6-Tetra-*O*-acetyl-1-thio-β-D-galactopyranoside 51

50 **51**

Peracetylated galactose **50** (2.22 g, 5.68 mmol) was stirred in dry dichloromethane (40 mL). Thiophenol (1.16 mL, 11.4 mmol) was added and the mixture was stirred at room temperature for 20 minutes. The mixture was cooled in an ice bath to 0 °C, after which boron trifluoride diethyl etherate (352 µL, 2.84 mmol) was added. The mixture was then refluxed overnight for 16 hours. Dichloromethane (125 mL) was

added to dilute the reaction mixture which was then quenched with saturated aqueous sodium bicarbonate (50 mL). The organic layer was separated and washed with saturated aqueous sodium bicarbonate (50 mL) and brine (50 mL) before being dried over magnesium sulfate. The crude mixture was then concentrated under vacuum and then purified with silica gel flash chromatography (2:1 petroleum spirit/ethyl acetate) to yield the product as a clear oil (1.58 g, 63%). R_f 0.22 (7:3 petroleum spirit/ethyl acetate) ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.49 - 7.46 (2H, m, 2 x Ar-H), 7.29 - 7.27 (3H, m, 3 x Ar-H), 5.39 - 5.38 (1H, m, H4), 5.20 (1H, t, *J* = 10.1, H2), 5.03 (1H, dd, *J* = 9.9, 3.3, H3), 4.70 (1H, d, *J* = 9.9, H1), 4.18 - 4.13 (1H, m, $H6_a$), 4.11 - 4.07 (1H, m, $H6_b$), 3.92 (1H, t, *J* = 7.0, H5), 2.08 (3H, s, CH_3), 2.06 (3H, s, CH_3), 2.00 (3H, s, CH_3), 1.94 (3H, s, CH_3). ^{13}C NMR (125 MHz, $CDCl_3$) δ 170.4, 170.2, 170.0, 169.5 (4 x C=O), 132.6, 132.5, 128.9, 128.2 (4 x Ar-C), 86.6 (C1), 74.5 (C5), 72.0 (C3), 67.3 (C4), 67.3 (C2), 61.7 (C6), 20.9, 20.7, 20.7, 20.6 (4 x CH_3). IR υ_{max} (cm^{-1}) neat 1739 (C=O). ESI$^+$ MS (*m/z*) calculated for $C_{20}H_{24}O_9S$ 440.11 found $[MNa]^+$ 463.10. m.p. 78-81 °C (lit. 69-73 °C).[270]

Phenyl 1-thio-β-D-galactopyranoside 52

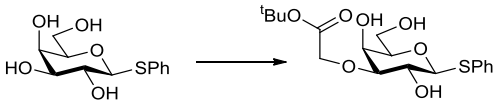


51 **52**

The acetylated thiogalactoside **51** (1.58 g, 3.59 mmol) was dissolved in dry methanol (20 mL) and stirred under nitrogen. To the stirred mixture was added sodium methoxide (10 mg, 190 µmol) and the reaction was then monitored by TLC. The pH was also monitored to ensure the catalyst had not been quenched. Further sodium methoxide was added when the pH of the reaction was found to be neutral. After 2 hours the reaction appeared complete by TLC, and Amberlyst 15 ion exchange resin was added until neutralisation of sodium methoxide. The resin was separated by filtration and the filtrate was evaporated under reduced pressure to afford a white foam (832 mg, 85%). R_f 0.08 (100% ethyl acetate). ^{1}H NMR (500 MHz, MeOD) δ 7.54 (2H, d, *J* = 7.3, 2 x Ar-H), 7.28 (2H, t, *J* = 7.3, 2 x Ar-H), 7.21 (1H, t, *J* = 7.3, Ar-H), 4.58 (1H, d, *J* = 9.6, H1), 3.89 (1H, d, *J* = 3.0, $H6_a$), 3.77 - 3.68 (2H, m, H5,

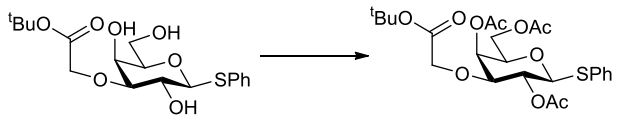
H6), 3.61 (1H, t, *J* = 9.6, H2), 3.56 (1H, t, *J* = 6.5, H4), 3.50 (1H, dd, *J* = 9.1, 3.3, H3). ^{13}C NMR (125 MHz, MeOD) δ 136.1, 132.1, 129.9, 128.0 (4 x Ar-C), 90.3 (C1), 80.6 (C3), 76.3 (C4), 71.0 (C5), 70.4 (C2), 62.6 (C6). IR υ_{max} (cm^{-1}) neat 3339 (OH). ESI$^+$ MS (*m/z*) calculated for $C_{12}H_{16}O_5S$ 272.07 found $[MNa]^+$ 295.06. m.p. 114-116 °C (lit. 105-107 °C).[271]

Phenyl 3-*O*-(*tert*-butyloxycarbonyl)methyl 1-thio-β-D-galactopyranoside 53



Thiogalactoside **52** (5.00 g, 18.4 mmol) was refluxed with dibutyltin oxide (6.86 g, 27.6 mmol) in dry methanol (50 mL) for 4 hours. The mixture was then evaporated to give a light yellow foam which was redissolved in dry tetrahydrofuran (150 mL) containing 4Å molecular sieves. The mixture was stirred and tetrabutyl ammonium bromide (3.00 g, 9.23 mmol) was added in one portion. After a further 30 minutes of stirring at room temperature, *tert*-butyl bromoacetate (16.4 mL, 111 mmol) was added with the resulting mixture stirred at 60 °C overnight. The yellow mixture was filtered and evaporated under reduced pressure before being purified by silica gel flash chromatography (2:1 petroleum spirit/ethyl acetate) to yield the product as a white crystalline solid (5.79 g, 82%). R_f 0.63 (100% ethyl acetate). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.56 (2H, d, *J* = 6.5, 2 x Ar-H), 7.32 - 7.27 (3H, m, 3 x Ar-H), 4.55 (1H, d, *J* = 9.8, H1), 4.22 (1H, d, *J* = 17.2, CH), 4.11 (1H, d, *J* = 17.2, CH), 4.03 (1H, d, *J* = 2.0, $H6_a$), 3.98 (1H, dd, *J* = 11.8, 6.6, H2), 3.87 - 3.79 (1H, m, $H6_b$, H4), 3.56 - 3.54 (1H, m, H3), 3.33 (1H, dd, *J* = 8.9, 3.2, H5), 1.47 (9H, s, $C(CH_3)_3$). ^{13}C NMR (125 MHz, $CDCl_3$) δ 171.4 (C=O), 132.5, 129.0, 127.9 (4 x Ar-C), 88.3 (C1), 85.0 (C5), 83.2 ($\underline{C}(CH_3)_3$), 78.3 (C3), 68.3 (C4), 67.9 (CH_2), 67.1 (C2), 62.8 (C6), 28.1 ($C(\underline{C}H_3)_3$). IR υ_{max} (cm^{-1}) neat 3405 (OH), 1720 (C=O). ESI$^+$ HR-MS (*m/z*) $C_{18}H_{26}O_7S$ found $[MNa]^+$ 409.1282, calculated 409.1297. m.p. 99-102 °C.

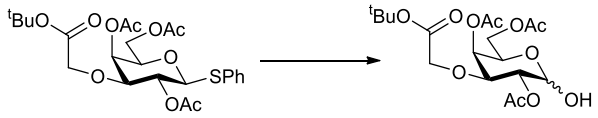
Phenyl 2,4,6-Tri-*O*-acetyl-3-*O*-(*tert*-butyloxycarbonyl)methyl 1-thio-β-D-galactopyranoside 54



53 **54**

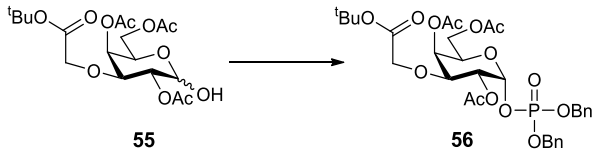
Thiogalactoside **53** (1.58 g, 4.10 mmol) was dissolved in dry pyridine (12 mL) and stirred at 0 °C in an ice bath. Acetic anhydride (6 mL) was added and the reaction was allowed to warm to room temperature and stirred overnight. The yellow mixture was then diluted with chloroform (60 mL) and added to crushed ice (20 g). The organic layer was then washed with water (50 mL), saturated aqueous sodium bicarbonate (50 mL), 2 M aqueous hydrochloric acid (50 mL), water (50 mL) and brine (50 mL). The organic extract was dried over anhydrous magnesium sulfate and then concentrated under vacuum to afford a yellow solid (2.03 g, 97%). R_f 0.48 (2:1 petroleum spirit/ethyl acetate). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.50 - 7.48 (2H, m, 2 x Ar-H), 7.27 - 7.26 (3H, m, 3 x Ar-H), 5.47 (1H, dd, *J* = 3.3, 1.1, H4), 5.12 (1H, t, *J* = 9.8, H2), 4.68 (1H, d, *J* = 10.1, H1), 4.14 - 4.12 (2H, m, $H6_a$, $H6_b$), 3.95 - 3.94 (2H, m, CH_2), 3.83 - 3.80 (1H, m, H5), 3.71 (1H, dd, *J* = 9.5, 3.3, H3), 2.15 (3H, s, CH_3), 2.07 (3H, s, CH_3), 2.03 (3H, s, CH_3), 1.42 (9H, s, $C(CH_3)_3$). ^{13}C NMR (125 MHz, $CDCl_3$) δ 170.6, 170.4, 169.9, 168.9 (4 x C=O), 149.1, 132.6, 128.9, 128.0 (4 x Ar-C), 86.4 (C1), 81.8 ($\underline{C}(CH_3)_3$), 79.0 (C3), 74.6 (C5), 69.0 (C2), 66.7 (CH_2), 66.2 (C4), 62.3 (C6), 28.1 ($C(\underline{C}H_3)_3$), 22.2, 21.1, 20.8 (3 x CH_3). IR υ_{max} (cm^{-1}) neat 1732 (C=O). ESI$^+$ HR-MS (*m/z*) $C_{24}H_{32}O_{10}S$ found $[MNa]^+$ 535.1621, calculated 535.1614. m.p. 85-86 °C.

2,4,6-Tri-*O*-acetyl-3-*O*-(*tert*-butyloxycarbonyl)methyl D-galactopyranoside 55



The acetylated thiogalactoside **54** (620 mg, 1.21 mmol) was stirred at room temperature in acetone (5.3 mL) and water (0.6 mL). To the stirred solution was added freshly recrystallised *N*-bromosuccinimide (650 mg, 3.65 mmol) which immediately turned the reaction mixture bright red (fading over time). The reaction was monitored by TLC and further *N*-bromosuccinimide was added when the reaction appeared to halt. TLC indicated completion within 2 hours. The mixture was diluted with ethyl acetate (25 mL) and quenched with saturated aqueous sodium thiosulfate (25 mL). The organic layer was then washed with water (25 mL) and then washed again with saturated aqueous sodium thiosulfate (25 mL). This process was repeated until no more precipitate formed, after which the organic layer was washed with water (25 mL) again. The crude product was purified by silica gel flash chromatography (1:1 petroleum spirit/ethyl acetate) to yield the product as a clear oil (411 mg, 81%). R_f 0.39 (1:1 petroleum spirit/ethyl acetate). IR υ_{max} (cm^{-1}) neat 1744 (C=O). ESI^+ HR-MS (*m/z*) $C_{18}H_{28}O_{11}$ found $[MNa]^+$ 443.1519, calculated 443.1529.

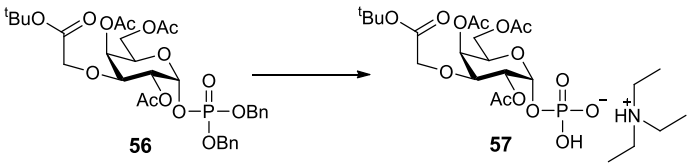
2,4,6-Tri-*O*-acetyl-3-*O*-(*tert*-butyloxycarbonyl)methyl 1-*O*-diphenoxyphosphoryl-α-D-galactopyranoside 56



Galactose derivative **55** (813 mg, 1.93 mmol) was dissolved in dry tetrahydrofuran (50 mL) and stirred under nitrogen in a flame dried flask at -78 °C. Lithium diisopropylamide (1.8 M solution in tetrahydrofuran/heptane/ethylbenzene) (1.28 mL, 2.32 mmol) was added and after a further 5 minutes of stirring, tetrabenzyl pyrophosphate (1.25 g, 2.32 mmol) in dry tetrahydrofuran (5.0 mL) was added. The

reaction was allowed to warm to room temperature and stirred for a further 2 hours. The reaction was quenched with ethyl acetate (300 mL) and transferred to a separating funnel. The mixture was then washed with water (150 mL), 2 M aqueous hydrochloric acid (150 mL), saturated aqueous sodium bicarbonate (150 mL) and brine (150 mL). The organic extract was then dried over anhydrous magnesium sulfate and then concentrated under vacuum to give a clear oil which was subsequently purified by silica gel flash chromatography (2:1 to 1:1 petroleum spirit/ethyl acetate) to yield a clear oil (952 mg, 72%). R_f 0.45 (1:1 petroleum spirit/ethyl acetate). ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.32 - 7.25 (10H, m, 10 x Ar-H), 5.86 (1H, dd, *J* = 6.8, 3.4, H1), 5.51 (1H, dd, *J* = 2.8, 0.9, H4), 5.08 (1H, dt, *J* = 10.2, 3.0, H2), 5.05 - 4.98 (4H, m, 2 x CH_2), 4.21 (1H, t, *J* = 6.6, H5), 4.06 - 4.03 (1H, m, $H6_a$), 3.98 - 3.89 (3H, m, CH_2, $H6_b$), 3.87 (1H, dd, *J* = 10.2, 3.4, H3), 2.04 (3H, s, CH_3), 1.86 (3H, s, CH_3), 1.83 (3H, s, CH_3), 1.40 (9H, s, $C(CH_3)_3$). ^{13}C NMR (150 MHz, $CDCl_3$) δ 170.5, 170.4, 170.2, 168.8 (4 x C=O), 128.9, 128.8, 128.2, 128.1 (4 x Ar-C), 95.0 (d, $J_{C\text{-}P}$ = 6.0, C1), 81.9 ($\underline{C}(CH_3)_3$), 73.5 (C3), 69.7 (C5), 69.2 (d, $J_{C\text{-}P}$ = 7.2, CH_2), 69.0 (C2), 67.8 (CH_2), 67.2 (C4), 62.0 (C6), 28.2 ($C(\underline{C}H_3)_3$), 20.9, 20.8, 20.7 (3 x CH_3). ^{31}P NMR (121MHz, $CDCl_3$) δ -1.44. IR υ_{max} (cm^{-1}) neat 1747 (C=O). ESI^+ HR-MS (*m/z*) $C_{32}H_{41}O_{14}P$ found $[MNa]^+$ 703.2184, calculated 703.2132.

