Synthesis and Structure of Ubiquitin

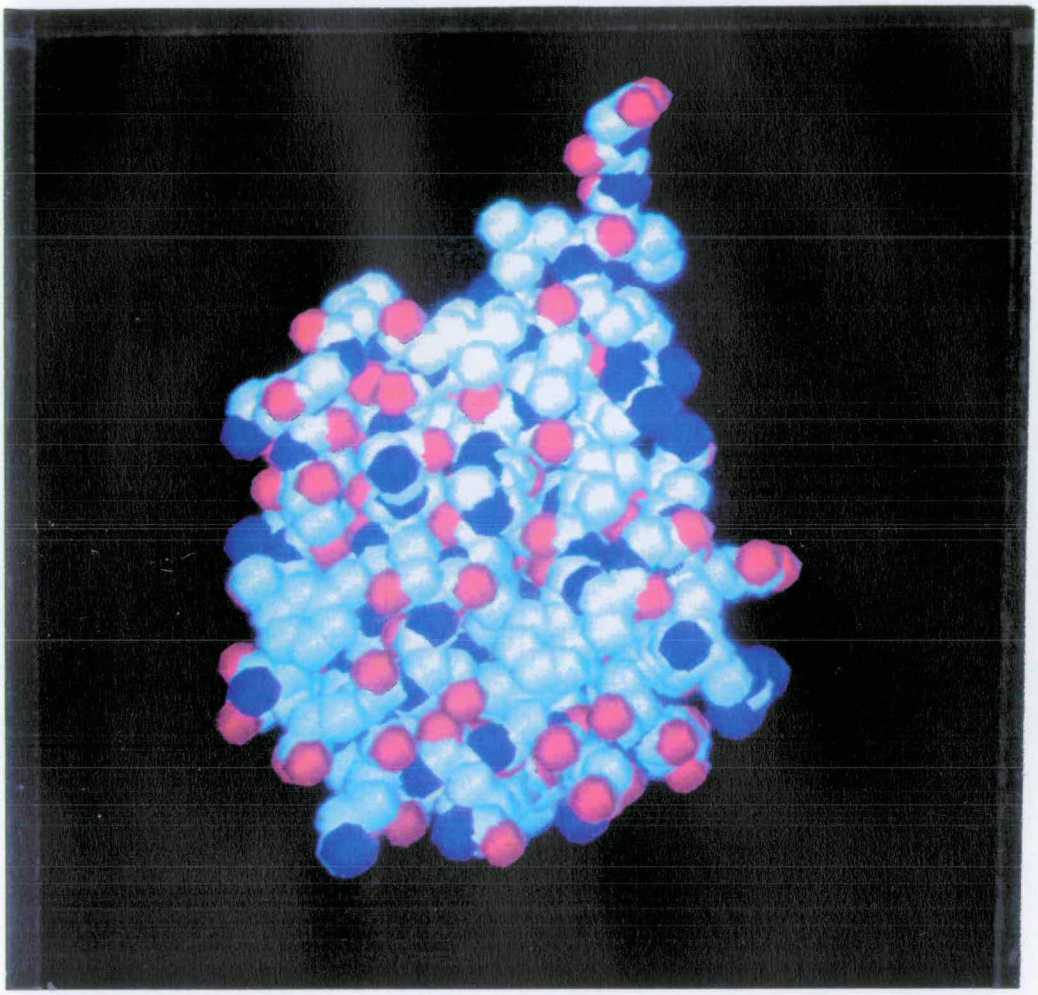
Jill Wilken

A thesis submitted for the degree of

Doctor of Philosophy

University of Edinburgh

October 1995



Spacefill diagram of ubiquitin: produced using Sybil molecular modelling program.

This thesis is submitted in part fulfilment of the requirements of 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
October 1995

To my parents,

James & Muriel Wilken

Acknowledgements

Firstly, sincere thanks to Professor R. Ramage for giving me the opportunity to study at Edinburgh and for constant encouragement and advice throughout the duration of this work.

Many thanks to Kevin Shaw and Brian Whigham for the chemical synthesis of proteins and for MALDI MS/amino acid analyses respectively and of course for helpful practical advice; to Dr A.R. Brown for the synthesis and purification of the ubiquitin fluorinated analogue, [FLeu^{67}]Ub; to colleagues past and present for useful advice.

Many thanks to Dr P. Barlow & Dr J.A. Parkinson for proton NMR analyses and to Dr I.H. Sadler for fluorine NMR analyses; to Helen MacSparron for proton NMR analyses and help in producing molecular diagrams using Insight II program; to all members of the departmental technical staff for their support.

I am grateful to Professor R.J. Mayor and Maureen Mee at the University of Nottingham for the biological assays on ubiquitin analogues; to Professor N.C. Price and Dr S. Kelly at the University of Stirling for circular dichroism analyses; to Professor D.W.Young and Bernard Starkmann at the University of Sussex for the provision of (2S,4S)-5-fluoroleucine.

Finally, thanks to my family and friends for their tremendous support, past and present, in particular thanks to Mark for all his help during my three years in Edinburgh.

Abstract

Ubiquitin (76 amino acids) has been used as a model for protein folding studies with the synthesis of analogues including both single and double substitutions. Chemical synthesis has allowed for the facile incorporation of both unnatural amino acids and a novel synthetic fluorinated amino acid. Purification of synthetic ubiquitin by a variety of protocols has identified important steps in the folding pathway. We have investigated the placement of fluorine within the hydrophobic core of ubiquitin through the synthesis of three double substitutions at positions 43&67, 50&67, 56&67 and one single substitution at position 67 using (2S,4S)-5-fluoroleucine to replace leucine. Methyl transfer across the hydrophobic core has been considered using the double substitution analogue [3-norleucine,43-norvaline]ubiquitin involving a modification in the distribution of alkyl groups. The importance of the chirality of the single histidine on the final strand of β-sheet has been examined in the synthesis of the analogue [68-Dhistidine]ubiquitin.

The contribution of secondary structural features in protein folding has been investigated by studying peptide fragments corresponding to the N and C-terminal halves of ubiquitin and three overlapping peptides on the final strand of β-sheet. The complementation of the N and C-terminal halves to regenerate the native fold has also been studied.

"Knowledge is wonderful but imagination is even better."

Einstein.

Abbreviations

AAA	amino acid analysis
ABI	Applied Biosystems
Ac	acetyl
Acm	acetoamidomethyl
ATP	adenosine triphosphate
Boc	tertiary-butyloxycarbonyl
Bum	butoxymethyl
Bu^t	tertiary-butyl
CD	circular dichroism
CM	carboxymethyl
D	dimensional
DCC	N,N'-dicyclohexylcarbodiimide
DCM	dichloromethane
DIC	N,N'-diisopropylcarbodiimide
DIEA	N,N-disopropylethylamine
DMF	dimethylformamide
DMSO	dimethyl sulphoxide
DTT	dithiothreitol
EDT	1,2-ethanedithiol
FAB MS	fast atom bombardment mass spectrometry
FLeu	(2S,4S)-5-fluoroleucine
Fmoc	9-fluorenylmethoxycarbonyl
FPLC	fast protein liquid chromatography
FTIR	fourier transform infrared
GdmCl	guanidinium chloride
HOAc	acetic acid
HOBt	1-hydroxybenzotriazole
HOCt	ethyl-1-hydroxy-1H-1,2,3-triazole-4-carboxylate
HPLC	high performance liquid chromatography
hrs	hours
MALDI MS	matrix assisted laser desorption ionisation mass spectrometry
Mbh	4,4'-dimethoxybenzhydryl
MG	molten globule
MHz	mega hertz
min	minute
MMA	N-methylmercaptoacetamide
MOPS	4-morpholinepropanesulphonic acid
mp	melting point
MQ	milli-Q
mRNA	messenger RNA
Mtr	4-methoxy-2,3,6-trimethylbenzenesulphonyl
MWCO	molecular weight cut off
N	native

ND	not determined
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser and exchange spectroscopy
PG	protecting group
PGC	porous graphitised carbon
Pmc	2,2,5,7,8-pentamethylchroman-6-sulphonyl
RP	reverse phase
R_t	retention time
RT	room temperature
SPPS	solid phase peptide synthesis
t	time
td	triplet of doublets
Tbfmoc	17-tetrabenzo[*a,c,g,i*]fluorenylmethoxycarbonyl
TEA	triethylamine
TFA	trifluoroacetic acid
TFE	2,2,2-trifluoroethanol
TIS	triisopropylsilane
TLC	thin layer chromatography
TMS	tetramethylsilane
Tris	tris(hydroxymethyl)aminomethane
Trt	trityl
Ub	ubiquitin
U	unfolded
UV	ultraviolet
VIS	visible
WT	wildtype
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
α-Aminobutyric acid	Abu	-
Norleucine	Nle	-
Norvaline	Nva	-
(2S,4S)-5-Fluoroleucine	FLeu	-

Contents

Chapter 1. Introduction

Section 1.1. Folding Studies on Ubiquitin.

1.1.1. How does a protein fold?

It is now over 90 years since Emil Fischer and Franz Hofmeister independently concluded that proteins are composed of α-amino acids covalently linked together.[1] However, although nature has encoded for these twenty simple building blocks it gives no clues as to how they assemble efficiently to give the unique three dimensional structure which is the defining feature of proteins.

The process of folding involves attainment of the native state from the primary sequence of amino acids through hydrophobic interactions and formation of specific secondary and tertiary structures. This is generally assumed to be under thermodynamic control as stated by Anfinsen (1973)[2] and reiterated by Kim & Baldwin (1990) stating "the evidence is good that the final three dimensional structure of a protein is under thermodynamic, not kinetic control."[3]

However, a simple calculation reveals that random search among all possible configurations is a highly ineffective search strategy for formation of the native fold. For example, consideration of a 100 amino acid protein and allowing three configurations for each amide bond with a sampling rate of $10^{13}s^{-1}$ results in a search time of 10^{27} years! Nevertheless, proteins do fold on a millisecond to second timescale and hence the famous Levinthal Paradox.[4]

It has been suggested that the problem of matching speed with stability of folding can be resolved by considering the shape rather than the size of conformational space available. This will be determined by effective search strategies and is succinctly stated by Dill : "searching for a needle in a haystack by random sampling would take forever, but with a big magnet it would be quick."[5] Similarly, it has been suggested

that nature has selected for sequences that have both the ability to fold rapidly and arrive at the thermodynamically stable structures.

Ultimately, the goal of the protein chemist would be the development of a folding code, analogous to the genetic code allowing for accurate prediction of 3D structure from the linear sequence of amino acids. Increased challenges in this area for "cracking the second half of the genetic code" result from the expected completion of the human genome project in 2005.[6] However, this task is ably assisted by advances in the chemical synthesis of proteins which allow for the specific modification of a sequence far beyond the twenty proteinogenic α-amino acids.

We have approached the study of protein folding through the chemical synthesis of both protein analogues and fragments of the protein ubiquitin. In particular we are interested in the modification of the primary sequence with unnatural and novel amino acids and the subsequent changes in the secondary and tertiary structure.

1.1.2. Ubiquitin.

Ubiquitin (Ub) is a small globular protein (76 amino acids, 8565 Da) found in all eukaryotic cells[7] and also recently reported in prokaryotes.[8] It is involved in a myriad of biological functions as reflected in a recent paper titled "Cell Cycle-Ubiquitin with Everything."[9] The functions of ubiquitin derives from its ability to mediate the responses of other proteins by marking or tagging them with either single or multiple ubiquitin molecules. This is most clearly demonstrated in the nonlysosomal intracellular degradation of proteins which involves the formation of multi-ubiquitin chains through isopeptide bonds between an ε amino group of lysine in the target protein and the C-terminal glycine in ubiquitin. Substrates which have been identified include cyclin which is involved in cell regulation,[10] the tumour repressor p53,[11] and the transcription factors myc and fos.[12]

However, ubiquitin conjugation does not always result in proteolysis suggesting other roles for ubiquitin ligation in influencing protein function. These relatively stable ubiquitin conjugates include ubiquitinated forms of the nuclear histones (H2A & H2B) which may be involved in regulating transcription, some cell surface receptors (lymphocyte homing & platelet derived growth factor receptors) are found as ubiquitin conjugates and a family of neurodegenerative diseases (Alzheimer's and Parkinson's) are linked by the presence of insoluble ubiquitin-protein conjugates.[13]

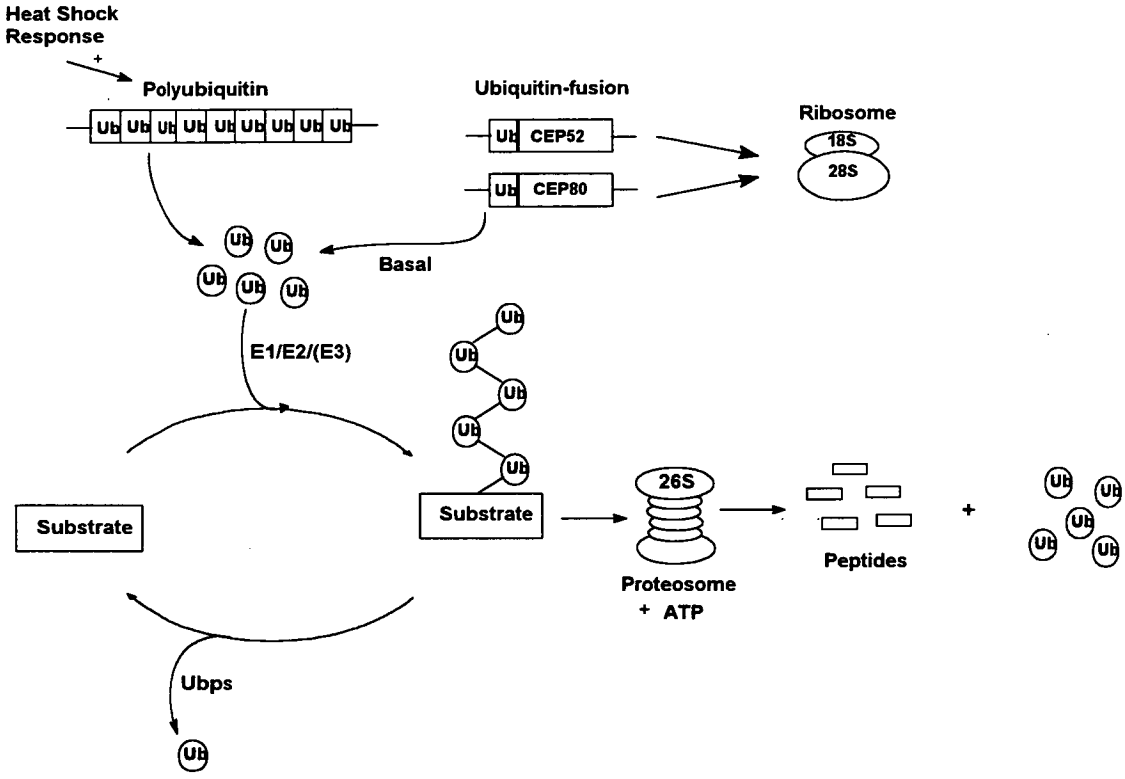


Figure 1.1. Simplified annotation of protein ubiquitination and degradation. Basal levels of ubiquitin are provided by ubiquitin carboxyl extension proteins (UbCEP's) while the polyubiquitin genes are only switched on during cellular stress and are a component of the heat shock response. Multi-ubiquitin chains that form on proteolytic substrates are dynamic structures with ubiquitinating (E1, E2 and E3) and deubiquitinating (Ubps) enzymes rapidly modifying these adducts. Ubiquitinated substrates are degraded by a large protease called the 26S protease complex.

The primary sequence (figure 1.2.) containing no cysteine or tryptophan exhibits unparalleled conservation between species being identical in human, bovine, chicken, *Xenopus* and *Drosophilia* while yeast and oat both have three amino acid changes.[14]

```
          1                                                  38
Human     MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPP
Yeast     ------------------S----D---S----------
Oat       ------------------S----D--------------
          39                                                 76
Human     DQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGG
Yeast     ---------------------------------------
Oat       ------------------A-------------------
```

Figure 1.2. Primary sequence of human, yeast and oat ubiquitin.

This remarkable level of conservation may be important for the many diverse functions of ubiquitin whereby a substitution that is harmless for one function may be detrimental in others. Alternatively the conservation pressure may be associated with some physical property, such as the ability to undergo conformational changes. At present three states of ubiquitin are known to exist at 298K native (N), unfolded (U) which is formed in 6M guanidinium chloride (GdmCl) or in 8M urea and the A-state which is formed in 60% aqueous methanol at pH 2. These all reflect the concept of a flexible molecule which interacts with a multitude of other proteins and co-factors.

1.1.3. Ubiquitin-a model for folding studies.

Ubiquitin is an excellent model for protein folding studies having a well defined secondary and tertiary structure characterised by both x-ray crystallography[15] and solution NMR.[16,17] These are in agreement and reveal an extremely compact globular structure which is stabilised by a pronounced hydrophobic core and a high degree of hydrogen bonding. Only the last 4 amino acids at the C-terminus protrude from the

core giving so called 'stinger or tail' which is critical for the 'tagging' properties of ubiquitin.

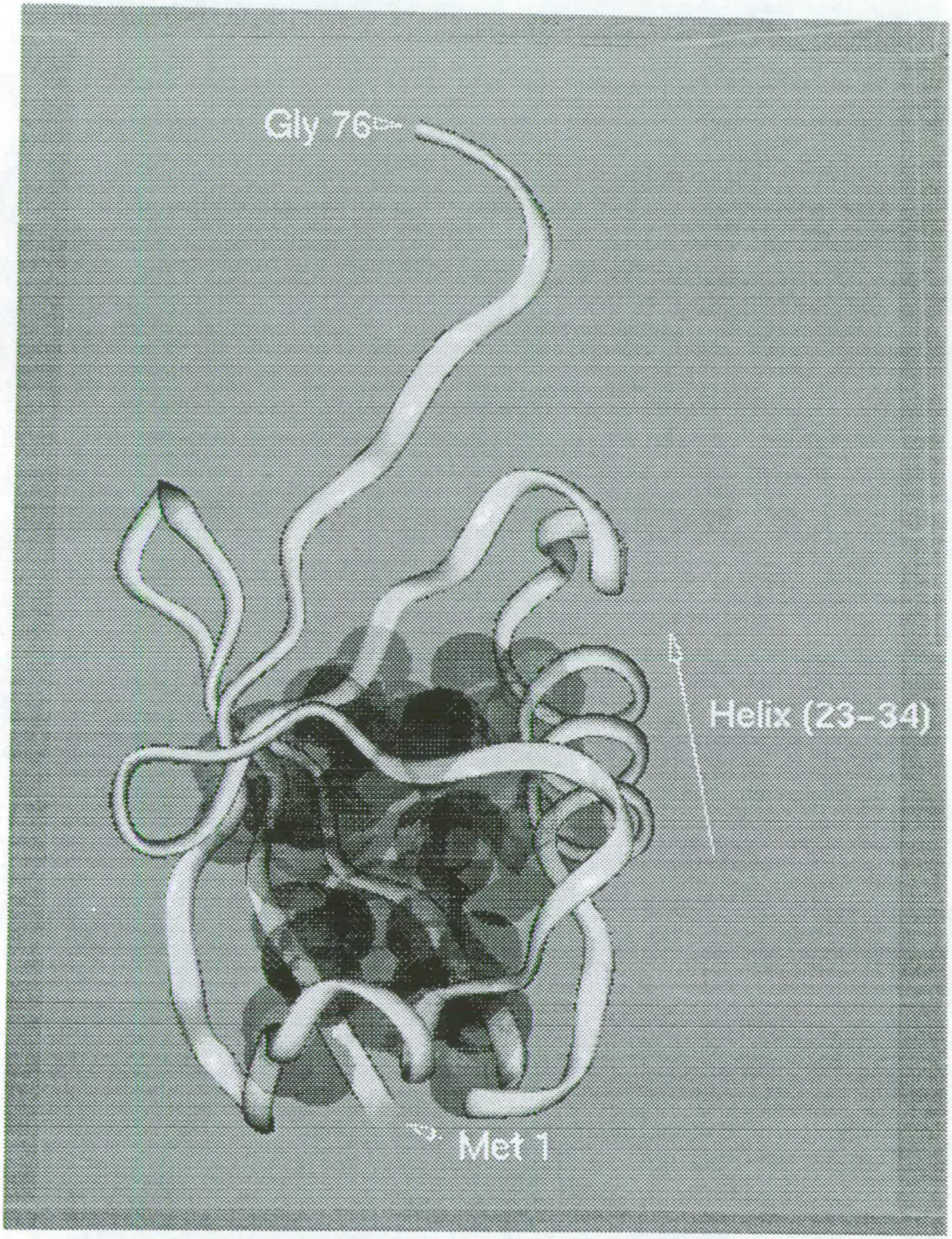


Figure 1.3. Ribbon diagram of ubiquitin with the hydrophobic residues forming the core shaded: produced using Insight II program.

The principle features of the structure include three and half turns of α-helix (residues 23-34), a short piece of 3_{10} helix (residues 56-59) and a five strand mixed β-sheet. Two of the inner strands of the β-sheet are parallel (residues 1-7 and 64-72) while the remaining three strands are antiparallel (residues 10-17, 40-45, and 48-50). Association of the α-helix and β-sheet give the prominent hydrophobic core which is formed by three residues from the α-helix (Ile23, Val26, & Ile30) and eleven of the thirteen hydrophobic residues from the β-sheet (Ile3, Val5, Ile13, Leu15, Val17, Ile36, Leu43, Leu50, Leu56, Ile61, & Leu67), only Ile44 and Val70 are somewhat exposed.

In comparison to the C-terminus the N-terminus is largely inaccessible due to the hydrogen bonding of the methionine sulphur to the backbone amide of Lys63. The S-N distance of 3·6A° is identical to that reported for a similar hydrogen bond in α-chymotrypsin.[18] Other important structural features include the stabilisation of the loop formed by residues 51-59 by a total of five hydrogen bond interactions (figure 1.4.). The first three are typical hydrogen bonds involving the N $\mapsto$ O bond of the reverse turn 51-54 and the two N $\mapsto$ O bonds of the 3_{10} helix. The remaining two are more unusual examples and involve the hydrogen bonds between the side-chains of Asp58 and Tyr59 with the backbone amides of Thr55 and Glu51 respectively.

The tightly packed hydrophobic core and extensive array of hydrogen bonding confers a high degree of stability to ubiquitin being stable in the temperature range 23-80°C and pH range 1·18-8·43.[19,20] The compact state of ubiquitin is also reflected in its resistance to proteolysis with complete digestion by trypsin only observed when the protein is unfolded in 6·5M urea.[21,22]

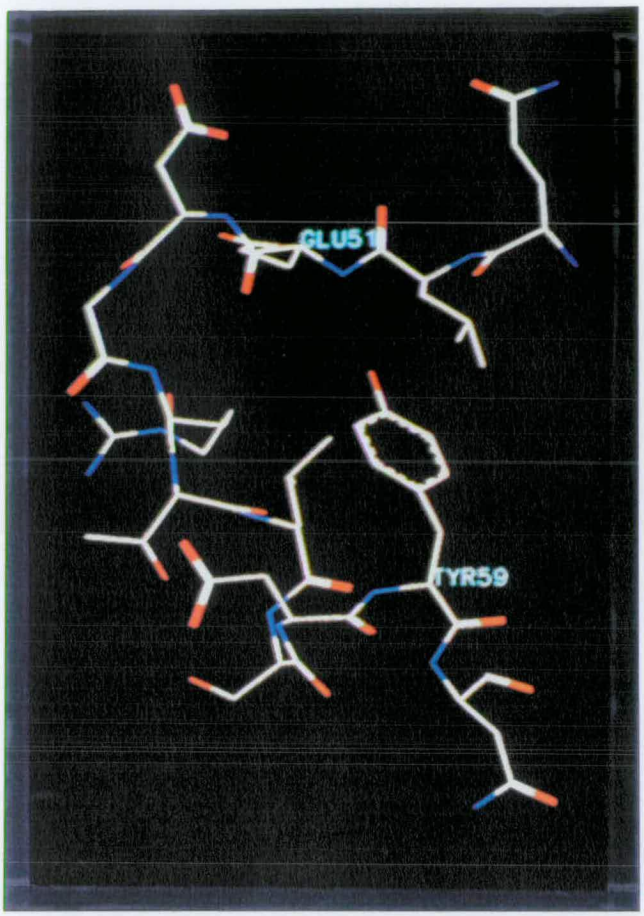


Figure 1.4. Loop containing residues 50-59: produced using Sybil molecular modelling program.

1.1.4. The ubiquitin α/β roll-a protein superfold.

The ubiquitin structure has been classified as one of the nine currently determined protein superfolds called the ubiquitin α/β roll.[23] Other proteins sharing this fold include the B1 domain (56 residues including the NH_2-terminal methionine) of streptococcal protein G[24,25] and the Ras binding domain (81 residues) of the serine/threonine kinase c-Raf1 known as Raf.[26,27] These are unrelated by sequence (<25% similarity) but all share a similar topology (figure 1.5.). This structure similarity in the absence of sequence similarity strongly suggests that the high conservation pressure exerted on the ubiquitin sequence is not simply a result to achieve the correct fold but must involve other factors.

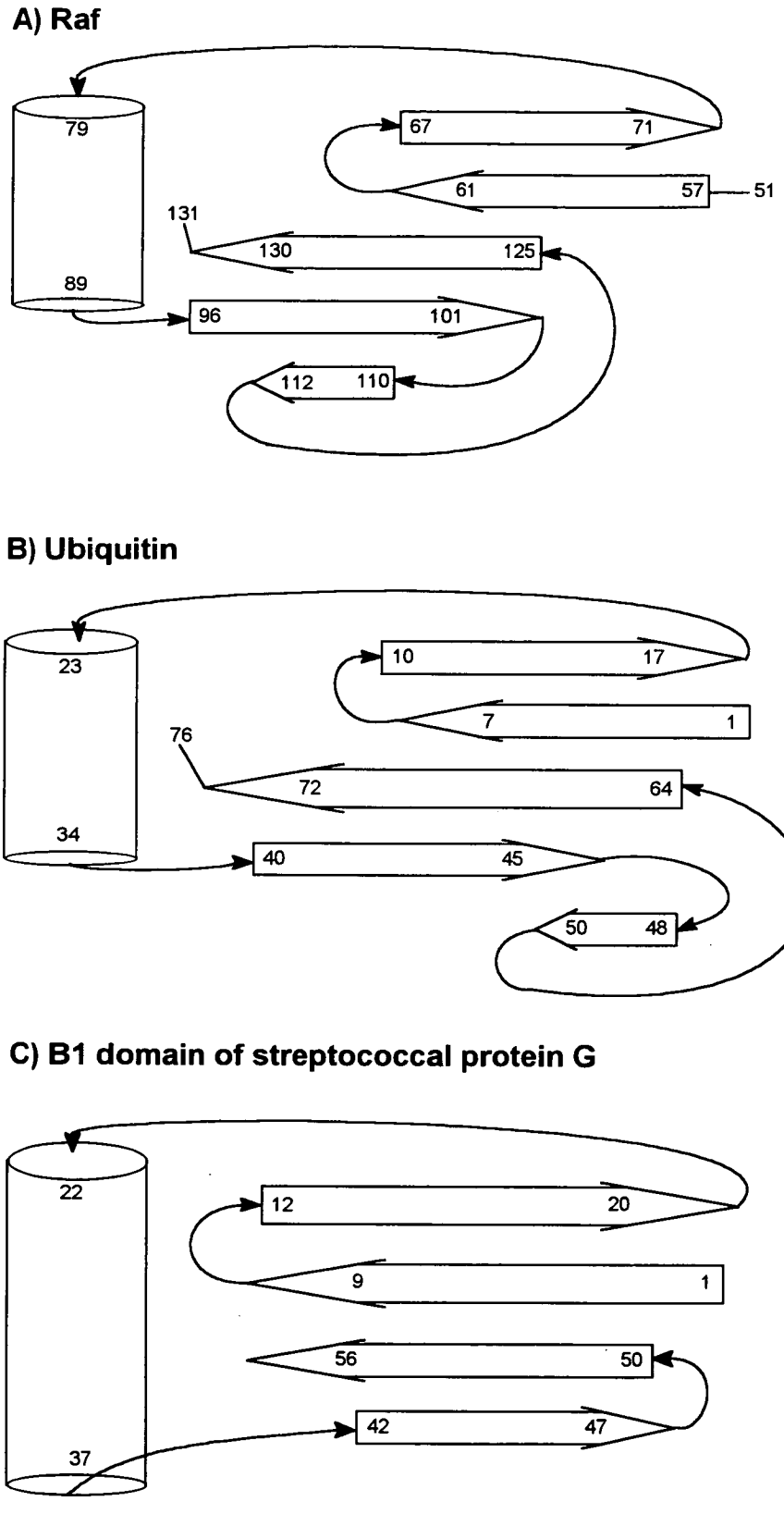


Figure 1.5. Diagram of the overall topologies of A) Raf, B) ubiquitin and C) B1 domain of streptococcal protein G. β-Strands are represented by arrows and α-helices by cylinders. The connecting loops are not drawn to scale.

1.1.5. Ubiquitin-like proteins.

In contrast to ubiquitin whose sequence of amino acids is highly conserved, there is a family of ubiquitin-like proteins that deviate from each other but have sufficient identity to ubiquitin to establish their relationship (figure 1.6.). The interesting question is whether these ubiquitin-like proteins retain some limited properties of ubiquitin that are important for their biological functions.

```
Protein  1                                                    38
   1     M--QIFVKTLTGKTITLEVEPS--DTIENVKAKIQDKEGIPP
   2     MGWDLTVKMLAGNEF--QVSLSSSMSVSELKAQITQKIGVHA
   3       LSILVRNNKGRSSTYEVRLT--QTVAHLKQQVSGLEGVQD
   4     M--QLTVKALQRGECSLQVPED--ELVSTLKQLVSEKLNVPV
   5     L--EVLVKTLDSQTRTFIVGAQ--MNVKEFKAHIRASVSIPS
   6     M--ELFIETLTGTCFELRVSPY--ETVTSVKSKIQRLEGIPV
   7     M--ELFIETLTGTCFELRVSPY--ETVTSVKSKIQRLEGIPV
   8     M--QIFIKTLTGKTITAETEPA--ETVADLKQKIADKEGVPV
   9                                          QDKEGIPP
  10     M--QIFVKTLTGKTITLEVEPS--DTIENVKAKIQDKEGIPP
  11     M--QIFVKTLTGKTITLEVEPS--DTIENVKAKIQDKEGIPP

Protein  39                                                   76
   1     DQQRL--IFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGG
   2     FQQRLAVHPSGVALQDRVPLASQGLGPGSTVLLVVDKCDEP
   3     DLFWL--TFEGKPLEDQLPLGEYGLKPLSTVFMNLRLRGG
   4     RQQRL--LFKGKALADGKRLSDYSIGPNSKLNLVVKPLEK
   5     EKQRL--IYQGRVLQDDKKLQEYNVGGKV-IHLVERAPPQ
   6     AQQHL--IRNNMELEDECSLSGYNISEGCTLKMVLAMRGG
   7     AQQHL--IWNNMELEDECSLSDYNISEGCTLKMVLAMRGG
   8     DQQRL--IFAGKQLEDSKTMADYNIQKESTLHMVLRLRGG
   9     DQQRL--IFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGG
  10     DQQRL--IFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGG
  11     DQQRL--IFAGKQLEDGRSLSDYNIQKESTLHLVLRLRGS
```

Figure 1.6. An alignment of the amino acid sequence from 1. Ub (human) and the ubiquitin-like domains of: 2. UCRP-A (Human), 3. UCRP-B (Human), 4. Gdx (human), 5. BAT3 (human), 6. An1a (*Xenopus laevis*), 7. An1b (*Xenopus laevis*), 8. v-Ubi an *Autographica californica* baculovirus encoded protein (AcMNPV), 9. BVDV CP1 (I), 10. BVDV CP1 (II) and 11. BVDV Osloss. Residues of ubiquitin-like proteins which are identical to ubiquitin (human) are shown in bold.

The first of these is an interferon induced protein known as ubiquitin cross reactive protein (UCRP)[28] which comprises two ubiquitin-like domains (76 and 80 residues) linked head to tail. The C-terminal domain of UCRP contains the -LRLRGG motif and is a functional ubiquitin analogue retaining the ability to conjugate with intracellular proteins.[29] The Gdx gene on the Human X chromosome encodes for a 157 amino acid protein[30] with the N-terminal 76 residues sharing 43% identity with ubiquitin while the C-terminal portion shows no homology. BAT3,[31] a gene from the human major histocompatibility complex class III encodes a single ubiquitin-like domain with 35% identity to ubiquitin. Two An1 mRNA isoforms from *Xenopus laevis* designated An1a (693 amino acids) and An1b (701 amino acids) have N-terminal ubiquitin-like domains (residues 28-104) with 50% sequence homology to ubiquitin.[32]

There are also several examples of ubiquitin-like proteins encoded by viral genes e.g. baculovirus and bovine viral diarrhoea viruses (BVDV). The baculovirus *Autographica californica* nuclear polyhedrosis virus (AcMNPV) encodes a 77 amino acid protein, v-ubi with an extra tyrosine at the C-terminus and shares 76% identity to ubiquitin.[33] Two types of BVDV are observed, CP1 which comprises one truncated and one full size ubiquitin moiety and Osloss which encodes a single ubiquitin domain with 97% similarity to human ubiquitin.[34]

There are important questions concerning the structural and functional aspects of these ubiquitin-like proteins. For example, recombinant expression of UCRP gave a large proportion of insoluble precipitate which is typical for a misfolded protein and would suggest that it lacks the structural stability of ubiquitin.[29]

1.1.6. Characterisation of the ubiquitin folding pathway.

The highly co-operative nature and rapidity of protein folding provides challenges for studying folding through characterisation of folding intermediates between the unfolded and native states. One of the main difficulties is the highly transient

existence of folding intermediates which often exist for only a few milliseconds and therefore cannot be isolated for study. Nevertheless, experiments have been devised to gain structural information on kinetic intermediates using stopped flow hydrogen-deuterium exchange labelling,[35,36], circular dichroism (CD),[37] fluorescence studies,[38] and the binding of the fluorescent probe, 1-anilino-napthalene-8-sulphonate.[39]

Detailed information on the folding of ubiquitin has been obtained using stopped flow techniques including both hydrogen exchange labelling in combination with 2D NMR and fluorescence studies. A schematic representation of the pulsed hydrogen exchange method is shown in figure 1.7.

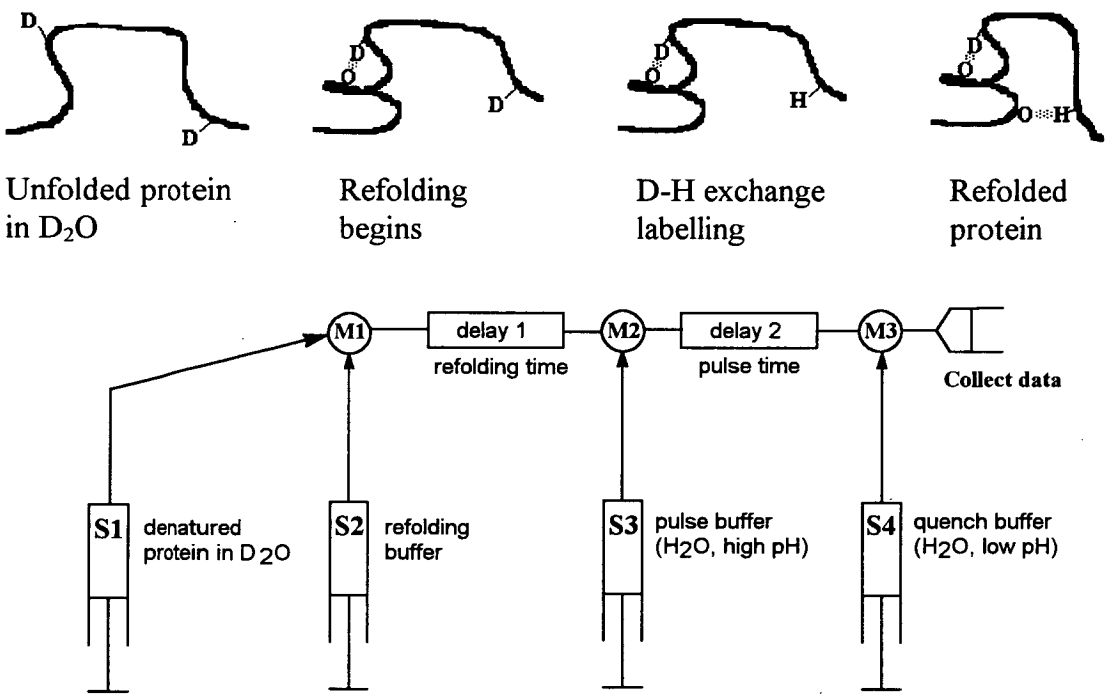


Figure 1.7. Schematic illustration of pulsed hydrogen exchange method. Two representative amide probes are shown in the protein, the first amide proton becomes protected against exchange and remains deuterated, while the second amide still exposed at this time becomes protonated with D to H exchange by an EX2 mechanism occurring during the labelling pulse. A diagram of the NMR quenched flow apparatus with three mixing stages (M1/M2/M3) and two variable time delays is shown. Syringes S1 and S2 are activated at time=0 to initiate refolding by dilution of the denaturant in M1. After a refolding time, the protein is mixed in M2 with H_2O buffer at high pH 9-10 to start the D to H exchange. After a pulse time, exchange is quenched by lowering the pH in M3 under conditions that favour rapid refolding. Adapted from reference.[40]

Using various refolding times, the folding intermediates sandwiched between the unfolded and native states can be mapped from the H/D protection pattern obtained. It is generally assumed that protection from exchange occurs only when native-like structure is formed. The main limitations are that only those peptide amide protons that are stable to exchange in the native protein can be used as probes because only these protons appear in the 2D NMR spectrum in D_2O and for H bonded amide protons the method does not give the location of the acceptor.

1.7. Hydrogen exchange in native ubiquitin.

Briggs & Roder used pulsed hydrogen-deuterium exchange labelling in conjunction with 2D NMR analysis to explore the kinetics of ubiquitin folding.[41] The protection pattern for 26 backbone amide protons that form stable H-bonds upon refolding and exchange only slowly under native conditions were measured using various refolding times from 3·4ms to 20s. A major co-operative folding event was observed with 80% of the amide probes including all of those in the α-helix and most of those in the β-sheet becoming protected within 8ms, followed by two minor phases with time constants (and amplitudes) of 100ms (12%) and 10s (8%). This simultaneous protection of amides from diverse structural elements suggests that both the secondary structure and the hydrophobic core of the protein is formed in a single folding event. Two to threefold slower protection rates were observed for the amide probes Tyr59 and Leu61 which are part of the surface loop and also Leu69 which is on the final strand of β-sheet indicating that these regions are stabilised in a subsequent step after formation of the hydrophobic core. The amide protons of Gln41 and Arg42 also exhibit more pronounced slower protection phases which can be attributed to non-native cis peptide bonds proceeding Pro37 & Pro38. Since the slow folding of residues 37 to 42 does not propagate to other parts of the molecule it is suggested that this region is structurally decoupled from the remainder of the protein. The folding of ubiquitin is extremely rapid compared to cytochrome c,[42] barnase[43] and hen lysozyme[44] also studied by pulsed hydrogen-deuterium exchange.

For example, in cytochrome c, rapid protection is observed for the amide probes in the N and C-terminal helices with 40% protection obtained in 30ms, however it is not until 200ms that 50% protection is obtained for all the amide probes and complete protection requires 10 seconds.

1.1.8. Hydrogen exchange in the ubiquitin mutant [Trp^{45}]Ub.

Further kinetic studies by Roder *et al* on the ubiquitin mutant [Trp^{45}]Ub incorporated tryptophan to replace phenylalanine as a sensitive chromophoric probe to monitor changes occurring specifically within the hydrophobic core.[45] The 2D ^{1}H NMR of [Trp^{45}]Ub is very similar to wild-type (WT) with only limited chemical shift changes observed for residues in the vicinity of Trp45 and, in addition, the biological activity of the mutant is indistinguishable from that of WT. Furthermore, NMR and circular dichroism measurements of the reversible GdmCl induced unfolding transition show that the mutation lowers the stability by less than 0·4kcalmol^{-1} (ΔG=7·1 compared with 7·5kcalmol^{-1}). This similarity means that structural information gained using the mutant can be directly related to WT ubiquitin. Trp45 is sandwiched between the α-helix/β-sheet at the edge of the hydrophobic core and is therefore strategically placed to provide information on unfolding. It is largely buried (97%) and exhibits a four fold increase in fluorescence upon unfolding with 6M GdmCl with the maximum shifted from 336 to 353nm which is expected of a solvent exposed free tryptophan. This compares to the chromophoric Tyr59 already present in WT which is part of the surface loop (residues 50-59) and is largely solvent exposed and consequently exhibits only small changes in fluorescence upon unfolding.[46]

The kinetics of unfolding and folding for [Trp^{45}]Ub were measured at two temperatures 8 & 25°C, over a wide range of GdmCl concentrations (0·5-6M) at pH 5 using fluorescence-detected stopped flow measurements. At both temperatures studied and at concentrations greater than 2M GdmCl the kinetic behaviour follows the expected two state model of unfolding[47,48] without significant population of intermediates between the unfolded and folded conformations, with a major

exponential phase accounting for 90% of the amplitude and a minor slow phase observed. This slow phase can be attributed to non-native proline isomers previously observed in hydrogen exchange with native ubiquitin.

At lower concentrations <2M GdmCl contrasting results were obtained at the two temperatures studied. The results at 8°C and at low concentrations of GdmCl again follow the expected two-state model i.e. at 1M GdmCl a major exponential phase with amplitude (and time constant) 75% (57ms) and two minor slow phases with 18% (0·5s) and 7% (100s) are observed. However, at 25°C and at low denaturant concentrations <2M GdmCl the folding kinetics are inconsistent with a simple two state model i.e. at 1M GdmCl a major burst phase with 50% amplitude occurs in the 2ms deadtime of the stopped flow measurement with three slower kinetic phases with amplitudes (and time constants), 36% (5ms), 11% (75ms) and 3% (20s) observed. This burst-phase is destabilised with increasing GdmCl concentrations and is no longer populated in the unfolding transition (midpoint=3·65M). This burst intermediate can be assigned to an early folding intermediate with a partially developed hydrophobic core involving Trp45. The destabilisation of the burst phase intermediate at low temperature suggests that the collapsed state is primarily stabilised by hydrophobic interactions which are known to be weakened at lower temperatures.[49]

1.1.9. A-state of ubiquitin.

The conformational flexibility of ubiquitin was first observed by Wilkinson & Mayer who detected a non-native structure with 50% helicity as monitored by circular dichroism in alcohol/water mixtures.[50] Harding *et al* further studied a partially structured form of ubiquitin termed the A-state which is formed in 60% aqueous methanol, pH 2 at 298K.[51] The A-state can be regarded as a putative molten globule (MG) and characterisation of this intermediate is fundamental for analysing the folding pathway of ubiquitin since the MG or a similar form may conceivably be on the folding pathway. Formation of the A-state is not observed using pH or temperatures

separately but requires both low pH and 60% aqueous methanol. In lower concentrations of methanol (30-60%) and at 298K equilibrium mixtures of the N and A-states of ubiquitin are observed. The effects of the conditions used to attain the A-state are twofold, the low pH will alter charge-charge interactions in particular salt bridges e.g. the interactions between Lys11/Glu34 and Lys27/Glu48 will be broken while aqueous methanol will diminish the driving force for excluding nonpolar residues compared to water.[52] Heating the A-state in the temperature range 300-335K suggests the protein is approaching a random coil with increased temperature but some ordered structure still persists at 335K. This compares to native ubiquitin at pH 5·8 which retains the native structure in the temperature range 298-335K but precipitates at higher temperatures. Complete assignment of the ^{1}H NMR spectrum of the A-state was not possible due to limited chemical shift dispersion in certain regions. However, some NOE's were able to be assigned in the β-sheet region as shown below (figure 1.8.).

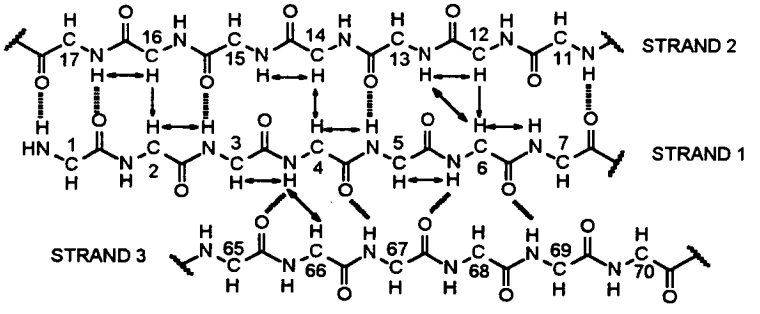


Figure 1.8. NOE data in the β-sheet structure of the A-State.

Additional information on the A-state was obtained from a hydrogen-deuterium exchange refolding experiment which allows identification of amide protons that are protected from exchange and hence postulated to be involved in 2° & 3° structure. This involved dissolving the protein in 60%CD_3OD/40%D_2O pH 2 for 30 minutes which is sufficient time to allow all amide protons not protected from the solvent to exchange with deuterons, and the methanol was then removed by centrifugation and the pH adjusted to 4·7. This resulted in the identification of 14 slowly exchanging amides comprising 11 amides from the first three strands of β-sheet (I3, F4, V5, K6,

T7, I13, L15, V17, L67, L69 & V70) and three amides from the α-helix (V26, A28 & E34). The conclusions drawn from these results are that the first three strands of β-sheet (1-7, 11-17 and 65-70) and a more flexible α-helix (22-34) are present in the A-state while the remaining residues 37-62 assume an essentially random coil.

Complete assignment for more than 90% of the residues in the A-state was achieved using 3D heteronuclear NMR analysis of uniformly ^{15}N enriched ubiquitin.[53] This confirmed the presence of the first two strands of β-sheet and also the α-helix. These elements can be regarded as associating to give a single folded domain which is stable in the absence of the C-terminal half of the molecule. No evidence was found for the presence of the third strand of β-sheet (65-70).

1.1.10. Hydrogen exchange in the native and A-states of ubiquitin.

Pan & Briggs studied hydrogen exchange rates in the putative MG of ubiquitin, the A-state which is formed at low pH in 40-60% aqueous methanol.[54] Protection factors for 41 slowly exchanging amides in the native state were compared to 30 observable amides in the A-state. The 11 protons not seen in the A-state include five amides (Q2, T12, E16, T22 and Q49) which are not hydrogen bonded in the crystal structure but which exchange slowly enough albeit with low protection factors to be detected in the native state. The other six amides which exchange too quickly to be observed in the A-state are mostly located on the periphery : T7 and L15 which are at the end and outer edge respectively of the N-terminal β-sheet, D32 at the C-terminal end of the α-helix, D58 and N60 in the 3_{10} helix and H68 in the central β-strand. The protection factors observed are shown in figure 1.9.

It is apparent that the same pattern of protection occurs in both the native and A-states differing only in their magnitude of protection. The protection factors are much lower in the A-state ranging from 2·6 to 42 compared to the native state which range from 22 to 10^5. The similar pattern of protection and increase in exchange rates for

the A-state suggests that it retains the major secondary structural elements present in the native state but that these are significantly destabilised.

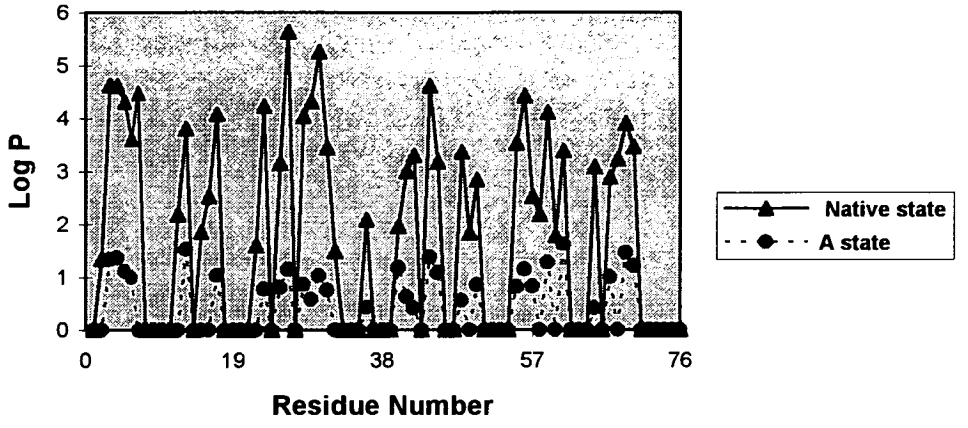


Figure 1.9. Protection factors in the native and A-states of ubiquitin. Protection factors (P) are defined as the ratio of k_c to k_{ex}, where k_c is the intrinsic exchange rate of a given amide in a random coil peptide and k_{ex} is the experimentally determined value. High protection factors are obtained for folded or partially folded structures and low values for unfolded structures.

The most strongly protected amides in both cases are those located in the interior of the hydrophobic core. This methanol induced state is similar to the previously discussed kinetic intermediate containing non-native cis peptide bonds proceeding P37 and P38 in that most of the native secondary structure is present and the association of the α-helix and β-sheet has occurred. The main difference between the two is for the region P37 to R42 which has non-native structure in the kinetic intermediate but native structure in the A-state. This can be explained in terms of the denaturing environment used in the kinetic folding experiment giving an unfolded state which allows cis-trans equilibration of proline residues, while in the A-state the structure is not sufficiently "loosened" to allow trans to cis isomerisation of peptidyl-prolyl bonds. Another difference for the A-state is seen at that the ends of the α-helix and β-sheet which are destabilised in the methanol induced state reflecting a more dynamic flexible structure expected of a MG.

