The Chemical Synthesis of Proteins

by

Gail Helen Morton

A thesis submitted for the
degree of Doctor of Philosophy

University of Edinburgh
September 1997
This thesis is submitted in part fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Edinburgh. Unless otherwise stated, the work described is original and has not been previously submitted in whole or in part, for any degree at this, or any other university.

University of Edinburgh
September 1997
To my parents and brother Alastair
for their endless love and support
Acknowledgements

First I would like to express my sincere thanks to Professor R. Ramage for the provision of research facilities and for his supervision, constant support and encouragement throughout the course of my PhD at Edinburgh.

I also wish to thank Dr. S.L. Irving (Pfizer Ltd, Sandwich, Kent) for his advice and assistance and Pfizer Ltd for the provision of funds.

I am extremely grateful the technical support provided by Kevin Shaw and Brian Whigham. Thanks must go to Kevin for his assistance with the chemical syntheses of the polypeptides and to Brian for the MALDI-TOF MS and amino acid analyses.

I am also indebted to a number of people outwith the research group for all of their valuable help and advice during my stromelysin work. Thanks must go to Dr. David Dryden for his advice and assistance with the fluorescence studies, Dr. Andrew Cronshaw (Welmet Protein Characterisation Facility) for the rapid and efficient N-terminal sequencing analyses and finally Dr. Sharon Kelly (University of Stirling) for her assistance and guidance with the circular dichroism studies.

Special thanks must go to Dr. Nicola Robertson and Dr. Lu Jiang for their practical assistance and support throughout the course of my PhD. Furthermore I am extremely grateful to Nicola, Alastair Hay and Craig Jamieson for proof reading this thesis - thanks to you all, your time and patience was greatly appreciated!

Finally, I would like to thank all my friends and colleagues, past and present, who have made my time in Edinburgh so enjoyable and memorable. Special thanks must go to all the members of the Ramage group, who have helped to make my three years in lab 29 so entertaining and unforgettable.
Abstract

The viability of extending the present methodology designed for the solid phase synthesis of peptides has been investigated. Using the catalytic domain of stromelysin (SCD, 173 residues) as the model system, a number of different factors affecting the preparation and purification of chemically synthesised proteins are examined.

The purification SCD, prepared using stepwise solid phase synthesis is described. Following characterisation, it is evident that protein of the correct primary sequence has been prepared, furthermore, preliminary studies centred on the enzymatic activity of SCD indicate that active protease has been isolated. However, comparison of the conformational and biophysical properties of chemically synthesised SCD with the recombinant counterpart, suggests that there are problems associated with the folding of the synthetic SCD.

The construction of chemically synthesised SCD via convergent protein synthesis is also described. Two different coupling strategies involving classical and azide fragment condensation are examined where it has been highlighted that the overall success of each fragment coupling is greatly dependent on the peptide length and sequence. As well as comparing methods for the preparation and coupling of fully and minimally protected peptides, general procedures for both solution and solid phase fragment coupling are discussed.

A novel strategy for the convergent synthesis of peptides and proteins has been investigated. In this total chemical synthesis approach, two minimally protected peptides are joined through unique, mutually reactive functional groups, yielding a peptide analogue with a thioether replacement for the native peptide bond at the site of ligation. A general route to C-terminal sulfhydryl and N-terminal haloacetylated peptides is presented, accompanied with results of the preliminary ligation studies.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA</td>
<td>amino acid analysis</td>
</tr>
<tr>
<td>ABI</td>
<td>applied biosystems</td>
</tr>
<tr>
<td>actSCD</td>
<td>activated SCD</td>
</tr>
<tr>
<td>Acm</td>
<td>acetamidomethyl</td>
</tr>
<tr>
<td>AcOH</td>
<td>acetic acid</td>
</tr>
<tr>
<td>AMPA</td>
<td>aminophenylmercuric acetate</td>
</tr>
<tr>
<td>Boc</td>
<td>tertiary-butoxycarbonyl</td>
</tr>
<tr>
<td>Bum</td>
<td>N\textsuperscript{-}tertiary-butoxymethyl</td>
</tr>
<tr>
<td>CD</td>
<td>circular dichroism</td>
</tr>
<tr>
<td>d</td>
<td>doublet</td>
</tr>
<tr>
<td>DCC</td>
<td>N,N-dicyclohexylcarbodiimide</td>
</tr>
<tr>
<td>DCM</td>
<td>dichloromethane</td>
</tr>
<tr>
<td>DEAE</td>
<td>diethylaminoethyl</td>
</tr>
<tr>
<td>DIC</td>
<td>N,N-diisopropylcarbodiimide</td>
</tr>
<tr>
<td>DIEA</td>
<td>N,N-diisopropylethylamine</td>
</tr>
<tr>
<td>DIU</td>
<td>N,N-diisopropylurea</td>
</tr>
<tr>
<td>DMAP</td>
<td>4-dimethylaminopyridine</td>
</tr>
<tr>
<td>DMF</td>
<td>dimethylformamide</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethylsulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>Dnp</td>
<td>dinitrophenyl</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiolthreitol</td>
</tr>
<tr>
<td>EDT</td>
<td>ethanedithiol</td>
</tr>
<tr>
<td>FAB</td>
<td>fast atom bombardment</td>
</tr>
<tr>
<td>Fmoc</td>
<td>9-fluorenylmethyloxycarbonyl</td>
</tr>
<tr>
<td>FPLC</td>
<td>fast protein liquid chromatography</td>
</tr>
<tr>
<td>Gdm.HCl</td>
<td>guanidinium hydrochloride</td>
</tr>
<tr>
<td>HBTU</td>
<td>O-benzotriazol-1-yl-N,N,N',N\textsuperscript{-}tetramethyluronium-hexafluorophosphate</td>
</tr>
<tr>
<td>HF</td>
<td>hydrogen fluoride</td>
</tr>
<tr>
<td>HFIP</td>
<td>hexafluoro-2-propanol</td>
</tr>
<tr>
<td>HIC</td>
<td>hydrophobic interaction chromatography</td>
</tr>
<tr>
<td>HIV</td>
<td>human immunodeficiency virus</td>
</tr>
<tr>
<td>HOAt</td>
<td>1-hydroxy-7-azabenzotriazole</td>
</tr>
<tr>
<td>HOBt</td>
<td>1-hydroxybenzotriazole</td>
</tr>
<tr>
<td>HOCt</td>
<td>ethyl-1-hydroxy-1\textsuperscript{H}-1,2,3-triazole-4-carboxylate</td>
</tr>
<tr>
<td>HOOBt</td>
<td>1-hydroxy-3-hydroxy-4-oxo-1,2,3-benzotriazine</td>
</tr>
<tr>
<td>HPLC</td>
<td>high performance liquid chromatography</td>
</tr>
<tr>
<td>HRMS</td>
<td>high resolution mass spectrometry</td>
</tr>
<tr>
<td>IEF</td>
<td>isoelectric focusing</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>IR</td>
<td>infrared</td>
</tr>
<tr>
<td>m</td>
<td>multiplet</td>
</tr>
<tr>
<td>MALDI</td>
<td>matrix assisted laser desorption ionisation</td>
</tr>
<tr>
<td>Mca</td>
<td>methoxycoumarin</td>
</tr>
<tr>
<td>MMA</td>
<td>N-methyl-mercaptoacetamide</td>
</tr>
<tr>
<td>MMP</td>
<td>matrix metalloproteinase</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger ribonucleic acid</td>
</tr>
<tr>
<td>MS</td>
<td>mass spectroscopy</td>
</tr>
<tr>
<td>Mtr</td>
<td>4-methoxy-2,3,6-trimethylbenzenesulfonyl</td>
</tr>
<tr>
<td>Mwt.</td>
<td>molecular weight</td>
</tr>
<tr>
<td>Mwt.CO</td>
<td>molecular weight cut off</td>
</tr>
<tr>
<td>nm</td>
<td>nanometre</td>
</tr>
<tr>
<td>NMP</td>
<td>N-methyl-2-pyrrolidinone</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PGC</td>
<td>porous graphitised carbon</td>
</tr>
<tr>
<td>Pmc</td>
<td>2,2,5,7,8-pentamethylchroman-6-sulfonyle</td>
</tr>
<tr>
<td>proSCD</td>
<td>preactivated SCD</td>
</tr>
<tr>
<td>recomSCD</td>
<td>recombinant SCD</td>
</tr>
<tr>
<td>RT</td>
<td>room temperature</td>
</tr>
<tr>
<td>R&lt;sub&gt;t&lt;/sub&gt;</td>
<td>retention time</td>
</tr>
<tr>
<td>s</td>
<td>singlet</td>
</tr>
<tr>
<td>SCD</td>
<td>stromelysin catalytic domain</td>
</tr>
<tr>
<td>s.c.p</td>
<td>side-chain protection</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulfonate</td>
</tr>
<tr>
<td>SPFC</td>
<td>solid phase fragment condensation</td>
</tr>
<tr>
<td>SPPS</td>
<td>solid phase peptide synthesis</td>
</tr>
<tr>
<td>t</td>
<td>triplet</td>
</tr>
<tr>
<td>t&lt;sub&gt;t&lt;/sub&gt;</td>
<td>tertiary</td>
</tr>
<tr>
<td>Tbfmoc</td>
<td>17-tetrabenzo[a,c,g,l]fluorenylmethoxycarbonyl</td>
</tr>
<tr>
<td>TEA</td>
<td>triethylamine</td>
</tr>
<tr>
<td>TFA</td>
<td>trifluoroacetic acid</td>
</tr>
<tr>
<td>TFE</td>
<td>trifluoroethanol</td>
</tr>
<tr>
<td>TFMSA</td>
<td>trifluoromethanesulfonic acid</td>
</tr>
<tr>
<td>TIPS</td>
<td>triisopropylsilane</td>
</tr>
<tr>
<td>TMSOTf</td>
<td>trimethylsilyltrifluoroacetate</td>
</tr>
<tr>
<td>THF</td>
<td>tetrahydrofuran</td>
</tr>
<tr>
<td>TLC</td>
<td>thin layer chromatography</td>
</tr>
<tr>
<td>TOF</td>
<td>time of flight</td>
</tr>
<tr>
<td>Tris</td>
<td>tris(hydroxymethyl)aminomethane</td>
</tr>
<tr>
<td>Trt, Trityl</td>
<td>triphenylmethyl</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>WSCI</td>
<td>water soluble carbodiimide</td>
</tr>
<tr>
<td>Z</td>
<td>benzyloxy carbonyl</td>
</tr>
</tbody>
</table>
Contents

Chapter One: Solid Phase Peptide Synthesis

1.1 Introduction 1
1.2 The Chemical Synthesis of Peptides 2
1.3 Solid Phase Peptide Synthesis (SPPS) - A General Overview 5
1.4 Protecting Group Strategies 6
  1.4.1 Boc strategy 6
  1.4.2 Fmoc strategy 7
    1.4.2.1 Fmoc strategy - UV monitoring 8
    1.4.2.2 Fmoc strategy - protecting groups 8
1.5 Nature of the Polymer Support 10
1.6 The Resin Linkage 11
1.7 Peptide Bond Formation 14
  1.7.1 Racemisation during activation 14
  1.7.2 Methods for carboxyl activation 15
1.8 The Chemical Synthesis of Proteins 18
  1.8.1 Preparation of the target molecule 18
  1.8.2 Specific site alteration 19
1.9 Chemically Synthesis Proteins - A Brief History 20
1.10 Purification of Chemically Synthesised Proteins 20
1.11 Convergent Protein Synthesis - Fragment Coupling 22
1.12 References 25
Chapter Two: The Chemical Synthesis of the Catalytic Domain of Stromelysin (SCD)

2.1 Background 28
2.2 Isolation of the active enzyme 29
2.3 Structural Analysis of SCD 30
2.4 Inhibition of SCD 31
2.5 Research Overview 34
2.6 The Chemical Synthesis of SCD 35
2.7 Purification and Characterisation of SCD 38
   2.7.1 Gel filtration chromatography 38
   2.7.2 Characterisation of SCD 39
      2.7.2.1 Molecular weight determination 39
      2.7.2.2 MALDI-TOF MS and N-terminal sequencing 39
      2.7.2.3 Isoelectric point (pI) determination 41
2.8 Folding and Further Purification of SCD 42
   2.8.1 Folding of SCD in solution 43
   2.8.2 Matrix assisted folding of SCD 43
   2.8.3 Characterisation of SCD after ion exchange 45
2.9 Enzymatic Activity of SCD 46
   2.9.1 Hydrolysis of Substance P 46
   2.9.2 Hydrolysis of fluorogenic substrates 47
   2.9.3 Chemical denaturation and refolding of SCD 49
2.10 Conformational Analysis of SCD 50
   2.10.1 Circular dichroism studies on SCD 50
   2.10.2 Chemical denaturation - CD analysis 52
   2.10.3 Chemical denaturation - fluorescence analysis 54
2.11 Final Investigation into the Activity of Synthetic SCD 56
   2.11.1 Chromatographic analysis of active SCD 56
   2.11.2 Enzymatic activity of chemically denatured recom SCD 57
Chapter Three: The Convergent Synthesis of SCD

3.1 Introduction

3.2 Solid Phase Fragment Condensation (SPFC)
   3.2.1 Preparation of protected peptide fragments
   3.2.2 Choice of peptide fragments
   3.2.3 Insolubility of protected peptides

3.3 Literature Examples of Fragment Condensation

3.4 Research Overview
   3.4.1 Coupling strategy

3.5 SPFC. Preliminary Investigation
   3.5.1 Preparation and characterisation of H₂N-Pro⁹⁰·Thr¹⁷³-OH
   3.5.2 Synthesis and characterisation of FmocNH-Gly⁷⁹·Gly⁸⁹-OH
   3.5.3 Coupling of FmocNH-Gly⁷⁹·Gly⁸⁹-OH to H₂N-Pro⁹⁰·Thr¹⁷³-resin
   3.5.4 Tbfmoc purification and characterisation of H₂N-Gly⁷⁹·Thr¹⁷³-OH
   3.5.5 Coupling of Fmoc-Gly⁷⁰·Gly⁹⁰-OH and Fmoc-Phe⁵⁰·Gly⁸⁹-OH

3.6 The Convergent Synthesis of SCD
   3.6.1 Preparation of H₂N-Asp⁷¹·Thr¹⁷³-resin
   3.6.2 Preparation of H₂N-Glu⁵⁷·Thr¹⁷³-resin
      3.6.2.1 Stepwise synthesis of Glu⁵⁷·Gly⁷⁰
      3.6.2.2 Dissection of Glu⁵⁷·Gly⁷⁰
      3.6.2.3 Synthesis of Glu⁵⁷·Ala⁶⁵
   3.6.3 Attempted synthesis of Leu⁴⁸·Ala⁵⁸

3.7 The Convergent Synthesis of SCD. Review of Initial Findings

3.8 Convergent Synthesis Using Minimally Protected Peptides
   3.8.1 Principles of azide condensation
   3.8.2 Generation of the C-terminal azide
Chapter Four: The Chemical Ligation of Peptides

4.1 Background
4.2 Protein Synthesis via Native Chemical Ligation
4.3 Convergent Synthesis of Backbone Engineered Proteins
  4.3.1 Convergent synthesis via thioester bond formation
  4.3.2 Chemical ligation of three or more fragments
4.4 Research Overview
  4.4.1 Proposed strategy for chemical ligation
  4.4.2 Design of effective coupling conditions
4.5 Preparation of Ligation Fragments
  4.5.1 Synthesis of the functionalised resin-linker
  4.5.2 Synthesis of a C-terminal modified peptide
  4.5.3 Preparation of bromoacetylated peptides
4.6 Peptide Coupling via Thioether Bond Formation
  4.6.1 Chemical ligation - Preliminary investigation
  4.6.2 Investigation into altering the rate of substitution
  4.6.3 Synthesis of an analogue of human proinsulin C-peptide
4.7 Final Conclusions and Recommendations
4.8 References
Chapter Five: Experimental

5.1 Notes 119
5.2 Solid Phase Peptide Synthesis 121
  5.2.1 Side-chain protecting groups 121
  5.2.2 Loading of C-terminal amino acids 121
  5.2.3 The Fmoc loading test 123
  5.2.4 Automated SPPS 123
  5.2.5 Preparation of Tbfmoc-peptides 125
  5.2.6 Acidolytic treatment of resin bound peptide 126
5.3 The Chemical Synthesis of SCD 127
5.4 Convergent Synthesis of SCD - Fragment Condensation 136
5.5 Convergent Synthesis of SCD - Azide condensation 147
5.6 Chemical Ligation 150
5.7 References 158
Chapter One

Solid Phase Peptide Synthesis

1.1 Introduction

Peptides and proteins play crucial roles in virtually all biological processes. The chemical and biophysical properties of these complex systems are completely diverse, thus the range of applications is extensive and extremely varied. For example, a number of these biologically active molecules are extremely toxic or responsible for a wide range of pathological diseases, while, on the other hand, other systems have been found to be of great therapeutic value.

The breadth and depth of our understanding of how these biologically significant systems function has increased at an extremely fast pace. For a number of years, biologically active peptides have been objects of intense interest in the research community and are the subjects of thousands of research programmes. More recently, there has been an upsurge in interest in the much larger, more complex protein molecules which have now become more attainable. This has been as a result of the developments in modern molecular biology which has led to the availability of amino acid sequence data determined by the cloning of and cDNA sequencing of genes.

The demand for these complex molecules is thus enormous and rising all of the time. Although the target molecule can normally be isolated from its native source or prepared using modern molecular biology, the development of an alternative, total synthetic approach has resulted in an effective and well-established method for the synthesis of biologically significant systems. Furthermore, chemical synthesis is now recognised as a highly versatile method, allowing systematic variation of the primary structure with the aim of developing second-generation products for therapeutic use.
1.2 The Chemical Synthesis of Peptides

The preparation of a peptide from its constituent amino acids using conventional solution phase chemistry is far from trivial. The problems encountered can easily be illustrated with the synthesis of a simple dipeptide AB (Fig. 1.1).

![Diagram of dipeptide synthesis]

This clearly underlines the fundamental problem associated with the synthesis of peptides. Without control over these highly functionalised species, the coupling of two amino acids would result in formation of four possible dipeptides, accompanied by the uncontrollable extension of the N and C-terminus (contamination with tripeptides, tetrapeptides etc). Furthermore a number of side-chain functionalities are also capable of reacting in the coupling conditions, thus it is obvious that the number of possible cross-reactions and resulting by-products would be extensive.

Fig. 1.1 Uncontrolled synthesis of a dipeptide AB
This problem can be avoided by the introduction of protecting groups (Fig. 1.2) which ensure the unambiguous synthesis of AB. In addition, if the protecting groups employed are labile under different conditions, selective deprotection of the α-amino or α-carboxyl moieties will allow subsequent elongation at either terminus of the peptide.

![Fig. 1.2 Controlled synthesis of a dipeptide AB](image)

Although the protecting group strategy permits the preparation of peptides with the correct sequence of amino acids, as with all solution phase chemistry, the intermediates require purification. As a result, even the assembly of a relatively short peptide sequence would be labour intensive, requiring the efforts of a number of highly skilled and experienced chemists.

The potential for peptide and protein synthesis was therefore revolutionised when, in 1963, Merrifield\(^1\) presented the brilliantly simple idea of attaching the carboxyl group of the C-terminal amino acid to an insoluble support from which the chain of amino acids could be elongated (Fig. 1.3).

Even at the onset of this solid phase peptide synthesis (SPPS), Merrifield\(^2\) identified the underlining advantages of employing the solid support -

- an excess of the incoming amino acid could be used to drive the reaction to completion.
- since the support and covalently attached peptide would be insoluble, excess reagents, by-products and solvents could be removed by simple filtration and washing procedures.
- all reactions could be conducted in a single reaction vessel so it seemed reasonable that the process would lend itself to automation.
Since its initial design, the intervening 35 years has seen extensive research efforts worldwide, which have concentrated on the development of methodology and the design of a wide range of reagents to improve the synthetic capabilities of SPPS. The result is a highly efficient automated system capable of preparing polypeptides and small proteins which would be insurmountable using solution phase chemistry. Unsurprisingly, considering the advantages of SPPS, the general principle of solid phase chemistry has been extended to other fields leading to the now established solid phase oligonucleotide synthesis as well as the development of methodology for the synthesis of oligosaccharides using insoluble supports. Furthermore, in organic synthesis, solid phase methods have led to the optimisation of a vast number of standard chemical reactions on insoluble supports. As well as providing a beneficial alternative to solution phase chemistry, solid phase organic chemistry is now considered to be of significant importance for the fast, simultaneous and multiple synthesis of many new organic compounds required in the search for lead drug candidates (combinatorial chemistry).
1.3 Solid Phase Peptide Synthesis (SPPS) - A General Overview

The following sections are designed to summarise briefly how SPPS has advanced over the past three decades and presents some of the reagents and strategies which are considered to have made a significant contribution to its overall development. Fig. 1.4 outlines the fundamental operations involved in one cycle of peptide chain extension.

![Diagram of SPPS process]

**Target Sequence**

*Fig. 1.4 General operations involved in SPPS*

The following aspects shall be individually considered -

- protecting group strategies
- nature of the polymer support
- the resin linkage
- peptide bond formation (carboxyl activation)
1.4 Protecting Group Strategies

One of the most important considerations in SPPS (outlined in fig. 1.4) is that the deprotection protocols required for the \( N^a \) and side-chain functionalities are orthogonal. The side-chain protecting groups must maintain complete integrity throughout, whilst the appropriate \( N^a \)-protection is cleaved at each cycle of the peptide synthesis. Fig. 1.5 summaries the two most effective protecting group strategies, Boc and Fmoc, both of which have been successfully utilised in automated SPPS.

![Diagram of Boc and Fmoc strategies]

**Fig. 1.5 Protecting group strategies in SPPS**

1.4.1 Boc strategy

The first of these two orthogonal protocols is based upon the acid lability of the \( t \)-butoxy-carbonyl (Boc) group, the principles of which were established by Merrifield with the synthesis of bradykinin.\(^6\) During assembly of the peptide chain, trifluoroacetic acid (TFA) treatment rapidly cleaves the Boc protection, ensuring the quantitative release of the \( N^a \)-amino functionality. Meanwhile, all side-chain functionalities remain blocked throughout by employing semi-permanent protection namely the benzyloxy carbonyl group (Z), benzyl ethers and esters. These benzyloxy groups are then cleaved from the side-chain functional groups at the end of the synthesis using much stronger acid conditions such as hydrogen fluoride (HF) or trifluoromethanesulfonic acid (TFMSA).
1.4.2 Fmoc strategy

The development of Fmoc SPPS arose out of the concern that repetitive TFA acidolysis in Boc group deprotection could lead to alteration of sensitive peptide bonds as well as acid catalysed side-reactions. In addition, the final deprotection and cleavage from the solid support requires very harsh acidolysis using highly corrosive HF. Not only do these harsh conditions clearly present the risk of affecting the peptide product, but there is also the requirement for specialised laboratory apparatus and very careful handling.

Thus an alternative to Boc was presented with the design of the base labile 9-fluorenymethoxycarbonyl (Fmoc) (1) by Carpino\(^7\)\(^,\)\(^8\) which was elegantly applied independently by Meienhofer\(^9\) and Sheppard\(^10\) to SPPS. This base labile protecting group not only provided an attractive alternative to Boc synthesis, but could also be described as a truly orthogonal approach to SPPS. In this milder, more specific strategy, the growing peptide is subjected to base treatment using piperidine which cleaves the Fmoc moiety via \(\beta\)-elimination (Fig. 1.6). On completion of the peptide’s synthesis, the base stable side-chain protection may then be removed by TFA treatment under relatively mild conditions.

![Fig. 1.6 Fmoc deprotection using piperidine](image-url)
1.4.2.1 Fmoc strategy - UV monitoring of coupling efficiency

In addition to the milder reaction conditions, Fig. 1.6 highlights a further advantage of the Fmoc strategy. Once released from the growing peptide chain, the dibenzofulvene (2) is quenched with excess piperidine, forming the fulvene-piperidine adduct (3). The isosbestic point of (3) occurs at 302nm, therefore, a useful estimation of the coupling efficiency of each amino acid can be made by measuring the absorbance of the deprotection effluent. This is of substantial benefit to the peptide chemist as it provides an overall profile of the peptide synthesis as well as highlighting possible problematic areas within the peptide sequence.

1.4.2.2 Fmoc strategy - protecting groups

With the introduction of base labile Nα protection, clearly the requirement for side-chain blocking is to use protecting groups which are base stable and extremely labile under mild acid conditions. Although the majority of functionalities are efficiently protected using triphenylmethyl (Trt), t-butyl ethers or t-butyl esters, special consideration is necessary for the protection of the side-chains of arginine, cysteine and histidine. Table 1.1 summarises the most common protecting groups used.

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Side-chain P.G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg</td>
<td>Mtr/Pmc</td>
</tr>
<tr>
<td>Asn/Gln</td>
<td>Trt</td>
</tr>
<tr>
<td>Asp/Glu</td>
<td>Bu' esters</td>
</tr>
<tr>
<td>Cys</td>
<td>Acm/Trt/SBu'</td>
</tr>
<tr>
<td>His</td>
<td>τ-Trt/β-But</td>
</tr>
<tr>
<td>Lys</td>
<td>Boc</td>
</tr>
<tr>
<td>Ser/Thr/Tyr</td>
<td>Bu' ethers</td>
</tr>
<tr>
<td>Trp</td>
<td>Boc</td>
</tr>
</tbody>
</table>

Table 1.1 Side-chain protection employed in Fmoc SPPS

Arginine protection - The trifunctional guanidino side-chain group of Arg is strongly basic thus is easily acylated during SPPS if not protected. Unlike other amino functionalities, blocking the side-chain with Boc\textsuperscript{11} is not sufficient with the
nucleophilicity of the guanidino function still capable of leading to undesired side-reactions and subsequent intramolecular decomposition during acid deprotection.\textsuperscript{12,13}

In contrast to urethanes, a single arene sulfonyl based Arg protecting group has proved to be a successful alternative leading to the design of derivatives such as 4-methoxy-2,3,6-trimethylbenzenesulfonyl (Mtr)\textsuperscript{14,15} (4), and 2,2,5,7,8-pentamethylchroman-6-sulfonyl (Pmc)\textsuperscript{16,17} (5). Considering Mtr requires stronger acid treatment (TFMSA) for cleavage of peptides containing multiple Arg residues, Pmc is presently the protecting group of choice in Fmoc SPPS.

\begin{center}
\begin{tabular}{c}
(4) Mtr \\
(5) Pmc
\end{tabular}
\end{center}

**Histidine protection** - The most serious problem encountered with His is racemisation due to the basicity of the \(\pi\)-nitrogen in the imidazole ring. Fig. 1.7 shows how the basic nitrogen is capable of proton abstraction from the \(\alpha\)-carbon at which point the stereochemical integrity is lost.

\begin{center}
\textbf{Fig. 1.7 Racemisation of histidine}
\end{center}

This stereochemical problem has been shown to be completely avoided by blocking the \(\pi\)-nitrogen with \(\tau\)-butoxymethyl (Bum).\textsuperscript{18} However due to its complex synthesis
and resulting cost, this derivative is not commonly employed. Alternatively, it has been found that the τ-nitrogen, blocked with the bulky trityl group\(^{19}\) significantly reduces the basicity of the imidazole. In addition, this proton abstraction can be suppressed even further by employing additives during synthesis.\(^{20}\)

**Cysteine protection** - The synthesis of cysteine containing peptides is often complicated because some peptide products require the cysteine residue to be in the free sulphydryl form while others require inter or intramolecular disulfide bonds. As a result, there are currently a number of protecting groups available, the choice depending on whether free or protected cysteines are required during purification. The most commonly used Cys protection is Trityl which generates the free thiol upon deprotection using TFA. Alternatively, Cys protection can be retained if the acetamidomethyl (Acm)\(^{22}\) or \(t\)-butylthio (SBut') groups are used. Removal of Acm is achieved by treatment with silver trifluoromethanesulfonate\(^{21}\) or mercury (II) acetate,\(^{22}\) while SBut' can be removed by reduction with thiols eg. \(\beta\)-mercaptoethanol\(^{23}\) or by using tributylphosphine.\(^{24}\)

### 1.5 Nature of the Polymer Support

The most fundamentally crucial consideration in SPPS has to be the solid support itself. In order for any chemistry to proceed effectively, the insoluble resin must possess a number of physical and chemical properties -

- the support should be inert towards, and insoluble in, all solvents and reagents used.
- functionality must be incorporated to allow attachment of the growing peptide chain.
- structurally, the support must be microporous, swelling to expose the large surface area allowing rapid access to all reagents and solvents.

The original resin used by Merrifield\(^1\) was a cross-linked form of polystyrene containing reactive chloromethyl groups. This was prepared by polymerisation of
styrene plus 1% divinylbenzene to produce a rigid cross-linked resin followed by chloromethylation. The C-terminal amino acid was fixed to the polymer via a nucleophilic substitution of the Cl by the carboxyl group (Fig. 1.8).

Following on from the success of the Merrifield polystyrene support, alternative resins have been designed and incorporated into SPPS. Possibly the most notable of these are the polyamide resins developed by Sheppard.\textsuperscript{2} The proposed advantage of such a polyamide support was that by increasing the polarity of the linker, this more hydrophilic functionalised resin should aid the solvation of the growing peptide chain, hence improving the efficiency of the synthesis.

1.6 The Resin Linkage

As previously discussed, the original design of SPPS involved attachment of the peptide directly to the functionalised resin. This benzyl ester link to the support was complementary to the initial Boc strategy in that the bond withstood the repetitive TFA treatment. Additionally, both the benzyl side-chain deprotection and release of the peptide would be effected by final strong acid (HF) treatment. However, considering that the Fmoc strategy employs a final, milder acid treatment, clearly the incorporation of a more acid labile linker system was viewed as having a significant advantage.
The introduction of linkers between the support and the peptide made it possible to alter the strength of the peptide-resin bond making it more or less labile depending on the synthetic strategy. The difference in acid lability can be demonstrated by considering two of the most commonly used linkers applied in Boc and Fmoc SPPS (Fig. 1.9).

![Fig. 1.9 PAM and Wang linkers used in SPPS](image)

Although with both the Fmoc and Boc methodologies the peptide cleavage mechanism is ester hydrolysis by alkyl-oxygen fission, the Wang linker has a greater acid lability. This is due to the additional $p$-oxygen which aids the stabilisation of the carbocation formed during the fission process (Fig. 1.10).

![Fig. 1.10 Carbocation stabilisation of the Wang linker](image)
Clearly the incorporation of these linkers makes it possible to closely control the acid lability of the peptide link. Subsequently a number of linkers have been designed which have a variety of applications. For example, depending on the linker, it is now possible to prepare peptides with a number of C-terminal modifications. In addition, a number of highly acid labile linkers have been developed permitting the preparation of peptides with the side-chain protection fully intact. Such side-chain protected, or C-terminally modified peptides are of great importance for a number of strategies applied to peptide fragment coupling (chapters 3 and 4). Fig. 1.11 summarises a selection of these versatile linkers, highlighting their properties and applications.

**Fig. 1.11 Examples of linkers employed in Fmoc SPPS.**
1.7 Peptide Bond Formation

As with all other aspects of peptide synthesis, the methodology associated with peptide bond formation has been extensively studied. Overall, optimal conditions should be such that amide bond formation is carried out rapidly and quantitatively under mild conditions, avoiding any side-reactions or changes in the stereochemical integrity.

In general, the carboxyl group of the amino acid being coupled is initially activated; subsequent reaction with the Nα-amino on the growing peptide chain results in amide formation (Fig. 1.12). The activation of the amino acid is simply the conversion of the carboxyl group to a more reactive derivative allowing the nucleophilic amine attack to proceed. The underlining requirement of such an ‘activating group’ (A) is that it has electron withdrawing properties thus increasing the electrophilicity of the carboxyl group making it more susceptible to nucleophilic attack.

![Fig. 1.12 Formation of the amide peptide link](image)

1.7.1 Racemisation during activation
Stereochemical considerations are the major underlining factor when designing effective activating and coupling procedures. Care must be taken to ensure that on activating, deprotonation of the α-carbon does not occur (Fig. 1.13). This would result in the racemisation of the amino acid (6) via the oxazolone intermediate (7, 9). Once formed, the chiral integrity is lost and subsequent reaction of the amine yields peptides containing both the L (8) and D (11) epimers. Fortunately this side reaction is not encountered to any extent in the chain assembly using the aforementioned Boc or Fmoc strategies where activation of the carboxylic acid is performed on the Nα-
1. Solid Phase Peptide Synthesis

urethane protected α-amino acid which, using standard conditions, does not form the oxazolone.

![Diagram of amino acid racemisation during activation and coupling]

**Fig. 1.13 Amino acid racemisation during activation and coupling**

1.7.2 Methods for carboxyl activation

Unsurprisingly, the range of suitable activating groups and procedures is extensive, however this section shall concentrate only on some of the more general methods applied to Fmoc SPPS. Fig. 1.14 highlights the more commonly applied coupling techniques.

![Diagram of general procedures for peptide bond formation]

**Fig. 1.14 General procedures for peptide bond formation**
1. Solid Phase Peptide Synthesis

Carbodiimides

Carbodiimides are undoubtedly the most important reagents of their kind used for carboxylic activation prior to amide bond formation. Ever since Sheehan and Hess presented the utility of $N,N'$-dicyclohexylcarbodiimide (DCC)\textsuperscript{32} (14), carbodiimide reagents have been applied extensively for carboxyl activation. Subsequently the preparation of $N,N'$-diisopropylcarbodiimide (DIC)\textsuperscript{33} (13) provided a preferred alternative to DCC for SPPS. This is simply because the $N,N'$-diisopropyl urea by-product (19) is fully soluble in the solvents used in SPPS, compared with the relatively insoluble $N,N'$-dicyclohexylurea.

As outlined in Fig. 1.14, these dicarbodiimides are employed for activation using one of two strategies. The initial activated carboxylic acid derivative (15) can either be reacted directly with an amine or with a variety of nucleophiles to produce more stable activated intermediates such as the corresponding symmetrical anhydride (17) or activated ester (16).

Symmetrical Anhydrides

Due to the general ease of preparation and rapid, unambiguous acylation properties, amino acid activation using the symmetrical anhydride (17) is a highly popular method for amide bond formation, used in both Boc and Fmoc SPPS.

