Native Chemical Thioesterification: Synthesis of Peptide and Protein Thioesters through an $N \rightarrow S$ Acyl Shift

Jaskiranjit Kang

A thesis submitted in partial fulfilment of the requirements for the degree award of:

Doctor of Philosophy

University College London

Department of Chemistry

2010

Declaration

I, Jaskiranjit Kang, 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.

Abstract

The total chemical synthesis of a protein provides atomic-level control over its covalent structure, however polypeptides prepared by solid phase peptide synthesis are limited to approximately fifty amino acid residues. This limitation has been overcome by 'Native Chemical Ligation', which involves amide bond formation between two unprotected polypeptides: a peptide with a C-terminal thioester and an N-terminal cysteinyl peptide. Synthesis of the required peptide thioester is difficult, particularly by Fmoc-chemistry. During our studies towards the semisynthesis of erythropoietin, we discovered reaction conditions that reversed Native Chemical Ligation and generated peptide and protein thioesters through an *N*→*S* acyl transfer.

A peptide with both a Gly-Cys and an Ala-Cys-Pro-glycolate ester sequence was selectively thioesterified between the Gly-Cys sequence upon microwave-heating at 80 °C with 30 % v/v 3-mercaptopropionic acid (MPA), to afford the peptide-Gly-MPA thioester (84 % yield). Recombinant erythropoietin containing 4(Xaa-Cys) sites was also selectively thioesterified between His-Cys and Gly-Cys sequences upon treatment with 20 % MPA at 60 °C to give the peptide-MPA thioesters. His-/Gly-/Cys-Cys sequences were found to be particularly prone to thioesterification but thioester hydrolysis and disulfide bond formation also occurred with MPA, and therefore optimisation studies were conducted using ^{13}C NMR spectroscopy. Sodium 2-mercaptoethanesulfonate (MESNa) was found to be a better thiol because it resulted in 100 % conversion into the peptide thioester and no thioester hydrolysis. The first biological application of this new method was then completed with the total synthesis of 45-amino acid Ser22His human *β*-defensin 3 (HBD3) for which a more complex 22-amino acid peptide thioester (22 % yield) was synthesised. Folded Ser22His HBD3 was assessed to have bioactivity equivalent to that reported: MBCs against *P. aeruginosa* = 0.9 µM, *S. aureus* = 1.8 µM; and optimal chemoattraction concentration = 0.2 nM. Posttranslationally modified native *N*-linked glycopeptide thioesters were also synthesised using the new thioesterification method, containing protected (20 % yield) or deprotected monosaccharides (64 % yield).

Contents

page

Acknowledgments……..…………...………………………………………………….….…8

List of Publications as a Result of the Work Presented in this Thesis…………….....…...…...9

Abbreviations……………………………………………………...……..…..............10

Chapter

1 Introduction…….………………………………………………...……………13

1.1 Total Chemical Protein Synthesis……………………...……..……..……...……13

1.2 Solution Phase Peptide Synthesis…………………………...…….…………..…15

1.3 Solid Phase Peptide Synthesis. ………………………………….……….…..……23

1.4 Convergent Synthesis………………………………………………….…………..26

1.5 Dibenzofuran-Assisted Chemoselective Ligation…………………….…………...27

1.6 Non-Native Bond Formation in Chemoselective Ligation……………...……………28

1.7 Native Chemical Ligation. ……………………………………...……...…...…29

1.8 Inteins……………………………………………………………………………...32

1.9 Kinetically Controlled Ligation…………………………………….…...………38

1.10 Sugar-Assisted Ligation……………………………………………….....……40

1.11 Phenolic Esters in the Synthesis of Posttranslationally Modified Peptides…………41

1.12 Other Strategies: Oxime-Forming Ligation and Staudinger Ligation………………..44

1.13 Peptide Thioesters………………………………………………………………45

1.14 *N→S* Acyl Transfer Using the Sulfonamide-Linker. ………………………………..47

1.15 *N→S* Shift in Acyl-Transfer Auxiliaries……………………………………………..48

1.16 3-Mercaptopropionic Acid (MPA)-Mediated Thioesterification……………………….52

1.17 *N→S* Alkyl Cysteine-Assisted Thioesterification……………………………………54

1.18 Acyl Oxazolidinone and Aniline Linker-Mediated Thioesterification………………55

1.19 Cysteinylprolyl Esters…………………………………………………………..………57

1.20 Conclusion…………………………………………………………………...……….60

1.21 Scope of Thesis……………………………………………………………………….61

2 Discovery of a Novel Peptide and Protein Thioesterification Reaction: Native Chemical Thioesterification..…………………………………………….…..…..…63

2.1 Introduction……………………………………………………………………..63

2.2 Results and Discussion……………………………………………….....…………66

2.3 Synthesis of a Model Peptide-CPE with Two Unprotected Cysteine Residues……………………………………………………………………………66

2.4 Attempted CPE Ligation…………………………………………….............……73

2.5 MPA-Mediated Peptide Thioesterification……………………...……………………77

2.6 Glycine-MPA Thioester Ligation……………………………………...……..…81

2.7 Synthesis of the Model Peptide-CPE with an Internally Capped Cysteine………..…82

2.8 CPE-Mediated Ligation with an Internally Capped Cysteine………………………..83

2.9 MPA-Mediated Thioesterification of a Peptide-CPE with an Internally Capped Cysteine……………………………………………………………...…………..86

2.10 Native Chemical Ligation of the Alanine-MPA Thioester…………………...……89

2.11 Protein Thioesterification………………………………...……………………..…89

2.12 Favoured Xaa-Cys Sequences Undergoing MPA-Mediated Thioesterification…………………………………………………………………..91

2.13 Racemisation…………………………………………………….……………..…94

2.14 MPA Thioester Characterisation by ^{1}H NMR Spectroscopy…………..……………94

2.15 Discussion………………………………………………………….…………………97

2.16 Conclusion……………………………………………………………………………..100

2.17 Acknowledgments…………………………………………………………………..101

3 Reaction Optimisation and First Biological Application of Native Chemical Thioesterification….……………………………………...………………...…...102

3.1 Introduction……………………………………………………………...…….102

3.2 Results and Discussion ………………………………………………………..……104

3.2.1 Optimisation of One-Pot Peptide Thioesterification………………………………..104

3.2.2 Optimisation Studies Using Conventional Heating………………………………...111

3.2.3 Other Observations from Conventional Heating Experiments……………………...112

3.2.15 Optimisation Studies Using Microwave Heating…………………………………...116

3.2.4 Summary of Optimisation Studies…………………………………………………118

3.3 The First Biological Application of Native Chemical Thioesterification: Human *β*-Defensin 3 (HBD3) ……………………………………………………………...119

3.3.1 Introduction……………………………………………………………………119

3.3.2 Results and Discussion………………………………………………………...120

3.3.3 Total Chemical Synthesis of the HBD3 Analogue…………………………………120

3.3.4 Synthesis of the Ser22His HBD3 N-Terminal Peptide……………………………..122

3.3.5 Synthesis of the Ser22His HBD3 C-Terminal Cysteinyl Peptide…………………..123

3.3.6 Synthesis of the N-Terminal HBD3 Thioester……………………………………...124

3.3.7 Synthesis of Full-Length Protected Ser22His HBD3(1-45) by NCL………………127

3.3.8 Synthesis of Full-Length Deprotected Ser22His HBD3(1-45) ……………………128

3.3.9 Synthesis of the Folded Ser22His HBD3(1-45) Analogue…………………………131

3.3.10 Biological Activity: Antimicrobial and Chemoattraction Assays…………………..134

3.4 Discussion……………………………………………………………………………..140

3.5 Conclusion……………………………………………………………………………..144

3.6 Acknowledgments……………………………………………………………………..145

4 The First Application of Native Chemical Thioesterification to the Synthesis of Native *N*-Linked Glycopeptide Thioesters………………………………………146

4.1 Introduction……………………………………………………………………………146

4.2 Results and Discussion………………………………………………………………...148

4.2.1 Synthesis of the *N*-Linked Glycoamino Acid………………………………………148

4.2.2 Synthesis of the Native *N*-Linked Glycopeptide……………………………………151

4.2.3 Native *N*-Linked Glycopeptide Thioester Formation………………………………153

4.2.4 Native Chemical Ligation of the *N*-Glycopeptide Thioester………………………..155

4.2.6 Native Chemical Thioesterification of a Fully Deprotected Native *N*-Linked Glycopeptide……………………………………………………………………………..156

4.3 Conclusion……………………………………………………………………...…158

5 Overall Conclusions and Future Work…………………………….……….…159

6 Experimental Section……………………………………………………...…161

7 References………..………………………………………………………...194

Acknowledgments

I would like to thank my supervisor Dr Derek Macmillan for giving me the opportunity to work on this exciting project. I would also like to extend my gratitude to the Macmillan group; Dr Jonathan P. Richardson for protein expression; the following members of "Team Defensin": Dr Julia R. Dorin, Natalie L. Reynolds, and Dr Christine Tyrrell for their help with biological assays; and Dr Abil E. Aliev and Dr Lisa Harris for spectroscopic and spectrometric instruction. Financial support from the E.P.S.R.C. and the UCL Graduate School for the Conference Fund Award is gratefully acknowledged. I would also like to thank the UCL Department of Chemistry for the 2010 Apley Prize for my thesis.

Publications as a Result of the Work Presented in this Thesis

1. "3-Mercaptopropionic acid-mediated synthesis of peptide and protein thioesters". Jaskiranjit Kang, Jonathan P. Richardson and Derek Macmillan, *Chemical Communications*, **2009**, (4), 407-409.

2. "Peptide thioester synthesis through *N→S* acyl-transfer: application to the synthesis of a *β*-defensin". Jaskiranjit Kang, Natalie L. Reynolds, Christine Tyrrell, Julia R. Dorin and Derek Macmillan, *Organic and Biomolecular Chemistry*, **2009**, *7*(23), 4918-4923.

3. "Peptide and protein thioester synthesis *via N→S* acyl transfer". Jaskiranjit Kang and Derek Macmillan, *Organic and Biomolecular Chemistry*, **2010**, *8*(9), 1993-2002.

Abbreviations

Abs	absorbance
Ac	acetyl
Acm	acetamidomethyl
Boc	*tert*-butoxycarbonyl
BME	*β*-mercaptoethanol
Bn	benzyl
CPE	cysteinylprolyl ester
DBU	1, 8-diazobicylco(5.4.0)undec-7-ene
DCC	*N, N′*-dicyclohexylcarbodiimide
DCM	dichloromethane
DIC	*N, N'*-1, 3-diisopropylcarbodiimide
DIPEA	*N, N*-diisopropylethylamine
DMF	*N, N*-dimethylformamide
DTT	1,4-dithio-DL-threitol
EDCI	*N*-(3-dimethylaminopropyl)-*N'*-ethylcarbodiimide
EDT	ethanedithiol
EEDQ	*N*-ethoxycarbonyl-2-ethoxy-1, 2-dihydroquinoline
EPL	expressed protein ligation
EPO	erythropoietin
Fmoc	9-fluorenylmethoxycarbonyl

Gc	glycolic acid
Glc*N*Ac	*N*-acetyl-D-glucosamine
Gu.HCl	guanidine hydrochloride
HBr	hydrogen bromide
HBTU	2-(1*H*-Benzotriazol-1-yl)-1,1,3,3-tetramethyluronium-hexafluorophosphate
HCl	hydrochloric acid
HF	hydrogen fluoride
HOBt	1-hydroxybenzotriazole
HOSu	*N*-hydroxysuccinimide
LC-MS	liquid chromatography-mass spectrometry
MAA	2-mercaptoacetic acid
Me	methyl
MeCN	acetonitrile
MESNa	sodium 2-mercaptoethanesulfonate
MHz	mega Hertz
MPA	3-mercaptopropionic acid
MPAA	4-mercaptophenylacetic acid
MS	mass spectrometry
NCL	Native Chemical Ligation
NCT	Native Chemical Thioesterification
NMP	*N*-methylpyrrolidone

NMR	nuclear magnetic resonance
PEG	polyethylene glycol
PyBOP	benzotriazol-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate salt
PyBrOP	bromo-tris-pyrrolidino-phosphonium hexafluorophosphate salt
(RP)HPLC	reversed-phase high performance liquid chromatography
SPPS	solid phase peptide synthesis
t-Bu	*tert*-butyl
TCEP	*tris*(2-carboxyethyl)phosphine
TFA	trifluoroacetic acid
THF	tetrahydrofuran
Thz	thiazolidine
TLC	thin layer chromatography
Trt	triphenylmethyl
UV	ultraviolet
Z (Cbz)	benzyloxycarbonyl

1 Introduction

1.1 Total Chemical Protein Synthesis

Proteins comprise one of the four major classes of biologically important macromolecules as well as nucleic acids, carbohydrates, and lipids, and have many diverse biochemical and structural roles. Each eukaryotic cell has been estimated *via* genomic sequencing to contain more than 20 000 genes[1,2] and whilst each gene encodes the amino acid sequence of a protein, protein diversity is further enhanced by mRNA splicing before translation and posttranslational modification of specific amino acid residues after the polypeptide chain has been assembled. This vastly increases the total number and diversity of proteins.[3]

In 1978, Werner Arber, Daniel Nathans, and Hamilton Smith received the Nobel Prize in Physiology or Medicine for the discovery of restriction enzymes and their application to the problems of molecular genetics.[4-16]

Recombinant DNA-based technology was illustrated by Cohen *et al.* in 1973 with the construction of new plasmid DNA species through the *in vitro* joining of restriction endonuclease-generated fragments of separate plasmids.[17] Specific sequences from prokaryotic or eukaryotic chromosomes or extrachromosomal DNA could be inserted by transformation into *Escherichia coli*, and the resulting recombinant plasmids were found to be biologically functional replicons with genetic properties and nucleotide base sequences present from both parent DNA molecules.[17] Functional plasmids could be obtained by the reassociation of endonuclease-generated fragments of larger replicons, in addition to joining plasmid DNA molecules from different origins.[17]

Since the early 1980s, protein engineering using recombinant DNA-based technology has become a useful tool to elucidate the molecular basis of protein function.[18-20] This has been achieved by sequentially changing the amino acid sequence of a polypeptide by expressing it in *Escherichia coli*, followed by folding the polypeptide and measuring the effect of the mutation on the properties of the mutant protein molecule.[21]

Molecular biology techniques that enable site-directed mutagenesis, for example, are extremely useful but they also have certain limitations: only the twenty genetically-encoded amino acids can be incorporated within the protein molecule.[21,22] The lack of atomic-level

control over the covalent structure of a protein molecule also makes it difficult to incorporate site-specific post-translational modifications or precise labels within a protein molecule.[21,22] Although Schultz and co-workers have described the use of engineered microorganisms and *in vitro* techniques to overcome these limitations by incorporating unnatural amino acid residues, these techniques are not yet commonly used to study the molecular basis of protein function.[23,24]

These limitations of molecular biology can be overcome through the total chemical synthesis of proteins. The total chemical synthesis of proteins provides atomic-level control over the chemical structure of a protein molecule, which in turn permits the installation of site-specific modifications[25] including: isotopic labels;[26-32] posttranslational modifications such as glycosylation[33-35] and phosphorylation;[36] non-genetically encoded amino acids;[37] and fluorophores.[38,39] The incorporation of such moieties enables us to further define the precise molecular basis of protein function through the use of various biophysical techniques,[40,41] including: nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography (Figure 1.1).

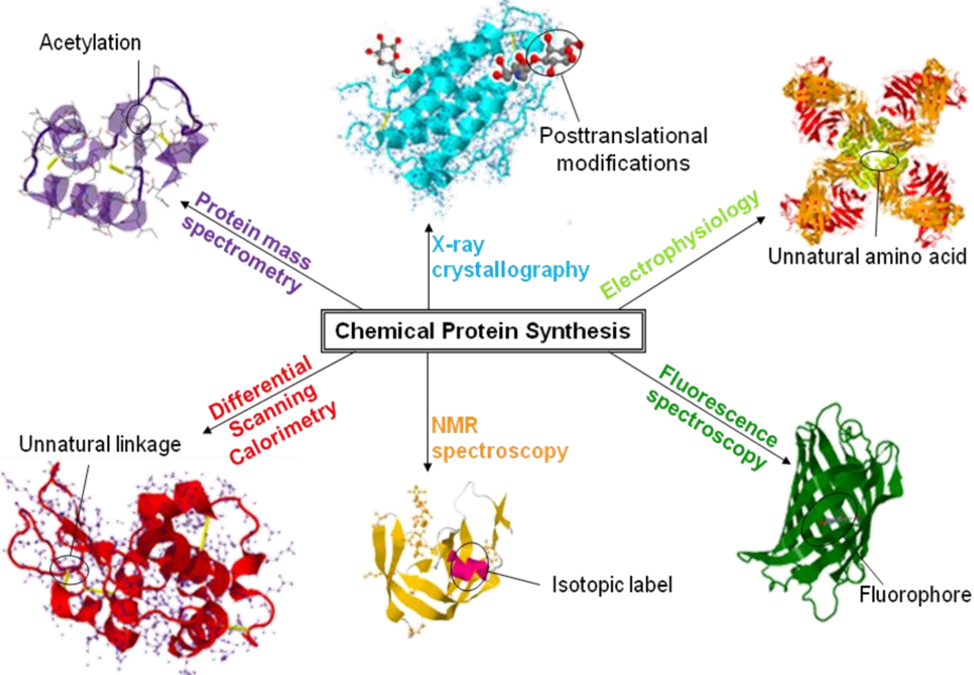


Figure 1.1 The total chemical synthesis of proteins enables the incorporation of various site-specific modifications, which aid the elucidation of the molecular basis of protein function through the use of biophysical techniques inlcuding X-ray crystallography.[25-41]

1.2 Solution Phase Peptide Synthesis

Early in the twentieth century, Emil Fischer began with the pursuit of the total chemical synthesis of enzymes and though he never succeeded in this goal, he did pioneer the field of chemical peptide synthesis.[42] It was during that time that Fischer and Hofmeister also developed their theory of proteins being linear polymers comprised of α-amino acid subunits, which were termed polypeptides.[43,44] Fischer developed methods for the chemical synthesis of peptides that contained even chiral amino acids, and this resulted in the synthesis of an 18 amino acid residue peptide which contained fifteen glycine and three leucine residues.[45,46]

The total chemical synthesis of proteins became a major objective worldwide and culminated in the development of many synthetic peptide chemistry methods. The first known biologically active peptide hormone to be synthesised was the 9 amino acid residue oxytocin by du Vigneaud and co-workers in 1953.[47] Kenner *et al.* described a notable advance towards the total chemical synthesis of proteins by using twelve protected peptides to synthesise a 129 amino acid residue lysozyme enzyme molecule.[48] The unambiguous total synthesis of the crystalline and fully biologically active 51 amino acid residue human insulin and its analogues,[49] and the 124 amino acid residue ribonuclease A were also demonstrated.[50,51]

'Classical' solution phase peptide synthesis usually involves the C→N terminal directional synthesis of peptides in organic solvents such as chloroform, dichloromethane, ethyl acetate, *N*,*N*-dimethylformamide, and methanol, for example.[52] The amino acid building blocks used are otherwise fully-protected except at the reacting terminus (Scheme 1.1).[52,53] An α-amino-protected amino acid with an activated α-carboxyl group reacts with a free amino group of an α-carboxyl-protected amino acid to afford a dipeptide. The α-amino protecting group is then selectively removed from the dipeptide and the next α-amino-protected amino acid with an activated α-carboxyl group is then coupled to form a tripeptide. This process is repeated until the desired fully-protected peptide chain has been assembled. In a final step all the protecting groups are removed to afford the desired peptide.[52,53]

Scheme 1.1 Classical solution phase peptide synthesis employing maximal protecting groups. All non-reacting functional groups present within the amino acid building blocks are reversibly protected, and are removed in the final step to afford the full-length peptide. 'PG' denotes protecting group and 'X' is a leaving group.[52,53] Purification *via* flash column chromatography and multiple re-crystallisations at each step prolong this synthetic route.

Protection is essential for the synthesis of peptides as the constituent amino acids are multifunctional. In 1932, Bergmann and Zervas introduced the first reversible protection of an α-amino group of an amino acid with the urethane derived protecting group: benzyloxycarbonyl 'Z'.[54] This group could be easily introduced and selectively removed under certain conditions (HBr/TFA or H_2/Pd-C), whilst being stable to other conditions (TFA or piperidine/DMF). This also led to the development of orthogonal protecting

groups by Barany and colleagues.[55,56] Orthogonality in this context translates as: two or more protecting groups that belong to independent classes of molecules which can be removed by distinct mechanisms. Thus, the protecting groups can be removed in any order and in the presence of other functional groups. In general, orthogonal protection strategies are milder because their selective deprotection is governed by different removal mechanisms, rather than different reaction rates.[57]

Some common protecting groups used for the protection of the α-amino group in solution syntheses, and their removal conditions and compatibility with other protecting groups are shown in Table 1.1.[57] Commonly used α-amino groups include: Z, Nps (2-nitrophenylsulfenyl), and Bpoc [2-(4-biphenyl)isopropoxycarbonyl] in combination with the *t*Bu (*tert*-butyl) side-chain protecting group, or the Boc (*tert*-butyloxycarbonyl) group in combination with the Bn (benzyl) side-chain protecting group.[57]

Equally, protection of the α-carboxyl group can be achieved by:[94]

- Reaction of the free acid of an α-amino acid with an alcohol under acidic conditions (usually *p*-TsOH or HCl).[95]

- Protecting the free acid with a *tert*-butyl group by reaction with isobutene under acidic conditions (usually *p*-TsOH or H_2SO_4).[96,97]

- Reaction of an α-amino protected amino acid in the presence of a base or as a caesium salt, with the corresponding halide (typically bromide).[98,99]

- Reaction of an α-amino protected amino acid with a condensing agent such as *N, N'*-dicylcohexylcarbodiimide (DCC) in the presence of DMAP and the alcohol derivative of the protecting group.[100]

Table 1.1 Common α-amino protecting groups employed during peptide synthesis are presented along with their removal conditions and potential orthogonal protecting groups.[57]

Name and structure	Removal conditions	Stable to removal of:	Ref.
Triphenylmethyl (Trityl, Trt)	1) 1 % TFA-DCM 2) 0.1 M HOBt-TFE 3) 0.2 % TFA, 1 % H_2O-DCM 4) 3 % TCA-DCM	1) Fmoc 2) Alloc	58, 59, 60, 61, 62, 63
2-(4-Biphenyl)isopropoxycarbonyl (Bpoc)	1) 0.2-0.5 % TFA	1) Fmoc 2) Alloc	64, 65, 66, 67, 68, 69
NO_2 **2-Nitrophenylsulfenyl (Nps)**	1) Diluted solutions of HCl-$CHCl_3$-AcOH 2) 2-Mercaptopyridine-AcOH-MeOH, DMF or DCM 3) Ni Raney column in DMF	1) Fmoc	70, 71, 72, 73
Benzyloxycarbonyl (Z)	1) cat. H_2 2) Strong acids i.e. -HBr in AcOH, -TFA at high temperatures, -TFA-thioanisole or liquid HF 3) BBr_3	1) Boc 2) Fmoc 3) Trt 4) Alloc 5) *p*NZ (except cat. hydrogenation)	54, 74, 75, 76, 77, 78
Allyloxycarbonyl (Alloc)	1) cat. $Pd(PPh_3)_4$, scavengers: $H_3N.BH_3$, $Me_2NH.BH_3$ or $PhSiH_3$ in organic solvents	1) Boc 2) Fmoc 3) Trt 4) *p*NZ (except cat. hydrogenation)	79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89
O_2N ***p*-Nitrobenzyloxycarbonyl (*p*NZ)**	1) 1-6 M $SnCl_2$, 1.6 mM $HCl_{(dioxane)}$ in DMF 2) cat. H_2	1) Boc 2) Fmoc 3) Trt 4) Alloc	89, 90, 91, 92, 93

Some α-carboxyl protecting groups are shown in Table 1.2 along with their removal conditions and compatibility with other orthogonal protecting groups.[57]

Table 1.2 α-Carboxyl protecting groups employed during peptide synthesis are presented along with their removal conditions and potential orthogonal protecting groups.[57]

Name and structure	Removal conditions	Stable to removal of:	Ref.
***tert*-Butyl (^{t}Bu)**	1) 90 % TFA-DCM 2) 4 M HCl in dioxane	1) Fmoc 2) Z (by cat. hydrogenation) 3) Trt 4) Alloc 5) *p*NZ	97, 101
2-Chlorotrityl (2-Cl-Trt)	1) 1 % TFA-DCM	1) Fmoc 2) Alloc	102
MeO, OMe **2, 4-Dimethoxybenzyl (Dmb)**	1) 1 % TFA-DCM	1) Fmoc 2) Alloc	103
Allyl (Al)	1) $Pd(PPh_3)_4$ (0.1 equiv.) and scavengers: $PhSiH_3$, 10 equiv., DCM	1) Boc 2) Fmoc 3) *p*NZ (except cat. hydrogenation) 4) Trt	104, 105
Benzyl (Bn)	1) HF 2) TFMSA 3) H_2 cat. 4) NaOH in aqueous organic solvents	1) Boc (except repetitive removal) 2) Fmoc 3) *p*NZ (except cat. hydrogenation) 4) Trt 5) Alloc	101

The formation of a peptide bond between two amino acids requires activation of the α-carboxylic acid of an amino acid to facilitate nucleophilic attack by the α-amino group of another amino acid. This is because at ambient temperatures, carboxylic acids simply form salts with amines.[106] These salts would require heating at high temperatures to generate an amide bond, and this is unfavourable due to the presence of other sensitive functional groups. Under heated acidic reaction conditions an equilibrium exists between the condensing amino acid starting materials and the amide bond-containing product. This

makes the peptide bond forming reaction difficult to drive to completion and increases the likelihood of peptide bond hydrolysis (Scheme 1.2i).[107,108]

Therefore, the formation of a peptide bond is favoured when the α-amino group of an amino acid is made more nucleophilic and the α-carboxyl group of another amino acid made more electrophilic.[106,109] Consequently, activation of the α-carboxyl group by transforming it into an acyl chloride is also unfavourable because it is unstable and hydrolyses easily, and this also makes the amino acid prone to racemisation (Scheme 1.2ii).[106,109] A molecule of hydrochloric acid is also formed during each peptide coupling and this is also unfavourable for acid-sensitive peptides.

Better amino acid coupling methods include:[106,109]

- Transformation of the α-carboxyl group into an acyl anhydride (either a mixed or a symmetrical anhydride), but along with the amide bonded product an equivalent of the free acid is also generated per coupling and this can be expensive (Scheme 1.2iii).

- Activation of the α-carboxyl group by transforming it into a pentafluorophenyl (Pfp) ester, which forms a stable anion leaving group during coupling with the α-amino group (Scheme 1.2iv).

- Generation of an active ester *in situ* by reaction of the α-carboxyl group with, for example, a coupling reagent mixture such as: 1-hydroxybenzotriazole (HOBt)/ 2-(1*H*-Benzotriazol-1-yl)-1,1,3,3-tetramethyluronium-hexafluorophosphate (HBTU) to generate an HOBt-active ester *in situ*, which is a good leaving group and is displaced upon nucleophilic attack by the α-amino group (Scheme 1.2v).

- Activation of the α-carboxyl group *in situ* with a carbodiimide derivative such as dicyclohexylcarbodiimide (DCC), which reacts with the α-amino group to generate the amide-containing product and urea by-product (Scheme 1.2vi).

i)

ii)

iii)

iv)

v)

vi)

Scheme 1.2 Methods of peptide bond formation tend to rely on α-carboxyl group activation of an amino acid followed by coupling with the free α-amino group of another amino acid. Generally, i and ii are disfavoured because they either use or generate acid, whereas iii-vi are more popular coupling methods.[106,109]

Classical solution phase peptide synthesis using organic solvents and maximal, reversible side-chain protecting groups to prevent undesired reactions has many associated shortcomings.[52,53] These include: racemisation *via* oxazolone formation of the α-carboxyl activated C-terminal amino acid of a protected peptide during its coupling with another protected peptide under basic conditions (Scheme 1.3).[106]

Scheme 1.3 Racemisation of the C-terminal amino acid during peptide bond formation *via* oxazolone formation under basic conditions.[106,109] Racemisation by this route is promoted by α-carboxyl activation, where 'X' denotes a leaving group, which facilitates cyclisation. Deprotonation of the α-carbon followed by re-protonation from either face results in a racemic mixture and hence loss of chiral integrity.

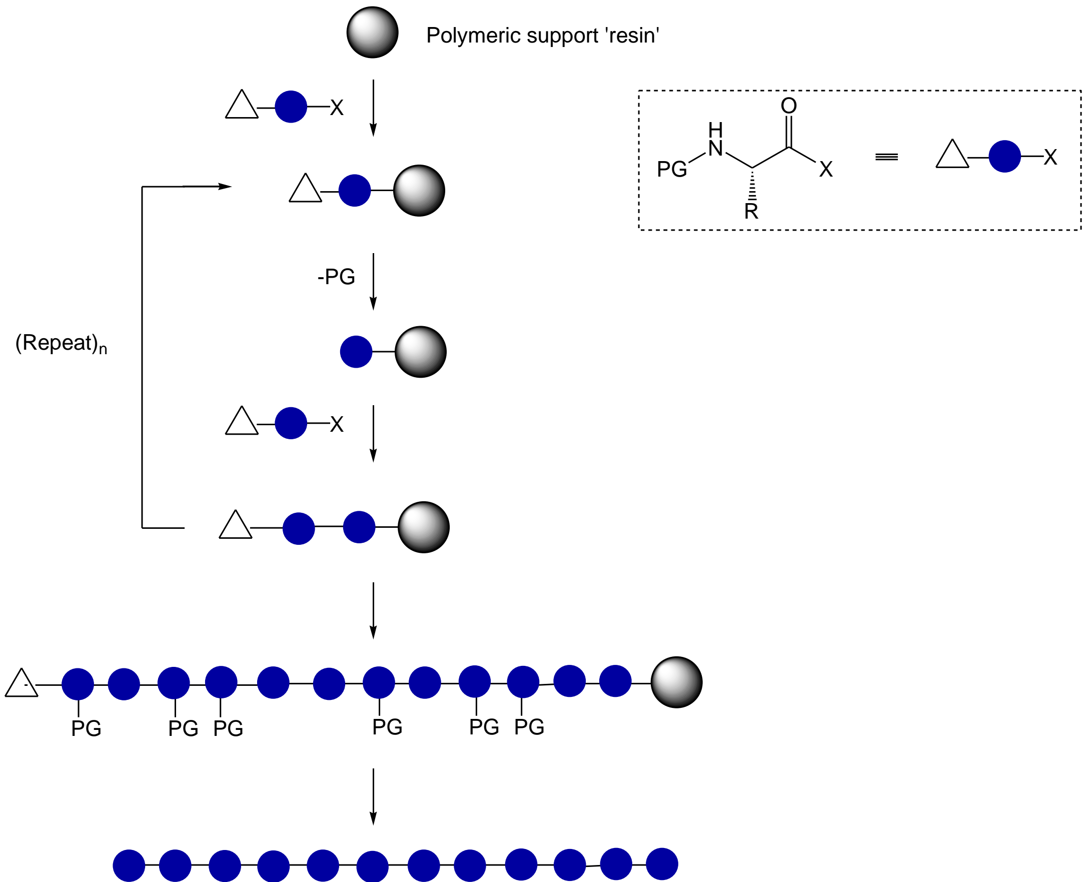
The necessary presence of multiple, usually hydrophobic, protecting groups also makes peptide purification by silica gel chromatography more difficult. Characterisation of the fully-protected peptide can be difficult, and multiple recrystallisations to provide a homogeneous product often prove insufficient. Therefore characterisation can usually be achieved once the peptides are deprotected.[52,53]

Fully-protected peptides often have a solubility problem,[48,51] and as well as being poorly soluble in aqueous media they can also be sparingly soluble in organic media. This has led to low peptide concentrations being used in reactions, which has been observed to reduce the reaction rate. This also results in incomplete reactions and accumulation of side-reaction products.[53]

Generally, these disadvantages remain associated with many solution phase peptide syntheses though this approach has been successfully used in the synthesis of several proteins including the 123-mer human angiogenin,[110] which is involved in angiogenesis and has ribonuclease enzymatic activity, and the largest protein synthesised to date the 238 amino acid residue *Aequoria victoria* green fluorescent protein (GFP).[111,112] Generally, solution phase synthesis requires highly-skilled chemists to carry out the laborious, and often time-consuming, syntheses.

1.3 Solid Phase Peptide Synthesis

In 1963, Merrifield introduced the total chemical synthesis of peptides through the revolutionary concept of 'solid phase peptide synthesis' (SPPS) (Scheme 1.4).[113] It has since become a reliable method for peptide synthesis, and has been estimated to be approximately fifty times less laborious than the solution phase synthesis of the same peptide.[53] SPPS involves the C→N terminal directional synthesis of a peptide, in order to avoid oxazolone formation and hence racemisation (Scheme 1.3),[106] whereas protein biosynthesis proceeds in the opposite N→C direction without racemisation.



Scheme 1.4 Merrifield's solid phase peptide synthesis.[113] A fully protected peptide chain is assembled stepwise on an insoluble polymer support in the C→N terminal direction. Protecting groups and the covalent linkage to the polymer support are cleaved in the final step to afford the full-length peptide. 'PG' denotes protecting group and 'X' is a leaving group. All reaction by-products can simply be filtered off after each coupling reaction.

During SPPS, an N-α protected amino acid is initially coupled through its C-terminus to a "solid support", which consists of a cross-linked polymer that is macroscopically insoluble in the reaction solvents but microscopically soluble to allow coupling (Scheme 1.4).[113] The N-α protecting group of the first amino acid is then removed and the excess amino acid and soluble by-products are easily removed through filtration and washing. The consecutive N-α protected carboxyl-activated amino acid is then reacted with the resin-bound amino acid to form a new amide bond. The resin-bound dipeptide is then filtered and washed, and these steps are then repeated until the desired resin-bound protected peptide sequence has been attained. At the end of a synthesis, the covalent linkage to the resin is cleaved as well as undesired N-α and orthogonal side-chain protecting groups to afford the crude product.[114]

The most common α-amino protecting groups employed during SPPS are: 9-fluorenylmethoxycarbonyl (Fmoc) and Boc, which are used in either Fmoc/*tert*-butyl (^{t}Bu) or Boc/benzyl (Bn) SPPS, respectively (Table 1.3).[115-117]
The C-terminus of an amino acid is usually linked to the solid support during SPPS, and therefore a linker or a handle usually acts as the α-carboxyl protecting group. However, in cases where the side-chain of the C-terminal amino acid is anchored to the resin, an α-carboxyl protecting group[57] is necessary and some examples of these have been shown in Table 1.2.

Coupling between consecutive amino acids during SPPS is usually achieved by activation of the α-carboxyl of the incoming amino acid into an active ester *in situ*. This is usually through generation of a benzotriazole ester with either: HOBt/HBTU and DIPEA, HOBt/PyBOP (benzotriazol-1-yl-oxy-tris-pyrrolidinophosphonium hexafluorophosphate salt) and DIPEA, or HOBt/DIC (*N, N'*-diisopropylcarbodiimide) mixtures.[131,132]

Merrifield championed the use of Boc for the protection of the α-amino group when using polystyrene resins,[113,114] whereas Sheppard favoured polyamide resins in combination with Fmoc as the α-amino protecting group.[120,121] This was due to the differences in the swelling properties of the resins in different reaction solvents, which was thought to affect the coupling of the molecule to the resin.[106]

Table 1.3 Common α-amino protecting groups employed during solution and solid phase peptide synthesis are presented with their removal conditions and potential orthogonal protecting groups.[57]

Name and structure	Removal conditions	Stable to removal of:	Ref.
9-Fluorenylmethoxycarbonyl (Fmoc)	**Solution phase:** 1) NH_3 (10 h) 2) Morpholine or piperidine in organic solvents (min) 3) Polymeric secondary amines (i.e. piperidine, piperazines) in organic solvents **Solid phase:** 1) 20 % piperidine-DMF 2) 1-5 % DBU-DMF 3) Morpholine-DMF (1:1) 4) 2 % HOBt, 2 % hexamethyleneimine, 25 % *N*-methylpyrrolidine in DMSO-NMP (1:1)	1) Boc 2) Z (except cat. hydrogenation) 3) Trt 4) Alloc 5) *p*NZ (except cat. hydrogenation)	118, 119, 120, 121, 122, 123, 124, 125
***tert*-Butyloxycarbonyl (Boc)**	1) 25-50 % TFA-DCM 2) 4 M HCl in dioxane 3) 2 M $MeSO_3H$ in dioxane 4) 1 M TMS-Cl, 1 M phenol-DCM	1) Fmoc 2) Z (by cat. hydrogenation) 3) Trt 4) Alloc 5) *p*NZ	113, 126, 127, 128, 129, 130

If the peptide synthesis has been designed and conducted efficiently then the majority of the crude product should be the desired product. The crude product can then be precisely characterised using high-resolution physical techniques such as: LC-MS (liquid chromatography-mass spectrometry) and NMR spectroscopy. However, if the crude peptide is analysed to contain by-products then it can be further purified by reversed-phase high performance liquid chromatography (RP)HPLC, which often results in a loss of yield, to better attain a homogeneous peptide product.[133]

Key advantages of SPPS are that linkage of the growing peptide chain to an insoluble resin facilitates simple purification through filtration and washing. Rapid and complete reactions are promoted by the addition of excess (typically five to ten equivalents) activated amino acids that can be easily removed after each amino acid coupling step, but this is expensive. Handling losses that are usually associated with solution phase chemistry are minimised and the inherent solubility problem in solution phase synthesis, where fully-protected peptide chains are assembled, is also chiefly avoided by SPPS. This is because a heterogeneous reaction mixture is required to perform the stepwise synthesis

of a protected peptide chain. The chiral integrity of peptide products is also usually retained.[113,114]

Once a polypeptide has been assembled, its purification can be precluded by the presence of by-products due to intermediates and incomplete reactions. In practical terms, this often translates into similar retention times for the by-products and the product polypeptide during (RP)HPLC purification. Preparative (RP)HPLC is a facile high resolution technique commonly used to purify unprotected peptides, provided they are transferable in HPLC solvent mixtures.[133] These are typically water: acetonitrile: 0.1 % trifluoroacetic acid. Deprotected peptides can also be readily characterised by electrospray ionisation mass spectrometry (ESI-MS).

A severe disadvantage of SPPS is that only small protein molecules that are comprised of a maximum of approximately fifty amino acids in length can be efficiently synthesised. This can be attributed to many factors including the swelling properties of the resin in the reaction solvent which can affect amino acid couplings, and also the coupling time allowed. Even if the synthetic route to the target polypeptide was highly optimised, the accretion of by-products attached to the resin due to incomplete reactions, and the presence of any remnant impurities within the reactants, reagents, or solvents used, renders SPPS impractical for the preparation of larger protein molecules in good yield and of satisfactory homogeneity, with unambiguously characterised covalent structures.[133] However, this limitation has not prevented the utilisation of SPPS in the synthesis of thousands of biologically active proteins.

Merrifield also developed an instrument for the automated solid phase synthesis of peptides[134] and utilised this new approach in the first synthesis of the enzyme ribonuclease A (RNase A), albeit in a low overall yield.[135,136]

1.4 Convergent Synthesis

The restrictions imposed by stepwise peptide syntheses led to the development of the convergent synthesis approach. This involves a coupling reaction between shorter synthetic peptides that join to produce a full-length polypeptide. This strategy is favoured as it overcomes the poorer quality of larger polypeptides attainable by stepwise peptide syntheses.[52,53] The precursor peptides are small and can therefore be highly pure and

precisely characterised molecules.[133]

A fully convergent strategy is generally considered better than stepwise peptide synthesis because an equal number of steps separate all the starting materials from the product. The product usually has different chemical properties to the starting material which makes isolation during (RP)HPLC easier. Another useful outcome of using convergent peptide synthesis is that it allows the synthesis of a variety of polypeptide analogues.

A successful coupling reaction should occur in the presence of all the other functional groups that are commonly found in proteinogenic amino acids, and proceed rapidly to give the product with no side-reactions. Unprotected peptides are also soluble at high (millimolar) concentrations in chaotropic solvents such as: 6 M guanidine hydrochloride and 8 M urea, which can be beneficial in promoting rapid and high yielding reactions.

Employing unprotected peptide segments in a convergent reaction route will provide intermediates that are also unprotected, and these are preferred because they are easier to handle, purify, and characterise than protected peptides.[133] This is because the presence of hydrophobic and bulky protecting groups tends to decrease the solubility of the peptide in typical HPLC solvents, and this makes it difficult to handle, purify, and characterise any intermediates and products.

Overall, a convergent synthesis strategy for polypeptide synthesis can overcome many of the shortcomings related to classical solution phase synthesis using protected peptides. Importantly, convergent synthesis can avoid the accumulation of resin-bound impurities that otherwise limits stepwise solid phase peptide synthesis to chains containing approximately fifty amino acid residues. For these reasons, chemoselective ligation forms the basis of modern chemical protein synthesis and has been used in the synthesis of classic protein molecules such as HIV-1 protease[38] and human lysozyme.[137]

1.5 Dibenzofuran-Assisted Chemoselective Ligation

Auxiliaries have been employed to assist chemoselective ligation by facilitating a high local concentration of both the N- and C-terminus of the peptide. Kemp and co-workers developed an auxiliary with an interjecting dibenzofuran group,[138] and the resulting proximity of the reactive termini was proposed to encourage ligation (Scheme 1.5).

The reactive α-amino group attacks the carbonyl group *via* a 12-membered-ring transition state, rather than a more conventional 5- or a 6-membered-ring transition state.[138]

Scheme 1.5 Kemp's prior thiol capture auxiliary incorporates a dibenzofuran scaffold to increase the proximity between the ligating termini. This in turn facilitates amide bond formation, and the dibenzofuran moiety can subsequently be removed with a reducing agent.[138]

1.6 Non-Native Bond Formation in Chemoselective Ligation

In 1992, the concept of chemoselective condensation of unprotected peptides was demonstrated by Schnölzer and Kent.[139] Principally, a chemoselective condensation reaction involves the reaction of two exclusive functional groups that are present on two different molecules. The functional groups are specifically designed to react with each other in the presence of all the other functional groups within the molecules, to give a sole product.

Schnölzer and Kent envisaged that a reaction that enabled the formation of a *non-native* covalent linkage between two reacting peptide segments greatly simplified the problem of joining two peptides, and also readily generated analogues of the native polypeptide.[139] Henceforth the term 'chemical ligation' was introduced to describe the chemoselective condensation of two unprotected peptides to produce a unique covalent polypeptide product.

Initial chemistry developed under the 'chemical ligation' banner included a nucleophilic substitution reaction between a peptide-α-carboxy thioacid (-$^{\alpha}$COSH) and a bromoacetyl peptide, which was conducted at low pH and in aqueous media (Scheme 1.6).[139] The resulting product was the thioester-linked analogue of the desired polypeptide and various

protein analogues have been synthesised using this non-native bond forming reaction,[140] including 'backbone engineered' HIV-1 protease which was used to determine the role of the backbone hydrogen bonds during enzymatic catalysis.[141] However despite these significant advances, the fact remained that a non-peptide bond was retained in the product.

Scheme 1.6 A non-native thioester linkage can be formed between two unprotected peptides through thioacid alkylation to provide access to protein analogues.[139-141]

1.7 Native Chemical Ligation

Chemical ligation is by definition a chemoselective reaction between unprotected peptides in solution, and was introduced for the total synthesis of proteins. A significant advantage in the total chemical synthesis of proteins is the potential to utilise organic chemistry to introduce precise modifications.

Therefore, an efficient chemoselective reaction should make use of the highly pure, fully characterised unprotected peptides that can be readily prepared by SPPS. It should form the full-length ligated product rapidly with no side-reactions, and allow the purification and characterisation of any intermediates. Wieland *et al.* had previously observed that an amide bond formed upon the reaction of Val-SPh with Cys,[142,143] and in 1994 'Native Chemical Ligation' (NCL) was introduced by the Kent laboratory as a method for polypeptide synthesis.[144] This was a significant advance because NCL overcame the SPPS limitation of peptide chains containing a maximum of approximately fifty amino acid residues.[52,53]

Native Chemical Ligation has since become an extremely powerful method for the routine synthesis of small to moderate-sized proteins (approximately 150 amino acids), and it involves a chemoselective reaction between two polypeptides in solution: an unprotected peptide-α-carboxy thioester (-$^{\alpha}$COSR) and an unprotected peptide with an N-terminal cysteine residue.[144] These two appropriately functionalised fragments react to form a native amide bond at the ligation junction (Scheme 1.7).[144,145] Key agreeable aspects of this reaction

include that it occurs in water, using unprotected peptides, at neutral pH and room temperature.[144,145] An early illustration of the utility of NCL was demonstrated with the synthesis of a 72 amino acid chemokine protein [Ala[33]]IL-8 which was prepared from two unprotected peptides.[144]

Scheme 1.7 The most extensively used and reliable method for protein synthesis: Native Chemical Ligation (NCL).[144] The first step of NCL is transthioesterification which is reversible and also rate-determining. The second step is an intramolecular *S→N* acyl transfer which results in a native amide linkage between the two ligated peptides.[144,145]

Hence, SPPS can be used to synthesise the appropriately functionalised peptides of up to approximately fifty amino acids in length. Two such synthetically pure and fully-characterised peptides can then be ligated in an NCL reaction to produce a polypeptide chain of about 100 amino acids in length. Convergent chemical ligations of unprotected peptides can also be conducted in aqueous media, to provide access to polypeptides that contain approximately 200 amino acids.[145]

Mechanistically, the first step in NCL is the intermolecular exchange of the Xaa-thioester with the thiol moiety of the N-terminal cysteine side chain to give a thioester-linked branched intermediate (Scheme 1.7).[144,145] Under the same reaction conditions throughout, this initial covalent product then undergoes a rapid intramolecular *S→N* acyl rearrangement to form the thermodynamically favoured more stable amide bond at the Xaa-Cys ligation site.[144,145] Under the NCL reaction conditions, the first thiol-thioester exchange is reversible but the subsequent rearrangement to the amide is irreversible, and this affords the ligated product in high yield.[144,145] Consequently, the rate of NCL depends upon the nature of the peptide-α-thioester thiol leaving group in the initial transthioesterification step and the nature of the amino acid thioester, whereas the rearrangement is rapid.[144,145] NCL also permits the use of high peptide concentrations (usually greater than 1 mM) under the denaturing reaction

conditions used, usually 6 M guanidine hydrochloride, to give clean ligation products in high yields.[144,145]

Many peptide thioesters have been shown to be successful in NCL reactions, however the synthetically most convenient is considered to be the benzyl thioester.[144] Alkylation of crude peptide thioacid with benzyl bromide in aqueous buffer at pH 4.0 can be used to generate the α-carboxy benzyl thioester. The utility of such a peptide benzyl thioester was demonstrated during the synthesis of interleukin 8 (IL-8).[144] The NCL reaction of the benzyl thioester peptide was shown to be about 60 % complete after 3 days. During the ligation, a large excess of benzyl mercaptan (BnSH) was employed to keep all the cysteine thiol side chains reduced and also reverse the formation of any unproductive thioesters that cannot further rearrange to produce the desired ligated product.[144,145]

As the first intermolecular transthioesterification step in NCL is the rate-determining step, the rate of NCL can be accelerated through the addition of an excess of a thiol additive which consequently increases the yields of the ligated product. Generally, aryl thioesters are more reactive than the alkyl thioesters due to their lower pKa values which make them better leaving groups. Therefore, a phenyl thioester is more reactive than a benzyl thioester and can be produced *in situ* by adding an excess of thiophenol (PhSH) which undergoes intermolecular thiol-thioester exchange.[146]

The transthioesterification of a model peptide with a highly reactive thioester Leu-Tyr-Arg-Ala-Gly-$^{\alpha}$COS-(2-nitrobenzoic acid) into the less activated thioester Leu-Tyr-Arg-Ala-Gly-$^{\alpha}$COSBn, was conducted by treatment with excess (2 % v/v) benzyl mercaptan in aqueous 100 mM sodium phosphate buffer at pH 6.5. After less than 30 min the transthioesterification was about 50 % complete, with negligible hydrolysis. Under the same reaction conditions but with excess (2 % v/v) thiophenol, a different peptide Ala-Glu-Ile-Ala-Ala-$^{\alpha}$COSBn was about 50 % converted within 3 h into the more reactive peptide phenyl thioester Ala-Glu-Ile-Ala-Ala-$^{\alpha}$COSPh. Additionally, a weakly activated peptide 2-mercaptoacetic acid (MAA) thioester Leu-Tyr-Arg-Ala-Gly-$^{\alpha}COSCH_2CO_2H$ was about 50 % converted into the more reactive peptide phenyl thioester Leu-Tyr-Arg-Ala-Gly-$^{\alpha}$COSPh within 1 h, under identical reaction conditions.[146] Hence, the reactivity of the peptide thioester can be tuned by the addition of a thiol additive to produce a better or worse leaving group in the initial transthioesterification step, to control the rate of

NCL.[146]

The synthesis of analogues of the enzyme barnase, which is a 110 amino acid residue microbial ribonuclease and is extensively used to study protein folding, were undertaken to test the application of *in situ* transthioesterification.[146] NCL requires a peptide thioester and an N-terminal cysteinyl peptide, however barnase contains no native cysteine residues and therefore ^{49}Lys was replaced by ^{49}Cys to create a ^{48}Gly-^{49}Cys ligation site, with little or no effect expected on protein folding or enzyme activity. Barnase(1-48)$^{\alpha}$COSBn and [Lys49Cys]barnase(49-110) were initially prepared by Boc-based SPPS.[147]

[Lys49Cys]barnase was synthesised by NCL using nearly equimolar quantities of barnase(1-48)$^{\alpha}$COSBn and [Lys49Cys]barnase(49-110), which were dissolved in 6 M guanidine hydrochloride to a concentration of 1 mM in each, with 100 mM sodium phosphate (pH 7.5) containing excess thiophenol (2 % v/v). After 4.5 h the ligation was judged to be complete, and (RP)HPLC purification followed by characterisation and folding of the synthetic [Lys49Cys]barnase allowed the enzyme activity to be assessed as being equivalent to that of wild-type barnase.[146]

Hence, NCL is a robust method that has been used to synthesise many proteins, however a problem that can be encountered is if a peptide were to contain both an unmasked N-terminal cysteine and a C-terminal thioester. Under typical NCL conditions the expected product would be the cyclic peptide.[148] Therefore, if multiple ligations are to be conducted for the synthesis of long polypeptide chains then the unreacting functionality needs to be temporarily protected to prevent undesirable intramolecular cyclisation. This includes, for example, the reversible protection of an N-terminal cysteine residue with a 1,3-thiazolidine group (Thz)[149] which is stable to treatment with TFA and anhydrous HF. Acylated thiazolidine rings can be opened with mercury(II) acetate to afford cysteine and non-acylated thiazolidine rings can be opened under mildly acidic conditions (pH 4.0) with 0.2 M methoxylamine hydrochloride.[149,150]

1.8 Inteins

In 1990, a biological *N→S* acyl shift was identified in the mechanism of natural protein splicing elements termed inteins (Scheme 1.8).[151-154] An intein is named by analogy to an

intron in nucleic acids and consists of an internal protein domain that is flanked by two polypeptides, commonly termed the N-extein and the C-extein.

Scheme 1.8 The accepted mechanism of protein splicing.[151-154] (a) An initial *N*→*S* acyl shift at the N-terminus of the intein results in a thioester-linked branched intermediate which then undergoes (b) intramolecular transthioesterification with a C-terminal cysteine residue. (c) At the C-terminus of the intein, asparagine cyclisation occurs to form a succinimide which detaches the intein from the ligating exteins. (d) The succinimide can then hydrolyse and (e) the N- and C-terminal exteins undergo an intramolecular *S*→*N* acyl shift to afford a native amide bond between the ligated exteins.

Inteins mediate protein splicing through a sequence of highly regulated acyl transfers beginning with a self-catalysed intramolecular rearrangement of the intein N-terminal

cysteine residue. This forms a thioester at the N-intein splice junction, which further undergoes transthioesterification by nucleophilic attack of the first residue (cysteine) in the C-extein to form a branched intermediate. The intein is then excised by cyclisation of asparagine (or glutamate) at the C-terminus of the intein splice junction. The resulting thioester linked exteins ligate through an NCL-like *S*→*N* acyl shift to form a stable amide bond (Scheme 1.8).[151-154] Over 350 members of the intein family are known and mutant inteins have been developed that participate only in the initial or final stages of protein splicing.[155]

An NCL reaction between an intein-generated protein thioester with a synthetic cysteinyl peptide to provide access to larger proteins (greater than about 150 amino acids) is termed Expressed Protein Ligation (EPL) (or semisynthesis), and was introduced by the Muir laboratory (Scheme 1.9).[156]

Scheme 1.9 Expressed Protein Ligation (EPL) is an NCL reaction between a recombinant protein thioester and a synthetic cysteinyl peptide, and it proceeds by resembling the initial stages of protein splicing.[156] (a) An initial *N*→*S* acyl shift occurs at the N-terminus of an intein which is bound through its C-terminus to a solid support, to afford a thioester-linked branched intermediate. (b) The presence of a thiol (i.e. MESNa) promotes transthioesterification to give the protein thioester. (c) The protein thioester can then participate in an NCL reaction with a synthetic cysteinyl peptide to afford a new amide bond.

Generally, EPL requires the expression of a protein of interest that is fused at its C-terminus to a mutant intein that is often associated with an affinity tag, such as a chitin binding domain

(CBD), for ease of purification. The protein expressed corresponds to the N-terminal extein and after an initial *N→S* acyl transfer, the protein thioester can be released from the intein *via* an intermolecular thiol-thioester exchange. This transthioesterification step is driven by the addition of a thiol (RSH) such as sodium 2-mercaptoethanesulfonate (MESNa). The resulting protein thioester can then participate in NCL.[157]

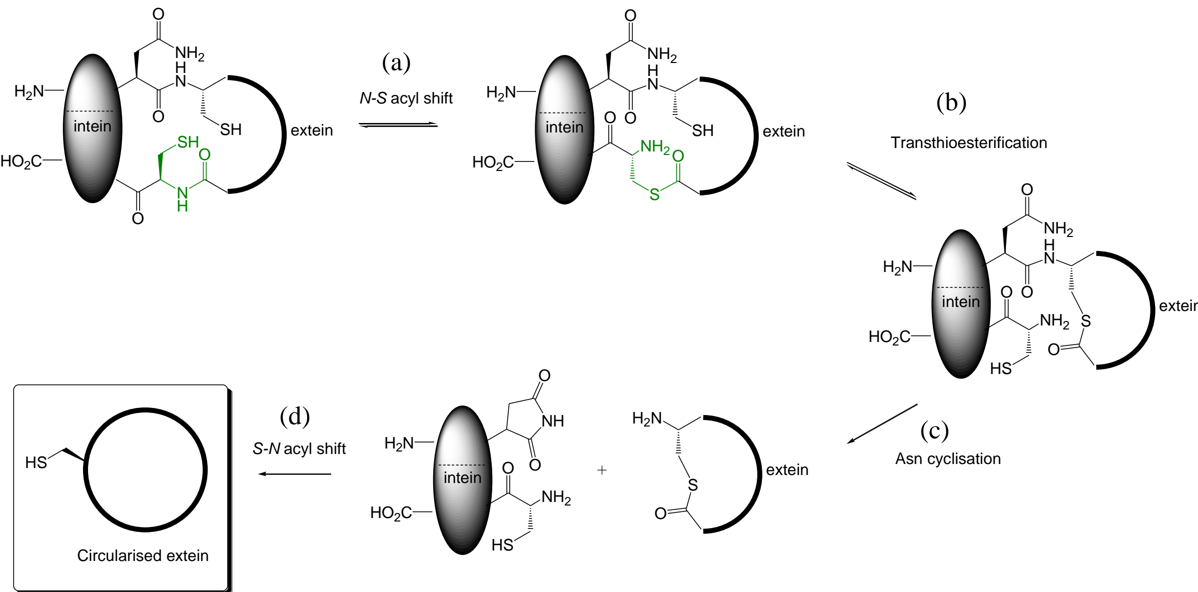
Alternatively, if the protein were to correspond to the C-extein, it would require N-terminal cleavage of the intein, which can be promoted by changing the pH and temperature of the solution, to generate a protein with an N-terminal cysteine.[158] Other methods to generate a recombinant N-terminal cysteinyl protein,[159] include:

- Cyanogen bromide (CNBr) mediated cleavage of chemically methylated cysteine residues by formic acid to expose the N-terminal cysteine residue.[160,161]
- An N-terminal methionine residue can be cleaved with an endogenous methionyl aminopeptidase to generate an N-terminal cysteinyl protein.[162,163]
- Factor Xa protease is commercially available and it requires a four amino acid residue recognition sequence (IEGR) after which it cleaves to release the N-terminal cysteinyl protein.[164]
- TEV protease can be over-expressed in *E. coli* and can be used to produce the N-terminal cysteinyl protein, though it does require a specific seven amino acid residue recognition sequence {ENLYFQ(S/G)} where S/G can be mutated to C.[165]
- The 3C protease ($3C^{pro}$) from Foot-and-Mouth Disease Virus (FMDV) has also been shown to produce an N-terminal cysteinyl protein for NCL reactions through selective proteolytic cleavage.[166]

Therefore, intein fusion systems provide a source of recombinant proteins that can be functionalised with either an N-terminal cysteine or a C-terminal thioester to be used as precursors in NCL or EPL,[167] and are commercially available.

In addition to the use of the so-called *cis*-splicing intein elements, protein *trans*-splicing (PTS) occurs through the use of split inteins.[168-171] These inteins are split (either naturally or engineered) into their complementary "halves" and each pair of halves has a different affinity for subsequent reassociation and this selectivity can be exploited to conduct multiple simultaneous ligations, as crossover reactions do not occur in PTS. Interestingly, split inteins have also been used to synthesise cyclic proteins (Scheme 1.10).[172]

Commercially available intein technology is often used in conjunction with labelled synthetic peptide fragments to act as biotechnology tools which provide valuable insight into the structure and function of a protein.[173-177]



Scheme 1.10 Protein cyclisation mediated by a split intein system.[151-154] (a) An *N→S* acyl shift occurs to afford the thioester-linked branched intermediate. (b) The intermediate is then transthioesterified. (c) Asparagine cyclisation occurs to detach the intein from the extein. (d) An *S→N* acyl shift occurs to provide a circularised extein.

High-resolution N-extein-intein crystallographic structures have provided insight into the mechanism of the first step of protein splicing. The *N→S* acyl shift occurs at a scissile amide bond at the N-terminus of a functional intein, which is situated close to highly conserved "block B" histidine and threonine residues. The imidazole side chain of the conserved histidine appears to function as a general acid or base catalyst to facilitate the acyl transfer.[31,178,179]

The conformation of the scissile amide bond itself has been the subject of considerable debate, and it has been shown to adopt many conformations including the normal and most stable *trans*, distorted *trans*, and *cis* configuration. Conformational backbone distortion from the ideal 180° may occur as a result of strain from specific interactions with conserved intein residues, which activate the scissile peptide bond and hence

promote nucleophilic attack by the adjacent cysteine thiolate anion due to loss of π symmetry to form a twisted amide bond,[178] and is generally referred to as ground-state destabilisation, which is thought to be the driving force for cleavage. The resulting tetrahedral intermediate can then be stabilised by other intein active site side-chain interactions such as backbone hydrogen bonding, analogous to those observed in an oxyanion hole.