The result obtained by Pan & Briggs differs from that of Harding *et al*[51] who observed many fewer exchanging amides; only 14 (2 of which are not observed by Pan &

Briggs) compared to the 30 reported by Pan & Briggs.[54] This discrepancy can be attributed to the different experimental conditions whereby Pan & Briggs used a series of exchange times ranging from 30s to 2 hours followed by 10 fold dilution with NMR buffer & subsequent ultrafiltration to concentrate the sample while Harding *et al* used only one 30 minute exchange period, following which the methanol was removed by vacuum centrifugation & then diluted with NMR buffer. This latter method effectively results in a much longer exchange period and may also have resulted in structural perturbation.

1.1.11. Protein dissection of the A-state.

The complexity of protein folding can be reduced by studying smaller segments or folded domains of a protein's sequence. Harding *et al* have done this for the A-state of ubiquitin by studying the synthetic peptides Ub(1-21), Ub(1-35) and Ub(62-72) in isolation from the remainder of the molecule to gauge their intrinsic stability's.[55] CD analysis (figure 1.10.) of the peptides Ub(1-21) and Ub(1-35) in 60% aqueous methanol pH 2 indicate maintenance of the β-sheet structure in the absence of wider interactions with the C-terminal half of ubiquitin. In contrast the A-state exhibits a large increase in helicity in 60% aqueous methanol pH 2 which can be attributed to promotion of random helicity in the C-terminal portion.

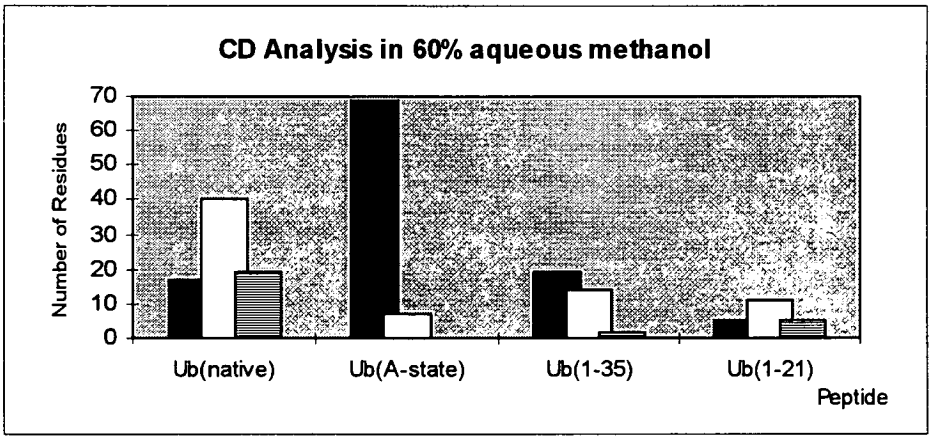


Figure 1.10. Deconvolutions of CD spectra into apparent secondary structure contributions expressed as the number of residues adopting each of the following structures: heavily shaded bars, α-helix; clear bars, β-sheet and patterned bars, remainder.

Comparison of the ^{1}H NMR spectra of Ub(1-21) and Ub(1-35) in 60% aqueous methanol pH 2 with the A-state show that corresponding resonance's have closely related chemical shifts confirming the existence of the N-terminal β-hairpin independently of wider interactions with the C-terminal portion of ubiquitin. Earlier work by Harding *et al*[51] also suggested the existence in the A-state of the β–strand comprising residues 62-72, however monitoring by ^{1}H NMR of 1:1 mixture Ub(1-35) and Ub(62-72) found no evidence to support any significant interaction between these peptides.

This strong tendency of the N-terminal region of ubiquitin to form the correct secondary structure suggests that folding might also occur from the N-terminus in the same way as the synthesis of the protein occurs on the ribosome. However, the kinetics of folding found no evidence to support this with only one major co-operative folding event observed. Folding from the N-terminus has been observed in several multi-domain proteins with folding occurring after completion of each domain while the remainder of the protein is synthesised.[56] The biological activity of ubiquitin is predominantly through conjugation to other proteins via the C-terminus so these results raise interesting questions regarding the structure of these C-terminal conjugates.

1.1.12. Molecular modelling of the A-state-reconciliation of theory and experiment.

Characterisation of the A-state by 2D & 3D NMR supports the idea that the N-terminal half of the molecule contains disrupted but native-like structure while the C-terminal half is largely unstructured and/or contains non-native structure. This is in conflict with the hydrogen exchange data for the A-state which concluded that the similar hydrogen exchange protection pattern combined with lower protection factors reflected a more dynamic and flexible native structure in the A-state. Molecular modelling simulations were used in an attempt to reconcile the interpretation of these different experimental findings.[57] The simulation protocols used all proceed from the well defined crystal structure of ubiquitin and are summarised in figure 1.11.

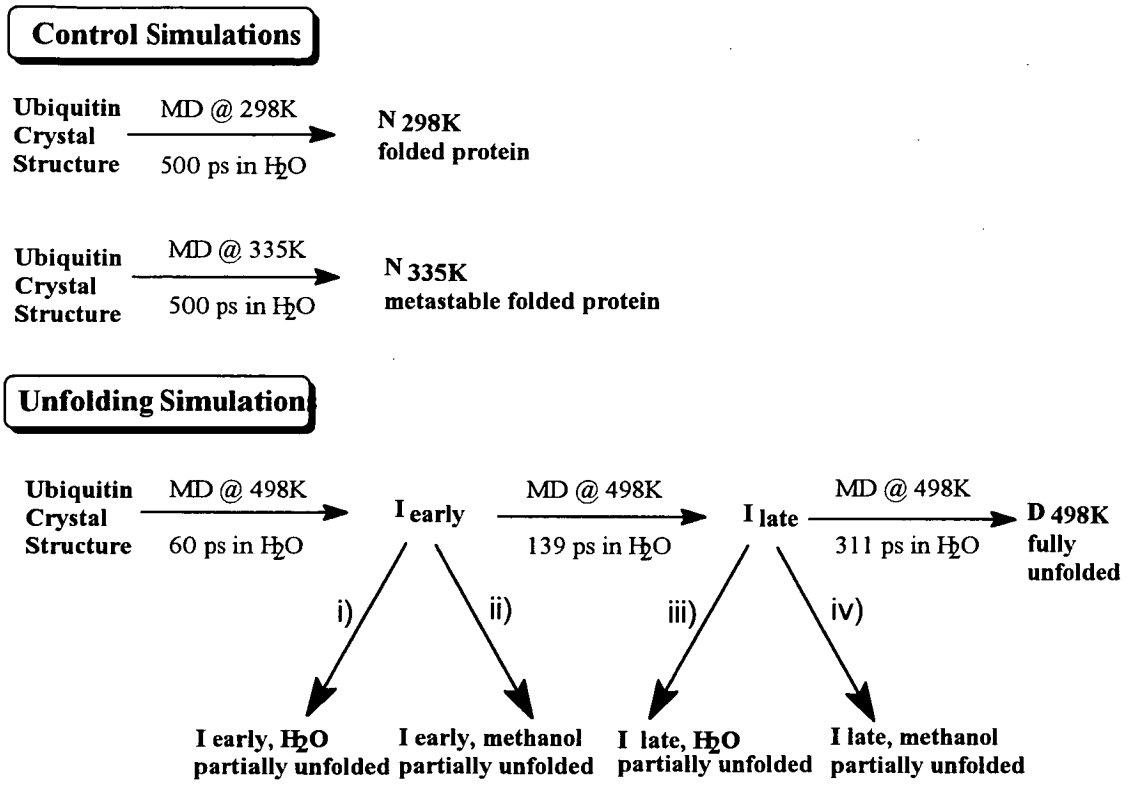


Figure 1.11. Simulation protocols and description of simulations done. Abbreviations used : MD, molecular dynamic simulations : N, native : I, intermediate : D, denatured. Simulation protocols for I(early) i) MD @ 335K, 500 ps in H_2O, ii) MD @ 335K, 500 ps in 60% aqueous methanol and for I(late) iii) MD @ 335K, 500 ps in H_2O and iv) MD @ 335K, 500 ps in 60% aqueous methanol. Adapted from reference.[57]

To allow greater conformational freedom during unfolding native ubiquitin was denatured at high temperature (498K) and in H_2O rather than methanol where reduced protein mobility is known to occur.[58] Two "snapshots" were taken during unfolding giving two intermediates I(early) and I(late) occurring at 60 and 198 ps respectively. These were further simulated in H_2O and 60% aqueous methanol at the lower and more meaningful temperature of 335K where the A-state is known to be stable. Control simulations with native ubiquitin were also performed at 298K and 335K. The tertiary structures obtained for the simulations are shown in figure 1.12.

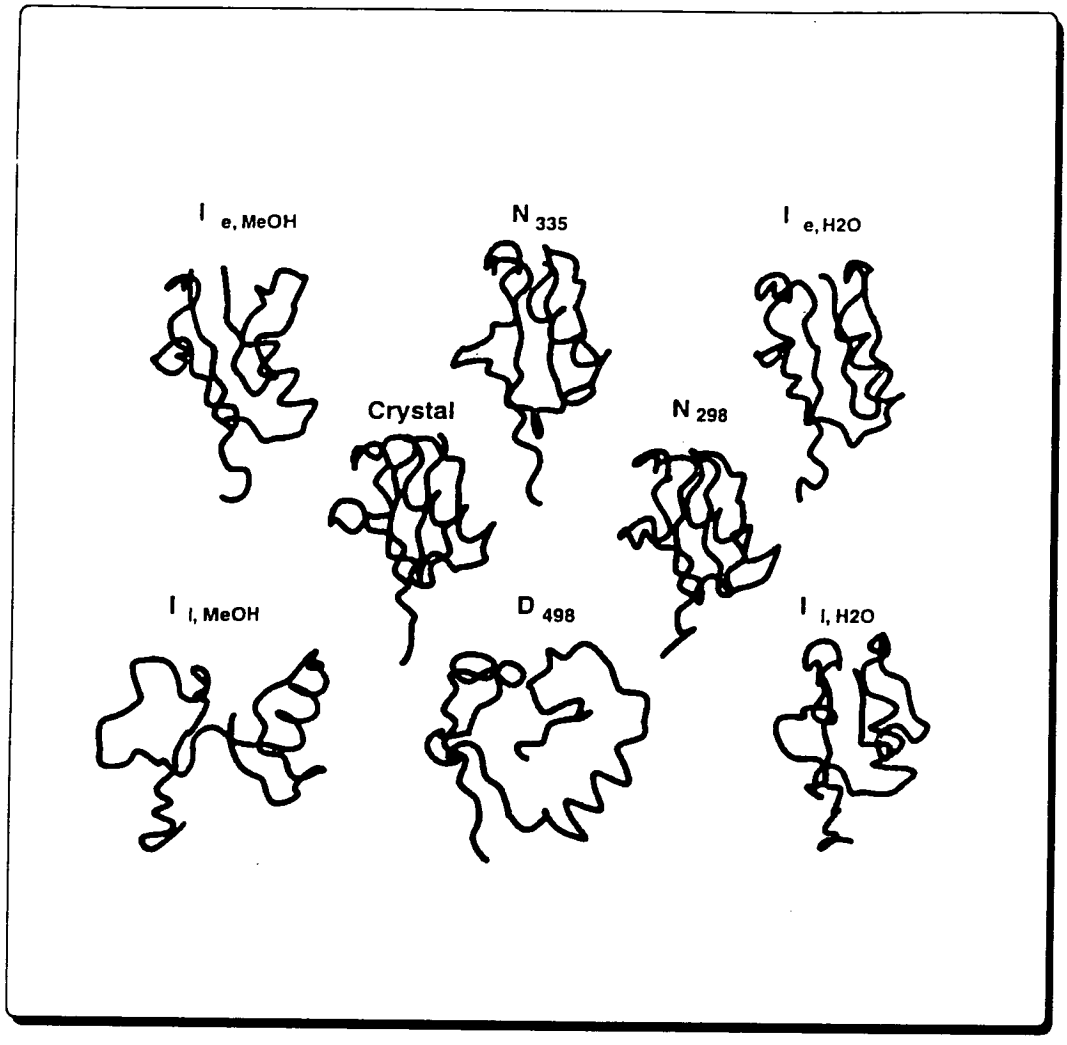


Figure 1.12. Representative structures for each of the 500 ps simulations. Abbreviations used : N, native : I, intermediate : D, denatured : e, early : l, late.

The simulations for N(335K) and the early intermediates, I(early H_2O) and I(early methanol) show only limited changes from the native fold being slightly less compact than N(298K). In contrast the later intermediates I(late H_2O) and I(late methanol) and D(498K) are considerably more disrupted with expanded structures. However, I(late methanol) does retain the β-hairpin comprising the first two strands of β-sheet and the α-helix.

Partial unfolding in these simulations is reflected in both an increase in the radius of gyration of the α-carbons and a concomitant increase in solvent accessible area (figure

1.13.). This is increased at higher temperatures and in organic solvents e.g. 60% aqueous methanol. The increase in solvent accessible area is accompanied with a corresponding exposure of hydrophobic residues from the core which are stabilised by favourable interactions with methanol.[59]

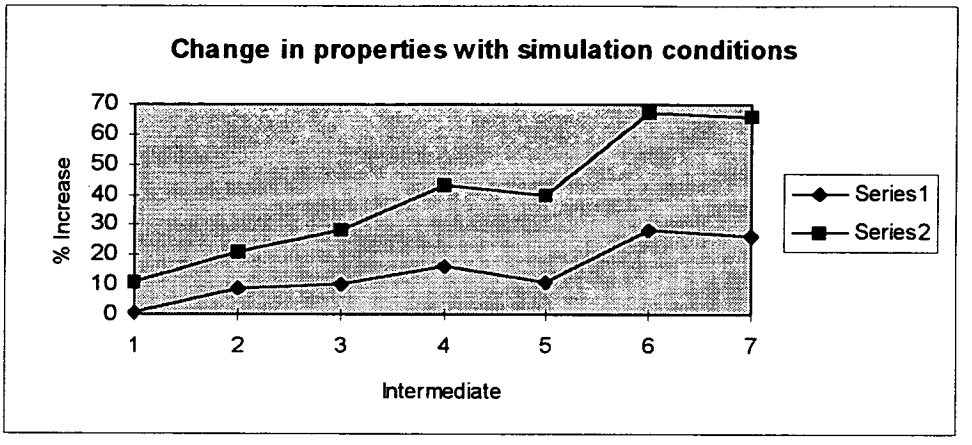


Figure 1.13. Series 1, % increase in the radius of gyration of the α-carbons and series 2, % increase in the solvent accessible surface area compared to the crystal structure. Intermediate 1, N(298K) : 2, N(335K) : 3, I(early H_2O) : 4, I(early methanol) : 5, I(late H_2O) : 6, I(late methanol) : 7, D(498K).

The conformational flexibility of I(early methanol) and I(late methanol) were accessed from two snapshots taken at 210 and 500 ps which shows that the N-terminal portion remains relatively constant in both while the C-terminal portion is very mobile in I(late methanol) but remains fixed in I(early methanol). Furthermore, comparison of the calculated NOE's in the β-strands 1, 2 and 5 shows that I(late methanol) is more similar to the A-state having the expected NOE's between the first two strands of β-sheet whereas I(early methanol) is more structured.

The hydrogen exchange pattern of N(298K), I(early methanol) and I(late methanol) were calculated and compared to the experimental results obtained by Pan & Briggs[54] and Harding *et al*[51] for the both the native and A-state of ubiquitin. For the simulations a 10% cut off was arbitrarily chosen as the maximum percentage of time

that an amide hydrogen can participate in hydrogen bonds with water and still remain protected. The overall results are summarised in table 1.1.

Table 1.1. Protection factors, P denotes protected to exchange with deuterium in : I) native Ub as reported by Pan & Briggs,[54] II) as calculated in molecular simulation of native Ub N(298K), III) A-state of Ub as reported by Pan & Briggs,[54] IV) A-state as reported by Harding *et al*,[51] V) as shown in molecular simulation of the A-state I(early, MeOH) and VI) as shown in molecular simulation of the A-state I(late MeOH).

Residue	I	II	III	IV	V	VI	Residue
M1							M1
Q2	P					P	Q2
I3	P	P	P	P		P	I3
F4	P	P	P	P	P	P	F4
V5	P	P	P	P	P	P	V5
K6	P	P	P	P	P	P	K6
T7	P	P		P	P	P	T7
L8					P		L8
T9					P		T9
G10		P			P		G10
K11		P			P		K11
T12	P					P	T12
I13	P	P	P	P		P	I13
T14							T14
L15	P	P		P			L15
E16	P				P		E16
V17	P	P	P	P	P	P	V17
E18		P					E18
P19							P19
S20							S20
D21		P			P		D21
T22	P						T22
I23	P	P			P		I23
E24		P			P		E24
N25	P	P	P		P	P	N25
V26	P	P	P	P	P	P	V26
K27		P			P	P	K27
A28	P	P	P	P	P	P	A28
K29	P	P	P		P	P	K29
I30	P	P	P		P	P	I30
Q31	P	P	P		P	P	Q30
D32	P	P			P	P	D32
K33		P			P	P	K33
E34		P		P	P	P	E34
G35		P			P	P	G35
I36	P	P			P		I36
P37							P37
P38							P38
D39							D39
Q40	P	P	P				Q40
Q41	P	P	P				Q41
R42	P	P	P		P		R42
L43		P			P		L43
I44	P	P	P		P	P	I44

F45	P	P	P		P		F45
A46		P			P		A46
G47					P		G47
K48	P	P	P		P	P	K48
Q49	P					P	Q49
L50	P	P	P		P	P	L50
E51					P	P	E51
D52							D52
G53							G53
R54					P		R54
T55	P		P				T55
L56	P	P	P		P	P	L56
S57	P	P	P		P	P	S57
D58	P	P			P	P	D58
Y59	P	P	P		P	P	Y59
N60	P	P			P	P	N60
I61	P	P	P			P	I61
Q62							Q62
K63					P		K63
E64		P			P	P	E64
S65	P	P	P		P	P	S65
T66					P		T66
L67	P	P	P	P	P	P	L67
H68	P	P			P		H68
L69	P	P	P	P	P		L69
V70	P	P	P	P	P	P	V70
L71					P	P	L71
R72		P			P		R72
L73					P		L73
R74					P		R74
G75							G75
G76							G76

Good correlation is seen for the native state, N(298K) with the experimental results of Pan and Briggs with a slight overestimation of protection in the N(298K) simulation suggesting that the arbitrarily assigned % cut-off of 10% is too high. The results for I (late methanol) compare favourably with the experimentally obtained values of Pan & Briggs for the A-state while for I(early methanol) an overestimation of protection is seen.

Overall from the data acquired (tertiary structure, NOE data, hydrogen exchange data), I(late methanol) is more similar to the A-state than I(early methanol). The simulation of I(late methanol) and the 2D and 3D ^{1}H NMR of the A-state clearly illustrate the non-native and unstructured C-terminal half of ubiquitin which gives the experimentally observed native-like protection. This has important implications for the interpretation of hydrogen exchange data where it is generally assumed that

protection from exchange signifies native like structure. The paradox can be resolved by consideration of the additional factors that contribute to protection from exchange which include steric shielding of amide hydrogen's in small hydrophobic clusters, preferential interactions with methanol and the formation of intraprotein hydrogen bonds with side or main chain groups. This anomaly has recently been demonstrated in monellin where the β-sheet adopts a non-native a-helical structure in alcohol as determined by 2D NMR but gives native-like protection patterns.[60]

1.1.13. Summary.

The folding of ubiquitin has been shown to be extremely rapid with a partial collapsed intermediate stabilised primarily by hydrophobic interactions formed within 2ms of initiation of folding. This partial condensed state brings residues in close proximity to each other resulting in the formation of secondary and tertiary structure on the 8ms timescale as demonstrated in hydrogen exchange studies on native ubiquitin.

The initial collapsed intermediate will significantly reduce the conformational space available to the protein and thus help in overcoming the Levinthal Paradox. These results strongly suggest a nonhierarchical folding model for ubiquitin in which the main driving force is the hydrophobicity of the amino acid sequence which results in the internal organisation of the protein.

Section 1.2. Chemical Synthesis of Proteins.

1.2.1. General.

The chemical synthesis of proteins requires the unambiguous formation of an amide bond between two α-amino acids to efficiently replicate the translational machinery of the ribosomes (figure 1.14.). This requires protection of the N^{α} amino group of one and the carboxyl group of the other amino acid to prevent unwanted coupling reactions. Furthermore, the unprotected carboxyl group requires to be activated with an electron withdrawing group to ensure nucleophilic attack by the unprotected N^{α} amino group. Additionally, protection is required for functional side-chains present in more than half the naturally occurring α-amino acids.

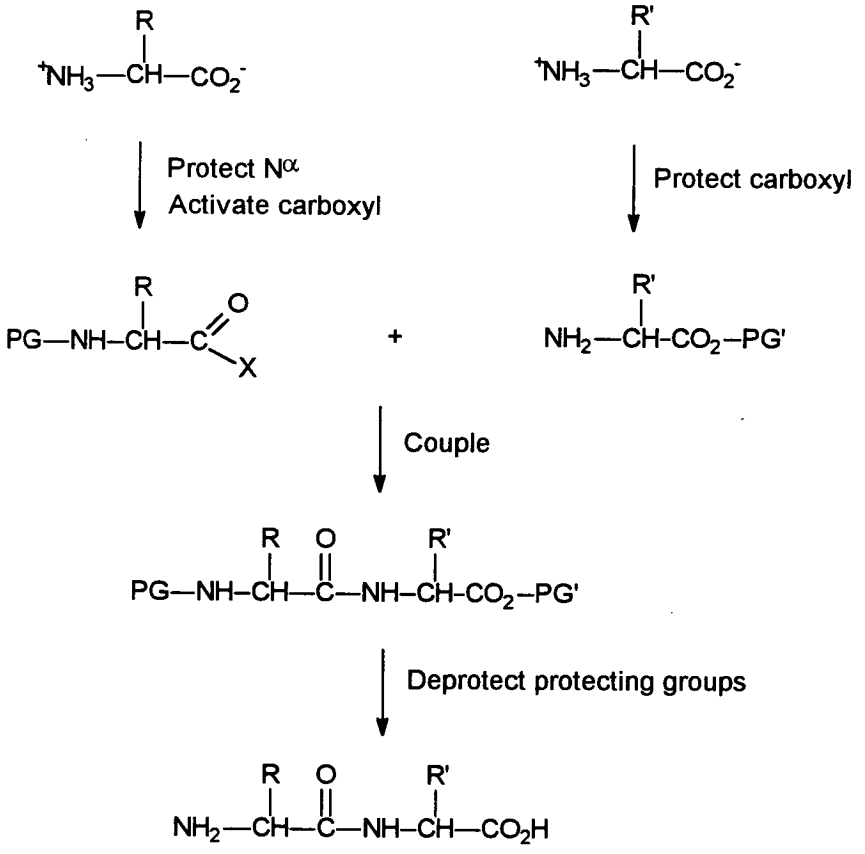


Figure 1.14. Unambiguous formation of a dipeptide from two α-amino acids.

1.2.2. Solid Phase Peptide Synthesis (SPPS).

The development of SPPS began over three decades ago when the method was first proposed by Merrifield[61] and for which he was later awarded the Nobel Prize.[62] The general principle involves the stepwise assembly of peptides from the C to the N-terminus while attached to an insoluble support. The final step involves the simultaneous cleavage of the peptide from the resin and the removal of amino acid side-chain protecting groups.

SPPS offers several advantages over solution synthesis of peptides and these include;

- no need to isolate and purify intermediates
- unreacted reagents and side products are removed by filtration & washing
- large excess of activated amino acid ensures high % couplings
- SPPS can be fully automated

There are two main strategies in SPPS, the first of which is associated with Merrifield and involves N^{α} tertiary-butyloxycarbonyl (Boc) protected amino acids in combination with benzyl based protecting groups (PG's) for functional amino acid side-chains. The Boc group is cleaved by trifluoracetic acid while the more acid stable PG's require cleavage by stronger acids such as hydrogen fluoride or trifluoromethanesulphonic acid. Disadvantages of Boc SPPS include the requirement of specialised apparatus for safe usage of strong acids and the repeated acidolysis to remove the N^{α} Boc group can result in acid catalysed side reactions involving cleavage of acid sensitive bonds and/or cleavage of the peptide resin linkage.

These concerns resulted in the development of a second approach to SPPS using a truly orthogonal strategy and involves N^{α} 9-fluorenylmethoxycarbonyl (Fmoc) protected amino acids[63,64] cleaved with the mild base, piperidine and tertiary-butyl based PG's cleaved with trifluoroacetic acid.

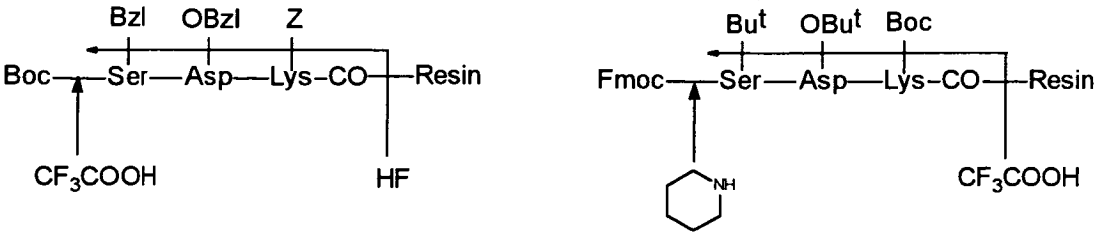
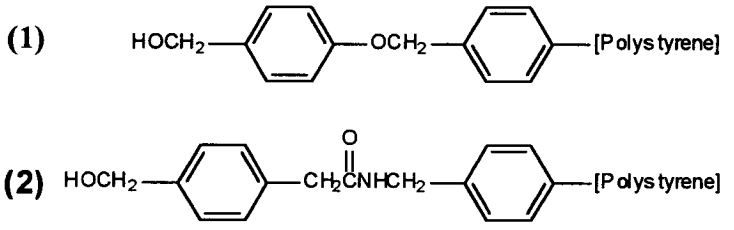


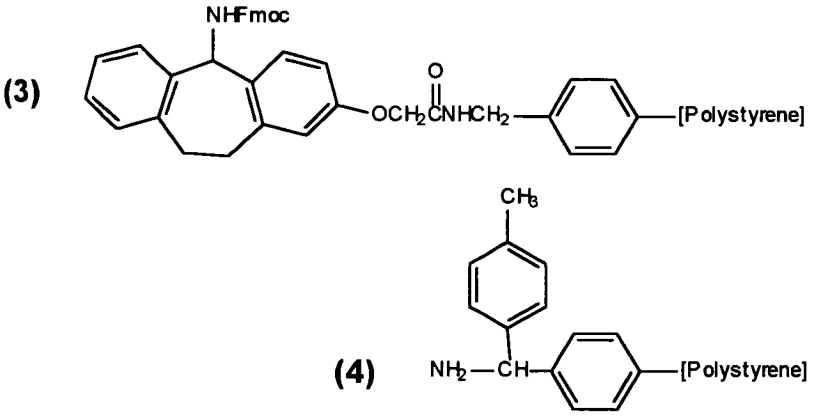
Figure 1.15. Protecting group strategies in Boc and Fmoc SPPS. Z group, benzyloxycarbonyl; Bzl, benzyl; Bu^t, tertiary-butyl; Boc, tertiary-butyloxycarbonyl and Fmoc, 9-fluorenylmethoxycarbonyl.

1.2.3. The solid support.

In FMOC SPPS a more acid labile peptide resin linkage can be used such as the 4-alkoxybenzylalcohol resin (Wang)[65] (1) compared to the phenylacetoamidomethyl resin (PAM)[66] (2) in Boc SPPS.



The above resins give the C-terminal carboxylic acid on cleavage, however other functionality's can be obtained by modifying the linker. For example C-terminal amides are obtained from using the methylaminocarbomethoxy-5-(9'Fmoc)aminodibenocycloheptadiene resin in Fmoc SPPS (3)[67] or 4-methyl benzylamine resin (MBHA) in Boc SPPS(4).[68]



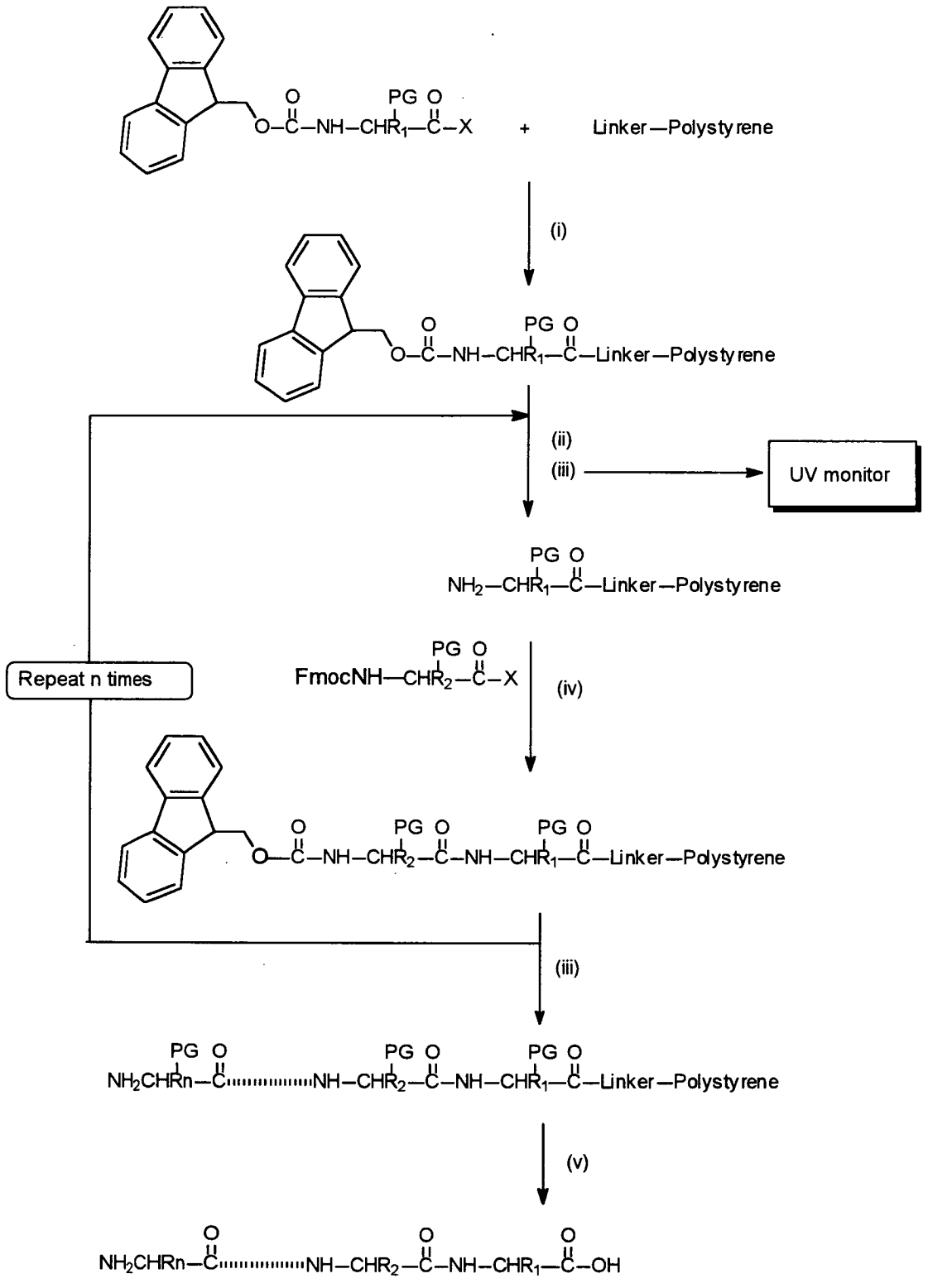


Figure 1.16. General scheme of N^{α} Fmoc SPPS; (i) attachment of first amino acid to the resin, (ii) capping with acetic anhydride, (iii) deprotection of N^{α} Fmoc with piperidine, (iv) coupling of the next amino acid and (v) simultaneous cleavage of PG's and peptide-resin linkage with aqueous trifluoroacetic acid.

1.2.4. Activation procedure and coupling methods.

For efficient coupling in SPPS, the protected carboxylic acid requires to be activated with a suitable electron-withdrawing substituent (X) to enhance the electrophilicity of the carbonyl carbon and thus facilitate nucleophilic attack by the amino group of a second α-amino acid.

The activating group needs to ensure high coupling efficiencies in the absence of unwanted side reactions.

One highly effective activation procedure currently employed in SPPS is symmetrical anhydrides. These are formed by the reaction of two equivalents of protected carboxylic acid with one equivalent of carbodiimide. In practice, di-isopropylcarbodiimide (DIC) is routinely used as the coupling agent, since the diisopropylurea by-product is fully soluble in the solvents used in SPPS in comparison to the insoluble dicyclohexylurea by-product of dicyclohexylcarbodiimide (DCC).

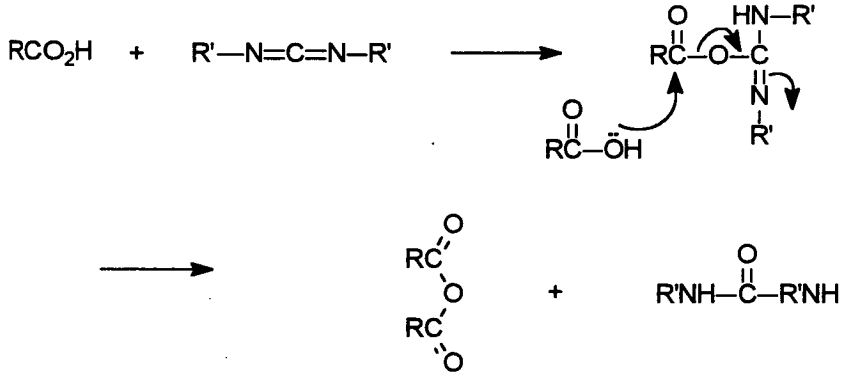
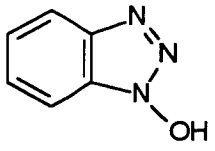


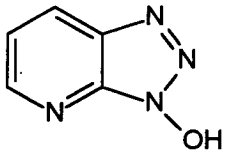
Figure 1.17. Formation of symmetrical anhydride by the action of carbodiimide.

Symmetrical anhydride activation is wasteful with only 50% of the carboxylic product incorporated into the peptide. Other disadvantages, also include the dehydration of Asn & Gln to the nitrile and high levels of racemisation with His. However, these undesirable side-reactions can be minimised by use of active esters in coupling reactions.

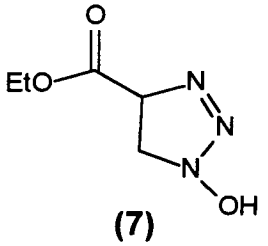
The active esters are prepared from reaction of one equivalent of carboxylic acid with one equivalent of the appropriate alcohol in DIC. A commonly used and extremely effective coupling agent is 1-hyroxybenzotriazole (HOBt) (5) whose catalytic activity was discovered by König and Geiger.[69]



(5)



(6)



(7)

Analogues developed as a consequence of the highly efficient coupling of HOBt esters include, 1-hydroxy-7-azabenzotriazole (6)[70] and ethyl-1-hydroxy-1H-1,2,3-triazole-4-carboxylate (HOCt) (7).[71]

1.2.5. Fmoc SPPS - side-chain protection.

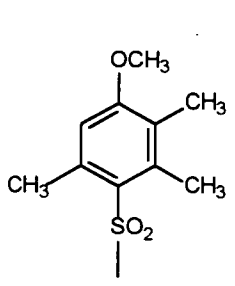
As mentioned previously side-chain protection in Fmoc SPPS is based on tertiary-butyl based PG's, the notable exceptions to this are Arg, Cys and His.

Amino acid	Protecting group
Arg	Mtr / Pmc
Asn/Gln	Mbh / Trt
Asp/Glu	Bu^t esters
Cys	Acm / Trt / SBu^t
His	π-Bum / τ-Trt
Lys	Boc
Ser/Thr/Tyr	Bu^t ethers
Trp	Boc

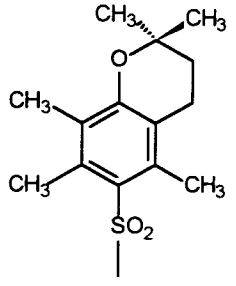
Table 1.2. Side-chain protection in Fmoc SPPS.

Protection for the strong nucleophilic guanidino side chain group of Arg have been developed on ring substituted arylsuphonyl derivatives such as the 4-methoxy-2,3,6-

trimethylbenzenesulphonyl (Mtr) (8).[72] However, the Mtr is somewhat resistant to TFA cleavage requiring extended cleavage times especially in multiple Arg sequences[73] and hence the more acid labile 2,2,5,7,8-pentamethylchroman-6-sulphonyl (Pmc) (9) is now preferred.[74,75]



(8) Mtr



(9) Pmc

His presents problems in that both the free and τ protected derivatives are prone to racemisation when activated due to intramolecular base catalysis by the π nitrogen of the imidazole ring. Trityl protection of the τ nitrogen significantly reduces the basicity of the π nitrogen and therefore suppresses proton abstraction from the α-carbon. However, racemisation is not completely eliminated but can be further reduced by coupling with the mildly acidic HOBt active ester.

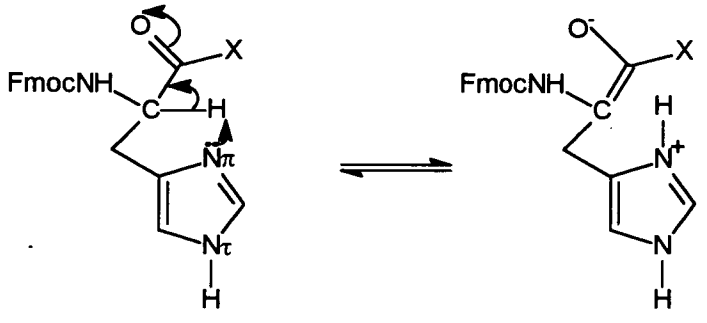


Figure 1.18. Histidine racemisation.

Blocking of the π nitrogen has been shown to virtually eliminate His racemisation[76] and π nitrogen protected with tertiary-butoxymethyl (Bum)[77] is available, however it is considerably more expensive than the τ nitrogen protected derivatives.

Differential protection is available for Cys on the basis of TFA lability and can be chosen on the criteria to have free or protected cysteines during the initial purification. Cys protected with trityl is TFA labile and is therefore cleaved to generate the free thiol during the standard cleavage. Alternatively, Cys protection can be retained if protected by the acetamidomethyl (Acm) or Bu^{t}thio (SBu^t) groups, both of which are stable to TFA. The Acm group is cleaved with silver trifluoromethane sulphonate[78] or mercury II ion[79] and SBu^t is cleaved by reduction with β-mercaptoethanol[80] or tributylphosphine.[81]

1.2.6. Fmoc SPPS-monitoring.

Coupling efficiencies cannot be assumed to go to completion and can often drop dramatically due to sequence effects such as steric hindrance e.g. the coupling of Cys(Trt) to His(Trt). Deprotection of Fmoc with base (20% piperidine/DMF) by a β-elimination mechanism results in a dibenzofulvene adduct which is quenched with excess piperidine giving an fulvene-piperidine adduct. This has a strong UV absorbance at 302nm and thus monitoring of the deprotection solution provides a quantitative estimation of the % efficiency at each cycle.

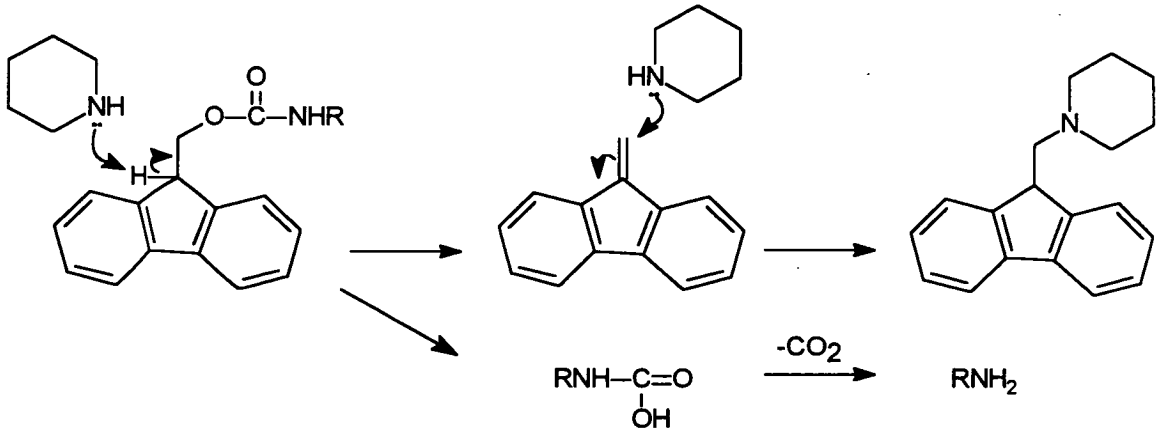


Figure 1.19. Deprotection of an Fmoc-protected peptide.

Although this technique provides a reliable estimation of coupling efficiencies at each cycle it does not importantly allow intervention to repeat the coupling or use extended

times if a particularly inefficient coupling has occurred. Towards this goal, monitoring of the activated amino acid by continuous circulation has been examined although this can be complicated by UV absorbance of the activating group. This therefore necessitates the development of a non UV active activating group which at the same time maintains high coupling efficiencies.

1.2.7. Fmoc SPPS-acidolytic cleavage.

The simultaneous cleavage of the peptide-resin linkage and removal of side-chain PG's can be achieved in most cases with 95% aqueous TFA by an S_N1 mechanism. This generates tertiary-butyl cations and tertiary-butyl trifluoroacetate which are quenched with scavengers (1,2-ethanedithiol/ethylmethylsulphide/phenol/thioanisole) to prevent tertiary-butylation of Met, Trp and Tyr. The scavengers are also important in shifting the cleavage mechanism from an S_N1 to S_N2 for certain less acid labile PG's such as the guanidino protection of Arg with Pmc.

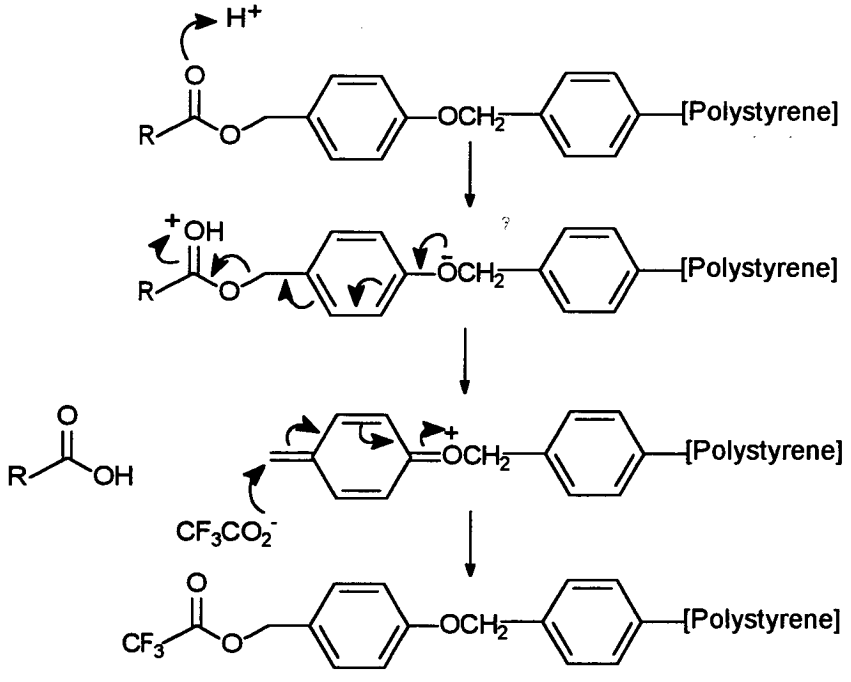


Figure 1.20. Acid catalysed cleavage of the peptide from the Wang resin.

1.2.8. Chemical ligation.

Advances in SPPS allow for the synthesis of moderately sized peptides of ca 100 amino acids in length before the accumulation of deletion products becomes prohibitively large.[82] To extend beyond this range, the chemical ligation of unprotected fragments to form larger proteins has been developed. This requires chemoselectivity in the unprotected peptides so that only the intended sites can react with each other in the presence of many side-chain functionality's.

Chemical ligation involving a native peptide bond between the two fragments has been developed by Kent[83] and is shown in figure 1.21. Ligation proceeds via nucleophilic attack by the thiol side chain of the N-terminus cysteine of peptide-2 with the thioester at the C-terminus of peptide-1. The ligation product then undergoes rapid intramolecular reaction via a favourable five membered ring and rearrangement to give the product with a native peptide bond at the ligation site. This methodology has been successfully used in the synthesis of the 72 amino acid human interleukin 8 which contains a total of four cysteines.

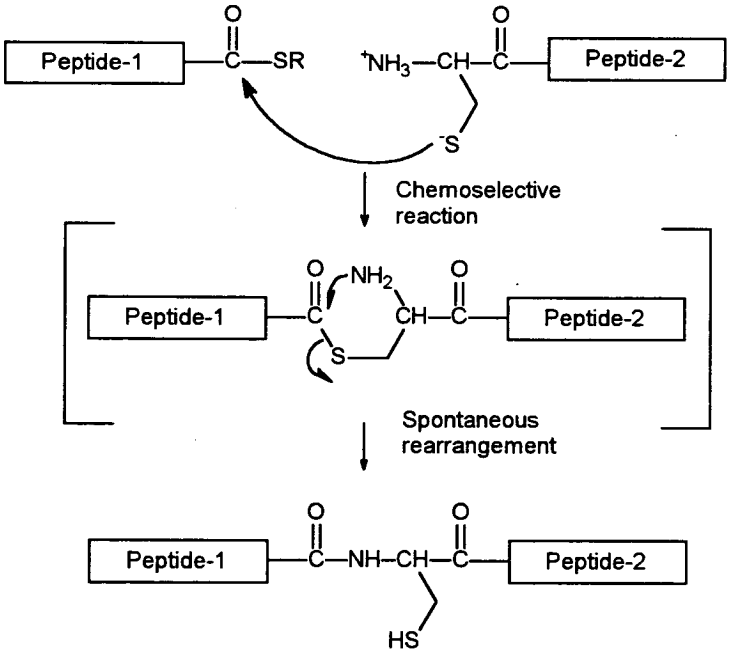


Figure 1.21. Chemical ligation between the C-terminal thioester of peptide-1 and the N-terminal cysteine of peptide-2.

However, ligation is not restricted to the native peptide bond with other ligation chemistries developed to include thioester-forming ligation[84,85,86] thioether formation,[87] thiazolidine ring formation[88] and oxime forming ligations.[89]

In a recent outstanding paper by Canne, three ligation reactions involving two thioester and one oxime forming ligations were performed to give the biologically active 20 kilodalton cMyc-Max heterodimer and the covalent Max homodimer which are important proteins for regulating gene expression through their DNA binding specificity's.[90]

1.2.9. Summary.

Therefore, continued advances in SPPS methodology allow for the synthesis of larger proteins, which in some instances may be difficult to obtain by recombinant DNA technology. Additionally, SPPS and chemical ligation allow for subtle changes in amino acid side-chains with unnatural or synthetic amino acids or indeed the peptide backbone through the use of non-native ligation procedures. Thus, an increasing array of tools are available to the protein chemist which provide new and exciting opportunities for the study of protein structure and function.

Chapter 2. Discussion

Section 2.1. Project statement.

The protein ubiquitin has been used as a model for folding studies with the synthesis of ubiquitin analogues including both single and double substitutions. Chemical synthesis has allowed for the facile incorporation of both unnatural amino acids and a novel synthetic fluorinated amino acid.

Purification of synthetic ubiquitin by a variety of protocols has identified important steps in the folding pathway of ubiquitin. These results have been used in the modification of subsequent purification's.

We have investigated the placement of fluorine within the hydrophobic core of ubiquitin through the synthesis of three double substitutions at positions 43&67, 50&67, 56&67 and one single substitution at position 67 using (2S,4S)-5-fluoroleucine to replace leucine at specific positions in the sequence. Methyl transfer across the hydrophobic core has been considered in the double substitution analogue [3-norleucine,43-norvaline]ubiquitin involving a modification in the distribution of alkyl groups. The importance of the chirality of the single histidine on the final strand of β-sheet has been examined in the synthesis of [68-Dhistidine]ubiquitin.

The contribution of secondary structural features in protein folding has been investigated by studying peptide fragments corresponding to the N and C-terminal halves of ubiquitin and three overlapping peptides on the final strand of β-sheet. The complementation of the N and C-terminal halves to regenerate the native fold has also been considered.

Structural information has been obtained using far UV circular dichroism, fluorescence studies, 1D & 2D ^{1}H NMR and where applicable ^{19}F NMR. The

biological activity has also been determined in order to correlate structure activity relationships.

Section 2.2. Synthetic ubiquitin.

During the purification of ubiquitin protein analogues it became apparent that some steps were crucial for the attainment of the native fold. In order to identify the best purification procedure synthetic ubiquitin was purified using several techniques. In particular it was found that high purity samples had a non-native fold and we were interested in purification/folding procedures which would allow access to the native fold.

The synthesis of ubiquitin gave an overall final yield of 66% for the N-terminal methionine on the resin as determined by quantitative Fmoc deprotection, which indeed is characteristic for the majority of ubiquitin analogues synthesised (figure 2.1.).