Unfortunately there are two major drawbacks associated with anhydride activation. Firstly, the preparation of symmetrical anhydrides is far from cost effective since two equivalents of amino acid are required to form one equivalent of activated species. Considering the expense of Fmoc protected amino acids, a minimum of 50\% loss of reagent will be costly. The other serious problem encountered with both the carbodiimide (15) and symmetrical anhydride (17) derivatives is the undesired side-reaction involving the side-chain of asparagine and glutamine. DIC is the sole reagent used for carboxyl activation yielding (15) and (17). Unfortunately DIC can also act as a dehydrating agent with the side-chain amides of Asn and Gln leading to the irreversible nitrile formation (Fig. 1.15).\textsuperscript{34}
1. Solid Phase Peptide Synthesis

![Chemical Structures]

**Fig. 1.15** DIC induced dehydration of the asparagine side-chain

**Activated Esters**

The final amino acid derivatives to be discussed in this review are the activated esters (16). Considering the disadvantages of aminolysis via carbodiimides or symmetrical anhydrides, namely side-chain dehydration and 50% reagent inefficiency, a practical alternative would be extremely beneficial.

Originally designed by Konig and Geiger to inhibit racemisation,\(^\text{35}\) electronegative 1-hydroxybenzotriazole (HOBt) (21) was soon employed as a highly effective coupling reagent. Such HOBt esters of protected amino acids are easily prepared *in situ* from one equivalent of Fmoc-amino acid, one equivalent DIC and one equivalent HOBt. Although less reactive than the corresponding symmetrical anhydride, these activated esters proceed unambiguously to form the peptide link on aminolysis without side-chain rearrangement.

Using the highly successful HOBt as a structural template, a number of analogues have been developed as coupling reagents, these currently include 1-hydroxy-7-azo-benzotriazole (HOAt)\(^\text{36}\) (20) and ethyl 1-hydroxy-1H-1,2,3-triazole-4-carboxylate (HOCt)\(^\text{20,37,38}\) (22).
1.8 The Chemical Synthesis of Proteins

Advances in synthetic and analytical methodology over the last 35 years has resulted in a highly optimised automated stepwise SPPS which is capable of the rapid and reproducible assembly of long protected peptide chains. In general, the coupling and deprotection efficiencies are >99% effective thus the total chemical synthesis of large polypeptides and small proteins is now considered a practical and accessible goal.

In recent years, the growth in our awareness and understanding of protein structure and function has been extensive. To date, the majority of proteins studied have been isolated from their natural source or prepared by exploiting recent advances in molecular biology. However, considering the continuous improvements in SPPS, it is clear that chemically synthesised proteins have a substantial contribution to make. In summary, the advantages of preparing proteins using chemical methods are -

- the reproducible construction of the target protein.
- ease of primary structure variation.

1.8.1 Preparation of the target molecule

To help build up the complete characterisation of a protein ideally sufficient quantities for full conformation and functional assessment (eg. three-dimensional NMR studies, x-ray crystal structure, biological assays etc) should be available. In practice this is not always possible. For example, consider the isolation of a protein from its native source. This is often a highly laborious task resulting in only minute quantities of material. However, provided the genetic information for the protein is available, more acceptable quantities may be prepared using recombinant technology. The isolated or synthesised gene can be translated into protein in large quantities by inserting the gene into an appropriate expression system. Once expressed, the protein is then isolated from the fermentation medium and purified. Although this is a significant improvement on native protein isolation, there are still problems encountered. Firstly, there is minimal control over protein expression. Complications may arise during translation resulting in the incorrect amino acid being incorporated into the peptide
1. Solid Phase Peptide Synthesis

chain. Secondly, isolation from the cell debris, DNA and host proteins may prove difficult, especially on large scale purification.

If however the protein is prepared from its constituent amino acids using chemical synthesis, the researcher has complete control over the construction of the target protein. Considering the ease and speed of synthesis, SPPS is often the most practical method for providing useful quantities of material.

1.8.2 Specific site alteration
To help provide further information on how individual residues contribute to the overall activity or conformation of a peptide or protein, alteration of the molecule at specific sites of interest is highly beneficial.

The sophisticated DNA technology now available is capable of making point mutations, leading to the expression of a range of modified proteins (site directed mutagenesis). However, the chemical synthesis approach to primary structure variation is a more flexible and controlled method which guarantees the incorporation of non-genetically encoded structures at the required site. For example by means of SPPS, artificial amino acids with isosteric side-chains can be used to investigate the functional importance of particular residues. In addition, NMR probe nuclei can be incorporated at any predetermined single atom site aiding NMR structure elucidation.

Finally, leading on from primary structure variation, more recently chemical synthesis has been employed for the preparation of completely unnatural molecules. The ultimate goal of this *de novo* design of protein synthesis is the construction of artificial proteins which, like the naturally occurring molecules, fold specifically, adopting predetermined conformations and exhibit tailor-made chemical and biological properties. This provides the peptide chemist with a greater understanding of the complex mechanisms associated with protein folding and scope for the development of a new generation of therapeutic products. (*de novo* reviews see refs 39,40)
1.9 Chemically Synthesised Proteins - A Brief History

Although very much still in the exploratory stages of development of SPPS, the first protein to be prepared using chemical methods was reported almost 30 years ago. This demonstrated that a protein molecule with true enzyme activity could be prepared from its constituent amino acids.

In the original publication, Gutte and Merrifield described the stepwise synthesis of ribonuclease A (RNAse).\textsuperscript{41} This was selected as the target because the 124 amino acid sequence had been determined\textsuperscript{42} and its crystal structure was known.\textsuperscript{43} The initial report was soon substantiated with a full and more detailed account which presented clear analytical data for protein homogeneity, composition and sequence.\textsuperscript{44} Not only did Gutte and Merrifield show that the synthetic RNAse A was indistinguishable from natural bovine pancreatic RNAse A, but also presented evidence that the synthetic counterpart had comparable specific activity (78%) and correct substrate specificity.

The chemical synthesis and characterisation of RNAse A was a considerable research achievement and this was reflected by the lack of publications in the following years describing the SPPS of other proteins of similar size. This clearly underlined the viability and limitations of the SPPS methodology 30 years ago. However the aim of protein synthesis has become much more accessible with the continuous advances in SPPS. This has been accompanied by the development of powerful purification strategies and highly accurate instrumentation for product analysis. In recent years numerous reports describing the stepwise synthesis of proteins have been published. Some of the more notable examples include ubiquitin (76 residues)\textsuperscript{45,46}, HIV-1 protease (99 residues),\textsuperscript{47,48} and interleukin-3 (140 residues).\textsuperscript{49}

1.10 Purification of Chemically Synthesised Proteins

The objective of protein synthesis is undoubtedly the production of homogeneous molecular species. However, in addition to obtaining protein with the correct primary
structure, to be biologically active the molecule must have defined secondary and tertiary structure. In the normal biologically active form, the protein chain is folded into a thermodynamically preferred three-dimensional structure, the conformation of which is maintained by relatively weak interatomic forces such as hydrogen bonding and hydrophobic and charge interaction.

Overall the target molecule can be separated from impurities by exploiting the differences in their physical and chemical properties, *ie* charge, molecular weight, hydrophobicity and biospecificity. Presently there is a vast range of techniques available which can be employed for chromatographic separation -

- ion exchange or isoelectric focusing
- gel filtration
- hydrophobic interaction
- HPLC
- affinity chromatography (eg. Tbfmoc purification - section 3.5.4)

Due to the diversity in their chemical and physical properties, no single purification strategy can be applied to all proteins. Instead the purification of each molecule is assessed individually with view to the development and optimisation of a suitable tailor-made protocol. Chapter 2 describes the isolation and purification of one such protein, the catalytic domain of stromelysin (SCD, 173 residues), prepared using stepwise solid phase methods. In addition to presenting the design of the purification strategy, the results of the protein conformational studies are discussed.

The most serious downfall of stepwise SPPS is that purification is only possible at the end of the synthesis. Consider the preparation of a 100 residue protein; every cycle comprises three chemical reactions, coupling, capping and deprotection. Following 300 chemical steps it is certain that the mixture will contain a number of side-products and impurities. Potentially the most problematic side-products are peptides or proteins containing deletion sequences. This occurs due to non-quantitative coupling, capping or N° deprotection giving rise to amino acid deficient sequences (Fig. 1.16) which can be too closely related to the target molecule in physical properties to allow separation.
1.11 Convergent Protein Synthesis - Fragment Coupling

One solution to such problems encountered with the stepwise construction of proteins involves the individual preparation of small peptide segments of the protein which are then linked together to produce the target sequence. This convergent synthesis has the advantage that the smaller peptide fragments are more manageable. Before coupling, each fragment can be purified and analysed confirming its sequence. Therefore, on coupling, the correct sequence is constructed in the absence of single amino acid deficiencies. Fig. 1.17 outlines a number of strategies for fragment coupling that have been previously applied successfully in the preparation of both peptide and protein chains.
**Fragment Condensation** - (Chapter 3).

- coupling conditions analogous to amino acid coupling (section 1.7)
- maximum side-chain protection necessary
- peptide fragment 1 can be in solution or fixed to a solid support
- due to low fragment solubility, couplings generally carried out in DMF or 1,4-dioxan

**Azide coupling** - (Chapter 3).

- coupling via amine attack of C-terminal azide
- minimum side-chain protection; only Lys and Cys side-chains require blocking
- improved solubility compared with original fragment condensation, couplings performed in DMF, 1,4-dioxan etc

**Chemical Ligation** - (Chapter 4).

- coupling reaction between two mutually reactive functional groups X and Y
- dependent on ligation design, side-chain protection may not be necessary
- generally fragments are soluble in aqueous conditions, buffer systems employed for coupling

*Fig. 1.17 Summary of the principal methods used for convergent synthesis*
The viability of such convergent methods has been clearly demonstrated with the successful syntheses of a number of proteins including prothymosin-α (109 residues, fragment condensation), ribonuclease A (124 residues, azide coupling), human interleukin-8 (72 residues, native chemical ligation) and analogues of HIV-1 protease (99 residues, chemical ligation). Examples in the literature describe methods using a variety of fragment lengths, using both protected and fully functional segments, with the couplings proceeding in solution or on a solid phase. Furthermore, a number of reports have illustrated that such methods are not restricted to amide bond formation but can be used to construct novel molecules with a wide range of possible backbone modifications.

In the proceeding chapters, different aspects concerning chemically synthesised proteins are discussed. Using the catalytic domain of stromelysin (SCD) as the model protein, chapters 2 and 3 deal with protein preparation and purification using stepwise and convergent SPPS respectively. In addition, chapter 2 describes the results of conformational and biological activity analysis of SCD compared with the recombinant counterpart. Finally, chapter 4 presents a new strategy, devised and developed at Edinburgh, for the chemical ligation of peptide and protein fragments, with the results of the preliminary studies.
1.12 References


1. Solid Phase Peptide Synthesis

47. Nutt R.F., Brady S.F., Darke P.L., Ciccarone T.M., Cotton C.D., Nutt E.M.,
    593.
    776-779.
Chapter Two

The Chemical Synthesis of the Catalytic Domain of Stromelysin (SCD)

2.1 Background

Stromelysin (MMP-3) is a member of the matrix metalloproteinase (MMP) family whose collective physiological function is the controlled extracellular matrix degradation during tissue remodelling and wound healing. However, an uncontrolled imbalance in metalloproteinase activity can lead to extensive tissue degradation resulting in the breakdown of cartilage and the eventual impairment of joint function associated with a number of inflammatory diseases. Although the precise role of each protein involved has not been determined, research has shown the concentration and proteolytic activity of stromelysin to be dramatically increased in joints afflicted by osteo and rheumatoid arthritis.

Inhibition of stromelysin and other members of the family has therefore become the focus of considerable therapeutic interest. Clearly the aim is to obtain detailed information on the three-dimensional structure of the protease from which a good understanding of the catalytic mechanism involved can be obtained. In addition, this structural template will be invaluable for the design of specific MMP inhibitors to combat a wide range of tissue degenerative diseases.
2. The Chemical Synthesis of SCD

2.2 Isolation of the Active Enzyme

Stromelysin is secreted from cells as a 57kDa monomeric proenzyme and consists of a signal peptide for secretion (Met'-Ala'), a propeptide for maintaining latency (Tyr'-His'), a catalytic domain (Phe'-Thr') and a C-terminal fragment (Pro'-Cys') believed to be involved in substrate and inhibitor binding (Fig. 2.1).

\[
\begin{align*}
\text{MKSLPILLLL} & \quad \text{CVAVCSAYPL} \\
\text{FVRRKDSGPV} & \quad \text{DGAAARGEDTS} \\
\text{RTFPGIPKWR} & \quad \text{LGLEVTKLD} \\
\text{RLEYEADIM} & \quad \text{SDTLEVMRKP} \\
\text{EQNTKDTTG} & \quad \text{DSAVEKALKY} \\
\text{LSQDDINGIQ} & \quad \text{LAHAYAPGPG} \\
\text{TLRGEILIFK} & \quad \text{NLFLVAHEI} \\
\text{LVFIFKNQNE} & \quad \text{KTHLTYRIVN} \\
\text{YFVEVDKYW} & \quad \text{ISFAVREHG} \\
\text{TGSSOLEDF} & \quad \text{SLYGPPDSP} \\
\end{align*}
\]

Fig. 2.1 Primary sequence of stromelysin (MMP-3)

The protease is activated on removal of the 82 residue propeptide. On activation, the 45kDa form of stromelysin further cleaves itself in an imprecise fashion yielding a set of heterogeneous forms of MMP-3 without loss of activity. Closer analysis, using N-terminal sequencing, highlighted that those lower molecular weight fragments had the same N-terminus (Phe'-Arg-Thr-Phe'...). From this observation it was postulated that a C-terminally truncated MMP-3 may be responsible and sufficient for the proteases activity.

In order to solve the problem of heterogeneity and determine the contribution of the C-terminal domain to the enzymatic activity, a 29kDa truncated form comprising the propeptide (82 residues) and catalytic domain (173 residues) was expressed. Following purification, it was found that on removal of the prosequence, the enzymatic activity of the 20kDa proteinase was comparable to that of the intact protein.
Finally, Ye et al. described the expression and purification of the 20kDa catalytic domain in absence of both the prosequence and C-terminal domain. In the same report, it was shown that not only was the catalytic domain stable, maintaining the homogeneity, but it clearly retained full enzymatic activity with no preactivation necessary. With this substantial evidence indicating active site preservation, the catalytic domain of stromelysin (SCD) was considered an appropriate model for the structure design of inhibitors of the full length enzyme. With a mass of 20kDa, SCD was now a challenging, but manageable size for three-dimensional structure determination. In addition, the use of molecular modelling could help visualise the active site, thus providing insight into the enzyme-inhibitor interactions.

2.3 Structural Analysis of SCD

The NMR structure of SCD has been solved. In addition, a number of reports have described the three-dimensional X-ray crystal structure of SCD complexed with a number of inhibitors or the catalytic domain with the propeptide intact. Fig. 2.2 shows a schematic representation of the inhibited protease, clearly displaying the secondary structure which comprises of a five stranded β-sheet (one antiparallel and four parallel strands) and three α-helices. Furthermore, the catalytic domain contains two tetrahedrally co-ordinated Zn ions (green) and three octahedrally co-ordinated Ca ions (red) that influence the enzymatic activity.

Situated at the top right-hand side of the ribbon drawing is the active site of SCD. Characterised by the groove with a catalytic zinc lying at the centre, structurally this region is of importance for inhibitor design. Illustrated in purple is an example of an inhibitor which is held in position by means of metal chelation. However, in order to enhance enzyme-inhibitor interactions, inhibitor design should also take advantage of the amino acid residues lining the groove. The ultimate aim is to obtain irreversible binding, which could be achieved by increasing the hydrogen bonding and hydrophobic interactions, in order to hold the molecule strongly in position.
2.4 Inhibition of SCD

Based on the three-dimensional structural information centred on the catalytic site, various substrates have been developed. This has been accompanied by a number of reviews, highlighting the advances in MMP inhibitor design.\textsuperscript{18,19,20} These studies have demonstrated that the most selective and effective substrates have been found to be small peptide derivatives.

Presenting the detailed crystal structure of the active site complexed with a N-carboxyalkyl peptide inhibitor\textsuperscript{21} (23), Becker \textit{et al} demonstrated why peptide derivatives displayed suitable properties as substrates.\textsuperscript{16} This was illustrated by comparing the structure of the active site complexed with the inhibitor to the active site of stromelysin in the latent, preactivated form. In the report, the three-dimensional structure of two forms of stromelysin were described\textsuperscript{16}: the catalytic domain with the 82 residue prosequence intact and the complex of SCD with (23) (Fig. 2.3).
Overall, the structure of SCD is not affected with the incorporation of the prosequence. The prodomain exists as a separate folding unit containing three α-helices with the peptide chain extending into the active site preventing the enzyme from functioning (Fig. 2.3B).

Present in the prosequence is a cysteine residue (Cys$^{92}$) which functions as the fourth ligand of the catalytic zinc: the other three ligands comprising the side-chains of His$^{218}$, His$^{222}$ and His$^{228}$. This cysteine is critical for maintaining latency, therefore protease activation can be achieved by disrupting the Cys-zinc bond. \textit{In vitro} SCD
activation has been found to proceed on treatment with a number of reagents such as proteases, organomercurials, oxidants and detergents. One of the most common reagents employed is aminophenylmercuric acetate (AMPA). In a mechanism described as the 'cysteine switch', mercurials activate the protease by complexing with the thiol releasing cysteine from the zinc, thereby exposing the active site.

Comparison of the alignment of the prosequence in the active site with the SCD-inhibitor complex highlights a number of structural and conformational similarities (Fig. 2.4). In both the proenzyme and inhibited protein, the groove is occupied by an extended peptide chain which interacts with the enzyme via several β-structure-like

![Fig. 2.4 A: Active site interactions with inhibitor. B: Active site of proenzyme. C and D: Specific residues involved in enzyme-inhibitor interactions. Dashed lines indicate hydrogen bonds between the substrate and enzyme.](image-url)
fourth ligand for the catalytic zinc ion. Closer analysis of the hydrogen bonds formed (Fig. 2.4 C and D) reveals that both the propeptide and inhibitor interact with the same residues in the catalytic core with the exception of Pro$^{238}$.

With the interactions of the inhibitor backbone considered, finally the attention is centred on the involvement of the side-chain groups P$^1$ and P$^2$ (Fig. 2.4C). Crystal structure analysis suggests favourable hydrophobic interactions with a number of residues$^{14,15,16}$. P$^1$ is surrounded by Leu$^{214}$, Val$^{215}$, Leu$^{235}$, Tyr$^{237}$, Pro$^{238}$ and Tyr$^{240}$, while P$^2$ is in close contact with residues such as Asn$^{179}$, Val$^{180}$ and Leu$^{181}$. Furthermore, the hydrophobic pocket in which P$^1$ lies is extensive in volume, spanning into the hydrophobic core$^{16}$. Considering the phenylalanine group (P$^1$) barely half fills this hydrophobic region, greater exploitation of this pocket may lead to the design of more potent and more specific inhibitors of SCD.

The availability of SCD has made it possible to obtain three-dimensional structures of the enzyme through NMR and X-ray analysis. As a result of detailed active site analysis, inhibitors with excellent in vitro potencies have been reported. Unfortunately, the majority of inhibitors developed share a common drawback; they are derived from peptides and therefore have limited oral and bioavailability. However, building from the extensive knowledge already obtained centred on the enzyme-inhibitor interactions, it is anticipated the near future developments will produce a new generation of inhibitors with excellent therapeutic application.

2.5 Research Overview

The previous sections have highlighted why SCD is of such great therapeutic interest. Considering the active enzyme has been isolated as a 20kDa stable and homogeneous protein, SCD appeared a viable target for chemical synthesis with respect to both structural and functional interests.
There are a number of reasons why SCD is viewed as a suitable target for stepwise SPPS. With recombinant SCD available, comparison of the synthetic protein with the recombinant counterpart would be possible throughout purification. In addition, a number of simple \textit{in vitro} experiments are available to assess enzymatic activity.\textsuperscript{23,24} Such simple assays are invaluable for protein purity and structure evaluation.

The initial aim of this study was to show that SCD could be prepared using SPPS. Subsequently, efforts concentrated on the development of a purification protocol for the isolation of homogeneous protein. Clearly the research objective was to show that the synthetic protein had comparable biophysical properties to its recombinant counterpart. Therefore studies comparing the structure and proteolytic activity of recombinant and synthetic SCD were performed. The eventual aim was to prepare analogues of SCD by chemical synthesis. Variation of single or multiple residues centred at the active site would be of significant interest, aiding inhibitor design and extending investigations into the mechanism of the enzyme action.

\subsection{2.6 The Chemical Synthesis of the Catalytic Domain of Stromelysin}

Illustrated by the successful preparation of human erthropoietin (166 residues),\textsuperscript{25} interferon-\(\gamma\) (143 residues)\textsuperscript{26} and leptin (146 residues),\textsuperscript{27} research at Edinburgh has demonstrated how by using stepwise chemical methods, the synthesis of proteins >150 amino acid residues is attainable. In each of these syntheses, the coupling reagent employed was HOCt (section 1.7 (22)), designed and developed at Edinburgh.\textsuperscript{28} Previous research has presented examples where HOCt has proven superior in coupling abilities to HOBt and as a result, the stepwise preparation of large polypeptides and proteins has become more feasible.

The primary sequence of SCD (Fig. 2.5) was prepared from the constituent amino acids using the aforementioned stepwise SPPS. Beginning with Fmoc-Thr functionalised Wang resin (0.15mmol/g), the synthesis involved the coupling of 173
Phe-Arg-Thr-Phe-Pro-Gly-Ile-Pro-Lys-Trp^{10}-Arg-Lys-Thr-His-Leu-Thr-Tyr-Arg-Ile-Val^{20}-Asn-Tyr-Thr-Pro-Asp-Leu-Pro-Lys-Asp-Ala^{30}-Val-Asp-Ser-Ala-Val-Glu-Lys-Ala-Leu-Lys^{40}-Val-Trp-Glu-Glu-Val-Thr-Pro-Leu-Thr-Phe^{50}-Ser-Arg-Leu-Tyr-Glu-Gly-Glu-Ala-Asp-Ile^{60}-Met-Ile-Ser-Phe-Ala-Val-Arg-Glu-His-Gly^{70}-Asp-Phe-Tyr-Pro-Phe-Asp-Gly-Pro-Gly-Asn^{80}-Val-Leu-Ala-His-Ala-Tyr-Ala-Pro-Gly-Pro^{90}-Gly-Ile-Asn-Gly-Asp-Ala-His-Phe-Asp-Asp^{100}-Asp-Glu-Gln-Trp-Thr-Lys-Asp-Thr-Thr-Gly^{110}-Thr-Asn-Leu-Phe-Leu-Val-Ala-Ala-His-Glu^{120}-Ile-Gly-His-Ser-Leu-Gly-Leu-Phe-His-Ser^{130}-Ala-Asn-Thr-Glu-Ala-Leu-Met-Tyr-Pro-Leu^{140}-Tyr-His-Ser-Leu-Thr-Asp-Leu-Thr-Arg-Phe^{150}-Arg-Leu-Ser-Gln-Asp-Asp-Ile-Asn-Gly-Ile^{160}-Gln-Ser-Leu-Tyr-Gly-Pro-Pro-Pro-Asp-Ser^{170}-Pro-Glu-Thr

**Fig. 2.5 Primary sequence of SCD**

amino acids via their HOCt activated esters and took approximately 18 days for completion. The progress of every coupling was monitored throughout using the Fmoc-UV method (section 1.4) enabling areas of unsatisfactory coupling to be identified. On completion of the synthesis, a reaction profile was prepared (Fig. 2.6). Overall the synthesis proceeded very well, with an estimated final percentage coupling of 71% (based on the initial functionality) and no significant drops in the coupling efficiency. The region between ~Gln^{102} and ~Lys^{40} where the percentage coupling rose to almost 200% has been previously observed in the synthesis of other peptide and proteins and is rationalised as a resin swelling effect.

![Deprotection profile](image)

**Fig. 2.6 Synthesis of SCD: N° Deprotection profile**

On completion of the synthesis, the ratios of each amino acid on the resin bound sample were assessed. The results obtained from amino acid analysis overall were in good agreement with the expected composition (Table 2.1).
Prior to purification, the protein was cleaved from the solid support and the side-chain protection removed by treating the resin with TFA in the presence of a number of scavengers (present to ‘mop up’ Bu and trityl carbocations). It is essential that the protein’s exposure to the acid conditions is minimal, therefore the optimum cleavage time was established by periodically removing samples from the TFA mixture. HPLC details showed no further change in the profile after 3-4 hours suggesting complete side-chain deprotection (Fig. 2.7).

**Table 2.1 Amino acid analysis of resin bound SCD**

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Calculated Value</th>
<th>Expected Value</th>
<th>Amino Acid</th>
<th>Calculated Value</th>
<th>Expected Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>13.54</td>
<td>13</td>
<td>Lys</td>
<td>4.21</td>
<td>6</td>
</tr>
<tr>
<td>Arg</td>
<td>5.3</td>
<td>7</td>
<td>Met</td>
<td>2.41</td>
<td>2</td>
</tr>
<tr>
<td>Asx</td>
<td>18.86</td>
<td>21</td>
<td>Phe</td>
<td>9.65</td>
<td>10</td>
</tr>
<tr>
<td>Glx</td>
<td>13.26</td>
<td>13</td>
<td>Pro</td>
<td>12.94</td>
<td>14</td>
</tr>
<tr>
<td>Gly</td>
<td>13.88</td>
<td>13</td>
<td>Ser</td>
<td>6.14</td>
<td>9</td>
</tr>
<tr>
<td>His</td>
<td>8.22</td>
<td>8</td>
<td>Thr</td>
<td>8.73</td>
<td>14</td>
</tr>
<tr>
<td>Ile</td>
<td>7.48</td>
<td>8</td>
<td>Tyr</td>
<td>6.42</td>
<td>8</td>
</tr>
<tr>
<td>Leu</td>
<td>14.18</td>
<td>16</td>
<td>Val</td>
<td>7.15</td>
<td>8</td>
</tr>
</tbody>
</table>

Fig. 2.7 Progress of SCD cleavage and deprotection
2.7 Purification and Characterisation of SCD

2.7.1 Gel filtration chromatography
Following chemical synthesis and TFA cleavage, the crude protein sample would undoubtedly contain a number of lower molecular weight impurities such as truncated peptide sequences, residual TFA and scavengers. The majority of these were removed using size exclusion chromatography where the rate of elution is dependent on molecular size. Crude SCD was loaded onto a Superdex™ 75 column, pre-equilibrated with 8M urea, 50mM Tris pH 7.5. All fractions absorbing at 280nm were assessed for protein content. HPLC analysis confirmed the initial fractions to elute contained protein of interest (shaded area Fig. 2.8 A). Compared with the crude SCD, the HPLC profile was now found to be considerably cleaner (Fig. 2.8 B and C).

![Graphs showing gel filtration chromatography and HPLC profiles](image)

**Fig. 2.8** A: FPLC profile of material eluting from Superdex™ 75. B: HPLC profile of crude SCD. C: HPLC profile of SCD after size exclusion.
2.7.2 Characterisation of SCD. 1. Molecular weight determination

Following size exclusion, the molecular weight of the synthetic protein was assessed. Molecular weight standards (ovalbumin Mwt. 43,000, chymotrypsin Mwt. 25,000 and ribonuclease Mwt. 13,700) and recombinant SCD were loaded onto the Superdex™ column and their individual elution volumes recorded. A graph was plotted displaying the logarithm of the molecular weight against the elution volume parameter ($k_{av}$) for the three standards, where:

\[ k_{av} = \frac{V_e - V_0}{V_t - V_0} \]

\[ V_e = \text{elution volume} \]
\[ V_t = \text{total column bed volume} \]
\[ V_0 = \text{mobile phase volume} \]

From the straight line plot the estimated $k_{av}$ for SCD was between 0.09 and 0.1 which was in good agreement with the experimental value of 0.095. In addition, the elution volumes of the recombinant and synthetic proteins were almost identical ($V_e$ 9.7 and 9.8 respectively). Both pieces of evidence gave a clear indication that the synthetic protein was of expected molecular weight, therefore the protein was analysed further to confirm that SCD had been isolated.

2.7.2.2 MALDI-TOF MS and N-terminal sequencing

Both the pro and activated recombinant SCD have been characterised by MALDI-TOF mass spectroscopy (Fig. 2.9). However, analysis of the synthetic counterpart proved difficult with no signal observed throughout purification.
Fig. 2.9 MALDI-TOF MS of the proenzyme (Mwt. 29,011.13) and activated SCD (Mwt. 19,632.31)

An alternative method for protein characterisation by mass spectrometry is to digest the sample using a proteolytic enzyme followed by examination of the lower molecular weight fragments. With the ability to withstand denaturing conditions (4M urea), and specific cleavage site of Lys and Arg, trypsin was used for the selective breakdown of SCD. Considering the number of sites for digestion (highlighted in Fig. 2.5), not all possible fragments were pinpointed, however fragments from throughout the protein’s sequence were identified (Table 2.2).

<table>
<thead>
<tr>
<th>Fragment</th>
<th>Molecular weight</th>
<th>Observed value</th>
<th>Fragment</th>
<th>Molecular weight</th>
<th>Observed value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thr^{13}-Lys^{37}</td>
<td>2.847.48</td>
<td>~2.820</td>
<td>Val^{61}-Arg^{67}</td>
<td>3.158.59</td>
<td>3.155.02</td>
</tr>
<tr>
<td>Ile^{19}-Lys^{40}</td>
<td>2.405.23</td>
<td>~2.400</td>
<td>Glu^{68}-Lys^{66}</td>
<td>4.256.86</td>
<td>4.254.08</td>
</tr>
<tr>
<td>Ile^{19}-Lys^{106}</td>
<td>9.801.74</td>
<td>~9.800</td>
<td>Asp^{107}-Arg^{151}</td>
<td>5.030.53</td>
<td>~5.050</td>
</tr>
<tr>
<td>Ala^{38}-Arg^{52}</td>
<td>1.775.97</td>
<td>1.870.54</td>
<td>Leu^{152}-Thr^{73}</td>
<td>2.343.10</td>
<td>2.344.15</td>
</tr>
</tbody>
</table>

Table 2.2 Tryptic digest of SCD: Analysis by MALDI-TOF mass spectrometry

From the digest, no fragments corresponding to the initial residues at the N-terminus were identified. However, this was unsurprising considering there are three sites of cleavage within Phe^{1}-Lys^{12} yielding a variety of low molecular weight fragments. N-terminal sequencing however clearly illustrated that the first three residues of the synthetic protein were Phe-Arg-Thr with no apparent secondary sequences present.
2.7.2.3 Isoelectric point (pI) determination

Finally, prior to further purification, the pI value of the synthetic SCD was determined. The isoelectric point (pI) of a protein is defined as the pH where all the individual charges of the protein exactly cancel out and the overall net charge is zero. Considering that SCD contains a large number of acidic amino acids (eg. 14 Asp and 10 Glu) compared with basic groups (eg. 6 Lys and 7 Arg), it is unsurprising that the estimated pI value is between 4 and 5.^{29}

Using the technique of isoelectric focusing, the pI of SCD was examined. The Biorad Rotofor® cell was used, which consists of a focusing chamber, where the protein sample and a selection of ampholites are exposed to an electric field. Ampholytes are aliphatic amino acids with a range of pI values which, on introduction of the charge across the cell, arrange to establish a pH gradient. With a pH gradient set up, the protein then moves within the charged solution until reaching a pH where the overall net charge is zero. The protein is then described as focused at its isoelectric point.

Initially ampholytes with a broad pH (3-10) range were used. After focusing, the pH of each fraction was measured and analysed for protein content (HPLC). The majority of protein was isolated from fractions of pH between 4 and 5 (Fig. 2.10 A). Subsequently, SCD was focused using a narrower ampholite range (pH 4-6). The pH difference between each of the 20 fractions was now much smaller, therefore an

![Fig. 2.10 Isoelectric focusing of SCD. Ampholyte range - A: 3-10. B: 4-6.](image-url)
accurate pI value could be determined. Once focused, from the HPLC analysis the estimated pI of SCD was 4.5, with the majority of protein accumulating in three fractions centred at pH 4.5 (Fig. 2.10B). Considering little material was found in the other fractions, isoelectric focusing also gave a valuable account of the overall purity of the protein after size exclusion.

2.8 Folding and Further Purification of SCD

After removal of the lower molecular weight species, efforts concentrated on the folding of SCD. It was proposed that in suitable conditions correctly sequenced SCD would fold into the defined conformation, meanwhile other incorrectly sequenced protein would remain unfolded. Subsequently, by means of ion exchange or hydrophobic interaction chromatography, isolation of the active protease would be feasible.

Following size exclusion, SCD was in a solution of 8M urea, 50mM Tris pH 7.5. The urea was employed here as a denaturant ensuring protein solubilisation. As a result, at this point SCD existed as the long chain of amino acids randomly fluctuating in solution without defined conformation.

The rate and extent of protein folding in solution is greatly dependent on the conditions used. Therefore, initially efforts centred on folding synthetic SCD using the conditions employed to refold the recombinant protein. For recombinant SCD, enzymatic activity was shown to be recovered when denatured protein was refolded in Tris buffer at pH 7.5, containing both calcium (10mM CaCl₂) and zinc (0.1mM ZnCl₂) ions. Furthermore, folding is normally performed at low protein concentrations (0.1-0.01mg/ml) in order to minimise potential side-reactions such as aggregation.
2.8.1 Folding of SCD in solution
Attempts to fold SCD in solution were examined. SCD in 8M urea solution was added dropwise to a large volume of folding buffer. This satisfied the general requirements for folding; the concentration of urea was at a minimum, and SCD concentration was sufficiently low allowing each molecule to arrange the groups of amino acids with minimal interactions from surrounding molecules.

The progress of folding was monitored using HPLC. It was predicted that on folding the original broad peak on HPLC would sharpen as the protein molecule arranged and folded into the defined conformation. Unfortunately this was not observed. Instead, the HPLC profile reduced in size over time, and, after two days the majority of protein had precipitated out of solution. Even at low protein concentrations, competitive intra- and intermolecular interactions and protein misfolding were apparent, clearly interfering with the folding process.

