Solid Phase Synthesis of Peptides and Proteins

by

Angus R. Brown

A Thesis submitted for the

degree of Doctor of Philosophy

University of Edinburgh

March 1995

for

Alice and Erin

ACKNOWLEDGEMENTS

I would like to thank Professor Robert Ramage FRS for the provision of research facilities and for his advice and constant encouragement throughout the course of this work and beyond.

Special thanks are due to Mr. Kevin Shaw for his very substantial input into the technical aspects of the peptide synthesis described herein and Mr. Brian Whigham for amino acid analysis and mass spectroscopy. I also thank those involved in the departmental analytical and spectroscopic services for their rapid and efficient work.

I wish to thank the Department of Chemistry (Ames Fund) and the Science and Engineering Research Council for the provision of funds.

Finally, I would like to thank all my friends and colleagues, both past and present, for their help and encouragement and also for making my time in LAB 29 so unforgettable and enjoyable.

This thesis is submitted in part fulfillment 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

March 1995

ABSTRACT

A strategy for the total chemical synthesis and purification of proteins has been investigated and applied to the 85 residue methylated DNA binding domain (MBD) from the chromosomal protein MeCP2, the 66 residue Restriction Alleviation (Ral) protein from bacteriophage λ and the 76 residue β-chemokine Monocyte Chemotactic protein (MCP-1). The hydrophobicity of the N^{α} protecting group tetrabenzo[a,c,g,i]fluorenyl-17-methoxycarbonyl (Tbfmoc) has been exploited to simplify the rapid purification of the 85 amino acid MBD protein by Hplc. Initial structural studies on the synthetic protein are also reported. In addition a comparative study of semi-permanent, temporary and enzyme cleavable thiol protection has resulted in the extention of this Tbfmoc methodology to the synthesis of cysteine containing proteins such as Ral and MCP-1.

A general route to C-terminal α-hydroxyglycine extended peptides via Fmoc/t-Bu based solid phase peptide synthesis is also described. Such peptides are the biosynthetic precursors of peptide amides in which the C-terminal carboxamide functionality is required for biological activity in a number of important hormones.

ABBREVIATIONS

A	adenosine
Aba	aminobutyric acid
Acm	acetamidomethyl
AcOH	acetic acid
α-AE	α-amidating enzyme
Boc	tert-butyloxycarbonyl
BOP	benzotriazole-1-yl-oxy-tris-(dimethylamino)-phosphonium hexafluorophosphate
tBu	tert-butyl
C	cytosine
CCK	cholecystokinin
CD	circular dichroism
CpG	deoxycytidyldeoxyguanosine dinucleotide
Da	dalton
DCCI	N,N'-dicyclohexylcarbodiimide
DIC	N,N'-diisopropylcarbodiimide
DIEA	N,N-diisopropylethylamine
DKP	diketopiperazine
DMF	N,N-dimethylformamide
DNA	deoxyribonucleic acid
DTNB	5,5'-dithiobis(2-nitrobenzoic acid)
DTT	dithiothreitol
EDT	ethane-1,2-dithiol
EDTA	ethylenediaminetetraacetic acid
FdC	5-fluoro-2'-deoxycytidine
Fmoc	9-fluorenylmethoxycarbonyl
G	guanosine
Gdm.HCl	guanidinium hydrochloride
HOBt	1-hydroxybenzotriazole
HPLC	high performance liquid chromatography
IL	interleukin
MBH	4,4'-dimethoxybenzhydryl
MBD	methylated DNA binding domain
MALDITOF	matrix assisted laser desorption ionisation time of flight
MCP-1	monocyte chemotactic peptide

MEM	2-methoxyethoxymethyl
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser enhancement
NOESY	nuclear Overhauser correlation spectroscopy
αOHG	α-hydroxyglycine
PAL	peptidylamidoglycolate lyase
Pfp	pentafluorophenyl
PGC	porous graphitised carbon
PHM	peptidyl-α-hydroxyglycine monooxygenase
pI	isoelectic point
Pmc	2,2,5,7,8-pentamethylchroman-6-sulphonyl
PSA	preformed symmetrical anhydride
PyBOP	benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate
RP-HPLC	reverse phase HPLC
Ral	restriction alleviation
Rt	retention time
SAM	S-adenosyl-L-methionine
SC	salmon calcitonin
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SKB	SmithKline Beecham
SPPS	solid phase peptide synthesis
T	thymidine
Tacm	trimethylacetamidomethyl
Tbfmoc	tetrabenzo[a,c,g,i]fluorenyl-17-methoxycarbonyl
TFA	trifluoroacetic acid
TFE	2,2,2-trifluoroethanol
THP	tetrahydropyranyl
t.l.c.	thin layer chromatography
TMSBr	trimethylsilyl bromide
TRIS	tris(hydroxymethyl)aminomethane
Trt	triphenylmethyl
UV	ultra violet
Z	benzyloxycarbonyl

AMINO ACIDS

Amino Acid	3 Letter Code	1 Letter Code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic Acid	Asp	D
Cysteine	Cys	C
Glutamic Acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

CONTENTS

CHAPTER ONE : INTRODUCTION

1.1 Introduction

The publication in 1963[1] of the first solid phase synthesis of a polypeptide, heralded the start of a technological revolution which has had a major influence on the progress of twentieth century chemical and biotechnological research. Prior to Bruce Merrifield's beautifully simple idea of covalently immobilising the growing peptide chain on an insoluble support, the synthesis of polypeptides in solution was a labour intensive and lengthy process requiring the purification and isolation of each intermediate compound. A consequence of this protracted synthetic route being that the assembly of relatively short peptide sequences required a major research effort from even the most experienced and well equipped laboratories.

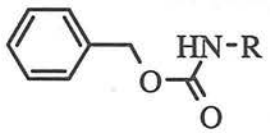
The intervening thirty years have seen technological advances in commercially available solid phase peptide synthesisers refined to such an extent, that many manufacturers claim the fully automated synthesis of peptides is "easy"[2] and that an operator only has to enter the required sequence, push "START" and walk away. Although the reality, as any peptide chemist will readily confirm, is that even a seemingly innocuous sequence of amino acids can hide synthetic difficulties which serve to remind us that the chemistry of amide bond formation and the associated protecting group strategies cannot be taken for granted simply because they are carried out under microprocessor control.

The advent of automated solid phase peptide synthesis (SPPS), was welcomed by those working in the fields of medicine and biological science who required peptides and were uninterested in how they were produced. To the organic chemistry purist, however, the lack of characterisation of intermediates was seen as the major worry.

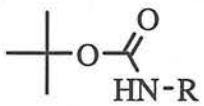
This concern has largely been overcome, due to the refinements introduced by Merrifield and others, which have resulted in the basic principle being extended to other areas such as oligonucleotide synthesis, peptide sequence analysis and related technologies.

1.2 The Merrifield Approach - Boc Strategy

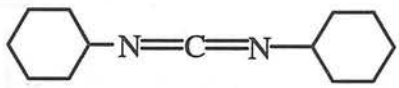
The basic principles reported in Merrifield's initial publication are outlined in Scheme 1. The second synthesis from Merrifield's laboratory[3] reported the replacement of the initially favoured benzyloxycarbonyl (Z) N^{α} protecting group (**1**) with the more acid labile t-butoxycarbonyl (Boc) group (**2**).



(1)



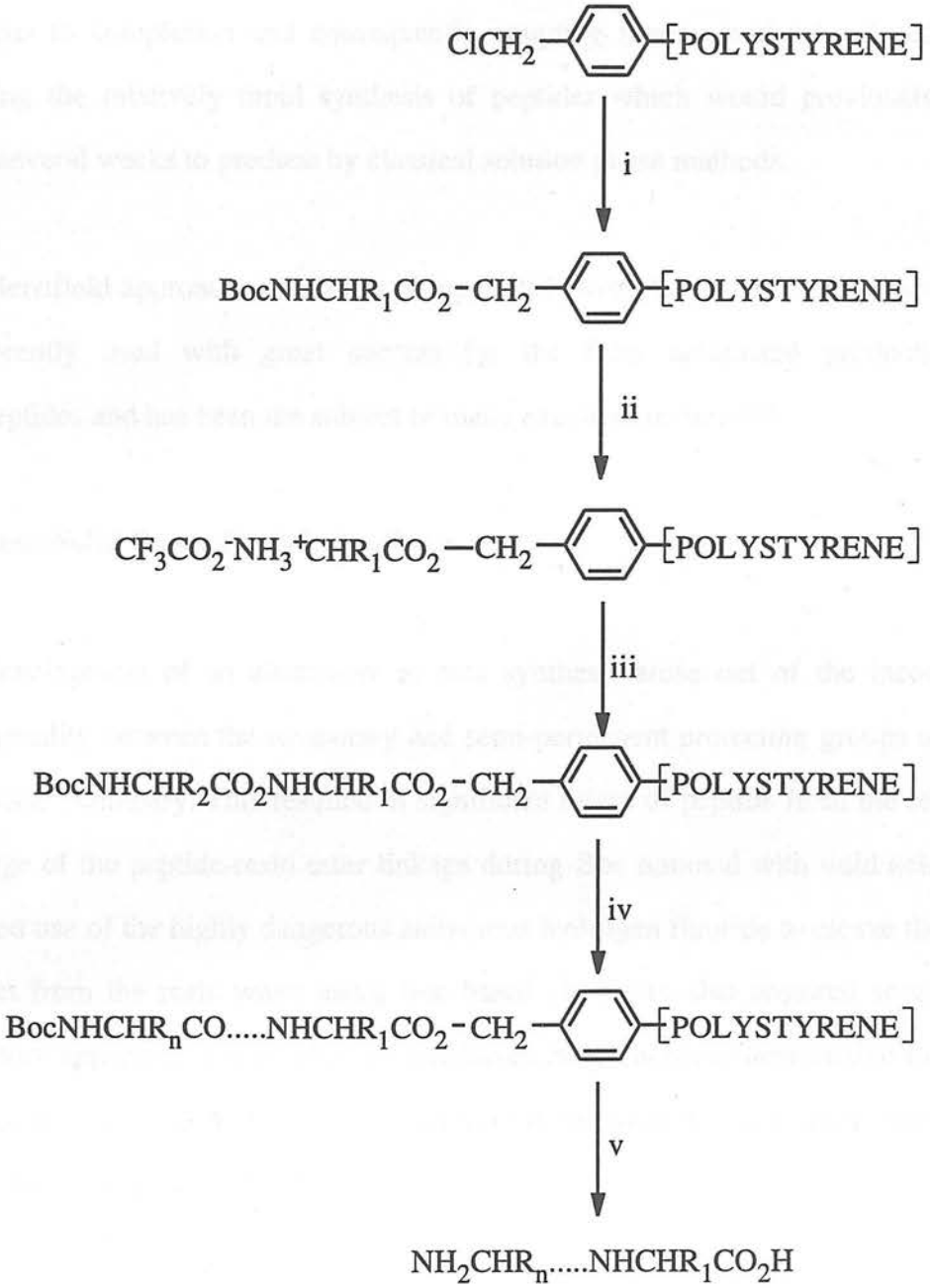
(2)



(3)

The C-terminal N^{α} protected amino acid residue was attached to the insoluble support, formed by crosslinking polystyrene with divinylbenzene and chloromethylating, by nucleophilic displacement. The Boc group was then removed by treatment with mild acid, followed by neutralisation, to give a free amino group to which the second N^{α} protected amino acid was coupled using N,N-dicyclohexylcarbodiimide (**3**). The deprotection, neutralisation and coupling steps were then repeated with other Boc protected amino acids to build up, in a stepwise fashion, the desired peptide sequence. During assembly all side chain functional groups were protected with semi-permanent protecting groups which are stable to the mild acid conditions required to remove the temporary N^{α} Boc protection and after

each step the insoluble polymer bound peptide was thoroughly washed to remove excess reagents and reaction byproducts. Finally, the peptide was released from the resin with simultaneous removal of all side chain protecting groups by treatment with anhydrous hydrogen fluoride.



Scheme 1. Conditions i, attachment of first amino acid as Cs salt by nucleophilic displacement ; ii, N^{α} deprotection with mild acid ; iii, neutralisation followed by DCCI mediated coupling of N^{α} protected amino acid ; iv, repetition of ii-iii as appropriate ; v, cleavage of peptide from resin with strong acid .

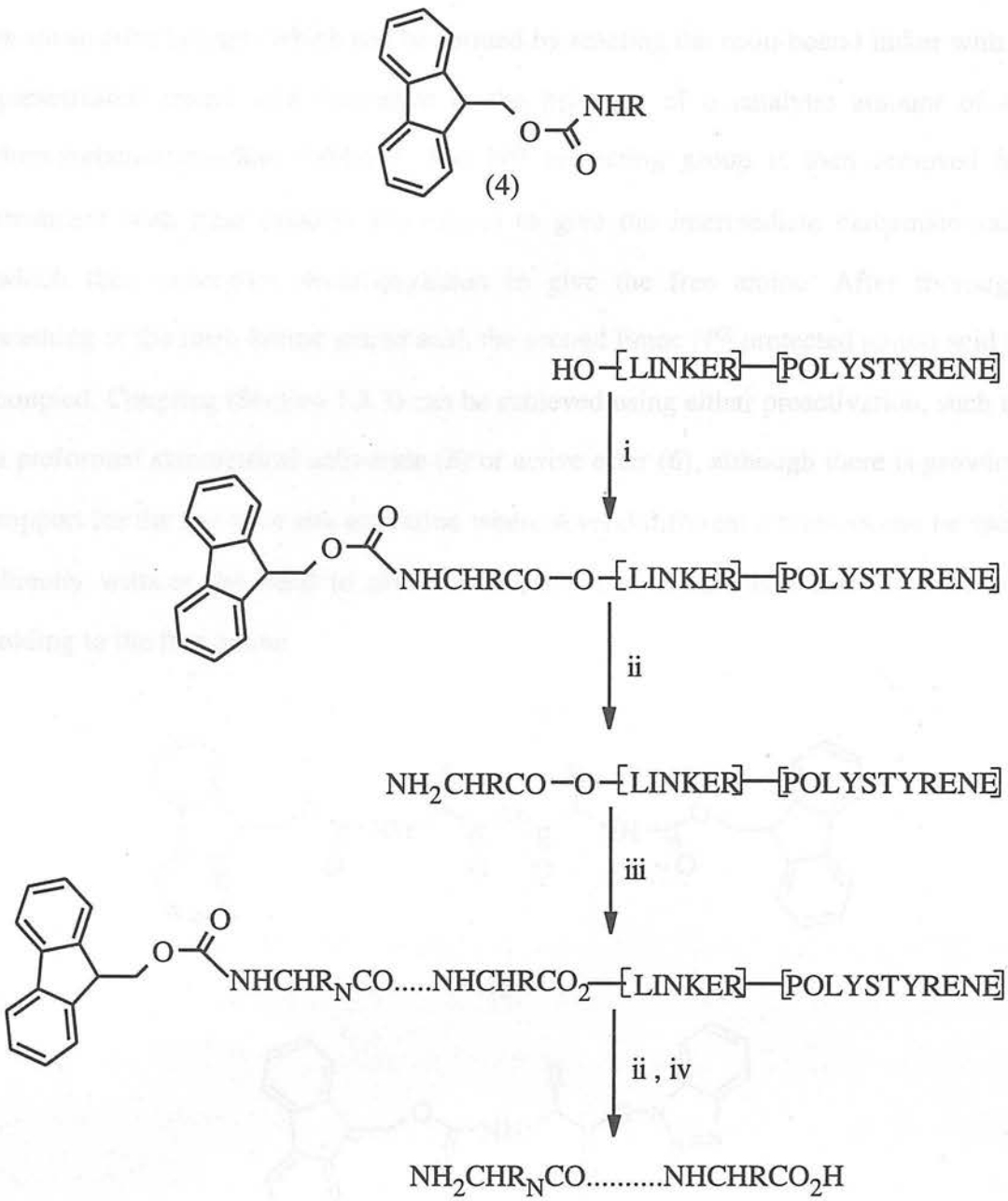
Throughout the synthetic operations the growing immobilised peptide-polymer conjugate was retained by a glass sinter which minimised mechanical losses and allowed for easy removal of reagents by filtration. The major benefits of this approach were that it allowed the use of large excesses of reagents in order to force reactions to completion and consequently coupling times could be reduced, thus allowing the relatively rapid synthesis of peptides which would previously have taken several weeks to produce by classical solution phase methods.

The Merrifield approach, which has changed little over the years since its inception, is currently used with great success for the fully automated production of polypeptides and has been the subject of many excellent reviews[4a,b].

1.3 Fmoc Solid Phase Peptide Synthesis

The development of an alternative to Boc synthesis arose out of the incomplete orthogonality between the temporary and semi-permanent protecting groups used in Boc based chemistry. This resulted in significant losses of peptide from the resin by cleavage of the peptide-resin ester linkage during Boc removal with mild acid. The required use of the highly dangerous anhydrous hydrogen fluoride to cleave the final product from the resin when using Boc based chemistry also required specialised laboratory apparatus, and resulted in Boc based methods being unattractive to many laboratories. In addition the use of strong acid in the final step may have deleterious effects on the peptide product.

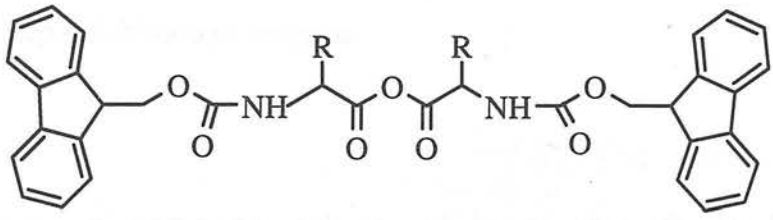
This situation changed with the introduction of the base labile N^{α} protecting group 9-fluorenylmethoxycarbonyl (Fmoc)[5] (**4**) and its incorporation into automated solid phase peptide synthesis[6,7,8]. The general strategy for Fmoc solid phase peptide synthesis (SPPS) is outlined in Scheme 2.



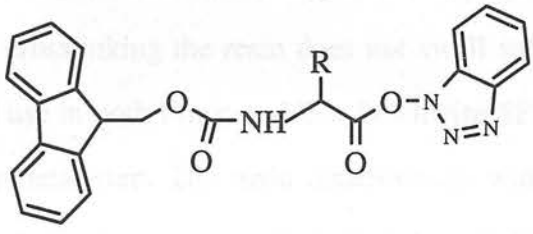
Scheme 2. Conditions i, attachment of first amino acid ; ii, N^{α} deprotect with piperidine ; iii, coupling of N^{α} protected amino acid ; iv, cleavage of the peptide from the solid support with trifluoroacetic acid .

The solid support used in standard Fmoc SPPS consists of a crosslinked polystyrene which is functionalised with an acid sensitive linker to which the first Fmoc N^{α} protected amino acid is coupled. Normally the attachment of amino acid to the linker

is *via* an ester linkage, which can be formed by reacting the resin bound linker with a preactivated amino acid derivative in the presence of a catalytic amount of 4-dimethylaminopyridine (DMAP). The N^{α} protecting group is then removed by treatment with base (usually piperidine) to give the intermediate carbamate salt, which then undergoes decarboxylation to give the free amine. After thorough washing of the resin-bound amino acid, the second Fmoc N^{α} protected amino acid is coupled. Coupling (Section 1.3.3) can be achieved using either preactivation, such as a preformed symmetrical anhydride (**5**) or active ester (**6**), although there is growing support for the use of *in situ* activation where several different activators can be used directly without the need to preform the activated amino acid derivative before adding to the free amine.



(**5**)



(**6**)

This cycle of coupling and deprotection is repeated until the desired sequence has been obtained and then the peptide is released from the resin with simultaneous removal of side chain protecting groups by treatment with 95% trifluoroacetic acid. Each of the above steps, as generally applied to automated Fmoc SPPS, are elaborated on in the following chapters.

1.3.1 The Solid Support

Early work on finding suitable resins for Boc based SPPS, identified several factors that were important in determining the optimum physical and chemical characteristics required for the solid support[9].These were:

(a) Functionality which allows attachment of the desired linker and subsequent coupling of amino acids.

(b) Stability to all reaction conditions.

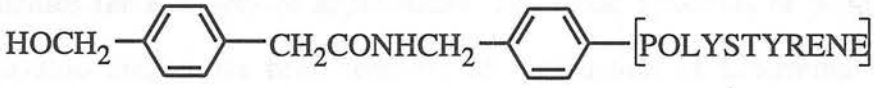
(c) The resin must swell sufficiently in the solvent used during synthesis, to allow rapid diffusion of reagents.

These characteristics are found in polymers of styrene when 1% crosslinked by m-divinylbenzene, with less than 1% crosslinking the resin produced is too fragile and with greater than 1% crosslinking the resin does not swell sufficiently[10]. This resin has found widespread use in both Fmoc and Boc batchwise SPPS, where the resin is filtered after each synthetic step. The main disadvantage with using a polystyrene based resin is the fact that polystyrene is relatively hydrophobic and it has therefore been argued[11] that a more hydrophilic support should help solvate the growing peptide chain and polymer backbone. This in turn should aid diffusion of reagents and lead to increased reaction rates and coupling yields for each step, hence producing a purer final product. This approach has been successfully applied by Atherton and Sheppard to the continuous flow method of Fmoc based SPPS using a polyamide resin. In this approach the peptide resin is continually solvated during synthesis and this requires a support which can withstand a continuous flow of

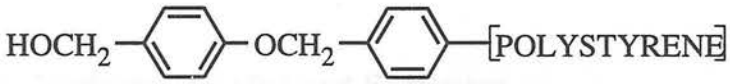
solvent. Polyamide polymers on their own are too fragile to withstand these conditions although, by chemically bonding the polyamide to the inorganic support kieselguhr, the desired physical characteristics can be achieved and this continuous flow polyamide/kieselguhr methodology has been well documented[11].

1.3.2 Resin Linkages

Final cleavage of peptides synthesised utilising Fmoc SPPS allows the use of resin linkers which are much more acid labile than the conventional polystyrene or Merrifield PAM[12] resin (**7**) used in Boc chemistry.

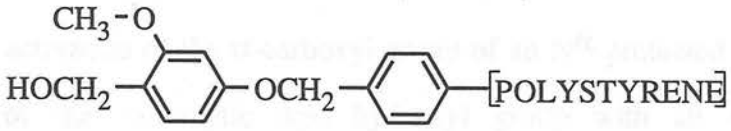


(**7**)

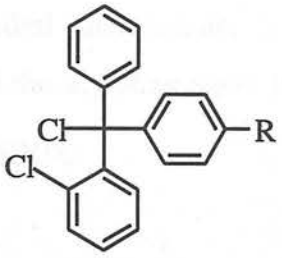


(**8**)

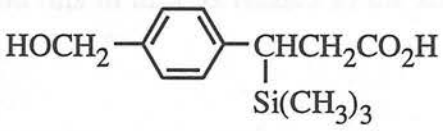
Chang and Meienhofer[6] utilised a p-benzyloxybenzylalcohol functionalised polystyrene resin (**8**), previously developed by Wang[12], for the first Fmoc solid phase synthesis of a peptide acid and this has remained popular due to its acid lability (50% trifluoroacetic acid in aprotic solvents). More recently the desire to produce fully protected fragments by using extremely mild cleavage conditions, which do not affect t-butyl based protecting groups, has led to the development of linkers which can be cleaved by very dilute TFA[13] (**9**), acetic acid[14] (**10**) and fluoride ion[15] (**11**) to give a free carboxylic acid C-terminus.



(9)



(10)

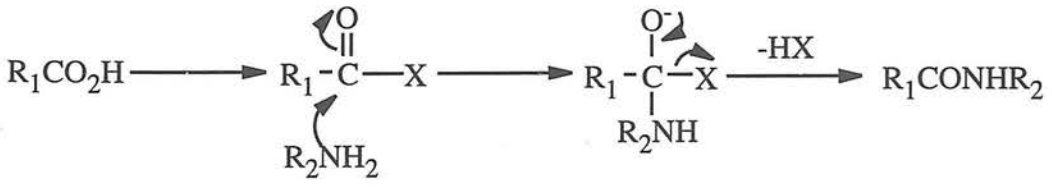


(11)

Besides free acids, linkers have also been used to generate alternative carboxyl functionalities for a variety of applications. The Fmoc synthesis of peptides using acid cleavable linkers has been reported for production of C-terminal amides[16], hydrazides[16,17] and carboxyl-linked peptide polymers[18] to name but a few.

1.3.3 Coupling Methods - Amide Bond Formation

The principal reaction in the synthesis of peptides involves the acylation of an amino acid by the carboxyl group of a second amino acid to form an amide bond. There have been a multitude of amide bond forming methods developed for the synthesis of peptides and with very few exceptions these methods are based on the direct electophilic acylation mechanism[19,20] (Scheme 3).



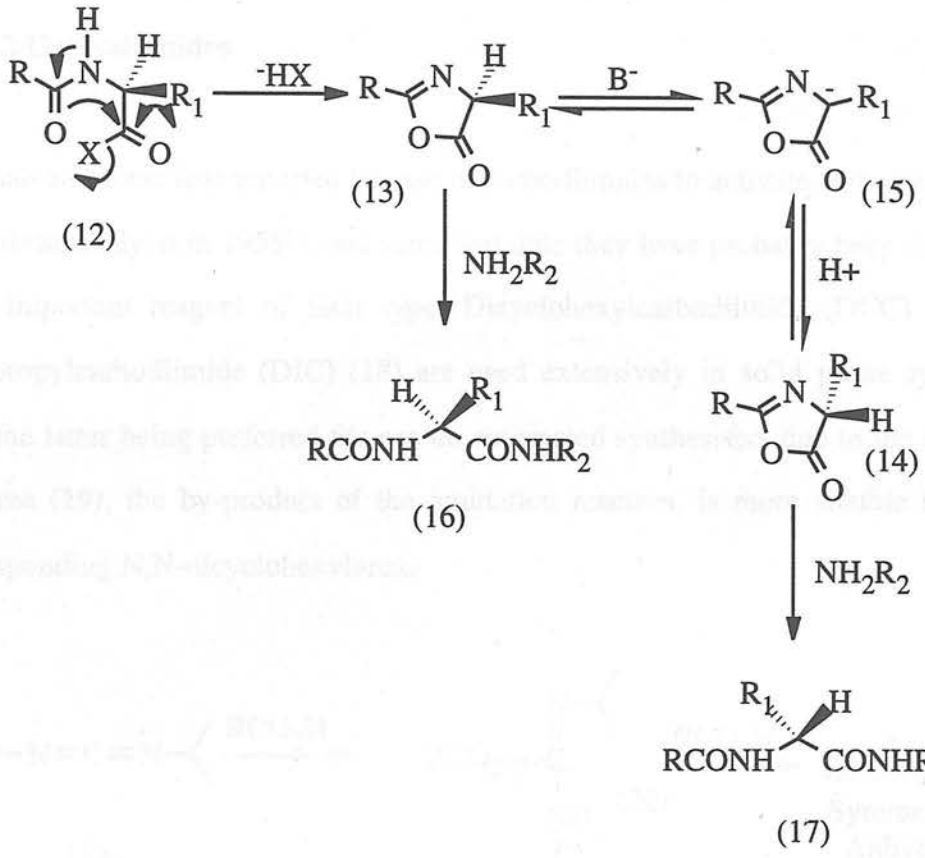
Scheme 3. Direct electrophilic acylation mechanism[19,20]

This requires activation of the α-carboxyl group of an N^{α} protected amino acid by replacement of the carboxylic acid hydroxyl group with an electronegative substituent X, which is susequently displaced by the incoming amine component via a tetrahedral intermediate. In general the better the leaving group X^-, the more powerful the acylating agent R_1COX is and this in turn, is related to the strength of the acid HX[19].

1.3.3.1 Racemisation During Activation/Coupling

One serious concern which must be considered at the activation and coupling stages of a synthesis is the danger of racemisation occurring. This side reaction has plagued the field of peptide synthesis throughout its history and much time and effort has been put into the study of conditions which minimise this undesirable side reaction[21].

The conditions employed during Fmoc SPPS tend to mean that, except for in the case of a few special amino acids (e.g. α-arylglycines) and during extremely slow couplings, the risk of racemisation occurring by direct enolisation is slight. The main cause of racemisation is the cyclisation of activated acyl amino acids (**12**) under the influence of base to give oxazolones (**13**) and (**14**) (Scheme 4).

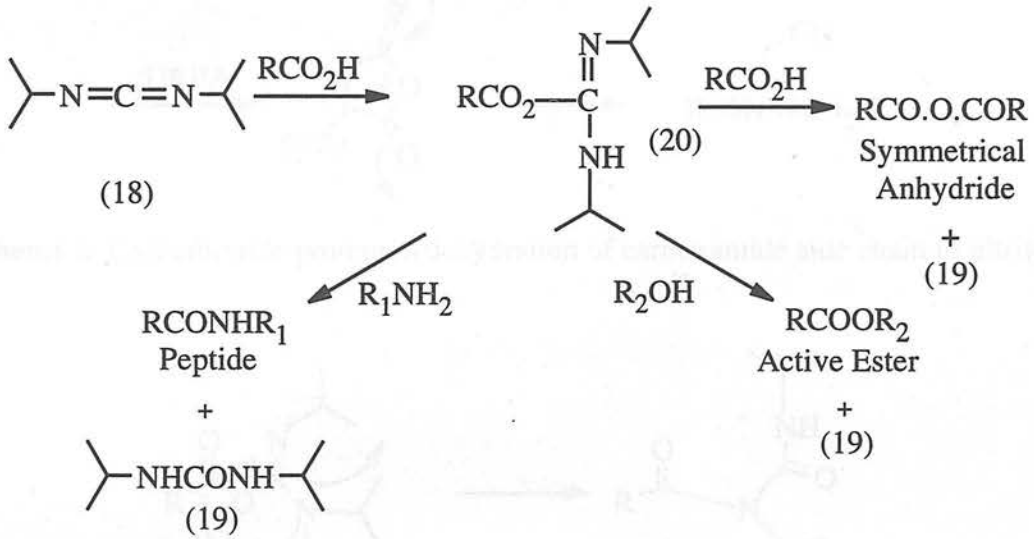


Scheme 4. Base catalysed racemisation of acyl amino acids via the oxazolone mechanism

The oxazolones (**13**) and (**14**) produced are themselves activated carboxylic acid derivatives and reaction with an amine can lead to formation of a peptide. However, due to the fact that racemisation via the stabilised anion (**15**) occurs more rapidly than the rate of aminolysis, both possible epimers (**16**) and (**17**) are formed. The extent of oxazolone mediated racemisation depends very much on the nature of the leaving group X^- and certain methods of coupling have been shown to produce lower levels of racemisation than others. The majority of amide bond forming methods have previously been reviewed[22,23] and will therefore not be covered in this text . Instead the narrower area of those which have been generally applied to Fmoc SPPS will be covered here.

1.3.3.2 Carbodiimides

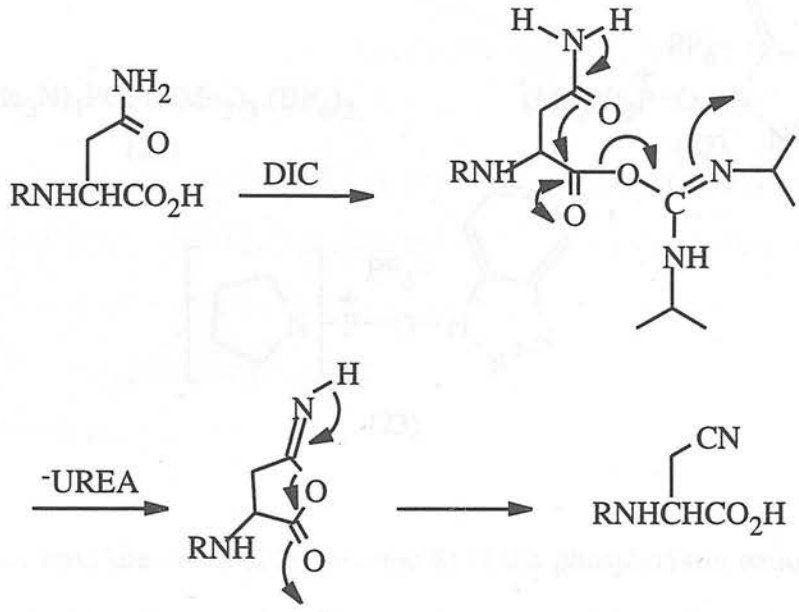
Sheehan and Hess first reported the use of carbodiimides to activate carboxylic acids towards aminolysis in 1955[24], and since that date they have probably been the single most important reagent of their type. Dicyclohexylcarbodiimide (DCC) (**3**) and diisopropylcarbodiimide (DIC) (**18**) are used extensively in solid phase synthesis, with the latter being preferred for use on automated synthesisers due to the fact that the urea (**19**), the by-product of the amidation reaction, is more soluble than the corresponding N,N-dicyclohexylurea.



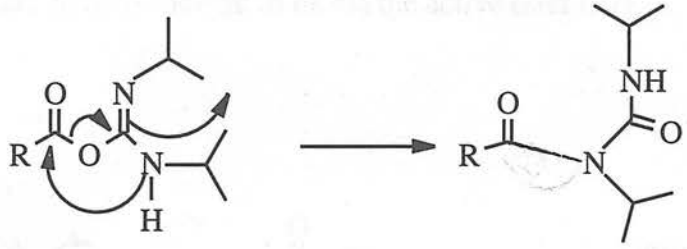
Scheme 5.

Addition of the carboxy group to the carbodiimide functionality (Scheme 5) gives an O-acylisourea (**20**). This very reactive acylating agent can then either react with a free amino group to form the peptide bond, or more usually with oxygen nucleophiles to produce an activated ester or symmetrical anhydride. Whilst the addition of oxygen nucleophiles generates an acylating reagent of lower reactivity with respect to aminolysis than (**20**), this is beneficial in that it also helps alleviate several serious side reactions, such as racemisation (Scheme 4), carboxamide

dehydration of asparagine/glutamine (Scheme 6) and formation of N-acylureas (Scheme 7), which have been found to accompany the generation of reactive species such as (**20**)[22].



Scheme 6. Carbodiimide promoted dehydration of carboxamide side chain to nitrile

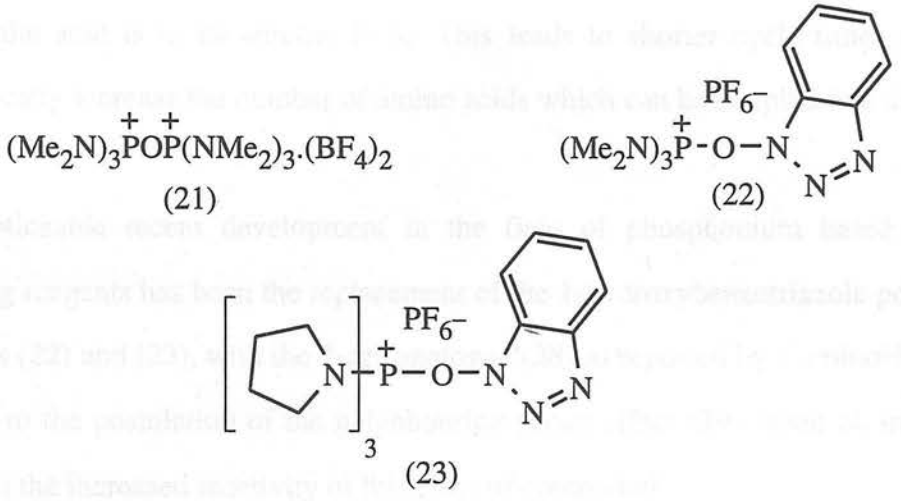


Scheme 7. N-acylurea formation

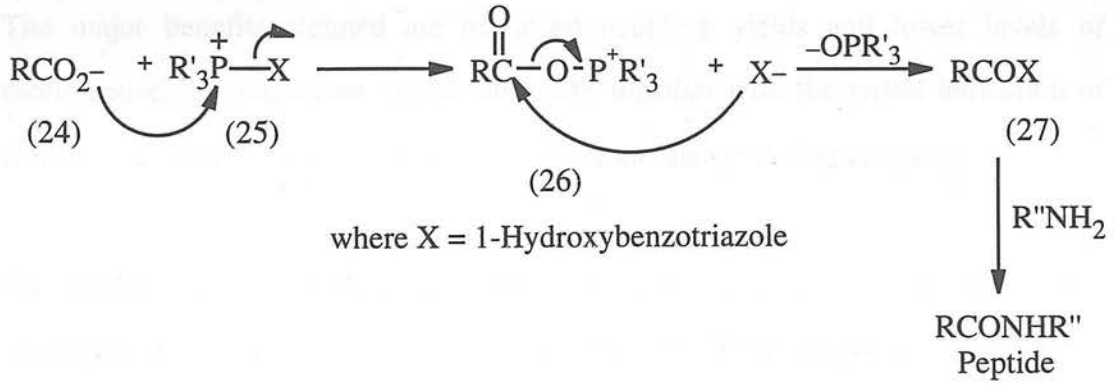
1.3.3.3 Phosphonium Reagents

Phosphonium reagents are becoming increasingly popular as direct coupling reagents in solid phase peptide synthesis. The di-tetrafluoroborate (**21**) initially reported by

Bates *et al*[25] has spawned several related reagents, the most noteable of which are the BOP[26] (**22**) and PyBOP[27] (**23**) reagents developed by Castro.



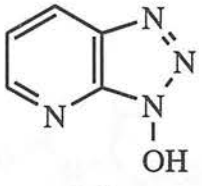
Attack of carboxylate anion (**24**) (Scheme 8) at the phosphonium cation (**25**) forms an acyloxyphosphonium species (**26**), which reacts readily with nucleophiles at the acyl carbon. Although several pathways which result in amide bond formation are possible, the main route is thought to be via the active ester (**27**).



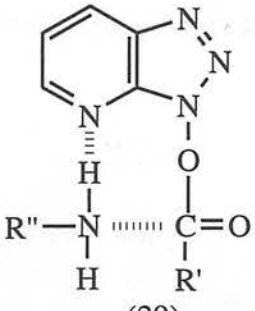
Scheme 8.

Typical coupling reactions utilising phosphonium reagents of this type require no need for preactivation. The coupling reagent is simply mixed together with the resin bound amino funtional group, the N^{α} protected amino acid and some tertiary base to ensure the acid is in its anionic form. This leads to shorter cycle times and can dramatically increase the number of amino acids which can be coupled in a day.

One noticeable recent development in the field of phosphonium based peptide coupling reagents has been the replacement of the 1-hydroxybenzotriazole portion of reagents (**22**) and (**23**), with the 7-aza analogue (**28**) as reported by Carpino[28,29]. This has led to the postulation of the neighbouring group effect (**29**) being an important factor in the increased reactivity of this class of compound.



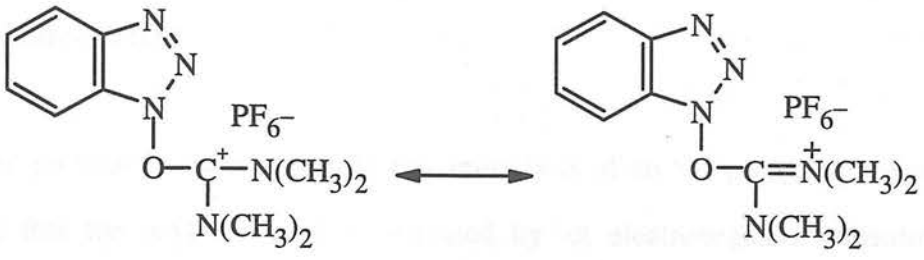
(28)



(29)

The major benefits claimed are enhanced coupling yields and lower levels of racemisation in comparison to (**22**) and (**23**), together with the visual indication of reaction endpoint by a yellow to colourless colour change during coupling.

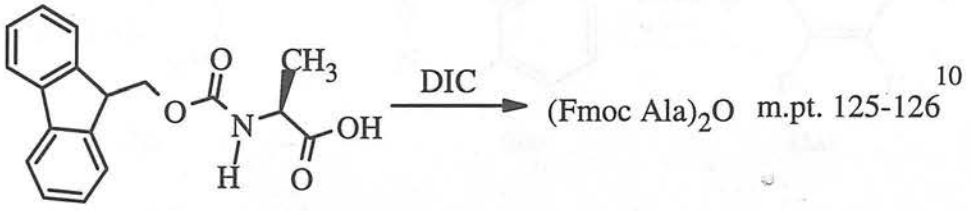
One further class of reagents which are related to phosphonium salts are the carbon analogues or uronium salts (**30**) described by Knorr[30]. These reagents show a similar reactivity to their BOP counterparts and the 7-aza analogues have also been reported[29].



(30)

1.3.3.4 Symmetrical Anhydrides

The symmetrical anhydrides of N^{α} protected amino acids are easily prepared from two equivalents of amino acid derivative and one equivalent of carbodiimide (Scheme 9). They are generally fairly stable crystalline solids, but are more routinely pre-formed automatically during synthesis and used without isolation.



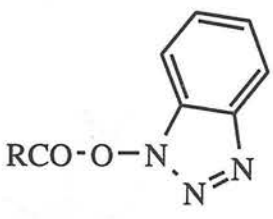
[10]

Scheme 9.

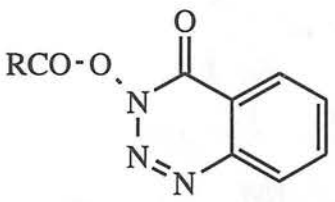
Pre-formed symmetrical anhydrides (PSA) have found widespread use in Fmoc SPPS due to their ease of preparation, reactivity and the unambiguous aminolysis they produce on reaction with a free amino group. The major drawbacks to their use is the fact that one equivalent of expensive N^{α} protected amino acid is discarded and also that certain amino acid side chains give serious side reactions with carbodiimides e.g. asparagine/glutamine - dehydration to nitrile (Section 1.3.3.2) and histidine - racemisation[31].

1.3.3.5 Active Esters

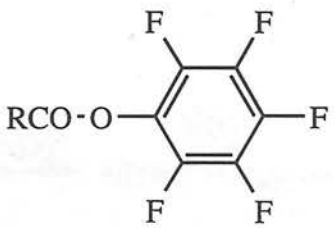
As stated previously (Section 1.3.3), the aminolysis of an N^{α} protected amino acid requires that the acyl carbonyl is activated by an electronegative substituent X (Scheme 3). Many types of active ester, which fulfil this requirement, have been reported[32] for the generation of amide bonds, although relatively few have been successfully applied to automated solid phase peptide synthesis. Those which are currently available commercially, are generally esters of phenols or other similar acidic functionalities which give stable crystalline derivatives e.g. esters of 1-hydroxybenzotriazole (HOBt)[33] (**31**), pentafluorophenol (Pfp)[34] (**32**) and 3,4-dihydro-3-hydroxy-4-oxo-1,2,3-benzotriazine[33,35] (**33**).



(31)



(33)



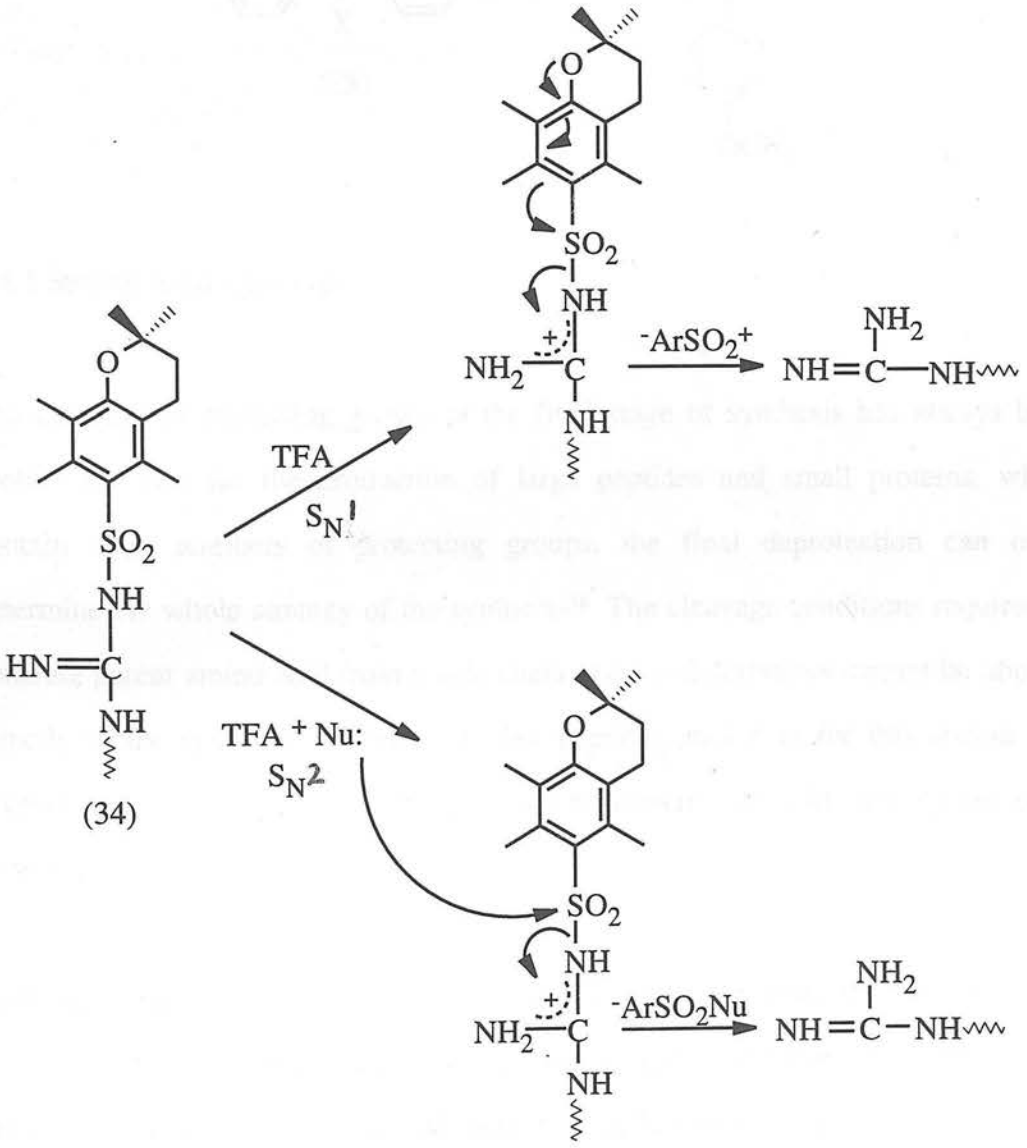
(32)

Because all of these active esters are less activated than the active intermediates discussed previously, side reactions which normally accompany the coupling step are less of a problem with this class of compounds.