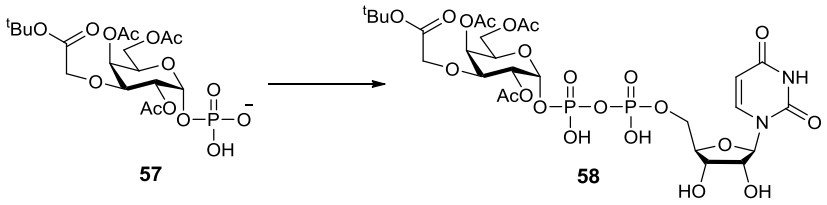
2,4,6-Tri-*O*-acetyl 3-*O*-(*tert*-butyloxycarbonyl)methyl 1-*O*-phosphoryl-α-D-galactopyranoside (triethylammonium salt) 57



The dibenzyl protected sugar monophosphate **56** (890 mg, 1.3 mmol) was stirred in dry methanol (25 mL) in the presence of Palladium on carbon (40 mg, 10% loading) under a hydrogen atmosphere at room temperature for 2 hours. Completion of the reaction was determined by TLC and the mixture was filtered through celite and washed through with methanol (2 x 25 mL). The filtrate was then stirred with triethylamine (250 µL) for 5 minutes and then concentrated *in vacuo* to give an off white amorphous solid which was further dried by co-evaporation with toluene to

yield a white solid (690 mg, 68% by NMR). R_f 0.5 (4:1 acetonitrile/water). ^{1}H NMR (600 MHz, $CDCl_3$) δ 5.68 - 5.66 (1H, br, H4), 5.53 (1H, d, *J* = 2.1, H1), 4.98 (1H, d, *J* = 10.3, H3), 4.42 - 4.39 (1H, m, H5), 4.09 - 3.94 (5H, m, H2, $H6_a$, $H6_b$, CH_2), 2.05 (6H, s, 2 x CH_3), 1.98 (3H, s, CH_3), 1.40 (9H, s, $C(CH_3)_3$). ^{13}C NMR (150 MHz, $CDCl_3$) δ 171.1, 170.8, 170.4, 169.1 (4 x C=O), 92.2 (C4), 81.6 ($\underline{C}(CH_3)_3$), 74.0 (C3), 70.2 (d, $J_{C\text{-}P}$ = 7.7, C2), 67.8 (d, $J_{C\text{-}P}$ = 9.5, C1), 67.2 (CH_2), 61.7 (C6), 28.2 ($C(\underline{C}H_3)$), 21.2, 20.9, 20.9 (3 x CH_3). ^{31}P NMR (121MHz, $CDCl_3$) δ -0.66. IR υ_{max} (cm^{-1}) neat 1740 (C=O). ESI^- HR-MS (*m/z*) $C_{18}H_{29}O_{14}P$ found $[M\text{-}H]^-$ 499.1257, calculated 499.1217.

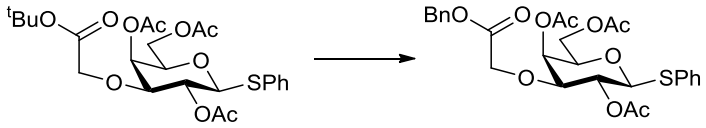
2,4,6-Tri-*O*-acetyl 3-*O*-(*tert*-butyloxycarbonyl)methyl 1-*O*-uridine-5'-diphosphoryl-α-D-galactopyranoside (Na salt) 58



Sugar monophosphate **57** (399 mg, 800 µmol) was dissolved in toluene and rotary evaporated to dryness. This process was repeated twice more and then dissolved in *N, N*-dimethylformamide (8 mL) under a nitrogen atmosphere. 4Å molecular sieves were added followed by uridine 5'-monophosphomorpholidate 4-morpholine-*N,N*'-dicyclohexyl carboxamidine salt (657 mg, 960 µmol). The mixture was stirred under nitrogen at 60 °C for 24 hours after which water (2.0 mL) was added. The mixture was stirred for a further ten minutes before being evaporated to dryness. The product was then purified using silica gel flash chromatography (1:20 to 1:4 water/acetonitrile) to yield the product as an off white crystalline solid (190 mg). The purified material was redissolved in water (5 mL) and passed through cation exchange resin (2 mL, Na^+ form). The flow through was collected and the resin was washed with water (10 mL) before the combined fractions were lyophilised. R_f 0.54 (4:1 acetonitrile/water). ^{1}H NMR (500 MHz, D_2O) δ 7.88 (1H, d, *J* = 8.1, Ur CH), 5.92 - 5.90 (2H, m, Rib H1, Ur CH), 5.71 - 5.68 (1H, dd, *J* = 7.7, 3.5, Gal H1), 5.59 - 5.56 (1H, m, Gal H4), 5.02 - 4.99 (1H, m, Gal H2), 4.45 (1H, q, Gal H5), 4.28 - 4.05 (10H, m, Rib H2, Rib H3, Rib H4, Rib $H5_a$, Rib $H5_b$, Gal H3, Gal $H6_a$, Gal $H6_b$, CH_2), 2.11 (6H, s, 2 x

CH_3), 2.01 (3H, s, CH_3), 1.39 (9H, s, $C(CH_3)_3$). ^{13}C NMR (125 MHz, D_2O) δ 173.9, 173.5, 170.9, 166.4, 158.2 152.1 (6 x C=O), 142.1 (Ur C=C), 103.1 (Ur C=C), 92.9 (Gal C1), 88.9 (Rib C1), 84.3 6 ($\underline{C}(CH_3)_3$), 83.5, 74.1, 70.0 (3 x CH), 69.7 (d, $J_{C\text{-}P}$ = 7.7, Rib C2), 68.4 (CH), 67.8 (d, $J_{C\text{-}P}$ = 7.7, Gal C2), 66.3 (CH), 65.3 (d, $J_{C\text{-}P}$ = 5.8, Gal C1), 62.3, 54.8, 27.6 (3 x CH), ($C(\underline{C}H_3)_3$), 20.8, 20.5, 20.5 (3 x CH_3). ^{31}P NMR (121MHz, D_2O) δ -9.03 (d, $J_{P\text{-}P}$ = 19.6), -11.01 (d, $J_{P\text{-}P}$ = 20.0). IR υ_{max} (cm^{-1}) neat 1744 (C=O), 1688 (C=O). ESI$^-$ HR-MS (*m/z*) $C_{27}H_{40}N_2O_{22}P_2$ found $[M\text{-}H]^-$ 805.1483, calculated 805.1470. m.p. 210 °C (decomposed).

Phenyl 2,4,6-Tri-*O*-acetyl-3-*O*-(benzyl)carboxymethyl 1-thio-β-D-galactopyranoside 59

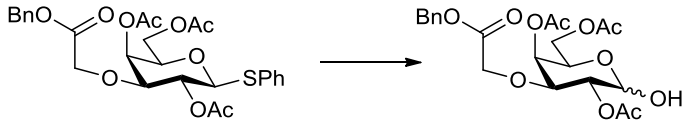


54 **59**

The protected thioglycoside **54** (2.96 g, 5.78 mmol) was dissolved with stirring in dichloromethane (20 mL) at room temperature before the addition of trifluoroacetic acid (5 mL). TLC indicated completion of the reaction within 3 hours. The crude reaction mixture was evaporated to dryness, followed by co-evaporation with toluene (3 x 10 mL) to yield a brown oil (2.50 g, 95%). The crude product was used without further purification and stirred in dry *N, N*-dimethylformamide (40 mL) with caesium carbonate (1.34 g, 4.11 mmol). Benzyl bromide (1.96 mL, 16.44 mmol) was added and the mixture was stirred under nitrogen at room temperature overnight. The following day, the reaction mixture was diluted with water (125 mL) and extracted with chloroform (2 x 125 mL). The aqueous layer was further acidified by the addition of 2 M aqueous hydrochloric acid to a final pH of 2, followed by a further extraction with chloroform (50 mL). The combined extracts were rotary evaporated to dryness and the crude material was purified by flash chromatography (3:1 petroleum spirit/ethyl acetate) to yield a clear oil (2.34 g, 72%). R_f 0.75 (1:1 petroleum spirit/ethyl acetate). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.52 - 7.50 (2H, m, 2 x Ar-H), 7.35 - 7.28 (8H, m, 8 x Ar-H), 5.47 (1H, dd, *J* = 3.3, 0.7, H4), 5.14 - 5.12 (3H, m, H2, CH_2), 4.67 (1H, d, *J* = 10.1, H1), 4.15 - 4.14 (4H, m, $H6_a$, $H6_b$,

CH_2), 3.81 - 3.78 (1H, m, H5), 3.70 (1H, dd, *J* = 9.5, 3.4, H3), 2.12 (3H, s, CH_3), 2.07 (3H, s, CH_3), 2.04 (3H, s, CH_3). ^{13}C NMR (125 MHz, $CDCl_3$) δ 170.5, 170.4, 169.8, 169.6 (4 x C=O), 135.3, 132.8, 132.7, 128.9, 128.7, 128.6, 128.5, 128.1, (8 x Ar-C), 86.5 (C1), 79.3 (C3), 74.6 (C5), 68.8 (C2), 66.7 (CH_2), 66.4 (CH_2), 66.0 (C4), 62.2 (CH_2), 21.1, 21.0, 20.8 (3 x CH_3). IR υ_{max} (cm^{-1}) neat 1750 (C=O). ESI^+ HR-MS (*m/z*) $C_{27}H_{30}O_{10}S$ found $[MNa]^+$ 569.1446, calculated 569.1457.

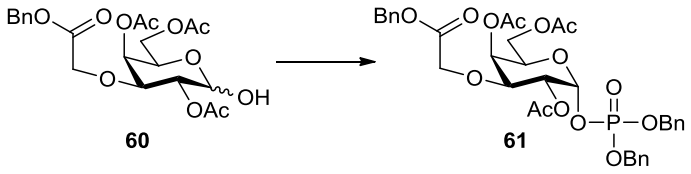
2,4,6-Tri-*O*-acetyl-3-*O*-(benzyl)carboxymethyl D-galactopyranoside 60



59 **60**

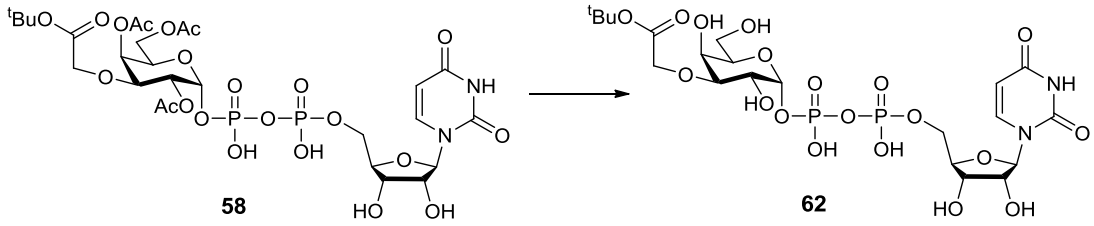
To a stirred solution of thioglycoside **59** (2.34 g, 4.28 mmol) in acetone (45 mL) and water (5 mL) was added freshly recrystallised *N*-bromosuccinimide (2.29 g, 12.9 mmol) at room temperature. TLC analysis suggested a halt in progress of the reaction after 2 hours and further *N*-bromosuccinimide (500 mg, 2.8 mmol) was added. After stirring at room temperature for another hour, the reaction mixture was concentrated under reduced pressure to 10 mL and the crude residue was taken up in ethyl acetate (150 mL). The extract was washed with water (100 mL), saturated sodium bicarbonate solution (100 mL) and water (100 mL). The organic layer was dried over magnesium sulfate before concentration *in vacuo*. The crude oil was purified by flash chromatography (3:1 petroleum spirit/ethyl acetate) to yield a clear oil (1.18 g, 61%). R_f 0.27 (1:1 petroleum spirit/ethyl acetate). IR υ_{max} (cm^{-1}) neat 1749 (C=O). ESI^+ HR-MS (*m/z*) $C_{21}H_{26}O_{11}$ found $[MNa]^+$ 477.1350, calculated 477.1373.

2,4,6-Tri-*O*-acetyl-3-*O*-(benzyl)carboxymethyl 1-*O*-diphenoxyphosphoryl-α-D-galactopyranoside 61



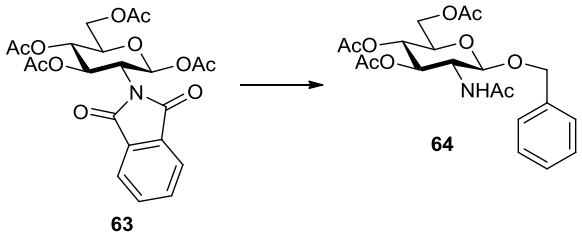
Galactose derivative **60** (508 mg, 1.12 mmol) was dissolved in dry tetrahydrofuran (15 mL) and stirred under nitrogen in a flame dried flask at -78 °C. Lithium diisopropylamide (1.8 M solution in tetrahydrofuran/heptane/ethylbenzene) (683 µL, 1.23 mmol) was added and after a further 5 minutes of stirring, tetrabenzyl pyrophosphate (662 mg, 1.23 mmol) in dry tetrahydrofuran (3.0 mL) was added. The reaction was allowed to warm to room temperature and stirred for a further 3 hours. The reaction was quenched with ethyl acetate (150 mL) and transferred to a separating funnel. The mixture was then washed with water (75 mL), 1 M aqueous hydrochloric acid (75 mL), saturated aqueous sodium bicarbonate (75mL) and brine (75 mL). The organic extract was then dried over anhydrous magnesium sulfate and then concentrated under vacuum to give a clear oil which was subsequently purified by silica gel flash chromatography (2:1 to 1:1 petroleum spirit/ethyl acetate) to yield a clear oil (540 mg, 68%). R_f 0.38 (1:1 petroleum spirit/ethyl acetate). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.37 - 7.34 (15H, m, 15 x Ar-H), 5.91 (1H, dd, *J* = 6.6, 3.4, H1), 5.55 (1H, dd, *J* = 2.2, H4), 5.17 - 5.09 (7H, m, H2, 3 x CH_2), 4.25 (1H, t, *J* = 6.4, H5), 4.17 (2H, q, *J* = 16.6, CH_2) 4.09 (1H, dd, *J* = 11.4, 5.9, $H6_a$), 3.97 (1H, dd, *J* = 11.4, 6.9, $H6_b$) 3.90 (1H, dd, *J* = 10.3, 3.3, H3), 2.11 (3H, s, CH_3), 1.91 (3H, s, CH_3), 1.88 (3H, s, CH_3). ^{13}C NMR (125 MHz, $CDCl_3$) δ 170.4, 170.3, 170.2, 169.5 (4 x C=O), 135.4, 135.3, 128.8, 128.7, 128.6, 128.5, 128.1, 128.0 (8 x Ar-C), 94.9 (d, $J_{C\text{-}P}$ = 6.0, C1), 73.8 (C3), 69.7 (d, $J_{C\text{-}P}$ = 5.0, CH_2), 69.0 (CH_2), 69.0 (C5) 68.9 (CH_2) 67.3 (C4), 67.0 (CH_2), 66.8 (C2), 61.9 (C6) 20.7, 20.6, 20.6 (3 x CH_3). ^{31}P NMR (121MHz, $CDCl_3$) δ -1.40. IR υ_{max} (cm^{-1}) neat 1753 (C=O). ESI$^+$ HR-MS (*m/z*) $C_{18}H_{29}O_{14}P$ found $[MNa^+(-PO_2(OBn)_2)]^+$ 460.1634, calculated 460.1340.

3-*O*-(*tert*-butyloxycarbonyl)methyl-1-*O*-uridine-5'-diphosphoryl-α-D-galactopyranoside (Na salt) 62



Sugar nucleotide **58** (13 mg, 19.8 µmol) was stirred in a mixture of methanol, water and triethylamine (5:3:1, 900 µL) at room temperature for 7 days. The reaction was diluted with water (20 mL) and freeze dried. The crude material was redissolved in water (1 mL) and passed through cation exchange resin (1 mL, Na^+ form). The flow through was collected and the resin was washed with water (5 mL) before the combined fractions were lyophilised. ESI^+ MS (*m/z*) calculated for $C_{21}H_{34}N_2O_{19}P_2$ 680.44 found $[MNa]^+$ 703.19. HR-MS attempted in ESI^+ and ESI^- modes, however no molecular ion was found by these methods.