Folding of SCD in solution was repeated examining the effect of protein concentration and temperature. In addition, an alternative approach, dialysis, where urea is gradually removed over time, was studied. Unfortunately, the recurring problem of aggregation and protein precipitation was encountered. Furthermore, litres of buffer were necessary for the folding of small quantities of protein.

2.8.2 Matrix assisted folding of SCD
A different approach for protein renaturation, where the potential for protein aggregation is minimised has been illustrated by Creighton. This involves fixation of the protein, in its urea-denatured state, onto an ion exchange matrix with subsequent elution using a gradient of folding buffer. Supplemented with a number of examples, Creighton has demonstrated the utility of matrix assisted protein folding for the recovery of sufficient quantities of active protein.

The recovery of non-covalently immobilised synthetic SCD from an anion exchange support (DEAE Sepharose) was examined. First, the protein in the urea solution was
mixed throughout a sample of DEAE matrix. By mixing the protein through a portion of matrix, concentration of protein molecules at the gel surface was avoided, preventing potential intermolecular interactions.

Once loaded, the gel was equilibrated using 8M urea, removing unbound components such as DTT (used for reducing Met) or impurities still present. The gradual removal of the denaturant and introduction of folding buffer proceeded overnight. In this slowly changing environment, it was envisaged that protein precipitation would be less likely. In addition, it was anticipated that removal of the denaturant would lead to the protein molecule arranging into a thermodynamically stable folded conformation (Fig. 2.11, stage 2). Once the urea solution had been completely replaced, the protein was finally eluted on introduction of a salt gradient (NaCl) (Fig. 2.11, stage 4).

![Schematic representation of matrix assisted protein folding](image)

The recovery of protein from the column was encouraging. On loading 32.5mg of protein, an estimated 21.5mg eluted in the salt gradient. Furthermore, protein containing fractions of ~1mg/ml were obtained without precipitation. Subsequent HPLC analysis displayed a peak which appeared to have sharpened compared with SCD in 8M urea (Fig. 2.12) suggesting that the protein had folded to some extent. Further evidence that the protein displayed folding characteristics was presented on studying the HPLC profile of SCD over time. The initial sharp peak was observed to
broaden over several days, indicating that over time there were conformational changes occurring. Clearly further structural assessment was necessary.

![HPLC profile of SCD before and after DEAE purification](image)

Fig. 2.12 HPLC profile of SCD before and after DEAE purification

2.8.3 Characterisation of SCD after ion exchange

Protein which eluted from the ion exchange column was assessed prior to structural analysis. N-Terminal sequencing confirmed that the initial residues were Phe-Arg-Thr with no apparent secondary sequences present. Amino acid analysis was in good agreement with the expected composition, however, considering the size of the molecule and the length of acid hydrolysis, it was unsurprising that overall a more accurate account for some residues was unattainable (Table 2.3).

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Expected Value</th>
<th>40 Hours Hydrolysis</th>
<th>50 Hours Hydrolysis</th>
<th>60 Hours Hydrolysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>13</td>
<td>13.39</td>
<td>13.51</td>
<td>13.15</td>
</tr>
<tr>
<td>Arg</td>
<td>7</td>
<td>5.23</td>
<td>5.05</td>
<td>5.06</td>
</tr>
<tr>
<td>Asx</td>
<td>21</td>
<td>22.01</td>
<td>20.73</td>
<td>21.01</td>
</tr>
<tr>
<td>Glx</td>
<td>13</td>
<td>14.07</td>
<td>14.01</td>
<td>12.42</td>
</tr>
<tr>
<td>Gly</td>
<td>13</td>
<td>/</td>
<td>13.24</td>
<td>13.59</td>
</tr>
<tr>
<td>His</td>
<td>8</td>
<td>9.83</td>
<td>8.76</td>
<td>8.82</td>
</tr>
<tr>
<td>Ile</td>
<td>8</td>
<td>7.85</td>
<td>7.63</td>
<td>7.54</td>
</tr>
<tr>
<td>Leu</td>
<td>16</td>
<td>16.61</td>
<td>15.43</td>
<td>15.84</td>
</tr>
<tr>
<td>Lys</td>
<td>6</td>
<td>5.76</td>
<td>5.68</td>
<td>4.42</td>
</tr>
<tr>
<td>Met</td>
<td>2</td>
<td>2.50</td>
<td>/</td>
<td>1.43</td>
</tr>
<tr>
<td>Phe</td>
<td>10</td>
<td>8.92</td>
<td>8.99</td>
<td>8.97</td>
</tr>
<tr>
<td>Pro</td>
<td>14</td>
<td>13.15</td>
<td>12.58</td>
<td>10.13</td>
</tr>
<tr>
<td>Ser</td>
<td>9</td>
<td>6.77</td>
<td>6.29</td>
<td>5.89</td>
</tr>
<tr>
<td>Thr</td>
<td>14</td>
<td>10.82</td>
<td>9.98</td>
<td>10.04</td>
</tr>
<tr>
<td>Tyr</td>
<td>8</td>
<td>/</td>
<td>6.69</td>
<td>6.15</td>
</tr>
<tr>
<td>Val</td>
<td>8</td>
<td>7.50</td>
<td>7.48</td>
<td>6.99</td>
</tr>
</tbody>
</table>

Table 2.3 Amino acid analysis of SCD following DEAE purification
Finally, samples of SCD were assessed using SDS gel electrophoresis (Fig 2.13). In general, the synthetic protein samples were inclined to streak down the gel, possibly due to incomplete removal of TFA, urea or NaCl. However, overall Fig. 2.13 gave an indication of how the purification was proceeding.

![SDS gel analysis of SCD. Lane 1 - molecular weight standards. 2 - Crude SCD. 3 - SCD after size exclusion. 4 - SCD after ion exchange. 5 - Recombinant SCD.]](image)

**2.9 Enzymatic Activity of SCD. Recombinant and synthetic compared**

**2.9.1 Hydrolysis of Substance P**

To help determine whether SCD had been isolated and was of correctly folded conformation, the protein's biological activity was assessed. If the synthetic protease had folded correctly, it was anticipated that the resulting activity would be comparable to that of the recombinant counterpart.

Substance P (24), a known substrate for the enzyme,\(^23\) is cleaved exclusively at Gln\(^6\)-Phe\(^7\) on incubation with SCD. Using HPLC analysis, both the recombinant and synthetic protein clearly displayed hydrolysing capabilities, with MALDI-TOF MS confirming the identity of the hydrolysed fragments (Fig. 2.14). Unfortunately it was also evident from HPLC analysis over time that the recombinant protein was distinctly more active, however an accurate assessment of the situation was not possible from the non-continuous HPLC assay.
2. The Chemical Synthesis of SCD

2.9.2 Hydrolysis of fluorogenic substrates

SCD is known to cleave at specific sites in various peptide sequences. Building from this, a number of fluorogenic substrates have been designed.\textsuperscript{24,31,32} The advantages over the Substance P assay are two-fold. Firstly, with the incorporation of a highly fluorescent marker, very dilute protein samples can be assessed. In addition, the change in fluorescence can be monitored continuously from initial incubation, therefore the extent and rate of hydrolysis can be calculated.

Fig. 2.14 Activity of SCD: Substance P hydrolysis (Arg\textsuperscript{1}-Met\textsuperscript{11} Mwt. 1,349.67, Arg\textsuperscript{1}-Gln\textsuperscript{6} Mwt. 613.80, Phe\textsuperscript{7}-Met\textsuperscript{11} Mwt. 753.88)

One of the most effective fluorogenic substrates for SCD is the novel coumarin-labelled peptide (25).\textsuperscript{24,31} In addition to the highly fluorescent 7-methoxycoumarin (Mca) and specific site of hydrolysis, this substrate has a 2,4-dinitrophenyl group incorporated that acts as a fluorescent quencher. It is proposed that prior to hydrolysis, the fluorescent energy is greatly dampened due to internal resonance transfer along the peptide backbone to the 2,4 DNP moiety.\textsuperscript{24,32} However, the
The quenching effect may simply be due to the interaction of the two groups through space considering their close proximity. Regardless of the interaction, on hydrolysis the coumarin is released from the quenching effect and a rise in fluorescence is observed, the rate of which is directly related to the activity of the protease.

Following ion exchange purification, a number of samples of synthetic SCD were incubated with the Mca-peptide. Protease activity was confirmed with a clear rise in fluorescence, the extent of which was shown to be concentration dependent. Furthermore, on mixing SCD with 8M urea, no further change in fluorescence was observed. Clearly in suitable conditions, synthetic SCD displayed properties that were comparable with those of the active enzyme.

Examination of the synthetic protease compared with recombinant SCD, however, demonstrated that problems associated with the folding of synthetic SCD were still apparent. Fig. 2.15 highlights the differences, with an estimated 150 fold increase in activity for recombinant SCD compared with the synthetic protein. Further assessment of the protease activity, examining the effect of the protein and substrate concentration underlined that the difference in activity was quite considerable.

One possible explanation for this was that the matrix bound protein was prevented from folding correctly. This protein misfolding could be the result of ionic and hydrophobic interactions between the unfolded or partially folded protein and the ion exchange support. The problem of protein folding was therefore examined further by denaturing and refolding the recombinant and synthetic proteases in solution. It was proposed that, in the absence of matrix interactions, renaturation in solution would allow correct folding of SCD to proceed.
2.9.3 Chemical denaturation and refolding of SCD

Recombinant and synthetic SCD were denatured by overnight incubation in a solution of 8M urea, 50mM Tris. Complete loss of enzymatic activity was confirmed by the Mca-peptide assay. Protein refolding was then performed in solution, adding the denatured SCD dropwise to folding buffer. Although slight precipitation on refolding of synthetic SCD was observed, after concentrating, the calculated loss was negligible. From the Mca-peptide assay, it was evident that proteolytic activity was recoverable for both proteins (Fig. 2.16).

**Fig. 2.15** Proteolytic activity of synthetic and recombinant SCD compared. Change in fluorescence on incubation with 0.1μM Mca-peptide at 37°C.

**Fig. 2.16** Refolding of SCD. Proteolytic activity compared.
These results displayed that enzymatic activity was recoverable after chemical denaturation. However, with only an estimated one third of the original activity, the folding problem for synthetic SCD was not resolved by refolding in solution. Clearly a more detailed study of the overall conformation of synthetic SCD was necessary.

2.10 Conformational Analysis of SCD

2.10.1 Circular Dichroism (CD) studies on SCD

CD is a convenient and widely used method for the quantitative assessment of the secondary structure content of a protein molecule. This technique involves exposing the protein sample to plane polarised light. As an asymmetric molecule, the left and right-handed circularly polarised light will be unequally absorbed, the extent of which depends on the optically active molecule being studied. The transmitted light which emerges from the sample will have rotated through an angle $\theta$ relative to the polarisation of the incident light, providing information on the chirality of the molecule.

The CD spectra for SCD was recorded between 260-190nm. The CD signals in this region are dominated by contribution of the amide bonds and will therefore provide information on the secondary structural characteristics of the protein. Fig. 2.17 shows the CD spectra obtained for two samples of SCD obtained after DEAE purification, compared in each case with the recombinant protein.

The far UV spectra for both the recombinant and synthetic samples appeared to be relatively similar. The secondary structure content in each case was determined using the CONTIN procedure, devised by Provencher and Glockner$^{33}$ (Table 2.4).

<table>
<thead>
<tr>
<th></th>
<th>$\alpha$-helix</th>
<th>$\beta$-sheet</th>
<th>Remainder</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recombinant ProSCD</td>
<td>21±0.53%</td>
<td>28±0.71%</td>
<td>51±1.0%</td>
</tr>
<tr>
<td>Recombinant ActSCD</td>
<td>8±0.47%</td>
<td>37±0.63%</td>
<td>54±0.9%</td>
</tr>
<tr>
<td>Synthetic SCD Fract. 1</td>
<td>6±0.92%</td>
<td>28±1.6%</td>
<td>65±2.3%</td>
</tr>
<tr>
<td>Synthetic SCD Fract. 2</td>
<td>22±0.57%</td>
<td>30±0.97%</td>
<td>49±1.4%</td>
</tr>
</tbody>
</table>

*Table 2.4 Calculated secondary structure content from CD results.*
The results presented two main points of interest. Firstly, considering the X-ray crystal structure for recombinant SCD displayed three large α-helices (Fig 2.2), it was surprising that the estimated α-helical content was, on average, less than 10%. This underestimated calculation was proposed to be due to the buffers used. Both Tris and Cl ions are known to absorb light in the wavelength range in which the experiment was carried out. The signal-to-noise ratio below 200nm was so strong that it was only possible to examine data from 260 to ~210-200nm. Unfortunately the region below 200nm is of importance to allow full calculation of the α-helical content. However it was still concluded that CD could provide a good representation of the conformational changes. For example, considering the prodomain of SCD is predominately α-helical in structure (Fig. 2.3B), the calculated value was clearly an underestimate (Table 2.4). However, on comparison with active SCD, the calculated decrease in α-helical content on activation helped illustrate the structural changes.
The second point of interest was the difference in the secondary structure content for the two synthetic SCD samples. Both fractions, isolated and analysed following DEAE purification, displayed comparable enzymatic activity. On analysis of the CD spectra however, the α-helical content was clearly different. Initially it was uncertain why two fractions of similar enzymatic activity displayed such structural diversity. However, proceeding conformational studies presented some interesting results providing further information on the overall structure of the synthetic SCD.

2.10.2 Conformational changes on chemical denaturation. CD analysis

CD is an extremely valuable tool for monitoring structural variations during progressive denaturation, thus changes in secondary structure became more apparent on further analysis by CD. On incubation of SCD with increasing concentrations of denaturant, the CD of each sample was recorded and compared (Fig. 2.18).

![Fig. 2.18 The effect of Gdm.HCl (0M black, 1M red, 2M blue, 4M green) on the CD signal of recombinant (top) and synthetic (bottom) SCD.](image-url)
It was envisaged that on introduction of a denaturant, the overall secondary structure content would decrease. With synthetic SCD, addition of Gdm.HCl, even at low concentrations, reduced the magnitude of the CD signal indicating that protein denaturation was occurring (Note - On mixing with 1M Gdm.HCl, a white precipitate was formed. The resulting signal for synthetic SCD will therefore be misleading).

The recombinant counterpart however displayed different structural characteristics. Surprisingly the initial CD signal, in the absence of Gdm.HCl, was clearly much smaller than that obtained for the synthetic protein under identical conditions. In addition, at low concentrations of denaturant (1M, 2M Gdm.HCl), the CD signal increased, indicating secondary structure enhancement. Finally, in the stronger denaturing conditions, the CD signal for recombinant was comparable to the synthetic protein and it was clear that, overall, the secondary structure of SCD had been disrupted.

It was concluded that although the synthetic SCD had clearly arranged to display secondary structural characteristics, the molecule had not folded completely to form the hydrophobic core with a defined tertiary structure. Evidence for this was based on a number of observations made from the CD spectra. The profile for recombinant SCD suggested that initial introduction of the denaturant opened up the hydrophobic core, exposing the centre of the molecule therefore a clearer estimation of the α-helical content could be made. This could explain why at 2M Gdm.HCl the intensity of the CD signal was similar to the signal for the synthetic protein at 0M Gdm.HCl. In the absence of these tertiary interactions and stabilising hydrophobic core however, the secondary structure of the synthetic SCD was clearly disrupted, even in low concentrations of Gdm.HCl. Furthermore, with CD analysis presenting two differing accounts of the α-helical content (Fig. 2.17), overall the three-dimensional structure of the synthetic molecule was concluded to be undefinable and unstable with obvious structural deviations. Further evidence that synthetic SCD had not completely folded was provided by the tryptophan emission spectra.
2.10.3 Conformational changes on chemical denaturation. Fluorescence analysis

The fluorescence of proteins originates from the three aromatic amino acids which can be used as spectroscopic probes to investigate conformational changes. Fluorescence is usually dominated by the contribution of the tryptophan residues which display the strongest absorbing and emitting properties. With the crystal structure of SCD illustrating that all three tryptophan residues are buried in the hydrophobic core (Fig. 2.19), it was predicted that on denaturing SCD, there would be an obvious change in the fluorescence spectra. On incubation with various concentrations of denaturant, the fluorescence spectra for recombinant and synthetic SCD were recorded and compared (Fig. 2.20).

![Fig. 2.19 Schematic representation of the carbon backbone of SCD (white) with the proenzyme intact (red). The three tryptophan side-chains (W\textsuperscript{82}, W\textsuperscript{124}, W\textsuperscript{186}) are shown in yellow.](image)

The results of the fluorescence experiment provided additional evidence that there was a problem with the folding of synthetic SCD. The fluorescence spectra obtained for recombinant SCD in the absence of denaturant displayed how, in a hydrophobic environment, the tryptophan emissions occur at shorter wavelengths. On exposure to solvent, the intensity of fluorescence was increasingly quenched, accompanied with a shift in wavelength from 328 to 354nm, which corresponds to the fluorescence of free tryptophan.
Fig. 2.20 Tryptophan emission spectra for SCD incubated with GuHCl. Top: Recombinant SCD - 0M red ($\lambda_{max}$ 328nm), 1M purple (327.5nm), 2M black (340nm), 4M green (354nm). Bottom: Synthetic SCD - 0M red ($\lambda_{max}$ 343nm), 1M black (346nm), 2M green (353.5nm), 4M purple (355.5nm).

The fluorescence spectra for synthetic SCD also showed an increase in the emission maximum (343-355nm), indicating that the environment of the tryptophan residues changed on denaturing. However, neither the wavelength nor intensity of the emission maxima of SCD without Gdm.HCl were altered to the same extent as the recombinant protein, confirming that the hydrophobic core was not fully established.
2. The Chemical Synthesis of SCD

2.11 Final Investigation into the Activity of Synthetic SCD

Evidence from the CD and fluorescence spectra clearly highlighted that synthetic SCD had not folded completely. As a result, the origin of the proteolytic activity of the synthetic protein was uncertain and required further assessment. Two proposed explanations were examined -

- a small proportion of synthetic SCD had folded completely, but was largely masked in spectroscopic examination by the partially folded material.
- although folding was incomplete, the active site was sufficiently formed to display enzymatic characteristics.

2.11.1 Chromatographic analysis of active SCD. Comparison of recombinant and synthetic protein

If a small proportion of correctly folded SCD was wholly responsible for enzymatic activity, it was expected that further chromatographic separation could isolate the active protein from the mixture.

Hydrophobic interaction chromatography (HIC) is based upon interactions between solvent accessible non-polar groups in the protein molecule and the hydrophobic ligands attached to the uncharged gel matrix. It was predicted that HIC would be sensitive enough to differentiate between the non-polar groups buried within the fully formed hydrophobic core of SCD and hydrophobic regions present on the protein surface due to incomplete folding.

On loading recombinant SCD onto a Phenyl Sepharose column (Resource Phe™) in a buffer of high ionic strength (50mM Tris pH 7.5, 10mM CaCl₂, 0.1mM ZnCl₂, 0.5M (NH₄)₂SO₄), the protein was found to elute, with an almost quantitative recovery, over a small volume during progressive removal of (NH₄)₂SO₄. However, when synthetic SCD was loaded using identical conditions, no material absorbing at 280nm was observed to elute. It was evident that with synthetic SCD, the majority of
2. The Chemical Synthesis of SCD

Hydrophobic regions were still largely exposed to solvent, binding strongly to the hydrophobic support.

Ion exchange chromatography was employed to provide further evidence that the synthetic protein had not folded. Once more, when loaded onto the ion exchange support (Mono Q™), recombinant SCD was found to elute almost quantitatively over a small volume. In addition, the recombinant protein was found to bind only very weakly to the matrix. On loading synthetic SCD using identical conditions, no protein was observed to elute in these low salt conditions. At higher salt concentrations, good protein recovery was obtained. Furthermore, fractions collected from this region all displayed enzymatic activity.

Both methods of chromatography confirmed the theory that synthetic SCD was not completely folded, with hydrophobic and ionic regions still largely exposed. This would explain the significant increase in the matrix binding properties for both of the chromatographic techniques. In addition, as no synthetic protein was found to co-elute with the recombinant SCD, there was no evidence that the proteolytic activity was a result of completely folded SCD.

2.11.2 Enzymatic activity of chemically denatured recombinant SCD

Following the extensive conformational analysis of synthetic SCD, the question of the origin of activity still remained unanswered. CD and fluorescence analysis of SCD during chemical denaturation demonstrated how, although the hydrophobic core was disrupted, recombinant SCD maintained the secondary structure at low concentrations of Gdm.HCl. It was therefore proposed that if these chemically denatured intermediates of recombinant SCD displayed activity, it could be concluded that the active site was still sufficiently formed in the partially unfolded molecule.

Incubated with increasing quantities of urea, the activity of recombinant SCD was assessed using the Mca-peptide fluorescence assay (Fig. 2.21). From the fluorescence spectra it was clear that the enzymatic activity decreased with increasing
concentration of denaturant as expected. Furthermore, it was evident that the active site of SCD was particularly robust, with the recombinant protein still clearly displaying enzymatic activity at concentrations where the tertiary structure would undoubtedly be disrupted. This could be attributed to the presence of the catalytic zinc at the enzymatic site which may help to maintain the structure of the groove by metal chelation. This finding substantiated the proposal that, although incompletely folded, the active site of synthetic SCD had arranged sufficiently to display the characteristic catalytic activity.

![Fig. 2.21 Proteolytic activity of recombinant SCD in urea. Change in fluorescence on incubation with 1μM Mca-peptide at 37°C.](image)

2.12 Folding of SCD. Conclusions and recommendations

This study has illustrated how the present methodology for the solid phase synthesis of peptides can now be extended, allowing the stepwise preparation of large polypeptides and proteins. The catalytic domain of stromelysin (SCD) has been prepared using chemical methods and, following chromatographic purification, under
suitable conditions proteolytic activity was clearly observed. Comparison studies with the recombinant protein has, however, highlighted the fundamental problems associated with the folding of the synthetic protein.

The question of how the amino acid sequence of a protein determines the three-dimensional structure and folding pathway is still the centre of extensive research efforts. Although a detailed account of the protein folding pathway have not been deduced, it has been established that folding occurs through a number of intermediate, non-random conformations, defined as the pre-folded state. Although unstable compared with the completely folded protein, this pre-folded state is envisaged to contain elements of conformation similar to these in the native structure.

It is clear from the enzymatic activity, CD and fluorescence spectra that the underlying problem with synthetic SCD is that the protein has not folded completely. It appears that the synthetic molecule has not fully established the hydrophobic core which is considered to be the most critical aspect for stability of the normal folded state. Considering the synthetic protein displays proteolytic activity and has comparable secondary structure to the recombinant counterpart, it could be concluded that synthetic SCD has arranged to form intermediates in the folding pathway.

One explanation for this folding problem could be that from their denatured forms, the two proteins adopt specific conformations, via different folding pathways. For example, on incubating recombinant SCD in strong denaturing conditions, it is uncertain whether the protein is in a fully denatured state. Conformations may still be present which, on removal of denaturant, initiate the folding process, following a definite pathway. Without these previous folding properties, the folding process for synthetic SCD may be inherently more random. Although the synthetic molecule arranges to display secondary structural characteristics, in the absence of a non-random folding pathway, the assembly of the final compact tertiary structure may prove unattainable.
Clearly a more detailed study of the folding pathway of SCD is necessary to help identify the underlining differences in the folding of the recombinant and synthetic proteins. With the proposal that the folding pathway for synthetic SCD may be highly random, it may prove beneficial to introduce agents into the folding mixture. In some cases, such additives promote the adoption of a protein's native structure by directing the initial conformations, and, on a whole, reduce incorrect folding which can result in aggregation and precipitation. For example, molecular chaperones such as GroEL have been invaluable in assisting the assembly of a number of proteins.\textsuperscript{35-37} Binding only to unfolded proteins, the conformation of GroEL-bound protein resembles that of the 'molten-globule state', an intermediate in the folding pathway where the hydrophobic groups are buried in the centre of the compact molecule. Considering the problem with synthetic SCD appears to be with the incomplete formation of the hydrophobic core, the introduction of such agents may prove invaluable, facilitating the correct folding of the synthetic protein.
2.13 References

2. The Chemical Synthesis of SCD


Chapter Three

The Convergent Synthesis of the Catalytic Domain of Stromelysin (SCD)

3.1 Introduction

It is evident from the previous chapter that the highly optimized SPPS can now be successfully applied to the assembly of proteins over 150 residues in length. However, in the case of the catalytic domain of stromelysin (SCD), the problem of protein folding still remains unresolved, with further investigation necessary to pinpoint the source of the protein's inability to fold completely.

One possible explanation could be that the chemically synthesised SCD, although purified to yield enzymatically active protein, may still be heterogeneous in nature, containing single or combination amino acid deletions (section 1.10). It is clear from the Fmoc deprotection profile and characterisation of chemically synthesised SCD that, overall, the protein molecule was of expected composition. However, it is not possible to completely rule out the presence of even a small proportion of amino acid deficient sequences which could hinder the protein folding. Unfortunately, due to their almost identical chromatographic behavior, these closely related impurities, if present, could remain undetected, therefore complete separation of the correctly sequenced SCD may prove impossible.

Considering the problems encountered with the conventional stepwise protein synthesis, the convergent approach (section 1.11) appears to be an attractive
alternative which avoids the problem of single amino acid deficiencies. The original and presently still the most flexible and extensively studied method for peptide coupling is the fragment condensation approach (Fig. 3.1). Here the target sequence is subdivided into a number of peptide segments which are individually synthesised using stepwise SPPS. The target peptide is then prepared by building up the peptide chain from smaller fragments using conditions analogous to single amino acid coupling.

![Fig. 3.1 The principle of peptide fragment condensation](image)

In theory this should improve the quality of the final product. Prior to peptide assembly, the purity of each fragment can be checked at this intermediate stage and purified if necessary. Furthermore, on building of the target sequence, incomplete coupling results in impurity peptides that should be sufficiently different from each other in physical properties, enabling chromatographic separation.

In summary, the protecting groups, coupling reagents and general strategies for fragment condensation are the same as those used for stepwise peptide synthesis. The necessary orthogonal protecting group strategy can be provided by both Fmoc and Boc methodologies, thus the range of available reagents and linkers is extensive, resulting in a highly flexible and adaptable convergent approach. Furthermore, as
3. The Convergent Synthesis of SCD

outlined in Fig. 3.1, fragment condensation can proceed with either both fragments in solution or with the growing fragment chain still attached to the solid support. As with stepwise SPPS, the solid phase fragment condensation (SPFC) is highly beneficial for peptide assembly, permitting the use of an excess of the activated peptide which is easily removed by simple filtration.

This review shall focus only on the SPFC of peptide segments using the Fmoc methodology which is of relevance to the research described in this chapter. For a much broader overview, describing a wide range of reagents, strategies and literature examples of peptides and proteins prepared using both solution and solid phase fragment condensation, reviews by Benz, Lloyd-Williams, and Sakakibara are recommended.

3.2 Solid Phase Fragment Condensation (SPFC)

This section deals with a number of strategic considerations which are of importance in SPFC. These are:

- the solid phase synthesis of fully protected peptides.
- division of the target sequence preventing racemisation.
- preparation of suitable coupling conditions which avoid peptide insolubility.

3.2.1 Preparation of protected peptide fragments

To ensure unambiguous fragment coupling in the absence of any side-reactions, it is essential that the side-chain functionalities are blocked to avoid interference. Therefore, unlike stepwise SPPS, on detachment from the solid support the side-chain protection on the fragment must be inherently stable in the resin cleavage conditions.

Section 1.6 describes the development of a number of linker-resins (Fig. 3.2) which allow extremely mild and selective acidolytic cleavage of the peptide-linker bond. Some of the most commonly used linkers which have been successfully employed in
the preparation of fully protected peptide derivatives include the SASRIN, 2-chlorotritylchloride and Rink linkers (Fig. 1.11). All three are highly acid labile and are therefore compatible with the Fmoc/Bu' strategy which avoids acid treatment throughout the peptide synthesis. Furthermore, during mild acid resin cleavage (e.g. 1% TFA, 10% AcOH), all of the side-chain protecting groups (Bu', Trityl, Pmc etc.) are retained.

![Diagram of peptide synthesis](image)

**Fig. 3.2 Synthesis of fully protected peptides**

### 3.2.2 Choice of peptide fragment

The most important consideration in SPFC is dissection of the target sequence into suitable fragments. In summary, the requirements are -

- synthesis and purification of the protected fragment must be feasible.
- fragment assembly onto the growing peptide chain should proceed rapidly and efficiently.
- racemisation of the C-terminal of each peptide must be suppressed.

This final point is of great importance and is normally the determining factor for fragment choice. In the absence of the N'-urethane protection, racemisation of the C-terminal amino acid on activation (via an oxazolone intermediate, section 1.7.1) is possible. This can be circumvented if the fragment contains a C-terminal glycine or proline. Unfortunately, depending on the target sequence, this may not always be possible without preparing fragments containing a large number of residues.
There are numerous literature examples of successful fragment condensation strategies where C-terminal residues other than Gly or Pro are used.\textsuperscript{1,2} The coupling efficiency and extent of racemisation has been shown to vary greatly depending on the C-terminal residue and the coupling reagents and conditions used.\textsuperscript{4-7} Generally, low temperature activation of fragments containing small, non-functionalised C-terminal residues (e.g. Ala, Leu and Ser) are considered to be the best approach. Furthermore, it has been demonstrated that the addition of such coupling reagents as 1-hydroxybenzotriazole (HOBt), originally developed to suppress racemisation,\textsuperscript{8,9} or 3,4-dihydroxy-3-hydroxy-4-oxo-1,2,3-benzotriazine (HOOBt), minimise the loss of chiral integrity during activation.\textsuperscript{3-7}

### 3.2.3 Insolubility of protected peptides

The other major obstacle in SPFC is that the protected fragments are often only sparingly soluble in the organic solvents used for coupling. It is important when coupling onto the solid phase that the solvent swells the microporous support, exposing all of the reactive sites. However, depending on the protected fragment sequence, solubility in solvents such as DMF, DCM and NMP can be low, leading to unsatisfactory coupling. It is now generally accepted that the increase in fragment insolubility occurs not only as a consequence of the bulky hydrophobic protecting groups but also as a direct result of the transition from a conformationally unordered solvated species to an ordered inter- and intra-molecular hydrogen-bonded species with a β-sheet like structure. To dissolve such peptides, the β-structure must be disturbed. Narita \textit{et al} investigated the influence of various solvents on hydrogen-bonding interaction.\textsuperscript{10,11} It was reported that low concentrations of fluorinated alcohols eg, hexafluoro-2-propanol (HFIP) and trifluoroethanol (TFE) have a strong potential for disrupting β-sheet aggregation, with 10-20% HFIP or TFE sufficient to enhance fragment solubility and improve the coupling efficiency.\textsuperscript{4,5}
3. The Convergent Synthesis of SCD

3.3 Literature Examples of Fragment Condensation

Accompanied with the growing understanding of suitable strategies and coupling procedures, the number of reports describing the successful preparation of larger polypeptide and protein chains have now been published. Table 3.1 summarises a selection of these syntheses which have employed either solid or solution phase fragment condensation.

### A. Solid phase fragment coupling

<table>
<thead>
<tr>
<th>Target Protein</th>
<th>No. of residues</th>
<th>No. of fragments</th>
<th>C-terminal residue</th>
<th>Coupling conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Androctonus australis hector Toxin II</td>
<td>64</td>
<td>8</td>
<td>Gly, Pro</td>
<td>DCC/HOBt DMF</td>
</tr>
<tr>
<td>λ-Cro. DNA binding protein</td>
<td>66</td>
<td>6</td>
<td>Gly, Ser</td>
<td>DCC/HOBt DMF</td>
</tr>
<tr>
<td>HIV-1&lt;sub&gt;tm&lt;/sub&gt; tat</td>
<td>72</td>
<td>5</td>
<td>Gly, Lys, Ser</td>
<td>DIC/HOBt DCM</td>
</tr>
<tr>
<td>Prothymosin (ProTα)</td>
<td>109</td>
<td>9</td>
<td>Glu, Gly, Lys</td>
<td>DCC/HOBt DMSO</td>
</tr>
</tbody>
</table>

### B. Solution phase fragment coupling

<table>
<thead>
<tr>
<th>Target Protein</th>
<th>No. of residues</th>
<th>No. of fragments</th>
<th>C-terminal residue</th>
<th>Coupling conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Porcine C5a anaphylatoxin</td>
<td>74</td>
<td>9</td>
<td>Ala, Glu, Gly</td>
<td>WSCI/HOBt DMF or NMP</td>
</tr>
<tr>
<td>Human parathyroid hormone</td>
<td>84</td>
<td>13</td>
<td>Ala, Gln, Glu, Gly, Leu, Pro</td>
<td>WSCI/HOBt DMF or NMP</td>
</tr>
<tr>
<td>Human midkine</td>
<td>121</td>
<td>13</td>
<td>Ala, Gln, Gly, Lys, Pro</td>
<td>WSCI/HOBt TFE/DCM (1:3)</td>
</tr>
<tr>
<td>Angigenin</td>
<td>123</td>
<td>15</td>
<td>Ala, Asn, Glu, Gly, Leu, Pro</td>
<td>DCC/HOBt or HOBt DMF or NMP</td>
</tr>
</tbody>
</table>

*Table 3.1* Outline of strategies used for protein fragment condensation
3.4 Research Overview

The catalytic domain of stromelysin (SCD) was considered a suitable target protein to investigate whether the present methodology for fragment condensation could be extended to prepare proteins over 150 amino acids in length. Having previously demonstrated that the 173 residue protein could be prepared using stepwise chemical synthesis, a direct comparison between the stepwise and convergent syntheses would be of great interest. In addition, considering the present folding problem with synthetic SCD, it would be invaluable to compare the purification and resulting biological activities of the two proteins. It is anticipated that in the absence of any single amino acid deficiencies, purification after convergent synthesis would lead to a homogeneous protein. If however the extent of proteolytic activity was still low it could be concluded that the folding of synthetic SCD was a mechanistic problem and not one of protein purity.