An unusual amide bond conformation has not been unanimously reported for all the N-extein-intein crystal structures solved to date,[179] but this may be explained by the use of modified inactive inteins. The conformation of the scissile amide bond within an *active* N-extein-intein precursor possesses a short half-life, and has been investigated by solution NMR spectroscopy and segmental isotopic labelling.[31]

Semisynthesis was used to generate active and inactive fused inteins. An N-extein peptide (AAMRF) was ligated to either wild-type or a mutant *Mycobacterium xenopi* DNA gyrase A (*Mxe* GyrA) miniintein lacking a homing endonuclease domain to afford two new constructs: (1) N-extein-WT *Mxe* GyrA intein with His-75 intact, or (2) N-extein-*Mxe* GyrA intein with a His75Ala mutation. The imidazole ring in the conserved His-75 was well positioned to donate a proton and break down the tetrahedral intermediate from an *N→S* acyl shift, and was shown to be hydrogen bonded to the amide nitrogen of the scissile peptide bond.[31]

The N-extein contained a single ^{13}C-1 labelled phenylalanine, whereas the recombinant intein was uniformly ^{15}N labelled. However, because only the scissile amide bond was dual-labelled (^{13}CO^{15}NH) for both ligated constructs, 2D-HNCO NMR spectra were compared and showed a shift in the amide signals from an unusally upfield chemical shift of δ_H 6.61 ppm ($^1J_{NC'}$ ~12 Hz) in the active intein, to δ_H 10.01 ppm ($^1J_{NC'}$ ~16 Hz) in the mutant construct.[31] In general, proteins have been found to have $^1J_{NC'}$ coupling constants between 13-17 Hz,[180,181] but hydrogen bonding to the amide NH has been shown to decrease the $^1J_{NC'}$ value.[31] This occurs as a result of the resonance stabilised amide bond containing highly polarised π electrons. Increased rotation about the C(O)-N bond has also been predicted to decrease the $^1J_{NC'}$ value through an increased C(O)-N bond length. This study revealed an unusually low $^1J_{NC'}$ coupling value for the scissile amide bond within an active intein.[31] It was attributed to unusual peptide bond polarisation for which the conserved His-75 was essential, and provided further evidence for the *N→S* acyl

transfer proceeding with a contribution from destabilisation of the scissile amide bond and an adjacent conserved histidine residue.[31]

Overall, despite the prominent role of inteins in protein engineering relatively little is understood about their mechanistic details at the atomic-level. This is presumably due to the practical difficulty in studying an active splicing precursor. Consequently, most mechanistic studies have employed inactive mutants. Several mutagenesis, biophysical and biochemical studies have served to elucidate the overall sequence of steps, and have identified conserved proximal residues (particularly histidine residues) to be essential for the early and late stages of the splicing process. These results have suggested that histidine plays an important role as a general acid and base catalyst, whilst the amide bond is believed to have atypical geometry.[31,178,179] It has therefore been proposed that protein splicing occurs as a consequence of a combination of intein-mediated amide distortion, generally termed ground state destabilisation (evidenced by the unusually low $^{1}J_{NC'}$ coupling constant of ~12 Hz),[31] and the catalytic features of proximal (mainly histidine) residues.[182]

1.9 Kinetically Controlled Ligation

Kinetically controlled ligation was also developed by the Kent laboratory and it exploits differences in thioester reactivity during convergent NCLs, and the order of the NCLs is influenced by the choice of the added thiol catalyst.[183] For example, the aryl thiol thiophenol has a lower sulfhydryl group pKa = 6.6, and therefore is a better leaving group during the initial rate-determining transthioesterification step than the alkyl thiol sodium 2-mercaptoethanesulfonate (MESNa), which has a higher sulfhydryl group $pK_a = 9.2$.[183]

Kent and colleagues showed that the incorporation of a new 4-mercaptophenyl acetic acid (MPAA)-based linker during Boc-SPPS enabled the quantitative installation of the aryl thioester at the C-terminus of a peptide (Scheme 1.11).[183] They also demonstrated the difference in reactivity between aryl and alkyl thioesters by conducting a one-pot experiment. They showed that a peptide with this MPAA-derived C-terminal aryl thioester preferentially reacted with an N-terminal cysteinyl peptide in high yield, without the use of aryl thiol additives, whilst the alkyl thioester could be recovered without having undergone ligation.[183]

Scheme 1.11 The 4-mercaptophenylacetic acid (MPAA) linker was incorporated during the Boc-based solid phase peptide synthesis of a peptide with a C-terminal MPAA-derived thioester.[183]

A study was conducted that screened several potential thiol catalysts and their effect on the rate of NCL. Of particular note was the water-soluble thiol 4-mercaptophenyl acetic acid, MPAA, which was also the main constituent of the linker that furnished the aryl thioester on solid support (Scheme 1.11), and was observed to promote the most rapid NCL reactions.[184]

The utility of kinetically controlled ligation (Scheme 1.12) was demonstrated with the synthesis of the 46 amino acid small protein crambin,[150] which was convergently assembled from six peptide segments. During the convergent ligations, only two of the N-terminal cysteine residues were protected with a thiazolidine (Thz) group to prevent undesired ligations. Unwanted side reactions resulting from the intramolecular formation of cyclic peptides or intermolecular reactions to form oligomeric peptides comprised less than 2 % of the reaction mixture. The use of methoxylamine hydrochloride at pH 4.0 (rather than pH 6.0) was sufficient to deprotect the cysteine residues when required without decomposition of the thioesters.[185]

Scheme 1.12 Kinetically controlled ligation allows the researcher to pre-organise the order of NCL reactions by strategically incorporating within an unprotected peptide a more reactive aryl thioester (-COSAr) where NCL should occur first, rather than an alkyl thioester (-COSR).[184]

1.10 Sugar-Assisted Ligation

Sugar-assisted ligation was established by Wong *et al.* and proceeds through an intramolecular *S→N* acyl transfer *via* a fourteen-membered-ring transition state (Scheme 1.13).[186] The enlarged fourteen-membered-ring transition state may be expected to proceed less efficiently than a five- or a six-membered ring transition state, however a templating effect of the carbohydrate is proposed to facilitate ligation. Ligation is achieved in a peptide that has a thioester substituted onto the *N*-acetyl group of a 2-acetamido-2-deoxy sugar moiety, which in turn is attached to an N-terminal penultimate serine residue of the peptide through a *β-O*-glycosidic linkage.[186]

The scope of sugar-assisted ligation was tested with glycine, histidine, alanine, and valine thioesters that were successfully ligated to an N-terminal glycine residue adjacent to the required glyco-serine.[186] Other less sterically demanding amino acids, alanine and serine, were also stated as acceptable at the N-terminus of the peptide.[186]

1. *S*→*N* acyl shift
2. desulfurisation

Scheme 1.13 Sugar-assisted ligation.[186] An *S→N* acyl rearrangement occurs initially to form a native amide bond, which is followed by desulfurisation to afford the *O*-linked glycopeptide.[186,188]

Certain issues have however been raised with regards to the necessity of the *β*-stereochemistry in the *O*-glycosidic linkage; the toleration of other carbohydrates in the allegedly template-driven ligation; and also whether the unsuccessful ligation of an N-terminal glycoamino acid can be attributed to steric or conformational strain.[187]

An important limitation of sugar-assisted ligation is that after ligation, the thiol moiety on the sugar needs to be desulfurised to release the native *N*-acetyl group, which is

incompatible with the presence of any free cysteine side-chain thiols within the peptide.[186] However, if a glycosidase can be used to remove the sugar that is linked to the peptide after ligation, then sugar-assisted ligation has the potential to become a traceless reaction.[187] Nonetheless, disadvantages of this method include the initial synthesis of the sugar bearing the 'thiol handle' and its subsequent stereospecific linkage to a serine residue.[187]

Sugar-assisted ligation was also found to be limited to certain types of extended glycosylation.[188] Extended glycosylation at the C-3 position of the existing sugar bearing the thiol handle prevented ligation from occurring, whereas it was possible to ligate with extended glycosylation at positions C-4 and C-6 of the existing sugar.[188]

1.11 Phenolic Esters in the Synthesis of Posttranslationally Modified Peptides

Danishefsky and co-workers employed a peptide with a C-terminal phenolic ester, which was *ortho-* substituted with a protected disulfide group, in chemical ligations (Scheme 1.14).[189,190] Solution-phase syntheses of short, complex glycopeptides (containing potentially acid-labile glycosidic linkages) were accomplished using the phenolic ester, which are otherwise considered incompatible with the acidic conditions used during Boc-based solid phase peptide synthesis.

In principle, native *N*-linked glycopeptides can be prepared through the incorporation of an *N*-linked glycoamino acid (i.e. glycosylated asparagine) during stepwise peptide synthesis, though the attached sugars are often (acetate) protected. Alternatively, oligosaccharides can be globally introduced after the peptide chain has been assembled. Danishefsky *et al.* utilised the latter approach to synthesise an *N*-linked glycopeptide with a C-terminal phenolic ester.[189,190]

A C-terminal phenolic ester was chosen because coupling of the glycosyl amine with the side-chain carboxylic acid of, for example, aspartic acid in the presence of a C-terminal thioester is disfavoured as it provides two competing sites of reaction. However, phenolic esters are comparatively less reactive than thioesters.[189]

After peptide chain assembly, the side-chain of the aspartic acid residue was then selectively deprotected.[189] The oligosaccharide was then coupled to the aspartic acid residue within the peptide sequence. Initially, the aspartic acid side-chain carboxylic acid

was converted into an active ester *in situ* (i.e. reaction conditions: HATU, DIPEA, DMSO, 72 % yield).[190] The glycosyl amine was then coupled to the active ester to afford the *N*-linked glycopeptide with a C-terminal phenolic ester.[190]

A trifunctional peptide was also prepared through sequential chemical ligations.[190] The disaccharide chitobiose was coupled to an aspartic acid residue in a peptide containing an N-terminal thiazolidine (Thz), which reversibly masked the cysteine, and a C-terminal phenolic ester bearing a protected *ortho*-disulfide substituent. The disulfide substituent on the phenyl ring was then reduced with the phosphine-based reducing agent *tris*(2-carboxyethyl)phosphine (TCEP), which allowed an *O→S* acyl shift to occur and produced the C-terminal thioester *in situ* (Scheme 1.14).

Scheme 1.14 Proposed mechanism of auxiliary-mediated ligation with a C-terminal phenolic ester rather than a thioester.[189,190] After reduction of the disulfide bond with TCEP, an intramolecular *O→S* acyl shift occurs to form a thioester. Transthioesterification then occurs between the thioester and the thiol group in the 4,5,6-trimethoxy-2-mercaptobenzyl (Tmmb) auxiliary. An intramolecular *S→N* acyl shift then occurs to form the native peptide bond, and the auxiliary is finally removed with TFA.

Another peptide with an N-terminal disulfide protected 4,5,6-trimethoxy-2-mercaptobenzyl (Tmmb) auxiliary present within the mixture was concomitantly reduced by TCEP to afford the free thiol, which then enabled auxiliary-mediated NCL to take place.[190] The ligated trifunctional glycopeptide contained: an N-terminal thiazolidine

group; two sites of *N*-linked glycosylation; and a glutamine residue bearing the Tmmb auxiliary (Scheme 1.15).[190] The thiazolidine group could be further deprotected with methoxylamine.HCl and reacted further in an NCL reaction. The solid-phase production of thioesters from esters through an *O→S* acyl transfer has also been shown though it is not commonly used.[191]

Scheme 1.15 A C-terminal phenolic ester was used in the synthesis of a trifunctional *N*-linked glycopeptide by Danishefsky *et al.* which contained an N-terminal thiazolidine group; two sites of *N*-linked glycosylation; and a glutamine residue bearing a Tmmb auxiliary.[190] *Reagents and conditions: a)* HATU, *i*Pr_2NEt, DMSO; *b)* TCEP, DMF, Na_2HPO_4; *c)* methyl *p*-nitrobenzene sulfonate; *d)* 95 % TFA.[190]

1.12 Other Strategies: Oxime-Forming Ligation and Staudinger Ligation

Alternative ligation chemistries have also been developed including oxime-bond forming ligation by Rose *et al.*, which involves the reaction of an amino-oxyacetyl-peptide with a glyoxylyl-peptide.[192] The reaction is conducted in aqueous solution at low pH and results in the formation of an oxime-linked ligated product that is stable at neutral pH. Oxime-forming ligation was used in conjunction with thioester-forming ligation to prepare topological analogues of the transcription factors cMyc and Max.[193]

A more recently developed chemoselective ligation method is Staudinger ligation which involves the formation of a native amide bond at approximately neutral pH. Initially, Saxon and Bertozzi showed that a phosphine derivative can function as an acyl donor when reacted with a glycosyl azide (Scheme 1.16i).[194] However, a phosphine oxide functionality remained in the ligated product. Therefore, a traceless version of the Staudinger ligation was developed for the chemical synthesis of proteins by Raines *et al.*, and it involved the reaction of a peptide carrying a C-terminal (phosphino)thioester with an N-terminal azido-peptide (Scheme 1.16ii).[195,196]

Scheme 1.16 (i) Staudinger ligation.[194] (ii) Traceless Staudinger ligation of peptides,[195,196] the loss of nitrogen and phosphine oxide drives the reaction. 'X' is a phenyl ring.[195]

Initially, an iminophosphorane is formed which allows the nitrogen of the iminophosphorane to attack the thioester to form an acyclic amidophosphonium salt. This salt is then hydrolysed to form an amide bond between the two peptides and generates phosphine oxide as a by-product.

Of all the chemical ligation strategies discussed here, by far the most extensively used method for the total chemical synthesis of proteins is Native Chemical Ligation.[144] However, whilst the N-terminal cysteinyl peptide for NCL can be readily prepared by standard stepwise peptide synthesis; preparation of the thioester is considerably more difficult.

1.13 Peptide Thioesters

Peptide and protein thioesters are extremely useful in chemistry and chemical biology, and are essential for the synthesis of site-specifically modified peptides and proteins by NCL.[144] Methods to synthesise thioesters have traditionally been based on activating the α-carboxylic acid of an amino acid, rather than the *N→S* acyl shift employed by inteins to produce thioesters.

Thioesters can be prepared by Boc-based solid-phase peptide synthesis on a thioester resin.[144] However, the harsh acidic conditions (usually anhydrous hydrogen fluoride, HF) employed in Boc-based SPPS used to cleave the peptide thioester from the solid support, despite being a generally high yielding method with demonstrated application to chemically fragile modifications such as *N*-linked glycopeptides,[197,198] has led to this chemistry being eclipsed by the more amenable Fmoc-based SPPS, which utilises a mild base, usually piperidine, during synthesis.[199-202] As a result, several methods to afford peptide thioesters using Fmoc-SPPS have been explored.

Since thioesters are unstable to piperidine, other non-nucleophilic bases have been investigated for the removal of the N-α-Fmoc group during peptide elongation with reasonable success; these have included 1-methylpyrrolidine/ hexamethyleneimine/HOBt[124,203] or DBU/ HOBt mixtures.[204] Use of low nucleophilic bases used during Fmoc-SPPS have however been shown to racemise chiral amino acids adjacent to the thioester.

A popular route for peptide thioester synthesis is a modification of Kenner's sulfonamide safety-catch linker method, and it employs standard Fmoc/*t*-Bu SPPS followed by sulfonamide alkylation to aid the release of the peptide thioester from the solid support (Scheme 1.17i).[199]

i)

ii)

iii)

Scheme 1.17 i) The typical sulfonamide safety-catch approach to peptide thioester synthesis employing Fmoc-based SPPS.[199] *Reagents and conditions: a)* ICH_2CN, DIPEA, DMF, 16 h r.t; *b)* RSH (often benzylmercaptan), NaSPh, DMF, 16 h, r.t; *c)* 95 % v/v TFA + scavengers. Recent developments in Fmoc-based peptide thioester formation: ii) *N*-acylisoureas.[210] *Reagents and conditions*: *d)* 0.5 M DIPEA in DMF, 15 min r.t; *e)* 95 % v/v TFA + scavengers; *f)* 200 mM MPAA (the thioester is generated *in situ*). iii) Peptide pyroglutamyl imide formation.[211] *Reagents and conditions: g)* 2 % v/v TFA + scavengers, DCM; *h)* PyBrOP, DIPEA, NMP, microwave 60 °C, 3 × 1 h; *i)* RSH, PhSNa, [15]crown-5, MeCN, 40 °C overnight; *j)* 95 % v/v TFA + scavengers. Pg = protecting group.

The resulting peptide thioesters are often isolated in low yield[205] and for post-translationally modified peptides such as glycopeptides, the activating alkylation step can also result in the undesirable alkylation of unprotected hydroxyl groups on an attached carbohydrate as well as on methionine residues.[199] Consequently, synthetic strategies that release fully protected peptides with C-terminal carboxylic acids or carboxamides from the solid-support prior to conversion to thioesters have been established. This is because preparation of such species is generally higher yielding and easier to monitor (Scheme 1.17).[206-211]

A backbone amide linker (BAL) with Fmoc-SPPS has also been used to prepare peptide thioesters with a masked thioester incorporated as a trithioortho ester.[212,213] Alternatively, peptide thioesters can be introduced after chain assembly to avoid exposure of the thioester to basic conditions during synthesis. This has been shown with a BAL extended by orthogonal allyl protection of the C-terminal α-carboxylic group.[214]

1.14 *N→S* Acyl Transfer Using the Sulfonamide-Linker

The past half a decade has seen an emergence in methods to produce peptide thioesters through an elegant *N→S* acyl transfer, as employed by inteins.[151-154] Melnyk and co-workers reported a solid-phase *N→S* acyl transfer for thioester synthesis after peptide chain assembly using Fmoc/*t*-Bu chemistry in combination with the sulfonamide safety-catch linker (Scheme 1.18).[215]

Scheme 1.18 Rink amide resin is loaded with 3-carboxypropanesulfonic acid before peptide assembly. After peptide assembly, the silyl protecting group is removed and this allows an *N→S* acyl shift to occur to afford the peptide thioester. *Reagents and conditions: a)* TBAF/AcOH; *b)* 95 % v/v TFA + scavengers.[215]

The acylsulfonamide group within a linker has already enjoyed widespread use in the modified safety-catch linker method for thioester synthesis,[199,216] but in this case it can undergo additional Mitsunobu alkylation with a mercaptoethanol derivative, which allows the

introduction of the prerequisite *β*-amino thiol component to facilitate the subsequent intramolecular *N*→*S* acyl shift.[215]

The isolated yield of the thioester was found to be dependent upon the degree of alkylation, which was independent of steric influence from the first amino acid due to the remoteness of the alkylation site from C-α.[215] This synthetic methodology also has the advantage of allowing NCL to potentially proceed on solid-support, though the *N*→*S* acyl shift can proceed with concomitant cleavage from the resin to allow isolation of the thioester.[215] A disadvantage of this method is that formation of the resin-bound acylsulfonamide can be inefficient and the subsequent Mitsunobu reaction can be relatively difficult to monitor, especially when using the commercially available 4-sulfamylbutyryl-linked resins.

1.15 *N→S* Shift in Acyl-Transfer Auxiliaries

The natural abundance of cysteine residues in proteins is relatively low (less than 2 %) and cysteine is therefore not always conveniently situated at desired ligation junctions. Strategies to overcome this limitation include ligation between thioester peptides and N-terminal homocysteinyl or homoselenocysteinyl peptides, followed by methylation to afford Xaa-Met or Xaa-Seleno-Met sequences respectively.[217,218] Similarly, the utilisation of an N-terminal cysteinyl peptide for NCL followed by desulfurisation has been shown to afford an Xaa-Ala ligated junction[219,220] and analogous strategies have evolved for several additional amino acid residues.[221-224]

NCL was further extended by the use of cleavable auxiliaries that could be appended to potentially any N-terminal amino acid, though most commonly at the least hindered glycine or alanine residues, to facilitate cysteine-free NCL.[225-231]

Peptides are synthesised to contain an N-terminal amino acid attached to the acyl transfer auxiliary *via* N-α. Under typical NCL conditions, the peptide thioester undergoes transthioesterification with the thiol moiety of the auxiliary-linked peptide, and an *S*→*N* acyl transfer ensues to afford the full-length polypeptide. Usually, removal of the auxiliary is accomplished upon treatment with acid[232] (or by photolysis)[233] to give the desired polypeptide. An important feature within the carbon skeleton of the auxiliary is that an NCL-like β, or γ amino-substituted thiol motif is retained to facilitate ligation *via*

a pre-organised five- or six-membered ring intermediate (Scheme 1.19a).[225,228]

The first transthioesterification step in NCL is the rate-determining step; however in auxiliary-mediated ligation, if the C-α of the thioester or the N-terminal amino acid attached to the auxiliary is substituted then the *S→N* acyl rearrangement becomes rate-determining.[187] Overall, the slower formation of the amide bond results in increased competition from thioester hydrolysis.[187] The sensitivity of thiol-containing auxiliaries to steric hindrance at the ligation site has also meant that cysteine-free NCL has not been as effective as NCL.[187]

Scheme 1.19 (a) Kent *et al.*'s 1-aryl-2-mercaptoethyl acyl transfer auxiliary is used in cysteine-free NCL,[225] which involves an initial transthioesterification step with the thiol moiety in the auxiliary. This is followed by an intramolecular *S→N* acyl shift to afford the native peptide bond, and the auxiliary is then removed. (b) Dawson *et al.*'s 4,5,6-trimethoxy-2-mercaptobenzyl (Tmmb) auxiliary is also used in cysteine-free NCL however the *S→N* acyl shift proceeds through a six- rather than a five-membered ring transition state.[228]

As described earlier (Section 1.11) Danishefsky *et al.* used the 4,5,6-trimethoxy-2-mercaptobenzyl (Tmmb) auxiliary disclosed by Dawson *et al.* (Scheme 1.19b)[228] in an auxiliary-mediated ligation across a glycine-glutamine junction, to synthesise a glycopeptide containing multiple asparagine-linked disaccharides in over 58 % yield.[190] However, after cysteine-free NCL of the glycopeptide Danishefsky *et al.* recounted an interesting observation upon cleaving the Tmmb auxiliary.[190] When the ligated product

was immersed in TFA to remove the Tmmb auxiliary, mass spectrometry revealed two species with identical molecular weight.[190] Interestingly, the two species were the desired auxiliary-free ligated '*N*-peptide' and also the '*S*-peptide', which was produced through an intramolecular *N→S* acyl shift and was driven by the liberation of the protonated amine (Scheme 1.20).[190] With no other free cysteine side chain thiols in the amino acid sequence, the *N→S* acyl transfer at the ligation junction was prohibited by methylation of the thiol in the auxiliary with methyl-*p*-nitrobenzenesulfonate before cleavage, and this allowed a glycopeptide comprised of 17 amino acid residues to be isolated.[190]

Scheme 1.20 An *N→S* acyl shift was observed during the cleavage of the 4,5,6-trimethoxy-2-mercaptobenzyl (Tmmb) auxiliary with TFA by Danishefsky *et al.*; however, prior methylation of the thiol group in the auxiliary before cleavage afforded the desired ligated product.[190]

This type of rearrangement was previously documented by Vorherr and Aimoto with the 2-mercapto-4,5-dimethoxybenzyl (Dmmb) auxiliary and has since been developed into a method for thioester synthesis (Scheme 1.21).[234,235]

Scheme 1.21 Peptide thioester synthesis using a solid-supported Dmmb auxiliary was shown to proceed through an *N→S* acyl shift under acidic conditions by Vorherr *et al.*[234] The solid-supported thioester was afterwards intercepted by 2-mercaptoethanesulfonic acid. *Reagents and conditions: a)* 88 % v/v TFA + scavengers, 1 h; *b)* 0.1 M 2-mercaptoethanesulfonic acid, DMF.[234,235]

The existence of the *N*→*S* acyl transfer was subsequently verified using ^{13}C NMR spectroscopy, (RP)HPLC, and mass spectrometry.[236] This was achieved through a two-part study of model systems: (i) a Dmmb containing dipeptide, Fmoc-Gly(^{13}*C*-1)-Ala(Dmmb)-OMe and (ii) a Cys containing pentapeptide, Fmoc-Ile-Ala-Gly(^{13}*C*-1)-Cys-Arg-NH_2, both of which were subjected to acid treatment.[236] Interestingly, while both peptides were shown to form thioesters by ^{13}C NMR, the glycinyl thioester formed in (ii) was shown to rapidly revert to the starting material upon attempted isolation, whereas the Dmmb-linked thioester could be isolated by (RP)HPLC.[236]

Dmmb auxiliary-assisted thioester synthesis, in solution or on solid phase, has recently been shown to produce peptide thioesters with up to 41 amino acid residues.[237,238] An Fmoc-Leu-D,L-[Dmmb(Trt)]-Ala motif was coupled to H-Phe-Alko-PEG resin and then extended by Fmoc-SPPS before being cleaved from the resin.[238] The Dmmb-attached peptide was then treated with dilute acid {0.25 M hydrochloric acid (HCl), pH ~0.6} containing 0.5 % TCEP at 37 °C for 3 h, followed by removal of the solvent under reduced pressure.[238] This created the *S*-peptide, which upon treatment with MESNa in an aqueous sodium acetate/ acetonitrile mixture, at elevated pH 6.0-7.0, produced the peptide-$SCH_2CH_2SO_3H$ thioester *via* intermolecular thiol-thioester exchange, with *N*-D,L-(Dmmb)Ala-Phe-OH as the by-product.[238]

The isolated yield for a bovine pancreatic trypsin inhibitor thioester BPTI(1-29)-$SCH_2CH_2SO_3H$ was 82 % based on the reacting *S*-peptide.[238] When 85 % v/v aqueous trifluoroacetic acid (TFA) was used in place of 0.25 M HCl, the peptide thioester yield was 44 % and the side products: [BPTI(1-29)-D,L-Ala-Phe-OH and BPTI(1-29)-D,L-(Dmmb)Ala-OH] were also identified by mass spectrometry.[238] The use of 0.25 M HCl suppressed such side reactions as well as deamination, also observed with 1.0 M HCl, which resulted in the conversion of Asn-24 to Asp-24 within the BPTI(1-29) sequence.[238] The milder acid appears to selectively protonate the Dmmb-linked amide which is again attributed to the conjugate acid of an *N,N*-dialkylamide having a higher pKa value than that for primary amines.

A similar yield was reported for the BPTI(1-29)-$SCH_2CH_2SO_3H$ thioester prepared with the *N*→*S* acyl shift occurring on-resin and no racemisation of the terminal Leu-29 residue was detected.[238] On-resin thioesterification with the Dmmb auxiliary was further used to synthesise a thrombopoietin receptor thioester (TpoR)(461-481)-$SCH_2CH_2SO_3H$ with

acid-sensitive linkages such as: tryptophan and the aspartate-proline sequence, in 34 % yield based on initial resin loading.[238] A telomere repeat-binding factor 2 thioester (TRF2)(1-41)-$SCH_2CH_2SO_3H$ was also prepared in 12 % yield.[238]

Additional peptides prepared also included modifications such as phosphorylation [Ser(PO_3H_2)^{10}]-histone H3(1-12)-$SCH_2CH_2SO_3H$ (31 %),[238] and methylation [Lys(Me_3)^{9}]-histone H3(1-33)-$SCH_2CH_2SO_3H$ (19 %).[238] Although some deamination of glutamine to glutamic acid could be observed in small amounts (less than 1.5 %) for certain peptides, the desired peptide thioesters could be separated by (RP)HPLC.[238] These results appear highly promising though drawbacks include the several synthetic steps required to prepare the auxiliary-linked resin, and loading of the first amino acid. Ideally, preloaded resins that allow the reasearcher to choose the amino acid thioester would become commercially available.

1.16 3-Mercaptopropionic Acid (MPA)-Mediated Thioesterification

Nakahara and co-workers described a post-peptide chain assembly thioesterification method that was used to produce a 61 amino acid glycopeptide from its 25 amino acid thioester.[239] The method involved a peptide bearing a C-terminal 5-mercaptomethylated proline derivative that was converted into a thioester upon treatment with aqueous 3-mercaptopropionic acid (MPA) (Scheme 1.22).[239]

Originally the peptide was prepared to comply with diketopiperazine (DKP) formation which is known to occur in peptides containing C-terminal proline-ester sequences.[239] The reaction occurs when a deprotected amine is able to react with an "activated" carboxylic acid to afford a bicyclic intermediate. However, the formation of the diketopiperazine tricyclic intermediate did not proceed as envisaged.[239] This was attributed to the diketopiperazine formation on-resin being slower than the reverse (*S*→*N*) acyl transfer, in weakly acidic or basic solution.[239]

The peptide with the C-terminal prolyl-prolyl ester motif was however successfully thioesterified *via* intermolecular thioester exchange with excess MPA.[239] At room temperature, the thioester of MPA could be observed within 12 h, using 40 % v/v MPA (pH 1.3) without any serious side reactions.[239] Extending the reaction time to 168 h resulted in complete consumption of the starting material.[239] Hydrolysis was reported to

be less than 15 %,[239] which nevertheless decreased the efficiency of the reaction. The reaction rate increased with microwave irradiation (150 W, 80 °C) and reactions were complete between 0.5-1 h, depending on the MPA concentration used.[239]

Decreased MPA concentrations (less than or equal to 20 % v/v) were correlated with increased reaction times, although the thioester purification was notably more facile.[239] Varying time and MPA concentration suggested that an impressive 70 % yield of thioester could be attained in only 10 min, using 40 % v/v MPA.[239]

Thus, the original peptide containing the C-terminal 'Pro-Pro ester' motif was experimentally concluded to be irrelevant because MPA transthioesterified all the *S*-peptide formed through an initial *N→S* acyl transfer.[239] The synthesis of an *N*-acetylglucosaminylated peptide thioester of emmprin(34-58) was achieved with 20 % v/v MPA and afforded yields that were over twice as high as those previously reported,[239] which showcased the efficiency of this MPA-thioesterification method.

A stated drawback of this method was the lengthy preparation of the initial 5-mercaptomethyl proline to be incorporated at the C-terminus of the peptide.[239] In addition, MPA thioesterification without microwave irradiation was observed to be considerably slower and increased the timescale to a week.[239] Furthermore, an aspartate-serine peptide bond was found to be sensitive to treatment with 10-20 % v/v MPA,[239] though a possible solution could be the employment of a suitable protecting group strategy. The driving force for the initial intramolecular *N→S* acyl transfer was stated as being the lability of the amide bond in the imino acid proline.[240]

Scheme 1.22 Nakahara *et al.*'s proposed thioesterification of a peptide with a C-terminal 5-mercaptomethyl prolyl-prolyl ester motif did not proceed as envisaged through route (a) *via* a diketopiperazine.[239] Instead, a peptide thioester was formed upon treatment with 3-mercaptopropionic acid (MPA). *Reagents and conditions: a)* mild H_3O^+ (or mild ^-OH); *b)* Reagent K, 60 °C, 4 h; *c)* 10 % v/v aqueous MPA, microwave, 40 °C, 1 h.[239]

1.17 *N*-Alkyl Cysteine-Assisted Thioesterification

Having accredited the *N→S* acyl shift to the presence of the imino acid proline, Nakahara *et al.* diverted their attention to the use of other imino acid "devices" that might promote thioesterification. *N*-alkylation of a C-terminal cysteine residue was devised as an alternative to facilitating the intramolecular *N→S* acyl migration under acidic conditions, which could be followed by intermolecular transthioesterification with MPA to afford the peptide thioester (Scheme 1.23).[240]

N-Ethyl cysteine was incorporated into the synthesis of emmprin(49-58) and evidence of an equilibrium between the *N*- and *S*-peptide was presented.[240] Both isomers could subsequently be transthioesterified with 5 % v/v MPA without any serious side reactions.[240]

The reaction rate with MPA was found to increase with increasing bulk of the R group attached to cysteine in the following order: Me< Et< *i*Bu.[240] The installation of an *N*-alkyl cysteine instead of a mercaptomethylated proline was shown to markedly increase the reaction rate of thioesterification using a lower MPA concentration (5 % v/v) and lower temperature (25 °C).[240] However, steric hindrance associated with the bulkier *N*-alkyl cysteine derivatives (*i*Bu and Bn) compromised the coupling of the adjacent glycine residue during synthesis, and hence the overall MPA thioester yield (28 % and 17 %) compared to *N*-Me Cys (33 %) and *N*-Et Cys (34 %).[240]

The *N*-ethyl cysteine derivative produced the highest peptide thioester yield and was relatively simple to prepare, although it did require HPLC purification.[240] This method was shown to be applicable for the preparation of peptide thioesters with chiral amino acids at the C-terminus in acceptable yields and levels of epimerisation.[240]

R = Me, 33 %
R = Et, 34 %
R = *i*Bu, 28 %
R = Bn, 17 %

Scheme 1.23 Nakahara *et al.* demonstrated substituent effects during *N*-alkyl cysteine-mediated thioesterification. *N*-Me Cys and *N*-Et Cys were found to be particularly favourable compared to the bulkier substituents *N*-*i*Bu Cys and *N*-Bn Cys. *Reagents and conditions: a)* 5 % v/v MPA, r.t., 2-3 days.[240]

N-Alkyl cysteine-mediated thioesterifcation was used to assemble a 95-amino acid chemokine, CCL27,[241] utilising the silver ion (Ag^{+}) promoted thioester method.[242] Notably, after cleavage from the solid-support followed by treatment with aqueous acetonitrile containing 6 M urea and 2 % MPAA, the more reactive MPAA aryl thioesters were obtained rather than the MPA-derived alkyl thioesters.[187] Recently, the *N*-alkyl cysteine thioesterification method has also been applied to the synthesis of a 22 kDa octameric glycopeptide dendrimer.[243]

The observed selectivity in *N*-alkyl cysteine-mediated thioesterification was attributed to a combination of distortion of the planarity of the amide bond due to *N*-alkylation, and liberation of a secondary amine which is a stronger base and thus favours protonation.[240]

1.18 Acyl Oxazolidinone and Aniline Linker-Mediated Thioesterification

Otaka and co-workers have additionally shown that a C-terminal cysteine residue when part of an acyl oxazolidinone moiety, can undergo an intramolecular *N→S* acyl transfer to yield thioesters (Scheme 1.24i).[244] In this case the ground state geometry of the exocyclic amide is considered compromised as a consequence of steric factors, imposed by the cysteine side chain, and electronic factors, since the $n_N \rightarrow \pi^*_{CO}$ interaction is additionally weakened by delocalisation of the nitrogen lone pair into the oxazolidinone carbonyl which lowers the energy barrier for the *N→S* acyl transfer.[244]

Despite the relatively long synthetic route employed to obtain the solid-supported oxazolidinone and its sensitivity to piperidine during peptide chain assembly prompting the use of Aimoto's reagent mixture [1-methylpyrrolidine/-hexamethyleneimine/1-hydroxybenzotriazole in NMP/ DMSO (1:1)] for Nα-Fmoc removal, this method was successfully applied to the synthesis of a 9-residue peptide thioester derived from human brain natriuretic peptide-32 (hBNP-32).[244] An NCL reaction between the 9-mer thioester and a synthetic N-terminal cysteine peptide, hBNP32-NH_2(10-32), was conducted in phosphate buffer at pH 7.6 containing 6 M guanidine hydrochloride in the presence of 1 % v/v thiophenol, and afforded hBNP32-NH_2 in 76 % yield for the NCL step.[244]

Racemisation of thioesters derived from chiral amino acids occurs with Aimoto's Fmoc removal reagent mixture during synthesis, and therefore only the Fmoc-glycyl oxazolidinone was successfully employed.[244]

Otaka *et al.* have also recently developed another device to facilitate an intramolecular *N*→*S* acyl shift for peptide thioester synthesis: an *N*-substituted aniline linker (Scheme 1.24ii).[245] It was hypothesised that anilide nitrogen rehybridisation could be effected by the introduction of two adjacent sp^2 hybridised atoms, ultimately causing activation of the scissile amide bond.[245] The core motif was based on *para*-(2-sulfanylethylamino)benzoic acid rather than the previously used acyloxazolidinone.

i) HS, peptide, N, O, O, O, $CONH_2$ — a) → [peptide, S, O, HN, O, O, $CONH_2$]

ii) HS, FmocHN, N, O, CO_2H — *N-S* acyl shift ⇌ FmocHN, S, O, N H, CO_2H

Scheme 1.24 Otaka *et al.* proposed ground state destabilisation to be a significant contributing factor to thioester formation using: (i) *N*-acyloxazolidinones[244] and ii) an *N*-substituted aniline linker.[245] *Reagents and conditions:* (i) *a)* phosphate buffer; pH 7.6, r.t. (ii) an isolable thioester is produced upon exposure to 4 M HCl/DMF, 1 % w/v TCEP, 8 h, 37 °C.

Consequently ground state destabilisation once more facilitated nucleophilic attack of the adjacent thiol onto the distorted amide carbonyl carbon, and the undesired reverse *S*→*N* acyl shift was only observed in small amounts.[245] This is because the aniline nitrogen is a poor nucleophile and this allows the equilibrium to favour the thioester.

The synthesis of a pentapeptide thioester (H-HRFAG-$SCH_2CH_2SO_3Na$) was demonstrated using an Fmoc-glycyl aniline derivative.[245] The *N*→*S* acyl shift for this was achieved by treatment with 4 M HCl/DMF in the presence of 1 % w/v TCEP for 8 h at 37 °C.[245] The resulting isolable thioester was then finally converted into the MESNa thioester.[245]

The aniline linker was also used to demonstrate an on-resin *N*→*S* acyl transfer.[245] This was shown with the synthesis of up to 9 amino acid residue peptide thioesters upon thiolysis in

approximately 70 % yields.[245] A small 9 residue phosphorylated serine-containing peptide thioester was also synthesised using the aniline linker in 67 % yield.[245] However, an unexplained partial racemisation occurred upon treating the H-FAA-*Ar* peptide with 4 M HCl/DMF, which was suppressed when TFA was used for the *N→S* acyl shift.[245]

1.19 Cysteinylprolyl Esters

Zanotti *et al.* initially observed that a *para*-nitrophenyl (Np) ester, $PhCH_2CO$-Cys(S*t*Bu)-Pro-ONp, when reacted with tributylphosphine produced a diketopiperazine thioester, *cyclo*(-Cys($COCH_2Ph$)-Pro-).[246] Kawakami and Aimoto then adopted this approach to generate a peptide thioester *in situ* by functionalising a peptide with a C-terminal “autoactivating” cysteinylprolyl ester (CPE) (Scheme 1.25).[247-249]

Scheme 1.25 Kawakami and Aimoto introduced the cysteinylprolyl ester (CPE) method for synthesising peptide thioesters *in situ* for use in NCL reactions.[247] An “autoactivating” *N→S* acyl shift occurs initially, which is then followed by diketopiperazine formation. Then, transthioesterification occurs with an added thiol to afford a thioester, which then participates in an NCL reaction to afford a new amide bond. *Reagents and conditions*: 0.1 M buffer, pH 8.2, 6 M guanidine.HCl, 20 mM *tris*-(hydroxypropyl)phosphine (THP), N-terminal cysteinyl peptide, r.t., 24 h.[247,248]

Mechanistically, Kawakami and Aimoto proposed that an intramolecular *N→S* acyl transfer occurs to form the *S*-peptide, which then promotes diketopiperazine formation.[247-249] The resulting diketopiperazine peptide can then participate in intermolecular thiol-mediated thioester exchange, which can then be followed by ligation with, for example, a cysteinyl peptide under standard NCL conditions to afford a full-length polypeptide and the liberation of *cyclo*(-Cys-Pro-) (Scheme 1.25).[247]

The proline residue within the CPE unit may be exchanged with sarcosine (*N*-methylglycine); however the ester is important as a C-terminal cysteinyl-prolyl-carboxylic acid or a cysteinyl-prolyl-carboxamide has been shown to not ligate under the same conditions.[247] The diketopiperazine peptide has not been isolated but the by-product *cyclo*(-Cys-Pro-) has been observed by mass spectrometry.[247] Glycolic acid is also typically incorporated to form the 'ester' component of the CPE.[247]

This CPE method circumvents the need for a peptide thioester building block and uses Fmoc-SPPS to potentially offer a higher yielding starting material.[247-249] This is because it relies instead upon the entrapment of the *S*-peptide generated at equilibrium by a thiol.[247] The reaction is reported to occur at room temperature and between pH 7.0 and 8.0, under denaturing conditions,[247] which is ideal for peptides that contain heat- or acid-sensitive linkages and are prone to aggregation. Whilst a pH lower than 7.0 would be expected to favour the initial *N→S* acyl transfer, the conditions adopted are reasoned to favour thioester formation at the Xaa-Cys site because the amino group released is then free to form the diketopiperazine in slightly basic solution. The *N→S* acyl transfer reaction rate in the cysteine-containing peptides appeared to be slow when analysed by (RP)HPLC in weakly acidic solution (aqueous acetonitrile containing 0.1 % TFA), but even if the *S*-peptide was produced it rapidly reverted back to the *N*-peptide upon attemped isolation.[247]

Ligation yields recorded when the C-terminal amino acid was varied in a model peptide-CPE from glycine to alanine, leucine, or valine, were similarly good (49-60 %), and epimerisation during ligation was found to be negligible, though CPE-mediated ligation was also accompanied by a small amount of peptide hydrolysis.[247]

A key advantage of the CPE device over the *N*-alkyl cysteine device is that it utilises standard, readily available materials that are typically associated with Fmoc/*t*Bu chemistry. A

minor synthetic inconvenience is that the Xaa-Cys motif (within the C-terminal Xaa-Cys-Pro-glycolyl sequence) needs to be introduced as a pre-prepared dipeptide in order to prevent premature on-resin diketopiperazine formation during peptide chain assembly.[247] The protecting groups for any free α-amino groups or side-chain lysine ε-amino groups are also usually retained throughout,[247] which is most likely to prevent unwanted intramolecular cyclisation or intermolecular reactions with other thioester intermediates. The removal of such protecting groups therefore involves an additional step after ligation.

Nonetheless, this CPE method does appear to overcome many disadvantages that are typically associated with thioester synthesis, because the thioester is only produced transiently under the reaction conditions at room temperature and at approximately neutral pH. After peptide synthesis, the peptide-CPEs are cleaved from the resin with a C-terminal carboxamide group,[247] which are typically high yielding.

Peptide-CPEs can also be used to directly generate thioesters rather than allowing them to participate in ligation. This has been achieved by exposure to a thiol (i.e. MESNa) at pH less than 8.0, to minimise hydrolysis, with an elevated temperature of 37 °C for 6 h, to give the peptide thioester in good yields.[248,249] In general, the degree of racemisation has been assessed to be low except when serine thioesters were synthesised from peptide-CPEs, in which case it was found to be greater than 20 %.[248,249]

More recently, Aimoto *et al.* have also demonstrated the transformation of a peptide containing a Cys-Pro-Cys (CPC) sequence into a peptide diketopiperazine thioester through an *N*→*S* acyl shift under acidic conditions.[250] The authors propose that since a Cys-Pro-Cys sequence can be readily encoded within proteins by recombinant DNA technology, it offers an alternative means to produce protein thioesters.[250]

Kawakami *et al.* recently disclosed a fascinating extension of the CPE method to provide a recombinant route to cyclic peptides called "backbone-cyclic peptide synthesis using recombinant elements" (bcPURE).[251] It involved the use of the CPE method in conjunction with codon reprogramming through *in vitro* transcription followed by translation to allow recombinant expression, which was then followed by non-enzymatic cyclic peptide synthesis.[251] Peptide cyclisation was promoted by the familiar formation of a *cyclo*(-Cys-Pro)- diketopiperazine thioester. Several target cyclic peptides were

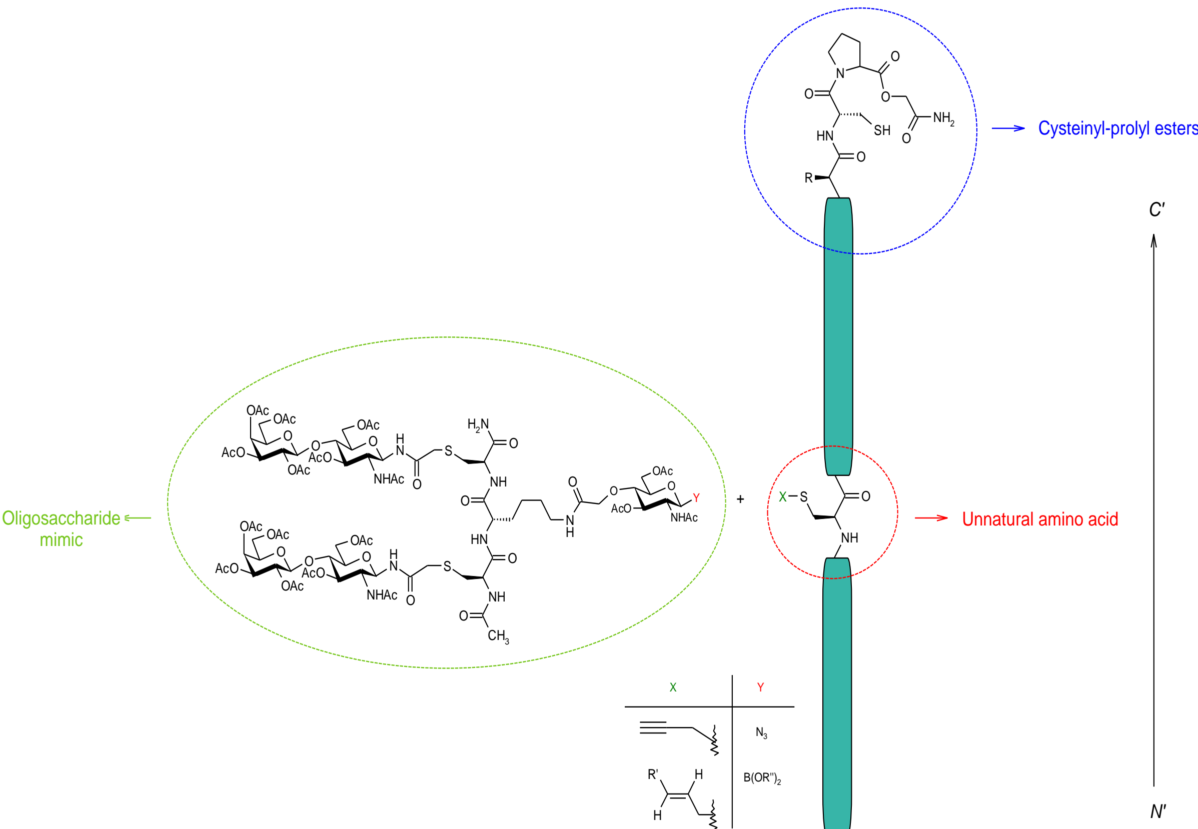
prepared including Eptidemnamide, Scleramide, Rhesus-θ-defensin-1, and Sunflower trypsin inhibitor.[251]

1.20 Conclusion

Whilst many methods exist for the preparation of thioesters, they often result in low or variable isolated yields of the thioesters and tend to require a specialised linker, resin, device, or a cleavage cocktail.[199,200,208,210,214,252-256] Recently, several laboratories have reported the synthesis of peptide thioesters based on an initial *N→S* acyl transfer, resembling the initial stages of protein splicing.[215,234-241,243-245,247-250] These approaches provide an alternative source of peptide thioester that may in the future prove a more viable route than conventional carboxylic acid activation chemistries.

1.21 Scope of Thesis

The initial aim of this project was to explore the structure-activity relationships of the glycoprotein hormone erythropoietin (EPO). This was to be achieved by synthesising analogues of EPO and subsequently studying the effect of the introduced amino acid modifications on protein function. EPO was to be semi-synthesised through an NCL reaction between a synthetic peptide thioester and a biologically-derived fragment. Consequently, the synthetic peptide thioester would allow us to incorporate unnatural, glycosylated amino acids (Scheme 1.26).



Scheme 1.26 The synthetic peptide thioester corresponds to the N-terminus of erythropoietin (EPO). It contains three points of variation: a C-terminal cysteinylprolyl ester (CPE) unit for *in situ* thioester formation, an oligosaccharide mimic attached to the peptide backbone *via* an unnatural linkage, and an unnatural amino acid.

- **Chapter 2** describes the unexpected discovery of a novel peptide and protein thioesterification method which was observed whilst experimenting with the cysteinylprolyl ester method for peptide thioester synthesis during our studies towards the semisynthesis of EPO. 3-Mercaptopropionic acid (MPA) was found to selectively fragment histidine-cysteine, cysteine-cysteine, and glycine-cysteine sequences to afford the corresponding MPA thioesters through an initial *N→S* acyl transfer. Selective thioesterification was demonstrated on several short EPO-based peptides and recombinant human EPO.[263]

- **Chapter 3** focuses on optimising the new thioesterification reaction discovered by introducing a site-specific ^{13}C isotopic label within a model peptide. The labelled peptide was then reacted under various reaction conditions and the resulting conversion into peptide thioesters was measured. The optimised thioesterification conditions were then used in the first application of a biologically active protein, a 45 amino acid residue analogue of human β-defensin 3.[267]

- **Chapter 4** demonstrates the compatibility of the new thioesterification procedure with the first synthesis of protected and deprotected native *N*-linked glycopeptide thioesters for potential use in the semisynthesis of glycoproteins.

- **Chapter 5** provides a summary of the main findings of the work presented in this thesis.[296]

2 Discovery of a Novel Peptide and Protein Thioesterification Reaction: Native Chemical Thioesterification

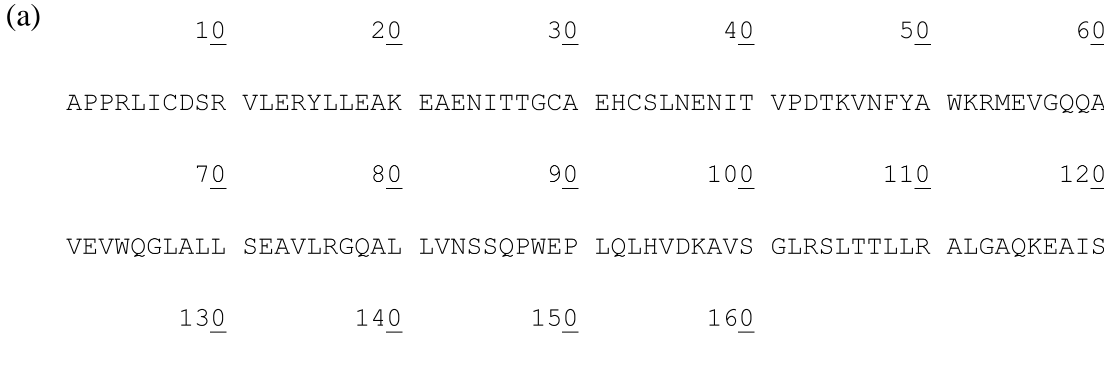
2.1 Introduction

Native Chemical Ligation is an extremely powerful method that is routinely used for the synthesis of small to moderate-sized proteins.[144,145,187,257] However, since approximately fifty amino acids represents the maximum for an efficient solid phase peptide synthesis, recombinant means to attain the required thioester and cysteine-containing components have been explored.[152,156,158,258]

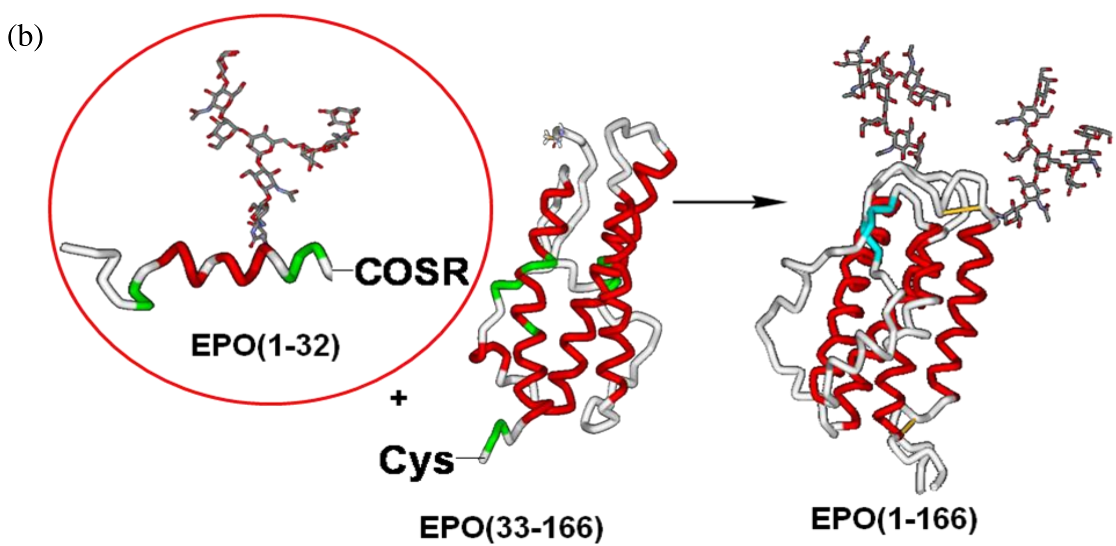
Larger polypeptides can be accessed through semisynthesis, which involves an NCL reaction between a biologically-derived fragment and a synthetic fragment. Our main interest was in using organic chemistry to try and understand biological processes and therefore, we geared our studies towards the semisynthesis of the glycoprotein hormone erythropoietin (EPO). Erythropoietin is comprised of 166 amino acid residues and its role in the regulation of erythrocyte production has led to its common use in the treatment of anaemia, but it has also gained notoriety for being the illegal doping agent used in many competitive sports.[259-262]

Our semisynthetic strategy towards erythropoietin involved the chemical synthesis of a glycosylated peptide thioester that corresponded to the N-terminus of erythropoietin: EPO(1-32)-COSR, which would then be ligated to a biologically expressed cysteinyl fragment that corresponded to the C-terminus of erythropoietin: EPO(33-166) (Scheme 2.1).

(a)



(b)



Scheme 2.1 (a) Amino acid sequence of human erythropoietin EPO(1-166).[259-262] Disulfide bonds exists between ^{7}Cys-^{161}Cys and ^{29}Cys-^{33}Cys; ^{24}Asn is *N*-linked to Glc*N*Ac; ^{38}Asn is *N*-linked to Glc*N*Ac; ^{83}Asn is *N*-linked to Glc*N*Ac; and ^{126}Ser is *O*-linked to Gal*N*Ac.[259-262] (b) The semisynthetic strategy towards EPO involved an NCL reaction between a synthetic N-terminal EPO(1-32) peptide thioester and a biologically-derived C-terminal cysteinyl EPO(33-166) fragment.

The number of methods available to synthesise the required thioester has risen dramatically, however these methods often result in low or variable isolated yields of the thioester and tend to require a specialised linker, resin, device, or a cleavage cocktail.[199,200,208,210,214,252-256] These requirements are disadvantageous because, as well as being synthetically inconvenient, they

often limit the amino acid sequence of the peptide thioester; limit the synthesis of posttranslationally modified peptide thioesters; and increase the number of synthetic steps from the commercially available starting materials to the product, which inevitably impacts on the overall yield and efficiency of the route. In general, the synthesis of peptide thioesters is notoriously difficult, particularly in larger quantities for use in industrial scale NCL reactions for the total synthesis of potentially therapeutic proteins. Therefore a more efficient and scalable route to produce (ideally posttranslationally modified) peptide thioesters using Fmoc-based chemistry was very much required.

Recently, several variations on the synthesis of thioesters *via* an *N→S* acyl shift that mechanistically resemble the initial stages of protein splicing using inteins have been reported (Scheme 2.2).[215,234-241,243-245,247-250]

Scheme 2.2 (a) The most extensively used method for protein synthesis: Native Chemical Ligation,[144] compared to (b) an acid-mediated *N→S* acyl transfer at cysteine residues used for thioester formation. R_1 denotes an acyl-transfer facilitating group such as: alkyl, benzyl, δ-mercaptomethyl prolyl.[215,234-241,243-245,247-250]

We began by looking at Kawakami and Aimoto's cysteinylprolyl ester (CPE) methodology.[247,248] Their reaction involved a peptide functionalised with a C-terminal CPE undergoing an initial "autoactivating" *N→S* acyl transfer to transiently produce thioesters for use in NCL reactions (Scheme 1.25). The term '*N→S* acyl shift' is used to describe an acyl

rearrangement from an amide bond (-CONH-) or '*N*-peptide', into a thioester (-COS-) or '*S*-peptide'. Mechanistically, CPE ligation proceeds *via* an initial reversible *N→S* acyl shift to produce the *S*-peptide. This is followed by diketopiperazine formation which is irreversible and hence drives the equilibrium to the right. Overall, the process is driven by the formation of the thermodynamically more stable amide bond after NCL.

We were intrigued by the CPE methodology and so we adopted it in our studies geared towards the semisynthesis of EPO. Peptide-CPEs had previously not been used for the semisynthesis of a protein and also seemed ideal because a peptide carboxamide is initially prepared (which is high yielding) through Fmoc-SPPS (which is preferable for appending acid-sensitive posttranslational modifications such as glycosylation), and the entire CPE mediated ligation reaction is conducted at approximately neutral pH (7.0-8.0).

2.2 Results and Discussion

2.2.1 Synthesis of a Model Peptide-CPE with Two Unprotected Cysteine Residues

Our ultimate aim was to prepare the N-terminal EPO fragment (amino acid sequence: 1-32) with a C-terminal CPE to facilitate semisynthesis (Scheme 2.1), and for this purpose we initially conducted some model CPE ligation experiments. We first synthesised a model peptide based on a short sequence of erythropoietin with a C-terminal cysteinylprolyl ester: EPO(22-32)-CPE, which encoded the following amino acid sequence: H-AENITTGCAEA-CPE. Notably, this 13 amino acid sequence contained two cysteine residues, one of which was required for the CPE, and the other was a native cysteine present within the full-length EPO(1-166) at position-29. The native cysteine at position-29 is involved in disulfide bonding within the folded full-length protein.[259-261] The Fmoc-based solid phase peptide synthesis strategy employed to prepare the model peptide-CPE **11** is shown in Scheme 2.3.

Scheme 2.3 Synthesis of the model peptide-cysteinylprolyl ester EPO(22-32)-CPE with the sequence: H-AENITTGCAEA-CPE **11**, for use in test CPE ligations with the ultimate aim of EPO(1-166) semisynthesis.[263]

The peptide-CPE was assembled using Fmoc-based SPPS on Rink Amide MBHA resin **7**, to produce the C-terminal carboxamide after peptide chain assembly. The synthesis was carried out on a 0.1 mmol scale. Once the Fmoc protecting group had been removed from the amino group of the resin with 20 % v/v piperidine in *N,N*-dimethylformamide (DMF), the "ester" component of the CPE was subsequently incorporated. This was achieved by initially forming the active ester of glycolic acid (Gc) using DIC and HOBt in DMF,[247,248] and this reaction mixture was then reacted with the resin to give **8** (Scheme 2.4).

The prolinyl component of the CPE was formed by coupling Fmoc-Pro-OH to **8**, using HBTU/HOBt as the coupling reagents, to give Fmoc-Pro-Gc-resin. Proline is an imino acid and therefore the removal of the N-α-Fmoc group from this secondary amino group was difficult to monitor by the standard Kaiser Ninhydrin colour-test.[115] The resin beads appeared red-orange in colour both before and after piperidine treatment, as opposed to the characteristic yellow (protected amine) and indigo (free amine).

Scheme 2.4 Mechanism of the standard coupling reaction between an HOBt-activated carboxylic acid ester and an amine to form a new amide bond in **8** and a urea by-product.[106]

Therefore, in order to quantify the amount of the first amino acid, proline, that had been successfully loaded onto the resin, a quick UV absorbance assay was conducted.[115] This involved the addition of a 20 % v/v piperidine in DMF solution to a few milligrams (1.5–3.0 mg) of Fmoc-Pro-Gc-resin, followed by measurement of the absorbance of the solution directly above the beads inside a quartz cuvette. Fmoc deprotection with piperidine is rapid and thus the loading of the first amino acid onto the resin was determined by Fmoc analysis to be 85 % (0.544 $mmolg^{-1}$/0.64 $mmolg^{-1}$) (Equation 1,[115] Table 2.1). This value was used after peptide chain assembly was complete, to calculate the final isolated yield of the peptide-CPE.

$$\text{Resin loading (mmolg}^{-1}\text{)} = \frac{Abs_{290\ nm}}{(1.75)\bullet(Mass_{resin}/mg)} \qquad (1)$$

Table 2.1 Absorbance data for the determination of the loading of the first amino acid, Fmoc-Pro-OH, onto the resin by Fmoc analysis using Equation 1.[115]

Sample	Mass /mg	$Abs_{290\ nm}$	Loading/ $mmolg^{-1}$
1	2.00	1.9720	0.563
2	2.10	1.9290	0.525
Ref	-	0.0000	-
Mean	2.05	1.9505	0.544

The N-α-Fmoc group was then removed from proline with 20 % v/v piperidine in DMF to afford **9**. In order to prevent premature on-resin diketopiperazine formation as a result of the sequential coupling of Fmoc-Cys(Trt)-OH to **9**, the cysteine residue of the growing peptide-CPE was incorporated as part of a pre-synthesised dipeptide **6**.

On-resin diketopiperazine formation could occur upon removal of the Fmoc group from N-α of cysteine (Scheme 2.5). The free α-amino group of cysteine can then cyclise onto the carbonyl group between the prolyl-glycolyl sequence to afford the undesired diketopiperazine. If the synthesis was to be continued in this way then the product would contain a C-terminal cysteine-proline deletion. Therefore, to circumvent the possibility of synthesising a peptide without the essential cysteine-proline sequence within the peptide-CPE, cysteine-containing dipeptides were synthesised.