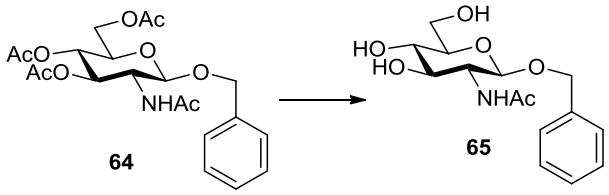
Benzyl 2-acetamido-2-deoxy-3,4,6-tri-*O*-acetyl-β-D-glucopyranoside 64



Phthalamido protected GlcNAc **63** (500 mg, 1.05 mmol) was stirred in dry dichloromethane (10 mL) with benzyl alcohol (109 µL, 1.05 mmol) under nitrogen and cooled to 0 °C. Boron trifluoride diethyl etherate (133 µL, 1.05 mmol) was added dropwise and the reaction was allowed to warm to room temperature. After stirring overnight, the reaction was diluted with dichloromethane (20 mL) before being washed with water (20 mL), saturated sodium bicarbonate (20 mL) and brine (20 mL) and dried over magnesium sulfate. The crude reaction mixture was concentrated *in vacuo* to give a yellow foam which was redissolved in n-butanol (20 mL). Ethylenediamine (232 µL, 1.57 mmol) was added, and the mixture was refluxed at 90 °C overnight. After concentration under reduced pressure, the

residue was dissolved in pyridine (20 mL) and acetic anhydride (10 mL) was added at 0 °C. After stirring under nitrogen overnight, the mixture was diluted with ethyl acetate (100 mL) before being washed with water (100 mL), saturated sodium bicarbonate (100 mL) and brine (100 mL) and dried over magnesium sulfate. The crude product was concentrated *in vacuo* and purified by silica gel flash chromatography (1:1 petroleum spirit/ethyl acetate to 100% ethyl acetate) to yield a white solid (205 mg, 45%). R_f 0.56 (100% ethyl acetate). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.45 - 7.18 (5H, m, 5 x Ar-H), 5.46 (1H, d, *J* = 8.9, NH), 5.20 (1H, dd, *J* = 10.5, 9.4, H3), 5.08 (1H, t, *J* = 9.6, H4), 4.88 (1H, d, *J* = 12.2, CH_{2a}), 4.64 - 4.58 (2H, m, H1, CH_{2b}), 4.27 (1H, dd, *J* = 12.2, 4.8, $H6_a$), 4.16 (1H, dd, *J* = 12.2, 2.4, $H6_b$), 3.97 (1H, dt, *J* = 10.5, 8.7, H2), 3.67 (1H, ddd, *J* = 9.9, 4.7, 2.5, H5) 2.09 (3H, s, CH_3), 2.01 (3H, s, CH_3), 2.00 (3H, s, CH_3), 1.90 (3H, s, CH_3). ^{13}C NMR (125 MHz, $CDCl_3$) δ 171.0, 170.8, 170.2, 169.4 (4 x C=O), 137.0, 128.6, 128.1, 128.0 (4 x Ar-C), 99.5 (C1), 72.5 (C3), 71.9 (C5), 70.7 (CH_2), 68.7 (C4), 62.2 (C6), 54.8 (C2), 23.4, 20.8, 20.7, 20.6 (4 x CH_3). IR υ_{max} (cm^{-1}) neat 1738 (C=O). ESI$^+$ MS (*m/z*) calculated for $C_{21}H_{27}NO_9$ 437.17 found $[MH]^+$ 438.17. m.p. 169-172 °C (lit. 165-166 °C).[272]

Benzyl 2-acetamido-2-deoxy-β-D-glucopyranoside 65



The acetylated glycoside **64** (100 mg, 228 µmol) was stirred in dry methanol (5 mL) with sodium methoxide (0.5 M, 50 µL) at room temperature for an hour. TLC indicated complete deacetylation and Amberlyst 15 ion exchange resin was added to neutralise the mixture. The resin was separated by filtration and the filtrate was evaporated under reduced pressure to afford **65** as a white solid (70 mg, 99%). ^{1}H NMR (600 MHz, DMSO-d_6) δ 7.76 (1H, d, *J* = 9.1, NH), 7.34 - 7.25 (5H, m, 5 x Ar-H), 4.78 (1H, d, *J* = 12.4, CH_{2a}), 4.51 (1H, d, *J* = 12.4, CH_{2b}), 4.35 (1H, d, *J* = 8.5, H1), 3.70 (1H, d, *J* = 10.8, $H6_a$), 3.55 - 3.42 (2H, m, H2, $H6_b$), 3.28 (1H, dd, *J* = 10.1, 8.1, H3), 3.10 - 3.08 (2H, m, H4, H5), 1.81 (3H, s, CH_3). ^{13}C NMR (150 MHz, DMSO-d_6) δ 169.18 (C=O), 138.1, 128.2, 127.4, 127.2 (4 x Ar-C), 100.7 (C1), 77.0 (C5), 74.0 (C3),

70.5 (C4), 69.5 (CH_2), 61.0 (C6), 55.3 (C2), 23.0 (CH_3). IR υ_{max} (cm^{-1}) neat 3299 (OH), 1640 (C=O). ESI^+ MS (*m/z*) calculated for $C_{15}H_{21}NO_6$ 311.14 found $[MH]^+$ 312.14. m.p. 191-193 °C (lit. 194-196 °C).[273]

9 References

1. R. Apweiler, H. Hermjakob and N. Sharon, *Biochim. Biophys. Acta. Gen. Subj.*, 1999, **1473**, 4-8.
2. R. A. Dwek, *Chem. Rev.*, 1996, **96**, 683-720.
3. A. Varki, *Glycobiology*, 1993, **3**, 97-130.
4. Y. Mimura, S. Church, R. Ghirlando, P. R. Ashton, S. Dong, M. Goodall, J. Lund and R. Jefferis, *Mol. Immunol.*, 2000, **37**, 697-706.
5. P. Van den Steen, P. M. Rudd, R. A. Dwek and G. Opdenakker, *Crit. Rev. Biochem. Mol. Biol.*, 1998, **33**, 151-208.
6. R. Kornfeld and S. Kornfeld, *Annu. Rev. Biochem.*, 1985, **54**, 631-664.
7. T. W. Rademacher, R. B. Parekh and R. A. Dwek, *Annu. Rev. Biochem.*, 1988, **57**, 785-838.
8. M. Wacker, D. Linton, P. G. Hitchen, M. Nita-Lazar, S. M. Haslam, S. J. North, M. Panico, H. R. Morris, A. Dell, B. W. Wren and M. Aebi, *Science*, 2002, **298**, 1790-1793.
9. C. Lizak, S. Gerber, S. Numao, M. Aebi and K. P. Locher, *Nature*, 2011, **474**, 350-355.
10. H. F. Lodish and N. Kong, *J. Cell Biol.*, 1984, **98**, 1720-1729.
11. W.-J. Ou, P. H. Cameron, D. Y. Thomas and J. Bergeron, *Nature*, 1993, **364**, 771-776.
12. C. Hammond, I. Braakman and A. Helenius, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 913-917.
13. E. J. Catherine, C. Seema, G. Natalio, J. H. Günter, L. Simon and J. B. Neil, *EMBO J.*, 2006, **26**, 28-40.
14. J. B. Huppa and H. L. Ploegh, *Cell*, 1998, **92**, 145-148.
15. A. J. Parodi, *Biochim. Biophys. Acta. Gen. Subj.*, 1999, **1426**, 287-295.
16. A. Tiziana and S. Roberto, *EMBO J.*, 2008, **27**, 315-327.
17. R. D. Cummings, *Mol. Biosyst.*, 2009, **5**, 1087-1104.
18. J. Kihlberg, M. Elofsson and L. A. Salvador, in *Methods in Enzymology*, ed. B. F. Gregg, Academic Press, 1997, pp. 221-245.
19. S. R. Hamilton and T. U. Gerngross, *Curr. Opin. Biotechnol.*, 2007, **18**, 387-392.
20. T. Ogawa, M. Sugimoto, T. Kitajima, K. K. Sadozai and T. Nukada, *Tetrahedron Lett.*, 1986, **27**, 5739-5742.
21. J. Seifert, M. Lergenmüller and Y. Ito, *Angew. Chem.Int. Ed.*, 2000, **39**, 531-534.
22. M. A. Walczak, J. Hayashida and S. J. Danishefsky, *J. Am. Chem. Soc.*, 2013, **135**, 4700-4703.
23. M. A. Walczak and S. J. Danishefsky, *J. Am. Chem. Soc.*, 2012, **134**, 16430-16433.
24. C. Unverzagt, *Angew. Chem.Int. Ed.*, 1996, **35**, 2350-2353.
25. Z. Wang, Z. S. Chinoy, S. G. Ambre, W. Peng, R. McBride, R. P. de Vries, J. Glushka, J. C. Paulson and G.-J. Boons, *Science*, 2013, **341**, 379-383.
26. A. Seko, M. Koketsu, M. Nishizono, Y. Enoki, H. R. Ibrahim, L. R. Juneja, M. Kim and T. Yamamoto, *Biochim. Biophys. Acta. Gen. Subj.*, 1997, **1335**, 23-32.
27. N. Yamamoto, Y. Ohmori, T. Sakakibara, K. Sasaki, L. R. Juneja and Y. Kajihara, *Angew. Chem.Int. Ed.*, 2003, **42**, 2537-2540.
28. M. Murakami, R. Okamoto, M. Izumi and Y. Kajihara, *Angew. Chem.Int. Ed.*, 2012, **51**, 3567-3572.
29. S. Mezzato and C. Unverzagt, *Carbohydr. Res.*, 2010, **345**, 1306-1315.
30. C. Piontek, D. Varón Silva, C. Heinlein, C. Pöhner, S. Mezzato, P. Ring, A. Martin, F. X. Schmid and C. Unverzagt, *Angew. Chem.Int. Ed.*, 2009, **48**, 1941-1945.
31. S. T. Cohen-Anisfeld and P. T. Lansbury, *J. Am. Chem. Soc.*, 1993, **115**, 10531-10537.

32. T. Conroy, K. A. Jolliffe and R. J. Payne, *Org. Biomol. Chem.*, 2010, **8**, 3723-3733.
33. R. Chen and T. J. Tolbert, *J. Am. Chem. Soc.*, 2010, **132**, 3211-3216.
34. J. D. Warren, J. S. Miller, S. J. Keding and S. J. Danishefsky, *J. Am. Chem. Soc.*, 2004, **126**, 6576-6578.
35. B. Aussedat, B. Fasching, E. Johnston, N. Sane, P. Nagorny and S. J. Danishefsky, *J. Am. Chem. Soc.*, 2012, **134**, 3532-3541.
36. P. Nagorny, N. Sane, B. Fasching, B. Aussedat and S. J. Danishefsky, *Angew. Chem.Int. Edit.*, 2012, **51**, 975-979.
37. P. Wang, X. C. Li, J. L. Zhu, J. Chen, Y. Yuan, X. Y. Wu and S. J. Danishefsky, *J. Am. Chem. Soc.*, 2011, **133**, 1597-1602.
38. P. Wang, B. Aussedat, Y. Vohra and S. J. Danishefsky, *Angew. Chem.Int. Ed.*, 2012, **51**, 11571-11575.
39. V. Ullmann, M. Rädisch, I. Boos, J. Freund, C. Pöhner, S. Schwarzinger and C. Unverzagt, *Angew. Chem.Int. Ed.*, 2012, **51**, 11566-11570.
40. S. T. Anisfeld and P. T. Lansbury, *J. Org. Chem.*, 1990, **55**, 5560-5562.
41. R. Subirós-Funosas, A. El-Faham and F. Albericio, *Tetrahedron*, 2011, **67**, 8595-8606.
42. A. J. Fairbanks, *Pure Appl. Chem.*, 2013, **85**, 1847-1863.
43. M. Fujita, S. Shoda, K. Haneda, T. Inazu, K. Takegawa and K. Yamamoto, *Biochim. Biophys. Acta. Gen. Subj.*, 2001, **1528**, 9-14.
44. T. W. D. F. Rising, T. D. W. Claridge, N. Davies, D. P. Gamblin, J. W. B. Moir and A. J. Fairbanks, *Carbohydr. Res.*, 2006, **341**, 1574-1596.
45. B. Li, Y. Zeng, S. Hauser, H. J. Song and L. X. Wang, *J. Am. Chem. Soc.*, 2005, **127**, 9692-9693.
46. H. Hojo, H. Tanaka, M. Hagiwara, Y. Asahina, A. Ueki, H. Katayama, A. Yoneshige, J. Matsuda, Y. Ito and Y. Nakahara, *J. Org. Chem.*, 2012, **77**, 9437-9446.
47. P. Wang, S. Dong, J.-H. Shieh, E. Peguero, R. Hendrickson, M. A. S. Moore and S. J. Danishefsky, *Science*, 2013, **342**, 1357-1360.
48. I. Sakamoto, K. Tezuka, K. Fukae, K. Ishii, K. Taduru, M. Maeda, M. Ouchi, K. Yoshida, Y. Nambu, J. Igarashi, N. Hayashi, T. Tsuji and Y. Kajihara, *J. Am. Chem. Soc.*, 2012, **134**, 5428-5431.
49. C. Piontek, P. Ring, O. Harjes, C. Heinlein, S. Mezzato, N. Lombana, C. Pöhner, M. Püttner, D. Varón Silva, A. Martin, F. X. Schmid and C. Unverzagt, *Angew. Chem.Int. Ed.*, 2009, **48**, 1936-1940.
50. S. B. H. Kent, *Annu. Rev. Biochem.*, 1988, **57**, 957-989.
51. S. Sakakibara, *Peptide Science*, 1999, **51**, 279-296.
52. R. B. Merrifield, *J. Am. Chem. Soc.*, 1963, **85**, 2149-2154.
53. L. A. Carpino and G. Y. Han, *J. Am. Chem. Soc.*, 1970, **92**, 5748-5749.
54. C.-D. Chang and J. Meienhofer, *Int. J. Pept. Protein Res.*, 1978, **11**, 246-249.
55. B. S. Jursic and Z. Zdravkovski, *Synth. Commun.*, 1993, **23**, 2761-2770.
56. B. Henkel, L. Zhang and E. Bayer, *Liebigs Ann.*, 1997, **1997**, 2161-2168.
57. J. C. Sheehan and G. P. Hess, *J. Am. Chem. Soc.*, 1955, **77**, 1067-1068.
58. N. L. Benoiton and F. M. F. Chen, *J. Chem. Soc., Chem. Commun.*, 1981, 543-545.
59. G. W. Anderson and F. M. Callahan, *J. Am. Chem. Soc.*, 1958, **80**, 2902-2903.
60. W. König and R. Geiger, *Chen. Ber.*, 1970, **103**, 788-798.
61. L. A. Carpino, *J. Am. Chem. Soc.*, 1993, **115**, 4397-4398.
62. A. El-Faham and F. Albericio, *Chem. Rev.*, 2011, **111**, 6557-6602.
63. B. Castro, J. R. Dormoy, G. Evin and C. Selve, *Tetrahedron Lett.*, 1975, **16**, 1219-1222.
64. J. Coste, D. Le-Nguyen and B. Castro, *Tetrahedron Lett.*, 1990, **31**, 205-208.
65. V. Dourtoglou, J.-C. Ziegler and B. Gross, *Tetrahedron Lett.*, 1978, **19**, 1269-1272.