3.4.1 Coupling strategy

It was uncertain whether the preparation of SCD from small fragments starting at the C-terminus would be feasible, however, the synthesis of a 173 residue protein from peptides averaging 10 residues in length would require an extensive research effort which was considered unattainable without any preliminary investigation. Obviously proteins up to \(~\)100 residues can be prepared using SPFC, therefore by starting with a significantly large peptide chain attached to the solid support, the optimum fragment length and coupling conditions at this more advanced stage could be investigated.

Stepwise synthesis of Pro\(^{96}\)-Thr\(^{173}\) As a starting point, the first 84 residues of SCD beginning from the C-terminus (Thr\(^{173}\) to Pro\(^{96}\)) were prepared using stepwise SPPS. This would act as a foundation onto which the remaining 89 residues could be assembled from a variety of protected fragments. Using the coupling reagent HOCt (section 1.7), the stepwise synthesis of peptides \(~\)80 residues in length is generally straightforward. At Edinburgh this has been illustrated with the synthesis of a variety of small proteins including ubiquitin\(^{20}\) (76 residues) and a number of chemokines\(^{21}\).
3. The Convergent Synthesis of SCD

(74-76 residues), all of which have been successfully purified to homogeneity and folded correctly.

**Dissection of the target sequence**  Fig. 3.3 shows the primary sequence of SCD from which a number of potential sites for condensation which completely avoided racemisation were identified. However at the onset of the project, it would be impossible to predict the outcome of any fragment synthesis or coupling, therefore these sites could only act as guidelines.

Phe-Arg-Thr-Phe-Pro-Gly-Ile-Pro-Lys-Trp\(^{10}\)-Arg-Lys-Thr-His-Leu-Thr-Tyr-Arg-Ile-Val\(^{20}\)-Asn-Tyr-Thr-Pro-Asp-Leu-Pro-Lys-Asp-Ala\(^{30}\)-Val-Asp-Ser-Ala-Val-Glu-Lys-Ala-Leu-Lys\(^{60}\)-Val-Trp-Glu-Glu-Val-Thr-Pro-Leu-Thr-Phe\(^{50}\)-Ser-Arg-Leu-Tyr-Glu-Gly-Glu-Ala-Asp-Ile\(^{60}\)-Met-Ile-Ser-Phe-Ala-Val-Arg-Glu-His-Gly\(^{70}\)-Asp-Phe-Tyr-Pro-Phe-Asp-Gly-Pro-Gly-Asn\(^{80}\)-Val-Leu-Ala-His-Ala-Tyr-Ala-Pro-Asp-Gly-Pro\(^{90}\)-Gly-Ile-Asn-Gly-Asp-Ala-His-Phe-Asp-Asp\(^{100}\)-Asp-Glu-Gln-Trp-Thr-Lys-Asp-Thr-Thr-Gly\(^{110}\)-Thr-Asn-Leu-Phe-Leu-Val-Ala-Ala-His-Glu\(^{120}\)-Ile-Gly-His-Ser-Leu-Gly-Leu-Phe-His-Ser\(^{130}\)-Ala-Asn-Thr-Glu-Ala-Leu-Met-Tyr-Pro-Leu\(^{140}\)-Tyr-His-Ser-Leu-Thr-Asp-Leu-Thr-Arg-Phe\(^{150}\)-Arg-Leu-Ser-Gln-Asp-Asp-Asn-Gly-Ile\(^{160}\)-Gln-Ser-Leu-Tyr-Gly-Pro-Pro-Pro-Asp-Ser\(^{170}\)-Pro-Glu-Thr

*Fig. 3.3 Primary sequence of SCD. The first 84 residues prepared using stepwise SPPS are underlined with the potential sites for fragment condensation in bold.*

3.5 Solid Phase Fragment Condensation. Preliminary Investigation.

Before commencing the convergent synthesis of SCD, a number of factors affecting fragment condensation were investigated with the view of developing effective coupling conditions. This section deals with the results of these preliminary studies

3.5.1 Preparation and characterisation of H\(_2\)N-Pro\(^{99}\)-Thr\(^{173}\)-OH (26)

The 84 residue peptide was prepared from the constituent amino acids using stepwise SPPS. Beginning with Fmoc-Thr functionalised Wang resin (0.2mmol/g), the peptide chain was built up, coupling each amino acid via the HOCT activated ester. Prior to the coupling study, the polypeptide was assessed to ensure the synthesis has been
Therefore, on completion of the solid phase assembly, the peptide was deprotected and cleaved from the solid support by stirring the resin in TFA in the presence of a number of scavengers. Following overnight incubation with N-methyl-mercaptoacetamide (MMA) to reduce any oxidised methionine, Pro\textsuperscript{90}-Thr\textsuperscript{173} was isolated from the crude mixture using preparative HPLC.

Purified Pro\textsuperscript{90}-Thr\textsuperscript{173} was readily characterised by MALDI-TOF MS and the calculated amino acid ratios were found to be in good agreement with the expected composition (Fig. 3.4). Furthermore, N-terminal sequencing clearly illustrated that the first five residues of the synthetic peptide were Pro-N/D-Ile-Asn-Gly with no apparent secondary sequences present.

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Calculated value</th>
<th>Expected value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>5.54</td>
<td>5</td>
</tr>
<tr>
<td>Arg</td>
<td>2.01</td>
<td>2</td>
</tr>
<tr>
<td>Asx</td>
<td>11.88</td>
<td>13</td>
</tr>
<tr>
<td>Gly</td>
<td>6.45</td>
<td>7</td>
</tr>
<tr>
<td>Glx</td>
<td>6.99</td>
<td>7</td>
</tr>
<tr>
<td>His</td>
<td>5.36</td>
<td>5</td>
</tr>
<tr>
<td>Ile</td>
<td>3.86</td>
<td>4</td>
</tr>
<tr>
<td>Leu</td>
<td>9.89</td>
<td>10</td>
</tr>
<tr>
<td>Lys</td>
<td>1.12</td>
<td>1</td>
</tr>
<tr>
<td>Met</td>
<td>0.89</td>
<td>1</td>
</tr>
<tr>
<td>Phe</td>
<td>4.15</td>
<td>4</td>
</tr>
<tr>
<td>Pro</td>
<td>6.20</td>
<td>6</td>
</tr>
<tr>
<td>Ser</td>
<td>4.41</td>
<td>6</td>
</tr>
<tr>
<td>Thr</td>
<td>6.31</td>
<td>8</td>
</tr>
<tr>
<td>Tyr</td>
<td>2.89</td>
<td>3</td>
</tr>
<tr>
<td>Val</td>
<td>1.12</td>
<td>1</td>
</tr>
</tbody>
</table>

**Fig. 3.4** Amino acid analysis and MALDI-TOF MS of H\textsubscript{2}N-Pro\textsuperscript{90}-Thr\textsuperscript{173}-OH (26) \textit{(Mwt. 9,454.71)}

3.5.2 Synthesis and characterisation of fully protected FmocGly\textsuperscript{79}-Gly\textsuperscript{89}-OH (27)

Following the successful preparation and analysis of Pro\textsuperscript{90}-Thr\textsuperscript{173}, a small peptide corresponding to residues 79-89 of SCD was prepared. This sequence contained a number of side-chain functionalities which required protection during fragment
condensation to ensure unambiguous amide bond formation, therefore the peptide was assembled on the extremely acid labile 2-chlorotritylchloride resin (Fig. 3.5), coupling each amino acid via the HOCt activated ester. Detachment of the fully protected peptide from the resin was then performed using AcOH/TFE/DCM (1:1:8).

Analysis of the crude peptide by HPLC displayed a single peak which absorbed at 214 and 300nm (Fmoc absorbs at 300nm). Eluting at ~80% acetonitrile, the peptide was clearly hydrophobic indicating that the side-chain protection was intact. Further evidence of this was provided by comparing the HPLC profile of the fully protected peptide with the same peptide treated with TFA (Fig. 3.6A). Finally, fully protected FmocGly<sup>79</sup>-Gly<sup>89</sup>-OH was characterised by FAB MS (Fig. 3.6B).

Fig. 3.5 Preparation of side-chain protected FmocGly<sup>79</sup>-Gly<sup>89</sup>-OH (27)

Fig. 3.6 A. HPLC profile of fully protected and side-chain deprotected FmocGly<sup>79</sup>-Gly<sup>89</sup>-OH (27). B. FAB MS of the protected peptide (Mwt. 1,589.78).
3.5.3 Coupling of FmocGly<sup>79</sup>-Gly<sup>89</sup>-OH to H<sub>2</sub>N-Pro<sup>90</sup>-Thr<sup>173</sup>-resin

After characterising both the 84 and 11 residue peptides, their ability to couple together was investigated. In general, optimum fragment coupling conditions involve an excess of the incoming fragment, dissolved in a minimal quantity of solvent. Therefore, five equivalents of FmocGly<sup>79</sup>-Gly<sup>89</sup>-OH were dissolved in DMF at a concentration of 150mg/ml. Converted to the corresponding HOCl activated ester, the peptide fragment was then mixed with the resin bound Pro<sup>90</sup>-Thr<sup>173</sup>. Throughout coupling, the mixture was sonicated under nitrogen ensuring that there was sufficient solvent to completely swell the resin.

To assess the reaction progress during coupling, samples of resin were removed. After sonication in 20% piperidine, an estimation of the Fmoc content of the resin-bound peptide was made by measuring the absorbance of the Fmoc-piperidine adduct at 302nm (section 1.4.2). Fig. 3.7 summarises the calculated change in the coupling efficiency over time.

<table>
<thead>
<tr>
<th>Coupling conditions</th>
<th>Coupling efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 hour sonication</td>
<td>41%</td>
</tr>
<tr>
<td>3 hours sonication</td>
<td>57%</td>
</tr>
<tr>
<td>overnight stirring</td>
<td>70%</td>
</tr>
<tr>
<td>fresh Gly&lt;sup&gt;79&lt;/sup&gt;-Gly&lt;sup&gt;89&lt;/sup&gt;</td>
<td>79%</td>
</tr>
<tr>
<td>3 hours sonication</td>
<td></td>
</tr>
</tbody>
</table>

*Fig. 3.7 Fragment coupling of FmocGly<sup>79</sup>-Gly<sup>89</sup>-OH to H<sub>2</sub>N-Pro<sup>90</sup>-Thr<sup>173</sup>-resin*

Clearly the incorporation of FmocGly<sup>79</sup>-Gly<sup>89</sup>-OH onto the solid support had been successful therefore further attempts to increase the coupling efficiency were investigated. These included changing the solvent from DMF to DCM, adding TFE to the reaction mixture and using HOBT as an alternative coupling reagent to HOCl. Unfortunately, the overall coupling efficiency could not be improved.
3.5.4 Tbfmoc purification and characterisation of H$_2$N-Gly$^{79}$-Thr$^{173}$-OH (28)

To confirm that the fragment condensation had been successful, efforts concentrated on the isolation and characterisation of the coupling product H$_2$N-Gly$^{79}$-Thr$^{173}$-OH. Affinity purification employing the N$^a$ protecting group tetrabenzo[a,c,g,i]fluorenyl-17-methoxycarbonyl (Tbfmoc) was used in order to separate the product from uncoupled H$_2$N-Pro$^{90}$-Thr$^{173}$-OH.

Designed and developed at Edinburgh, Tbfmoc is incorporated onto the N-terminus of the completed peptide or protein chain and possesses a number of properties which makes this base labile group a valuable tool for peptide and protein purification (Fig. 3.8).

![Fig. 3.8 Structure and characteristic properties of Tbfmoc](image)

Tbfmoc purification has a great potential for fragment condensation having the ability to isolate the product from uncoupled peptide which remains in the crude mixture. Fig. 3.9 presents a schematic representation for the Tbfmoc purification of H$_2$N-Gly$^{79}$-Thr$^{173}$-OH, outlining the main steps involved.

On completion of the coupling, the resin was first sonicated in the presence of acetic anhydride, capping any uncoupled H$_2$H-Pro$^{90}$-Thr$^{173}$-OH still present in the mixture. The N$^a$-Fmoc protection was then removed and replaced with the Tbfmoc which was selectively incorporated onto the free N-terminus of the coupling product only. After deprotection and detachment from the solid support using the standard cleavage
conditions, the coupling mixture, dissolved in 6M Gdm.HCl/i-propanol (1:1), was mixed with PGC onto which the Tbfmoc-peptide was adsorbed. All uncoupled Pro\textsuperscript{99}-Thr\textsuperscript{173} and other impurities remained in solution and were therefore easily removed by simply washing the solid support before H\textsubscript{2}N-Gly\textsuperscript{79}-Thr\textsuperscript{173}-OH was finally released by sonicating the carbon with 10-20% piperidine.

![Diagram](image-url)

_Fig. 3.9 Tbfmoc purification of H\textsubscript{2}N-Gly\textsuperscript{79}-Thr\textsuperscript{173}-OH (28)_
Following affinity chromatography and HPLC purification, \( \text{H}_2\text{N-Gly}^{79}-\text{Thr}^{173}-\text{OH} \) was characterised by MALDI-TOF MS and the amino acid ratios determined (Fig. 3.10). On comparison of the amino acid analysis before, and after coupling (original amino acid ratios that altered on coupling are given in parenthesis, Fig. 3.10) it was evident that the product isolated after purification was of the expected composition for the coupling product and not uncoupled \( \text{Pro}^{90}-\text{Thr}^{173} \).

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Calculated value</th>
<th>Expected value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>7.89</td>
<td>8 (5)</td>
</tr>
<tr>
<td>Arg</td>
<td>2.04</td>
<td>2</td>
</tr>
<tr>
<td>Asx</td>
<td>12.17</td>
<td>14 (13)</td>
</tr>
<tr>
<td>Gly</td>
<td>9.14</td>
<td>9 (7)</td>
</tr>
<tr>
<td>Glx</td>
<td>7.60</td>
<td>7</td>
</tr>
<tr>
<td>His</td>
<td>6.19</td>
<td>6 (5)</td>
</tr>
<tr>
<td>Ile</td>
<td>4.07</td>
<td>4</td>
</tr>
<tr>
<td>Leu</td>
<td>11.14</td>
<td>11 (10)</td>
</tr>
<tr>
<td>Lys</td>
<td>/</td>
<td>1</td>
</tr>
<tr>
<td>Met</td>
<td>0.86</td>
<td>1</td>
</tr>
<tr>
<td>Phe</td>
<td>4.31</td>
<td>4</td>
</tr>
<tr>
<td>Pro</td>
<td>7.23</td>
<td>7 (6)</td>
</tr>
<tr>
<td>Ser</td>
<td>5.83</td>
<td>6</td>
</tr>
<tr>
<td>Thr</td>
<td>7.32</td>
<td>8</td>
</tr>
<tr>
<td>Tyr</td>
<td>3.05</td>
<td>4 (3)</td>
</tr>
<tr>
<td>Val</td>
<td>2.14</td>
<td>2 (1)</td>
</tr>
</tbody>
</table>

**Fig. 3.10** Amino acid analysis and MALDI-TOF MS (TFA salt) of \( \text{H}_2\text{N-Gly}^{79}-\text{Thr}^{173}-\text{OH} \) (28) (Mwt. 10,500.44)

3.5.5 Coupling of FmocGly\(^{70}\)-Gly\(^{89}\)-OH (29) and FmocPhe\(^{50}\)-Gly\(^{89}\)-OH (30)

The previous sections have illustrated how it was possible to couple a fully protected peptide, 11 residues in length, to a 84 residue polypeptide fixed to a solid support. Efforts then concentrated on investigating whether, using the same coupling conditions, it would be possible to couple larger protected fragments.

Two peptides corresponding to the next 20 (Gly\(^{70}\)-Gly\(^{89}\)) and 40 (Phe\(^{50}\)-Gly\(^{89}\)) residues in the SCD primary sequence were synthesised. Beginning with Fmoc-Gly
functionalised 2-chlorotriylchloride resin, both peptide chains were prepared from the corresponding HOCl activated amino acids. On completion of the syntheses, the two fully protected peptides, FmocGly\textsuperscript{70}-Gly\textsuperscript{89}-OH (29) and FmocPhe\textsuperscript{50}-Gly\textsuperscript{89}-OH (30), were detached from the solid support by stirring the resin in AcOH/TFE/DCM (1:1:8) for 30 minutes.

Following characterisation, the coupling abilities of the two peptides were investigated. Using the same conditions as for FmocGly\textsuperscript{70}-Gly\textsuperscript{89}-OH, the fragments were individually sonicated with the peptide-resin. After coupling overnight, the resin was filtered and washed before freshly activated protected peptide was introduced and the coupling continued. The final percentage coupling for each reaction was then determined, examining the Fmoc content of the resin-bound peptide and measuring the ratios of the N-terminal amino acids. Table 3.2 summarises the estimated coupling efficiencies for the 20 and 40 residue peptides, comparing their results with the calculated fragment condensation efficiency of FmocGly\textsuperscript{70}-Gly\textsuperscript{89}-OH.

<table>
<thead>
<tr>
<th>Fragment</th>
<th>Coupling efficiency</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Fmoc</td>
<td>N-terminal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>deprotection</td>
<td>sequencing</td>
</tr>
<tr>
<td>Gly\textsuperscript{70}-Gly\textsuperscript{89} (11 residues) (27) Mwt. 1,589.78</td>
<td>79%</td>
<td>86%</td>
<td></td>
</tr>
<tr>
<td>Gly\textsuperscript{70}-Gly\textsuperscript{89} (20 residues) (29) Mwt. 2,996.49</td>
<td>14%</td>
<td>17%</td>
<td></td>
</tr>
<tr>
<td>Phe\textsuperscript{50}-Gly\textsuperscript{89} (40 residues) (30) Mwt. 6,632.11</td>
<td>/</td>
<td>/</td>
<td></td>
</tr>
</tbody>
</table>

*Table 3.2 Effect of fragment length on the coupling efficiency*

From these results it was concluded that protected fragments much greater than 10 residues in length would not couple effectively. By doubling the size of the protected fragment, the coupling efficiency dropped considerably to an estimated 15-20%. As for the coupling of the 40 residue fragment, there was no detectable coupling observed. Both couplings were repeated varying the reaction conditions, however, there was no improvement on these initial results.
3. The Convergent Synthesis of SCD

Considering the bulk associated with these protected peptides, it was not highly surprising that, as the fragment length and molecular weight increased, the coupling efficiency dropped to zero. Furthermore, as the size of any protected fragment increases, accompanied with a greater number of protecting groups, the generally low solubility is reduced further which would clearly affect the coupling efficiency. This was observed with FmocGly^{50}-Gly^{89}-OH which, on introduction of DMF, initially formed a 'gel-like' solution and required a large volume of solvent to completely dissolve the peptide.

Overall literature examples where fragments of more than 15 amino acids in length are effectively coupled onto the solid support are extremely rare, with severe problems encountered with larger fragment purification and coupling.

This is also the case with solution phase fragment condensation of larger polypeptides. Generally the target sequence is assembled by initially coupling protected peptides less than 10 residues in length together yielding blocks of peptide containing 20-30 amino acids. The target sequence is then finally prepared by coupling these larger 'blocks' in solution. It is proposed that in the absence of the bulky solid phase, the two large protected fragments are less hindered and more flexible, improving the overall coupling ability. However, examples in the literature reporting the solution phase preparation of polypeptides generally do not record the final coupling efficiency, therefore it is uncertain how successful this method of coupling is.

3.6 The Convergent Synthesis of SCD

After establishing suitable coupling conditions and investigating the effect of fragment length on the overall coupling efficiency, the preparation of SCD via convergent synthesis was studied.
3.6.1 Preparation of H$_2$N-Asp$^{71}$-Thr$^{173}$-resin

The first two fragments to be synthesised corresponded to the next 10 (Asn$^{80}$-Gly$^{89}$) and following 9 (Asp$^{71}$-Gly$^{79}$) residues of SCD respectively, both of which contained a C-terminal Gly, thus the problem with racemisation was avoided. On detachment from the 2-chlorotritylchloride resin, both peptides were individually characterised before they were successively coupled onto the polypeptide-resin, incorporating a capping cycle and Fmoc N$^\alpha$ deprotection between the two condensation reactions. Fig. 3.11 outlines the strategy and presents the resulting coupling efficiencies.

Both peptide fragments were successfully incorporated onto the polypeptide chain. Unfortunately an accurate measure of the percentage coupling for either fragment was not provided by the two methods used. Although measuring the Fmoc content of the resin bound sample provided an almost instant estimation of any coupling reaction, the overall accuracy of this method when dealing with such large polypeptides was uncertain. N-terminal sequencing analysis however generally provided a more accurate account of the coupling, measuring the concentrations of the two N-terminal amino acids (coupled:uncoupled). Overall these results were found to be more consistent and reliable (Table 3.3 A and B).
3. The Convergent Synthesis of SCD

A. Coupling of FmocAsn\[^{80}\]-Gly\[^{89}\]-OH (31) to H\_2N-Pro\[^{90}\]-Thr\[^{173}\]-resin

<table>
<thead>
<tr>
<th>Residue</th>
<th>Cycle 1</th>
<th>Cycle 2</th>
<th>Cycle 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major</td>
<td>Asn[^{80}] 610.51 pmol</td>
<td>Val[^{87}] 887.88 pmol</td>
<td>Leu[^{82}] 570.72 pmol</td>
</tr>
<tr>
<td>Minor</td>
<td>Pro[^{90}] 66.56 pmol</td>
<td>Gly[^{97}] 95.82 pmol</td>
<td>Ile[^{92}] 39.34 pmol</td>
</tr>
<tr>
<td>% coupling</td>
<td>90%</td>
<td>90%</td>
<td>94%</td>
</tr>
</tbody>
</table>

B. Coupling of FmocAsp\[^{71}\]-Gly\[^{79}\]-OH (32) to H\_2N-Asn\[^{80}\]-Thr\[^{173}\]-resin

<table>
<thead>
<tr>
<th>Residue</th>
<th>Cycle 1</th>
<th>Cycle 2</th>
<th>Cycle 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major</td>
<td>Asp[^{71}] 151.03 pmol</td>
<td>Phe[^{72}] 119.86 pmol</td>
<td>Tyr[^{73}] 75.18 pmol</td>
</tr>
<tr>
<td>Minor</td>
<td>Asn[^{80}] 51.44 pmol</td>
<td>Val[^{87}] 41.41 pmol</td>
<td>Leu[^{82}] 37.14 pmol</td>
</tr>
<tr>
<td>% coupling</td>
<td>75%</td>
<td>74%</td>
<td>67%</td>
</tr>
</tbody>
</table>

Table 3.3 Results from N-terminal sequencing following fragment condensation

3.6.2 Preparation of H\_2N-Glu\[^{57}\]-Thr\[^{173}\]-OH

In order to avoid any problems with racemisation, the next fragment to be coupled was a peptide 14 residues in length, corresponding to residues 57 to 70 in the primary sequence of SCD (Fig. 3.12). Considering the difficulties associated with coupling a 20 and 40 residue peptide, it was uncertain whether a fragment of this size would couple effectively, however it was concluded that proceeding in one fragment would be preferable to dividing the fragment further.

\[
\text{But} \quad \text{But} \quad \text{But} \quad \text{Pmc} \quad \text{But} \quad \text{Trt}
\]

\[
\text{Fmoc-Glu}^{57}\text{-Ala-Asp-Ile}^{60}\text{-Met-Ile-Ser-Phe-Ala}^{65}\text{-Val-Arg-Glu-His-Gly}^{70}\text{-OH}
\]

**Fig. 3.12 The primary sequence of FmocGlu\[^{57}\]-Gly\[^{70}\]-OH**

3.6.2.1 Stepwise synthesis of Glu\[^{57}\]-Gly\[^{70}\]

As with the other syntheses of protected peptides, Glu\[^{57}\]-Gly\[^{70}\] was prepared on Fmoc-Gly 2-chlorotritylchloride resin. Unfortunately, from the Fmoc N\[^{a}\] deprotection profile, it was clear that the stepwise synthesis was unsatisfactory, highlighting a number of considerable drops in the amino acid coupling efficiency (Fig 3.13). Attempts to separate FmocGlu\[^{57}\]-Gly\[^{70}\]-OH from the deletion sequences by HPLC proved problematic, hindered by the insolubility of the sample. As a result, it was not
possible to isolate sufficient quantities of the peptide to proceed with the fragment condensation using this 14 residue segment.

3.6.2.2 Dissection of Glu$^{57}$-Gly$^{70}$

Considering the problems encountered with the synthesis of Glu$^{57}$-Gly$^{70}$, efforts concentrated on the preparation of smaller fragments within this sequence assessing their coupling abilities. Two smaller peptides Val$^{66}$-Gly$^{70}$ and Phe$^{64}$-Gly$^{70}$ were synthesised and, following complete characterisation, both peptides were individually activated and sonicated with H$_2$N-Asp$^{71}$-Thr$^{73}$-resin using identical conditions. On completion, their coupling efficiencies were compared (Table 3.4).

**A. Coupling of FmocVal$^{66}$-Gly$^{70}$-OH (33) to H$_2$N-Asp$^{71}$-Thr$^{73}$-resin**

<table>
<thead>
<tr>
<th>Residue</th>
<th>Cycle 1</th>
<th>Cycle 2</th>
<th>Cycle 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major</td>
<td>Val$^{66}$ 638.71pmol</td>
<td>Arg$^{67}$ 122.78pmol</td>
<td>Glu$^{65}$ 171.51pmol</td>
</tr>
<tr>
<td>Minor</td>
<td>Asp$^{71}$ 17.53pmol</td>
<td>Phe$^{72}$ 20.33pmol</td>
<td>Tyr$^{73}$ 9.74pmol</td>
</tr>
<tr>
<td>% coupling</td>
<td>97%</td>
<td>86%</td>
<td>95%</td>
</tr>
</tbody>
</table>

**B. Coupling of FmocPhe$^{64}$-Gly$^{70}$-OH (34) to H$_2$N-Asp$^{71}$-Thr$^{73}$-resin**

<table>
<thead>
<tr>
<th>Residue</th>
<th>Cycle 1</th>
<th>Cycle 2</th>
<th>Cycle 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major</td>
<td>Phe$^{64}$ 38.32pmol</td>
<td>Ala$^{65}$ 25.65pmol</td>
<td>Val$^{66}$ 15.58pmol</td>
</tr>
<tr>
<td>Minor</td>
<td>Asp$^{71}$ 17.56pmol</td>
<td>Phe$^{72}$ 15.96pmol</td>
<td>Tyr$^{73}$ 7.71pmol</td>
</tr>
<tr>
<td>% coupling</td>
<td>69%</td>
<td>62%</td>
<td>67%</td>
</tr>
</tbody>
</table>

*Table 3.4 Results of N-terminal sequencing following fragment condensation*
Initially it was surprising that the coupling efficiencies of the two peptides were so different considering the 5 residues at the C-terminus were identical. However the two additional residues attached onto Phe$^{64}$-Gly$^{70}$ are both hydrophobic, therefore it was concluded that the overall solubility of the fragment was reduced, hindering coupling onto the solid support. Attempts to further solubilise Phe$^{64}$-Gly$^{70}$ were examined, measuring the coupling efficiency in different solvent systems. Unfortunately, the overall percentage coupling could not be improved with an estimated 20% drop in the coupling efficiency observed when DMF was replaced with DCM containing 10% TFE.

3.6.2.3 Synthesis of Glu$^{57}$-Ala$^{65}$

Because of problems encountered with the fragment condensation of FmocPhe$^{64}$-Gly$^{70}$-OH onto the solid support, FmocVal$^{66}$-Gly$^{70}$-OH was coupled to H$_2$N-Asp$^{71}$-Thr$^{173}$-OH (estimated coupling efficiency was 93%) and the next fragment for coupling was prepared. However, the problems associated with the synthesis of Glu$^{57}$-Gly$^{70}$ were still apparent in the stepwise synthesis of Gly$^{57}$-Ala$^{65}$, with a considerable drop in the amino acid coupling efficiencies observed in the region between Met$^{61}$ and Asp$^{59}$ (Fig. 3.14).

![Graph](image)

**Fig. 3.14 Stepwise synthesis of Glu$^{57}$-Ala$^{65}$. Fmoc N° deprotection profile**

From the HPLC and FAB MS of this hydrophobic peptide, it was evident that a large number of deletion sequences were present. Furthermore, due to the insolubility of the
peptide sample, the isolation of sufficient quantities of protected FmocGly^{57}-Ala^{65}-OH for a coupling reaction proved problematic.

This region of the SCD sequence, which was proving difficult to assemble using fragment condensation, was finally prepared by initially coupling the small peptide FmocIle^{62}-Ala^{65}-OH (35) onto the solid support (estimated coupling efficiency was 89%). Met^{61}, Ile^{60} and Asp^{59} were then individually coupled onto the polypeptide chain, thus the coupling efficiency of each amino acid could be measured. Fig. 3.15 summarises the final coupling strategy used for the preparation of H_{2}N-Asp^{59}-Thr^{73}-resin (Fig. 3.15).

3.6.3 Attempted synthesis of Leu^{48}-Ala^{58}

It was intended that once the amino acids centered at the problematic region had been coupled, fragment coupling could recommence with the preparation and coupling of the next protected fragment. Unfortunately, from the reaction profile for the synthesis of FmocLeu^{48}-Ala^{58}-OH (Fig. 3.16), it was obvious that the problems associated with the synthesis of these small peptides still remained, with the overall coupling efficiency dropping by an estimated 60% during the coupling of 4 residues (Arg^{52} to Thr^{59}).
Once again the convergent assembly of SCD was hindered by the fragment preparation. To help provide an explanation of why certain peptide sequences proved difficult to prepare, the hydrophobicity index for SCD was examined (Fig. 3.17). Although stepwise chemical synthesis was performed in organic solvent and not in aqueous conditions, the solubilising effect of different amino acid side-chains still varied greatly in solvents such as DMF, affecting the overall solubility of individual peptide sequences.

Fig. 3.16 Stepwise synthesis of Leu$^{48}$-Ala$^{58}$. Fmoc N$^\alpha$ deprotection profile

Fig. 3.17 Hydrophobicity index of SCD
Fig. 3.17 highlights that the section centred between residues \( \sim 50-70 \) of SCD is hydrophobic in nature. This would explain why such difficulties were encountered on synthesising fragments from this region. During the stepwise synthesis of the full length SCD, the amino acid couplings in this area did not appear to be problematic, with no significant drops in the coupling efficiency observed. In this case, the growing polypeptide chain on the solid support contained a number of hydrophilic regions which would aid the overall solubility of the protein chain. However, in the absence of this solubilising 'arm', the synthesis of the small peptide fragments, consisting mainly of hydrophobic residues, has proved problematic.

As with the fragment preparation and coupling of the region between Glu\(^{57}\)-Glu\(^{70}\), the strategy for the coupling of Ile\(^{48}\)-Ala\(^{58}\) would have to be revised, possibly dividing the peptide further into two smaller fragments, each of which would have to be synthesised and characterised before coupling could proceed. Considering the number of difficulties already encountered with the synthesis and coupling of protected peptides within a small region of SCD, it was decided to suspend this study.

3.7 The Convergent Synthesis of SCD. Review of Initial Findings

At the end of this preliminary investigation, it was concluded that when designing a convergent solid phase strategy for any peptide or protein, although guidelines for the division of a target sequence into fragments could be followed, changes in strategy as a direct result of unpredictable fragment preparation, solubility or low coupling efficiencies often have to be made.

With the fragment assembly of the primary sequence of SCD, the continuous problems encountered all required close examination, which unfortunately protracted the research attention away from the initial project aim - to prepare SCD using chemical methods, in order to study the ability of the synthetic protein to fold correctly, and display the characteristic enzymatic activity.
On reflection, the underlying problem associated with the fragment coupling is the necessity for fully protected peptides. As a direct result of these bulky groups, this study has illustrated that fragment length is restricted in order to avoid severe problems associated with peptide insolubility. Furthermore it has been demonstrated that the overall solubility is highly unpredictable even for the smaller protected peptides.

### 3.8 Convergent Synthesis Using Minimally Protected Fragments

Arising from the inherent problems associated with fragment insolubility, alternative coupling strategies where the requirement for side-chain protection is minimal or completely unnecessary have been investigated. The remaining sections in this chapter shall briefly deal with one such strategy which was applied to the convergent synthesis of SCD.

The azide condensation method (Fig. 3.18) is one of the most successful approaches for the coupling of minimally protected fragments via the native amide link. The flexibility of this approach for the convergent synthesis of proteins was initially demonstrated in 1969, when Denkwalter et al reported the total synthesis of bovine pancreatic ribonuclease A (RNAse A) S-protein (104 residues) by the assembly of small minimally protected fragments via azide condensation. Unfortunately, the product displayed little enzymatic activity and was not fully characterised. However, in 1980, Yamjima et al described the preparation of full length Rnase A (124 residues) using azide condensation to assemble 30 peptide fragments. Not only did this synthetic protein display full enzymatic activity, but it was also obtained in a crystalline form.

#### 3.8.1 Principles of azide condensation. 1. Side-chain protection

Due to the design of the azide condensation, once the C-terminal azide or azide precursor has been incorporated, it is no longer necessary to block the carboxylic side
3. The Convergent Synthesis of SCD

3.8.2 Generation of the C-terminal azide

Due to the reactivity and instability of azides generally, their preparation and isolation prior to coupling is not considered a practical approach. Instead, the C-terminal hydrazide, a precursor of the azide, is normally prepared. This C-terminally modified peptide is inherently more stable than the azide, allowing the fragment to be characterised and stored, if necessary, prior to coupling.

The method that is generally adopted for conversion of the peptide hydrazide to the corresponding azide uses the conditions developed and optimised by Honzl and Rudinger\textsuperscript{30} (Fig. 3.19). The conversion is carried out at low temperatures using anhydrous conditions in order to minimise potential side-reactions. The peptide hydrazide (36) is reacted with an alkyl nitrite such as \textit{i}-butyl nitrite which generates an N-nitroso intermediate (37) before forming the corresponding azide (38) on loss of water. Due to its instability, the peptide azide is then reacted immediately \textit{in situ} with the amino functionality on the other peptide fragment.
3. The Convergent Synthesis of SCD

3.8.3 Preparation of the C-terminal hydrazide

C-terminal hydrazides can be obtained by treating the corresponding C-terminal alkyl ester with hydrazine. Therefore hydrazinolysis can be incorporated into the solid phase preparation of the fragment as highlighted in Fig. 3.20. Reaction with the peptide-resin (39) would release the C-terminal hydrazide from the solid support which, upon standard TFA treatment, would then deprotect the side-chain groups (40).