Scheme 2.5 An undesirable side reaction: premature on-resin diketopiperazine formation resulting in a peptide containing a C-terminal cysteine-proline deleted sequence. This can be overcome by incorporating Fmoc-Xaa-Cys(Trt)-OH as a dipeptide during solid-phase synthesis, rather than sequential Fmoc amino acids.

Kawakami and Aimoto have demonstrated the scope of the CPE method with various C-terminal amino acid residues.[247,248] However, we chose to prepare a peptide-CPE that contained one of the least hindered amino acids (glycine or alanine) at its C-terminus. The dipeptide Fmoc-Xaa-Cys(Trt)-OH where Xaa was either glycine or alanine, was initially synthesised for incorporation into the growing peptide-CPE chain (Scheme 2.6). This was

achieved through triphenylmethyl (Trt) protection of the side-chain sulfhydryl group of cysteine. L-Cysteine hydrochloride, **1**, was reacted with triphenylmethyl chloride to afford the side-chain protected amino acid **2** in 70 % yield (Scheme 2.6). Trityl-protected cysteine **2** was then coupled to the pre-activated α-carboxylic acid group of **3** to give **4** in 18 % yield, or coupled with **5** to give **6** in 78 % yield. Therefore, peptide-CPE synthesis was continued with **6**.

Scheme 2.6 Synthesis of the cysteine-containing Fmoc-Xaa-Cys(Trt)-OH dipeptides to avoid premature diketopiperazine formation on-resin during peptide-CPE solid phase synthesis. Xaa represents a sterically less hindered amino acid: either glycine **4** or alanine **6**.

The cysteine-containing dipeptide **6** was then coupled to the proline-linked resin **9** using HBTU/HOBt, to give the resin-bound tripeptide **10**. The peptide chain was then elongated by automated Fmoc-SPPS, with double-coupling of isoleucine, to give the protected resin-bound peptide: AE(O^{t}Bu)N(Trt)IT(^{t}Bu)T(^{t}Bu)GC(Trt)AE(O^{t}Bu)AC(Trt)P-Gc-resin. This resin-bound peptide was then deprotected and concomitantly cleaved from the resin by treatment with the standard cleavage cocktail comprised of 95 % trifluoroacetic acid, 2.5 % ethanedithiol and 2.5 % water (v/v/v),[115] to afford the crude peptide-CPE with a C-terminal carboxamide. During the cleavage reaction ethanedithiol acts as a scavenger for the cleaved by-products of acid-labile protecting groups (Scheme 2.7).

Scheme 2.7 An example of the cleavage mechanism of the *tert*-butyl side-chain protecting group of glutamic acid using a cleavage cocktail comprised of 95 % trifluoroacetic acid, 2.5 % ethanedithiol, and 2.5 % water (v/v/v).[106] Ethanedithiol acts as a scavenger of any electrophiles generated.

The 13 amino acid deprotected peptide-CPE corresponding to the sequence EPO(22-32)-CPE: H-AENITTGCAEA-CPE was then purified by (RP)HPLC and rather interestingly, but not surprisingly, the peptide-CPE eluted as an asymmetric 'doublet' at 23.0 min (Figure 2.1). The doublet peptide fractions were found to have an identical mass by ESI$^+$ LC-MS, which corresponded to the expected mass of the desired model peptide EPO(22-32)-CPE **11**. Therefore, the observed doublet was most likely due to the presence of diastereoisomers in the peptide; one possibility is cysteine diastereoisomers which could have formed upon coupling **6** to **9** (Scheme 2.8); another possibility is proline diastereoisomers. However, because the CPE unit is excised during ligation, the diastereoisomers were not separated.

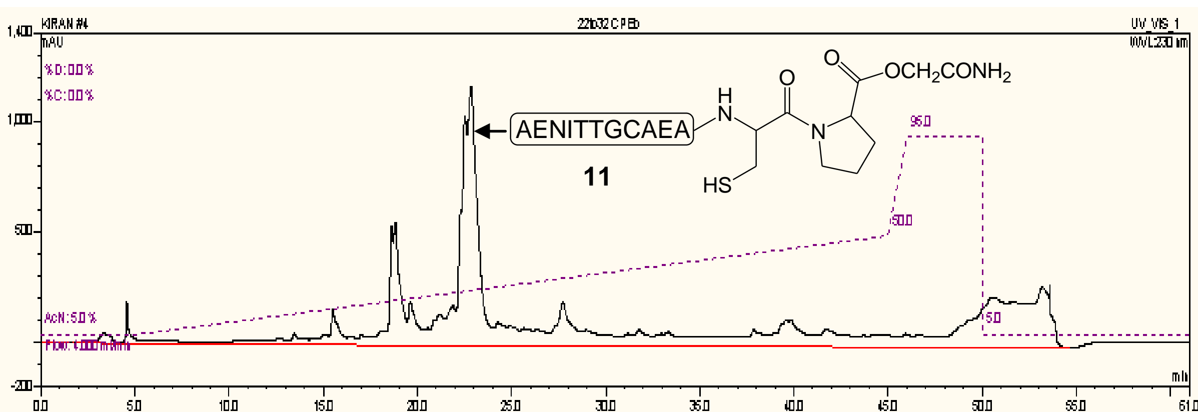


Figure 2.1 (RP)HPLC purification chromatogram of the proposed diastereoisomer-containing peptide-CPE **11**, t_R = 23.0 min. ESI$^+$ MS *m/z* calculated for EPO(22-32)-CPE [M]: 1335.6 Da; found [MH]$^+$: 1336.7 Da.

Scheme 2.8 Mechanism of possible cysteine diastereoisomerisation during peptide-CPE synthesis. The activation of the α-carboxyl group of the dipeptide Fmoc-Ala-Cys(Trt)-OR during its coupling to H-Pro-Gc-resin under basic conditions can cause racemisation of the cysteine residue, and both diastereoisomers can continue in Fmoc-SPPS.

Diastereoisomerisation of the C-terminal cysteine may have occurred as a consequence of pre-activation of the cysteinyl α-carboxylic acid to the HOBt-ester under basic reaction conditions (Scheme 2.8).[106] The lone pair of electrons on the cysteinyl α-amino group can participate in oxazolone formation. Nucleophilic attack of the carbonyl oxygen onto the activated ester can form a five-membered oxazolone ring. Base-mediated deprotonation of the α-hydrogen atom, followed by reprotonation of the α-carbon from the opposite face can cause diastereoisomer formation. The oxazolone ring can then be re-opened by nucleophilic attack of the secondary amino group of proline, to afford the dipeptide as a mixture of diastereoisomers. Hence, the peptide-CPE product may contain cysteine diastereoisomers.

The model peptide-CPE **11** was isolated in 22 % overall yield (Figure 2.2) and it was then reacted in model CPE ligations.

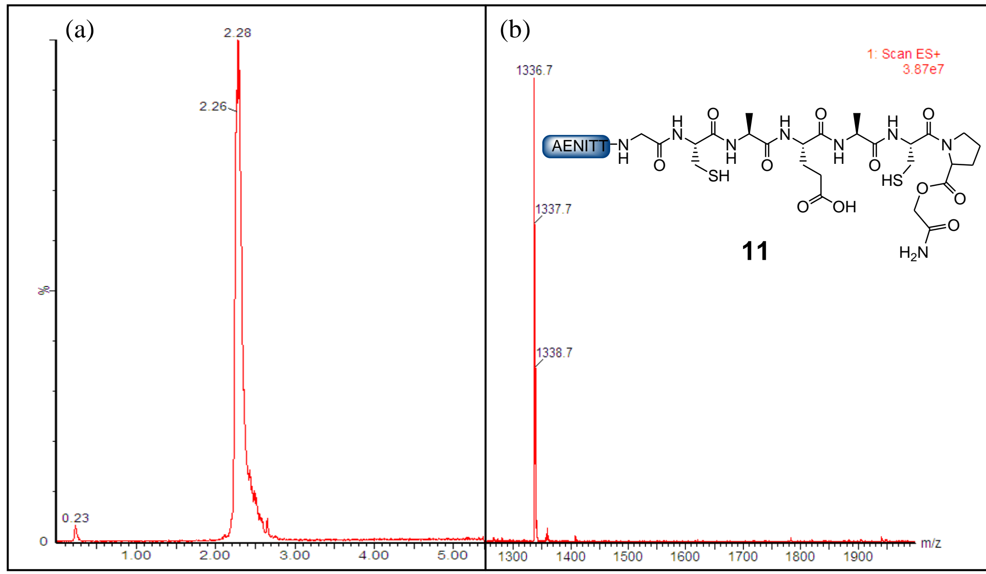


Figure 2.2 Characterisation of the synthesised, isolated model peptide-CPE **11** (sequence = AENITTGCAEA-CPE) for subsequent CPE-mediated ligation studies. (a) LC-MS UV absorbance trace of EPO(22-32)-CPE **11**, absorbance was detected at 254 nm. (b) ESI^+ MS *m/z* calculated for EPO(22-32)-CPE [M]: 1335.6 Da, observed $[MH]^+$: 1336.7 Da.[263]

2.2.2 Attempted CPE ligation

Model CPE ligations were conducted between **11** and an N-terminal cysteinyl peptide that corresponded to a short sequence of erythropoietin EPO(33-40): H-CSLNENIT-OH, **50** (Scheme 2.9). A new Ala-Cys amide bond was expected in the full-length ligated peptide EPO(22-40): H-AENITTGCAE<u>AC</u>SLNENIT-OH, **12**.

Scheme 2.9 Attempted CPE ligation between the model peptide-CPE: EPO(22-32)-CPE **11** and the N-terminal cysteinyl peptide: EPO(33-40) **50** to give the ligated product: EPO(22-40) **12**, under typical NCL conditions: 6 M guanidine hydrochloride, 0.3 M sodium phosphate buffer; pH 8.0, 2 % MESNa, and 10 mM TCEP.[263] The reaction was monitored by LC-MS between t = 1 and 122 h, however the ligated product **12** was not observed.[263]

Three different CPE ligation conditions were devised in which the thiol additive and pH were varied. The thiol additive is known to influence the ligation rate[146] and was chosen to be either the alkyl thiol sodium 2-mercaptoethanesulfonate (MESNa), or the aryl thiol 4-mercaptophenyl acetic acid (MPAA).[184] The ligation was conducted at either pH 7.0 or 8.0. The higher pH facilitates deprotonation of the cysteine sulfhydryl within the CPE, and therefore promotes the *N→S* acyl shift and increases the rate of the diketopiperazine formation (Scheme 1.25).[247,248] Conversely, a slightly lower pH also promotes the initial *N→S* acyl shift because it increases the amount of the protonated amine present at equilibrium. A lower pH would however mean that the second diketopiperazine formation step would be expected to proceed slower than at pH 8.0.

Analytical scale ligation reactions were set up between **11** and **50** at 25 °C in aqueous buffer. The reaction conditions for three parallel experiments were as follows: (A) 50 mM MPAA, 10 mM TCEP, pH 7.0; (B) 50 mM MESNa, 10 mM TCEP, pH 7.0; or (C) 50 mM MPAA, 10 mM TCEP, pH 8.0. The reactions were monitored by LC-MS at regular intervals from t = 1 to 122 h. It was expected that the rate of ligation would be fastest when the thiol was MPAA and the pH 8.0.

After 18 h, an aliquot was removed from each of the three reaction mixtures for LC-MS analysis. The LC-MS traces for all three experiments looked noticeably different after 18 h than those after 1 h. Reaction vessel 'A' contained 50 mM MPAA at pH 7.0 and after 18 h it showed the presence of two major species with longer retention times than the starting material model peptides. However, neither of these peaks corresponded to the ligated product EPO(22-40) **12**. The earlier eluting of these peaks had an observed mass which corresponded to the calculated value for the cysteinyl peptide disulfide bonded to MPAA (Figure 2.3). It was possible that the reducing agent *tris*(2-carboxyethyl)phosphine (TCEP) had been rapidly consumed since the reaction was not conducted under strictly inert conditions and both starting peptides contained a total of three cysteine residues. Hence more TCEP was added to give a final concentration of 40 mM.

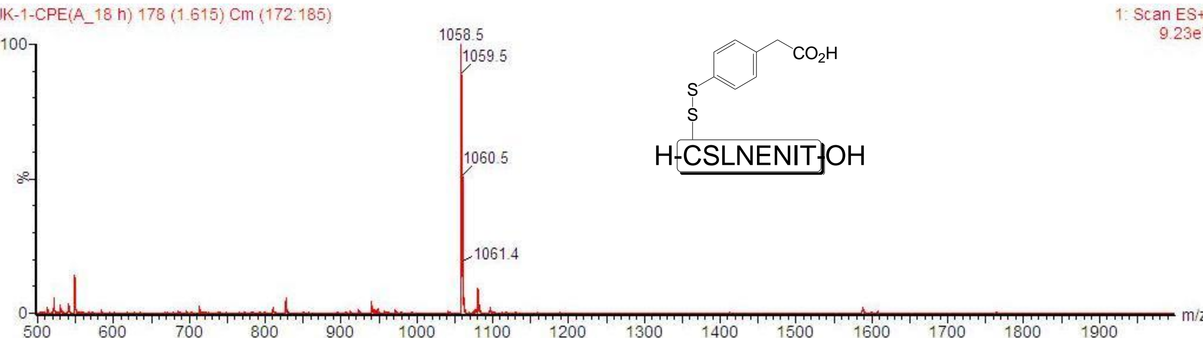


Figure 2.3 ESI^{+} mass spectrum of the main component of the test CPE ligation between **11** and **50**, analysed after 18 h; the reaction contained 50 mM MPAA, 10 mM TCEP; pH 7.0. Calculated mass for the cysteinyl-peptide EPO(33-40) **50** being disulfide-bonded to MPAA: 1058.4; observed $[MH]^{+}$: 1058.5.

The proposed mechanism for the reduction of disulfide bonds by TCEP is shown in Scheme 2.10. Initially, the nucleophilic phosphorus(III) attacks the more electropositive sulfur atom of the disulfide and releases a free sulfide group. Nucleophilic attack from an oxygen atom of water then occurs onto the phosphonium ion to release the other free sulfide. The formation of the strong phosphonium oxide P=O double bond in the co-product drives the reaction to afford the two free sulfhydryl groups.

Scheme 2.10 The proposed mechanism of reduction of disulfide bonds using *tris*(2-carboxyethyl)phosphine (TCEP) to afford two free sulfhydryl groups. The reaction is proposed to be driven by the formation of the strong phosphorus-oxygen double bond.

Therefore, CPE ligation to give **12** was unsuccessful when 50 mM MPAA and 10 mM TCEP (pH 7.0) were used. This was likely due to the formation of intermolecular disulfide bonds between MPAA molecules and the three cysteine residues present within **11** and **50**. If the sulfhydryl group of the C-terminal cysteine residue within peptide-CPE **11** was oxidised, it would therefore be unable to undergo an initial *N→S* acyl shift. Similarly, if N-terminal cysteinyl peptide **50** was oxidised it would also be unable to undergo transthioesterification with any existing peptide thioester.

Reaction vessel 'B' contained 50 mM MESNa at pH 7.0 and after 18 h it showed the presence of unreacted **11** and **50** by LC-MS analysis (Figure 2.4). The peptide-CPE **11** was observed to have a slightly longer retention time of 1.32 min than the shorter cysteinyl peptide **50** (1.11 min), using a 5–95 % acetonitrile gradient over 8 min during LC-MS analysis. It appeared that at pH 7.0, MESNa was less prone to oxidation than MPAA. After 18 h both starting material peptides were still available for CPE-mediated ligation but no ligated product was observed. Therefore, for reasons unknown, CPE ligation to give **12** was also unsuccessful when 50 mM MESNa and 10 mM TCEP (pH 7.0) were employed.

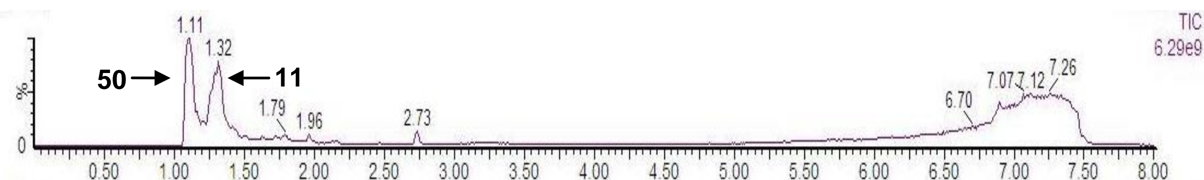


Figure 2.4 The LC-MS (TIC) chromatogram of the test CPE ligation reaction between EPO(22-32)-CPE **11** and EPO(33-40) **50** in vessel 'B', which contained 50 mM MESNa; pH 7.0, and was analysed after 18 h. No ligated product **12** was observed, only unreacted starting peptide-CPE **11** and N-terminal cysteinyl peptide **50** were observed.

After 18 h, reaction vessel 'C' which contained 50 mM MPAA at pH 8.0 was also shown by LC-MS analysis to be mainly comprised of **50** being disulfide bonded to MPAA. Oxidation of **11** was also observed, therefore the amount of the reducing agent TCEP was increased to 40 mM to allow ligation to prevail. However, CPE ligation to give **12** was unsuccessful when 50 mM MPAA and 10 mM TCEP (pH 8.0) were used.

Overall, the ligated product **12** was not observed by LC-MS analysis under any of the three reaction conditions studied, even after 122 h, which is an uncommonly long duration of time for a ligation reaction because of potential side-reactions that can occur, including thioester hydrolysis. Both of the starting materials remained intact after 122 h in the reactions that contained MPAA or MESNa at pH 7.0. However peptide-CPE **11** was found to degrade after 122 h in the reaction that contained 50 mM MPAA at pH 8.0.
A plausible explanation is that ligation at the alanine-cysteine junction is slow. The presence of multiple cysteine residues within both of the starting materials also likely hindered the ligation due to undesirable oxidation, particularly in reactions containing MPAA. It was also proposed that the presence of an additional internal cysteine and its position within the sequence of **11** may have inhibited the reaction.

2.2.3 MPA-Mediated Peptide Thioesterification

The CPE ligation reactions were unsuccessful possibly due to the two cysteine residues present within **11**, so we turned our attention to "rescuing" **11** by using the same forcing conditions that were employed by Nakahara *et al*.[239]

Recently, Nakahara and co-workers reported the thioesterification of peptides that were furnished with a C-terminal 5-mercaptomethyl prolyl-prolyl ester, upon treatment with 3-mercaptopropionic acid (MPA) (Scheme 1.22).[239] They had originally proposed a route *via* diketopiperazine formation at the C-terminal prolyl-prolyl-ester sequence, however this was not observed.[239] Instead, when their peptide carrying the C-terminal 5-mercaptomethyl prolyl-prolyl-ester moiety was treated with 40 % aqueous MPA and heated at 80 °C using microwave irradiation for 48 h, they observed the formation of the peptide-$SCH_2CH_2CO_2H$ MPA thioester.[239] Therefore, we decided to treat our model peptide-CPE **11**, which had two free cysteine residues, with the same forcing reaction conditions that Nakahara *et al.* employed (Scheme 2.11).

30 % MPA, microwave irradiation, 80 °C, 16 h, 84 %

Scheme 2.11 Unexpectedly, when the model peptide-CPE **11** was treated with 30 % v/v 3-mercaptopropionic acid (MPA) using microwave irradiation, it selectively fragmented between the internal glycine-cysteine sequence rather than at the expected C-terminal alanine-cysteine sequence, to afford the glycine-MPA thioester **13**.[263]

Therefore, **11** was treated with 40 % v/v aqueous MPA at an elevated temperature of 80 °C under microwave irradiation, at an apparent pH of 1.0, and the reaction progress was monitored by LC-MS analysis. We expected **11** to undergo thioesterification at the C-terminal alanine-cysteine-proline-ester sequence, analogous to the observation that had been made with Nakahara *et al.*'s 5-mercaptomethyl prolyl-prolyl-ester upon treatment with MPA.[239]

After four hours of treatment of **11** with MPA, we found that the main component of the reaction mixture was unreacted **11**. However, we also detected fragmentation of **11** between the internal ^{28}Gly-^{29}Cys sequence to afford the glycine MPA thioester **13**: H-AENITTG-MPA. Surprisingly, thioesterification had occurred selectively at the internal glycine-cysteine sequence rather than at the C-terminal alanine-cysteine sequence in **11** to give the MPA thioester **13**.[263]

The thioesterification of the model peptide-CPE **11** with 40 % v/v aqueous MPA after 12 h showed three main peaks by LC-MS analysis in an approximately 1:1:1 ratio (Figure 2.5). The earliest eluting peak was characterised to be the hydrolysed peptide thioester H-AENITTG-OH, which was followed by the glycine-MPA thioester **13**, and finally the co-eluting unreacted starting material **11** and the full-length alanine-MPA thioester H-AENITTGCAEA-MPA. Peptide thioester hydrolysis is an undesirable side-reaction that decreased the overall yield of the desired peptide thioester.

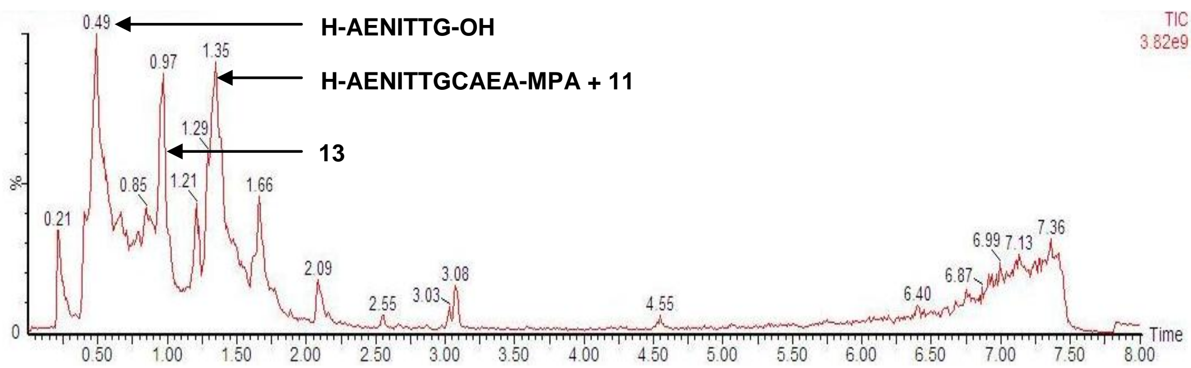


Figure 2.5 LC-MS (TIC) chromatogram of the reaction of EPO(22-32)-CPE **11** with 40 % v/v aqueous MPA with microwave heating at 80 °C. LC-MS analysis showed the presence of the hydrolysed thioester H-AENITTG-OH with an observed mass of 705.5 Da, the glycine-MPA thioester H-AENITTG-$SCH_2CH_2CO_2H$ **13** with an observed mass of 793.4 Da, the alanine-MPA thioester H-AENITTGCAEA-$SCH_2CH_2CO_2H$ with an observed mass of 1167.6 Da, and the unreacted peptide-CPE H-AENITTGCAEA-CPE **11** with an observed mass of 1336.8.

After elongating the reaction time to 24 h, LC-MS analysis revealed the main component of the reaction mixture to be the undesirable hydrolysed thioester. Therefore, to prevent thioester hydrolysis becoming the main product, thioesterification under these forcing reaction conditions should be terminated before 12 h.

This new thioesterification reaction was discovered using an excess of MPA (40 % v/v). However, MPA was found to have several strong absorbance peaks in UV chromatograms during HPLC purification, which were attributed to both the free MPA thiol and its disulfide-bonded dimer (HO_2C-CH_2-CH_2-S-S-CH_2-CH_2-CO_2H), and these peaks often precluded product thioester isolation.

The accumulation of a white precipitate within the reaction mixture was also observed over time, and this was isolated and characterised by NMR spectroscopy to be oxidised MPA. The precipitate produced was unproductive because it decreased the amount of the free MPA thiol available to participate in the reaction. The use of lower MPA concentrations was then explored to convert **11** into thioester **13**.

Hence, thioesterification of **11** was tested with the lower, though still in excess, concentration of 30 % v/v aqueous MPA and the reaction mixture was heated at 80 °C with microwave irradiation for 16 h. The glycine-MPA thioester **13** was isolated in 84 % yield (0.5 mg).[263]

Overall, the glycine-MPA thioester **13** appeared to be a kinetically favoured product. The expected C-terminal alanine-MPA thioester was also detected within the reaction mixture though only in trace amounts by LC-MS. It appeared that over time the alanine-MPA thioester was further "processed" into the glycine-MPA thioester **13**.

Under these reaction conditions **13** was found to be undesirably hydrolysed over time. We also observed undesirable disulfide bond formation between the cysteine residues in **11** and MPA by LC-MS. Therefore the reaction required optimisation, but it was the first time we unexpectedly observed thioester formation selectively at a glycine-cysteine sequence rather than at an alanine-cysteine sequence. This suggested that the CPE device,[247,248] or any other device such as 5-mercaptomethyl proline[239] or *N*-alkyl cysteine,[240,241] was not required to promote thioesterification, because the internal "device-free" glycine-cysteine sequence in **11** was thioesterified. The preference of MPA-mediated thioesterification taking place at the internal glycine-cysteine site can be explained in terms of steric hindrance, because the amide bond between the least hindered glycine and cysteine was broken.

Therefore, this observation of "device-free thioesterification" was an exciting new discovery that appeared to triumph on account of its sheer simplicity. However, we first needed to verify that the glycine-MPA thioester **13** had indeed been formed. To do this NCL reactions were conducted.

2.2.4 Glycine-MPA Thioester Ligation

One of the ways in which a thioester can be verified is through its participation in an NCL reaction. The isolated glycine-MPA thioester **13** was therefore reacted in an NCL reaction with the model cysteinyl peptide **50** (Scheme 2.12). The reaction was monitored by LC-MS and the ligated product **14** was observed in the first LC-MS analysis after 25 h (Figure 2.6). The observed mass for **14**, AENITTGCSLNENIT, was 1578.7 Da which was in good agreement with the expected value.

AENITT SLNENIT 50
6 M Guanidine.HCl,
0.3 M Na Phosphate; pH 8.0,
2 % MESNa, 10 mM TCEP
13 14

Scheme 2.12 Successful ligation between the glycine-MPA thioester: EPO(22-28)-MPA **13** and the cysteinyl peptide: EPO(33-40) **50**, under typical NCL conditions, to afford the ligated peptide: EPO(22-28)-(33-40) **14**.[263]

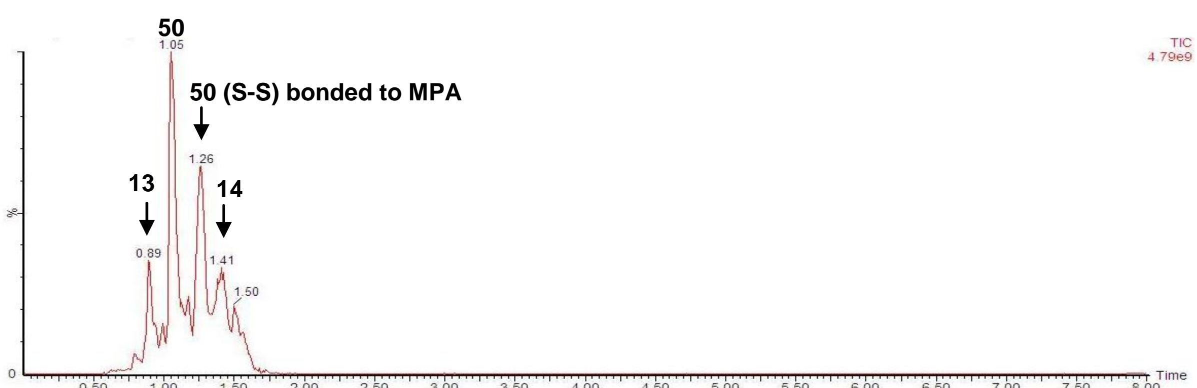


Figure 2.6 LC-MS analysis at t = 25 h of the NCL reaction between the glycine-MPA thioester: EPO(22-28)-MPA **13** and excess cysteinyl peptide: EPO(33-40) **50** to give the ligated product: EPO(22-28)-(33-40) **14** with an observed mass of 1578.7 Da.

2.2.5 Synthesis of the Model Peptide-CPE with an Internally Capped Cysteine

The model peptide-CPEs employed by Kawakami and Aimoto in ligation reactions contained a sole free cysteine residue within the C-terminal CPE moiety.[247,248] Our attempted CPE ligations were unsuccessful which we proposed could have been attributed to the presence of, or the distance between, the *two* cysteine residues in **11**. This proposition was supported to a certain extent by the unexpected, selective fragmentation of **11** at the less hindered glycine-cysteine sequence compared to the relatively more hindered alanine-cysteine sequence upon treatment with MPA.

To deduce whether the internal cysteine residue was indeed problematic during ligation, the same model EPO peptide-CPE **11** was synthesised, though this time an acetamidomethyl (Acm)-protected internal cysteine was incorporated within the same EPO(22-32) sequence, to prevent any side reactions at the internal ^{28}Gly-^{29}Cys site. The resulting peptide EPO(22-^{29}C(Acm)-32)-CPE had the sequence: H-AENITTGC(Acm)AEA-CPE (**16**) and was prepared by Fmoc-SPPS in 20 % yield, based on the resin loading of the first amino acid proline (Figure 2.7).

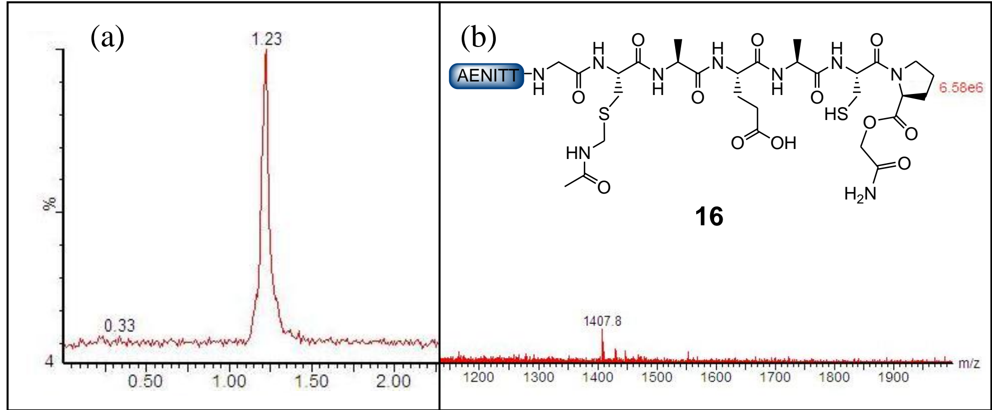


Figure 2.7 Characterisation of the synthesised, internal cysteine acetamidomethyl-protected model peptide-CPE **16** with the sequence: H-AENITTGC(Acm)AEA-CPE. (a) LC-MS UV absorbance trace of EPO(22-^{29}C(Acm)-32)CPE: **16**, absorbance detected at 254 nm. (b) ESI$^+$ MS *m/z* calculated for **16** [M]: 1406.6, observed [MH]$^+$: 1407.8.

2.2.6 CPE-Mediated Ligation with an Internally Capped Cysteine

The Cys(Acm) protected peptide-CPE **16** was then tested in a model CPE-mediated ligation reaction with the cysteinyl peptide **50** (Scheme 2.13). The reaction conditions were: 50 mM MPAA, 10 mM TCEP, and aqueous buffer at pH 8.0, 25 °C. The reaction was monitored by LC-MS analysis and continued until the CPE had been consumed.

6 M Guanidine.HCl,
0.3 M Na Phosphate; pH 8.0,
2 % MESNa, 10 mM TCEP

Scheme 2.13 Successful CPE ligation between the model peptide-CPE containing an internally acetamidomethyl (Acm) protected cysteine **16** and the N-terminal cysteinyl peptide **50** to give the ligated product **19**.

Interestingly, this time when the internal cysteine was capped, the expected full-length ligated product: EPO(22-^{29}C(Acm)-40) **19** was detected by LC-MS after 24 h. The observed mass for the ligated product was in good agreement with the expected value, and it had a retention time during LC-MS analysis that was longer than either **16** or **50** (Figure 2.8). This provided some evidence to support the hypothesis that the internal free cysteine obstructed the previous attempted CPE-mediated ligation (Scheme 2.9).

However, as well as the desired ligated product **19**, a further unexpected peak corresponding to a species that was 200 mass units larger than the ligated product was also observed (Figure 2.8).

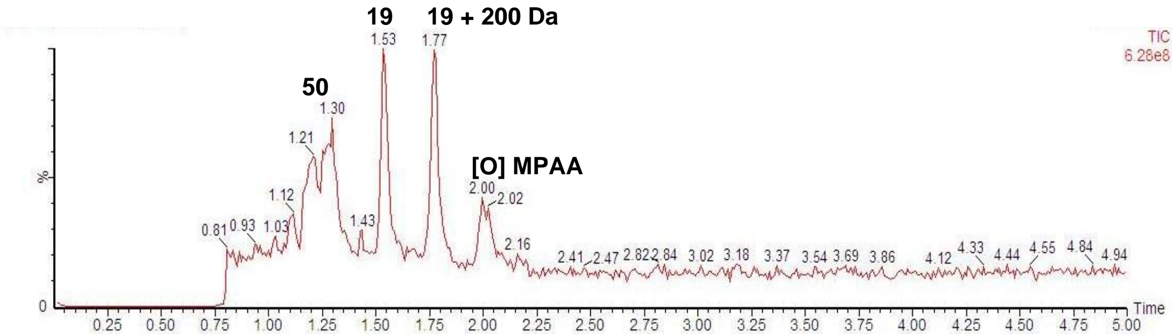
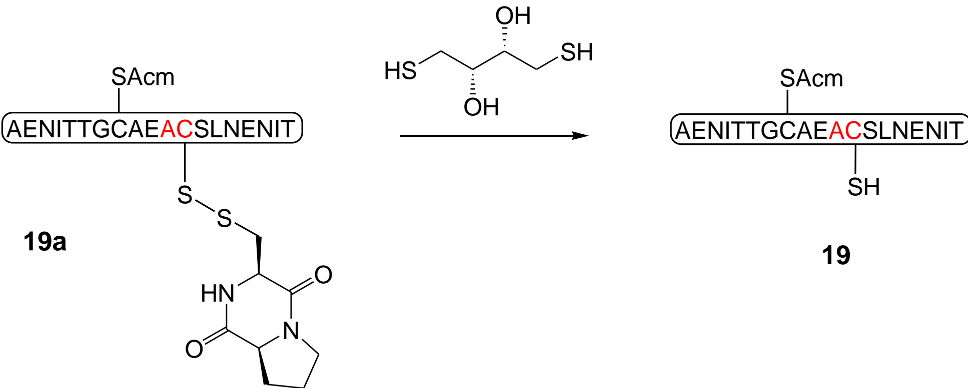


Figure 2.8 LC-MS (TIC) analysis of the ligation reaction between the internally capped cysteine containing peptide-CPE **16** (AENITTGC(Acm)AEA-CPE) and the cysteinyl peptide EPO(33-40) **50** (CSLNENIT) after 69 h to give **19** (AENITTGC(Acm)AE<u>AC</u>SLNENIT). The ligated peptide **19** had an observed mass of $[MNa + 2H]^{2+}$ 1024.3 Da, which was in good agreement with the calculated mass for the ligated peptide **19** [M] = 2023.9 Da. However, a species that was 200 Da larger than **19** with [M] = 2223.9 Da was also observed at 1.77 min.

The additional product had a longer retention time during analytical liquid chromatography than the full-length ligated product **19**, and had an observed mass that was 200 Da larger than **19**. This had not been previously reported by the Aimoto laboratory.[247,248] We initially hypothesised that this species could be the full-length ligated product **19** disulfide bonded to the excised diketopiperazine to give **19a** (Scheme 2.14).

Therefore, the more powerful reductant DTT was added to the isolated 200 Da larger species to reduce any potential disulfide bonds present and produce the cyclic-DTT by-product. However, we found that the 200 Da larger species persisted despite treatment with DTT, which suggested that an irreversible covalent bond had been formed.



Scheme 2.14 Reduction of the proposed disulfide bond in **19a** by DTT, which may be present between the ligated peptide **19** and the excised diketopiperazine by-product. Expected mass of the disulfide-containing **19a** [M] = 2221.9 Da, however this was not in agreement with the observed mass $[M + 2H]^{2+}$ 1113.3 Da which corresponded to [M] = 2223.0 Da.

Another plausible explanation for the presence of this 200 Da larger species was that a competing nucleophilic substitution reaction was taking place at the C-terminal prolyl-glycolyl site in **16** (Scheme 2.15, pathway *b*), instead of transthioesterification of any *S*-peptide produced (Scheme 2.15, pathway *a*).

The sulfhydryl group of the cysteinyl peptide **50** can attack the electrophilic carbonyl carbon of the C-terminal ester in **16** to displace the alkoxy group, which results in the formation of a full-length product but with an additional cysteine-proline insertion between the two reacting peptides (Scheme 2.15, pathway *b*). The calculated mass of the product expected from reaction pathway *b* ([M] = 2223.9 Da) was in good agreement with that observed for the 200 Da larger species, whereas the calculated mass expected from the ligated product being disulfide bonded to the excised diketopiperazine ([M] = 2221.9 Da) was not observed.

This competing substitution side-reaction appeared to act as fast as the CPE ligation proceeding *via* pathway *a*, as determined by LC-MS analysis, because after 89 h the mixture was comprised of an approximately 1:1 ratio of the desired ligated product **19** and the ligated product containing the cysteine-proline insertion. However, we did not analyse the sequence of the 200 Da larger species *via* MS-MS to verify that it was the product of pathway *b*.

Scheme 2.15 Two possible competing reaction pathways were proposed to occur during CPE-mediated ligation: *a)* an initial *N→S* acyl shift occurs to produce the expected ligation product **19** and the diketopiperazine by-product; however *b)* an addition-elimination reaction can occur between the thiol of the cysteinyl peptide **50** and the ester of the peptide-CPE **16** to provide an undesired product that is 200 Da larger than the expected product **19**, and contains a Cys-Pro insertion.

2.2.7 MPA-Mediated Thioesterification of a Peptide-CPE with an Internally Capped Cysteine

A further verification of selective fragmentation of peptides at Xaa-Cys sequences upon treatment with MPA was achieved by reacting the protected internal cysteine containing peptide-CPE **16** with MPA (Scheme 2.16). The peptide-CPE **16** was treated with 20 % aqueous MPA and the reaction was heated at 80 °C with microwave irradiation.

Scheme 2.16 Reaction of the partially protected model peptide-CPE **16** (AENITTGC(Acm)AEA-CPE) with 3-mercaptopropionic acid (MPA) afforded the full-length alanine-MPA thioester **17** (AENITTGC(Acm)AEA-MPA).[263]

After 18 h LC-MS analysis showed the presence of two major species, the desired full-length alanine-MPA thioester **17** and the unreacted starting material **16**. However, a small amount of peptide thioester hydrolysis was also observed as a result of initial thioesterification at the internal ^{28}Gly-^{29}Cys site, which occurred due to the cleavage of the acetamidomethyl group from the internal cysteine under the heated acidic reaction conditions used (80 °C, an apparent pH of 1.0).

Although the desired alanine-MPA thioester **16** was still the major product after 42 h, and some starting material remained unconverted, the amount of the undesirable glycine-MPA thioester side-product EPO(22-28)-MPA had begun to increase (Figure 2.9). Therefore, the reaction was terminated and purified by RP(HPLC) to afford the isolated alanine-MPA thioester **17** in 57 % yield.

Another detected undesirable side-product was the full-length hydrolysed peptide thioester: H-AENITTGC(Acm)AEA-OH, for which a mass of 1150.7 Da was observed (Figure 2.9). This occurred either as a consequence of hydrolysis of **17** or, and equally as plausible, hydrolysis of the *S*-peptide produced *via* an *N*→*S* acyl shift at the C-terminal alanine-cysteine sequence of **16**.

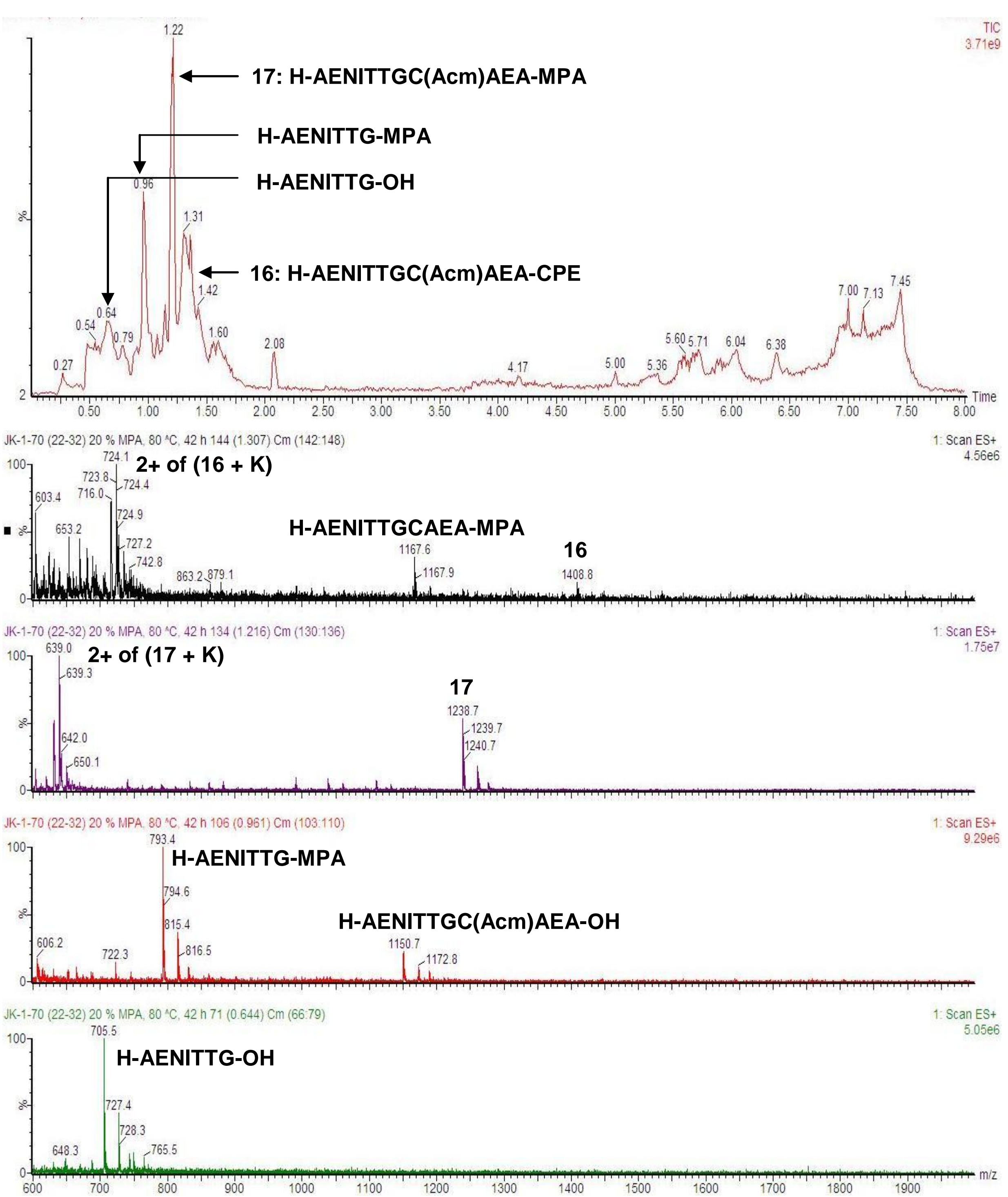


Figure 2.9 LC-MS analysis after 42 h of MPA-mediated thioesterification of the partially protected peptide-CPE **16** (AENITTGC(Acm)AEA-CPE) with 20 % v/v MPA at 80 °C to give the alanine-MPA thioester **17** (AENITTGC(Acm)AEA-MPA) as the main component.

2.2.8 Native Chemical Ligation of the Alanine-MPA thioester

The production of the full-length alanine-MPA thioester **17** was also verified in an analytical NCL reaction with the cysteinyl peptide **50** (Scheme 2.17). This produced the full-length ligated peptide **18** and further substantiated the hypothesis that internal cysteine residues, especially those comprising glycine-cysteine sequences were problematic during the initial attempted CPE ligation (Scheme 2.9).

6 M Guanidine.HCl,
0.3 M Na Phosphate; pH 8.0,
2 % MESNa, 10 mM TCEP

Scheme 2.17 An NCL reaction between the full-length alanine-MPA thioester EPO(22-^{29}C(Acm)-32)-MPA: (AENITTGC(Acm)AEA-$SCH_2CH_2CO_2H$) **17** and the cysteinyl peptide EPO(33-40): (CSLNENIT) **50** was conducted and it afforded the ligated product (AENITTGC(Acm)AEACSLNENIT) **18**, under typical ligation conditions.

2.2.9 Protein Thioesterification

In a fascinating extension of the MPA-mediated peptide thioesterification method described above,[263] wider application of MPA-mediated thioesterification was investigated with a recombinant 21 kDa sample of His-tagged wild-type (His_{10}-WT) recombinant human erythropoietin (recombinant human erythropoietin was provided by Dr Jonathan P. Richardson). Erythropoietin is comprised of 166 amino acids and contains four Xaa-Cys sequences: ^{6}Ile-^{7}Cys, ^{28}Gly-^{29}Cys, ^{32}His-^{33}Cys, and ^{160}Ala-^{161}Cys (Figure 2.10a).[259-261]

Data provided by Dr Derek Macmillan showed that when erythropoietin was treated with 20 % v/v aqueous MPA, at temperatures ranging from 40-80 °C, fragmentation of the protein was observed at temperatures as low as 40 °C. Over time, accumulation of the His_{10}-EPO(1-28)-Gly-MPA and His_{10}-EPO(1-32)-His-MPA peptide thioesters were observed (Figure 2.10).[263] Of these two thioesters the His_{10}-EPO(1-32)-histidine-MPA

peptide thioester appeared to be the most abundant.[263] The glycine- and histidine-MPA thioesters were subsequently isolated and ligated in an NCL reaction to the model cysteinyl peptide EPO(33-40) **50** using: 6 M guanidine.HCl, 300 mM sodium phosphate buffer; pH 7.4, 25 mM MPAA, 20 mM TCEP, to afford the ligated peptides.[263] The ligation experiments served to verify that the protein had indeed selectively fragmented upon exposure to MPA, to produce the corresponding thioesters.[263]

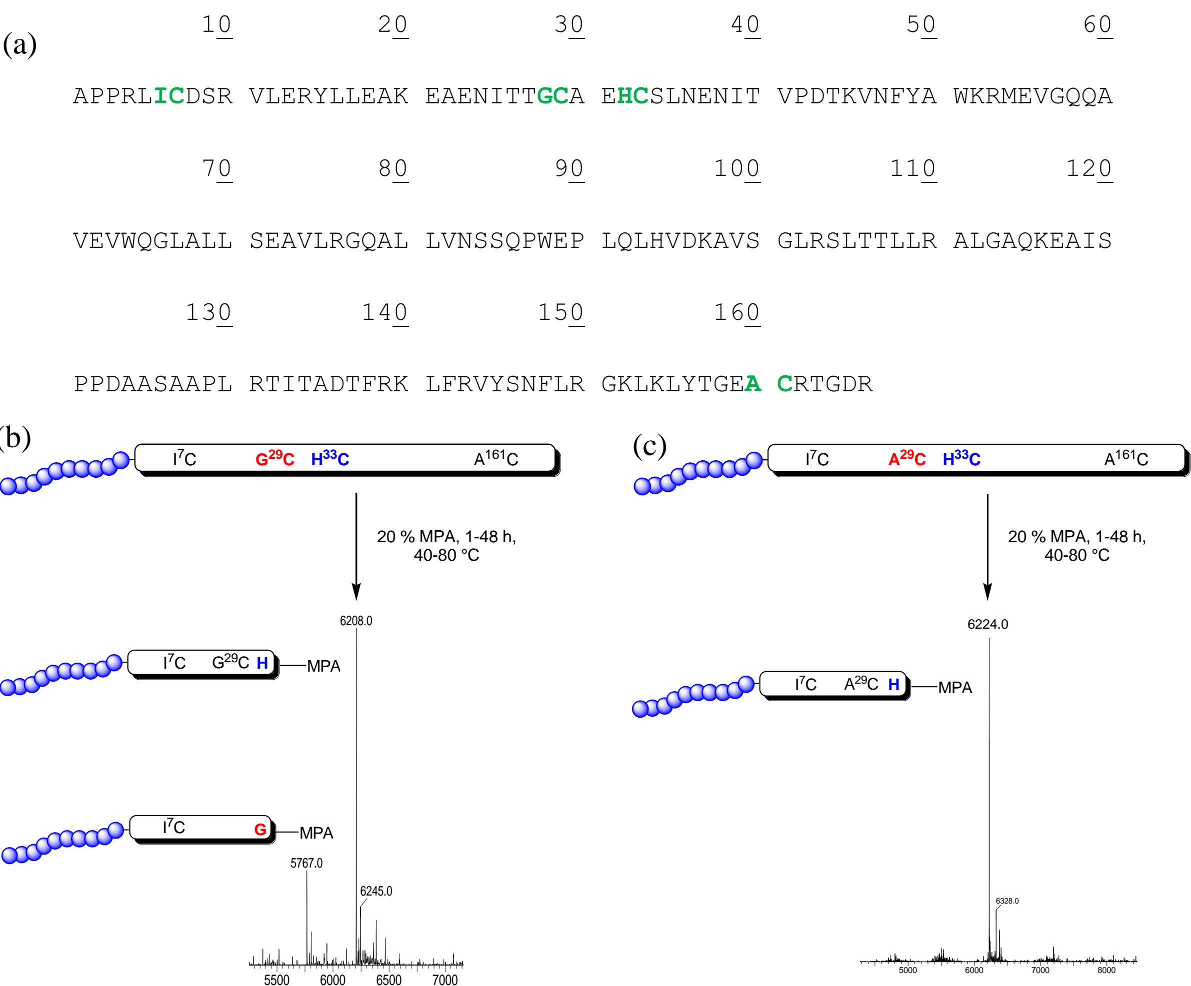


Figure 2.10 (a) Amino acid sequence of human EPO(1-166) showing the four XC sequences. His_{10}-tagged wild type recombinant human erythropoietin was observed to selectively fragment upon treatment with 20 % v/v 3-mercaptopropionic acid (MPA) at 50 °C for 16 h.[263] (b) ^{28}Gly-^{29}Cys and ^{32}His-^{33}Cys sequences within EPO were selectively thioesterified by MPA rather than ^{6}Ile-^{7}Cys or ^{160}Ala-^{161}Cys sequences. (c) Gly28Ala His_{10}-tagged wild type recombinant human erythropoietin was selectively thioesterified between the histidine-cysteine sequence by MPA as expected (data provided by Dr Derek Macmillan).[263]

Of the four possible Xaa-Cys sequences within erythropoietin, the least hindered ^{28}Gly-^{29}Cys site and the ^{32}His-^{33}Cys site were favourably thioesterified by MPA, rather than the more hindered ^{6}Ile-^{7}Cys and ^{160}Ala-^{161}Cys sites (Figure 2.10a).[263] This further supported our working selectivity hypothesis and seemed to suggest that the sites within peptides and proteins could be "tuned" to favourably fragment and provide MPA thioesters, depending on the nature of the amino acid adjacent to cysteine. Thus, amino acid residues adjacent to cysteine were thioesterified by MPA in the following order: His≥ Gly> Ala> Ile.[263]

Dawson and co-workers showed a similar trend in the reactivity of peptide thioesters that had a C-terminal histidine, glycine, or a cysteine residue in NCL reactions,[264] the reverse process. Fast ligation rates of peptide thioesters with a C-terminal histidine, glycine, or a cysteine residue were attributed to the electron withdrawing nature of these residues which resulted in activation of the thioester. This increased the reactivity of the thioester towards nucleophilic attack of the cysteinyl sulfhydryl group during the initial NCL transthioesterification step. Furthermore, the reactivity of histidine and cysteine residues can be evidenced by them being the two amino acids that are most prone to racemisation.[106]

Therefore, we proposed that MPA-mediated thioesterification may also be limited to same three residues that were observed to promote particularly fast kinetics in regular NCL reactions: histidine, glycine, and cysteine.[263]

2.2.10 Favoured Xaa-Cys Sequences Undergoing MPA-Mediated Thioesterification

Data provided by Dr Derek Macmillan showed that peptides could be specifically tuned to fragment at pre-determined Xaa-Cys sequences (Table 2.2).[263] Short model peptides based on the sequence of EPO(22-32): H-AENITTXCAEXC-NH_2 containing two potential competing Xaa-Cys cleavage sites were synthesised to test the scope of MPA-mediated thioesterification. Peptides were treated with 20 % v/v MPA at 50 °C, and as predicted histidine, glycine, and cysteine-cysteine sequences were found to preferably fragment, rather than the isoleucine-cysteine sequence also present within the same peptide, to provide the corresponding MPA thioesters.[263] Thioesters were isolated in approximately 30 % yield when two equally favoured sites for fragmentation were present

within the same peptide, such as: glycine-cysteine and histidine-cysteine or glycine-cysteine and cysteine-cysteine sequences (Entries 1 and 2, Table 2.2), and 60 % yield when only one favoured fragmentation site was present, such as in the peptide containing both isoleucine-cysteine and histidine-cysteine (Entry 3, Table 2.2).[263] Hence, isoleucine-cysteine sequences were found to not fragment, whereas glycine-cysteine, histidine-cysteine, and cysteine-cysteine sequences readily fragmented upon treatment with MPA.[263]

Table 2.2 Preferred sites of MPA-mediated thioesterification within a single peptide were further explored by varying the amino acid adjacent to cysteine. Model peptides were based on a short sequence of EPO(22-32): H-AENITT**XC**AE**XC**-NH_2, and contained two XC competing sites of thioesterification. These model peptides were treated with 20 % v/v MPA at 50 °C for 36 h. Reactions were monitored by LC-MS for the cleavage ratio at 36 h, and after 36 h were purified by semi-preparative (RP)HPLC for the isolated thioester yields (data provided by Dr Derek Macmillan).[263]

Entry	**Peptide sequence**	**Cleavage Ratio**	**Isolated yield** [a] **%**	**Thioester formation**		**Thioester formation**	
	H-AENITT^{1}XCAE^{2}XC-NH_2	^{1}XC : ^{2}XC		^{1}X-MPA		^{2}X-MPA	
				calc. *m/z*	obs. *m/z*	calc. *m/z*	obs. *m/z*
1	H-AENITTGCAEHC-NH_2	≈ 1: 1	28	793.3	793.5	1233.5	1233.6
2	H-AENITTGCAECC-NH_2	≈ 1: 1	28	793.3	793.5	1199.4	1199.6
3	H-AENITTICAEHC-NH_2	> 9: 1	60	849.4	n/o	1289.5	1289.7

[a]Isolated yield for the full-length 11-mer peptide MPA thioester.[263]

In a bid to further probe amino acid residues that were favoured or unfavoured in MPA-mediated thioesterification, another small screen of model peptides was undertaken by Dr Derek Macmillan (Table 2.3).[263] The model peptides were based on the erythropoietin sequence: H-AENITTXC-NH_2 where X was glycine, alanine, proline, glutamic acid, valine, serine, and aspartic acid; and were subjected to forcing thioesterification conditions: 20 % v/v MPA and heating at 80 °C for 24 h.[263] After 20 h, glycine-MPA thioester formation was found to be almost complete (Entry 1, Table 2.3).[263] However, only trace amounts of proline

and valine-MPA thioesters were formed (Entries 3 and 5, Table 2.3), and lysine and tryptophan were the least efficient amino acids adjacent to cysteine in promoting thioesterification.[263] An approximately 1:1 mixture of the starting material to peptide thioester resulted after 20 h for peptides that contained a C-terminal glutamate-cysteine, alanine-cysteine, serine-cysteine, or phenylalanine-cysteine sequence (Entries 2, 4, and 6, Table 2.3).[263] Both the serine-containing starting peptide and product thioester also appeared to form an MPA ester on the primary hydroxyl group of serine (Entry 6, Table 2.3).[263] Interestingly, the aspartate-cysteine containing model peptide did not give rise to the MPA thioester, instead the major product was analysed by LC-MS to correspond to the loss of both aspartate and cysteine residues (Entry 7, Table 2.3).[263]

Table 2.3 MPA-Mediated selective thioesterification of model peptides was further explored by varying the amino acid adjacent to cysteine. Model peptides were based on a short sequence of EPO(22-29): H-AENITT**XC**-NH_2, and X was varied. The model peptides were treated with 20 % v/v MPA at 80 °C for 20 h. Reactions were monitored by LC-MS, and were purified by semi-preparative (RP)HPLC for the isolated yields (data provided by Dr Derek Macmillan).[263,a]

Entry	**Peptide sequence** H-AENITTXC-NH_2	**Isolated yield of starting material** %	**Isolated yield of thioester** %	**Thioester formation** X-MPA calc. *m/z*	obs. *m/z*
1	H-AENITTGC-NH_2	0	33[b]	793.3	793.5
2	H-AENITTAC-NH_2	40	42	807.4	807.6
3	H-AENITTPC-NH_2	49	0	833.4	833.5
4	H-AENITTEC-NH_2	n.d.	n.d.	865.4	865.5
5	H-AENITTVC-NH_2	67	0	835.4	835.6
6	H-AENITTSC-NH_2	0	n.d.	823.4	911.5[c]
7	H-AENITTDC-NH_2	0	n/o[d]	851.3	n/o

[a]When XC = FC, KC, or WC, product isolation was difficult. [b]Loss in yield due to thioester hydrolysis. [c]Starting material and product thioester appeared further esterified by MPA. [d]Peptide had been consumed during reaction but no thioester was observed.[263]

2.2.11 Racemisation

Data provided by Dr Derek Macmillan showed that when a short model peptide: Ac-Ala-(L)His-Cys-NH_2 was treated with 20 % v/v MPA at 55 °C for 48 h, the resulting MPA thioester Ac-Ala-His-$SCH_2CH_2CO_2H$ exhibited undetectable racemisation by HPLC, when compared to the thioester that resulted upon reacting Ac-Ala-(D)His-Cys-NH_2 to identical thioesterification conditions.[263]

Racemisation *via* oxazolone formation upon activation of the C-terminus of the peptide during MPA-mediated thioesterification was not expected to be problematic as the reactions were conducted at acidic pH (~1.0).

2.2.12 MPA Thioester Characterisation by ^{1}H NMR Spectroscopy

A model study was conducted to characterise the peptide MPA thioesters formed by ^{1}H NMR spectroscopy.[263] For this purpose two model peptides were synthesised based on the erythropoietin sequence EPO(22-29): H-AENITT$\underline{X}$C-NH_2, where X was either glycine or histidine. These peptides were treated with 20 % v/v MPA at 60 °C for 48 h. MPA-mediated thioesterification was monitored by LC-MS and after 48 h the reaction mixtures were centrifuged to remove the MPA-derived precipitate, and the peptide-containing supernatant was purified by semi-preparative (RP)HPLC.

Data provided by Dr Derek Macmillan showed that the model peptide that terminated in a histidine-cysteine sequence was readily converted into the histidine-MPA thioester and could also be isolated by HPLC in several milligram quantities. Comparison of the ^{1}H NMR spectrum of the histidine-cysteine containing starting peptide **54** into the histidine-MPA product thioester **55** showed the expected disappearance of the cysteine β-CH_2 signals in the product, as well as the characteristic appearance of the MPA 2(CH_2) signals (Figure 2.11 provided by Dr Derek Macmillan).[263] A downfield shift in the α-CH signal of histidine in the ^{1}H NMR spectra of **55** was also observed.