66. J. Hachmann and M. Lebl, *Peptide Science*, 2006, **84**, 340-347.
67. M. Schnolzer and S. Kent, *Science*, 1992, **256**, 221-225.
68. P. E. Dawson and S. B. H. Kent, *Annu. Rev. Biochem.*, 2000, **69**, 923-960.
69. S. Kent, *Peptide Science*, 2010, **94**, iv-ix.
70. P. E. Dawson, T. W. Muir, I. Clarklewis and S. B. H. Kent, *Science*, 1994, **266**, 776-779.
71. T. Wieland, E. Bokelmann, L. Bauer, H. U. Lang and H. Lau, *Liebigs Ann.*, 1953, **583**, 129-149.
72. E. C. B. Johnson and S. B. H. Kent, *J. Am. Chem. Soc.*, 2006, **128**, 6640-6646.
73. S. Aimoto, *Peptide Science*, 1999, **51**, 247-265.
74. H. Hojo, Y. Matsumoto, Y. Nakahara, E. Ito, Y. Suzuki, M. Suzuki, A. Suzuki and Y. Nakahara, *J. Am. Chem. Soc.*, 2005, **127**, 13720-13725.
75. E. Saxon and C. R. Bertozzi, *Science*, 2000, **287**, 2007-2010.
76. G. Chen, J. D. Warren, J. Chen, B. Wu, Q. Wan and S. J. Danishefsky, *J. Am. Chem. Soc.*, 2006, **128**, 7460-7462.
77. J. W. Bode, R. M. Fox and K. D. Baucom, *Angew. Chem.Int. Ed.*, 2006, **45**, 1248-1252.
78. R. J. Payne, S. Ficht, W. A. Greenberg and C.-H. Wong, *Angew. Chem.Int. Ed.*, 2008, **47**, 4411-4415.
79. H. Mao, S. A. Hart, A. Schink and B. A. Pollok, *J. Am. Chem. Soc.*, 2004, **126**, 2670-2671.
80. P. E. Dawson, M. J. Churchill, M. R. Ghadiri and S. B. H. Kent, *J. Am. Chem. Soc.*, 1997, **119**, 4325-4329.
81. P. E. Dawson, *Isr. J. Chem.*, 2011, **51**, 862-867.
82. J. P. Tam and Q. Yu, *Biopolymers*, 1998, **46**, 319-327.
83. L. Z. Yan and P. E. Dawson, *J. Am. Chem. Soc.*, 2001, **123**, 526-533.
84. Q. Wan and S. J. Danishefsky, *Angew. Chem.Int. Ed.*, 2007, **46**, 9248-9252.
85. N. Metanis, E. Keinan and P. E. Dawson, *Angew. Chem.Int. Ed.*, 2010, **49**, 7049-7053.
86. L. R. Malins, N. J. Mitchell and R. J. Payne, *J. Pept. Sci.*, 2014, **20**, 64-77.
87. D. Crich and A. Banerjee, *J. Am. Chem. Soc.*, 2007, **129**, 10064-10065.
88. J. Chen, Q. Wan, Y. Yuan, J. Zhu and S. J. Danishefsky, *Angew. Chem.Int. Ed.*, 2008, **47**, 8521-8524.
89. C. Haase, H. Rohde and O. Seitz, *Angew. Chem.Int. Ed.*, 2008, **47**, 6807-6810.
90. R. Yang, K. K. Pasunooti, F. Li, X.-W. Liu and C.-F. Liu, *J. Am. Chem. Soc.*, 2009, **131**, 13592-13593.
91. K. S. Ajish Kumar, M. Haj-Yahya, D. Olschewski, H. A. Lashuel and A. Brik, *Angew. Chem.Int. Ed.*, 2009, **48**, 8090-8094.
92. J. Chen, P. Wang, J. Zhu, Q. Wan and S. J. Danishefsky, *Tetrahedron*, 2010, **66**, 2277-2283.
93. S. Shang, Z. Tan, S. Dong and S. J. Danishefsky, *J. Am. Chem. Soc.*, 2011, **133**, 10784-10786.
94. Z. Tan, S. Shang and S. J. Danishefsky, *Angew. Chem.Int. Ed.*, 2010, **49**, 9500-9503.
95. Z. Harpaz, P. Siman, K. S. A. Kumar and A. Brik, *Chembiochem*, 2010, **11**, 1232-1235.
96. P. Siman, S. V. Karthikeyan and A. Brik, *Org. Lett.*, 2012, **14**, 1520-1523.
97. L. R. Malins, K. M. Cergol and R. J. Payne, *Chembiochem*, 2013, **14**, 559-563.
98. R. E. Thompson, B. Chan, L. Radom, K. A. Jolliffe and R. J. Payne, *Angew. Chem.Int. Ed.*, 2013, **52**, 9723-9727.
99. K. M. Cergol, R. E. Thompson, L. R. Malins, P. Turner and R. J. Payne, *Org. Lett.*, 2013, **16**, 290-293.
100. L. R. Malins, K. M. Cergol and R. J. Payne, *Chem. Sci.*, 2014, **5**, 260-266.

101. A. Brik, Y.-Y. Yang, S. Ficht and C.-H. Wong, *J. Am. Chem. Soc.*, 2006, **128**, 5626-5627.
102. P. Botti, M. R. Carrasco and S. B. H. Kent, *Tetrahedron Lett.*, 2001, **42**, 1831-1833.
103. J. Offer, C. N. C. Boddy and P. E. Dawson, *J. Am. Chem. Soc.*, 2002, **124**, 4642-4646.
104. D. Macmillan and D. W. Anderson, *Org. Lett.*, 2004, **6**, 4659-4662.
105. T. Kawakami and S. Aimoto, *Tetrahedron Lett.*, 2003, **44**, 6059-6061.
106. C. Marinzi, J. Offer, R. Longhi and P. E. Dawson, *Bioorg. Med. Chem.*, 2004, **12**, 2749-2757.
107. J. Offer, *Peptide Science*, 2010, **94**, 530-541.
108. R. J. Payne, *Angew. Chem.Int. Ed.*, 2013, **52**, 505-507.
109. H. Kunz, *Angew. Chem.Int. Ed.*, 1987, **26**, 294-308.
110. X. Bu, G. Xie, C. W. Law and Z. Guo, *Tetrahedron Lett.*, 2002, **43**, 2419-2422.
111. A. B. Clippingdale, C. J. Barrow and J. D. Wade, *J. Pept. Sci.*, 2000, **6**, 225-234.
112. X. Li, T. Kawakami and S. Aimoto, *Tetrahedron Lett.*, 1998, **39**, 8669-8672.
113. H. Hojo, E. Haginoya, Y. Matsumoto, Y. Nakahara, K. Nabeshima, B. P. Toole and Y. Watanabe, *Tetrahedron Lett.*, 2003, **44**, 2961-2964.
114. S. Futaki, K. Sogawa, J. Maruyama, T. Asahara, M. Niwa and H. Hojo, *Tetrahedron Lett.*, 1997, **38**, 6237-6240.
115. A. R. Mezo, R. P. Cheng and B. Imperiali, *J. Am. Chem. Soc.*, 2001, **123**, 3885-3891.
116. R. von Eggelkraut-Gottanka, A. Klose, A. G. Beck-Sickinger and M. Beyermann, *Tetrahedron Lett.*, 2003, **44**, 3551-3554.
117. Y. Kajihara, A. Yoshihara, K. Hirano and N. Yamamoto, *Carbohydr. Res.*, 2006, **341**, 1333-1340.
118. N. Yamamoto, Y. Tanabe, R. Okamoto, P. E. Dawson and Y. Kajihara, *J. Am. Chem. Soc.*, 2007, **130**, 501-510.
119. N. Stuhr-Hansen, T. S. Wilbek and K. Strømgaard, *Eur. J. Org. Chem.*, 2013, **2013**, 5290-5294.
120. K. Nakamura, N. Hanai, M. Kanno, A. Kobayashi, Y. Ohnishi, Y. Ito and Y. Nakahara, *Tetrahedron Lett.*, 1999, **40**, 515-518.
121. L. Li and P. Wang, *Tetrahedron Lett.*, 2007, **48**, 29-32.
122. D. Lelièvre, P. Barta, V. Aucagne and A. F. Delmas, *Tetrahedron Lett.*, 2008, **49**, 4016-4019.
123. S. Ficht, R. J. Payne, R. T. Guy and C.-H. Wong, *Chem. Eur. J.*, 2008, **14**, 3620-3629.
124. J. Alsina, T. S. Yokum, F. Albericio and G. Barany, *J. Org. Chem.*, 1999, **64**, 8761-8769.
125. R. Raz and J. r. Rademann, *Org. Lett.*, 2011, **13**, 1606-1609.
126. B. J. Backes and J. A. Ellman, *J. Org. Chem.*, 1999, **64**, 2322-2330.
127. S. Mezzato, M. Schaffrath and C. Unverzagt, *Angew. Chem.Int. Ed.*, 2005, **44**, 1650-1654.
128. D. Macmillan and C. R. Bertozzi, *Angew. Chem.Int. Ed.*, 2004, **43**, 1355-1359.
129. Y. Shin, K. A. Winans, B. J. Backes, S. B. H. Kent, J. A. Ellman and C. R. Bertozzi, *J. Am. Chem. Soc.*, 1999, **121**, 11684-11689.
130. J. P. Richardson, C. H. Chan, J. Blanc, M. Saadi and D. Macmillan, *Org. Biomol. Chem.*, 2010, **8**, 1351-1360.
131. J. A. Camarero, B. J. Hackel, J. J. de Yoreo and A. R. Mitchell, *J. Org. Chem.*, 2004, **69**, 4145-4151.
132. J. B. Blanco-Canosa and P. E. Dawson, *Angew. Chem.Int. Ed.*, 2008, **47**, 6851-6855.
133. A. P. Tofteng, K. K. Sørensen, K. W. Conde-Frieboes, T. Hoeg-Jensen and K. J. Jensen, *Angew. Chem.Int. Ed.*, 2009, **48**, 7411-7414.
134. P. Wang, R. Layfield, M. Landon, R. J. Mayer and R. Ramage, *Tetrahedron Lett.*, 1998, **39**, 8711-8714.

135. G.-M. Fang, Y.-M. Li, F. Shen, Y.-C. Huang, J.-B. Li, Y. Lin, H.-K. Cui and L. Liu, *Angew. Chem.Int. Ed.*, 2011, **50**, 7645-7649.
136. G.-M. Fang, J.-X. Wang and L. Liu, *Angew. Chem.Int. Ed.*, 2012, **51**, 10347-10350.
137. P. Botti, M. Villain, S. Manganiello and H. Gaertner, *Org. Lett.*, 2004, **6**, 4861-4864.
138. J.-S. Zheng, W.-X. Xi, F.-L. Wang, J. Li and Q.-X. Guo, *Tetrahedron Lett.*, 2011, **52**, 2655-2660.
139. J. Vizzavona, F. Dick and T. Vorherr, *Bioorg. Med. Chem. Lett.*, 2002, **12**, 1963-1965.
140. T. Kawakami, M. Sumida, K. i. Nakamura, T. Vorherr and S. Aimoto, *Tetrahedron Lett.*, 2005, **46**, 8805-8807.
141. A. Romanelli, A. Shekhtman, D. Cowburn and T. W. Muir, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 6397-6402.
142. D. A. Evans, J. C. Anderson and M. K. Taylor, *Tetrahedron Lett.*, 1993, **34**, 5563-5566.
143. Y. Ohta, S. Itoh, A. Shigenaga, S. Shintaku, N. Fujii and A. Otaka, *Org. Lett.*, 2006, **8**, 467-470.
144. S. Tsuda, A. Shigenaga, K. Bando and A. Otaka, *Org. Lett.*, 2009, **11**, 823-826.
145. N. Ollivier, J. Dheur, R. Mhidia, A. Blanpain and O. Melnyk, *Org. Lett.*, 2010, **12**, 5238-5241.
146. J. Dheur, N. Ollivier, A. l. Vallin and O. Melnyk, *J. Org. Chem.*, 2011, **76**, 3194-3202.
147. W. Hou, X. Zhang, F. Li and C.-F. Liu, *Org. Lett.*, 2010, **13**, 386-389.
148. F. Nagaike, Y. Onuma, C. Kanazawa, H. Hojo, A. Ueki, Y. Nakahara and Y. Nakahara, *Org. Lett.*, 2006, **8**, 4465-4468.
149. H. Hojo, Y. Onuma, Y. Akimoto, Y. Nakahara and Y. Nakahara, *Tetrahedron Lett.*, 2007, **48**, 25-28.
150. C. Ozawa, H. Katayama, H. Hojo, Y. Nakahara and Y. Nakahara, *Org. Lett.*, 2008, **10**, 3531-3533.
151. R. K. Sharma and J. P. Tam, *Org. Lett.*, 2011, **13**, 5176-5179.
152. G. Zanotti, F. Pinnen and G. Lucente, *Tetrahedron Lett.*, 1985, **26**, 5481-5484.
153. T. Kawakami and S. Aimoto, *Tetrahedron Lett.*, 2007, **48**, 1903-1905.
154. J.-S. Zheng, H.-N. Chang, F.-L. Wang and L. Liu, *J. Am. Chem. Soc.*, 2011, **133**, 11080-11083.
155. J. Kang, J. P. Richardson and D. Macmillan, *Chem. Comm.*, 2009, 407-409.
156. J. Masania, J. J. Li, S. J. Smerdon and D. Macmillan, *Org. Biomol. Chem.*, 2010, **8**, 5113-5119.
157. D. Macmillan, M. De Cecco, N. L. Reynolds, L. F. A. Santos, P. E. Barran and J. R. Dorin, *Chembiochem*, 2011, **12**, 2133-2136.
158. A. L. Adams and D. Macmillan, *J. Pept. Sci.*, 2013, **19**, 65-73.
159. S. C. Makrides, *Microbiol. Rev.*, 1996, **60**, 512-538.
160. M.-Q. Xu and F. B. Perler, *EMBO J.*, 1996, **15**, 5146.
161. S. Chong, F. B. Mersha, D. G. Comb, M. E. Scott, D. Landry, L. M. Vence, F. B. Perler, J. Benner, R. B. Kucera, C. A. Hirvonen, J. J. Pelletier, H. Paulus and M.-Q. Xu, *Gene*, 1997, **192**, 271-281.
162. T. W. Muir, D. Sondhi and P. A. Cole, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 6705-6710.
163. W. Lu, D. Gong, D. Bar-Sagi and P. A. Cole, *Mol. Cell*, 2001, **8**, 759-769.
164. T. Durek, K. Alexandrov, R. S. Goody, A. Hildebrand, I. Heinemann and H. Waldmann, *J. Am. Chem. Soc.*, 2004, **126**, 16368-16378.
165. M. Vila-Perelló, M. R. Pratt, F. Tulin and T. W. Muir, *J. Am. Chem. Soc.*, 2007, **129**, 8068-8069.
166. M. E. Hahn, J.-P. Pellois, M. Vila-Perelló and T. W. Muir, *Chembiochem*, 2007, **8**, 2100-2105.