1.4 Acidolytic Cleavage - Mild Acid Cleavage

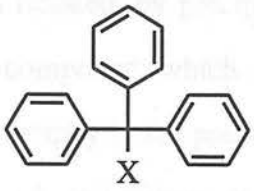
At the completion of the assembly phase of SPPS, it is common practice to simultaneously deprotect the majority of side chain protecting groups and remove the peptide from the resin in a single step. This process usually involves generation of stable cations which can attack electron rich side chains of the amino acids

tryptophan , tyrosine and methionine. Additives such as water, 1,2-ethanedithiol (EDT), thioanisole, anisole and phenol can be added to trifluoroacetic acid (TFA) based cleavage mixtures in order to minimise these side reactions, by scavenging the TFA liberated carbocations before they can interact with susceptible side chains. These scavengers act not only by trapping carbocations, but also as nucleophiles in the cleavage step by shifting the mechanism of cleavage from S_N1 towards S_N2 for certain less acid labile protecting groups e.g. 2,2,5,7,8-pentamethylchroman-6-sulphonyl (Pmc) protected guanidino side chain of arginine (**34**)[36] (Scheme 10).

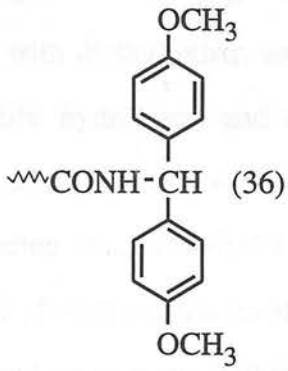


Scheme 10. Acidolytic cleavage of Arg(Pmc)

As a general rule, cleavage of peptides which do not contain sensitive amino acids and only require removal of t-butyl based protecting groups, can be accomplished by treating with 95% TFA/5%H_2O. But for sequences which contain these aforementioned residues or trityl (**35**) , 4,4'-dimethoxybenzhydryl (MBH) (**36**) and Pmc protected residues, carbocation scavengers must be included[37].



(35)



(36)

1.4.1 Strong Acid Cleavage

The cleavage of protecting groups at the final stage of synthesis has always been problematic and for the production of large peptides and small proteins, which contain large numbers of protecting groups, the final deprotection can often determine the whole strategy of the synthesis[38]. The cleavage conditions required to generate parent amino acid from a side chain protected derivative cannot be applied directly to the synthesis of large complex peptides, and it is for this reason that several groups have tried to ensure complete deprotection on acid cleavage by using strong acids.

Both HF[39] and TFMSA[40] have been reported for cleavage from the resin and side chain deprotection, after assembly using Fmoc SPPS. TFMSA offers the added advantage that it is compatible with conventional laboratory glassware, is an easily handled liquid and its use in cleavage and side chain deprotection with respect to

Boc based chemistry has been well reviewed[41,42]. More recently Yajima *et al* have reported the use of trimethylsilylbromide (TMSBr)[43] as a hard acid[44] in combination with thioanisole as a soft nucleophile to effect complete side chain deprotection. This reagent efficiently removes all side chain protecting groups commonly used in Fmoc SPPS (with the exception of the acetamidomethyl derivative of cysteine) including benzyl esters and ethers, pMeBz(Cys) and Z(Lys). After cleavage using TMSBr/TFA the peptide is then isolated, by precipitation with diethyl ether, as the intermediate trimethylsilylated compound which is readily hydrolysed and desalted by size exclusion chromatography in the presence of acetic acid to give the pure deprotected product. Yajima and co-workers also reported that TMSBr/TFA quantitatively reduces Met(O) to Met and that the amount of succinimide formation found in a model sequence was reduced from 4.6%, found when using TFMSOTf/TFA, to nil when using TMSBr/TFA. This reagent is therefore an attractive alternative to TFA for the cleavage of peptides synthesised using Fmoc SPPS.

1.5 Chemical Synthesis of Proteins

Six years after the initial publication outlining the solid phase method of peptide synthesis[1], Gutte and Merrifield reported the first solid phase synthesis of a protein[45] and, although the biologically active product was by no means homogeneous, this proved that it was possible to assemble such a complex molecule in a stepwise manner and that it would fold into its active form. Since that pioneering synthesis the solid phase synthesis of large peptides and small proteins has been the ultimate challenge of SPPS methodology.

The intervening 20 or so years have seen improvements in both the chemistry and separation science which now make the production of homogeneous, crystalline proteins by stepwise solid phase synthesis a realistic goal. In particular, the chemical synthesis of proteins and analogues containing residue specific labels and non-coded amino acids is now seen as a complementary approach to recombinant DNA techniques. This has most recently been exemplified by the synthesis and X-ray crystal structure determination of Kent's L-(Aba^{67,95,167,195})HIV-1 protease analogue[46], using Boc SPPS, and the Fmoc SPPS and crystallisation of ubiquitin[47,48] and a closely related analogue[49].

The total chemical synthesis approach to such molecules demands highly optimised protocols, both during synthesis and purification, in order to produce the desired target molecule in pure form. Several important factors which influence the success of such syntheses have been identified.

1. As a general rule the initial loading of the first amino acid onto the solid support should be lower than would normally be employed in the synthesis of a short peptide (i.e. <0.25mM/g of resin).

2. Each coupling must use an excess of activated amino acid derivative in order to force the coupling to completion and double coupling cycles are preferable.

3. After each coupling step and before the subsequent deprotection of the N^{α} protecting group, any remaining free amino groups must be "capped" by a suitable acetylating agent (e.g.acetic anhydride) before the next amino acid in the sequence is introduced.

4. The progress of the synthesis must be monitored in order to assess the effectiveness of couplings and deprotections as well as the possibility of side reactions or mechanical failure during automated synthesis.

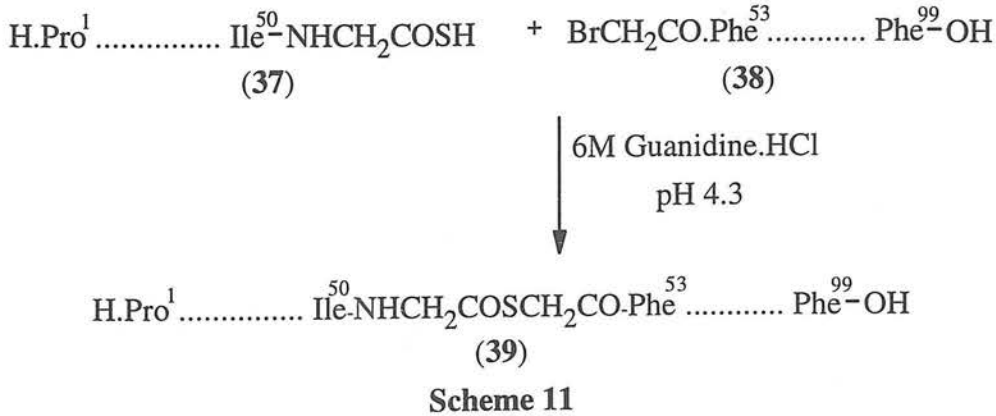
5. At the completion of synthesis, after side chain deprotection and release from the resin, the crude cleaved protein should then be sequentially purified using a variety of techniques which allow differentiation between the desired sequence and any accumulated acylated truncation sequences. The protein should then be allowed to fold into its native form and purified to homogeneity under non-denaturing conditions.

6. Finally the pure product must be thoroughly characterised using techniques such as amino acid analysis, isoelectric focussing, mass spectrometry and N-terminal sequencing, in order to confirm the composition and purity of the product. It is only once these criteria have been satisfied that the protein structure can be studied with any degree of certainty by either non-destructive techniques such as nuclear magnetic resonance (NMR) and X-ray crystallography, or by enzymatic degradation if applicable.

1.5.1 Chemical Synthesis of Proteins - Convergent Strategy

An attractive approach to overcoming the inadequacies of stepwise SPPS as a method of protein synthesis, is the use of a convergent strategy in which smaller fragments of a protein are synthesised and purified separately before being coupled together to form the desired molecule. In principle, the advantage of this approach lies in the fact that the difference in size between product and reactants should allow easy purification of the target molecule. Although in practice, the low solubility of large protected peptide fragments in organic solvents has meant that few laboratories have used this approach successfully.

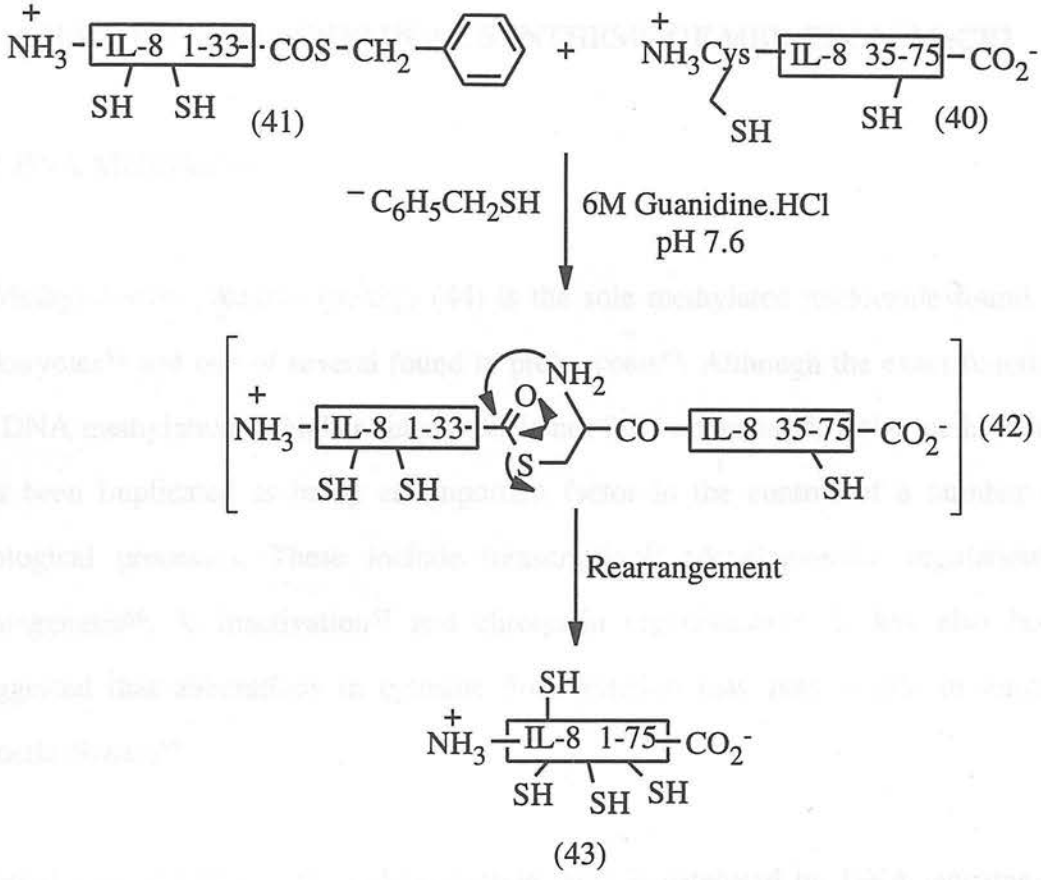
An interesting variation on this convergent approach to protein synthesis has recently been reported by Schnolzer and Kent[50] in which the two halves of the target polypeptide were synthesised incorporating a mutually reactive functional group in each segment. The N-terminal fragment contained a thioacid functional group at the C-terminus (**37**) (Scheme 11), whilst the C-terminal fragment incorporated an N-terminal alkyl bromide (**38**). The two segments (**37**) and (**38**) were then ligated together at pH4.3 in guanidine solution *via* an S_N2 nucleophilic displacement to give a fully biologically active analogue of HIV-1 protease (**39**) containing a thioester pseudo-peptide bond at the ligation site.



Scheme 11

This highly selective ligation reaction means that the coupling of the two fragments can be carried out with no need for side chain protection, even of cysteine residues[50]. The main disadvantages of this method are due to the instability of the thioester pseudo peptide bond at physiological pH and the fact that this structural motif results in the introduction of an altered backbone into the protein. Whilst these factors are unimportant in the case of HIV-1 protease, the general application of this method remains to be demonstrated in other systems.

Kent's group have recently reported a further variation of their chemoselective ligation strategy which may prove to have all the advantages of the thioester method without the previously stated disadvantages. Chemoselective ligation to give a native peptide bond has been reported for the synthesis of IL-8 (72 residues)[51]. The ligation reaction (Scheme 12) involves nucleophilic attack by the thiol group of an N-terminal cysteine residue of one fragment (**40**), the C-terminus of the target protein, on a second fragment (N-terminus of target protein) which contains a carboxy terminal thioester functional group (**41**). Displacement of the thioester results in the formation of an intermediate S-acyl compound (**42**), which spontaneously rearranges to give a native amide bond and the regenerated thiol of the cysteine residue next to the ligation site in the product protein (**43**). Again this ligation reaction is carried out in guanidine solution and at pH 7.6 with no need for any side chain protection, even in the presence of other cysteine residues which may be present in the sequence[51].



Scheme 12. Schematic diagram of chemoselective ligation to give native amide bond

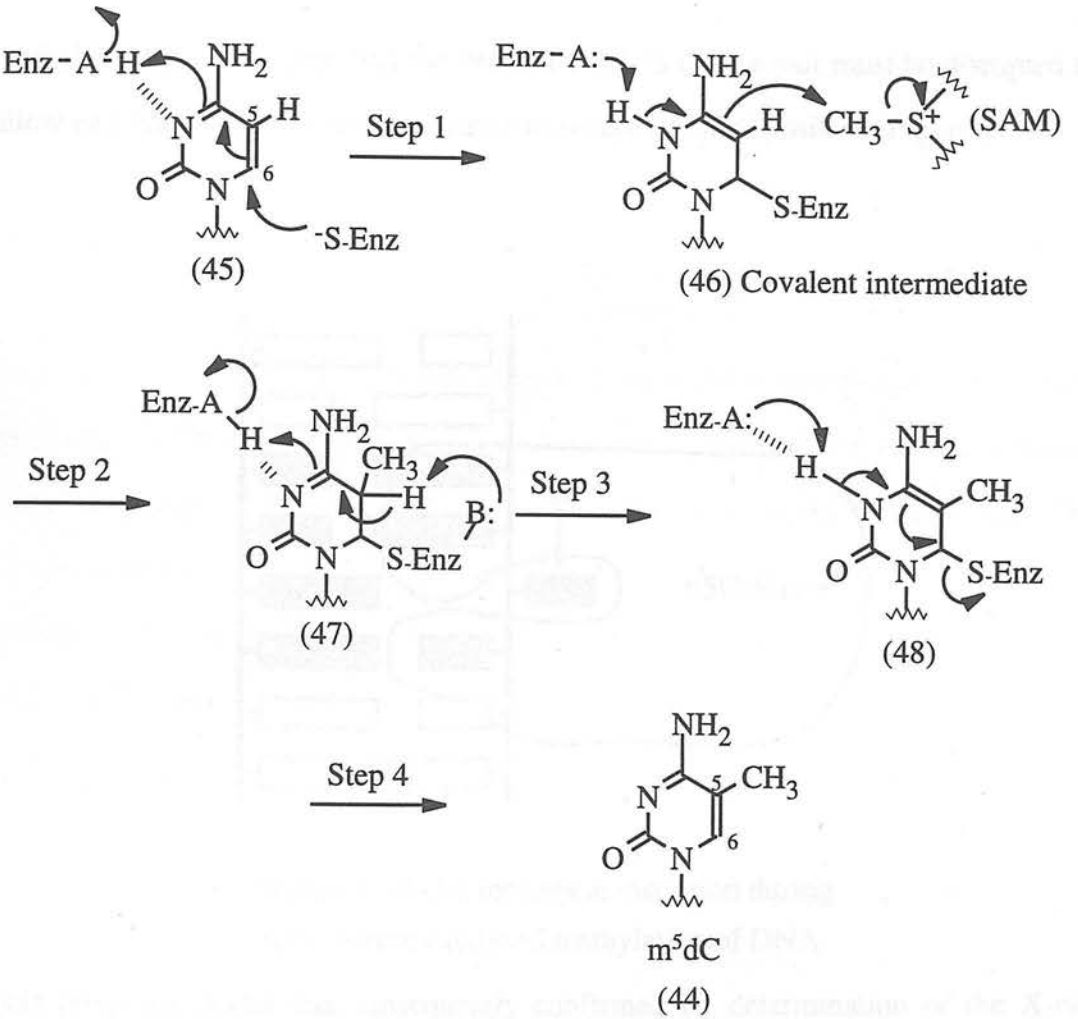
This method offers a departure from conventional approaches towards the chemical synthesis of proteins and should allow construction of large synthetic proteins (>200 amino acids) in a controlled and reproducible manner and in yields which should rival recombinant DNA technology.

CHAPTER TWO : CHEMICAL SYNTHESIS OF MBD FROM MeCP2

2.1 DNA Methylation

5-Methyl-2'-deoxycytidine (m^5dC) (**44**) is the sole methylated nucleoside found in eukaryotes[52] and one of several found in prokaryotes[53]. Although the exact function of DNA methylation in higher eukaryotes is not fully understood, DNA methylation has been implicated as being an important factor in the control of a number of biological processes. These include transcription[54], developmental regulation[55], mutagenesis[56], X inactivation[57] and chromatin organisation[58]. It has also been suggested that abberations in cytosine 5-methylation may play a role in human genetic disease[59].

Methylation of DNA at the polynucleotide level is catalysed by DNA cytosine-5-methyltransferase (m5C-Mtase) enzymes which mediate delivery of a methyl group from S-adenosyl-L-methionine (SAM) to the 5-position of a cytosine residue in DNA. The mechanism employed by m5C-Mtase's as proposed by Verdine *et al*[52] is given in Scheme 13.



Scheme 13. Mechanism of enzyme mediated cytosine methylation[52] .

The first step of the reaction involves attack of a cysteine thiolate on C6 of the cytosine substrate (**45**) with simultaneous protonation at N3 by an enzyme-derived acid to give an enamine covalently attached to the enzyme (46). The enamine then attacks the methyl group of S-adenosyl-L-methionine (SAM), transferring it to C5 (**47**) (Step 2, Scheme 13). Abstraction of the proton at C5 (Step3) by some basic residue on the enzyme again yields an enamine (**48**) which then undergoes conjugate elimination (Step 4) to yield the product (**44**) and the re-generated m5C-Mtase.

Verdine and co-worker proposed[60] that because protons were transferred to and from N3 of the cytosine residue, and also that co-valently linked duplexes were still

methylated by the enzyme, that the Watson-Crick G·C base pair must be disrupted to allow extrusion of the cytosine substrate to an extra-helical position (Figure 1).

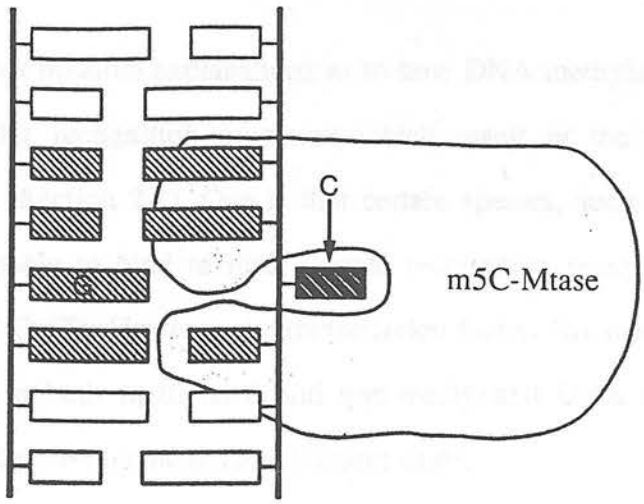
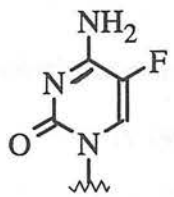


Figure 1. Model for helical distortion during m5C-Mtase catalysed methylation of DNA

This proposed model was subsequently confirmed by determination of the X-ray crystal structure of a covalent reaction intermediate from the reaction of a m5C-Mtase, SAM and a 13mer DNA duplex containing 5-fluoro-2'-deoxycytidine (FdC) (**49**) as it target[61].



(49)

2.2 Synthesis and Structural Studies of the Methyl-CpG Binding Domain from the Chromosomal Protein MeCP2

2.2.1 General

There have been two possible explanations as to how DNA methylation serves as a signal for molecular recognition processes which result in the aforementioned biological changes (Section 2.1). One is that certain species, such as transcription factors, may be unable to bind to their normal recognition sequences when they contain methylated CpG's. However the transcription factor Sp1 has been shown to bind equally well to both methylated and non-methylated DNA and can indeed stimulate transcription from a methylated binding site[62].

The second possibility is that nuclear proteins which bind to methylated CpG, could form protein-DNA complexes which do not allow access to transcription factors or other species which normally recognise the "naked" DNA. Recently, several proteins have been identified which show a high affinity for methylated DNA and this has given support to the indirect mechanism.

Two chromatin associated proteins which bind to double stranded DNA containing symmetrically methylated CpG's have recently been reported and are referred to as methyl-CpG-binding proteins or MeCP's. MeCP1 requires at least 12 methylated sites for efficient binding[63] whereas the second member of this family of proteins, MeCP2, can bind specifically to a single methyl-CpG pair. Neither protein requires specific sequences to either side of the symmetrically methylated CpG sequence in double stranded DNA, and this highly specific binding distinguishes them from other proteins which bind to methylated DNA[64,65].

As well as differing in the number of methyl-CpG's required for binding, MeCP-1 and MeCP-2 exhibit functional differences. There is some evidence that MeCP-1 can inhibit transcription from methylated promoters both *in vitro* and *in vivo*[54], whilst initial reports by Meehan and co-workers[66] indicate that MeCP-2 is not involved in transcriptional repression. However, MeCP-2 may play a role in the protection of methylated genome DNA against nucleases, due to the fact that it is much more abundant than MeCP-1 and also due to its distribution in the genome[66].

MVAGMLGLRKEKSEDQDLQGLKEKPLKFKKVKKDKKEDKE
GKHEPLQPSAHHSAEPAEAGKAETSESSGSAPAVPEASAS
PKQRRSIIRDRGPMYDDPTLPEGWTRKLKQRKSGRSAGKY
DVYLINPQGKAFRSKVELIAYFEKVGDTSLDPNDFDFTVT
GRGSPSRREQKPPKKPKSPKAPGTGRGRGRPKGSGTGRPK
AAASEGVQVKRVLEKSPGKLLVKMPFQASPGGKGEGGGAT
TSAQVMVIKRPGRKRKAEADPQAIPKKRGRKPGSVVAAAA
AEAKKKAVKESSIRSVQETVLPIKKRKTRETVSIEVKEVV
KPLLVSTLGEKSGKGLKTCKSPGRKSKESSPKGRSSSASS
PPKKEHHHHHHHAESPKAPMPLLPPPPPPEPQSSEDPISP
PEPQDLSSSICKEEKMPRAGSLESDGCPKEPAKTQPMVAA
AATTTTTTTTTVAEKYKHRGEGERKDIVSSSNPRPNREEP
VDSRTPVTERVS

Figure 2. Deduced amino acid sequence of MeCP-2 (MBD underlined)

The deduced amino acid sequence of MeCP-2 is shown in Figure 2. MeCP-2 is a 492 amino acid 84Kd protein which is rich in basic residues and, although it contains short motifs which have been implicated in minor groove binding of A-T rich DNA, shows no extensive similarities to other DNA binding proteins[67]. Bird and co-

workers[67] were able to show, by constructing deletion mutants of MeCP-2, that an 85 amino acid section from the N-terminus of MeCP-2 (78-162) (underlined in Figure 2) was responsible for the specificity of the protein for methylated DNA. This methylated-CpG-binding domain, or MBD protein, has been overexpressed and shown to bind *in vitro*, as a monomer, to double stranded DNA containing a single symmetrically methylated CpG pair with an approximate dissociation constant of 10^{-9}M[67]. MBD has negligible affinity for either non-methylated or hemi-methylated DNA and the amino acid sequence of MBD contains no recognisable DNA binding motifs.

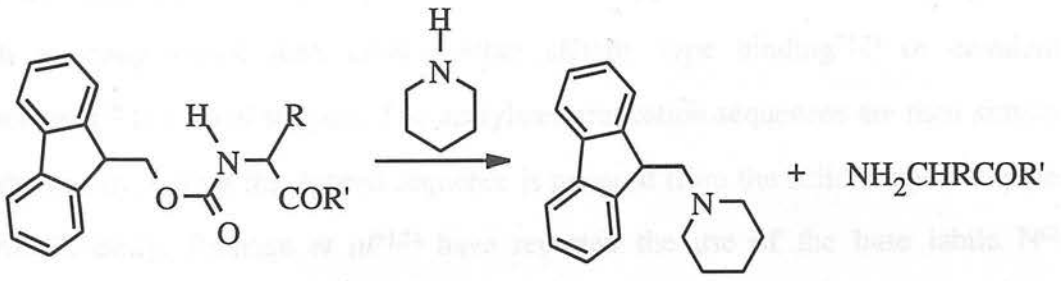
In order to study the interaction of MBD with methylated DNA at the molecular level, sufficient amounts of protein for NMR studies were required. However, initial attempts to produce MBD using recombinant DNA technology gave only low levels of the desired protein which proved to be extremely insoluble, even at relatively low concentration in aqueous solution[68]. It was therefore decided to attempt the total chemical synthesis of MBD protein using batch stepwise Fmoc SPPS.

2.2.2 Chemical Synthesis of MBD

As has been stated previously (Section 1.5), the total chemical synthesis of a protein requires a different strategy from that normally employed for the synthesis of short peptide sequences. In particular the increased possibility of secondary structure formation during synthesis often leads to abrupt drops in coupling and deprotection efficiencies, which in turn may ultimately result in very low yield of the desired sequence along with the inherent problems of isolation of the protein from the accumulated truncation sequences at the completion of the assembly. It was for these reasons that an initial trial synthesis was attempted as a "scouting run" which, it was

hoped, would highlight any sections of the sequence which required extended couplings or deprotections in order to optimise the final synthesis.

This first synthesis was undertaken on an initial substitution of 0.15mM/g of Fmoc.Arg(Pmc) on 4-alkoxybenzylalcohol functionalised polystyrene resin (**8**) and utilising single couplings of 3.5 equivalents of preformed HOBt esters. All side chains were protected as described in the experimental notes (Chapter 7) with the exception of the carboxamide side chains of asparagine and glutamine which were incorporated unprotected. The coupling efficiencies for each amino acid residue were estimated by automatically taking an aliquot of the deprotection solution, containing the generated fulvene-piperidine adduct (**50**) (Scheme 14), passing it through a UV detector set at 302nm and integrating the peak obtained[69].



(50)

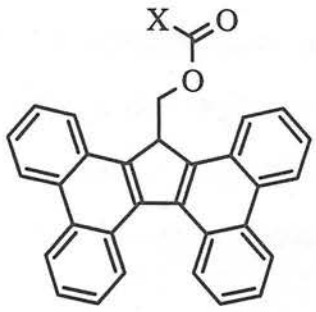
Scheme 14.

Thus, the on line coupling monitor showed a fall off in coupling efficiency over the regions corresponding to residues 69 to 56 and 41 to 25. The coupling efficiency (by UV of the deprotection peak) for this trial synthesis dropped to <10% of the initial peak height for the first amino acid on the resin, and therefore no attempt was made to isolate any of the expected sequence.

The 85 amino acid sequence of MBD was then resynthesised using optimised coupling cycles over the areas which had proved to be problematic during the initial

synthesis. At the completion of the assembly phase of this optimised synthesis the Fmoc deprotection peak corresponding to the penultimate residue, indicated an overal coupling efficiency of approximately 70% with respect to the value obtained for the deprotection of the first residue. The N-terminal Fmoc group was left on at the end of the synthesis and the resin bound product was then treated with a mixture of acetic anhydride/ HOBt/ DIEA in a sonic bath, in order to "cap" any remaining free amino groups present on the resin.

One of the main obstacles to the stepwise chemical synthesis of proteins, is the difficulty in purification of the desired sequence from the accumulated deletion sequences which form, due to incomplete couplings during synthesis. Several methods have been reported, which allow differentiation between the target sequence and any acetylated truncations present, by derivatising the N-terminus of the protein with a group which then allows either affinity type binding[70,71] or covalent attachment[72] to a solid support. The acetylated truncation sequences are then simply washed away, before the desired sequence is released from the solid support in pure form. Recently, Ramage *et al*[73,74] have reported the use of the base labile N^{α} protecting group tetrabenzo(a,c,g,i)fluorenyl-17-methoxycarbonyl (Tbfmoc) (**51**) for the purification of synthetic peptides and small proteins. This method allows rapid and efficient separation of the Tbfmoc-peptide from any acetylated truncation sequences, due to the high affinity of the polycyclic aromatic label for porous graphitised carbon (PGC)[73,74]. Alternatively, the hydrophobic nature of the Tbfmoc group and its specific UV absorbance at 364nm can be exploited to simplify peptide purification by RP-Hplc. This method was therefore chosen to help aid the purification of the synthetic 85 amino acid MBD protein.



(51) X= NHR
(52) X= Cl

After removal of the N-terminal Fmoc group with 20% piperidine in DMF, the Tbfmoc group was introduced by treatment of the resin bound peptide with chloroformate (**52**)[75] and DIEA in dichloromethane with sonication. After three hours the resin bound Tbfmoc N^{α} protected peptide was thoroughly washed with dichloromethane before being dried, cleaved from the resin (TFA, water, EDT, thioanisole,phenol), precipitated with diethyl ether and gel filtered (Sephadex G50, 30% AcOH). Figure 3(a) shows the analytical RP-Hplc trace of the crude Tbfmoc-MBD after size exclusion chromatography.

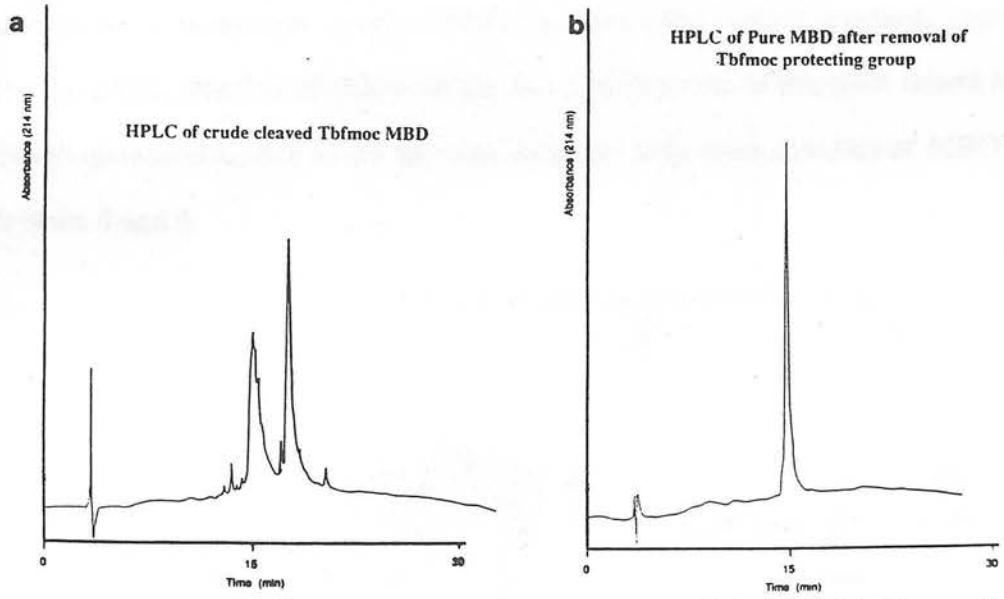


Figure 3. HPLC traces of (a) crude cleaved MBD and (b) pure MBD protein after removal of the Tbfmoc group

Semi-preparative Hplc, monitoring at 364nm, gave pure Tbfmoc-MBD which was then dissolved in guanidine solution and treated with piperidine to remove the Tbfmoc group, before being neutralised (AcOH,0°C), desalted by gel filtration (Sephadex G50, 30% AcOH) and purified to homogeneity by RP-HPLC (Figure 3(b)).

A comparison of the crude Tbfmoc-MBD and the purified protein (Figure 3) illustrates that the deletion sequences, which do not react with reagent (**52**), elute at the same retention time on Hplc as the pure MBD protein. Hence, without use of the Tbfmoc group to aid purification, Hplc purification on its own would have been unable to separate the truncated sequences from the desired product.

Figure 4 shows an SDS polyacrylamide gel of the crude cleaved MBD, Tbfmoc purified material and the deletion sequences which did not react with the reagent (**52**). It is clear from these results that the truncations (lanes 4 and 5) consist mainly of material of molecular weight < 6380 Da.(molecular weight standards lanes 1 and 8). Therefore, reaction of (**52**) with the free amino group of the resin bound MBD is almost quantitative, due to the fact that there are only trace amounts of MBD visible in lanes 4 and 5.

Figure 4. SDS Phastgel (Homogeneous 20), Lanes 1+8 Mwt.Std.
Lanes 2+3 Crude cleaved MBD. Lanes 4+5 Truncations. Lanes 6+7 Pure MBD

From an initial loading of resin bound Fmoc.Arg(Pmc) of 0.14mM(1.0g), the synthesis yielded 2.28g of fully protected resin bound product. Typical yields of protein from the above synthesis gave 30-35mg of pure protein from a 0.5g portion of the fully protected resin bound product. This equates to an overall yield of 13% for over 170 chemical steps.

The purified MBD protein has also been fully characterised by mass spectrometry (Figure 5)[76],amino acid analysis (Table 1) and automated Edman degradation[77] confirmed that the first twenty six residues were in the correct sequence.

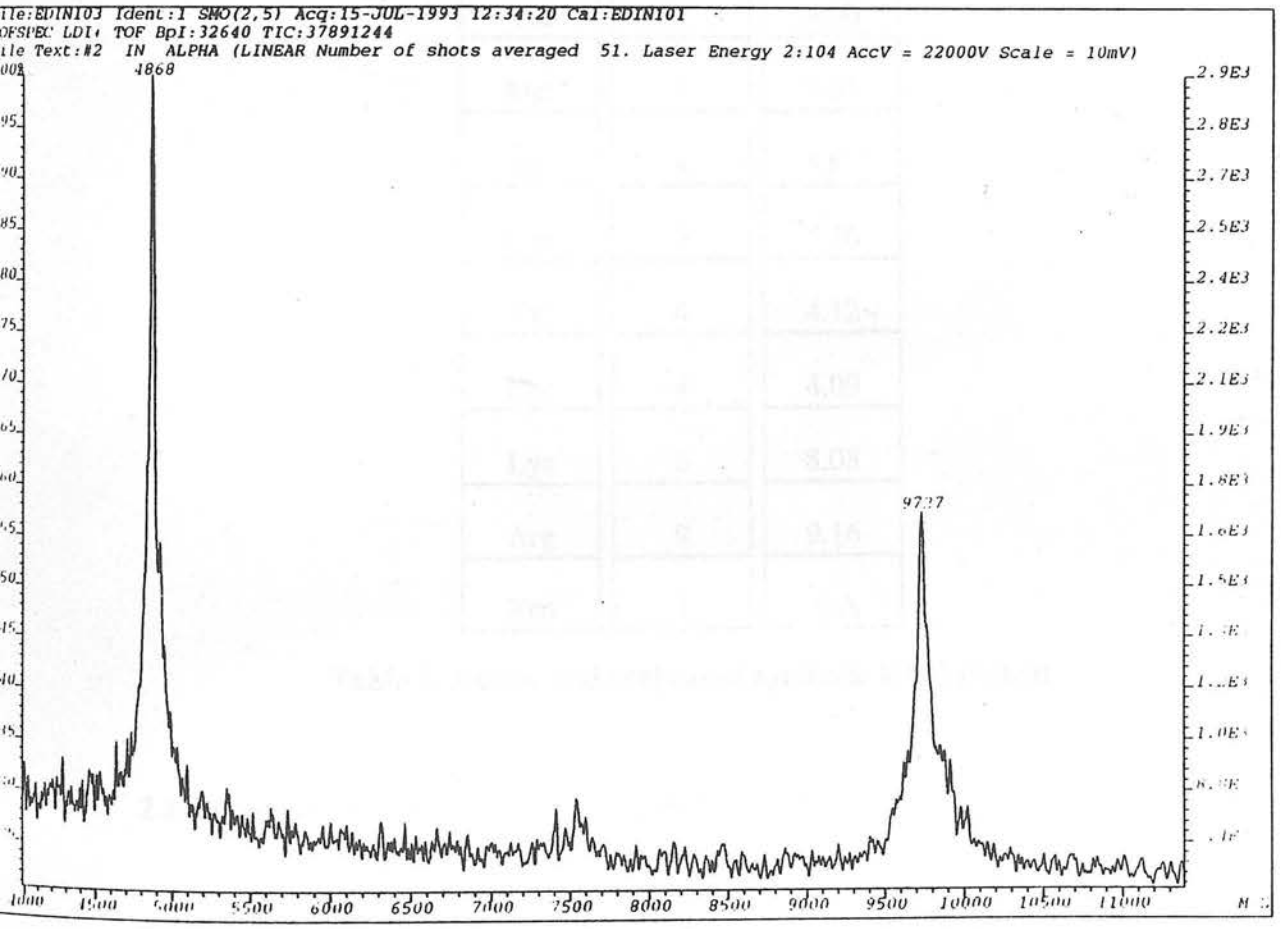


Figure 5. Laser desorption time of flight mass spectrum of synthetic MBD[76]

C_{431} H_{683} N_{126} O_{129} S_1 requires 9726.0 amu

Residue	Expect	Found
Asx	10	9.53
Thr	5	4.44
Ser	7	5.80
Glx	6	7.09
Pro	6	5.82
Gly	7	7.27
Ala	4	4.43
Val	4	4.00
Met	1	1.05
Ile	4	3.87
Leu	5	5.26
Tyr	4	4.12
Phe	4	4.09
Lys	8	8.08
Arg	9	9.16
Trp	1	N/A

Table 1. Amino acid analysis of synthetic MBD protein

2.2.3 Binding of Synthetic MBD to Methylated DNA

The simplest and perhaps the most widely used method for investigating protein-nucleic acid interactions is by use of the gel retardation or band-shift assay[78]. It is based on the observation that binding of a protein to a DNA fragment usually leads

to a reduction in the electrophoretic mobility of the fragment in non-denaturing polyacrylamide or agarose gels.

Figure 6 shows such a band-shift assay obtained for native MBD when incubated with the DNA sequences indicated. This shows that binding of MBD to methylated DNA requires m^5dC (44) to be present symmetrically in the CpG sequence. This assay, when repeated with chemically synthesised MBD, gave identical results to those obtained for the expressed material[68], thus proving that both samples had the same affinity and specificity for DNA sequences containing a symmetrically methylated CpG sequence.

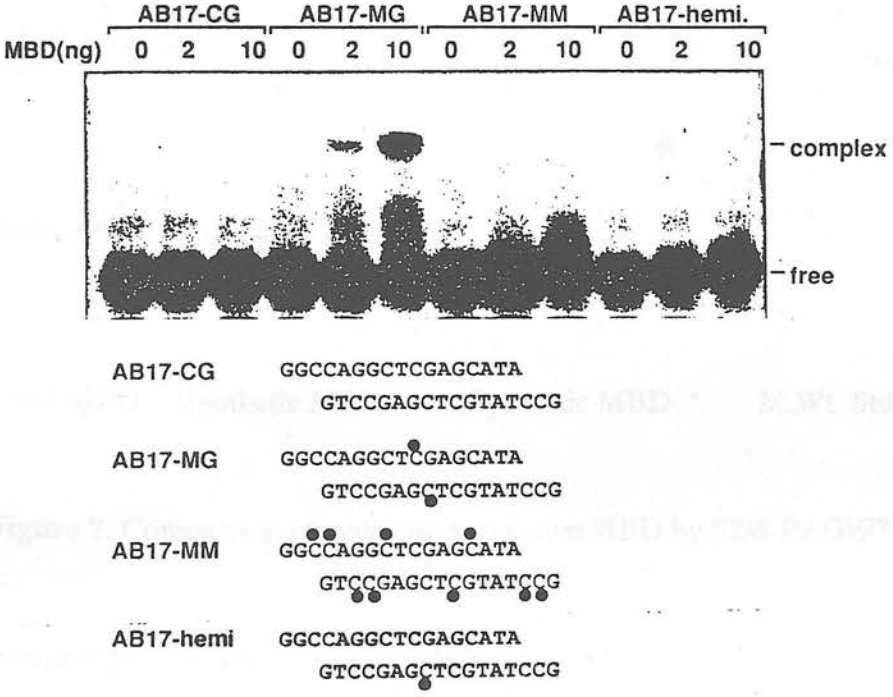
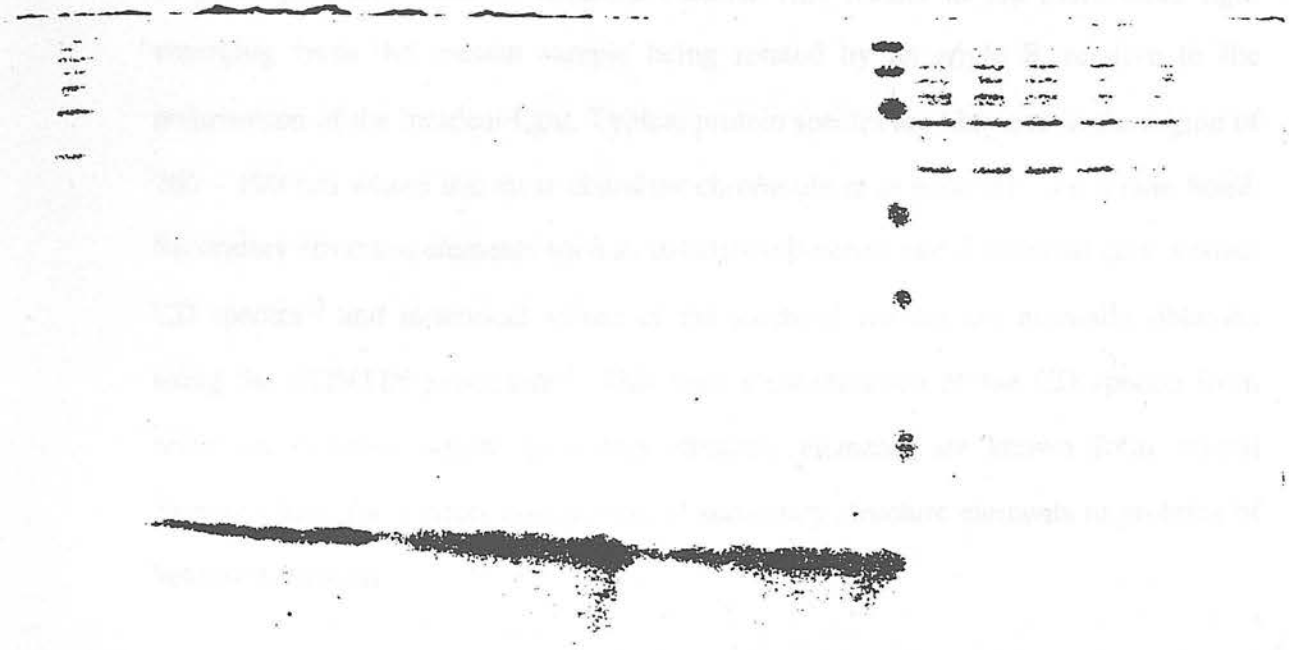


Figure 6.

Surprisingly the chemically synthesised material was much more soluble in aqueous solution than the expressed protein, although comparisons of synthetic and native protein both by band-shift (Figure 6) and denaturing SDS PAGE (Figure 7)[79] seemed to indicate that both materials were essentially identical. One possible explanation for the differing solubilities of the synthetic and native MBD could be due to the fact

that one is folded whilst the other is denatured. Alternatively the solubility of the synthetic MBD may be due to residual trifluoroacetate from Hplc buffers, which results in solutions of acidic pH when the protein is dissolved in water and thus solubilises the basic protein (pI 10.7).



Recombinant MBD Synthetic MBD Synthetic MBD M.Wt. Std.