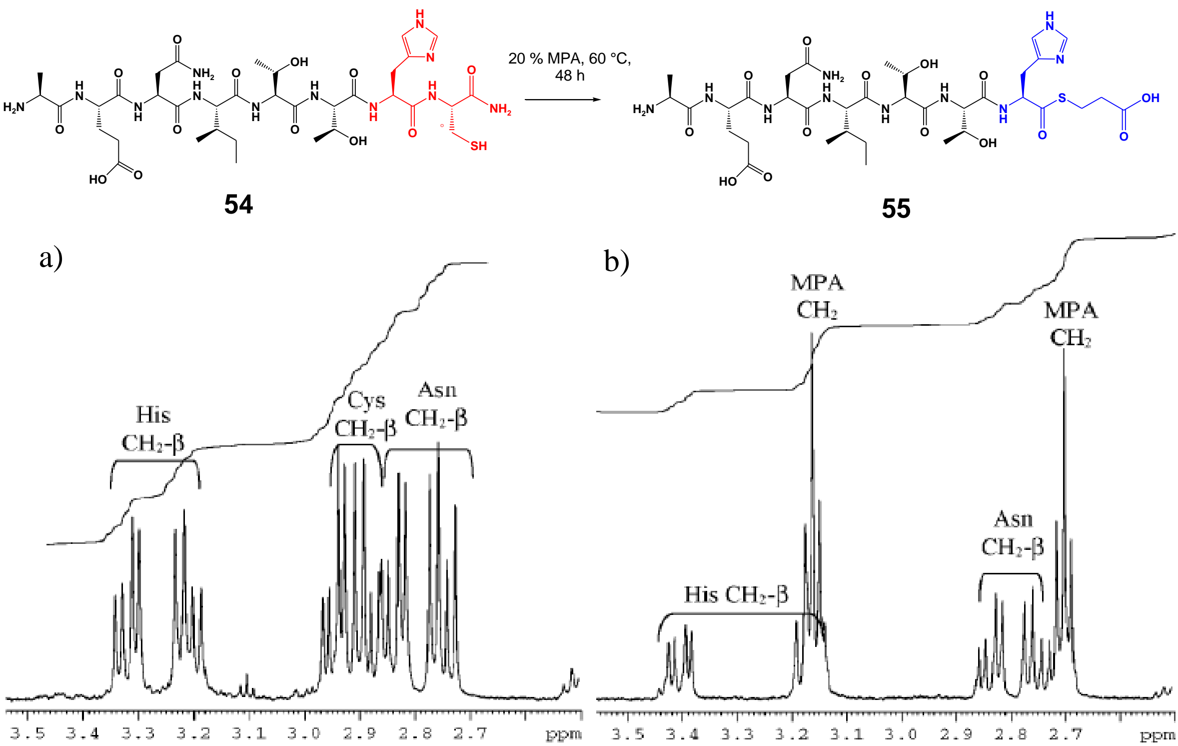


Figure 2.11 ^{1}H NMR spectra (500 MHz, solvent = D_2O, number of scans = 16, room temperature) of the EPO(22-29) model peptide functionalised with a C-terminal histidine-cysteine sequence: H-AENITTHC-NH_2 **54** (a) before and (b) after MPA-mediated thioesterification to afford the histidine-MPA thioester: H-AENITTH-$SCH_2CH_2CO_2H$ **55** (data provided by Dr Derek Macmillan).[263]

Similarly, comparison of the ^{1}H NMR spectrum of the glycine-cysteine containing starting peptide **20** into the glycine-MPA product thioester **21** also showed the expected disappearance of the cysteine β-CH_2 signals in the ^{1}H NMR spectrum of **21**, as well as the characteristic appearance of the MPA 2(CH_2) signals (Figure 2.12a, b). The thioesterification reaction also resulted in the hydrolysis of **21** and the ^{1}H NMR of the isolated hydrolysed thioester **21b** showed the characteristic disappearance of the MPA 2(CH_2) signals (Figure 2.12c).

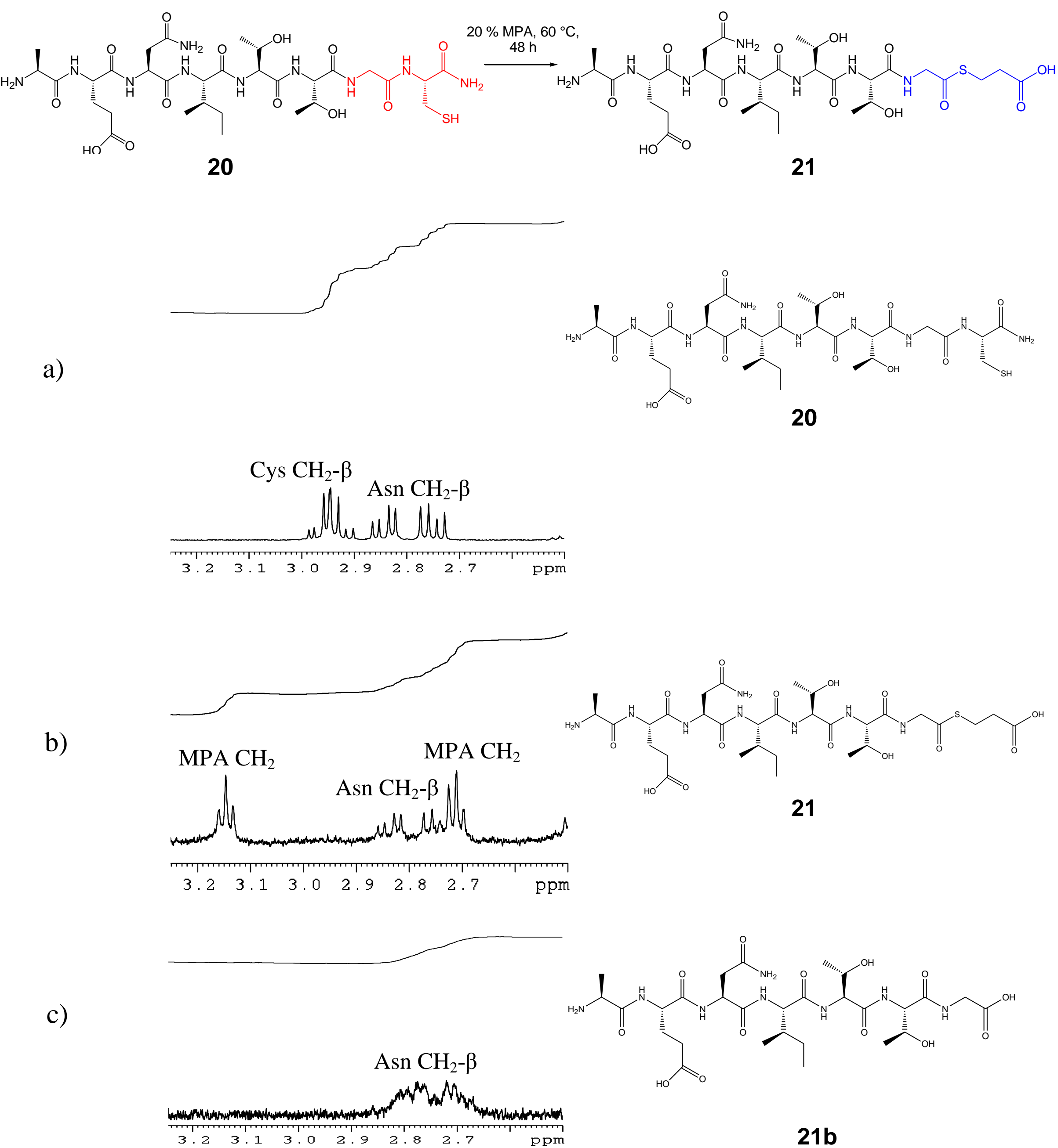


Figure 2.12 ^{1}H NMR spectra (500 MHz, solvent = D_2O, number of scans = 16, room temperature) of MPA-mediated thioesterification of a short model EPO(22-29) peptide functionalised with a C-terminal glycine-cysteine sequence. (a) Starting material: H-AENITTGC-NH_2 **20** (~5.0 mg/500 µL). (b) Glycine-MPA thioester: H-AENITTG-$SCH_2CH_2CO_2H$ **21** (~3.0 mg/500 µL). (c) Undesirable side-product, the hydrolysed glycine thioester: H-AENITTG-OH **21b** (~0.5 mg/500 µL).

2.2.13 Discussion

During our endeavours towards the semisynthesis of the glycoprotein erythropoietin, we experimented with the cysteinylprolyl ester method.[247,248] We observed thioesterification as a consequence of the following model experiment. We had devised a parallel set of CPE ligation reactions employing identical peptide sequences (sequence: AENITTGCAEA-CPE). The difference between the two peptides was that one contained free cysteine sulfhydryls throughout, and the other contained a single acetamidomethyl (Acm) protected internal cysteine. The partially protected peptide **16** was found to ligate to the corresponding cysteinyl peptide (CSLNENIT) **50** to afford the full-length peptide: AENITTGC(Acm)AEACSLNENIT **19** whereas the fully deprotected peptide did not. We reasoned that the presence, and more importantly the position,[265] of the additional free internal cysteine residue may interfere with the desired ligation enough to limit this application of the CPE method. However, this was not further explored by changing the position of the internal cysteine.

Furthermore, two competing sites of reaction were proposed during CPE-mediated ligation of the partially protected peptide-CPE **16** with **50**, which decreased the amount of the desired ligated product **19**. One reaction pathway was the desired route: an *N→S* acyl shift followed by diketopiperazine formation; and the other was an undesired route: nucleophilic attack of the N-terminal cysteine of **50** onto the ester which resulted in a cysteine-proline insertion between **50** and **16**. These findings led us to explore alternative strategies to attain peptide thioesters from the fully deprotected peptide-CPE **11**.

In doing so, we recalled that Nakahara and co-workers had reported the use of aqueous 3-mercaptopropionic acid (MPA) to produce peptide thioesters by means of an *N→S* acyl transfer.[239] Upon reaction of our model peptide-CPEs **11** and **16** with MPA, we found thioesterification occurred selectively at the internal glycine-cysteine motif rather than at the expected C-terminal alanine-cysteine sequence (Scheme 2.18).[263]

Scheme 2.18 *Upper route:* attempted CPE ligation between EPO(22-32)-CPE: AENITTGCAEA-CPE **11** and EPO(33-40): CSLNENIT **50** to afford the ligated product EPO(22-40) **12**. *Lower route:* an unexpected, novel MPA-mediated thioesterification reaction discovered. The model peptide-CPE **11**, which contained two free cysteine residues, was selectively transformed into the glycine-MPA thioester: AENITTG-$SCH_2CH_2CO_2H$ **13**.[263]

Delving deeper into the profile of favoured Xaa-Cys motifs led us to conclude that, rather surprisingly, peptide thioesters could be selectively formed between histidine-cysteine, glycine-cysteine, and cysteine-cysteine sequences upon exposure to MPA.[263] Coincidentally, peptide thioesters with a C-terminal histidine, glycine, or a cysteine residue are also known to accelerate the rate of NCL,[264] the reverse process. Selective thioesterification at histidine-cysteine and glycine-cysteine sequences was also illustrated in the context of the recombinant protein erythropoietin (EPO), which contained 166 amino acids.[263] Consequently, installation of a specific sequence potentially allows us to program the chemoselective fragmentation of peptides and proteins in the presence of aqueous MPA to afford the highly desirable thioesters.

Our studies resulted in the discovery of a new thioesterification method, termed "Native Chemical Thioesterification" (NCT), which could in the future provide an alternative to the intein-mediated method of preparing protein thioesters, as it requires no additional stimulus other than the appropriate sequence. Existing methods to prepare synthetic

peptide thioesters often require the use of a specialised resin, linker, device, or a cleavage cocktail which tends to limit the general applicability of those methods. Hence, preparing peptide thioesters by simply installing a histidine-cysteine, glycine-cysteine, or a cysteine-cysteine sequence can be synthetically more convenient.

Mechanistically, it is likely that an initial *N*→*S* acyl shift occurs by reversible nucleophilic attack of the cysteine thiol onto the carbonyl carbon of the amide across the histidine-cysteine, glycine-cysteine or cysteine-cysteine junction to give a proposed hydroxythiazolidine intermediate,[266] which then releases the protonated amine and undergoes an intermolecular thioester exchange with MPA to give the isolable thioester (Scheme 2.19). Thus, NCL is reversed.

Scheme 2.19 A novel thioester-forming reaction discovered, termed "Native Chemical Thioesterification" (NCT), which proceeds through an *N*→*S* acyl shift site-specifically at histidine-cysteine, glycine-cysteine, and cysteine-cysteine sequences.[263] 'RSH' denotes a small molecular weight thiol such as 3-mercaptopropionic acid (MPA).[263]

In this thioesterification reaction (Scheme 2.19), an amide bond is selectively disrupted to form a thioester through an *N*→*S* acyl shift. Energetically, this is an uphill process because a more stable and lower in energy amide bond is broken to form a less stable, higher in energy thioester. The half-life of a typical amide bond in aqueous solution to hydrolysis is extremely slow, approximately 10 to 1000 years in water at neutral pH and room temperature in the absence of a catalyst,[21,22,107] despite being thermodynamically favoured. The driving force for the breakdown of an amide bond to afford a thioester can be explained by considering that the reaction overall is conducted at acidic pH. This favours breakdown of the amide bond through protonation, though whether backbone amide protonation is the first step in the reaction mechanism or not is unknown.

Peptide thioester hydrolysis was also observed to increase over time with MPA.[263] Hence further optimisation of this new method was required to make it a more efficient and less wasteful process.

MPA-mediated thioesterification at glycine-cysteine sequences can be attributed to glycine being the least hindered amino acid, but the reason for preferred fragmentation at histidine-cysteine and cysteine-cysteine sites remains unclear. It can be postulated that the histidine or cysteine residues adjacent to cysteine act as a general acid or base to protonate or deprotonate reacting species, which is acknowledged to be the role of histidine residues located proximal to the *N→S* acyl shift site within inteins.[182] Alternatively, and rather intriguingly, the amide bond between histidine-cysteine, glycine-cysteine, and cysteine-cysteine sequences may adopt an atypical geometry, particularly under the reaction conditions used, which could cause ground-state destabilisation and result in "activation" of the scissile amide bond.[31,178,179] Preferential fragmentation may also be attributable to the electron withdrawing nature of these amino acid residues which can activate the scissile amide bond towards nucleophilic attack.[264]

2.3 Conclusion

Our initial aim was to test Kawakami and Aimoto's cysteinylprolyl ester methodology in model peptide ligations,[247,248] with an ultimate aim of utilising these CPEs for the semisynthesis of erythropoietin. A model peptide bearing a C-terminal CPE that was based on a short sequence of erythropoietin: H-AENITTGCAEA-CPE **11** was synthesised, and test ligations were conducted with the corresponding cysteinyl peptide: H-CSLNENIT-OH **50**.[263]

When CPE-mediated ligation was not observed under archetypal ligation conditions it was concluded that the presence and also the position of the internal free cysteine residue within the model peptide-CPE played a part in prohibiting the ligation.[265] Further supporting evidence for the internal cysteine being an inhibitory factor during model CPE ligations was provided upon treatment of the same peptide-CPE **11**, which contained two free cysteine residues, with 3-mercaptopropionic acid (MPA).

MPA-mediated thioesterification was unexpectedly observed between the internal glycine-cysteine site rather than at the expected C-terminal alanine-cysteine site,[263] and this selectivity was attributed to glycine being the sterically least hindered amino acid.

Further studies were conducted on the preferred sites of MPA-mediated thioesterification and those found to be favoured were: histidine-cysteine, glycine-cysteine, and cysteine-cysteine sequences.[263] Of these three, histidine-cysteine sequences were found to be particularly prone to thioesterification.[263] Placement of these three residues (histidine, glycine, or cysteine) at the C-terminus of a peptide thioester has also been previously shown to increase the rate of NCL,[264] the reverse process. When bulkier amino acids (proline, valine, phenylalanine, isoleucine, tryptophan and lysine) were placed adjacent to the cysteine residue in model peptides, MPA-mediated thioesterification was not as efficient.[263]

An overall reaction pathway was proposed to consist of an initial intramolecular *N→S* acyl shift occurring favourably at histidine-cysteine, glycine-cysteine, and cysteine-cysteine sequences to afford the *S*-peptide,[215,234-241,243-245,247-250] which then undergoes intermolecular transthioesterification with MPA to form the isolable MPA thioester.[263] These initial experiments enlightened us to the possibility of extending this reaction to tune the sites within both peptides and proteins to selectively fragment and provide the highly desirable thioesters.[263] However, optimisation of the reaction conditions was required since we observed thioester hydrolysis which decreased the yield of the thioester.[263]

2.4 Acknowledgments

Data for Sections 2.2.9-2.2.11 and the ^{1}H NMR spectra for the thioesterification of H-AENITTH-NH_2 **54** in Section 2.2.12 was provided by Dr Derek Macmillan. Recombinant human erythropoietin for Section 2.2.9 was expressed by Dr Jonathan P. Richardson.

3 Reaction Optimisation and First Biological Application of Native Chemical Thioesterification

3.1 Introduction

We came across an interesting observation whereby synthetic peptides and a recombinant protein containing a histidine-cysteine, glycine-cysteine, or a cysteine-cysteine sequence selectively fragmented to produce thioesters upon exposure to 3-mercaptopropionic acid (MPA). The reaction was termed Native Chemical Thioesterification (NCT) since it reversed the process of Native Chemical Ligation (NCL) (Scheme 3.1).[263]

(a)

-RSH

pH 6.5-8.0

(b)

N-S acyl shift

Scheme 3.1 (a) Native Chemical Ligation (NCL)[144] compared to (b) Native Chemical Thioesterification (NCT), our novel thioester-forming reaction which proceeds *via* an initial *N→S* acyl shift followed by transthioesterification with a thiol (i.e. 3-mercaptopropionic acid) to afford the isolable thioester.[263]

However, certain disadvantages of using MPA as the thiol were observed. Thioester hydrolysis was a side-reaction that was observed to increase over time when samples were heated with aqueous MPA.[263] Intermolecular disulfide bond formation between MPA and the starting material was also observed (Figure 3.1).[263] MPA had several strong UV absorbance

peaks during HPLC purification that potentially precluded analysis and product isolation.[263] MPA was also observed to precipitate out of the reaction mixture over time due to disulfide bond formation, which decreased the effective concentration of the thiol available to implement the second transthioesterification step within our proposed mechanism (Scheme 3.1 b).[263] Therefore, this new thioesterification method required further optimisation.

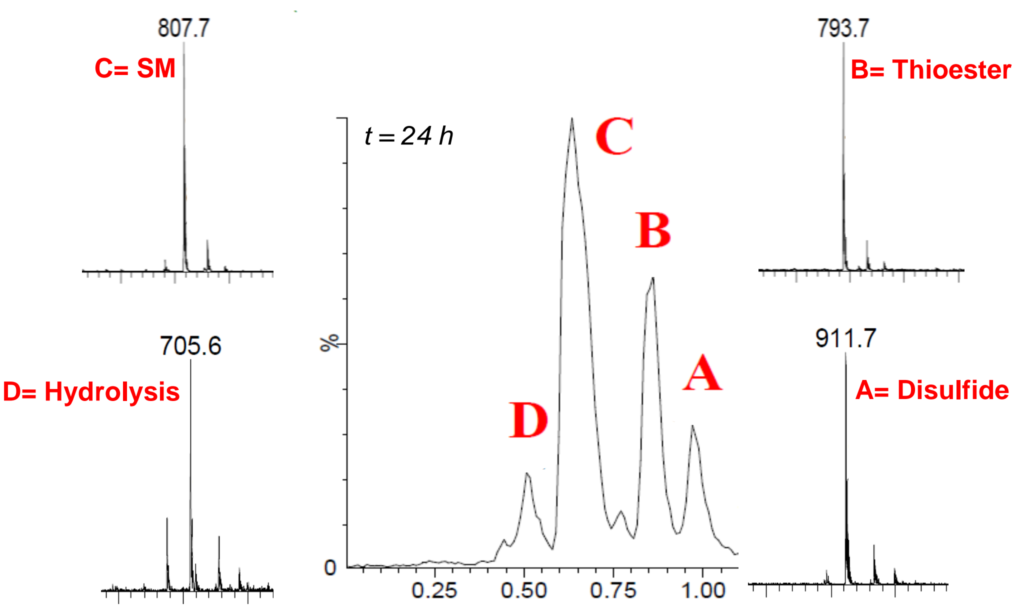


Figure 3.1 Typical LC-MS analysis of MPA-mediated thioesterification proceeding *via* an *N→S* acyl shift, analysed after t = 24 h. The starting material based on a short sequence of EPO(22-29): H-AENITTGC-NH_2 (C) was converted into the MPA thioester: H-AENITTG-$SCH_2CH_2CO_2H$ (B). Undesirable thioester hydrolysis: H-AENITTG-OH (D) and disulfide bond formation: H-AENITTGC(MPA)-NH_2 (A) were also observed.[263]

Another important goal we aimed to achieve was the application of optimised thioesterification reaction conditions to the synthesis of a biologically active protein, prepared from fully-characterised starting materials and intermediates.

3.2 Results and Discussion

3.2.1 Optimisation of One-Pot Peptide Thioesterification

Our aims for optimising MPA-mediated Native Chemical Thioesterification (NCT) included the following:

- Decrease the amount of the unproductive side-products: peptide thioester hydrolysis and peptide-disulfide bond formation.
- Increase the isolated yields of simple and more complex peptide thioesters.
- Increase the overall rate of peptide thioesterification so that peptides and proteins are not heated under acidic conditions for prolonged periods.
- Decrease the reaction temperature to expand the potential application of NCT to peptides and proteins containing heat-sensitive moieties.

Preliminary experiments were conducted with the model peptide EPO(22-29): H-AENITTGC-NH_2 **20** (Scheme 3.2). MPA-mediated peptide thioesterification was initially investigated by reacting **20** with 10 % v/v aqueous MPA and heating at 60 °C. The reaction was monitored at regular intervals by LC-MS analysis.

AENITT H N O N H SH NH2 O 10 % MPA, 60 °C AENITT H N O S CO2H

20 **21**

Scheme 3.2 Initial optimisation studies were conducted to determine the optimum reaction time of MPA-mediated peptide thioesterification. The model peptide H-AENITTGC-NH_2 **20** was reacted with 10 % v/v MPA, at 60 °C, and the reaction was monitored by LC-MS.

After 24 h, the major component of the reaction was unreacted starting material **20** and whilst a significant amount of the product MPA thioester H-AENITTG-$SCH_2CH_2CO_2H$ **21** had been formed, peptide thioester hydrolysis to give H-AENITTG-OH **21b** and undesirable intermolecular disulfide bond formation H-AENITTGC(MPA)-NH_2 **21c** were also detected. Elongating the reaction time to 48 h resulted in an increase in the amount of the peptide MPA thioester **21** formed, and the ratio of **20**: **21** was observed to be approximately 1:1. However, increasing the reaction time to 72 h, to force the reaction to completion, resulted in a significant amount of peptide thioester hydrolysis **21b**. The ratio of the unreacted starting material **20** to product thioester **21** and hydrolysed product **21b** was approximately 1:1:1 after 72 h (Figure 3.2). Hence, the reaction appeared to reach maximal conversion by 48 h, after which peptide hydrolysis became significant.

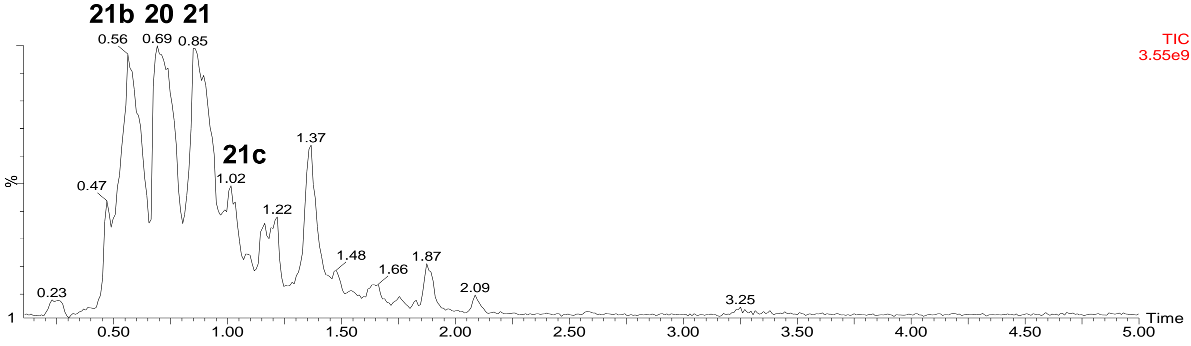


Figure 3.2 LC-MS analysis of peptide thioesterification of **20** upon exposure to 10 % v/v MPA at 60 °C after 72 h. The starting peptide (H-AENITTGC-NH_2) **20**, thioester (H-AENITTG-$SCH_2CH_2CO_2H$) **21**, and hydrolysed product (H-AENITTG-OH) **21b** were formed in an approximately 1:1:1 ratio. The starting peptide **20** was also found to be disulfide bonded to MPA through its C-terminal cysteine residue (**21c**).

^{1}H NMR spectroscopy was previously used to characterise the peptide-MPA thioester which, although essential for characterisation, offered limited information about the connectivity of the MPA molecule to the peptide (Section 2.2.12).[263] Therefore, ^{13}C NMR spectroscopy was employed as a tool to optimise peptide thioester formation under various reaction conditions.

For this purpose, a model peptide was synthesised with a site-specific ^{13}C isotopic label (Scheme 3.3).[267] ^{13}C-Labelled glycine **22** was incorporated into the Fmoc-SPPS of the EPO-based peptide **26** with the sequence: H-AENITTG(^{13}*C*-1)C-NH_2.[267] This resulted in a C-terminal labelled amide bond within the C-terminal glycine-cysteine sequence in **26**, which was chosen because we had previously shown that peptides containing histidine-cysteine, glycine-cysteine, and cysteine-cysteine sequences readily fragmented (Chapter 2).[263]

H_2N–CH_2–$^{13}C(O)$–OH **22** → FmocOSu, Et_3N (67 %) → FmocHN–CH_2–$^{13}C(O)$–OH **23**

FmocHN–resin **24** → 1. piperidine; 2. Fmoc-Cys(Trt)-OH, DIPEA, HOBt/HBTU, DMF; 3. piperidine → **25**

1. FmocHN–CH_2–$^{13}C(O)$–OH, DIPEA, HOBt/HBTU, DMF (**23**); 2. Fmoc-SPPS; 3. 95 % TFA → **26** (83 %)

Scheme 3.3 Synthesis of the ^{13}C-labelled model peptide **26** (AENITTG(^{13}*C*-1)C-NH_2) with a C-terminal glycine-cysteine sequence for use in thioester optimisation studies.[267] The α-amino group of ^{13}C-labelled glycine **22** was initially Fmoc-protected to afford **23**, and this was then incorporated during the Fmoc-based solid phase peptide synthesis of **26** using standard coupling methodology (see Experimental Section 6 for further details).

It was reasoned that if the role of MPA was solely to act as a transthioesterification reagent in the latter part of the proposed mechanism (Scheme 3.1 b), then a peptide thioester could potentially be prepared by any thiol *via* an intermolecular thiol-thioester exchange. To test this hypothesis, thioesterification was investigated with a range of thiols and the significance

of having both acid and thiol functionalities present within the same reagent was also explored. The readily available, water-soluble transthioesterification reagents tested and certain notable chemical properties of these thiols are summarized in Table 3.1.

Table 3.1 Selected readily-available, water-soluble thiols that were used in the optimisation studies are listed. The thiol, carbon chain length, and presence of a carboxylic acid functionality within the thiol was varied.

Thiol (RSH)	Structure	Abbreviation	Carbon chain length	pKa[268]
2-Mercaptoacetic acid	HS–CH2–C(=O)OH	MAA	2	10.3 (SH), 3.7 (CO_2H)
3-Mercaptopropionic acid	HS–CH2–CH2–C(=O)OH	MPA	3	10.3 (SH), 4.3 (CO_2H)
4-Mercaptobutyric acid*	HS–CH2–CH2–CH2–C(=O)OH	MBA	4	10.3 (SH), 4.7 (CO_2H)
β-Mercaptoethanol	HS–CH2–CH2–OH	BME	2	9.5 (SH)
Sodium 2-mercaptoethanesulfonate	HS–CH2–CH2–S(=O)(=O)–O⁻Na⁺	MESNa	2	9.1 (SH)

* Synthesised from 4-bromobutyric acid (see Experimental Section 6).

MPA is comprised of a three carbon chain and in order to determine whether the distance between the acid and thiol functionalities within this thiol was important, the carbon chain length was varied with 2-mercaptoacetic acid (MAA) and 4-mercaptobutyric acid (MBA). The importance of the presence of the carboxylic acid functionality within MPA was also investigated by testing thiols that contained either a sulfonate group: sodium 2-mercaptoethanesulfonate (MESNa), or a hydroxyl group: *β*-mercaptoethanol (BME).

The general procedure for the optimisation studies involved dissolving the ^{13}C-labelled model peptide **26** in a solution of 10 % thiol in D_2O at a peptide concentration of 6 mM, with or without acid, to study a range of pH values (1.0 to 8.0) (Scheme 3.4).[267] The reactions were heated at the fixed temperature of 60 °C by conventional heating and the reactions were continued for a total time of 72 h.[267]

Scheme 3.4 The optimisation studies involved thioesterification of the ^{13}C-labelled model peptide H-AENITTG(^{13}*C*-1)C-NH_2 **26** terminating in a glycine-cysteine sequence, under various reaction conditions to afford the corresponding ^{13}C-labelled thioesters **29**.[267] The reactions were conducted in D_2O and monitored by ^{13}C NMR spectroscopy (125 MHz, number of scans = 256) at regular intervals between t = 0 and 72 h. The relative intensities of the ^{13}C-labelled carbonyl carbon peaks were then used to determine the percentage conversion of the starting peptide **26** into the product(s).

Aliquots of the reaction mixture were removed at specific intervals for ^{13}C NMR spectroscopic analysis. During that time, disappearance of the starting material amide chemical shift (**C**ONH) δ_C ~171 ppm was observed, along with the appearance of the product thioester (**C**OSR) δ_C ~200 ppm and/or hydrolysed peptide (**C**O_2H) δ_C ~173 ppm.[267] Examples of ^{13}C NMR spectra acquired for various reaction conditions studied during the optimisation studies are shown in Figure 3.3.[267]

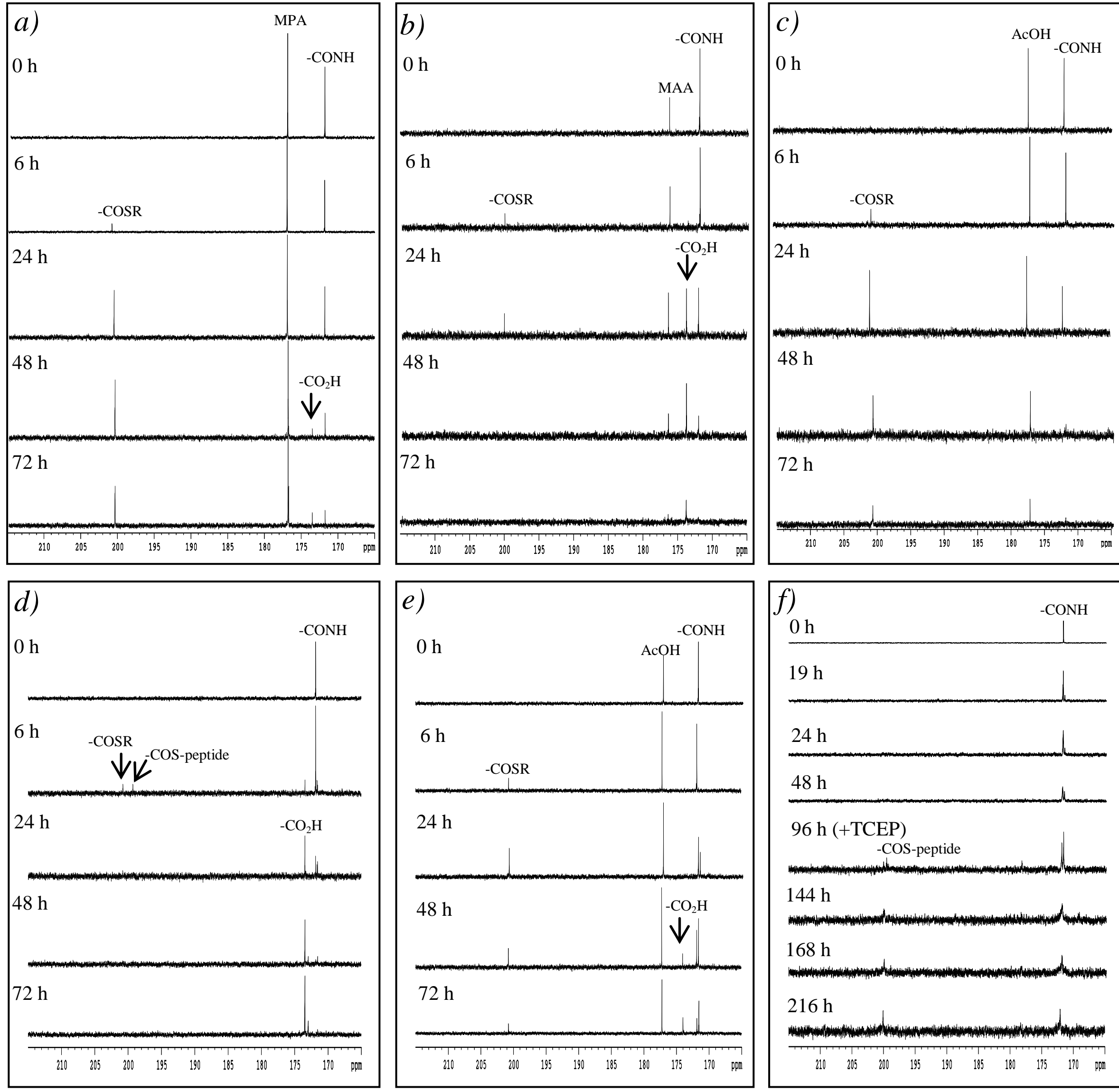


Figure 3.3 Peptide thioesterification of the model peptide (H-AENITTG(^{13}*C*-1)C-NH_2) **26** was monitored under various reaction conditions by ^{13}C NMR spectroscopy. The labelled peptide **26** was dissolved in D_2O at a concentration of 6 mM, and the thioesterification reaction was heated using conventional heating at 60 °C. ^{13}C NMR spectra were acquired at regular intervals (125 Hz, number of scans = 256, room temperature). *Reagents and conditions: a)* 10 % MPA, pH ~1.0; *b)* 10 % MAA, pH 1.0-2.0; *c)* 10 % MESNa, 1 % AcOH, pH ~2.0; *d)* 10 % BME, 1 % TFA, pH ~1.0; *e)* 10 % BME, 1 % AcOH, pH ~2.0; *f)* 71 % TFA-*d*, 29 % $CDCl_3$, (0.5 % TCEP.HCl added after 48 h).[267]

The relative intensities of the peptide-derived carbonyl carbon signals during thioesterification of the ^{13}C-labelled model peptide **26** under various reaction conditions, were then used to calculate the percentage of starting material converted into the peptide thioester **29** and the hydrolysed product (H-AENITTG(^{13}*C*-1)-OH) **29b**.[267] These conversions were plotted in the graphs shown in Figure 3.4.[267]

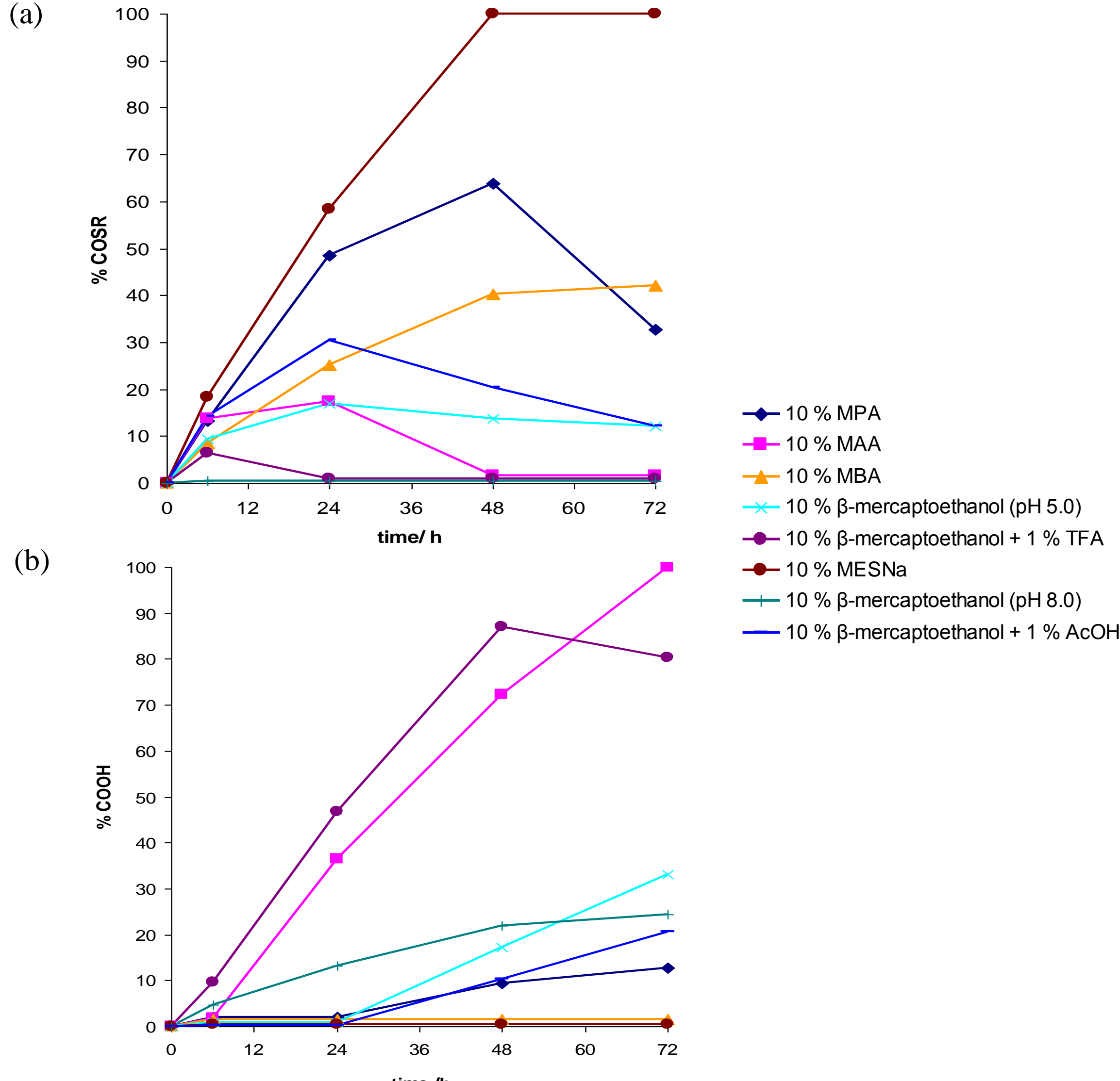


Figure 3.4 Peptide thioesterification of H-AENITTG(^{13}*C*-1)C-NH_2 **26** was monitored by ^{13}C NMR spectroscopy (125 MHz, D_2O, 256 scans) under various reaction conditions. Reactions were conducted at 60 °C using conventional heating and 6 mM labelled model peptide **26**, 10 % thiol in D_2O ± acid, between pH 1.0 and 8.0. The relative intensities of the ^{13}C-labelled carbonyl carbons were then used from the ^{13}C NMR spectra acquired to determine the percentage conversion of the starting material **26** into the product: (a) peptide thioester; (b) hydrolysed peptide thioester.[267]

3.2.2 Optimisation Studies Using Conventional Heating

Key observations from the optimisation studies conducted using conventional heating and relating to Figure 3.4, are discussed in Table 3.2.

Table 3.2. Key observations from the optimisation studies conducted by reacting the labelled model peptide H-AENITTG(^{13}C-1)C-NH_2 **26** under various reaction conditions at 60 °C by conventional heating, and relating to Figure 3.4 are described.

Conditions	Observations
10 % MPA (pH ~ 1.0)	The maximum amount of peptide MPA thioester: H-AENITTG-$SCH_2CH_2CO_2H$ (200 ppm) attained was about 60 % after 48 h. After 48 h, thioester hydrolysis (173 ppm) was detected and it increased to approximately 10 % after 72 h.
10 % MAA (pH 1.0-2.0)	The maximum amount of MAA thioester: H-AENITTG-SCH_2CO_2H detected was approximately 20 % after 24 h. After 72 h, 100 % of the starting peptide had been hydrolysed to H-AENITTG-OH.
10 % MBA (pH ~ 2.0)	No peptide thioester hydrolysis was observed. Peptide-MBA thioester H-AENITTG-$SCH_2CH_2CH_2CO_2H$ formation reached approximately 40 % conversion after 48 h. The side-reaction of MBA thiolactone (210 ppm) precipitate formation was observed to increase over time (Figure 3.5).
10 % BME, 1 % TFA (pH ~ 1.0)	Thioester hydrolysis rapidly increased throughout the reaction to about 80 % after 72 h. TFA appeared to accelerate acid-mediated peptide thioester hydrolysis.
10 % BME, 1 % AcOH (pH 2.0)	More thioester was detected under these conditions than in the reaction containing TFA, but peptide thioester hydrolysis was about 20 % after 72 h.
10 % BME (pH 5.0)	The hydrolysed product increased to about 30 % after 72 h.
10 % BME (pH 8.0)	No peptide-SCH_2CH_2OH BME thioester formation was observed. Base-mediated peptide thioester hydrolysis was observed to be more than 20 % after 72 h. The reactions conducted with 10 % BME at pH 1.0, 2.0, 5.0, or 8.0, all

	resulted in hydrolysis.
10 % MESNa, 1 % AcOH (pH 2.0)	100 % Conversion of the starting material into the product-$SCH_2CH_2SO_3H$ thioester was detected after 48 h. No peptide thioester hydrolysis was observed.
10 % AcOH (pH 2.0)	No peptide thioester hydrolysis as a result of breakdown of the thioester from an *N→S* acyl shift was observed.
1 % TFA (pH 0.8)	The "*S*-peptide" from an *N→S* acyl shift was observed at 199 ppm. About 70 % of the hydrolysed product was detected after 72 h.
71 % TFA-*d* + 29 % $CDCl_3$	The *S*-peptide was present at room temperature[236] but it was observed after 96 h, when 0.5 % TCEP.HCl had been added which suggested that intramolecular disulfide bond formation had prevented the *N→S* acyl shift from occurring.

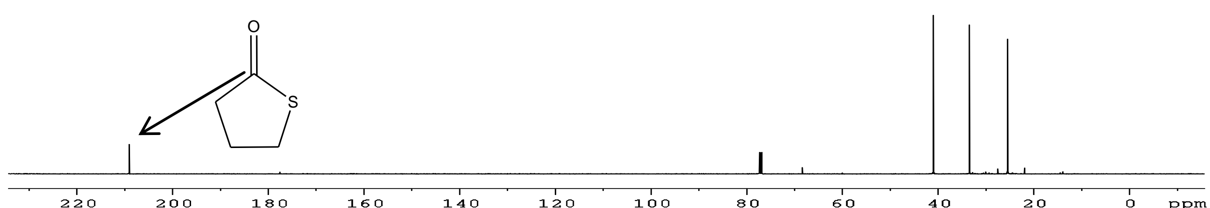


Figure 3.5 ^{13}C NMR spectrum of the undesirable 4-mercaptobutyric acid (MBA) thiolactone with the characteristic chemical shift of ~210 ppm was isolated upon attempted distillation of the desired linear MBA. MBA thiolactone formation was also detected in ^{13}C NMR spectra monitoring the thioesterification of peptide **26** with 10 % MBA in D_2O at 60 °C.

3.2.12 Other Observations from Conventional Heating Experiments

Overall, the rate of peptide thioester hydrolysis with 10 % MAA was significantly higher than with 10 % MPA, and no hydrolysis was observed with 10 % MBA. For this observation we proposed a 'neighbouring group participation' hypothesis for mercaptocarboxylic acids, which stated that the intramolecular acid functionality within these molecules can facilitate hydrolysis once the peptide thioester had been formed *via* anhydride formation[269] (Scheme 3.5).

The proposed anhydride has not been isolated, nor has it been detected by ^{13}C NMR. Neighbouring group participation was proposed as an explanation for the increased rates of

hydrolysis observed with MAA and MPA thioesters, which are proposed to undergo anhydride formation *via* a five- or six-membered ring transition state. It follows that the absence of detectable hydrolysis in the reaction containing 10 % MBA could be attributed to decreased neighbouring group participation due to the slower formation of the larger seven-membered ring transition state. Peptide-BME thioesters were also shown to hydrolyse, which can also be explained by participation of the neighbouring alcohol group to form an ester that subsequently hydrolyses. The increased stability of the sulfonate group within MESNa relative to the carboxylic acid group of MPA, could also account for the absence of detectable hydrolysis when MESNa was used as the thiol.

Scheme 3.5 Neighbouring group participation to form anhydrides[269] during peptide thioesterification is proposed to contribute to the trend observed for peptide thioester hydrolysis: MAA> MPA> MBA. The MAA thioester is proposed to undergo anhydride formation *via* a five-membered ring transition state; the MPA thioester is shown to undergo anhydride formation *via* a six-membered ring transition state; and the MBA thioester is proposed to undergo anhydride formation *via* a large seven-membered ring transition state.

Therefore, from these optimisation studies conducted with conventional heating, sodium 2-mercaptoethanesulfonate (MESNa) was found to be a superior transthioesterification reagent to MPA, whilst MPA was observed to be a better reagent than 4-mercaptobutyric acid (MBA), 2-mercaptoacetic acid (MAA), and β-mercaptoethanol (BME).[267] The increased polarity of MESNa compared to MPA also greatly simplified HPLC purification.[267] MESNa was found to elute within the first 5 min during HPLC purification and remained water-soluble throughout the course of the reaction without precipitation.

Overall, the highest amount of hydrolysis was observed with 10 % BME + 1 % TFA at pH 0.8 (about 80 % hydrolysed after 72 h), and also with 10 % MAA at pH 1.6 (100 % hydrolysed after 72 h).[267] It appeared that over time peptide thioesterification became significantly disfavoured compared to peptide hydrolysis at pH lower than 2.0. No

background *S*-peptide hydrolysis was observed with 1-10 % aqueous AcOH alone (pH 2.0), however it was observed with 1 % aqueous TFA (pH 0.8).

The percentage conversions of the starting material **26** into the respective thioesters **29** under various thioesterification conditions using conventional heating, which were used to plot Figure 3.4a, are shown in Table 3.3. These percentages were used to determine the extent of peptide thioesterification at each time point (Table 3.4, Figure 3.6). In general, the extent of thioesterification of the labelled peptide **26** (H-AENITTG(^{13}C-1)C-NH_2) with an added thiol using conventional heating was observed to increase until 6 h, after which it decreased until the reaction was terminated after 72 h (Figure 3.6). Thus, the majority of the transformation of the starting material **26** into the thioester **29** (H-AENITTG(^{13}C-1)-SR) occurred within the first 6 h of the reaction.

Table 3.3 Percentage conversions of the starting labelled peptide **26** (H-AENITTG(^{13}C-1)C-NH_2) into the peptide thioester **29** (H-AENITTG(^{13}C-1)-SR) under various reaction conditions.[a] These percentages were used to plot Figure 3.4a.

Reaction conditions	% COSR at t = 0 h	% COSR at t = 6 h	% COSR at t = 24 h	% COSR at t = 48 h	% COSR at t = 72 h
10 % MPA	0.00	13.22	48.6	63.78	32.69
10 % MAA	0.00	13.64	17.48	0.00	0.00
10 % MBA	0.00	8.40	25.09	40.25	42.14
10 % β-mercaptoethanol (pH 5.0)	0.00	9.29	16.85	13.67	12.13
10 % β-mercaptoethanol + 1 % TFA	0.00	6.51	0.00	0.00	0.00
10 % MESNa	0.00	18.37	58.25	100.00	100.00
10 % β-mercaptoethanol (pH 8.0)	0.00	0.00	0.00	0.00	0.00
10 % β-mercaptoethanol + 1 % AcOH	0.00	14.22	30.5	20.47	12.06

[a]The relative intensities of the ^{13}C-labelled carbonyl carbons in the ^{13}C NMR spectra acquired to follow the conversion of **26** (H-AENITTG(^{13}C-1)C-NH_2) into **29** (H-AENITTG(^{13}C-1)-SR) over time, were used to determine the percentage conversions (% COSR).

Table 3.4 The extent of conversion of the starting labelled peptide **26** (H-AENITTG(^{13}C-1)C-NH_2) into the peptide thioester **29** (H-AENITTG(^{13}C-1)-SR) under various reaction conditions, was determined using the percentages shown in Table 3.3.[a]

Reaction conditions	t = 0 h	t = 6 h	t = 24 h	t = 48 h	t = 72 h
10 % MPA	0.00	2.20	2.03	1.33	0.45
10 % MAA	0.00	2.27	0.73	0.00	0.00
10 % MBA	0.00	1.40	1.05	0.84	0.59
10 % β-mercaptoethanol (pH 5.0)	0.00	1.55	0.70	0.28	0.17
10 % β-mercaptoethanol + 1 % TFA	0.00	1.09	0.00	0.00	0.00
10 % MESNa	0.00	3.06	2.43	2.08	1.39
10 % β-mercaptoethanol (pH 8.0)	0.00	0.00	0.00	0.00	0.00
10 % β-mercaptoethanol + 1 % AcOH	0.00	2.37	1.27	0.43	0.17

[a] For each reaction condition, the percentage of thioester formed (from Table 3.3) at, for example, 24 h was divided by the time (24 h), and this ratio was plotted against that time point (i.e. 24 h) in Figure 3.6.

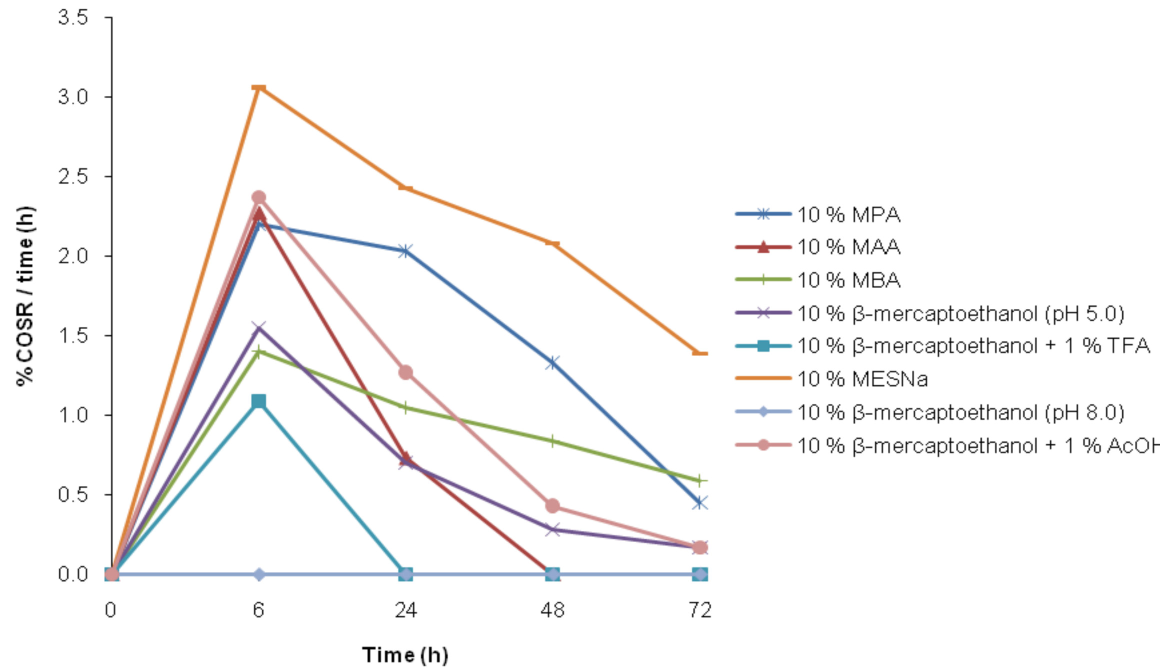


Figure 3.6 Extent of peptide thioesterification. The percentage conversions of the starting labelled peptide **26** relative to time = 0 h were taken from Table 3.4.

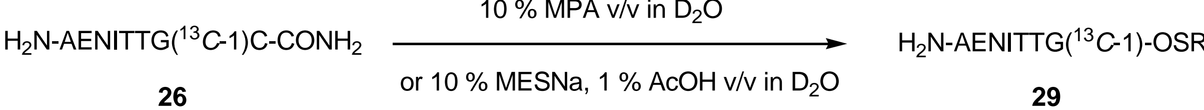
3.2.15 Optimisation Studies Using Microwave Heating

The two best thiols from the initial optimisation studies using conventional heating were found to be MESNa and MPA,[267] and these were then further investigated in microwave-mediated reactions in an attempt to increase the reaction rate (Figure 3.7).[267]

Microwave-mediated thioesterification was expected to be useful if the reaction rate, when 10 % MESNa and 1 % AcOH were used as the reagents, increased to result in 100 % conversion of the starting material into the product thioester within 24 h. Therefore, labelled peptide **26** was reacted with either 10 % MESNa and 1 % AcOH, or 10 % MPA at three different temperatures 40, 50 and 60 °C, and the reactions were monitored at regular intervals over 24 h (Figure 3.7).[267] Microwave-mediated thioesterification reactions were conducted with low magnetic stirring and were also monitored by ^{13}C NMR spectroscopy.

However, microwave irradiation was found to offer no advantage to peptide thioesterification since the results were comparable to those obtained with conventional heating (Figures 3.4a and 3.5).[267] The percentage conversion into the peptide thioester with 10 % MESNa in the presence of 1 % AcOH was found to be approximately 60 % in either case, when thioesterification was conducted at 60 °C after 24 h, with either microwave or conventional heating.[267]

The role of certain additives was also explored and the addition of 0.5 % w/v TCEP was observed to be beneficial. This was concluded after it was found that microwave-mediated thioesterification with 10 % MESNa and 1 % AcOH at 50 °C resulted in approximately 50 % of the peptide thioester, but upon addition of 0.5 % TCEP the amount of the peptide thioester was noticeably higher than 50 % after the same duration of time (Figure 3.7).[267] This was attributed to the reduction of unproductive intermolecular disulfide bonds formed during the reaction between the starting material and the thiol, since the reaction contained 0.7 M MESNa.[267]



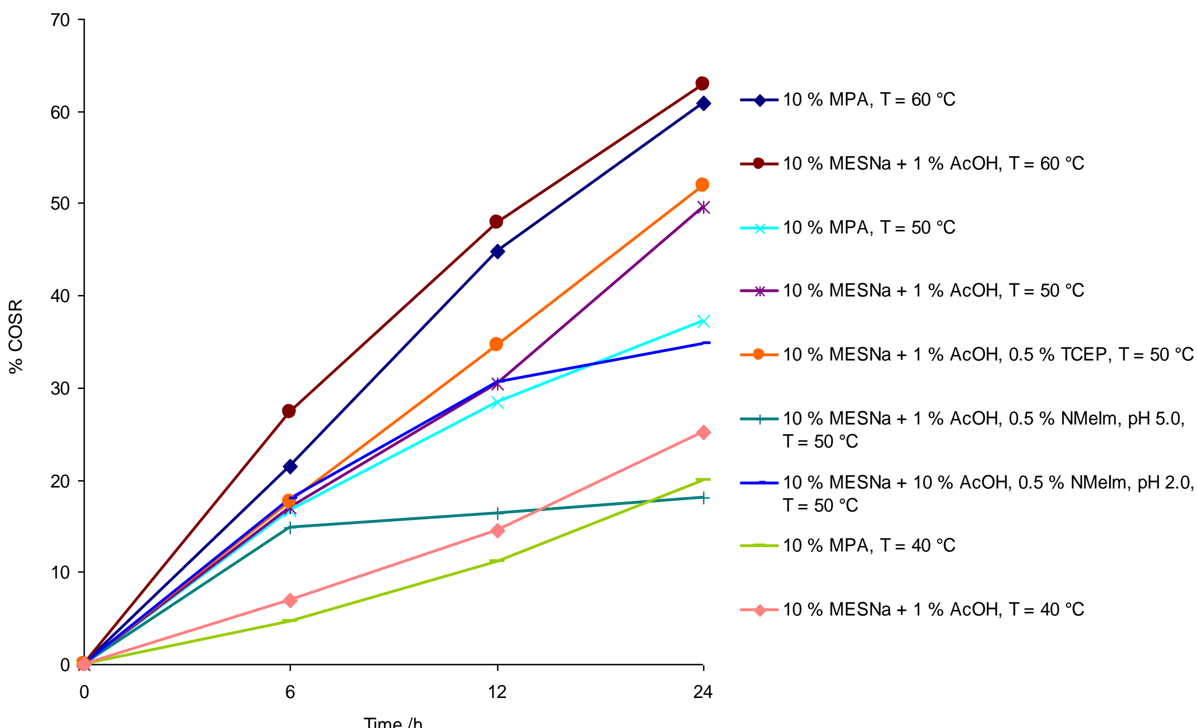


Figure 3.7 Peptide thioesterification of the labelled peptide H-AENITTG(^{13}C-1)C-NH_2 **26** was conducted using microwave irradiation and was monitored by ^{13}C NMR spectroscopy (125 MHz, D_2O, 256 scans), under various reaction conditions. Reactions were conducted with 6 mM labelled peptide **26** in D_2O, 10 % MPA or MESNa, at 40/ 50/ 60 °C.[267] The relative intensities of the ^{13}C-labelled carbonyl carbons were then used from the ^{13}C NMR spectra acquired, to determine the percentage conversion of the starting material **26** into the product peptide thioester **29**.[267]

The role of the general base *N*-methyl imidazole (Me-Im) was also investigated. Microwave-mediated thioesterification at 50 °C of the starting peptide **26** with 10 % MESNa and 1 % AcOH, 0.5 % Me-Im; pH 5.0, resulted in approximately 20 % thioester formation. Whereas identical reaction conditions but with the pH adjusted to 2.0 by the addition of 10 % AcOH resulted in greater than 30 % thioester formation. The same reaction conducted at pH 2.0 without Me-Im as the additive resulted in about 50 % thioester. Thus, both of the reactions

with 0.5 % Me-Im, at either pH 2.0 or 5.0, produced less thioester than without this additive (Figure 3.7).[267] This suggested that the addition of a general base to deprotonate the cysteine sulfhydryl group at the glycine-cysteine junction within the starting peptide, for example, was not required.

Peptide thioesterification was also conducted at the lower temperature of 40 °C and though thioester formation occurred, it was slower than that observed at 50 °C or 60 °C (Figure 3.7).[267] Nonetheless thioesterification at 40 °C may prove useful for some peptides and proteins that contain heat-sensitive moieties.

3.2.13 Summary of Optimisation Studies

All of the optimisation studies were carried out using a starting peptide concentration of 6 mM.[267] Data supplied by Dr Derek Macmillan showed that a higher concentration of the starting peptide (16 mM) resulted in less thioester formation than when the reaction was conducted with a 4 mM concentration of the peptide.[267] Hence, the peptide concentration was important since a lower peptide concentration resulted in a higher amount of thioester formation. This was likely due to an increased propensity of disulfide bond formation with the thiol at higher peptide concentrations, which was supported by the beneficial effect of the addition of the reducing agent TCEP.[267]

In summary, the optimal conditions for peptide thioesterification from these studies were deduced to be the following: 6 mM [peptide], 10 % MESNa, 1 % AcOH, 0.5 % TCEP, heating at 60 °C at pH 2.0.[267] These optimisation studies were conducted in deuterated water, however replacement of the reaction solvent from water to 6 M guanidine hydrochloride has been observed to be beneficial for insoluble samples.[267] Having deduced optimised peptide thioesterification conditions, they were then applied to the total chemical synthesis of a biologically active protein.[267]

3.3 The First Biological Application of Native Chemical Thioesterification: Human *β*-Defensin 3 (HBD3)

3.3.1 Introduction

Defensins are characteristically small proteins with molecular weights ranging from 2-5 kDa.[270,271] Defensins are cationic, due to a net charge calculated from the total sum of the charged amino acids at neutral pH, and have also been described as "cysteine-rich" peptides. Defensins are found in mammalian leukocytes and epithelial tissues and based on their disulfide connectivity are categorised into three different structural families: *α, β,* and *θ*. In 2009, six human *α*-defensins: human neutrophil peptides (HNPs 1-4) and enteric defensins (HD5 and HD6) were identified. However, a larger number of *β*-defensins are believed to exist, predominantly in epithelia.

Distinct from *α* and *β*-defensins, *θ*-defensins are macrocyclic and are expressed in leukocytes and the bone marrow of specific non-human primates.[270,271] Mammalian *α*- and *β*-defensins differ in their amino acid compositions, cysteine disulfide topology, and tissue distribution, but their three-dimensional structure is conserved. They adopt a three-stranded *β*-sheet core structure that is stabilised by three intramolecular disulfide bonds. Despite their overall structural similarity, defensins exhibit rather diverse functions.