167. R. K. McGinty, J. Kim, C. Chatterjee, R. G. Roeder and T. W. Muir, *Nature*, 2008, **453**, 812-816.
168. K. A. Pickin, S. Chaudhury, B. C. R. Dancy, J. J. Gray and P. A. Cole, *J. Am. Chem. Soc.*, 2008, **130**, 5667-5669.
169. S. Frutos, M. Goger, B. Giovani, D. Cowburn and T. W. Muir, *Nat. Chem. Biol.*, 2010, **6**, 527-533.
170. N. H. Shah and T. W. Muir, *Chem. Sci.*, 2014, **5**, 446-461.
171. J. Thom, D. Anderson, J. McGregor and G. Cotton, *Bioconjugate Chem.*, 2011, **22**, 1017-1020.
172. F. I. Valiyaveetil, R. MacKinnon and T. W. Muir, *J. Am. Chem. Soc.*, 2002, **124**, 9113-9120.
173. S. K. Mazmanian, G. Liu, H. Ton-That and O. Schneewind, *Science*, 1999, **285**, 760-763.
174. J. J. Ling, R. L. Policarpo, A. E. Rabideau, X. Liao and B. L. Pentelute, *J. Am. Chem. Soc.*, 2012, **134**, 10749-10752.
175. Y.-M. Li, Y.-T. Li, M. Pan, X.-Q. Kong, Y.-C. Huang, Z.-Y. Hong and L. Liu, *Angew. Chem.*, 2014, **126**, 2230-2234.
176. J. M. Antos, G.-L. Chew, C. P. Guimaraes, N. C. Yoder, G. M. Grotenbreg, M. W.-L. Popp and H. L. Ploegh, *J. Am. Chem. Soc.*, 2009, **131**, 10800-10801.
177. K. Piotukh, B. Geltinger, N. Heinrich, F. Gerth, M. Beyermann, C. Freund and D. Schwarzer, *J. Am. Chem. Soc.*, 2011, **133**, 17536-17539.
178. C. Noren, S. Anthony-Cahill, M. Griffith and P. Schultz, *Science*, 1989, **244**, 182-188.
179. L. Wang, A. Brock, B. Herberich and P. G. Schultz, *Science*, 2001, **292**, 498-500.
180. J. Guo, J. Wang, J. C. Anderson and P. G. Schultz, *Angew. Chem.Int. Ed.*, 2008, **47**, 722-725.
181. Y.-M. Li, M.-Y. Yang, Y.-C. Huang, Y.-T. Li, P. R. Chen and L. Liu, *ACS Chem. Biol.*, 2012, **7**, 1015-1022.
182. J. P. Waller, *J. Mol. Biol.*, 1963, **7**, 483-IN481.
183. A. Ben-Bassat, K. Bauer, S.-Y. Chang, K. Myambo, A. Boosman and S. Chang, *J. Bacteriol.*, 1987, **169**, 751-757.
184. J. A. Camarero, D. Fushman, D. Cowburn and T. W. Muir, *Bioorg. Med. Chem.*, 2001, **9**, 2479-2484.
185. A. Looman, J. Bodlaender, L. Comstock, D. Eaton, P. Jhurani, H. De Boer and P. Van Knippenberg, *EMBO J.*, 1987, **6**, 2489.
186. P. H. Hirel, M. J. Schmitter, P. Dessen, G. Fayat and S. Blanquet, *Proc. Natl. Acad. Sci. U. S. A.*, 1989, **86**, 8247-8251.
187. D. A. Erlanson, M. Chytil and G. L. Verdine, *Chem. Biol.*, 1996, **3**, 981-991.
188. T. J. Tolbert and C.-H. Wong, *Angew. Chem.Int. Ed.*, 2002, **41**, 2171-2174.
189. R. B. Kapust, J. Tözsér, T. D. Copeland and D. S. Waugh, *Biochem. Biophys. Res. Commun.*, 2002, **294**, 949-955.
190. G. K. Busch, E. W. Tate, P. R. J. Gaffney, E. Rosivatz, R. Woscholski and R. J. Leatherbarrow, *Chem. Comm.*, 2008, 3369-3371.
191. T. C. Evans, J. Benner and M.-Q. Xu, *J. Biol. Chem.*, 1999, **274**, 18359-18363.
192. D. Macmillan and L. Arham, *J. Am. Chem. Soc.*, 2004, **126**, 9530-9531.
193. A. Kuliopulos and C. T. Walsh, *J. Am. Chem. Soc.*, 1994, **116**, 4599-4607.
194. T. M. Hackeng, J. H. Griffin and P. E. Dawson, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 10068-10073.
195. L. Urge, E. Kollat, M. Hollosi, I. Laczko, K. Wroblewski, J. Thurin and L. Otvos Jr, *Tetrahedron Lett.*, 1991, **32**, 3445-3448.
196. Y. Asahina, M. Kanda, A. Suzuki, H. Katayama, Y. Nakahara and H. Hojo, *Org. Biomol. Chem.*, 2013, **11**, 7199-7207.

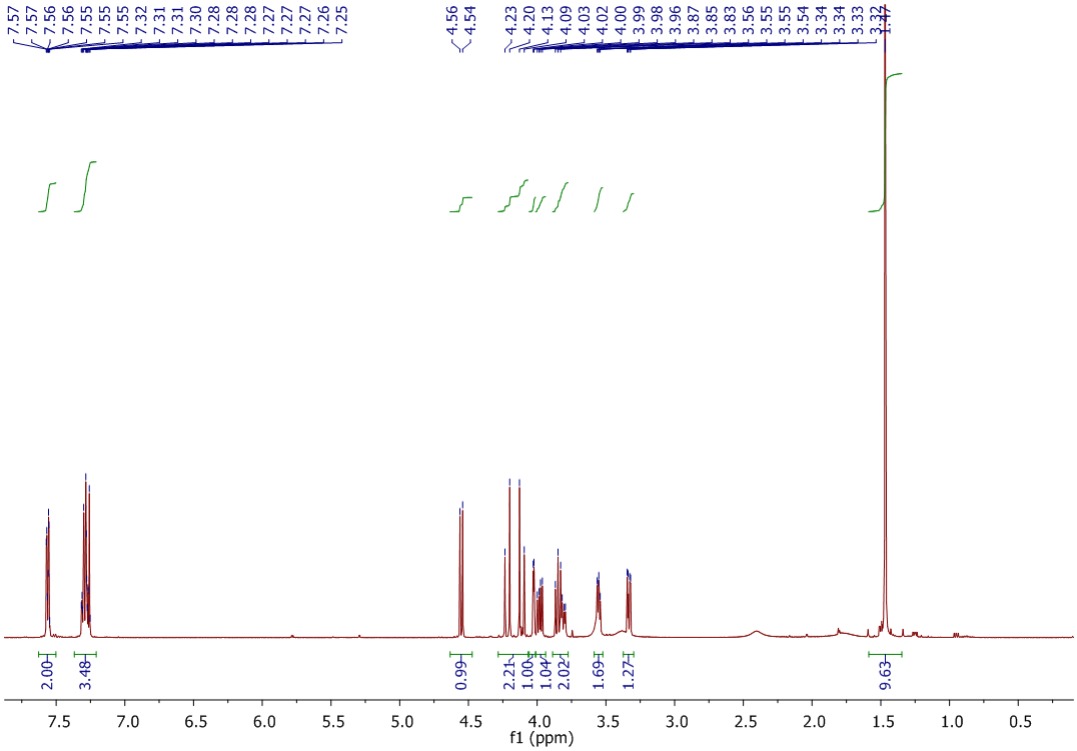
197. T. Conroy, K. A. Jolliffe and R. J. Payne, *Org. Biomol. Chem.*, 2009, **7**, 2255-2258.
198. K. J. Doores, Y. Mimura, R. A. Dwek, P. M. Rudd, T. Elliott and B. G. Davis, *Chem. Comm.*, 2006, 1401-1403.
199. F. Schwarz, W. Huang, C. Li, B. L. Schulz, C. Lizak, A. Palumbo, S. Numao, D. Neri, M. Aebi and L.-X. Wang, *Nat. Chem. Biol.*, 2010, **6**, 264-266.
200. F. Liu, B. Vijayakrishnan, A. Faridmoayer, T. A. Taylor, T. B. Parsons, G. J. L. Bernardes, M. Kowarik and B. G. Davis, *J. Am. Chem. Soc.*, 2013, **136**, 566-569.
201. D. Horton, *Org. Synth.*, 1966, 1-1.
202. F. D. Tropper, F. O. Andersson, S. Braun and R. Roy, *Synthesis*, 1992, **1992**, 618-620.
203. D. Macmillan, A. M. Daines, M. Bayrhuber and S. L. Flitsch, *Org. Lett.*, 2002, **4**, 1467-1470.
204. P. Kirsch, N. Kusunose, J.-i. Aikawa, T. Kigawa, S. Yokoyama and T. Ogawa, *Bioorg. Med. Chem.*, 1995, **3**, 1631-1636.
205. K. M. Harris, S. Flemer and R. J. Hondal, *J. Pept. Sci.*, 2007, **13**, 81-93.
206. H. Rohde, J. Schmalisch, Z. Harpaz, F. Diezmann and O. Seitz, *Chembiochem*, 2011, **12**, 1396-1400.
207. T. Haack and M. Mutter, *Tetrahedron Lett.*, 1992, **33**, 1589-1592.
208. J. Kang, N. L. Reynolds, C. Tyrrell, J. R. Dorin and D. Macmillan, *Org. Biomol. Chem.*, 2009, **7**, 4918-4923.
209. A. L. Adams, B. Cowper, R. E. Morgan, B. Premdjee, S. Caddick and D. Macmillan, *Angew. Chem.Int. Ed.*, 2013, **52**, 13062-13066.
210. J.-S. Zheng, S. Tang, Y. Guo, H.-N. Chang and L. Liu, *Chembiochem*, 2012, **13**, 542-546.
211. M.-Q. Xu and T. C. Evans Jr, *Methods*, 2001, **24**, 257-277.
212. E. R. Lavallie, E. A. DiBlasio, S. Kovacic, K. L. Grant, P. F. Schendel and J. M. McCoy, *Nat. Biotech.*, 1993, **11**, 187-193.
213. C. Aslanidis and P. J. de Jong, *Nucleic Acids Res.*, 1990, **18**, 6069-6074.
214. J. Porath and B. Olin, *Biochemistry*, 1983, **22**, 1621-1630.
215. J. P. Richardson and D. Macmillan, *Org. Biomol. Chem.*, 2008, **6**, 3977-3982.
216. Y. A. Andreev, S. A. Kozlov, A. A. Vassilevski and E. V. Grishin, *Anal. Biochem.*, 2010, **407**, 144-146.
217. Q. Zhang, E. V. Johnston, J.-H. Shieh, M. A. S. Moore and S. J. Danishefsky, *Proc. Natl. Acad. Sci. U. S. A.*, 2014.
218. A. L. Tarentino and F. Maley, *Arch. Biochem. Biophys.*, 1969, **130**, 295-303.
219. A. Tarentino, T. H. Plummer and F. Maley, *J. Biol. Chem.*, 1970, **245**, 4150-4157.
220. A. L. Tarentino, T. H. Plummer and F. Maley, *J. Biol. Chem.*, 1972, **247**, 2629-2631.
221. A. Kobata, *Anal. Biochem.*, 1979, **100**, 1-14.
222. K. Takegawa, S. Yamaguchi, A. Kondo, H. Iwamoto, M. Nakoshi, I. Kato and S. Iwahara, *Biochem. Int.*, 1991, **24**, 849-855.
223. K. J. Yamamoto, S. Kadowaki, J. Watanabe and H. Kumagai, *Biochem. Biophys. Res. Commun.*, 1994, **203**, 244-252.
224. H. Muramatsu, H. Tachikui, H. Ushida, X.-j. Song, Y. Qiu, S. Yamamoto and T. Muramatsu, *J. Biochem.*, 2001, **129**, 923-928.
225. K. Fujita, T. Miyamura, M. Sano, I. Kato and K. Takegawa, *J. Biosci. Bioeng.*, 2002, **93**, 614-617.
226. K. Takegawa, M. Tabuchi, S. Yamaguchi, A. Kondo , I. Kato and S. Iwahara, *J. Biol. Chem.*, 1995, **270**, 3094-3099.
227. B. Li, H. Song, S. Hauser and L.-X. Wang, *Org. Lett.*, 2006, **8**, 3081-3084.
228. T. W. D. F. Rising, C. D. Heidecke, J. W. B. Moir, Z. Ling and A. J. Fairbanks, *Chem. Eur. J.*, 2008, **14**, 6444-6464.

229. T. B. Parsons, J. W. B. Moir and A. J. Fairbanks, *Org. Biomol. Chem.*, 2009, **7**, 3128-3140.
230. W. Huang, C. Li, B. Li, M. Umekawa, K. Yamamoto, X. Zhang and L.-X. Wang, *J. Am. Chem. Soc.*, 2009, **131**, 2214-2223.
231. W. Huang, Q. Yang, M. Umekawa, K. Yamamoto and L.-X. Wang, *Chembiochem*, 2010, **11**, 1350-1355.
232. M. Umekawa, W. Huang, B. Li, K. Fujita, H. Ashida, L.-X. Wang and K. Yamamoto, *J. Biol. Chem.*, 2008, **283**, 4469-4479.
233. C. D. Heidecke, Z. L. Ling, N. C. Bruce, J. W. B. Moir, T. B. Parsons and A. J. Fairbanks, *Chembiochem*, 2008, **9**, 2045-2051.
234. G. M. Watt and G.-J. Boons, *Carbohydr. Res.*, 2004, **339**, 181-193.
235. M. V. Chiesa and Richard R. Schmidt, *Eur. J. Org. Chem.*, 2000, **2000**, 3541-3554.
236. I. Matsuo, M. Isomura, R. Walton and K. Ajisaka, *Tetrahedron Lett.*, 1996, **37**, 8795-8798.
237. M. Sakagami and H. Hamana, *Tetrahedron Lett.*, 2000, **41**, 5547-5551.
238. M. Noguchi, T. Tanaka, H. Gyakushi, A. Kobayashi and S.-i. Shoda, *J. Org. Chem.*, 2009, **74**, 2210-2212.
239. M. Colon, M. M. Staveski and J. T. Davis, *Tetrahedron Lett.*, 1991, **32**, 4447-4450.
240. K. Fujita, N. Tanaka, M. Sano, I. Kato, Y. Asada and K. Takegawa, *Biochem. Biophys. Res. Commun.*, 2000, **267**, 134-138.
241. K. Fujita and K. Takegawa, *Biochem. Biophys. Res. Commun.*, 2001, **283**, 680-686.
242. J. J. Holst, *Physiol. Rev.*, 2007, **87**, 1409-1439.
243. T. Ueda, K. Tomita, Y. Notsu, T. Ito, M. Fumoto, T. Takakura, H. Nagatome, A. Takimoto, S.-I. Mihara, H. Togame, K. Kawamoto, T. Iwasaki, K. Asakura, T. Oshima, K. Hanasaki, S.-I. Nishimura and H. Kondo, *J. Am. Chem. Soc.*, 2009, **131**, 6237-6245.
244. A. G. Morell, G. Gregoriadis, I. H. Scheinberg, J. Hickman and G. Ashwell, *J. Biol. Chem.*, 1971, **246**, 1461-1467.
245. C. T. Yuen, A. M. Lawson, W. Chai, M. Larkin, M. S. Stoll, A. C. Stuart, F. X. Sullivan, T. J. Ahern and T. Feizi, *Biochemistry*, 1992, **31**, 9126-9131.
246. H. Huang and C.-H. Wong, *J. Org. Chem.*, 1995, **60**, 3100-3106.
247. J. C. Prodger, M. J. Bamford, M. I. Bird, P. M. Gore, D. S. Holmes, R. Priest and V. Saez, *Bioorg. Med. Chem.*, 1996, **4**, 793-801.
248. W. S. Somers, J. Tang, G. D. Shaw and R. T. Camphausen, *Cell*, 2000, **103**, 467-479.
249. J. C. Paulson and K. J. Colley, *J. Biol. Chem.*, 1989, **264**, 17615-17618.
250. L. F. Leloir, *Science*, 1971, **172**, 1299-1303.
251. J. Seibel, R. Beine, R. Moraru, C. Behringer and K. Buchholz, *Biocatal. Biotransfor.*, 2006, **24**, 157-165.
252. C. Breton, L. Šnajdrová, C. Jeanneau, J. Koča and A. Imberty, *Glycobiology*, 2006, **16**, 29R-37R.
253. B. W. Murray, S. Takayama, J. Schultz and C.-H. Wong, *Biochemistry*, 1996, **35**, 11183-11195.
254. L. L. Lairson and S. G. Withers, *Chem. Comm.*, 2004.
255. L. L. Lairson, B. Henrissat, G. J. Davies and S. G. Withers, in *Annu. Rev. Biochem.*, Annual Reviews, Palo Alto, 2008, pp. 521-555.
256. C. J. Thibodeaux, C. E. Melancon and H.-w. Liu, *Nature*, 2007, **446**, 1008-1016.
257. G. K. Wagner, T. Pesnot and R. A. Field, *Natural Product Reports*, 2009, **26**, 1172-1194.
258. M. S. Motawia, J. Marcussen and B. L. Muller, *J. Carbohydr. Chem.*, 1995, **14**, 1279-1294.
259. C. Xu, H. Liu and X. Li, *Carbohydr. Res.*, 2011, **346**, 1149-1153.
260. J. G. Moffatt and H. G. Khorana, *J. Am. Chem. Soc.*, 1961, **83**, 649-658.