Unfortunately, there are a number of side-reactions associated with hydrazinolysis. These include removal of the Nα-Fmoc protection and interference at the t-butyl protected Asp and Glu residues. However, the development of a number of hydrazine resin linkers has completely eliminated the need for hydrazinolysis, thus avoiding these accompanying side-reactions.
At Edinburgh, the linker-resin used for the preparation of C-terminal hydrazides is the functionalised dibenzocycloheptadiene system (41) (Fig. 3.21). As with other hydrazide resin linkers, the hydrazide moiety is selectively incorporated onto the C-terminus of the peptide chain at the beginning of the synthesis. On completion of the peptide assembly, the peptide is cleaved and deprotected from the solid support with the hydrazide functionality intact.

![Fig. 3.21 Preparation of C-terminal hydrazides using a functionalised linker](image)

### 3.9 Research Overview

This chapter has already highlighted the problems associated with the convergent synthesis of SCD via the fragment coupling of fully protected peptides. The azide method of fragment condensation employs a minimal protection strategy, therefore it was proposed that the fragment solubility should be greatly improved and the coupling of larger fragments should be feasible.

At Edinburgh, the utility of the azide method has already been illustrated with the successful preparation of salmon calcitonin I (SCT) a 32 residue peptide. Full-length SCT was prepared by coupling together two fragments FmocCys\(^1\)-Gly\(^{10}\)-NHNH\(_2\) and H\(_2\)N-Lys\(^{11}\)-Pro\(^{32}\)-NHNH\(_2\) using the azide method.\(^{34,35}\)
3. The Convergent Synthesis of SCD

The aim of this study was to investigate whether the azide condensation methodology could be extended, permitting the successful coupling of larger peptide fragments. Two fragments of SCD were therefore separately prepared using stepwise SPPS and their ability to couple using azide condensation conditions was examined.

3.10 The Azide Condensation of SCD

Fig. 3.22 highlights the intended coupling strategy between two fragments of SCD, 89 and 48 residues in length. Prior to coupling, both fragments were individually purified and characterised.

3.10.1 Preparation of H$_2$N-Pro$^{90}$-Thr$^{173}$-OH (Lys$^{106}$-(2-Cl-Z)) (26)

The 89 residue peptide was identical to that prepared for the original fragment condensation study (see section 3.5.1 for the purification and characterisation), however, the sequence contains one lysine residue (Lys$^{106}$) which would be required to be protected during the azide coupling. The N$^\delta$-amino functionality of Lys$^{106}$ was blocked using 2-chlorobenzoxycarbonyl (2-Cl-Z) protection. This moiety is stable to both the base and acid cleavage conditions used in the Fmoc SPPS methodology and can be removed following coupling using stronger acid conditions eg. trimethylsilyl trifluoroacetate (TMSOTf) or trifluoromethanesulfonic acid (TFMSA).
3.10.2 Preparation and characterisation of FmocTrp\textsuperscript{42}-Gly\textsuperscript{89}-NHNH\textsubscript{2} (43)

The C-terminal hydrazide peptide was assembled on the dibenzocycloheptadiene derived linker (41) (Fig. 3.23). The first amino acid FmocGly\textsuperscript{89} was manually coupled to the resin via the acid chloride after which the remaining 47 residues were individually coupled using the standard Fmoc methodology. On completion, the peptide was deprotected and cleaved from the solid support by stirring the resin in TFA. Following overnight incubation with MMA, (43) was purified using preparative HPLC.

![Fig. 3.23 Solid phase synthesis of FmocTrp\textsuperscript{42}-Gly\textsuperscript{89}-NHNH\textsubscript{2} (43)](image)

Purified FmocTrp\textsuperscript{42}-Gly\textsuperscript{89}-NHNH\textsubscript{2} was easily characterised by MALDI-TOF MS where it was evident from the mass obtained that the peptide isolated had the hydrazide moiety intact. Furthermore, the amino acid ratios were found to be in good agreement with the expected composition (Fig. 3.24).

3.10.3 Coupling of FmocTrp\textsuperscript{42}-Gly\textsuperscript{89}-NHNH\textsubscript{2} (43) to H\textsubscript{2}N-Pro\textsuperscript{90}-Thr\textsuperscript{173}-OH (26)

After characterising the two peptide fragments, their ability to couple together was investigated. The first step, conversion of the hydrazide to the azide, was carried out at -15°C to suppress possible side-reactions once the azide had formed. The HPLC
<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Calculated value</th>
<th>Expected value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>4.86</td>
<td>5</td>
</tr>
<tr>
<td>Arg</td>
<td>2.07</td>
<td>2</td>
</tr>
<tr>
<td>Asx</td>
<td>4.20</td>
<td>4</td>
</tr>
<tr>
<td>Glx</td>
<td>4.93</td>
<td>5</td>
</tr>
<tr>
<td>Gly</td>
<td>5.14</td>
<td>5</td>
</tr>
<tr>
<td>His</td>
<td>2.39</td>
<td>2</td>
</tr>
<tr>
<td>Ile</td>
<td>1.81</td>
<td>2</td>
</tr>
<tr>
<td>Leu</td>
<td>3.00</td>
<td>3</td>
</tr>
<tr>
<td>Met</td>
<td>0.30</td>
<td>1</td>
</tr>
<tr>
<td>Phe</td>
<td>3.84</td>
<td>4</td>
</tr>
<tr>
<td>Pro</td>
<td>4.47</td>
<td>4</td>
</tr>
<tr>
<td>Ser</td>
<td>1.97</td>
<td>2</td>
</tr>
<tr>
<td>Thr</td>
<td>1.83</td>
<td>2</td>
</tr>
<tr>
<td>Tyr</td>
<td>1.44</td>
<td>3</td>
</tr>
<tr>
<td>Val</td>
<td>2.72</td>
<td>3</td>
</tr>
</tbody>
</table>

Fig. 3.24 Amino acid analysis and MALDI-TOF MS of FmocTrp^{42}-Gly^{89}-NHNH$_2$ (43) (Mwt. 5,607.64)

retention time of the peak corresponding to FmocTrp^{42}-Gly^{89}-NHNH$_2$ shifted on activation, indicating that the hydrazide had been converted. This conversion appeared to occur rapidly, therefore after 10 minutes activation, the amine fragment was added.

Reaction conditions were maintained at 0°C and the coupling progress was monitored using HPLC. After 3 hours coupling little change was observed in the HPLC profile, thus a second portion of hydrazide was activated and added to the reaction mixture. Following overnight incubation at 0°C and addition of a third portion of azide, there was still little change in the HPLC profile with no new signal corresponding to the coupling product observed.

Following 48 hours of coupling, the reaction mixture was analysed using MALDI-TOF MS and SDS gel electrophoresis. From the crude mixture, there was no evidence that any coupling had occurred. All components of the reaction mixture...
were then separated using preparative HPLC and individually assessed using amino acid analysis, MALDI-TOF MS and gel electrophoresis. Unfortunately, there was still no proof that the polypeptide FmocTrp\textsuperscript{47}-Thr\textsuperscript{73}-OH with a molecular weight of $\sim$15,000 had been prepared.

The azide coupling was repeated, taking care to ensure that the low temperature and anhydrous conditions were maintained throughout. Unfortunately, following detailed analysis of the reaction mixture, it was evident that the azide coupling was not proceeding to any extent where even small quantities of the product could be detected or isolated.

3.11 The Convergent Synthesis of SCD. Conclusions and Recommendations

As part of the continuing study centred on the chemical synthesis of enzymatically active SCD, the feasibility of preparing the 173 residue protein using convergent synthesis was investigated. Two different strategies were examined which involved the coupling of fully and minimally protected peptide fragments. Both methods have been successfully applied for the preparation of small peptide sequences. Furthermore, there have been numerous reports in the literature describing the convergent synthesis of proteins, prepared by coupling small peptides using either of these methods of condensation.

This study has highlighted a number of problems associated with convergent synthesis when the coupling of larger peptide fragments was attempted. Initially efforts concentrated on the coupling of fully protected peptides onto the solid phase, however the study was hindered not only by the general insolubility of the protected peptide which restricted the fragment length, but also by difficulties encountered with the stepwise preparation of a number of the peptide sequences. Consequently, the coupling strategy had to be constantly revised and it was envisaged that the complete
convergent synthesis would require a greater number of peptide fragments than was initially intended.

Azide fragment condensation which involved the coupling of minimally protected peptides was then studied. It was anticipated that, in the absence of the side-chain protection, improved peptide solubility would result in more effective fragment coupling. Although this method has been successfully applied for the coupling of small peptides, the methodology could not be extended to large fragments with no coupling observed. The main reasons for the unsuccessful coupling were concluded to be due not only to the size of the two peptides, but also as a result of the low temperatures used to avoid azide decomposition over time.

Despite the problems experienced, it is clear that both methods of convergent synthesis possess a number of properties which are beneficial for the chemical synthesis of proteins. Thus the design and development of a new, improved strategy which is a combination of these two methods is now the subject of ongoing research at Edinburgh\textsuperscript{20} (Fig. 3.25).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig3_25}
\caption{Outline of proposed coupling strategy}
\end{figure}

Similar to the azide condensation, minimal protection is required as the C-terminal hydrazide functionality is incorporated at the beginning of the synthesis. Although the hydrazide (44) is converted to the azide (45) as before, HOCt present in the coupling
mixture, displaces the C-terminal azide affording the HOCT activated ester (46). Finally, coupling with an other fragment (47) proceeds with the usual Nα-amino attack.

Findings from the preliminary model studies are very encouraging. It has been shown that coupling does not need the low temperature conditions required for the original azide coupling, thus increasing coupling potential. Furthermore, the results show that when HOCT is present in the reaction mixture, the coupling proceeds more rapidly and the overall coupling efficiency is obviously improved compared with the azide coupling without HOCT.

This study has now been extended to the coupling of small, minimally protected peptides where it has been clearly demonstrated once again that the addition of HOCT improves the rate and extent of coupling. With the present research efforts concentrating on the preparation of di-ubiquitin by coupling together two single ubiquitin units (76 residues in length), it is anticipated that this coupling strategy will soon prove invaluable for the preparation of large protein sequences.
3.12 References

1292.
18. Inui T., Bodi J., Kubo S., Nishio H., Kimura T., Kojima S., Maruta H.,
Biochem. Soc. Trans., 1990, 18, 1297-1299.
1996, 2, 40-46.
Jiang L., Morton G.H., Robertson N., Shaw K.T., Tennant G., Urquart K. and
Wilken J., Innovation and Perspectives in Solid Phase Synthesis and

98
Chapter Four

The Chemical Ligation of Proteins

4.1 Background

The previous chapter has demonstrated how although the classical methods of convergent synthesis have been successfully applied for the chemical synthesis of proteins, there are still a number of significant problems associated with the coupling of large peptide sequences. Important advances in peptide synthesis over recent years however, have now made it possible to join two large, completely unprotected peptides in a specific and controlled fashion without any significant side reactions.

The concept of this fragment ligation, although simple, has proved to be of substantial benefit for the convergent synthesis of peptides and proteins. By substituting the classical coupling approach used in peptide synthesis, namely the amination of the activated carboxylate, with an alternative strategy involving two unique, mutually reactive functional groups, fragment coupling can now proceed via a controlled chemoselective reaction which does not involve any other functional group present.

One of the main distinctions from the aforementioned methods of convergent synthesis is that, in the absence of side-chain protection or anhydrous conditions, these chemoselective reactions can generally be performed in aqueous conditions. Therefore the coupling mixture can be buffered at a pH which complements the selective reaction and avoids side-chain involvement. Furthermore, denaturing agents
such as urea or Gdm.HCl can be used to enhance the efficiency of the coupling by solubilising and denaturing the peptide fragment.

Several research groups have reported a number of powerful ligation strategies which have been applied successfully for the coupling of large unprotected peptides. This chapter shall initially concentrate on a selection of these methods which clearly demonstrate both the flexibility and diversity of the chemical ligation, before presenting an outline of the ligation strategy designed and developed at Edinburgh with results of the preliminary studies. For a more extensive account of the various ligation methods that have recently been developed, reviews by Muir, Kent, and Hilvert are recommended.

4.2 Protein Synthesis via Native Chemical Ligation

Kent et al, considered to be leaders in the field of chemical ligation, have designed and developed a number of chemoselective strategies which have made the convergent synthesis of a number of proteins accessible. One highly powerful approach, referred to as native chemical ligation, results in the straightforward generation of proteins with native backbone structures from fully unprotected peptide fragments (Fig. 4.1).

The initial step involves the chemoselective reaction at an unprotected C-terminal thioester peptide (48), with another unprotected fragment containing an N-terminal cysteine residue (49). The initial product of this attack is a thioester (50) which, without changes in the reaction conditions, undergoes a spontaneous intramolecular rearrangement to form the native amide bond with the regenerated Cys side-chain next to the site of ligation (51).
Model studies established that this native chemical ligation was applicable to peptides containing any of the natural functionalities, even the side-chain sulfhydryl function of cysteine. Although other internal cysteines can undergo the initial nucleophilic attack, there is no possible rearrangement to the amide bond. Furthermore, as the formation of the thioester is readily reversible in the conditions used (pH 7.5), this side reaction shall not hinder the coupling process. Further evidence that this side reaction was unprogressive was illustrated by the successful preparation of interleukin-8, a 72 residue cytokine containing 4 cysteines and turkey ovomucoid third domain (51 residues) which contains 6 cysteine residues.

A second scheme for the selective coupling of unprotected peptides which results in amide bond formation was recently reported by Tam et al. As with the preceding method, fragment coupling employed the nucleophilic side-chain functionality on the N-terminal amino acid which was initially attracted to the activated C-terminal thiocarboxylic acid, essentially holding the two peptides together permitting rearrangement to yield the peptide bond. Here the imidazole group on the side-chain
4. The Chemical Ligation of Proteins

of histidine, a weak base in acidic pH, was used instead of the sulphydryl moiety of cysteine. Following intramolecular rearrangement, the target peptide was generated with the recovered His side-chain at the site of ligation.

Presently the success of this second coupling strategy has only been demonstrated using small unprotected peptides. However, it is anticipated that this fast, selective coupling can be repeated for the preparation of larger polypeptide chains. In summary, both these methods have illustrated the great potential for the native ligation of unprotected peptides by incorporating a nucleophilic moiety at the N-terminus of the peptide fragment.

4.3 Convergent Synthesis of Backbone Engineered Proteins

One of the important advantages of chemical ligation is that the fragment coupling is not restricted to ligation via amide bond formation. A number of reports in the literature have clearly demonstrated that peptide and protein fragments can be joined by linkages other than the usual peptide bond, resulting in the construction of novel molecules with a wide range of possible backbone modifications. Replacement of the usual amide bond makes it possible to investigate how regions of the peptide backbone contribute to the protein's conformation and activity. Furthermore, with the development of a number of different strategies for the convergent synthesis of peptides and proteins, the choice of site for ligation within the target sequence should become more flexible.

4.3.1 Convergent synthesis via thioester bond formation

One of the initial methods of chemical ligation to be developed by Kent et al reported the coupling of peptide fragments, replacing the usual amide link with a thioester bond. Here the target sequence (52) was constructed by combining a completely unprotected peptide possessing a C-terminal thioester (53) with a second fragment comprising an N-terminal bromoacetyl moiety (54) (Fig. 4.2).
The success of this ligation was demonstrated with the preparation of an analogue of HIV-1 protease, where a pseudo-peptide bond was incorporated into the molecule between Gly\textsuperscript{51}-Gly\textsuperscript{52}. The two unprotected fragments, both ~50 residues in length, were selectively joined rapidly and in high yield through the mutually reactive functional groups preparing an analogue which exhibited full enzymatic activity.\textsuperscript{7}

Once again the ligation strategy was shown to be compatible with all the side-chain functionalities. A further investigation performed by Kent et al\textsuperscript{8} confirmed that even the nucleophilic sulfhydryl groups were compatible in the coupling conditions, ruling out the requirement for cysteine protection. This selectivity between the thiocarboxylate and the side-chain sulfhydryl was attributed to the acidic ligation conditions (pH 4.3).\textsuperscript{8} The thiocarboxylate group has a much larger pKa value, therefore under the conditions typically used for the thioester ligation, the thioacid was predominantly deprotonated. The sulfhydryl group in the same conditions however remained protonated and was therefore not sufficiently nucleophilic to attack the bromoacetyl moiety.

Fig. 4.2 Chemical ligation of backbone modified HIV-1 protease

Fig. 4.3 pKa values for the thiocarboxyl and sulfhydryl functionalities
4.3.2 Chemical ligation three or more fragments

Kent et al took this thioester ligation one step further in order to construct one of the largest synthetic protein analogues prepared completely using chemical methods (172 residues). This was made possible by assembling a total of four unprotected peptide fragments through an approach involving the use of the two mutually compatible thioester and oxime ligation strategies.

The target sequence encompassed the two small protein units cMyc and Max that together play a central role in controlling mRNA transcription, generally associating to form the noncovalent heterodimer. Although it has been established that for full activity, both proteins were required to be present, preparation of this heterodimer by means of standard DNA technology has not been possible. However this unique, covalently bound cMyc-Max protein was chemically synthesised from the corresponding peptide fragments using the novel ligation strategy (Fig. 4.4).

Firstly, using thioester-forming ligation, the two individual cMyc (55) and Max (56) domains, both ~ 90 residues in length, were separately prepared. The two synthetic analogues were then joined in a regioselective manner via an oxime linkage to yield the heterodimer containing two N-termini (57). Once assembled, this unique cMyc-Max heterodimer displayed the specific DNA binding properties as anticipated.
The findings of this study were invaluable, illustrating how powerful a technique chemical ligation was. It is clear that chemical ligation can be successfully applied for the synthesis of significantly large proteins, coupling relatively large peptide fragments. Furthermore, it is evident that chemical methods can prove suitable for the preparation of protein analogues with novel sequences or unusual backbone structure, which are not always attainable by recombinant technology.

4.4 Research Overview

4.4.1 Proposed strategy for chemical ligation

After reviewing the literature reports describing a number of successful ligation strategies, a coupling approach was devised which was compatible with Fmoc peptide synthesis and of interest and relevance to ongoing research at Edinburgh (Fig. 4.5). Here the chemoselective reaction involves the coupling of a peptide fragment containing a C-terminal sulfhydryl (58) with a second peptide comprising an N-terminal bromoacetyl functionality (59). The nucleophilic attack occurred via an $S_N^2$ mechanism, yielding the target sequence with a thioether replacement (60) at the site of ligation.

![Chemical ligation via thioether bond formation](image)

*Fig. 4.5 Chemical ligation via thioether bond formation*
The principle of this chemoselective reaction is similar to the thioester ligation designed by Kent\(^7\) (section 4.3.1), however it was anticipated that the thioether replacement would be inherently more stable at a higher pH (it was reported that the HIV-1 protease analogue containing the thioester link, although stable in acidic conditions, had a half-life of 2 hours at pH 7.5, rapidly undergoing thioester hydrolysis).\(^7\)

The success of the thioether replacement has already been illustrated using a number of ligation strategies including the polymerisation of peptides,\(^1^1\) the preparation of a novel protein comprising two C-termini\(^1^2\) and the synthesis of a second analogue of HIV-1 protease\(^1^3\) with the surrogate peptide link again incorporated between Gly\(^5^1\)-Gly\(^5^2\).

### 4.4.2 Design of effective coupling conditions

As with the other ligation methods discussed, the choice of a suitable pH, resulting in efficient and selective coupling, was crucial. Previous studies had revealed that no coupling occurred between the cysteine side-chain and bromoacetyl moieties at low pH\(^8\) (section 4.3.1), the pH was clearly required to be sufficiently high to render the sulfhydryl group nucleophilic. In summary, thioether coupling have been reported to proceed effectively in solutions buffered between pH 7 and 8.\(^1^1^-^1^3\)

Finally, due to the nature of the C-terminal moiety and the coupling conditions used, cysteine protection was now necessary. Furthermore, there was concern that the N\(^ε\)-amino group of lysine would be suitably nucleophilic to attack the N\(^α\)-bromoacetyl, however, it has been demonstrated that the thioether ligation proceeds selectively in the presence of unprotected lysine at various pHs between 7 and 8.5 without side-chain interference.\(^1^2^-^1^3\)
4.5 Preparation of Ligation Fragments

4.5.1 Synthesis of the functionalised resin-linker

The thiol moiety (-NH-(CH₂)₂-SH) situated at the C-terminus of the ligation peptide was incorporated into the fragment using the versatile dibenzocycloheptadiene-5-ol linker (61), designed and developed by Ramage and co-workers.¹⁴

![Fig. 4.6 Preparation of the functionalised linker](image)

N-Fluorenlymethoxycarbonyl-aminoethyl-2-thiol (62) was prepared starting with aminoethanethiol (64). To ensure that peptides with the intended C-terminus were prepared it was essential that the -SH functionality was selectively attached to the linker with the amine function available to couple with the first amino acid. Therefore the amine function was selectively protected with Fmoc (Fig. 4.7).

![Fig. 4.7 Synthesis of FmocNH-(CH₂)₂-SH (62)](image)
Before incorporation of the thiol moiety, the linker was initially loaded onto the solid support. This was achieved by coupling the dibenzocycloheptadiene linker via its caesium salt (67) to chloromethylpolystyrene resin (Merrifield) (68), forming a stable ether link. By assessing the chlorine content of the resin, an estimation of the coupling efficiency was made, from which it was concluded that after 4 days, the loading was almost quantitative. Finally, after reduction of the ketone (69) to the corresponding alcohol (61), FmocNH-(CH₂)₂-SH was introduced onto the solid support (Fig. 4.8).

Acid catalysis was used to load (62) onto the linker. Reaction conditions applied were analogous to those used previously to derivatise the resin-linker with protected hydrazine and amine functionalities. To assess the coupling progress, samples of resin were removed during coupling. After sonication in 20% piperidine, an estimation of the Fmoc content of the resin-linker was made, by measuring the absorbance of the Fmoc-piperidine adduct at 302nm. From this it was concluded that coupling was almost 100% effective after one hour (estimated resin functionality - 0.77mmol/g), thus completing the synthesis of the linker.

Fig. 4.8 Synthesis of the FmocNH-(CH₂)₂-S-functionalised resin-linker
4.5.2 Synthesis of a C-terminal modified peptide

In order to confirm that this derivatised linker could feasibly be used to prepare peptides with the intended C-terminal functionality, a small peptide sequence (70) was prepared on the modified solid support (63). Once assembled, the peptide-resin was stirred in aqueous TFA releasing the peptide from the solid support (Fig. 4.9).

![Stepwise peptide assembly](Image)

**Fig. 4.9 Solid phase assembly of H₂H-Gly-Phe-Ala-NH-(CH₂)₂-SH (70)**

The resulting crude peptide was easily characterised using FAB MS confirming that the small peptide had been prepared with the C-terminal functionality intact. Furthermore HPLC and amino acid analysis confirmed that (70) was synthesised in the absence of any side reactions. However, there was evidence from both the HPLC and MS that, once the peptide was cleaved from the solid support, peptide dimerisation via disulfide bond formation had occurred to a small extent.

4.5.3 Preparation of bromoacetylated peptides

The bromoacetyl functionality was incorporated into the peptide at the end of the stepwise synthesis using bromoacetic acid, which was first converted to the symmetrical anhydride using DIC, before being coupled with the N-terminal free amide. Bromoacetylation of a small peptide sequence (71) assembled on Wang resin was successfully achieved (Fig. 4.10). Following removal of the final N°-Fmoc
4. The Chemical Ligation of Proteins

protection, the peptide resin was sonicated with bromoacetic anhydride for 1 hour after which time the reaction was complete. This was confirmed by the ninhydrin test for free amines which gave a negative result.

$$\text{Br-CH}_2\text{-CO-NH-Leu-Arg-Leu-Arg-Gly-Gly-OH}$$

Following deprotection and cleavage from the solid support, HPLC and FAB-MS analysis of the bromoacetylated peptide (72), confirmed that the correct peptide sequence had been prepared.

4.6 Peptide Coupling via Thioether Bond Formation

4.6.1 Chemical ligation - Preliminary investigation

Following successful preparation of peptides containing the necessary functionalities, their ability to couple under suitable conditions was examined. Therefore two peptides were synthesised incorporating the bromoacetyl (72) and sulfhydryl (73) functionalities. Furthermore, Tbfmoc (section 3.5.4) was attached onto the N-terminus of the C-terminal modified peptide (73). This should not only simplify purification at the end of ligation, but Tbfmoc is also a useful chromatographic probe (strongly absorbing at 364nm) which would help pinpoint both the sulfhydryl peptide
(73) and ligation product (74) while monitoring the ligation progress using HPLC analysis.

\[
\text{Tbfo}c\text{NH-Thr}^{1}-\text{Asp-Glu-Thr-Leu-His-Leu-Val}^{8}\text{-NH-(CH}_{2}\text{)}_{2}\text{-SH (73)}
\]

\[
+ \quad \text{Br-CH}_{2}\text{-CO-Leu}^{11}\text{-Arg-Leu-Arg-Gly-Gly}^{16}\text{-OH (72)}
\]

Gdm.HCl pH 7/
trifluoroethanol
(1:1)

\[
\text{Tbfo}c\text{NH-Thr}^{1}-\text{Asp-Glu-Thr-Leu-His-Leu-Val}^{8}\text{-NH-(CH}_{2}\text{)}_{2}\text{-S-CH}_{2}\text{-CO-Leu}^{11}\text{-Arg-Leu-Arg-Gly-Gly}^{16}\text{-OH (74)}
\]

**Fig. 4.11 Preparation of peptide analogue (74) via chemical ligation**

Once both peptides had been synthesised and characterised, they were separately dissolved in 6M Gdm.HCl pH 7/TFE (1:1). The two solutions were then mixed and the progress of ligation monitored by analytical HPLC (Fig. 4.12).

**Fig. 4.12 The chemical ligation of a bromoacetylated (72) and C-terminal modified (73) peptide. The ligation progress was assessed using dual wavelength HPLC, monitoring at 214 (solid line) and 364 (broken line) nm.**
The ligation occurred rapidly. Even the first sample, injected when the solutions were initially mixed, clearly displayed product formation. From the HPLC traces it was concluded that the majority of the coupling proceeded within the first few hours of reaction time. Furthermore, it was demonstrated that the coupling could be driven to near completion by adding an excess of one of the peptide fragments (a further portion of (72) was added after 3 hours of coupling).

On completion of the coupling, the ligation peptide (74) was easily isolated from the ligation mixture by semi-preparative HPLC. The product was then characterised by MALDI-TOF MS, before and after removal of Tbfmoc, and assessed using amino acid analysis (Fig. 4.13) verifying that the ligation had proceeded successfully.

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Calculated value</th>
<th>Expected value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg</td>
<td>2.46</td>
<td>2</td>
</tr>
<tr>
<td>Asx</td>
<td>1.03</td>
<td>1</td>
</tr>
<tr>
<td>Glx</td>
<td>1.03</td>
<td>1</td>
</tr>
<tr>
<td>Gly</td>
<td>2.09</td>
<td>2</td>
</tr>
<tr>
<td>His</td>
<td>1.02</td>
<td>1</td>
</tr>
<tr>
<td>Leu</td>
<td>3.71</td>
<td>4</td>
</tr>
<tr>
<td>Thr</td>
<td>1.94</td>
<td>1</td>
</tr>
<tr>
<td>Val</td>
<td>0.91</td>
<td>1</td>
</tr>
</tbody>
</table>

Fig. 4.13 Amino acid analysis and MALDI-TOF MS of ligation product

\( \text{H}_2\text{N}-\text{Thr}^\text{I}-\text{Gly}^\text{I}^\text{I}-\text{OH} \) (74)

4.6.2 Investigation into altering the rate of substitution

Although the success of fragment coupling via the nucleophilic attack on a bromoacetylated peptide had been demonstrated, it was proposed that an increase in the rate of substitution could be achieved by substituting Br\(^-\) with a better leaving group.

An investigation into the varying rates of substitution was carried out, comparing the coupling of two haloacetylated peptides with aminoethanethiol (64). Two peptides of identical sequence containing the bromo (72) and iodo (75) functionalities were
synthesised, deprotected and cleaved from the solid support and finally separately dissolved in ligation buffer. A measured excess of the thiol (64) was then added to both solutions and the progress monitored using HPLC (Fig. 4.14).

\[
\begin{align*}
\text{H}_2\text{N}-(\text{CH}_2)_2-\text{SH} & \quad + \quad \text{Br-CH}_2-\text{CO-Leu-Arg-Leu-Arg-Gly-Gly-OH} \quad (72; \\
(64) & \quad \text{I-CH}_2-\text{CO-Leu-Arg-Leu-Arg-Gly-Gly-OH} \quad (75; \\
\text{6M Gdm.HCl pH } 7/ & \quad \text{TFE (1:1)} \\
\downarrow & \\
\text{H}_2\text{N}-(\text{CH}_2)_2-\text{S-CH}_2-\text{CO-Leu-Arg-Leu-Arg-Gly-Gly-OH} & \quad (76)
\end{align*}
\]

**Fig. 4.14** HPLC progress of the nucleophilic substitution reaction involving aminoethanethiol (64) and A. the bromoacetylated peptide (72) B. the iodoacetylated peptide (75)

From the HPLC traces it was evident that the substitution of the iodide ion proceeded more rapidly, with complete conversion to the thiol modified peptide (76) in less than 30 minutes. Meanwhile, over the same period of time, only an estimated 50% of the bromoacetylated fragment had been converted. It was therefore easily concluded that the iodide ion was a faster, more effective leaving group.
4.6.3 Synthesis of an analogue of human proinsulin C-peptide

To complete this study, an analogue of the human proinsulin C-peptide, a 31 residue peptide corresponding to position 33-63 of human proinsulin (77) was prepared. Containing a Gly \(^{43}\)-Gly-\(^{47}\) situated near the centre of the sequence, this peptide was considered a suitable model for the thioether chemical ligation.

\[ \text{H}_2\text{N-Gly}^{33}\text{-Ala-Glu-Asp-Leu-Gln-Val-Gly-Gln-Val-Glu-Leu-Gly-Gly-Gly-Pro-Gly-Ala-}\]
\[ \text{Gly-Ser-Leu-Gln-Pro-Leu-Ala-Leu-Glu-Gly-Ser-Leu-Gln}^{63}\text{-OH (77)} \]

The two peptide fragments (78) and (79) were synthesised using the derivatised dibenzocycloheptadiene and Wang resin respectively. Once the iodoacetyl and Tbfmoc functionalities had been incorporated, the two peptides were individually deprotected and cleaved. On precipitation, the crude fragments were dissolved in the ligation buffer and the coupling progress was monitored (Fig. 4.15).

From the HPLC profile, it was evident that, compared with the previous ligation, the rate of coupling was considerably slower to the extent that even after 25 hours of coupling, there were still appreciable quantities of both peptide fragments present (Fig. 4.15). It was proposed that the rate of coupling was reduced due to the quality of the peptide fragments.

Unfortunately, the stepwise synthesis of the C-terminal modified peptide (78) had proved problematic. A number of drops in the coupling efficiency were observed during peptide synthesis, as a result, the final product was contaminated with various deletion sequences. Due to the problems associated with C-terminal dimerisation, the crude peptide (78) was mixed with the iodoacetylated fragment (79) without purification. From the HPLC profiles, it was evident that a number of impurities were present which would not only hinder the coupling, but also make it difficult to estimate the actual concentration of (78) in solution. Clearly further investigation into the purification of the C-terminally modified peptide prior to coupling was necessary before the coupling of larger fragments could be examined.
Despite all these problems, isolation of the target analogue (80) was relatively simple. With the incorporation of Tbfmoc, both the ligation product (80) and the C-terminal modified fragment (78) were easily removed from the reaction mixture on the addition of PGC. Then, on removal of Tbfmoc, the two peptides were separated using preparative HPLC and the proinsulin C-peptide was characterised by MALDI-TOF MS and amino acid analysis.
4.7 Final Conclusions and Recommendations

Previously it has been established that by incorporating unique, mutually reactive functional groups into the peptide molecule, minimally protected peptides can be selectively coupled in the absence of side-chain interference. Furthermore, it has been demonstrated that large peptide fragments can be coupled together and that it is possible to construct peptide sequences \textit{via} bonds other than the native amide link.

This chapter has highlighted the great potential that the thioether ligation strategy presents for the total chemical synthesis of polypeptide and protein chains. Firstly, the preparation of two peptides incorporating the mutually reactive sulfhydryl and haloacetyl functionalities has been shown to be straightforward and applicable to any peptide sequence. Furthermore, it has been demonstrated that the coupling of the tailored peptide fragments can proceed rapidly and selectively in the presence of a number of side-chain functionalities. Finally, with the incorporation of the Tbfmoc moiety, the final step involving ligation product isolation has been greatly simplified.

In summary, by employing the highly versatile dibenzocycloheptadiene linker and the novel Tbfmoc system, the design of this particular thioether coupling strategy has become straightforward and manageable. It is anticipated that the present success with chemical ligation of small peptide sequences will be reproduced in future thioether ligations involving larger peptide fragments (>50 residues).