Defensins are natural peptide antibiotics and they operate in the early stages of the innate immune defence system against invading infectious pathogens including: bacteria, fungi, and viruses.[270,271] The cationic property of defensins is generally acknowledged as being responsible for killing bacteria due to the electrostatic attraction between the cationic defensins and the anionic microbial membrane. This results in cell death due to the disturbance of the microbial membrane which leads to leakage of cell contents. Defensins inhibit enveloped and non-enveloped viruses through diverse molecular mechanisms.

The synthesis of an analogue of human *β*-defensin 3 (HBD3) was undertaken, which is comprised of 45 amino acid residues and is an antimicrobial and chemoattractant polypeptide (Figure 3.8).[270,271] HBD3 contains six cysteine residues which form three intramolecular

disulfide bonds in the characteristic β-defensin arrangement: Cys^{I}-Cys^{V}, Cys^{II}-Cys^{IV}, and Cys^{III}-Cys^{VI}.[270,271]

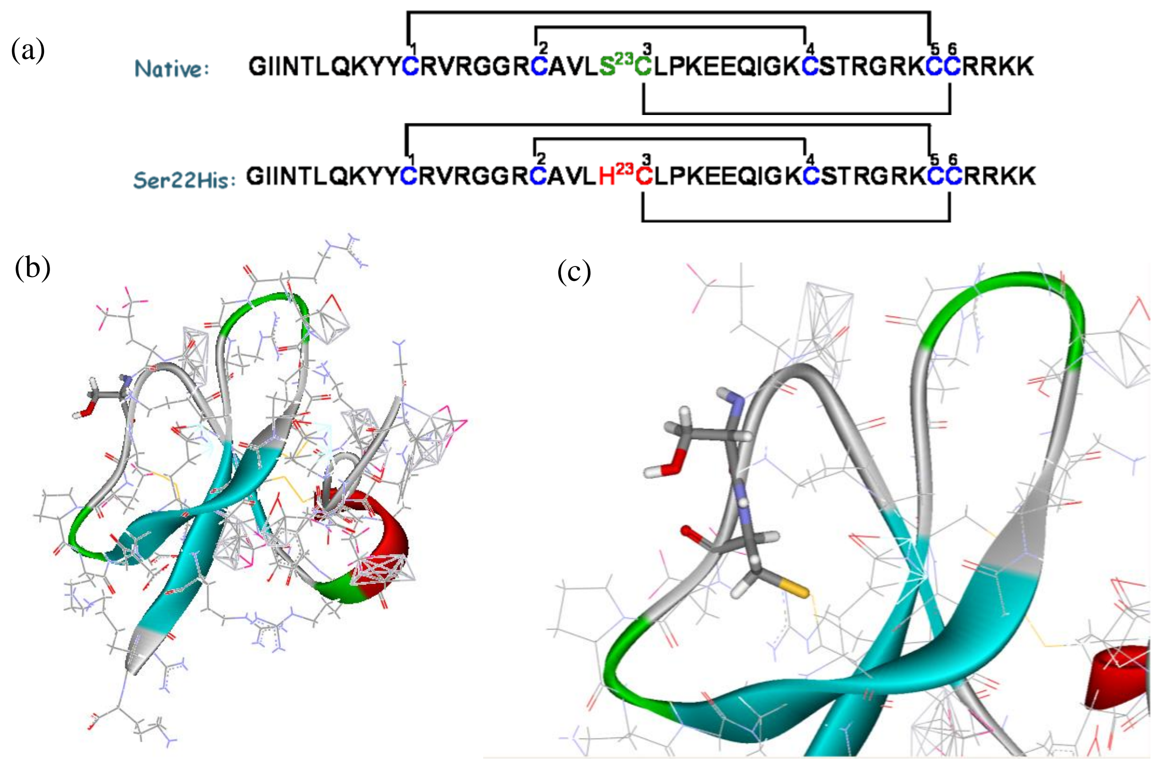


Figure 3.8 (a) Amino acid sequence of human β-defensin 3 (HBD3) and the Ser22His analogue. (b) Known NMR structure of HBD3. (c) The native ^{22}Ser-^{23}Cys site is located on the surface of HBD3.[270,271]

3.6.2 Results and Discussion

3.3.3 Total Chemical Synthesis of the HBD3 Analogue

The synthetic strategy for the total chemical synthesis of HBD3 involved the Fmoc-based SPPS of two "halves" of the full-length HBD3 polypeptide: HBD3(1-22) and HBD3(23-45) (Scheme 3.6). The N-terminal fragment was to be functionalised with a C-terminal thioester and the C-terminal fragment would have an N-terminal cysteine. Histidine-cysteine

sequences were found to be particularly prone to thioesterification (Chapter 2),[263] therefore a histidine-cysteine motif could simply appended to the N-terminal HBD3 fragment by introducing a Ser22His mutation. Each half of HBD3 could then be ligated through an NCL reaction,[144] and leave an amide bond at the ligation site.

Scheme 3.6 The synthesis of the folded analogue of HBD3: Ser22His HBD3(1-45) **36**.[267] Ser22His HBD3 was assembled from two synthetic 23 amino acid peptide "halves". Ser22His HBD3(1-21)-His-Cys-NH_2 **30** was initially thioesterified using the optimised conditions to afford a 22 amino acid thioester: Ser22His HBD3(1-21)-His-$SCH_2CH_2SO_3H$ **31**, which was then ligated to the cysteinyl peptide: HBD3(23-45) **33** to afford **34**. The acetamidomethyl cysteine side-chain protecting groups present in the full-length polypeptide **34** were then removed to give deprotected polypeptide **35**, which was then oxidatively folded to afford **36** for subsequent biological evaluation.[267]

3.3.4 Synthesis of the Ser22His HBD3 N-Terminal Peptide

Theoretically, HBD3 can be synthesised in a single SPPS run, however HBD3 was prepared from two fragments to enable the demonstration of the biocompatibility of the thioesterification method (Scheme 3.6).[267] The N-terminal HBD3 fragment was assembled after the initial manual coupling of Fmoc-Cys(Trt)-OH onto Rink Amide MBHA resin. The peptide was then elongated in an automated fashion and purified by (RP)HPLC to afford the N-terminal HBD3(1-21)-His-Cys-NH_2 peptide **30** in 50 % isolated yield, based on 100 % loading of the first amino acid onto the resin.

The N-terminal HBD3 fragment **30** had two other internal cysteine residues, apart from the cysteine within the C-terminal histidine-cysteine sequence, which were protected with an acetamidomethyl (Acm) group. These other cysteine residues that were Acm protected formed part of either a $^{10}Y^{11}C$ or $^{17}R^{18}C$ sequence, therefore side-reactions at these sequences were expected to be minimal during the thioesterification step.[263] However, the internal cysteine residues were anticipated to decrease the rate of the initial transthioesterification step in NCL and potentially render thioester hydrolysis a significant side-reaction. Therefore, internal cysteine residues within both the N- and C-terminal fragments of HBD3 were Acm protected.

The solid-phase peptide synthesis of **30** was notably efficient and provided mainly the desired peptide carboxamide (Figure 3.9). LC-MS analysis of the isolated N-terminal HBD3 peptide **30** confirmed the correct mass of the desired peptide.

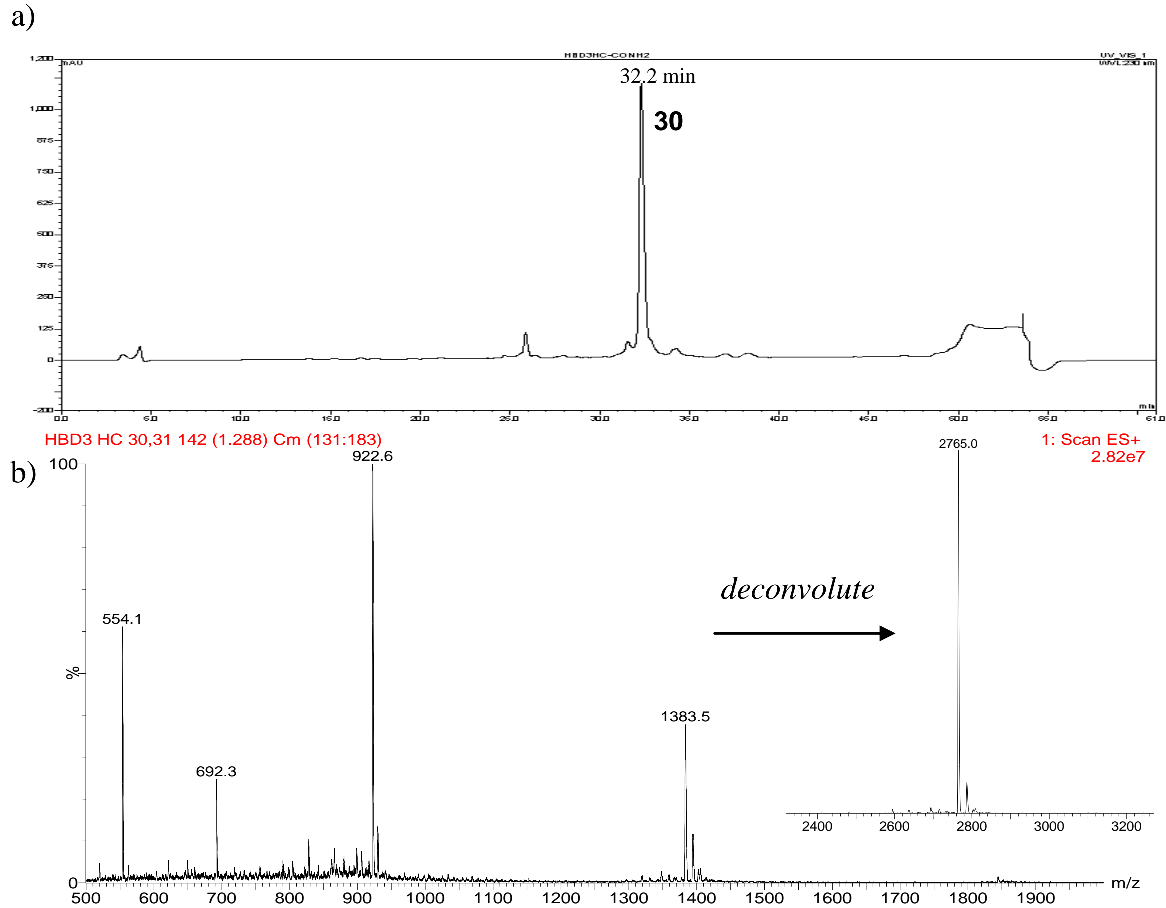


Figure 3.9 (a) (RP)HPLC purification chromatogram of **30**, absorbance detected at 230 nm. The isolated N-terminal HBD3 peptide **30** was found to have a retention time of t_R = 32.3 min, using the gradient: 5-50 % MeCN over 45 min. (b) LC-MS analysis of isolated **30**. Calculated average mass for [M]: 2765.3, found: 2765.0.[267]

3.3.5 Synthesis of the Ser22His HBD3 C-Terminal Cysteinyl Peptide

The C-terminal HBD3(23-45) peptide **33** was synthesised on preloaded Fmoc-Lys(Boc)-Novasyn® TGT resin **32** and extended using automated Fmoc-SPPS. After cleavage and (RP)HPLC purification, **33** was isolated in 37 % yield.[267] Synthetic C-terminal HBD3(23-45) peptide **33** contained an equal number of amino acids to the N-terminal HBD3 peptide **30** (23 amino acids each). However **33**, which contained three Acm protected cysteine residues, was considerably earlier eluting during (RP)HPLC purification than **30** (Figure 3.10). This was

attributed to the presence of the majority of the polar, positively charged (arginine and lysine) and negatively charged (glutamate), amino acid residues residing within this "half".

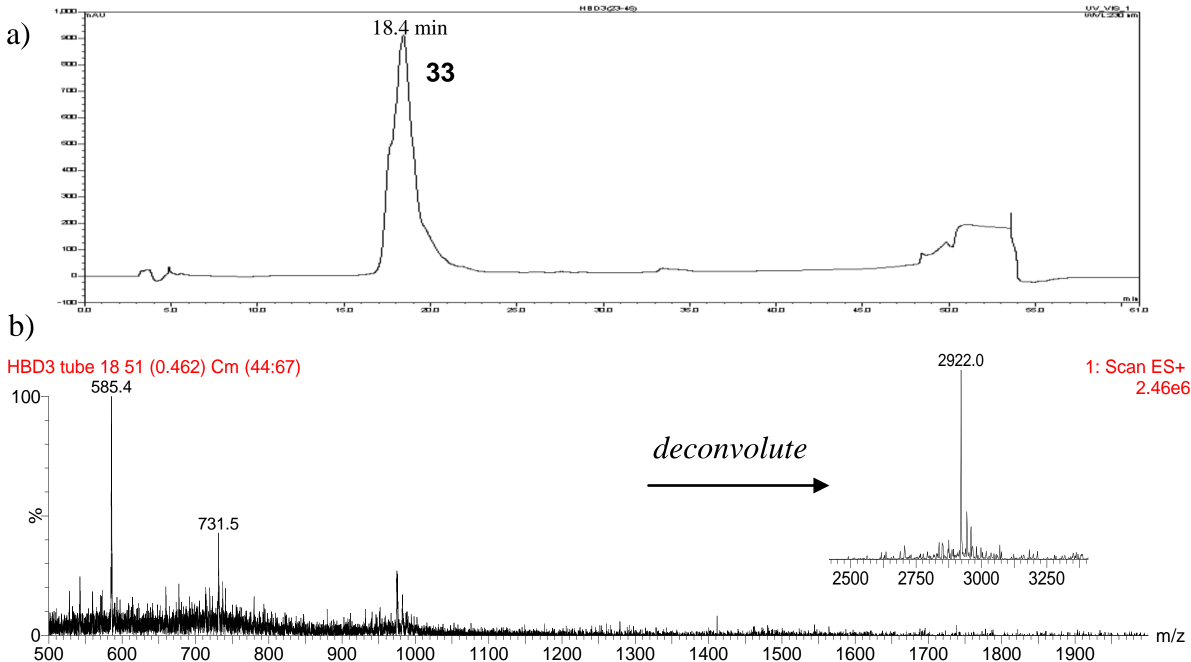


Figure 3.10 (a) (RP)HPLC purification chromatogram of **33**, absorbance detected at 230 nm. The isolated HBD3 cysteinyl peptide **33** was found to have a retention time of t_R = 18.4 min, using the gradient: 5-50 % MeCN over 45 min. (b) LC-MS analysis of isolated **33**. Calculated average mass for [M]: 2921.5, found: 2922.0.[267]

3.3.6 Synthesis of the N-Terminal HBD3 Thioester

Using the optimised thioesterification conditions (Section 3.2.16),[267] the more complex 22 amino acid thioester **31** was synthesised.[267] Thioesterification of 15.8 mg of the histidine-cysteine terminating N-terminal peptide **30** was conducted at a concentration of 4 mM in 6 M guanidine hydrochloride, which contained 10 % w/v MESNa, 10 % v/v acetic acid, and 0.5 % w/v TCEP, and was reacted for a period of 48 h at 60 °C.[267] The desired 22 amino acid thioester **31** was co-purified with the 23 amino acid starting material **30** by semi-preparative (RP)HPLC purification of the reaction mixture. These two species could not be separated, even if the solvent gradient during (RP)HPLC purification was modified. However, it was

reasoned that the starting material **30** would not participate in the subsequent NCL reaction and could therefore be recovered after ligation. Thus, the lyophilised product which was comprised of **30** and **31** was used in the NCL reaction.

MESNa-mediated thioesterification of the longer 23 amino acid peptide **30** did not proceed to completion. It appeared as if the impetus for thioester formation was higher in the initial stages of the reaction, with a slower rate of conversion into the product thioester after the initial 6 h, analogous to that observed in the optimisation studies (Figure 3.6).

An attempt was also made to differentiate between the elution times of the starting material **30** and the product thioester **31** by extending the number of amino acids within the peptide chain of **30**. Hence, to allow for a better separation of the starting material N-terminal HBD3 peptide and the product thioester, the N-terminal HBD3 fragment was synthesised with four extra C-terminal amino acids: HBD3(1-21)-HC<u>SSGG</u>-NH_2.

The isolated yield of the longer N-terminal peptide after Fmoc-SPPS was lower than that of the original peptide **30** which did not contain the extra four residues. Nonetheless, the isolated peptide was subsequently subjected to the same optimised thioesterification conditions to afford two distinguishable peaks that again co-eluted though to a lesser extent (Figure 3.11b). Overall, the loss in yield for the synthesis of this new N-terminal peptide bearing an extra four amino acids compared to the simpler histidine-cysteine terminating starting peptide **30** was considered too great, and the resulting resolution of the product thioester after thioesterification insufficient, to be useful. Also, separation of the two species was not ultimately required because **30** could be recovered after NCL.

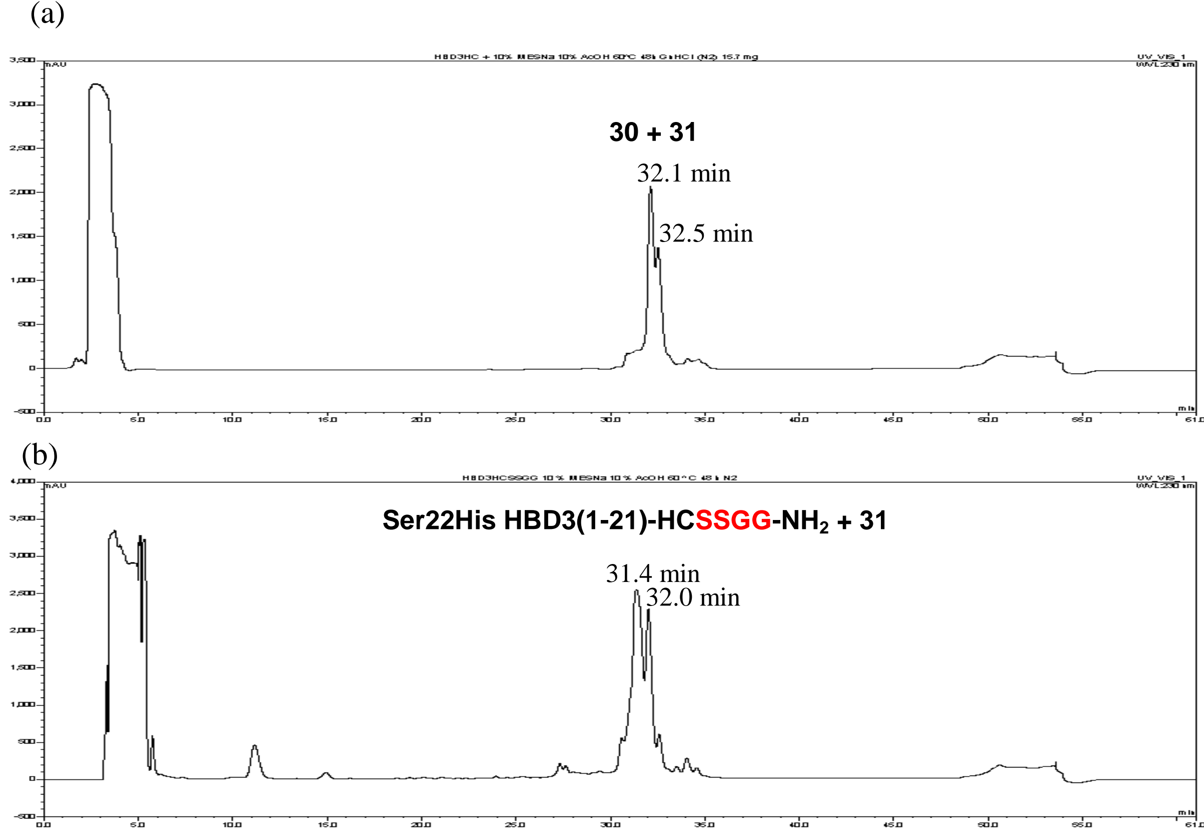


Figure 3.11 (RP)HPLC purification traces compared after thioesterification of two different N-terminal HBD3 peptides, absorbance detected at 230 nm. (a) The co-eluting N-terminal Ser22His HBD3(1-21)-HC-NH_2 terminating peptide **30** and product thioester Ser22His HBD3(1-21)-H-$SCH_2CH_2SO_3H$ **31**. (b) The co-eluting N-terminal Ser22His HBD3(1-21)-HCSSGG-NH_2 terminating peptide with four extra amino acids and the product thioester **31**.[267]

Multi-milligram MESNa-mediated thioesterification reactions conducted under a positive nitrogen atmosphere produced the most efficient conversion into the desired product thioester **31**, and were also coupled with the least amount of undesirable disulfide bond formation to give HBD3(1-21)-His-Cys(MESNa)-NH_2.[267] It was generally observed that multi-milligram reactions (approximately 10 mg scale) provided sufficient quantities of the co-eluting starting material and thioester for subsequent ligation.

3.3.7 Synthesis of Full-Length Protected Ser22His HBD3(1-45) by NCL

The synthetic peptide thioester **31** (which was co-purified with **30**) and the cysteinyl peptide **33** were then ligated. An initial NCL was successful though the ligated product had to be re-purified employing a shallower gradient of aqueous acetonitrile, to afford the homogenous full-length HBD3. Generally, it was observed that the larger the NCL reaction scale, the larger the imposition of the ligated product peak onto the starting material **30** peak during (RP)HPLC purification. Hence, the scale of the NCL reaction needed to be large enough to expect a multi-milligram yield but also result in minimal peak overlap during HPLC purification.

Therefore, multi-milligram scale NCL reactions were simultaneously conducted. The co-purified peptide thioester **31** (10.6 mg) was ligated to excess cysteinyl peptide **33** (11.1 mg) under standard NCL conditions: 3.8 mM [assumed peptide thioester] in 6 M guanidine hydrochloride, 200 mM sodium phosphate buffer; pH 7.0, 100 mM MESNa, 40 mM TCEP, for 48 h at 25 °C. The reaction was not exposed to air during the ligation until after 48 h when it was analysed by LC-MS, which revealed the presence of the ligated product **34**. The reaction mixture was then purified by (RP)HPLC.

Pleasingly, **34** could be isolated from the unreacted starting material **30** to afford the full-length Acm-protected defensin (6.8 mg). A presumed single equivalent of the cysteinyl peptide **33** was added into the NCL reaction, and the recovered yield of this cysteinyl peptide (6.3 mg) and the unreacted starting material **30** (7.1 mg) allowed the yield of the consumed peptide thioester **31** to be estimated as a modest 22 % (40 % based on the recovered starting material **30**),[267] which is comparable to other less straightforward peptide thioester synthesis methods.[199,200,208,210,214,252-256] This NCL also resulted in an excellent 97 % yield of the ligated product **34** (Figure 3.12).[267]

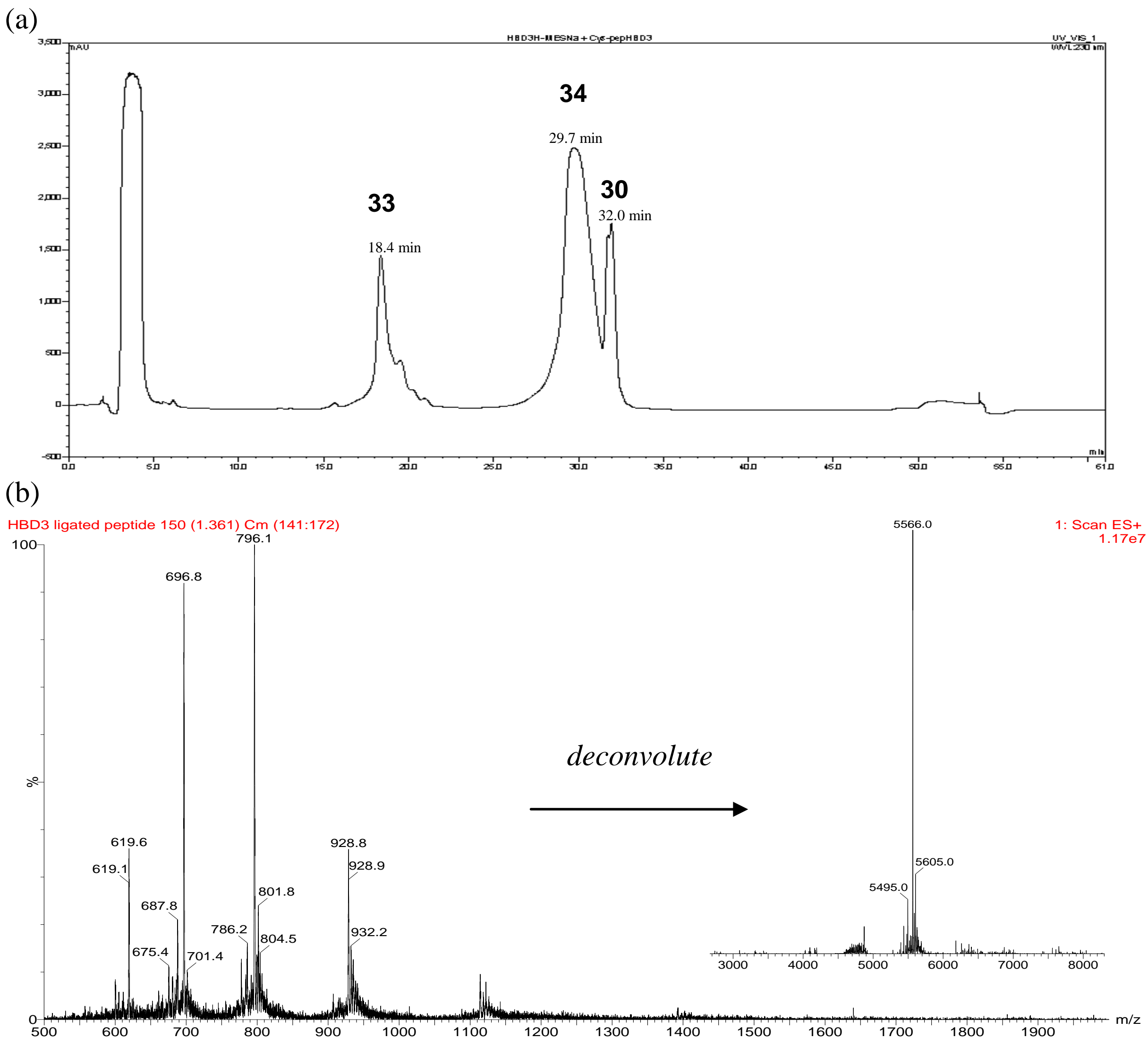


Figure 3.12 (a) (RP)HPLC purification chromatogram of the NCL reaction between co-eluting (starting material **30** and thioester **31**) and cysteinyl peptide **33**, absorbance detected at 230 nm. The isolated ligated full-length HBD3 polypeptide **34** was found to have a retention time of t_R = 29.7 min, using the gradient: 5-50 % MeCN over 45 min. (b) LC-MS analysis of isolated **34**. Calculated average mass for [M]: 5566.6, found: 5566.0.[267]

3.3.8 Synthesis of Full-Length Deprotected Ser22His HBD3(1-45)

Apart from the free cysteine at position-23 of the full-length HBD3(1-45) polypeptide **34**, the remaining cysteine residues were orthogonally protected with an Acm group. The

presence of several free cysteine residues would have hindered the previous NCL step, because the side-chain sulfhydryl groups could all potentially participate in the initial NCL transthioesterification step but then could not subsequently undergo an *S→N* acyl rearrangement. This would be unproductive because only the cysteine at position-23 with a free α-amino group can rearrange through an *S→N* acyl shift to provide the desired ligated product. The reversible nature of the initial transthioesterification step would eventually lead to the correct ligated product, however the slower first step would likely bring into force thioester hydrolysis as a significant side-reaction. Thus, the Acm-protected five remaining cysteine residues within **34** were deprotected after NCL but before protein folding.

Deprotection was achieved using 10 equivalents of silver(I) acetate per each Acm group present within **34** in 10 % v/v acetic acid.[267] Silver acetate was used rather than the standard mercury(II) acetate to avoid the toxic mercury by-product. Acetic acid acted to solubilise the polypeptide and the reaction mixture was shaken for 6 h, before 10 % w/v DTT was added to break down the silver-sulfur salts (Scheme 3.7).

Scheme 3.7 Mechanism of silver acetate (AgOAc)-mediated removal of an acetamidomethyl (Acm) protecting group from the side-chain of cysteine, the silver-sulfur salt produced is broken down by DTT.

The reaction mixture was then centrifuged and the supernatant was concentrated under reduced pressure to remove the aqueous acetic acid. The concentrate was then re-dissolved in water and purified by semi-preparative (RP)HPLC to afford the deprotected defensin, Ser22His HBD3(1-45) **35** in 95 % yield. (Figure 3.13).[267]

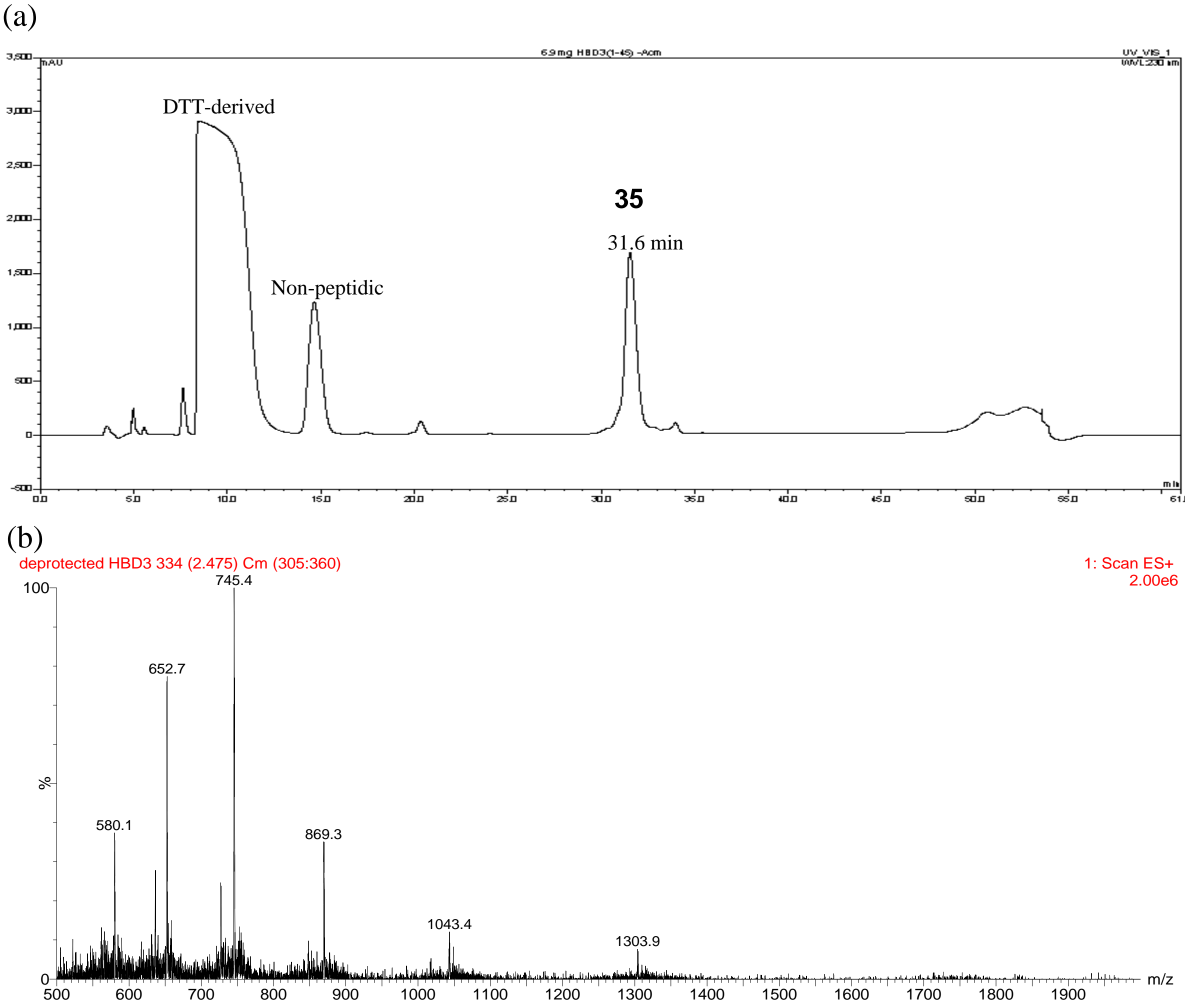


Figure 3.13 (a) (RP)HPLC purification chromatogram of **35**, absorbance detected at 230 nm. The isolated deprotected full-length HBD3 polypeptide **35** was found to have a retention time of t_R = 31.6 min, using the gradient: 5-50 % MeCN over 45 min. (b) LC-MS analysis of isolated **35**. Calculated average mass for [M]: 5211.2, found: 5211.0.[267]

3.3.9 Synthesis of the Folded Ser22His HBD3(1-45) Analogue

The deprotected full-length Ser22His HBD3(1-45) **35** was then oxidatively folded using conditions that were anticipated to form the thermodynamically most stable, native isoform out of a total sum of fifteen possible disulfide connectivities for the defensin.[271]

Initially, protein folding was attempted on a small scale (using 0.6 mg of the deprotected defensin **35**) with 20 % v/v DMSO, 5 % AcOH, ammonium carbonate $(NH_4)_2CO_3$ buffer; pH 6.0, at 25 °C for 7 days.[271] Upon reaction completion, an LC-MS of the reaction mixture confirmed the loss of six daltons that corresponded to the formation of three disulfide bonds. However, (RP)HPLC purification showed the presence of two main oxidised species that corresponded to two different topological isoforms of HBD3 (Figure 3.14).

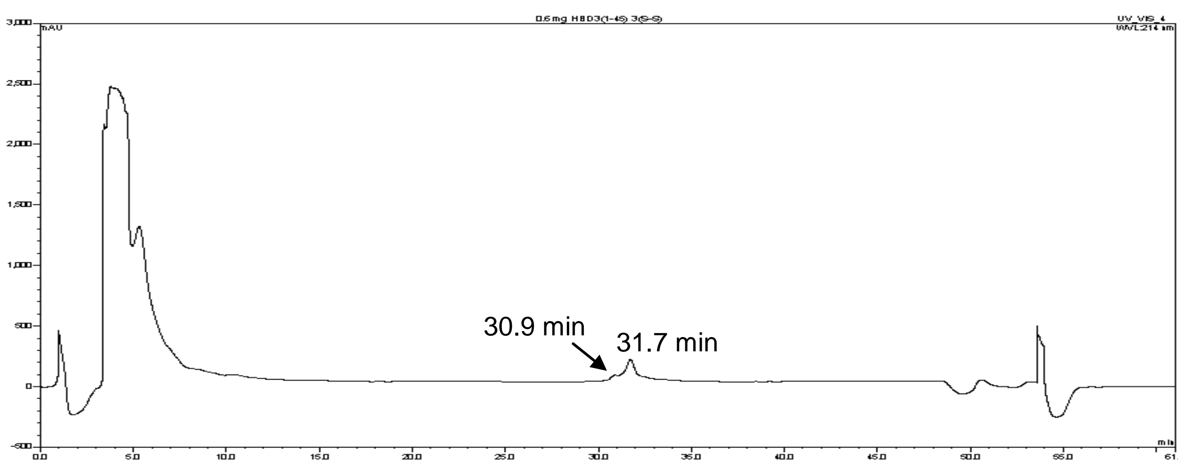


Figure 3.14 (RP)HPLC purification chromatogram of a small-scale folding reaction of Ser22His HBD3 **35** (0.6 mg, 0.12 μmol) conducted using 20 % v/v DMSO as the oxidant at pH 6.0 for 7 days. Two main isoforms with retention times of t_R = 30.9 and 31.7 min were observed, using the gradient: 5-50 % MeCN over 45 min, absorbance detected at 214 nm.

Formation of the three native disulfide bonds present within the *β*-defensin was then attempted with the standard protocol of Wu *et al.*[272] which included a cysteine-cystine redox pair. The oxidative folding conditions included dissolving 10.6 mg of the deprotected peptide **35** in 0.1 M sodium bicarbonate, 1 M guanidine hydrochloride; pH 8.1, 3 mM cysteine, 0.3 mM cystine, with shaking at 25 °C for 24 h; and a final deprotected peptide **35** concentration of 0.25 mg mL^{-1}.[267] Upon reaction completion, the reaction mixture was analysed by LC-MS

to show the expected loss of six daltons which corresponded to the formation of three disulfide bonds. The reaction mixture was then concentrated and purified by semi-preparative (RP)HPLC. The resulting three different oxidised forms of the folded defensin were then combined and lyophilised to afford **36** (Figure 3.15).[267]

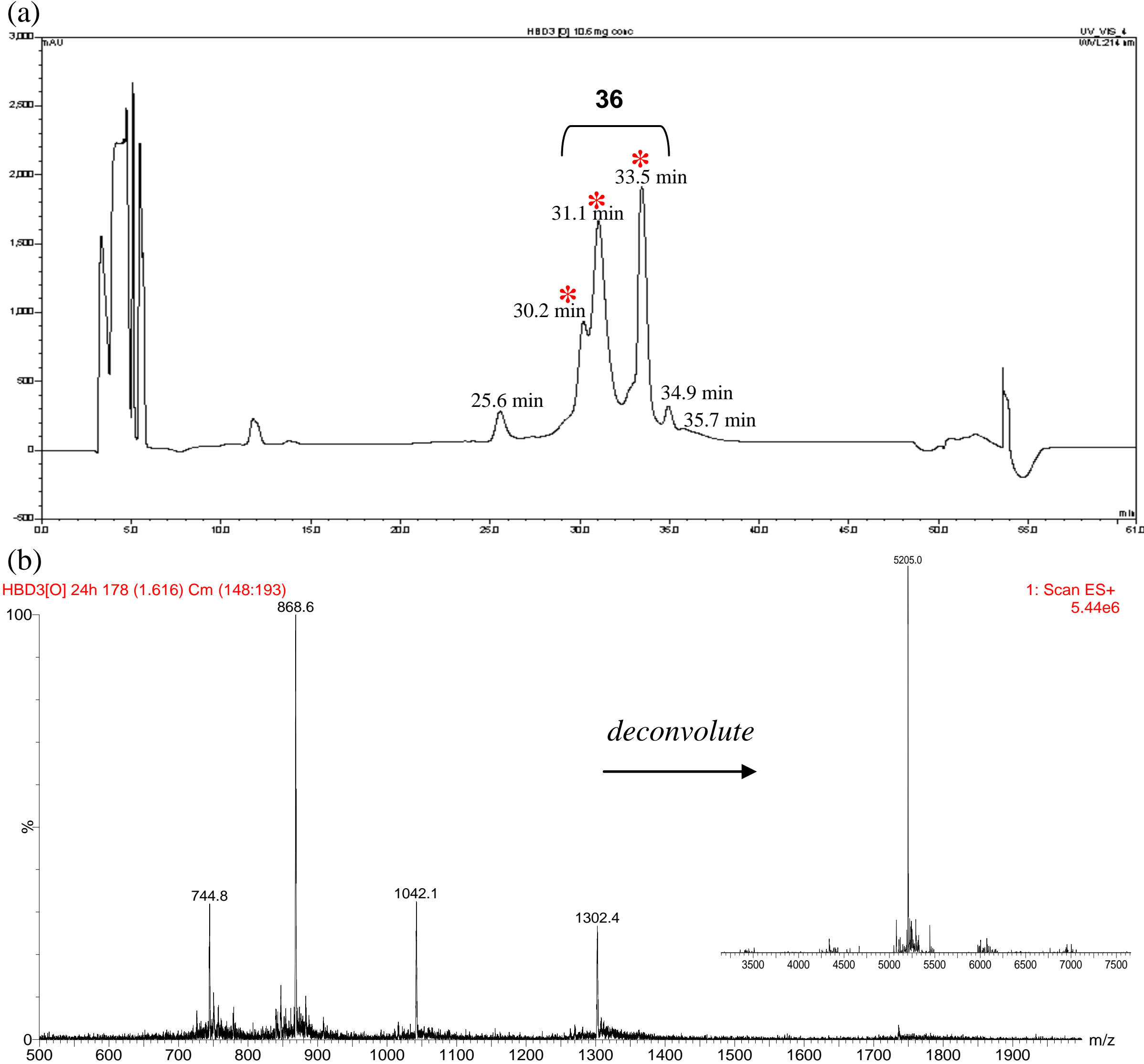


Figure 3.15 (a) (RP)HPLC purification chromatogram of the Ser22His HBD3 folding reaction conducted using **35** (10.6 mg, 2.03 μmol) and the cysteine-cystine redox system at pH 8.1 for 24 h. **36** was pooled from the fractions illustrated by an asterisk, absorbance detected at 230 nm. **36** comprised three main isoforms with retention times of t_R = 30.2, 31.1, and 33.5 min, respectively, using the gradient: 5-50 % MeCN over 45 min. (b) LC-MS analysis of isolated **36**. Calculated average mass for [M]: 5205.2, found: 5205.0.[267]

Protein folding often results in the formation of an earlier eluting compound during (RP)HPLC purification. This is usually attributed to the burial of the hydrophobic amino acid residues within the protein, which prevents them from interacting with the C-18 HPLC column. Interestingly, the three major folded HBD3 species had similar HPLC retention times (30.2-33.5 min) to that of the unfolded species (31.6 min).[267] This could therefore be the effect of exposed patches of hydrophobic amino acids within the folded molecule interacting with the HPLC column media.

The presence of the cysteine-cystine redox couple should have promoted the reversible formation of disulfide bonds until the thermodynamically most stable topological isoform was attained. Wu *et al.* have also found that whilst the β-defensins: HBD1 and HBD2 fold to give a single thermodynamically most stable native species; under the same oxidation conditions however, HBD3 folds to give more than one major topological isomer.[272]

The cysteine-cystine redox system folding procedure resulted in the formation of three major oxidised species, which were pooled to give **36**, and the disulfide connectivities within this pooled mixture were then analysed by tryptic and chymotryptic digestion. However, since the Cys^{V} and Cys^{VI} residues have no intervening amino acid sequence between them for digestion to occur (Figure 3.8a), the disulfide connectivity of the defensin could not be precisely determined. The presence of the Cys^{II}-Cys^{III} disulfide bond within **36** by tryptic digestion was detected by LC-MS (Figure 3.16a), and it was reduced upon addition of DTT and resulted in the loss of two daltons (Figure 3.16b). However, since all topological isoforms of HBD3 have been shown to be biologically active,[272,273] the mixture was not resolved prior to its biological evaluation.

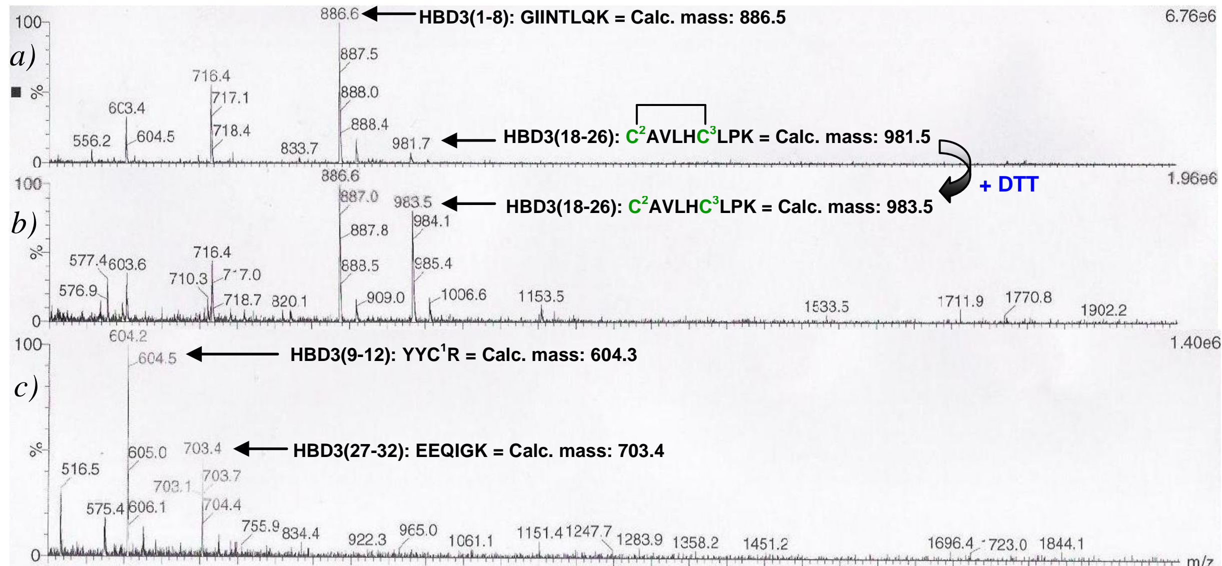


Figure 3.16 LC-MS data for the trypsin enzymatic digestion of the synthesised 45 amino acid residue folded *β*-defensin: Ser22His HBD3(1-45) **36**. ESI$^+$ mass spectra are shown of the polypeptide mapping experiments after treatment of **36** (0.5-1.0 mgmL^{-1}) with trypsin (0.01 mgmL^{-1}) {and DTT (5 % w/v) added after 1 h}; pH 8.0, 37 °C. The reaction was monitored at regular intervals and the resulting peptide fragments of **36** suggested that disulfide bonding between Cys^{II}-Cys^{III} existed. (a) Calculated mass for oxidised Cys^{II}-Cys^{III} [M]: 981.5 Da, observed [M]$^+$: 981.7 Da. (b) After addition of DTT, the calculated mass for reduced Cys^{II}-Cys^{III} [M]: 983.5 Da; observed [M]$^+$: 983.5 Da. The digested sequences were not confirmed by isolation followed by sequence analysis.

3.3.10 Biological Activity

Antimicrobial Assays

Minimum bactericidal concentration (MBC) values, which are the lowest concentrations of peptide required to kill 99.99 % of bacteria, were determined against Gram-negative *Pseudomonas aeruginosa* strain PA01 and Gram-positive *Staphylococcus aureus* strain ATCC 25923 bacteria (see Experimental Section 6 for further details).[267]

Bacterial stocks were stored at -80 °C in either glycerol or skimmed milk. Typically, 800 μL of bacterial cells growing in the exponential phase were frozen with 200 μL of glycerol. The bacterial stock from the freezer was removed and stored on ice. A sterile loop was used to streak out bacteria onto Iso-sensitest agar plates (the media was from Oxoid) and the plates were incubated at 37 °C overnight. The agar plates were comprised of: 4 × (7.85 g agar/ 250 mL dH_2O) and these mixtures were initially autoclaved, and then melted in a microwave. The 4 × 250 mL agar solutions were then poured to provide approximately eighty agar plates.

The following morning, the plates were removed from the incubator and inspected for any contaminating bacterial colonies and then stored at 4 °C. In the early evening, a sterile toothpick was used to pick off a single colony from the plate and it was added to 10.0 mL of Iso-sensitest broth (ISB) (from Oxoid) in a 50.0 mL falcon tube (twice for each organism). The 50.0 mL falcon tube was preferred rather than a 15.0 mL falcon tube as bacteria grow better with a larger air supply. The falcon tubes containing the bacteria were then incubated at 37 °C overnight, with shaking at approximately 150 rpm.

The next morning, bacterial cultures were removed for dilution; 1.0 mL of cells was added to fresh ISB (9.0 mL) for each organism in separate falcon tubes. The resulting suspensions were incubated at 37 °C for 4 h, with shaking at approximately 150 rpm, to afford the bacteria in the desired mid-log phase.

Meanwhile, sterile phosphate buffer (10 mM) was prepared, which was comprised of 1 M K_2HPO_4 (4.01 mL), 1 M KH_2PO_4 (0.99 mL), glucose (0.5 g), and dH_2O (500.0 mL); and 99.0 mL of this solution was removed and ISB (1.0 mL) was added to afford 100.0 mL of phosphate buffered saline (PBS) buffer. This buffer was filtered through a fritted-filter containing syringe into sterile falcon tubes. The defensins were also prepared for testing in doubling dilutions using the PBS buffer. Initially the peptides were diluted to a concentration of 1.0 mg/mL and then to the starting concentration of 1.0×10^{-4} M (100 μM). The starting peptide concentration of 100 μM was then further diluted to afford: 10.0 μM, 5.0 μM, 2.5

μM, 1.25 μM, and 0.625μM, and the negative control 0 μM. Since 100 μL of bacterial cells were to be added to each peptide concentration tested, the starting peptide concentration was ten times that of the required final concentration. The negative control contained PBS buffer only and 100 μL of bacterial cells.

The test peptide concentrations were further corroborated through an absorbance assay. The absorbance of each test peptide that was dissolved at a concentration of 1 $mgmL^{-1}$ was measured using a spectrophotometer. The absorbance (A); molar extinction coefficient (ε, units = $Lmol^{-1}cm^{-1}$) taken from the web server ExPASy; and the sample path length (ℓ = 1 cm) were then inserted into the Beer-Lambert Law (Equation 2) to calculate the actual peptide concentrations (c, units = $molL^{-1}$). These peptide concentrations were then used in the antimicrobial assays to determine the MBCs.

$$A = \varepsilon . c . \ell \qquad (2)$$

After 4 h the organisms were removed from the incubator and centrifuged at 5000 rpm for 20 min. The supernatant was discarded and the cells were resuspended, to remove any residual salts, in PBS (1.0 mL) and broth in an Eppendorf tube, and the cells were bench-top centrifuged at 13000 rpm for 5 min. The supernatant was again discarded and the pellet of cells was resuspended in fresh PBS (1.0 mL) and broth and then centrifuged and decanted again. The bacterial cells were then diluted with PBS and broth in the ratio 1: 1000-2000 in a falcon tube to afford the initial inoculum. The cells (100 μL) were then vortexed and added to randomised test peptides in Eppendorfs five at a time before being vortexed again. The Eppendorfs containing the resulting test mixtures were then briefly centrifuged and then incubated at 37 °C for 3 h, with low level shaking at 50 rpm. As a growth control for each organism, bacteria from the negative control was also incubated at 37 °C for 3 h, with low level shaking at 50 rpm. A second aliquot of 10 μL was removed from the negative control (1×10^6 CFU "colony forming units") and diluted in a 1: 10 ratio with PBS three times to afford 1×10^5 CFU, 1×10^4 CFU, and 1×10^3 CFU dilutions, which were used to determine the number of bacterial cells present within the initial inoculum in CFU/mL. These dilutions were also incubated at 37 °C for 3 h, with low level shaking at 50 rpm. After incubation, each Eppendorf that contained a test suspension was vortexed and 10 μL of each was plated out using disposable sterile spreaders, and the plates were incubated at 37 °C overnight.

The following day, the plates were removed from the incubator and the colonies on each of the plates that were less than 200 were counted using a colony counter. Where a lawn of bacteria was observed it was recorded as growth and indicated no bacterial killing. The MBC value for each peptide was determined as the peptide concentration which resulted in greater than 99.99 % killing of the initial inoculum (Table 3.5). This was the peptide concentration that resulted in less than or an equal number of colonies to those observed for the negative control for either *P. aeruginosa* or *S. aureus*. All antimicrobial assays were performed in triplicate and on three independent occasions. The MBC value was obtained by taking the mean of all results, and all experimental errors were within one doubling dilution.

The N-terminal HBD3 peptide **30** and C-terminal HBD3 peptide **33** "halves" that were used to synthesise the full-length defensin; the ligated defensin with five Acm protected cysteine residues **35**; and the folded full-length defensin **36** were all tested for bactericidal activity. Of those peptides tested, only the full-length folded defensin **36** was found to possess broad spectrum antimicrobial activity (Table 3.5).[267] The folded defensin had MBC values equivalent to those previously reported for the defensin,[271] and was found to be as potent as the positive control mouse orthologue of HBD3, Defb14-1Cys^{V}, which had five out of its six cysteine residues mutated to alanine and only retained the native Cys^{V} residue.[273] The MBC value for **36** against *S. aureus* (1.8 µM) was one-doubling dilution higher than the MBC against *P. aeruginosa* (0.9 µM).[267]

Table 3.5 Minimum bactericidal concentrations (MBCs) of the synthesised defensin analogue Ser22His HBD3 **36** and synthetic peptide fragments thereof (antimicrobial assays were conducted with and completed by Natalie L. Reynolds).[267]

Microbial Organism	Minimum Bactericidal concentration (µM)[a]				
	30	**33**	**35**	**36** HBD3[O]	Defb14-1Cys^{V}
Pseudomonas aeruginosa	2.7	11.9	2.7	0.9	1.25
Staphylococcus aureus	>11	>11	>3.2	1.8	2.5

[a]Details of how the antimicrobial assays were conducted are described in Section 3.3.10 and in Experimental Section 6.

Notably, the N-terminal HBD3 peptide **30** terminating in a histidine-cysteine sequence was active against *P. aeruginosa* (2.7 μM) but not *S. aureus* (> 11 μM), and it was less potent than the full-length folded defensin **36** (Table 3.5).[267] The ligated full-length defensin with five Acm protected cysteine residues **35** was also found to be active against *P. aeruginosa* (2.7 μM).[267] Conversely, the C-terminal HBD3(23-45) peptide **33** was found to not possess bactericidal activity against either the Gram negative (11.9 μM) or Gram positive (>11 μM) bacteria.[267] This supported what was known with regards to the antibacterial activity residing within the N-terminal "half" of HBD3.[271]

Chemotaxis Assays

A correct three-dimensional structure is not required for the antimicrobial activity of HBD3 however its chemoattractant ability does depend upon a correct three-dimensional structure.[272] Therefore, a chemoattraction assay was conducted with the folded defensin **36** and other defensin-derived peptides to test its chemoattractant ability (see Experimental Section 6 for a description of the chemotaxis assay and Figure 3.17).[267] HBD3 chemoattracts cells that express the cell-surface chemokine receptor 6 (CCR6). Therefore, human embryonic kidney-293 (HEK293) cells that were transfected with CCR6 were used in the assay. Figure 3.17 illustrates the number of CCR6-expressing HEK293 cells that migrated towards the folded defensin preparation **36**.[267]

The chemotaxis graph is a bell-shaped curve which is characteristic of chemoattraction by Ser22His HBD3 **36** (Figure 3.17b).[267] The optimal chemoattraction concentration was found to be 0.2 nM,[267] which is equivalent to that reported for HBD3 prepared by traditional methodology.[272] This suggested that introduction of the Ser22His mutation, or exposing the N-terminal fragment of HBD3 to the optimised thioesterification conditions, did not compromise the biological activity of the defensin.[267] Neither the N- or C-terminal halves of the defensin (**30** and **33**) were found to be chemoattractant peptides, which can be attributed to the lack of the correct three-dimensional structure, and these findings are in agreement with those of previous studies on the mouse orthologue of HBD3, Defb14.[273]

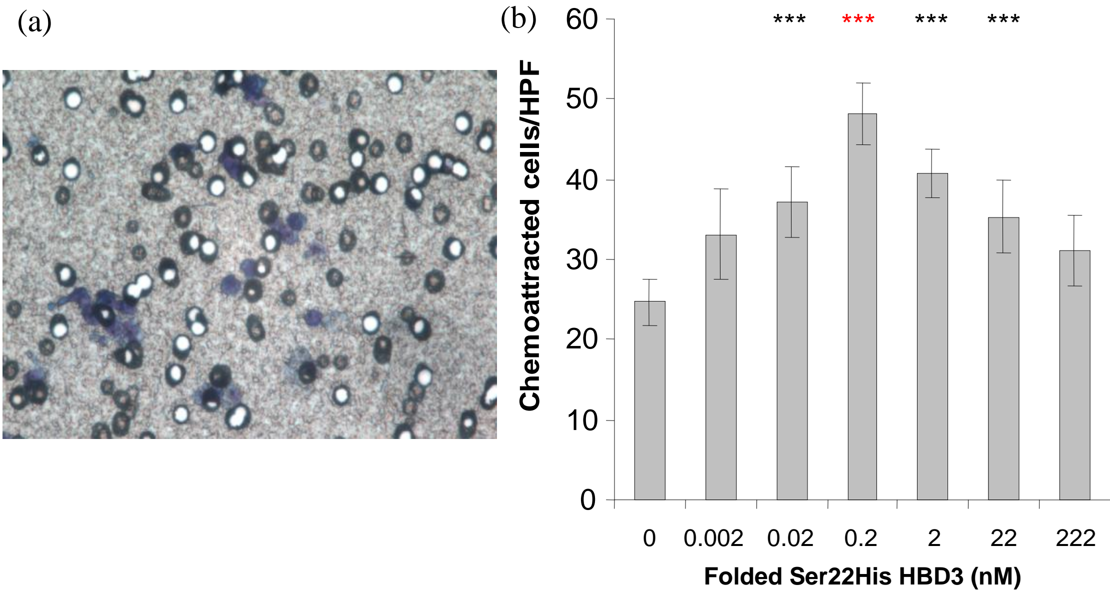


Figure 3.17 (a) The micrograph represents the chemoattraction of human embryonic kidney-293 (HEK293) cells that were transfected with the chemokine receptor CCR6 by the folded defensin Ser22His HBD3 synthesised **36** (chemotaxis assays were conducted with and completed by Dr Christine Tyrrell).[267] The assay was conducted using a 96 well microchemotaxis chamber where the central wells were used and the outer wells (buffer zone) were filled with chemotactic media alone (RPMI/ 1 % bovine serum albumin). The bottom chamber central wells contained the test peptide in concentrations increasing from 0 (chemotactic media only) to 222 nM (30 μL each). The central wells in the top chamber contained the transfected HEK293 CCR6 cells at 1 × 10^5 cells/mL in chemotactic media and the vacant outer wells were filled with chemotactic media alone (200 μL each). The positive control was human macrophage inflammatory protein-3α (hMIP-3α) and the negative control was the chemotactic media alone. The two chambers were separated by a collagen-coated polycarbonate filter (pore size 10 μm). [The filter was coated groove side down in a solution of 5 mL human collagen and 45 mL sterile water, which was then incubated at 37 °C overnight and then air-dried.] The chamber was placed in a humid, 5 % CO_2 incubator at 37 °C for 4 h. The filter was then removed and the side with the non-migrating cells was washed with PBS buffer and then left to air dry. The filter was stained with solution A (methanol fixative) of DiffQuick for 2 min, solution B (Azure B) for 2 min, and solution C (Eosin) for 2 min, and then gently rinsed in sterile water and air dried. The filter was then observed under a microscope with three random fields of view taken per well using a high power field with × 40 magnification, and the migrated cells were counted in each view. The mean number of migrated cells was then calculated as well as the standard error of means (SEM). Each test peptide dilution was tested in triplicate and each experiment was repeated a minimum of three times. Data was analysed using analysis of variance (ANOVA) with post-test at $p < 0.05$. The chemotactic activity was determined as the optimal chemoattraction concentration of the test peptide that produced the highest chemotactic index. (b) The chemotaxis graph shows the number of CCR6-transfected HEK293 cells that were chemoattracted at different concentrations of folded Ser22His HBD3 **36**. Data is presented as the numerical mean ± standard deviation of the number of cells per high power field of view at × 40 magnification. Each assay was repeated three times, in triplicate with three fields of view per repeat. (***)Asterisks represent significant chemoattraction compared to the positive control hMIP-3α at $p < 0.001$. The optimal chemoattraction concentration of the folded defensin analogue **36** was found to be 0.2 nM. The results were approximately equivalent to those of the CCR6 chemokine ligand hMIP-3α and murine *β*-defensin Defb14.[267]

3.7 Discussion

Thioester hydrolysis was observed when MPA was used as the thioester-forming reagent,[263] therefore further optimisation of the reaction was undertaken.[267] The reaction of a ^{13}C-labelled model peptide (H-AENITTG(^{13}*C*-1)C-NH_2) **26** with varying thiols and reaction conditions was followed by ^{13}C NMR spectroscopy.[267] The optimisation studies revealed water-soluble MESNa to be a superior transthioesterification reagent to MPA.[267] MESNa fully converted the starting peptide amide into the desired peptide thioester (H-AENITTG(^{13}*C*-1)-$SCH_2CH_2SO_3H$) within 48 h, compared to only approximately 60 % conversion with MPA.[267] Furthermore, no peptide thioester hydrolysis occurred when MESNa (at pH 2.0) was employed as the thiol.[267] The optimal thioesterification conditions were found to be heating at 60 °C with excess thiol (10 %) in aqueous acidic media (pH 2.0-6.0), with 0.5 % (w/v) TCEP to reduce any potential disulfide bonds.[267]

The concentration of the starting peptide was also important as lower concentrations (less than 6 mM) were found to increase the reaction rate.[267] This was likely because of a reduction in unwanted intermolecular disulfide bonds, which acted to decrease the effective concentration of peptide available to perform the *N→S* acyl transfer.