Figure 7. Comparison of synthetic and native MBD by SDS-PAGE[79]

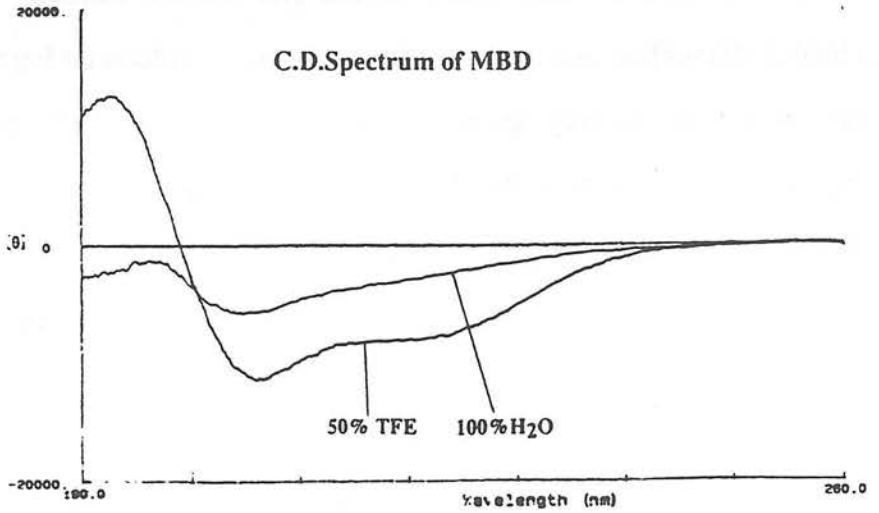
2.2.4 Structural Studies on Synthetic MBD

One of the major goals in the field of protein synthesis, is to be be able to deduce the three dimensional structure of interesting target molecules. Once the desired protein is obtained pure in milligram amounts, this can be approached by utilising a variety of methods such as circular dichroism(CD), nuclear magnetic resonance (NMR) and X-ray crystallography.

2.2.4.1 Circular Dichroism (CD)

CD is useful in providing quantitative information in regard to the secondary structure content of an asymmetric molecule, such as a folded protein. The technique utilises the ability of chiral substances to absorb the right and left hand components of plane polarised light to different extents. This results in the transmitted light emerging from the protein sample being rotated by an angle θ, relative to the polarisation of the incident light. Typical protein spectra are obtained in the region of 260 - 190 nm where the most abundant chromophore is generally the amide bond. Secondary structure elements such as α-helices, β-sheets and β-turns all give distinct CD spectra[80] and numerical values of the stuctural content are normally obtained using the CONTIN procedure[81]. This uses a combination of the CD spectra from reference proteins, whose secondary structure elements are known from crystal structure data, for a direct comparison of secondary structure elements in proteins of unknown structure.

CD measurements on synthetic MBD were carried out at the Scottish CD Facility at Stirling University[82] and the results are given in Figure 8 and Table 2 below.



C.D. measurements were carried out at the Department of Biological and Molecular Sciences, University of Stirling.

Figure 8. CD spectra of synthetic MBD in H_2O and H_2O/TFE (1:1)

Solvent	α-Helix	β-Sheet	Remainder
H_2O pH4.5	8% +/- 0.43%	51% +/- 0.52%	41% +/- 0.83%
50% TFE	30% +/- 1.1%	52% +/- 1.1%	18% +/- 2.0%

Table 2. Predicted structural content of synthetic MBD from CD spectra in the solvents indicated

The sample was prepared simply by dissolving the pure synthetic MBD in water to give a solution of pH 4.5. The CD spectra obtained from this sample (Figure 8) indicated only low levels (8%) of α-helical structure, although addition of 2,2,2-trifluoroethanol (TFE) resulted in induced structure formation which correlated to 30% α-helix. This increase in helicity occurs with no loss of β-sheet structure as calculated by the CONTIN[81] secondary structure prediction procedure. However, previous studies on the ubiquitin system[83,84] have also given increased α-helical content in the presence of TFE, although the validity of these figures remains to be confirmed by other methods.

None the less, it was encouraging that the synthetic protein seemed to have some degree of structure without any attempt being made at folding and, as had been proved by gel retardation assay, the synthetic MBD was sufficiently folded to bind to methylated DNA. Hence it is possible that the protein domain is only loosely structured on its own in solution, but that binding to its recognition site stabilises some of the more flexible regions of the protein due to electrostatic interactions with the methylated DNA.

2.2.4.2 Nuclear Magnetic Resonance (NMR)

NMR spectra of synthetic MBD were obtained from a sample of 25mg at pH 6.3 and at various temperatures. At pH values above this the protein sample very quickly aggregated, and even at pH 6.3 the instability of the sample resulted in poor quality spectra. Although these spectra were not of the quality required to allow sequential assignment using standard techniques[85], they did indicate that the synthetic protein was spontaneously folding in solution.

Recently, the solution structure of recombinant MBD has been solved[86]. An analogue of MBD, extended at the N-terminus with a sequence of amino acids to allow affinity purification and also extended at the C-terminus, was overexpressed and purified in sufficient quantities to allow the determination of its three dimensional structure in solution (Figure 9).

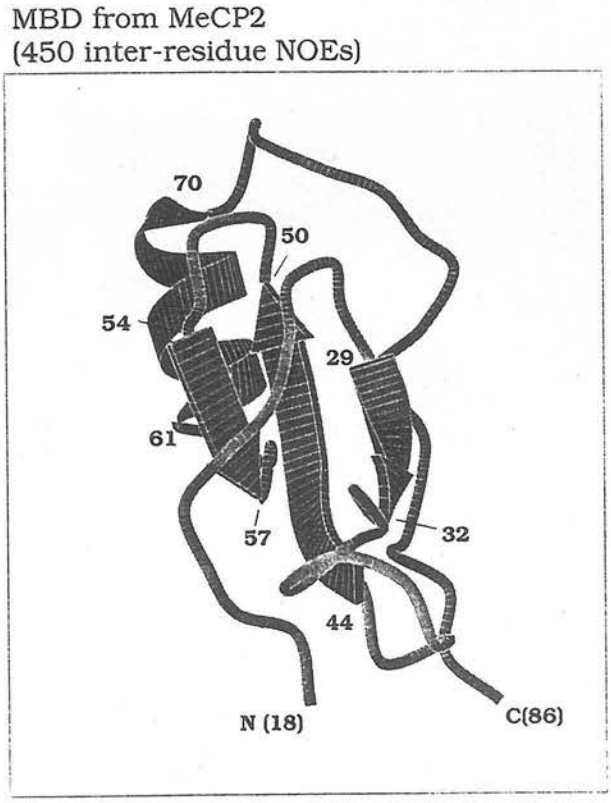


Figure 9. Solution structure of recombinant MBD[86].

The core of the structure is made up of a short α-helix and 3 β-strands, whilst the remainder of the backbone consists of relatively unstructured loops. Both the N and the C-terminal regions of the structure could not be resolved and are probably flexible in solution when not bound to methylated DNA.

A comparison of the synthetic and recombinant protein by 1D NMR (Figure 10) clearly illustrates that both adopt a similar structure in solution. Both spectra have the same pattern of methyl resonances between 0-0.6ppm, where the upfield shift is due to their proximity to electron rich aromatic residues within the protein core. Furthermore, both proteins exhibit similar paterns of signals around 6ppm, due to increased shielding of an aromatic residue in the core, as well as between 7.5 and 9.7ppm due to the backbone amide protons of the folded protein.

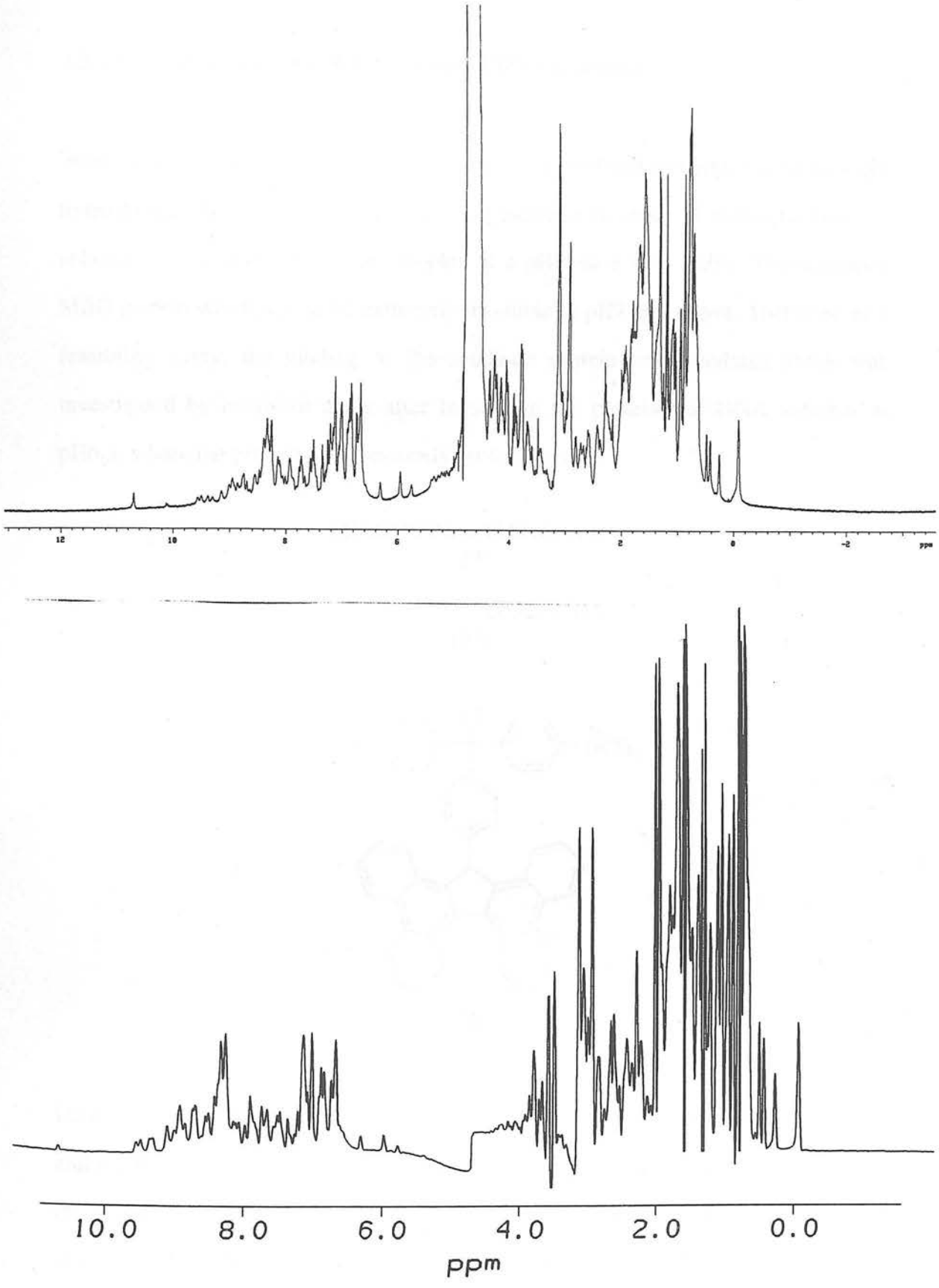


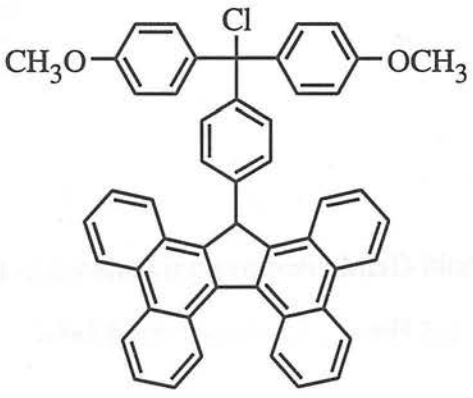
Figure 10. NMR spectra of synthetic MBD (top) and recombinant protein incorporating the MBD domain (bottom).

2.2.4.3 NMR Studies of MBD-Methylated DNA Complex

Whilst it was clear that both the synthetic and recombinant proteins bound strongly to methylated DNA (Figure 6). The limiting factor to the study of such a complex in solution, is the solubility of the complex at a pH where it is stable. The synthetic MBD protein was found to be extremely insoluble at pH7 and above. Therefore as a feasibility study, the binding of the synthetic protein to methylated DNA was investigated by bandshift assay after incubating the protein and DNA together at pH6.3, where the protein had previously proved soluble.

$^{5'}$CAGGCT5mCGAGCATA$^{3'}$
(53)

$^{5'}$TATGCT5mCGAGCCTG$^{3'}$
(54)



(55)

DNA sequences (**53**) and (**54**) were synthesised (OSWEL, University of Edinburgh) and purified using the method of Ramage and Wahl[87a] by labelling the 5' end of the oligonucleotide with 4-(17-tetrabenzo[a,c,g,i]fluorenylmethyl)-4',4''-dimethoxytrityl chloride **55**[87b]. The purified oligonucleotides were then annealed to form the corresponding duplex which was subsequently used in the band shift assay (Figure 10). This confirmed that at pH 6.3, the protein would still bind to the methylated

DNA. Further studies are currently in progress to enable NMR studies on the protein-DNA complex, in an attempt to solve the structure of the methylated DNA binding domain of MeCP2 when bound to methylated DNA.

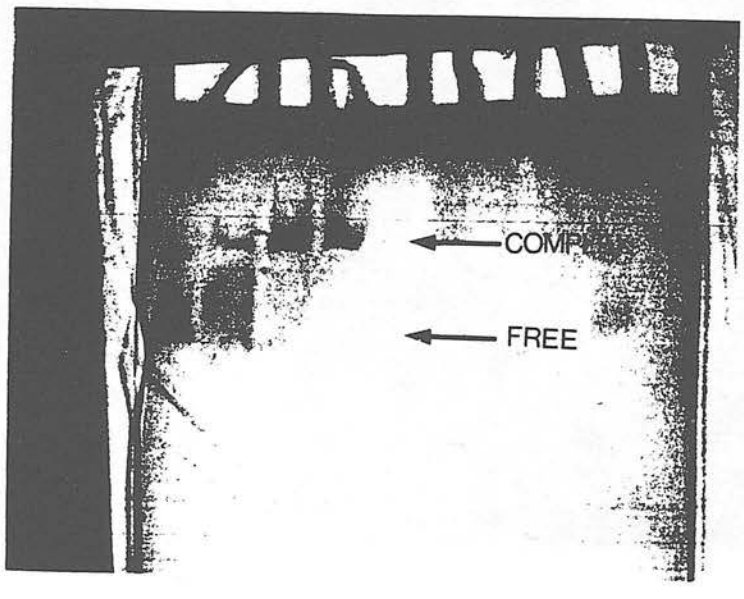


Figure 10. Bandshift assay showing synthetic MBD binding to DNA duplex of **53+54** after incubation at pH 6.3

2.2.4.4 Crystallisation studies on MBD

As part of a collaboration between our group and the laboratory of Dr.L. Sawyer (Department of Biochemistry, University of Edinburgh), crystallisation studies have been carried out on both the synthetic MBD and the protein-methylated DNA complex. Extensive investigations have failed to produce any crystalline MBD protein on its own[88]. However, initial results of crystal screening carried out on the protein-DNA complex have yielded micro-crystals (Figure 11). Although these

micro-crystals are too small to be useful for structure determination, efforts are being made to increase the size of the crystals by varying the conditions of crystallisation[88].

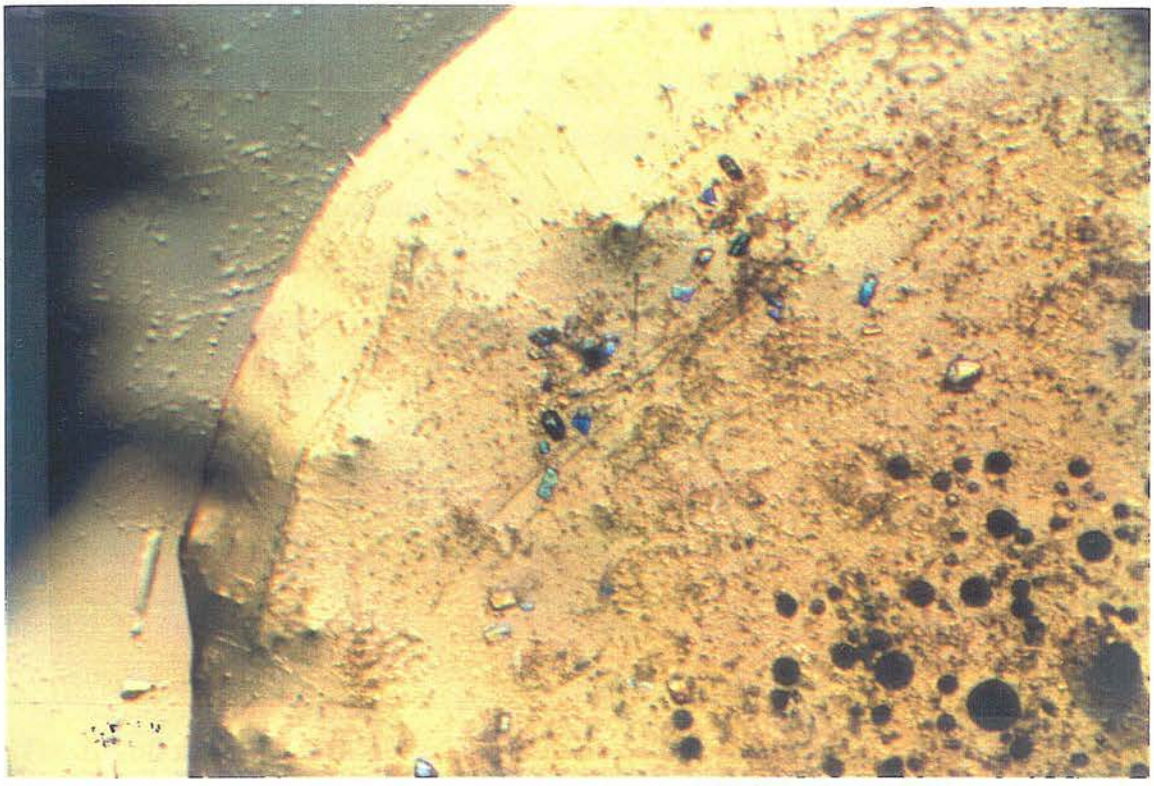
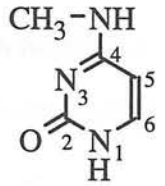


Figure 11. Micro-crystals obtained by hanging drop experiment from synthetic MBD protein-synthetic methylated DNA mixture.

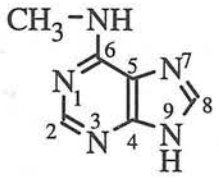
CHAPTER THREE : RESTRICTION ALLEVIATION PROTEIN

3.1 General

Bacteria use three kinds of methylation to protect their DNA against the action of restriction enzymes. As well as 5-methyl-2'-deoxycytidine (**44**), which is widespread in eukaryotes, lower organisms have been found to contain DNA methylated at the N4 position of cytosine (m4C) (**53**) and at the N6 position of adenosine (m6A) (**54**)[89].



(53)



(54)

The *E.Coli* strain K12 chromosome encodes a type I endonuclease known as EcoK, which is a complex multifunctional system made up of three subunits. The subunits are coded for by three contiguous gene sequences known as hsdR (host specific for DNA restriction), hsdM (modification) and hsdS (specificity)[90], which are responsible for the nuclease action, methylation and sequence specificity of the endonuclease.

When foreign DNA from an infecting phage enters bacterial cells, a process called host controlled restriction is initiated by the restriction endonuclease EcoK. In this process the EcoK acts either as a methylase or a nuclease, and the mode of action is triggered by the level of DNA methylation at the EcoK recognition site (Figure 9).

```
      *
5'-AACNNNNNNGTGC-3'
3'-TTGNNNNNNCACG-3'
            *
```

Figure 9. EcoK recognition site

If the recognition site is methylated at the adenines marked with an asterisk (Figure 9), the enzyme recognises the DNA sequence as being host derived and there is no reaction. If neither adenine is methylated the enzyme recognises this as foreign DNA, which signals the enzyme to act as a nuclease and the DNA is cleaved at a site distant from the recognition site. When the recognition site is hemimethylated, i.e. methylated at only one of the asterisked adenines (Figure 9) as would be the case immediately after replication, the EcoK enzyme is triggered into its methylase mode and the unmodified adenine is methylated.

Several phages are known to protect themselves against host controlled restriction by encoding a protein which completely blocks the EcoK activity[91], but phage λ utilises a different mechanism in order to protect its unmodified genome. Zabeau *et al*[92] were the first to report mediation of restriction by the ral gene product of phage λ and also the location of the ral gene in the λ genome, although SDS-PAGE failed to reveal the expected product of the ral gene. Determination of the ral gene sequence[93] identified a 66 amino acid protein as the ral gene product which will subsequently be referred to as Ral for Restriction Alleviation (Figure 10).

MTTTIDKNQWCGQFKRCNGC

KLQSECMVKPEEMFPVMEDG

KYVDKWAIRTTAMIARELGK

QNNKAA

Figure 10. Deduced amino acid sequence of Ral

Loenen and Murray[94] subsequently reported the cloning and expression of the ral gene product, which gave rise to a 1000-fold reduction in restriction and a 100-fold increase in modification *in vivo*. However, attempts to produce Ral protein in amounts sufficient for characterisation and *in vitro* assay of the postulated action of

ral have proved unsuccessful[95]. Therefore, in order to investigate the action of Ral at the molecular level it was decided to attempt the total chemical synthesis of the 66 amino acid protein deduced from the coding sequence (Figure 10).

3.2 Choice of Cysteine Protection in Protein Synthesis

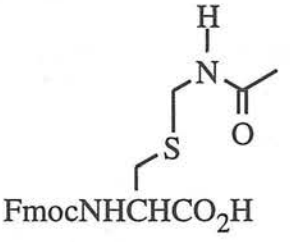
Perhaps one of the most difficult trifunctional amino acid residues to deal with during the chemical synthesis of polypeptides is cysteine. The stepwise SPPS of cysteine containing proteins poses added problems due to the fact that some proteins require the cysteine residues to be in the free thiol form, while others form disulphide bonds between cysteine residues in order to stabilise secondary/tertiary structural conformations.

The choice of Fmoc-cysteine derivative will vary depending on the synthetic strategy. However, synthesis of cysteine containing peptides usually involves either cleavage of the thiol protecting group as the peptide is removed from the solid support, or later by an orthogonal method once the polypeptide has been purified to some extent. Alternatively, a combination of temporary and semi-permanent cysteine protecting groups offers the possibility of regioselective disulphide formation[96].

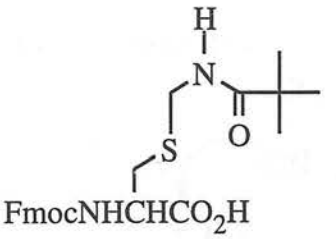
Opinion seems to be divided on the best approach towards the batch SPPS of cysteine containing proteins. Boc based strategy has in the past relied mainly on temporary thiol protection which is removed on cleavage from the resin. The crude protein can then be folded and disulphides formed using a variety of techniques[97], before being purified to homogeneity. This approach has been successfully demonstrated for the synthesis of hTGF-α[98] (50 residues, 3 disulphide bridges), Interleukin-3[99] (140 residues, 2 disulphide bridges), Interleukin-8[51,100] (72 residues, 2 disulphide bridges) and neutrophil activating peptide 2[100] (70 residues, 2

disulphide bridges).

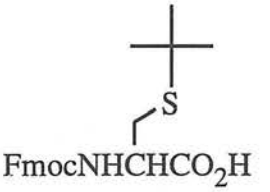
The Fmoc based stepwise SPPS of large, cysteine containing polypeptides has on the other hand relied on the use of either the semi-permanent acetamidomethyl (Acm)[101] (**55**) protecting group, as described for the synthesis of hTGF–α[102] (50 residues, 3 disulphide bridges) or, a combination of Acm (**55**), with the semipermanent Tacm (**56**) and t-Butyl (**57**), as reported recently by Akaji *et al* for the synthesis of human insulin[96] (51 residues, two peptide chains, 3 disulphide bridges). It cannot be stressed too strongly that a pre-requisite for the use of such a regioselective disulphide bridge forming strategy is that in subsequent chemical manipulations, the disulphide bridge or bridges are not exposed to conditions which promote disulphide exchange, scrambling, cleavage or modification[103].



(55)



(56)



(57)

3.3 Chemical synthesis of the *ral* gene product of phage λ

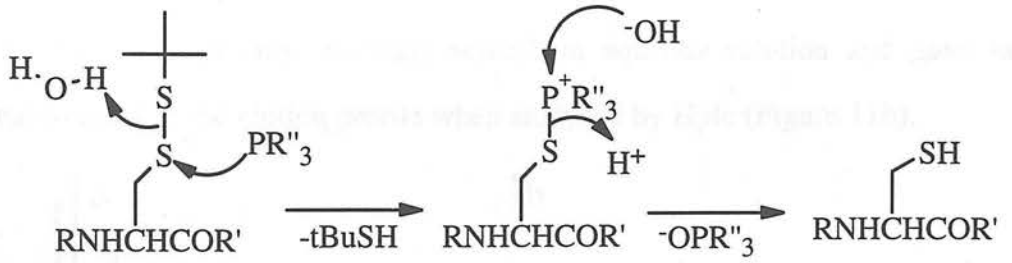
As has been stated previously (Section 3.1), some proteins require free thiols for biological activity, whilst others require disulphide bridges. In the case of Ral, little was known about whether the cysteine residues were reduced or disulphide bridged, although initially it was proposed[95] that Ral contained a putative zinc finger-like motif[104]. The thiols in Ral protein would therefore need to be present in the reduced

state in order to be able to co-ordinate tetrahedrally to a Zn^{2+} ion.

Ral has subsequently been chemically synthesised four times, and each time the thiol protection has been varied in order to assess the best general approach for the synthesis of multiple cysteine containing proteins. The various synthetic approaches are discussed in the following chapters.

3.3.1 Synthesis A - Cys S-S^{t}Butyl protection.

For the initial synthetic strategy of Ral, it was hoped that by using the acid and base stable t-butylsulphenyl (S^{t}Bu) derivative (Scheme 14), it would be possible to purify the protein with the cysteine thiol groups fully protected. Then, as the final step it was planned to effect removal of the S^{t}Bu cysteine protection by treating the pure protein with thiols[105] or phosphines[106] (Scheme 14).



Scheme 14. Removal of StBu group from cysteine by the action of trialkylphosphine/water.

Starting from an inital functionality of 0.2mM resin bound Fmoc alanine, the desired sequence (Figure 10) was assembled on an Applied Biosystems 430A peptide synthesiser. All residues were treble coupled, with the exception of glycine which was single coupled *via* the symmetrical anhydride. Monitoring of the coupling efficiency during synthesis by UV (Section 2.2.2), indicated a low coupling of

lysine^{60} to glutamine^{61}, although the remainder of the synthesis proceeded without further major drops in coupling.

The cysteine protected protein was obtained by acidolytic cleavage using a mixture of TFA, EDT, phenol and water for five hours under an atmosphere of nitrogen, followed by evaporation of the TFA and trituration with diethyl ether to give the crude protein which was isolated by filtration. The crude protein was then applied to the top of a column of sephadex G-50 and eluted with aqueous acetic acid (30%). The protein containing fractions were then pooled and lyophilised to give crude StButyl protected Ral protein. This material proved to be only slightly soluble in aqueous solution and when analysed by Hplc (Figure 11a), eluted as a broad peak which suggested the presence of several similar species. In an attempt to improve the solubility of the material, the hydrophobic S^{t}Bu cysteine protecting groups were removed by treating a solution of the protected protein in 95% trifluoroethanol/5% water with tributylphosphine (Scheme 14). However, the crude fully reduced protein obtained remained only sparingly soluble in aqueous solution and gave no real improvement in the elution profile when analysed by Hplc (Figure 11b).

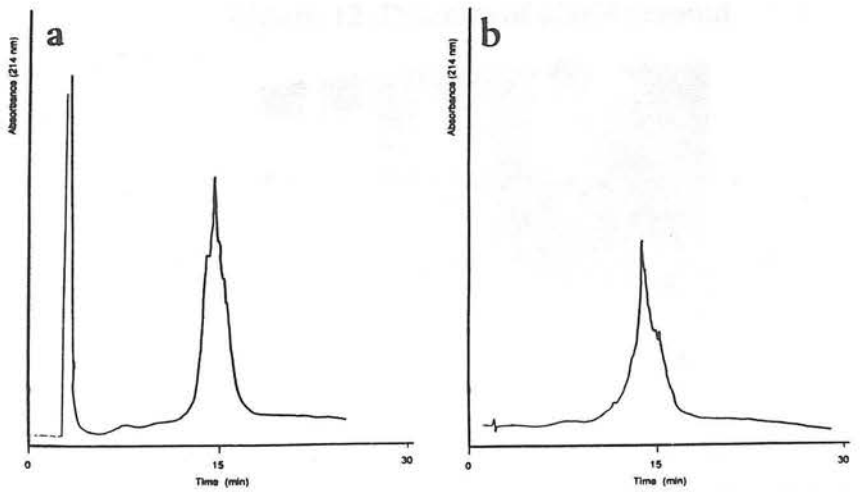


Figure 11. Hplc analysis of crude (a) StButyl protected and (b) reduced Ral from synthesis A

Attempts to further purify this material by ion exchange chromatography failed to give any separation as most of the material failed to bind to either cation or anion exchangers at various pH values, but simply eluted in the void volume. However, a sample of this crude synthetic Ral protein was subsequently used to raise anti-Ral antibodies, which were used to monitor expression of recombinant Ral from plasmid pGK4 (Figure 12) by western blot analysis (Figure 13). This then allowed comparison of the synthetic and recombinant material (Section 3.3.2.1).

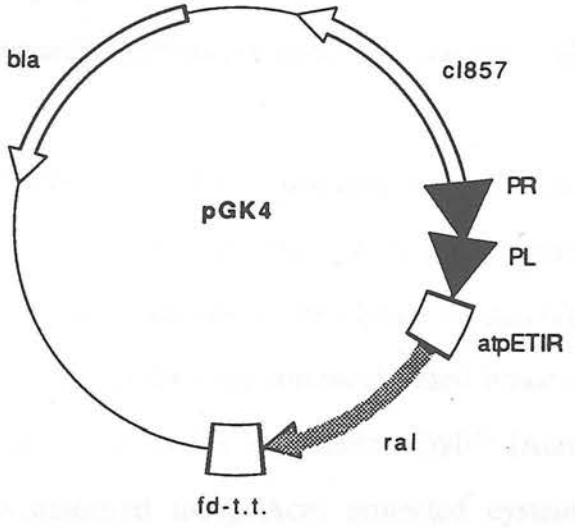
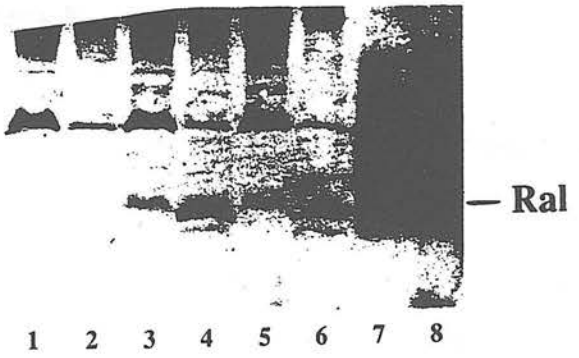


Figure 12. Diagram of pGK4 plasmid



Lanes 1&2 - products from plasmid without Ral gene
Lanes 3&5 - products from plasmid pGK4 before heat induction
Lanes 4&6 - products from plasmid pGK4 after heat induction
Lanes 7&8 - synthetic Ral

Figure 13. Western blot of synthetic and recombinant Ral (from pGK4)

3.3.2 Synthesis B - Cys(Acm) protection

Insolubility of synthetic peptides can be due to many primary and secondary structural factors such as isoelectric point or aggregation at high concentration. However, in the case of Ral it was felt that the insolubility of the fully reduced crude protein, obtained from synthesis A, could only mean that cleavage from the resin with TFA had not removed all of the side chain protecting groups incorporated during the assembly phase. One notoriously difficult group to remove using TFA is the 4,4'-dimethoxybenzhydryl (MBH) group (**36**) used to block the carboxamide side chains of asparagine and glutamine, of which Ral contains eight in total.

Funakoshi *et al*[107] have reported the use of trimethylsilyl bromide (TMSBr)/TFA to effect final cleavage and side chain deprotection of human pancreastatine-52 (52 residues, 6 of them glutamine). However, the choice of cleavage conditions also dictates the synthetic strategy, as the only commonly used Fmoc cysteine derivative which is stable to TMSBr/TFA[43] is the acetamidomethyl[101] (Acm) derivative (**55**). Ral was therefore resynthesised using Acm protected cysteine and optimised synthetic cycles in order to overcome the areas of the synthesis which had proved problematic during synthesis A.

The first twelve residues (55-66) were coupled in a glass vessel, immersed in an ultrasound bath, and connected to the ABI430A peptide synthesiser to allow delivery of reagents. The protected peptide-resin was then transferred back into a conventional reaction vessel, before the assembly was allowed to continue until the end of the synthesis. At the completion of the assembly phase, the N-terminal Fmoc group was removed by treatment with piperidine and replaced by the afforementioned Tbfmoc group (**51**)[73,74]. The Tbfmoc labelled, Acm protected protein was then cleaved from the resin by treatment with

TMSBr/EDT/thioanisole/m-cresol/TFA for 5 hours at 0°C, under an atmosphere of dry nitrogen and protected from the light. After rapid evaporation of the TFA and precipitation with diethyl ether, the crude Tbfmoc Acm-protected Ral protein was applied to the top of a column of sephadex G-50 and eluted with aqueous acetic acid (30%). The fractions containing Tbfmoc labelled protein were then pooled (Figure 14a) and the pure Tbfmoc Acm-protected Ral was isolated by Hplc (semi-prep Vydac C18, λ364nm). Pure Acm protected Ral protein was then obtained after removal of the Tbfmoc group with piperidine and isolation of the protein derivative by Hplc (Figure 14b). The pure cysteine protected protein was then characterised by mass spectrometry (Figure 15) and amino acid analysis.

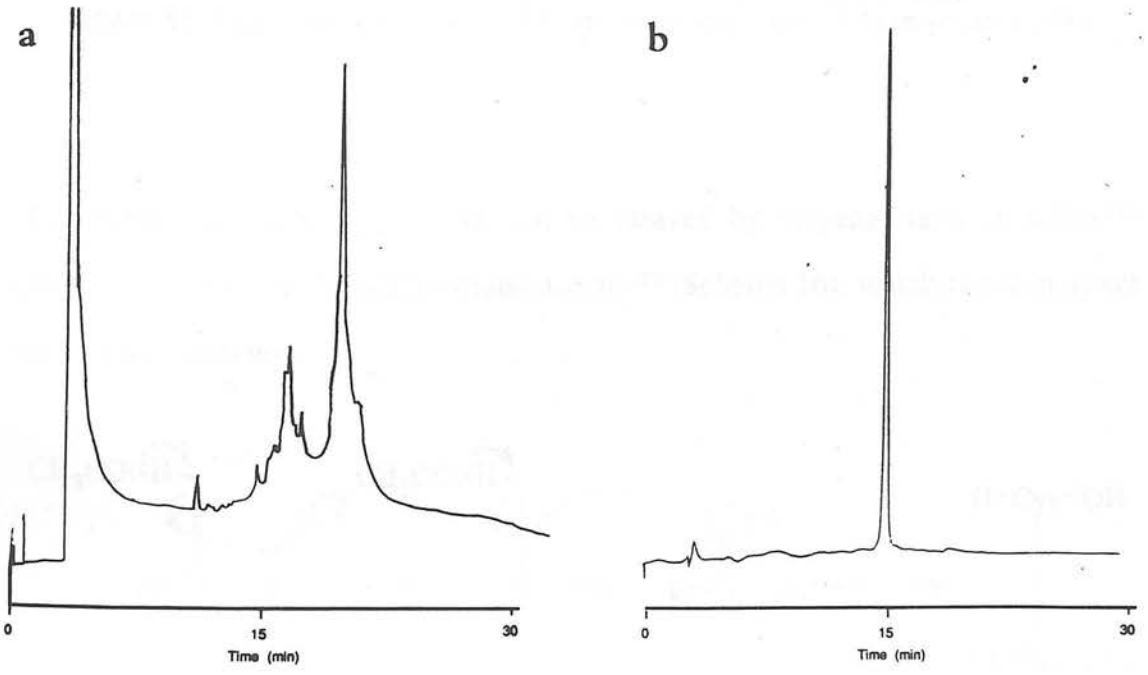


Figure 14. Hplc of (a) crude Tbfmoc Ral(Acm)$_4$ after size exclusion chromatography and (b) pure Ral(Acm)$_4$

Ral (Cys-Acm)4 M.Wt.=7888

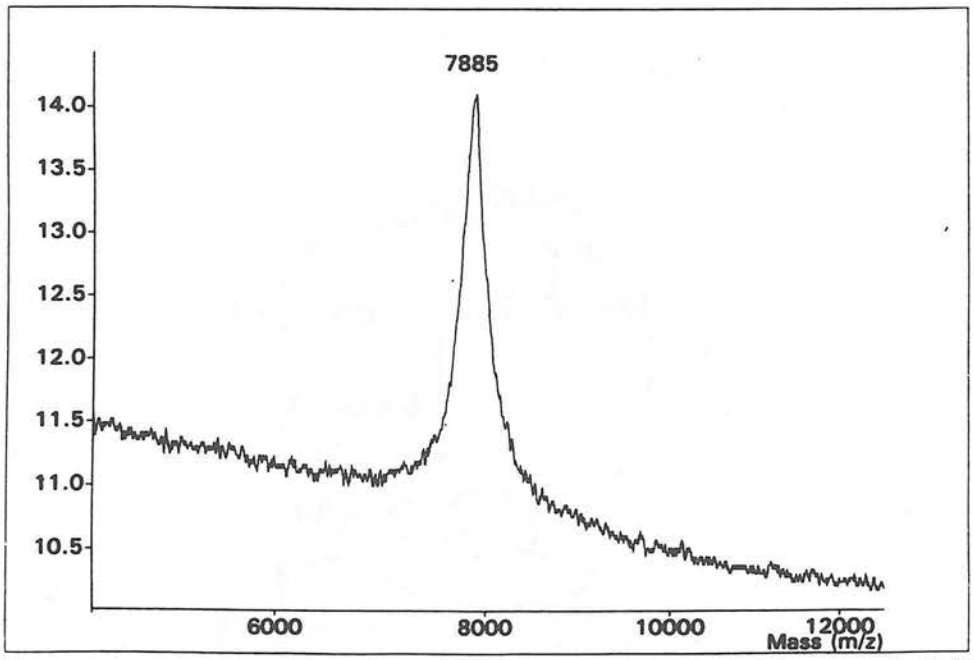
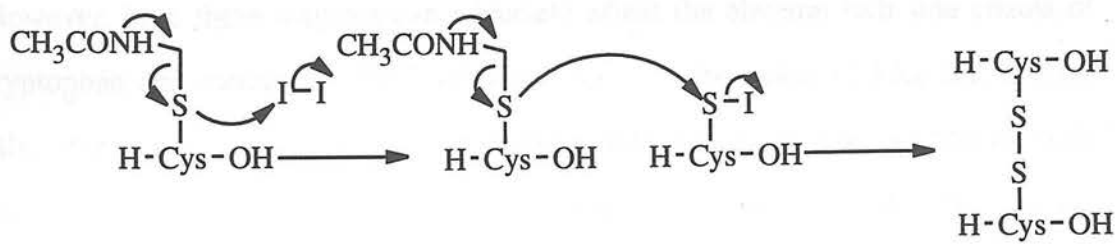
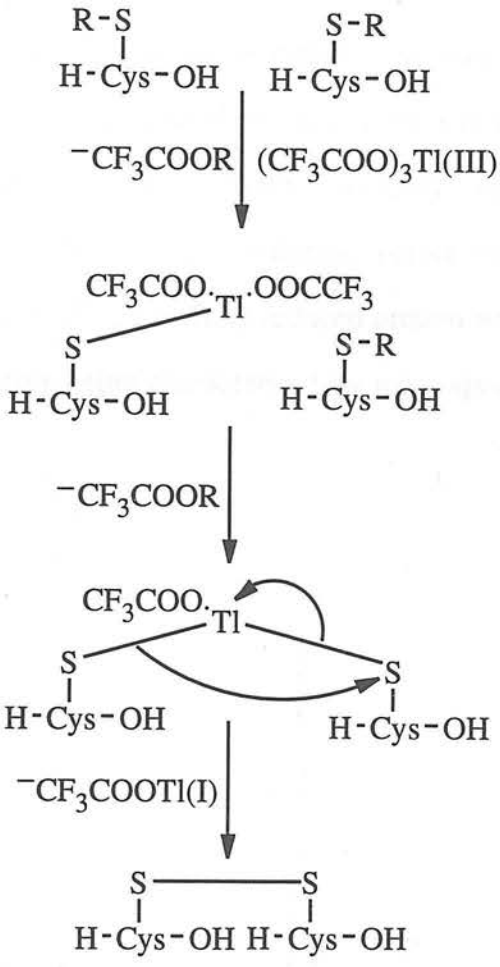


Figure 15. Laser desorption time of flight mass spectrum of Acm protected Ral

The stable Acm protecting group can be cleaved by reagents such as iodine[108] (Scheme 15) and thallium(III) trifluoroacetate[109] (Scheme 16), which result in direct disulphide formation.



Scheme 15.



Scheme 16

However, both these reagents can adversely affect the electron rich side chains of tryptophan and methionine[109,110], of which Ral contains seven (5 Met and 2 Trp). Also, due to the fact that we required the fully reduced form of Ral, in order to study its zinc binding properties, we were limited to the use of either silver trifluoromethanesulphonate[111] or mercury (II) acetate[101] to remove the thiol protecting groups. Several proceedures for the cleavage of the Acm groups were examined using both the above reagents, and it was found that mercury acetate was superior for accomplishing complete removal of the Acm groups, as determined by mass spectrometry of the reaction products.

Hence the conditions used to remove the stable Acm groups on a large scale were by treatment of the pure Acm protected protein at pH 4.0 with mercury (II) acetate, followed by treatment of the resulting mercury salt of the protein with mercaptoethanol under denaturing conditions, before desalting by gel filtration (sephadex G-50, 30% AcOH). The fully reduced protein was then further purified by Hplc (Figure 16), before being characterised by mass spectrometry (Figure 17) and amino acid analysis.

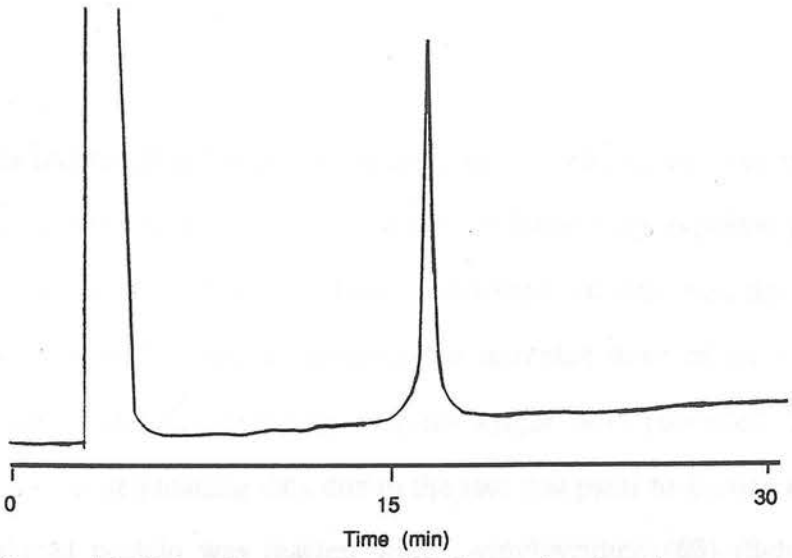


Figure 16. Hplc of pure reduced Ral protein after Acm removal.

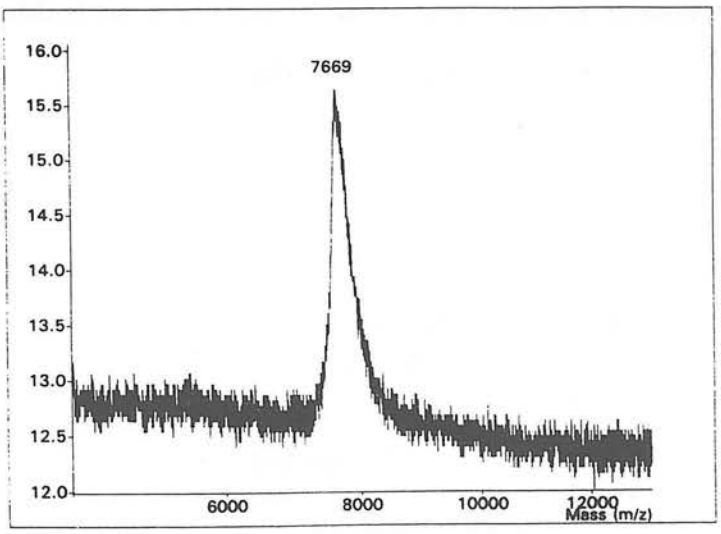
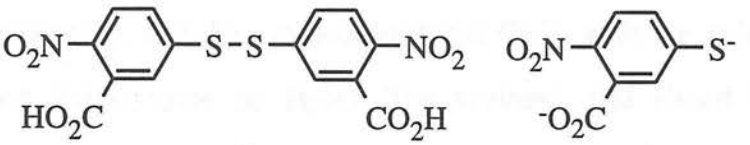


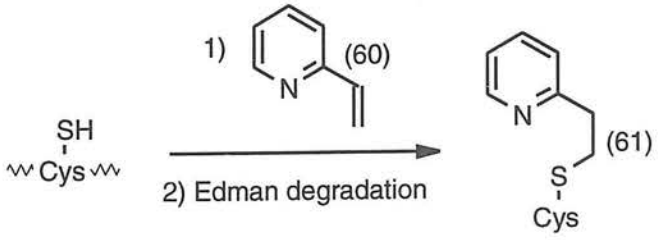
Figure 17. Laser desorption mass spectrum of reduced Ral protein.