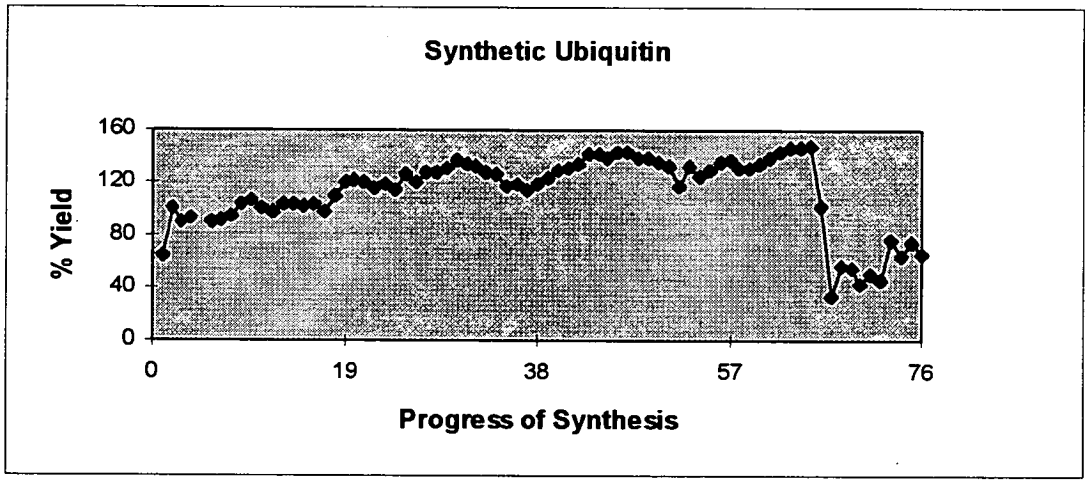


Figure 2.1. Progress of Ub(syn) synthesis as monitored by UV at 302nm of the deprotection solution containing the piperidine-fulvene adduct.

The observed % yields during chain assembly were observed to be greater than 100% which can be attributed to the swelling properties of the polystyrene resin that vary according to both the type of solvents used and the peptide sequence being synthesised. It is thought that increased resin swelling could result in the absorption

of more solvent which would increase the concentration of the piperidine-fulvene adduct in solution and hence result in observed % yields greater than 100%.[91]

The different purification protocols used in the purification of synthetic ubiquitin are summarised in figure 2.2.

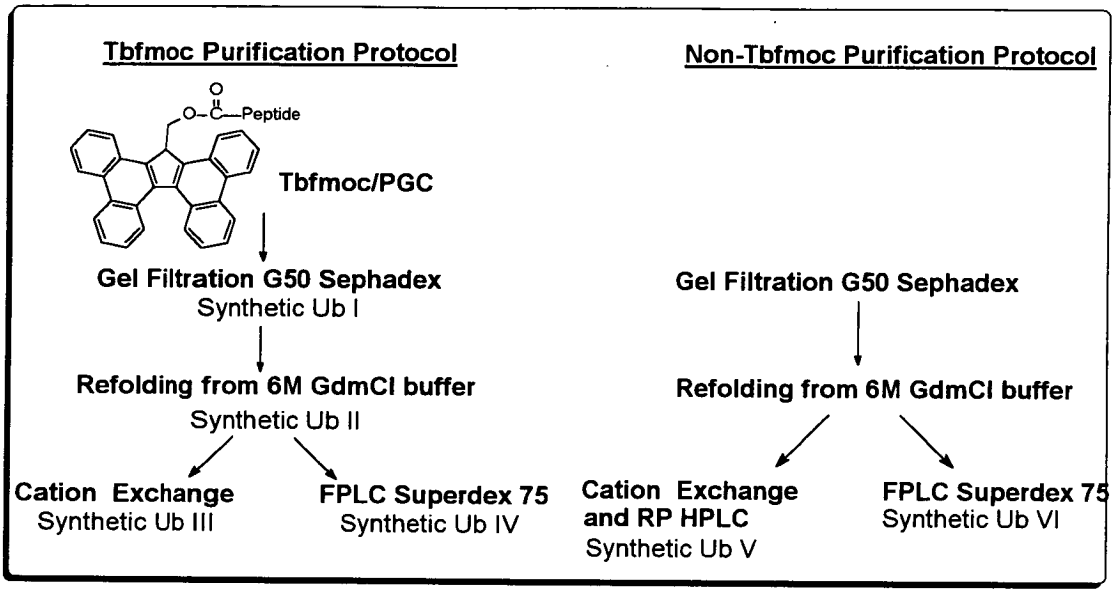


Figure 2.2. Purification protocols used in the study of synthetic ubiquitin.

2.2.1. Tbfmoc Purification Protocol.

During solid phase peptide synthesis acetylated truncates accumulate on the resin due to incomplete coupling steps and need to be separated from the desired peptide sequence. This can be achieved by derivatising the N-terminus with the base labile 17-tetrabenzo[*a,c,g,i*]fluorenylmethoxycarbonyl (Tbfmoc) group and exploiting its hydrophobicity to enable separation of the Tbfmoc-peptide from the truncates. Two methods of purification have been developed using either affinity purification on porous graphitised carbon (PGC) through the non covalent binding of the Tbfmoc-peptide or alternatively reverse phase HPLC which enables separation through the increased retention time of the Tbfmoc-peptide.[92,93,94,95] For the purification of ubiquitin protein analogues the Tbfmoc/PGC purification (figure 2.3.) was chosen in preference to reverse phase HPLC since initial studies had shown limited recovery of ubiquitin analogues in using the latter method.

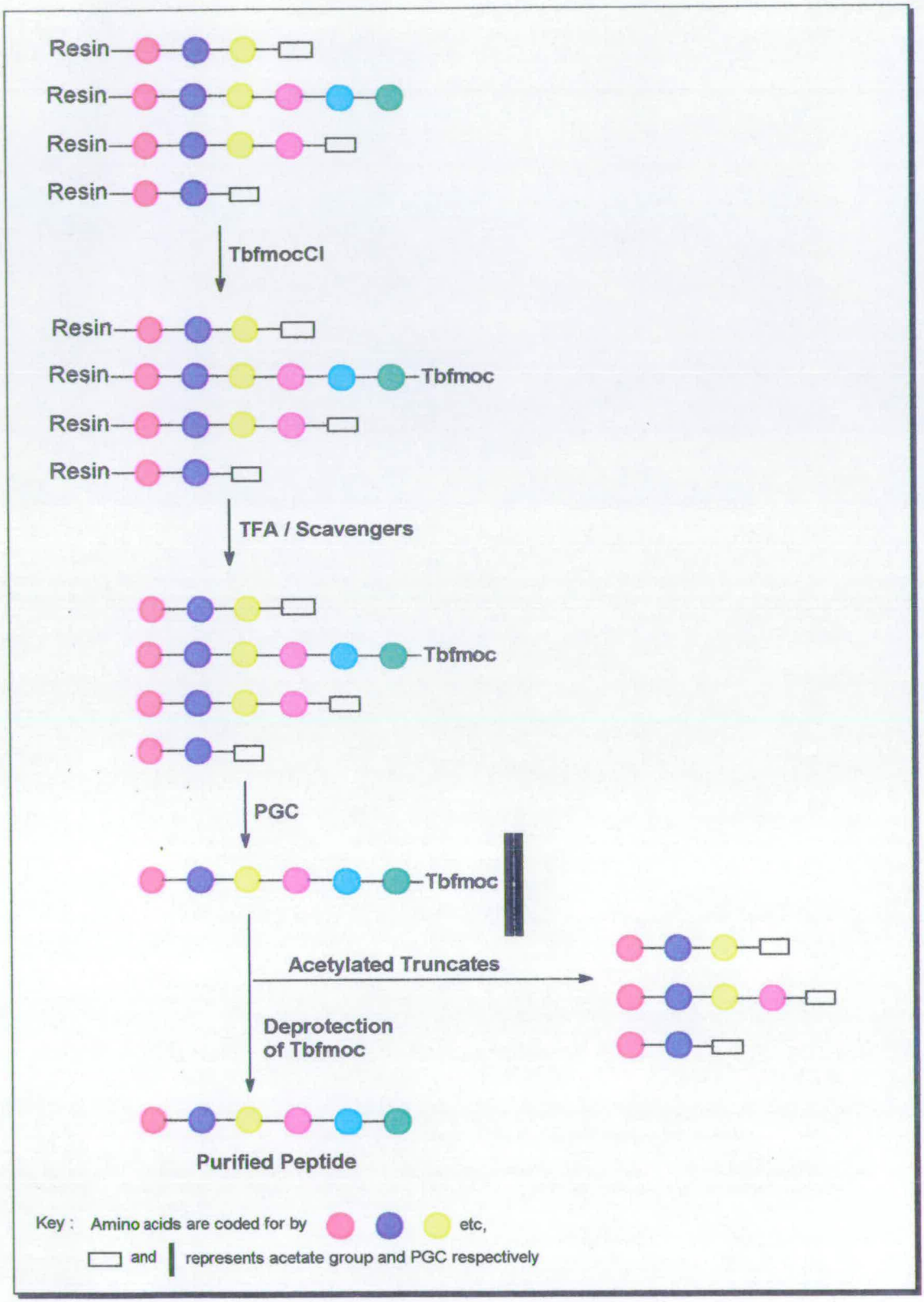


Figure 2.3. Schematic diagram of PGC/Tbfmoc purification.

The solvent chosen for the procedure was a 1:1 mixture of an aqueous 6M guanidinium chloride & isopropanol. The reasons for this were twofold, the guanidinium chloride was used to ensure the solubility of the Tbfmoc-protein which tended to be rather insoluble and the isopropanol because it had previously been shown that 50% aqueous isopropanol was a good solvent to ensure complete recovery of the peptide from the PGC after cleavage of the Tbfmoc moiety with base.[96]

Initial purification of synthetic ubiquitin used the Tbfmoc/PGC protocol followed by desalting by gel filtration giving Ub(syn) I. The AAA and MALDI MS for Ub(syn) I were as required, however the analytical RP HPLC of this was much broader compared to Ub(bovine) (figure 2.6).

A portion of Ub(syn) I was refolded from 6M guanidinium chloride to give Ub(syn) II and the 1D NMR spectra were recorded at 600 MHz on both Ub(syn) I (figure 2.4.) and Ub(syn) II to assess the degree of folding. Comparison of the 1D ^{1}H NMR's of Ub(syn) I & II indicated they were similar to Ub(bovine) but with broader NMR signals and lack of chemical shift dispersion indicating a non-native fold with perhaps more than one interconverting conformation.

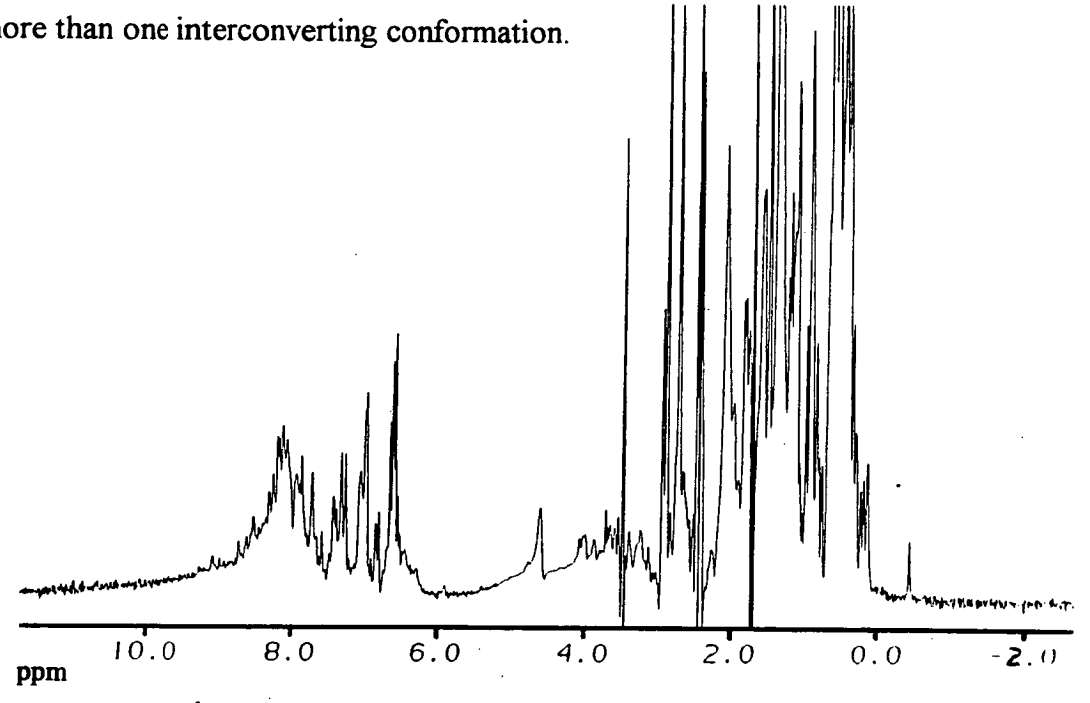


Figure 2.4. 1D ^{1}H NMR of Ub(syn) I at 600 MHz, pH 4·8 and 25°C.

The secondary structure content of Ub(syn) I was investigated using far UV circular dichroism (figure 2.5.).

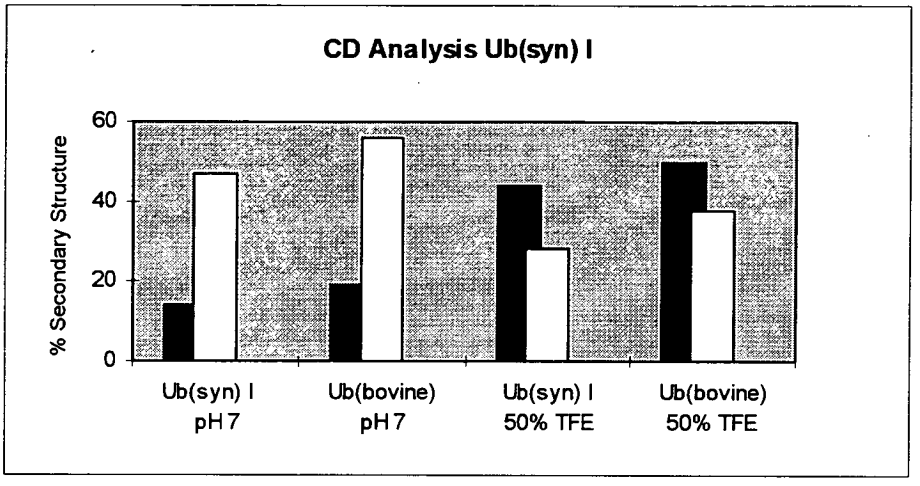


Figure 2.5. Deconvolutions of CD spectra into % secondary structure expressed as one of the following structures : heavily shaded bars, α-helix and clear bars, β-sheet

The CD analysis at pH 7 indicate both reduced α-helicity from 19% in Ub(bovine) to 14% in Ub(syn) I, and similarly reduced β-sheet to 47% in Ub(syn) from 56% in Ub(bovine). Addition of 50% TFE increases the α-helicity and decreases the β-sheet by approximately the same proportion in both, indicating similar structural preferences. These results indicate a similar secondary structure in Ub(syn) I and Ub(bovine), however CD analysis does not allow for identification of the actual residues involved in secondary structure.

A portion of Ub(syn) II was further purified by cation exchange chromatography on CM-Sepharose CL6B. The separation is based on the reversible adsorption of positively charged protein molecules to a negatively charged matrix. The protein sample initially binds to the matrix and selective separation and elution of protein is achieved by application of a pH gradient followed by a salt gradient. This resulted in the isolation of five protein fractions, of which the analytical RP HPLC profile of Ub(syn) III F4 was much sharper compared to Ub(syn) I & II previously obtained (figure 2.6).

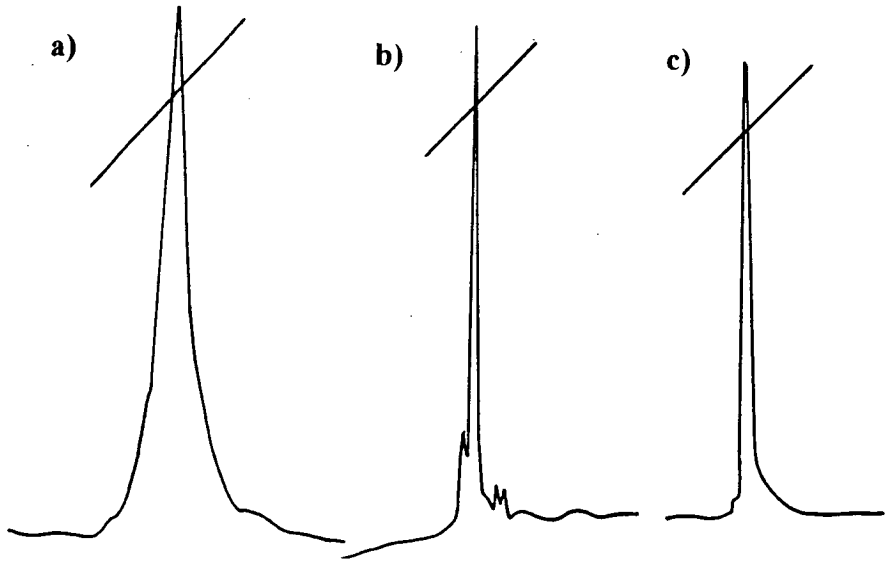


Figure 2.6. Analytical RP HPLC's of (a) Ub(syn) I, (b) Ub(syn)III F4 and (c)Ub(bovine).

Furthermore, the 1D ^{1}H NMR spectrum of Ub(syn) III F4 (figure 2.7.)was indistinguishable from that of Ub(bovine). This clearly shows that cation exchange is crucial for the folding of ubiquitin and in this instance should be viewed as a folding rather than a purification protocol.

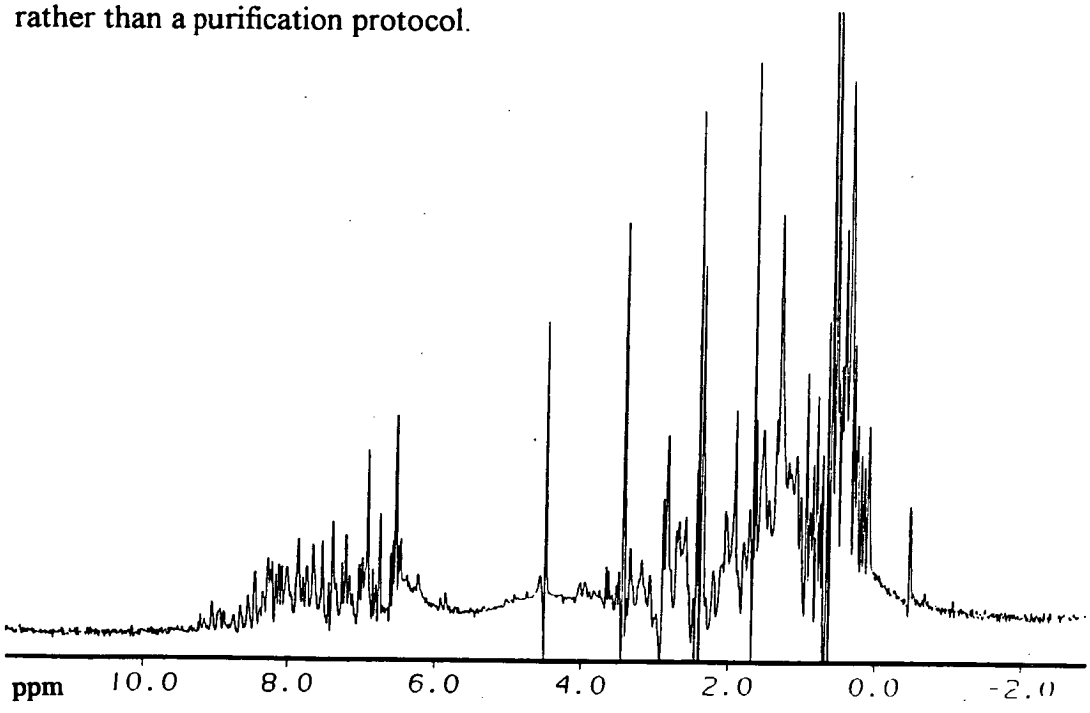


Figure 2.7. 1D ^{1}H NMR of Ub(syn) III F4 at 600 MHz, pH 4·8 and 25°C.

In order to quantify the importance of cation exchange chromatography a second portion of Ub(syn) II was further purified using FPLC Superdex 75 with the separation based on size differences. Two fractions were isolated, Ub(syn) IV F1 and F2, with the second fraction co-eluting with Ub(bovine). The analytical RP HPLC's of these were sharper than Ub(syn) I & II but less sharp than synthetic Ub(syn) III F4. The 1D ^{1}H NMR of Ub(syn) IV F2 was less well defined with reduced chemical shift dispersion compared to Ub(syn) III F4 indicating a non-native fold.

From these results we can conclude that the Tbfmoc group is perturbing the folding of the protein at a very early stage. The Tbfmoc group is highly hydrophobic and could be postulated to act as a template for the clustering of hydrophobic amino acids around it.

Another plausible explanation is that the N-terminus of ubiquitin plays an important role in directing the folding to the native state. Attachment of the Tbfmoc group to the N-terminus significantly distorts the folding to a non-native fold. The process of cation exchange chromatography does allow formation of the native fold and is obviously an important step in the folding of ubiquitin. Further work is required on the structure of Tbfmoc-ubiquitin, although difficulties would be encountered in working with it due to limited solubility.

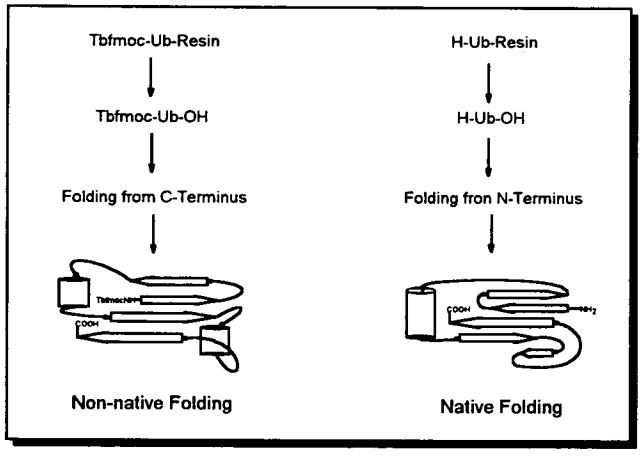


Figure 2.8. Ubiquitin folding pathways (secondary structure shown for non-native folding is only a schematic representation of a possible structure for Tbfmoc-ubiquitin)

The non-native folding is postulated to occur predominantly from the C-terminus prior to release of the purified protein from the PGC (figure 2.8.).

The conditions used in the Tbfmoc/PGC purification were investigated to see if they caused any structural changes. In particular, we studied the deprotection of the Tbfmoc group which used high pH to evaluate any structural perturbation. Therefore, Ub(bovine) was treated with 10% piperidine in 1:1 mixture of guanidinium chloride/isopropanol and subsequently desalted by semipreparative RP HPLC & lyophilised. The 1D ^{1}H NMR of this material was identical to non-treated Ub(bovine) confirming that the actual experimental conditions caused no structural changes but rather the Tbfmoc group resulted in non-native folding.

Several other proteins have been purified using the Tbfmoc purification protocol with no perturbation of structure observed. One example is the helix substitution analogue [26-α-aminobutyric,30-norvaline]ubiquitin which was purified using the Tbfmoc/RP HPLC protocol and showed no overall change in architecture apart from the slight rotation of Nva30 as determined by x-ray crystallography.[97] Although, interestingly the ^{1}H NMR of this analogue showed less dispersion of chemical shift suggesting that perhaps the physical process of crystallisation favoured formation of the native structure.[98]

2.2.2. Non-Tbfmoc Purification Protocol.

The earlier work of Muir *et al*[99] was repeated in order to confirm the folding of ubiquitin using a non-Tbfmoc purification protocol. Three step purification by gel filtration, cation exchange chromatography and RP HPLC gave Ub(syn) V. The AAA and MALDI MS were as required and the analytical RP HPLC was very sharp. Furthermore, the 1D ^{1}H NMR of Ub(syn) V was indistinguishable from Ub(bovine) with good chemical dispersion in the range up to 9ppm and two lower frequency signals at ca -0·5ppm (figure 2.9.).

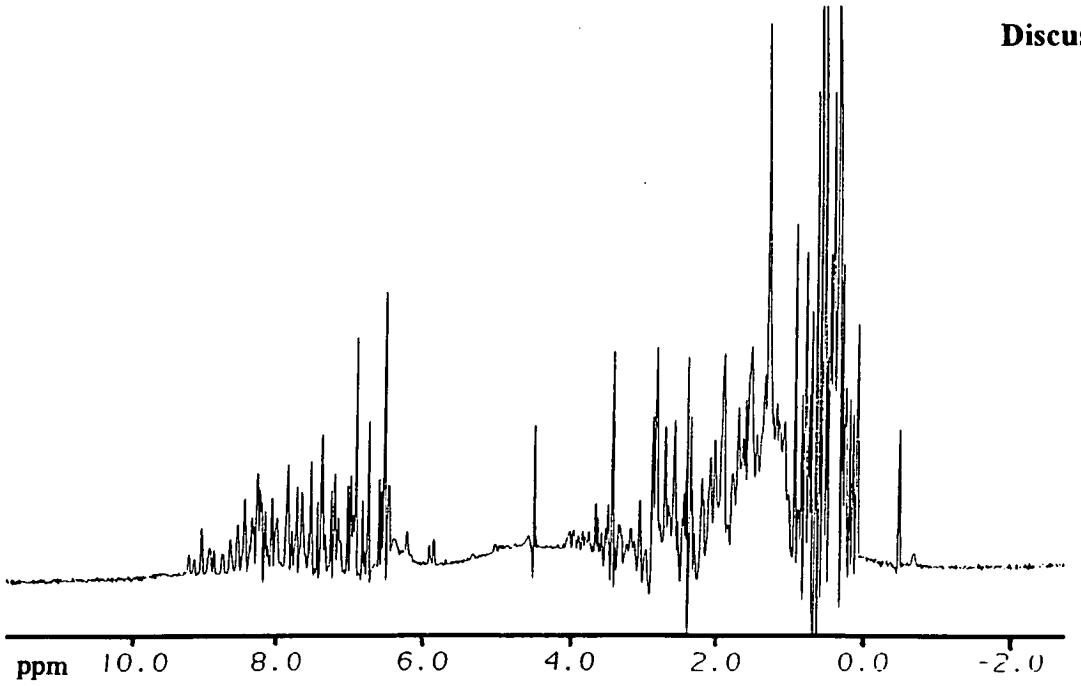


Figure 2.9. 1D ^{1}H NMR of Ub(syn) V at 600 MHz, pH 4·8 and 25°C.

Additional purification of a second portion by a non-Tfmoc protocol used FPLC Superdex 75 column instead of cation exchange chromatography, giving synthetic Ub(syn) VI. The AAA and MALDI MS were as required but the analytical RP HPLC of this was less sharp. The 1D ^{1}H NMR of Ub(syn) VI showed a similar dispersion of chemical shifts to that of Ub(syn) V, but with reduced definition of signal indicating a less well formed tertiary structure.

Conclusion for Synthetic Ubiquitin.

It has been shown that Tbfmoc/PGC purification results in a non-native structure for ubiquitin. The native fold can be accessed via the use of cation exchange chromatography which has important implications for selection of purification techniques. This serendipitous result has shown the importance of the N-terminus in determining the native fold at a very early stage of protein folding.

Section 2.3. Ubiquitin-(1-35), (36-76)-Peptides.

The Tbfmoc purification protocol has shown that the N-terminus is crucial for the attainment of the native fold in ubiquitin. To this end we have studied the intrinsic

structural properties of the N and C-terminal halves of ubiquitin. This dissection of proteins into fragments allows interpretation of the individual structural preferences of the excised fragment and identification of any intrinsic folding domains. A previous study by Harding *et al*[55] on the peptides Ub(1-21) & Ub(1-35) showed that the native structure was maintained in the absence of the C-terminal half in 60% aqueous methanol pH 2. However, although it is known that native-like structure is found in the N-terminal half under these conditions, no information is available for the C-terminal portion. The peptide Ub(1-35) comprises the β-hairpin (1-7 & 10-17) and the α-helix (23-34) while Ub(36-76) contains the remaining three strands of β-sheet (40-45, 48-50 & 64-72) and the short piece of 3_{10} helix (56-59).

2.3.1. Synthesis and Purification of Ub(1-35) and Ub(36-76).

Synthesis and purification of both peptides were straightforward giving final yields for the N-terminal amino acid on the resin of 70% for Ub(1-35) and a somewhat lower yield on the resin of 42% for Ub(36-76). The purification of Ub(1-35) used successive semi-preparative RP HPLC purifications. In contrast, due to the reduced % yield for Ub(36-76), a three step purification was employed using Tbfmoc/PGC, desalting by gel filtration and semi-preparative RP HPLC. The purity of the products were confirmed by AAA and MALDI MS (figure 2.10.).

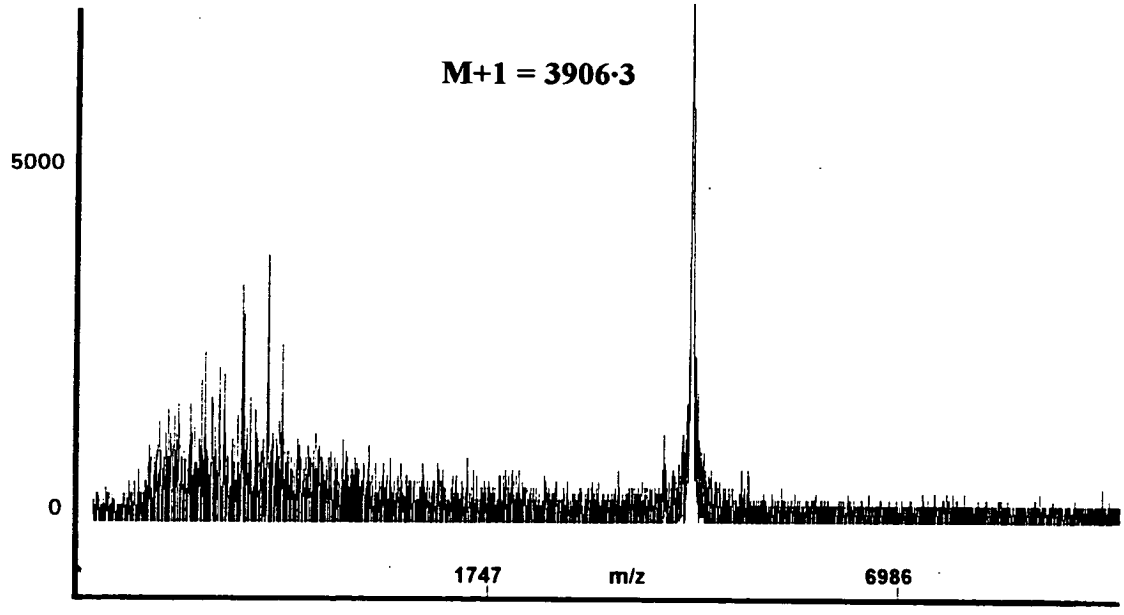


Figure 2.10. MALDI MS for Ub(1-35), M+1 requires 3906·5 Da.

2.3.2. Circular Dichroism Analysis on Ub(1-35) and Ub(36-76).

The far UV circular dichroism spectra of both peptides were measured in both methanol/water pH 2 and TFE/water pH 4·6 mixtures. For *in vitro* folding studies the solvent is very important in determining the secondary structure of the peptide and might be compared to the microsolvent environment created by nonlocal amino acids during formation of the tertiary structure. Alcohols such as methanol and TFE are known to promote the formation of α-helices in peptides although the precise mechanism of structure enhancement is unknown.[100] The helix promoting ability (HPA) of these solvent mixtures have been measured for the alanine tripeptide (Ac-$(Ala)_3$-NHMe) and the HPA order was calculated to be TFE/water = methanol (HPA=12) > methanol/water (HPA=4·5) > water (HPA=1) > TFE (HPA=0·3).

The far UV circular dichroism results for Ub(1-35) are shown in figure 2.11.

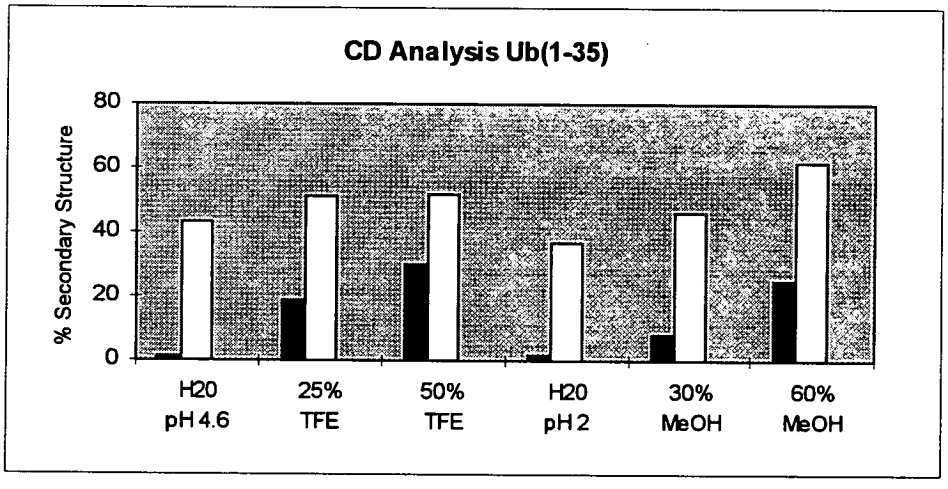


Figure 2.11. Deconvolutions of CD spectra into % secondary structure expressed as one of the following structures : heavily shaded bars, α-helix and clear bars, β-sheet

The calculated secondary structure for Ub(1-35) as present in the native state is α-helix (34%), β-sheet (46%) with remainder (20%). In water pH 4·6 the % of α-helix is negligible with the estimated structure divided almost equally between β-sheet and random structure. Addition of TFE greatly increases the α-helix content accompanied

by only a slight increase in β-sheet in both 25 & 50% TFE/water mixtures. In water pH 2, the results mirror those of water alone with a slight decrease in β-sheet observed with a corresponding increase in random structure. Addition of methanol increases the % of α-helix with 30% methanol giving only a modest increase to 8% helicity and 60% methanol increasing the helicity to 25% which is comparable to that observed in 50% TFE/water pH 4·6. The reduced helicity observed in 60% methanol pH 2 may result from increased flexibility in this portion of the molecule. In all of the solutions studied the % of β-sheet remains roughly constant thus supporting the strong intrinsic conformational preference of the β-hairpin in Ub(1-35).

The far UV circular dichroism results for Ub(36-76) are shown in figure 2.12.

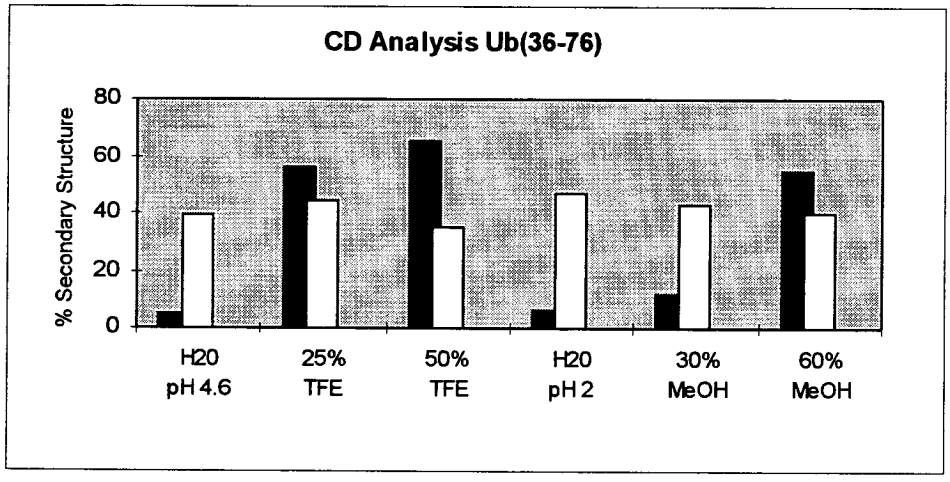


Figure 2.12. Deconvolutions of CD spectra into % secondary structure expressed as one of the following structures : heavily shaded bars, α-helix and clear bars, β-sheet.

The calculated secondary structure for Ub(36-76) as present in the native state is α-helix (0%), β-sheet (49%) with remainder (52%). In water pH 4·6 the % of α-helix is low (5%) but higher than for Ub(1-35), and as expected increases sharply with the addition of 25 & 50% TFE to 56 & 65% respectively. Surprisingly this increase in α-helicity is not accompanied by a corresponding decrease in β-sheet which remains reasonably constant. Addition of methanol again increases the α-helicity, with 30% methanol resulting in only a small increase to 12% α-helicity while 60% methanol

increases the α-helicity to 55%. Again the β-sheet structure is maintained with the increase in α-helicity correlating with a decrease in the % random coil structure.

From the actual CD spectra for Ub(36-76) we observe the expected correlation between the mean residue ellipticity at 222nm [θ_{222nm}] to the proportion of helical content. The large difference between the CD spectra in water and alcohol/water mixtures shown in figure 2.13. reflects the increased α–helicity in alcohol/water mixtures.

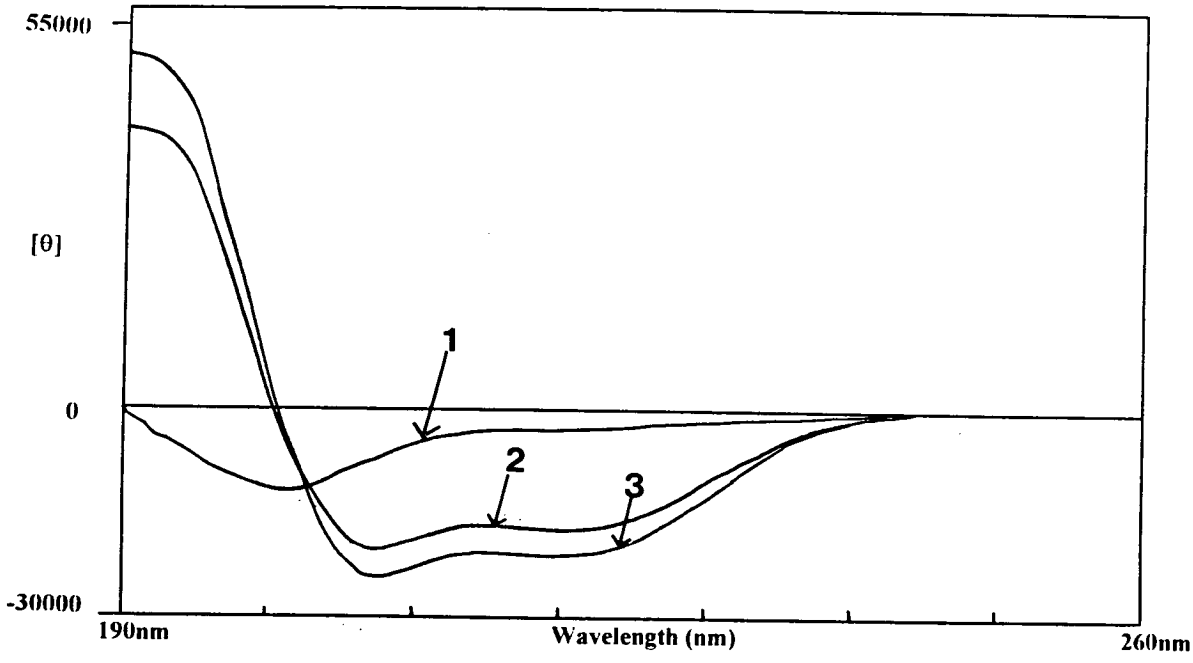


Figure 2.13. Far UV circular dichroism spectra for Ub(36-76) in (1) H_2O pH 4·6, (2) +25% TFE and (3) +50% TFE as indicated.

The results for Ub(1-35) and Ub(36-76) show that TFE or methanol mixtures promote the formation of α-helicity in both peptides studied. In Ub(1-35) the α-helix may be regarded as native-like but with increased flexibility while maintenance of β-sheet is also native as indicated by the earlier work of Harding *et al.*[55] The increase in α-helicity for Ub(36-76) can be regarded as non-native and results entirely from the

specific helix promoting abilities of the aqueous alcohol solvents. Further work is required to determine whether the β-sheet structure maintained is native or non-native.

2.3.3. Complementation of Ub(1-35) and Ub(36-76).

The analysis of peptide fragments is important as it is postulated they could form structures which are important early events in protein folding.[101] There are several examples of fragments of proteins associating together to regenerate the native-like fold as shown for the Trp repressor,[102] barnase[103] and chymotrypsin inhibitor-2.[104] Therefore, we were interested to study the interaction between the two peptides Ub(1-35) and Ub(36-76).

The 1D ^{1}H NMR's of Ub(1-35) and Ub(36-76) were recorded at the lower pH of 3·1 since it was found that Ub(1-35) formed an insoluble gel at higher pH's. The results for both are indicative of a random coil structure with no dispersion of chemical shift observed at either below 0·6ppm or above 8·5ppm.

In order to evaluate the association of the two peptides, an equimolar amount of Ub(36-76) was mixed with Ub(1-35) and left at 25°C. After 24 hours the 1D ^{1}H NMR was recorded and showed increased chemical shift dispersion at both the lower and higher frequencies, which was further enhanced after 96 hours (figure 2.14.). These results indicate that some association has occurred between the two peptides, resulting in the increased chemical shift dispersion. However, further 2D NMR experiments are required to discern the exact details of association

This experiment between the two peptides was performed in aqueous conditions and it would be interesting to investigate the influence of other solvents such as aqueous alcohols on directing the association between the two peptides.

A recent publication has shown that complementary assembly to give the native fold occurs in the ubiquitin α/β roll structure of the streptococcal protein G B1 domain,

between the two fragments (1-40) and (41-56).[105] Importantly no association was seen for other pairs of fragments studied which included (1-10) & (11-56), (1-20) & (21-56) and (1-30) & (31-56).

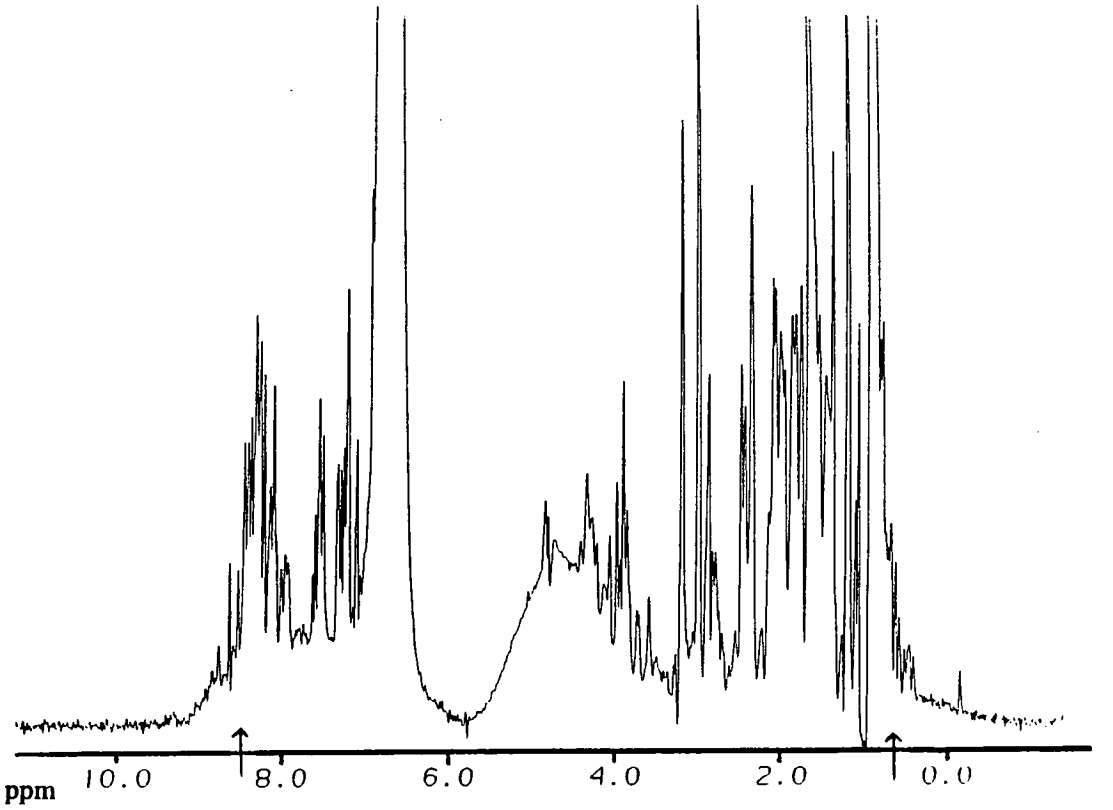


Figure 2.14. 1D ^{1}H NMR for equimolar mix of Ub(1-35) & Ub(36-76) after 96 hours at 600 MHz, pH 3·1 and 25°C: arrows indicate 0·6 and 8·5ppm respectively.

Section 2.4. Ubiquitin-(68-76), (66-76), (64-76)-Peptides.

64 76
Ac-GluSerThrLeuHisLeuValLeuArgLeuArgGlyGly-OH

Ac-ThrLeuHisLeuValLeuArgLeuArgGlyGly-OH

Ac-HisLeuValLeuArgLeuArgGlyGly-OH

Three overlapping peptides were synthesised corresponding to the final strand of β-sheet (64-72) and purified by semi-preparative RP HPLC. Following purification, the purity of the peptides were confirmed by AAA and FAB MS. To access the structural preferences of each peptide circular dichroism analysis were recorded in both TFE/water pH 4·6 and methanol/water pH 2 mixtures (figure 2.15.).

The results for Ub(68-76) indicate that the β-sheet remains approximately constant in all the solutions studied. Addition of TFE or methanol increases the α-helicity only slightly accompanied by a corresponding decrease in β-sheet.

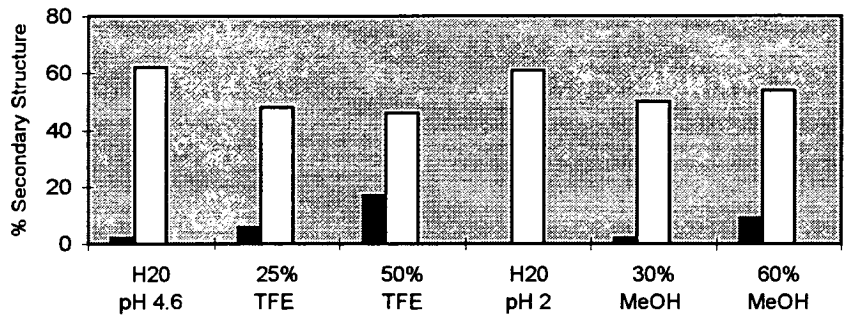
For Ub(66-76) the β-sheet content remains reasonably constant in all but the 30% methanol solution where random structure prevails. Addition of 25% or 50% TFE increases the α-helicity to a greater extent than in Ub(68-76) and importantly with no decrease in β-sheet observed. 60% Methanol results in only a minor increase in α–helicity with a corresponding decrease in β-sheet seen.

For Ub(64-76) the β-sheet structure shows more variation than in the other solutions studied. Addition of TFE or methanol increase the α-helicity to the same extent as in Ub(66-76) but this time with decreased β-sheet structure suggesting more flexibility in the structure.

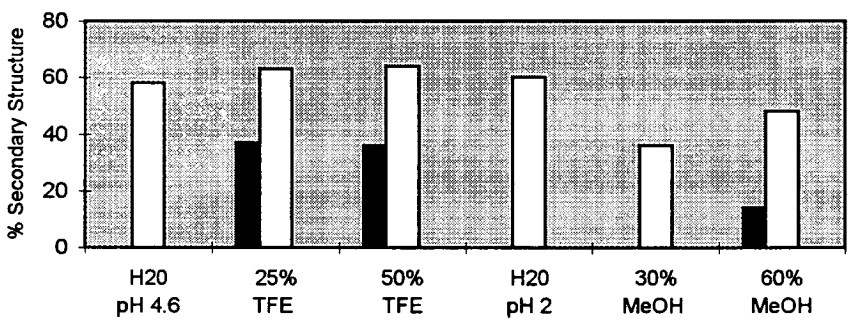
Therefore, overall there is limited helix forming tendencies in these small peptides studies but increased α-helicity is observed with increased chain length. The β-sheet

shows limited variation indicating a strong preference for the β-sheet structure which has formed.

a) CD analysis Ub(68-76)



b) CD analysis Ub(66-76)



c) CD analysis Ub(64-76)

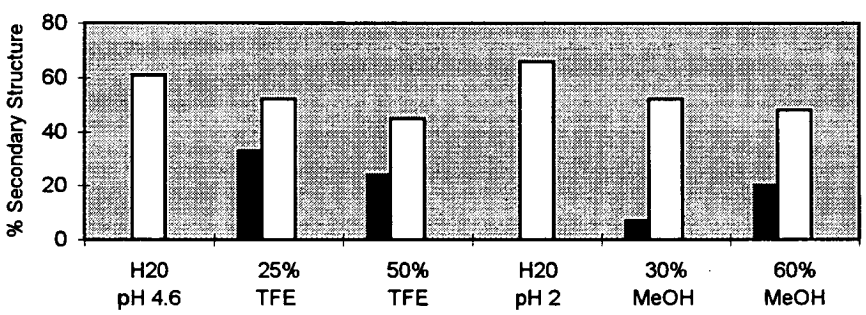
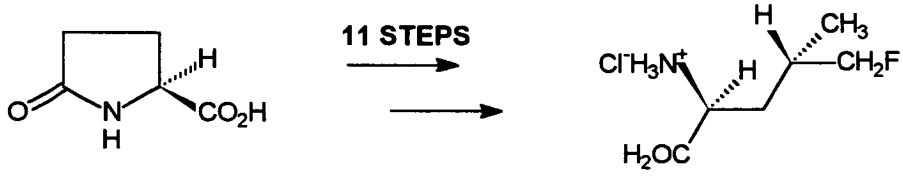


Figure 2.15. Deconvolutions of CD spectra for a) Ub(68-76), b) Ub(66-76) and c) Ub(64-76) into % secondary structure expressed as one of the following structures : heavily shaded bars, α-helix ; clear bars, β-sheet.

Section 2.5. Ubiquitin Fluorinated Analogues.