The optimised thioester-forming conditions were then applied to the total chemical synthesis of the biologically active defensin analogue Ser22His HBD3,[267] which was found to have antimicrobial and chemoattractant activity equivalent to that reported.[271]

Thioesterification of the N-terminal 23 amino acid residue peptide (**30**) of HBD3 did not proceed to completion with 10 % MESNa and 1 % AcOH after 48 h at 60 °C, as was expected from the optimisation studies conducted with the 8 amino acid residue ^{13}C-labelled model peptide. However the 22 % yield of the 22 amino acid thioester **31** obtained is comparable to the synthesis of such complex thioesters by other methods.[199,200,208,210,214,252-256] This raised several questions about potential factors that may be contributing to peptide thioesterification including: the reaction time, pH, solvent, reducing agent, hydrophobicity, peptide chain length, peptide concentration, side-chain substituents, and even the decreased accessibility of the thiol due to formation of a hydration sphere.

Overall, the initial proposed reaction route (Scheme 3.1b) appears to be an oversimplification and several factors appear to contribute to the formation of the thioester through an initial *N*→*S* acyl shift. An alternative scheme for thioester formation through an initial *N*→*S* acyl shift, followed by transthioesterification to give the thioester (NCT) is shown in Scheme 3.8.

A histidine-cysteine, glycine-cysteine, or cysteine-cysteine sequence present in **I** initially undergoes a reversible intramolecular *N*→*S* acyl shift under acidic conditions to give **IV**, the '*S*-peptide' (Scheme 3.8). This *N*→*S* acyl transfer can proceed through a five-membered ring transition state, or through a proposed five-membered ring hydroxythiazolidine intermediate **II**.[266] The intermediate **II** has not been observed during this study but it can release the amine **III** which is then protonated to give **IV**. The *S*-peptide **IV** then undergoes intermolecular transthioesterification to afford the isolable thioester **V** and the N-terminal cysteinyl fragment **VI**.

Previously, the Aimoto laboratory had attempted to isolate **IV** but found that it rapidly reverted back into the '*N*-peptide' **I**, even under acidic HPLC purification conditions (aqueous acetonitrile containing 0.1 % trifluoroacetic acid).[236] However, our studies have shown that the addition of a thiol under acidic conditions in one-pot, promotes thioesterification and directly generates **V** without the need to isolate **IV**.[263] Furthermore, detachment of the N-terminal cysteinyl fragment **VI** upon transthioesterification prevents the reverse transformation into **I**, and is thus irreversible.

Certain reaction conditions were found to promote undesirable thioester hydrolysis of either **IV** or **V** into **VII**. Thioester hydrolysis was found to be favoured when the pH was lower than 1.0, when 10 % 2-mercaptoacetic acid (MAA) was employed as the thiol, or when the reaction contained aqueous trifluoroacetic acid.[267] The starting material **I** was also found to be reversibly oxidised with the external thiol present within the one-pot reaction to afford **VIII**, which could be reduced back to **I** upon addition of 0.5 % w/v TCEP.[267]

Scheme 3.8 Proposed intermediates and products of Native Chemical Thioesterification involving thioester formation *via* an *N→S* acyl transfer.

The formation of the *S*-peptide is favoured by the protonation of the released α-amino group of cysteine, and can be further assisted by deprotonation of the cysteine sulfhydryl group which attacks the 'scissile' amide bond. A lower pH promotes the release of the α-amino group of cysteine (pKa Cys $-NH_2 \approx 9$) to exist as the ammonium species ($-NH_3^+$), whereas a higher pH (pKa Cys -SH ≈ 8) promotes the formation of the cysteinyl thiolate ($-CH_2-S^-$).

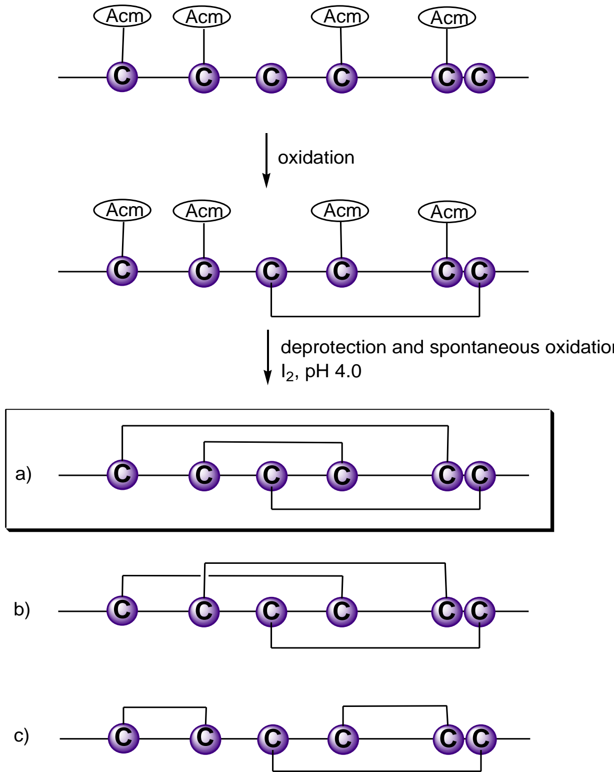
The optimum pH for thioesterification was found to reside towards the acidic side of the pH scale between 2.0 and 6.0.[267] Protonation of the α-amino group of cysteine and deprotonation of the sulfhydryl group of cysteine are factors that promote the shift in the *N-*/*S*-peptide equilibrium towards the *S*-peptide. The optimised thioesterification procedure deduced with MESNa was conducted at pH 2.0, and therefore the cysteine sulfhydryl was mostly protonated. However at pH 6.0 approximately 1 % of the cysteine sulfhydryl group is deprotonated which can assist its nucleophilic attack onto the carbonyl group of the scissile amide bond.

Thioesters are energetically less stable than amides and therefore it is reasonable to suspect that the equilibrium lies to the far left towards the *N*-peptide at neutral pH. However, under

acidic conditions the more basic amide nitrogen can be protonated and this shifts the equilibrium towards the *S*-peptide. Amides are stabilised by 2s-2p orbital overlap between the carbonyl carbon and the nitrogen atom and by resonance stabilisation. Comparatively, the 2s-3p orbital overlap between the carbonyl carbon and the sulfur atom in thioesters is weaker due to the less tightly held electrons in the larger 3p orbital. The transformation from the *N*-peptide to the *S*-peptide during the initial *N→S* acyl transfer step is proposed to be fast because it proceeds through a five-membered ring transition state or intermediate, and therefore the process appears to be governed by kinetics.

The initial *N→S* acyl transfer was observed to occur at room temperature, but the second transthioesterification step was found to require heat energy.[267] An accumulation of the *S*-peptide was also not observed during NCT. Increasing the concentration of the added thiol was observed to increase the reaction rate and therefore the second transthioesterification step was proposed to be the rate-determining step.[267]

During the first biological application of the new thioesterification reaction, three main topological isoforms of Ser22His HBD3 were observed during (RP)HPLC purification of the folded defensin preparation,[267] though they were not resolved because all isoforms have been shown to be biologically active.[271] However, potential future work for the preparation of the native topological isoform of Ser22His HBD3 could be based on the alternative strategy depicted in Scheme 3.9. It can be employed to reduce the number of possible disulfide connectivities from fifteen to three combinations. This optional synthetic route would require the formation of one of the three desired disulfide bonds first, whilst the remaining four cysteine residues are acetamidomethyl (Acm) protected. This is then followed by concomitant deprotection of the remaining cysteine residues and spontaneous oxidation with iodine at mildly acidic pH to form the remaining two disulfide bonds. As our chosen ligation site was between ^{22}His and ^{23}Cys^{III}, therefore the ^{23}Cys^{III}-^{41}Cys^{VI} disulfide bond could be installed first.



Scheme 3.9 Proposed future work: a synthetic route to reduce the number of possible topological isomers of HBD3 during folding from fifteen possibilities to three. “Acm” denotes an acetamidomethyl orthogonal protecting group.

3.8 Conclusion

MPA-mediated thioesterification was optimised and MESNa was identified as a superior transthioesterification reagent to MPA. MESNa fully converted a ^{13}C-labelled model peptide terminating in a glycine-cysteine sequence into the thioester within 48 h and did not result in detectable peptide thioester hydrolysis at pH 2.0.[267] Through a series of optimisation experiments, the following reaction conditions were deduced to be excellent for the production of stable thioesters: 10 % MESNa in combination with 1-10 % acetic acid and 0.5 % TCEP at 60 °C for 48 h.[267] Thus, the simple appendage of a histidine-cysteine, glycine-cysteine, or a cysteine-cysteine sequence to the C-terminus of a peptide can potentially be used to synthesise the desired peptide thioester using these optimised conditions.[263,267]

The optimised conditions were subsequently employed in the first biological application of the novel thioesterification method, in the assembly of Ser22His HBD3,[267] which was found to have antimicrobial and chemoattractant activity equivalent to that known for HBD3.[271] The simplicity of the new thioesterification method is significant, and also forms the basis of its potential application in the production of designed protein thioesters from recombinant samples.

3.9 Acknowledgments

Biological evaluation of the *β*-defensin-derived peptides (Section 3.3.10) was conducted with the help of Dr Julia R. Dorin and her group at the MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, Western General Hospital, Edinburgh. The antimicrobial assays were conducted with and completed by Natalie L. Reynolds. The chemotaxis assays were conducted with and completed by Dr Christine Tyrrell.

4 The First Application of Native Chemical Thioesterification to the Synthesis of Native *N*-Linked Glycopeptide Thioesters

4.1 Introduction

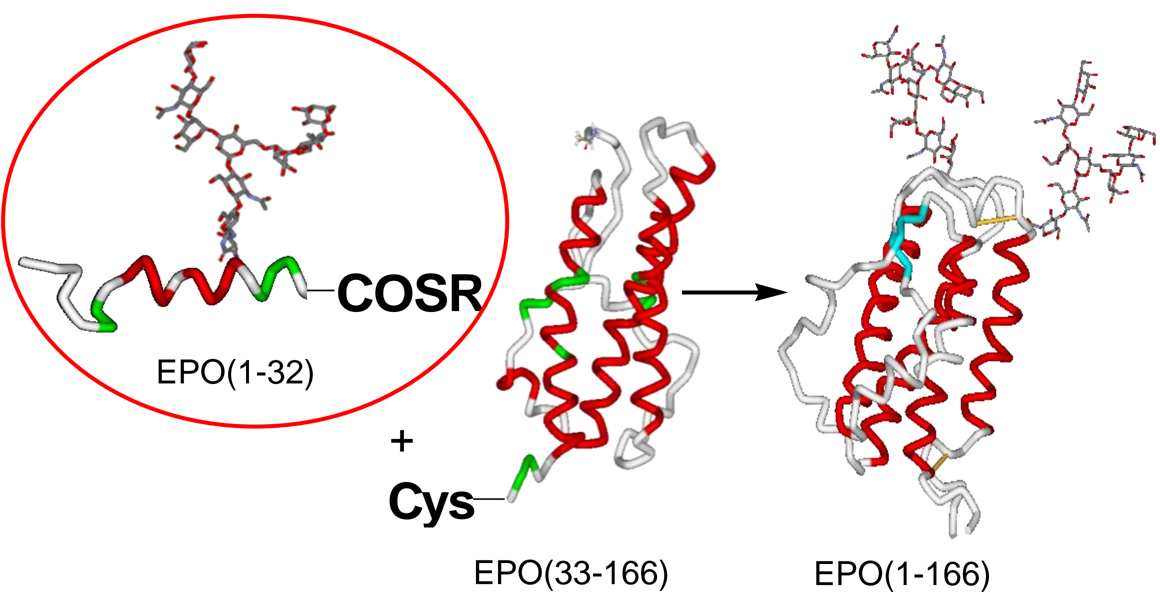
Approximately 50 % of human proteins are estimated to be glycosylated[274] and appending oligosaccharides increases the structural diversity of proteins. Glycosylation is responsible for many biological recognition events including: cell adhesion, cell differentiation, and cell growth.[275,276] Aberrant protein glycosylation has often been shown to alter the recognition events, and is thought to be responsible for many disorders including: autoimmune diseases, infectious diseases and cancer. Therefore, glycoproteins are highly sought-after for the development of drug therapeutics, diagnostic tools, and vaccines.[277]

O-linked glycans (oligosaccharides) are α- or β-linked to the hydroxyl group of a serine, threonine, or tyrosine residue within a native protein sequence, whereas *N*-linked glycans depend upon *N*-acetyl-D-glucosamine (Glc*N*Ac) being β-linked to the amide side-chain of an asparagine residue within an Asn-Xaa-Ser/Thr consensus sequence.[275-277] Oligosaccharides that are attached to a native protein backbone are often complex and branched due to different points of attachment to the existing glycan free hydroxyl groups, and either α- or β-stereochemistry. A natural glycoprotein can have many glycosylated forms because protein glycosylation is not under direct genetic control,[275-277] and therefore the (semi)synthesis of proteins bearing specific glycoforms allows us to determine the role of glycosylation at a molecular level.[156,158] NCL is generally preferred for the synthesis of posttranslationally modified proteins because sufficient quantities of homogeneous posttranslationally modified proteins are not easily obtained from biological sources.[144,153,187,278,279]

Recombinant human erythropoietin (rhEPO) has four sites of glycosylation: *O*-linked glycan at serine 138, and three *N*-linked glycans at asparagines 24, 38 and 83. Erythropoietin exists as a heterogeneous mixture of glycoforms, which has made it difficult to analyse the effect of specific types of glycosylation on the structure and function of the protein.[280]

The *O*-linked glycan at serine 138 in EPO has been found to not be required for *in vivo* or *in vitro* biological activity; however the *N*-linked glycans have been assessed to be essential for *in vivo* activity, because they increase the circulatory half-life of the glycoprotein by being sialylated. The *N*-linked glycans also allow the glycoprotein to fold correctly and be normally secreted.[281-291]

The Danishefsky laboratory has been engaged in the laborious, yet potentially extremely valuable, quest towards the total chemical synthesis of homogeneous EPO.[292-294] We adopted an alternative approach that was based on the semisynthesis of EPO analogues. Semisynthesis allows us to employ NCL to potentially incorporate multiple specific glycans at pre-determined sites within the synthetic fragment to afford native glycosidic linkages.[190,207,295] Our semisynthetic strategy involved the chemical synthesis of a glycopeptide thioester followed by an NCL reaction with a biologically expressed cysteinyl fragment to provide access to EPO analogues (Scheme 4.1).



Scheme 4.1 Our semisynthetic strategy towards the 166 amino acid glycoprotein erythropoietin (EPO). The encircled glycosylated peptide thioester EPO(1-32) is chemically-derived and is ligated to the biologically-derived N-terminal cysteinyl fragment EPO(33-166) to afford the full-length site-specifically modified EPO(1-166).

We began this project with the synthesis of native *N*-linked glycopeptide thioesters in mind (Chapter 2) and though we had not anticipated the discovery of a new thioesterification

reaction,[263,267] it seemed logical to apply it to the synthesis of these highly desirable posttranslationally modified thioesters towards the semisynthesis of erythropoietin. Here, we describe the first chemical synthesis of the posttranslationally modified native *N*-linked glycopeptide thioesters *via* "Native Chemical Thioesterification" (NCT). NCT is used to describe thioester synthesis through an initial *N*→*S* acyl shift which occurs site-specifically at histidine-cysteine, glycine-cysteine, and cysteine-cysteine sequences to afford the *S*-peptide intermediate, which is then transthioesterified to give the isolable thioester (Scheme 4.2).[296]

HS, O, N, H, O, H_3O^+, $H_3\overset{+}{N}$, O, O, S, RSH, O, SR

Scheme 4.2 Native Chemical Thioesterification (NCT). One-pot peptide thioester synthesis is instigated by an initial *N*→*S* acyl transfer at histidine-cysteine, glycine-cysteine, or cysteine-cysteine sequences under mildly acidic conditions. The resulting thioester-linked branched intermediate is transthioesterified by a small-molecular weight thiol to provide the isolable thioester.[296]

4.2 Results and Discussion

4.2.1 Synthesis of the *N*-Linked Glycoamino Acid

The preparation of a model *N*-linked glycopeptide thioester containing a natural glycopeptide linkage began with the synthesis of the *N*-linked glycoamino acid **42** (Scheme 4.3). An *N*-linked glycoamino acid was synthesised for incorporation into a growing peptide, rather than the method developed by Lansbury which involved the selective glycosylation of deprotected aspartic acid residues within an assembled peptide.[297]

Initially, *N*-acetyl-D-glucosamine **37** was treated with acetyl chloride to form the peracetylated *α*-chloro-glycoside **38**, which was purified using silica-gel column chromatography in 47 % yield. Mechanistically, the sugar alcohol functional groups react with acetyl chloride to displace the chloride anion and produce four equivalents of hydrochloric acid overall. Following the complete acetylation of the sugar, acetic acid is lost from the anomeric carbon *via* an oxocarbenium ion. The chloride anions produced then participate in nucleophilic attack at the anomeric carbon to produce the *α*-chloro-glycoside **38**, which is stabilised by the anomeric effect. At best, this first step is found to provide half-maximal conversion into the product. Ideally, escape of the hydrochloric acid produced from the reaction flask would be limited, in order to endorse the formation of the desired product by increasing the availability of chloride anions that can participate in nucleophilic attack at the anomeric carbon.

Scheme 4.3 Synthesis of the *N*-linked glycoamino acid **42** Fmoc-Asn(Ac_3-Glc*N*Ac)-OH for incorporation into the Fmoc based solid phase peptide synthesis of a native *N*-linked glycopeptide **48** (H-AEN(Ac_3-Glc*N*Ac)ITTGC-OH). *N*-acetyl-D-glucosamine **37** was initially transformed into *α*-chloro-glycoside **38**, and then into the *β*-azido-glycoside **39**. The *β*-azide group was then reduced to afford the *β*-amino-glycoside **40**, which was then coupled to side-chain active ester of Fmoc-Asp-O^{t}Bu to afford the new amide bond in **41**. The *tert*-butyl group was then removed from the α-carboxyl group of Asn to afford **42**.

The peracetylated *α*-chloro-glycoside **38** was then subjected to a biphasic (dichloromethane/ aqueous saturated sodium bicarbonate) solution to react with sodium azide and afford the peracetylated *β*-azido-glycoside **39** in an excellent 95 % yield. This process is known to be an example of *phase transfer catalysis*.[298] This is because the peracetylated **38** resides within the dense dichloromethane layer, and tetrabutylammonium hydrogen sulfate (TBAHS) and sodium azide are soluble in the aqueous layer. TBAHS initially reacts with sodium azide to form "TBA-N_3", which then transfers into the organic layer to allow the direct S_N2 attack of the azide at the anomeric carbon to displace the chloride anion.

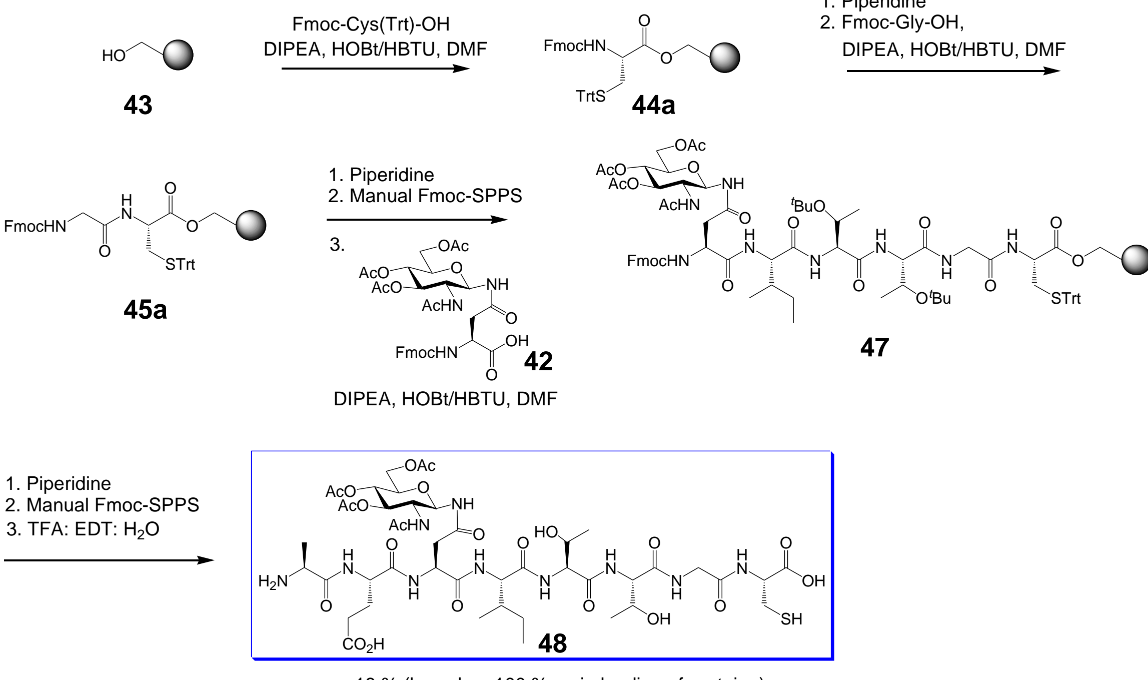
The *β*-azido-glycoside **39** was then hydrogenated in the presence of platinum oxide to afford the reduced *β*-amino-glycoside **40** in 97 % yield. The amino sugar **41** was then coupled to the side-chain of Fmoc-Asp-O^{t}Bu, using standard methodology.[299,300] The side-chain of aspartic acid was pre-activated with EEDQ (*N*-Ethoxycarbonyl-2-ethoxy-1,2-dihydroquinoline)[301,302] to afford the mixed carbonic anhydride of aspartic acid, which reacted with **40** to afford the coupled *N*-linked glycoamino acid **41** after work-up. Finally, the *tert*-butyl group was removed from the *α*-carboxyl group of **41** with 95 % v/v aqueous TFA, to afford the desired *N*-linked glycoamino acid building block **42** in quantitative yield. Throughout this five-step glycoamino acid synthetic route, a single flash column chromatography purification step was required in the first transformation of **37** into **38**. The remaining steps within this strategy are high-yielding, and thus render this an attractive route.

An alternative approach for obtaining deprotected glycosylated L-asparagine derivatives includes the direct coupling of a *β*-azido-glycoside onto the side-chain carboxylic acid of aspartic acid, using a tertiary phosphine. A simple three component strategy of this sort was reported by Inazu and co-workers for the synthesis of *N*-*β*-glycosylated L-asparagine derivatives through a Staudinger reaction between a carboxylic acid and a protected glycosyl azide, in the presence of a tertiary phosphine, to give the *N*-linked glycoamino acid in yields ranging from 23 to 77 %, depending upon the solvent used.[303,304] Davis and co-workers recently reported the synthesis of *N*-linked glycoamino acids and glycopeptides through a Staudinger ligation using a *deacetylated β*-azido-glycoside, an activated side-chain carboxylic acid of aspartic acid, and tributylphosphine, using mainly acetonitrile as the solvent, in yields ranging from 47 to 87 %.[305] Another recently reported approach to generate *N*-glycosylated

L-asparagine for *N*-glycopeptide synthesis involved the reaction of *N*-Fmoc aspartic anhydride with a per-*O*-acetylated glycosyl amine in dimethylsulfoxide.[306]

4.2.2 Synthesis of the Native *N*-Linked Glycopeptide

A model *N*-linked glycopeptide that was based on a short peptide sequence of erythropoietin EPO(22-29): H-AEN(Ac$_3$-Glc*N*Ac)ITTGC-NH$_2$ was synthesised (Scheme 4.4). The glycopeptide synthesised contained a C-terminal glycine-cysteine sequence to enable thioesterification, and a native glycosylated asparagine at position-24 within the EPO sequence.



Scheme 4.4 Manual Fmoc-based solid phase peptide synthesis of the native *N*-linked glycopeptide **48** (H-AEN(Ac$_3$-Glc*N*Ac)ITTGC-OH). NovaSyn TGA® resin **43** was initially coupled with Fmoc-Cys(Trt)-OH using 0.45 M HOBt/HBTU, DIPEA, DMF. The peptide chain was manually elongated and 2 equivalents of Fmoc-glycoamino acid **42** were coupled under standard conditions, though for an extended coupling time of 30 h, to give **47**. The α-amino Fmoc group was removed from **47** with 20 % piperidine/DMF and peptide chain elongation was continued to afford **48**.

The glycopeptide chain was manually assembled using Fmoc-based solid phase peptide synthesis on NovaSyn TGA® resin **43** using the standard coupling reagents 0.45 M HOBt/HBTU in the presence of the tertiary amine *N,N*-diisopropylethylamine, and five equivalents of each *N*-Fmoc protected amino acid (except the glycoamino acid **42**). The resin **43** was chosen to furnish a C-terminal free carboxylic acid upon cleavage of the peptide after synthesis. This was because a peptide functionalised with a C-terminal free carboxylic acid had been shown to promote an increased amount of thioesterification, when compared to a peptide with a C-terminal carboxamide, though the reason for this was unknown.[307]

Two equivalents of the *N*-Fmoc protected glycoamino acid **42** were used during coupling to **45a**, along with a prolonged reaction time of 30 h. An advantage of increasing the coupling reaction time is that less equivalents of the synthetic glycoamino acid can be used. Once the final amino acid, Boc-Ala-OH, had been coupled to **47**, it was then treated with the standard cleavage cocktail comprised of 95 % TFA: 2.5 % ethanedithiol: 2.5 % water (v/v/v) for 3 h, to remove the protecting groups and concomitantly cleave it from the resin. After (RP)HPLC purification, the *N*-linked glycopeptide **48** was isolated in 18 % yield (based on 100 % loading of cysteine) (Figure 4.1).

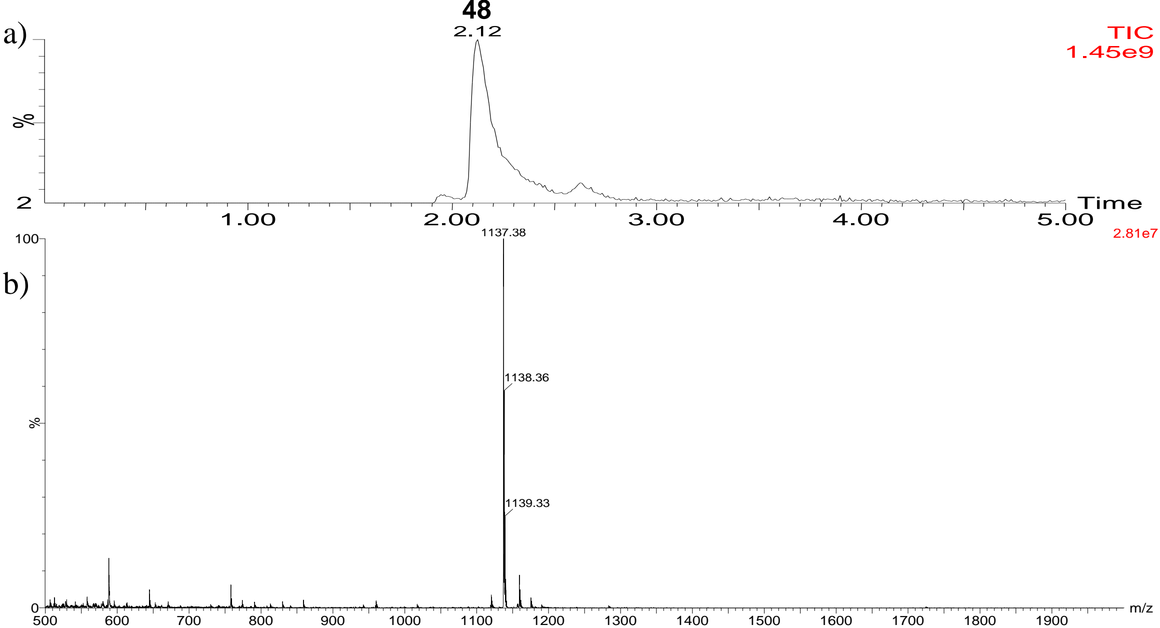


Figure 4.1 LC-MS analysis of the synthesised *N*-linked glycopeptide **48** (H-AEN(Ac_3-Glc*N*Ac)ITTGC-OH) (Scheme 4.4) for glycopeptide thioester formation. (a) TIC chromatogram and (b) ESI^+ mass spectrum, calculated mass for [M]: 1136.18, observed $[MH]^+$: 1137.38.

4.2.3 Native *N*-Linked Glycopeptide Thioester Formation

The *N*-linked glycopeptide **48** terminating in a glycine-cysteine sequence was then reacted in a Native Chemical Thioesterification reaction. The reaction was conducted in the chaotropic reagent 6 M guanidine hydrochloride, though it was a short peptide and not prone to aggregation, in the presence of excess thiol 10 % w/v sodium 2-mercaptoethanesulfonate, 10 % acetic acid, and 20 mM TCEP. The final pH of the reaction mixture was 5.0, which was possibly the result of a buffering effect of the slightly acidic guanidine hydrochloride reaction medium. Thioesterification was carried out at 60 °C and monitored by LC-MS analysis.

The glycopeptide thioester and starting material were observed to co-elute under a single broad peak in the LC-MS total ion count (TIC) chromatograms. However, the glycopeptide thioester **49** was observed to have a strong UV absorbance during LC-MS analysis at 230 nm relative to the starting glycopeptide **48** (Figure 4.2). After 6 h, the main component of the reaction mixture was the unreacted starting material. However, a small peak that could be attributed to the starting glycopeptide with a sodium ion $[M + Na]^+$ or the glycopeptide thioester was also observed. After 24 h, LC-MS analysis showed an increased amount of glycopeptide thioester **49**, which further increased after 48 h. After 48 h, **49** appeared to be the main product by UV absorbance at 230 nm; however the LC-MS TIC analysis showed unreacted starting material **48** to be the main product.

The *N*-linked glycopeptide thioester **49** was isolated in 20 % yield after (RP)HPLC purification. Unreacted **48** was also isolated in 47 % yield (based on starting material used). Interestingly, during LC-MS analysis the product thioester **49** was observed to elute after the starting material **48**, however during HPLC purification **49** was earlier eluting than the starting peptide. This can be attributed to the increased polarity of the 7 amino acid residue thioester product **49** compared to the 8 amino acid residue starting material **48**. This reversal has been frequently observed throughout the thioesterification studies.

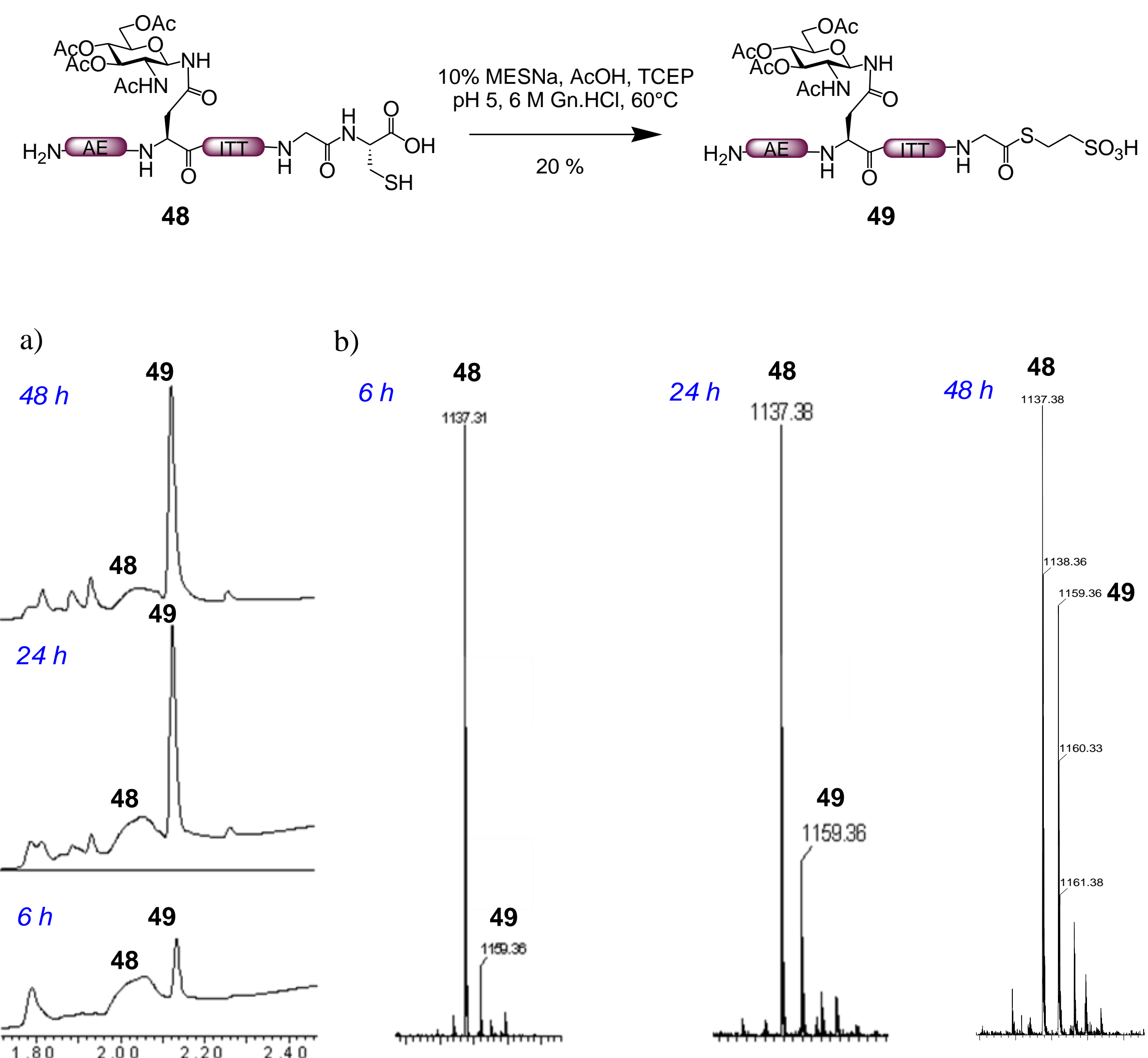


Figure 4.2 Native Chemical Thioesterification of the native *N*-glycopeptide **48** (H-AEN(Ac_3-Glc*N*Ac)ITTGC-OH) to afford the native *N*-linked glycopeptide thioester **49** (H-AEN(Ac_3-Glc*N*Ac)ITTG-$SCH_2CH_2SO_3H$). The *N*-glycopeptide **48** was treated with 10 % w/v MESNa in 6 M guanidine hydrochloride, 10 % v/v AcOH, 0.5 % w/v TCEP; pH 5.0, at 60 °C for 48 h. The reaction was monitored by LC-MS at t = 6, 24, and 48 h, and purified after 48 h by semi-preparative (RP)HPLC to afford the *N*-linked glycopeptide thioester **49** in 20 % isolated yield. (a) LC-MS UV chromatograms of the thioesterification after 6, 24, and 48 h. Absorbance detected at 230 nm. (b) ESI^+ mass spectra following thioester formation after 6, 24, and 48 h. Calculated average mass for **49** [M]: 1158.21, observed $[MH]^+$: 1159.36.

4.2.4 Native Chemical Ligation of the *N*-Glycopeptide Thioester

The isolated native *N*-linked glycopeptide thioester **49** (sequence: H-AEN(Ac_3-Glc*N*Ac)ITTG-$SCH_2CH_2SO_3H$) was then reacted in an NCL reaction with excess model N-terminal cysteinyl peptide derived from a short sequence of erythropoietin (H-CSLNENIT-OH) **50** (Scheme 4.5). Standard ligation conditions were employed, 2.6 mM of the glycopeptide thioester **49** was dissolved in 6 M guanidine hydrochloride containing 200 mM sodium phosphate buffer; pH 7.0, 100 mM sodium 2-mercaptoethanesulfonate, and 40 mM *tris*(2-carboxyethyl)phosphine. The reaction was conducted at 25 °C and it was monitored at regular intervals by LC-MS analysis. After 14 h, the expected ligated product **51** was observed. The 15 amino acid residue *N*-glycopeptide **51** was observed to have a longer retention time than both the cysteinyl peptide **50** and the *N*-glycopeptide thioester **49** by LC-MS.

6M Gu.HCl, 200mM Na phosphate, pH7, 40mM TCEP, 100mM MESNa, 25°C

30 %

50 **49** **51**

Scheme 4.5 An NCL reaction using typical ligation conditions between the isolated *N*-linked glycopeptide thioester **49** (H-AEN(Ac_3-Glc*N*Ac)ITTG-$SCH_2CH_2SO_3H$) and excess cysteinyl peptide **50** EPO(33-40): H-CSLNENIT-OH, to afford the 15 amino acid residue *N*-linked glycopeptide **51** (H-AEN(Ac_3-Glc*N*Ac)ITTGCSLNENIT-OH). Calculated mass for the ligated product **51** [M]: 1907.83, observed $[MH]^+$: 1909.17.

After 62 h, LC-MS analysis of the NCL reaction still showed the presence of unreacted thioester **49**, however the relative amount of the ligated product remained unchanged since the first analysis after 14 h. Therefore, the NCL reaction mixture was purified by semi-preparative (RP)HPLC to afford the ligated *N*-glycopeptide **51** in 30 % yield. This experiment verified that the native *N*-linked glycopeptide thioester **49** had been formed.

4.2.6 Native Chemical Thioesterification of a Fully Deprotected Native *N*-Linked Glycopeptide

Shin *et al.* carried out the synthesis of the antimicrobial *O*-linked glycoprotein diptericin using acetate protected *O*-linked monosaccharides during the Fmoc-based solid phase synthesis of the glycopeptide thioester, *via* a modified alkanesulfonamide safety-catch linker.[199] This was to avoid alkylation by iodoacetonitrile of the free hydroxyl groups within the carbohydrate during activation of the amide nitrogen within the linker. Consequently, after the glycopeptide thioester had been synthesised, the carbohydrate acetate protecting groups were removed. This added an additional step in the synthetic route to their target glycoprotein, after the glycopeptide had been assembled.

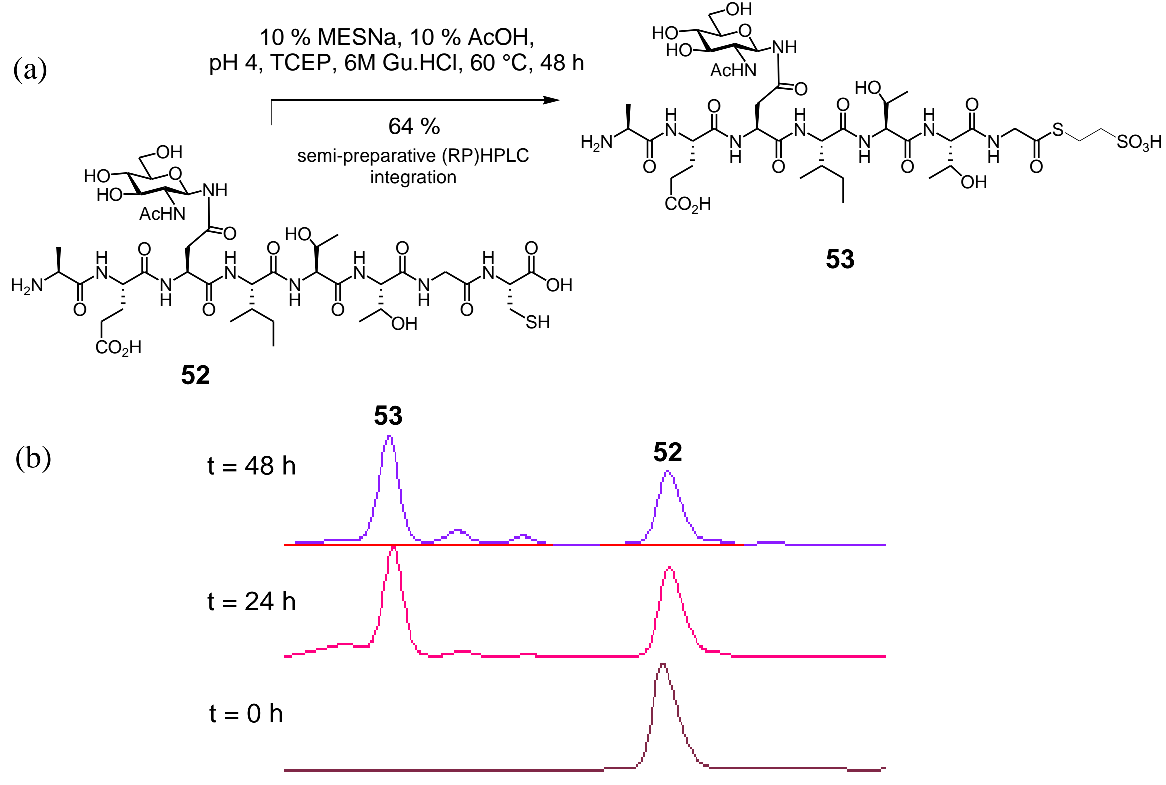
Acetate protecting groups within a carbohydrate would at some point have to be removed. Whether it is before the monosaccharide has been coupled to the amino acid, or after the glycoamino acid itself has been synthesised, or after the glycoamino acid has been incorporated into the growing peptide chain. Native glycopeptides do not contain protected carbohydrates and therefore there are several stages at which the carbohydrate protecting groups can be removed. Whilst it was necessary in Shin *et al.*'s thioester synthesis procedure to incorporate fully protected carbohydrates to avoid the alkylation side-reaction,[199] we investigated the compatibility of our thioesterification method with peptides containing deprotected native *N*-linked carbohydrates.

For this purpose, the acetate protecting groups from the carbohydrate in the *N*-linked glycopeptide **48** were removed with 5 % v/v hydrazine in water[199] (Scheme 4.6). An excess amount of 1,4-dithio-DL-threitol (DTT) was used to prevent the formation of disulfide dimers

as a result of oxidation of the C-terminal cysteine residue in **48**.[199] After 3 h, LC-MS analysis showed that the reaction was complete and therefore the reaction mixture was centrifuged and the supernatant was purified by semi-preparative (RP)HPLC. The deacetylated *N*-glycopeptide **52** was isolated in 87 % yield.

Scheme 4.6 Deprotection of the peracetylated Glc*N*Ac in **48** (H-AEN(Ac_3-Glc*N*Ac)ITTGC-OH) was achieved with 5 % v/v aqueous hydrazine and excess 10 % w/v 1,4-dithio-DL-threitol, to prevent dimerisation of the starting material, to afford **52** (H-AEN(Glc*N*Ac)ITTGC-OH) for subsequent thioesterification studies.

The deprotected *N*-glycopeptide **52** was thioesterified in 6 M guanidine hydrochloride which contained 10 % w/v MESNa, 10 % v/v acetic acid, 0.5 % w/v TCEP; pH 4.0, at 60 °C for 48 h (Scheme 4.7). The reaction was monitored by analytical (RP)HPLC in conjunction with LC-MS. After 24 h, the deprotected *N*-glycopeptide thioester **53** was detected, and after a total time of 48 h unreacted starting glycopeptide **52** remained. The relative amount of the thioester produced after 48 h compared to that after 24 h had not increased significantly, and therefore the reaction was terminated. The deprotected *N*-glycopeptide thioester **53** was formed in 64 % yield (by semi-preparative (RP)HPLC integration) after 48 h.



Scheme 4.7 (a) Thioesterification of the deprotected *N*-glycopeptide **52** (H-AEN(Glc*N*Ac)ITTGC-OH) to afford **53** (H-AEN(Glc*N*Ac)ITTG-$SCH_2CH_2SO_3H$). (b) Analytical (RP)HPLC traces following the thioesterification of the deprotected *N*-glycopeptide **53** at t = 0, 24, and 48 h. Absorbance was detected at 230 nm.

4.3 Conclusion

N-Glycosylation of proteins is very important as it is often crucial for correct protein folding, protein stability, and protein function.[276] We have demonstrated that our simple thioesterification method is compatible with the synthesis of the highly sought-after native *N*-linked glycopeptide thioesters, which can contain either protected or deprotected sugars, in good yields. Glycopeptide thioester formation was verified through an NCL reaction with a model cysteinyl peptide to give a ligated 15 amino acid *N*-linked glycopeptide. Native Chemical Thioesterification provides a milder alternative to existing thioester synthesis approaches because the integrity of the attached carbohydrate is retained. Therefore, synthetic native *N*-linked glycopeptide thioesters can be prepared in this way for use in the semisynthesis of glycoproteins.

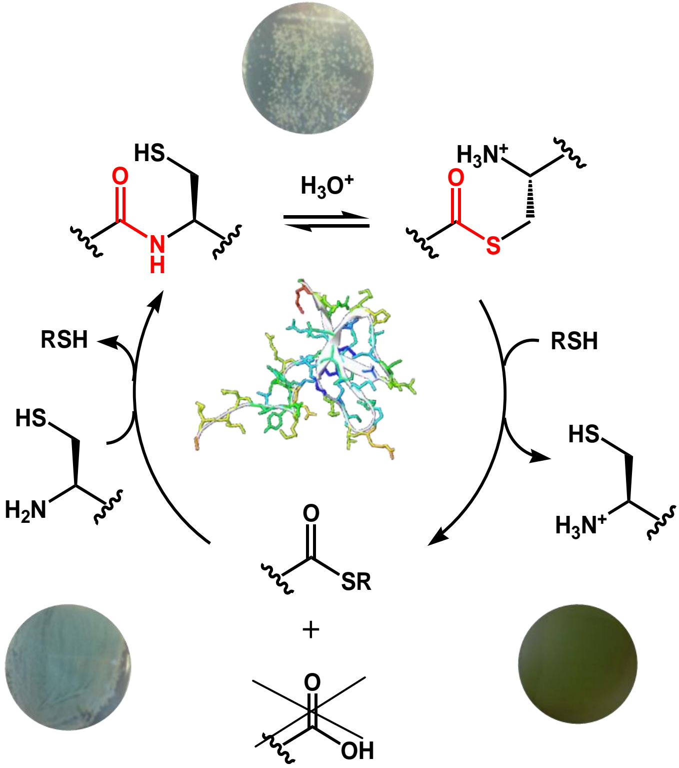
5 Overall Conclusions and Future Work

A new thioesterification reaction was discovered upon attempting to utilise peptides functionalised with C-terminal cysteinylprolyl esters. Peptides and a recombinant protein that contained histidine-cysteine, glycine-cysteine, and cysteine-cysteine sequences were found to selectively fragment in the presence of 3-mercaptopropionic acid (MPA) to form thioesters.[263] Mechanistically, it was proposed that an initial *N→S* acyl transfer occurs to afford a thioester-linked branched intermediate, which then undergoes transthioesterification with a thiol to produce an isolable thioester (Scheme 5.1).[263] However peptide thioester hydrolysis was also observed when MPA was used as the transthioesterification reagent.[263] Therefore, optimisation studies employing ^{13}C NMR spectroscopy were undertaken.

Sodium 2-mercaptoethanesulfonate (MESNa) in the presence of acetic acid and *tris*(2-carboxyethyl)phosphine were found to provide excellent conditions for thioesterification.[267] The thioesterification method discovered was then first applied to the total chemical synthesis of a biologically active protein.[267] An analogue of human *β*-defensin 3 was synthesised using the optimised thioesterification conditions, and the folded defensin was found to have antimicrobial and chemoattractant activity equivalent to that reported.[267]

The first application of the new thioesterification method to the synthesis of the highly sought-after native *N*-linked glycopeptide thioesters was also demonstrated, in which the glycopeptides contained either protected or deprotected monosaccharides.

Potential future work includes a detailed investigation into the mechanism of the thioesterification reaction, and determination of the role of the histidine, glycine, and cysteine residues adjacent to cysteine at the site of the *N→S* acyl transfer.



Scheme 5.1 In summary, a novel peptide and protein thioesterification reaction was discovered which reversed the process of Native Chemical Ligation. It was proposed to proceed through an initial *N*→*S* acyl shift selectively at histidine-, glycine-, and cysteine-cysteine sequences to form a thioester-linked branched intermediate, which can then be transthioesterified by a small molecular-weight thiol, such as sodium 2-mercaptoethanesulfonate (MESNa), to form an isolable thioester for use in Native Chemical Ligation reactions. This simple new method was termed Native Chemical Thioesterification, and was utilised to synthesise a biologically active analogue of the 45 amino acid residue human *β*-defensin 3, and native *N*-linked glycopeptide thioesters.

6 Experimental Section

6.1 General experimental information

All protected amino acids and the preloaded resin for solid phase peptide synthesis (SPPS) were purchased from Novabiochem Merck Biosciences. Water and acetonitrile for (RP)HPLC purification were of reagent or HPLC grade and used without further purification.

^{1}H and ^{13}C NMR spectra were recorded on Bruker AMX 300 and 75, 400 and 100, 500 and 125 MHz instruments. Reference NMR solvent signals $CDCl_3$: δ_H (ppm) = 7.26 ppm, δ_C (ppm) = 77.0 ppm; D_2O: δ_H (ppm) = 4.79 ppm, δ_C: sr = -32.10 Hz; $(CD_3)_2SO$: δ_H (ppm) = 2.50 ppm, δ_C (ppm) = 39.5 ppm; CD_3OD: δ_H (ppm) = 3.31 ppm, δ_C (ppm) = 49.0 ppm; unless otherwise stated. Electrospray ionisation was carried out by direct infusion or liquid chromatography on Waters uPLC/SQD-LC mass spectrometer. A linear gradient from 95 % solvent A (water containing 0.1 % formic acid) to 95 % solvent B (acetonitrile containing formic acid) over 8 min was employed.

Semi-preparative (RP)HPLC was carried out using a C_{18} column (diameter = 10 mm × 250 mm). The HPLC method used involved a semi-preparative column: flow rate 3.0 mL/min, detection at 214, 230, 254, and 280 nm. Gradient: from 5 % MeCN/ 95 % H_2O (0.1 % TFA) to 50 % MeCN/ 50 % H_2O (0.1 % TFA) over 45 min.

Analytical thin layer chromatography (TLC) analysis was conducted by visualisation of Merck-aluminium backed plates coated with silica gel under ultraviolet light (254 nm), followed by staining with *p*-anisaldehdye. Flash column chromatography was carried out with silica, SiO_2, 60 Å particle size, 35-70 µm in diameter. Where petroleum ether is stated, the fraction that boils at 40-60 °C was used.

Automated solid phase peptide synthesis

Solid phase peptide elongation, where stated, was conducted on an Applied Biosystems 433A automated synthesiser. Fulvene conductivity was monitored as a quality control step to delay subsequent activation and coupling steps until the conductivity was within 10 % error. Coupling reagents used: 0.45 M HBTU/HOBt in DMF, NMP for washes, 20 % v/v piperidine in DMF for deprotection.

Solid phase peptide synthesis

Manual SPPS on a 0.05 mmol scale was performed using MBHA Rink amide resin (loading = 0.64 mmolg^{-1}); Fmoc-amino acids (0.50 mmol, 10 equivalents) were used with HBTU/HOBt (1.1 mL, 0.45 M), *N, N*-diisopropylethylamine (150 μL) and *N, N*-dimethylformamide (1.1 mL); unless otherwise stated. The average coupling time was 3 h, and the reactions were monitored *via* Kaiser Ninhydrin test. Fmoc deprotection from the resin-bound peptide chain was achieved by treatment with 20 % v/v piperidine in DMF (1.0 mL) for 5 min. The resin was washed exhaustively with *N, N*-dimethylformamide (1.0 mL) then dichloromethane (1.0 mL) for 5 min each, between each coupling and deprotection step.

Analytical Kaiser Ninhydrin test

To the resin-bound peptide (~1.0 mg) two drops of Kaiser Ninhydrin test solutions I (0.001 M KCN, 2 % v/v pyridine), II (5 % w/v Ninhydrin in *t*-BuOH), and III (80 % v/v phenol in *t*-BuOH) were added and the reaction vessel placed in a heat block at 100 °C for 5 min.[115] An indigo colour indicated a free amine, whereas yellow indicated a protected amine.[115]

Analytical peptide cleavage form solid-support

A TFA solution comprised of 100 μL of trifluoroacetic acid: ethanedithiol: water = 95: 2.5: 2.5 v/v/v was added to the resin-bound peptide (~1.0 mg) and left to stand at room temperature for 1 h. The beads were then filtered and diethyl ether (1.0 mL) was added to the

solution to initiate precipitation, which was then centrifuged. The diethyl ether was decanted and centrifugation was repeated with diethyl ether (1.0 mL) to afford a white solid.

6.1 Synthesis of cysteinylprolyl esters

2. H-(L)-Cys(Trt)-OH

(L)-Cysteine hydrochloride **1** (2.00 g, 12.69 mmol) and triphenylmethyl chloride (5.40 g, 19.37 mmol) were dissolved in anhydrous *N, N*-dimethylformamide (8.0 mL) and stirred under nitrogen for 24 h. Aqueous sodium acetate (10 % w/v) (70.0 mL) was added to the reaction mixture. The white precipitate that formed was filtered under vacuum and washed with water. The white precipitate was then stirred in acetone (50.0 mL) at 50 °C for 0.5 h. The resulting solution was cooled to room temperature, filtered with suction, washed with a small volume of acetone and dried under high vacuum to afford **2** (3.22 g, 70 %) as a white amorphous solid. R_f 0.1 (8: 2 ethyl acetate: methanol); ^{1}H NMR (400 MHz, $CDCl_3$ 7.26 ppm, T = 298 K) δ_H (ppm) 7.34 (6 H, br asym d, J = 7.7, Trt Ph-H_a); 7.15 (6 H, br asym t, J = 6.5, Trt Ph-H_b); 7.08 (3 H, br asym t, Trt Ph-H_c); 2.77 (1 H, br m, Cys $^{\alpha}$CH); 2.52 (1H, br m, Cys $^{\beta}$CH); 2.24 (1 H, br m, Cys $^{\beta}$CH); ^{13}C NMR (125 MHz, DMSO-d_6 40.5 ppm) δ_C (ppm) 145.2, 130.1, 129.0, 127.7 (Ph); 67.0; 54.3; 31.6; IR (υ_{max}) neat: 3053, 1595, 1546, 1489, 1441, 1394, 1335, 1300 cm^{-1}; ESI$^+$ MS (*m/z*) calculated for $C_{22}H_{21}O_2NS$: 363.13, observed $[MH]^+$ 364.14.

4. Fmoc-Gly-Cys(Trt)-OH

(i) HOSu, dry DMF, EDCI.HCl
(ii) H_2N-Cys(STrt)-COOH, DIPEA

3 → 4

Fmoc-Gly-OH **3** (595 mg, 2.00 mmol) and *N*-hydroxysuccinimide (HOSu) (230 mg, 2.00 mmol) were dissolved in anhydrous *N, N*-dimethylformamide (7.0 mL). *N*-(3-dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride (383 mg, 2.00 mmol) was then added, and the resulting mixture was stirred under nitrogen for 24 h. H-Cys(Trt)-OH, **2**, (800 mg, 2.20 mmol) and diisopropylethylamine (348 μL) were added to the reaction mixture and stirring continued for a further 6 h. The reaction was quenched with 1 M aqueous hydrochloric acid (2.0 mL) and the solution was concentrated *in vacuo*. Extraction ensued with ethyl acetate (50.0 mL), followed by washing with 1 M aqueous hydrochloric acid (2 × 10.0 mL) and brine (10.0 mL). The combined organic phase was dried with anhydrous magnesium sulfate, filtered and concentrated *in vacuo* to give the crude dipeptide as an off-white foam solid. The crude dipeptide was purified by flash column chromatography using (i) ethyl acetate; (ii) 8: 2 ethyl acetate: methanol to afford **4** (0.23 g, 18 %) as an off-white foam solid. R_f = 0.1 (8:2 ethyl acetate: methanol); ^{1}H NMR (400 MHz, $CDCl_3$ 7.29 ppm) δ_H (ppm) 7.66 (2 H, br m, Fmoc Ph-H); 7.48 (2 H, br m, Fmoc Ph-H); 7.27 (19 H, br m, Fmoc 4 × Ph-H, Trt 15 × Ph-H); 4.12 (4 H, br m, Fmoc CH_2, Fmoc CH, Cys $^{\alpha}$CH); 3.72 (2 H, br m, Gly $^{\alpha}CH_2$); 2.73 (2 H, br m, Cys $^{\beta}CH_2$); IR (υ_{max}) neat: 3371, 3058, 2960, 2925, 2853, 1711, 1654, 1595, 1520, 1445, 1405, 1260 cm^{-1}; ESI$^+$ MS (*m/z*) calculated for $C_{39}H_{34}O_5N_2S$: 642.22, observed $[MNa]^+$ 665.34.

6. Fmoc-Ala-Cys(Trt)-OH

(i) HOSu, dry DMF, EDCI.HCl
(ii) H-Cys(STrt)-OH (H_2N, COOH), DIPEA

5 → **6**

Fmoc-Ala-OH **5** (623 mg, 2.00 mmol) and *N*-hydroxysuccinimide (230 mg, 2.00 mmol) were dissolved in anhydrous *N,N*-dimethylformamide (7.0 mL, 0.29 M). *N*-(3-dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride (383 mg, 2.00 mmol) was then added to the reaction mixture, which was then stirred under nitrogen for 24 h. H-Cys(Trt)-OH, **2**, (800 mg, 2.20 mmol) and diisopropylethylamine (348 μL) were then added to the reaction mixture and stirring continued for a further 6 h. The reaction was quenched with 1 M aqueous hydrochloric acid (2.0 mL) and the solution was concentrated *in vacuo*. Extraction ensued with ethyl acetate (50.0 mL), followed by washing with 1 M aqueous hydrochloric acid (2 × 10.0 mL) and brine (10.0 mL). The combined organic phase was dried with anhydrous magnesium sulfate, filtered and concentrated *in vacuo* to give the crude dipeptide. The crude dipeptide was purified by flash column chromatography using (i) ethyl acetate; (ii) 8: 2 ethyl acetate/ methanol afforded **6** (1.02 g, 78 %) as an off-white foam solid. R_f = 0.2 (8:2 ethyl acetate/ methanol); ^{1}H NMR (400 MHz, $CDCl_3$ 7.24 ppm, T = 328 K) δ_H (ppm) 7.72 (2 H, d, *J* = 7.9, Fmoc Ph-H); 7.52 (2 H, d, *J* = 6.3, Fmoc Ph-H); 7.25 (19 H, m, Fmoc 4 × Ph-H, Trt 15 × Ph-H); 6.31 (1 H, br m, CONH); 5.28 (1 H, br m, CONH); 4.36 (3 H, br m, Fmoc CH_2, Fmoc CH); 4.11 (2 H, br m, Cys $^{\alpha}$CH, Ala $^{\alpha}$CH); 2.73 (2 H, br m, Cys $^{\beta}CH_2$); 1.27 (3 H, br d, *J* = 6.0, Ala $^{\beta}CH_3$); ^{13}C NMR (100 MHz, $CDCl_3$, T = 328 K) δ_C (ppm) 170.8, 165.3 (C=O); 144.7, 144.0, 141.3, 134.0, 129.6, 129.6, 127.8, 127.5, 127.0, 126.6, 125.1, 119.8 (Ph); 67.2; 63.0; 60.2; 47.2; 20.8; 14.1 (CH_3); IR (υ_{max}) neat: 3389, 3292, 3058, 2927, 1711, 1670, 1491, 1446, 1319, 1243 cm^{-1}; ESI$^+$ MS (*m/z*) calculated for $C_{40}H_{36}O_5N_2S$: 656.30, observed $[MNa]^+$ 679.52.

8. HO-Gc-Rink amide resin

Fmoc-Rink amide AM resin **7** (0.156 g, 0.10 mmol) was subjected to 20 % piperidine in DMF (v/v) (1.0 mL) for 5 min, and then washed exhaustively with *N, N*-dimethylformamide then dichloromethane. Glycolic acid (Gc) (38 mg, 0.50 mmol) and 1-hydroxy-1*H*-benzotriazole (HOBt) (40 mg, 0.50 mmol) were dissolved in anhydrous *N, N*-dimethylformamide (3.0 mL) and DIC (47 μL, 0.50 mmol) was added before being transferred to the dry resin. The reaction mixture was shaken at 410 rpm, at room temperature for 3 h, then filtered and washed with *N, N*-dimethylformamide (1.0 mL) and dichloromethane (2 × 1.0 mL) to give **8**.