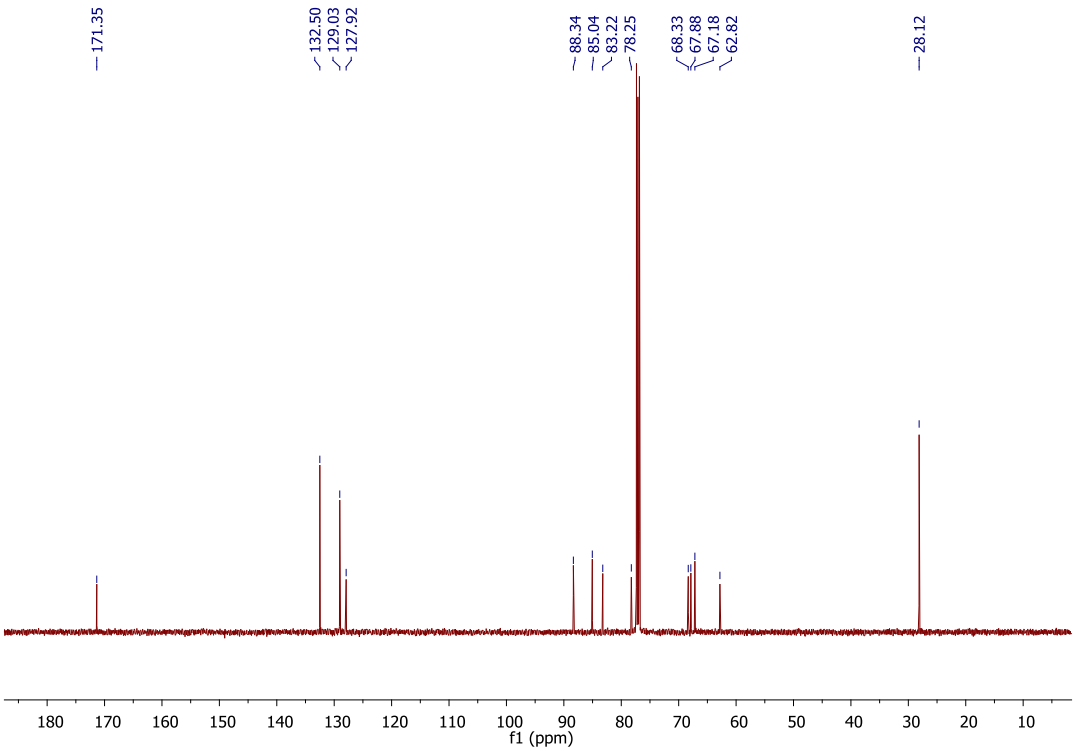
261. S. Roseman, J. J. Distler, J. G. Moffatt and H. G. Khorana, *J. Am. Chem. Soc.*, 1961, **83**, 659-663.
262. C. Dini, N. Drochon, P. Ferrari and J. Aszodi, *Bioorg. Med. Chem. Lett.*, 2000, **10**, 143-145.
263. A. Babič and S. Pečar, *Tetrahedron Lett.*, 2007, **48**, 4403-4405.
264. K. Furuhata, K. Komiyama, H. Ogura and T. Hata, *Chem. Pharm. Bull.*, 1991, **39**, 255-259.
265. S. Marchesan and D. Macmillan, *Chem. Comm.*, 2008, 4321-4323.
266. J. Debenham, R. Rodebaugh and B. Fraser-Reid, *Liebigs Ann.*, 1997, **1997**, 791-802.
267. G. S. Marks, R. D. Marshall and A. Neuberger, *Biochem. J.*, 1963, **87**, 274-270.
268. T. Koide, H. Itoh, A. Otaka, H. Yasui, M. Kuroda, N. Esaki, K. Soda and N. Fujii, *Chem. Pharm. Bull.*, 1993, **41**, 502-506.
269. L. M. Lerner, *Carbohydr. Res.*, 1996, **282**, 189-192.
270. Y. E. Tsvetkov, N. É. Byramova and L. V. Backinowsky, *Carbohydr. Res.*, 1983, **115**, 254-258.
271. J. Ohlsson and G. Magnusson, *Carbohydr. Res.*, 2000, **329**, 49-55.
272. B. Aguilera, A. Fernández-Mayoralas and C. Jaramillo, *Tetrahedron*, 1997, **53**, 5863-5876.
273. Y. Meng, Y. Guo, Y. Ling, Y. Zhao, Q. Zhang, X. Zhou, F. Ding and Y. Yang, *Bioorg. Med. Chem.*, 2011, **19**, 5577-5584.

Appendix - NMR spectra of novel compounds

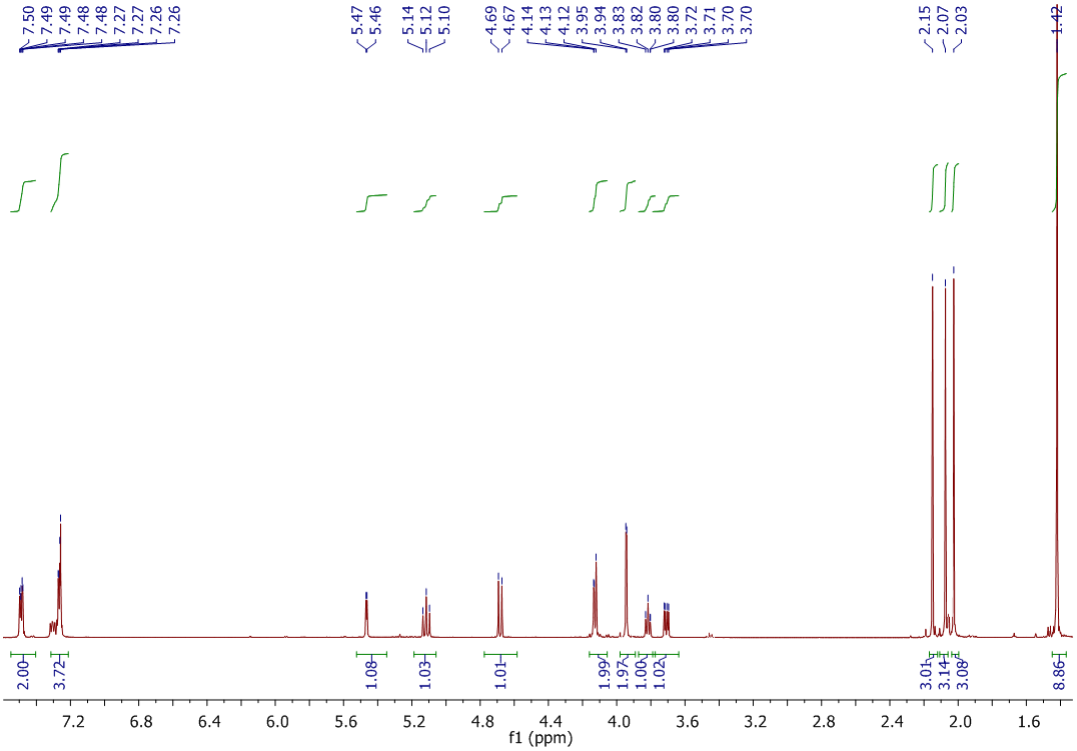
^{1}H NMR (500 MHz, $CDCl_3$) Phenyl 3-*O*-(*tert*-butyloxycarbonyl)methyl 1-thio-β-D-galactopyranoside 53



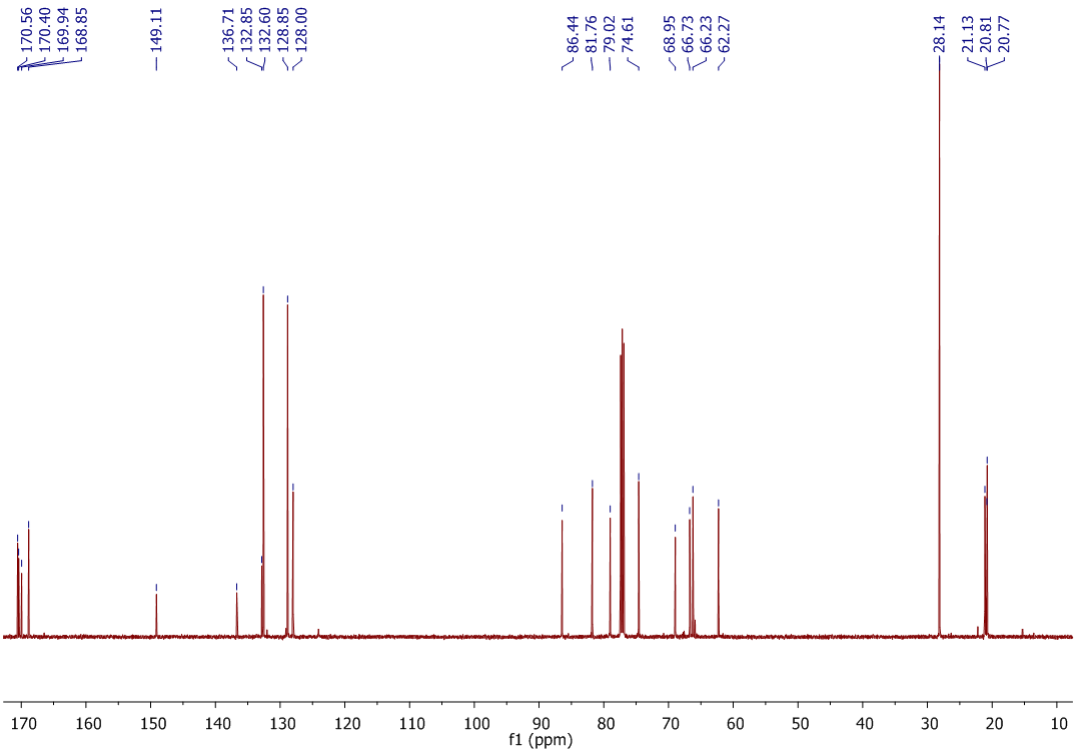
^{13}C NMR (125 MHz, $CDCl_3$) Phenyl 3-*O*-(*tert*-butyloxycarbonyl)methyl 1-thio-β-D-galactopyranoside 53



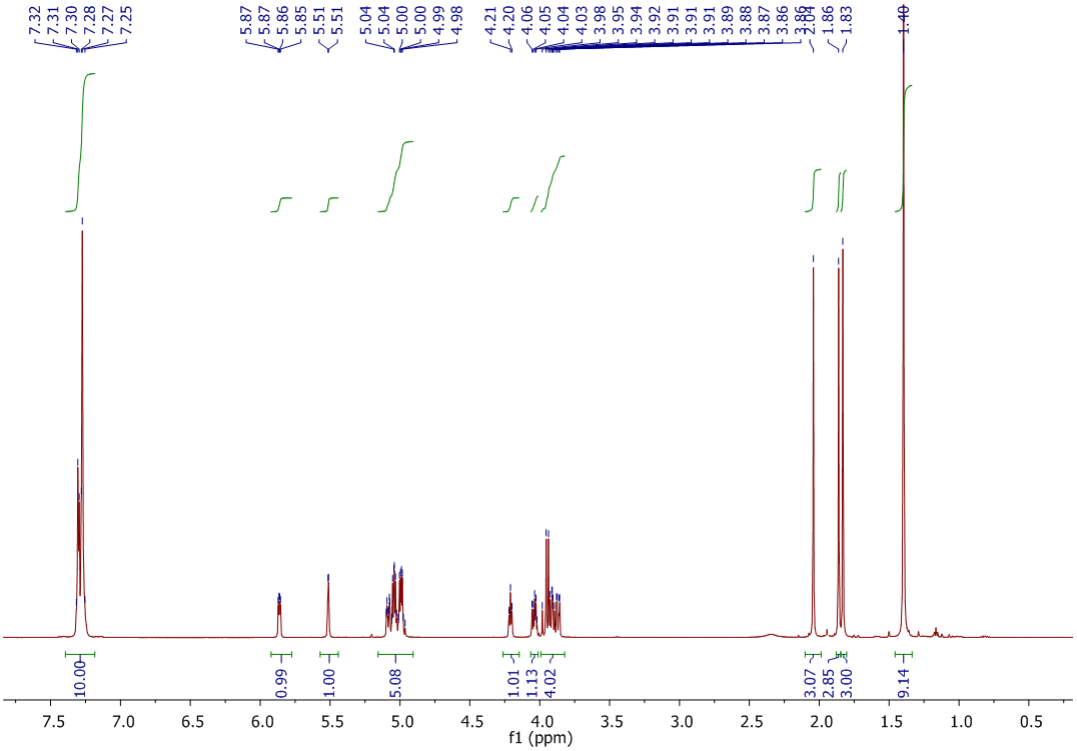
^{1}H NMR (500 MHz, $CDCl_3$) Phenyl 2,4,6-Tri-*O*-acetyl-3-*O*-(*tert*-butyloxycarbonyl)methyl 1-thio-β-D-galactopyranoside 54



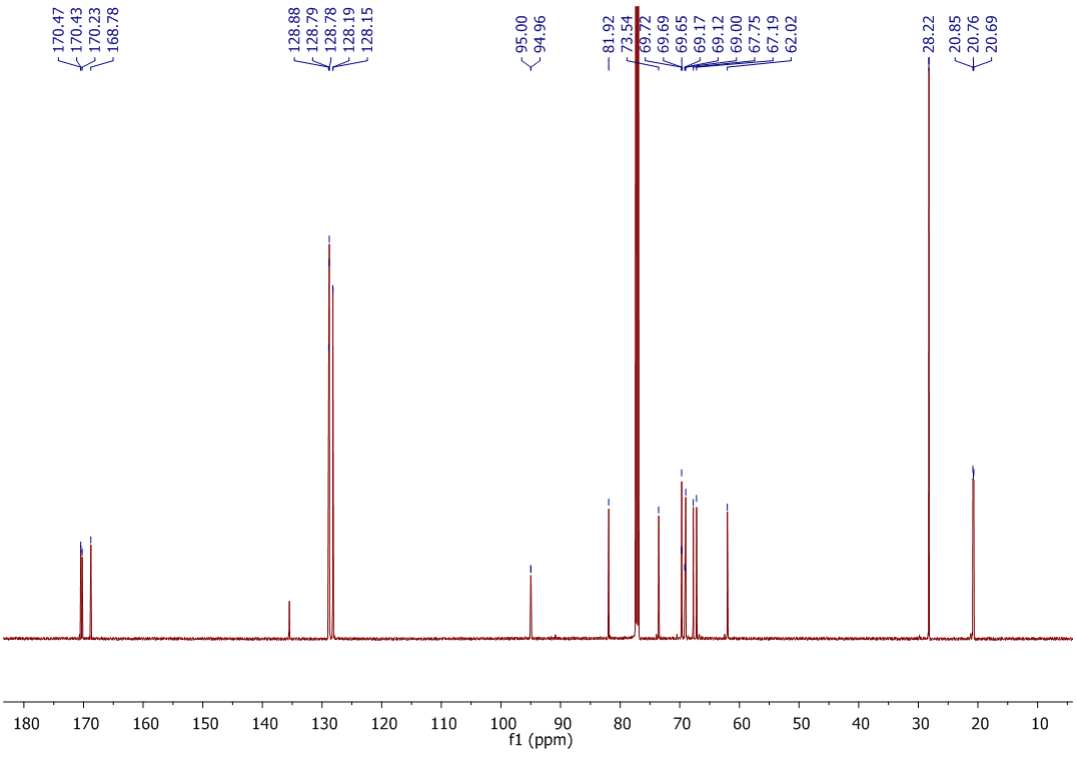
^{13}C NMR (125 MHz, $CDCl_3$) Phenyl 2,4,6-Tri-*O*-acetyl-3-*O*-(*tert*-butyloxycarbonyl)methyl 1-thio-β-D-galactopyranoside 54