Several fluorine containing ubiquitin analogues have been made in order to investigate the structural and biological consequences of fluorine substitution within the primary sequence. Chemical synthesis offers the advantage of specific incorporation of the fluorine containing amino acid at any predetermined position. This compares to biosynthetic means in which all positions in the sequence occupied by a given amino acid become fluorine labelled to some degree. Computer graphics indicate that incorporation of fluorine into a peptide or protein results in only minor structural variation and often no effect on biochemical properties is observed unless a critical amino acid has been substituted.[106] The NMR properties of fluorine are particularly useful having a spin of ½, present in 100% natural abundance and a similar sensitivity to that of hydrogen. Furthermore, the fluorine nucleus is very sensitive to its environment as reflected in the large chemical shift range for fluorine containing compounds exceeding 900ppm.[107] Hence, fluorine chemical shifts vary between the unfolded and native states of a protein and therefore conformational changes can be monitored by ^{19}F NMR studies as shown for the native and unfolded forms of the intestinal fatty acid-binding protein.[108] A further advantage of ^{19}F NMR is that we can monitor specifically the change in environment for the individual amino acid side-chain substituted compared to the complexity of ^{1}H NMR relating to all amino acids present.

In order to gain information on the hydrophobic core of ubiquitin we have constructed both double and single substitutions using (2S,4S)-5-fluoroleucine (FLeu) (figure 2.16.) to replace leucine at specific positions in the sequence.



(2S)-Pyroglutamic acid **(2S,4S)-5-Fluoroleucine**

Figure 2.16. Synthesis of (2S,4S)-5-fluoroleucine by Professor D.W. Young and B.A. Starkmann, University of Sussex, UK.

These are isosteric replacements and hence should result in negligible steric effects. The major change in replacing hydrogen with fluorine is the difference in electronegativities between the two and how the highly electronegative fluorine will be accommodated within the core.

Ubiquitin contains a total of nine leucine residues in the primary sequence (figure 2.17.), five of which are in reasonable proximity to each other and of interest for substitution.

```
1                                                   38
MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPP

39                                                  76
DQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGG
```

Figure 2.17. Primary sequence of human ubiquitin with leucine residues highlighted at positions 8, 15, 43, 50, 56, 67, 69, 71 and 73.

The relative distance between the five residues 15, 43, 50, 56 & 67 which form part of the hydrophobic core are shown in table 2.1.

Table 2.1. Relative distance (A°) between δ1 carbon of Leu residues in ubiquitin based on the 3D structure.[15]

-	Leu 15	Leu 43	Leu 50	Leu 56	Leu 67
Leu 15	-				
Leu 43	**9.067**	-			
Leu 50	**12.154**	**4.825**	-		
Leu 56	**9.447**	**8.419**	**7.571**	-	
Leu 67	**8.488**	**4.485**	**4.182**	**6.281**	-

The following double substitution analogues, [FLeu^{43},FLeu^{67}]Ub, [FLeu^{50},FLeu^{67}]Ub, [FLeu^{56},FLeu^{67}]Ub using position 67 as the banker residue and the mono-substitution [FLeu^{67}]Ub^{109} were chosen for synthesis. Residues 43, 50 & 67 are on the third,

fourth and fifth strands of β-sheet respectively while residue 56 is part of the surface loop comprising residues 51-59. Their related spatial positions are shown in figure 2.18.

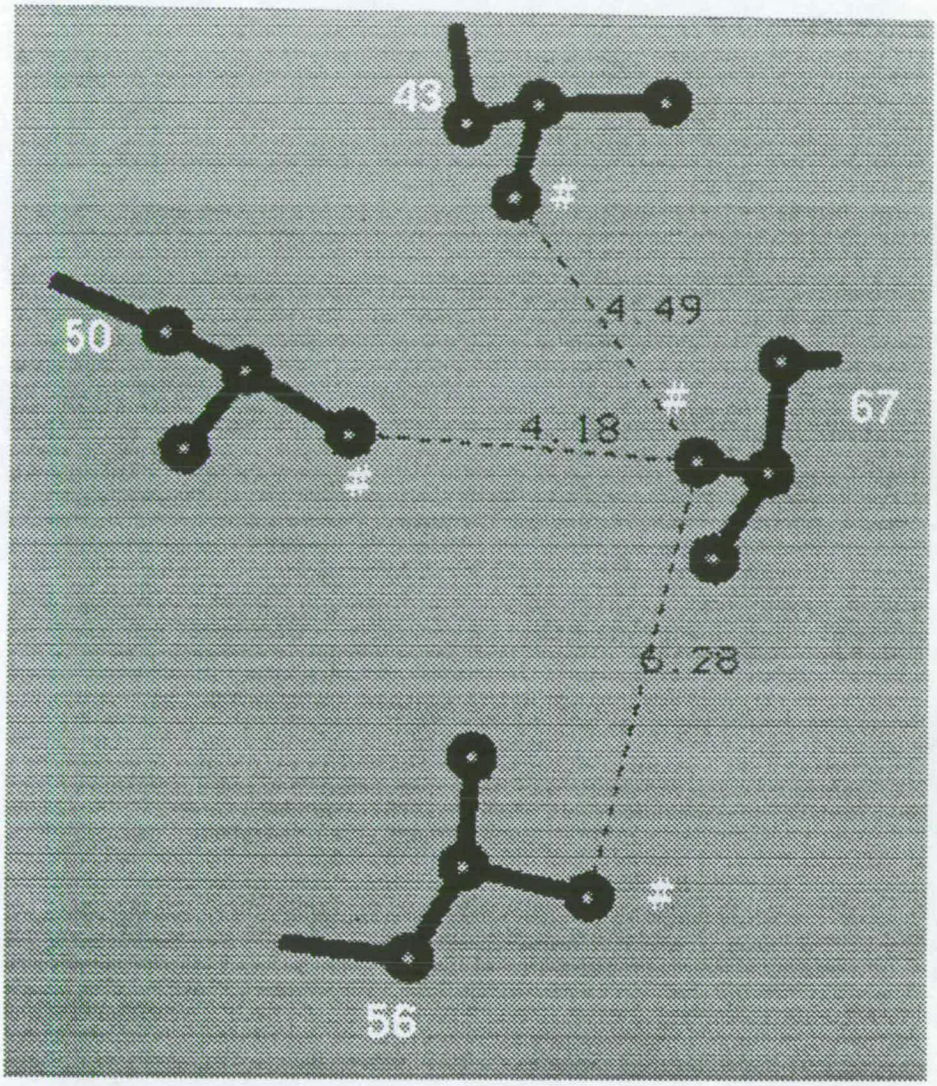


Figure 2.18. Diagram of FLeu side-chains at positions 43, 50, 56 and 67 with δ - CH_2F indicated by #: produced using Insight II program.

All analogues were synthesised using standard solid phase synthesis procedure except for the coupling of (2S,4S)-5-fluoroleucine which utilised sonication to maximise the reaction. The synthesis of [FLeu^{43},FLeu^{67}]Ub and [FLeu^{50},FLeu^{67}]Ub gave excellent overall yields with 72% and 83% respectively for the N-terminal methionine on the resin. The synthesis of [FLeu^{56},FLeu^{67}]Ub was less good due to the inefficient coupling of the FLeu residues giving an overall yield of 32% for the N-terminal methionine on the resin (figure 2.19.).

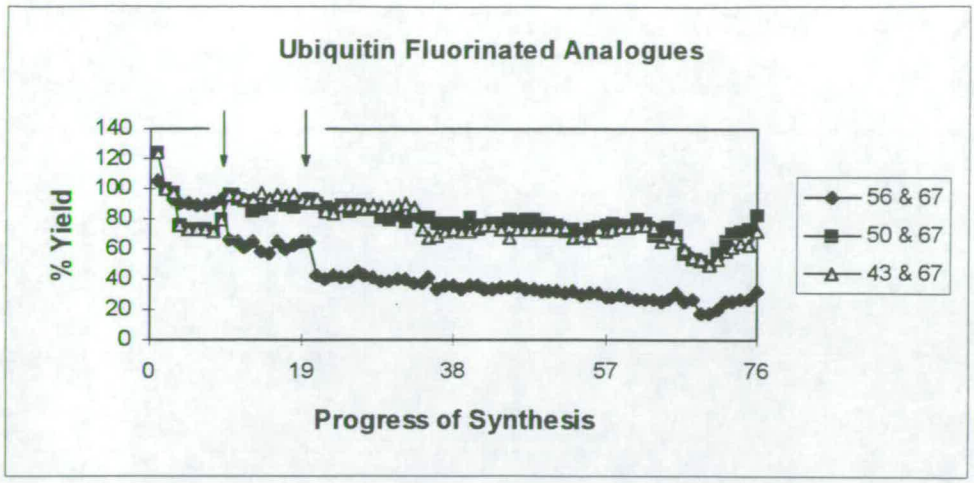


Figure 2.19. Progress of ubiquitin fluorinated analogues syntheses as monitored by UV at 302nm of the deprotection solution containing the piperidine-fulvene adduct. The inefficient coupling of FLeu67 & FLeu56 in the synthesis of [FLeu^{56},FLeu^{67}]Ub are indicated by arrows.

Each ubiquitin fluorinated analogue will be discussed separately apart from an overall discussion of the fluorine NMR data from all the protein analogues at the end of the section.

Section 2.5.1. [10,35,47&53-^{15}N-Glycine, 56&67-(2S,4S)-5-fluoroleucine] ubiquitin

Purification of this analogue by standard chromatographic techniques would have proved extremely difficult due to the low overall yield resulting in a high quantity of truncates as shown in figure 2.20.

However, the high capability of the Tbfmoc group selectively separates the required peptide from the truncates, thus simplifying the purification. Two step purification by Tbfmoc/PGC and desalting by RP HPLC gave synthetic [FLeu^{56},FLeu^{67}]Ub I in reasonable quantities. Characterisation by AAA and MALDI MS confirmed the authenticity of the purified product. Prior to starting structural analysis, [FLeu^{56},FLeu^{67}]Ub I was dissolved in 8M urea buffer and sequentially dialysed against decreasing urea concentrations to allow refolding from the denatured state.

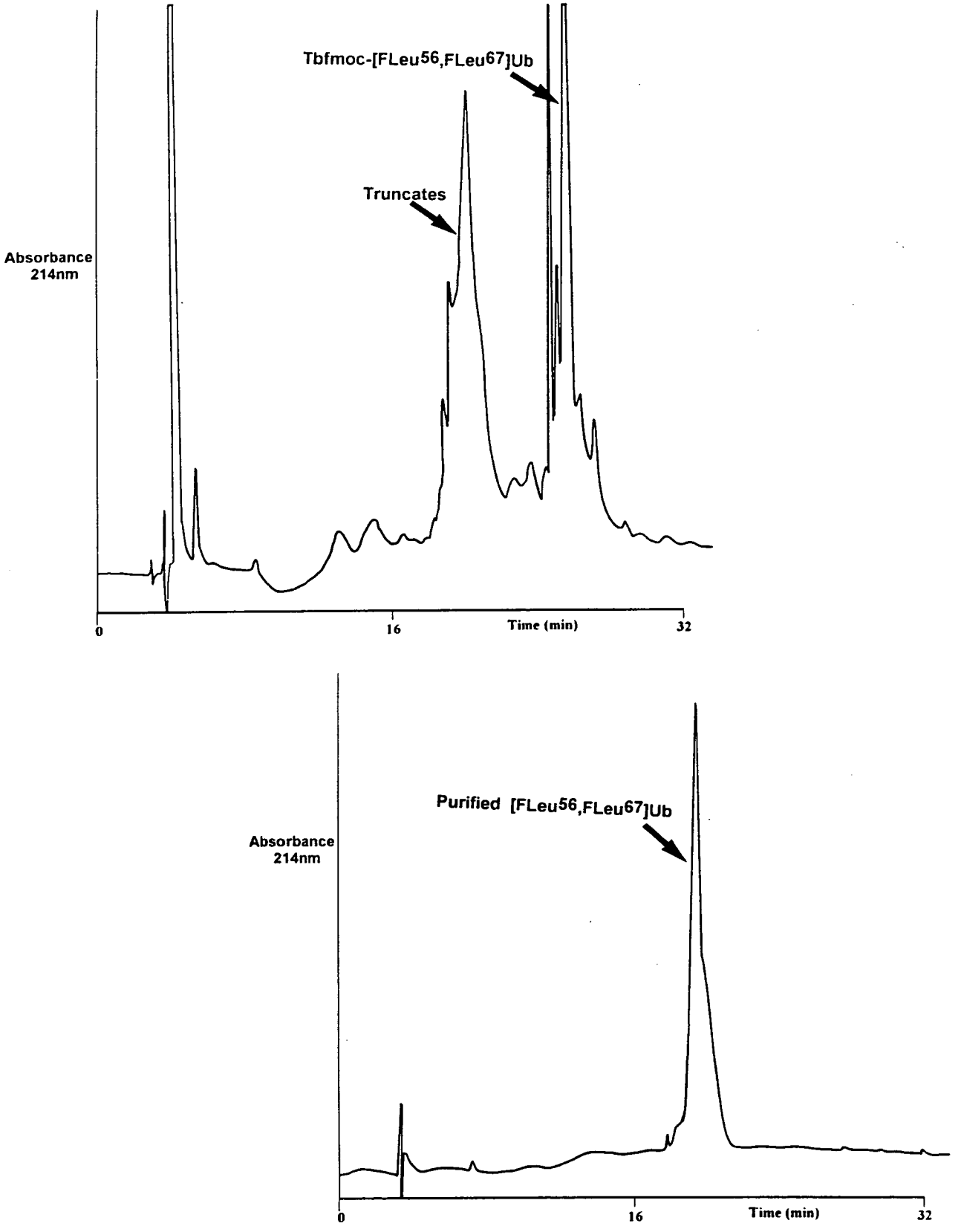


Figure 2.20. Analytical RP HPLC's of crude Tbfmoc-[FLeu^{56},FLeu^{67}]Ub (upper) and purified [FLeu^{56},FLeu^{67}]Ub I (lower).

The far UV circular dichroism results for [FLeu^{56},FLeu^{67}]Ub I are shown in figure 2.21.

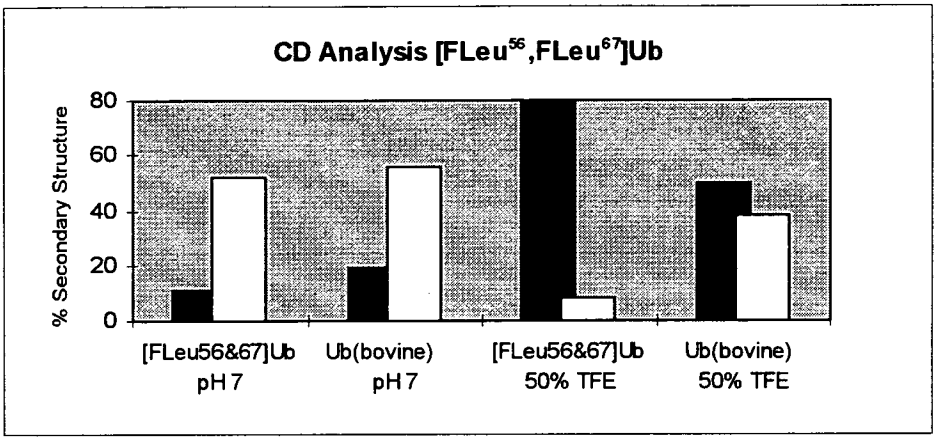


Figure 2.21. Deconvolutions of CD spectra into % secondary structure expressed as one of the following structures : heavily shaded bars, α-helix and clear bars, β-sheet.

The calculated secondary structure for [FLeu^{56},FLeu^{67}]Ub I indicates a decrease in the α-helicity from 19% as measured in Ub(bovine) to only 11% in the analogue. This compares with the β-sheet structure which is approximately equal in both with 52% observed in the analogue compared to 56% in Ub(bovine). Addition of 50% TFE increases the α–helicity as expected, but the increase is far greater in the analogue, 80% compared to only 50% in Ub(bovine). This large increase in α–helicity is accompanied by a corresponding decrease in β-sheet which suggests the overall β-sheet structure is less rigid in the analogue being easily promoted to α-helicity.

The environment of the single tyrosine at position 59 was investigated using fluorescence studies. In Ub(bovine) the fluorescence is partially quenched by the surrounding amide bonds in the folded state but increases upon unfolding in denaturing conditions. The disadvantages of using Tyr59 to monitor structural changes is its lack of sensitivity compared to Trp and also its position on the periphery

of the structure which means information gained is really only applicable to the loop comprising residues 51-59.

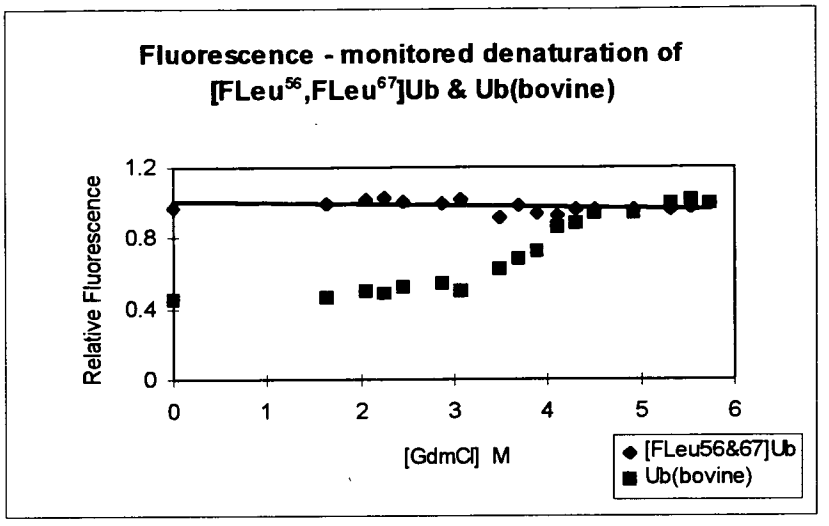


Figure 2.22. Plot of relative fluorescence for [FLeu^{56},FLeu^{67}]Ub I and Ub(bovine) versus [GdmCl].

From figure 2.22. we can clearly observe the expected unfolding curve for Ub(bovine) with increasing concentration of denaturant. No unfolding curve is seen for [FLeu^{56},FLeu^{67}]Ub I suggesting that Tyr59 is already exposed to the solvent in non-denaturing conditions. The data set for the guanidinium chloride unfolding curve of Ub(bovine) can be fitted with the non-linear regression analysis equation 3, which is derived from equations 1 & 2 and assumes a two state model of unfolding.[48]

Equilibrium constant for unfolding:

$$K_U = (F_N\text{-}F) / (F\text{-}F_U) = \exp(-\Delta G_U / RT) \quad (1)$$

Free energy of unfolding:

$$\Delta G = \Delta G_{H2O} - m[GdmCl] \quad (2)$$

Observed fluorescence:

$$F = F_N - (F_N - F_U) \times \frac{\exp(m[GdmCl] - \Delta G_{H2O} / RT)}{1 + (\exp(m[GdmCl] - \Delta G_{H2O} / RT)} \quad (3)$$

where K_U is the equilibrium constant for unfolding, ΔG_U is the free energy of unfolding at a particular concentration of denaturant, ΔG_{H2O} is the free energy of unfolding in water, F is the observed fluorescence and F_N and F_U are the values of fluorescence in the native and denatured states, m is a constant and R is the gas constant.

The free energy of stabilisation for Ub(bovine) was calculated by Grafit, Microsoft using nonlinear regression analysis to be ΔG_{H2O}=7·1kcalmol^{-1} which compares favourably with the expected value of ca 7kcalmol^{-1}.[40]

The 1D ^{1}H NMR of [FLeu^{56},FLeu^{67}]Ub I was recorded and shows both less dispersion and definition of chemical shifts, although the overall shape of the Ub(bovine) is clearly visible (figure 2.23.).

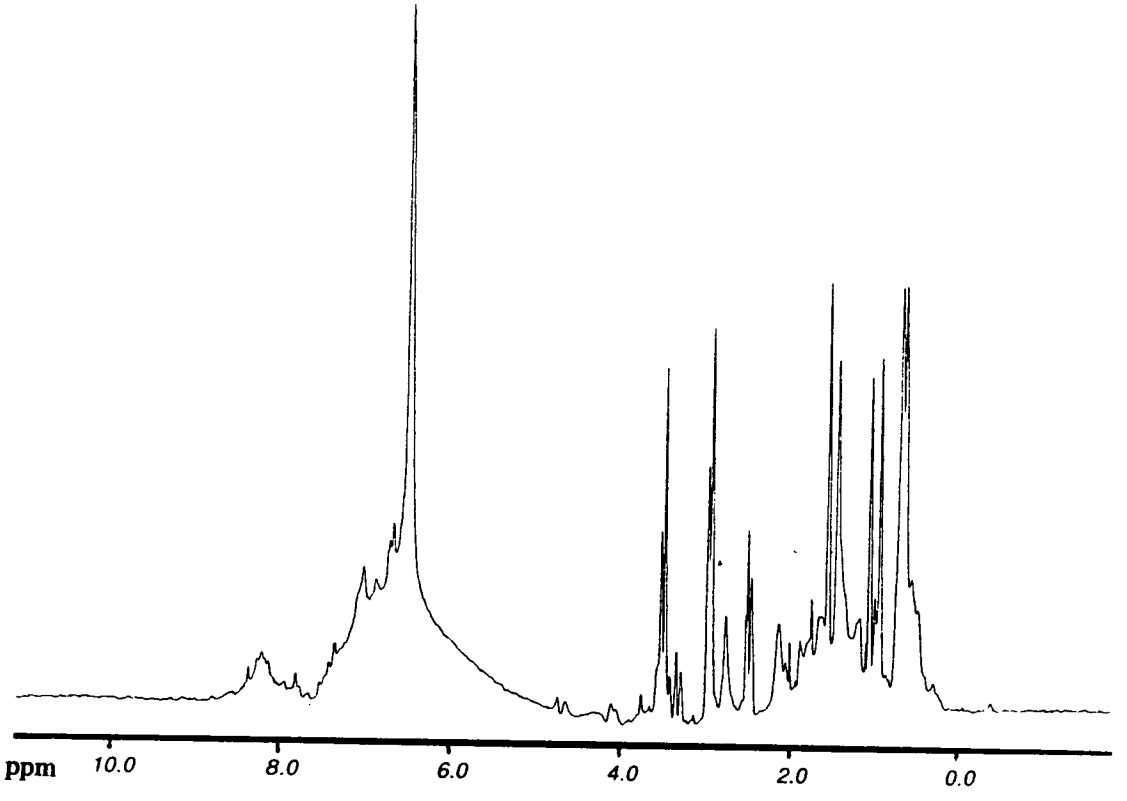


Figure 2.23. 1D ^{1}H NMR of [FLeu^{56},FLeu^{67}]Ub I at 600 MHz, pH 5·1 and 15°C.

Section 2.5.2. [50&67-(2S,4S)-5-Fluoroleucine]ubiquitin.

Initially half of the resin bound product was purified using the Tbfmoc/PGC protocol and desalted by gel filtration to give [FLeu^{50},FLeu^{67}]Ub I. The MALDI MS and AAA were as required indicating a pure sample but the analytical RP HPLC was much broader compared to Ub(bovine) (figure 2.24.). Furthermore the 1D ^{1}H NMR of [FLeu^{50},FLeu^{67}]Ub I had a narrower dispersion of chemical shifts in the region ca >

8ppm region compared with Ub(bovine), although one lower frequency signal at -0·5ppm is seen (figure 2.24.). The overall shape of the NMR spectra is the same in both the analogue and Ub(bovine), but with reduced definition in the analogue suggesting slow tumbling between more than one conformation. No improvement in the ^{1}H NMR occurred after denaturation and refolding from 8M urea or from subsequent lyophilisation of the refolded solution and dissolution into the NMR buffer again.

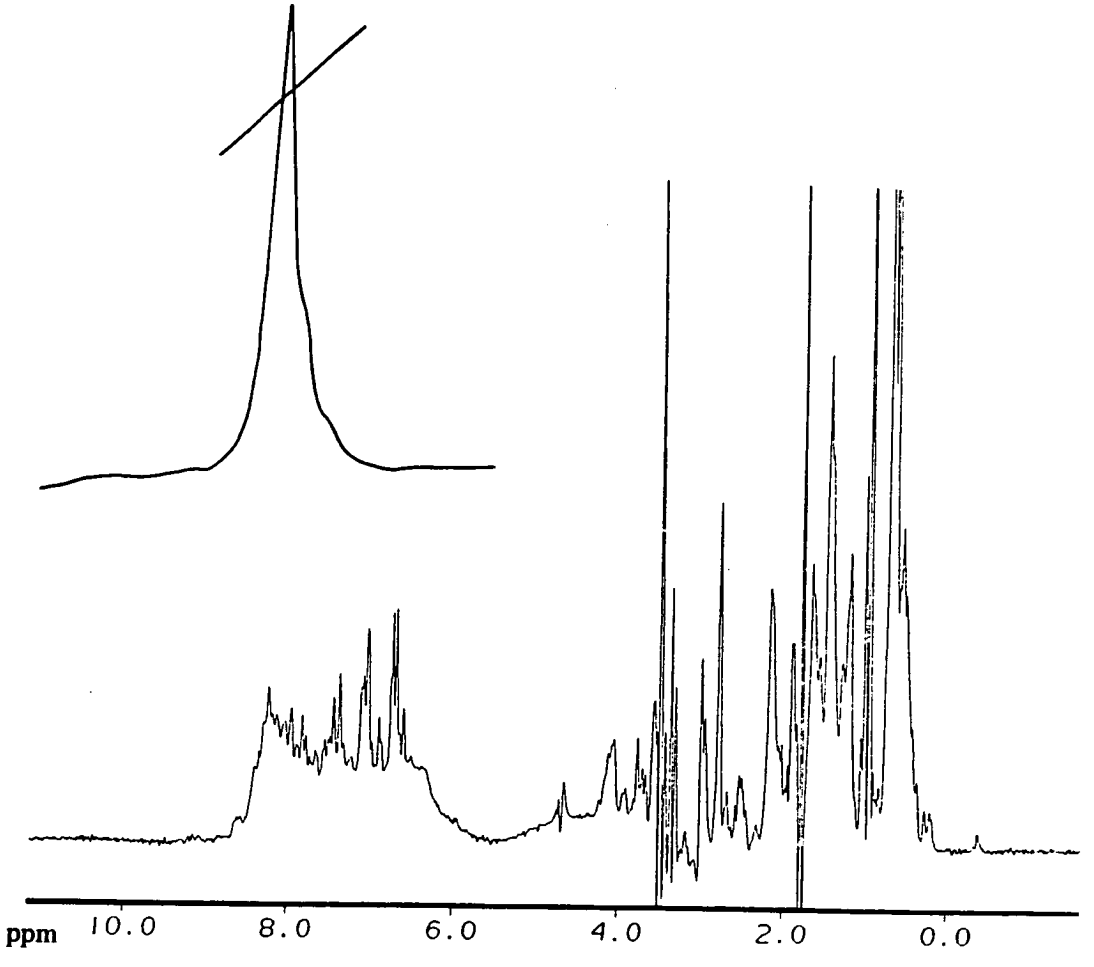


Figure 2.24. 1D ^{1}H NMR of [FLeu^{50},FLeu^{67}]Ub I at 600 MHz, pH 5·1 and 15°C with inset of analytical RP HPLC.

Further purification of [FLeu^{50},FLeu^{67}]Ub I by FPLC Superdex 75 gave synthetic [FLeu^{50},FLeu^{67}]Ub II F1 & F2. During FPLC purification it was noted that that two closely spaced protein peaks were observed in the later eluting fraction perhaps

indicating two slowly interconverting conformations. Again, the analytical RP HPLC's of [FLeu^{50},FLeu^{67}]Ub II F1 & F2 were broad and in this instance the MALDI MS was much more difficult to obtain. In general, we have noticed that it is far easier to obtain MALDI MS data for ubiquitin protein analogues that have a well defined tertiary structure as indicated by ^{1}H NMR.[110]

The far UV circular dichroism results for [FLeu^{50},FLeu^{67}]Ub II F2 are shown in figure 2.25.

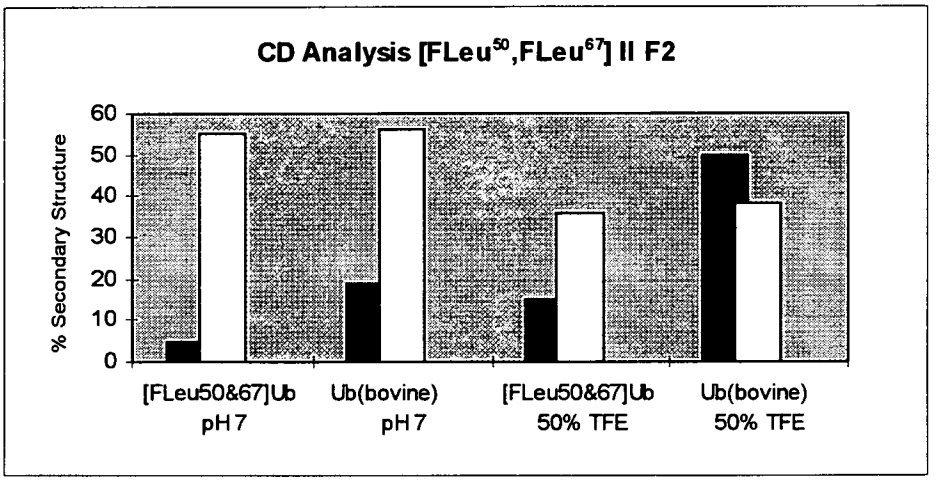


Figure 2.25 Deconvolutions of CD spectra into % secondary structure expressed as one of the following structures : heavily shaded bars, α-helix and clear bars, β-sheet.

The CD results indicate the α-helicity is much reduced from 19% in Ub(bovine) to only 5% α-helicity in [FLeu^{50},FLeu^{67}]Ub II F2, while the β-sheet content is virtually identical in both. Addition of 50% TFE results in a small increase in α-helicity to 15% in [FLeu^{50},FLeu^{67}]Ub II F2 compared to 50% in Ub(bovine), at the same time the β-sheet content decreases to ca 37% in both.

Previous work on Ub(syn) had shown that the native fold could be attained by cation exchange, however application of [FLeu^{50},FLeu^{67}]Ub II F1/F2 to the cation exchange column resulted in only poor recovery of [FLeu^{50},FLeu^{67}]Ub III F1/F2/F3/F4 and no further work was done on these fractions.

The second half of resin bound product was purified by a two step protocol using gel filtration and cation exchange chromatography giving [FLeu^{50},FLeu^{67}]Ub IV. Good correlation was seen for the AAA, much improved compared to those samples purified using the Tbfmoc/PGC protocol. We believe this reflects the different 'folded' structures present in the samples rather than the overall purity. For example, previous attempts to optimise the hydrolysis time for [FLeu^{50},FLeu^{67}]Ub I with a timed hydrolysis at 24, 40, 60 & 70 hours gave no overall improvement in the AAA results.

The analytical RP HPLC of [FLeu^{50},FLeu^{67}]Ub IV was as sharp as Ub(bovine) although it eluted earlier with a R_t=17·4 minutes compared to a R_t=17·6 minutes for Ub(bovine) (figure 2.27.). The MALDI MS was 17·1Da higher than expected, and furthermore [FLeu^{50},FLeu^{67}]Ub IV co-eluted with [MetS-oxide^{1}]Ub(bovine) confirming that the methionine had oxidised to the sulphoxide during the purification protocol.

The 1D ^{1}H NMR of [FLeu^{50},FLeu^{67}]Ub IV was much sharper than those obtained previously and indeed with good dispersion of chemical shifts seen albeit with reduced definition in the region 8-9ppm indicating formation of a looser native fold (figure 2.27.). This is expected and was noted by Breslow *et al* during their study on [MetS-oxide^{1}]Ub because the two possible sulphoxide isomers weaken the stability of the hydrogen bond between Met1 and Lys63 (figure 2.26.).[111]

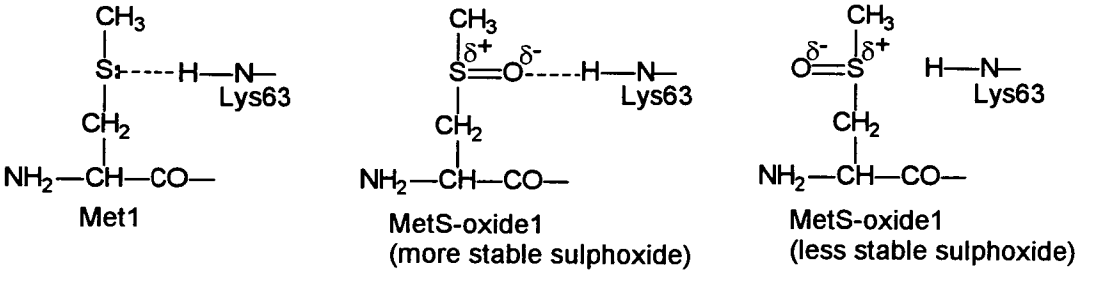


Figure. 2.26. Schematic representation of the hydrogen bond between Met1 and the two MetS-oxide1 isomers with the backbone amide of Lys63.

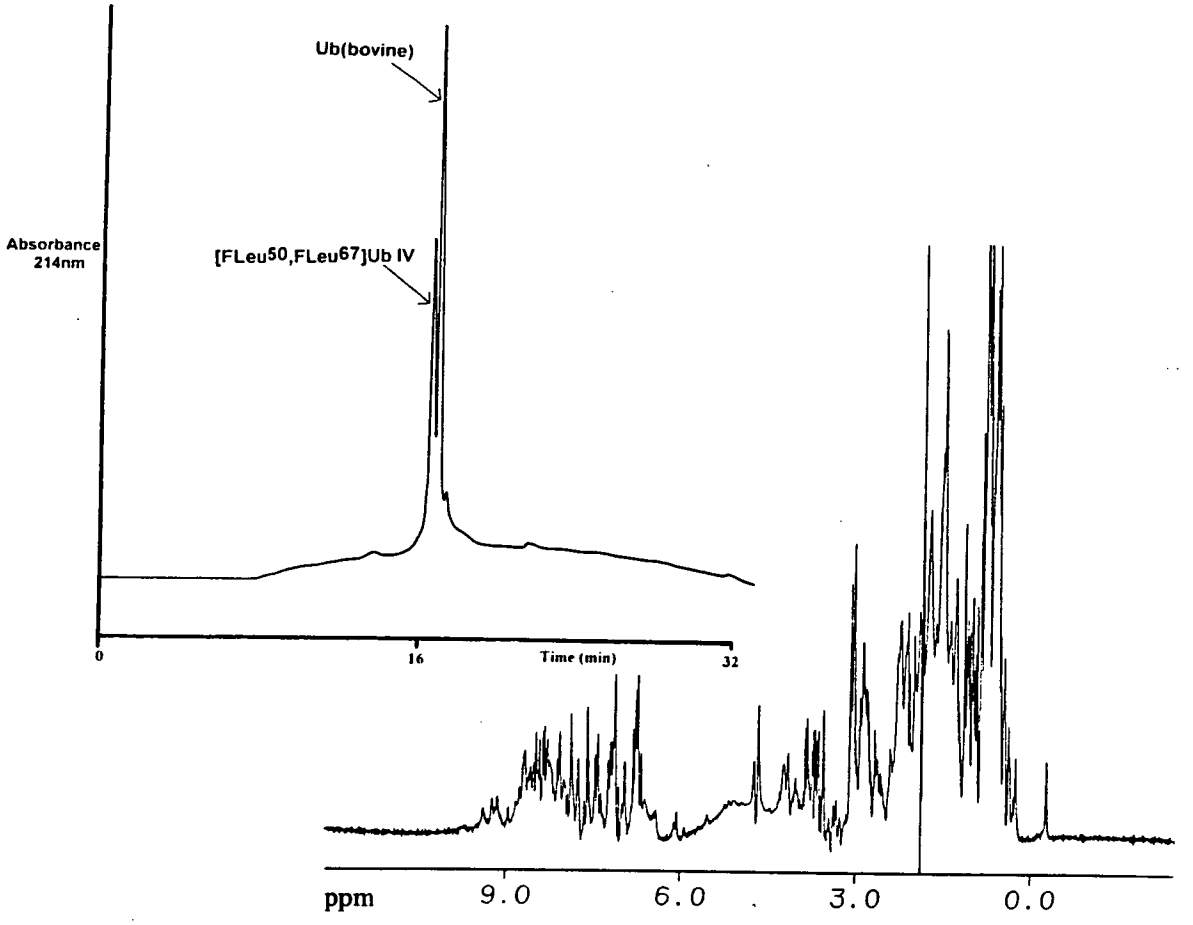


Figure 2.27. Analytical RP HPLC of [FLeu^{50},FLeu^{67}]Ub IV co-injection with Ub(bovine) (upper) and 1D ^{1}H NMR of [FLeu^{50},FLeu^{67}]Ub IV at 600 MHz, pH 4·8 and 25°C (lower).

The methionine sulphoxide was reduced to methionine by treatment with N-methylmercaptoacetamide in 10% aqueous acetic acid over 48 hours at 37°C. Initially, the reduction reaction was attempted using 50mM NH_4OAc pH 4·5 instead of 10% aqueous acetic acid but the reduction reaction was observed to be only negligible after 24 hours and therefore 10% aqueous acetic acid was used. Other reagents reported for the reduction of methionine sulphoxide include β-mercaptoethanol and dithiothreitol but these were not considered.[112] Cation exchange of the reduction reaction to maintain the folded state and lyophilisation gave [FLeu^{50},FLeu^{67}]Ub V. The analytical RP HPLC of this co-eluted with Ub(bovine) and the MALDI MS was as expected indicative of the successful reduction reaction (figure 2.28.).

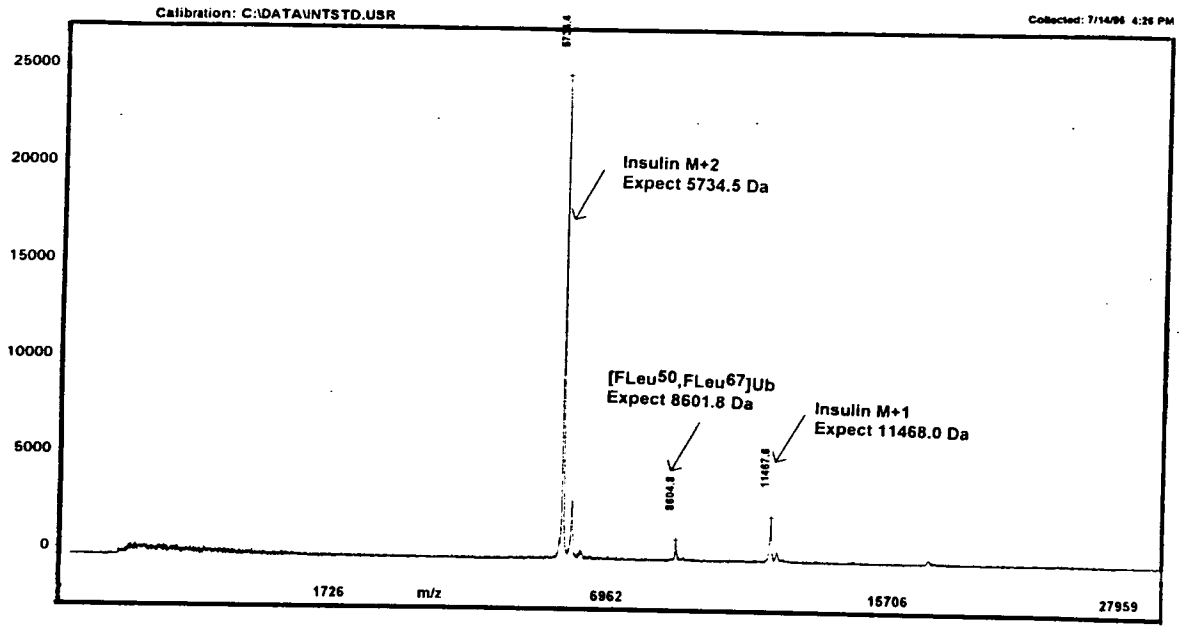


Figure 2.28. MALDI MS of reduced product [FLeu^{50},FLeu^{67}]Ub V.

Furthermore, the 1D ^{1}H NMR of [FLeu^{50},FLeu^{67}]Ub V was virtually indistinguishable from that of Ub(bovine) with both excellent definition of chemical shifts throughout, but importantly in the region 8-9ppm indicating formation of the native fold (figure 2.29.).

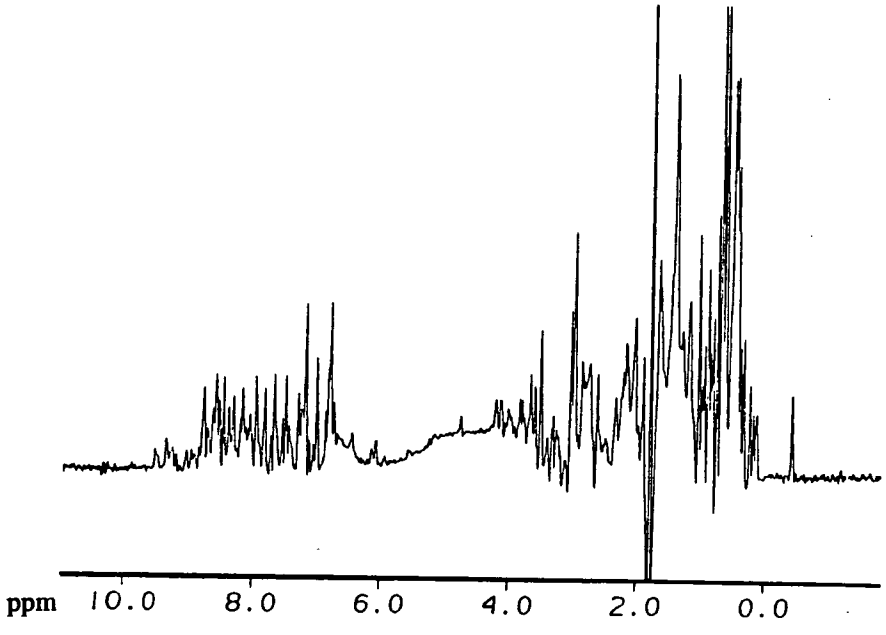


Figure 2.29. 1D ^{1}H NMR of [FLeu^{50},FLeu^{67}]Ub V at 600 MHz, pH 4·8 and 25°C.

The far UV circular dichroism results for [FLeu^{50},FLeu^{67}]Ub V are shown in figure 2.30.

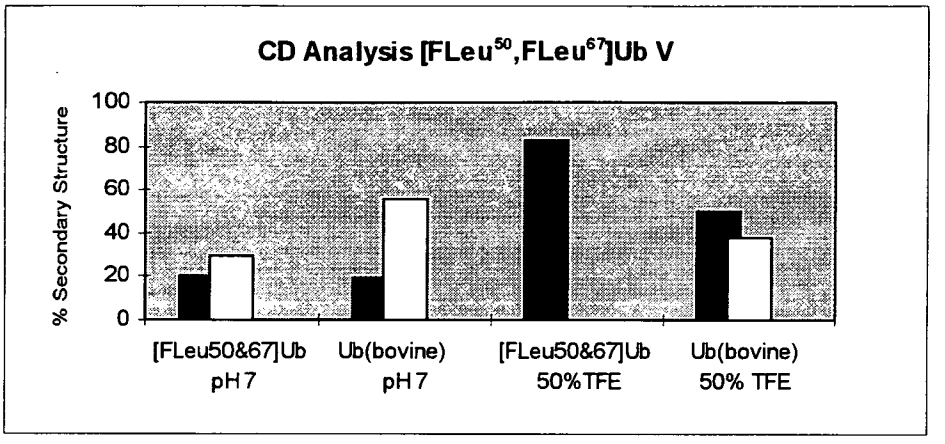


Figure 2.30. Deconvolutions of CD spectra into % secondary structure expressed as one of the following structures : heavily shaded bars, α-helix and clear bars, β-sheet.

At pH 7, good correlation is seen for the α-helicity content at ca 20% for both [FLeu^{50},FLeu^{67}]Ub V and Ub(bovine). However, the β-sheet content is much reduced from 56% in Ub(bovine) to only 30% in [FLeu^{50},FLeu^{67}]Ub V. Addition of 50% TFE results in a large promotion of α-helicity in [FLeu^{50},FLeu^{67}]Ub V with complete disappearance of β-sheet structure indicative of weaker tertiary interactions present in the β-sheet.

Section 2.5.3. [43&67-(2S,4S)-5-Fluoroleucine]ubiquitin.

The results for this analogue closely mirror those obtained for the [FLeu^{50},FLeu^{67}]Ub analogue previously discussed. Initial purification of half the resin bound product by the Tbfmoc/PGC protocol gave [FLeu^{43},FLeu^{67}]Ub I. From the analytical RP HPLC of the starting Tbfmoc protein solution several peaks elute between 68-78% aqueous CH_3CN indicating that even at this early stage more than one conformation is present (figure 2.31.). The MALDI MS and AAA were as required but again the analytical RP HPLC was broader compared to Ub(bovine) (figure 2.31.).

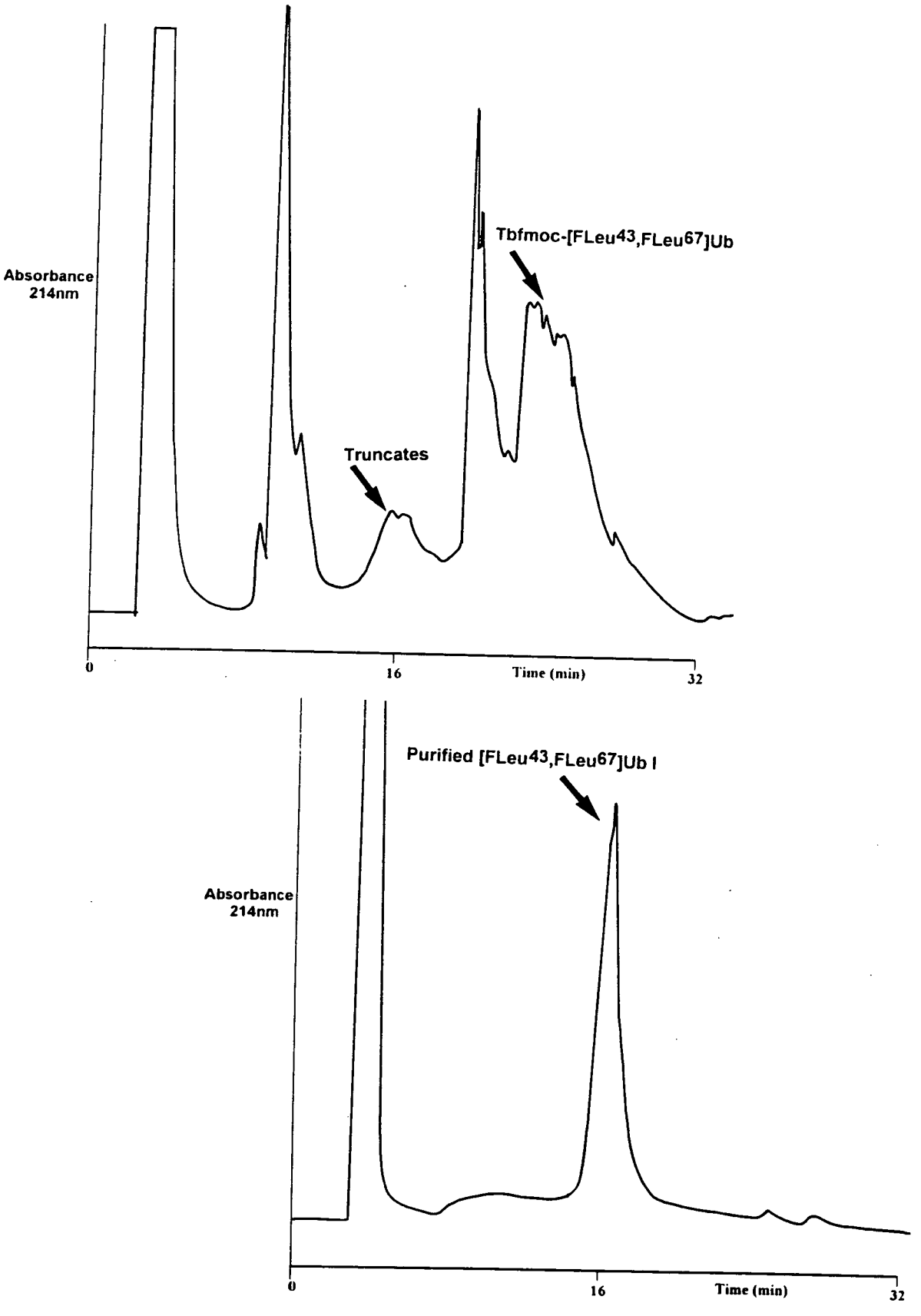


Figure 2.31. Analytical RP HPLC's of Tbfmoc-[FLeu^{43},FLeu^{67}]Ub (upper) and after purification on PGC and desalting to give [FLeu^{43},FLeu^{67}]Ub I (lower).

The 1D ^{1}H NMR of [FLeu^{43},FLeu^{67}]Ub I was essentially identical to that obtained for [FLeu^{50},FLeu^{67}]Ub I with both a reduced dispersion and definition of chemical shifts obtained indicating a non-native fold (figure 2.32.).