9. H-Pro-Gc-Rink amide resin

Fmoc-Pro-OH (0.169 g, 0.50 mmol) was then added to **8**, followed by HOBt/HBTU (1.1 mL, 0.45 M) and *N, N*-diisopropylethylamine (150 μL) in *N, N*-dimethylformamide (1.1 mL), the heterogeneous mixture was shaken for 4 h. The vessel was subsequently drained and washed with *N, N*-dimethylformamide (1.0 mL) and dichloromethane (2 × 1.0 mL), and the Fmoc-Pro-OH coupling procedure was repeated with shaking overnight. Fmoc-analysis was then carries out. This was achieved by measuring out (1) 2.0 mg of Fmoc-Pro-Gc-resin synthesised in 20 % piperidine in DMF (v/v) (3.0 mL), and (2) 2.1 mg of Fmoc-Pro-Gc-resin

synthesised in 20 % piperidine in DMF (v/v) (3.0 mL). A blank solution of 20 % piperidine in DMF (v/v) (3.0 mL) was initially placed in the UV spectrophotometer and the absorbance taken as the reference value. Sequentially the two aforementioned solutions (1) and (2) were also placed in the photometer until a stable reading was taken. The absorbance at 290 nm for solution (1) was 1.972 and for (2) was 1.929. These values were substituted in Equation 1 and determined the resin loading to be 0.54 mmolg^{-1} (85 %). Fmoc-Pro-Gc-Resin (0.0817 g, 0.05 mmol), **9**, was deprotected with 20 % piperidine in DMF (v/v) (1.0 mL) for 5 min, and then washed exhaustively with *N, N*-dimethylformamide then dichloromethane.

10. H-Ala-Cys(Trt)-Pro-Gc-Resin

Fmoc-Ala-Cys(Trt)-OH

HBTU/HOBt, DMF, DIPEA

9

10

Fmoc-Ala-Cys(Trt)-OH **6** (0.161 g, 0.25 mmol) was then coupled to the *N*-terminus of **9** using HOBt/HBTU (0.55 mL, 0.45 M), *N, N*-dimethylformamide (0.55 mL) and diisopropylethylamine (75 µL), and was shaken for 5 h. The resulting mixture was then filtered, washed with *N, N*-dimethylformamide (1.0 mL) then dichloromethane (2 × 1.0 mL). TFA solution [100 µL of trifluoroacetic acid: ethanedithiol: water (95: 2.5: 2.5)] was added to the resin-bound peptide (1.0 mg), which was left to stand at room temperature for 1h. The beads were then filtered and diethyl ether (1.0 mL) was added to the filtrate to initiate a white precipitate, which was then centrifuged. The diethyl ether was decanted and centrifugation was repeated with diethyl ether (1.0 mL) to afford a white solid. The solid was then dissolved in water for analysis. ESI+ LC-MS (m/z) calculated for Fmoc-ACP-Gc-CONH$_2$ 568.2, found [MNa]$^+$ 591.3. The remaining resin-bound peptide was then subjected to 20 % piperidine in DMF (v/v) (1.0 mL) for 8 min, filtered and washed with *N, N*-dimethylformamide (1.0 mL) then dichloromethane (2 × 1.0 mL) to afford the deprotected, resin-bound tripeptide **10**, *H-Ala-Cys(Trt)-Pro-Gc-Resin*.

11. H-AENITTGCAEACP-Gc-CONH$_2$: EPO(22-32)-CPE

(i) Automated Peptide Synthesis (10 eq) Fmoc-Xaa-OH; (ii) TFA : EDT : H_2O 95 : 2.5 : 2.5

10 → H-AENITTGCAEACP-Gc-CONH$_2$ **11**

H-Ala-Cys(Trt)-Pro-Gc-Resin, **10**, was transferred to an Applied Biosystems 433A reaction vessel and peptide elongation was performed in an automated fashion on an Applied Biosystems 433A automated synthesiser, using the Fastmoc protocol. Fmoc-Xaa-OH used (10 eq of each): Fmoc-Ala-OH, Fmoc-Glu(O^{t}Bu)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Ile-OH, Fmoc-Thr(^{t}Bu)-OH, Fmoc-Gly-OH and Fmoc-Cys(Trt)-OH. Ile was double coupled. The peptide was then cleaved from the solid-support by treatment with 95 % trifluoroacetic acid solution [3 mL (trifluoroacetic acid: ethanedithiol: water = 95: 2.5: 2.5)] for 4.5 h. The mixture was filtered, and finally rinsed with neat TFA (1.0 mL). Diethyl ether (20.0 mL) was added to the filtrate and centrifuged (15 min), this was then repeated with diethyl ether (20.0 mL). The crude peptide was a white precipitate and it was lyophilised from 1: 1 = acetonitrile: water, and then purified *via* semi-preparative (RP)HPLC (gradient: 5-50 % aqueous acetonitrile/ 45 min, t_R = 23.0 min). The collections were lyophilised to afford **11**, *H-AENITTGCAEACP-Gc-CONH$_2$* (12.5 mg, 22 % calculated from resin loading) as a fluffy white solid. ESI$^+$ LC-MS (*m/z*) calculated for **11** 1335.5, observed [MH]$^+$ 1336.5.

12. Attempted EPO(22-40)

H-A-E-N-I-T-T-G-C-A-E-A-C-P-Gc-CONH$_2$ **11** + H-C-S-L-N-E-N-I-T-OH —(6 M Gu.HCl, TCEP, MESNa/ MPAA, pH 7.0/ 8.0, 25 °C)—✗→ H-A-E-N-I-T-T-G-C-A-E-A-C-S-L-N-E-N-I-T-OH **12**

EPO(22-32)-CP-Gc-CONH$_2$, **11** (3 × 1.0 mg, 0.75 μmol) and EPO(33-40)-OH (3 × 1.0 mg, 1.12 μmol) were dissolved in 6 M guanidinium hydrochloride buffer (3 × 500 μL), adjusted to either pH 7.0 or 8.0, before the addition of the thiol additive 1 M MESNa (25 μL, 50 mM)

or 1 M MPAA (25 µL, 50 mM) and reductant 1 M TCEP (5 µL,10 mM) (Table 6.1). The reactions were conducted in a thermomixer at 25 °C and monitored at regular intervals; from t = 1 to 122 h, via LC-MS. No ligation product was observed.

Table 6.1 Model peptide ligation conditions

Reagent	**Ligation conditions 1***	**Ligation conditions 2***	**Ligation conditions 3***
EPO(22-32)-CPE	1.0 mg, 0.75 µmol	1.0 mg, 0.75 µmol	1.0 mg, 0.75 µmol
EPO(33-40)-OH	1.0 mg, 1.12 µmol	1.0 mg, 1.12 µmol	1.0 mg, 1.12 µmol
6 M Guanidinum.HCl ($Na_xH_yPO_4$) buffer	500 µL	500 µL	500 µL
pH	7.0	8.0	7.0
MPAA	25 µL, 50 mM	-	25 µL, 50 mM
MESNa	-	25 µL, 50 mM	-
TCEP	5 µL, 10 mM	5 µL, 10 mM	5 µL, 10 mM

* Ligation between EPO(22-32)CPE + EPO(33-40) at 25 °C with LC-MS monitoring from t = 1 to 122 h.

13a. H-EPO(22-28)-$SCH_2CH_2CO_2H$: EPO(22-28)-MPA

H-AENITTGCAEACP–C(=O)–O–CH_2–C(=O)–NH_2 **11** → (40 % MPA (aq), M.W. 150 W, 80 °C) → H-AENITTG–C(=O)–S–CH_2CH_2–C(=O)OH **13a**

EPO(22-32)-CP-Gc-$CONH_2$ **11** (1.0 mg, 0.75 µmol) was dissolved in 3-mercaptopropionic acid (MPA) (200 µL) and water (300 µL) and placed in a microwave reactor at 80 °C, 150 W, for 30 min. LC-MS indicated the presence of starting material and hence the reaction was continued for a further 25.5 h. The reaction mixture was subsequently filtered and centrifuged prior to purification *via* (RP)HPLC (gradient: 5-50 % aqueous acetonitrile/ 45 min). Fractions collected were lyophilised to afford **13a**, *EPO(22-28)-$SCH_2CH_2CO_2H$* (< 1 %) as a white solid. LC-MS ESI^+ MS (*m/z*) calculated for [M] 792.3, observed $[MH]^+$ 793.4.

13b. EPO(22-28)-$SCH_2CH_2CO_2H$: EPO(22-28)-MPA

H-AENITTGCAEACP–C(=O)–O–CH2–C(=O)–NH2 (**11**) → [30 % MPA (aq); M.W. 150 W; 80 °C] → H-AENITTG–C(=O)–S–CH2CH2–C(=O)OH (**13b**)

EPO(22-32)-CP-Gc-$CONH_2$, **11** (1.0 mg, 0.75 µmol) was dissolved in 3-mercaptopropionic acid (MPA) (150 µL) and water (350 µL) and placed in a microwave reactor at 80 °C, 150 W, for 4 h. LC-MS indicated the presence of starting material and hence the reaction was continued for a further 12 h. The reaction mixture was subsequently filtered and centrifuged and then purified *via* (RP)HPLC (gradient: 5-50 % aqueous acetonitrile/ 45 min). Fractions collected were lyophilised to afford **13b**, *EPO(22-28)-MPA* (0.5 mg, 84 %) as a white solid. LC-MS ESI^+ MS (*m/z*) calculated for [M] 792.3, observed $[MH]^+$ 793.4.

14. H-AENITTGCSLNENIT-OH: EPO(22-28)-(33-40)

H-AENITTG–C(=O)–S–CH2CH2–C(=O)OH (**13**) → [6 M Gu.HCl buffer; EPO(33-40); 2 % MESNa; 30 mM TCEP; pH 8.0] → H-AENITTG–C(=O)–NH–CSLNENIT (**14**)

EPO(22-28)-$SCH_2CH_2CO_2H$, **13a** (< 1 %) was dissolved in guanidinium hydrochloride buffer (6 M, pH 7.0) (100 µL), and added to EPO(33-40) (1.0 mg, 1.12 µmol). The reaction was conducted in a thermomixer at 25 °C, 600 rpm, initially reacted for 24 h and monitored via LC-MS, which indicated presence of product at t = 25 h but also the oxidised product. MESNa (1 M) (2 % v/v) and TCEP (1 M) (30 mM) were added to the reaction mixture, the pH was adjusted to 8.0 and the reaction continued for a further 27 h. LC-MS, t = 51 h, indicated total consumption of the starting material, and the reaction was terminated to afford the sequence **14**, *EPO(22-28)-(33-40)*. LC-MS ESI^+ MS (*m/z*) calculated for $[MH]^+$ 1578.8, observed $[MH]^+$ 1578.8.

16. H-AENITTGC(Acm)AEACP-Gc-$CONH_2$: EPO(22-^{29}C(Acm)- 32)-CPE

(i) Automated Peptide Synthesis (10 eq) Fmoc-Xaa-OH
(ii) TFA : EDT : H_2O 95 : 2.5 : 2.5

15

H-A-E-N-I-T-T-G-C(Acm)-A-E-A-C-P-Gc-$CONH_2$

16

H-Ala-Cys(Trt)-Pro-Gc—Resin, **15**, was transferred to an Applied Biosystems 433A reaction vessel and peptide elongation was performed in an automated fashion on an Applied Biosystems 433A automated synthesiser, using the Fastmoc protocol. Fmoc-Xaa-OH used (10 eq of each): Fmoc-Ala-OH, Fmoc-Glu(O^{t}Bu)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Ile-OH, Fmoc-Thr(^{t}Bu)-OH, Fmoc-Gly-OH, Fmoc-Cys(Acm)-OH. Ile was double coupled. The peptide was then cleaved from the solid-support by treatment with 95 % trifluoroacetic acid solution [3.0 mL (trifluoroacetic acid: ethanedithiol: water = 95: 2.5: 2.5)] for 4.5 h. The mixture was filtered, and finally rinsed with neat TFA (1.0 mL). Diethyl ether (20.0 mL) was added to the filtrate and centrifuged (15 min), and this was then repeated with diethyl ether (20.0 mL). The crude peptide was a white precipitate and it was lyophilised from 1: 1 = acetonitrile: water, and then purified *via* semi-preparative (RP)HPLC (gradient: 5-50 % aqueous acetonitrile/ 45 min, t_R = 22.0 min). The collections were lyophilised to afford **16**, *EPO(22-^{29}C(Acm)- 32)-CPE* (11.7 mg, 20 % calculated from resin loading) as a fluffy white solid. LC-MS ESI$^+$ MS (*m/z*) calculated for **16**: 1406.6, observed $[MH]^+$ 1407.8.

17. H-AENITTGC(Acm)AEA-$COSCH_2CH_2CO_2H$: EPO(22-^{29}C(Acm)- 32)-MPA

16 → 17: 20 % MPA (aq), 80 °C

EPO(22-^{29}C(Acm)- 32)-CPE, **16** (1.0 mg, 0.71 μmol) was dissolved in 3-mercaptopropionic acid (MPA) (100 μL) and water (400 μL) and placed in a thermomixer at 80 °C, pH 1.0, until total consumption of starting material. The reaction mixture was subsequently filtered and

centrifuged and then purified *via* (RP)HPLC (gradient: 5 to 50 % acetonitrile in water, semi-preparative column). Fractions collected were lyophilised to afford **17**, *EPO(22-^{29}C(Acm)32)-MPA* (0.5 mg, 57 %) as a white solid. LC-MS ESI^{+} MS (*m/z*) calculated for [M] 1237.5, observed $[MH]^{+}$ 1238.7.

18. H-AENITTGC(Acm)AEACSLNENIT = EPO(22-^{29}C(Acm)- 40)

H-AENITTGC(Acm)AEA-C(O)-S-CH2CH2-COOH (**17**) → [6 M Gu.HCl buffer, EPO(33-40), 50mM MPAA, 40 mM TCEP, pH 8, 25 °C] → H-AENITTGC(Acm)AEA-C(O)-NH-CSLNENIT (**18**)

EPO(22-^{29}C(Acm)-32)-$SCH_2CH_2CO_2H$, **17** (0.7 mg, 0.50 µmol) was dissolved in guanidinium hydrochloride buffer (6 M, pH 7) (500 µL), and added to EPO(33-40) (1.0 mg, 1.12 µmol). MPAA (1 M) (50 mM) and TCEP (1 M) (40 mM) were added to the reaction mixture, the pH was adjusted to 8.0, the reaction was carried out at 25 °C in a thermomixer, and the reaction continued until starting material was consumed to afford the sequence **18**, *EPO(22-^{29}C(Acm)- 40).* LC-MS ESI^{+} MS (*m/z*) calculated for $[MH_2]^{2+}$ 1013.0, observed $[MH_2]^{2+}$ 1013.0.

19. H-AENITTGC(Acm)AEACSLNENIT-OH = EPO(22-^{29}C(Acm)- 40)

H-A-E-N-I-T-T-G-C(Acm)-A-E-A-C-P-Gc-$CONH_2$ (**16**) → [6 M Gu.HCl buffer, EPO(33-40), 50mM MPAA, 40 mM TCEP, pH 8.0, 25 °C] → H-AENITTGC(Acm)AEA-C(O)-NH-CSLNENIT (**19**)

EPO(22-^{29}C(Acm)-32)-CPE **16** (5.0 mg, 0.36 µmol) was dissolved in guanidine hydrochloride buffer (6 M, pH 7.0) (500 µL), and added to EPO(33-40) (1.0 mg, 1.12 µmol). MPAA (1 M) (50 mM) and TCEP (1 M) (40 mM) were added to the reaction mixture, the pH was adjusted to 8.0, the reaction was carried out at 25 °C in a thermomixer, and it was continued until starting material was deemed undetectable by LC-MS to afford the sequence

19, *EPO(22-^{29}C(Acm)- 40).* LC-MS ESI$^+$ MS (*m/z*) calculated for $[MH_2]^{2+}$ 1013.0, observed $[MH_2]^{2+}$ 1013.0.

20. H_2N-AENITTGC-NH_2

The model peptide was synthesised by solid phase peptide synthesis using MBHA Rink amide resin (0.15 g, 0.1 mmol), which was initially deprotected at its N-terminus by treatment with 20 % piperidine (v/v) in DMF (1.0 mL) and shaken for 5 min. The solution was then filtered and the beads were washed with DMF (1.0 mL) for 5 min and DCM (2 × 1.0 mL) for 5 min each and the resin dried before the second addition of 20 % piperidine (v/v) in DMF (1.0 mL), the resulting mixture was shaken for 15 min. The deprotected resin was washed successively with DMF (1.0 mL) for 5 min and DCM (2 × 1.0 mL) for 5 min each, and dried. A Kaiser-Ninhydrin test of a sample of resin beads (< 1.0 mg) indicated positive for a free amine. The first amino acid was then coupled manually to the solid support with the addition of Fmoc-Cys(Trt)-OH (293 mg, 0.5 mmol) to the resin followed by dry DMF (2.0 mL), HOBt/HBTU (0.45 M solution in dry DMF) (2.2 mL) and DIPEA (300 μL), the resulting heterogeneous mixture was shaken for 4 h, after which it was filtered, then washed with DMF (1.0 mL) for 5 min and DCM (2 × 1.0 mL) for 5 min each and the resin dried. Ninhydrin test indicated a protected N-terminus and the resin was transferred into a reaction vessel for peptide elongation by automated solid phase peptide synthesis, using 10 equivalents of the following: Fmoc-Ala-OH, Fmoc-Glu(O^{t}Bu)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Ile-OH (double coupled), Fmoc-Thr(^{t}Bu)-OH and Fmoc-Gly-OH. The elongated peptide was then cleaved from the solid support and the orthogonal protecting groups were simultaneously removed by treating the resin with a cocktail of TFA: EDT: H_2O = 95: 2.5: 2.5 (total volume 5.0 mL) for 5 h. The mixture was then filtered into two tubes, one then had the majority of the peptide and the other contained a final rinse of the resin beads with TFA (1.0 mL). Diethyl ether (20.0 mL) was then added to each of the tubes containing the filtrate, and the resulting solutions were vortexed and centrifuged (3000 rpm, 4 °C, 15 min).The supernatant diethyl ether was decanted and the process was repeated with more diethyl ether (20.0 mL) added to both tubes, and afforded the crude peptide as a white solid. The crude peptide was then dissolved in H_2O: MeCN (3:1) and purified by semi-preparative (RP)HPLC (gradient: 5-50 % MeCN/45 min), and lyophilised to afford the model peptide amide **20** (79.2

mg, 98 %) as a white solid. ^{1}H NMR (500 MHz, D_2O) δ_H (ppm) 4.65 (1 H, m, Cys $^{\alpha}$CH); 4.45 (1 H, dd, J = 5.1, 5.2, Asn $^{\alpha}$CH); 4.38 (1 H, d, J = 4.9, Ile-Thr $^{\alpha}$CH); 4.33 (1 H, dd, J = 5.8, Glu $^{\alpha}$CH); 4.30 (1 H, d, J = 4.4, Thr-Thr $^{\alpha}$CH); 4.21 (1 H, d, J = 7.7, Ile $^{\alpha}$CH); 4.18 (4 H, m, 2 × Thr $^{\beta}CH_2$); 4.04 (1 H, q, J = 7.1, Ala $^{\alpha}$CH); 3.94 (2 H, m, Gly $^{\alpha}CH_2$); 2.86 (2 H, m, Cys $^{\beta}CH_2$); 2.71 (2 H, m, Asn $^{\beta}CH_2$); 2.38 (2 H, m, Glu $^{\gamma}CH_2$); 1.97 (2H, m, Glu $^{\beta}CH_2$); 1.84 (1 H, m, Ile $^{\beta}$CH); 1.46 (3 H, d, J = 7.1, Ala $^{\beta}CH_3$); 1.39 (1 H, m, Ile $^{\gamma}CH_2$);1.12 (7 H, m, 2 × Thr CH_3, Ile $^{\gamma}CH_2$); 0.84 (3 H, d, J = 6.8, Ile $^{\gamma}CH_3$); 0.79 (3 H, t, J = 7.4, Ile $^{\delta}CH_3$); LC-MS ESI$^+$ MS (*m/z*) calculated for $C_{31}H_{54}N_{10}O_{13}S$: 806.4, observed: 807.4 $[MH]^+$.

21. H-AENITTG-$SCH_2CH_2CO_2H$ = H-AENITTG-MPA

The model peptide amide **20** (11.7 mg, 14.5 µmol) was dissolved in water (450 µL) and 3-mercaptopropionic acid (50 µL, 10 % v/v) was added. The resulting acidic solution (pH 1.3) was vortexed and placed in a thermomixer at 60 °C for 48 h. After which the mixture was directly purified by (RP)HPLC (gradient: 5-50 % MeCN/45 min) and the eluted peptide thioester (t_R = 16.1 min) and unreacted peptide **20** (t_R = 17.3 min) were lyophilised to afford the unreacted starting peptide amide (3.1 mg, 26 % based on starting peptide amide) as a white solid, and the desired peptide thioester **21** (3.3 mg, 29 %) as a white solid. ^{1}H NMR (500 MHz, D_2O) δ_H (ppm) 4.33 (1 H, d, J = 5.2, Thr $^{\alpha}$CH); 4.29 (2 H, m, Glu $^{\alpha}$CH, Thr $^{\alpha}$CH); 4.15 (5 H, m, Ile $^{\alpha}$CH, 2 × Thr $^{\beta}CH_2$); 4.04 (1 H, br d, J = 2.8, Ala $^{\alpha}$CH); 3.99 (2 H, m, Gly $^{\alpha}CH_2$); 3.01 (2 H, t, J = 6.5, MPA CH_2); 2.70 (2 H, m, Asn $^{\beta}CH_2$); 2.58 (2 H, t, J = 7.3, MPA CH_2); 2.40 (2 H, m, Glu $^{\gamma}CH_2$); 2.01 (2H, m, Glu $^{\beta}CH_2$); 1.83 (1 H, m, Ile $^{\beta}$CH); 1.41 (3 H, d, J = 7.1, Ala $^{\beta}CH_3$); 1.38 (1 H, m, Ile $^{\gamma}CH_2$); 1.10 (7 H, br m, 2 × Thr CH_3, Ile $^{\gamma}CH_2$); 0.84 (3 H, d, J = 6.8, Ile $^{\gamma}CH_3$); 0.78 (3 H, t, J = 7.5, Ile $^{\delta}CH_3$); LC-MS ESI$^+$ MS (*m/z*) calculated for $C_{31}H_{52}N_8O_{14}S$: 792.3, observed: 793.4 $[MH]^+$.

Optimisation and First Biological Application

23. Fmoc-Gly(^{13}C-1)-OH

Glycine(^{13}C-1) **22** (0.50 g, 6.57 mmol) was dissolved in water (6.26 mL), and triethylamine (0.92 mL, 6.57 mmol) was added whilst stirring. Fmoc-succinimide (2.13 g, 6.31 mmol) was then dissolved at 50 °C in acetonitrile (6.26 mL), and this solution was then added to the amino acid solution in one portion. Further triethylamine (0.50 mL) was added to maintain a pH between 8.5 and 9.0. The combined reaction mixture was stirred at room temperature for 0.5 h, then it was concentrated *in vacuo* and the off-white concentrate was poured whilst stirring into 1.5 M hydrochloric acid (26.3 mL). The solution was then filtered under gravity and the white solid was collected. The aqueous filtrate was further extracted with dichloromethane (3 × 12.5 mL). To these combined organic extracts was added the white solid that was collected after filtration, and this organic layer was then washed with 1.5 M hydrochloric acid (12.5 mL), water (12.5 mL), and saturated aqueous sodium chloride (12.5 mL). The organic extract was then dried with anhydrous magnesium sulfate and the solvent was removed *in vacuo*. Petroleum ether (5.0 mL) was added to the concentrated product and formed a white solid after trituration. The product was then washed with petroleum ether and dried under high vacuum to afford Fmoc-Gly(^{13}C-1)-OH **23** (1.26 g, 67 %) as a white solid which was used without further purification. R_f = 0.1 (4 : 1 ethyl acetate: petroleum ether); m.p. 164-165 °C; ^{1}H NMR (500 MHz, $CDCl_3$) δ_H (ppm) 7.77 (2 H, d, J = 7.5, Fmoc Ph-H);

7.60 (2 H, br d, J = 7.3, Fmoc Ph-H); 7.40 (2 H, t, J = 7.3, J = 7.5, Fmoc Ph-H); 7.32 (2 H, m, Fmoc Ph-H); 5.26 (1 H, br s, CONH); 4.43 (2 H, d, J = 6.9, Fmoc C$\underline{H_2}$-O); 4.24 (1 H, asym t, J = 6.8, J = 6.6, Fmoc C$\underline{H}$-CH_2); 4.06 (2 H, br m, Gly $^{\alpha}CH_2$); ^{13}C NMR (125 MHz, $CDCl_3$) δ_C (ppm) 172.4 (CO_2H); IR (υ_{max}) neat: 3313, 2964, 2923, 1703, 1677, 1538, 1446, 1400, 1385, 1272 cm^{-1}; calculated mass for $[MH]^+$ 299.10793, observed (high resolution accurate mass) (CI): $[MH]^+$ 299.10844.

26. H-AENITTG($^{13}C_1$)C-NH_2

General peptide synthesis procedure: Rink amide-MBHA resin, loading = 0.64 $mmolg^{-1}$, (156 mg, 0.1 mmol) was initially treated with 20 % piperidine in DMF (v/v) (1.0 mL, 5 min). The solution was then filtered off and the resin was washed with DMF (1.0 mL, 5min) and twice with DCM (1.0 mL, 5 min). This was then repeated with 20 % piperidine in DMF (v/v) (1.0 mL, 15 min). The dried resin was then coupled to Fmoc-Cys(Trt)-OH (293 mg, 0.5 mmol) with DIPEA (300 μL), 0.45 M HOBt/HBTU (2.2 mL) and dry DMF (2.0 mL), and the mixture was left to shake for 4 h. The solution was then filtered off and the resin was washed once with DMF (1.0 mL, 5 min) and twice with DCM (1.0 mL, 5 min). The resin was then treated with 20 % piperidine in DMF (v/v) (1.0 mL, 5 min then 15 min), washed and then dried. The ^{13}C-labelled amino acid **23** (149 mg, 0.50 mmol) was then coupled using DIPEA (300 μL), 0.45 M HOBt/HBTU (2.2 mL) and dry DMF (2.0 mL), and the mixture was left to shake overnight. The resin was washed and dried, and then transferred to an automated peptide synthesiser reaction vessel for peptide chain elongation employing 10 equivalents of Fmoc-Xaa-OH = Fmoc-Ala-OH, Fmoc-Glu(O^{t}Bu)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Ile-OH (double coupled) and Fmoc-Thr(^{t}Bu)-OH. The resin was then treated with trifluoroacetic acid: ethanedithiol: water (95: 2.5: 2.5 = 5.0 mL) for 5 h. After which, the resin was filtered off and to the filtrate cold diethyl ether (20.0 mL) was added, which induced precipitation, and this was centrifuged at 3000 rpm, 4 °C for 15 min. The ether layer was then decanted and

this was repeated with cold diethyl ether (20.0 mL). The white precipitate was then dissolved in 20 % acetonitrile and purified by semi-preparative (RP)HPLC (gradient: 5-50 % acetonitrile/ 45 min). The labelled peptide fractions were lyophilised to afford peptide amide, H-AENITTG(^{13}C-1)C-NH_2, **26** (67 mg, 83 %; t_R = 17.2 min) as a white solid. ^{13}C NMR (125 MHz, $CDCl_3$) δ_C (ppm) 171.8 (CONH). LC-MS ESI$^+$ MS (*m/z*) calculated mass for **26** 807.4, observed $[MH]^+$ 808.4.

28. 4-Mercaptobutyric acid (MBA)

4-Bromobutyric acid **27** (4.0 g, 24.10 mmol) was refluxed for 3 h with thiourea (2.06 g, 27.10 mmol) in ethanol (7.30 mL). Aqueous 4 M sodium hydroxide (6.10 mL) was then added and refluxing was continued for another 2 h. The solution was concentrated *in vacuo*, extracted with ether and the ether extract was discarded, acidified with aqueous 6 M hydrochloric acid, and extracted with ether. The ethereal extract was then dried over anhydrous sodium sulfate, the ether was evaporated to give **28** (2.89 g, 100 %) as a pungent colourless oil in quantitative yield. ^{1}H NMR (500 MHz, $CDCl_3$) δ_H (ppm): 11.60 (1 H, br s, CO_2H), 2.52 (2 H, m, C$\underline{H}_2$-CO_2H), 2.44 (2 H, asym t, J = 7.3 Hz, CH_2-C$\underline{H}_2$-CH_2), 1.87 (2 H, m, C$\underline{H}_2$-SH), 1.31 (1 H, t, J = 8.1 Hz, SH); ^{13}C NMR (125 MHz, $CDCl_3$) δ_C (ppm): 179.5 (CO_2H), 32.2 ($\underline{C}H_2$-CO_2H), 28.4 (CH_2-$\underline{C}H_2$-CH_2), 23.6 ($\underline{C}H_2$-SH); LC ESI$^+$ MS (*m/z*) calculated for $C_4H_8O_2S$: 120.0, observed: $[MNa]^+$ 143.2.

29. H-AENITTG($^{13}C_1$)-COSR

General procedure for optimisation studies: ^{13}C-labelled peptide amide **26** (6.0 mg, 7.43 μmol) was dissolved in a total reaction volume of 1.2 mL, which was comprised of D_2O and

10 % thiol (RSH = 3-mercaptopropionic acid (MPA), sodium 2-mercaptoethanesulfonate (MESNa), 2-mercaptoacetic acid (MAA), 4-mercaptobutyric acid (MBA), or *β*-mercaptoethanol (BME)). Certain reactions also required additives: 1 % (v/v) trifluoroacetic acid (TFA), 1 % (v/v) acetic acid (AcOH), 0.5 % (w/v) *tris*(2-carboxyethyl)phosphine (TCEP), or 0.5 % (v/v) *N*-methyl imidazole (Me-Im). After acquisition of a ^{13}C NMR at 0 h, the reaction mixture was placed in a thermomixer at 60 °C, or in a microwave reactor at either 40, 50, or 60 °C (150 W) with low magnetic stirring, and the reaction was monitored by ^{13}C NMR at regular intervals: 6, 24, 48, 72 h, and analysed by LC-MS upon reaction completion. The ^{13}C NMR spectra for the thioester optimisation studies were acquired at 125 MHz, number of scans = 256, reference = -32.10 Hz (D_2O).

Biological application of novel peptide thioesterification method: synthesis of Ser22His human *β*-defensin 3 (HBD3)

1. FmocHN-Cys(Trt)-OH, *i*-Pr_2EtN, 0.45 M HOBt/HBTU, dry DMF
2. Fmoc-SPPS
3. TFA : EDT : H_2O, 5 h (95 : 2.5 : 2.5)
50 %

30 HBD3 (1-21)

10 % (w/v) MESNa, 10 % (v/v) AcOH, 0.5 % (w/v) TCEP, 6 M Guanidine.HCl, 60 °C, 48 h, N_2 (g) (83 % = **7** + **8**)

31 HBD3 (1-21)

32

1. Fmoc-SPPS
2. TFA : EDT : H_2O, 5 h (95 : 2.5 : 2.5)
37 %

33 HBD3 (24-45)

6 M Guanidine.HCl, 0.2 M Na phosphate, 0.1 M MESNa, 40 mM TCEP, pH 7.0, 25 °C, 48 h
97 %

34 HBD3 (1-45)

1. 10 % AcOH, 50 equiv Ag(I)OAc, 25 °C, 6 h
2. 10 % (w/v) DTT, 25 °C
95 %

35 HBD3 (1-45)

0.1 M $NaHCO_3$, 1 M Guanidine.HCl, 3.0 mM cysteine, 0.3 mM cystine, pH 8.1, 25 °C, 24 h
80 %

36 HBD3 (1-45)

30. HBD3(1-21)HC-NH$_2$: H-GIINTLQKYYC(Acm)RVRGGRC(Acm)AVLHC-NH$_2$

NHFmoc

1. 20 % piperidine in DMF (v/v)
2. Fmoc-Cys(Trt)-OH, *i*-Pr$_2$EtN, HOBt/HBTU, dry DMF, 4h
3. Fmoc-SPPS
4. TFA: EDT: H$_2$O = 95: 2.5: 2.5, 5 h

50 %

Acm Acm HBD3 (1-21) HS NH$_2$ NH **30**

Rink amide-MBHA low loading (LL) resin (147 mg, 0.05 mmol, loading 0.34 mmolg^{-1}) was treated with 20 % piperidine in DMF (v/v) (1.0 mL, 5 min). The resin was then filtered, washed once with DMF (1.0 mL, 5min) and twice with DCM (1.0 mL, 5 min), then dried and the process was repeated with 20 % piperidine in DMF (v/v) (1.0 mL, 15 min). The dried resin was then coupled to Fmoc-Cys(Trt)-OH (146 mg, 0.25 mmol) using DIPEA (150 µL), 0.45 M HOBt/HBTU (1.1 mL) and dry DMF (1.0 mL). The resulting mixture was left to shake for 4 h; the resin was then filtered, washed with DMF (1.0 mL, 5 min) and twice with DCM (1.0 mL, 5 min). The dried resin was then transferred to an automated peptide synthesiser (433 A) reaction vessel for peptide elongation employing 10 equivalents of Fmoc-Xaa-OH = Fmoc-Gly-OH, Fmoc-Ile-OH, Fmoc-Asn(Trt)-OH, Fmoc-Thr(^{t}Bu)-OH, Fmoc-Leu-OH, Fmoc-Gln(Trt)-OH, Fmoc-Lys(Boc)-OH, Fmoc-Tyr(^{t}Bu)-OH, Fmoc-Cys(Acm)-OH, Fmoc-Val-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Ala-OH, and Fmoc-His(Trt)-OH. All *β*-branched amino acids were double coupled. The resin was then treated with TFA: EDT: H$_2$O (95: 2.5: 2.5 = 5.0 mL) for 5 h, and the solution was then filtered and precipitation was induced upon addition of cold diethyl ether (20.0 mL) to the filtrate, which was then vortexed and centrifuged (3000 rpm, 4 °C, 15 min). After decanting the ether layer, this process was repeated with cold diethyl ether (20.0 mL). The crude peptide was then dissolved in 20 % acetonitrile and purified by semi-preparative (RP)HPLC (gradient: 5-50 % acetonitrile/ 45 min). The desired peptide fractions were lyophilised to afford the N-terminal defensin **31** (69 mg, 50 %, t_R = 32.3 min) as a white solid. LC-MS ESI$^+$ MS (*m/z*) calculated average mass [M] 2765.3, observed 2765.0.

31. HBD3(1-21)H-$SCH_2CH_2SO_3H$:

H-GIINTLQKYYC(Acm)RVRGGRC(Acm)AVLH-$SCH_2CH_2SO_3H$

30 → 31: 10 % (w/v) MESNa, 10 % (v/v) AcOH, 0.5 % (w/v) TCEP, 6 M Guanidine.HCl, 60 °C, 48 h, N_2 (g), 83 %

HBD3(1-21)HC-$CONH_2$ **30** (12.5 mg, 4.52 µmol, 4.0 mM) was dissolved in a solution of 6 M guanidine hydrochloride (1.02 mL) containing 10 % (w/v) sodium 2-mercaptoethanesulfonate (113 mg, 0.69 mmol) and 10 % (v/v) acetic acid (113 µL). Then, 0.5 % (w/v) tris(2-carboxyethyl)phosphine (5.7 mg, 19.88 µmol) was added and the resulting mixture, pH 2.0, was stirred under nitrogen at 60 °C for 48 h. LC-MS analysis at 48 h of an aliquot of the reaction mixture showed ~1:1 of **30**: **31**. The reaction mixture was then purified by semi-preparative (RP)HPLC (gradient: 5-50 % acetonitrile/ 45 min). The peptidic components were collected and lyophilised to afford a mixture of the unreacted starting material that co-eluted with the desired peptide thioester (**30** + **31**) (10.5 mg, 83 %, t_R = 32.1 and 32.5 min) as a white solid. Oxidised starting material (0.6 mg, 5 % (based on **30**), t_R = 34.1 min) was also recovered as a white solid. LC-MS ESI$^+$ MS (*m/z*) calculated average mass of **31** [M] 2787.3, observed 2787.0.

33. HBD3(23-45): H-CLPKEEQIGKC(Acm)STRGRKC(Acm)C(Acm)RRKK-OH

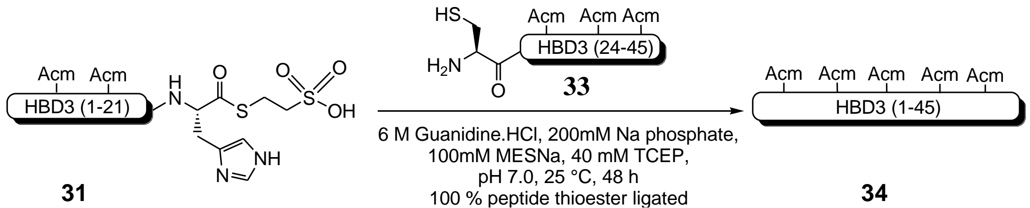
32 → 33: 1. Automated Fmoc-SPPS; 2. TFA: EDT: H_2O = 95: 2.5: 2.5, 5 h; 37 %

Preloaded Fmoc-Lys(Boc)-Novasyn® TGT resin **32** (250 mg, 0.05 mmol, loading 0.20 mmolg^{-1}) was transferred to an automated peptide synthesiser (433 A) for peptide elongation employing 10 equivalents of Fmoc-Xaa-OH = Fmoc-Cys(Acm)-OH, Fmoc-Leu-OH, Fmoc-Pro-OH, Fmoc-Lys(Boc)-OH, Fmoc-Glu(O^{t}Bu)-OH, Fmoc-Gln(Trt)-OH, Fmoc-Ile-OH, Fmoc-Gly-OH, Fmoc-Ser(^{t}Bu)-OH, Fmoc-Thr(^{t}Bu)-OH and Fmoc-Arg(Pbf)-OH. Boc-Cys(Trt)-OH was incorporated at the N-terminus and *β*-branched amino acids were double

coupled. The resin was then treated with TFA: EDT: H_2O (95: 2.5: 2.5 = 5.0 mL) for 5 h, and the solution was then filtered and precipitation was induced upon addition of cold diethyl ether (20.0 mL) to the filtrate, which was then vortexed and centrifuged (3000 rpm, 4 °C, 15 min). After decanting the ether layer, this process was repeated with cold diethyl ether (20.0 mL). The crude peptide was then dissolved in 20 % acetonitrile and purified by semi-preparative (RP)HPLC (gradient: 5-50 % acetonitrile/ 45 min). The desired peptide fractions were lyophilised to afford the C-terminal defensin **33** (54 mg, 37 %, t_R = 18.4 min) as a white solid. LC-MS ESI$^+$ MS (*m/z*) calculated average mass [M] 2921.5, observed 2922.0.

34. HBD3(1-45) + Acm_5:

H-GIINTLQKYYC(Acm)RVRGGRC(Acm)AVLHCLPKEEQIGKC(Acm)STRGRKC(Acm)C(Acm)RRKK-OH



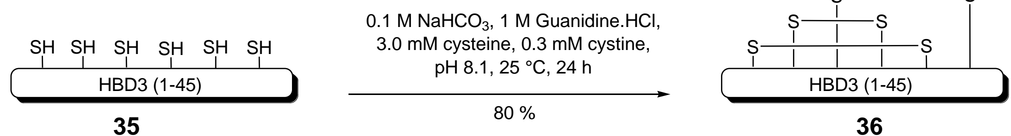
Co-purified starting amide **30** and peptide thioester **31** (10.6 mg, 3.80 mM), and excess cysteinyl peptide **34** (11.1 mg, 3.80 mM) were dissolved in degassed 6 M guanidine hydrochloride (660 µL).To this solution 1 M sodium phosphate pH 7.0 buffer (200 µL, 200 mM) was added, followed by 1 M sodium 2-mercaptoethanesulfonate (100 µL, 100 mM) and 1 M tris(2-carboxyethyl)phosphine (40 µL, 40 mM). The resulting mixture was vortexed and placed in a thermomixer at 25 °C, 800 rpm, for 48 h. At 48 h LC-MS/MS analysis showed the ligated full length peptide and the reaction mixture was purified by semi-preparative (RP)HPLC (gradient: 5-50 % acetonitrile/ 45 min). Defensin peptide fractions were lyophilised to afford unreacted cysteinyl peptide **33** (6.3 mg, 57 % (based on **33**), t_R = 18.4 min) as a white solid; ligated peptide **34** (6.8 mg, 97 %, t_R = 29.7 min) as a white solid; and unreacted starting amide **30** (7.1 mg, 67 % (based on **30**), t_R = 32.0 min) as a white solid. LC-MS ESI$^+$ MS (*m/z*) calculated average mass of **34** [M] 5566.6, observed 5566.0.

35. HBD3(1-45) [H]: H-GIINTLQKYYCRVRGGRCAVLHCLPKEEQIGKCSTRGRKCCRRKK-OH

Acm Acm Acm Acm Acm
HBD3 (1-45)
34
1. 10 % AcOH, 50 equiv Ag(I)OAc, 25°C, 6 h
2. 10 % (w/v) DTT, 25 °C
95 %
SH SH SH SH SH SH
HBD3 (1-45)
35

To a solution of full-length peptide **34** (6.8 mg, 1.22 μmol) in 10 % (v/v) acetic acid (120 μL), silver(I) acetate (10.2 mg, 61.11 μmol) (10 equivalents/Acm group) was added and the mixture was left to shake in a thermomixer at 25 °C for 6 h. Then 10 % (w/v) 1,4-dithio-DL-threitol (120 mg, 0.78 mmol) was added and the mixture was left to shake at 25 °C overnight. The resulting yellow solution was centrifuged and the supernatant was concentrated under reduced pressure. The precipitate formed *in vacuo* was dissolved in water and purified by semi-preparative (RP)HPLC (gradient: 5-50 % acetonitrile/ 45 min). The deprotected peptide was lyophilised to afford **35** (6.1 mg, 95 %, t_R = 31.6 min) as a white solid. LC-MS ESI^+ MS (*m/z*) calculated average mass [M] 5211.2, observed 5211.0.

36. Ser22His HBD3(1-45) [O]



The reduced peptide **35** (10.6 mg, 2.03 μmol) was dissolved in 6 M guanidine hydrochloride (7.1 mL, 0.29 mM), and then further diluted 6-fold with water (35.5 mL) to give final peptide concentration = 0.25 $mgmL^{-1}$, containing guanidine hydrochloride (1 M), 0.1 M sodium bicarbonate (358 mg, 4.26 mmol), 3.0 mM L-cysteine hydrochloride (20.1 mg, 0.13 mmol), 0.3 mM L-cystine (3.1 mg, 12.78 μmol), pH 8.1. The solution was gently shaken at 25 °C in a sealed vessel for 24 h. The reaction mixture was then concentrated by size-exclusion filtration (molecular weight cut-off = 3 kDa) to 1.0 mL and then diluted to 3.0 mL with water. LC-MS/MS analysis indicated that the reduced peptide had fully oxidised with an observed difference of 6 mass units. The concentrate was then purified by semi-preparative (RP)HPLC (gradient: 5-50 % acetonitrile/ 45 min). The three major peaks corresponding to folded peptide were collected and lyophilised to afford **36** (8.5 mg, 80 %, t_R = 30.2, 31.1, and

33.5 min) as a white solid. LC-MS ESI^+ MS (*m/z*) calculated average mass [M] 5205.2, observed 5205.0.

Antimicrobial activity assay

Test organisms (*Pseudomonas aeruginosa* strain PA01 and *Staphylococcus aureus* strain ATCC 25923) were grown to mid-logarithmic phase in Iso-Sensitest broth (Oxoid) growth media and then diluted to 1-5 × 10^6 colony forming units (CFU)/mL in 10 mM potassium phosphate containing 1 % (v/v) Iso-sensitest broth, pH 7.4. Different concentrations of test peptides were incubated in 100 µL of cells (1-5 × 10^6 CFU) at 37 °C for 3 h. Serial dilutions (10-fold) of the incubation mixture were spread on Iso-sensitest plates and incubated at 37 °C, and the CFU levels were determined the following day. The minimum bactericidal concentration (MBC) is the concentration of peptide at which we observed >99.99 % killing of the initial inoculum. All assays were performed in triplicate on three independent occasions. The MBC was obtained by taking the mean of all results, and all experimental errors were within one doubling dilution.

Chemotaxis assay

Chemotaxis was performed using a 96 well microchemotaxis chamber (Neuro Probe Inc, Gaithersburg USA). 30 µL of sample was loaded into the bottom chamber. The positive control human macrophage inflammatory protein-3α (hMIP-3α), also known as chemokine ligand-20 (CCL20), was purchased from PeproTech EC (London, UK) and dissolved in RPMI/ 1 % bovine serum albumin (BSA). The negative control was RPMI/ 1 % BSA (chemotactic media) alone. Stock solutions were prepared at 1 × 10^{-6} M. A polycarbonate filter (pore size 10 µm) (Neuro Probe Inc, Gaithersburg, USA) coated overnight with collagen (5 µg/mL) at 37 °C, was added to the top chamber against the silicon gasket and carefully placed over the bottom chamber to eliminate air bubbles. Human embryonic kidney cells transfected with chemokine receptor CCR6 were prepared in RPMI/ 1 % BSA and 200 µL was added to the wells in the top chamber at 1-5 × 10^5 cells/mL. The chamber was incubated in a humid 5 % CO_2 incubator for 4 h. From the chamber the filter was removed

and the side with the non-migrating cells was washed in PBS. The filter was allowed to air dry and was then stained with: solution A (methanol fixative) of DiffQuick for 2 min, solution B (Azure B) for 2 min, solution C (Eosin) for 2 min, and gently rinsed in sterile water and allowed to air dry. Three random fields of view were taken per well from which the migrated cells were counted, which allowed the mean number of migrated cells and the standard error of means (SEM) to be calculated. Each dilution was tested in triplicate and experiments were carried out a minimum of three times. Data was analysed by ANOVA with post-test when $p < 0.05$. Chemotactic activity was determined as the optimal concentration of test compound at which the highest chemotactic index was achieved.

The First Application of Native Chemical Thioesterification to the Synthesis of Native *N*-Linked Thioesters

38. 2-Acetamido-2-deoxy-3, 4, 6-tri-*O*-acetyl-*α*-D-glucopyranosyl chloride.

N-acetyl-D-glucosamine **37** (20.0 g, 90.46 mmol) was added to acetyl chloride (30.0 mL) and the solution was stirred overnight. The reaction was then diluted with chloroform (50.0 mL) and the resulting solution was washed with ice-water (50.0 mL) and ice-saturated aqueous sodium bicarbonate (50.0 mL). The organic phase was dried with anhydrous magnesium sulfate, filtered and concentrated under reduced pressure. The crude product was purified by flash column chromatography (the column was packed with 1:1 petroleum ether: ethyl acetate; mobile phase 1: 2 petroleum ether: ethyl acetate) to afford the glycosyl chloride **38** (15.52 g, 47 %) as a flaky white solid. R_f 0.4 (ethyl acetate); ^{1}H NMR (300 MHz, $CDCl_3$) δ_H (ppm) 6.15 (1 H, d, J = 3.7, H_{1e}); 6.03 (1 H, d, J = 8.7, NHAc); 5.30 (1 H, dd, J = 9.9, H_{3a}); 5.17 (1 H, dd, J = 9.9, H_{4a}); 4.51 (1 H, ddd, J = 9.9, J = 8.7, J = 3.7, H_{2a}); 4.24 (2 H, m, H_{5a}, H_{6a}); 4.09 (1 H, m, H_{6b}); 2.06, 2.00, 2.00, 1.95 (12 H, 4 × s, 4 × CH_3CO); ^{13}C NMR (75 MHz, $CDCl_3$) δ_C (ppm) 171.4, 170.6, 170.2, 169.1 (C=O); 93.7, 70.9, 70.1, 67.0 (CH); 61.1 (CH_2); 53.4 (CH); 23.0, 20.7, 20.7, 20.6 (CH_3); ESI$^+$ MS (*m/z*) calculated for $C_{14}H_{20}ClO_8N$: 365.09, observed $[MH]^+$ 366.13.

39. 2-Acetamido-2-deoxy-3, 4, 6-tri-*O*-acetyl-*β*-D-glucopyranosyl azide.

To a solution of the glucosyl chloride **38** (10.0 g, 27.39 mmol), *tetra*-butylammonium hydrogen sulfate (9.30 g, 27.39 mmol) and sodium azide (5.34 g, 82.17 mmol) in dichloromethane (110.0 mL), was added saturated aqueous sodium bicarbonate solution (110.0 mL). The resulting biphasic solution was stirred vigorously at room temperature for 1 h. Ethyl acetate (200.0 mL) was then added and the organic layer was separated and washed with saturated aqueous sodium bicarbonate (100.0 mL) and water (2 × 100 mL). The organic phase was then dried with anhydrous magnesium sulfate, filtered and concentrated under reduced pressure to afford the glycosyl azide **39** (8.36 g, 82 %) as a fluffy white solid. R_f 0.5 (2: 8 methanol: ethyl acetate); m.p. 169-170 °C [lit.[308] 166-167 °C]; ^{1}H NMR (500 MHz, $CDCl_3$) δ_H (ppm) 5.71 (1 H, d, J = 8.9, N$\underline{H}$Ac); 5.24 (1 H, dd, J = 9.5, J = 10.5, H_{3a}); 5.09 (1 H, dd, J = 9.6, J = 9.8, H_{4a}); 4.76 (1 H, d, J = 9.3, H_{1a}); 4.26 (1 H, asym dd, J = 4.9, H_{6b}); 4.16 (1 H, asym dd, J = 2.3, H_{6a}); 3.91 (1 H, ddd, J = 9.1, J = 10.5, J = 10.3, H_{2a}); 3.79 (1 H, ddd, J = 10.1, J = 4.8, J = 2.3, H_{5a}); 2.10, 2.03, 2.03 (3 H, 3 × s, 3 × CH_3CO); 1.97 (3 H, s, $C\underline{H_3}CONH$); ^{13}C NMR (125 MHz, $CDCl_3$) δ_C (ppm) 171.0, 170.7, 170.4, 169.2 (C=O); 88.4, 74.0, 72.1, 68.0 (CH); 61.8 (CH_2); 54.2 (CH); 23.2, 20.7, 20.6, 20.6 (CH_3); IR (υ_{max}) neat: 3342, 2942, 2881, 2103, 1746, 1660, 1529, 1371, 1226 cm^{-1}; ESI$^+$ MS (*m/z*) calculated for $C_{14}H_{20}O_8N_4$: 372.30, observed $[MNa]^+$ 395.12.

40. 2-Acetamido-2-deoxy-3, 4, 6-tri-*O*-acetyl-*β*-D-glucopyranosyl amine.

A solution of the glycosyl *β*-azide **39** (500 mg, 1.34 mmol) in anhydrous tetrahydrofuran (9.05 mL) was catalytically hydrogenated at atmospheric pressure in the presence of platinum(II) oxide (23 mg) for 3 h. The catalyst was then removed by filtration through

Celite® and the filtrate was concentrated under reduced pressure to afford the glycosyl *β*-amine **40** (448 mg, 97 %) as an off-white/grey solid. R_f 0.3 (2: 8 methanol: ethyl acetate); m.p. decomposed at 159 °C [lit.[308] 159 °C]; ^{1}H NMR (500 MHz, $CDCl_3$) δ_H (ppm) 5.83 (1 H, d, *J* = 9.2, N<u>H</u>Ac); 5.03 (2 H, m, H_{3a}, H_{4a}); 4.19 (1 H, asym dd, *J* = 4.9, *J* = 12.3, H_{6b}); 4.09 (2 H, m, H_{6a}, H_{2a}); 3.98 (1 H, d, *J* = 9.9, H_{1a}); 3.63 (1 H, m, H_{5a}); 2.11 (2 H, br s, NH_2); 2.07, 2.02, 2.00, 1.95 (12 H, 4 × s, 4 × CH_3CO); ^{13}C NMR (125 MHz, $CDCl_3$) δ_C (ppm) 171.5, 170.7, 170.7, 169.3 (C=O); 86.6, 73.4, 72.7, 68.5 (CH); 62.4 (CH_2); 54.8 (CH); 23.2, 20.7, 20.7, 20.6 (CH_3); IR (υ_{max}) neat: 3400, 3342, 2935, 1749, 1733, 1660, 1538, 1429, 1367, 1232 cm^{-1}; high resolution accurate mass (TOF MS ES$^+$) (*m/z*) calculated for $C_{14}H_{22}N_2O_8Na$: $[MNa]^+$ 369.1274, observed $[MNa]^+$ 369.1266.

41. *N*-*α*-9-Fluorenylmethoxycarbonyl-*N*-*β*-(2-*N*-acetylamido-2-deoxy-*β*-D-gluco-pyranosyl)-L-asparagine *tert*-butyl ester.

The glycosyl *β*-amine **40** (238 mg, 0.69 mmol) was added to a solution of Fmoc-Asp-O^{t}Bu (293 mg, 0.69 mmol) and *N*-ethoxycarbonyl-2-ethoxy-1,2-dihydroquinoline (171 mg, 0.69 mmol) in dry dichloromethane (21.0 mL, 0.025 M). The solution was stirred at room temperature for 3.5 h under nitrogen. TLC analysis indicated that the reaction was complete. The resulting pale-yellow/grey, almost brown, reaction mixture was then concentrated to afford a white solid. Hot ethyl acetate: petroleum ether 1: 1 (10.0 mL) was then added to the white solid and the suspension was filtered. The quinoline containing filtrate was discarded but the filtered solid was dried under vacuum to give the desired glycoamino acid **41** (399 mg, 78 %) as a white/ pale-grey solid. R_f 0.5 (ethyl acetate); m.p. decomposed at 210 °C; ^{1}H NMR [500 MHz, 17 % v/v $CDCl_3$ (7.258 ppm) + 83 % v/v MeOD (4.197 ppm)] δ_H (ppm) 7.54 (2 H, br d, *J* = 7.2, Fmoc Ph-H); 7.39 (2 H, br d, *J* = 7.2, Fmoc Ph-H); 7.17 (2 H, br t, *J*

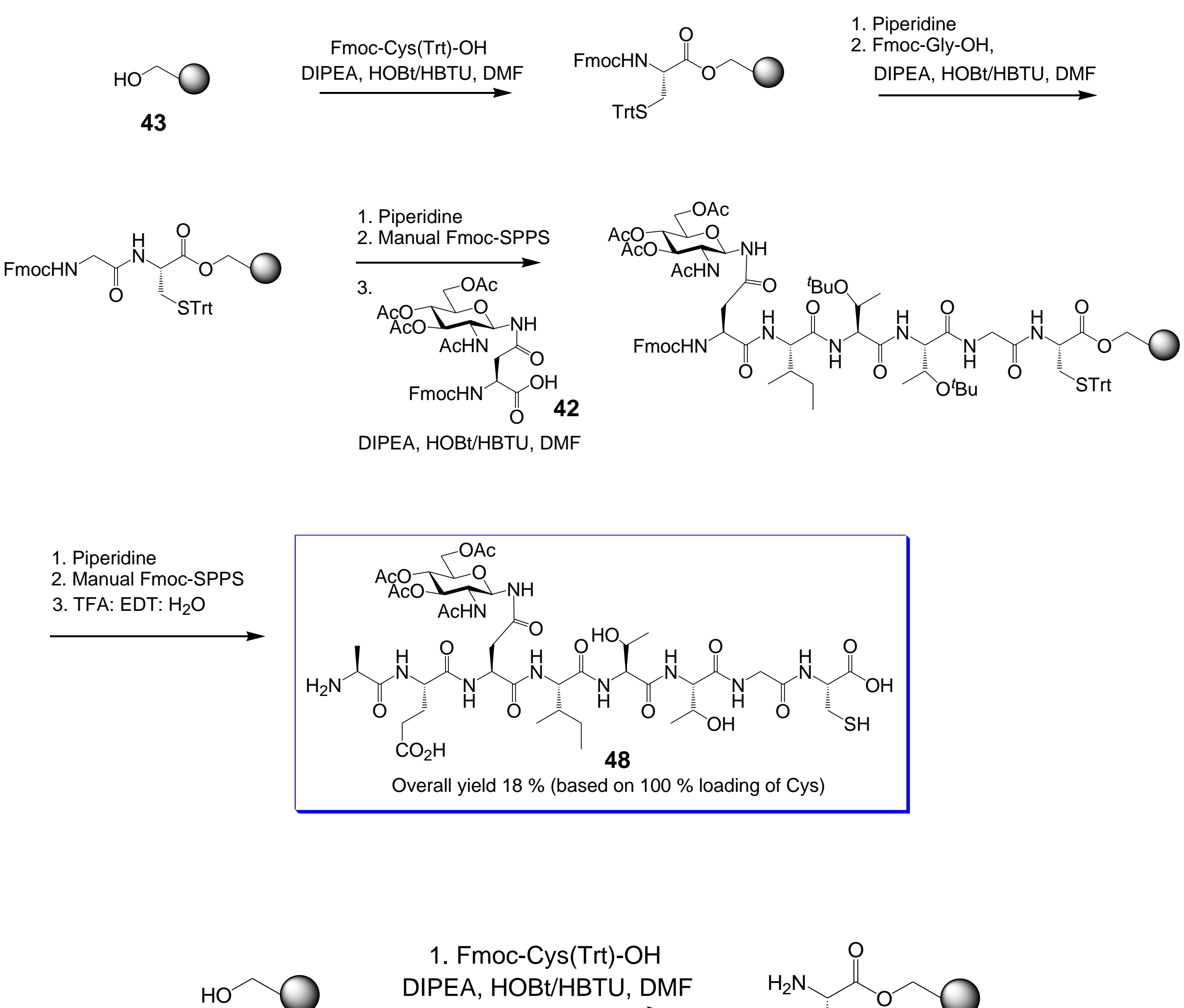
= 7.0, *J* = 7.3, Fmoc Ph-H); 7.09 (2 H, br t, *J* = 7.0, Fmoc Ph-H); 4.90 (2 H, m, H_{3a}, H_{1a}); 4.82 (1 H, asym br t, *J* = 9.7, H_{4a}); 4.25 (1 H, br s, Asn $^{\alpha}$CH); 4.05 (4 H, br m, Fmoc CH_2, H_{6a}, H_{6b}); 3.84 (2 H, br m, Fmoc CH, H_{2a}); 3.57 (1 H, br m, H_{5a}); 2.48 (2 H, m, Asn $^{\beta}CH_2$); 1.81, 1.81, 1.79 (9 H, 3 × s, 3 × CH_3CO); 1.66 (3 H, s, $\underline{H_3}$CONH); 1.22 (9 H, s, O^{t}Bu: 3 × CH_3); ^{13}C NMR [125 MHz, $CDCl_3$ (78.7 ppm) + MeOD (49.9 ppm)] δ_C (ppm) 174.0, 172.8, 172.5, 172.1, 171.7, 171.2 (C=O); 145.1, 145.0, 142.5; 129.0, 128.3, 126.3, 121.2 (Fmoc C=C); 83.6, 79.9, 74.6, 74.5, 69.7 (CH); 68.3 (CH_2); 63.3 (CH_2); 53.8, 52.4, 52.4, 48.3 (CH); 38.7 (CH_2), 28.9 (O^{t}Bu $\underline{C}H_3$); 23.6 ($\underline{C}H_3$CONH); 21.6, 21.6, 21.5 ($\underline{C}H_3$CO); IR (υ_{max}) neat: 3301, 3067, 2971, 1742, 1696, 1660, 1531, 1450, 1368, 1225 cm^{-1}; accurate mass (TOF MS ES$^+$) (*m/z*) calculated for $C_{37}H_{45}N_3O_{13}Na$: $[MNa]^+$ 762.2850, observed $[MNa]^+$ 762.2830.

42. *N*-*α*-9-Fluorenylmethoxycarbonyl-*N*-*β*-(2-*N*-acetylamido-2-deoxy-*β*-D-glucopyranosyl)-L-asparagine.