However, assay of the number of free thiols present in the protein using the method of Ellman[112] gave only 1.0 thiol group per molecule. Ellman's reagent, 5,5'-dithiobis(2-nitrobenzoic acid) (DTNB) (**58**), reacts with thiol groups and releases a chromophoric anion (**59**), which can be exploited to enable spectroscopic measurement of the number of thiol groups present in the protein of interest.



(58) (59)

However, this technique is limited to the detection of unhindered, solvent accessible thiols and hence some proteins are known to give lower than expected values[113,114]. N-terminal sequencing, on the other hand, confirmed not only that the first twenty two residues were in the correct sequence, but also that three of the four cysteine residues present in the Ral sequence were no longer Acm protected. This can be confirmed from the sequencing data due to the fact that prior to Edman degradation, the fully reduced protein was reacted with 2-vinylpyridine (**60**) (Scheme 17) to produce the pyridylethylated cysteine derivative (**61**). Compound (**61**) was subsequently identified as a product of the degradation by comparison with an external mixed standard containing (**61**).



Scheme 17.

3.3.2.1 Comparison of synthetic and recombinant Ral protein

At this point in time, some recombinant protein, produced from plasmid pGK4 (Section 3.3.1) and which cross reacted with anti-Ral antibodies, became available[95] and was compared with the synthetic protein. Figure 18 shows an overlay of the Hplc traces obtained for the recombinant protein, both before and after reduction/denaturation, and the synthetic material (N.B. both the reduced and fully Acm protected Ral co-elute on Hplc). The synthetic Ral eluted later than the recombinant material and, even after denaturation and reduction of disulphide bridges, the recombinant material was much less hydrophobic than the synthetic Ral, hence proving that the difference in retention times was not due to the recombinant material being folded differently. Therefore, in an attempt to identify if the recombinant material was actually the fully extended protein, it was subjected to analysis by N-terminal sequencing, amino acid analysis and mass spectrometry.

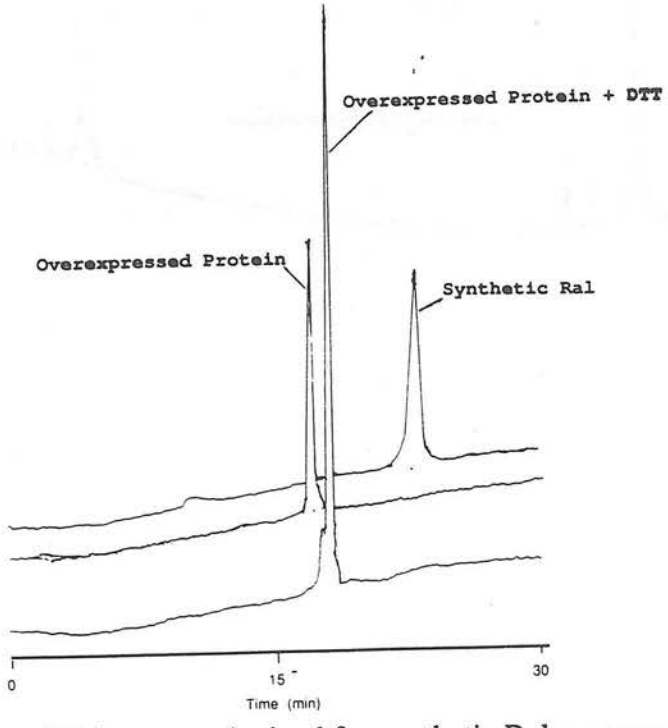


Figure 18. Overlay of Hplc traces obtained for synthetic Ral compared to the traces obtained for the recombinant protein before and after reduction.

Amino acid analysis of the recombinant material indicated that the value for alanine was lower than expected for the predicted sequence of Ral (Figure 10), which contains 5 alanine residues in the first 20 amino acids of the hydrophobic C-terminus. Attempts to perform N-terminal sequencing on the material[115] produced no degradation and therefore indicated that the N-terminus was blocked. However, mass spectral analysis of the recombinant protein[115] before and after reduction and pyridylethylation (Figure 19) indicated that the material was of lower molecular weight than expected for the predicted sequence (7605amu), although it was clear that it did contain 4 cysteine residues due to the mass increase obtained on pyridylethylation (Figure 19). Hence from the combination of the preceeding evidence, the recombinant protein was identified as N-formyl Ral(1-50) (requires MH^+= 5935, found = 5938.5) which had been isolated in an oxidised, disulphide bridged form.

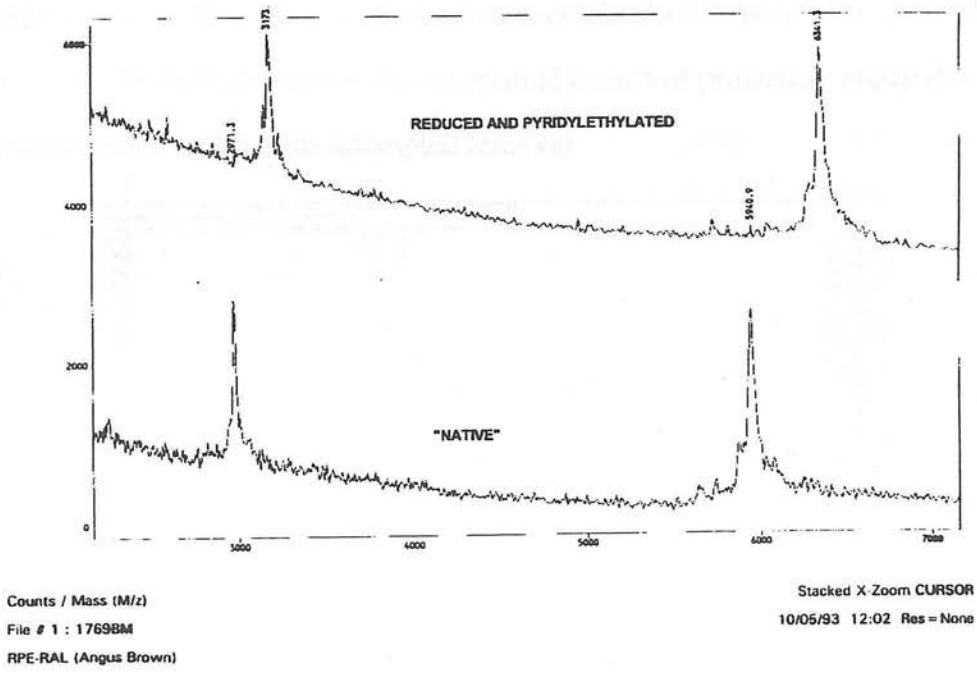


Figure 19. Laser desorption mass spectrum of recombinant protein before (bottom) and after (top) reduction and pyridylethylation.

Attempts to introduce Co^{2+} ions into the protein, as a test of whether the protein adopted a classical zinc finger structure, failed to give any UV absorbance characteristic of a tetrahedral Co species[116] with either the synthetic Ral protein or the reduced recombinant N-formyl(1-50). Similarly, attempts to assay for the expected biological action of Ral gave no increase in the level of DNA modification for either the synthetic Ral or the recombinant material in the presence of various DNA substrates and the EcoK enzyme[117].

However, subsequent to this work being carried out, a sample of the synthetic Ral protein was re-analysed on a mass spectrometer[76] of higher resolution to that used previously. The spectrum produced (Figure 20) clearly indicates that, although there is a large proportion of fully deprotected Ral present (requires MH^+ = 7605, found MH^+ = 7620), there is an approximately equal amount of mono-Acm protected Ral present (requires MH^+ = 7676.1, found MH^+ = 7694). This therefore suggested that although the cysteine protected protein had been obtained in pure form after Tbfmoc purification, the Acm group was not an optimal choice of protecting group due to the difficulties encountered in its attempted removal.

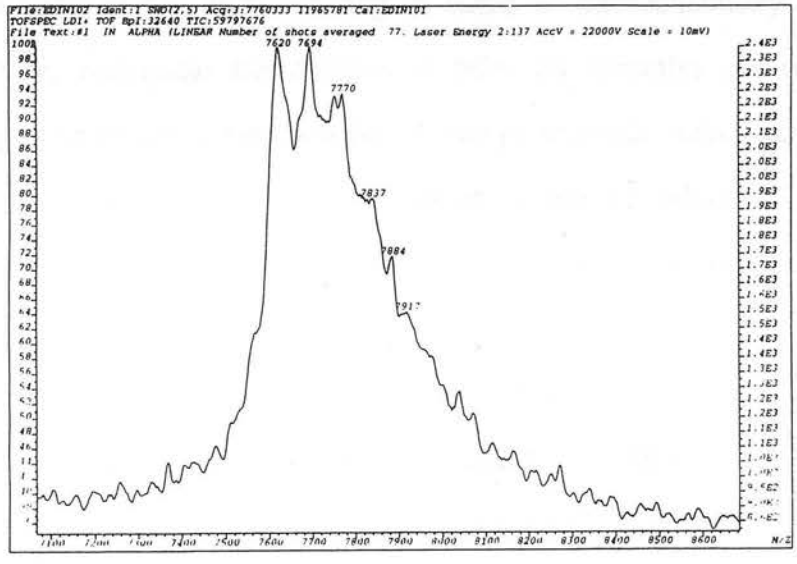


Figure 20. Laser desorption mass spectrum[76] of synthetic Ral protein showing incomplete removal of Acm protection.

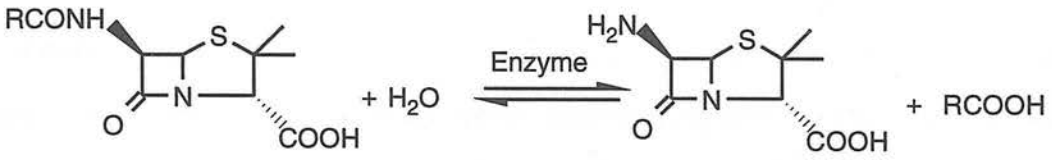
3.3.3 Synthesis C - Design of an enzyme cleavable cysteine protecting group and investigations of its application to the synthesis of proteins.

As has been described above, our synthetic route to cysteine protected Ral protein, using Acm protecting groups and the Tbfmoc N^{α} protecting group to aid purification, successfully results in the rapid isolation of the cysteine protected protein. However, the Acm group is not removed easily using literature procedures[101,111], as has been confirmed previously, both in these[118,119] and other laboratories[120,121]. Therefore it seemed likely that if the thiol group of cysteine could be protected with a group which incorporated the stability and solubility properties of the Acm group, but which at the same time could be unambiguously removed under the mildest of conditions, then this would aid the synthesis of cysteine containing proteins.

Recently there has been a renewed interest in the application of biocatalysts to the problem of functional group protection, for use in the field of peptide synthesis[122] and in other areas[123,124]. Because enzymes generally operate under physiological conditions and are highly discriminating in terms of stereochemistry, regio and chemoselectivity, enzymatic transformations offer an attractive solution to the problems of protecting group manipulation. At the present time, several enzymes are commercially available in pure, stable forms, some of which are supplied immobilised on a solid support to aid handling and recovery after use.

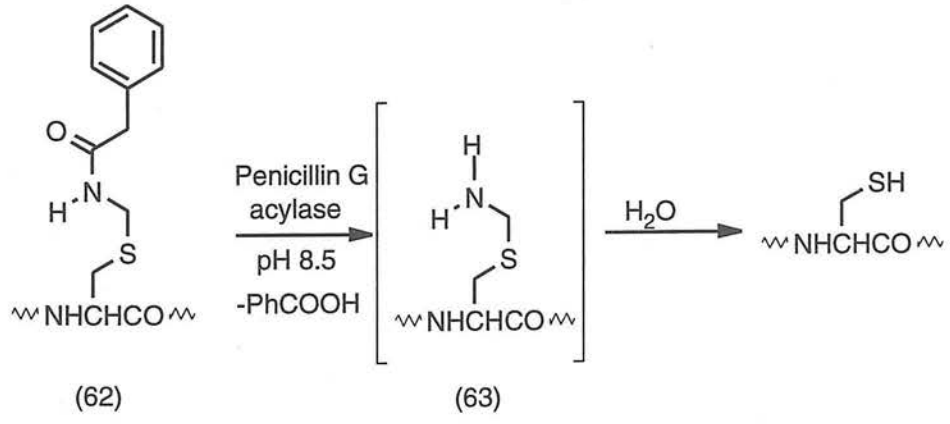
Penicillin G acylase (E.C. 3.5.1.11) from *Escherichia coli* is an enzyme that has found wide spread use in the industrial preparation of semi-synthetic penicillins (Scheme 18). This enzme exists as an 80KDa heterodimer, which is made up of A and B polypeptide chains which contain 209 and 566 amino acids respectively, and has been shown to hydrolyse phenylacetyl amides with no specific requirements for

the amide portion of the substrate[124,125,126].



Scheme 18. Penicillin acylase catalysed reversible reaction

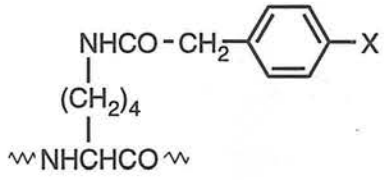
Its use in peptide synthesis was first reported by Brtnik *et al*[127], who utilised the enzyme to remove the N^{ε}-phenylacetyl group from the side chain of lysine during the solution synthesis of vasopressin (9 residues). Greiner and Hermann[128] subsequently extended the use of penicillin acylase to deblock the side chain of phenylacetamidomethyl (Phacm) protected cysteine residues (**62**) during the synthesis of oxidised glutathione. The enzyme mediated cleavage reaction (Scheme 19) initially generates an intermediate S-aminomethyl compound (**63**) (Scheme 19), on cleavage of the acyl component, which spontaneously hydrolyses to liberate the free thiol. The phenylacetamidomethyl (Phacm) protecting group was subsequently applied by Royo *et al*[129] to the solid phase synthesis of peptides, where they reported that the Phacm group had the same stability/lability as Acm, but that Phacm could additionally be cleaved by the action of penicillin G acylase.



(62) (63)

Scheme 19.

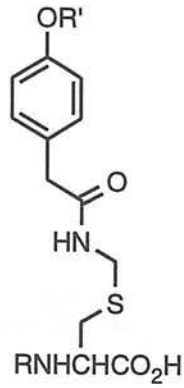
This methodology was therefore attractive to us due to our desire to carry the cysteine protected peptide through to the final stage of purification, before deblocking the thiol functional groups under mild conditions. Previous work in this laboratory had identified several problems in the application of this phenyacetylamide protecting group when used to selectively protect the N^{ε} side chain of lysine (**64**)[130]. It was found that the insolubility of peptides containing this phenylacetylamide group caused problems both in purification and final enzyme mediated cleavage. Modification of the protecting group (**64**) by introduction of a para hydroxyl group (**65**) was shown to impart greater solubility to the protected peptide and also resulted in a substrate which was hydrolysed at a faster rate by the enzyme. Therefore it was also desireable to include this feature in any subsequent enzyme cleavable protecting group for cysteine, based on the Phacm group.



(64) X=H
(65) X=OH

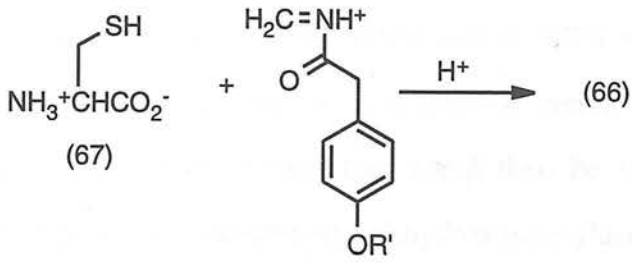
3.3.3.1 Synthesis and evaluation of an enzyme cleavable cysteine protecting group for use in solid phase peptide synthesis.

Incorporation of the cysteine derivative (**66**) during the assembly phase of automated SPPS requires that the phenolic hydroxy group is blocked by a protecting group (R') which should ideally be removed during TFA cleavage of the peptide from the resin at the completion of synthesis.



(**66**)

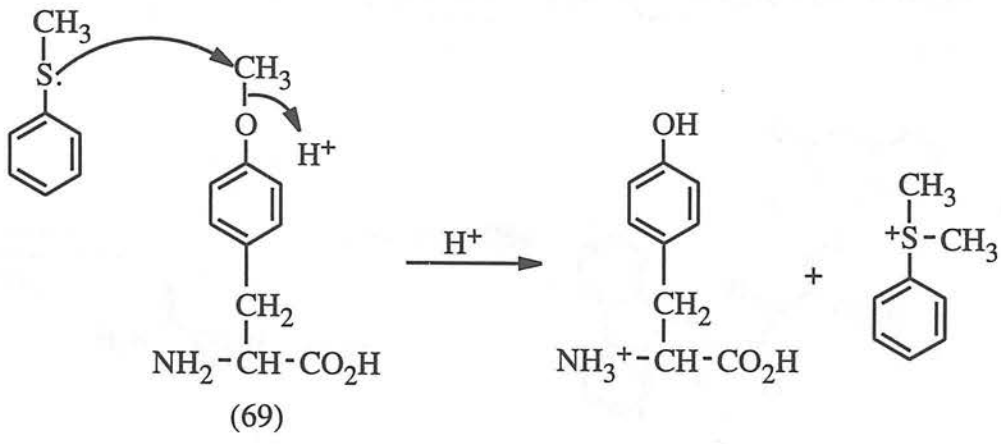
However, the synthesis of such a protected cysteine derivative generally involves reaction of a cysteine thiol (**67**) with the stabilised cation (**68**) under acidic conditions (Scheme 20).



Scheme 20.

Hence, any protecting group for the phenolic oxygen of (**66**) must therefore be sufficiently acid stable to survive the conditions required for formation of (**68**) but, ultimately must be removable at a later stage after incorporation into the polypeptide.

Kiso *et al*[131] utilised the O-methyltyrosine derivative (**69**) as an acid stable phenol protected amino acid during the synthesis of an enkephalin analogue. Removal of the methyl ether protecting group at the completion of synthesis was then accomplished using a thioanisole-trifluoromethanesulphonic acid (TFMSA) system, via a "push-pull" mechanism (Scheme 21).

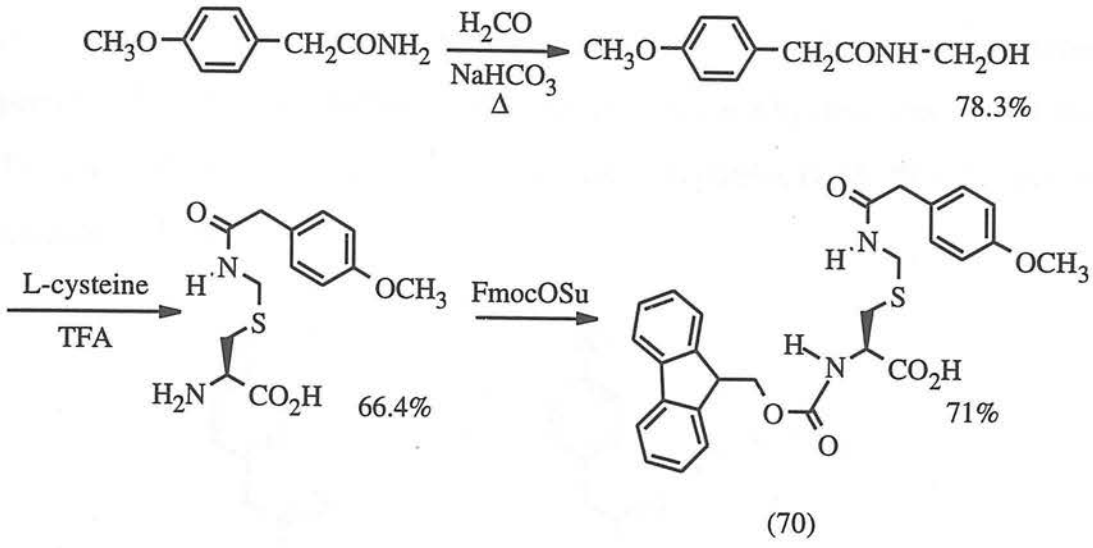


(69)

Scheme 21.

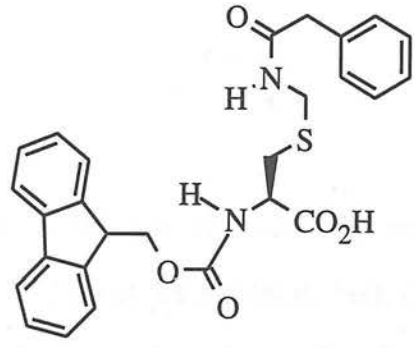
It was therefore decided to synthesise the 4-methoxyphenylacetamidomethyl (4-MeOPhacm) protected cysteine derivative (**66**) with the expectation of incorporating the protected amino acid (**66**) into the assembly phase of SPPS which, on cleavage from the resin with TFA would yield the polypeptide containing 4-MeOPhacm protected cysteine residues. The methyl ether could then be removed using the method of Kiso[131] to give the corresponding 4-hydroxyphenylacetamidomethyl (4-OHPhacm) cysteine protected peptide, which was the desired target molecule to allow investigation of the enzyme mediated deprotection of the 4-OHPhacm protected cysteine residues.

The Fmoc N^{α} protected 4-methoxyphenylacetamidomethyl cysteine derivative (**70**) was obtained as outlined in Scheme 22 in an overall yield of 37%. The corresponding phenylacetamidomethyl protected cysteine derivative (**71**) as described by Royo *et al*[129], was also prepared to allow a direct comparison of the properties of (**70**) and (**71**) once incorporated into polypeptides, and on subsequent treatment with penicillin G acylase.



(70)

Scheme 22.



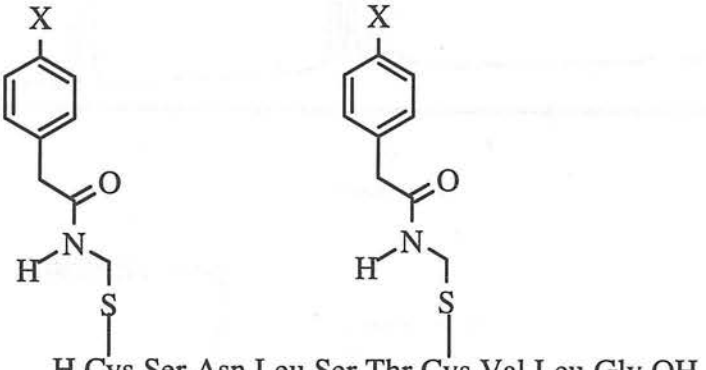
(71)

H.Cys.Ser.Asn.Leu.Ser.Thr.Cys.Val.Leu.Gly.OH

(72)

Both cysteine derivatives (**70**) and (**71**) were then incorporated into a decapeptide test sequence, corresponding to the N-terminal residues of salmon calcitonin (SC 1-10) (**72**). Synthesis of the cysteine protected SC 1-10 was carried out on an ABI 430A peptide synthesiser, and the products from both syntheses were then cleaved

from the resin by treatment with TFA/EDT/water to give the cysteine protected peptides (**73**) and (**74**). Derivative (**75**) was then obtained by treatment of (**74**) with Thioanisole/TFMSA[131] followed by gel filtration (Sephadex G-15, 30% AcOH) and isolation by Hplc.



(**73**) X=H
(**74**) X=OMe
(**75**) X=OH

Each of the SC 1-10 derivatives (**73**), (**74**) and (**75**) were dissolved/suspended in ammonium acetate buffer (0.1M) at pH 8.5 (N.B. both (74) and (75) were soluble, but (73) was insoluble), before resin bound penicillin G acylase (kindly supplied by SmithKline Beecham Pharmaceuticals Ltd.) was added and the reactions were then incubated at 37°C with gentle agitation. Hplc analysis of each reaction (Figure 21) showed that after 24Hrs both (**73**) and (**74**) had been consumed to give deprotected product, whilst (**74**) required a further 192Hrs incubation for the enzyme mediated cleavage reaction to go to completion.

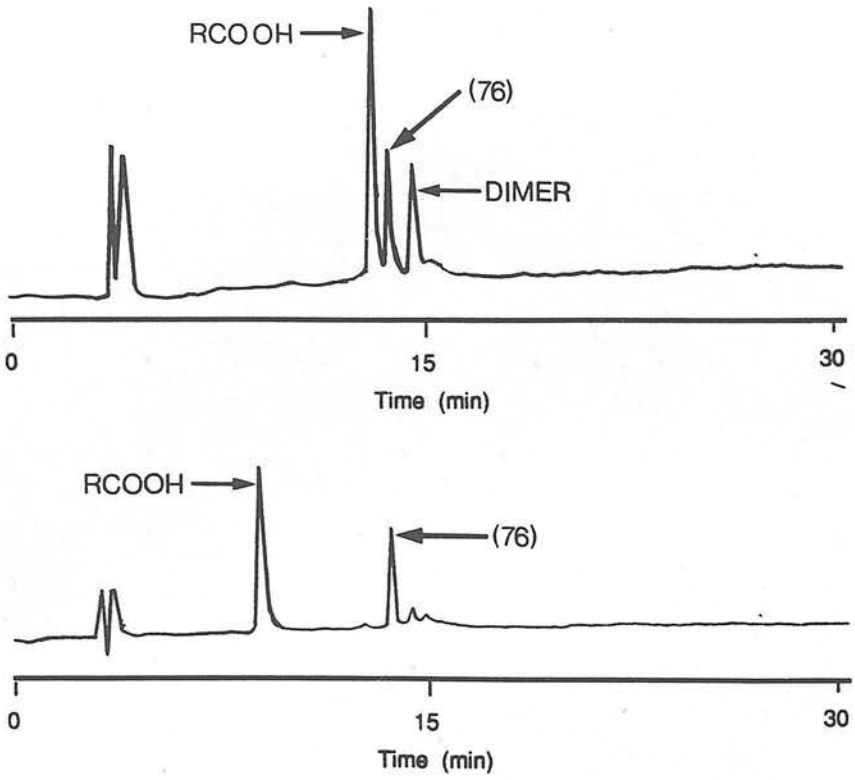


Figure 21. Hplc analysis of the reaction of (**73**) (top) and (**75**) (bottom) with resin bound penicillin G acylase to give the products indicated.

H.Cys.Ser.Asn.Leu.Ser.Thr.Cys.Val.Leu.Gly.OH

(**76**)

The main product of the reaction of (**73**) with the enzyme was identified as the expected cyclic disulphide bridged peptide (**76**), although a small amount of intermolecularly disulphide bridged dimer was also isolated, whilst (**75**) gave only the desired oxidised product (**76**). Subsequent enzyme mediated cleavage reactions carried out under an atmosphere of argon and with the addition of 2-mercaptoethanol resulted in the production of the reduced linear product (**72**).

It was therefore clear that 4-MeOPhacm protected cysteine could be incorporated into a model system which, after treatment with thioanisole/TFMSA, yielded the

corresponding 4-OHPhacm cysteine protected polypeptide. This same model system was then successfully deprotected by the action of penicillin G acylase to give either the fully reduced or oxidised peptide as required.

3.3.3.2 Ral synthesis C - Incorporation of 4-methoxyphenylacetamidomethyl protected cysteine into a synthetic protein.

In order to ascertain the usefulness of this 4-MeOPhacm protected cysteine derivative (**70**) in the synthesis of cysteine containing proteins, restriction alleviation protein (Ral) was resynthesised as described previously, but incorporating the cysteine residues as the Fmoc derivative (**70**). Purification of the 4-MeOPhacm cysteine protected protein was again accomplished with the aid of the Tbfmoc protecting group (**51**) to give essentially pure cysteine protected Ral protein (Figure 22).

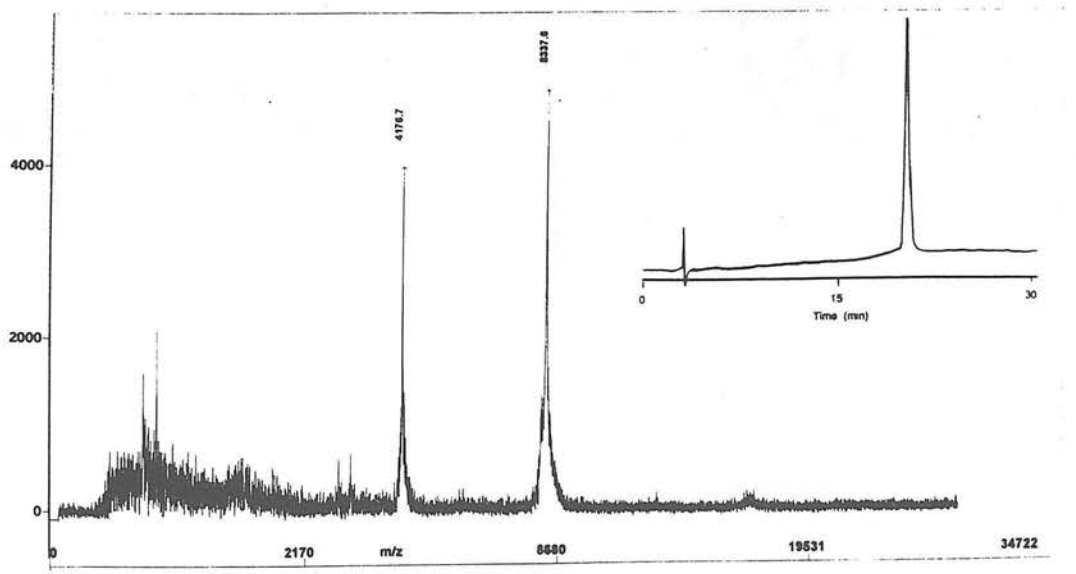


Figure 22. Laser desorption mass spectrum of 4-MeOPhacm protected Ral protein after Tbfmoc purification with Hplc trace inset.

This material was then treated with thioanisole/TFMSA to remove the methyl groups from the phenolic oxygen of the 4-MeOPhacm cysteine protection and desalted by gel filtration to give the 4-OHPhacm protected protein. The 4-hydroxy derivative, which eluted earlier than the corresponding 4-methoxyphenyacetamidomethyl protected protein on Hplc, was then incubated with immobilised penicillin G acylase in an attempt to generate the deprotected Ral protein.

Initial attemps to remove the cysteine protection using the immobilised enzyme supplied by SKB, resulted in complete absorption of the protein onto the solid support used to immobilise the enzyme. This absorption was subsequently shown to be due to the fact that the solid support was based on an ion exchange material which very quickly bound any free basic protein present in solution[49]. Therefore a second immobilised enzyme was obtained, which utilised a polymeric solid support, and the removal of the cysteine protecting groups was attempted once more. Hplc monitoring (Figure 23) indicated that after 24 hours incubation at 37°C three main peaks were produced.

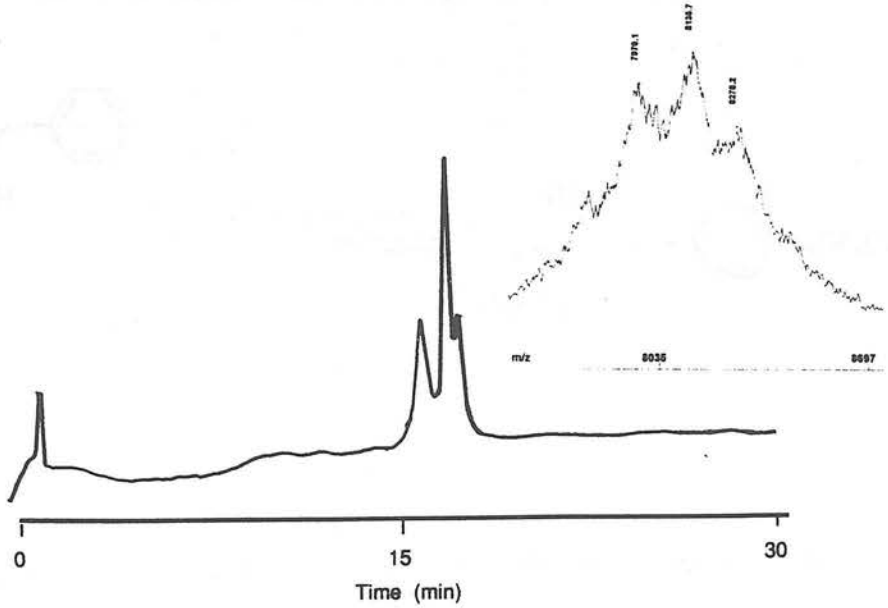
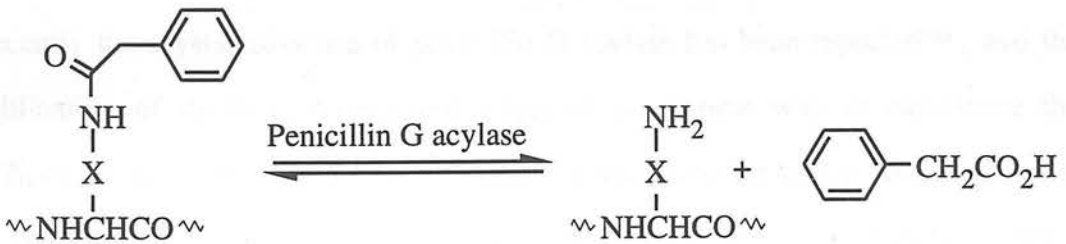


Figure 23. Hplc of 4-OHPhacm protected Ral after treatment with Penicillin G acylase, with inset of MALDITOF mass spectrum obtained from each isolated peak.

Each of these peaks was isolated and analysed by mass spectrometry. This revealed that each peak, although resolvable by Hplc, actually contained a mixture of incompletely deprotected species (Cys protected Ral protein requires MH^+=8258, Ral requires MH^+=7605) which gave essentially identical mass spectra (Figure 23). The presence of free thiol groups in each separate peak was confirmed by Ellman assay[112]. Further attempts to deprotect the cysteine residues completely by adding soluble enzyme gave no improvement in Hplc profile, although analysis of the reaction products by isoelectric focussing gel electrophoresis did indicate that the initially basic protein (pI 8.5) was losing basic residues to give a more acidic protein. This result can only be explained if the rate of the forward reaction (Scheme 19), to produce the free thiol and phenylacetic acid, is sufficiently slowed down due to steric interactions between the synthetic protein and the enzyme, to allow the reverse reaction (Scheme 23), between phenylacetic acid and an amine group to predominate. This reverse reaction, which is normally negligible at pH 8.5 and 37°C, would result in amide bond formation between lysine side chains and the liberated phenylacetic acid, thus resulting in production of a more acidic protein of similar molecular weight to the initial 4-OHPhacm cysteine protected Ral.



Scheme 23.

In an attempt to overcome this shift in the equilibrium, a search of the literature revealed that Wang *et al*[132] had also encountered difficulties in the deprotection of N-phenylacetylated insulin using immobilised penicillin G acylase. They attributed

the incomplete deprotection of the protected insulin derivative to the fact that the size of the substrate would result in a lower collision frequency with the active site of the enzyme compared to that of the liberated phenylacetic acid and hence the equilibrium (Scheme 23) would shift to favour the reverse reaction. However, they were successful in overcoming the reversibility of the enzyme mediated deprotection reaction by adding a column of strong basic ion exchange resin to the experimental apparatus used to perform the enzyme reaction. This successfully sequestered the liberated phenylacetic acid which resulted in isolation of fully deprotected insulin.

Attempts to reproduce the reported reaction conditions[132] with 4-OHPhacm protected Ral protein resulted in almost complete absorption of protein onto the ion exchanger, and no measureable removal of enzyme cleavable cysteine protection. Subsequent attempts at varying the reaction conditions by altering the ionic strength and temperature of the enzyme reaction failed to give any further improvement and at this point the enzyme work was abandoned.

Recently the crystal structure of penicillin G acylase has been reported[133] , and the publication of its three dimensional structure goes some way to explaining the difficulties encountered in its use as a reagent for protecting group removal. Figure 24 shows the three dimentional crystal structure of the enzyme, in which the B chain N-terminal serine, single amino acid catalytic centre, is just discernable at the apex of the β-strand buried deep in the centre of the molecule. Clearly the relative inaccessibility of the active site means that, whilst small unstructured molecules are allowed access to the binding pocket and are rapidly deacetylated, larger protein molecules (which will adopt some degree of structure under the reaction conditions) will have a lower collision frequency with the active site of the enzyme. This could

result in either incomplete deprotection or a shift in the equilibrium (Scheme 23) to favour the reverse reaction, which in turn could lead to enzyme catalysed acylation of free amino groups during deprotection of cysteine protected proteins.

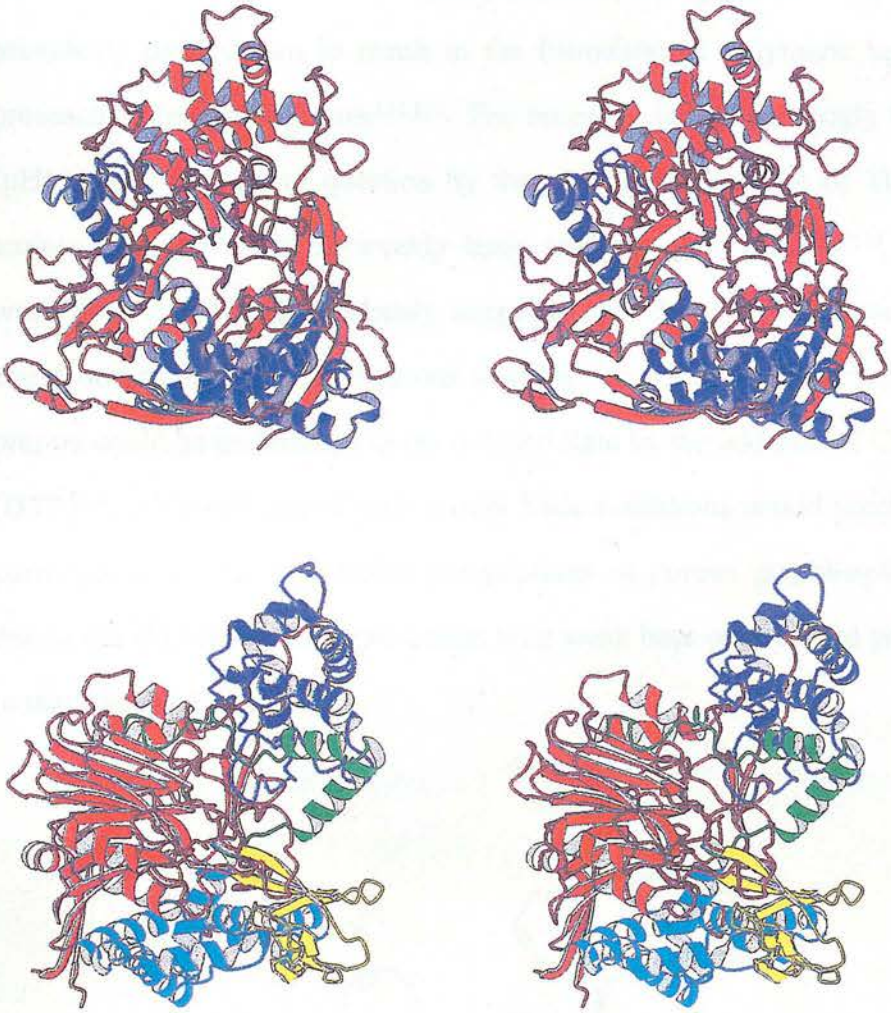
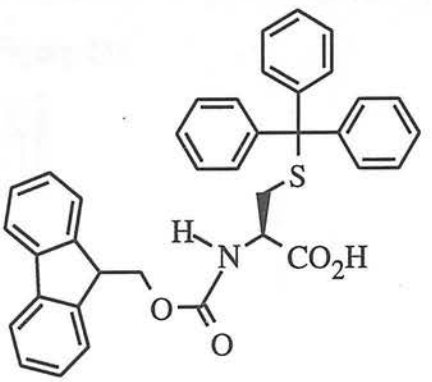


Figure 24. MOLSCRIPT representation of the three-dimensional crystal sructure of penicillin acylase (top) Blue=A chain, Red=B chain showing ball and stick active site serine side chain (B1) and (bottom) an orthogonal view, coloured as described previously[133].

3.3.4 Synthesis D - Trityl protection

Our desired synthetic strategy utilising the affinity based protecting group Tbfmoc[73,74] dictated the necessity for semi-permanent thiol protection, due to the fact that removal of the Tbfmoc group involved strongly basic conditions which had previously been shown to result in the formation of polymeric aggregates in the presence of free thiol groups[134,135]. The necessity for such strongly basic conditions (pH13) was called into question by the observed instability of Tbfmoc protected amino acid derivatives in weakly basic solvents such as DMF[135]. Indeed, earlier work by Irving[135] had previously suggested that the removal of the Tbfmoc group could be accomplished in aqueous solution, at pH values where any thiol groups present could be maintained in the reduced state by the addition of Clelland's reagent (DTT)[136]. Although use of such mildly basic conditions would preclude the affinity purification of Tbfmoc labelled polypeptides on porous graphitised carbon (PGC), due to the slower rate of deprotection with weak base on the solid phase[74] compared to that observed in solution.



(77)

In order to test this proposed methodology for the synthesis of cysteine containing proteins, Ral protein was again synthesised with the cysteine residues being

incorporated as the triphenylmethyl (trityl) derivative (**77**)[137]. The assembly was carried out on a 0.25mM scale using an ABI 430A peptide synthesiser and at the completion of the assembly phase, the Tbfmoc group (**51**) was introduced to the free N-terminus of the resin bound Ral by treatment with chloroformate (**52**) and DIEA with sonication. The resin bound Tbfmoc labelled Ral protein was then thoroughly washed with dichloromethane and dried to give 2.2g of resin bound product. Cleavage of 0.5g of the resin bound product (TFA/EDT/thioanisole/phenol/TIS /water), followed by rapid evaporation of TFA and precipitation with diethyl ether gave the crude Tbfmoc-protein, which was applied to the top of a column of Sephadex G-50 and eluted with aqueous acetic acid.

The Tbfmoc-protein containing samples were then pooled and purified by Hplc, monitoring at 364nm to isolate only the Tbfmoc labelled protein, which was subsequently lyophilised to yield 140mg of solid. Removal of the Tbfmoc group under reducing conditions was accomplished by dissolving the solid material (140mg) in 6M guanidine.HCl (0.1M TRIS, pH 8.5) containing a large excess of dithiothreitol (Clellands reagent, DTT). This solution was then incubated at 37°C and monitored by Hplc (Figure 25).

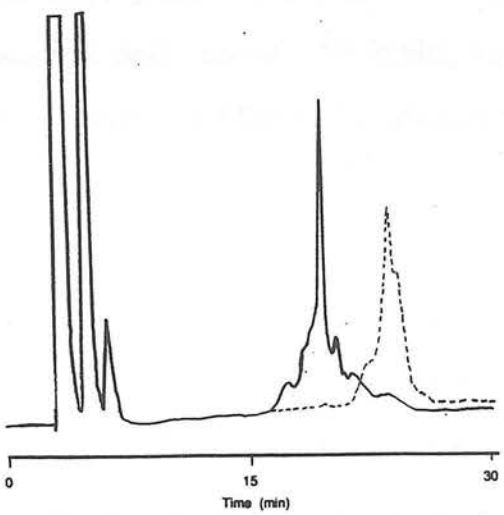


Figure 25. Hplc monitoring of Tbfmoc deprotection in 6M Gdm.HCl at pH 8.5 under reducing conditions to give crude fully reduced Ral protein (solid line).

After 3 hours incubation the deprotection of Tbfmoc was judged to have gone to completion and the fully reduced Ral protein was isolated after desalting (Sephadex G-50, 30% AcOH) and Hplc to give 40mg of pure protein (Figure 26). Ellman assay[112] for free thiols gave a value of 3.7 thiols/molecule which compares well with the expected value of 4 thiols/molecule. This material was then fully characterised by mass spectrometry and amino acid analysis.

Spectroscopic studies on the synthetic fully reduced Ral protein failed to show any binding of cobalt ions, and NMR spectra run in the presence and absence of Zn^{2+} ions indicated that the protein was not undergoing any structural change due to metal ion binding.

3.3.4.1 Folding of synthetic Ral

Since Ral showed no structural features in the reduced state when in the presence of metal ions, it was decided to investigate if oxidative folding to give a disulphide bridged structure would result in any secondary structure formation. The fully reduced protein was therefore folded using the method of Jaenicke and Rudolph[136]. Hplc monitoring of the folding reaction showed no change in elution profile, although isolation of the single peak material after folding for 24 hours and assay for free thiol groups using the method of Ellman[112] confirmed that there were no free thiols present in the protein.

Although the folded Ral protein appeared homogenous by Hplc (Figure 26), the presence of differently disulphide bridged isomers was not precluded. Mapping of the disulphides was therefore carried out by digesting the folded Ral with trypsin and analysing the total digest by mass spectrometry (Figure 27).

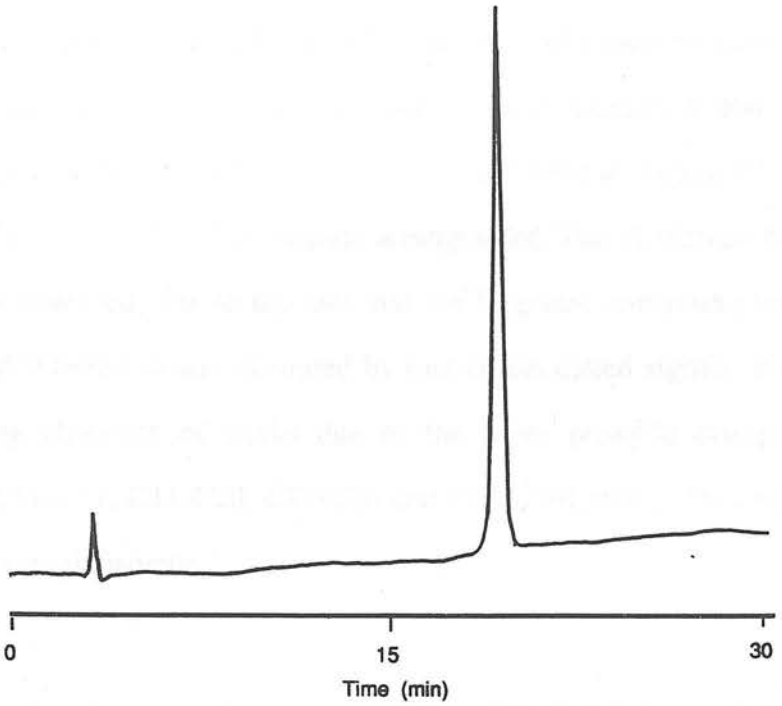


Figure 26. Hplc of pure folded Ral.