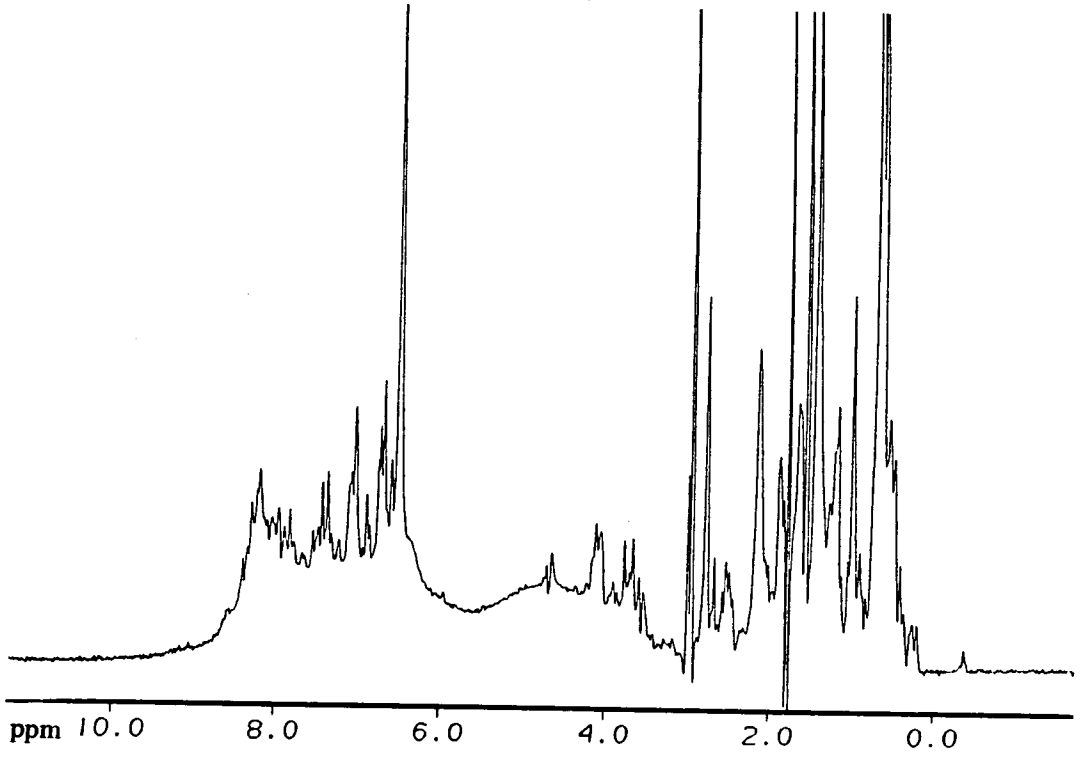


Figure 2.32. 1D ^{1}H NMR of [FLeu^{43},FLeu^{67}]Ub I at 600 MHz, pH 5·1 and 15°C.

Further purification of [FLeu^{43},FLeu^{67}]Ub I by FPLC Superdex 75 gave [FLeu^{43},FLeu^{67}]Ub II F1 & F2. The analytical RP HPLC's of these were broad, the AAA were as expected and again the MALDI MS were difficult to obtain.

The far UV circular dichroism results for [FLeu^{43},FLeu^{67}]Ub II F2 are shown in figure 2.33.

At pH 7, the α–helicity is much reduced to only 6% in [FLeu^{43},FLeu^{67}]Ub II F2 from 19% in Ub(bovine), while the β-sheet content is similar at 48% for [FLeu^{43},FLeu^{67}]Ub II F2 compared to 56% in Ub(bovine). Addition of 50% TFE to [FLeu^{43},FLeu^{67}]Ub

II F2 results in only a modest increase in the α-helicity to 19% with a corresponding decrease in the β-sheet to 35%.

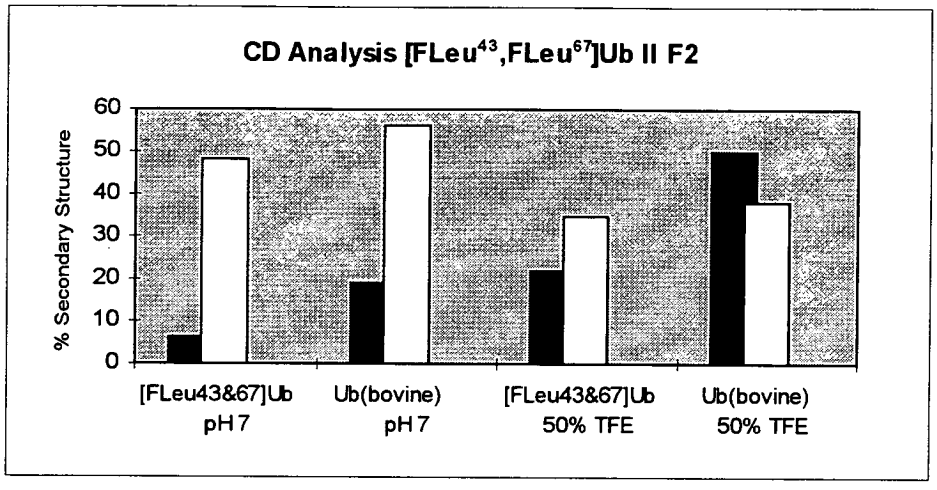


Figure 2.33. Deconvolutions of CD spectra into % secondary structure expressed as each of the following structures : heavily shaded bars, α-helix; clear bars, β-sheet.

Attempted refolding of [FLeu^{43},FLeu^{67}]Ub II F1/F2 by cation exchange chromatography gave increased recovery compared to [FLeu^{50},FLeu^{67}]Ub. Analytical RP HPLC of the third fraction, [FLeu^{43},FLeu^{67}]Ub III F3 from the cation exchange column was very sharp with two closely eluting peaks. Previous results would indicate that these correspond to the methionine and methionine sulphoxide of [FLeu^{43},FLeu^{67}]Ub. The MALDI MS gave 8607·6 and 8832·4Da corresponding to the methionine protein and matrix adduct. The 1D ^{1}H NMR of this material was essentially identical to that obtained for [FLeu^{50},FLeu^{67}]Ub IV purified by the non-Tbfmoc/PGC protocol, therefore cation exchange has successfully folded the material. Again, the 1D ^{1}H NMR was much sharper with a good dispersion of chemical shifts seen albeit with reduced definition in the region 8-9ppm indicating formation of a looser native fold.

The second half of the resin bound product was purified by a two step protocol using gel filtration and cation exchange chromatography giving [FLeu^{43},FLeu^{67}]Ub IV. This was confirmed as the methionine sulphoxide by both co-elution with [MetS-

oxide[1]]Ub(bovine) on analytical RP HPLC and the higher (+17·1Da) MALDI MS. The 1D ^{1}H NMR of this material was sharp with good dispersion albeit with reduced definition in the 8-9ppm region (figure 2.34.).

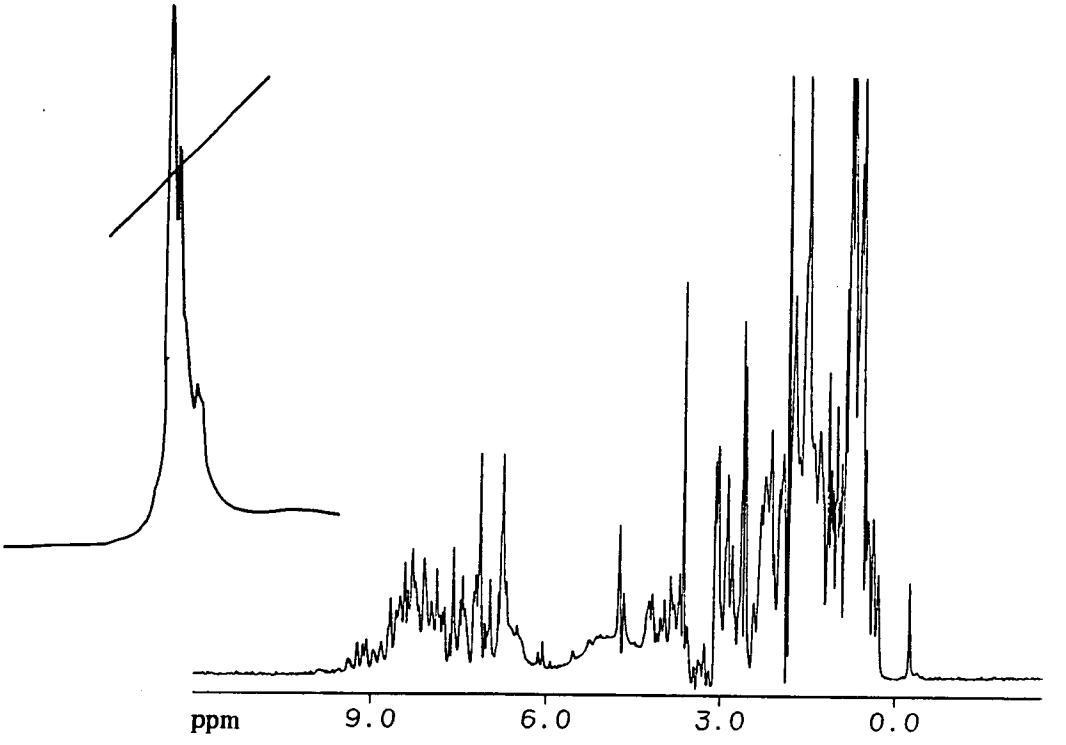


Figure 2.34. 1D ^{1}H NMR of [FLeu^{43},FLeu^{67}]Ub IV at 600 MHz, pH 4·8 and 25°C with inset of analytical RP HPLC.

Reduction of the methionine sulphoxide with N-methylmercaptoacetamide and cation exchange chromatography gave [FLeu^{43},FLeu^{67}]Ub V. However, the MALDI MS of this was higher (+20·4) than expected and has been assigned as sodium adduct and not the methionine sulphoxide since the product co-eluted with Ub(bovine). The 1D ^{1}H NMR indicates that the sample is weak but a reasonable dispersion of chemical shifts is seen.

The far UV circular dichroism results for [FLeu^{43},FLeu^{67}]Ub V are shown in figure 2.35. The results at pH 7 indicate an increased α-helicity to 29% for [FLeu^{43},FLeu^{67}]Ub V compared to 19% in Ub(bovine). In comparison the β-sheet content is much reduced from 56% in Ub(bovine) to 32% in [FLeu^{43},FLeu^{67}]Ub V.

Addition of 50% TFE results in 100% α-helicity indicative of a strong helix forming tendency perhaps reflecting a less rigid β-sheet.

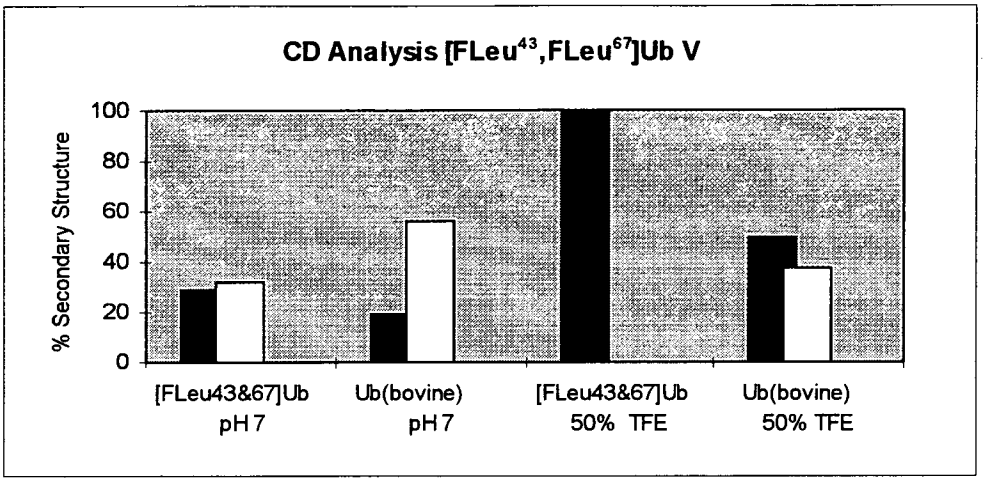


Figure 2.35. Deconvolutions of CD spectra into % secondary structure expressed as one of the following structures : heavily shaded bars, α-helix and clear bars, β-sheet.

Section 2.2.5. [67-(2S,4S)-5-Fluoroleucine]ubiquitin.

Structural analysis on this analogue used material purified by Dr A.R.Brown in a similar manner to that used in the non-Tbfmoc purification of [FLeu^{50},FLeu^{67}]Ub and [FLeu^{43},FLeu^{67}]Ub. The analytical RP HPLC of [FLeu^{67}]Ub co-eluted with Ub(bovine) although it was much broader than Ub(bovine) (figure 2.36.). The MALDI MS was more difficult to obtain giving MH^+= 8605·7Da which is 27Da higher than expected and has been assigned as the sodium adduct. The integrity of the methionine was confirmed by a trial N-methylmercaptoacetamide reduction in which no change in the analytical RP HPLC profile was seen after 48hrs incubation confirming no methionine sulphoxide was present.

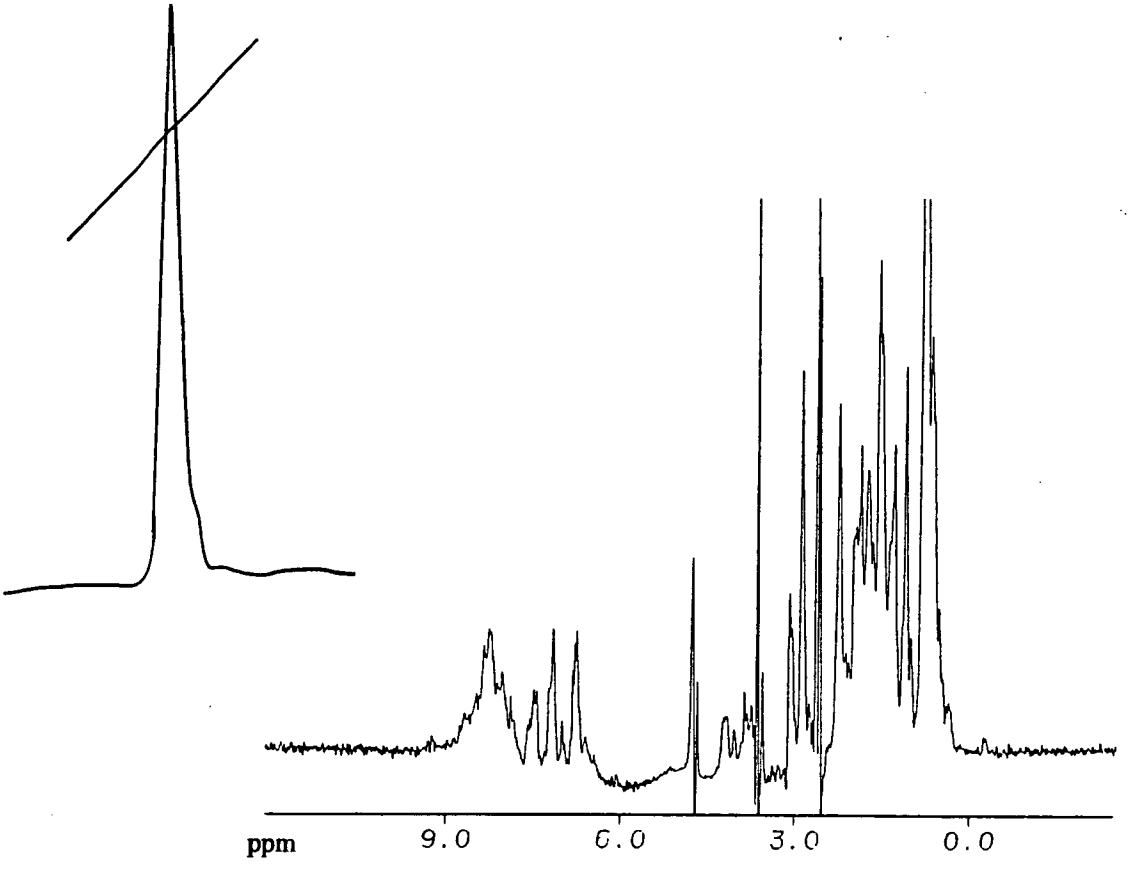


Figure 2.36. 1D ^{1}H NMR of [FLeu^{67}]Ub at 600 MHz, pH 4·8 and 25°C with inset of analytical RP HPLC.

The 1D ^{1}H NMR is very similar to that obtained for the Tbfmoc purified analogues i.e. a reduced definition of signal but the same overall shape (figure 2.36.). The chemical shift dispersion is slightly increased to 9ppm compared to 8ppm seen in the Tbfmoc purified analogues indicative of some tertiary structure formation.

The far UV circular dichroism analysis are shown in figure 2.37. Excellent correlation is seen for the β-sheet content in [FLeu^{67}]Ub and Ub(bovine) at both pH 7 and 50% TFE. In contrast the α–helicity is much reduced to only 8% at pH 7 compared to 19% in Ub(bovine) and exhibits only a modest helix forming tendency to 22% in 50% TFE.

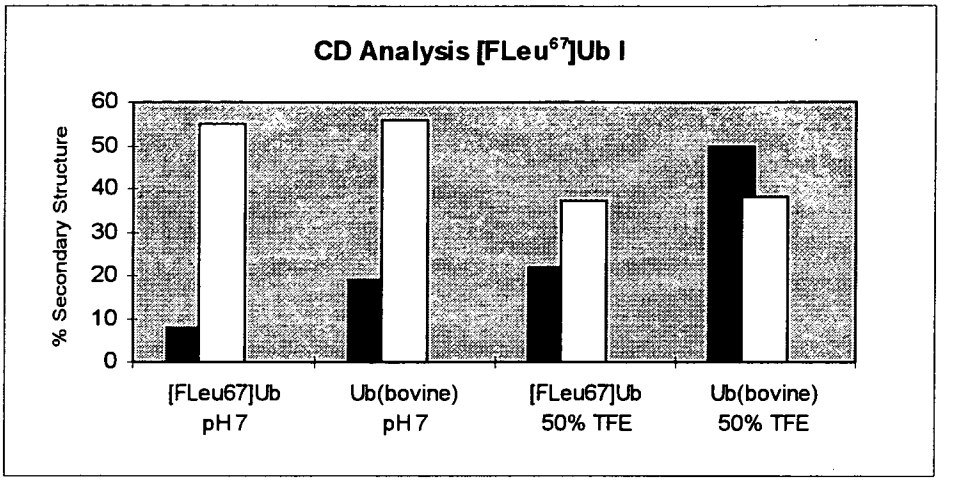


Figure 2.37. Deconvolutions of CD spectra into % secondary structure expressed as one of the following structures : heavily shaded bars, α-helix and clear bars, β-sheet.

2.5.5. Fluorine NMR.

The basis of this work on ubiquitin fluorinated analogues was to incorporate fluorine within the primary sequence and assess the effect of this substitution on the tertiary structure. The difficulties encountered in folding these analogues has allowed us to map out the folding with both ^{1}H NMR and ^{19}F NMR studies starting from a non-native tertiary fold to a native fold.

Table 2.2. Summary of ^{19}F NMR data for ubiquitin fluorinated analogues with tertiary fold as indicated by 1D ^{1}H NMR also shown: data recorded at 600MHz, 25°C and referenced to $CFCl_3$ assigned to zero.

	^{1}H	^{19}F 50*	^{19}F	^{19}F 67*	^{19}F 56*	^{19}F 43*
FLeu amino acid*	NA		-223·7			
[FLeu^{56},FLeu^{67}]Ub I	non-native		-221·5	-221·8	-223·2	
[FLeu^{50},FLeu^{67}]Ub I	non-native		-221·3			
[FLeu^{50},FLeu^{67}]Ub IV	native-like	-217·4	-221·3	-222·1		
[FLeu^{50},FLeu^{67}]Ub V	native	-217·1		-221·0		
[FLeu^{43},FLeu^{67}]Ub I	non-native		-221·4			
[FLeu^{43},FLeu^{67}]Ub IV	native-like		-221·7	-222·1		-226·8 -226·4
[FLeu^{43},FLeu^{67}]Ub V	native			-222·0		-226·3
[FLeu^{67}]Ub	non-native		-221·5	-221·9		

The ^{19}F NMR results clearly illustrate the large variation in chemical shifts of ca 10ppm between both the different analogues and different folded states of the analogues, and correlates well with the expected dispersion of 5-15ppm for a typical fluorinated amino acid incorporated into a protein.[113] This is expected since a FLeu

residue which is experiencing a different environment as a result of folding will resonate at a different chemical shift.

In the non-native structures a degenerate signal is present at -221·3 to -221·7ppm which can be assigned to FLeu residue in an unfolded environment. The results of [FLeu^{56},FLeu^{67}]Ub, [FLeu^{43},FLeu^{67}]Ub, and [FLeu^{67}]Ub would indicate that FLeu67 moves to the lower frequency of ca -222·0ppm in a folded environment although this is not seen in [FLeu^{50},FLeu^{67}]Ub whereby it moves to the higher frequency of -221·0ppm. Therefore, it appears that the folded chemical shift frequency of FLeu67 is similar to the frequency of the FLeu residues in a non-native fold.

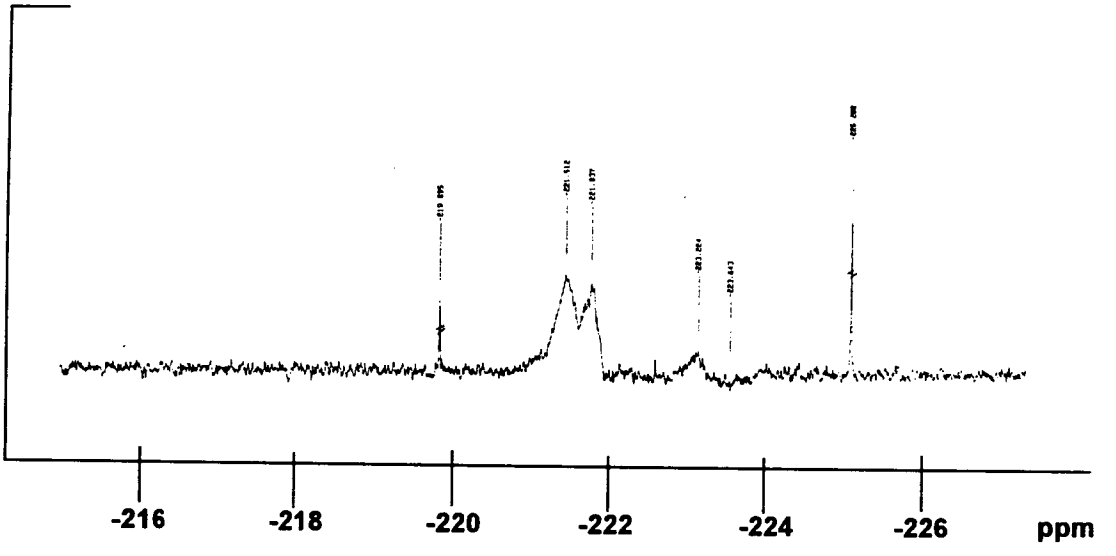


Figure 2.38. ^{19}F NMR spectra of [FLeu^{56},FLeu^{67}]Ub I.

The three signals from the non-native [FLeu^{56},FLeu^{67}]Ub I (figure 2.38.) suggest limited folding in a proportion of protein molecules resulting in the extra signals at the lower frequencies of -221·8ppm assigned to FLeu67 in a folded environment and at -223·2ppm assigned to FLeu56.

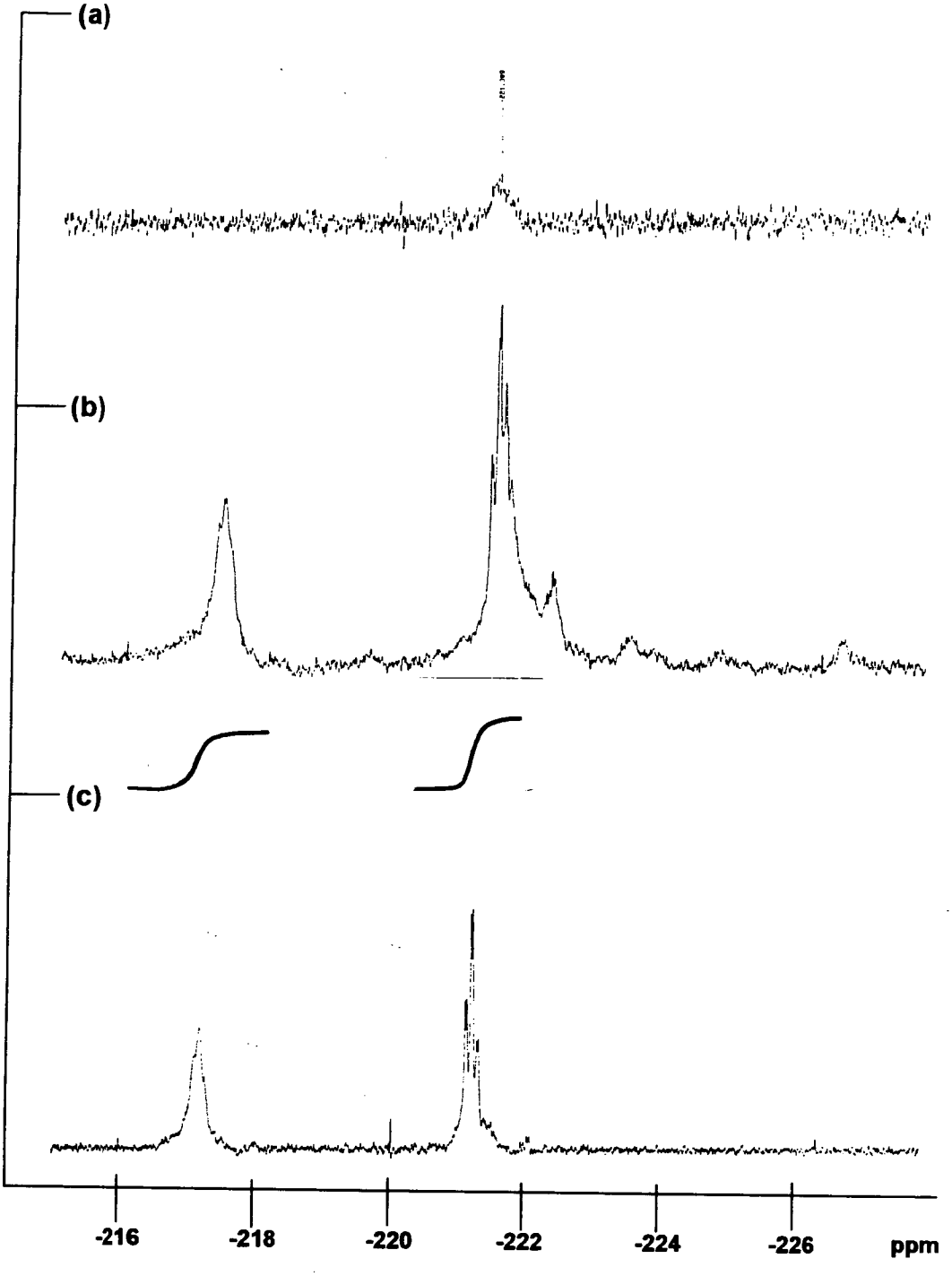


Figure 2.39. ^{19}F NMR spectra of (a) [FLeu^{50},FLeu^{67}]Ub I, (b) [FLeu^{50},FLeu^{67}]Ub IV and (c) [FLeu^{50},FLeu^{67}]Ub V.

From the results of [FLeu^{50},FLeu^{67}]Ub I (figure 2.39a.) only one signal is obtained at -221·3ppm from the non-native state signifying the same environment for both FLeu residues. However, in the native-like [FLeu^{50},FLeu^{67}]Ub IV (figure 2.39b.) which was assigned as the MetS-oxide1, two main signals at -217·4 & -221·3ppm and a

smaller signal at -222·1ppm are observed and reflects the less compact state present. Formation of the native fold by reduction of the methionine to give [FLeu^{50},FLeu^{67}]Ub V (figure 2.39c.) results in the expected two signals at -221·0ppm assigned to FLeu67 and at the higher frequency of -217·1ppm assigned to FLeu50.

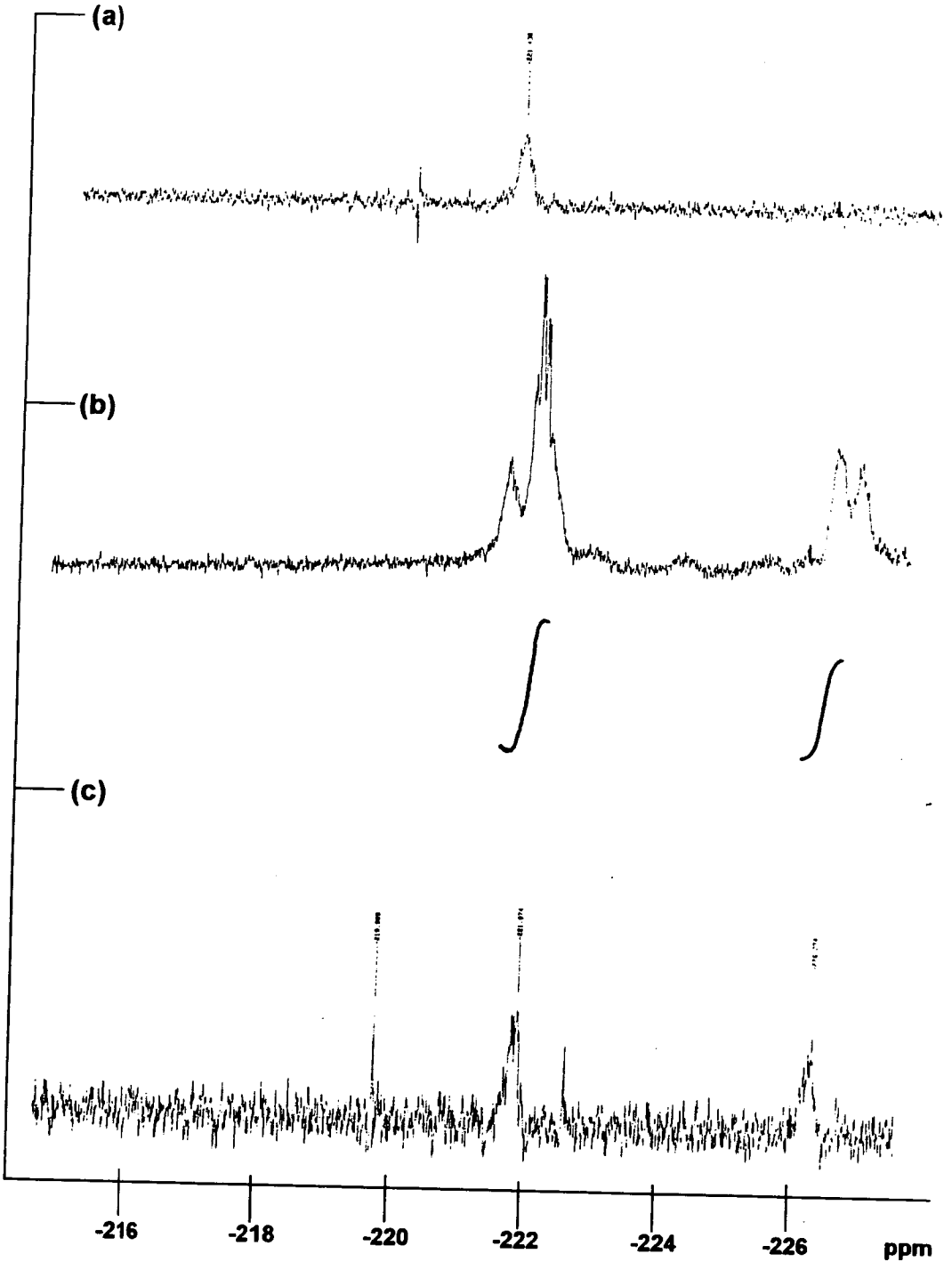


Figure 2.40. ^{19}F NMR spectra of (a) [FLeu^{43},FLeu^{67}]Ub I, (b) [FLeu^{43},FLeu^{67}]Ub IV and (c) [FLeu^{43},FLeu^{67}]Ub V.

The non-native fold of [FLeu^{43},FLeu^{67}]Ub I (figure 2.40a.) gives only one signal at -221·4ppm. In the native-like [FLeu^{43},FLeu^{67}]Ub IV (figure 2.40b.) assigned as the MetS-oxidel, one main signal at -222·1ppm and three smaller signals at -221·7, -226·4 and -226·8ppm are present. Following reduction of the methionine to give [FLeu^{43},FLeu^{67}]Ub V (figure 2.40c.) we obtain the expected two signals, at -222·0 assigned to FLeu67 and at the lower frequency of -226·3ppm assigned to FLeu43.

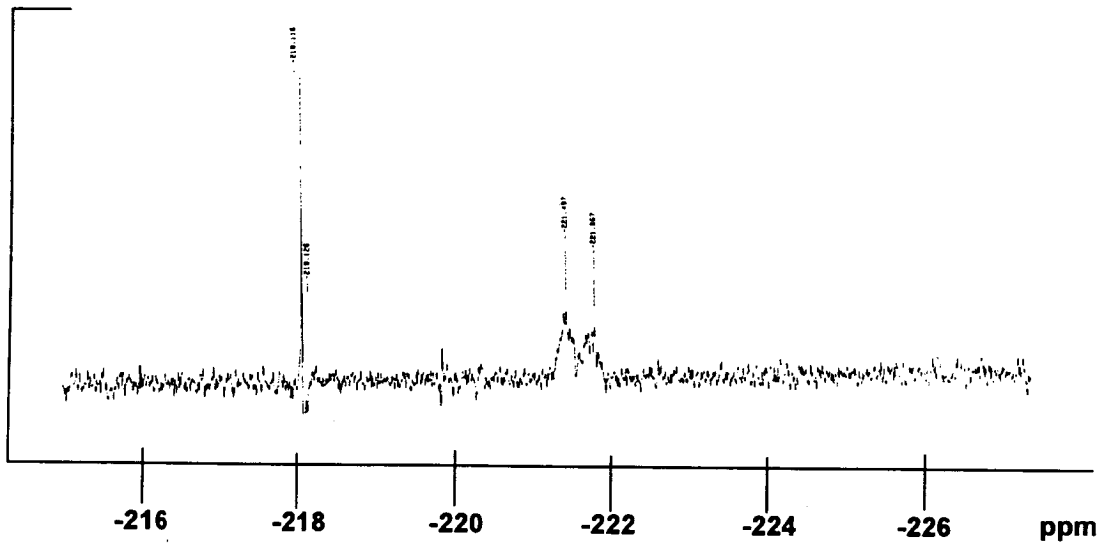


Figure 2.41. ^{19}F NMR spectra of [FLeu^{67}]Ub.

The results of the mono-substituted [FLeu^{67}]Ub (figure 2.41.) allow assessment of the stabilising or otherwise influence of two fluorines compared to one fluorine in the hydrophobic core, and also identification of the chemical shift due to the FLeu67 residue. Two signals are seen at -221·5ppm and -221·9ppm indicating the presence of more than one conformation in which the environment of the FLeu residue is only marginally different.

It is interesting to note the shape of the ^{19}F NMR peaks, all of which are broad with no detectable splitting pattern except for the native-like & native folds of [FLeu^{50},FLeu^{67}]Ub IV & V. The splitting pattern in the latter shows the expected triplet of doublets observed in the free amino acid with coupling constants of 14·0 and 47·6 Hz compared to 20·8 and 47·6 Hz in the amino acid.

It is important to identify the environment experienced by each FLeu residue and hence explain the distribution of chemical shifts and further NMR analysis is continuing on the structure determination of these ubiquitin fluorinated analogues. The actual origins of these tertiary structure induced fluorine chemical shifts remain to be investigated further but are known to include the anisotropic shielding from the peptide bond system & aromatic rings, Van der Waals interactions and hydrogen bonding.[113]

Section 2.6. [3-Norleucine,43-norvaline]ubiquitin.

A nonhierarchical model of protein folding has been proposed by Dill in which the driving force for protein folding is the hydrophobicity of the amino acid sequence.[1] The resultant hydrophobic collapse gives compact conformations which direct the subsequent formation of secondary and tertiary structure. This is in contrast to the framework or hierarchical model of folding in which the early formation of secondary structure guides the progressive formation of further secondary and tertiary structure.[114]

The importance of hydrophobicity as the main driving force for protein folding was noted by Kauzmann in 1959[1] although there is some confusion over the actual meaning of the term hydrophobicity.[115] At least three different meanings have been used:

- transfer of nonpolar solutes into an aqueous solution
- transfer of nonpolar solutes into an aqueous solution with a particular temperature dependence
- ordering of water molecules around nonpolar solutes (molecular model)

The pictorial representation of the core of a protein is shown in figure 2.42.

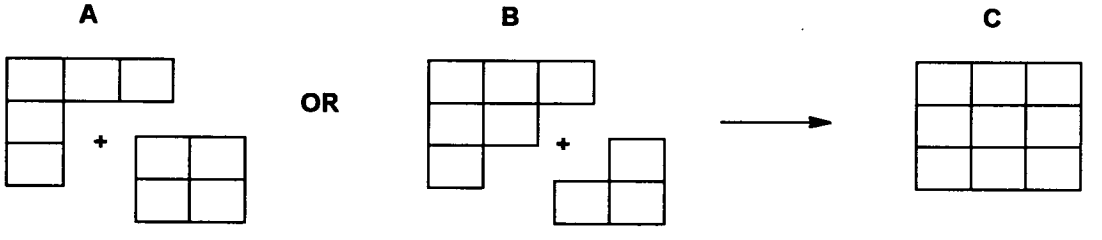


Figure 2.42. Structures A and B nucleate to form the 3x3 matrix C.

This somewhat simplified diagram of the hydrophobic core illustrates that there are many ways, of which two are shown in forming a 3x3 matrix and similarly the same argument applies to the core of a protein.

We were interested in maintaining the overall volume of the hydrophobic core while at the same time modifying the distribution of alkyl groups around the core. The analogue chosen for synthesis was [Nle^{3},Nva^{43}]Ub in which the linear norleucine and norvaline replace isoleucine and leucine respectively (figures 2.43. & 2.44.). The effective result is the transfer of one methyl group across the hydrophobic core and due to the commercial availability of amino acids we chose to remove the side-chain methyl on the β–carbon of residue 3. However, removal of this methyl group is a surface change and would be expected to be less destabilising than a core deletion.

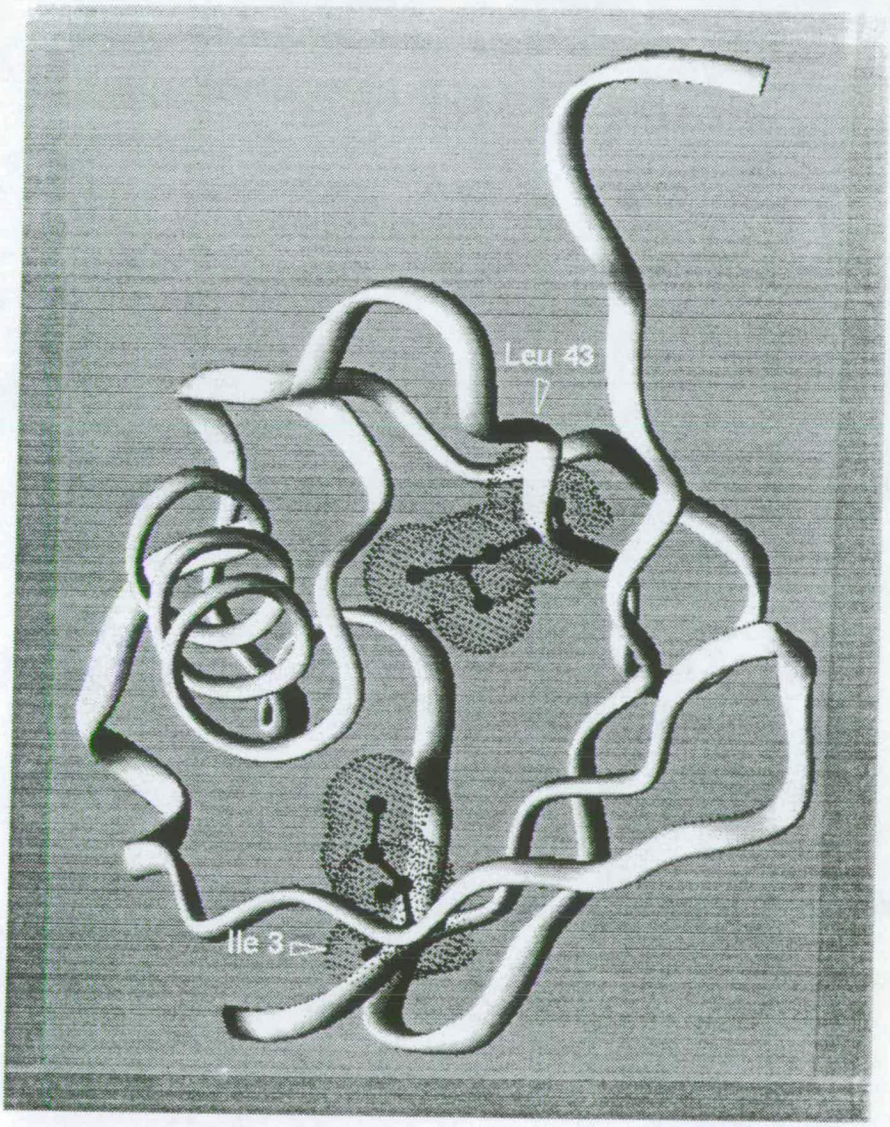


Figure 2.43. Ribbon diagram of ubiquitin with Ile3 & Leu43 shown: produced using Insight II program.

Residues 3 & 43 are not in contact with each other with the distances between the δ1 & δ2 carbons of Leu43 to the δ carbon of Ile3 measured at 7·27 and 7·37A° respectively.

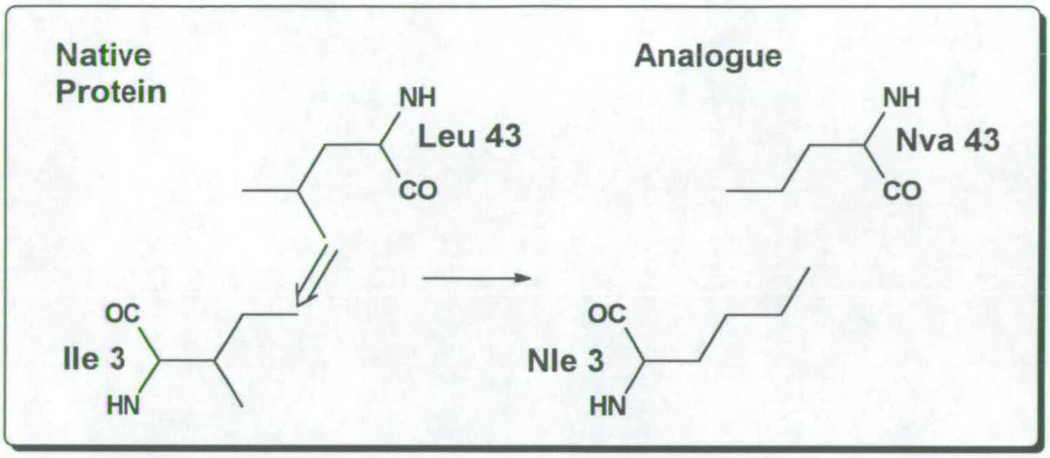


Figure 2.44. Representation of substitutions within the core of ubiquitin.

This means that the modification results in the redistribution of the internal space across the core. Surprisingly, although ubiquitin is a highly compact molecule there are several gaps within the core and these could be modified in future ubiquitin model studies (figure 2.45.).[116]

Ile-3	**changed to**	$-C_\alpha-C_\beta(-C_{\gamma2})-C_\gamma-C_\delta-C_\varepsilon$
Ile-23	**changed to**	$-C_\alpha-C_\beta-C_\gamma-C_\delta-C_\varepsilon-C_\zeta$
Leu-50 & Leu-67	**changed to**	$-C_\alpha-C_\beta(-C_{\delta1}-C_\varepsilon)(-C_{\delta2})-C_\gamma$
Ile-61	**changed to**	$-C_\alpha-C_\beta(-C_{\gamma2}-C_{\delta2})-C_\gamma-C_\delta-C_\varepsilon$

Figure 2.45. Possible modifications within the hydrophobic core of ubiquitin as suggested by Alexeev.[116]

The suggestions noted by Aleexev involve α-amino acids that are not commercially available and therefore a synthetic protocol would need to be developed to allow easy access of these novel α–amino acids. Recently, the synthesis of the suggested Ile3 modification to (2S,3S)-2-amino-3-methylhexanoic acid has been achieved by Professor Young & co-workers at the University of Sussex.[117]

Hydrophobic core modifications such as the one described above are introducing the possibility of separating the parameters of thermodynamics and kinetics. For example, maintenance of the folded hydrophobic core but with a redistribution of the alkyl side-chains around the core would result in the same overall thermodynamics of the folded protein but the kinetics in obtaining the native fold from the unfolded state may be altered.

The synthesis of [Nle3,Nva43]Ub gave excellent yields of 67% for the N-terminal methionine on the resin (figure 2.46.).

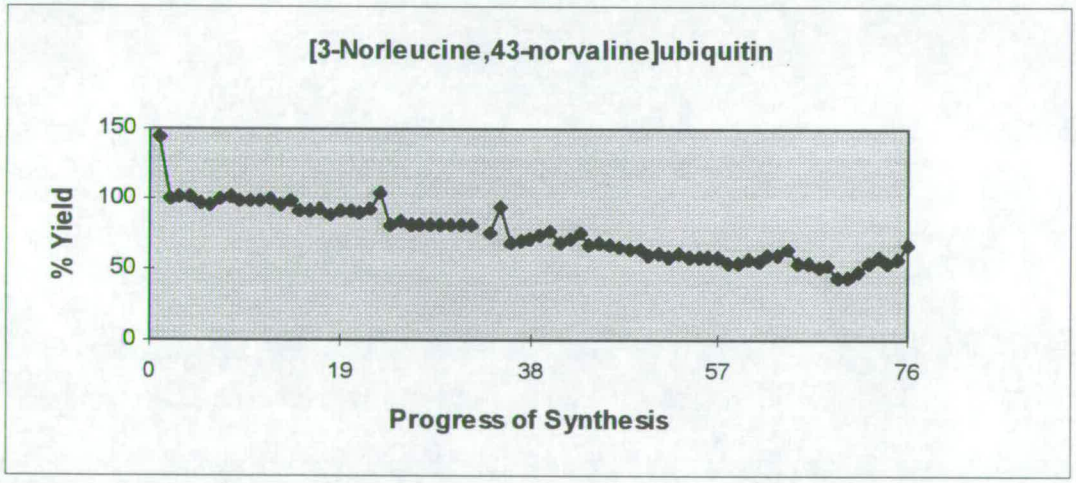


Figure 2.46. Progress of [Nle^{3},Nva^{43}]Ub synthesis as monitored by UV at 302nm of the deprotection solution containing the piperidine-fulvene adduct.

Purification was straightforward using the standard Tbfmoc/PGC protocol, desalting by gel filtration and semi-preparative RP HPLC gave [Nle^{3},Nva^{43}]Ub II in good

yields. The purity of the product was confirmed by AAA and MALDI MS and furthermore it was noted that the analytical RP HPLC was relatively sharp (figure 2.47.). Before starting structural analysis [Nle^{3},Nva^{43}]Ub II was dissolved in 8M urea and dialysed against decreasing concentrations of urea to allow refolding from the denatured state.

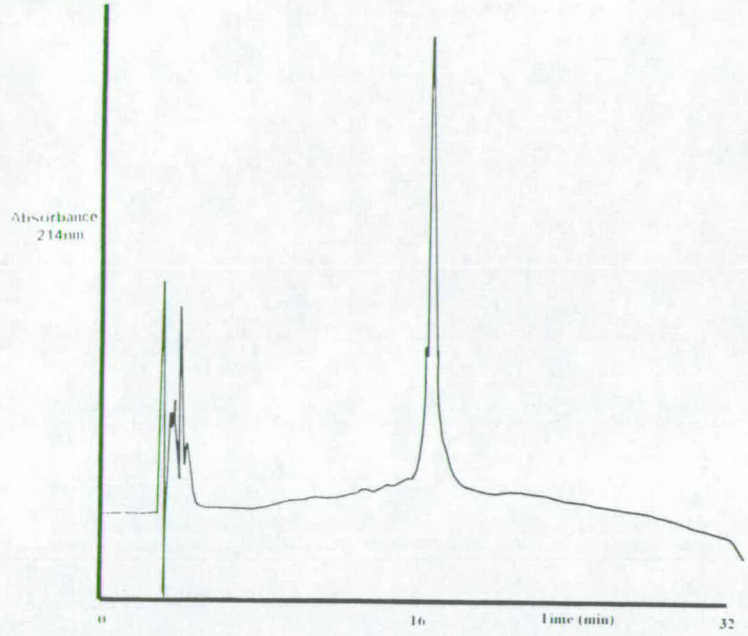


Figure 2.47. Analytical RP HPLC of [Nle^{3},Nva^{43}]Ub II.

The results of the far UV circular dichroism analysis are shown in figure 2.48.

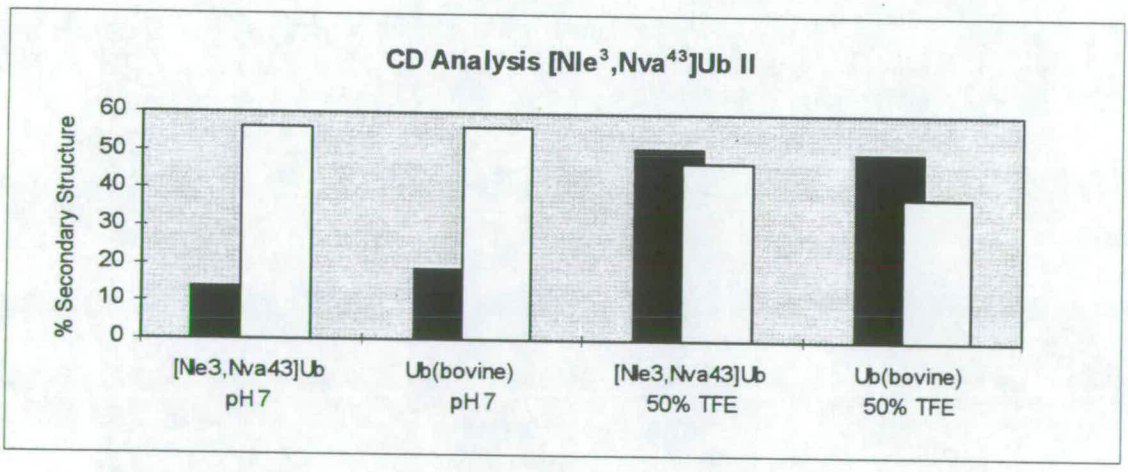


Figure 2.48. Deconvolutions of CD spectra into % secondary structure expressed as each of the following structures : heavily shaded bars, α-helix and clear bars, β-sheet.