The *tert*-butyl protecting group was removed by treating **41** (251 mg, 0.34 mmol) with 95 % (v/v) TFA (4.75 mL) in water (250 μL), and the resulting solution was stirred for 3 h at room temperature, with monitored evolution of isobutylene. T.L.C. analysis indicated that the reaction was complete. The solvent was then removed under reduced pressure to afford the desired glycoamino acid **42** (232 mg, 100 %) as an off-white solid. Crude NMR analysis corresponded to the pure product and hence the glycoamino acid was used without further purification. R_f 0.1 (2: 8 methanol: ethyl acetate); m.p. decomposed at 166-169 °C; ^{1}H NMR [300 MHz, 22 % v/v $CDCl_3$ (7.258 ppm) + 78 % v/v MeOD (4.097 ppm)] δ_H (ppm) 7.59 (2 H, d, *J* = 7.5, Fmoc Ph-H); 7.45 (2 H, d, *J* = 7.3, Fmoc Ph-H); 7.23 (2 H, t, *J* = 7.4, Fmoc Ph-H); 7.14 (2 H, t, *J* = 7.3, *J* = 7.4, Fmoc Ph-H); 4.91 (3 H, m, H_{1a}, H_{3a}, H_{4a}); 4.39 (1 H, br t, *J* = 4.7, Asn $^{\alpha}$CH); 4.22 (2 H, m, H_{6a}, H_{6b}); 4.06 (2 H, d, *J* = 5.3, Fmoc CH_2); 3.93 (2 H, m, Fmoc CH, H_{2a}); 3.63 (1 H, ddd, *J* = 4.1, *J* = 2.1, *J* = 9.8, H_{5a}); 2.64 (2 H, m, Asn $^{\beta}CH_2$); 1.88 (3 H,

s, CH_3CO); 1.87 (3 H, s, CH_3CO); 1.85 (3 H, s, CH_3CO); 1.72 (3 H, s, $C\underline{H}_3CONH$); ^{13}C NMR [125 MHz, $CDCl_3$ (78.7 ppm) + MeOD (49.9 ppm)] δ_C (ppm) 174.8, 174.4, 173.2, 172.9, 172.3, 171.7 (C=O); 158.6, 145.5, 145.4, 142.9, 142.9, 129.3; 128.6, 126.7, 126.6, 121.4 (Fmoc C=C); 80.1, 75.0, 75.0, 70.2 (CH); 68.6 (CH_2); 63.7 (CH_2); 54.3, 52.1, 48.7 (CH); 38.8 (Asn $^{\beta}CH_2$), 23.5 ($\underline{C}H_3CONH$); 21.6, 21.6, 21.5 ($\underline{C}H_3CO$); IR (υ_{max}) neat: 3308, 3067, 2944, 1743, 1696, 1660, 1532, 1372, 1226 cm^{-1}; LC-MS ESI^+ MS (*m/z*) calculated for $C_{33}H_{37}N_3O_{13}$: [M] 683.2, observed $[MNa]^+$ 706.5.

48. H-AEN(Ac$_3$-Glc*N*Ac)ITTGC-OH.



NovaSyn® TGA resin **43** (384 mg, 0.10 mmol) was initially swollen in dichloromethane (1.0 mL) for 15 min, the solvent was then filtered off and the resin was dried. The dry resin was then reacted with 5 equivalents of Fmoc-Cys(Trt)-OH (293 mg, 0.5 mmol),

diisopropylethylamine (300 μL), 0.45 M HOBt/HBTU (2.2 mL), and anhydrous *N*, *N*'-dimethylformamide (2.0 mL), which was left to shake for 2 h. The heterogeneous mixture was filtered off and washed once with DMF (1.0 mL) for 5 min, followed by twice with DCM (1.0 mL) for 5 min each, and the resin was then dried. The N^{α}-Fmoc group was then removed by the addition of 20 % (v/v) piperidine/DMF (1.0 mL) to the dried resin and the mixture was left to shake for 5 min. The mixture was then filtered off and washed once with DMF (1.0 mL) for 5 min, followed by twice with DCM (1.0 mL) for 5 min each, and finally the resin was dried. More 20 % (v/v) piperidine/DMF (1.0 mL) was added to the dried resin and the mixture was shaken for 15 min. The mixture was then filtered off and washed once with DMF (1.0 mL) for 5 min, followed by twice with DCM (1.0 mL) for 5 min each, and finally the resin was dried.

H_2N TrtS **44** — 1. Fmoc-Gly-OH, DIPEA, HOBt/HBTU, DMF; 2. Piperidine → H_2N STrt **45**

The resin **44** was then coupled to Fmoc-Gly-OH (74 mg, 0.25 mmol), with *N, N*-diisopropylethylamine (150 μL), 0.45 M HOBt/HBTU (1.1 mL), and anhydrous *N*, *N*-dimethylformamide (1.0 mL), and was left to shake for 2.5 h. The reaction mixture was filtered off and washed once with *N*,*N*-dimethylformamide (1.0 mL) for 5 min, followed by twice with DCM (1.0 mL) for 5 min each, and the resin was then dried. The N^{α}-Fmoc group was then removed by the addition of 20 % (v/v) piperidine/ *N*, *N*-dimethylformamide (1.0 mL) to the dried resin and the mixture was left to shake for 5 min. The mixture was then filtered off and washed once with *N*, *N*-dimethylformamide (1.0 mL) for 5 min, followed by twice with DCM (1.0 mL) for 5 min each, and then the resin was dried. More 20 % (v/v) piperidine/ *N*, *N*-dimethylformamide (1.0 mL) was added to the dried resin and the mixture was shaken for 15 min. The mixture was then filtered off and washed once with *N*, *N*-dimethylformamide (1.0 mL) for 5 min, followed by twice with DCM (1.0 mL) for 5 min each, and finally the resin was dried.

Manual Fmoc-SPPS was continued to afford the sequence: H-IT(^{t}Bu)T(^{t}Bu)GC(Trt)-resin, using 5 equivalents of Fmoc-Ile-OH and Fmoc-Thr(^{t}Bu)-OH, with double-coupling of Fmoc-Ile-OH.

The *N*-linked glycoamino acid **42** (68 mg, 0.10 mmol) was then added to the peptidyl-resin along with diisopropylethylamine (150 μL), 0.45 M HOBt/HBTU (1.1 mL), and anhydrous *N*,*N*'-dimethylformamide (1.0 mL), and the mixture was left to shake for 30 h. The reaction mixture was then filtered off and washed once with *N*, *N*-dimethylformamide (1.0 mL) for 5 min, followed by twice with DCM (1.0 mL) for 5 min each, and the resin was then dried. The N^{α}-Fmoc group was then removed by the addition of 20 % (v/v) piperidine/ *N*, *N*-dimethylformamide (1.0 mL) to the dried resin and the mixture was left to shake for 5 min. The mixture was then filtered off and washed once with DMF (1.0 mL) for 5 min, followed by twice with DCM (1.0 mL) for 5 min each, and then the resin was dried. More 20 % (v/v) piperidine/ *N*, *N*-dimethylformamide (1.0 mL) was added to the dried resin and the mixture was shaken for 15 min. The mixture was then filtered off and washed once with *N*, *N*-dimethylformamide (1.0 mL) for 5 min, followed by twice with DCM (1.0 mL) for 5 min each, and finally the resin was dried.

Peptide synthesis was then completed manually using 5 equivalents of Fmoc-Glu(O^{t}Bu)-OH and Boc-Ala-OH to afford: Boc-AE(O^{t}Bu)N(Ac$_3$-Glc*N*Ac)IT(^{t}Bu)T(^{t}Bu)GC(Trt)-resin. The protected peptidyl-resins were then treated with a cleavage cocktail comprised of TFA:

ethanedithiol: water (95: 2.5: 2.5) with total volume of 3.0 mL. The mixture was then filtered and ice-cold diethyl ether (20.0 mL) was added, which induced precipitation. This mixture was then vortexed and centrifuged (3000 rpm, 15 min, 4 °C); then the ether layer was carefully decanted. More diethyl ether (20.0 mL) was added and the mixture was again vortexed, centrifuged, and decanted to afford the crude peptide product as a white solid. The crude peptide was dissolved in water (1.0 mL) and acetonitrile (0.5 mL), centrifuged and the supernatant was then purified by (RP)HPLC (gradient: 5-50 % MeCN over 45 min). The desired peptide product fractions were pooled and lyophilised to afford the unlabelled glycopeptide **48**: H-AEN(Ac_3-Glc*N*Ac)ITTGC-OH (10.1 mg, 18 %) as a fluffy white solid. LC-MS ESI^+ MS (*m/z*) calculated for **48** 1136.18, observed $[MH]^+$ 1137.38.

49. H-AEN(Ac_3-Glc*N*Ac)ITTG-$SCH_2CH_2SO_3H$.

10 % MESNa, 10 % AcOH, TCEP, 6M Gu.HCl, pH 3.0, 60 °C, 48 h

20 %

48 → **49**

The peracetylated glycopeptide **48** (3.0 mg, 2.64 mmol) was dissolved in degassed 6 M guanidine hydrochloride (331 µL), 10 % (w/v) sodium 2-mercaptoethanesulfonate (43 mg, 0.26 mmol), 10 % (v/v) acetic acid (43 µL), 1 M neutralised solution of TCEP (9 µL, 20 mM), to give a solution with pH ~3.0. The resulting mixture was vortexed and placed in an Eppendorf thermomixer at 60 °C for 48 h. The reaction was monitored by LC-MS at t = 6, 24, and 48 h. After 48 h the pH of the reaction appeared closer to 5.0, a crystal of TCEP.HCl was added to the reaction mixture, which was then vortexed, then centrifuged and finally purified by (RP)HPLC (gradient: 5-50 % MeCN over 45 min) to afford the desired peracetylated glycopeptide thioester **49** (0.6 mg, 20 %) as a white solid and unreacted starting material (1.4 mg, 47 % based on starting material) as a white solid. LC-MS ESI^+ MS (*m/z*) calculated for **49** 1158.21, observed $[MH]^+$ 1159.36.

50. H-CSLNENIT-OH: EPO(33-40)

NovaSyn® TGA resin **43** (455 mg, 0.1 mmol; loading 0.22 $mmolg^{-1}$; 90 µm) was transferred to an automated peptide synthesiser for peptide elongation. 10 Equivalents of each of the

following Fmoc-amino acids were employed: Fmoc-Cys(Trt)-OH, Fmoc-Ser(^{t}Bu)-OH, Fmoc-Leu-OH, Fmoc-Asn(Trt)-OH, Fmoc-Glu(O^{t}Bu)-OH, Fmoc-Ile-OH, and Fmoc-Thr(^{t}Bu)-OH. The resin was then treated with TFA: EDT: H_2O (95: 2.5: 2.5 = 5.0 mL) for 5 h, and the solution was then filtered and precipitation was induced upon addition of cold diethyl ether (20.0 mL) to the filtrate, which was then vortexed and centrifuged (3000 rpm, 4 °C, 15 min). After decanting the ether layer, this process was repeated with cold diethyl ether (20.0 mL). The crude peptide was then dissolved in 20 % aqueous acetonitrile, centrifuged, and the supernatant was purified by semi-preparative (RP)HPLC (gradient: 5-50 % acetonitrile/ 45 min). The desired peptide fractions were lyophilised to afford the N-terminal cysteinyl peptide **50** (63 mg, 71 %) as a white solid. LC-MS ESI$^+$ MS (*m/z*) calculated mass for $C_{35}H_{60}N_{10}O_{15}S$: [M] 892.4, observed 892.5.

51. H-AEN(Ac$_3$-Glc*N*Ac)ITT<u>GC</u>SLNENIT-OH.

The peracetylated glycopeptide thioester **49** (0.6 mg, 0.52 µmol) and excess cysteinyl-peptide EPO(33-40) **50** (0.80 mg, 0.90 µmol) were dissolved in a solution of 6 M guanidine hydrochloride (132 µL), sodium phosphate buffer (40 µL, 200 mM) at pH 7.0, sodium 2-mercaptoethanesulfonate (20 µL, 100 mM), and *tris*(2-carboxyethyl)phosphine (8 µL, 40 mM). The reaction mixture was then vortexed and placed in an Eppendorf thermomixer and left to shake at 25 °C, 650 rpm for 48 h. The ligation reaction was monitored by removing aliquots of the reaction mixture and analysing the sample by analytical (RP)HPLC and LC-MS. After 14 h the ligation product was observed, and more TCEP (5 µL of 1 M solution) was added to the reaction mixture. After 62 h no further progess of the reaction had occurred and it was terminated. A crystal of TCEP.HCl was added to the reaction mixture and it was purified by (RP)-HPLC (gradient: 5-50 % MeCN over 45 min) to afford the desired ligated glycopeptide **51** (0.3 mg, 30 %) as a white solid. LC-MS ESI$^+$ MS (*m/z*) calculated for **51** $C_{77}H_{125}O_{35}N_{19}S$: [M] 1907.83, observed $[MH]^+$ 1909.17.

52. H-AEN(Glc*N*Ac)ITTGC-OH.

5 % (v/v) Hydrazine (aq), DTT

87 %

48 → 52

The peracetylated glycopeptide **48** (6.7 mg, 5.90 μmol) was dissolved in 5 % (v/v) aqueous hydrazine.monohydrate (50 μL) in water (850 μL), with a 10 % (w/v) excess of 1,4-dithio-DL-threitol (100 mg). The resulting mixture was vortexed to give a solution with pH 9.0-10.0. It was then placed in an Eppendorf thermomixer and allowed to shake at 25 °C, 650 rpm, for 3 h. LC-MS analysis indicated that the reaction was complete, hence the reaction mixture was centrifuged and purified by (RP)HPLC (gradient: 5-50 % MeCN over 45 min) to afford the desired deprotected glycopeptide **52** (5.2 mg, 87 %) as a white solid. LC-MS ESI$^+$ MS (*m/z*) calculated for **52** $C_{39}H_{66}O_{19}N_{10}S$: 1010.42, observed $[MH]^+$ 1011.30.

53. H-AEN(Glc*N*Ac)ITTG-SCH$_2$CH$_2$SO$_3$H.

10 % MESNa, 10 % AcOH, 0.5 % TCEP, pH 4.0, 60 °C, 48 h

64 %

52 → 53

The deacetylated glycopeptide **52** (4.5 mg, 2.64 μmol) was dissolved in water (800 μL), 10 % (w/v) sodium 2-mercaptoethanesulfonate (100 mg, 0.61 mmol), 10 % (v/v) acetic acid (100 μL), 0.5 % (w/v) TCEP.HCl (5.0 mg, 17.44 μmol), to give a solution with pH 1.0. To the mixture was added 10 M sodium hydroxide (20 μL) to raise the pH to 4.0. The resulting mixture was vortexed and placed in an Eppendorf thermomixer at 60 °C for 48 h. The reaction was monitored by LC-MS, and analytical (RP)HPLC, at regular intervals: t = 0, 24, and 48 h. After 48 h, a crystal of TCEP.HCl was added to the reaction mixture which was then vortexed, centrifuged and purified by (RP)HPLC (gradient: 5-50 % MeCN over 45 min) to afford the deacetylated glycopeptide thioester **53** (64 % by semi-preparative HPLC integration) and unreacted starting material **52** (0.8 mg, 18 % based on **52**) as a white solid.

7. References

1. E. S. Lander, L. M. Linton, B, Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K.Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Harkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J.-F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, *Nature*, **2001**, *409*, 860.
2. Z. Abdellah, A. Ahmadi, S. Ahmed, M. Aimable, R. Ainscough, J. Almeida, C. Almond, A. Ambler, K. Ambrose, K. Ambrose, R. Andrew, D. Andrews, N. Andrews, D. Andrews, E. Apweiler, H. Arbery, B. Archer, G. Ash, K. Ashcroft, J. Ashurst, R. Ashwell, D. Atkin, A. Atkinson, B. Atkinson, J. Attwood, K. Aubin, K. Auger, T. Avis, A. Babbage, S. Babbage, J. Bacon, C. Bagguley, J. Bailey, A. Baker, R. Banerjee, S. Bardill, D. Barker, G. Barker, D. Barker, K. Barlow, L. Baron, A. Barrett, R. Bartlett, D. Basham, V. Basham, A. Bateman, K. Bates, C. Baynes, L. Beard, S. Beard, D. Beare, A. Beasley, H. Beasley, O. Beasley, S. Beck, E. Bell, D. Bellerby, T. Bellerby, R. Bemrose, J. Bennett, D. Bentley, A. Bentley, M. Berks, M. Berks, G. Bethel, C. Bird, E. Birney, H. Bissell, S. Blackburne-Maze, S. Blakey, C. Bolton, J. Bonfield, R. Bonnett, R. Border, A. Bradley, N. Brady, J. Bray, S. Bray-Allen, A. Bridgeman, J. Brook, S. Brooking, A. Brown, C. Brown, J. Brown, M. Brown, M. Brown, R. Bruskiewich, *Nature*, **2004**, *431*, 931.

3. R. D. Unwin, S. J. Gaskell, C. A. Evans and A. D. Whetton, *Exp. Hematol.*, **2003**, *31*, 1147.
4. W. Arber, *Arch. Sci. (Genève)*, **1958**, *11,* 259.
5. W. Arber, *Virology*, **1960**, *11,* 273.
6. W. Arber, *Path. Microbiol.*, **1962**, *25*, 668.
7. W. Arber, *J. Mol. Biol.*, **1965**, *1 I,* 247.
8. W. Arber and D. Dussoix, *J. Mol. Biol.*, **1962**, *5*, 18.
9. D. Dussoix and W. Arher, *J. Mol. Biol.*, **1962**, 5, 37.
10. W. Arber, *Annu. Rev. Microbiol.*, **1965**, *19*, 365.
11. H. O. Smith and K. W. J. Wilcox, *J. Mol. Biol.*, **1970**, *51*, 379.
12. T. J. Kelly, Jr. and H. O. Smith, *J. Mol. Biol.*, **1970**, *51*, 393.
13. H. O. Smith and D. Nathans, *J. Mol. Biol.*, **1973**, *81*, 419.
14. P. H. Roy and H. O. Smith, *J. Mol. Biol.*, **1973**, *81*, 427.
15. P. H. Roy and H. O. Smith, *J. Mol. Biol.*, **1973**, *81*, 445.
16. D. Nathans and H. O. Smith, *Annu. Rev. Biochem.*, **1975**, *44*, 273.
17. S. N. Cohen, A. C. Y. Chang, H. W. Boyer and R. B. Helling, *Proc. Natl. Acad. Sci. USA*, **1973**, *70*, 3240.
18. I. S. Johnson, *Science*, **1983**, *219*, 632.
19. S. Lee-Huang, *Proc. Natl. Acad. Sci. USA*, **1984**, *81*, 2708.
20. M. Chalfie, Y. Tu, G. Euskirchen, W. W. Ward and D. C. Prasher, *Science*, **1994**, *263*, 802.
21. B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter, *Molecular Biology of the Cell*, Fourth Edition, Garland Science, New York, USA, **2002**.
22. J. M. Berg, J. L. Tymoczko and L. Stryer, *Biochemistry*, Fifth Edition, W. H. Freeman and Company, New York, USA, **2002**.
23. D. Mendel, V. W. Cornish and P. G. Schultz, *Annu. Rev. Biophys. Biomol. Struct.*, **1995**, *24*, 435.
24. J. Xie and P. G. Schultz, *Curr. Opin. Chem. Biol.*, **2005**, *9*, 548.
25. D. S. Y. Yeo, R. Srinivasan, M. Uttamchandani, G. Y. J. Chen, Q. Zhu and S. Q. Yao, *Chem. Commun.*, **2003**, 2870.
26. R. Smith, I. M. Brereton, R. Y. Chai and S. B. H. Kent, *Nat. Struct. Biol.*, **1996**, *3*, 946.

27. U. K. Blaschke, G. J. Cotton and T. W. Muir, *Tetrahedron*, **2000**, *56*, 9461.
28. T. Yamazaki, T. Otomo, N. Oda, Y. Kyogoku, K. Uegaki, N. Ito, Y. Ishino and H. Nakamura, *J. Am. Chem. Soc.*, **1998**, *120*, 5591.
29. R. Xu, B. Ayers, D. Cowburn and T. W. Muir, *Proc. Natl. Acad. Sci. USA*, **1999**, *96*, 388.
30. J. A. Camarero, A. Shekhtman, E. A. Campbell, M. Chlenov, T. M. Gruber, D. A. Bryant, S. A. Darst, D. Cowburn and T. W. Muir, *Proc. Natl. Acad. Sci. USA*, **2002**, *99*, 8536.
31. A. Romanelli, A. Shekhtman, D. Cowburn and T. W. Muir, *Proc. Natl. Acad. Sci. USA*, **2004**, *101*, 6397.
32. H. Yagi, T. Tsujimoto, T. Yamazaki, M. Yoshida and H. Akutsu, *J. Am. Chem. Soc.*, **2004**, *126*, 16632.
33. G. G. Kochendoerfer, S. Y. Chen, F. Mao, S. Cressman, S. Traviglia, H. Y. Shao, C. L. Hunter, D. W. Low, E. N. Cagle, M. Carnevali, V. Gueriguian, P. J. Keogh, H. Porter, S. M. Stratton, M. C. Wiedeke, J. Wilken, J. Tang, J. J. Levy, L. P. Miranda, M. M. Crnogorac, S. Kalbag, P. Botti, J. Schindler-Hovat, L. Savatski, J. W. Adamson, A. Kung, S. B. H. Kent and J. A. Bradburne, *Science*, **2003**, *299*, 884.
34. J. S. Miller, V. Y. Dudkin, Gholson J. Lyon, T. W. Muir and S. J. Danishefsky, *Angew. Chem. Int. Ed.*, **2003**, *42*, 431.
35. H. C. Hang and C. R. Bertozzi, *Acc. Chem. Res.*, **2001**, *34*, 727.
36. J.-W. Wu, M. Hu, J. Chai, J. Seoane, M. Huse, C. Li, D. J. Rigotti, S. Kyin, T. W. Muir, R. Fairman, J. Massagué and Y. Shi, *Mol. Cell*, **2001**, *8*, 1277.
37. F. I. Valiyaveetil, R. MacKinnon and T. W. Muir, *J. Am. Chem. Soc.*, **2002**, *124*, 9113.
38. V. Y. Torbeev and S. B. H. Kent, *Angew. Chem. Int. Ed.*, **2007**, *46*, 1667.
39. H. Iwai, A. Lingel and A. Pluckthun, *J. Biol. Chem.*, **2001**, *276*, 16548.
40. S. A. Tatulian, S. Qin, A. H. Pande and X. M. He, *J. Mol. Biol.*, **2005**, *351*, 939.
41. D. Clayton, G. Shapovalov, J. A. Maurer, D. A. Dougherty, H. A. Lester and G. G. Kochendoerfer, *Proc. Natl. Acad. Sci. USA*, **2004**, *101*, 4764.
42. E. Fischer and E. Fourneau, *Ber. Dtsh. Chem. Ges.*, **1901**, *34*, 2868.
43. F. Hofmeister, *Ergeb. Physiol.*, **1902**, *1*, 759.
44. E. Fischer, *Chemiker Zeitung*, **1902**, *26*, 939.

45. E. Fischer, *Ber. Dtsh. Chem. Ges.*, **1904**, *37*, 3062.
46. E. Fischer, *Ber. Dtsh. Chem. Ges.*, **1907**, *40*, 1754.
47. V. Du Vigneaud, C. Ressler, J. M. Swan, C. W. Roberts, P. G. Katsoyannis and S. Gordon, *J. Am. Chem. Soc.*, **1953**, *75*, 4879.
48. G. W. Kenner, *Proc. R. Soc. London, Ser. A*, **1977**, *353*, 441.
49. P. Sieber, K. Eisler, B. Kamber, B. Riniker, W. Rittel, F. Marki and M. Degasparo, *Hoppe-Seyler's Z. Physiol. Chem.*, **1978**, *359*, 113.
50. H. Yajima and N. Fujii, *Biopolymers*, **1981**, *20*, 1859.
51. D. S. Kemp, *Biopolymers*, **1981**, *20*, 1793.
52. S. B. H. Kent, *Annu. Rev. Biochem.*, **1988**, *57*, 957.
53. S. B. H. Kent, *Chem. Soc. Rev.*, **2009**, *38*, 338.
54. M. Bergmann and L. Zervas, *Ber. Dtsch. Chem. Ges.*, **1932**, *65B*, 1192.
55. G. Barany and R. B. Merrifield, *J. Am. Chem. Soc.*, **1977**, *99*, 7363.
56. G. Barany and F. Albericio, *J. Am. Chem. Soc.*, **1985**, *107*, 4936.
57. A. Isidro-Llobet, M. Álvarez and F. Albericio, *Chem. Rev.*, **2009**, *109*, 2455.
58. K. Barlos, P. Mamos, D. Papaioannou, S. Patrianakou, C. Sanida and W. Schaefer, *Liebigs Ann.Chem.*, **1987**, *12*, 1025.
59. M. Bodanszky, M. A. Bednarek and A. Bodanszky, *Int. J. Pept. Prot. Res.*, **1982**, *20*, 387.
60. J. Alsina, E. Giralt and F. Albericio, *Tetrahedron Lett.*, **1996**, *37*, 4195.
61. B. G. de la Torre, M. A. Marcos, R. Eritja and F. Albericio, *Lett. Pept. Sci.*, **2001**, *8*, 331.
62. A. Floersheimer and B. Riniker, *Peptides 1990: Proceedings of the 21st European Symposium*; E. Giralt and D. Andreau, eds. ESCOM Sci. Publ.: Leiden, The Netherlands, **1991**, p 131.
63. K. Barlos, D. Papaioannou, S. Patrianakou and T. Tsegenidis, *Liebigs Ann. Chem.*, **1986**, *11*, 1950.
64. S. S. Wang, C. C. Yang, I. D. Kulesha, M. Sonenberg and R. B. Merrifield, *Int. J. Pept. Prot. Res.*, **1974**, *6*, 103.
65. S. Mojsov and R. B. Merrifield, *Biochemistry*, **1981**, *20*, 2950.
66. F. Albericio, *Biopolymers*, **2000**, *55*, 123.

67. R. I. Carey, L. W. Bordas, R. A. Slaughter, B. C. Meadows, J. L. Wadsworth, H. Huang, J. J. Smith and E. Furusjo, *J. Pept. Res.*, **1997**, *49*, 570.
68. T. J. Attard, E. C. Reynolds and J. W. Perich, *Org. Biomol. Chem.*, **2007**, *5*, 664.
69. S. Zaramella, E. Yeheskiely and R. Stroemberg, *J. Am. Chem. Soc.*, **2004**, *126*, 14029.
70. L. Zervas, D. Borovas and E. Gazis, *J. Am. Chem. Soc.*, **1963**, *85*, 3660.
71. V. A. Najjar and R. B. Merrifield, *Biochemistry*, **1966**, *5*, 3765.
72. A. Tun-Kyi, *Helv. Chim. Acta*, **1978**, *61*, 1086.
73. J. Meienhofer, *Nature*, **1965**, *205*, 73.
74. D. Ben-Ishai and A. Berger, *J. Org. Chem.*, **1952**, *17*, 1564.
75. A. R. Mitchell and R. B. Merrifield, *J. Org. Chem.*, **1976**, *41*, 2015.
76. Y. Kiso, K. Ukawa and T. Akita, *J. Chem. Soc. Chem. Commun.*, **1980**, *3*, 101.
77. S. Sakakibara, Y. Shimonishi, Y. Kishida, M. Okada and H. Sugihara, *Bull. Chem. Soc. Jpn.*, **1967**, *40*, 2164.
78. A. M. Felix, *J. Org. Chem.*, **1974**, *39*, 1427.
79. C. M. Stevens and R. Watanabe, *J. Am. Chem. Soc.*, **1950**, *72*, 725.
80. J. Tsuji, *Tetrahedron*, **1986**, *42*, 4361.
81. B. M. Trost and D. L. van Vranken, *Chem. Rev.*, **1996**, *96*, 395.
82. F. Guibé, *Tetrahedron*, **1997**, *53*, 13509.
83. F. Guibé, *Tetrahedron*, **1998**, *54*, 2967.
84. D. Fernandez-Forner, G. Casals, E. Navarro, H. Ryder and F. Albericio, *Tetrahedron Lett.*, **2001**, *42*, 4471.
85. M. Dessolin, M.-G. Guillerez, N. Thieriet, F. Guibé and A. Loffet, *Tetrahedron Lett.*, **1995**, *36*, 5741.
86. P. Gómez-Martínez, M. Dessolin, F. Guibé and F. Albericio, *J. Chem. Soc. Perk. Trans. 1*, **1999**, 22871.
87. N. Thieriet, P. Gómez-Martínez and F. Guibé, *Tetrahedron Lett.*, **1999**, *40*, 2505.
88. N. Thieriet, J. Alsina, E. Giralt, F. Guibé and F. Albericio, *Tetrahedron Lett.*, **1997**, *38*, 7275.
89. C. Gracia, A. Isidro-Llobet, L. J. Cruz, G. A. Acosta, M. Álvarez, C. Cuevas, E. Giralt and F. Albericio, *J. Org. Chem.*, **2006**, *71*, 7196.
90. F. H. Carpenter and D. T. Gish, *J. Am. Chem. Soc.*, **1952**, *74*, 3818.

91. A. Isidro-Llobet, J. Guasch-Camell, M. Álvarez and F. Albericio, *Eur. J. Org. Chem.*, **2005**, 3031.
92. W. Liao, C. F. Piskorz, R. D. Locke and K. L. Matta, *Bioorg. Med. Chem. Lett.*, **2000**, *10*, 793.
93. J. Tulla-Puche, N. Bayó-Puxan, J. A. Moreno, A. M. Francesch, C. Cuevas, M. Álvarez and F. Albericio, *J. Am. Chem. Soc.*, **2007**, *129*, 5322.
94. H.-J. Kohlbau, R. Thürmer and W. Voelter. *Protection of the α-Carboxy Group.* In *Synthesis of Peptides and Peptidomimetics (Houben-Weyl E22a: Methods of Organic Chemistry)*; M. Goodman, A. M. Felix, L. Moroder, C. Toniolo, eds., Georg Thieme Verlag: Stuttgart and New York, **2002**, p 193-237.
95. H. Waldmann and H. Kunz, *Liebigs Ann. Chem.*, **1983**, *10*, 1712.
96. R. Roeske, *J. Org. Chem.*, **1963**, *28*, 1251.
97. G. W. Anderson and F. M. Callahan, *J. Am. Chem. Soc.*, **1960**, *82*, 3359.
98. J. A. Maclaren, *Aust. J. Chem.*, **1971**, *24*, 1695.
99. S.-S. Wang, B. F. Gisin, D. P. Winter, R. Makofske, I. D. Kulesha, C. Tzougraki, J. Meienhofer, *J. Org. Chem.*, **1977**, *42*, 1286.
100. J. L. Matthews, K. Gademann, B. Juan and D. Seebach, *J. Chem. Soc. Perk. Trans. 1*, **1998**, *20*, 3331.
101. P. Llyod-Williams, F. Albericio and E. Giralt, *Chemical Approaches to the Synthesis of Peptides and Proteins*, CRC Press, Boca Raton, FL, **1997**.
102. D. Gatos, P. Athanassopoulos, C. Tzavara and K. Barlos, *Peptides 1998: Proceedings of the 25th European Peptide Symposium*, S. Bajusz, F. Hudecz, eds. Akademiai Kiado, Budapest, Hungary, **1999**, p 146.
103. J. S. McMurray, *Tetrahedron Lett.*, **1991**, *32*, 7679.
104. K. J. Jensen, J. Alsina, M. F. Songster, J. Vágner, F. Albericio and G. Barany, *J. Am. Chem. Soc.*, **1998**, *120*, 5441.
105. A. Guzman-Martinez, R. Lamer, M. S. VanNieuwenhze, *J. Am. Chem. Soc.*, **2007**, *129*, 6017.
106. J. Jones, *Amino Acid and Peptide Synthesis*, Second edition, Oxford Science Publications, Oxford University Press, Great Britain, **2002**.
107. D. Kahne and W. C. Still, *J. Am. Chem. Soc.*, **1988**, *110*, 7529.
108. R. B. Martin, *Biopolymers*, **1988**, *45*, 351.

109. S. Doonan, *Peptides and Proteins.* Royal Society of Chemistry, Great Britain, **2002**.
110. S. Sakakibara, *Biopolymers*, **1995**, *37*, 17.
111. Y. Nishiuchi, T. Inui, H. Nishio, J. Bódi, T. Kimura, F. I. Tsuji and S. Sakakibara, *Proc. Natl. Acad. Sci. USA*, **1998**, *95*, 13549.
112. S. Sakakibara, *Biopolymers*, **1999**, *51*, 279.
113. R. B. Merrifield, *J. Am. Chem. Soc.*, **1963**, *85*, 2149.
114. R. B. Merrifield, *Methods Enzymol.*, **1997**, *289*, 3.
115. W. C. Chan and P. D. White, eds., *Fmoc solid phase peptide synthesis. A practical approach*. Oxford University Press, Oxford, Great Britain, **2000**.
116. D. A. Wellings and E. Atherton, *Methods Enzymol.*, **1997**, *289*, 44.
117. B. L. Nilsson, M. B. Scoellner and R. T. Raines, *Annu. Rev. Biophys. Biomol. Struct.*, **2005**, *34*, 91.
118. L. A. Carpino and G. Y. Han, *J. Am. Chem. Soc.*, **1970**, *92*, 5748.
119. L. A. Carpino and G. Y. Han, *J. Org. Chem.*, **1972**, *37*, 3404.
120. E. Atherton, H. Fox, D. Harkiss, C. J. Logan, R. C. Sheppard and B. J. Williams, *J. Chem. Soc., Chem. Commun.*, **1978**, *13*, 537.
121. J. D. Wade, J. Bedford, R. C. Sheppard and G. W. Tregear, *Pept. Res.*, **1991**, *4*, 194.
122. M. Meldal, T. Bielfeldt, S. Peters, K. J. Jensen, H. Paulsen and K. Bock, *Int. J. Pept. Prot. Res.*, **1994**, *43*, 529.
123. B. Liebe and H. Kunz, *Angew. Chem. Int. Ed.*, **1997**, *36*, 618.
124. X. Li, T. Kawakami and S. Aimoto, *Tetrahedron Lett.*, **1998**, *39*, 8669.
125. J. Martinez and M. Bodanszky, *Int. J. Pept. Prot. Res.*, **1978**, *12*, 277.
126. L. A. Carpino, *J. Am. Chem. Soc.*, **1957**, *79*, 4427.
127. G. W. Anderson and N. F. Alberston, *J. Am. Chem. Soc.*, **1957**, *79*, 6180.
128. R. B. Merrifield, *Adv. Enzymol.*, **1969**, *32*, 221.
129. E. Kaiser, F. Picart, T. Kubiak, J. P. Tam and R. B. Merrifield, *J. Org. Chem.*, **1993**, *58*, 5167.
130. J. M. Stewart and J. D. Young, *Solid Phase Peptide Synthesis*, Second edition, Pierce Chemical Company, Rockford, IL, **1984**.
131. F. Albericio, *Curr. Opin. Chem. Biol.*, **2004**, *8*, 211.
132. J. Humphrey and R. Chamberlin, *Chem. Rev.*, **1997**, *97*, 2243.
133. P. D. Bailey, *An Introduction to Peptide Chemistry*, Wiley, Germany, **1990**.

134. R. B. Merrifield, *Anal. Chem.*, **1966**, *38*, 1905.
135. B. Gutte and R. B. Merrifield, *J. Am. Chem. Soc.*, **1969**, *91*, 501.
136. B. Gutte and R. B. Merrifield, *J. Biol. Chem.*, **1971**, *246*, 1922.
137. T. Durek, V. Y. Torbeev and S. B. H. Kent, *Proc. Natl. Acad. Sci. USA*, **2007**, *104*, 4846.
138. D. S. Kemp, N. G. Galakatos, B. Bowen and K. Tan, *J. Org. Chem.*, **1986**, *51*, 1829.
139. M. Schnölzer and S. B. H. Kent, *Science*, **1992**, *256*, 221.
140. R. C. D. Milton, S. C. F. Milton and S. B. H. Kent, *Science*, **1992**, *256*, 1445.
141. M. Baca and S. B. H. Kent, *Proc. Natl. Acad. Sci. USA*, **1993**, *90*, 11638.
142. V. T. Wieland, E. Bokelmann, L. Bauer, H. U. Lang and H. Lau, *Liebigs Ann. Chem.*, **1953**, *583*, 129.
143. V. T. Wieland, R. Lamber, H. L. Lang and M. Günther Schramm, *Liebigs Ann. Chem.*, **1956**, *591*, 181.
144. P. E. Dawson, T. W. Muir, I. Clark-Lewis and S. B. H. Kent, *Science*, **1994**, *266*, 776.
145. P. E. Dawson and S. B. H. Kent, *Annu. Rev. Biochem.*, **2000**, *69*, 923.
146. P. E. Dawson, M. J. Churchill, M. R. Ghadiri and S. B. H. Kent, *J. Am. Chem. Soc.*, **1997**, *119*, 4325.
147. M. Schnölzer, P. Alewood, A. Jones, D. Alewood and S. B. H. Kent, *Int. J. Pept. Protein. Res.*, **1992**, *40*, 180.
148. L. Zhang and J. P. Tam, *J. Am. Chem. Soc.*, **1997**, *119*, 2363.
149. M. Villain, J. Vizzavona and K. Rose, *Chem Biol.*, **2001**, *8*, 673.
150. D. Bang and S. B. H. Kent, *Angew. Chem. Int. Ed.*, **2004**, *43*, 2534.
151. H. Paulus, *Chem. Soc. Rev.*, **1998**, *27*, 375.
152. C. J. Noren, J. M. Wang and F. B. Perler, *Angew. Chem. Int. Ed.*, **2000**, *39*, 450.
153. J.-P. Pellois and T. W. Muir, *Curr. Opin. Chem. Biol.*, **2006**, *10*, 487.
154. L. Saleh and F. B. Perler, *Chem. Rec.*, **2006**, *6*, 183.
155. www.neb.com
156. T. W. Muir, D. Sondhi and P. A. Cole, *Proc. Natl. Acad. Sci. USA,* **1998**, *95*, 6705.
157. T. C. Evans, Jr., J. Benner and M. Q. Xu, *Protein Sci.*, **1998**, *7*, 2256.
158. V. Muralidharan and T. W. Muir, *Nat. Methods,* **2006**, *3*, 429.
159. I. E. Gentle, D. P. De Souza and M. Baca, *Bioconjug. Chem.*, **2004**, *15*, 658.
160. D. Macmillan and L. Arham, *J. Am. Chem. Soc.*, **2004**, *126*, 9530.

161. J. P. Richardson and D. Macmillan, *Org. Biomol. Chem.*, **2008**, *6*, 3977.
162. H. Iwai and A. Pluckthun, *FEBS Lett.*, **1999**, *459*, 166.
163. J. A. Camarero, D. Fushman, D. Cowburn and T. W. Muir, *Bioorg. Med. Chem.*, **2001**, *9*, 2479.
164. D. A. Erlanson, M, Chytil and G. L. Verdine, *Chem. Biol.*, **1996**, *3*, 981.
165. T. J. Tolbert and C.-H. Wong, *Angew. Chem. Int. Ed.*, **2002**, *41*, 2171.
166. G. K. Busch, E. W. Tate, P. R. J. Gaffney, E. Rosivatz, R. Woscholski and R. J. Leatherbarrow, *Chem. Commun.*, **2008**, 3369.
167. C. Piontek, D. V. Silva, C. Heinlein, C. Pöhner, S. Mezzato, P. Ring, A. Martin, F. X. Schmid and C. Unverzagt, *Angew. Chem. Int. Ed.*, **2009**, *48*, 1941.
168. H. D. Mootz, *Chembiochem*, **2009**, *10*, 2579.
169. C. Ludwig, M. Pfeiff, U. Linne and H. D. Mootz, *Angew. Chem. Int. Ed.*, **2006**, *45*, 5218.
170. J. X. Shi and T. W. Muir, *J. Am. Chem. Soc.*, **2005**, *127*, 6198.
171. K. V. Mills, B. M. Lew, S.-q. Jiang and H. Paulus, *Proc. Natl. Acad. Sci. USA*, **1998**, *95*, 3543.
172. C. P. Scott, E. Abel-Santos, M. Wall, D. C. Wahnon and S. J. Benkovic, *Proc. Natl. Acad. Sci. USA*, **1999**, *96*, 13638.
173. F. B. Perler and E. Adam, *Curr. Opin. Biotechnol.*, **2000**, *11*, 377.
174. T. Ozawa, A. Kaihara, M. Sato, K. Tachihara and Y. Umezawa, *Anal. Chem.*, **2001**, *73*, 2516.
175. J.-Y. Yang and W. Y. Yang, *J. Am. Chem. Soc*., **2009**, *131*, 11644.
176. J. H. Appleby, K. Zhou, G. Volkmann and X.-Q. Liu, *J. Biol. Chem.*, **2009**, *284*, 6194.
177. S. W. Lockless and T. W. Muir, *Proc. Natl. Acad. Sci. USA*, **2009**, *106*, 10999.
178. Q. Xu, D. Buckley, C. Guan and C.-H. Guo, *Cell*, **1999**, *98*, 651.
179. T. Klabunde, S. Sharma, A. Telenti, W. R. Jacobs Jr. and J. C. Sacchettini, *Nat. Struct. Biol.*, **1998**, *5*, 31.
180. F. Delaglio, D. A. Torchia and A. Bax, *J. Biomol. NMR*, **1991**, *1*, 439.
181. N. Juranic, M. Moncrieffe, V. A. Likic, F. G. Prendergast and S. Macura, *J. Am. Chem. Soc.*, **2002**, *124*, 14221.
182. Z. Du, P. T. Shemella, Y. Liu, S. A. McCallum, B. Pereira, S. K. Nayak, G. Belfort, M. Belfort and C. Wang, *J. Am. Chem. Soc.*, **2009**, *131*, 11581.

183. D. Bang, B. L. Pentelute, Z. P. Gates and S. B. H. Kent, *Org. Lett.*, **2006**, *8*, 1049.
184. E. C. B. Johnson and S. B. H. Kent, *J. Am. Chem. Soc.*, **2006**, *128*, 6640.
185. D. Bang, B. L. Pentulute and S. B. H. Kent, *Angew. Chem. Int. Ed.*, **2006**, *45*, 3985.
186. A. Brik, Y. Y. Yang, S. Ficht and C.-H. Wong, *J. Am. Chem. Soc.*, **2006**, *128*, 5626.
187. D. Macmillan, *Angew. Chem. Int. Ed.*, **2006**, *45*, 7668.
188. C. S. Bennett, S. M. Dean, R. J. Payne, S. Ficht, A. Brik and C.-H. Wong, *J. Am. Chem. Soc.*, **2008**, *130*, 11945.
189. J. D. Warren, J. S. Miller, S. J. Keding and S. J. Danishefsky, *J. Am. Chem. Soc.*, **2004**, *126*, 6576.
190. B. Wu, J. Chen, J. D. Warren, G. Chen, Z. Hua and S. J. Danishefsky, *Angew. Chem. Int. Ed.*, **2006**, *45*, 4116.
191. P. Botti, M. Villain, S. Manganiello and H. Gaertner, *Org. Lett.*, **2004**, *6*, 4861.
192. K. Rose, *J. Am. Chem. Soc.*, **1994**, *116*, 30.
193. L. E. Canne, A. R. Ferredamare, S. K. Burley and S. B. H. Kent, *J. Am. Chem. Soc.*, **1995**, *117*, 2998.
194. E. Saxon and C. R. Bertozzi, *Science*, **2000**, *287*, 2007.
195. B. L. Nilsson, L. L. Kiessling and R. T. Raines, *Org. Lett.*, **2000**, *2*, 1939.
196. B. L. Nilsson, L. L. Kiessling and R. T. Raines, *Org. Lett.*, **2001**, *3*, 9.
197. E. Haginoya, H. Hojo, Y. Nakahara, Y. Nakahara, K. Nabeshima, B. P. Toole and Y. Watanabe*, Biosci. Biotechnol. Biochem*., **2006**, *70*, 1338.
198. N. Yamamoto, Y. Tanabe, R. Okamoto, P. E. Dawson and Y. Kajihara*, J. Am. Chem. Soc*., **2007**, *130*, 501.
199. 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
200. R. Ingenito, E. Bianchi, D. Fattori and A. Pessi, *J. Am. Chem. Soc.*, **1999**, *121*, 11369.
201. F. Mende and O. Seitz, *Angew. Chem. Int. Ed.*, **2007**, *46*, 4577.
202. S. Mezzato, M. Schaffrath and C. Unverzagt, *Angew. Chem. Int. Ed.*, **2005**, *44*, 1650.
203. H. Hojo, Y. Matsumoto, Y. Nakahara, E. Ito, Y. Suzuki, M. Suzuki, A. Suzuki and Y. Nakahara, *J. Am. Chem. Soc*., **2005**, *127*, 13720.
204. A. B. Clippingdale, C. J. Barrow and J. D. Wade, *J. Pept. Sci*., **2000**, *6*, 225.
205. R. Quaderer and D. Hilvert, *Org. Lett*., **2001**, *3*, 3181.

206. R. von Eggelkraut-Gottanka, A. Klose, A. G. Beck-Sickinger and M. Beyermann, *Tetrahedron Lett*., **2003**, *44*, 3551.
207. C. P. R. Hackenberger, C. T. Friel, S. E. Radford and B. Imperiali, *J. Am. Chem. Soc*., **2005**, *127*, 12882.
208. D. Swinnen and D. Hilvert, *Org. Lett*., **2000**, *2*, 2439.
209. T. Murase, T. Tsuji and Y. Kajihara, *Carbohydrate Res*., **2009**, *344*, 762.
210. J. B. Blanco-Canosa and P. E. Dawson, *Angew. Chem. Int. Ed.*, **2008**, *47*, 6851.
211. A. P. Tofteng, K. S. Kasper, W. C.-F. Kilian , H.-J. Thomas and J. J. Knud, *Angew. Chem. Int. Ed*., **2009**, *48*, 7411.
212. J. Brask, F. Albericio and K. J. Jensen, *Org. Lett*., **2003**, *5*, 2951.
213. U. Boas, J. Brask and K. J. Jensen*, Chem. Rev*., **2009**, *109*, 2092.
214. S. Ficht, Richard J. Payne, Richard T. Guy and C.-H. Wong, *Chem. Eur. J*., **2008**, *14*, 3620.
215. N. Ollivier, J.-B. Behr, Q. El-Mahdi, A. Blanpain and O. Melnyk, *Org. Lett*., **2005**, *7*, 2647.
216. Y. He, J. P. Wilkins and L. L. Kiessling, *Org. Lett.*, **2006**, *8*, 2483.
217. K. Pachamuthu and R. R. Schmidt, *Synlett.*, **2003**, 659.
218. G. Roelfes and D. Hilvert, *Angew. Chem.*, **2003**, *115*, 2377.
219. B. L. Pentelute and S. B. H. Kent*, Org. Lett*., **2007**, *9*, 687.
220. L. Z. Yan and P. E. Dawson, *J. Am. Chem. Soc.*, **2001**, *123*, 526.
221. D. Crich and A. Banerjee, *J Am. Chem. Soc.*, **2007**, *129*, 10064.
222. R. Okamoto and Y. Kajihara, *Angew. Chem. Int. Ed.*, **2008**, *47*, 5402.
223. C. Haase, H. Rohde and O. Seitz, *Angew. Chem. Int. Ed.*, **2008**, *47*, 6807.
224. J. Chen, Q. Wan, Y. Yuan, J. Zhu and S. J. Danishefsky, *Angew. Chem. Int. Ed.*, **2008**, *47*, 8521.
225. P. Botti, M. R. Carrasco and S. B. H. Kent, *Tetrahedron Lett.*, **2001**, *42*, 1831.
226. L. E. Canne, S. J. Bark and S. B. H. Kent, *J. Am. Chem. Soc.*, **1996**, *118*, 5891.
227. C. Marinzi, J. Offer, R. Longhi and P. E. Dawson, *Bioorg. Med. Chem*., **2004**, *12*, 2749.
228. J. Offer, C. N. C. Boddy and P. E. Dawson*, J. Am. Chem. Soc*., **2002**, *124*, 4642.
229. J. Offer and P. E. Dawson, *Org. Lett.*, **2000**, *2*, 23.
230. D. Macmillan and D. W. Anderson, *Org. Lett*., **2004**, *6*, 4659.

231. A. Dirksen and P. E. Dawson, *Curr. Opin. Chem. Biol.*, **2008**, *12*, 760.
232. D. W. Low, M. G. Hill, M. R. Carrasco, S. B. H. Kent and P. Botti, *Proc. Natl. Acad. Sci. USA*, **2001**, *98*, 6554.
233. T. Kawakami and S. Aimoto, *Tetrahedron Lett.*, **2003**, *44*, 6059.
234. J. Vizzavona, F. Dick and T. Vorherr, *Bioorg. Med. Chem. Lett.*, **2002**, *12*, 1963.
235. T. Kawakami, M. Sumida, K. Nakamura, T. Vorherr and S. Aimoto, *Tetrahedron Lett.*, **2005**, *46*, 8805.
236. K. Nakamura, M. Sumida, T. Kawakami, T. Vorherr and S. Aimoto, *Bull. Chem. Soc. Jpn.*, **2006**, *79*, 1773.
237. K. Nakamura, H. Mori, T. Kawakami, H. Hojo, Y. Nakahara and S. Aimoto, *Int. J. Pept. Res. Ther.*, **2007**, *13*, 191.
238. K. Nakamura, T. Kanao, T. Uesugi, T. Hara, T. Sato, T. Kawakami and S. Aimoto, *J. Pept. Sci.*, **2009**, *15*, 731.
239. F. Nagaike, Y. Onuma, C. Kanazawa, H. Hojo, A. Ueki, Y. Nakahara and Y. Nakahara, *Org. Lett.*, **2006**, *8*, 4465.
240. H. Hojo, Y. Onuma, Y. Akimoto, Y. Nakahara and Y. Nakahara, *Tetrahedron Lett.*, **2007**, *48*, 25.
241. H. Hojo, Y. Murasawa, H. Katayama, T. Ohira, Y. Nakahara and Y. Nakahara, *Org. Biomol. Chem.*, **2008**, *6*, 1808.
242. H. Hojo and S. Aimoto, *Bull. Chem. Soc. Jpn.*, **1991**, *64*, 111.
243. C. Ozawa, H. Katayama, H. Hojo, Y. Nakahara and Y. Nakahara, *Org. Lett.*, **2008**, *10*, 3531.
244. Y. Ohta, S. Itoh, A. Shigenaga, S. Shintaku, N. Fujii and A. Otaka, *Org. Lett.*, **2006**, *8*, 467.
245. S. Tsuda, A. Shigenaga, K. Bando and A. Otaka, *Org. Lett.*, **2009**, *11*, 823.
246. G. Zanotti, F. Pinnen and G. Lucente, *Tetrahedron Lett.*, **1985**, *26*, 5481.
247. T. Kawakami and S. Aimoto, *Chem. Lett.*, **2007**, *36*, 76.
248. T. Kawakami and S. Aimoto, *Tetrahedron Lett.*, **2007**, *48*, 1903.
249. T. Kawakami and S. Aimoto, *Tetrahedron*, **2009**, *65*, 3871.
250. T. Kawakami, S. Shimizu and S. Aimoto, *Bull. Chem. Soc. Jpn.*, **2010**, *83*, 570-574.
251. T. Kawakami, A. Ohta, M. Ohuchi, H. Ashigai, H. Murakami and H. Suga, *Nat. Chem. Biol.*, **2009**, *5*, 888.

252. H. Hojo, E. Haginoya, Y. Matsumoto, Y. Nakahara, K. Nabeshima, B. P. Toole and Y. Watanabe, *Tetrahedron Lett.*, **2003**, *44*, 2961.
253. J. A. Camarero, B. J. Hackel, J. J. D. Yoreo and A. R. Mitchell, *J. Org. Chem.*, **2004**, *69*, 4145.
254. O. S. Franziska Mende, *Angew. Chem. Int. Ed.*, **2007**, *46*, 4577.
255. S. Manabe, T. Sugioka and Y. Ito, *Tetrahedron Lett.*, **2007**, *48*, 849.
256. T. J. Hogenauer, Q. Wang, A. K. Sanki, A. J. Gammon, C. H. L. Chu, C. M. Kaneshiro, Y. Kajihara and K. Michael, *Org. Biomol. Chem.*, **2007**, *5*, 759.
257. C. Haase and O. Seitz, *Angew. Chem. Int. Ed.*, **2008**, *47*, 1553.
258. R. K. McGinty, J. Kim, C. Chatterjee, R. G. Roeder and T. W. Muir, *Nature*, **2008**, *453*, 812.
259. T. B. Krantz, *Blood*, **1991**, *77*, 419.
260. F. K. Lin, S. Suggs, C. H. Lin, J. K. Browne, R. Smalling, J. C. Egrie, K. Chen, G. M. Fox, F. Martin, Z. Stabinsky, S. M. Badrawi, P. H. Lai and E. Goldwasser, *Proc. Natl. Acad. Sci. USA*, **1985**, *82*, 7580.
261. R. S. Syed, S. W. Reid, C. Li, J. C. Cheetham, K. H. Aoki, B. Liu, H. Zhan, T. D. Osslund, A. J. Chirino, J. Zhang, J. Finer-Moore, S. Elliott, K. Sitney, B. A. Katz, D. J. Matthews, J. J. Wendoloski, J. Egrie and R. M. Stroud, *Nature*, **1998**, *395*, 511.
262. S. Gillis and D. E. Williams, *Curr. Opin. Immunol.*, **1998**, *10*, 501.
263. J. Kang, J. P. Richardson and D. Macmillan, *Chem. Commun.*, **2009**, 407.
264. T. M. Hackeng, J. H. Griffin and P. E. Dawson, *Proc. Natl. Acad. Sci. USA*, **1999**, *96*, 10068.
265. C. Haase and O. Seitz, *Eur. J. Org. Chem.*, **2009**, 2096.
266. C. Ludwig, D. Schwarzer and H. D. Mootz, *J. Biol. Chem.*, **2008**, *283*, 25264.
267. J. Kang, N. L. Reynolds, C. Tyrrell, J. R. Dorin and D. Macmillan, *Org. Biomol. Chem.*, **2009**, *7*, 4918.
268. http://chemweb.unp.ac.za/chemistry/Physical_Data/pKa_compilation.pdf
269. E. T. Batchelar, R. B. Hamed, C. Ducho, T. D. W. Claridge, M. J. Edelmann, B. Kessler and C. J. Schofield, *Angew. Chem. Int. Ed.*, **2008**, *47*, 9322.
270. R. I. Lehrer, *Nature Rev. Microbiol.*, **2004**, *2*, 727.
271. D. M. Hoover, Z. Wu, K. Tucker, W. Lu and J. Lubkowski, *Antimicrob. Agents Chemother.*, **2003**, *47*, 2804.

272. Z. Wu, D. M. Hoover, D. Yang, C. Boulegue, F. Santamaria, J. J. Oppenheim, J. Lubkowski and W. Lu, *Proc. Natl. Acad. Sci. USA*, **2003**, *100*, 8880.
273. K. Taylor, D. J. Clarke, B. McCullough, W. Chin, E. Seo, D. Yang, J. Oppenheim, D. Uhrin, J. R. W. Govan, D. J. Campopiano, D. MacMillan, P. Barran and J. R. Dorin, *J. Biol. Chem.*, **2008**, *283*, 6631.
274. R. Apweiler, H. Hermjakob and N. Sharon, *Biochim. Biophys. Acta*, **1999**, *1473*, 4.
275. R. A. Dwek, *Chem. Rev.*, **1996**, *96*, 683.
276. A. Varki, *Glycobiology*, **1993**, *3*, 97.
277. R. J. Payne and C.-H. Wong, *Chem. Commun.*, **2010**, 21.
278. D. Rauh and H. Waldmann, *Angew. Chem. Int. Ed.*, **2007**, *46*, 826.
279. C. Chatterjee, R. K. McGinty, J.-P. Pellois and T. W. Muir, *Angew. Chem. Int. Ed.*, **2007**, *46*, 2814.
280. R. S. Rush, P. L. Derby, D. M. Smith, C. Merry, G. Rogers, M. F. Rohde and V. Katta, *Anal. Chem.*, **1995**, *67*, 1442.
281. H. Saaki, B. Bothner, A. Dell and M. Fukuda, *J. Biol. Chem.*, **1987**, *262*, 12059.
282. M. Takeuchi, N. Inoue, T. W. Strickland, M. Kubota, M. Wada, R. Shimizu, S. Hoshi, H. Kozutsumi, S. Takasaki and A. Kobata, *Proc. Natl. Acad. Sci. USA*, **1989**, *86*, 7819.
283. L. O. Narhi, T. Arakawa, K.-H. Aoki, R. Elmore, M. Rohde, T. Boone and T. W. Strickland, *J. Biol. Chem.*, **1991**, *266*, 23022.
284. K.Yamaguchi, K. Akai, G. Kawanishi, M. Ueda, S. Masuda and R. Sasaki, *J. Biol. Chem.*, **1991**, *266*, 20434.
285. L. C. Wasley, G. Timony, P. Murtha, J. Stoudemire, A. J. Dorner, J. Caro, M. Krieger and R. J. Kaufman, *Blood*, **1991**, *77*, 2624.
286. M. Higuchi, M. Oh-eda, H. Kuboniwa, K. Tomonoh, Y. Shimonaka and N. Ochi, *J. Biol. Chem.*, **1992**, *267*, 7703.
287. J.-P. Boissel, W.-R. Lee, S. R. Presnell, F. E. Cohen and H. F. Bun, *J. Biol. Chem.*, **1993**, *268*, 15983.
288. M. R. Fibi, P. Hermentin, J. U. Pauly, L. Lauffer and G. Zettlmeissl, *Blood*, **1995**, *85*, 1229.
289. T. Misaizu, S. Matsuki, T. W. Strickland, M. Takeuchi, A. Kobata and S. Takasaki, *Blood*, **1995**, *86*, 4097.

290. S. Elliott, D. Chang, E. Delorme, C. Dunn, J. Egrie, J. Griffin, T. Lorenzini, C. Talbot and L. Hesterberg, *Blood*, **1996**, *87*, 2714.
291. P. L. Storring, R. J. Tiplady, R. E.G. Das, B. Rafferty and G. Misty, *J. Endocrinol.*, **1996**, *150*, 401.
292. Z. Tan, S. Shang, T. Halkina, Y. Yuan and S. J. Danishefsky, *J. Am. Chem. Soc.*, **2009**, *131*, 5424.
293. Y. Yuan, J. Chen, Q. Wan, Z. Tan, G. Chen, C. Kan and S. J. Danishefsky, *J. Am. Chem. Soc.*, **2009**, *131*, 5432.
294. C. Kan, J. D. Trzupek, B. Wu, Q. Wan, G. Chen, Z. Tan, Y. Yuan and S. J. Danishefsky, *J. Am. Chem. Soc.*, **2009**, *131*, 5438.
295. Y.-Y. Yang, S. Ficht, A. Brik and C.-H. Wong, *J. Am. Chem. Soc.*, **2007**, *129*, 7690.
296. J. Kang and D. Macmillan, *Org. Biomol. Chem.*, **2010**, *8*, 1993.
297. S. T. Cohen-Anisfeld and P. T. Lansbury Jr., *J. Am. Chem. Soc.*, **1993**, *115*, 10531.
298. R. Roy and F. Tropper, *Syn. Commun.*, **1990**, *20*, 2097.
299. M. Bejugam and S. L. Flitsch, *Org. Lett.*, **2004**, *6*, 4001.
300. N. J. Davis and S. L. Flitsch, *Tetrahedron Lett.*, **1991**, *32*, 6793.
301. B. Belleau, R. Martel, G. Lacasse, M. Ménard, N. L. Weinberg and Y. G. Perron, *J. Am. Chem. Soc.*, **1968**, *90*, 823.
302. B. Belleau and G. Malek, *J. Am. Chem. Soc.*, **1968**, *90*, 1651.
303. T. Inazu and K. Kobayashi, *Synlett*, **1993**, *11*, 869.
304. M. Mizuno, I. Muramoto, K. Kobayashi, H. Yaginuma and T. Inazu, *Synthesis*, **1999**, *1*, 162.
305. K. J. Doores, Y. Mimura, R. A. Dwek, P. M. Rudd, T. Elliott and B. G. Davis, *Chem. Commun.*, **2006**, 1401.
306. F. M. Ibatullin and S. I. Selivanov, *Tetrahedron Lett.*, **2009**, *50*, 6351.
307. J. P. Richardson, C.-H. Chan, J. Blanc, M. Saadi and D. Macmillan, *Org. Biomol. Chem.*, **2010**, *8*, 1351.
308. D. Macmillan, A. M. Daines, M. Bayrhuber and S. L. Flitsch, *Org. Lett.*, **2002**, *4*, 1467.