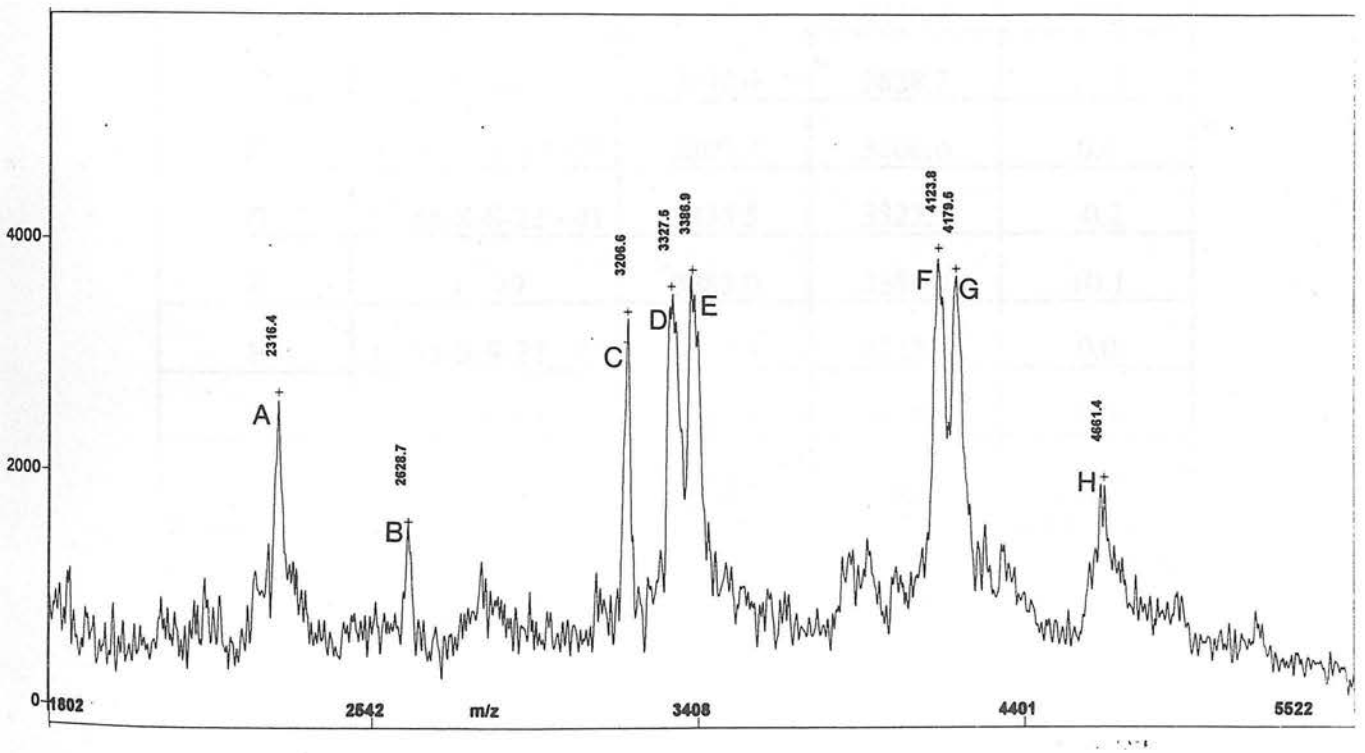


Figure 27. Laser desorption mass spectrum of major fragments produced on tryptic digestion of folded Ral protein.

The results of the disulphide mapping are summarised in Table 3, and indicate the presence of a single disulphide bridged folded form. The identification and location of the disulphide pairings between the four cysteine residues in Ral was deduced from the signals at m/z 3206.6, 3327.5, 4123.8 and 4661.4 (Figure 27 and Table 3) which establish the C11-C26 disulphide arrangement. The disulphide between C17-C20 was not observed, due to the fact that the fragment containing this disulphide (CNGCK M.Wt=524.4) was obscured by matrix associated signals in the spectum. However, the abscence of peaks due to the other possible disulphide bridged fragments (C11-C17, C11-C20, C17-C26 and C20-C26) gives added support for the presence of a single isomer.

Peak label	Predicted Fragment*	Calc. MH^+	Found MH^+	% Error
A	22 - 41	2327.0	2316.4	-0.5
B	8 - 29	2592.0	2628.7	+1.4
C	1 - 15-S-S-17 -29	3207.9	3206.6	0.0
D	8 - 15-S-S-22 - 41	3335.5	3327.5	-0.2
E	1 - 29	3383.0	3386.9	+0.1
F	1 - 15-S-S-22 - 41	4123.9	4123.8	0.0
G	F + Fe**	4179.8	4179.5	0.0
H	1 - 15-S-S-22 - 45	4631.1	4661.4	+0.6

Table 3. *Predicted fragments are numbered according to the sequence given in Figure 10. S-S denotes fragments linked by a disulphide bridge. **Residual iron from the mass spectrometer sample plate.

3.3.4.2 Comparison of synthetic and recombinant Ral.

At this point in time, a new batch of recombinant Ral became available[117] and was compared against the folded synthetic material obtained from synthesis D. Previous attempts to overexpress the protein resulted in production of a truncated sequence (section 3.3.2.1). However, the new batch of recombinant protein co-eluted with the folded synthetic Ral protein when co-injected on Hplc (Figure 28), and when analysed by mass spectrometry gave a molecular ion corresponding to the expected Ral sequence minus the N-terminal methionine. N-terminal sequencing confirmed that there was no N-terminal methionine present in the recombinant protein, although the protein did contain the expected sequence of amino acids[117].

Mapping of the disulphides, by mass spectroscopy of the protein after partial tryptic digest, also confirmed that the same fragments (minus the N-terminal Met) were generated from the recombinant protein, as were found in the folded synthetic Ral (Figure 29). Therefore it was unambiguously established that, with the exception of the post-translational removal of methionine, the recombinant and synthetic proteins were identical in structure and contained the same disulphide pairings. Disappointingly, neither material has exhibited any of the biological activity attributed to the ral gene product[94], when assayed for by monitoring the effect of EcoK mediated DNA modification in the presence of Ral.

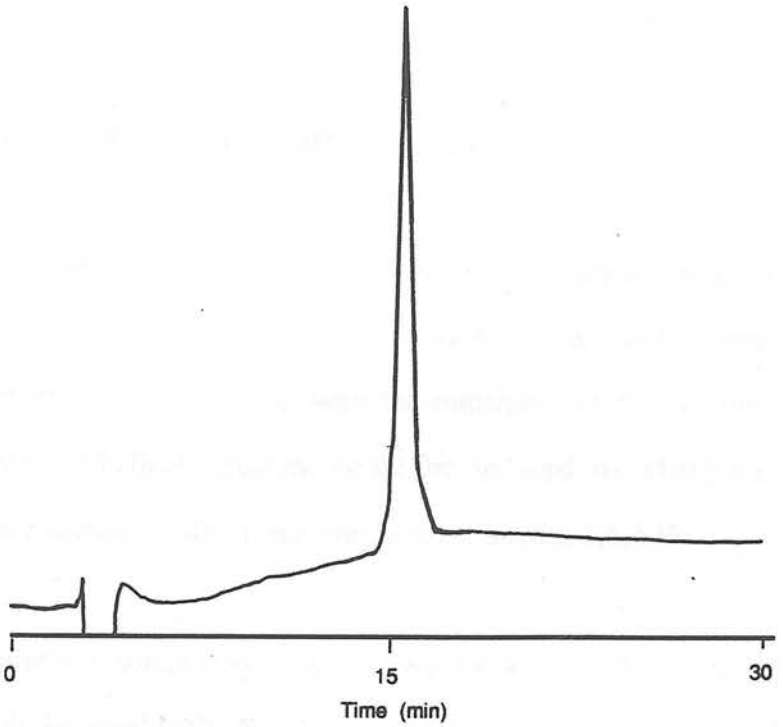


Figure 28. Hplc co-injection of synthetic and recombinant Ral.

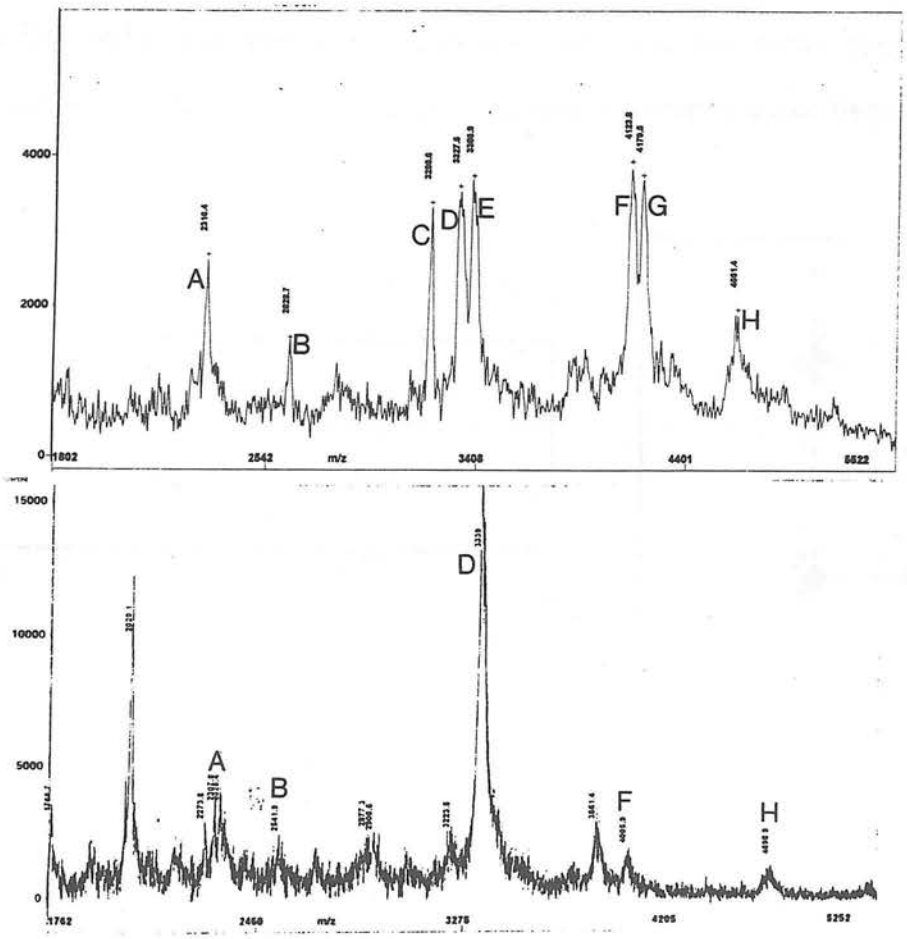


Figure 29. Comparison of the fragments obtained by MALDITOF on partial tryptic digest of synthetic (top) and recombinant (bottom) Ral protein.

3.3.4.3 CD and NMR studies on synthetic Ral

The circular dichroism (CD) spectra obtained from synthetic folded Ral protein are illustrated in Figure 30(a). In aqueous solution the oxidised protein shows no α-helical structure and 46% β-sheet with the remainder of the protein being random coil. However, α-helical structure could be induced by changing the dielectric constant of the solvent as described previously (Section 2.2.4.1).

Nuclear magnetic resonance spectra acquired from a sample of the folded synthetic Ral protein show very little dispersion of backbone amide protons (Figure 30(b)), which is characteristic of an unstructured random coil. Similar spectra were obtained for the fully reduced protein in the presence of zinc ions, thus further supporting the argument that Ral does not contain a motif capable of forming a zinc finger.

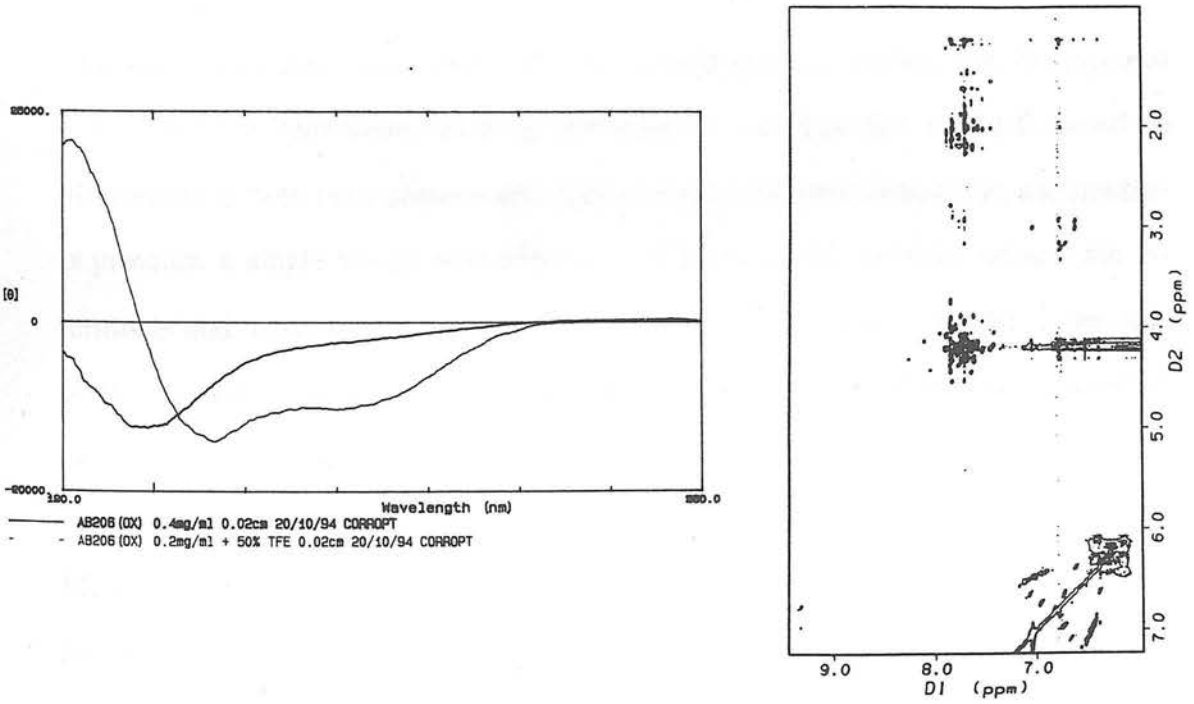


Figure 30. (a) CD spectra obtained from folded synthetic Ral.
(b) Part of the 2D NOESY spectrum obtained from a sample of folded synthetic Ral protein, indicating the lack of dispersion on NH signals consistent with a random coil structure

CHAPTER FOUR : MONOCYTE CHEMOTACTIC PROTEIN (MCP-1)

4.1 General

Our approach towards the synthesis of Ral protein has allowed us to study the most effective route to obtaining a pure, completely deprotected cysteine containing protein, which can then be folded into a single conformation. Our studies suggest that incorporation of cysteine residues as the triphenylmethyl (trityl) protected derivative (**77**) results in unambiguous removal of the thiol protection, which seems to be an important factor during oxidative folding of the reduced protein. As a test of the generality of this approach we then chose to synthesise a cysteine containing protein, that could then be compared against native or recombinant protein of known biological activity, in order to assess the success of our synthetic methodology in producing a fully active folded protein. We therefore chose to attempt the total chemical synthesis of one of the chemokine family of proteins, namely MCP-1.

Chemokines are proinflammatory cytokines that attract and activate specific types of leukocytes[138]. Chemokines can be grouped into two subfamilies, α and β, based on differences in both their primary structure and chromosomal location. In chemokine-α proteins, a single amino acid separates the two cysteine residues nearest the N-terminus and hence they are often referred to as C-X-C proteins. Whilst in the case of the β-chemokines the two cysteines nearest the amino terminus are adjacent to each other and are referred to as C-C proteins.

Monocyte chemoattractant protein-1 (MCP-1) is a member of the β-chemokine family of cytokines and the amino acid sequence of this 76 amino acid protein[139] is given in Figure 31.

EPDAINAPVTCCYNFTNRKI

SVQRLASYRRITSSKCPKEA

VIFKTIVAKEICADPKQKWV

QDSMDHLDKQTQTPKT

Figure 31. Amino acid sequence of MCP-1

In contrast to α-chemokines such as IL-8, which attract and activate neutrophils, β-chemokines tend to attract and activate monocytes. MCP-1 has been shown to be active *in vitro* at sub-nanomolar concentrations[140] and has also been detected in a variety of pathological conditions.

Crystal[141] and solution[142] structures for IL-8 have been solved and indicate that IL-8 is active as a homodimer. Much less is known about the structures of β-chemokines such as MCP-1, although modelling[143] and gel-filtration results[144] have predicted the existence of MCP-1 as a dimer. Further evidence that β-chemokines exist as dimers has recently been supported by the publication of the solution structure of the β-chemokine hMIP-1β (human macrophage inflammatory protein - 1β) by Lodi *et al*[145]. The three dimensional structure of this member of the β-subfamily of chemokines shows that hMIP-1β exists as a symmetrical homodimer under the experimental conditions of the structural investigations. However, whilst the hMIP-1β monomer is similar in structure to that of IL-8, the quarternary structures of the hMIP-1β and IL-8 dimers are entirely different and hence the previously reported modelled structure of MCP-1[143] is incorrect.

4.2 Chemical synthesis of MCP-1

The chemical synthesis of the α-chemokine IL-8 has previously been reported[100], although to our knowledge no β-chemokine has ever been successfully assembled and purified. Thus when we were approached by Dr. Rodney Kelly (MRC Reproductive Biology Unit, Edinburgh) and asked to chemically synthesise MCP-1, it seemed the ideal test of our synthetic methodology for the production of cysteine containing proteins.

Assembly of the MCP-1 sequence (Figure 31) was initially carried out on a 0.086mM scale as described in the experimental section. However, in our initial synthesis, monitoring of coupling efficiency by UV[69] indicated that Asn^{6} had failed to couple to the resin bound MCP-1^{7-76} and hence the amino terminus was subsequently acetylated under automated control of the synthesis. This meant that the protein could not be purified using our Tbfmoc strategy. However, upon cleavage of the protein from the resin and gel filtration of the crude material, Hplc analysis of the resulting fractions (Figure 32) indicated that there was sufficient of the expected Ac-MCP-1^{7-76} present to allow purification by conventional methods in order to study the folding and biological activity of this analogue.

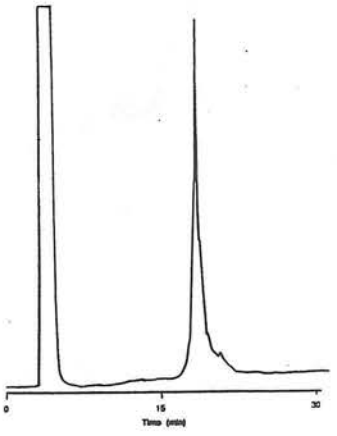


Figure 32. Hplc of crude Ac-MCP-1^{7-76} after gel filtration.

The protein containing fractions were pooled and lyophilised to give 55mg of crude acetylated MCP-1^{7-76}. The protein was then folded as described previously[136], during which Hplc monitoring indicated the earlier elution of the folded, disulphide bridged protein. After 24 hours the folding solution was acidified (AcOH) and loaded onto a semi-preparative Hplc column. Elution with a linear gradient of acetonitrile (0.1% TFA) gave a single main peak which was collected and lyophilised to yield 50mg of pure Ac-MCP-1^{7-76} (Figure 33), which was fully characterised by mass spectometry (Figure 33) and amino acid analysis.

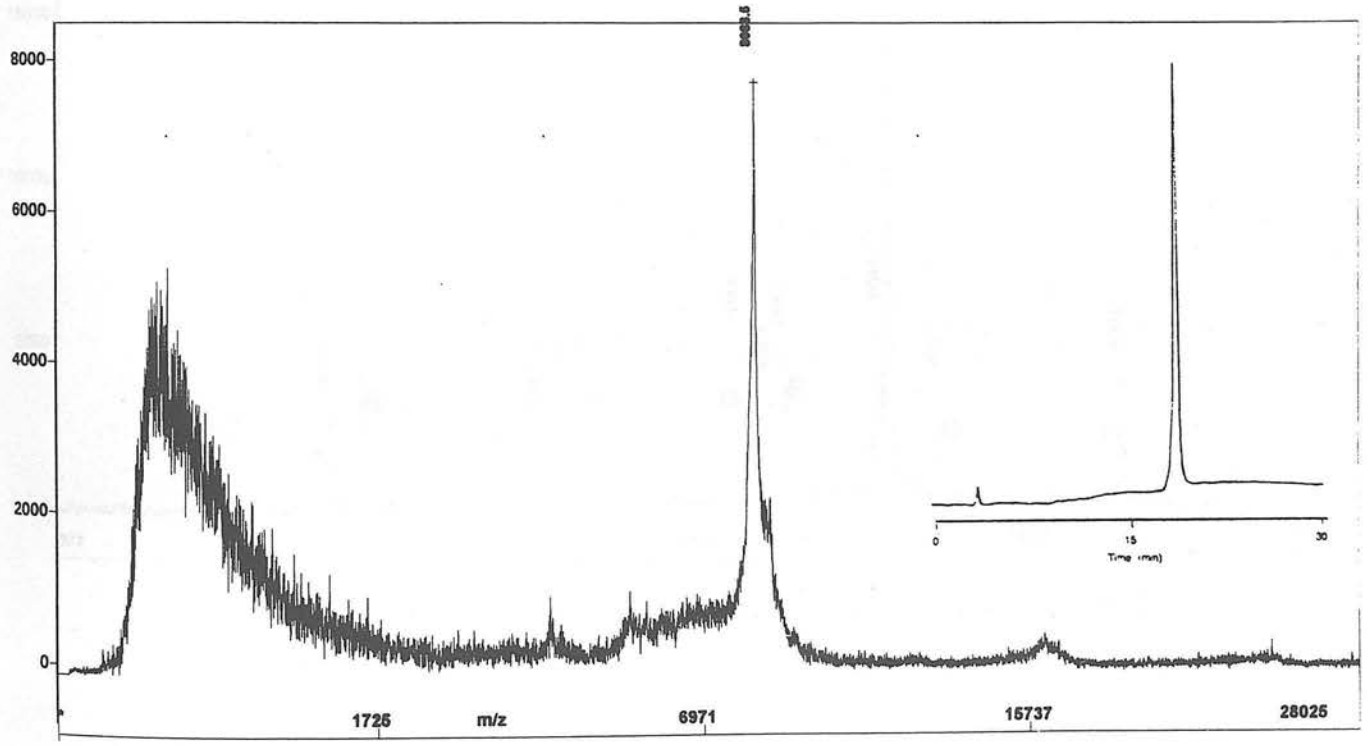


Figure 33. Laser desorption mass spectrum of Ac-MCP-1^{7-76} with Hplc inset.

In addition to this, a sample of the pure folded Ac-MCP-1$^{7-76}$ was subjected to partial digestion by trypsin, and the fragments generated were then analysed by mass spectrometry. The fragments obtained (Figure 34 and Table 4) give support for the presence of only the expected C11-C36 and C12-C52 disulphide bridged species, due to the fact that no fragment from the possible unnatural mixed disulphide pairing of C36-C52 was observed.

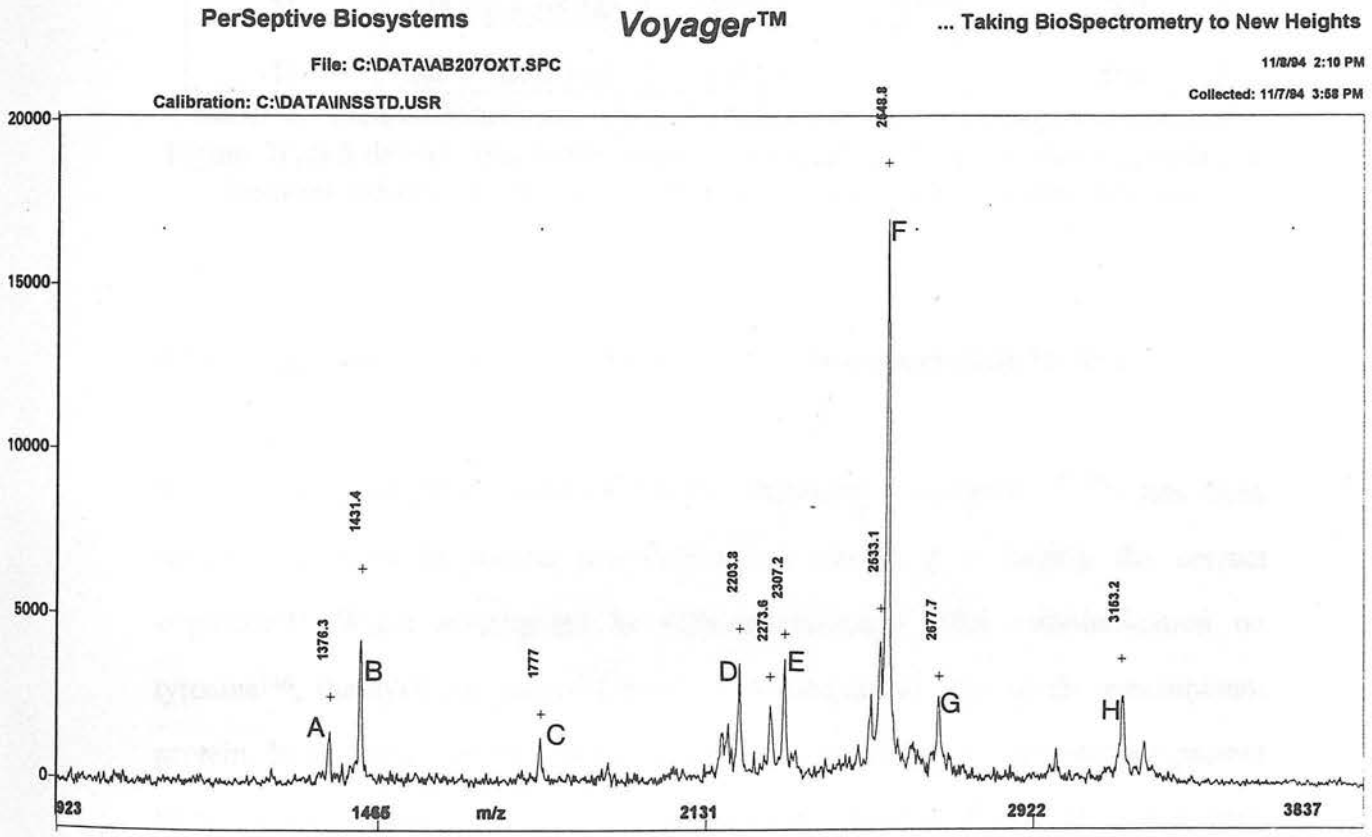


Figure 34. Mass spectrum of the fragments obtained from Ac-MCP-1 on partial tryptic digest

Peak Label	Predicted Fragment*	Calc. MH^+	Found MH^+	% Error
A	59 - 69	1373.6	1376.3	+0.2
B	Ac-7 - 19	1431.6	1431.4	0.0
C	B-S-S-36 - 38	1774.8	1777.0	+0.1
D	B-S-S-50 - 56	2203.0	2203.8	0.0
E	B-S-S-31 - 38	2293.0	2307.2	+0.6
F	Ac-7-18-S-S-(36-38 + 50-56)	2549.2	2548.8	0.0
G	Ac-7-19-S-S-(36-38 + 50-56)	2677.3	2677.7	0.0
H	Ac-7-38Δ20-24	3151.0	3153.2	0.0

Table 4. *Predicted fragments are numbered according to the sequence given in Figure 31. S-S denotes fragments linked by a disulphide bridge, whilst fragments in brackets indicate the presence of more than one disulphide bridge from the preceeding fragment.

4.3 Comparison of synthetic Ac-MCP-1^{7-76} with recombinant MCP-1

The synthetic, slightly shortened MCP-1 derivative (Ac-MCP-1^{7-76}) has been shown to behave in several test systems as though it is largely the correct sequence[146]. When investigated by radioimmunoassay, after radioiodination on tyrosine[146], the synthetic material binds in a comparable way to the recombinant protein. In addition, comparison of the synthetic analogue and recombinant protein by western blot analysis[146] shows that the synthetic material (Figure 35, lane 4) runs as a dimer of ~14KDa. This dimer is also slightly lower in molecular weight when compared to the recombinant protein dimer (Figure 35, lane 5) as would be expected for the truncated sequence.

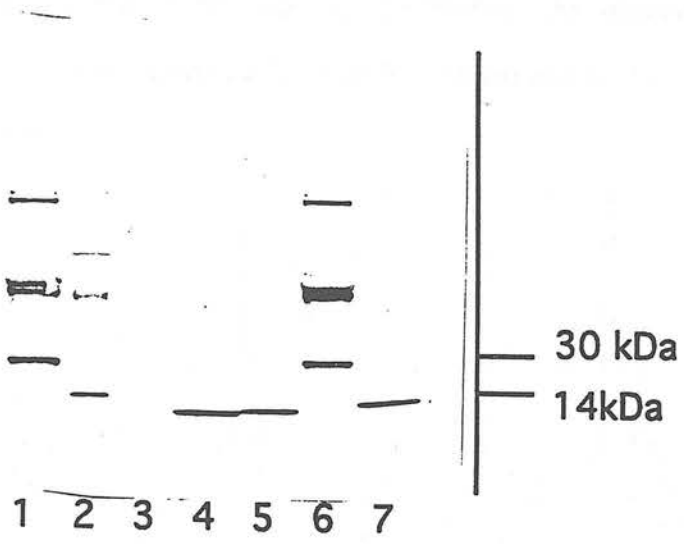


Figure 35. Western blott showing Ac-MCP-1 (lane 4) and recombinant MCP-1 (lane 5).

4.4 Second chemical synthesis of MCP-1

Although the results from the truncated MCP-1 derivative (Ac-MCP-1^{7-76}) were encouraging, previous work by Zhang *et al*[147] had confirmed that MCP-1, like IL-8, requires the N-terminal amino acid sequence for full biological activity. We therefore resynthesised the desired sequence (Figure 31) using optimised coupling cycles over the areas of the assembly where data from the previous synthesis had indicated slow coupling. At the completion of the assembly phase, the N-terminal Fmoc group was replaced by Tbfmoc and the pure Tbfmoc-MCP-1 was subsequently isolated, after cleavage (TFA/EDT/thioanisole/water/TIS/phenol) and size exclusion chromatography (Sephadex G-50, 30%AcOH), by semi-preparative Hplc (monitoring at 364nm). The solid Tbfmoc-MCP-1, obtained after lyophilisation, was then deprotected under reducing conditions (6M Gdm.HCl, 0.1M TRIS pH 8.5, DTT, 37°C) for four hours, before being desalted (Sephadex G-50, 30%AcOH) and lyophilised. The crude reduced protein (Hplc Figure 36) was then folded by stirring for three days in 1M guanidine.HCl (0.1M TRIS, pH 8.5, 250ml) containing 1mM

EDTA, 0.3mM oxidised and 3mM reduced glutathione. As observed for the truncated Ac-MCP-1^{7-76}, the oxidised folded MCP-1 eluted earlier than the reduced protein on Hplc (Figure 36).

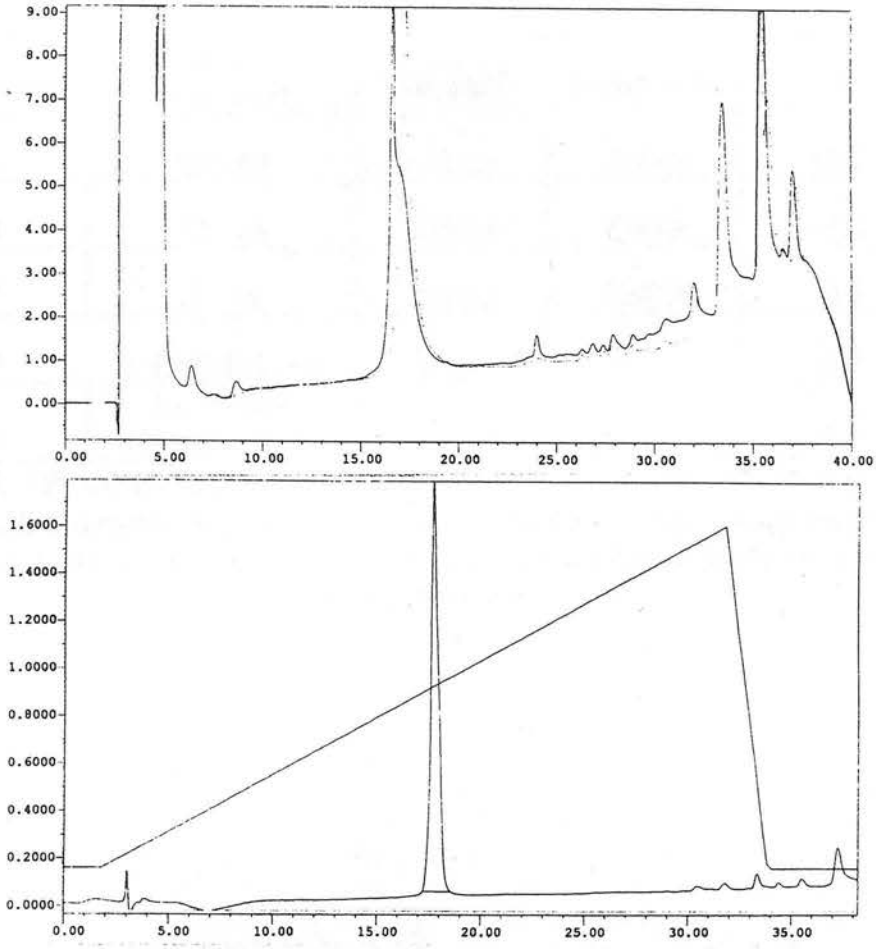


Figure 36. Hplc traces (top) of the crude reduced MCP-1 (dashed line) overlayed with the crude folded protein (solid line) and (bottom) the pure folded MCP-1.

The folding solution was then acidified (AcOH) and loaded onto a semi-preparative Hplc column, before being eluted with a linear gradient of acetonitrile (0.1%TFA). The pure folded MCP-1 (Hplc Figure 36) was collected and lyophilised, before being characterised by mass spectrometry and amino acid analysis. Thus, cleavage of 0.5g Tbfmoc labelled resin bound MCP-1 yielded 4mg of purified folded protein. In addition to this mapping of the disulphides, by mass spectrometry of the fragments obtained on partial tryptic digest (Figure 37 and table 5), again confirmed that only

the expected C11-C36 and C12-C52 fragments were observed. Further support for the presence of only the correctly folded disulphide bridged MCP-1 is given by the lack of any fragments due to the unnatural C36-C52 mixed disulphide.

Peak Label	Predicted Fragment*	Calc.MH+	Found MH+	% Error
A	59 - 69	1373.6	1375.8	+0.2
B	57 - 69	1629.8	1648.6	+1.2
C	1 - 19	2156.5	2142.5	-0.6
D	1-19-S-S-31-38	3020.5	3099.9	+2.6
E	1 - 18-S-S-(36-38 + 5-56)	3151.6	3155.3	+0.1

Table 5. *Predicted fragments are numbered according to the sequence given in Figure 31. S-S denotes fragments linked by a disulphide bridge, whilst fragments in brackets indicate the presence of more than one disulphide bridge from the preceeding fragment.

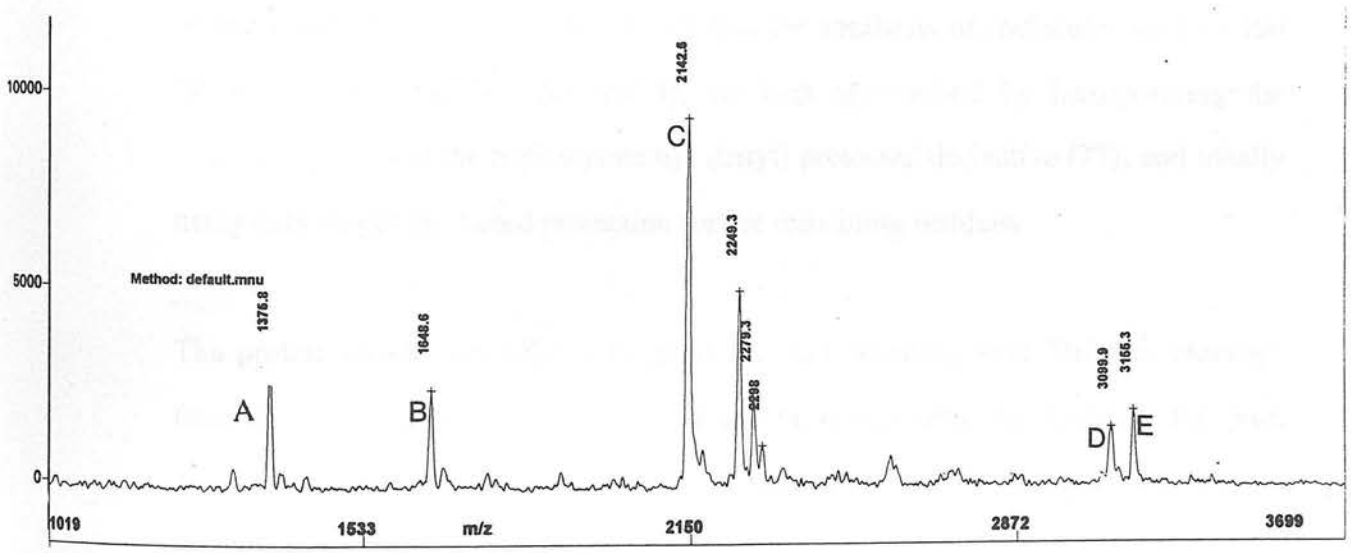


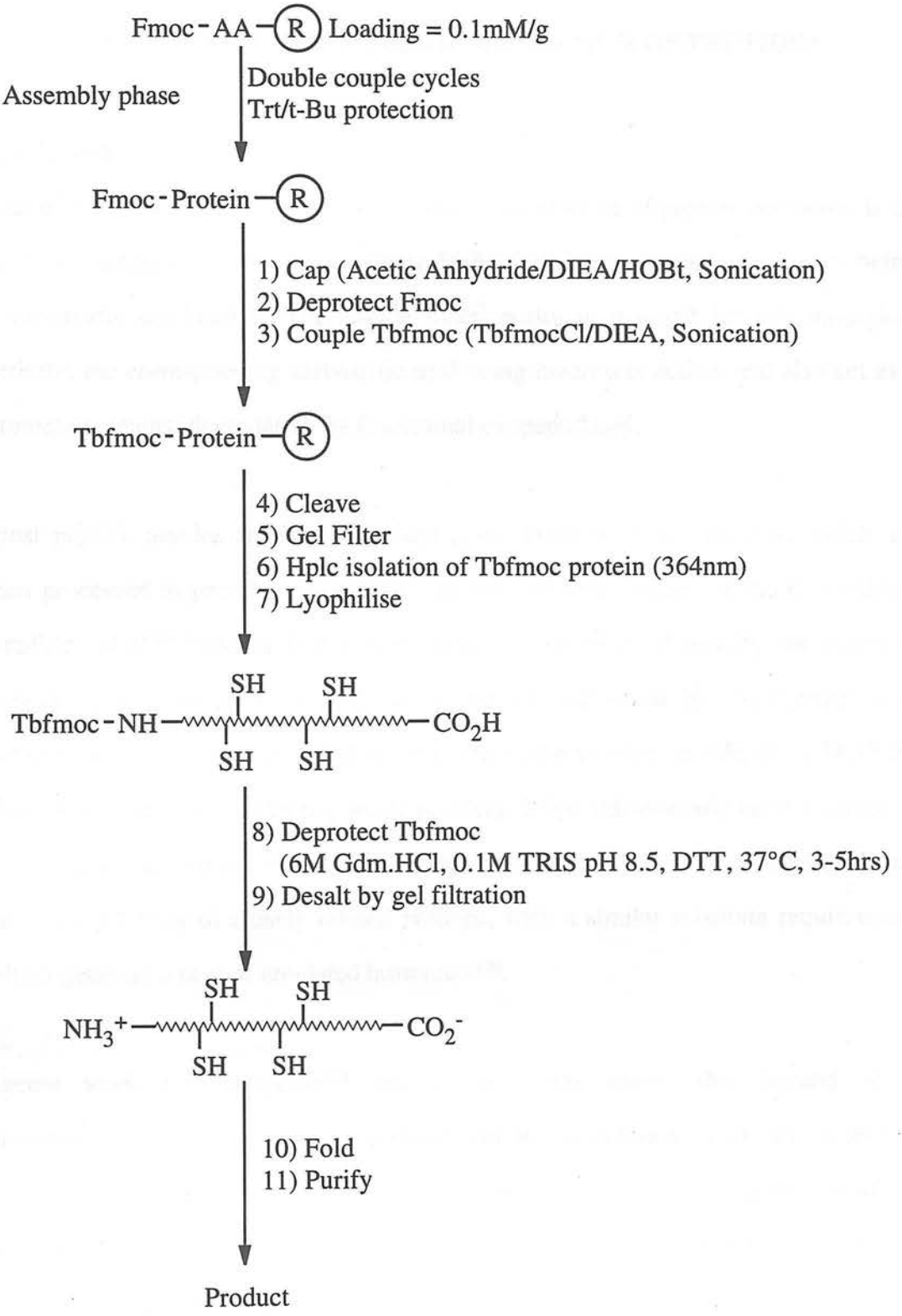
Figure 37. Mass spectrum of the fragments obtained from MCP-1 on partial tryptic digest

This chemically synthesised MCP-1 is currently undergoing testing for biological activity[146] in comparison with the recombinant protein.

CHAPTER FIVE : CONCLUSIONS AND RECOMMENDATIONS

The preceding chapters have covered the chemical synthesis and purification of proteins of biological interest. Use of our strategy for the chemical synthesis and purification of proteins which do not contain cysteine or disulphide bridges, such as MBD (Section 2.2), has been shown to result in the rapid isolation of homogeneous protein in high yield. Tbfmoc[73,74] based purification of this type of molecule can be accomplished either by affinity binding the Tbfmoc labelled protein to PGC[75], or by utilising the hydrophobicity and specific absorbance of the Tbfmoc group (λ_{max} 364nm) to aid Hplc purification. However, the choice of thiol protection for the synthesis of cysteine containing proteins remains an unresolved problem. Although penicillin G acylase cleavable cysteine protection can be successfully applied to short peptide sequences, initial studies indicate that this type of protecting group strategy will not be generally applicable to the production of chemically synthesised proteins. At the present time our studies indicate that the synthesis of molecules such as Ral (Section 3) and MCP-1 (Section 4), are best approached by incorporating the cysteine residues as the triphenylmethyl (trityl) protected derivative (**77**), and ideally using only trityl/t-Bu based protection for the remaining residues.

The protein should subsequently be purified, after labelling with Tbfmoc, cleavage from the solid support and size exclusion chromatography, by isolating the pure Tbfmoc labelled protein by reverse phase Hplc (monitoring at 364nm). The Tbfmoc group should then be removed under mildly basic reducing conditions, which preserve the protein in the reduced state, before allowing the protein to fold and purifying to homogeneity. A generalised scheme for the purification of cysteine containing proteins, based on the work contained in this manuscript, is given in Scheme 24.


Fmoc-AA-R Loading = 0.1mM/g
Assembly phase
Double couple cycles
Trt/t-Bu protection
Fmoc-Protein-R
1) Cap (Acetic Anhydride/DIEA/HOBt, Sonication)
2) Deprotect Fmoc
3) Couple Tbfmoc (TbfmocCl/DIEA, Sonication)
Tbfmoc-Protein-R
4) Cleave
5) Gel Filter
6) Hplc isolation of Tbfmoc protein (364nm)
7) Lyophilise
Tbfmoc-NH
SH SH SH SH
CO2H
8) Deprotect Tbfmoc
(6M Gdm.HCl, 0.1M TRIS pH 8.5, DTT, 37°C, 3-5hrs)
9) Desalt by gel filtration
NH3+
SH SH SH SH
CO2-
10) Fold
11) Purify
Product

Scheme 24.

CHAPTER SIX : BIOLOGICAL AMIDATION OF PEPTIDES

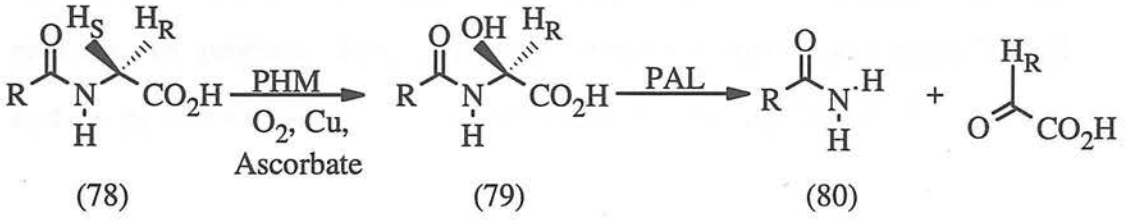
6.1 General

One of the most prevalent post-translational modifications of peptide hormones is C-terminal amidation, with approximately 50% of all known peptide hormones being C-terminally amidated[148]. This amide functionality is required for full biological activity, the corresponding carboxylic acid being much less active, and also act as a protection against degradation by C-terminal exopeptidases.