Excellent correlation are seen for the results at pH 7 and 50% TFE. At pH 7, the α-helicity is marginally reduced to 14% in [Nle3,Nva43]Ub II compared to 19% in Ub(bovine). Addition of 50% TFE increases the α–helicity to the same extent in both but with increased maintenance of β-sheet structure in [Nle3,Nva43]Ub II compared to Ub(bovine). These results strongly suggest that the overall tertiary structure in [Nle3,Nva43]Ub II is similar to that in Ub(bovine).

With these encouraging results the fluorescence of the single Tyr59 was investigated using fluorescence studies. However, no unfolding curve was observed with only a general increase in fluorescence noted (figure 2.49.).

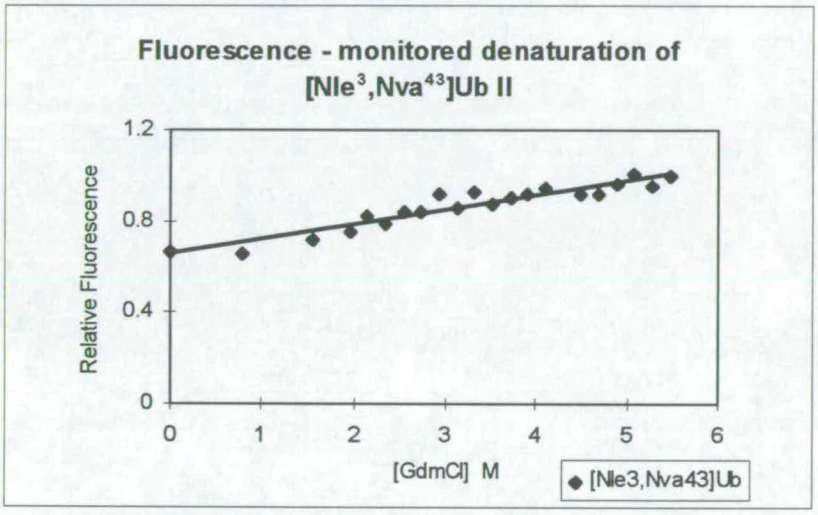


Figure 2.49. Plot of relative fluorescence of [Nle^{3},Nva^{43}]Ub II versus [GdmCl] .

Since the Tyr residue is located on the surface loop (51-59) it could flip out slightly resulting in no unfolding curve but still maintaining the overall native fold. For example, studies by Ecker *et al* on the mutant [Phe^{45}]Ub in which the Tyr is substituted with Phe gave a similar 2D NOESY spectrum to that of WT ubiquitin and was 70-100% biologically active as determined by *in vitro* protein degradation assays.[118] This NMR similarity in the presence of biological activity suggests that the hydrogen bond between the side-chain of Tyr59 and the backbone amide of Glu51 is not crucial for attainment of the native fold. However, in contradiction to a similar

native fold for the mutant [Phe[45]]Ub Ecker *et al* did note that it was only sparingly soluble at less than 1mg/ml which is characteristic for a misfolded protein.

Initial 1D ^{1}H NMR studies on [Nle3,Nva43]Ub II indicated a reasonable dispersion of chemical shifts up to ca 9ppm but with slightly reduced definition. However, a much improved NMR spectrum was obtained after the sample was lyophilised and redissolved in the NMR buffer at pH 4·8. Part of the 2D NOESY spectrum is shown below and this is presently being assigned.[119]

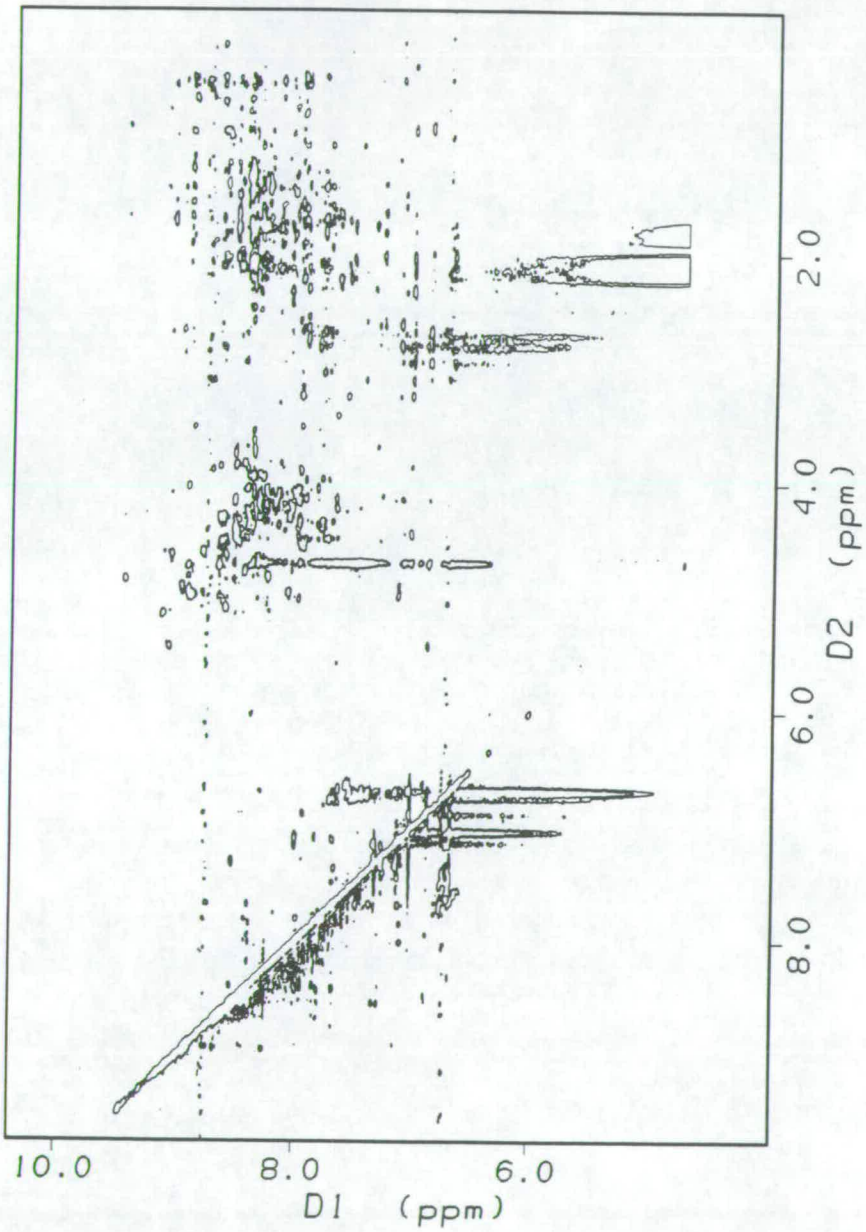


Figure 2.50. Region from the 2D NOESY spectrum of [Nle3,Nva43]Ub II at 600 MHz, pH 4·8 and 25°C.

Section 2.7. [68-DHistidine]ubiquitin.

The synthesis of [DHis^{68}]Ub was performed in order to investigate the importance of the chirality of the single histidine at position 68. This involves a surface change on the central strand of β-sheet in which the non-natural D-histidine was substituted for L-histidine (figure 2.51.).

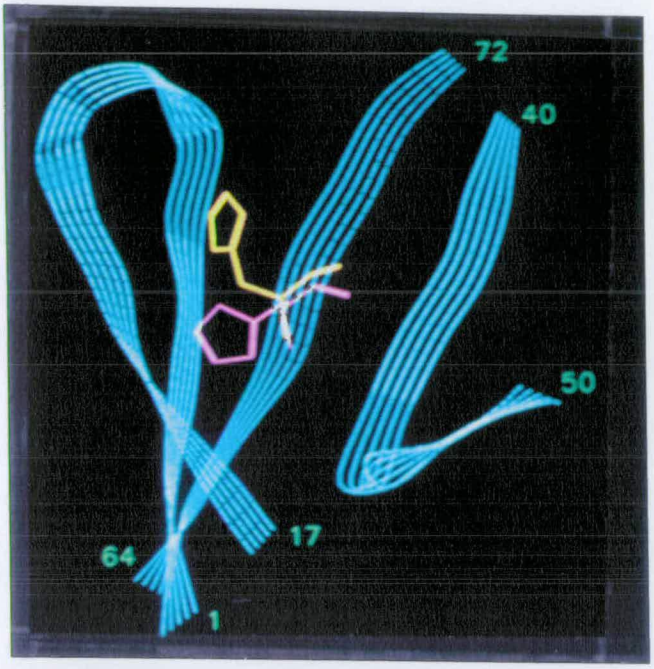


Figure 2.51. Ribbon diagram of β-sheet structure of ubiquitin with L-histidine shown in yellow and D-histidine shown in pink: produced using Sybil molecular modelling program.

No problems were encountered in the synthesis with a final yield of 58% for the N-terminal methionine on the resin (figure 2.52.).

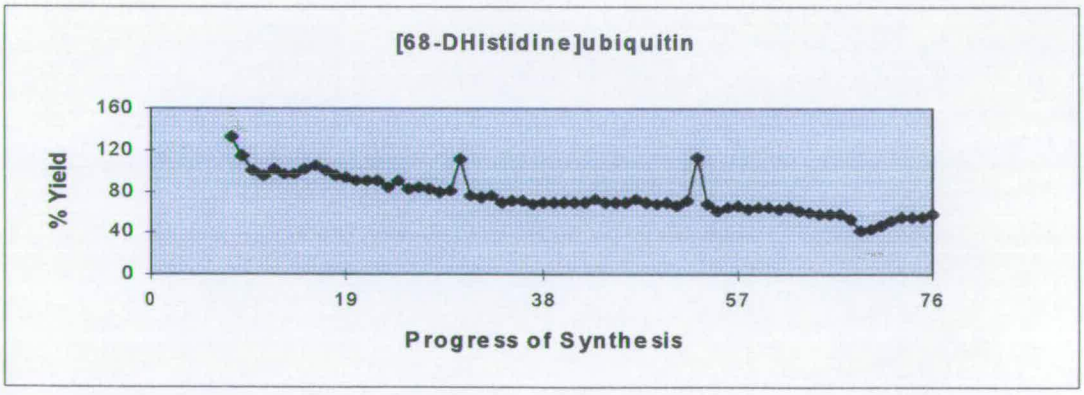


Figure 2.52. Progress of [DHis^{68}]Ub synthesis as monitored by UV at 302nm of the deprotection solution containing the piperidine-fulvene adduct.

A timed cleavage on a small quantity of resin bound product was performed in order to optimise the hydrolysis time. The results (figure 2.53.) clearly indicated that 4 hours is sufficient for the hydrolysis of the peptide resin-linkage and removal of side-chain protection.

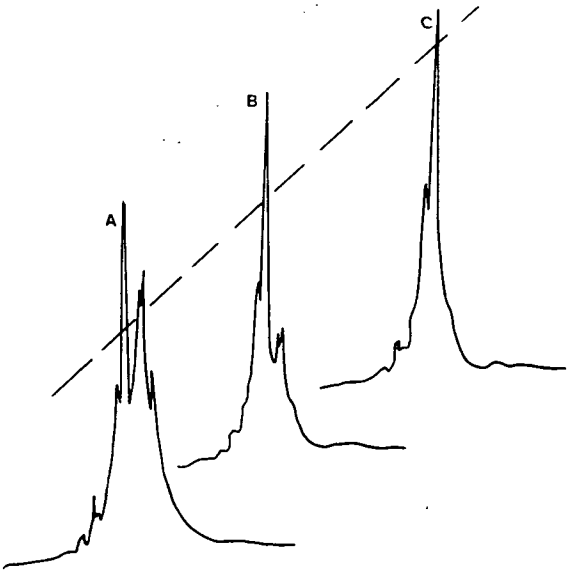


Figure 2.53. Analytical RP HPLC's for the timed hydrolysis of resin bound [DHis^{68}]Ub at (A) 2 hrs, (B) 3 hrs and (C) 4 hrs.

Two step purification using Tbfmoc/PGC and desalting by gel filtration gave [DHis^{68}]Ub I in good yields. The purity of the product was confirmed by AAA and MALDI MS (figure 2.54.) and it was noted that the analytical RP HPLC of [DHis^{68}]Ub I was relatively sharp. Prior to structural analysis [DHis^{68}]Ub I was dissolved in 8M urea and dialysed against decreasing concentrations of urea to permit refolding from the denatured state.

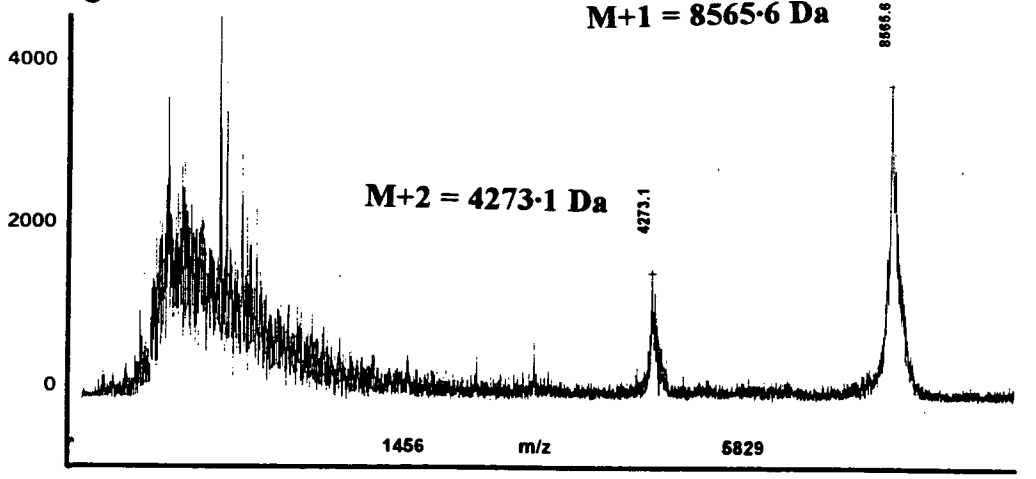


Figure 2.54. MALDI MS of [DHis^{68}]Ub I, M+1 requires 8565·6 Da.

The results for the far UV circular dichroism analysis are shown in figure 2.55.

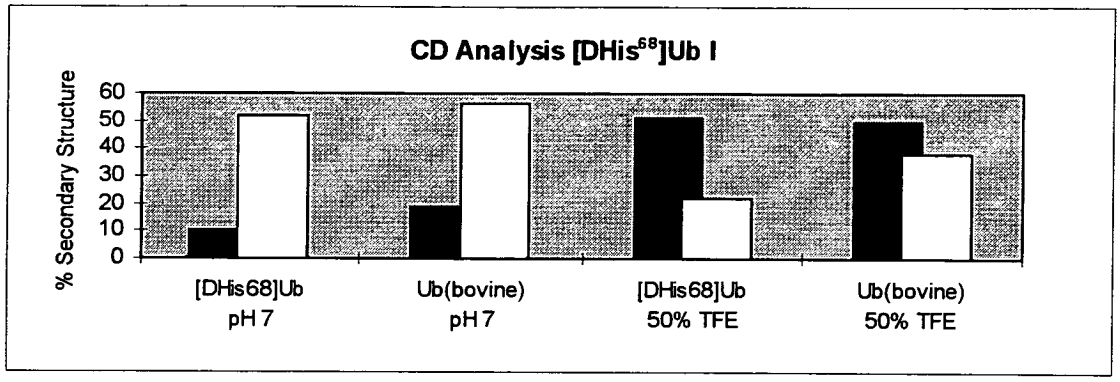


Figure 2.55. Deconvolutions of CD spectra into % secondary structure expressed as each of the following structures : heavily shaded bars, α-helix and clear bars, β-sheet.

The circular dichroism results at pH 7 indicate the α-helicity is much reduced to 10% in [DHis^{68}]Ub I compared to 19% in Ub(bovine). However, good correlation is observed for the β-sheet content at 50 and 52% for [DHis^{68}]Ub I and Ub(bovine) respectively. Addition of 50% TFE results in promotion of α-helicity to ca 50% in both cases with an increased reduction in β-sheet for [DHis^{68}]Ub I.

Fluorescence studies on Tyr59 indicated that it is already exposed to the solvent even in non-denaturing conditions with only a slight increase in fluorescence observed with increasing concentrations of guanidinium chloride (figure 2.56.).

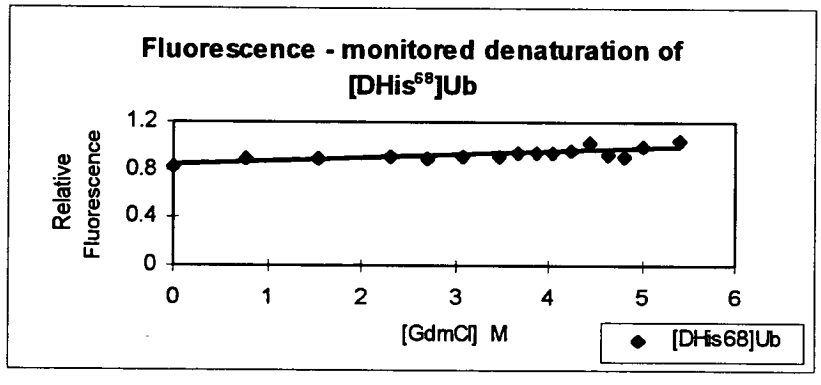


Figure 2.56. Plot of relative fluorescence of [DHis^{68}]Ub I versus [GdmCl].

The 1D ^{1}H NMR spectrum of [DHis^{68}]Ub I shows both limited chemical shift dispersion and a much reduced definition of signal, which is indicative of a non-native fold. As seen in the 2D NOESY spectrum (figure 2.57.), there is a much reduced dispersion of chemical shifts which is characteristic of a random coil structure.

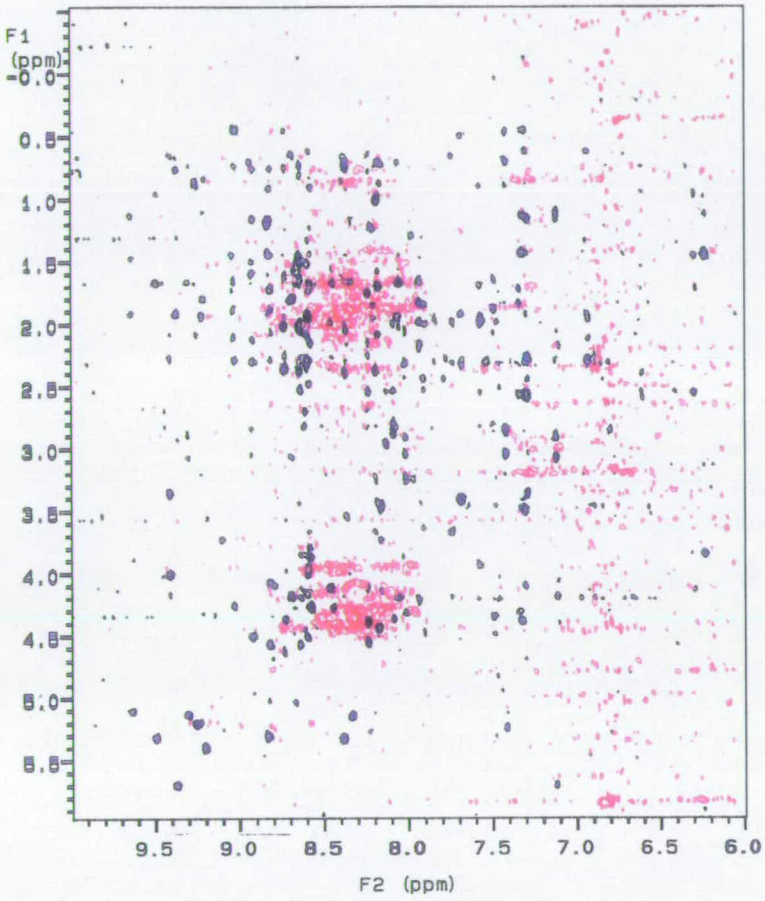


Figure 2.57. Region from the 2D NOSEY spectra of [DHis^{68}]Ub I shown in red and of Ub(bovine) shown in black , both recorded at 600 MHz, pH 4·8 and 25°C.

Although this analogue has been purified by the Tbfmoc/PGC protocol which is known to induce non-native folding in native ubiquitin, this is clearly not always the case as seen for [Nle^{3},Nva^{43}]Ub in which a native-like fold is formed. Therefore, it seems probable that this is a true denatured state rather than a Tbfmoc induced non-native fold. It could be postulated that the replacement with D-histidine would result in the twisting of final strand of β–sheet disrupting the formation of the hydrophobic core and thus resulting in the non-native fold.

2.8. Biological Activity.

The biological activity of synthetic ubiquitin and analogues is presently being determined by Professor R.J. Mayor and Maureen Mee at the University of Nottingham. The assay involves measuring the conjugation of ubiquitin radiolabelled with ^{125}iodine to rabbit reticulocytes (fraction II) in the presence and absence of ATP.[120]

The results from the assay will allow assessment of the tertiary structure requirements for the biological activity. In particular, the different folds of synthetic ubiquitin will permit a direct comparison between biological activity and protein fold. For example, in the study of disulphide engineered ubiquitin mutants it was found that the activity of the conformationally restricted mutant [Cys^{4},Cys^{66}]Ub was much reduced compared to the fully active [Cys^{4},Cys^{14}]Ub mutant, indicating that conformational mobility is important for the activity of ubiquitin.[121]

2.9. Summary.

Solid phase peptide synthesis allows for the facile incorporation of unnatural and novel synthetic α-amino acids into the primary sequence of proteins, as demonstrated for the ubiquitin analogues discussed. Continued advances in SPPS methodology, in particular more effective coupling procedures result in improved chain assembly which provides the opportunity for the synthesis of larger proteins. Alternatively, access to multi-domain proteins can be gained by using the chemical ligation of unprotected peptide fragments.

A major challenge for protein chemists is the purification of the correct sequence from the unwanted truncates which have accumulated during peptide synthesis. This task is simplified by utilising the hydrophobic properties of the Tbfmoc group for purification of the required peptide from the deletions. However, our results for synthetic ubiquitin demonstrate that this approach may in some instances result in non-native

folding and that optimisation of subsequent folding procedures are required to gain access to the native state.

Synthesis of the ubiquitin analogues incorporating (2S,4S)-5-fluoroleucine have opened up the possibility of using fluorine NMR as a specific 'tag' for characterising folding intermediates without the complexity of proton NMR which often gives poor dispersion of chemical shifts for these largely unstructured intermediates.

The hydrophobic core modification analogue of [Nle^{3},Nva^{43}]Ub has shown the capacity for folding to the native-like fold following the redistribution of methyl groups within the core notwithstanding the deletion of one methyl group on the surface of the protein. Development of a synthetic protocol to enable access to a wide variety of hydrophobic α–amino acids would allow the synthesis of various hydrophobic substitutions to probe the importance of hydrophobicity as the main driving force for protein folding.

Thus, an endless array of modifications are available for use in SPPS to investigate folding or biological studies in proteins. The design of these substitutions in model proteins is ultimately limited by the boundaries of our imagination.

Chapter 3. Experimental

Section 3.1. Notes.

9-Fluorenylmethoxycarbonyl (Fmoc) protected amino acids were purchased from either Bachem or Novabiochem. ^{15}N-Glycine was obtained from Sigma and Fmoc protected using the method of Shute and Rich.[122] (2S,4S)-5-Fluoroleucine (FLeu) was synthesised by B.A. Starkmann, University of Sussex, UK.[123] All amino acids used were of the L-configuration unless otherwise stated. Bachem supplied the 4-alkoxybenzylalcohol functionalised polystyrene resin (Wang). Peptide synthesis grade dimethylformamide (DMF), 1-4-dioxan and piperidine were supplied by Rathburn Chemicals. The N-N-diisopropylethylamine (DIEA) and trifluoroacetic acid (TFA) used were peptide synthesis grade and purchased from Applied Biosystems (ABI). 1-Hydroxybenzotriazole (HOBt) was obtained from Aldrich and ethyl-1-hydroxy-1H-1,2,3-triazole-4-carboxylate (HOCt) was synthesised within the research group. (17-Tetrabenzo[*a,c,g,i*]fluorenyl)methylchloroformate (TbfmocCl) used was also synthesised within the research group. Porous graphitised carbon (PGC) was supplied by Shandon Scientific. All other reagents used were AR grade or better and were purchased from either Aldrich, Fluka or Sigma. High performance liquid chromatography (HPLC) was carried out using either an ABI system comprising 2 x 1406A solvent delivery systems, a 1480A injection/mixer and a 783A detector/controller, or a Gilson system comprising 2 x 306 solvent delivery systems, a 811C dynamic mixer, a 805 manometric module, and a 119 UV/VIS detector. HPLC grade acetonitrile (CH_3CN) was supplied by Rathburn Chemicals and HPLC grade TFA was obtained from Fisons. Fast protein liquid chromatography (FPLC) was carried out on a Pharmacia FPLC system with Liquid Chromatography Controller LCC-501 Plus using a Superdex™ 75 HR 10/30 column and the conditions stated in the text. Gel filtration and cation exchange chromatography were carried out using Pharmacia/LKB apparatus comprising 2 x LKB 2138 UVICORDS, a Pharmacia 2132 Microperpex peristaltic pump, an LKB 2112 redirac fraction collector and where appropriate a Pharmacia Gm gradient mixer. Sonication of samples used a Decon

FS300b sonic bath. High and low resolution fast atom bombarbment mass spectra (FAB MS) were measured on a Kratos MS50TC using a combination of 3-nitrobenzylalcohol and thioglycerol as matrix. Matrix assisted laser desorption ionisation mass spectra (MALDI MS) were measured on a Perspective Biosystems Voyager™ Biospectrometry™ Workstation using either α-cyano-4-hydroxycinnamic acid or 3,5-dimethoxy-4-hydroxycinnamic acid as matrix. Calculated masses were based on average isotope composition and were derived using the program ADDMASS. Amino acid analysis (AAA) were performed on an LKB 4151 amino acid analyser equipped with an LKB 2220 Recording Integrator following hydrolysis of the peptide (0·2 to 0·5mg) in 6M HCl (containing $5mgml^{-1}$ Na_2SO_3) at 110°C in sealed Carius tubes at the times indicated in the text. Proton NMR spectra were recorded on either a Bruker WP200 (200 MHz) or a Varian VXR 5000 (600 MHz) instrument in the solvent indicated and the chemical shifts were measured relative to tetramethylsilane (TMS) assigned to zero. Fluorine NMR spectra were recorded on a Varian VXR 5000 (600 MHz) instrument in the solvent indicated and the chemical shifts were measured relative to $CFCl_3$ assigned to zero. Circular dichroism (CD) spectra were recorded at 20°C on a JASCO J600 spectropolarimeter in the solvents indicated. Fluorescence spectra were recorded on a Perkin Elmer LS50 luminescence spectrometer fitted with a Grant LTD-6 thermostatic control. Melting points were determined using a Koffler hot stage melting point apparatus and are uncorrected. Analytical thin layer chromatography (TLC) used aluminuim sheets precoated with silica gel 60 F_{254} (Merck) in the solvent systems indicated in the text. Visualisation of the compounds was achieved either by ultraviolet (UV) absorption at 254nm or with ninhydrin for compounds with free amino groups. FTIR spectra were recorded on a BIO-RAD FTS-7/SPC 3200 spectrophotometer using polystyrene ($1603cm^{-1}$) as the standard. Elemental analyses were performed on a Perkin-Elmer 2400 CHN elemental analyser. UV spectra were recorded on a Varian Cary 210 spectrophotometer in the solvents indicated. Optical rotations were measured on an AA1000 polarimeter (Optical Activity Ltd) using a 10cm cell in the solvents indicated. Refractive indices were measured on an Abbe 60 refractometer. A Corning ion analyser 150 was used for all pH measurements. Dichloromethane (DCM) was dried

using calcium hydride when required. Ubiquitin(bovine) was purchased from Sigma. Throughout the text ether refers to diethyl ether.

Section 3.2. Solid Phase Peptide Synthesis.

All peptides were synthesised on an ABI 430A automated peptide synthesiser. The efficiency of chain assembly was monitored from measuring the UV absorbance at 302nm of the deprotection solution containing the piperidine-fulvene adduct.[124] An orthogonal synthetic strategy was employed utilising the base labile Fmoc group for N^{α}amino protection and complementary acid labile side-chain protection and acid labile peptide-resin linkage. The following side-chain protecting groups were used: 2,2,5,7,8-pentamethylchroman-6-sulphonyl (Pmc) for the guanidino group of arginine; 4,4'-dimethoxybenzhydryl (Mbh) for asparagine and glutamine; τ-triphenylmethyl (Trt) for histidine; tertiary-butyl (Bu^t) ethers for serine, threonine and tyrosine; Bu^t esters for aspartic and glutamic acid and tertiary-butyloxycarbonyl (Boc) for lysine. The first (C-terminal) amino acid was attached to the 4-alkoxybenzylalcohol functionalised polystyrene resin manually and then transferred to the reaction vessel for automated peptide synthesis. The resin was first swollen in either DMF or DMF/1,4-dioxan (1:1) prior to assembly of the peptide in a stepwise fashion through recurrent capping, deprotection and coupling steps, using one of the two programmed methods summarised below.

Programmed Method 1

1. Capping. The unreacted functional sites are capped using acetic anhydride (0·5M), DIEA (0·125M) and HOBt (0·2% w/v) in DMF (10ml) with vortexing for 10 minutes. The reaction vessel is then drained and washed with six portions of DMF/1,4-dioxan (1:1).

2. Deprotection. The N^{α} Fmoc protecting group is cleaved using 20% v/v piperidine in DMF (10ml) with vortexing for 3 minutes. Deprotection is then repeated for a

second time with vortexing for 1·5 minutes. Finally the resin is washed with six portions of DMF/1,4-dioxan (1:1).

3. Coupling of the next amino acid. Amino acids are coupled either as the symmetrical anhydride or the active ester of HOBt or HOCt. The symmetrical anhydrides are formed from reacting the Fmoc amino acid (2 equivalents) with N-N'-diisopropylcarbodiimide (DIC) (1 equivalent) in DMF/1,4-dioxan (1:1) (8ml). The active esters are formed from reaction of the Fmoc amino acid (1 equivalent) with either HOBt or HOCt (1 equivalent) and DIC (1 equivalent) in DMF/1,4-dioxan (1:1) (8ml). After vortexing for 30 minutes (standard cycle), the reaction vessel is drained and washed with six portions of DMF/1,4-dioxan (1:1).

Programmed Method 2

1. Capping. The unreacted functional sites are capped using acetic anhydride (0·5M), DIEA (0·125M) and HOBt (0·2% w/v) in DMF/1,4-dioxan (1:1) (10ml) with vortexing for 10 minutes. The reaction vessel is then drained and washed with six portions of DMF/1,4-dioxan (1:1).

2. Deprotection. The N^{α} Fmoc protecting group is cleaved using 20% v/v piperidine in DMF/1,4-dioxan (1:1) (10ml) with vortexing for 6 minutes. Two further deprotections are repeated to ensure complete removal of the Fmoc group with vortexing for 1·5 minutes each. Finally the resin is washed with six portions of DMF/1,4-dioxan (1:1).

3. Coupling of the next amino acid. All amino acids except His are coupled as the HOCt ester formed from reaction of the Fmoc amino acid (1 equivalent) with HOCt (1 equivalent) and DIC (1 equivalent) in DMF/1,4-dioxan (1:1) (8ml). His is coupled in a similar manner but as the HOBt ester. After vortexing for 30 minutes (standard cycle), the reaction vessel is drained and washed with six portions of DMF/1,4-dioxan (1:1).

Method 1 uses only DMF for the capping and deprotection stages while the coupling of amino acids uses a 1:1 mixture of DMF/1,4-dioxan. This change in solvents is known to effect resin swelling, therefore in method 2 a 1:1 mixture of DMF/1,4-dioxan is used throughout all three stages. However, the Fmoc deprotection is much slower in DMF/1,4-dioxan (1:1) compared to DMF alone and therefore an extra deprotection step is required.[125]

Section 3.3. Analtyical and Protein Structural Analysis.

High Performance Liquid Chromatography (HPLC). Separations were achieved using linear gradients of Buffer B in A (A=0·1% TFA in MQ-grade water, B=0·09% TFA in CH_3CN). Sample absorbance was recorded at either 214nm (amide bonds) or at 364nm (Tbfmoc group) unless otherwise stated. The columns and conditions used are shown below.

Analytical RP HPLC.

Columns : one of the following columns was used:

Column A ABI aquapore RP300 RP silica 7μm spherical silica, 220 x 4·6mm C8

Column B Vydac RP silica 5μm spherical silica, 250 x 4·6mm C8

Column C Vydac RP phase silica 5μm spherical silica, 250 x 4·6mm C18

Column D HICHROM RP base deactivated spherical silica, 250 x 4·6mm C8/C18

Gradients and conditions : all analytical HPLC's used the following gradient in combination with a 2ml loop and a flow rate of 1ml/minute:

Gradient I (t,%B)(0,10)(2,10)(32,90)(34,10) : 1ml/minute

Semi-Preparative RP HPLC.

Columns : one of the following columns was used:

Column E ABI aquapore RP300 RP silica 7μm spherical silica, 250 x 9·2mm C8

Column F Vydac RP silica 10μm spherical silica, 250 x 22mm C18

Gradients and conditions : one of the following gradients was used in combination with a 5ml loop and a flow rate as described:

Gradient II (t,%B)(0,10)(2,20)(22·5,50)(24·5,90)(26·5,10) : 5ml/minute

Gradient III (t,%B)(0,10)(2,10)(32,90)(34,10) : 10ml/minute

Gradient IV (t,%B)(0,10)(25,70)(27,90)(29,10) : 10ml/minute

Gradient V (t,%B)(0,10)(5,10)(25,60)(27,10) : 9ml/minute

Gradient VI (t,%B)(0,10)(5,10)(35,60)(37,10) : 10ml/minute

Gradient VII (t,%B)(0,10)(2,10)(32,90)(34,10) : 5ml/minute

Gradient VIII (t,%B)(0,10)(5,10)(35,90)(37,10) : 5ml/minute

Circular Dichroism Spectra and Analysis. All samples were prepared from the lyophilised peptide unless otherwise described in the text. Stock protein solutions were usually at concentrations of 0·5mgml^{-1} and prepared in 50mM Na borate buffer pH 7 or MQ-grade water as described in the text. Protein concentrations were determined either by weight (ubiquitin peptides) or from duplicate HPLC integrations using a known concentration (typically 2mgml^{-1}) of purified ubiquitin(bovine) supplied by Sigma as the standard. Additional samples were prepared using aliquots of the stock solution with either 2,2,2-trifluroethanol (TFE) or methanol as described in the text. Samples were either filtered (Sartorius Minisart RC4, 4·5μm) or centrifuged (13000 rpm x 5 minutes) prior to analysis. CD spectra were measured between 190-240nm at 20°C and the secondary structure estimated using the CONTIN procedure.[126]

Fluorescence Studies and Analysis. All samples were prepared from the lyophilised peptide unless otherwise described in the text. Stock protein solutions were usually at concentrations of 100μM and prepared in 50mM 4-morpholinepropanesulphonic acid (MOPS) pH 7 using MQ-grade water. The guanidinium chloride (GdmCl) (≥99%, Life Technologies, Inc) used was typically 7·8M as determined by refractive index measurements and the following equation:[127]

$$\text{Molarity} = \{57{\cdot}147x(\Delta N)\} + \{38{\cdot}68x(\Delta N)^2\} - \{91{\cdot}60x(\Delta N)^3\}$$

where ΔN is the difference in refractive index between the GdmCl and MOPS solutions at the sodium D line.

In general 15-20 data points were measured over the range 0 to 6M GdmCl. For each data point 100µl of stock protein solution was added to the appropriate volume of GdmCl stock solution and made up to 1000µl with MOPS buffer pH 7. The solutions were centrifuged (13000 rpm x 5 minutes) and incubated at 25°C for at least 3 hours prior to fluorescence measurements. Fluorescence emission was monitored at 303nm with excitation at 280nm (emmision/excitation slit width=5nm) in a thermostatted cell at 25°C. The data are presented as relative fluorescence versus concentration of GdmCl.

NMR Spectra and Analysis. Protein solutions were prepared in either MQ-grade water at pH 3·1, 4·8, 4·9 or 5·1 or in 25mM $C_2D_3O_2Na$ buffer pH 4·8 with concentrations in the range 2 to 6mgml^{-1}. The pH was adjusted when required using 2M NaOH or 2M HCl. All samples were centrifuged (13000 rpm x 5 minutes) prior to analysis with the NMR spectra collected at either 15 or 25°C as stated in the text.

Tryptic Mapping.[128] Protein (2mg) was dissolved in 1ml of buffer containing 50mM Tris/HCl pH 8·2, 6·5M urea and 10mM $CaCl_2$. Trypsin (5% w/w with protein) was added and the solution was incubated at 37°C for 4 hours. Additional trypsin (5% w/w with protein) was added and the digestion allowed to proceed for a further 4 hours. The solution was then acidified to pH 2 with 6M HCl and stored at -20°C until required. The peptide fragments obtained were separated by analytical RP HPLC (Column C or D) and gradient (t,%B)(0,10)(2,10)(50,70)(52·5,100)(57·5,100)(65,10) at 1ml/minute, monitoring at 205nm and using the following buffers:

Buffer A 0·1% TFA ,0·1% TEA in MQ-grade water

Buffer B 0·1% TFA ,0·1% TEA in 40% aqueous CH_3CN

The peptide fragments were collected and lyophilised prior to AAA (24hrs) in the normal procedure.

Section 3.4. Experimental (general).

Coupling of the C-Terminal Amino Acid onto the 4-Alkoxybenzylalcohol Functionalised Polystyrene Resin. Two methods of activation were used to couple the first amino acid onto the resin.

1. Symmetrical Anhydride. Fmoc-Gly-OH (0·57g, 1·92mmol) was dissolved in DMF (10ml) and DIC (150µl, 0·96mmol) added. The solution was sonicated at RT for 15 minutes before being added to 4-alkoxybenzylalcohol functionalised polystyrene resin (1·00g, 0·96mmol) previously swollen in DMF (10ml). A catalytic amount of 4-dimethylaminopyridine (10mg, 0·08mmol) was added and the resin was sonicated at RT for 60 minutes. The funtionalised resin was then separated by filtration, washed with DMF and ether and dried overnight in a desicator. The resin substitution was then determined. This procedure generally gave resin loadings with a functionality of ca 0·20mmolg^{-1}.

2. Acid Chloride. Fmoc-Gly-OH (0·939g, 3·16mmol) was suspended in dry DCM (50ml) and freshly distilled thionyl chloride (2·20ml, 30·16mmol) added dropwise under nitrogen. The mixture was heated under reflux at 50°C for 2 hours after which the DCM and thionyl chloride were removed *in vacuo.* The solid obtained was redissolved in dry DCM and concentrated *in vacuo.* This was repeated to give Fmoc-Gly-Cl : FTIR ν_{max} (DCM) 1810 (acid chloride) and 1698cm^{-1} (urethane). The Fmoc-Gly-Cl obtained was dissolved in dry DCM (10ml) and added to 4-alkoxybenzylalcohol functionalised polystyrene resin (2·00g, 1·60mmol) previously swollen in dry DCM (25ml) and pyridine (3ml). The mixture was sonicated at RT for 90 minutes before the functionalised resin was separated by filtration, washed with DCM and ether and dried in a desicator overnight. The resin substitution was then determined. This procedure generally gave resin loadings with a functionality of ca 0·60mmolg^{-1}.

Determination of Resin Loading. The loading efficiency was determined by sonicating duplicate samples of the dried resin (3 to 5mg) in 20% v/v piperidine in DMF (10ml) for 10 minutes. The UV absorbance between 350 to 290nm of the solutions were then measured and the resin loading calculated by a computer program which utilises the absorbance at 302nm of the piperidine-fulvene adduct ($\varepsilon_{\text{adduct 302nm}}$ = 15400 $dm^3\,mol^{-1}\,cm^{-1}$). The values for the resin loading obtained are listed in the text.

Glycylglycine Content in the Loaded Resin. Fmoc-Gly-resin (20mg) was sonicated in 20% v/v piperidine in DMF (10ml) for 10 minutes. The resin was then separated by filtration and washed with DMF, DCM and ether. Following cleavage with 95% aqueous TFA (10ml) for 1 hour, the solution was filtered, concentrated *in vacuo* and the residue triturated with ether (50ml). The supernatant was decanted off and the solid product dissolved in citric acid buffer pH 2·2 (2ml) for amino acid analysis. The retention time (R_t) of the amino acid in the sample was then compared with glycine (R_t=5·86 minutes) and glycylglycine (R_t=10·46 minutes) standards. No glycylglycine was detected in any Fmoc-Gly-resin prepared.

Determination of the Fmoc Content in the Completed Peptide-Resin. Following completion of peptide assembly the Fmoc protecting group was left on the final (N-terminal) amino acid to allow for quantitative Fmoc deprotection. Duplicate samples of the dried Fmoc-peptide-resin (5-10mg) were sonicated in 20% v/v piperidine in DMF (10ml) for 10 minutes. The UV absorbance between 350 to 290nm of the solutions were then measured and the Fmoc loading calculated by a computer program which utilises the absorbance at 302nm of the piperidine-fulvene adduct ($\varepsilon_{\text{adduct 302nm}}$ = 15400 $dm^3\,mol^{-1}\,cm^{-1}$). The values for the quantitative Fmoc deprotection obtained are listed in the text.

Determination of Tbfmoc Loading in the Resin Bound Peptide. Two methods were used to calculate Tbfmoc loading which was found to be quantitative in all cases.

1. Absorbance at 364nm. Duplicate samples of the dried Tbfmoc-peptide-resin (3-5mg) were sonicated in 20% v/v piperidine in 1,4-dioxan (10ml) for 10 minutes. The UV absorbance between 400 to 350nm of the solutions were then measured and the loading estimated using the equation:[129]

$$\text{Tbfmoc functionality (mmolg}^{-1}) = \frac{0{\cdot}613 \times \text{Absorbance at 364nm}}{\text{mass of Tbfmoc-peptide-resin (mg)}}$$

2. Trial Cleavage of Tbfmoc-peptide-resin. Tbfmoc-peptide-resin (10-15 mg) was cleaved using the conditions stated for each peptide in the text and analytical RP HPLC of the lyophilised peptide provided both a qualitative Tbfmoc loading test and a trial cleavage test.

Fmoc-(2S,4S)-5-Fluoroleucine. (2S,4S)-5-Fluoroleucine hydrochloride (1·00g, 5·39mmol) and sodium carbonate (1·14g, 10·75mmol) were dissolved in water (21ml) and acetone (9ml). Following addition of Fmoc-N-hydroxysuccinimide (1·82g, 5·39mmol), the solution was stirred overnight. Ethyl acetate (200ml) was added and the solution acidified to pH 2 with 6M HCl. The ethyl acetate layer was separated, washed with water (2x80ml), dried ($MgSO_4$) and the solution concentrated *in vacuo* to give a yellow oil. Trituration of this oil with hexane gave the *title compound* as a white solid (1·82g, 4·90mmol, 91%).

mp 145-153°C; CHN : Found C 67·89, H 5·70, N 3·86, $C_{21}H_{22}NO_4F$ requires C 67·91, H 5·97, N 3·77%; TLC : R_f ($CHCl_3$:MeOH:AcOH, 85:10:5) 0·74, R_f (isopropanol:pyridine:AcOH:H_2O, 15:10:3:12) 0·84; FTIR : ν_{max} ($CHBr_3$ mull) 3340 (NH), 3022 (CH aromatic), 2960 (CH aliphatic), 1719 (carboxyl), 1680cm^{-1} (urethane); λ_{max} (DMF) 302nm (ε=6720·0 dm^3 mol^{-1} cm^{-1}); FAB MS : 372·16111 $C_{21}H_{22}NO_4F$ requires 372·15920; ^{1}H NMR (200MHz, d_6-DMSO) : 7·28-7·88 (m, 8H, aromatic), 4·25-4·47 (m, 6H, αCH, δCH_2F, FmocCHCH$_2$), 1·53-1·97 (m, 3H, βCH_2, γCH), 0·89-0·91ppm (d, J=6·60Hz, 3H, δCH_3); ^{19}F NMR (600MHz, d_6-DMSO, ^{1}H coupled) : 223·57-223·77ppm (td, J = 47·6, J = 20·8Hz); $[\alpha]_D^{25}$ = -22·5 (c=0·107, DMF).

Section 3.5. Ubiquitin(synthetic).

The synthesis was performed on a 0·1mmol scale using the functionalised resin Fmoc-Gly-($OCH_2C_6H_4OR$) (0·444g, 0·225mmolg^{-1}) and method 2 of the programmed cycles with online monitoring of the deprotection solution at 302nm. All amino acids were single coupled using the HOCt ester (1mmol) except His which was double coupled as the HOBt ester (1mmol each). The side-chains of amino acids were protected as previously described. Following completion of assembly the N-terminal Fmoc was left on and the resin bound product was capped by sonication for 1 hour in acetic anhydride (0·5M), DIEA (0·125M), and HOBt (0·2% w/v) in DCM (20ml). The peptide-resin was separated by filtration, washed with DMF and 1,4-dioxan and divided into two equal portions which were stored in 1,4-dioxan at 4°C until required. Quantitative Fmoc deprotection gave 0·038mmolg^{-1}, 66% of the theoretical maximum. One portion was purified starting with the Tbfmoc/PGC protocol while the second portion was purified using a non-Tbfmoc/PGC protocol.

3.5.1. Tbfmoc/PGC Protocol. Approximately one half of the peptide-resin was sonicated in 20% v/v piperidine/DMF at RT for 20 minutes and subsequently washed with DMF, DCM and ether and dried under water suction for 30 minutes. The dry resin (950mg, 0·036mmols) was then added to a suspension of TbfmocCl (49·7mg, 0·108mmol) and DIEA (18·9µl, 0·108mmol) in dry DCM (10ml) and sonicated at RT for 3 hours in the dark. The Tbfmoc-peptide-resin was separated by filtration and washed thoroughly with DCM and additionally with ether. The Tbfmoc-peptide-resin was then swollen in phenol (0·75g), 1,2-ethanedithiol (EDT) (2ml) and thioanisole (0·5ml) and triisopropylsilane (TIS) (200µl) for 30 minutes under nitrogen whilst protected from the light. Water (0·5ml) and TFA (10ml) were then added and the cleavage mixture stirred for a further 4 hours. The resin was separated by filtration, washed with TFA (5ml) and the combined filtrate and washings concentrated *in vacuo* to give an oil, which yielded the crude Tbfmoc-peptide on trituration with 2% v/v β-mercaptoethanol in ether (100ml). The ether was decanted off and the crude Tbfmoc-protein was dissolved in 20% aqueous CH_3CN (50ml), diluted with water and lyophilised.

3.5.1.1. Tbfmoc/PGC and Gel Filtration G50 Sephadex. The crude Tbfmoc-protein was dissolved in 6M GdmCl (20ml) and diluted with isopropanol (20ml), and centrifuged to remove a small amount of insoluble material. PGC (HPLC grade, 7μm, 1280mg) was added and the solution was vortexed and then centrifuged. Adsorption of the Tbfmoc-protein was monitored by analytical RP HPLC of the supernatant at 364nm. The PGC was then subjected to the following wash protocol with alternate vortexing and centrifugation:

1. 6M GdmCl/isopropanol (1:1) (80ml)
2. 6M GdmCl/isopropanol (1:1) (80ml)
3. 6M GdmCl/isopropanol (1:1) (80ml)
4. 10% piperidine in 6M GdmCl/isopropanol (1:1) (40ml)
5. 10% piperidine in 6M GdmCl/isopropanol (1:1) (32ml)

The supernatant from washes 4 & 5 were combined, neutralised to pH 4·5 with acetic acid (HOAc) and the isopropanol removed *in vacuo* prior to being loaded onto a Sephadex G50 (fine) column (660 x 30mm) pre-equilibriated with 30% aqueous HOAc. The column was then eluted at 39 mlh^{-1} with monitoring at 254 and 277nm and collecting 20 minute fractions. Protein fractions 10-17 inclusive were combined, diluted with water and lyophilised twice to give **Ub(syn) I** (75 mg).

- **Ub(syn) I**

HPLC (Column A, Gradient I) : R_t 17·6 minutes;

MALDI MS : 8588·7 (Na adduct), $C_{378}H_{630}N_{105}O_{118}S_1$ requires 8565·9 Da;

AAA (40hrs) : Asp_7 7·23, Thr_7 6·51, Ser_3 2·34, Glu_{12} 12·41, Pro_3 3·68, Gly_6 5·55, Ala_2 2·35, Val_4 3·90, Met_1 0·85, Ile_7 5·73, Leu_9 7·38, Tyr_1 0·59, Phe_2 1·98, His_1 0·63, Lys_7 7·10, Arg_4 3·40;

AAA (60hrs) : Asp_7 8·38, Thr_7 6·56, Ser_3 2·82, Glu_{12} 12·03, Pro_3 3·71, Gly_6 5·92, Ala_2 2·23, Val_4 3·67, Met_1 0·71, Ile_7 6·82, Leu_9 6·97, Tyr_1 0·66, Phe_2 2·04, His_1 0·64, Lys_7 6·85, Arg_4 2·58.