Unfortunately, in the case of the human proinsulin C-peptide, a suitable biological assay was not available to compare the activity of the native peptide with the analogue containing the thioether replacement. However, it is generally accepted that the biological activity of any peptide or protein analogue can be affected by an unnatural peptide link, depending on the site of ligation. This has been illustrated by various literature reports where the unnatural bond is generally positioned in a region where the peptide bond is not critical for the protein's structure or activity.\textsuperscript{7,9,13}
Finally, at present the site of ligation is restricted to that between two glycines, however, considering the simplicity of the sulphydryl and haloacetyl moities, it should be possible to design analogues corresponding to some of the other simple amino acids. For example, the bromo/iodoacetic acid could be replaced with another α-substituted bromo/iodoacetic acid. The major consideration of course is that all other amino acids are optically active, therefore care would have to be taken to ensure that the stereochemical integrity was maintained throughout the synthesis of the bromo or iodo components. If a number of amino acid derivatives containing either of the functionalities could be prepared, this thioether chemoselective ligation would become an extremely flexible and practical method for the convergent synthesis of a wide range of protein analogues.
4. The Chemical Ligation of Proteins

4.8 References

3. Hilvert D., Chemistry and Biology, 1994, 1, 201-203.
Chapter Five

Experimental

5.1 Notes

All Fmoc amino acids were purchased from either Bachem or Novabiochem and are of L-configuration. Chloromethylpolystyrene-divinylbenzene resin (Merrifield) and 2-chlorotritylchloride resin (Cl-Trt) were both purchased from Novabiochem, while the p-alkoxybenzylalcohol resin (Wang) was supplied by Bachem. Peptide synthesis grade dimethylformamide (DMF), 1,4-dioxan and piperidine were supplied by Rathburn Chemicals. N,N-diisopropylethylamine (DIEA) and trifluoroacetic acid (TFA) were peptide synthesis grade and purchased from Applied Biosystems (ABI). 1-Hydroxy-4-ethoxycarbonyl-1,2,3-triazole (HOCt),\textsuperscript{1} tetrabenzo[a,c,g,i]fluorenyl-17-methoxy-carbonyl (Tbfmoc),\textsuperscript{2} and Substance P were all synthesised within the research group. Recombinant proMMP-3 was supplied by Dr S.L. Irving, Pfizer Ltd, Kent. 7-Methoxycoumarin-4-acetyl-Pro-Leu-Gly-Leu-β-(2,4-dinitrophenylamino)-Ala-Ala-Arg (Mca-peptide), 7-methoxycoumarin-4-acetyl-Pro-Leu-Gly and p-aminophenyl-mercuric acetate (APMA) were purchased from Sigma.

Melting points were recorded in open capillaries using a Buchi 510 oil immersion melting point apparatus. Analytical thin layer chromatography (TLC) was performed using aluminium sheets precoated with silica gel 60 F\textsubscript{254} (Merck) using the solvent systems indicated in text. Infrared (IR) spectra were recorded on a Bio-Rad SPC 3200 instrument. Ultraviolet (UV) spectra were recorded on a Varian Cary 210 double beam spectrophotometer, a Unicam UV/Vis double beam spectrophotometer or a Perkin Elmer single beam spectrophotometer in the solvents described in the text.
Proton nuclear magnetic resonance (NMR) spectra were recorded on either a Joel FX-60 (60MHz) or a Brucker WP-200 (200MHz).

High and low resolution fast atom bombardment mass spectra (FAB MS) were measured on a Kratos MS50TC instrument, using either thioglycerol, 3-nitrobenzyl alcohol or glycerol as matrix. Matrix assisted laser desorption ionisation mass spectra (MALDI) time of flight mass spectra (TOF MS) were recorded on a PerSeptive Biosystems Voyager™ Biospectrometry™ Workstation using either α-cyan-4-hydroxycinnamic acid or 3,5-dimethoxy-4-hydroxycinnamic acid as matrix. Amino acid analysis (AAA) were performed on a Pharmacia Biochrom 20 LKB 4150 alpha acid analyser on the hydrolysate obtained after heating samples in 6M HCl at 110°C in sealed Carius tubes for the length of time indicated in the text. Fluorescence spectra were recorded on a Perkin Elmer LS50 luminescence spectrometer fitted with a Grant LTD-6 thermostatic control. Protein N-terminal sequencing was performed on an ABI 477A sequencer at the Welnet Protein Characterisation Facility (University of Edinburgh). This was performed by Edman degradation, labelling the N-terminus with phenyl isothiocyanate. Circular dichroism (CD) spectra were recorded at 20°C on a JASCO J600 spectropolarimeter (University of Stirling) in buffers stated in the text.

All buffers described were prepared using Milli-Q grade water, Biochemika Microselect grade guanidine hydrochloride (Gdm.HCl) and urea (Fluka). High performance liquid chromatography (HPLC) was carried out using either an ABI system comprising 2x1406A solvent delivery systems, a 1480 injector/mixer and a 1783 detector/controller, or a Gilson system comprising 2x306 solvent delivery systems, an 811c dynamic mixer, an 805 manometric module, a 119 UV/Vis detector and a Gilson 715 software-driven gradient controller. Peptides were eluted from various columns using a linear gradient of acetonitrile (HPLC grade, Rathburn Chemicals) in Milli-Q grade water, where both solvents contained 0.1% v/v of HPLC grade TFA (Fisons). Fast protein liquid chromatography (FPLC) was carried out on a Pharmacia FPLC system with Liquid Chromatography Controller LCC-501 Plus using a Superdex™ 75 HR 10/30 or 26/60, Mono-Q® HR 5/5 or Resource Phe™ column.
Anion exchange media diethylaminoethyl (DEAE) Sepharose was purchased from Pharmacia. Ion exchange was carried out using Pharmacia LKB apparatus comprising 2xLKB 2138 UVCords, a Pharmacia 2132 Microperpex peristaltic pump, a LKB 2112 redirac fraction collector and a Pharmacia GM-1 gradient mixer. Isoelectric focusing (IEF) was carried out using the Bio-Rad Rotofo r® system with ampholites purchased from Bio-Rad. Polyacrylamide molecular weight gels were run using 20% homogeneous SDS gels (Pharmacia) on a LKB Pharmacia PhastGel electrophoresis system. Dialysis tubing used, purchased from Spectrum®, were Spectra/Por® CE (cellulose ester) membranes with molecular weight cut-offs (MwtCO) as described in the text. Samples were centrifuged using MSE Mistral 2000R (Sanyo). Sample concentration employed Amicon centicon® cells or the Amicon® Ultrafiltration cell 8050 using Spectrum Molecular/Por® membranes, MwtCO as indicated in the text.

5.2 Solid Phase Peptide Synthesis

All peptides were synthesised on an ABI 430A automated peptide synthesiser with on-line UV monitoring using an ABI 758A detector. All peptides were synthesised using the Fmoc strategy of Nα protection which involved the complementary use of orthogonal acid labile side chain protection and acid labile peptide-resin linkers.

5.2.1 Side-chain protecting groups

The side-chain protecting groups used were as follows: no protection for Ala, Gly, Ile, Leu, Met, Phe, Pro & Val; t-butyl (Bu') ethers for Ser, Thr & Tyr; t-butyl (Bu') esters for Asp & Glu; t-butoxycarbonyl (Boc) for Lys & Trp; τ-triphenylmethyl (Trt) for Asn, Gln & His; 2,2,5,7,8-pentamethylchroman-6-sulphonyl (Pmc) for Arg.

5.2.2 Loading of the C-terminal amino acid onto the solid support

A number of resins were used depending on the particular peptide requirements. For example, it was necessary to use differing functionalised linkers for the preparation of
C-terminal thiol, hydrazide and amide modified peptides. These more specific procedures are described later (sections 5.5.2 and 5.6.3).

However, all C-terminal acids were synthesised using either:

1. 4-alkoxybenzylalcohol (Wang) resin - for preparation of fully deprotected peptides.
2. 2-chlorotritylchloride (Cl-Trityl) resin - for preparation of side-chain protected peptides.

1. **Wang resin - a. For resin loadings 0.5 - 0.6 mmol/g**

The C-terminal Fmoc amino acid (6 equivalents) was dissolved in DMF (10ml) before DIC (3 equivalents) was added and the solution sonicated for 15 minutes. Following the formation of the symmetrical anhydride, the solution was added to Wang resin (1g) preswollen in a minimum quantity of DMIF, mixed with a catalytic amount of DMAP (approx. 10mg). The mixture was then sonicated for 2 hours at room temperature. The resin was separated by filtration and sequentially washed with copious amounts of DMF, 1,4-dioxan, DCM and ether before being dried under vacuum.

b. **For resin loadings 0.1 - 0.2 mmol/g**

If a low resin loading was required, 2 equivalents of Fmoc amino acid and 1 equivalent of DIC were used. Following addition to the preswollen resin, the mixture was left to stand for 1 hour before it was washed and dried as before.

2. **Cl-Trityl resin - For resin loadings 0.5 - 0.6 mmol/g**

The C-terminal Fmoc amino acid (1 equivalent) was dissolved in the minimal quantity of DMF (up to 1ml) before DCM (10ml) was added. The solution was the added to Cl-Trityl resin (1g) before DIEA (2.5 equivalents) was gradually added dropwise over 1-2 minutes. The mixture was then sonicated for 1 hour before HPLC grade MeOH (0.8ml) was added to the mixture and sonicated for a further 5 minutes (this ensured capping of remaining uncoupled Trityl groups). The resin was then sequentially
washed with copious amounts of DCM, DMF, DCM, MeOH and ether before being dried under vacuum.

**Note** - The CI-Trityl resin is moisture sensitive therefore, for satisfactory coupling, it was found necessary to use dried glassware, dry DCM (CaCl₂) and the CI-Trityl resin and amino acid were dried under vacuum over KOH pellets.

### 5.2.3 The Fmoc loading test

The loading of the functionalised resin was determined by treating an accurately measured quantity of resin (typically ~5mg) with 20% piperidine/DMF/1,4-dioxan (10ml). This solution was sonicated for 10 minutes before the UV absorbance of the supernatant was recorded between 280 and 320nm, measuring the absorbance at 302nm. The coupling efficiency and resulting resin functionality (mmol/g) was calculated by applying the Beer-Lambert law \( (e = 15,400 \text{ for the Fmoc-piperidine adduct})\).

### 5.2.4 Automated SPPS

All peptides and proteins described were synthesised using an Applied Biosystems 430A automated peptide synthesiser. Every cycle, resulting in the coupling of a single amino acid, involved the following sequence of events -

1. **Capping** - to irreversibly block any unreacted amino groups.
2. **Deprotection** - removal of the Fmoc \( N^\alpha \) protection releasing the amino functionality for coupling.
3. **Coupling** - 1. activation of the required amino acid.

2. reaction with growing peptide on the solid support.

After every step, the resin was washed repeatedly with copious amounts of solvent to ensure removal of all unreacted reagents.

**1. Capping**

The resin was vortexed with a solution of acetic anhydride (0.5M), DIEA (0.125M) and HOBt (0.2% w/v) in DMF/1,4 dioxan (1:1, 10ml) for 10 minutes before the
solution was drained from the reaction vessel and the resin washed with 6 portions of DMF/1,4 dioxan (1:1).

2. Deprotection
The resin was vortexed with a solution of 20% piperidine DMF/1,4 dioxan (1:1, 10ml) for 10 minutes before being drained, passing the solution through a UV detector*. After washing with DMF/1,4 dioxan (1:1, 4 times), the resin was treated with a further two portions of piperidine solution (10ml) vortexing for 1.5 minutes in each case. Finally, the resin was washed with 6 portions of DMF/1,4 dioxan (1:1).

*Note - In order to monitor the progress of the synthesis, following every deprotection, an aliquot of filtrate was passed through the UV detector with on-line integration. The absorbance at 302nm is directly related to the concentration of the Fmoc-piperidine adduct and thus makes it possible to estimate the coupling efficiency of every amino acid.

3. Coupling
1. Activation - prior to coupling, every amino acid was converted to the corresponding activated ester using HOBt, HOCl or HBTU.
   Formation of HOBt ester - Fmoc amino acid (1mmol) was treated with HOBt (1mmol) and DIC (1mmol) in DMF/1,4 dioxan (1:1, 8ml). After 15 minutes, the activated ester was transferred to the reaction vessel.
   Formation of HOCl ester - the procedure was as for the HOBt activation for all amino acids with the exception of His. In the case of His, additional HOCl (2mmol) was added to the cartridge containing His amino acid.
   Formation of HBTU ester - Fmoc amino acid (1mmol) was treated with HBTU (1mmol), DIEA (2mmol) and HOBt (1mmol) in DMF (8ml). After 15 minutes, the activated ester was transferred to the reaction vessel.
2. Coupling - on addition of the activated amino acid to the resin, the mixture was vortexed for 30 minutes before the solution was drained and the resin washed with DMF/1,4 dioxan (1:1, 4 times).
For the majority of syntheses, the above single coupling cycle was sufficient. However in the cases where problematic residues or sequences were encountered, resulting in unsatisfactory coupling efficiencies, a second coupling cycle was incorporated. For this double coupling cycle, an additional portion of Fmoc amino acid (1mmol) was activated and the above coupling procedure repeated.

5.2.5 Preparation of Tbfmoc-peptides

1. Loading of Tbfmoc onto peptide-resin

On completion of the peptide's synthesis, the Nα-Fmoc protection was removed (section 5.2.4.2) and the resin washed and swollen in 1,4 dioxan. Tbfmoc chloroformate (3 equivalents*), dissolved in 1,4 dioxan (10ml for 1g peptide-resin), was added to the resin followed by DIEA (1 equivalent to peptide-resin). The mixture was then sonicated in the dark for 2-3 hours before the resin was filtered and washed with copious amounts of 1,4 dioxan, DCM and ether. The resin was then stored under vacuum in the dark until required.

*Note - on completion of the synthesis, the final Fmoc-amino acid loading was calculated using the Fmoc loading test. Based on the absorbance at 302nm, the overall resin functionality was determined and this compared with the theoretical final loading given 100% yield in every step of the synthesis (information provided using computer programme). This, in turn, provided an estimation of the overall success of the synthesis.

2. The Tbfmoc loading test

The loading of the functionalised resin was determined by treating an accurately measured quantity of resin (typically ~5mg) with 20% piperidine/1,4 dioxan (10ml). After sonicating the solution for 10 minutes, the UV absorbance of the supernatant was recorded between 320 and 400nm. The Tbfmoc-peptide-resin functionality and resulting coupling efficiency was calculated from the following equations:

\[
\text{Resin functionality (Tbfmoc-peptide - mmol/g)} = 0.613 \times \frac{\text{Abs}_{364\text{nm}}}{\text{weight of resin (mg)}}
\]
Coupling efficiency of Tbfmoc (%) = resin funct. mmol/g (Tbfmoc-peptide) / resin funct. mmol/g (peptide)

3. Tbfmoc purification

**Adsorption** - Tbfmoc-peptide, dissolved in 6M Gdm.HCl/i-propanol (1:1, ~10mg/ml), was mixed with washed activated charcoal*. The solution was then stirred until HPLC indicated that all Tbfmoc-peptide had adsorbed onto charcoal (disappearance of all material absorbing at 365nm). The mixture was then centrifuged and the supernatant removed.

**Wash** - the charcoal was washed using 6M Gdm.HCl/i-propanol (1:1, 3 times) in order to remove all deletions and material still in solution. In each case, the mixture was shaken, centrifuged and the supernatant discarded.

**Cleavage** - the charcoal was shaken vigorously in 10% piperidine/6M Gdm.HCl/i-propanol (1:1) for 15 minutes, releasing the peptide from the Tbfmoc moiety. Following centrifugation, i-propanol was removed *in vacuo* and the pH of the solution was altered using AcOH (pH 6-7). Depending on the purification strategy, the product was either dialysed in order to remove the Gdm.HCl followed by lyophilisation or purified directly using preparative HPLC.

*Note* - the quantities of carbon used depended greatly on the size of peptide/protein being purified and varied from 1-5mg carbon:peptide. The best approach involved using a small quantity of carbon initially, and, depending on the extent of adsorbance (indicated by HPLC), further carbon could then be added.

5.2.6 Acidolytic treatment of resin bound peptide

**1. Wang resin** - this general procedure for deprotection and cleavage was optimised depending on the particular peptide. N° deprotected peptide-resin (1g) was swollen and stirred in a mixture of scavengers for 30 minutes. The choice of scavengers varied depending on the peptide being treated, overall the scavengers employed were: anisole (0.5ml); thioanisole (0.5ml); EDT (2ml); phenol (750mg); TIPS (0.25ml). Aqueous TFA (95%, 10ml) was then added to the resin and the mixture stirred under nitrogen at room temperature for the required length of time. The resin was then
removed by filtration and the cleavage mixture precipitated by adding dropwise to ice-cold ether. The resulting white precipitate was then washed in ether (3 times), filtered, dissolved in aqueous AcOH (10-20%) and lyophilised.

2. Cl-Trityl resin - general procedure for the cleavage of side-chain protected peptides. Peptide resin (1g), previously swollen in DCM, was stirred in a solution of AcOH/TFE/DCM (1:1:8, 20ml) at room temperature for 30 minutes under nitrogen. The resin was then filtered and washed with the same solvent system before the solution was concentrated to 3-5ml in vacuo. The crude peptide was then precipitated by trituration in ice-cold ether, washed and filtered. 

Note - 1. When concentrating, the mixture was not evaporated to near dryness. This would expose the protected peptide to neat AcOH which may result in side-chain deprotection. 2. The crude peptide was not lyophilised due to the general insolubility of protected peptides in aqueous media.

5.3 The Chemical Synthesis of Stromelysin (SCD)

Notes

1. Protein concentration determination - Samples were analysed for their protein content using the method described by Edelhoch. By measuring the absorbance at 280 and 288nm, it was possible to determine the content of tryptophan and tyrosine in the protein sample which relates to the concentration of the protein. The protein concentration could then be calculated by applying the following equations -

\[
\varepsilon = \frac{A}{C}
\]

\[
A = \text{UV absorbance}
\]

\[
C = \text{protein concentration}
\]

\[
\varepsilon_{280} = N_{\text{Trp}} 5690 + M_{\text{Tyr}} 1280 \quad N \text{ and } M = \text{no. of moles of Trp and Tyr per mole of protein.}
\]

\[
\varepsilon_{288} = N_{\text{Trp}} 4815 + M_{\text{Tyr}} 385
\]
Prior to the absorbance measurement, a blank sample (containing protein buffer) was run. All protein concentrations were converted to mg/ml.

2. HPLC - Analysis of all the protein samples in this section employed a C₄ column. The conditions used were as listed -
   - Aquapore C₄, 100x4.6mm, 7μm.
   - A=H₂O, B=CH₃CN, 0.1%TFA.
   - 2ml loop, 1ml/min.
   - 0-2min 10%B, 2-32min 10-90%B.
   - λ=214nm.

3. Activation of proSCD using APMA - To a 1ml solution of proSCD (1mg, 34.49mmol) was added APMA (20μl, from a 0.1M stock solution (DMSO)). On mixing, the protein sample was incubated at 37°C for 4 hours. The sample was then centrifuged to remove any precipitates before the lower molecular weight fragments and APMA were removed by spinning down the solution (Mwt.CO. 3,000), washing with 50mM Tris pH 7.5 containing 10mM CaCl₂ and 0.1mM ZnCl₂ (2ml, 3 times). The volume of the protein sample was then adjusted to 1ml using the same buffer system and stored at -20°C until use (est. conc. 0.67mg/ml).

m/z (MALDI-TOF) ~28,500-29,500 (MH⁺ proSCD C₁₂₉₅H₁₉₇₇N₃₄₄O₃₈₉S₆Ca₃Zn₂ requires 29,011.13), ~19,200-19,800 (MH⁺ actSCD C₈₈₂H₁₃₁₁N₂₂₈O₂₆₄S₂Ca₃Zn₂ requires 19,632.31).

4. Calculation of SCD activity - The extent of protein activity was determined by measuring the change in fluorescence over time on incubation of Mca-peptide with SCD. All assays were performed at 37°C in a 100μl cuvette containing:

   10μl Mca-peptide in DMSO (10⁻⁶M stock solution)
   Xμl SCD in 50mM Tris, 10mM CaCl₂, 0.1mM ZnCl₂
   90-Xμl Tris buffer

On mixing, the sample was excited at 325nm and the change in fluorescence emission at 393nm was monitored (Ex./Em. slit width - 15 and 20nm respectively).
Activity calculation - The change in fluorescence as a result of the complete hydrolysis of 0.1\mu M Mca-peptide was determined by separately measuring the fluorescence of the fully intact Mca-peptide and completely hydrolysed Mca-fragment.

1. 10\mu l Mca-peptide (10^{-6} M DMSO)  
   90\mu l Tris buffer  
   (average fluorescence ~20 units) 
2. 10\mu l Mca-fragment (10^{-6} M DMSO)  
   90\mu l Tris buffer  
   (average fluorescence ~270 units) 

Neither samples show any change in fluorescence over time. The protein activity can then be determined:

\[
\text{Activity (units)} = \frac{\text{change in fluorescence/min}}{\text{change in fluorescence for total hydrolysis of Mca-peptide (0.1\mu M)}} 
\]

The activity was quoted as units/mg of protein in the 100\mu l sample.

5. Substance P hydrolysis - To a solution of Substance P (4mg/ml) in 50mM Tris pH 7.5, containing 10mM CaCl\textsubscript{2} and 0.1mM ZnCl\textsubscript{2} was added the activated sample of SCD. The rate of hydrolysis was monitored using HPLC analysis.

HPLC (Vydac C\textsubscript{18}, 250x4.6mm, 5\mu m; 2ml loop, 1ml/min; 0-2min 10\%B, 2-20min 10-50\%B, 20-24min 50-90\%B; \lambda = 214nm) R\textsubscript{i} = 12.2min, 32.5\%B (Arg\textsuperscript{1}-Met\textsuperscript{11}); R\textsubscript{i} = 10.8min, 29.5\%B (Phe\textsuperscript{7}-Met\textsuperscript{11}); R\textsubscript{i} = 3.1min, 10.5\%B (Arg\textsuperscript{1}-Gln\textsuperscript{6}). m/z (MALDI-TOF) 1,346.9 (MH\textsuperscript{+} Arg\textsuperscript{1}-Met\textsuperscript{11}, 1,349.67), 610.98 (MH\textsuperscript{+} Phe\textsuperscript{7}-Met\textsuperscript{11}, 613.80), 753.28 (MH\textsuperscript{+} Arg\textsuperscript{1}-Gln\textsuperscript{6}, 753.88).

Chemical synthesis of SCD (H\textsubscript{2}N-Phe\textsuperscript{1}-Thr\textsuperscript{173}-OH)  
The synthesis was carried out on a 0.13mmol scale using Fmoc-Thr(Bu\textsuperscript{1}) functionalised Wang resin (850mg, 0.15mmol/g). The amino acids used had side-chain protection as previously described and were coupled via the corresponding HOCl activated ester. The first 153 amino acids were attached onto the solid support via single coupling cycles with double coupling cycles being employed for the last 20 residues (with the exception of Gly\textsuperscript{6}). Approximately one half of the resin was removed after 72 residues (Asp\textsuperscript{101}) which was stored in 1,4-dioxan until required. On
completion of the synthesis, the resin was washed with copious amounts of DMF, 1,4-dioxan, DCM then ether and dried under vacuum. Fmoc-SCD-resin was then stored in 1,4-dioxan at 5°C until required.

**Amino acid analysis** (64 hours hydrolysis) Ala_{13} 13.54, Arg_{7} 5.3, Asx_{21} 18.86, Glx_{13} 13.26, Gly_{13} 13.88, His_{8} 8.22, Ile_{8} 7.48, Leu_{16} 14.18, Lys_{6} 4.21, Met_{2} 2.41, Phe_{10} 9.65, Pro_{14} 12.94, Ser_{9} 6.14, Thr_{14} 8.73, Trp_{3} N/A, Tyr_{8} 6.42, Val_{8} 7.15.

**Deprotection and cleavage of SCD**

Prior to cleavage, the N^a-Fmoc protection was removed by sonicating resin (300mg) in 20% piperidine/DMF/1,4-dioxan (1:1, 10ml) for 15 minutes. On removal, the resin was filtered and washed using DMF, 1,4-dioxan, DCM then ether. The protein was then deprotected and cleaved from the solid support by stirring the resin in 95%aq. TFA (10ml) containing phenol (750mg), EDT (2ml), thioanisole (0.5ml) and TIPS (0.25ml). The mixture was stirred under nitrogen for 4 hours after which time the resin was filtered and washed with TFA (5ml). The filtrate was then added dropwise to ice-cold ether and the resulting precipitate centrifuged and washed with ether (3 times). The crude protein was dried, dissolved in 20% AcOH and lyophilised to yield crude SCD as a white solid (152mg).

**HPLC** (Aquapore C₄) Rt=18-20.5min, 54-59%B. **Amino acid analysis** (64 hours hydrolysis) Ala_{13} 13.03, Arg_{7} 4.27, Asx_{21} 18.18, Glx_{13} 14.00, Gly_{13} 12.97, His_{8} 7.6, Ile_{8} 6.60, Leu_{16} 13.04, Lys_{6} 3.83, Met_{2} 1.61, Phe_{10} 8.87, Pro_{14} 11.43, Ser_{9} 4.63, Thr_{14} 7.24, Trp_{3} N/A, Tyr_{8} 6.95, Val_{8} 6.89. **N-terminal sequencing** (predominant peaks): cycle 1 - Phe 67.55pmol; cycle 2 - Arg 19.35pmol; cycle 3 - Thr 46.94pmol.

**FPLC Size exclusion purification of SCD**

**Column conditions** - Gel filtration was carried out using Superdex™ 75 HR 26/60 column. Prior to protein purification, the column was equilibrated with 8M urea, 50mM Tris pH 7.5 buffer using a flow rate of 3ml/min. Following cleavage from the solid support, SCD (40mg) was dissolved in the 8M urea buffer (pH 7.5 - 1ml), loaded onto the column and a flow rate of 3ml/min applied.
5. Experimental

After 100ml, fractions (9ml) were collected, monitoring the progress at 280nm. Fractions absorbing at 280nm were examined using HPLC (Aquapore C4).

**HPLC** - Fractions 5-10 absorbed strongly at 280nm and when analysed using HPLC, eluted with the correct retention time (Rt=18-20min, 54-59%B), however fractions 6 and 7 gave the sharpest, cleanest HPLC profiles. As a result, fractions 8-10 were combined, concentrated (Mwt.CO 5,000) and re-eluted. All protein which eluted between fractions 6 and 7 was combined for subsequent analysis and purification.

UV concentration determination (section 5.3.1) est. 0.45mg/ml, thus 8mg of protein recovered from 40mg crude. **Amino acid analysis** (64 hours hydrolysis) Ala13 13.27, Arg7 6.2, Asx21 22.8, Glx13 14.9, Gly13 13.82, His8 8.6, Ile8 8.45, Leu16 14.25, Lys6 6.55, Met2 1.75, Phe10 11.10, Pro14 11.95, Ser9 8.5, Thr14 10.7, Trp3 N/A, Tyr8 8.2, Val8 9.1. **N-terminal sequencing** (predominant peaks): cycle 1 - Phe 47.67pmol; cycle 2 - Arg 15.85pmol; cycle 3 - Thr 16.49pmol.

**FPLC Molecular weight analysis of SCD**

**Column conditions** - Analysis was carried out using a Superdex™ 75 HR 10/30 column which was equilibrated with 8M urea, 50mM Tris pH 7.5 buffer using a flow rate of 0.5ml/min. Blue Dextrin (Mwt. 2,000,000) was loaded to measure the column void volume (V0 8.2).

0.5mg of protein was dissolved in 8M urea buffer (200μl), loaded onto the column and the elution volumes (Ve) recorded by monitoring at 280nm. The molecular weight standards used were ovalbumin (Mwt. 43,000, Ve 9.10±0.1) chymotrypsin (Mwt. 25,000, Ve 9.45±0.1) and ribonuclease (Mwt. 13,700, Ve 10.5±0.1). The elution volumes for the recombinant and chemically synthesised SCD were 9.70±0.1 and 9.80±0.1 respectively.

**Tryptic digest of SCD**

To a sample of SCD in 8M urea buffer (pH 7.5) (250μl, est. conc. 0.45mg/ml) was added trypsin (0.25mg dissolved in 250μl 50mM Tris pH 7.5). After mixing, the solution was incubated at 37°C, monitoring the progress of the digest using HPLC.
5. Experimental

(Vydac C8, 250x4.6mm, 5μm; 2ml loop, 1ml/min; 0-2min 10%B, 2-42min 10-70%B, 42-44min 70-90%; λ=214nm). After 5 hours, the solution was separated using analytical HPLC and a number of samples were analysed by mass spectrometry.

m/z (MALDI-TOF) 1,870.54 (MH⁺ Ala³⁸-Arg⁵², 1,775.97, TFA salt), ~2,160 (MH⁺ Ile¹⁹-Lys⁵⁰, 2,074.07, TFA salt), 2,344.15 (MH⁺ Leu¹⁵²-Thr¹⁷³, 2,343.10), ~2,400 (MH⁺ Ile¹⁹-Lys⁴⁰, 2,405.23), ~2,820 (MH⁺ Thr¹³-Lys¹⁷, 2,847.48), 3,155.02 (MH⁺ Val¹¹-Arg⁶⁷, 3,158.59), 4,254.08 (MH⁺ Glu⁶⁸-Lys¹⁰⁶, 4,256.86), ~5,050 (MH⁺ Asp¹⁰⁷-Arg¹⁵¹, 5,030.53), ~9,800 (MH⁺ Ile¹⁹-Lys¹⁰⁶, 9,801.74).

Isoelectric point (pI) determination using IEF

1. pH range 3-10 - As low salt conditions were necessary, SCD (10mg), dissolved in 4M urea (50ml) was initially desalted overnight using dialysis (4M urea, Mwt.CO 5,000). Ampholytes (Bio-Lyte®, pH range 3-10, 40% w/v, 1ml) were added and, following mixing, the protein solution was loaded into the focusing chamber of the Rotofor® cell. The cell was first allowed to rotate without applying the power so the system could reach thermal equilibrium at the cooling temperature of ~5°C. After 10 minutes, the power supply was connected, running at a constant power of 15W. A large increase in the voltage was initially observed which gradually stabilised over 3.5 hours. Once the voltage had completely levelled out, the 20 fractions were harvested and their individual pHs measured to ensure that a pH gradient had been established. All fractions were then examined by HPLC (Aquapore C8) to assess the protein content. The majority of protein was found in fractions 5 and 6.

<table>
<thead>
<tr>
<th>Fract.</th>
<th>pH</th>
<th>Fract.</th>
<th>pH</th>
<th>Fract.</th>
<th>pH</th>
<th>Fract.</th>
<th>pH</th>
<th>Fract.</th>
<th>pH</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2-2.5</td>
<td>5</td>
<td>4.5</td>
<td>9</td>
<td>6</td>
<td>13</td>
<td>7</td>
<td>17</td>
<td>8-8.5</td>
</tr>
<tr>
<td>2</td>
<td>2.5-3</td>
<td>6</td>
<td>5</td>
<td>10</td>
<td>6-6.5</td>
<td>14</td>
<td>7.5-7.5</td>
<td>18</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>3-3.5</td>
<td>7</td>
<td>5.25</td>
<td>11</td>
<td>6.5</td>
<td>15</td>
<td>7.5</td>
<td>19</td>
<td>9.5-10</td>
</tr>
<tr>
<td>4</td>
<td>3.5-4</td>
<td>8</td>
<td>5.5</td>
<td>12</td>
<td>7</td>
<td>16</td>
<td>7.5-8</td>
<td>20</td>
<td>10</td>
</tr>
</tbody>
</table>

2. pH range 4-6 - The same procedure as part 1 was applied using a second batch of SCD (10mg) in 4M urea (50ml). The ampholytes ranged from pH 4-6 to employ a narrower pH gradient. After 3 hours, the fractions were harvested and the individual
pH values measured before being analysed by HPLC. The majority of protein was found in fraction 5 indicating a pI of 4.5.

<table>
<thead>
<tr>
<th>Fract.</th>
<th>pH</th>
<th>Fract.</th>
<th>pH</th>
<th>Fract.</th>
<th>pH</th>
<th>Fract.</th>
<th>pH</th>
<th>Fract.</th>
<th>pH</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>5</td>
<td>4-4.5</td>
<td>9</td>
<td>4.5-5</td>
<td>13</td>
<td>5-5.5</td>
<td>17</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>6</td>
<td>4.5</td>
<td>10</td>
<td>4.5-5</td>
<td>14</td>
<td>5-5.5</td>
<td>18</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>3.5</td>
<td>7</td>
<td>4.5</td>
<td>11</td>
<td>5</td>
<td>15</td>
<td>5-5.5</td>
<td>19</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>8</td>
<td>4.5-5</td>
<td>12</td>
<td>5</td>
<td>16</td>
<td>5.5-6</td>
<td>20</td>
<td>10</td>
</tr>
</tbody>
</table>

**DEAE ion exchange purification of SCD**

**Adsorption onto matrix** - Following purification by FPLC size exclusion, SCD in 8M urea, 50mM Tris pH 7.5 (72ml, est. conc. 0.45mg/ml, 32.5mg), was mixed with DEAE Sepharose (10ml, previously washed with 8M urea, 50mM Tris pH 7.5). After shaking for 10 minutes, the mixture was centrifuged and the supernatant examined using HPLC (Aquapore C4) to confirm that all protein had adsorbed onto the gel media.

**Removal of urea** - Following adsorption, the supernatant was discarded and the gel loaded onto a column which was then run isocratically in 8M urea, 50mM Tris pH 7.5 (0.4ml/min) for 2 hours. A gradient of 8M urea, 50mM Tris pH 7.5 to 50mM Tris pH 7.5 containing 10mM CaCl2 and 0.1mM ZnCl2 (500ml) was introduced to gradually remove urea. This was run over 20 hours at a flow of 0.4ml/min monitoring at 226 and 277nm.

**Protein elution** - On removal of urea, the column was run isocratically of 50mM Tris pH 7.5, 10mM CaCl2 and 0.1mM ZnCl2 for 4 hours. Finally, a salt gradient of 0M-1M NaCl in 50mM Tris pH 7.5 containing 10mM CaCl2 and 0.1mM ZnCl2 (100ml, 1ml/min) was applied to elute the protein. 10ml fractions were collected and their protein content determined using HPLC (Aquapore C4). Protein containing fractions (5 and 6) were combined and repeatedly dialysed (3 times, Mwt.CO 10,000) against 50mM Tris pH 7.5, 10mM CaCl2, 0.1mM ZnCl2 to remove salt.

**UV concentration** (section 5.3.1) est. 1.07mg/ml, thus 21.5mg recovered from 32.5mg. **Amino acid analysis** - see section 2.8.3. **N-terminal sequencing** (predominant peaks): cycle 1 - Phe 21.30pmol; cycle 2 - Arg 6.34pmol; cycle 3 - 15.39pmol.
Denaturing and refolding of SCD

Denature - SCD in 50mM Tris pH 7.5, containing 10mM CaCl₂ and 0.1M ZnCl₂ (250μl) was gradually added to a solution of 8M urea, 50mM Tris pH 7.5 (2.5ml). After mixing, the solution was left to stand overnight at RT.