Most peptide amides are biosynthesised from larger precursor proteins which are then processed to peptides extended by an extra glycine residue at the C-terminus. Bradbury *et al*[149] were the first to show that the conversion of this glycine extended peptide to the corresponding peptide amide was catalysed by an enzyme, now commonly referred to as peptidylglycine α-amidating enzyme (α-AE, EC 1.14.17.3), which they had extracted from porcine pituitary. It has subsequently been shown that similar enzymes are present in many tissue sources and it has been suggested that these are a family of closely related proteins, with a similar substrate requirement, which generate a host of amidated hormones[150].

Recent work by Katopodis[151] and others[152] has shown that instead of a monofunctional enzyme being responsible for the conversion of glycine extended peptides to peptide amides, there is in fact a two step mechanism in operation which utilises a bifunctional enzyme (derived from a single structural gene), which can either catalyse the full amidation reaction or may be cleaved into two monofunctional protein domains which can then act separately[153]. These two monofunctional enzymes are referred to as peptidyl-α-hydroxyglycine monooxygenase (PHM, EC 1.4.17.3) and peptidylamidoglycolate lyase (PAL, EC 4.3.2.5).

The overall mechanism of amidation[153] (Scheme 25) has been shown to involve PHM catalysed removal of the pro-S hydrogen from the C-terminal glycine residue (**78**) and simultaneous addition of a hydroxyl group, with retention of configuration, to give an S-α-hydroxyglycine residue (**79**) in a process dependant on oxygen, copper and ascorbate. The second enzyme, PAL, then catalyses the decomposition of the α-hydroxyglycine extended intermediate (**79**) to the amide (**80**) and glyoxylic acid. This second step occurs spontaneously at alkaline pH[154], but under physiological conditions requires the action of the PAL enzyme and has been shown to be stereospecific[153,155].



(78) (79) (80)

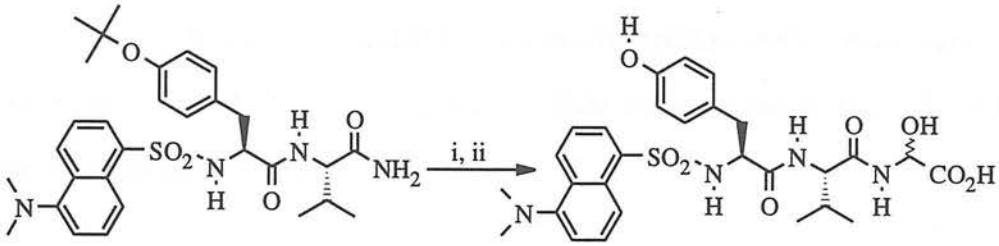
Scheme 25.

Because many peptide amides play an important role in the regulation and activation of biochemical processes, the discoverey that the amidation of these molecules proceeds via an α-hydroxyglycine extended intermediate raises the question as to whether these intermediates have a biological function or, indeed, could be used as a basis for control of *in vivo* hormone production.

6.2 Synthesis of α-hydroxyglycine extended peptides.

Previous reports of synthetic α-hydroxyglycine extended peptides have relied on the condensation of short peptide amides with glyoxylic acid at elevated temperatures, as described by Young and Tamburini[155] (Scheme 26) for the synthesis of N-dansyl-

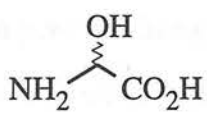
Tyr-Val-α(OH)Gly.



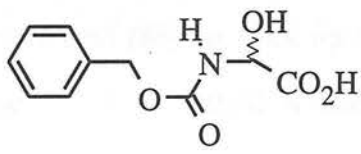
Scheme 26. i) Glyoxylic acid monohydrate ii) TFA

The synthesis of α-hydroxyglycine extended peptides has also been described more recently by the reaction of recombinant PHM enzyme on synthetic glycine extended peptides, to produce α-hydroxyglycine extended human calcitonin[156] and α-hydroxyglycine extended human growth hormone releasing factor[157].

We sought to find a general route to α-hydroxyglycine extended peptides by introducing the α-hydroxyglycine residue (**81**), or some protected derivative of (**81**) as the C-terminal residue in the assembly phase of Fmoc based SPPS. It was hoped that this would then allow the synthesis of peptide sequences of varying length and complexity.



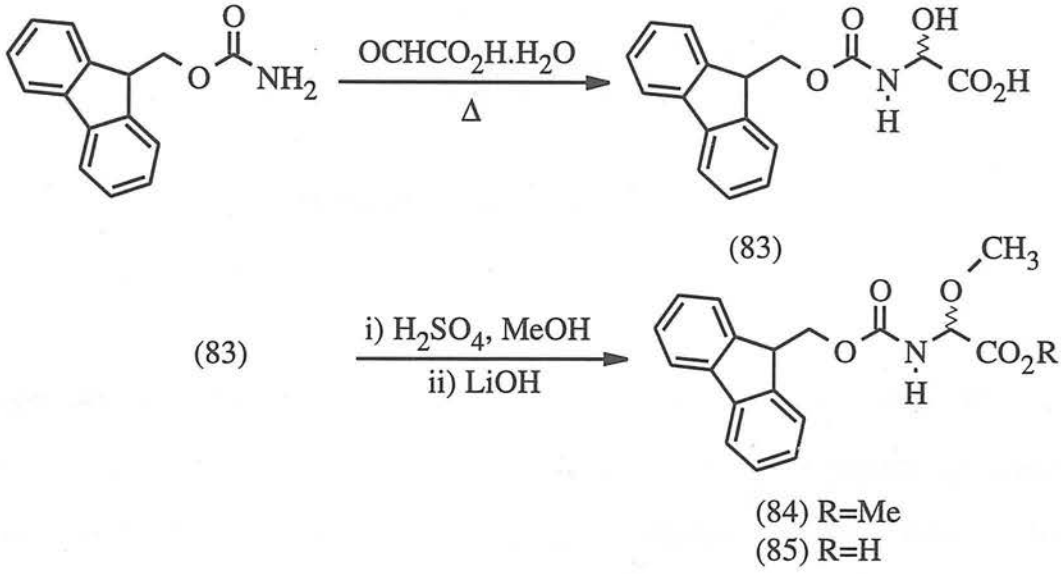
(81)



(82)

The synthesis of N^{α}-9-fluorenylmethoxycarbonyl-α-hydroxyglycine (**83**) was achieved by a variation of the method reported by Ben-Ishai and Zoller[158] for the synthesis of the corresponding Z protected derivative (**82**). Fmoc carbamate was treated with glyoxylic acid (5 eq.) in gently refluxing ethyl acetate for 6 hours to give racemic (**83**) as a white powder in 73% yield (Scheme 27). Attempts to protect

the secondary alcohol in (**83**) as the corresponding t-Bu, MEM or THP ethers failed to yield any of the desired products, although dehydration of (**83**) with concentrated sulphuric acid in methanol[159] yielded the α-methoxyglycine methyl ester derivative (**84**) by addition of methanol to the intermediate iminium anion formed on acid catalysed dehydration of (**83**).

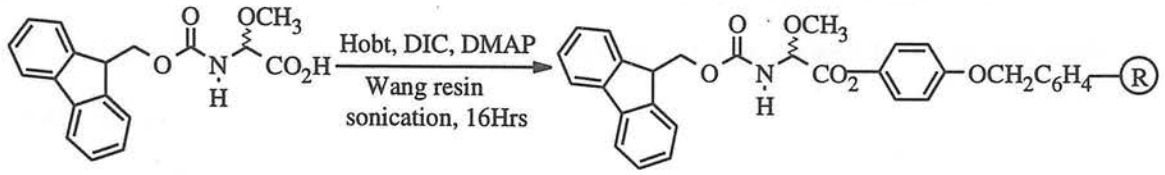


Scheme 27.

Treatment of (**84**) with lithium hydroxide in aqueous acetone allowed hydrolysis of the methyl ester to give the free carboxylic acid (**85**) on work up as a white powder in 68% overall yield from (**83**) (Scheme 27). This method of ester hydrolysis gave only low levels of Fmoc deprotection as monitored by t.l.c. and is an attractive alternative to the use of orthogonal carboxylic acid protection for the synthesis of Fmoc derivatives.

The α-methoxyglycine derivative (**85**) was then immobilised for use in SPPS by esterification of the free carboxylic acid with resin bound 4-benzyloxybenzyl alcohol (Wang resin[12]). Optimum conditions required overnight sonication of an excess of (**85**), DMAP, HOBt and DIC with the Wang resin in DMF. This gave a 61% yield of

functionalised resin (**86**) (Scheme 28), as determined by UV spectrometry.



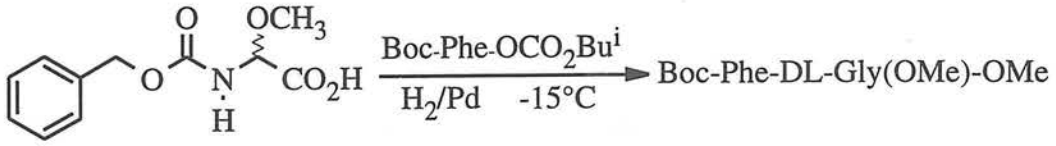
(85) (86)

Scheme 28

H.Trp.Met.Asp.Phe.α-(OH)Gly.OH

(**87**)

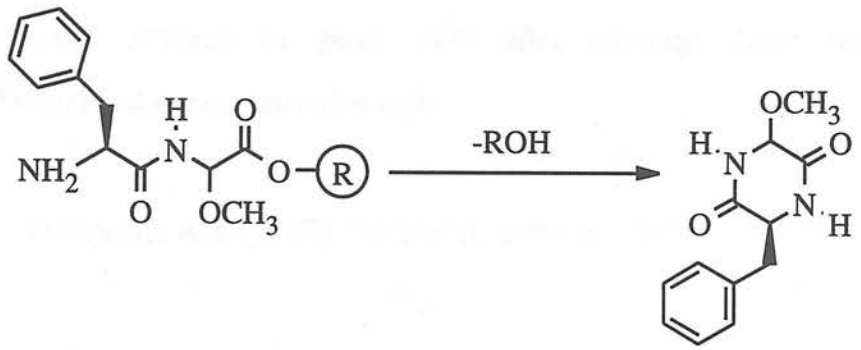
Resin (**86**) was then employed for the synthesis of gastrin derivative (**87**). The synthesis was carried out on a 0.486mM scale on an ABI 430A peptide synthesiser. Fmoc deprotection monitoring during the synthesis indicated only a 60% incorporation of phenylalanine. This was expected as Kawai *et al*[159] had previously reported that the methyl ester of α-methoxyglycine was unstable, could not be isolated, and could only be incorporated into a dipeptide by generation *in situ* from the corresponding Z derivative in the presence of a mixed anhydride of Boc-phenylalanine (Scheme 29).



Scheme 29

A more worrying aspect was that the incorporation of aspartic acid was only 34.6% of the initial substitution. This drop in coupling at the dipeptide stage is

characteristic of diketopiperazine (DKP) formation. This well known side reaction in peptide synthesis[160] involves a base catalysed cyclisation of the dipeptide to form a six membered ring and results in self cleavage of the dipeptide from the resin (Scheme 30).



Scheme 30

The remaining residues were successfully incorporated without further decrease in coupling efficiency, and the desired diastereomeric pentapeptides were isolated in low yield after cleavage (TFA/EDT/water) and semi-preparative Hplc. The diastereomers, due to the R and S configuration of the racemic α-hydroxyglycine residue, were separable by Hplc and gave identical mass spectra and amino acid analysis. Transformation of the C-terminal α-methoxyglycine to α-hydroxyglycine is thought to proceed via the corresponding imine intermediate, which is generated on acidolysis. This imine is then thought to be hydrolysed to the α-hydroxyglycine extended peptide on aqueous work up.

Whilst the initial synthesis confirmed that this was a viable route to the desired α-hydroxyglycine extended peptides, the problem of DKP formation resulted in a low yield of the desired compound. Several solutions to the problem of DKP formation during solid phase synthesis have been reported, the most elegant and generally applicable of which involves the use of 2-chlorotrityl chloride functionalised resin

(**10**)[14,161]. This resin is reported to completely suppress DKP formation[162] by exploiting the relatively large volume of the trityl group to sterically block any cyclisation. We therefore employed this resin for the synthesis of gastrin derivative (**88**) and cholecystokinin (CCK) derivative (**89**). In both cases deprotection monitoring confirmed that no DKP formation had occured during synthesis, and both compounds were obtained in good yield after cleavage from the resin (TFA/EDT/water) and semi-preparative Hplc.

H.Glu.Glu.Ala.Tyr.Gly.Trp.Met.Asp.Phe.α-(OH)Gly.OH

(**88**)

H.Asp.Tyr.Met.Gly.Trp.Met.Asp.Phe.α-(OH)Gly.OH

(**89**)

The diastereomeric peptides (**88**) and (**89**) were inseparable by Hplc (Figure 38), although the presence of two diastereomers was confirmed by NMR. The ratio of the diasteromers was found to be 1:1 in (**88**) and 0.8:1 in (**89**), as determined by integration of the signal due to the αH of the α-hydroxyglycine residue in each of the above sequences (Figure 39).

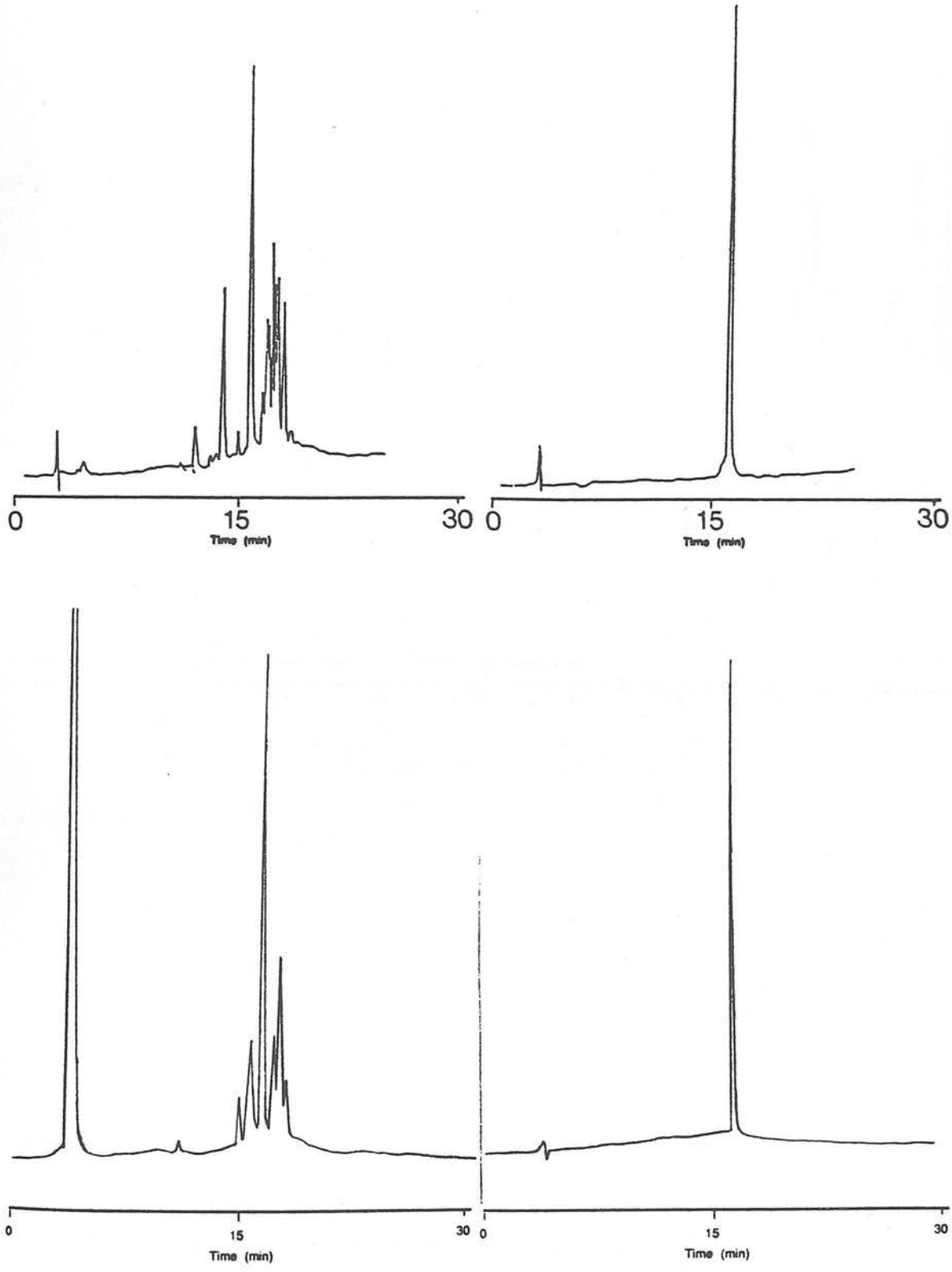


Figure 38. Hplc traces of crude and purified (**88**) (top) and (**89**) (bottom).

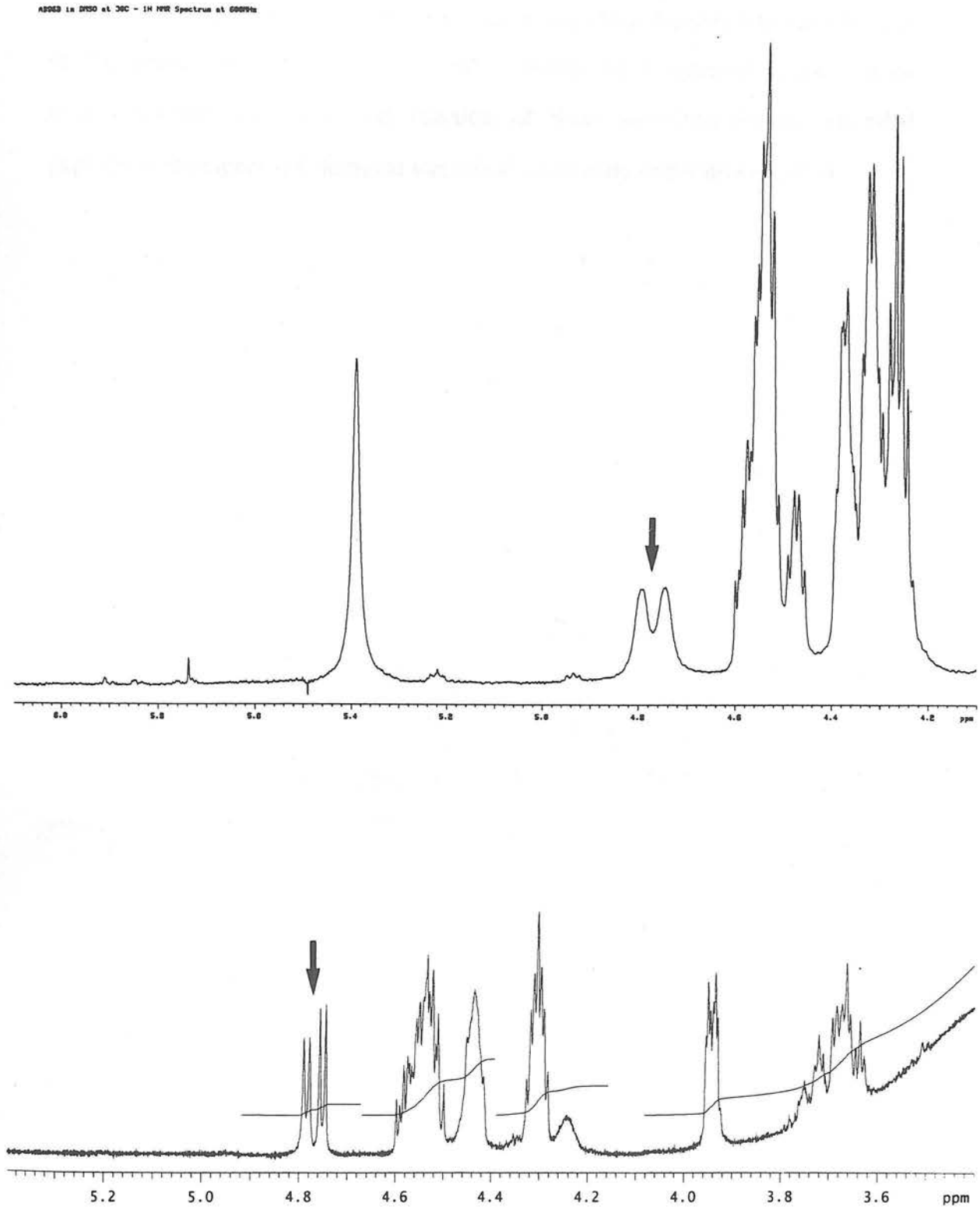


Figure 39. NMR spectra showing the αH signal (arrowed) due to the racemic α-hydroxyglycine residue in (**88**) (top) and (**89**) (bottom).

In conclusion, we have shown that it is possible to introduce α-methoxyglycine as the C-terminal residue in the assembly phase of Fmoc SPPS. Subsequent acid treatment and aqueous work up not only causes side chain deprotection and cleavage of the peptide from the resin, but also converts the C-terminal residue to α-hydroxyglycine. The biological function of these α-hydroxyglycine extended peptides with respect to C-terminal amidation is currently under investigation.

CHAPTER SEVEN : EXPERIMENTAL

7.1 Notes

All amino acids were purchased from either Bachem, Novabiochem, Raylo or were synthesised as described in the text, and are of the L-configuration unless otherwise stated. Melting points were determined using a Koffler hot stage melting point apparatus and are uncorrected. Optical rotations were measured on an AA1000 polarimeter (Optical Activity Ltd.) using a 10.0cm cell in the solvent indicated in the text. Analytical thin layer chromatography (t.l.c.) was carried out using a 0.3mm layer of silica (Merck, kieselgel 609) containing 0.5% woelm fluorescent indicator, on foil plates in the solvents indicated in the text. The components were observed under ultra-violet light, by reaction with iodine vapour and by charing of the plate after spraying with 20% sulphuric acid in methanol. Infrared specta were recorded on either a Perkin-Elmer 781 or Bio-RAD SPC 3200 instrument. Ultra-violet spectra were recorded on a Varian Cary 210 spectrophotometer in the solvents indicated in the text. High and low resolution fast atom bombardment (FAB) spectra were measured on a Kratos MS50TC instrument, using either thioglycerol, 3-nitrobenzyl alcohol or glycerol as matrix. Routine laser desorption time of flight mass spectra were measured on a PerSeptive Biosystems LaserTec Benchtop II system. Circular dichroism (CD) spectra were recorded on a JASCO J600 spectropolarimeter (University of Stirling). Nuclear magnetic resonance (NMR) spectra were recorded on either a Jeol FX-60 (60MHz), a Brucker WP-80 (80MHz), a Brucker WP-200 (200MHz), a Brucker AC-250 (250MHz), a Brucker WH-360 (360MHz) or a Varian VXR5000 (600MHz) instrument in the solvents stated. Elemental analyses were performed on a Perkin-Elmer 2400CHN elemental analyzer. Amino acid analysis was performed on an LKB 4150 alpha amino acid analyser on the hydrolysate obtained after heating samples at 110°C for 24 hours in a sealed Carius tube, followed by evaporation to dryness. Peptide sequencing was carried out on an

Applied Biosystems 477A sequencer at the Welmet sequencing facility, University of Edinburgh. Polyacrylamide molecular weight gels were run using 20% homogenious SDS gels (Pharmacia) on an LKB Pharmacia PhastGel electrophoresis system, and developed using coomassie blue staining.

All solvents were distilled before use and the following were dried using the reagents given in parenthesis when required : dichloromethane (calcium hydride), diethyl ether (sodium wire), tetrahydrofuran (sodium,benzophenone). Peptide synthesis grade dimethylformamide (DMF), 1,4-dioxan and piperidine were obtained from Rathburn Chemicals, Walkerburn Scotland. Peptide synthesis grade trifluoroacetic acid was obtained from Applied Biosystems (ABI). High performance liquid chromatography (Hplc) was carried out using either an ABI system, comprising 2X 1406A solvent delivery systems, an 1480A injector/mixer and an 1783A detector/controller, or a Gilson system, comprising 2X 306 solvent delivery systems, an 811C dynamic mixer, an 805 manometric module, a 119 UV/VIS detector and Gilson 715 software-driven gradient controller. Components were eluted from various columns, as described in the text, by a linear gradient of acetonitrile (far UV grade, Rathburn Chemicals) in Milli-Q grade water, where both solvents contained 0.1% v/v of Hplc grade trifluoroacetic acid (Fisons).

7.2 Solid phase peptide synthesis

The polypeptides described were synthesised on an Applied biosystems 430A automated peptide synthesiser fitted with a UV monitoring system as described previously[69]. All peptides were synthesised using the 9-fluorenylmethoxycarbonyl (Fmoc) strategy of N^{α} protection. This involves the complementary use of orthogonal acid labile side chain protection and an acid labile peptide-resin linker. The side chain protecting groups used were as follows : t-butyl (Bu^{t}) ethers for

serine, threonine and tyrosine; t-Bu esters for aspartic and glutamic acid; t-butoxycarbonyl (Boc) for lysine; τ-triphenylmethyl (trityl) for histidine; 2,2,5,7,8-pentamethylchroman-6-sulphonyl (Pmc) for arginine. The carboxamide side chains of asparagine and glutamine were incorporated as either the 4,4'-dimethoxybenzhydryl (MBH) or trityl derivatives, or without side chain protection as indicated in the text. Cysteine was incorporated as either the t-butylsulphenyl (S^tBu), acetamidomethyl (Acm), phenylacetamidomethyl (Phacm), 4-methoxyphenylacetamidomethyl (4-MeOPhacm) or trityl derivatives as indicated in the text. TbfmocCl[74] was prepared using the method of Irving[75].

7.2.1 Coupling of the C-terminal amino acid onto 4-alkoxybenzylalcohol (Wang) resin.

A solution of Fmoc amino acid (1mM) and N,N'-diisopropylcarbodiimide (DIC) (0.5mM) in DMF (20ml) was stirred for 15 minutes at room temperature, then 4-benzyloxybenzylalcohol functionalised polystyrene resin (Wang) (1.0g, 0.8mM) was added, together with a catalytic amount of 4-(N,N'-dimethylamino)-pyridine and the mixture was then sonicated for 1-2 hours. The functionalised resin was then removed by filtration and sequentially washed with DMF, 1,4-dioxan and dichloromethane, before being dried under vacuum. The loading of the functionalised resin was then determined by treating a known weight of resin with 20% piperidine/DMF in a 10ml volumetric flask, for 20 minutes in a sonic bath. The UV absorbance of the supernatant was then measured at 302nm and the loading calculated using the Beer-Lambert law (ε_{302} = 15400 for fulvene-piperidine adduct). This procedure generally gives a resin with a functionality in the region of 0.1mM/g. To obtain higher resin loadings (0.5-0.6mM/g), the amounts of Fmoc amino acid and DIC were increased to 5.0mM and 2.5mM respectively.

7.2.2 Automated SPPS

Synthetic proceedures were pre-programmed into the ABI 430A synthesiser prior to the comencement of synthesis. Routinely most residues were incorporated using double couple cycles, in which the first coupling cycle utilised a preformed symmetrical anhydride, followed by the second coupling using a preformed 1-hydroxybenzotriazole (HOBt) ester. The exceptions to this were the amino acids asparagine, glutamine and histidine, which were coupled twice via their HOBt esters and glycine, which was coupled singly as a symmetrical anhydride.

Each synthetic cycle involved (1) a capping step to block any unreacted amino groups. (2) Deprotection of the base labile Fmoc group. (3) Coupling of the next N^{α} protected amino acid. Each of these steps was followed by thorough washing of the resin and each cycle was then repeated with the chosen amino acid in order to build up the desired sequence of amino acids in a stepwise manner from the C-terminus to the N-terminus. The preprogrammed synthetic cycles are summarised below.

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 (10ml) for 10 minutes, before the capping solution was drained from the vessel and the resin washed by six portions of DMF.

2. Deprotection - The resin was vortexed with a solution of 20% piperidine/DMF for 3 minutes before being drained. An aliquot of the deprotection solution was then sent to a UV detector in order to quantify the amount of fulvene-piperidine adduct present and hence subsequently gave an indication of the percentage incorporation of each residue. Deprotection was then repeated for a second time for 1 minute in order

to establish if the N^{α} protecting group had been completely removed. Finally the resin was washed with six portions of DMF/1,4-dioxan (1;1).

3. Coupling - The resin was vortexed with a solution of 0.5mM Fmoc amino acid preformed symmetrical anhydride (formed from 1mM Fmoc AA and 0.5mM DIC in the activator vessel). The first coupling cycle was allowed to continue for 30 minutes before the solution was drained from the reaction vessel and the resin washed with two portions of DMF. The resin was then vortexed for a second 30 minute period with 0.5mM of Fmoc amino acid HOBt active ester (preformed from 0.5mM Fmoc AA, 0.5mM HOBt and 0.5mM DIC) before being drained and washed with four portions of DMF. For certain difficult couplings the coupling cycle time was extended or alternatively a third coupling of HOBt ester was incuded as required.

7.3 Experimental

Methylated DNA Binding Domain of MeCP2 (MBD) (MeCP2 78-162)

The synthesis was carried out on an initial scale of 0.144mM using the functionalised resin Fmoc Arg(PMC)-($OCH_2C_6H_4OR$) (1.0192g, 0.142mM/g). All amino acid side chains were protected as described previously with the exception of asparagine and glutamine, which were incorporated without side chain protection. All amino acids (with the exception of glycine) were double coupled, apart from residues 69 to 56 which were treble coupled with the third coupling cycle time extended, and residues 41 to 25 which were allowed to couple for an extended period during the second coupling cycle. The N-terminal Fmoc group was left on at the completion of the assembly phase and the resin bound product was capped for 30 minutes in a sonic bath before being washed sequentially with DMF, 1,4-dioxan, dichloromethane and dried to give 2.8178g of resin bound product.

500mg of the Fmoc protected, resin bound product was treated with 20%

piperidine/DMF (10mls) for 10 minutes in a sonic bath, before the resin bound product was filtered off, washed with DMF, 1,4-dioxan, dichloromethane and dried. This dry resin was then added to a suspension of TbfmocCl (**52**) (230mg, 0.5mM) in dichloromethane (10ml) and DIEA (64μL, 0.37mM) was added before the reaction flask was sealed, covered in aluminium foil and sonicated for three hours with occasional mixing. The resin bound Tbfmoc-protein was then washed thoroughly with dichloromethane, dried (functionality by UV[75] =0.04mM/g) and cleaved from the resin by stirring for 4 hours in the dark, with a mixture of TFA (10ml), Phenol (0.75g), EDT (1.5ml), thioanisole (0.5ml) and water (0.5ml). The resin was then removed by filtration and the filtrate concentrated *in vacuo* to give an oil, which yielded solid Tbfmoc-protein on trituration with diethyl ether. The solid Tbfmoc-protein was then filtered off, washed with diethyl ether and dissolved in aqueous acetic acid (50%) before being applied to the top of a column of Sephadex G-50 (10 x 800mm) and eluted with aqueous acetic acid (30%). Pure Tbfmoc-protein was obtained by semi-preparative Hplc (Vydac C18, 250 x 22mm, 10μM, A=H_2O, B=CH_3CN, 0.1% TFA; 9ml/min. 10-70% B over 40 min. λ=364nm) and concentrated to a small volume (5-10ml) by lyophilisation. The Tbfmoc was then deprotected by adding piperidine (1ml) and mixing for 5 minutes, before the solution was cooled on ice, acidified (AcOH), applied to the top of a column of Sephadex G-50 (10 x 800mm)and eluted with aqueous acetic acid (30%). The protein containing fractions were then pooled and pure MBD protein was obtained after semi-preparitive Hplc (Vydac C18, 250 x 22mm, 10μM, A=H_2O, B=CH_3CN, 0.1% TFA; 9ml/min. 10-70% B over 40 min. λ=214nm) and lyophilisation as a white solid (32.8mg); amino acid analysis (48Hr. hydrolysis): Asx_{10} 9.53, Thr_5 4.44, Ser_7 5.80, Glx_6 7.09, Pro_6 5.83, Gly_7 7.27, Ala_4 4.43, Val_4 4.00, Met_1 1.05, Ile_4 3.87, Leu_5 5.26, Tyr_4 4.12, Phe_4 4.09, Lys_8 8.08, Arg_9 9.16, Trp_1 (N/A); m/z (laser desorption) 9727, $C_{431}H_{683}N_{126}O_{129}S_1$ requires 9726.0; Hplc (Vydac C18, 250 x 4.6mm, 5μM, A=H_2O, B=CH_3CN, 0.1% TFA; 1ml/min. 10-70% B over 45 min. λ=214nm), Rt=

24.0 min., 42.5%B. In addition, the first twenty six amino acid residues were sequenced by Edman degradation and were found to be in agreement with the required sequence.

Oligodeoxyribonucleotide Synthesis and Purification

Synthetic DNA sequences (**53**) and (**54**) were obtained from OSWEL, University of Edinburgh, and purified using the method of Ramage and Wahl[87a,b]. DNA concentrations, after Hplc and Sephadex G-25 purification, were determined at 260nm and ambient temperature in aqueous solution.

Duplex formation was acheived by mixing equimolar amounts of each oligonucleotide strand, this solution was subsequently heated at 70°C for 5minutes and then allowed to cool very gradually to room temperature.

Bandshift Assay

Protein-DNA binding reactions were carried out in 20mM MES (pH6.3), 1mM EDTA, 3mM $MgCl_2$, 10 mM mercaptoethanol, 10% glycerol, 0.1% TRITON X-100 and 60mM NaCl. Increasing amounts of protein were incubated with the DNA on ice for 30 minutes, then loaded onto a 5% polyacrylamide gel and run for 4 hours. at 100v and 4°C. After 4 hours the gel was placed onto cling-film and then onto a t.l.c. plate which allowed visualisation of the bands by exposure to UV light (254nm).

Restriction Alleviation Protein - Synthesis A

The synthesis was carried out on a 0.2mM scale using the functionalised resin Fmoc Ala-($OCH_2C_6H_4OR$). All amino acid side chains were protected as described previously, with the exception of asparagine and glutamine which were incorporated as the corresponding 4,4'-dimethoxybenzhydryl (MBH) derivatives, and cysteine

which was incorporated as the t-butylthio (S^tBu) derivative. All residues (with the exception of glycine) were treble coupled and at the completion of the assembly the resin bound material was washed sequentially with DMF, 1,4-dioxan, dichloromethane and dried.

Cleavage of the protein from the resin with simultaneous deprotection of all side chain protecting groups, with the exception of Cys-S^tBu, was accomplished by stirring the 250mg of the dry resin bound protein with a mixture of TFA (30ml), EDT (0.75ml), water (1.5ml) amd phenol (2.25g) for 5 hours under an atmosphere of dry nitrogen. The protein was then isolated by evaporation of the TFA under reduced pressure and trituration with diethyl ether. The crude cleaved protein was applied to a column of Sephadex G-50 (10 x 800mm), eluted with aqueous acetic acid (30%) and the desired fractions were then pooled and lyophilised to give 40mg of solid protein. This material was then dissolved in 95%TFE/5% water (10ml) and to this stirring solution was added an excess of tributylphosphine (0.1ml). Hplc monitoring (Aquapore C8, 200 x 4.6mm, 7μM; A=H_2O, B=CH_3CN, 0.1% TFA, 1ml/min. 10-90%B over 25 minutes, λ=214nm) showed complete disappearance of starting material (Rt=14.2min., 57%B) and appearance of a less hydrophobic product (Rt=12.8min., 53%B) after 5 hours. The reduced protein was isolated by evaporation to a small volume (0.5ml) and trituration with diethyl ether. The product was isolated by filtration, dissolved in aqueous acetic acid (30%) and lyophilised to give 29mg of white solid; amino acid analysis: Asx_7 6.2, Thr_5 3.5, Ser_1 1.0, Glx_9 10.5, Pro_2 3.0, Gly_4 3.9, Ala_5 5.3, Cys_4 2.4, Val_3 3.2, Met_5 3.8, Ile_3 2.5, Leu_2 2.0, Tyr_1 1.2, Phe_2 2.0, Lys_8 7.5, Arg_3 2.5, Trp_2 N/A.

This material proved extremely insoluble in a variety of buffers, and attempts to further purify this synthetic protein proved unsuccessful. Polyacrylamide gel electrophoresis showed a broad smear which was indicative of the presence of multiple products of similar size. This was probably due to incomplete coupling

and/or capping during assembly and resulted in an accumulation of similar deletion sequences which co-purified with the desired sequence.

Restriction Alleviation Protein - Synthesis B

The synthesis was carried out on a 0.22mM scale using Fmoc-Alanine functionalised 2-methoxy-4-alkoxybenzylalcohol (SASRIN) resin (490mg, 0.45mM/g). All amino acid side chains were protected as described previously, with the exception of asparagine and glutamine which were incorporated as the corresponding 4,4'-dimethoxybenzhydryl (MBH) derivatives, and cysteine which was incorporated as the acetamidomethyl (Acm) derivative. All amino acids (with the exception of glycine) were double coupled, with the first 12 residues being coupled in a glass vessel in a sonic bath connected to the ABI 430A peptide synthesiser. The resin bound product (Fmoc Ral 66-55) was then transferred back into a conventional reaction vessel and the synthesis continued using standard cycles except for the regions 54-45 and 30-1 which were allowed to couple for an extended period of time. At the end of the synthesis the Fmoc was left on the N-terminal methionine, and the resin bound product was capped for 30 minutes in a sonic bath. The resin was then washed with DMF, 1,4-dioxan, dichloromethane and dried to give 1.6g of Fmoc resin bound protein.

A 650mg portion of the resin was then treated with 20% piperidine/DMF for 30 minutes in a sonic bath befor being filtered off, washed with DMF, 1,4-dioxan, dichloromethane and dried. The dry resin bound protein was then added to a suspension of TbfmocCl (100mg, 0.22mM) in dichloromethane (10ml) and DIEA (30µl, 0.17mM) was added before the reaction flask was sealed, covered in aluminium foil and then sonicated for 5 hours with occasional mixing. The Tbfmoc-peptide-resin was then filtered off, washed thoroughly with dichloromethane and dried. The dry Tbfmoc resin bound protein (functionality by UV[75] =0.054mM/g) was

then added to a mixture of TMSBr (1.32ml), EDT (2ml), thioanisole (1.2ml), m-cresol (0.1ml) and TFA (10ml) at 0°C under an atmosphere of dry nitrogen, and allowed to stir at 0°C for 5 hours whilst protected from daylight. The resin was then removed by filtration and the filtrate concentrated *in vacuo* to give an oil, which yielded solid Tbfmoc-protein on trituration with diethyl ether. The solid Tbfmoc-protein was then filtered off, washed with diethyl ether and dissolved in aqueous acetic acid (50%) before being applied to the top of a column of Sephadex G-50 (10 x 800mm)and eluted with aqueous acetic acid (30%). Pure Tbfmoc-protein was obtained by semi-preparative Hplc (Vydac C18, 250 x 22mm, 10μM, A=H_2O, B=CH_3CN, 0.1% TFA; 10ml/min. 10-90% B over 30 min. λ=364nm) and concentrated to a small volume (5-10ml) by lyophilisation. The Tbfmoc was then deprotected by adding piperidine (1ml) and mixing for 5 minutes, before the solution was cooled on ice,acidified (AcOH), applied to the top of a column of Sephadex G-50 (10 x 800mm)and eluted with aqueous acetic acid (30%). The protein containing fractions were then pooled and pure tetra Acm protected Ral protein was obtained after semi-preparitive Hplc (Vydac C18, 250 x 22mm, 10μM, A=H_2O, B=CH_3CN, 0.1% TFA; 10ml/min. 10-90% B over 30 min. λ=214nm) and lyophilisation as a white solid (19mg); amino acid analysis: Asx_7 6.58, Thr_5 4.68, Ser_1 0.99, Glx_9 10.37, Pro_2 3.95, Gly_4 3.99, Ala_5 5.00, Cys_4 1.88, Val_3 3.10, Met_5 5.00, Ile_3 2.86, Leu_2 2.11, Tyr_1 0.48, Phe_2 2.01, Lys_8 7.82, Arg_3 2.71, Trp_2 (N/A); m/z (laser desorption) 7885, $C_{338}H_{541}N_{97}O_{102}S_9$ requires 7889.34; Hplc (Vydac C18, 250 x 4.6mm, 5μM, A=H_2O, B=CH_3CN, 0.1% TFA; 1ml/min. 10-90% B over 30 min. λ=214nm), Rt= 15.4 min., 59%B.

The stable Acm protecting groups on the four cysteine residues were removed by dissolving the pure $Ral(Acm)_4$ (35mg) in water (20ml) and adjusting the pH to 4.0 with ammonium hydroxide (2M). Mercury (II) acetate (100mg) was then added and the mixture was allowed to stir overnight at room temperature. This solution was

then lyophilised, before the protein was fully reduced and denatured by dissolving the residue in 6M Gdm.HCl (50mM TRIS pH8.0) containing 20% v/v mercaptoethanol and stirring overnight. This solution was then applied to a column of Sephadex G-50 (10 x 800mm) and eluted with aqueous acetic acid (30%). Further purification of the protein containing fractions by Hplc (Aquapore C18, 250 x 10mm, 10μM, A=H_2O, B=CH_3CN, 0.1% TFA; 5ml/min. 0-50% B over 30 min. λ=214nm) gave 10mg white solid after lyophilisation; amino acid analysis: Asx_7 7.04, Thr_5 4.74, Ser_1 1.32, Glx_9 10.70, Pro_2 2.44, Gly_4 4.05, Ala_5 4.77, Cys_4 2.43, Val_3 3.11, Met_5 4.42, Ile_3 3.13, Leu_2 2.06, Tyr_1 1.14, Phe_2 2.11, Lys_8 7.87, Arg_3 3.33, Trp_2 (N/A); m/z (laser desorption) 7669*, $C_{326}H_{526}N_{93}O_{98}S_9$ requires 7604.94; Hplc (Aquapore C18, 220 x 4.6mm, 7μM, A=H_2O, B=CH_3CN, 0.1% TFA; 1ml/min. 10-50% B over 30 min. λ=214nm), Rt= 26.2 min., 45%B. The first 22 amino acid residues were sequenced by Edman degradation and found to be in agreement with the required sequence. *Ellman assay for free thiols[112] gave a value of one thiol per molecule (requires 4), and although initially this was thought to be due either to rapid oxidation of the reduced protein, recent evidence from high resolution mass spectra suggests that the cysteine residues were incompletely deprotected (see Section 3.3.2.1).

N-Hydroxymethyl-4-methoxyphenylacetamide

This compound was prepared by the general method of Ben-Ishai *et al*[163]. A mixture of 4-methoxyphenylacetamide (5.38g, 0.032M), aqueous formaldehyde (40% w/v, 12.5ml) and aqueous sodium hydrogen carbonate (5% w/v, 7.5ml) were heated on a boiling water bath until a clear solution was obtained (~10 min.). The reaction mixture was cooled for one hour and then extracted with ethyl acetate (3 x 25ml). The organic layer was dried ($MgSO_4$), filtered and evaporated to give a white solid which was triturated with diethyl ether, filtered off and dried to yield 4.89g (78.3%) of the *title compound* as a white powder; m.p.117-118°C; (found:

C,61.39;H,6.90;N,7.13, $C_{10}H_{13}NO_3$ requires: C,61.52; H,6.71; N,7.18%);t.l.c. ($CHCl_3$/MeOH/AcOH, 85:10:5) R_f 0.55, ($CH_3CN/CHCl_3$/AcOH, 8:1:1) R_f 0.6; υ_{max}(KBr disc) 3320 (amide NH), 3254 (OH), 1655 (C=O), 1612, 1543, 1511 cm^{-1} (aromatic); λ_{max}(MeOH) 277nm (ε 17800 $dm^3mol^{-1}cm^{-1}$), 284nm (16000); δH[$(CD_3)_2SO$, 200MHz] 8.60 (1H, Br t, NH); 7.21-6.82 (4H, q, aromatics); 5.59 (1H, t, J=6.81Hz, OH); 4.50 (2H,t, J=6.60Hz, NH$\underline{CH_2}$OH);3.71 (3H, s, OCH_3); 3.35 (2H, s, CH_2); δC[$(CD_3)_2SO$, 50MHz] 130.26 (aromatic CH); 128.21 (aromatic quarternary); 113.78 (aromatic CH); 62.61 (NH$\underline{CH_2}$OH); 55.15 (OCH_3); 41.84 (CH_2); m/z (FAB) 196 (MH^+), HRMS found 196.09737, $C_{10}H_{14}NO_3$ requires 196.09736 therefore <1ppm.

S-(4-Methoxyphenylacetamidomethyl)-L-cysteine

TFA (12ml) was added to a stirring suspension of L-cysteine (free base) (0.9g, 7.4mM) in dichloromethane (60ml) at 0°C, and a suspension of N-hydroxymethyl-4-methoxyphenylacetamide (1.46g, 7.5mM) in dichloromethane (50ml) was then added portionwise over a 10 minute period. The reaction mixture was then evaporated and the residue chased with diethyl ether to give a white solid. This solid was then isolated by filtration, washed with water, EtOH and dried to give 1.46g of the *title compound* as a white powder; m.p.200°C (dec.); (found: C,52.51;H,6.18;N,9.59, $C_{13}H_{18}N_2O_4S$ requires: C, 52.33; H,6.08; N,9.39%); $[\alpha]_D^{24}$ = -49.1°(c=1 in TFA); t.l.c. (n-BuOH/pyridine/AcOH/water, 15:10:3:2) R_f 0.6; υ_{max}(KBr disc) 3440 (OH), 3200 (NH), 1639 (C=O), 1613, 1542, 1513 cm^{-1} (aromatic); δH[$CF_3CO_2D/(CD_3)_2SO$, 200MHz] 8.39 (1H, Br s, NH); 6.89 (4H, AB q, J=8Hz, aromatics); 4.35 (1H, d, J=13.5Hz, αCH); 4.11 (2H,t, J=14Hz, NH$\underline{CH_2}$S); 3.59 (3H, s, OCH_3); 3.36 (2H, s, CH_2); 3.0 (2H, m, βCH_2); δC[$CF_3CO_2D/(CD_3)_2SO$, 50MHz] 175.02 (acid C=O); 171.59 (amide C=O); 160.67 (aromatic quarternary); 132.22 (aromatic CH); 129.19 (aromatic quarternary); 115.85 (aromatic CH); 56.40 (OCH_3); 54.39 (αCH); 43.35,42.98 (CH_2);32.33 (βCH_2); m/z (FAB) 299 (MH^+),

HRMS found 299.10657, $C_{13}H_{19}N_2O_4S$ requires 299.10654 therefore <1ppm.