CD results:	α-helix (%)	β-sheet (%)	remainder (%)
50mM Na borate pH 7	14 +/- 0·94	47 +/- 0·98	39 +/- 1·70
+50% TFE	44 +/- 0·52	28 +/- 0·75	28 +/- 0·91

3.5.1.2. Refolding from 6M GdmCl, 50mM NH_4OAc, 5mM DTT at pH 4·5. Ub(syn) I (50mg) was dissolved in 6M GdmCl, 50mM NH_4OAc, 5mM DTT at pH 4·5 (40ml) and sequentially dialysed (MWCO=2000, Cellu Sep H1) against nitrogen saturated solutions of 50mM NH_4OAc, 5mM DTT at pH 4·5 (2000ml), 50mM NH_4OAc, 2mM DTT at pH 4·5 (3000ml), 50mM NH_4OAc at pH 4·5 (3000ml) for 24 hours at each step. One half of the protein solution was lyophilised twice to give **Ub(syn) II** (24·8mg). Following removal of 10mg for ^1H NMR, the remainder of Ub(syn) II was purified by FPLC (section 3.5.1.4.). The second half of the refolded protein solution was further purified by cation exchange chromatography (section 3.5.1.3.).

- **Ub(syn) II**

HPLC (Column A, Gradient I) : R_t 17·6 minutes;

MALDI MS : 8588·7 (Na adduct), $C_{378}H_{630}N_{105}O_{118}S_1$ requires 8565·9 Da.

3.5.1.3. Cation Exchange/CM-Sepharose CL6B. The second half of refolded Ub(syn) II was concentrated by ultrafiltration (Amicon, MWCO=3000) from 40ml to 15ml and loaded onto a CM-Sepharose CL6B column (280 x 16mm) pre-equilibriated with 50mM NH_4OAc pH 4·5. A pH gradient of 50mM NH_4OAc pH 4·5 to 50mM NH_4OAc pH 5·5 (500ml) was run at $30mlh^{-1}$ overnight with monitoring at 277 and 226nm. The column was then run isocratically at 50mM NH_4OAc pH 5·5 (60ml) for 2 hours after which a salt gradient of 50mM NH_4OAc pH 5·5 to 0·3M NH_4OAc at pH 5·5 (300ml) for 10 hours was applied to elute the protein from the column. Protein fractions (510-555ml), (555-600ml), (600-645ml), (645-675ml) and (675-705ml) were combined and repeatedly lyophilised to give **Ub(syn) III F1** (0·4mg), **F2** (0·3mg), **F3** (0·6mg), **F4** (2·0mg) and **F5** (0·7mg).

- **Ub(syn) III F1**

HPLC (Column A, Gradient I) : R_t 18·4 minutes;

- **Ub(syn) III F2**

HPLC (Column A, Gradient I) : R_t 17·6 minutes;

- **Ub(syn) III F3**

HPLC (Column A, Gradient I) : R_t 17·6 minutes;

- **Ub(syn) III F4 (used in ^{1}H NMR studies)**

HPLC (Column A, Gradient I) : R_t 17·6 minutes;

MALDI MS : 8588·7 (Na adduct), $C_{378}H_{630}N_{105}O_{118}S_1$ requires 8565·9 Da;

AAA (40hrs) : Asp_7 7·18, Thr_7 5·72, Ser_3 2·36, Glu_{12} 13·96, Pro_3 2·79, Gly_6 5·81, Ala_2 2·25, Val_4 3·81, Met_1 1·07, Ile_7 7·05, Leu_9 8·14, Tyr_1 0·98, Phe_2 2·08, His_1 0·76, Lys_7 6·93, Arg_4 4·31.

- **Ub(syn) III F5**

HPLC (Column A, Gradient I) : R_t 18·4 minutes;

3.5.1.4. FPLC/Superdex™ 75. Ub(syn) II (14mg) was dissolved in 50mM HOAc, 1·0M NaCl pH 4·5 (800µl) and purified by FPLC using a Superdex™ 75 HR 10/30 column eluting with 50mM HOAc, 1·0M NaCl pH 4·5, 200µl loop and monitoring at 280nm. Protein fractions (12·51-14·00ml) and (14·01-15·00ml) were combined and each dialysed (MWCO=2000, CelluSep H1) against nitrogen saturated solutions of 50mM NH_4OAc at pH 4·5 (500ml) x 2 for 24 hours at each step. Repeated lyophilisation gave **Ub(syn) IV F1** (2·4mg) and **Ub(syn) IV F2** (2·6mg). Ub(bovine) eluted in fractions (14·01-15·00ml) using the same column and conditions described above.

- **Ub(syn) IV F1**

HPLC (Column A, Gradient I) : R_t 17.4 minutes;

- **Ub(syn) IV F2 (used in ^{1}H NMR studies)**

HPLC (Column A, Gradient I) : R_t 17·6 minutes;

MALDI MS : 8568·1, $C_{378}H_{630}N_{105}O_{118}S_1$ requires 8565·9 Da;

AAA (40hrs) : Asp_7 7·71, Thr_7 6·98, Ser_3 2·67, Glu_{12} 14·28, Pro_3 3·00, Gly_6 5·91, Ala_2 2·34, Val_4 4·24, Met_1 0·60, Ile_7 6·86, Leu_9 7·96, Tyr_1 0·65, Phe_2 2·20, His_1 0·57, Lys_7 6·75, Arg_4 4·00.

3.5.2. Non-Tbfmoc/PGC Protocol. The second half of peptide-resin was sonicated in 20% v/v piperidine/DMF at RT for 20 minutes and subsequently washed with DMF, DCM and ether and dried under water suction for 30 minutes. The dry resin (950mg, 0·036mmols) was then swollen in phenol (0·75g), EDT (2ml), thioanisole (0·5ml) and TIS (200μl) for 30 minutes under nitrogen. Water (0·5ml) and TFA (10ml) were then added and the cleavage mixture stirred for a further 4 hours. The resin was separated by filtration, washed with TFA (5ml) and the combined filtrate and washings concentrated *in vacuo* to give an oil, which yielded the crude protein on trituration with 2% v/v β-mercaptoethanol in ether (100ml). The ether was decanted off and the crude protein dissolved in 20% aqueous CH_3CN (50ml), diluted with water and lyophilised.

3.5.2.1. Gel Filtration G50 Sephadex and Refolding from 6M GdmCl, 50mM NH_4OAc, 5mM DTT at pH 4·5. The lyophilised protein (from 3.5.2.) was dissolved in 6M GdmCl, 50mM NH_4OAc, 5mM DTT pH 4·5 (20ml) and left stirring at RT overnight. The protein solution was then divided equally and taken through in two batches. Protein solution (10ml) was loaded onto a Sephadex G50 (fine) column (660 x 30mm) pre-equilibriated with 6M GdmCl, 50mM NH_4OAc, 5mM DTT pH 4·5. The column was then eluted at 36mlh^{-1} with monitoring at 277nm and collecting 20 minute fractions. Protein fractions 12-15 inclusive were combined and dialysed (MWCO=2000, CelluSep H1) against nitrogen saturated solutions of 50mM NH_4OAc, 5mM DTT at pH 4·5 (2000ml), 50mM NH_4OAc, 2mM DTT at pH 4·5 (3000ml) and 50mM NH_4OAc at pH 4·5 (3000ml) for 24 hours at each step. The

procedure was then repeated for the second batch. One portion of the refolded protein solution was further purified by cation exchange chromatography (3.5.2.2.) while the second portion was purified by FPLC (3.5.2.3.).

3.5.2.2. Cation Exchange/CM-Sepharose CL6B and Semi-Preparative RP HPLC. One portion of the refolded protein solution was concentrated by ultrafiltration (Amicon, MWCO=3000) from 105ml to 15ml and loaded onto a CM-Sepharose CL6B column (280 x 16mm) pre-equilibriated with 50mM NH_4OAc pH 4·5. A pH gradient of 50mM NH_4OAc at pH 4·5 to 50mM NH_4OAc pH 5·5 (500ml) was run at $30mlh^{-1}$ overnight with monitoring at 277 and 226nm. The column was then run isocratically at 50mM NH_4OAc pH 5·5 (60ml) for 2 hours after which a salt gradient of 50mM NH_4OAc pH 5·5 to 0·3M NH_4OAc at pH 5·5 (300ml) for 10 hours was applied to elute the protein from the column. Protein fractions (661-675ml) were combined and repeatedly lyophilised. Further purification of the synthetic material obtained by semi-preparative RP HPLC (Column E, Gradient II) and subsequent lyophilisation gave **Ub(syn) V** (4·0mg).

- **Ub(syn) V**

HPLC (Column A, Gradient I) : R_t 17·6 minutes;

MALDI MS : 8588·7 (Na adduct), $C_{378}H_{630}N_{105}O_{118}S_1$ requires 8565·9 Da;

AAA (40hrs) : Asp_7 6·78, Thr_7 6·02, Ser_3 2·52, Glu_{12} 11·71, Pro_3 3·27, Gly_6 6·07, Ala_2 2·15, Val_4 4·06, Met_1 0·98, Ile_7 6·70, Leu_9 7·90, Tyr_1 1·00, Phe_2 2·15, His_1 0·77, Lys_7 6·78, Arg_4 2·72.

3.5.2.3. FPLC/Superdex™ 75. The second portion of refolded protein solution was lyophilised and then dissolved in 50mM HoAc, 1·0 M NaCl pH4·5 (1500µl) and purified by FPLC using a Superdex™ 75 HR10/30 column eluting with 50mM HOAc, 1·0M NaCl pH 4·5, 200µl loop and monitoring at 280nm. Protein fractions (14·01-15·00ml) were combined and subsequently dialysed (MWCO=2000,CelluSep H1) against nitrogen saturated solutions of 50mM NH_4OAc at pH 4·5 (1000ml) and 50mM NH_4OAc at pH 4·5 (2000ml) for 24 hours at each step and lyophilised to give

Ub(syn) VI (9·1mg). Ub(bovine) eluted in fractions (14·01-15·00ml) using the same column and conditions described above.

- **Ub(syn) VI**

HPLC (Column A, Gradient I) : R_t 17·6 minutes;

MALDI MS : 8588·7 (Na adduct), $C_{378}H_{630}N_{105}O_{118}S_1$ requires 8565·9 Da;

AAA (40hrs) : Asp_7 7·25, Thr_7 6·59, Ser_3 2·49, Glu_{12} 12·54, Pro_3 2·21, Gly_6 5·62, Ala_2 2·25, Val_4 3·79, Met_1 0·84, Ile_7 6·96, Leu_9 8·21, Tyr_1 0·80, Phe_2 2·05, His_1 1·00, Lys_7 7·13, Arg_4 3·77.

Section 3.6. Ubiquitin-(1-35)-Peptide.

H-Met-Gln-Ile-Phe-Val-Lys-Thr-Leu-Thr-Gly-Lys-Thr-Ile-Thr-Leu-Glu-Val-Glu-Pro-Ser-Asp-Thr-Ile-Glu-Asn-Val-Lys-Ala-Lys-Ile-Gln-Asp-Lys-Glu-Gly-OH

The synthesis was performed on a 0·125mmol scale using the functionalised resin Fmoc-Gly-($OCH_2C_6H_4OR$) (0·306g, 0·406mmolg^{-1}) and method 1 of the programmed cycles with online monitoring of the deprotection solution at 302nm. The side-chains of amino acids were protected as previously described with the exception of Gln/Asn which were incorporated unprotected. Most amino acids were double coupled using the symmetrical anhydride (0·5mmol) and the HOBt ester (0·5mmol). The exceptions to this were Gly which was single coupled using the symmetrical anhydride (1mmol), and Asn/Gln which were double coupled using the HOBt ester (0·5mmol each). Following completion of assembly the N-terminal Fmoc was left on and the resin stored in 1,4-dioxan at 4°C until required. Quantitative Fmoc deprotection gave 0·069mmolg^{-1}, 70% of the theoretical maximum.

3.6.1. Semi-Preparative RP HPLC. Approximately one half of the peptide-resin (400mg) was sonicated in 20% v/v piperidine/DMF at RT for 20 minutes and subsequently washed with DMF, DCM and ether and dried under water suction for 30 minutes. The dried peptide-resin was then cleaved using EDT (1ml), thioanisole (0·5ml), water (0·5ml) and TFA (12ml) with stirring under nitrogen for 4 hours. The

resin was separated by filtration, washed with TFA (5ml) and the combined filtrate and washings concentrated *in vacuo* to give an oil, which yielded crude peptide on trituration with 2% v/v β-mercaptoethanol in ether (100ml). The ether was decanted off and the crude peptide dissolved in 20% aqueous HOAc (20ml), diluted with water and lyophilised. Purification of the crude peptide by semi-preparative RP HPLC (Column F, Gradient III) and lyophilisation gave 78·0mg. Further purification by semi-preparative RP HPLC (Column F, Gradient IV) and repeated lyophilisation gave **Ub(1-35)** (59·1mg).

- **Ub(1-35)**

HPLC : (Column C, Gradient I) : R_t 17·3 minutes;

MALDI MS: 3906·3: FAB MS: 3906·1417, $C_{172}H_{292}N_{43}O_{57}S_1$ requires 3906·5244Da;

AAA (40hrs) : Asp_3 3·08, Thr_5 4·43, Ser_1 0·96, Glu_6 6·59, Pro_1 0·95, Gly_2 2·09, Ala_1 1·12, Val_3 2·91, Met_1 0·70, Ile_4 4·06, Leu_2 2·23, Phe_1 1·00, Lys_5 4·79.

CD results:	α-helix (%)	β-sheet (%)	remainder (%)
H_20 pH 4·6	1 +/- 1·1	43 +/- 1·1	56 +/- 2·0
+25% TFE	19 +/- 1·2	51 +/- 1·2	30 +/- 2·2
+50% TFE	30 +/- 1·1	52 +/- 1·1	17 +/- 2·0
H_20 pH 2	2 +/- 0·3	37 +/- 0·5	60 +/- 0·6
+30% MeOH	8 +/- 0·6	46 +/- 0·7	45 +/- 1·2
+60% MeOH	25 +/- 1·5	62 +/- 1·5	13 +/- 2·7

Section 3.7. Ubiquitin-(36-76)-Peptide.

H-Ile-Pro-Pro-Asp-Gln-Gln-Arg-Leu-Ile-Phe-Ala-Gly-Lys-Gln-Leu-Glu-Asp-Gly-Arg-Thr-Leu-Ser-Asp-Tyr-Asn-Ile-Gln-Lys-Glu-Ser-Thr-Leu-His-Leu-Val-Leu-Arg-Leu-Arg-Gly-Gly-OH

The synthesis was performed on a 0·125mmol scale using the functionalised resin Fmoc-Gly-($OCH_2C_6H_4OR$) (0·196g, 0·635mmolg^{-1}) and using method 1 of the programmed cycles with online monitoring of the deprotection solution at 302nm.

The side-chains of amino acids were protected as previously described. Most amino acids were double coupled using the symmetrical anhydride (0·5mmol) and the HOBt ester (0·5mmol). The exceptions to this were Gly which was single coupled using the symmetrical anhydride (1mmol), His was treble coupled as the HOBt ester (1mmol each) and Asn/Gln were double coupled as the HOBt ester (0·5mmol each). Extended coupling times were used for residues Gln41 to Ile36. Following completion of assembly the N-terminal Fmoc was left on and the resin stored in 1,4-dioxan at 4°C until required. Quantitative Fmoc deprotection gave 0·052mmolg^{-1}, 42% of the theoretical maximum.

3.7.1. Tbfmoc/PGC, Gel Filtration Sephadex G25 and Semi-Preparative RP HPLC. Approximately one half of the peptide-resin (420mg, 0·023mmol) was sonicated in 20% v/v piperidine/DMF at RT for 20 minutes and subsequently washed with DMF, DCM and ether and dried under water suction for 30 minutes. The dried peptide-resin was then added to a suspension of TbfmocCl (30·3mg, 0·066mmol) and DIEA (12µl, 0·066mmol) in DCM (6ml) and sonicated at RT for 3 hours in the dark. The Tbfmoc-peptide-resin was separated by filtration and thoroughly washed with DCM and additionally with ether. The dried Tbfmoc-peptide-resin was cleaved for 4 hours with stirring using phenol (0·75g), EDT (2ml), thioanisole (0·5ml), water (0·5ml) and TFA (10ml) under nitrogen whilst protected from the light. The resin was separated by filtration, washed with TFA (5ml) and the combined filtrate and washings concentrated *in vacuo* to give an oil, which yielded solid Tbfmoc-peptide on trituration with ether (100ml). The ether was decanted off and the Tbfmoc-peptide dissolved in 6M GdmCl (10ml) and diluted with isopropanol (10ml). PGC (HPLC grade, 7µm, 600mg) was added and the solution was vortexed and then centrifuged. Adsorption of the Tbfmoc-protein was monitored by analytical RP HPLC of the supernatant at 364nm. The PGC was then subjected to the following wash protocol with alternate vortexing and centrifugation:

1. 6M GdmCl/isopropanol (1:1) (40ml)
2. 6M GdmCl/isopropanol (1:1) (40ml)
3. 6M GdmCl/isopropanol (1:1) (40ml)

4. 10% piperidine in 6M GdmCl/isopropanol (1:1) (20ml)
5. 10% piperidine in 6M GdmCl/isopropanol (1:1) (20ml)

The supernatant from washes 4 & 5 were combined, neutralised to pH 4·5 with HOAc and the isopropanol removed *in vacuo* prior to being loaded onto a Sephadex G25 (fine) column (800 x 15mm) pre-equilibriated with 30% aqueous HOAc. The column was then eluted with at $48mlh^{-1}$ with monitoring at 226 and 277nm and collecting 20 minute fractions. Protein fractions 9-14 were combined, diluted with water and lyophilised to give 60mg. Further purification by semi-preparative RP HPLC (Column F, Gradient V) and repeated lyophilisation gave **Ub(36-76)** (41·9mg).

- **Ub(36-76)**

HPLC (Column C, Gradient I): R_t 16·6 minutes;

MALDI MS : 4679·4, FAB MS : 4678·5842, $C_{206}H_{341}N_{62}O_{62}$ requires 4678·3517 Da;

AAA (40hrs) : Asp_4 3·93, Thr_2 1·91, Ser_2 1·83, Glu_6 5·89, Pro_2 2·34, Gly_4 4·02, Ala_1 0·95, Val_1 0·99, Ile_3 2·47, Leu_7 7·22, Tyr_1 0·60, Phe_1 0·97, His_1 0·88, Lys_2 2·03, Arg_4 3·56.

CD results:	α-helix (%)	β-sheet (%)	remainder (%)
H_20 pH4·6	5 +/- 0·7	39 +/ -0·8	56 +/- 1·3
+25% TFE	56 +/- 0·75	44 +/- 0·76	$0 +/-7·3x10^{-7}$
+50% TFE	65 +/- 0·96	35 +/- 0·96	$0 +/- 9·3x10^{-7}$
H_20 pH 2	6 +/- 0·65	47 +/- 0·75	46 +/- 1·2
+30% MeOH	12 +/- 1·7	43 +/- 1·7	45 +/- 3·1
+60% MeOH	55 +/- 2·1	40 +/- 2·2	5 +/- 4·0

Section 3.8. Ubiquitin-(68-76), (66-76), (64-76)-Peptides.

Ub (68-76) Ac-His-Leu-Val-Leu-Arg-Leu-Arg-Gly-Gly-OH
Ub (66-76) Ac-Thr-Leu-His-Leu-Val-Leu-Arg-Leu-Arg-Gly-Gly-OH
Ub (64-76) Ac-Glu-Ser-Thr-Leu-His-Leu-Val-Leu-Arg-Leu-Arg-Gly-Gly-OH

Ub(68-76), (66-76),(64-76) contd. The synthesis was started on a 0·25mmol scale using the functionalised resin Fmoc-Gly-($OCH_2C_6H_4OR$) (0·379g, 0·660mmolg^{-1}) and method 1 of the programmed cycles with online monitoring of the deprotection solution at 302nm. The side-chains of amino acids were protected as previously described. All amino acids were treble coupled, once as the symmetrical anhydride (0·5mmol) and twice as the HOBt ester (0·5mmol each). The exceptions to this were Gly which was single coupled as the symmetrical anhydride (1mmol) and His which was treble coupled as the HOBt ester (1mmol each). One third of the resin was removed after the coupling of His 68 and the N-terminus acetylated by sonication at RT for 20 minutes in acetic anhydride (0·5M), DIEA (0·125M), and HOBt (0·2% w/v) in DMF (20ml). The resin was separated by filtration, washed with DMF and ether and dried in a desicator overnight to give Ub(68-76)-resin (300mg). The synthesis was continued until Thr 66 when half of the remaining resin was removed and the N-terminus acetylated in an identical manner to that above giving Ub(66-76)-resin (270mg). Following the coupling of Glu 64 the Fmoc group was removed and the N-terminus acetylated on the peptide synthesiser, the resin was then washed with DMF and ether and dried in a desicator overnight to give Ub(64-76)-resin (316mg). All three resins were stored at -20°C until required.

3.8.1. Semi-Preparative RP HPLC Purification of Ub(68-76). Ub(68-76)-resin (140mg) was swollen in a mixture of EDT (0·5ml) and thioanisole (0·5ml) under nitrogen for 30 minutes. Water (0·5ml) and TFA (10ml) were added and the mixture was stirred for 3 hours. The resin was then separated by filtration, washed with TFA (5ml) and the combined filtrate and washings concentrated *in vacuo* to give an oil, which yielded the crude peptide on trituration with ether (100ml). The ether was decanted off and the crude peptide dissolved in 20% aqueous HOAc (20ml), diluted with water and lyophilised twice to give crude peptide (52mg). Purification by semi-preparative RP HPLC (Column F, Gradient VI) and lyophilisation gave pure **Ub(68-76)** (34mg).

- **Ub(68-76)**

HPLC : (Column A, Gradient I) : R_t 15·5 minutes;

FAB MS : 1062·65265 , $C_{47}H_{84}N_{17}O_{11}$ requires 1062·6536 Da;

AAA (20hrs) : Gly_2 2·26, Val_1 1·03, Leu_3 2·91, His_1 0·89, Arg_2 2·13.

CD results:	α-helix (%)	β-sheet (%)	remainder (%)
H_2O pH 4·6	2 +/- 0·67	62 +/ 0·69	36 +/- 1·2
+25% TFE	6 +/- 0·87	48 +/- 1·0	46 +/- 1·7
+50% TFE	17 +/- 0·4	46 +/- 0·57	37 +/- 0·7
H_2O pH 2	0 +/- $3·7E^{-7}$	61 +/- 0·59	39 +/- 0·59
+30% MeOH	2 +/- 0·91	50 +/- 1·0	48 +/- 1·7
+60% MeOH	9 +/- 0·83	54 +/-0·96	37 +/- 1·6

3.8.2. Semi-Preparative RP HPLC Purification of Ub(66-76). Ub(66-76)-resin (140mg) was purified in an identical manner to that described in section 3.8.1 to give crude peptide (46mg) and pure **Ub(66-76)** (28mg).

- **Ub(66-76)**

HPLC : (Column A, Gradient I) : R_t 17·2 minutes;

FAB MS : 1276·79541, $C_{57}H_{102}N_{19}O_{14}$ requires 1277·5585 Da;

AAA (20hrs) : Thr_1 1·04, Gly_2 2·23, Val_1 1·14, Leu_4 377, His_1 0·88, Arg_2 1·99.

CD results:	α-helix (%)	β-sheet (%)	remainder(%)
H_2O pH 4·6	0 +/- $3·4E^{-7}$	58 +/ 0·55	42 +/- 0·55
+25% TFE	37 +/- 0·64	63 +/- 0·64	0 +/- $6·2E^{-7}$
+50% TFE	36 +/- 1·0	64 +/- 1·0	0 +/- $9·9E^{-7}$
H_2O pH 2	0 +/- $3·0E^{-7}$	60 +/- 0·48	40 +/- 0·48
+30% MeOH	0 +/- 1·0	36 +/- 1·1	64 +/- 1·9
+60% MeOH	14 +/- 0·91	48 +/- 1·0	37 +/- 1·7

3.8.3. Semi-Preparative RP HPLC Purification of Ub(64-76). Ub(64-76)-resin (140mg) was purified in an identical manner to that described in section 3.8.1 to give crude peptide (25mg) and pure **Ub(64-76)** (11mg).

- **Ub(64-76)**

HPLC : (Column A, Gradient I) : R_t 17·2 minutes;

FAB MS : 1492·84850 , $C_{65}H_{114}N_{21}O_{19}$ requires 1493·7521 Da;

AAA (20hrs) : Thr_1 0·98, Ser_1 0·95, Glu_1 1·16, Gly_2 2·17, Val_1 1·01, Leu_4 3·73, His_1 0·89, Arg_2 2·00.

CD results:	α-helix (%)	β-sheet (%)	remainder (%)
H_20 pH 4·6	0 +/- $3·7E^{-7}$	61 +/ 0·52	39 +/- 0·52
+25% TFE	33 +/- 1·6	52 +/-1·6	39 +/- 0·52
+50% TFE	24 +/- 1·9	45 +/- 2·0	31 +/- 3·6
H_20 pH 2	0 +/- $3·5E^{-7}$	66 +/- 0·57	34 +/- 0·57
+30% MeOH	7 +/- 0·81	52 +/- 0·85	41 +/- 1·5
+60% MeOH	20 +/- 1·1	48 +/- 1·1	33 +/- 1·9

Section 3.9. [10,35,47&53-^{15}N-Glycine, 56&67-(2S,4S)-5-fluoroleucine] ubiquitin.

The synthesis was performed on a 0·1mmol scale using the functionalised resin Fmoc-Gly-($OCH_2C_6H_4OR$) (0·246g, 0·406mmolg^{-1}) and method 1 of the programmed cycles with online monitoring of the deprotection solution at 302nm. The side-chains of amino acids were protected as previously described. Most amino acids were double coupled using the symmetrical anhydride (0·5mmol) and the HOBt ester (1mmol). The exceptions to this were, Gly 75 single coupled using the symmetrical anhydride (1mmol), ^{15}N-Gly 10,5,47 & 53 single coupled using the HOBt ester (1mmol), His treble coupled as the HOBt ester (0·5, 1·0, 1·0mmol) and Asn/Gln double coupled as the HOBt ester (0·5, 1·0mmol). FLeu was single coupled as the HOBt ester (0·4mmol) by sonication at RT for 24 hours. Following completion of assembly the

N-terminal Fmoc was left on and the resin stored in 1,4-dioxan at 4°C until required. Quantitative Fmoc deprotection gave 0·021mmolg^{-1}, 34% of the theoretical maximum.

3.9.1. Tbfmoc/PGC and Semi-Preparative RP HPLC. Approximately one quarter of the peptide-resin was sonicated in 20% v/v piperidine/DMF for 20 minutes and subsequently washed with DMF, DCM and ether and dried under water suction for 30 minutes. The dry resin (255mg, 5·4x10^{-3}mmol) was then added to a suspension of TbfmocCl (7·5mg, 0·016mmol) and DIEA (2·8µl, 0·16mmol) in DCM (5ml) and sonicated at RT for 3 hours in the dark. The Tbfmoc-peptide-resin was separated by filtration and washed thoroughly with DCM and additionally with ether. The dried Tbfmoc-peptide-resin was swollen in a mixture of EDT (0·25ml), ethyl methyl sulphide (0·25ml) and thioanisole (0·25ml) for 30 minutes under nitrogen whilst protected from the light. Water (0·25ml) and TFA (9ml) were then added and the cleavage mixture stirred for a further 4 hours. The resin was then separated by filtration, washed with TFA (5ml) and the combined filtrate and washings concentrated *in vacuo* to give an oil, which yielded crude Tbfmoc protein on trituration with 2% v/v β-mercaptoethanol in ether (100ml). The ether was decanted off and the Tbfmoc protein dissolved in 6M GdmCl (5ml) and diluted with isopropanol (5ml) and the solution centrifuged to remove a small amount of insoluble material. PGC (HPLC grade, 7µm, 525mg) was added and the solution was vortexed and then centrifuged. Adsorption of the Tbfmoc-protein was monitored by analytical RP HPLC of the supernatant at 364nm. The PGC was then subjected to the following wash protocol with alternate vortexing and centrifugation:

1. 6M GdmCl/isopropanol (1:1) (40ml)
2. 6M GdmCl/isopropanol (1:1) (40ml)
3. 6M GdmCl/isopropanol (1:1) (40ml)
4. 10% piperidine in 6M GdmCl/isopropanol (1:1) (10ml)
5. 10% piperidine in 6M GdmCl/isopropanol (1:1) (10ml)

The supernatant from washes 4 and 5 were combined, neutralised to pH 4·5 with HOAc and the isopropanol removed *in vacuo*. The protein solution was then desalted

by semi-preparative RP HPLC (Column E, Gradient VII) and lyophilised to give **[FLeu^{56}, FLeu^{67}]Ub I**. Purification of all four portions in a similar manner gave 54·8mg overall.

- **[FLeu^{56}, FLeu^{67}]Ub I**

HPLC (Column A, Gradient I) : Rt=17·6 minutes;

MALDI MS : 8606·7, $C_{378}H_{628}N_{105}O_{118}S_1F_2$ requires 8605·8 Da;

AAA (40hrs) : Asp_7 7·34, Thr_7 6·18, Ser_3 2·76, Glu_{12} 13·21, Pro_3 2·61, Gly_6 6·41, Ala_2 1·85, Val_4 4·09, Met_1 0·83, Ile_7 6·79, Leu_7 7·07, Tyr_1 0·99, Phe_2 2·03, His_1 1·02, Lys_7 6·74, Arg_4 3·95.

Tryptic digestion fragments separated by analytical RP HPLC (Column C, 2ml loop) and characterised by AAA (24hrs):

Fragment	Sequence	R_t (minutes)
T1(30-42)	IQDKEGIPPDQQR	7·2
T2 (7-11)	TLTGK	12·8
T3 (49-54)	QLEDGR	15·8
T4 (55-63)	T†LSDYNIQK	29·9
T5 (43-48)	LIFAGK	37·6
T6 (1-6)	MQIFVK	41·2
T7 (64-72)	EST†LHLVLR	43·4
T8 (12-27)	TITLEVEPSDTIENVK	46·0

†denotes FLeu residue.

3.9.2. Refolding from 8M Urea, 50mM NH_4OAc, 10mM DTT at pH4·5. [FLeu^{56}, FLeu^{67}]Ub (25mg) I was dissolved in 8M urea, 50mM NH_4OAc, 10mM DTT at pH 4·5 (10ml) and stirred at RT for 2 hours. The protein solution was then sequentially dialysed (MWCO=2000, Spectropor) against nitrogen saturated solutions of 6M urea, 50mM NH_4OAc, 10mM DTT at pH 4·5 (800ml), 4M urea, 50mM NH_4OAc, 10mM DTT at pH 4·5 (800ml), 2M urea, 50mM NH_4OAc at pH 4·5 (800ml), 50mM NH_4OAc at pH 4·5 (3000ml) and 50mM NH_4OAc at pH 4·5 (3000ml) for 24 hours at

each step. The protein solution was then lyophilised and used in structural studies. A second batch (25mg) was refolded in a similar manner.

CD results :	α-helix (%)	β-sheet (%)	remainder (%)
50mM Na borate pH 7	11 +/- 0·84	52 +/- 0·88	37 +/- 1·6
+50% TFE	80 +/- 1·0	8 +/- 1·5	11 +/- 1·8

Section 3.10. [50&67-(2S,4S)-5-Fluoroleucine]ubiquitin.

The synthesis was started on a 0·25mmol scale using the functionalised resin Fmoc-Gly-($OCH_2C_6H_4OR$) (1·30g, 0·193mmolg^{-1}) and method 1 of the programmed cycles with online monitoring of the deprotection solution at 302nm. The side-chains of amino acids were protected as previously described. Most amino acids were double coupled using the symmetrical anhydride (0·5mmol) and the HOBt ester (1mmol). The exceptions to this were Gly which was single coupled using the symmetrical anhydride (1mmol), His was treble coupled as the HOBt ester (0·5, 1·0, 1·0mmol respectively) and Asn/Gln were double coupled as the HOBt ester (0·5, 1·0mmol). FLeu 67&50 were single coupled as the HOBt ester (1·0 and 0·5mmol respectively) by sonication at RT for 4 hours with quantitative Fmoc deprotection indicating 100% loading at both these residues. Following the coupling of FLeu67, the peptide-resin was divided equally with one half continuing on for the synthesis of [FLeu^{50},FLeu^{67}]Ub with the remaining half stored in 1,4-dioxan at 4°C prior to being used in the synthesis of [FLeu^{43},FLeu^{67}]Ub. Extended coupling times were used for residues Lys6 to Met1. Following completion of assembly the N-terminal Fmoc was left on and the resin stored in 1,4-dioxan at 4°C until required. Quantitative Fmoc deprotection gave 0·046mmolg^{-1}, 83% of the theoretical maximum. The resin was divided into four equal portions with half purified starting with the Tbfmoc/PGC protocol and the second half purified by a non-Tbfmoc/PGC protocol.

3.10.1. Tbfmoc/PGC Protocol. Approximately one quarter of the peptide-resin was sonicated with 20% v/v piperidine/DMF (20ml) for 20 minutes and subsequently

washed with DMF, DCM, and ether and dried under water suction for 30 minutes. The dry resin (463mg, 0·021mmol) was then added to a suspension of TbfmocCl (28·9mg, 0·063mmol) and DIEA (11μl, 0·063mmol) in DCM (5ml) and sonicated at RT for 3 hours in the dark. The Tbfmoc-protein-resin was separated by filtration and washed thoroughly with DCM and additionally with ether. The dried Tbfmoc-protein-resin was swollen in a mixture of phenol (0·75g), EDT (2ml) and thioanisole (0·5ml) under nitrogen for 30 minutes whilst protected from the light. Water (0·5ml) and TFA (10ml) were then added and the cleavage mixture stirred for a further 4 hours. The resin was separated by filtration, washed with TFA (5ml) and the combined filtrate and washings concentrated *in vacuo* to give an oil, which yielded solid Tbfmoc-protein on trituration with 2% v/v β-mercaptoethanol in ether (100ml). The ether was decanted off and the Tbfmoc-protein dissolved in 6M GdmCl (12·5ml) and diluted with isopropanol (12·5ml) and the solution centrifuged to remove a small amount of insoluble material. PGC (HPLC grade, 7μm, 800mg) was added and the solution was vortexed and then centrifuged. Adsorption of the Tbfmoc-protein was monitored by analytical RP HPLC of the supernatant at 364nm. The PGC was then subjected to the following wash protocol with alternate vortexing and centrifugation:

1. 6M GdmCl/isopropanol (1:1) (40ml)
2. 6M GdmCl/isopropanol (1:1) (40ml)
3. 6M GdmCl/isopropanol (1:1) (40ml)
4. 10% piperidine in 6M GdmCl/isopropanol (1:1) (16ml)
5. 10% piperidine in 6M GdmCl/isopropanol (1:1) (16ml)

The supernatant from washes 4 & 5 were combined, neutralised to pH 4·5 with HOAc and concentrated *in vacuo* prior to being loaded onto a Sephadex G50 (fine) column (660 x 30mm) pre-equilibriated with 30% aqueous HOAc. The column was then eluted at $36mlh^{-1}$ with monitoring at 226 and 277nm and collecting 20 minute fractions. Protein fractions 14-22 inclusive were combined, diluted with water and lyophilised twice. Purification of a second quarter of peptide-resin in an identical manner gave **[FLeu^{50},FLeu^{67}]Ub I** (87·1mg overall).

- **[FLeu^{50},FLeu^{67}]Ub I**

HPLC (Column B, Gradient I) : R_t 18 minutes;

MALDI MS : 8602·2, $C_{378}H_{628}N_{105}O_{118}S_1F_2$ requires 8601·8 Da;

AAA (40hrs) : Asp_7 6·71, Thr_7 5·43, Ser_3 2·05, Glu_{12} 13·42, Pro_3 3·17, Gly_6 6·40, Ala_2 2·12, Val_4 3·40, Met_1 1·01, Ile_7 6·53, Leu_7 6·07, Tyr_1 0·63, Phe_2 1·88, His_1 0·51, Lys_7 6·50, Arg_4 2·75;

AAA (60hrs) : Asp_7 8·43, Thr_7 6·63, Ser_3 2·26, Glu_{12} 11·70, Pro_3 3·45, Gly_6 6·00, Ala_2 1·87, Val_4 3·68, Met_1 0·73, Ile_7 7·09, Leu_7 6·63, Tyr_1 0·67, Phe_2 2·04, His_1 0·55, Lys_7 7·12, Arg_4 3·65;

Tryptic digestion fragments separated by analytical RP HPLC (Column D, 2ml loop) and characterised by AAA (24hrs):

Fragment	Sequence	R_t (minutes)
T1 (30-33)	IQDK	8·8
T2 (73-76)	LRGG	10·4
T3 (49-54)	Q†LEDGR	13·8
T4(7-11)	TLTGK	14·8
T5 (34-42)	EGIPPDQQR	24·6
T6 (55-63)	TLSDYNIQK	32·2
T7 (43-48)	LIFAGK	40·0
T8 (1-6)	MQIFVK	43·0
T9 (64-72)	EST†LHLVLR	44·4
T10 (12-27)	TITLEVEPSDTIENVK	46·0

† denotes FLeu residue.

3.10.1.1. Initial NMR Analysis. [FLeu^{50},FLeu^{67}]Ub I (6·4mg) was dissolved in MQ-grade water (700µl) and D_2O (70µl) added giving pH=5·1. The ^1H NMR spectrum was recorded prior to dialysis (MWCO=2000, Spectropor) against nitrogen saturated solutions of 8M urea, 50mM NH_4OAc, 5mM DTT at pH 4·5 (200ml), 50mM NH_4OAc, 1mM DTT at pH 4·5 (200ml) and finally 1mM DTT (200ml) for 24 hours at each step. Following dialysis the pH of the protein solution was adjusted to 5·1

and ^{1}H NMR recorded. The protein solution was then lyophilised and subsequently dissolved in 25mM $C_2D_3O_2Na$/1mM DTT/0·1 w/v Na azide/10% D_2O pH 4·8 (800µl) and the ^{1}H NMR repeated.

3.10.1.2. FPLC/Superdex™ 75. [FLeu^{50},FLeu^{67}]Ub I (67mg) was dissolved in 6M GdmCl (2ml) and purified by FPLC using a Superdex™ 75 HR 10/30 column eluting with 6M GdmCl, 200µl loop and monitoring at 270 and 280nm. Protein fractions (10·51-11·50ml) and (11·51-13·00ml) were combined and extensively dialysed (MWCO=2000, Spectropor) against water and lyophilised to give **[FLeu^{50},FLeu^{67}]Ub II F1** (5·7mg) and **[FLeu^{50},FLeu^{67}]Ub II F2** (27·3mg). Ub(bovine) eluted in fraction (11·51-12·50ml) using the same column and conditions described above.

- **[FLeu^{50},FLeu^{67}]Ub II F1**

HPLC (Column B, Gradient I) : R_t 18 minutes;

MALDI MS : 8626·0 (Na adduct), $C_{378}H_{628}N_{105}O_{118}S_1F_2$ requires 8601·8 Da;

AAA (40hrs) : Asp_7 5·98, Thr_7 5·69, Ser_3 2·34, Glu_{12} 10·96, Pro_3 3·10, Gly_6 6·74, Ala_2 2·14, Val_4 3·74, Met_1 0·86, Ile_7 5·83, Leu_7 5·68, Tyr_1 0·59, Phe_2 1·78, His_1 0·69, Lys_7 6·13, Arg_4 3·10;

AAA (60hrs) : Asp_7 6·91, Thr_7 5·68, Ser_3 2·11, Glu_{12} 11·20, Pro_3 3·39, Gly_6 6·53, Ala_2 2·02, Val_4 3·93, Met_1 0·85, Ile_7 6·28, Leu_7 5·84, Tyr_1 0·70, Phe_2 1·78, His_1 0·70, Lys_7 6·46, Arg_4 3·08.

- **[FLeu^{50},FLeu^{67}]Ub II F2**

HPLC (Column B, Gradient I) : R_t 18 minutes;

MALDI MS : 8628·1 (Na adduct), $C_{378}H_{628}N_{105}O_{118}S_1F_2$ requires 8601·8 Da;

AAA (40hrs) : Asp_7 6·80, Thr_7 5·20, Ser_3 2·35, Glu_{12} 10·95, Pro_3 3·30, Gly_6 6·44, Ala_2 2·20, Val_4 3·60, Met_1 0·84, Ile_7 6·85, Leu_7 5·13, Tyr_1 0·51, Phe_2 1·90, His_1 0·69, Lys_7 6·36, Arg_4 3·03;

AAA (60hrs) : Asp_7 7·21, Thr_7 5·90, Ser_3 1·98, Glu_{12} 11·70, Pro_3 3·08, Gly_6 6·22, Ala_2 2·05, Val_4 3·39, Met_1 0·75, Ile_7 6·86, Leu_7 6·00, Tyr_1 0·58, Phe_2 2·10, His_1 0·73, Lys_7 6·74, Arg_4 3·43.

CD results:	α-helix (%)	β-sheet (%)	remainder (%)
50mM Na borate pH 7	5 +/- 0·86	55 +/- 0·9	40 +/- 1·6
+50% TFE	15 +/- 1·7	36 +/- 1·8	50 +/- 3·2

3.10.1.3. Cation Exchange/CM-Sepharose CL6B. [FLeu^{50},FLeu^{67}]Ub II F1/F2 (33mg) were dissolved in 8M urea, 50mM NH_4OAc, 5mM DTT at pH 4·5 (50ml) and stirred at RT for 2 hours. The protein solution was then sequentially dialysed (MWCO=2000, Cellu Sep H1) against nitrogen saturated solutions of 50mM NH_4OAc, 5mM DTT at pH 4·5 (2000ml), 50mM NH_4OAc, 2mM DTT at pH 4·5 (3000ml), 50mM NH_4OAc at pH 4·5 (3000ml) for 24 hours at each step. The protein solution was then loaded onto the CM-Sepharose CL6B column (280 x16 mm) pre-equilibriated with 50mM NH_4OAc pH 4·5. A pH gradient of 50mM NH_4OAc at pH 4·5 to 50mM NH_4OAc pH 5·5 (500ml) was run at 30mlh^{-1} overnight with monitoring at 277 and 226nm. The column was then run isocratically at 50mM NH_4OAc pH 5·5 (60ml) for 2 hours after which a salt gradient of 50mM NH_4OAc pH 5·5 to 0·3M NH_4OAc at pH 5·5 (300ml) for 10 hours was applied to elute the protein from the column. Protein fractions (555-585ml), (585-660ml) and (660-705ml) and (705-750ml) were combined and repeatedly lyophilised to give **[FLeu^{50},FLeu^{67}]Ub III F1** (0·4mg), **F2** (0·5mg), **F3** (0·9mg) and **F4** (0·4mg). No further work was done on these fractions.

3.10.2. Non-Tbfmoc/PGC Protocol. Approximately one half of the peptide-resin was sonicated in 20% v/v piperidine/DMF (20ml) for 20 minutes and subsequently washed with DMF, DCM, and ether and dried under water suction for 30 minutes. The dried protein-resin (926mg, 0·042mmol) was then swollen in a mixture of phenol (0·75g), EDT (2ml), thioanisole (0·5ml) and TIS (200μl) under nitrogen for 30 minutes. Water (0·5ml) and TFA (10ml) were then added and the cleavage mixture

stirred for a further 4 hours. The resin was separated by filtration, washed with TFA (5ml) and the combined filtrate and washings concentrated *in vacuo* to give an oil, which yielded crude protein on trituration with 2% v/v β-mercaptoethanol in ether (100ml). The ether was decanted off and the protein dissolved in 20% aqueous CH_3CN (50ml), diluted with water and lyophilised.

3.10.2.1. Gel Filtration Sephadex G50 and Refolding from 8M Urea, 50mM NH_4OAc, 5mM DTT at pH 4·5. The lyophilised protein from above was dissolved in 30% HOAc (35ml) and divided equally and taken through in two batches. Protein solution (17·5ml) was loaded onto a Sephadex G50 (fine) column (660 x 30mm) pre-equilibriated with 30% aqueous HOAc. The column was then eluted at 37·5mlh^{-1} with monitoring at 226 and 277nm and collecting 20 minute fractions. Protein fractions 14-18 inclusive were combined, diluted with water and lyophilised twice. Purification of the second portion in an identical manner gave 263·6mg overall. Protein (260mg) was dissolved in 8M urea, 50mM NH_4OAc, 5mM DTT at pH 4·5 (50ml), stirred at RT for 2 hours and subsequently divided into two equal portions. Each portion was sequentially dialysed (MWCO=2000, Cellu Sep H1) against nitrogen saturated solutions of 50mM NH_4OAc, 5mM DTT at pH 4·5 (2000ml), 50mM NH_4OAc, 2mM DTT at pH 4·5 (3000ml) and 50mM NH_4OAc at pH 4·5 (3000ml) for 24 hours at each step.

3.10.2.2. Cation Exchange/CM-Sepharose CL6B. The combined protein solutions from both dialysis were then loaded onto the CM-Sepharose CL6B column (280 x16 mm) pre-equilibriated with 50mM NH_4OAc pH 4·5. A pH gradient of 50mM NH_4OAc at pH 4·5 to 50mM NH_4OAc pH 5·5 (500ml) was run at 30mlh^{-1} overnight with monitoring at 277 and 226nm. The column was then run isocratically at 50mM NH_4OAc pH 5·5 (60ml) for 2 hours after which a salt gradient of 50mM NH_4OAc pH 5·5 to 0·3M NH_4OAc at pH 5·5 (300ml) for 10 hours was applied to elute the protein from the column. Protein fractions (660-705ml) were combined and repeatedly lyophilised to give **[FLeu^{50},FLeu^{67}]Ub IV** (32·2mg).

- **[FLeu^{50},FLeu^{67}]Ub IV**

HPLC (Column A, Gradient I) : R_t 17·4 minutes;

MALDI MS : 8618·9, $C_{378}H_{628}N_{105}O_{118}S_1F_2$ requires 8601·8 Da;

AAA (40hrs) : Asp_7 6·81, Thr_7 5·73, Ser_3 2·32, Glu_{12} 11·19, Pro_3 2·92, Gly_6 6·36, Ala_2 1·98, Val_4 4·11, Met_1 0·95, Ile_7 6·43, Leu_7 6·83, Tyr_1 0·87, Phe_2 1·90, His_1 0·99, Lys_7 6·66, Arg_4 3·92;

AAA (60hrs) : Asp_7 6·58, Thr_7 5·33, Ser_3 1·95, Glu_{12} 11·19, Pro_3 3·87, Gly_6 6·46, Ala_2 2·01, Val_4 3·54, Met_1 0·68, Ile_7 5·92, Leu_7 6·23, Tyr_1 0·66, Phe_2 1·68, His_1 0·96, Lys_7 6·53, Arg_4 3·76.

3.10.2.3. MMA Reduction and Cation Exchange/CM-Sepharose CL6B. [FLeu^{50},FLeu^{67}]Ub IV (32mg) was dissolved in 10% aqueous HOAc (3ml) and N-methylmercaptoacetamide (MMA) (251µl, 2·85mmol) added giving a 10% w/v solution, which was purged with nitrogen and incubated at 37°C for 18 hours. Additional MMA (125µl, 1·43mmol) was added and the solution was incubated at 37°C for a further 30 hours. The reduction reaction was monitored by analytical RP HPLC at 214nm. The protein solution was then diluted to 60ml with 50mM NH_4OAc at pH 4·5 and applied to the CM-Sepharose CL6B column and eluted with the same conditions as described in 3.10.2.2. This gave **[FLeu^{50},FLeu^{67}]Ub V** (16mg estimated).

- **[FLeu^{50},FLeu^{67}]Ub V**

HPLC (Column A, Gradient I) : R_t 17·6 minutes;

MALDI MS : 8604·8, $C_{378}H_{628}N_{105}O_{118}S_1F_2$ requires 8601·8 Da.

CD results:	α-helix (%)	β-sheet (%)	remainder (%)
50mM Na borate pH 7	20 +/-1·7	30 +/-1·7	50 +/- 3·1
+50% TFE	82 +/- 1·5	0 +/-6·6E^{-7}	18 +/- 1·5

Section 3.11. [43&67-(2S,4S)-5-Fluoroleucine]ubiquitin.

The synthesis was performed on a 0·125mmol scale starting at residue 66 with residues 76 to 67 already coupled to the resin. (section 3.10.) The side-chains of amino acids were protected as previously described. Method 1 of the programmed cycles was used with most amino acids double coupled using the symmetrical anhydride (0·5mmol) and the HOBt ester (1mmol). The exceptions to this were Gly which was single coupled using the symmetrical anhydride (1mmol) and Asn/Gln which were double coupled as the HOBt ester (0·5, 1·0mmol). FLeu43 was coupled as the HOBt ester (0·5mmol) by sonication at RT for 4 hours with quantitative Fmoc deprotection indicating a 100% loading. Following completion of assembly the N-terminal Fmoc was left on and the resin stored in 1,4-dioxan at 4°C until required. Quantitative Fmoc deprotection gave 0·040mmolg^{-1}, 72% of the theoretical maximum. The resin was divided into four equal portions with half purified starting with the Tbfmoc/PGC protocol and the second half purified by a non-Tbfmoc/PGC protocol.