Refold - Following incubation, the urea solution containing protein was introduced, dropwise at RT, to folding buffer (15ml, 50mM Tris pH 7.5, 10mM CaCl₂, 0.1mM ZnCl₂) while ensuring the solution was thoroughly mixed during addition. Once complete, the solution was left overnight at 5°C after which time it was concentrated (Mwt.CO 3,000) to a final volume of ~250μl. The extent of the protein's activity was then measured using the Mca-peptide fluorescence assay (section 5.3.4).

<table>
<thead>
<tr>
<th></th>
<th>Initial conc. in solution</th>
<th>Calc. activity (units/mg/min)</th>
<th>Conc. after refolding</th>
<th>Calc. Activity (units/mg/min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recom.</td>
<td>0.028mg/ml</td>
<td>28.57</td>
<td>0.03mg</td>
<td>14.67</td>
</tr>
<tr>
<td>Synth.</td>
<td>0.37mg/ml</td>
<td>0.17</td>
<td>0.32mg</td>
<td>0.053</td>
</tr>
</tbody>
</table>

Fluorescence and CD analysis of chemically denatured SCD

Denaturing experiments were all carried out using ultrapure grade guanidine hydrochloride (Gdm.HCl), preparing a stock solution of 8M Gdm.HCl, 50mM Tris pH7.5. Stock solutions of recombinant (1.4mg/ml) and synthetic SCD (1mg/ml) were used. A number of samples were individually prepared and incubated at room temperature for 15 minutes prior to analysis.

<table>
<thead>
<tr>
<th>Gdm.HCl Conc.</th>
<th>SCD (recom or synth)</th>
<th>Buffer</th>
<th>Gdm.HCl</th>
</tr>
</thead>
<tbody>
<tr>
<td>0M</td>
<td>90μl</td>
<td>410μl</td>
<td>0μl</td>
</tr>
<tr>
<td>1M</td>
<td>90μl</td>
<td>347.5μl</td>
<td>62.5μl</td>
</tr>
<tr>
<td>2M</td>
<td>90μl</td>
<td>285μl</td>
<td>125μl</td>
</tr>
<tr>
<td>4M</td>
<td>90μl</td>
<td>160μl</td>
<td>250μl</td>
</tr>
</tbody>
</table>

Fluorescence measurements were made by exciting the sample at 290nm, monitoring the fluorescence emission between 300 and 400nm (Ex./Em. Slit width - 5nm). The wavelength and value of maximum fluorescence for each sample was recorded. The values for the fluorescence of the synthetic protein were adjusted, based on the concentration difference.
CD spectra were measured between 260-210nm at 20°C from which the molar ellipticity at 226nm was calculated. The values for synthetic SCD were adjusted based on the concentration difference.

<table>
<thead>
<tr>
<th>Gdm.HCl Conc.</th>
<th>Recom. ( \lambda_{\text{max}} ) (nm)</th>
<th>Recom. Fluorescence</th>
<th>Synth. ( \lambda_{\text{max}} ) (nm)</th>
<th>Synth. Fluorescence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0M</td>
<td>328</td>
<td>854.3</td>
<td>343</td>
<td>410.4</td>
</tr>
<tr>
<td>1M</td>
<td>327.5</td>
<td>687.0</td>
<td>346</td>
<td>308.3</td>
</tr>
<tr>
<td>2M</td>
<td>340</td>
<td>409.3</td>
<td>353.5</td>
<td>328.9</td>
</tr>
<tr>
<td>4M</td>
<td>354</td>
<td>310</td>
<td>355.5</td>
<td>357.8</td>
</tr>
</tbody>
</table>

**Mono-Q analytical ion exchange of SCD**

**Column conditions** - Analysis was carried out using a Mono Q® HR 5/5 column which was equilibrated with 50mM Tris pH 7.5 containing 10mM CaCl\(_2\) and 0.1mM ZnCl\(_2\).

Protein samples, dissolved in the 50mM Tris buffer were loaded and the column was run isocratically for 5 minutes (1ml/min). A salt gradient of 0-1M NaCl was then introduced over 30 minutes, monitoring the progress at 280nm. Fractions absorbing at 280nm were examined using HPLC (Aquapore C\(_4\)).

**Resource Phe analysis of SCD**

**Column conditions** - Analysis was carried out using a Resource Phe™ column which was equilibrated with 50mM Tris pH 7.5, 10mM CaCl\(_2\), 0.1mM ZnCl\(_2\) containing 0-XM (NH\(_4\))\(_2\)SO\(_4\) (a number of runs were attempted applying varying concentrations (X) of (NH\(_4\))\(_2\)SO\(_4\)).

Protein samples, dissolved in 50mM Tris buffer and containing XM (NH\(_4\))\(_2\)SO\(_4\) were loaded onto the column which was run isocratically for 5 minutes (1ml/min). A gradient of X-0M (NH\(_4\))\(_2\)SO\(_4\) was then introduced over 30 minutes, the progress monitored at 280nm.
Activity analysis of chemically denatured recombinant SCD

Activated recombinant SCD (10μl, conc. 0.6mg/ml) was incubated with various concentrations of urea, 50mM Tris pH 7.5 for 1 hour at room temperature. To each sample was then added Mca-peptide (10μl, 10⁻⁵M stock) and the change in fluorescence over 5 minutes was measured (Ex. 325nm/Em. 393nm).

<table>
<thead>
<tr>
<th>Conc. of urea</th>
<th>Change in fluorescence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0M</td>
<td>132.5-903.4 (154 units/min.)</td>
</tr>
<tr>
<td>1M</td>
<td>188.7-804.4 (123 units/min.)</td>
</tr>
<tr>
<td>2M</td>
<td>116.6-749.1 (126 units/min.)</td>
</tr>
<tr>
<td>4M</td>
<td>125.7-627.0 (100 units/min.)</td>
</tr>
<tr>
<td>6M</td>
<td>85.9-386.2 (60 units/min)</td>
</tr>
<tr>
<td>8M</td>
<td>62.8-73.7 (~2 units/min.)</td>
</tr>
</tbody>
</table>

5.4 Convergent Synthesis of SCD - Fragment Condensation

Notes

1. Solid support - The general procedure used for preparation of all Fmoc amino acid functionalised CI-Trityl resins is as described in section 5.2.2.2.

2. Fragment preparation - The syntheses and coupling of all the side-chain protected peptides used in the study were all very similar in that -
   - each synthesis employed single coupling cycles using HOCl.
   - all resins were washed and dried using the same solvent systems.
   - peptides were cleaved from the solid support using AcOH/TFE/DCM (1:1:8) and the work-up involved TFE/H₂O reprecipitation.
   - the couplings were carried out in the minimal quantity of DMF using an excess (3-5 equivalents) of the protected fragment.

As a result, the following procedures are brief, for detailed experimental information refer to the preparation of H₂N-Gly⁷⁹-Thr⁶³⁴-OH (26).
3. Coupling conditions - During coupling, the reaction mixture was maintained at a pH between 6 and 7. This was normally achieved by the addition of a molar equivalent of DIEA however, in some cases, further portions of base were necessary. In order to monitor the pH of the DMF solution, an aliquot was removed and spotted on moistened pH paper. This gave a good indication of the reaction conditions.

4. HPLC - Conditions used for the analysis of the majority of side-chain protected peptides were identical. The conditions used are described below -

- Vydac C18, 250x4.6mm, 5μm. or Aquapore C4, 100x4.6mm, 7μm.
- A=H2O, B=CH3CN, 0.1%TFA.
- 2ml loop, flow rate 1ml/min.
- 0-2min 10%B, 2-32min 10-90%B.
- λ=214 and 300nm (determine Fmoc content).

Preparation of $\text{H}_2\text{N-Pro}^{90}$-$\text{Thr}^{73}$-$\text{OH} (\text{Lys}^{106}$-$\text{(2-Cl-Z)})$ (26)

The synthesis was carried out on a 0.125mmol scale using Fmoc-Thr functionalised Wang resin (650mg, 0.2mmol/g). The amino acids were all coupled in single cycles using the HOCl method and their side-chain functionalities protected as previously described with the exception of Lys$^{106}$ where 1mmol of Fmoc-Lys(2-Cl-Z)-OH was used. The completed Fmoc-peptide-resin (1.65g, 0.047mmol/g) was treated with 20% piperidine/DMF/1,4-dioxan (1:1, 10ml) removing the Nα-Fmoc protection before it was filtered and washed with DMF, 1,4-dioxan, DCM then ether and dried in vacuo. The peptide-resin was then stored in 1,4-dioxan at 5°C until further use.

The peptide-resin (200mg) was deprotected and cleaved from the solid support by stirring the resin in 95% aq. TFA (10ml) containing thioanisole (0.5ml), EDT (2ml) and phenol (750mg). After 4 hours stirring under nitrogen, the resin was removed by filtration and the peptide precipitated by adding dropwise to ice-cold ether. The crude precipitate was washed in ether and dried under vacuum before being dissolved in 10% AcOH and lyophilised (76mg). $\text{H}_2\text{N-Pro}^{90}$-$\text{Thr}^{73}$-$\text{OH}$ was then redissolved in 10% AcOH (5mg/ml) and incubated overnight with N-methylmercapto-acetamide.
(MMA, 5 equivs., 1.7\mu l) at 37°C (26) was then purified by preparative HPLC (ABI aquapore RP300 C8, 250x10mm, 20\mu m; 5ml loop, 8ml/min; 0-2min 10%B, 2-42min 10-90%B; \lambda=214nm) R_t=31min, 69%B, 46mg.

HPLC (Vydac C8) R_t=21.4min, 62%B. m/z (MALDI-TOF) 9,455.68 (MH^+, C_{420}H_{619}N_{109}O_{137}S_{1}Cl requires 9,454.71), 9,481.99 (MNa^+ 9477.70), 9,499.77 (MK^+ 9,493.81). Amino acid analysis (30 hours hydrolysis) Ala5 5.54, Arg2 2.01, Asx_{13} 11.88, Gly_7 6.45, Glx_7 6.99, His_5 5.36, Ile_4 3.86, Leu_{10} 9.89, Lys_{1} 1.12, Met_{1} 0.89, Phe_4 4.15, Pro_6 6.20, Ser_6 4.41, Thr_8 6.31, Trp N/D, Tyr_3 2.89, Val_1 1.12.

Preparation of H_2N-Gly^{79}-Thr^{173}-OH (28)

1. FmocGly-Asn(Trt)-Val-Leu-Ala-His(Trt)-Ala-Tyr(Bu')-Ala-Pro-Gly-OH (27)
The synthesis was carried out on a 0.25mmol scale using Fmoc-Gly functionalised Cl-Trityl resin (650mg, 0.4mmol/g). The amino acids were all single coupled using the HOCl method with no extended coupling necessary. On completion of the synthesis, the Fmoc-peptide-resin (1.05g, 0.29mmol/g) was washed with DMF, 1,4-dioxan, DCM then ether. The resin was then stored under vacuum until use.

(27) (200mg, 0.29mmol/g) was cleaved from the solid support by sonicating the resin in AcOH/TFE/DCM (1:1:8, 10ml) at room temperature for 30 minutes. The resin was then filtered and washed with the same solvent system as used for coupling before the filtrate was concentrated to 3-5ml in vacuo. The crude peptide was then precipitated by triturating in ice-cold ether. In order to remove trace AcOH the dried, crude product was dissolved in TFE (0.5ml), reprecipitated (H_2O) and washed with copious amounts of H_2O and finally dried using ether to give the desired peptide as a white solid (79mg).

HPLC 1. Side-chain protection (s.c.p) intact (Vydac C_{18}) R_t=27.8min, 78.5%B; 2. s.c.p removed R_t=21.8min, 62.5%B. m/z (MALDI-TOF) 1,347.16 (MH^+-Trt 1,347.67); (FAB) 1,590 (MH^+ C_{85}H_{105}N_{14}O_{16} requires 1,589.78), 1,348 (MH^+-Trt). Amino acid analysis (24 hours hydrolysis) Ala_3 3.14, Asn_1 0.88, Gly_2 1.85, His_1 0.96, Leu_1 0.98, Pro_1 0.95, Tyr_1 0.95, Val_1 0.99.
5. Experimental

2. Coupling of FmocGly\textsuperscript{79}-Gly\textsuperscript{89}-OH (27) to H\textsubscript{2}N-Pro\textsuperscript{90}-Thr\textsuperscript{173}-resin

(27) (37mg, 23.5\textmu mol) was dissolved in DMF (200\textmu l) and mixed with HOCl (3.7mg, 23.5\textmu mol) dissolved in DMF (50\textmu l). The solution was then added to H\textsubscript{2}N-Pro\textsuperscript{90}-Thr\textsuperscript{173}-resin (100mg, 0.047mmol/g, 4.7\mu mol) and, after mixing, DIC (3.7\mu l, 23.5\textmu mol) was added and the mixture was sonicated at room temperature, with overnight stirring. DIEA (4.0\mu l, 23.5\textmu mol) was added to the reaction, dropwise over the first 15 minutes of coupling and the pH of the solution was monitored during and after addition. The peptide-resin was then filtered and washed with copious amounts of DMF, 1,4-dioxan, DCM then ether.

In order to monitor the coupling progress, Fmoc-peptide-resin (~5mg) was treated with 20\% piperidine/DMF/1,4-dioxan (1:1, 2ml - in order to avoid using large quantities of peptide-resin, 2ml of solution was used. Obviously the change in quantities was considered when the calculation was made) for 10 minutes. The coupling efficiency and resulting resin functionality was calculated as before (see section 5.2.3).

Coupling efficiency (based on Fmoc deprotection) 79%.

N-Terminal sequencing - estimated 86% coupling.

<table>
<thead>
<tr>
<th>Residue</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major</td>
<td>Gly 12.2\textmu mol</td>
<td>/</td>
<td>/</td>
<td>Leu 15.9\textmu mol</td>
<td>/</td>
</tr>
<tr>
<td>Minor</td>
<td>Pro 1.4\textmu mol</td>
<td>/</td>
<td>/</td>
<td>Asn 3.2\textmu mol</td>
<td>/</td>
</tr>
</tbody>
</table>

3. Purification of H\textsubscript{2}N-Gly\textsuperscript{79}-Thr\textsuperscript{173}-OH (28)

On completion of coupling, the N\textsuperscript{3}-Fmoc protection was removed and the resin (80mg, 0.034mmol/g) was treated with Tbfmoc-Cl (section 5.2.5). The peptide was then cleaved and deprotected by treating the resin with aq. TFA (95\%, 5ml), EDT (1ml) and thioanisole (0.25ml) for 4 hours under nitrogen. The resin was removed by filtration and the peptide precipitated from ice-cold ether and dried before the crude mixture (40mg) was dissolved in 6M Gdm.HCl/i-propanol (1:1, 5ml) and purified using Tbfmoc-carbon affinity purification (section 5.2.5, 30mg carbon). On removal of the Tbfmoc moiety and charcoal, the supernatant was concentrated \textit{in vacuo} to remove i-propanol and Gdm.HCl was then removed by dialysis (Mwt.CO 1,000,
against \( \text{H}_2\text{O} \) and lyophilised before (28) was isolated as a fluffy white solid after semi-preparative HPLC (Aquapore C8, 250x7.0mm, 7\( \mu \)m; 2ml loop, 2ml/min; 0-2min 10%B, 2-42min 10-90%B; \( \lambda =214\text{nm} \)) \( R_t=24\text{min}, 54\%\text{B}, (-3\text{mg}) \).

**HPLC (Aquapore C4)** \( R_t=19.5\text{min}, 56\%\text{B} \). \( m/z \) (MALDI-TOF) 10,598.8 (MH\(^+\) \( C_{468}H_{690}N_{123}O_{152}S_1Cl \) requires 10,500.44, TFA salt). **Amino acid analysis** (30 hours hydrolysis) Ala\(_8\) 7.89, Arg\(_2\) 2.04, Asx\(_{14}\) 12.17, Gly\(_9\) 9.14, Gln\(_7\) 7.60, His\(_6\) 6.19, Ile\(_4\) 4.07, Leu\(_{11}\) 11.14, Lys\(_{1}\) N/A, Met\(_1\) 0.86, Phe\(_4\) 4.31, Pro\(_7\) 7.23, Ser\(_6\) 5.83, Thr\(_8\) 7.32, Tyr\(_4\) 3.05, Val\(_2\) 2.14.

### Preparation of \( \text{H}_2\text{N-Gly}^{70}\text{-Thr}^{173}\)-resin

1. **Fmoc-Gly-Asp(Bu')-Phe-Tyr(Bu')-Pro-Phe-Asp(Bu')-Gly-Pro-Gly-Asn(Trt)-Val-Leu-Ala-His(Trt)-Ala-Tyr(Bu')-Ala-Pro-Gly-OH** (29)

The synthesis was carried out on a 0.22mmol scale, using Fmoc-Gly functionalised CI-Trityl resin (560mg, 0.4mmol/g). On completion, the Fmoc-peptide-resin (200mg, 0.18mmol/g) was treated with AcOH/TFE/DCM (1:1:8, 20ml) before the crude (29) was precipitated and dried to yield the title peptide as a white solid (97mg).

**HPLC** 1. side-chain protection (s.c.p) intact (Aquapore C4) \( R_t=32\text{min}, 90\%\text{B} \); 2. s.c.p removed (Vydac C18) \( R_t=25.2\text{min}, 71.8\%\text{B} \). \( m/z \) 1. s.c.p intact (FAB) 2,998 (MH\(^+\) 2,996.49), 2,774 (MH\(^+\)-Fmoc 2,773.42), 2,755 (MH\(^+\)-Trt 2,753.37), 2,532 (MH\(^+\)-Fmoc-Trt 2,530.30). 2. s.c.p removed (FAB) 2,288 (MH\(^+\) 2,287.01), (MALDI-TOF) 2,288.25 (MH\(^+\)), 2,064.94 (MH\(^+\)-Fmoc 2,064.95). **HRMS** (FAB) 2,996.49204 (MH\(^+\) \( C_{166}H_{201}N_{23}O_{30} \) requires 2,996.49087). **Amino acid analysis** (24 hours hydrolysis) Ala\(_3\) 2.90, Asx\(_3\) 3.06, Gly\(_4\) 4.10, His\(_1\) 0.97, Leu\(_1\) 0.85, Phe\(_2\) 1.70, Pro\(_3\) 2.71, Tyr\(_2\) 1.79, Val\(_1\) 0.98.

2. **Coupling of FmocGly\(^70\)-Gly\(^89\)-OH (29) to \( \text{H}_2\text{N-Pro}^{90}\text{-Thr}^{173}\)-resin**

(29) (31.7mg, 11.5\( \mu \)mol) was dissolved in DMF (100\( \mu \)l) and mixed with HOCl (1.8mg, 11.5\( \mu \)mol) dissolved in DMF (50\( \mu \)l). The solution was then added to \( \text{H}_2\text{N-Pro}^{90}\text{-Thr}^{173}\)-resin (50mg, 0.047mmol/g, 2.3\( \mu \)mol) and, after mixing, DIC (1.8\( \mu \)l, 11.5\( \mu \)mol) was added. DIEA (1.9\( \mu \)l, 11.5\( \mu \)mol) was gradually added over the first 15
minutes of coupling which was then sonicated at room temperature, stirring the reaction overnight.

Coupling efficiency (based on Fmoc deprotection) 14%.

N-Terminal sequencing - estimated 17% coupling.

<table>
<thead>
<tr>
<th>Residue</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major</td>
<td>Pro</td>
<td>/</td>
<td>Ile 3.6pmol</td>
<td>Asn</td>
<td>Gly</td>
</tr>
<tr>
<td>Minor</td>
<td>Gly</td>
<td>/</td>
<td>Phe 0.74pmol</td>
<td>Tyr</td>
<td>Pro</td>
</tr>
</tbody>
</table>

**Preparation of H$_2$N-Phe$^{50}$-Thr$^{173}$-resin**

1. FmocPhe-Ser(Bu')-Arg(Pmc)-Leu-Tyr(Bu')-Glu-(Bu')-Gly-Glu(Bu')-Ala-Asp(Bu')-Ile-Met-Ile-Ser(Bu')-Phe-Ala-Val-Arg(Pmc)-Glu(Bu')-His(Trt)-Gly-Asp(Bu')-Phe-Tyr(Bu')-Pro-Phe-Asp(Bu')-Gly-Pro-Gly-Asn(Trt)-Val-Leu-Ala-His(Trt)-Ala-Tyr(Bu')-Ala-Pro-Gly-OH (30)

The synthesis was carried out on a 0.11mmol scale using FmocGly$^{70}$-Gly$^{89}$-Cl-Trt-resin (600mg, 0.18mmol/g). On completion, Fmoc-peptide-resin (380mg, 0.12mmol/g) was treated with AcOH/TFE/DCM (1:1:8, 20ml), precipitated and dried to yield the title peptide as a white solid (231mg).

**HPLC** (Aquapore C$_4$) R$_t$=32min, 90%B. m/z No peaks were observed by either MALDI-TOF or FAB. **Amino acid analysis** (24 hours hydrolysis) Ala, 4.9, Arg, 1.32, Asx, 3.58, Glu, 2.45, Gly, 5.1, His, 2.05, Ile, 1.46, Leu, 1.72, Met, 0.6, Phe, 3.32, Pro, 3.35, Ser, 1.12, Tyr, 2.60, Val, 2.09.

2. **Coupling of FmocPhe$^{50}$-Gly$^{89}$-OH (30) to H$_2$N-Pro$^{90}$-Thr$^{173}$-resin**

(30) (155mg, 23.5μmol) was dissolved in DMF (1.0ml) and mixed with HOCl (3.7mg, 23.5μmol) as a solution in DMF (50μl). The solution was then added to H$_2$N-Pro$^{90}$-Thr$^{173}$-resin (100mg, 0.047mmol/g, 4.7μmol) and, after mixing, DIC (3.7μl, 23.5μmol) was added. At this point, a further 0.5ml of DMF was added (to dilute 'gel-like' solution) and, following addition of DIEA (4.0μl, 23.5μmol), the mixture was sonicated at room temperature, stirring the reaction overnight.
N-Terminal sequencing - Phe-Ser-Arg was not observed, hence coupling was deemed unsuccessful.

**Preparation of H$_2$N-Asn$^{80}$-Thr$^{173}$-resin**

1. FmocAsn(Trt)-Val-Leu-Ala-His(Trt)- Ala-Tyr(Bu')- Ala-Pro-Gly-OH (31)

The synthesis was carried out on a 0.26mmol scale using Fmoc-Gly functionalised Cl-Trityl resin (850mg, 0.3mmol/g). On completion, Fmoc-peptide-resin (400mg, 0.18mmol/g) was treated with AcOH/TFE/DCM (1:1:8, 10ml) before (31) was precipitated from ice-cold ether, reprecipitated (TFE/H$_2$O) and dried to yield the title peptide as a white solid (182mg).

HPLC 1. side-chain protection (s.c.p) intact (Vydac C$_{18}$) $R_t$=24min, 68.5%B; 2. s.c.p removed $R_t$=14.7min, 44%B. m/z 1. s.c.p intact (MALDI-TOF) 1,801.16 (MH$^+$ 1,775.88, Na$^+$ salt), 1,816.65 (K$^+$ salt); (FAB) 1,775 (MH$^+$), 1,533 (MH$^+$-Trt 1,531.75). 2. s.c.p removed (MALDI-TOF) 1,239.99 (MH$^+$ 1,235.60); (FAB) 1,235 (MH$^+$). HRMS (FAB) 1,775.87613 (MH$^+$ C$_{103}$H$_{116}$N$_{13}$O$_{15}$ requires 1,775.87915).

Amino acid analysis (24 hours hydrolysis) Ala: 2.89, Asx: 1.02, Gly: 1.13, His: 0.93, Leu: 1.12, Pro: 0.99, Tyr: 1.06, Val: 1.02.

2. Coupling of FmocAsn$^{80}$-Gly$^{89}$-OH (31) to H$_2$N-Pro$^{90}$-Thr$^{173}$-resin

(31) (41.3mg, 23.3µmol) was dissolved in DMF (150µl) and mixed with HOCl (3.6mg, 23.3µmol) as a solution in DMF (50µl). The solution was then added to H$_2$N-Pro$^{90}$-Thr$^{173}$-resin (75mg, 0.062mmol/g, 4.7µmol), and, after mixing, DIC (3.6µl, 23.3µmol) was added. DIEA (4.8µl, 27.9µmol) was gradually added over the first 15 minutes of coupling which was then sonicated at room temperature, stirring the reaction overnight.

Coupling efficiency (based on Fmoc deprotection) 73%.

N-Terminal sequencing - estimated 91% coupling.

<table>
<thead>
<tr>
<th>Residue</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major</td>
<td>Asn 610.51µmol</td>
<td>Val 887.88µmol</td>
<td>Leu 570.72µmol</td>
</tr>
<tr>
<td>Minor</td>
<td>Pro 66.56µmol</td>
<td>Gly 95.82µmol</td>
<td>Ile 39.34µmol</td>
</tr>
</tbody>
</table>
5. Experimental

Preparation of H$_2$N-Asp$^{71}$-Thr$^{73}$-resin

1. FmocAsp(Bu')-Phe-Tyr(Bu')-Pro-Phe-Asp(Bu')-Gly-Pro-Gly-OH (32)

Initial attempts to synthesise FmocAsp$^{71}$-Gly$^{79}$-OH failed, with a large drop in the coupling efficiency (70%) observed between Gly$^{77}$ and Pro$^{78}$. This has been observed in the synthesis of other peptides with Pro-Gly-OH situated at the C-terminus. It is postulated that, on removal of the Fmoc-Pro$^{78}$ N$^a$-protection, the resulting amine attacks the C-terminal carboxyl and releases the dipeptide from the solid support. In order to avoid this diketopiperazine formation from occurring, Gly$^{77}$-Pro$^{78}$ was separately prepared and coupled as the dipeptide to Gly$^{79}$-resin.

FmocGly$^{77}$-Pro$^{78}$-OH (350mg, 0.89mmol) and HOCt (139.7mg, 0.89mmol) was dissolved in DMF (2ml) before DIC (140μl, 0.89mmol) was added. Following 15 minutes sonication, the solution was added to Gly-Ci-Trityl-resin (850mg, 0.59mmol/g, 0.5mmol), mixed, and DIEA added dropwise (154μl, 0.89mmol). The mixture was then sonicated for 3 hours and the coupling efficiency measured using the Fmoc UV method (89% coupling, resin functionality 0.41mmol/g). The resin was then capped (section 5.2.4.1) and washed with DMF, 1,4-dioxan and DCM before being transferred to the reaction vessel ready for the coupling of the remaining amino acids. On completion of the synthesis, Fmoc-peptide-resin (1.1g, 0.25mmol/g) was treated with AcOH/TFE/DCM (1:1:8, 20ml) for 30 minutes before (32) was precipitated from ether, reprecipitated (TFE/H$_2$O) and dried to yield the title peptide as a white solid (387mg).

**HPLC**
1. side-chain protection (s.c.p) intact (Vydac C$_{18}$) R$_t$=21.7min, 62.5%B; 2. s.c.p removed Rt=16.8min, 49%B. m/z 1. s.c.p intact (FAB) 1,405 (MH$^+$ C$_{76}$H$_{94}$N$_{90}$O$_{17}$ requires 1,404.68), 1,389 (MH$^+$-H$_2$O 1,386.67), 1,354 (MH$^+$-Bu' 1,347.61), 1,297 (MH$^+$-2Bu' 1,290.54). 2. s.c.p removed (FAB) 1,250 (MH$^+$ 1,236.49).

**Amino acid analysis** (24 hours hydrolysis) Asx$_2$ 2.03, Gly$_2$ 1.09, Phe$_2$ 2.00, Pro$_2$ 1.06, Tyr$_1$ 1.10.

2. Coupling of FmocAsp$^{71}$-Gly$^{79}$-OH (32) to H$_2$N-Asn$^{80}$-Thr$^{73}$-resin

(32) (70.7mg, 50.4μmol) was dissolved in DMF (600μl) and mixed with HOCt (7.9mg, 50.4μmol) dissolved in DMF (100μl). The solution was then added to H$_2$N-
Asn\textsuperscript{80}-Thr\textsuperscript{73}-resin (400mg, 0.042μmol/g, 16.8μmol) and, after mixing, DIC (7.9μl, 50.4μmol) was added. DIEA (8.6μl, 50μmol) was gradually added over the first 15 minutes of coupling, monitoring the pH throughout, and the coupling mixture was then sonicated at room temperature, stirring the reaction overnight.

Coupling efficiency (based on Fmoc deprotection) 77%.

N-Terminal sequencing - estimated 72% coupling.

<table>
<thead>
<tr>
<th>Residue</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major</td>
<td>Asp 151.03pmol</td>
<td>Phe 119.86pmol</td>
<td>Tyr 75.18pmol</td>
</tr>
<tr>
<td>Minor</td>
<td>Asn 51.44pmol</td>
<td>Val 41.41pmol</td>
<td>Leu 37.14pmol</td>
</tr>
</tbody>
</table>

Preparation of H\textsubscript{2}N-Val\textsuperscript{66}-Thr\textsuperscript{73}-resin

1. FmocVal-Arg(Pmc)-Glu(Bu\textsuperscript{t})-His(Trt)-Gly-OH (33)

The synthesis was carried out on a 0.32mmol scale, using Fmoc-Gly functionalised CI-Trityl resin (700mg, 0.46mmmol/g). On completion, Fmoc-peptide-resin (500mg, 0.21mmol/g) was treated with AcOH/TFE/DCM (1:1:8, 10ml) before the crude peptide was precipitated from ether, reprecipitated and dried using ether to yield the title peptide as a white solid (217mg).

HPLC 1. side-chain protection (s.c.p) intact (Vydac C\textsubscript{18}) R\textsubscript{t}=22min, 63.5%B; 2. s.c.p removed R\textsubscript{t}=11min, 34%B. m/z 1. s.c.p intact (MALDI-TOF) 1,386.2 (MH\textsuperscript{+} 1,384.65), 1,139.59 (MH\textsuperscript{+}-Trt 1,140.53); (FAB) 1,385 (MH\textsuperscript{+}), 1,142 (MH\textsuperscript{+}-Trt), 1,118 (MH\textsuperscript{+}-Pmc 1,117.55), 876 (MH\textsuperscript{+}-Pmc-Trt 877.46). 2. s.c.p removed (MALDI-TOF) 817.88 (MH\textsuperscript{+} 819.38); (FAB) 819 (MH\textsuperscript{+}). HRMS (FAB) 1,384.64986 (MH\textsuperscript{+} C\textsubscript{76}H\textsubscript{91}N\textsubscript{10}O\textsubscript{13}S\textsubscript{1} requires 1,384.65214). Amino acid analysis (24 hours hydrolysis) Arg\textsubscript{1} 0.82, Glx\textsubscript{1} 1.11, Gly\textsubscript{1} 1.19, His\textsubscript{1} 0.87, Val\textsubscript{1} 0.95.

2. Coupling of FmocVal\textsuperscript{66}-Gly\textsuperscript{70}-OH (33) to H\textsubscript{2}H-Asp\textsuperscript{71}-Thr\textsuperscript{72}-OH (33) (12mg, 8.7μmol) was dissolved in DMF (100μl) and mixed with HOCt (1.37mg, 8.7μmol) dissolved in DMF (20μl). The solution was then added to H\textsubscript{2}N-Asp\textsuperscript{71}-Thr\textsuperscript{72}-resin (50mg, 0.035mmol/g, 1.75μmol) and, after mixing, DIC (1.4μl, 8.7μmol)
was added. DIEA (1.5μl, 8.7μmol) was the added to the coupling mixture which was then sonicated at room temperature, stirring the reaction overnight.

Coupling efficiency (based on Fmoc deprotection) 73%.

**N-Terminal sequencing** - estimated 93% coupling.

<table>
<thead>
<tr>
<th>Residue</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major</td>
<td>Val 638.71μmol</td>
<td>Arg 122.78μmol</td>
<td>Glu 171.51μmol</td>
</tr>
<tr>
<td>Minor</td>
<td>Asp 17.53μmol</td>
<td>Phe 20.33μmol</td>
<td>Tyr 9.74μmol</td>
</tr>
</tbody>
</table>

**Preparation of H$_2$N-Phe$^{64}$-Thr$^{73}$-resin**

1. **FmocPhe-Ala-Val-Arg(Pmc)-Glu(Bu')-His(Trt)-Gly-OH (34)**

The synthesis was carried out on a 0.3mmol scale, using Fmoc-Gly functionalised Cl-Trityl-resin (594mg, 0.52mmol/g). On completion, Fmoc-peptide-resin (820mg, 0.16mmol/g) was treated with AcOH/TFE/DCM (1:1:8, 20ml) before (34) was precipitated from ether, reprecipitated (TFE/H$_2$O) and dried to yield the title peptide as a white solid (375mg).

**HPLC**
1. side-chain protection (s.c.p) intact (Vydac C$18$) $R_t$=22.5%, 65%B; 2. s.c.p removed $R_t$=14min, 42%B. m/z 1. s.c.p intact (MALDI-TOF) 1,621.85 (MH$^+$+H$_2$O 1,620.77), 1,354.71 (MH$^+$-Trt 1,359.64), 1,335.52 (MH$^+$-Pmc 1,335.66); (FAB) 1,603 (MH$^+$), 1,361 (MH$^+$-Trt), 1,336 (MH$^+$-Pmc), 1,094 (MH$^+$-Pmc-Trt 1,092.54). 2. s.c.p removed (MALDI-TOF) 1,037.8 (MH$^+$ 1,037.48); (FAB) 1,038 (MH$^+$).

**HRMS** (FAB) 1,602.75433 (MH$^+$ C$_{88}$H$_{106}$N$_{12}$O$_{15}$S$_1$ requires 1,602.76213). **Amino acid analysis** (24 hours hydrolysis) Ala 1.08, Arg i N/D, G1x 1.20, Gly 1.16, His 0.83, Phe 0.93, Val 0.99.

2. **Coupling of FmocPhe$^{64}$-Gly$^{70}$-OH (34) to H$_2$N-Asp$^{71}$-Thr$^{73}$-resin**

(34) (14mg, 8.7μmol) was dissolved in DMF (100μl) and mixed with HOCl (1.36mg, 8.7μmol) dissolved in DMF (20μl). The solution was then added to H$_2$N-Asp$^{71}$-Thr$^{73}$-resin (50mg, 0.035mmol/g, 1.7μmol) and, after mixing, DIC (1.4μl, 8.7μmol) was added. DIEA (1.5μl, 8.7μmol) was then added and the coupling mixture was sonicated at room temperature, stirring the reaction overnight.
Coupling efficiency (based on Fmoc deprotection) 35%.