N$^{\alpha}$(9-Fluorenylmethoxycarbonyl)-S-(4'-methoxyphenylacetamidomethyl)-L-cysteine (**70**)

4-Methoxyphenylacetamidomethyl-L-cysteine (0.69g, 2.31mM) was suspended in a mixture of 1,4-dioxan/water (1:1, 50ml) and the pH adjusted to between 8 and 9 with solid sodium carbonate. A solution of FmocOSu (0.56g, 1.67mM) in acetonitrile (25ml) was then added and allowed to stir overnight. The volume was then reduced by evaporation under reduced pressure, and then the pH was adjusted to 3 with 2M HCl. Extraction with ethyl acetate, washing with water, drying ($MgSO_4$) and evaporation yielded a white solid which, after trituration with hot ethyl acetate and cooling was isolated by filtration. The solid was then washed with ethyl acetate, diethyl ether, and dried to give 0.62g (71% based on FmocOSu) of the *title compound* as a white powder; m.p.152-154°C; (found: C,64.71;H,5.58;N,5.50, $C_{28}H_{28}N_2O_6S$ requires: C,64.60;H,5.42;N,5.38%); $[\alpha]_D^{24}$ = -32.8°(c=1 in DMF); t.l.c. ($CHCl_3$/MeOH/AcOH, 90:8:2) R_f 0.20, ($CHCl_3$/MeOH/AcOH, 85:10:5) R_f 0.6; υ_{max}(KBr disc) 3358 (OH), 3316 (NH), 3064 (aromatic CH), 2927 (aliphatic CH), 1697 (C=O), 1616 cm^{-1} (aromatic); λ_{max}(MeOH) 266nm (ε 26000$dm^3mol^{-1}cm^{-1}$), 290nm (6800), 300nm (7800); δH[$(CD_3)_2SO$, 200MHz] 12.80 (1H, Br s CO_2H); 8.69 (1H, Br s, NH); 7.9-7.28 (9H, m, Fmoc aromatics + NH); 7.19 (4H, AB q, J=8Hz, aromatics); 4.27 (6H, Br m, Fmoc CH_2CH+αCH+S$\underline{CH_2}$N); 3.71 (3H, s, OCH_3); 3.39 (2H, s, CH_2); 2.94 (2H, m, βCH_2); δC[$(CD_3)_2SO$, 50MHz] 172.56 (acid C=O); 170.92 (amide C=O); 158.08 (aromatic quarternary); 156.22 (urethane C=O); 143.95, 140.89 (aromatic quarternary); 130.16 (aromatic CH); 128.06 (aromatic quarternary); 127.83, 127.26, 125.47, 120.29, 113.82 (aromatic CH); 65.95 (Fmoc CH_2); 55.13 (OCH_3); 54.23 (αCH); 46.77 (Fmoc CH); 41.56, 40.61 (CH_2);31.84 (βCH_2); m/z (FAB) 521 (MH^+), HRMS found 521.17464, $C_{28}H_{29}N_2O_6S$ requires 521.17462 therefore <1ppm.

N$^{\alpha}$(9-Fluorenylmethoxycarbonyl)-S-(phenylacetamidomethyl)-L-cysteine (**71**)

This compound was prepared using the general method of Albericio *et al*[129,164]. To a suspension of N-hydroxymethylphenylacetamide[165] (6.9g, 42mM) and cysteine.HCl monohydrate (7.33g, 42mM) in water (15ml) under argon at 0°C was added a mixture of TFMSA/TFA (1:19, 50ml). After stirring at 0°C for 90 minutes the reaction mixture was evaporated and repeatedly chased with diethyl ether (5 x 50ml) to give an oil which was used immediately for the next step.

The above oil was suspended in water/1,4-dioxan (1:1, 500ml) and the pH adjusted to 9-10 with solid sodium carbonate. FmocOSu (11.33g, 33.6mM) in acetonitrile (100ml) was then added and the reaction was allowed to stir overnight before being poured into water (1000ml), acidified to pH 2 (2M HCl) and extracted with ethyl acetate. The organic layer was then washed with brine, dried ($MgSO_4$) and evaporated to give a white solid. Trituration with diethyl ether and drying under vacuum gave the *title compound* 14.17g (86% based on FmocOSu) as a white powder; m.p. 156-158°C; (found: C,66.64;H,5.97;N,5.71, $C_{27}H_{26}N_2O_5S$ requires: C,66.10;H,5.34;N,5.78%); $[\alpha]_D^{24}$ = -35.2°(c=1 in DMF); t.l.c.($CHCl_3$/MeOH/AcOH, 90:10:5) R_f 0.54; υ_{max}(KBr disc) 3370 (OH), 3000-2900 (CH), 1716 (urethane C=O), 1685 (acid C=O), 1626cm^{-1} (aromatic); λ_{max}(MeOH) 266nm (ε 24500dm^{3}mol^{-1}cm^{-1}), 290nm (5800), 300nm (7600); δH[$(CD_3)_2SO$, 200MHz] 13.00 (1H, Br s CO_2H); 8.77 (1H, t, J=6.16Hz, NH); 7.9-7.19 (14H, m, aromatics + NH); 4.37-4.21 (6H, Br m, Fmoc CH_2CH+αCH+S$\underline{CH_2}$N); 3.48 (2H, s, CH_2); 3.03 (2H, m, βCH_2); δC[$(CD_3)_2SO$, 50MHz] 172.54 (acid C=O); 170.57 (amide C=O); 156.22 (urethane C=O); 143.95, 140.89 ,136.18 (aromatic quarternary); 129.16, 128.40, 127.83, 127.26, 126.56, 125.47, 120.28 (aromatic CH); 65.95 (Fmoc CH_2); 54.22 (αCH); 46.78 (Fmoc CH); 42.44, 40.62 (CH_2); 31.86 (βCH_2); m/z (FAB) 492 (MH^+), HRMS found 491.16408, $C_{27}H_{27}N_2O_5S$ requires 491.16405 therefore <1ppm.

Salmon Calcitonin 1-10 (CSNLSTCVLG)

The above sequence was synthesised twice on a 0.25mM scale using the functionalised resin Fmoc-Gly-($OCH_2C_6H_4OR$). All amino acid side chains were protected as described previously with the exception of cysteine. In synthesis A the cysteine residues were incorporated as the phenylacetamidomethyl derivative (**71**), and in synthesis B as the 4-methoxyphenylacetamidomethyl derivative (**70**). All residues were double coupled, and at the completion of the assembly phase the resin bound product was washed sequentially with DMF, 1,4-dioxan, dichloromethane and dried.

Synthesis A

Cleavage of the peptide from the resin and simultaneous removal of all side chain protecting groups, with the exception of the acid stable phenylacetamidomethyl cysteine protection, was accomplished by stirring 650mg of the resin bound peptide with a mixture of EDT (2ml) and TFA/water (95:5, 10ml) for 3 hours at room temperature. Evaporation of the TFA and precipitation with diethyl ether gave the bis-phenylacetamidomethyl cysteine protected peptide (**73**) as a white solid (80mg); Hplc (Vydac C18, 250 x 4.6mm, 5μM, A=H_2O, B=CH_3CN, 0.1% TFA; 1ml/min. 10-90% B over 30 min. λ=214nm), Rt= 16.4 min., 54%B; m/z (FAB) 1330 (MK^+), 1315 (MNa^+),1291 (MH^+), HRMS found 1290.58699, $C_{57}H_{88}N_{13}O_{17}S_2$ requires 1290.58626 therefore <+/- 1ppm.

Synthesis B

Cleavage of the peptide from the resin and simultaneous removal of all side chain protecting groups, with the exception of the acid stable 4-methoxyphenylacetamidomethyl cysteine protection, was accomplished by stirring 1g of the resin bound peptide with a mixture of EDT (2ml) and TFA/water (95:5, 10ml) for 3 hours at room temperature. Evaporation of the TFA and precipitation with diethyl ether gave the bis-phenylacetamidomethyl cysteine protected peptide (**74**) as a white solid (150mg); Hplc (Vydac C18, 250 x 4.6mm, 5μM, A=H_2O,

B=CH_3CN, 0.1% TFA; 1ml/min. 10-90% B over 30 min. λ=214nm), Rt= 19.8 min., 63.5%B; m/z (FAB) 1372 (MNa^+), 1350 (M^+), HRMS found 1350.61782, $C_{59}H_{91}N_{13}O_{19}S_2$ requires 1350.60739 therefore <+/- 10ppm.

Conversion of the 4-methoxyphenylacetamidomethyl cysteine protected peptide (**74**) to the 4-hydroxyphenylacetamidomethyl protected peptide (**75**) was carried out according to the method of Kiso *et al*[131]. A 30mg portion of the 4-methoxyphenylacetamidomethyl cysteine protected peptide (**74**) was treated with a mixture of TFMSA (0.9ml, 10mM), thioanisole (0.6ml, 5mM) and TFA (5ml) for 30 minutes at 0°C and then for a further 2 hours at room temperature. Evaporation of the TFA and gel filtration (Sephadex G-15, 30% AcOH) to remove TFMSA salts gave peptide containing fractions which were further purified by semi-preparative Hplc (Vydac C18, 250 x 22mm, 10μM, A=H_2O, B=CH_3CN, 0.1% TFA; 9ml/min. 10-70% B over 30 min. λ=214nm) to yield 21mg of 4-hydroxyphenylacetamidomethyl cysteine protected peptide (**75**) on lyophilisation; Hplc (Vydac C18, 250 x 4.6mm, 5μM, A=H_2O, B=CH_3CN, 0.1% TFA; 1ml/min. 10-90% B over 30 min. λ=214nm), Rt= 14.8 min., 50%B; m/z (FAB) 1322 (MH^+), HRMS found 1322.57690, $C_{57}H_{88}N_{13}O_{19}S_2$ requires 1322.57609 therefore <+/- 1ppm.

Enzyme Mediated Deprotection of Cysteine Protected Salmon Calcitonin 1-10

As a general method, a 15mg portion each of the bis-cysteine protected peptides (**73**), (**74**),and (**75**) was added to ammonium acetate buffer (20ml, 0,1M,pH8.5) and the pH re-adjusted to 8.5. Resin bound penicillin acylase (0.5g, 1325mega units/Kg, supplied by SmithKline Beecham Pharmaceuticals Ltd.) was added and the reaction mixture was allowed to gently mix whilst incubating at 37°C. The progress of the reaction was monitored by Hplc (figure 21, section 3.3.3.1). Substrates (**73**) and (**75**) were fully deprotected after overnight incubation, whilst substrate (**74**) required 9 days incubation for full deprotection. All substrates yielded the expected cyclic

disulphide bridged peptide (**76**) as the major product of the enzyme reaction, although only substrate (**75**), the 4-hydroxyphenylacetamidomethyl protected derivative, gave virtually quantitative conversion to (**76**).

At the completion of the reaction, the resin bound enzyme was removed by filtration and washed with water. The combined filtrate was then lyophilised and gel filtered (Sephadex G-15, 30% AcOH) before the desired product (**76**) was isolated by Hplc. Hence 15mg of bis-4-hydroxymethylphenylacetamidomethyl cysteine protected SC1-10 yielded 4.2mg of purified SC1-10 (oxidised) after lyophilisation; amino acid analysis: Asx_1 1.25, Thr_1 0.95, Ser_2 1.68, Gly_1 1.02, Cys_2 1.89, Val_1 1.03, Leu_2 1.84; m/z (FAB) 1032 (MK^+), 1016 (MNa^+), 994(M^+), HRMS found 994.44320, $C_{39}H_{67}N_{11}O_{15}S_2$ requires 994.43378 therefore < +/- 10ppm.; Hplc (Vydac C18, 250 x 4.6mm, 5μM, A=H_2O, B=CH_3CN, 0.1% TFA; 1ml/min. 10-90% B over 30 min. λ=214nm), Rt= 13.4 min., 47%B.

Restriction Alleviation Protein - Synthesis C

The synthesis was carried out on a 0.153mM scale using Fmoc-Alanine functionalised 2-methoxy-4-alkoxybenzylalcohol (SASRIN) resin (340mg, 0.45mM/g). All amino acid side chains were protected as described previously, with the exception of asparagine and glutamine which were incorporated as the corresponding triphenylmethyl (trityl) derivatives, and cysteine which was incorporated as the 4-methoxyphenylacetamidomethyl (4-MeOPhacm) derivative (**70**). All amino acids (with the exception of glycine) were double coupled, with the regions 54-45 and 30-1 being coupled for an extended period of time during the second coupling cycle. At the end of the synthesis the Fmoc was left on the N-terminal methionine, and the resin bound product was capped for 30 minutes in a sonic bath. The resin was then washed with DMF, 1,4-dioxan, dichloromethane and

dried to give 1.28g of Fmoc resin bound protein.

A 400mg portion of the resin was then treated with 20% piperidine/DMF for 15 minutes in a sonic bath befor being filtered off, washed with DMF, 1,4-dioxan, dichloromethane and dried. The dry resin bound protein was then added to a suspension of TbfmocCl (200mg, 0.44mM) in dichloromethane (10ml) and DIEA (30µl, 0.17mM) was added before the reaction flask was sealed, covered in aluminium foil and then sonicated for 3 hours with occasional mixing. The Tbfmoc-peptide-resin was then filtered off, washed thoroughly with dichloromethane and dried. The dry Tbfmoc resin bound protein (functionality by UV[75] =0.069mM/g) was then added to a mixture of TMSBr (1.32ml), EDT (2ml), thioanisole (1.2ml), m-cresol (0.1ml) and TFA (10ml) at 0°C under an atmosphere of dry nitrogen, and allowed to stir at 0°C for 5 hours whilst protected from daylight. The resin was then removed by filtration and the filtrate concentrated *in vacuo* to give an oil, which yielded solid Tbfmoc-protein on trituration with diethyl ether. The solid Tbfmoc-protein was then filtered off, washed with diethyl ether and dissolved in aqueous acetic acid (50%) before being applied to the top of a column of Sephadex G-50 (10 x 800mm)and eluted with aqueous acetic acid (30%). Pure Tbfmoc-protein was obtained by semi-preparative Hplc (Vydac C18, 250 x 22mm, 10µM, A=H_2O, B=CH_3CN, 0.1% TFA; 9ml/min. 30-90% B over 25 min. λ=364nm) and concentrated to a small volume (5-10ml) by lyophilisation. The Tbfmoc was then deprotected by adding piperidine (1ml) and mixing for 5 minutes, before the solution was cooled on ice,acidified (AcOH), applied to the top of a column of Sephadex G-50 (10 x 800mm)and eluted with aqueous acetic acid (30%). The protein containing fractions were then pooled and lyophilised to give 108mg tetra 4-MeOPhacm protected Ral protein as a white solid; amino acid analysis: Asx_7 6.38, Thr_5 4.52, Ser_1 1.02, Glx_9 11.26, Pro_2 2.41, Gly_4 3.94, Ala_5 5.50, Cys_4 3.29, Val_3 3.09, Met_5 4.29, Ile_3 3.35, Leu_2 2.08, Tyr_1 0.41, Phe_2 2.02, Lys_8 7.54, Arg_3 2.90, Trp_2 (N/A);

m/z (laser desorption) 8337.6 (Na adduct), $C_{367}H_{574}N_{97}O_{106}S_9$ requires 8313.8; Hplc (Vydac C18, 250 x 4.6mm, 5μM, A=H_2O, B=CH_3CN, 0.1% TFA; 1ml/min. 10-90% B over 30 min. λ=214nm), Rt= 18.4 min., 58.5%B.

Cleavage of the methyl group at the phenolic oxygen of the 4-MeOPhacm cysteine protecting group was carried out according to the method of Kiso *et al*[131]. TFMSA (0.9ml) was added to an ice cold stirring solution of 4-MeOPhacm protected Ral in a mixture of TFA/thioanisole (100:6, 10.6ml) under an atmosphere of nitrogen. After stirring at 0°C for 60 minutes the reaction mixture was allowed to warm up to room temperature and was stirred for a further 2 hours. The TFA was removed by evaporation and the crude 4-hydroxyphenylacetamidomethyl cysteine protected protein was precipitated with diethyl ether (250ml) and filtered off to give a hygroscopic solid. This solid was then redissolved in 6M Gdm.HCl (50mM TRIS pH 8.0) containing DTT (200mg) and stirred for 30 minutes at 37°C, before being applied to a column of Sephadex G-50 (10 x 800mm) and eluted with aqueous acetic acid (30%). The protein containing fractions were then pooled and lyophilised to give a white solid; Hplc (Vydac C18, 250 x 4.6mm, 5μM, A=H_2O, B=CH_3CN, 0.1% TFA; 1ml/min. 10-90% B over 30 min. λ=214nm), Rt= 17.0 min., 55%B.

The freeze dried 4-hydroxyphenylacetamidomethyl cysteine protected protein was dissolved in ammonium acetate buffer (0.1M, 50ml, pH 8.5) and the pH re-adjusted to 8.5 (2M ammonium hydroxide). To this solution was then added 2-mercaptoethanol (10ml),0.5g of resin bound penicillin acylase (Sigma). The flask was then flushed with argon and sealed, before being incubated at 37°C with gentle agitation. After 72 hours an aliquot of soluble enzyme (0.1ml, Sigma) was added and incubated for a further 72 hours. The resin bound enzyme was then removed by filtration and the filtrate lyophilised to give solid material. This solid was redissolved in 6M Gdm.HCl (0.1M TRIS pH8.5) containing DTT (excess) and incubated at 37°C

for 2hours, before being desalted by gel filtration (Sephadex G-50, 30% AcOH). Hplc analysis of the Ral containing fractions showed three main peaks which, on isolation by Hplc, gave identical mass specta corresponding to incompletely deprotected protein. This suggested that the enzyme reaction had failed to completely remove the enzyme cleavable cysteine protection, although further treatment of the partially deprotected protein with fresh enzyme gave no improvement in Hplc profile.

Restriction Alleviation Protein - Synthesis D

The synthesis was carried out on a 0.25mM scale using Fmoc-Alanine functionalised 4-alkoxybenzylalcohol resin (400mg, 0.614mM/g). All amino acid side chains were protected as described previously, with the exception of asparagine, glutamine and cysteine which were incorporated as the corresponding triphenylmethyl (trityl) derivatives. All amino acids (with the exception of glycine) were double coupled, with the regions 54-45 and 30-1 being coupled for an extended period of time during the second coupling cycle. At the end of the synthesis the Fmoc was left on the N-terminal methionine, and the resin bound product was capped for 30 minutes in a sonic bath. The resin was then washed witn DMF, 1,4-dioxan, dichloromethane and dried to give 2.18g of Fmoc resin bound protein.

The resin was then treated with 20% piperidine/DMF for 15 minutes in a sonic bath befor being filtered off, washed with DMF, 1,4-dioxan, dichloromethane and dried. The dry resin bound protein was then added to a suspension of TbfmocCl (400mg, 0.88mM) in dichloromethane (20ml) and DIEA (60µl, 0.34mM) was added before the reaction flask was sealed, covered in aluminium foil and then sonicated for 3 hours with occasional mixing. The Tbfmoc-peptide-resin was then filtered off, washed thoroughly with dichloromethane and dried.

A 500mg portion of the dry Tbfmoc resin bound protein was then added to a mixture of EDT (2ml), thioanisole (0.5ml), phenol (0.75g), water (0.5ml), TIS (0.5ml) and stirred for 15 minutes, before TFA (10ml) was added and the mixture stirred for a further 4 hours hours whilst protected from daylight. The resin was then removed by filtration and the filtrate concentrated *in vacuo* to give an oil, which yielded solid Tbfmoc-protein on trituration with diethyl ether. The solid Tbfmoc-protein was then filtered off, washed with diethyl ether and dissolved in aqueous acetic acid (50%) before being applied to the top of a column of Sephadex G-50 (10 x 800mm)and eluted with aqueous acetic acid (30%). Pure Tbfmoc-protein was obtained by semi-preparative Hplc (Vydac C18, 250 x 22mm, 10μM, A=H_2O, B=CH_3CN, 0.1% TFA; 9ml/min. 30-90% B over 25 min. λ=364nm) and lyophilised to give 140mg white solid. The Tbfmoc was then deprotected under reducing conditions (6M Gdm.HCl, pH 8.5 TRIS, 0.1M, containing excess DTT) for 4 hours at 37°C before being desalted (Sephadex G-50, 30% AcOH) and purified by Hplc to give 40mg of fully reduced Ral protein after lyophilisation; Asx_7 5.11, Thr_5 4.74, Ser_1 1.07, Glx_9 8.08, Pro_2 2.02, Gly_4 4.37, Ala_5 5.52, Cys_4 0.99, Val_3 3.17, Met_5 4.48, Ile_3 3.12, Leu_2 2.31, Tyr_1 0.71, Phe_2 2.10, Lys_8 8.17, Arg_3 3.11, Trp_2 (N/A); m/z (laser desorption) 7607.3, $C_{326}H_{526}N_{93}O_{98}S_9$ requires 7604.49; Hplc (Vydac C18, 250 x 4.6mm, 5μM, A=H_2O, B=CH_3CN, 0.1% TFA; 1ml/min. 10-90% B over 30 min. λ=214nm), Rt= 17 min., 56%B.

Folded oxidised Ral was obtained by dissolving 29mg of fully reduced protein in 1M guanidine.HCl (0.1M, TRIS, pH 8.5, 250ml) containing 0.1mM EDTA, 0.3mM reduced and 3mM oxidised glutathione. After stirring for 2 days at room temperature, the single product was isolated by Hplc to give 12mg of folded protein after lyophilisation; Asx_7 4.26, Thr_5 3.86, Ser_1 1.07, Glx_9 8.08, Pro_2 1.44, Gly_4 4.91, Ala_5 5.32, Cys_4 2.30, Val_3 3.42, Met_5 4.29, Ile_3 3.17, Leu_2 2.39, Tyr_1 0.80, Phe_2 2.22, Lys_8 7.39, Arg_3 3.10, Trp_2 (N/A); m/z (laser desorption) 7624.1 (Na adduct),

$C_{326}H_{522}N_{93}O_{98}S_9$ requires 7600.9; Hplc (Vydac C18, 250 x 4.6mm, 5μM, A=H_2O, B=CH_3CN, 0.1% TFA; 1ml/min. 10-90% B over 30 min. λ=214nm), Rt= 17 min., 56%B.

MCP-1 (Ac7-76)

The synthesis was carried out on a 0.086mM scale using Fmoc-Thr(O^tBu) functionalised 4-alkoxybenzylalcohol resin (1.06g, 0.081mM/g). All amino acid side chains were protected as described previously, with the exception of asparagine, glutamine and cysteine which were incorporated as the corresponding triphenylmethyl (trityl) derivatives, and all amino acids were double coupled.

The resin (1.8g) was then added to a mixture of EDT (2ml), thioanisole (0.5ml), phenol (0.75g), water (0.5ml), TIS (0.5ml) and stirred for 15 minutes, before TFA (10ml) was added and the mixture stirred for a further 4 hours. The resin was then removed by filtration and the filtrate concentrated *in vacuo* to give an oil, which yielded solid protein on trituration with diethyl ether. The solid protein was then filtered off, washed with diethyl ether and dissolved in aqueous acetic acid (50%) before being applied to the top of a column of Sephadex G-50 (10 x 800mm)and eluted with aqueous acetic acid (30%). The protein containing fractions were then pooled and lyophilised prior to folding.

Folded oxidised Ac-MCP-1^{7-76} was obtained by dissolving 55mg of fully reduced protein in 1M guanidine.HCl (0.1M, TRIS, pH 8.5, 250ml) containing 0.1mM EDTA, 0.3mM reduced and 3mM oxidised glutathione. After stirring for 3 days at room temperature the folding solution was acidified (AcOH) and loaded onto a semi-preparative Hplc column (Vydac C18, 250 x 22mm, 10μM). Elution with a linear gradient of CH_3CN/0.1%TFA gave pure folded Ac-MCP-1^{7-76} (50mg) as a white solid after lyophilisation; Asx_6 5.88, Thr_7 6.39, Ser_5 4.28, Glx_7 8.14, Pro_4 3.30, Ala_5

5.00, Cys_4 3.81, Val_5 4.72, Met_1 1.72, Ile_5 4.79, Leu_2 3.35, Tyr_2 1.34, Phe_2 1.96, His_1 1.08,Lys_9 9.38, Arg_4 3.92, Trp_1 (N/A); m/z (laser desorption) 8088.5, $C_{352}H_{572}N_{101}O_{104}S_5$ requires 8086; Hplc (Vydac C18, 250 x 4.6mm, 5μM, A=H_2O, B=CH_3CN, 0.1% TFA; 1ml/min. 10-90% B over 30 min. λ=214nm), Rt= 16.6 min., 54%B.

MCP-1

The synthesis was carried out as described above. However, coupling of residues 46-39 and 10-1 was extended. All amino acids were incorporated as described above, with the exception of glutamine[1] which was incorporated unprotected. At the completion of the assembly the N-terminal Fmoc group was removed automatically on the synthesiser and the resin bound product washed with DMF, 1,4-dioxan, dichloromethane and dried. The dry resin bound protein was then added to a suspension of TbfmocCl (400mg, 0.88mM) in dichloromethane (20ml) and DIEA (70μl, 0.4mM) was added before the reaction flask was sealed, covered in aluminium foil and then sonicated for 3 hours with occasional mixing. The Tbfmoc-peptide-resin was then filtered off, washed thoroughly with dichloromethane and dried.

A 500mg portion of the dry Tbfmoc resin bound protein was then added to a mixture of EDT (2ml), thioanisole (0.5ml), phenol (0.75g), water (0.5ml), TIS (0.5ml) and stirred for 15 minutes, before TFA (10ml) was added and the mixture stirred for a further 4 hours hours whilst protected from daylight. The resin was then removed by filtration and the filtrate concentrated *in vacuo* to give an oil, which yielded solid Tbfmoc-protein on trituration with diethyl ether. The solid Tbfmoc-protein was then filtered off, washed with diethyl ether and dissolved in aqueous acetic acid (50%) before being applied to the top of a column of Sephadex G-50 (10 x 800mm)and eluted with aqueous acetic acid (30%). Pure Tbfmoc-protein was obtained by semi-preparative Hplc (Vydac C18, 250 x 22mm, 10μM, A=H_2O, B=CH_3CN, 0.1% TFA;

9ml/min. 30-90% B over 25 min. λ=364nm) and lyophilised. The Tbfmoc was then deprotected under reducing conditions (6M Gdm.HCl, pH 8.5 TRIS, 0.1M, containing excess DTT) for 4 hours at 37°C before being desalted (Sephadex G-50, 30% AcOH). The protein containing fractions were then pooled and lyophilised prior to folding.

Folded oxidised MCP-1 was obtained by dissolving the fully reduced protein in 1M guanidine.HCl (0.1M, TRIS, pH 8.5, 250ml) containing 0.1mM EDTA, 0.3mM reduced and 3mM oxidised glutathione. After stirring for 3 days at room temperature the folding solution was acidified (AcOH) and loaded onto a semi-preparative Hplc column (Vydac C18, 250 x 22mm, 10μM). Elution with a linear gradient of CH_3CN/0.1%TFA gave pure folded MCP-1 (4mg) as a white solid after lyophilisation; Asx_8 7.84, Thr_7 6.49, Ser_5 4.78, Glx_8 9.90, Pro_5 4.93, Ala_6 6.77, Cys_4 3.44, Val_5 4.73, Met_1 1.02, Ile_6 5.76, Leu_2 2.29, Tyr_2 1.37 , Phe_2 2.23, His_1 1.06, Lys_9 8.27, Arg_4 4.13, Trp_1 (N/A); m/z (laser desorption) 8681.2, $C_{379}H_{614}N_{109}O_{114}S_5$ requires 8682; Hplc (Vydac C18, 250 x 4.6mm, 5μM, A=H_2O, B=CH_3CN, 0.1% TFA; 1ml/min. 10-90% B over 30 min. λ=214nm), Rt= 15.8 min., 52.5%B.

Disulphide Mapping by Partial Tryptic Digest

As a general method, the protein of interest (0.5mg) was dissolved in ammonium acetate buffer (0.1M, pH 7.5) containing 0.1mg of trypsin. This sample was then incubated at 37°C for 4-7 hours before the digestion was stopped by adding 6M HCl (5μl). The sample was then lyophilised and analysed by laser desorption mass spectrometry.

N$^{\alpha}$-9-Fluorenylmethoxycarbonyl-α-hydroxyglycine (**83**)

9-Fluorenymethylcarbamate[166] (1g, 4.2mM) and glyoxylic acid monohydrate (2g, 0.022M) were gently refluxed in ethyl acetate (100ml) for 6 hours. The reaction mixture was then cooled, washed with water (5 x 50ml), dried ($MgSO_4$) and evaporated to give a white solid. Recrystallisation from ethyl acetate/hexane gave the *title compound* as a white solid, (1.19g, 91.9%); m.p. 214-216.5°C; (found: C,65.3;H,5.06;N,4.53, $C_{17}H_{15}NO_5$ requires: C,65.2;H,4.83;N,4.47%); t.l.c. ($CH_3CN/CHCl_3$/AcOH, 8:1:1) R_f 0.22; υ_{max}(KBr disc) 3600-3100 br (CO_2H and OH), 3320 (NH), 1730 (acid C=O), 1700cm^{-1} (urethane C=O); λ_{max}(MeOH) 266nm (ε 22900dm^{3}mol^{-1}cm^{-1}), 289nm (6000), 300nm (7200); δH[$(CD_3)_2SO$, 80MHz] 8.0-7.21 (8H, m, aromatics); 5.28 (1H, d, J=8.8Hz, αCH); 4.27 (3H, Br s, Fmoc CH_2CH); δC[$(CD_3)_2SO$, 50MHz] 169.6 (acid C=O); 154.8 (urethane C=O); 143.10, 140.30 (aromatic quarternary); 126.8, 126.2, 124.4, 119.0 (aromatic CH); 76.2 (αCH); 65.80 (Fmoc CH_2);46.00 (Fmoc CH); m/z (FAB) 314 (MH^+), HRMS found 313.09504, $C_{17}H_{15}NO_5$ requires 313.09502 therefore <1ppm.

N$^{\alpha}$-9-Fluorenylmethoxycarbonyl-α-methoxyglycine methyl ester(**84**)

To an ice cold solution of Fmoc-α-hydroxyglycine (**83**) (13.1g, 0.04M) in absolute methanol (150ml) was added conc. sulphuric acid (2ml). This solution was then allowed to warm up to room temperature and stirring was continued for 2 days. The resulting suspension was then poured into ice cold sodium hydrogen carbonate (saturated, 200ml) and extracted with ethyl acetate (100ml). The organic layer was then washed with water, dried ($MgSO_4$) and evaporated to give a white solid. Recrystallisation from ethanol gave the *title compound* as a white solid (11.6g, 82%); m.p. 123-125.5°C; (found: C,66.55;H,5.49;N,4.08, $C_{19}H_{19}NO_5$ requires: C,66.84;H,5.61;N,4.10%); t.l.c. (ethyl acetate/hexane, 1:1) R_f 0.44; υ_{max}(nujol) 3290 (NH), 1760 (ester C=O), 1700cm^{-1} (urethane C=O); λ_{max}(MeOH) 266nm (ε 22000dm^{3}mol^{-1}cm^{-1}), 290nm (5600), 300nm (6800); δH[$CDCl_3$, 200MHz] 7.78-

7.25 (8H, m, aromatics); 5.97 (1H,d, J=9.54Hz, NH);5.36 (1H, d, J=9.48Hz, αCH); 4.55-4.38 (2H, m, Fmoc CH_2); 4.23 (1H, t,J=6.70Hz, Fmoc CH); 3.80 (3H, s, ester OMe); 3.43 (3H, s, ether OMe); δC[$CDCl_3$, 50MHz] 167.83 (ester C=O); 155.54 (urethane C=O); 143.39, 143.26, 141.07 (aromatic quarternary); 127.56,126.87, 124.74, 119.80 (aromatic CH); 80.37 (αCH); 66.99 (Fmoc CH_2); 55.96 (ether OMe); 52.69 (ester OMe); 46.79 (Fmoc CH); m/z (FAB) 342 (MH^+), HRMS found 342.13416, $C_{19}H_{20}NO_5$ requires 342.13414 therefore <1ppm.

N$^{\alpha}$-9-Fluorenylmethoxycarbonyl-α-methoxyglycine (**85**)

A solution of $LiOH.H_2O$ (0.615g, 0.015M) in water (20ml) was slowly added to a stirring ice sold solution of Fmoc-α-methoxyglycine methyl ester (**84**) (5g, 0.015M) in acetone/water (4:1, 400ml). The resulting suspension was then stirred for a further 30 minutes before evaporating to a small volume (~80ml) to remove most of the acetone. Extraction with ethyl acetate (200ml) to remove any unreacted starting material was followed by acidification to pH 3.0 (2M sulphuric acid) to give a solid which was extracted into ethyl acetate (100ml), washed with water (2 x 50ml), dried ($MgSO_4$) and evaporated to give an oil which solidified on standing. Recrystallisation from ethyl acetate/hexane gave the *title compound* as a white solid (4g, 83%); m.p. 134-137°C; (found: C,65.94;H,5.26;N,4.19, $C_{18}H_{17}NO_5$ requires: C,66.03;H,5.24;N,4.28%); t.l.c. (methanol/dichloromethane, 1:1) R_f 0.70; υ_{max}(nujol) 3281 (NH), 1739 (acid C=O), 1696cm^{-1} (urethane C=O); λ_{max}(MeOH) 266nm (Ɛ 24000$dm^3mol^{-1}cm^{-1}$), 290nm (6100), 300nm (7500); δH[$(CD_3)_2SO$, 80MHz] 13.00 (1H, br s, CO_2H); 8.25 (1H,d, NH); 7.95-7.23 (8H, m, aromatics);5.06 (1H, d, αCH); 4.38-4.25 (3H, m, Fmoc CH_2CH); 3.24 (3H, s, ether OMe); δC[$(CD_3)_2SO$, 50MHz] 169.05 (acid C=O); 156.21 (urethane C=O); 143.92, 143.82, 140.91 (aromatic quarternary); 127.84,127.24, 125.49, 120.27 (aromatic CH); 80.68 (αCH); 66.13 (Fmoc CH_2); 54.75 (ether OMe); 46.71 (Fmoc CH); m/z (FAB) 328 (MH^+), HRMS found 328.11850, $C_{18}H_{18}NO_5$ requires 328.11849

therefore <1ppm.

Coupling of N^{α}-9-Fluorenylmethoxycarbonyl-α-methoxyglycine onto 4-Alkoxybenzylalcohol (Wang) Resin (**86**)

N,N'-Diisopropylcarbodiimide (250µl, 0.8mM) was added to a solution of N^{α}-9-Fluorenylmethoxycarbonyl-α-methoxyglycine (**85**) (0.52g, 1.6mM) and HOBt (0.25g) in DMF (10ml). After sonicating for 15 minutes at room temperature this solution was added to a slurry of 4-alkoxybenzylalcohol functionalised resin (0.5g, 0.8mM/g, 0.4mM), which had been pre-swollen in the minimum amount of DMF together with a catalytic amount of DMAP (~1mg). This mixture was then sonicated overnight at room temperature before being filtered, washed sequentially with DMF, 1,4-dioxan, dichloromethane and dried to give 0.5778g of functionalised resin. The resin functionality was determined by UV to be 0.427mM/g (61% loading).

Trp-Met-Asp-Phe-α-(OH)Gly (**87**)

The synthesis was carried out on a 0.486mM scale using the Fmoc-α-(MeO)Gly functionalised 4-alkoxybenzylalcohol resin (**86**). All side chains were protected as described previously and all residues were coupled using standard double coupling cycles. At the completion of the assembly phase, the resin bound peptide was sequentially washed with DMF, 1,4-dioxan, dichloromethane and dried to give 1.0759g of yellow solid. The peptide was then cleaved from the resin by treatment with TFA/water/EDT (9.5:0.5:2) for 30 minutes at room temperature. The resin was then removed by filtration and washed with TFA. Evaporation of the filtrate and trituration with diethyl ether yielded the crude peptide which was isolated by filtration, before being dissolved in aqueous acetic acid (20%) and lyophilised to yield 80mg white solid. The peptide was purified by semi-preparative Hplc (Vydac C18, 250 x 22mm, 10µM, A=H_2O, B=CH_3CN, 0.1% TFA; 10ml/min. 10-30% B over 28 min. λ=214nm). The main peak eluting at 26%B was isolated and

lyophilised to give 8mg (2.5%) of the diastereomeric *title compound.* Analytical Hplc (Vydac C18, 250 x 4.6mm, 5μM, A=H_2O, B=CH_3CN, 0.1% TFA; 1ml/min. isocratic elution 18%B. λ=214nm) gave two peaks A and B (A Rt=16.2 min., B Rt=17.6 min), with the ratio of A:B = 2:1. Both these peaks were isolated seperately using the above analytical conditions, and were shown to be identical by mass spectrometry and amino acid analysis; amino acid analysis: Asp_1 0.97, Met_1 0.82, Phe_1 1.03,Trp_1 N/A; m/z (FAB) 671 (MH^+), HRMS $C_{31}H_{38}N_6O_9S$ requires 670.24208, found 670.24215 therefore < 1ppm.

Coupling of N^{α}-9-Fluorenylmethoxycarbonyl-α-methoxyglycine onto 2-Chlorotrityl chloride Resin

A suspension of N^{α}-9-Fluorenylmethoxycarbonyl-α-methoxyglycine (2.1g, 6.4mM) in dry dichloromethane (40ml) was added to 2-chlorotritylchloride functionalised resin (1g, 1.6mM). Diisopropylethylamine (1.4ml) was added and the mixture was allowed to vortex for 20 minutes before a further aliquot of DIEA (1.4ml) was added and agitation continued for a further 30 minutes. The resin was then capped by adding methanol (2ml) and gently mixing for a further 20 minutes. The rein was then removed by filtration and washed sequentially with DMF, IPA, MeOH and ether before being dried to give 1.41g of functionalised resin. The resin functionality was determined by UV to be 0.754mM/g (63% loading).

Glu-Glu-Ala-Tyr-GlyTrp-Met-Asp-Phe-α-(OH)Gly (**88**)

The synthesis was carried out on a 0.486mM scale using the Fmoc-α-(MeO)Gly functionalised 2-chlorotritylchloride resin. All side chains were protected as described previously and all residues were coupled using standard double coupling cycles. At the completion of the assembly phase, the resin bound peptide was sequentially washed with DMF, 1,4-dioxan, dichloromethane and dried to give 0.69g of solid. The peptide was then cleaved from the resin in two steps. Firstly the

protected peptide was cleaved from the resin by treatment with AcOH/TFE/dichloromethane (2:2:6, 20ml) for 60 minutes at room temperature. The resin was then removed by filtration and the filtrate evaporated to a small volume. The protected peptide was then precipitated with diethyl ether, filtered off and washed with water, then diethyl ether and dried to give 118mg of yellow solid. This yellow solid was immediately mixed with EDT (2ml) before TFA/water (9.5:5, 10ml) was added and allowed to stir for 30 minutes at room temperature. Evaporation of the TFA and trituration with diethyl ether gave a pale yellow solid which was isolated by filtration, washed with diethyl ether and dried. Purification of the crude peptide by Hplc (Vydac C18, 250 x 22mm, 10μM, A=H_2O, B=CH_3CN, 0.1% TFA; 10ml/min. 10-50% B over 28 min. λ=214nm) yielded 13mg (3.75%) of the pure diastereomeric *title compound* as a white solid after lyophilisation; Asp_1 0.97, Glx_2 2.65, Gly_1 1.18, Ala_1 0.98, Met_1 0.87, Tyr_1 0.84, Phe_1 0.99,Trp_1 N/A; m/z (FAB) 1220 (MH^+), HRMS $C_{55}H_{69}N_{11}O_{19}S$ requires 1219.44915, found 1219.44916 therefore < +/-1ppm; Hplc (Vydac C18, 250 x 22mm, 5μM, A=H_2O, B=CH_3CN, 0.1% TFA; 1ml/min. 10-65% B over 30 min. λ=214nm) Rt = 17.8 min., 43.5%B.

Nuclear Magnetic Resonance Data for (**88**)

The diastereomeric isomers of (**88**) proved inseparable by Hplc, although 2D NMR did allow assignment of all the signals present in the specta and indicated the presence of diastereomeric forms I and II in a ratio of 1:1. The full NMR assignment is given in table 6.

Residue	NH	α	β	Other
E1		3.80	1.89	γCH_2 2.32
E2	8.59	4.32	1.64,1.88	γCH_2 2.30
A3	8.12	4.25	1.15	
Y4	(I) 7.94 (II) 7.96	4.38 4.38	2.68,2.85 2.68,2.85	3,5H 6.60; 2,6H 6.99
G5	(I) 8.22 (II) 8.11	3.58,3.72 3.58,3.72		
W6	(I) 8.13 (II) 8.02	4.53 4.55	2.96,3.12 2.93,3.12	2H7.12;4H7.57; 5H6.946H7.04; 7H7.29;N10.74
M7	(I) 8.25 (II) 8.15	4.28 4.30	1.79,1.84 1.75,1.83	$\gamma CH_2$2.36,$\varepsilon CH_3$1.99 $\gamma CH_2$2.35,$\varepsilon CH_3$1.99
D8	8.16	4.46	2.48,2.62	
F9	(I) 8.03 (II) 7.93	4.54 4.52	2.74,3.01 2.84,3.00	2,6H 7.16; 3,5H 7.24 4H 7.20
αOHG10	(I) 8.67 (II) 8.65	4.79 4.75		

Table 6. NMR assignments for diastereomers I and II of (**88**)

δ_H [$(CD_3)_2SO$, 600Mhz]

Asp-Tyr-Met-Gly-Trp-Met-Asp-Phe-α-(OH)Gly (**89**)

The synthesis was carried out on a 0.53mM scale using the Fmoc-α-(MeO)Gly functionalised 2-chlorotritylchloride resin. All side chains were protected as described previously and all residues were coupled using standard double coupling cycles. At the completion of the assembly phase, the resin bound peptide was sequentially washed with DMF, 1,4-dioxan, dichloromethane and dried to give 0.69g of solid. The peptide was then cleaved from the resin by treatment with

TFA/EDT/water (9:2:1, 12ml) for 60 minutes at room temperature before the resin was removed by filtration. Evaporation of the TFA and trituration with diethyl ether gave a pale yellow solid which was isolated by filtration, washed with diethyl ether and dried. Purification of the crude peptide by Hplc (Vydac C18, 250 x 22mm, 10μM, A=H_2O, B=CH_3CN, 0.1% TFA; 9ml/min. 10-70% B over 30 min. λ=230nm) yielded 36mg (6%) of the pure diastereomeric *title compound* as a white solid after lyophilisation; Asp_2 1.92, Gly_1 1.28, Met_2 1.67, Tyr_1 1.06, Phe_1 1.06,Trp_1 N/A; m/z (FAB) 1136 (M^+), HRMS $C_{57}H_{65}N_{10}O_{16}S_2$ requires 1137.40211, found 1137.40207 therefore < +/-1ppm; Hplc (Vydac C18, 250 x 22mm, 5μM, A=H_2O, B=CH_3CN, 0.1% TFA; 1ml/min. 10-90% B over 30 min. λ=214nm) Rt = 17 min., 56%B.

Nuclear Magnetic Resonance Data for (**89**)

The diastereomeric isomers of (**89**) proved inseparable by Hplc, although 2D NMR did allow assignment of all the signals present in the specta and indicated the presence of diastereomeric forms I and II in a ratio of 0.8:1. The full NMR assignment is given in table 7.

Residue	NH	α	β	Other
D1		4.55	2.75,3.03	
Y2	8.54	4.40	2.62,2.93	3,5H 6.51; 2,6H 7.01
M3	8.25	4.28	1.75,1.87	$\gamma CH_2$2.40,$\varepsilon CH_3$2.00
G4	(I) 8.02	3.63,3.72		
	(II) 7.97	3.65		
W5	(I) 8.15	4.52	2.97,3.10	2H7.12;4H7.57; 5H6.946H7.04; 7H7.29;N10.74
	(II) 8.07	4.52	2.94,3.10	
M6	(I) 8.20	4.28	1.75,1.82	$\gamma CH_2$2.33,$\varepsilon CH_3$2.00
	(II) 8.31	4.21	1.75,1.84	$\gamma CH_2$2.33,$\varepsilon CH_3$2.00
D7	(I) 8.25	4.42	2.45,2.65	
	(II) 8.15	4.50	2.45,2.65	
F8	(I) 7.92	4.55	2.87,3.00	2,6H 7.13; 3,5H 7.22
	(II) 7.87	4.32	2.80,2.98	4H 7.19
αOHG 9	(I) 8.66	4.73		
	(II) 8.70	4.76		

Table 7. NMR assignments for diastereomers I and II of (**89**)

δ_H [$(CD_3)_2SO$, 600Mhz]

REFERENCES

1. R.B. Merrifield, *J. Am. Chem. Soc.*, 1963, **85**, 2149.
2. Anon, *Biosystems Reporter*, September 1992. Applied Biosystems Inc. Foster City, California.
3. R.B. Merrifield,, *Biochemistry*, 1964, **3**, 1385.