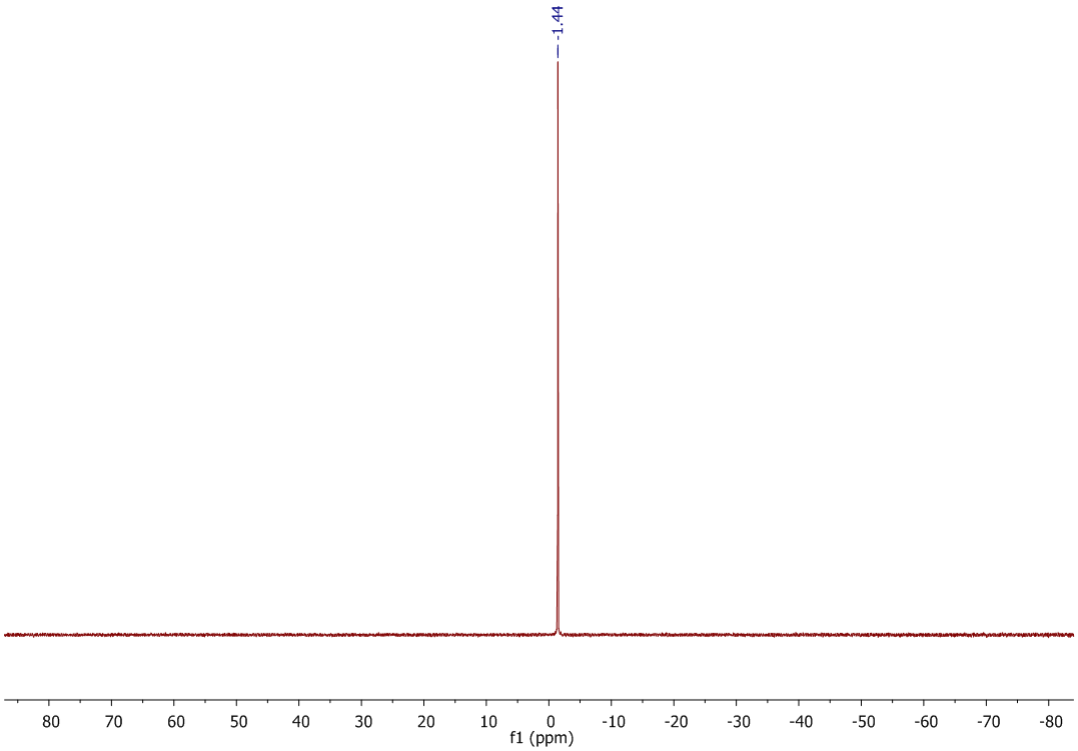
^{1}H NMR (600 MHz, $CDCl_3$) 2,4,6-Tri-*O*-acetyl-3-*O*-(*tert*-butyloxycarbonyl)methyl 1-*O*-diphenoxyphosphoryl-α-D-galactopyranoside 56



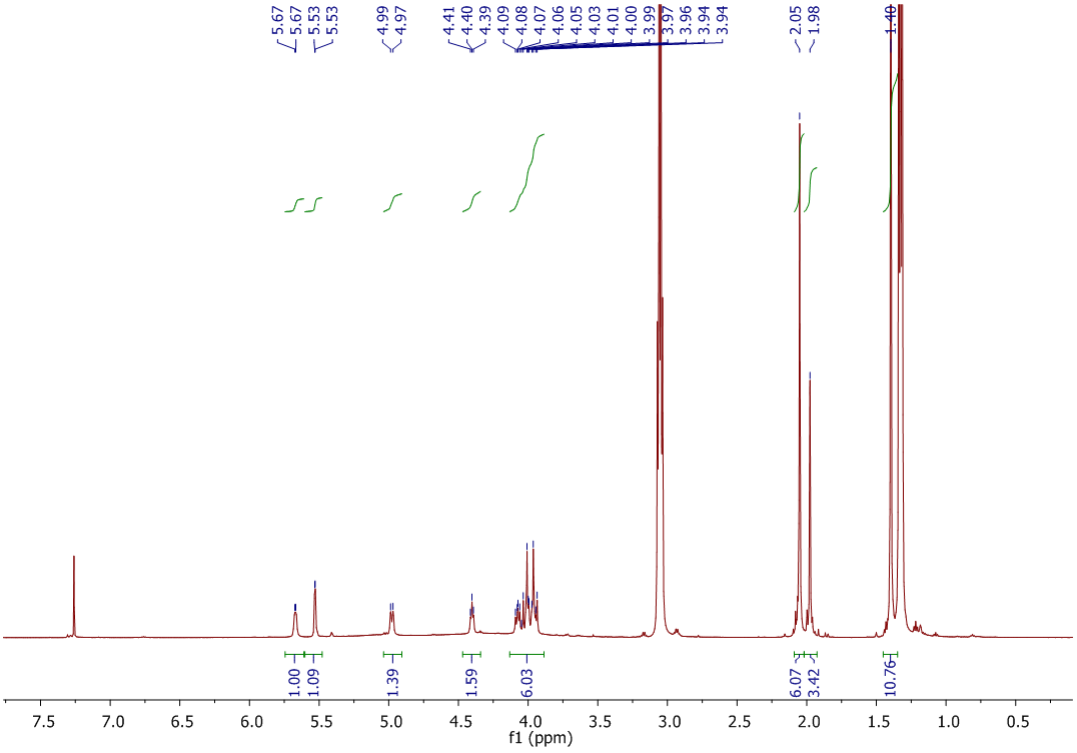
^{13}C NMR (150 MHz, $CDCl_3$) 2,4,6-Tri-*O*-acetyl-3-*O*-(*tert*-butyloxycarbonyl)methyl 1-*O*-diphenoxyphosphoryl-α-D-galactopyranoside 56



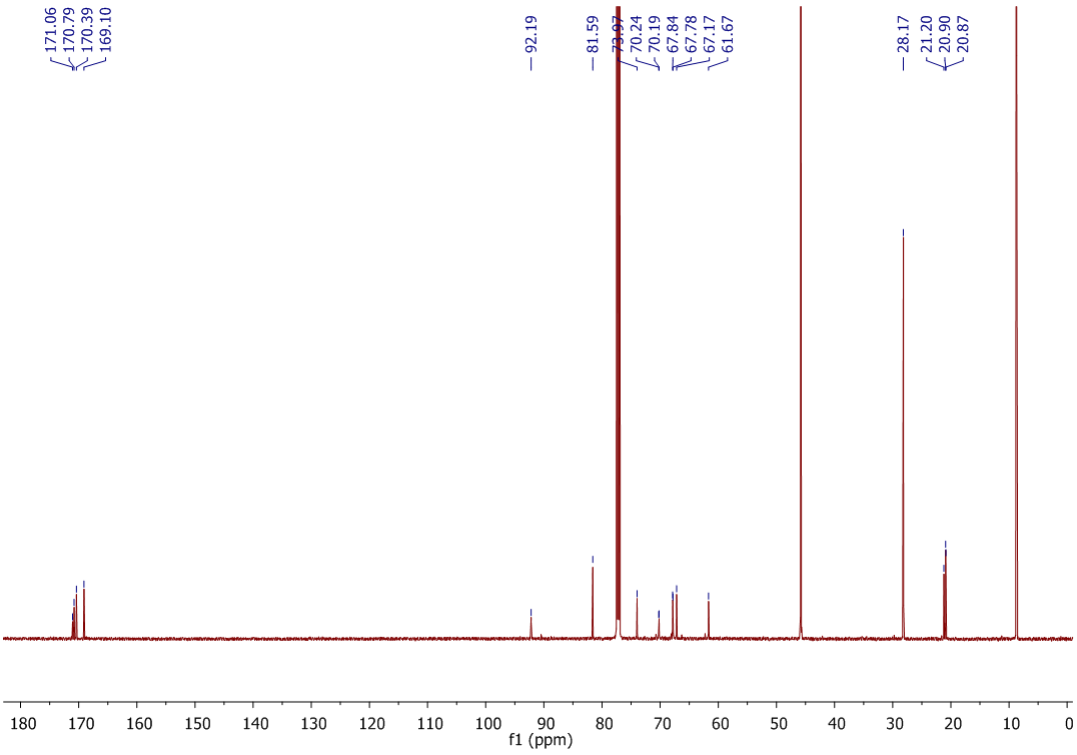
^{31}P NMR (121 MHz, $CDCl_3$) 2,4,6-Tri-*O*-acetyl-3-*O*-(*tert*-butyloxycarbonyl)methyl 1-*O*-diphenoxyphosphoryl-α-D-galactopyranoside 56



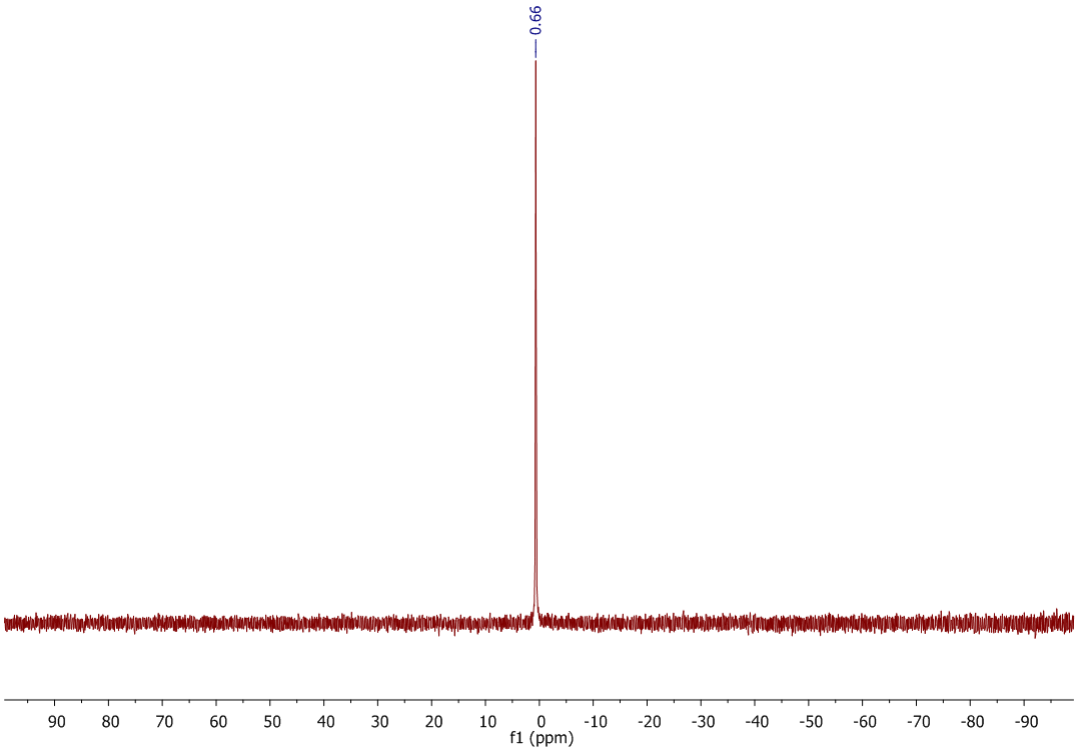
^{1}H NMR (600 MHz, $CDCl_3$) 2,4,6-Tri-*O*-acetyl 3-*O*-(*tert*-butyloxycarbonyl)methyl 1-*O*-phosphoryl-α-D-galactopyranoside (triethylammonium salt) 57



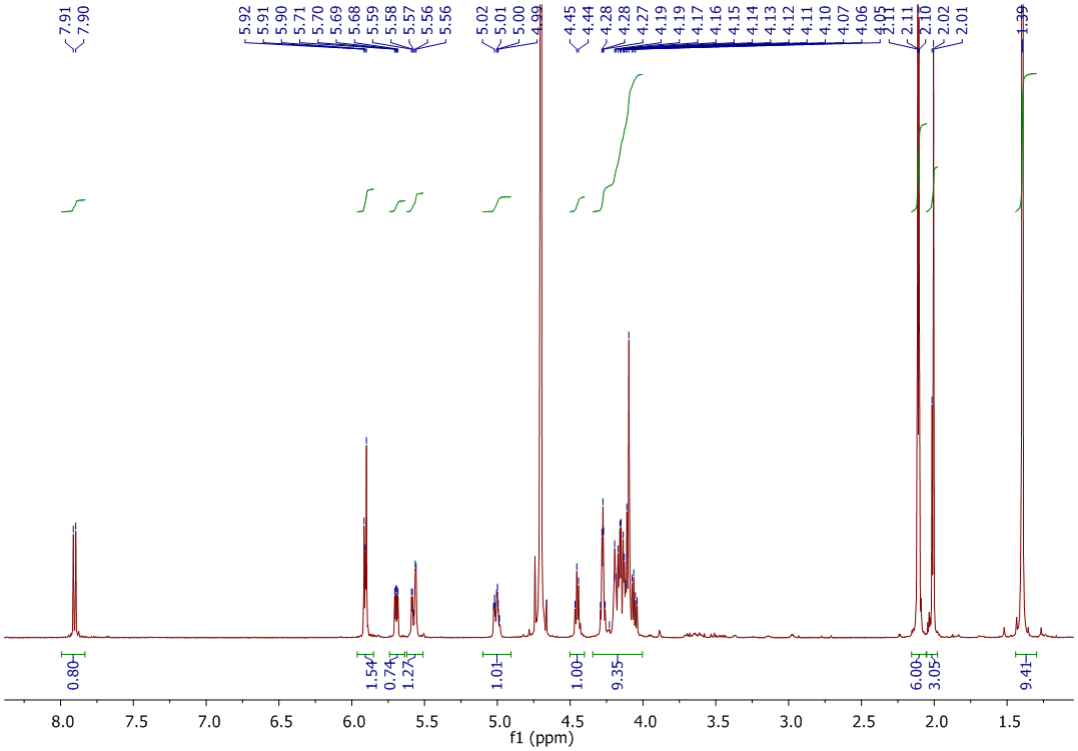
^{13}C NMR (150 MHz, $CDCl_3$) 2,4,6-Tri-*O*-acetyl 3-*O*-(*tert*-butyloxycarbonyl)methyl 1-*O*-phosphoryl-α-D-galactopyranoside (triethylammonium salt) 57



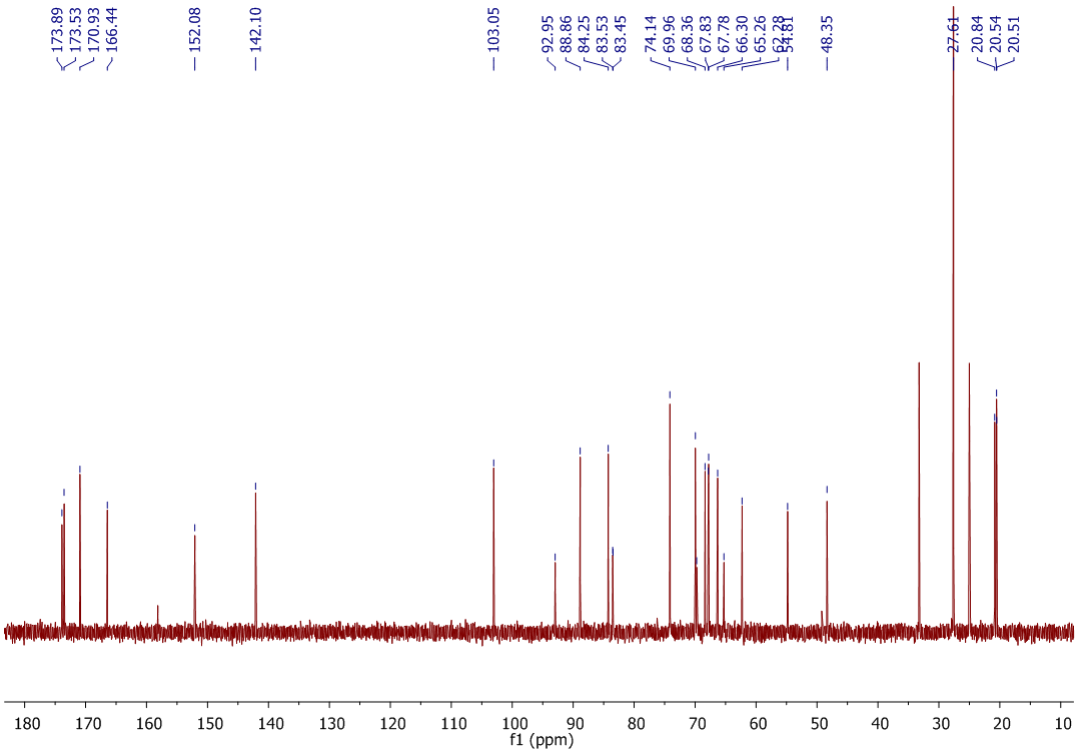
^{31}P NMR (121 MHz, $CDCl_3$) 2,4,6-Tri-*O*-acetyl 3-*O*-(*tert*-butyloxycarbonyl)methyl 1-*O*-phosphoryl-α-D-galactopyranoside (triethylammonium salt) 57