3.11.1. Tbfmoc/PGC Protocol. Approximately one quarter of the peptide-resin was sonicated with 20% v/v piperidine/DMF (20ml) for 20 minutes and subsequently washed with DMF, DCM, and ether and dried under water suction for 30 minutes. The dry resin (475mg, 0·019mmol) was then added to a suspension of TbfmocCl (26·1mg, 0·057mmol) and DIEA (10µl, 0·057mmol) in DCM (5ml) and sonicated at RT for 3 hours in the dark. The Tbfmoc-protein-resin was separated by filtration and washed thoroughly with DCM and additionally with ether. The dried Tbfmoc-protein-resin was swollen in a mixture of phenol (0·75g), EDT (2ml) and thioanisole (0·5ml) under nitrogen for 30 minutes whilst protected from the light. Water (0·5ml) and TFA (10ml) were then added and the cleavage mixture stirred for a further 4 hours. The resin was separated by filtration, washed with TFA (5ml) and the combined filtrate and washings concentrated *in vacuo* to give an oil, which yielded crude Tbfmoc-protein on trituration with 2% v/v β-mercaptoethanol in ether (100ml). The ether was decanted off and the Tbfmoc-protein dissolved in 6M GdmCl (12·5 ml) and diluted with isopropanol (12·5ml). PGC (HPLC grade, 7µm, 800mg) was added

and the solution was vortexed and then centrifuged. Adsorption of the Tbfmoc-protein was monitored by analytical RP HPLC of the supernatant at 364nm. The PGC was then subjected to the following wash protocol with alternate vortexing and centrifugation:

1. 6M GdmCl/isopropanol (1:1) (40ml)
2. 6M GdmCl/isopropanol (1:1) (40ml)
3. 6M GdmCl/isopropanol (1:1) (40ml)
4. 10% piperidine in 6M GdmCl/isopropanol (1:1) (16ml)
5. 10% piperidine in 6M GdmCl/isopropanol (1:1) (16ml)

The supernatant from washes 4 & 5 were combined, neutralised to pH 4·5 with HOAc and the isopropanol removed *in vacuo* prior to being loaded onto a Sephadex G50 (fine) column (660 x 30mm) pre-equilibriated with 30% aqueous HOAc. The column was then eluted at $36mlh^{-1}$ with monitoring at 226 and 277nm and collecting 20 minute fractions. Protein fractions 10-14 inclusive were combined, diluted with water and lyophilised twice. Purification of a second quarter of peptide-resin in an identical manner gave **[FLeu^{43},FLeu^{67}]Ub I** (75·4mg overall).

- **[FLeu^{43},FLeu^{67}]Ub I**

HPLC : (Column B, Gradient I) : R_t 18 minutes;

MALDI MS : 8603·5, $C_{378}H_{628}N_{105}O_{118}S_1F_2$ requires 8601·8 Da;

AAA (40hrs) : Asp_7 6·59, Thr_7 5·99, Ser_3 2·42, Glu_{12} 13·64, Pro_3 3·71, Gly_6 6·32, Ala_2 2·10, Val_4 3·56, Met_1 0·78, Ile_7 6·63, Leu_7 6·15, Tyr_1 0·80, Phe_2 1·86, His_1 0·55, Lys_7 6·58, Arg_4 3·60;

AAA (60hrs) : Asp_7 7·13, Thr_7 5·12, Ser_3 2·62, Glu_{12} 12·77, Pro_3 3·77, Gly_6 6·62, Ala_2 2·36, Val_4 3·52, Met_1 ND, Ile_7 6·16, Leu_7 5·67, Tyr_1 0·58, Phe_2 1·91, His_1 0·66, Lys_7 6·02, Arg_4 3·97.

Tryptic digestion fragments separated by analytical RP HPLC (Column D, 2ml loop) and characterised by AAA (24hrs):

Fragment	Sequence	R_t (minutes)
T1 (30-33)	IQDK	8·8
T2 (73-76)	LRGG	10·4
T3 (7-11)	TLTGK	14·8
T4 (49-54)	QLEDGR	17·0
T5 (34-42)	EGIPPDQQR	24·6
T6 (55-63)	TLSDYNIQK	32·2
T7 (43-48)	†LIFAGK	36·2
T8 (1-6)	MQIFVK	43·0
T9 (64-72)	EST†LHLVLR	44·4
T10 (12-27)	TITLEVEPSDTIENVK	46·0

† denotes FLeu residue.

3.11.1.1. Initial NMR Analysis. [FLeu^{43},FLeu^{67}]Ub I (6·8mg) was dissolved in MQ-grade water (700µl) and D_2O (70µl) added giving pH=5·1. The ^1H NMR spectrum was recorded prior to dialysis (MWCO=2000, Spectropor) against nitrogen saturated solutions of 8M urea, 50mM NH_4OAc, 5mM DTT at pH 4·5 (200ml), 50mM NH_4OAc, 1mM DTT at pH 4·5 (200ml) and finally 1mM DTT (200ml) for 24 hours at each step. Following dialysis the pH of the protein solution was adjusted to 5·1 and ^1H NMR recorded. The protein solution was then lyophilised and subsequently dissolved in 25mM $C_2D_3O_2Na$/1mM DTT/0·1 w/v Na azide/10% D_2O pH 4·8 (800µl) and the ^1H NMR repeated.

3.11.1.2. FPLC/Superdex™ 75. [FLeu^{43},FLeu^{67}]Ub I (60mg) was dissolved in 6M GdmCl (2ml) and purified by FPLC using a Superdex™ 75 HR 10/30 column eluting with 6M GdmCl, 200µl loop and monitoring at 270 and 280nm. Protein fractions (10·01-11·50ml) and (11·51-13·50ml) were combined and extensively dialysed (MWCO=2000, Spectropor) against water and lyophilised to give **[FLeu^{43},FLeu^{67}]Ub II F1** (6·1mg) and **[FLeu^{43},FLeu^{67}]Ub II F2** (43·5mg).

Ub(bovine) eluted in fractions (11·51-12·50ml) using the same column and conditions described above.

- **[FLeu^{43},FLeu^{67}]Ub II F1**

HPLC (Column B, Gradient I) : R_t 18·minutes;

MALDI MS : 8633·8 (K adduct), $C_{378}H_{628}N_{105}O_{118}S_1F_2$ requires 8601·8 Da;

AAA (40hrs) : Asp_7 6·86, Thr_7 5·99, Ser_3 2·38, Glu_{12} 12·01, Pro_3 3·84, Gly_6 6·52, Ala_2 2·09, Val_4 3·32, Met_1 0·32, Ile_7 6·72, Leu_7 6·03, Tyr_1 0·62, Phe_2 1·87, His_1 0·60, Lys_7 63·9, Arg_4 3·53;

AAA (60hrs) : Asp_7 7·96, Thr_7 6·54, Ser_3 2·55, Glu_{12} 13·42, Pro_3 4·18, Gly_6 6·00, Ala_2 1·83, Val_4 3·93, Met_1 0·68, Ile_7 7·29, Leu_7 6·69, Tyr_1 0·67, Phe_2 2·09, His_1 0·67, Lys_7 7·19, Arg_4 4·16.

- **[FLeu^{43},FLeu^{67}]Ub II F2**

HPLC (Column B, Gradient I) : R_t 18 minutes;

MALDI MS : 8642·3 (K adduct), $C_{378}H_{628}N_{105}O_{118}S_1F_2$ requires 8601·8 Da;

AAA (40hrs) : Asp_7 7·13, Thr_7 5·94, Ser_3 2·31, Glu_{12} 11·71, Pro_3 4·02, Gly_6 6·18, Ala_2 2·16, Val_4 3·43, Met_1 ND, Ile_7 6·82, Leu_7 5·63, Tyr_1 0·54, Phe_2 1·84, His_1 0·53, Lys_7 6·66, Arg_4 2·95;

AAA (60hrs) : Asp_7 7·95, Thr_7 6·34, Ser_3 2·23, Glu_{12} 12·37, Pro_3 3·54, Gly_6 5·69, Ala_2 1·96, Val_4 3·79, Met_1 0·62, Ile_7 7·15, Leu_7 6·03, Tyr_1 ND, Phe_2 2·06, His_1 0·61, Lys_7 7·35, Arg_4 3·78.

CD results:	α-helix (%)	β-sheet (%)	remainder (%)
50mM Na borate pH 7	6 +/- 0·63	48 +/- 0·73	47 +/- 1·2
+50% TFE	22 +/- 1·5	35 +/- 1·6	44 +/- 2·8

3.11.1.3. Cation Exchange/CM-Sepharose CL6B. [FLeu^{43},FLeu^{67}]Ub II F1/F2 (47·4mg) were dissolved in 8M urea, 50mM NH_4OAc, 5 mM DTT at pH 4·5 (50ml) and stirred at RT for 2 hours. The protein solution was then sequentially dialysed (MWCO=2000, Cellu Sep H1) against nitrogen saturated solutions of 50mM

NH_4OAc, 5mM DTT at pH 4·5 (2000ml), 50mM NH_4OAc, 2mM DTT at pH 4·5 (3000ml), 50mM NH_4OAc at pH 4·5 (3000ml) for 24 hours at each step. The protein solution was then loaded onto the CM-Sepharose column CL6B (280 x 16mm) pre-equilibriated with 50mM NH_4OAc pH 4·5. A pH gradient of 50mM NH_4OAc at pH 4·5 to 50mM NH_4OAc pH 5·5 (500ml) was run at 30 mlh^{-1} overnight with monitoring at 277 and 226nm. The column was then run isocratically at 50mM NH_4OAc pH 5·5 (60ml) for 2 hours after which a salt gradient of 50mM NH_4OAc pH 5·5 to 0·3M NH_4OAc at pH 5·5 (300ml) for 10 hours was applied to elute the protein from the column. Protein fractions (570-630ml), (630-675ml) and (675-720ml) were combined and repeatedly lyophilised to give **[FLeu^{43},FLeu^{67}]Ub III F1** (4·0mg), **F2** (1·5mg) and **F3** (3·6mg).

- **[FLeu^{43},FLeu^{67}]Ub III F3 (used in ^{1}H NMR studies).**

HPLC (Column A, Gradient I) : R_t 17·4 & 17·6 minutes;

MALDI MS : 8607·6 (Na adduct) and 8832·4 (matrix adduct) $C_{378}H_{628}N_{105}O_{118}S_1F_2$ requires 8601·8 Da;

3.11.2. Non-Tbfmoc/PGC Protocol. Approximately one half of the peptide-resin was sonicated in 20% v/v piperidine/DMF (20ml) for 20 minutes and subsequently washed with DMF, DCM, and ether and dried under water suction for 30 minutes. The dried protein-resin (714mg, 0·029mmol) was then swollen in a mixture of phenol (0·75g), EDT (2ml), thioanisole (0·5ml) and TIS (200µl) under nitrogen for 30 minutes. Water (0·5ml) and TFA (10ml) were then added and the cleavage mixture stirred for a further 4 hours. The resin was separated by filtration, washed with TFA (5ml) and the combined filtrate and washings concentrated *in vacuo* to give an oil, which yielded crude protein on trituration with 2% v/v β-mercaptoethanol in ether (100ml). The ether was decanted off and the crude protein dissolved in 20% aqueous CH_3CN (50ml), diluted with water and lyophilised.

3.11.2.1 Gel Filtration Sephadex G50 and Refolding from 8M Urea, 50mM NH_4OAc, 5mM DTT at pH 4·5. The lyophilised protein from above was dissolved

in 30% aqueous HOAc (35ml) and divided equally and taken through in two batches. Protein solution (17·5ml) was loaded onto a Sephadex G50 (fine) column (660 x 30mm) pre-equilibriated with 30% aqueous HOAc. The column was then eluted at 37·5mlh^{-1} with monitoring at 226 and 277nm and collecting 20 minute fractions. Protein fractions 13-17 inclusive were combined, diluted with water and lyophilised twice. Purification of the second portion in an identical manner gave 175·6mg overall. Protein (175mg) was dissolved in 8M urea, 50mM NH_4OAc, 5mM DTT at pH 4·5 (50ml), stirred at RT for 2 hours and subsequently divided into two equal portions. Each portion was sequentially dialysed (MWCO=2000, Cellu Sep H1) against nitrogen saturated solutions of 50mM NH_4OAc, 5mM DTT at pH 4·5 (2000ml), 50mM NH_4OAc, 2mM DTT at pH 4·5 (3000ml), 50mM NH_4OAc at pH 4·5 (3000ml) for 24 hours at each step.

3.11.2.2 Cation Exchange/CM-Sepharose CL6B. The combined protein solutions from both dialysis were then loaded onto the CM-Sepharose column CL6B column (280 x 16mm) pre-equilibriated with 50mM NH_4OAc pH 4·5. A pH gradient of 50mM NH_4OAc at pH 4·5 to 50mM NH_4OAc pH 5·5 (500ml) was run at 30mlh^{-1} overnight with monitoring at 277 and 226nm. The column was then run isocratically at 50mM NH_4OAc pH 5·5 (60ml) for 2 hours after which a salt gradient of 50mM NH_4OAc pH 5·5 to 0·3M NH_4OAc at pH 5·5 (300ml) for 10 hours was applied to elute the protein from the column. Protein fractions (645-690ml), these were combined and repeatedly lyophilised to give **[FLeu^{43},FLeu^{67}]Ub IV** (11·3mg).

- **[FLeu^{43},FLeu^{67}]Ub IV**

HPLC (Column A, Gradient I): R_t 17·4 minutes;

MALDI MS : 8618·9, $C_{378}H_{628}N_{105}O_{118}S_1F_2$ requires 8601·8 Da;

AAA (40hrs) : Asp_7 6·52, Thr_7 6·01, Ser_3 2·43, Glu_{12} 11·07, Pro_3 2·55, Gly_6 6·61, Ala_2 1·96, Val_4 3·34, Met_1 0·59, Ile_7 6·43, Leu_7 6·67, Tyr_1 0·91, Phe_2 1·73, His_1 1·01, Lys_7 6·43, Arg_4 3·63;

AAA (60hrs) : Asp_7 7·04, Thr_7 5·76, Ser_3 2·36, Glu_{12} 10·71, Pro_3 2·25, Gly_6 6·60, Ala_2 1·88, Val_4 3·33, Met_1 0·57, Ile_7 7·02, Leu_7 6·88, Tyr_1 0·64, Phe_2 2·27, His_1 1·04, Lys_7 6·52, Arg_4 5·04.

3.11.2.3. MMA Reduction and Cation Exchange/CM-Sepharose CL6B. [FLeu^{43},FLeu^{67}]Ub IV (11mg) was dissolved in 10% aqueous HOAc (3ml) and (MMA) (251µl, 2·85mmol) added giving a 10% w/v solution, which was purged with nitrogen and incubated at 37°C for 18 hours. Additional MMA (125µl, 1·43mmol) was added and the solution incubated at 37°C for a further 30 hours. The reduction reaction was monitored by analytical RP HPLC at 214nm. The protein solution was then diluted to 60ml with 50mM NH_4OAc at pH 4·5 and applied to the CM-sepharose CL6B column and eluted with the same conditions as described in 3.11.2.2. This gave **[FLeu^{43},FLeu^{67}]Ub V** (5mg estimated).

- **[FLeu^{43},FLeu^{67}]Ub V**

HPLC (Column A, Gradient I) : R_t 17·6 minutes;

MALDI MS : 8622·2 (Na adduct), $C_{378}H_{628}N_{105}O_{118}S_1F_2$ requires 8601·8 Da.

CD results:	α-helix (%)	β-sheet (%)	remainder (%)
50mM Na borate pH 7	29 +/- 0·79	32 +/- 1·1	39 +/- 1·4
+50% TFE	100 +/-7·7E^{-2}	0 +/- 1·6E^{-7}	0 +/- 5·6E^{-7}

Section 3.12. [3-Norleucine,43-norvaline]ubiquitin.

The synthesis was performed on a 0·125mmol scale using the functionalised resin Fmoc-Gly-($OCH_2C_6H_4OR$) (0·308g, 0·406mmolg^{-1}) and method 2 of the programmed cycles with online monitoring of the deprotection solution at 302nm. The side-chains of amino acids were protected as previously described. Most amino acids were double coupled using the HOCt ester (1·0 and 0·5mmol). The exceptions to this were Gly which was single coupled using the symmetrical anhydride (1mmol) and His which was treble coupled as the HOBt ester (0·5, 1·0, 1·0mmol). Following

completion of assembly the N-terminal Fmoc was left on and the resin stored in 1,4-dioxan at 4°C until required. Quantitative Fmoc deprotection gave 0·042mmolg^{-1}, 67% of the theoretical maximum.

3.12.1. Tbfmoc/PGC and Gel Filtration Sephadex G50. Approximately 40% of the peptide-resin was sonicated with 20% v/v piperidine/DMF for 20 minutes and subsequently washed with DMF, DCM, and ether and dried under water suction for 30 minutes. The dry resin (500mg, 0·021mmol) was then added to a suspension of TbfmocCl (28·9mg, 0·063mmol) and DIEA (11µl, 0·063mmol) in DCM (5ml) and sonicated at RT for 3 hours in the dark. The Tbfmoc-peptide-resin was separated by filtration and washed thoroughly with DCM and additionally with ether. The dried Tbfmoc-peptide-resin was swollen in a mixture of EDT (0·25ml), ethyl methyl sulphide (0·25ml) and thioanisole (0·25ml) for 30 minutes under nitrogen whilst protected from the light. Water (0·25ml) and TFA (9ml) were then added and the cleavage mixture stirred for a further 4 hours. The resin was separated by filtration, washed with TFA (5ml) and the combined filtrate and washings concentrated *in vacuo* to give an oil, which yielded crude Tbfmoc-protein on trituration with 2% v/v β-mercaptoethanol in ether (100ml). The ether was decanted off and the Tbfmoc-protein further washed with ether (50ml). The Tbfmoc-protein was dissolved in 6M GdmCl (15ml) and diluted with isopropanol (15ml) and the solution centrifuged to remove a small amount of insoluble material. PGC (HPLC grade, 7µm, 720mg) was added and the solution was vortexed and then centrifuged. Adsorption of the Tbfmoc protein was monitored by analytical RP HPLC at 364nm. The PGC was then subjected to the following wash protocol with alternate vortexing and centrifugation:

1. 6M GdmCl/isopropanol (1:1) (64ml)
2. 6M GdmCl/isopropanol (1:1) (64ml)
3. 6M GdmCl/isopropanol (1:1) (64ml)
4. 10% piperidine in 6M GdmCl/isopropanol (1:1) (16ml)
5. 10% piperidine in 6M GdmCl/isopropanol (1:1) (24ml)

The supernatant from washes 4 & 5 were combined, neutralised to pH 4·5 with HOAc and the isopropanol removed *in vacuo* prior to being loaded onto a Sephadex G50

(fine) column (800 x 30mm) pre-equilibriated with 30% aqueous HOAc. The column was then eluted at 40mlh^{-1} with monitoring at 226 and 277nm and collecting 15 minute fractions. Protein fractions 14-19 inclusive were combined, diluted with water and lyophilised twice to give **[Nle^{3},Nva^{43}]Ub I** (68·9)mg.

- **[Nle^{3},Nva^{43}]Ub I**

HPLC (Column B, Gradient I) : R_t 18 minutes;

MALDI MS : 8576·6 (Na adduct), $C_{377}H_{628}N_{105}O_{118}S_1$ requires 8551·8260 Da;

‡AAA (40hrs) : Asp_7 6·96, Thr_7 5·65, Ser_3 2·16, Glu_{12} 14·07, Pro_3 2·92, Gly_6 6·17, Ala_2 2·18, Val_4 3·83, Met_1 0·78, Ile_6 5·58, Leu_8 7·94, Tyr_1 0·73, Phe_2 2·02, His_1 0·81, Lys_7 6·60, Arg_4 3·63.

‡Nle/Nva seen at distinctive R_t's but no integration performed.

3.12.2. Semi-Preparative HPLC. Further purification by semi-preparative HPLC (Column E, Gradient VI) and repeated lyophilisation gave **[Nle^{3},Nva^{43}]Ub II** (54·5mg).

- **[Nle^{3},Nva^{43}]Ub II**

HPLC (Column B, Gradient I) : R_t 18 minutes;

MALDI MS : 8552·3, $C_{377}H_{628}N_{105}O_{118}S_1$ requires 8551·8260 Da;

AAA (40hrs) : Asp_7 7·25, Thr_7 6·72, Ser_3 2·66, Glu_{12} 13·81, Pro_3 2·93, Gly_6 6·10, Ala_2 2·23, Val_4 4·89, Met_1 0·70, Ile_6 6·06, Leu_8 7·46, Tyr_1 0·66, Phe_2 2·11, His_1 0·95, Lys_7 6·67, Arg_4 3·94.

Tryptic digestion fragments separated by analytical RP HPLC (Column C, 2ml loop) and characterised by AAA (24hrs):

Fragment	Sequence	R_t (minutes)
T1 (30-33)	IQDK	7·2
T2 (73-76)	LRGG	8·2
T3 (7-11)	TLTGK	12·0
T4 (49-54)	QLEDGR	16·4

T5 (34-42)	EGIPPDQQR	23·2
T6 (55-63)	TLSDYNIQK	34·2
T7 (43-48)	‡NvaIFAGK	36·2
T8 (1-6)	MQ‡NleFVK	47·2
T9 (64-72)	ESTLHLVLR	48·6
T10 (12-24)	TITLEVEPSDTIENVK	49·4

‡Nle/Nva seen at distinctive R_t's but no integration performed.

3.12.3. Refolding from 8M Urea, 50 mM NH_4OAc, 10 mM DTT at pH4·5.

[Nle3,Nva43]Ub II (20mg) was dissolved in 8M urea, 50mM NH_4OAc, 10mM DTT at pH 4·5 (20ml) and stirred at RT for 2 hours. The protein solution was then sequentially dialysed (MWCO=2000, Spectropor) against nitrogen saturated solutions of 6M urea, 50mM NH_4OAc, 10mM DTT at pH 4·5 (800ml), 4M urea, 50mM NH_4OAc, 10mM DTT at pH 4·5 (800ml), 2M urea, 50mM NH_4OAc, 10mM DTT at pH 4·5 (800ml), 50mM NH_4OAc at pH 4·5 (3000ml) and 50mM NH_4OAc at pH 4·5 (3000ml) for 24 hours at each step. The protein solution was stored at 4°C until required. A second portion of [Nle^3,Nva^{43}]Ub II (20mg) was refolded in a similar manner.

3.12.4. CD Analysis. [Nle^3,Nva^{43}]Ub II (2ml) of the refolded protein solution from above was by dialysised (MWCO=2000, Spectropor) against 50mM Na borate pH 7 (200ml) x 2 for 24 hours at each step.

CD results:	α-helix (%)	β-sheet (%)	remainder (%)
50mM Na borate pH 7	14 +/- 1·1	56 +/- 1·2	30 +/- 2·0
+50% TFE	51 +/- 1·9	47 +/- 2·0	2 +/- 3·6

3.12.5. Fluorescence Studies. [Nle^3,Nva^{43}]Ub II (4ml) of the refolded protein solution from above was exchanged by ultrafiltration (MWCO=1000, Filtron) into 50mM MOPS pH 7.

3.12.6. NMR Studies. [Nle^{3},Nva^{43}]Ub II (10ml) of the refolded protein solution from above was exchanged by ultrafiltration (MWCO=1000, Filtron) into 25mM $C_2D_3O_2Na$, 1mM DTT at pH 4·8 and the ^{1}H NMR recorded.

Section 3.13. [68-DHistidine]ubiquitin.

The synthesis was started on a 0·25mmol scale using the functionalised resin Fmoc-Gly-($OCH_2C_6H_4OR$) (0·694g, 0·360mmolg^{-1}) and method 1 of the programmed cycles with online monitoring of the deprotection solution at 302nm. The side-chains of amino acids were protected as previously described. For residues 75-69 inclusive all amino acids were double coupled as the HOCt ester (0·5mmol each) except Gly which was single coupled as the HOCt (1mmol). Following the coupling of residue 69 the resin was divided equally with half the resin continuing on for the synthesis of [DHis68]Ub. For residues 68-1 inclusive, most amino acids were double coupled using the symmetrical anhydride (0·5mmol) and HOBt ester (1·0mmol). The exceptions to this were Gly which was single coupled using the symmetrical anhydride (1mmol), DHis was coupled four times as the HOBt ester (0·5mmol each) and Asn/Gln were double coupled as the HOBt ester (0·5 and 1·0mmol). Following completion of assembly the N-terminal Fmoc was left on and the resin stored in 1,4-dioxan at 4°C until required. Quantitative Fmoc deprotection gave 0·036mmolg^{-1}, 58% of the theoretical maximum.

3.13.1. Trial Timed Cleavage of [DHis68]Ub. Peptide-resin (40mg) was sonicated in 20% v/v piperidine/DMF (20ml) for 20 minutes and subsequently washed with DMF, DCM, and ether and dried. The dried peptide-resin was then swollen in a mixture of anisole (0·25ml), thioanisole (0·25ml) and ethyl methyl sulphide (0·25ml) under nitrogen for 30 minutes. Water (0·25ml) and TFA (4ml) were then added. Aliquots (0·5ml) were removed after 2 & 3 hours and triturated with 2% β-mercaptoethanol in ether (25ml). The ether was decanted off and the protein dissolved in 20% aqueous HOAc, diluted with water and lyophilised. After 4 hours the resin was separated by filtration, washed with TFA (5ml) and the combined filtrate

and washings concentrated *in vacuo* to give an oil, which yielded crude protein on trituration with 2% v/v β-mercaptoethanol in ether (100ml). The ether was decanted off and the protein dissolved in 20% aqueous HOAc, diluted with water and lyophilised. Analytical RP HPLC's of the lyophilised products obtained after 2, 3 and 4 hours cleavage indicated that a 4 hour cleavage was sufficient time to remove side-chain protection and cleave the peptide-resin linkage.

3.13.2. Tbfmoc/PGC and Gel Filtration Sephadex G50 Approximately 25% the peptide-resin was sonicated in 20% v/v piperidine/DMF at RT for 20 minutes and subsequently washed with DMF, DCM, and ether and dried under water suction for 30 minutes. The dry resin (410mg, 0·015mmol) was then added to a suspension of TbfmocCl (20·6mg, 0·045mmol) and DIEA (7·8μl, 0·045mmol) in DCM (2ml) and sonicated at RT for 3 hours in the dark. The Tbfmoc-peptide-resin was separated by filtration and washed thoroughly with DCM and additionally with ether. The Tbfmoc-peptide-resin was swollen in a mixture of anisole (0·5ml), thioanisole (0·5ml) and ethyl methyl sulphide (0·5ml) under nitrogen for 30 minutes whilst protected from the light. Water (0·5ml) and TFA (8ml) were then added and the cleavage mixture stirred for a further 4 hours. The resin was separated by filtration, washed with TFA (5ml) and the combined filtrate and washings concentrated *in vacuo* to give an oil, which yielded crude Tbfmoc-protein on trituration with 2% v/v β-mercaptoethanol in ether (100ml). The ether was decanted off and the Tbfmoc-protein was dissolved in 6M GdmCl (6ml), diluted with isopropanol (6ml) and the solution centrifuged to remove a small amount of insoluble material. PGC (HPLC grade, 7μm, 500mg) was added and the solution was vortexed and then centrifuged. Adsorption of the Tbfmoc-protein was monitored by analytical RP HPLC at 364nm. The PGC was then subjected to the following wash protocol with alternate vortexing and centrifugation:

1. 6M GdmCl/isopropanol (1:1) (24ml)
2. 6M GdmCl/isopropanol (1:1) (24ml)
3. 6M GdmCl/isopropanol (1:1) (24ml)
4. 10% piperidine in 6M GdmCl/isopropanol (1:1) (24ml)
5. 10% piperidine in 6M GdmCl/isopropanol (1:1) (24ml)

6. 10% piperidine in 6M GdmCl/isopropanol (1:1) (24ml)

The supernatant from washes 4, 5 & 6 were combined, neutralised to pH 4·5 with HOAc and the isopropanol removed *in vacuo* prior to being loaded onto a Sephadex G50 (fine) column (800 x 30mm) pre-equilibriated with 30% aqueous HOAc. The column was then eluted at 40ml h^{-1} with monitoring at 226 and 277nm and collecting 15 minute fractions. Protein fractions 15-22 inclusive were combined, diluted with water and lyophilised twice to give **[DHis^{68}]Ub I** (36mg). Purification of a second batch of peptide-resin (270mg) in a similar manner gave an additional 30mg.

- **[DHis^{68}]Ub I**

HPLC (Column B, Gradient I) : R_t 18.0 minutes;

MALDI MS : 8567·8, $C_{378}H_{630}N_{105}O_{118}S_1$ requires 8565·6 Da;

AAA (40hrs) : Asp_7 6·99, Thr_7 6·22, Ser_3 2·28, Glu_{12} 14·27, Pro_3 3·25, Gly_6 5·97, Ala_2 2·09, Val_4 4·04, Met_1 0·94, Ile_7 6·52, Leu_9 8·88, Tyr_1 0·56, Phe_2 2·10, His_1 0·91, Lys_7 6·94, Arg_4 3·25.

Tryptic digestion fragments separated by analytical RP HPLC (Column C, 2ml loop) and characterised by AAA (24hrs):

Fragment	Sequence	R_t (minutes)
T1 (30-33)	IQDK	5·4
T2 (73-76)	LRGG	6·0
T3 (7-11)	TLTGK	9·0
T4 (49-54)	QLEDGR	11·4
T5 (34-42)	EGIPPDQQR	17·2
T6 (55-63)	TLSDYNIQK	28·6
T7 (43-48)	LIFAGK	34·2
T8 (1-6)	MQIFVK	38·0
T9 (64-72)	ESTL$^{\#}$HLVLR	39·8
T10 (12-24)	TITLEVEPSDTIENVK	43·4

$^{\#}$ denotes D-histidine.

3.13.3. Refolding from 8M Urea, 50mM NH_4OAc, 10mM DTT at pH 4·5. [DHis^{68}]Ub I (36mg) was dissolved in 8M urea, 50mM NH_4OAc, 10mM DTT at pH 4·5 (20ml) and stirred at RT for 2 hours. The protein solution was then sequentially dialysed (MWCO=2000, Spectropor) against nitrogen saturated solutions of 6M urea, 50mM NH_4OAc, 10mM DTT at pH 4·5 (800ml), 4M urea, 50mM NH_4OAc, 10mM DTT at pH 4·5 (800ml), 2M urea, 50mM NH_4OAc, at pH 4·5 (800ml) and finally 50mM NH_4OAc at pH 4·5 (3000ml) x 3 for 24 hours at each step. Approximately half of the protein solution (10ml) was stored at 4°C until required for NMR analysis, with the remaining half lyophilised and used in the other structural studies. A second portion (30mg) was refolded in a similar manner and subsequently lyophilised.

CD results:	α-helix (%)	β-sheet (%)	remainder(%)
50mM Na borate pH 7	10 +/- 1·1	52 +/ -1·2	38 +/- 2·1
+50% TFE	51 +/- 1·1	22 +/- 1·3	27 +/- 2·1

3.13.4. NMR Studies. [DHis^{68}]Ub (10ml) of the refolded protein solution from above was exchanged by ultrafiltration (MWCO=2000, Amicon) into MQ-grade water, concentrated to ca 1ml, the pH adjusted to 4·8 and the ^1H NMR recorded.

Section 3.14. Miscellaneous Analyses on Ubiquitin(bovine).

3.14.1. Treatment of Ubiquitin(bovine) with 10% v/v Piperidine in 6M GdmCl/isopropanol (1:1). Ub(bovine) (9·5mg) was dissolved in 10% v/v piperidine in 6M GdmCl/isopropanol (1:1) (4ml) and stirred at RT for 40 minutes. The solution was then neutralised to pH 4·5 with HOAc and stored at 4°C overnight. Purification by semi-preparative RP HPLC (Column E, Gradient VIII) and subsequent lyophilisation gave 9·0mg. This was dissolved in MQ-grade water (600µl), D_2O (67µl) added, the pH adjusted to 4·9 and the ^1H NMR recorded.

3.14.2 Oxidation of [Met[1]]Ub(bovine) to [MetS-oxide[1]]Ub(bovine).[111]

Ub(bovine) (5·0mg) was dissolved in 0·2M aqueous HOAc (3ml) and 30% aqueous H_2O_2 (90µl) added. The solution was stirred at RT for 1 hour, diluted with H_2O and lyophilised to give [MetS-oxide[1]]Ub(bovine) (3·5mg).

- **[MetS-oxide[1]]Ub(bovine)**

HPLC (Column A, Gradient I) : R_t 17·4 minutes;

MALDI MS : 8581·2, $C_{378}H_{630}N_{105}O_{119}S_1$ requires 8581·8 Da;

AAA (40hrs) : Asp_7 6·80, Thr_7 6·17, Ser_3 2·49, Glu_{12} 11·61, Pro_3 2·36, Gly_6 5·98, Ala_2 2·28, Val_4 4·18, †Met_1 0·82, Ile_7 6·76, Leu_9 8·29, Tyr_1 0·76, Phe_2 1·89, His_1 0·94, Lys_7 6·74, Arg_4 3·84.

†The $(Na)_2SO_3$ present in the 6M HCl used in the hydrolysis of the protein reduces the methionine sulphoxide to the methionine and hence it is seen in the AAA trace.

3.14.2. CD Analysis on Ub(bovine).

CD results:	α-helix (%)	β-sheet (%)	remainder (%)
H_20 pH 4·6	19 +/- 1·4	56 +/ -1·4	25 +/- 2·5
+25% TFE	32 +/- 1·5	65 +/- 1·6	3 +/- 2·8
+50% TFE	50 +/- 1·9	38 +/- 2·0	11 +/- 3·6
H_20 pH 2	29 +/- 1·2	33 +/- 1·3	38 +/- 2·2
+30% MeOH	24 +/- 1·7	24 +/- 2·0	53 +/- 3·3
+60% MeOH	79 +/- 1·1	21 +/- 1·1	0+/- $1·1x10^{-6}$

4. References

[1] K.A. Dill, *Biochemistry*, 1990, **29**, 7133.

[2] C. Anfinsen, *Science*, 1973, **181**, 223.

[3] P.S. Kim and R.L. Baldwin, *Annu. Rev. Biochem.*, 1990, **59**, 631.

[4] C. Levinthal, *J. Chim. Phys.*, 1968, **65**, 44.

[5] K.A. Dill, *Curr. Opin. Struct. Biol.*, 1993, **3**, 99.

[6] F. Collins and D. Galas, *Science*, 1993, **262**, 43.

[7] G. Goldstein, M. Scheid, U. Hammerling, E.A. Boyse, D.H. Schlesinger and H.D. Niall, *Proc. Natl. Acad. Sci., USA*, 1975, **72**, 11.

[8] S. Wolf, F. Lottspeich and W. Baumeister, *FEBS Lett.*, 1993, **326**,42.

[9] J. Pines, *Nature*, 1994, **371**,742.

[10] M. Glotzer, A.W. Murray and M.W. Kirschner, *Nature*, 1991, **349**, 132.

[11] M. Scheffner, B.A. Werness, J.M. Hulbregtse, A.J. Levine and P.M. Howley, *Cell*, 1990, **63**, 1129.

[12] A. Ciechanover, J.A. Di Giuseppe, B. Bercovich, A. Orian, J.D. Richter, A.L. Schwartz and G.M. Brodeur, *Proc. Natl. Acad. Sci., USA*, 1991, **88**, 139.

[13] R.J. Mayer, J. Arnold, L. László, M. Landon and J. Lowe, *Biochimica et Biophysica Acta*, 1991, **1089**, 141.

[14] S. Jentsch, W. Seufert and H.P. Hauser, *Biochim. Biophys. Acta.*, 1991, **1089**, 127.

[15] S. Vijay-Kumar, C.E. Bugg and W.J. Cook, *J. Mol. Biol.*, 1987, **194**, 531.

[16] D.L. Di Stefano and A.J. Wand, *Biochemistry*, 1987, **26**, 7272.

[17] P.L. Weber , S.C. Brown and L. Mueller, *Biochemistry*, 1987, **26**, 7282.

[18] H. Tsukada and D.M. Blow, *J. Mol. Biol.*, **184**, 703.

[19] R.E. Lenkinski, D.M. Chen, J.D. Glickson and G. Goldstein, *Biochim. Biophys. Acta.*, 1977, **494**, 126.

[20] P.D. Cary, D.S. King, C. Crane-Robinson, E.M. Bradbury, A. Rabbani, G.H. Goodwin and E.W. Johns, *Eur. J. Biochem.*, 1980, **112**, 577.

[21] D.H. Schlesinger, G. Goldstein, and H.D. Niall, *Biochemistry*, 1975, **14**, 2214.

[22] K.D. Wilkinson and T.K. Audhya, *J. Biol. Chem.*, 1981, **256**, 9235.

[23] C.A. Orenga, D.T. Jones and J.M. Thorton, *Nature,* 1994, **372**, 631.

[24] A.M. Gronenborn, D.R. Filpula, N.Z. Essig, A. Achari, M. Whitlow, P.T. Wingfield and G.M. Clore, *Science*, 1991, **253**, 657.

[25] T. Gallagher, P. Alexander, P. Bryan and G.L. Gilliland, *Biochemistry*, 1994, **33**, 4721.

[26]S.D. Emerson, D.S. Waugh, J.E. Scheffler, K-T. Tsao, K.M. Prinzo and D.C. Fry, *Biochemistry*, 1994, **33**, 7745.

[27] N. Nasser, G. Horn, C. Herrmann, A. Scherer, F. McCormick and A. Wittinghofer, *Nature,* 1995, **375**, 554.

[28] A.L. Haas, P. Ahrens, P.M. Bright and H. Ankel, *J. Biol. Chem.*, 1987, **262**, 11315.

[29] K.R. Loeb and A.L. Haas, *J. Biol. Chem., 1992,* **267**, *7806.*

[30] D. Toniolo, M. Perisco, M. Alcalay, *Proc. Natl. Acad. Sci. USA*, 1988, **85**, 851.

[31] J. Banerji, J. Sands, J.L. Strominger and T. Spies, *Proc. Natl. Acad. Sci. USA*, 1990, **87**, 2374.

[32] J.M. Linnen, C.P. Bailey and D.L. Weeks, *Gene,* 1993, **128**, 181.

[33] L.A. Guarino, *Proc. Natl. Acad. Sci. USA*, 1990, **87**, 409.

[34] G. Meyers, N. Tautz, E.J. Dubovi and H.J. Theil, *Virology*, 1991, **180**, 602.

[35] S.W. Englander and L. Mayne, *Annu. Rev. Biophys. Biomol. Struct.*, 1992, **21**, 243.

[36] R.L. Baldwin, *Curr. Opin. Struct. Biol.*, 1993, **3**, 84.

[37] K. Kuwajima, H. Yamaya, S. Miwa, S. Sugai and T. Nagamura, *FEBS Lett.*, 1987, **221**, 115.

[38] G.V. Semisotnov, N.A. Rodionova, V.P. Kutyshenko, B. Ebert, J. Blanck, O.B. Ptitsyn, *FEBS Lett.*, 1987, **224**, 9.

[39] G.V. Semisotnov, N.A. Rodionova, O.I. Razgulyaev and V.N. Uversky, *Biopolymers*, 1991, **31**, 119.

[40] H. Roder and G.A. Elöve in *Mechanisms of Protein Folding,* 1994, p. 26-48 (ed. R.H. Pain), IRL Press, Oxford.

[41] M.S. Briggs and H. Roder, *Proc. Natl. Acad. Sci., USA*, 1992, **89**, 2017.

[42] H. Roder, G.A. Elöve and S.W. Englander, *Nature*, 1988, **335**, 700.

[43] M. Bycroft, A. Matouschek, J.T. Kellis Jr, L. Serrano and A.R. Fersht, *Nature*, 1990, **346**, 488.

[44] S.E. Radford, C.M. Dobson and P.E. Evans, *Nature*, 1992, **358**, 302.

[45] S. Khorasanizadeh, I.D. Peters, T.R. Butt and H. Roder, *Biochemistry*, 1993, **32**, 7054.

[46] J. Jensen, G. Goldstein, E. Breslow, *Biochim. Biophys. Acta.*, 1991, **624**, 378.

[47] C. Tanford, *Adv. Protein Chem.*, 1970, **24**, 1.

[48] S.E. Jackson and A.R. Fersht, *Biochemistry*, 1991, **30**, 10428.

[49] P.L. Privalov, *Adv. Protein Chem.*, 1979, **33**, 167.

[50] K.D. Wilkinson and A.N. Mayer, *Arch. of Biochem. Biophys.*, 1986, **250,** 390.

[51] M.M. Harding, D.H. Williams and D.N. Woolfson, *Biochemistry*, 1991, **30**, 3120.

[52] R.G. Biringer and A.L. Fink, *Biochemistry*, 1982, **21**, 4748.

[53] B.J. Stockman, A. Euvrard and T.A. Scahill, 1993, *J. Biomol. NMR*, **3**, 285.

[54] Y. Pan and M.S. Briggs, *Biochemistry*, 1992, **31**, 11405.

[55] J.P.L. Cox, P.A. Evans, L.C. Packman, D.H. Williams and D.N. Woolfson, *J. Mol. Biol.*, 1993, **234**, 483.

[56] T. Peters Jn and L.K. Davidson, *J. Biol. Chem.*, 1982, **257**, 8847.

[57] D.O.V. Alonso and V. Daggett, *J. Mol. Biol.*, 1995, **247**, 501.

[58] A. Zaks and A.M. Klibanov, *J. Biol. Chem.*, 1988, **263**, 3194.

[59] T. Nakano and A.L. Fink, *J. Biol. Chem.*, 1990, **265**, 12356.

[60] P. Fan, C. Bracken and J. Baum, *Biochemistry*, 1993, **32**, 1573.

[61] R.B. Merrifield, *J. Am. Chem. Soc.*, 1963, **85**, 2149.

[62] R.B. Merrifield, *Angew. Chemie Int. Ed.*, 1985, **24**, 799.

[63] L.A. Carpino and G.Y. Han, *J. Org. Chem.*, 1972, **37**, 3404.

[64] G.B. Fields and R.L. Noble, *Int. J. Pept. Res.* , 1990, **35**, 161.

[65] S.S. Wang, *J. Am. Chem. Soc.*, 1973, **95**, 1328.

[66] A.R. Mitchell, B.W. Erickson, M.N. Ryabtsev, R.S. Hodges and R.B. Merrifield, *J. Am. Chem. Soc.*, 1976, **98**, 7357.

[67] R. Ramage, S.L. Irving and C. McInnes, *Tetrahedron Lett.*, 1993, **34**, 6599.

[68] G.R. Matsueda and J.M. Stewart, *Peptides*, 1981, **2**, 45.

[69] W. König and R. Geiger, *Chem. Ber.*, 1973, **186**, 3626.

[70] L.A. Carpino, *J. Am. Chem. Soc.*, 1993, **115**, 4397.

[71] A. Davidson, PhD Thesis, University of Edinburgh, 1993.

[72] M. Fujino, M. Wakimusa and C. Kitada, *Chem. Pharm. Bull.*, 1981, **29**, 2825.

[73] E. Atherton, R.C. Sheppard and J.D. Wade, *J. Chem. Soc. Chem. Commun.*, 1983, 1060.

[74] R. Ramage and J. Green, *Tetrahedron Lett.*, 1987, **28**, 2287.

[75] J. Green, O. Ogunjobi, R. Ramage and A.S.J. Stewart, *Tetrahedron Lett.*, 1988, **29**, 4341.

[76] A.R. Fletcher, J.H. Jones, W.I. Ramage and A.V. Stachulski, *J. Chem. Soc. Perkin Trans. I,* 1979, 2261.

[77] R. Columbo, *J. Chem. Soc. Chem. Commun.*, 1984, 292.

[78] N. Fujii, A. Otaka, T. Watanabe, A. Okamachi, H. Tamamura, H. Yajima, Y. Inagaki, M. Nomizu and K. Asano, *J. Chem. Soc. Chem. Commun.*, 1989, 283.

[79] D.G. Veber, J.D. Milkowski, S.L. Vargax, D.L. Denkewalter and R. Hirschmann, *J. Am. Chem. Soc.*, 1972, **94**, 5456.

[80] R. Eritja, J.P. Ziehler-Martin, P.A. Walker , T.D. Lee, K. Legesse, F. Alberico and B.E. Kaplan, *Tetrahedron Lett.*, 1987, **43**, 2675.

[81] E. Atherton, R.C. Sheppard and P. Ward, *J. Chem. Soc. Perkin Trans. I, 1985*, 2065.

[82] T.W. Muir and S.B.H. Kent, *Curr. Opin. Biotechnol.* 1993, **4**, 420.

[83] P.E. Dawson, T.W. Muir, I. Clark-Lewis and S.B.H. Kent, *Science*, 1994, **266**, 776.

[84] M. Schölzer and S.B.H. Kent, *Science*, 1993, **256**, 221.

[85] M. Baca and S.B.H. Kent, *Proc. Natl. Acad. Sci. USA, 1993,* **90**, 11638.

[86] P.E. Dawson and S.B.H. Kent, *J. Am. Chem. Soc.*,1993, **115**, 7263.

[87] T.W. Muir, M.J. Williams, M.H. Ginsberg and S.B.H. Kent, *Biochemistry*, 1994, **33**, 7701.

[88] C.F. Lui and J.P. Tam, *Proc. Natl. Acad. Sci. USA, 1994,* **91**, 6584.

[89] K.J. Rose, *J. Am. Chem. Soc.*,1994, **116**, 30.

[90] L.E. Cannes, A.R. Ferré-D'Amare, S.K. Burley and S.B.H. Kent, *J. Am. Chem. Soc.*,1995, **117**, 2998.

[91] K.T. Shaw, Personal Communication.

[92] R. Ramage and G. Raphy, *Tetrahedron Lett.*, 1992, **33**, 385.

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

[94] A.R. Brown, M. Covington, R.C. Newton, R. Ramage and P.Welch, Manuscript in Preparation.

[95] A.R. Brown, S.L. Irving, R. Ramage and G. Raphy, Manuscript in Preparation.

[96] S.L. Irving, PhD Thesis, University of Edinburgh 1993.

[97] S.G. Love, R. Ramage, K.T. Shaw, D. Alexeev and L. Sawyer, Manuscript in Preparation.

[98] S.G. Love, Personal Communication.

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

[100] J.W. Nelson, N.R. Kallenbach, *Proteins: Struct., Funct. Genet.*, 1986, **1**, 211.

[101] H.J. Dyson, G. Merutka, J.P. Waltho, R.A. Lerner and P.E. Wright, *J. Mol. Biol.*, 1992, **226**, 795.

[102] M.L. Tasayco and J. Carey, *Science*, 1992, **255**, 594.

[103] A.D. Kippen, J. Sancho and A.R. Fersht, *Biochemistry*, 1994, **33**, 3778.

[104] G. de Prat Gay, J. Ruiz-Sanz and A.R. Fersht, *Biochemistry*, 1994, **33**, 7964.

[105] N. Kobayashi, S. Honda, H. Yoshii, H. Uedaira, E. Munekata, *FEBS Lett.*, 1995, **366**, 99.

[106] K.L. Kirk, in *Fluorine-containing Amino Acids: Synthesis and Properties*, 1995, p. 343, (ed. V.P. Kukhar & V.A. Soloshonok), John Wiley & Sons Ltd.

[107] J.T.Gerig, *Prog. Nucl. Magn. Reson. Spectrosc.*,1994, **26 4-B**, 293.

[108] I.J. Ropson and C. Frieden, *Proc. Natl. Acad. Sci. USA*, 1992, **89**, 7222.

[109] Synthesis and initial purification of [FLeu^{67}]Ub performed by Dr. A.R. Brown.

[110] B. Whigham, Personal Communication.

[111] S. Bamezai, M.A.T. Banez and E. Breslow, *Biochemistry*, 1990, **29**, 5389.

[112] R.A. Houghton and C.H. Li, *Methods Enzymol.*, **91**, 1983, 549.

[113] S.E. Chambers, E.Y. Lau and J.T. Gerig, *J. Am. Chem. Soc.*, 1994, **116**, 3603.

[114] P.S. Kim and R.L. Baldwin, 1982, *Annu. Rev. Biochem.*, **5**, 459.

[115] P.L. Privalov, S.J. Gill and K.P. Murphy, 1990, *Science*, **250**, 297.

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

[117] D.W. Young, Personal Communication.

[118] D.J. Ecker, T.R. Butt, J. Marsh, E.J. Sternberg, N. Margolis, B.P. Monia, S. Jonnalagadda, M.I. Khan, P.L. Weber, L. Mueller and S.T. Crookes, *J. Biol. Chem.*, 1987, **262**, 14213.

[119] H.M. MacSparron, Personal Communication.

[120] M. Mee, Personal Communication.

[121] D.J. Ecker, T.R. Butt, J. Marsh, E. Sternberg, A. Shatzman, J.S. Dixon, P.L. Weber and S.T. Crookes, *J. Biol. Chem.*, 1987, **264**, 1887.

[122] R.E. Shute and D.H. Rich, *Synthesis*, 1987, 346.

[123] C.M. Moody, B.A. Starkmann and D.W. Young, *Tetrahedron Lett.*, 1994, **35**, 5485-5488.

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

[125] G.J. Cotton, PhD Thesis, University of Edinburgh, 1994.

[126] S.W. Provencher and J. Glöckner, *Biochemistry*, 1981, **20**, 33-37.

[127] C.N. Pace, B.A. Shirley and J.A.Thomson in *Protein Structure-a practical approach,* 1990, p. 311-330, (ed. T.E. Creighton), IRL Press, Oxford.

[128] M.J. Cox, R. Shapira and K.D. Wilkinson, *Anal. Biochem.*, 1986, **154**, 345.

Courses Attended

Departmental Colloquia, University of Edinburgh, 1992-95, various speakers.

Organic Research Seminars, University of Edinburgh, 1992-95, various speakers.

Royal Society of Chemistry Perkin Division Scottish Meeting, Edinburgh 1992, Aberdeen 1993 and Dundee 1994, various speakers: presented a poster (Aberdeen 1993) - "Exploring the Structure of Two Ubiquitin Analogues."

Royal Society of Chemistry Bio-Organic Postgraduate Symposium, University of Exeter, UK, 1993, various speakers: presented a seminar "The Chemical Synthesis, Purification and Structural Characterisation of Ubiquitin Analogues."

SCI Graduate Symposium Novel Organic Chemistry, Heriot-Watt University 1994 and Strathclyde University 1995, various speakers: presented a seminar "A Synthetic Approach to the Study of Protein Folding."

"Medicinal Chemistry", Prof. R. Baker and colleagues, Merck Sharp & Dohme, Terling's Park, UK, 1993-95.

"Chemical Development in the Pharmaceutical Industry", various speakers, SmithKline Beecham, UK, 1993-94.

Solid Phase Synthesis & Complementary Technologies, 3rd International Symposium, Oxford, UK, 1993, various speakers.

The Biochemical Society Meeting No. 651, University of Kent at Canterbury, UK, 1994, various speakers.

EPSRC Graduate School, University of Durham, 1994, various speakers and workshops.

Zeneca Vacation Course, Manchester, 1994, various speakers and workshops.