4(a). J.P. Tam, in *Macromolecular Sequencing and Synthesis*, ed. D.H. Schlesinger, 1988, Alan R. Liss, New York, pp153-184.

(b) T. Geisler,, H. Beilan, B.J. Bergst and K.M. Otteson, in *Macromolecular Sequencing and Synthesis*, ed. D.H. Schlesinger, 1988, Alan R. Liss, New York, pp.199-218.

5. L.A. Carpino and G.Y. Han, *J. Am. Chem. Soc.*, 1970, **92**, 5748.
6. C.D. Chang and J. Meienhofer, *Int. J. Peptide Protein Res.*, 1978, **11**, 246.
7. E. Atherton, D. Harkiss, C.J. Logan, R.C. Sheppard and B.J. Williams, *J. Chem. Soc. Chem. Commun.*, 1978, 537.
8. E. Atherton, D. Harkiss, and R.C. Sheppard, *J. Chem. Soc. Chem. Commun.*,1978, 539.
9. B.W. Erickson and R.B. Merrifield, in *The Proteins*, eds. H. Neurath and R.L. Hill, 1976, Vol.2, 3rd edn., Academic Press, New York, pp.255-257.
10. G.B. Fields and R.L. Noble, *Int. J. Peptide Protein Res.*,1990, **35**, 161.
11. E. Atherton and R.C. Sheppard, *Solid Phase Synthesis, A Practical Approach,*Oxford University Press, Oxford, 1989.
12. S.S. Wang, *J. Am. Chem. Soc.*,1973, **95**, 1328.
13. M. Mergler, R. Tanner, J. Gosteli and P. Grogg, *Tetrahedron Lett.* ,1988, **29**, 4005.
14. K. Barlos, D. Gatos, S. Kapolos, G. Paphotiu, W. Schafer and Y. Wenquing, *Tetrahedron Lett.* ,1989, **30**, 3947.
15. R. Ramage, C.A. Barron, S. Bielecki and D.W. Thomas,*Tetrahedron Lett.* ,1987, **28**, 4105.

16. R. Ramage, S.L. Irving and C.M. McInnes, *Tetrahedron Lett.*, 1993, 34, 6599.

17. S.S. Wang, *J. Org. Chem,*, 1975, **49**, 1235.

18. P. Goddard, J.S. McMurray, R.C. Sheppard and P. Esmond, *J. Chem. Soc. Chem. Commun.*, 1988, 1025.

19. D.P.N. Satchell, *Chem. Ind.* ,1974, 683.

20. J.H. Jones, *Chem. Ind.* ,1974, 723.

21. D.S. Kemp, in *The Peptides : Analysis, Synthesis and Biology;* eds. E. Gross and J. Meienhofer, Vol.1, Academic Press, New York, pp.315-382.

22. *The Peptides : Analysis, Synthesis and Biology;* eds. E. Gross and J. Meienhofer, Vol.1, Academic Press, New York, pp.315-382.

23. J. Jones, *The Chemical Synthesis of Peptides*, 1991. Clarendon Press, Oxford.

24. J.C. Sheehan and G.P. Hess, *J. Am. Chem. Soc.*, 1955, **77**, 1067.

25. A.J. Bates, I.G. Galpin, A. Hallet, D. Hudson, G.W. Kenner and R. Ramage, *Helv. Chim. Acta*, 1975, **58**, 688.

26. B. Castro, J.-R. Dormoy, B. Dourtoglon, G. Erin, C. Selve and J.-C. Zeigler, *Synthesis*, 1976, 751.

27. J. Coste, D. Le-Nguyen and B. Castro, *Tetrahedron Lett.*, 1990, **31**, 205.

28. L.A. Carpino, *J. Am. Chem. Soc.*, 1993, **115**, 4397.

29. L.A. Carpino, A. El-Faham, C.A. Minor and F. Albericio, *J. Chem. Soc. Chem. Commun.*, 1994, 201.

30. R. Knorr, A. Trzeciak, W. Banworth and D. Gillessen, *Tetrahedron Lett.*, 1989, **30**, 1927.

31. H. Rink and B. Riniker, *Helv. Chim. Acta*, 1974, **57**, 831.

32. M. Bodansky, in *The Peptides : Analysis, Synthesis and Biology*, 1979, eds. E. Gross and J. Meienhofer, Vol.1, Academic Press, New York, pp.105-196.

33. W. Konig and R. Geiger, *Chem. Ber.*, 1970, **103**, 788.

34. L. Kisfaludy and I. Shon, *Synthesis*, 1983, 325.

35. E. Atherton, J.L. Holder, M. Meldal, R.C. Sheppard and R.M. Valerio, *J. Chem. Soc.Perkin Trans. 1*, 1988, 2887.

36. R. Ramage and J. Green, *Tetrahedron Lett.*, 1987, **28**,2287.

37. Anon. *Introduction to Cleavage Techniques*, 1990. Applied Biosystems Inc., Foster City, California.

38. H. Yajima and N. Fujii, *J. Am. Chem. Soc.*, 1981, **103**, 5867.

39. E. Pedroso, A. Grandas, M.A. Saralegui, E. Giralt, C. Granier and J. Van Rietschoten, *Tetrahedron*, 1982, **38**, 1183.

40. H. Beilan, R.L. Noble, C.P. Ashton and D.P. Hadfield, in *Peptides : Chemistry and Biology*, 1988, ed. G.R. Marshall, Escom, Leiden. pp.198-201.

41. H. Yajima and N. Fujii, in *The Peptides : Analysis, Synthesis and Biology*, 1979, eds. E. Gross and J. Meienhofer, Vol.5, Academic Press, New York, pp.65-109.

42. J.P. Tam and R.B. Merrifield, in *The Peptides : Analysis, Synthesis and Biology*, 1987, eds. S. Udenfriend and J. Meienhofer, Vol.9, Academic Press, New York, pp.65-109.

43. H. Yajima, N. Fujii, S. Funakoshi, T. Watanabe, E. Murayama and A. Otaka, *Tetrahedron*, 1988, **3**, 805.

44. R.G. Pearson and J. Songstad, *J. Am. Chem. Soc.*, 1967, **89**, 1827.

45. B. Gutte and R.B. Merrifield, *J. Am. Chem. Soc.*, 1969, **91**, 501.

46. A. Wlodawer, M. Miller, M. Jaskolski, B.K. Sathyanarayana, E. Baldwin, I.T. Weber, L.M. Selk, L. Clawson, J. Schneider and S.B.H. Kent, *Science*, 1989, **245**, 616.

47. R. Ramage, J. Green, T.W. Muir, O.M. Ogunjobi, S. Love and K. Shaw, *Biochem. J.*, 1994, **299**, 151.

48. D. Alexeev, S.M. Bury, M.A. Turner, O.M. Ogunjobi, T.W. Muir, R. Ramage and L. Sawyer, *Biochem. J.*, 1994, **299**, 159.

49. S. Love *personal communication.*

50. M. Schnolzer and S.B.H. Kent, *Science*,1992, **256**, 221.

51. P.E. Dawson, T.W. Muir, I. Clarke-Lewis and S.B.H. Kent, *Science*, 1994, **266**, 776.

52. L. Chen, A.M. MacMillan and G.L. Verdine, *J. Am. Chem. Soc.*,1993, **115**, 5318.

53. R. Adams and R. Burdon, *Molecular Biology of DNA Methylation*, 1985, Springer-Verlag, New York.

54. J. Boyes and A. Bird, *Cell*, 1991, **64**, 1123.

55. F. Antequera, D. MacLeod and A.P. Bird, *Cell*, 1989, **58**, 509.

56. D.N. Cooper and M. Youssoufian, *Hum. Genet.*, 1988, **78**, 151.

57. G.P. Pfeifer, S.D. Steigerwald, R.S. Hansen, S.M. Gartler and A.D. Riggs, *Proc. Natl. Acad. Sci. USA*, 1990, **87**, 8252.

58. J. Lewis and A. Bird, *FEBS Lett.*, 1991, **285**, 155.

59. I. Oberle, F. Rousseau, D. Heitz, C. Kretz, D. Devys, A. Hanauer, J. Boue, M.F. Bertheas and J.L. Mandel, *Science*, 1991, **252**, 1097.

60. D.A. Erlanson, L. Chen and G.L. Verdine, *J. Am. Chem. Soc.*, 1993, **115**, 12583.

61. S. Kilmasauskas, S. Kumar, R.J. Roberts and X. Cheng, *Cell*, 1994, **76**, 357.

62. M.A. Harington, P.A. Jones, M. Imagawa and M. Karin, *Proc. Natl. Acad. Sci. USA.*, 1988, **85**, 2066.

63. R.R. Meehan, J.D. Lewis, S. McKay, E.L. Kleiner and A.P. Bird, *Cell*, 1989, **58**, 499.

64. R. Wang, X.-Y. Zhang, R. Khan, Y. Zhou, L.-H. Huang and M. Ehrlich, *Nucl. Acids Res.*, 1986, **14**,9843.

65. A. Pawcak, M. Bryans and J.-P. Jost, *Nucl. Acids Res.*, 1991, **19**,1029.

66. R.R. Meehan, J.D. Lewis and A.P. Bird, *Nucl. Acids Res.*, 1992, **20**, 5085.

67. X. Nan, R.R. Meehan and A. Bird, *Nucl. Acids Res.*, 1993, **2**, 4886.

68. A. Bird *personal communication.*

69. K.M. Otteson, R.L. Noble, P,D. Hoeprich, K.T. Shaw and R. Ramage, *Applied Biosystems Research News*, June 1993.

70. T.J. Lobl, M.R.Jr. Deibel and A.W. Yem, *Anal. Biochem.*, 1988, **170**, 502.

71. M. Wilchek, *Adv. Exp. Med. Biol.*, 1974, **42**, 15.

72. S. Funakoshi, H. Fukuda and N. Fujii, *Proc. Natl. Acad. Sci. USA.*, 1991, **88**, 6981.

73. R. Ramage and G. Raphy, *Tetrahedron Lett.*, 1992, **33**, 385.

74. A.R. Brown, S.L. Irving and R. Ramage, *Tetrahedron Lett.*, 1993, **34**, 7129.

75. S.L. Irving, PhD Thesis, University of Edinburgh, 1993.

76. Mass spectra run by V.G Analytical Ltd.

77. Sequencing carried out by Welmet Unit, Department of Biochemistry, University of Edinburgh.

78. D. Lane , P. Prentki and M. Chandler, *Microbiological Reviews*, 1992, **56**, 509.

79. X. Nan *personal communication.*

80. R.W. Woody, in *The Peptides : Analysis, Synthesis and Biology*, 1985, eds. S. Udenfriend and J. Meienhofer, Vol.7, Academic Press, New York, pp.15-114.

81. S.W. Provencher and J. Glockner, *Biochemistry*, 1981, **20**, 33.

82. S. Kelly and N. Price *personal communication.*

83. T.W. Muir, PhD Thesis, University of Edinburgh, 1992.

84. K.D. Wilkinson and A.N. Mayer, *Arch. Biochem. Biophys.*, 1986, **250**, 390.

85. K. Wuthrich, *NMR of Proteins and Nucleic Acids*, 1986, John Wiley and sons, USA.

86. P.N. Barlow *personal communication.*

87(a) R. Ramage and F.O. Wahl, *Tetrahedron Lett.*,1993, **34**, 7133.

87(b) Provided by F.O. Wahl.

88. S. Bury *personal communication.*

89. R.L. Adams, J.T. Knowler and D.P. Leader, in *The Biochemistry of the Nucleic Acids*,11th Edition, 1992, Chapman and Hall, London.

90. B. Sain and N.E. Murray, *Mol. Gen. Genet.*, 1980, **180**, 35.

91. N. Spoerel, P. Herrlich and T.A. Bickle, *Nature*(London), 1979, **278**, 30.

92. M. Zabeau, S. Freidman, M. Van Montagu and J. Schell, *Mol. Gen. Genet.*, 1980, **179**, 63.

93. J. Ineichen, J.C.W. Sheperd and T.A. Bickle, *Nucleic Acids Res.*, 1981, **9**, 4639.

94. W.A.M. Loenen and N.E. Murray, *J. Mol. Biol.*, 1986, **190**, 11.

95. D.T.F. Dryden *personal communication.*

96. K. Akaji, K. Fujino, T. Tatsumi and Y. Kiso, *J. Am. Chem. Soc.*, 1993, **115**, 11384.

97. J.P. Tam, C.-R. Wu, W.Liu and J.-W. Zhang, *J. Am. Chem. Soc.*, 1991, **113**, 6657.

98. J.P. Tam, M.A. Sheikh, D.S. Solomon and L. Ossowski, *Proc. Natl. Acad. Sci. USA.*, 1986, **83**,8082.

99. I. Clark-Lewis, R. Aebersold, H. Ziltener, J.W. Schrader, L.E. Hood and S.B.H. Kent, *Science*, 1986, **231**, 134.

100. I. Clark-Lewis,B. Moser, A. Walz, M. Baggiolini, G.C. Scott and R. Aebersold, *Biochemistry*, 1991, **30**, 3128.

101. D.F. Veber, J.D. Mikowski, S.L. Varga, R.G. Denkewalter and R.J. Hirschmann, *J. Am. Chem. Soc.*, 1972, **94**, 5456.

102. D.B. Scanlon, M.A. Eefting, C.J. Lloyd, A.W. Burgess and R.J. Simpson, *J. Chem. Soc. Chem. Commun.*, 1987, 516.

103. G. Barany and R.B. Merrifield, in *The Peptides : Special Methods in Peptide Synthesis*, eds. E. Gross and J. Meienhofer, Vol.2, Academic Press, New York, pp.233-247.

104. J.W.R. Schwabe and A. Klug, *Structural Biology*, 1994, **1**, 345.

105. U. Weber and P. Hartler, *Z. Physiol. Chem.*, 1970, **351**, 1384.

106. U.T. Ruegg and H.G. Gattner, *Z. Physiol. Chem.*, 1975, **356**, 1527.

107. S. Funakoshi, H. Tamamura, N. Fuji, K. Yoshizawa, K. Miyasaka, A. Funakoshi, M. Ohta, Y. Inagaki and L.A. Carpino, *J. Chem. Soc. Chem. Commun.*, 1988, 1588.

108. B. Kamber and W. Rittel, *Helv. Chim. Acta*, 1968, **51**, 2061.

109. N. Fujii, A. Otaka, S. Funakoshi, K. Bessho and H. Yajima, *J. Chem. Soc. Chem. Commun.*, 1987, 183.

110. P. Seiber, B. Kamber, A. Hartmann, A. Johl, B. Riniker and W. Rittel, *Helv. Chim. Acta*, 1970, **60**, 27.

111. N. Fujii, A. Otaka, T. Watanabe, A. Okamachi, H. Tamamura and H. Yajima, *J. Chem. Soc. Chem. Commun.*, 1989, 283.

112. G.L. Ellman, *Arch. Biochem. Biophys.*, 1959, **82**, 70.

113. M.J. Fernandez-Diez, D.T. Osuga and R.E. Feeney, *Arch. Biochem. Biophys.*, 1964, **107**, 449.

114. M. Torchinsky, in *Sulfur in Proteins*, 1981, Pergamon Press, Oxford.

115. J. Fothergill *personal communication.*

116. B.A. Krizek, B.T. Amann, V.J. Kilfoil, D.L. Merkle and J.M. Berg, *J. Am. Chem. Soc.*, 1991, **113**, 4518.

117. G. King, *personal communication.*

118. K. Urquhart, *personal communication.*

119. L. Jiang, *personal communication.*

120. I.J. Galpin, F.E. Hancock, B.K. Handra, A.G. Jackson, G.W. Kenner, P. McDowell, P. Noble and R. Ramage, *Tetrahedron*, 1981, **37**, 3043.

121. R. Tomatis, M. Guarneri, A. Guggi, S. Salvadori and R. Rocchi, *Int. J. Pept. Protein Res.*, 1979, **14**, 347.

122. J.D. Glass, in *The Peptides : Analysis, Synthesis and Biology*, 1987, eds. S. Udenfriend and J. Meienhofer, Vol.9, Academic Press, New York, pp.167-184.

123. A. Reidel and H. Waldmann, *J. Prakt. Chem.*, 1993, **335**, 109.

124. H. Waldmann and D. Sebastian, *Chem. Rev.*, 1994, **94**, 911.

125. M. Van Der Mey and E. De Vroom, *Bioorganic and Medicinal Chem. Lett.*, 1994, **4**, 345.

126. M.A. Dineva, B. Galunsky, V. Kasche and D.D. Petkov, *Bioorganic and Medicinal Chem. Lett.*, 1993, **3**, 2781.

127. F. Brtnik, T. Barth and K. Jost, *Coll. Czech. Chem. Comm.*, 1981, **46**, 1983.

128. G. Greiner and P. Hermann, in *Peptides 1990*, eds. E. Giralt and D. Andreu, Escom, Amsterdam, 1991.

129. M. Royo, J. Alsina, E. Giralt, U. Slomczynska and F. Albericio, in *Proceedings of the 13th American Peptide Symposium*, eds. R.S. Hodges and J.A. Smith, 1993,p.116.

130. T.D. Pallin, PhD Thesis, University of Edinburgh, 1993.

131. Y. Kiso, S. Nakamura, K. Ito, K. Ukawa, K. Kitagawa, T. Akita and H. Moritoki, *J. Chem. Soc. Chem. Commun.*, 1979, 971.

132 Q.-C. Wang, J. Fei, D.-F. Cui, S.-G. Zhu and L.-G. Xu, *Biopolymers*, 1986, **25**, S109.

133. H.J. Duggleby, S.P. Tolley, C.P. Hill, E.J. Dodson, G. Dodson and P.C.E. Moody, *Nature*(London), 1995, **373**, 264.

134. A. Kelly, *personal communication.*

135. S.L. Irving, *personal communication.*

136. R. Jaenicke and R. Rudolph, in *Protein Structure - A Practical Approach*, 1990, ed. T.E. Creighton, IRL Press, Oxford, pp.191-223.

137. S.N. McCurdy, *Peptide Research*, 1989, **2**, 147.

138. T.J. Schall, *Cytokine*, 1991, **3**, 165.

139. E.A. Robinson, T. Yoshimura, E.J. Leonard, S. Tanaka, P.R. Griffin, J. Shabanowitz, D.F. Hunt and E. Appella, *Proc. Natl. Acad. Sci. USA.*, 1989, **86**, 1850.

140. E.J. Leonard and T. Yoshimura, *Immunol. Today*, 1990, **11**, 97.

141. E.T. Baldwin, I.T. Weber, R. St.Charles, J.-C. Xuan, E. Appella, M. Yamada, K. Matsushima, B.F.P. Edwards, G.M. Clore, A.M. Gronenborn and A. Wlodawer, *Proc. Natl. Acad. Sci. USA.*,1991, **88**, 502.

142. G.M. Clore, E. Appella, M. Yamada, K. Matsushima and A.M. Gronenborn, *J. Biol. Chem.*, 1989, **264**, 18907.

143. A.M. Gronenborn and G.M. Clore, *Protein Eng.*, 1991, **4**, 263.

144. B.J. Rollins, T. Yoshimura, E.J. Leonard and J.S. Pober, *Am. J. Pathol.*, 1990, **136**, 1229.

145. P.J. Lodi, D.S. Garrett, J.Kuszewski, M.L.-S. Tsang, J.A. Weatherbee, W.J. Leoanard, A.M. Gronenborn and G.M. Clore, *Science*, 1994, **263**, 1762.

146. R. Kelly, *personal communication.*

147. Y.J. Zhang, B.J. Rutledge and B.J. Rollins, *J. Biol. Chem.*, 1994, **269**,15918.

148. B.A. Eipper and R.E. Mains, *Annu. Rev. Physiol.*, 1988, **50**, 333.

149. A.F. Bradbury, M.D.A. Finnie and D.G. Smyth, *Nature* (London), 1982, **298**, 686.

150. D.A. Stoffers, C.B.R. Green and B.A. Eipper, *Proc. Natl. Acac. Sci. USA.*, 1988, **86**, 735.

151. D. Ping, A.G. Katopodis and S.W. May, *J. Am. Chem. Soc.*,1992, **114**, 3998.

152. S.N. Perkins, E.J. Husten and B.A. Eipper, *Biochem. Biophys. Res. Commun.*, 1990, **171**,926.

153. T. Kawahara, K. Suzuki, Y. Iwasaki, H. Shimoi, M. Akita, Y. Moro-Oka and Y. Nishikawa, *J. Chem. Soc. Chem. Commun.*,1992, 625.

154. H. Bundgaard and A. Kahns, *Peptides*, 1991, **12**, 745.

155. S.D. Young and P.P. Tamburini, *J. Am. Chem. Soc.*, 1989, **111**, 1933.

156. K. Suzuki, H. Shimoi, Y. Iwasaki and Y. Nishikawa, *European Patent Application*No. 91810163.5.

157. J. Bongers, A.M. Felix, R.M. Campbell, Y. Lee, D.J. Merkler and E.P. Heimer, *Pept. Res.*, 1992, **5**, 183.

158. U. Zoller and D. Ben-Ishai, *Tetrahedron*, 1975, **31**, 863.

159. M. Kawai, P. Neogi, P.S. Khattri and Y. Butsugan, *Chem. Lett.*, 1990, 577.

160. M. Bodansky, in *Principles of Peptide Synthesis*, 1984, Springer-Verlag, Berlin.

161. K. Barlos, D. Gatos, J. Kallitsis, G. Papaphotiou, Y. Wenquing and W. Schafer, *Tetrahedron Lett.*, 1989, **30**, 3943.

162. K. Barlos, D. Gatos and W. Schafer, *Angew. Chem. Int. Edn. Engl.*, 1991, **30**, 590.

163. D. Ben-Ishai, I. Sataty, N. Peled and R. Goldshare, *Tetrahedron*, 1987, **43**, 439.

164. F. Albericio, A. Grandas, A. Porta, E. Pedroso and E. Giralt, *Synthesis*, 1987, 271.

165. Y. Watanabe, Y. Kamochi and T. Miyaraki, *Heterocycles*, 1981, **16**, 609.

166. L.A. Carpino, E.M.E. Mansuor, C.H. Cheng, J.R. Williams, R. MacDonald, J. Knapczyk and E. Carman, *J. Org. Chem.*, 1983, **48**, 661.

Courses and Conferences Attended

Organic Research Seminars, various speakers, 1991-1994.

The Royal Society of Chemistry Perkin Division, Regional Meetings, various speakers, Heriot-Watt 1991, Edinburgh 1992, Dundee 1994.

The Royal Society of Edinburgh Symposium on Protein Engineering, various speakers, Edinburgh 1992.

3rd International Symposium on Solid Phase Synthesis, various speakers, Oxford 1993.

Medicinal Chemistry, Professor R Baker, Merck Sharp and Dohme, 1992-93

NMR Spectroscopy, Dr I.H. Sadler, Edinburgh University, 1993.

Chemical Development in the Pharmaceutical Industry, SmithKline Beecham, various speakers, University of Edinburgh, 1993.

23rd European Peptide Symposium, various Speakers, Braga, Portugal, 1994.

Departmental German Course, passed 1992.

Tetrahedron Letters, Vol. 34, No. 44, pp. 7129-7132, 1993
Printed in Great Britain

0040-4039/93 $6.00 + .00
Pergamon Press Ltd

Affinity Purification of Synthetic Peptides and Proteins on Porous Graphitised Carbon

Angus R. Brown, Stephen L. Irving and Robert Ramage*

Department of Chemistry, University of Edinburgh, West Mains Road, Edinburgh EH9 3JJ, Scotland

Key words: affinity chromatography; Tbfmoc N$^{\alpha}$-protection; porous graphitised carbon; tetrabenzo[a,c,g,i]fluorenyl-17-methanol; protein purification

Abstract: The affinity of the Tbfmoc group for porous graphitised carbon has been exploited for the purification of synthetic peptides containing up to 85 amino acids. The hydrophobicity of the group has also been used to simplify peptide purification by HPLC.

Over the last three decades great progress has been made in solid phase peptide synthesis. This, combined with the widespread use of HPLC for peptide purification has resulted in the synthesis of peptides containing up to 20-30 amino acid residues becoming almost routine. In contrast, only a few proteins, e.g. HIV-1 protease[1] (99 residues) and ubiquitin[2] (76 residues), have succumbed to chemical synthesis and and crystallisation hence this area still represents a considerable challenge.

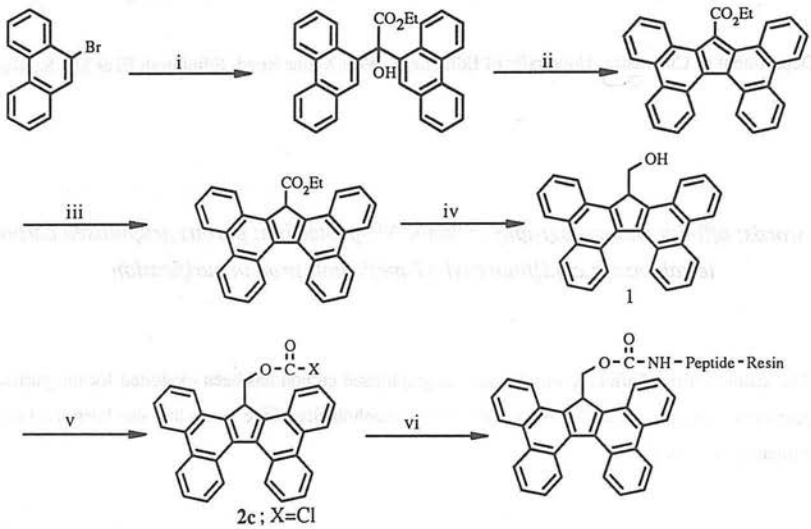
One of the main obstacles to the stepwise chemical synthesis of proteins is the difficulty in purification of the final product, due to the accumulation of truncated peptides on the resin. Truncated peptides are formed when the coupling of an amino acid fails to go to completion and the N$^{\alpha}$-termini of these truncations are routinely capped with acetic anhydride to ensure they play no further part in synthesis. Several methods have been suggested for separating the desired peptide from the acetylated truncated peptides by derivatising the peptide N-terminus with a group capable of an affinity-type binding to a solid support. Examples include the binding of N$^{\alpha}$–biotinylated peptides to immobilised avidin[3], N$^{\alpha}$-dinitrophenyl (DNP) peptides to immobilised DNP antibody[4] and the reaction of an N$^{\alpha}$-thiol group with an iodoacetamide resin[5].

Recently[6], we reported the potential of the base labile N$^{\alpha}$-protecting group, tetrabenzo[*a,c,g,i*]fluorenyl-17-methoxycarbonyl (Tbfmoc) for affinity purification of peptides on porous graphitised carbon (PGC)[7]. We now report an improved synthesis of the Tbfmoc precursor, tetrabenzo[*a,c,g,i*]fluorenyl-17-methanol (**1**) (Scheme 1), without recourse to chromatographic purification of intermediates, and the application of this methodology to the purification of peptides containing up to 85 amino acids

Previously, the Tbfmoc group had been introduced using Tbfmoc-Gly-OH as the N-terminal amino acid in solid phase peptide synthesis. Although a number of N$^{\alpha}$-Tbfmoc amino acids have been prepared, it was thought to be more efficient if a single Tbfmoc reagent could be used to introduce the group directly onto the N-termini of resin-bound peptides. Model studies of the reaction of a resin-bound pentapeptide, Leu.Ile.Phe.Ala.Gly-Resin, with two tetrabenzo[*a,c,g,i*]fluorenyl-17-methyl mixed

carbonates (**2a**; X=O-nitrophenyl, **2b**; X=O-pentafluorophenyl) were carried out. The extent of reaction was monitored by HPLC of the crude cleaved product and showed inefficient incorporation of the Tbfmoc group. However, better results were obtained using chloroformate (**2c**; X=Cl) in CH_2Cl_2 when virtually quantitative incorporation of the Tbfmoc group was observed by sonicating in the presence of diisopropylethylamine. A range of peptides of varying length (Table 1) were then synthesised by the Fmoc/t-Bu method of solid phase synthesis[8]. After reaction of the resin-bound peptides with chloroformate **2c**, the Tbfmoc peptides were cleaved from the resin with TFA + scavengers (peptides **3**, **4** and **6**) or TFA/TMSBr[9] (peptide **5**).

Scheme 1



i. Mg, THF; $(CO_2Et)_2$; ii. TFA/CH_2Cl_2; iii. Et_3N/CH_2Cl_2; iv. Diisobutylaluminium hydride/CH_2Cl_2; v. triphosgene/ N,N-dimethylaniline/CH_2Cl_2; vi. Peptide-Resin/diisopropylethylamine/CH_2Cl_2

Table 1: Peptides purified using Tbfmoc group

	Peptide	No. of Amino Acids
3	Hepatitis B surface antigen PreS1 (1-23) ayw[10]	23
4	Gastrin releasing peptide[11]	27
5	Bacteriophage λ Ral $(Acm)_4$[12]	66
6	MeCP2 methylated DNA binding domain (78-162)[13]	85

Table 2: Tbfmoc **3** deprotection on PGC in 70% CH_3CN

Deprotection Solution	Deprotection Time
1% piperidine	60 min
10% piperidine	10 min
10% triethylamine	>3 h
10% pyrrolidine	10 min

Initial trials were carried out on peptide **3**, when a solution of the crude Tbfmoc peptide in 70% aqueous acetonitrile was applied to a short column of PGC (50-100 μm). HPLC of the eluent showed complete retention of the Tbfmoc peptide, whilst the acetylated truncated peptide impurities eluted. Deprotection of the Tbfmoc group and release of the purified peptide (Figure 1) was then effected by washing the column with 1% piperidine/70% aqueous acetonitrile until no further material eluted. A similar purification could also be achieved by simply adding PGC (either 7 μm or 50-100 μm) to a solution of the crude peptide in 70%

aqueous acetonitrile, whereby the Tbfmoc peptide is adsorbed onto the surface of the PGC and the truncated impurities remained in solution. Adsorption of the Tbfmoc peptide onto PGC was monitored by the uv absorbance of the Tbfmoc group at 364 nm. The supernatant (containing the impurities) was then removed and the purified peptide released from the PGC by suitable base treatment. Using this procedure, the base-deprotection of the adsorbed Tbfmoc peptide was studied more closely. Several deprotection mixtures were tried (Table 2) and the release of the peptide from the carbon was monitored by HPLC. Surprisingly, deprotection of the adsorbed Tbfmoc peptide was much slower than the same reaction in homogeneous solution which is complete within 1 minute in 1% piperidine. This is presumably due to steric hindrance of the acidic 17-H of the Tbfmoc group by the surface of the PGC. Consequently a piperidine concentration of at least 10% (v/v) is recommended for rapid deprotection of Tbfmoc peptides on PGC.

Having established the criteria for efficient deprotection of the Tbfmoc group, purification of larger peptides **4** and **6** on PGC was attempted. In the case of peptide **4**, while a reasonably good purification of the peptide was achieved (Figure 2), variable recoveries of product were obtained. This situation became even worse for peptide **6**, when, although preferential adsorption of the Tbfmoc peptide **6** occurred, subsequent treatment with piperidine/aqueous acetonitrile failed to release the purified peptide from the PGC. These problems were traced to poor solubility of peptides **4** and **6** in piperidine/aqueous acetonitrile mixtures. Alternative solvent systems were tried and some success was achieved using aqueous acetonitrile solutions containing urea or guanidine hydrochloride. However, best results were obtained with 10%piperidine/50% aqueous isopropanol which gave good recoveries of purified products. These results indicate that it is crucial to the success of this method to maintain conditions in which the peptide is soluble and recent experience with other synthetic proteins has shown 1:1 6M guanidine

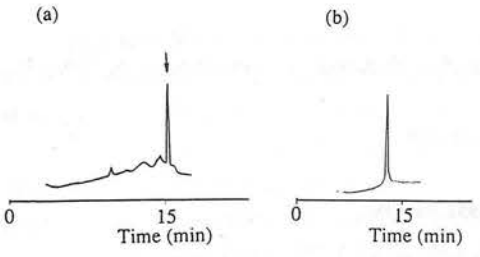


Figure 1: HPLC chromatograms[14] of **3**: (a) crude Tbfmoc peptide; (b) after purification on PGC

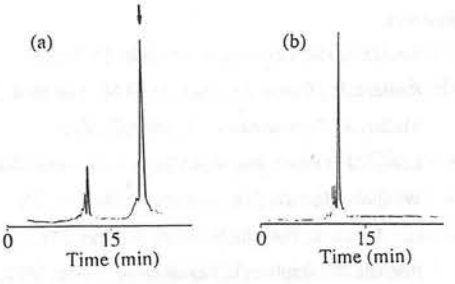


Figure 2: HPLC chromatograms[15] of **4**: (a) crude Tbfmoc peptide; (b) after purification on PGC

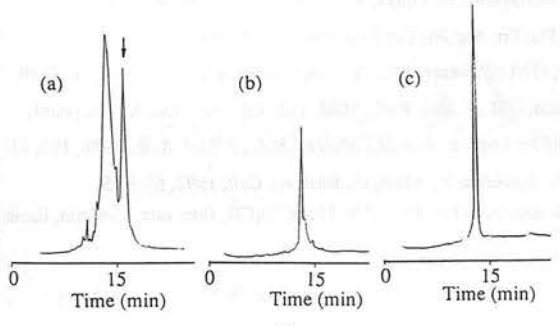


Figure 3: HPLC chromatograms[15] of **6**: (a) crude Tbfmoc peptide; (b) after purification on PGC; (c) after purification by preparative HPLC

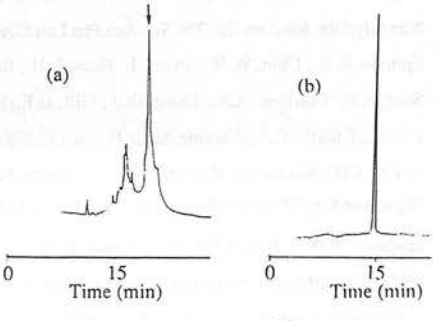
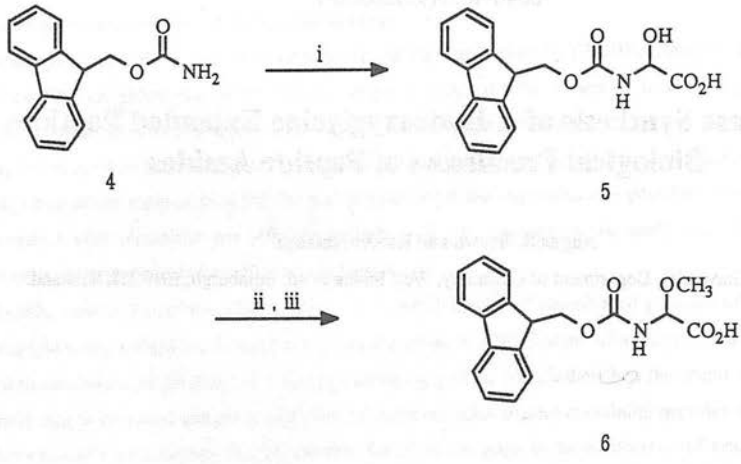


Figure 4: HPLC chromatograms[16] of **5**: (a)crude Tbfmoc peptide; (b) after purification by preparative HPLC



Scheme 2. i, glyoxylic acid; ii, MeOH, sulphuric acid; iii, LiOH

Fmoc carbamate[6] **4** (Scheme 2) was treated with glyoxylic acid (5 equiv.) in gently refluxing ethyl acetate for 6 hours to give racemic 9-fluorenylmethoxycarbonyl-α-hydroxyglycine **5** in 73% yield. The secondary alcohol was protected as the methyl ether **6** by O-methylation of the α-hydroxyglycine derivative[7] **5** followed by ester hydrolysis using LiOH in aqueous acetone to give **6** in 68% overall yield. Activation of **6** (DIC/HOBT/DMAP) was used to functionalise Wang resin[8] which was employed for the solid phase synthesis of the gastrin derivative **7** on an Applied Biosystems 430A peptide synthesiser.

H.Trp.Met.Asp.Phe.α-OH Gly.OH

7

H.Glu.Glu.Ala.Tyr.Gly.Trp.Met.Asp.Phe.α-OH Gly.OH

8

Fmoc deprotection monitoring[9] during synthesis showed low incorporation of Phe due to the instability of the O-methylated α-hydroxyglycine residue[10]. Coupling of Asp to Phe also showed a low incorporation which is characteristic of diketopiperazine (DKP) formation[11]. This resulted in the isolation of only small amounts of the diastereomeric peptide **7** after cleavage (TFA/H_2O/EDT) and RP-HPLC. The diastereomers of **7**, due to the R and S configuration of the terminal α-hydroxyglycine system, were separable by HPLC and were shown to be identical by FAB mass spectrometry and amino acid analysis. Transformation of the α-OMe substituent to the α-OH during acid treatment probably proceeds via the corresponding imine intermediate.

Synthesis of the decapeptide gastrin analogue **8** was then carried out using 2-chlorotrityl chloride resin[12] in order to prevent DKP formation. Fmoc deprotection monitoring during synthesis of **8** showed no evidence of DKP formation and the target peptide was isolated in 3.8% overall yield (13mg) after cleavage and HPLC purification.

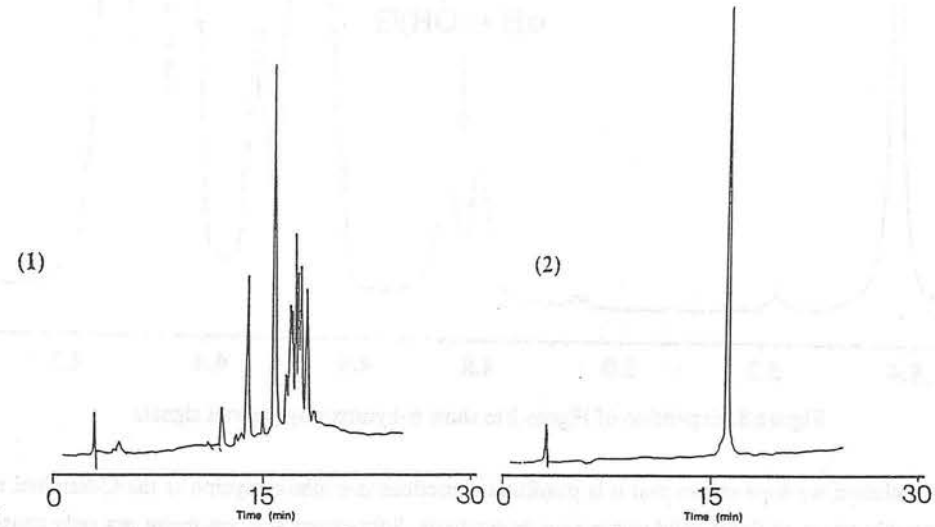


Figure 1. HPLC of crude (1) and pure (2) peptide 8

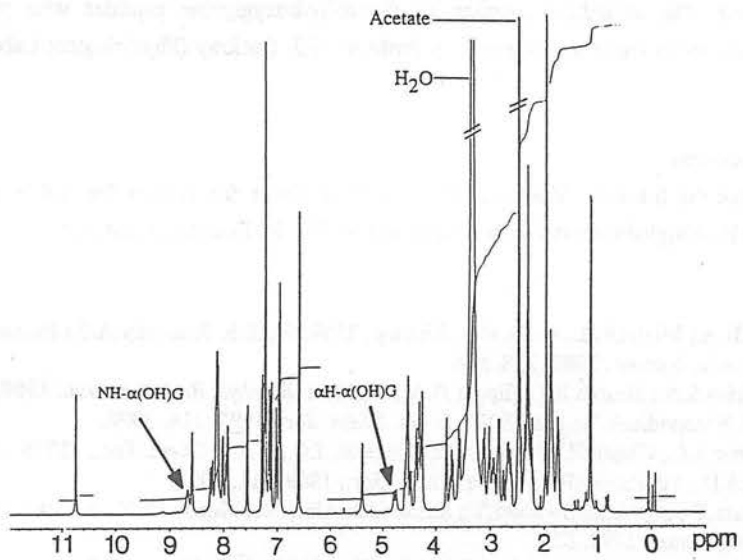


Figure 2. 600Mhz ^{1}H NMR of peptide 8

The diastereomeric decapeptides **8** were inseparable by RP-HPLC (Figure 1) but ^{1}H NMR (600Mhz,D^6 DMSO) showed a 1:1 ratio of diastereomers with the resonances at 4.79 ppm and 4.75 ppm (Figures 2 and 3) assigned to the αH of an α-hydroxyglycine residue. DQF-COSY (Double Quantum Filtered Chemical Shift Correlation Spectroscopy) allowed a complete assignment to be made of the ^{1}H NMR (Figure 2) and showed that the αH signals of the diastereoisomeric α-hydroxyglycine did not have any other crosspeaks except with the NH signal at 8.66 ppm and the resonance of these αH's showed a downfield shift with respect to the other αH's.

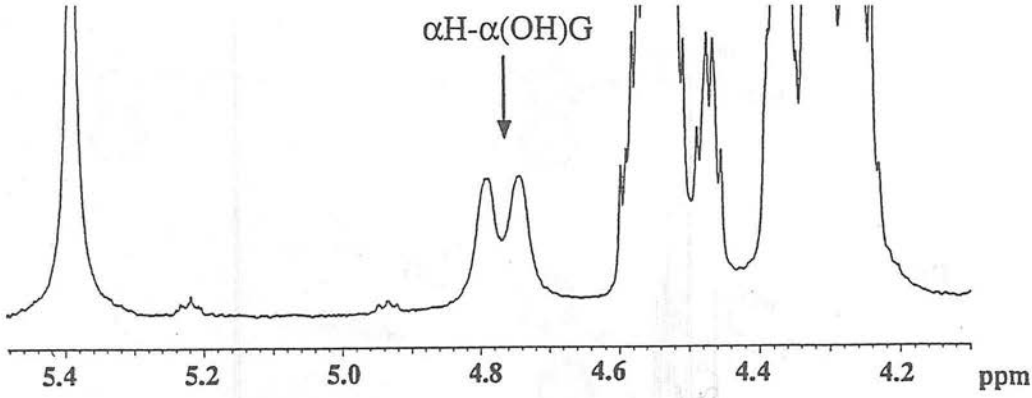


Figure 3. Expansion of Figure 2 to show α-hydroxyglycine αH signals

In conclusion,we have shown that it is possible to introduce α-methoxyglycine as the C-terminal residue in the assembly phase of Fmoc solid phase peptide synthesis. Subsequent acid treatment not only causes side chain deprotection and cleavage of the peptide from the resin but also converts the C-terminal residue to α-hydroxyglycine. The biological function of the α-hydroxyglycine peptides with respect to C-terminal amidation is currently under investigation by Professor G.J. Dockray (Physiological Laboratory, University of Liverpool).

Acknowledgements

We thank the S.E.R.C. Molecular Recognition Initiative for support for A.R.B. In addition we thank K.Shaw, and B. Whigham for technical support and M.R.C. for financial assistance.

References

1. Eipper,B.A.; Mains.R.E., *Annu. Rev. Physiol.*, **1988**, 50, 333. Bradbury,A.F.; Finnie,M.D.A.; Smyth,D.G., *Nature* , **1982**, 298, 686
2. (a) Perkins,S.N.; Husten,E.J.; Eipper,B.A., *Biochem. Biophys. Res. Commun.*, **1990**, 171, 926. (b) Ping,D; Katopodis,A.G.; May,S.W., *J. Am. Chem. Soc.*, **1992**, 114, 3998.
3. (a) Ramer,S.E.; Cheng,H.; Palcic,M.M.; Vederas, J.C., *J. Am. Chem. Soc.*, **1988**, 110, 8526. (b) Young,S.D.; Tamburini,P.P., *J. Am. Chem. Soc.*, **1989**, 111, 1933.
4. Kawahara,T.; Suzuki,K.;Iwasaki,Y.; Shimoi,H.; Akita,M.; Moro-oka,Y.; Nishikawa,Y., *J. Chem. Soc. Chem. Commun.*, **1992**, 625.
5. Bongers,J.; Felix,A.M.; Campbell,R.M.; Lee,Y.; Merkler,D.J.; Heimer,E.P., *Pept. Res.*, **1992**, 5, 183.
6. Carpino,L.A.; Mansuor,E.M.E.; Cheng,C.H.; Williams,J.R.; MacDonald,R.; Knapczyk,J.; Carman,E., *J.Org. Chem.*, **1983**, 48, 661.
7. Zoller,U.; Ben-Ishai,D., *Tetrahedron*, **1975**, 31, 863.
8. Wang,S.S., *J. Am. Chem. Soc.*, **1973**, 95, 1328.
9. Otteson,K.M.; Noble,R.L.; Hoeprich,P.D.jr.; Shaw,K.T.; Ramage,R., *Proceedings of the 13th American Peptide Symposium 1993*, in press. Ramage,R.; Stewart,A.S.J., *J. Chem. Soc. Perkin Trans. 1*, **1993**, 17, 1947.
10. Kawai,M.; Neogi,P.; Khattri,P.S.; Butsugan,Y., *Chem. Lett.*,**1990**, 577.
11. Bodansky,M., *Principles of Peptide Synthesis* ;Springer-Verlag: Berlin, 1984.
12. Barlos,K.; Gatos,D.;Kallitsis,J.;Papaphotiou,G.: Wenquing,Y.;Shafer,W., *Tetrahedron Lett.*, **1989**, 30, 3943.

(*Received in UK* 15 *November* 1993; *accepted* 26 *November* 1993)