N-Terminal sequencing - estimated 66% coupling.

<table>
<thead>
<tr>
<th>Residue</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major</td>
<td>Phe 38.32pmol</td>
<td>Ala 25.65pmol</td>
<td>Val 15.58pmol</td>
</tr>
<tr>
<td>Minor</td>
<td>Asp 17.56pmol</td>
<td>Phe 15.96pmol</td>
<td>Tyr 7.71pmol</td>
</tr>
</tbody>
</table>

**Preparation of H$_2$N-Ile$^{62}$-Thr$^{73}$-resin**

1. **Fmocle-Ser(Bu')-Phe-Ala-OH (35)**

The synthesis was carried out on a 0.36mmol scale, using Fmoc-Ala functionalised Cl-Trityl resin (850mg, 0.42mmol/g). On completion, Fmoc-peptide-resin (1.08g) was treated with AcOH/TFE/DCM (1:1:8, 20ml), precipitated from ether, reprecipitated (TFE/H$_2$O) and dried to yield the title peptide as a white solid (265mg).

**HPLC**

1. side-chain protection (s.c.p) intact (Vydac C$_8$, 250x4.6mm, 5μm; 1ml/min, 2ml loop; 0-2min 10%B, 2-32min 10-90%B; λ=214&300nm) R$_t$=25.5min, 72.5%B; 2. s.c.p removed R$_t$=18.5min, 53.5%. m/z 1. s.c.p intact (FAB) 715 (MH$^+$ 715.37), 493 (MH$^+$-Fmoc 492.29), 437 (MH$^+$-Bu'-Fmoc 435.22) . 2. s.c.p removed (MALDI-TOF) 659.97 (MH$^+$ 659.31); (FAB) 659 (MH$^+$). **HRMS** (FAB) 715.37025 (MH$^+$ C$_{40}$H$_{51}$N$_{10}$ requires 715.37069). **Amino acid analysis** (8 hours hydrolysis) Ala 1.10, Ile 0.30, Phe 1.00, Ser 0.80.

2. **Coupling of Fmocle$^{62}$-Ala$^{65}$-OH (35) to H$_2$N-Val$^{66}$-Thr$^{73}$-resin**

(35) (7.5mg, 10.5μmol) was dissolved in DMF (100μl) and mixed with HOCt (1.65mg, 10.5μmol) dissolved in DMF (50μl). The solution was then added to H$_2$N-Val$^{66}$-Thr$^{73}$-resin (50mg, 0.021mmol/g, 1.05μmol) and, after mixing DIC (1.7μl, 10.5μmol) was added. DIEA (1.8μl, 10.5μmol) was then added and the coupling mixture was sonicated at 0°C.

Coupling efficiency (based on Fmoc deprotection) 93%.

N-Terminal sequencing - estimated 89%.

<table>
<thead>
<tr>
<th>Residue</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major</td>
<td>Ile 124.95pmol</td>
<td>Ser 82.12pmol</td>
<td>Phe 87.03pmol</td>
</tr>
<tr>
<td>Minor</td>
<td>Val 36.04pmol</td>
<td>Arg 2.90pmol</td>
<td>Glu 6.95pmol</td>
</tr>
</tbody>
</table>
5.5 Convergent Synthesis of SCD - Azide Condensation

Notes
1. HPLC - Conditions employed for the analysis of the peptide purifications and couplings are described below -
   • Vydac C_8, 250x4.6mm, 5μm.
   • A=H_2O, B=CH_3CN, 0.1%TFA.
   • 2ml loop, 1ml/min.
   • 0-2min 10%B, 2-32min 10-90%B.
   • λ=214 and 300nm (determine Fmoc content).

2. Preparation of C-terminal hydrazide - The C-terminal hydrazide functionality was incorporated into the peptide using the versatile dibenzocycloheptadiene linker (section 5.6.2). Modification of the linker is described below:

Dibenzo[a,d]cycloheptadien-5-Boc-hydrazide-(2-oxy-methylpolystyrene) resin (41)
t-Butyloxycarbonyl carbonate (800mg, 6.06mmol), dissolved in DCM (75ml), was added to dibenzo[a,d]cycloheptadien-5-ol-(2-oxy-methylpolystyrene) resin (750mg, 1.23mmol/g, 0.92mmol). Following addition of benzenesulfonic acid (155mg, 0.97mmol), the reaction was refluxed, while stirring under nitrogen, for 24 hours. The resin was then filtered, washed using DCM, DMF, DCM then ether and dried in vacuo (1.02g).

IR : (KBr disc) ν_{max} = 3419 (NH), 3100-2950 (CH arom.), 2900-2750 (CH_2), 1736 (C=O, amide), 1602, 1493cm^{-1} (C=C).

Dibenzo[a,d]cycloheptadien-5-Fmoc-Gly-hydrazide-(2-oxy-methylpolystyrene) resin (42)
To a solution of Fmoc-Gly-OH (2.17g, 7.3mmol), dissolved in DCM (75ml), was added freshly distilled thionyl chloride (5ml, 73mmol) and DMF (500μl). The mixture
was stirred at room temperature until a clear solution was obtained (~30-60 minutes, indicating complete dissolution of all reactants) before the excess thionyl chloride was removed in vacuo. In order to remove trace quantities of SOCl₂, the resulting residue was washed in twice in DCM (30ml), the solvent in each case was removed in vacuo. The residue was then dissolved in DCM (10ml), added to (41) (1g, 0.92mmol), previously swollen in DCM (50ml) and pyridine (2.6ml), and the mixture stirred at room temperature for 1 hour under nitrogen. The resin was then filtered and washed using copious amounts of DCM, DMF, DCM then ether and dried in vacuo (1.46g, estimated functionality 0.61mmol/g).

IR: (KBr disc) ν_max = 3420-3270 (NH), 3100-2950 (CH arom.), 2900-2750 (CH₂), 1720 (C=O, amide), 1685 (CO₃ ester), 1602, 1502, 1449cm⁻¹ (C=C).

**Attempted preparation of H₂N-Trp⁴²-Thr¹⁷³-OH (azide coupling)**

1. **Preparation of H₂N-Pro⁹⁹-Thr¹⁷³-OH (Lys¹⁰⁶-(2-Cl-Z)) (26)**
The synthesis, purification and characterisation for (26) is described in section 5.4.

2. **Fmoc Trp⁴²-Gly⁸⁹-NHNH₂ (43)**
The synthesis was carried out on a 0.25mmol scale using Fmoc-Gly functionalised hydrazide resin (42) (450mg, 0.61mmol/g). All amino acids were coupled in single cycles via the corresponding HBTU activated esters with the side-chain functionalities protected as previously described. On completion, the resin (930mg) was washed using DMF, DCM then ether. Fmoc-peptide-resin was then stored in 1,4-dioxan at 5°C until further use.

(43) (490mg) was deprotected and cleaved from the solid support by stirring the resin in 95%aq. TFA (10ml) containing thioanisole (0.5ml) and EDT (0.5ml) for 3 hours under nitrogen. The resin was then removed by filtration and the peptide precipitated by adding to ice-cold ether. The crude peptide was then washed in ether, dried, dissolved in 10% AcOH and lyophilised to yield the title peptide as a fluffy white solid (290mg). (43) was then isolated using preparative HPLC (ABI Aquapore C18,
5. Experimental

250x10mm, 20µm; 5ml loop, 8ml/min; 0-2min 30%B, 2-32min 30-80%B; λ=214&300nm) Rᵦ=24.8min, 68%B, 32.5mg.

**HPLC** (Vydac C₈) Rᵦ=22.5min, 64.7%B. m/z (MALDI-TOF) 5,606.07 (MH⁺\(\text{C}_{263}\text{H}_{365}\text{N}_{62}\text{O}_{74}\text{S}_{1}\)) requires 5,607.64). **Amino acid analysis** (30 hours hydrolysis) Ala₅ 4.86, Arg₂ 2.07, Asx₄ 4.20, G1x₅ 4.93, G1y₅ 5.14, His₂ 2.39, Ile₂ 1.81, Leu₃ 3.00, Met₁ 0.3, Phe₄ 3.84, Pro₄ 4.47, Ser₂ 1.97, Thr₂ 1.83, Trp₁ N/D, Tyr₃ 1.44, Val₃ 2.72.

3. Coupling of FmocTrp⁴²-Gly⁸⁹-NHNH₂ (43) and H₂N-Pro⁹⁰-Thr¹⁷₃-OH (26)

**Azide formation** - (43) (5mg, 0.89µmol), dissolved in DMF (100µl), was cooled to -20°C in an acetone/dry-ice bath. 4M HCl in 1,4-dioxan (0.62µl, 2.67µmol - 6µl of 4M HCl was diluted to 100µl using 1,4-dioxan, 10µl of this was used) was then added and the solution was warmed to -15°C. 1,4-dioxan (0.15µl, 0.89µmol - 15µl of 1,4-dioxan diluted to 1ml DMF, 10µl used) was then introduced, the temperature was raised to -10°C and left stirring for 10 minutes, allowing the formation of the azide. HPLC analysis was used to confirm that complete conversion had occurred. **HPLC** (Vydac C₈) Rᵦ=22.9min, 65.7%B (azide).

**Coupling** - Following formation of the azide, (26) (5mg, 0.52µmol), dissolved in DMF (100µl) and cooled to -10°C, was carefully added dropwise to the azide solution. A stock solution of DIEA (15µl in 1ml DMF) was prepared from which 5µl was added to the reaction every 30 minutes over the first 3 hours (2.67µmol of DIEA added). A second portion of azide and base (2.5mg hydrazide was converted using the above conditions) were added after 3 hours, using the same procedure, and the temperature of the reaction was maintained at 0°C overnight. After 24 hours coupling, a third portion of azide and base (2.5mg hydrazide, 1.3µmol DIEA) were added and the mixture was incubated at 0°C for a further 24 hours before the coupling mixture was analysed.
5.6 Chemical Ligation

Notes

1. HPLC - analysis of all the peptides in this section employed a C\textsubscript{18} column with one of 3 gradient systems -
   - Vydac C\textsubscript{18}, 250x4.6mm, 5\textmu m.
   - A=H\textsubscript{2}O, B=CH\textsubscript{3}CN, 0.1\%TFA.
   - 2ml loop, flow rate 1ml/min, \lambda=214nm.
     1. 0-10min 10\%B, 10-40min 10-90\%B.
     2. 0-5min 10\%B, 5-35min 10-90\%B.
     3. 0-2min 0\%B, 2-32min 0-60\%B, 32-34min 60-90\%B.

2. General procedure for the bromo/iodoacetylation of peptides\textsuperscript{5}

1. Preparation of anhydride - Bromo/iodoacetic acid (2mmol) in DCM (5ml), treated with DIC (1mmol), was stirred for 15 minutes at room temperature. During this time a white precipitate formed (N,N-diisopropyl urea) which was removed before the filtrate was evaporated to approximately 2ml by bubbling the solution with nitrogen gas. The volume was then adjusted to 6ml using DMF before the remaining DCM was removed.

2. Acetylation - To a solution of peptide-resin (Fmoc removed) in ME (5ml) was added the freshly prepared solution containing bromo/iodoacetic anhydride (1mmol). The mixture was then sonicated for 1 hour at room temperature before the resin was filtered and washed with copious amounts of DMF, DCM then ether and dried \textit{in vacuo}.

3. Ninhydrin Assay - In order to confirm that the acetylation was 100\% complete, a sample of resin was assayed for the presence of free amines.

To 1mg of resin bound peptide was added -
   1. Phenol/aqueous ethanol 76\% w/w (75\mu l).
   2. 0.0002M potassium cyanide/pyridine (100\mu l).
   3. 0.28M ninhydrin/ethanol (75\mu l).
This mixture was then heated for 5 minutes (100-105°C). Blue colouration indicated the presence of free amines.

3. C-terminal modified peptides - All peptides prepared containing the C-terminal thiol functionality were synthesised using the modified dibenzocycloheptadiene linker, the synthesis of which is described below:

S-[2(Triphenylmethyl)thio]ethylamine (65)

2-Mercaptoethylamine hydrochloride (5g, 44mmol) and triphenylmethanol (11.5g, 44.5mmol) were stirred in TFA (40ml) at room temperature for 30 minutes before the mixture was evaporated to a brown oil. Trituration of the oil with ether (250ml) gave the trifluoracetic salt as a white precipitate that was filtered and washed with ether. The washing were cooled to give a second crop. The salt was then separated between NaOH (1M, 250ml) and ether (250ml). Evaporation of the organic layer in vacuo followed by recrystallisation (ether/hexane) yielded the product as a white solid (8.35g, 26.15mmol, 59%).

TLC Rf 0.56-0.70 (n-butanol/AcOH/H₂O, 3:1:1). m/z (FAB) 320 (MH⁺ 320.15). HRMS (FAB) 320.14647 (C₂₁H₂₁N₁S₁ requires 320.14730). MP 94-96°C (lit⁶ 93-94°C). ¹H NMR (CDCl₃, 200MHz): δ 1.31 (s, 2H, NH₂), 2.31 (t, 2H, J=6.4Hz, CH₂S), 2.58 (t, 2H, J=5.7Hz, NH₂CH₂), 7.15-7.5 (m, 15H, arom.). IR: νmax = 3305 (NH), 3055-3000 (CH, arom.), 2970-2850 (CH₂), 1488, 1443cm⁻¹ (CH).

S-[2(Triphenylmethyl)thio]N-fluorenlymethoxycarbonyl-ethylamine (66)

A solution of (65) (3g, 9.39mmol) in triethylamine (80ml) was cooled in an ice bath before the addition of Fmoc-OSu (3g, 9.3mmol) in 1,4-dioxan (80ml). The reaction mixture was stirred for 2 hours at 0°C then overnight at room temperature after which time any precipitate present was removed before the solution was concentrated in vacuo. The resulting yellow oil was then dissolved in DCM (150ml), washed with H₂O (2x50ml) and dried (MgSO₄) before solvent was removed to yield the required compound as a yellow oil which was used without any further purification.
TLC Rf 0.68-0.75 (DCM/MeOH, 95:0.5). m/z (FAB) 564 (MH+ 542.22, Na+ salt). HRMS (FAB) 542.21468 (C36H32N1O2S1 requires 542.21538). 1H NMR (CDCl3, 200MHz): δ 2.41 (t, 2H, J=6.5Hz, CH2), 3.01 (dd, 2H, J=12.4 and 6.2Hz, NHCH2), 4.20 (t, 1H, J=6.8Hz, CH), 4.35 (d, 2H, J=6.7Hz, CH2), 4.80 (t, 1H, J=5.7Hz, NH), 7.16-7.35 (m, 15H, Trt arom. CH), 7.35-7.77 (m., 8H, Fmoc arom. CH). IR: νmax = 3431, 3337 (NH), 3050-3000 (CH, arom.), 2950-2820 (CH2), 1715 (C=O), 1509 (C=C), 1244,1142cm⁻¹ (C-O).

N-Fluorenylmethoxycarbonylaminoethyl-2-thiol (62)

Crude (66) (3g, 5.53mmol) was added to a solution containing TFA/H2O (9:1, 20ml) and TIPS (2.26ml, 11.08mmol) and the reaction mixture was stirred at room temperature for 1 hour. The white precipitate formed was removed and the filtrate concentrated in vacuo. The resulting yellow oil was dissolved in ether (100ml), washed (H2O, 2x100ml), the organic layer dried (MgSO4) and concentrated in vacuo to yield the crude product. The product was then purified by wet flash chromatography eluting with 2% MeOH in DCM (680mg, 2.23mmol, 40%).

TLC Rf 0.55-0.66 (DCM/MeOH, 9.5:0.5). m/z (FAB) 300 (MH+ 300.11). HRMS (FAB) 300.10618 (C17H18N1O2S1 requires 300.10583). 1H NMR (CDCl3, 200MHz): δ 2.64 (dd, 2H, J=14.5 and 6.6Hz, CH2), 3.35 (dd, 2H, J=12.5 and 6.5Hz, NHCH2), 4.21 (t, 1H, J=6.7Hz, CH), 4.44 (d, 2H, J=6.7Hz, CH2), 5.24 (b, 1H, NH), 7.25-7.7 (m, 8H, Fmoc arom. CH). IR: νmax = 3430, 3348 (NH), 2941 (CH), 1707 (C=O), 1513 (C=C), 1249,1144cm⁻¹ (C-O).

Dibenzo[a,d]cycloheptadiene-5-one-(2-oxy-methylpolystyrene) resin (69)

To a solution containing 2-hydroxydibenzo[a,d]cycloheptadien-5-one (2.0g, 8.92mmol), dissolved in t-butanol/H2O (1:1, 80ml), was added caesium hydroxide (1.5g, 8.8mmol). After stirring for 10 minutes, t-butanol was removed in vacuo. The caesium salt was dried by azeotropic distillation using pyridine (2x80ml) followed by DMF (2x80ml). The salt was then dissolved in DMF (30ml) and added to chloromethylpolystyrene resin (CMP) (1.5g, 1.6mmol/g, 2.4mmol) previously swollen in DMF (20ml). The reaction mixture was stirred mechanically under nitrogen at 60°C.
for 4 days. After this time, the resin was filtered off, washed with copious amounts of DMF, i-propanol, H₂O, DMF and finally i-propanol and dried in vacuo to yield (69) as an off-white solid (1.95g, 100%).

**IR:** (KBr disc) νₘₐₓ = 3100-3000 (CH, arom.), 3000-2850 (CH₂), 1635 (C=O), 1603, 1582, 1492 (C=C), 1445 cm⁻¹ (CH₂). C.I analysis - found less than 0.55%Cl (CMP 4.2%). est. functionality 1.108mmol/g.

**Dibenzo[a,d]cycloheptadien-5-ol-(2-oxy-methylpolystyrene) resin (61)**

Linker-resin (69) (0.5g, 1.108mmol/g, 0.55mmol) was added to dry THF (30ml) along with LiBH₄ (100mg, 5.58mmol) and the mixture was heated, while stirring, at reflux under nitrogen for 1 hour. The resin was then filtered, washed with THF, MeOH, H₂O, MeOH and ether and dried in vacuo to yield (61) as white solid (0.48g).

**IR:** (KBr disc) νₘₐₓ = 3566, 3404 (OH), 3100-2980 (CH, arom.), 2980-2840 (CH₂), 1602, 1581, 1491 (C=C), 1445 cm⁻¹ (CH₂).

**Dibenzo[a,d]cycloheptadien-5-(Fmoc-aminoethy-l-2'-thio)-(2-oxyethyl-polystyrene) resin (63)**

Benzenesulfonic acid (78mg, 0.49mmol) was added to (62) (1g, 3.34mmol) along with (61) (0.45g, 1.108mmol/g, 0.49mmol) previously swollen in DCM (25ml) and the reaction mixture was stirred for 1 hour, at room temperature, under an atmosphere of nitrogen. The resin was then filtered and washed with DCM and ether and dried in vacuo yielding (63) as a pale yellow solid (0.63g, estimated functionality 0.77mmol/g).

**IR:** (KBr disc) νₘₐₓ = 3412, 3333 (NH), 3075-2980 (CH, arom.), 2980-2840 (CH₂), 1703 (C=O), 1602, 1488 (C=C), 1446 cm⁻¹ (CH₂).

**H₂N-Gly-Phe-Ala-NH-(CH₂)₂-SH (70)**

The peptide was synthesised using the thiol derivatised cycloheptadiene linker (63) (300mg, 0.89mmol/g, 0.27mmol). This peptide was synthesised manually using the nitrogen bubbler apparatus, however the sequence of events did not differ from peptides synthesised on the machine. Each amino acid was coupled via the activated
HOBt ester (coupling time 1.5 hours) and the coupling efficiency was monitored using Fmoc UV analysis (section 5.2.3). On completion of the synthesis, the peptide was deprotected and cleaved from the solid support by stirring the resin in 95%aq. TFA (10ml) containing thioanisole (0.5ml) for 1 hour. On removal of the resin by filtration, ether precipitation yielded the crude peptide as a white solid.

HPLC (gradient system 3) Rₜ=13.5 min, 23%B; Rₜ=23.5 min, 43%B (dimer). m/z (MALDI-TOF) 353.72 (M⁺ 353.16). HRMS (FAB) 353.16367 (MH⁺ C₁₆H₂₃N₄O₃ requires 353.16474). Amino acid analysis (8 hours hydrolysis) Ala₁ 1.04, Gly₁ 1.04, Phe₁ 0.92.

Preparation of H₂N-Thr¹-Asp-Glu-Thr-Leu-His-Leu-Val-NH-(CH₂)₂-S-CH₂-CO-Leu-Arg-Leu-Arg-Gly-Gly¹⁶-OH (74)

1. TbfmocThr¹-Asp-Glu-Thr-Leu-His-Leu-Val⁶-NH-(CH₂)₂-SH (73)

The synthesis was carried out on a 0.25mmol scale using thiol derived resin (63) (320mg, 0.78mmol/g). The amino acids were all coupled in single cycles using HOBt activated esters. On completion, Fmoc-peptide-resin (573mg, 0.32mmol/g) was washed with DMF then DCM before the N⁰-Fmoc protection was removed (20% piperidine) and the resin treated with Tbfmoc-Cl (section 5.2.5). The peptide was then deprotected and cleaved from the solid support by stirring resin (200mg) in 95%aq. TFA (5ml) containing thioanisole (0.25ml) and EDT (0.25ml) for 1 hour. On ether precipitation, the crude peptide was yielded as a white solid (84mg).

HPLC (gradient system 1) Rₜ=38.5 min, 86%B. m/z (MALDI-TOF) 1,411.7 (M⁺ 1,408.63); (FAB) 1,410 (M⁺). HRMS (FAB) 1,408.62776 (MH⁺ C₇₃H₹₀N₁₁O₁₆S₁ requires 1,408.62877). Amino acid analysis (24 hours hydrolysis) Asp₁ 1.10, Glu₁ 1.02, His₁ 1.01, Leu₂ 2.18, Thr₂ 1.93, Val₁ 0.96.


The synthesis was carried out on a 0.31mmol scale using Fmoc-Gly functionalised Wang resin (500mg, 0.63mmol/g). The peptide was assembled on the solid support using HOBr activated amino acids (single coupling). On completion of the synthesis,
5. Experimental

Nα-Fmoc protection was removed (20% piperidine) and the peptide was bromoacetylated (section 5.6.2). A sample of the peptide-resin (200mg) was cleaved and deprotected by stirring the resin in 95%aq. TFA (5ml) containing anisole (0.25ml) for 1 hour. The resin was then removed by filtration, the crude peptide precipitated using ether and dried in vacuo (78mg).

**HPLC (gradient system 1)** $R_t=23.2$min, 45.2%B. m/z (MALDI-TOF) 792.3 (MH$^+$ C$_{30}$H$_{56}$N$_{12}$O$_8$Br$_1$ requires 791.35). Prior to bromoacetylation (FAB) 671 (MH$^+$ 671.43). **HRMS** (FAB) 671.43166 (MH$^+$ C$_{28}$H$_{55}$N$_{12}$O$_7$ requires 671.43167). **Amino acid analysis** (20 hours hydrolysis) Arg$_2$ 2.18, Gly$_2$ 2.01, Leu$_2$ 2.10.

3. Ligation of TbfmocThr'-Val$^8$-SH (73) and Br-Leu$^{14}$-Gly$^{16}$-OH (72)

TbfmocThr'-Val$^8$-SH (10mg) was dissolved in TFE (2ml) before 6M Gdm.HCl (2ml), buffered at pH 7.0 (0.1M sodium phosphate), was added. To this solution was added Br-Leu$^{14}$-Gly$^{16}$-OH (10mg), previously dissolved in TFE/6M Gdm.HCl pH 7.0 (1:1. 2ml). The ligation mixture was then sonicated overnight at room temperature under nitrogen. The progress of the ligation was monitored throughout using HPLC analysis. On completion, the ligation product was isolated as a fluffy white solid after semi-preparative HPLC (Aquapore C$_8$, 250x7.0mm, 7µm; 5ml loop, 2ml/min; 0-5min 10%B, 5-10min 10-30%B, 10-30min 30-60%B, 30-32min 60-90%B; $\lambda=214$&365nm), $R_t=25.4$min, 53%B, 8.3mg.

**HPLC (gradient system 1)** $R_t=34.6$min, 75.6%B. m/z (MALDI-TOF) 2,121.3 (MH$^+$ C$_{103}$H$_{143}$N$_{23}$O$_{24}$S$_1$ requires 2,120.06). **Amino acid analysis** (40 hours hydrolysis) Arg$_2$ 2.46, Asp$_1$ 1.03, Glu$_1$ 1.03, Gly$_2$ 2.09, His$_1$ 1.02, Leu$_4$ 3.71, Thr$_2$ 1.94, Val$_1$ 0.91.

**Tbfmoc removal** - Tbfmoc-peptide (6mg), dissolved in 6M Gdm.HCl/i-propanol (1:1, 2ml), was mixed with piperidine (0.2ml) and sonicated for 15 minutes before the mixture was neutralised using AcOH. The product was then isolated using semi-preparative HPLC (same column as above, 0-5min 10%B, 5-35min 10-60%B, 35-37min 60-90%B) $R_t=15.8$min, 29%B, ~3mg.

m/z (MALDI-TOF) 1,701.9 (MH$^+$ C$_{72}$H$_{128}$N$_{23}$O$_{22}$S$_1$ requires 1,698.93).

155
Nucleophilic substitution of (72) and (75)

The preparation of the bromo (72) and iodoacetylated (75) peptides were identical, refer to the synthesis of Br-Leu^{11}-Gly^{16}-OH (72) for experimental details. Following deprotection and cleavage from the solid support, both peptides (10mg) were separately dissolved in 6M Gdm.HCl, 0.1M sodium phosphate pH 7/TFE (1:1, 2ml), before aminoethanethiol (10mg, 88.02μmol) was added to each solution. The coupling progress was then monitored using HPLC analysis.

HPLC (gradient system 2) R=20.2 min, 50.5%B (peptides (72) and (75)), R=19.1min, 47%B (peptide (76)).


1. TbfmocGlu$^1$-Ala-Glu*-Asp*-Leu*-Gln*-Val-Gly-Gln*-Val-Glu-Leu$^{12}$-NH-(CH$_2$)$_2$-SH (78)

The synthesis was carried out on a 0.25mmol scale using thiol derived resin (63) (320mg, 0.79mmol/g). All amino acids were coupled via the HOCT activated ester employing double coupling cycles where indicated (*). On completion, Fmoc-peptide-resin (593mg, 0.26mmol/g) was washed before the N$^\alpha$-Fmoc protection was removed (20% piperidine) and the resin treated with Tbfmoc-Cl (section 5.2.5). The peptide was then deprotected and cleaved from the solid support by treating the resin (150mg) with 95% aq. TFA (5ml) containing thioanisole (0.25ml) and EDT (0.25ml), stirring the solution for 2 hours. On precipitation, the crude peptide was yielded as a white solid (64mg).

HPLC (gradient system 2) R=32min, 82%B. m/z (FAB) 1,812 (MH$^+$ C$_{89}$H$_{117}$N$_{15}$O$_{24}$S$_1$ requires 1,811.81). Amino acid analysis (20 hours hydrolysis) Ala$_1$ 1.06, Asp$_1$ 0.92, Gln$_5$ 4.85, Gly$_1$ 1.15, Leu$_2$ 1.93, Val$_2$ 1.89.
5. Experimental

2. I-CH_2-CO-Gly^{12}-Pro-Gly-Ala-Gly-Ser-Leu-Gln-Pro-Leu-Ala-Leu-Glu-Gly-Ser-Leu-Gln^{31}-OH (79)

The synthesis was carried out on a 0.25mmol scale using Fmoc-Gln functionalised Wang resin (500mg, 0.52mmol/g). All amino acids were coupled via the HO_Ct activated ester employing double coupling cycles (except Gly - single coupling). On completion, Fmoc-peptide-resin (785mg, 0.18mmol/g) was washed before the N^α-Fmoc protection was removed (20% piperidine) and the peptide iodoacetylated (section 5.6.2). A sample of resin (150mg) was then cleaved and deprotected using 95%aq. TFA (5ml) containing anisole (0.5ml), stirring the mixture for 2 hours before the filtrate was isolated and the crude peptide precipitated using ice-cold ether (57mg).

HPLC (gradient system 2) R_t=20.4min, 51%B. m/z (MALDI-TOF) 1,760.33 (MH^+ C_71H_{119}N_{19}O_{25}I requires 1,762.75). Amino acid analysis (28 hours hydrolysis) Ala_2 1.85, Gln_3 2.93, Gly_4 4.24, Leu_4 4.07, Pro_2 2.31, Ser_2 1.91.

3. Ligation of TbfmocGlu^{1}-Leu^{12}-SH (78) and I-Gly^{15}-Gln^{31}-OH (79)

TbfmocGlu^{1}-Leu^{12}-SH (60mg), dissolved in TFE (4ml), was added to 6M Gdm.HCl buffered at pH 7.5 (4ml, 0.1M sodium phosphate). After mixing, I-Gly^{15}-Gly^{31}-OH (30mg) was added and the ligation solution was sonicated in an atmosphere of nitrogen overnight. The progress of the ligation was monitored throughout using HPLC analysis.

HPLC (gradient system 2) R_t=29.2, 74.5%B. m/z (MALDI-TOF) 3,478.4 (MH^+ C_{160}H_{233}N_{34}O_{49}S_1 requires 3,446.65).

Tbfmoc purification: 1. Adsorption - On completion, PGC (200mg) was added to ligation mixture. The solution was then stirred, monitoring the adsorption using HPLC (λ=365nm). Once all of the Tbfmoc was adsorbed, the mixture was centrifuged and the supernatant removed.

2. Wash (2 times) - PGC was washed using i-propanol/6M Gdm.HCl (1:1, 10ml). After mixing, the solution was centrifuged and the supernatant removed.
3. **Tbfmoc removal** - 10% piperidine in i-propanol/6M Gdm.HCl (1:1, 10ml) was added to the carbon and mixed vigorously for 15 minutes. Following centrifugation, the supernatant was collected and i-propanol removed in vacuo. The pH of the solution was then altered using AcOH (pH 6-7), before separating the ligation product (80) from the peptide fragment (78) using preparative HPLC (Aquapore C8, 250x7.0mm, 7μm; 2ml loop, 5ml/min; 0-5min 10%B, 5-35min 10-60%B, 35-37min 60-90%B; λ=214nm) Rf=19.4min, 34.0%B, 27mg (product), Rf=21.4min, 36.8%B, 21mg (fragment).

**HPLC** (gradient system 2) Rf 20.5min, 51.3%B (product); Rf 21.4min, 53.8%B (fragment). m/z (MALDI-TOF) 3,025.4 (MH+ C₁₂₉H₂₁₅N₃₄O₄₇S₁ requires 3,024.52 - product), 2,773.3 (MH+ 1,389.45 - dimer of fragment). **Amino acid analysis** (28 hours) Ala₃ 3.08, Asp₁ 1.12, Glx₈ 9.05, Gly₅ 4.92, Leu₆ 4.91, Pro₂ 1.96, Ser₂ 1.67, Val₂ 2.19.

**5.7 References**

Courses and Conferences Attended


NMR Spectroscopy, Drs I. Sadler and P. Barlow, University of Edinburgh, 1996.


"The Discovery of Agrochemicals", Drs C. Godfrey and T. Perrior (Zeneca Agrochemicals), University of Edinburgh, 1996.


Edinburgh Centre for Protein Technology (ECPT) lectures, University of Edinburgh, 1997, various speakers.
# The Naturally Occurring Amino Acids

![Amino Acid Structure](image)

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>3 letter code</th>
<th>1 letter code</th>
<th>R group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>Ala</td>
<td>A</td>
<td>-CH₃</td>
</tr>
<tr>
<td>Arginine</td>
<td>Arg</td>
<td>R</td>
<td>-(CH₂)₃NHC(=NH)NH₂</td>
</tr>
<tr>
<td>Asparagine</td>
<td>Asn</td>
<td>N</td>
<td>-CH₂CONH₂</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>Asp</td>
<td>D</td>
<td>-CH₂CO₂H</td>
</tr>
<tr>
<td>Cysteine</td>
<td>Cys</td>
<td>C</td>
<td>-CH₂SH</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>Glu</td>
<td>E</td>
<td>-(CH₂)₂CO₂NH₂</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Gln</td>
<td>Q</td>
<td>-(CH₂)₂CO₂H</td>
</tr>
<tr>
<td>Glycine</td>
<td>Gly</td>
<td>G</td>
<td>-H</td>
</tr>
<tr>
<td>Histidine</td>
<td>His</td>
<td>H</td>
<td>-CH₂(4-imidazolyl)</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>Ile</td>
<td>I</td>
<td>-CH(CH₃)CH₂CH₃</td>
</tr>
<tr>
<td>Leucine</td>
<td>Leu</td>
<td>L</td>
<td>-CH₂CH(CH₃)₂</td>
</tr>
<tr>
<td>Lysine</td>
<td>Lys</td>
<td>K</td>
<td>-(CH₂)₄NH₂</td>
</tr>
<tr>
<td>Methionine</td>
<td>Met</td>
<td>M</td>
<td>-(CH₂)₂SCH₃</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>Phe</td>
<td>F</td>
<td>-CH₃Ph</td>
</tr>
<tr>
<td>Proline</td>
<td>Pro</td>
<td>P</td>
<td>pyrrolidine-2L-carboxylic acid</td>
</tr>
<tr>
<td>Serine</td>
<td>Ser</td>
<td>S</td>
<td>-CH₂OH</td>
</tr>
<tr>
<td>Threonine</td>
<td>Thr</td>
<td>T</td>
<td>-CH(CH₃)OH</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>Trp</td>
<td>W</td>
<td>-CH₂(3-indolyl)</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Tyr</td>
<td>Y</td>
<td>-CH₂(4-hydroxyphenyl)</td>
</tr>
<tr>
<td>Valine</td>
<td>Val</td>
<td>V</td>
<td>-CH(CH₃)₂</td>
</tr>
</tbody>
</table>