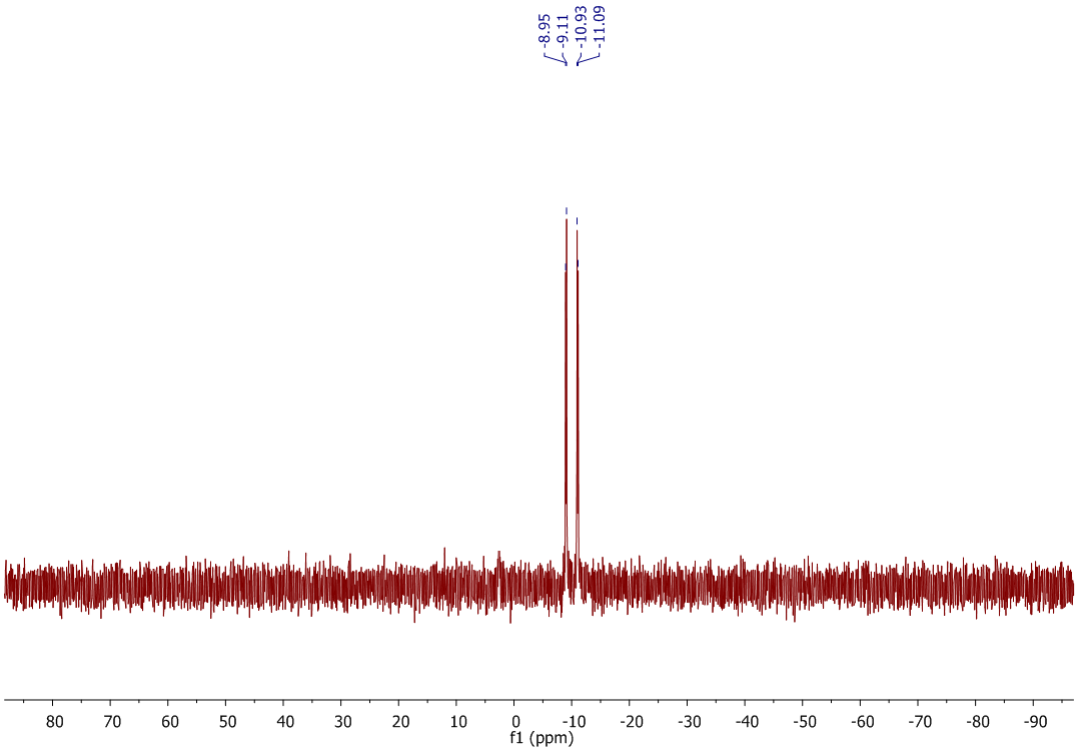
^{1}H NMR (500 MHz, D_2O) 2,4,6-Tri-*O*-acetyl 3-*O*-(*tert*-butyloxycarbonyl)methyl 1-*O*-uridine-5'-diphosphoryl-α-D-galactopyranoside (Na salt) 58



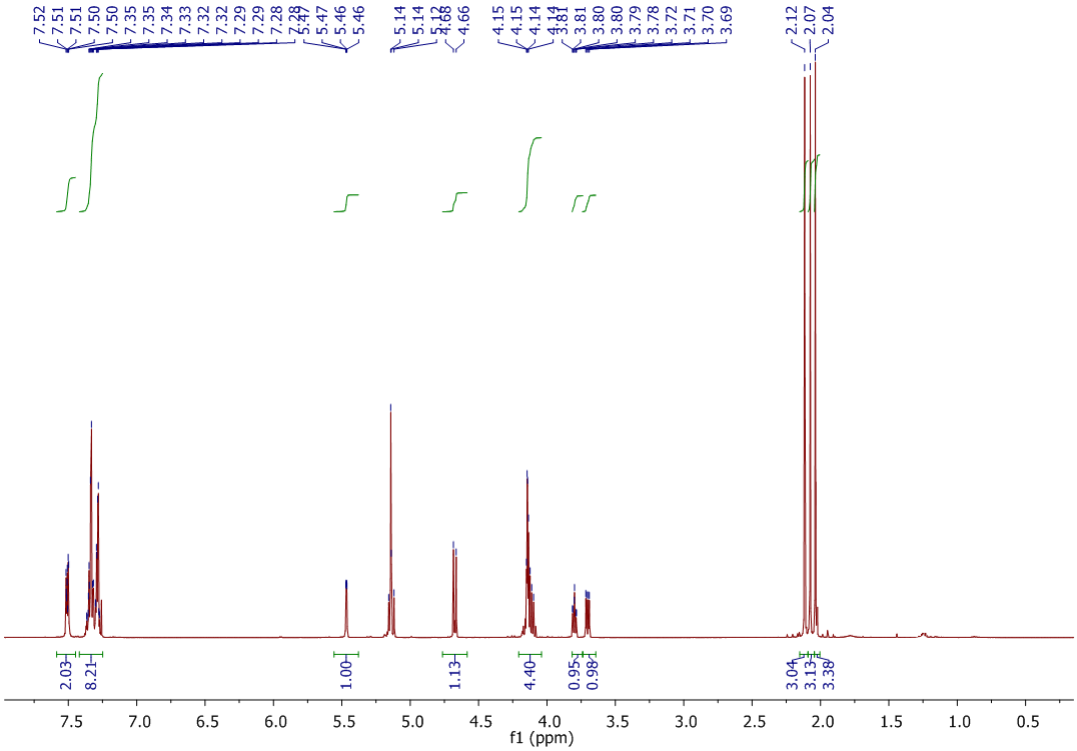
^{13}C NMR (125 MHz, D_2O) 2,4,6-Tri-*O*-acetyl 3-*O*-(*tert*-butyloxycarbonyl)methyl 1-*O*-uridine-5'-diphosphoryl-α-D-galactopyranoside (dicyclohexylcarboxamide salt) 58



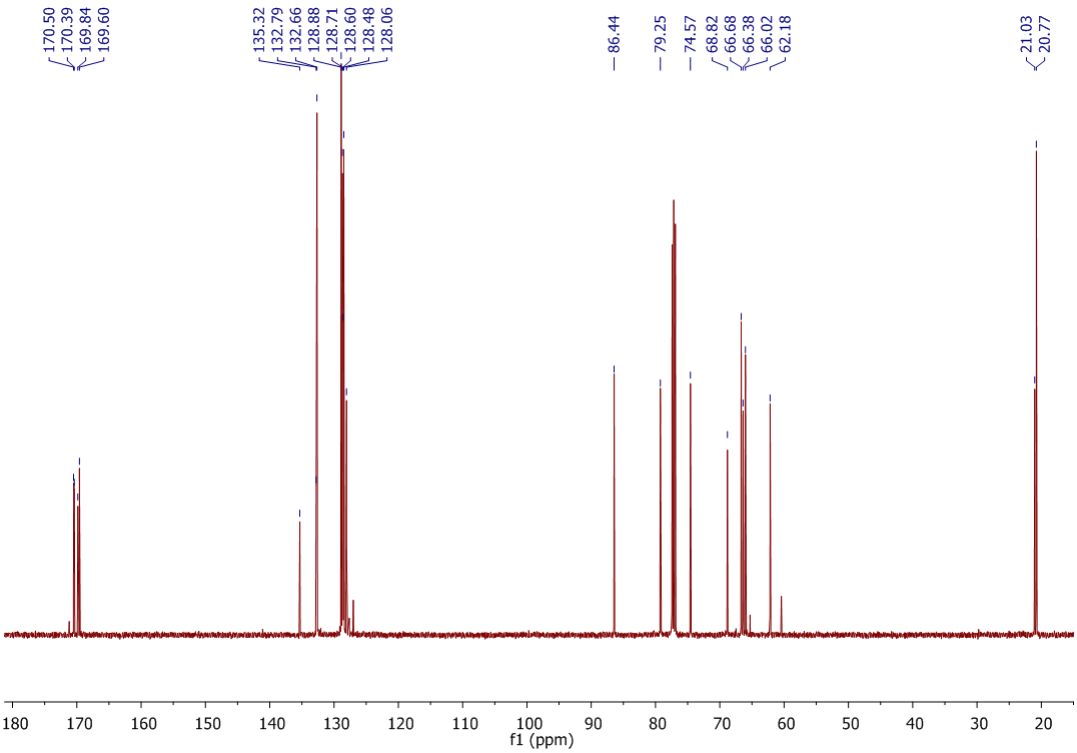
^{31}P NMR (121 MHz, D_2O) 2,4,6-Tri-*O*-acetyl 3-*O*-(*tert*-butyloxycarbonyl)methyl 1-*O*-uridine-5'-diphosphoryl-α-D-galactopyranoside (Na salt) 58



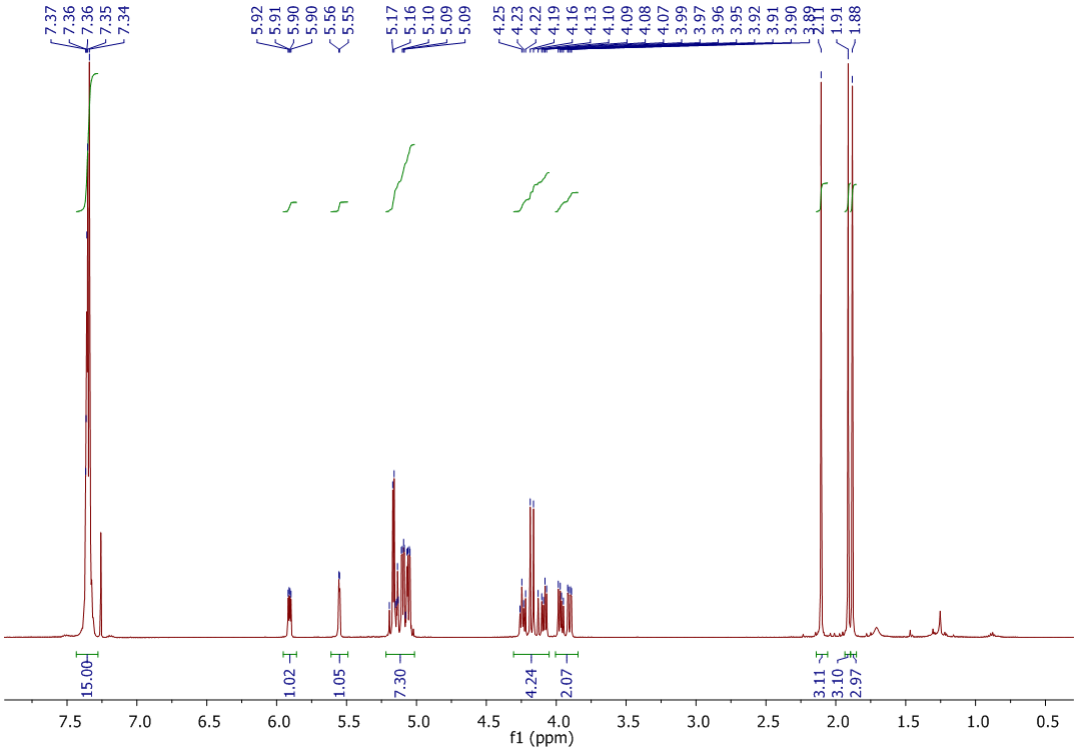
^{1}H NMR (500 MHz, $CDCl_3$) Phenyl 2,4,6-Tri-*O*-acetyl-3-*O*-(benzyl)carboxymethyl 1-thio-β-D-galactopyranoside 59



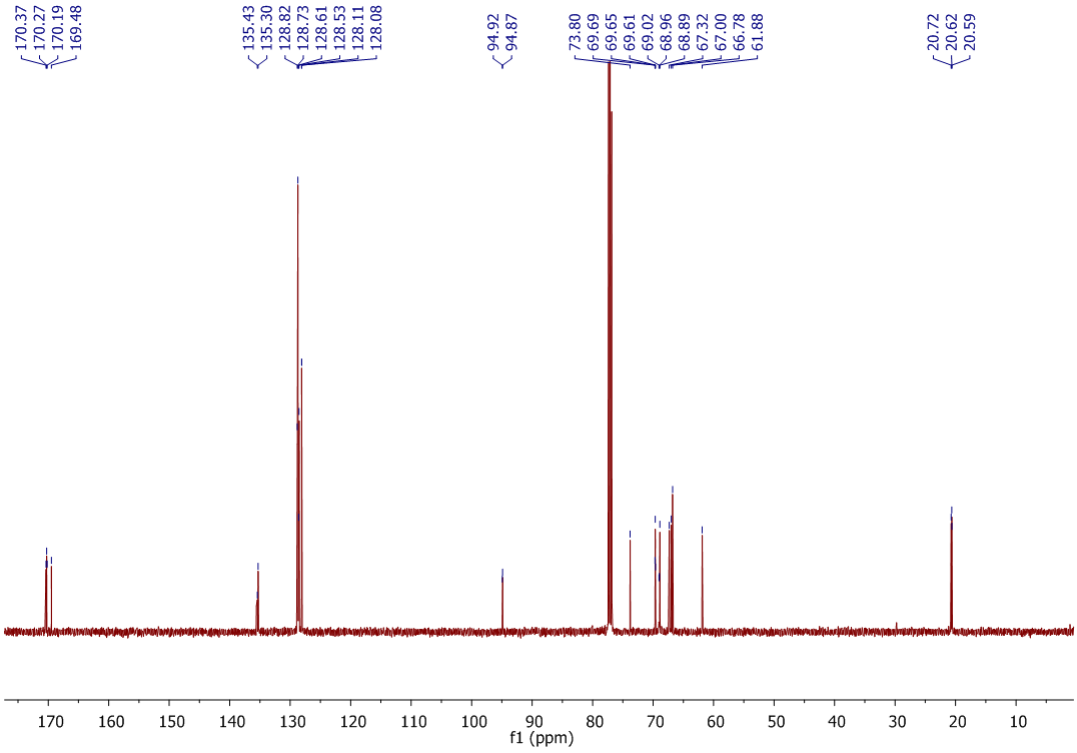
^{13}C NMR (125 MHz, $CDCl_3$) Phenyl 2,4,6-Tri-*O*-acetyl-3-*O*-(benzyl)carboxymethyl 1-thio-β-D-galactopyranoside 59



^{1}H NMR (500 MHz, $CDCl_3$) 2,4,6-Tri-*O*-acetyl-3-*O*-(benzyl)carboxymethyl 1-*O*-diphenoxyphosphoryl-α-D-galactopyranoside 61



^{13}C NMR (125 MHz, $CDCl_3$) 2,4,6-Tri-*O*-acetyl-3-*O*-(benzyl)carboxymethyl 1-*O*-diphenoxyphosphoryl-α-D-galactopyranoside 61



^{31}P NMR (121 MHz, $CDCl_3$) 2,4,6-Tri-*O*-acetyl-3-*O*-(benzyl)carboxymethyl 1-*O*-diphenoxyphosphoryl-α-D-galactopyranoside 61